

Nonideal Transistor Theory

Outline

- Nonideal Transistor Behavior
 - High Field Effects
 - Mobility Degradation
 - Velocity Saturation
 - Channel Length Modulation
 - Threshold Voltage Effects
 - Body Effect
 - Drain-Induced Barrier Lowering
 - Short Channel Effect
 - Leakage
 - Subthreshold Leakage
 - Gate Leakage
 - Junction Leakage
- Process and Environmental Variations

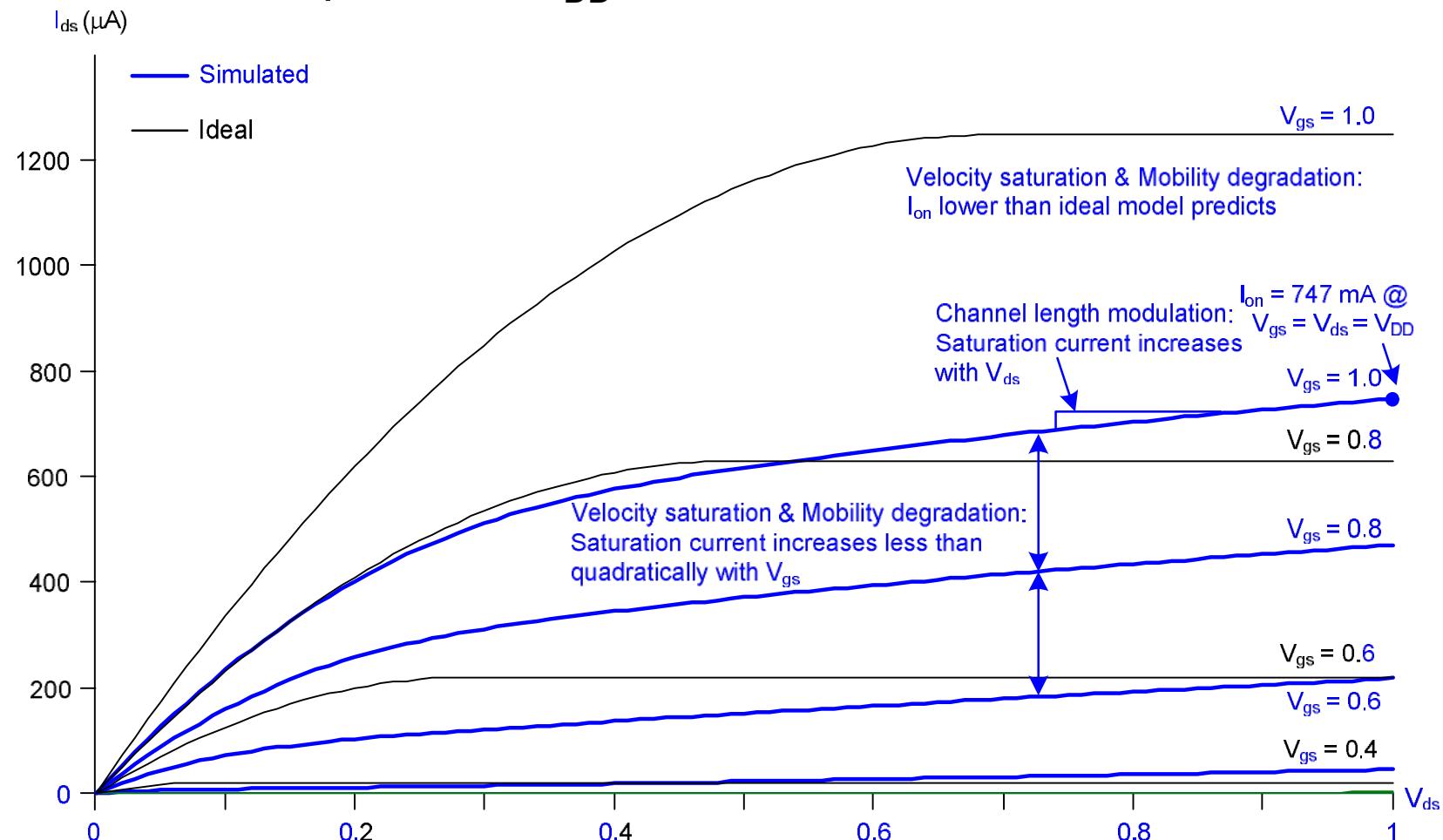
Ideal Transistor I-V

- Shockley long-channel transistor models

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{cutoff} \\ \beta \left(V_{gs} - V_t - \frac{V_{ds}}{2} \right) V_{ds} & V_{ds} < V_{dsat} \quad \text{linear} \\ \frac{\beta}{2} \left(V_{gs} - V_t \right)^2 & V_{ds} > V_{dsat} \quad \text{saturation} \end{cases}$$

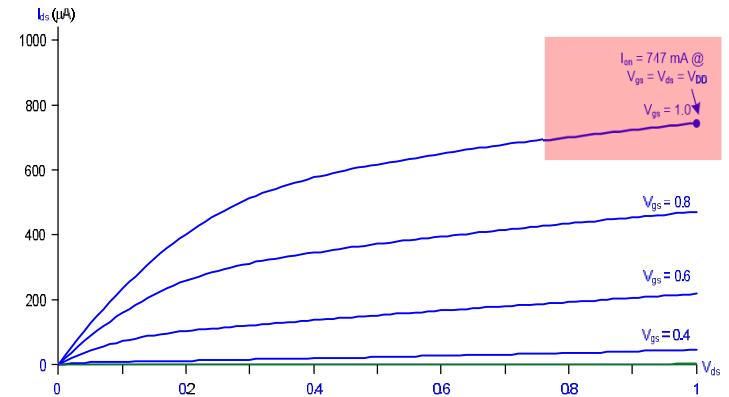
Ideal vs. Simulated nMOS I-V Plot

- 65 nm IBM process, $V_{DD} = 1.0$ V

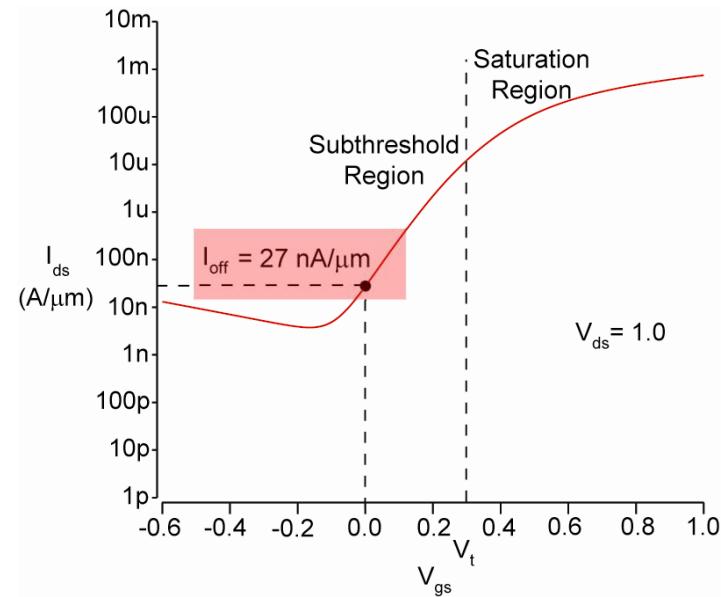


ON and OFF Current

- $I_{on} = I_{ds} @ V_{gs} = V_{ds} = V_{DD}$
 - Saturation

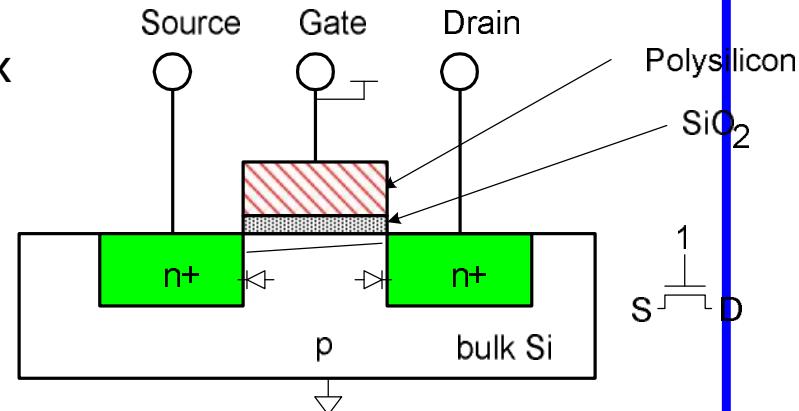


- $I_{off} = I_{ds} @ V_{gs} = 0, V_{ds} = V_{DD}$
 - Cutoff



Electric Fields Effects

- Vertical electric field: $E_{\text{vert}} = V_{\text{gs}} / t_{\text{ox}}$
 - Attracts carriers into channel
- Lateral electric field: $E_{\text{lat}} = V_{\text{ds}} / L$
 - Accelerates carriers from drain to source
 - Long channel: $v = \mu E_{\text{lat}}$
 - Long channel model assumes μ is constant
 - But at high fields μ is NOT constant



Coffee Cart Analogy

- Tired student runs from VLSI lab to coffee cart
- Freshmen are pouring out of the physics lecture hall
- V_{ds} is how long you have been up
 - Your velocity = fatigue \times mobility
- V_{gs} is a wind blowing you against the glass (SiO_2) wall
- At high V_{gs} , you are buffeted against the wall
 - *Mobility degradation*
- At high V_{ds} , you scatter off freshmen, fall down, get up
 - *Velocity saturation*
 - Don't confuse this with the saturation region

Mobility Degradation

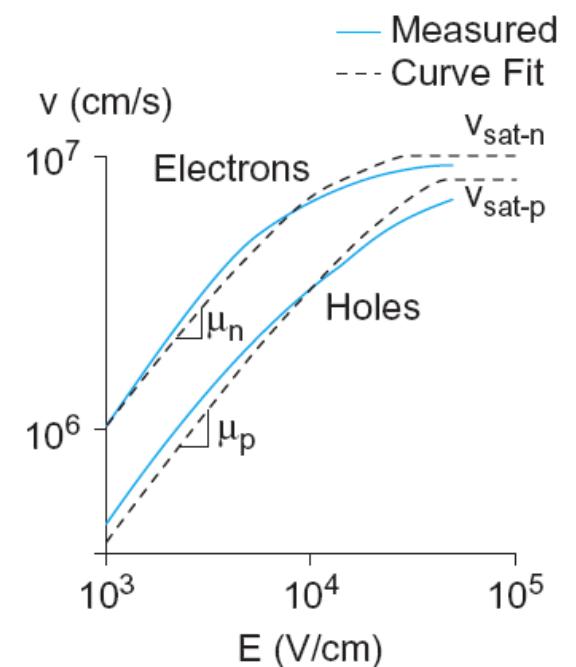
- High E_{vert} effectively reduces mobility
 - Collisions with oxide interface

$$\mu_{\text{eff-}n} = \frac{540 \frac{\text{cm}^2}{\text{V} \cdot \text{s}}}{1 + \left(\frac{V_{gs} + V_t}{0.54 \frac{\text{V}}{\text{nm}} t_{\text{ox}}} \right)^{1.85}}$$
$$\mu_{\text{eff-}p} = \frac{185 \frac{\text{cm}^2}{\text{V} \cdot \text{s}}}{1 + \left(\frac{V_{gs} + 1.5V_t}{0.338 \frac{\text{V}}{\text{nm}} t_{\text{ox}}} \right)^{1.85}}$$

Velocity Saturation

- At high E_{lat} , carrier velocity rolls off
 - Carriers scatter off atoms in silicon lattice
 - Velocity reaches v_{sat}
 - Electrons: 10^7 cm/s
 - Holes: 8×10^6 cm/s
 - Better model

$$v = \begin{cases} \frac{\mu_{eff}E}{E} & E < E_c \\ 1 + \frac{E}{E_c} & \\ v_{sat} & E \geq E_c \end{cases} \quad E_c = \frac{2v_{sat}}{\mu_{eff}}$$



Note that E_c (critical electric field) is assumed to be independent from E_{lat} , V_{DS} , and L

Vel Sat I-V Effects

- Ideal transistor ON current increases with V_{DD}^2

$$I_{ds} = \mu C_{ox} \frac{W}{L} \frac{(V_{gs} - V_t)^2}{2} = \frac{\beta}{2} (V_{gs} - V_t)^2$$

- Velocity-saturated ON current increases with V_{DD}

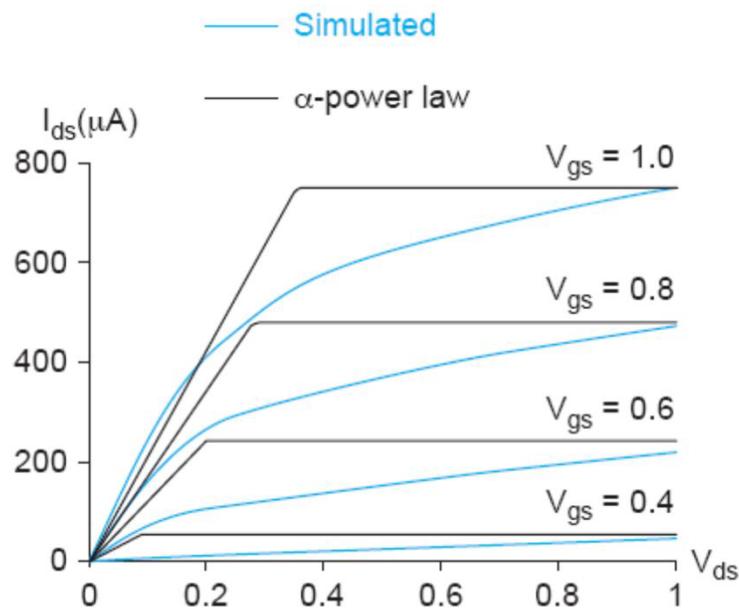
$$I_{ds} = C_{ox} W (V_{gs} - V_t) v_{max}$$

- Real transistors are partially velocity saturated
 - Approximate with α -power law model
 - $I_{ds} \propto V_{DD}^\alpha$
 - $1 < \alpha < 2$ determined empirically (≈ 1.3 for 65 nm)

α -Power Model

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \\ I_{dsat} \frac{V_{ds}}{V_{dsat}} & V_{ds} < V_{dsat} \\ I_{dsat} & V_{ds} > V_{dsat} \end{cases} \quad \begin{matrix} \text{cutoff} \\ \text{linear} \\ \text{saturation} \end{matrix}$$

$$I_{dsat} = P_c \frac{\beta}{2} (V_{gs} - V_t)^\alpha$$
$$V_{dsat} = P_v (V_{gs} - V_t)^{\alpha/2}$$



V-Sat Example

Example 2.4

Find the critical voltage for fully ON nMOS and pMOS transistors using the effective mobilities from Example 2.3.

SOLUTION: Using EQ(2.25)

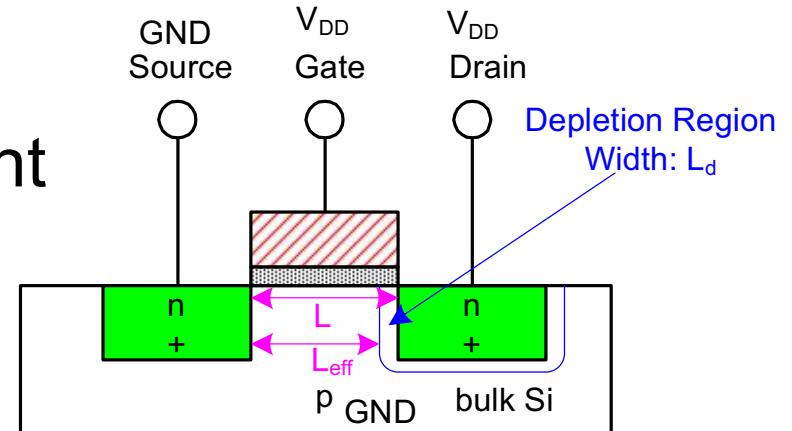
$$V_{c-n} = \frac{2\left(10^7 \frac{\text{cm}}{\text{s}}\right)}{96 \frac{\text{cm}^2}{\text{V} \cdot \text{s}}} \left(5 \times 10^{-6} \text{cm}\right) = 1.04 \text{ V}$$

$$V_{c-p} = \frac{2\left(8 \times 10^6 \frac{\text{cm}}{\text{s}}\right)}{36 \frac{\text{cm}^2}{\text{V} \cdot \text{s}}} \left(5 \times 10^{-6} \text{cm}\right) = 2.22 \text{ V}$$

The nMOS transistor is velocity saturated in normal operation because V_{c-n} is comparable to V_{DD} . The pMOS transistor has lower mobility and thus is not as badly velocity saturated.

Channel Length Modulation

- Reverse-biased p-n junctions form a *depletion region*
 - Region between n and p with no carriers
 - Width of depletion L_d region grows with reverse bias
 - $L_{eff} = L - L_d$
- Shorter L_{eff} gives **more current**
 - I_{ds} **increases** with V_{ds}
 - Even in saturation



Chan Length Mod I-V

$$I_{ds} = \frac{\beta}{2} (V_{gs} - V_t)^2 (1 + \lambda V_{ds})$$

- λ = *channel length modulation coefficient*
 - Not feature size
 - Empirically fit to I-V characteristics

Threshold Voltage Effects

- V_t is V_{gs} for which the channel starts to invert
- Ideal models assumed V_t is constant
- Really depends (weakly) on almost everything else:
 - Body voltage: *Body Effect*
 - Drain voltage: *Drain-Induced Barrier Lowering*
 - Channel length: *Short Channel Effect*

Body Effect

- Body is a fourth transistor terminal
- V_{sb} affects the charge required to invert the channel
 - Increasing V_s or decreasing V_b increases V_t
$$V_t = V_{t0} + \gamma \left(\sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s} \right)$$
- ϕ_s = *surface potential at threshold*
 - Depends on doping level N_A , thermal voltage,
 - And intrinsic carrier concentration n_i
- γ = *body effect coefficient (typically 0.4 to 1 v^{1/2})*

$$\gamma = \frac{t_{ox}}{\epsilon_{ox}} \sqrt{2q\epsilon_{si}N_A} = \frac{\sqrt{2q\epsilon_{si}N_A}}{C_{ox}}$$

Body Effect Cont.

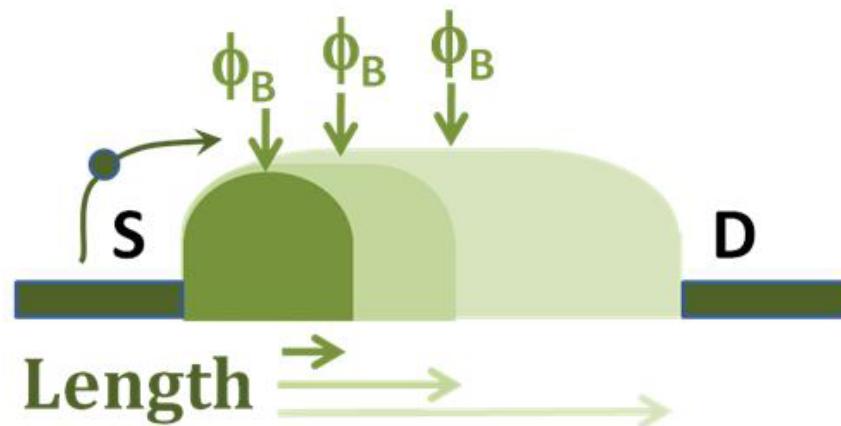
- For small source-to-body voltage, treat as linear

$$V_t = V_{t0} + k_\gamma V_{sb}$$

$$k_\gamma = \frac{\gamma}{2\sqrt{\phi_s}} = \frac{\sqrt{\frac{q\epsilon_{si}N_A}{v_T \ln \frac{N_A}{n_i}}}}{2C_{ox}}$$

DIBL

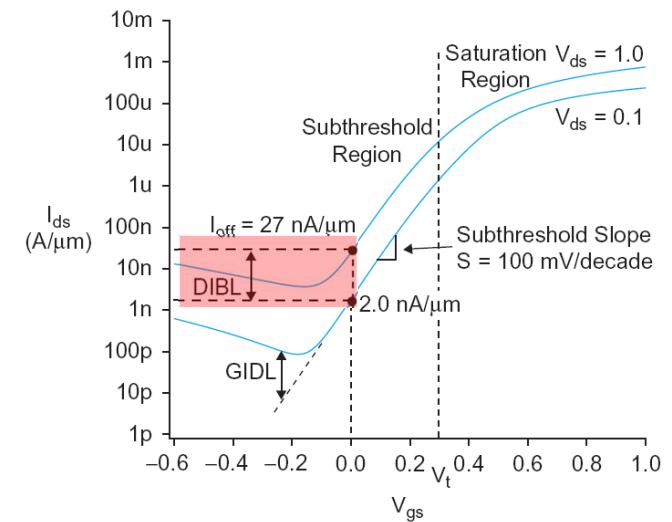
- Drain-Induced Barrier Lowering:
 - Reduction of threshold voltage of the transistor at higher drain voltages
- Electric field from drain affects channel
- More pronounced in small transistors where the drain is closer to the channel



DIBL (Cont.)

- Drain-Induced Barrier Lowering
 - Drain voltage also affects V_t

$$V'_t = V_t - \eta V_{ds}$$



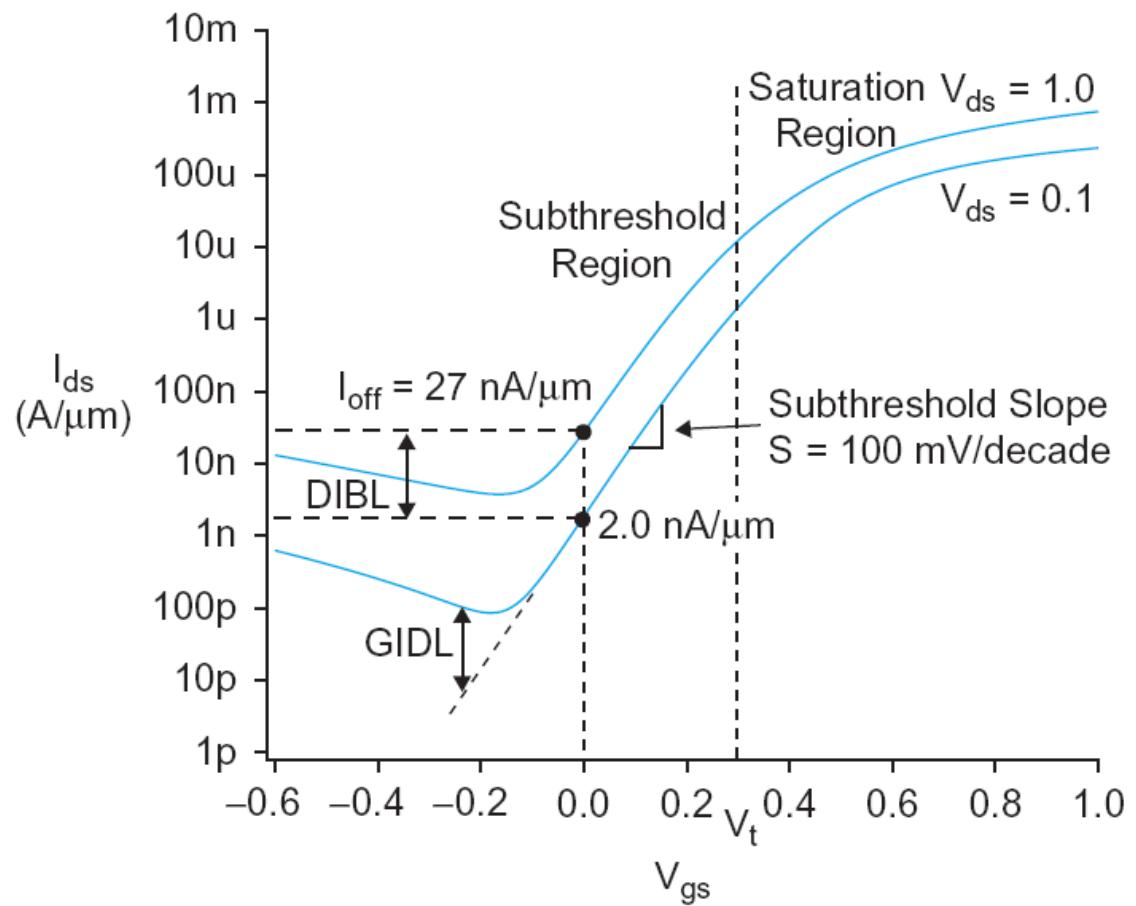
- High drain voltage causes current to **increase**.

Short Channel Effect

- In small transistors, source/drain depletion regions extend into the channel
 - Impacts the amount of charge required to invert the channel (recall a capacitor's relationship with area)
 - And thus makes V_t a function of channel length
- Short channel effect: V_t increases with L
 - Some processes exhibit a reverse short channel effect in which V_t decreases with L

Leakage

- ❑ What about current in cutoff?
- ❑ Simulated results
- ❑ What differs?
 - Current doesn't go to 0 in cutoff



Leakage Sources

- Subthreshold conduction
 - Caused by thermal emission of carriers over the potential barrier
 - Means transistors can't be completely OFF
 - Means transistors can't abruptly turn ON or OFF
 - Dominant source in contemporary transistors
- Gate leakage
 - Tunneling through ultrathin gate dielectric
- Junction leakage
 - Reverse-biased PN junction diode current

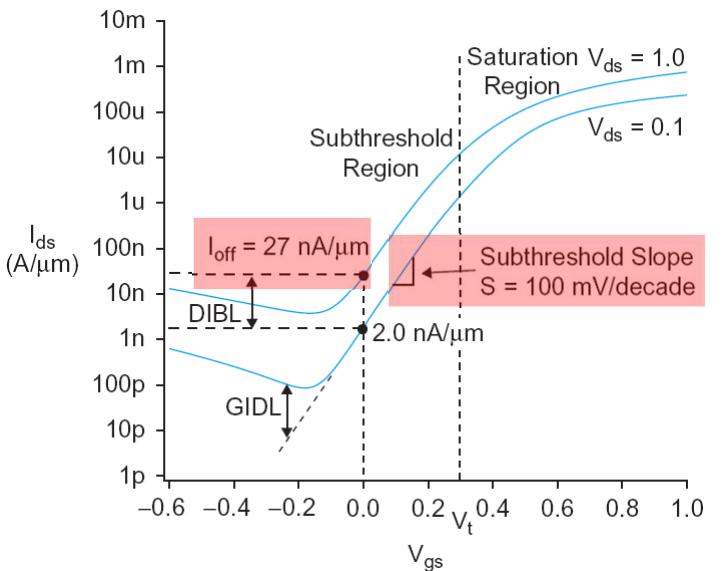
Subthreshold Leakage

- Subthreshold leakage exponential with V_{gs}

$$I_{ds} = I_{ds0} e^{\frac{V_{gs} - V_{t0} + \eta V_{ds} - k_\gamma V_{sb}}{nv_T} \left(1 - e^{\frac{-V_{ds}}{v_T}} \right)}$$

- n is process dependent
 - typically 1.3-1.7
- Note the effect of all parameters
- Rewrite relative to I_{off} on log scale

$$I_{ds} = I_{off} 10^{\frac{V_{gs} + \eta(V_{ds} - V_{dd}) - k_\gamma V_{sb}}{S} \left(1 - e^{\frac{-V_{ds}}{v_t}} \right)}$$



$$S = \left[\frac{d(\log_{10} I_{ds})}{dV_{gs}} \right]^{-1} = nv_T \ln 10$$

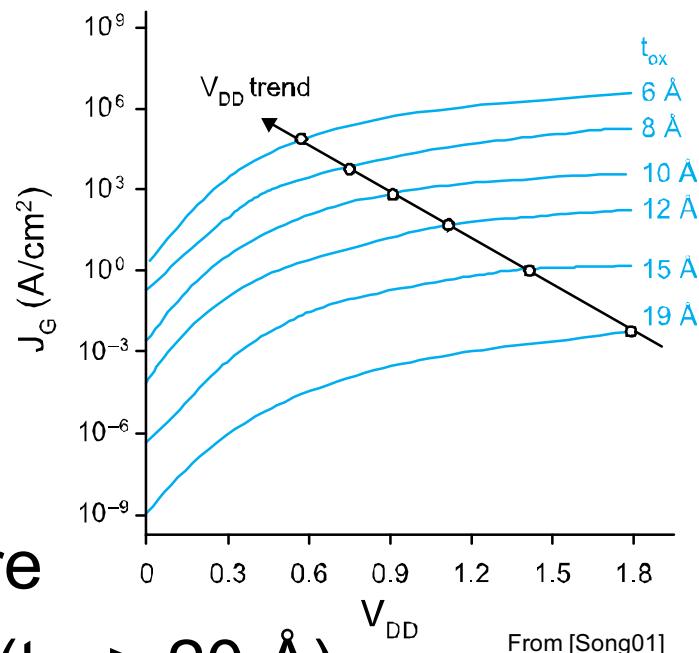
- $S \approx 100 \text{ mV/decade}$ @ room temperature

Gate Leakage

- Carriers tunnel through very thin gate oxides
- Exponentially sensitive to t_{ox} and V_{DD}

$$I_{\text{gate}} = WA \left(\frac{V_{DD}}{t_{ox}} \right)^2 e^{-B \frac{t_{ox}}{V_{DD}}}$$

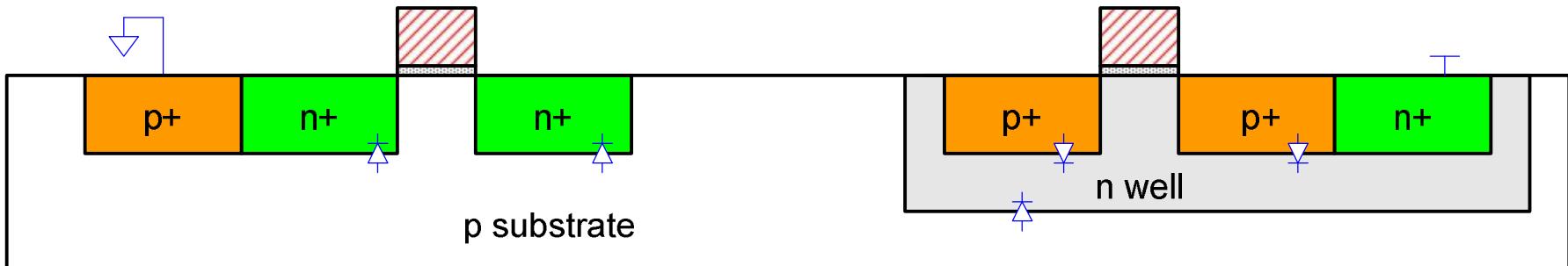
- A and B are tech constants
- Greater for electrons
 - So nMOS gates leak more
- Negligible for older processes ($t_{ox} > 20 \text{ \AA}$)
- Critically important at 65 nm and below ($t_{ox} \approx 10.5 \text{ \AA}$)



From [Song01]

Junction Leakage

- Reverse-biased p-n junctions have some leakage
 - Ordinary diode leakage
 - Band-to-band tunneling (BTBT)
 - Gate-induced drain leakage (GIDL)



Diode Leakage

- Reverse-biased p-n junctions have some leakage

$$I_D = I_S \left(e^{\frac{V_D}{v_T}} - 1 \right)$$

- I_D : diode current, $-I_s$: diode max reverse bias current
- At any significant negative diode voltage, $I_D = -I_s$
- I_s depends on doping levels
 - And area and perimeter of diffusion regions
 - Typically $< 1 \text{ fA}/\mu\text{m}^2$ (negligible)

Band-to-Band Tunneling

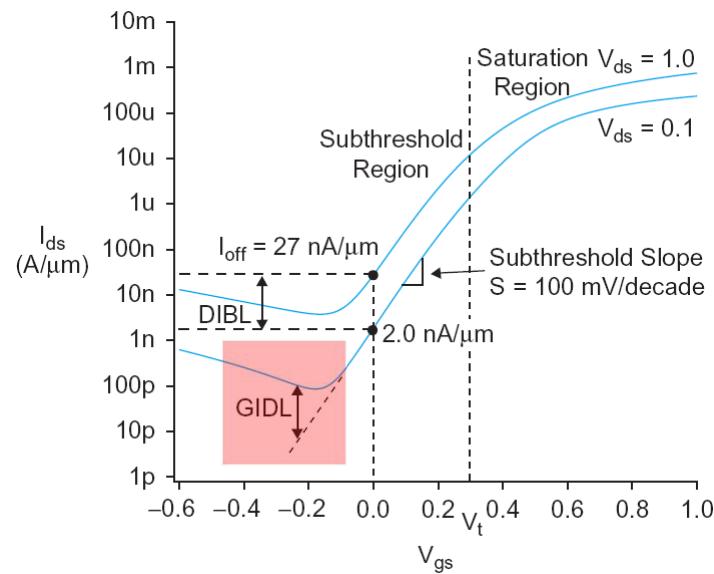
- Tunneling across heavily doped p-n junctions
 - Especially sidewall between drain & channel when halo doping is used
 - Halo doping: increasing the doping of the substrate or well near the ends of the channels to increase V_t
- Increases junction leakage to significant levels

$$I_{BTBT} = W X_j A \frac{E_j}{E_g^{0.5}} V_{dd} e^{-B \frac{E_g^{1.5}}{E_j}}$$
$$E_j = \sqrt{\frac{2qN_{halo}N_{sd}}{\varepsilon(N_{halo} + N_{sd})}} \left(V_{DD} + v_T \ln \frac{N_{halo}N_{sd}}{n_i^2} \right)$$

- X_j : sidewall junction depth
- E_g : bandgap voltage
- A, B: tech constants

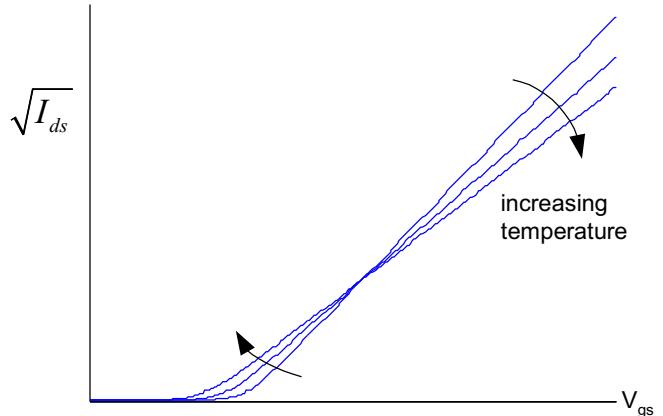
Gate-Induced Drain Leakage

- ❑ Occurs at overlap between gate and drain
 - Most pronounced when drain is at V_{DD} , gate is at a negative voltage
 - Thwarts efforts to reduce subthreshold leakage using a negative gate voltage



Temperature Sensitivity

- Increasing temperature
 - Reduces mobility
 - Reduces V_t
- I_{ON} decreases with temperature
- I_{OFF} increases with temperature
- Overall, is it better to operate in high temperature or low?

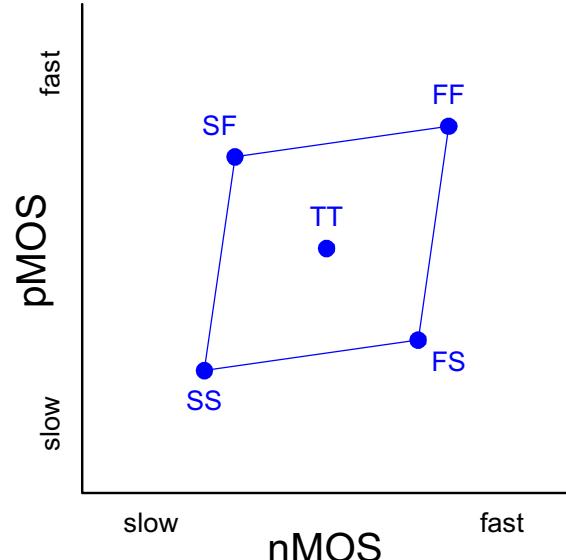


So What?

- So what if transistors are not ideal?
 - They still behave like switches.
- But these effects matter for...
 - Supply voltage choice
 - Logical effort
 - Quiescent power consumption
 - Pass transistors
 - Temperature of operation

Parameter Variation

- Transistors have uncertainty in parameters
 - Process: L_{eff} , V_t , t_{ox} of nMOS and pMOS
 - Vary around typical (T) values
- Fast (F)
 - L_{eff} : **short**
 - V_t : **low**
 - t_{ox} : **thin**
- Slow (S): opposite
- Not all parameters are independent for nMOS and pMOS



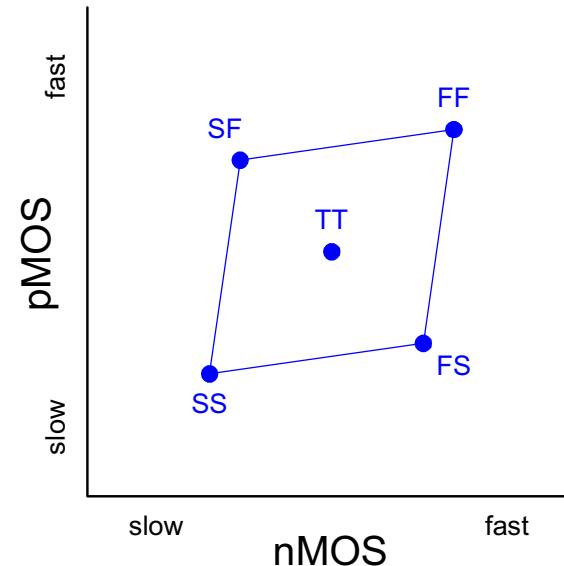
Environmental Variation

- V_{DD} and T also vary in time and space
- Fast:
 - V_{DD} : high
 - T: low

Corner	Voltage	Temperature
F	1.98	0 C
T	1.8	70 C
S	1.62	125 C

Process Corners

- Process corners describe worst case variations
 - If a design works in all corners, it will probably work for any variation.
- Describe corner with four letters (T, F, S)
 - nMOS speed
 - pMOS speed
 - Voltage
 - Temperature



Important Corners

- Some critical simulation corners include

Purpose	nMOS	pMOS	V_{DD}	Temp
Cycle time	S	S	S	S
Power	F	F	F	F
Subthreshold leakage	F	F	F	S

Geometry Dependence

- Designer's desired L and W: L_{drawn} and W_{drawn}
- Fabrication variation: X_L and X_W
- Lateral diffusion: L_D and W_D
- Effective length and width:
 - $L_{\text{eff}} = L_{\text{drawn}} + X_L - 2L_D$ and $W_{\text{eff}} = W_{\text{drawn}} + X_W - 2W_D$
- Effects:
 - A transistor drawn twice as long may have an effective length that is more than twice as great
→ current is less than half
 - A transistor drawn twice as wide may have an effective width that is more than twice as great → current is more than double

Geometry Dep. Cont.

- Other effects:
 - V_t increases for longer transistors → less current, slower speed
 - Long transistors experience less channel length modulation → less saturation current
- For current matching, use the same L and W for all devices

Homeworks

- Chapter 2 exercises: 2.2, 2.3, 2.6, 2.7, 2.8, 2.9, 2.10
 - Due date: 1402/8/23
-