Chapter 2
MOS Device Physics and Models

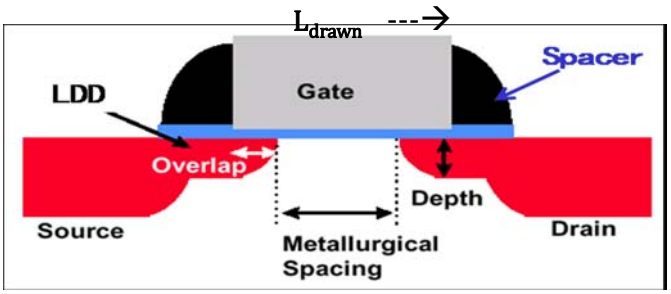
Reading Assignment

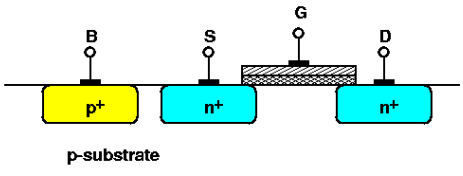
- Razavi, 2017
  - Chapter 2 : Basic MOS device physics
  - Chapter 17 : Short-channel effects and device models

Chapter 2
MOS Device Physics and Models

2.1 General Considerations

2.1.1 N-MOSFET Structure





$L_{eff} = L_{drawn} - 2L_D$ For $0.18\ \mu m$ technology :

$L_{eff} \approx 0.148\ \mu m$ (nMOS)

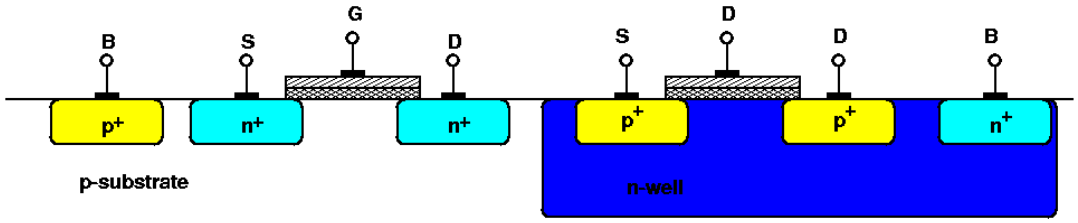
$0.16\ \mu m$ (pMOS) $t_{ox} \approx 42\ \text{Å}$

- L_{eff} is the effective length, L_{drawn} is the total drawn (layout) length. L_D is the amount of lateral (side diffusion. t_{ox} is the oxide thickness.
- In typical operations, the source /drain junction diodes must be reverse-biased or zero-biased.
- In a general case, the p-substrate must be connected to the most negative supply in the system.

Chapter 2
MOS Device Physics and Models

2.1 General Considerations

2.1.2 Complementary MOS Technology



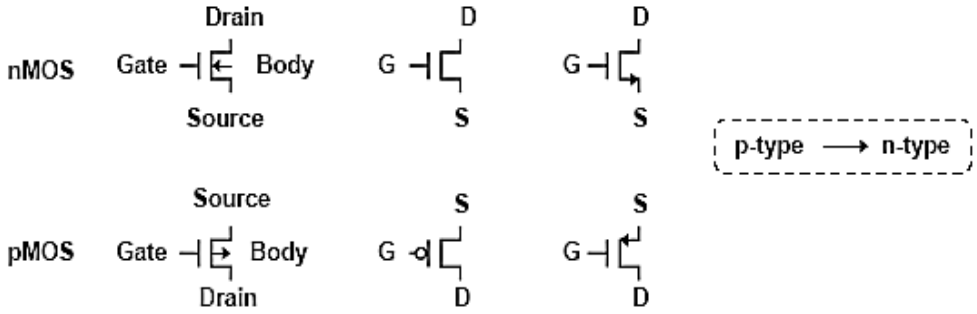
- The PMOS device is fabricated in an n-well.
- The n-well is tied to a potential such that the S/D junction diodes of the PMOS transistor remain reverse-biased under all conditions.
  (In most circuits, the most positive supply voltage)

Chapter 2
MOS Device Physics and Models

2.1 General Considerations

2.1.3 MOS Transistors

- MOS Symbols



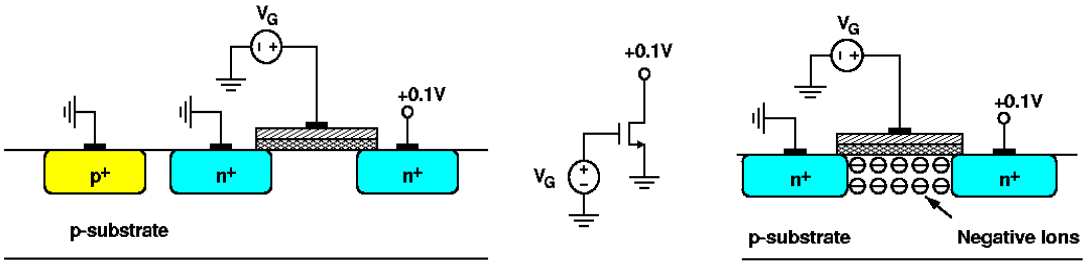
- In nMOS (pMOS), the n^+ (p^+) region with more positive (negative) voltage is defined as the drain. The other one is as the source. The drain/source is not fixed to a region. It is defined by its voltage.

Chapter 2
MOS Device Physics and Models

2.2 MOS I/V Characteristics

2.2.1 Threshold Voltage

Formation of depletion region



Onset of Inversion

Formation of inversion layer

Chapter 2
MOS Device Physics and Models

2.2 MOS I/V Characteristics

2.2.1 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} : electron (hole) concentration in the inversion layer is the same as that in the p-substrate (n-well).
- For enhancement-mode devices, $V_{THn} > 0$, and $V_{THp} < 0$.

Chapter 2
MOS Device Physics and Models

2.2 MOS I/V Characteristics

2.2.1 Threshold Voltage

$$V_{TH} = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep}}{C_{ox}} + \frac{Q_{SS}}{C_{ox}} \qquad C_{ox}: \textit{oxide capacitance per unit area}$$

$$= \Phi_{MS} + 2\Phi_F + \frac{Q_{dep0}}{C_{ox}} + \frac{Q_{SS}}{C_{ox}} + \frac{Q_{dep} - Q_{dep0}}{C_{ox}}$$

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

where γ is the body effect factor

- Φ_{MS} is the work function difference between the gate metal and the silicon interface.
- Φ_F is the Fermi level, when the potential in silicon reaches $2\Phi_F$, inversion occurs.
- Q_{dep} is the charge stored in the depletion region with substrate bias V_{SB}.

$$Q_{dep} = \sqrt{2qN_A\epsilon_{si}(2\Phi_F + V_{SB})}$$

Q_{dep0} is the charge stored in the depletion region with substrate bias $V_{SB} = 0$.

$$Q_{dep0} = \sqrt{2qN_A\epsilon_{si}2\Phi_F}$$

- Q_{ss} is the positive charge density in the oxide at the silicon interface.
- $$\gamma = \frac{1}{C_{ox}}\sqrt{2q\epsilon_{si}N_{sub}}$$
- $$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$$

ε_{si} : dielectric constant of silicon

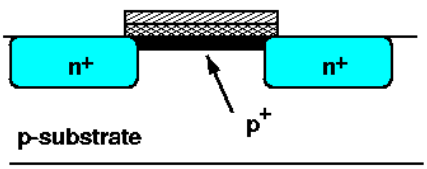
For $t_{ox} \approx 42\,\mathring{A}$ $\quad C_{ox} \approx 8.214\,fF/\mu m^2$

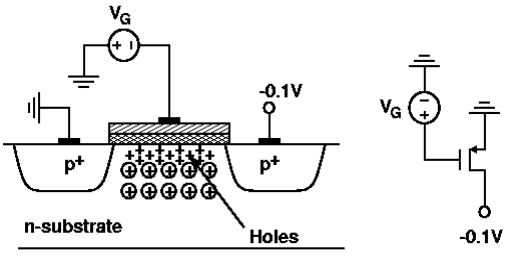
Chapter 2
MOS Device Physics and Models

2.2 MOS I/V Characteristics

2.2.1 Threshold Voltage - Adjustment

- Implantation of p^+ dopants to alter the threshold.
- If a thin sheet of p^+ is created, the gate voltage required to deplete the region increases. (N_A increases → Q_{dep} increases)
- The turn-on phenomenon in a PMOS device is similar to that of nMOS device, but with all of the polarities reversed.
- Formation of inversion layer in a pMOS device.





Chapter 2
MOS Device Physics and Models

2.2 MOS I/V Characteristics

2.2.2 Derivation of I/V Characteristics

Fig. 2.9 (a) A semiconductor bar carrying a current *I*. (b) Snapshots of the the carriers one second apart.

Fig. 2.10(a) Channel charge with equal source and drain voltages.

$I = Q_d \cdot v$

- Q_d (coulombs/m) : Charge density per meter length along the direction of current.
- v (m/sec) : velocity of the charge. $v = \mu E$ $E(x) = -dV/dx$ μ: carrier mobility
- 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 $Q_d(x) = WC_{ox}(V_{GS} - V(x) - V_{TH})$

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

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

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

Chapter 2
MOS Device Physics and Models

2.2 MOS I/V Characteristics

2.2.2 Derivation of I/V Characteristics

- Channel charge with unequal source and drain voltage

Fig. 2.11 Drain current versus drain-source voltage in the triode region.

Fig. 2.10(a) Channel charge with unequal source and drain voltages.

- The peak current of I_D can be found by calculating

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

- V_{GS}-V_{TH} : Overdrive voltage
- W/L : Aspect ratio
- If $V_{DS} \le V_{GS} - V_{TH}$

  the device is operated in the triode (linear) region

2.2 MOS I/V Characteristics

2.2.2 Derivation of I/V Characteristics - Linear Operation in the 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

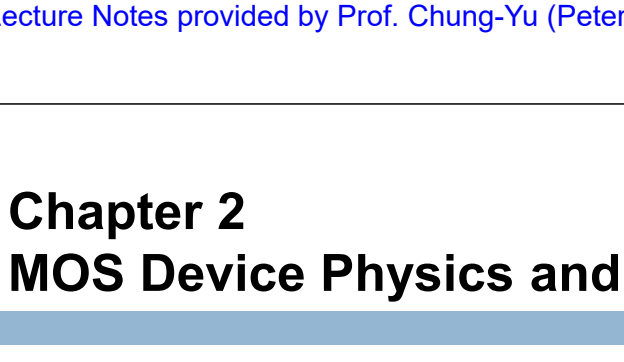


Fig. 2.12 Linear operation in deep triode region.

- Formation of inversion layer

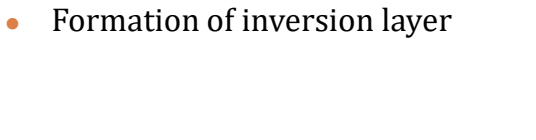


Fig. 2.13 MOSFET as a controlled linear resistor.

2.2 MOS I/V Characteristics

2.2.2 Derivation of I/V Characteristics - Linear Operation in the Deep Triode Region

Example 2.1:

Fig. 2.14

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

$$R_{on} = \infty$$

- For $V_G > 1\ V + V_{TH} = 1.7$ V, M1 is in the deep triode region.

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

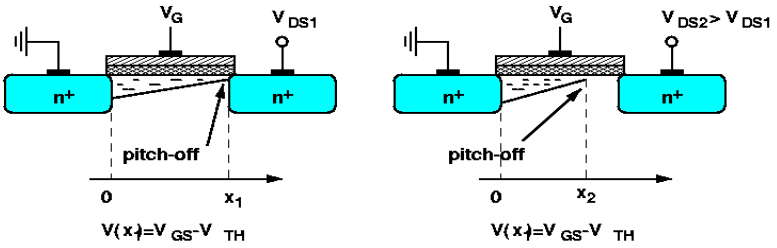
2.2 MOS I/V Characteristics

2.2.2 Derivation of I/V Characteristics - Saturation of Drain Current

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

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

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



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

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

Chapter 2
MOS Device Physics and Models

2.2 MOS I/V Characteristics

2.2.2 Derivation of I/V Characteristics - MOS I-V Relation

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

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

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

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

- Saturated MOSFETs operating as current source

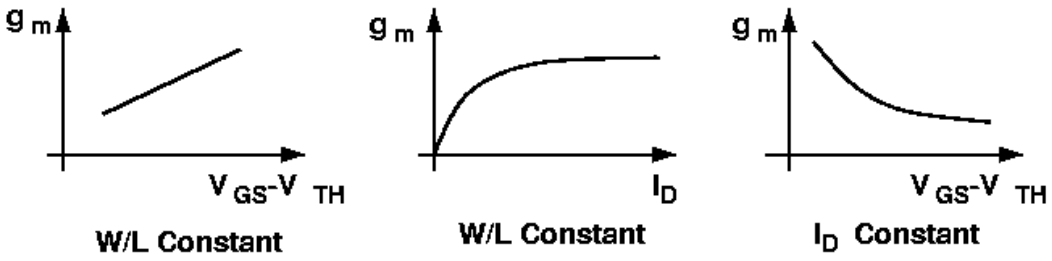
Fig. 2.17 Saturated MOSFETs operating as current sources.

2.2 MOS I/V Characteristics

2.2.2 Derivation of I/V Characteristics - Transconductance

- MOS in the saturation region.

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



Chapter 2
MOS Device Physics and Models

2.2 MOS I/V Characteristics

2.2.2 Derivation of I/V Characteristics - Transconductance

Example 2.3:

Fig. 2.21

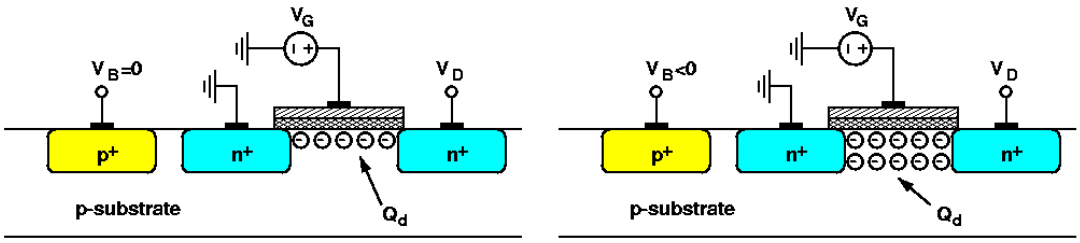
- Plot the transconductance as a function of V_{DS}
- M1 is operated in saturation region for $V_{DS} \geq (V_b - V_{TH})$

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

- M1 is operated in triode region for $V_{DS} < (V_b - V_{TH})$

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

2.3 Second-Order Effect – Body Effect (Back-Gate Effect)



$$V_{TH} = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep}}{C_{ox}} + \frac{Q_{ss}}{C_{ox}} \qquad V_{TH} = V_{TH0} + \gamma\left(\sqrt{2\Phi_F + V_{SB}} - \sqrt{2\Phi_F}\right)$$

$$\gamma = \sqrt{2q\varepsilon_{si}N_{sub}}\,/\,C_{ox}$$

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

2.3 Second-Order Effect – Body Effect (Back Gate Effect)

Example 2.4:

Fig. 2.24

- Plot the drain current if V_X varies from -∞ to 0. Assume $V_{TH0} = 0.6$ V, $\gamma = 0.4\ V^{0.5}$, and $2\Phi_F = 0.7$ V
- If V_X is sufficiently negative, the threshold voltage of M1 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 $-4.76\,V < V_{X1} < 0$

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

Chapter 2
MOS Device Physics and Models

2.3 Second-Order Effect – Body Effect (Back Gate Effect)

Example : Body Effect on source follower

Fig. 2.25 (a) A circuit in which the source-bulk voltage varies with input level; (b) input and output voltages with no body effect; (c) input and output voltages with body effect.

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

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

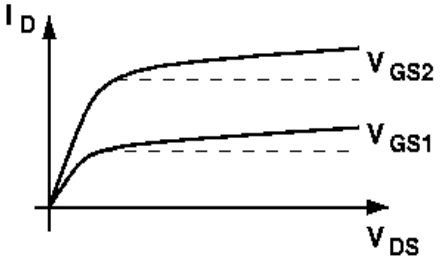
Chapter 2
MOS Device Physics and Models

2.3 Second-Order Effect – Channel-Length 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} .

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

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



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

- 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} .

Chapter 2
MOS Device Physics and Models

2.3 Second-Order Effect – Channel-Length Modulation Effect

Example 2.7

Fig. 2.7 Effect of doubling channel length.

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

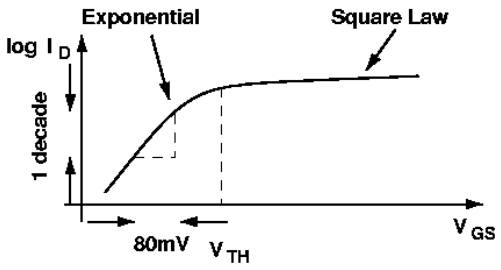
$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS}-V_{TH})^2(1+\lambda V_{DS}) \qquad \lambda \propto \frac{1}{L} \qquad \frac{\partial I_D}{\partial V_{DS}} \propto \frac{\lambda}{L} \propto \frac{1}{L^2}$$

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

Chapter 2
MOS Device Physics and Models

2.3 Second-Order Effect – Subthreshold Conduction

- **Weak Inversion** : For $V_{GS} \approx V_{TH}$, a weak inversion layer exists and some current flow from D to S.
- **Subthreshold Conduction** : For $V_{GS} < V_{TH}$, I_D is finite, but 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(zeta)≈1.5 at room temperature, V_{GS} must decrease by approximately 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 $10^{300/80} = 10^{3.75}$ when V_{GS} is reduced to zero.
  - In rough calculations, we often view V_{TH} as the gate source voltage yielding $I_D/W = 1\mu A/\mu m$-width
  - The transconductance of MOSFET in subthreshold region is $g_m = \frac{I_D}{\zeta V_T}$, which is inferior to that of bipolar transistors.



2.3 Second-Order Effect – Voltage Breakdown and Punch Through

- **Gate-Source (Drain) breakdown** : At high gate-source (drain) voltages, the gate oxide breaks down irreversibly, damaging the transistor.
- **Drain-Source breakdown**

1. 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. →Steep (nearly vertical) $I_D - V_{DS}$ characteristics.

2. **Drain junction breakdown or BJT-like breakdown:**

Drain-substrate junction breakdown before punch through. drain-substrate-source BJT (lateral bipolar junction transistor) breakdown before punch through. → Soft rising $I_D - V_{DS}$ characteristics.

2.4 MOS Device Models

2.4.1 MOS Device Layout

Example 2.9

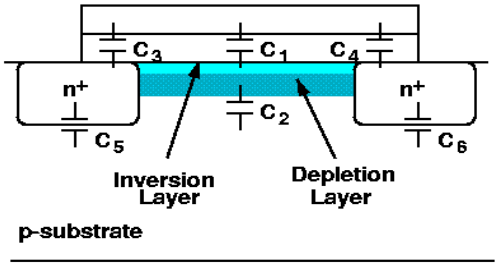
Fig. 2.30

- 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.

Chapter 2
MOS Device Physics and Models

2.4 MOS Device Models

2.4.2 MOS Device Capacitance



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

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

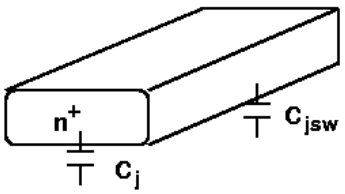
- C_3, C_4 : Overlap capacitance of the gate poly with the source and drain areas. Owing to fringing effect, they can be obtained by more elaborate calculations. C_{ov} is the overlap capacitance per unit width. $C_{GDov} = C_{GSov} = C_{ov}W \approx WL_D C_{ox}$

Chapter 2
MOS Device Physics and Models

2.4 MOS Device Models

2.4.2 MOS Device Capacitance

- Junction capacitance between the source/drain and the substrate.
  - C_j (F/m^2) :
    bottom plate capacitance associated with the bottom of the junction.
    $C_j = C_{j0}/[1+V_R/\Phi_B]^{mj}$
    where CJ0, PHI, and MJ are HSPICE parameters.
  - C_{jsw} (F/m) :
    side-wall capacitance due to the perimeter of the junction.
    $C_{jsw} = C_{j0sw}/[1+V_R/\Phi_B]^{mjsw}$
    where CJSW0, PHI and MJSW are HSPICE parameters.



Chapter 2
MOS Device Physics and Models

2.4 MOS Device Models

2.4.2 MOS Device Capacitance

Example

Fig. 2.33

- For the transistor in (a)
  - $C_{DB} = C_{SB} = WEC_j + (2W + 2E)C_{jsw}$
- For the transistor in (b), smaller C_{DB} and larger C_{SB}. C_{SB} is not important since S and B are usually shorted.

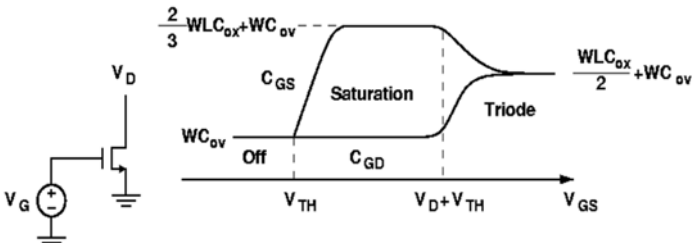
$$C_{DB} = \frac{W}{2}EC_j + 2(\frac{W}{2} + E)C_{jsw} \qquad C_{SB} = 2\left[\frac{W}{2}EC_j + 2\left(\frac{W}{2} + E\right)C_{jsw}\right] = WEC_j + 2(W + 2E)C_{jsw}$$

Chapter 2
MOS Device Physics and Models

2.4 MOS Device Models

2.4.2 MOS Device Capacitance - C_{GS} and C_{GD} vs. V_{GS}

- If the device is off
  - $C_{GD} = C_{GS} = C_{ov}W$
  - C_{GB} consists of overlap capacitance $C_{GB,ov}$ and the series combination of the gate oxide capacitance and the depletion region capacitance .



$$C_{GB} = \frac{(WL_{eff}C_{ox})C_d}{WL_{eff}C_{ox} + C_d} + C_{GB,ov} \qquad C_d = WL_{eff}\sqrt{q\varepsilon_{si}N_{sub}/4\Phi_F}$$

- If the device is in the deep triode region : $C_{GB} = C_{GB,ov}$

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

- If the device is in the saturation region : $C_{GB} = C_{GB,ov}$
  - The potential difference between the gate and channel varies from V_{GS} at the source to V_{GS}-V_{TH} at the pinch-off point.

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

- C_{GB} : gate to substrate (bulk) capacitance.

Chapter 2
MOS Device Physics and Models

2.4 MOS Device Models

2.4.2 MOS Device Capacitance - C_{GS} and C_{GD} vs. V_{GS}

Example

Fig. 2.35 **Fig. 2.36**

- 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 $V_X \sim 0$, M_1 is in the triode region

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

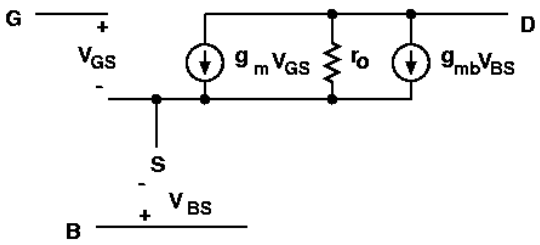
- For $V_X > 1V$, the role of the source and drain is exchanged
- M_1 is out of triode region for

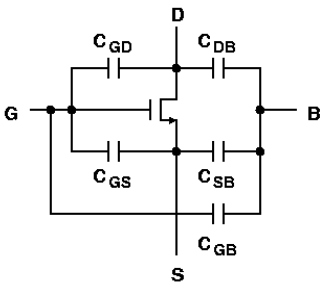
$$V_X \geq 2V - 0.6V$$

Chapter 2
MOS Device Physics and Models

2.4 MOS Device Models

2.4.3 MOSFET Small-Signal Model





- $g_{mb}V_{BS}$ has the same (opposite) polarity with that of g_mV_{GS} if $V_{BS} > 0$ (<0), forward (reverse biased) B-S junction.

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

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

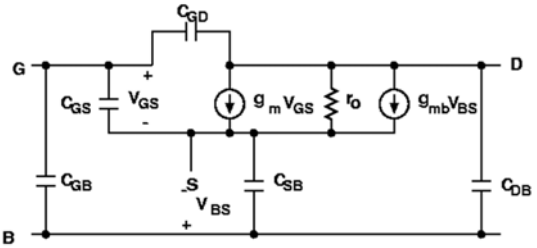
where $\frac{\partial V_{TH}}{\partial V_{BS}} = -\frac{\partial V_{TH}}{\partial V_{SB}} = -\frac{\gamma}{2}(2\Phi_F + V_{SB})^{-1/2}$ $\qquad \eta = \frac{g_{mb}}{g_m} \approx 0.1 - 0.3$

Chapter 2
MOS Device Physics and Models

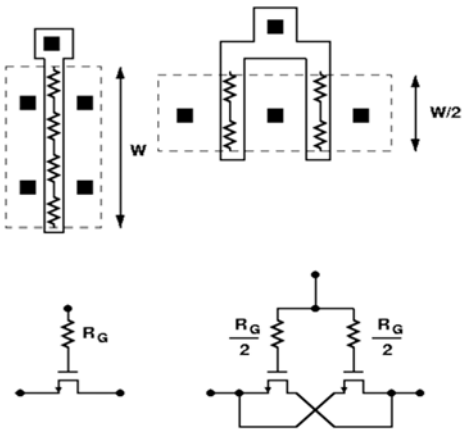
2.4 MOS Device Models

2.4.3 MOSFET Small-Signal Model

- Complete MOS small-signal model



- Reduction of gate resistance by folding



Chapter 2
MOS Device Physics and Models

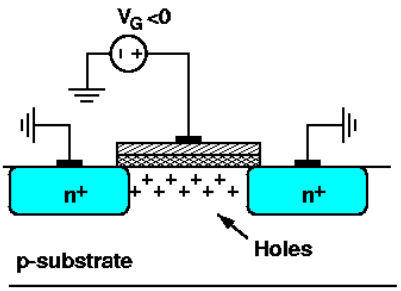
2.4 MOS Device Models

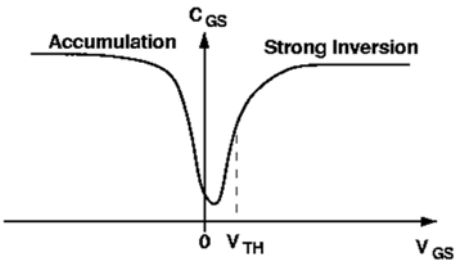
2.4.4 NMOS vs. PMOS Devices

- $\mu_p C_{ox} \approx 0.25\, \mu_n C_{ox}$
- For given dimensions and bias currents, NMOS transistors exhibit a higher output resistance because of a smaller .

2.4.5 MOS Device as a Capacitor

- NMOS device as a capacitor with S, D, and B connected together.
  - **Accumulation** ($V_{GS} < 0$, large negative voltage)
    $C_{GS} + C_{GD} + C_{GB} = C_{ox} W L_{eff} + 2WC_{ov} + C_{GB,ov}$
  - **Weak inversion** ($V_{GS} < V_{TH}$), a depletion region begins to form under the oxide.
    $C_{GS} + C_{GD} + C_{GB} = C_{ox} W L_{eff} C_d/(C_{ox} W L_{eff} + C_d) + 2WC_{ov} + C_{GB,ov}$
  - **Strong Inversion** ($V_{GS} > V_{TH}$, $V_{DS} = 0$), the oxide-silicon interface sustains a channel. (Deep triode region)
    $C_{GS} + C_{GD} + C_{GB} = C_{ox} W L_{eff} + 2WC_{ov} + C_{GB,ov}$





Chapter 2
MOS Device Physics and Models

2.5 Scaling Theory

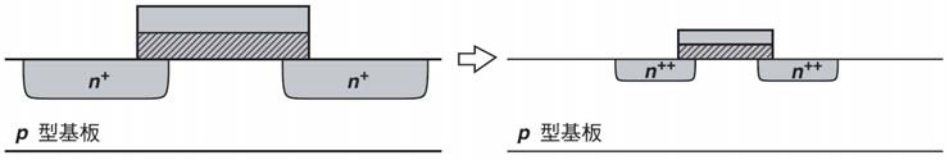
- The square-law characteristic provide accuracies for devices with minimum channel lengths of greater than 4 μm.
- The two principal reasons for the dominance of CMOS technology in today's semiconductor industry
  - Zero static power dissipation
  - Scalability
- Scaling theory follows three rules: $\alpha>1$

  1) Reduce all lateral and vertical dimensions by α
     - W, L, t_{ox}, depth and perimeter of the source /drain junctions
  2) Reduce the threshold voltage V_{TH} and supply voltage V_{DD} by α.

     →Constant field scaling : dimensions and voltage are scaled together
  3) Increase all the doping level by α

Chapter 2
MOS Device Physics and Models

2.5 Scaling Theory – Ideal Scaling of MOS Transistor



- The current capability of the transistor drops by a factor of α

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

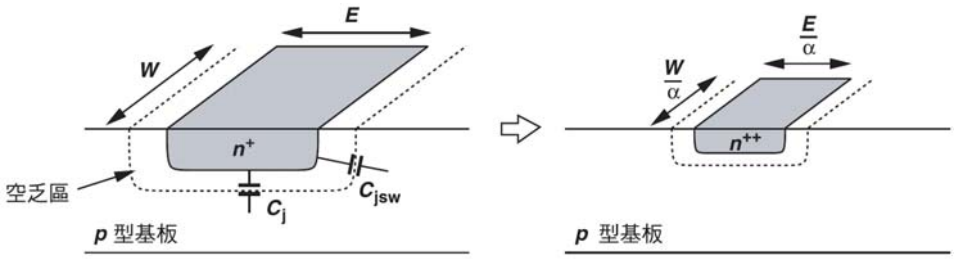
- Reduction of capacitances and power dissipation
  - The channel capacitance **drops** by a factor of α

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

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

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

2.5 Scaling Theory – 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 because the junction depth D_{depth} is reduced by α.

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

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

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

2.5 Scaling Theory – Scaling of the 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

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

Fig. 17.3 CMOS inverter.

- The dynamic power dissipation

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

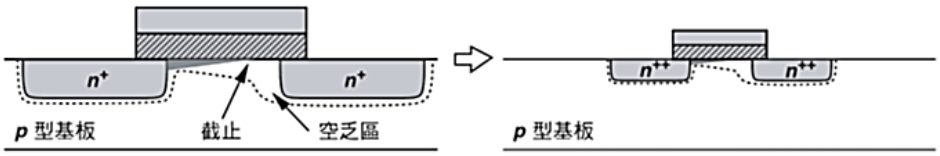
- The layout density (the number of transistors per unit area)

$$Density = \alpha^2$$

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

Chapter 2
MOS Device Physics and Models

2.5 Scaling Theory – Effect of Ideal Scaling in Analog Circuits



- If all of the dimensions and voltages (and currents) are scaled down, (Ideal scaling)

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

- If dimensions scale down while overdrive voltage remains constant (modified law)

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

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

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

Chapter 2
MOS Device Physics and Models

2.5 Scaling Theory – Greatest Impact of Scaling on Analog Circuit

- Reduction of the supply voltage
  - Maximum allowable voltage swings decreases by α
  - g_m constant and hence input-referred thermal noise power remain constant ($4kT\gamma g_m^{-1}$).
    - The lower end of the dynamic range is limited by thermal noise.
    - Dynamic range=Maximum allowable voltage/Noise voltage.
  - The dynamic range is decreased by α. (Disadvantage)
  - If both the voltage and current are scaled by α, the power is decreased by α^2 (advantage):

$$\frac{V_{DD}}{\alpha}\frac{I_{DD}}{\alpha} = \left(\frac{V_{DD}I_{DD}}{\alpha^2}\right)$$

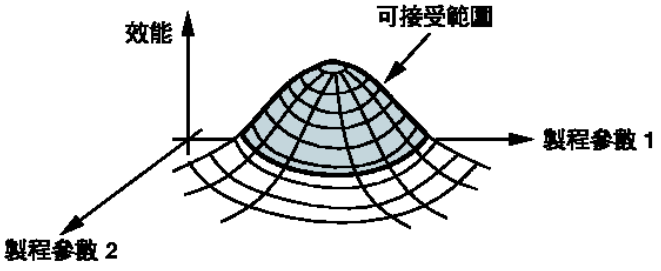
2.5 Scaling Theory – Greatest Impact of Scaling on Analog Circuit

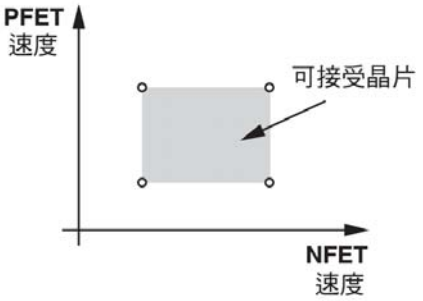
- To restore the dynamic range
  - The transconductance of the transistors must be increased by α^2
  - The thermal noise voltages and currents are scaled with $\sqrt{1/g_m}$

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

  - V_{ov} decreases by a factor of α, I_D must increase by a factor of α, and power dissipation remains 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 analog circuits requires a **constant power dissipation** and a **higher device capacitance** (αW).

2.6 Process Corners



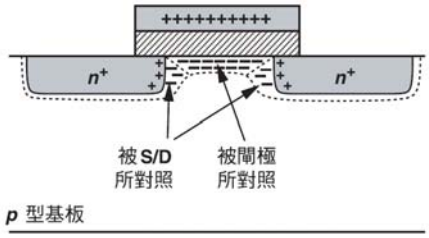


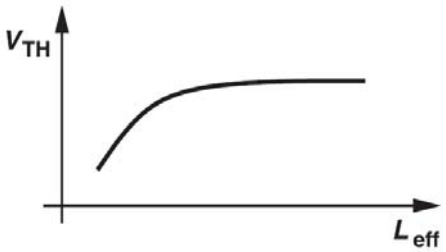
- Unlike bipolar transistors, MOSFETs suffer from substantial parameter variations from wafer to wafer and from lot to lot.
- It is common in today's CMOS technologies to obtain a gate delay that varies by a factor of two to one with process and temperature.
- Transistors having a thinner gate oxide 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 determine the yield.

Chapter 2
MOS Device Physics and Models

2.7 Short Channel Effect

2.7.1 Threshold Voltage Variation





- The threshold voltage exhibits a temperature coefficient of approximately -1mV/°K, yielding a 50-mV change across the commercial temperature range (0-50°C).
- Process induced variation is also in the vicinity of 50 mV, raising the margin to approximately 100 mV.
- It is difficult to reduce V_{TH} below several hundred millivolts.

Chapter 2
MOS Device Physics and Models

2.7 Short Channel Effect

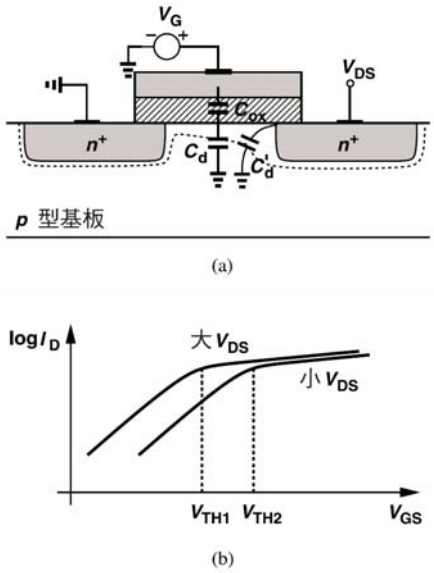
2.7.1 Threshold Voltage Variation

- Transistors fabricated on the same wafer but with different lengths yield lower V_{TH} as L decreases.
  - The depletion regions associated with the source and drain junctions protrude into the channel are considerably, thereby reducing the immobile charge that must be imaged by the charge on the gate.
  - Part of the immobile charge in the substrate is now imaged by the charge inside the source and drain areas rather than by the charge on the gate.
  - Gate voltage required to create an inversion layer decreases.
  - Length variation also introduces additional variations in V_{TH}.
- If the length of a device is increased so as to achieve a higher output impedance, then the threshold voltage also increases by as much as 100 to 200 mV.

2.7 Short Channel Effect

2.7.1 Threshold Voltage Variation

Drain Induced Barrier Lowering (DIBL) in a Short Channel Device



- 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 charges and hence the threshold voltage are decreased.
- The principal impact of DIBL on circuit design is the degraded output impedance.

Chapter 2
MOS Device Physics and Models

2.7 Short Channel Effect

2.7.2 Mobility Degradation with Vertical Field

- High electric field developed 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 vertical filed affects the device transconductance as well.

2.7 Short Channel Effect

2.7.3 Velocity Saturation

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

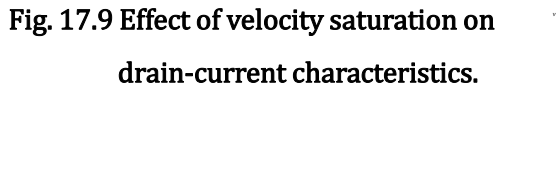


Fig. 17.9 Effect of velocity saturation on drain-current characteristics.

- The mobility of carriers also depends on the lateral electric field in the channel, beginning to drop as the field reaches levels of 1V/μm.
- υ approaches a saturated value about 10^7 cm/s.

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

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

Chapter 2
MOS Device Physics and Models

2.7 Short Channel Effect

2.7.3 Velocity Saturation

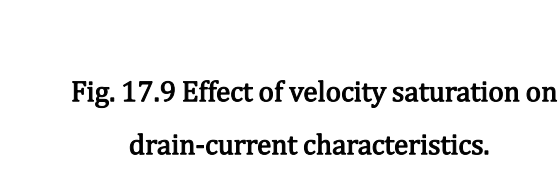


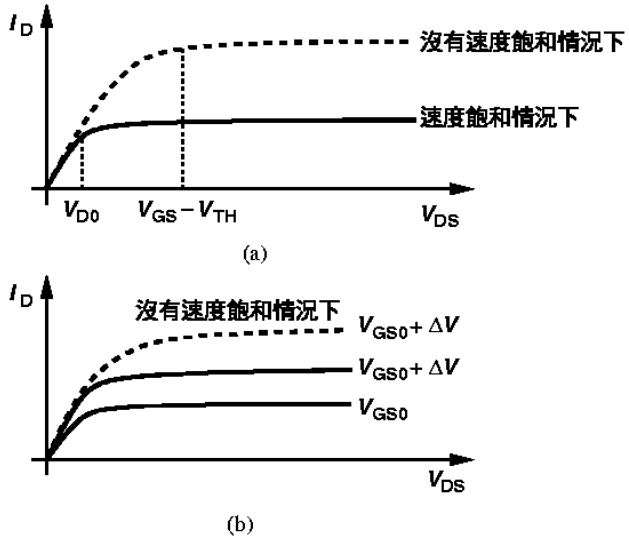
Fig. 17.9 Effect of velocity saturation on drain-current characteristics.

- The current is linearly proportional to the overdrive voltage 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 drain current and channel length in the velocity saturation region. $g_m = v_{sat} W C_{ox}$

2.7 Short Channel Effect

2.7.3 Velocity Saturation - Effect of Velocity Saturation



- Under 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 quite lower than that obtained if the device is saturated by channel pinch-off.
- The transconductance is also lower than that predicted by the square law.

Chapter 2
MOS Device Physics and Models

2.7 Short Channel Effect

2.7.4 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.