ECEN474/704: (Analog) VLSI Circuit Design
Spring 2018

Lecture 2: MOS Transistor Modeling

Sam Palermo

Analog & Mixed-Signal Center

Texas A&M University

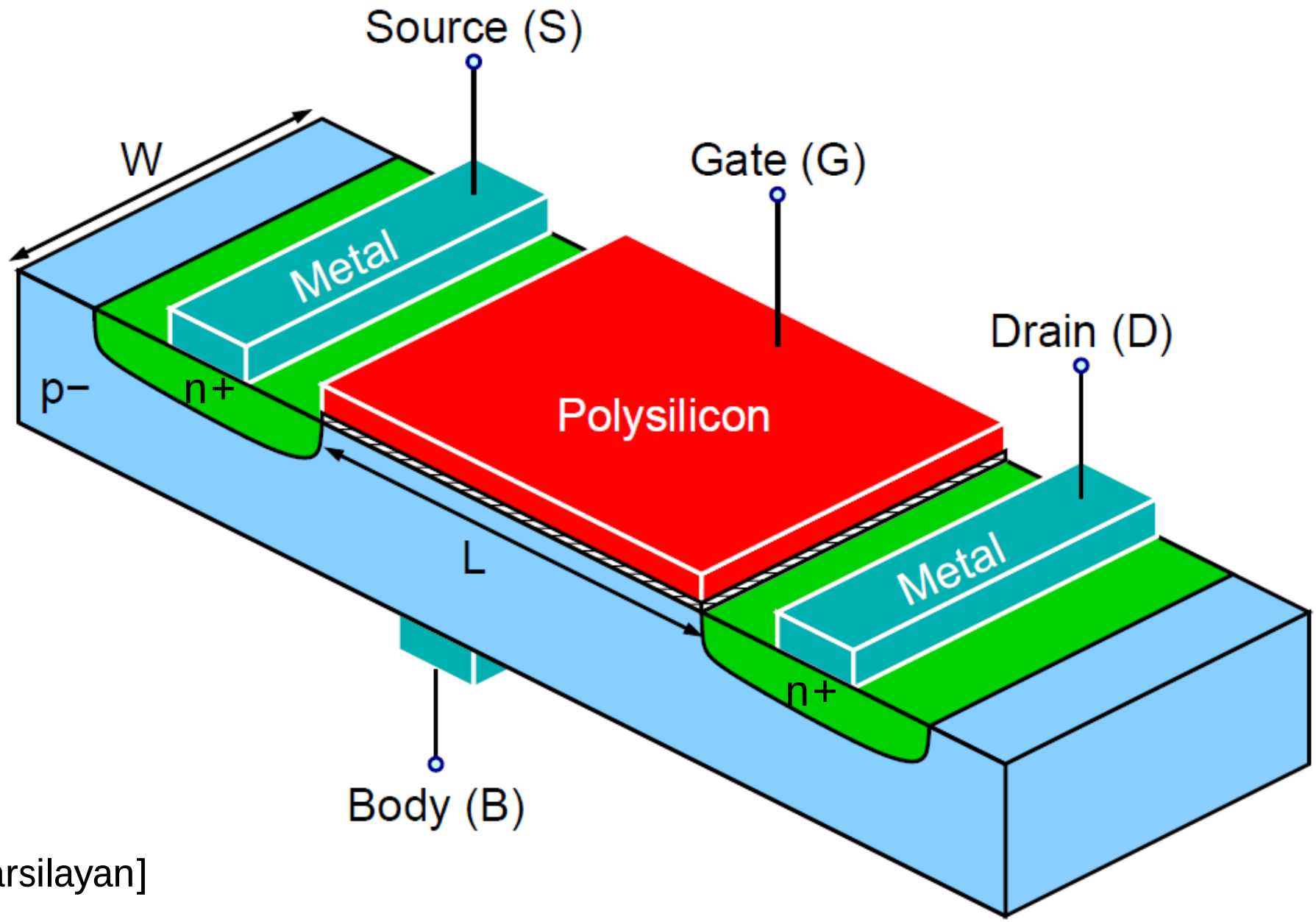
Announcements

- If you haven’t already, turn in your 0.18um NDA form ASAP
- Lab 1 starts Jan 31
- Current Reading
  - Razavi Chapters 2 & 17

Agenda

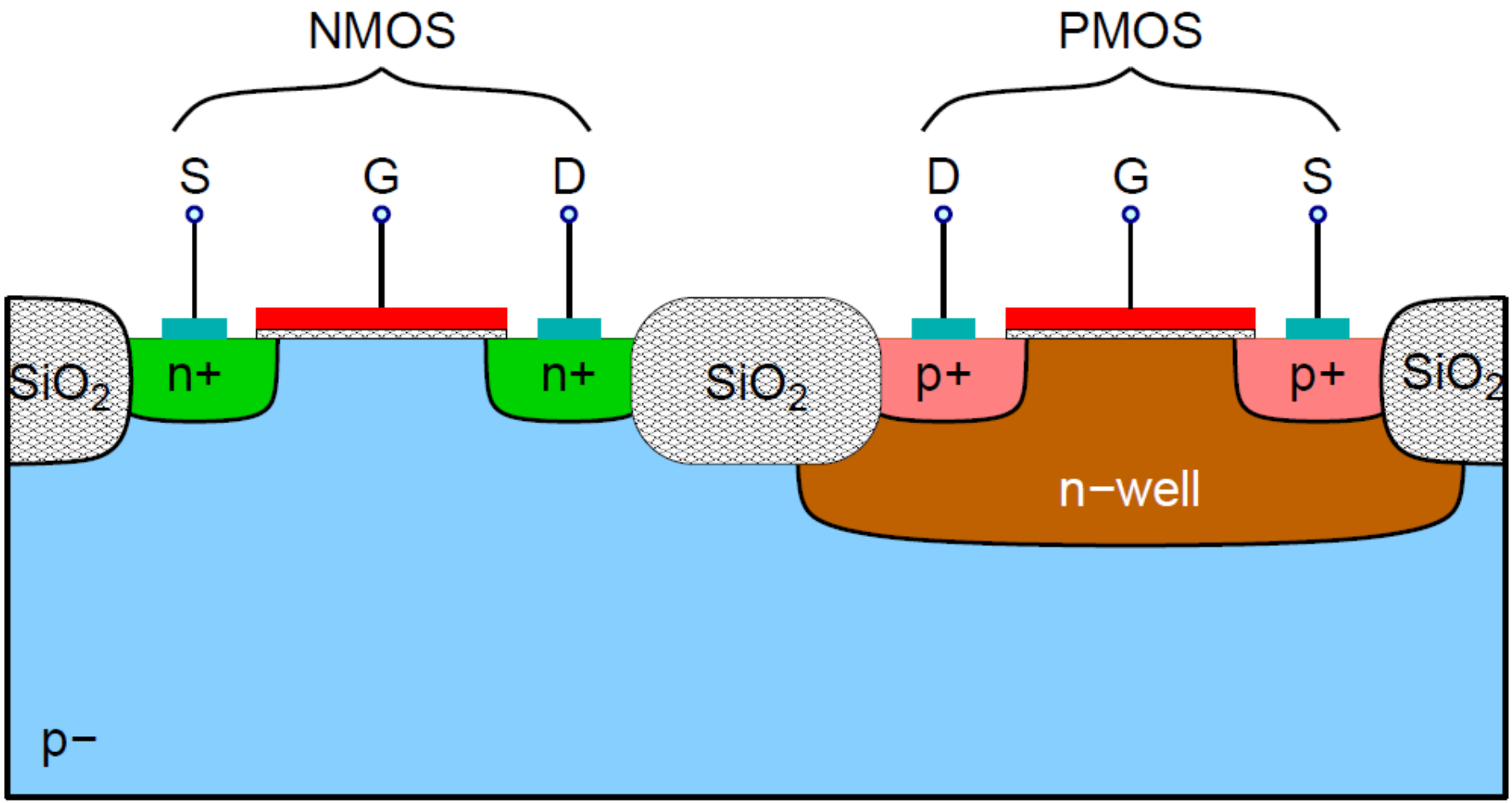
- MOS Transistor Modeling
  - Physical Structure
  - Threshold Voltage, V_T
  - DC I-V Equations
  - Body Effect
  - Subthreshold Region

NMOS Physical Structure



[Karsilayan]

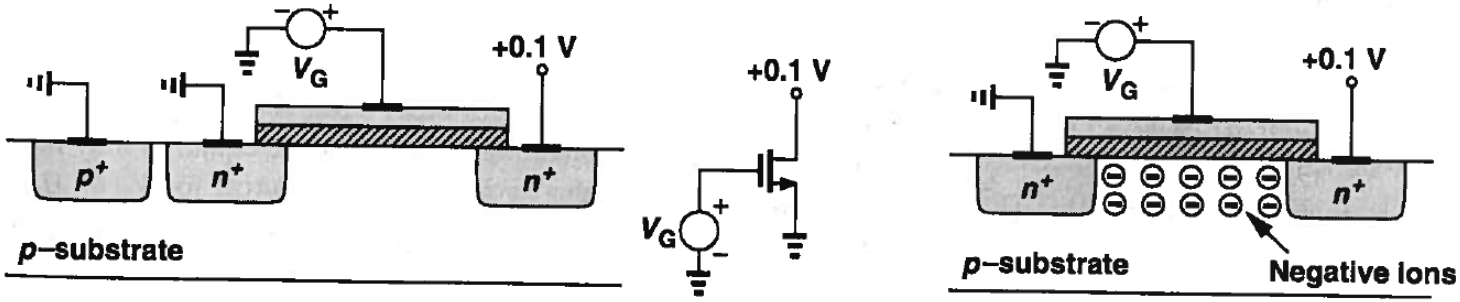
CMOS Physical Structure



[Karsilayan]

Threshold Voltage, V_T

[Razavi]

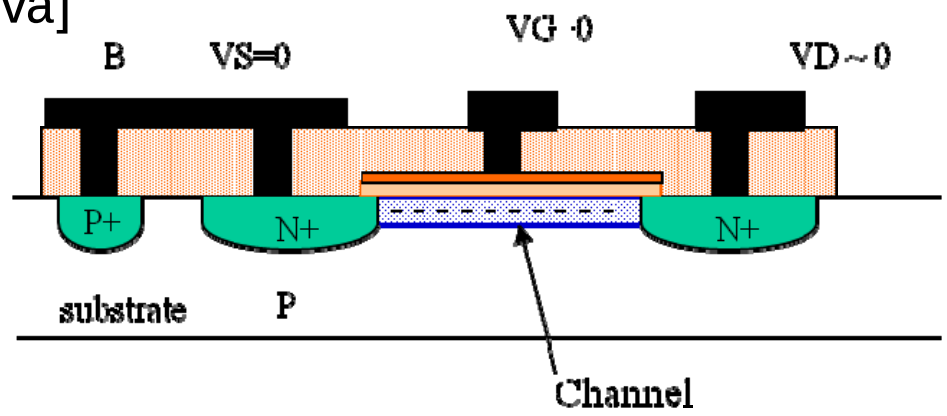


- Applying a positive voltage to the gate repels holes in the p-substrate under the gate, leaving negative ions (depletion region) to mirror the gate charge

- Before a “channel” forms, the device acts as 2 series caps from the oxide cap and the depletion cap
- If V_G is increased to a sufficient value the area below the gate is “inverted” and electrons flow from source to drain

V_T Definition

[Silva]



- The threshold voltage, V_T, is the voltage at which an "inversion layer" is formed
  - For an NMOS this is when the concentration of electrons equals the concentration of holes in the p- substrate

$$V_T = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep}}{C_{ox}} = \Phi_{MS} + 2\Phi_F + \gamma\sqrt{2\Phi_F}$$

Φ_{MS} is the difference between the work functions of the polysilicon gate and the silicon substrate

Φ_F is the Fermi potential, $\Phi_F = \frac{kT}{q}\ln\left(\frac{N_{sub}}{n_i}\right)$

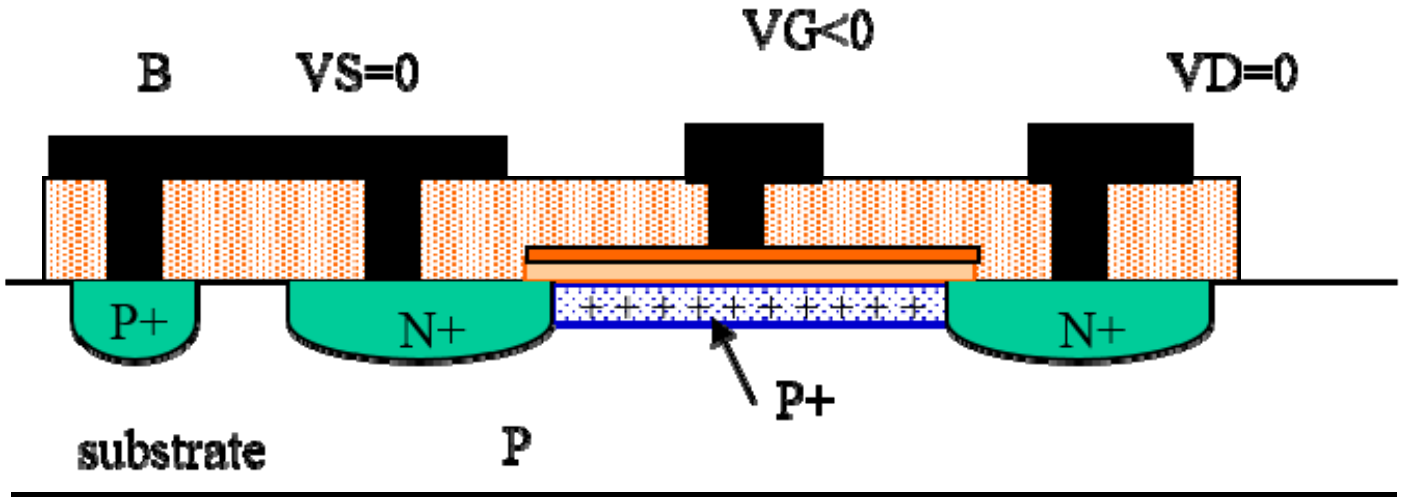
N_{sub} is substrate doping density, n_i is undoped silicon electron density

Q_{dep} is the depletion region charge, $Q_{dep} = \sqrt{4q\varepsilon_{si}\left|\Phi_F\right|N_{sub}}$

C_{ox} is the gate cap/area, $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$

Note, γ will be defined later

MOSFET in Accumulation Mode



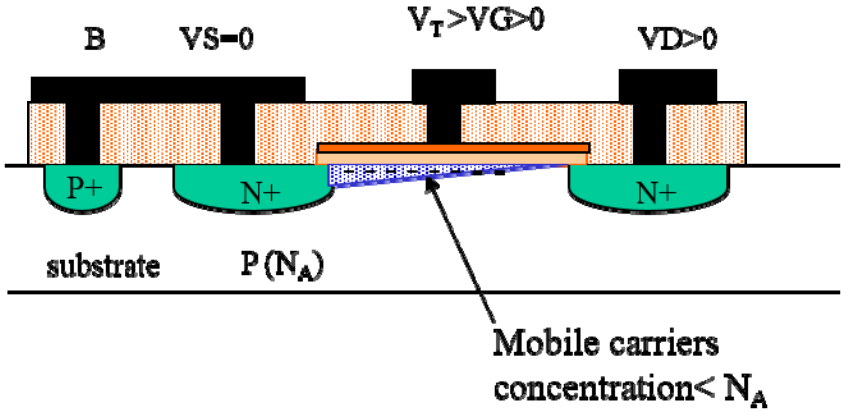
[Silva]

- If a negative gate voltage is applied w.r.t. the source, then positive charge “accumulates” below the gate
- In this Accumulation Mode, no current flows and the device is often used as a capacitor with the drain shorted to the source
- This capacitor consists of parallel plate capacitance below the gate and overlap/fringing capacitance near the drain/source regions

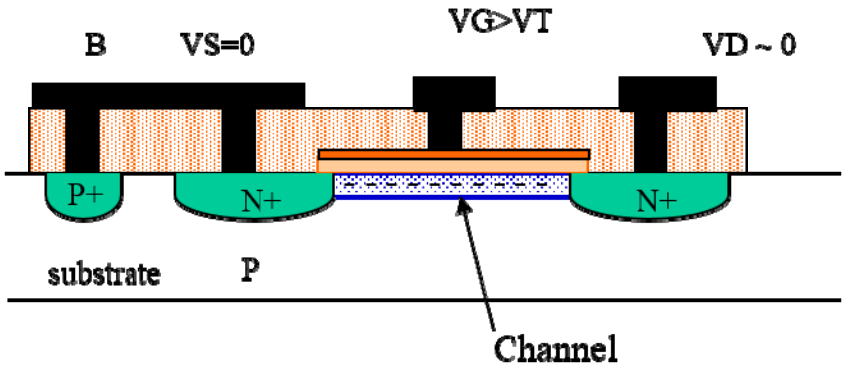
$$C_{G,acc} = WL_{eff}C_{ox} + 2WC_{ov}$$

MOSFET in Inversion Mode

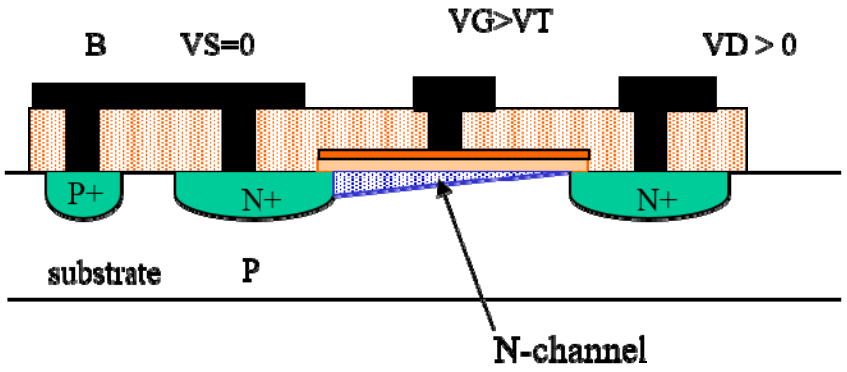
Subthreshold
$0<V_G<V_T$
$V_{DS}>0$



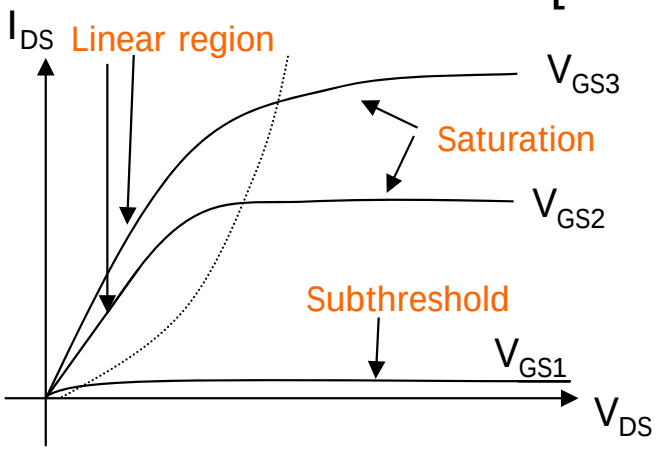
Triode / Linear
$V_G>V_T$
Small V_{DS}



Saturation
$V_G>V_T$
$V_{DS}>V_{GS}-V_T$



N-type transistor [Silva]



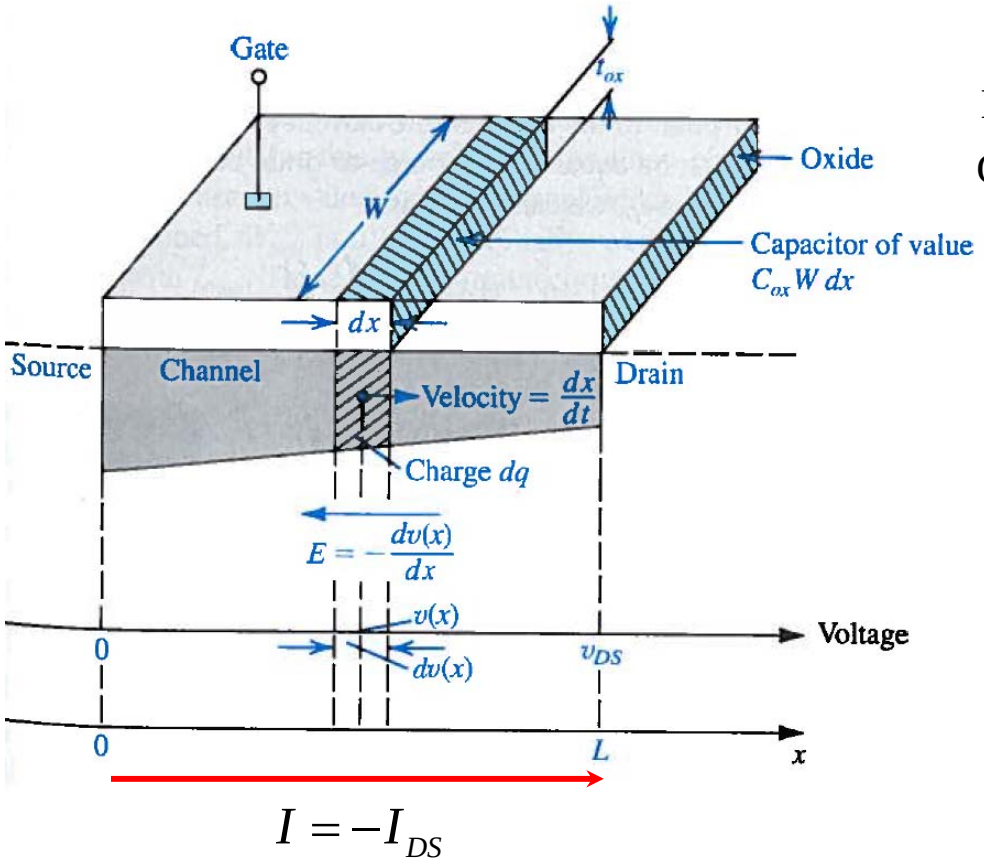
Subthreshold (extremely low-voltage low-power applications)

Linear region (voltage controlled resistor, linear OTA's, multipliers, switches)

Saturation region (Amplifiers)

MOS Equations in Triode Region (Small V_{DS})

[Sedra/Smith]



Current from Source to Drain : $I = \frac{dQ}{dt} = \frac{dQ}{dx}\frac{dx}{dt} = Q_d(x)\upsilon$

Incremental Charge Density : $Q_d(x) = -C_{OX}W(V_{GC}(x) - V_T)$

Gate - to - Channel Voltage : $V_{GC}(x) - V_T = V_{GS} - V_{CS}(x) - V_T$

Electron Velocity : $\upsilon = -\mu_n E(x) = \mu_n \frac{dv(x)}{dx}$

$$I_{DS} = -I = C_{OX}W(V_{GS} - V_{CS}(x) - V_T)\mu_n \frac{dv(x)}{dx}$$

$$\int_0^L I_{DS}\,dx = \int_0^{V_{DS}} \mu_n C_{OX}W(V_{GS} - V_T - V_{CS}(x))\,dv(x)$$

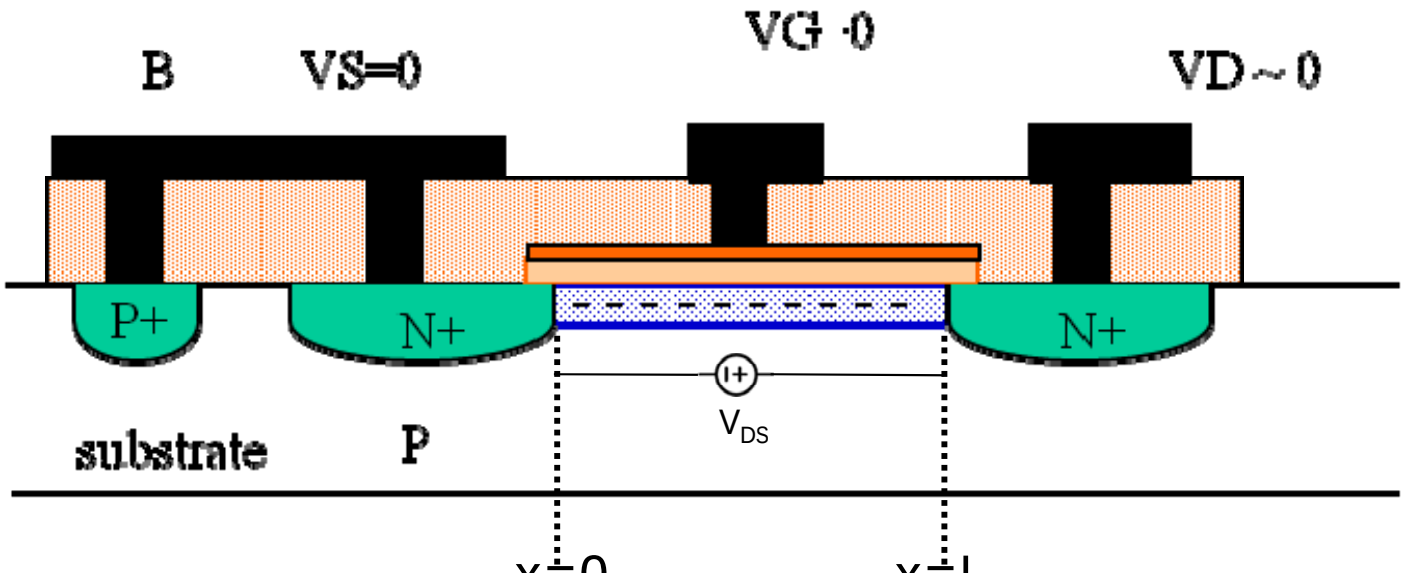
$$I_{DS}L = \mu_n C_{OX}W(V_{GS} - V_T)V_{DS} - \mu_n C_{OX}W\left(\frac{1}{2}V_{DS}^2\right)$$

$$\boxed{I_{DS} = \mu_n C_{OX}\frac{W}{L}\left(V_{GS} - V_T - \frac{1}{2}V_{DS}\right)V_{DS}}$$

Capacitance per unit gate area : $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$

Electron mobility : μ_n

Triode or Linear Region



[Silva]

$$V(0)=0 \qquad V(x)=V_{DS}\frac{x}{L} \qquad V(L)=V_{DS}$$

$$V_{GC}(x)=V_{GS}-V(x)=V_{GS}-V_{DS}\frac{x}{L}$$

V(x) = Channel-Source Voltage

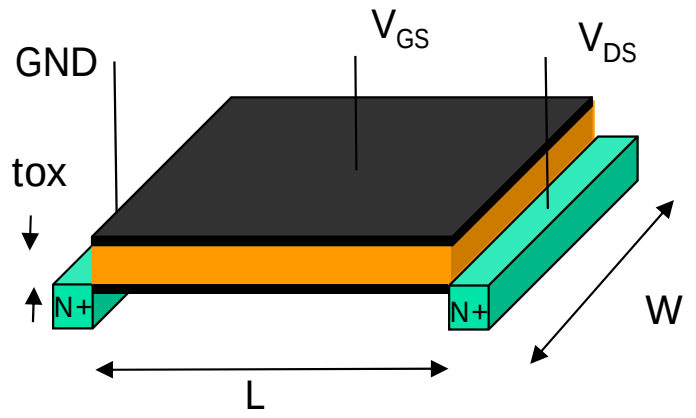
- Channel depth and transistor current is a function of the overdrive voltage, V_{GS}-V_T, and V_{DS}
- Because V_{DS} is small, V_{GC} is roughly constant across channel length and channel depth is roughly uniform

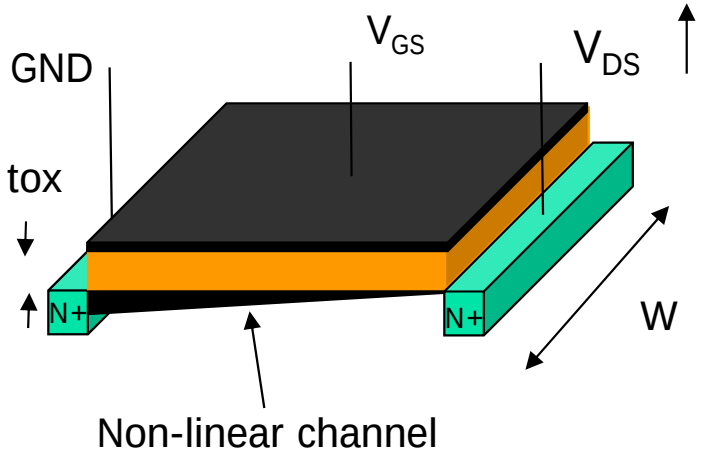
$$I_{DS}=\frac{W}{L}\mu_n C_{OX}\left(V_{GS}-V_{Tn}-0.5V_{DS}\right)V_{DS}$$

For small V_{DS}

$$R_{DS}\approx\frac{1}{\frac{W}{L}\mu C_{ox}\left(V_{GS}-V_{Tn}\right)}$$

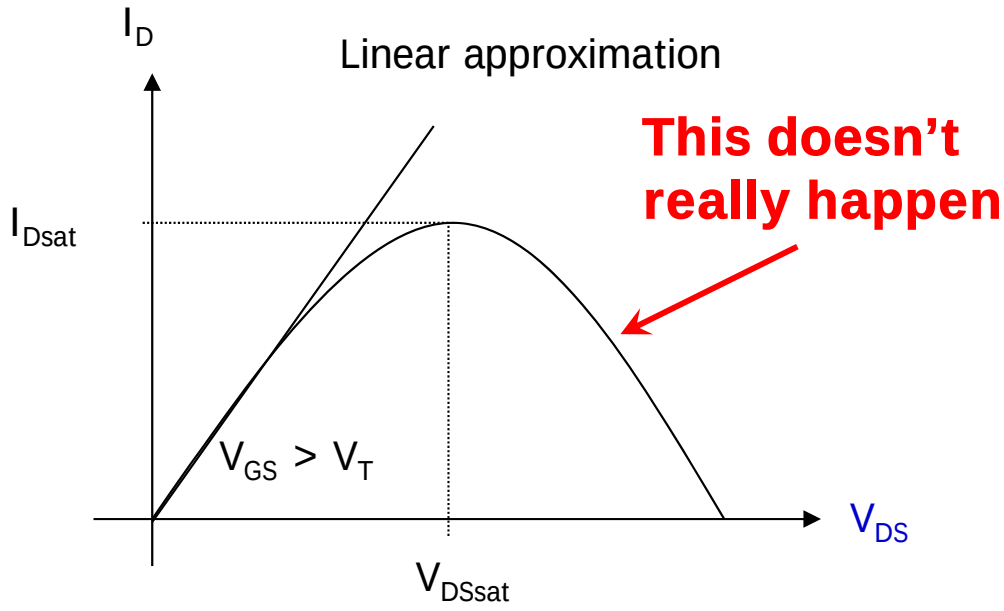
MOS Equations in Triode Region (Large V_{DS})





Drain current: Expression used in SPICE level 1

$$I_{DS} = \frac{W}{L}\mu_n C_{OX}\left(V_{GS} - V_{Tn} - 0.5V_{DS}\right)V_{DS}$$

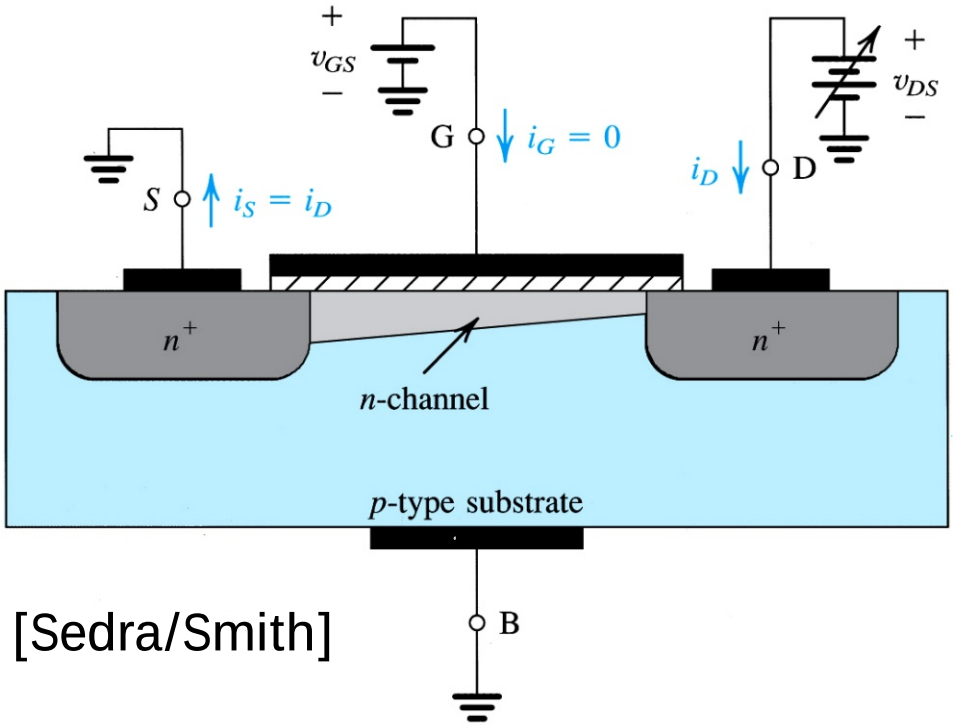


$$V_{DSsat} = V_{GS} - V_{Tn}$$

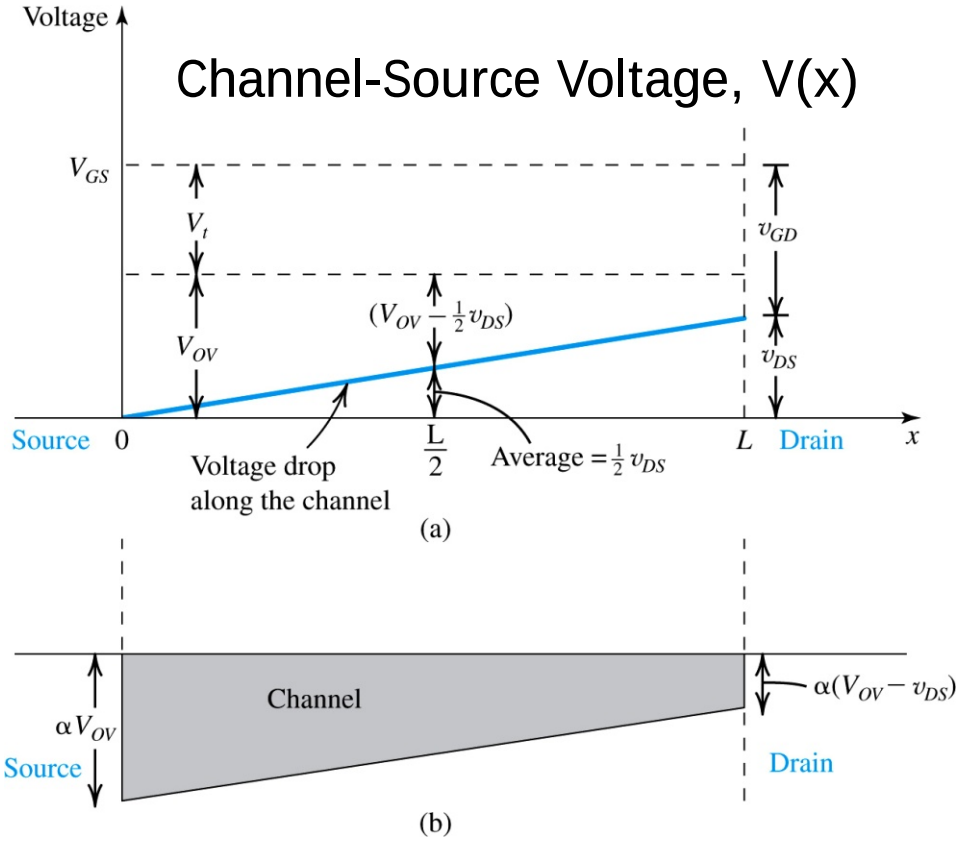
Triode Region Channel Profile

- Recall that the channel charge density is $\propto (V_{GC}(x) - V_T)$

$$V_{GC}(x) - V_T = V_{GS} - V(x) - V_T = V_{OV} - V_{DS}\frac{x}{L}$$

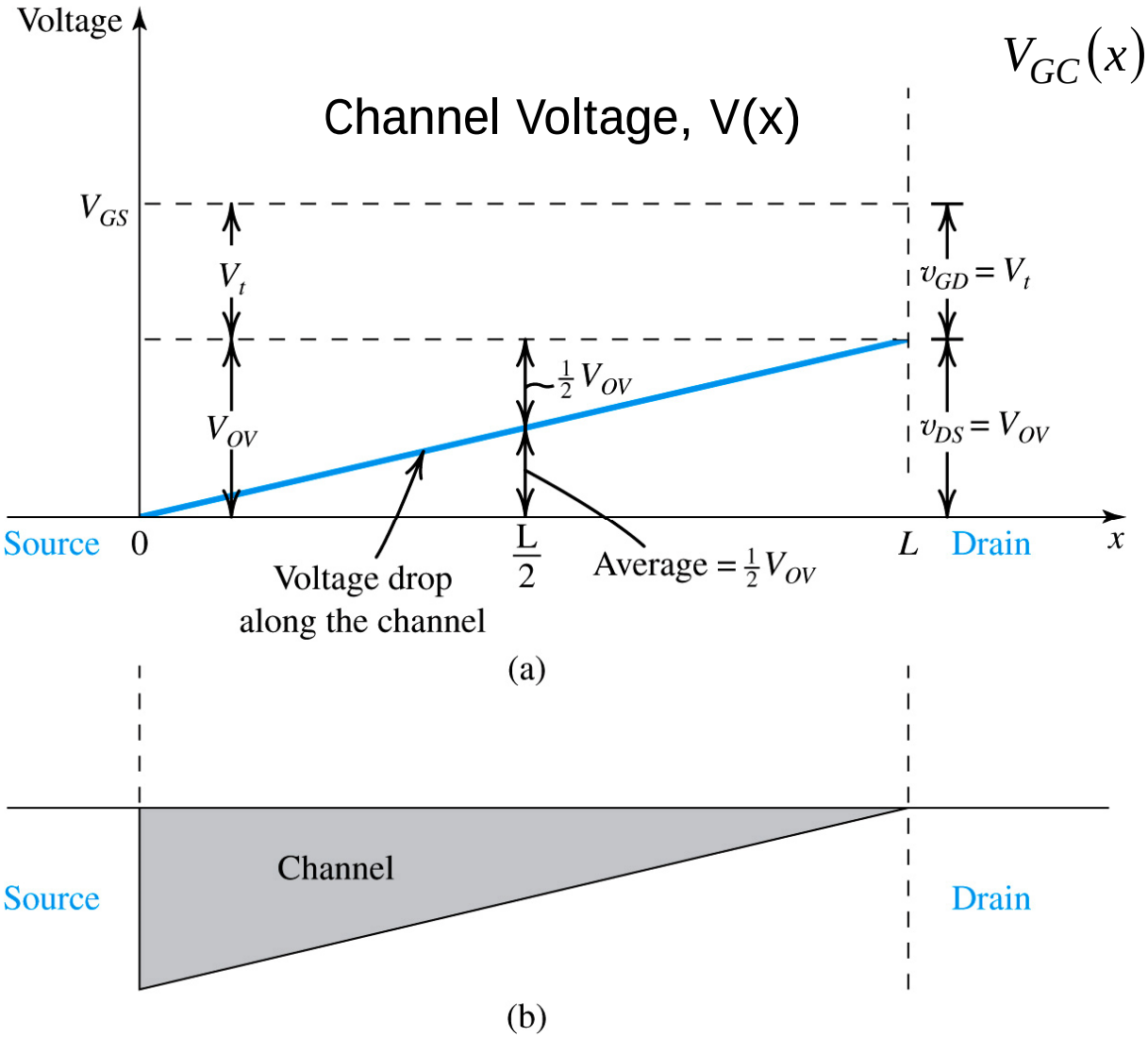


[Sedra/Smith]



- If V_{GC} is always above V_T throughout the channel length, the transistor current obeys the triode region current equation

Saturation Region Channel Profile

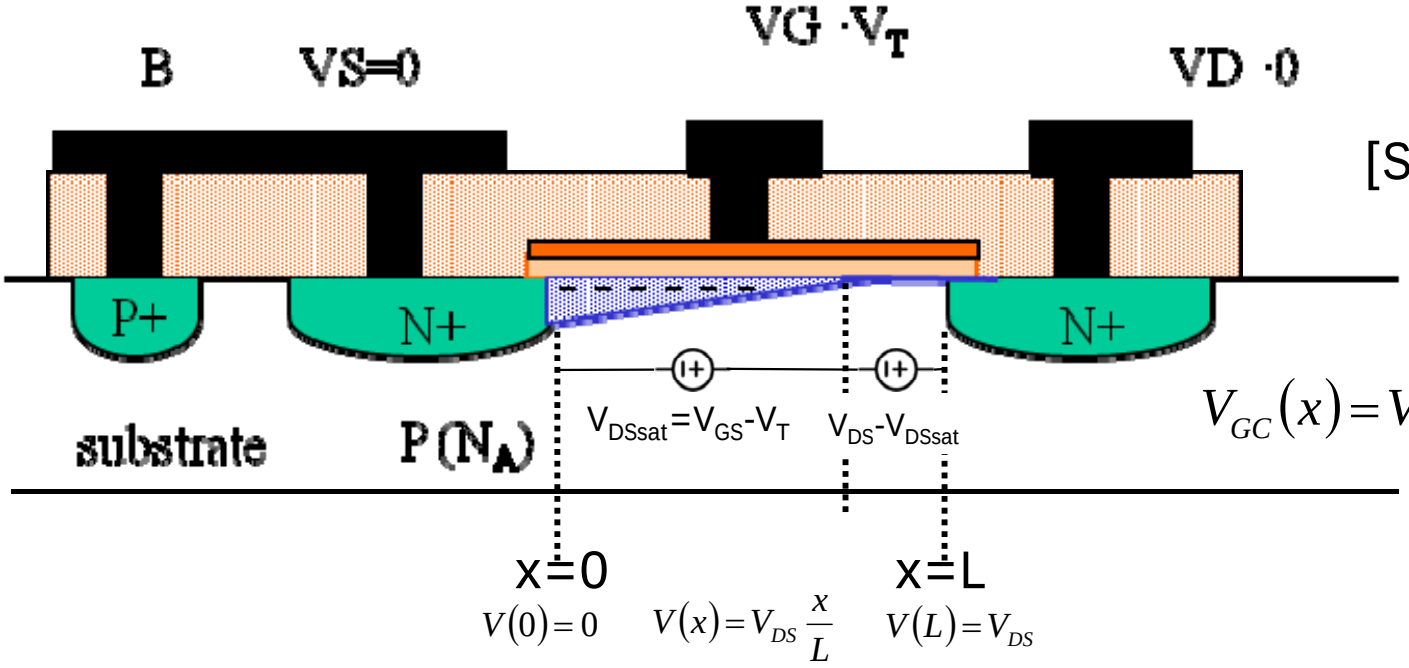


[Sedra/Smith]

$$V_{GC}(x) - V_T = V_{GS} - V(x) - V_T = V_{OV} - V_{DS}\frac{x}{L}$$

- When $V_{DS} \geq V_{GS}-V_T=V_{OV}$, V_{GC} no longer exceeds V_T, resulting in the channel "pinching off" and the current saturating to a value that is no longer a function of V_{DS} (ideally)

Saturation Region



[Silva]

$$V_{GC}(x) = V_{GS} - V(x) = V_{GS} - V_{DS}\frac{x}{L}$$

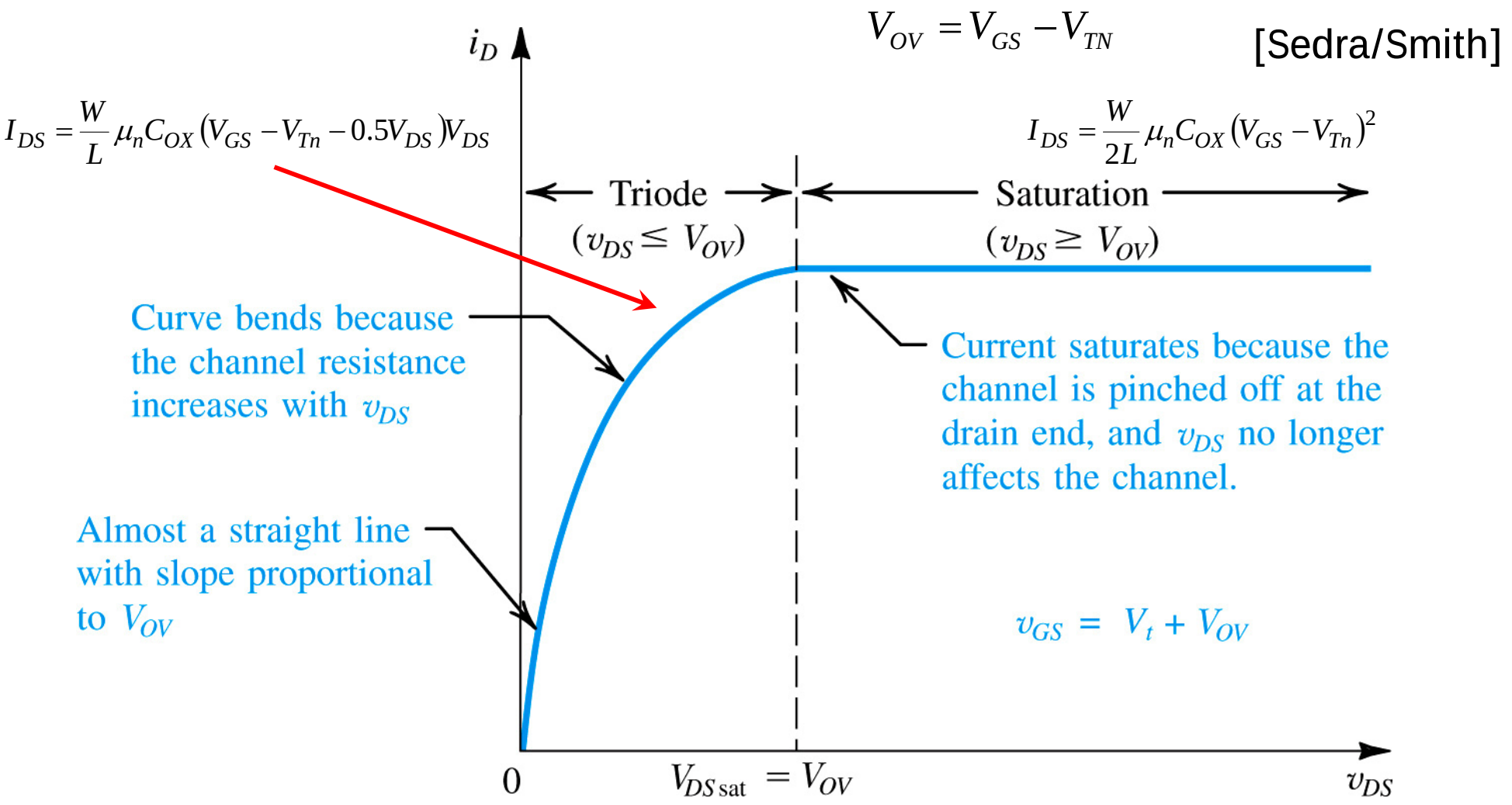
- Channel "pinches-off" when $V_{DS}=V_{GS}-V_T$ and the current saturates
- After channel charge goes to 0, the high lateral field "sweeps" the carriers to the drain and drops the extra V_{DS} voltage

$$I_{DS} = \mu_n C_{OX}\frac{W}{L}\left(V_{GS} - V_{Tn} - \frac{V_{DS}}{2}\right)V_{DS}\Bigg|_{V_{DS}=V_{GS}-V_{Tn}}$$

$$\boxed{V_{DSsat} = V_{GS} - V_{Tn}}$$

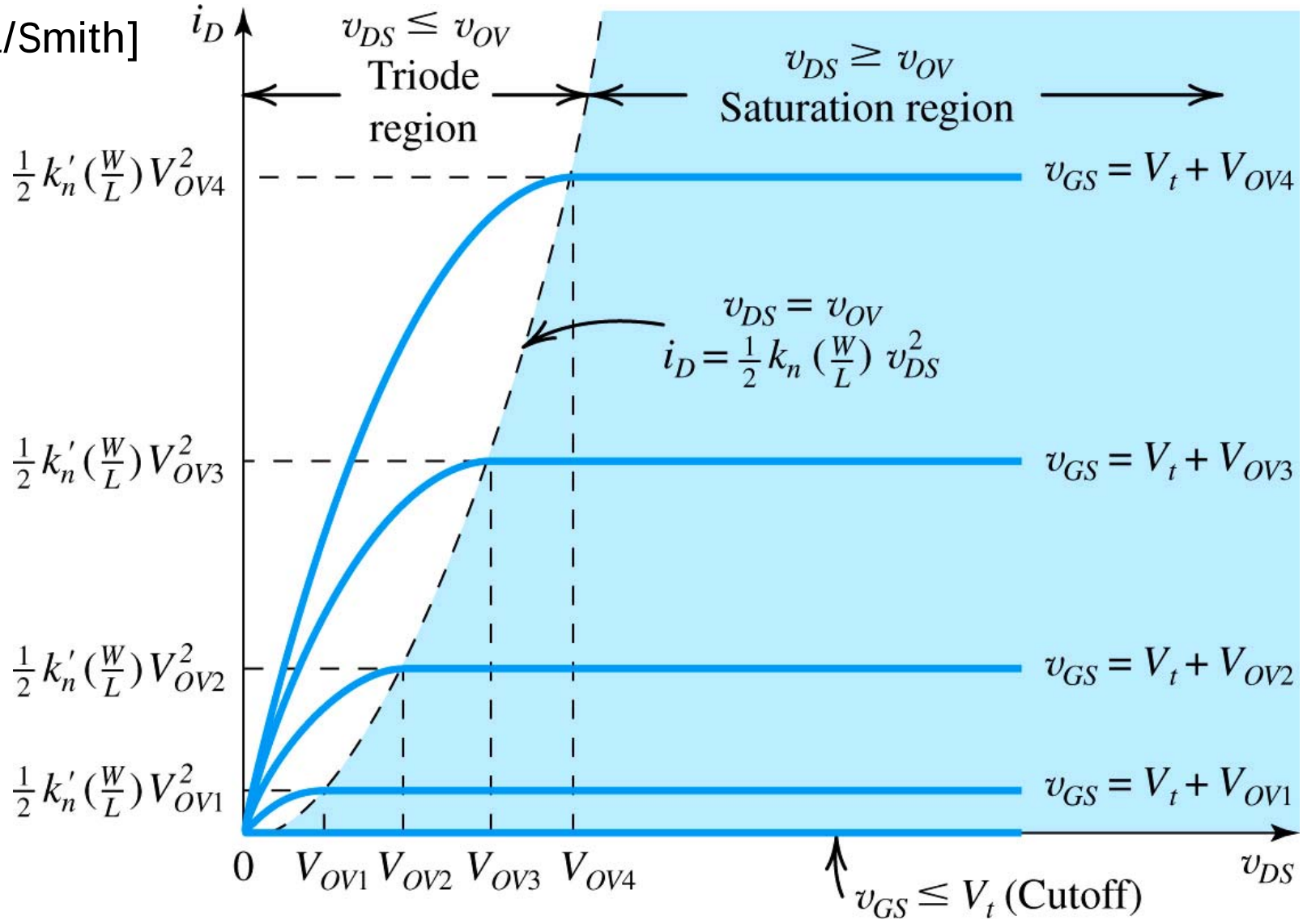
$$\boxed{I_{DS} = \frac{\mu_n C_{OX}}{2}\frac{W}{L}(V_{GS} - V_{Tn})^2}$$

NMOS $I_D - V_{DS}$ Characteristics



MOS "Large-Signal" Output Characteristic

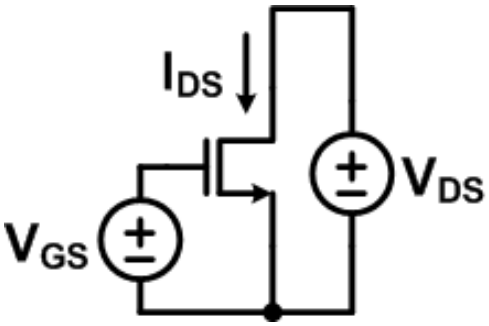
[Sedra/Smith]



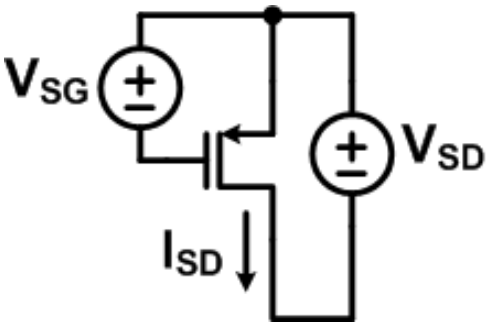
Note: $V_{OV} = V_{GS} - V_T$ and $k_n' = \mu_n C_{ox}$

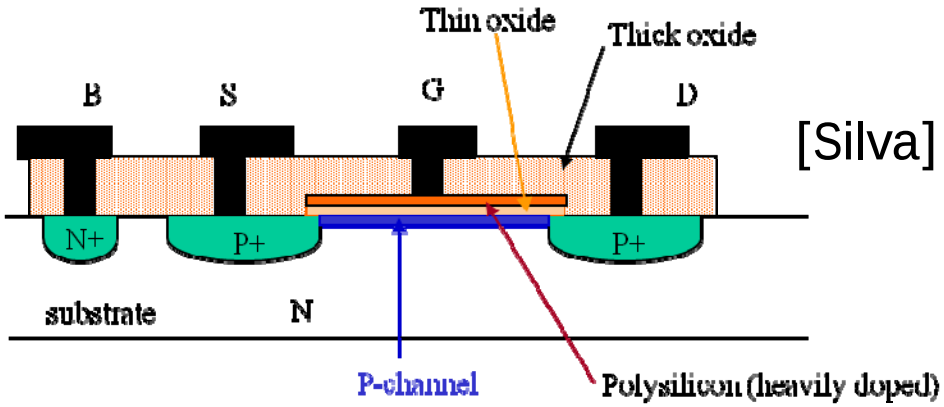
What about the PMOS device?

NMOS



PMOS





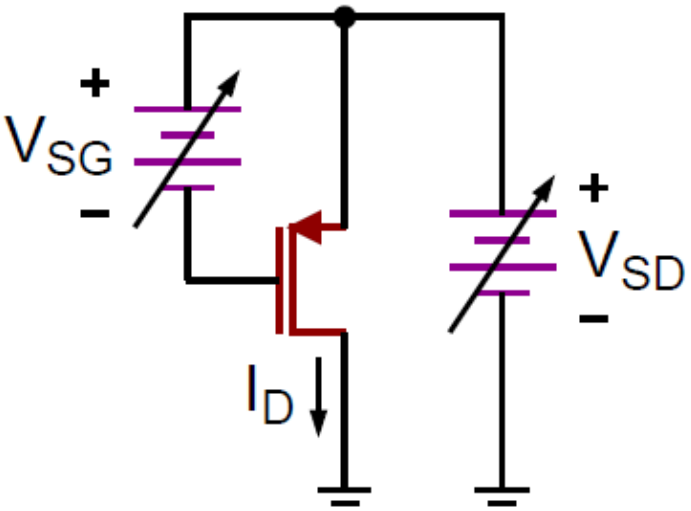
[Silva]

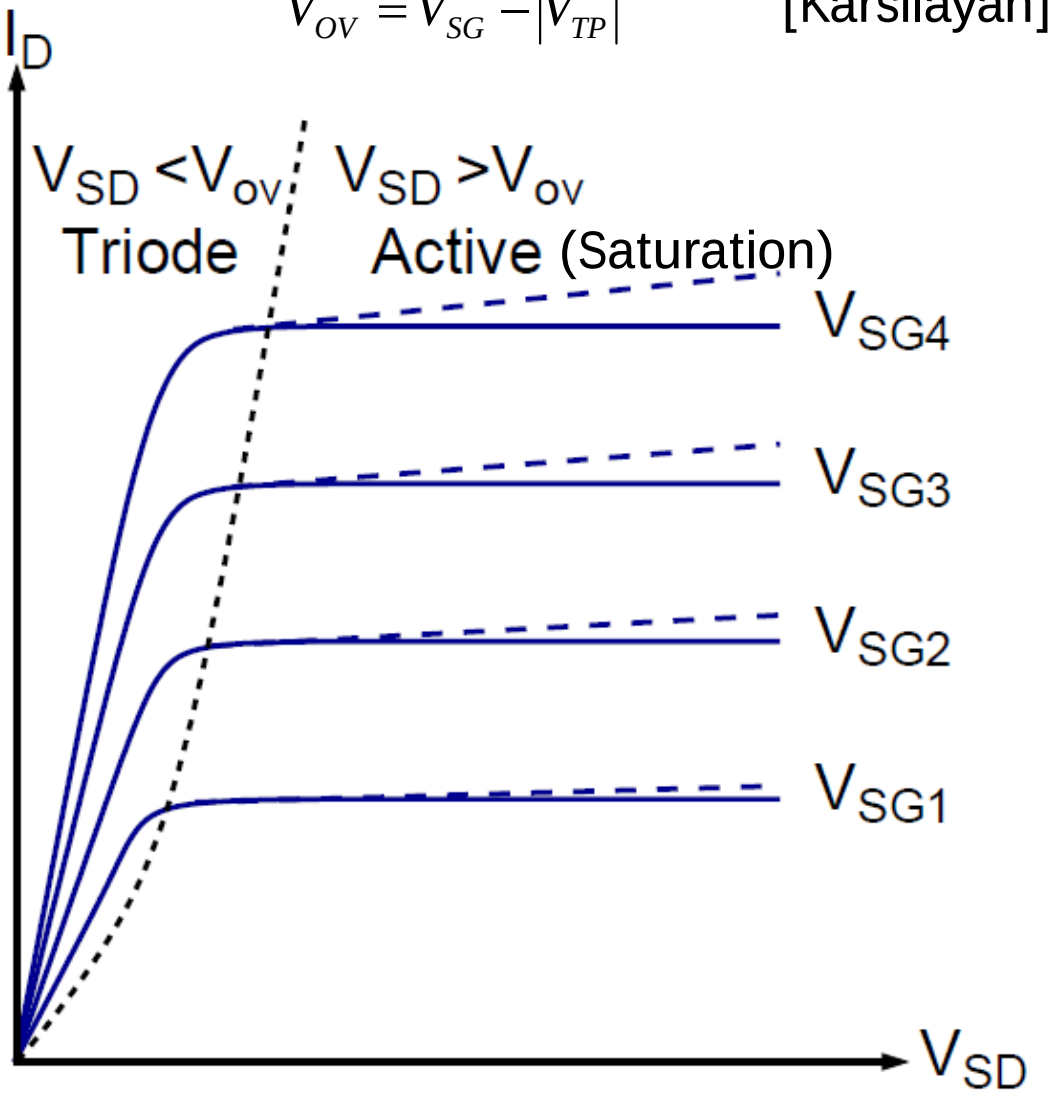
- The current equations for the PMOS device <u>are the same</u> as the NMOS **EXCEPT** you swap the current direction and all the voltage polarities

	NMOS	**PMOS**
Linear:	$I_{DS} = \frac{W}{L}\mu_n C_{OX}\left(V_{GS} - V_{Tn} - 0.5V_{DS}\right)V_{DS}$	$I_{SD} = \frac{W}{L}\mu_p C_{OX}\left(V_{SG} - \lvert V_{Tp}\rvert - 0.5V_{SD}\right)V_{SD}$
Saturation:	$I_{DS} = \frac{W}{2L}\mu_n C_{OX}\left(V_{GS} - V_{Tn}\right)^2$	$I_{SD} = \frac{W}{2L}\mu_p C_{OX}\left(V_{SG} - \lvert V_{Tp}\rvert\right)^2$

PMOS I_D – V_{SD} Characteristics

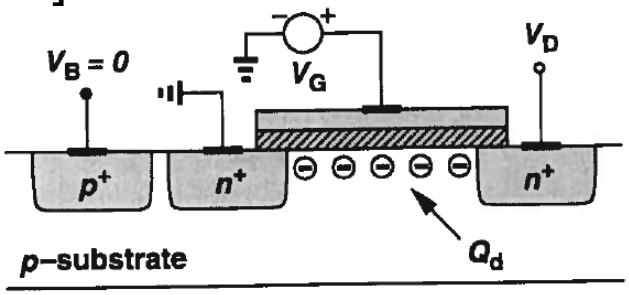
$$V_{OV} = V_{SG} - |V_{TP}|$$ [Karsilayan]

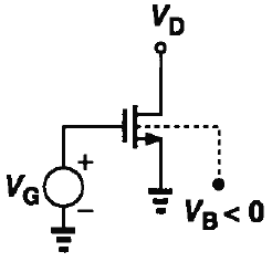


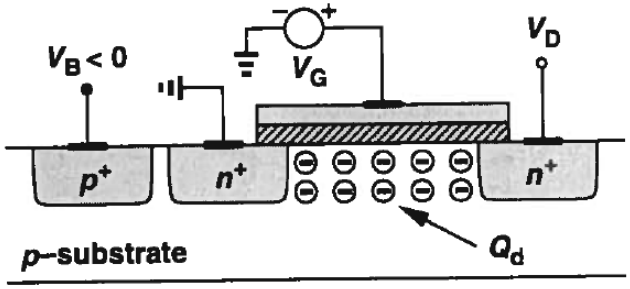


Body Effect

[Razavi]







- If the body and source potential are equal, a certain $V_G=V_{T0}$ is required to form an inversion layer

$$V_{T0} = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep0}}{C_{ox}} = \Phi_{MS} + 2\Phi_F + \gamma\sqrt{2\Phi_F}$$

- As V_S becomes positive w.r.t. V_B, a larger depletion region forms, which requires a higher V_G to form a channel
- The net result is that V_T increases due to this "body effect"
- Note, it also works in reverse, as if you increase V_B w.r.t. V_S, then V_T lowers

$$V_T = V_{T0} + \gamma\left(\sqrt{|2\Phi_F + V_{SB}|} - \sqrt{|2\Phi_F|}\right)$$

$$\text{Body effect coefficient, } \gamma = \frac{\sqrt{2q\varepsilon_{si}N_{sub}}}{C_{ox}}$$

γ typically ranges from 0.3 to 0.4$V^{1/2}$

MOS MODEL: SPICE LEVEL-II

- **Drain current, Triode region**

$$\text{NMOS: } I_{DS} = \frac{W}{L}\mu_n C_{OX}\left(V_{GS} - V_{Tn} - 0.5V_{DS}\right)V_{DS}$$

$$\text{PMOS: } I_{SD} = \frac{W}{L}\mu_p C_{OX}\left(V_{SG} - |V_{Tp}| - 0.5V_{SD}\right)V_{SD}$$

- **Drain Current, Saturation region**

$$\text{NMOS: } I_{DS} = \frac{W}{2L}\mu_n C_{OX}\left(V_{GS} - V_{Tn}\right)^2$$

$$\text{PMOS: } I_{SD} = \frac{W}{2L}\mu_p C_{OX}\left(V_{SG} - |V_{Tp}|\right)^2$$

- **Threshold voltage (zero bias)**

$$V_{T0} = \Phi_{MS} + 2\Phi_F + \gamma\sqrt{2\Phi_F}$$

- **Threshold voltage**

$$V_T = V_{T0} + \gamma\left[\sqrt{2\Phi_F + V_{SB}} - \sqrt{2\Phi_F}\right] \Rightarrow V_{T0}\big|_{V_{SB}=0}$$

- **KP and γ (Spice Model)**

$$KP = \mu C_{OX}; \quad \gamma = \frac{\sqrt{2q\varepsilon_{si}N_{sub}}}{C_{OX}}$$

Subthreshold Region

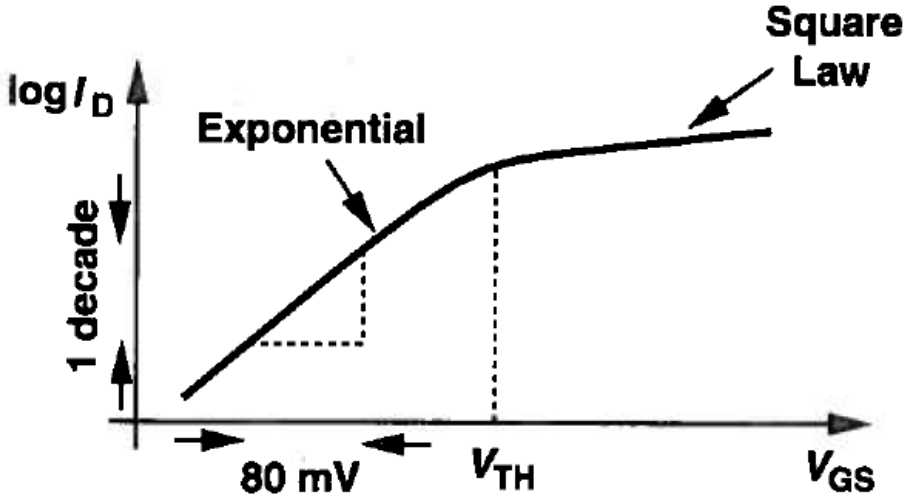
- So far we have assumed that $I_D=0$ when $V_{GS}<V_T$
- However, in reality an exponentially decreasing current exists for $V_{GS}<V_T$

In subthreshold region : $I_D = I_0 \exp\left(\frac{V_{GS} q}{\zeta kT}\right)$

where I_0 is a scale current

$\zeta > 1$ is a nonideality factor

The steepest subthreshold slope is 1dec./60mV with $\zeta = 1$



[Razavi]

- V_T values are often set by extrapolating above threshold data to current values of zero or infinite R_{on}
- A rough value often used is the V_{GS} which yields $I_D/W=1\mu A/\mu m$

Subthreshold Current & V_T Scaling

- This subthreshold current prevents lowering V_T excessively
- Assuming V_T=300mV and has an 80mV subthreshold slope, then the I_{on}/I_{off} ratio is only on the order of 10^(300/80)=5.6e3
- Reducing V_T to 200mV drops the I_{on}/I_{off} ratio to near 316
- If we have a large number of “off” transistors on our chip these subthreshold currents add up quickly, resulting in significant power dissipation
- This is a huge barrier in CMOS technology scaling and one of the main reasons Vdd scaling has slowed

Next Time

- MOS Transistor Modeling
  - Small-Signal Model
  - Spice Models