

---

# Final Product Demo Notes

---

*Authors*

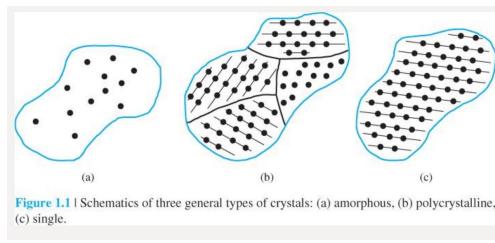
Joshua LEVY

# Lecture 2: Crystal Structure of Solids

---

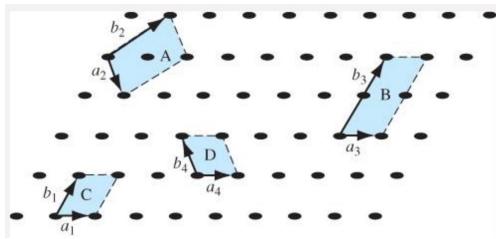
## Topics

- Lattice, basis
- Amorphous, polycrystalline, crystalline
- Cubic unit cells: SC, BCC, FCC
- Silicon crystal structure - diamond
- Crystallographic notation for planes, directions
- Atomic bonding
- Imperfections
- Neamen Chapter 1
- Crystal Lattices:
  - Periodic arrangement of atoms
  - Repeated unit cells (solid-state)
  - Stuffing atoms into unit cells



**Figure 1.1** | Schematics of three general types of crystals: (a) amorphous, (b) polycrystalline, (c) single.

	IIIA	IVA	vA	VIA
	$B^5$	$C^6$	$N^7$	$O^8$
IIB	$A ^{13}$	$Si^{14}$	$P^{15}$	
Zn <sup>30</sup>	Ga <sup>31</sup>	Ge <sup>32</sup>	As <sup>33</sup>	
Cd <sup>48</sup>	In <sup>49</sup>	Sn <sup>50</sup>	Sb <sup>5</sup>	Te <sup>52</sup>



### The periodic lattice

- Unit cell: A small volume of the crystal that can be used to reproduce the entire crystal
- Primitive cell: The smallest unit cell that can be repeated to form the lattice

Figure 1.3 I Two-dimensional representation of a single-crystal lattice showing various possible unit cells.

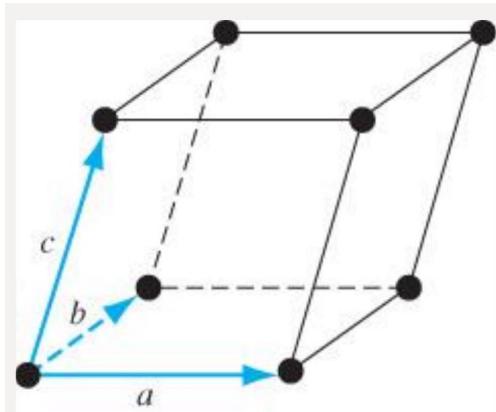


Figure 1.4 I A generalized primitive unit cell.

### Basic crystal structures

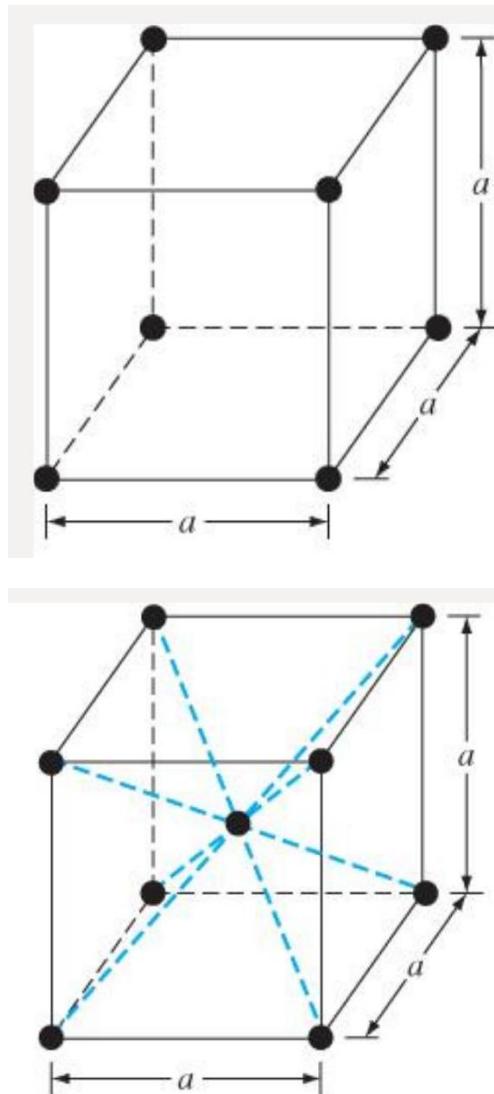
- (a)
- (b)
- (c)

Figure 1.5 Three lattice types: (a) simple cubic, (b) body-centered cubic, (c) face-centered cubic.

- Volume Density =  $\frac{\# \text{ atoms per unit cell}}{\text{volume of unit cell}}$

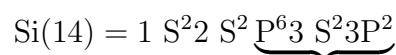
### Volume density of atoms in a crystal

- Exercise:
- The lattice constant of a face centered cubic lattice is 4.25. Determine the
  - (a) effective number of atoms per unit cell
  - (b) volume density of atoms.



### Silicon crystal structure

- Unit cell of silicon crystal is cubic.
- Each Si atom has 4 nearest neighbors.
- Diamond: two fcc lattices interpenetrating, displaced by 1/4 length of body diagonal

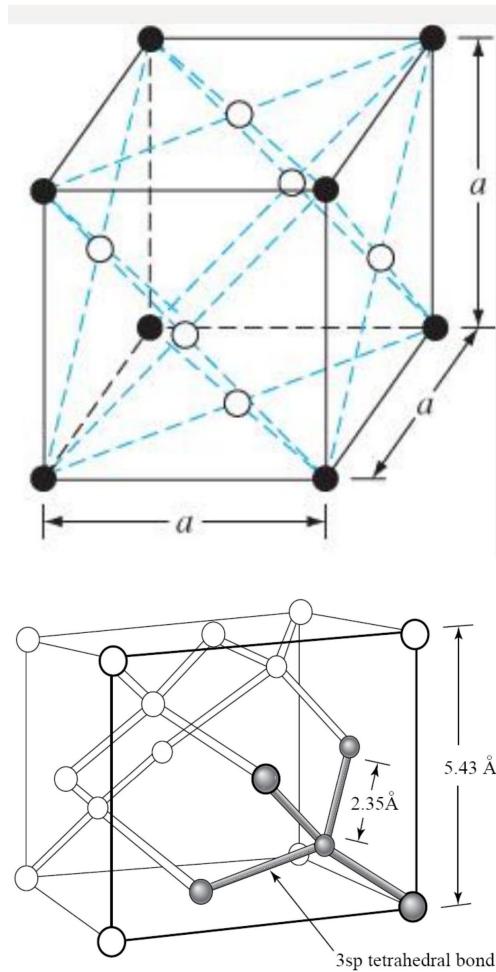


4 valence electrons

Each atom shares 2 electrons with 4 nearest neighbors to form a covalent bond

### How many silicon atoms per cm<sup>3</sup> ?

- Total number of atoms within a unit cell: 8



- Number of atoms completely within the cell: 4
- *Number* of corner atoms (1/8 per cell):  $8 \times 1/8 = 1$
- Number of atoms on the faces ( 1/2 per cell):  $6 \times 1/2 = 3$
- Cell volume:  $(0.543 \text{ nm})^3$

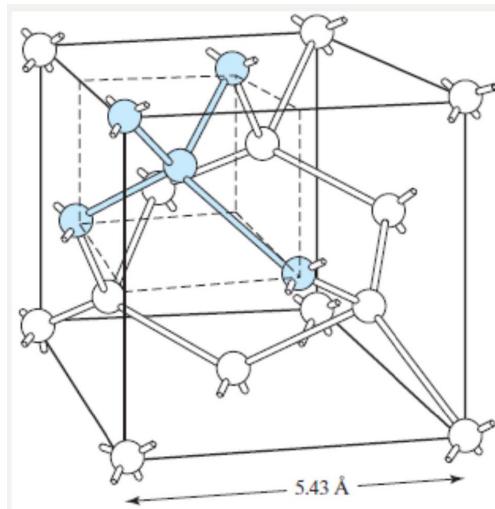
$$= 1.6 \times 10^{-21} \text{ cm}^3$$

- Density of silicon atoms:

$$8/1.6 \times 10^{-21} = 5 \times 10^{22}$$

### Zinc blende lattice (GaAs, AlAs, InP):

- Two intercalated fcc lattices



	IIIA	IVA	VA	VIA
	$B^5$	$C^6$	$N^7$	0
IIB	$Al^{13}$	$Si^{14}$	$P^{11}$	
Zn <sup>3</sup>	$Ga^{31}$	$Ge^{32}$	As	Se
Cd	$In^{49}$	$Sn^{50}$	Sb	Te

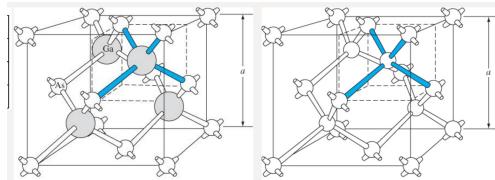


Figure 1.14 | The zincblende (sphalerite) lattice of GaAs.

Figure 1.11 I The diamond structure.

## Crystallographic notation

- Indexing procedure for planes

(1) After setting up coordinate axes along the edges of the unit cell, note where the plane to be indexed intercepts the axes. Divide each intercept value by the unit cell length along the respective coordinate axis. Record the resulting normalized (pure number) intercept set in the order  $x, y, z$ .

(2) Invert the intercept values - that is, form  $1 / \text{intercept}$

(3) Using an appropriate multiplier, convert the  $1 / \text{intercept}$  set to the smallest possible set of whole numbers

(4) Enclose the whole-number set in curvilinear

Notation	Interpretation
$(hkl)$	crystal plane
$\{hkl\}$	equivalent planes
$[hkl]$	crystal direction
$< hkl >$	equivalent directions

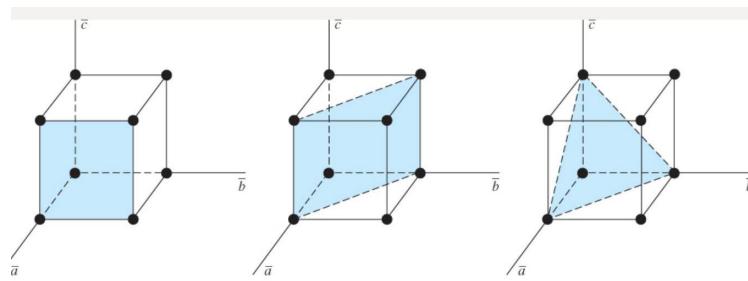
brackets

### Miller indices

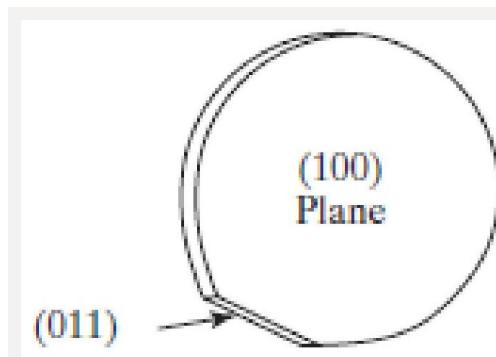
- $(h k l)$  crystal plane
- $\{hkl\}$  equivalent planes

$h k l$  crystal direction

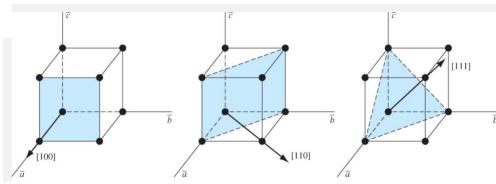
- equivalent direction



Silicon wafers are usually cut along the (100) plane with a flat or notch to help orient the wafer during IC fabrication.



- Flat  
(a)  
(b)  
(c)



### Exercise

- The lattice constant of a FCC structure is 4.25. Calculate the surface density of atoms for a
- (100) plane
- (110) plane

### Atomic bonding

- Covalent
- Ionic
- Metallic
- Van der Waals

### Imperfections and Impurities in solids

- Perfect crystalline solid may appear to be pure academic fiction
- In real semiconductor crystal, there are unavoidable disruptions:
- Vacancies
- Interstitials
- Antisites in compound semiconductors
- Foreign atoms
- At finite temperatures, the atoms in a semiconductor are not standing still but randomly jiggling around
- Finite device size implies there are surfaces that break perfect atomic periodicity
- Doping
- Si crystal: 1 defect in  $10^6$  atoms! How many atoms /cm<sup>3</sup> ?

# Lecture 3: Introduction to Quantum Mechanics

---

## Topics

- Neamen Chapter 2
- What gave rise to quantum?
- Wave particle duality
- Schrödinger's wave equation
- Electron in an infinite quantum well

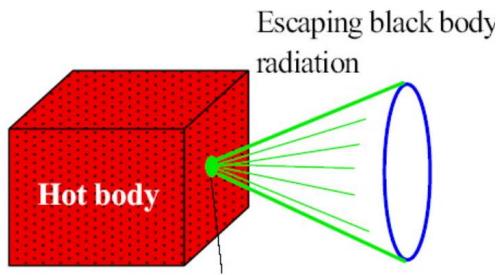
## Black Body Radiation

- Black body: an idealized object that absorbs and emits all frequencies.
- Classical electromagnetism - intensity of a black body radiation as a function of frequency for a fixed temperature:
- $B_\lambda(T) = \frac{2ck_B T}{\lambda^4}$
- Ultraviolet catastrophe!
- Max Planck in 1900 explained black body radiation by assuming that energy was emitted in quanta:
- $E = nh\nu = nh\frac{c}{\lambda}$
- $h = \text{Planck's constant} = 6.626 \times 10^{-34} \text{ J} - \text{s}$

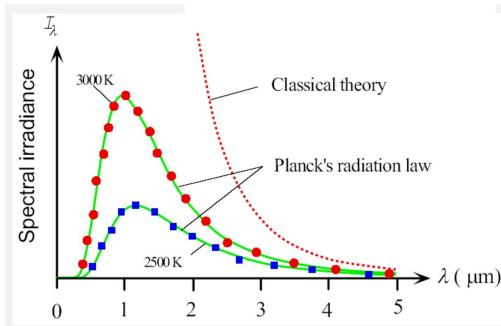
Planck's equation:

$$u(\nu, T) = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{e^{h\nu/kT} - 1}$$

Escaping black body



Small hole acts as a black body



## Max Planck

The Nobel Prize in Physics 1918 was awarded to Max Planck "in recognition of the services he rendered to the advancement of Physics by his discovery of energy quanta".

- Max Planck (1858-1947), a German theoretical physicist, was one of the originators of quantum theory, and won the Nobel Prize in Physics in 1918. His Nobel citation is "in recognition of the services he rendered to the advancement of Physics by his discovery of energy quanta".
- C Alpha Historica/Alamy Stock Photo



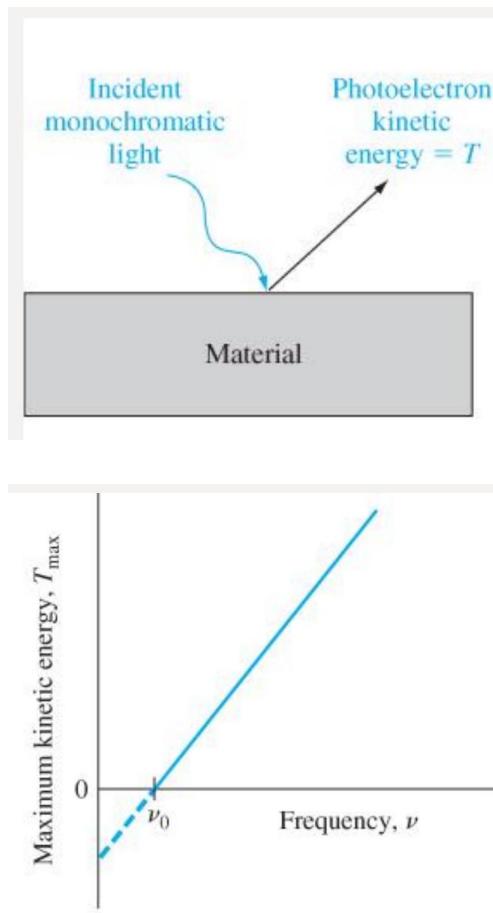
## Photoelectric Effect

(a)

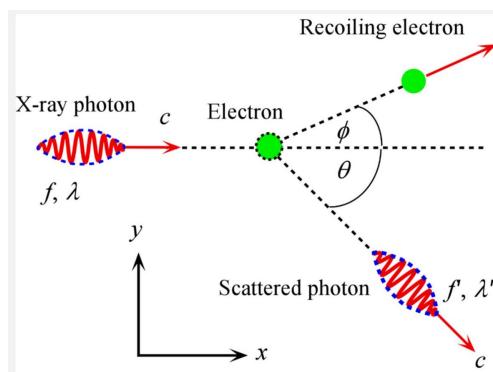
(b)

Figure 2.1 I (a) The photoelectric effect and (b) the maximum kinetic energy of the photoelectron as a function of incident frequency.

- Einstein in 1905 interpreted photoelectric effect by suggesting that energy in a light wave is also contained in discrete packets = photon with  $E = h\nu$
- Maximum kinetic energy of photoelectron
- $T = \frac{1}{2}mv^2 = h\nu - \Phi$
- $\Phi$  = workfunction, minimum energy required to remove an electron from the surface

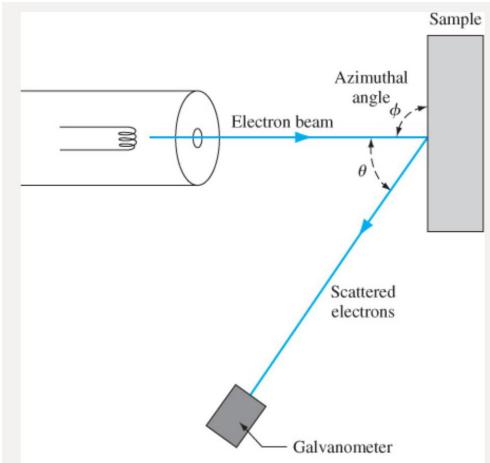


### Compton scattering: particle nature of light



Arthur Holly Compton (1892-1962) at the University of Chicago won the Nobel prize in physics in 1927 for his discovery of the Compton effect with C. T. R. Wilson in 1923. The January 13, 1936 issue of the Time magazine featured Arthur Compton holding a cosmic ray detector.

C) Imagno/Hulton Archive/Getty Images.



### Wave nature of electrons: Davisson and Germer (1927)

Figure 2.2 I Experimental arrangement of the DavissonGermer experiment.

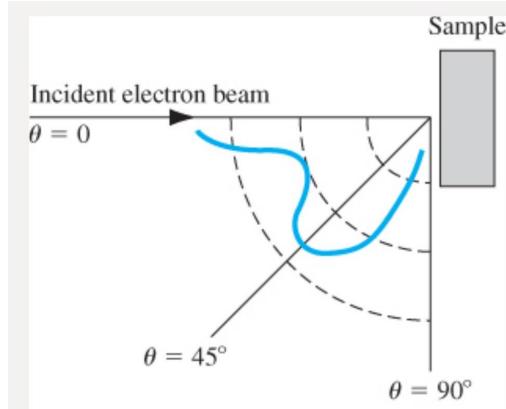
Figure 2.3 I Scattered electron flux as a function of scattering angle for the Davisson-Germer experiment.

### Wave-particle duality de Broglie Relationship

Wavelength  $\lambda$  of the electron depends on its momentum  $p$

$$\lambda = \frac{h}{p}$$

de Broglie relations



$$\lambda = \frac{h}{p} \quad \text{or} \quad p = \frac{h}{\lambda}$$

### Heisenberg's Uncertainty Principle

Heisenberg uncertainty principle for position and momentum

$$\Delta X D_X > \eta_n = \frac{(\hbar k_n)^2}{2m} = \frac{p_x^2}{2m}$$

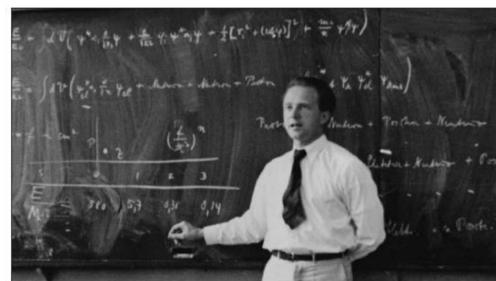
Momentum  $p_x = \pm \hbar k_n$

Heisenberg uncertainty principle for energy and time

$$\Delta E \Delta t \geq \hbar$$

Consequence of uncertainty principle: We cannot determine the exact position of an electron. Instead, we determine the probability of finding an electron at a particular position.

### Werner Heisenberg (1901-1976)



- Werner Heisenberg (1901-1976) received the Nobel prize in physics in 1932 for the uncertainty principle. This photo was apparently taken in 1936, while he was lecturing on quantum mechanics. "An expert is someone who knows some of the worst mistakes that can be made in his subject, and how to avoid them." W. Heisenberg.
- (C) AIP/Science Source

### Schrödinger's wave equation

$$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi(x, t) + V(x)\Psi(x, t) = j\hbar \frac{\partial}{\partial t} \Psi(x, t)$$

### Time-Independent Schrödinger Equation

$$\frac{d^2\psi}{dx^2} + \frac{2m_e}{\hbar^2} [E - V(x)]\psi = 0$$

- $E$  = Energy of the electron
- $\psi(x)$  = Wavefunction of the electron
- $|\psi|^2$  = Probability of finding the electron in  $dx$

### Physical meaning of wave function

- $|\Psi(x, t)|^2$  is the probability density function
- $|\Psi(x, t)|^2 dx$  is the probability of finding the particle between  $x$  and  $x + dx$  at a given time

### Problem 2.21

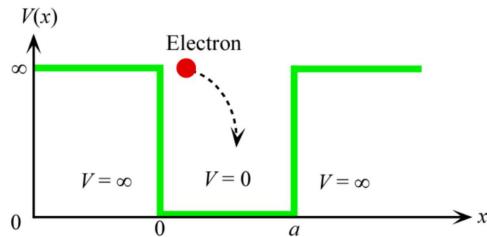
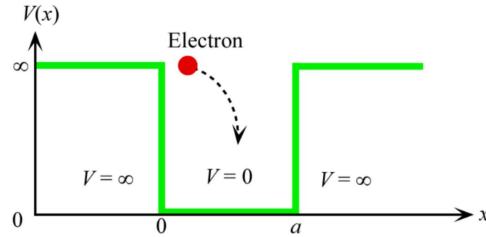
An electron is described by a wave function given by  $\Psi(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{2\pi x}{a}\right)$  for  $-\frac{a}{2} < x < \frac{a}{2}$ . The wave function is zero elsewhere. Calculate the probability of finding the electron between (a)  $0 < x < a/4$ , (b)  $a/4 < x < a/2$ , and (c)  $-a/2 < x < a/2$ .

### Electron in a one-dimensional infinite PE well

#### The electron is confined

- QUESTIONS

- Where is the electron?
- What is the electron's energy?



### Electron in a one-dimensional infinite PE well

At point 0  $\psi(x) = \psi(0) = 0$

Between points 0 and  $a$

At point  $a$   $\psi(x) = \psi(a) = 0$

$$\frac{d^2\psi}{dx^2} + \frac{2m_e}{\hbar^2} E \psi = 0$$

$$\psi(x) = A \exp(+jkx) + B \exp(-jkx)$$

Can also be written as:  $\psi(x) = A_1 \cos kx + B_1 \sin kx$

$$k = \sqrt{\frac{2mE}{\hbar^2}}$$

Electron in a one-dimensional infinite PE well

Boundary condition:  $\psi(x=0) = \psi(x=a) = 0$

$$\psi(x) = A_1 \cos kx + B_1 \sin kx$$

$$\psi(0) = A_1 + B_1 \cdot 0 = 0 \Rightarrow A_1 = 0$$

$$\psi(a) = B_1 \sin ka = 0$$

For nontrivial solutions:  $\sin ka = 0 \Rightarrow ka = n\pi, n = 1, 2, 3 \dots$

$$\therefore k = \frac{n\pi}{a} = k_n \quad k \text{ is quantized}$$

Electron in a one-dimensional infinite PE well

$$B_1 = ? \quad \int_{-\infty}^{\infty} \psi(x) \psi^*(x) dx = 1$$

Assume  $\psi(x)$  is real. Then  $\psi(x) = \psi^*(x)$

$$\begin{aligned} \therefore \int_0^a (B_1 \sin kx)^2 dx &= 1 \Rightarrow B_1^2 \int_0^a \frac{1}{2}(1 - \cos 2kx)dx = 1 \\ \Rightarrow B_1^2 \left[ \frac{1}{2} \cdot a - 0 \right] &= 1 \\ \Rightarrow B_1 &= \sqrt{2/a} \end{aligned}$$

So:  $\psi(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right); n = 1, 2, 3 \dots$

### Confinement leads to quantization

$$k_n = \frac{n\pi}{a}$$

$n = 1, 2, 3 \dots$  a quantum number (QN)  
 $k$  is quantized

$$\psi_n(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right)$$

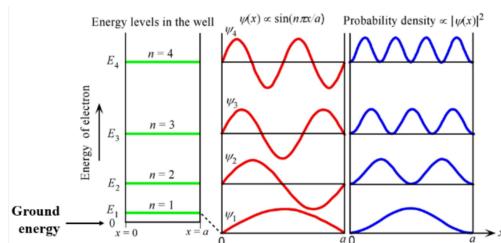
The electron can only have certain wavefunctions: Eigenfunctions

$\psi_n(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right)$	Electron eigenfunctions
$\frac{d^2\psi}{dx^2} + \frac{2m_e}{\hbar^2} E\psi = 0$	Energy of the electron

### Confinement leads to quantization

$$E_n = \frac{\hbar k_n}{2m} = \frac{\hbar^2 n^2 \pi^2}{2ma^2}$$

The energy of the electron is quantized: eigenenergies



## Summary of Infinite Potential Well

Wavefunction in an infinite PE well

$$\psi_n(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right)$$

Electron energy in an infinite PE well

$$E_n = \frac{\hbar^2(\pi n)^2}{2ma^2} = \frac{\hbar^2 n^2 \pi^2}{2ma^2}$$

Energy separation in an infinite PE well

$$\Delta E = E_{n+1} - E_n = \frac{\hbar^2(2n+1)}{8ma^2}$$

### Problem 2.25

An electron is bound in a one-dimensional infinite potential well with a width of 75. Determine the electron energy levels (in eV) for  $n = 1, 2, 3$ .

### Next class

- Energy band theory

### Two important differential equations in quantum mechanics

Tania Roy

ECE

University

Solution to 2<sup>nd</sup> order differential equation of the form:  $\frac{d^2y}{dx^2} + k^2y = 0$

Assume  $y = Ae^{+jkx} + Be^{-jkx}$  [ $j = \sqrt{-1}$ ]

Then,

$$\begin{aligned} \frac{dy}{dx} &= jk(Ae^{+jkx} - Be^{-jkx}) \\ \frac{d^2y}{dx^2} &= j^2k^2 \underbrace{[Ae^{+jkx} + Be^{-jkx}]}_{=y} \\ &= -k^2y \end{aligned}$$

$$\frac{dx^2}{y} = -k^2y = y$$

Thus the proposed solution  $y = Ae^{+jkx} + Be^{-jkx}$  holds.

Solution to 2 nd order differential equation of the form:

$$\frac{d^2y}{dx^2} - k^2y = 0.$$

Assume  $y = Ae^{kx} + Be^{-kx}$

$$\begin{aligned}y &= Ae \\ \frac{dy}{dx} &= k(Ae^{kx} - Be^{-kx}) \\ d^2y &= k^2(Ae^{kx} + Be^{-kx}) \\ d^2x^2 &= k^2y\end{aligned}$$

Thus the proposed solution works.

Solution is  $y = Ae^{kx} + Be^{-kx}$

# Lecture 4: Quantum theory-Band structure

---

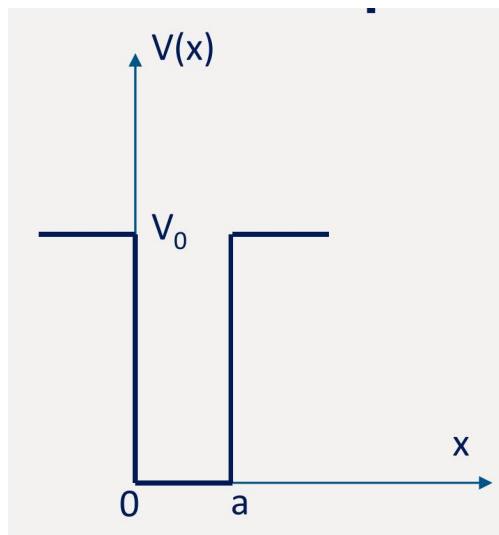
## Topics

- Neamen Chapter 3
- Electrons in a crystal
- Origin of bandgap

## Recap

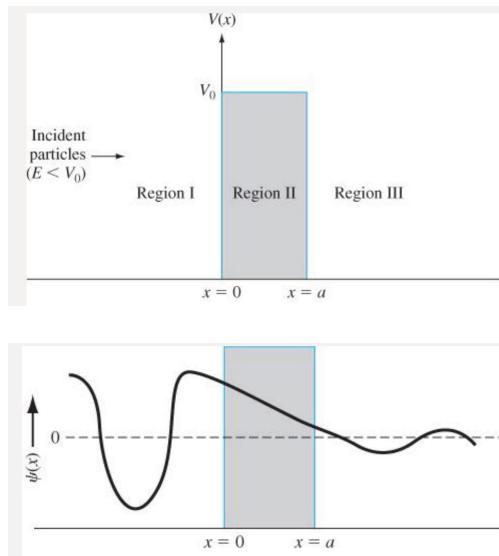
- Schrödinger wave equation
- Particle in infinite quantum well

Finite potential well



## Wavefunction through the potential barrier

Figure 2.9 I The potential barrier function.



### Extension of wave theory to atoms

Table 2.1 | Initial portion of the periodic table

Element	Notation	<i>n</i>	<i>l</i>	<i>m</i>	<i>s</i>
Hydrogen	$1s^1$	1	0	0	$+\frac{1}{2}$ or $-\frac{1}{2}$
Helium	$1s^2$	1	0	0	$+\frac{1}{2}$ and $-\frac{1}{2}$
Lithium	$1s^2 2s^1$	2	0	0	$+\frac{1}{2}$ or $-\frac{1}{2}$
Beryllium	$1s^2 2s^2$	2	0	0	$+\frac{1}{2}$ and $-\frac{1}{2}$
Boron	$1s^2 2s^2 2p^1$	2	1		
Carbon	$1s^2 2s^2 2p^2$	2	1		
Nitrogen	$1s^2 2s^2 2p^3$	2	1		
Oxygen	$1s^2 2s^2 2p^4$	2	1		
Fluorine	$1s^2 2s^2 2p^5$	2	1		$s = +\frac{1}{2}, -\frac{1}{2}$
Neon	$1s^2 2s^2 2p^6$	2	1		

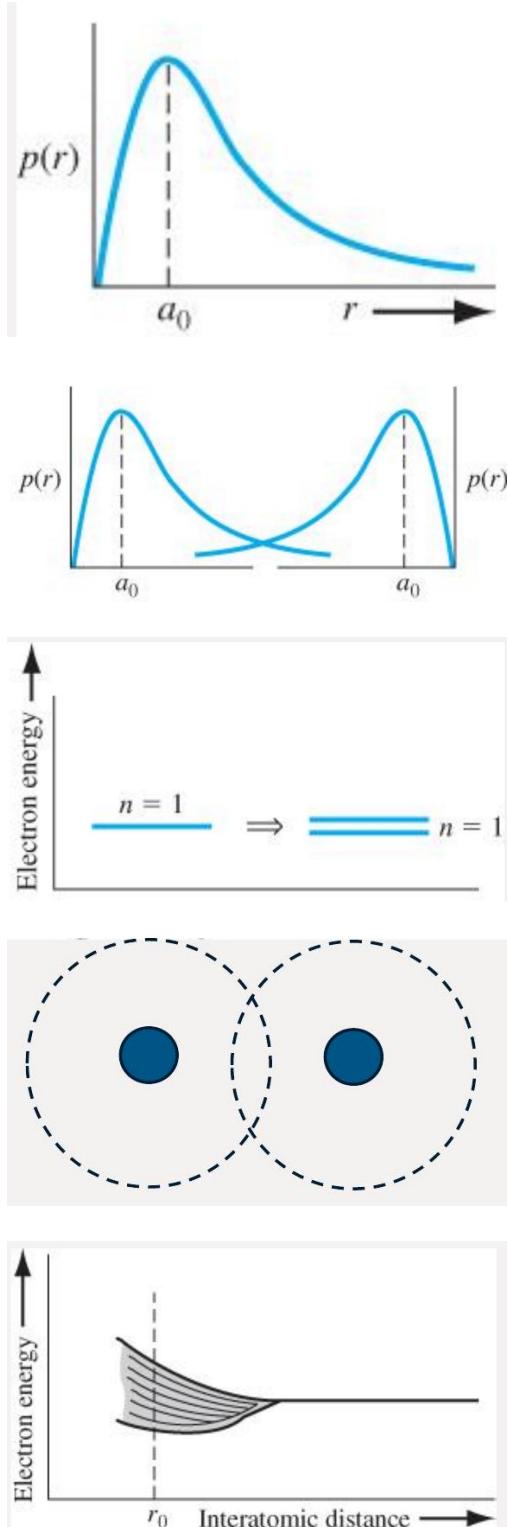
### Formation of energy bands

What happens when atoms are brought close enough together so that their electronic orbitals interact?

- Energy bands are formed by splitting of their energy levels

- (a)
- (b)
- (c)

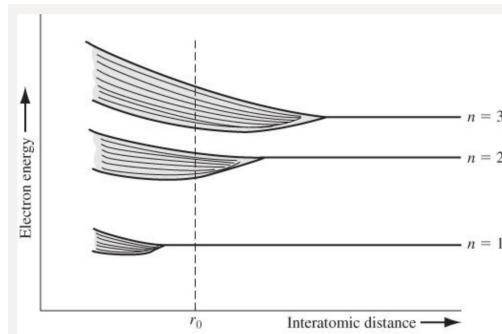
Figure 3.1 I (a) Probability density function of an isolated hydrogen atom. (b) Overlapping probability density functions of two adjacent hydrogen atoms. (c) The splitting of the  $n = 1$  state.



**What happens when  $n$  atoms are brought close to each other?**

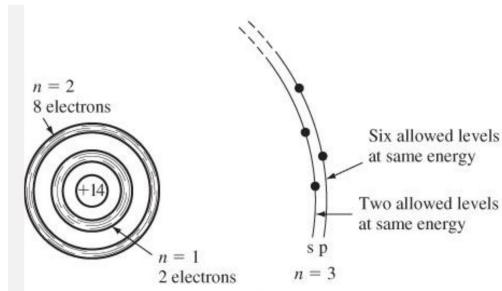
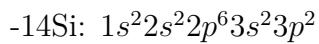
Figure 3.2 I The splitting of an energy state into a band of allowed energies.

Figure 3.3 I Schematic showing the splitting of three energy states into allowed bands of

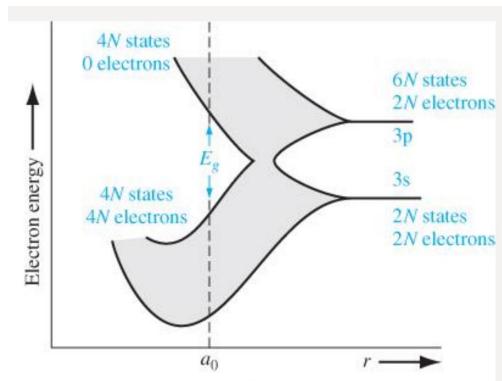


energies.

**What happens when many silicon atoms are brought close to each other?**



(a)



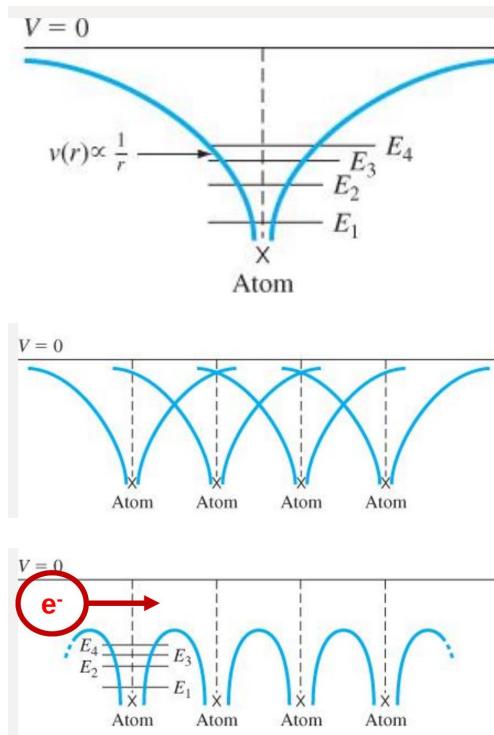
(b)

Figure 3.4 I (a) Schematic of an isolated silicon atom. (b) The splitting of the  $3s$  and  $3p$  states of silicon into the allowed and forbidden energy bands. (From Shockley [6].)

### Kronig-Penney model

(a)

(b)



(c)

Figure 3.5 I (a) Potential function of a single isolated atom. (b) Overlapping potential functions of adjacent atoms. (c) Net potential function of a one-dimensional single crystal.

- Energy bands tell us where electrons are allowed to be, but what dictates how they move from one place to another?
- Most influential is their attraction to the atomic nuclei
- Kronig-Penney model approximates with square step functions and solves Schrödinger's equation

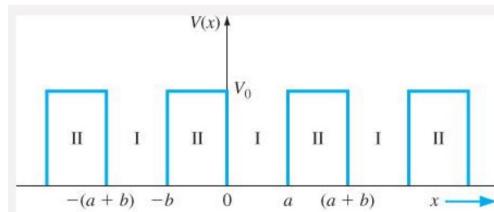


Figure 3.6 I The one-dimensional periodic potential function of the Kronig-Penney model.

### Kronig-Penney model: results

- A solution for energy of an electron with relation to  $k$
- The relation between  $E$  and  $k$  is also known as "dispersion relation"

- $k$  = wave number
- Momentum  $p$  of electron
- $p = \hbar k$
- The most generalized solution gives this relation for energy:

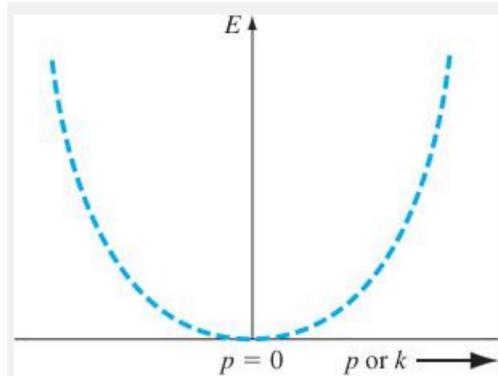


Figure 3.7 | The parabolic  $E$  versus  $k$  curve for the free electron.

- $E = \frac{p^2}{2m^*} = \frac{\hbar^2 k^2}{2m^*} \Rightarrow$  parabolic dispersion relation for "free particle"

### Kronig-Penney model contd.

- If the periodic conditions of the crystal are applied, then the "allowed" values of  $k$  for which the wave equation has a solution give you your E-k diagram:

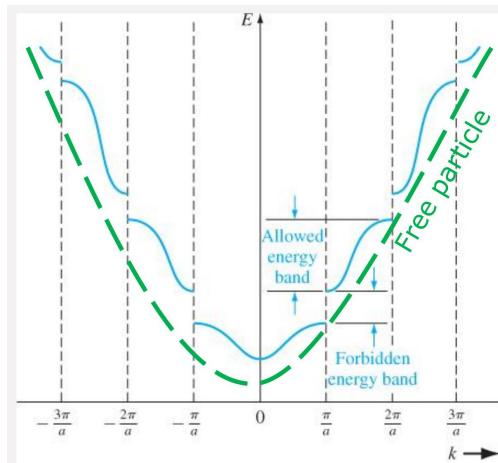
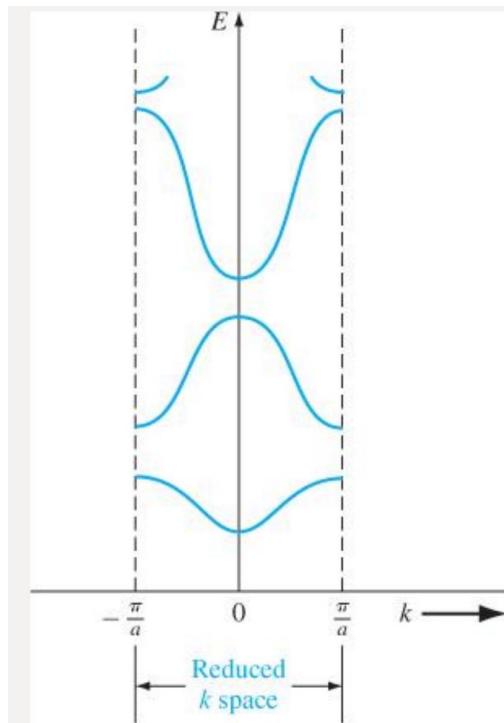


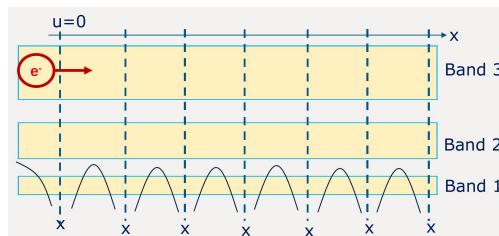
Figure 3.9 | The  $E$  versus  $k$  diagram generated from Figure 3.8. The allowed energy bands and forbidden energy bandgaps are indicated.

Due to the periodicity of  $2\pi$ , we can translate the allowed energy bands to a reduced zone "k-space" representation

Figure 3.11 | The  $E$  versus  $k$  diagram in the reduced-zone representation.



Visualizing these energy bands in a crystal



- Generally, an electron will be in an "upper" or higher energy band, thus effectively like a free particle
- $k = \sqrt{\frac{2mE}{\hbar^2}}; E = \frac{\hbar^2 k^2}{2m}$

Where are electrons allowed to be in a semiconductor crystal?

Atom: electrons in orbitals

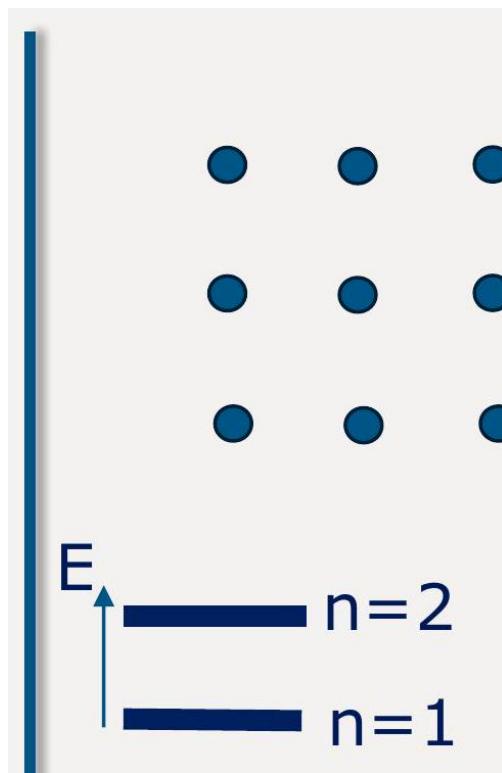
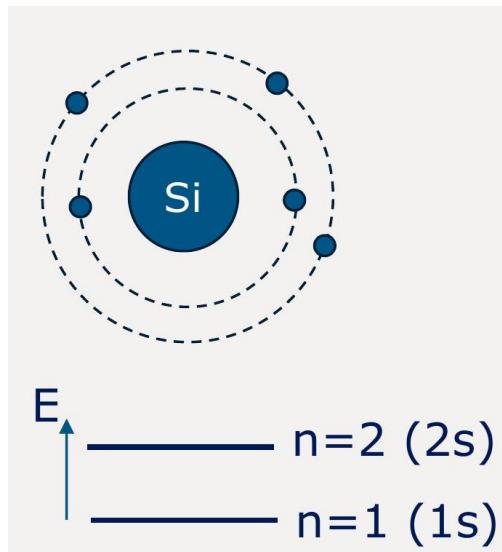
N atoms brought together cause E-levels to split N times

3D crystal lattice

Energy bands are formed in atomic crystal

An electron can only be on line in k-space. A change in E is a movement through k space or momentum space, and vice versa.

Solving the wave equation in periodic crystal yields relationship between E and k (wavenumber, crystallographic direction vector)



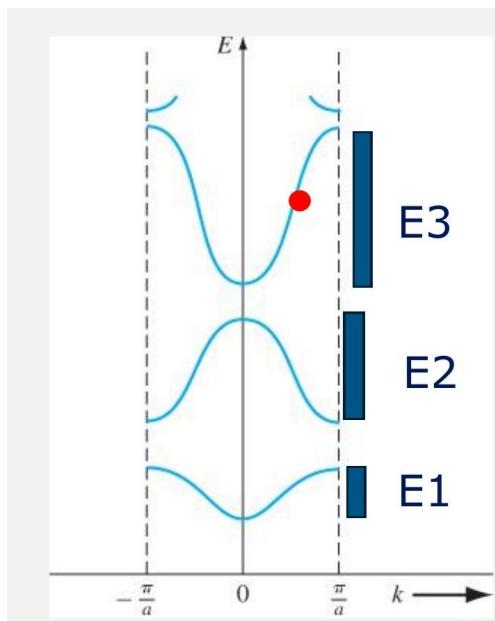
### Band gap

(a)

(b)

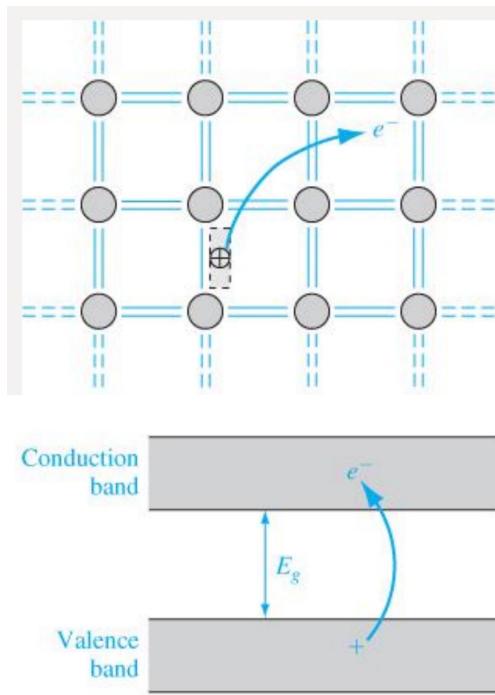
Figure 3.13 I (a) Two-dimensional representation of the breaking of a covalent bond.

(b) Corresponding line representation of the energy band and the generation of a negative and positive charge with the breaking of a covalent bond.



Next class

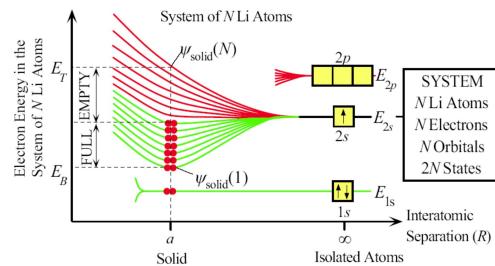
- Conduction and Density of States



### How energy bands form in Silicon (from "Principles of Electronic Materials and Devices" by Kasap)

Introduction to Microelectronic Devices and Circuits  
Spring 2024

#### Example: Formation of Energy Bands in Li

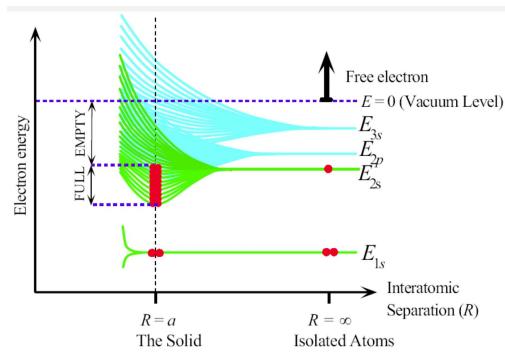


The formation of 2s energy band from the 2s orbitals when  $N$  Li atoms come together to form the Li solid.

There are  $N$  2s electrons, but  $2N$  states in the band. The 2s band is therefore only half full. The atomic 1s orbital is close to the Li nucleus and remains undisturbed in the solid. Thus, each Li atom has a closed K shell (full 1 s orbital).

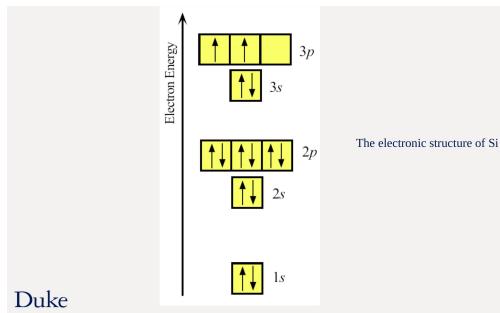
#### Example: Formation of Energy Bands 2

As Li atoms are brought together from infinity, the atomic orbitals overlap and give rise to bands. Outer orbitals overlap first. The 3s orbitals give rise to the 3s band, 2p orbitals to

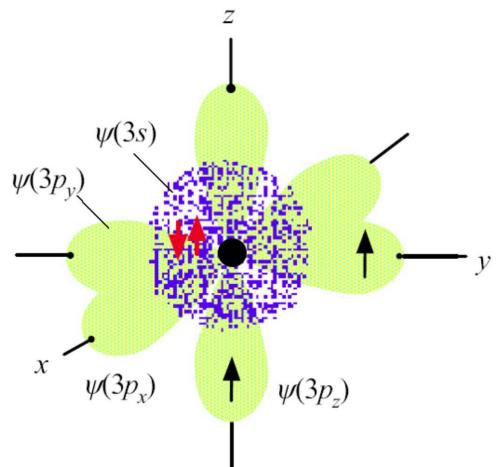


the  $2p$  band, and so on. The various bands overlap to produce a single band in which the energy is nearly continuous.

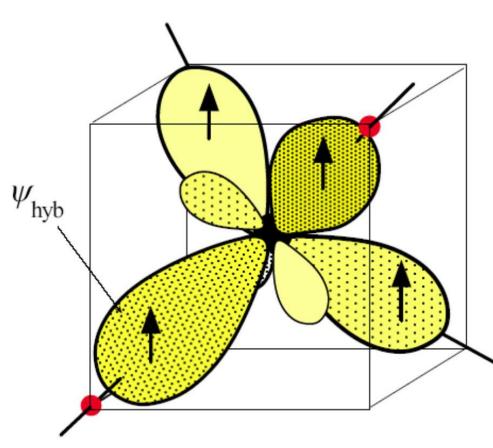
### Isolated Silicon Atom



### Hybridization in Si



- (a) Isolated Si
- (b) Si preparing to bond



(a) Si is in Group IV in the Periodic Table. An isolated Si atom has two electrons in the  $3s$  and two electrons in the  $3p$  orbitals.

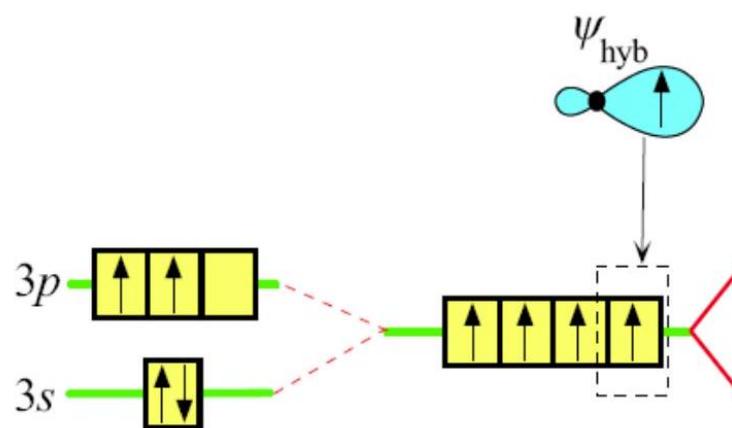
(b) When Si is about to bond, the one  $3s$  orbital and the three  $3p$  orbitals become perturbed and mixed to form four hybridized orbitals,  $\psi_{\text{hyb}}$ , called  $sp^3$  orbitals, which are directed toward the corners of a tetrahedron. The  $\psi_{\text{hyb}}$  orbital has a large major lobe and a small back lobe. Each  $\psi_{\text{hyb}}$  orbital takes one of the four valence electrons.

### $sp^3$ hybridization

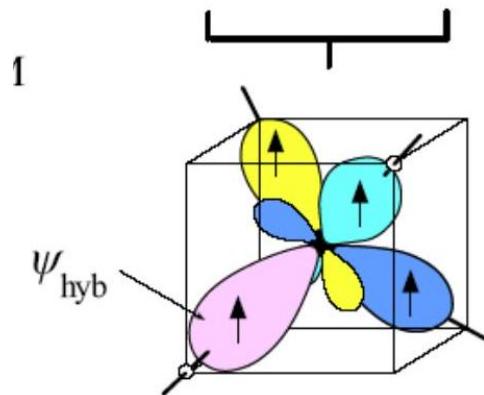
The  $3s$  and  $3p$  energy levels are quite close, and when five Si atoms approach each other, the interaction results in the four orbitals  $\psi(3s)$ ,  $\psi(3p_x)$ ,  $\psi(3p_y)$  and  $\psi(3p_z)$  mixing together to form four new hybrid orbitals, which are directed in tetrahedral directions; that is, each one is aimed as far away from the others as possible.

### Formation of energy bands in the Si crystal 1

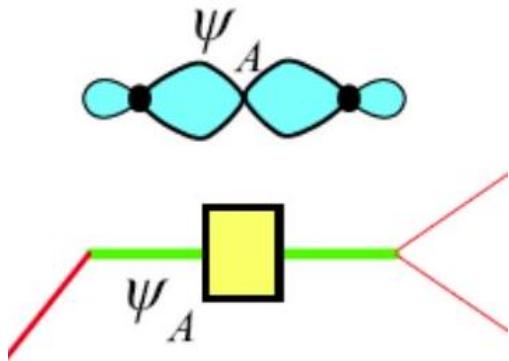
(a)



## Si ATOM



(c)

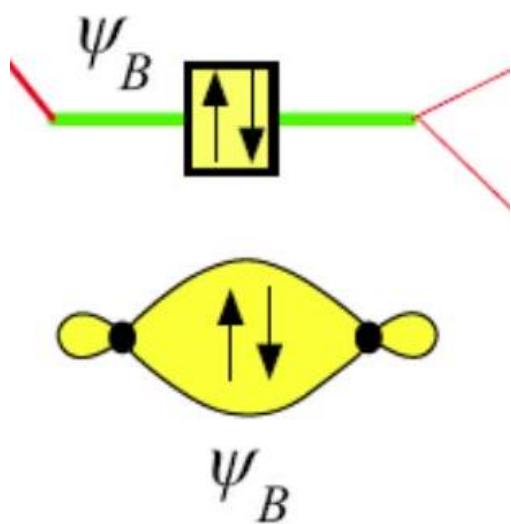


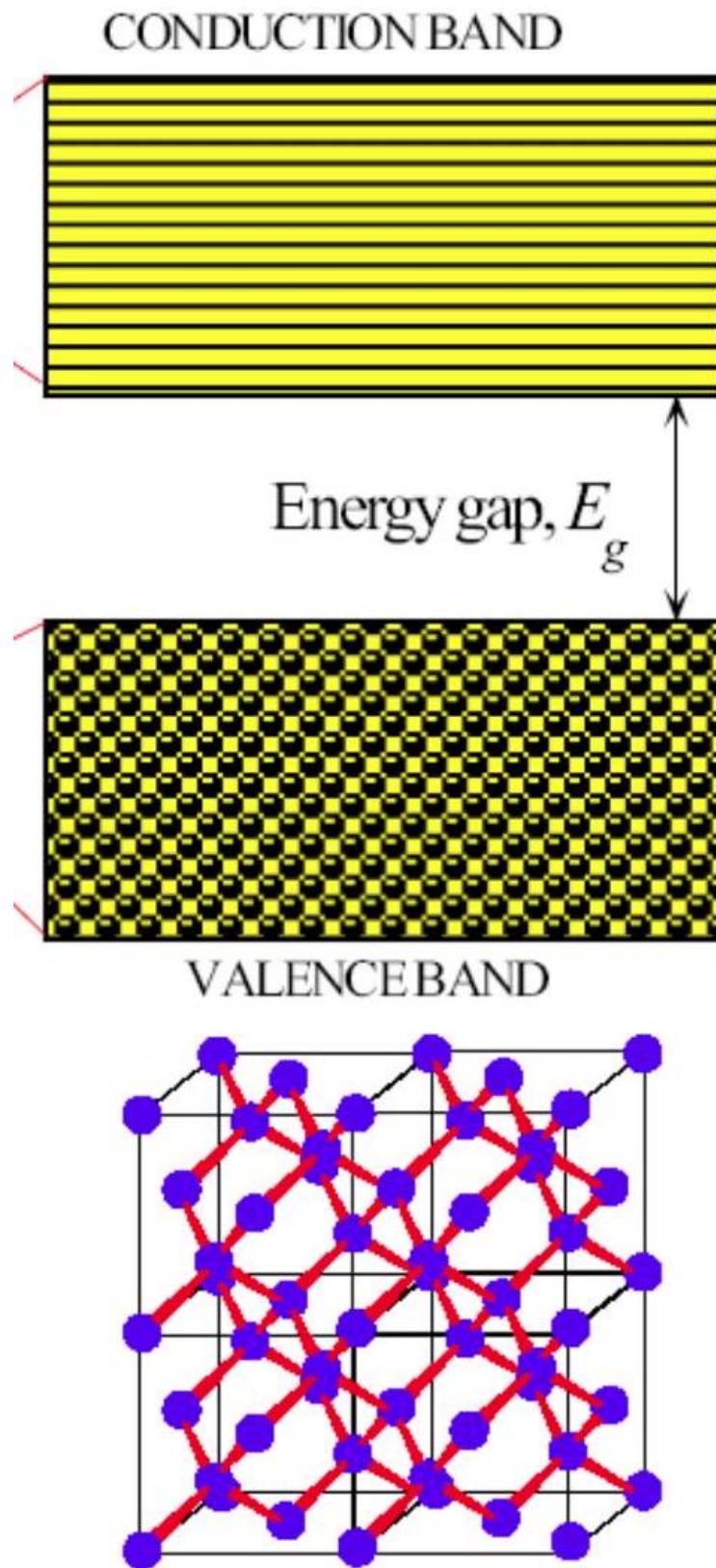
(d)

## Si CRYSTAL

(a) Formation of energy bands in the Si crystal first involves hybridization of  $3s$  and  $3p$  orbitals to four identical  $\psi_{\text{hyb}}$  orbitals which make  $109.5^\circ$  with each other as shown in (b).

(c)  $\psi_{\text{hyb}}$  orbitals on two neighboring Si atoms can overlap to form  $\psi_B$  or  $\psi_A$ . The first is a bonding orbital (full) and the second is an antibonding orbital (empty). In the crystal  $\psi_B$  overlap to give the valence band (full) and  $\psi_A$  overlap to give the conduction band (empty)





---

# Lecture 5: Quantum theory-Conduction and DOS

---

## Topics

- Neamen Chapter 3
- E-k dispersion relation
- Effective mass
- Concept of holes
- Density of states
- Fermi function

## E-k dispersion relation

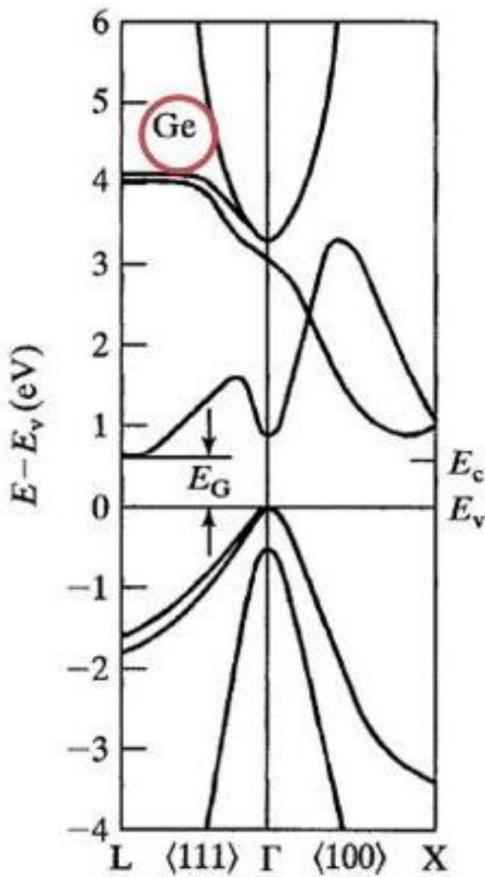
- So far this was developed for 1D (Kronig-Penney model); how does it look in a 3D crystal?
- Direct bandgap
- Indirect bandgap

## Effective mass ( $m^*$ )

- Consider a parabolic relation from free electron:  $E = \frac{\hbar^2 k^2}{2m}$
- $\frac{dE}{dk} = \frac{\hbar^2 k}{m} = \frac{\hbar p}{m}$  (based on  $p = \hbar k$ )
- Relate p to velocity by  $v = \frac{p}{m} = \frac{1}{\hbar} \frac{dE}{dk}$
- Take 2<sup>nd</sup> derivative:  $\frac{d^2 E}{dk^2} = \frac{\hbar^2}{m} \implies m^* = \frac{\hbar^2}{\frac{d^2 E}{dk^2}}$  THE EFFECTIVE MASS

$$m_0 = 9.11 \times 10^{-31} \text{ kg}$$

$m^* = m_n m_0$     Example: effective mass of an electron in silicon = 0.3!



### Effective mass

- For a free electron, mass is constant.
- But when bound in a crystal with periodic potential, the electron mass depends on inverse of  $E-k$  curvature
- $m^*$  ties the classical world to quantum:
- $F = m^*a = -qE$

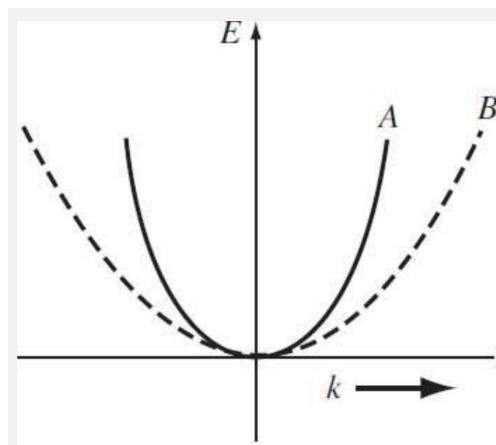
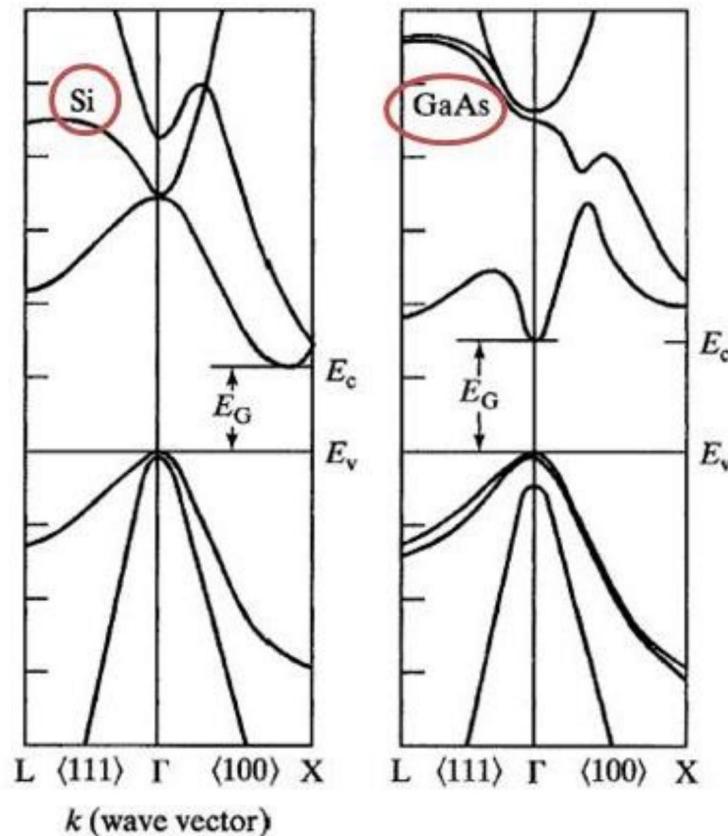
### Exercise

Two possible conduction bands are shown in the  $E$  versus  $k$  diagram given in Figure P3.13. State which band will result in the heavier electron effective mass; state why.

Figure P3.13 I Conduction bands for Problem 3.13.

Two possible valence bands are shown in the  $E$  versus  $k$  diagram given in Figure P3.14. State which band will result in the heavier hole effective mass; state why.

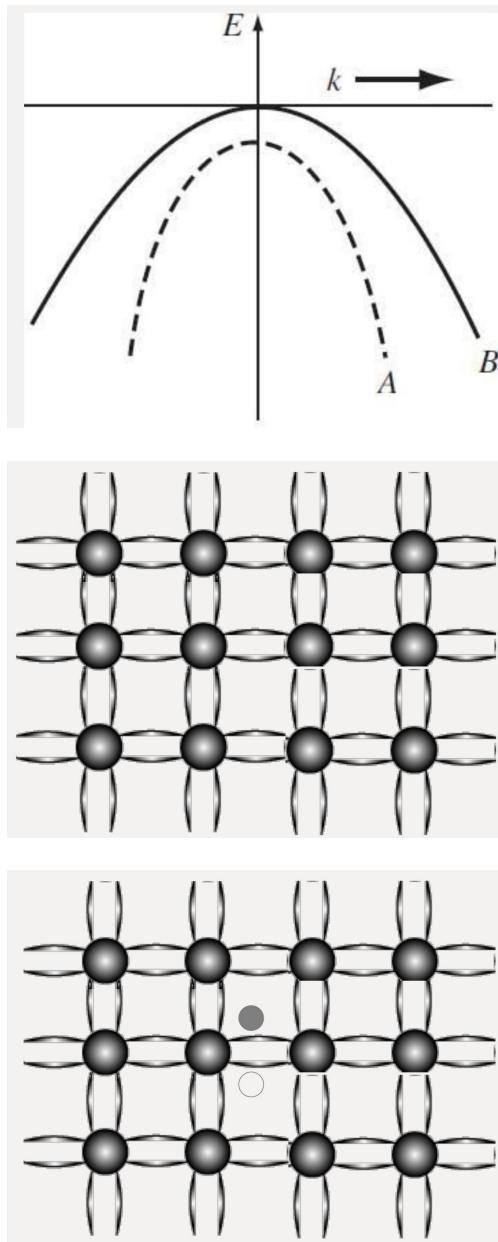
Figure P3.14 I Valence bands for Problem 3.14.



### Intrinsic Si: The Bond Model

Each atom shares 2 electrons with 4 nearest neighbors to form a covalent bond

At  $T = 0$  K, all bonds are satisfied, there are no free carriers, no current flows, looks like an insulator

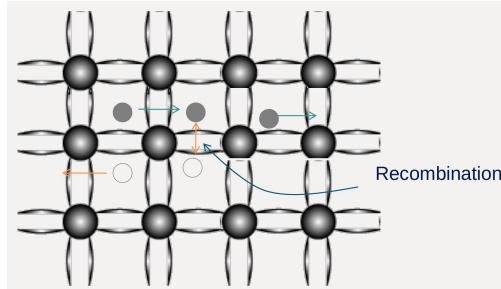


### Intrinsic Si: The Bond Model: Electrons

At finite temperature, an electron may gain enough energy to break the covalent bond, become free and move around. As a result, current will flow if a voltage is applied to Si. As the electrons move, they leave behind a hole, which itself can move.

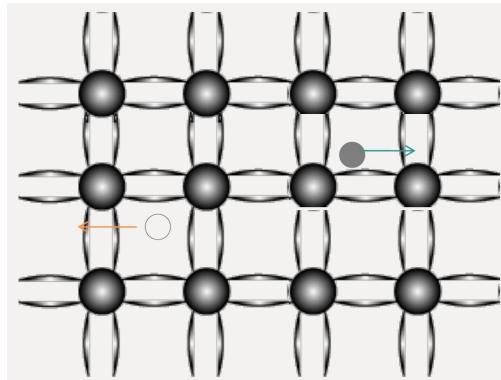
### Intrinsic Si: The Bond Model: Holes

Movement of electrons is equivalent to opposite movement of holes. How? If the electron moves towards right and there is enough thermal energy in the system to dislodge electrons, after some time another electron from left will come and recombine with the hole. The net result is that the hole moves from right to left. For all practical purposes, it is easy to



consider holes as equivalent to electrons with positive charge .

### Si: The Bond Model: Holes



Note that each electron is associated with a complementary hole. So the material is charge neutral.

However, electrons and holes are not exactly the same quantity with opposite charges. We shall see why when we discuss band model.

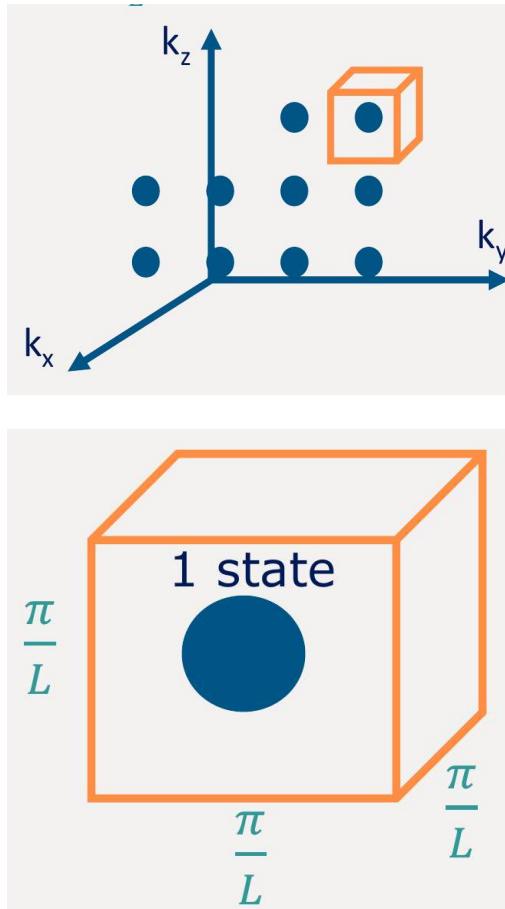
### Where are the electrons? Density of states (DOS)

- How many states are there at a certain energy E ?
  - NOTE: There are several different ways to derive DOS. This one is different from the textbook (so you have options!):
  - 3D-DOS derivation for parabolic  $E - k^2$  ( $E = \frac{\hbar^2 k^2}{2m^*}$ ) and crystal that is  $L$  in length  $x, y$  and  $z$
1. Determine the # of states per unit volume of k-space

From Schrödinger's equation, allowed wave numbers in periodic potential are:  $k = \frac{\pi}{L}n$  (n is an integer)

With  $\frac{\pi}{L}$  periodicity, a box around one state in k-space would be:

Volume of the entire crystal is  $V = L^3$



$$\Rightarrow \frac{1 \text{ state}}{\left(\frac{\pi}{L}\right)^3} = \frac{1}{\frac{\pi^3}{V}} = \frac{V}{\pi^3} \times 2 = \frac{2V}{\pi^3} \quad \left( \frac{\text{allowed energy states}}{\text{unit volume } k - \text{ space}} \right)$$

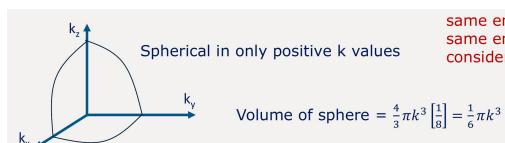
For spin up and spin down

3D-DOS derivation for parabolic E-k ( $E = \frac{\hbar^2 k^2}{2m^*}$ ) and crystal that is  $L$  in length  $x, y$  and  $z$

## 2. Find total # of states up to a value $k$

$$N(k) = \left( \frac{\text{allowed energy states}}{\text{unit volume } k - \text{ space}} \right) (\text{total volume up to } k)$$

$$N(k) = \left( \frac{2V}{\pi^3} \right) \left( \frac{1}{6} \pi k^3 \right) = \frac{V k^3}{3 \pi^2}$$



3D-DOS derivation for parabolic E-k ( $E = \frac{\hbar^2 k^2}{2m^*}$ ) and crystal that is  $\mathbf{L}$  in length  $\mathbf{x}, \mathbf{y}$  and  $\mathbf{z}$

3. Convert to total number of states up to a certain energy  $E$

Use this conversion for E-k relation:  $k = \sqrt{\frac{2mE}{\hbar^2}}$

$$N(E) = \left( \frac{V}{3\pi^2} \right) \left[ \sqrt{\frac{2mE}{\hbar^2}} \right]^3 = \frac{V}{3\pi^2 \hbar^3} (2mE)^{3/2}$$

3D-DOS derivation for parabolic E-k ( $E = \frac{\hbar^2 k^2}{2m^*}$ ) and crystal that is  $L$  in length  $\mathbf{x}, \mathbf{y}$  and  $\mathbf{z}$

4. Differentiate w.r.t. energy and divide by volume for DOS per unit volume:

$$\text{DOS}_{3D} = g(E) = \frac{1}{V} \frac{d}{dE} N(E) = \frac{m\sqrt{2mE}}{\pi^2 \hbar^3}$$

Using  $\hbar = \frac{h}{2\pi}$  gives:  $g(E) = \frac{4\pi}{h^3} (2m)^{3/2} \sqrt{E}$   
(format in book)

3D-DOS derivation for parabolic E-k ( $E = \frac{\hbar^2 k^2}{2m^*}$ ) and crystal that is  $\mathbf{L}$  in length  $\mathbf{x}, \mathbf{y}$  and  $\mathbf{z}$

Applying to conduction and valence bands (approx. parabolic):

$$\text{Conduction Band: } E = E_C + \frac{\hbar^2 k^2}{2m_n^*} \rightarrow g_C(E) = \frac{4\pi}{h^3} (2m_n^*)^{3/2} \sqrt{E - E_C}$$

$$\text{Valence Band: } E = E_V - \frac{\hbar^2 k^2}{2m_p^*} \rightarrow g_V(E) = \frac{4\pi}{h^3} (2m_p^*)^{3/2} \sqrt{E_V - E}$$

## Density of states

$$g_C(E) = \frac{4\pi}{h^3} (2m_n^*)^{3/2} \sqrt{E - E_C}$$

$$g_V(E) = \frac{4\pi}{h^3} (2m_p^*)^{3/2} \sqrt{E_V - E}$$

## What is the probability of finding an electron at a particular energy?

### Fermi level and Fermi-Dirac statistics:

The probability of finding an electron at a particular energy level,  $E$ , is given by the Fermi-Dirac statistics, most often called the Fermi Function:

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

$E_f$  is the Fermi level which is defined as the energy at which the probability of finding an electron is exactly 1/2.

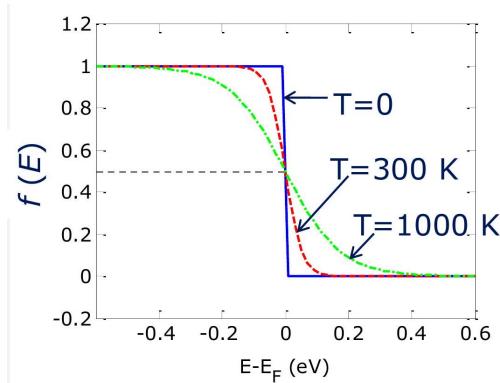
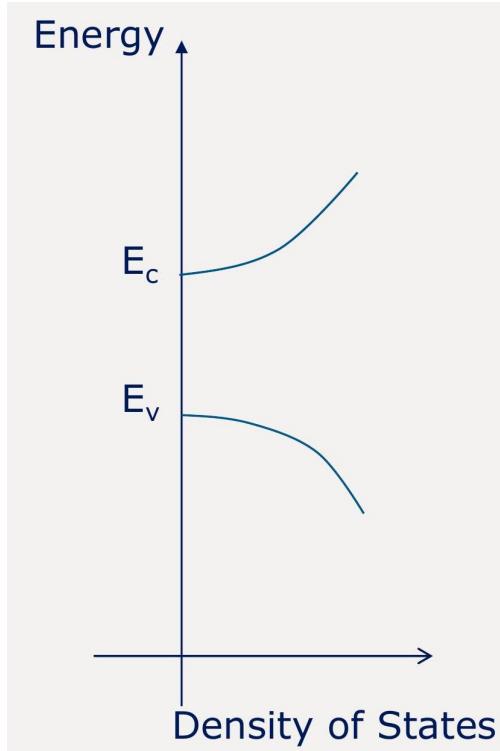
### Fermi Function

Boltzmann approximation:

$$f(E) \approx e^{-(E-E_f)/kT} \quad E - E_f \gg kT$$

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

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



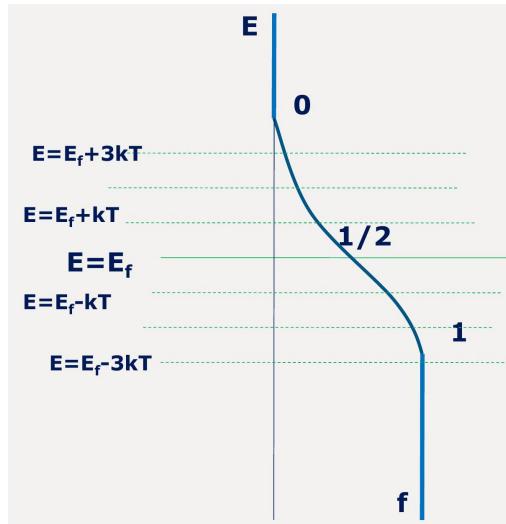
### Example

- Assume  $E_F$  is 0.3eV below  $E_c$ .  $T = 300$  K. (a) Determine the probability of a state being occupied by an electron at  $E = E_c + kT/4$ . (b) same for  $E = E_c + kT$

$$\begin{aligned}
 f_F(E) &\cong \exp\left[\frac{-(E-E_F)}{kT}\right] = \exp\left[\frac{-(E_c+kT/4-E_F)}{kT}\right] \\
 &= \exp\left[\frac{-(0.30+0.0259/4)}{0.0259}\right] = 7.26 \times 10^{-6} \\
 f_F(E) &\cong \exp\left[\frac{-(E-E_F)}{kT}\right] = \exp\left[\frac{-(0.30+0.0259)}{0.0259}\right] = 3.43 \times 10^{-6}
 \end{aligned}$$

### Next class

- Number of carriers in equilibrium



- In-class Quiz (based on all topics covered so far)
- Easy!!

# Lecture 6: Quantum theory-Equilibrium carrier Statistics

---

## Topics

- Neamen Chapter 4
- Equilibrium distribution of electrons and holes
- Intrinsic semiconductor
- Intrinsic carrier concentration  $n_i$
- Intrinsic Fermi level  $E_i$
- Expressions for equilibrium carrier concentrations,  $n_0, p_0$
- Quiz

**What makes a crystal a semiconductor VS. insulator vS. metal?**

Insulator

$$E_g > 3\text{eV}$$

Semiconductor

$$0.5\text{eV} < E_g < 3\text{eV}$$

Metal

$E_g \sim 0\text{eV}$ , or  $E_F$  goes through a band - there are already states for electrons to move through

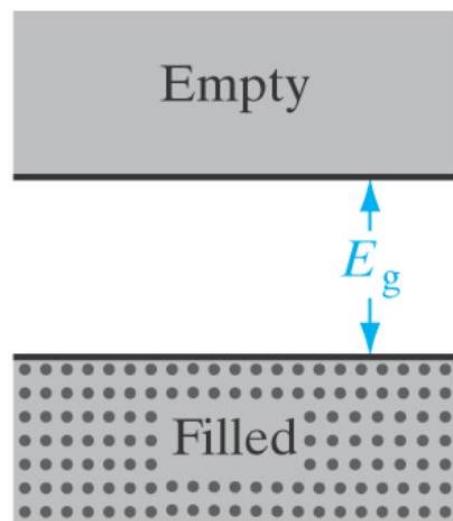
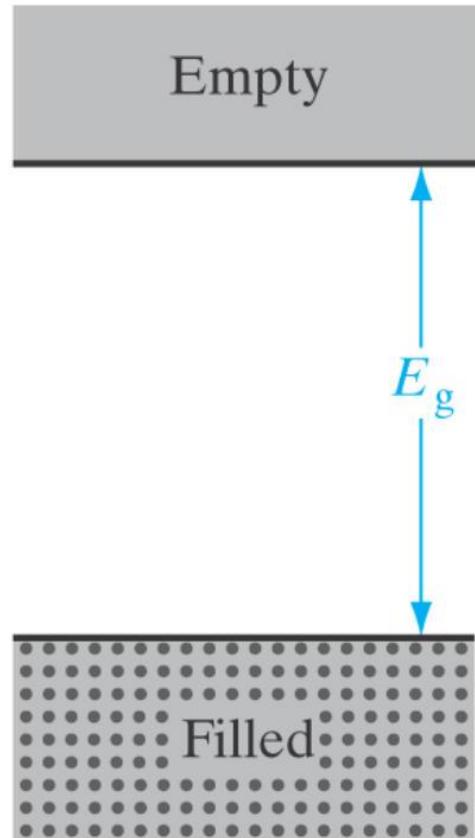
**Semiconductor, insulator, metal bandgap**

Si, Semiconductor

$\text{SiO}_2$ , insulator

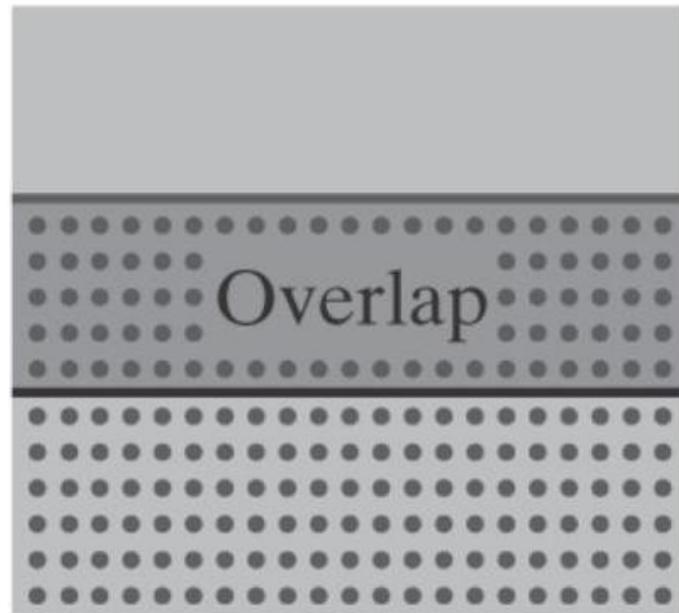
Conductor

- Band structure explains why  $\text{SiO}_2$  (diamond, etc) is insulating, silicon is semiconducting, copper is a metal
- For electrons to be accelerated in an electric field they must be able to move into new, unoccupied energy states.
- Water bottle flow analogy (empty vs. full)
- So, what is a hole then?



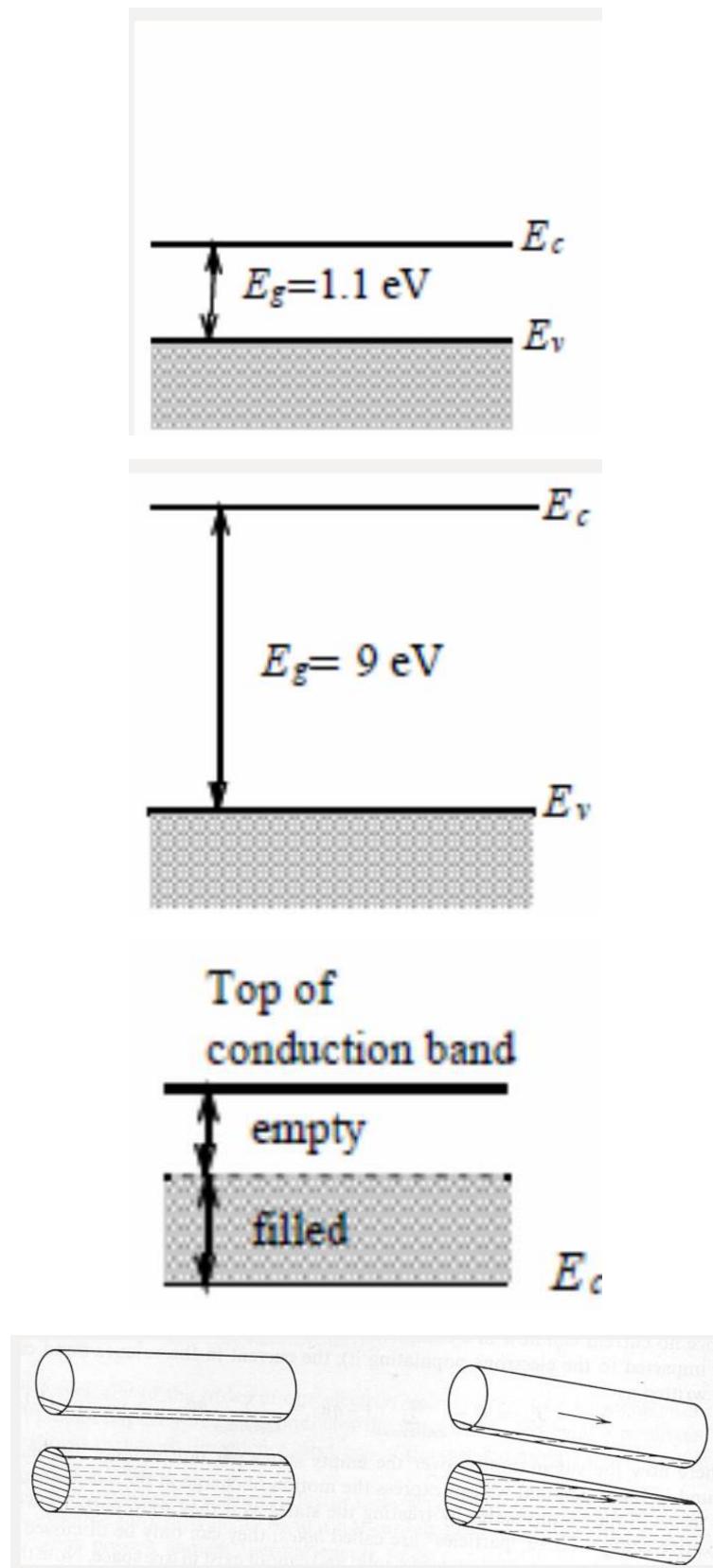
**Going from  $E - k$  to  $E - x$**

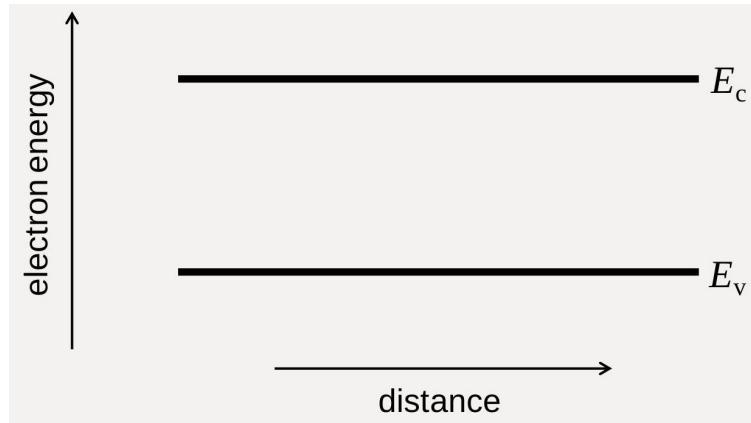
- In devices we usually draw:
- Simplified version of energy band model, indicating



- Top edge of valence band ( $E_V$ )
- Bottom edge of conduction band (  $E_C$  )
- Their separation, i.e. band gap energy

$$(E_G)$$





### Number of carriers

- We would like to find out the number of carriers in the semiconductor
- The game plan:
  - (i) Find out the number of states available per unit energy (Density of states)
  - (ii) Find out the probability of finding an electron in the states (Fermi Function)
  - (iii) Multiply (i) and (ii) to find out number of electrons per unit energy
  - (iv) Integrate over the relevant energy

### Recall: Density of states

$$g_C(E) = \frac{4\pi}{h^3} (2m_n^*)^{3/2} \sqrt{E - E_C}$$

$$g_V(E) = \frac{4\pi}{h^3} (2m_p^*)^{3/2} \sqrt{E_V - E}$$

In other places, you may also see the following forms for DOS:

$$D_c(E) \equiv \frac{m_n^* \sqrt{2m_n^*(E - E_c)}}{\pi^2 \hbar^3}; E \geq E_c \quad D_v(E) \equiv \frac{m_p^* \sqrt{2m_p^*(E_v - E)}}{\pi^2 \hbar^3}; E \leq E_v$$

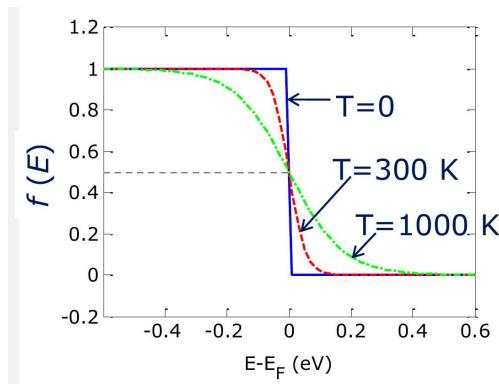
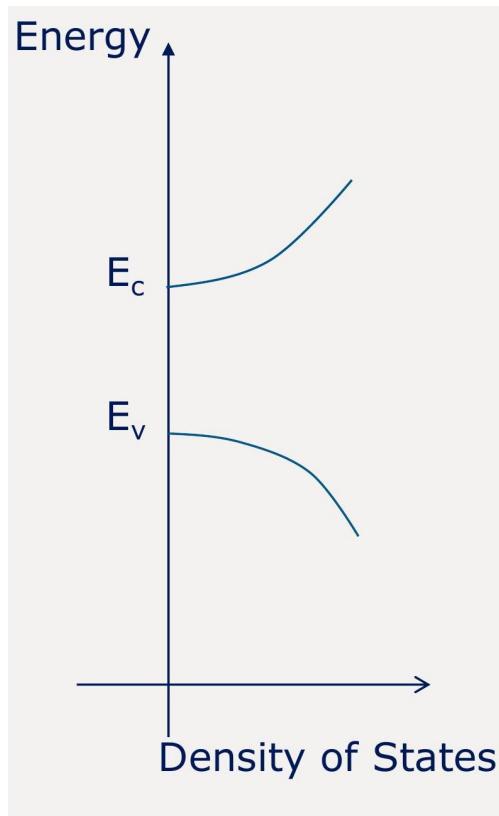
### Recall: What is the probability of finding an electron at a particular energy?

Fermi level and Fermi-Dirac statistics:

The probability of finding an electron at a particular energy level,  $E$ , is given by the Fermi-Dirac statistics, most often called the Fermi Function:

$E_f$  is the Fermi level which is defined as the energy at which the probability of finding an electron is exactly 1/2.

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



Boltzmann approximation:

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

for

$$E - E_f \gg kT$$

### Charge carrier distribution

- How do we calculate electron & hole concentrations knowing the density of states and probability of occupation?

$$n_0 = \int_{E_C}^{\infty} f(E) N(E) dE$$

- This is the density of electrons in the C-band. What about holes in the V-band?

$$p_0 = \int_{-\infty}^{E_V} [1 - f(E)] N_V(E) dE$$

- This is usually a tough numerical integral, but we can approximate it if  $E_F$  is well inside the band gap

$$n_0 \approx N_C e^{-(E_C - E_F)/kT} \quad N_C = 2 \left( \frac{2\pi m_n^* kT}{h^2} \right)^{3/2}$$

$$p_0 \gg N_V e^{-(E_F - E_V)/kT} \quad N_V = 2 \left( \frac{2pm_p^* kT}{h^2} \right)^{3/2}$$

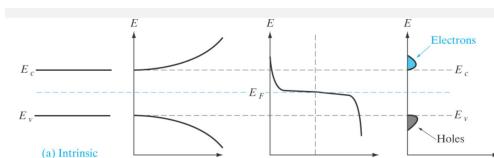
### Exercise

- Calculate the thermal equilibrium hole concentration in silicon at (a)  $T = 400$  K, (b)  $T = 250$  K. Given:  $N_v = 1.04 \times 10^{19} \text{ cm}^{-3}$  at 300 K. Assume Fermi energy is 0.27eV above valence band energy.

### Exercise

- Calculate the thermal equilibrium electron and hole concentration in silicon at  $T = 300$  K for the case when the Fermi energy level is 0.22eV below the conduction band energy  $E_c$ .  $E_G$  of SI = 1.12 eV

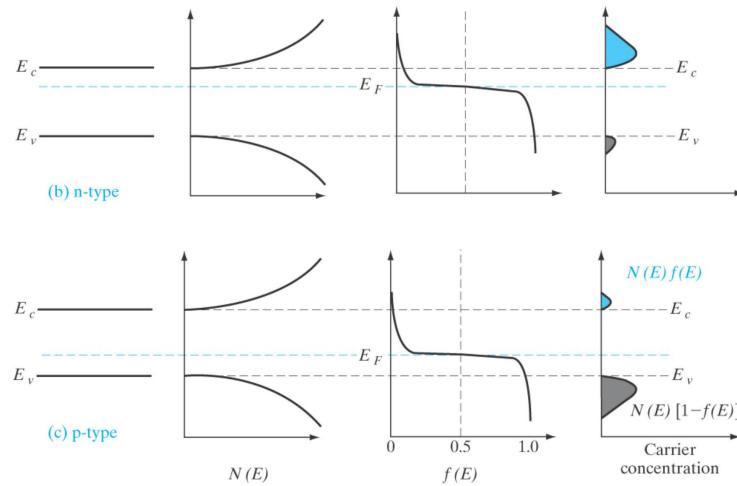
### Charge carrier distribution



(b) n-type

### Intrinsic semiconductor

- Ideally a semiconductor with no impurities, but most accurately defined as having an equal concentration of electrons and holes ( $E_F$  near middle of band gap)
- $n_0 = p_0 = n_i$



- The product  $n_0 p_0 = N_C e^{-(E_C - E_F)/kT} N_V e^{-(E_F - E_V)/kT} = N_C N_V e^{-(E_C - E_V)/kT}$
- Intrinsic carrier concentration  $n_i = \sqrt{N_C N_V} e^{-E_G/2kT}$

For Si,  $n_i =$  at 300 K

Table 4.2 I Commonly accepted values of  $n_i$  at  $T = 300$  K

Silicon

Gallium arsenide  $n_i = 1.5 \times 10^{10} \text{ cm}^{-3}$  Germanium  $n_i = 1.8 \times 10^6 \text{ cm}^{-3}$   $n_i = 2.4 \times 10^{13} \text{ cm}^{-3}$

Check Appendix B. 4 of Neamen

### Charge carrier distribution

$$n_i = \sqrt{N_C N_V} e^{-E_G/2kT}$$

- $n_0$  and  $p_0$  can be written as:

$$n_0 = n_i e^{(E_F - E_i)/kT} \quad p_0 = n_i e^{(E_i - E_F)/kT}$$

### What is intrinsic Fermi level $E_i$

- If the material is intrinsic,  $E_F = E_i =$  where?

$$\rightarrow E_C - E_i = E_G/2$$

$$n_0 \approx N_C e^{-(E_C - E_F)/kT} \quad p_0 \gg N_V e^{-(E_F - E_V)/kT}$$

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

$$n_0 = n_i e^{(E_C - E_i)/kT} e^{(E_F - E_C)/kT} = n_i e^{(E_F - E_i)/kT} \quad p_0 = n_i e^{(E_i - E_V)/kT} e^{(E_V - E_F)/kT} = n_i e^{(E_i - E_F)/kT}$$

For intrinsic semiconductor,  $n_o = p_o$   
Solving for the Fermi level, we find:

$$E_i = \frac{E_C + E_V}{2} + kT \ln \sqrt{\frac{N_V}{N_C}} = \frac{E_C + E_V}{2} + \frac{3}{4}kT \ln \frac{m_p^*}{m_n^*}$$

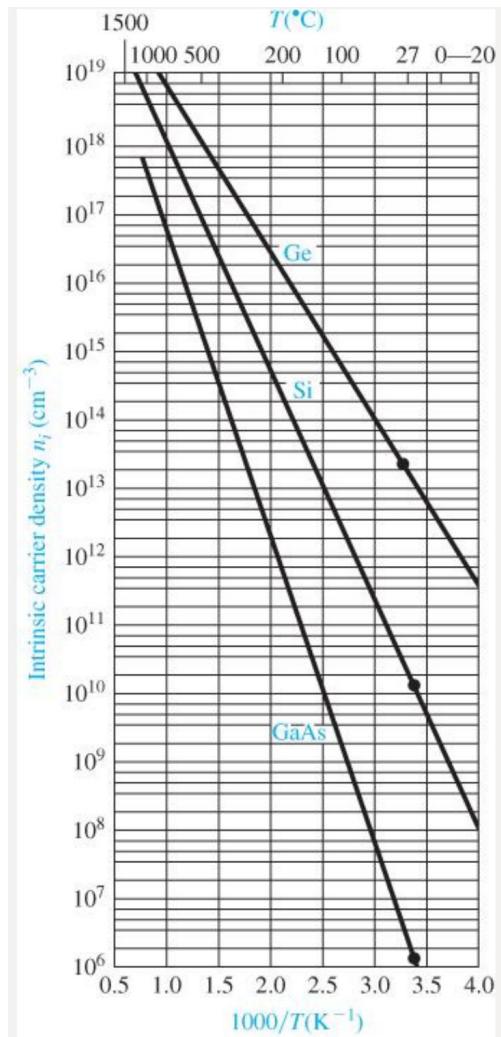
### Exercise

- Calculate the intrinsic carrier concentration in Si at T = 200 K, T = 450 K.

### Problem 4.5

Two semiconductor materials have exactly the same properties except material A has a bandgap energy of 0.90 eV and material B has a bandgap energy of 1.10eV. Determine the ratio of  $n_i$  of material B to that of material A for (a) T = 200 K, (b) T = 300 K, and (c) T = 400 K.

Figure 4.2 I The intrinsic carrier concentration of Ge, Si, and GaAs as a function of temperature.  
(From Sze [14].)



# Lecture 7: Quantum theory-Carrier statistics, Doping

---

## Topics

- Neamen Chapter 4
- $N_i$ ,  $E_i$
- Dopants
- Carrier concentration

## Intrinsic semiconductor

- Ideally a semiconductor with no impurities, but most accurately defined as having an equal concentration of electrons and holes ( $E_F$  near middle of band gap)
- $n_0 = p_0 = n_i$
- The product  $n_0 p_0 = N_C e^{-(E_C - E_F)/kT} N_V e^{-(E_F - E_V)/kT} = N_C N_V e^{-(E_C - E_V)/kT}$
- Intrinsic carrier concentration  $n_i = \sqrt{N_C N_V} e^{-E_G/2kT}$

For Si,  $n_i$  = at 300 K

Table 4.2 I Commonly accepted values of  $n_i$  at  $T = 300$  K

Silicon

Gallium arsenide  $n_i = 1.5 \times 10^{10} \text{ cm}^{-3}$  Germanium  $n_i = 1.8 \times 10^6 \text{ cm}^{-3}$   $n_i = 2.4 \times 10^{13} \text{ cm}^{-3}$

Check Appendix B. 4 of Neamen

## What is intrinsic Fermi level $E_i$

- If the material is intrinsic,  $E_F = E_i$  = where?

$$\rightarrow E_C - E_i = E_G/2$$

$$n_0 \approx N_C e^{-(E_C - E_F)/kT} \quad p_0 \gg N_V e^{-(E_F - E_V)/kT}$$

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

$$n_0 = n_i e^{(E_C - E_i)/kT} e^{(E_F - E_C)/kT} = n_i e^{(E_F - E_i)/kT} \quad p_0 = n_i e^{(E_i - E_V)/kT} e^{(E_V - E_F)/kT} = n_i e^{(E_i - E_F)/kT}$$

For intrinsic semiconductor,  $n_o = p_o$   
 Solving for the Fermi level, we find:

$$E_i = \frac{E_C + E_V}{2} + kT \ln \sqrt{\frac{N_V}{N_C}} = \frac{E_C + E_V}{2} + \frac{3}{4}kT \ln \frac{m_p^*}{m_n^*}$$

### Charge carrier distribution

$$n_i = \sqrt{N_C N_V} e^{-E_G/2kT}$$

- $n_0$  and  $p_0$  can be written as:

$$n_0 = n_i e^{(E_F - E_i)/kT} \quad p_0 = n_i e^{(E_i - E_F)/kT}$$

### Exercise

- Calculate the intrinsic carrier concentration in Si at  $T = 200$  K,  $T = 450$  K.

### Problem 4.5

Two semiconductor materials have exactly the same properties except material A has a bandgap energy of 0.90 eV and material B has a bandgap energy of 1.10eV. Determine the ratio of  $n_i$  of material B to that of material A for (a)  $T = 200$  K, (b)  $T = 300$  K, and (c)  $T = 400$  K.

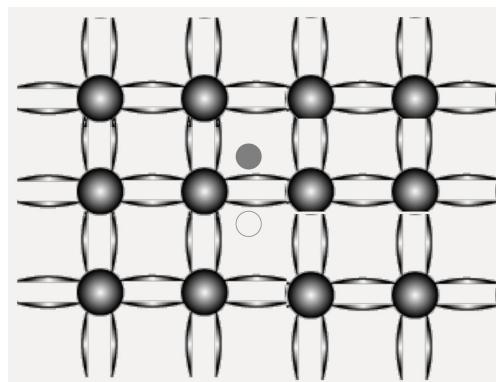
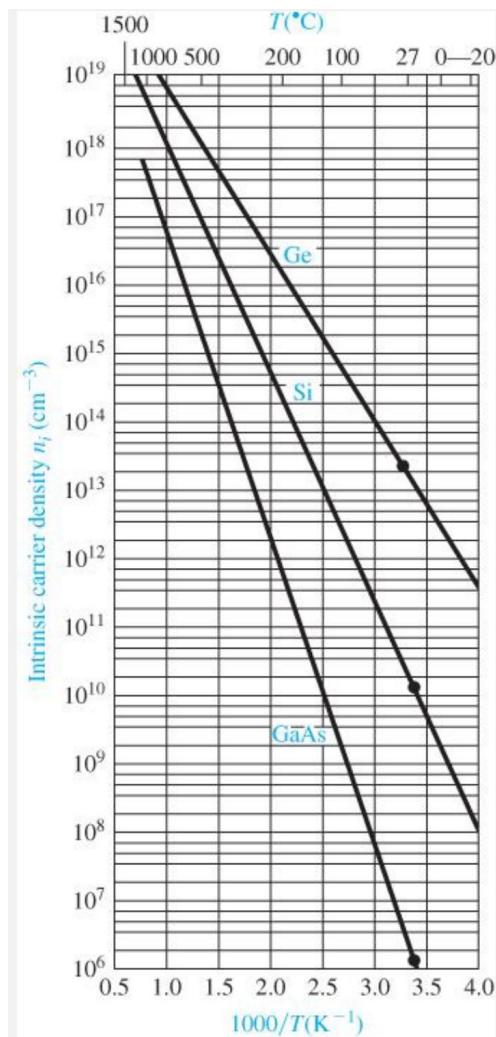
Figure 4.2 I The intrinsic carrier concentration of Ge, Si, and GaAs as a function of temperature.  
 (From Sze [14].)

### Intrinsic Si: The Bond Model: Electrons (Recap)

At finite temperature, an electron may gain enough energy to break the covalent bond, become free and move around. As a result, current will flow if a voltage is applied to Si. As the electrons move, they leave behind a hole, which itself can move.

### Intrinsic Si: The Bond Model: Holes (Recap)

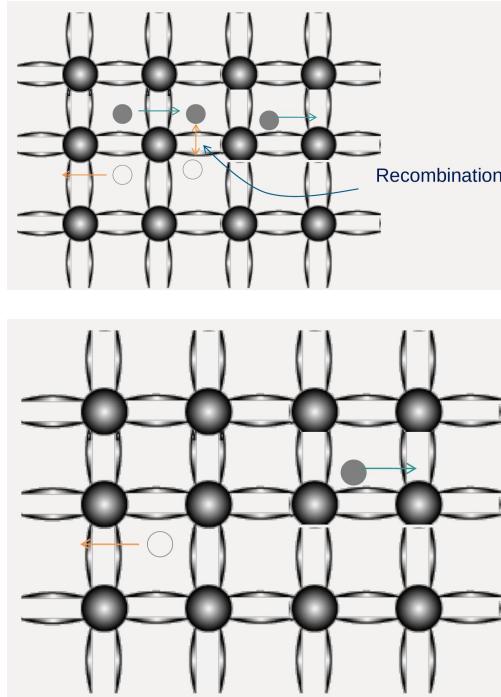
Movement of electrons is equivalent to opposite movement of holes. How? If the electron moves towards right and there is enough thermal energy in the system to dislodge electrons, after some time another electron from left will come and recombine with the hole. The net result is that the hole moves from right to left. For all practical purposes, it is easy to consider holes as equivalent to electrons with positive charge.



### Si: The Bond Model: Holes (Recap)

Note that each electron is associated with a complementary hole. So the material is charge neutral.

However, electrons and holes are not exactly the same quantity with opposite charges. We shall see why when we discuss band model.



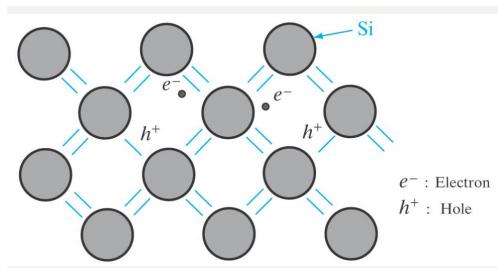
- How many electron-hole pairs (EHPs) are created when bonds are broken by thermal agitation?

Intrinsic carrier concentration  $n_i$

In Si,  $n_i \sim 10^{10} \text{ cm}^{-3}$

- In thermal equilibrium
- generation = recombination
- Simple probability:

$n_o, p_o$



- Recombination driven by... Availability of carriers  $n.p$
- Generation intrinsically driven only by... Temperature, light, etc.
- Hence the  $np$  product at thermal equilibrium:

$$n_o \cdot p_o = n_i^2$$

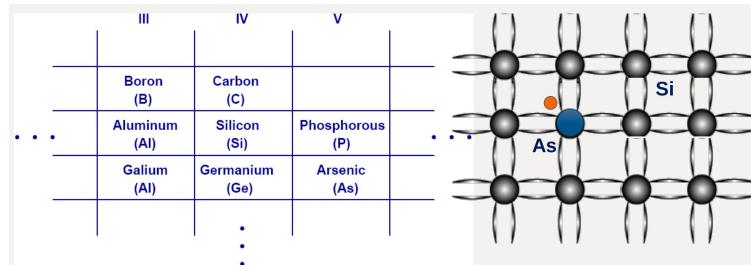
- At room temperature ( $T \approx 300$  K) intrinsic concentrations:
- $n_i \sim 2 \times 10^6$  electrons and holes per  $\text{cm}^3$  in GaAs
- $\sim 1 \times 10^{10}$   $\text{cm}^{-3}$  in Si
- $\sim 2 \times 10^{13}$  in Ge

-What about the band gaps of these materials?

- To get a sense of scale, how do these numbers:
- compare with  $N_A$  ?
- compare with number of Si atoms per  $\text{cm}^3$  ( $\times 4$  valence electrons per atom)?

-What if temperature is increased?

		$E_g$ [eV])
Si	$(i/D)$	1.11
$G_0$	$ V/D $	0.67
SiC( $\alpha$ )	IAM	-2.86
AlP	$[i/Z]$	2.45
AlAs	$(i/Z)$	2.16
AlSb	$[i/Z]$	1.6
GaP	$(i/Z)$	2.26
GaAs	$(d/Z)$	1.43
GaN	$d/Z, W$	-3.4
Gasb	$(d/Z)$	0.7
$\ln P$	$(d/Z)$	1.35
inAs	$[d/Z]$	0.36
InSb	$(d/Z)$	0.18
ZnS	$(d/Z, m$	3.6
ZnSe	$(d/Z)$	2.7
ZnTe	$[d/Z)$	2.25
CdS	$(d/W, Z)$	2.42
CdSe	$(d/M$	1.73
CdTe	$(d/Z)$	1.58
PbS	$[i/H)$	0.37
PbSe	$(i/H)$	0.27
PbTe	$(i/H)$	0.29

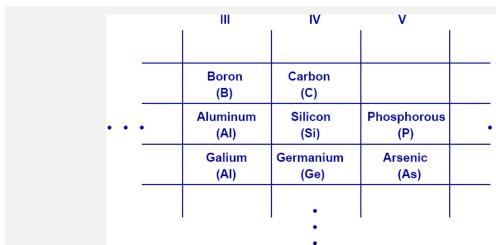


### Adding Dopants: Donors (non-intrinsic Si)

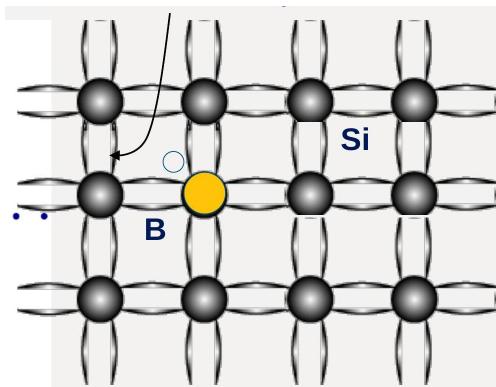
Donors

Group V elements have 5 valence electrons. Therefore, if one Si atom is replaced by one group V atom (e.g. P or As), then 4 electrons are shared among the 4 nearest neighbors which complete the shell, leaving one electron to freely wonder around. Thus each dopant 'donates' an extra electron to the material. Thus the name 'Donors'.

### Dopants: Acceptors



A bond breaks to produce the necessary electron

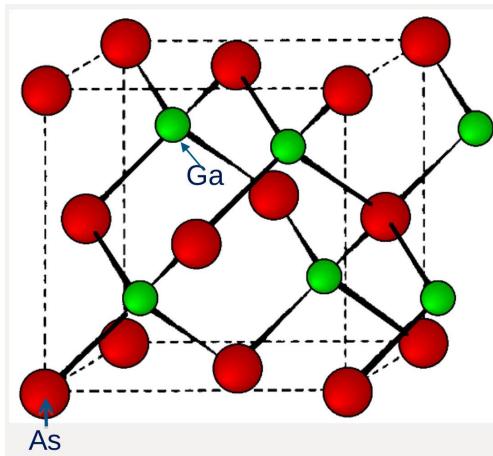


Group III elements have 3 valence electrons. Therefore, if one Si atom is replaced by one group III atom (e.g. B), then 4 electrons are shared among the 4 nearest neighbors which complete the shell, leaving one hole to freely wonder around. Thus each dopant accepts an extra electron leaving one free hole to the material. Thus the name 'Acceptors'.

## Dopants

- When a material is doped with Donors, it has many more mobile electrons than holes. It is called a 'n-type' semiconductor.
- When a material is doped with Acceptors, it has many more mobile holes than electrons. It is called a 'p-type' semiconductor.
- The material is charge neutral. The dopants are charge neutral too. Therefore, even after doping, the material system as a whole remains charge neutral.

### Some examples of other important semiconductors



#### Amphoteric Dopants:

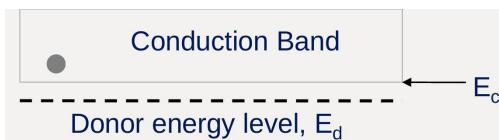
Interestingly, group IV elements are used as dopants. For example, Si replaces smaller Ga atoms which have three valence electrons. Thus Si acts as a donor. Similarly Ge replaces As (5 valence electrons) and acts as an acceptor.

III-V semiconductors such as GaAs are commonly used for high speed and opto-electronics. These materials have similar crystal structure as Si

**What is a reasonable number (to the higher side) of dopants in #/ cm<sup>3</sup> ?**

0.1% of the atomic density

### Dopants in the energy band model



Acceptor energy level,  $E_a$



$$E_c - E_d \sim E_a - E_v \sim 50\text{meV} \sim 2kT$$

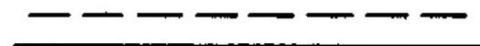
This difference is called the ionization energy.

Due to the ionization energy being close to thermal energy, all dopants are practically ionized

### Dopants and Free Carriers

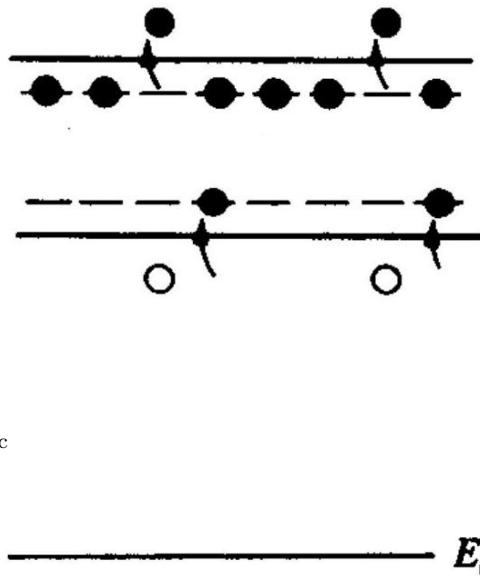


$T \rightarrow 0\text{ K}$



$T \rightarrow 0\text{ K}$

Increasing  $T$   
(a)



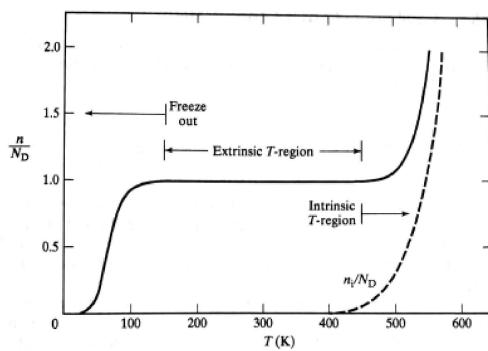
000000  $E_A$   
Room temperature

- Extrinsic (doped) material:
- $n$ -type semiconductor if...
- p-type semiconductor if...

$$\begin{aligned} N_D &\gg n_i \\ N_A &\gg n_i \end{aligned}$$

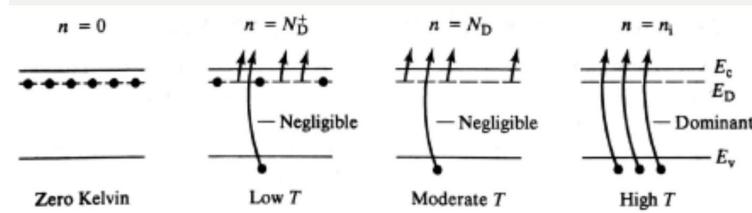
- Ex: what are the electron & hole concentrations in a Si wafer doped with  $10^{15} \text{ cm}^{-3}$  As atoms (at room T). Is this n-or p-type Si?

### Temperature dependence of carrier concentration



(a)

- Assume Si sample doped with  $N_D = 10^{15} \text{ cm}^{-3}$  (n-type)
- How does  $n$  change with  $T$ ? Recall the band diagram, including the donor level.
- Note three distinct regions:
- Low, medium, and hightemperature



(b)

### General Theory of $n$ and $p$

I.  $N_d - N_a \gg n_i$  (i.e., N-type)  $n = N_d - N_a$   
 $p = n_i^2/n$

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

### Compensation Doping

More donors than acceptors  $N_d - N_a \gg n_i$

$$n = N_d - N_a \quad p = \frac{n_i^2}{n} = \frac{n_i^2}{N_d - N_a}$$

More acceptors than donors

$$N_a - N_d \gg n_i$$

$$p_{N_d} = N_a -$$

$$n = \frac{n_i^2}{p} = \frac{n_i^2}{N_a - N_d}$$

## Charge neutrality

- most generally, what are the carrier concentrations in thermal equilibrium, if we have both donor and acceptor doping?

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

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

$$n_0 + N_a = p_0 + N_d$$

$$n_0 + N_a = \frac{n_i^2}{n_0} + N_d$$

$$n_0^2 - (N_d - N_a) n_0 - n_i^2 = 0$$

- And how do these simplify if we have  $N_D \gg N_A$  (n-type doping dominates)?  $n = N_d$   $p = n_i^2/N_d$

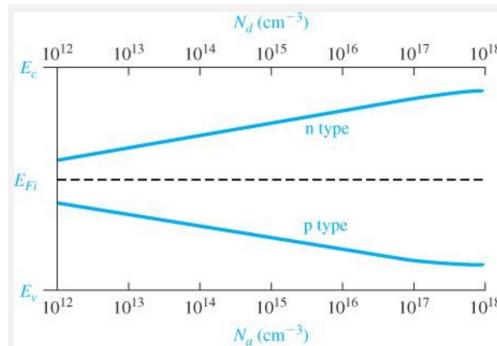


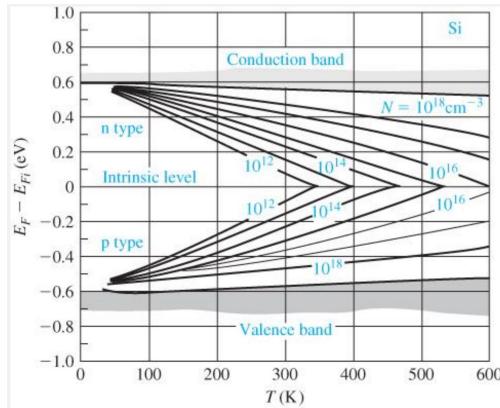
Figure 4.18 I Position of Fermi level as a function of donor concentration (n type) and acceptor concentration (p type).

Figure 4.19 I Position of Fermi level as a function of temperature for various doping concentrations.

(From Sze [14].)

## Problem 4.54

Silicon at  $T = 300$  K contains acceptor atoms at a concentration of  $N_a = 5 \times 10^{15}$  cm<sup>-3</sup>. Donor atoms are added forming an n-type compensated semiconductor such that the Fermi level is 0.215eV below the conductionband edge. What concentration of donor atoms are added?



Summary of key equations for semiconductor carrier concentration calculations:

#### Density of States and Fermi Function

$$g_c(E) = \frac{m_n^* \sqrt{2m_n^*(E - E_c)}}{\pi^2 \hbar^3}, \quad E \geq E_c$$

$$g_v(E) = \frac{m_p^* \sqrt{2m_p^*(E_v - E)}}{\pi^2 \hbar^3}, \quad E \leq E_v$$

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

#### Carrier Concentration Relationships

$$n = N_C \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_c)$$

$$p = N_V \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_v)$$

$$N_{C,V}(T_B) = 2 \left[ \frac{m_p^* k T}{2\pi \hbar^2} \right]^{3/2}$$

$$n_i, \text{ np-Product, and Charge Neutrality} \quad n_i = N_C e^{(E_F - E_c)/kT} \quad n_o = N_C e^{(E_F - E_c)/kT}$$

$$p_o = N_V e^{(E_v - E_F)/kT} \quad n_{i,ve300k} \left( \frac{T_B}{300k} \right)^{3/2} \quad p_i e^{(E_F - E_p)/kT}$$

$$n_i = \sqrt{N_C N_V} e^{-E_G 2kT}$$

$$n p_0 = n_i^2 \quad p_0 = n_0 + N_D - N_A = 0$$

#### $n, p$ , and Fermi Level Computational Relationships

$$n_o = \frac{N_D - N_A}{2} + \left[ \left( \frac{N_D - N_A}{2} \right)^2 + n_i^2 \right]^{1/2} \quad E_i = \frac{E_c + E_v}{2} + \frac{3}{4} kT \ln \left( \frac{m_p^*}{m_n^*} \right)$$

$$n_0 \simeq N_D$$

$$p_0 \simeq n_i^2 / N_D$$

$$N_D \gg N_A, N_D \gg n_i \quad E_F - E_i = kT \ln(n_0/n_i) = -kT \ln(p_0/n_i)$$

$$p_0 \simeq N_A \quad N_A \gg N_D, N_A \gg n_i \quad E_F - E_{fi} = kT \ln(N_D/n_i) \quad N_D \gg N_A, N_D \gg n$$

$$n_0 \simeq n_i^2/N_A \quad E_{fi} - E_F = kT \ln(N_A/n_i) \quad N_A \gg N_D, N_A \gg n$$

Si parameters:  
 $E_g = 1.12\text{eV}$ ,  $n_i = 1.5 \times 10^{10} \text{cm}^{-3}$ ,  $m_e^* = 1.08m_0$ ,  $m_h^* = 0.56m_0$   
 $(@300K)$

 $N_c = 2.3 \times 10^{19} \text{cm}^{-3}$ ,  $N_v = 1.04 \times 10^{19} \text{cm}^{-3}$

$$k = 8.62 \times 10^{-5} \text{eV/k} = 1.38 \times 10^{-23} \text{ J/k}$$

constants:  $h = 4.14 \times 10^{-15} \text{eV} \cdot \text{s} = 6.63 \times 10^{-34} \text{ J} \cdot \text{s}$ ,  $\hbar = \frac{h}{2\pi}$

$$q = e = 1.602 \times 10^{-19} \text{C}$$

$$m_0 = 9.11 \times 10^{-31} \text{ kg}$$

## Next class

- Carrier transport: drift, diffusion

# Lecture 8: Carrier Transport: Drift and Diffusion

---

## Topics

- Neamen Chapter 5
- Drift
- Mobility
- Velocity saturation
- Diffusion
- Total current

## Key Points

### Energy Band Diagram

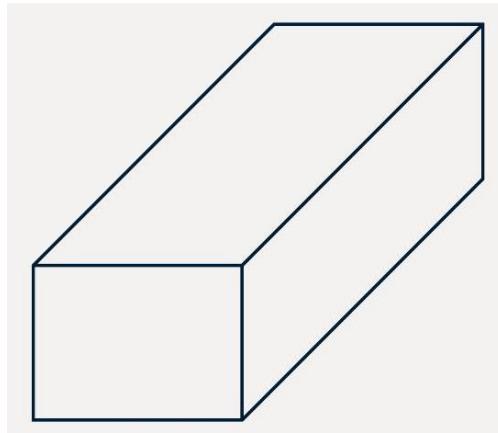
Concept of Fermi level, Donor levels, Acceptor Levels

$$n = N_c e^{(E_F - E_C)/kT} \quad p = N_v e^{(E_V - E_F)/kT}$$

$$(\text{N-type}) \quad n = N_d - N_a$$

$$(\text{P-type}) \quad p = N_a - N_d$$

$$np = n_i^2$$



Review of some basic electrical behavior:

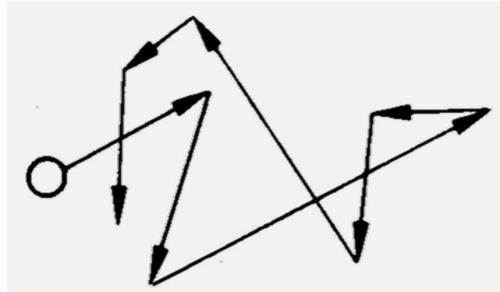
- Ohm's law:  $V = IR$
- Current density:  $J = I/A = \sigma \varepsilon$
- Electric field:  $\mathcal{E} = V/L$
- Resistivity:  $\rho = \frac{RA}{L}$  or  $R = \frac{\rho L}{A}$
- Conductivity:  $\sigma = \frac{1}{\rho}$

Direction of current flow (convention):

- Holes flow in the direction of the electric field, hence  $J_p$  is in the direction of  $\varepsilon$
- Electrons flow opposite to the direction of the electric field, but  $J_n$  still is in the direction of  $\varepsilon$

What happens in zero electric field?

### Thermal Velocity



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

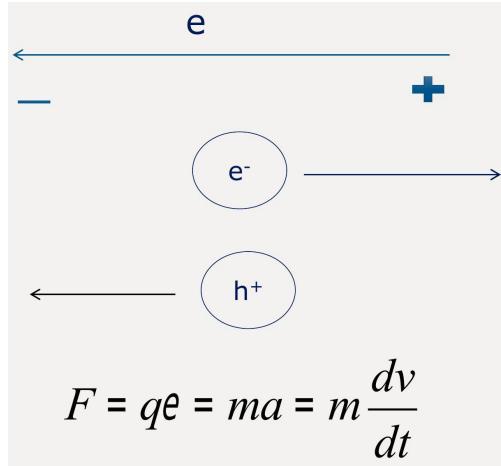
Average electron or hole kinetic energy =  $\frac{3}{2}kT = \frac{1}{2}mv_{th}^2$

$$\begin{aligned} v_{th} &= \sqrt{\frac{3kT}{m_{eff}}} = \sqrt{\frac{3 \cdot 1.38 \cdot 10^{-23} \text{ J K}^{-1} \cdot 300 \text{ K}}{0.26 \cdot 9.1 \cdot 10^{-31} \text{ kg}}} \\ &= 2.3 \cdot 10^5 \text{ m/s} = 2.3 \cdot 10^7 \text{ cm/s} \end{aligned}$$

What happens in an electric field?

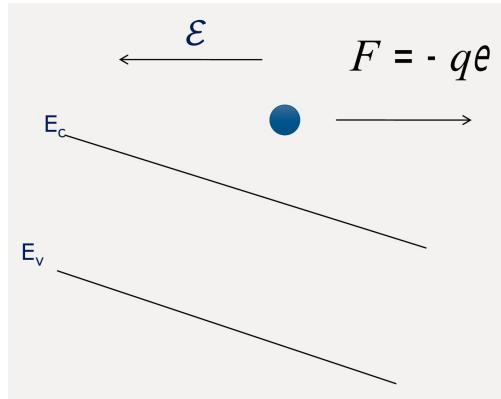
### Drift

When there is an electric field, it gives the carriers a net velocity in the direction of the field.



### Electric Field and Energy Band Diagram

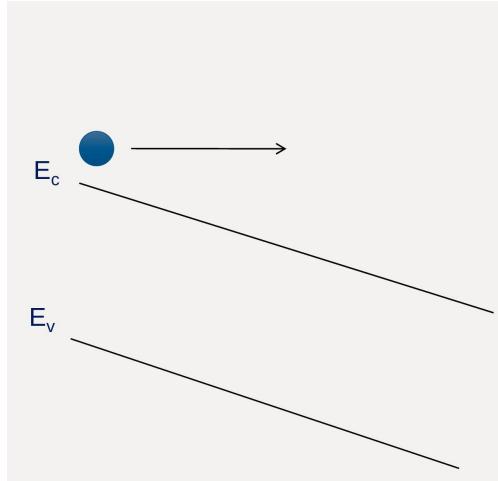
Maxwell's equations dictate that the energy band diagram will bend in presence of an electric field



Electrons gaining more kinetic energy as they go, From Maxwell's equations:

$$\begin{aligned} e &= -\frac{dV}{dx} \\ &= -\frac{1}{q} \frac{d(qV)}{dx} \\ &= \frac{1}{q} \frac{dE_c}{dx} \end{aligned}$$

Remember: Bands are higher where voltage is lower



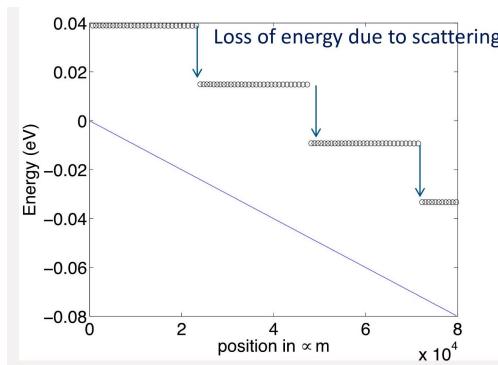
### What happens in an electric field?

$$\begin{aligned} m \frac{dv}{dt} &= qe \\ v &= \frac{qe}{m} t \\ E_{\text{kinetic}} &= \frac{1}{2}mv^2 = \frac{1}{2}m \left( \frac{qe}{m} t \right)^2 \end{aligned}$$

If the kinetic energy is large enough, the electron can scatter strongly by colliding with lattice and leading to a lattice vibration. We say, a phonon has been emitted. In Si, minimum energy needed to emit an optical phonon is 63meV.

Therefore the velocity and kinetic energy cannot increase incessantly.

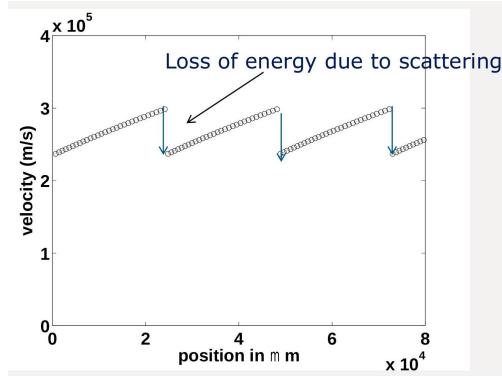
### What happens in an electric field?



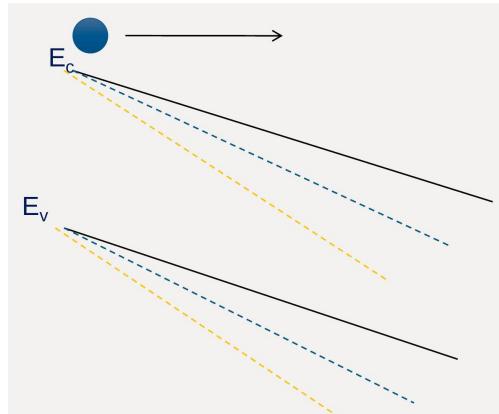
### What happens in an electric field?

The velocity only increases, i.e., acceleration only happens within the time between two scattering events. This time is called the collision time, scattering time or mean free time.

Drift



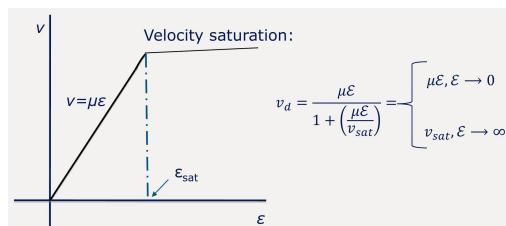
What happens if electric field is continuously increased?



$$v = \frac{qe}{m} t = \zeta \frac{q}{e} \frac{t}{m} \ddot{\theta} = me$$

After some threshold, the electrons always have enough energy to emit a phonon. So any incremental energy goes to emit more phonons and not in increasing velocity. As a result velocity saturates.

## Drift



## Drift

### Electron and hole mobilities

$$\text{Electron mobility } m_n = \frac{q t_m}{m_n^*}$$

$$\text{Hole mobility } m_p = \frac{q t_m}{m_p^*}$$

Mobility: one of the most quoted properties of a semiconductor material to indicate material quality for devices

$$\text{Hole velocity } \vec{v}_h = \mu_p \vec{E}$$

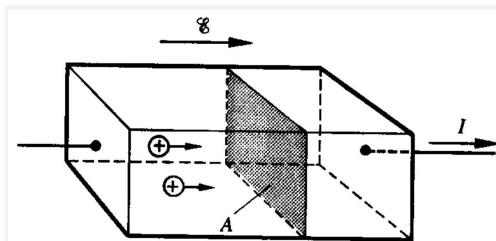
$$\text{Electron velocity } \vec{v}_e = -\mu_n \vec{E}$$

## Drift Current

- Drift current is proportional to the carrier velocity and carrier concentration:

$$\text{Total current } J_{p, \text{drift}} = Q/t$$

$Q$  = total charge contained in the volume shown to the right  $t$  = time taken by  $Q$  to cross the volume



$$Q = qp \text{ (in cm}^3\text{)} \times \text{Volume} = qpAL = qpAv_h t$$

→ Hole current per unit area (i.e. current density)

$$J_{p, \text{drift}} = qp v_h$$

## Conductivity and Resistivity

- In a semiconductor, both electrons and holes conduct current:

$$J_{p, \text{drift}} = qpm_p\varepsilon$$

Hole Current

$$J_{n, \text{drift}} = -qn(-m_n\varepsilon)$$

Electron Current

$$J_{\text{tot,drift}} = J_{p, \text{drift}} + J_{n, \text{drift}} = qpm_p\varepsilon + qnm_n\varepsilon$$

$$J_{\text{tot,drift}} = q(pm_p + nm_n) \quad 0 \leq \varepsilon \quad \sigma \equiv qp\mu_p + qn\mu_n$$

- The conductivity of a semiconductor is  $\rho \equiv \frac{1}{\sigma}$
- Unit: mho/cm
- The resistivity of a semiconductor is
- Unit: ohm-cm

### Mobility of different materials

Electron and hole mobilities of selected semiconductors

	Si	Ge	GaAs	InAs
$\mu_n$ ( cm <sup>2</sup> /V · s)	1400	3900	8500	30000
$\mu_p$ ( cm <sup>2</sup> /V · s)	470	1900	400	500

This is why GaAs is used for making high speed devices. In the recent days, much research is being done on high mobility materials such as InAs, graphene etc.

### Drift Velocity Calculations

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

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

$$\begin{aligned} t_{mp} &= m_p m_p / q = 470 \text{ cm}^2/\text{V} \cdot \text{s} \times 0.39 \times 9.1 \times 10^{-31} \text{ kg}/1.6 \times 10^{-19} \text{ C} \\ &= 0.047 \text{ m}^2/\text{V} \cdot \text{s} \times 2.2 \times 10^{-12} \text{ kg/C} = 1 \times 10^{-13} \text{ s} = 0.1 \text{ ps} \end{aligned}$$

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

$$= 2.2 \times 10^{-6} \text{ cm} = 220 = 22 \text{ nm}$$

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

### More on mobility

#### Mechanisms of Carrier Scattering

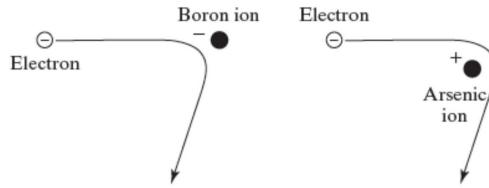
There are two main causes:

1. Phonon Scattering (lattice-induced scattering)
2. Impurity (Dopant) Ion Scattering

Phonon scattering mobility decreases when temperature rises:

$$\mu = qv/m$$

phonon density  $\frac{1}{T}$   
 carrier thermal velocity  $\frac{1}{T} \cdot T^{1/2} \propto T^{-3/2}$   
 $v_{th} \propto T^{1/2}$



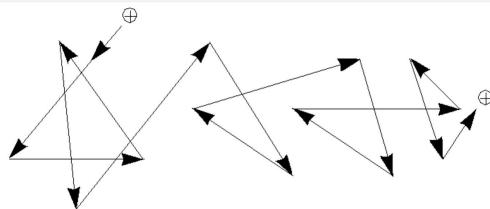
### More on mobility

#### 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.

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

### Putting it all together



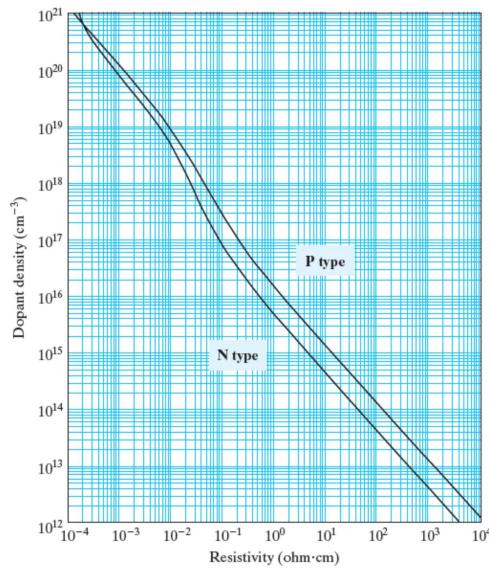
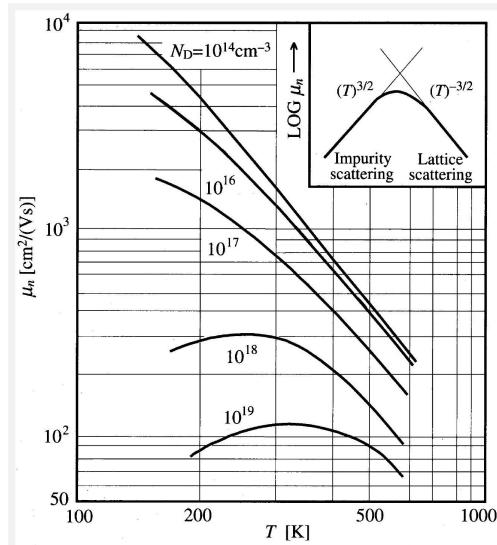
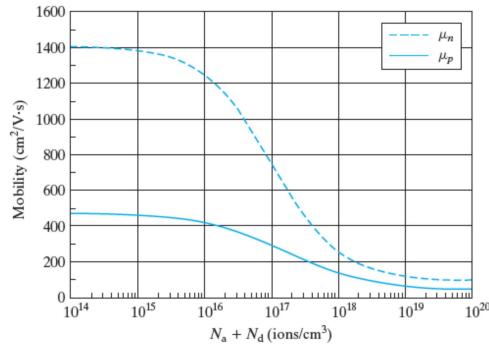
Different scattering processes could be thought of sources of resistances to the motion of carriers. Thus if the resistance due to phonons is  $R_1$  and due to ionized impurities is  $R_2$ , the total resistance can be written as

$$\begin{aligned}
 R &= R_{\text{phonon}} + R_{\text{impurity}} \\
 P \frac{1}{s} &= \frac{1}{s_{\text{phonon}}} + \frac{1}{s_{\text{impurity}}} \\
 P \frac{1}{m} &= \frac{1}{m_{\text{phonon}}} + \frac{1}{m_{\text{impurity}}}; \quad \frac{1}{t} = \frac{1}{t_{\text{phonon}}} + \frac{1}{t_{\text{impurity}}}
 \end{aligned}$$

Known as Matthiessen's rule

### Dependence of mobility of doping

Dependence of mobility on temperature

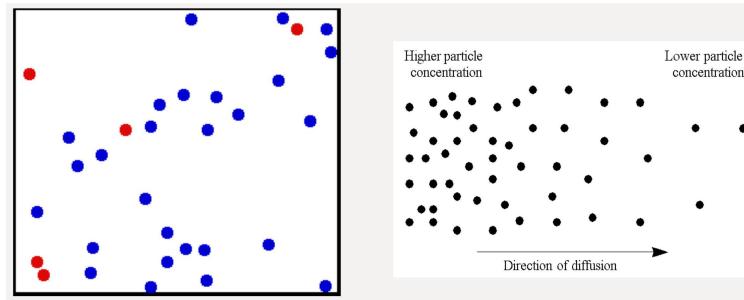


### Dopant density and resistivity

The last three plots are not exactly consistent! This comes from difficulty in measuring mobility accurately.

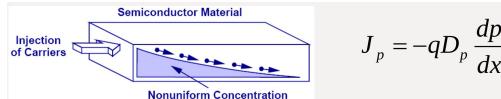
## Diffusion

Diffusion: Due to thermally induced random motion, mobile particles tend to move from a region of high concentration to a region of low concentration.



## Diffusion current

- Current flow due to mobile charge diffusion is proportional to the carrier concentration gradient.
- The proportionality constant is the diffusion constant.



Notation:

$D_p \equiv$  hole diffusion constant ( $\text{cm}^2/\text{s}$ )

$D_n \equiv$  electron diffusion constant ( $\text{cm}^2/\text{s}$ )

## Diffusion current

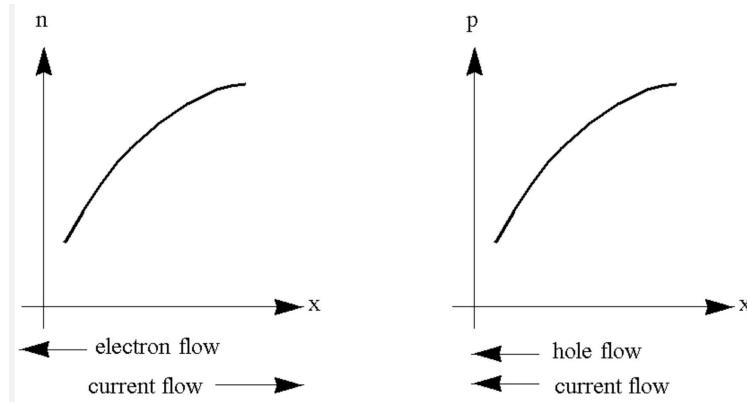
$$J_{n, \text{diffusion}} = qD_n \frac{dn}{dx} \quad J_{p, \text{diffusion}} = -qD_p \frac{dp}{dx}$$

$D$  is called the diffusion constant. Signs explained:

## Review of four current components

$$\begin{aligned} J_{\text{TOTAL}} &= J_n + J_p \\ J_n &= J_{n, \text{drift}} + J_{n, \text{diffusion}} = qn\mu_n E + qD_n \frac{dn}{dx} \\ J_p &= J_{p, \text{drift}} + J_{p, \text{diffusion}} = qp\mu_p E - qD_p \frac{dp}{dx} \end{aligned}$$

Summary of key equations for semiconductor carrier concentration calculations:



### Density of States and Fermi Function

$$g_c(E) = \frac{m_n^* \sqrt{2m_n^*(E - E_c)}}{\pi^2 \hbar^3}, \quad E \geq E_c$$

$$g_v(E) = \frac{m_p^* \sqrt{2m_p^*(E_v - E)}}{\pi^2 \hbar^3}, \quad E \leq E_v$$

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

### Carrier Concentration Relationships

$$n = N_C \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_c)$$

$$p = N_V \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_v)$$

$$N_{C,V}(T_B) = 2 \left[ \frac{m^*_p k T}{2\pi \hbar^2} \right]^{3/2}$$

$n_i$ , np-Product, and Charge Neutrality

$$n_i = \sqrt{N_C N_v} e^{-E_G 2kT}$$

$$N_C = 2 \left[ \frac{m_n^* k T}{2\pi \hbar^2} \right]^{3/2}$$

$$p_o = N_V e^{(E_v - E_F)/kT}$$

$$n_{c,ve300k} \left( \frac{T_B}{300k} \right)^{3/2}$$

$$n_i e^{(E_i - E_F)/kT}$$

$$n_o = N_C e^{(E_F - E_c)/kT}$$

$$p_i e^{(E_F - E_p)/kT}$$

$$np_0 = n_i^2 \quad p_0 - n_0 + N_D - N_A = 0$$

### $n, p$ , and Fermi Level Computational Relationships

$$n_o = \frac{N_D - N_A}{2} + \left[ \left( \frac{N_D - N_A}{2} \right)^2 + n_i^2 \right]^{1/2} \quad E_i = \frac{E_c + E_v}{2} + \frac{3}{4} k T \ln \left( \frac{m_p^*}{m_n^*} \right)$$

$$n_0 \simeq N_D$$

$$p_0 \simeq n_i^2 / N_D$$

$$N_D \gg N_A, N_D \gg n_i \quad E_F - E_i = kT \ln(n_0/n_i) = -kT \ln(p_0/n_i)$$

$$p_0 \simeq N_A \quad N_A \gg N_D, N_A \gg n_i \quad E_F - E_{fi} = kT \ln(N_D/n_i) \quad N_D \gg N_A, N_D \gg n$$

$$n_0 \simeq n_i^2 / N_A \quad E_{fi} - E_F = kT \ln(N_A/n_i) \quad N_A \gg N_D, N_A \gg n$$

Si parameters:  
 $E_g = 1.12 \text{ eV}$ ,  $n_i = 1.5 \times 10^{10} \text{ cm}^{-3}$ ,  $m_e^* = 1.08 m_0$ ,  $m_h^* = 0.56 m_0$   
 $(@300\text{K})$   
 $N_C = 2.3 \times 10^{19} \text{ cm}^{-3}$ ,  $N_V = 1.04 \times 10^{19} \text{ cm}^{-3}$

$$k = 8.62 \times 10^{-5} \text{ eV/k} = 1.38 \times 10^{-23} \text{ J/k}$$

constants:  $h = 4.14 \times 10^{-15} \text{ eV} \cdot \text{s} = 6.63 \times 10^{-34} \text{ J} \cdot \text{s}$ ,  $\hbar = \frac{h}{2\pi}$

$$q = e = 1.602 \times 10^{-19} \text{ C}$$

$$m_0 = 9.11 \times 10^{-31} \text{ kg}$$

## Next class

- Current continuity, band diagrams

# Lecture 9: Band diagrams

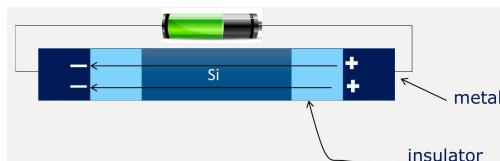
## Topics

- Neamen Chapter 5
- Einstein's relation
- Fermi level at equilibrium
- Information from band diagrams
- Revision of concepts

## Near-Equilibrium Concepts

### Fermi Level at Equilibrium

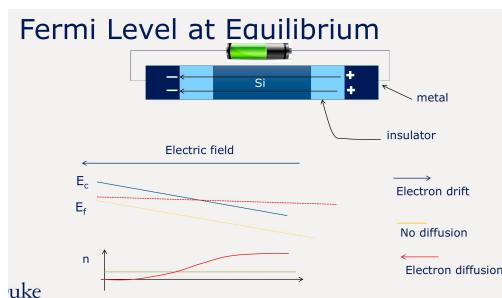
- At equilibrium, there is no exchange of carriers  $\rightarrow$  any type. So current carried by electrons and holes is individually zero
- If the current has to be zero, the drift and diffusion components for electrons and holes must individually cancel each other.



Electric field

$$E_c \quad e = \frac{1}{q} \frac{dE_c}{dx}$$

How does the  $E_f$  look like?



## Fermi Level at Equilibrium

### Remember:

- An electric field acts on the potential; Fermi level is determined by chemical potential. So having an electric field is not sufficient to do anything to the alignment of the Fermi level.
- When a current flows, electrons can come from outside and go to outside of the material → this outside is a battery where a specific chemical potential difference is maintained. Therefore, only when a current is flowing, there is a change in the Fermi level. If there is no current, Fermi level will be flat throughout the device

### Einstein's Relationship between D and $\mu$

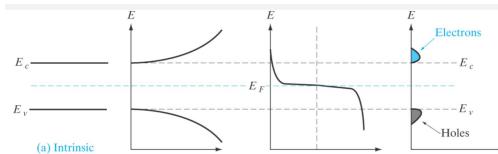
No current flows at equilibrium. The electron and hole currents are individually zero. This allows to draw a relationship between the diffusion constant and mobility at equilibrium.

$$\frac{J_p = qpm_p e - qD_p \frac{dp}{dx} = 0 \quad \frac{dp}{dx}}{\frac{p \frac{D_p}{m_p} = \frac{pe}{dp/dx} = \frac{1}{kT} \left( \frac{dE_V}{dx} - \frac{dE_f}{dx} \right) N_V e^{(E_V - E_f)/kT}}{p \frac{D_p}{m_p} = \frac{kT}{q} = \frac{q}{kT} ep}} = N_V \frac{d}{dx} e^{(E_V - E_f)/kT}$$

### Einstein's Relationship

- Strictly applicable at equilibrium. But we still use it when current is flowing saying the situation is near or quasi-equilibrium
- No longer valid for very small devices where quasi equilibrium is not valid

### Review



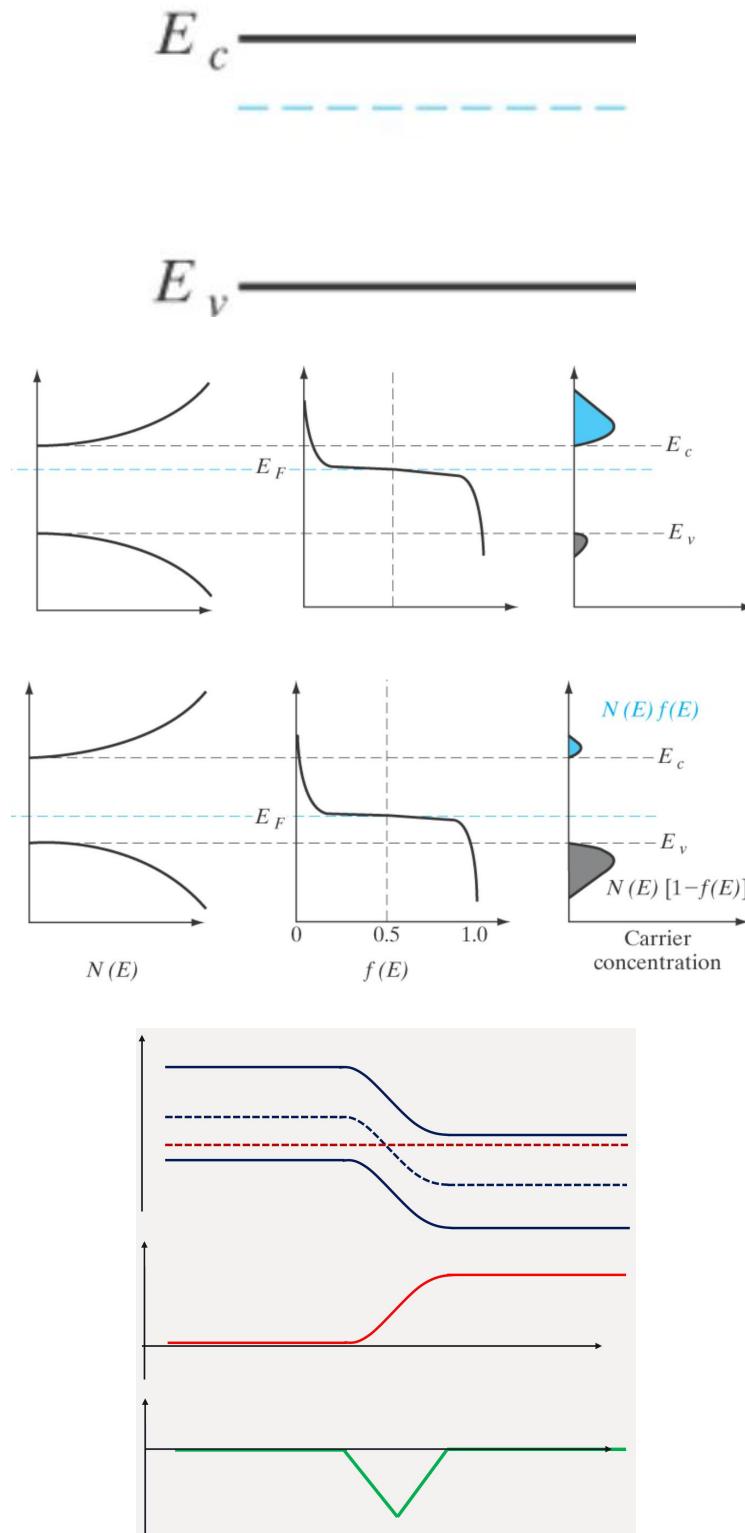
(b) n-type

## Band diagrams

### Example

A semiconductor has a bandgap of 1eV, and effective density-of-states,  $N = 10^{19} \text{ cm}^{-3}$

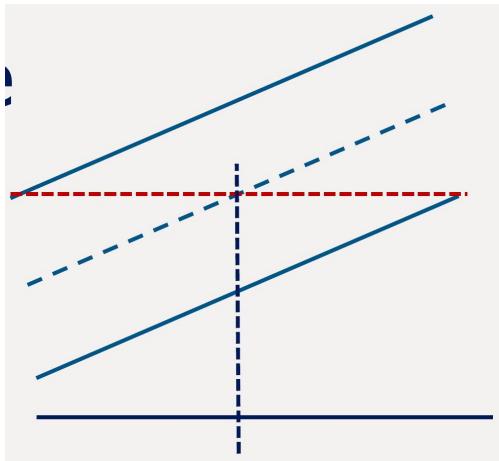
$N_y = 4 \times 10^{19} \text{ cm}^{-3}$  electron and hole mobilities of  $= 4000$  and  $2500 \text{ cm}^2/\text{V} - \text{s}$ , respectively. It is subjected to the following potentials at the various locations as follows (assume linear variation of potentials between locations):



Point A at  $x = 0$  microns,  $V = 0V$ ;

Point B at  $x = 2$  microns,  $V = -2V$

Point C at 4 microns,  $V = +4V$



Point  $D$  is at 8 microns; electric field is zero between  $C$  and  $D$ .

Sketch the simplified band diagram, properly labeling the positions, energies, and directions of electric fields. If the electron concentration at location  $B$  is  $10^{18} \text{ cm}^{-3}$ , and assuming things are close to equilibrium, what is the hole concentration there? If an electron at the conduction band edge at  $B$  goes towards  $C$ , and how long does it take to get there? If there is negligible scattering at low electric fields, how long does it take to go from  $C$  to  $D$ ?

## Problem 2

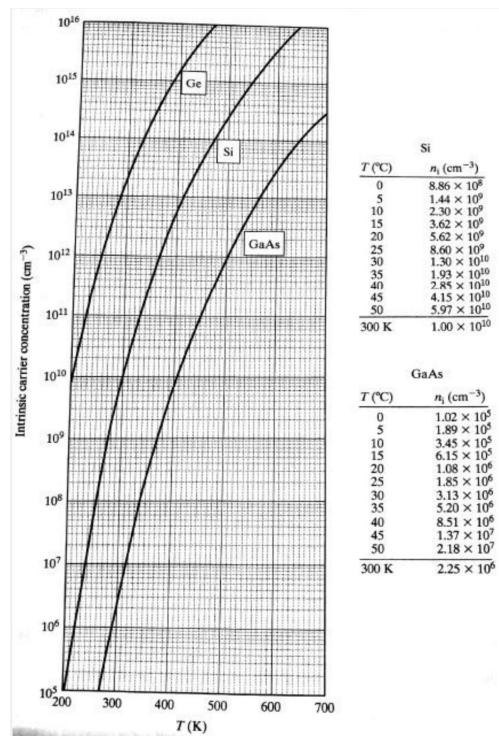
The electron concentration in a silicon material is  $5 \times 10^2 / \text{cm}^3$  at room temperature under equilibrium conditions. (a) What is the hole concentration? (b) Where is EF positioned relative to  $E_i$ ? (c) Draw the energy band diagram for the material.

(d) Repeat for parts (a), (b) and (c) for the same sample if the temperature is raised to 400 K. Refer to chart at the back to obtain the intrinsic carrier concentration at 400 K. Also, the band gap energy is reduced to 1.08eV at 400 K.

Quiz

## Next class

- P-n junctions



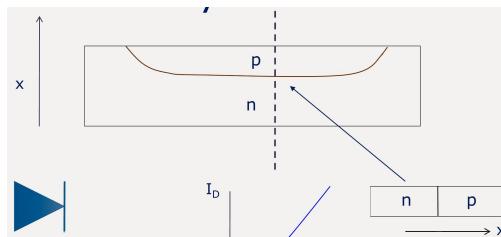
# Lecture 10: pn junctions in equilibrium

---

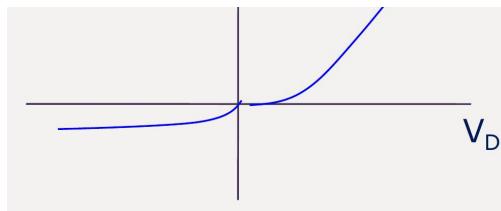
## Topics

- Neamen Chapter 7
- Built-in potential,  $V_{bi}$
- Electric field
- Space charge width
- One-sided junction

## Qualitative Analysis



Circuit Symbol



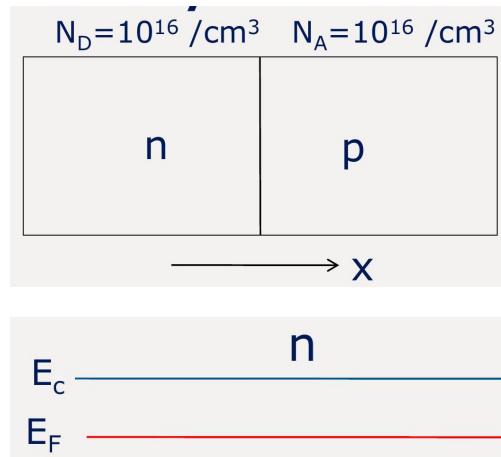
Rectifier: Basis of all devices

We shall use this structure for all our analysis

## Qualitative Analysis

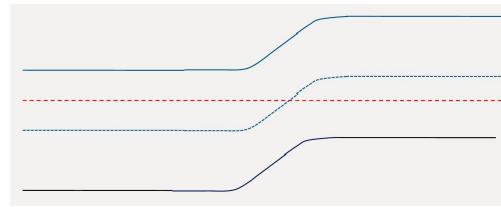
Assumptions:

1. Abrupt junctions (step junction)
2. Uniform doping
3. Non-degenerate (Boltzmann approximation holds)
4. Fully ionized dopants (  $T \sim 300$  K )



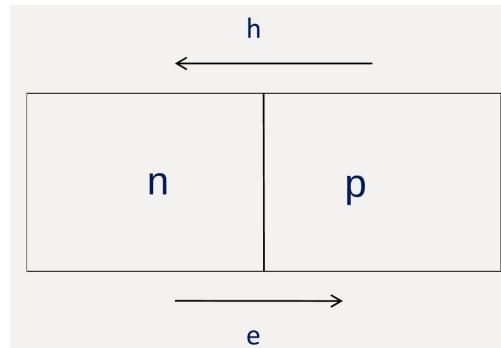
### Qualitative Analysis

$V_{bi}$  in energy band diagram



$$\begin{aligned} V_{bi} &= (E_{Fn} - E_i) + (E_i - E_{Fp}) = kT \ln \frac{n}{n_i} + kT \ln \frac{p}{n_i} \\ &= kT \ln \frac{N_D}{n_i} + kT \ln \frac{N_A}{n_i} = kT \ln \frac{N_A N_D}{n_i^2} \end{aligned}$$

### Qualitative Analysis



### Qualitative Analysis: Depletion Region

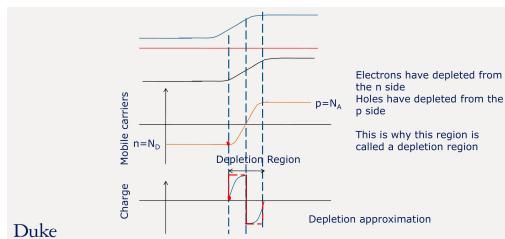
together ensure charge neutrality



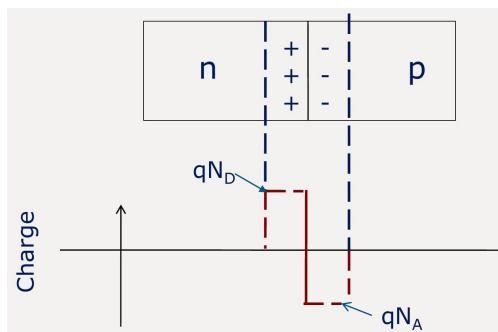
- The Acceptor ions and holes together ensure charge neutrality
- In the region where  $n < N_D$  and  $p < N_A$ , a net charge will exist. This region is called the space charge region or depletion region

- Note that the total system is still charge neutral

### Qualitative Analysis: Depletion Approximation



### Qualitative Analysis: Depletion Region



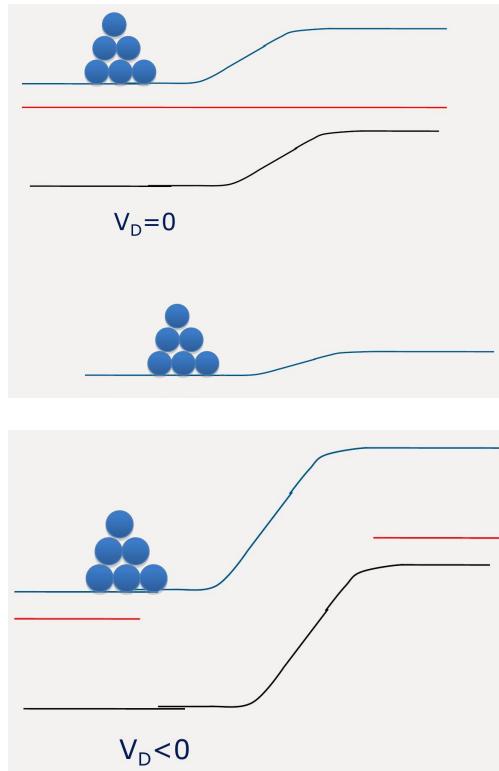
The uncompensated charges at the junction provide the electric field. Since these dopant ions are fixed in the lattice and cannot move, this electric field (the built in field) is always present.

### How does current flow?

$$V_D > 0$$

When negative terminal is connected to 'n' side and positive terminal is connected to 'p' side, we call it a forward bias

- External Field oppose the built in Field.
- Number of carriers go up that can diffuse go up exponentially.
- Thus current goes up exponentially  $I_D \sim \exp(V_D)$



### What happens in the reverse bias?

When negative terminal is connected to 'p' side and positive terminal is connected to 'n' side, we call it a negative bias

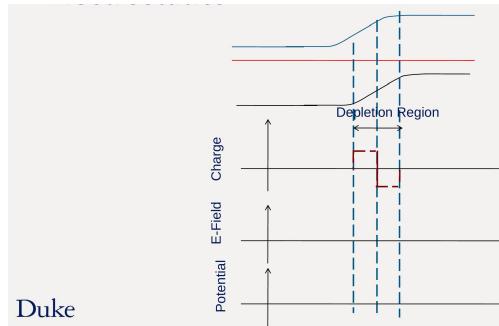
### Recap

- Diffusion of carriers at the junction gives the depletion region.
- Depletion of carriers leave behind uncompensated dopant ions which give built in electric field in the depletion region
- This built in field eventually stops further diffusion.
- Depletion region is charged and usually the depletion approximation is used to describe the charge profile

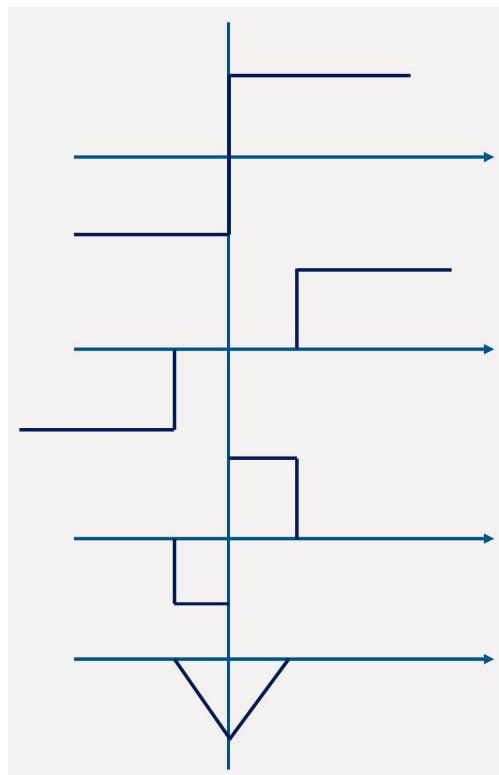
### Electrostatics

#### Electrostatics

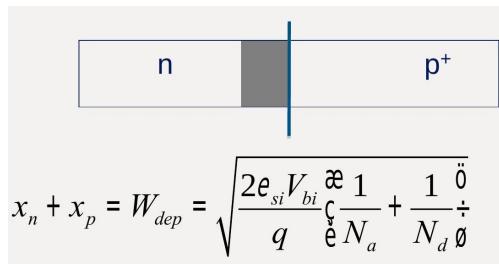
- Poisson's equation:  $\frac{d^2\Phi}{dx^2} = -\frac{\rho}{\epsilon_S} = -\frac{q}{\epsilon_S} (p - n + N_D - N_A)$
- Depletion approximation
- Abrupt junction:  $\frac{d^2\Phi}{dx^2} = -\frac{q}{\epsilon_S} (N_D - N_A)$



## Electrostatics



## One Sided Junction



If  $N_a \gg N_d$ , as in a P+N junction,

$$W_{dep} = \sqrt{\frac{2e_s V_{bi}}{qN_d}} \gg x_n \longrightarrow \text{What about a N N\textsuperscript{+}P junction?}$$

$$x_p = x_n N_d / N_a @ 0$$

$$W_{dep} = \sqrt{2e_s V_{bi} / qN} \text{ where } \frac{1}{N} = \frac{1}{N_d} + \frac{1}{N_a} \gg \frac{1}{\text{lighter dopant density}}$$

### Example

A  $P^+N$  junction has  $N_a = 10^{20} \text{ cm}^{-3}$  and  $N_d = 10^{17} \text{ cm}^{-3}$ . What is a) its built in potential,

b)  $W_{dep}$ , c)  $x_n$ , and d)  $x_p$ ?

Solution:

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

$$\text{b) } W_{dep} \sqrt{\frac{2e_s V_{bi}}{qN_d}} = \frac{\pi_2' 12' 8.85' 10^{-14'} 10^{-1/2} \dot{\phi}}{1.6' 10^{-19'} 10^{17}} = 0.12 \text{ mm}$$

$$\text{c) } x_n \approx W_{dep} = 0.12 \mu\text{m}$$

$$\text{d) } x_p = x_n N_d / N_a = 1.2 \times 10^{-4} \mu\text{m} = 1.2 \approx 0$$

# Lecture 11: p n junctions - under bias

---

## Topics

- Neamen Chapter 7-8
- Reverse bias band diagram
- Junction capacitance
- Junction breakdown
- Forward bias
- Current in forward bias
- Diode current equation

## Recap

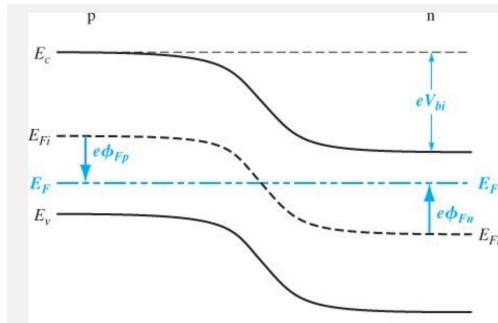


Figure 7.3 I Energy-band diagram of a pn junction in thermal equilibrium.

$$V_{bi} = kT \ln \frac{N_A N_D}{n_i^2}$$

Figure 7.6 I Electric potential through the space charge region of a uniformly doped pn junction.

### pn junctions in equilibrium - equations

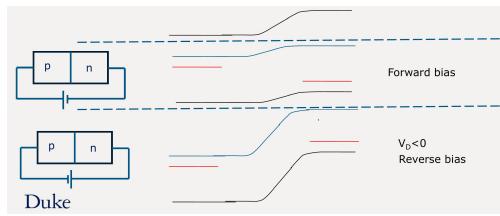
$$V_{bi} = \frac{q}{2\epsilon_s} (N_D x_n^2 + N_A x_p^2) = kT \ln \frac{N_A N_D}{n_i^2}$$

$$x_n = \sqrt{\frac{2\epsilon_s}{q} V_{bi} \frac{N_A}{N_D(N_A+N_D)}} \quad x_p = \sqrt{\frac{2\epsilon_s}{q} V_{bi} \frac{N_D}{N_A(N_A+N_D)}}$$

$$W_{dep} = x_n + x_p = \sqrt{\frac{2\epsilon_s}{q} V_{bi} \left( \frac{1}{N_A} + \frac{1}{N_D} \right)}$$

### Biased p-n junction

$$V_D = 0$$



### Biasing a p-n junction: reverse bias

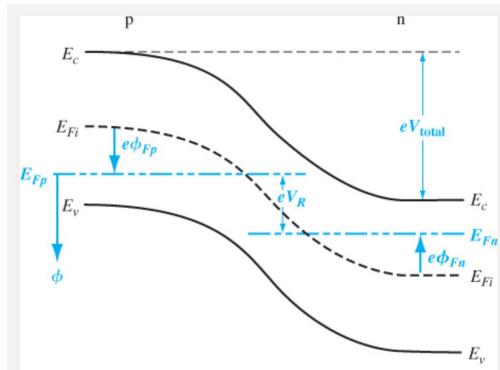


Figure 7.7 I Energy-band diagram of a pn junction under reverse bias.

$$V_{total} = |\Phi_{Fn}| + |\Phi_{Fp}| + V_R$$

$$V_{total} = V_{bi} + V_R$$

$$W_{dep} = \sqrt{\frac{2\epsilon_s}{q} (V_{bi} + V_R) \left( \frac{1}{N_A} + \frac{1}{N_D} \right)}$$

Maximum electric field still occurs at the metallurgical junction:

$$\varepsilon_{max} = -\frac{2(V_{bi} + V_R)}{W_{dep} (\text{reverse-bias})}$$

### Exercise

- The maximum electric field in a reverse-biased GaAs pn junction at T = 300 K is to be limited to  $|\varepsilon_{\max}| = 7.2 \times 10^4$  V/cm. The doping concentrations are  $N_d = 5 \times 10^{15}$  cm $^{-3}$  and  $N_a = 3 \times 10^{16}$  cm $^{-3}$ . Determine the maximum reverse-biased voltage that can be applied.

$$\varepsilon_{\max} = -\frac{2(V_{bi} + V_R)}{W_{dep} \text{ (reverse-bias)}}$$

$$W_{dep} = \sqrt{\frac{2\epsilon_s}{q}(V_{bi} + V_R) \left( \frac{1}{N_A} + \frac{1}{N_D} \right)}$$

### Junction capacitance

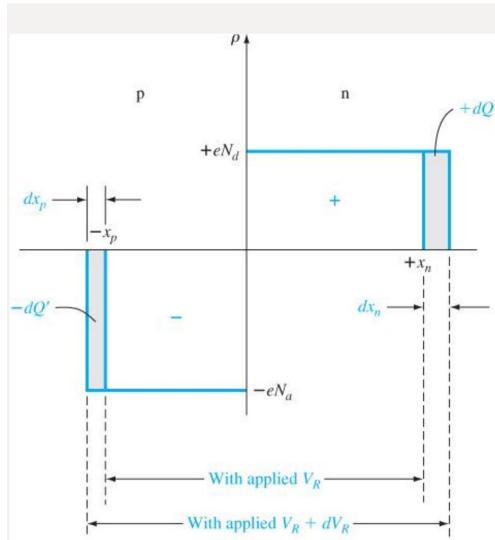


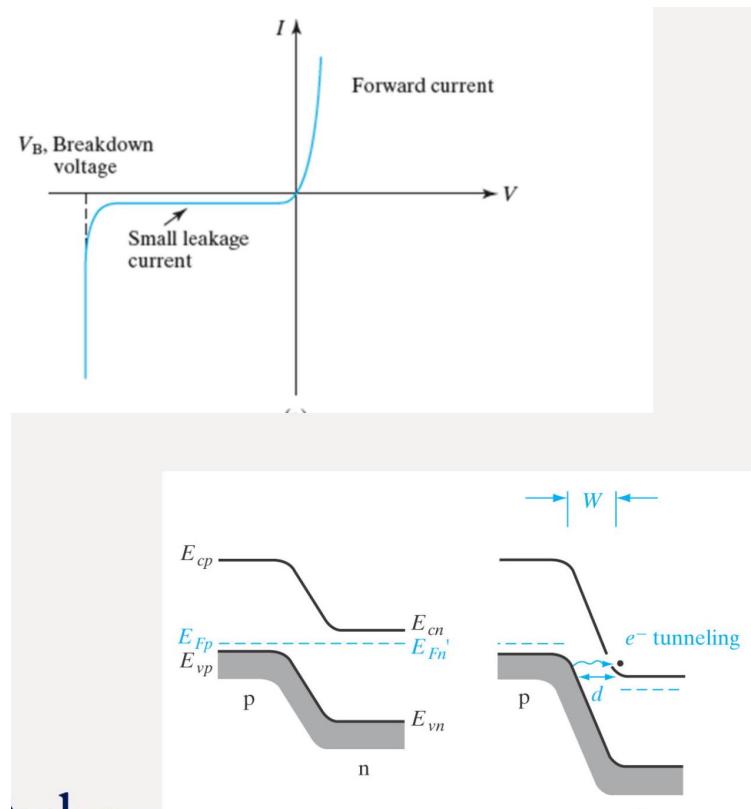
Figure 7.9 I Differential change in the space charge width with a differential change in reverse-biased voltage for a uniformly doped pn junction.

- The junction capacitance is also referred to as the depletion layer capacitance.
- Equation resembles parallel plate capacitor

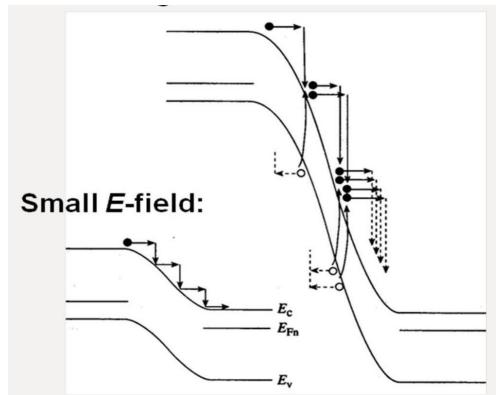
### Junction breakdown

Reverse bias current is due to minority carrier drift - Hence small  
With increasing reverse bias, electrons gain sufficient energy, collide with lattice atoms

- Knocks off more electrons - electron-hole pairs created
- Avalanche effect - more electron-hole pairs generated



- Avalanche breakdown occurs
  - Zener tunneling
- High  $E$ -field:



## Overview of different bias conditions

- (a)
- (b)
- (c)

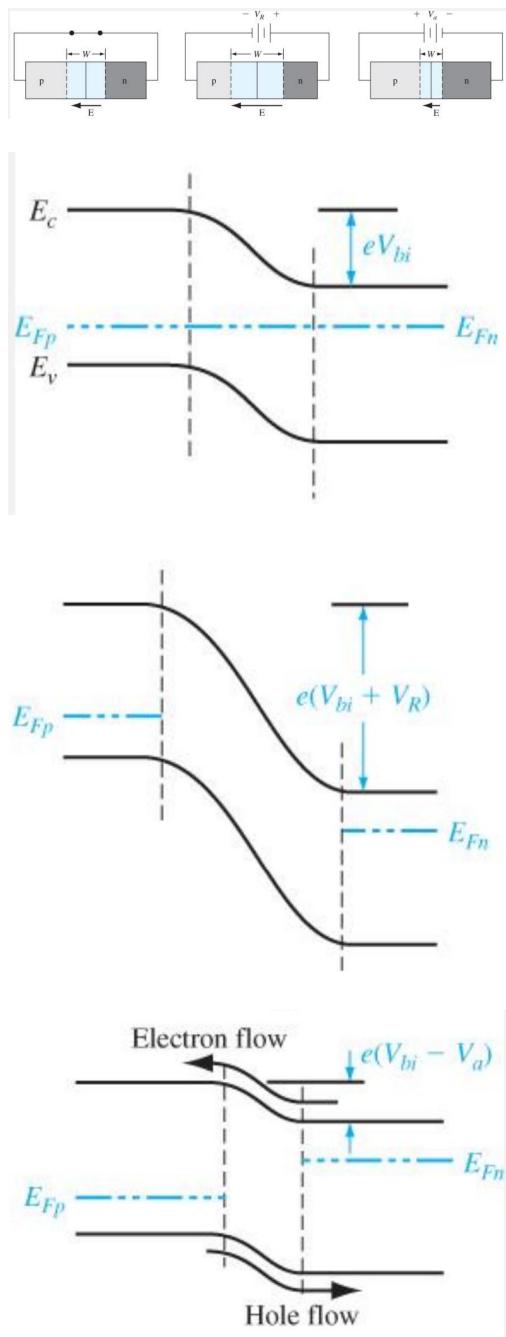


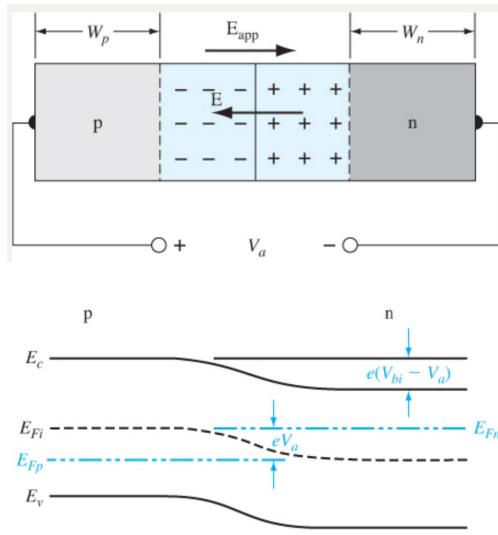
Figure 8.1 | A pn junction and its associated energy-band diagram for (a) zero bias, (b) reverse bias, and (c) forward bias.

### Focusing on forward bias

(a)

(b)

Figure 8.3 I (a) A pn junction with an applied forward-bias voltage showing the directions of the electric field induced by  $V_a$  and the space charge electric field. (b) Energy-band diagram



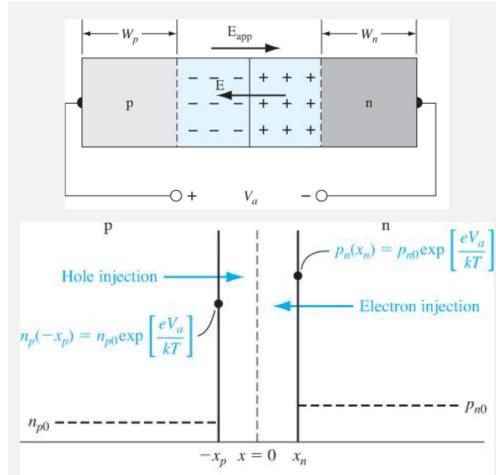
of the forward-biased pn junction.

### Current in forward bias

Table 8.1 I Commonly used terms and notation for this chapter

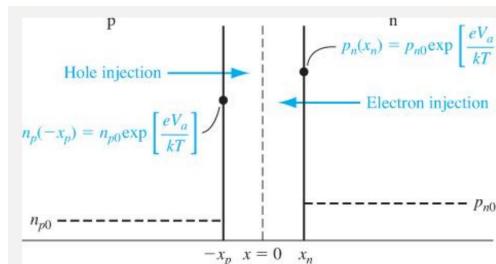
Term	Meaning
$N_a$	Acceptor concentration in the p region of the pn junction
$N_d$	Donor concentration in the n region of the pn junction
$n_{n0} = N_d$	Thermal-equilibrium majority carrier electron concentration in the n region
$p_{p0} = N_a$	Thermal-equilibrium majority carrier hole concentration in the p region
$n_{p0} = n_i^2/N_a$	Thermal-equilibrium minority carrier electron concentration in the p region
$p_{n0} = n_i^2/N_d$	Thermal-equilibrium minority carrier hole concentration in the n region
$n_p$	Total minority carrier electron concentration in the p region
$p_n$	Total minority carrier hole concentration in the n region
$n_p(-x_p)$	Minority carrier electron concentration in the p region at the space charge edge
$p_n(x_n)$	Minority carrier hole concentration in the n region at the space charge edge
$\delta n_p = n_p - n_{p0}$	Excess minority carrier electron concentration in the p region
$\delta p_n = p_n - p_{n0}$	Excess minority carrier electron concentration in the p region

Figure 8.4 I Excess minority carrier concentrations at the space charge edges generated by the forward-bias voltage.



### Exercise

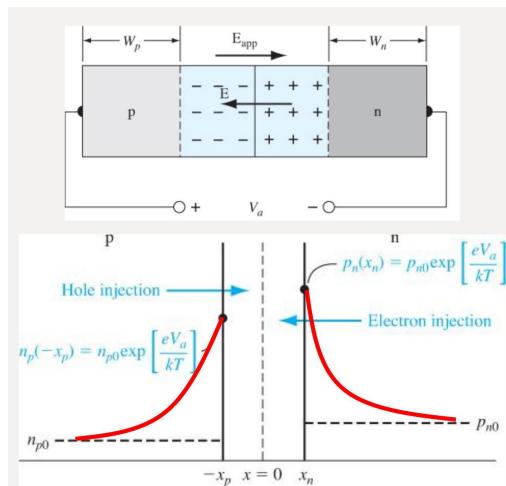
A silicon pn junction at  $T = 300$  K is doped with impurity concentrations of  $N_d = 2 \times 10^{16} \text{ cm}^{-3}$  and  $N_a = 5 \times 10^{16} \text{ cm}^{-3}$ . The junction is forward biased at  $V_a = 0.650 \text{ V}$ . Determine the minority carrier concentrations at the space charge edges. Does low injection still apply?



### Current in forward bias

Table 8.1 I Commonly used terms and notation for this chapter

Term	Meaning
$N_a$	Acceptor concentration in the p region of the pn junction
$N_d$	Donor concentration in the n region of the pn junction
$n_{n0} = N_d$	Thermal-equilibrium majority carrier electron concentration in the n region
$p_{p0} = N_a$	Thermal-equilibrium majority carrier hole concentration in the p region
$n_{p0} = n_i^2/N_a$	Thermal-equilibrium minority carrier electron concentration in the p region
$p_{n0} = n_i^2/N_d$	Thermal-equilibrium minority carrier hole concentration in the n region
$n_p$	Total minority carrier electron concentration in the p region
$p_n$	Total minority carrier hole concentration in the n region
$n_p(-x_p)$	Minority carrier electron concentration in the p region at the space charge edge
$p_n(x_n)$	Minority carrier hole concentration in the n region at the space charge edge
$\delta n_p = n_p - n_{p0}$	Excess minority carrier electron concentration in the p region
$\delta p_n = p_n - p_{n0}$	Excess minority carrier hole concentration in the n region



### Forward bias band diagram

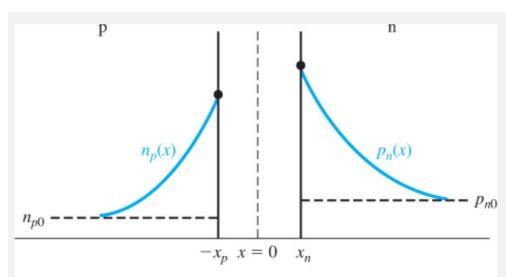


Figure 8.5 I Steady-state minority carrier concentrations in a pn junction under forward bias.

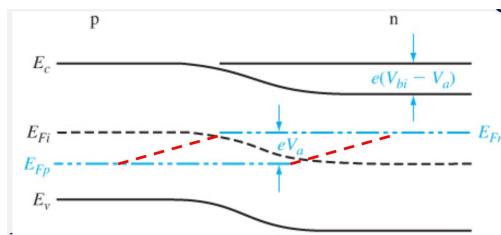
$$n_p(-x_p) = n_{p0} \exp\left(\frac{qV_a}{kT}\right) \quad p_n(x_n) = p_{n0} \exp\left(\frac{qV_a}{kT}\right)$$

Excess minority carrier concentrations:

$$\delta n_p(x) = n_p(x) - n_{p0} = n_{p0} \exp\left(\frac{qV_a}{kT} - 1\right) \exp\left(\frac{x_p + x}{L_n}\right)$$

$$\delta p_n(x) = p_n(x) - p_{n0} = p_{n0} \exp\left(\frac{qV_a}{kT} - 1\right) \exp\left(\frac{x_n - x}{L_p}\right)$$

### Quasi Fermi levels

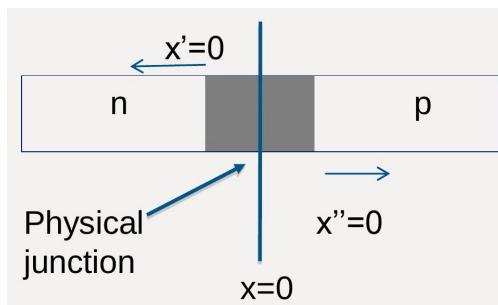


$$E_{Fn} = E_{Fi} + kT \ln\left(\frac{n}{n_i}\right)$$

$$E_{Fp} = E_{Fi} - kT \ln\left(\frac{p}{n_i}\right)$$

$$n_p(-x_p) p_n(x_n) = n_i^2 \exp\left(\frac{E_{Fn} - E_{Fp}}{kT}\right)$$

### Current flow in a $p - n$ junction



We know two facts:

- (1) Electrons will diffuse from n-side and give a current
- (2) Holes will diffuse from the p-side and give another component of the current

In addition

Current inside the depletion region is constant

Total current is constant - individual hole and electron currents are continuous functions and constant in a depletion region.

Note that the diffusion process is injecting minority carriers (electrons into the p-side and holes into the n side). This is known as the minority carrier injection

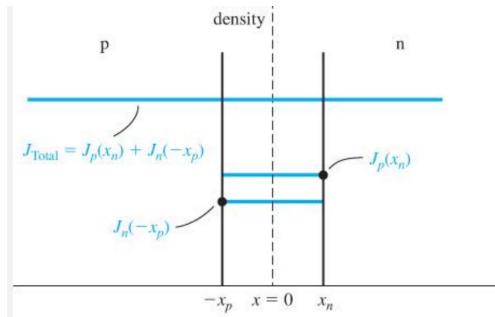


Figure 8.7 | Electron and hole current densities through the space charge region of a pn junction.

### Current flow in a p-n junction

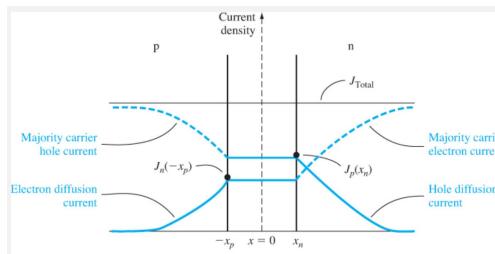


Figure 8.10 I Ideal electron and hole current components through a pn junction under

$$J_p(x_n) = -qD_p \frac{dp_n}{dx} \text{ at } x = x_n$$

$$J_n(-x_p) = qD_n \frac{dn_p}{dx} \text{ at } x = -x_p$$

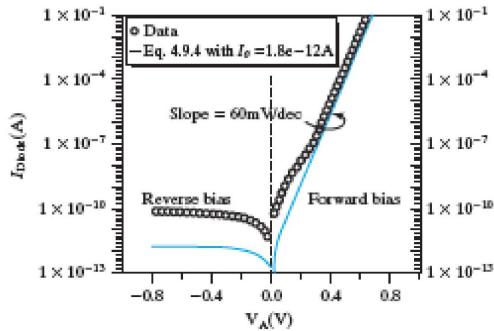
$$J_{\text{ideal}} = \left[ \frac{qD_p p_{n_0}}{L_p} + \frac{qD_n n_{p_0}}{L_n} \right] \left( \exp \left( \frac{qV_a}{kT} \right) - 1 \right)$$

$J_s$  = reverse saturation current forward bias.

\section\*{Current-voltage characteristics}

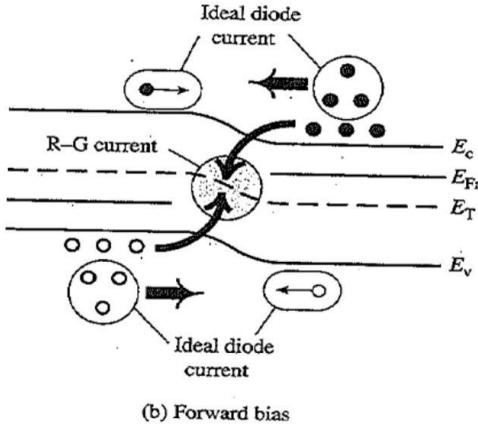
$$= J \text{ at all } x$$

$$J(0) = J_{pN}(0) + J_{nP}(0) = \frac{q}{L_p} D_p p_{n0} + q \frac{D_n}{L_n} n_{p0} \frac{\partial}{\partial} e^{qV_D/kT} - 1)$$



### Recombination-generation current

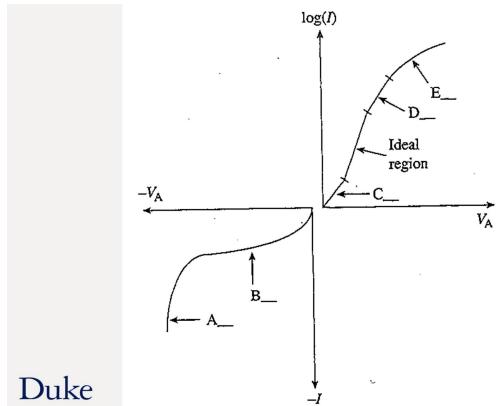
- Some electrons and holes crossing the depletion region can recombine at a rate:
- $R = \frac{np - n_i^2}{\tau_{p0}(n+n') + \tau_{n0}(p+p')}$
- $R_{\max} = \frac{n_i}{2\tau_0} \exp\left(\frac{qV_a}{2kT}\right)$
- $J_{\text{rec}} = \frac{qWn_i}{2\tau_0} \exp\left(\frac{qV_a}{2kT}\right) = J_{r0} \exp\left(\frac{qV_a}{2kT}\right)$



- Recombination-generation occurs at a "trap" level created by defects in the crystal lattice and existing around mid-gap.

### Deviations from ideal (the complete picture)

- A: reverse bias breakdown
- B: thermal generation in depletion region
- C: Thermal recombination in depletion region
- \*\* Ideal (diffusion)



D: High-level injection

$$(\delta n_p > p_p)$$

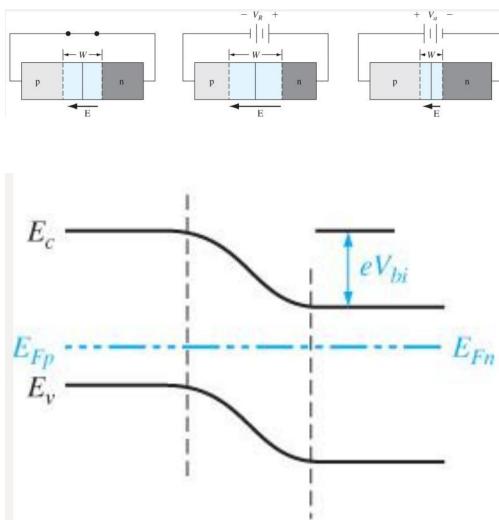
E: Series resistance

# Lecture 12: p n junctions current, small signal model

## Topics

- Neamen Chapter 8
- Ideal p-n junction current
- Recombination-generation in the depletion region
- Non-ideal diode characteristics
- Diode small signal model

## Overview of different bias conditions



(a)

(b)

(c)

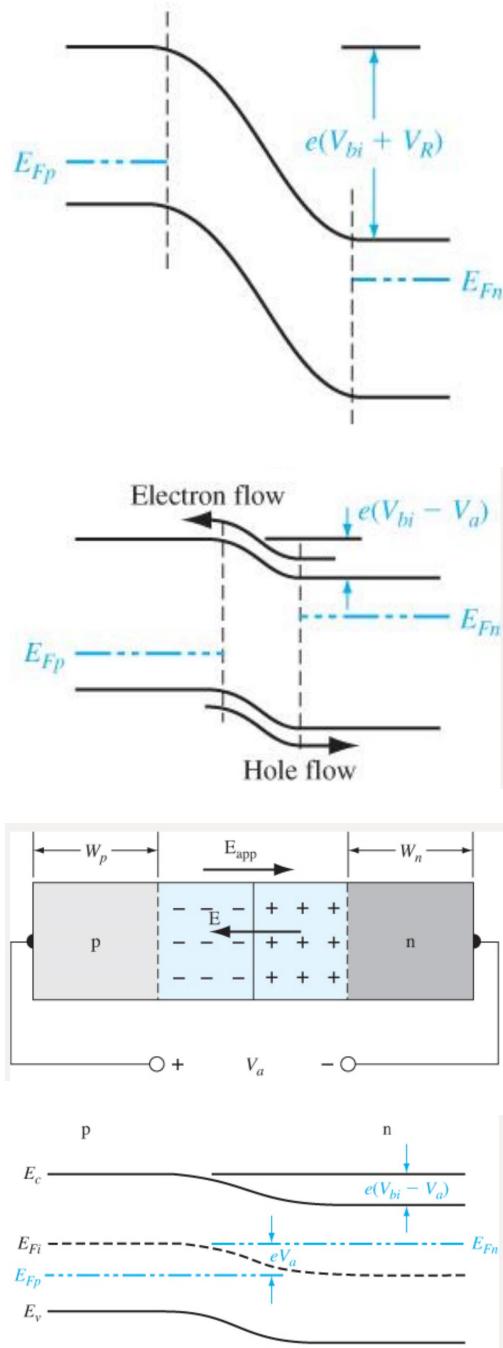
Figure 8.1 A pn junction and its associated energy-band diagram for (a) zero bias, (b) reverse bias, and (c) forward bias.

## Focusing on forward bias

(a)

(b)

Figure 8.3 I (a) A pn junction with an applied forward-bias voltage showing the directions of the electric field induced by  $V_a$  and the space charge electric field. (b) Energy-band diagram of the forward-biased pn junction.



## Law of the junction

- $$V_{bi} = \frac{kT}{e} \ln \frac{N_a N_d}{n_i^2}$$

- Rewriting:

- $$\frac{N_a N_d}{n_i^2} = \exp\left(\frac{eV_{bi}}{kT}\right)$$

- $n_{p0} = n_{n0} \exp\left(-\frac{eV_{bi}}{kT}\right)$
- Relates the minority carrier concentration on the p-side to the majority carrier concentration on the n-side

### Current in forward bias

Table 8.1 I Commonly used terms and notation for this chapter

Term	Meaning
$N_a$	Acceptor concentration in the p region of the pn junction
$N_d$	Donor concentration in the n region of the pn junction
$n_{n0} = N_d$	Thermal-equilibrium majority carrier electron concentration in the n region
$p_{p0} = N_a$	Thermal-equilibrium majority carrier hole concentration in the p region
$n_{p0} = n_i^2/N_a$	Thermal-equilibrium minority carrier electron concentration in the p region
$p_{n0} = n_i^2/N_d$	Thermal-equilibrium minority carrier hole concentration in the n region
$n_p$	Total minority carrier electron concentration in the p region
$p_n$	Total minority carrier hole concentration in the n region
$n_p(-x_p)$	Minority carrier electron concentration in the p region at the space charge edge
$p_n(x_n)$	Minority carrier hole concentration in the n region at the space charge edge
$\delta n_p = n_p - n_{p0}$	Excess minority carrier electron concentration in the p region
$\delta p_n = p_n - p_{n0}$	Excess minority carrier hole concentration in the n region

$$n_p = n_{n0} \exp\left(-\frac{e(V_{bi} - V_a)}{kT}\right) = n_{p0} \exp\left(\frac{eV_a}{kT}\right)$$

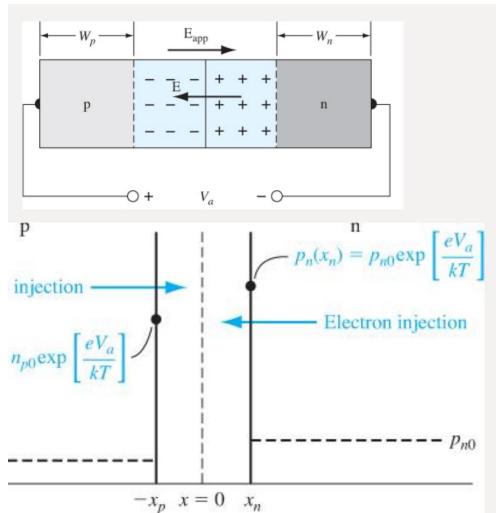
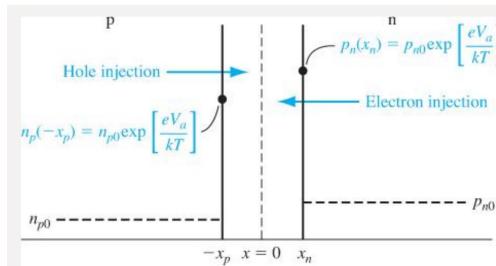


Figure 8.4 I Excess minority carrier concentrations at the space charge edges generated by the forward-bias voltage.

### Exercise

A silicon pn junction at  $T = 300$  K is doped with impurity concentrations of  $N_d = 2 \times 10^{16}$  cm $^{-3}$  and  $N_a = 5 \times 10^{16}$  cm $^{-3}$ . The junction is forward biased at  $V_a = 0.650$  V. Determine the minority carrier concentrations at the space charge edges. Does low injection still apply?



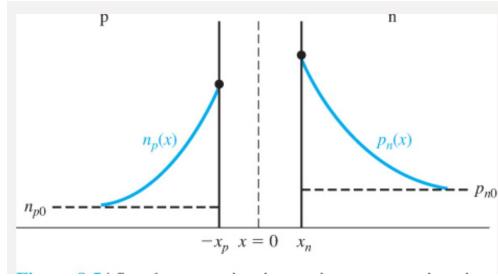
### Current in forward bias

Table 8.1 I Commonly used terms and notation for this chapter

Term	Meaning
$N_a$	Acceptor concentration in the p region of the pn junction
$N_d$	Donor concentration in the n region of the pn junction
$n_{n0} = N_d$	Thermal-equilibrium majority carrier electron concentration in the n region
$p_{p0} = N_a$	Thermal-equilibrium majority carrier hole concentration in the p region
$n_{p0} = n_i^2/N_a$	Thermal-equilibrium minority carrier electron concentration in the p region
$p_{n0} = n_i^2/N_d$	Thermal-equilibrium minority carrier hole concentration in the n region
$n_p$	Total minority carrier electron concentration in the p region
$p_n$	Total minority carrier hole concentration in the n region
$n_p(-x_p)$	Minority carrier electron concentration in the p region at the space charge edge
$p_n(x_n)$	Minority carrier hole concentration in the n region at the space charge edge
$\delta n_p = n_p - n_{p0}$	Excess minority carrier electron concentration in the p region
$\delta p_n = p_n - p_{n0}$	Excess minority carrier hole concentration in the n region

### Forward bias band diagram

Figure 8.5 I Steady-state minority carrier concentrations in a pn junction under forward bias.

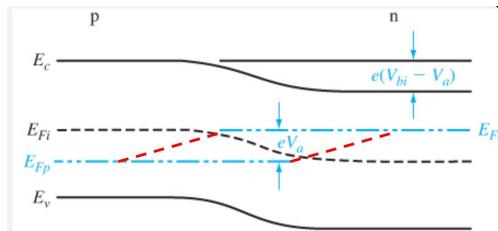


$$n_p(-xp) = n_{p0} \exp\left(\frac{qV_a}{kT}\right) \quad p_n(x_n) = p_{n0} \exp\left(\frac{qV_a}{kT}\right)$$

Excess minority carrier concentrations:

$$\begin{aligned} \delta n_p(x) &= n_p(x) - n_{p0} = n_{p0} \exp\left(\frac{qV_a}{kT} - 1\right) \exp\left(\frac{x_p + x}{L_n}\right) \\ \delta p_n(x) &= p_n(x) - p_{n0} = p_{n0} \exp\left(\frac{qV_a}{kT} - 1\right) \exp\left(\frac{x_n - x}{L_p}\right) \end{aligned}$$

## Quasi Fermi levels



$$\begin{aligned} E_{Fn} &= E_{Fi} + kT \ln\left(\frac{n}{n_i}\right) \\ E_{Fp} &= E_{Fi} - kT \ln\left(\frac{p}{n_i}\right) \\ n_p(-x_p) p_n(x_n) &= n_i^2 \exp\left(\frac{E_{Fn} - E_{Fp}}{kT}\right) \end{aligned}$$

## vuKe

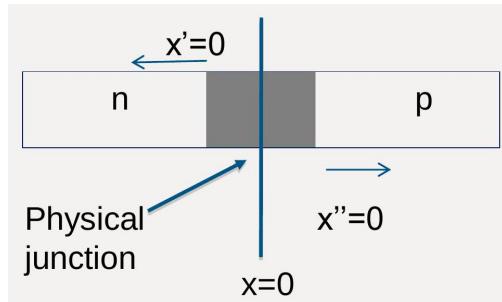
### Current flow in a $p - n$ junction

We know two facts:

- (1) Electrons will diffuse from n-side and give a current
- (2) Holes will diffuse from the p-side and give another component of the current

In addition

Current inside the depletion region is constant



Total current is constant - individual hole and electron currents are continuous functions and constant in a depletion region.

Note that the diffusion process is injecting minority carriers (electrons into the p-side and holes into the n side). This is known as the minority carrier injection

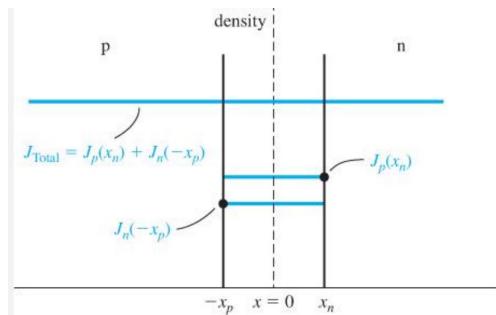


Figure 8.7 | Electron and hole current densities through the space charge region of a pn junction.

### Current flow in a p-n junction

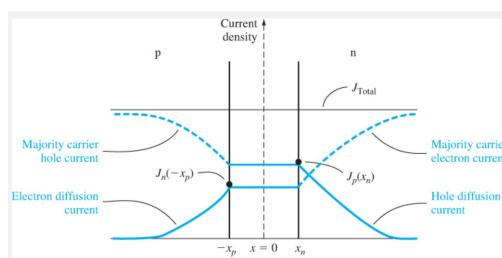


Figure 8.10 I Ideal electron and hole current components through a pn junction under forward bias.

### Problem 8.7

An ideal germanium pn junction diode has the following parameters:  $N_a = 4 \times 10^{15} \text{ cm}^{-3}$ ,  $N_d = 2 \times 10^{17} \text{ cm}^{-3}$ ,  $D_p = 48 \text{ cm}^2/\text{s}$ ,  $D_n = 90 \text{ cm}^2/\text{s}$ ,  $\tau_{p0} = \tau_{n0} = 2 \times 10^{-6} \text{ s}$ , and  $A = 10^{-4} \text{ cm}^2$ . Determine the diode current for (a) a forward-bias voltage of 0.25 V and (b) a reverse-bias voltage of 0.25 V.

## Recombination-generation current

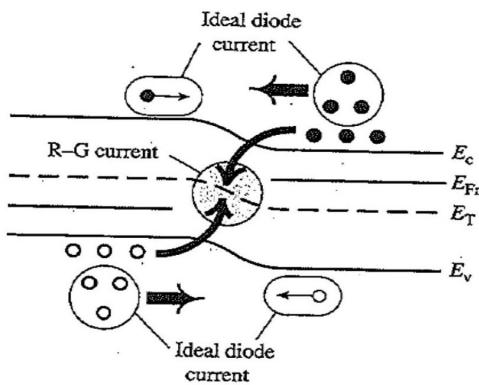
- Some electrons and holes crossing the depletion region can recombine at a rate:

$$\bullet R = \frac{np - n_i^2}{\tau_{p_0}(n+n') + \tau_{n_0}(p+p')}$$

$$\bullet R_{\max} = \frac{n_i}{2\tau_0} \exp\left(\frac{qV_a}{2kT}\right)$$

$$\bullet J_{rec} = \frac{qWn_i}{2\tau_0} \exp\left(\frac{qV_a}{2kT}\right) = J_{r0} \exp\left(\frac{qV_a}{2kT}\right)$$

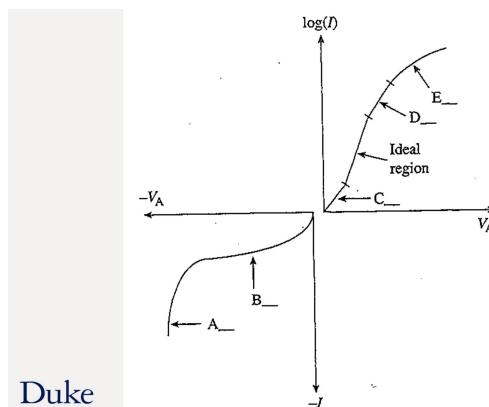
$$J_{gen} = \frac{en_i W}{2\tau_0}$$



(b) Forward bias

- Recombination-generation occurs at a "trap" level created by defects in the crystal lattice and existing around mid-gap.

## Deviations from ideal (the complete picture)



A: reverse bias breakdown

B : thermal generation in depletion region

C: Thermal recombination in depletion region

\*\* Ideal (diffusion)

D: High-level injection

$(\delta n_p > p_p)$

E: Series resistance

### Exercise 8.6

Consider a GaAs pn junction diode at  $T = 300$  K with parameters  $N_d = 8 \times 10^{10} \text{ cm}^{-3}$ ,  $N_a = 2 \times 10^{15} \text{ cm}^{-3}$ ,  $D_n = 207 \text{ cm}^2/\text{s}$ ,  $D_p = 9.80 \text{ cm}^2/\text{s}$ , and  $\tau_0 = \tau_{p0} = \tau_{n0} = 5 \times 10^{-8} \text{ s}$ .

(a) Calculate the ideal reverse-biased saturation current density. (b) Find the reverse-biased generation current density if the diode is reverse biased at  $V_R = 5 \text{ V}$ .

(c) Determine the ratio of  $J_{\text{gen}}$  to  $J_s$ .

$$(a) J_s = e n_i^2 \left[ \frac{1}{N_a} \sqrt{\frac{D_n}{\tau_{n0}}} + \frac{1}{N_d} \sqrt{\frac{D_p}{\tau_{p0}}} \right] = (1.6 \times 10^{-19}) (1.8 \times 10^6)^2 \times \left[ \frac{1}{2 \times 10^{15}} \sqrt{\frac{207}{5 \times 10^{-8}}} + \frac{1}{8 \times 10^{16}} \sqrt{\frac{9.8}{5 \times 10^{-8}}} \right]$$

or  $J_s = 1.677 \times 10^{-17} \text{ A/cm}^2$

$$(b) V_{bi} = (0.0259) \ln \left[ \frac{(2 \times 10^{15})(8 \times 10^{16})}{(1.8 \times 10^6)^2} \right]$$

$$= 1.174 \text{ V}$$

$$\begin{aligned} W &= \left\{ \frac{2 \in_s (V_{bi} + V_R)}{e} \left( \frac{N_a + N_d}{N_a N_d} \right) \right\}^{1/2} \\ &= \left\{ \frac{2(13.1)(8.85 \times 10^{-14})(1.174 + 5)}{1.6 \times 10^{-19}} \right. \\ &\quad \left. \times \left[ \frac{2 \times 10^{15} + 8 \times 10^{16}}{(2 \times 10^{15})(8 \times 10^{16})} \right] \right\}^{1/2} \\ \text{or } W &= 2.141 \times 10^{-4} \text{ cm} \end{aligned}$$

$$\begin{aligned} J_{\text{gen}} &= \frac{en_i W}{2\tau_0} \\ &= \frac{(1.6 \times 10^{-19})(1.8 \times 10^6)(2.141 \times 10^{-4})}{2(5 \times 10^{-8})} \end{aligned}$$

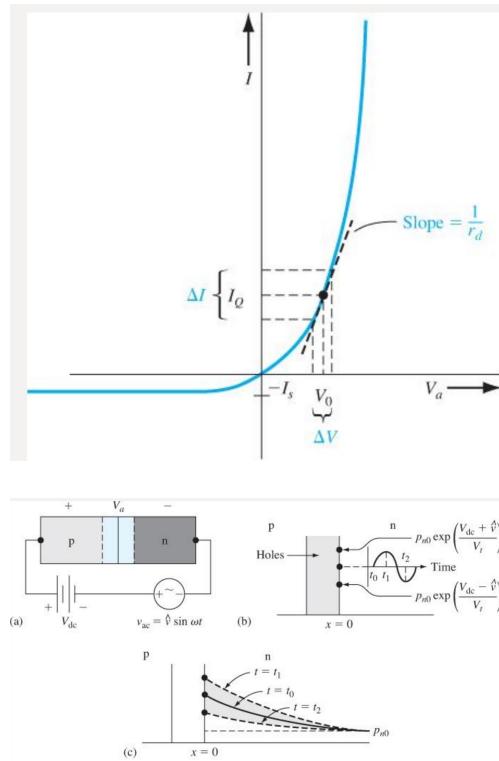
$$J_{\text{gen}} = 6.166 \times 10^{-10} \text{ A/cm}^2$$

$$(a) \frac{J_{\text{gen}}}{J_s} = \frac{6.166 \times 10^{-10}}{1.677 \times 10^{-17}} = 3.68 \times 10^7$$

### Small signal model

$$r_d = \frac{V_t}{I_{DQ}}$$

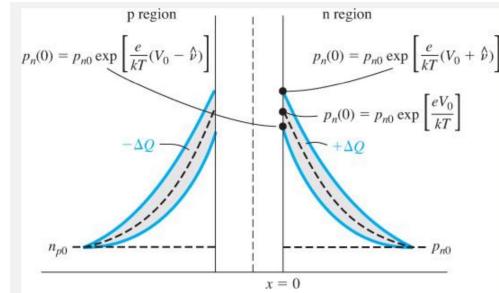
Figure 8.18 I Curve showing the concept of the small-signal diffusion resistance.



### Small signal model: diffusion capacitance

Figure 8.19l (a) A pn junction with an ac voltage superimposed on a forward-biased dc value; (b) the hole concentration versus time at the space charge edge; (c) the hole concentration versus distance in the n region at three different times.

### Small signal model

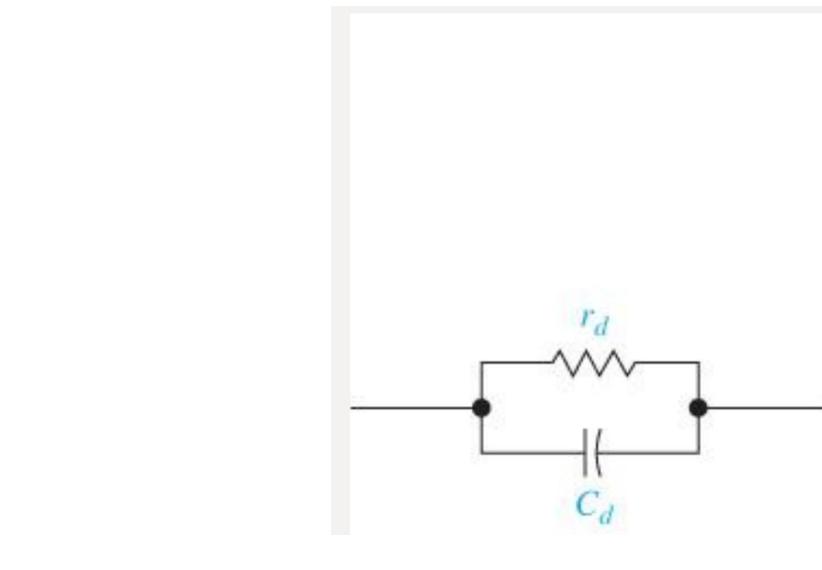


Diffusion capacitance

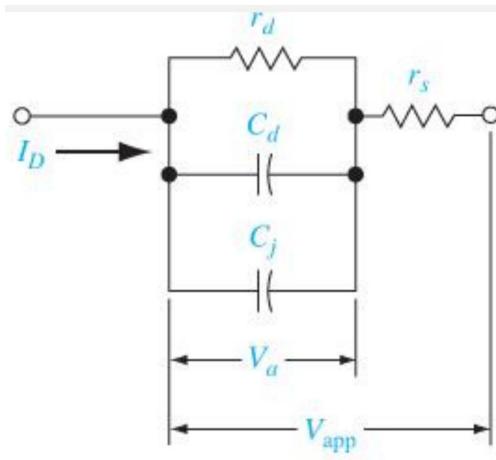
$$C_d = \left( \frac{1}{2V_t} \right) (I_{p0}\tau_{p0} + I_{n0}\tau_{n0})$$

Figure 8.21| Minority carrier concentration changes with changing forward-bias voltage.

### Small signal model



(a)



(b)

Figure 8.22 I (a) Small-signal equivalent circuit of ideal forwardbiased pn junction diode; (b) complete small-signal equivalent circuit of pn junction.

### Problem 8.39

Consider a  $p^+$ n silicon diode at  $T = 300$  K. The diode is forward biased at a current of 1 mA. The hole lifetime in the n region is  $10^{-7}$  s. Neglecting the depletion capacitance, calculate the diode impedance at frequencies of 10 kHz, 100kHz, 1MHz, and 10MHz.

# Lecture 13: p n junctions current, small signal model, metal-semiconductor junctions

---

## Topics

- Neamen Chapter 8
- Non-ideal diode characteristics
- Diode small signal model
- Neamen Chapter 9
- Schottky and Ohmic contacts

**Current flow in a p-n junction**

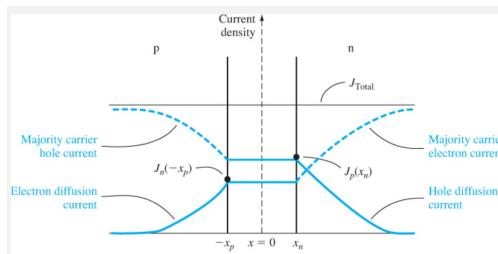


Figure 8.10 I Ideal electron and hole current components through a pn junction under forward bias.

$$F, B \cdot J_{\text{ideal}} \longrightarrow J_S e^{aV_a/kT}$$

$$R \cdot B \cdot J_{\text{ideal}} \rightarrow -J_s$$

## Problem 8.7

An ideal germanium pn junction diode has the following parameters:  $N_a = 4 \times 10^{15} \text{ cm}^{-3}$ ,  $N_d = 2 \times 10^{17} \text{ cm}^{-3}$ ,  $D_p = 48 \text{ cm}^2/\text{s}$ ,  $D_n = 90 \text{ cm}^2/\text{s}$ ,  $\tau_{p0} = \tau_{n0} = 2 \times 10^{-6} \text{ s}$ , and  $A = 10^{-4} \text{ cm}^2$ . Determine the diode current for (a) a forward-bias voltage of 0.25 V and (b) a reverse-bias voltage of 0.25 V.

## Recombination-generation current

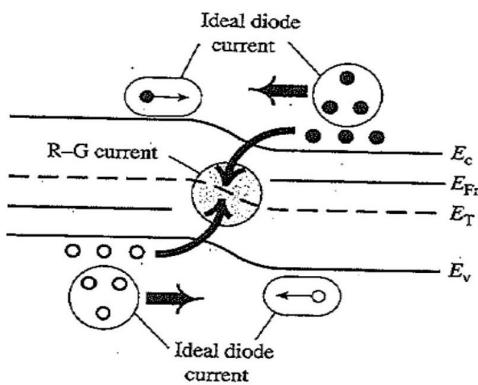
- Some electrons and holes crossing the depletion region can recombine at a rate:

$$\bullet R = \frac{np - n_i^2}{\tau_{p_0}(n+n') + \tau_{n_0}(p+p')}$$

$$\bullet R_{\max} = \frac{n_i}{2\tau_0} \exp\left(\frac{qV_a}{2kT}\right)$$

$$\bullet J_{\text{rec}} = \frac{qWn_i}{2\tau_0} \exp\left(\frac{qV_a}{2kT}\right) = J_{r0} \exp\left(\frac{qV_a}{2kT}\right)$$

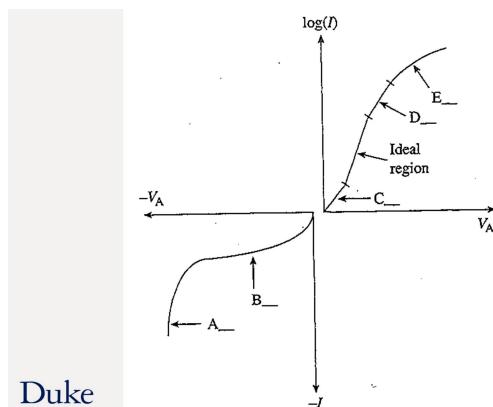
$$J_{\text{gen}} = \frac{en_i W}{2\tau_0}$$



(b) Forward bias

- Recombination-generation occurs at a "trap" level created by defects in the crystal lattice and existing around mid-gap.

## Deviations from ideal (the complete picture)



A: reverse bias breakdown

B : thermal generation in depletion region

C: Thermal recombination in depletion region

\*\* Ideal (diffusion)

D: High-level injection

$(\delta n_p > p_p)$

E: Series resistance

### Exercise 8.6

Consider a GaAs pn junction diode at  $T = 300$  K with parameters  $N_d = 8 \times 10^{10} \text{ cm}^{-3}$ ,  $N_a = 2 \times 10^{15} \text{ cm}^{-3}$ ,  $D_n = 207 \text{ cm}^2/\text{s}$ ,  $D_p = 9.80 \text{ cm}^2/\text{s}$ , and  $\tau_0 = \tau_{p0} = \tau_{n0} = 5 \times 10^{-8} \text{ s}$ .

(a) Calculate the ideal reverse-biased saturation current density. (b) Find the reverse-biased generation current density if the diode is reverse biased at  $V_R = 5 \text{ V}$ .

(c) Determine the ratio of  $J_{\text{gen}}$  to  $J_s$ .

$$(a) J_s = e n_i^2 \left[ \frac{1}{N_a} \sqrt{\frac{D_n}{\tau_{n0}}} + \frac{1}{N_d} \sqrt{\frac{D_p}{\tau_{p0}}} \right] = (1.6 \times 10^{-19}) (1.8 \times 10^6)^2 \times \left[ \frac{1}{2 \times 10^{15}} \sqrt{\frac{207}{5 \times 10^{-8}}} + \frac{1}{8 \times 10^{16}} \sqrt{\frac{9.8}{5 \times 10^{-8}}} \right]$$

or  $J_s = 1.677 \times 10^{-17} \text{ A/cm}^2$

$$(b) V_{bi} = (0.0259) \ln \left[ \frac{(2 \times 10^{15})(8 \times 10^{16})}{(1.8 \times 10^6)^2} \right]$$

$$= 1.174 \text{ V}$$

$$W = \left\{ \frac{2 \in_s (V_{bi} + V_R)}{e} \left( \frac{N_a + N_d}{N_a N_d} \right) \right\}^{1/2}$$

$$= \left\{ \frac{2(13.1)(8.85 \times 10^{-14})(1.174 + 5)}{1.6 \times 10^{-19}} \right.$$

$$\left. \times \left[ \frac{2 \times 10^{15} + 8 \times 10^{16}}{(2 \times 10^{15})(8 \times 10^{16})} \right] \right\}^{1/2}$$

or  $W = 2.141 \times 10^{-4} \text{ cm}$

$$J_{\text{gen}} = \frac{en_i W}{2\tau_0}$$

$$= \frac{(1.6 \times 10^{-19})(1.8 \times 10^6)(2.141 \times 10^{-4})}{2(5 \times 10^{-8})}$$

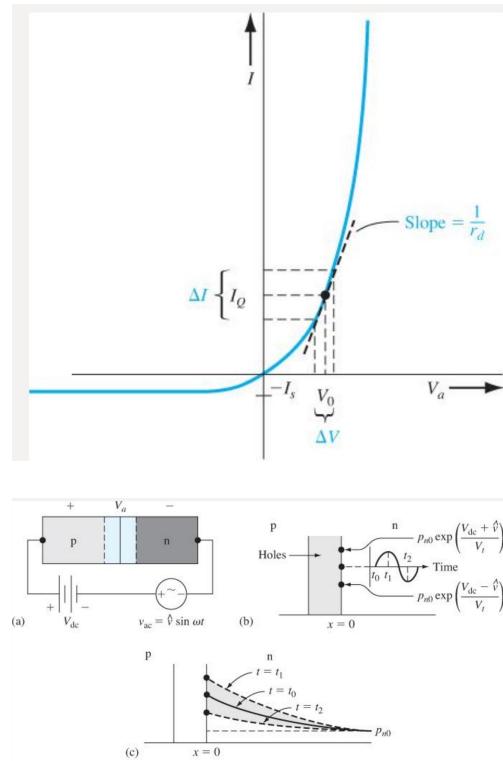
$$J_{\text{gen}} = 6.166 \times 10^{-10} \text{ A/cm}^2$$

$$(a) \frac{J_{\text{gen}}}{J_s} = \frac{6.166 \times 10^{-10}}{1.677 \times 10^{-17}} = 3.68 \times 10^7$$

### Small signal model

$$r_d = \frac{V_t}{I_{DQ}}$$

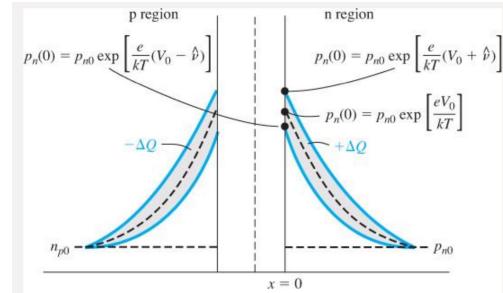
Figure 8.18 I Curve showing the concept of the small-signal diffusion resistance.



### Small signal model: diffusion capacitance

Figure 8.19l (a) A pn junction with an ac voltage superimposed on a forward-biased dc value; (b) the hole concentration versus time at the space charge edge; (c) the hole concentration versus distance in the n region at three different times.

### Small signal model

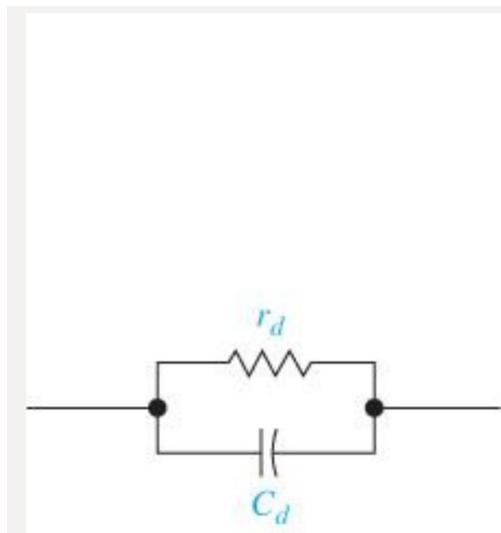


Diffusion capacitance

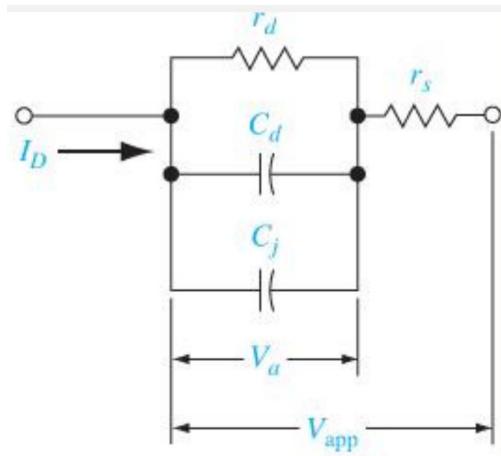
$$C_d = \left( \frac{1}{2V_t} \right) (I_{p0}\tau_{p0} + I_{n0}\tau_{n0})$$

Figure 8.21| Minority carrier concentration changes with changing forward-bias voltage.

### Small signal model



(a)

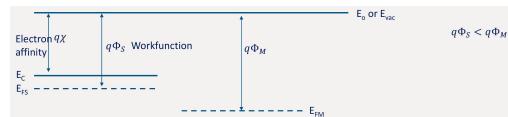


(b)

Figure 8.22 I (a) Small-signal equivalent circuit of ideal forwardbiased pn junction diode; (b) complete small-signal equivalent circuit of pn junction.

### Problem 8.39

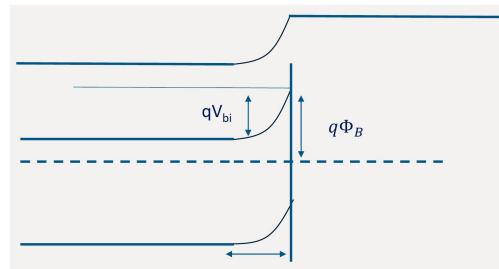
Consider a p<sup>+</sup>n silicon diode at  $T = 300$  K. The diode is forward biased at a current of 1 mA. The hole lifetime in the n region is  $10^{-7}$  s. Neglecting the depletion capacitance, calculate the diode impedance at frequencies of 10 kHz, 100kHz, 1MHz, and 10MHz.



N-type semiconductor

$E_V$

N-type semiconductor



$$q\Phi_S < q\Phi_M$$

Depletion region

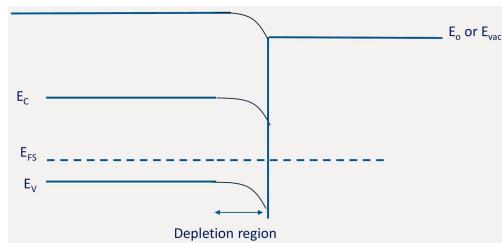
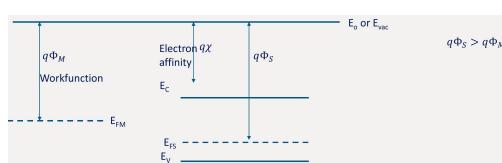
# Lecture 14: MOS Capacitor

---

## Topics

- Neamen Chapter 9
- Metal-semiconductor junction (quick)
- Neamen Chapter 10
- MOS band diagram at  $V_G = 0$  V
- Flatband voltage

P-type semiconductor

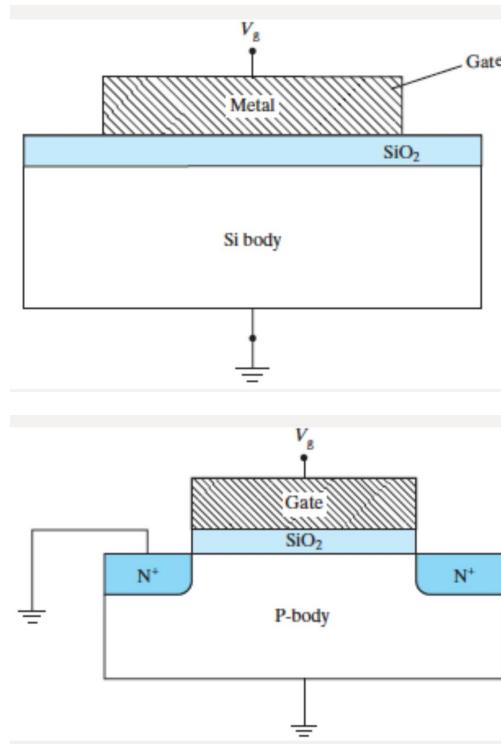


\*9.37 Both Schottky barrier diodes and ohmic contacts are to be fabricated by depositing a particular metal on a silicon integrated circuit. The work function of the metal is 4.5 V. Considering the ideal metal-semiconductor contact, determine the allowable range of doping concentrations for each type of contact. Consider both p- and n-type silicon regions.

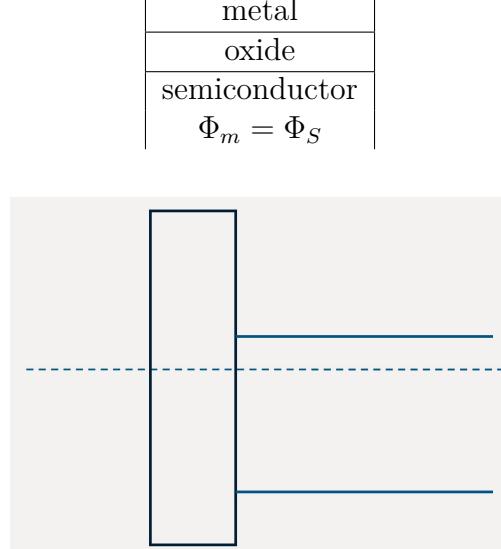
## MOS Capacitors

MOS: Metal-Oxide-Semiconductor

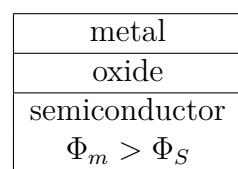
MOS  
capacitor  
MOS transistor

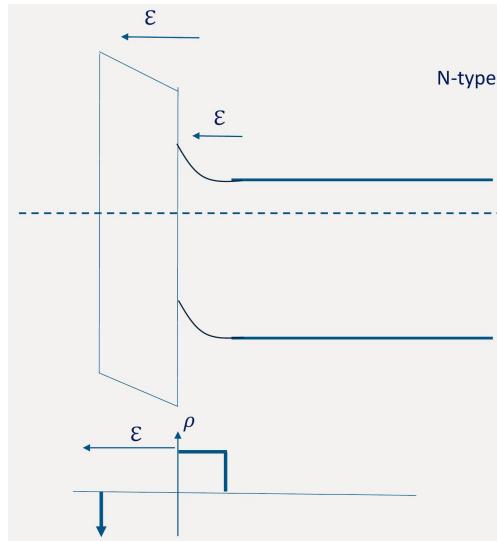


Equilibrium band diagram



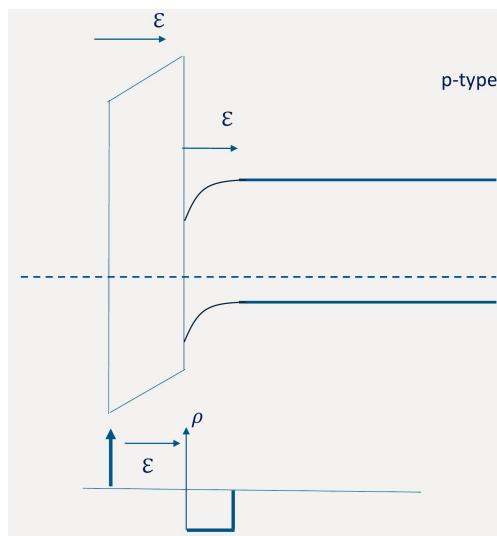
Equilibrium band diagram





### Equilibrium band diagram

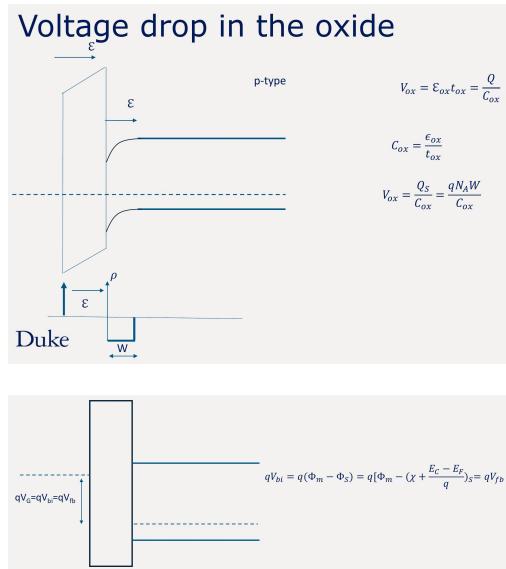
metal
oxide
semiconductor

$$\Phi_m < \Phi_S$$


### Flatband voltage

#### Applying voltage ( $V_G$ ) to MOS capacitors

- Always applied to metal (M)
- Will move  $E_{Fm}$  up ( $-V_G$ ) or down ( $+V_G$ )



- Edge of semiconductor bands ( $E_C$  and  $E_V$ ) at oxide/semiconductor interface will move with  $E_{Fm}$

# Lecture 15: MOS Capacitor 2

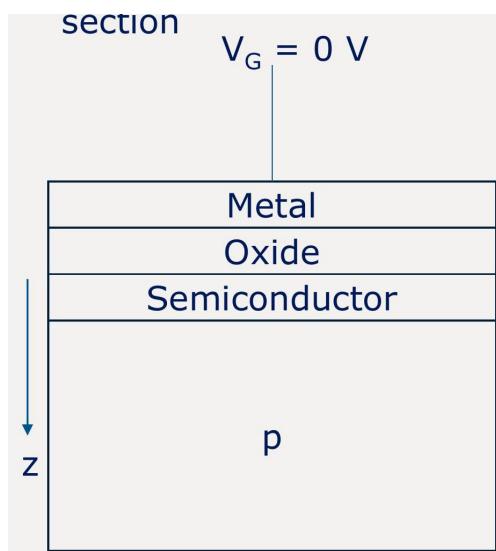
---

## Topics

- Neamen Chapter 10
- MOS capacitors operational modes
- MOS Capacitor accumulation, depletion, inversion

$$V_G = 0 \text{ V}$$

Capacitor Crosssection



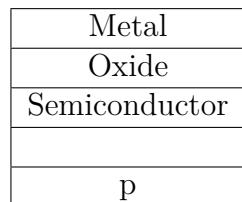
Energy band diagram

Block charge diagram

Key equations:

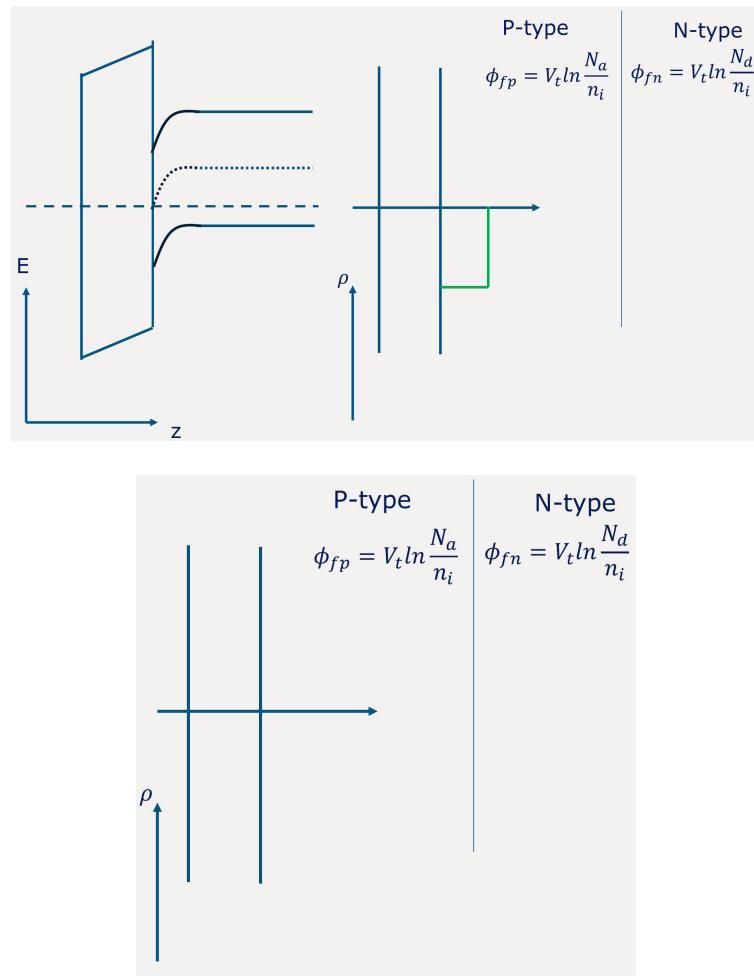
$$V_G = V_{FB}$$

Capacitor Crosssection  $V_G = V_{FB} < 0 \text{ V}$



Energy band diagram Block charge diagram

Key equations



$V_G < 0$  V accumulation

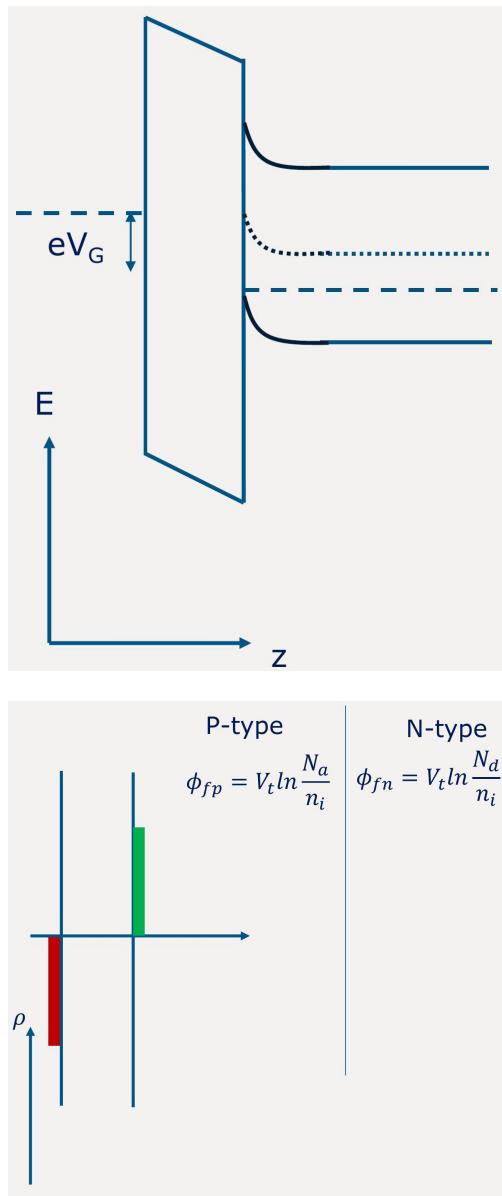
Capacitor Crosssection  $G < V_{FB} < 0$  V



Energy band diagram:

Block charge diagram

Key equations



$V_G > 0$  V depletion

Capacitor Crosssection  $v > 0$  V

Energy band diagram

Block charge diagram

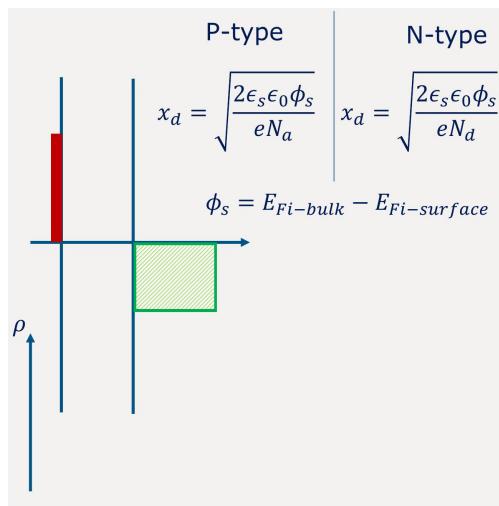
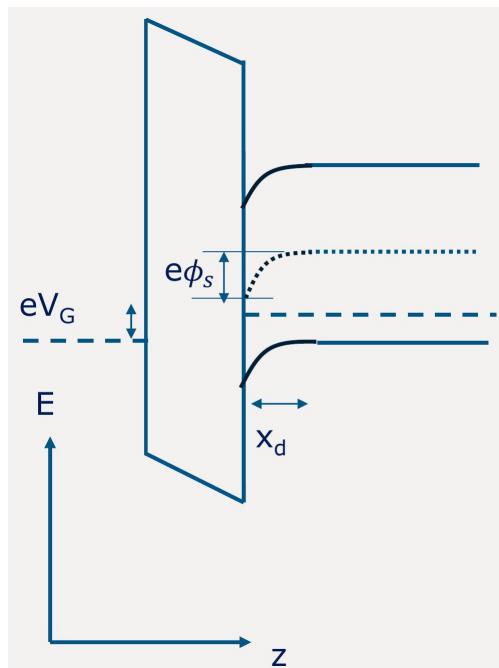
Key equations

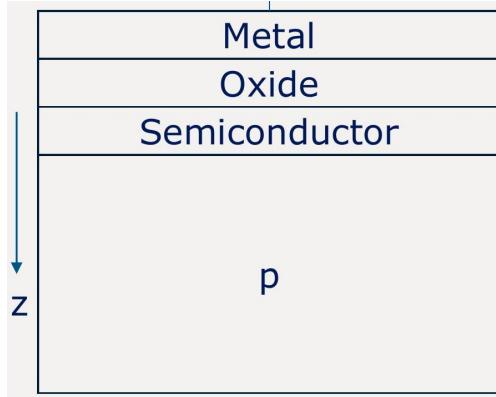
$V_G = V_T$  inversion

Key equations

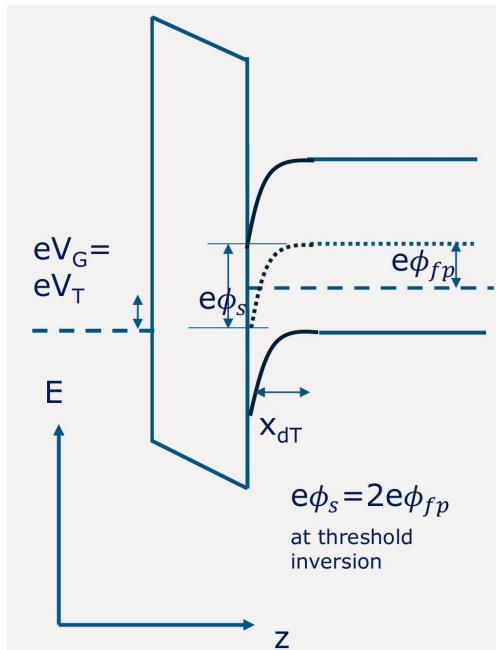
P-type

Capacitor Cross-section:  $V_G = V_T > 0V$





Energy band diagram Block charge diagram:



$$x_{dT} = \sqrt{\frac{4\epsilon_s \epsilon_0 \phi_{fp}}{eN_a}}$$

$V_G > V_T$  inversion

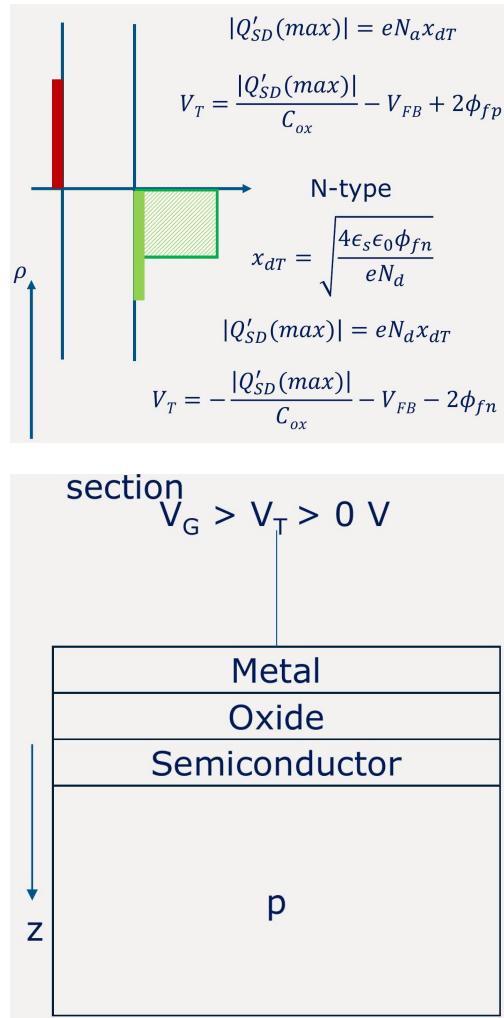
Capacitor Crosssection

Energy band diagram:

Block charge diagram

Key equations

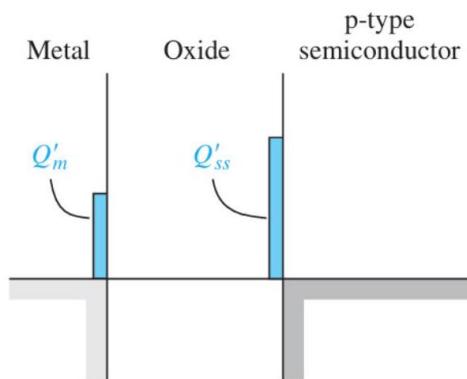
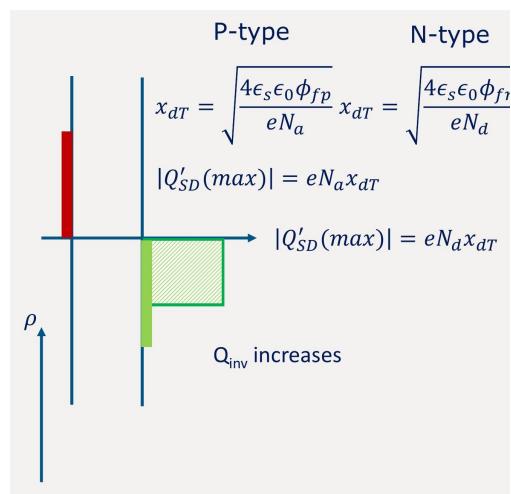
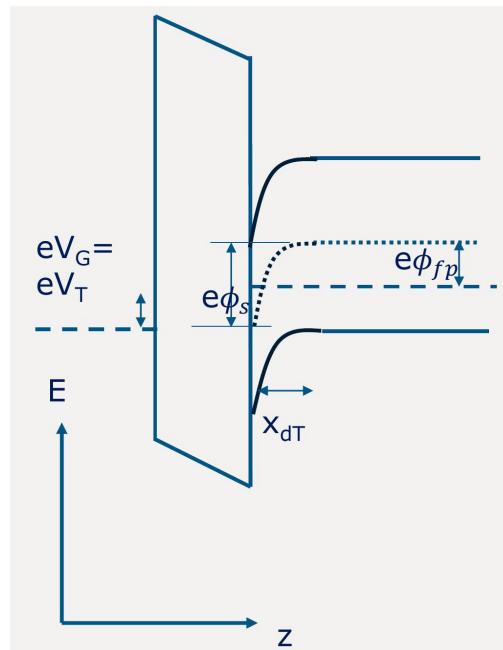
- Consider a MOS capacitor with an n-type silicon substrate. A metal-semiconductor work function difference of  $\phi_{ms} = -0.30$  V is required. Determine the silicon doping concentration required to meet this specification when the gate is (a)n<sup>+</sup>polysilicon, (b)p<sup>+</sup>polysilicon, and (c) aluminum. If a particular gate cannot meet this specification, explain why.



When charge is trapped in the oxide

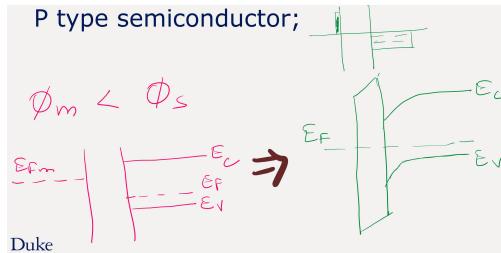
$$V_{ox} = \frac{Q'_m}{C_{ox}} \quad V_{ox} = \frac{-Q'_{ss}}{C_{ox}}$$

$$V_G = V_{FB} = \phi_{ms} - \frac{Q'_{ss}}{C_{ox}}$$



# Lecture 15\*: MOS Capacitor 2 clarifications

P type semiconductor;



**P type semiconductor;**

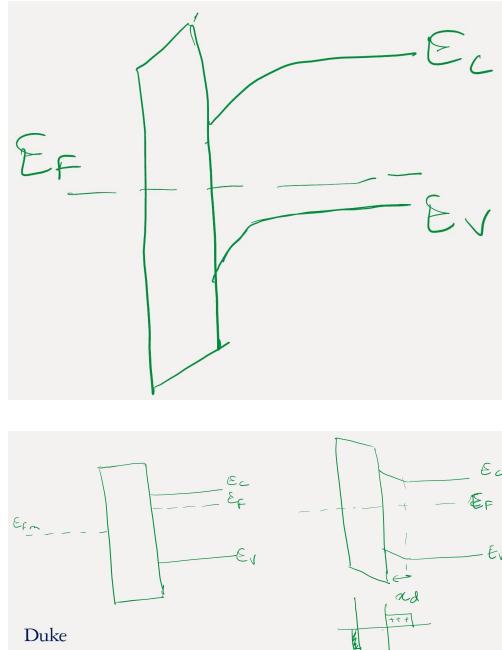
$\phi_m$

- Workfunction of metal is less than that of semiconductor
- In case of metal-semiconductor junction, holes would move from semiconductor to metal
- Depletion charge (negative) would be left behind
- Here there is an oxide in between metal and semiconductor
- So holes cannot move to metal. Instead they are pushed away from the surface.
- Depletion region formed at semiconductor surface
- Consider  $\phi_m - \phi_s$  as a positive potential being applied on gate (metal).
- This pushes holes away from semiconductor surface

N-type semiconductor;  $\phi_m > \phi_s$

**At  $V_G = 0$  V, depletion**

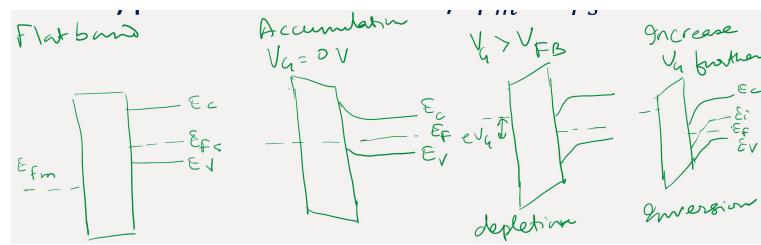
- In both the cases described in previous slides:
- P-type semiconductor;  $\phi_m < \phi_s$
- N-type semiconductor;  $\phi_m > \phi_s$
- At equilibrium (Fermi level constant throughout), we have depletion condition in semiconductor.



- The purpose of the MOSCAP is to obtain inversion charge
- If depletion is obtained at  $V_G = 0$  V (no voltage applied), then it is much easier to get inversion
- Apply  $V_T > 0$  V for p-type semiconductor
- Apply  $V_T < 0$  V for n-type semiconductor
- Now imagine the cases where due to the metalsemiconductor workfunction differences, there is accumulation charge at  $V_G = 0$  V
- To get inversion, first you need to push the accumulation charge away from semiconductor surface, then deplete the semiconductor surface, then invert it.
- Thus, much higher voltages will be required for inversion, which is inefficient
- So usually we only see the following cases for MOSCap:
- P-type semiconductor;  $\phi_m < \phi_s$
- N-type semiconductor;  $\phi_m > \phi_s$

P-type semiconductor;  $\phi_m > \phi_s$   
Flat band

$$V_{FB} = \phi_m - \phi_s > 0 \text{ V}$$



# Lecture 16: MOS Capacitor 3

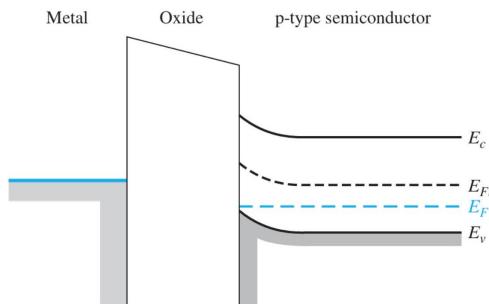
## Topics

- Neamen Chapter 10
- MOS capacitors C-V

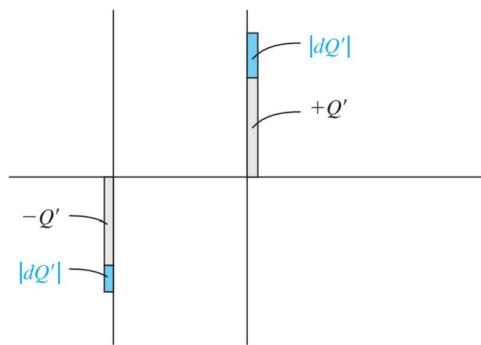
Problem 10.23

An ideal MOS capacitor with an n<sup>+</sup> polysilicon gate has a silicon dioxide thickness of  $t_{ox} = 12 \text{ nm} = 120$  on a p-type silicon substrate doped at  $N_a = 10^{16} \text{ cm}^{-3}$ . Determine the capacitance  $C_{ox}$ ,  $C'_{FB}$ ,  $C'_{min}$ , and  $C'$  (inv) at (a)  $f = 1 \text{ Hz}$  and (b)  $f = 1 \text{ MHz}$ . (c) Determine  $V_{FB}$  and  $V_T$ . (d) Sketch  $C'/C_{ox}$  versus  $V_G$  for parts (a) and (b).

## Capacitance at accumulation



(a)

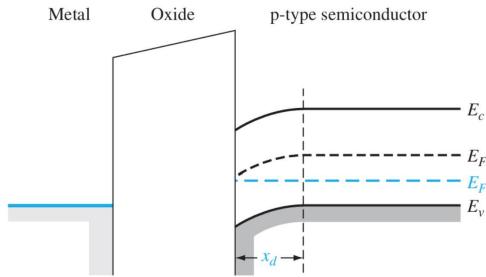


(b)

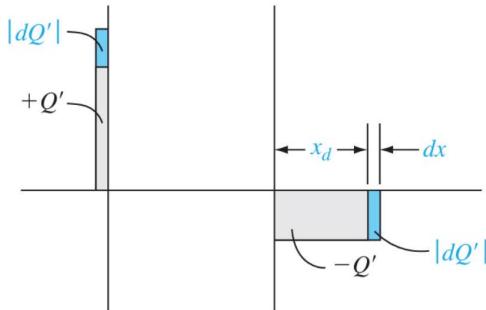
$$C'(\text{acc}) = C_{ox} = \frac{\epsilon_{\text{ox}}}{t_{\text{ox}}}$$

Figure 10.23 I (a) Energy-band diagram through a MOS capacitor for the accumulation mode. (b) Differential charge distribution at accumulation for a differential change in gate voltage.

### Capacitance at depletion



(a)



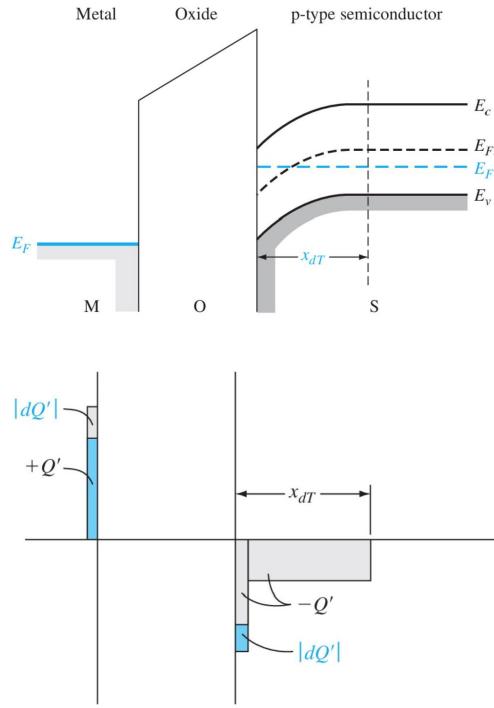
(b)

Figure 10.24 I (a) Energy-band diagram through a MOS capacitor for the depletion mode. (b) Differential charge distribution at depletion for a differential change in gate voltage.

$$\begin{aligned} \frac{1}{C'(\text{depl})} &= \frac{1}{C_{\text{ox}}} + \frac{1}{C'_{SD}} \\ C'(\text{depl}) &= \frac{C_{\text{ox}}}{1 + \frac{C_{\text{ox}}}{C'_{SD}}} = \frac{\epsilon_{\text{ox}}}{t_{\text{ox}} + \left(\frac{\epsilon_{\text{ox}}}{\epsilon_s}\right)x_d} \\ C'_{\min} &= \frac{\epsilon_{\text{ox}}}{t_{\text{ox}} + \left(\frac{\epsilon_{\text{ox}}}{\epsilon_s}\right)x_{dT}} \end{aligned}$$

### Capacitance at inversion

(a)

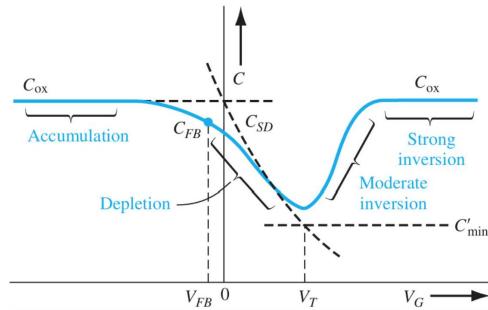


(b)

$$C'(\text{ inv }) = C_{\text{ox}} = \frac{\epsilon_{\text{ox}}}{t_{\text{ox}}}$$

Figure 10.25 I (a) Energy-band diagram through a MOS capacitor for the inversion mode. (b) Differential charge distribution at inversion for a low-frequency differential change in gate voltage.

### C-V curve



$$C'_{FB} = \frac{\epsilon_{\text{ox}}}{t_{\text{ox}} + \left( \frac{\epsilon_{\text{ox}}}{\epsilon_s} \right) \sqrt{\left( \frac{kT}{e} \right) \left( \frac{\epsilon_s}{eN_a} \right)}}$$

Figure 10.26 | Ideal low-frequency capacitance versus gate voltage of a MOS capacitor with a p-type substrate. Individual capacitance components are also shown.

### Low frequency vs. high frequency CV

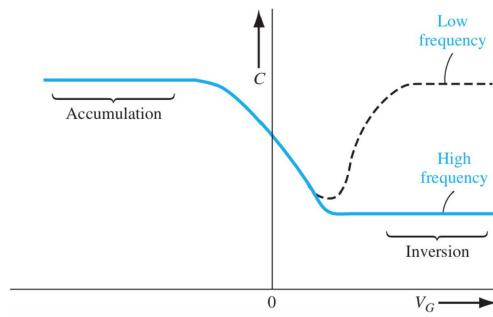
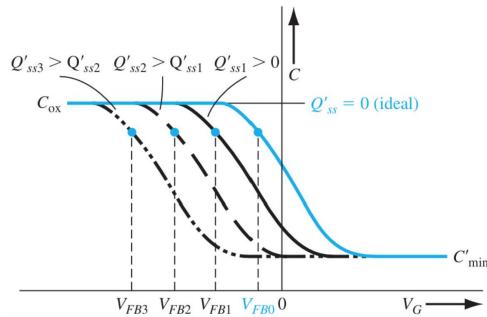


Figure 10.29 | Low-frequency and high-frequency capacitance versus gate voltage of a MOS capacitor with a p-type substrate.

### Effect of oxide traps



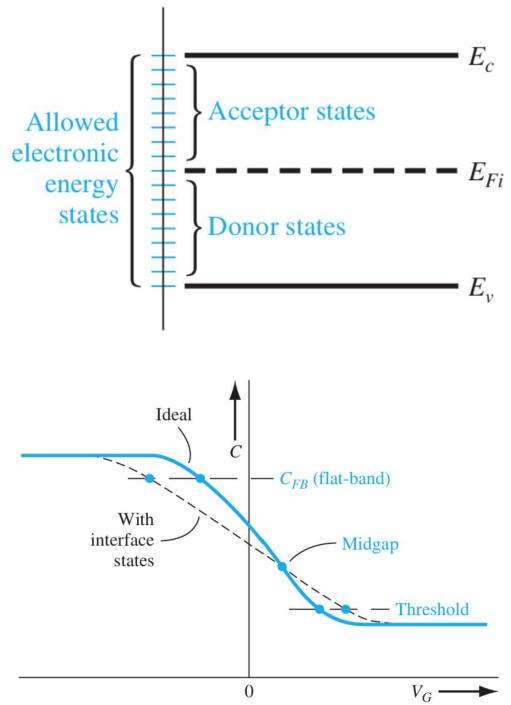
$$V_{FB} = \phi_{ms} - \frac{Q'_ss}{C_{ox}}$$

Figure 10.30 | High-frequency capacitance versus gate voltage of a MOS capacitor with a p-type substrate for several values of effective trapped oxide charge.

### Effect of interface traps

Figure 10.31 | Schematic diagram showing interface states at the oxidesemiconductor interface.

Figure 10.33 | High-frequency C-V characteristics of a MOS capacitor showing effects of interface states.



An ideal MOS capacitor is fabricated by using intrinsic silicon and an n polysilicon gate.

- Sketch the energy-band diagram through the MOS structure under flat-band conditions.
- Sketch the low-frequency  $C - V$  characteristics from negative to positive gate voltage.

# Lecture 17: MOSFETs 1

---

## Topics

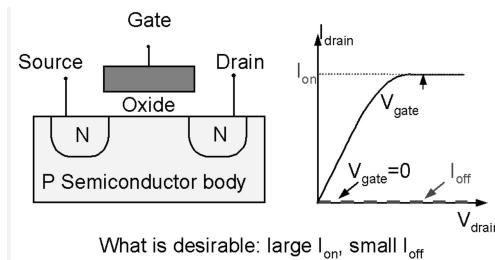
- Neamen Chapter 10
- MOSFET basic operation

## The MOSFET

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

## The MOSFET

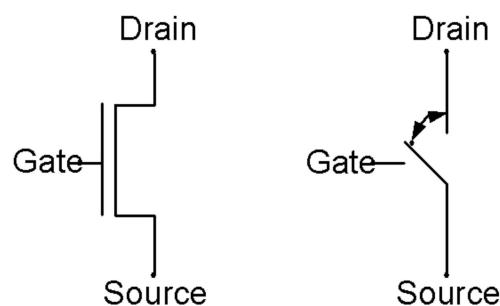
Basic MOSFET structure and IV characteristics

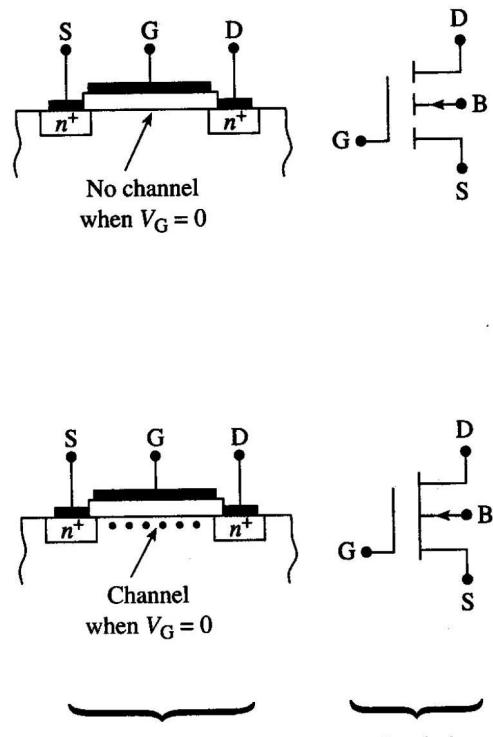


## The MOSFET

Two ways of representing a MOSFET:

### Circuit Symbol Simple Switch





## Types of MOSFETs

### Structure

Most common in logic families

### Characteristics

Used as load resistors

## The MOSFET

Jan. 23, 1930 .

J. E LILJENFELD

1,745,175

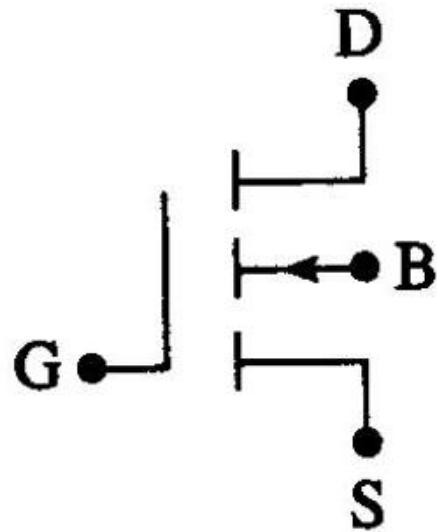
Tiled Dot. 8, 1 ges

## The MOSFET

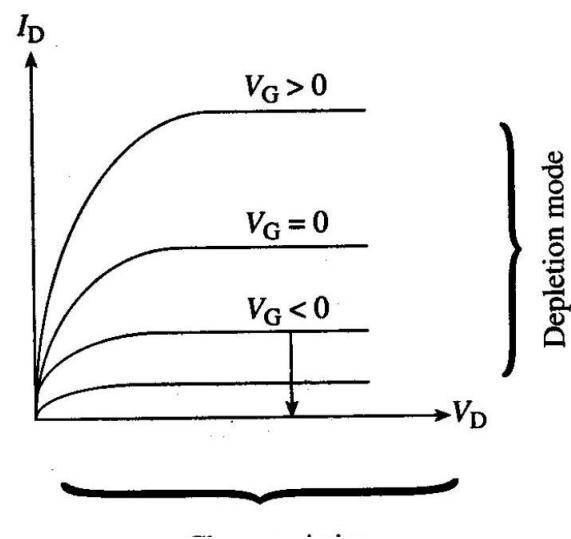
In 1935, a British patent was issued to Oskar Heil. A working MOSFET was not demonstrated until 1955.

## A modern day MOSFET

Gate oxides as thin as 1.2 nm can be manufactured reproducibly.



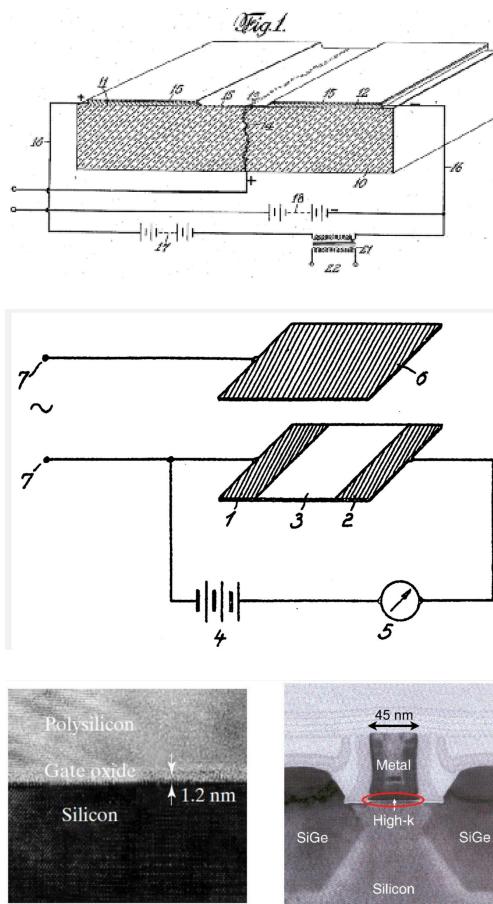
Most common in logic families



METHOD AND APPARATUS FOR CONTROLLING ELECTRIC CURRENTS

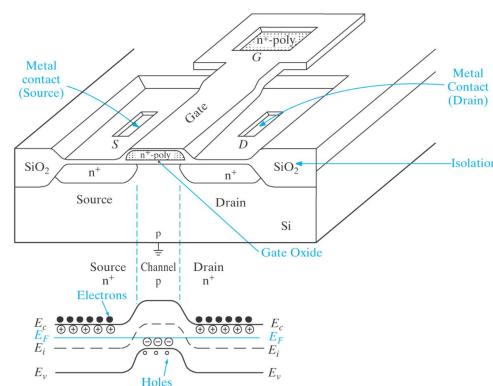
Recall:

- MOSFET is a switch that is turned ON/OFF by gate voltage ( $V_{GS}$ )



- MOS Capacitor is the core of MOSFET and determines on-state (threshold voltage)  $V_T$
- Drain voltage (  $V_{DS}$  ) determines operating regime in onstate

### MOSFET basic operation

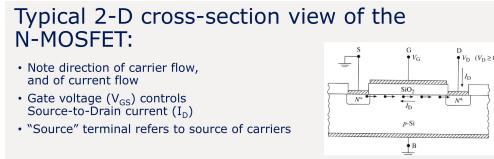


A modern "n-type" MOSFET (N-MOSFET):  
How does it work?

If  $V_G = 0$ , any current between source-drain ( $I_D$ ) ?

If  $V_G > 0$  what happens (assume source grounded,  $V_S = 0$  )

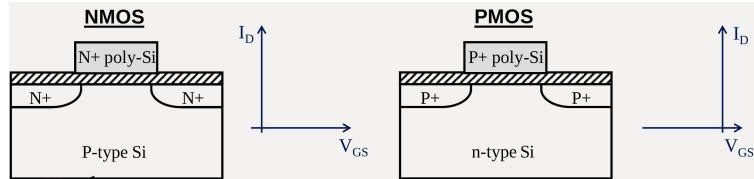
If  $V_{GS} \gg 0$  and  $V_{DS} > 0$  what happens?



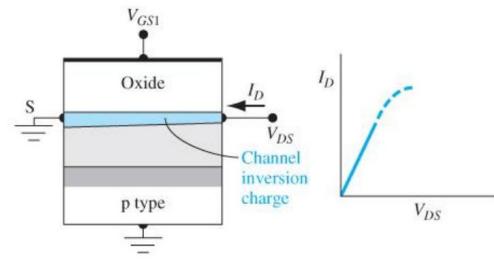
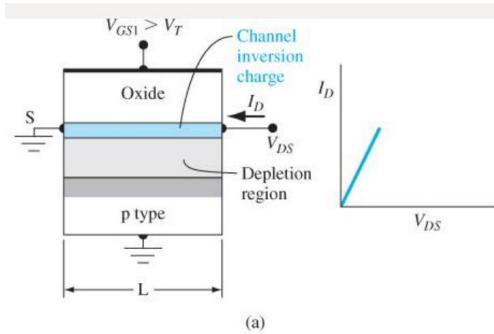
- Note direction of carrier flow, and of current flow

Gate voltage (  $V_{GS}$  ) controls Source-to-Drain current (  $I_D$  )

- "Source" terminal refers to source of carriers

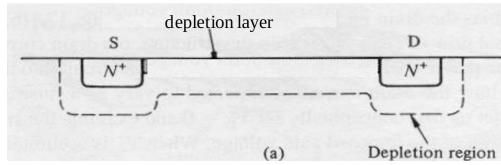


## Linear regime

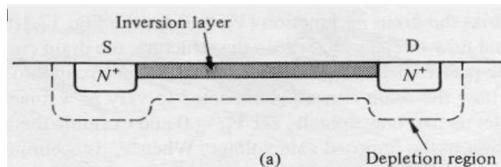


(b)  $V_{GS} > V_T$ ,  $V_{DS}$  is small

- Inversion layer is formed in channel, connecting the source and drain
- Applied  $V_{DS}$  creates electric field to generate carriers flowing from source to drain
- $I_D = \frac{W}{L} \mu_n C_{ox} (V_{GS} - V_T) V_{DS}$



(a)  
Depletion region  
 $V_{GS} > V_T, V_{DS}$  close to  $V_{DS(\text{sat})}$



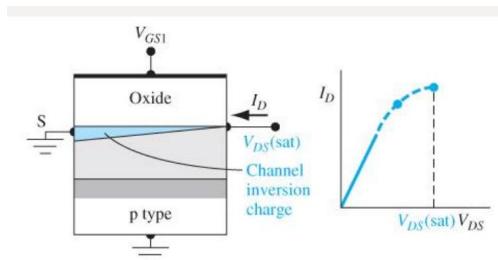
- Voltage drop across oxide (band movement at Si/SiO<sub>2</sub> interface) decreases near the drain, causing inversion charge near drain to decrease as well

Use this equation for linear regime if unsure

$$I_D = \frac{W}{2L} \mu_n C_{ox} [2(V_{GS} - V_T) V_{DS} - V_{DS}^2]$$

you're at low enough  $V_{DS}$

### Saturation regime

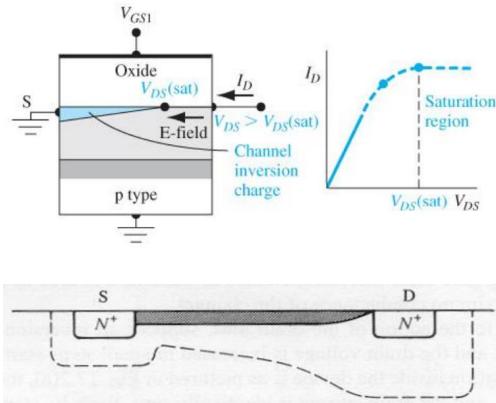


$$V_{GS} > V_T, V_{DS} = V_{DS(\text{sat})}$$

- Inversion charge density seems zero at drain, "pinching off" inversion channel

(d)  
 $V_{GS} > V_T, V_{DS} > V_{DS(\text{sat})}$

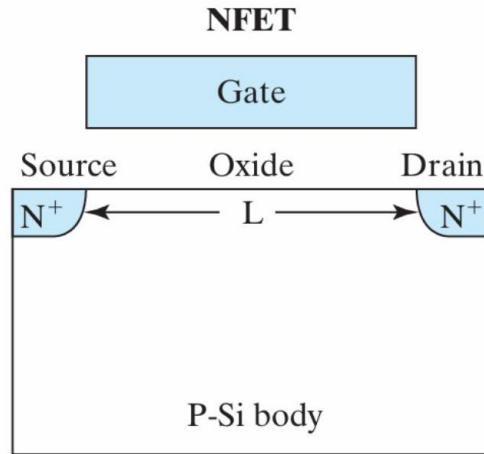
- "pinched off" region of  $L(\Delta L)$  becomes larger and corresponding electric field in  $\Delta L$  also increases



- Electrons are swept across  $\Delta L$  by electric field
- If  $\Delta L \ll L$ , then  $I_D$  is constant
- $I_D = I_D(sat) = \frac{W}{2L} \mu_n C_{ox} (V_{GS} - V_T)^2$

### Notations in your book

- $k'_n = \mu_n C_{ox} \equiv$  process control parameter
- $k_n = \frac{W\mu_n C_{ox}}{2L} \equiv$  conduction parameter



(a)

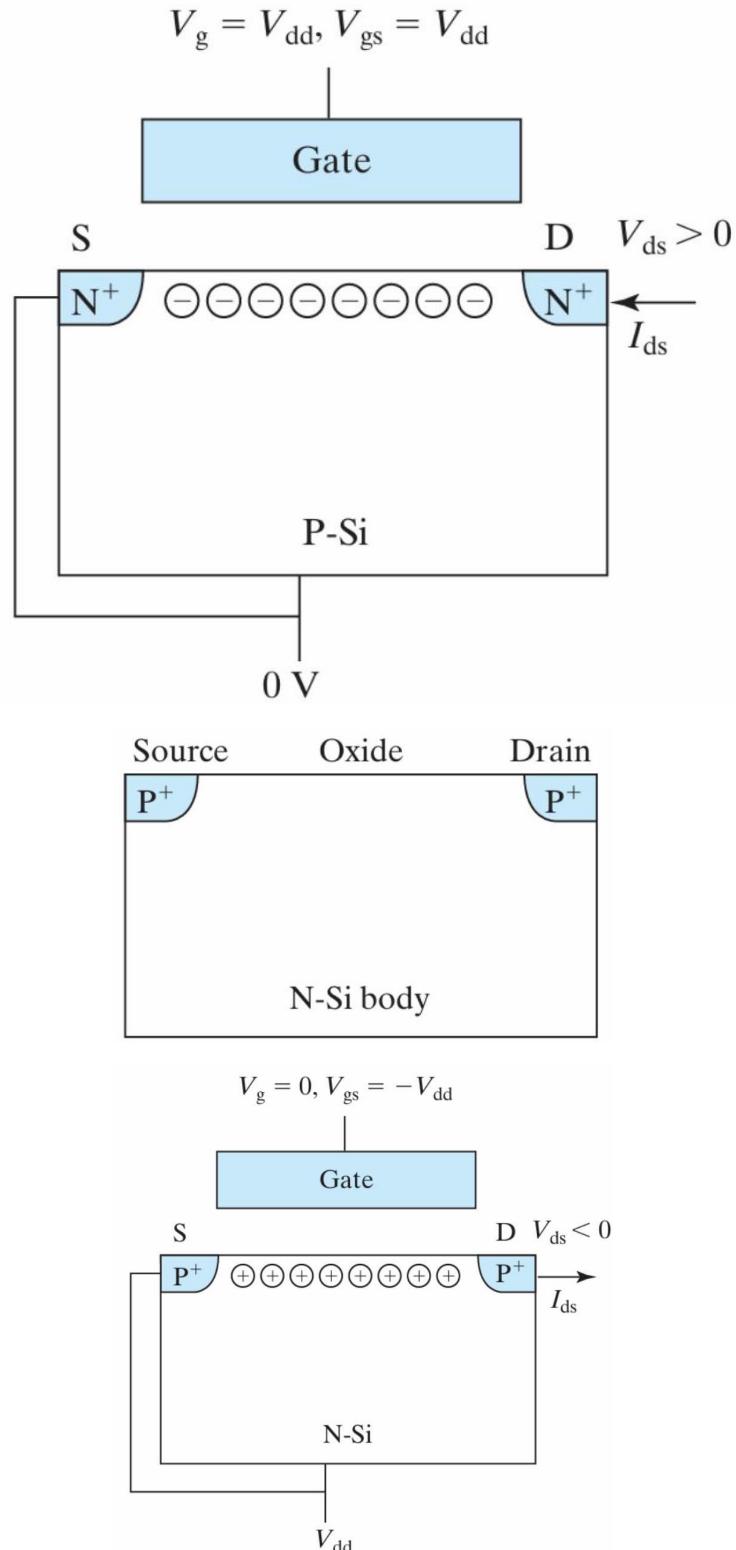
(b)

PFET

Gate

(c)

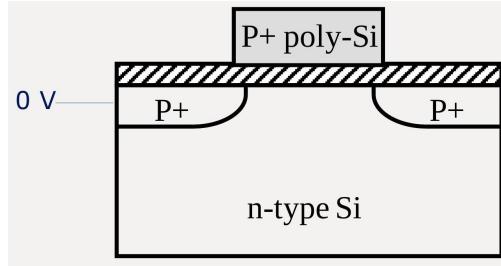
(d) state (b). (c) and (d) show a P-channel MOSFET in the off and the on states.



## PMOSFET

PMOS

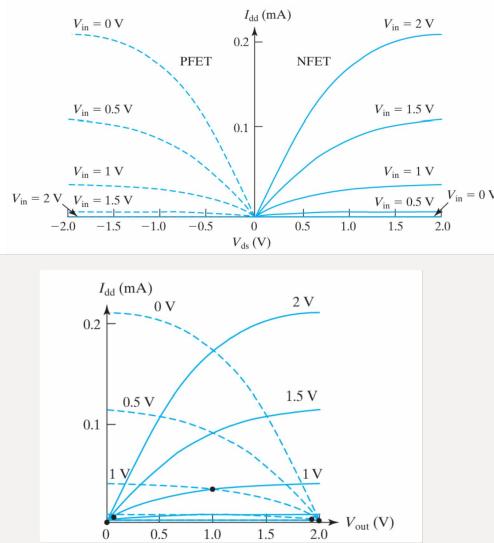
$$\text{Linear: } I_D = \frac{W}{L} \mu_p C_{ox} (V_{SG} + V_T) V_{SD}$$



$$\text{Saturation: } I_D = I_{D(\text{sat})} = \frac{W}{2L} \mu_p C_{ox} (V_{SG} + V_T)^2$$

$$k'_p = \mu_p C_{ox}$$

$$k_p = \frac{W \mu_p C_{ox}}{2L}$$



## Example

An n-channel Si MOSFET has the following parameters:  $W = 6 \text{ um}$ ,  $L = 1.5\text{um}$  and  $t_{ox} = 8 \text{ nm}$ . When the transistor is biased in the saturation region, the drain current is  $I_D(\text{sat}) = 0.132 \text{ mA}$  at  $V_{GS} = 1.0 \text{ V}$  and  $I_D(\text{sat}) = 0.295 \text{ mA}$  at  $V_{GS} = 1.25 \text{ V}$ . Determine the electron mobility and the threshold voltage.

$$I_D(\text{sat}) = \frac{W}{2L} \mu_n C_{ox} (V_{GS} - V_T)^2 \quad C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{(3.9)(8.85 \times 10^{-14})}{80 \times 10^{-8}} = 4.314 \times 10^{-7} \text{ F/cm}^2$$

$$\sqrt{I_{D_2}(\text{stat})} - \sqrt{I_{I_1}(\text{sat})} = \sqrt{\frac{W\mu_n C_\alpha}{LL}} (V_{G_{S22}} - V_{CS1})$$

$$\sqrt{0.295 \times 10^{-3}} - \sqrt{0.132 \times 10^{-3}}$$

$$= 1.7176 \times 10^{-2} - 1.1489 \times 10^{-2} = 5.687 \times 10^{-3}$$

Then

$$5.687 \times 10^{-3} = \sqrt{\frac{(6)\mu_n (4.314 \times 10^{-7})}{2(1.5)}} (1.25 - 1.0)$$

Or

$$\left(\frac{5.687 \times 10^{-3}}{0.25}\right)^2 = 8.628 \times 10^{-7} \cdot \mu_n$$

$$\Rightarrow \mu_n \cong 600 \text{ cm}^2/\text{V} \cdot \text{s}$$

We now find

$$0.132 \times 10^{-3} = \frac{(6)(600)(4.314 \times 10^{-7})}{2(1.5)} (1.0 - V_T)^2$$

$$\left(\frac{0.132 \times 10^{-3}}{5.1768 \times 10^{-4}}\right)^{1/2} = 1.0 - V_T$$

$$V_T = 0.495 \text{ V}$$

### Problem 10.34

A p-channel MOSFET has the following parameters:  $k'_p = 0.10 \text{ mA/V}^2$ ,  $W/L = 15$ , and  $V_T = -0.4 \text{ V}$ . Calculate the drain current  $I_D$  for (a)  $V_{SG} = 0.8 \text{ V}$ ,  $V_{SD} = 0.25 \text{ V}$ ; (b)  $V_{SG} = 0.8 \text{ V}$ ,  $V_{SD} = 1.0 \text{ V}$ ; (c)  $V_{SG} = 1.2 \text{ V}$ ,  $V_{SD} = 1.0 \text{ V}$ ; and (d)  $V_{SG} = 1.2 \text{ V}$ ,  $V_{SD} = 2.0 \text{ V}$ .

$$(a) I_D = \frac{k'_p}{2} \cdot \frac{W}{L} [2(V_{SG} + V_T)V_{SD} - V_{SD}^2]$$

$$= \left(\frac{0.10}{2}\right)(15)[2(0.8 - 0.4)(0.25) - (0.25)^2]$$

$$I_D = 0.103 \text{ mA}$$

$$(b) I_D = \frac{k'_p}{2} \cdot \frac{W}{L} (V_{SG} + V_T)^2$$

$$= \left(\frac{0.10}{2}\right)(15)(0.8 - 0.4)^2$$

$$= 0.12 \text{ mA}$$

$$(c) I_D = \frac{k'_p}{2} \cdot \frac{W}{L} (V_{SG} + V_T)^2$$

$$= \left(\frac{0.10}{2}\right)(15)(1.2 - 0.4)^2$$

$$= 0.48 \text{ mA}$$

(d) Same as (c),  $I_D = 0.48 \text{ mA}$

# Lecture 18: MOSFETs 2

---

## Topics

- Neamen Chapter 10
- MOSFET I-Vs and parameter extraction

## Transconductance

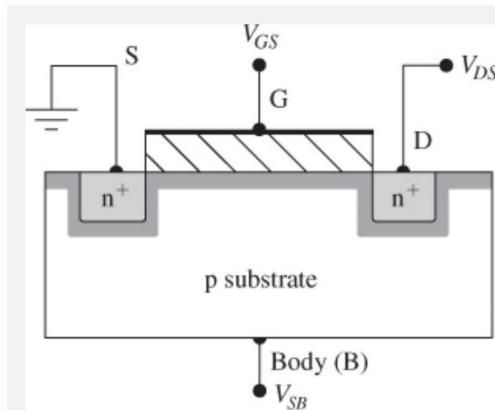
$$\bullet \quad g_m = \frac{\partial I_D}{\partial V_{GS}} = \frac{\mu_n C_{ox} W}{L} (V_{GS} - V_T)$$

For saturation region

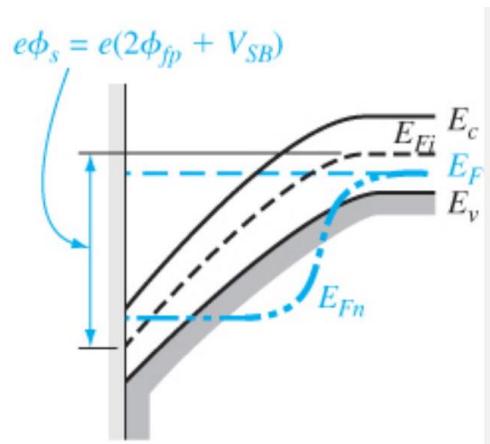
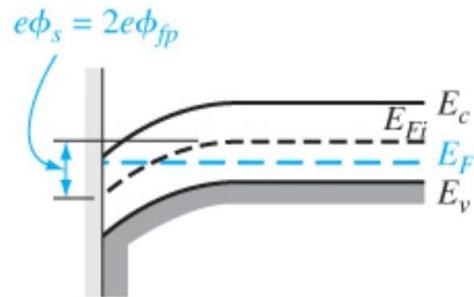
- The transconductance is a function of:
- geometry of the device:  $W, L, t_{ox}$
- carrier mobility
- threshold voltage

## Substrate bias effects

- Applying a  $V_{BS}$  to the body (B) will shift  $V_T$  such that:
- $\Delta V_T = V_T(V_{SB} > 0) - V_T(V_{SB} = 0) = \frac{\sqrt{2e\epsilon_s\epsilon_0 N_a}}{C_{ox}} [\sqrt{2\phi_{fp} + V_{SB}} - \sqrt{2\phi_{fp}}]$



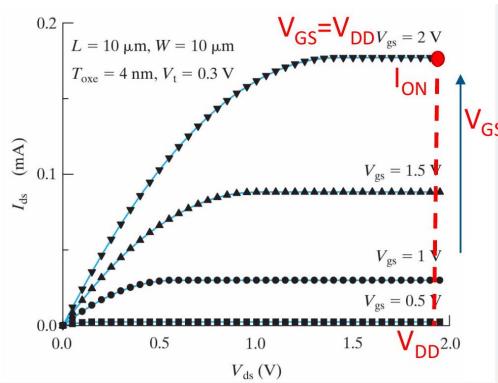
- (a)  
(b)  
(c)



### I-V curves and parameter extraction

- Note:  $V_{DD} \equiv$  operating voltage ( $V_{GS} = V_{DS} = V_{DD}$ )

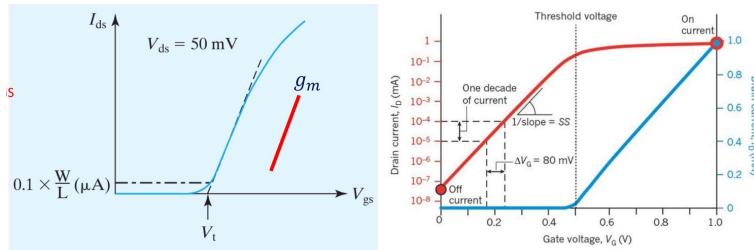
$I_D - V_{DS}$  (output curves)



On-current  $I_{ON}$  ( $I_D$  at  $V_{DD}$ )

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}} \left[ \frac{S}{\mu m} \right]$$

$I_D - V_{GS}$  (transfer curves)



- OFF current  $-I_{\text{off}}$  ( $I_D @ V_{GS} = 0 \text{ V}, V_{DS} = V_{DD}$ )
- Subthreshold swing (SS) or inverse

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \left[ \frac{S}{\mu m} \right] \text{ subthreshold slope (SS}^{-1})$$

- $SS = \left( \frac{d(\log D)}{dV_{GS}} \right)^{-1} = 60 \frac{\text{mV}}{\text{dec}}$  (best possible)
- How many  $\text{mV}$  of  $V_{GS}$  to change  $I_D$  by one decade (order of magnitude)

### Small-signal equivalent circuit

- Just as for diode, a small ac signal is added on top of some dc bias operating the transistor
- Must consider all of the impedance effects (capacitances and resistances) in the MOSFET

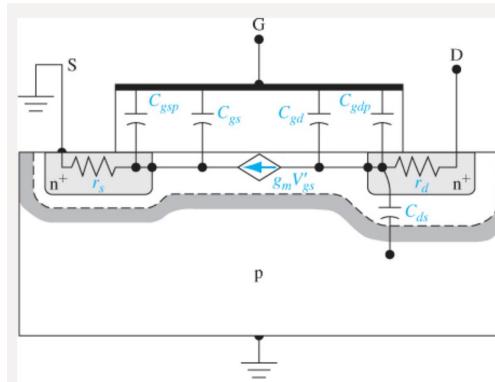
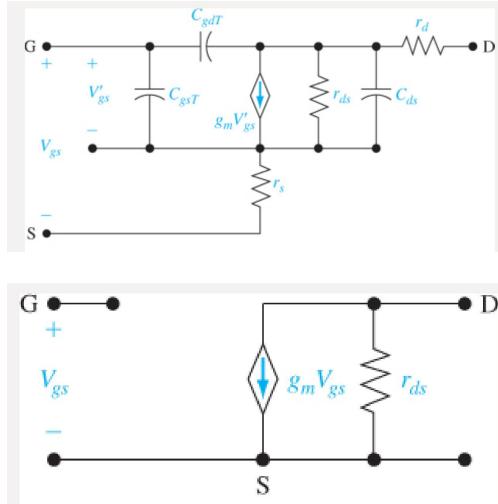


Figure 10.52 | Inherent resistances and capacitances in the n-channel MOSFET structure.

### Small signal equivalent circuit

- $r_{ds}$  due to channel length modulation

Figure 10.53 | Small-signal equivalent circuit of a commonsource n-channel MOSFET.



### Small signal model at low frequencies

- At low f, all capacitances are open circuit
- Neglect  $r_s$  and  $r_d$
- Input gate impedance is infinite

Figure 10.54 | Simplified, low-frequency small-signal equivalent circuit of a common-source n-channel MOSFET.

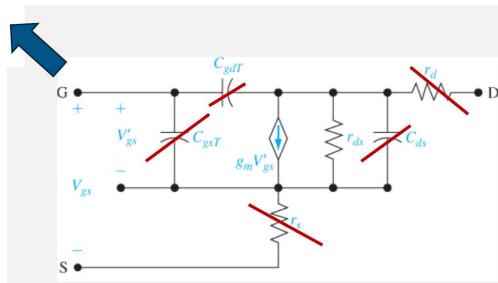


Figure 10.55 | Simplified, lowfrequency small-signal equivalent circuit of common-source n-channel MOSFET including source resistance  $r_s$ .

$$\begin{aligned} I_d &= g_m V'_{gs} \\ V_{gs} &= V'_{gs} + (g_m V'_{gs}) r_s = (1 + g_m r_s) V'_{gs} \\ I_d &= \left( \frac{g_m}{1 + g_m r_s} \right) V_{gs} = g'_m V_{gs} \end{aligned}$$

The source resistance reduces the effective transconductance or transistor gain!

- Channel transit time - not a problem!
- Gate or capacitance charging time

## Small signal model include $r_s$

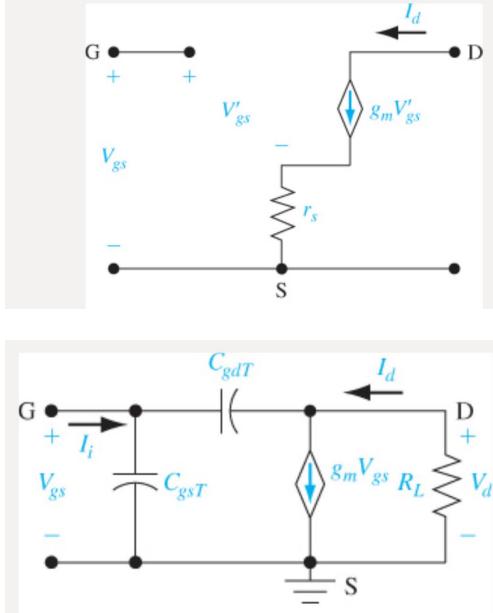


Figure 10.56 | High-frequency smallsignal equivalent circuit of commonsource n-channel MOSFET.

$$I_i = j\omega C_{gsT} V_{gs} + j\omega C_{gdT} (V_{gs} - V_d)$$

$$\frac{V_d}{R_L} + g_m V_{gs} + j\omega C_{gdT} (V_d - V_{gs}) = 0$$

$$I_i = j\omega [C_{gsT} + C_{gdT} (1 + g_m R_L)] V_g$$

$$C_M = C_{gdT} (1 + g_m R_L)$$

$$I_i = j\omega \left[ C_{gs:T} + C_{gdT} \left( \frac{1 + g_m R_L}{1 + j\omega R_L C_{grt}} \right) \right] V_{gs} \quad I_i = j\omega [C_{gsT} + C_{gdT} (1 + g_m R_L)] V_{gs}$$

- drain overlap capacitance!
- When the transistor is operating in the saturation region,  $C_{gd}$  essentially becomes zero, but  $C_{gdP}$  is a constant. This parasitic capacitance is multiplied by the gain of the transistor and can become a significant factor in the input impedance.

### Cutoff frequency

- The cutoff frequency  $f_T$  is defined to be the frequency at which the magnitude of the current gain of the device is unity, or when the magnitude of the input current  $I_i$  is equal to the ideal load current  $I_d$ .

$$f_T = \frac{g_m}{2\pi(C_{gsT} + C_M)} = \frac{g_m}{2\pi C_G} \quad C_G \text{ is the equivalent input gate capacitance}$$

Ideal:

$$g_{ms} = \frac{W\mu_n C_{ox}}{L} (V_{GS} - V_T)$$

$$f_T = \frac{g_m}{2\pi C_G} = \frac{\frac{W\mu_n C_{ox}}{L} (V_{GS} - V_T)}{2\pi (C_{ox}WL)} = \frac{\mu_n (V_{GS} - V_T)}{2\pi L^2}$$

### Small-signal equivalent circuit

- Just as for diode, a small ac signal is added on top of some dc bias operating the transistor
- Must consider all of the impedance effects (capacitances and resistances) in the MOSFET

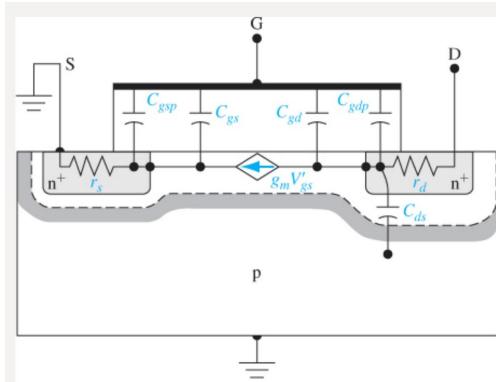
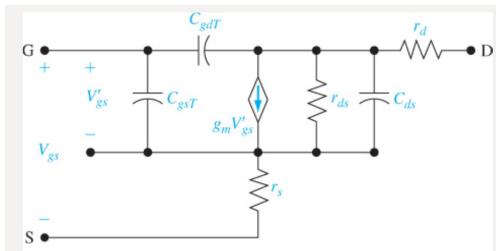


Figure 10.52 | Inherent resistances and capacitances in the n-channel MOSFET structure.

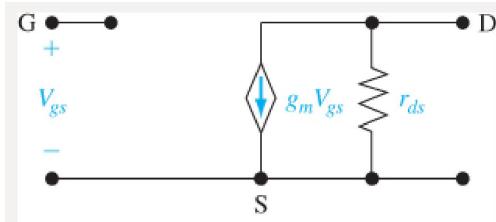
### Small signal equivalent circuit



- $r_{ds}$  due to channel length modulation

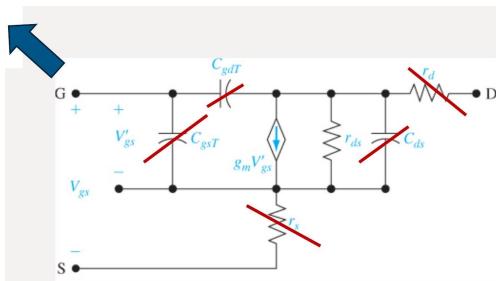
Figure 10.53 | Small-signal equivalent circuit of a commonsource n-channel MOSFET.

### Small signal model at low frequencies



- At low f, all capacitances are open circuit
- Neglect  $r_s$  and  $r_d$
- Input gate impedance is infinite

Figure 10.54 | Simplified, low-frequency small-signal equivalent circuit of a common-source n-channel MOSFET.



### Small signal mode include $r_s$

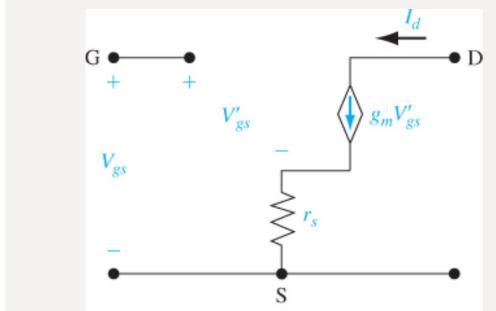


Figure 10.55 | Simplified, lowfrequency small-signal equivalent circuit of common-source n-channel MOSFET including source resistance  $r_s$ .

$$\begin{aligned}
 I_d &= g_m V'_{gs} \\
 V_{gs} &= V'_{gs} + (g_m V'_{gs}) r_s = (1 + g_m r_s) V'_{gs} \\
 I_d &= \left( \frac{g_m}{1 + g_m r_s} \right) V_{gs} = g'_m V_{gs}
 \end{aligned}$$

The source resistance reduces the effective transconductance or transistor gain!

- Channel transit time - not a problem!
- Gate or capacitance charging time

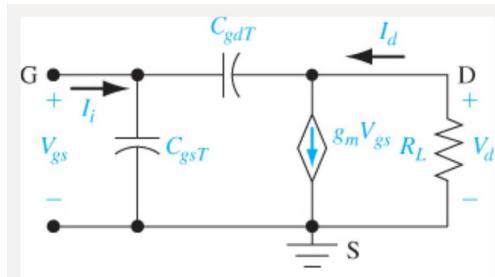


Figure 10.56 | High-frequency small-signal equivalent circuit of commonsource n-channel MOSFET.

$$\begin{aligned}
 I_i &= j\omega C_{gsT} V_{gs} + j\omega C_{gdT} (V_{gs} - V_d) \\
 \frac{V_d}{R_L} + g_m V_{gs} + j\omega C_{gdT} (V_d - V_{gs}) &= 0 \\
 I_i &= j\omega [C_{gsT} + C_{gdT} (1 + g_m R_L)] V_g \\
 C_M &= C_{gdT} (1 + g_m R_L)
 \end{aligned}$$

$$I_i = j\omega \left[ C_{gs:T} + C_{gdT} \left( \frac{1 + g_m R_L}{1 + j\omega R_L C_{grt}} \right) \right] V_{gs} \quad I_i = j\omega [C_{gsT} + C_{gdT} (1 + g_m R_L)] V_{gs}$$

- drain overlap capacitance!
- When the transistor is operating in the saturation region,  $C_{gd}$  essentially becomes zero, but  $C_{gdP}$  is a constant. This parasitic capacitance is multiplied by the gain of the transistor and can become a significant factor in the input impedance.

### Cutoff frequency

- The cutoff frequency  $f_T$  is defined to be the frequency at which the magnitude of the current gain of the device is unity, or when the magnitude of the input current  $I_i$  is equal to the ideal load current  $I_d$ .

$$f_T = \frac{g_m}{2\pi(C_{gsT} + C_M)} = \frac{g_m}{2\pi C_G} \quad C_G \text{ is the equivalent input gate capacitance}$$

Ideal:

$$g_{ms} = \frac{W\mu_n C_{ox}}{L} (V_{GS} - V_T)$$

$$f_T = \frac{g_m}{2\pi C_G} = \frac{\frac{W\mu_n C_{ox}}{L} (V_{GS} - V_T)}{2\pi (C_{ox}WL)} = \frac{\mu_n (V_{GS} - V_T)}{2\pi L^2}$$

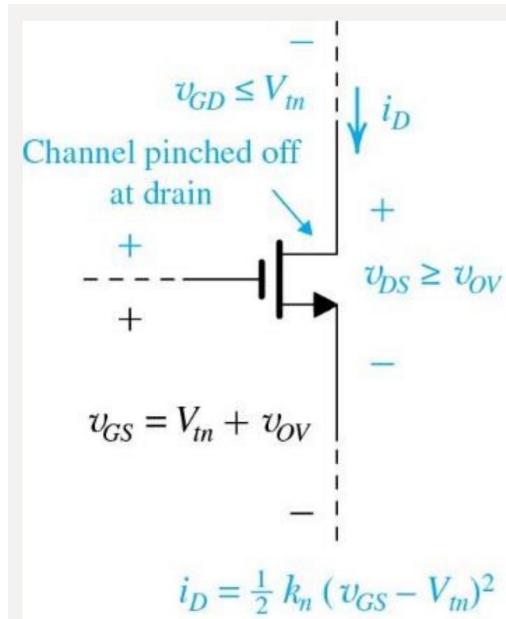
# Lecture 21: MOSFET - common source amplifier

---

## Topics

- Class notes
- MOSFET as an amplifier
- DC bias point - quiescent point
- Small signal operation
- Gain and transconductance
- Unity gain frequency
- Miller effect
- Ref: Sedra Smith 8<sup>th</sup> edition (Chapters 7, 10)

## MOSFET as an amplifier



- The basis for using the transistor (a MOSFET) in amplifier design is that when we operate the device in the active region, we create a voltage-controlled current source.
- Operating an NMOS transistor in the active mode.

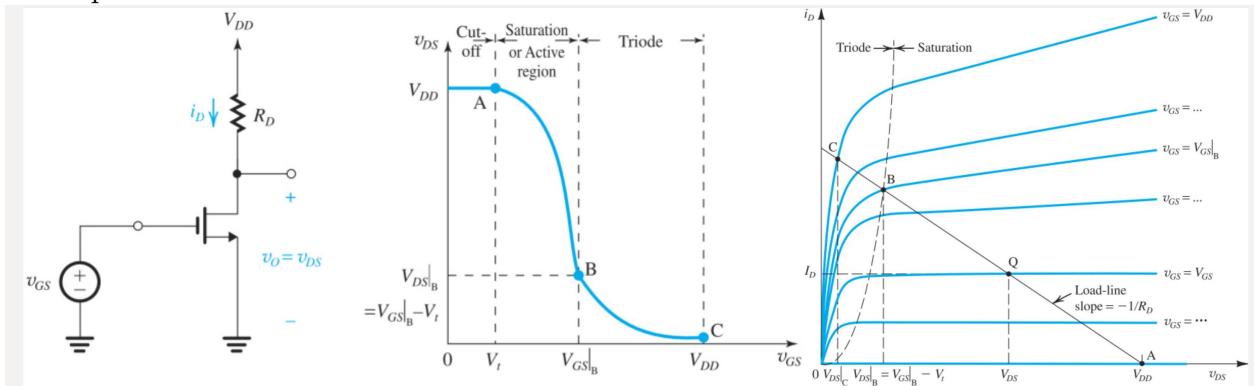
- $v_{GS} = V_{in} + v_{OV}$  and  $v_{DS} \geq v_{OV}$
- Thus,  $v_{GD} \leq V_{in} \Rightarrow$  channel is pinched - off at drain end
- $v_{OV} = v_{GS} - V_{tn}$
- The control relationship is nonlinear

$$i_D = \frac{1}{2}k_n(v_{GS} - V_{tn})^2$$

- Transconductance amplifier: input is voltage, output is current

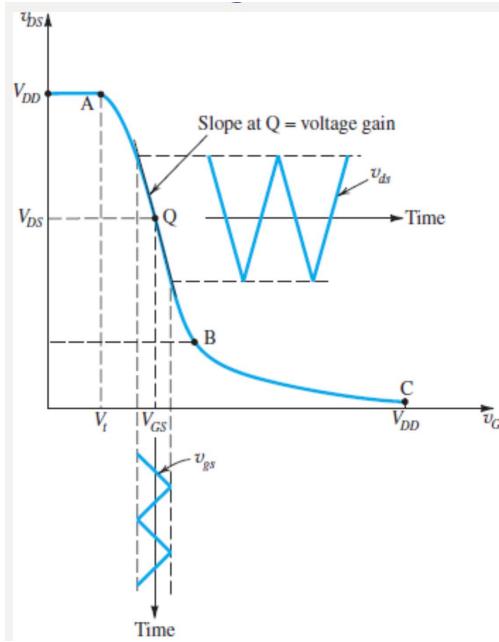
## NMOS amplifier: voltage transfer characteristics

- A simple way to convert a transconductance amplifier to a voltage amplifier is to pass the output current through a resistor and take the voltage across the resistor as the output



$$v_{DS} = V_{DD} - i_D R_D$$

## Obtaining Linear Amplification by Biasing the Transistor



$$V_{DS} = V_{DD} - \frac{1}{2}k_n R_D (V_{GS} - V_t)^2$$

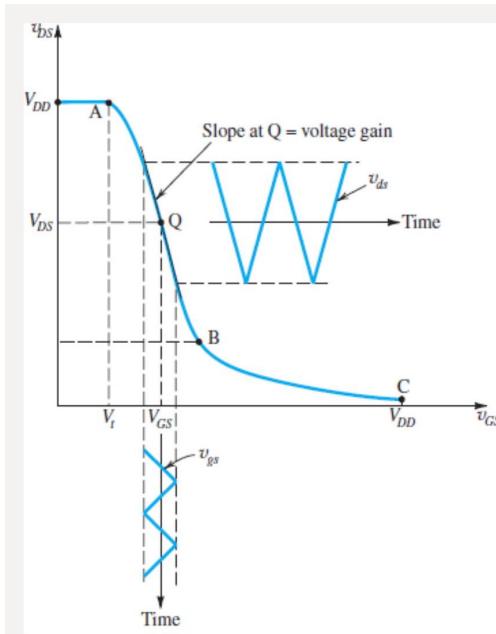
- Point Q is known as the bias point or the dc operating point
- Since at Q no signal component is present, it is also known as the quiescent point (hence the symbol Q).
- The signal to be amplified,  $v_{gs}$ , a function of time  $t$ , is superimposed on the bias voltage  $V_{GS}$

$$v_{GS}(t) = V_{GS} + v_{gs}(t)$$

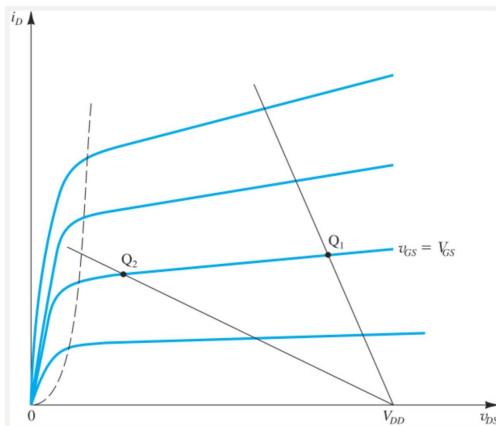
- The MOSFET operates on a short, almost-linear segment of the VTC around the bias point Q and provides an output voltage  $v_{ds} = A_v v_{gs}$ .
- Where should Q be?
- What is  $v_{gs}$  is too high?
- Small signal voltage gain:  $A_v = \left. \frac{dv_{DS}}{dv_{GS}} \right|_{v_{GS}=V_{GS}}$

### Deciding on the bias point Q

- Q is determined by  $V_{GS}$  and  $R_D$



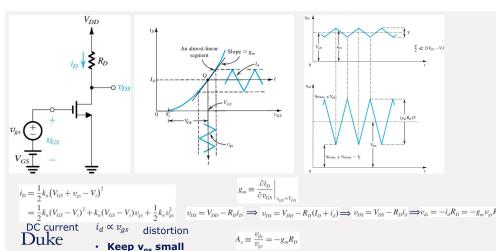
- For a given  $R_D$ , locate  $Q$  as close to the triode region (point B) as possible to obtain high gain but sufficiently distant to allow for the required negative signal swing.



Q1: Not sufficient "headroom"

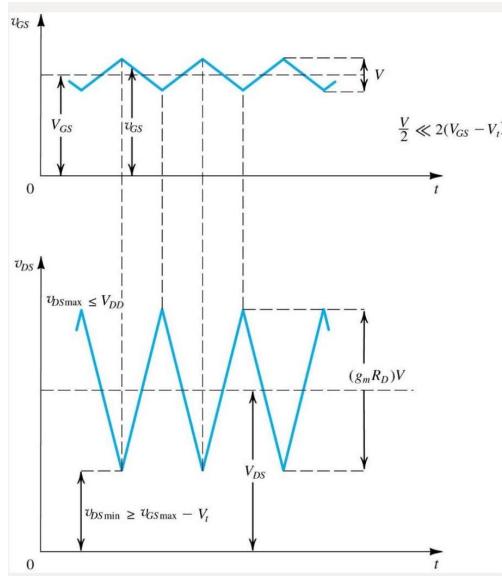
Q2: Not sufficient "legroom"

## Small signal operation of a MOSFET amplifier



$$\begin{aligned}
 i_D &= \frac{1}{2}k_n (V_{GS} + v_{gs} - V_t)^2 \\
 &= \frac{1}{2}k_n (V_{GS} - V_t)^2 + k_n (V_{GS} - V_t) v_{gs} + \frac{1}{2}k_n v_{gs}^2 \quad v_{DS} = V_{DD} - R_D i_D \Rightarrow v_{DS} = V_{DD} - \\
 R_D (I_D + i_d) \Rightarrow v_{DS} &= V_{DS} - R_D i_d \Rightarrow v_{ds} = -i_d R_D = -g_m v_{gs} R_D \\
 \text{DC current}
 \end{aligned}$$

$$A_v \equiv \frac{v_{ds}}{v_{gs}} = -g_m R_D$$



ent  $i_d \propto v_{gs}$

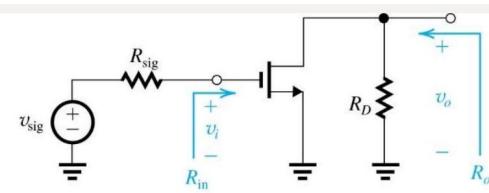
- Keep  $v_{gs}$  small

\$ \\$ \\ \$

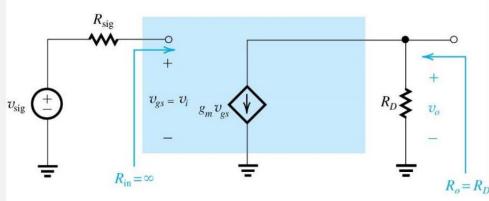
\$ \\$ \\ \$

## Common source configuration

- connect the source terminal to ground, apply the input voltage signal  $v_i$  between the gate and ground, and take the output voltage signal  $v_o$  between the drain and ground, across the resistance  $R_D$ .
- grounded-source or common-source (CS) amplifier
- determine  $R_{in}$ ,  $A_{vo}$  and  $R_o$      $R_{in} = \infty$



(a)



(b)

$$v_i = v_{\text{sig}}$$

$$v_{gs} = v_i$$

$$v_o = -g_m v_{gs} R_D$$

$$R_o = R_D$$

$$A_{vo} = \frac{v_o}{v} = -g_m R_D$$

- If a load resistance  $R_L$  is connected across  $R_D$ , the voltage gain  $A_v$  can be obtained from:

$$\begin{aligned} A_v &= A_{vo} \frac{R_L}{R_L + R_o} \\ A_v &= -g_m (R_D \| R_L) \\ G_v \equiv \frac{v_o}{v_{\text{sig}}} &= -g_m (R_D \| R_L) \end{aligned}$$

## Common source amplifier summary

- The CS amplifiers are the most useful of all transistor amplifier configurations. They exhibit a moderate-to-high input resistance (infinite for the CS), a moderate-to-high output resistance, and reasonably high voltage gain.

- Reducing  $R_D$  to lower the output resistance of the CS amplifier, is usually not a viable proposition because this also reduces the voltage gain. Alternatively, if we need a very low output resistance (in the ohms or tens-of-ohms range), we can use a source-follower stage between the output of the CS amplifier and the load resistance
- Although the CS configurations are the workhorses of transistor amplifiers, they suffer from a limitation on their high-frequency response.

## High frequency transistor model

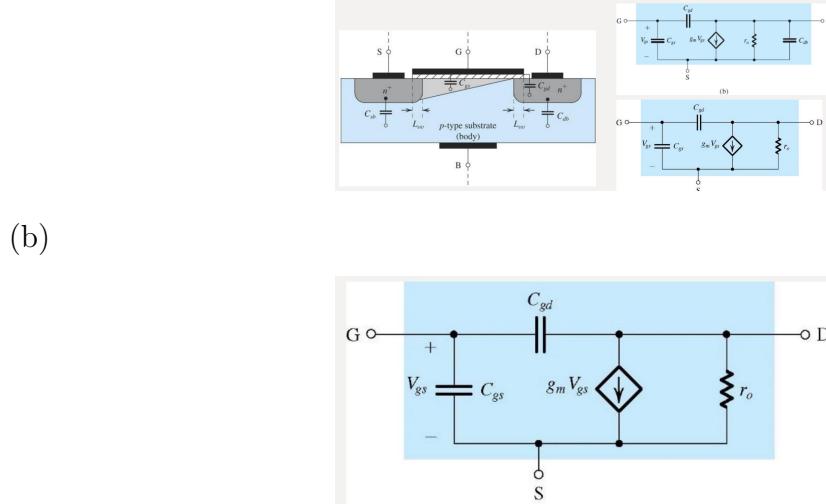


Figure 10.4 (a) High-frequency, equivalent-circuit model for the MOSFET. (b) The equivalent circuit for the case in which the source is connected to the substrate (body). (c) The equivalent-circuit model of (b) with  $C_{db}$  neglected (to simplify analysis).

## Unity gain frequency $f_T$

The MOSFET Unity-Gain Frequency ( $f_T$ ) :

A figure of merit for the high-frequency operation of the MOSFET as an amplifier is the unity-gain frequency,  $f_T$ , also known as the transition frequency, which gives rise to the subscript T. This is defined as the frequency at which the short-circuit current gain of the common-source configuration becomes unity.

## $f_T$ related to device parameters

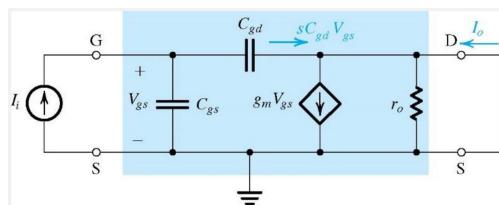


Figure 10.5 Determining the short-circuit current gain  $I_o/I_i$ .

$$I_o = g_m V_{gs} - sC_{gd}V_{gs} = (g_m - sC_{gd})V_{gs}$$

$$V_{gs} = \frac{I_i}{s(C_{gs} + C_{gd})}$$

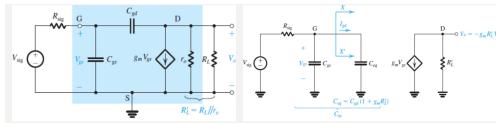
$$\frac{I_o}{I_i} = \frac{g_m - sC_{gd}}{s(C_{gs} + C_{gd})} \Rightarrow \left| \frac{I_o}{I_i} \right| \simeq \frac{g_m}{\omega(C_{gs} + C_{gd})}$$

For unity gain:

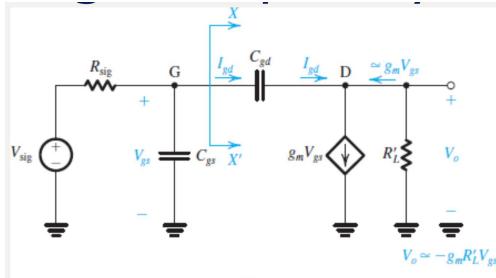
$$\omega_T = g_m / (C_{gs} + C_{gd}) \Rightarrow f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})}$$

Since  $f_T$  is proportional to  $g_m$ , which determines the midband gain, and inversely proportional to the MOSFET internal capacitances, which limit the amplifier bandwidth, the higher the value of  $f_T$ , the more effective the MOSFET becomes as an amplifier.

High frequency response of CS amplifier



## High frequency response of CS amplifier



(e)

- Focus on the input side of the circuit
- Simplify the input circuit to a simple RC low pass network
- Replace bridging capacitor  $C_{gd}$  by  $C_{eq}$  between node G and ground

$$V_o \simeq -(g_m V_{gs}) R'_L = -g_m R'_L V_{gs}$$

$$I_{gd} = sC_{gd}(V_{gs} - V_o)$$

$$= sC_{gd}[V_{gs} - (-g_m R_L V_{gs})]$$

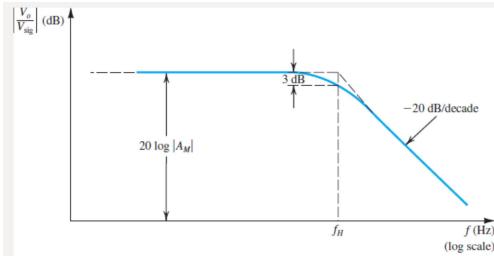
$$= sC_{gd}(1 + g_m R'_L) V_g$$

$$sC_{eq} V_{gs} = sC_{gd}(1 + g_m R'_L) V_{gs}$$

$$C_{eq} = C_{Rd} (1 + g_m R'_L)$$

Thus  $C_{gd}$  gives rise to a much larger capacitance  $C_{eq}$ , which appears at the amplifier input. The multiplication effect that  $C_{gd}$  undergoes comes about because it is connected between circuit nodes  $G$  and  $D$ , whose voltages are related by a large negative gain. This effect is known as the Miller effect, and  $1 + g_m R'_L$  is known as the Miller multiplier.

## Frequency response of gain



(e)

The frequency-response plot, which is that of a low-pass, Singletime-constant circuit.

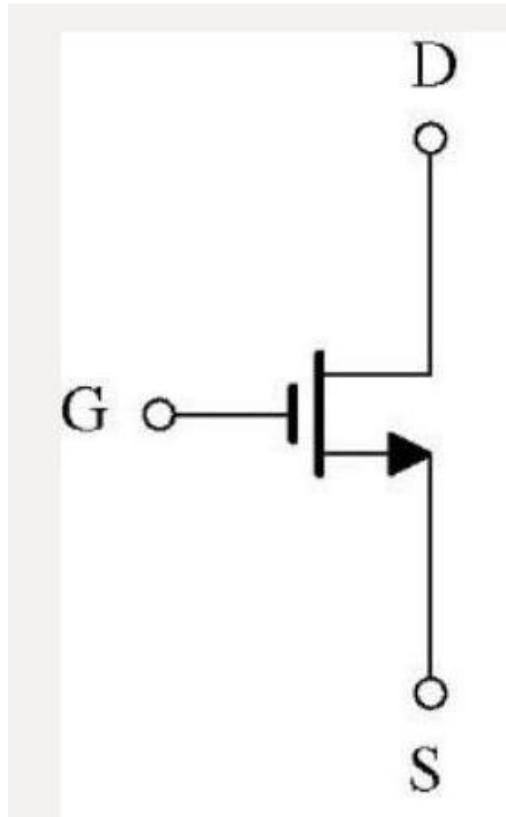
# Lecture 22: Inverter

---

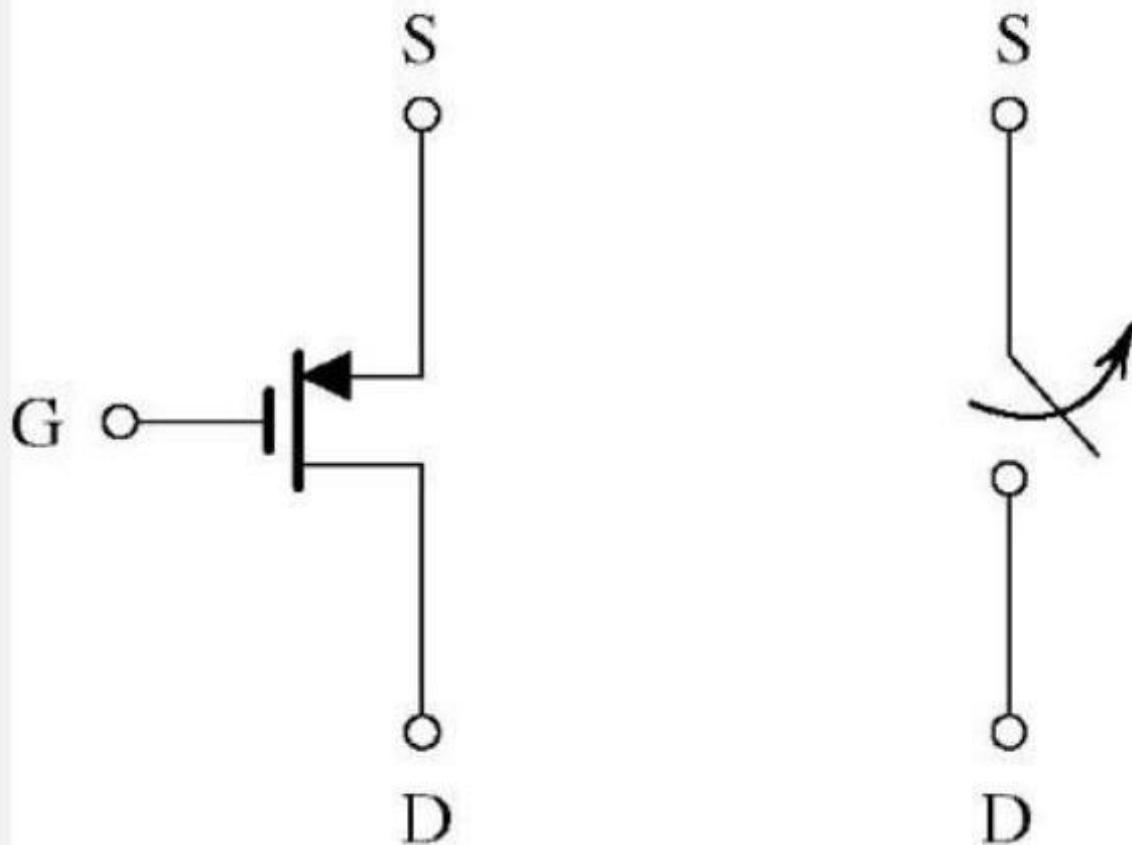
## Topics

- Class notes
- Digital logic
- Pull-up and pull-down network
- Inverter VTC parameters
- CMOS inverter
- Ref: Sedra Smith 8<sup>th</sup> edition (Chapter 16)

## NMOS and PMOS transistors as switch



(a)



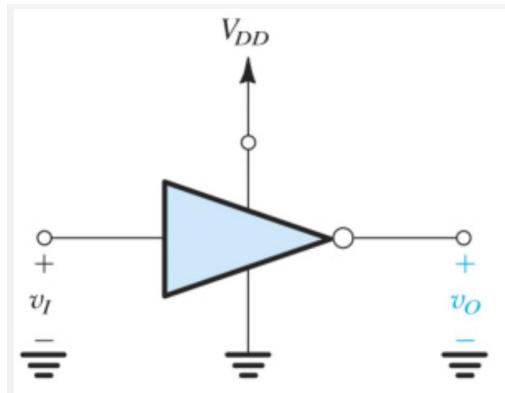
$$V_G = V_{DD} \\ (G = 1)$$

(b)

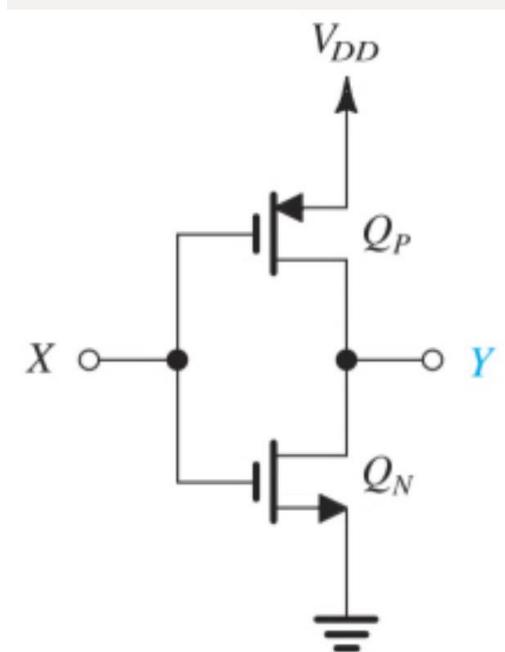
- CMOS circuits for combinational logic functions
- NMOS and PMOS are used as switches
- Gate voltage - input

Figure 16.1 Operation of the (a) NMOS and (b) PMOS transistor as an on/off switch. The gate voltage controls the operation of the transistor switch, with the voltage  $V_{DD}$  representing a logic 1 and 0 V representing a logic 0 . Note that the connections of the drain and source terminals are not shown.

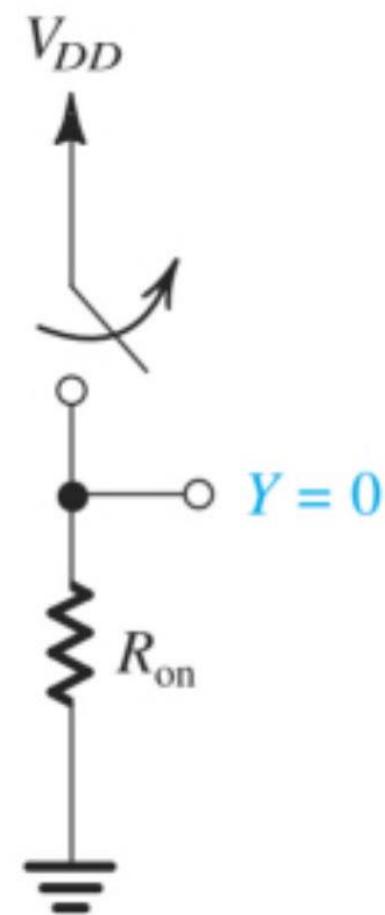
CMOS inverter design  $Y = \bar{X}$



(a)

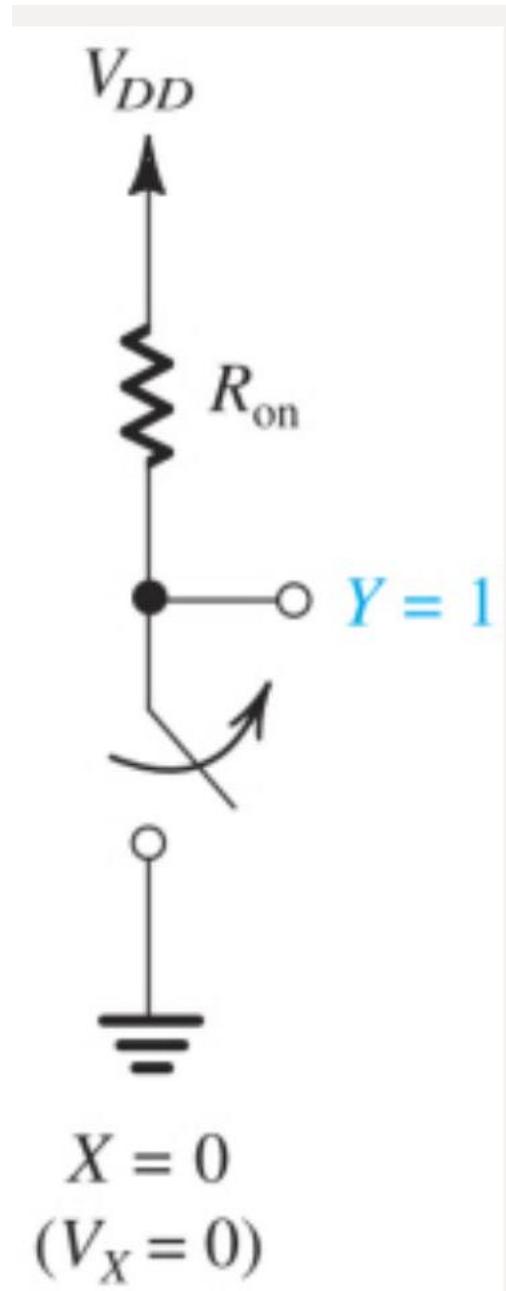


(b)



$X = 1$   
 $(V_X = V_{DD})$

(c)

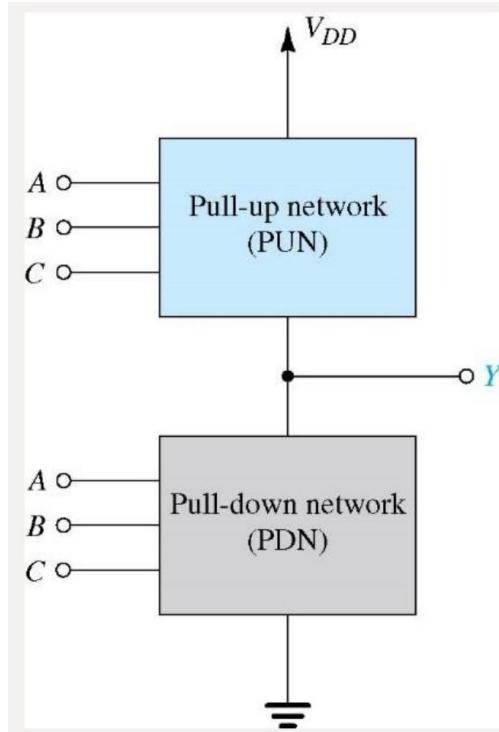


(d)

Figure 16.2 (a) Block representation of the logic inverter; (b) its CMOS realization; (c) operation when the

input is a logic 1; (d) operation when the input is a logic 0 .

## Pull-up and pull-down network

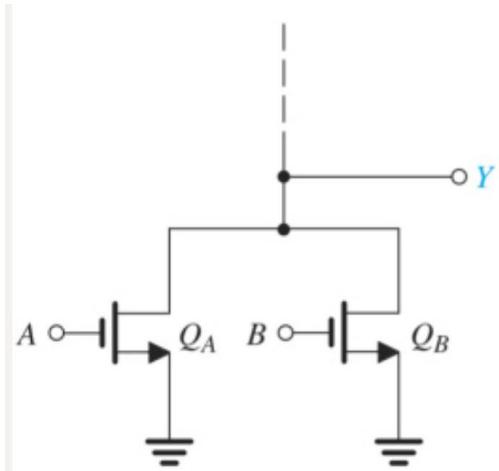


- A CMOS logic circuit is in effect an extension, or a generalization, of the CMOS inverter
- The inverter consists of an NMOS pull-down transistor and a PMOS pull-up transistor, operated by the input voltage in a complementary fashion.
- The CMOS logic gate consists of two networks:
  - the pull-down network (PDN) constructed of NMOS transistors
  - the pull-up network (PUN) constructed of PMOS transistors

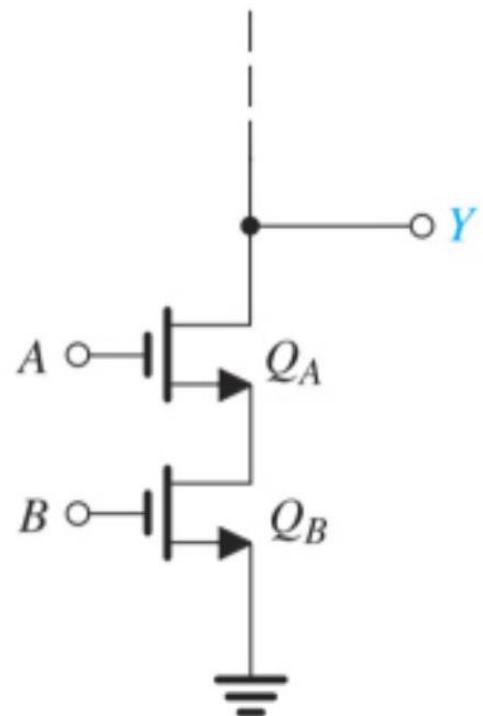
Figure 16.3 Representation of a three-input CMOS logic gate. The PUN comprises PMOS transistors, and the PDN comprises NMOS transistors.

### Pull down network (PDN)

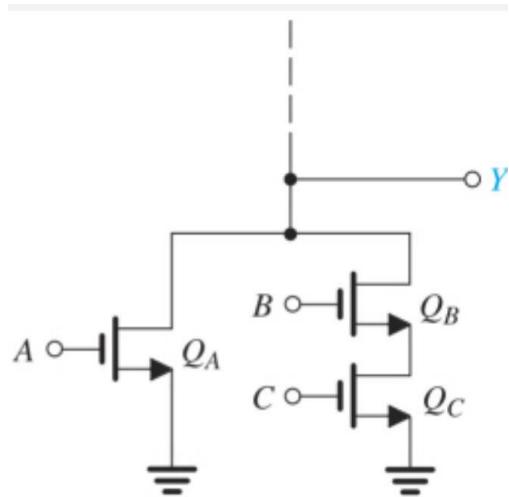
- Consists of NMOS transistor
- Activated when inputs are high
- The PDN utilizes devices in parallel to form an OR function and devices in series to form an AND function



(a)



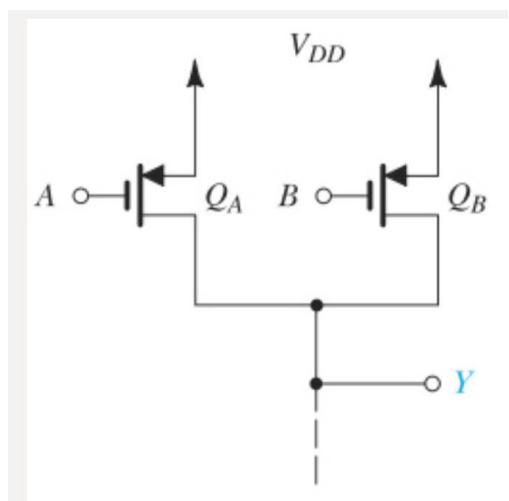
(b)



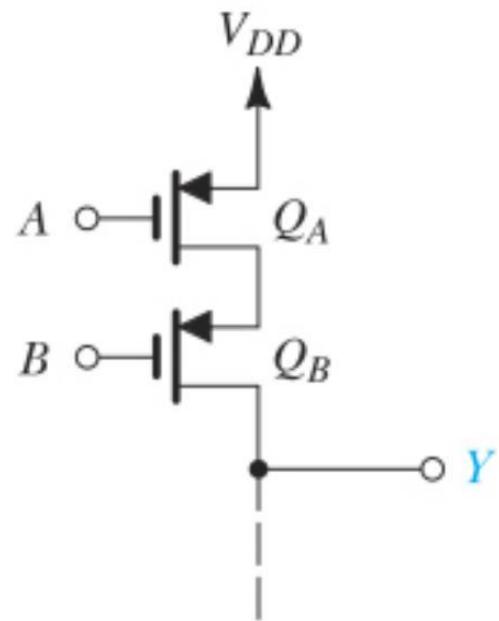
(c)

### Pull up network (PUN)

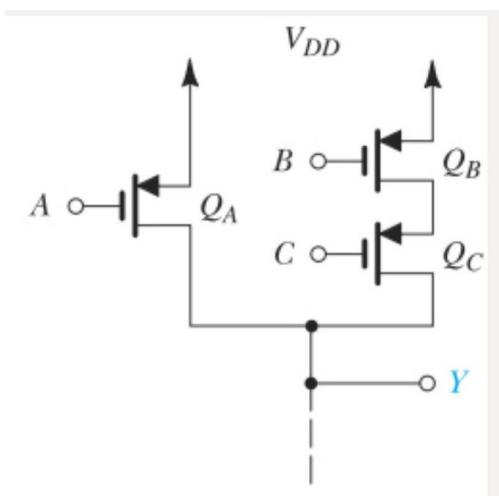
- Consists of PMOS transistor
- Activated when inputs are low
- The PUN utilizes devices in parallel to form an OR function and devices in series to form an AND function



(a)



(b)



(c)

The two-input NOR gate

## The two-input CMOS NAND gate

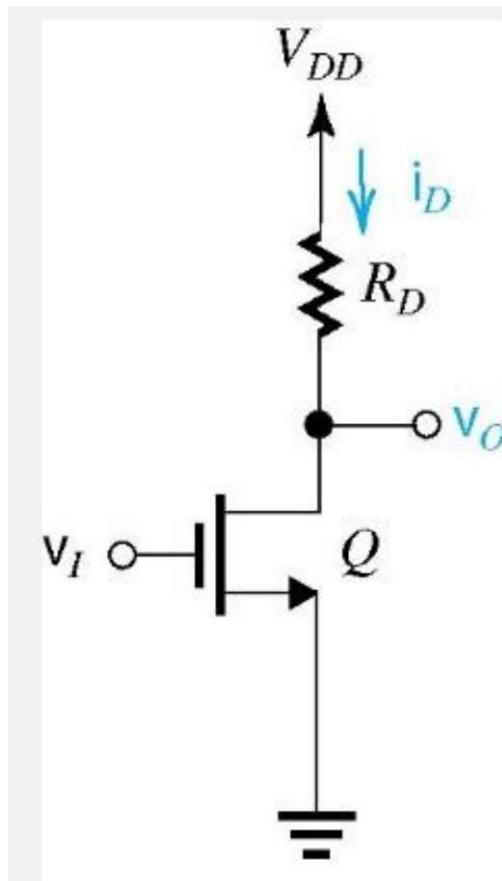
**Example**  $Y = \overline{A(B + CD)}$

**Example: XOR gate**

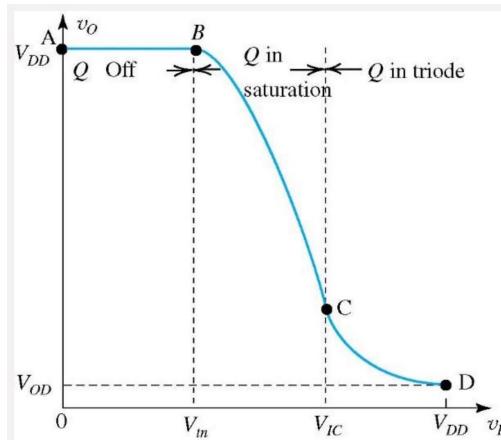
**Review: PDN and PUN**

- The PDN can be most directly synthesized by expressing as a function of the uncomplemented variables. If complemented variables appear in this expression, additional inverters will be required to generate them.
- The PUN can be most directly synthesized by expressing Y as a function of the complemented variables and then applying the uncomplemented variables to the gates of the PMOS transistors. If uncomplemented variables appear in the expression, additional inverters will be needed.
- The PDN can be obtained from the PUN (and vice versa) using the duality property.

## Inverter with NMOS



(a)



(b)

- For a logic-0 input,  $v_I$  is close to 0 V and specifically lower than the MOSFET threshold voltage  $V_{tn}$
- the transistor will be off
- $i_D = 0$
- $V_O = V_{DD}$ , which is a logic 1
- For a logic-1 input,  $v_I = V_{DD}$
- the transistor will be conducting and operating in the linear region (at point D on the VTC)
- the output voltage will be low (logic 0)

Thus, to use this amplifier as a logic inverter, we utilize its extreme regions of operation. This is exactly the opposite to its use as a signal amplifier, where it would be biased at the middle of the transfer characteristic segment  $BC$  and the signal kept small enough to restrict operation to a short, almost linear, segment of the transfer curve. Digital applications, on the other hand, make use of the gross nonlinearity exhibited by the VTC.

## Parameters of VTC for inverter

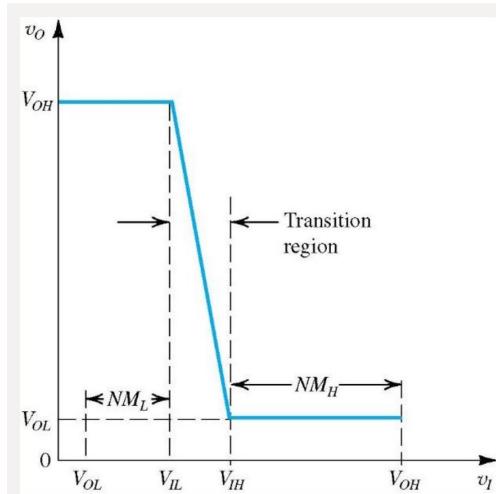


Figure 16.13 Voltage-transfer characteristic of an inverter. The VTC is approximated by three straight-line segments. Note the four parameters of the VTC ( $V_{OH}$ ,  $V_{OL}$ ,  $V_{IL}$ , and  $V_{IH}$ ) and their use in determining the noise margins ( $NM_H$  and  $NM_L$ ).

Table 16.1 Important Parameters of the VTC of the Logic Inverter

$V_{OL}$  : Output low level

$V_{OH}$  : Output high level

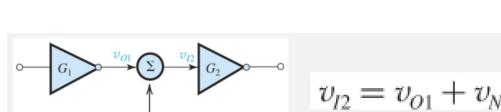
$V_{IL}$  : Maximum value of input interpreted by the inverter as a logic 0

$V_{IH}$  : Minimum value of input interpreted by the inverter as a logic 1

$NM_L$  : Noise margin for low input =  $V_{IL} - V_{OL}$

$NM_H$  : Noise margin for high input =  $V_{OH} - V_{IH}$

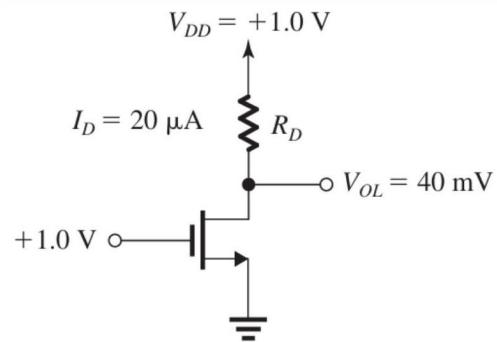
Observe that changes in the input signal level within the noise margins are rejected by the inverter. Thus noise is not allowed to propagate further through the system, a definite advantage of digital over analog circuits.



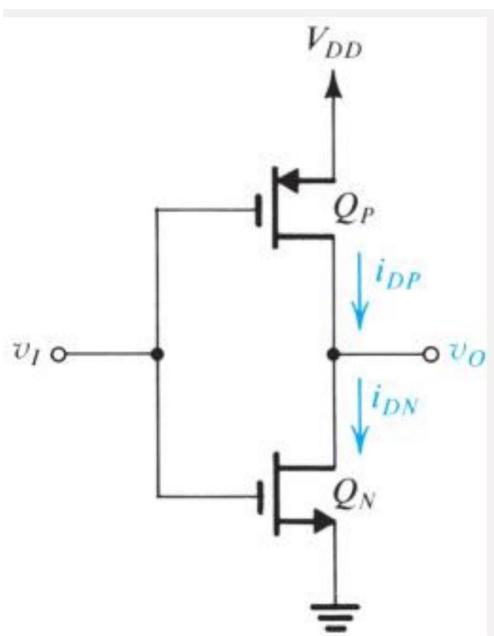
Design the inverter in Fig. 16.12(a) to provide  $V_{OL} = 40\text{mV}$  and to draw a supply current of  $20 \mu\text{A}$  in the low-output state. Let the transistor be specified to have  $V_t = 0.35 \text{ V}$ ,  $\mu_n C_{ox} = 540 \mu\text{A/V}^2$ , and  $\lambda = 0$ . The power supply  $V_{DD} = 1.0 \text{ V}$ . Specify the required values of  $W/L$  and  $R_D$ . How much power is drawn from  $V_{DD}$  when the switch is open? Closed?

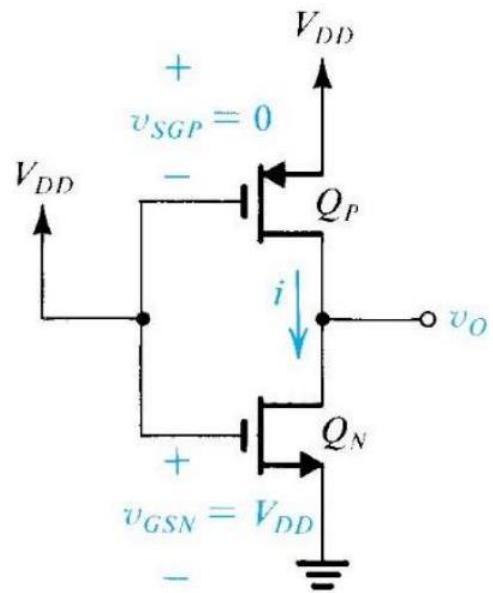
Hint: Recall that for small  $v_{DS'}$

$$r_{DS} \simeq 1 / \left[ (\mu_n C_{ox}) \left( \frac{W}{L} \right) (V_{GS} - V_t) \right]$$

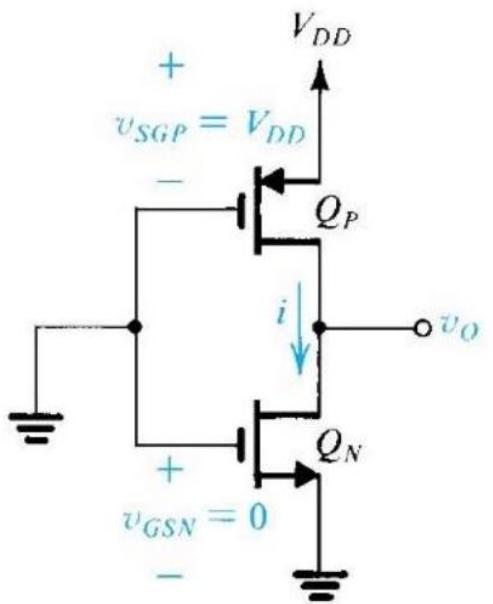


## The CMOS inverter

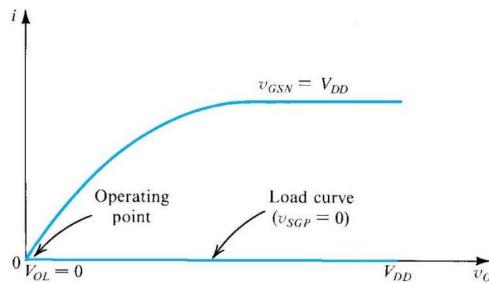




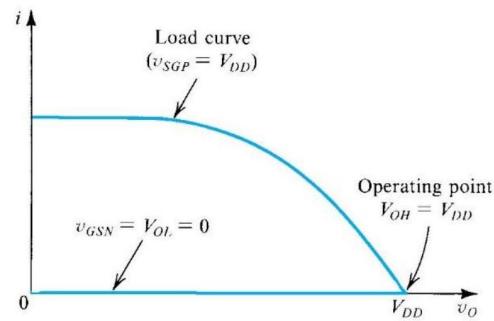
(a)



(a)



(b)  
 $v_1$  high



(b)

$$r_{DSN} = 1 / \left[ k'_n \left( \frac{W}{L} \right)_n (V_{DD} - V_{tn}) \right]$$

Power drawn = 0

(c)

\$\$\begin{gathered} \\ \text{\textbackslash begin\{gathered\}} \\ \text{\textbackslash V\\_\{1\}} \end{gathered}\$\$

=

\end{gathered} \\ \$\$

(c)

$$r_{DSP} = 1 / \left[ k'_p \left( \frac{W}{L} \right)_p (V_{DD} - |V_{tp}|) \right]$$

Power drawn = 0

**Note**

- In spite of the fact that the quiescent current is zero, the load-driving capability of the CMOS inverter is high.
- For instance, with the input high, transistor  $Q_N$  can sink a relatively large load current. This current can quickly discharge the load capacitance.
- Because of its action in sinking load current and thus pulling the output voltage down toward ground, transistor  $Q_N$  is known as the pull-down device.
- Similarly, with the input low, transistor  $Q_p$  can source a relatively large load current. This current can quickly charge up a load capacitance, thus pulling the output voltage up toward  $V_{DD}$ .
- Hence,  $Q_p$  is known as the pull-up device.

**CMOS logic inverter review**

1. The output voltage levels are 0 and  $V_{DD}$ , and thus the signal swing is the maximum possible. This, coupled with the fact that the inverter can be designed to provide a symmetrical voltage-transfer characteristic, results in wide noise margins.
2. The static power dissipation in the inverter is zero (neglecting the dissipation due to leakage currents) in both of its states. This is because no dc path exists between the power supply and ground in either state.
3. A low-resistance path exists between the output terminal and ground (in the low-output state) or  $V_{DD}$  (in the high-output state). These low-resistance paths ensure that the output voltage is 0 or  $V_{DD}$  independent of the exact values of the W/L ratios or other device parameters. Furthermore, the low output resistance makes the inverter less sensitive to the effects of noise and other disturbances.
4. The active pull-up and pull-down devices provide the inverter with high output driving capability in both directions. This speeds up the operation considerably.
5. The input resistance of the inverter is infinite (because  $I_G = 0$ ). Thus the inverter can drive an arbitrarily large number of similar inverters with no loss in signal level. Of course, each additional inverter increases the load capacitance on the driving inverter and slows down the operation. Later we will consider the inverter switching times.

$$i_{DN} = k'_n \left( \frac{W}{L} \right)_n \left[ (v_l - V_m) v_O - \frac{1}{2} v_O^2 \right] \quad \text{for } v_O \leq v_l - V_m$$

and

$$i_{DN} = \frac{1}{2} k'_n \left( \frac{W}{L} \right)_n (v_l - V_m)^2 \quad \text{for } v_o \geq v_l - V_m$$

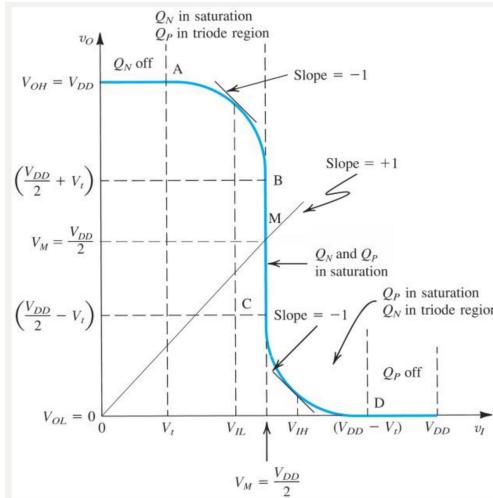
For  $Q_p$

$$i_{DP} = k'_p \left( \frac{W}{L} \right)_p [(V_{DD} - v_l - |V_{tp}|)(V_{DD} - v_o) - \frac{1}{2}(V_{DD} - v_o)^2] \quad \text{for } v_o \geq v_l + |V_{tp}|$$

The CMOS inverter is usually designed to have  $V_{tn} = |V_{tp}| = V_t$ . Also, although this is not always the case, we shall assume that  $Q_N$  and  $Q_p$  are matched; that is,  $k'_n(W/L)_n = k'_p(W/L)_p$ . It should be noted that since  $\mu_p$  is often 0.25 to 0.5 times the value of  $\mu_n$  to make  $k'(W/L)$  of the two devices equal, the width of the  $p$ -channel device is made two to four times that of the  $n$ -channel device. More specifically, the two devices are designed to have equal lengths, with widths related by

$$\frac{W_p}{W_n} = \frac{\mu_n}{\mu_p} \quad (16.32)$$

$$i_{DP} = \frac{1}{2} k'_p \left( \frac{W}{L} \right)_p (V_{DD} - v_l - |V_{tp}|)^2 \quad \text{for } v_o \leq v_l + |V_{tp}|$$



$$V_M = \frac{V_{DD}}{2}$$

$$(v_I - V_t)v_O - \frac{1}{2}v_O^2 = \frac{1}{2}(V_{DD} - v_I - V_t)^2 \quad (16.33)$$

Differentiating both sides relative to  $v$ , results in

$$(v_I - V_t) \frac{dv_O}{dv_I} + v_O - v_O \frac{dv_O}{dv_I} = -(V_{DD} - v_I - V_t)$$

in which we substitute  $v_I = V_{IH}$  and  $dv_O/dv_I = -1$  to obtain

$$v_O = V_{IH} - \frac{V_{DD}}{2} \quad (16.34)$$

Substituting  $v_l = V_{IH}$  and for  $v_O$  from Eq. (16.34) in Eq. (16.33) gives

$$V_{IH} = \frac{1}{8} (5V_{DD} - 2V_t) \quad (16.35)$$

$V_{IL}$  can be determined in a manner similar to that used to find  $V_{IH}$ . Alternatively, we can use the symmetry relationship

$$V_{IH} - \frac{V_{DD}}{2} = \frac{V_{DD}}{2} - V_{IL}$$

together with  $V_{IH}$  from Eq. (16.35) to obtain

$$V_{IL} = \frac{1}{8} (3V_{DD} + 2V_t) \quad (16.36)$$

The noise margins can now be determined as follows:

$$\begin{aligned} NM_H &= V_{OH} - V_{IH} \\ &= V_{DD} - \frac{1}{8} (5V_{DD} - 2V_t) \\ &= \frac{1}{8} (3V_{DD} + 2V_t) \end{aligned} \quad (16.37)$$

$$\begin{aligned} NM_L &= V_{IL} - V_{OL} \\ &= \frac{1}{8} (3V_{DD} + 2V_t) - 0 \\ &= \frac{1}{8} (3V_{DD} + 2V_t) \end{aligned} \quad (16.38)$$

As expected, the symmetry of the voltage-transfer characteristic results in equal noise margins. Of course, if  $Q_N$  and  $Q_P$  are not matched, the voltage-transfer characteristic will no longer be symmetric, and the noise margins will not be equal.

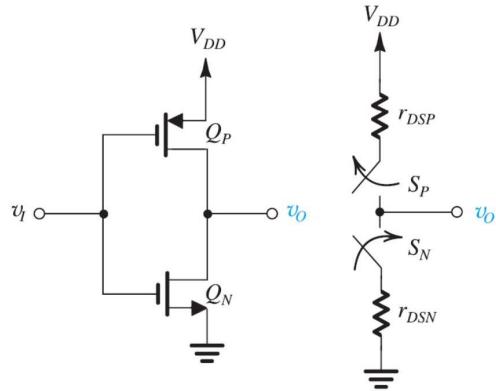
Table 16.2 Summary of Important Static Characteristics of the CMOS Logic Inverter  
Inverter Output Resistance

When  $v_o$  is low (current sinking):

$$r_{DSN} = 1 / \left[ k'_n \left( \frac{W}{L} \right)_n (V_{DD} - V_{tn}) \right]$$

When  $v_O$  is high (current sourcing):

$$r_{DSP} = 1 / \left[ k'_p \left( \frac{W}{L} \right)_p (V_{DD} - |V_{tp}|) \right]$$



### Inverter VTC and Noise Margins

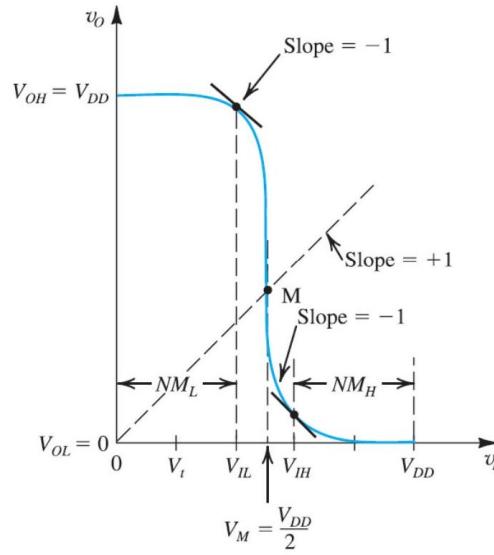
$$V_M = \frac{r(V_{DD} - |V_{tp}|) + V_{tn}}{1+r} \quad \text{where} \quad r = \sqrt{\frac{k'_p(W/L)_p}{k'_n(W/L)_n}}$$

For matched devices, that is,  $\mu_n \left(\frac{W}{L}\right)_n = \mu_p \left(\frac{W}{L}\right)_p$ , and  $V_{tn} = -V_{tp} = V_t$   
 $r = 1$

$$V_M = V_{DD}/2$$

$$V_{IL} = \frac{1}{8}(3V_{DD} + 2V_t)$$

$$NM_H = NM_L = \frac{1}{8}(3V_{DD} + 2V_t)$$



# Lecture 23: MOSFET - Digital circuits

---

## Topics

- Class notes
- Digital circuits
- Inverter VTC
- CMOS inverter
- Propagation delay
- Transistor sizing
- Power dissipation
- Ref: Sedra Smith 8<sup>th</sup> edition (Chapter 16-17)

## Parameters of VTC for inverter

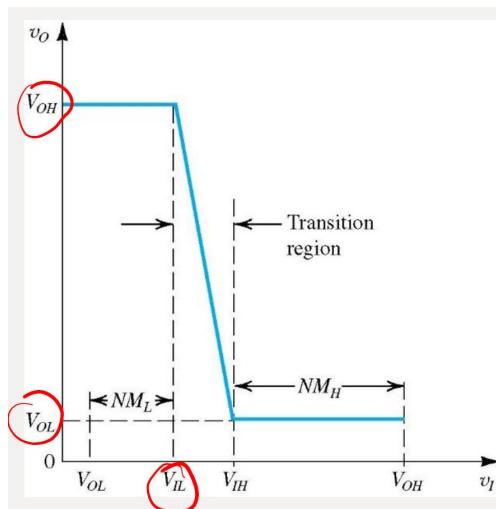
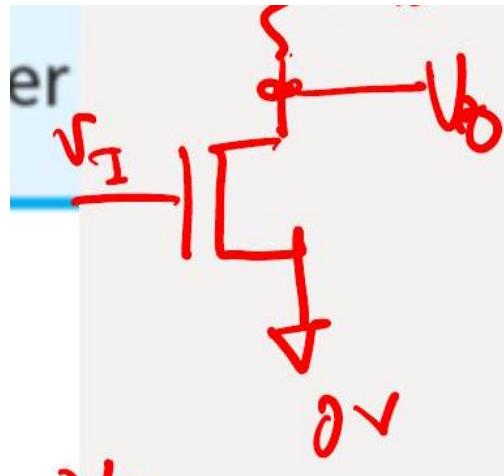


Figure 16.13 Voltage-transfer characteristic of an inverter. The VTC is approximated by three straight-line segments. Note the four parameters of the VTC ( $V_{OH}$ ,  $V_{OL}$ ,  $V_{IL}$ , and  $V_{IH}$ ) and their use in determining the noise margins ( $NM_H$  and  $NM_L$ )

Table 16.1 Important Parameters of the VTC of the Logic Inverter

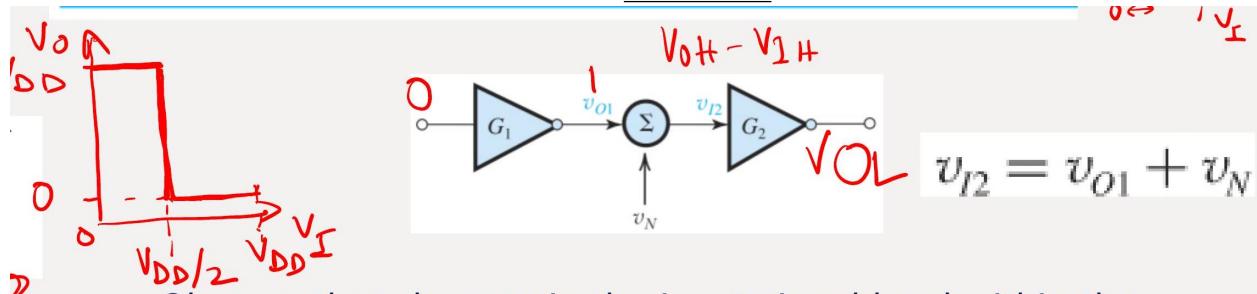
VIL: Maximum value of input interpreted by the inverter as a logic 0



$V_{IH}$  : Minimum value of input interpreted by the inverter as a logic 1

$NM_L$  : Noise margin for low input =  $V_{LL} - V_{OL}$

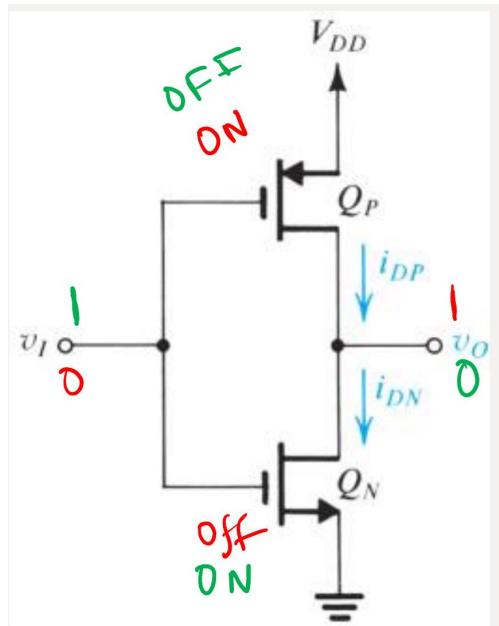
$\overline{NM}_H$  : Noise margin for high input =  $\underline{V_{OH} - V_{IH}}$



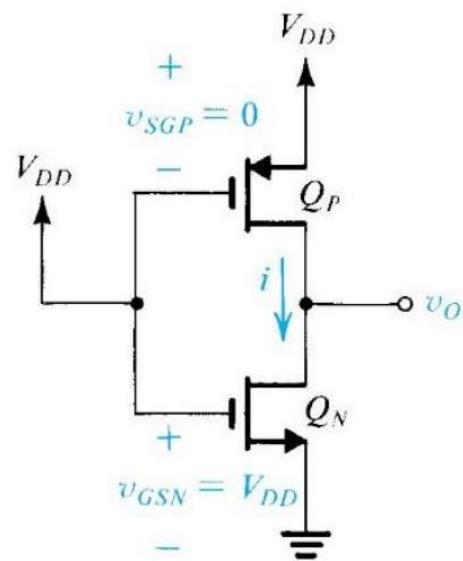
Observe that changes in the input signal level within the noise margins are rejected by the inverter. Thus noise is not allowed to propagate further through the system, a definite advantage of digital over analog circuits.

What should ideal VTC look like?

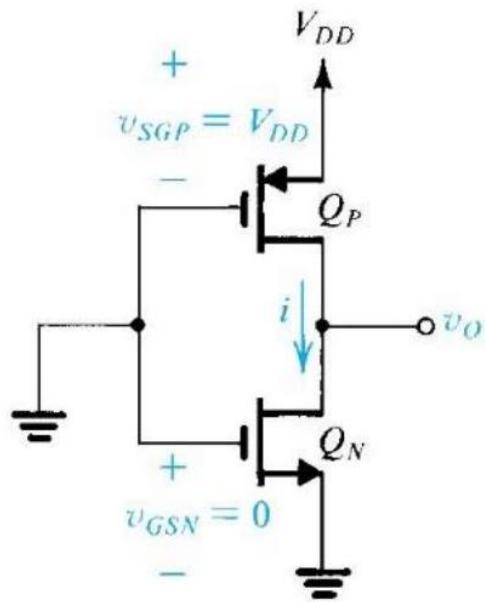
## The CMOS inverter



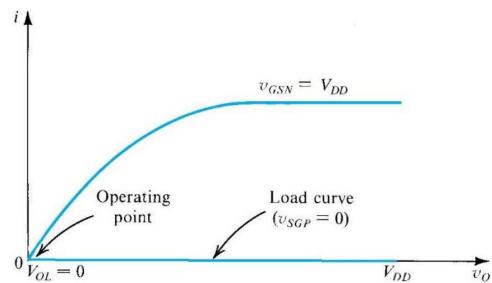
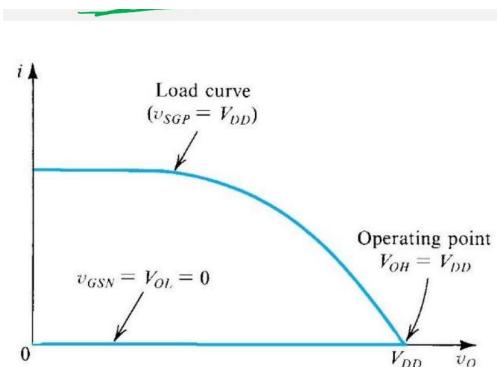
Static power dissipation = 0



(a)

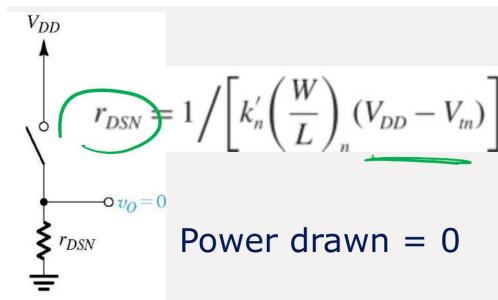


(a)

(b)  
v, high

(b)

$$I_{D_{NMOS}} = \mu_n \operatorname{Cor} \frac{\omega}{L} (V_{CS} \cdot V_{TN})$$



(c)

v, low

Power drawn = 0

### Note

- In spite of the fact that the quiescent current is zero, the load-driving capability of the CMOS inverter is high.
- For instance, with the input high, transistor  $Q_N$  can sink a relatively large load current. This current can quickly discharge the load capacitance.
- Because of its action in sinking load current and thus pulling the output voltage down toward ground, transistor  $Q_N$  is known as the pull-down device.
- Similarly, with the input low, transistor  $Q_p$  can source a relatively large load current. This current can quickly charge up a load capacitance, thus pulling the output voltage up toward  $V_{DD}$ .
- Hence,  $Q_p$  is known as the pull-up device.

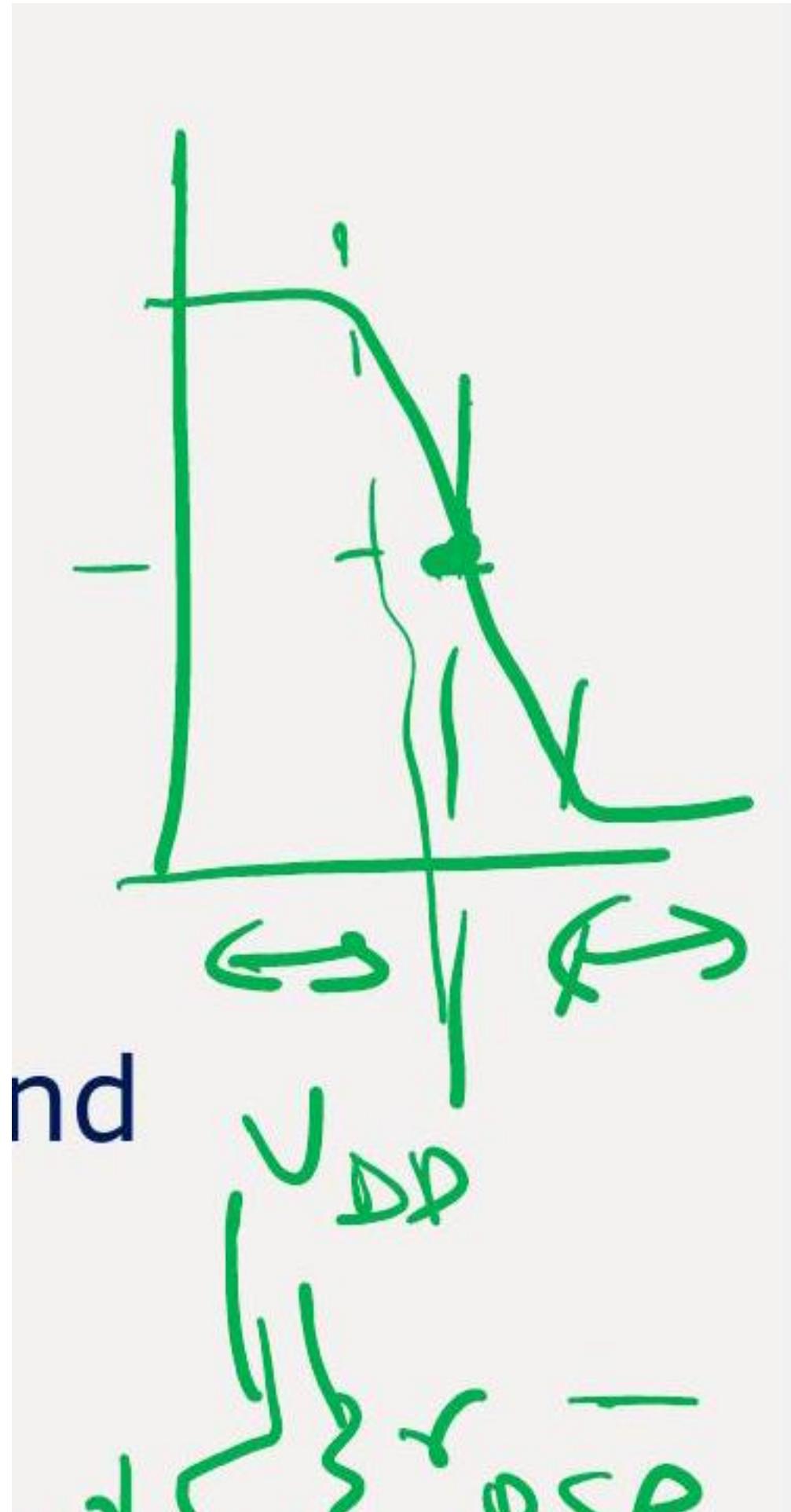
### CMOS logic inverter review

1. The output voltage levels are 0 and  $V_{DD}$ , and thus the signal swing is the maximum possible. This, coupled with the fact that the inverter can be designed to provide a symmetrical voltage-transfer characteristic, results in wide noise margins.

2. The static power dissipation in the inverter is zero (neglecting the dissipation due to leakage currents) in both of its states. This is because no dc path exists between the power supply and ground in either state.

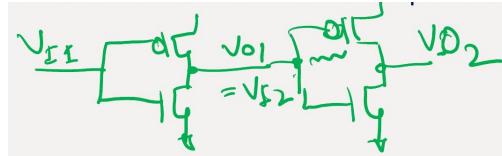
3. A low-resistance path exists between the output terminal and ground (in the low-output state) or  $V_{DD}$  (in the high-output state). These low-resistance paths ensure that the output voltage is 0 or  $V_{DD}$  independent of the exact values of the W/L ratios or other device parameters. Furthermore, the low output resistance makes the inverter less sensitive to the effects of noise and other disturbances.

4. The active pull-up and pull-down devices provide the inverter with high output driving capability in

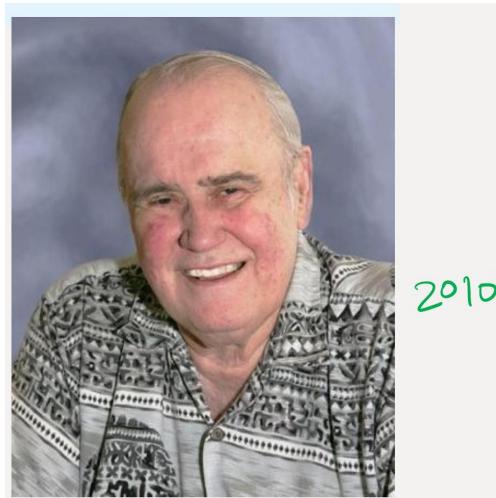


both directions. This speeds up the operation considerably.

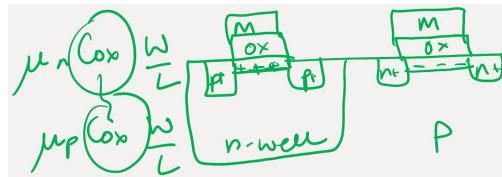
- The input resistance of the inverter is infinite (because  $I_G = 0$ ). Thus the inverter can drive an arbitrarily large number of similar inverters with no loss in signal level. Of course, each additional inverter increases the load capacitance on the driving inverter and slows down the operation. Later we will consider the inverter switching times.



While working for Fairchild Semiconductor in 1963, Frank Wanless filed the first patent on CMOS logic, heralding the new age of zero-static-power logic. In 1964, as director of research and engineering at General Microelectronics (a start-up later bought by Philco-Ford), he created the first commercial CMOS integrated circuit. The symmetry of the logic form Wanless had invented was at first emphasized by the use of the name COnplementary Symmetry MOS, or COS-MOS, but the simpler CMOS shorthand soon prevailed.



$$i_{DN} = k'_n \left( \frac{W}{L} \right)_n \left[ (v_I - V_{tn}) v_O - \frac{1}{2} v_O^2 \right]$$



for  $v_o \leq v_I - V_{tn}$   
and

$$i_{DN} = \frac{1}{2} k'_n \left( \frac{W}{L} \right)_n (v_t - V_{tn})^2 \quad \text{for } v_O \geq v_l - V_{tn}$$

For  $Q_p$

$$i_{DP} = k'_p \left( \frac{W}{L} \right)_p [(V_{DD} - v_l - |V_{tp}|)(V_{DD} - v_O) - \frac{1}{2}(V_{DD} - v_O)^2] \quad \text{for } v_O \geq v_l + |V_{tp}|$$

and

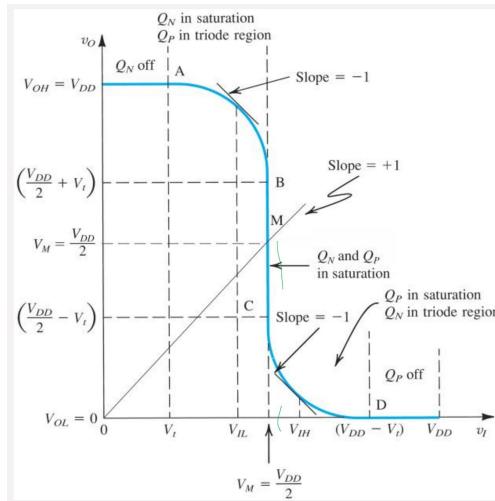
The CMOS inverter is usually designed to have  $V_{tn} = |V_{tp}| = V_t$ . Also, although this is not always the

$$i_{DP} = \frac{1}{2} k'_p \left( \frac{W}{L} \right)_p (V_{DD} - v_I - |V_{tp}|)^2 \quad \text{for } v_O \leq v_I + |V_{tp}|$$

case, we shall assume that  $Q_N$  and  $Q_p$  are matched; that is,  $k'_n(W/L)_n = k'_p(W/L)_p$ . It should be noted that since  $\mu_p$  is often 0.25 to 0.5 times the value of  $\mu_n$  to make  $k'(W/L)$  of the two devices equal, the width of the  $p$ -channel device is made two to four times that of the  $n$ -channel device. More specifically, the two devices are designed to have equal lengths, with widths related by

$$\frac{W_p}{W_n} = \frac{\mu_n}{\mu_p} \quad (16.32)$$

This will result in  $k'_n(W/L)_n = k'_p(W/L)_p$  and the inverter will have a symmetric transfer characteristic and equal current-driving capability in both directions (pull-up and pull-down).



$$V_M = \frac{V_{DD}}{2}$$

$$(v_I - V_t)v_O - \frac{1}{2}v_O^2 = \frac{1}{2}(V_{DD} - v_I - V_t)^2 \quad (16.33)$$

Differentiating both sides relative to  $v$ , results in

$$(v_I - V_t) \frac{dv_O}{dv_I} + v_O - v_0 \frac{dv_O}{dv_I} = -(V_{DD} - v_I - V_t)$$

in which we substitute  $v_I = V_{IH}$  and  $dv_O/dv_I = -1$  to obtain

$$v_o = V_{IH} - \frac{V_{DD}}{2} \quad (16.34)$$

Substituting  $v_I = V_{IH}$  and for  $v_O$  from Eq. (16.34) in Eq. (16.33) gives

$$V_{IH} = \frac{1}{8} (5V_{DD} - 2V_t) \quad (16.35)$$

$V_{IL}$  can be determined in a manner similar to that used to find  $V_{IH}$ . Alternatively, we can use the symmetry relationship

$$V_{IH} - \frac{V_{DD}}{2} = \frac{V_{DD}}{2} - V_{IL}$$

together with  $V_{IH}$  from Eq. (16.35) to obtain

$$V_{IL} = \frac{1}{8} (3V_{DD} + 2V_t) \quad (16.36)$$

The noise margins can now be determined as follows:

$$NM_H = V_{OH} - V_{IH} \quad (16.37)$$

$$\begin{aligned} &= V_{DD} - \frac{1}{8} (5V_{DD} - 2V_t) \\ &= \frac{1}{8} (3V_{DD} + 2V_t) \end{aligned}$$

$$NM_L = V_{IL} - V_{OL} \quad (16.38)$$

$$\begin{aligned} &= \frac{1}{8} (3V_{DD} + 2V_t) - 0 \\ &= \frac{1}{8} (3V_{DD} + 2V_t) \end{aligned}$$

As expected, the symmetry of the voltage-transfer characteristic results in equal noise margins. Of course, if  $Q_N$  and  $Q_p$  are not matched, the voltage-transfer characteristic will no longer be symmetric, and the noise margins will not be equal.

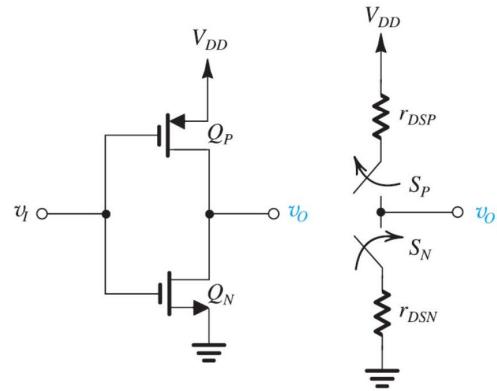
Table 16.2 Summary of Important Static Characteristics of the CMOS Logic Inverter  
Inverter Output Resistance

When  $v_o$  is low (current sinking):

$$r_{DSN} = 1 / \left[ k'_h \left( \frac{W}{L} \right)_n (V_{DD} - V_{tn}) \right]$$

When  $v_o$  is high (current sourcing):

$$r_{DSP} = 1 / \left[ k'_p \left( \frac{W}{L} \right)_p (V_{DD} - |V_{tp}|) \right]$$



## Inverter VTC and Noise Margins

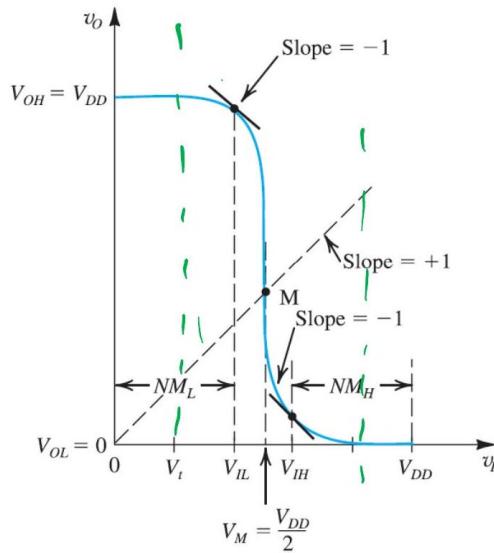
$$V_M = \frac{r(V_{DD} - |V_{tp}|) + V_{tn}}{1+r} \quad \text{where} \quad r = \sqrt{\frac{k'_p(W/L)_p}{k'_n(W/L)_n}}$$

For matched devices, that is,  $\mu_n \left(\frac{W}{L}\right)_n = \mu_p \left(\frac{W}{L}\right)_p$ , and  $V_{tn} = -V_{tp} = V_t$   
 $r = 1$

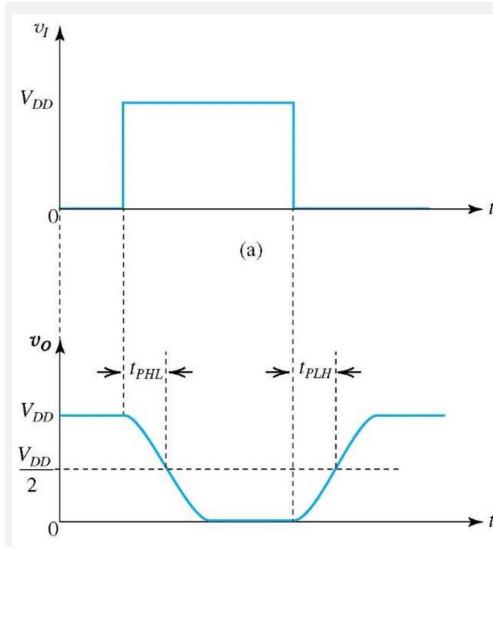
$$V_M = V_{DD}/2$$

$$V_{IL} = \frac{1}{8}(3V_{DD} + 2V_t)$$

$$NM_H = NM_L = \frac{1}{8}(3V_{DD} + 2V_t)$$



## Propagation Delay



(b)

- The propagation delay is the time the inverter takes to respond to a change at its input.
- The inverter propagation delay  $t_p$  is defined as the average of the two

$$t_p \equiv \frac{1}{2} (t_{PLH} + t_{PHL})$$

- It is simply a result of the time needed to charge and discharge the various capacitances in the circuit. These include the MOSFET capacitances, the wiring capacitance, and the input capacitances of all the logic gates driven by the inverter.

Figure 17.1 An inverter fed with the ideal pulse in (a) provides at its output the pulse in (b). Two delay times are defined as indicated.

## Transistor sizing

- To minimize area, the length of all channels is usually made equal to the minimum length permitted by the given technology.
- In a given inverter, if our interest is strictly to minimize area,  $(W/L)n$  can be selected in the range from 1 to 1.5, which minimizes driving current. However, in more recent technologies it has become common to use slightly higher values, closer to 2 and above.

3. The selection of  $(W/L)_n$  relative to  $(W/L)_p$  affects the noise margins and  $t_{PLH}$ . Both are optimized by matching  $Q_p$  and  $Q_N$ . Historically, the value of  $k'_n$  has significantly exceeded that of  $k'_p$ , often by a factor of 3 to 5. In those technologies, matching  $Q_p$  and  $Q_N$  is often wasteful of area and has the potential to increase the effective capacitance  $C$ , and although  $t_{PHL}$  and  $t_{PLH}$  are matched, delay may in fact end up higher than in the case without matching. To avert these problems, selecting  $(W/L)_p = (W/L)_n$  is a possibility, and  $(W/L)_p = 2(W/L)_n$  is a frequent compromise. In emerging technologies (e.g., the 28 – nm process),  $k'_n$  and  $k'_p$  are closer in value, and it has become more common to use  $(W/L)_p = (W/L)_n$ .
4. Having settled on an appropriate ratio of  $(W/L)_p$  to  $(W/L)_{n'}$  we still have to select  $(W/L)_n$  to reduce  $t_p$  and thus allow higher speeds of operation. Any increase in  $(W/L)_n$  and proportionally in  $(W/L)_p$  will of course increase area, and hence the inverter contribution to the value of the equivalent capacitance  $C$ . To be more precise we express  $C$  as the sum of an intrinsic component  $C_{int}$  contributed by  $Q_N$  and  $Q_p$  of the inverter, and an extrinsic component  $C_{ext}$  resulting from the wiring and the input capacitance of the driven gates,

$$C = C_{int} + C_{ext} \quad (17.21)$$

Increasing  $(W/L)_n$  and  $(W/L)_p$  of the inverter by a factor  $S$  relative to that of a minimum-size inverter for which  $C_{int} = C_{int0}$  results in

$$C = SC_{int0} + C_{ext} \quad (17.22)$$

Now, if we use the equivalent-resistances approach to compute  $t_p$  and define an equivalent inverter resistance  $R_{eq}$  as

$$R_{eq} = \frac{1}{2}(R_N + R_P) \quad (17.23)$$

then,

$$t_P = 0.69R_{eq}C \quad (17.24)$$

Further, if for the minimum-size inverter  $R_{eq}$  is  $R_{eq0}$  increasing  $(W/L)_n$  and  $(W/L)_p$  by the factor  $S$  reduces  $R_{eq}$  by the same factor:

$$R_{eq} = R_{eq0}/S \quad (17.25)$$

## Transistor sizing

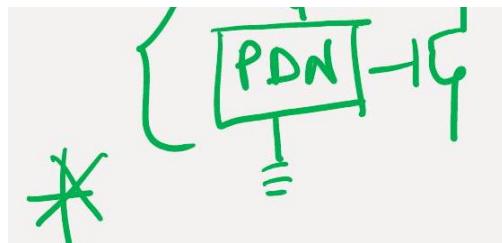
Combining Eqs. (17.24), (17.25), and (17.22), we obtain

$$\begin{aligned} t_P &= 0.69 \left( \frac{R_{eq0}}{S} \right) (SC_{int0} + C_{ext}) \\ t_P &= 0.69 \left( R_{eq0}C_{int0} + \frac{1}{S}R_{eq0}C_{ext} \right) \end{aligned} \quad (17.26)$$

We can see, then, that scaling the  $W/L$  ratios does not change the component of  $t_p$  caused by the capacitances of  $Q_N$  and  $Q_p$ . It does, however, reduce the component of  $t_p$  that results from capacitances external to the inverter itself. It follows that we can use Eq. (17.26) to decide on a suitable scaling factor  $S$  that keeps  $t_p$  below a specified maximum value, keeping in mind of course the effect of increasing  $S$  on silicon area.

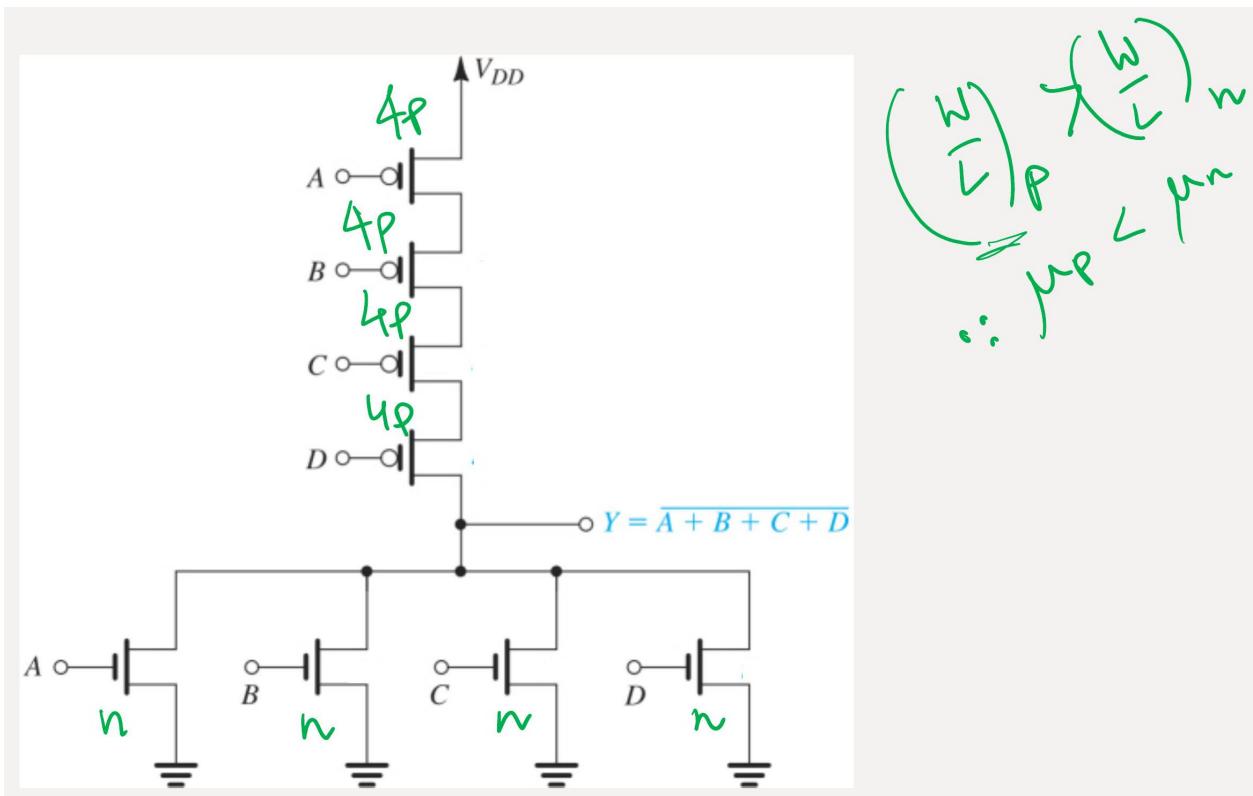
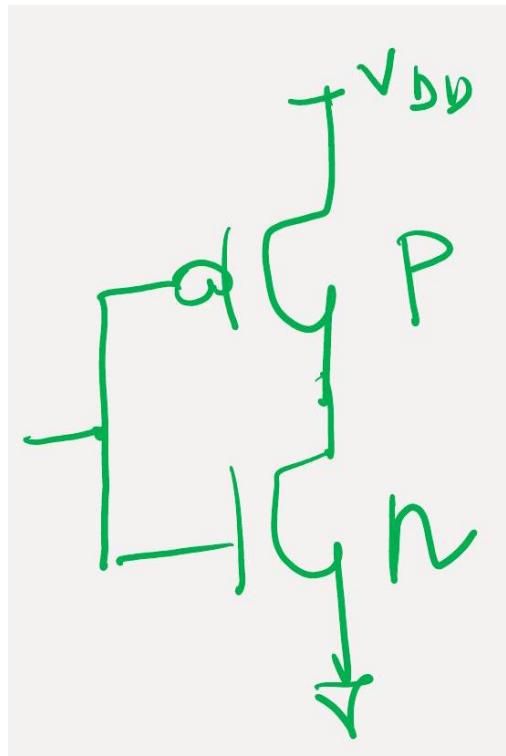
## Transistor Sizing in CMOS Logic Gates

- After designing circuit, determine ( $W/L$ ) for all transistors
- $W/L$  ratios are selected to provide the gate with current-driving capability in both directions equal to that of the basic inverter.

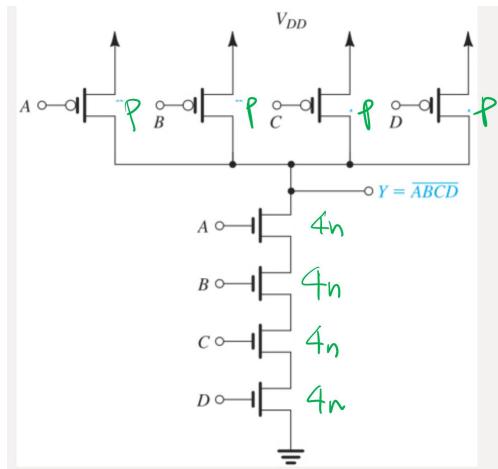


- For the basic inverter design, denote  $(W/L)_n = n$  and  $(W/L)_p = p$ , where  $n$  is usually 1 to 3, depending on technology (using larger numbers for newer technologies), and, for a matched design,  $p = (\mu_n / \mu_p) n$
- It should be noted, however, that often  $p = 2n$  and for minimum area  $p = n$ .
- Select individual  $W/L$  ratios for all transistors in a logic gate so that the PDN should be able to provide a capacitor discharge current at least equal to that of an NMOS transistor with  $W/L = n$ , and the PUN should be able to provide a charging current at least equal to that of a PMOS transistor with  $W/L = p$ . This will guarantee a worst-case gate delay equal to that of the basic inverter

## Transistor sizing for 4-input NOR gate

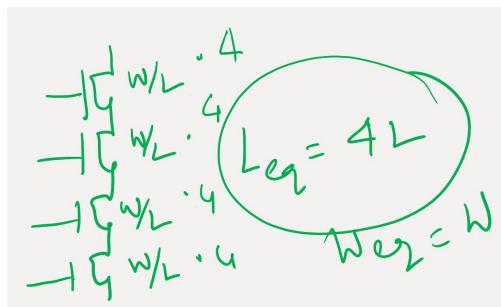


## Transistor sizing for 4-input NAND gate



Which requires higher area? NAND or NOR?

- Connecting MOS transistors in series is equivalent to adding the lengths of their channels while the width does not change; connecting MOS transistors in parallel does not change the channel length but increases the width to the sum of the W's

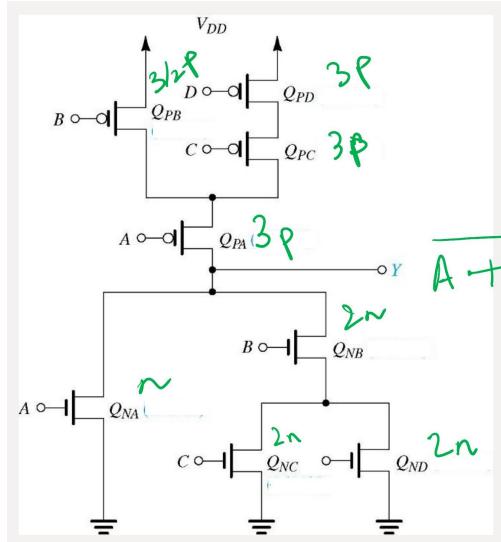


Provide transistor W/L ratios for the logic circuit shown below. Assume that for the basic inverter  $n = 2$  and  $p = 4$  and that the channel length is 65 nm. Round all values of W to the nearest 10 nm.

channel length is 65 nm. Round all values of W to the nearest 10 nm.

$$n=2 = \frac{2 \times 65}{65} = \frac{130}{65}$$

$$P = \frac{4 \times 65}{65} = \frac{260}{65}$$



## Power dissipation

- NMOS inverter: dissipates no power when  $v_I$  is low and the switch is open. In the other state, the power dissipation is approximately  $V_{DD^2}/R$  and can be substantial. This power dissipation occurs even if the inverter is not switching and is thus known as static power dissipation.
- CMOS inverter exhibits no static power dissipation, a definite advantage.
- Unfortunately, however, another component of power dissipation arises when a capacitance exists between the output node of the inverter and ground.
- Now, as the inverter is switched from one state to another, current must flow through the switch(es) to charge (and discharge) the load capacitance. These currents give rise to power dissipation in the switches, called dynamic power dissipation.

## Power dissipation

$$P_{\text{dyn}} = fCV_{DD}^2$$

- To minimize the dynamic power dissipation, reduce the value of  $C$ .
- However,  $C$  is largely determined by the transistors of the inverter itself and cannot be substantially reduced.
- Reducing  $V_{DD}$  reduces  $P_{\text{dyn}}$  significantly.
- This has been a major motivating factor behind the reduction

V<sub>DD</sub> is proportional to  
dependent threshold  
subthreshold switch

of V<sub>DD</sub> with every technology generation

- Thus, while the  $0.5 - \mu\text{m}$  CMOS process utilized a  $5 - \text{V}$  power supply, the power-supply voltage used with the  $0.13 - \mu\text{m}$  process is only  $1.2 \text{ V}$ , and that for the  $28 - \text{nm}$  process is even lower at  $0.9 \text{ V}$ .
- We may be tempted to reduce  $P_{dyn}$  by reducing  $f$ .
- However, we don't want to reduce speed!
- The dynamic power dissipation of high-density chips can be over 100 W.

# Lecture 24: MOSFET scaling

---

## Top

- MOSFET scaling
- Why scale MOSFETs?
- Short channel effects
- DIBL
- Mobility degradation
- Velocity saturation
- Gate leakage
- GIDL
- Evolution of MOSFET design

## Course evaluation!

- Your investment in completing this evaluation thoughtfully for about 10-15 minutes today will not only help instructors improve their courses, but it will help other students, too.
- Most of the ratings and responses you provide in your evaluations this semester will be shared with future students.
- The more students that contribute evaluations, the more access every student will have

to course evaluation data when making future course enrollment decisions.

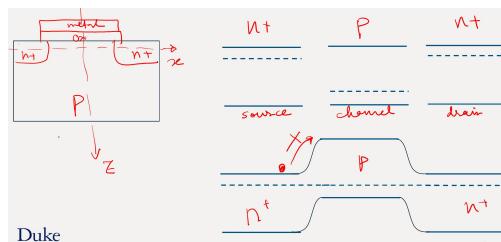
## Summary of key performance metrics

	SS	$I_{ON}$	$g_m$	$\mu$	$V_{DD}$	$I_{off}$
Ideally	$\downarrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\downarrow$	$\downarrow$

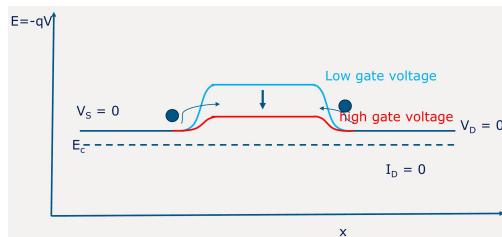
## Scaling of MOSFETs

- Dennard scaling (1974) was based on reducing the key dimensions while maintaining essentially equivalent electric field within the device
- Why do we want to make transistors smaller?
- Moore's law: double number of transistors on IC every 18 months
- Increased computing performance (has driven the technological revolution)
- Scaling of dimensions, power supply voltage, capacitance
- Implies reduction in power consumption, speed, cost!

Energy band view of MOSFET



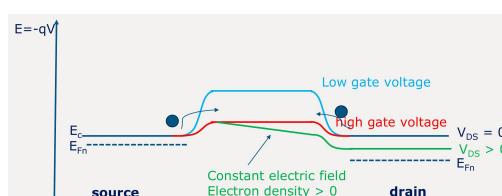
## Energy band view of MOSFET



## Energy band view of MOSFET

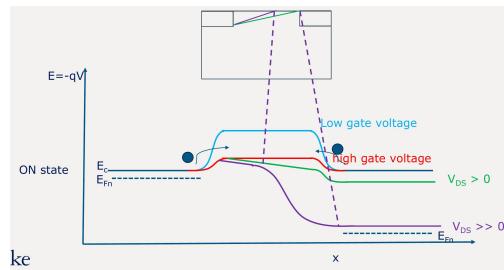
What if we apply a small voltage to the drain?

- The Fermi level in the drain is lowered
- The conduction band is lowered too, but electron density stays the same



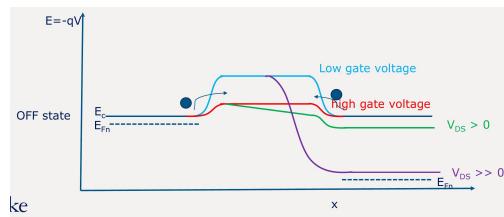
## Energy band view of MOSFET

Now increase the drain voltage



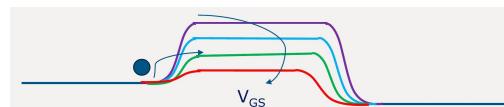
## Energy band view of MOSFET

Now remove the gate voltage



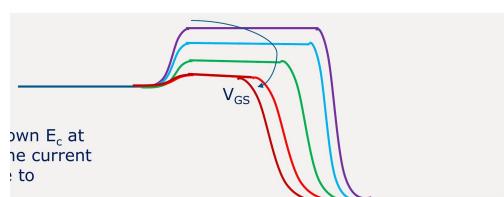
## Energy band view of MOSFET

Why does the current saturate?

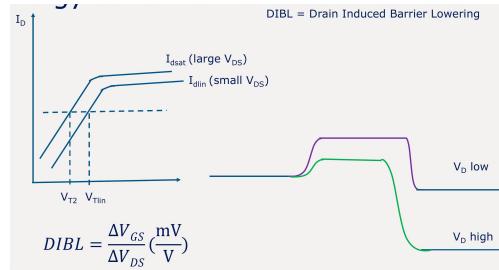


Small  $V_{DS}$

Current saturates because bringing down  $E_c$  at drain end does not change current. The current is limited by the injection from source to channel.



## Energy band view of MOSFET



## Short channel effect: mobility degradation

- An interesting consequence of the universal mobility dependence on effective vertical field is that mobility degradation generally gets worse as the vertical dimensions of the device scale down. This is because, all things being equal, for a thinner oxide, the effective vertical field increases. Since lateral scaling and vertical scaling have to proceed in a harmonious way, MOSFET scaling brings along a degradation in inversion layer mobility. Hence, mobility degradation is a short channel effect.

## Universal mobility curves

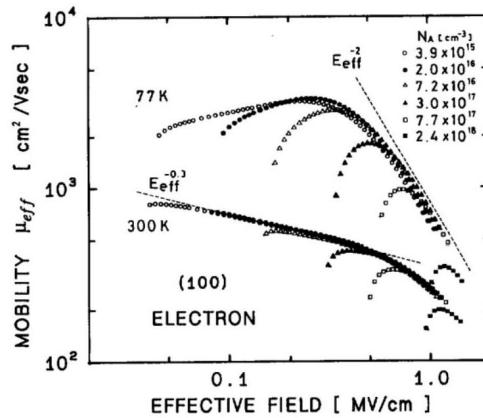


Fig. 1. Electron mobility in inversion layer at 300 K and 77 K versus effective field  $E_{\text{eff}}$ , as a parameter of substrate acceptor concentration,  
Here,  $E_{\text{eff}}$  is defined by  $E_{\text{eff}} = q \cdot (N_{\text{dpl}} + \eta \cdot N_s) / \varepsilon_{\text{si}}$  with  $\eta$  of 1/2.

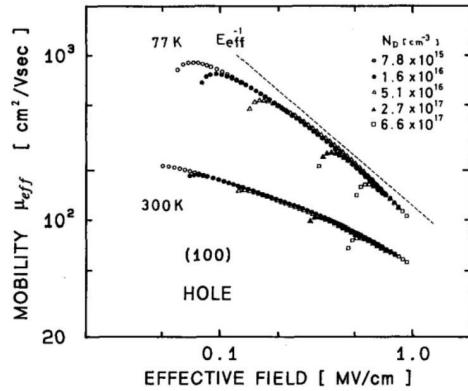
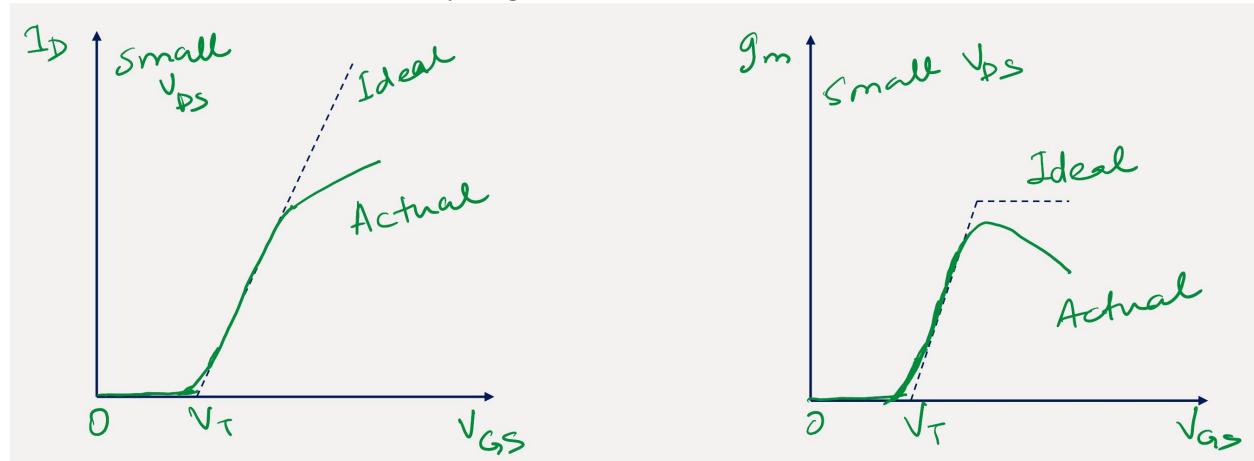


Fig. 2. Hole mobility in inversion layer at 300 K and 77 K versus effective field,  $E_{eff}$  as a parameter of substrate donor concentration,  $N_D$ . Here,  $E_{eff}$  is defined by  $E_{eff} = q \cdot (N_{dpp} + \eta \cdot N_s) / \varepsilon_{si}$  with  $\eta$  of 1/3.  $\varepsilon_{eff} = \frac{Q_{dmax} + \eta Q_{inv}}{\varepsilon_s}$   $\eta$  varies depending on surface etc.

S. Takagi et al., IEEE TED, vol. 41, 1994

Short channel effect: mobility degradation



NOTE: Mobility degradation is due to vertical electric field

### Short channel effects: velocity saturation

- Mobility degradation due to high vertical electric fields
- Velocity saturation due to high longitudinal electric fields in the channel

Electric field in short channels:

$$\varepsilon = \frac{1 \text{ V}}{100 \text{ nm}} = 10^5 \text{ V/cm}$$

## $V_t$ roll-off

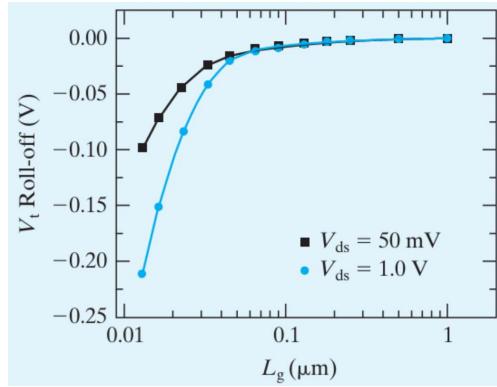


FIGURE 7 – 4  $|V_t|$  decreases at very small  $L_g$ . This phenomenon is called  $V_t$  roll-off. It determines the minimum acceptable  $L_g$  because  $I_{off}$  is too large when  $V_t$  becomes too low or too sensitive to  $L_g$ .

## Subthreshold swing dependence on gate length and $V_{DS}$

$$S = \left(1 + \frac{C_{dep}}{C_{ox}}\right) \frac{kT}{q} \ln 10$$

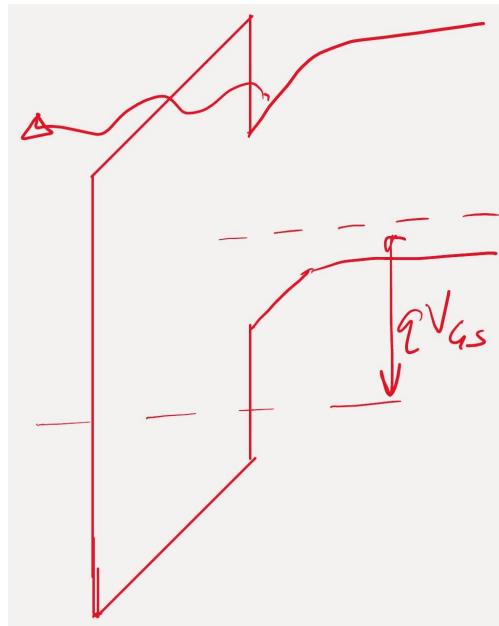
- Competition between the gate and the drain for electrostatic control of the energy barrier present in the channel below threshold
- The energy barrier sets the subthreshold current and therefore the threshold voltage and the subthreshold swing
- Lateral field from source and drain raises the potential along the channel wrt long channel MOSFET
- This weakens the control that the gate exerts over the channel
- As a result negative shift in V, slower change in the interface electron charge with  $V_{GS}$
- Subthreshold swing is a measure of this
- From quantum mechanics, the probability that an electron penetrates through a forbidden barrier is exponentially dependent on the inverse of the barrier thickness and the square root of its height.

$$J_g = A \varepsilon_{ox}^2 \exp\left(-\frac{B}{\varepsilon_{ox}}\right)$$

- A rule of thumb for device designers: tolerable limit for gate leakage is a current density of  $1 \text{ A/cm}^2$  at  $1 \text{ V}$  at room temperature.

- For SiO<sub>2</sub> - this level of current is reached for SiO<sub>2</sub> thickness of 1.7-1.8 nm (attained at mid-late 2000s)
- Move to high-K dielectrics

$$EOT_{high-K} = \frac{\epsilon_{SiO2}}{\epsilon_{high-K}} x_{high-K} = \frac{\kappa_{SiO2}}{\kappa_{high-K}} x_{high-K}$$



### Gate induced drain leakage (GIDL)

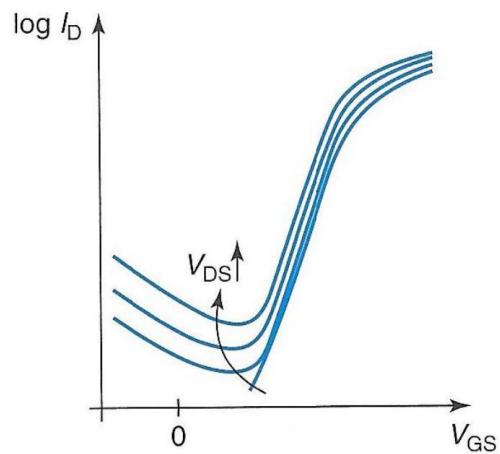
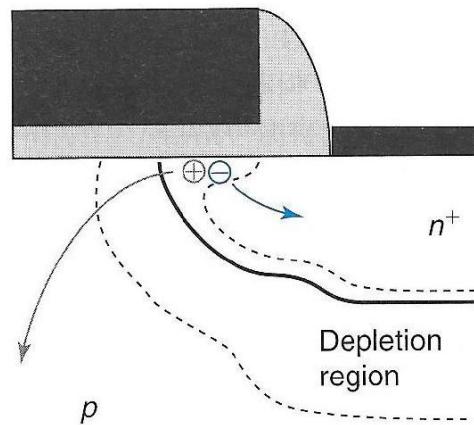
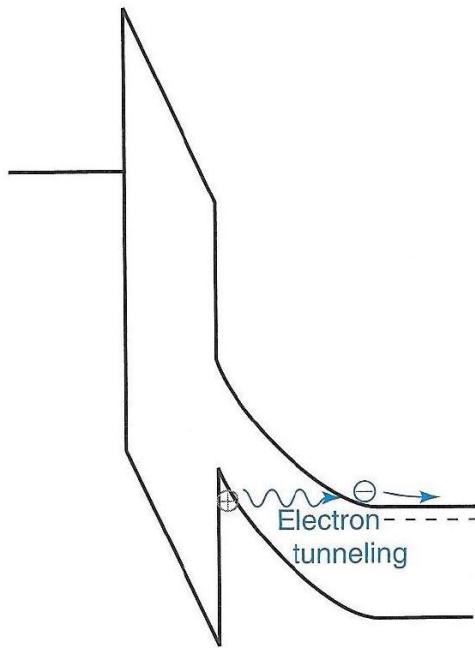


FIGURE 10.32

Sketch of subthreshold characteristics of a short MOSFET showing the characteristic signature of GIDL. This is an increase in the drain current for values of  $V_{GS}$  much below threshold.



(a)



(b)

FIGURE 10.33

Illustration of gate-induced drain leakage (GIDL) at the gate-drain overlap region of a MOSFET biased in the subthreshold regime with a large drain-to-gate voltage. (a) In the overlap region, there is a depletion region at the semiconductor surface with a large vertical electric field. Band-to-band tunneling of electrons from the substrate to the drain contributes to drain current. The generated holes flow to the substrate. (b) Energy band diagram across the vertical dimension in the gate-drain overlap region showing band-to-band tunneling of electrons.

MOSFET scaling

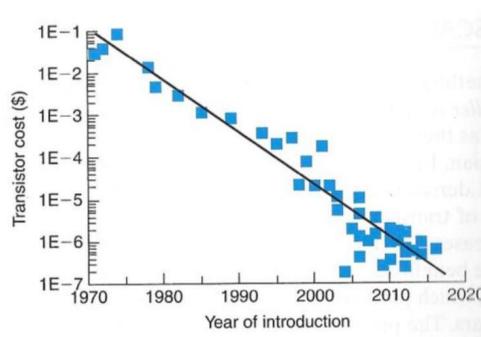
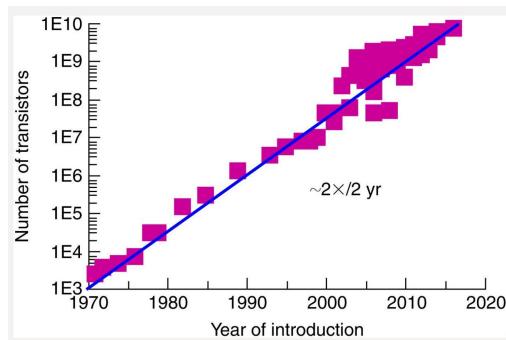
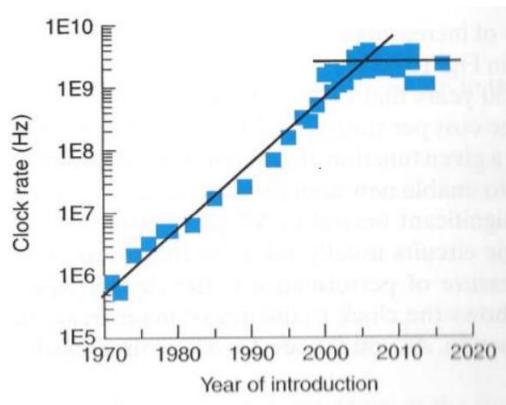
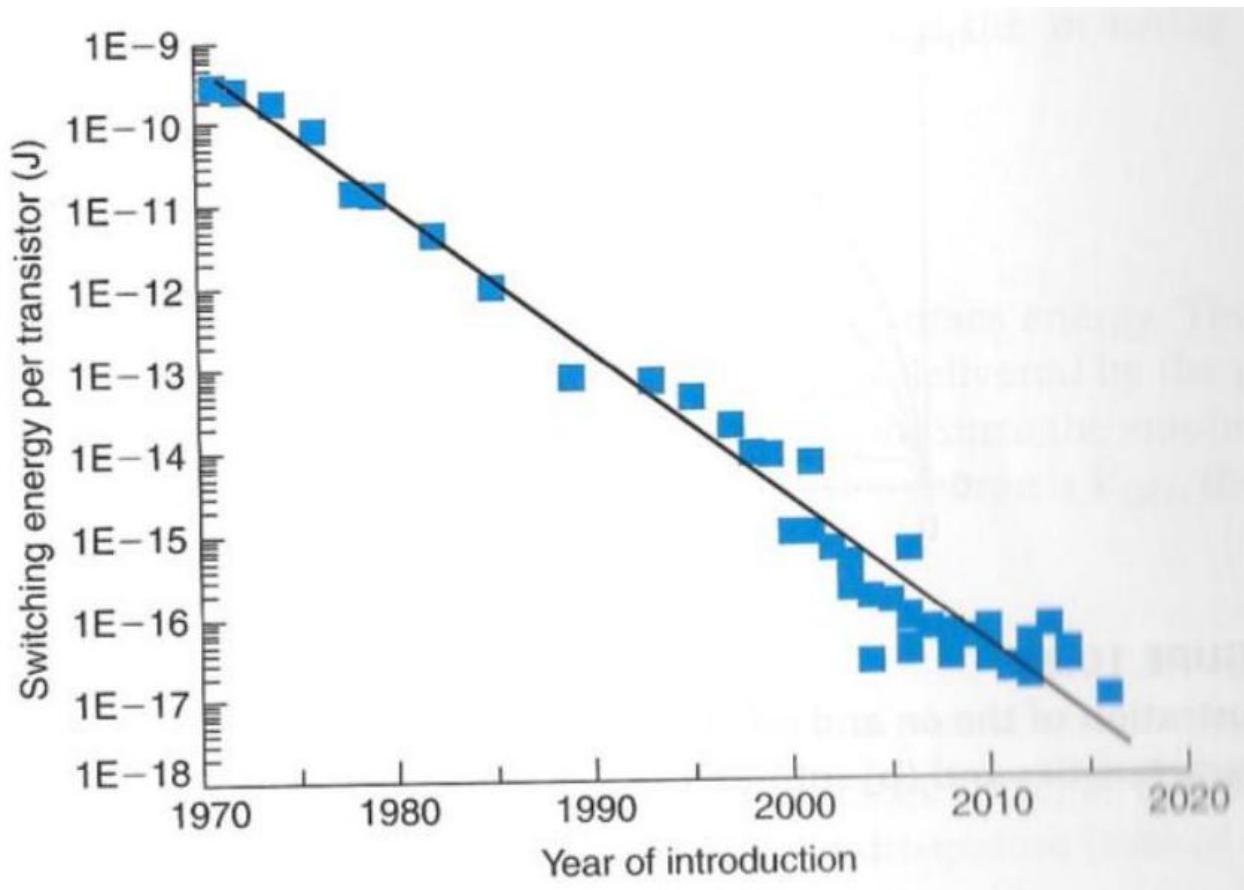


FIGURE 10.36

Average transistor cost in mainstream microprocessors as a function of year of introduction.

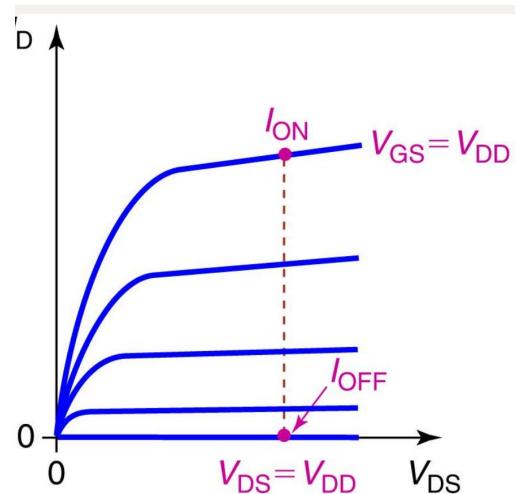


Evolution of clock frequency in mainstream microprocessors as a

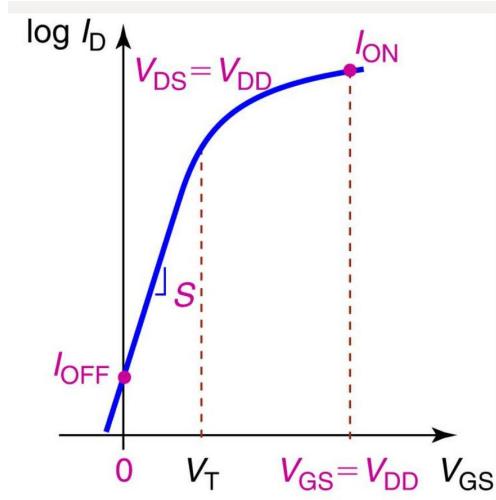


unction of year of introduction.

### The MOSFET as a switch



(a)



(b)

$$I_{ON} = I_D (V_{GS} = V_{DD}, V_{DS} = V_{DD})$$

$$I_{OFF} = I_D (V_{GS} = 0, V_{DS} = V_{DD})$$

$$\tau_d \cong \frac{C_G V_{DD}}{I_{ON}}$$

Time taken to switch  
between ON and OFF  
states - create

states - create

inversion charge

$$E_d \cong C_G V_{DD}^2$$

Switching from off state to  
on state and back - power-  
delay product

$$P_d = f E_d = f C_G V_{DD}^2$$

$$= \frac{E_d}{\tau_d} \cong I_{ON} V_{DD}$$

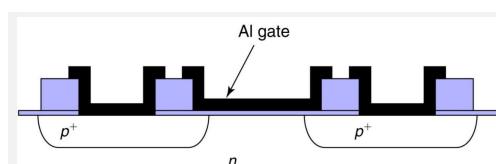
Dynamic power

$$P_S = I_{OFF} V_{DD}$$

Static power

## Evolution of MOSFET design

### Early PMOS used in the first MOSFET ICs



Circa ~ mid 1960 s

$L \sim 20\mu\text{m}$

$x_{\text{ox}} \sim 100 \text{ nm}$

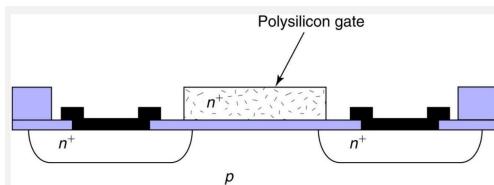
$$x_i \sim 3\mu\text{m}$$

$$V_{DD} = 15 \text{ V}$$

$$V_T = 2 \text{ V}$$

Why p-type MOSFETs only?

## First commercial n-channel MOSFET



Circa early 1970 s

$$L \sim 5\mu\text{m}$$

$$x_{0x} \sim 60 \text{ nm}$$

$$x_j \sim 2\mu\text{m}$$

$$V_{DD} = 12 \text{ V}$$

- Polysilicon gate!

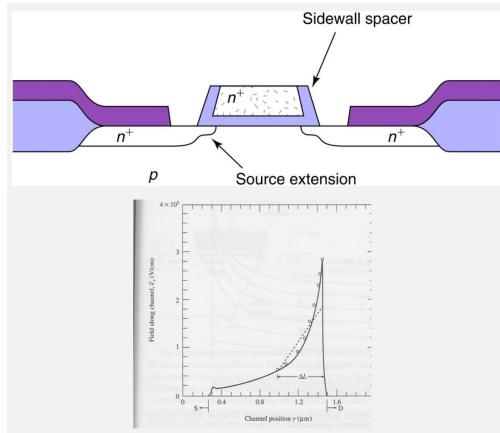
$$V_T = 1 \text{ V}$$

- Al cannot withstand high temperature needed for drain/source dopant diffusion
- Gate definition comes later in the process
- Large overlap between gate and source/drain
- Large parasitic capacitances
- Polysilicon gate - self-aligned process

## Early 1960s - CMOS

- Early logic circuits used two types of FETs of the same polarity (both  $n$  or both  $p$ )
- Threshold voltages were different
- One enhancement mode with normally OFF channel
- One depletion mode with normally ON channel
- CMOS invented by Frank Wanlass at Fairchild
- Absence of standby power consumption while still enabling fast switching

## Sidewall spacers, S/D extensions



Circa ~ early 1960 s

$L \sim 1\mu\text{m}$

$x_{\text{ox}} \sim 40 \text{ nm}$

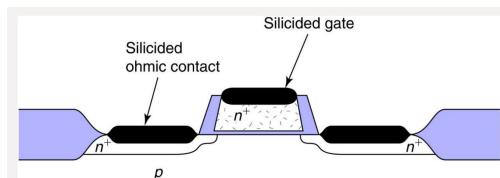
$x_i \sim 0.3 \text{ nm}$

$V_{DD} = 5 \text{ V}$

$V_T \sim 1 \text{ V}$

- Dielectric spacers are placed on both sides of the polysilicon gate
- Shallow lightly-doped source and drain regions that are self-aligned to the polysilicon gate
- Mitigates high electric field at drain end
- Also known as lightly doped drain (LDD)
- Reduces short channel effects

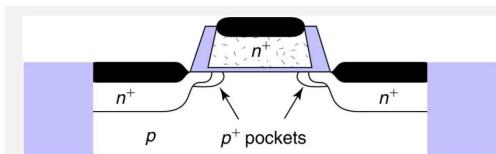
## Self-aligned silicide - salicide



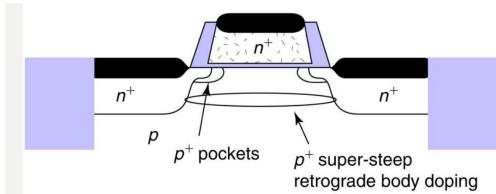
$p$

- In the previous case, source and drain metal contacts are not self aligned
- This means source and drain regions have to be made large enough to accommodate misalignments of the positioning of the metal contacts

- This limits the ability to scale the footprint of the device and results in high source and drain resistances
- Silicides are metallic compounds that only form when certain metals are deposited on Si and annealed; e.g. NiSi, PtSi, TiSi, WSi etc
- Deposit metal such as Ti or Ni on previous device structure without the metal layers.
- Metal only comes in contact with Si at the exposed S/D and gate regions
- Silicide only forms where there is silicon underneath. Metal is unreacted in other regions
- Now selectively etch metal away



(a)



(b)

- Scaling theory suggests short-channel effects can be improved

Circa early 1990 s

$$L \sim 0.18\mu\text{m}$$

$$x_{0x} \sim 3.2 \text{ nm}$$

$$x_j \sim 100 \text{ nm}$$

$$V_{DD} = 2.5 \text{ V}$$

$$V_T \sim 0.6 \text{ V}$$

Circa late 1990s

$$L \sim 0.13\mu\text{m}$$

$$x_{0x} \sim 2 \text{ nm}$$

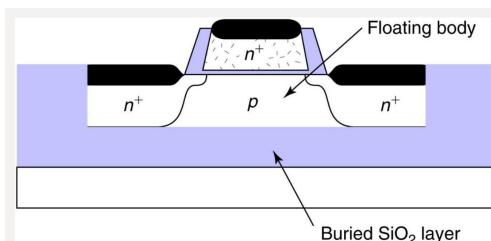
$$x_j \sim 60 \text{ nm}$$

$$V_{DD} = 1.8 \text{ V}$$

$$V_T \sim 0.5 \text{ V}$$

- But mobility will degrade due to ionized impurity scattering

- Introduction of  $p+$  pockets (halo doping):
- High doping at source and drain ends of the channel helps to screen the electric field lines emanating from drain at high VDS.
- Can be fabricated in a self-aligned way by using angled ion implantation of the p-type species once the gate is formed
- But reverse short channel effect!
- Super-steep retrograde (SSR) body doping
- The idea is to form a relatively highly doped buried layer under the gate while preserving a low doping level at the surface where the channel is induced
- Buried layer formed by blanket ion-implantation prior to gate formation
- $P+$  doping layer compensated by S/D doping
- By introducing these 2 techniques, MOSFET successfully scaled to sub-0.1 um gate length regime without short channel effects



Buried SiO<sub>2</sub> layer

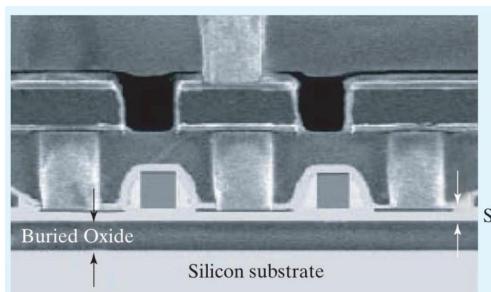
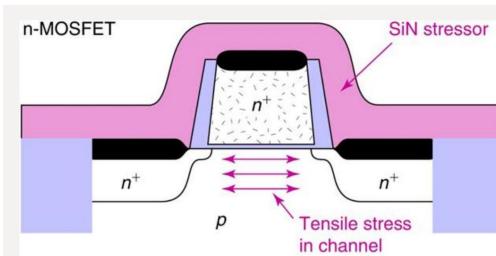


FIGURE 7-17 The cross-sectional electron micrograph of an SOI integrated circuit. The lower level structures are transistors and contacts. The upper two levels are the vias and the interconnects, which employ multiple layers of materials to achieve better reliability and etch stops.

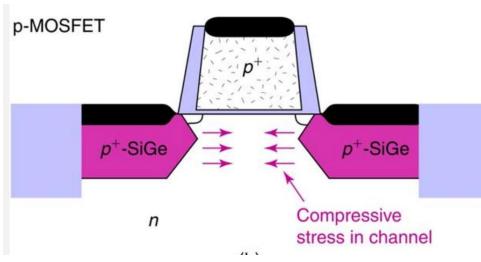
- Because of buried oxide, MOSFET on SOI enjoys reduced junction capacitance
- Means improved switching speed

- Floating body
- Under normal operating conditions, the body gets forward biased due to holes produced through impact ionization
- The holes cannot escape and pile up
- This reduces  $V_{th}$  and enhances gate overdrive, leading to more drain current
- Bipolar effect
- Disadvantage:
- $V_{th}$  is unpredictable due to floating body

## Strained silicon



(a)



(b)

Circa early 2000s

$$L \sim 50 \text{ nm}$$

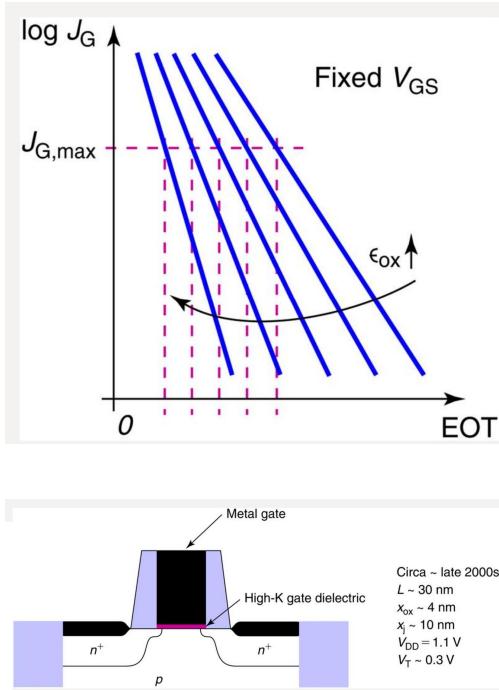
$$x_{\text{ox}} \sim 1.2 \text{ nm}$$

$$x_j \sim 20 \text{ nm}$$

$$V_{DD} = 1.4 \text{ V}$$

$$V_T \sim 0.4 \text{ V}$$

## High-k/metal gate



- Re-engineer the "holiest" element of a traditional MOSFET - the Si/SiO<sub>2</sub> interface!!
- Also need new metals
- Got rid of polysilicon depletion
- Two metals needed to be introduced for  $V_t$  adjustments - for PMOS and NMOS

## FinFET to Gate-all-around

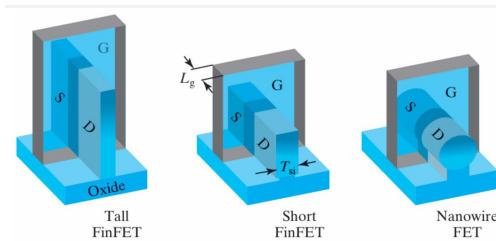


FIGURE 7-19 Variations of FinFET. Tall FinFET has the advantage of providing a large  $W$  and therefore large  $I_{on}$  while occupying a small footprint. Short FinFET has the advantage of less challenging lithography and etching. Nanowire FET gives the gate even more control over the transistor body by surrounding it. FinFETs can also be fabricated on bulk Si substrates.

## Introducing new channel materials

- Ge for p-channel
- InGaAs, InAs, InSb for n channel
- Dielectric?

# Lecture 25: Evolution of MOSFET design

---

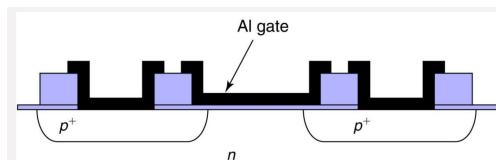
## Top

- MOSFET scaling
- Why scale MOSFETs?
- Short channel effects
- DIBL
- Mobility degradation
- Velocity saturation
- Gate leakage
- GIDL
- Evolution of MOSFET design

## Course evaluation!

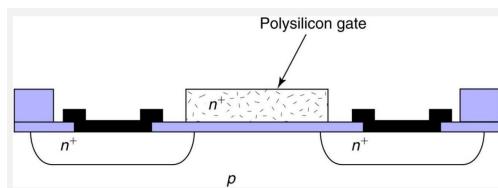
- Your investment in completing this evaluation thoughtfully for about 10-15 minutes today will not only help instructors improve their courses, but it will help other students, too.
- Most of the ratings and responses you provide in your evaluations this semester will be shared with future students.
- The more students that contribute evaluations, the more access every student will have to course evaluation data when making future course enrollment decisions.

## Early PMOS used in the first MOSFET ICs



Circa ~ mid 1960 s  
 $L \sim 20\mu\text{m}$   
 $x_{\text{ox}} \sim 100 \text{ nm}$   
 $x_i \sim 3\mu\text{m}$   
 $V_{\text{DD}} = 15 \text{ V}$   
 $V_T = 2 \text{ V}$   
 Why p-type MOSFETs only?

## First commercial n-channel MOSFET



Circa early 1970 s  
 $L \sim 5\mu\text{m}$   
 $x_{0x} \sim 60 \text{ nm}$   
 $x_j \sim 2\mu\text{m}$   
 $V_{\text{DD}} = 12 \text{ V}$

- Polysilicon gate!

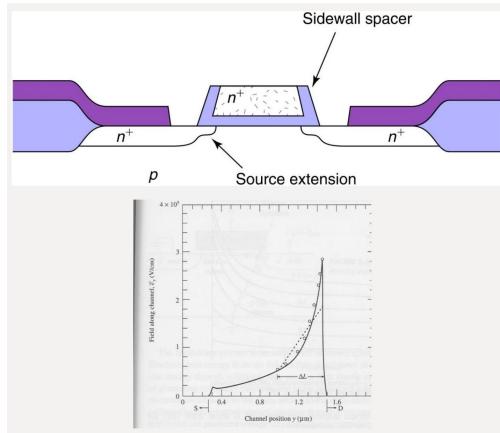
$V_T = 1 \text{ V}$

- Al cannot withstand high temperature needed for drain/source dopant diffusion
- Gate definition comes later in the process
- Large overlap between gate and source/drain
- Large parasitic capacitances
- Polysilicon gate - self-aligned process

## Early 1960s - CMOS

- Early logic circuits used two types of FETs of the same polarity (both  $n$  or both  $p$ )
- Threshold voltages were different
- One enhancement mode with normally OFF channel
- One depletion mode with normally ON channel
- CMOS invented by Frank Wanlass at Fairchild
- Absence of standby power consumption while still enabling fast switching

## Sidewall spacers, S/D extensions



Circa ~ early 1960 s

$L \sim 1\mu\text{m}$

$x_{\text{ox}} \sim 40 \text{ nm}$

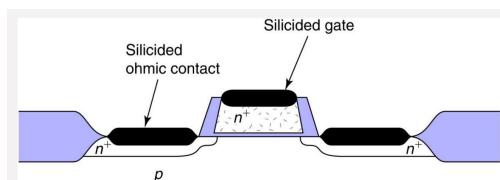
$x_i \sim 0.3 \text{ nm}$

$V_{DD} = 5 \text{ V}$

$V_T \sim 1 \text{ V}$

- Dielectric spacers are placed on both sides of the polysilicon gate
- Shallow lightly-doped source and drain regions that are self-aligned to the polysilicon gate
- Mitigates high electric field at drain end
- Also known as lightly doped drain (LDD)
- Reduces short channel effects

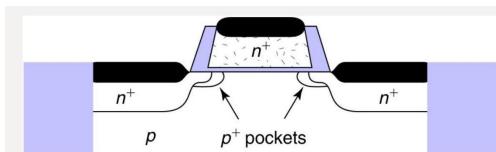
## Self-aligned silicide - salicide



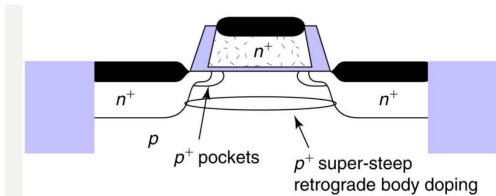
$p$

- In the previous case, source and drain metal contacts are not self aligned
- This means source and drain regions have to be made large enough to accommodate misalignments of the positioning of the metal contacts

- This limits the ability to scale the footprint of the device and results in high source and drain resistances
- Silicides are metallic compounds that only form when certain metals are deposited on Si and annealed; e.g. NiSi, PtSi, TiSi, WSi etc
- Deposit metal such as Ti or Ni on previous device structure without the metal layers.
- Metal only comes in contact with Si at the exposed S/D and gate regions
- Silicide only forms where there is silicon underneath. Metal is unreacted in other regions
- Now selectively etch metal away



(a)



(b)

- Scaling theory suggests short-channel effects can be improved

Circa early 1990s

$$L \sim 0.18\mu\text{m}$$

$$x_{\text{ox}} \sim 3.2 \text{ nm}$$

$$x_j \sim 100 \text{ nm}$$

$$V_{DD} = 2.5 \text{ V}$$

$$V_T \sim 0.6 \text{ V}$$

Circa late 1990s

$$L \sim 0.13\mu\text{m}$$

$$x_{\text{ox}} \sim 2 \text{ nm}$$

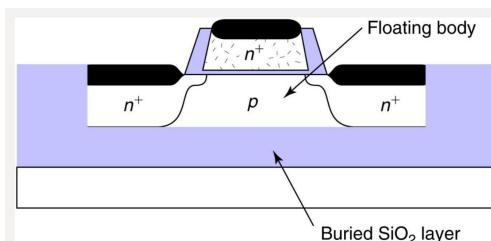
$$x_j \sim 60 \text{ nm}$$

$$V_{DD} = 1.8 \text{ V}$$

$$V_T \sim 0.5 \text{ V by raising body doping}$$

- But mobility will degrade due to ionized impurity scattering

- Introduction of  $p+$  pockets (halo doping):
- High doping at source and drain ends of the channel helps to screen the electric field lines emanating from drain at high VDS.
- Can be fabricated in a self-aligned way by using angled ion implantation of the p-type species once the gate is formed
- But reverse short channel effect!
- Super-steep retrograde (SSR) body doping
- The idea is to form a relatively highly doped buried layer under the gate while preserving a low doping level at the surface where the channel is induced
- Buried layer formed by blanket ion-implantation prior to gate formation
- $P+$  doping layer compensated by S/D doping
- By introducing these 2 techniques, MOSFET successfully scaled to sub-0.1 um gate length regime without short channel effects



Buried SiO<sub>2</sub> layer

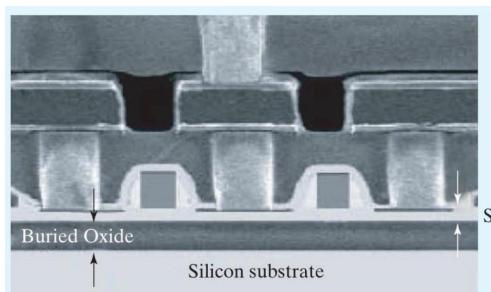
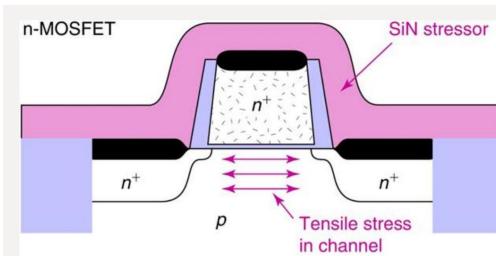


FIGURE 7-17 The cross-sectional electron micrograph of an SOI integrated circuit. The lower level structures are transistors and contacts. The upper two levels are the vias and the interconnects, which employ multiple layers of materials to achieve better reliability and etch stops.

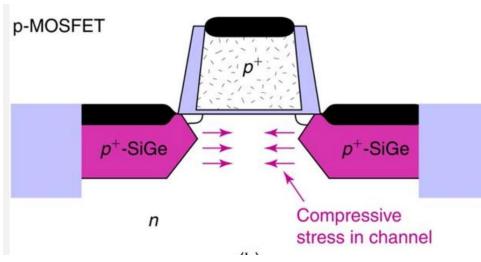
- Because of buried oxide, MOSFET on SOI enjoys reduced junction capacitance
- Means improved switching speed

- Floating body
- Under normal operating conditions, the body gets forward biased due to holes produced through impact ionization
- The holes cannot escape and pile up
- This reduces  $V_{th}$  and enhances gate overdrive, leading to more drain current
- Bipolar effect
- Disadvantage:
- $V_{th}$  is unpredictable due to floating body

## Strained silicon



(a)



(b)

Circa early 2000s

$$L \sim 50 \text{ nm}$$

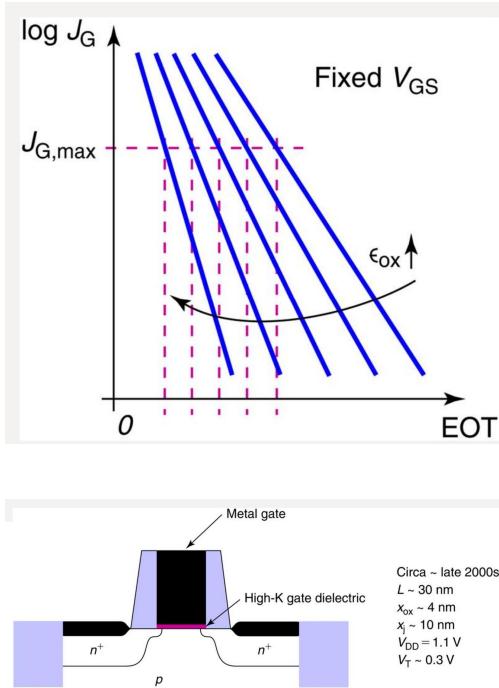
$$x_{\text{ox}} \sim 1.2 \text{ nm}$$

$$x_j \sim 20 \text{ nm}$$

$$V_{DD} = 1.4 \text{ V}$$

$$V_T \sim 0.4 \text{ V}$$

## High-k/metal gate



- Re-engineer the "holiest" element of a traditional MOSFET - the Si/SiO<sub>2</sub> interface!!
- Also need new metals
- Got rid of polysilicon depletion
- Two metals needed to be introduced for  $V_T$  adjustments - for PMOS and NMOS

## FinFET to Gate-all-around

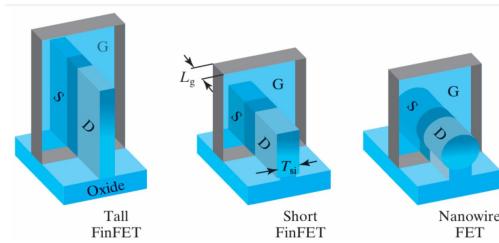


FIGURE 7-19 Variations of FinFET. Tall FinFET has the advantage of providing a large  $W$  and therefore large  $I_{on}$  while occupying a small footprint. Short FinFET has the advantage of less challenging lithography and etching. Nanowire FET gives the gate even more control over the transistor body by surrounding it. FinFETs can also be fabricated on bulk Si substrates.

## Introducing new channel materials

- Ge for p-channel
- InGaAs, InAs, InSb for n channel
- Dielectric?