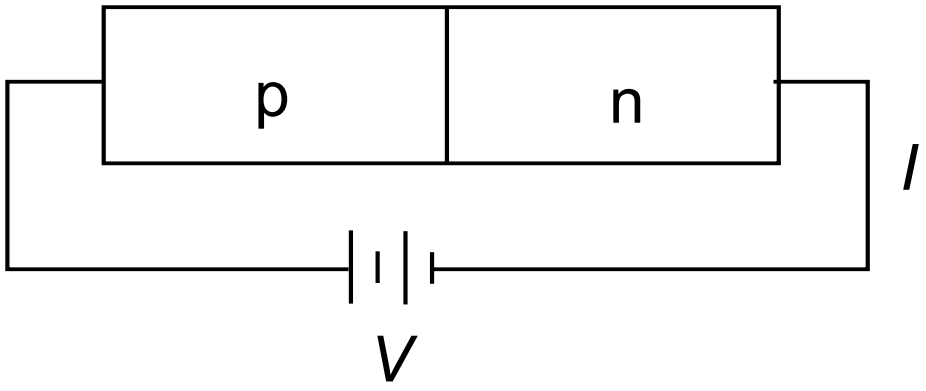
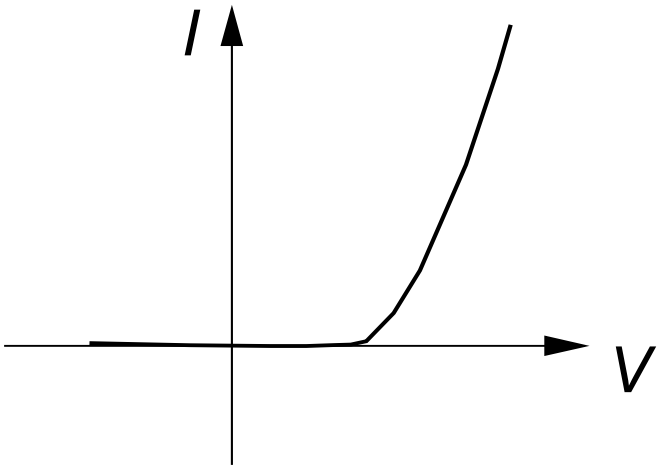
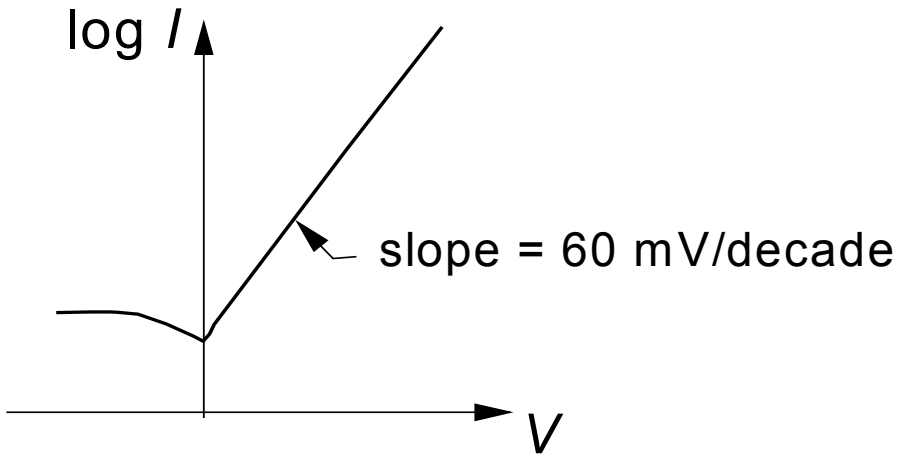
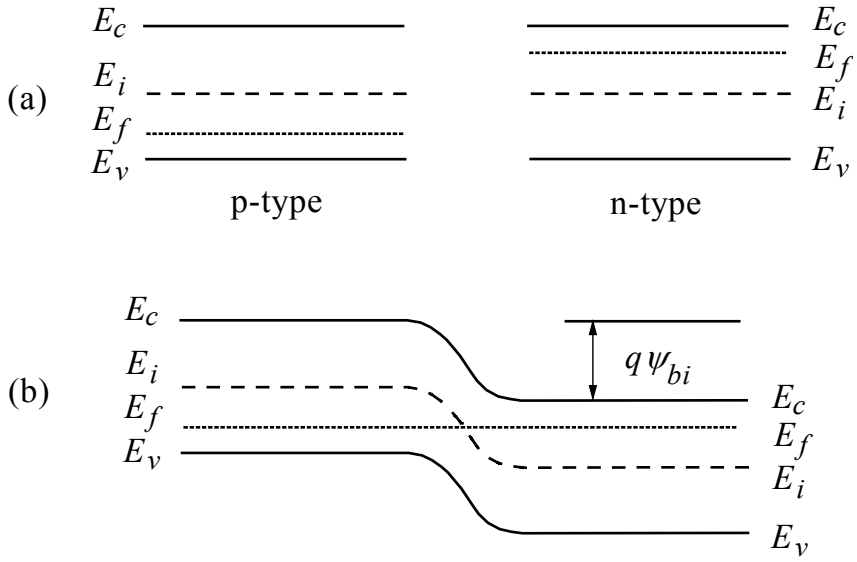
P-N Junctions/Diodes







Static Properties



No net current flow at thermal equilibrium:

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

$$\mathcal{E} = -\frac{d\psi_i}{dx}$$

$$n = n_i e^{(E_f - E_i)/kT} = n_i e^{q(\psi_i - \psi_f)/kT}$$

$$\Rightarrow \quad \frac{dE_f}{dx} = 0$$

Built-in Potential

• Fermi level E_f is spatially constant (flat), causing a built-in potential difference across the diode.

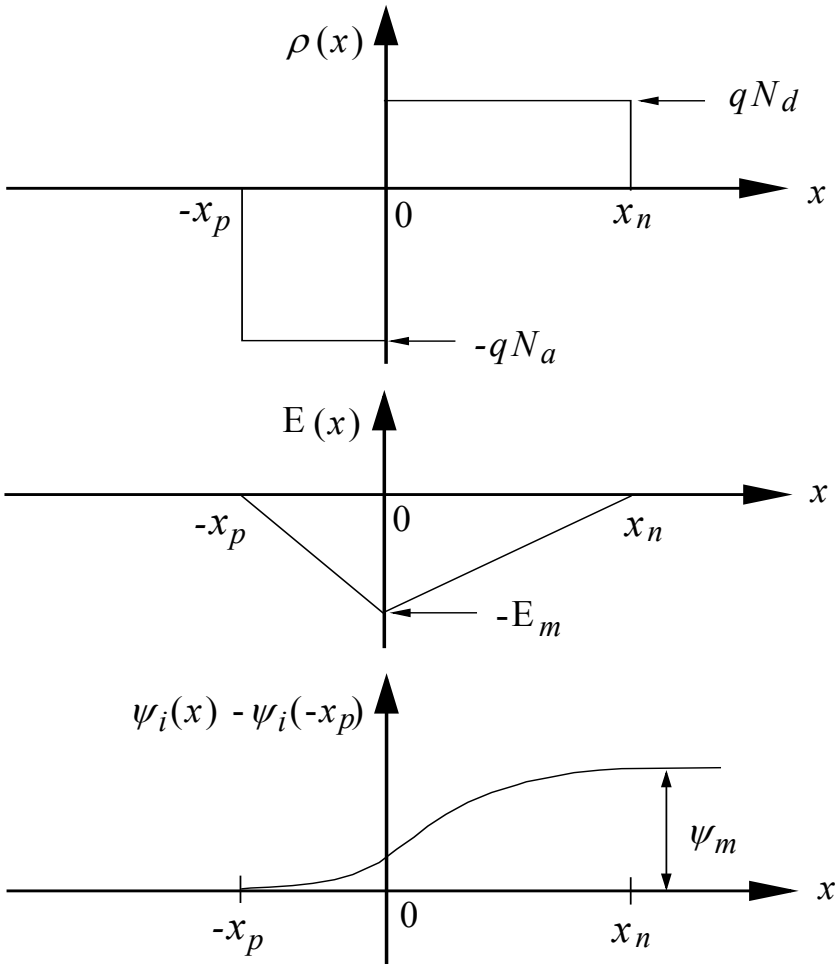
• Built-in potential:

$$q\psi_{bi} = kT\ln\left(\frac{N_a N_d}{n_i^2}\right) = kT\ln\left(\frac{p_{p0}}{p_{n0}}\right) = kT\ln\left(\frac{n_{n0}}{n_{p0}}\right)$$

p_p = (majority) hole density on p-side $\approx N_a$
p_n = (minority) hole density on n-side
n_n = (majority) electron density on n-side $\approx N_d$
n_p = (minority) electron density on p-side

Abrupt Junctions: Depletion Approx.

Quasi-neutral p-region | Depletion region | Quasi-neutral n-region



$$-\frac{d^2\psi_i}{dx^2} = \frac{qN_d}{\varepsilon_{si}} \qquad \text{for } 0 \le x \le x_n$$

$$-\frac{d^2\psi_i}{dx^2} = -\frac{qN_a}{\varepsilon_{si}} \qquad \text{for } -x_p \le x \le 0$$

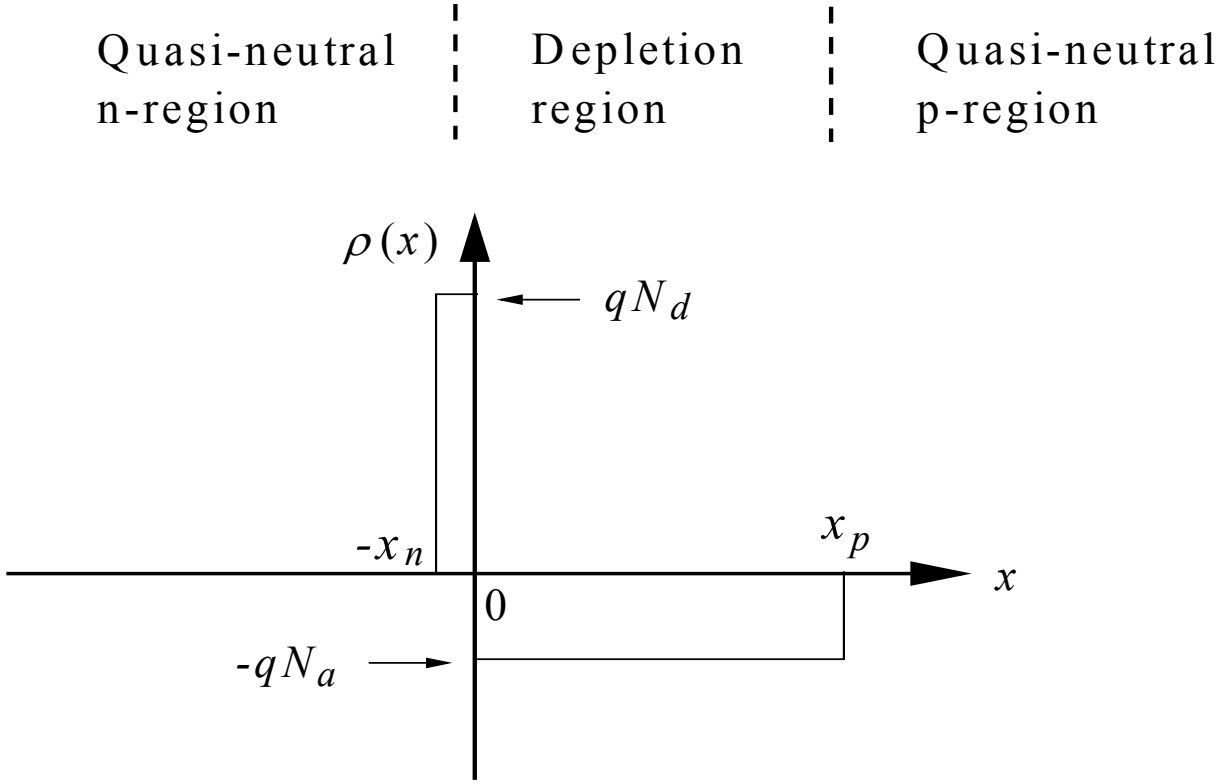
$$\mathrm{E}_m \equiv \left|\frac{-d\psi_i}{dx}\right|_{x=0} = \frac{qN_d x_n}{\varepsilon_{si}} = \frac{qN_a x_p}{\varepsilon_{si}}$$

$$\psi_m = \frac{\mathrm{E}_m(x_n + x_p)}{2} = \frac{\mathrm{E}_m W_d}{2}$$

$$W_d = \sqrt{\frac{2\varepsilon_{si}(N_a + N_d)\psi_m}{qN_a N_d}}$$

$$\psi_m = \psi_{bi} \pm V_{app}$$

One-Sided n^+-p Diode

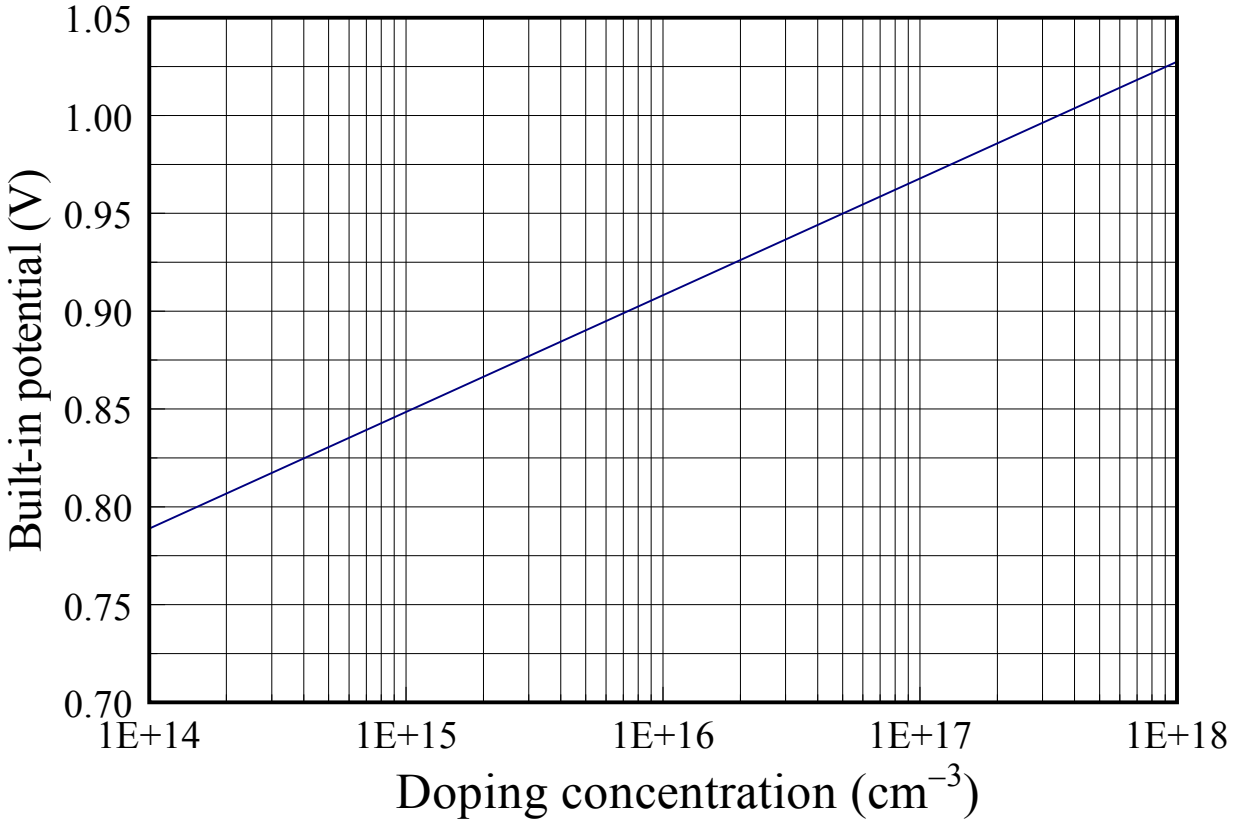


- Charge neutrality: $N_d x_n = N_a x_p$, if $N_d >> N_a \Rightarrow x_p >> x_n$,

i.e., depletion layer and voltage drop primarily appear on the lightly doped side.

One-Sided n^+-p Diode

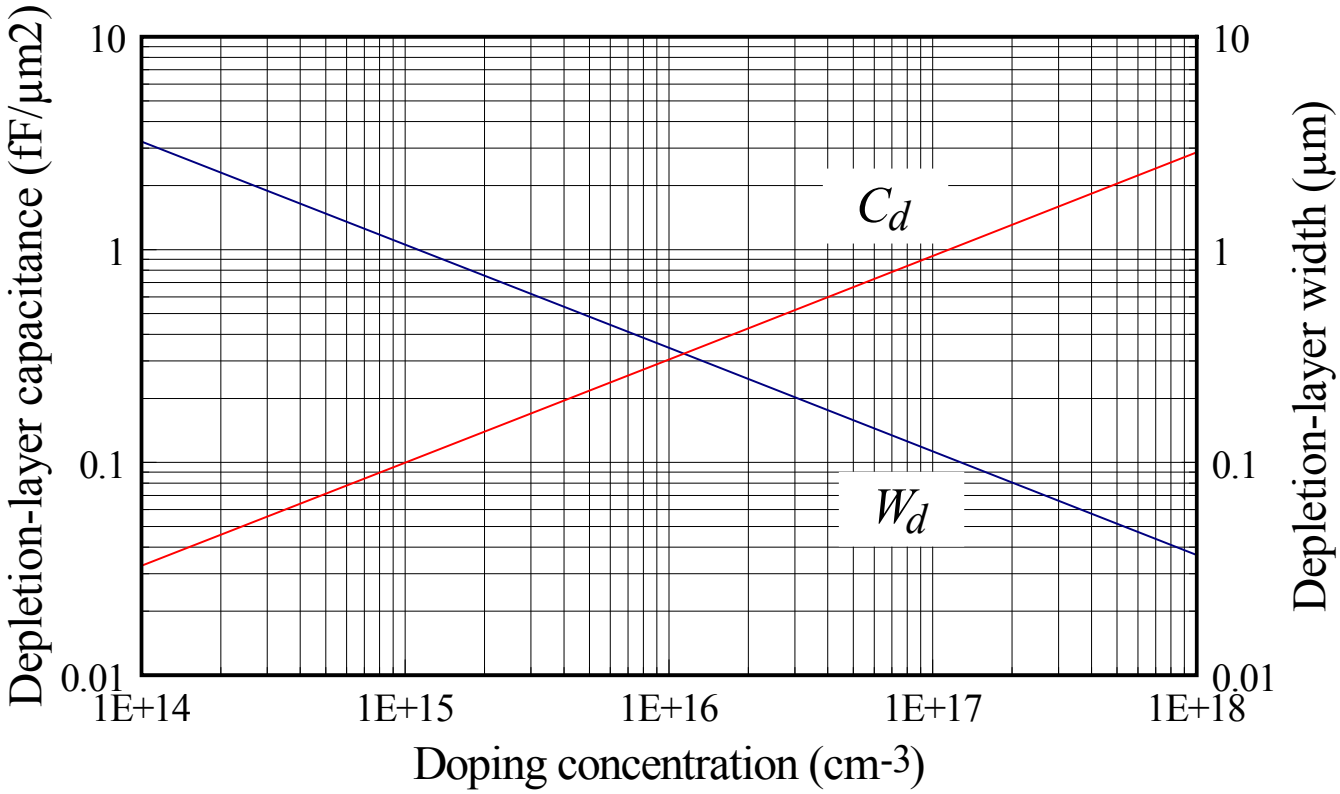
- Built-in potential: $q\psi_{bi} \approx \frac{E_g}{2} + kT \ln\left(\frac{N_a}{n_i}\right)$



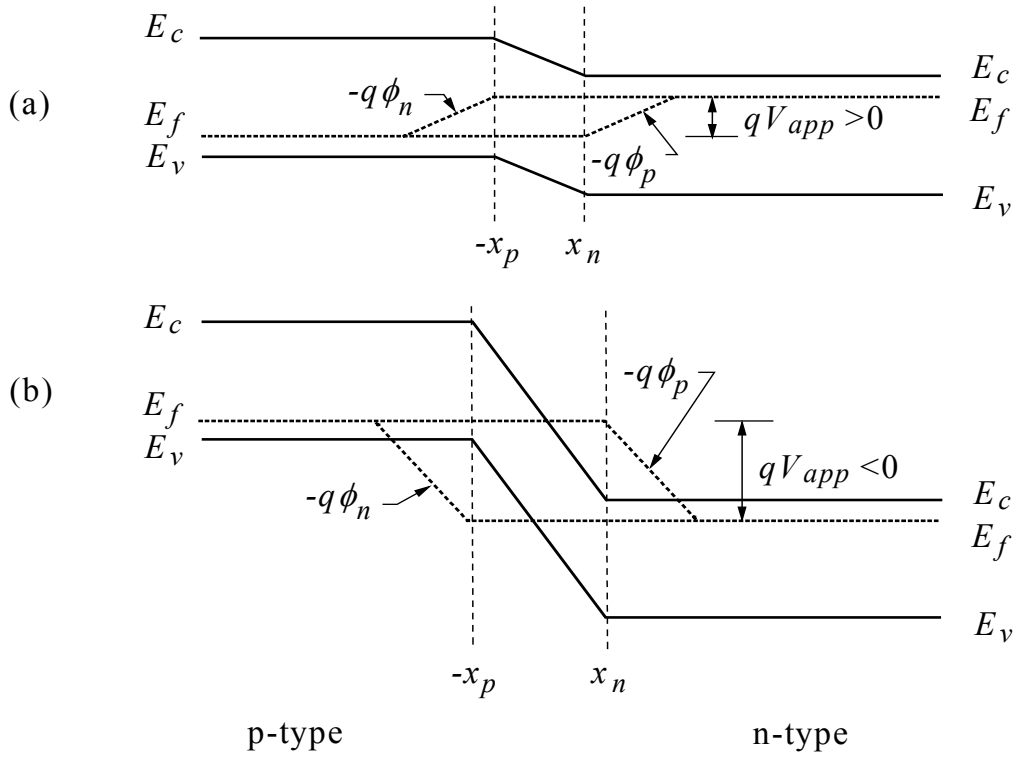
$\Rightarrow$ Built-in potential a weak function of doping conc.

One-Sided n^+-p Diode

- Depletion-layer width: $W_d = \sqrt{\frac{2\varepsilon_{si}(\psi_{bi} \pm V_{app})}{qN_a}} = x_n + x_p \approx x_p$
- Depletion-layer capacitance: $C_d \equiv dQ_d/dV_{app} = \varepsilon_{si}/W_d$



Quasi-Fermi Potentials ϕ_n and ϕ_p



$$\phi_n \equiv \psi_i - \frac{kT}{q}\ln\left(\frac{n}{n_i}\right)$$

$$\phi_p \equiv \psi_i + \frac{kT}{q}\ln\left(\frac{p}{n_i}\right)$$

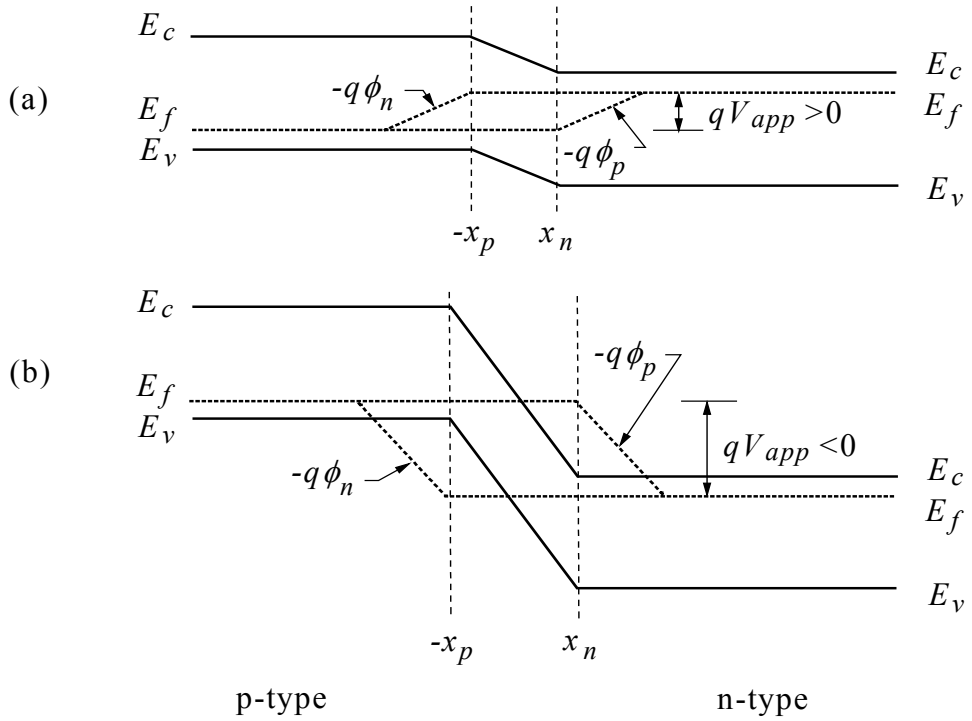
Nonequilibrium near the junction, $pn \neq n_i^2$.

$$pn = n_i^2 \exp\left[q(\phi_p - \phi_n)/kT\right]$$

$$J_n = -qn\mu_n\left(\frac{d\psi_i}{dx} - \frac{kT}{qn}\frac{dn}{dx}\right) = -qn\mu_n\frac{d\phi_n}{dx}$$

$$J_p = -qp\mu_p\left(\frac{d\psi_i}{dx} + \frac{kT}{qp}\frac{dp}{dx}\right) = -qp\mu_p\frac{d\phi_p}{dx}$$

Spatial Variations of ϕ_n and ϕ_p



$V_{app} = \phi_p - \phi_n$ at junction boundaries.

Inside the space-charge region: J_n is constant (neglect *G-R* currents)

$\Rightarrow$ $[n_n \mu_n d\phi_n/dx]_{xn} = [n_p \mu_n d\phi_n/dx]_{-xp}$

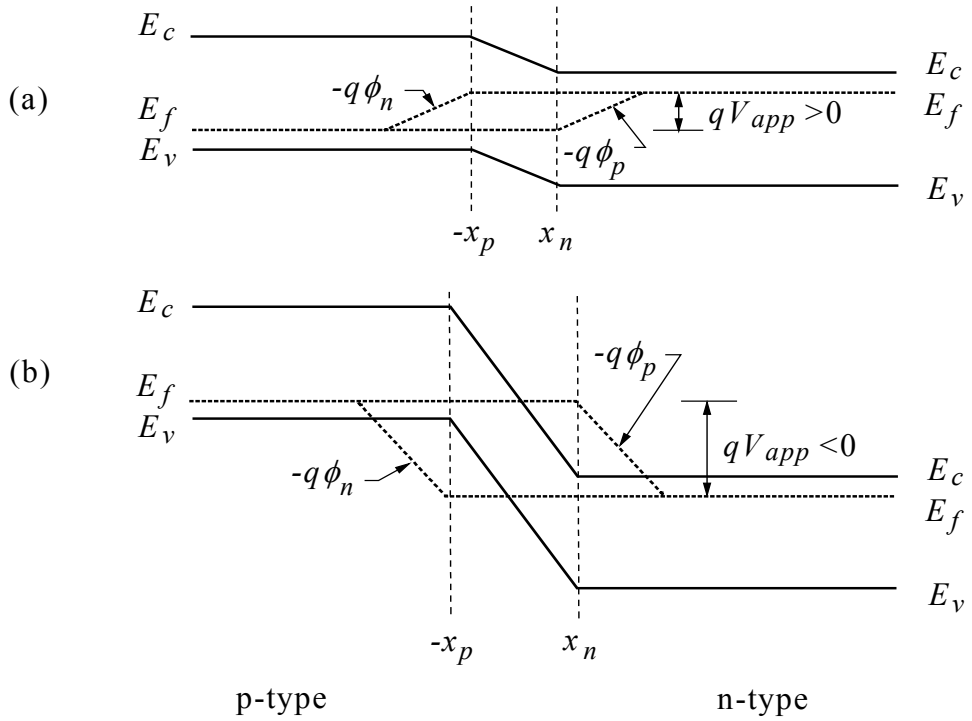
$\Rightarrow$ $[d\phi_n/dx \text{ at } x_n] << [d\phi_n/dx \text{ at } -x_p]$

$\Rightarrow \phi_n \sim$ constant inside space-charge region

Practically all spatial variation in ϕ_n occurs in p-region,

Likewise, all spatial variation in ϕ_p occurs in n-region.

Spatial Variations of ϕ_n and ϕ_p



$V_{app} = \phi_p - \phi_n$ at junction boundaries.

$pn = n_i^2 \exp[q(\phi_p - \phi_n)/kT]$

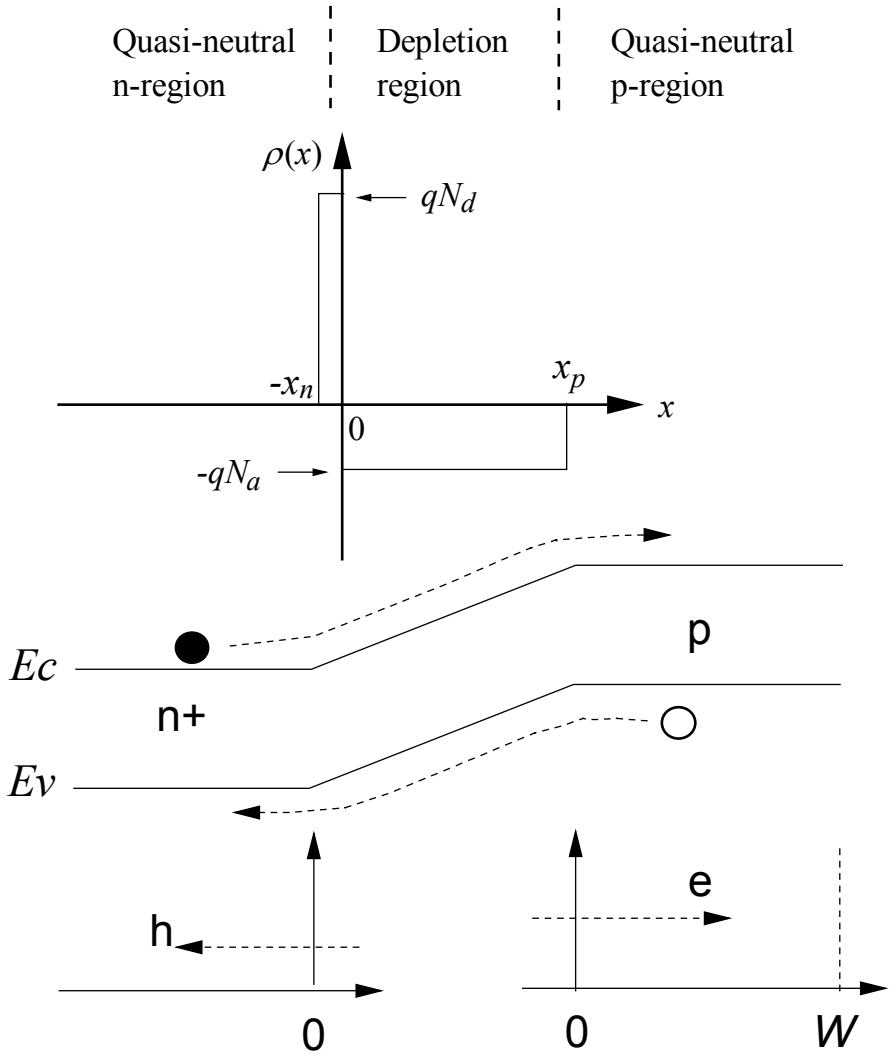
$V_{app} > 0$ for forward bias,
$V_{app} < 0$ for reverse bias.

(1) $n_p(x = -x_p) \approx n_{p0}(x = -x_p) \exp(qV_{app}/kT)$

(2) $p_n(x = x_n) \approx p_{n0}(x = x_n) \exp(qV_{app}/kT)$

(1) and (2) are the most important boundary conditions governing a p-n diode.

Currents in a p-n Junction



- Generation-recombination currents in space-charge region are usually negligible.

⇒ Electron current leaving n-side = electron current entering p-side.
⇒ Hole current leaving p-side = hole current entering n-side.

- ***Need to consider minority carriers and currents only.***

- Total current in diode = electron current + hole current.

Excess Electrons in the p-Region

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial x} - R_n + G_n = 0 \qquad J_n = -qn\mu_n\left(\frac{d\psi}{dx} - \frac{kT}{qn}\frac{dn}{dx}\right) = \mu_n kT\frac{dn}{dx} \qquad R_n - G_n = \frac{n - n_0}{\tau_n}$$

$$\Rightarrow \quad \frac{d^2 n_p}{dx^2} - \frac{n_p - n_{p0}}{L_n^2} = 0, \qquad \text{where} \quad L_n \equiv \sqrt{\tau_n D_n} = \sqrt{\frac{kT\mu_n\tau_n}{q}} \quad \text{is the}$$

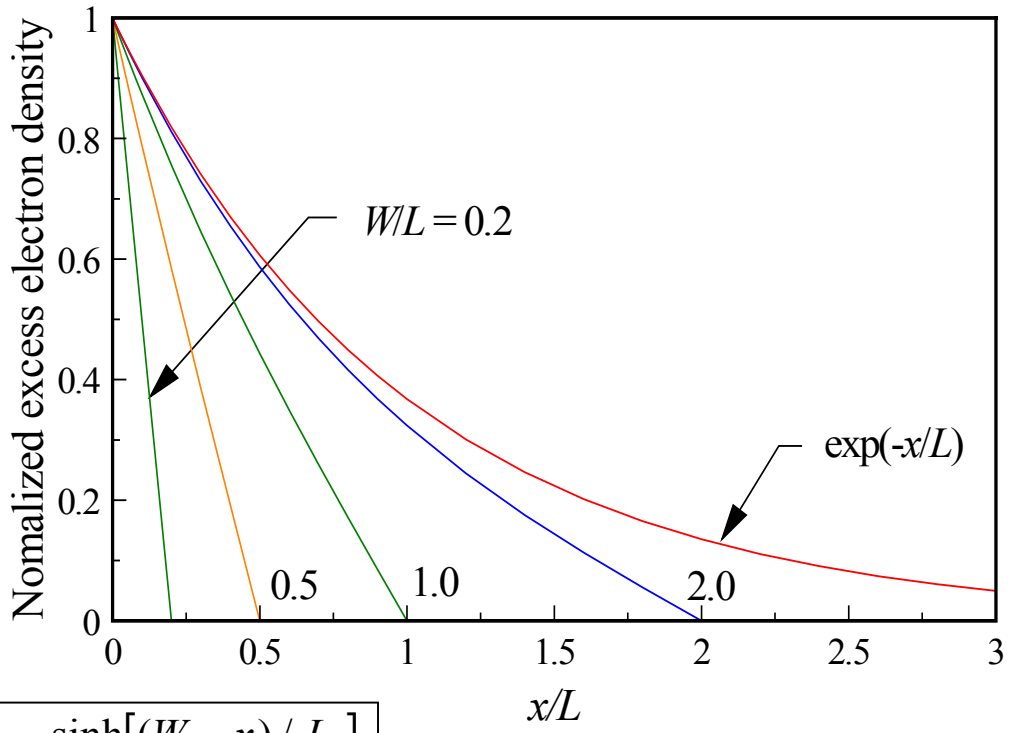
minority carrier diffusion length.

Boundary conditions:

$n_p = n_{p0}\exp\left(qV_{app} / kT\right)$ at x=0,

and

$n_p = n_{p0}$ at x=W (ohmic contact).



$$n_p - n_{p0} = n_{p0}\left[\exp(qV_{app} / kT) - 1\right]\frac{\sinh\left[(W - x) / L_n\right]}{\sinh(W / L_n)}.$$

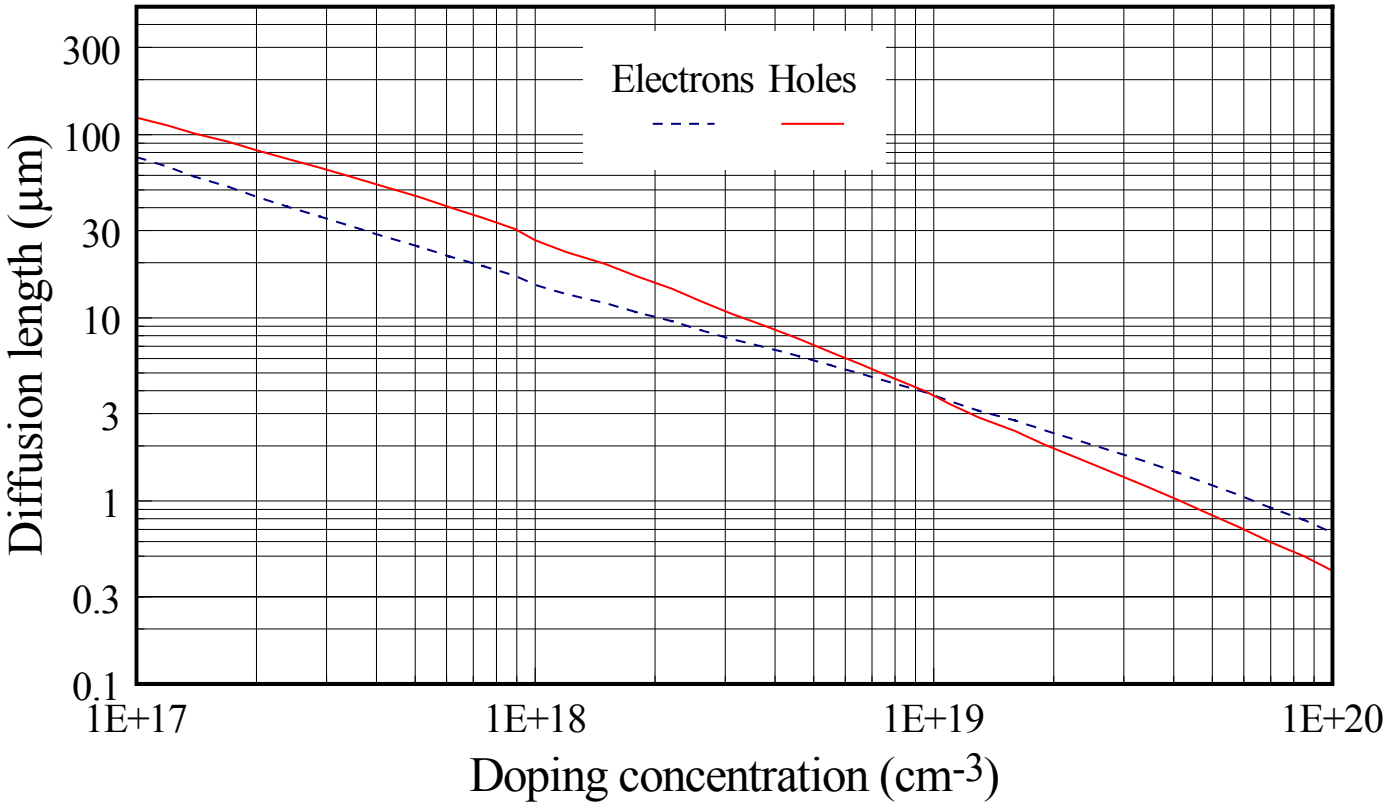
Wide-Base and Narrow-Base Diodes

$$J_n(x=0) = qD_n\left(\frac{dn_p}{dx}\right)_{x=0} = -\frac{qD_n n_{p0}\left[\exp(qV_{app}/kT)-1\right]}{L_n \tanh(W/L_n)} = -\frac{qD_n n_i^2\left[\exp(qV_{app}/kT)-1\right]}{p_{p0}L_n \tanh(W/L_n)}$$

- Wide-base: $W >> L_n$
  - Forward-bias: $J_n(x = 0) = -[qD_n n_i^2/N_a L_n]\exp(qV_{app}/kT)$
    Current increases exponentially with V_{app}, at 60 mV per decade at RT.
  - Reverse-bias: $J_n(x = 0) = +[qD_n n_i^2/N_a L_n]$
    Electrons on p-side but within a diffusion length of the depletion-region boundary diffuse towards n-side.

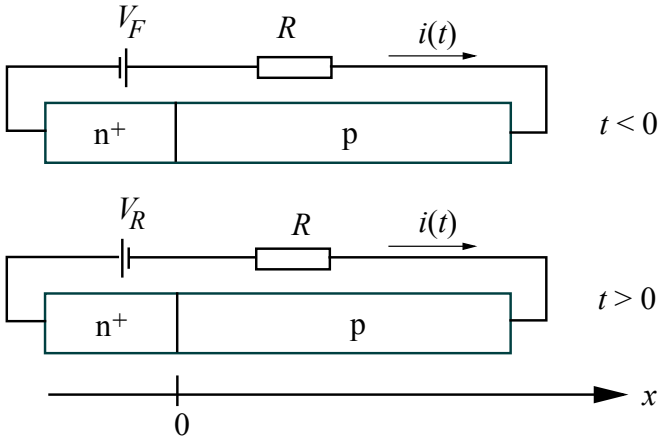
- Narrow-base: $W << L_n$
  - Forward-bias: $J_n(x = 0) = -[qD_n n_i^2/N_a W]\exp(qV_{app}/kT)$
  - Reverse-bias: $J_n(x = 0) = +[qD_n n_i^2/N_a W]$
  - ***Currents increase rapidly as W decreases!***

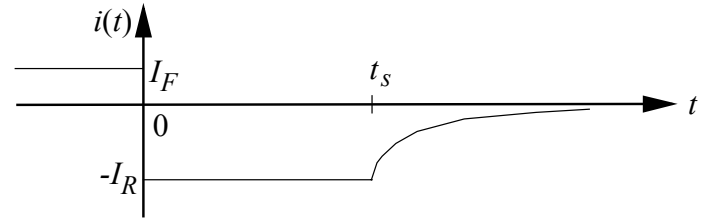
Minority-Carrier Diffusion Length

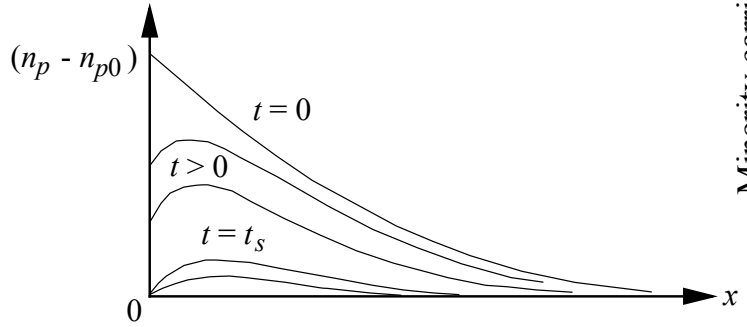


Most p-n junctions in silicon technology are narrow-base diodes. (e.g., source and drain in MOSFETs, emitter and base in bipolars.)

Turning Off a p-n Diode



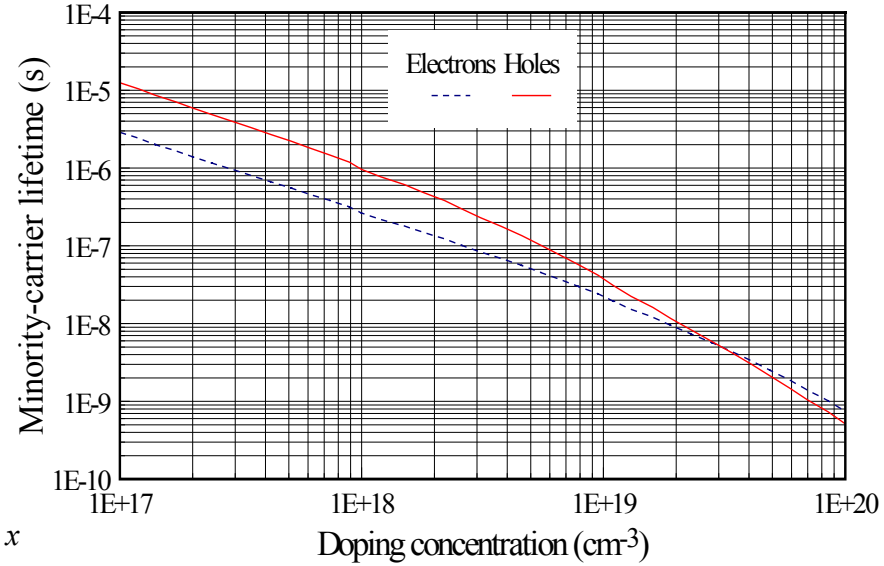




- Excess electrons in p-region

$$Q_B = -q\int_0^W (n_p - n_{p0})dx$$

- Effective turn off starts only after most excess minority carriers have recombined or have drained off.



Diffusion Capacitance

- Diffusion capacitance C_D is due to stored minority carriers responding to applied voltage.

- C_D due to electrons stored in p-type region:

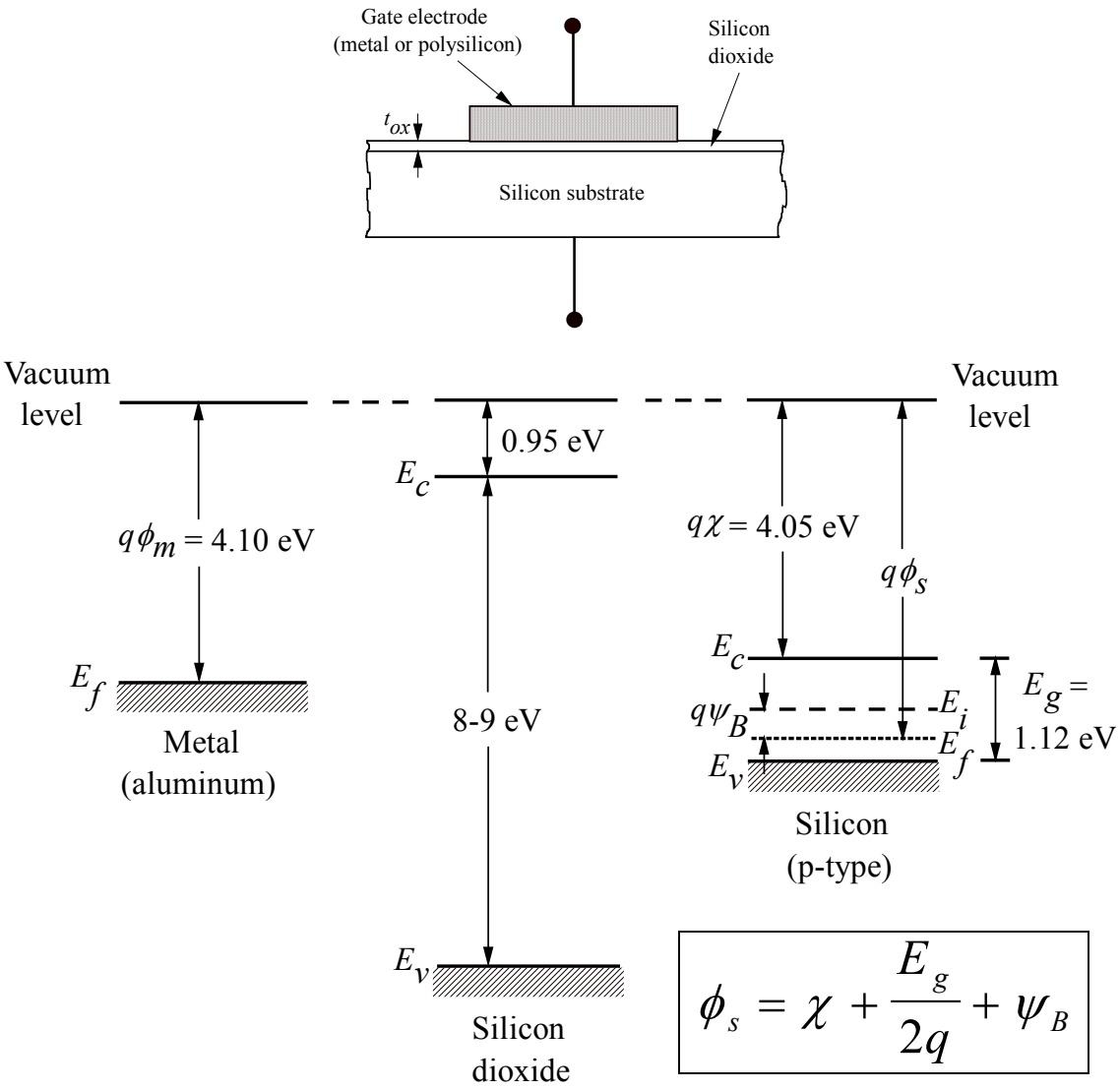
$$C_{Dn} = dQ_B/dV_{app} \propto \exp(qV_{app}/kT)$$

- For a diode or bipolar transistor to switch fast, it must have minimal diffusion capacitance.

- To minimize diffusion capacitance: ***increase doping concentration and minimize charge-storage volume.***

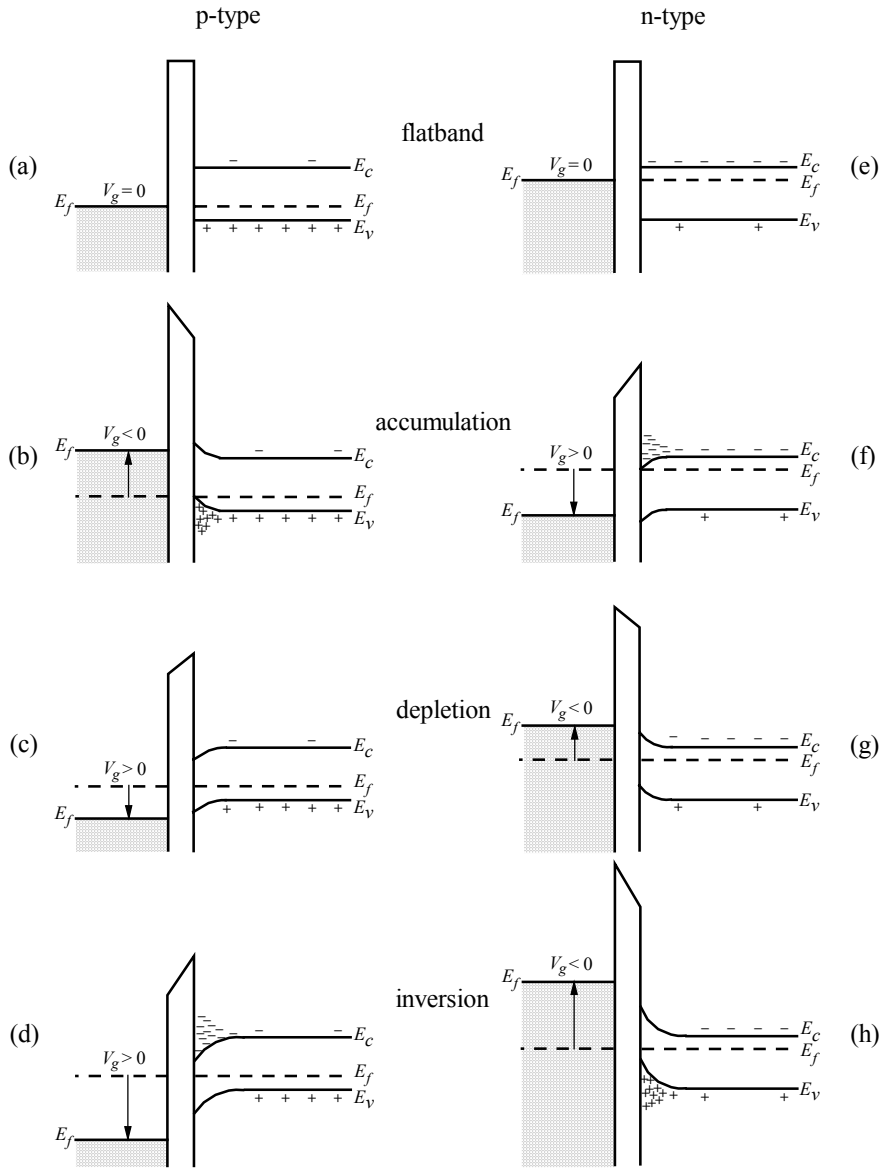
- Modern high-speed bipolar transistors require very thin base.

MOS Device

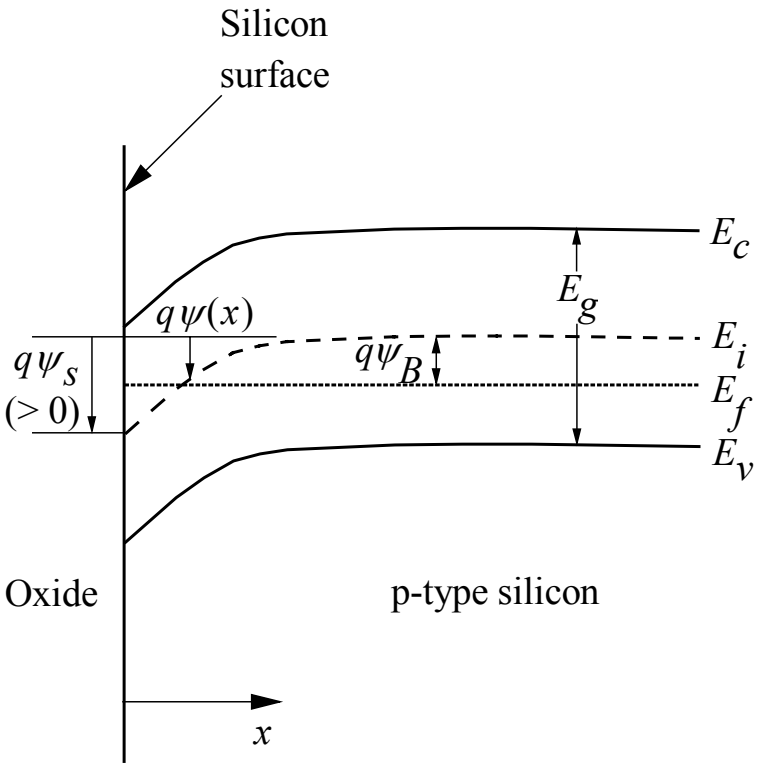

Gate electrode
(metal or polysilicon)
Silicon
dioxide
t_{ox}
Silicon substrate
Vacuum
level
$q\phi_m = 4.10$ eV
E_f
Metal
(aluminum)
0.95 eV
E_c
8-9 eV
E_v
Silicon
dioxide
Vacuum
level
$q\chi = 4.05$ eV
$q\phi_s$
E_c
$q\psi_B$
E_i
E_f
E_v
$E_g =$
1.12 eV
Silicon
(p-type)
$\phi_s = \chi + \frac{E_g}{2q} + \psi_B$

Accumulation, Depletion, Inversion

Assume $\phi_m = \phi_s$:



Poisson's Equation



$$\frac{d^2\psi}{dx^2} = -\frac{d\mathcal{E}}{dx} = -\frac{q}{\varepsilon_{si}}\left[p(x) - n(x) + N_d^{+}(x) - N_a^{-}(x)\right]$$

$$p(x) = n_i e^{q(\psi_f - \psi_i)/kT} = n_i e^{q(\psi_B - \psi)/kT} = N_a e^{-q\psi/kT}$$

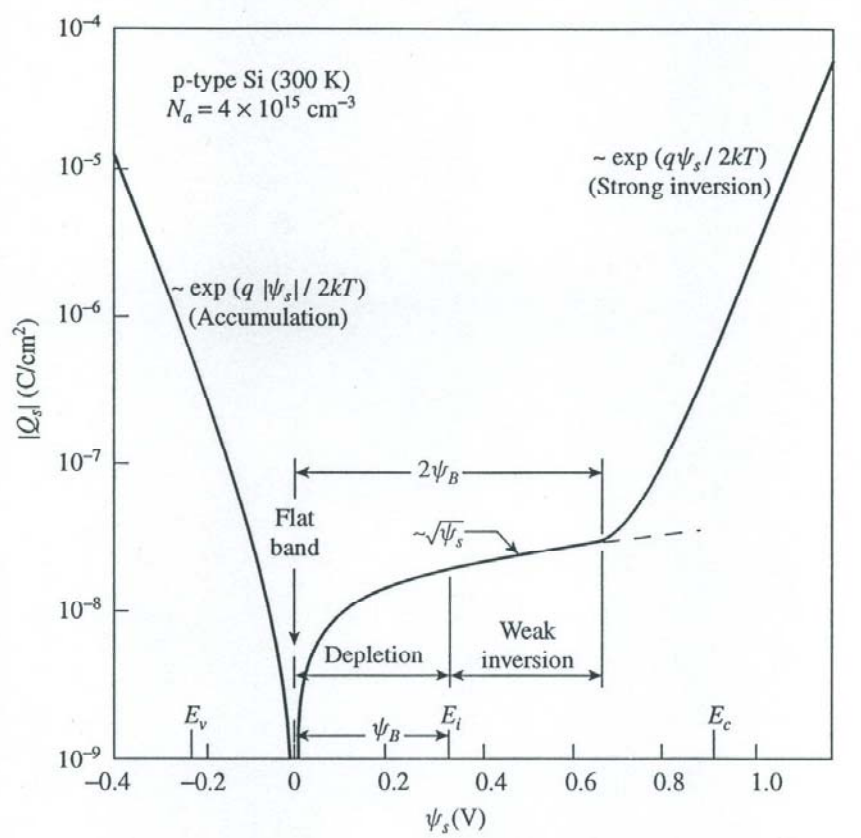
$$n(x) = n_i e^{q(\psi_i - \psi_f)/kT} = n_i e^{q(\psi - \psi_B)/kT} = \frac{n_i^2}{N_a} e^{q\psi/kT}$$

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

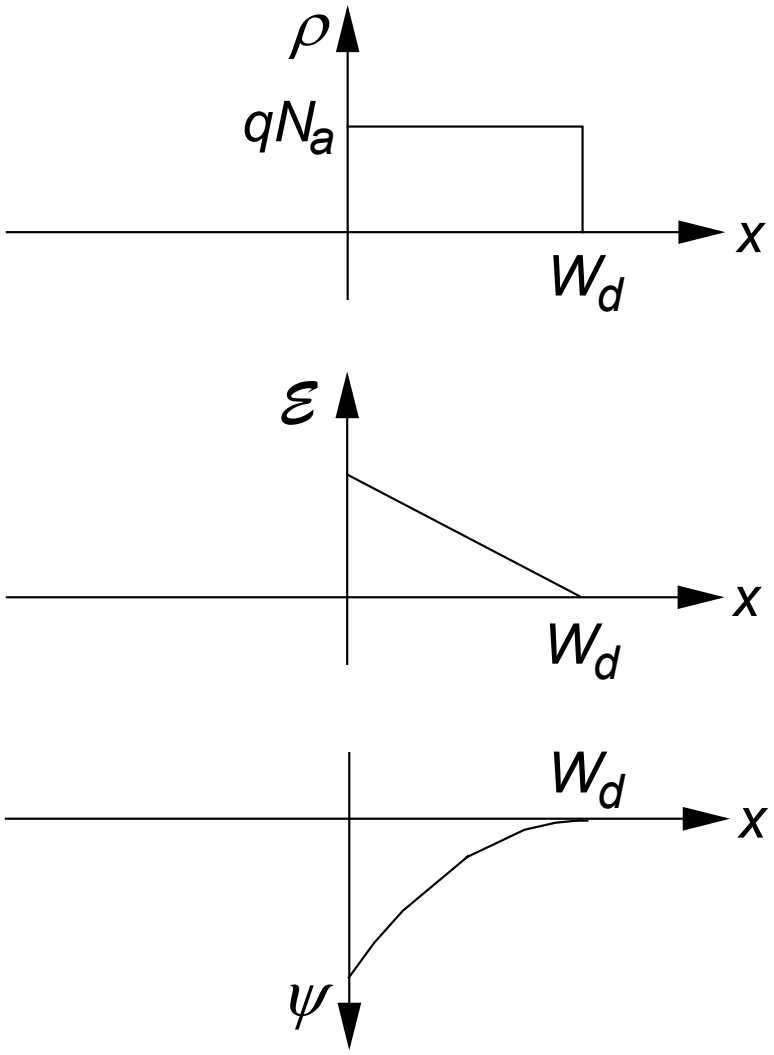
Solving Poisson's Equation

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

$$Q_s = -\varepsilon_{si}\mathcal{E}_s = \pm\sqrt{2\varepsilon_{si}kTN_a}\left[\left(e^{-q\psi_s/kT} + \frac{q\psi_s}{kT} - 1\right) + \frac{n_i^2}{N_a^2}\left(e^{q\psi_s/kT} - \frac{q\psi_s}{kT} - 1\right)\right]^{1/2}$$



Depletion Approximation (1-D Uniform Doping)



$$Q_d = qN_aW_d$$

$$\mathcal{E} = qN_a(W_d - x)/\varepsilon_{si}$$

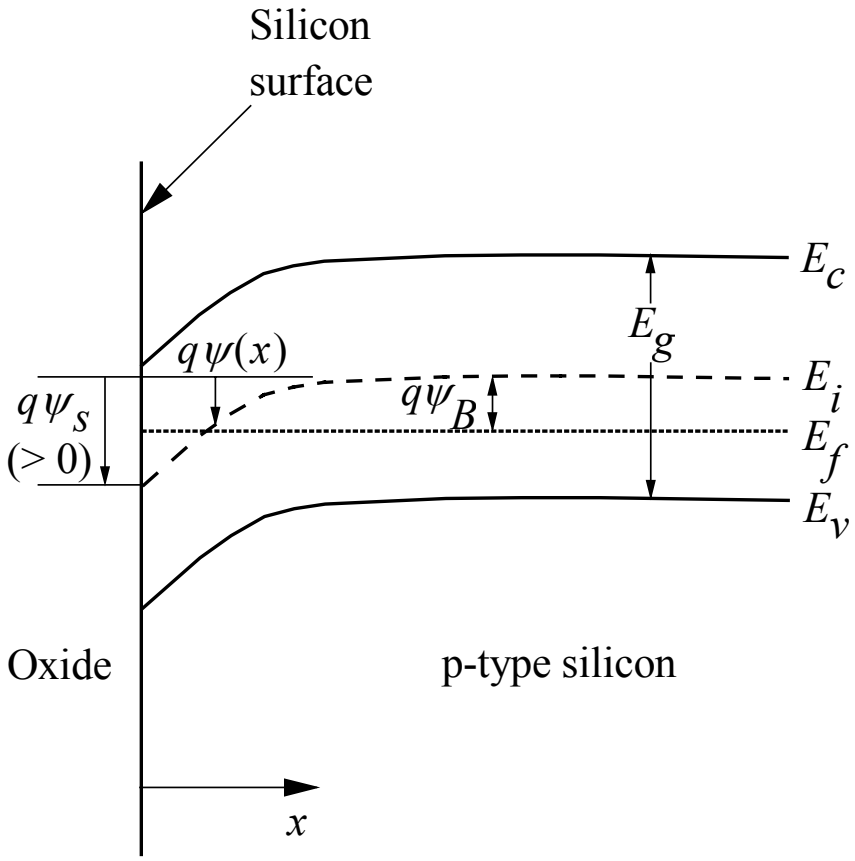
$$\psi = qN_a(W_d - x)^2/2\varepsilon_{si}$$

$$\Rightarrow \quad \psi_s = qN_aW_d^2/2\varepsilon_{si}$$

$$W_d = \sqrt{\frac{2\varepsilon_{si}\psi_s}{qN_a}}$$

$$\psi = \psi_s\left(1 - \frac{x}{W_d}\right)^2$$

Condition for Strong Inversion



$$\psi_s(inv) = 2\psi_B = 2\frac{kT}{q}\ln\left(\frac{N_a}{n_i}\right)$$

i.e., $(n_i^2/N_a^2)\exp(q\psi_s/kT) = 1.$

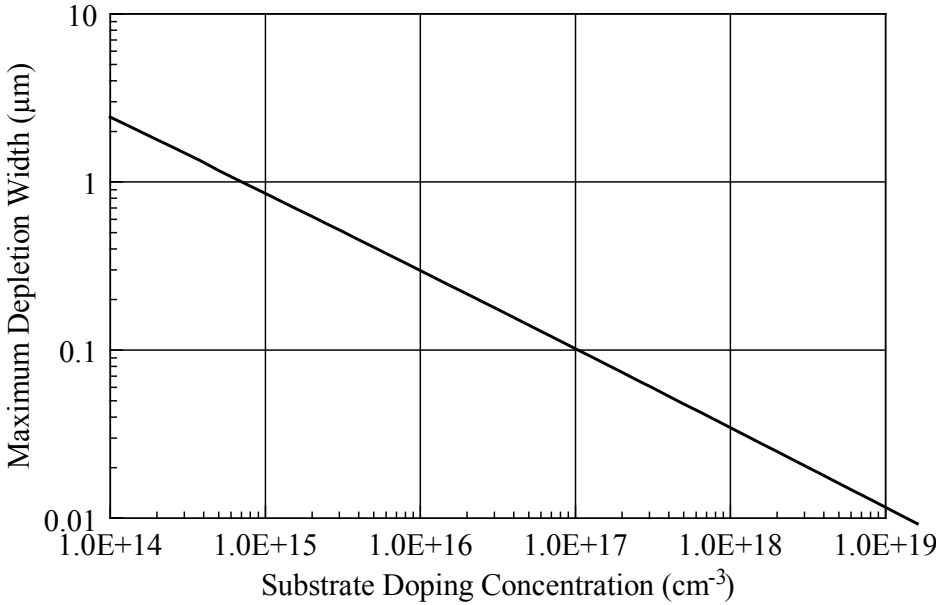
And the electron concentration at the surface equals the hole concentration in the bulk Si.

Max. Depletion Width in MOS (1-D Uniform Doping)

In contrast to p-n junctions, W_d reaches a maximum value W_{dm} at the onset of strong inversion when

$$\psi_s = 2\psi_B = 2(kT/q)\ln(N_a/n_i):$$

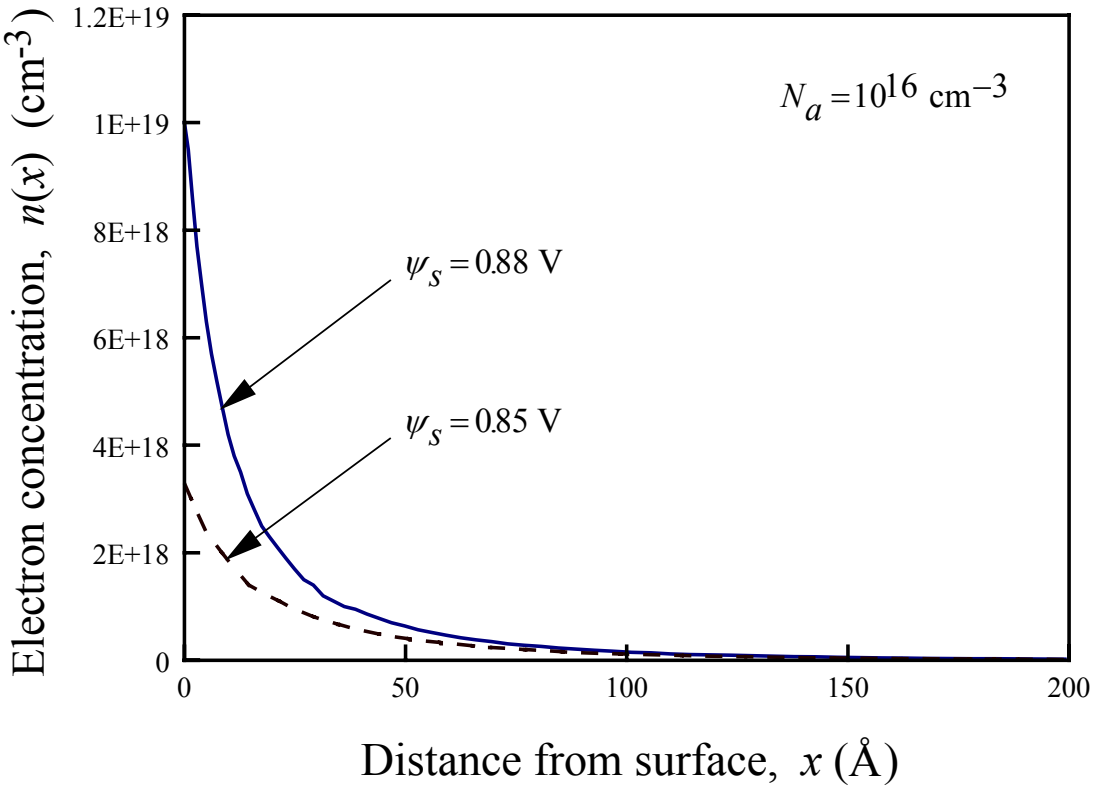
$$W_{dm} = \sqrt{\frac{4\varepsilon_{si} kT \ln(N_a / n_i)}{q^2 N_a}}$$



This defines the threshold condition of a MOSFET. W_{dm} also plays a key role in the short-channel scaling of a MOSFET, namely, $L_{min} \propto W_{dm}$.

Strong Inversion

$$\frac{d\psi}{dx} = -\sqrt{\frac{2kTN_a}{\varepsilon_{si}}\left(\frac{q\psi}{kT} + \frac{n_i^{\,2}}{N_a^{\,2}}e^{q\psi/kT}\right)}$$



Charge per area:

$$Q_i = -\sqrt{\frac{2\varepsilon_{si}kTn_i^{\,2}}{N_a}}e^{q\psi_s/2kT}$$

Electron conc. at surface:

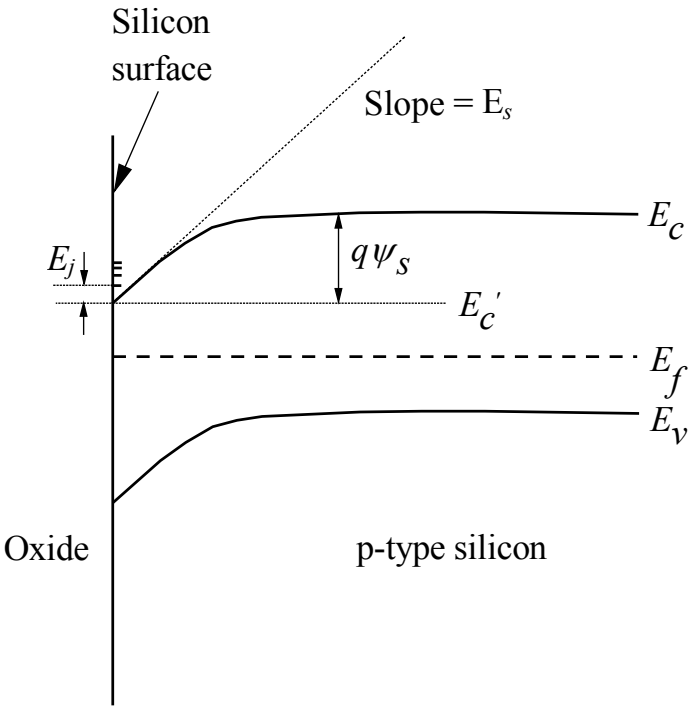
$$n(0) = \frac{n_i^{\,2}}{N_a}e^{q\psi_s/kT}$$

Inversion layer thickness:

$Q_i/qn(0) = 2\varepsilon_{si}kT/qQ_i$

Quantum Effect in MOS Inversion

In an MOS inversion layer, carriers are confined in the direction perpendicular to the surface and therefore need be treated quantum mechanically (2-D).



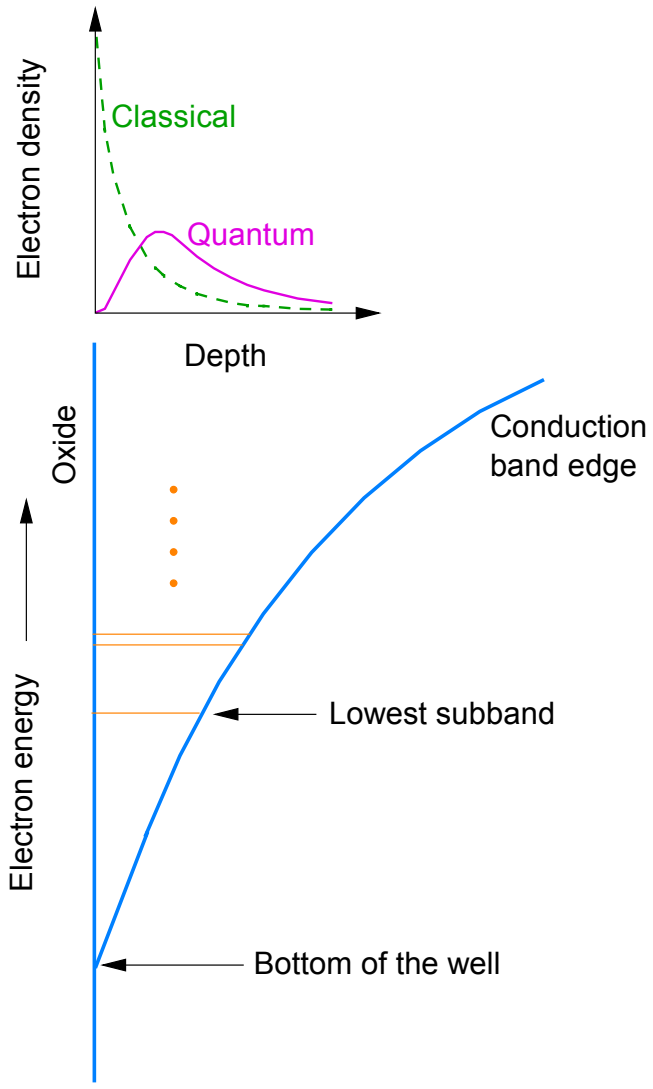
Discrete energy levels:

$$E_j = \left[\frac{3hq\mathrm{E}_s}{4\sqrt{2m_x}}\left(j+\frac{3}{4}\right)\right]^{2/3}$$

Average distance of inversion layer from the surface:

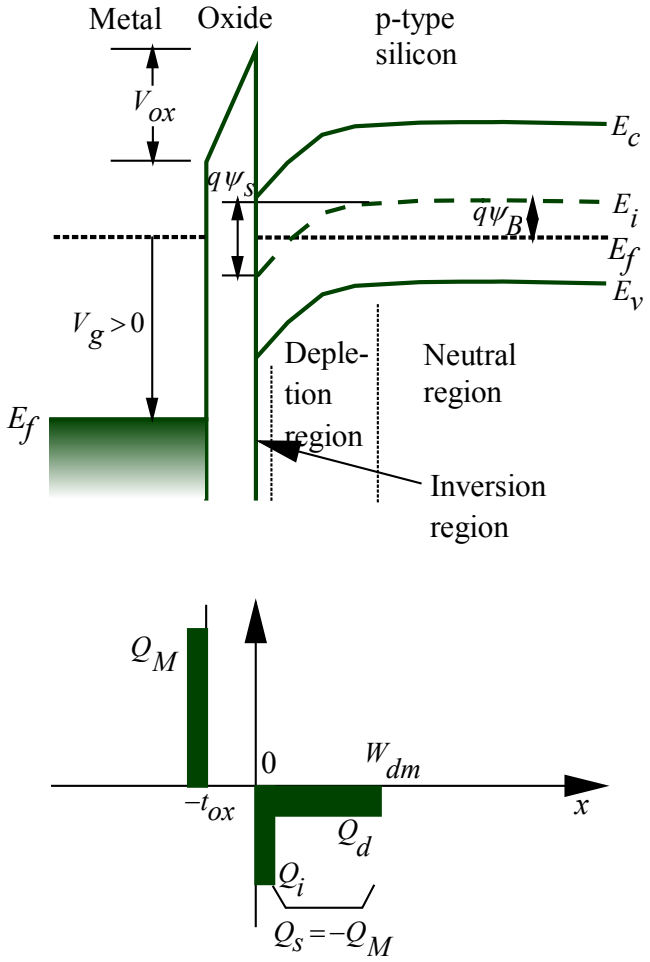
$$x_j = \frac{2E_j}{3q\mathrm{E}_s}$$

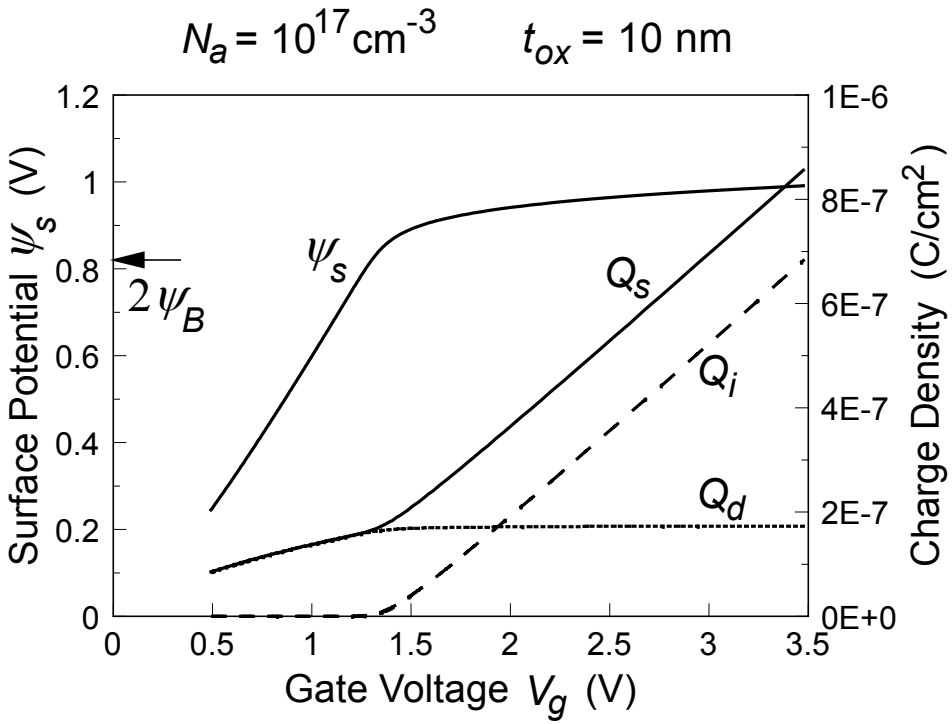
Self-Consistent QM Solution



- Electron ground state is at some finite energy above the bottom of the conduction band.
- Band bending must exceed $2\psi_B$ to invert surface.
- The centroid of inversion layer is farther away from the surface than in the classical case.

MOSFET Charge and Potential





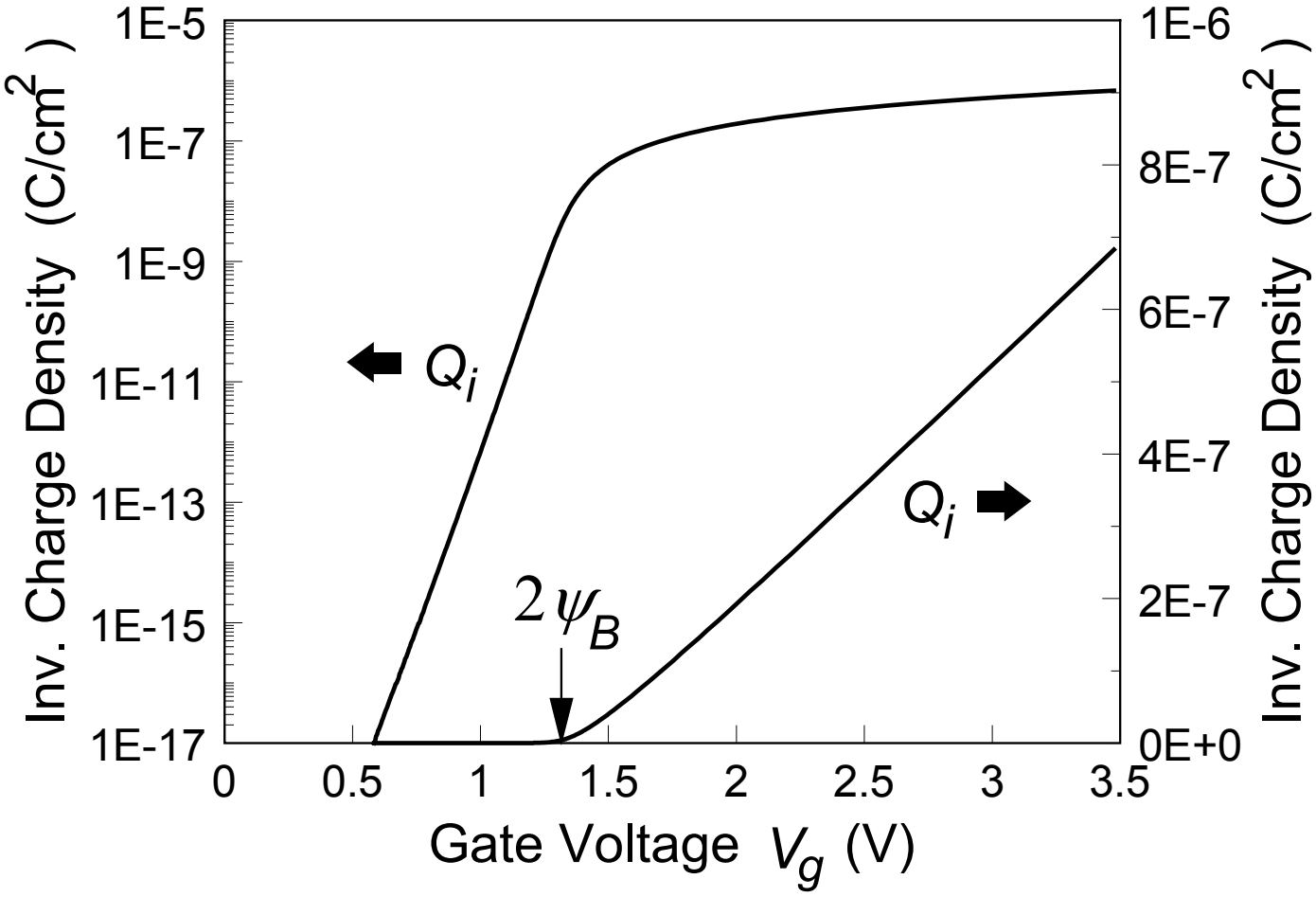
Gate voltage equation (V_{fb}=0):

$$V_g = V_{ox} + \psi_s = \frac{-Q_s}{C_{ox}} + \psi_s$$

Note: $C_{ox}=\varepsilon_{ox}/t_{ox}$ and $\varepsilon_{ox}\mathcal{E}_{ox}=\varepsilon_{si}\mathcal{E}_s$

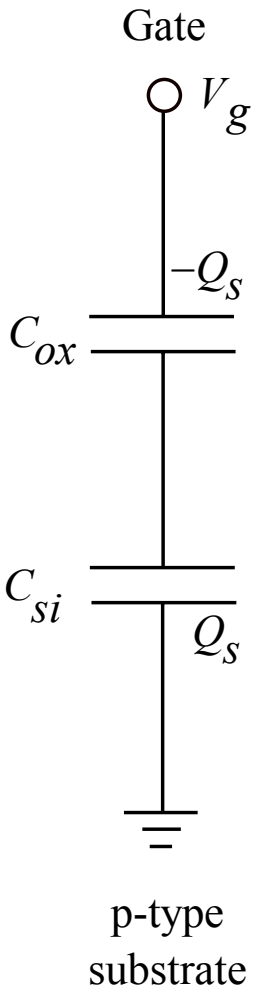
$$Q_s = -\varepsilon_{si}\mathcal{E}_s = \pm\sqrt{2\varepsilon_{si}kTN_a}\left[\left(e^{-q\psi_s/kT} + \frac{q\psi_s}{kT} - 1\right) + \frac{n_i^2}{N_a^2}\left(e^{q\psi_s/kT} - \frac{q\psi_s}{kT} - 1\right)\right]^{1/2}$$

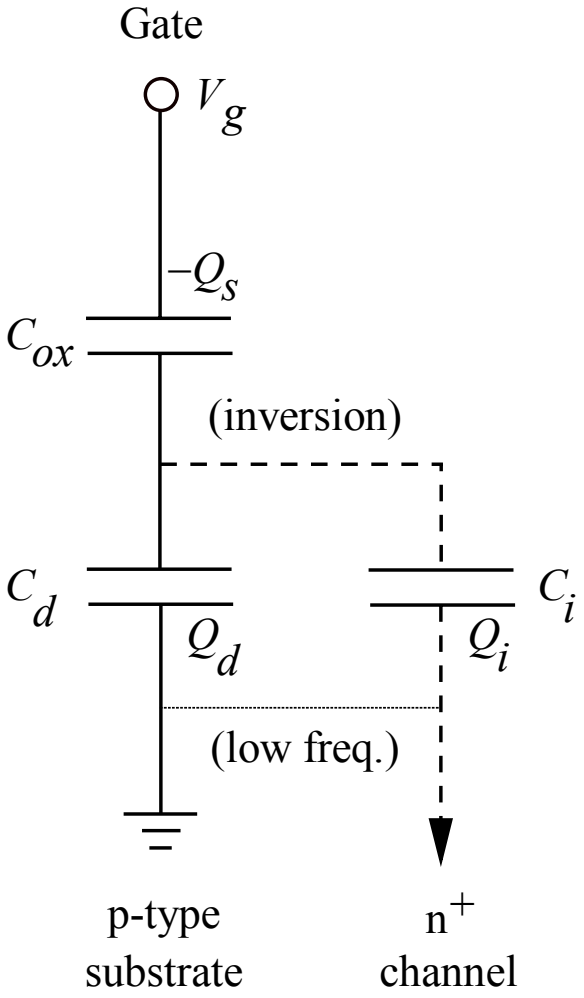
Inversion Charge in Log Scale



$N_a = 10^{17}\text{cm}^{-3}$ $t_{ox} = 10$ nm

MOS Capacitances





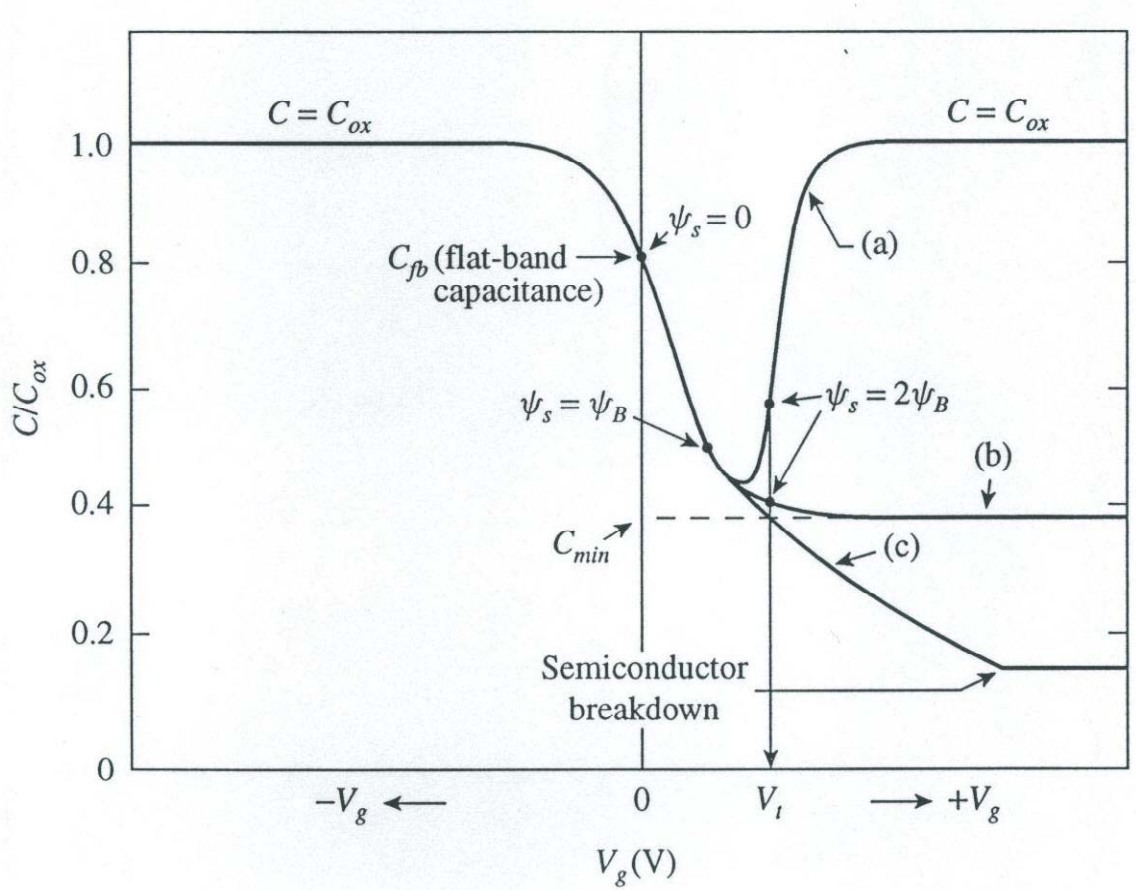
$$C = \frac{d(-Q_s)}{dV_g}$$

$$C_{si} = \frac{d(-Q_s)}{d\psi_s}$$

$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{d\psi_s}{d(-Q_s)}$$

$$= \frac{1}{C_{ox}} + \frac{1}{C_{si}}$$

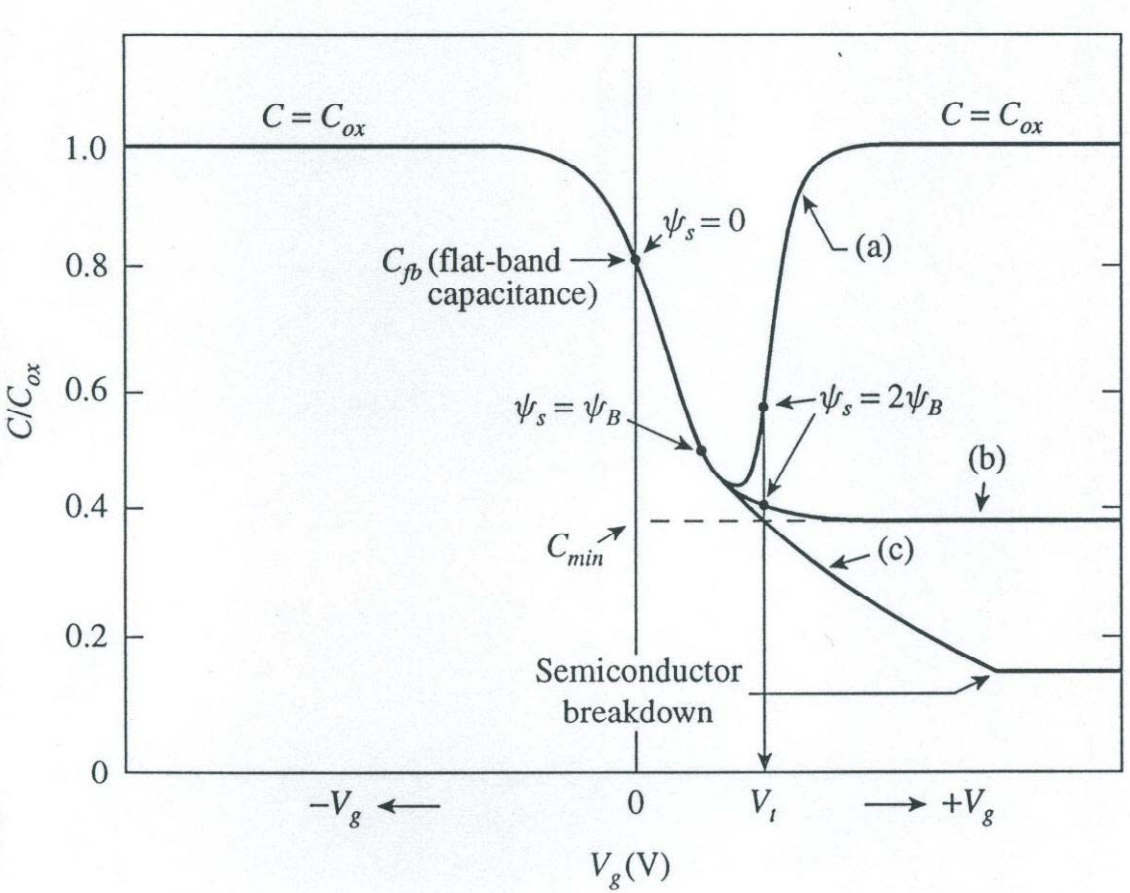
Capacitance-Voltage Characteristics



- In accumulation, $Q_s \propto \exp(-q\psi_s/2kT)$, so $C_{si} = -dQ_s/d\psi_s = (q/2kT)Q_s = (q/2kT)C_{ox}|V_g - \psi_s|$.

$$\frac{1}{C} = \frac{1}{C_{ox}}\left[1 + \frac{2kT/q}{|V_g - \psi_s|}\right]$$

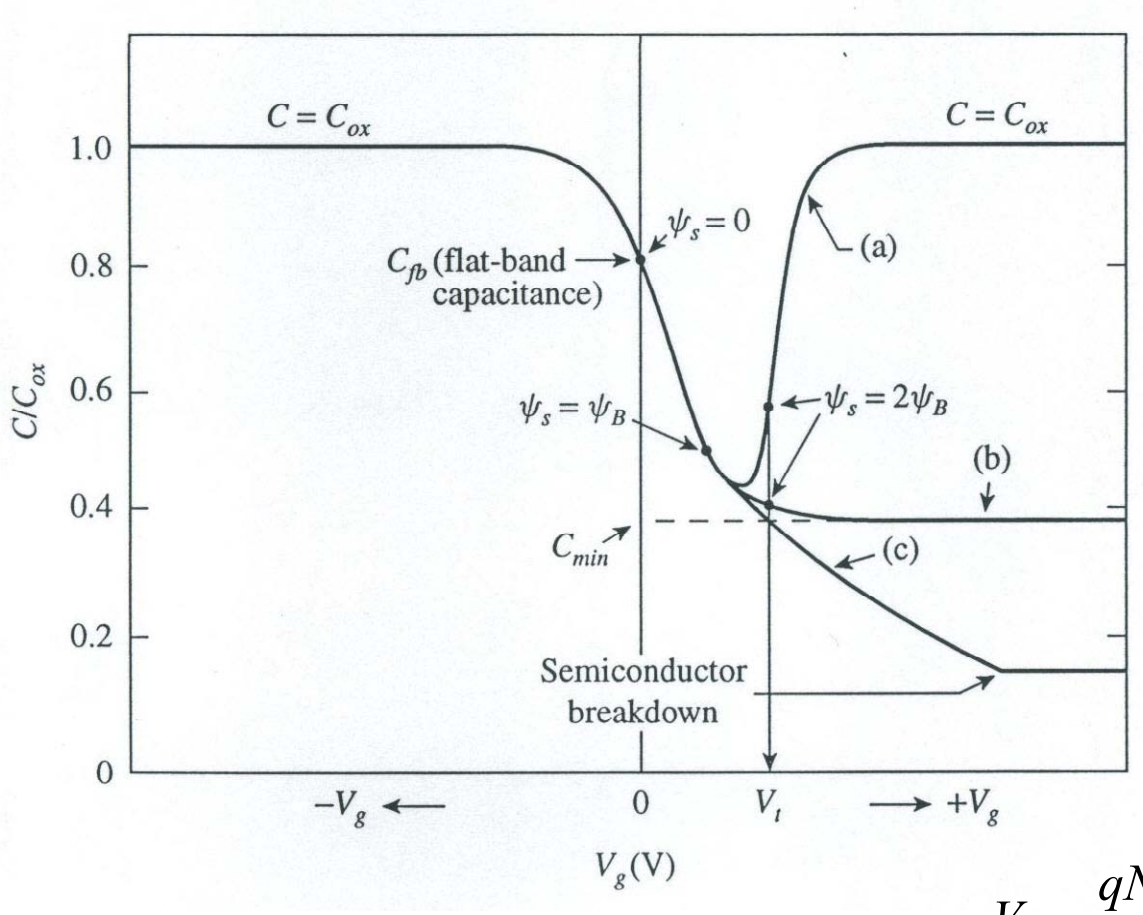
Capacitance-Voltage Characteristics



- At flatband voltage, $q\psi_s/kT<<1$, therefore, $Q_s=-(\varepsilon_{si}q^2N_a/kT)^{1/2}\psi_s$.

$$\frac{1}{C_{fb}} = \frac{1}{C_{ox}} + \sqrt{\frac{kT}{\varepsilon_{si}q^2N_a}} = \frac{1}{C_{ox}} + \frac{L_D}{\varepsilon_{si}}$$

Capacitance-Voltage Characteristics



- In depletion,

$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_d}$$

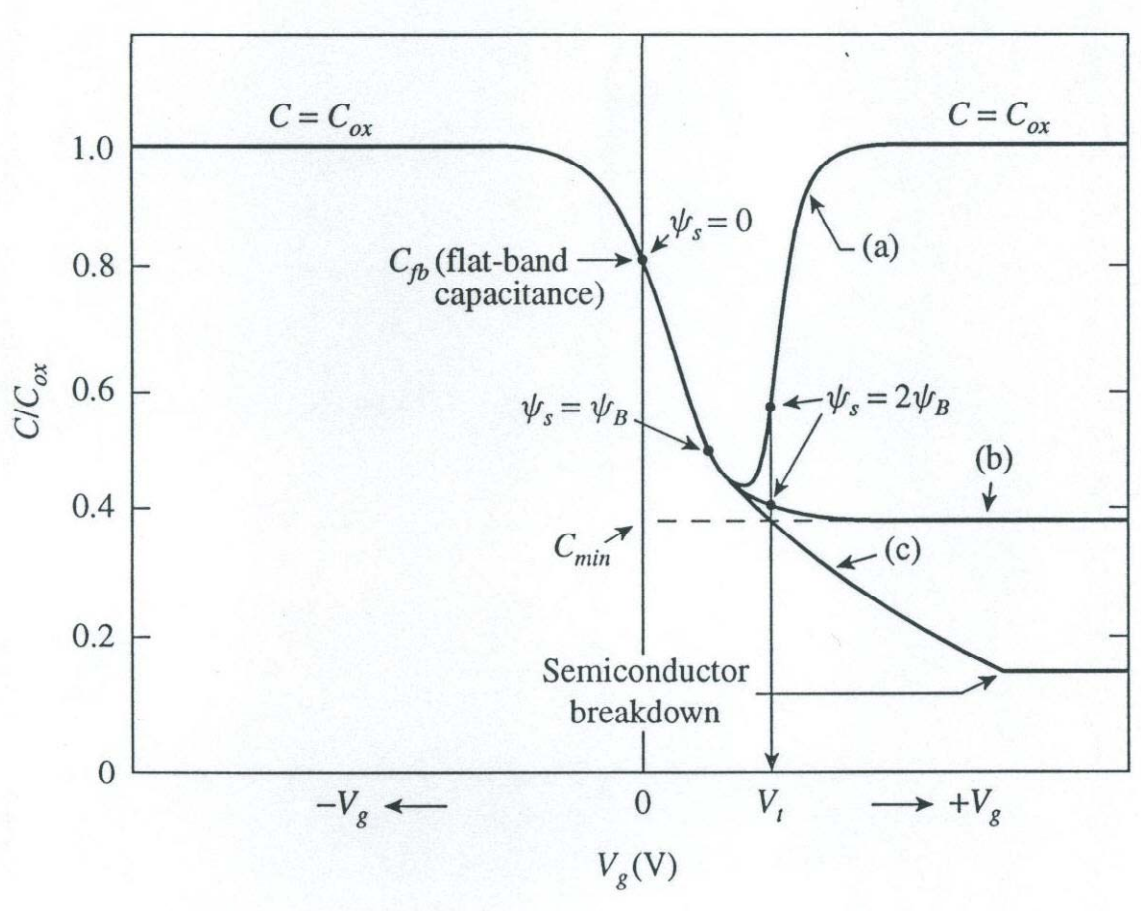
where

$$C_d = \frac{d(-Q_d)}{d\psi_s} = \sqrt{\frac{\varepsilon_{si} q N_a}{2\psi_s}} = \frac{\varepsilon_{si}}{W_d}$$

Note that

$$V_g = \frac{qN_a W_d}{C_{ox}} + \psi_s = \frac{\sqrt{2\varepsilon_{si} qN_a \psi_s}}{C_{ox}} + \psi_s$$

Capacitance-Voltage Characteristics



- Inversion, high freq.: Inversion charge cannot respond,

$$\frac{1}{C_{\min}} = \frac{1}{C_{ox}} + \sqrt{\frac{4kT\ln(N_a / n_i)}{\varepsilon_{si} q^2 N_a}}$$

- Inversion, low freq., or connected to a reservoir:

$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_d + C_i}$$

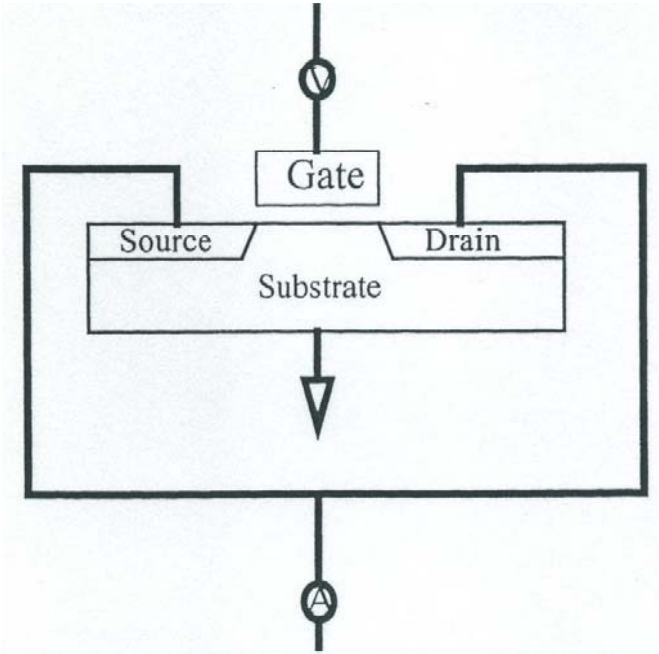
where

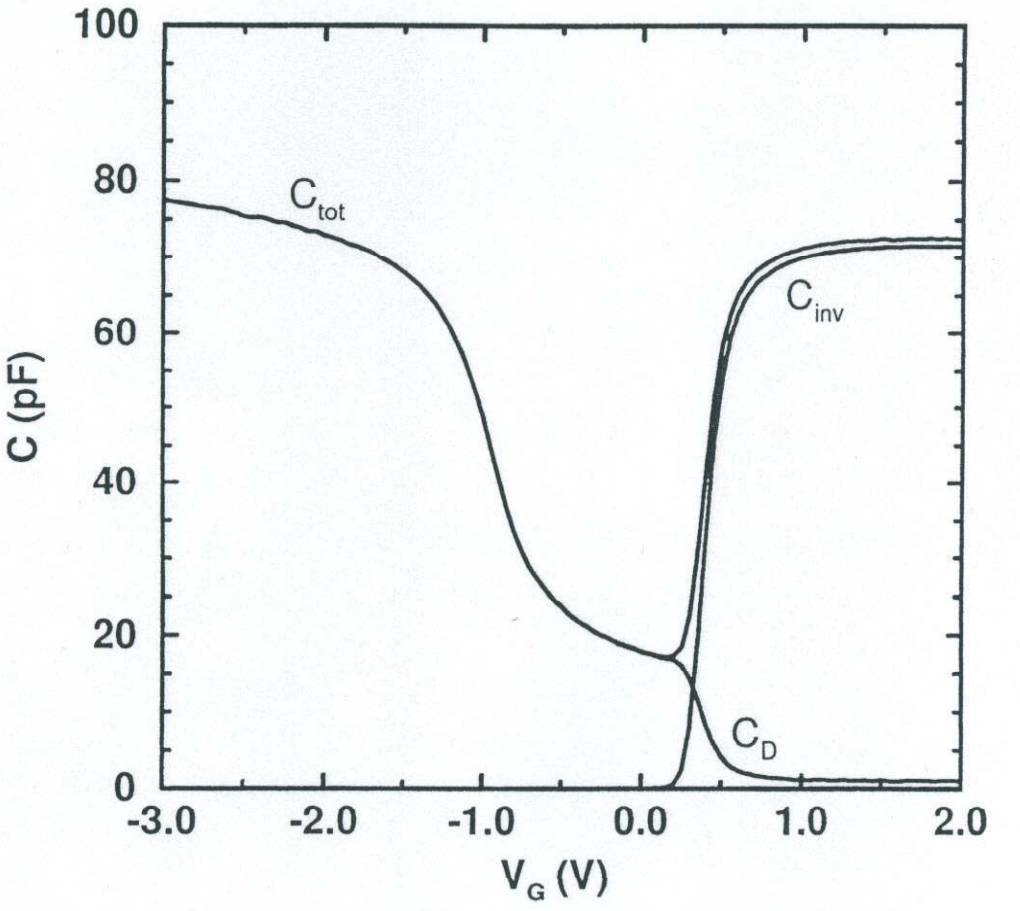
$$C_i = \frac{d(-Q_i)}{d\psi_s} = \frac{|Q_i|}{2kT/q}$$

is the inv. layer cap.

Like accumulation, $\frac{1}{C} = \frac{1}{C_{ox}}\left[1 + \frac{2kT/q}{|V_g - \psi_s|}\right]$

Split C-V Measurement

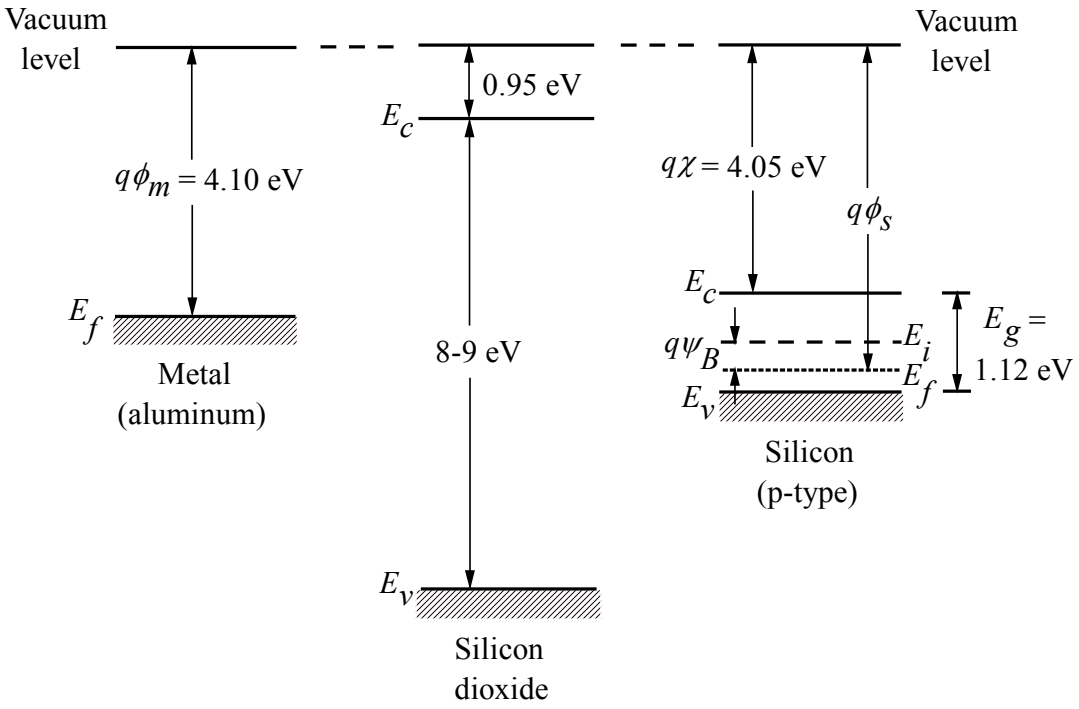




Effect of Gate Work Function

$$V_t = V_{fb} + 2\psi_B + V_{ox} = V_{fb} + 2\psi_B + \frac{Q_d}{C_{ox}}$$

$$V_{fb} = (\phi_m - \phi_s) - \frac{Q_{ox}}{C_{ox}}$$



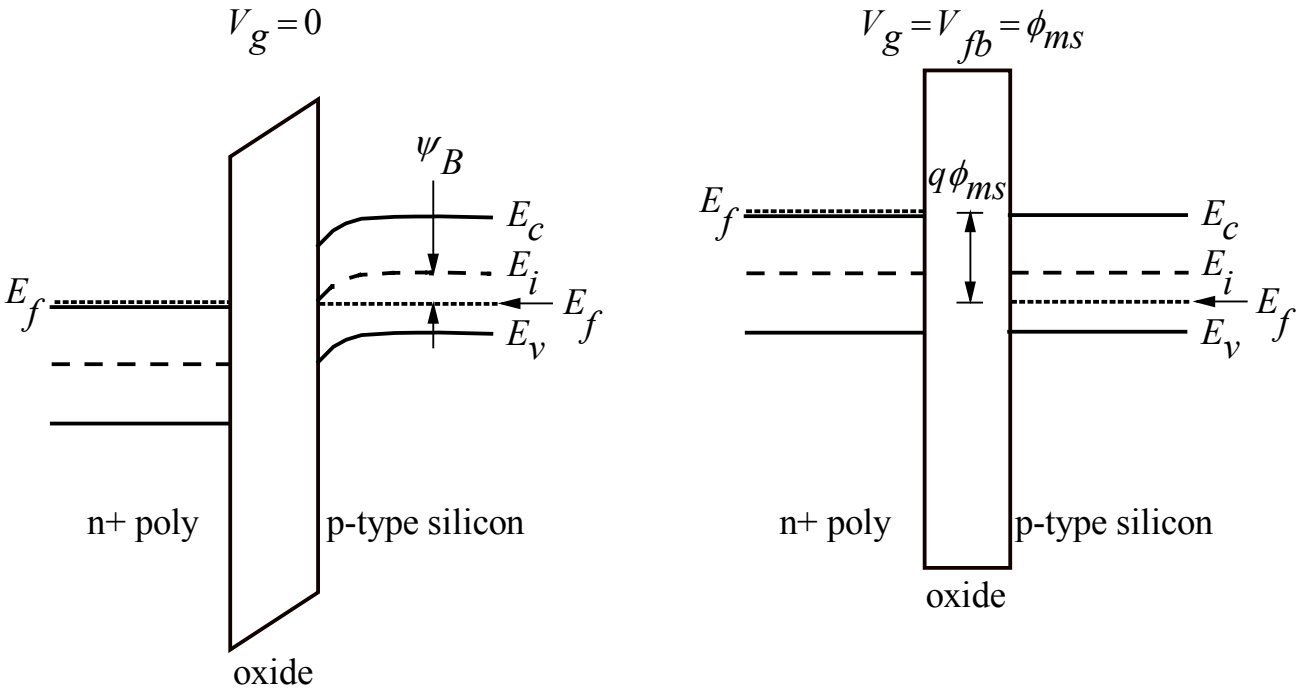
$$\phi_s = \chi + \frac{E_g}{2q} + \psi_B$$

$$\begin{cases} \phi_m = \chi \ \text{(n}^+\text{ poly)} \\ \phi_m = \chi + \dfrac{E_g}{2q} \ \text{(midgap)} \\ \phi_m = \chi + \dfrac{E_g}{q} \ \text{(p}^+\text{ poly)} \end{cases}$$

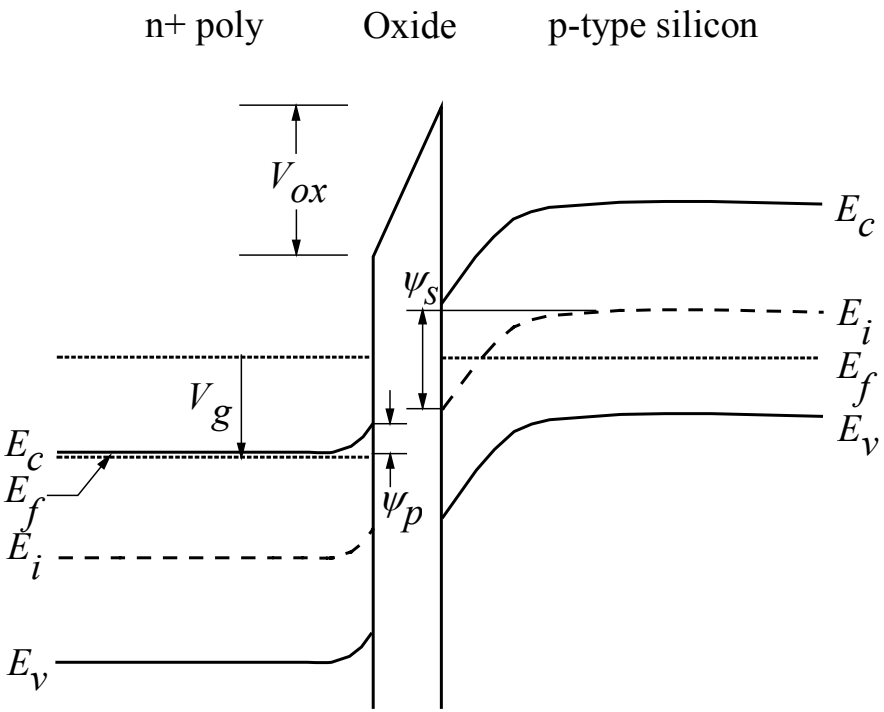
Effect of Gate Work Function

Example: n+ polysilicon gate on p-type silicon

$$\phi_{ms} = -\frac{E_g}{2q} - \psi_B = -0.56 - \frac{kT}{q}\ln\left(\frac{N_a}{n_i}\right)$$



Poly-Si Gate Depletion Effect

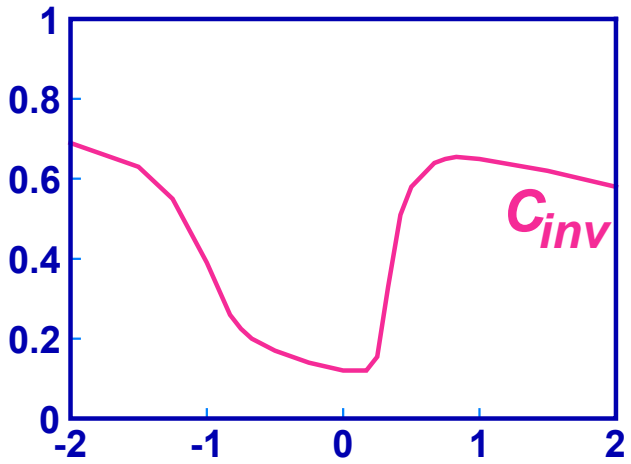


Gate eq. becomes:

$$V_g = V_{fb} + \psi_s + \psi_p - \frac{Q_s}{C_{ox}}$$

and,

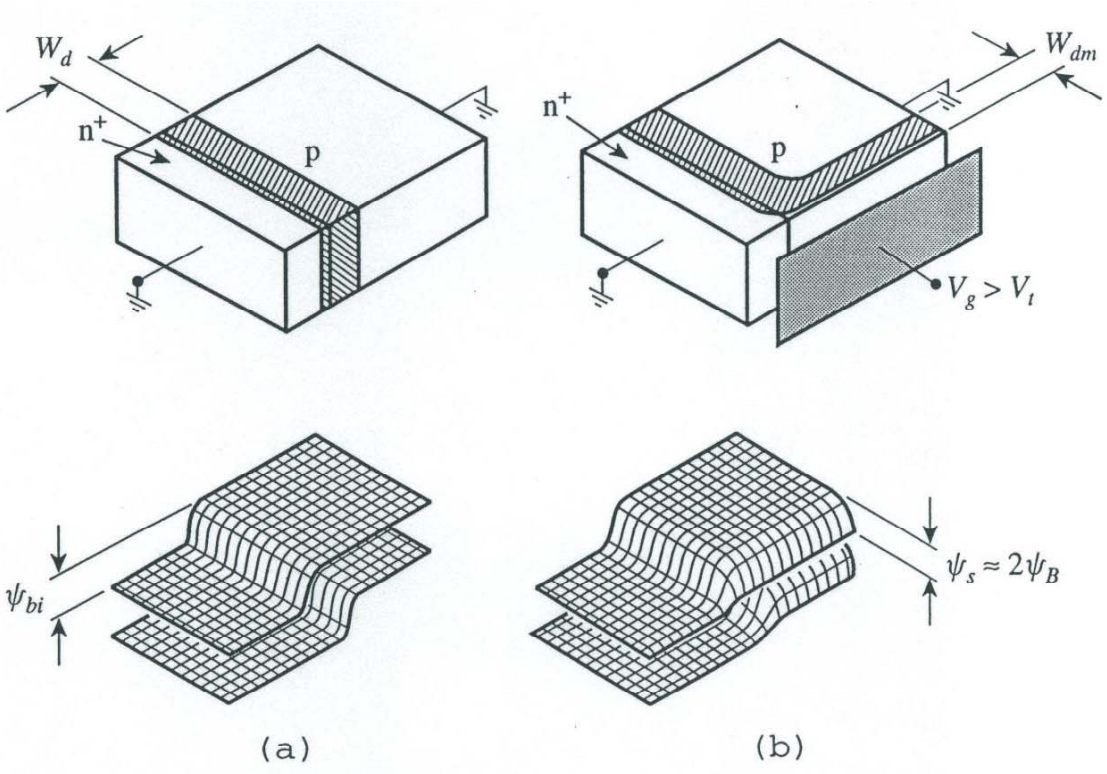
$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_{si}} + \frac{1}{C_p}$$



Typically, t_{inv} is 0.8-1.0 nm thicker than t_{ox}.

Gated-Diode: MOS + p-n Junction

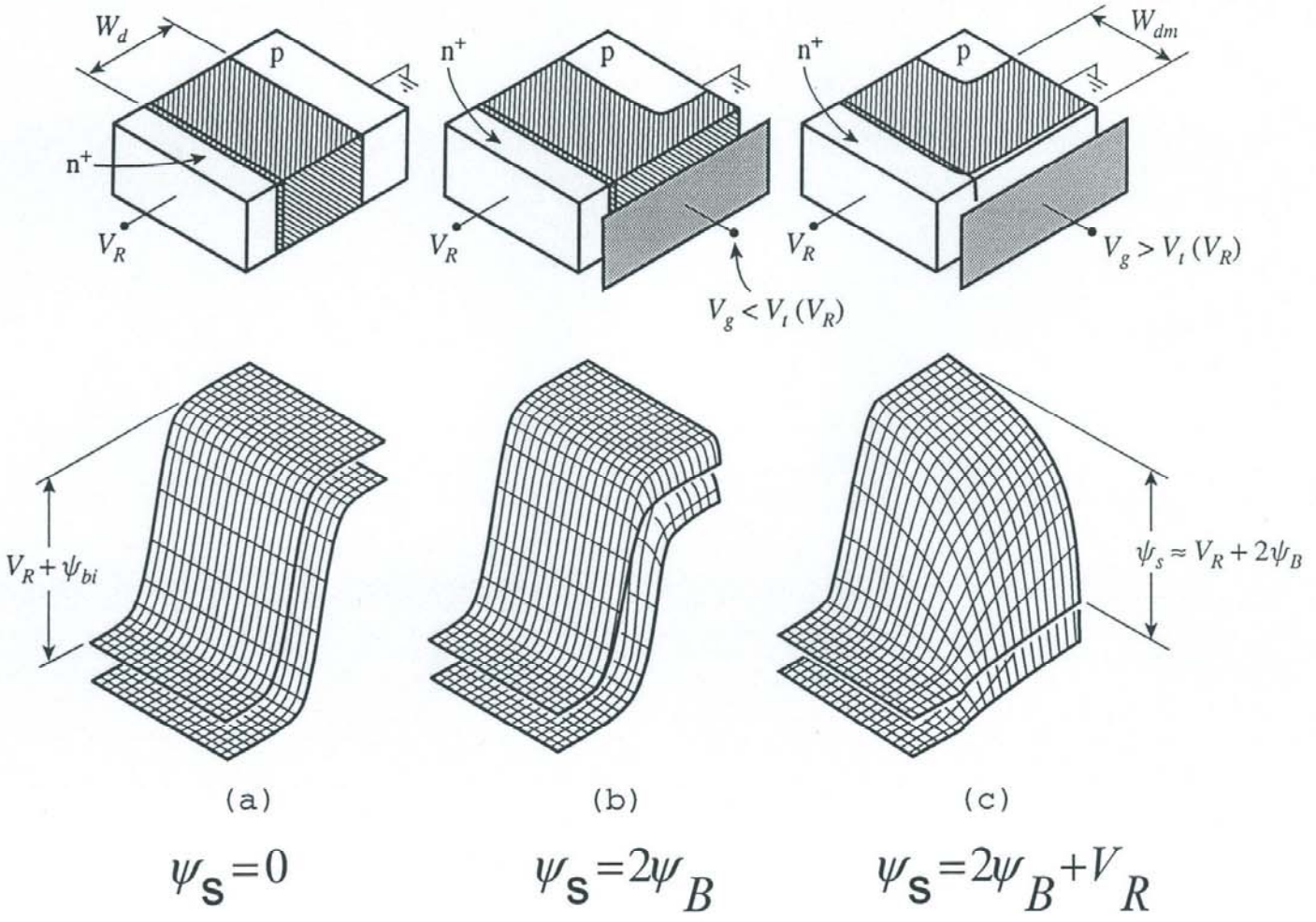
Zero-bias on the p-n junction (equilibrium):



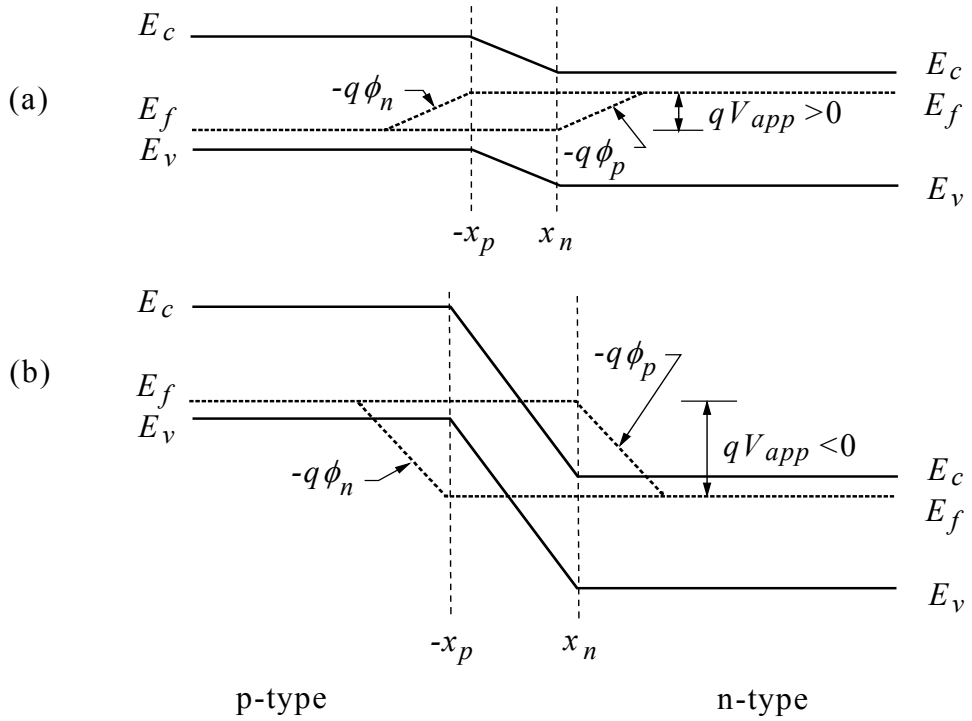
The electron quasi-Fermi level in the MOS is the same as the Fermi level of the p-type Si.

$\Rightarrow$ Inversion occurs when $\psi_s = 2\psi_B$.

Gated-Diode: Reverse Biased (Nonequilibrium)



Spatial Variations of ϕ_n and ϕ_p



$V_{app} = \phi_p - \phi_n$ at junction boundaries.

$pn = n_i^2 \exp[q(\phi_p - \phi_n)/kT]$

$V_{app} > 0$ for forward bias,
$V_{app} < 0$ for reverse bias.

(1) $n_p(x = -x_p) \approx n_{p0}(x = -x_p) \exp(qV_{app}/kT)$

(2) $p_n(x = x_n) \approx p_{n0}(x = x_n) \exp(qV_{app}/kT)$

(1) and (2) are the most important boundary conditions governing a p-n diode.

MOS under Nonequilibrium

For a p-n junction reverse-biased at a voltage V_R, the electron concentration on the p-side of the junction is

$$n = \frac{n_i^2}{N_a} e^{-qV_R/kT}$$

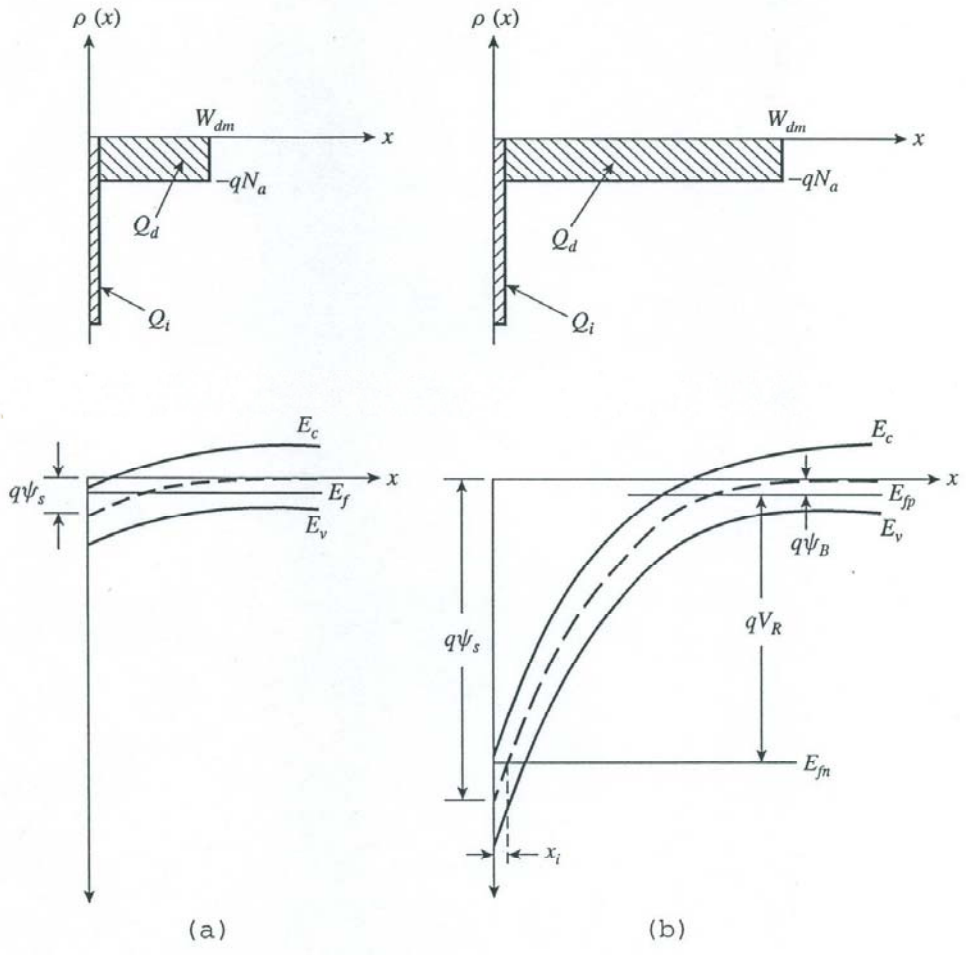
If a gate voltage is applied to bend the p-type bands by ψ_s, the electron concentration at the surface is

$$n = \frac{n_i^2}{N_a} e^{q\psi_s/kT} e^{-qV_R/kT}$$

For surface inversion to occur, i.e., $n = N_a$, Need

$$\psi_s(inv) = V_R + 2\psi_B$$

MOS under Nonequilibrium



Maximum depletion width at inversion is $$W_{dm} = \sqrt{\frac{2\varepsilon_{si}(V_R + 2\psi_B)}{qN_a}}$$