Status Quo and future of IC technologies

IC/semiconductors(Logic/Memory/Analog/Micro +OSD)

3D x3D x3D

Chee Wee Liu (劉致為)

Distinguished/Chair Professor / IEEE Fellow,
National Taiwan University

cliu@ntu.edu.tw

http://nanosioe.ee.ntu.edu.tw

High mobility/high K/FinFET

PPACR (performance/power/area/ cost/reliability)

Google Scholar Page

Citation :8305/ H index:41

i10 index:201

Handout content coverage

- Chapter 1 Electrons and Holes in Semiconductors
- Chapter 2 Motion and Recombination
- Chapter 4 PN and Metal-Semiconductor Junctions
- Chapter 5 MOS Capacitor
- Chapter 6 MOSFET

The slides are based on the class notes of Prof. Chenming Hu (胡正明), UC Berkely

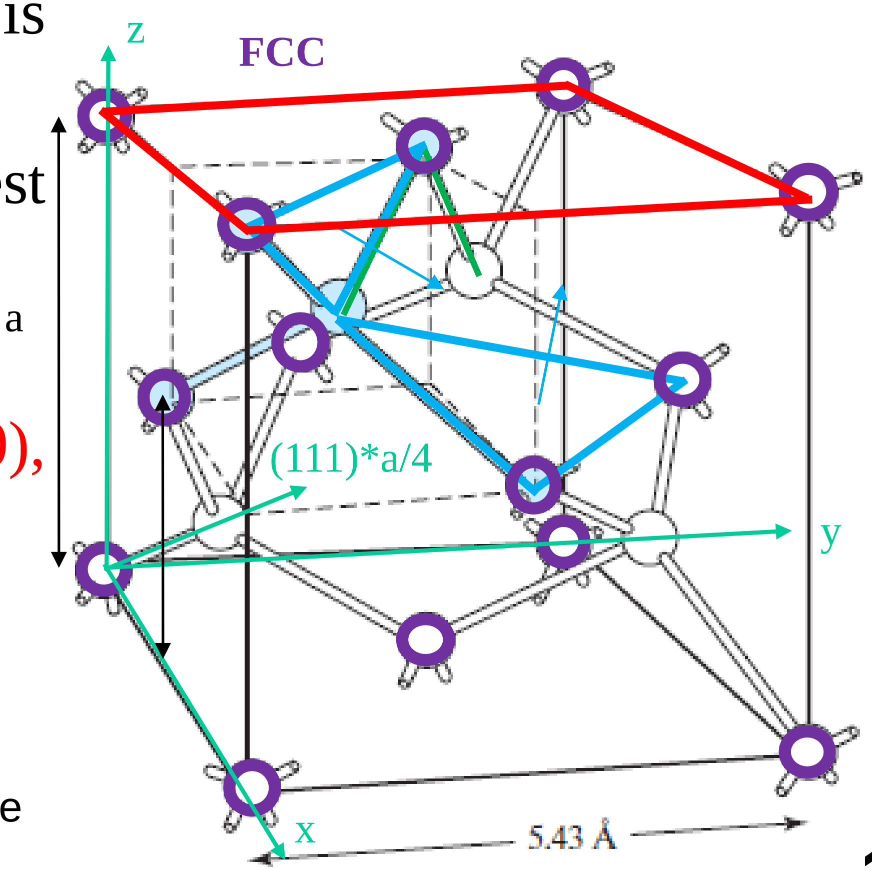
Prof. Chee Wee Liu (劉致為), Email: cliu@ntu.edu.tw
Tao Chou (周韜), Email: tommyc1031@gmail.com

There are some “Practices” in the content and exercise chapters for readers to exam themselves.

Chapter 1   Electrons and Holes in Semiconductors

1.1 Silicon Crystal Structure

- ***Unit cell*** of silicon crystal is cubic
- Each Si atom has 4 nearest neighbors
- Monolayer = a/4
- Two basis (dumbbell): (000), (111)*a/4

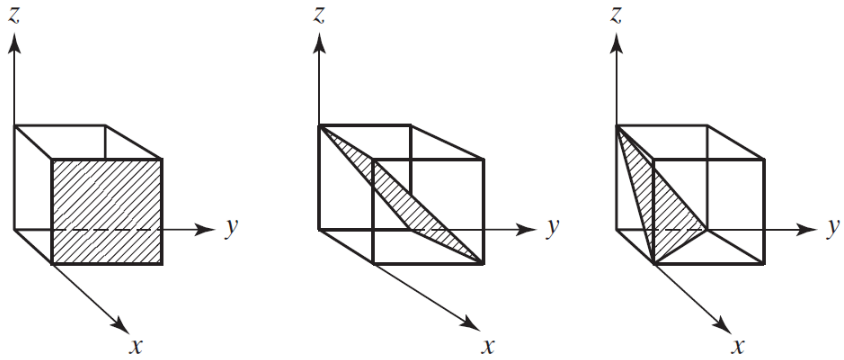


Some rotatable online Si crystal model

1. Diamond crystal structure
2. materialsproject.org
3. sketchfab.com

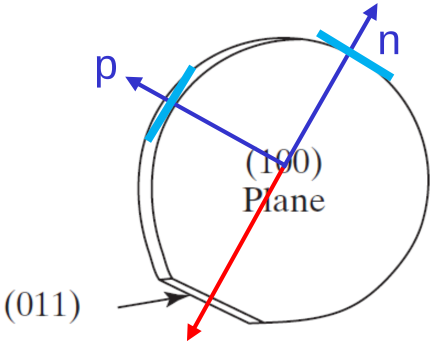
Practice1: draw the diamond structure

1.2.1 Silicon Wafers and Crystal Planes

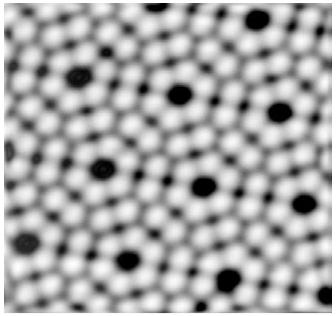


(100) (011) (111)

hexagonal sysmetry
Example: GaN on 111 Si substrate



90 degree: p-type
180 degree: n-type

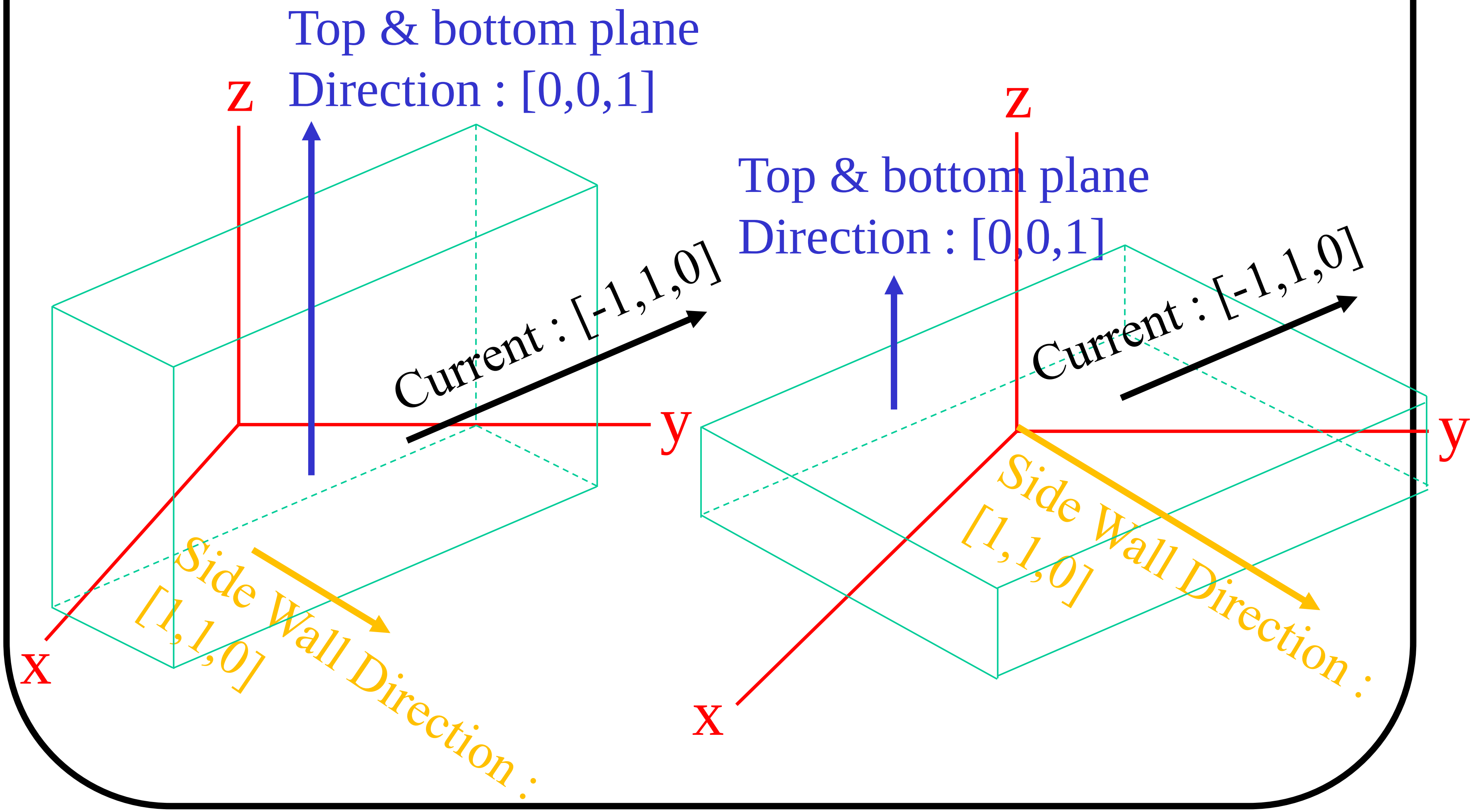


Si (111) plane

- The standard notation for crystal planes is based on the cubic unit cell.

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

Practice2: draw the FinFET and nanosheet channels



FinFET

Top & bottom plane

Direction : [0,0,1]

Current : [-1,1,0]

Side Wall Direction :

[1,1,0]

Nanosheet

Top & bottom plane

Direction : [0,0,1]

Current : [-1,1,0]

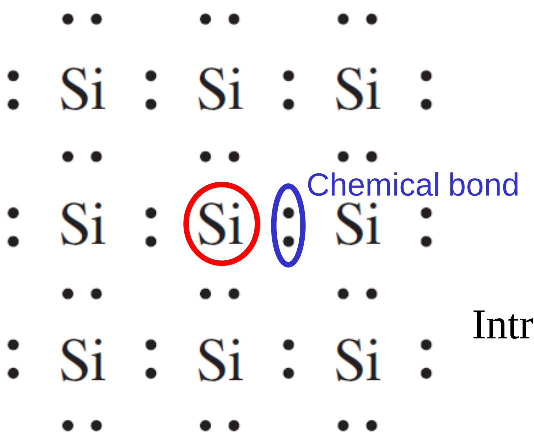
Side Wall Direction :

[1,1,0]

1.2.2 Bond Model of Electrons and Holes

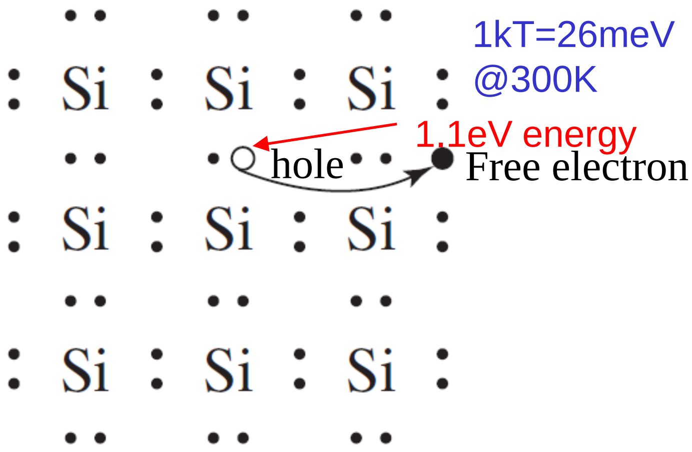
Si-Si bond ~2eV

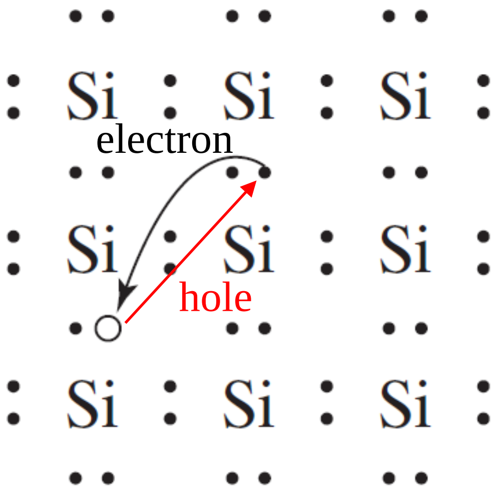
group 4 semiconductor



- Silicon crystal in a two-dimensional representation

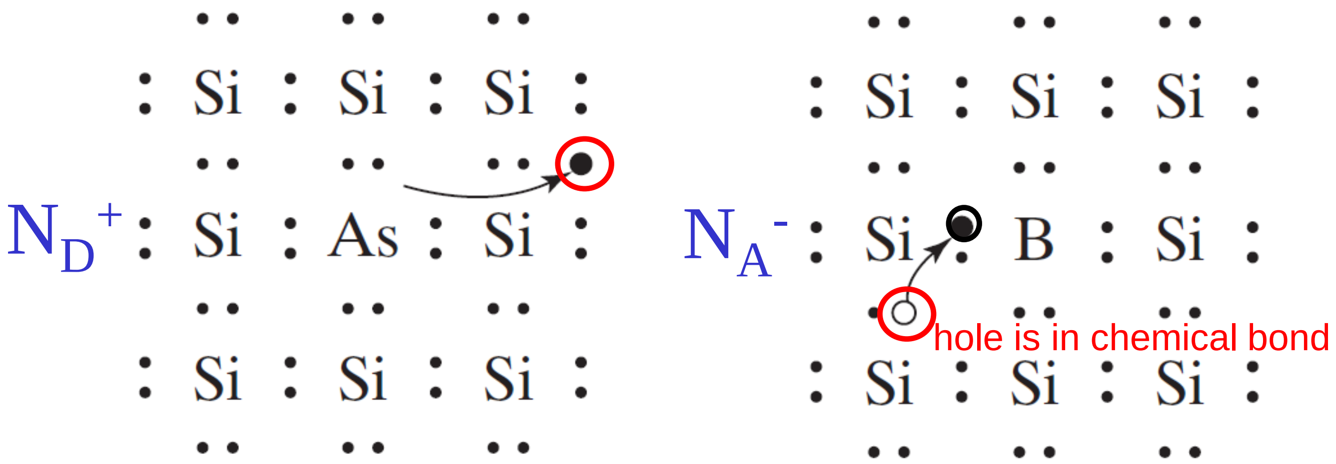
$$\text{Intrinsic}: \frac{10^{10} cm^{-3}}{10^{23} cm^{-3}} = 10^{-13} = \frac{1}{10^{4} billion}$$





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

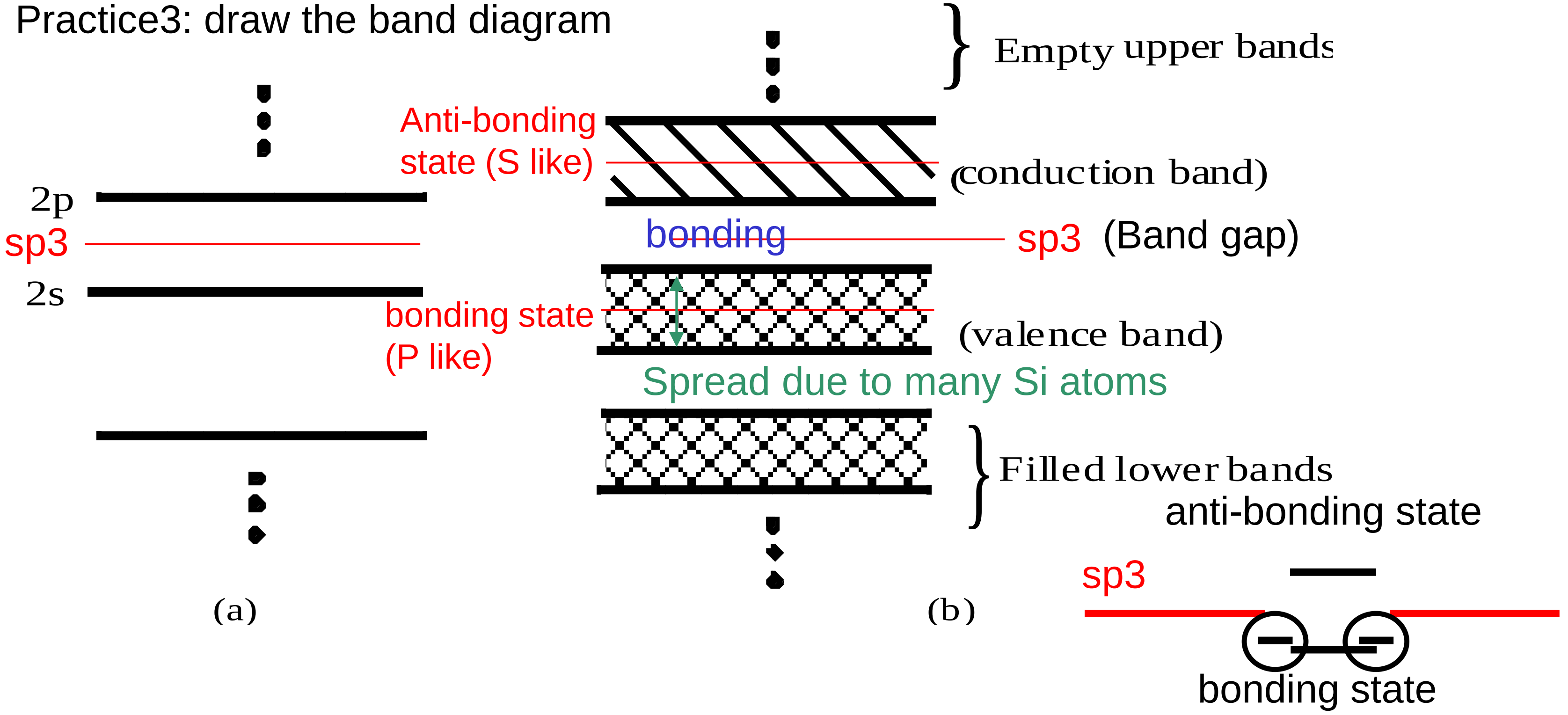
Extrinsic Dopants in Silicon



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

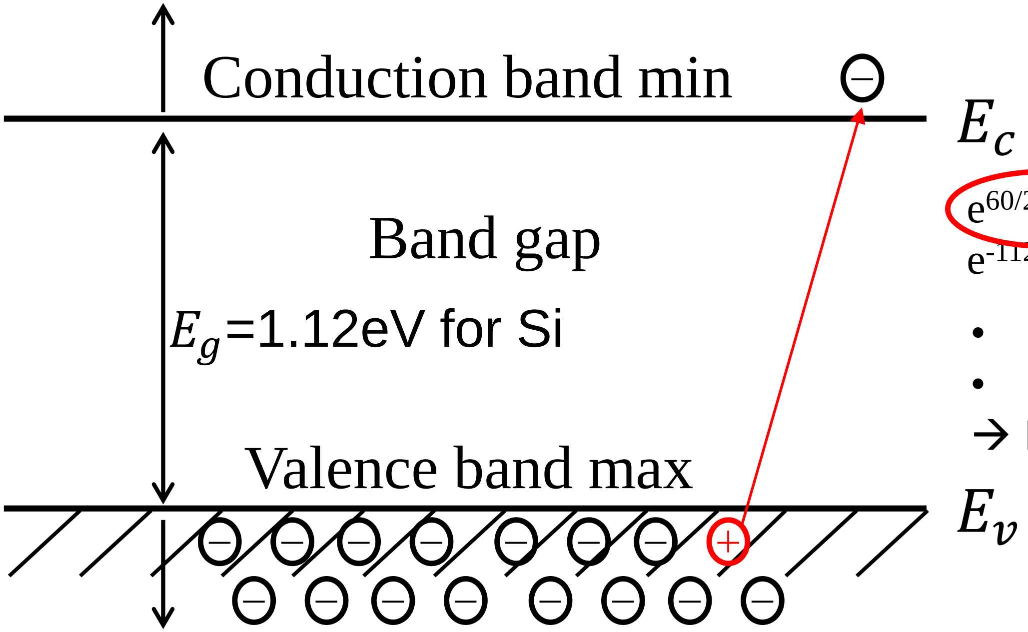
1.3 Energy Band Model

Practice3: draw the band diagram



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

1.3.1 Energy Band Diagram



$e^{60/26} = 10$ → 60meV = 1decade

$e^{-1120/26} = 10^{-1120/60} \sim 10^{-19}$

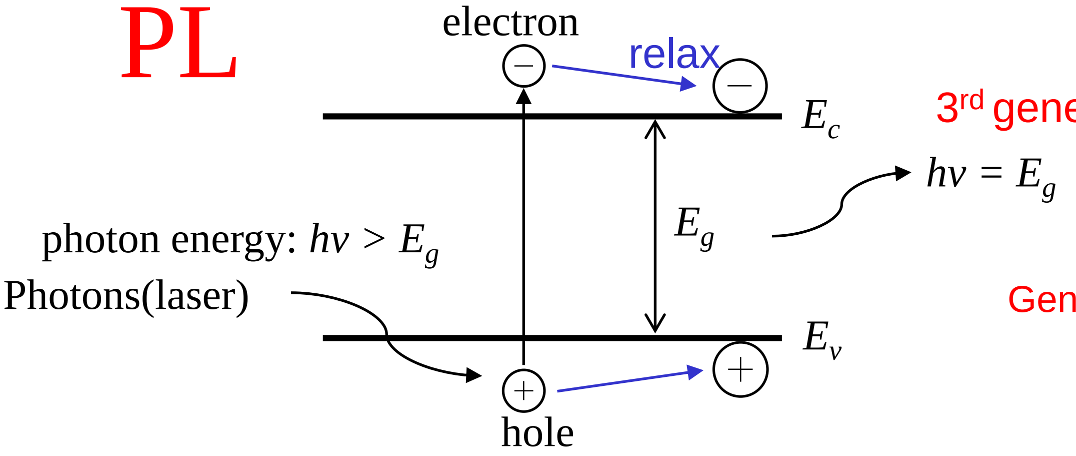
- Intrinsic $n_i = 10^{10}\text{cm}^{-3}$ @300K
- $n_i(400\text{K}) > 10^{10}\text{cm}^{-3}$

→ Higher kT

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

Measuring the Band Gap Energy by Light Absorption

PL



3rd generation (3rd kind)

SiC

GaN

Generation → kind

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

Bandgap energies of selected semiconductors

GaO

Semi-conductor	InSb	Ge	Si	GaAs	GaP	ZnSe	Diamond
Eg (eV)	0.18	0.66	1.12	1.42	2.25	2.7	6

2nd generation (2nd kind)

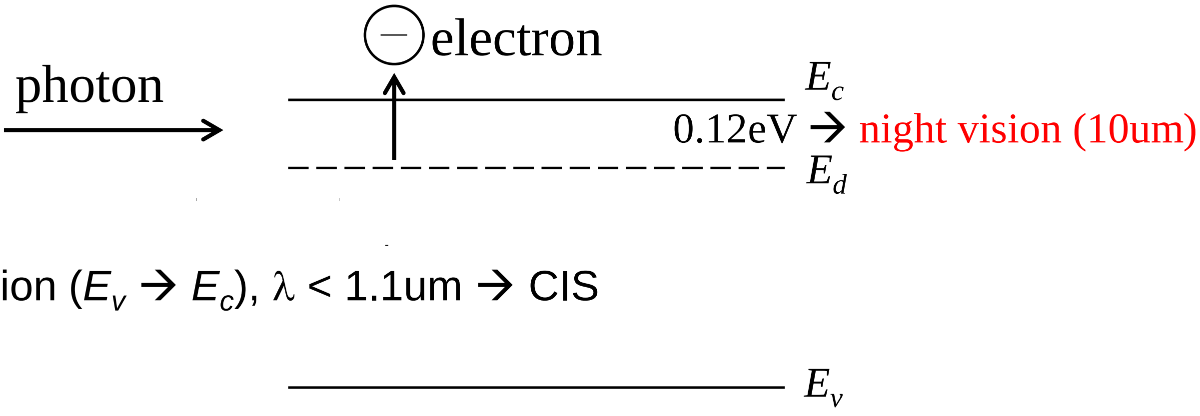
1st generation (1st kind)

4th generation 4th kind)

Infrared Detector Based on Freeze-out

- To image the black-body radiation emitted by tumors requires a photodetector that responds to $h\nu$'s around 0.1 eV.
- In doped Si operating in the freeze-out mode (T=77K), conduction electrons are created when the infrared photons provide the energy to ionized the donor atoms.

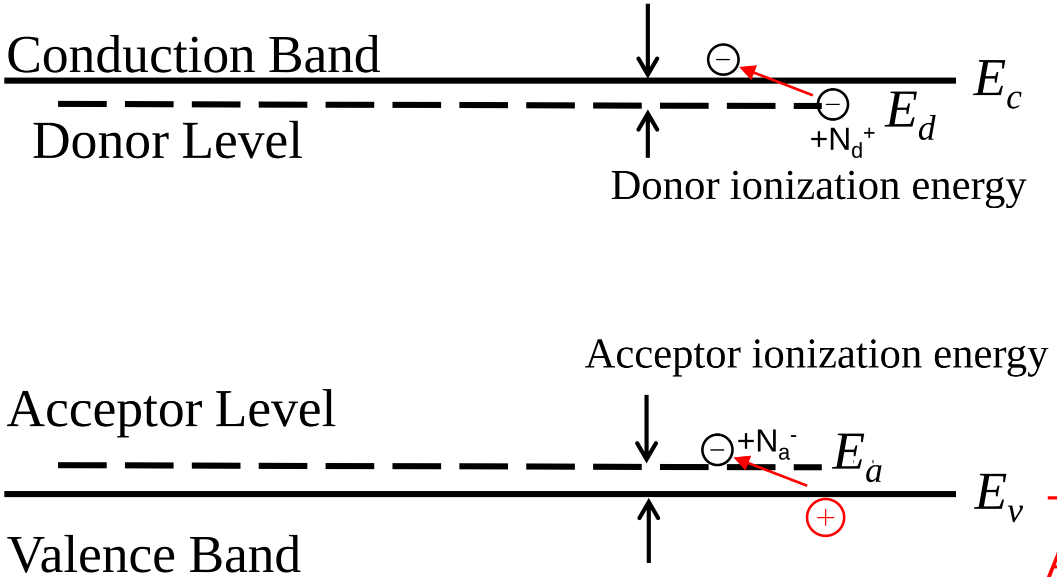
Black body radiation of human: $\lambda = 10$ um



$\lambda = 1.24$eV-um/E

Si: inter-band transition (E_v → E_c), $\lambda < 1.1$um → CIS

1.3.2 Donor and Acceptor in the Band Model



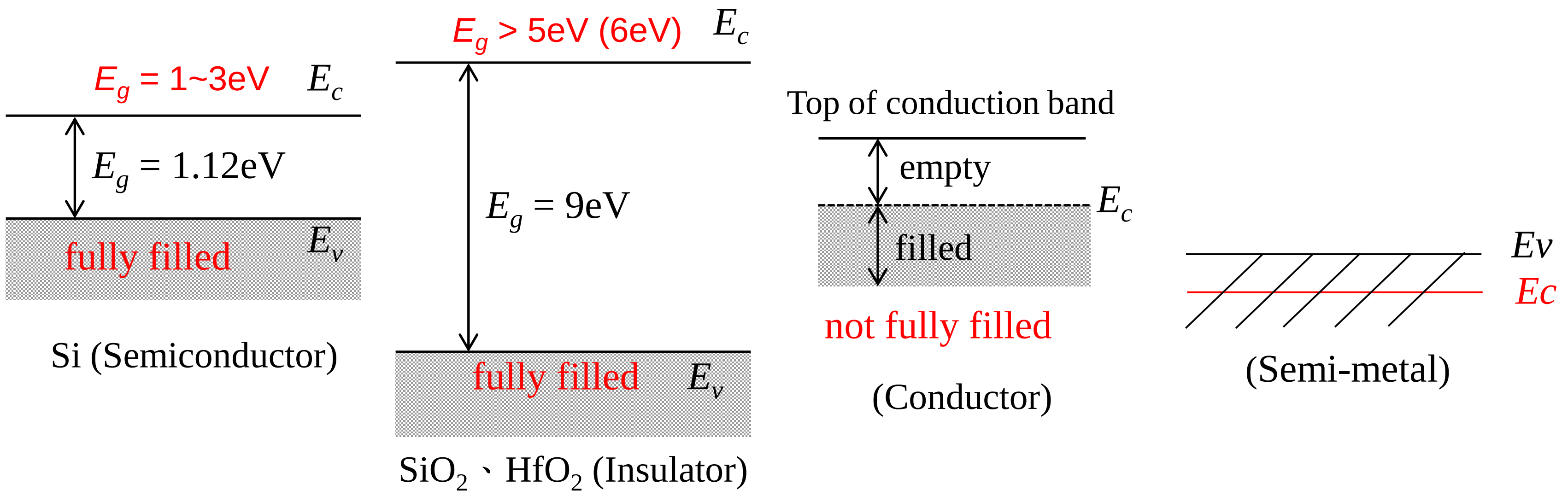
Toxic!
AsH_3 is also toxic
Asadin →As2O3

Ionization energy of selected donors and acceptors in silicon

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

1kT=26meV @300K

1.4 Semiconductors, Insulators, Conductors, and Semi-metal



E_g = 1~3eV E_c

$E_g = 1.12$eV

fully filled E_v

Si (Semiconductor)

E_g > 5eV (6eV) E_c

$E_g = 9$eV

fully filled E_v

SiO_2 、HfO_2 (Insulator)

Top of conduction band

empty

E_c

filled

not fully filled

(Conductor)

Ev

Ec

(Semi-metal)

- Totally filled bands and totally empty bands do not allow current flow. (Just as there is no motion of liquid in a totally filled or totally empty bottle.)
- Metal conduction band is half-filled.
- Semiconductors have lower E_g's than insulators and can be doped.
- α-Sn : $E_C = E_V$ → zero bandgap semiconductor, not semi-metal

1.5 Electrons and Holes

1.5.1 Effective Mass

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

No scattering

$$a = \frac{-eE}{m_n} \quad \text{electrons}$$

$$a = \frac{-eE}{m_p} \quad \text{holes}$$

J = nev

I = J*A

τ = CV/I

Electron and hole effective masses

bad reliability

SiO_2 is good insulator
and GeO_2 is bad

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

SiGe in 5nm node

When n is small, conduction band is a circle

1.5.1 Effective Mass

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

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

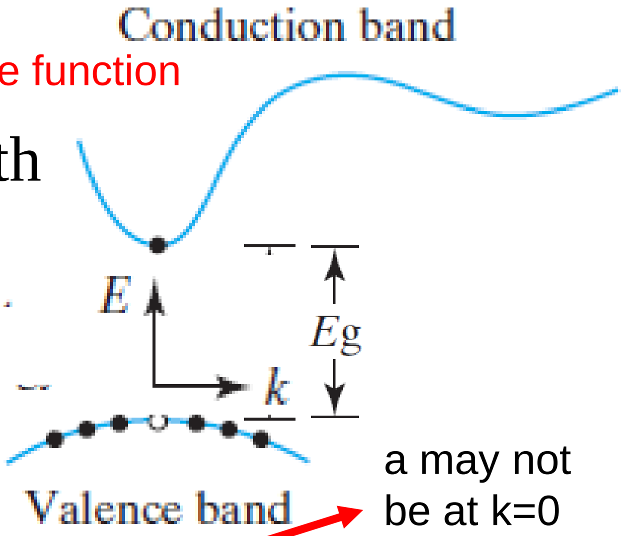
Bloch’s theorem

$\Psi = U(r)e^{\pm ik\cdot r}$

Free electron

$\Psi = e^{\pm ik\cdot r}$

Envelope function



The solution is of the form $e^{\pm ik\cdot r}$

k = wave vector = 2π/electron wavelength

For each k, there is a corresponding E(k) (eigen energy). Band structure: E is function of k.

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

a may not be at k=0

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

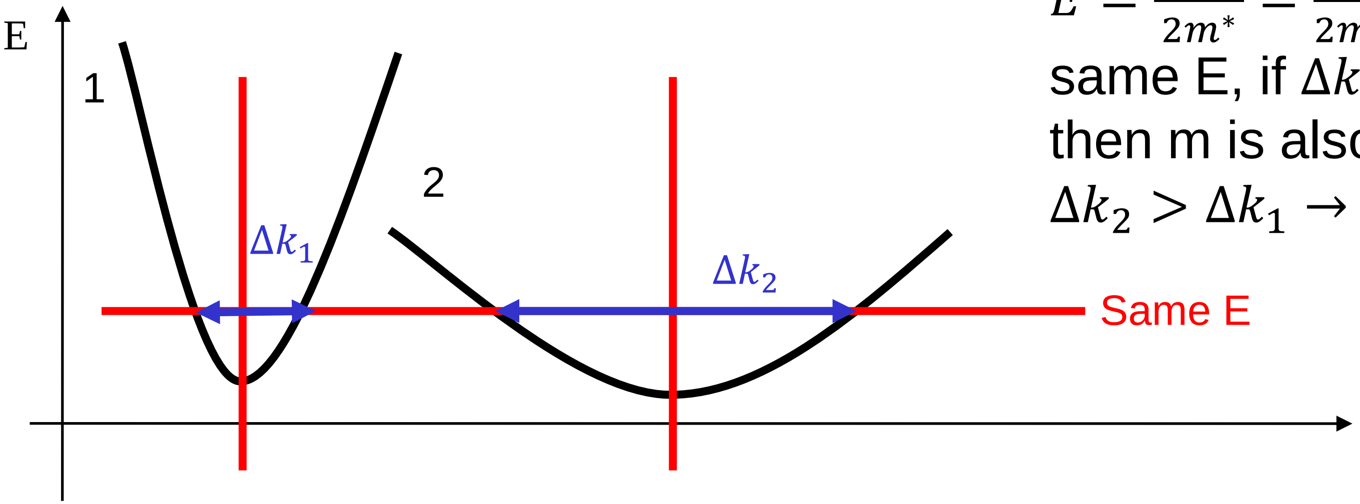
f(x)= $f(a) + \cancel{\frac{f'(a)}{1!}(x-a)} + \frac{f''(a)}{2!}(x-a)^2 + \cancel{\frac{f'''(a)}{3!}(x-a)^3 + \cdots},$

Free electron : $E = \dfrac{\hbar^2k^2}{2m^*}$

Practice4: derive formula of effective mass

- $E = E_c + ak + bk^2 + ck^3 + dk^4 + \cdots$
- Assume parabolic band: $E = E_c + bk^2 \rightarrow E = E_c + \frac{\hbar^2 k^2}{2m^*}$
- $\frac{d^2E}{dk^2} = \frac{\hbar^2}{m^*} \rightarrow m^* = \frac{\hbar^2}{d^2E/dk^2}$
- The electron's behavior near approximation point is just like an electron with effective mass m^*.

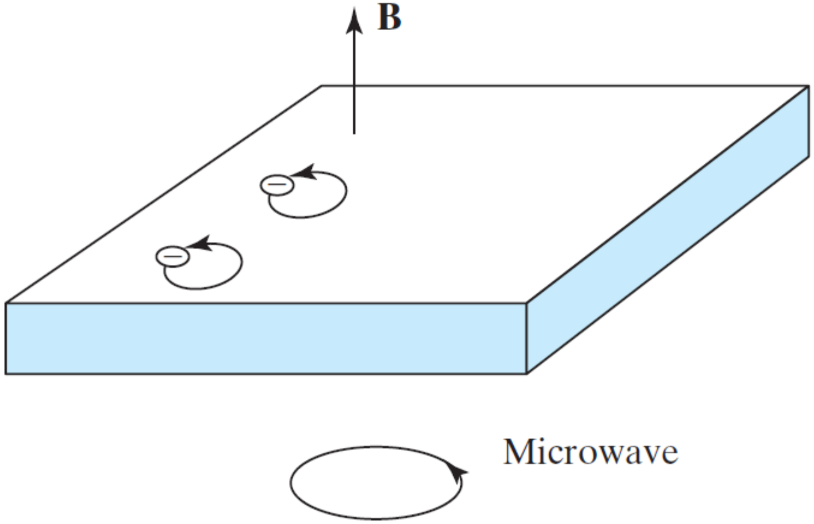
Q: Which one has larger effective mass?



$E = \frac{\hbar^2 k^2}{2m^*} = \frac{P^2}{2m^*} \rightarrow$ For same E, if Δk is larger, then m is also larger

$\Delta k_2 > \Delta k_1 \rightarrow m_2^* > m_1^*$

1.5.2 How to Measure the Effective Mass



Cyclotron Resonance Technique

Centripetal force = Lorentzian force

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

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

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

$$\omega = 2\pi f_{cr} = \frac{eB}{m_n}$$

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

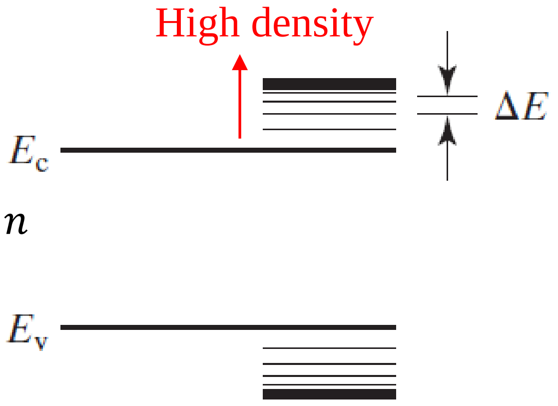
Quantum Computing: MRS bulletin July, 2021
Low temperature similar to Cyclotron Resonane: < 4K

1.6 Density of States To determine n

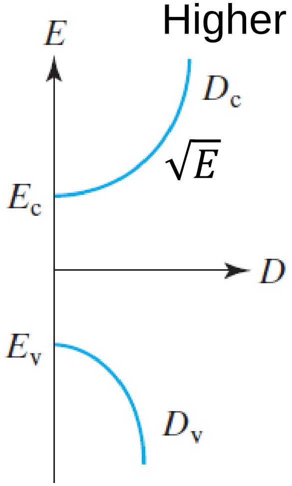
Fermi-dirac distribution similar to boltzmann distribution

$$E = \frac{\hbar^2 k^2}{2m^*}$$

$$k(k_x, k_y, k_z) = \frac{2\pi}{l} * n$$



Higher energy, higher density



$\sqrt{E}$ for 3D
Constant for 2D

The higher energy, the lower occupation probability

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

3D density of state

$$\propto \sqrt{m_x m_y m_z}$$

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

Conduction band edge (minimum energy)

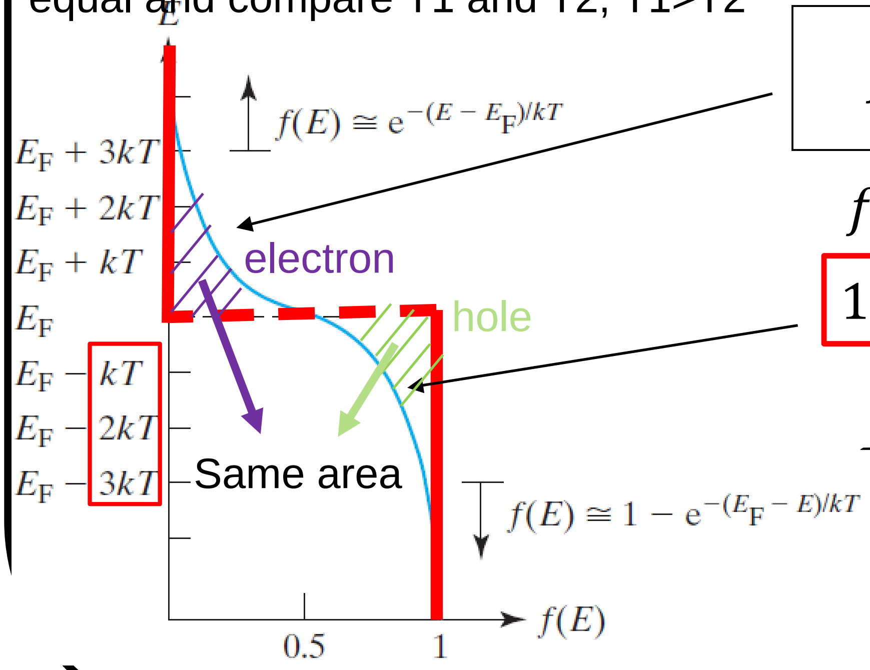
density of state effective mass

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

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

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

Practice5: Plot f(E) and prove the area above and below fermi-level are equal and compare T1 and T2, T1>T2



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

Boltzmann approximation:

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

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

$$\boxed{1 - f(E) \sim e^{-(E_f-E)/kT}}$$ hole distribution

[redacted]

1.8 Electron and Hole Concentrations

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

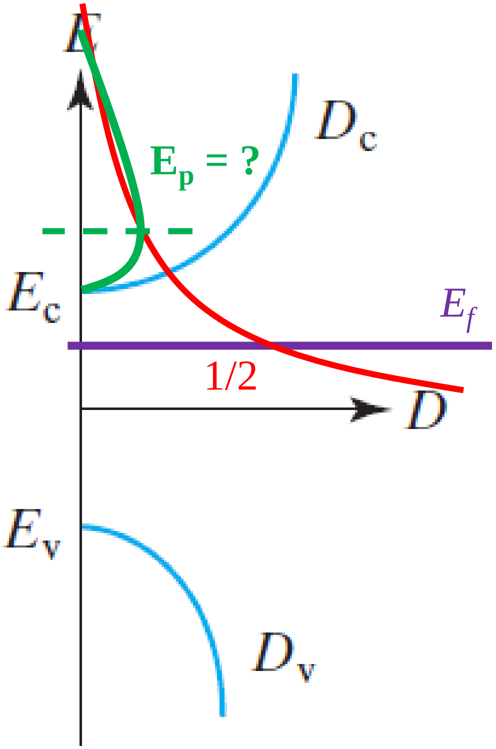
$$\boxed{n = \int_{E_c}^{top\ of\ conduction\ band} f(E) D_c(E) dE \ \ (cm^{-3})}$$

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

Boltzmann approximation

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

Practice6: find E_p(relative to E_c), where f(E)*D(E) is maximum.



Electron and Hole Concentrations

Practice7: write function of n and p

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

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

Density of state effective mass

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

60meV→10X concentration

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

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

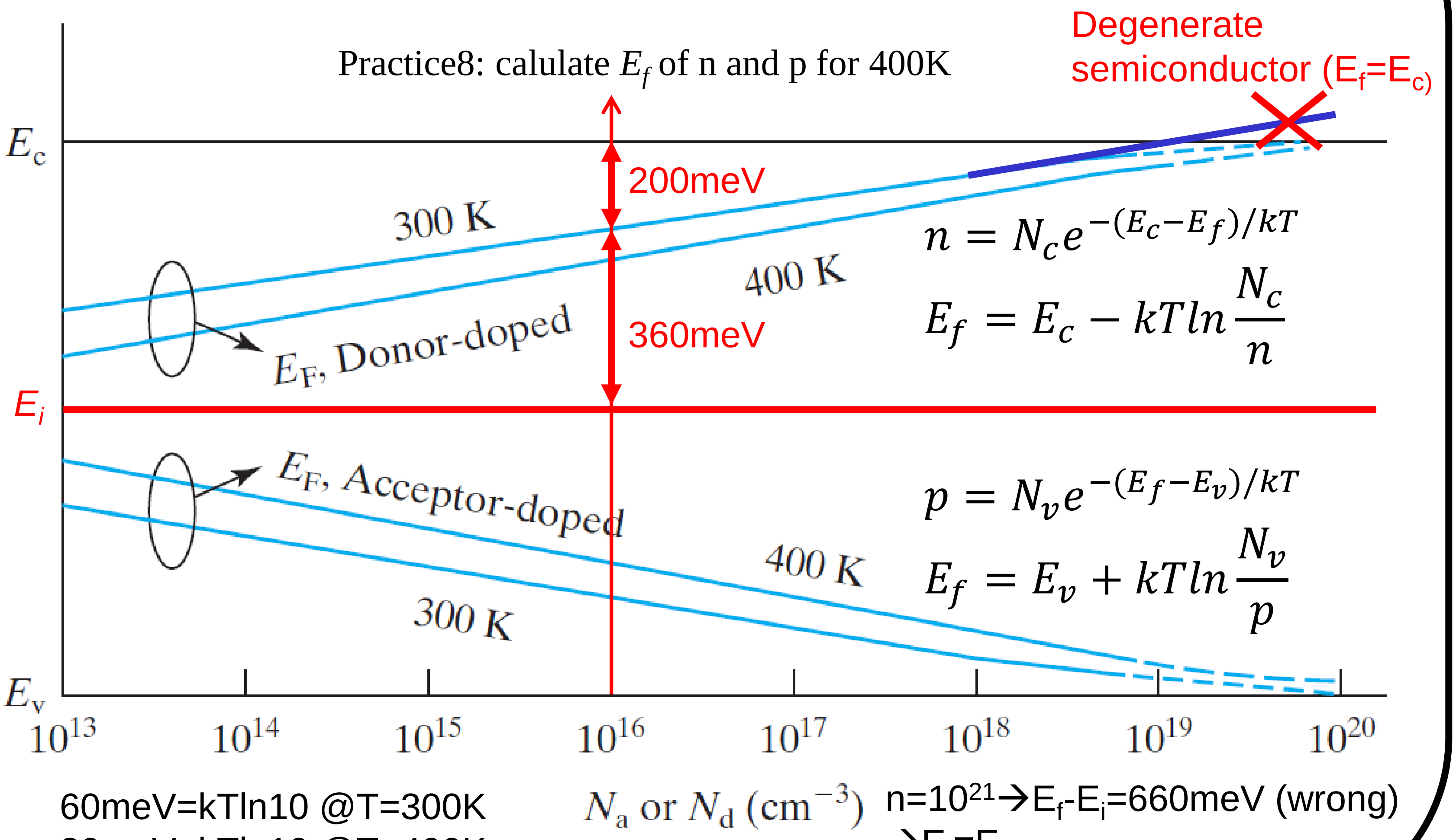
N_v is called the ***effective density of states of the valence band.***

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

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

1.8.2 The Fermi Level and Carrier Concentrations

Practice8: calulate E_f of n and p for 400K



Degenerate semiconductor (E_f=E_c)

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

$$E_f = E_c - kT\ln\frac{N_c}{n}$$

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

$$E_f = E_v + kT\ln\frac{N_v}{p}$$

60meV=kTln10 @T=300K
80meV=kTln10 @T=400K

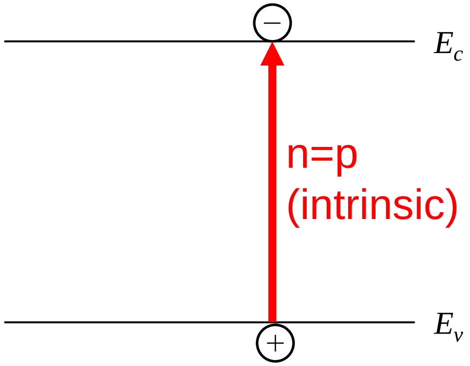
n=10^{21}→E_f-E_i=660meV (wrong)
→E_f=E_c

1.8.3 The *np* Product and the Intrinsic Carrier Concentration

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

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

Practice9: calculate the exact location of E_{fi}

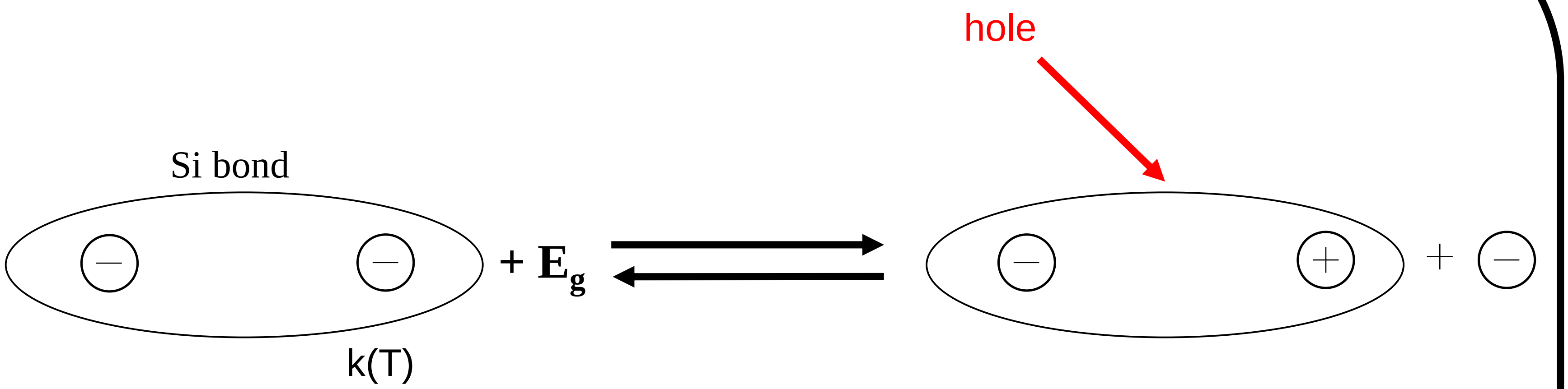


$$\boxed{np = n_i^2}$$

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

$$n_i^2 \propto T^3 e^{-E_g/kT}$$

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



$[A][B] = k[C] = \text{constant}$

$np = \text{constant} = n_i^2 = 10^{20} (\text{cm}^{-6})$

$n_i^2 \propto \boxed{T^3 e^{-E_g/kT}}$

Practice10: What is n_i in 0^oK

$H_2O \leftrightarrows H^+ + OH^-$

$[H^+][OH^-] = k[H_2O] = 10^{-14}(M^2)$ @300K

Ex: n_i @T=0K?

→n_i=0 since no enegy

If A+B ⇆ C+D is at equilibrium

→ [C][D]/[A][B] is a constant (function of T)

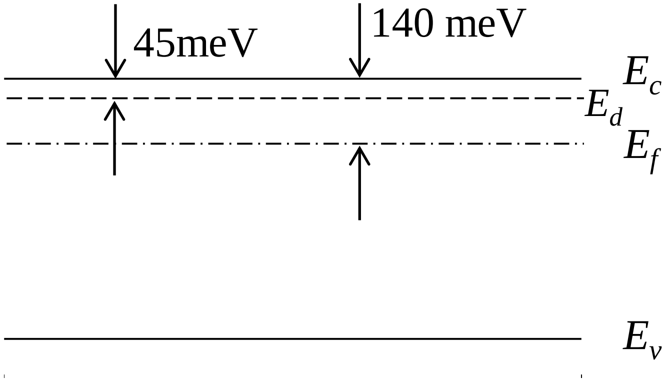
1.9 General Theory of n and p

EXAMPLE: Complete ionization of the dopant atoms

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

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

$$n = N_D^+ = 10^{17} cm^{-3} \rightarrow E_f = E_c - 140meV$$



Practice11: compare ionization rate w/ and w/o coefficient due to spin

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

Probability of not being ionized

$$\approx \frac{1}{1 + \underset{\text{due to spin}}{\boxed{\frac{1}{2}}} e^{(E_d-E_f)/kT}} = \frac{1}{1 + \frac{1}{2} e^{(140-45)meV/26meV}} = 0.04$$

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

1.9 General Theory of n and p

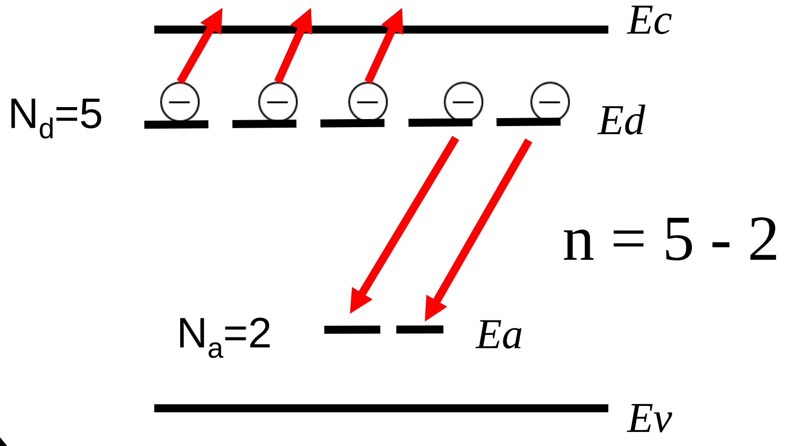
Charge neutrality:

Practice12: derive this two eq.

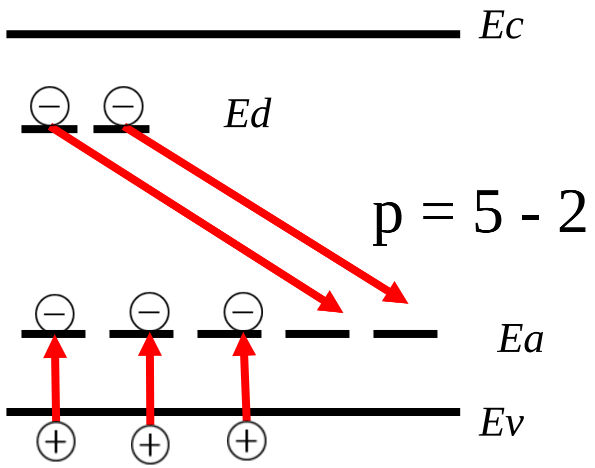
$n + N_a^- = p + N_d^+ \qquad np = n_i^2$

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

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



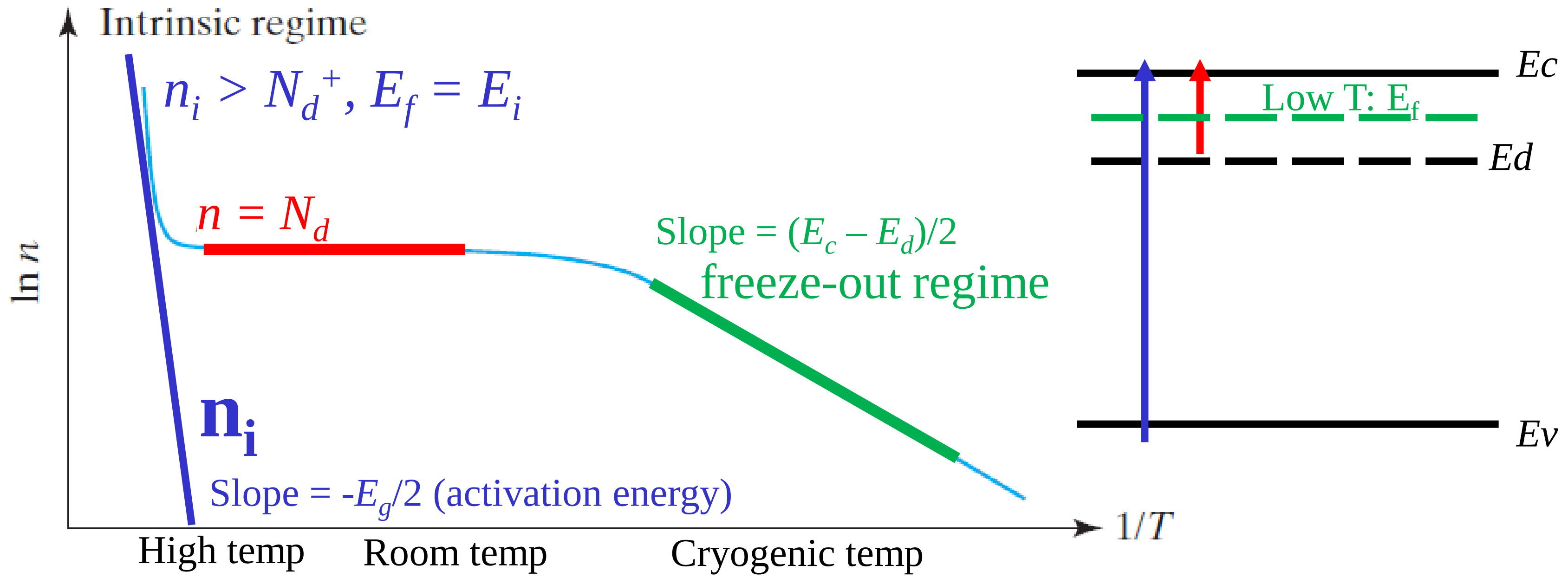
n = 5 - 2



p = 5 - 2

1.10 Carrier Concentrations at Extremely High and Low Temperatures

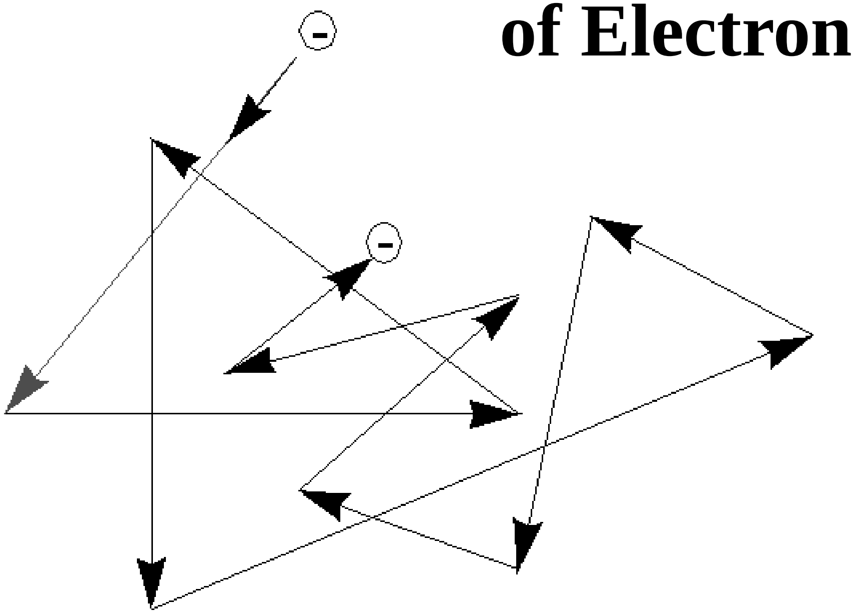
Arrhenius plot:
$\text{Log}_{10}y$ vs 1/T



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

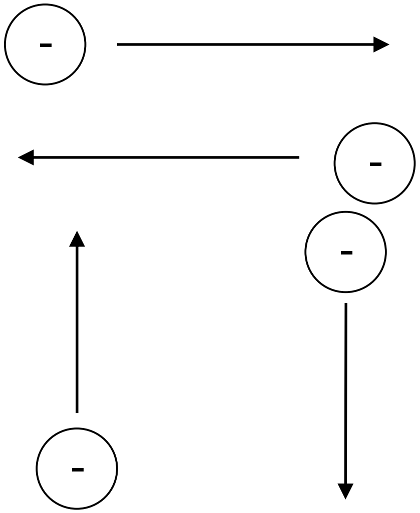
$$\text{Low T: } n = \left[\frac{N_c N_d}{2}\right]^{1/2} e^{-(E_c - E_d)/2kT}$$

Chapter 2 Motion and Recombination of Electrons and Holes



E change → inelastic scattering

E not change → elastic scattering



Ensemble statistic: average $<V_{th}> = 0$

- Zig-zag motion is due to collisions or scattering with imperfections in the crystal.
- Net thermal velocity <**V**> is zero (Ensemble average).
- Mean time between collisions (Scattering time) is $\tau_m \sim 0.1\text{ps}$

2.1 Thermal Motion

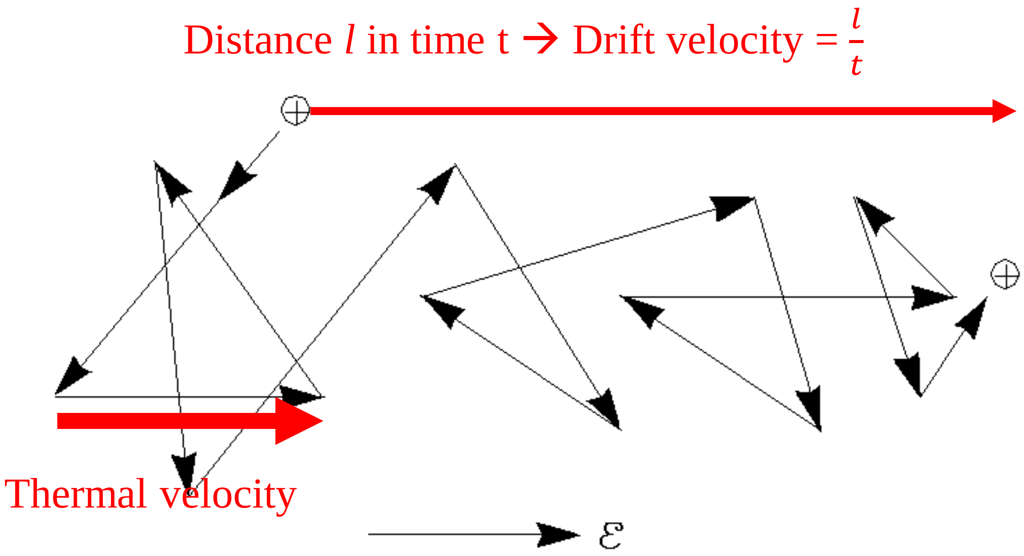
$$\text{Average electron or hole kinetic energy}\ \overset{<V^2>}{} = \frac{3}{2}kT = \frac{1}{2}mv_{th}^2$$

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

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

Saturation Velocity ~1E7 cm/s

2.2 Drift velocity

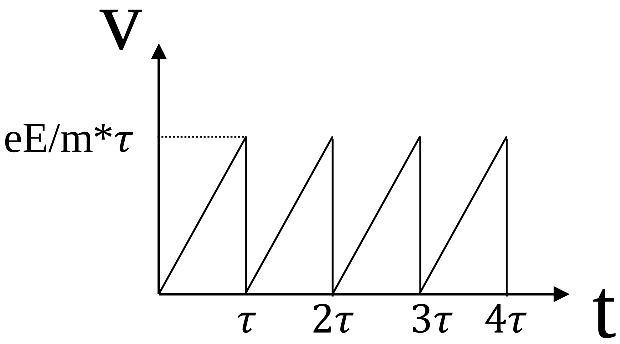


Distance l in time t → Drift velocity = $\frac{l}{t}$

Thermal velocity

$a=F/m=eE/m$

$\Delta V = a*\tau = eE\tau/m$



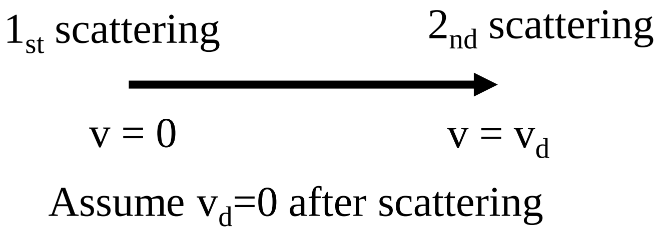
$<V> = V_{drift} = ½*eE/m*\tau =$
$eE/m*\tau$ (stochastic process)
$V_{drift} = \mu E$

- ***Drift*** is the motion caused by an electric field.
- If E=0 → $< l > = 0$ (ensemble) → $v_{drift} = 0$

Real $V = V+\Delta V$,
Drift Velocity=$<V> = <V_{th}> +eE\tau/m^{*}= 0+eE\tau/m^{*} = \mu E$

2.2.1 Electron and Hole Mobilities

$$\Delta p = F\Delta t \qquad m^{*}\overset{\nearrow\; v = v_d}{v} = eE\tau_m \qquad \mu = \frac{e\tau_m}{m^{*}}$$



1_{st} scattering

2_{nd} scattering

$v = 0$

$v = v_d$

Assume $v_d = 0$ after scattering

$$v = \frac{eE\tau_m}{m^{*}}$$

- High mobility channel (SiGe for pFET) used in N5
- 90nm: strained Si for both nFET and pFET

$$v_n = \mu_n E$$

$$\mu_n = \frac{e\tau_{mn}}{m_n^{*}}$$

$$v_p = \mu_p E$$

$$\mu_p = \frac{e\tau_{mp}}{m_p^{*}}$$

- μ_p is the hole mobility and μ_n is the electron mobility
- Mobility is the most important knob

2.2.1 Electron and Hole Mobilities

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

$J = ne\mu E$

Electron and hole mobilities of selected semiconductors

area of pFET > area of nFET due to smaller μ_p

Due to small n

	Si	**Ge**	~~**GaAs**~~	~~**InAs**~~
μ_n (cm^2/V·s)	1400	3900	~~8500~~	~~30000~~
μ_p (cm^2/V·s)	470	1900	~~400~~	~~500~~

Si is most used

→SiO_2 is a good oxide

$\mu_{n,ntu}$ = 2,400,000 @<1K

Not For CMOS

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

Name	Symbol	Germanium	Silicon
Band minimum at $k = 0$			
Minimum energy	$E_{g,\mathrm{direct}}$ [eV]	0.8	3.2
Effective mass Sphere (m_x=m_y=m_z)	$m_e{}^*/m_0$	0.041	?0.2?
Band minimum *not* at $k = 0$		L(111)$\frac{\pi}{a}$	Δ0.85$\frac{\pi}{a}$(001)
Minimum energy	$E_{g,\mathrm{indirect}}$ [eV]	0.66	1.12
Longitudinal effective mass ellipsoid	$m_{e,l}{}^*/m_0$	1.64	0.98
Transverse effective mass	$m_{e,t}{}^*/m_0$	0.082	0.19
Wavenumber at minimum	k [1/nm]	xxx	xxx
Longitudinal direction		(111)	(100)
Heavy hole valence band maximum at $E = k = 0$			
Effective mass	$m_{hh}{}^*/m_0$	0.28	0.49
Light hole valence band maximum at $k = 0$			
Effective mass	$m_{lh}{}^*/m_0$	0.044	0.16
Split-off hole valence band maximum at $k = 0$			
Split-off band valence band energy	$E_{v,so}$ [eV]	-0.028	-0.044
Effective mass	$m_{h,so}{}^*/m_0$	0.084	0.29

E(k) Space: band structure

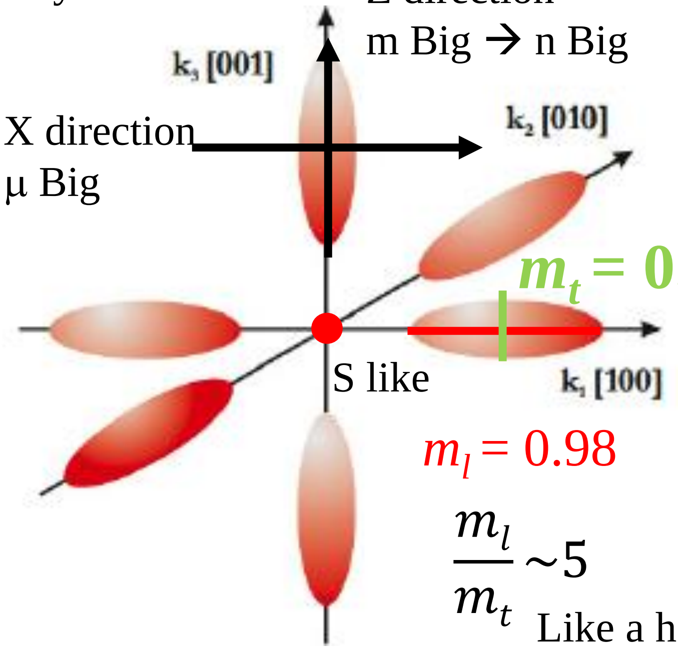
Si: Δ point, $0.85\frac{\pi}{a}(001)$

All the valleys are available

Ge: **Γ** point, $\frac{\pi}{a}(111)$

Only half of the valleys can be used

→ Share with other bw zone



Z direction

m Big → n Big

X direction

μ Big

$m_t = 0.19$

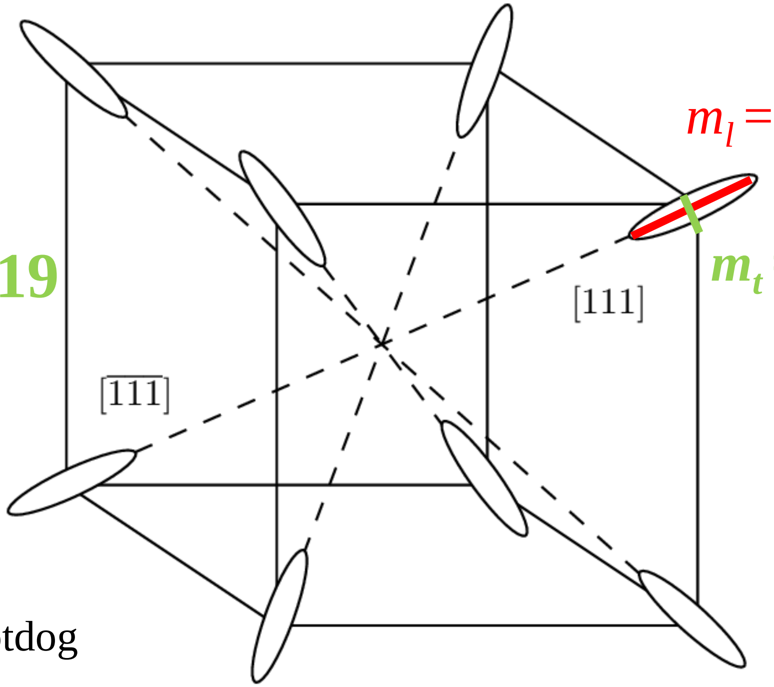
S like

$m_l = 0.98$

$\frac{m_l}{m_t} \sim 5$

Like a hotdog

Degeneracy = 6



$m_l = 1.64$

$m_t = 0.082$

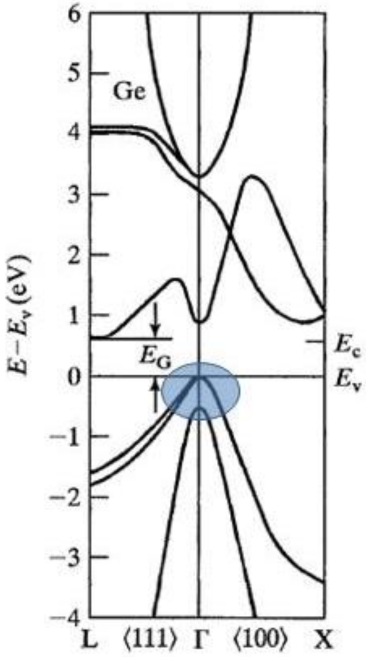
Degeneracy = 8

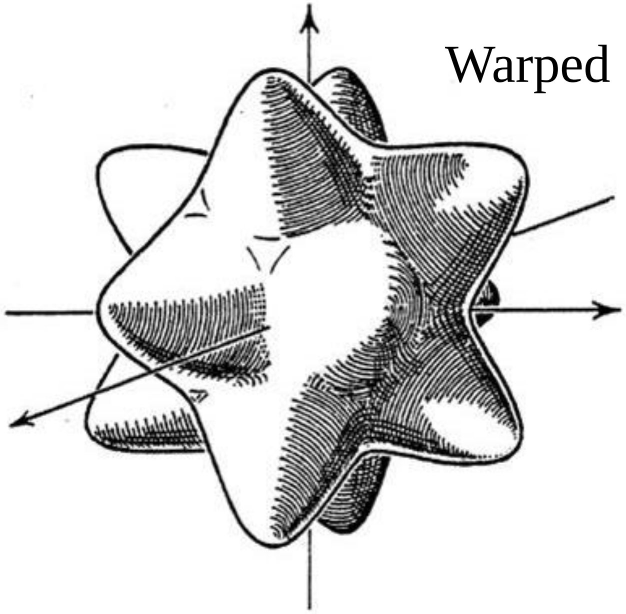
$\frac{m_l}{m_t} \sim 20$

Like a bamboo

III-V : $\frac{m_l}{m_t} \sim 1$ like a sphere

Constant E-surface for Valence Band



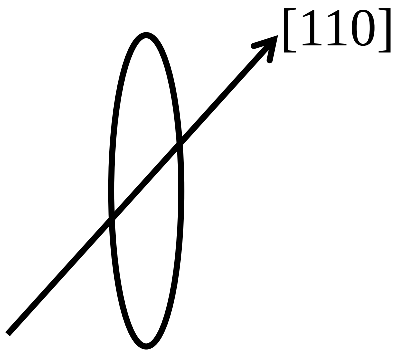


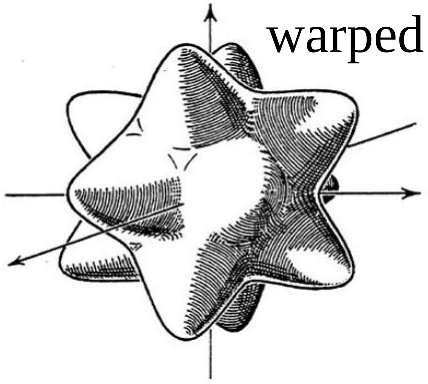
Warped Surface

$$E = E_v - Ak^2 \mp \sqrt{[B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)]}$$

Si: A=4.29, B=0.68, C=4.87; Ge: A=13.38, B=8.48, C=13.15

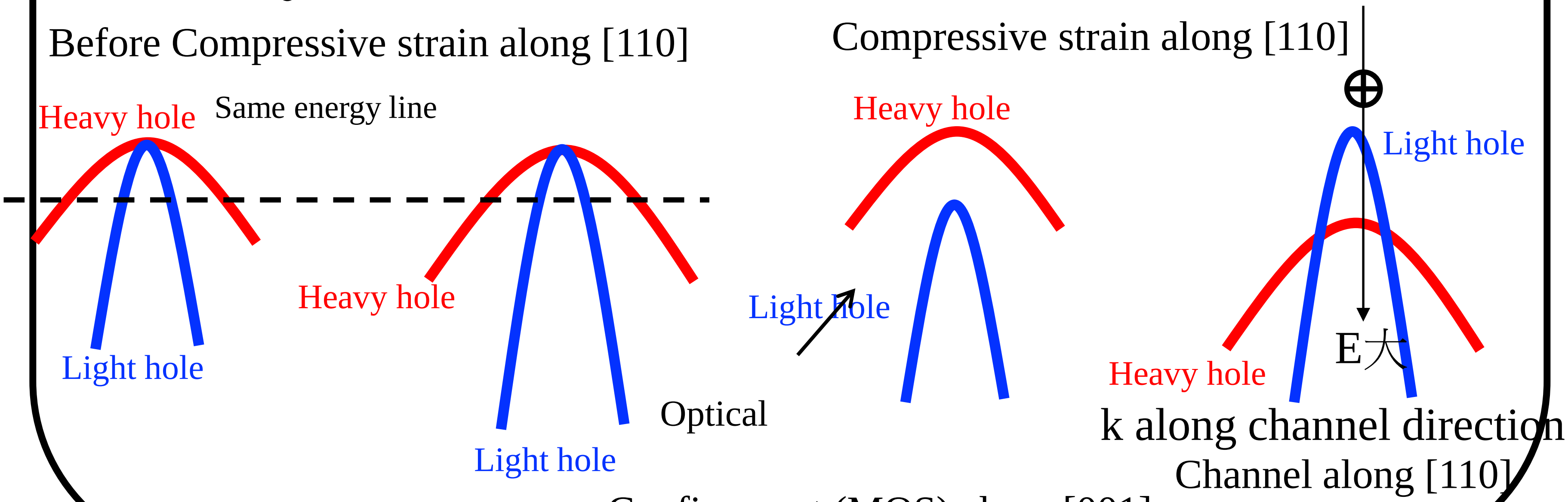
I_{ON} Improvement -- strain





Before Compressive strain along [110]

Compressive strain along [110]



Confinement (MOS) along [001]

I_{ON} Improvement -- strain

$\boxed{J = ne\mu E}$

- $Channel\ direction\ transport\ mass:\ m^* \downarrow \rightarrow \mu \uparrow \rightarrow J \uparrow$
- $In - plane\ confinement\ mass: m^* \uparrow \rightarrow D(E) \uparrow \rightarrow J \uparrow$
- $Confinement\ mass \uparrow \rightarrow ground\ energy \downarrow$ $\rightarrow wavefunction\ more\ localize$
- $nFET(electron):$ $tensile\ strain$ $along\ channel$→ Δ2 valley energy becomes lower→ $transport\ mass = 0.19m_0/confinement\ mass = 0.92m_0$ ☺
- $pFET(hole):$ $compressive\ strain$ $along\ channel$→ Light hole energy becomes higher. ☺

$\rightarrow transport\ mass = 0.16m_0/confinement\ mass = 0.49m_0$

2.2.2 Mechanisms of Carrier Scattering

There are two main causes of carrier scattering:

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

Phonon = lattice wave: T=0K minimum

T↑, phonon scattering ↑

Phonon scattering mobility decreases when temperature rises:

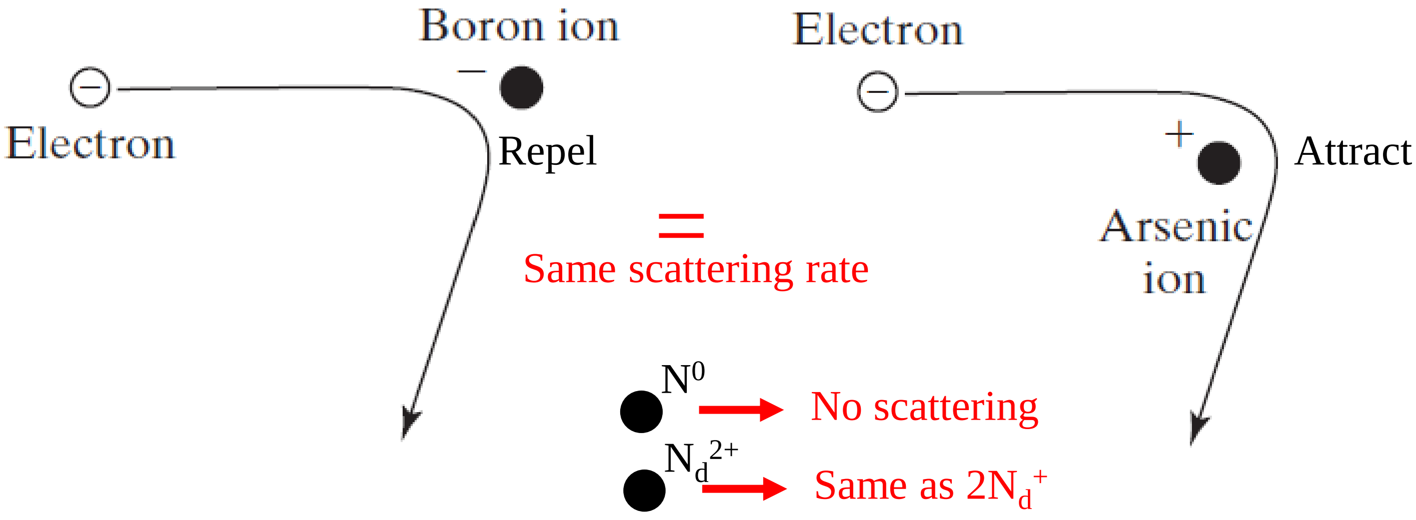
$$\mu_{phonon} \propto \tau_{phonon} \propto \frac{1}{phonon\ \ density \times carrier\ \ thermal\ \ velocity} \propto \frac{1}{T \times T^{1/2}} \propto T^{\boxed{-3/2}}$$

$$\mu = \frac{e\tau_m}{m^*}$$

$$\propto T$$

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

Impurity (Dopant)-Ion Scattering or Coulomb Scattering



=

Same scattering rate

N^0 → No scattering

N_d^{2+} → Same as $2N_d^+$

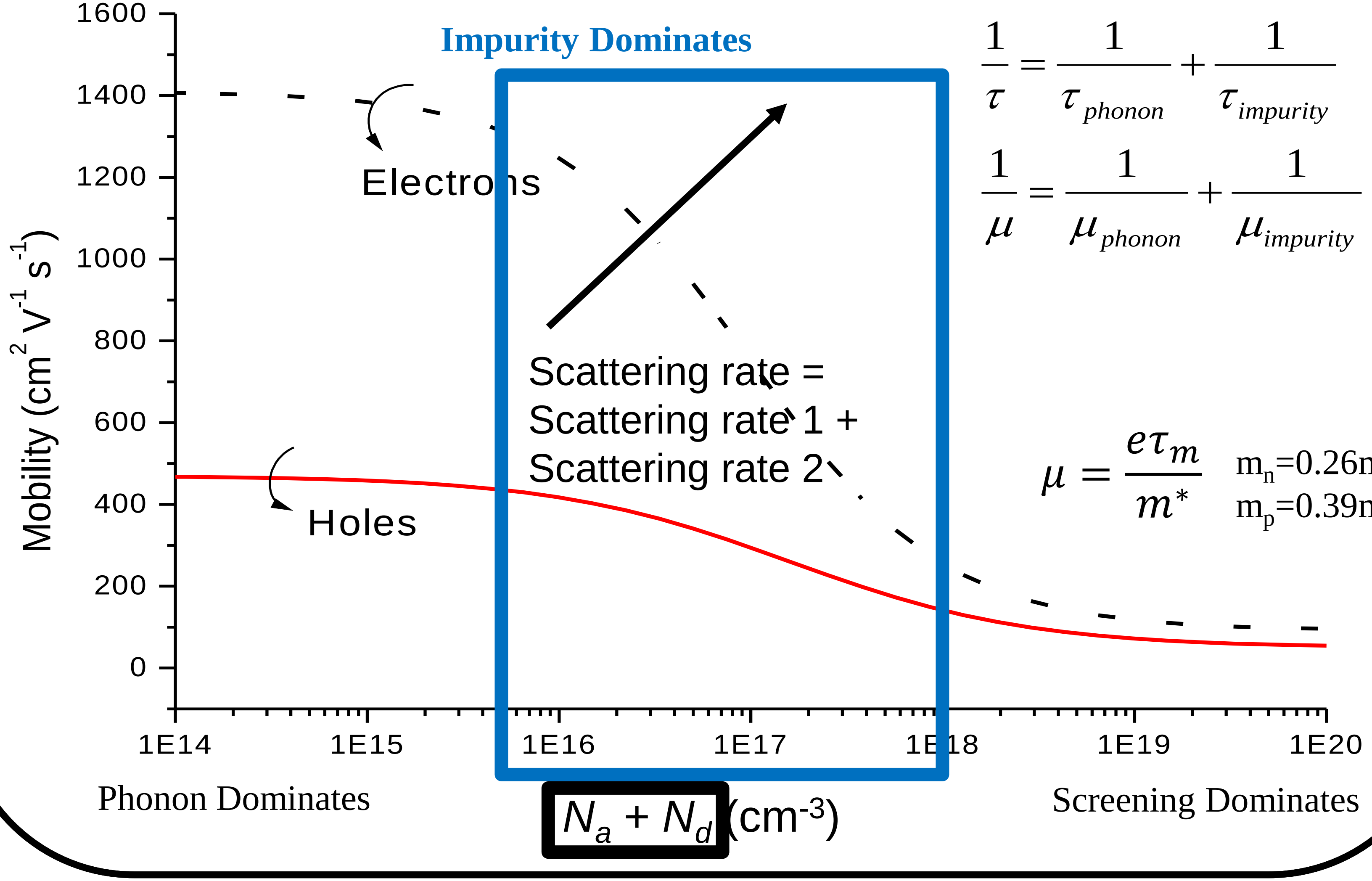
There is less change in the direction of travel if the electron zips by the ion at a higher speed.

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

Same scattering rate

Total Mobility

Bulk mobility: no surface roughness scattering



Impurity Dominates

Scattering rate =
Scattering rate 1 +
Scattering rate 2

Phonon Dominates

Screening Dominates

Matthiessen's rule

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

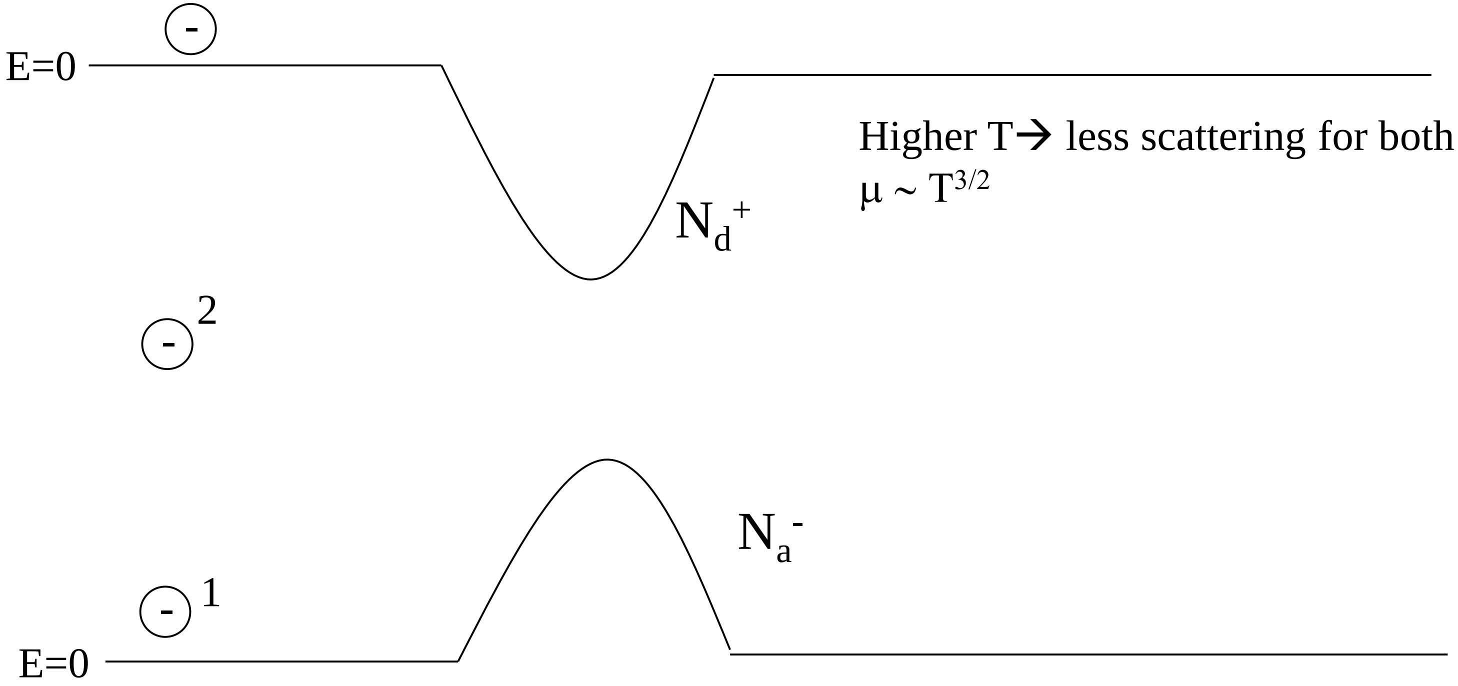
$$\frac{1}{\mu} = \frac{1}{\mu_{phonon}} + \frac{1}{\mu_{impurity}}$$

$$\mu = \frac{e\tau_m}{m^*}$$

$m_n=0.26m_0$
$m_p=0.39m_0$

$N_d^+ + N_a^-$ Not $N_d^+ - N_a^-$

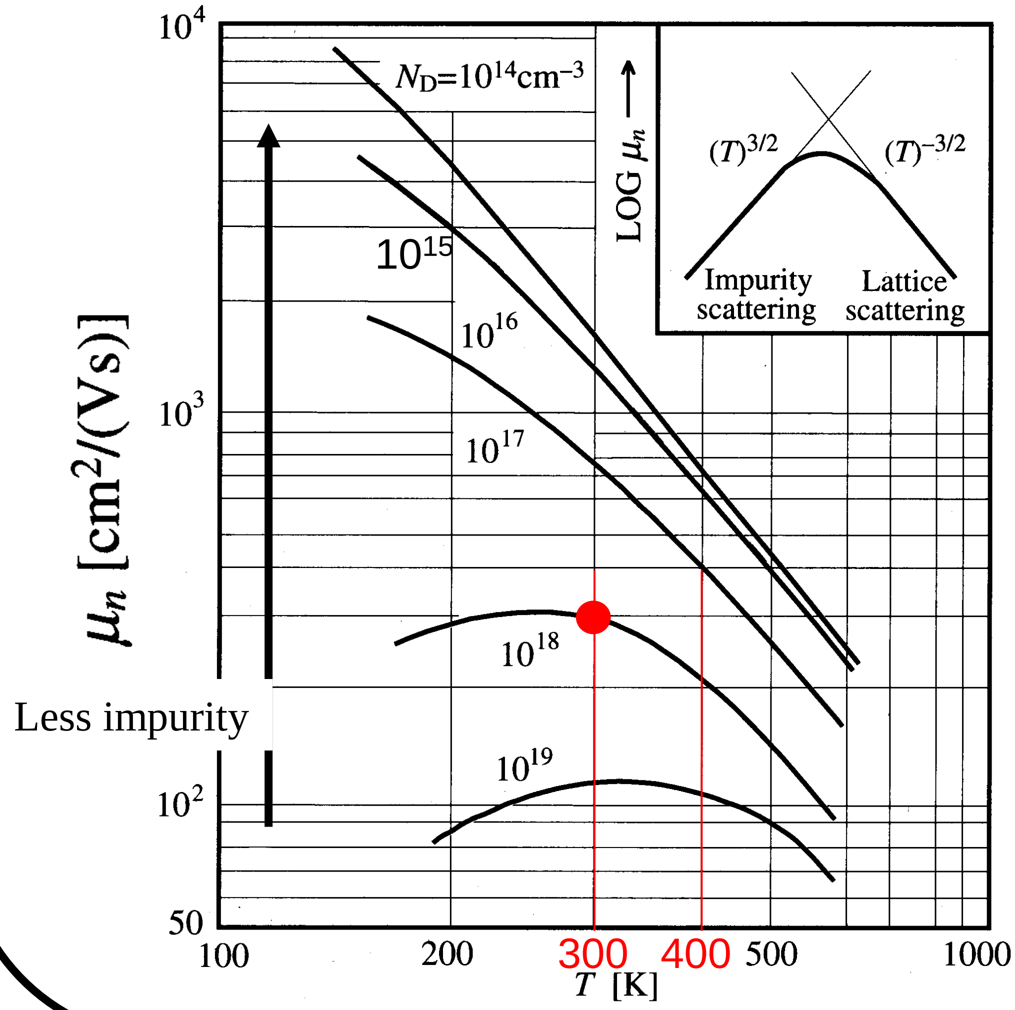
repulsive scattering rate = attractive scattering rate



Higher T→ less scattering for both cases, $\mu \sim T^{3/2}$

Bulk Silicon, not MOSFET (including surface roughness scattering)

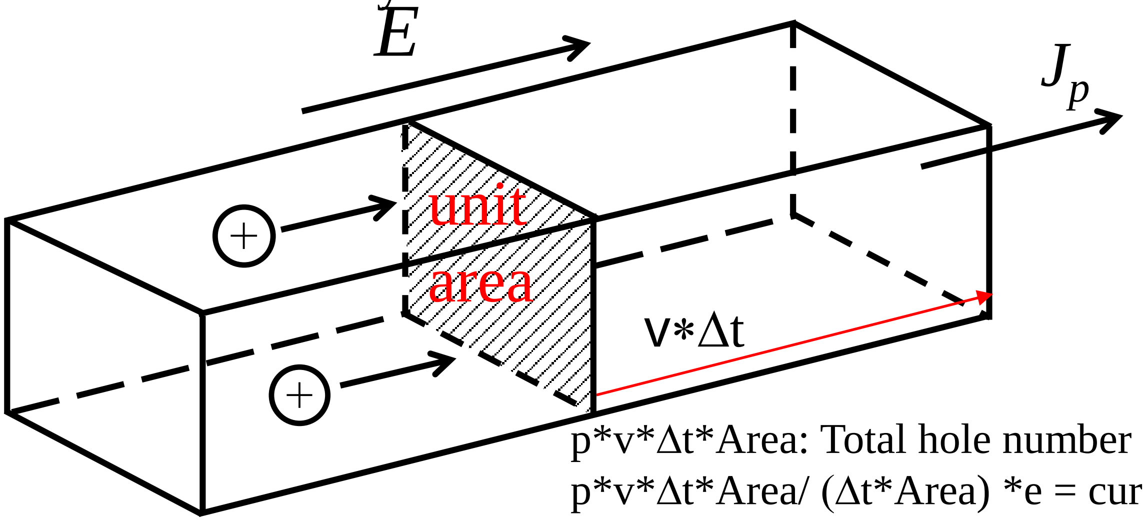
Temperature Effect on Mobility



Practice13:
What Nd will make dμ_n/dT = 0 at room temperature?
Practice14: draw the bulk mobility vs T

2.2.3 Drift Current and Conductivity

Practice15: derive current density formula.



p*v*Δt*Area: Total hole number passing unit area

p*v*Δt*Area/ (Δt*Area) *e = current density

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

$\boxed{J_n = -nev}$

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

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

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

2.2.3 Drift Current and Conductivity

$$J_{n,drift} = -nev = ne\mu_n E \qquad v_n = -\mu_n E$$

$$J_{p,drift} = pev = pe\mu_p E \qquad v_p = \mu_p E$$

$$J_{drift} = J_{n,drift} + J_{p,drift} = \sigma E = (ne\mu_n + pe\mu_p)E$$

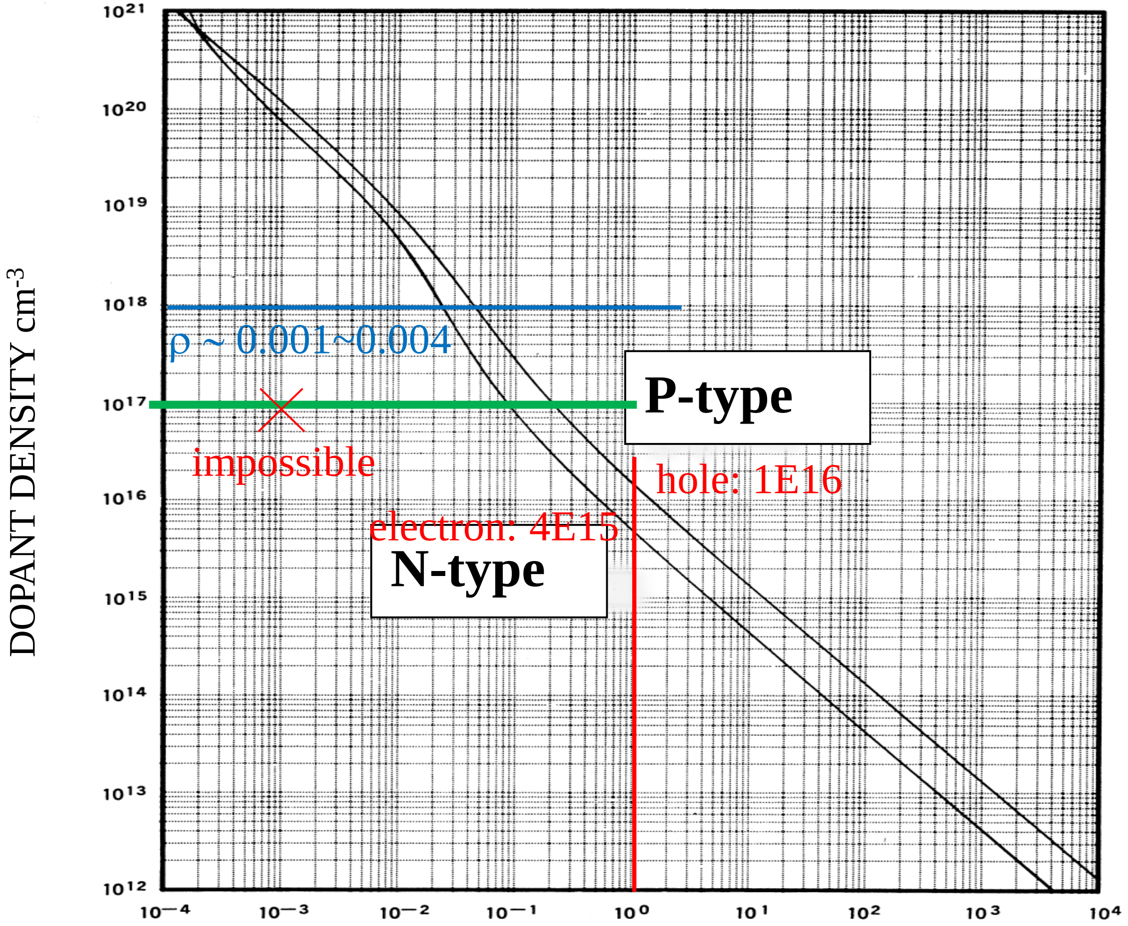
$\therefore$ ***conductivity*** *(1/ohm-cm) of a semiconductor is*

$$\sigma = ne\mu_n + pe\mu_p$$

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

Wafer spec: 1~10 ohm-cm

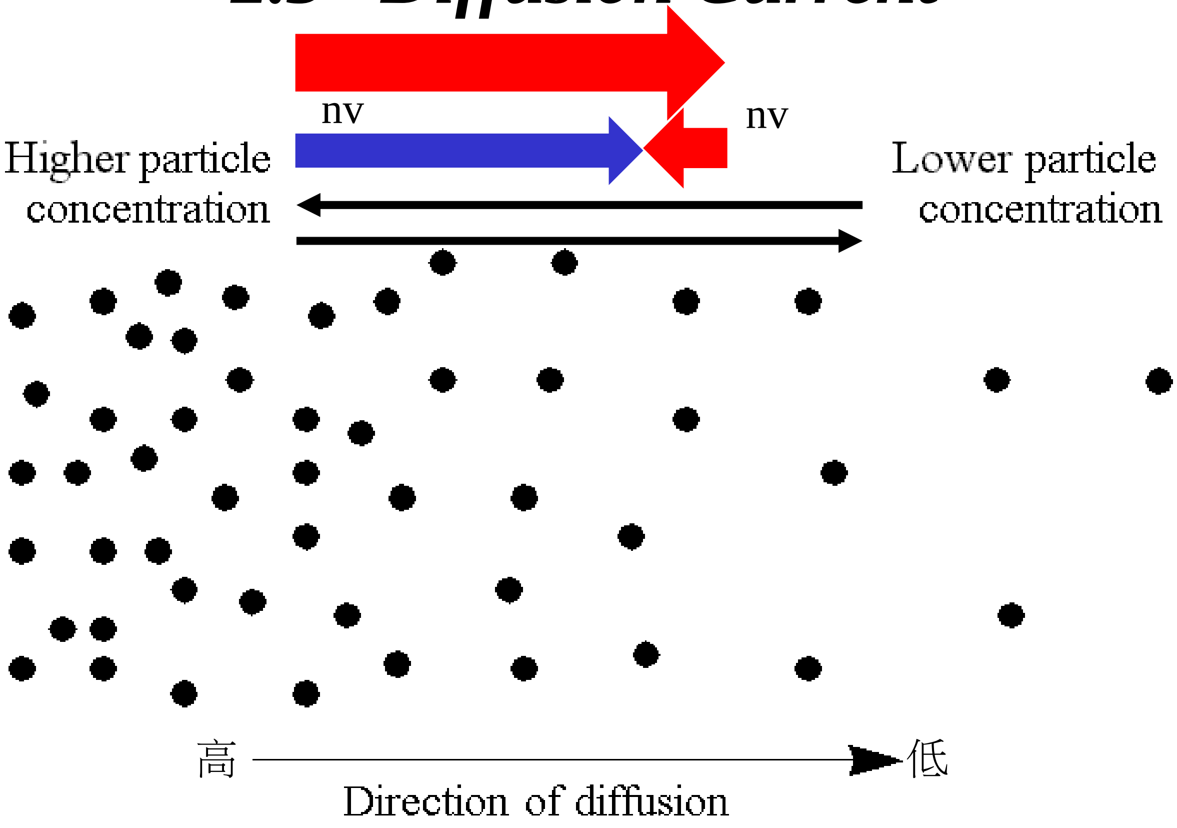
Relationship between Resistivity and Dopant Density



$$\rho = 1/\sigma = 1/ne\mu(N_d+N_a) = \rho(N_d^+=n) = \rho(N_{d=}N_d^+)$$

$$n=N_d=N_d^+$$

2.3 Diffusion Current

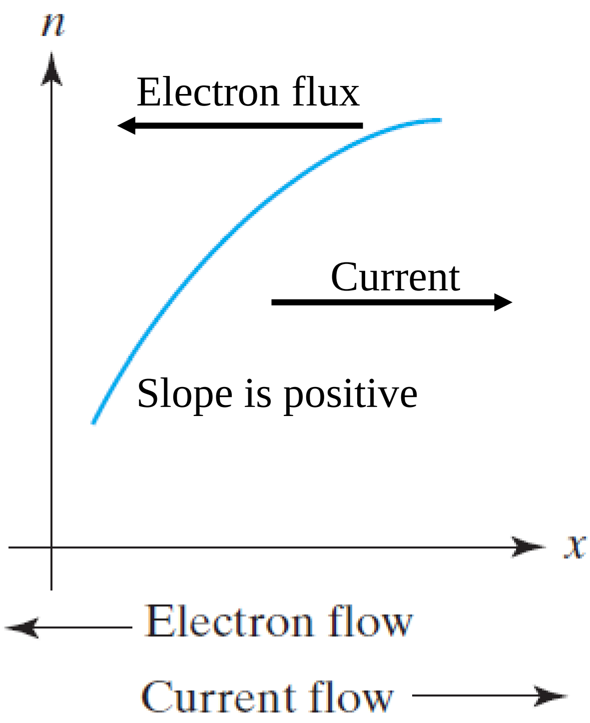


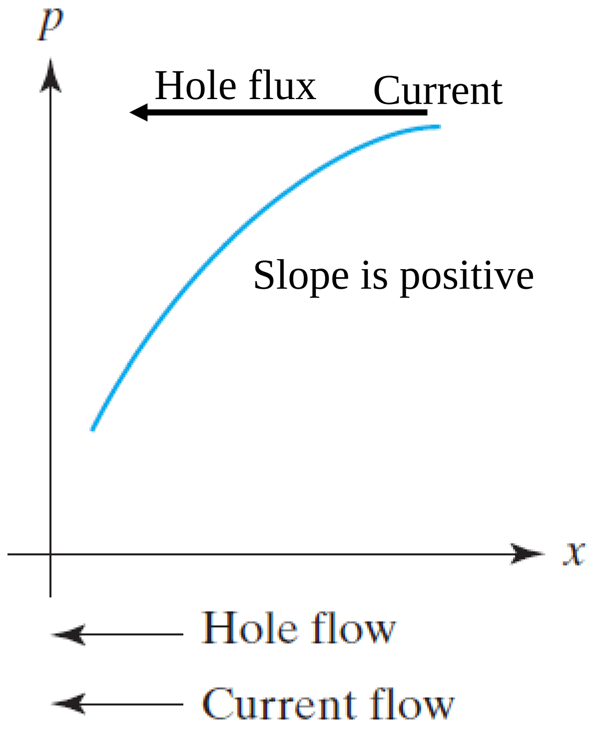
Particles diffuse from a higher-concentration location to a lower-concentration location.

2.3 Diffusion Current

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

D is called the diffusion constant. Signs explained:





Total Current – Review of Four Current Components

$$\boxed{J_{total} = J_n + J_p}$$

$$J_n = J_{n,drift} + J_{n,diffusion} = ne\mu_n E + eD_n \frac{dn}{dx}$$

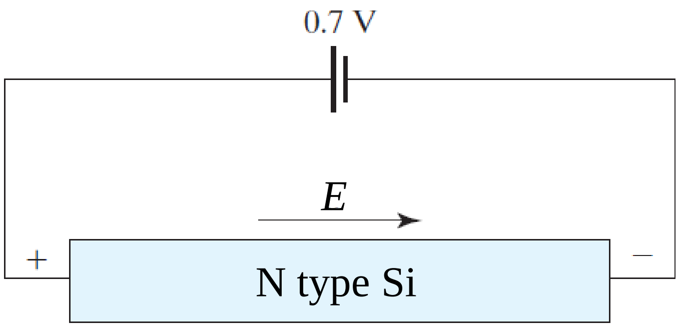
$$J_p = J_{p,drift} + J_{p,diffusion} = pe\mu_p E - eD_p \frac{dp}{dx}$$

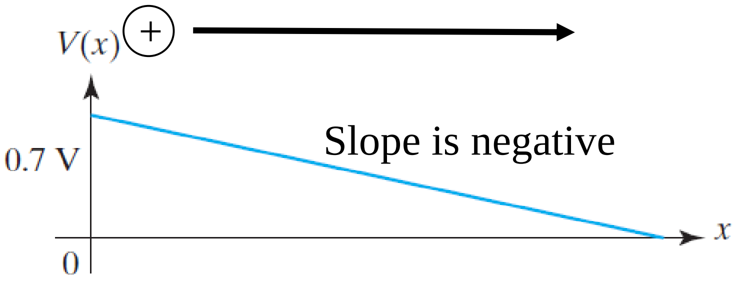
Drift current

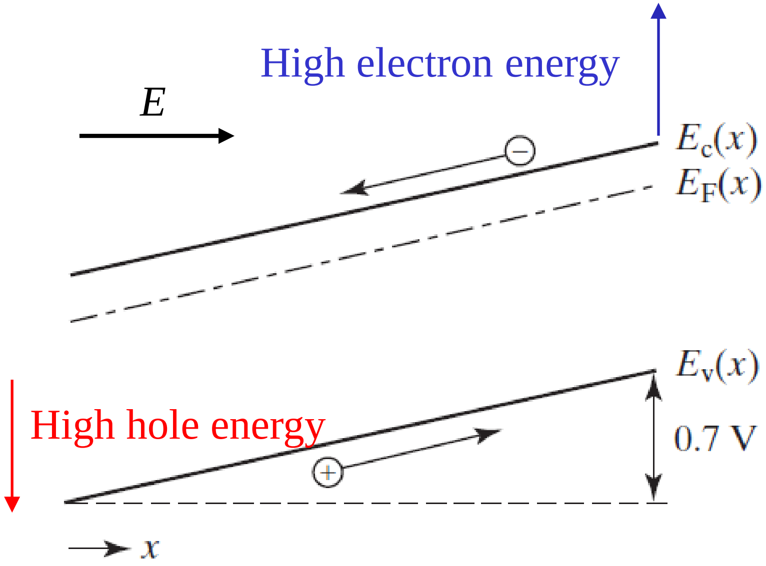
Diffusion current

TCAD: diffusion-drift model

2.4 Relation Between the Energy Diagram and V, E







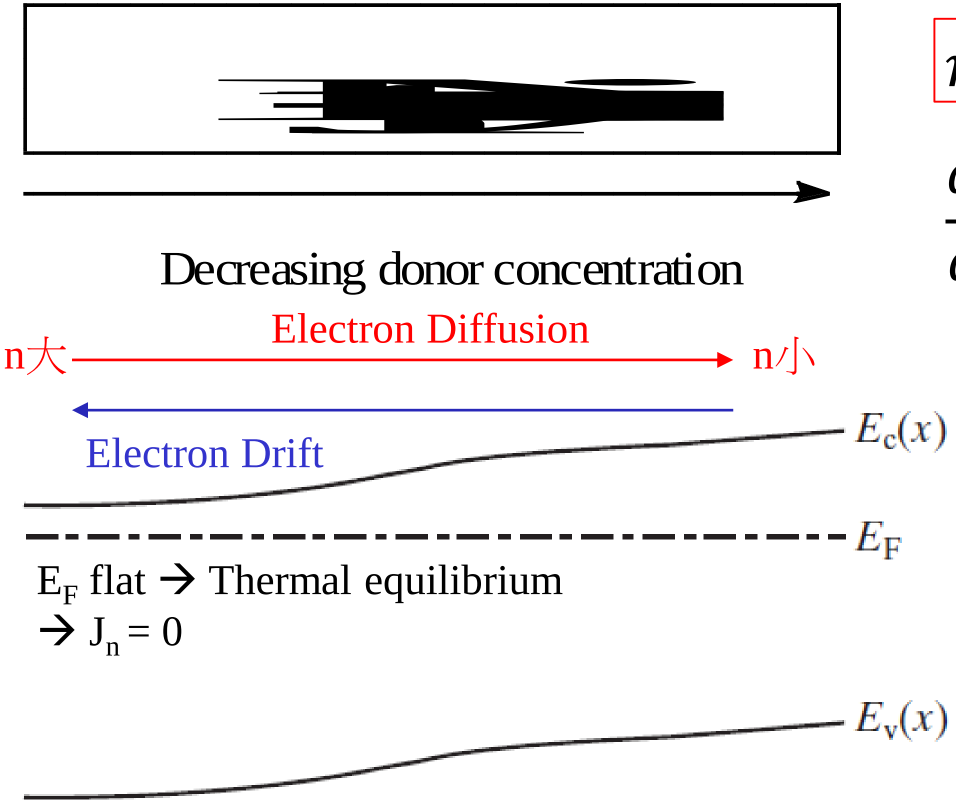
E_c and E_v vary in the opposite direction from the voltage. That is, E_c and E_v are higher where the voltage is lower.

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

Defination

2.5 *Einstein Relationship between D and μ*

Consider a piece of non-uniformly doped semiconductor.



E_F flat → Thermal equilibrium
→ $J_n = 0$

For non-degenerate

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

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

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

$$= -\frac{n}{kT} eE$$

2.5 Einstein Relationship between D and μ

$$\frac{dn}{dx} = -\frac{n}{kT} eE$$

$$J_n = ne\mu_n E + eD_n \frac{dn}{dx} = 0 \; at \; equilibrium.$$

When i=0

$$\rightarrow ne\mu_n E - en\frac{eD_n}{kT} E = 0$$

$$\boxed{D_n = \frac{kT}{e}\mu_n}$$ Similarly, $$\boxed{D_p = \frac{kT}{e}\mu_p}$$ $$\frac{D}{\mu} = \frac{kT}{e}$$

Know μ to get D

These are known as the Einstein relationship.

2.6 Electron-Hole Recombination

- The equilibrium carrier concentrations are denoted with n_0 and p_0 .
- The total electron and hole concentrations can be different from n_0 and p_0 .
- The differences are called the ***excess carrier concentrations*** n' and p'.

$$\boxed{\begin{aligned} n &\equiv n_0 + \boxed{n'} \\ p &\equiv p_0 + \boxed{p'} \end{aligned}}$$

Excess carrier can be generated by electrical field or light

Charge Neutrality

- Charge neutrality is satisfied at equilibrium ($n' = p' = 0$). $N_d^+ + p_0 = N_a^- + n_0$
- When a non-zero n' is present, an equal p' may be assumed to be present to maintain charge equality and vice-versa. $N_d^+ + p_0 + p' = N_a^- + n_0 + n'$
- If charge neutrality is not satisfied, the net charge will attract or repel the (majority) carriers through the drift current until neutrality is restored.

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

But $n_0 \neq p_0$, or $n_0 = p_0 = n_i$

Recombination Lifetime

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

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

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

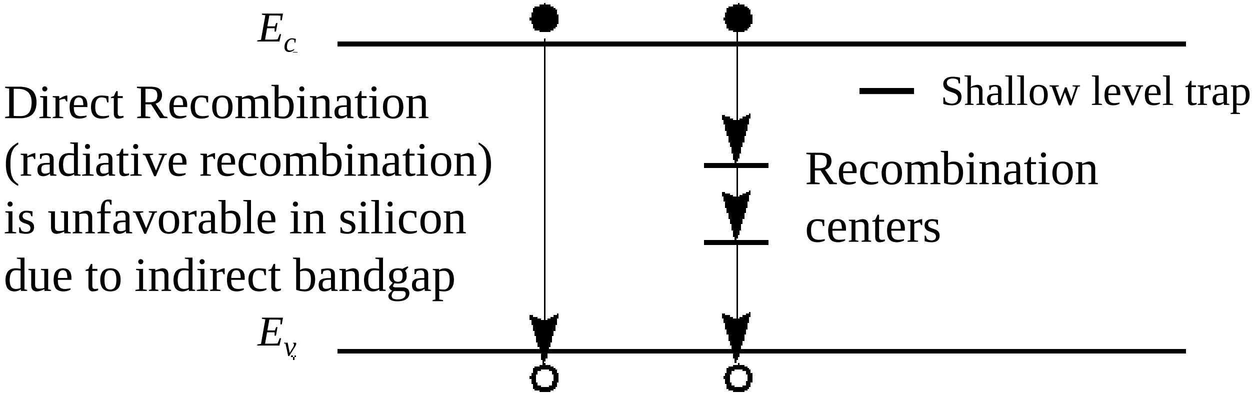
Steady state: dn’/dt =0→n’ =g*τ

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

$$n' = p'$$

Recombination Lifetime

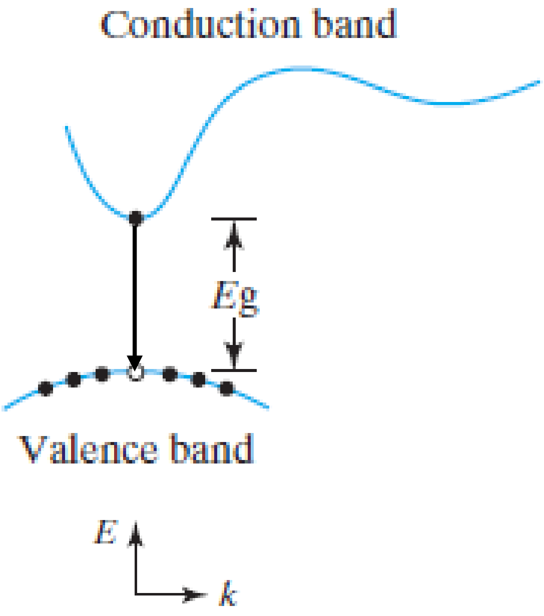
- τ ranges from 1ns to 1ms in Si and depends on the density of metal impurities (contaminants) such as Au and Pt.
- These ***deep traps*** (E_t around midgap) capture electrons and holes to facilitate recombination and are called ***recombination centers.***

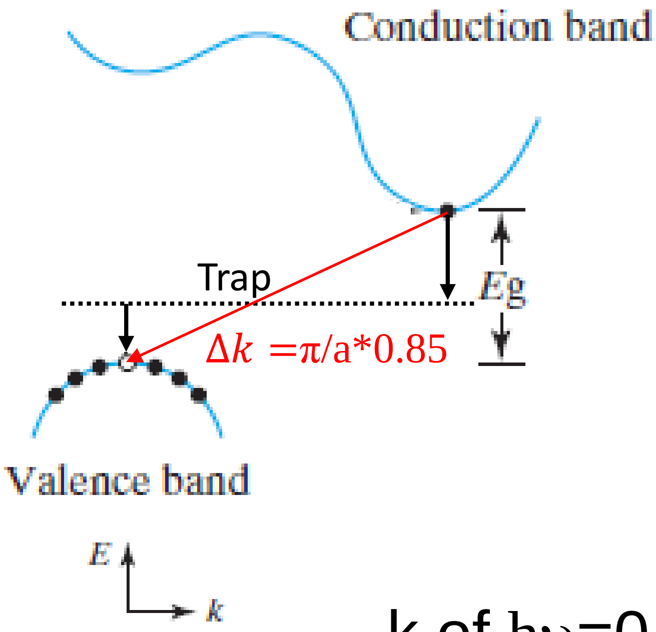


Direct Recombination (radiative recombination) is unfavorable in silicon due to indirect bandgap

Direct and Indirect Band Gap

Practice16: what is k of 1.1eV photon





k of hυ=0

Direct band gap
Example: GaAs

Direct recombination is efficient as k conservation is satisfied.

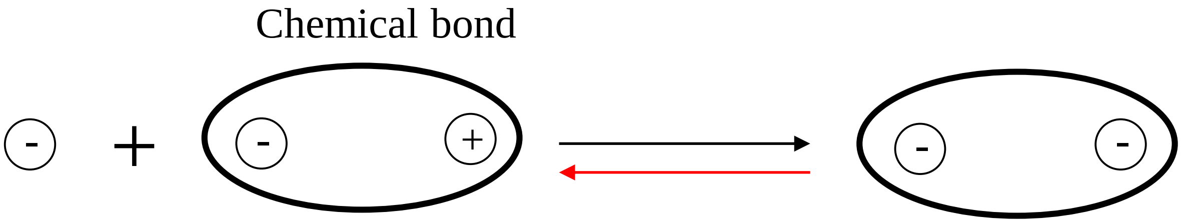
Indirect band gap
Example: Si

Direct recombination is rare as k conservation is not satisfied

2.7 Thermal Generation

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

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



$np > n_i^2$ → recombination

$np < n_i^2$ → generation

2.8 Quasi-equilibrium and Quasi-Fermi Levels

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

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

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

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

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

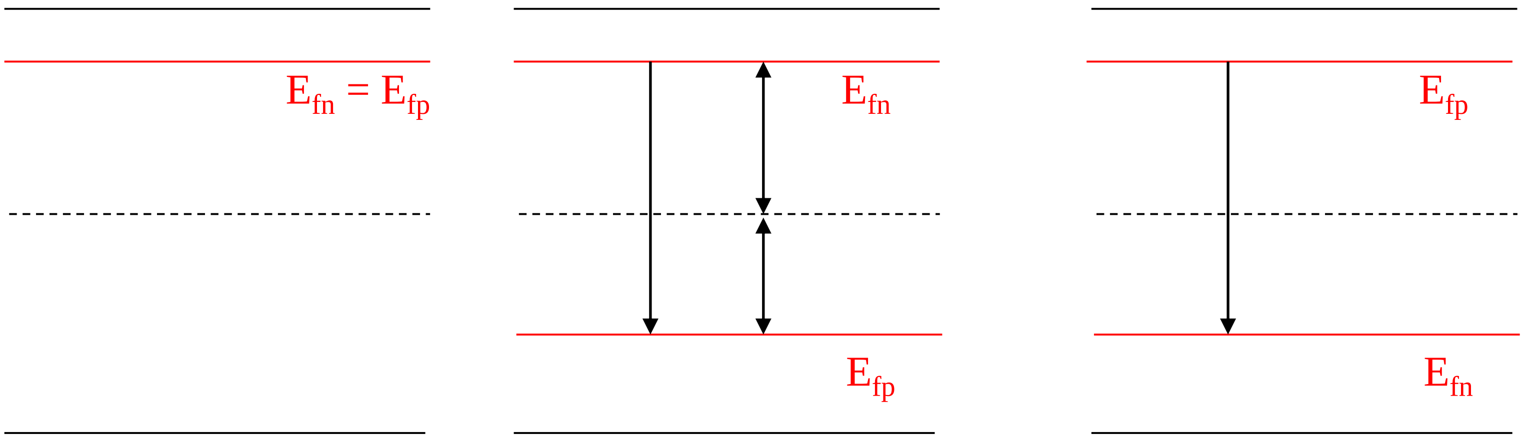
$$p = N_v e^{-(\boxed{E_{fp}} - E_v)/kT}$$

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

Thermal equilibrium

Quasi equilibrium

Quasi equilibrium



$np = n_i^2$

$np > n_i^2$

→recombination

E_{fn} is above E_{fp}

$np < n_i^2$

→generation

E_{fn} is lower than E_{fp}

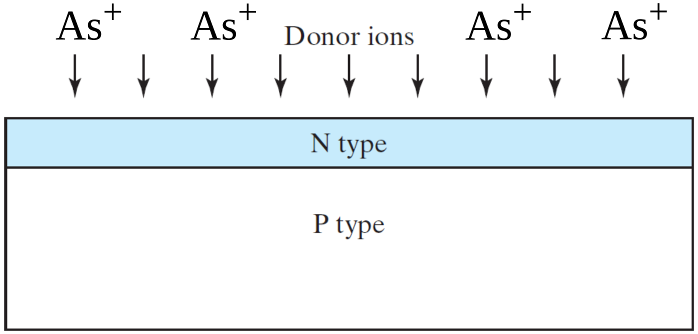
Chapter 4 PN and Metal-Semiconductor Junctions

4.1 Building Blocks of the PN Junction Theory

I/I: ion implantaion

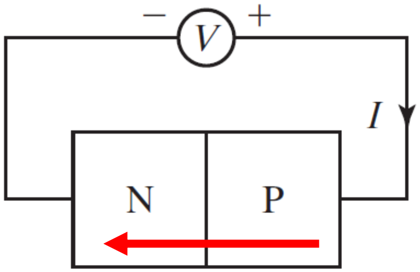
Tens of keV, I controls dose

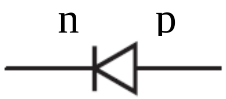
Implant annealing afterward



Now: epi, in-situ doping

Forward bias:



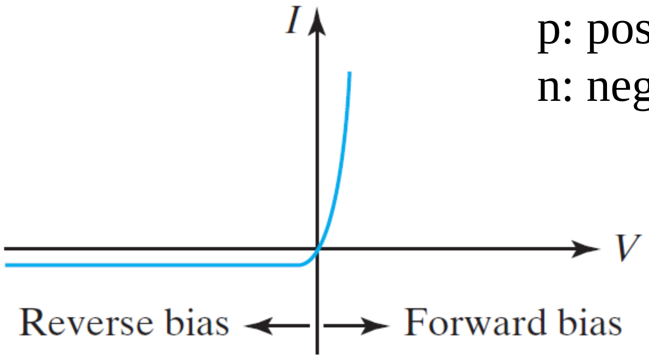


diode symbol

Forward bias:

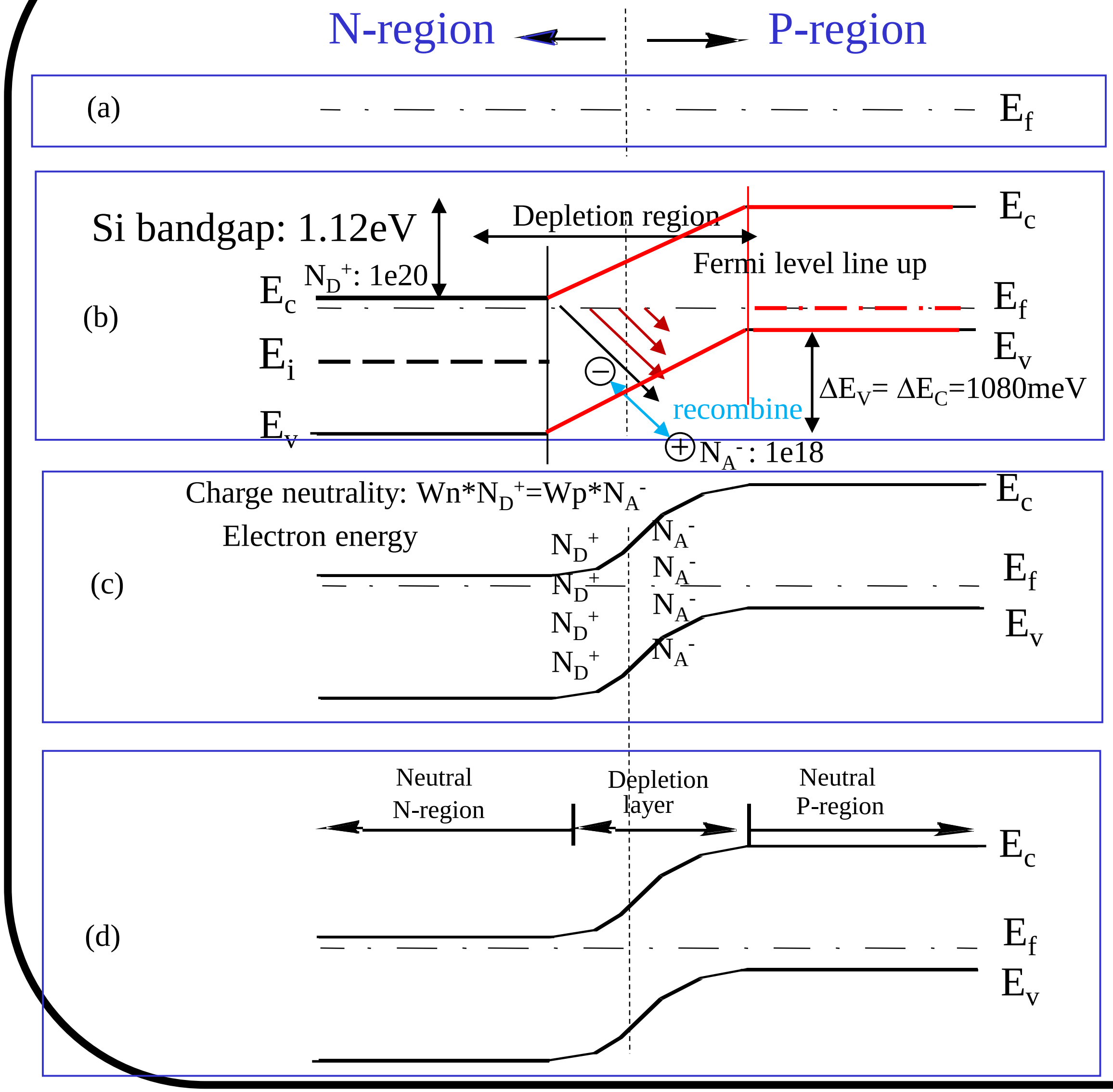
p: positive potential

n: negative potential



PN junction is present in perhaps every semiconductor device.

4.1.1 Energy Band Diagram of a PN Junction



E_f is constant at equilibrium

E_c and E_v are known relative to E_f

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

A depletion layer exists at the PN junction where n ≈ 0 and p ≈ 0.

Not exactly 0, ~1/10 n_0

Practice17: plot band diagram for pn junction

4.1.2 Built-in Potential

N-region

$$n = N_d = N_c e^{-eA/kT} \rightarrow A = \frac{kT}{e} ln \frac{N_c}{N_d}$$

P-region

$$n = \frac{n_i^2}{N_a} = N_c e^{-eB/kT} \rightarrow B = \frac{kT}{e} ln \frac{N_c N_a}{n_i^2}$$

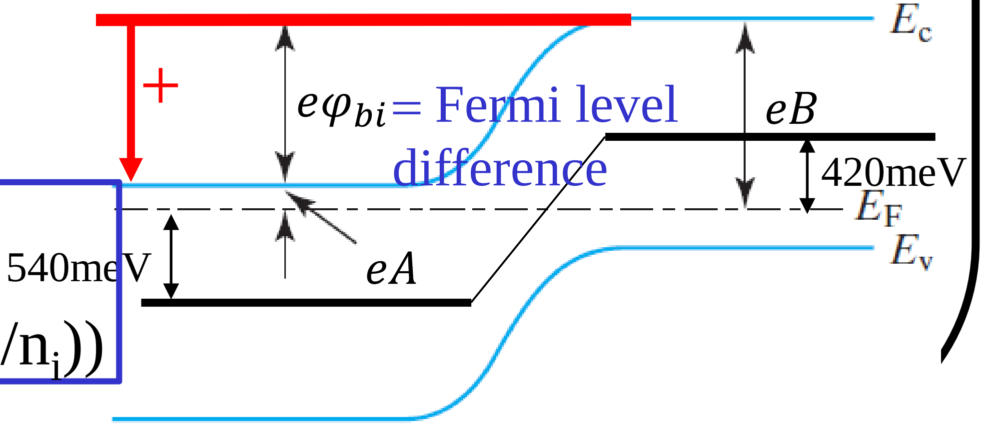
$$\phi_{bi} = B - A = \frac{kT}{e}(ln\frac{N_c N_a}{n_i^2} - ln\frac{N_c}{N_d})$$

Example: n=10^{19}, p=10^{17}, find ϕ_{bi}

A: $(\mathbf{9+7}) \cdot \mathbf{60}\boldsymbol{mV} = \mathbf{960}\boldsymbol{mV}$

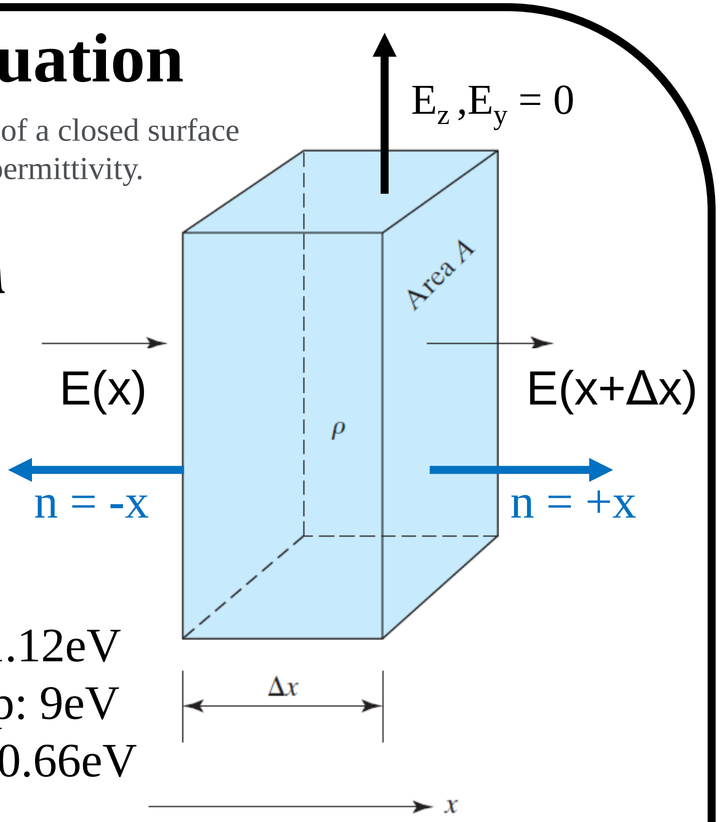
$$\phi_{bi} = \frac{kT}{e}(ln\frac{N_a}{n_i} + ln\frac{N_d}{n_i})$$

=60mV*(log(N_a/n_i)+log(N_d/n_i))



4.1.3 Poisson’s Equation

Gauss’s Law: The total of the electric flux out of a closed surface is equal to the charge enclosed divided by the permittivity.



Gauss’s Law:

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

(dielectric constant)

ε_s(k): permittivity (k~11.9ε_o for Si)

ρ: charge density (C/cm^3)

Si: 11.9ε_o,
SiO2: 3.9ε_o
Ge: 16ε_o
Si bandgap: 1.12eV
SiO2 bandgap: 9eV
Ge bandgap: 0.66eV

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

$$\rightarrow \frac{dE}{dx} = \frac{\rho}{\varepsilon_s} \qquad \boxed{\frac{dV}{dx} = -E}$$

$$\boxed{\frac{d^2V}{dx^2} = -\frac{dE}{dx} = -\frac{\rho}{\varepsilon_s}}$$

Poisson’s equation

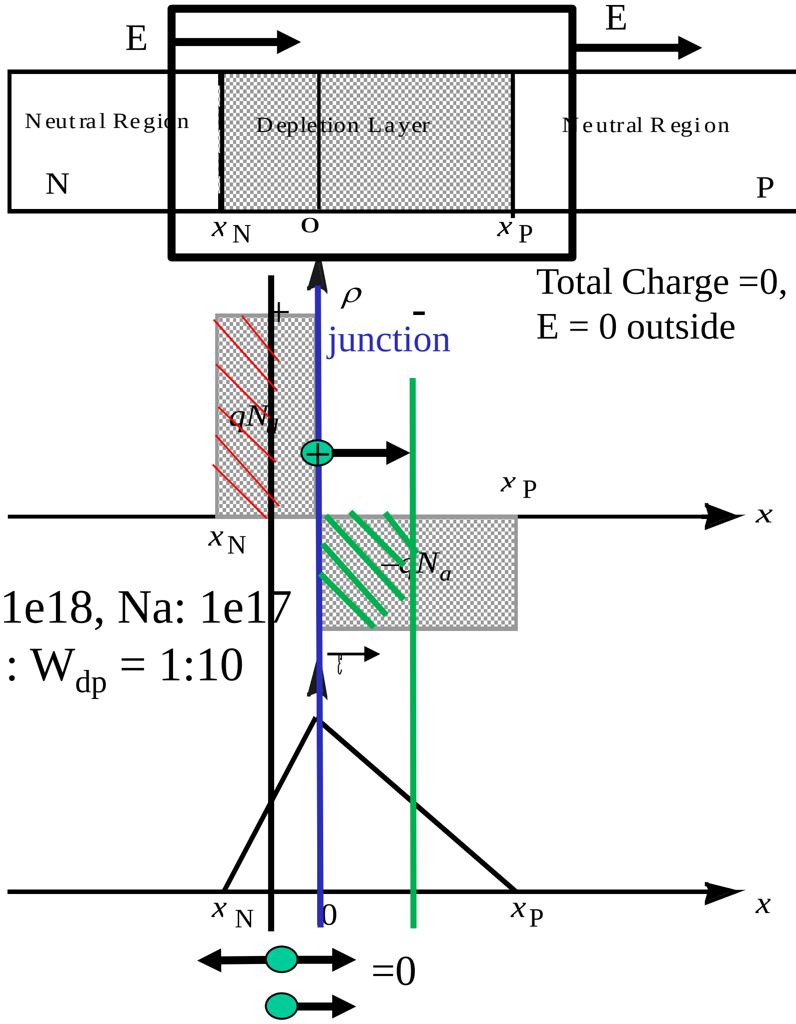
TSMC 5nm
pFET channel: SiGe (guessed)
nFET channel: Si

k: SiGe > Si

Practice18: derive Poisson’s equation

4.2 Depletion-Region Model

4.2.1 Field and Potential in the Depletion region



Total Charge =0,
E = 0 outside

Nd: 1e18, Na: 1e17
W_{dn} : W_{dp} = 1:10

On the ***P-side*** of the depletion region, $\rho = -eN_a$

$$\frac{dE}{dx} = \frac{\rho}{\varepsilon_s} = -\frac{qN_a}{\varepsilon_s}$$

$E_{p,crit}=qN_a/\varepsilon * x_p$

$$\boldsymbol{E = -\int \frac{\rho}{\varepsilon_s} dx}$$

$$E(x) = -\frac{eN_a}{\varepsilon_s}x + C = \boxed{\frac{eN_a}{\varepsilon_s}(x_P - x)}$$

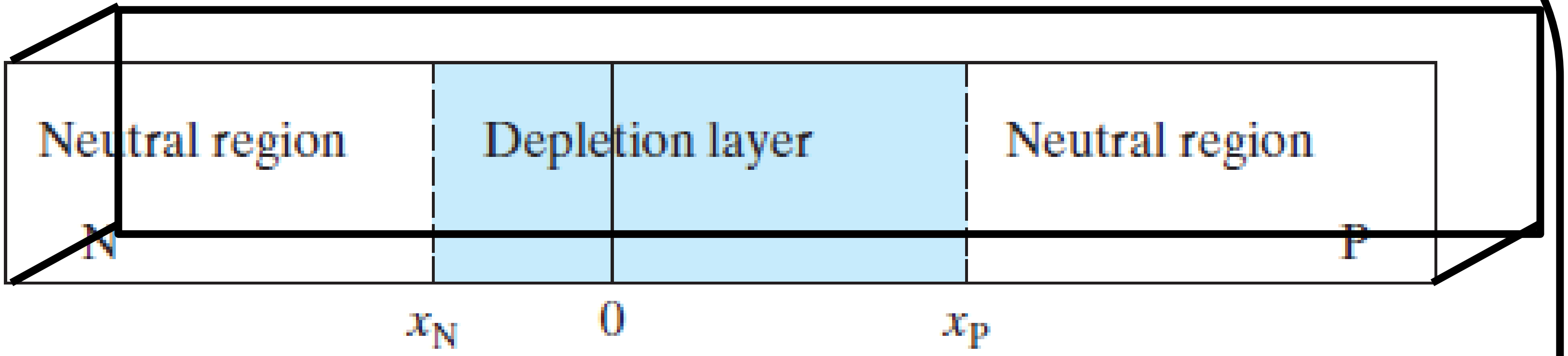
On the ***N-side***, $\rho = eN_d$

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

$$\boldsymbol{E_{max} = E(0) = \frac{N_d^+ W_n}{\varepsilon} = \frac{N_a^- W_p}{\varepsilon}}$$

Practice19: draw the electric field and potential vs position for PN junction

4.2.1 Field and Potential in the Depletion Layer



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

Charge neutrality

unit: $cm^{-3} * cm = cm^{-2}$

$$\boxed{N_a^- |x_P| = N_d^+ |x_N|}$$

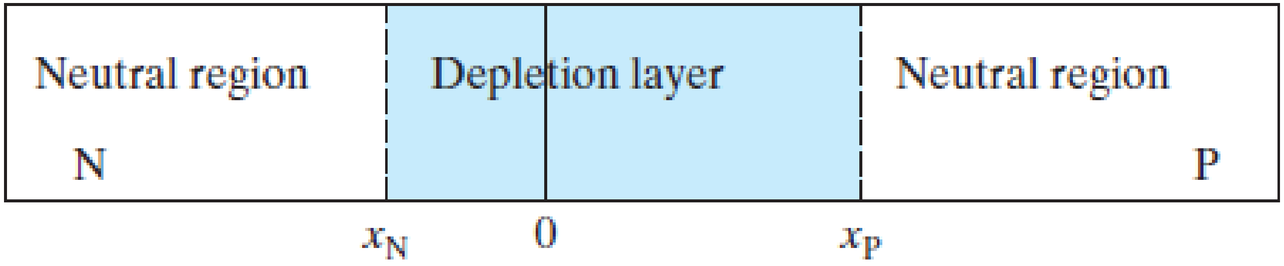
Which side of the junction is depleted more?

One sided abrupt junction ：N_d^+ / N_a^- **or** $N_a^- / N_d^+ \geq \mathbf{10}$

Depletion region in lower doping side

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

4.2.2 Depletion-Layer Width



V is continuous at x = 0

→ $x_P - x_N = W_{dep} = \sqrt{\frac{2\varepsilon_s\phi_{bi}}{e}\left(\frac{1}{N_a}+\frac{1}{N_d}\right)}$ $\quad V_{bi} = \frac{\boldsymbol{W E_{max}}}{\mathbf{2}} = \frac{\boldsymbol{qNW^2}}{\boldsymbol{2\varepsilon}}$

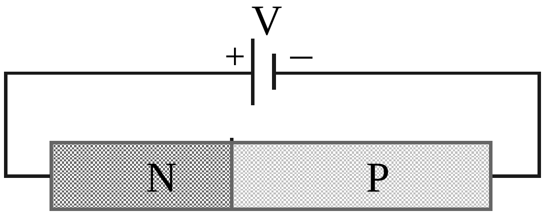
If $N_a >> N_d$, as in a P^+N junction,

$$\boxed{W_{dep} = \sqrt{\frac{2\varepsilon_s\phi_{bi}}{eN_d}} \approx |x_N|} \qquad \boxed{|x_P| = |x_N|\frac{N_d}{N_a} \cong 0}$$

What about a N^+P junction? $\boldsymbol{N_B = \frac{N_d^+ N_a^-}{N_d^+ + N_a^-}}$ Practice20: derive N_B

$$W_{dep} = \sqrt{\frac{2\varepsilon_s\phi_{bi}}{eN_B}} \text{ where } \frac{1}{N_B} = \frac{1}{N_a} + \frac{1}{N_d} \approx \frac{1}{lighter\ dopant\ density}$$

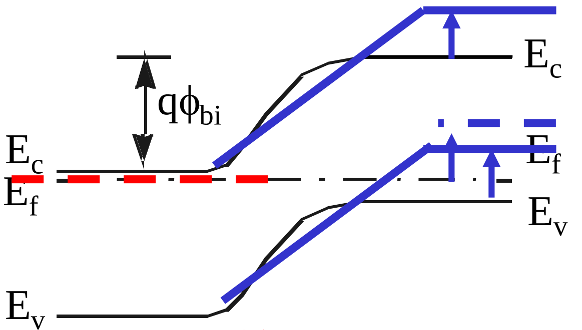
4.3 Reverse-Biased PN Junction



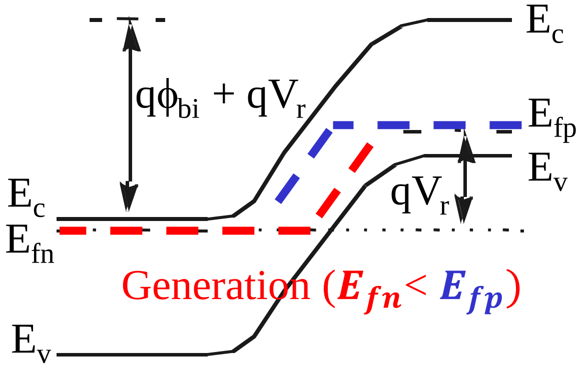
Forward: p connects with +, n connects with –
Reverse: p connects with -, n connects with +

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

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



(a) V = 0



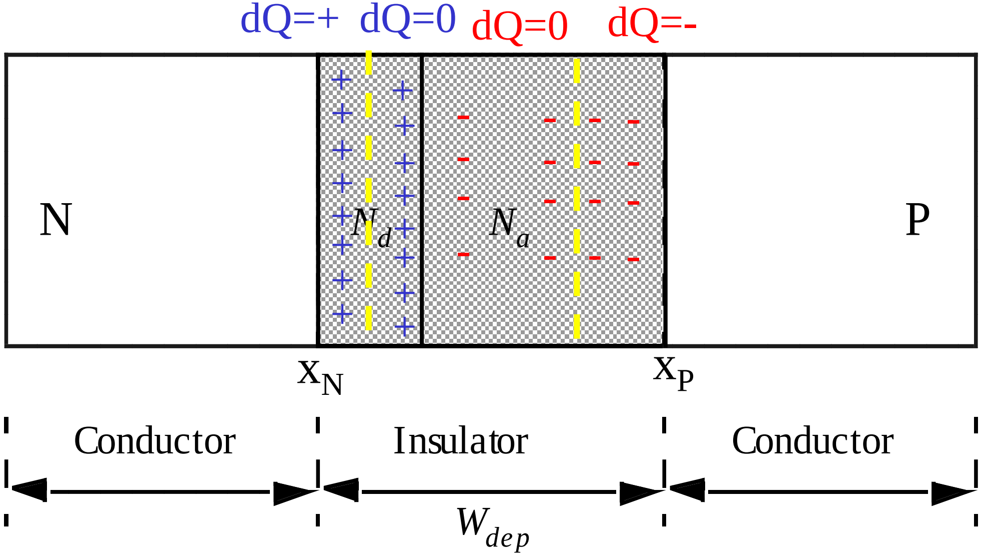
Generation ($E_{fn} < E_{fp}$)

(b) reverse-biased

Q: Does the depletion layer widen or shrink with increasing reverse bias?

A: widen

4.4 Capacitance-Voltage Characteristics



$$\mathbf{W_{dep}} = \sqrt{\frac{2\varepsilon_s(\phi_{bi}+V_{rev})}{eN_B}}$$

Small signal capacitance

Differential charges are on the edge of W_{dep}

Reverse biased PN junction is a capacitor. $$\boxed{C_{dep} = \frac{\varepsilon_s A}{W_{dep}}} = \frac{dQ}{dV}$$

- Is C_{dep} a good thing? Capacitance is controlled by Voltage
- Varactor: voltage-controlled capacitance
- How to minimize junction capacitance?

→ Minimize W_d → Low doping ~ 1e14,15

$$\omega = \frac{1}{\sqrt{LC}}$$

(the principle of adjusting the freq. of radio)

4.4 Capacitance-Voltage Characteristics

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

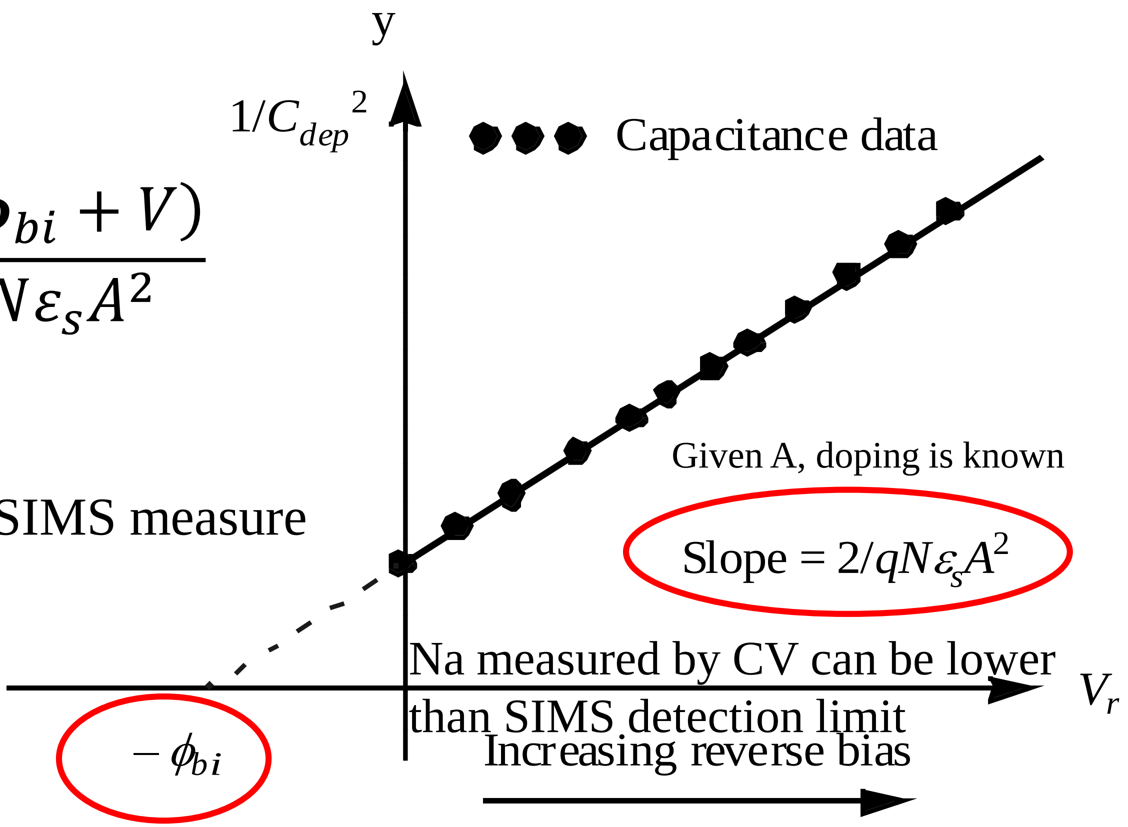
$$\frac{1}{C_{dep}^2} = \frac{W_{dep}^2}{A^2\varepsilon_s^2} = \frac{2(\boldsymbol{\phi}_{bi} + V)}{qN\varepsilon_s A^2}$$

Si atom density: 5e22cm^{-3}

[B]:5e17cm^{-3}→ 10ppm→ SIMS measure

$\Phi_{bi} \leq E_g$

If V > cut-in voltage (0.6V), then the current is significant



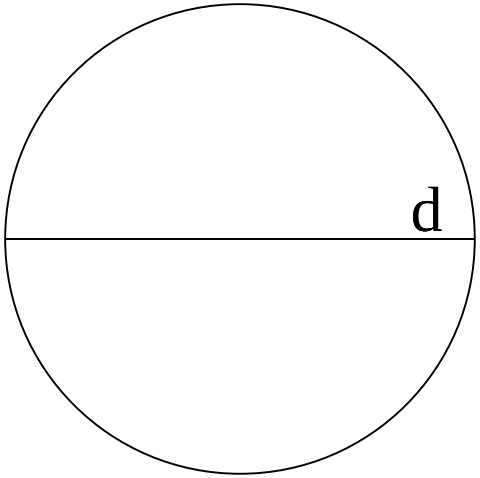
Given A, doping is known

Slope = $2/qN\varepsilon_s A^2$

Na measured by CV can be lower than SIMS detection limit

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

Considering variation in area measurement



$$(\textbf{slope is fixed})$$

$$\boldsymbol{N \times A^2 = const.}$$

$$\boldsymbol{N \times d^4 = const.}$$

$$\boldsymbol{ln(N) + 4ln(d) = const.}$$

$$\boldsymbol{\frac{\Delta N}{N} + 4\frac{\Delta d}{d} = 0}$$

→ length error results in 4 times doping error.

- Large area capacitance is preferred.

$\Delta \boldsymbol{d}$ small, d should be “large” enough for precise measurement.

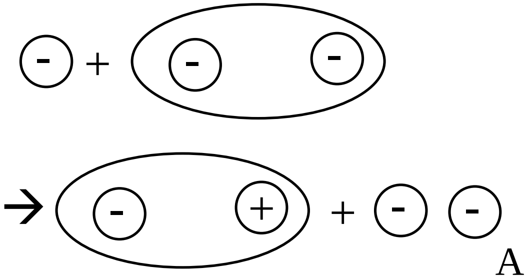
4.5 Junction Breakdown

for core logic V_{dd} for advanced node ≦ 0.7V

→V_{BD} is not important

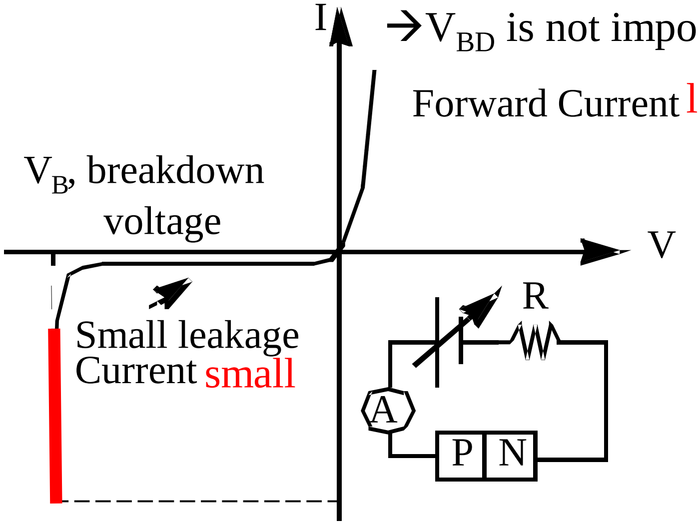
Breakdown:

1. tunneling
2. impact ionization

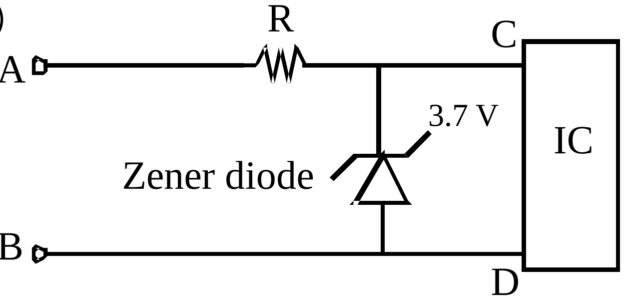


Input electron energy

> 1.5X E_g



Power amplifier needs larger breakdown voltage



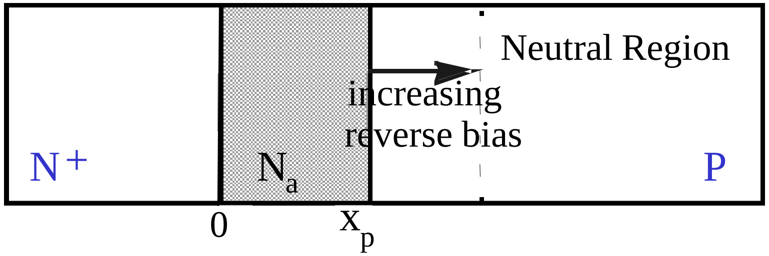
Application: regulator(voltage stabilizer)

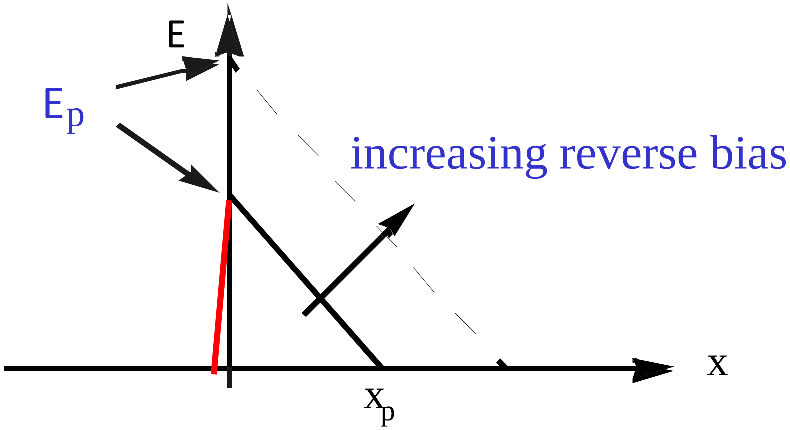
Different current

Same voltage

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

4.5.1 Peak Electric Field





$$V_r + \phi_{bi} = \frac{1}{2} \times x_p \times E_p$$

$$W_d = \sqrt{\frac{2\varepsilon_s(\phi_{bi} + |V_r|)}{qN}}$$

One-sided junction → $x_p \sim W_d$

$$E_p = \frac{2(\phi_{bi} + |V_r|)}{W_d} \quad N_a \uparrow \rightarrow E_p \uparrow$$

$$E_p = E(0) = \left[\frac{2eN_a}{\varepsilon_s}(\phi_{bi} + |V_r|)\right]^{\frac{1}{2}}$$

Assume $E_p = E_{crit}$ at breakdown

$$E_g \uparrow \rightarrow E_{crit} \uparrow \quad E_p = E_{crit} \sim 10^6 V/cm$$

$$V_B = \frac{\varepsilon_s E_{crit}^2}{2eN_a} - \phi_{bi}$$

(3rd generation semiconductor) (impact ionization)

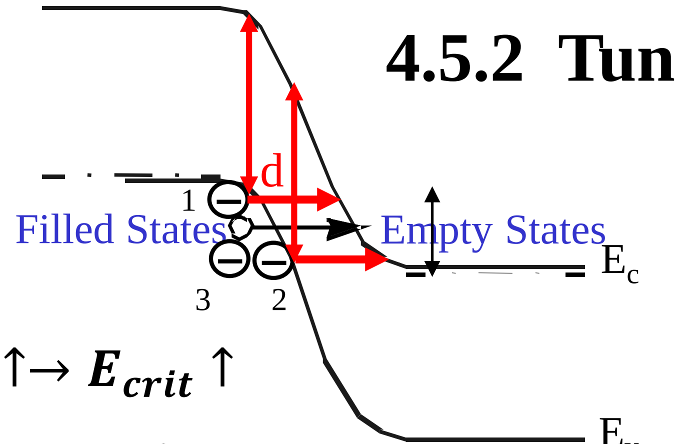
GaN, SiC : E_{crit}= Si 10X

→ V_B = Si 100X

→ High voltage applications

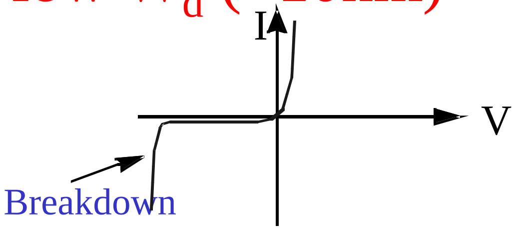
→ However, Si IGBT can still gave high V_B

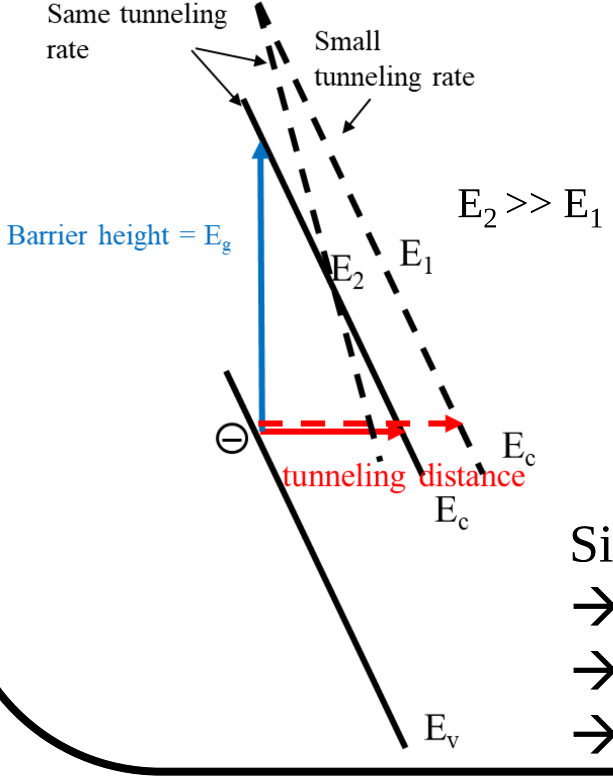
4.5.2 Tunneling Breakdown



$E_g \uparrow \rightarrow E_{crit} \uparrow$

Dominant if both sides of a junction are very heavily doped.
→ low W_d (<10nm)





$$J = G \times e^{-H/E_p}$$ (empirical formula)

WKB can provide better calculation

H is material dependent $H_{SiC} > H_{Si}$

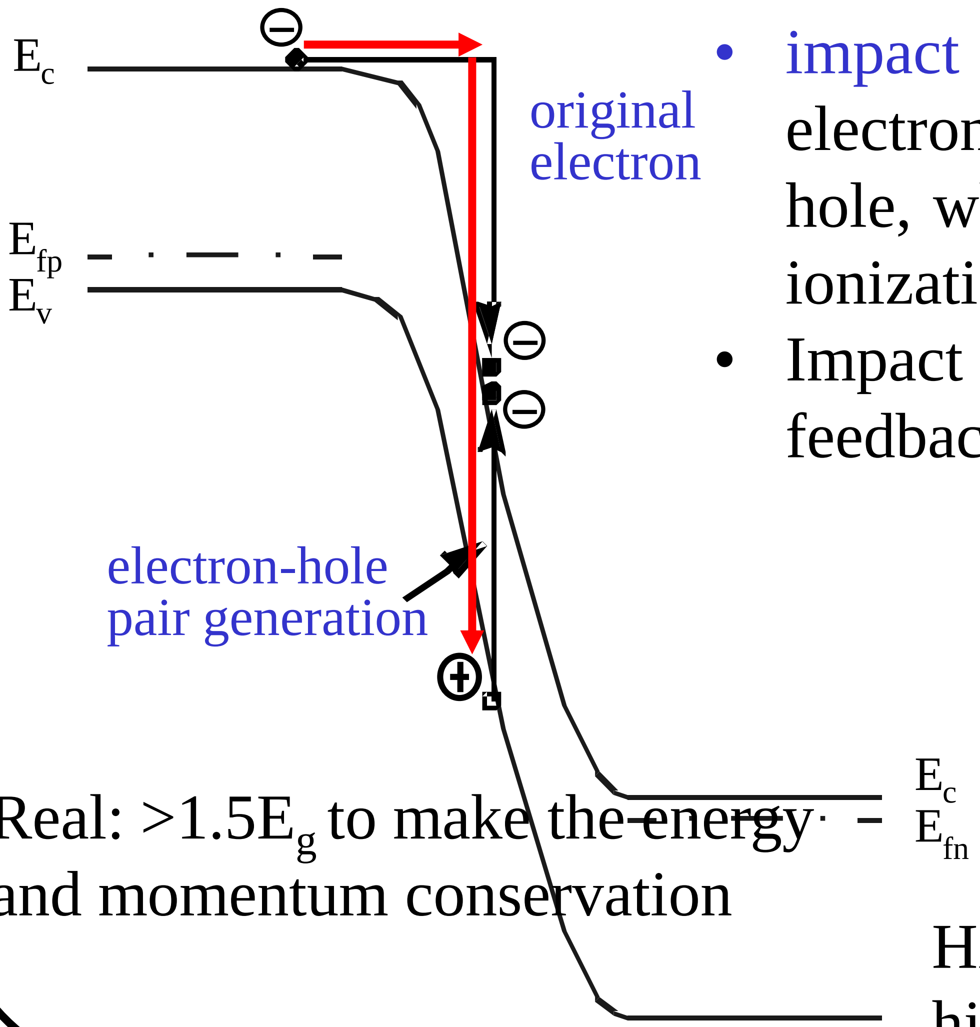
Practice21: explain why we need large enough reverse bias to have tunneling current?

SiC/GaN have large bandgap
→ Large barrier height
→ Large tunneling distance
→ Small tunneling rate

Tunneling distance<10nm

4.5.3 Avalanche Breakdown

electron (primary)→ electron (secondary) + hole



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

$$V_B = \frac{\varepsilon_s E_{crit}^2}{2eN}$$

Real: >1.5E_g to make the energy and momentum conservation

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

Higher bandgap can have higher critical field

4.6 Forward Bias – Carrier Injection (dark)

Practice22: draw reverse bias band diagram and validate drift and diffusion current (bigger or smaller)

V=0 Short p-n

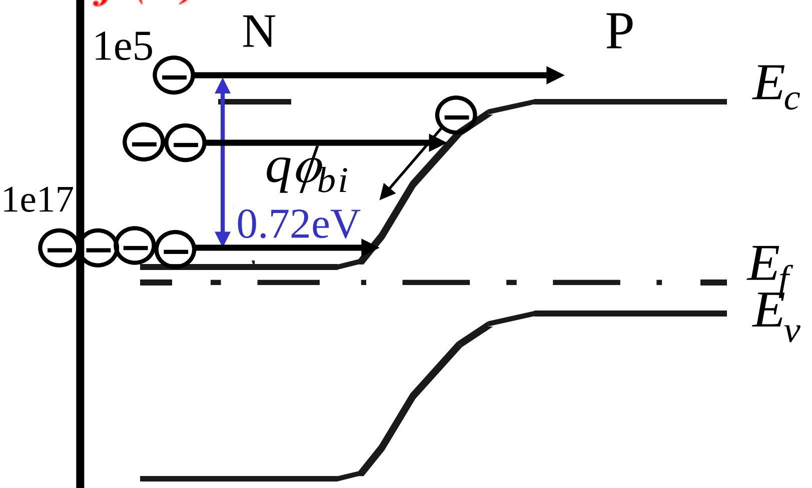
I=0

60meV/dec

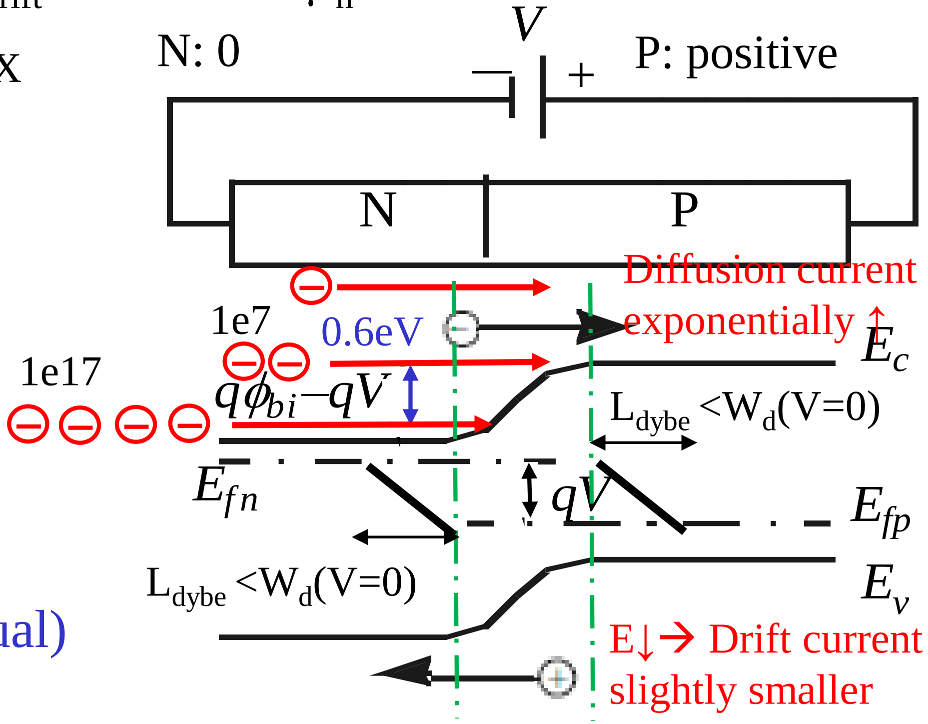
↑60meV→ density 0.1X

$J_{n,drift} = -env = en\mu_n E$

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



Forward biased



Diffusion current exponentially ↑

E↓→ Drift current slightly smaller

Drift and diffusion cancel out (equal)

Electron: Diffusion: electron n→ p

Electron: Drift: electron p→ n

Hole drift and diffusion current also cancel out

Minority carrier injection

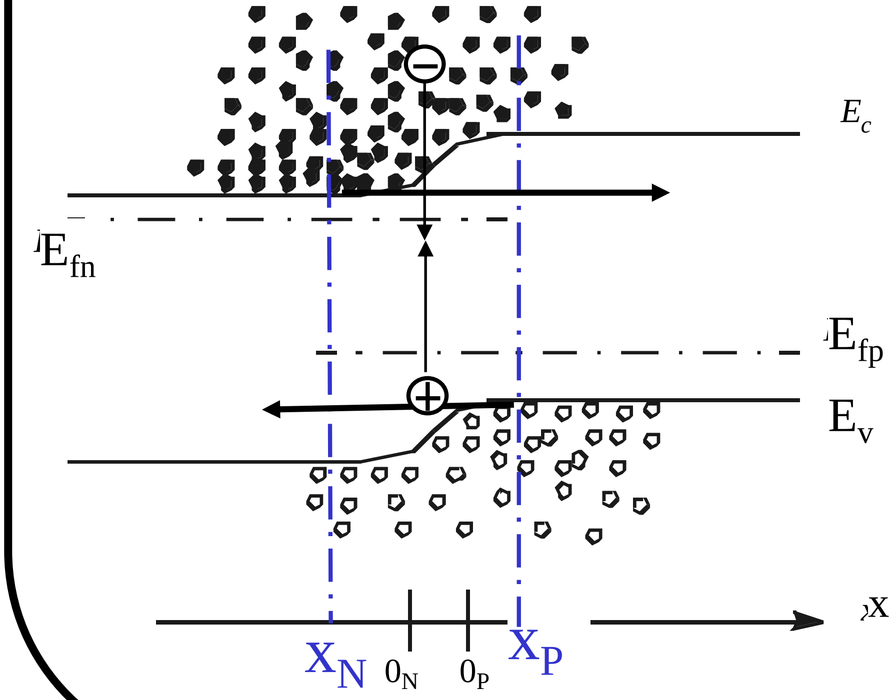
Practice23: draw the hole case

4.6 Forward Bias –Quasi-equilibrium Boundary Condition

Law of junction: Fermi levels are flat in depletion region

Practice24: derive $np = n_i^2 e^{eV/kT}$ by band diagram with law of junction

$p(x_{P+}) = p_0$ $n(x_{P+}) = n_i^2/p_0 * e^{eV/kT} = n_0 e^{eV/kT}$



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

4.6 Forward Bias–Quasi-equilibrium Boundary Condition

Law of junction (low level injection)

$$n(x_P) = n_{P0}e^{eV/kT} = \frac{n_i^2}{N_a}e^{eV/kT}$$ (assume p=N_A n'=p'<N_A)

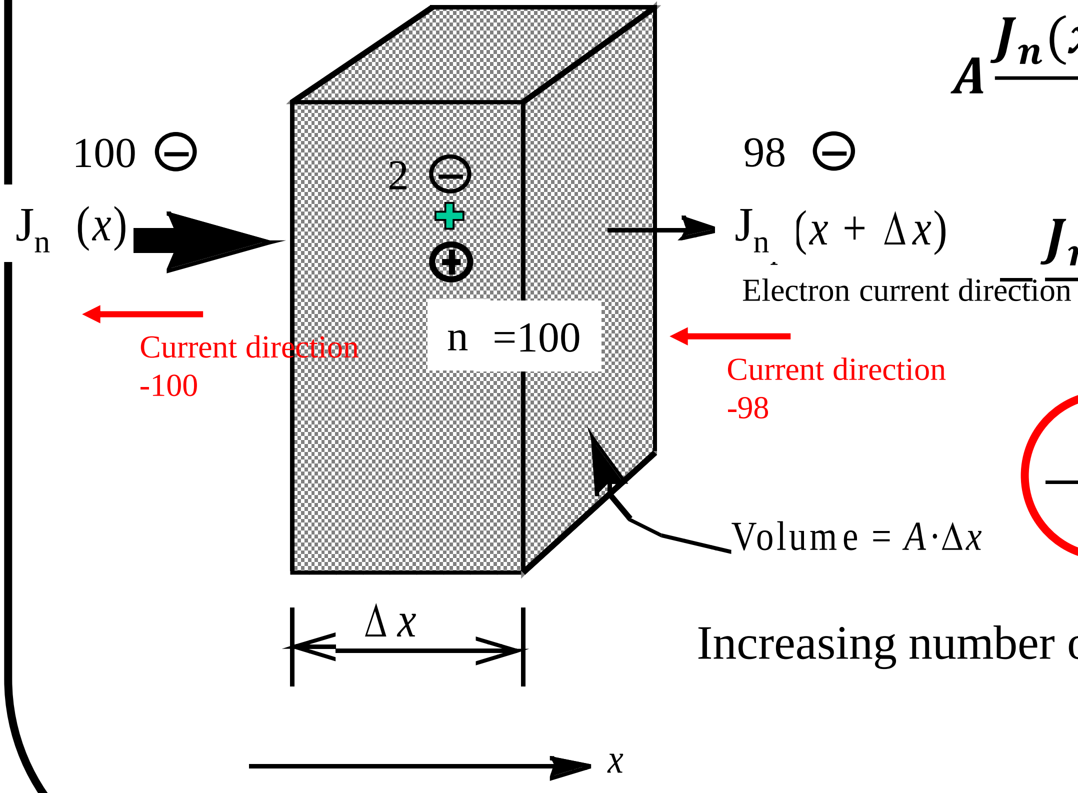
$$p(x_N) = p_{N0}e^{eV/kT} = \frac{n_i^2}{N_d}e^{eV/kT}$$ (assume n=N_d n'=p'<N_d)

excess carrier ($\Delta \boldsymbol{n} = \boldsymbol{n} - \boldsymbol{n_0}, \Delta \boldsymbol{p} = \boldsymbol{p} - \boldsymbol{p_0}$)

$$n'(x_P) \equiv n(x_P) - n_{P0} = n_{P0}\left(e^{eV/kT} - 1\right)$$ P side

$$p'(x_N) \equiv p(x_N) - p_{N0} = p_{N0}\left(e^{eV/kT} - 1\right)$$ N side

4.7 Current Continuity Equation



$$A\frac{J_n(x)}{e} = A\frac{J_n(x+\Delta x)}{e} + A\Delta x\frac{n'}{\tau}$$

$$A\frac{J_n(x+\Delta x)}{e} - A\frac{J_n(x)}{e} = -A\Delta x\frac{n'}{\tau}$$

$$-\frac{J_n(x+\Delta x) - J_n(x)}{\Delta x} = e\frac{n'}{\tau}$$

$$\boxed{-\frac{dJ_n}{dx}} = e\frac{n'}{\tau}$$

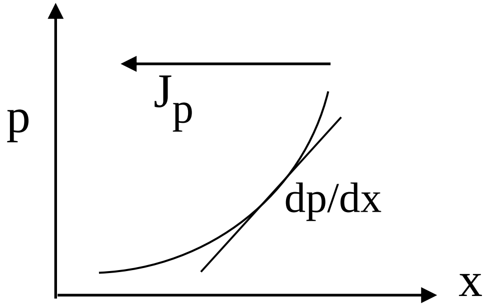
Increasing number of electron per unit time in Δx

4.7 Current Continuity Equation

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

Minority drift current is negligible;

$\therefore J_p = -eD_p dp/dx$



$$eD_p\frac{d^2p}{dx^2} = e\frac{p'}{\tau_p}$$

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

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

L_p and L_n are the diffusion lengths

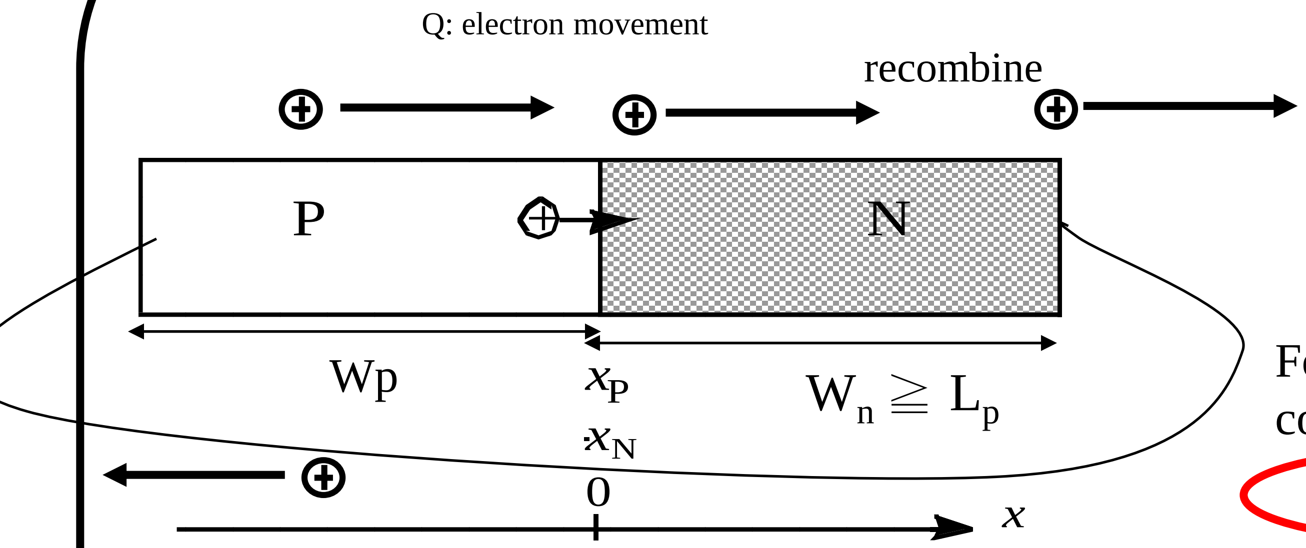
$$\boxed{L_p \equiv \sqrt{D_p\tau_p}}$$

Solar cell needs large life time (ms)

VLSI small life time (μs)

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

4.8 Forward Biased Junction-- Excess Carriers



Practice25: solve the differential equation

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

For solar cell, the boundary condition is different

$$\boxed{p'(\infty) = 0}$$

$$np = n_i^2 e^{eV/kT}$$

$$p'(x_N) = p_{N0}\left(e^{eV/kT} - 1\right)$$

$$p(x_N) = \frac{n_i^2}{N_d} e^{eV/kT}$$

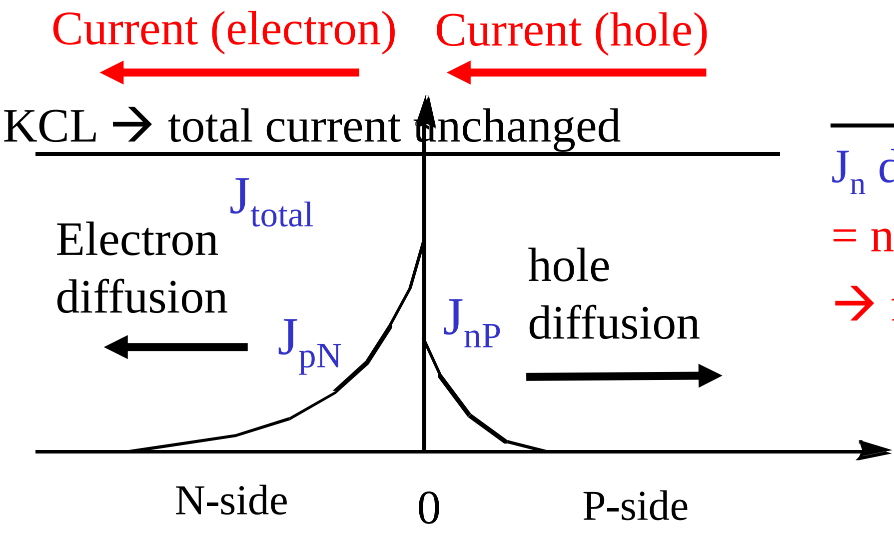
$$p'(x) = Ae^{x/L_p} + Be^{-x/L_p}$$

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

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

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

4.9 PN Diode I-V Characteristics



Current (electron) Current (hole)

KCL → total current unchanged

J_{total}

Electron diffusion

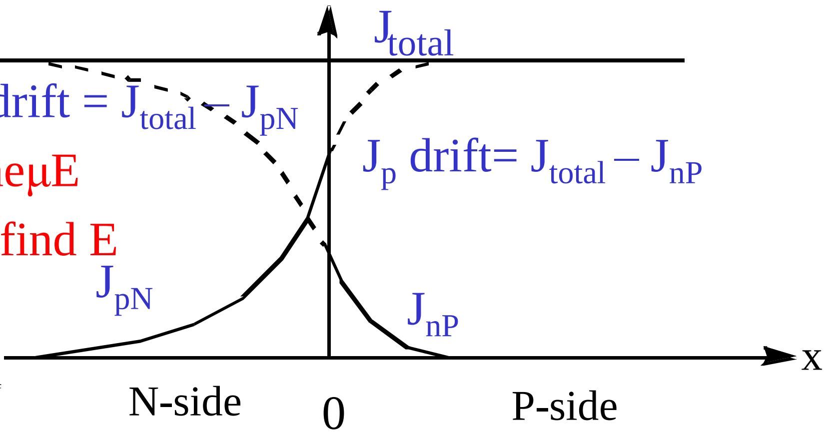
J_{pN}

J_{nP}

hole diffusion

N-side 0 P-side

Negligible for depletion region



J_{total}

J_n drift = $J_{total} - J_{pN}$

= neμE

→ find E

J_p drift= $J_{total} - J_{nP}$

J_{pN}

J_{nP}

x

N-side 0 P-side

x=x_N

$$J_{pN} = -eD_p\frac{dp'(x)}{dx} = \boxed{e\frac{D_p}{L_p}p_{N0}\left(e^{eV/kT}-1\right)}e^{-(x-x_N)/L_p}$$

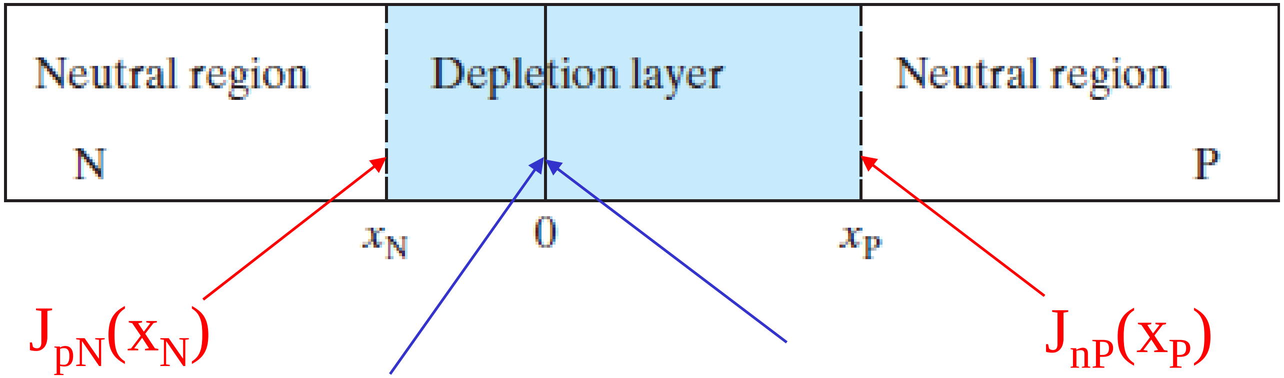
$$J_{nP} = eD_n\frac{dn'(x)}{dx} = e\frac{D_n}{L_n}n_{P0}\left(e^{eV/kT}-1\right)e^{(x-x_P)/L_n}$$

Practice26: Electric field at Neutral region =?

$$J_{total} = J_{pN}(x_N) + J_{nP}(x_P) = \left[e\frac{D_p}{L_p}p_{N0} + e\frac{D_n}{L_n}n_{P0}\right]\left(e^{eV/kT}-1\right)$$

$$= J\ at\ all\ x$$

V=0 → I=0



$J_{pN}(x_N)$

$J_{nP}(x_P)$

$J_{pN}(0)=J_{pN}(x_N)$ $J_{nP}(0)=J_{nP}(x_P)$

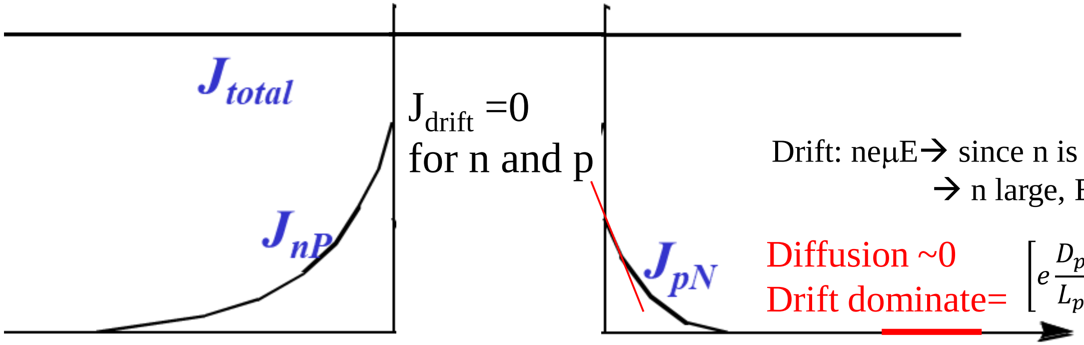
$J_{total}=J_{pN}(x_N)+J_{nP}(x_P)$

- No drift electron/hole current in the depletion region
- No recombination in the depletion region

→ diffusion current is the same inside depletion region

Q: x=-x_p : electron drift and diffusion,
x=-∞: electron drift and diffusion

Neutral Region has sufficient majority carrier has sufficient drift current



J_{drift} =0 for n and p

Drift: neμE→ since n is majority carrier (n-type)
→ n large, E small

Diffusion ~0
Drift dominate

$=\left[e\frac{D_p}{L_p}p_{N0}+e\frac{D_n}{L_n}n_{P0}\right]\left(e^{eV/kT}-1\right)$ **P-side**

-x_p x_N **N-side**

Diffusion ~0
Drift dominate= $\left[e\frac{D_p}{L_p}p_{N0}+e\frac{D_n}{L_n}n_{P0}\right]\left(e^{eV/kT}-1\right)$

$$At\ x=x_N\quad Drift: peV=pe\mu E\sim 0$$

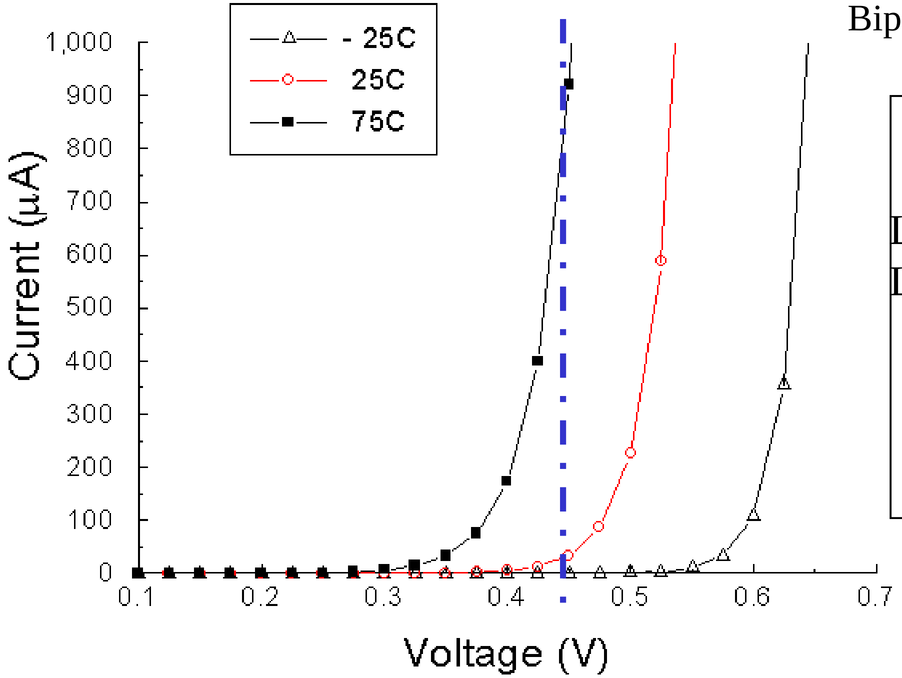
$$J_{pN}=-eD_p\frac{dp'(x)}{dx}=e\frac{D_p}{L_p}p_{N0}\left(e^{eV/kT}-1\right)$$

$$\boxed{\left[e\frac{D_p}{L_p}p_{N0}+e\frac{D_n}{L_n}n_{P0}\right]\left(e^{eV/kT}-1\right)}$$

1kT

The PN Junction as a Temperature Sensor

temperature↑→electrons are more likely jumping from Ev to Ec → current↑

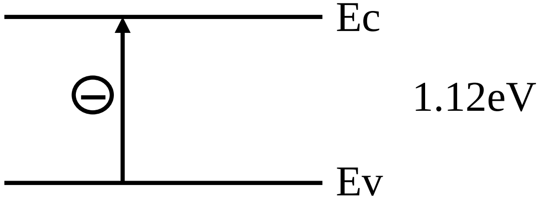


PN junction is not a good temperature sensor
Bipolar is a good temperature sensor

$$I = I_0\left(e^{eV/kT} - 1\right)$$

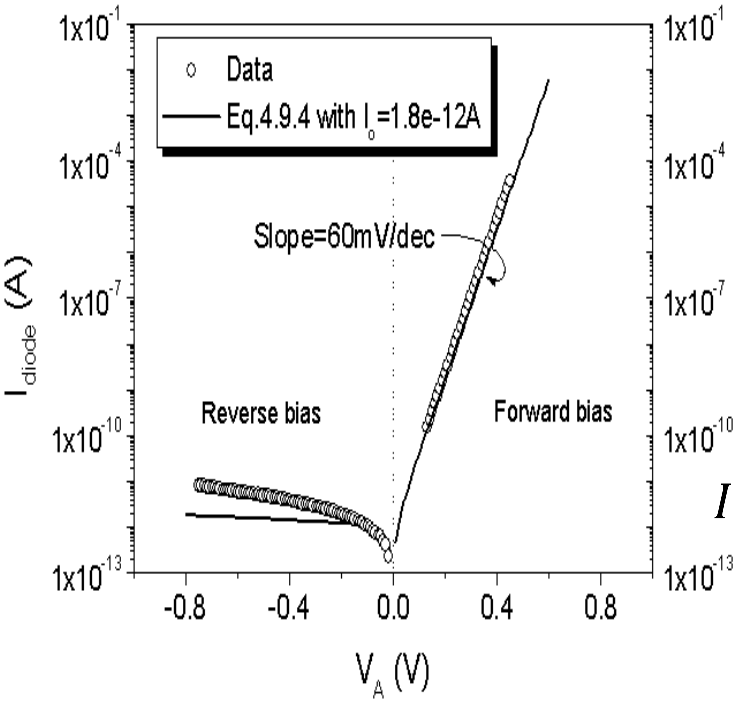
Ln<< Wp
Lp<< Wn

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



Practice27: What causes the IV curves to shift to lower V at higher T ?

4.9.1 Contributions from the Depletion Region



$$\boxed{n{\sim}p{\sim}n_i e^{eV/2kT}}\ ?$$

Net recombination (generation) rate:

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

Assume there are defects in W_{dep}

$$I = \boxed{I_0\left(e^{eV/kT}-1\right)} + \boxed{A\frac{en_iW_{dep}}{\tau_{dep}}\left(e^{eV/2kT}-1\right)}$$

1kT

2kT

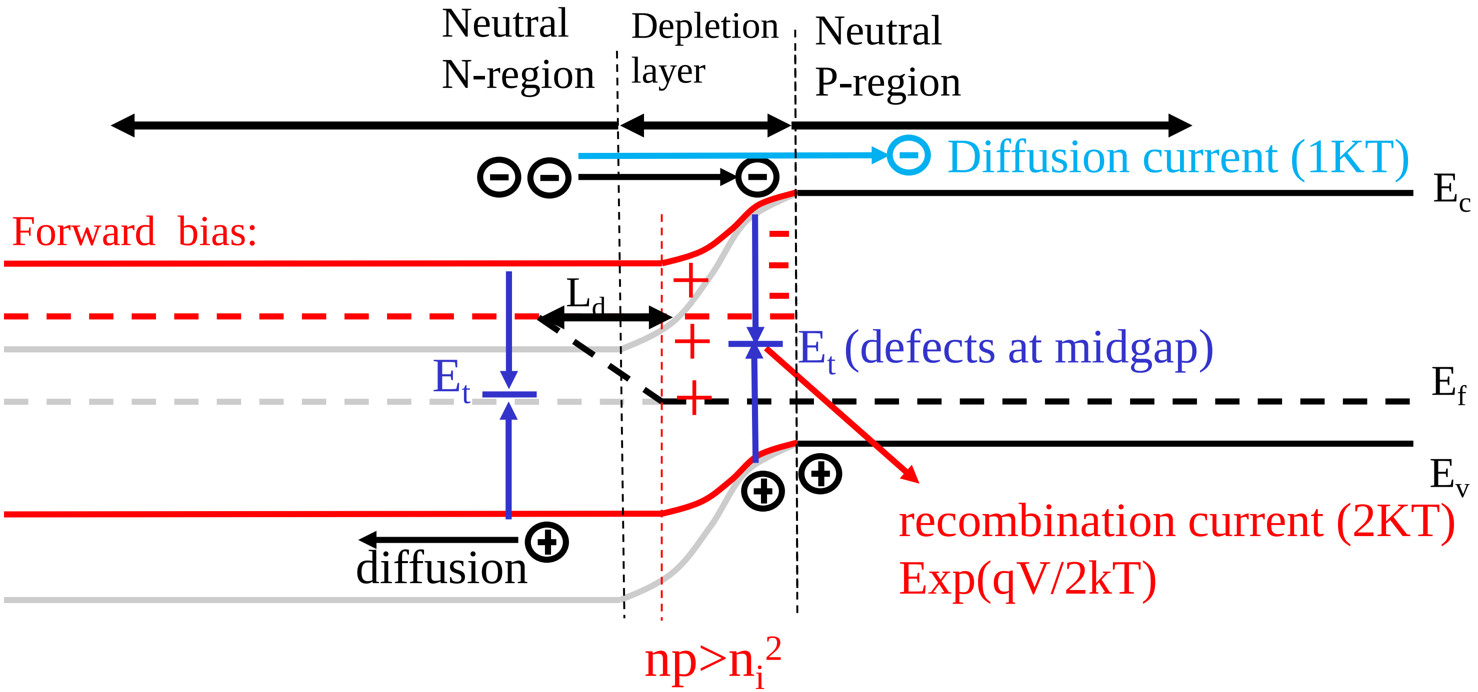
ηkT, 1< η <2

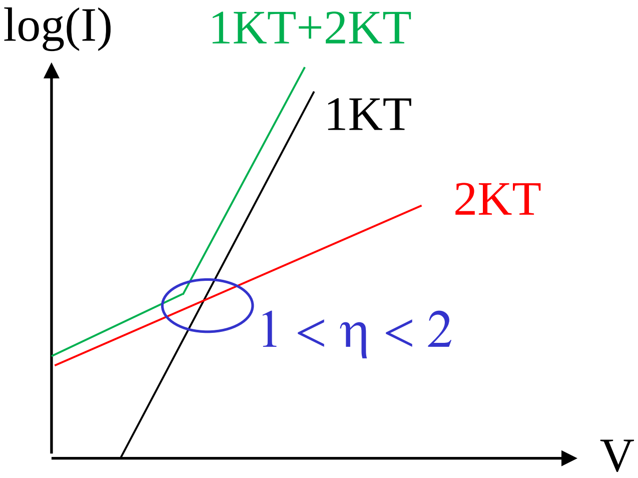
Space-Charge Region (SCR) current

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

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

- For good device: no 2kT current
- For sollar cell η >>2





- low current → 2KT
- η = ? if E_t not in the midgap
- high current → 1KT

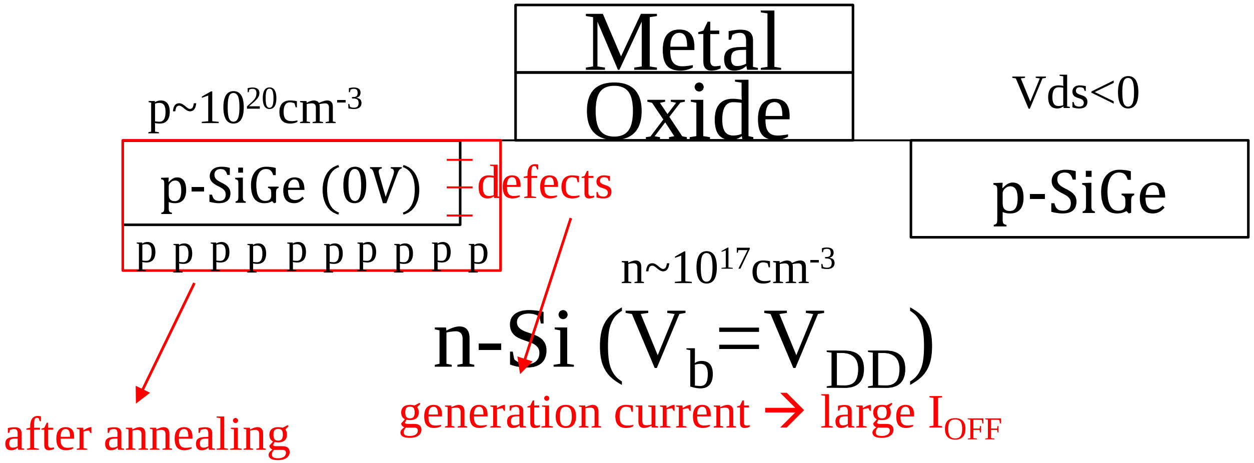
PN junction at V=0:
Electron current=0 (diffusion+drift)
Hole current =0 (diffusion+drift)

PN junction at at forward biased:
Electron current + Hole current(Both diffusion)

Bipolar device: electron + hole + diffusion

FET: majority electron (NFET) or majority hole (PFET)
drift (more important)

90nm node



Metal

Oxide

$p\sim10^{20}cm^{-3}$

p-SiGe (0V)

p p p p p p p p p p

defects

Vds<0

p-SiGe

$n\sim10^{17}cm^{-3}$

n-Si ($V_b=V_{DD}$)

after annealing

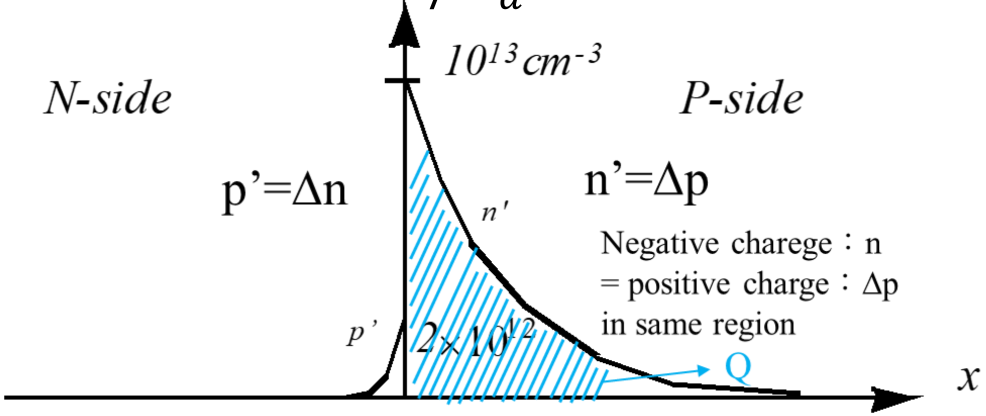
generation current → large I_{OFF}

- After annealing, the dopants of p-SiGe (B) diffuse out, the defects are no longer inside the depletion region

→ Less recombination current

4.10 Charge Storage

reverse bias: $C = \varepsilon / W_d$



Negative charege：n
= positive charge：Δp
in same region

$$Q \propto I \qquad \boldsymbol{Q = \int_0^\infty n' dx}$$

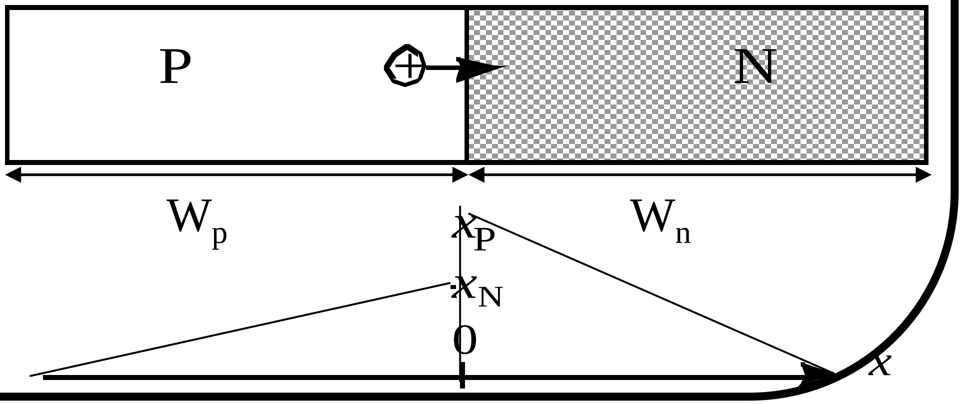
$$I = Q/\tau_s$$

$$Q = I\tau_s$$

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

Practice28: caculate τ_s for long diode

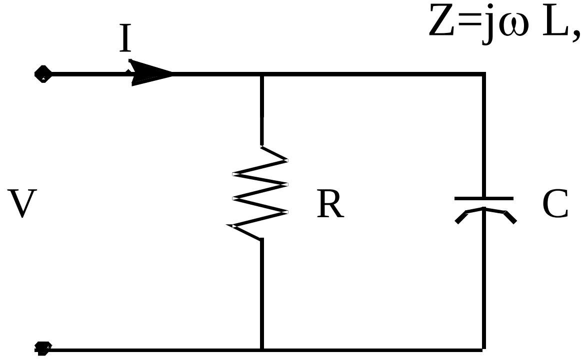
Practice29: caculate τ_s for short diode, namely physical size < diffusion length



Hint: n'(x) : linear

4.11 Small-signal Model of the Diode

Z=jω L, ω is lower 1GHz, L can be negligible



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

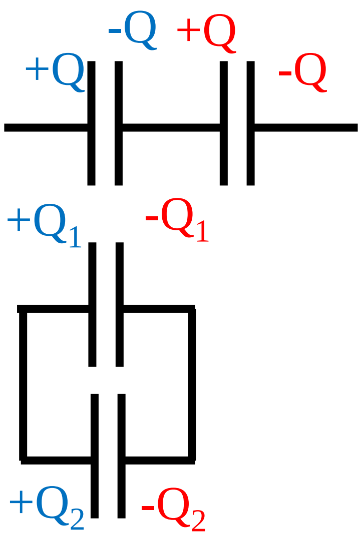
$$= \frac{e}{kT}\boldsymbol{I_0} e^{eV/kT} = \frac{I_{DC}}{kT/e}$$

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

$$\boldsymbol{C_{dep} = \frac{\varepsilon_s}{W_d} = \varepsilon_s / debye\ legnth(flat\ band)}$$

Diffusion Capacitance (small signal):

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



Which is larger, diffusion or depletion capacitance?

C_{dep} and C_{diff} are parallel, C_{diff} larger

4.12 Metal-Semiconductor Junction

Two kinds of metal-semiconductor(MS junction) contacts:

- Rectifying Schottky diodes: metal on lightly doped silicon

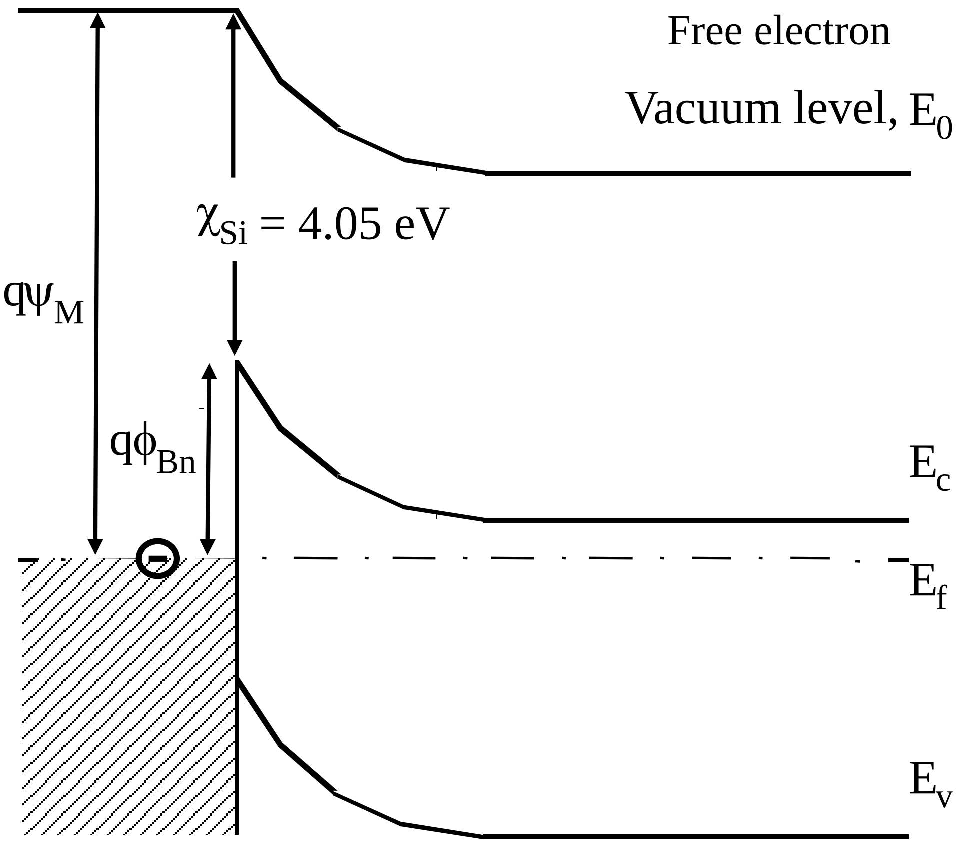
One bias voltage has current, while the other doesn't. There is sufficient current only at forward bias

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

There is sufficient current at both forward and reverse bias

Depletion region 小 → tunneling

ϕ_{Bn} Increases with Increasing Metal Work Function



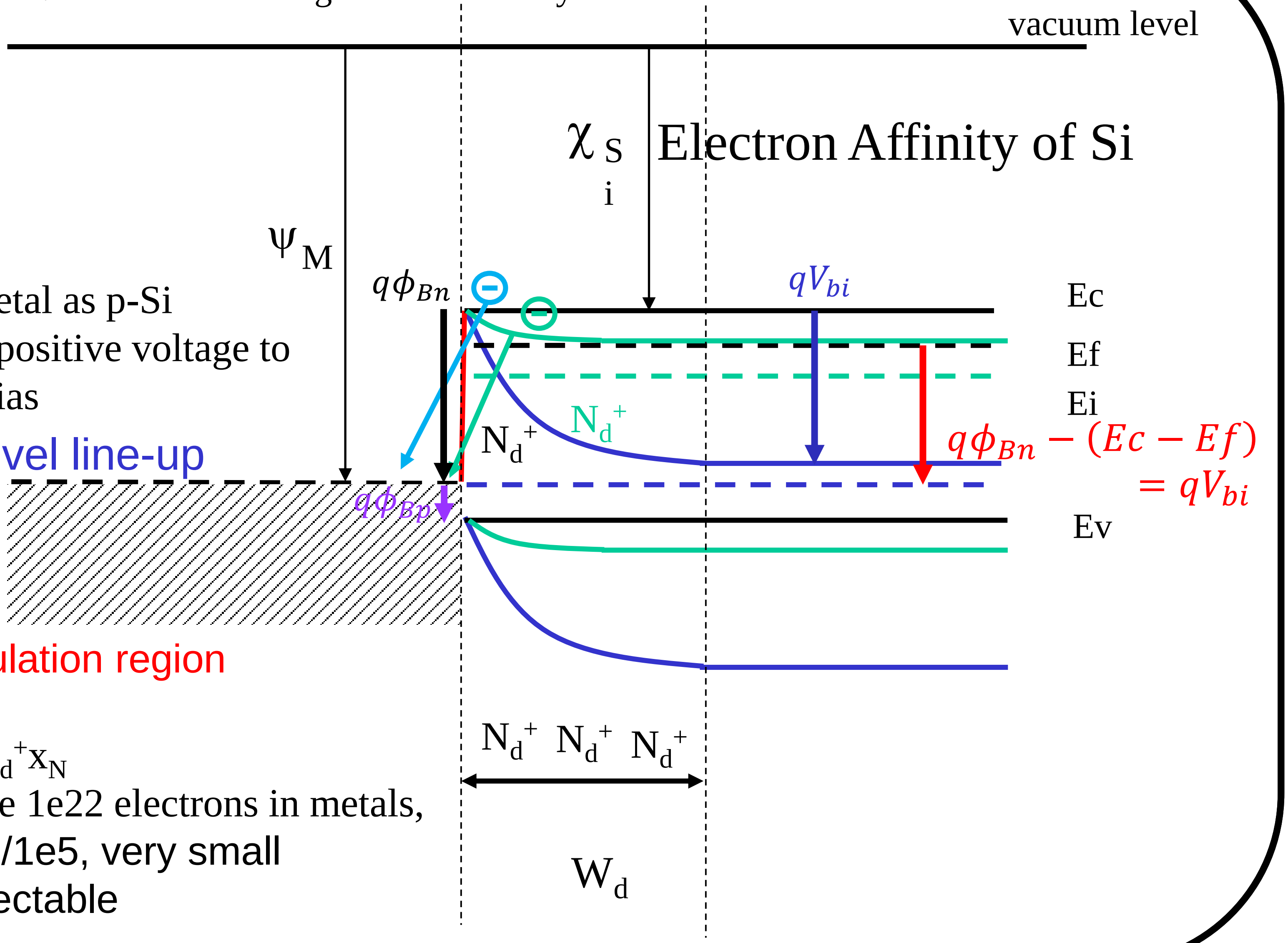
ψ_M:Work Function of metal

χ_{Si}:Electron Affinity of Si

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

- defect/MIGS of interface are not taken into consideration (Metal Induced Gap States)
- Fermi-level pinning

Practice30: draw the band diagram for Schottky diodes



Regard metal as p-Si
→ Apply positive voltage to forward bias

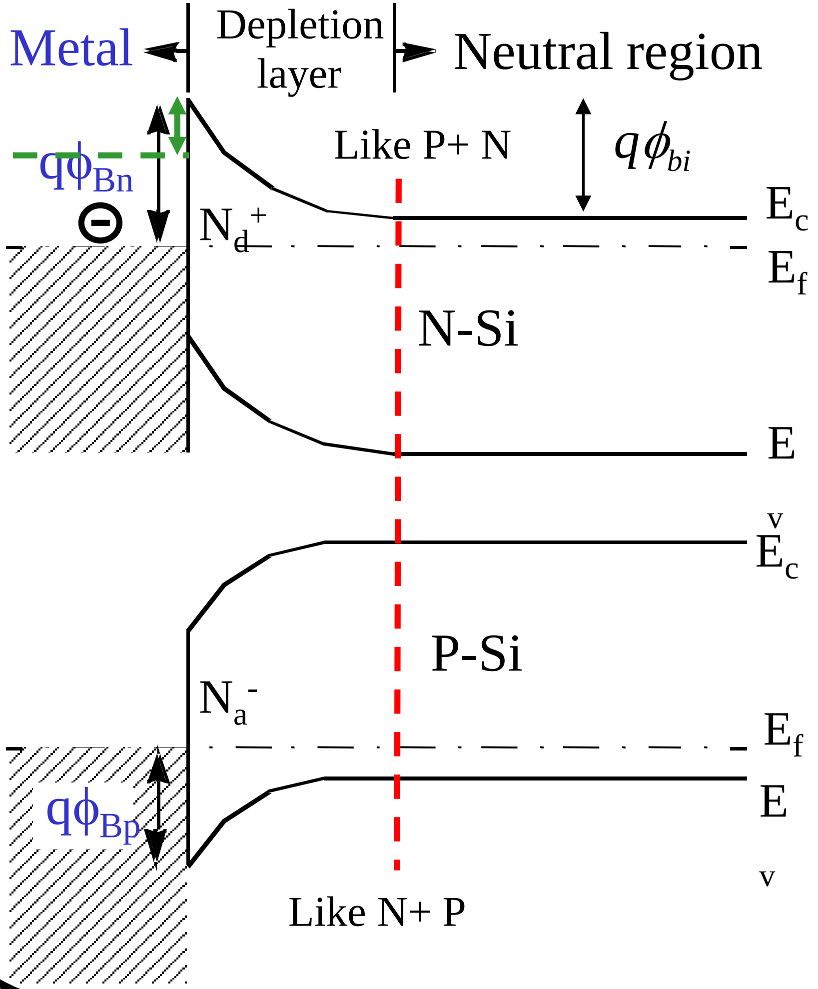
Fermi level line-up

if any,
Accumulation region

$N_a^- x_P = N_d^+ x_N$
There are 1e22 electrons in metals, only W_d/1e5, very small
→ neglectable

4.13 Schottky Barriers

Energy Band Diagram of Schottky Contact $q\phi_{Bn}$ =?



- Schottky barrier height, ϕ_B, is a function of the metal material.
- ϕ_B is the most important parameter. The sum of $q\phi_{Bn}$ and $q\phi_{Bp}$ is equal to E_g .

Schottky barrier heights for electrons and holes

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

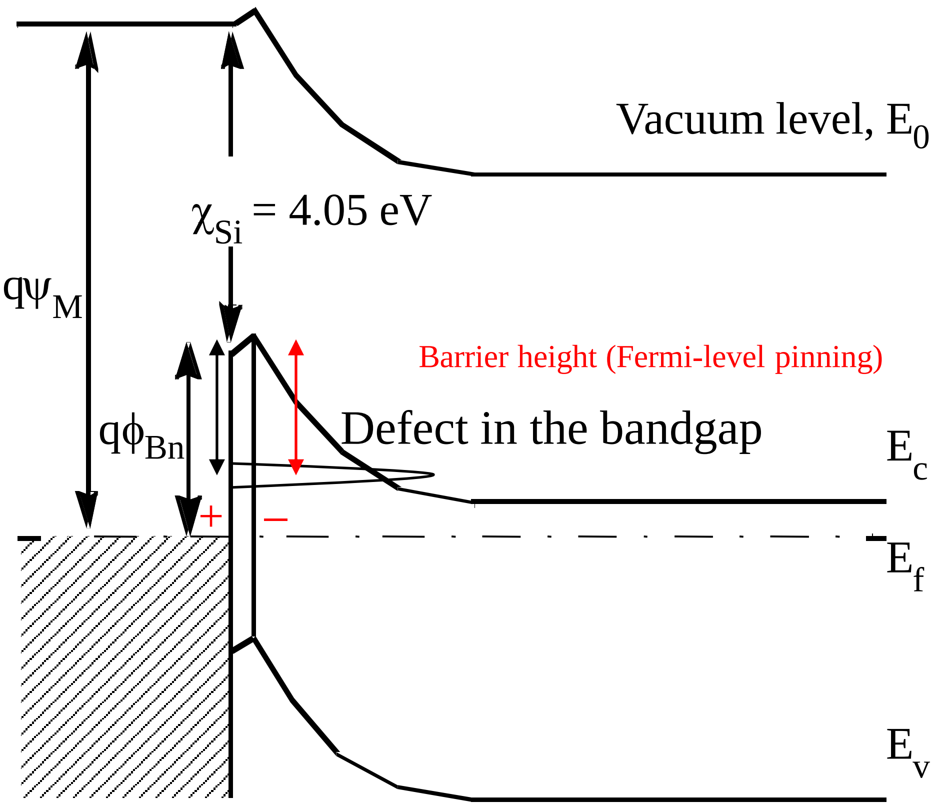
Ni

$$\phi_{Bn} + \phi_{Bp} \approx E_g \text{(even considering Fermi-level pinning)}$$

ϕ_{Bn} increases with increasing metal work function

- Due to large D_{it} at M/S interface
- Small barrier height metal can be used in night vision goggles
- MOSFET needs ohmic contact (small barrier height)

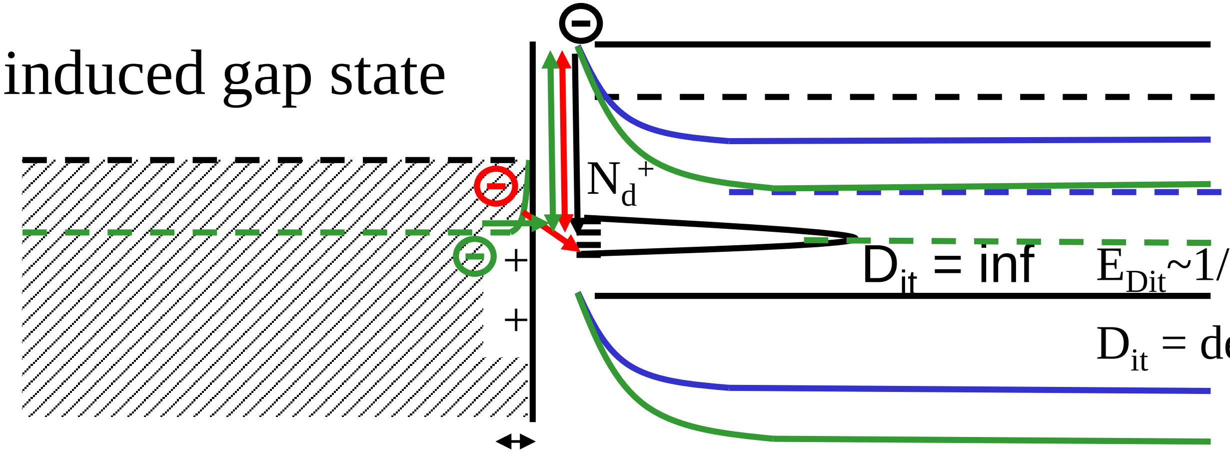
Fermi Level Pinning



- A high density of energy states in the bandgap at the metal-semiconductor interface pins E_f to a narrow range and ϕ_{Bn} is typically 0.4 to 0.9 V
- Question: What is the typical range of ϕ_{Bp}? 0.7 to 0.2 V

Metal induced gap state

$$q\phi_{Bn} = E_c - E_{Dit} \sim \frac{2}{3} E_g$$



N_d^+

D_{it} = inf

E_{Dit}~1/3E_g, close to Ev

D_{it} = defect + MIGS

W_d within metal ~ 0
electrons can tunnel

$$q\phi_{Bp} = E_{Dit} - \mathrm{E}_v \sim \frac{1}{3} E_g$$

- Fermi-level pinning at E_{Dit}

→ Barrier height is indepent of metal work function

Schottky Contacts of Metal Silicide on Si

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

M+Si → MSi due to thermal budget

Planar technology: After sputtering Source/Drian anneal: 1050°C

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

TiN is used for metal gate

Erbium

platinum

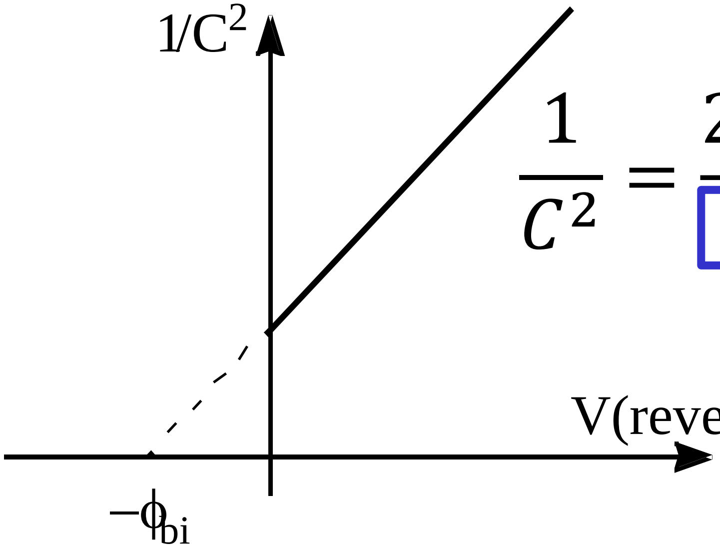
Silicide	$ErSi_{1.7}$	HfSi	$MoSi_2$?	$ZrSi_2$	$TiSi_2$	$CoSi_2$	WSi_2	$NiSi_2$	Pd_2Si	PtSi
ϕ_{Bn} (V)	0.28	0.45	0.55	0.55	0.61	0.65	0.67	0.67	0.75	0.87
ϕ_{Bp} (V)			0.55	0.49	0.45	0.45	0.43	0.43	0.35	0.23

R_c :contact resistance

Al: workfunction ~ E_c

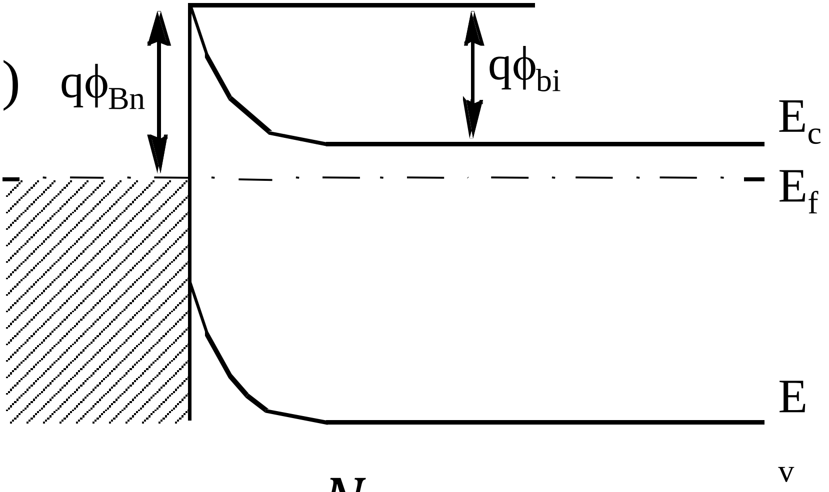
Using CV Data to Determine ϕ_B

Diameter: 5%
Area: 10%
N_D^+: 20%



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

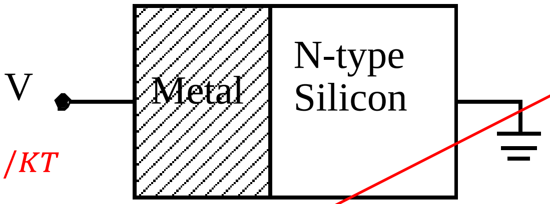
Slope: N_D^+
Intercept: ϕ_{bi}

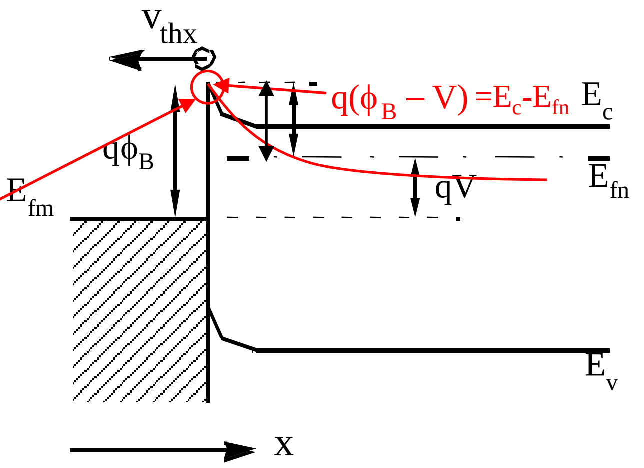


Once ϕ_{bi} is known, ϕ_B can be determined using

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

4.17 Thermionic Emission Theory





$n = N_c e^{-(Ec-Efn)/KT}$

n(energy above energy tip)=

$N_c e^{-(Ec-Efn)/KT} e^{-q\phi_{bi}/KT}$

$= N_c e^{-q\phi B/KT}$

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

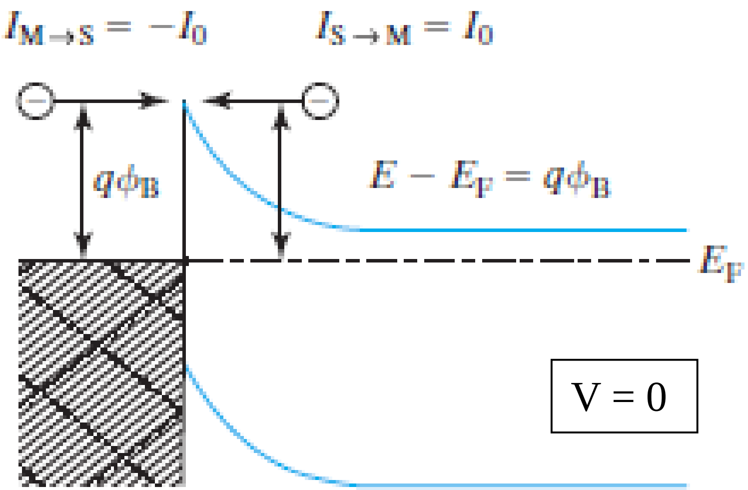
$$v_{th} = \sqrt{3kT/m_n} \qquad v_{thx} = -\sqrt{2kT/\pi m_n} \ ?$$

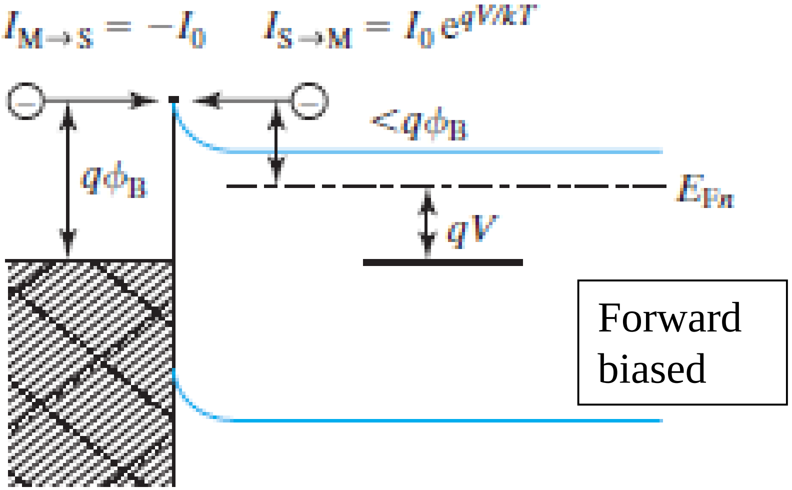
Half of electrons go left

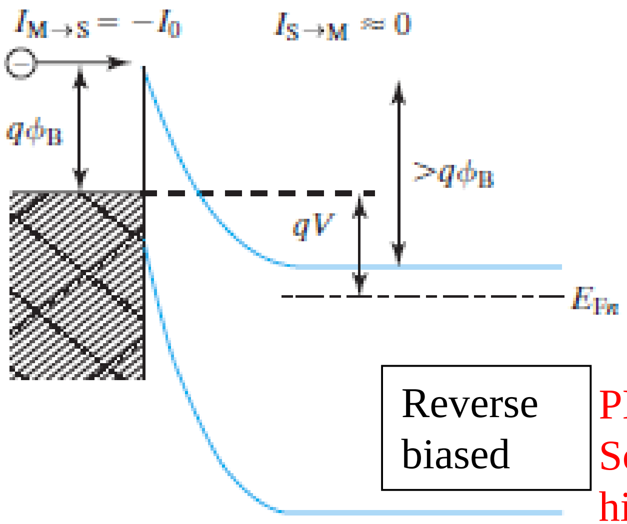
$$J_{S\to M} = -\frac{1}{2}qnv_{thx} = \frac{4\pi q m_n k^2}{h^3}T^2 e^{-q\phi_B/kT} e^{qV/kT}$$

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

4.18 Schottky Diodes







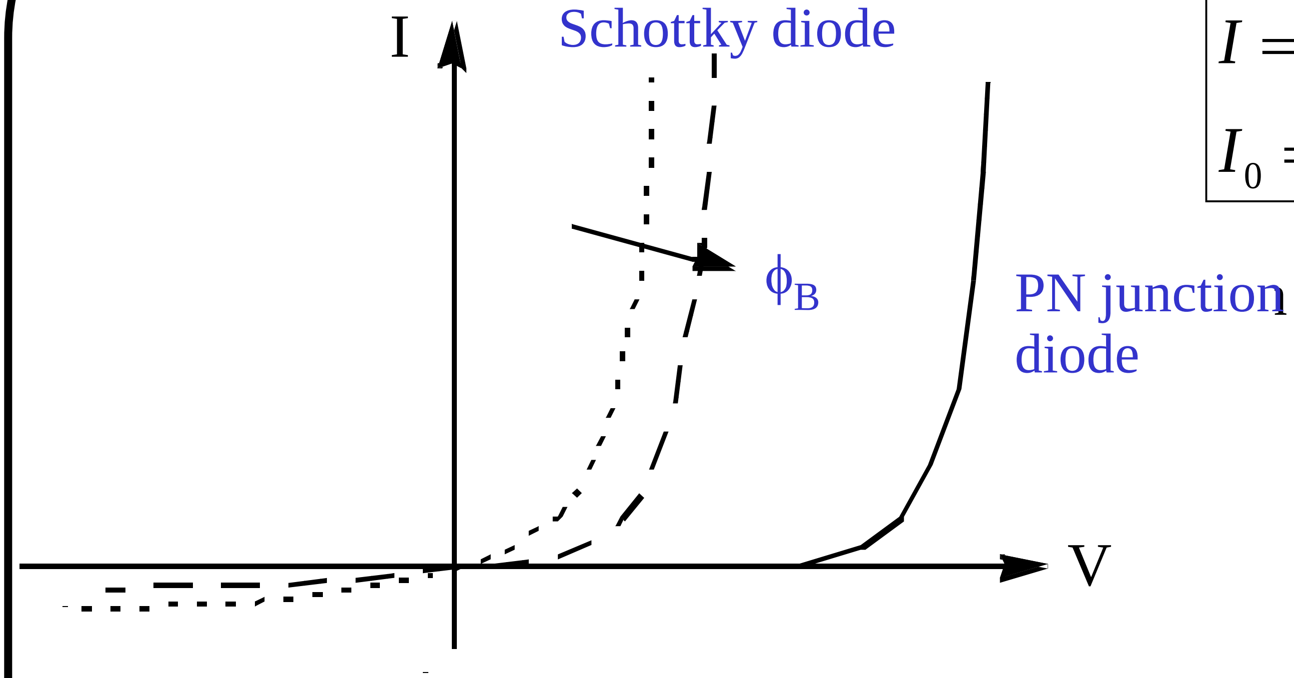
$$I_0 = AKT^2 e^{-q\varphi_B/kT}$$

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

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

PN: diffusion (minority carrier)
Schottky: thermionic emission (majority carrier)→
high current, high speed

4.19 Applications of Schottly Diodes

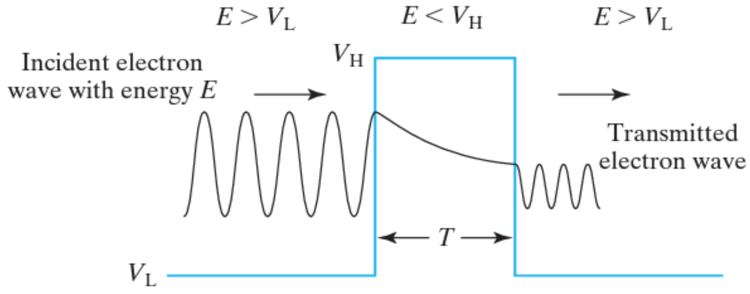


$$I = I_0(e^{qV/kT} - 1)$$

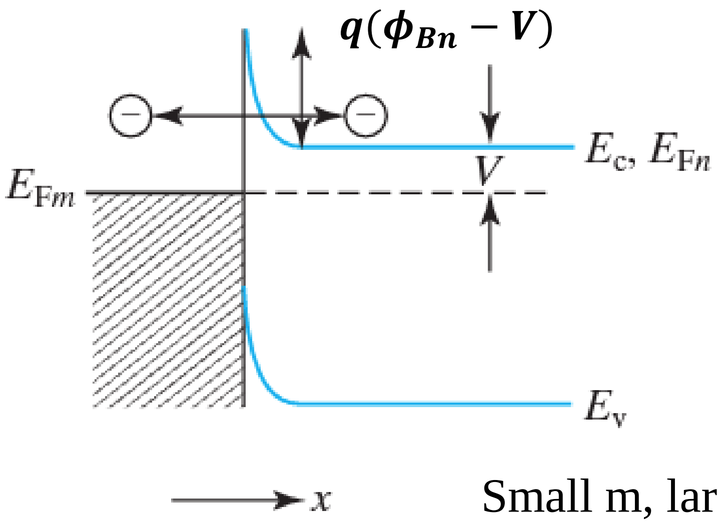
$$I_0 = AKT^2 e^{-q\phi_B/kT}$$

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

4.20 Quantum Mechanical Tunneling



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



$$\begin{cases} V_H - E = q(\phi_{Bn} - V) \\ T \approx \dfrac{W_{dep}}{2} = \sqrt{\dfrac{\epsilon_s(\phi_{Bn} - V)}{2qN_d}} \end{cases}$$

Small m, large P

$$A = \frac{2}{\hbar}\sqrt{m\epsilon_s}$$

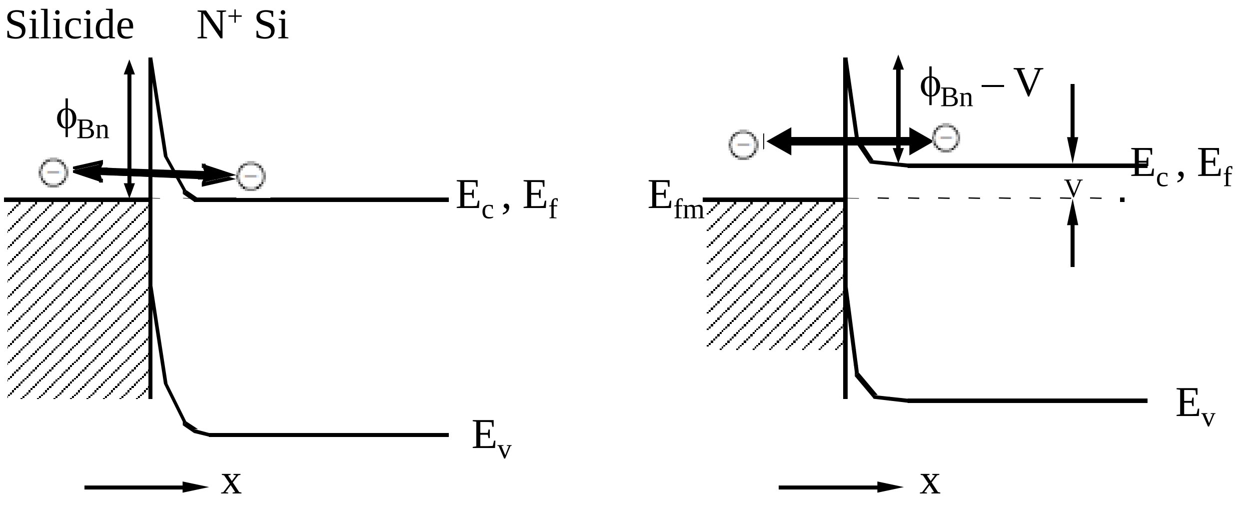
$$P = exp\left(-2\sqrt{\frac{8\pi^2 m[q(\phi_{Bn} - V)]}{h^2}\frac{\epsilon_s(\phi_{Bn} - V)}{2qN_d}}\right) = exp\left(-\frac{2}{\hbar}(\phi_{Bn} - V)\sqrt{\frac{m\epsilon_s}{N_d}}\right) = exp(-A\frac{(\phi_{Bn} - V)}{\sqrt{N_d}})$$

$$\boxed{P = exp(-A\frac{(\phi_{Bn} - V)}{\sqrt{N_d}}) \propto exp(-\frac{(\phi_{Bn} - V)}{\sqrt{N_d}})}$$

N_d: Doping Concentration
ϕ_{Bn}: Barrier Height
V: Applied Voltage

4.21 Ohmic Contacts

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



Tunneling probability:

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

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

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

thermionic emission : E > barrier height

tunneling (W_{dep} < 10nm) : E < barrier height

tunneling current

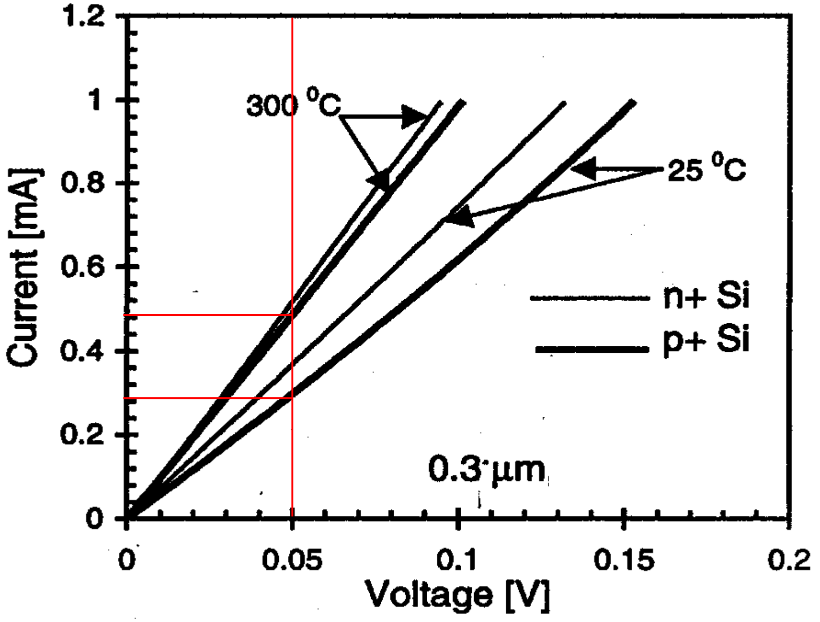
$$J_{S\to M} \approx \frac{1}{2} qN_d v_{thx} P = qN_d \sqrt{\boxed{kT}/2\pi m_n}\, e^{-H(\phi_{Bn}-V)/\sqrt{N_d}}$$

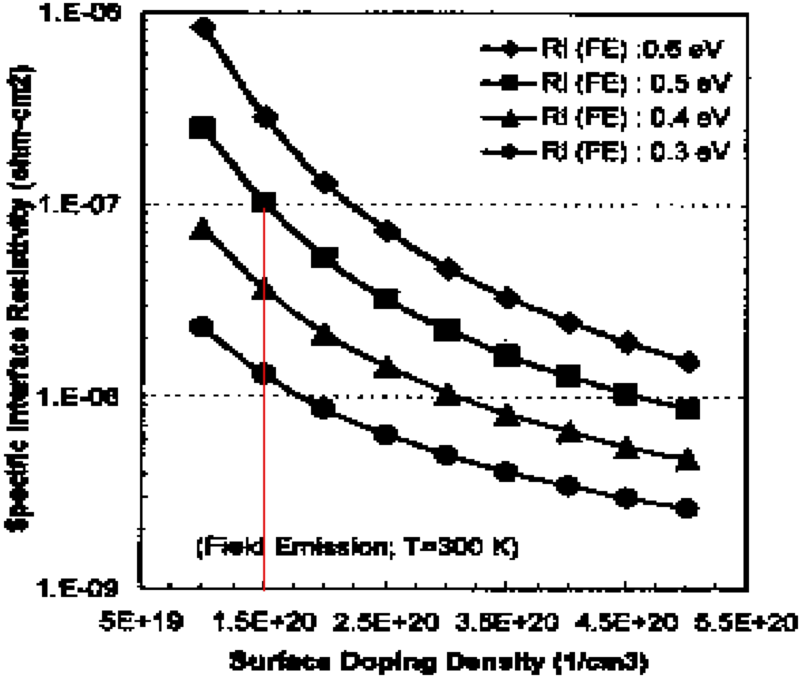
weak dependence on temperature

4.21 Ohmic Contacts

$R_c(2D) : \Omega \cdot cm$

R_c : specific contact resistivity $(\Omega \cdot cm^2)$





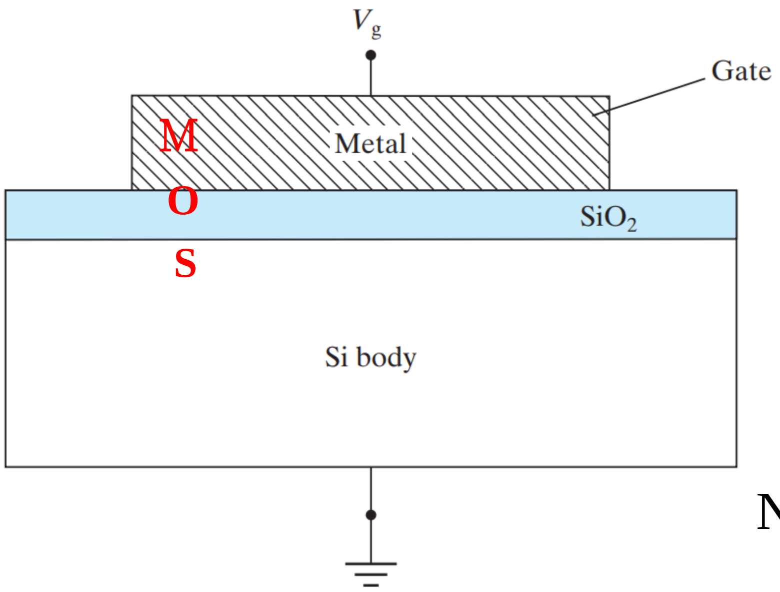
Target: 10^{-10}

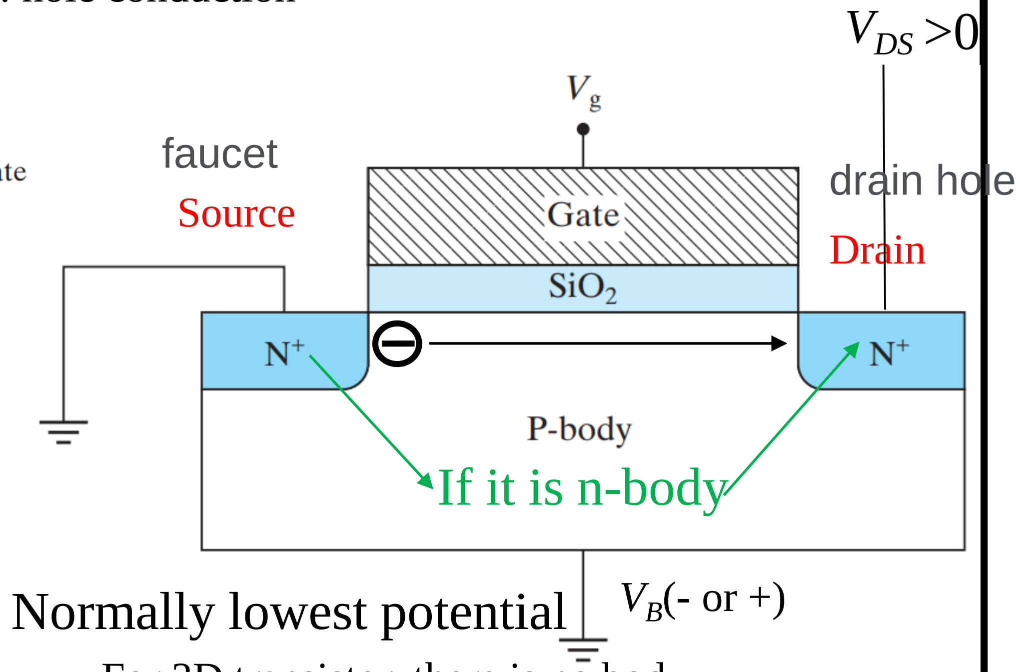
$$R_c \equiv \left(\frac{dJ_{S \to M}}{dV} \right)^{-1} = \frac{2e^{H\phi_{Bn}/\sqrt{N_d}}}{qv_{thx}H\sqrt{N_d}} \propto e^{H\phi_{Bn}/\sqrt{N_d}} \;\; \Omega \cdot \mathrm{cm}^2$$

Chapter 5 MOS Capacitor

MOS: Metal-Oxide-Semiconductor

- For bulk planar, p-body means nFET.
- nFET: electron conduction, pFET: hole conduction





Normally lowest potential

For 3D transistor, there is no body

MOS capacitor : MOSCAP

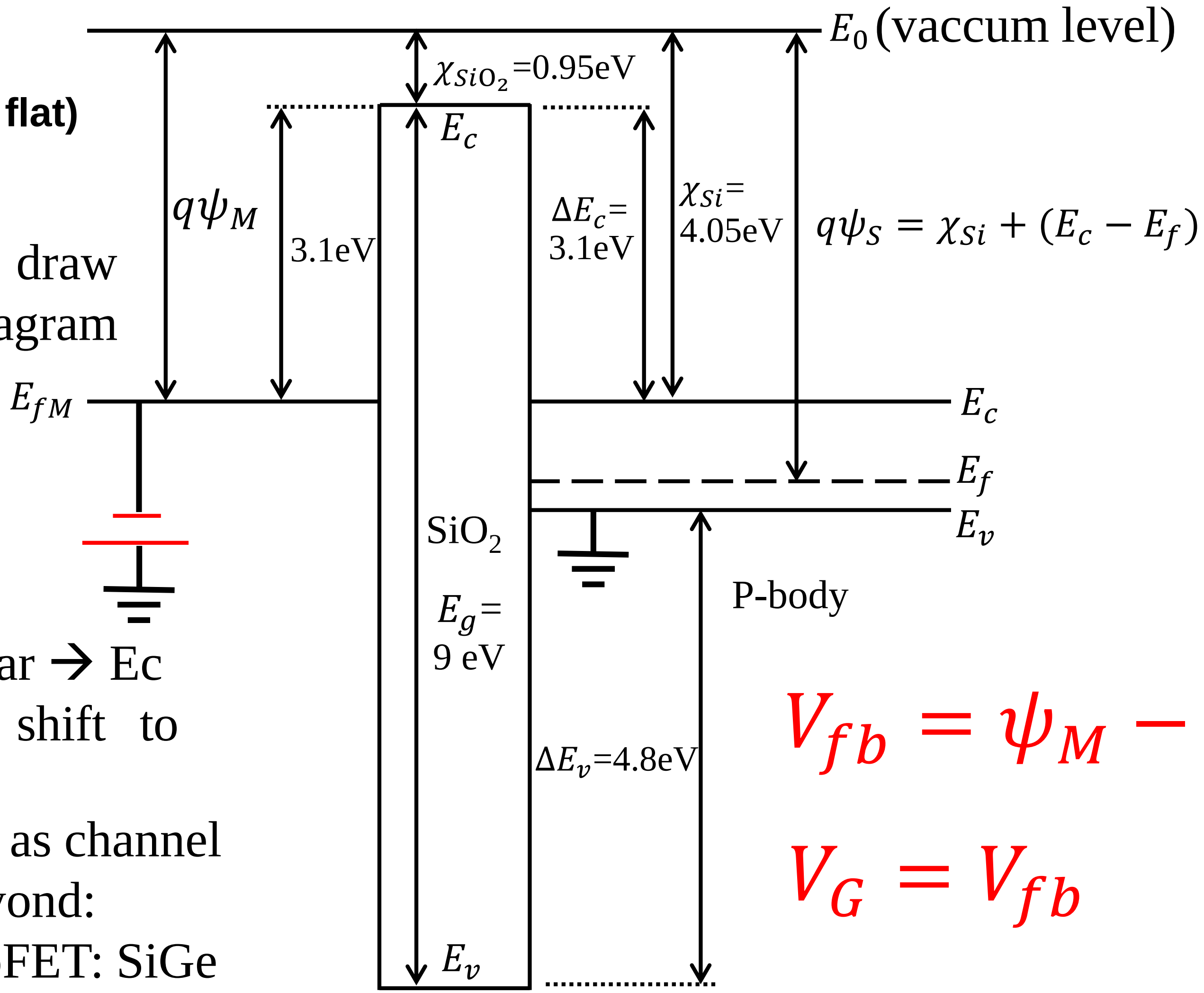
MOS transistor: MOSFET

MOSFET (field-effect transistor):drift current

5.1 Flat-band Condition and Flat-band Voltage

ϕ_s=0
(E_c/E_v at Si are flat)

Practice31: draw the band diagram for flat band



$$V_{fb} = \psi_M - \psi_S$$

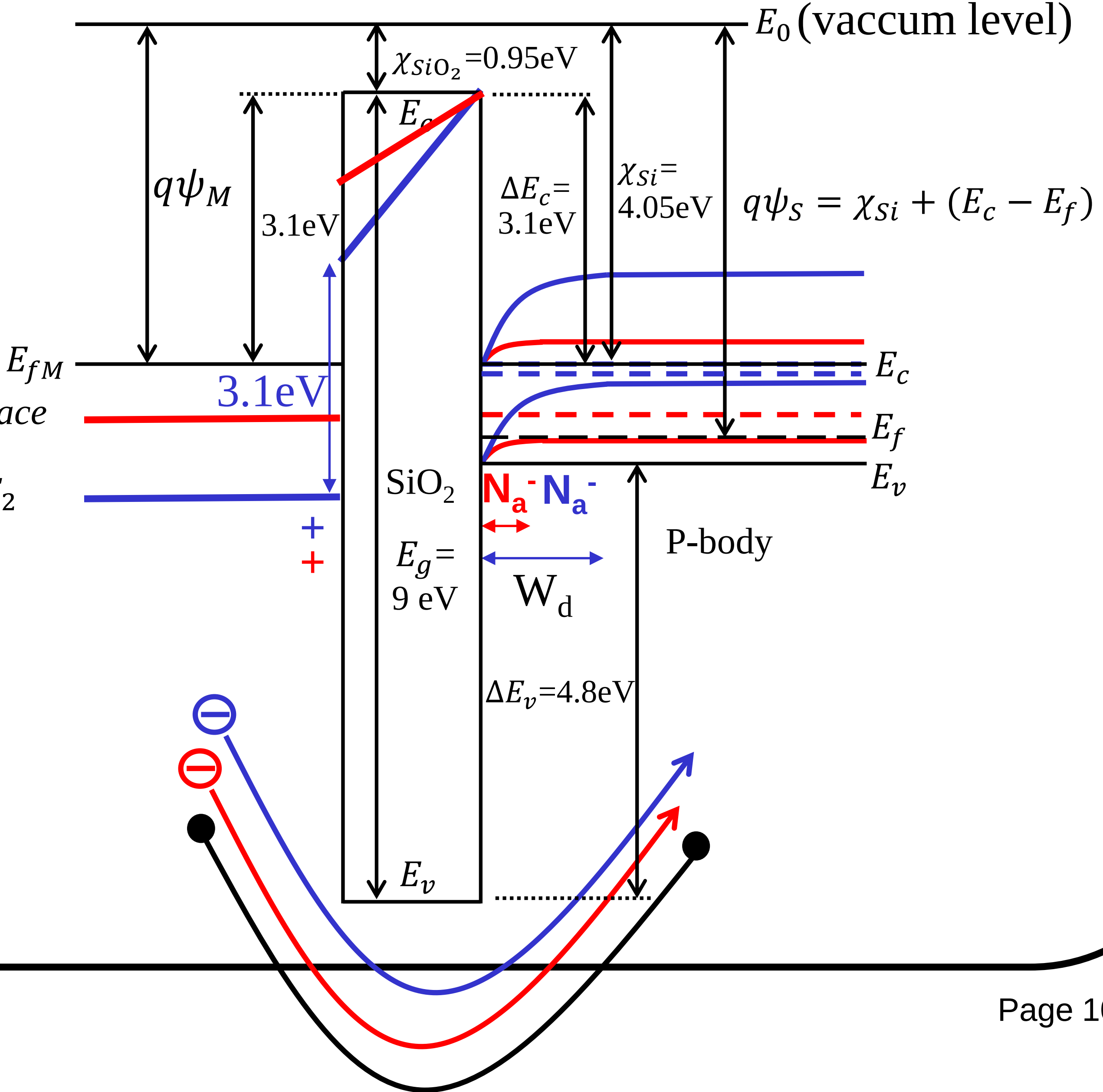
$$V_G = V_{fb}$$

nMetal: planar → Ec
FinFET Ec shift to near midgap

Si is defined as channel
5nm and beyond:
nFET: Si pFET: SiGe

Short: $V_G = 0$

χ : electron affinity



No charge at interface

$\rightarrow \varepsilon_1 E_1 = \varepsilon_2 E_2$

$\varepsilon_{Si} = 11.9$

$\varepsilon_{SiO_2} = 3.9$

$\varepsilon_{Si}/\varepsilon_{SiO_2} = 3$

5.1.1 Oxide Charge–A Modification to V_{fb}

Types of oxide charge:

- Fixed oxide charge, Si^+

Pure DI water: 18MΩ-cm
Due to CO_2 in air → 5-6MΩ-cm

- Mobile oxide charge, due to Na^+ contamination
- Interface traps (D_{it}), neutral or charged depending on Vg.

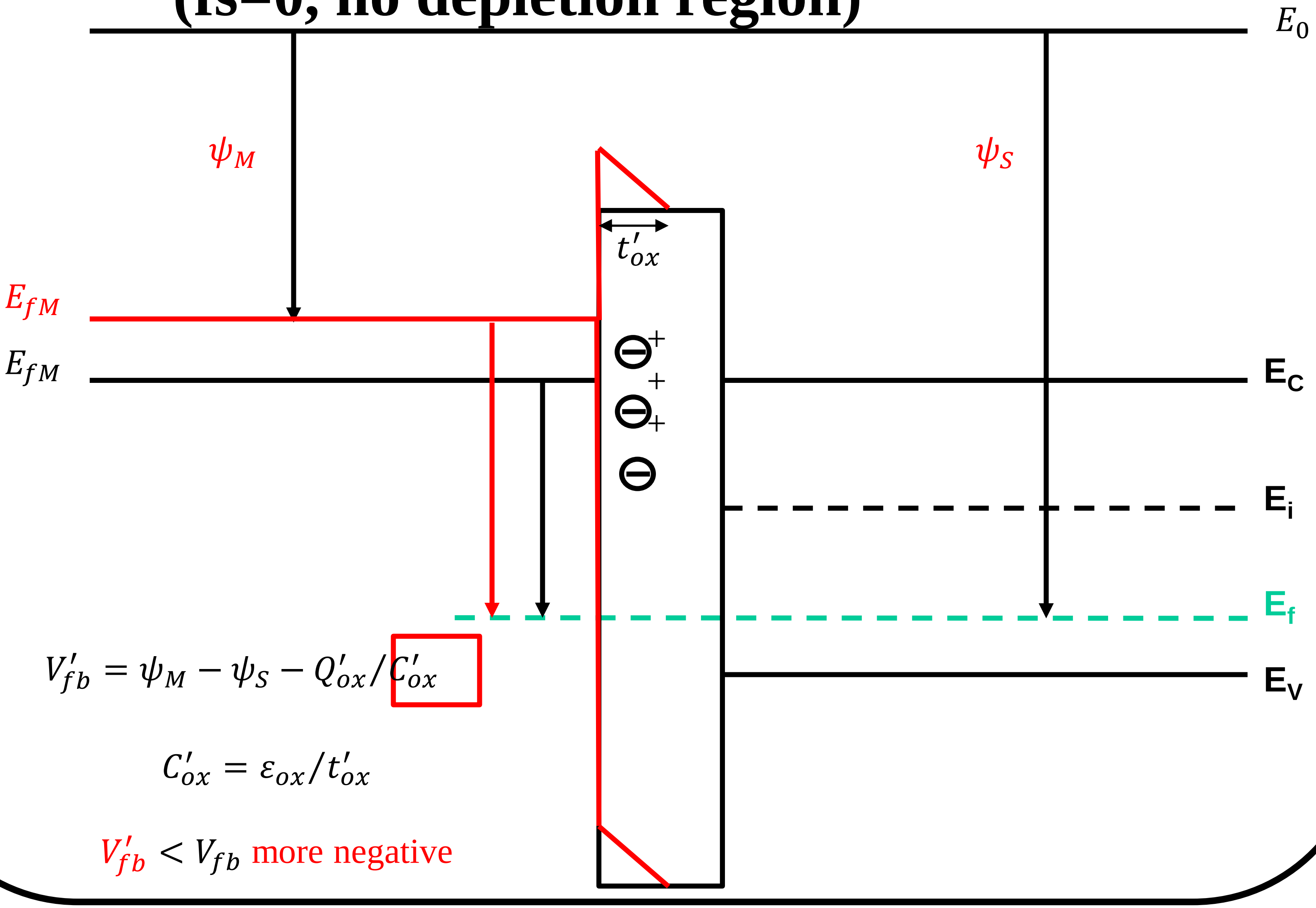
Defect + electron + hole

- Voltage/temperature/current stress induced charge and traps →a reliability issue

cause defects

  - Si cap on SiGe (5nm) solve the reliability issue

Flat band considering fixed charge (fs=0, no depletion region)

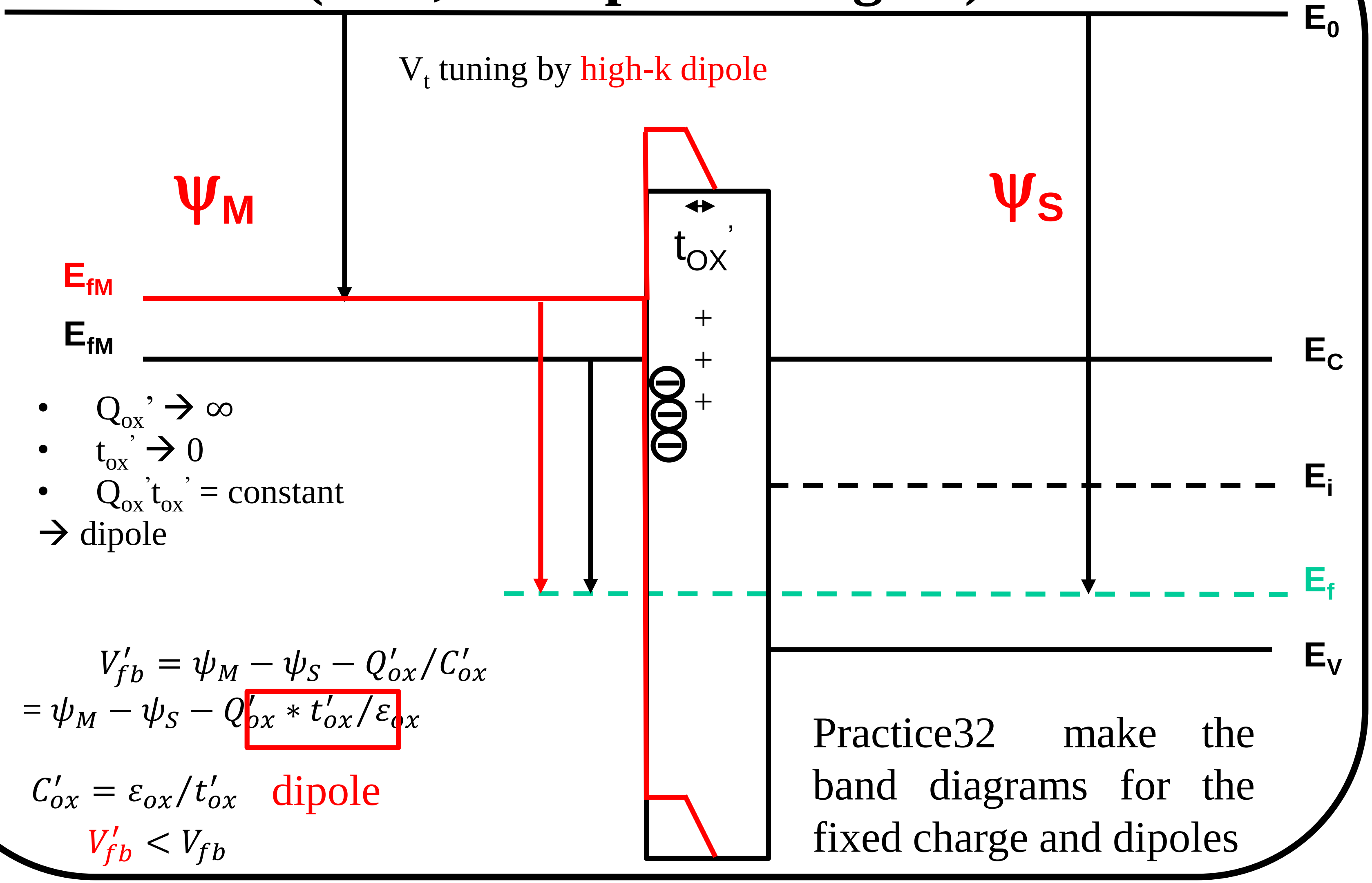


$V'_{fb} = \psi_M - \psi_S - Q'_{ox}/\boxed{C'_{ox}}$

$C'_{ox} = \varepsilon_{ox}/t'_{ox}$

$V'_{fb} < V_{fb}$ more negative

Flat band considering dipole
(fs=0, no depletion region)



- $Q_{ox}' \rightarrow \infty$
- $t_{ox}' \rightarrow 0$
- $Q_{ox}' t_{ox}' =$ constant

→ dipole

$$V'_{fb} = \psi_M - \psi_S - Q'_{ox}/C'_{ox}$$

$$= \psi_M - \psi_S - \boxed{Q'_{ox} * t'_{ox}/\varepsilon_{ox}}$$

$$C'_{ox} = \varepsilon_{ox}/t'_{ox}$$

dipole

$$V'_{fb} < V_{fb}$$

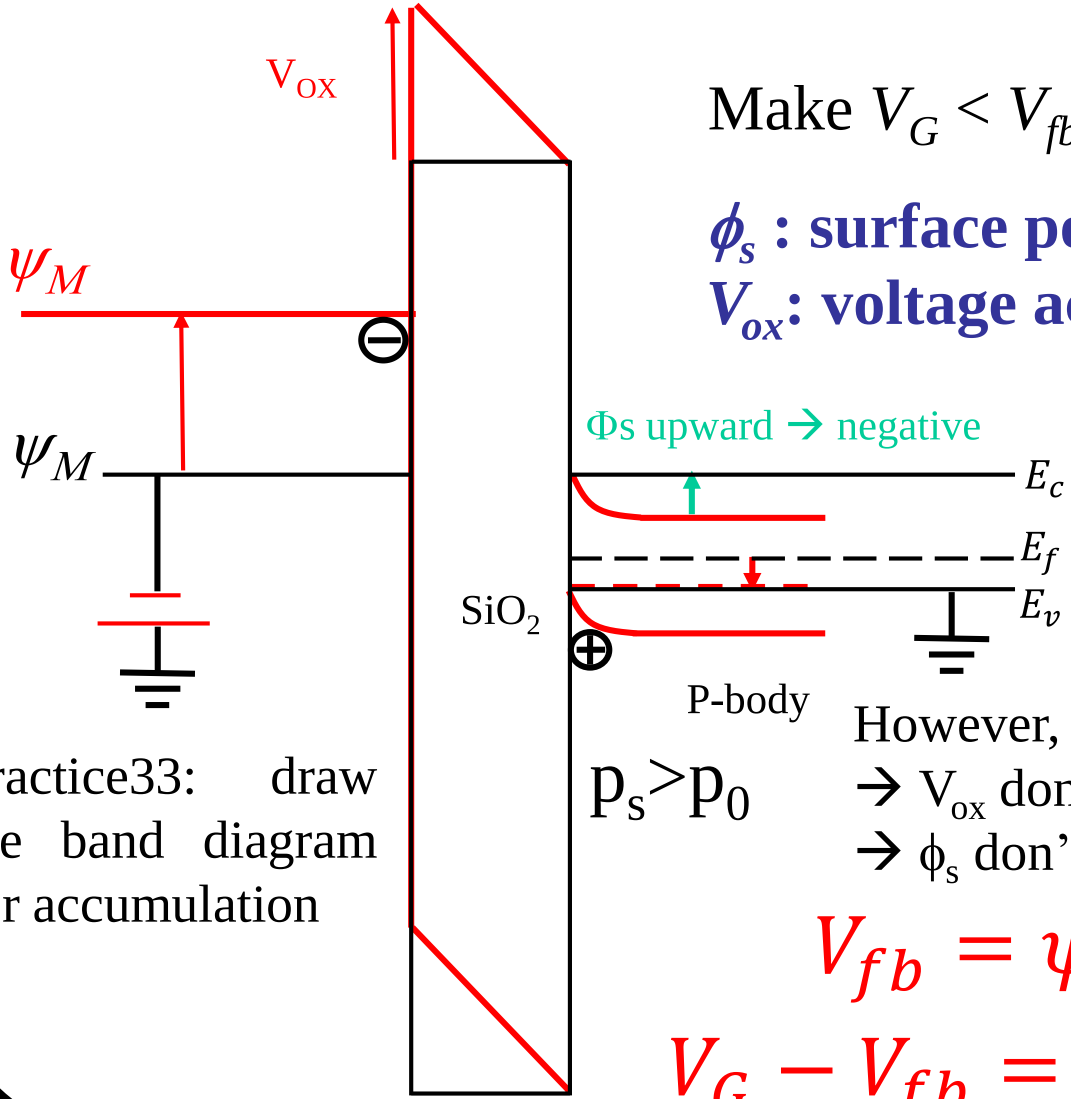
Practice32 make the band diagrams for the fixed charge and dipoles

5.2 Surface Accumulation

Make $V_G < V_{fb}$

ϕ_s : surface potential, band bending

V_{ox}: voltage across the oxide



ϕ_S is negligible when the surface is in accumulation.

$$p_s = p_0 e^{-q\phi_S/kT}$$

$$V_{ox} \propto e^{\phi_S/kT}$$

However, V_G don’t increase exponentially.
→ V_{ox} don’t increase exponentially, either.
→ ϕ_s don’t change too much.

Practice33: draw the band diagram for accumulation

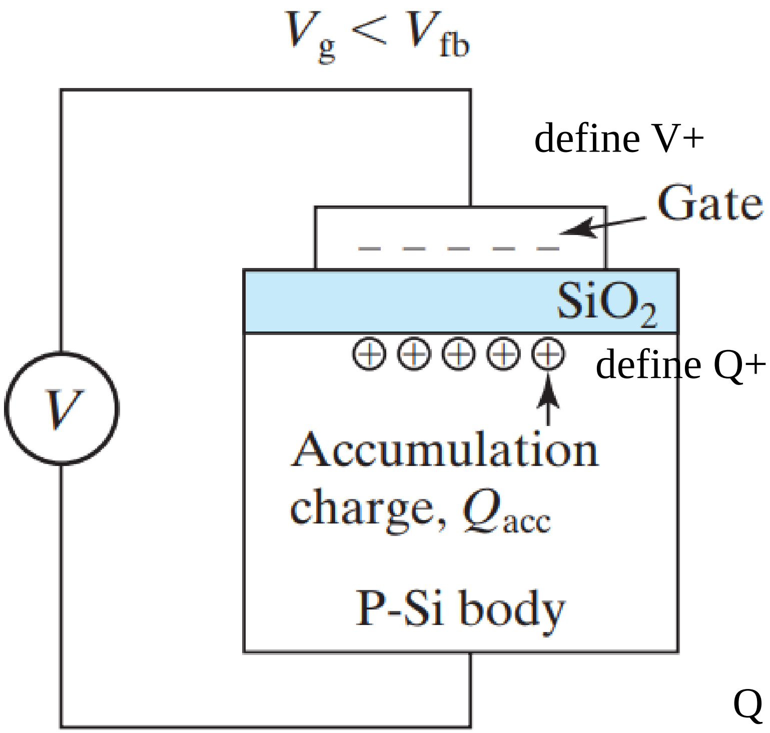
$$V_{fb} = \psi_M - \psi_S$$

$$V_G - V_{fb} = V_{OX} + \phi_S$$

5.2 Surface Accumulation

$V_G - V_{fb} = V_{OX} + \phi_S$

ϕ_s is small, negligible → $V_{OX} = V_G - V_{fb}$



Gauss's Law → $V_{OX} = -\frac{Q_{acc}}{C_{ox}}$

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

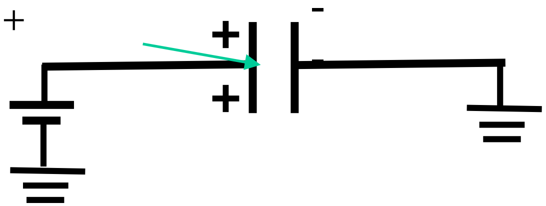
positive (negative)

C_{OX} must be positive

$V_{OX} = -\frac{Q_s}{C_{ox}}$ ⟶ Silicon

Q = CV

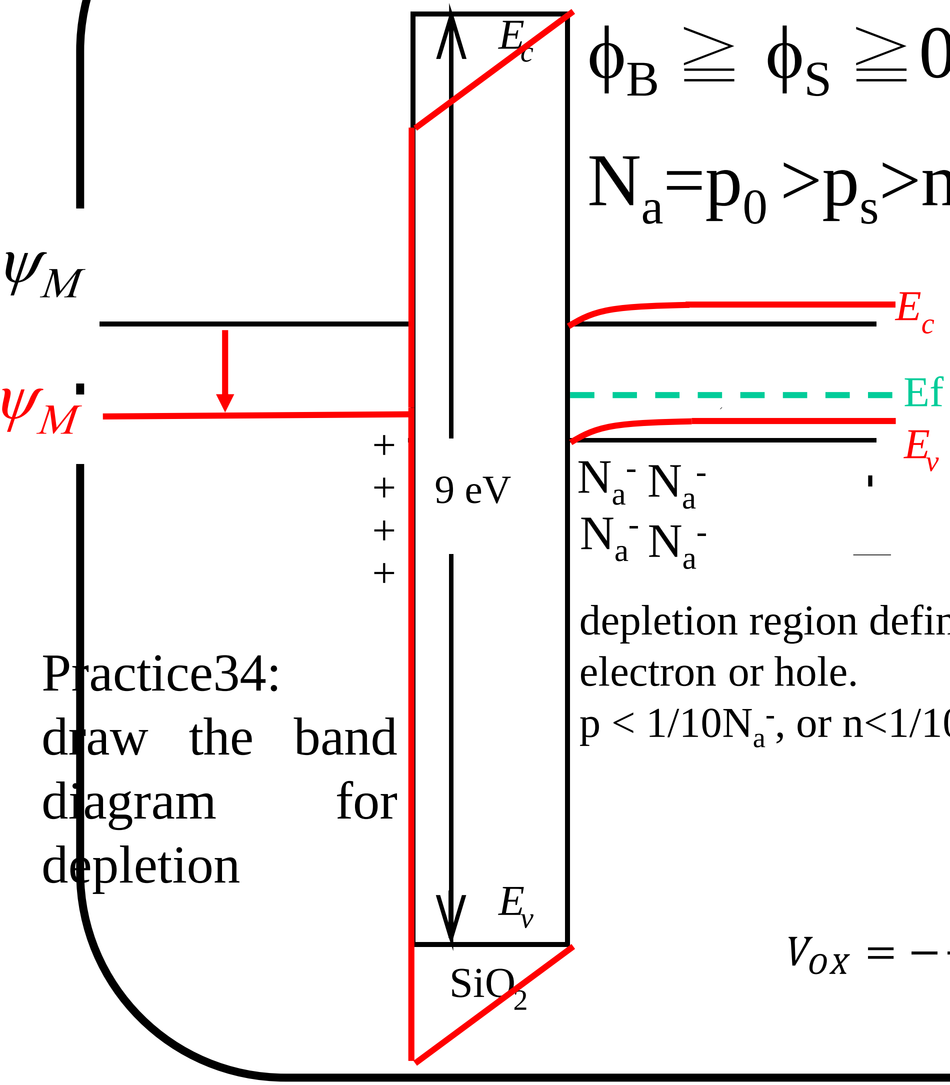
define Q,V +



5.3 Surface Depletion (V_g > V_{fb}) for Off State

$$\phi_B \geqq \phi_S \geqq 0$$

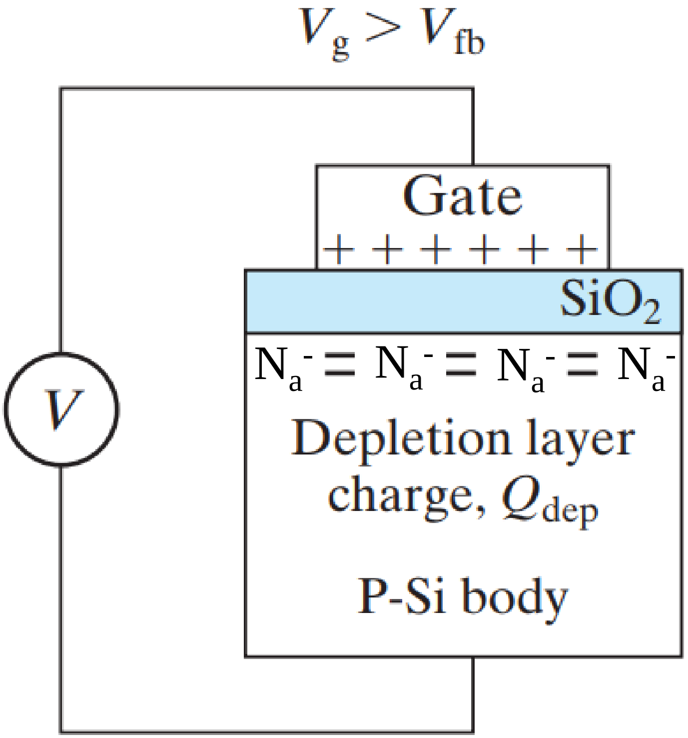
$$N_a = p_0 > p_s > n_s > n_0$$



depletion region define as no electron or hole.

$p < 1/10N_a^-$, or $n<1/10N_d^+$

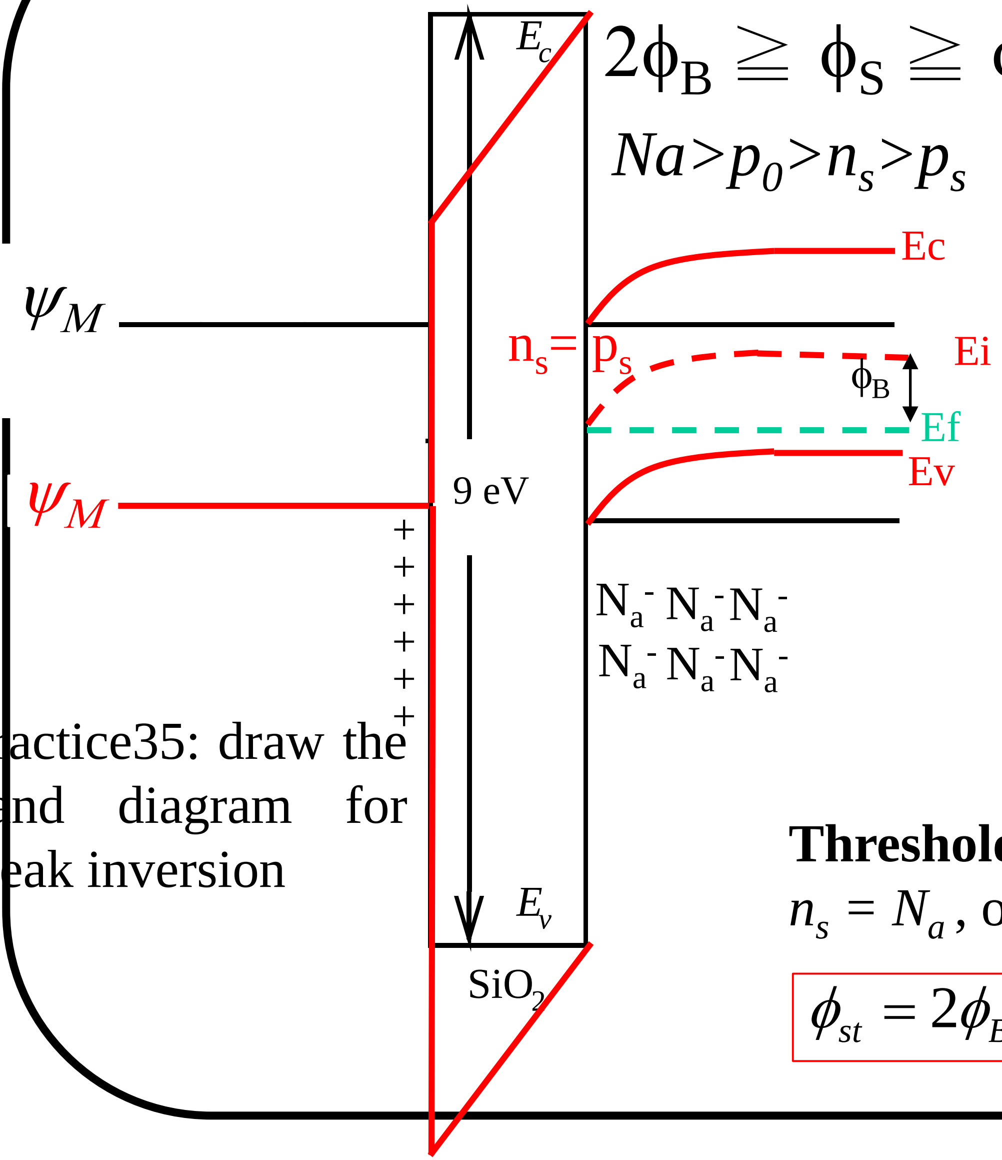
Practice34: draw the band diagram for depletion



$$C_{ox} = \varepsilon_{ox}/t_{ox}(F/cm^2)$$

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

5.4 Weak Inversion and Threshold Condition



$$2\phi_B \geqq \phi_S \geqq \phi_B$$

$$Na > p_0 > n_s > p_s$$

Surface concentration can be controlled by gate voltage

$$q\phi_B = \frac{E_g}{2} - (E_f - E_v)|_{bulk}$$

$$= \frac{kT}{q}\ln\left(\frac{N_v}{n_i}\right) - \frac{kT}{q}\ln\left(\frac{N_v}{N_a}\right)$$

$$= \frac{kT}{q}\ln\left(\frac{N_a}{n_i}\right)$$

Practice35: draw the band diagram for weak inversion

Threshold (of inversion):

$n_s = N_a$, or

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

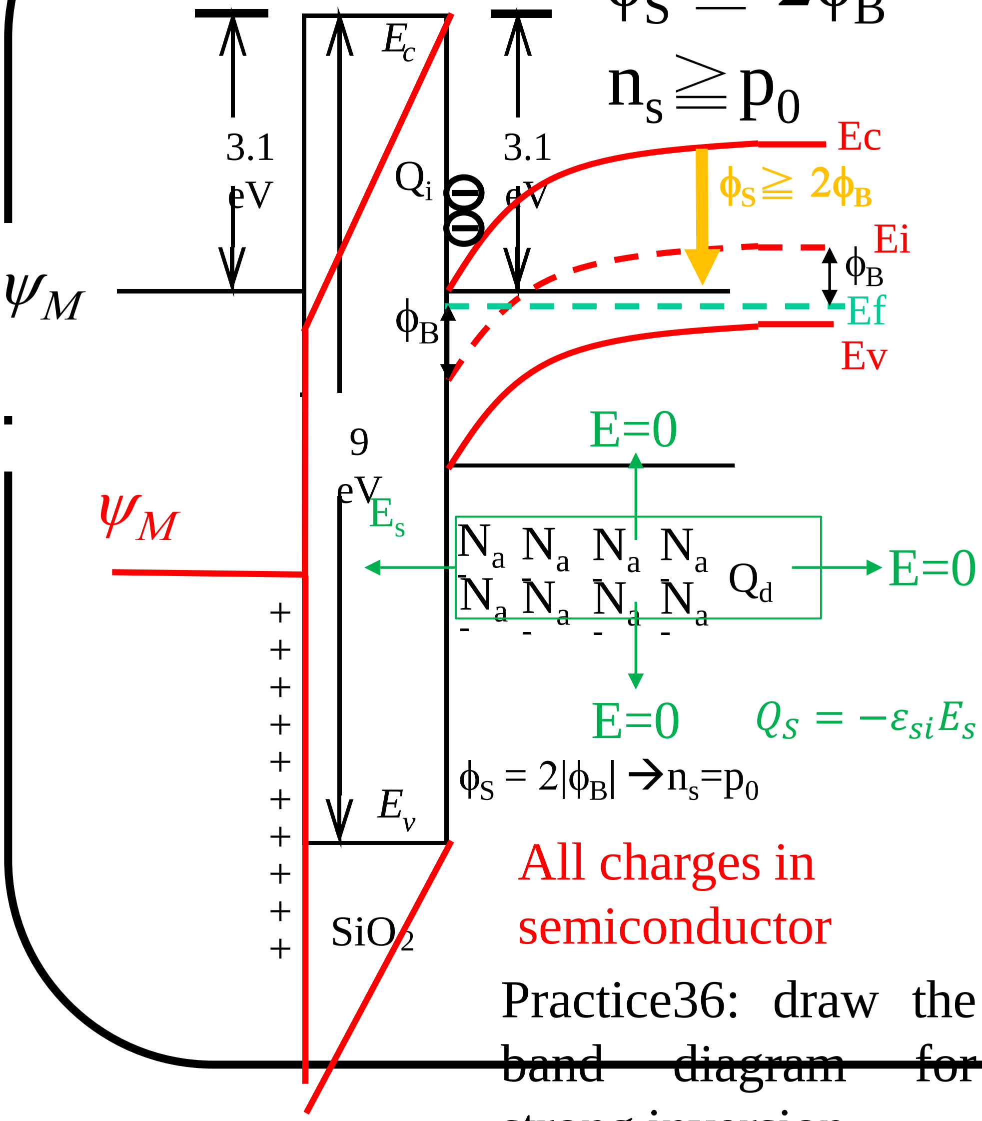
5.5 Strong Inversion–Beyond Threshold

$$\phi_S \geqq 2\phi_B$$

$$n_s \geqq p_0$$

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

$$V_{ox}=V_g-V_{fb}-2\phi_B$$



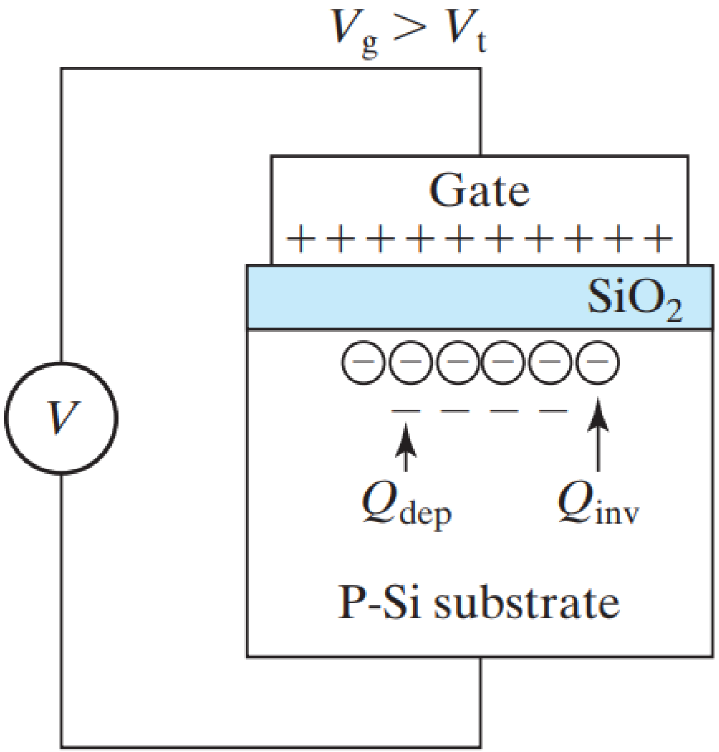
$Q_S = -\varepsilon_{si} E_s$

$\phi_S = 2|\phi_B| \rightarrow n_s=p_0$

All charges in semiconductor

Practice36: draw the band diagram for strong inversion

$V_g > V_t$



Poisson’s Equation

Physically not exist

$$\frac{d\phi}{dx} = -\mathrm{E} \qquad \frac{d^2\phi}{dx^2} = -\frac{dE}{dx} = -\frac{q}{\varepsilon_{si}}[p(x) - n(x) + \boxed{N_d^+(x)} - N_a^-(x)]$$

n_0

$$\frac{dE}{dx} = \frac{\rho}{\varepsilon_s}$$

$N_a^- = N_a$

$N_d^+ = N_d$

$$\mathrm{n}(x) = n_i e^{-(E_i - E_f)/kT} = n_i e^{(E_f - E_i)/kT} = n_i e^{q(\phi(x) - \phi_B)/kT} = \frac{n_i^2}{N_a} e^{q\phi(x)/kT}$$

$$\mathrm{p}(x) = n_i e^{(E_i - E_f)/kT} = n_i e^{q(\phi_B - \phi(x))/kT} = N_a e^{-q\phi(x)/kT}$$

$N_d^+(x)$

$$\frac{d^2\phi}{dx^2} = -\frac{q}{\varepsilon_{si}}\left[N_a\left(e^{-q\phi/kT} - 1\right) - \frac{n_i^2}{N_a}\left(e^{q\phi/kT} \boxed{- 1}\right)\right]$$

$$\frac{d^2\phi}{dx^2} * \frac{d\phi}{dx} = -\frac{q}{\varepsilon_{si}}\left[N_a\left(e^{-\frac{q\phi}{kT}} - 1\right) - \frac{n_i^2}{N_a}\left(e^{\frac{q\phi}{kT}} - 1\right)\right] * \frac{d\phi}{dx}$$

Practice37 : derive the Qs formula $\frac{d}{dx}(\frac{d\phi}{dx})^2 = 2\frac{d\phi}{dx}\frac{d^2\phi}{dx^2}$

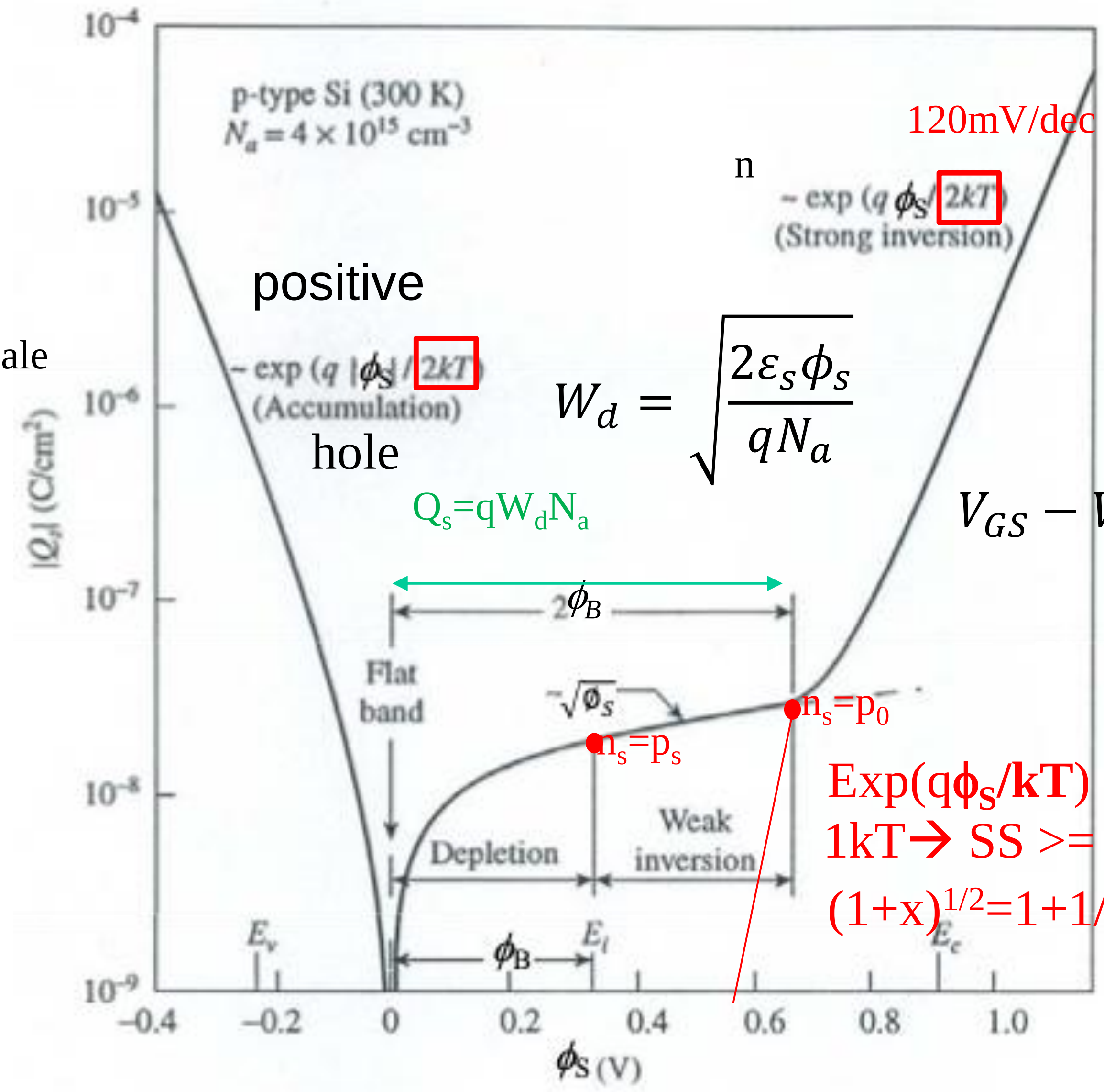
$$\frac{d^2\phi}{dx^2} * \frac{d\phi}{dx} = -\frac{q}{\varepsilon_{si}}\left[N_a\left(e^{-\frac{q\phi}{kT}} - 1\right) - \frac{n_i^2}{N_a}\left(e^{\frac{q\phi}{kT}} - 1\right)\right] * \frac{d\phi}{dx}$$

$$\frac{1}{2}\frac{d}{\cancel{dx}}(\frac{d\phi}{dx})^2 = -\frac{q}{\varepsilon_{si}}\left[N_a\left(e^{-\frac{q\phi}{kT}} - 1\right) - \frac{n_i^2}{N_a}\left(e^{\frac{q\phi}{kT}} - 1\right)\right] * \frac{d\phi}{\cancel{dx}}$$

integration

$$E^2(x) = (\frac{d\phi}{dx})^2 = \frac{2kTN_a}{\varepsilon_{si}}[\left(e^{-\frac{q\phi}{kT}} + \frac{q\phi}{kT} - 1\right) + \frac{n_i^2}{N_a^2}(e^{\frac{q\phi}{kT}} + \frac{q\phi}{kT} - 1)]$$

$$Q_S = -\varepsilon_{si}E_s = \pm\sqrt{2\varepsilon_{si}kTN_a}[\left(e^{-\frac{q\phi_S}{kT}} + \frac{q\phi_S}{kT} - 1\right) + \frac{n_i^2}{N_a^2}(e^{\frac{q\phi_S}{kT}} + \frac{q\phi_S}{kT} - 1)]^{\frac{1}{2}}$$



Log scale

positive

hole

n

120mV/dec

$$W_d = \sqrt{\frac{2\varepsilon_s\phi_s}{qN_a}}$$

$Q_s=qW_dN_a$

J=nev

$$V_{GS} - V_{fb} = V_{OX} + \phi_S$$

$n_s=p_s$

$n_s=p_0$

Exp($q\phi_S$/**kT**)

1kT→ SS >= 60mV/dec

$(1+x)^{1/2}=1+1/2x$

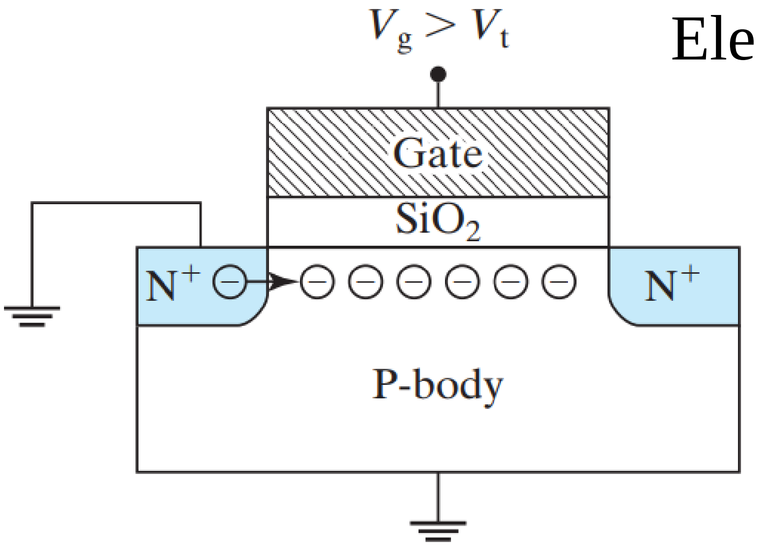
Inversion Layer Charge, Q_{inv} (C/cm^2)

Strong Inversion/I_{ON}

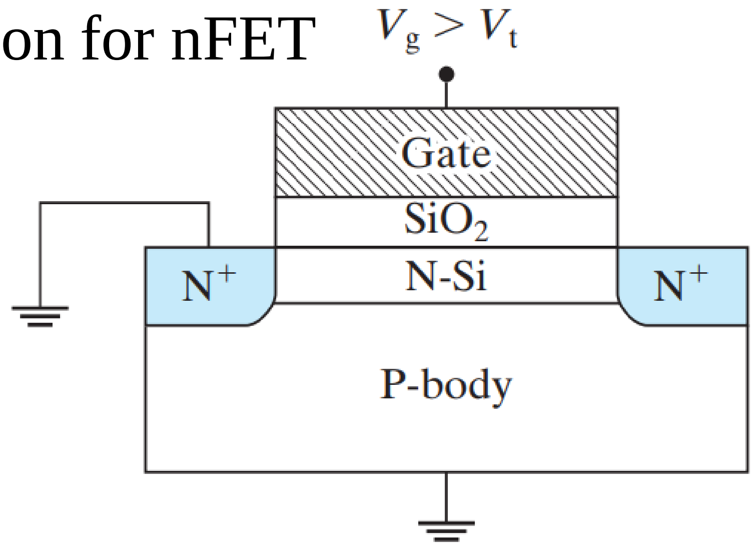
$$Q_S = -\varepsilon_{si}E_s = \pm\sqrt{2\varepsilon_{si}kTN_a}\left[\left(e^{-\frac{q\phi_S}{kT}} + \frac{q\phi_S}{kT} - 1\right) + \frac{n_i^2}{N_a^2}\left(e^{\frac{q\phi_S}{kT}} + \frac{q\phi_S}{kT} - 1\right)\right]^{\frac{1}{2}} \sim Q_{dep} + Q_{inv}$$

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

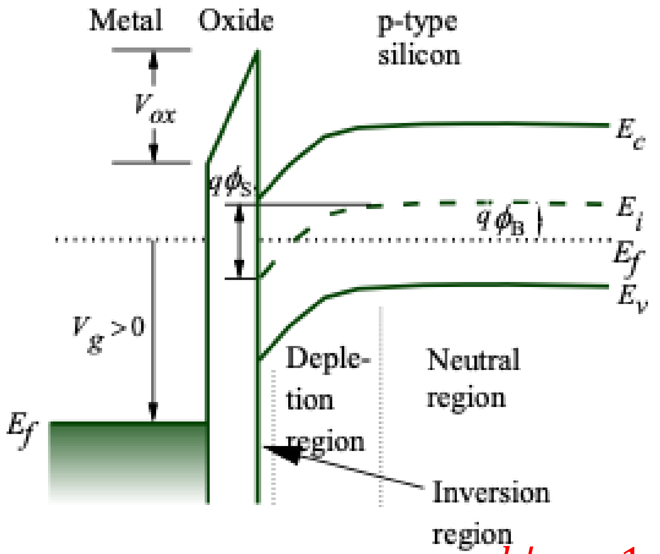
$$= V_t - \frac{Q_{inv}}{C_{ox}} \qquad \therefore \quad \boxed{Q_{inv} = -C_{ox}(V_g - V_t)}$$



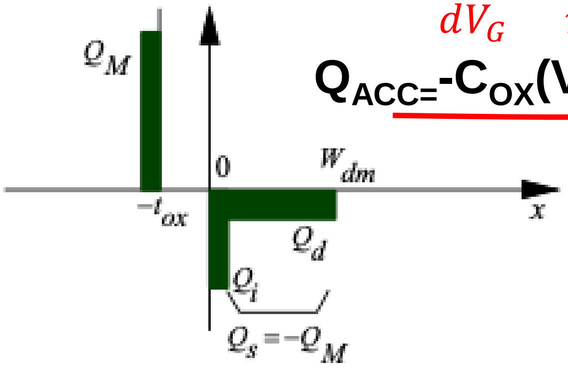
Electron for nFET



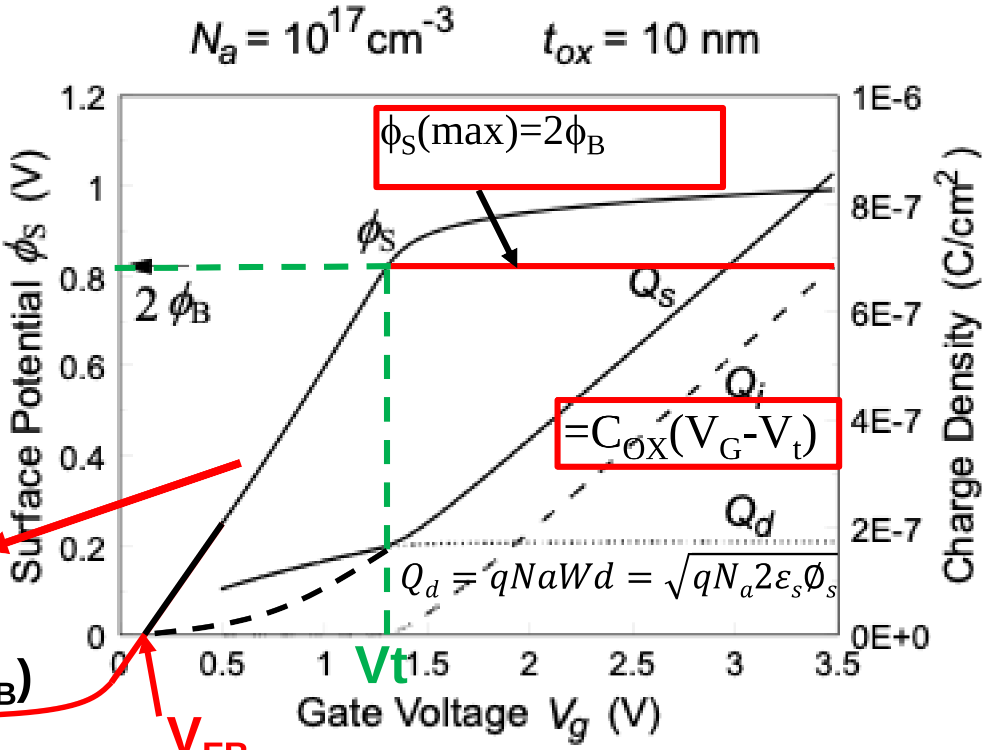
MOSFET Charge and Potential



$$\frac{d\phi_s}{dV_G} = \frac{1}{m} \le 1$$



$Q_{ACC} = -C_{OX}(V_G - V_{FB})$



$\phi_S(\max) = 2\phi_B$

$= C_{OX}(V_G - V_t)$

$Q_d = qNaWd = \sqrt{qN_a 2\varepsilon_s \phi_s}$

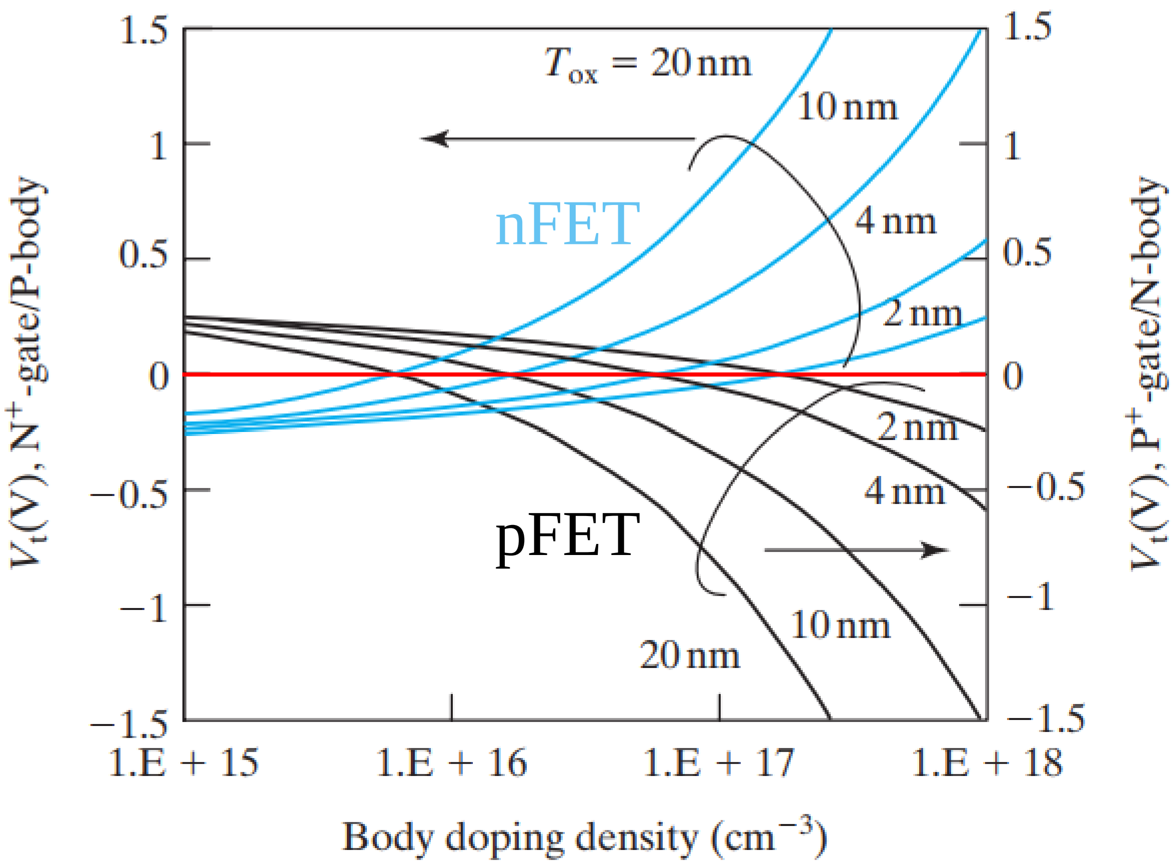
V_{FB}

Gate voltage equation:

$$V_G - V_{FB} = V_{OX} + \boxed{\phi_S} = \frac{-Q_S}{C_{OX}} + \boxed{\phi_S}$$

Threshold Voltage

90nm node:
EOT=1nm
N_B > 1E17
$V_{tn}=-V_{tp}>0$



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

+ for P-body(nFET)
– for N-body(pFET)

Planar (modify doping)

$$\mathbf{V}_{t} = \mathbf{V}_{\mathbf{fb}} + 2\phi_{\mathbf{B}} - \frac{\mathbf{q}\mathbf{N}_{\mathbf{B}}\mathbf{W}_{\mathbf{dm}}}{\mathbf{C}_{\mathbf{OX}}}$$

$$\varphi_{\mathbf{m}} - \varphi_{\mathbf{s}} - \frac{\mathbf{Q}_{\mathbf{OX}}}{\mathbf{C}_{\mathbf{OX}}}$$

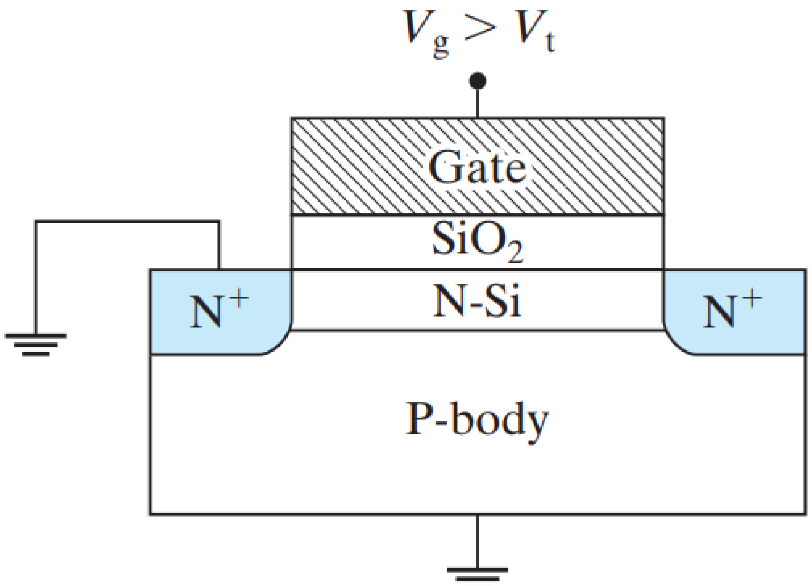
Mutiple Vt:
FinFET→ φ_{m}, $\mathbf{Q}_{\mathbf{OX}}$(dipole)
Nanosheet→ $\mathbf{Q}_{\mathbf{OX}}$

$$\mathrm{V}_{\mathrm{tn}} = V_{FB} + \frac{qW_{dm}N_{A}}{C_{ox}} + 2\phi_{B}$$

$$\mathrm{V}_{\mathrm{tp}} = V_{FB} + \frac{-qW_{dm}N_{d}}{C_{ox}} - 2\phi_{B}$$

Practice38: what is the metal WF choice for n/pFET?

5.5.1 Choice of V_t and Gate Doping Type



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

nFET

nS/D, n metal (planar)

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

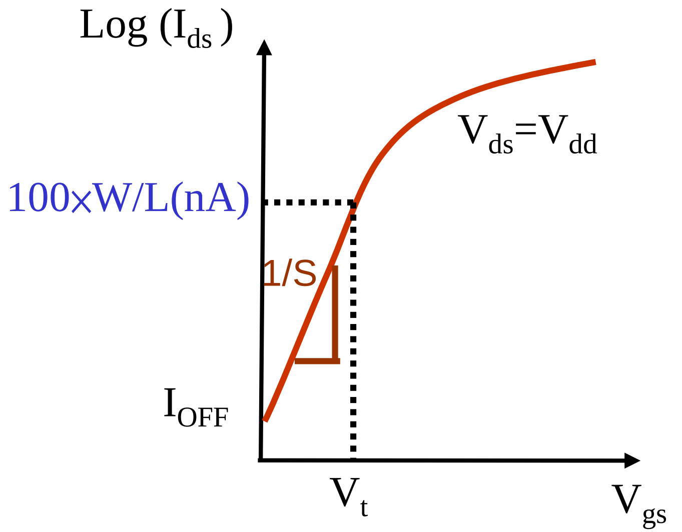
pS/D, p metal (planar)

pFET

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

- Al: work function near Si E_c → nMetal
- Pt: work function near Si E_v → pMetal
- For planar FET, nMetal is needed for Vt adjustment of nFET, pMetal is needed for Vt adjustment od pFET
- For FinFET, nMetal and pMetal have to move toward to midgap due to undoped channel, dual metal gate

Constant current V_t and Transconductance Vt



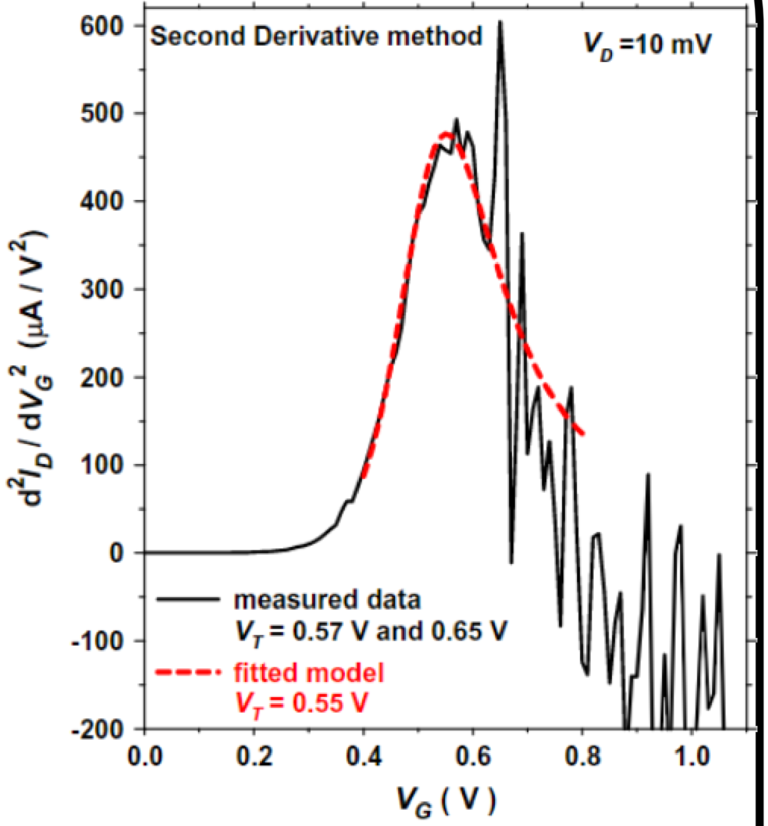
Practical definition of V_t : the V_{gs} at which I_{ds}= $\boxed{\text{100nA}}$×W/L

(Const. current)

$$I_{subthreshold}(nA) \approx 100 \times \frac{W}{L} \times e^{\frac{q(V_{gs}-V_t)}{mkT}}$$

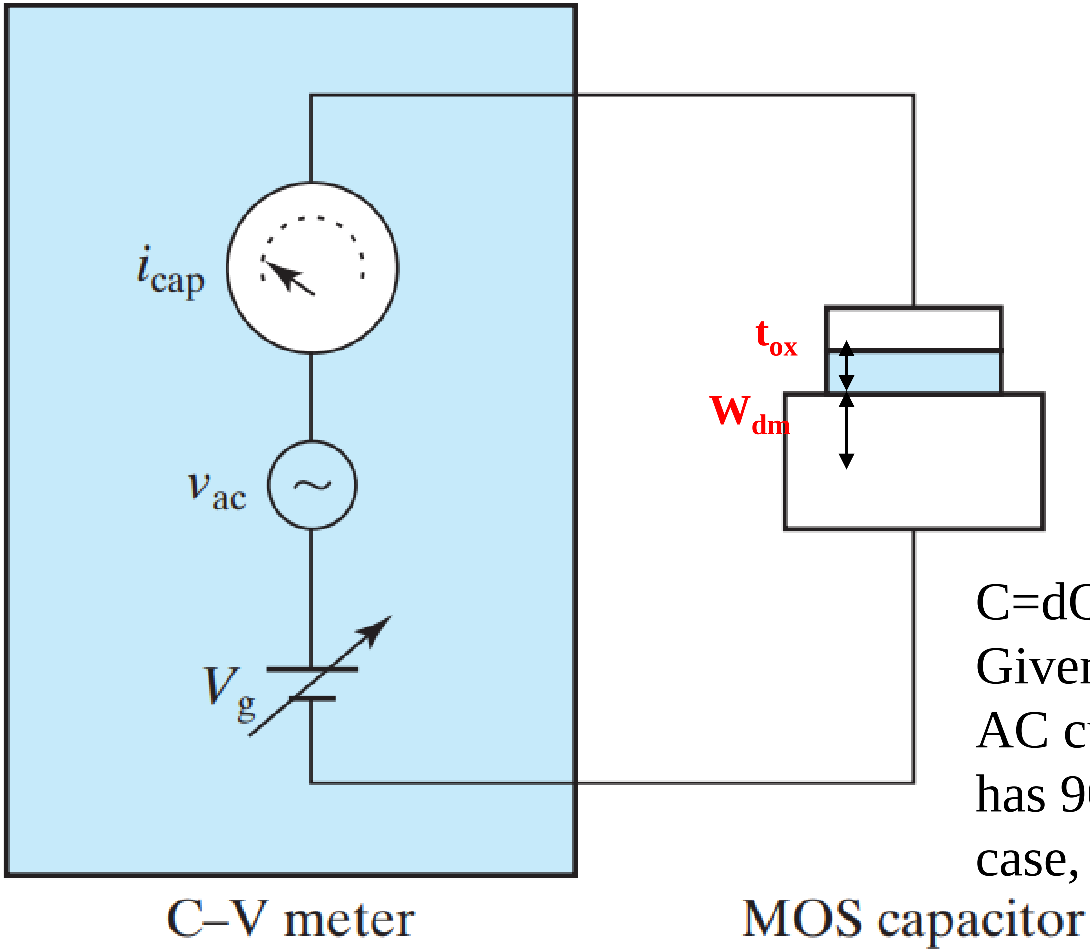
$$= 100 \times \frac{W}{L} \times 10^{(V_{gs}-V_t)/SS}$$

$$I_{OFF}(nA) = 100 \times \frac{W}{L} \times 10^{-V_t/SS}$$



V_t can be found at the maximum of d^2I_D/dV_G^2

5.6 MOS CV Characteristics



CsRs vs CpRp modes

Small signal cap

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

Vg > 0 → Qg > 0

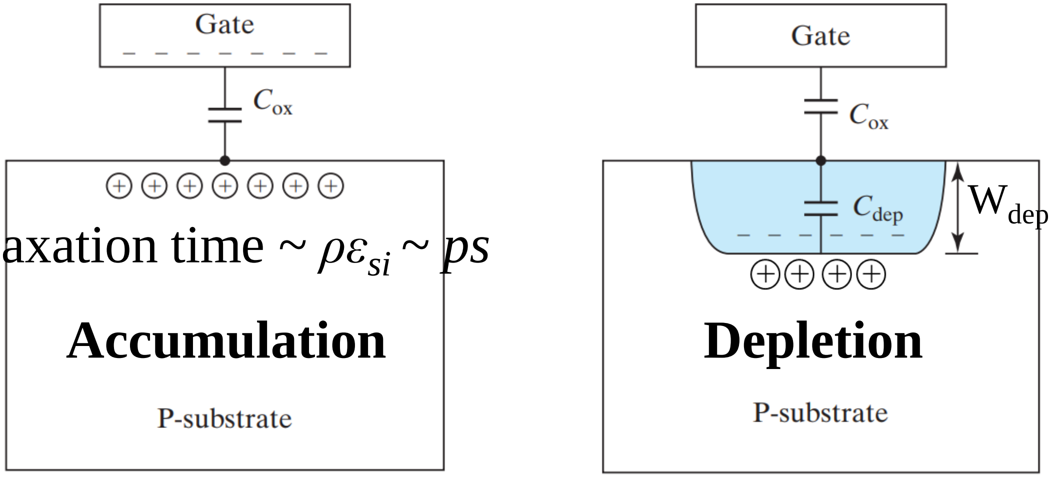
$Q_s = -Q_g$

C=dQ/dV=dQ/dt*dt/dV

Given the AC voltage, and measure the AC current. We can find out the current has 90 degree phase difference,. In this case, we the calculate cap.

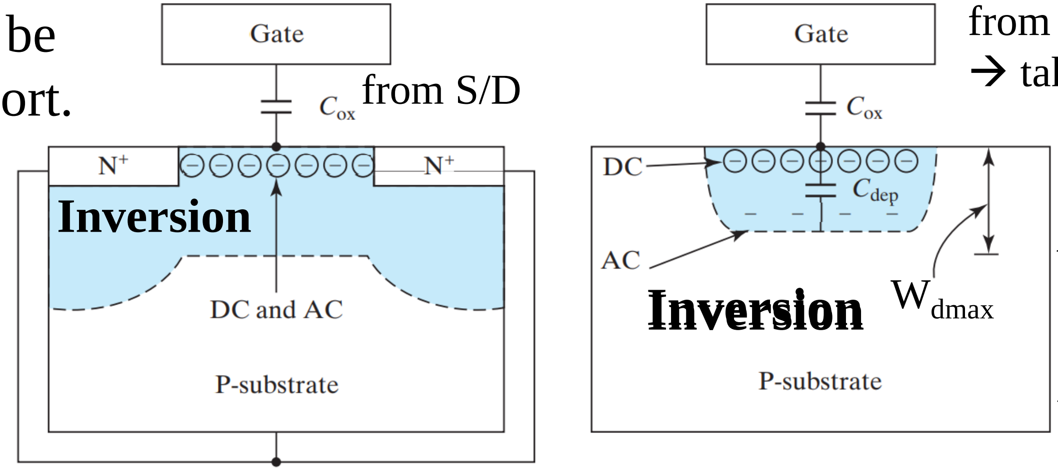
Supply of Inversion Charge May be Limited

C=dQ/dV



relaxation time ~ $\rho\varepsilon_{si}$ ~ ps

Electrons supplied by S/D would be fast if Lg is short.



from S/D

from valence band
→ takes time

If D_{it} is small, the speed of S/D supply electrons is slow.

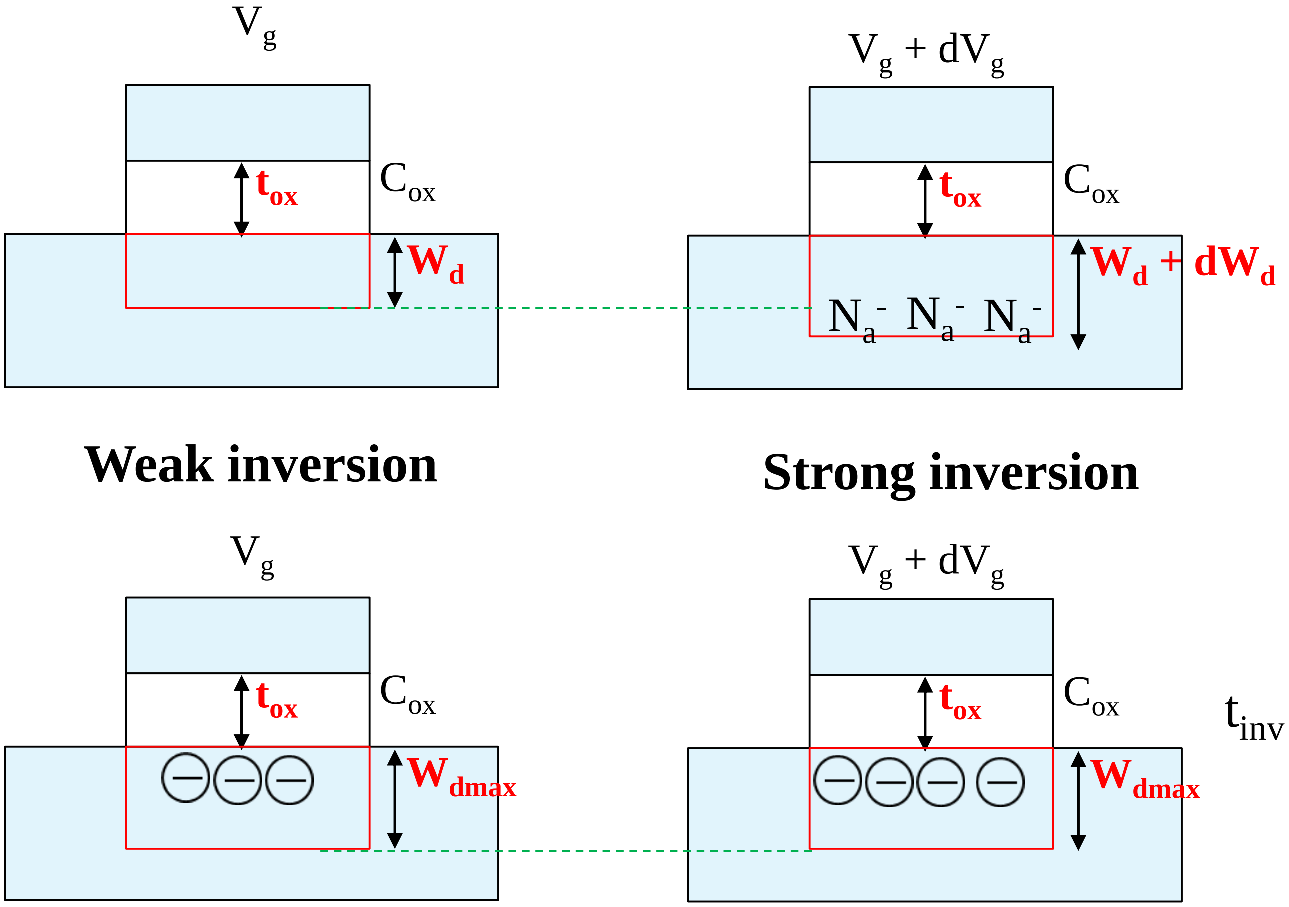
Electrons can follow AC voltage

Electrons can not follow AC voltage

$$dQ = N_a dW_d$$

Accumulation

Depletion



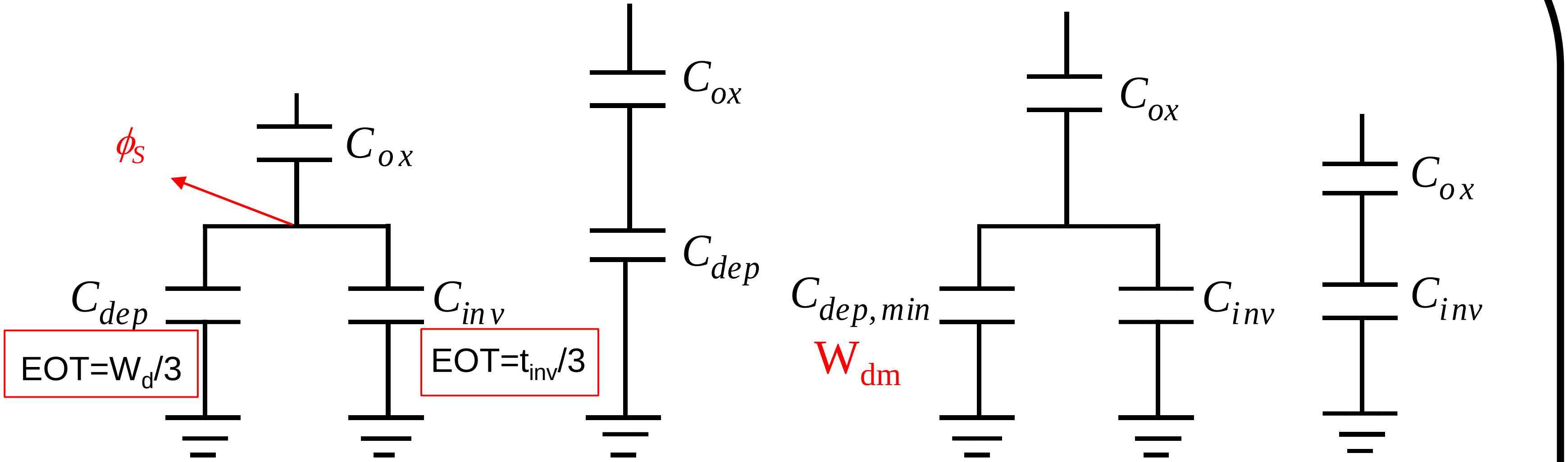
Weak inversion

Strong inversion

t_{inv} ~ 1nm

$$dQ = dN_a^- W_{dmax} = 0$$

Equivalent circuit in the depletion and the inversion regimes



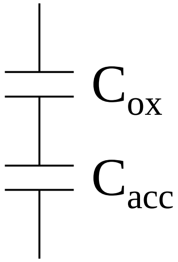
(a) General case for both depletion and inversion regions.

(b) In the depletion regions+ weak inversion

(c) $V_g \approx V_t$

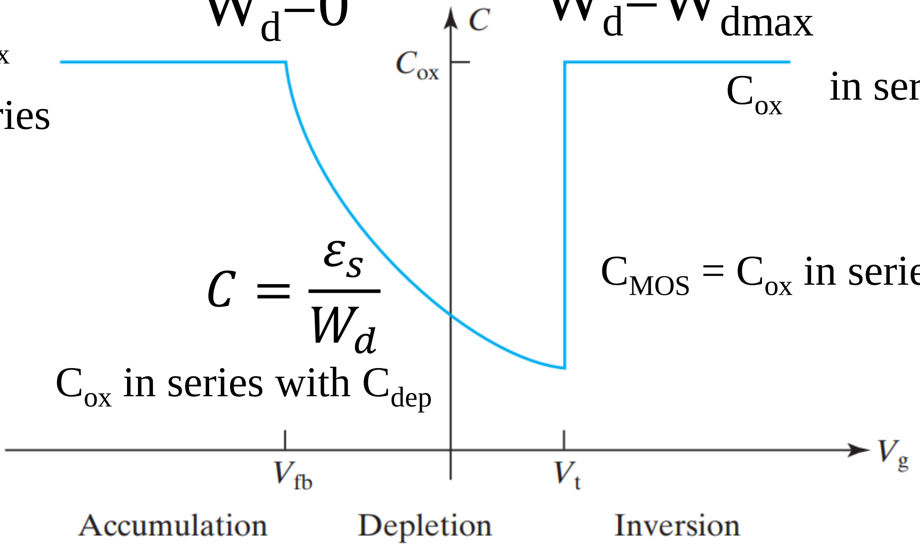
(d) Strong inversion

(e) Accumulation



Q: what about C_{Dit} ?

Low frequency



C_{ox}

C_{acc} in series

$W_d=0$

$W_d=W_{dmax}$

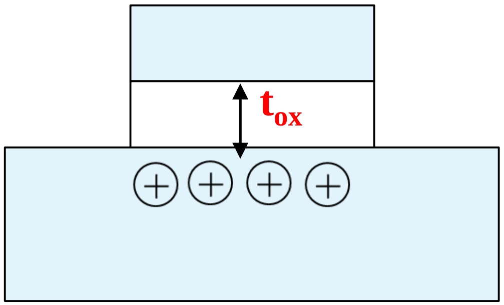
C_{ox} in series with C_{inv}

$$C = \frac{\varepsilon_s}{W_d}$$

C_{ox} in series with C_{dep}

C_{MOS} = C_{ox} in series with C_{si}

$V_g < V_{fb}$



t_{acc} ~ 1.2nm

t_{ox} = 1nm, t_{acc} not negligible

EOT=CET-0.4nm

measured

C_{si} = C_{inv} in parallel with C_{dep}=C_{inv}(large)+C_{dep}

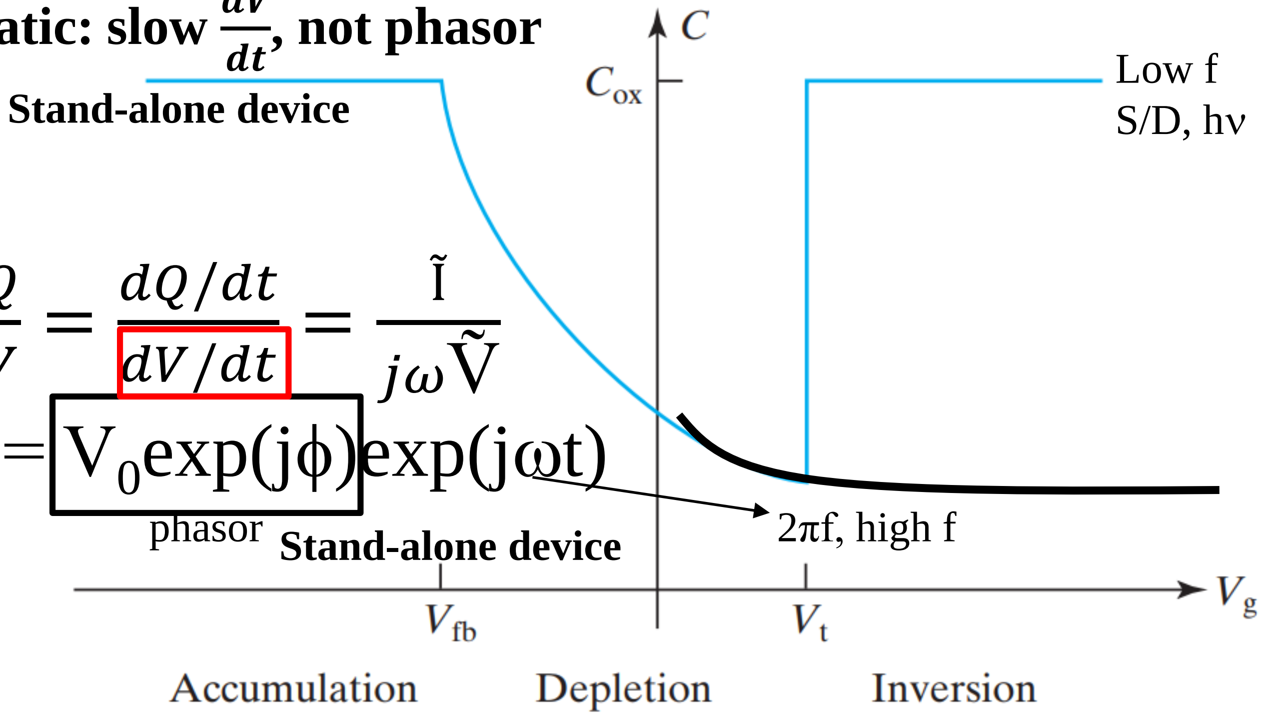
Practice39: draw the CV curve for n/pFET

Quasi-Static CV of MOS Capacitor

quasi-static: slow $\frac{dV}{dt}$, not phasor

$$C = \frac{dQ}{dV} = \frac{dQ/dt}{\boxed{dV/dt}} = \frac{\tilde{I}}{j\omega\tilde{V}}$$

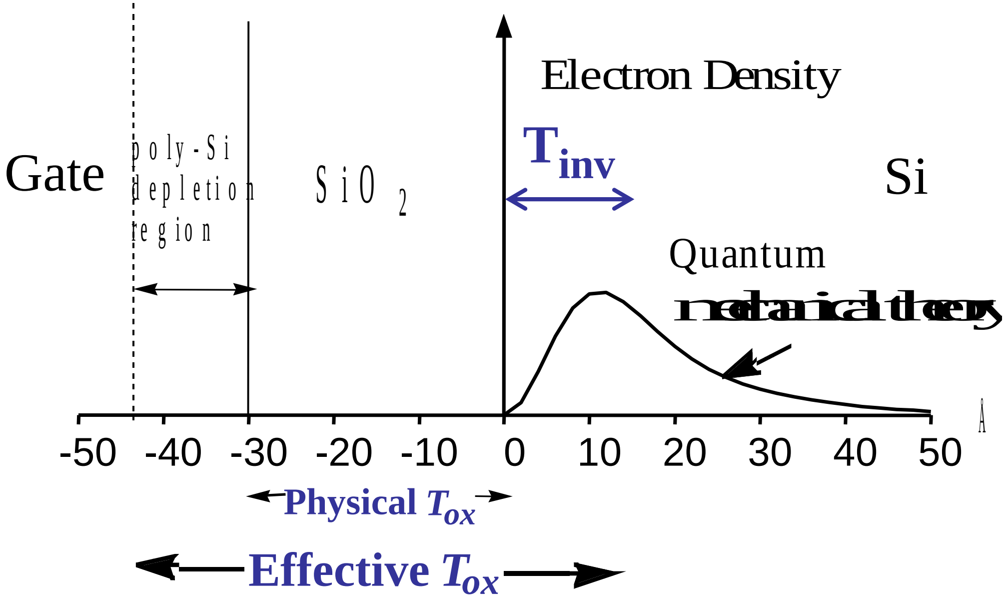
$$V = \tilde{V} = \underset{\text{phasor}}{\boxed{V_0\exp(j\phi)}}\exp(j\omega t)$$



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

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

Average inversion-layer location below the Si/SiO$_2$ interface is called the ***inversion-layer thickness, T_{inv}***.



$$\int T * |\Psi(x)|^2$$

CET = EOT + $T_{inv}/3$

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

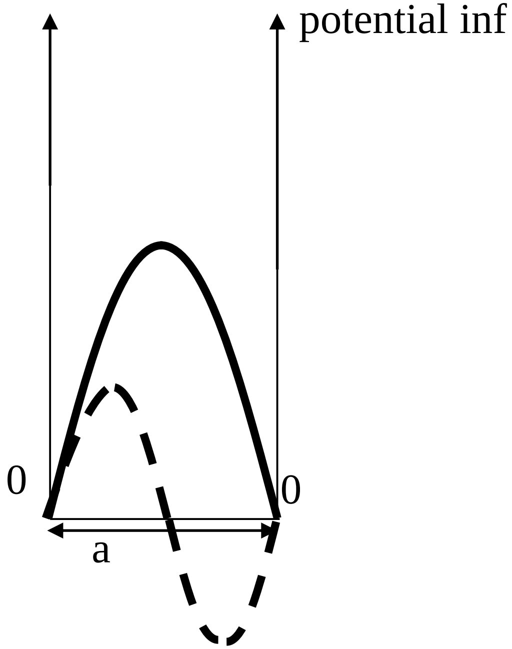
potential inf

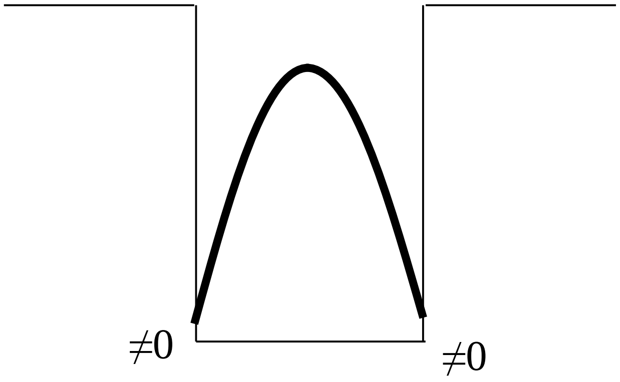
$$E = \frac{\hbar^2 k^2}{2m}$$

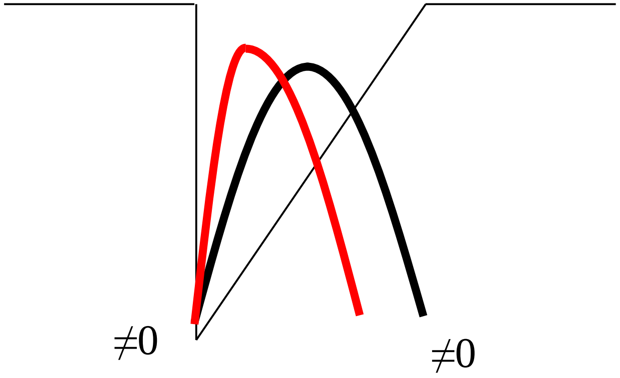
$$k = \frac{2\pi}{\lambda} = \frac{\pi}{a}$$

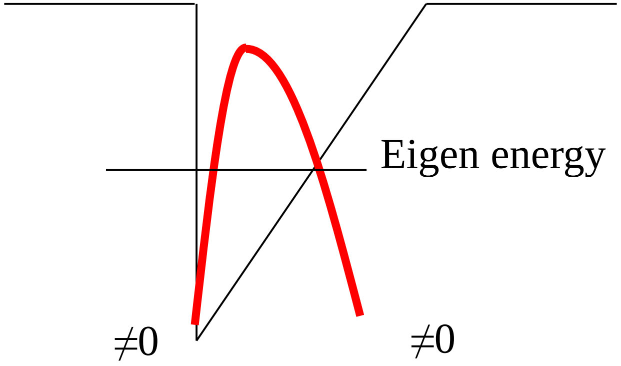
$$\lambda = 2a, a, \frac{2a}{n}$$

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



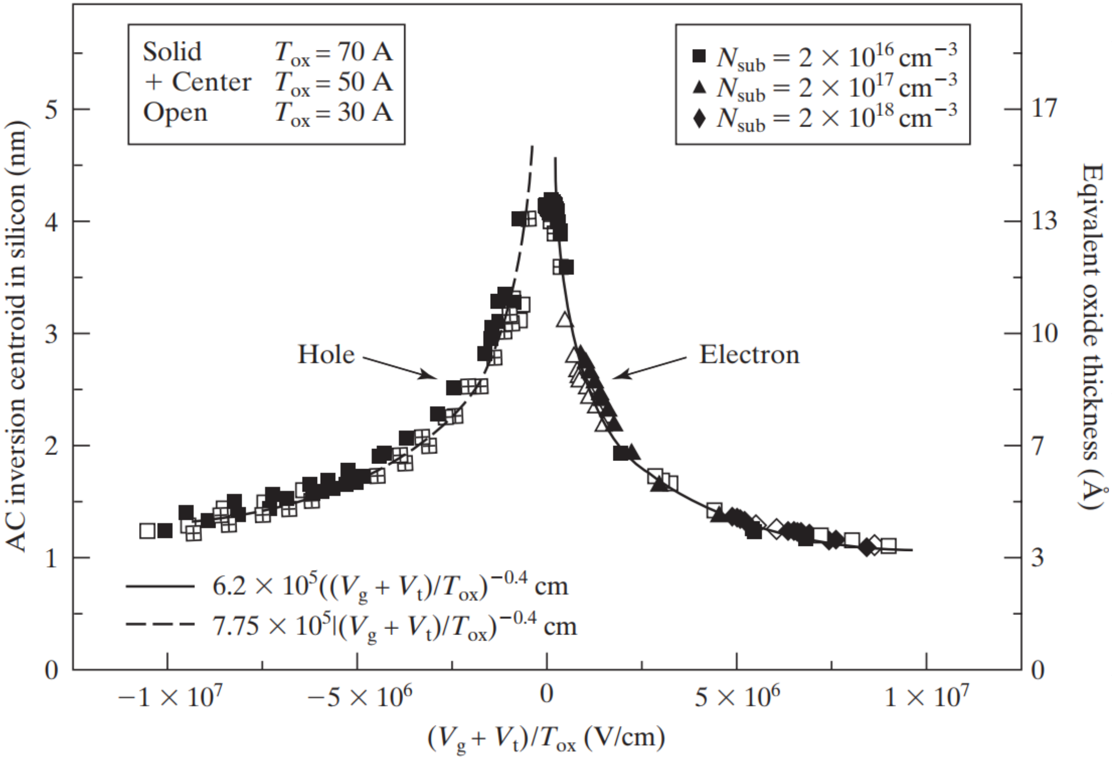






solution: if inf potential, Airy function

Electrical Oxide Thickness, T_{oxe}



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

$$\boxed{EOT = t_{ox} + t_{inv}/3}$$

The definition is different from the previous one.

$$T_{oxe} = T_{ox} + W_{poly}/3 + T_{inv}/3 \;\; at \; V_G = V_{dd}$$

Chapter 6 MOSFET

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

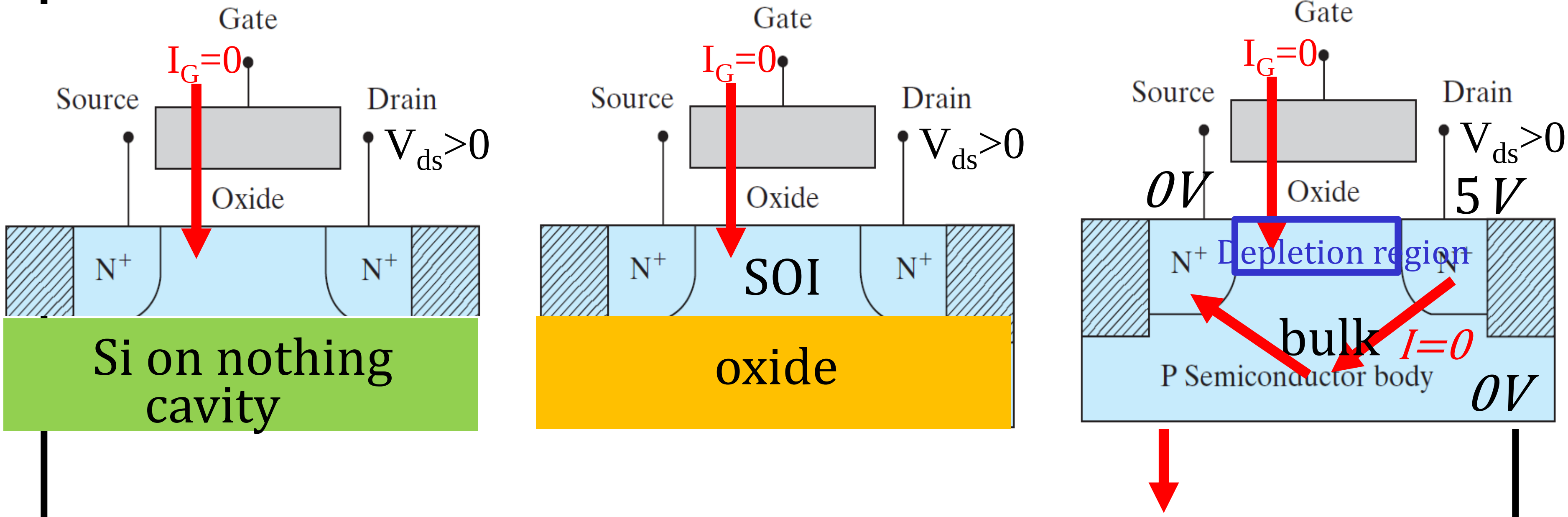
Field-Effect: inversion charge induced by gate voltage

Match the following MOSFET characteristics with their applications:

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

6.1 Introduction to the MOSFET

Basic MOSFET structure and IV characteristics



Body contact, 4 terminal device for planar
For FinFET and GAA there is no body contact

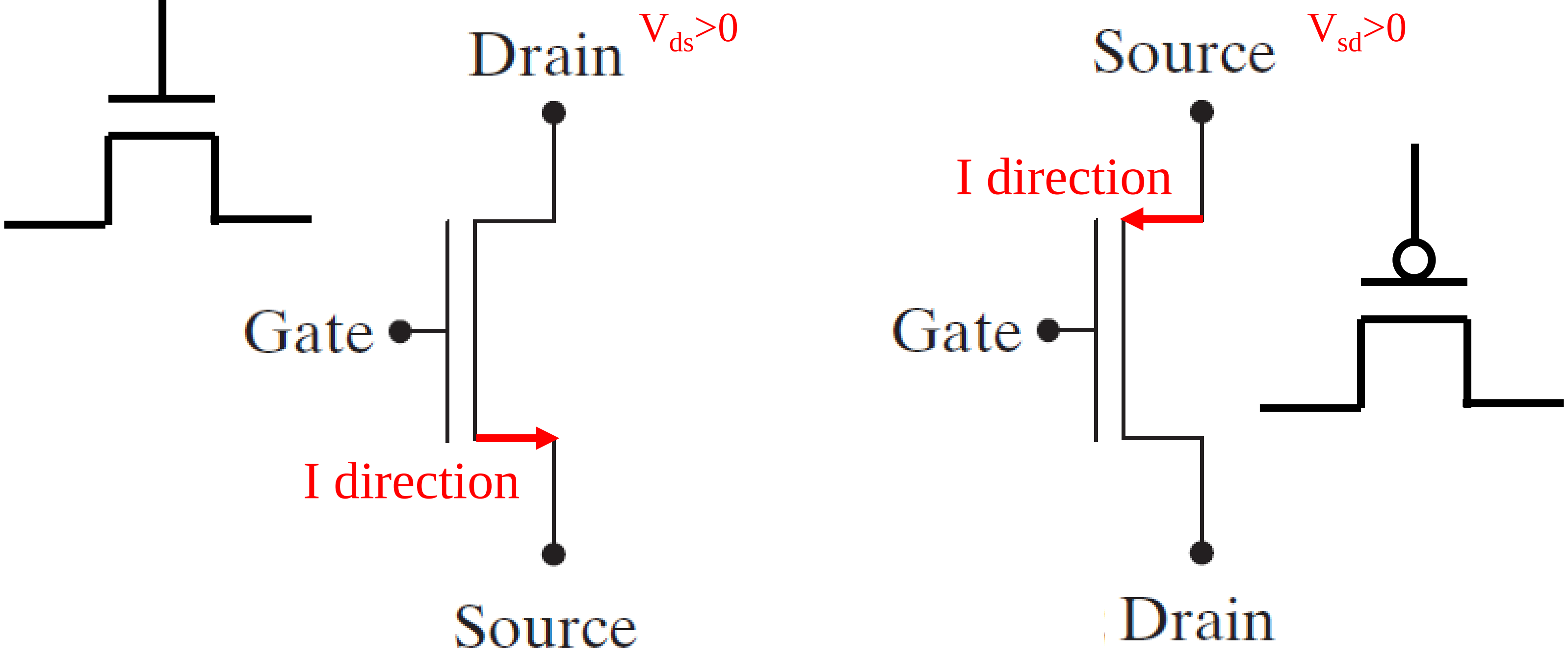
nFET or pFET is determined by S/D, not body

6.1 Introduction to the MOSFET

nFET

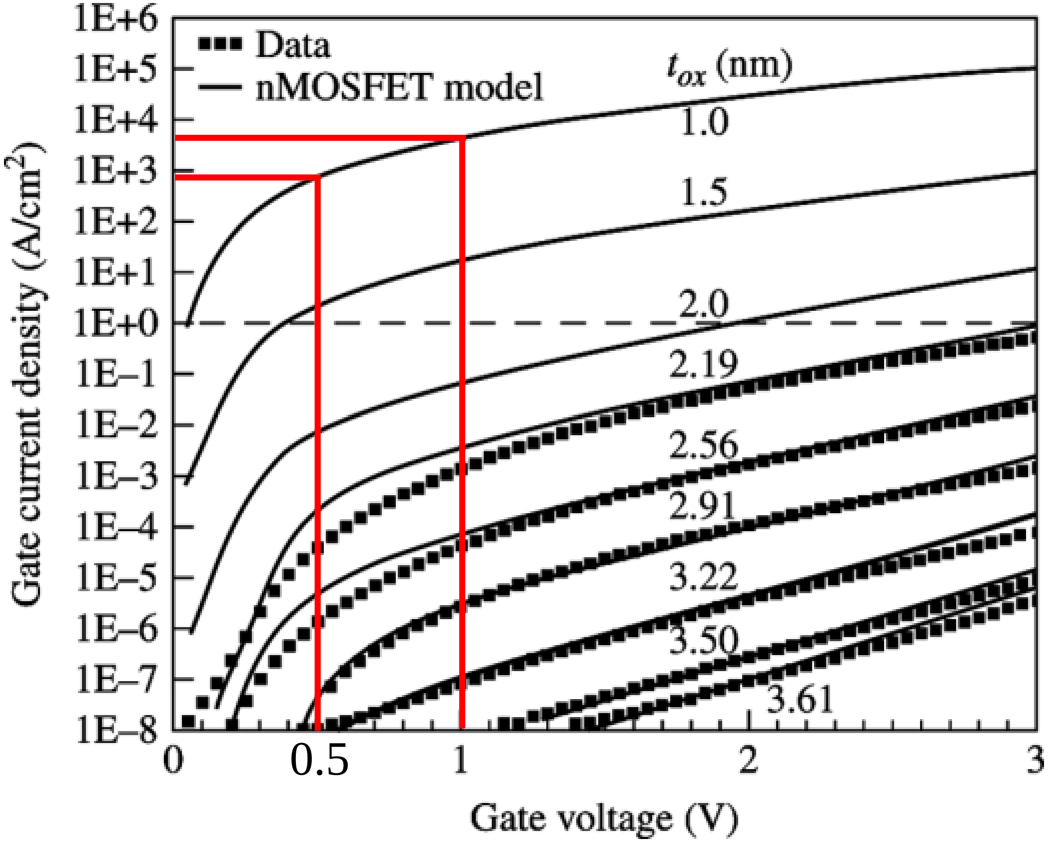
pFET

Usually high voltage is up above



- S/D is defined by circuit designer
- DTCO: Design Technology Co-Optimization

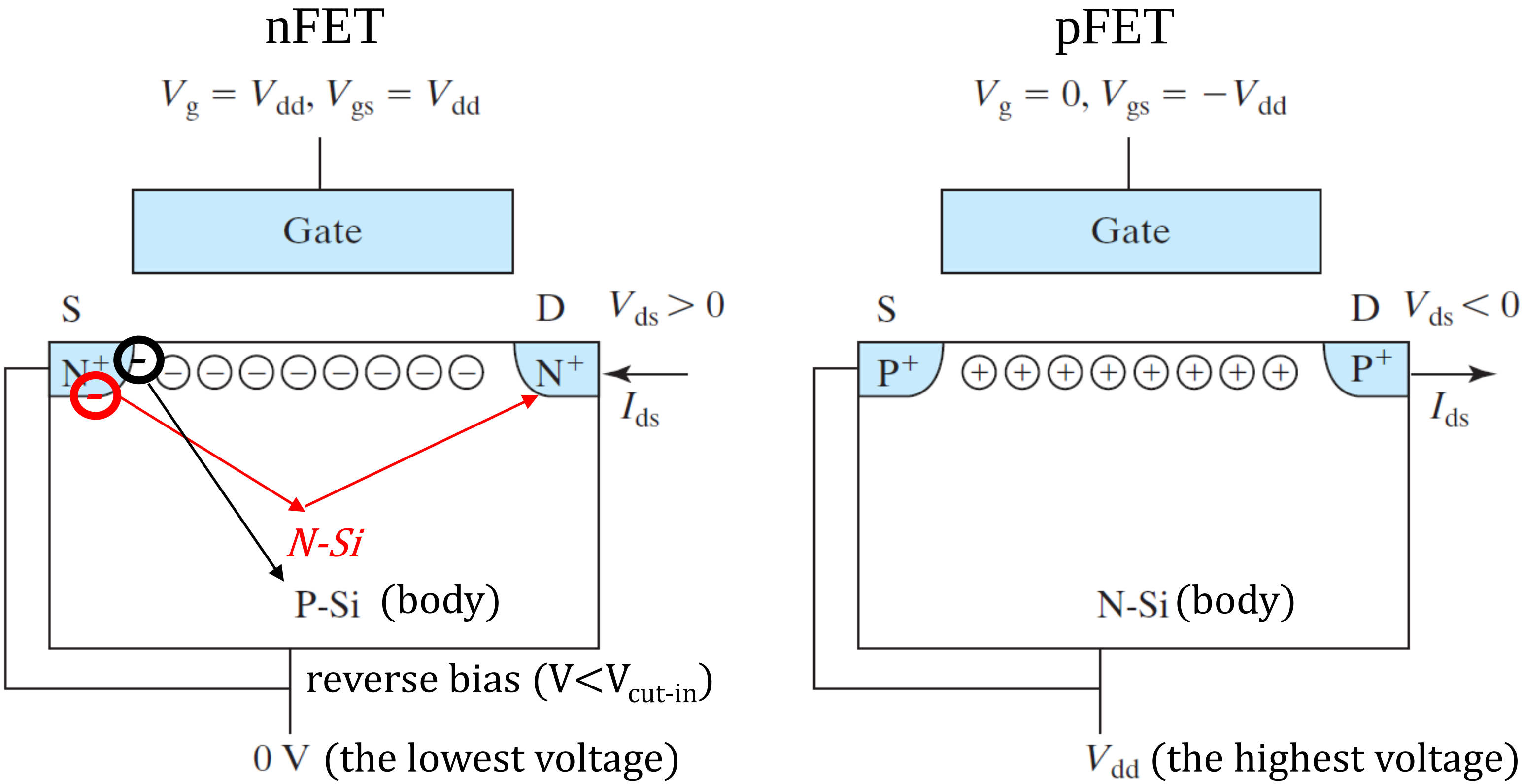
Gate leakage current



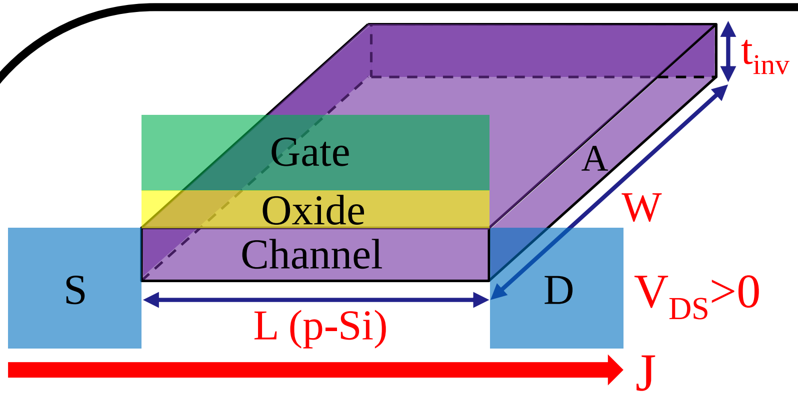
Ref : Yuan Taur

- $J_G \sim 10^3 A/cm^2$ at $V_G = 0.5V$ and $t_{ox} = 1nm$.
- $J_G \sim 10^4 A/cm^2$ at $V_G = 1V$ and $t_{ox} = 1nm$.

6.2 Complementary MOSFETs (CMOS) Technology

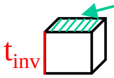


- When $V_g = V_{dd}$, the nFET is on and the pFET is off.
- When $V_g = 0$, the pFET is on and the nFET is off.
- Schottky barrier FET: Pt for pFET, Al for nFET



Practice: derive Id formula

$The\ unit\ of\ Q_{inv} = e \times n \times t_{inv}$



$\times t_{inv} \rightarrow C/cm^3 \times cm = C/cm^2$

$I_D = \boldsymbol{J} \times A = \boldsymbol{nev} \times W t_{inv}, n:\ cm^{-3}, e: C, v:\ cm/s\ , J: A/cm^2$

$$\because Q_{inv} = e \times n \times t_{inv}, v = \mu E = \mu \frac{V_{DS}}{L} \rightarrow I_D = Q_{inv} \times \mu \frac{W}{L} \times V_{DS}$$

$$Also, Q_{inv} = C_{ox}(V_{GC} - V_t), C_{ox}:\ F/cm^2$$

$$\because V_{GS} > V_{GD} \rightarrow Q_{inv}(S) > Q_{inv}(D) \rightarrow C_{ox}(V_{GS} - V_t) > C_{ox}(V_{GD} - V_t)$$

$$Average\ Q_{inv} = \frac{1}{2}[Q_{inv}(S) + Q_{inv}(D)] = \frac{1}{2}[C_{ox}(V_{GS} - V_t) + C_{ox}(V_{GD} - V_t)]$$

$$Take\ source\ as\ ref \rightarrow V_{GD} = V_{GS} - V_{DS} \rightarrow Q_{inv} = \frac{C_{ox}}{2}[2(V_{GS} - V_t) - V_{DS}]$$

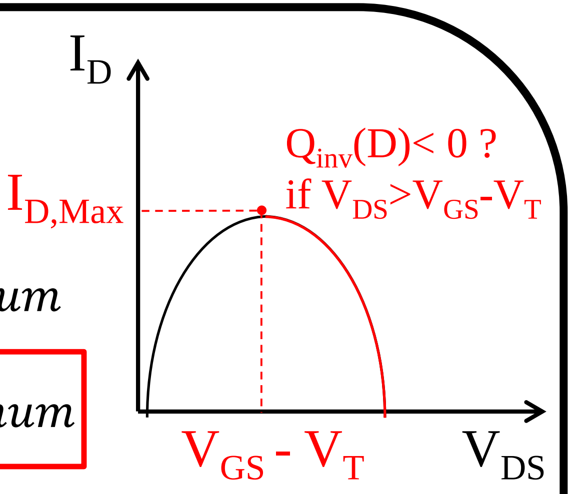
$\therefore I_D = \frac{1}{2}\mu C_{ox}\frac{W}{L}[2(V_{GS} - V_t)V_{DS} - V_{DS}^2]$ source reference (SPICE)

Practice: explain saturation region and channel length modulation

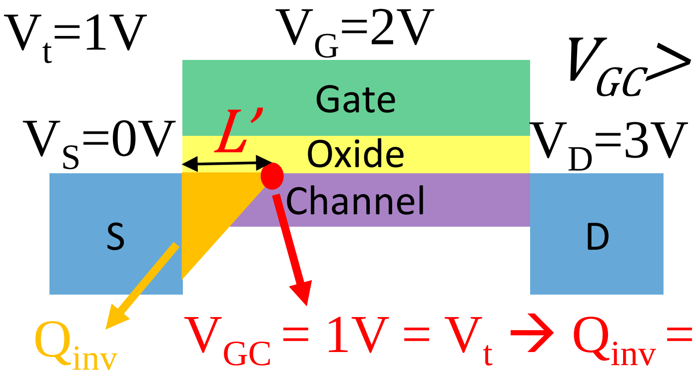
$$I_D=\frac{1}{2}\mu C_{ox}\frac{W}{L}[2(V_{GS}-V_t)V_{DS}-V_{DS}^2]$$

$$\because f(x)=ax^2+bx+c\rightarrow at\ x=-\frac{b}{2a}, f(x)\ has\ maximum$$

$$\therefore At\ V_{DS}=V_{GS}-V_t, \boxed{I_D=\frac{1}{2}\mu C_{ox}\frac{W}{L'}(V_{GS}-V_t)^2\ is\ maximum}$$



Let V_t=1V



$V_{GC} > V_t$ → Q_{inv} *exists at* $V_{DS} < V_{GS} - V_t$

At source: V_{GC} = 2V → Q_{inv} exists

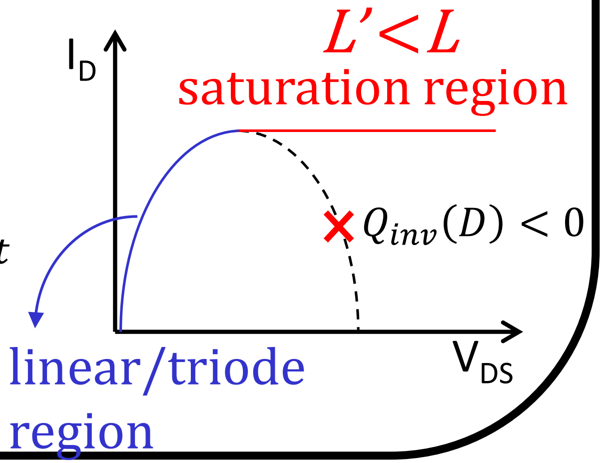
At drain: V_{GC} = -1V → Q_{inv} = 0(accumulation)

$$\because Q_{inv}(D)<0\ is\ wrong\ (Charge\ can't\ be<0)$$

$$\therefore I_D\ will\ keep\ at\ maximum\ after\ V_{DS}=V_{GS}-V_t$$

V_{GS}-V_t is define as V_{ov} (overdrive voltage)

V_{DC} /(L-L’) is large enough, drift velocity is large



I_{ON} Improvement -- Overview

Practice40: repeat this page

$$I_d = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{gs} - V_t)^2 \times \text{n (floor number)}$$

→ Nanosheet (N2)

FinFET (16nm) : $2H_{FIN}+W_{FIN}$

Intel : 22nm

High mobility $\frac{\varepsilon_{ox}}{t_{ox}} \rightarrow HK(45nm\ node)$

$$\frac{\varepsilon_{ox}}{t_{ox}} = \frac{\varepsilon_{HK}}{t_{HK}} = \frac{\varepsilon_{ox}}{EOT} \quad EOT = t_{HK}\frac{\varepsilon_{ox}}{\varepsilon_{HK}} \qquad \varepsilon = k\varepsilon_0$$

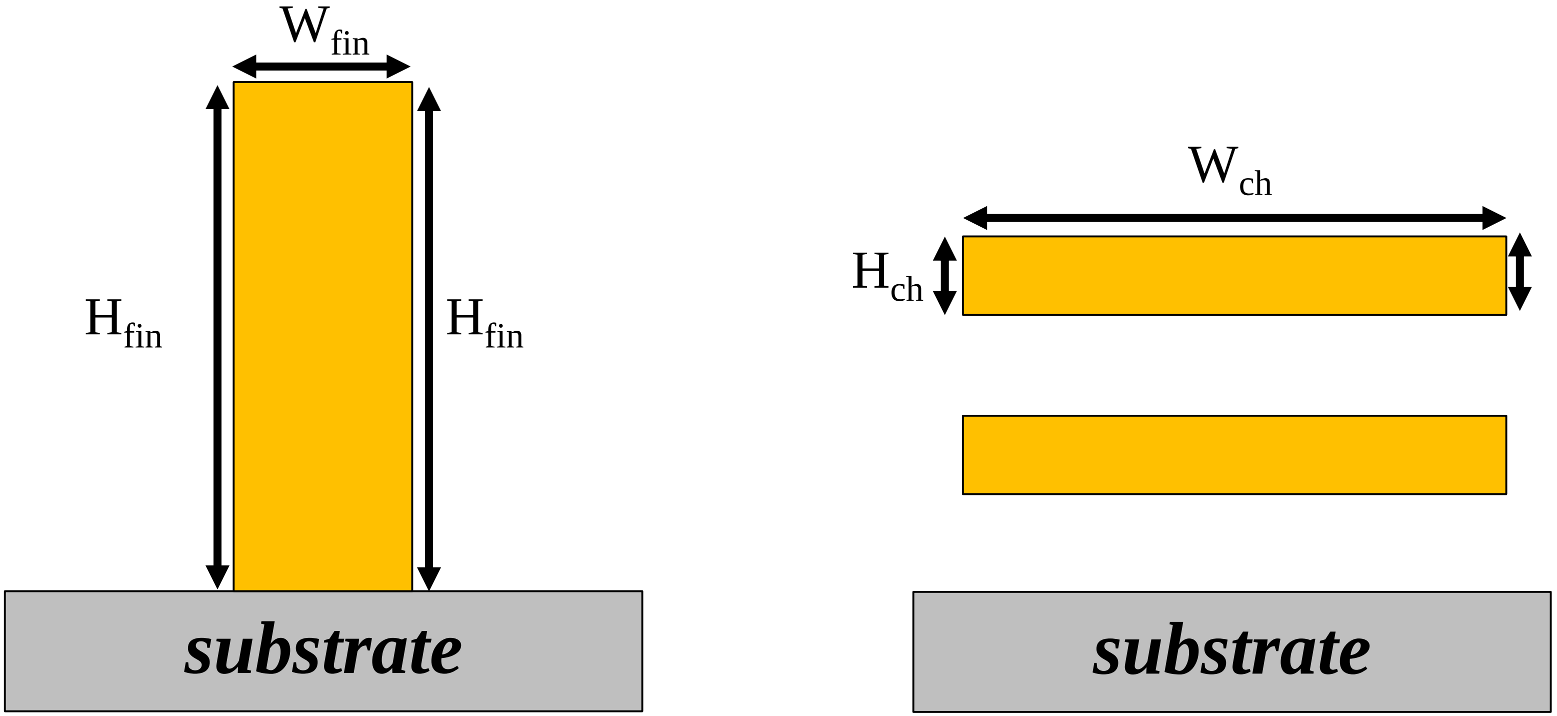
t_{ox} >1nm. Otherwise, I_g is significant

1. Strained Si (90nm)
2. High mobility channel (5nm SiGe p-channel)

I_{ON} Improvement -- W_{eff}

FinFET = Tri-gate

Nanosheet = GAA

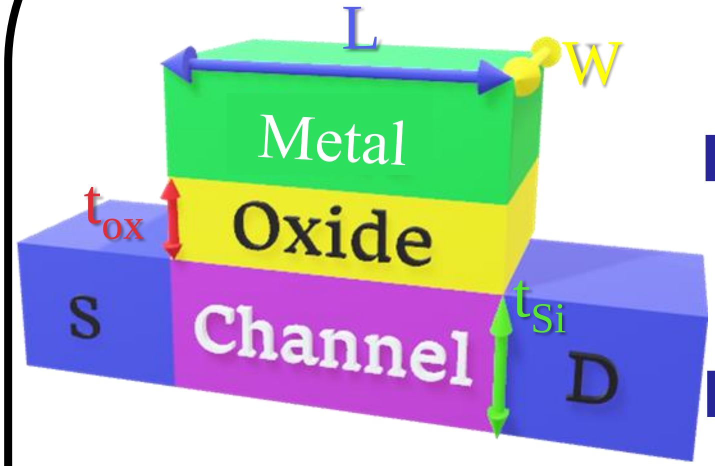


$W_{eff} = 2H_{fin} + W_{fin}$

$W_{eff} = (2H_{ch} + 2W_{ch})$ x floor number

Footprint = W_{fin} x L

Scaling length



Ref: J.-P. COLINGE, ROMJIST, Volume 11, Number 1, 2008, 3–15

$$\boxed{C_{GC} = \frac{\varepsilon_{ox} \times A}{d} = \frac{\varepsilon_{ox} \times W \times L}{t_{ox}}}$$

$$\boxed{C_{\mathrm{DS}} = \frac{\varepsilon_{Si} \times A'}{d'} = \frac{\varepsilon_{Si} \times W \times t_{Si}}{L}}$$

$$\therefore \frac{\varepsilon_{ox} \times W \times L}{t_{ox}} > \frac{\varepsilon_{Si} \times W \times t_{Si}}{L}$$

$$\rightarrow L^2 > \frac{\varepsilon_{Si}}{\varepsilon_{ox}} \times t_{Si} \times t_{ox}$$

∵Gate control needs to > S/D effect

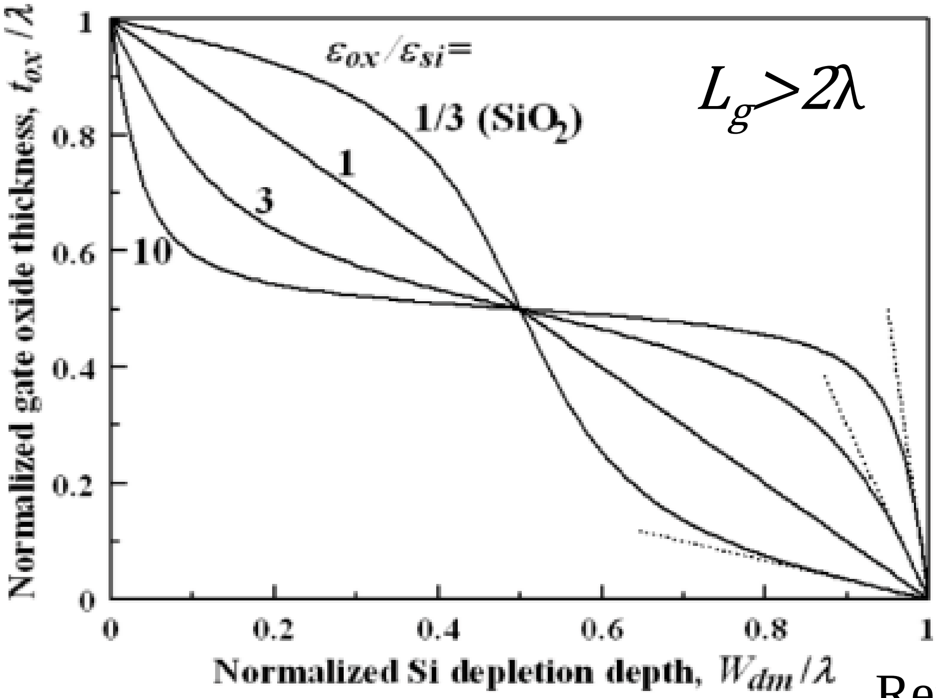
$\rightarrow C_{GC} > C_{SD}$

Vertical E field > Lateral E field

$$\boxed{L > \sqrt{\frac{\varepsilon_{Si}}{\varepsilon_{ox}} \times t_{Si} \times t_{ox}} = l(scaling\ length)}$$

L > l the more excess→ DIBL ($\frac{\Delta V_t}{\Delta VDS}$), SCE (L↓→$V_t$↓), SS (small) better

Scaling length



$L_g > 2\lambda$

$W_{dm} = t_{si}$

- In the lower right corner, $\lambda \approx W_{dm} + \overset{=3}{\boxed{(\varepsilon_{si}/\varepsilon_{ox})}} t_{ox}$

So high-k gate insulator helps.

- But it is only valid when $t_{ox} << \lambda$.

- t_{ox}/λ is limited to ½ , i.e., $\lambda > 2\, t_{ox}$.

Ref: Yuan Taur

$$\boldsymbol{W_{dm} = \sqrt{\frac{2\varepsilon_{si} 2\phi_B}{qN_B}}}$$

$N_B = 10^{17} cm^{-3} \rightarrow W_{dm} \sim 100nm$
$N_B = 10^{19} cm^{-3} \rightarrow W_{dm} \sim 10nm$
$N_B \uparrow \rightarrow \ \mu \downarrow \rightarrow \textit{FinFET is undoped}$

How will the scaling length change if the channel material is Ge?

$$C_{GC} = \frac{\varepsilon_{ox} \times W \times L_{Ge}}{t_{ox}} \quad C_{SD} = \frac{\varepsilon_{Ge} \times W \times t_{Ge}}{L}$$

$$\because C_{GC} > C_{SD}$$

$$\rightarrow \frac{\varepsilon_{ox} \times W \times L_{Ge}}{t_{ox}} > \frac{\varepsilon_{Ge} \times W \times t_{Ge}}{L_{Ge}}$$

$$\rightarrow L_{Ge}^2 > \frac{\varepsilon_{Ge}}{\varepsilon_{ox}} \times t_{Ge} \times t_{ox}$$

$$\rightarrow L_{Ge} > \sqrt{\frac{\varepsilon_{Ge}}{\varepsilon_{ox}} \times t_{Ge} \times t_{ox}}$$

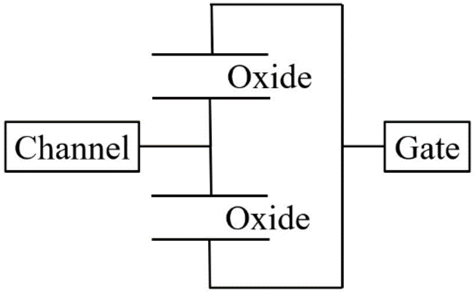
$$\because \varepsilon_{Ge}(= 16) > \varepsilon_{Si}(= 11.9)$$

$$L_{Ge} = \sqrt{\frac{16}{11.9}} \times L_{Si} > L_{Si}$$

$$\therefore \mathit{channel\ length\ will\ be\ longer}$$

How will the scaling length change in the double gate structure?

$\because$ *Capacitance in parallel*



$$C_{GC(double\ gate)} = \boxed{2} C_{GC(single\ gate)} = \boxed{2} \frac{\varepsilon_{ox} \times W \times L}{t_{ox}}$$

$$\because C_{GC(double\ gate)} > C_{SD}$$

$$\rightarrow 2 \times \frac{\varepsilon_{ox} \times W \times L_{double}}{t_{ox}} > \frac{\varepsilon_{Si} \times W \times t_{Si}}{L_{double}}$$

$$\rightarrow L_{double}^2 > \frac{1}{2} \times \frac{\varepsilon_{Si}}{\varepsilon_{ox}} \times t_{Si} \times t_{ox}$$

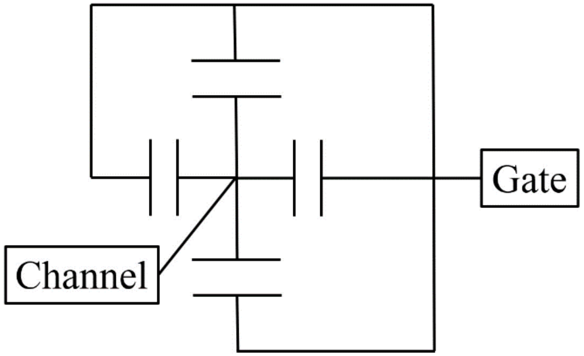
$$\rightarrow L_{double} > \sqrt{\frac{1}{2} \times \frac{\varepsilon_{Si}}{\varepsilon_{ox}} \times t_{Si} \times t_{ox}} \sim 0.707 L_{single}$$

$\therefore$ *channel length can be shorter*

How will the scaling length change in the GAA structure?

$\because$ *Capacitance in parallel*

Assume the four side length of the channel are equal.



$$C_{GC(GAA)} = 4C_{GC(single\ gate)} = 4 \times \frac{\varepsilon_{ox} \times W \times L}{t_{ox}}$$

$$\because C_{GC(GAA)} > C_{SD}$$

$$\rightarrow 4 \times \frac{\varepsilon_{ox} \times W \times L_{GAA}}{t_{ox}} > \frac{\varepsilon_{Si} \times W \times t_{Si}}{L_{GAA}}$$

$$\rightarrow L_{GAA}^2 > \frac{1}{4} \times \frac{\varepsilon_{Si}}{\varepsilon_{ox}} \times t_{Si} \times t_{ox}$$

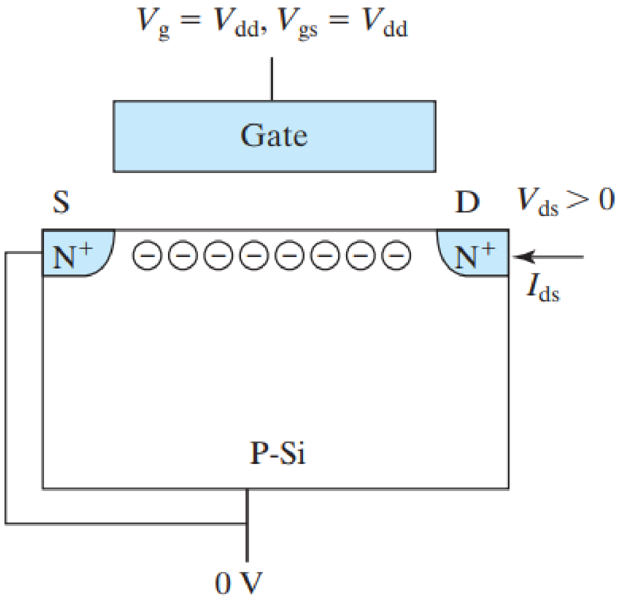
$$\rightarrow L_{GAA} > \sqrt{\frac{1}{4} \times \frac{\varepsilon_{Si}}{\varepsilon_{ox}} \times t_{Si} \times t_{ox}} = \frac{1}{2} l_{single}$$

$\therefore$ *channel length can be shorter*

6.3 Surface Mobilities and High-Mobility FETs

6.3.1 Surface Mobilities < Bulk mobility

due to surface roughness scattering



$\boxed{split\ CV}: J = ne\mu E = \frac{V_{ds}}{L}$

$+IV$

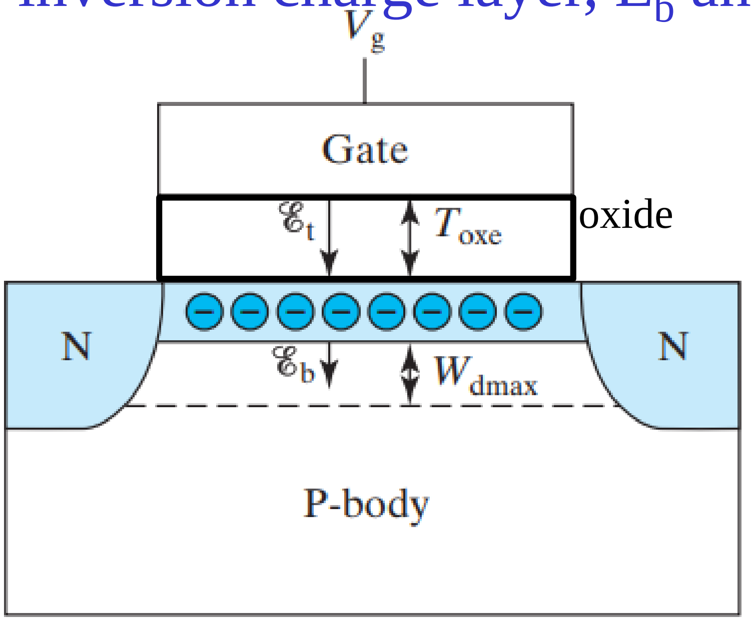
How to measure the surface mobility:

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

$$= WC_{oxe}\left(V_{gs} - V_t\right)\mu_{ns}V_{ds}/L$$

Field effect mobility

Mobility is a function of the average of the fields at the bottom and the top of the inversion charge layer, E_b and E_t .



From Gauss’s Law,

$$E_b = -Q_{dep}/\varepsilon_{si}$$

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

Therefore,

$$E_b = \frac{C_{ox}}{\varepsilon_{si}}(V_t - V_{fb} + \phi_s)$$

$$E_t = -\left(Q_{dep} + Q_{inv}\right)/\varepsilon_{si}$$

$$= E_b - Q_{inv}/\varepsilon_{si} = E_b + \frac{C_{ox}}{\varepsilon_{si}}(V_{gs} - V_t)$$

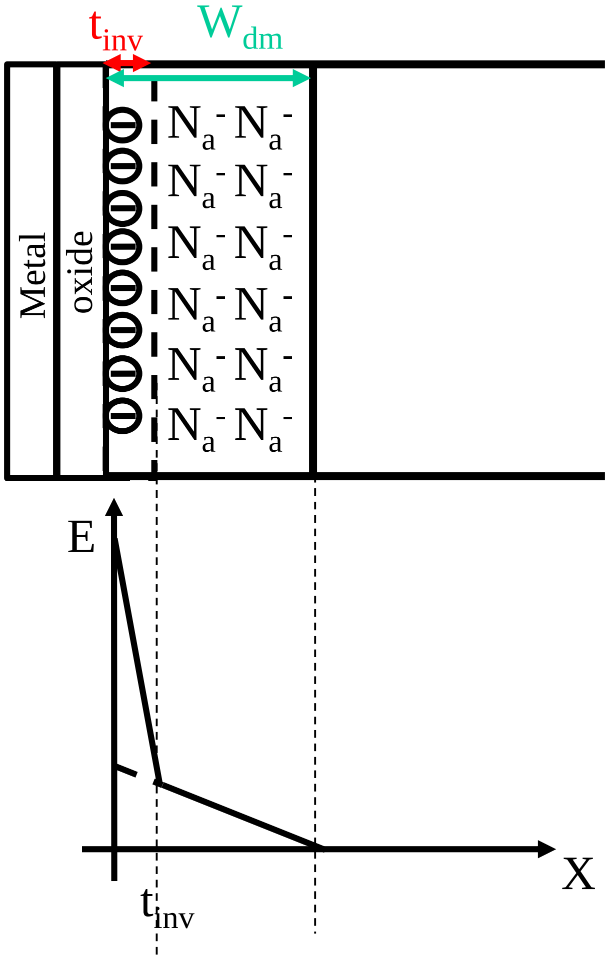
$$= \frac{C_{ox}}{\varepsilon_{si}}(V_{gs} - V_{fb} - \phi_s)$$

$$\therefore \frac{1}{2}(E_b + E_t) = \frac{C_{ox}}{2\varepsilon_{si}}(V_{gs} + V_t - 2V_{fb} - 2\phi_s)$$

$$\approx \frac{C_{ox}}{2\varepsilon_{si}}(V_{gs} + V_t + 0.2V)$$

$$= \frac{V_{gs} + V_t + 0.2V}{6t_{ox}}$$

Simple $E_{average}$ model for On state



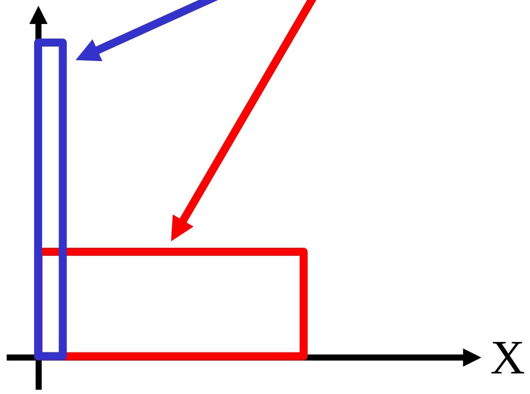
$$E(0) = (Q_{dep} + Q_{inv})/\varepsilon_{si}$$

$$E(t_{inv}) = Q_{dep}/\varepsilon_{si}$$

$$E_{avreage} = (E(0) + E(t_{inv}))/2$$

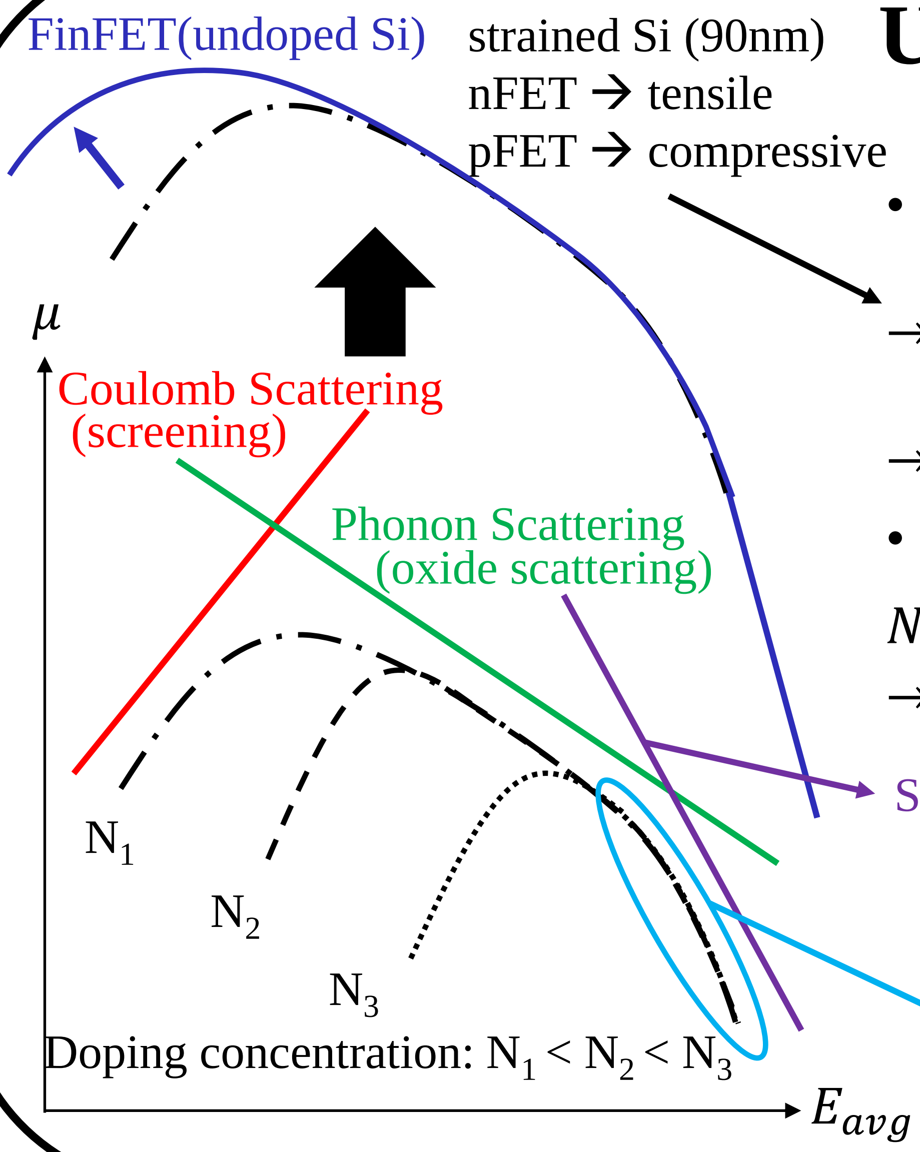
$$= \frac{2Q_{dep} + Q_{inv}}{2\varepsilon_{si}}$$

By split CV



$$\nabla E = -\rho/\varepsilon_{si} = dE/dx$$

Universal Mobility



strained Si (90nm)
nFET → tensile
pFET → compressive

- *As device scaling*

$$\rightarrow L > \sqrt{\frac{\varepsilon_{Si}}{\varepsilon_{ox}} \times t_{Si} \times t_{ox}}$$

$$\rightarrow N_B \uparrow \rightarrow W_{dm} = t_{Si} \downarrow \rightarrow L \downarrow$$

- $E_{avg} = \frac{2Q_{dep}+Q_{inv}}{2\varepsilon_{si}}$

$$N_B \uparrow \rightarrow Q_{dep} \uparrow \rightarrow E_{avg} \uparrow \mu \downarrow$$

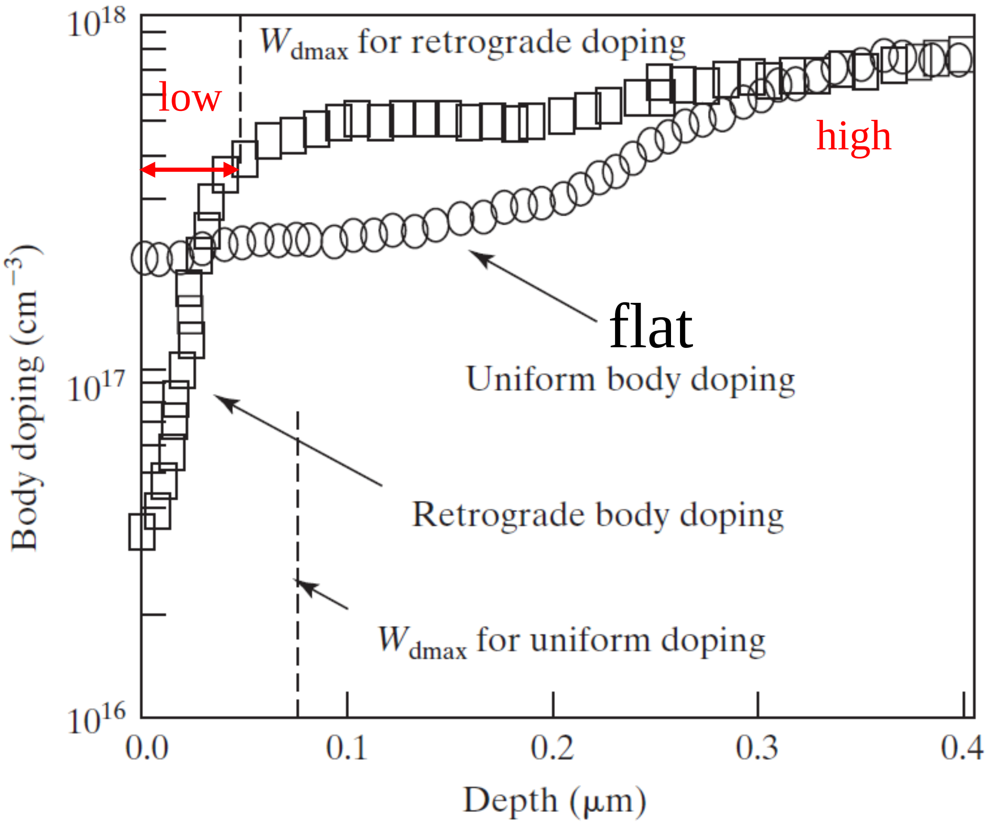
→ *Using strained Si to enhance* μ

Surface Roughness Scattering
(wave function at interface)

- At high electric field, the mobility merge together

→ Universal mobility

Retrograde Body Doping Profiles

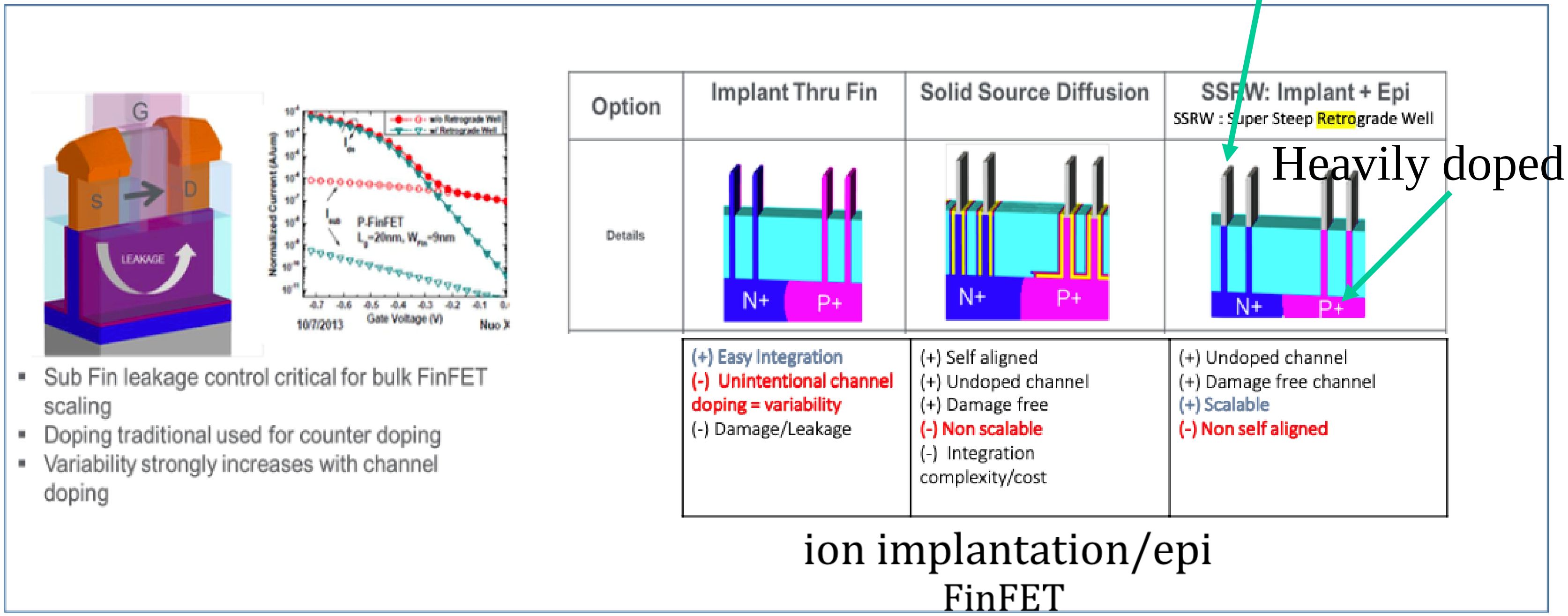


- W_{dep} does not vary with V_{sb} .
- Retrograde doping is popular because it reduces off-state leakage and allows higher surface mobility.

Sub-Fin Isolation Pathways

Towards un-doped channels

Fin undoped by Epi growth



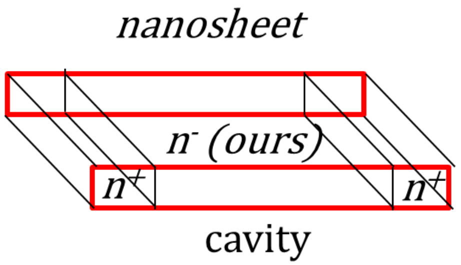
- Sub Fin leakage control critical for bulk FinFET scaling
- Doping traditional used for counter doping
- Variability strongly increases with channel doping

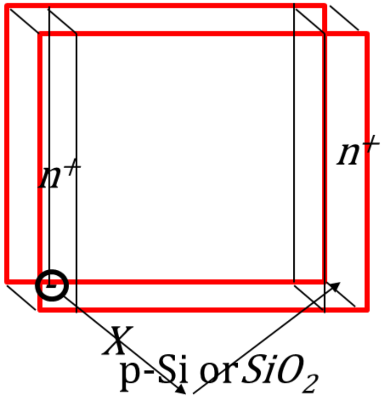
Option	Implant Thru Fin	Solid Source Diffusion	SSRW: Implant + Epi SSRW : Super Steep Retrograde Well
Details	N+ P+	N+ P+	N+ P+
	(+) Easy Integration (-) Unintentional channel doping = variability (-) Damage/Leakage	(+) Self aligned (+) Undoped channel (+) Damage free (-) Non scalable (-) Integration complexity/cost	(+) Undoped channel (+) Damage free channel (+) Scalable (-) Non self aligned

Heavily doped

ion implantation/epi

FinFET





CMOS vs CFET

- CMOS: nFET + pFET
  - Planar (≥N20): area pFET > nFET

    → Because hole mobility < electron mobility
  - FinFET: area pFET = nFET

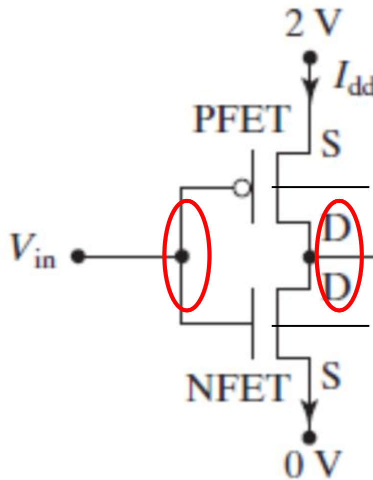
    → Because of undoped channel

- CFET: pFET on nFET or nFET on pFET→ 2X transistor density
- CFET are expected to use nanosheet

6.7.1 CMOS Inverter--voltage transfer curve

For 5nm FinFET, Wn~Wp

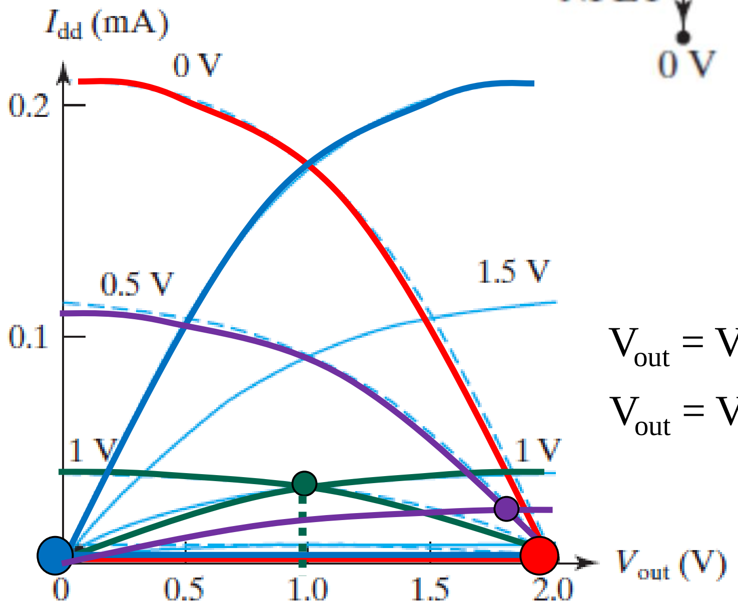
$W_p = 2W_{\imath}$



$V_{gs} = V_g - V_s$

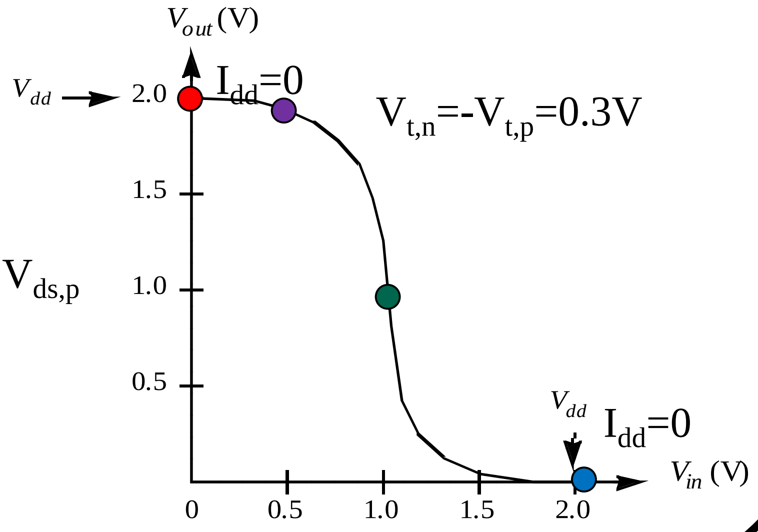
P-Body V_{DD}

N-Body ground= 0



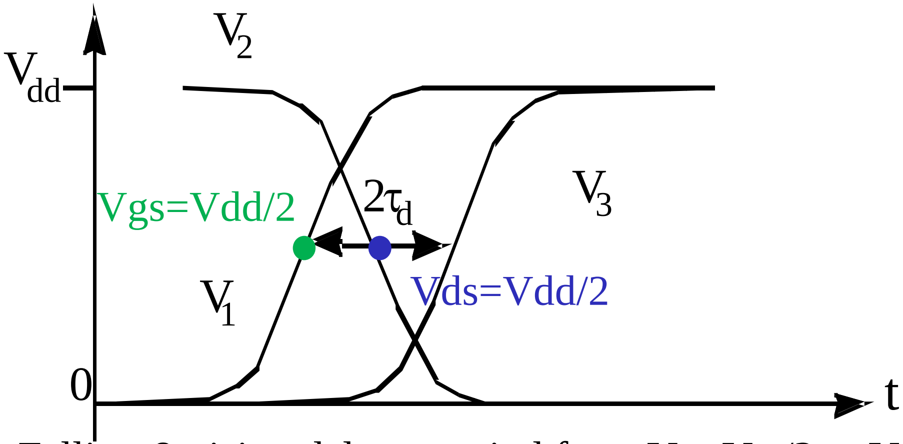
$V_{out} = V_{dd} + V_{ds,p}$

$V_{out} = V_{ds,n}$

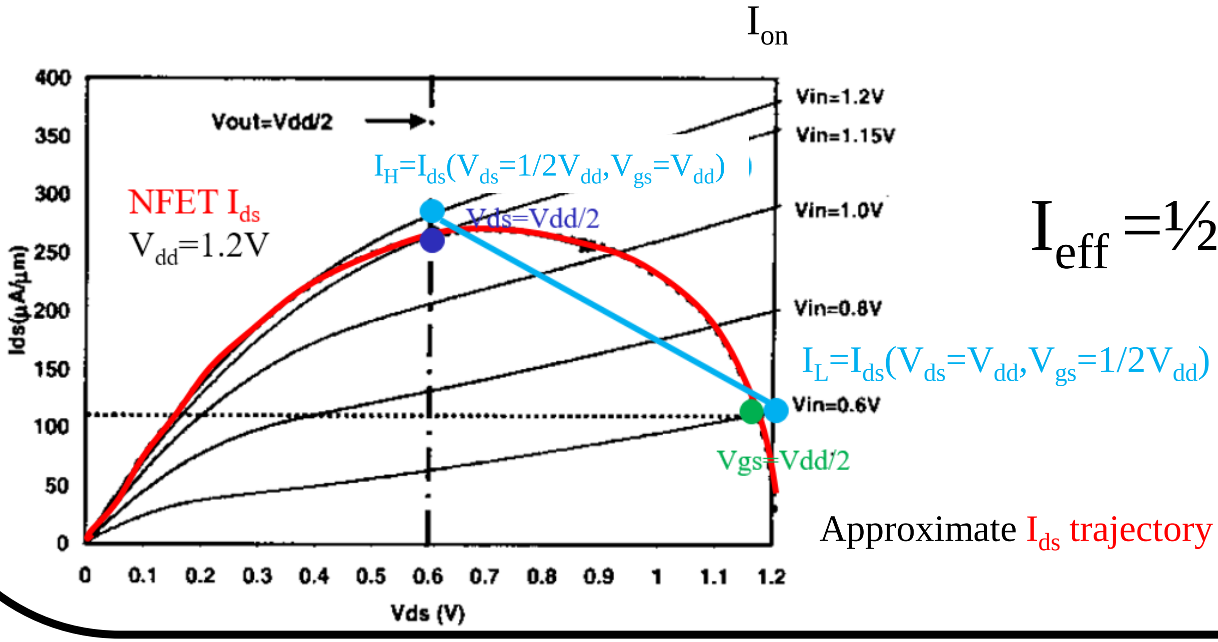


$V_{t,n} = -V_{t,p} = 0.3V$

The Effective Drive Current in CMOS Inverters



- Falling & rising delays: period from $V_{in}=V_{dd}/2$ to $V_{out}=V_{dd}/2$



$$I_{eff}=½(I_L+I_H)$$

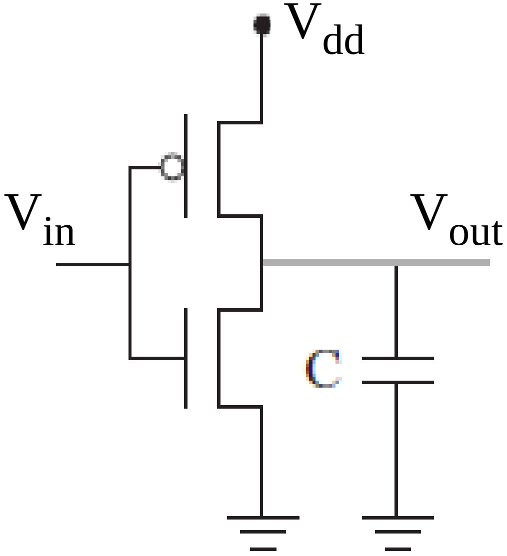
Approximate I_{ds} trajectory with linearize I_{ds}(Vgs,Vds)

6.7.2 Inverter Speed - Impact of I_{on}

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

$$pull-up\,delay \approx \frac{CV_{dd}}{2\ Ip_{eff}}$$

$$pull-down\,delay \approx \frac{CV_{dd}}{2.\ In_{eff}}$$



$I_{eff} = ½(I_L+I_H)$

$I_L=I_{ds}(V_{ds}=V_{dd},V_{gs}=1/2V_{dd})$

$I_H=I_{ds}(V_{ds}=1/2V_{dd},V_{gs}=V_{dd})$

τ_d = Max(pull-up delay, pull down delay)

How can the speed of an inverter circuit be improved?

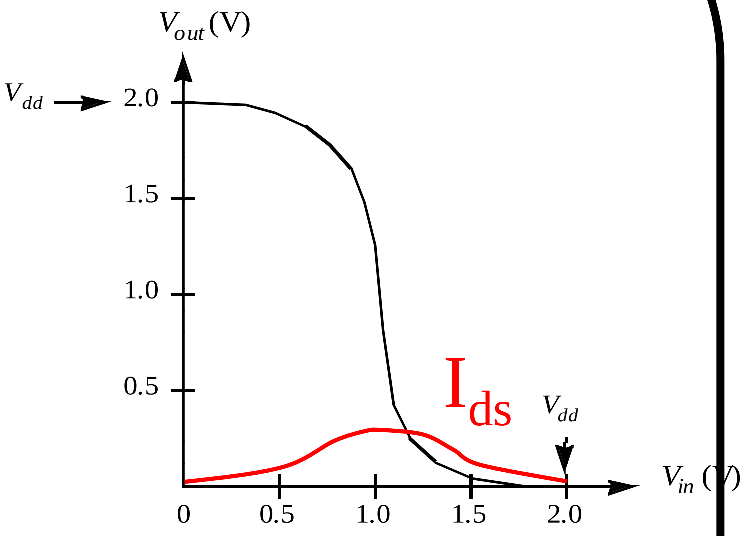
6.7.3 Power Consumption

$$P_{dynamic} = V_{DD} \times I_{avg} + kCV_{DD}^2 f$$

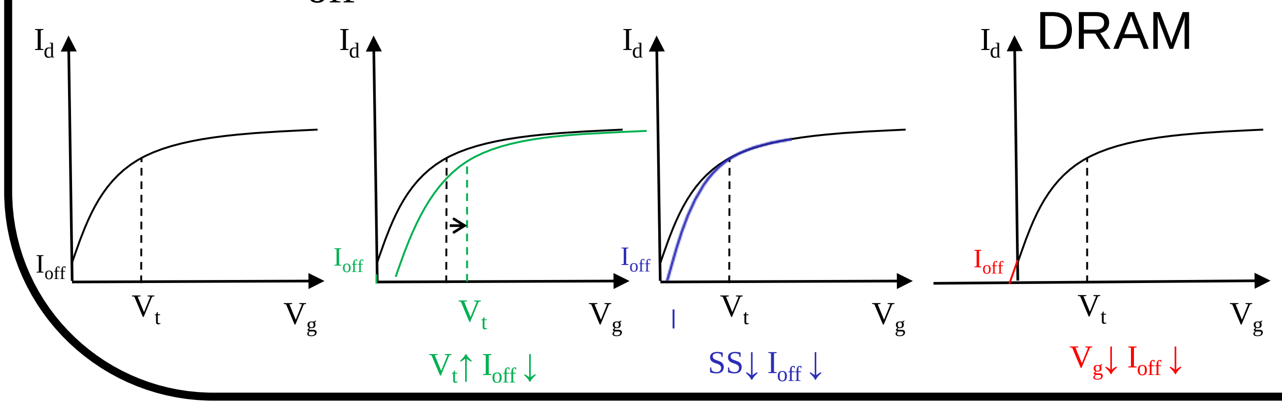
$$P_{static} = V_{DD} I_{OFF}$$

Total power consumption

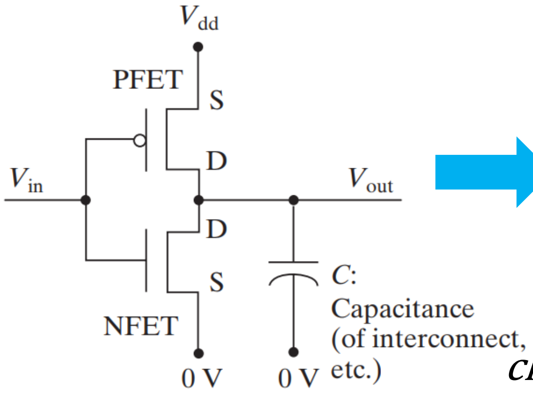
$$P_{total} = P_{dynamic} + P_{static}$$

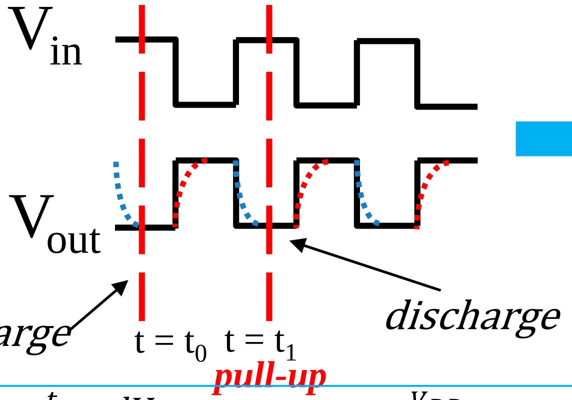


To reduce I_{off}



RC Circuit Energy Consumption





$$\begin{cases} t = t_0 \rightarrow V_{out} = 0 \\ t = t_1 \rightarrow V_{out} = V_{DD} \end{cases}$$

$$\boxed{E_{total} = \int_{t_0}^{t_1} I(t) \times V_{DD} \times dt = V_{DD} \int_{t_0}^{t_1} C\frac{dV_{out}}{dt} dt = CV_{DD} \int_{0}^{V_{DD}} dV_{out} = CV_{DD}^2}$$

$$\boxed{E_C = \int_{t_0}^{t_1} I(t) \times V_{out} \times dt = \int_{t_0}^{t_1} C\frac{dV_{out}}{dt} V_{out} dt = C\int_{0}^{V_{DD}} V_{out} dV_{out} = \frac{1}{2}CV_{DD}^2}$$

$\text{Loss} = \frac{1}{2}CV_{DD}^2 \ (pFET)$

$$\because Kirchhoff's\ law: V = Q/C = IR \rightarrow I = \frac{dQ}{dt} = \frac{Q}{RC}$$

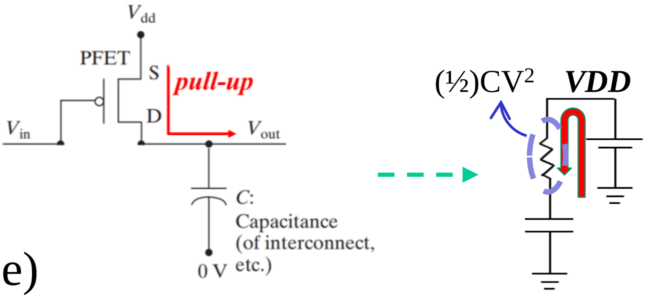
$$\rightarrow \int \frac{dQ}{Q} = \int \frac{dt}{RC} \rightarrow Q = qe^{t/RC} \rightarrow I = \frac{q}{RC}e^{t/RC}$$

$$+ \begin{cases} t = t_0 \rightarrow Q_C = 0 \\ t = t_1 \rightarrow Q_C = Q \end{cases}$$

$$\boxed{E_R = \int_{t_0}^{t_1} I(t)^2 \times R \times dt = \int_{t_0}^{t_1} \frac{q^2}{R^2C^2} e^{2t/RC} R dt = \frac{q^2}{RC^2}[\frac{RC}{2}(e^{2t_1/RC} - e^{2t_0/RC})]}$$

$$\boxed{= \frac{1}{2C}(Q_{t_0}^2 - Q_{t_1}^2) = \frac{1}{2C}Q^2 = \frac{1}{2C}C^2V_{DD}^2 = \frac{1}{2}CV_{DD}^2 = E_{total} - E_C}$$

pull-up (V_{out} from 0 → V_{DD})



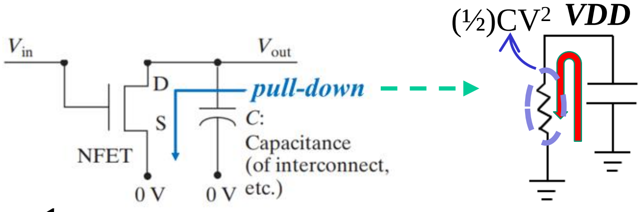
$$E_{total} = CV_{DD}^2$$ (total energy from source)

$$E_C = \frac{1}{2}CV_{DD}^2$$ (energy stored in capacitance)

$$E_R = E_{total} - E_C = \frac{1}{2}CV_{DD}^2$$ (energy dissipation in the pFET)

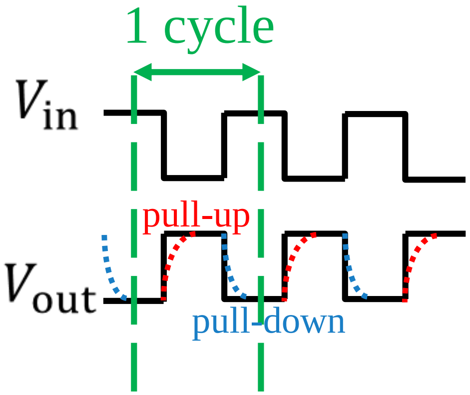
pull-down (V_{out} from V_{DD} → 0)

$$E_{total} = CV_{DD}^2$$



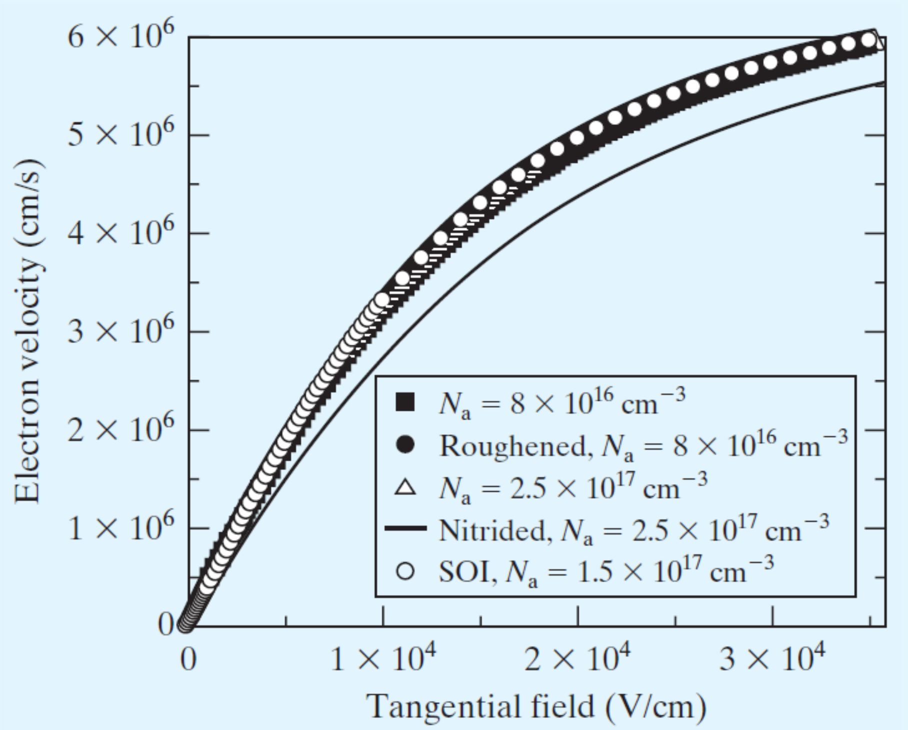
$$E_C = \frac{1}{2}CV_{DD}^2$$

$$E_R = E_{total} - E_C = \frac{1}{2}CV_{DD}^2$$ (energy dissipation in the nFET)



Energy consumption of one cycle = $E_{R,up} + E_{R,down} = CV_{DD}^2$

6.8 Velocity Saturation



Along current direction

$$v = \frac{\mu_n E}{1 + E/E_{sat}}$$

$$E \ll E_{sat}: v = \mu_n E$$

$$E = E_{sat}: v_{sat} = \frac{\mu_n E_{sat}}{2}$$

$$E \gg E_{sat}: v_{sat} = \mu_n E_{sat}$$

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

6.9 MOSFET IV Model with Velocity Saturation

$$I_{ds} = WQ_{inv}v \qquad \boxed{v = \frac{\mu_n E}{1 + E/E_{sat}}} \qquad \boxed{E = dV_{cs}/dx}$$

$$\rightarrow I_{ds} = WC_{ox}(V_{gs} - mV_{cs} - V_t)\frac{\mu_n\, dV_{cs}/dx}{1 + \frac{dV_{cs}}{dx}/E_{sat}}$$

$$\rightarrow \int_0^L I_{ds}dx = \int_0^{V_{ds}} [WC_{ox}\mu_n(V_{gs} - mV_{cs} - V_t) - I_{ds}/E_{sat}]dV_{cs}$$

$$\rightarrow I_{ds}L = WC_{ox}\mu_n\left(V_{gs} - V_t - \frac{m}{2}V_{ds}\right)V_{ds} - I_{ds}V_{ds}/E_{sat}$$

6.9 MOSFET IV Model with Velocity Saturation

$$\text{solving } \frac{dI_{ds}}{dV_{ds}} = 0,$$

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

A simpler and more accurate V_{dsat} is:

$$\frac{1}{V_{dsat}} = \frac{m}{V_{gs} - V_t} + \frac{1}{E_{sat}L}$$

$$E_{sat} \equiv 2v_{sat}/\mu_n$$

- No velocity saturation → $V_{dsat} = \left(V_{gs} - V_t\right)/m$

I_{dsat} with Velocity Saturation

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

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

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

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

$$I_{dsat} = W v_{sat} C_{oxe} (V_{gs} - V_t)$$ Ballistic transport → $I = Q_{source} V_{inj}$

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

Saturation Current and Transconductance

- linear region, saturation region

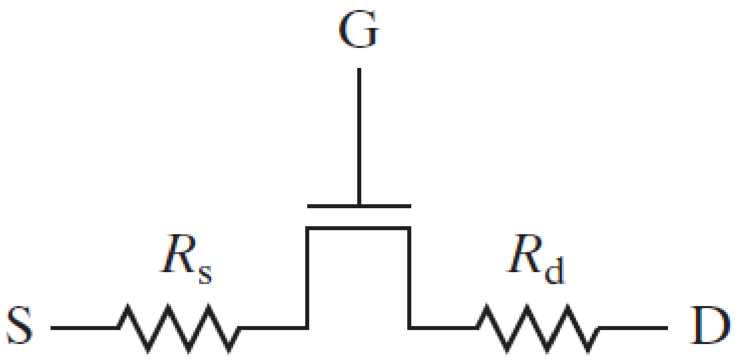
$$I_{dsat} = \frac{1}{2m}\mu_n C_{ox}\frac{W}{L}(V_{gs} - V_t)^{\alpha}$$

(short channel) 1 < a < 2 (long channel)

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

$$\rightarrow g_{msat} = \frac{\alpha}{2m}\mu_n C_{ox}\frac{W}{L}(V_{gs} - V_t)^{\alpha-1}$$

6.10 Parasitic Source-Drain Resistance



(if R_s large)

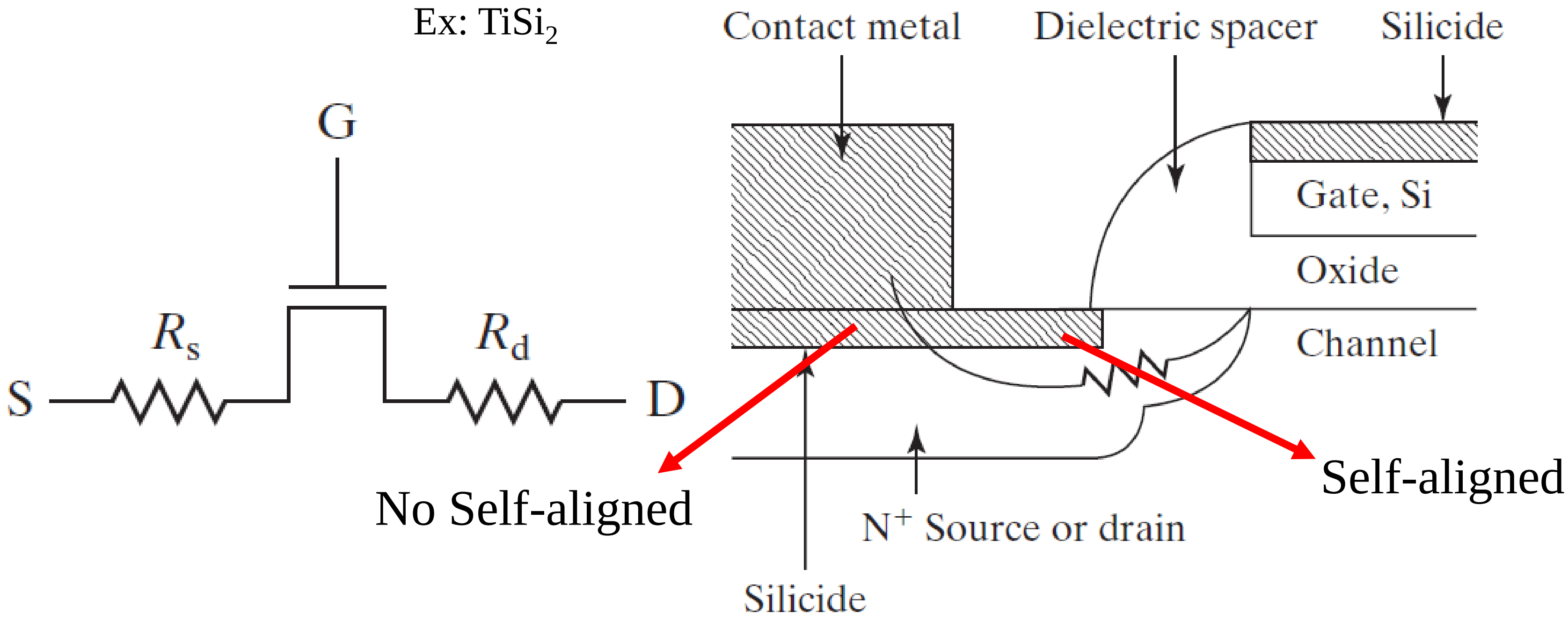
- If $I_{dsat0} \propto V_{gs} - V_t$, $I_{dsat} = \dfrac{I_{dsat0}}{1 + \dfrac{I_{dsat0}\boxed{R_s}}{(V_{gs} - V_t)}} = \dfrac{V_{gs} - V_t}{R_s}$

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

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

SALICIDE (Self-Aligned Silicide) Source/Drain

Ex: $TiSi_2$



No Self-aligned

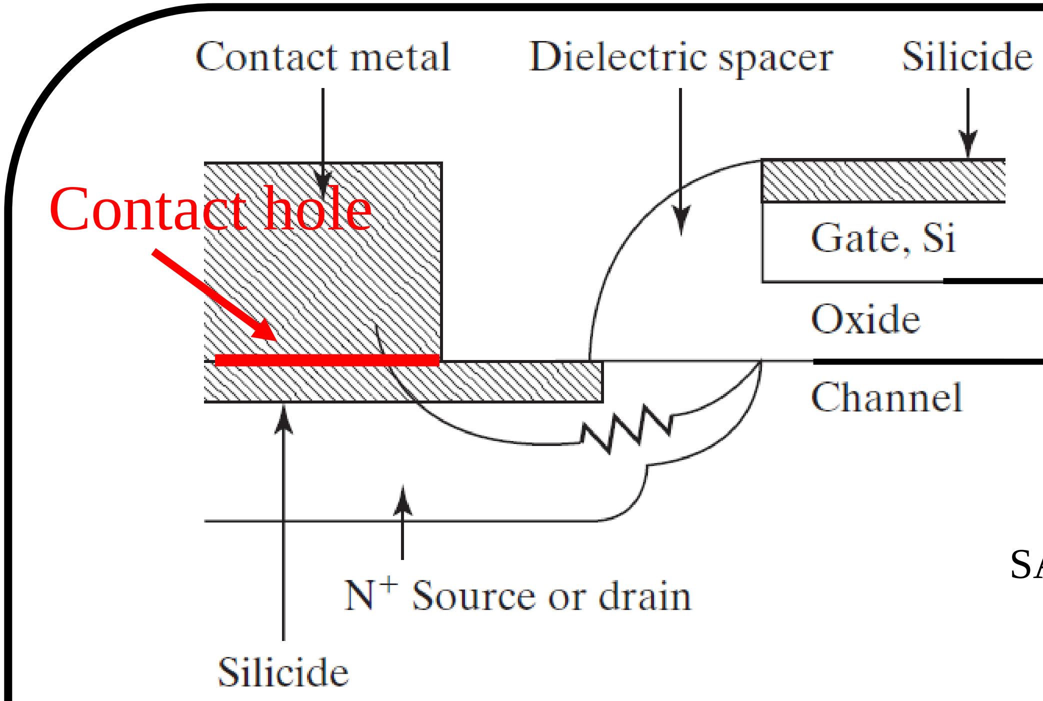
Self-aligned

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

Question:

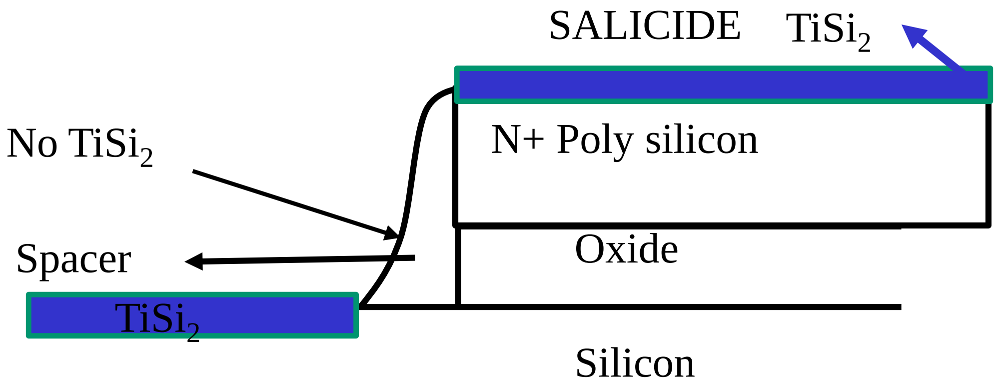
- What is the purpose of siliciding the source/drain/gate?
- What is self-aligned to what?

N+ poly silicon gate For nFET



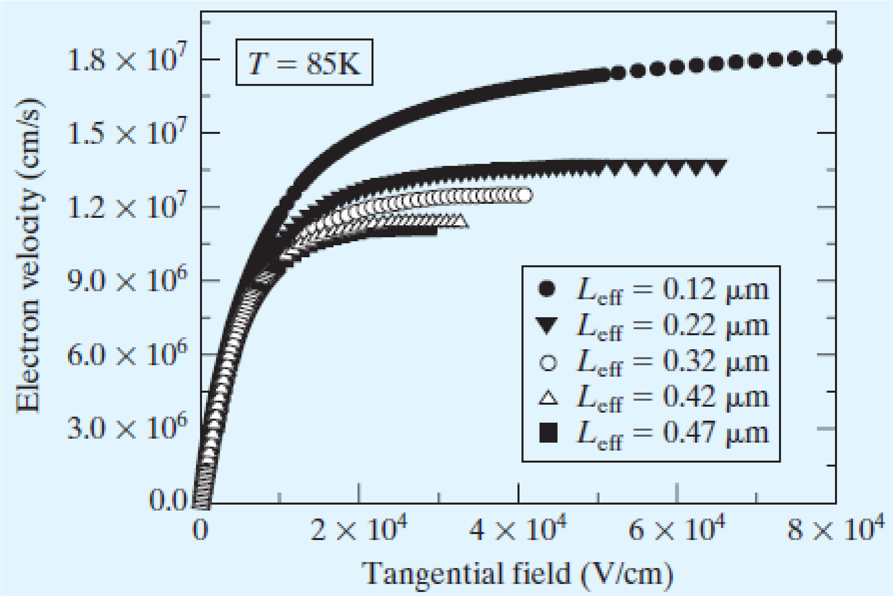
SALICIDE process to reduce R_s/R_d

1. Define Gate length
2. Deposit SiO_2
3. RIE SiO_2 (Spacer formation)
4. Deposit Ti
5. RTA to form $TiSi_2$
6. Etching Ti



6.12 Velocity Overshoot

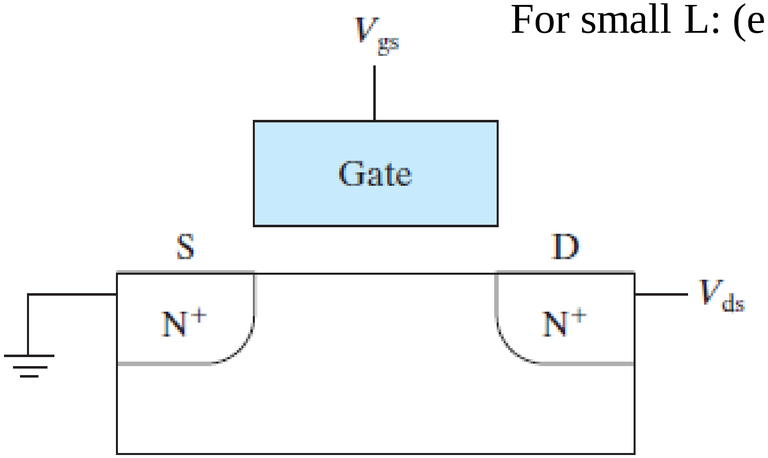
Has been replaced by injection velocity at source



- Velocity saturation should not occur in very short MOSFETs.
- This velocity overshoot could lift the limit on I_{ds}. But…

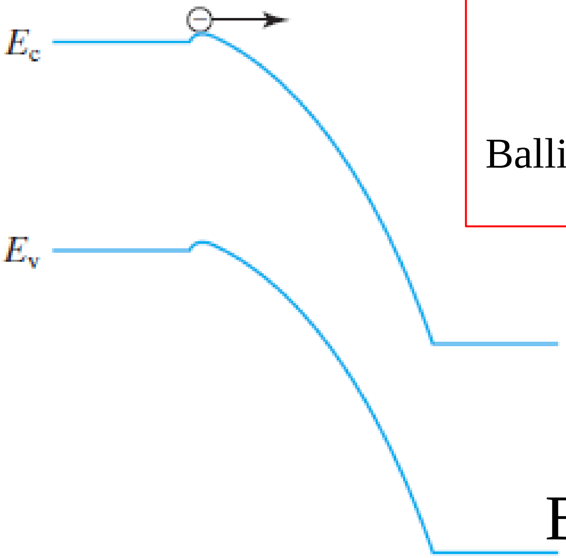
6.12 Source Velocity Limit

For small L: (e.g.: 10~20nm)



$I_d = W v_{inj} C_{ox} (V_{gs} - V_t)$ (injection velocity)

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



cm cm/s C/cm²

$$I_{dsat} = WBv_{thx}Q_{inv}$$

Ballistic factor

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

•Similar to

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

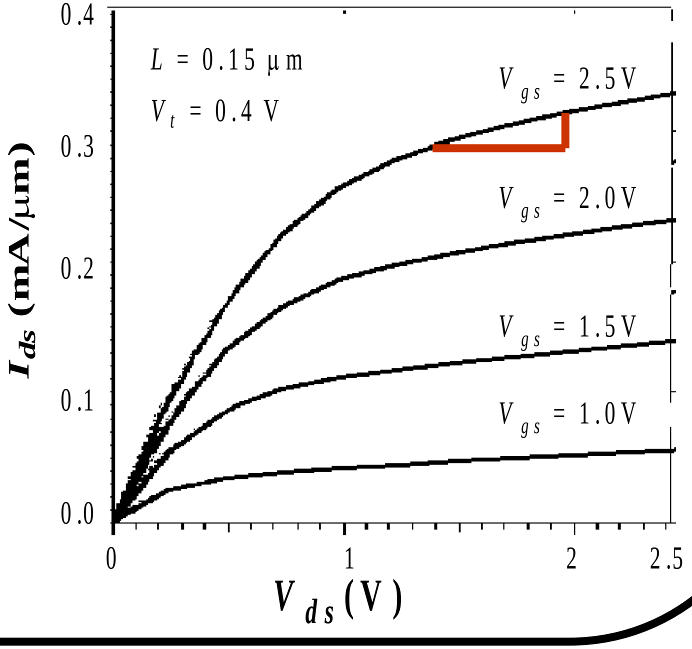
B^*v_{thx} = injection velocity at Source

6.13 Output Conductance

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

$$\boxed{g_{ds} \equiv \frac{dI_{dsat}}{dV_{ds}}} \quad =1/r_0$$

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



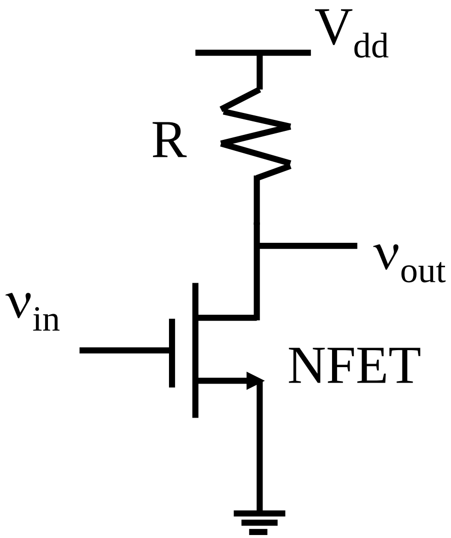
Example of an Amplifier

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

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

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

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



➡

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

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

$g_m r_0$

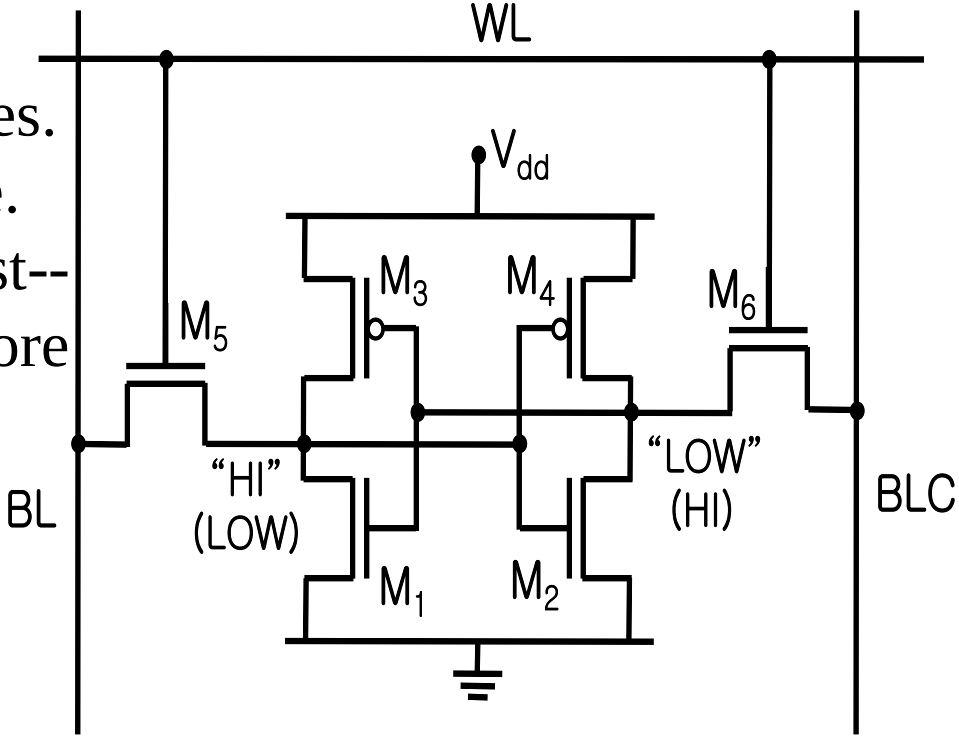
r_0 depends on channel length modulation

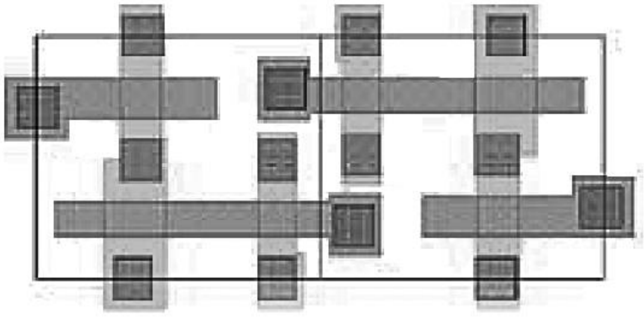
6.16 Memory Devices

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

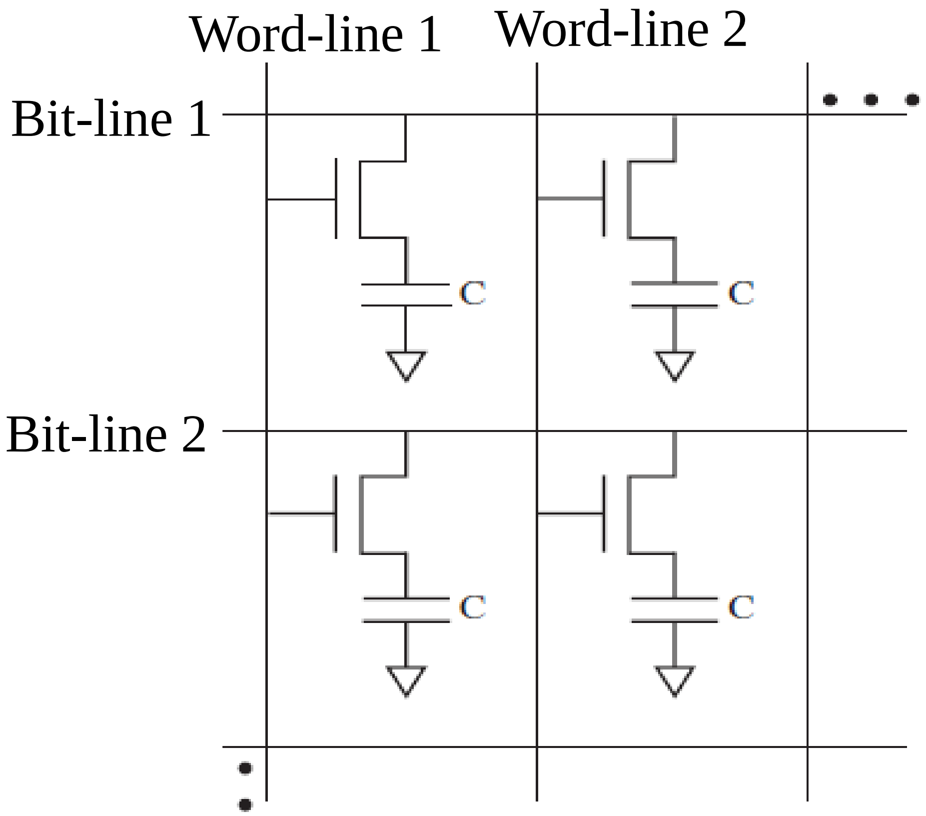
6.16.1 SRAM

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



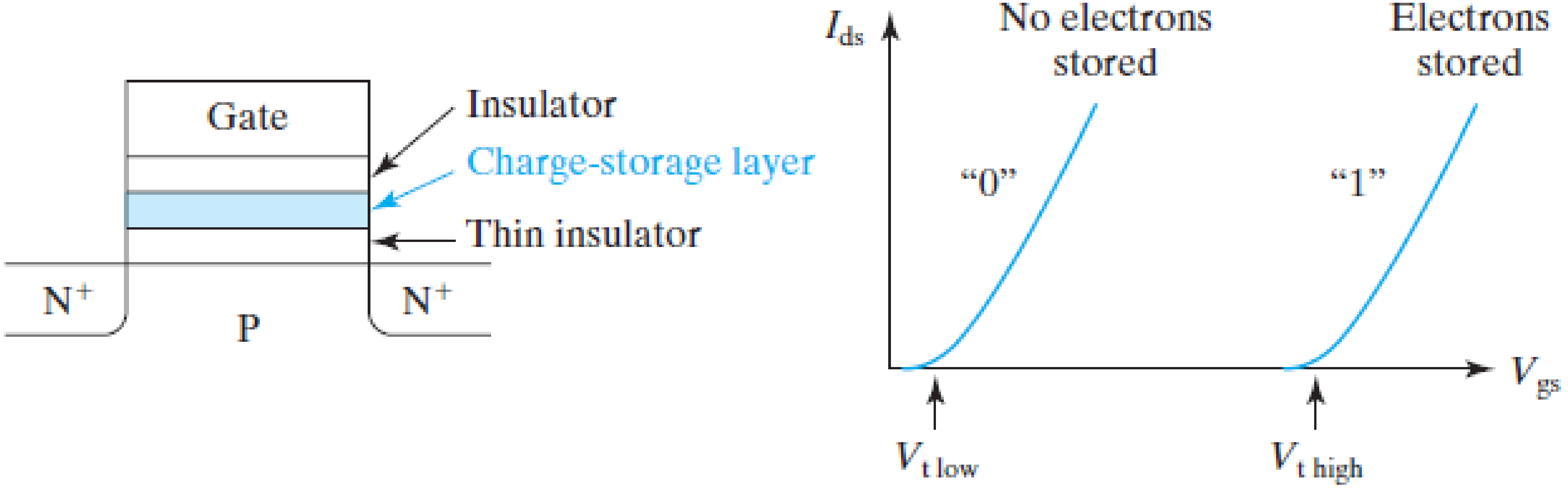


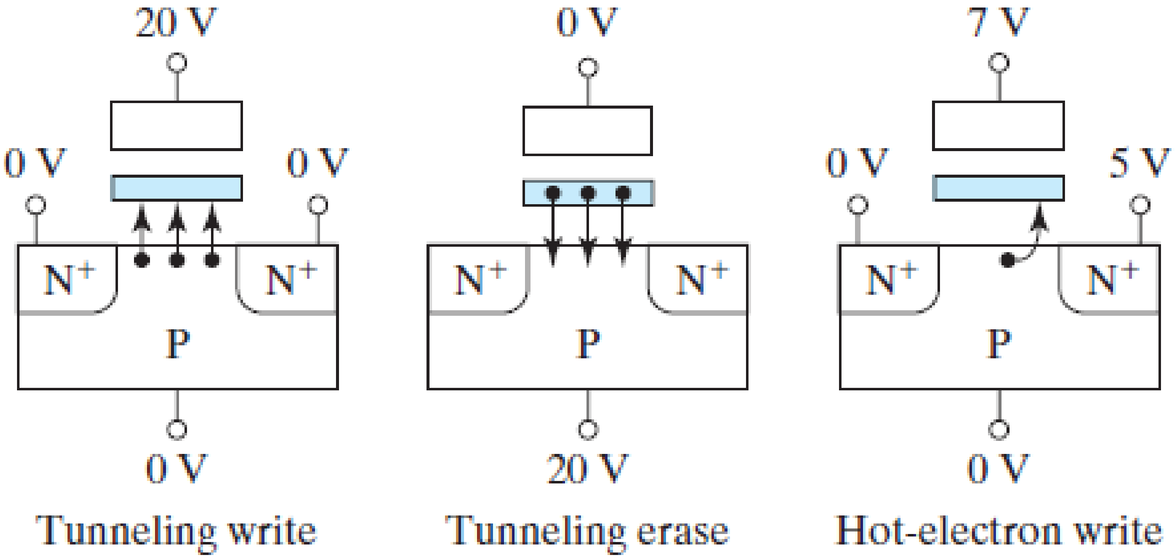
6.16.2 DRAM



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

6.16.3 Nonvolatile (Flash) Memory

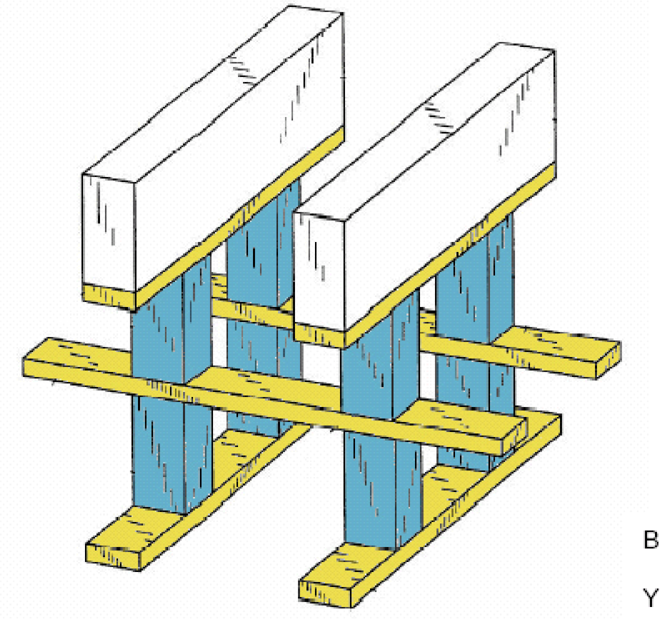


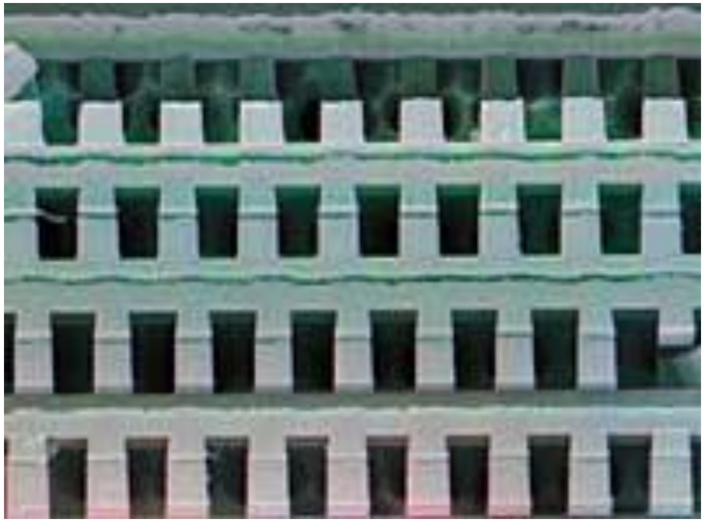


- Floating gate (poly-Si)
- Charge trap (SONOS)
- Nanocrystal

3D (Multi-layer) Memory

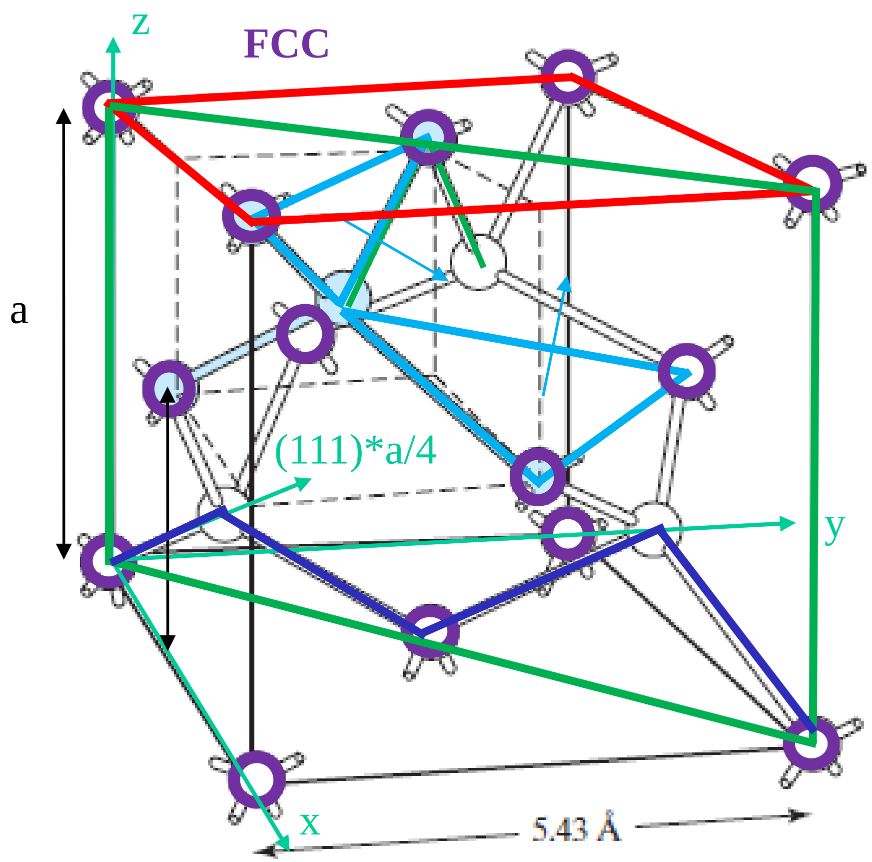
- Epitaxy from seed windows can produce Si layers.
- Ideally memory element is simple and does not need single-crystalline material.





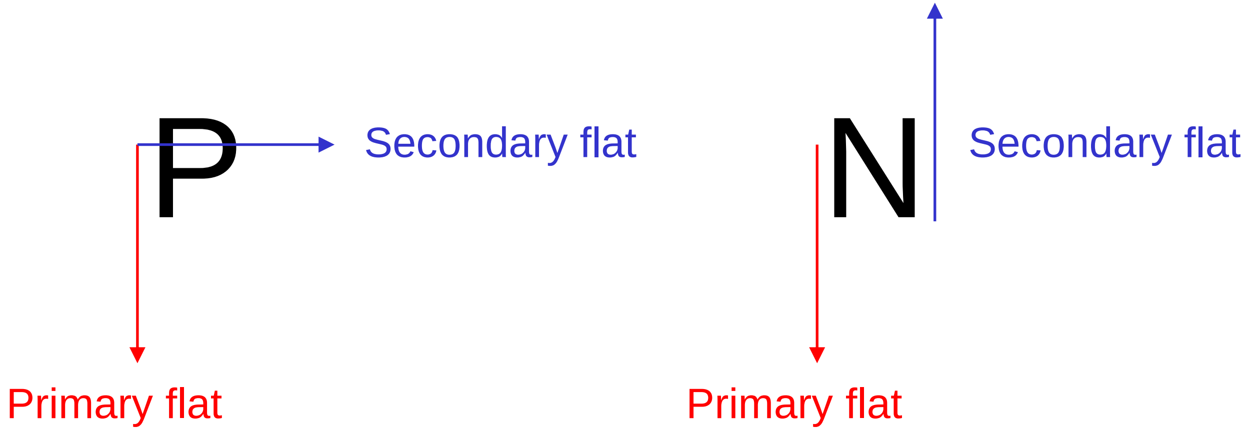
1.1 Silicon Crystal Structure

- **Q1: How many atoms in FCC?**
- **A1: $8 * \frac{1}{8} + 6 * \frac{1}{2} = 4$**
- **Q2: How many atoms in diamond structure?**
- **A2: 2*FCC=8**
- **Q3:Find out Zigzag**
- **A3: on the {110}**
- **Q4: What are the sidewall of nanosheet?**
- **A4: {110}**
- **Q5 What's the thickness of monolayer**



1.1 Silicon Crystal Structure

Q: How to memorize the type of Si (100) wafer by flat?



P

Secondary flat

Primary flat

N

Secondary flat

Primary flat

Q: How about 12 inch wafer (no flat to save area)?
How to distinguish between n and p wafer?
A: Using thermal electric experiment.

EXAMPLE: Temperature Dependence of Resistance

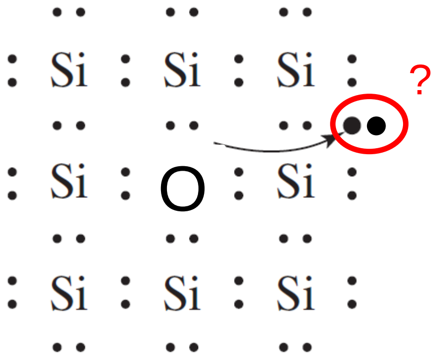
By what factor will R increase or decrease from T=300 K to T=400 K?

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

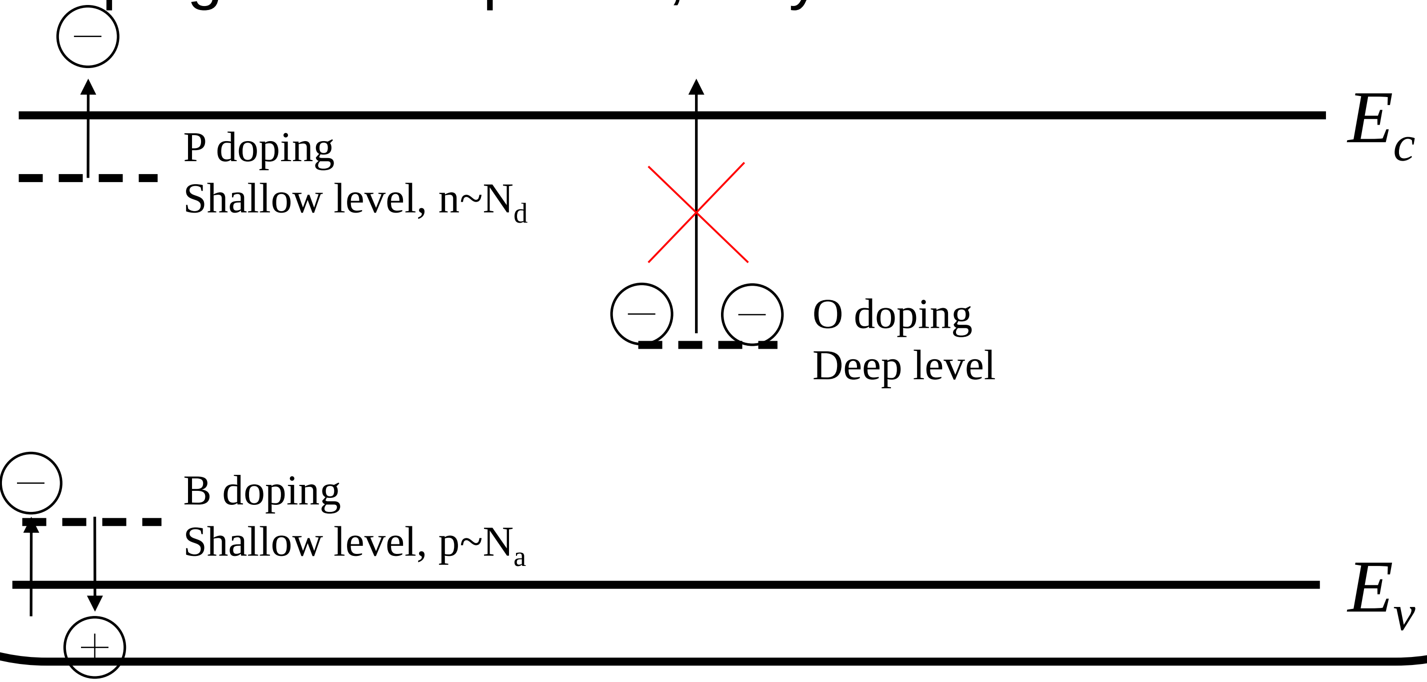
$$\frac{770}{400} = 1.93$$

1.3 Energy Band Model

Q: Why not use O as donor?
There can be 2 free electrons.



A: O doping is at deep level, only few free electrons at 300K



P doping
Shallow level, $n \sim N_d$

O doping
Deep level

B doping
Shallow level, $p \sim N_a$

E_c

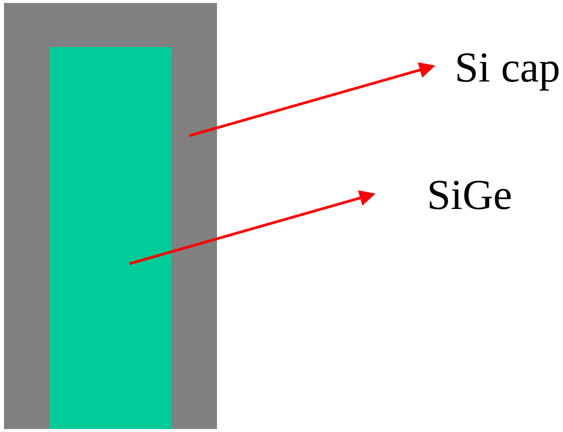
E_v

1.5 Electrons and Holes

Why not Ge? Ge has low effective mass for both electron and hole

GeO_2 is not good enough

SiO_2 is the best

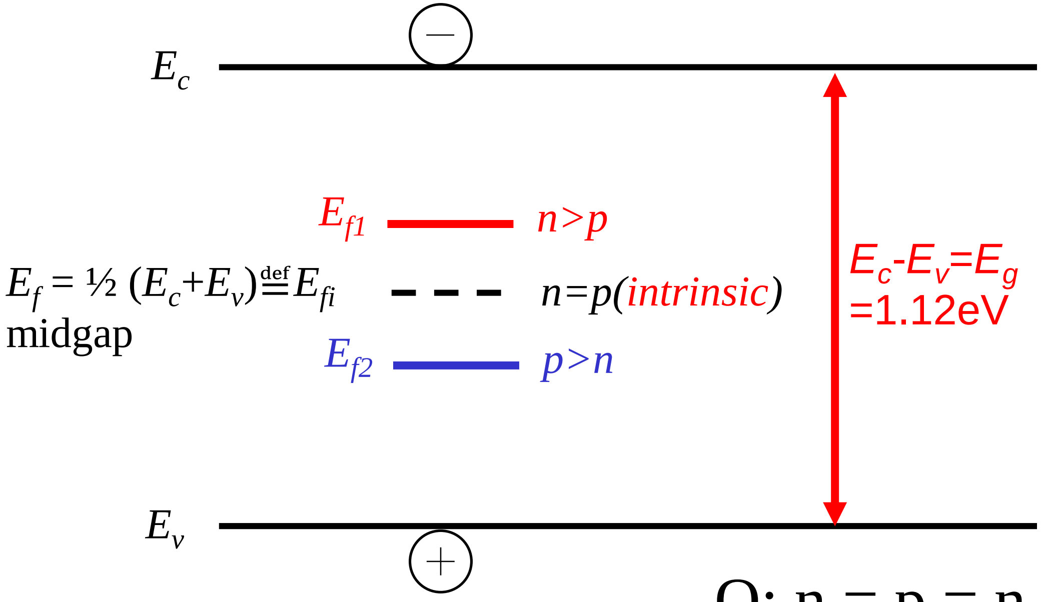


Solution: Si cap on Ge → Good compatibility with SiO_2

SiGe is in production for pFET in 5nm node, but the insulator is still SiO_2

1.8 The Fermi Level and Carrier Concentrations

$$n = N_c e^{-\boxed{E_c - E_f}/kT} \qquad p = N_v e^{-\boxed{E_f - E_v}/kT}$$



- At E_i, $n = p = n_i = 10^{10}$ cm^{-3}

→ $n = n_i e^{(Ef-Ei)/kT}$

- At T=300K, $e^{60/26}$~10

→ 60 meV/decade

$N_c>N_v$: E_f will below E_i
E_f near E_v p↑
E_f near E_c n↑

Q: $n = p = n_i$
Assume $N_c = N_v$, E_f = ?

A: $n = p = n_i$

Q: If N_c not equal N_v, where is E_f ?
(relative to E_i)

1.8 The Fermi Level and Carrier Concentrations

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

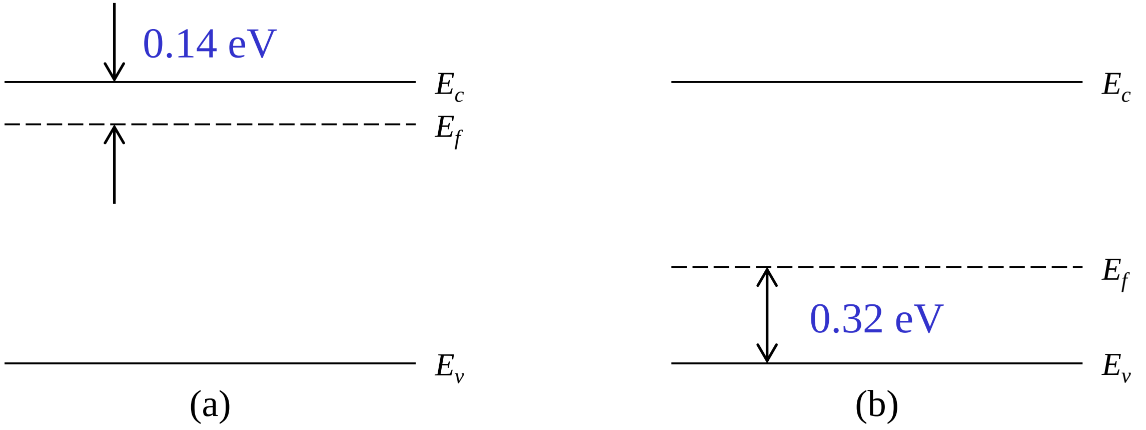
Hint :1decade = 60meV

Solution: (a) $n/n_i = 10^7 \rightarrow$ 7 decade $\rightarrow \Delta E = 7 \cdot 60$ meV

$$E_f - E_i = 0.42\text{eV} \rightarrow E_c - E_f = 0.56\text{-}0.42 = 0.14\text{eV}$$

(b) $p/n_i = 10^4 \rightarrow$ 4 decade $\rightarrow \Delta E = 4 \cdot 60$ meV

$$E_i - E_f = 0.24\text{eV} \rightarrow E_f - E_v = 0.56\text{-}0.24 = 0.32\text{eV}$$



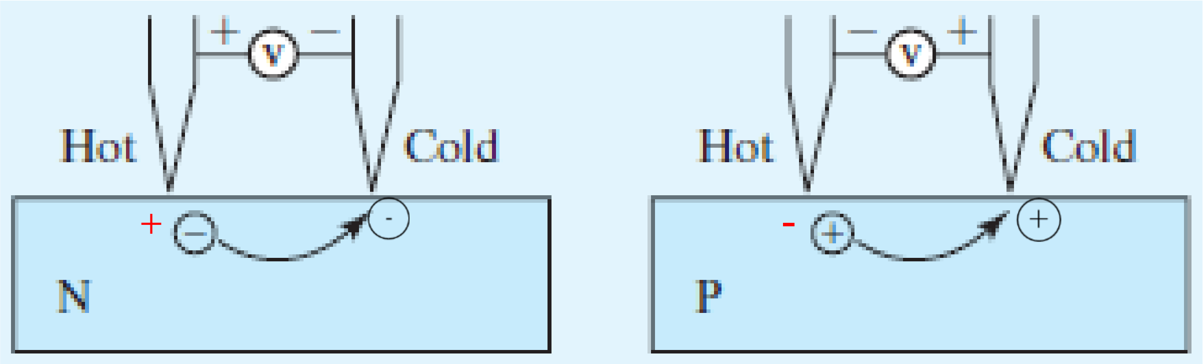
1.8 The Fermi Level and Carrier Concentrations

Q: for Si, n_i = 1.4x10^{10} , for Ge, n_i = ?
Bandgap: Si = 1.12eV, Ge = 0.66eV

$$n_i^2 \propto T^3 e^{-E_g/kT}$$

A: n_i^2(Ge) = 1E28, n_i(Ge) = 1E14

Hot-point Probe can determine sample doing type

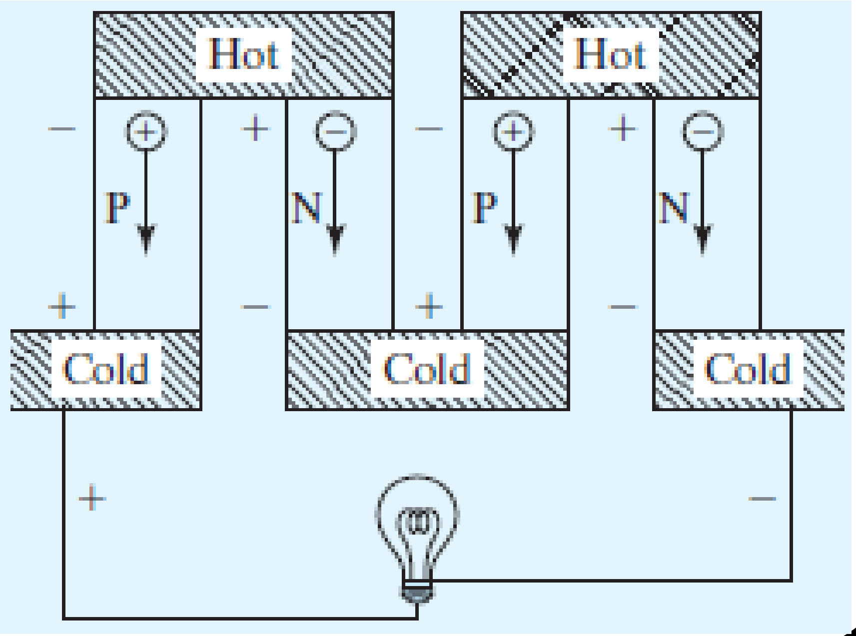


***Hot-point Probe* distinguishes N and P type semiconductors.**

Hole is different from positron

Thermoelectric cooler:

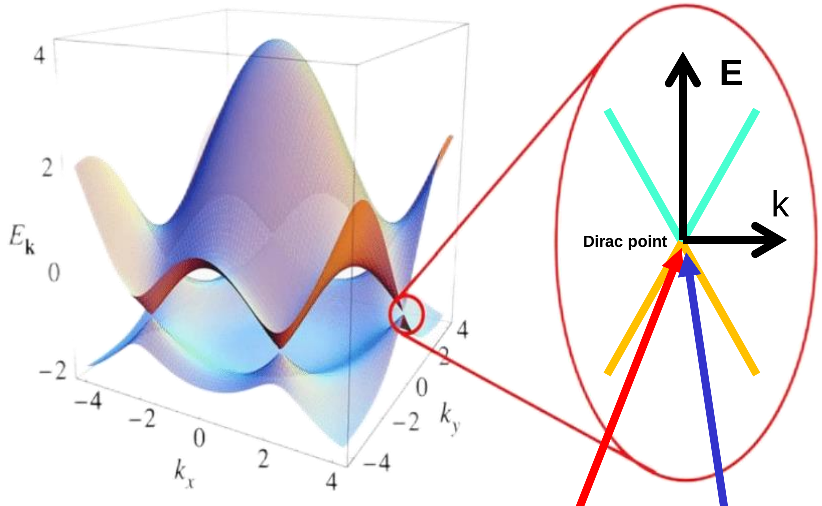
carriers go from A to B with energy



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

Thermoelectric Cooler

The effective mass in graphene



- Slope discontinuity at dirac point

→ Can not apply Taylor expansion

- $E_g = 0$ → Large I_{off}
- E_g can not be widen by quantum confinement because of 2D property

$$p = \hbar k = m^* v_g \qquad \textbf{(1)}$$

$$v_g = \frac{1}{\hbar}\frac{dE(k)}{dk} \qquad \textbf{(2)}$$

Substituting (1) into (2) gives the definition (3)

$$m^* = \hbar^2 k \left(\frac{dE(k)}{dk}\right)^{-1} \qquad \textbf{(3)}$$

We write the linear dispersion (like a photon) for graphene as $E(k) = \hbar c_g k$, where c_g is the speed of the electrons in graphene.

$$E(k) = \hbar c_g k \qquad \textbf{(4)}$$

Substituting (4) into (3) gives the definition (5)

$$\boxed{m^* = \hbar \frac{1}{c_g} k} \qquad \textbf{(5)}$$

For k = 0, m*=0

2D, Graphene: p = mv

- Cannot use E's second differential to calculate m^*

Dirac point:

- Linear relation between E and k → $m^* = 0$
- GeSn has Dirac point if [Sn] is large enough

2.2 Drift velocity

Drift Velocity, Mean Free Time, Mean Free Path

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

Solution: $v = \mu_p E$ = 470 cm^2/V·s × 10^3 V/cm = 4.7× 10^5 cm/s

$\tau_{mp} = \mu_p m_p / e$ =470 cm^2/V ·s × 0.39 × 9.1×10^{-31} kg/1.6×10^{-19} C

= 0.047 m^2/V ·s × 2.2×10^{-12} kg/C = 1×10^{-13}s = 0.1 ps

mean free path = $\tau_{mh} v_{th}$ ~ 1× 10^{-13} s × 2.2×10^7 cm/s

= 2.2×10^{-6} cm = 220 Å = 22 nm

If $v_{drift} = v_{th}$, the mean free path is equal

mean free path = $\tau_{mh} v_d$ ~ 1× 10^{-13} s × 4.7×10^5 cm/s

= 4.7×10^{-8} cm = 4.7 Å = 0.47 nm

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

2.2 Drift velocity

EXAMPLE: Temperature Dependence of Resistance

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

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

Solution:

(a) Using the N-type curve in the previous figure, we find that $\rho = 0.084\ \Omega$-cm.

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

$= 0.084\ \Omega\text{-cm} \times 10^{-4}\ cm / 10^{-9}\ cm^2$

$= 8.4 \times 10^3\ \Omega$

2.3 Diffusion Current

EXAMPLE: Diffusion Constant

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

Solution:

$$D_p = \frac{kT}{e}\mu_p = 26mV \cdot 410cm^2/V \cdot s = 11cm^2/s$$

Remember: kT/e = 26 mV at room temperature (300K).

2.6 Electron-Hole Recombination

EXAMPLE: Photoconductors

A bar of Si is doped with boron at 10^{15}*cm*$^{-3}$*. It is exposed to light such that electron-hole pairs are generated throughout the volume of the bar at the rate of* 10^{20}*/s·cm*^{3}*. The recombination lifetime is 10μs.*

What are (a) p_0 *, (b)* n_0 *, (c) p', (d) n', (e) p , (f) n, and (g) the np product?*

2.6 Electron-Hole Recombination
EXAMPLE: Photoconductors

Solution:

(a) What is p_0*?*

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

(b) What is n_0 *?*

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

(c) What is p'*?*

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

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

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

2.6 Electron-Hole Recombination
EXAMPLE: Photoconductors

(d) What is n'?

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

(e) What is p?

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

(f) What is n?

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

(g) What is np?

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

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

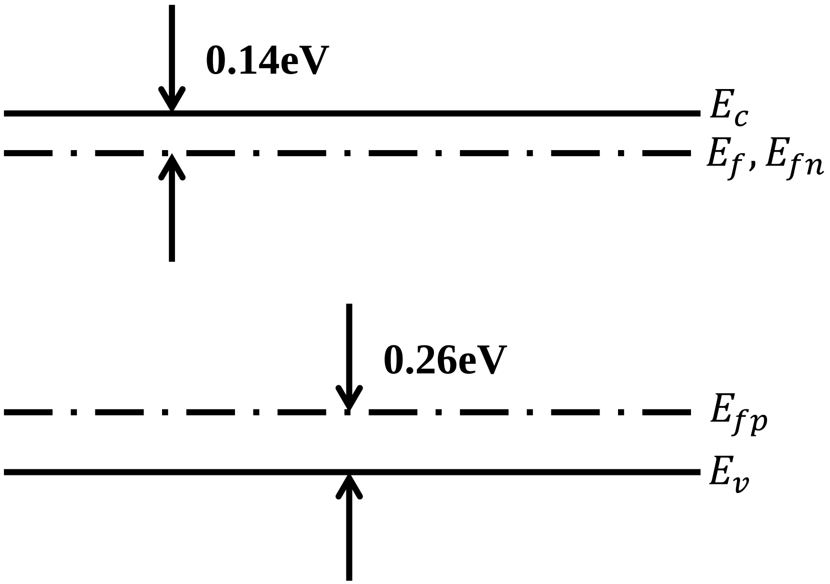
2.8 Quasi-equilibrium and Quasi-Fermi Levels

EXAMPLE: Quasi-Fermi Levels

N_d=$10^{17,}$ n'=p'= 10^{15}

$10^{15}/10^{10}=10^5$ → 5*60 = 300meV

$\therefore E_{fp}$–E_v = 0.56 – 0.3 = 0.26eV



recombination

2.8 Quasi-equilibrium and Quasi-Fermi Levels

EXAMPLE: Quasi-Fermi Levels and Low-Level Injection

Consider a Si sample with $N_d=10^{17}cm^{-3}$.
Find out E_{fn} *and* E_{fp} *if* $n'=p'=10^{18}$ *(high injection)*

(a) Find E_f .

$$10^{17}/10^{10}=10^{7} \rightarrow 7*60 = 420\text{meV}$$
$$\therefore E_c\text{-}E_f = 0.56 - 0.42 = 0.14\text{eV}$$

Note: n' and p' are much less than the majority carrier concentration. This condition is called ***low-level injection. (n' , p' < n_0 , p_0)***

2.8 Quasi-equilibrium and Quasi-Fermi Levels

EXAMPLE: Quasi-Fermi Levels and Low-Level Injection

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

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

$n = 1.01*10^{17}\text{cm}^{-3} = N_c e^{-(E_c - E_{fn})/kT}$

$1.01*10^{17}/10^{10} \approx 10^7 \rightarrow 7*60 = 420\text{meV}$

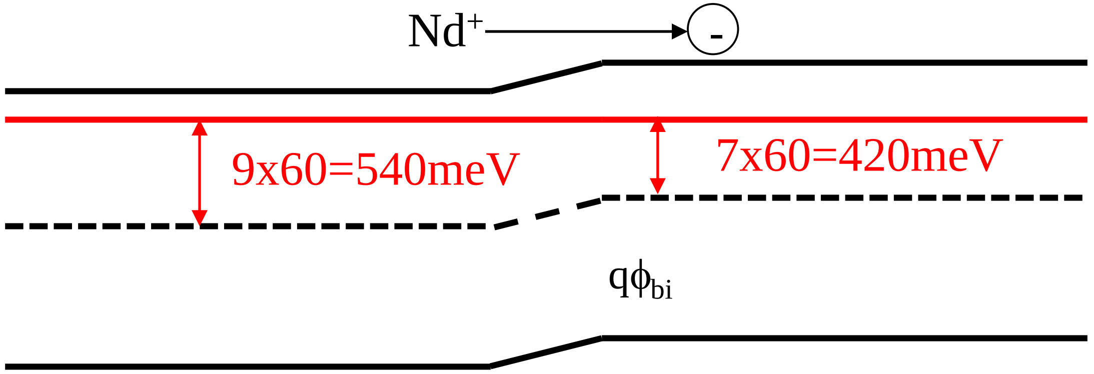
$\therefore E_c – E_{fn} = 0.56 – 0.42 = 0.14\text{eV}$

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

4.1 Building Blocks of the PN Junction Theory

Junctionless transistor

$n^+ = 1E19$, $n = 1E17$, $q\phi_{bi} = ?$ Plot the band diagram



4.2 Depletion-Region Model

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

n=10^{17}, p=10^{20}

$$\left(\boldsymbol{log_{10}\frac{10^{17}}{10^{10}}+log_{10}\frac{10^{20}}{10^{10}}}\right)\cdot\boldsymbol{60mV=1020mV}$$

Solution:

$$(a)\ \phi_{bi}=\frac{kT}{e}ln\frac{N_dN_a}{n_i^2}=0.026V*ln\frac{10^{20}*10^{17}cm^{-6}}{10^{20}cm^{-6}}\approx 1V$$

$$\phi_{bi}=\frac{eN_a^-W_p^2}{2\varepsilon_s}$$

1e-12F/cm

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

$$(c)\ |x_N|\approx W_{dep}=0.12\mu m$$

$$(d)\ |x_P|=|x_N|\frac{N_d}{N_a}=1.2*10^{-4}\mu m\approx 0.1nm$$

$$\boldsymbol{\frac{10^{17}}{10^{20}+10^{17}}}$$

By charge neutrality

1E15: W_{dep}: 1um

1E17: W_{dep}: 0.1um= 100nm

1E19: W_{dep}: 10nm

1E21: W_{dep}: 1nm

4.3 Reverse-Biased PN Junction

EXAMPLE: If the slope of the line in the previous slide is $2\text{x}10^{23}$ F^{-2} V^{-1}, the intercept is 0.84V, and A is 1 μm^2, find the lighter and heavier doping concentrations N_l and N_h .

Solution: $\boldsymbol{slope = \frac{2}{eN_B\varepsilon_s A^2}}$ $\quad \frac{1}{N_B} = \frac{1}{N_a} + \frac{1}{N_d} \approx \frac{1}{lighter\ dopant\ density}$

$$\boldsymbol{N_B \sim N_L = \frac{2}{slope \times e\varepsilon_s A^2}}$$

$$\boldsymbol{= \frac{2}{2 \times 10^{23} \times 1.6 \times 10^{-19} \times 12 \times 8.85 \times 10^{-14} \times 10^{-8}}}$$

$$\boldsymbol{= 6 \times 10^{15} cm^{-3}}$$

60mV error→ N_h 10X error

$$\phi_{bi} = \frac{kT}{e} \ln \frac{N_h N_l}{n_i^2} \rightarrow \boxed{N_h = \frac{n_i^2}{N_l} e^{\frac{e\phi_{bi}}{kT}}} = \frac{10^{20}}{6 \times 10^{15}} e^{\frac{0.84}{0.026}} = 1.8 \times 10^{18} cm^{-3}$$

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

Slope → lightly dope

$\phi_{\text{lightly dope}} + \phi_{\text{heavily dope}}$ → ϕ_{bi}

4.6 Forward Bias – Carrier Injection (dark)

EXAMPLE: Carrier Injection

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

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

Solution:
$$n(x_P) = n_{P0}e^{eV/kT} = \frac{10^{20}}{10^{19}}e^{0.6/0.026} = 10^{11}cm^{-3}$$
$$p(x_N) = p_{N0}e^{eV/kT} = \frac{10^{20}}{10^{16}}e^{0.6/0.026} = 10^{14}cm^{-3}$$

Question: What are the excess minority carrier concentrations?

Solution:
$$n'(x_P) = n(x_P) - n_{P0} = 10^{11} - 10 = 10^{11}cm^{-3}$$
$$p'(x_N) = p(x_N) - p_{N0} = 10^{14} - 10^4 = 10^{14}cm^{-3}$$

4.8 Excess Carrier Distributions

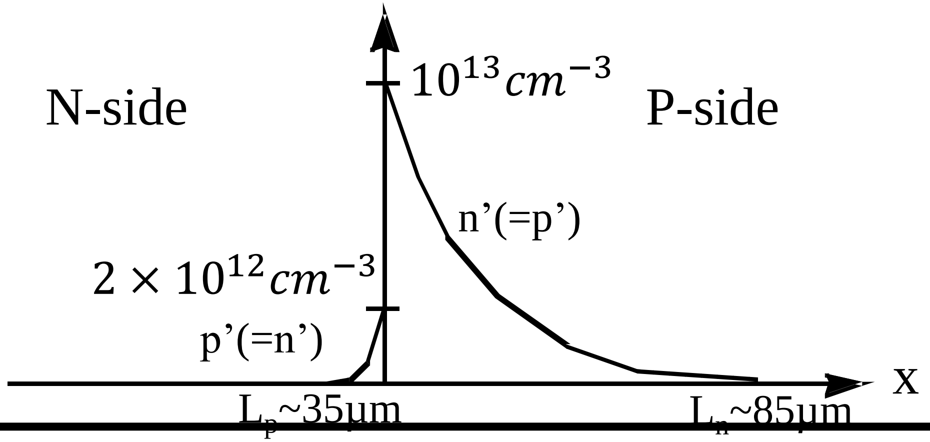
EXAMPLE: Carrier Distribution in Forward-biased PN Diode

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

Forward bias: 0.6V

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

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



4.8 Excess Carrier Distributions

EXAMPLE: Carrier Distribution in Forward-biased PN Diode

- How does L_n compare with a typical device size?

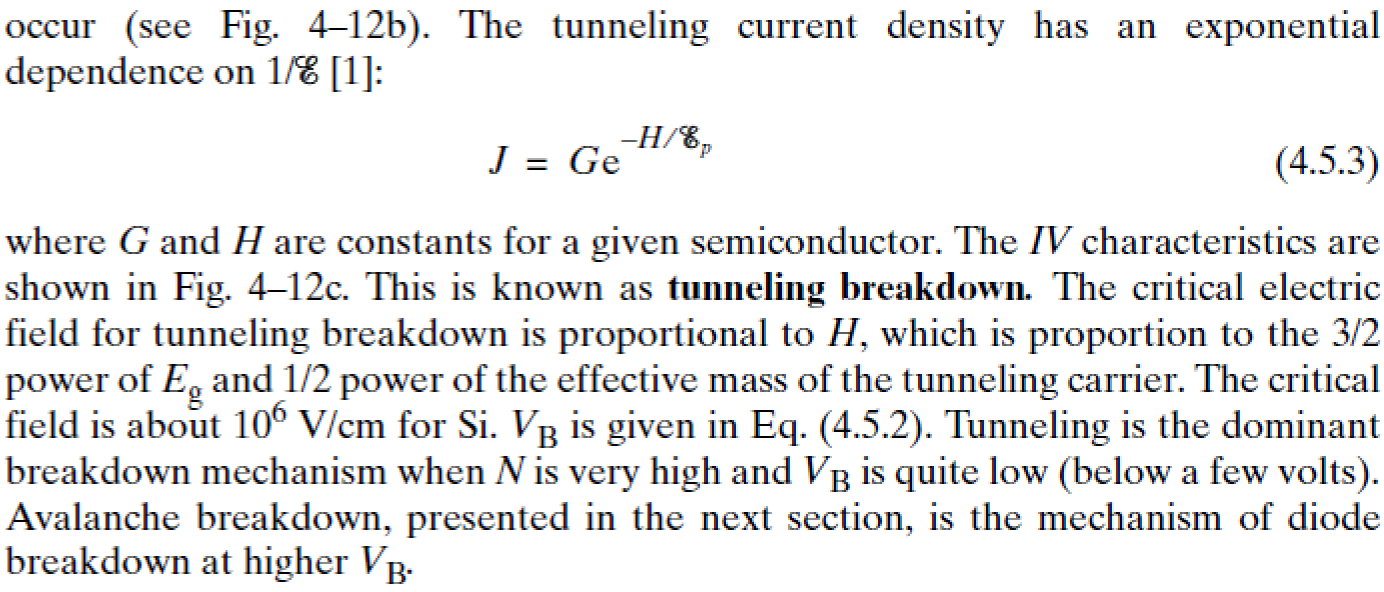
$$L_n = \sqrt{D_n \tau_n} = \sqrt{36 \times 2 \times 10^{-6}} = 85\mu m$$

In VLSI, the device is in nm-size

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

Same as n’(x)

Junction tunneling rate



occur (see Fig. 4–12b). The tunneling current density has an exponential dependence on $1/\mathscr{E}$ [1]:

$$J = G\mathrm{e}^{-H/\mathscr{E}_p} \tag{4.5.3}$$

where G and H are constants for a given semiconductor. The IV characteristics are shown in Fig. 4–12c. This is known as **tunneling breakdown.** The critical electric field for tunneling breakdown is proportional to H, which is proportion to the 3/2 power of E_g and 1/2 power of the effective mass of the tunneling carrier. The critical field is about 10^6 V/cm for Si. V_B is given in Eq. (4.5.2). Tunneling is the dominant breakdown mechanism when N is very high and V_B is quite low (below a few volts). Avalanche breakdown, presented in the next section, is the mechanism of diode breakdown at higher V_B.

5.1 Flat-band Condition and Flat-band Voltage

Q: Gate is n^+ poly silicon, body is p-Si (N_a=1E18)
What is V_{fb}? draw the band diagram

A: Eg/2 + E_i-E_f = 0.56+0.48 = 1.04V, negative
→ V_{FB} = -1.04V

Q: derive ϕ_s as a function of V_{GS}

1. Depletion (including weak inversion)
2. Strong inversion
3. Accumulation

Q : For nFET(p-body) , find Vt

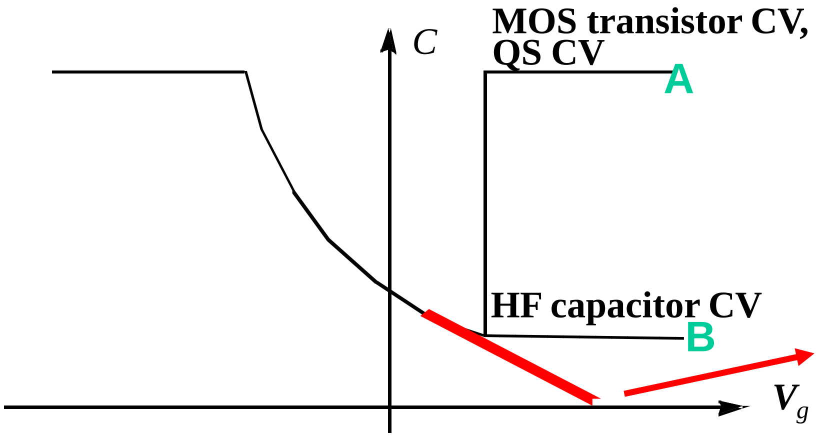
(a) if $\psi_M = E_c$ (b) if $\psi_M = E_v$

t_{ox} = 1nm , N_a=1E18

- PolySi: Vt tuning by body doping and t_{ox} (>45nm node)
- High-k metal gate (45nm and beyond) : Vt tuning by metal work function/high-k dipole

$$V_t = \boxed{V_{fb}} \pm 2\phi_B \pm \boxed{\frac{\sqrt{qN_B 2\varepsilon_s 2\phi_B}}{C_{ox}}}$$

EXAMPLE : CV of MOS Capacitor and Transistor



Does the QS CV or the HF capacitor CV apply?

Deep depletion ($W_d > W_{dm}$)

CCD camera

(1) MOS transistor, 1kHz. (Answer: A *).*

(2) MOS transistor, 2MHz. (Answer: A *).*

(3) MOS capacitor, 2MHz. (Answer: B *).*

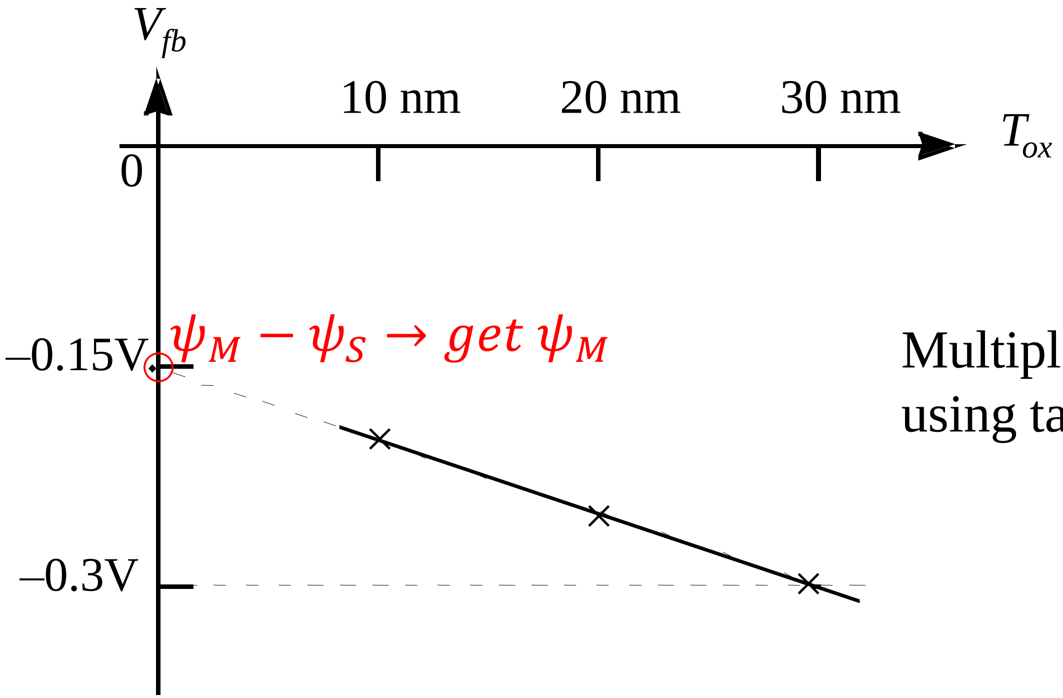
(4) MOS capacitor, 1kHz. (Answer: A *).*

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

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

(7) MOS capacitor, fast V_g *ramp. (Answer: deep depletion)*

EXAMPLE: *To determine* ψ_M.



Multiple oxide thickness
using tapered oxide

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

Solution: $V_{fb} = \psi_M - \psi_S - Q_{ox}T_{ox}/\varepsilon_{ox}$

by CV by doping by CV

Equivalent Oxide Thickness (EOT):

$$I_d = \frac{1}{2}\mu C_{ox}\frac{W}{L}(VG - Vt)^{\alpha}$$

ε_0 = 8.85e-14 F/cm

C_{OX}: area capacitance (fF/μm^2), (F/cm^2)

Ex1: oxide is SiO_2, 1nm, $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$, ε_{ox} =3.9ε_0

C_{OX} ~ 34.5 fF/μm^2

Ex:2 oxide is High K, 1nm C_{OX} $\boxed{= \frac{\varepsilon_{HK}}{t_{HK}} = \frac{\varepsilon_{ox}}{EOT}}$

$\varepsilon_{HK} \geq 20\varepsilon_0, EOT \geq t_{HK}/5$

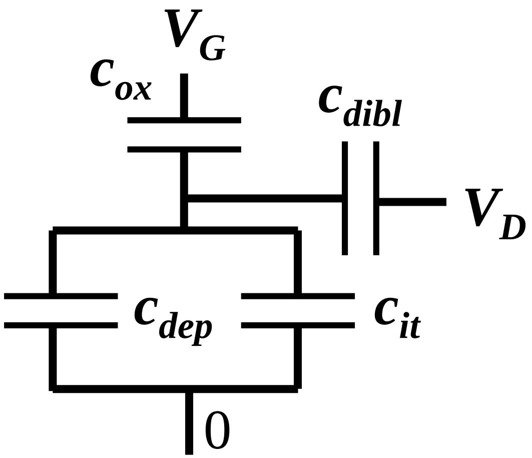
Ex:3 for Silicon EOT, $C_{OX} = \frac{\varepsilon_{si}}{t_{si}} = \frac{\varepsilon_{ox}}{EOT}$

ε_{Si} =11.9ε_0 ~ 3ε_{ox}, EOT ~ t_{si}/3

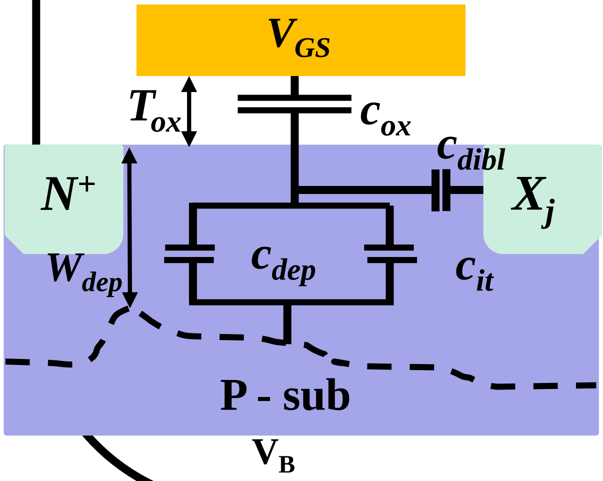
Ex:4 t_{ox}=1nm (minimum value of oxide thickness), I_G=? (Yuan Taur)

Ex:5 EOT=0.7nm, ε_{HK} =20ε_0, → t_{HK}= 3.5nm

How to Suppress the Short-Channel-Effects(SCE)?



1. $C_{dibl} \propto \exp(-\frac{L_{ch}}{l})$, characteristic length $l \equiv \sqrt{\varepsilon_s T_{ox} W_{dmax}/\varepsilon_{ox}}$, $\varepsilon_s = 13.9$ for $\mathrm{In_{0.53}Ga_{0.47}As}$
2. $|\Delta Vth| = \frac{C_{dibl}}{C_{ox}} \times VDS \propto (VDS + c) \times \exp(-\frac{L_{ch}}{l})$
3. $SS = 60 \times (1 + \frac{C_{dep} + C_{dibl} + Cit}{C_{ox}})$



A. Vertical Scaling : $l \propto T_{ox}^{\frac{1}{3}} W_{dmax}^{\frac{1}{3}} X_j^{\frac{2}{3}}$ **by TCAD**

1. **T_{ox}↓, improve the gate controllability**
2. **X_j↓, reduce the electric force from the drain side**
3. **W_{dmax}↓(N_A↑,V_B↓),enhance the ground plane that decrease the electric force from the drain side**

B. Reduce D_{it} :

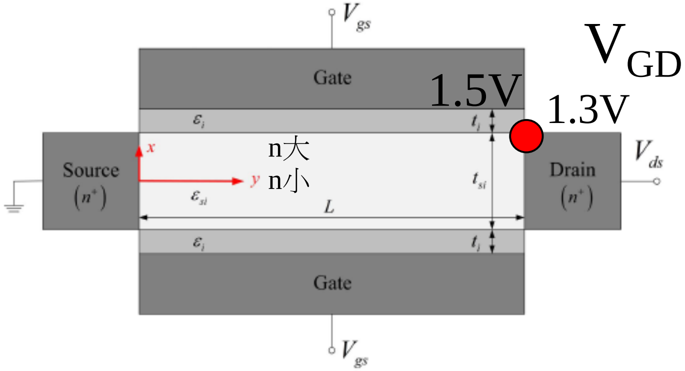
$$\mathbf{W_{dmax}} = \sqrt{\frac{2\varepsilon_{si}(2\varphi_B + V_B)}{qN_A}}$$

$SS = 60 \times (1 + \frac{C_{dep} + C_{dibl} + C_{it}}{C_{ox}})$,

$C_{it} = qDit$ → **improve oxide quality**

Modeling of DG MOSFET I–V Characteristics in the Saturation Region

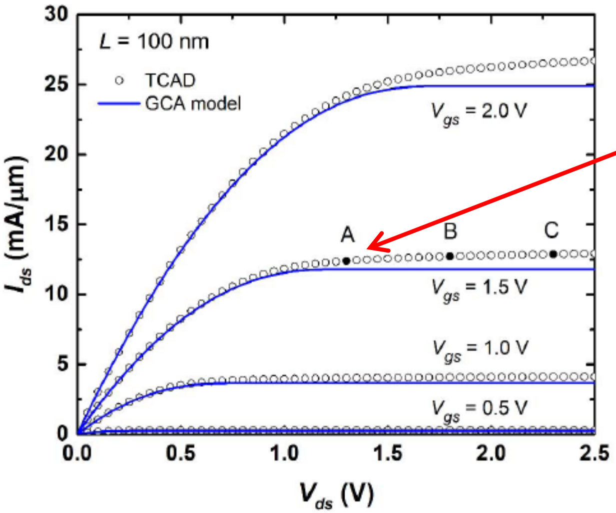
Yuan Taur, *Life Fellow, IEEE*, and Huang-Hsuan Lin, *Student Member, IEEE*

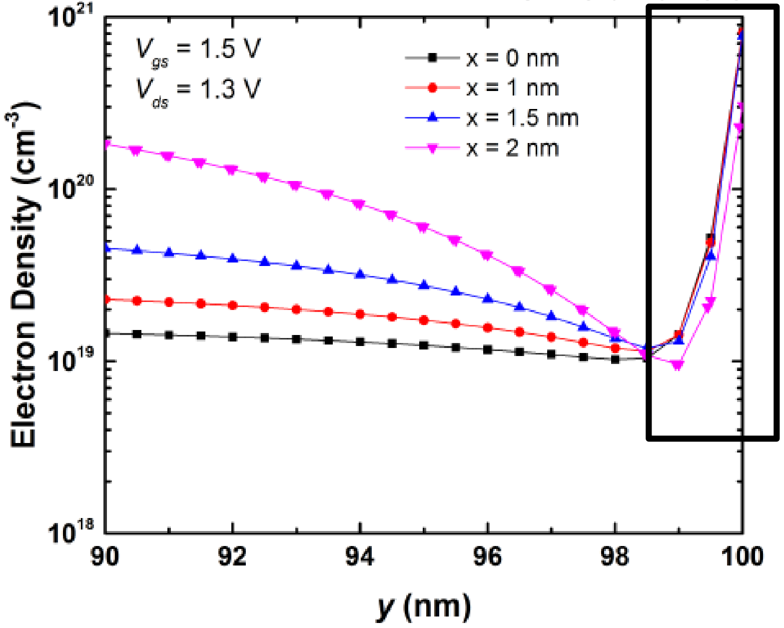


$V_{GD}=0.2V < V_{FB}$ → accumulation

III. “PINCHOFF”—A MYTH IN THE TEXTBOOKS

Should be hole

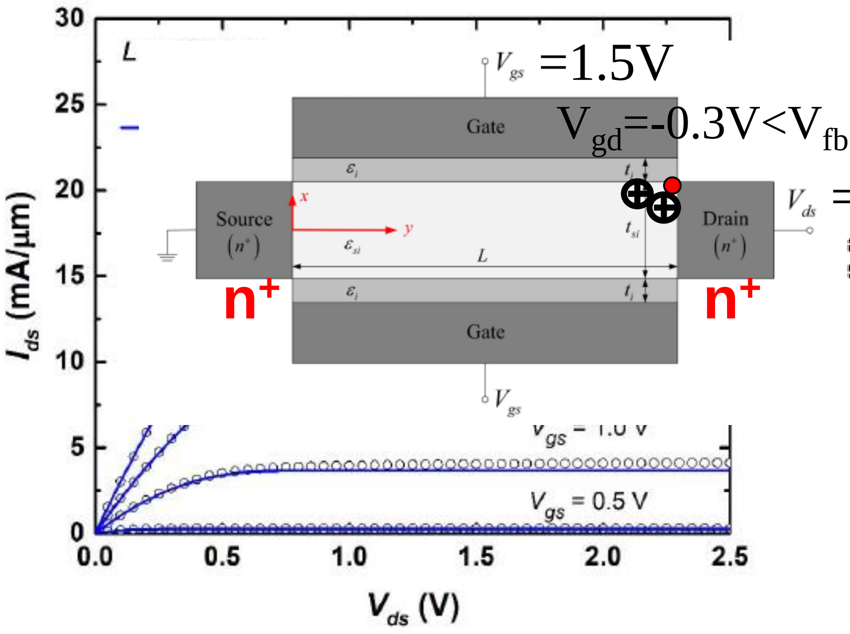


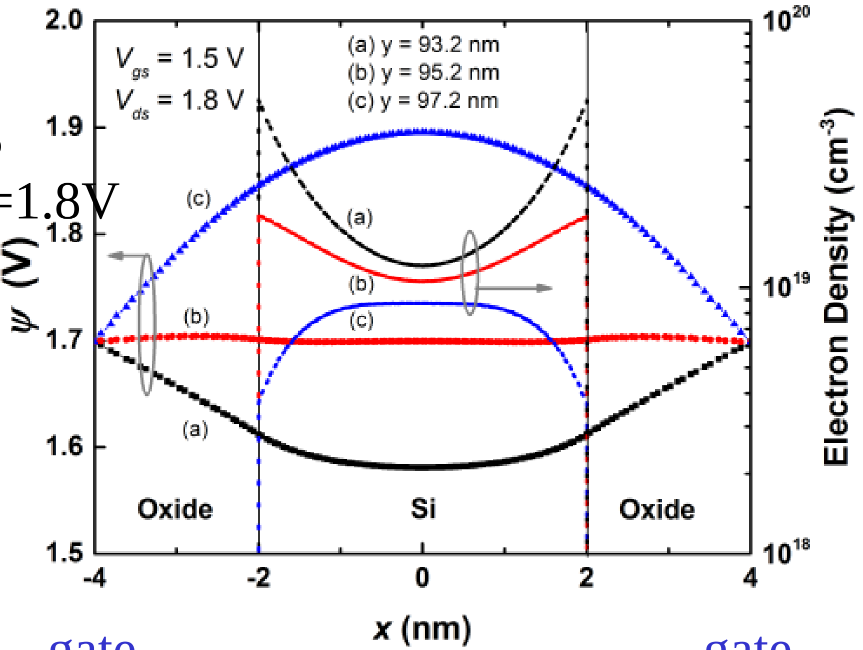


J = neμE

- n of drain ~ 0 → E→inf.

Potential and Carrier Density





gate gate

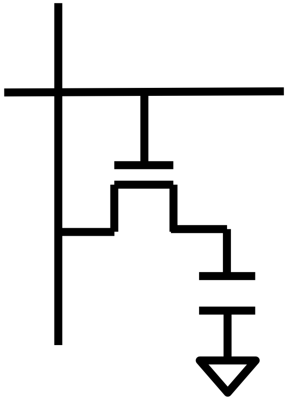
- At the saturation point, what goes to zero is the field in the gate direction, not the electron density.
- Beyond the saturation point, the field in the gate direction becomes negative, so does the curvature $\partial^2 \psi/\partial x^2$.

$\boldsymbol{Q1}: I = 10^{-7}A\ at\ V_t = 0.3V, SS = 100\, mV/dec\, , I_{OFF} = ?$
$\boldsymbol{A1}: I_{OFF} = 10^{-10}A$

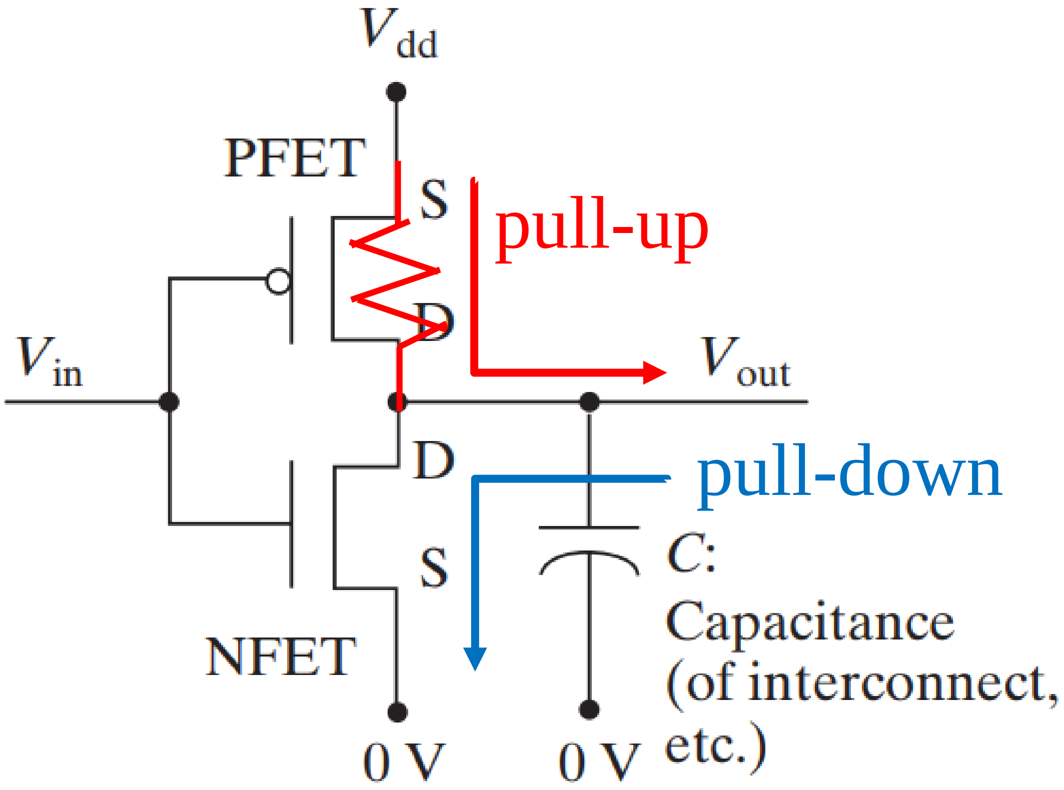
$\boldsymbol{Q2}: Total\ I_{OFF}\ for\ 50\ billion\ transistors = ?$
$\boldsymbol{A2}: Total\ I_{OFF} = 10^{-10}A \times 50 \times 10^9 = 5A$

$\boldsymbol{Q3}: I = 10^{-7}A\ at\ V_t = 0.3V, SS = 60\, mV/dec\, , I_{OFF} = ?$
$\boldsymbol{A3}: I_{OFF} = 10^{-12}A, total\ I_{OFF} = 10^{-12}A \times 50 \times 10^9 = 50mA$

$Logic: V = 0{\sim}V_{DD}, DRAM: I_{OFF} < Logic\ I_{OFF}$
$\rightarrow DRAM\ uses\ negative\ word\ line\ (-0.3{\sim}1V)$
Vgs =-0.3V, Repeat Q1-3



CMOS (Complementary MOS) Inverter



A CMOS inverter is made of a pFET pull-up device and a nFET pull-down device.

Q: V_{out} = ? if V_{in} = 0V. V_{out} = ? if V_{in} = V_{DD} V.

I_{ON} Improvement -- C_{ox}

Equivalent Oxide Thickness (EOT):

C_{OX}: area capacitance (fF/μm^2) or (F/cm^2)

Ex1: oxide is 1nm SiO_2, $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}, \varepsilon_{ox} = 3.9\varepsilon_0$

$C_{ox}\sim 34.5\, fF/\mu m^2$

Ex:2 oxide is 1nm High K, $\boxed{C_{ox} = \frac{\varepsilon_{HK}}{t_{HK}} = \frac{\varepsilon_{ox}}{EOT}}$

$\varepsilon_{HK} = 20\varepsilon_0 \rightarrow EOT = t_{HK}/5$

Ex:3 for 1nm Si, $C_{ox} = \frac{\varepsilon_{si}}{t_{si}} = \frac{\varepsilon_{ox}}{EOT}$

$\varepsilon_{si} = 11.9\varepsilon_0 \sim 3\varepsilon_{ox} \rightarrow EOT \sim t_{si}/3$

Ex:4 $EOT = 0.7nm, \varepsilon_{HK} = 20\varepsilon_0 \rightarrow t_{HK} = 3.5nm$

Ex:5 t_{ox}=1nm (minimum oxide thickness), I_G=?

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

$$Solution: (V_{gs} + V_t + 0.2)/6t_{ox}$$
$$= 1.5V/(12 \times 10^{-7}cm)$$
$$= 1.25\,MV/cm$$

$$E_{avg} = \frac{2Q_{dep} + Q_{inv}}{2\varepsilon_{si}}$$
$$Q_{dep} = qN_B W_{dm}$$
$$Q_{inv} = C_{ox}(V_{gs} - V_t)$$

1MV = 10^6V. From the mobility figure, μ_{ns}=190 cm^2/Vs, which is several times smaller than the bulk mobility (μ_n=1350 cm^2/Vs).

$Q1: If\ V_b < 0 \rightarrow V_t?$ more positive or more negative?

$A1: V_t$ more positive $\rightarrow I_{ds}$ smaller

$Q2: If\ V_\gamma > V_b > 0 \rightarrow V_t?$ more positive or more negative?

$A2: V_t$ more negative $\rightarrow I_{ds}$ bigger

- Body effect can be used for multiple V_T

$$V_t = V_{FB} + \frac{\sqrt{qN_B 2\varepsilon_{si}(2\phi_B + V_{sb})}}{C_{ox}} + 2\phi_B$$

- Tuning V_{FB} for multiple V_T

$$V_{FB} = \boxed{\phi_m} - \phi_s - \boxed{Q_{ox}}/C_{ox}$$

EXAMPLE: Drain Saturation Voltage

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

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

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

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

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

EXAMPLE: Drain Saturation Voltage

$$V_{dsat} = (\frac{m}{V_{gs} - V_t} + \frac{1}{E_{sat}L})^{-1}$$

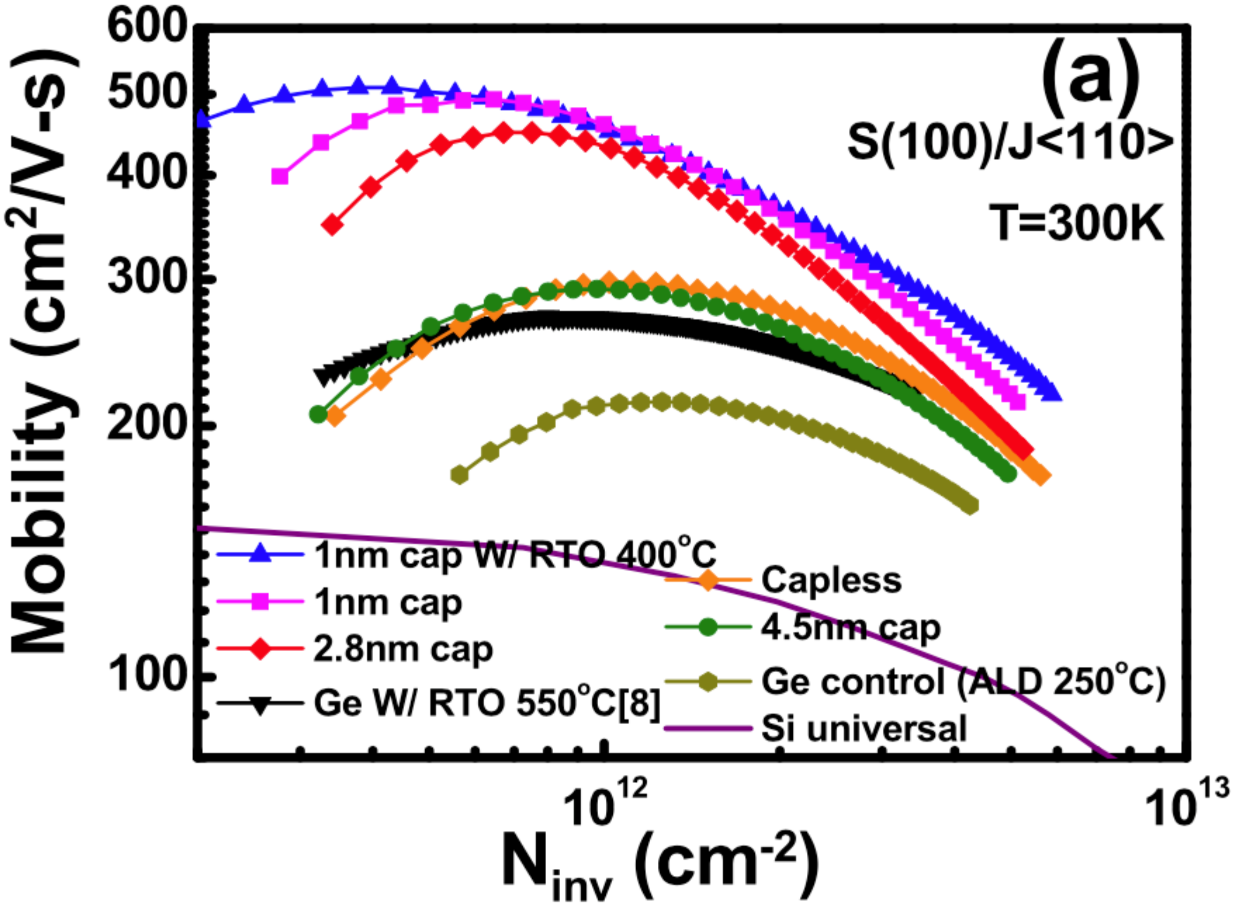
(a) L = 10 μm, V_{dsat}= $(1/1.3V + 1/80V)^{-1}$ = 1.3 V

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

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

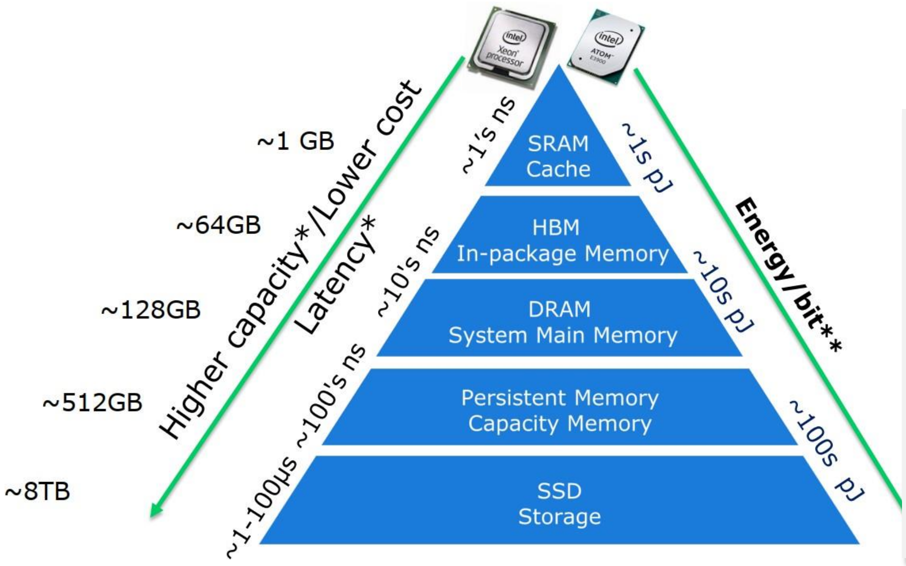
(d) L = 0.05 μm, V_{dsat}= $(1/1.3V + 1/.4V)^{-1}$ = 0.3 V

Mobility of GeSn



Ref: Y.-S. Huang, 2017, TED

Modern Memory Hierarchy



This hierarchical organization of memory works because of the ***Principle of Locality.*** Programs access a relatively small portion of the address space at any moment. There are two different types of locality:

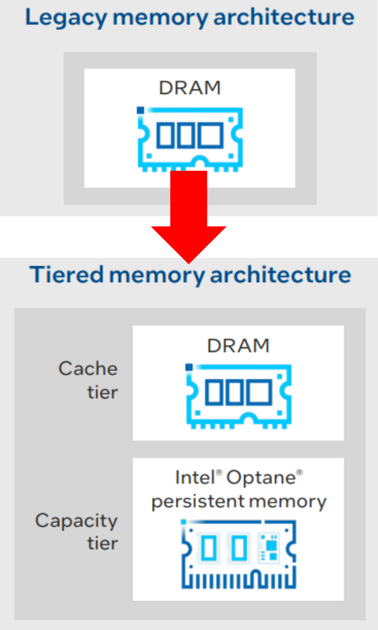
- ***Temporal Locality***: If an item is referenced, it will tend to be referenced again soon
- ***Spatial Locality***: If an item is referenced, items whose addresses are close by tend to be referenced soon

Source: Shanthi, Web

- Modern hardware relies upon temporal and spatial locality of memory accesses to create a high-capacity, high performance and energy efficient virtual memory system

E. Karl, 2023 IEEE International Solid-State Circuits Conference T7:
Fundamentals of Ultra-Low Voltage Embedded Memory Design

Persistent Memory



Ref: S. Yu et al., IEEE Solid-State Circuits Magazine 2016

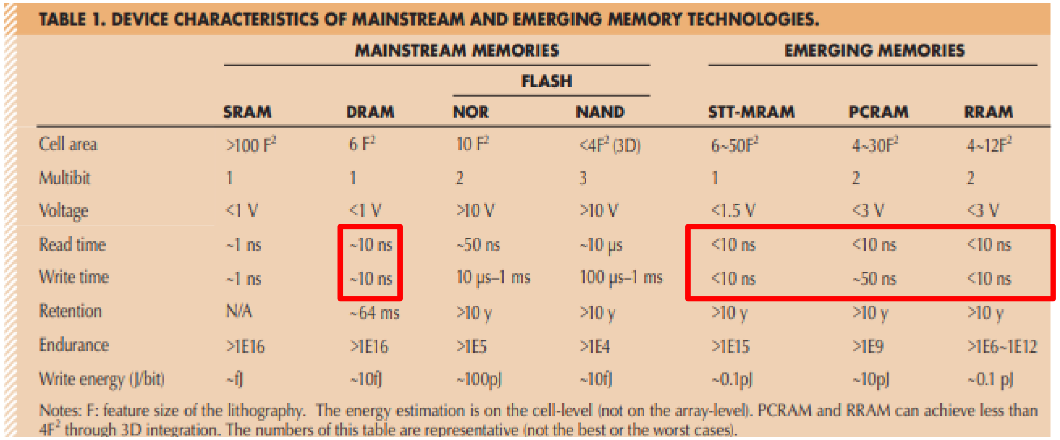


TABLE 1. DEVICE CHARACTERISTICS OF MAINSTREAM AND EMERGING MEMORY TECHNOLOGIES.

	MAINSTREAM MEMORIES				EMERGING MEMORIES		
			FLASH				
	SRAM	DRAM	NOR	NAND	STT-MRAM	PCRAM	RRAM
Cell area	>100 F^2	6 F^2	10 F^2	<4F^2 (3D)	6~50F^2	4~30F^2	4~12F^2
Multibit	1	1	2	3	1	2	2
Voltage	<1 V	<1 V	>10 V	>10 V	<1.5 V	<3 V	<3 V
Read time	~1 ns	~10 ns	~50 ns	~10 μs	<10 ns	<10 ns	<10 ns
Write time	~1 ns	~10 ns	10 μs–1 ms	100 μs–1 ms	<10 ns	~50 ns	<10 ns
Retention	N/A	~64 ms	>10 y	>10 y	>10 y	>10 y	>10 y
Endurance	>1E16	>1E16	>1E5	>1E4	>1E15	>1E9	>1E6~1E12
Write energy (J/bit)	~fJ	~10fJ	~100pJ	~10fJ	~0.1pJ	~10pJ	~0.1 pJ

Notes: F: feature size of the lithography. The energy estimation is on the cell-level (not on the array-level). PCRAM and RRAM can achieve less than 4F^2 through 3D integration. The numbers of this table are representative (not the best or the worst cases).

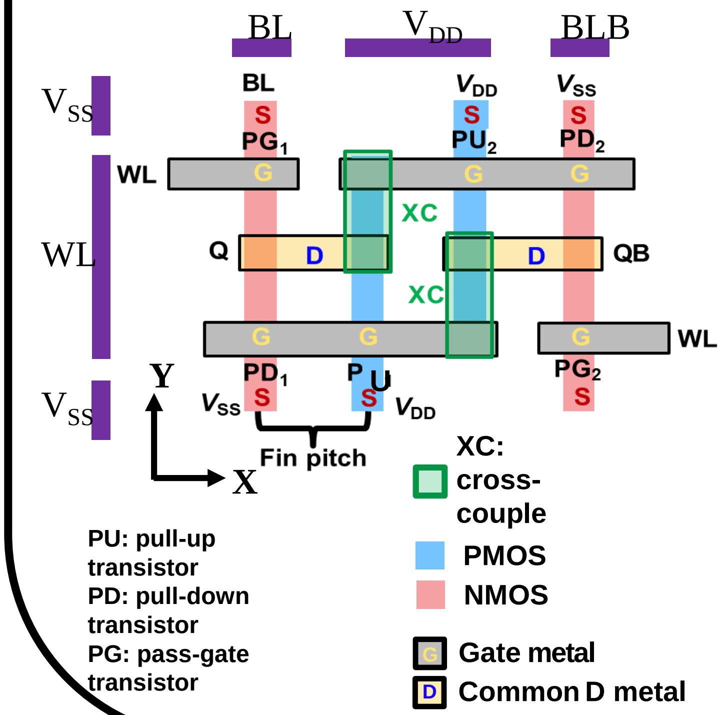
- **To meet the growing gap between DRAM density and demand [1]**
- **Goal:**
  - **Nonvolatile and faster than flash SSDs.**
  - **Nearly the same speed and latency of DRAM but cheaper than it**
  - **Support to write data immediately to the flash when power fails. (NVDIMM and Intel 3D XPoint DIMMs technology)**
- **Candidates: RRAM, PCRAM, MRAM,**

[1] "Solve Data Challenges with Memory Tiering" whitepaper published on the official website of intel

7nm SRAM F^2 calculation

[1] https://en.wikichip.org/wiki/7_nm_lithography_process
[2] Q. Dong et al., ISSCC, 2020 (TSMC)
[3] G. Yeap, et al., IEDM, 2019 (TSMC)

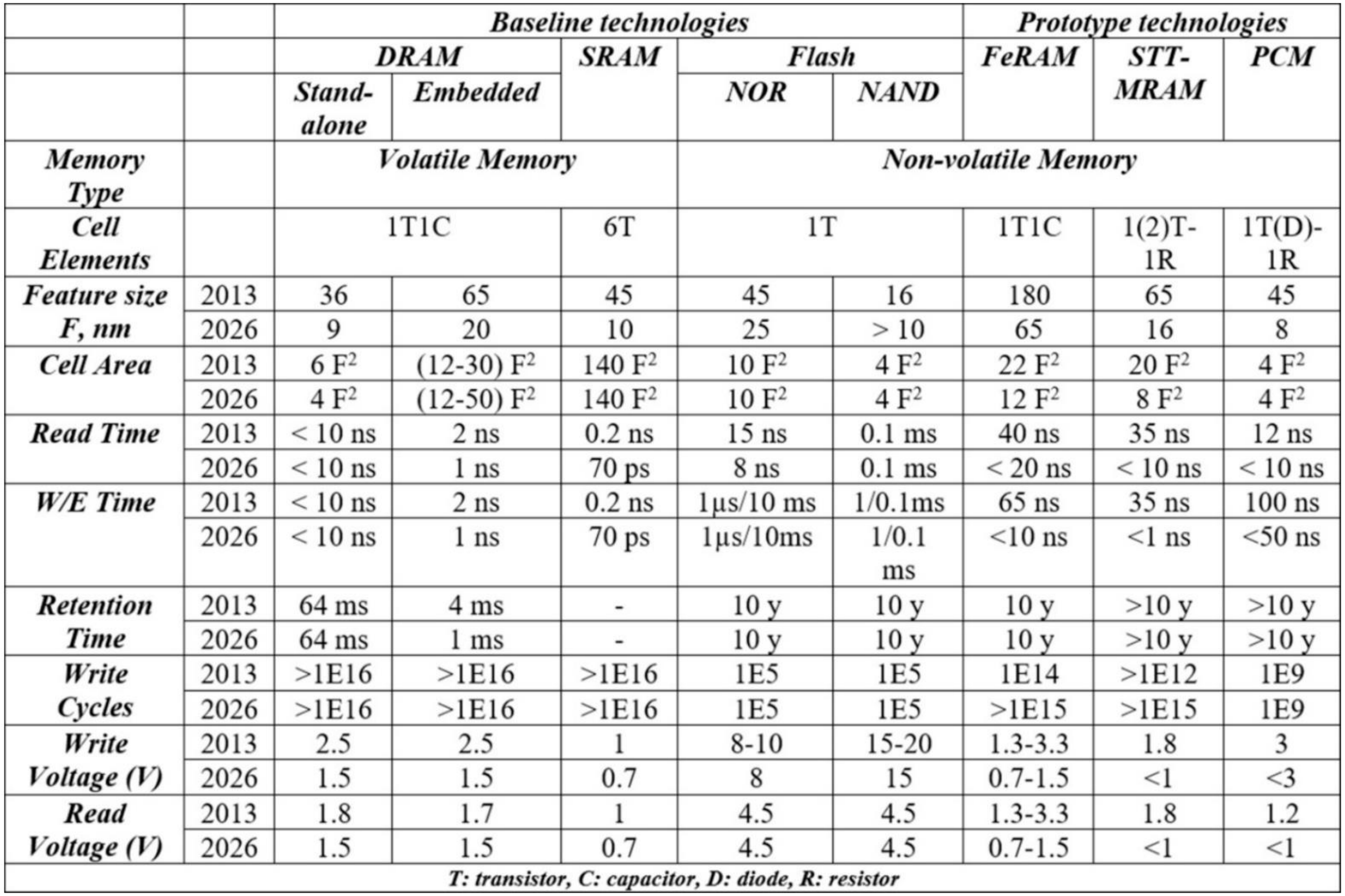
- From wikichip [1], minimum metal pitch is 40nm
- F means half metal pitch=20nm
- 8T CIM SRAM [2] with bitcell area=0.053um^2=53000nm^2 = 53000/20/20 F^2 = 132.5 F^2
- 6T HD SRAM [3] with bitcell area=0.027um^2=27000nm^2= 27000/20/20 F^2 = 67.5 F^2
- Actually, 6T HD SRAM needs 3 metal track in Y direction (BL, V_{DD}, BLB) and 2 metal track in X direction (WL, V_{SS}), but SRAM area is mainly limited by NP isolation.



PU: pull-up transistor
PD: pull-down transistor
PG: pass-gate transistor

🟩 XC: cross-couple
🟦 PMOS
🟥 NMOS
▨ Gate metal
▨ Common D metal

- **Thus, designer would choose wider power rail and WL to boost SRAM performance since metal pitch is not bottleneck of SRAM area.**
- The area consumption of memory implemented by single transistor (such as DRAM and FeFET) is much lower compared to SRAM.



		Baseline technologies					***Prototype technologies***		
		DRAM		***SRAM***	***Flash***		***FeRAM***	***STT-MRAM***	***PCM***
		Stand-alone	***Embedded***		***NOR***	***NAND***			
Memory Type		***Volatile Memory***			***Non-volatile Memory***				
Cell Elements		1T1C		6T	1T		1T1C	1(2)T-1R	1T(D)-1R
Feature size F, nm	2013	36	65	45	45	16	180	65	45
	2026	9	20	10	25	> 10	65	16	8
Cell Area	2013	6 F^2	(12-30) F^2	140 F^2	10 F^2	4 F^2	22 F^2	20 F^2	4 F^2
	2026	4 F^2	(12-50) F^2	140 F^2	10 F^2	4 F^2	12 F^2	8 F^2	4 F^2
Read Time	2013	< 10 ns	2 ns	0.2 ns	15 ns	0.1 ms	40 ns	35 ns	12 ns
	2026	< 10 ns	1 ns	70 ps	8 ns	0.1 ms	< 20 ns	< 10 ns	< 10 ns
W/E Time	2013	< 10 ns	2 ns	0.2 ns	1µs/10 ms	1/0.1ms	65 ns	35 ns	100 ns
	2026	< 10 ns	1 ns	70 ps	1µs/10ms	1/0.1 ms	<10 ns	<1 ns	<50 ns
Retention Time	2013	64 ms	4 ms	-	10 y	10 y	10 y	>10 y	>10 y
	2026	64 ms	1 ms	-	10 y	10 y	10 y	>10 y	>10 y
Write Cycles	2013	>1E16	>1E16	>1E16	1E5	1E5	1E14	>1E12	1E9
	2026	>1E16	>1E16	>1E16	1E5	1E5	>1E15	>1E15	1E9
Write Voltage (V)	2013	2.5	2.5	1	8-10	15-20	1.3-3.3	1.8	3
	2026	1.5	1.5	0.7	8	15	0.7-1.5	<1	<3
Read Voltage (V)	2013	1.8	1.7	1	4.5	4.5	1.3-3.3	1.8	1.2
	2026	1.5	1.5	0.7	4.5	4.5	0.7-1.5	<1	<1
T: transistor, C: capacitor, D: diode, R: resistor									

**Ref: Writam Banerjee,
Electronics 2020**