Device Modeling

Outline

1. **P-N Junction**
2. Bipolar Junction Transistor
3. MOS Field-Effect Transistor
    - I/V characteristic
    - Second-order effect
    - Small-signal model
    - Scaling & Short-channel effects
    - Simulation models

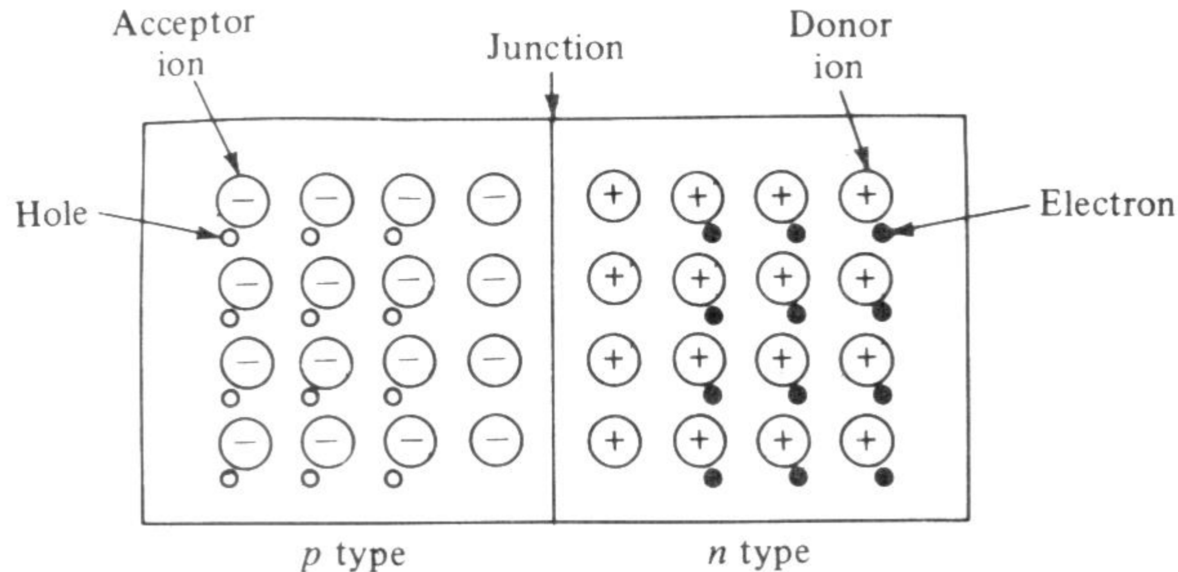
Symbol Convention

V_{DS}, I_D → bias and DC quantities

v_{ds}, i_d → small signal quantities

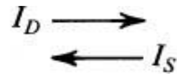
I_d → sum of bias and signal

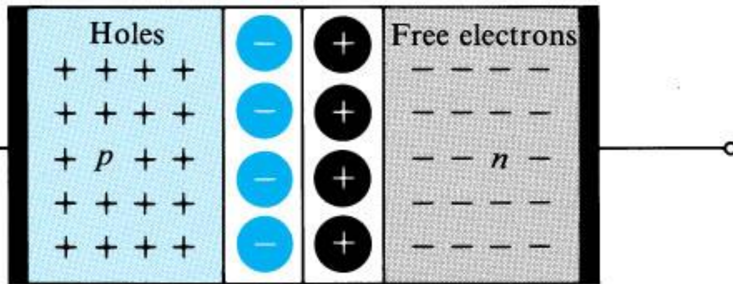
Junction Formation

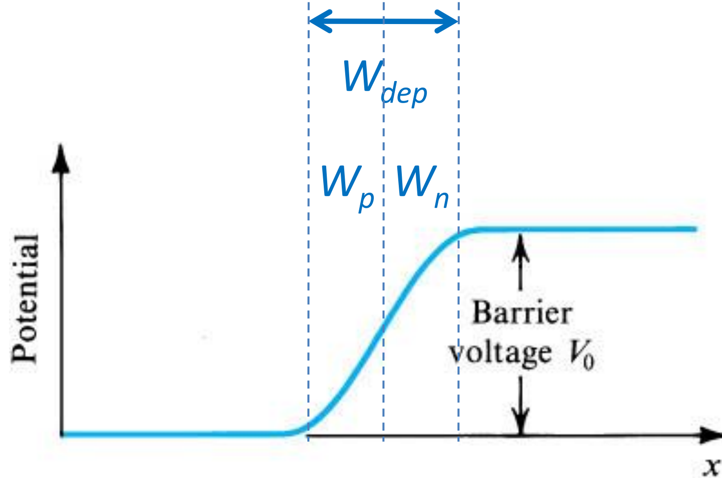


- P-N junction is the basis of semiconductor operation.
- There are a variety of methods for junction formation, like alloying, epitaxy, diffusion, and implantation.

P-N Junction - Open







- $I_D = I_S$
  - I_D = Majority diffusion due to concentration gradient
  - I_S = Minority drift driven by junction electric field
- Junction Built-in Voltage V_0 : ≈0.6~0.8V

$$V_0 = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right)$$

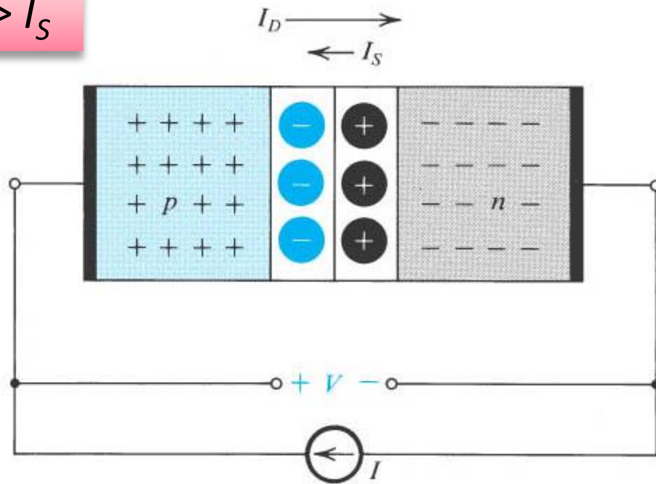
(see p.2-10)

- Width of Depletion Region W_{dep} ≈ 0.1~1um

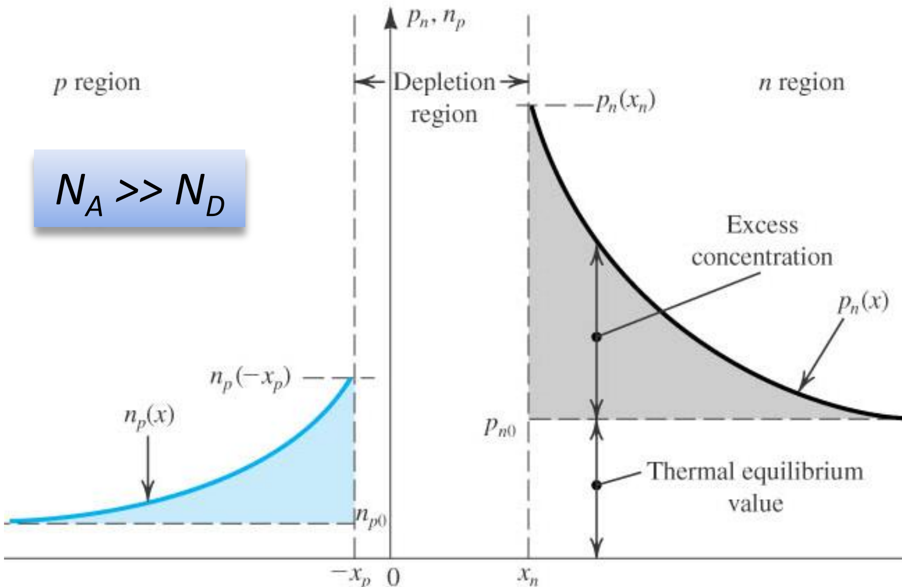
$$W_n/W_p = N_A/N_D$$

P-N Junction - Forward

$I_D >> I_S$



$N_A >> N_D$



- $I_D >> I_S$ ($I = I_D - I_S$)

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

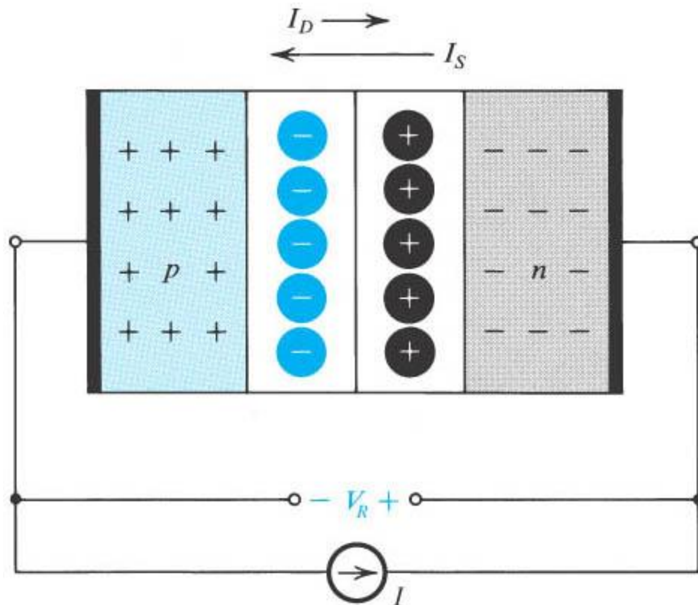
$$I_0 = Aqn_i^2\left(\frac{D_p}{L_pN_D} + \frac{D_n}{L_nN_A}\right)$$

- Diffusion length $L_p = \sqrt{D_p\tau_p}$
- Diffusivity D_p = *hole diffusivity in n-type*
- τ_p = excess minority carrier lifetime

P-N Junction - Reverse

$I_D << I_S$

$V_R < V_{ZK}$



- $I_D << I_S$ ($I = I_S - I_D$)
  - Reverse bias V_R↑, Majority leave ↑, Depletion region width ↑
  - $I \approx I_S \approx 10^{-9}A$ (Thermal leak)
- Junction charge q_j

$$q_J = qAW_{dep} \frac{N_A N_D}{N_A + N_D}$$

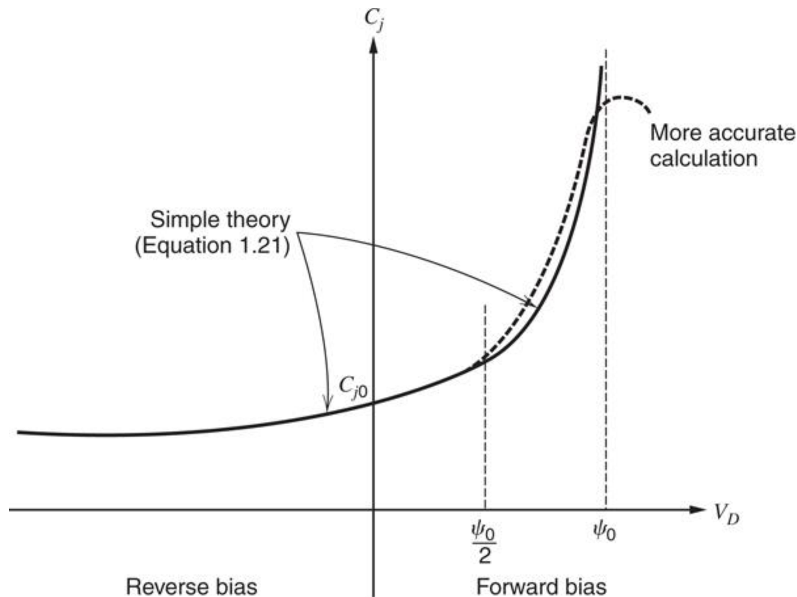
- Depletion width with V_R

$$W_{dep} = x_n + x_p = \sqrt{\frac{2\varepsilon_s}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)(V_0 + V_R)}$$

- Junction capacitance

$$C_J = \left.\frac{dq_j}{dV_R}\right|_{V_R = V_Q} = \frac{\varepsilon_S A}{W_{dep}}$$

Small Signal Junction Capacitance



- small signal junction capacitance C_j

$$C_j = \frac{dQ_j}{dV_R} = A \cdot \left[\frac{q\varepsilon N_A N_D}{2(N_A + N_D)}\right]^{1/2} \cdot \frac{1}{\sqrt{\varphi_0 + V_R}}$$

in general, C_j can be expreased as

$$C_j = \frac{C_{j0}}{\left(1 - \frac{V_D}{\varphi_0}\right)^m}$$

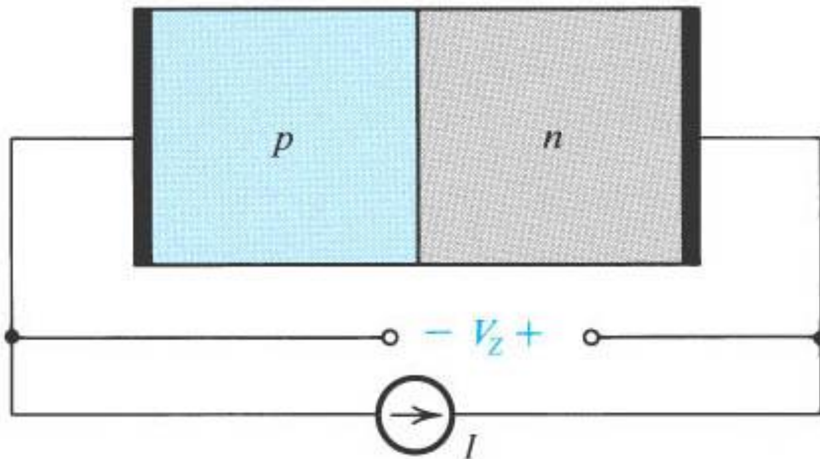
$m = \frac{1}{2}$ for abrupt junction, $m = \frac{1}{3}$ for graded junction

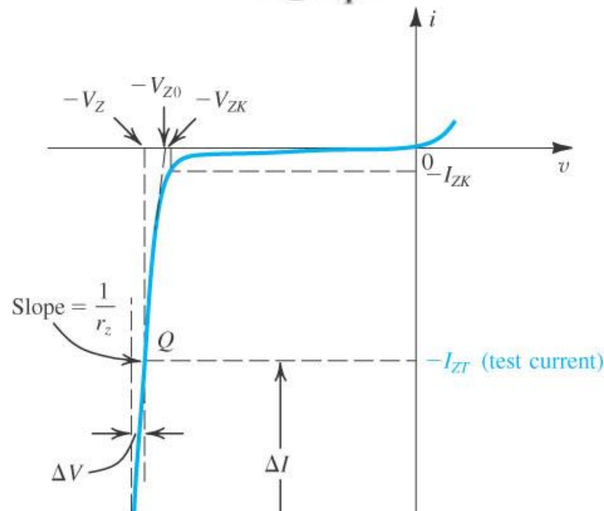
$$C_{j0} = A\left[\frac{q\varepsilon N_A N_D}{2(N_A + N_D)\varphi_0}\right]^{1/2}$$

V_D : positive for F.B., V_D : negative for R.B.

P-N Junction - Breakdown

$V_R > V_{ZK}$





- Zener breakdown
  - Reverse bias V_R↑, E of depletion region ↑, Covalent bond break
  - Hole-electron pair generation in depletion region, electrons swept to n-type (holes swept to p-type)
- Avalanche breakdown
  - Reverse bias V_R↑, minority swept by electric filed E , kinetic energy break covalent bond
  - Ionizing collision , Hole-electron pair generation in depletion region
- Punch through
  - Two neighboring junction depletion regions meet.

Junction Built-in Voltage

*Open-Circuit **p-n** Junction :*

$$J_{p-Total} = J_{p-diff} + J_{p-drift} = -qD_p \frac{dp}{dx} + qp\mu_p E = 0$$

$$D_p \frac{dp}{dx} = p\mu_p E, \quad E = -\frac{dv}{dx}, \quad dv = -Edx = -\frac{D_p}{\mu_p}\frac{dp}{p}$$

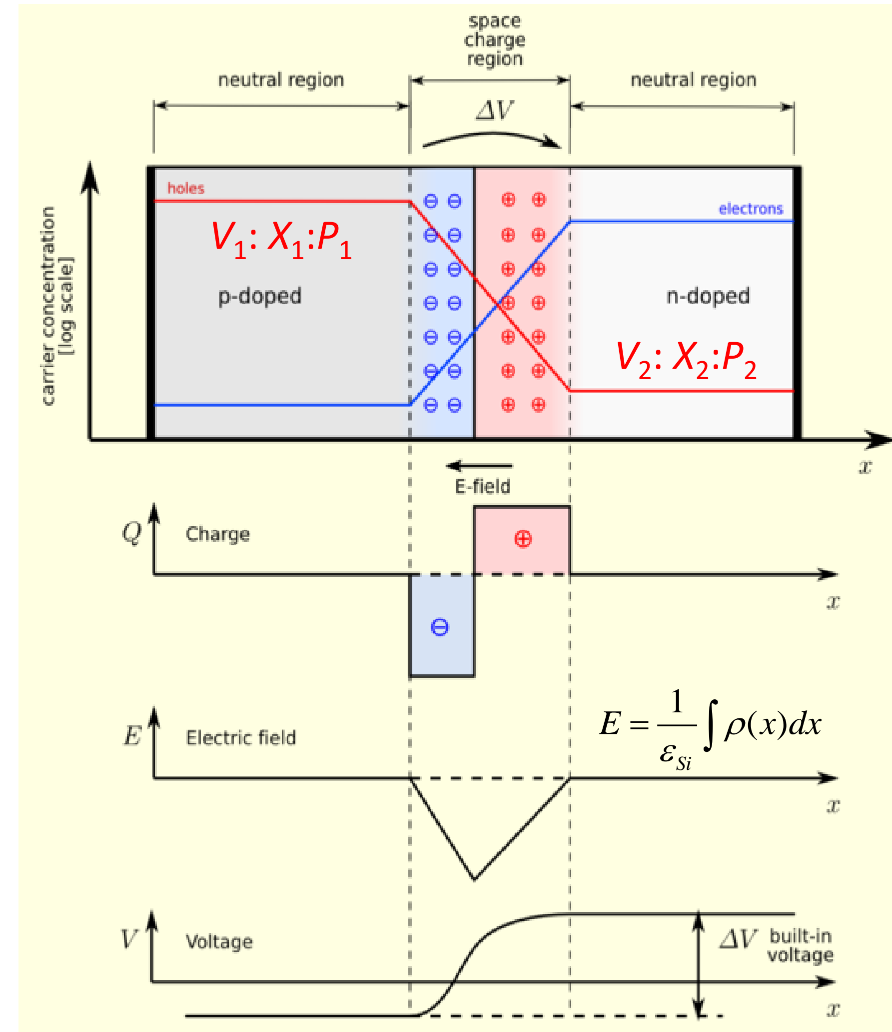
$$V_2 - V_1 \equiv \int_{V_1}^{V_2} dv = -\int_{X_1}^{X_2} Edx = -\int_{P_1}^{P_2} \frac{D_p}{\mu_p}\frac{dp}{p}$$

$$\frac{D_p}{\mu_p} = V_T, \quad -\int_{P_1}^{P_2} \frac{dp}{p} = \ln\frac{P_1}{P_2}$$

$$P_1 = N_A, \; P_2 = \frac{n_i^2}{N_D}, \; \ln\frac{P_1}{P_2} = \ln\frac{N_A N_D}{n_i^2}$$

Junction Built-in Voltage :

$$V_0 = V_2 - V_1 = V_T \ln\frac{P_1}{P_2} = V_T \ln\frac{N_A N_D}{n_i^2}$$

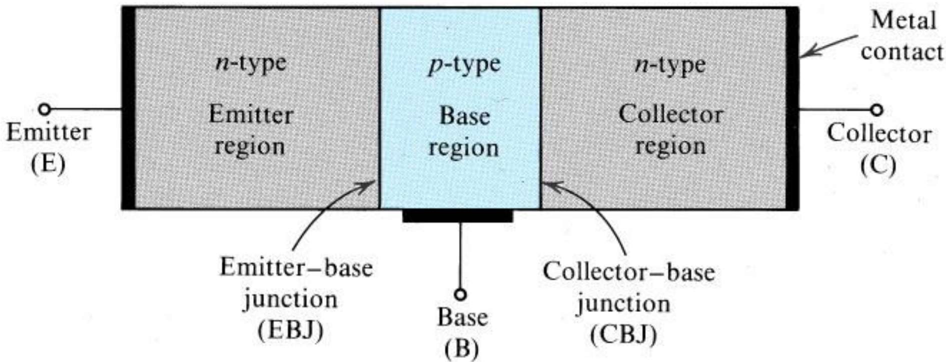


Outline

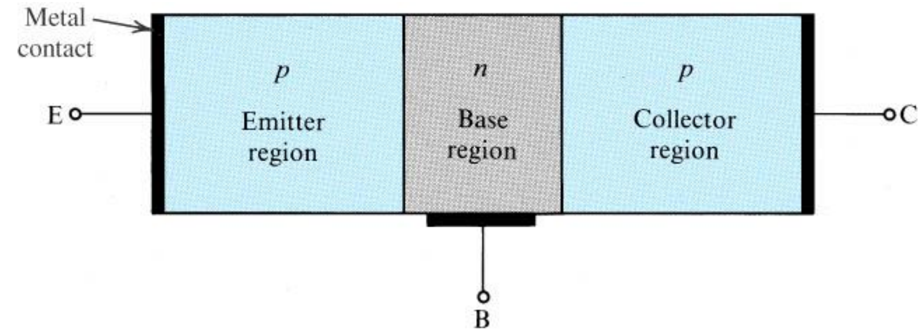
1. P-N Junction
2. **Bipolar Junction Transistor**
3. MOS Field-Effect Transistor
    - I/V characteristic
    - Second-order effect
    - Small-signal model
    - Scaling & Short-channel effects
    - Simulation models

BJT operation mode

npn bipolar

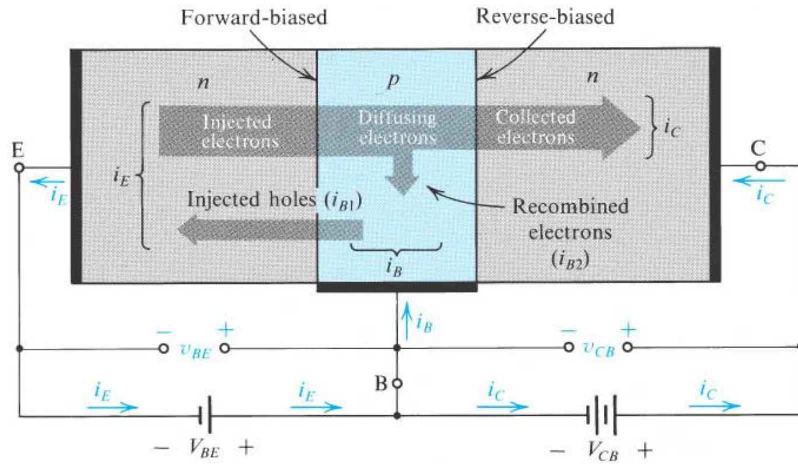


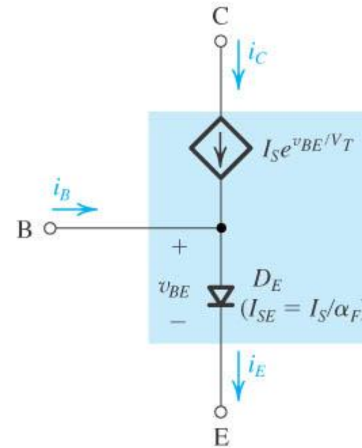
pnp bipolar

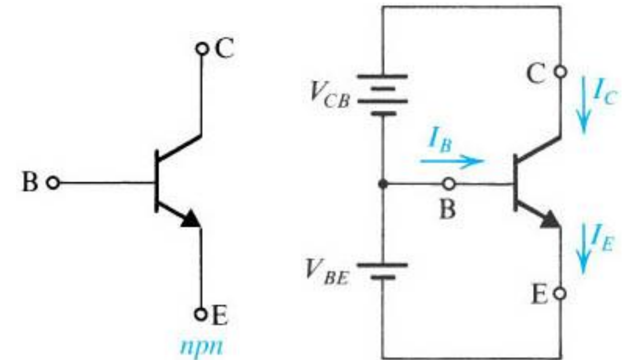


BJT Modes of Operation			
Mode	EBJ	CBJ	Operation
Cutoff	Reverse	Reverse	Switch OFF
Active	Forward	Reverse	Amplifier
Reverse active	Reverse	Forward	
Saturation	Forward	Forward	Switch ON

NPN vs PNP



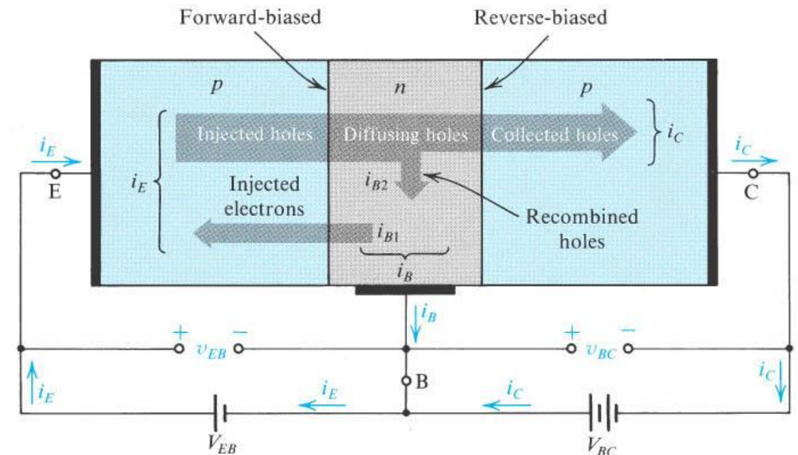


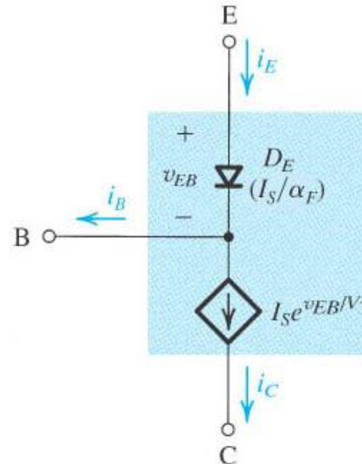


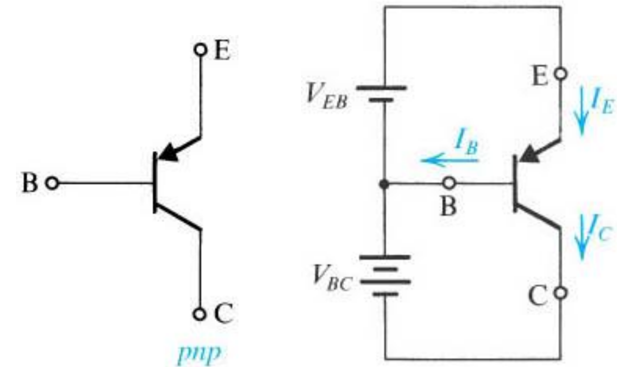
Carrier flow

Large signal model

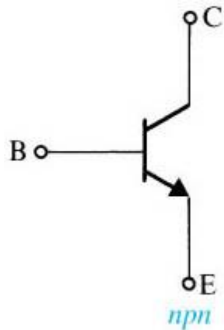
Circuit symbol and polarity

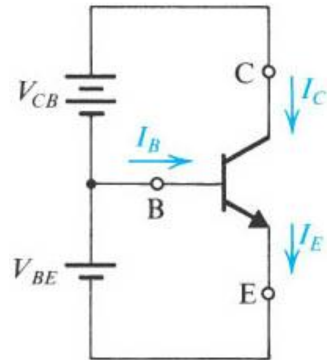






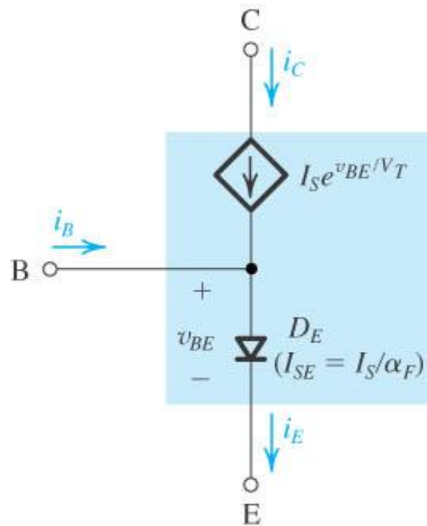
Current Relationship

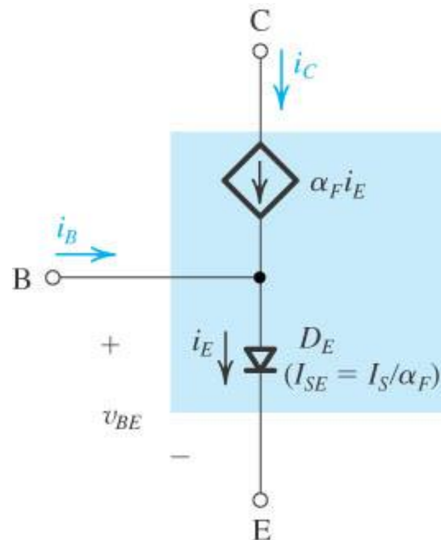




VCCS

CCCS





$$i_C = \beta\, i_B$$

$$i_E = i_C + i_B = (1+1/\beta)\, i_C$$

$$i_C = \alpha\, i_E\,,\ \alpha = \beta/(1+\beta) \approx 1$$

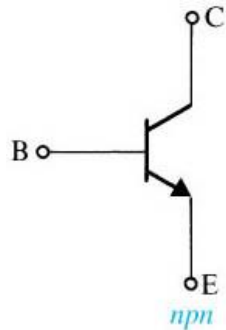
Common-Emitter Current Gain β

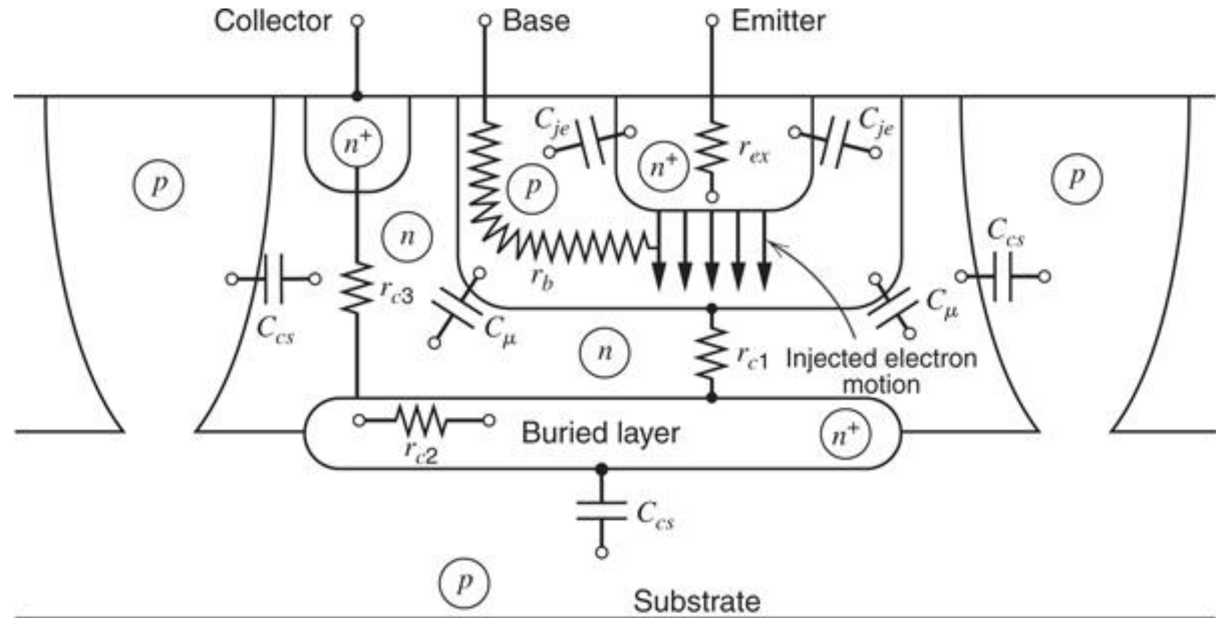
$$\beta = \left(\frac{D_p}{D_n} \frac{N_A}{N_D} \frac{W}{L_p} + \frac{1}{2} \frac{W^2}{D_n \tau_b} \right)^{-1}$$

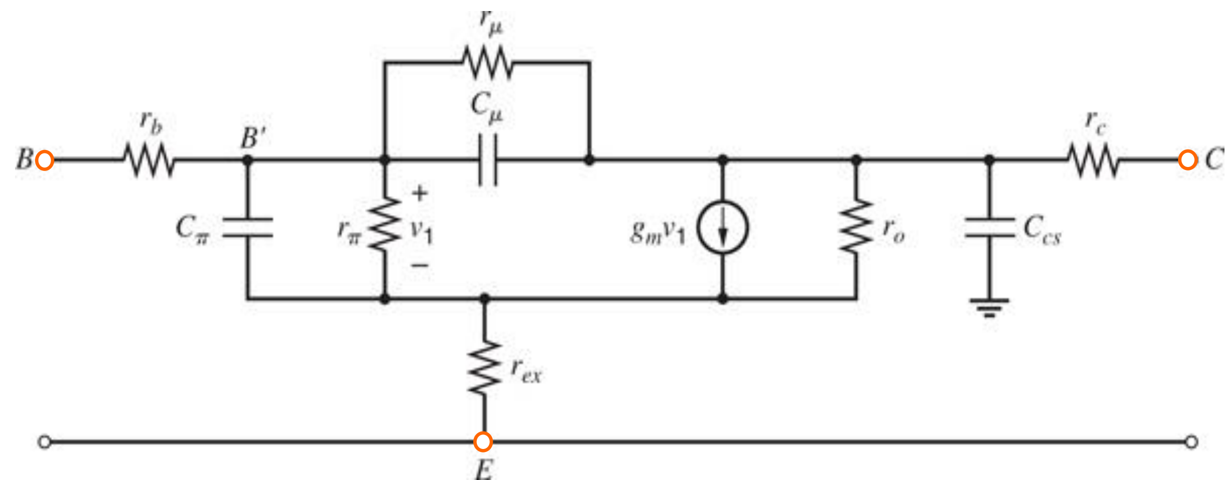
1.Width of base region $W\downarrow, \beta\uparrow$
   Thinner B can get a higher gain.

2. Relative doping $N_A(B)/N_D(E)\downarrow, \beta\uparrow$
   lighter B and heavier E get a higher gain

Bipolar Device Small Signal Model







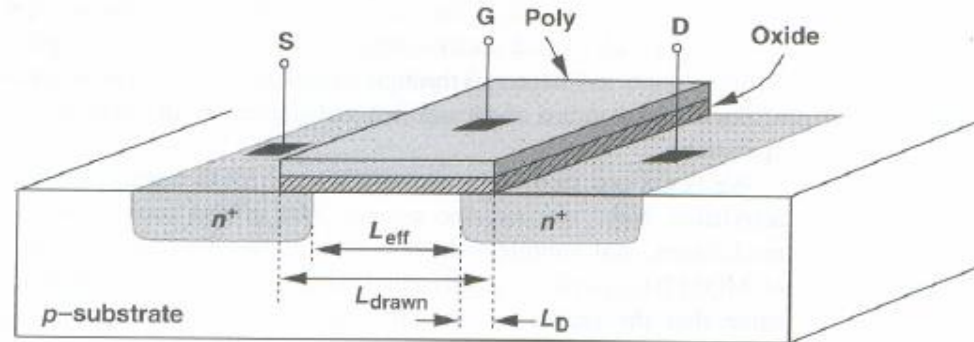
Bipolar Transistor Parameters

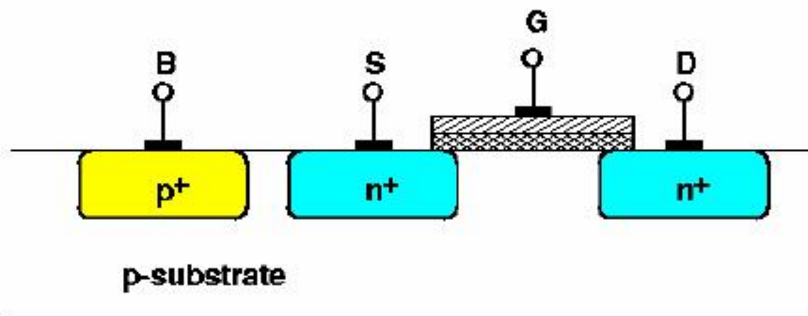
Quantity	Formula
Large-Signal Forward-Active Operation	
Collector current	$I_c = I_S \exp \frac{V_{be}}{V_T}$
Small-Signal Forward-Active Operation	
Transconductance	$g_m = \frac{qI_C}{kT} = \frac{I_C}{V_T}$
Transconductance-to-current ratio	$\frac{g_m}{I_C} = \frac{1}{V_T}$
Input resistance	$r_\pi = \frac{\beta_0}{g_m}$
Output resistance	$r_o = \frac{V_A}{I_C} = \frac{1}{\eta g_m}$
Collector-base resistance	$r_\mu = \beta_0 r_o$ to $5\beta_0 r_o$
Base-charging capacitance	$C_b = \tau_F g_m$
Base-emitter capacitance	$C_\pi = C_b + C_{je}$
Emitter-base junction depletion capacitance	$C_{je} \simeq 2C_{je0}$
Collector-base junction capacitance	$C_\mu = \frac{C_{\mu 0}}{\left(1 - \frac{V_{BC}}{\psi_{0c}}\right)^{n_c}}$
Collector-substrate junction capacitance	$C_{cs} = \frac{C_{cs0}}{\left(1 - \frac{V_{SC}}{\psi_{0s}}\right)^{n_s}}$
Transition frequency	$f_T = \frac{1}{2\pi} \frac{g_m}{C_\pi + C_\mu}$
Effective transit time	$\tau_T = \frac{1}{2\pi f_T} = \tau_F + \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m}$
Maximum gain	$g_m r_o = \frac{V_A}{V_T} = \frac{1}{\eta}$

Outline

1. P-N Junction
2. Bipolar Junction Transistor
3. **MOS Field-Effect Transistor**
   - I/V characteristic
   - Second-order effect
   - Small-signal model
   - Scaling & Short-channel effects
   - Simulation models

NMOS Structure





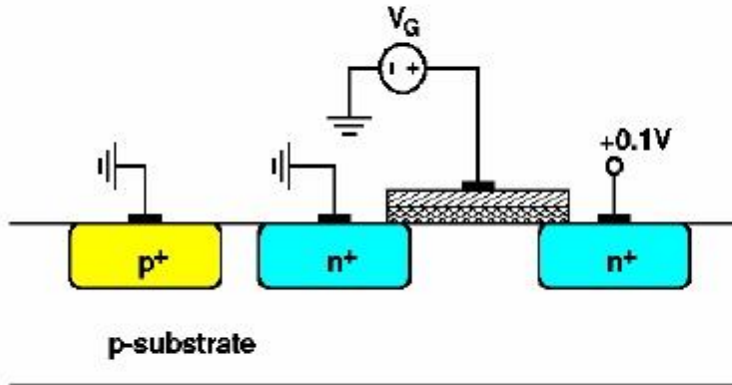
$$L_{eff} = L_{drawn} - 2L_D$$

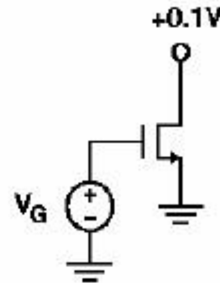
For $L_{eff} \approx 0.15\ \mu m \quad t_{ox} \approx 50\,\mathring{A}$

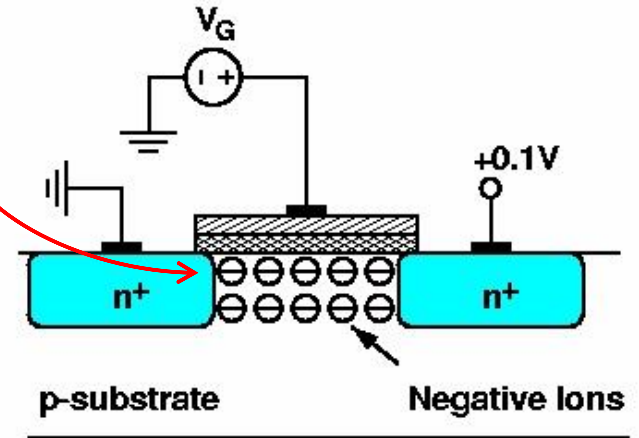
- L_{eff} is the effective length, L_{drawn} is the total length. L_D is the amount of side diffusion. t_{ox} is the oxide thickness.
- In a general case, the substrate is connected to the most negative supply in the system.

Threshold Voltage

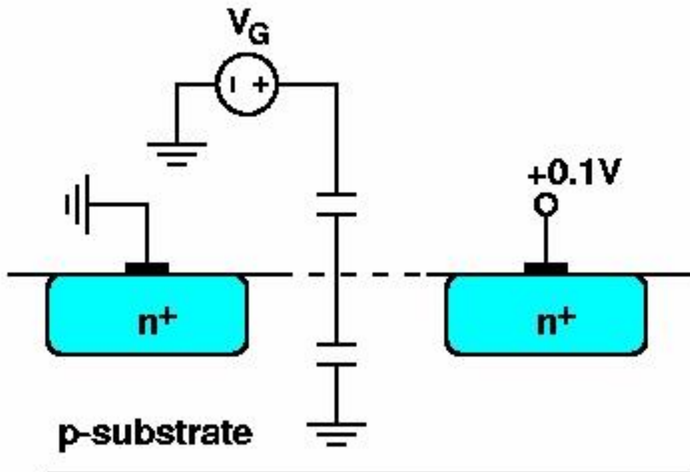
Formation of depletion region



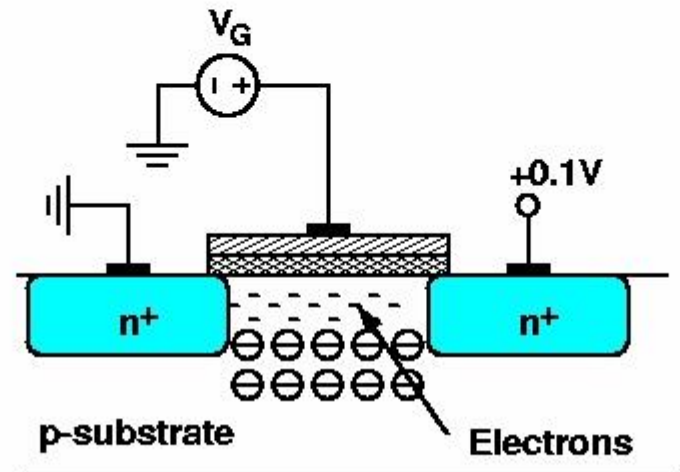




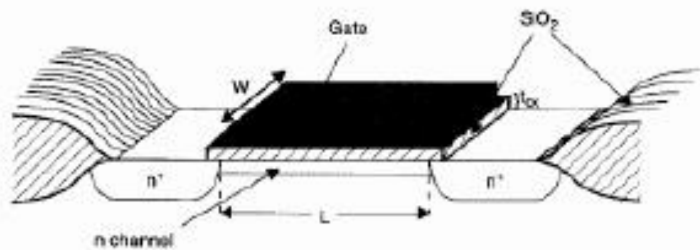
Onset of Inversion

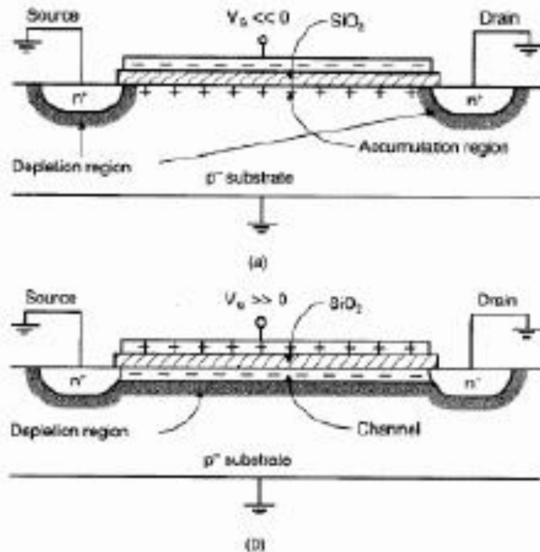


Formation of inversion layer



MOS Transistors





- $L_{gate} = L_{layout} - 2_{LD}$
- For nMOS, $V_D > V_S > V_B$,
- For pMOS, $V_D<V_S<V_B$
- For enhancement mode device, $V_{tn} > 0$, and $V_{tp}< 0$.

Threshold Voltage

- As V_G becomes more positive, the holes in the substrate are repelled from the gate area, leaving negative ions behind so as to mirror the charge on the gate. → depletion region
- As V_G increases, so do the width of the depletion region and the potential at the oxide-silicon interface.
- If V_G rises further, the charge in the depletion region remain relatively constant while the channel charge density continue to increase, providing a greater current from S to D.
- Threshold voltage V_{TH} (for the interface is as much n-type as the substrate is p-type).

Threshold Voltage

$$V_T = \phi_{ms} + 2\phi_f + \frac{Q_b}{C_{ox}} - \frac{Q_{SS}}{C_{ox}} = \phi_{ms} + 2\phi_f + \frac{Q_{b0}}{C_{ox}} - \frac{Q_{SS}}{C_{ox}} + \frac{Q_b - Q_{b0}}{C_{ox}}$$

$$= V_{t0} + \gamma\left(\sqrt{2\phi_f + V_{SB}} - \sqrt{2\phi_f}\right)$$

$\phi_{ms} =$ work function difference between metal and this silicon

$2\phi_f + \frac{Q_b}{C_{ox}} =$ cause inversion layer and sustain depletion layer change

$Q_{SS} =$ Si-SiO_2 surface change

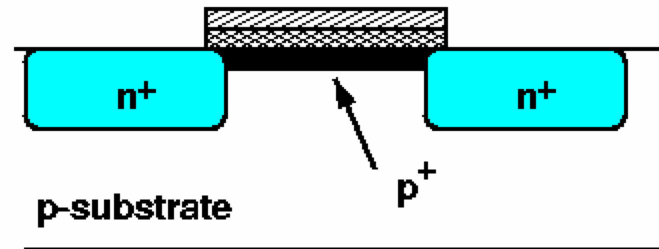
$$\phi f = {}^{kT}/_{q} \ln(N_{sub}/ni) \cdots Fermilevel$$

$$\gamma = \sqrt{2q\epsilon N_{sub}}/Cox \ \cdots typical\ 0.5\ V^{\frac{1}{2}}$$

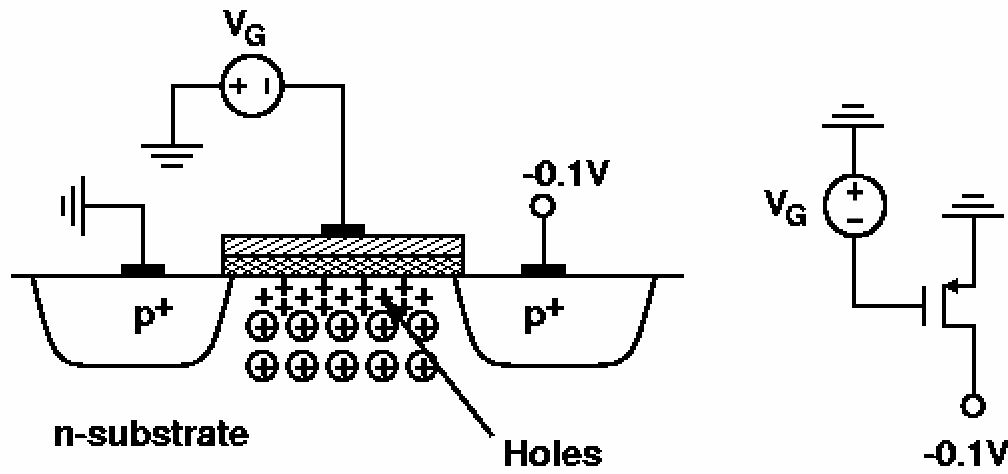
$$\mathrm{Cox} = \epsilon_{\mathrm{ox}}/\mathrm{t_{ox}} \cdots 3.5\ fF/\mu m^2\ for\ 100\text{Å}$$

Threshold Voltage Adjustment

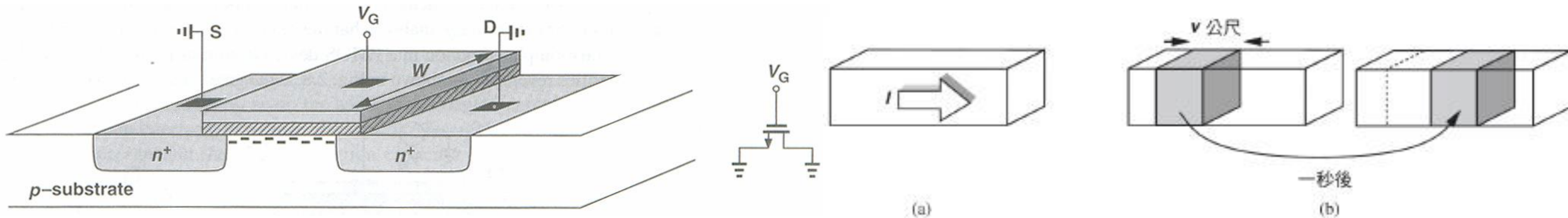
- Implementation of ***p+*** dopants to alter the threshold.



- If a thin sheet of *p+* is created, the gate voltage required to deplete the region increases , V_{tp} ↑
- Formation of inversion layer in a PFET (similar to NMOS with reversed polarities)



I/V Characteristic



- Q_d (coulombs/m) : Charge density along the direction of current. $I = Q_d \bullet v$
- *v* (m/sec) : velocity of the charge.
- The inversion charge density produced by the gate oxide capacitance is proportional to $V_{GS} - V_{TH}$.

$$Q_d = WC_{ox}(V_{GS} - V_{TH})$$

- *Let V(x) is the channel potential at x* $Q_d(x) = WC_{ox}(V_{GS} - V(x) - V_{TH})$

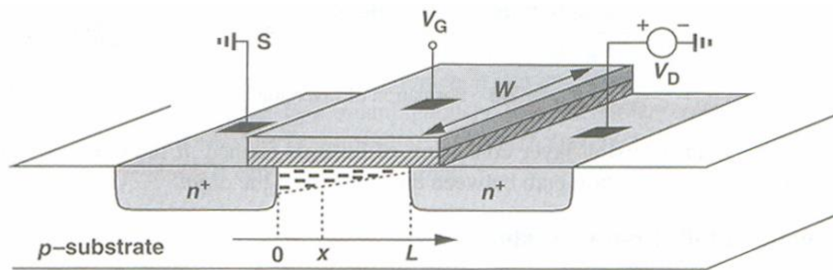
$$I_D = -WC_{ox}\left[V_{GS} - V(x) - V_{TH}\right]v \qquad v = \mu E \qquad E(x) = -dV/dx$$

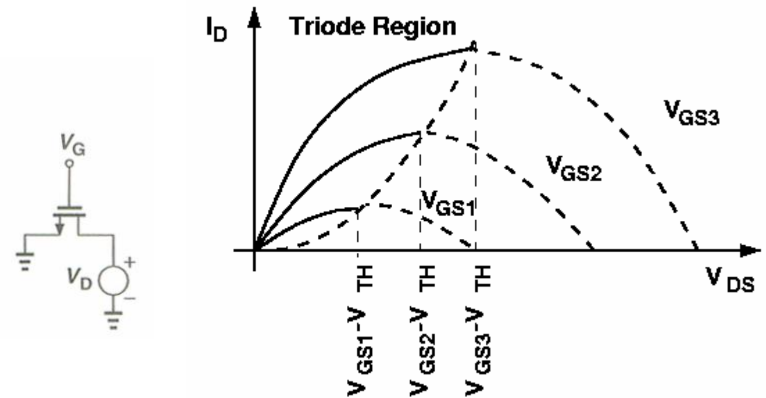
$$I_D = -WC_{ox}\left[V_{GS} - V(x) - V_{TH}\right]\mu_n \frac{dV(x)}{dx} \qquad \int_{x=0}^{L} I_D dx = \int_{V=0}^{V_{DS}} WC_{ox}\mu_n\left[V_{GS} - V(x) - V_{TH}\right]dV$$

$$I_D = \mu_n C_{ox}\frac{W}{L}\left[\left(V_{GS} - V_{TH}\right)V_{DS} - \frac{1}{2}V_{DS}^2\right]$$

I/V Characteristic

- Channel charge with unequal source and drain voltage





- The peak current of I_D can be found by calculating

$$\frac{\partial I_D}{\partial V_{DS}} = 0 \quad \Rightarrow \quad V_{DS} = V_{GS} - V_{TH} \qquad I_{D,\max} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{TH})^2$$

- $V_{GS} - V_{TH}$: Overdrive voltage V_{ov}, W/L: Aspect ratio
- If

$$V_{DS} \leq V_{GS} - V_{TH}$$

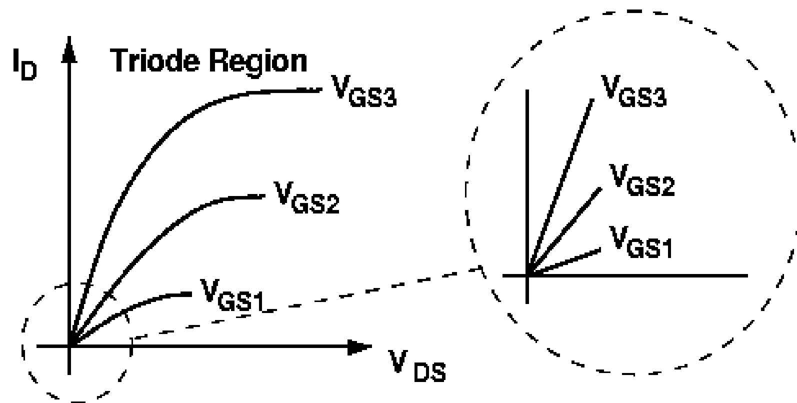
the device is operated in the triode region.

Deep Triode Region

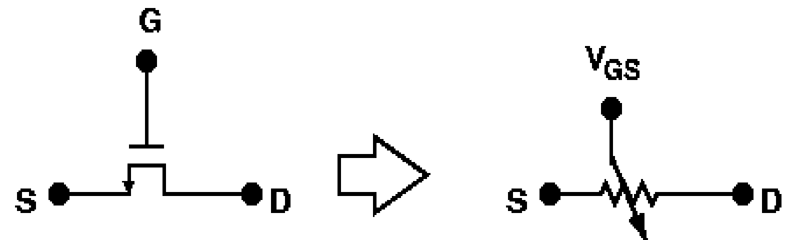
- If $V_{DS} << 2(V_{GS} - V_{TH})$

$$I_{D,\max} \approx \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{TH})V_{DS} \quad \Rightarrow \quad R_{on} = \frac{1}{\mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{TH})}$$

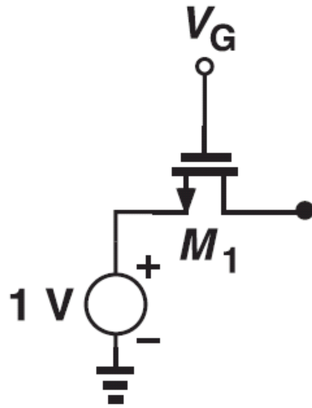
- I/V Characteristic

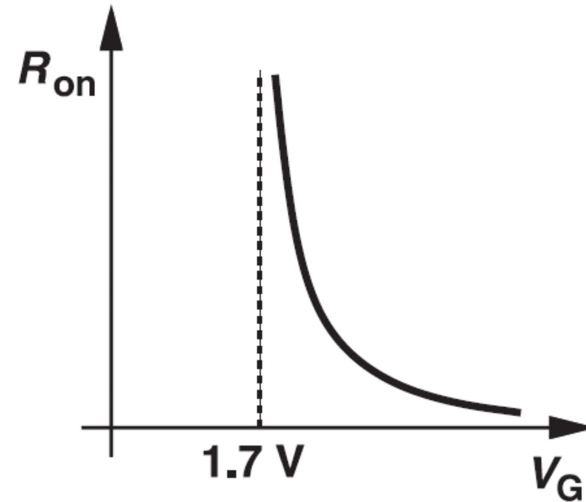


- MOSFET as a controlled linear resistor



Example





- "ON" resistance of M1 as a function of V_G with $\mu_n C_{ox}$= 50 μA/V^2, W/L= 10, V_{TH} = 0.7V. Note that the drain terminal is open.
- For V_G < 1 V + V_{TH} = 1.7 V

$$R_{on} = \infty$$

- For V_G > 1 V + V_{TH} = 1.7 V

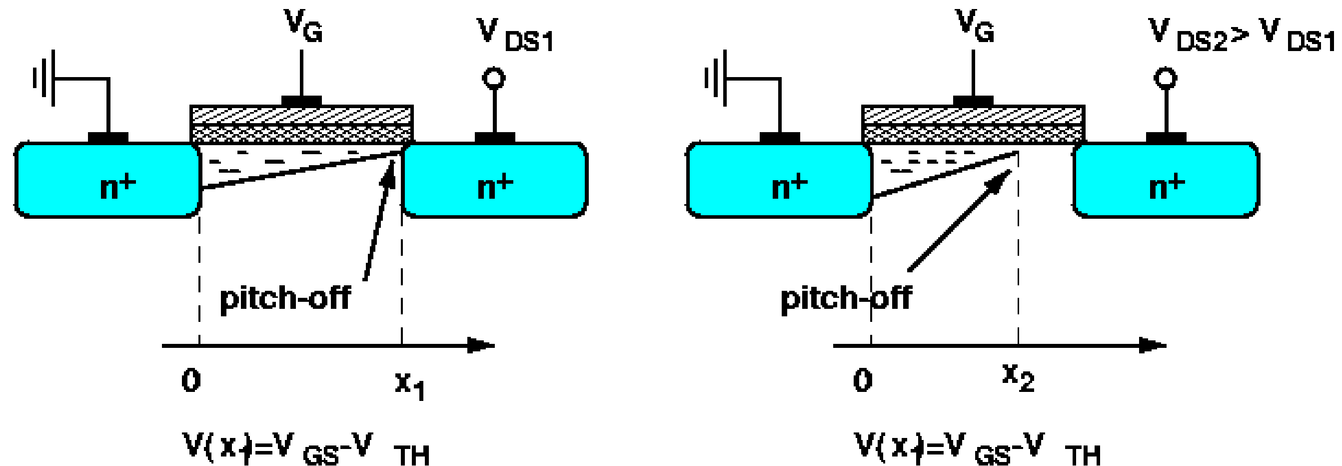
$$R_{on} = \frac{1}{50\ \mu\text{A}/\text{V}^2 \times 10(V_G - 1\ \text{V} - 0.7\ \text{V})}$$

Saturation of Drain Current

- As the local charge density of inversion layer is proportional to

$$Q_d(x) = WC_{ox}(V_{GS} - V(x) - V_{TH}) \quad \Rightarrow \text{ If } V(x) = V_{GS} - V_{TH} \text{ then } Q_d = 0 \quad \text{(pinched-off)}$$

- The inversion layer stops at $x \leq L$
- As V_{DS} increases further, the point at which Q_d equals zero gradually moves toward the source.



- Where *L'* is the point at which Q_d drops to zero ($L' \approx L$).

$$\int_{x=0}^{L'} I_D dx = \int_{V=0}^{V_{GS}-V_{TH}} WC_{ox}\mu_n[V_{GS} - V(x) - V_{TH}]dV \qquad I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L'}(V_{GS} - V_{TH})^2$$

MOS I-V Relation

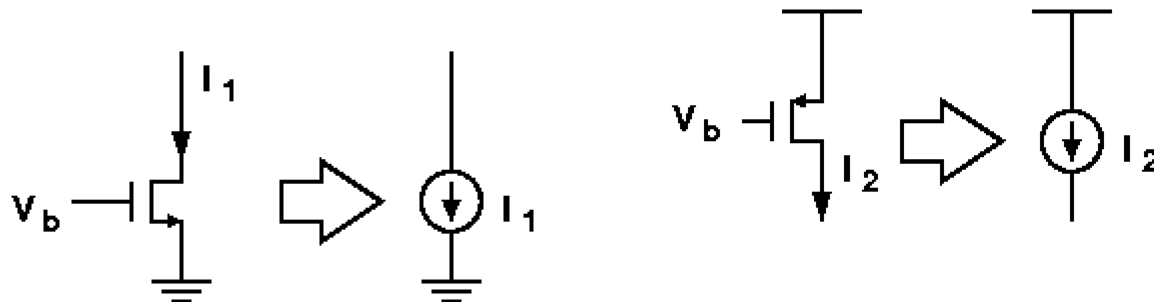
- Triode region $\left(V_{GD} > V_{TH} \Rightarrow V_{GS} - V_{DS} > V_{TH} \Rightarrow V_{GS} - V_{TH} > V_{DS}\right)$

$$I_{DS} = \mu_n C_{ox} \frac{W}{L}\left[(V_{GS} - V_{TH})V_{DS} - \frac{1}{2}V_{DS}^2\right] \quad \text{(N)} \qquad I_{SD} = \mu_p C_{ox} \frac{W}{L}\left[(V_{SG} - |V_{TH}|)V_{SD} - \frac{1}{2}V_{SD}^2\right] \quad \text{(P)}$$

- Saturation region $\left(V_{GD} < V_{TH} \Rightarrow V_{GS} - V_{DS} < V_{TH} \Rightarrow V_{GS} - V_{TH} < V_{DS}\right)$

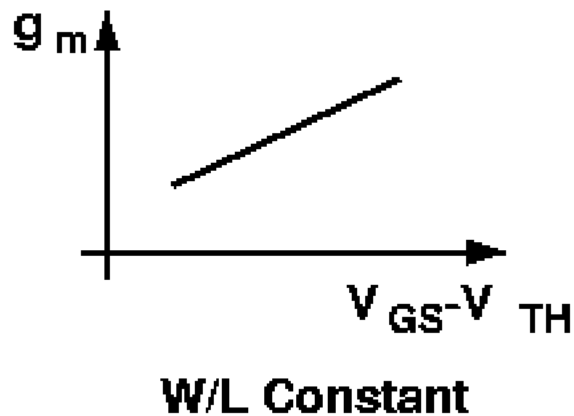
$$I_{DS} = \frac{1}{2}\mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{TH})^2 \quad \text{(N)} \qquad I_{SD} = \frac{1}{2}\mu_p C_{ox} \frac{W}{L}(V_{SG} - |V_{TH}|)^2 \quad \text{(P)}$$

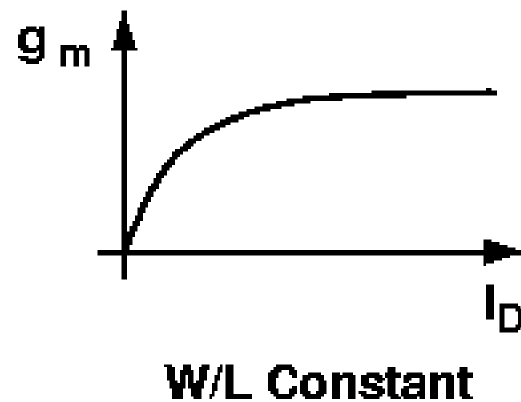
- Saturated MOSFETs operating as current source

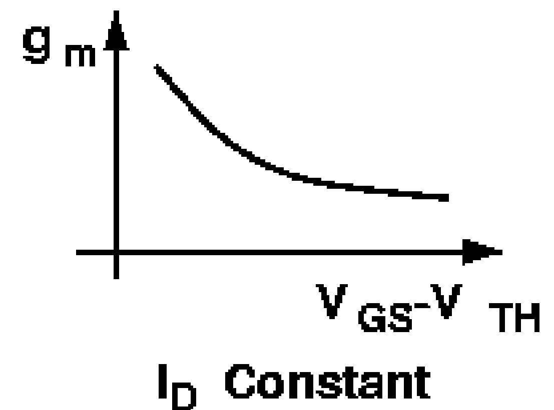


Transconductance (in Sat.)

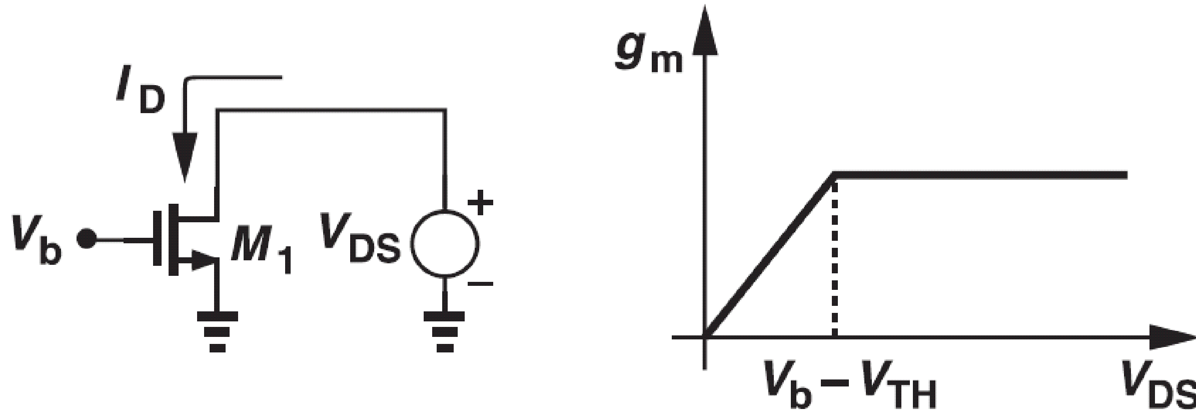
$$g_m = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{VDS,const} = \mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{TH}) = \sqrt{2\mu_n C_{ox}\frac{W}{L}I_D} = \frac{2I_D}{V_{GS} - V_{TH}}$$







Transconductance (in Tri.)



- Plot the transconductance as a function of V_{DS}
- M1 is operated in saturation region for $V_{DS} \geq V_b - V_{TH}$

$$g_m = \mu_n C_{ox} \frac{W}{L} \left(V_b - V_{TH} \right)$$

- M1 is operated in triode region for

$$g_m = \frac{\partial}{\partial V_{GS}} \left\{ \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \left[2\left(V_{GS} - V_{TH} \right) V_{DS} - V_{DS}^2 \right] \right\} = \mu_n C_{ox} \frac{W}{L} V_{DS}$$

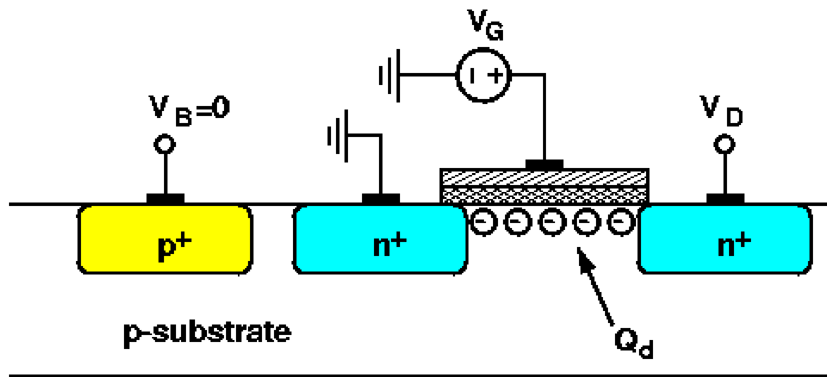
- g_m drops if device enters the triode region, therefore, we usually employ MOSFET in ***saturation*** for amplification.

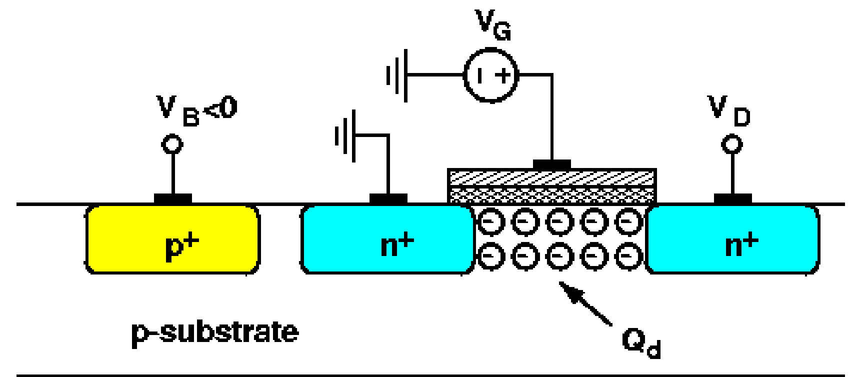
Outline

1. P-N Junction
2. Bipolar Junction Transistor
3. **MOS Field-Effect Transistor**
    - I/V characteristic
    - Second-order effect
    - Small-signal model
    - Scaling & Short-channel effects
    - Simulation models

Second Order Effect

- Body Effect (Back Gate Effect)





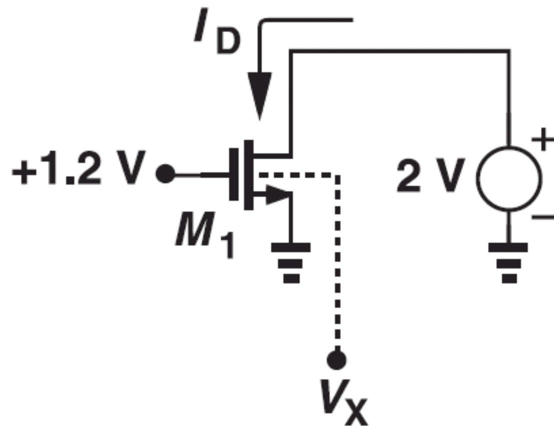
$$V_{TH} = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep}}{C_{ox}} + \frac{Q_{ss}}{C_{ox}}$$

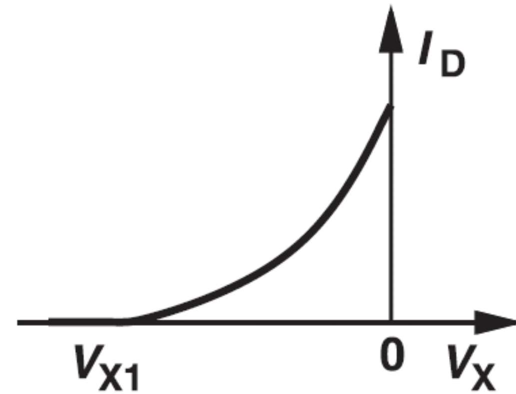
$$V_{TH} = V_{TH0} + \gamma\left(\sqrt{|2\Phi_F + V_{SB}|} - \sqrt{|2\Phi_F|}\right)$$

$$\gamma = \sqrt{2q\varepsilon_{si}N_{sub}} / C_{ox}$$: Body effect coefficient

- As $V_{BS} < 0$
  - More holes are attracted to the substrate connection.
  - The depletion region becomes wider (Q_{dep} *increases)*
  - V_{TH} also increases.

Example: Body Effect





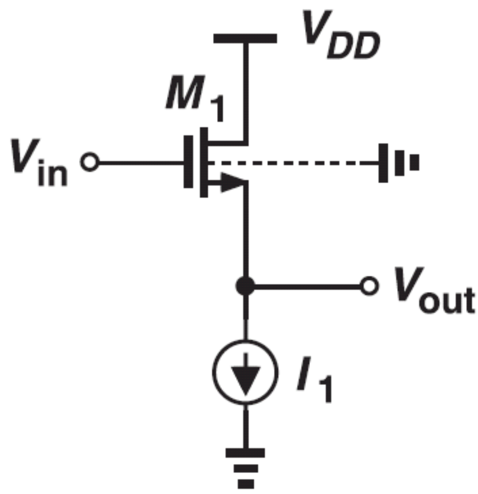
- Plot the drain current if V_X varies from -∞ to 0. Assume V_{TH0} = 0.6 V, γ= 0.4 $V^{0.5}$, and $2\Phi_F$ = 0.7 V
- If V_X is sufficiently negative, the threshold voltage of M_1 exceeds 1.2 V and the device is off.

$$1.2V = 0.6 + 0.4\left(\sqrt{0.7 - V_{X1}} - \sqrt{0.7}\right) \Rightarrow V_{X1} = -4.76V$$

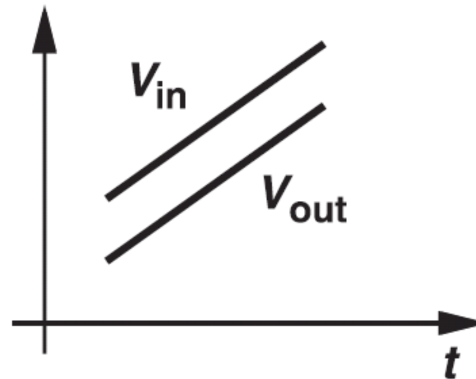
- For $-4.76V < V_{X1} < 0$

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left[V_{GS} - V_{TH0} - \gamma\left(\sqrt{2\phi_F - V_X} - \sqrt{2\phi_F}\right)\right]^2$$

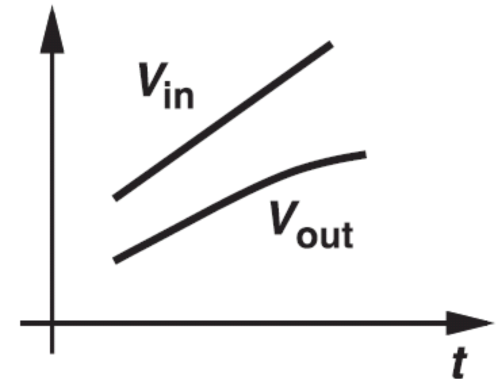
Body Effect of Source Follower



(a)



(b)



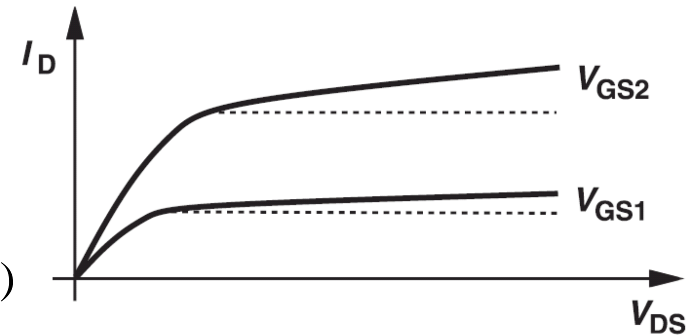
(c)

- (b) without body effect
- (c) with body effect
  - The source and bulk increases, raising the value of V_{TH}.

$$I_1 = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{in} - V_{out} - V_{TH})^2$$

Channel Modulation Effect

- The actual length of the inverted channel gradually decreases as the potential difference between the gate and the drain increases.
- L' is a function of V_{DS}.



$$L' = L - \Delta L \quad \Rightarrow \quad \frac{1}{L'} \approx \frac{1 + \Delta L / L}{L}$$

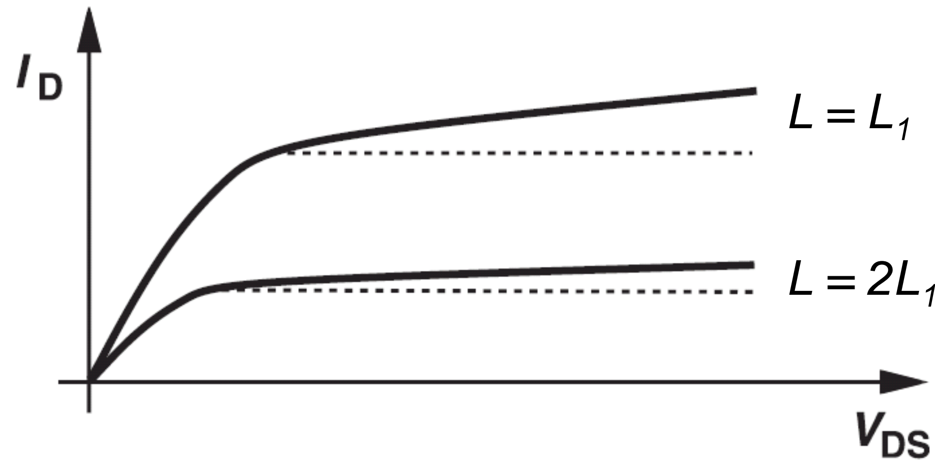
$$\text{if} \quad \frac{\Delta L}{L} \approx \lambda V_{DS} \quad \Rightarrow I_D \approx \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2 (1 + \lambda V_{DS})$$

$$g_m = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})(1 + \lambda V_{DS}) = \sqrt{2 \mu_n C_{ox} (W/L) I_D (1 + \lambda V_{DS})}$$

λ : Channel-length modulation coefficient

- The linear approximation $\frac{\Delta L}{L} \propto V_{DS}$ becomes less accurate in short-channel transistors, resulting in a variable slope in the saturated I_D / V_{DS}.
- Result in a nonideal current source.

Example



- Keeping all other parameters constant, plot I_D/V_{DS} characteristic of a MOSFET for $L = L_1$ and $L = 2L_1$

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_{GS} - V_{TH}\right)^2\left(1 + \lambda V_{DS}\right) \qquad \lambda \propto \frac{1}{L} \qquad \frac{\partial I_D}{\partial V_{DS}} \propto \frac{\lambda}{L} \propto \frac{1}{L^2}$$

- If the length is doubled, the slop of I_D vs V_{DS} is divided by four.
- For a given V_{OV}, a larger L gives a more ideal current source, and W need to be increased proportionally to keep the current capability.

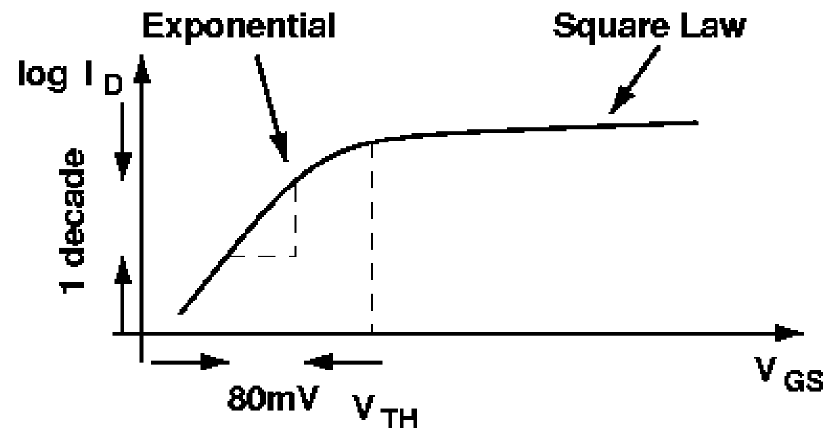
Subthreshold Conduction

- ***Weak Inversion*** : $V_{GS} \approx V_{TH}$, a weak inversion layer exists, small I_D .
- ***Subthreshold Conduction*** : $V_{GS} < V_{TH}$, I_D is finite, it exhibits an *exponential dependence* on V_{GS}. For V_{DS} *greater than roughly 200 mV,*

$$I_D \approx I_0 \exp \frac{V_{GS}}{\zeta V_T}$$

  - With typical values of ζ, at room temperature V_{GS} must decrease ~80 mV for I_D to decrease by one decade.
  - If $V_{TH} = 0.3\ V$, the drain current decreases by only a factor of $10^{3.75}$ when V_{GS} *is reduced to zero.*
  - In rough calculations, we often view V_{TH} as the gate source voltage yielding

    $I_D / W = 1\mu A / \mu m$
  - The transconductance of MOSFET in subthreshold region is $g_m = \frac{I_D}{\zeta V_T}$
  - Which is inferior to that of bipolar transistors.



Voltage Breakdown/Punch Through

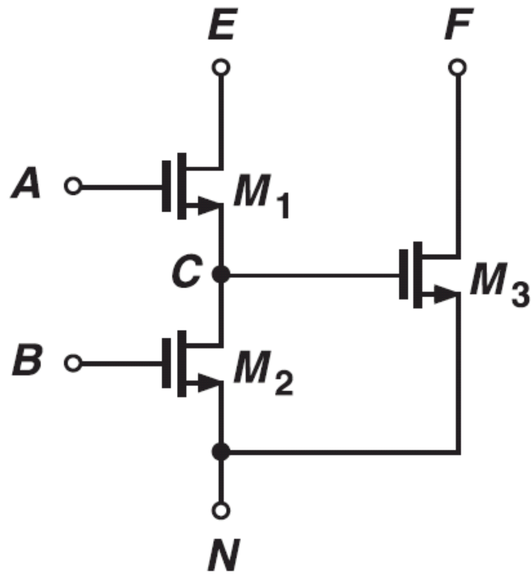
- *Voltage breakdown* : At high gate-source voltages, the gate oxide breaks down irreversibly, damaging the transistor.

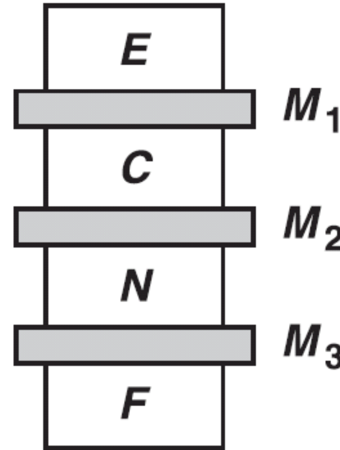
- *Punch through* : In short channel devices, an excessively large drain source voltage widens the depletion region around the drain so much that it touches around the source, creating a very large drain current.

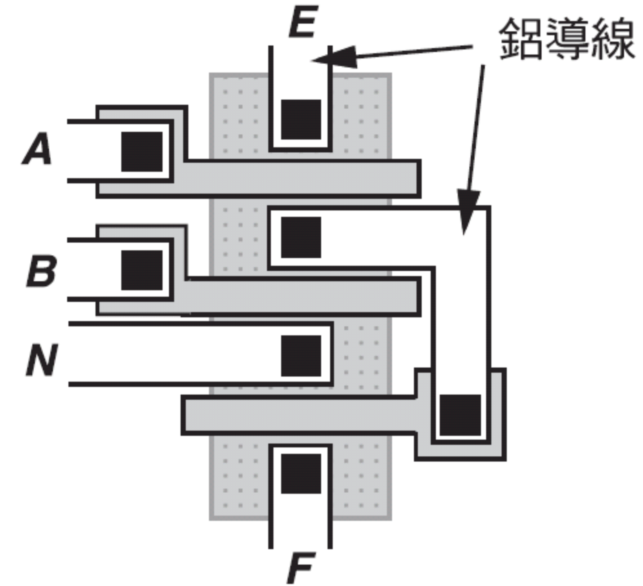
Outline

1. P-N Junction
2. Bipolar Junction Transistor
3. **MOS Field-Effect Transistor**
   - I/V characteristic
   - Second-order effect
   - Small-signal model
   - Scaling & Short-channel effects
   - Simulation models

Layout Example

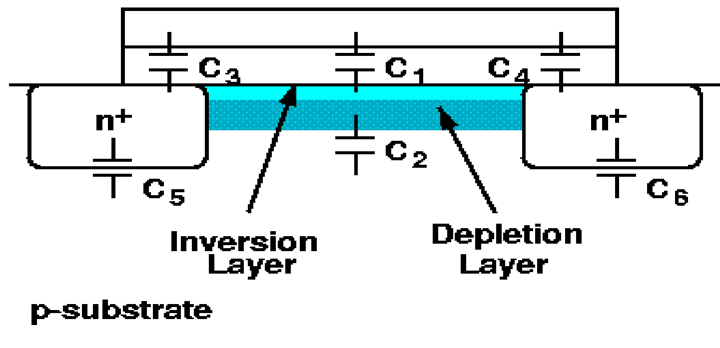


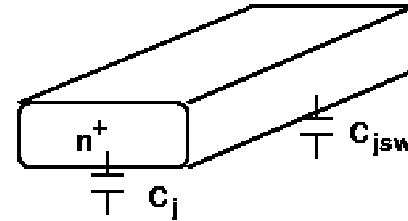


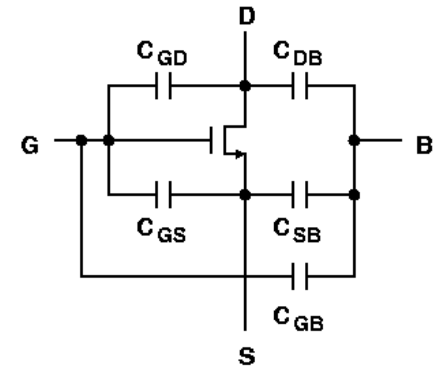


- Share the source drain junction at node C and M2 and M3 also do so at node N .
- The gate polysilicon of M3 is connected to C by metal interconnect.

MOS Device Capacitance







- C_1 : Oxide capacitance between the gate and the channel, $C_1 = WLC_{ox}$
- C_2 : Depletion capacitance between the channel and the substrate

$$C_2 = WL\sqrt{q\varepsilon_{si}N_{sub}/4\Phi_F}$$

- C_3, C_4 : Overlap capacitance of the gate poly with the source and drain areas

$$C_{GD} = C_{GS} = C_{ov}W \approx WL_D C_{ox}$$

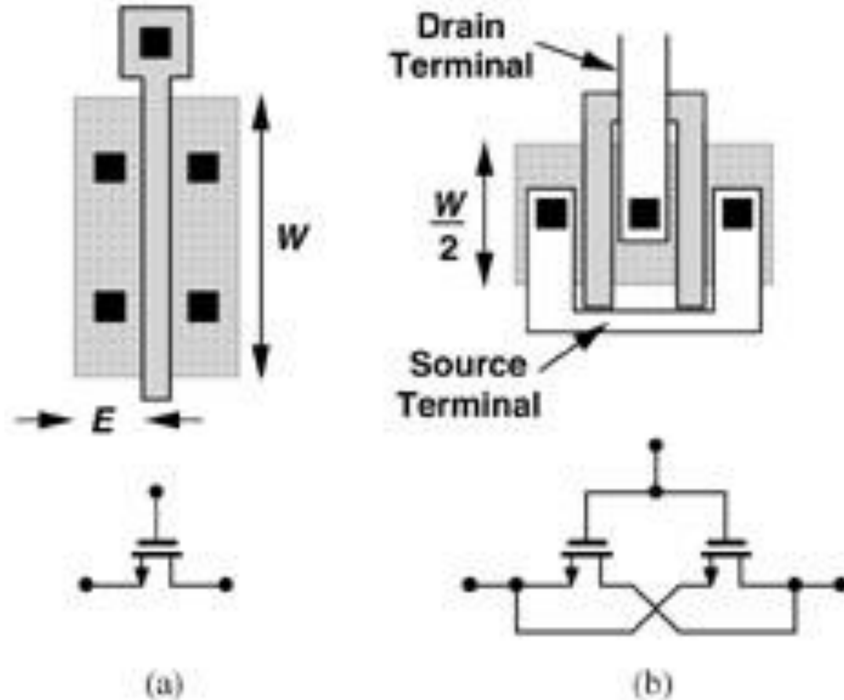
- Junction capacitance between the source drain areas and the substrate.
  - C_j (F/m^2) : bottom plate capacitance from the bottom of the junction.
  - C_{jsw} (F/m) : side-wall capacitance due to the perimeter of the junction.

$$C_j = C_{j0}/\left[1 + V_R/\Phi_B\right]^m$$

Example

- For the transistor in (a)

$$C_{DB} = C_{SB} = WEC_j + 2(W+E)C_{jsw}$$

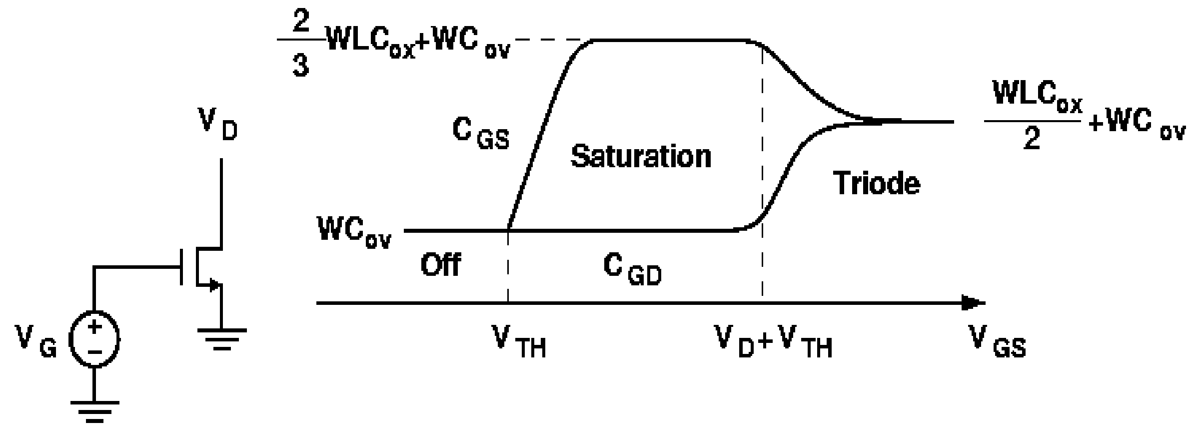


- For the transistor in (b) : "folding"
- Less "drain" junction cap.

$$C_{DB} = \frac{W}{2}EC_j + 2(\frac{W}{2}+E)C_{jsw}$$

$$C_{SB} = 2\left[\frac{W}{2}EC_j + 2\left(\frac{W}{2}+E\right)C_{jsw}\right] = WEC_j + 2(W+2E)C_{jsw}$$

C_{GS} and C_{GD} v.s. V_{GS}



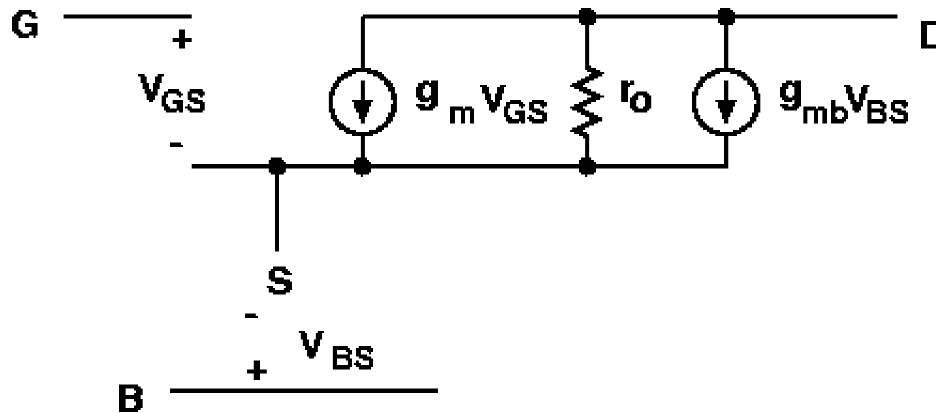
- If the device is off
  - $C_{GD} = C_{GS} = C_{ov}W.$
  - $C_{GB} = C_{ox}WL$: gate to substrate overlap capacitance.
- If the device is weakly on: C_{GB} consists of the series comb. of C_{gate} and C_{dep}.
- If the device is in the deep triode region

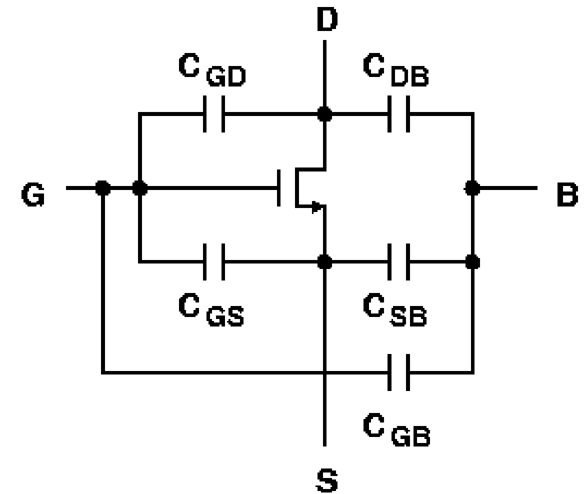
$$C_{GD} = C_{GS} = WLC_{ox}/2 + WC_{ov}$$

- If the device is in the saturation region
  - The potential difference $V_{Gate\ to\ channel}$ varies from V_{GS} at the source to $V_{GS} - V_{TH}$ at the pinch-off point.

$$C_{GS} = \frac{2}{3}WLC_{ox} + WC_{ov} \qquad C_{GD} = WC_{ov}$$

MOSFET Small Signal Model





- $g_m V_{GS}$ and $g_{mb} V_{BS}$ have the same polarity.

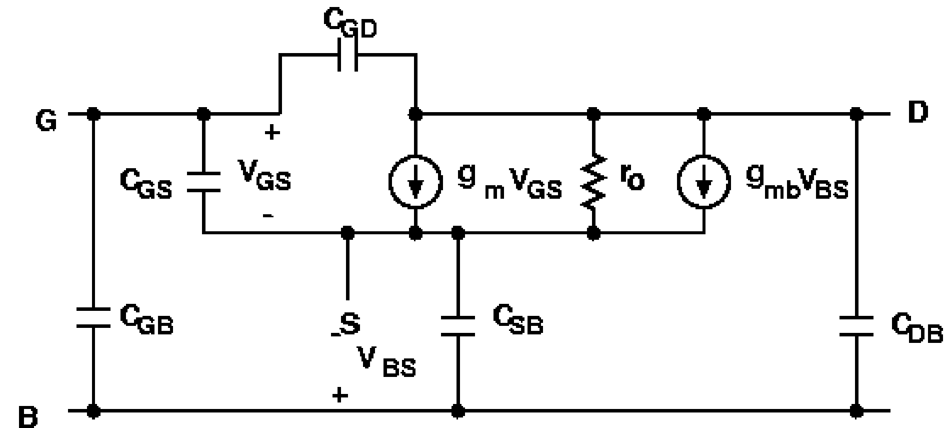
$$r_O = \frac{\partial V_{DS}}{\partial I_D} = \frac{1}{\partial I_D / \partial V_{DS}} = \frac{1}{\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS}-V_{TH})^2 \cdot \lambda} \approx \frac{1}{\lambda I_D}$$

$$g_{mb} = \frac{\partial I_D}{\partial V_{BS}} = \mu_n C_{ox}\frac{W}{L}(V_{GS}-V_{TH})\left(-\frac{\partial V_{TH}}{\partial V_{BS}}\right) = g_m \frac{\gamma}{2\sqrt{2\Phi_F + V_{SB}}} = \eta g_m$$

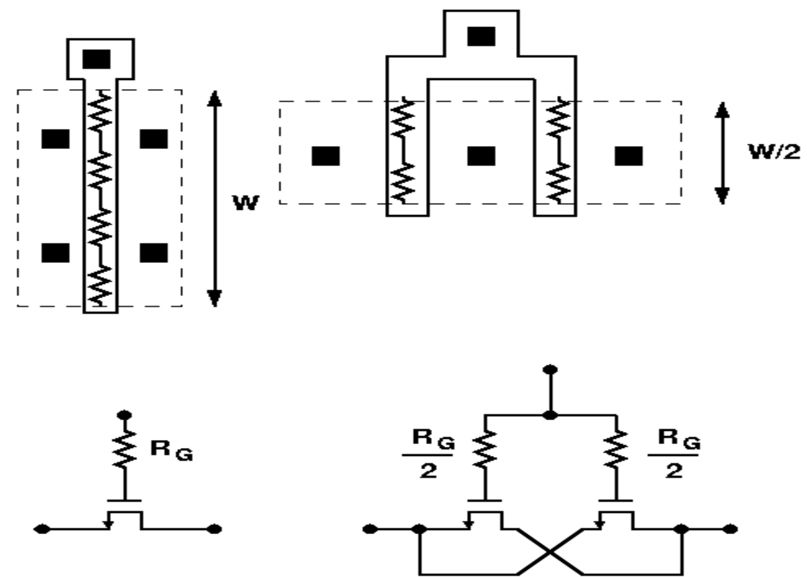
where $\frac{\partial V_{TH}}{\partial V_{BS}} = -\frac{\partial V_{TH}}{\partial V_{SB}} = -\frac{\gamma}{2}(2\Phi_F + V_{SB})^{-1/2}$ $\quad \eta = \frac{g_{mb}}{g_m} \approx 0.1\text{-}0.3$

Complete MOS Small-Signal Model

- Complete MOS small signal model



- Reduction of gate resistance by folding



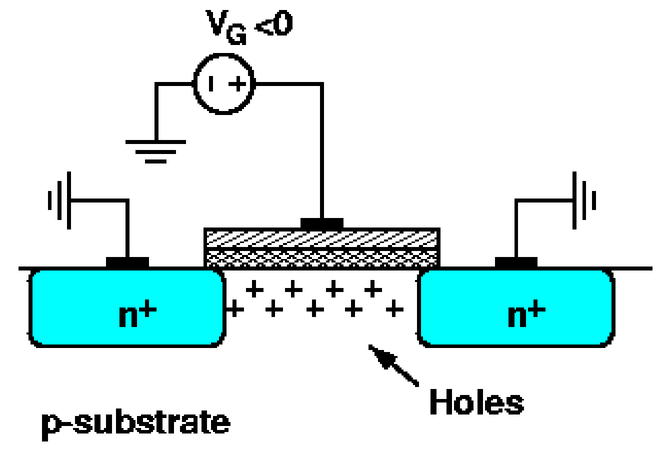
NMOS v.s. PMOS Devices

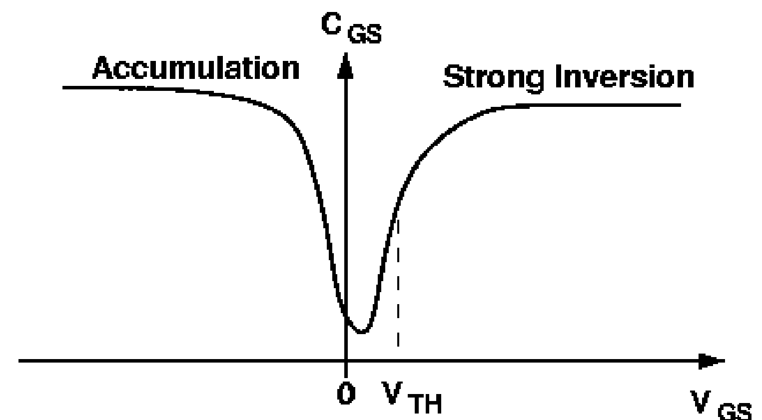
- $\mu_p C_{ox} \approx 0.25 \mu_n C_{ox}$
- For given W/L and I_{bias}, NMOS exhibit a higher output resistance.
- MOS device as a capacitor
  - ***Accumulation*** ($V_{GS} < 0$)

    $C_{GS} = C_{ox} WL$
  - ***Weak inversion*** ($V_{GS} < V_{TH}$), a depletion region begins to form under the oxide.

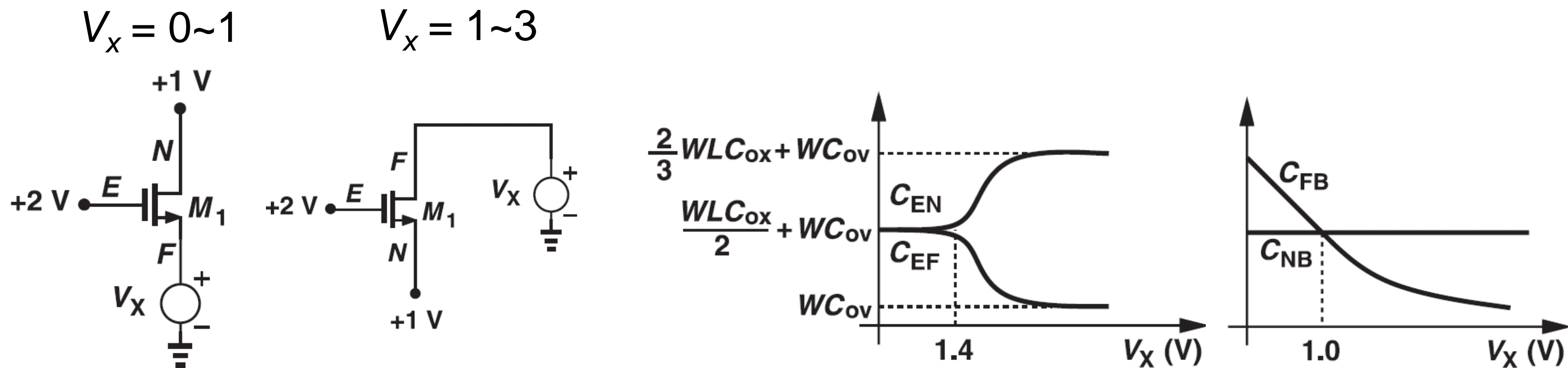
    $C_{GS} = C_{ox} WL C_{dep} / (C_{ox} WL + C_{dep})$
  - ***Strong Inversion*** ($V_{TH} < V_{GS}$), the oxide-silicon interface sustains a channel.

    $C_{GS} = C_{ox} WL$





Example



- Sketch the capacitance of M_1 as V_X varies from zero to 3V. Assume V_{TH} = 0.6V and $\lambda = \gamma = 0$.
- C_{NB} is independent of V_X
- For V_x ~ 0, M1 is in the triode region

$$C_{EN} \approx C_{EF} = \frac{1}{2} WLC_{ox} + WC_{ov} \qquad\qquad C_{FB} \ \text{maximum}$$

- For V_x > 1V, the role of the source and drain is exchanged
- M_1 is out of triode region for $V_X \geq 2V - 0.6V$

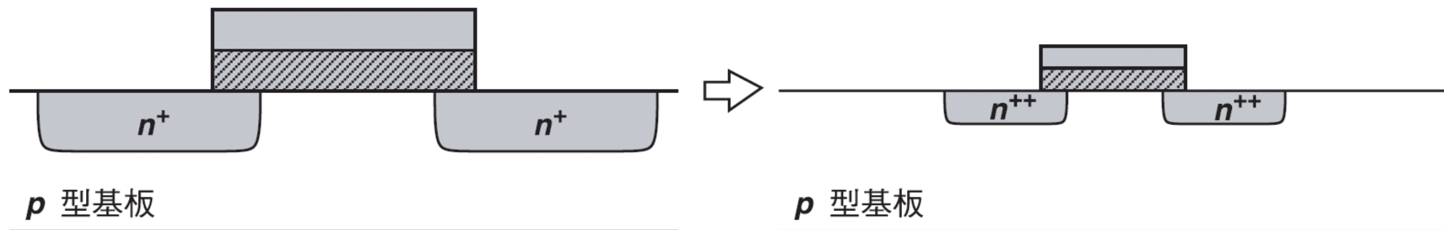
Outline

1. P-N Junction
2. Bipolar Junction Transistor
3. **MOS Field-Effect Transistor**
    - I/V characteristic
    - Second-order effect
    - Small-signal model
    - Scaling & Short-channel effects
    - Simulation models

Scaling Theory

- The square-law characteristic provide accuracies for devices with minimum channel lengths of greater than 4 μm.
- The two principal reasons for the dominance of CMOS technology in today's semiconductor industry
  - Zero static power dissipation
  - Scalability
- Scaling theory follows three rules
  - Reduce all lateral and vertical dimensions by α ($\alpha > 1$)
    - W, L, t_{ox}, depth and perimeter of the source drain junctions
  - Reduce the threshold voltage and supply voltage by α (V_{DD}, V_{TH})
    - Constant field scaling : dimensions and voltage scale together
  - Increase all the doping level by α

Ideal Scaling of MOS Transistor



- The current capability of the transistor *drops* by a factor of α (>1)

$$I_{D,scaled} = \frac{1}{2}\mu_n(\alpha C_{ox})\left(\frac{W/\alpha}{L/\alpha}\right)\left(\frac{V_{GS}}{\alpha} - \frac{V_{TH}}{\alpha}\right)^2 = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{TH})^2\frac{1}{\alpha}$$

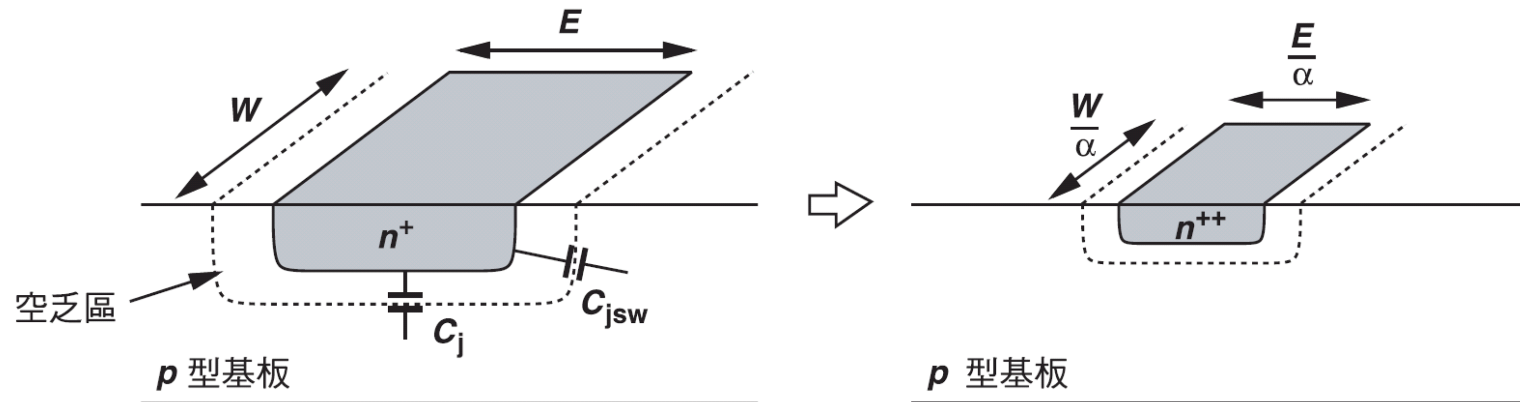
- Reduction of capacitances and power dissipation
  - The channel capacitance *drops* by a factor of α

$$C_{ch,scaled} = \frac{W}{\alpha}\frac{L}{\alpha}(\alpha C_{ox}) = \frac{1}{\alpha}WLC_{ox}$$

  - The depletion width *drops* by a factor of α
    - The depletion region capacitance per unit area *increases* by a factor of α

$$W_{d,scaled} \approx \sqrt{\frac{2\varepsilon_{si}}{q}\left(\frac{1}{\alpha N_A} + \frac{1}{\alpha N_D}\right)\frac{V_R}{\alpha}} \approx \frac{1}{\alpha}\sqrt{\frac{2\varepsilon_{si}}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)V_R}$$

Scaling of S/D Junction Capacitance



- The bottom-plate capacitance of the S/D junction (per unit area), C_j, *increases* by a factor of α.

$$C_{j,scaled} \propto C_{deplete} \propto \alpha$$

- The side-wall capacitance (per unit width) remains constant

$$C_{jsw,scaled} \propto C_{deplete}\frac{D_{depth}}{\alpha} = \text{constant}$$

- In summary, the source drain junction capacitance *drops* by a factor of α

$$C_{S/D,scaled} = \frac{W}{\alpha}\frac{E}{\alpha}(\alpha C_j) + 2\left(\frac{W}{\alpha} + \frac{E}{\alpha}\right)(C_{jsw}) = \left[WEC_j + 2(W+E)C_{jsw}\right]\frac{1}{\alpha}$$

Gate Delay and Power Dissipation

- The gate delay T_d

$$T_{d,scale} = \frac{C}{I} V_{DD} = \frac{C/\alpha}{I/\alpha} \frac{V_{DD}}{\alpha} = \left(\frac{C}{I} V_{DD} \right) \frac{1}{\alpha}$$

- The speed of the digital circuit

$$f \propto \frac{1}{T_{d,scale}} \propto \alpha$$

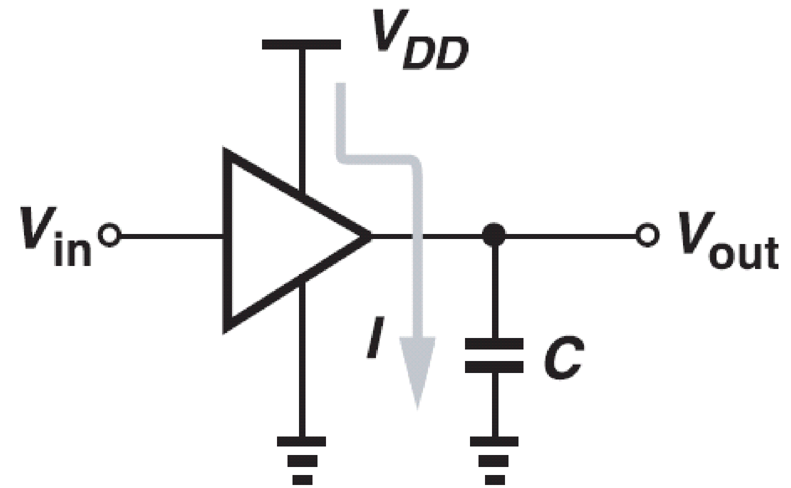
- The dynamic power dissipation

$$P = fCV_{DD}^2 = f \frac{C}{\alpha} \left(\frac{V_{DD}}{\alpha} \right)^2 = \frac{fCV_{DD}^2}{\alpha^3}$$

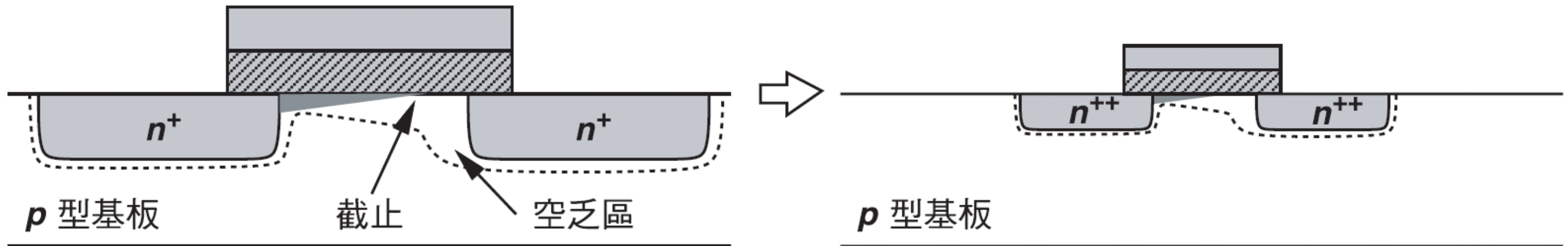
- The layout density (the number of transistors per unit area)

$$Density = \alpha^2$$

- Reduction of power and delay while increase the circuit density
  - Extremely attractive for digital systems



Effect of Ideal Scaling in Analog Circuits



- If all of the dimensions and voltages (and currents) scale down.

$$g_{m,scaled} = \mu(\alpha C_{ox})\frac{W/\alpha}{L/\alpha}\frac{V_{GS}-V_{TH}}{\alpha} = \mu C_{ox}\frac{W}{L}(V_{GS}-V_{TH})$$

- If dimensions scale down while overdrive voltage remain constant

$$g_{m,scaled} = \mu(\alpha C_{ox})\frac{W/\alpha}{L/\alpha}(V_{GS}-V_{TH}) = \alpha\mu C_{ox}\frac{W}{L}(V_{GS}-V_{TH})$$

- Consider output impedance
  - The width of the depletion region around the drain decreases by α

$\frac{\Delta L/\alpha}{L/\alpha}$ remains constant, $\lambda = \frac{\frac{\Delta L/\alpha}{L/\alpha}}{V_{DS}/\alpha}$ increases by α, $r_{O,scaled} = \frac{1}{\alpha\lambda\frac{I_D}{\alpha}} = \frac{1}{\lambda I_D}$, $g_m r_O$ remains constant.

Impact of Scaling on Analog Circuit

- Reduction of the supply voltage
  - Maximum allowable voltage swings decreases by α
  - g_m constant and hence thermal noise remain constant
    - The lower end of the dynamic range is limited by thermal noise.
  - The dynamic range (SNR) is decreased by α.
    - $V_{sig}' = V_{sig}/\alpha$, $V_{noise}' = V_{noise}$, $S/N' = V_{sig}'/V_{noise}' = (S/N)/\alpha$
  - How to recover dynamic range (SNR)?
    - Since $i_n^2 = 4kTg_m = g_m^2 v_n^2, \quad v_n \propto \sqrt{kT/g_m}$
    - $V_{noise}' = V_{noise}/\alpha$, $g'_m = \alpha^2 g_m$

Impact of Scaling on Analog Circuit

- To restore the dynamic range
  - The transconductance of the transistors must be increased by α^2
  - The thermal noise voltages and currents scale with $\sqrt{g_m}$

$$g_m = \frac{2I_D}{V_{ov}} = \mu C_{ox}\left(\frac{W}{L}\right)(V_{GS} - V_{TH})$$

  - V_{ov} decreases by a factor of α, I_D must increase by a factor of α, power dissipation remain constant.
  - C_{ox} is scaled up by α, L and V_{ov} are scaled down by α, then W must increase by α.
  - For a constant (***thermal noise limited***) dynamic range, ideal scaling of linear circuits requires a ***constant power dissipation*** and a ***higher device capacitance***.

$$(\frac{V_{DD}}{\alpha})(\alpha I_{DD}) = V_{DD} I_{DD} \qquad (\alpha W)(L/\alpha)(\alpha C_{ox}) = \alpha W L C_{ox}$$

Short Channel Effect

- Small geometry effects arise because
  - The electric fields tend to increase because the supply voltage has not scaled proportionally.
  - The built in potential is neither scalable nor negligible.
  - The depth of S/D junctions cannot be reduced easily.
  - The mobility decreases as the substrate doping increases.
  - The subthreshold slope is not scalable.

Threshold Voltage Variation

- For the threshold voltage
  - The upper bound is roughly equal to VDD/4 to avoid degrading the speed of digital CMOS gates.
  - The lower bound is determined by ***subthreshold behavior, variation with temperature***, ***process***, and ***dependence upon the channel length***.
- temperature coefficient ~-1mV/^{o}K, yielding a 50-mV change across the commercial temperature range (0-50°C).
- Process induced variation is ~50 mV, raising the margin to approximately 100 mV.
- It is difficult to reduce V_{TH} below several hundred millivolts.

Threshold Voltage Variation

- For **long-channel** devices, the subthreshold drain current

$$I_D = \mu C_d \left(\frac{W}{L}\right) V_T^2 \left(\exp\frac{V_{GS} - V_{TH}}{\zeta V_T}\right)\left(1 - \exp\frac{-V_{DS}}{V_T}\right)$$

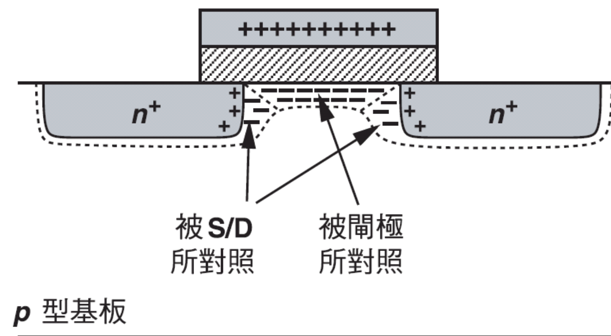
$$C_d = \sqrt{\frac{\varepsilon_{si} q N_{sub}}{4\phi_B}} \quad \text{depletion capacitance under the gate area} \qquad V_T = \frac{kT}{q} \qquad \zeta = 1 + \frac{C_d}{C_{ox}}$$

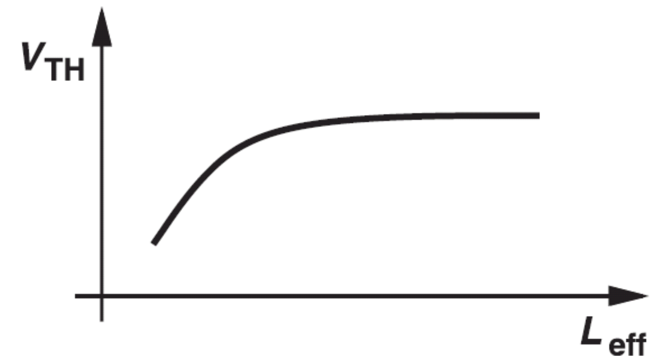
  - As V_{DS} exceeds a few V_T, I_D becomes ***independent*** of the drain source voltage.
  - The slope of I_D on a logarithmic scale becomes

$$\frac{\partial(\log_{10} I_D)}{\partial V_{GS}} = (\log_{10} e)\frac{1}{\zeta V_T}, \quad \text{Sub threshold slope: S} = \left[\frac{\partial(\log_{10} I_D)}{\partial V_{GS}}\right]^{-1} = 2.3 V_T \left(1 + \frac{C_d}{C_{ox}}\right) \text{ V/dec}$$

- $C_d = 0.67C_{ox}$, S = 100mV/dec, a change of 100 mV in V_{GS} leads to a ten-fold reduction in the drain current.
- To turn off the transistor by lowering V_{GS} below V_{TH}
  - S must be as small as possible, C_d/C_{ox} must be minimized.
  - Constant magnitude of S severely limits the scaling of the V_{TH}.

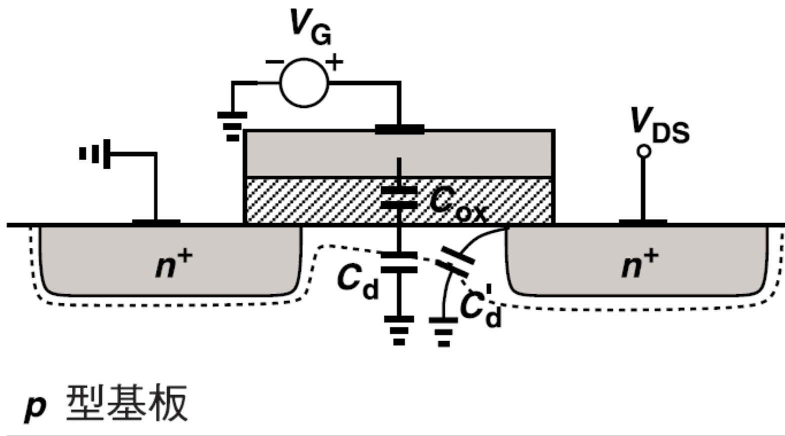
Threshold Voltage Variation



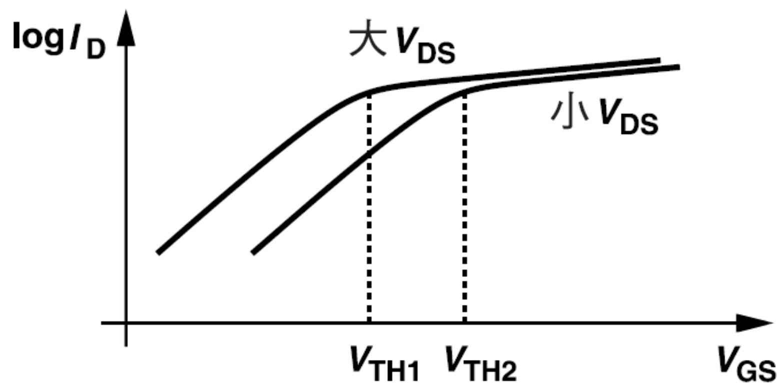


- Different lengths yield lower V_{TH} as L decreases.
  - The depletion regions associated with the source/drain is comparable to channel area.
  - Gate voltage required to create an inversion layer decreases.
  - Length variation also introduces additional variations in V_{TH}.
- If the length is increased for a higher output impedance, then the threshold voltage also increases by as much as 100 ~ 200 mV.

DIBL



(a)



(b)

- DIBL: Drain Induced Barrier Lowering
  - In short channel devices, the ***drain voltage*** also makes the surface more positive by creating a two-dimensional field in the depletion region.
  - The drain introduces a capacitance C_d' that raises the surface potential in a manner similar to C_d.
  - The barrier to the flow charge and hence the threshold voltage are decreased.
  - The principal impact of DIBL on circuit design is the degraded output impedance. (p.2-67)

Mobility Degradation

- High vertical electrical field E_{ver} filed between the gate and the channel confines the charge carriers to a narrow region below the oxide-silicon interface, leading to more scattering and hence lower mobility.
- Small geometry devices experience significant mobility degradation.

$$\mu_{eff} = \frac{\mu_0}{1+\theta(V_{GS}-V_{TH})}$$

  - μ_o denotes the low-field mobility and θ is a fitting parameter ~ $10^{-7}/t_{ox}$ (V^{-1}).
  - θ rises as t_{ox} drops.
  - If t_{ox} =*100 Å,* θ~ $1V^{-1}$, the mobility begins to fall considerably as the overdrive exceeds 100 mV.

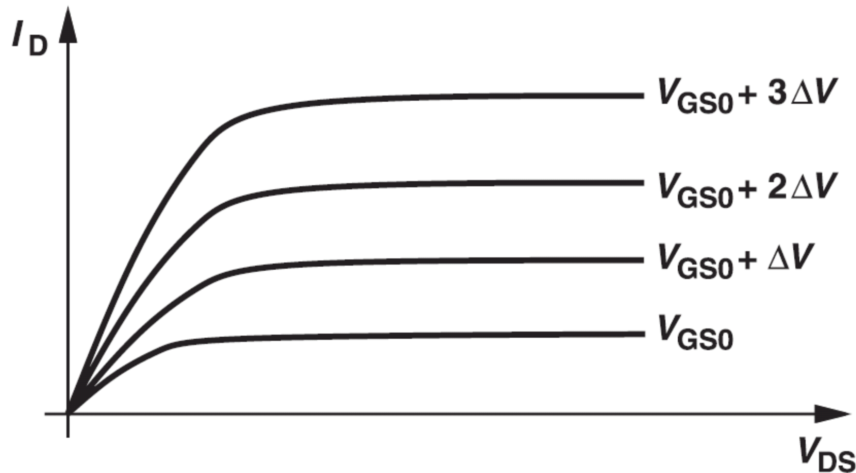
$$I_D = \frac{1}{2}\frac{\mu_0 C_{ox}}{1+\theta(V_{GS}-V_{TH})}\frac{W}{L}(V_{GS}-V_{TH})^2$$

Assume $\theta(V_{GS}-V_{TH}) << 1$

$$I_D \approx \frac{1}{2}\mu_0 C_{ox}\frac{W}{L}[1-\theta(V_{GS}-V_{TH})](V_{GS}-V_{TH})^2 \approx \frac{1}{2}\mu_0 C_{ox}\frac{W}{L}\left[(V_{GS}-V_{TH})^2-\theta(V_{GS}-V_{TH})^3\right]$$

  - Lead to higher harmonics in the drain current.
  - The mobility degradation with the E_{ver} affects the device g_m as well.

Velocity Saturation



$v = \mu E$ approaches a saturated value $10^7 cm/s$

- The μ also depends on the lateral electric filed in the channel, beginning to drop as the field reaches 1V/µm.
- v : saturated value about 10^7 cm/s.

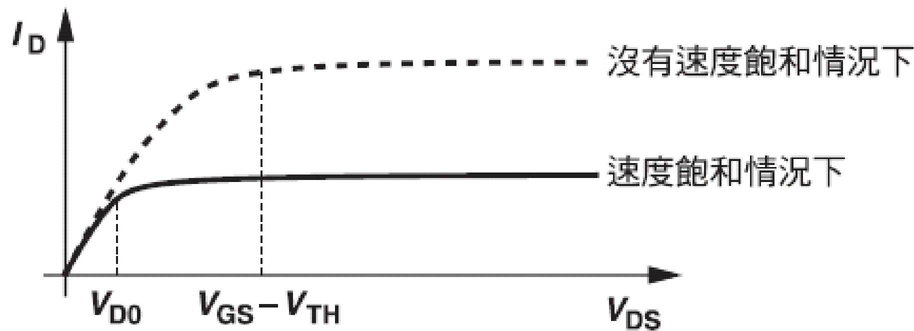
- Carriers may reach a saturated velocity at some point along the channel.
- In the case carriers experience velocity saturation along the entire channel

$$I_D = v_{sat} Q_d = v_{sat} W C_{ox} (V_{GS} - V_{TH})$$

- The current is proportional to the V_{ov} only and does not depend on the length.
- Devices with L < 1 µm reveal velocity saturation because equal increments in V_{GS}-V_{TH} result in roughly equal increments in I_D.
- The transconductance is a weak function of the I_D and channel length.

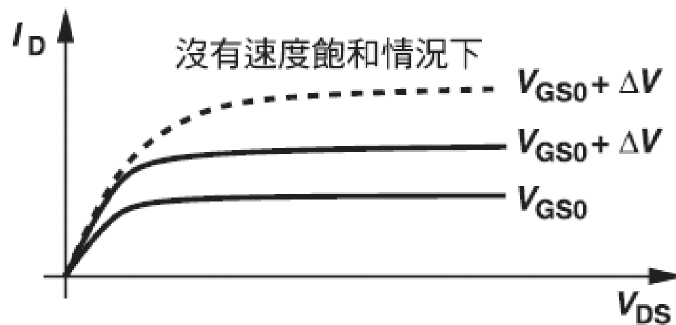
$$g_m = v_{sat} W C_{ox}$$

Effect of Velocity Saturation



$$I_D = WC_{ox}v_{sat}\frac{(V_{GS}-V_{TH})^2}{V_{GS}-V_{TH}+2\frac{v_{sat}L}{\mu_{eff}}}$$

$$V_{D0} = V_{DS,sat} = \frac{2\mu_{eff}L(V_{GS}-V_{TH})}{2\mu_{eff}L+V_{GS}-V_{TH}}$$



$$I_D = WC_{ox}v_{sat}\frac{(V_{GS}-V_{TH})^2}{V_{GS}-V_{TH}+\frac{2v_{sat}L}{\mu_0}\left[1+\theta(V_{GS}-V_{TH})\right]}$$

$$= WC_{ox}v_{sat}\frac{(V_{GS}-V_{TH})^2}{\frac{2v_{sat}L}{\mu_0}+\left(1+\frac{2v_{sat}L\theta}{\mu_0}\right)(V_{GS}-V_{TH})}$$

$$= \frac{1}{2}\mu_0 C_{ox}\frac{W}{L}\frac{(V_{GS}-V_{TH})^2}{1+\left(\frac{\mu_0}{2v_{sat}L}+\theta\right)(V_{GS}-V_{TH})}$$

- Typical bias conditions, MOSFETs experience some velocity saturation, displaying a characteristic between linear and square-law behavior.
- As V_{GS} increases, the drain current saturates well before pinch-off occurs, yielding a constant current lower than that obtained if the device saturated.
- The transconductance is also lower than that predicted by the square law.

Effect of Velocity Saturation

- Put mobility degradation & velocity saturation together

$$I_D = WC_{ox}v_{sat}\frac{(V_{GS}-V_{TH})^2}{V_{GS}-V_{TH}+2\dfrac{v_{sat}L}{\mu_{eff}}} = \frac{1}{2}\mu_0 C_{ox}\frac{W}{L}\frac{(V_{GS}-V_{TH})^2}{1+\left(\frac{\mu_0}{2v_{sat}L}+\theta\right)(V_{GS}-V_{TH})}$$

- If L or v_{sat} is large, the expression reduces to the square law relationship.
- If the overdrive voltage is so small that the denominator is approximated as

$$2\frac{v_{sat}L}{\mu_{eff}} \quad \text{and} \quad \mu_{eff} \approx \mu_0$$

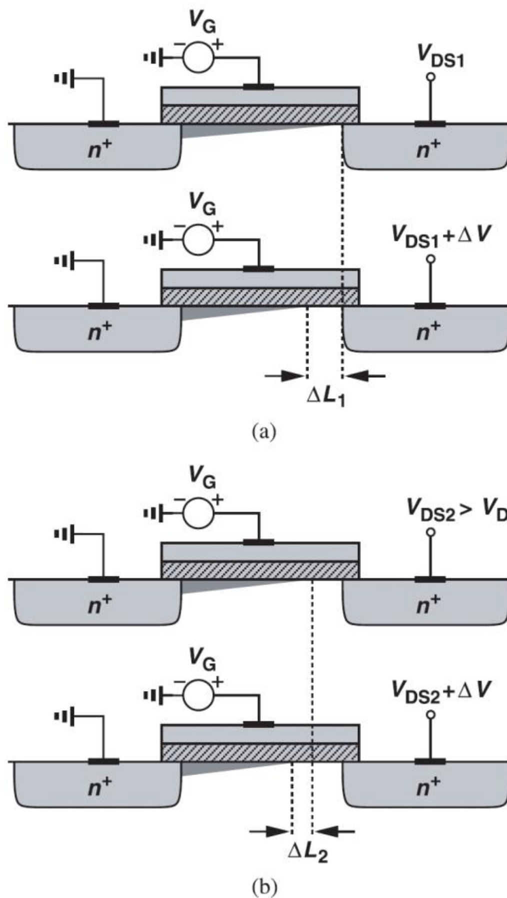
  - the device still follows the square law behavior even if L is relatively small.
- The degradation of the mobility with both lateral and vertical fields can be represented by adding the terms

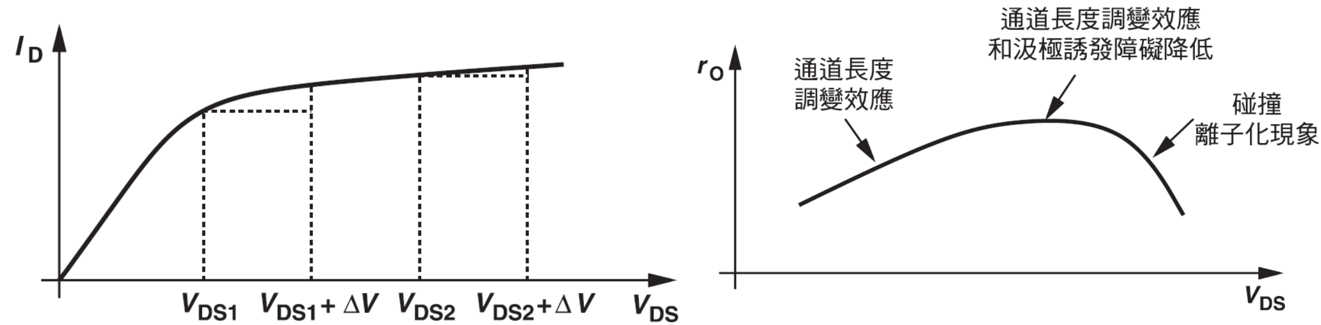
$$\frac{\mu_0}{2v_{sat}L} \quad \text{and} \quad \theta$$

Hot Carrier Effect

- The instantaneous velocity and hence the kinetic energy of carriers continue to increase as they accelerate towards the drain.
- Hot carriers may hit the silicon atoms at high speeds, thereby creating impact ionization.
  - New electrons and holes are generated, with the electrons absorbed by the drain and the holes by the substrate.
  - A finite drain substrate current appears.
  - If the carriers acquire a very high energy, they may be injected into the gate oxide and even flow out the gate terminal, introducing the gate current.
  - Supply voltage scaling becomes inevitable.

r_o Variation with V_{DS}





- As V_{DS} increases and the pinch off point moves toward the source, the rate at which the depletion region around the drain becomes wider decreases, resulting in a higher incremental output impedance.
- The W_{dep} is a strong function of the voltage of small $V_{reverse}$ and becomes weak with large $V_{reverse}$.
- ***Impact ionization*** limits the maximum gain that can be obtained from cascode structures because it introduces a small-signal resistance from ***the drain to the substrate*** rather than ***to the source***.

Outline

1. P-N Junction
2. Bipolar Junction Transistor
3. **MOS Field-Effect Transistor**
    - I/V characteristic
    - Second-order effect
    - Small-signal model
    - Scaling & Short-channel effects
    - Simulation models

Why Modeling?

- Analog circuits more sensitive to detailed transistor behavior
  - Precise currents, voltages, etc. matter
  - Digital circuits have much larger “margin of error”
- Models allow us to reason about circuits
  - Provide window into the physical device and process
  - “Experiments” with SPICE much easier to do

Levels of Abstraction

- Best abstraction depends on questions you want to answer
- Digital functionality:
  - MOSFET is a switch
- Digital performance:
  - MOSFET is a current source and a switch
- Analog characteristics:
  - MOSFET described by BSIM with 100's of parameters?
  - MOSFET described by measurement results?

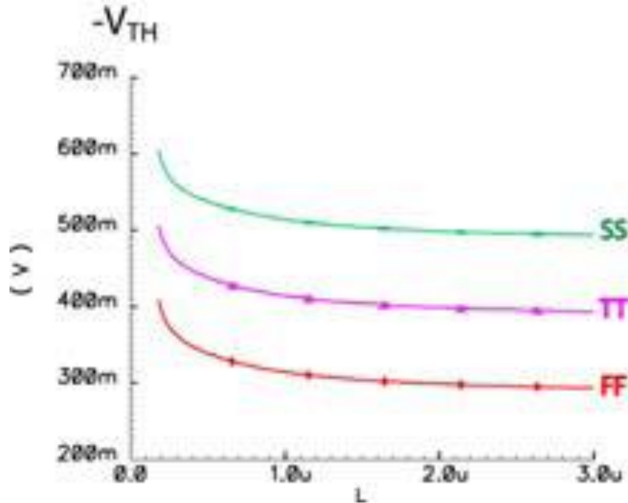
Device Corners

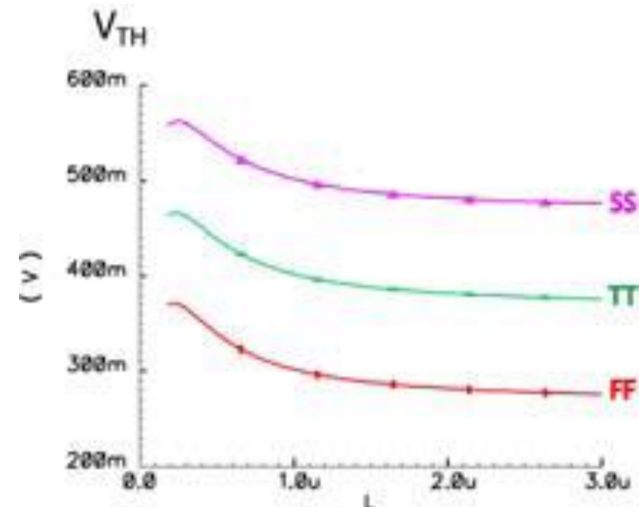
- Run-to-run parameter variations:
  - E.g. implant doses, layer thickness, dimensions
  - Affect V_{TH}, μ, C_{ox}, $R_{\square}$, …
  - How model in SPICE?
- Nominal / slow / fast parameters (tt, ss, ff)
  - E.g. fast: low V_{TH}, high μ, high C_{ox}, low $R_{\square}$
  - Combine with supply & temperature extremes
  - Pessimistic but numerically tractable

    → improves chances for working Silicon

Corner Example: V_{TH}

- Corners will shift V_{TH}
  - V_{TH} depends on channel length and process corner as well
- Variations probably bigger than reality too
  - Fab wants you to buy everything they make





Why not Square Law?

- Square law model most widely known:

$$I_{D,sat} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{th})^2$$

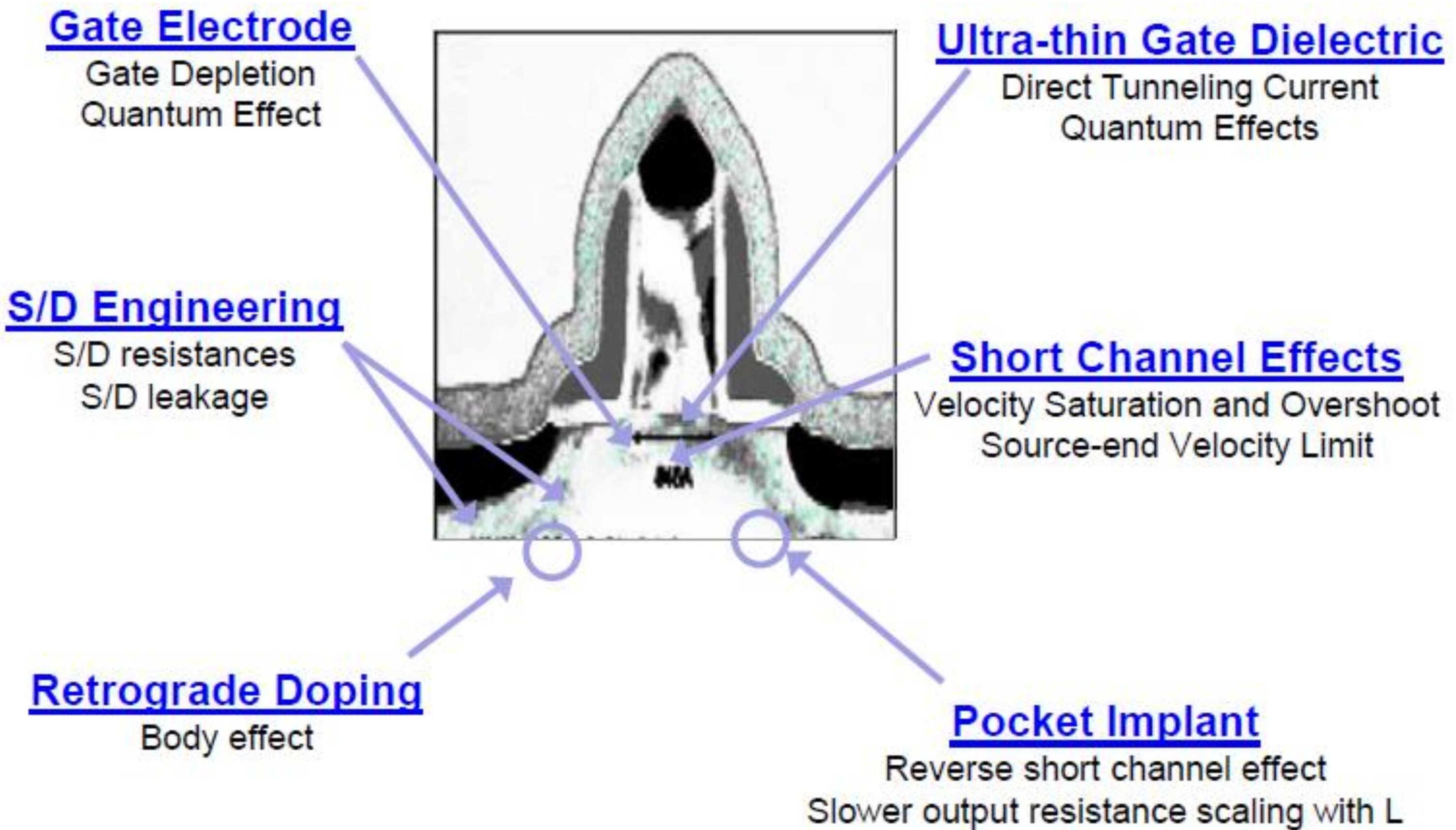
- But, totally inadequate for “short-channel” behavior

- Also doesn’t capture *moderate inversion*
  - (i.e., in between sub-threshold and strong inversion)

Square Law Model Assumptions

- Charge density determined only by vertical field
- Drift velocity set only by lateral field
- Neglect diffusion currents (“magic” V_{th})
- Constant mobility
- And many more…

A Real Transistor



Now What?

- Rely purely on simulator to tell us how devices behave?
  - Models not always based on real measurements
  - Model extraction is hard
  - Models inherently compromise accuracy for speed

- Need to know about important effects
  - So that know what to look for
  - Model might be wrong, or doesn’t automatically include some effects
    - E.g., gate leakage

Level I

- **Level 1** : Deployed during mid-1960s~late-1970s (Shichman and Hodges Model).

$$I_D = \frac{1}{2} K_P \frac{W}{L - 2L_D}\left[2(V_{GS} - V_{TH})V_{DS} - V_{DS}^2\right](1 + \lambda V_{DS})$$

$$I_D = \frac{1}{2} K_P \frac{W}{L - 2L_D}(V_{GS} - V_{TH})^2(1 + \lambda V_{DS})$$

Where $K_P = \mu C_{ox}$ and $V_{TH} = V_{TH0} + \gamma\left(\sqrt{2\varphi_B - V_{BS}} - \sqrt{2\varphi_B}\right)$

- Does not include subthreshold conduction or any short-channel effects.
- For the capacitance to change value continuously from one region to another

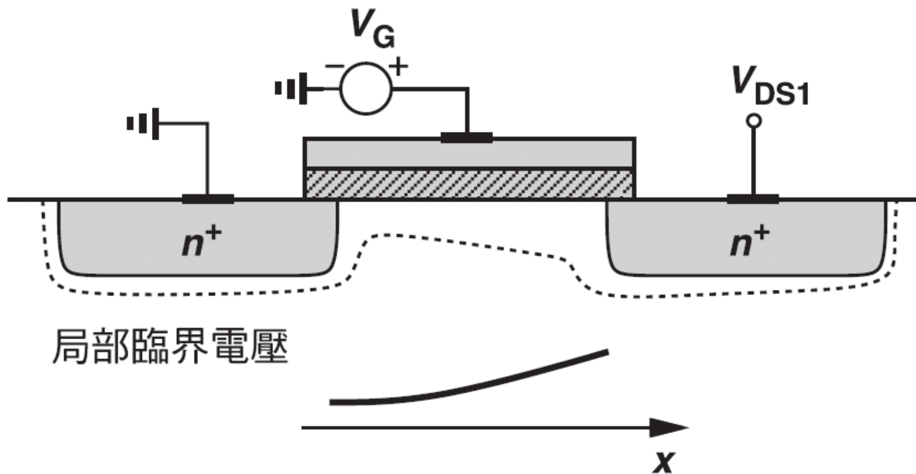
$$C_{GS} = \frac{2}{3} WLC_{ox}\left\{1 - \frac{(V_{GS} - V_{DS} - V_{TH})^2}{\left[2(V_{GS} - V_{TH}) - V_{DS}\right]^2}\right\} + WC_{ov}$$

$$C_{GB} = 0$$

$$C_{GD} = \frac{2}{3} WLC_{ox}\left\{1 - \frac{(V_{GS} - V_{TH})^2}{\left[2(V_{GS} - V_{TH}) - V_{DS}\right]^2}\right\} + WC_{ov}$$

- It predicts the output impedance of transistors in saturation quite poorly.

Level II



- L <= 4 μm, need level II for higher order effect.
- Not constant threshold.

- With a varying threshold voltage

$$I_D = \mu C_{ox}\frac{W}{L}\left\{(V_{GS}-V_{TH0})V_{DS}-\frac{V_{DS}^2}{2}-\frac{2}{3}\gamma\left[(V_{DS}-V_{BS}+2\varphi_F)^{3/2}-(-V_{BS}+2\varphi_F)^{3/2}\right]\right\}$$

- Even for $V_{BS} = 0$, I_D exhibits some dependence on γ.
- For small V_{DS}, the equation reduces to that of the Level 1 model, but for large V_{DS} the drain current is less than that predicted by the square law.

$$V_{D,sat} = V_{GS}-V_{TH0}-\phi_F+\gamma^2\left[1-\sqrt{1+\frac{2}{\gamma^2}\left(V_{GS}-V_{TH0}+\phi_F\right)}\right]$$

- In the saturation region, the drain current is $I_{DS} = I_{D,sat}\dfrac{1}{1-\lambda V_{DS}}$

Level II: Channel Length Modulation

- The depletion region of a PN junction

$$\Delta L = \sqrt{\frac{2\varepsilon_{si}}{qN_{sub}}\left[\varphi_B + (V_{DS} - V_{D,sat})\right]}$$

  - $V_{D,sat}$: pinch-off voltage, I_D is discontinuous at the edge of the triode region.
- A fixed-up equation

$$\Delta L = \sqrt{\frac{2\varepsilon_{si}}{qN_{sub}}\left(V_1 + \sqrt{1 + V_1^2}\right)} \qquad V_1 = \frac{V_{DS} - V_{D,sat}}{4} \qquad \lambda = \frac{\Delta L}{LV_{DS}}$$

  - The output conductance of the transistor varies as V_{DS} increases.
- The degradation of the mobility with the vertical field in the channel

$$\mu_s = \mu_0 \left(\frac{\varepsilon_{si}}{C_{ox}} \cdot \frac{U_c}{V_{GS} - V_{TH} - U_t V_{DS}}\right)^{U_e}$$

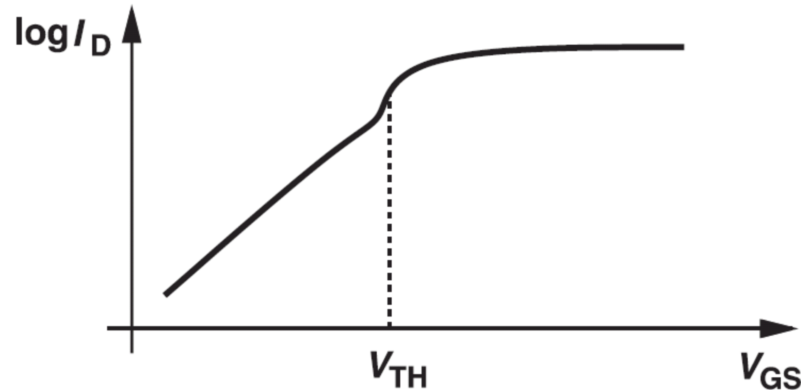
U_c:gate-channel critical electric field, U_t: A fitting parameter (~0-0.5), U_e: (~0.15)

- The subthreshold behavior defines a voltage V_{on}

$$V_{on} = V_{TH} + \zeta V_T \qquad \zeta = 1 + \frac{qN_{FS}}{C_{ox}} + \frac{C_d}{C_{ox}} \qquad N_{FS} \text{ is an empirical constant}$$

$$I_{DS} = I_{on} \exp\frac{V_{GS} - V_{on}}{\zeta V_T} \qquad I_{on} = I_{DS} \mid V_{GS} = V_{on} \text{ in strong inversion}$$

Level II: Kink in I_D



$$I_{DS} = I_{on} \exp \frac{V_{GS} - V_{on}}{\zeta V_T}$$

$$I_{on} = I_{DS} | V_{GS} = V_{on} \quad \text{in strong inversion}$$

- I_{on} is the drain current calculated in strong inversion for $V_{GS} = V_{on}$.
- The discontinuity in the slope of I_D from the subthreshold to strong inversion.
- Provide reasonable I/V accuracy for wide/short devices in the saturation region with L ~ 0.7 μm.
- Represent two short channel phenomena
  - (1)The variation of V_{TH} with L. (2) Velocity saturation
- Suffers from substantial error in representing the output impedance and the transition point between saturation and triode regions.
- For narrow or long devices, the model is quite inaccurate.

Level III

- Many empirical constants introduced to improve the accuracy for channel lengths as small as 1 μm.
- The threshold voltage is expressed as

$$V_{TH} = V_{TH0} + F_s \gamma \sqrt{2\varphi_F - V_{BS}} + F_n (2\varphi_F - V_{BS}) + \xi \frac{8.15 \times 10^{-22}}{C_{ox} L_{eff}^3} V_{DS}$$

  – where F_s and F_n represent short channel and narrow channel effects, and ξ models drain induced barrier lowering.
- The mobility equation involves both vertical and lateral field effects:

$$\mu_1 = \frac{\mu_{eff}}{1 + \dfrac{\mu_{eff} V_{DS}}{v_{max} L_1}} \qquad \text{where} \qquad \mu_{eff} = \frac{\mu_0}{1 + \theta (V_{GS} - V_{TH})}$$

v_{max} denotes the maximum velocity of carriers in the channel

- μ_{eff} models the effect of the vertical field, μ_1 adds that of the lateral field as well.

Level III

- The drain current is realized as

$$I_D = \mu_1 C_{ox} \frac{W_{eff}}{L_{eff}} \left[V_{GS} - V_{TH0} - \left(1 + \frac{F_s \gamma}{4\sqrt{2\varphi_F - V_{BS}}} + F_n \right) \frac{V_{DS}^{'}}{2} \right] V_{DS}^{'}$$

  - where V_{DS}' = $V_{D,sat}$ if the device is in saturation.

- Exhibits moderate accuracy for wide, short channels but suffers from large errors for longer channels.
- Drawback: Discontinuity of the derivative of I_D with respect to V_{DS} at the edge of the triode region.

- Poorly model of r_o for a short-channel device.

BSIM (Berkeley Short-channel IGFET Model)

- It became more difficult to introduce physically meaningful equations that would be both accurate and computationally efficient.
- BSIM adopted numerous (~50) empirical parameters to simply the equations
  - At the cost of losing touch with the actual device operation
  - Represent the geometry dependence of many of the device parameters

$$\text{general form: } P = P_0 + \frac{\alpha_P}{L_{eff}} + \frac{\beta_P}{W_{eff}} \qquad \text{mobility: } \mu = \mu_0 + \frac{\alpha_u}{L_{eff}} + \frac{\beta_u}{W_{eff}}$$

  - where P_0 is the value of the parameter for a long, wide transistor, and α_p and β_p are fitting factors.
  - Become less accurate at small dimension.

BSIM, BSIM2, BSIM3

- Improvement
  - The dependence of mobility upon the vertical field includes the substrate voltage
  - The threshold voltage is modified for substrates with nonuniform doping.
  - The currents in the weak and strong inversion regions are derived such that their values and first derivatives are continuous.
  - New equations are devised for velocity saturation, dependence of mobility upon lateral field, and the saturation voltage.
- Its accuracy for narrow, short transistors is somewhat poor.
- $L < 0.25\mu m$, BSIM3, has returned to the physical principles of device operation while maintaining many of the useful features of BSIM and BSIM2.
  - Still suffering from large errors in predicting the output impedance.

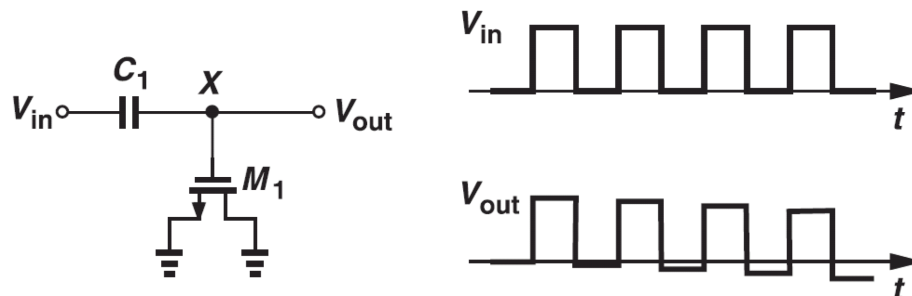
Other Models

- HSPICE Level 28
  - Improves the dependence of accuracy upon device dimensions by expressing the parameters as

$$P = P_0 + \alpha\left(\frac{1}{L} - \frac{1}{L_{ref}}\right) + \beta\left(\frac{1}{W} - \frac{1}{W_{ref}}\right) + \gamma\left(\frac{1}{L} - \frac{1}{L_{ref}}\right)\left(\frac{1}{W} - \frac{1}{W_{ref}}\right)$$

  - where L_{ref} and W_{ref} denote the dimensions of a measured reference.
  - The dependence is expressed in terms of "***increments***" with respect to characterized transistors than the absolute value of dimensions.
  - Proportional to the product of the L and W increments helps curve fitting.
- MOS 9
- Enz-Krummenacher-Vittoz (EKV) model
  - Considering the bulk rather than the source as the reference point.
  - Introduce a single drain-source current that is both valid for subthreshold and saturation region.

Charge and Capacitance Modeling



- Level 1 gate capacitance model does not conserve charge
  - Integrating voltages respect to time, small accumulated errors.
  - Droop at the output because in every period some charge at node X is lost.
- The following assumptions are inaccurate for short channel devices
  - In the triode region

$$C_{GS} = C_{GD} = \frac{1}{2} WLC_{ox} + WC_{ov}$$

  - In the saturation region

$$C_{GS} = \frac{2}{3} WLC_{ox} + WC_{ov} \qquad C_{GD} = WC_{ov}$$

  - Require flexible partitioning for ease of curve fitting (40%/60%, 50%/50%, and 0%/100% for BSIM and BSIM3)

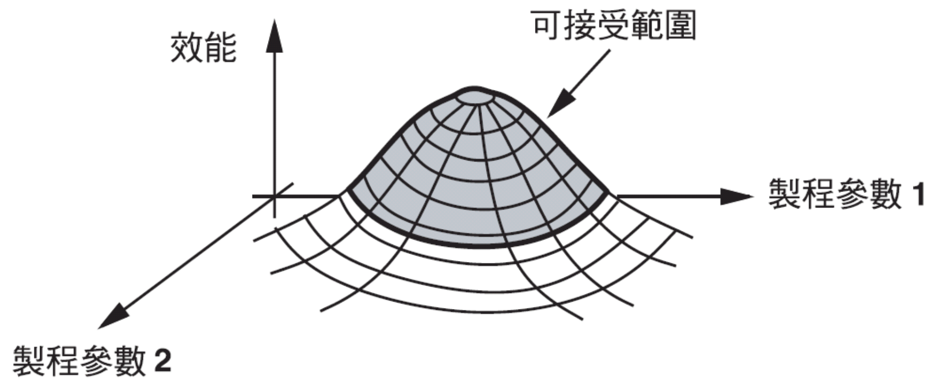
Temperature Dependence

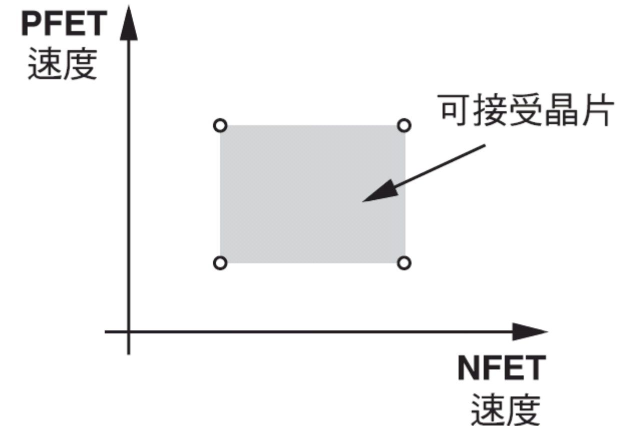
- In Level 1-3 models as well as BSIM and BSIM2, the following parameters have temperature dependence
  - V_{TH}
  - Built-in potential of S/D junctions
  - The intrinsic carrier concentration of silicon (n_i)
  - The bandgap energy (E_G)
  - The mobility
- Most equations are empirical

$$E_g = 1.16 - \frac{7.02 \times 10^{-4} T^2}{T + 1108} \qquad \mu = \mu_0 \left(\frac{300}{T} \right)^{3/2}$$

  - where $\mu_o = \mu\ (T = 300^o\ K)$
- BSIM 3 incorporates a few more parameters to represent the T dependence
  - Velocity saturation
  - The effect of subthreshold voltage on V_{TH}.

Process Corners





- MOSFETs suffer from substantial parameter variations from wafer to wafer and from lot to lot.
- Transistors having a thinner gate oxide and lower threshold voltage fall near the fast corner.
- Four corners (FF, FS, SF, SS)
- Simulation of circuits for various process corners and temperature extremes is essential to determining the yield.

Reference

1. Analog Integrated Circuits, Spring 2010, by Prof. WZ Chen, NCTU
2. Analog IC Analysis and Design, Spring 2010, by Prof. PC Huang, NTHU
3. Advanced Analog IC Design, Spring 2009, by Prof. Elad Alon, Berkeley
4. Advanced Analog IC Design, Fall 2009, by Prof. Y. Chiu, UIUC