وَمَا أُوتِيتُمْ مِنَ الْعِلْمِ إِلَّا قَلِيلًا

Analog IC Design

Lecture 04
MOSFET Large Signal Model

Dr. Hesham A. Omran

Integrated Circuits Lab (ICL)
Electronics and Communications Eng. Dept.
Faculty of Engineering
Ain Shams University

Why is the Transistor Different?

- We are used to two-terminal electronic devices
  - Diodes, resistors, capacitors, inductors
- The transistor is a three-terminal device
  - The voltage between two terminals controls the current flowing in the third terminal
  - Voltage controlled current source (VCCS)
- This feature enabled a multitude of applications that changed our life!
  - Analog signal amplification and processing
  - Digital logic and memory circuits

VCCS as an Amplifier

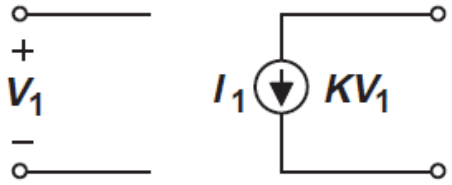
- Voltage controlled current source (VCCS): V_1 controls I_1

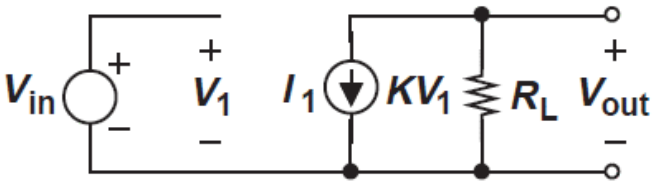
$$Transconductance = \frac{I_1}{V_1} = K$$

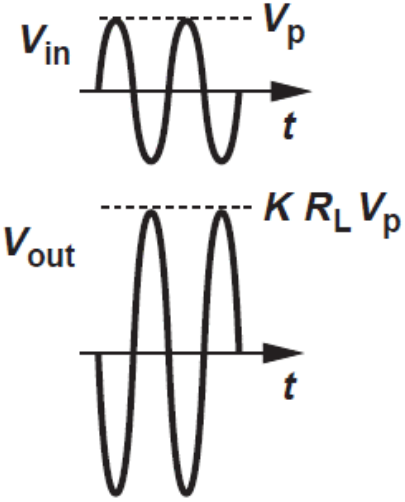
$$V_{in} = V_p \sin(\omega t)$$

$$V_{out} = -KV_1 \times R_L = -(KR_L)V_{in} = -(KR_L)V_p \sin(\omega t)$$

$$Voltage\ Gain = A_v = \frac{V_{out}}{V_{in}} = -KR_L$$





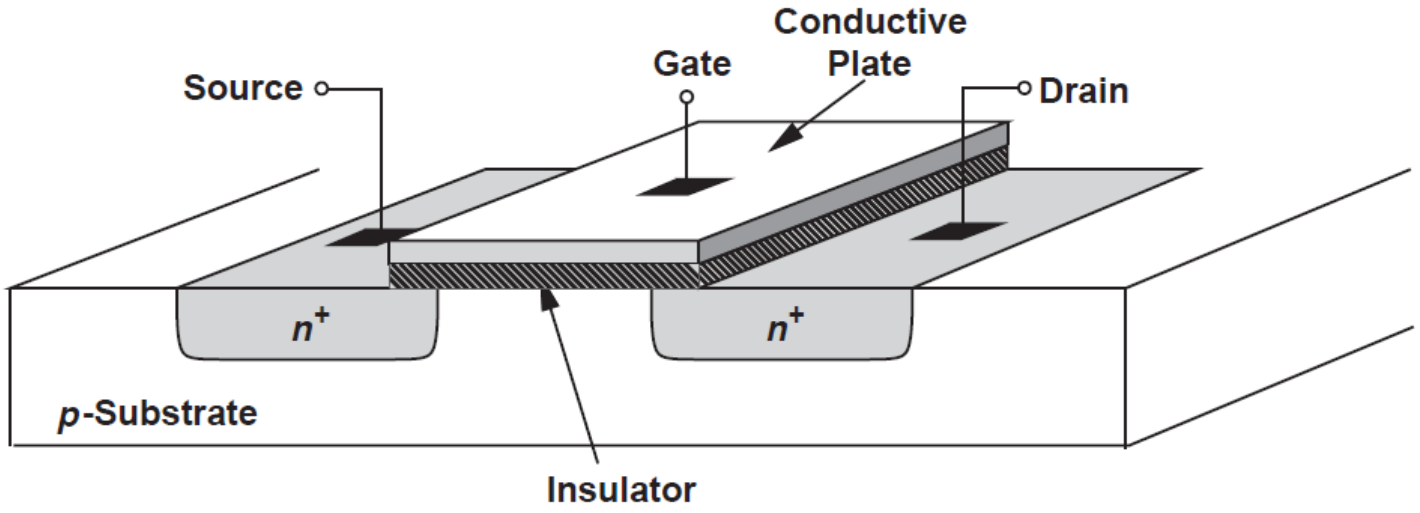


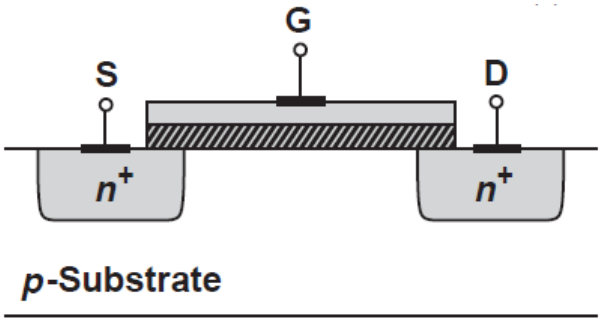
MOSFET

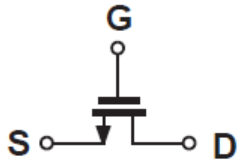
- MOSFET: Metal-oxide-semiconductor field-effect transistor
  - N-channel MOSFET: NMOS
  - P-channel MOSFET: PMOS
  - Complementary MOS (CMOS) technology: NMOS + PMOS
- A.k.a. insulated-gate FET or IGFET
- Simply, a VCCS
- The concept of MOSFET was patented in 1925
- But it was not successfully fabricated till 1960s
- CMOS technology became the dominant IC fabrication technology by the 1980s

N-Channel MOSFET Structure

- MOSFET: Metal-oxide-semiconductor field-effect transistor
- Three-terminal device: Gate (G), Source (S), and Drain (D)
- Substrate/Bulk/Body (S/B) can be treated as a fourth terminal



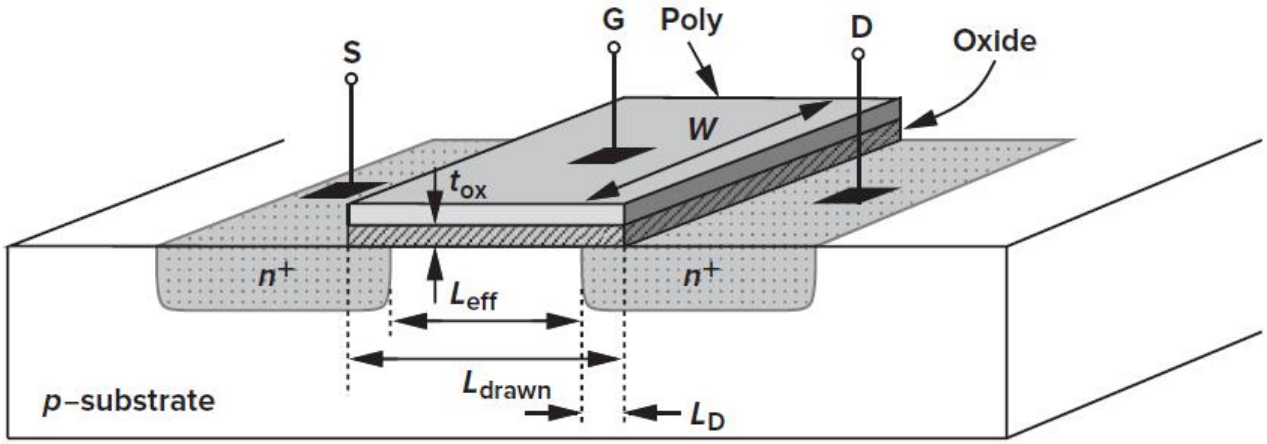




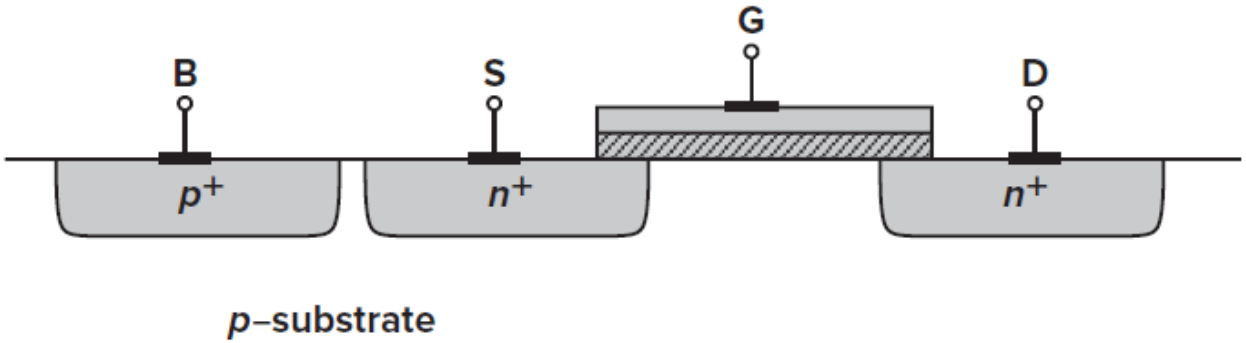
[Razavi, 2014]

N-Channel MOSFET Structure

- The transistor is primarily a VCCS



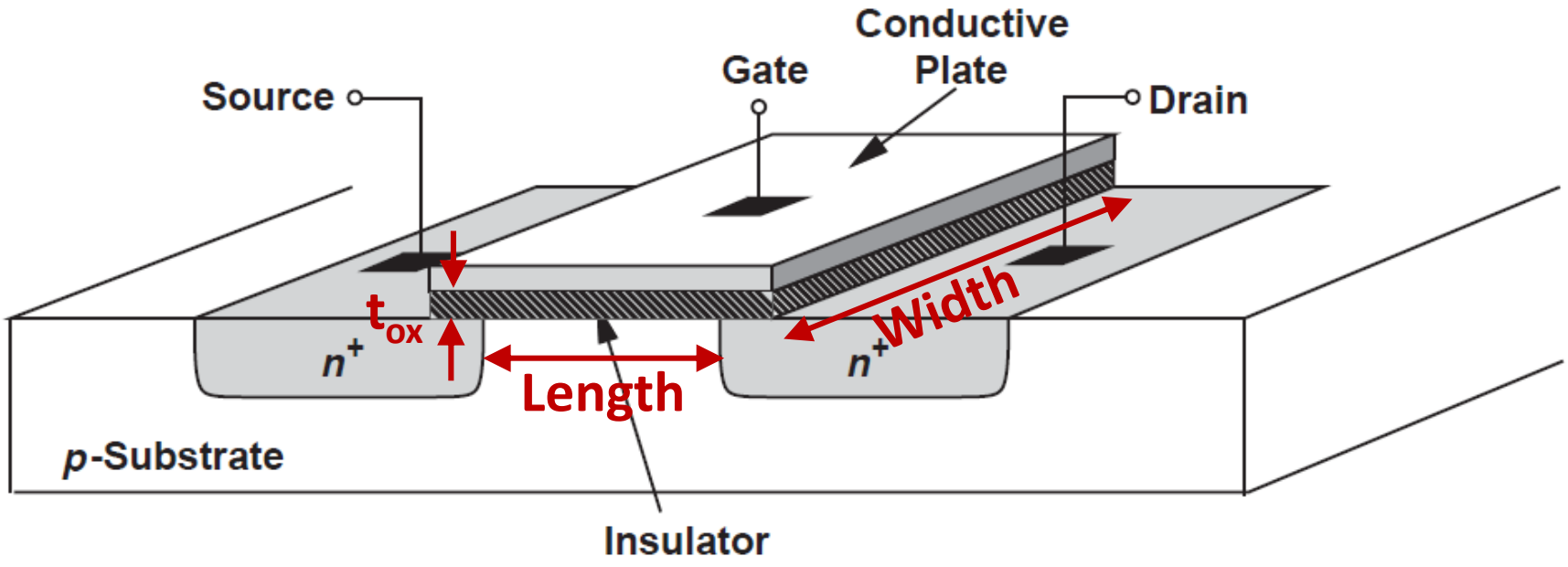
- MOSFET is a four-terminal device



[Razavi, 2017]

MOSFET Dimensions

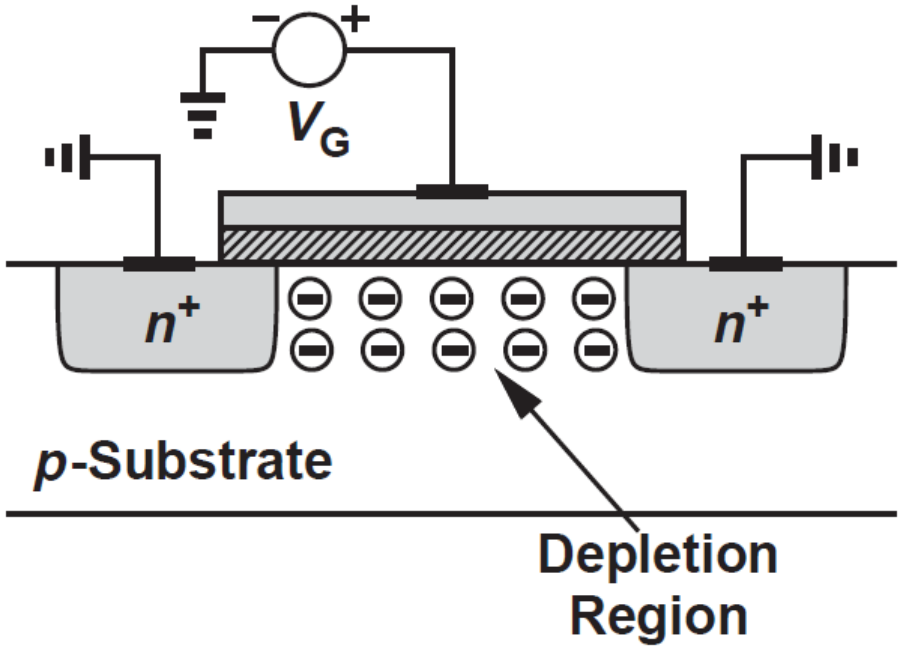
- Channel length: $L \sim 10nm - 10\mu m$
- Channel width: $W \sim 50nm - 100\mu m$
- Oxide thickness: $t_{ox} \sim 1nm - 10nm$
- Gate formed of metal or polysilicon



[Razavi, 2014]

Depletion

- The device acts as a capacitor: positive charge on the gate is mirrored by negative charge in the substrate
- The positive charge on the gate repels the holes in the substrate
  - Fixed negative ions are exposed (uncovered)
  - A depletion region is created



[Razavi, 2014]

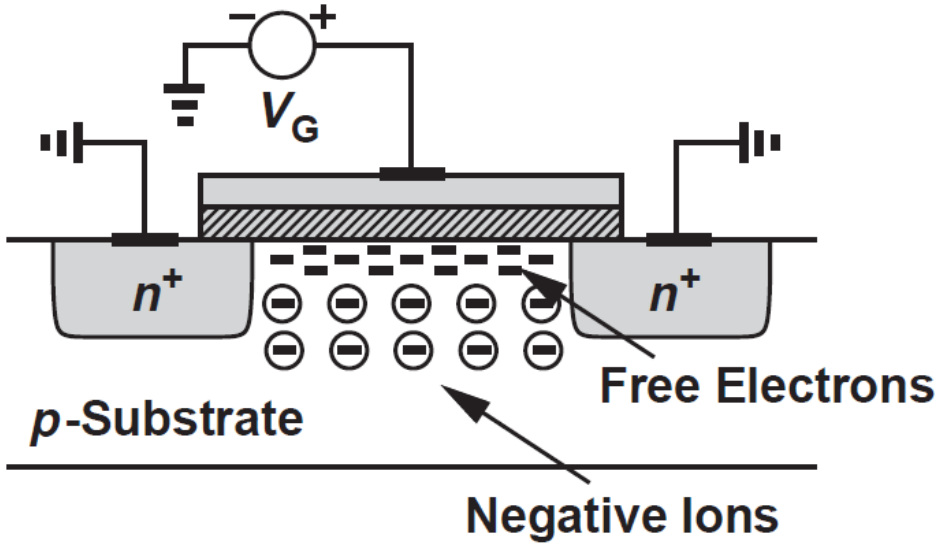
Inversion and Channel Formation

- N-type channel region (inversion layer) formed at

$$V_{GS} > V_{TH}$$

$$V_{GS} = V_{TH} + V_{ov}$$

  - Threshold voltage: $V_{TH} \sim 0.3V - 1V$
  - Overdrive voltage: $V_{ov} \sim 0.05V - 0.5V$ (for analog circuits)
  - Electrons are provided by the n+ source and drain regions



[Razavi, 2014]

Charge in Channel

$$C_{gate} = \frac{\epsilon_{ox} A}{d} = \frac{\epsilon_{ox} WL}{t_{ox}} = C_{ox} WL$$

- For SiO_2

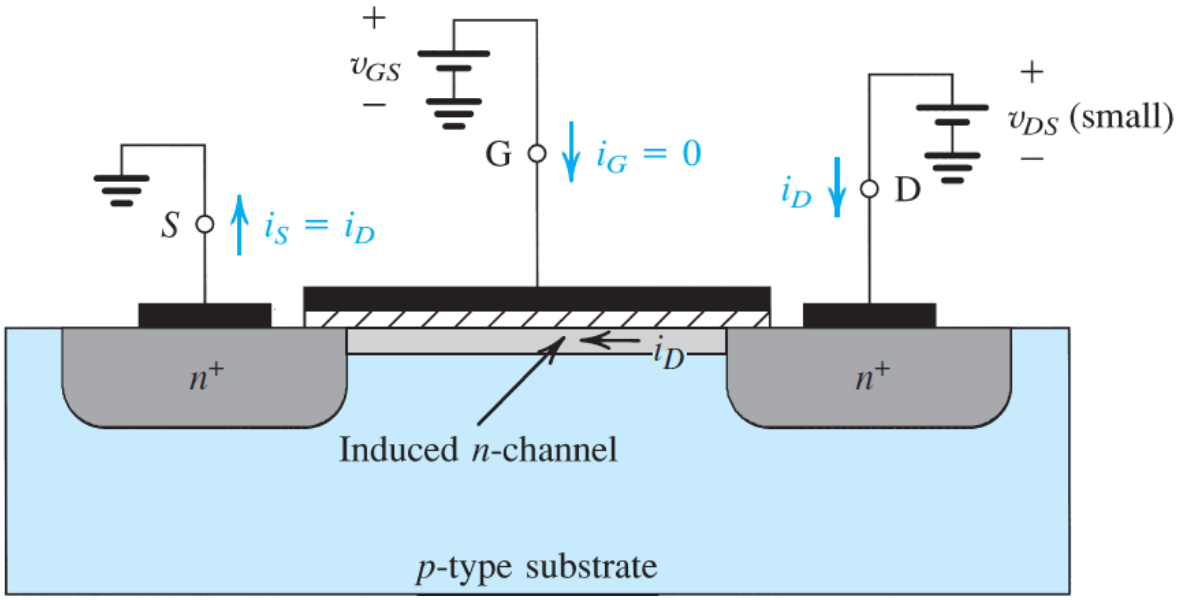
$$\epsilon_{ox} = \epsilon_r \epsilon_o = 3.9 \times 8.854 \times 10^{-12} \frac{F}{m}$$

- Example: if $t_{ox} = 4nm \rightarrow C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \approx 8.6 \frac{fF}{\mu m^2}$

$$|Q| = CV = C_{ox} WL \cdot (V_{GS} - V_{TH}) = C_{ox} WL \cdot V_{ov}$$

Linear Region

- Small V_{DS}: We assume the channel is uniform
- MOSFET acts as a voltage controlled resistor (VCR)



[Sedra/Smith, 2015]

Linear Region

$$|Q| = CV = C_{ox}WL \cdot (V_{GS} - V_{TH}) = C_{ox}WL \cdot V_{ov}$$

$$Electric\ Field = |E| = \frac{V_{DS}}{L}$$

$$Carrier\ Velocity = |v| = \mu_n |E| = \mu_n \frac{V_{DS}}{L}$$

$$Drain\ Current = I_D = \frac{Q}{t} = C_{ox}W\left(\frac{L}{t}\right) \cdot V_{ov} = C_{ox}W \cdot v \cdot V_{ov}$$

$$I_D = \mu_n C_{ox} \frac{W}{L} \cdot V_{ov} \cdot V_{DS} = \frac{V_{DS}}{R_{DS}}$$

$$R_{DS} = \frac{1}{\mu_n C_{ox} \frac{W}{L} \cdot V_{ov}} = \frac{1}{k_n' \frac{W}{L} V_{ov}} = \frac{1}{k_n V_{ov}} = \frac{1}{\beta_n V_{ov}}$$

$$Aspect\ Ratio = \frac{W}{L}$$

Linear Region (Deep Triode)

$$I_D = \mu_n C_{ox} \frac{W}{L} \cdot V_{ov} \cdot V_{DS} = \frac{V_{DS}}{R_{DS}}$$

$$R_{DS} = \frac{1}{\mu_n C_{ox} \frac{W}{L} \cdot V_{ov}} = \frac{1}{k_n' \frac{W}{L} V_{ov}} = \frac{1}{k_n V_{ov}} = \frac{1}{\beta_n V_{ov}}$$

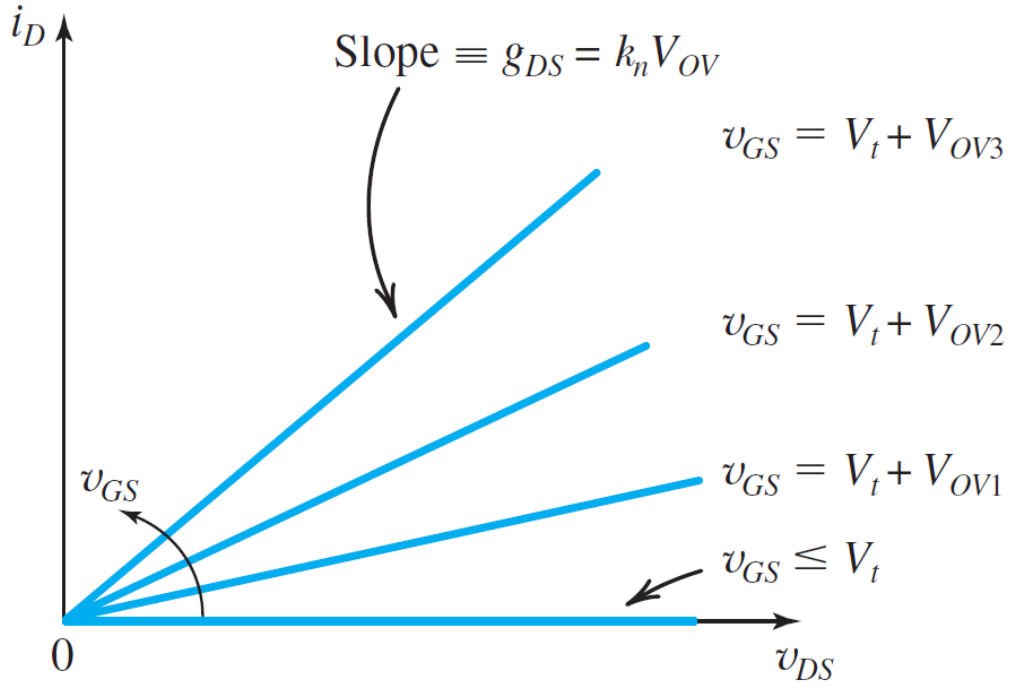
$$Aspect\ Ratio = \frac{W}{L}$$

[Razavi, 2017]

Linear Region (Deep Triode)

- Small V_{DS}: We assume the channel is uniform
- MOSFET acts as a voltage controlled resistor (VCR)

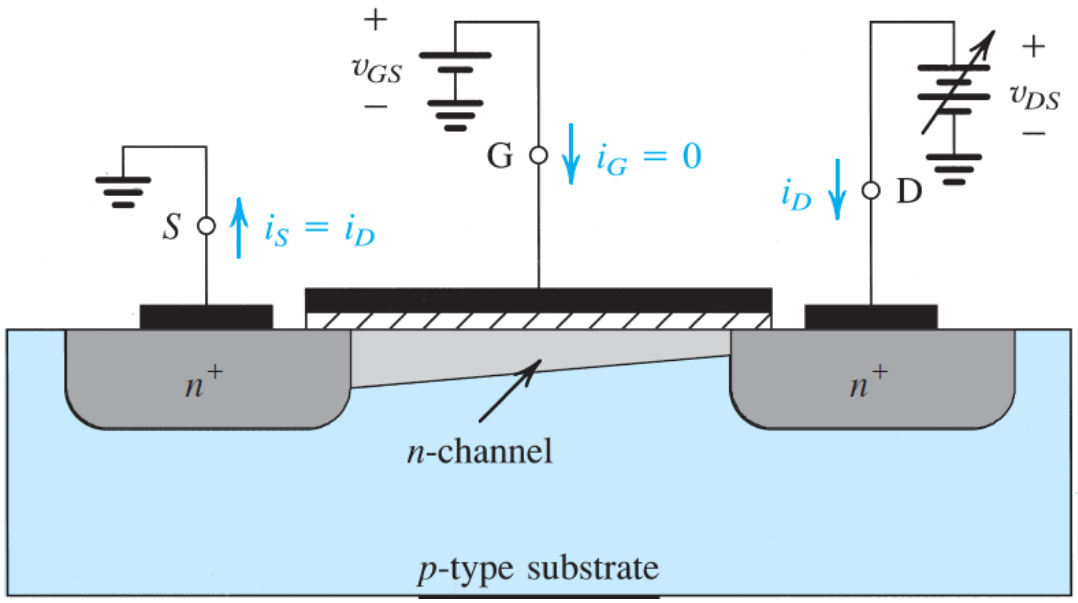
$$R_{DS} = \frac{1}{G_{DS}} = \frac{1}{\mu_n C_{ox} \frac{W}{L} \cdot V_{ov}} = \frac{1}{k_n' \frac{W}{L} V_{ov}} = \frac{1}{k_n V_{ov}} = \frac{1}{\beta_n V_{ov}}$$



[Sedra/Smith, 2015]

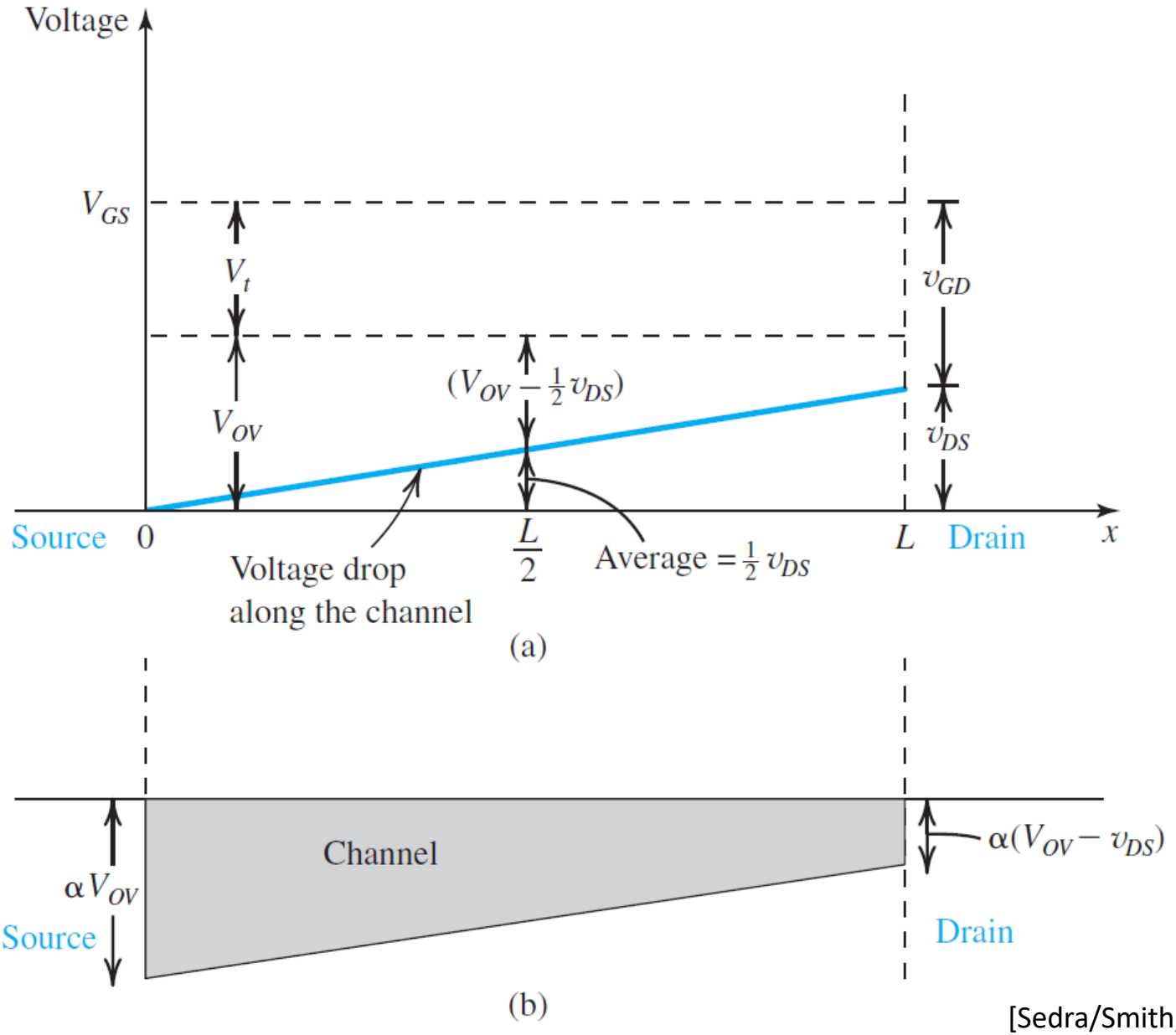
Triode Region

- V_{DS} increases: The channel becomes tapered
- Voltage at source side: $V_{GS} - 0 = V_{GS} = V_T + V_{ov}$
  - If $V_{GS} > V_T$ or $V_{ov} > 0$: The channel exists at source
- Voltage at drain side: $V_{GS} - V_{DS} = V_{GD} = V_T + (V_{ov} - V_{DS})$
  - If $V_{GD} > V_T$ or $V_{ov} > V_{DS}$: The channel exists at drain



[Sedra/Smith, 2015]

Triode Region

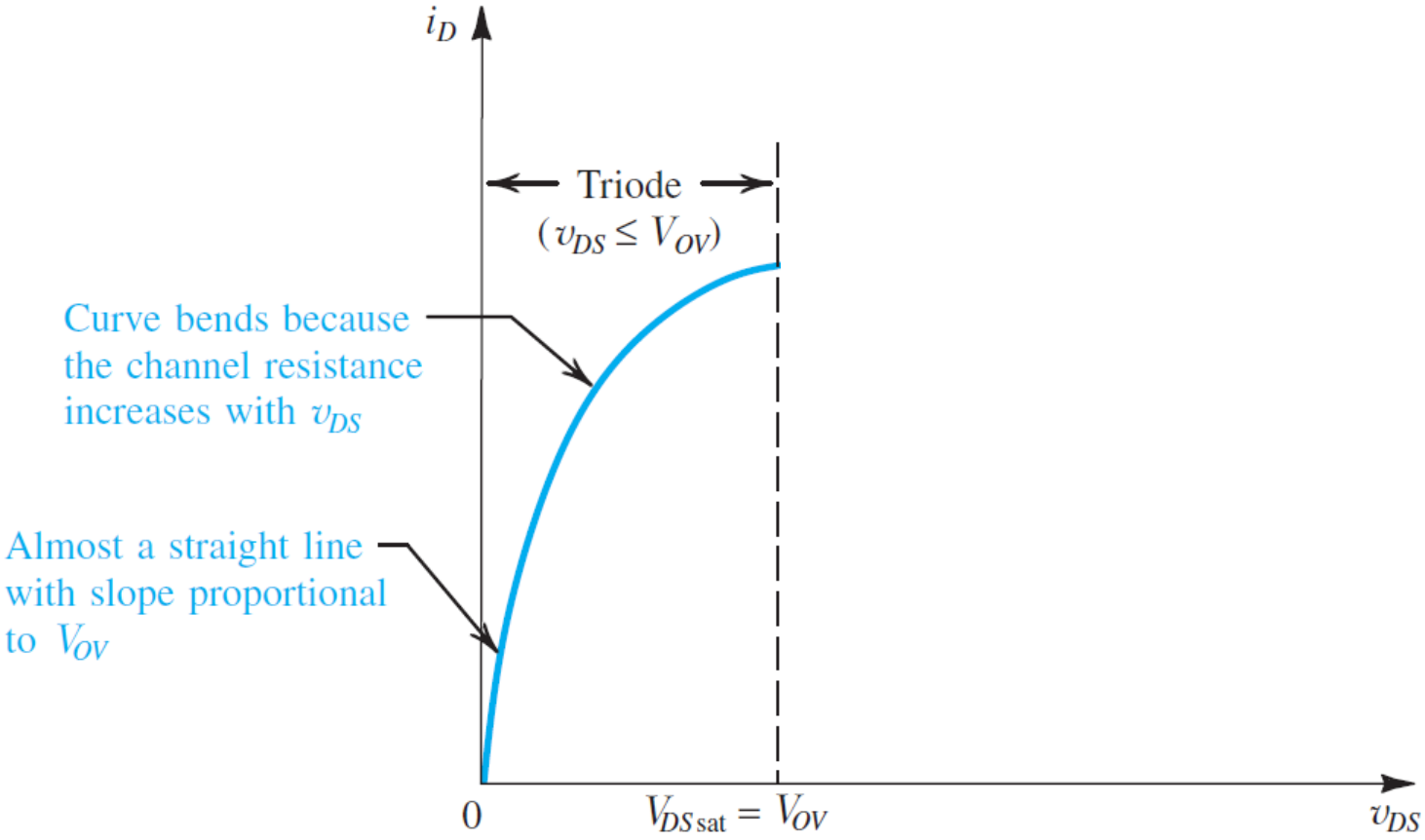


[Sedra/Smith, 2015]

Triode Region

- Replace V_{ov} with $(V_{ov})_{average} = V_{ov} - \frac{V_{DS}}{2}$

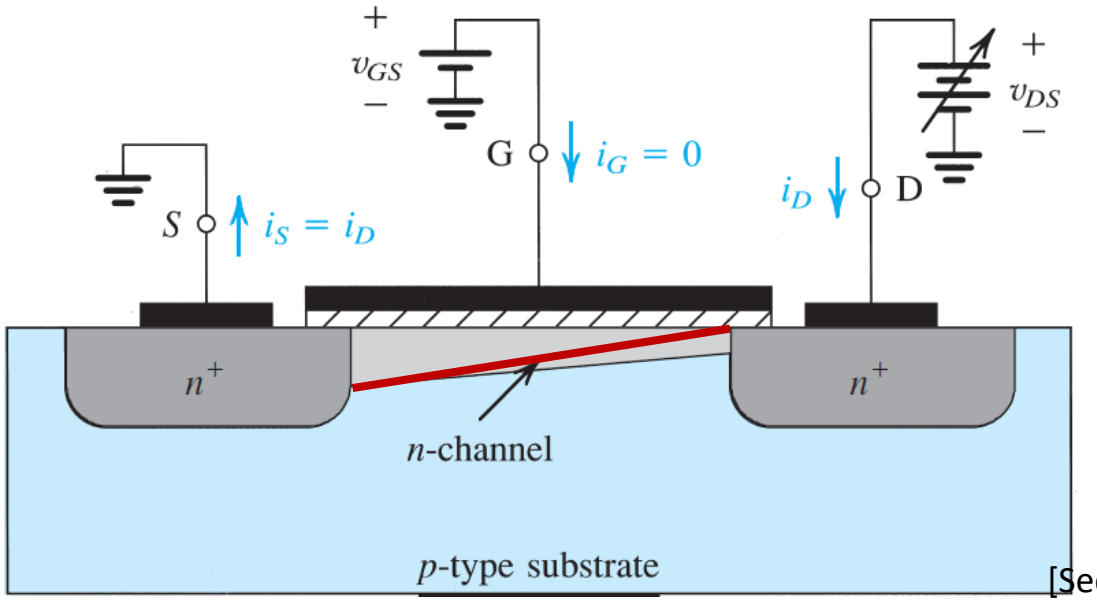
$$I_D = \mu_n C_{ox} \frac{W}{L} \cdot \left(V_{ov} - \frac{V_{DS}}{2}\right) \cdot V_{DS} = \mu_n C_{ox} \frac{W}{L} \cdot \left(V_{ov} V_{DS} - \frac{V_{DS}^2}{2}\right)$$



[Sedra/Smith, 2015]

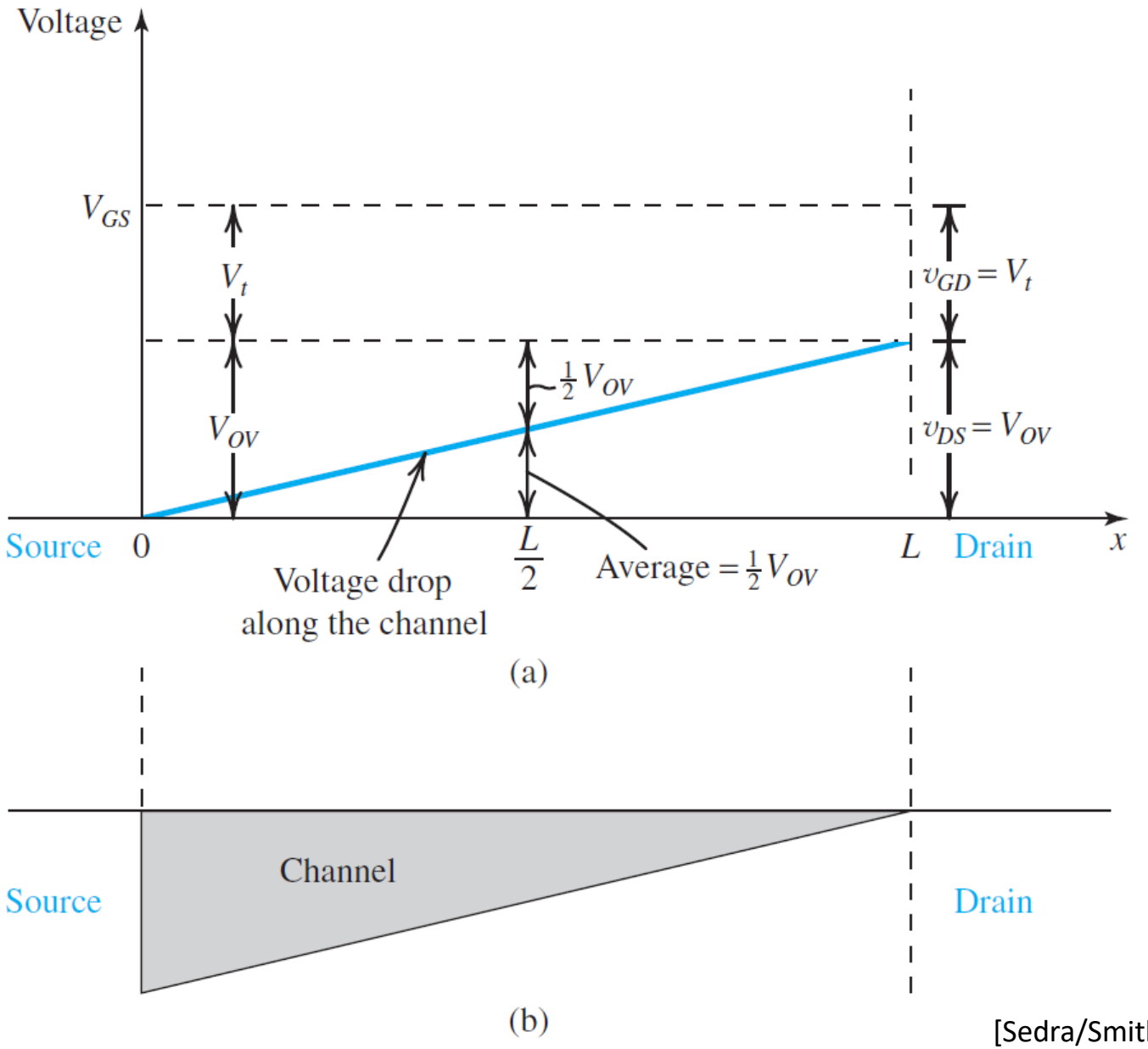
Pinch-Off (Saturation)

- $V_{GD} = V_{GS} - V_{DS} \leq V_{TH}$ → $V_{DS} \geq V_{GS} - V_{TH} = V_{ov}$
  - No channel at drain side
  - V_{DS} has no more control on the shape and charge of the channel
- Voltage across channel is constant $= V_{GS} - V_{TH} = V_{ov}$
  - Extra V_{DS} falls on the small region between channel and drain
  - Current remains constant (saturates) → **VCCS**



[Sedra/Smith, 2015]

Pinch-Off (Saturation)

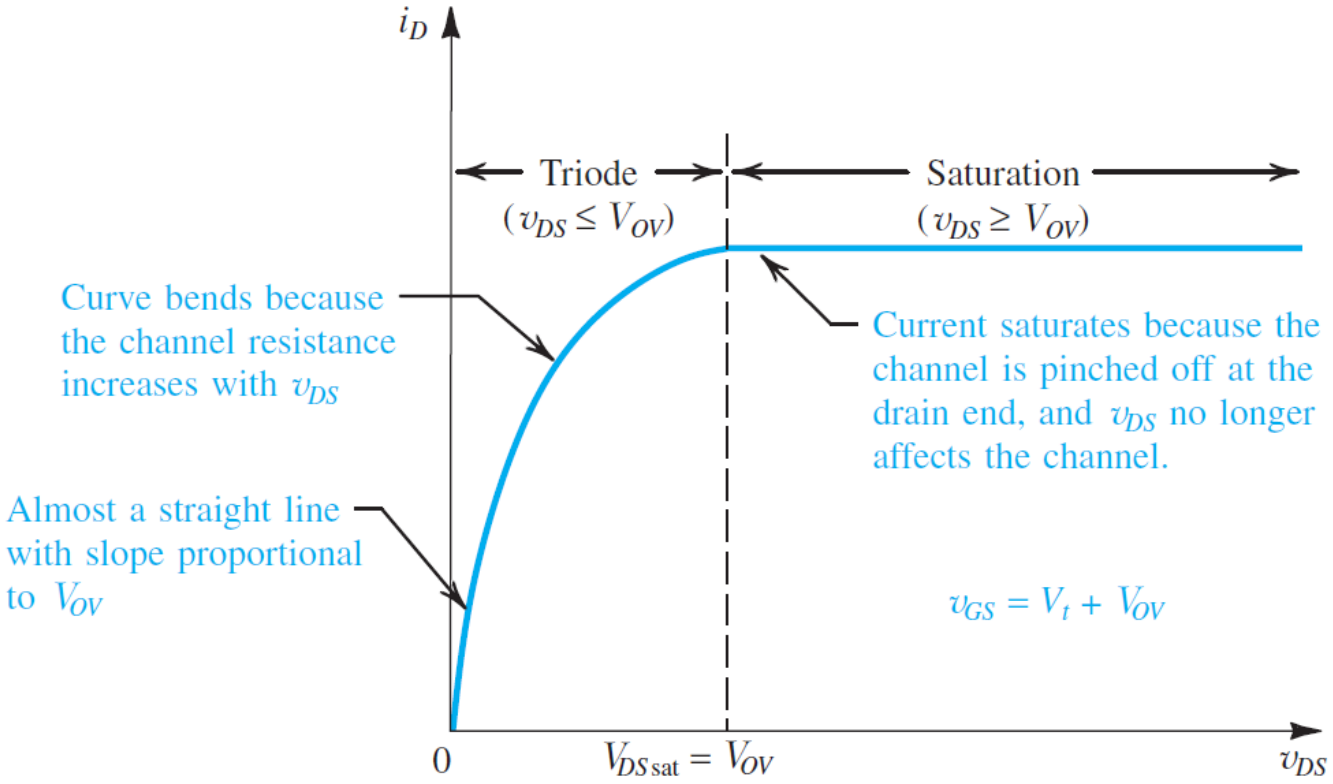


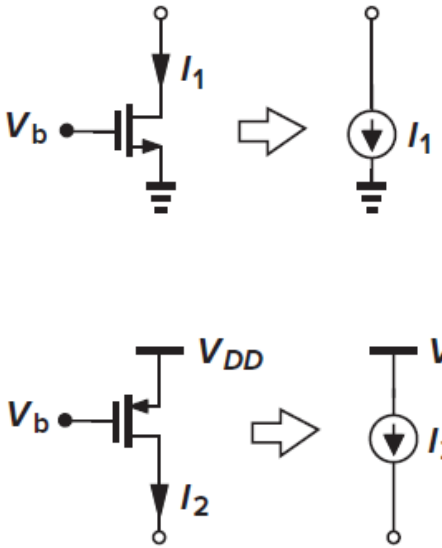
[Sedra/Smith, 2015]

Pinch-Off (Saturation)

- Replace V_{DS} with $(V_{GS} - V_{TH}) = V_{ov}$

$$I_D = \mu_n C_{ox} \frac{W}{L} \cdot \left(V_{ov} V_{DS} - \frac{V_{DS}^2}{2} \right) = \frac{\mu_n C_{ox}}{2} \frac{W}{L} \cdot V_{ov}^2$$





[Sedra/Smith, 2015]

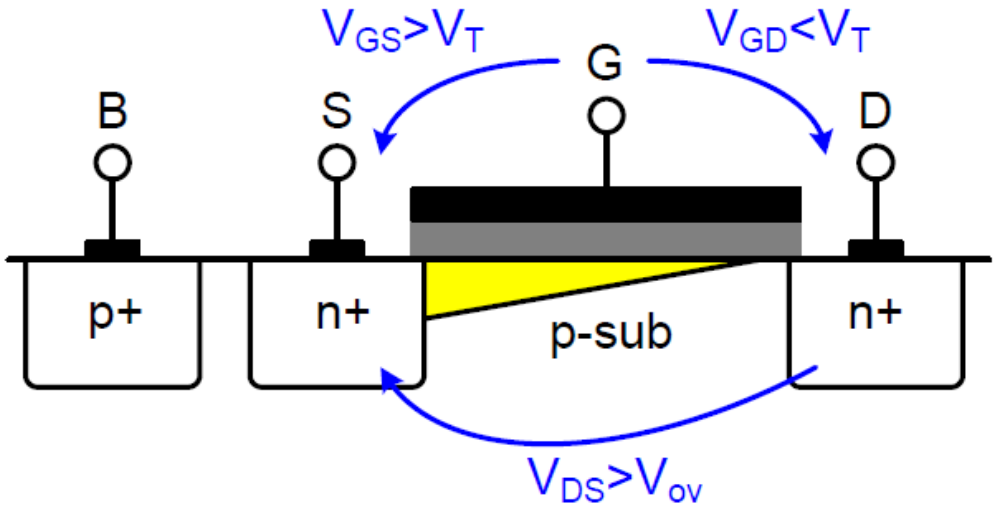
Pinch-Off (Saturation)

- The channel is pinched off if the difference between the gate and drain voltages is not sufficient to create an inversion layer

$$V_{DS} \geq V_{ov} = (V_{GS} - V_{TH})$$

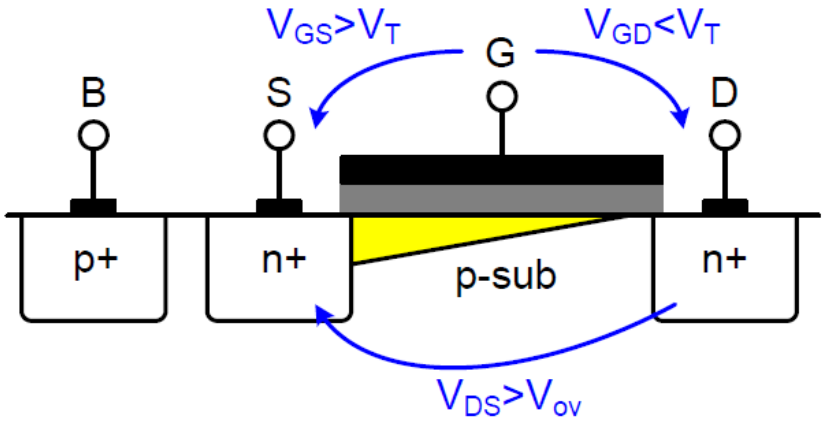
- Square-law (long channel MOS)

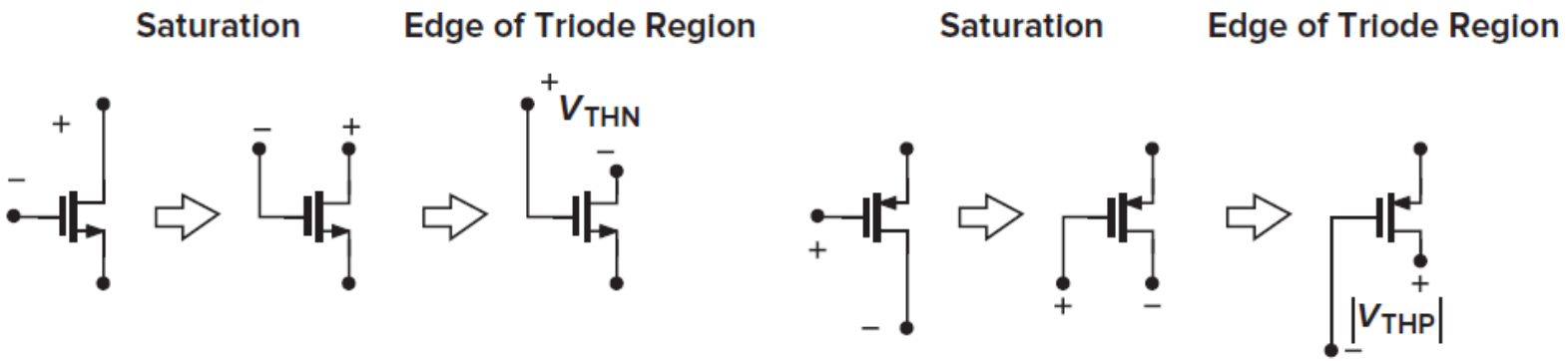
$$I_D = \frac{\mu_n C_{ox}}{2} \frac{W}{L} \cdot V_{ov}^2$$



Edge of Saturation / Triode

- The channel is pinched off if the difference between the gate and drain voltages is not sufficient to create an inversion layer
  - For NMOS: $V_G - V_D = V_{GD} \leq V_{THN}$
  - For PMOS: $V_D - V_G = V_{DG} \leq |V_{THP}|$

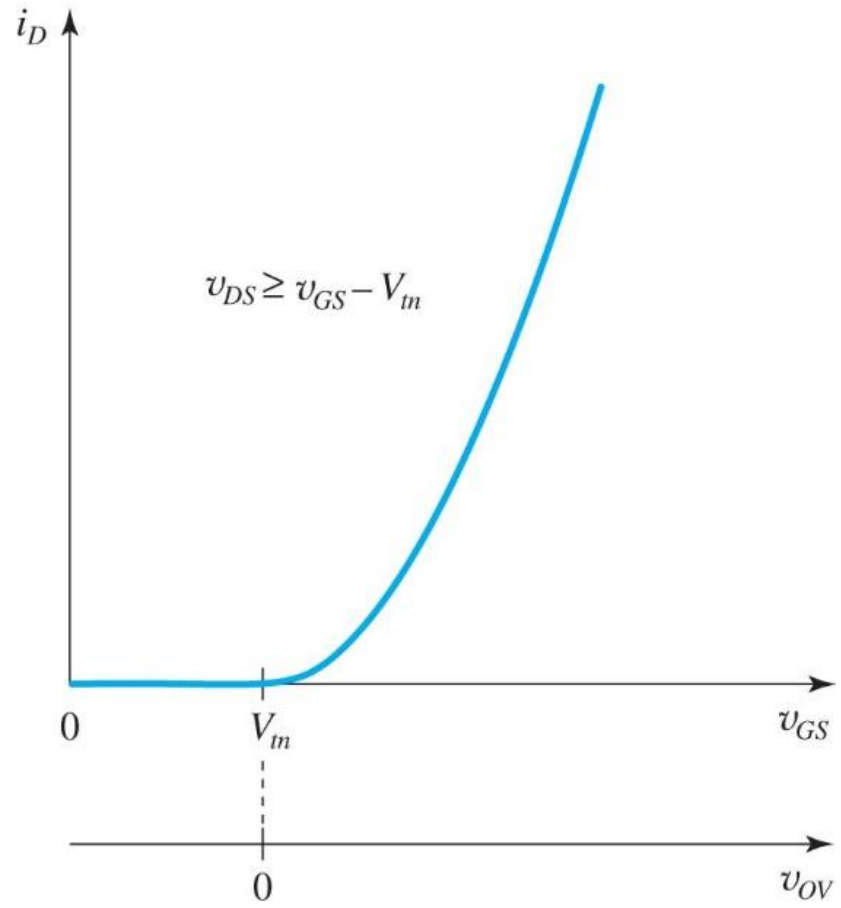


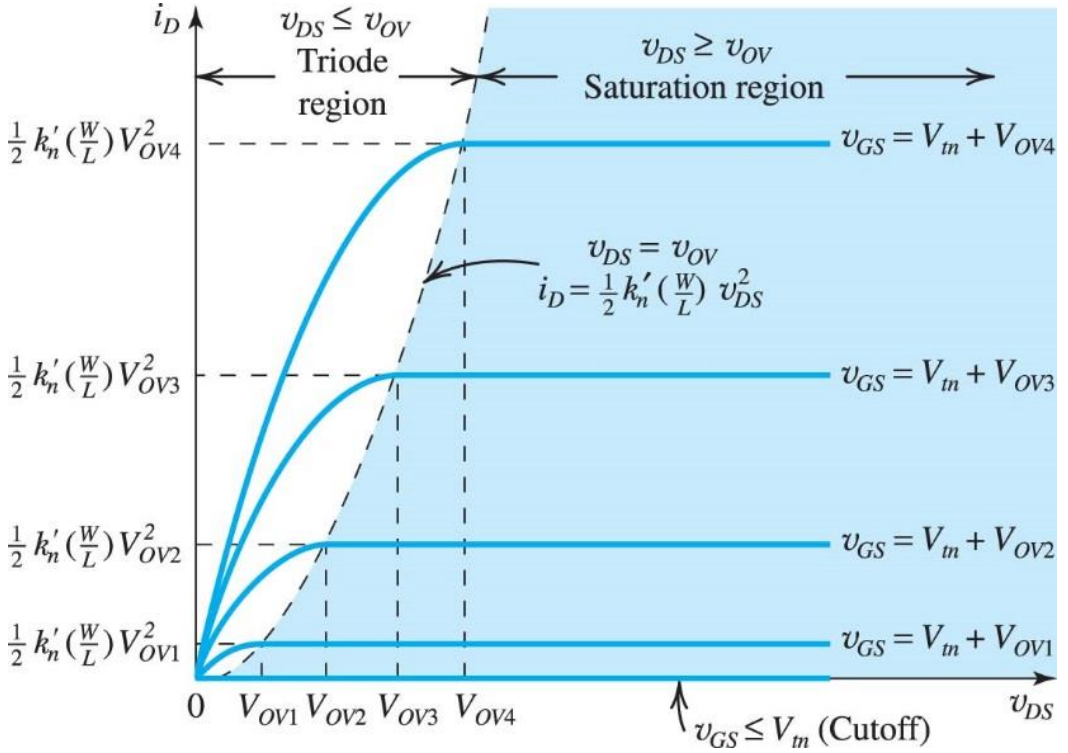


[Razavi, 2014]

IV Characteristics

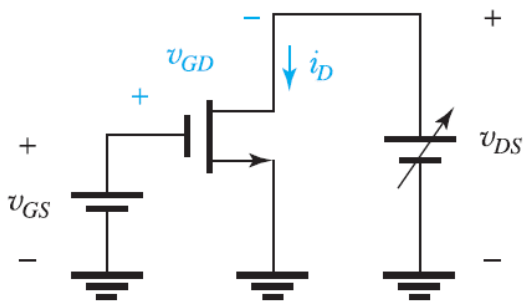
$$I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}\cdot V_{ov}^2 = \frac{\mu_n C_{ox}}{2}\frac{W}{L}\cdot (V_{GS}-V_T)^2$$

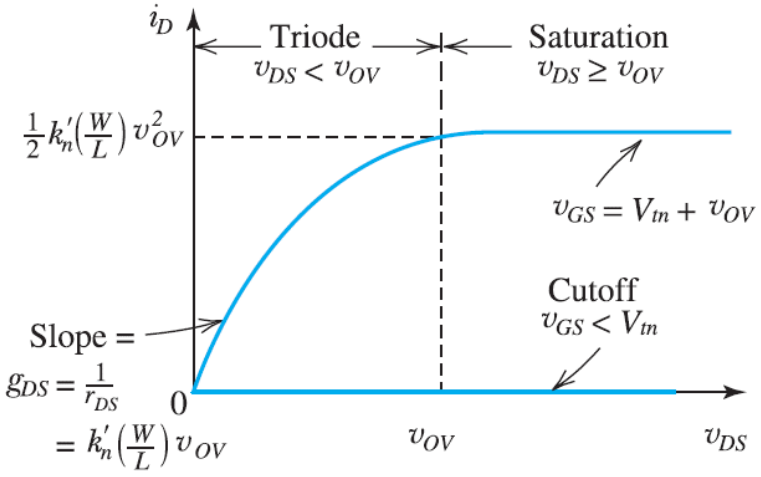




[Sedra/Smith, 2015]

Regions of Operation Summary





- $v_{GS} < V_{tn}$: no channel; transistor in cutoff; $i_D = 0$
- $v_{GS} = V_{tn} + v_{OV}$: a channel is induced; transistor operates in the triode region or the saturation region depending on whether the channel is continuous or pinched off at the drain end;

Triode Region

Continuous channel, obtained by:

$$v_{GD} > V_{tn}$$

or equivalently:

$$v_{DS} < v_{OV}$$

Then,

$$i_D = k_n'\left(\frac{W}{L}\right)\left[(v_{GS} - V_{tn})v_{DS} - \frac{1}{2}v_{DS}^2\right]$$

or equivalently,

$$i_D = k_n'\left(\frac{W}{L}\right)\left(v_{OV} - \frac{1}{2}v_{DS}\right)v_{DS}$$

Saturation Region

Pinched-off channel, obtained by:

$$v_{GD} \leq V_{tn}$$

or equivalently:

$$v_{DS} \geq v_{OV}$$

Then

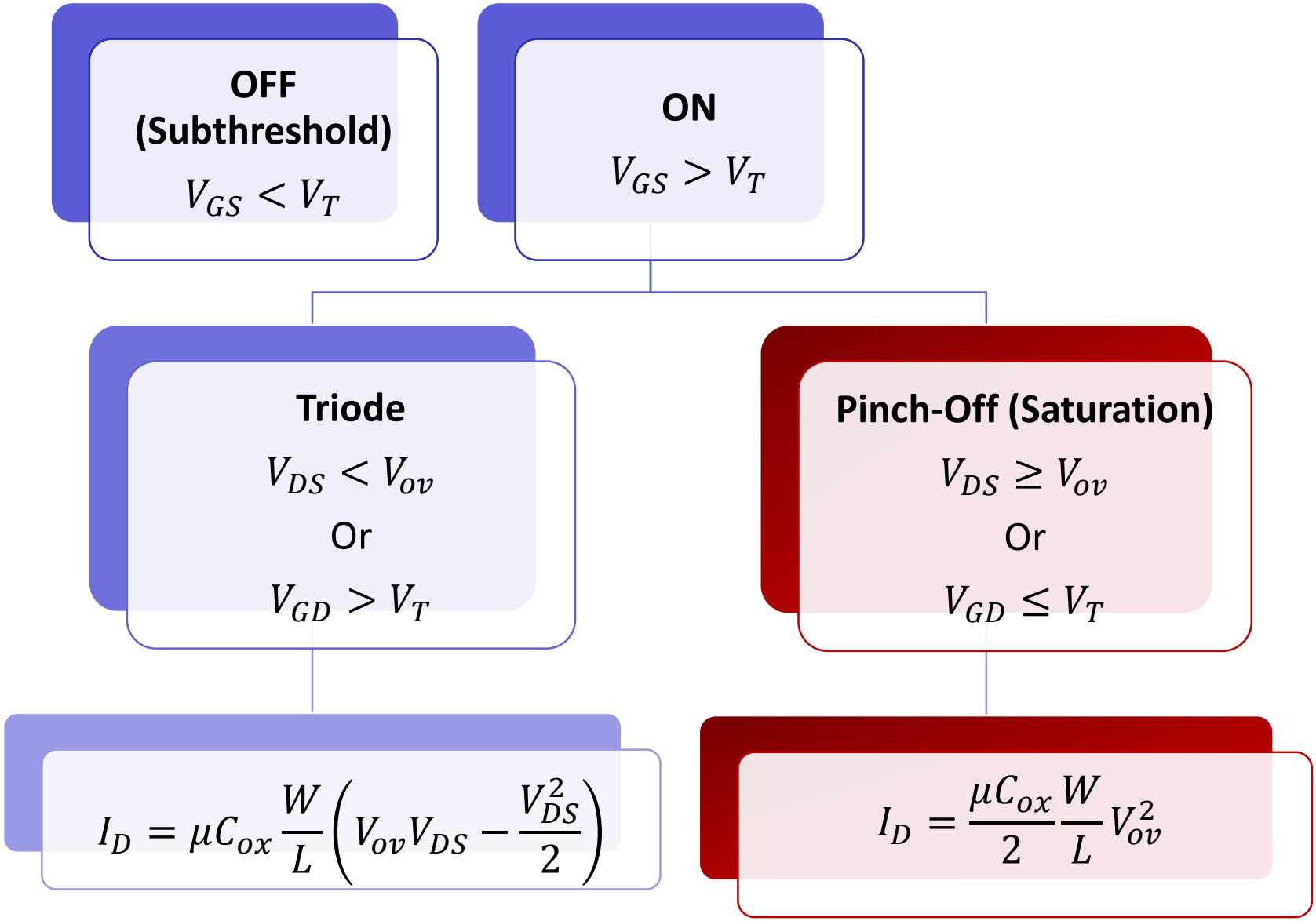
$$i_D = \frac{1}{2}k_n'\left(\frac{W}{L}\right)(v_{GS} - V_{tn})^2$$

or equivalently,

$$i_D = \frac{1}{2}k_n'\left(\frac{W}{L}\right)v_{OV}^2$$

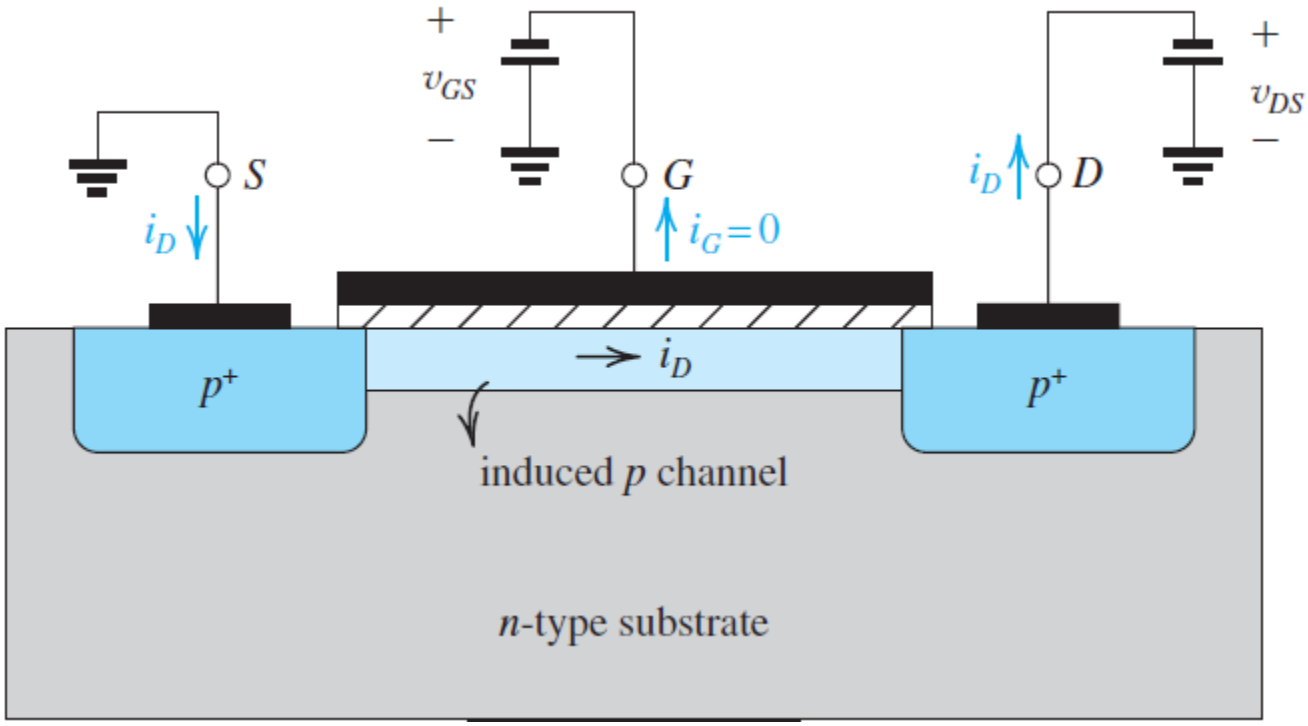
[Sedra/Smith, 2015]

Regions of Operation Summary



P-Channel MOSFET (PMOS)

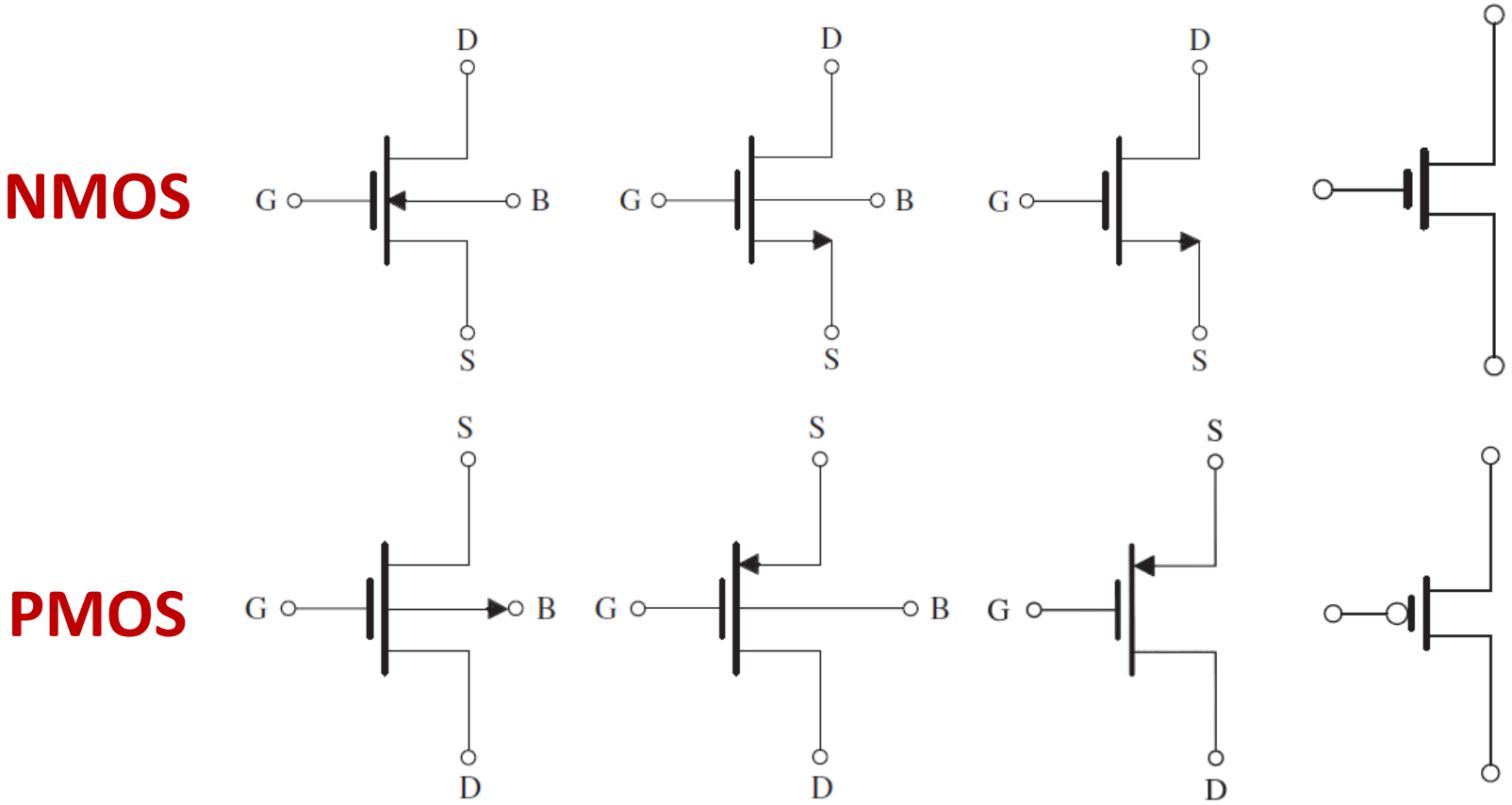
- Electrons have higher mobility than holes
- For same W/L, NMOS current is several times higher than PMOS



[Sedra/Smith, 2015]

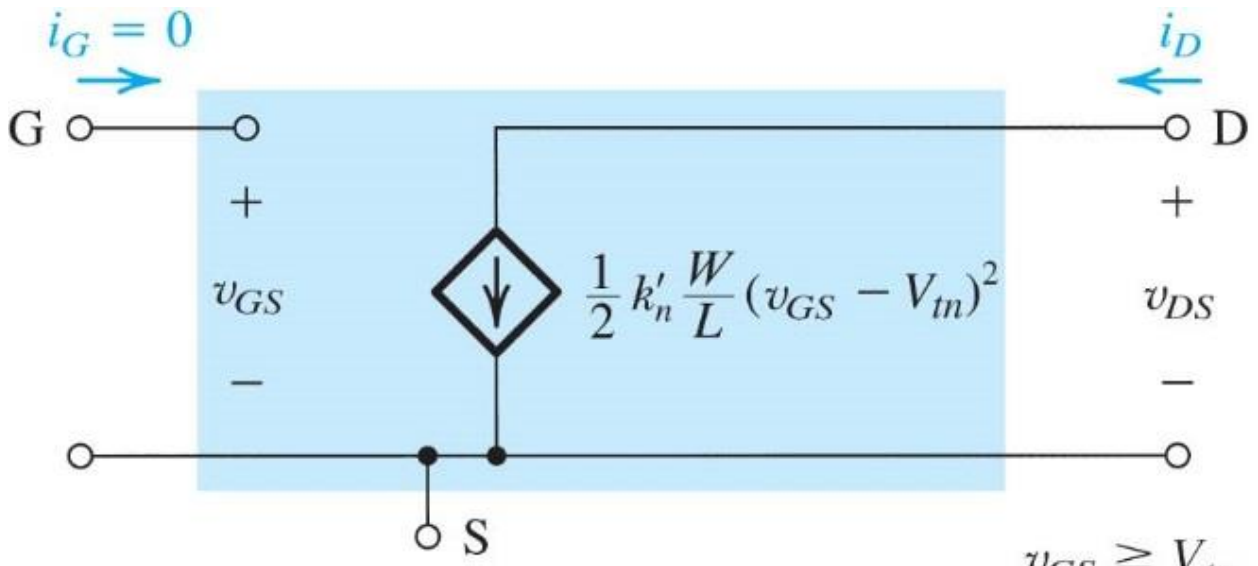
MOSFET Symbols

- S/D junction diodes must be reverse-biased under all conditions
  - NMOS bulk connected to most negative potential (ground)
  - PMOS bulk connected to most positive potential (VDD)



[Sedra/Smith, 2015]

Large Signal Model in Saturation



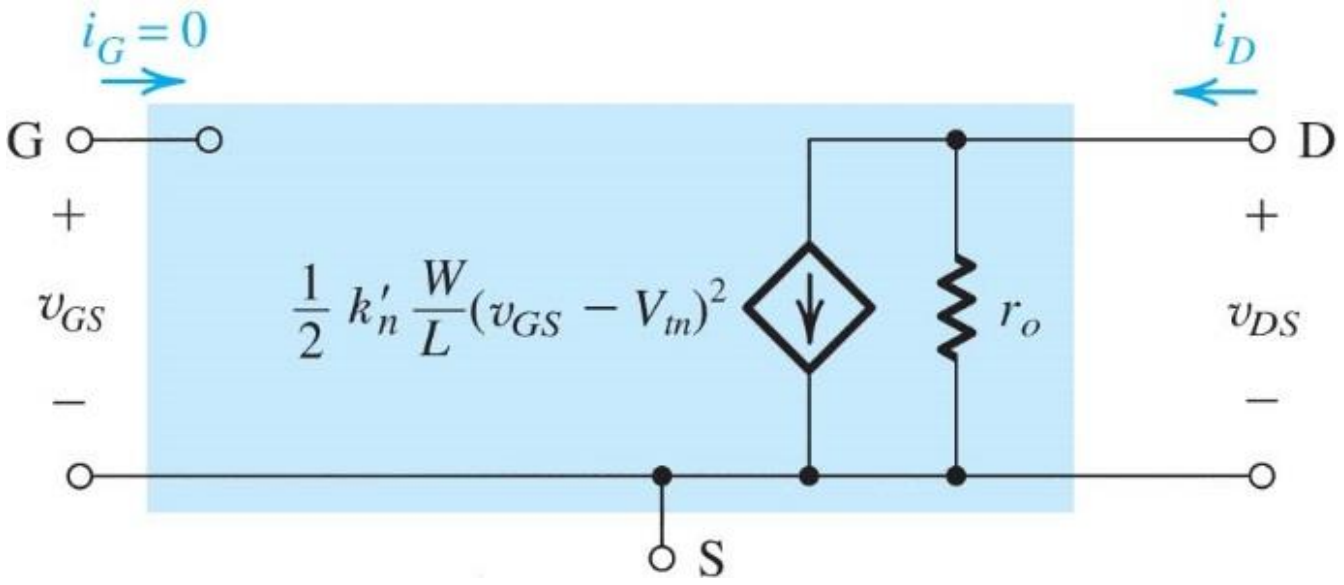
[Sedra/Smith, 2015]

Large Signal Model with Finite Output Res

- The VCCS is not ideal: There is some dependence on V_{DS}

$$I_D = I_{DS} + \frac{V_{DS}}{r_o} = I_{DS}\left(1 + \frac{V_{DS}/I_{DS}}{r_o}\right)$$

$$I_{DS} = \frac{\mu C_{ox}}{2}\frac{W}{L}V_{ov}^2$$



[Sedra/Smith, 2015]

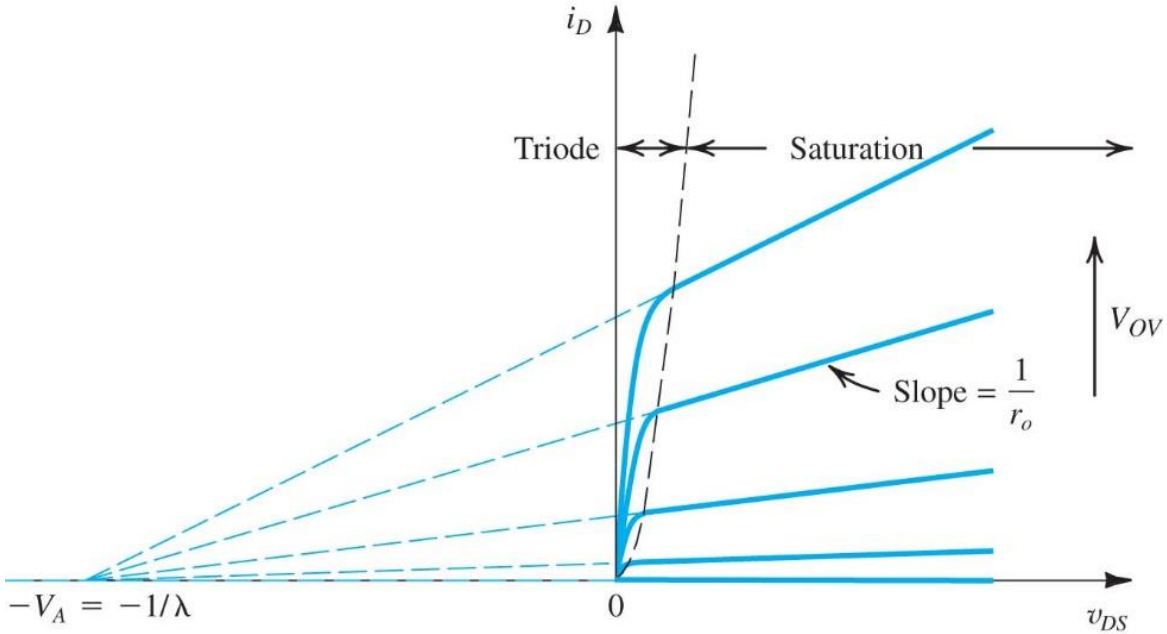
Channel Length Modulation (CLM)

- The VCCS is not ideal: There is some dependence on V_{DS}

$$r_o = \frac{V_A}{I_{DS}} = \frac{1}{\lambda I_{DS}}$$

λ: Channel length modulation coefficient ($\lambda \propto 1/L$)

$$I_D = I_{DS} + \frac{V_{DS}}{r_o} = I_{DS}\left(1 + \frac{V_{DS}/I_{DS}}{r_o}\right) = \frac{\mu C_{ox}}{2}\frac{W}{L}V_{ov}^2(1 + \lambda V_{DS})$$

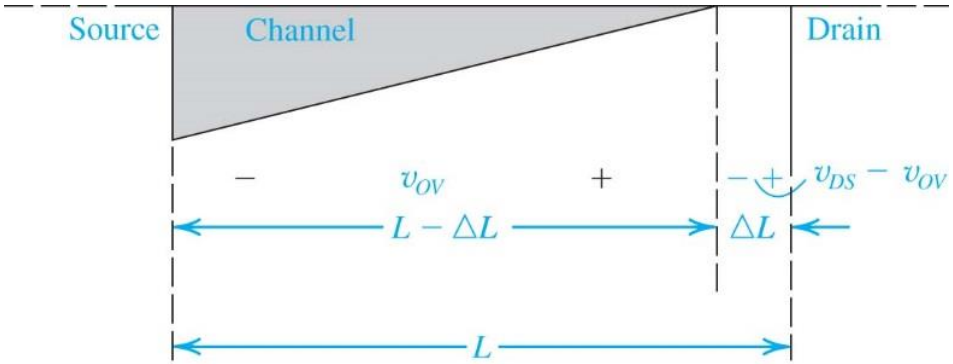


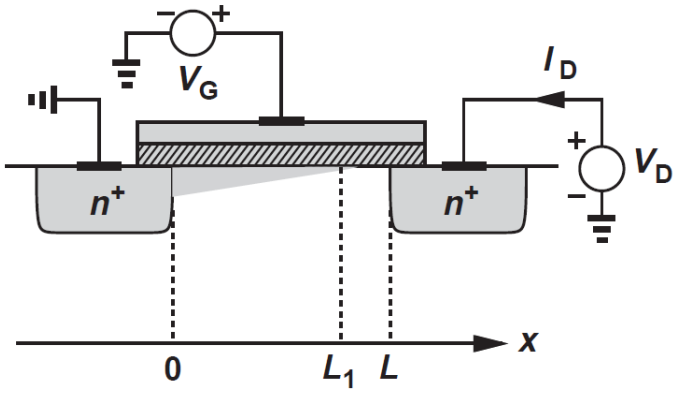
[Sedra/Smith, 2015]

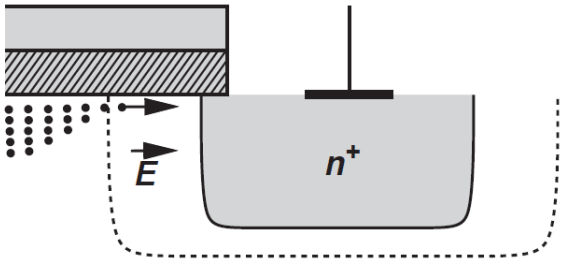
Channel Length Modulation (CLM)

- L_{eff} decreases with V_{DS} → Shorter L gives more current
- λ: Channel length modulation coefficient ($\lambda \propto 1/L$)

$$I_D = \frac{\mu C_{ox}}{2}\frac{W}{L}V_{ov}^2(1+\lambda V_{DS})$$

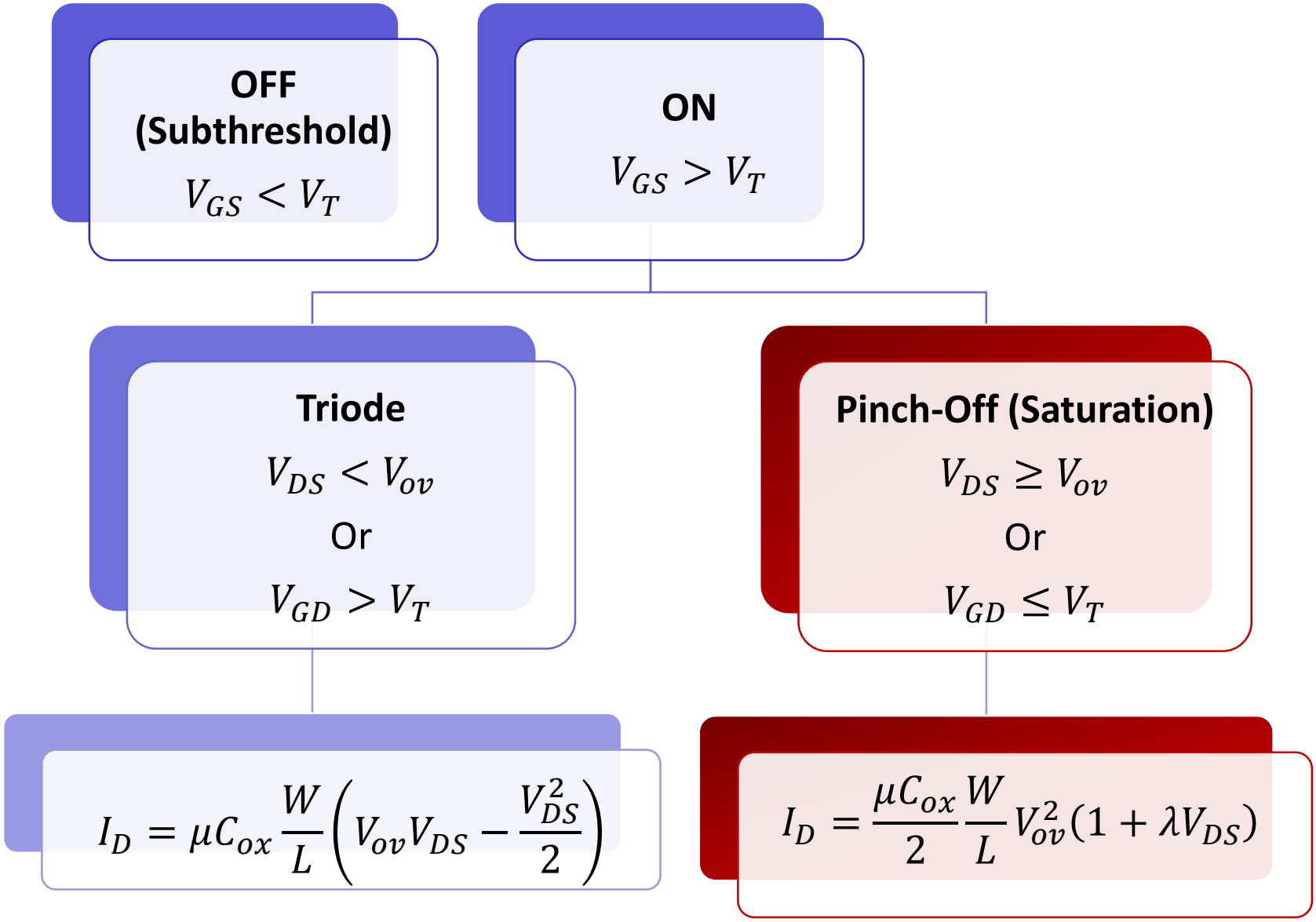






[Razavi, 2014] [Sedra/Smith, 2015]

Regions of Operation Summary

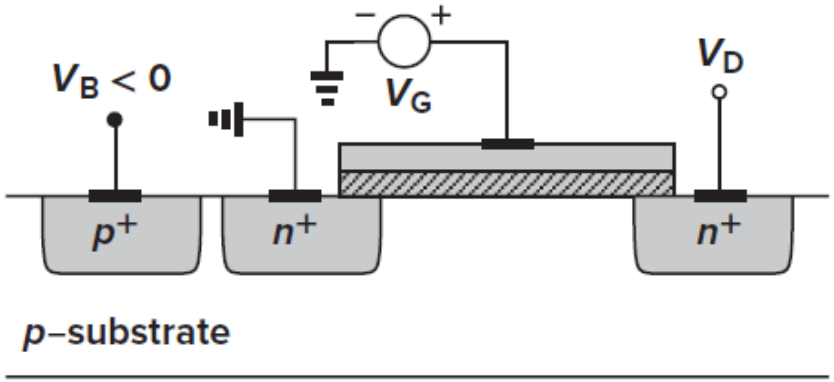


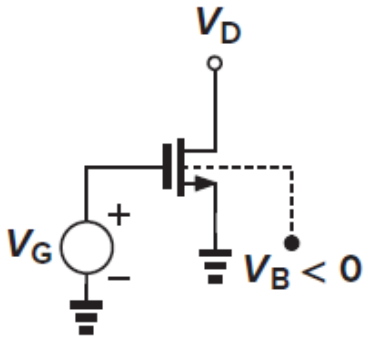
Body Effect

- V_{SB} affects the charge required to invert the channel
  - Increasing V_S or decreasing V_B increases V_{TH}

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

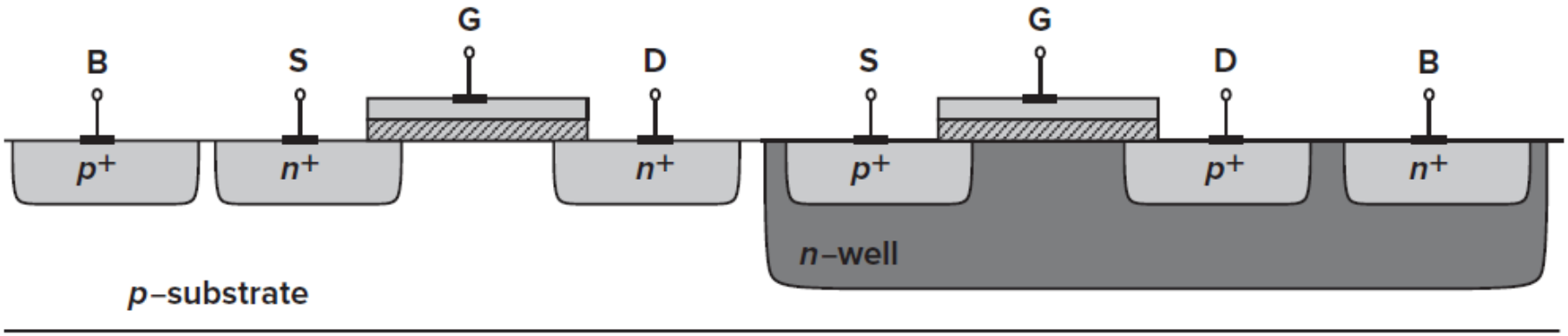
  - Φ_F = surface potential at threshold
    - Depends on doping level and intrinsic carrier concentration n_i
  - γ = *body effect coefficient*
    - Depends on C_{ox} and doping





CMOS

- CMOS = NMOS + PMOS on the same substrate
- All NFETs share the same substrate
- Each PFET can have an independent n-well
  - Useful in some analog circuits
- S/D junction diodes must remain reverse-biased under all conditions
  - NMOS bulk connected to most negative potential (ground)
  - PMOS bulk connected to most positive potential (VDD)



Why CMOS?

- Early integrated circuits primarily used bipolar transistors (BJTs)
- CMOS technologies dominated the digital market since the 1980s
  - Consumed negligible static power
    - Was indeed negligible in the past
    - But not negligible any more…
  - Required very few devices per gate
  - Can be scaled down more easily
  - Lower fabrication cost

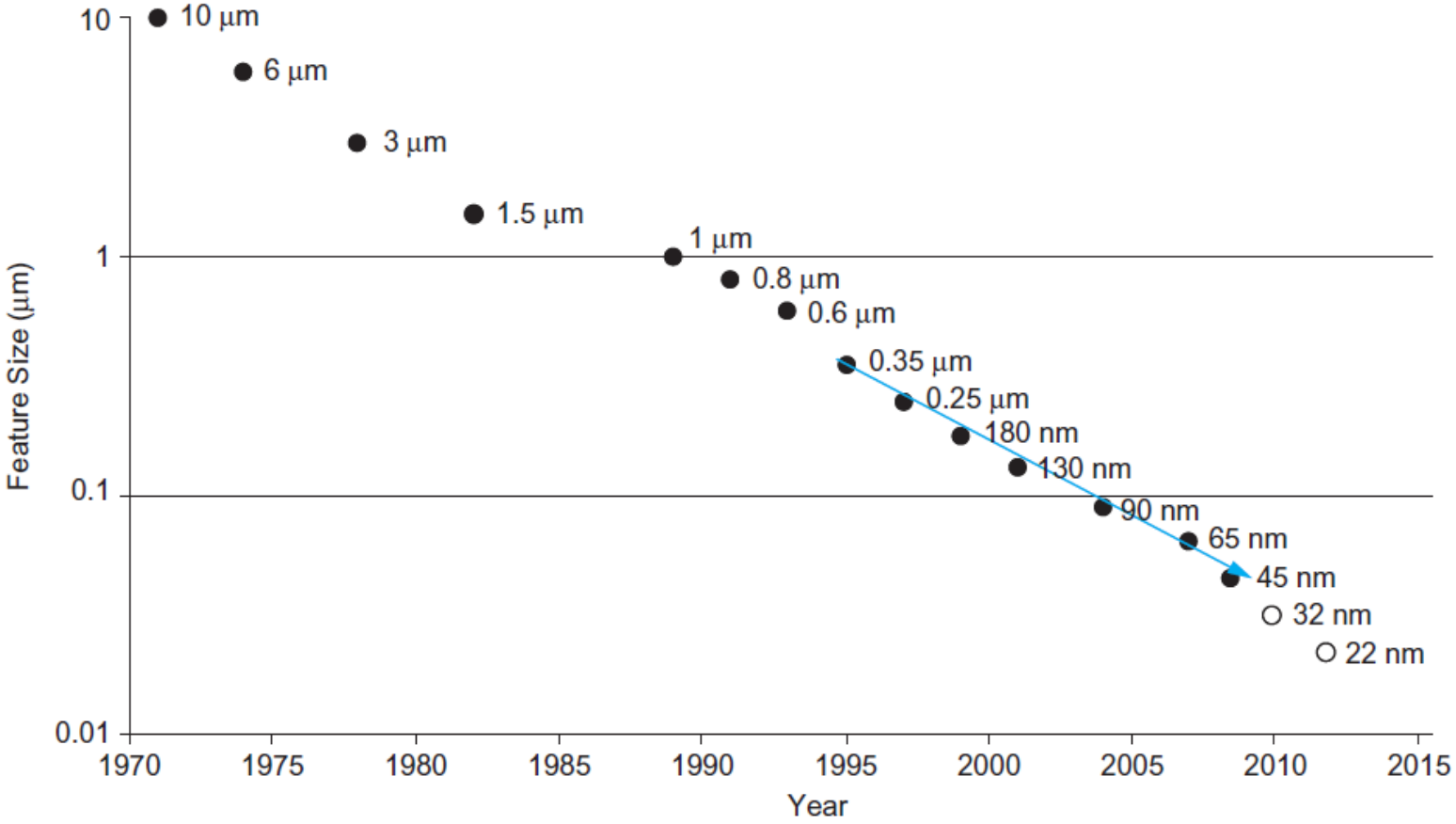
- For analog design, BJTs used to be much better than MOSFETs
  - Faster, less noisy, less variations, more energy efficient
- Then why analog CMOS?

Why Analog CMOS?

- ICs market is driven primarily by memories and microprocessors
  - The analog designer needs to survive in a digital driven market
- We want to integrate analog and digital on the same chip
  - Mixed-signal design and system-on-a-chip
- BJTs used to be faster, but with continuous scaling, MOSFET speed exceeded BJT
- MOSFET can operate with lower supply voltage

CMOS Technology Scaling: Moore's Law

- Min feature size (L_{min}) shrinking 30% ($\approx 1/\sqrt{2}$) every 2-3 years
  - Transistor area (and cost) are reduced by a factor of 2
- Device scaling brings new challenges in circuit design



[Weste and Harris, 2010]

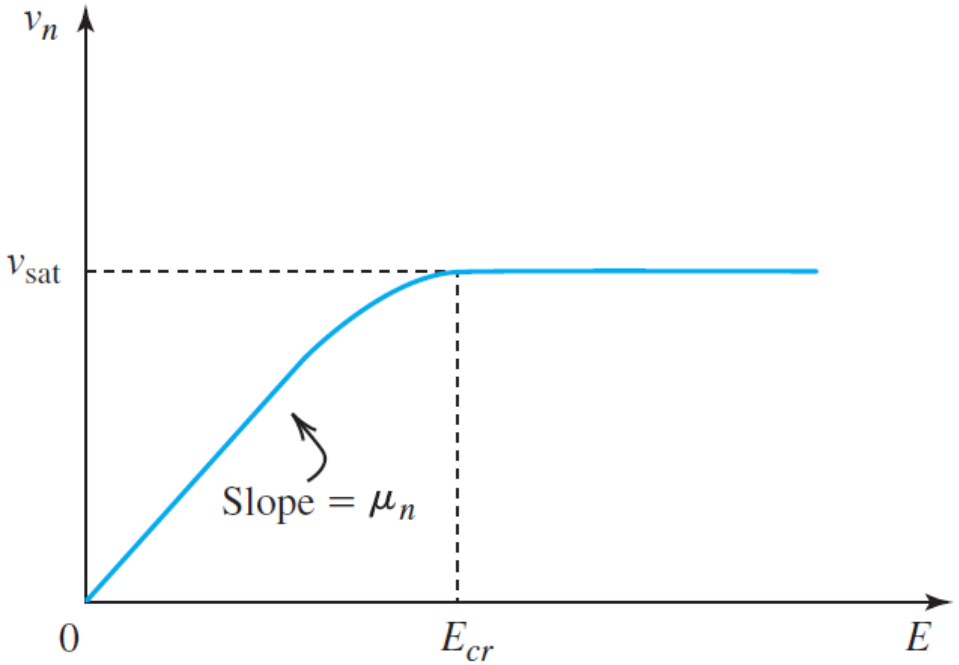
Short Channel Effects: Velocity Saturation

- For deep-submicron MOSFET with short channel length ($L < 0.25\mu m$) the electric field is very high

$$E = \frac{V_{DS}}{L}$$

- @ $E = E_{cr}$ ($V_{DS} = V_{DSsat}$) the velocity of the carriers saturates

$$v_{sat} = \mu E_{cr} = \mu \frac{V_{DSsat}}{L} \approx 10^7 cm/s$$



[Sedra/Smith, 2015]

Velocity Saturation: IV Characteristics

- Long channel: Triode region

$$I_D = \mu_n C_{ox} \frac{W}{L} \cdot \left(V_{ov} - \frac{V_{DS}}{2} \right) \cdot V_{DS}$$

- Velocity sat happens before pinch-off if $V_{ov} > V_{DSsat}$
  - Replace V_{DS} with V_{DSsat} and $v = \mu_n \frac{V_{DS}}{L}$ with $v_{sat} = \mu_n \frac{V_{DSsat}}{L}$

$$I_D = \mu_n C_{ox} \frac{W}{L} \cdot \left(V_{ov} - \frac{V_{DSsat}}{2} \right) \cdot V_{DSsat} = C_{ox} W v_{sat} \cdot \left(V_{ov} - \frac{V_{DSsat}}{2} \right)$$

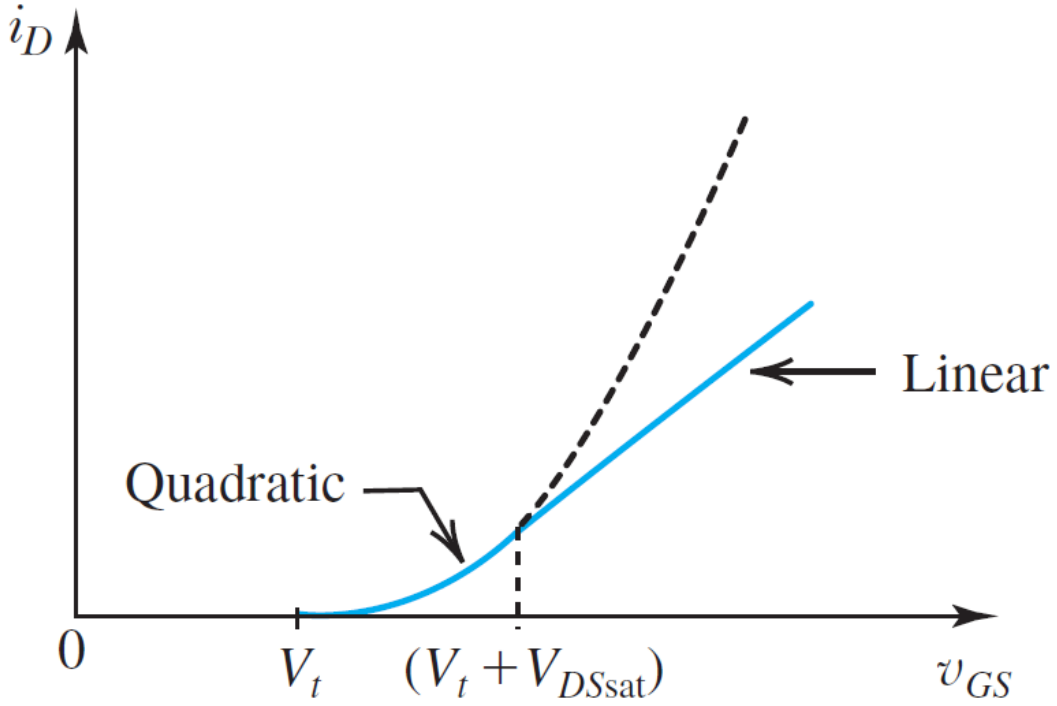
  - Including channel length modulation effect (the physical reason is different, but the effect on I_D is the same)

$$I_D = C_{ox} W v_{sat} \cdot \left(V_{ov} - \frac{V_{DSsat}}{2} \right) (1 + \lambda V_{DS})$$

Velocity Saturation: IV Characteristics

- Velocity sat happens before pinch-off if $V_{ov} > V_{DSsat}$

$$I_D = C_{ox} W v_{sat} \cdot \left(V_{ov} - \frac{V_{DSsat}}{2} \right) (1 + \lambda V_{DS})$$

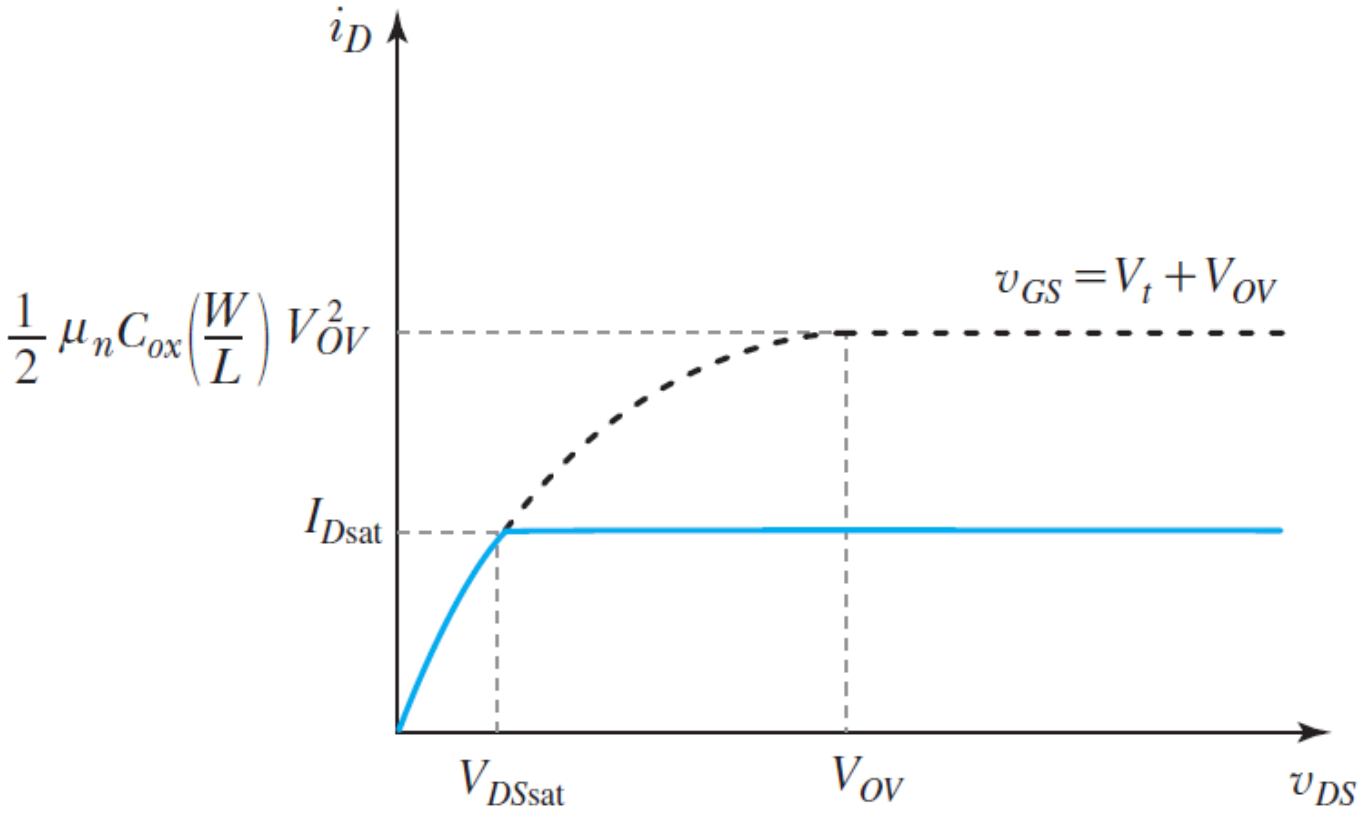


[Sedra/Smith, 2015]

Velocity Saturation: IV Characteristics

- Velocity sat happens before pinch-off if $V_{ov} > V_{DSsat}$

$$I_D = C_{ox} W v_{sat} \cdot \left(V_{ov} - \frac{V_{DSsat}}{2} \right) (1 + \lambda V_{DS})$$

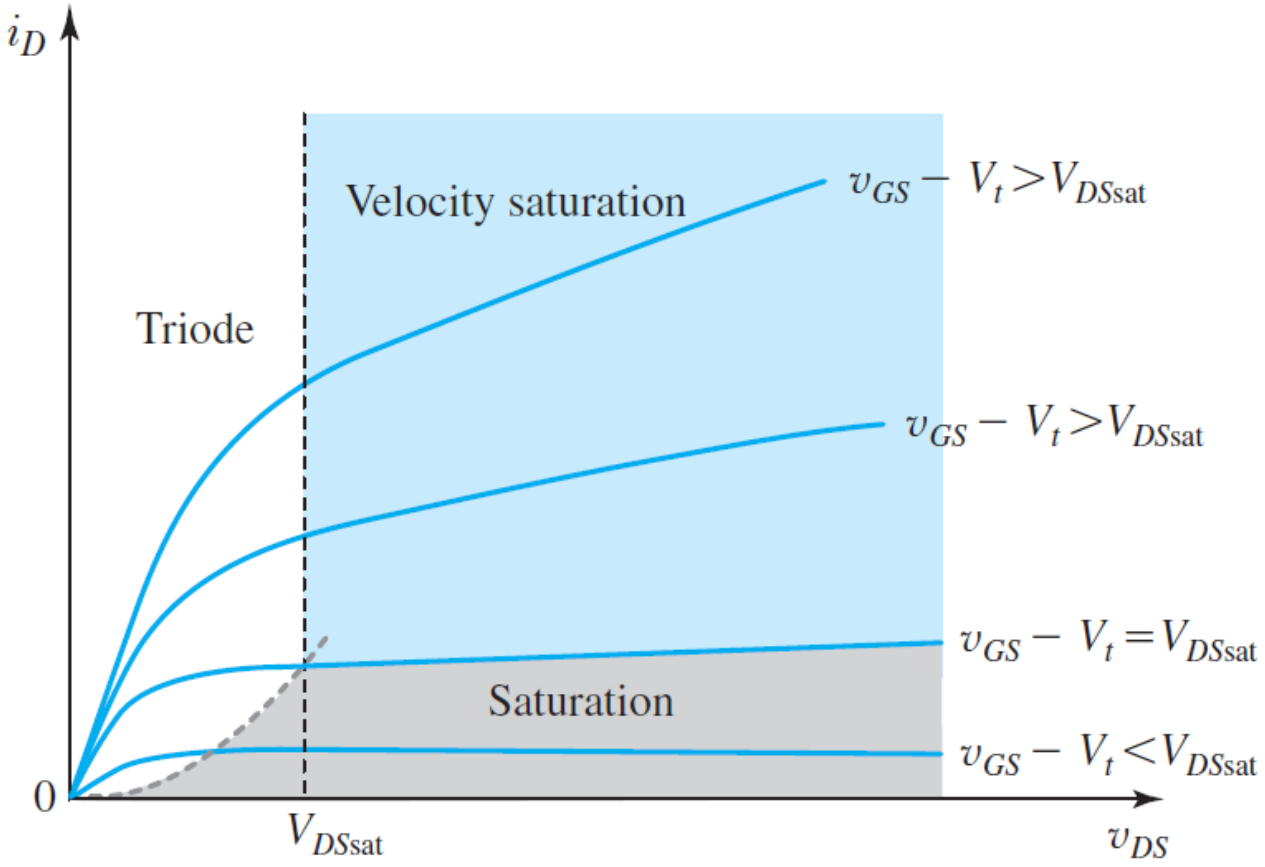


[Sedra/Smith, 2015]

Velocity Saturation: IV Characteristics

- Velocity sat happens before pinch-off if $V_{ov} > V_{DSsat}$

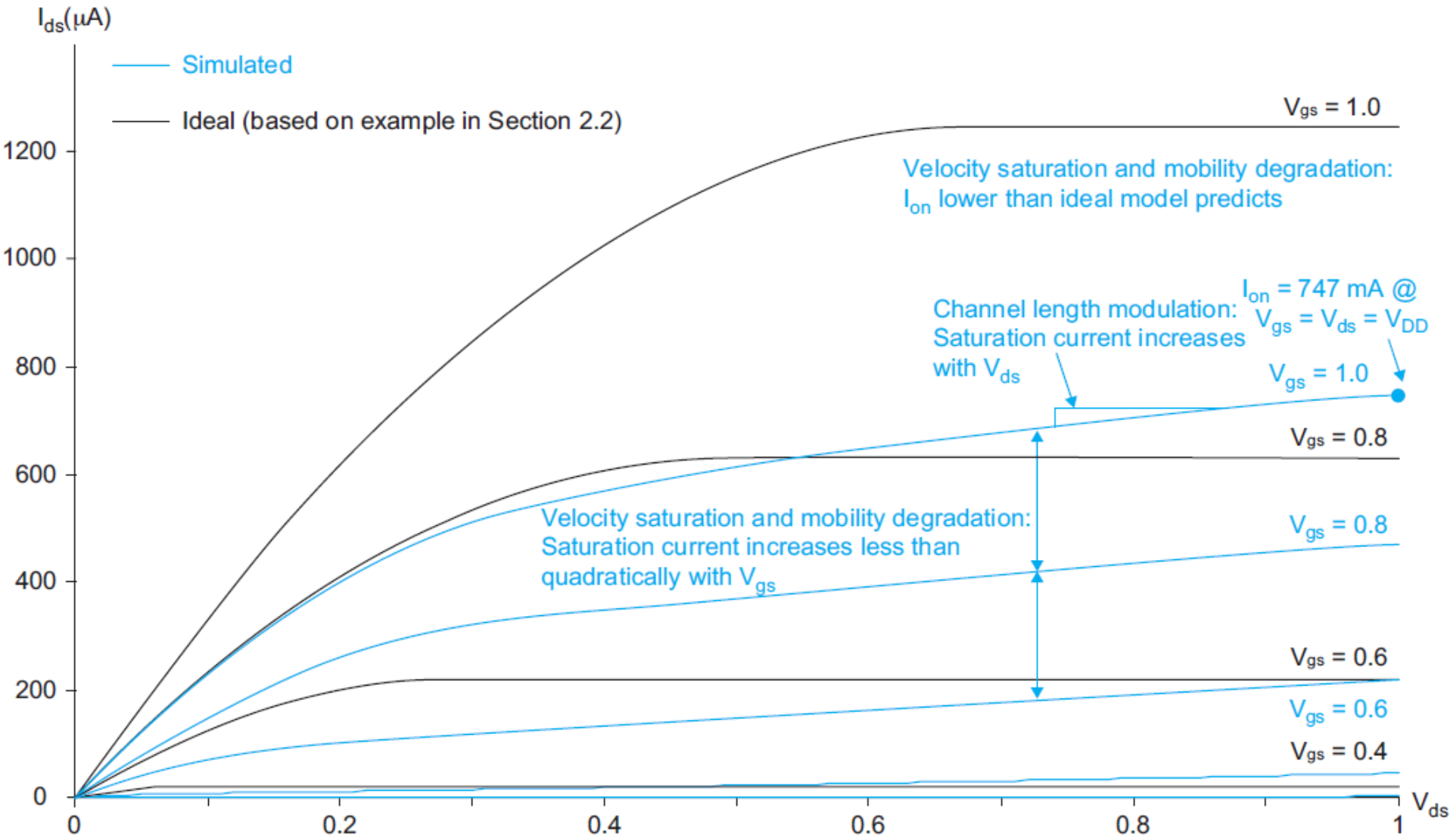
$$I_D = C_{ox} W v_{sat} \cdot \left(V_{ov} - \frac{V_{DSsat}}{2} \right) (1 + \lambda V_{DS})$$



[Sedra/Smith, 2015]

Short-Channel MOSFET I-V Ccs

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



[Weste & Harris]

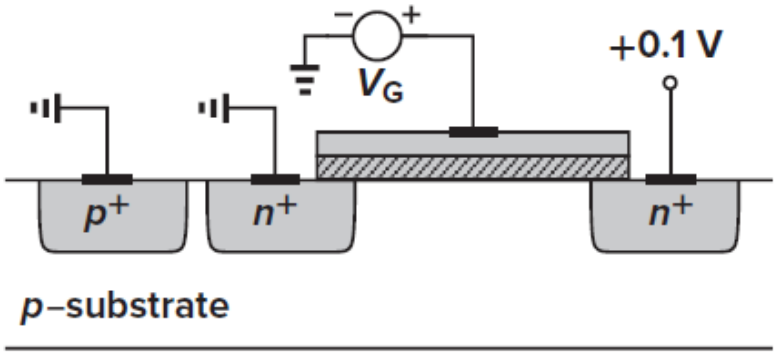
Why Do We Still Learn Square-Law?

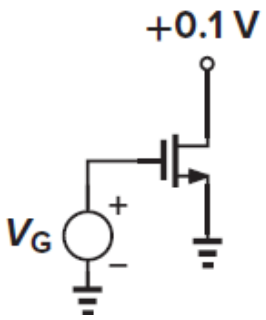
- For digital and RF, use min L
  - You care most about speed and power
- For analog, we use relatively long L
  - We care about matching, gain, and low-frequency noise
- For digital Vov = VDD – VTH
  - Short channel effects (e.g., velocity sat.) are more pronounced
- For analog Vov is relatively low
  - Short channel effects (e.g., velocity sat.) are less pronounced
- Simple model provides a great deal of intuition that is necessary in analog design
  - We must simulate the circuit to get more accurate results
- **However, graphical design approaches (e.g., gm/ID) are becoming more popular**

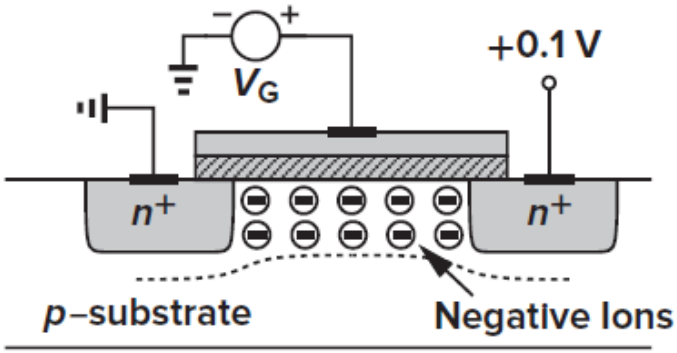
Thank you!

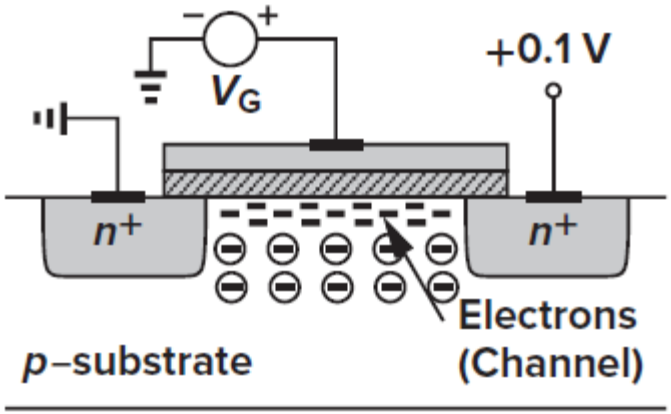
Threshold Voltage

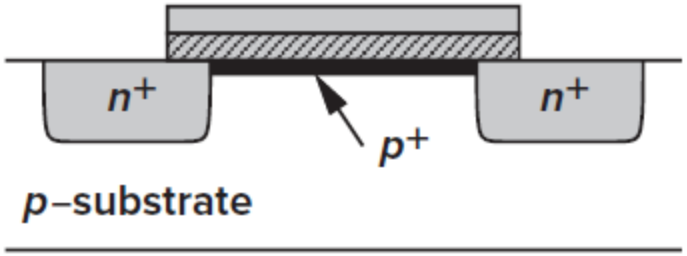
- Physicists usually define VTH of NFET as the gate voltage for which the inversion layer is as much n-type as the substrate is p-type
- P+ implant is used beneath the gate to adjust VTH











Extra Material

The following content is mainly based on "CMOS VLSI Design", 4th edition, by N. Weste and D. Harris and its accompanying lecture notes

More Non-Ideal MOSFET Behavior

- High Field Effects
  - Mobility Degradation
  - Velocity Saturation
- Threshold Voltage Effects
  - Drain-Induced Barrier Lowering (DIBL)
  - Short Channel Effect
- Leakage
  - Subthreshold Leakage
  - Gate Leakage
  - Junction Leakage

Non-Ideal MOSFET Behavior

- **High Field Effects**
  - **Mobility Degradation**
  - **Velocity Saturation**
- Threshold Voltage Effects
  - Drain-Induced Barrier Lowering (DIBL)
  - Short Channel Effect
- Leakage
  - Subthreshold Leakage
  - Gate Leakage
  - Junction Leakage

Mobility Degradation

- Vertical electric field: $E_{vert} = V_{gs}/t_{ox}$
  - Attracts carriers into channel
  - Long channel: $Q_{channel} \propto E_{vert}$
- At high vertical field strengths (V_{gs}/t_{ox})
  - The carriers scatter off the oxide interface more often
  - Scattering slows carrier progress
  - leads to less current than expected at high V_{gs}

Mobility Degradation

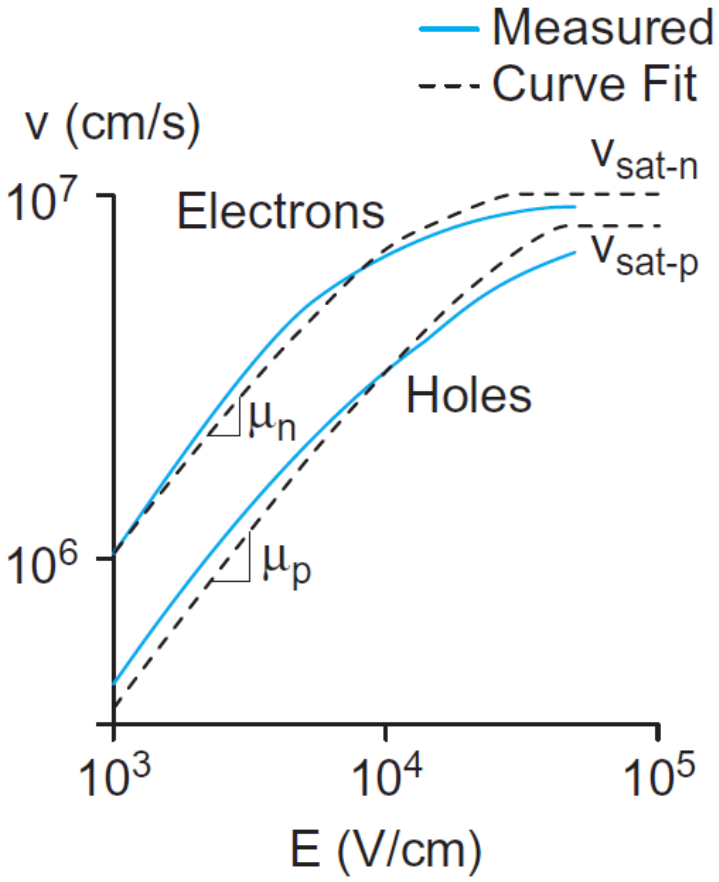
- Mobility degradation can be modeled by replacing μ with a smaller μ_{eff} that is a function of V_{gs}.
- Example: $V_T = 0.3V$, $t_{ox} = 1.05nm$

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

$$\mu_{\text{eff-n}}(V_{gs}=1.0) = 96\ \text{cm}^2/\text{V}, \mu_{\text{eff-p}}(V_{gs}=1.0) = 36\ \text{cm}^2/\text{V}$$

Velocity Saturation

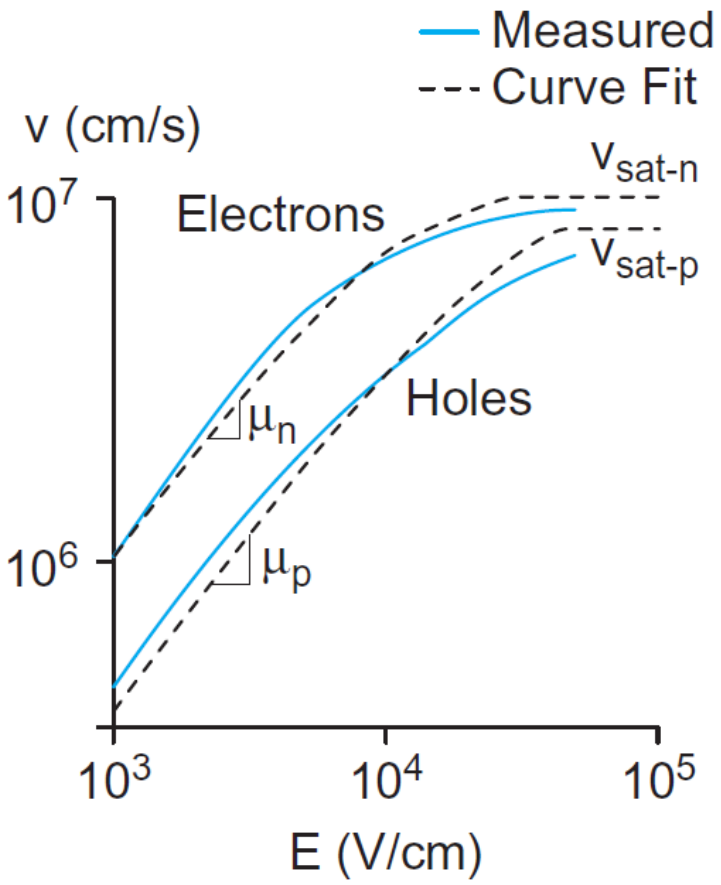
- Lateral electric field: $E_{lat} = V_{ds}/L$
  - Accelerates carriers from drain to source
  - Long channel: $v = \mu E_{lat}$
- At high lateral field (V_{ds}/L)
  - Carriers scatter off atoms in silicon lattice
  - Carrier velocity ceases to increase linearly with field strength
  - Leads to less current than expected at high V_{ds}



Velocity Saturation cont'd

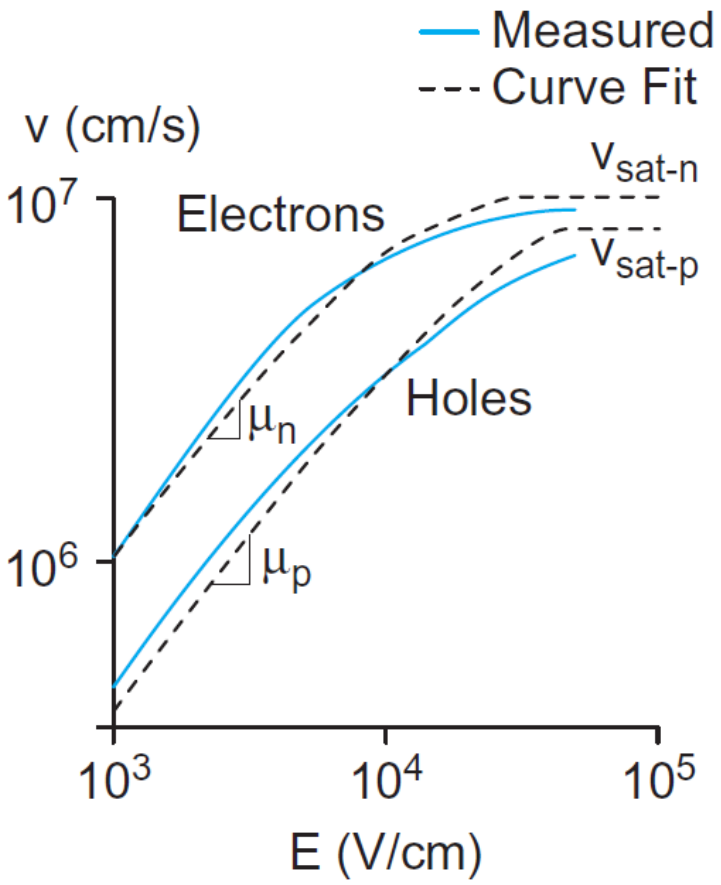
$$v = \begin{cases} \dfrac{\mu_{\text{eff}} E}{1 + \dfrac{E}{E_c}} & E < E_c \\ v_{\text{sat}} & E \geq E_c \end{cases}$$

- $v_{sat-n} \approx 10^7 \frac{cm}{s}$
- $v_{sat-p} \approx 8 \times 10^6 \frac{cm}{s}$
- Critical electric field: $E_c = \frac{2v_{sat}}{\mu_{eff}}$
- Critical $V_{ds} = V_c = E_c L$



Velocity Saturation cont'd

- $\mu_{eff-n} = 96\frac{cm^2}{V \cdot s}, \mu_{eff-p} = 36\frac{cm^2}{V \cdot s}$
- Example:
  - $E_c = \frac{2v_{sat}}{\mu_{eff}}$
  - $E_{c-n} = 20\frac{mV}{nm}, E_{c-p} = 44\frac{mV}{nm}$
  - $V_c = E_c L$
  - If $L = 50nm$
    - $V_{c-n} = 1V, V_{c-p} = 2.2V$
- $V_{GT} = V_{GS} - V_T = V_{ov}$
- $V_c < V_{GT}$: Velocity sat first
- $V_c > V_{GT}$: Pinch-off first



Short Channel I-V Model

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

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

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

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

  - Independent of L!

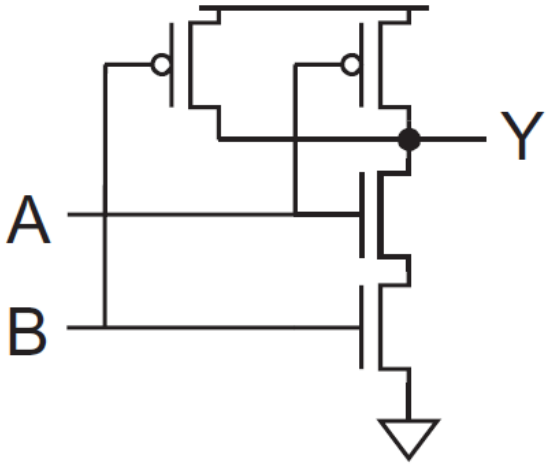
- Real transistors are partially velocity saturated
  - Approximate with α-power law model
  - $I_{DS} \propto (V_{GS} - V_T)^{\alpha}$
  - $1 < \alpha < 2$ determined empirically (≈ 1.3 for 65 nm)

Velocity Sat.: V_{DD} Scaling

- Velocity saturation and mobility degradation result in less current than expected at high voltage
- This means that there is no point in trying to use a high V_{DD} to achieve fast transistors
- V_{DD} has been decreasing with process generation to reduce power consumption
- Moreover, the very short channels and thin gate oxides would be damaged by high V_{DD}

Velocity Sat.: Series Ts

- Transistors in series drop part of the voltage across each transistor
  - Thus experience smaller fields and less velocity saturation
  - Series transistors tend to be a bit faster than a simple model would predict
- Two NMOS transistors in series deliver more than half the current of a single NMOS transistor of the same width
- This effect is more pronounced for NMOS
  - NMOS has higher mobility and thus is more velocity saturated
  - NAND gates perform better than first order estimates



Non-Ideal MOSFET Behavior

- High Field Effects
  - Mobility Degradation
  - Velocity Saturation
- **Threshold Voltage Effects**
  - **Drain-Induced Barrier Lowering (DIBL)**
  - **Short Channel Effect**
- Leakage
  - Subthreshold Leakage
  - Gate Leakage
  - Junction Leakage

DIBL

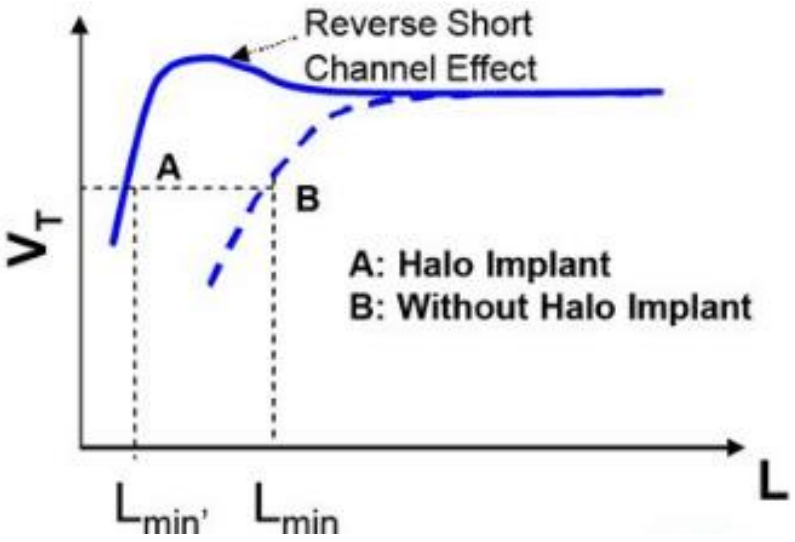
- DIBL: Drain-Induced Barrier Lowering
- Electric field from drain affects threshold voltage
  - More pronounced in short channel devices

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

  - η: DIBL coefficient $\sim 100mV/V$
- High drain voltage causes current to increase (similar to channel length modulation)
- **Gate is losing control over the channel**

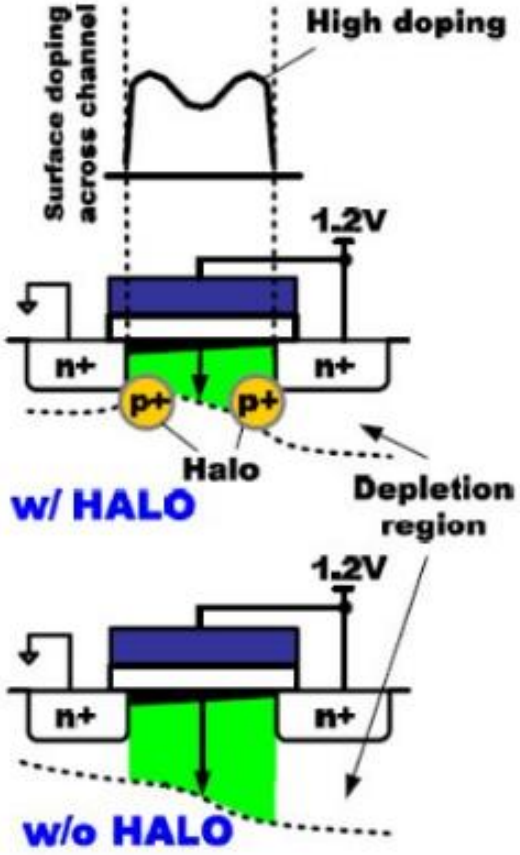
Short Channel Effect (V_t roll-off)

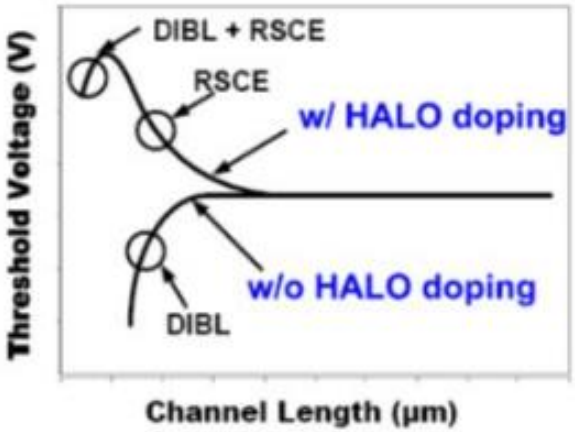
- In short channel devices, S/D depletion regions extend into the channel
  - Impacts the amount of charge required to invert the channel
  - And thus makes V_t a function of channel length
  - Somewhat similar to DIBL
- Short channel effect (SCE):
  - V_t decreases with smaller L
- Reverse short channel effect (RSCE):
  - Halo doping is used to fix DIBL
  - V_t increases then decreases with smaller L



[Sharma, UCSD]

SCE and RSCE





- HALO pocket implants used to mitigate the Short Channel Effect (SCE)
- Reverse Short Channel Effect (RSCE) observed due to HALO

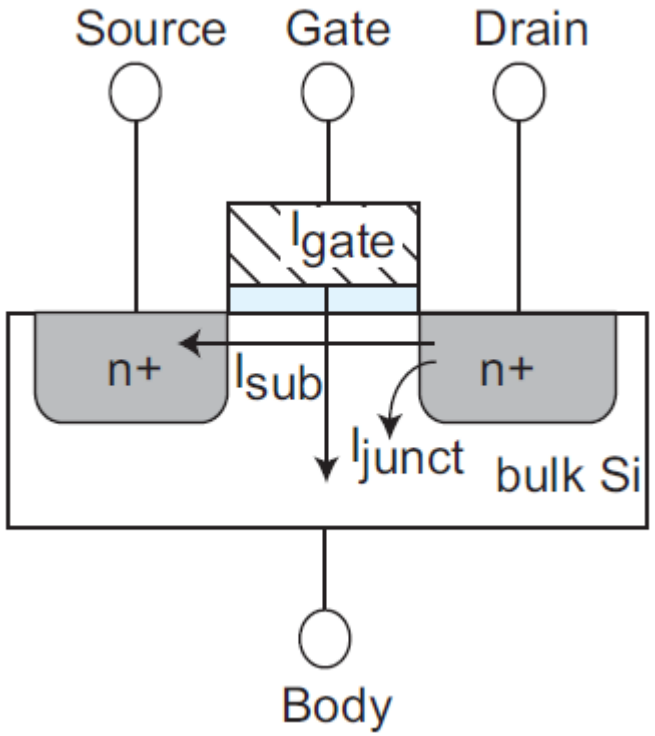
[M. Singh]

Non-Ideal MOSFET Behavior

- High Field Effects
  - Mobility Degradation
  - Velocity Saturation
- Threshold Voltage Effects
  - Drain-Induced Barrier Lowering (DIBL)
  - Short Channel Effect
- **Leakage**
  - **Subthreshold Leakage**
  - **Gate Leakage**
  - **Junction Leakage**

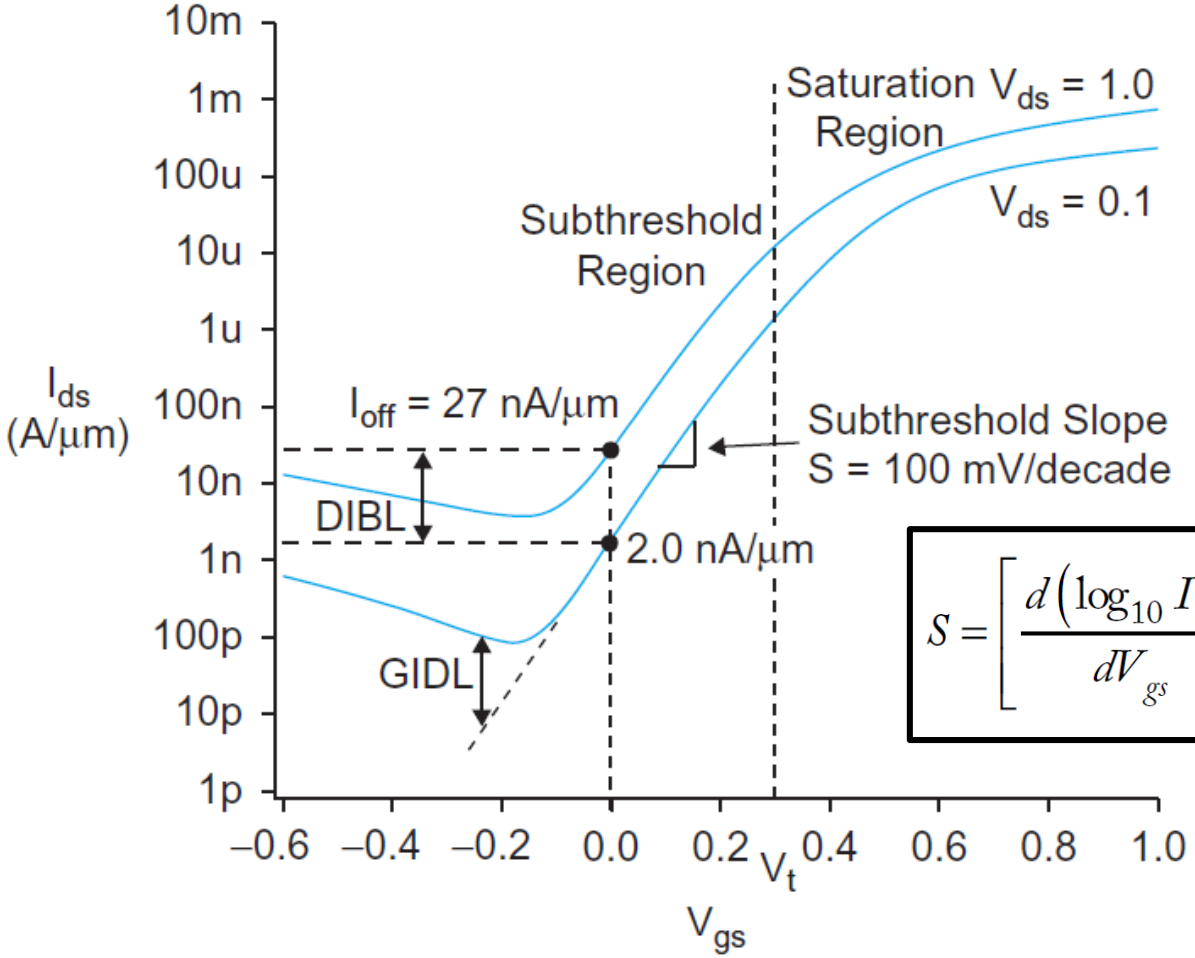
Leakage Currents

- Subthreshold Leakage
  - Was not important for $2\lambda \geq 180nm$
  - For 90nm and 65nm:
    - 1s to 10s of nA per transistor
    - Significant for billions of Ts
- Gate Leakage
  - Tunneling through the extremely thin gate oxide
  - Important for 65nm and beyond
- Junction Leakage
  - Reverse biased pn-junctions



Subthreshold Leakage

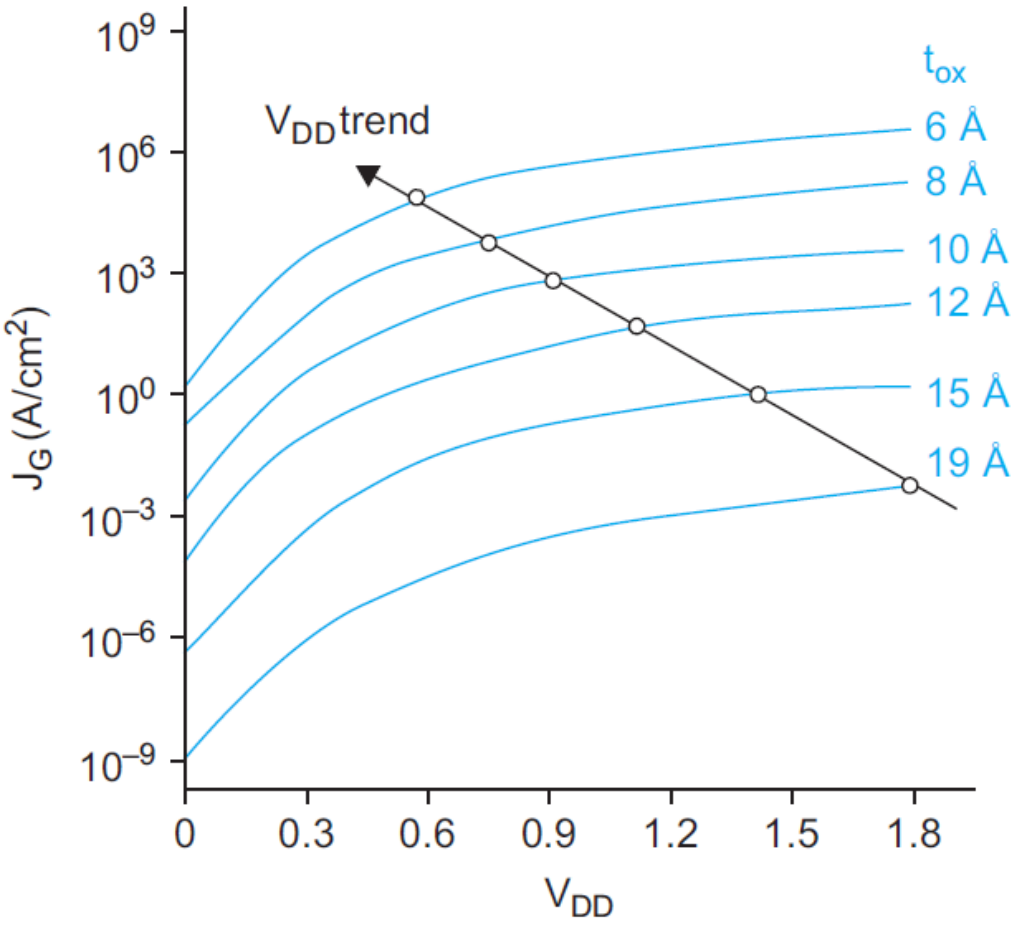
- Subthreshold current exponential with V_{gs} (and V_t and temp.)
- DIBL makes it worse
- Use high V_t devices to mitigate leakage (in non-critical paths)



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

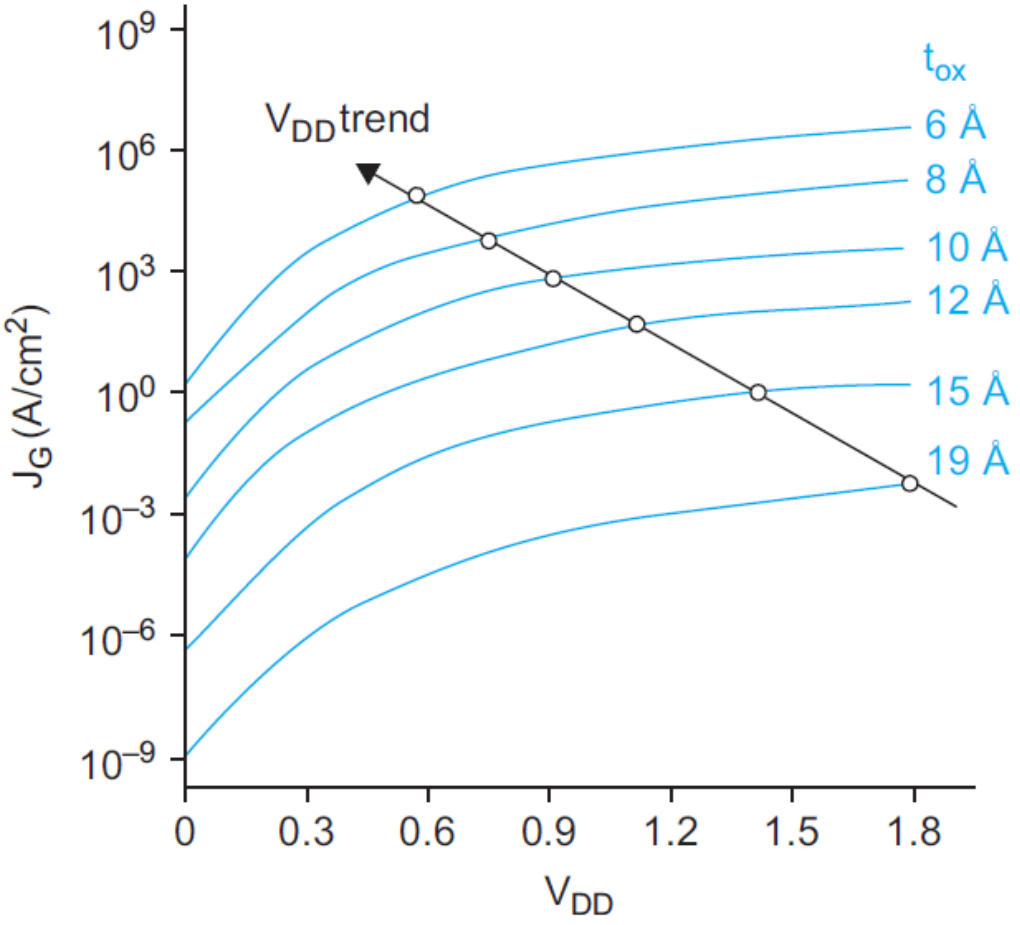
Gate Leakage

- Fowler-Nordheim (FN) tunneling (Field emission):
  - Most important at high voltage and moderate oxide thickness
  - Used to program EEPROM memory
- Direct tunneling:
  - Most important at lower voltage with thin oxides
  - The dominant gate leakage component



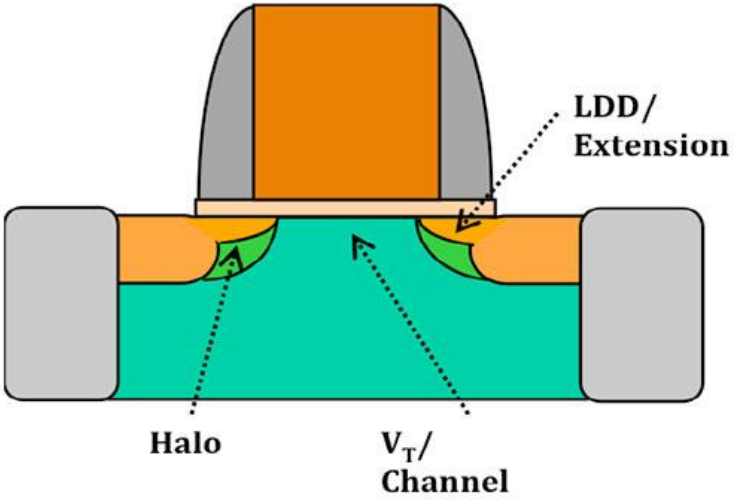
Gate Leakage cont'd

- Exponentially sensitive to t_{ox} and V_{DD}
- Independent of temperature
- For SiO_2 NMOS leakage is one order of magnitude higher
- Negligible for older processes (t_{ox} > 20 Å)
- Critically important at 65 nm and below
  - t_{ox} must be > 10.5 Å for gate leakage < 100 A/cm^2
  - SiO_2 atomic layer ≈ 3 Å

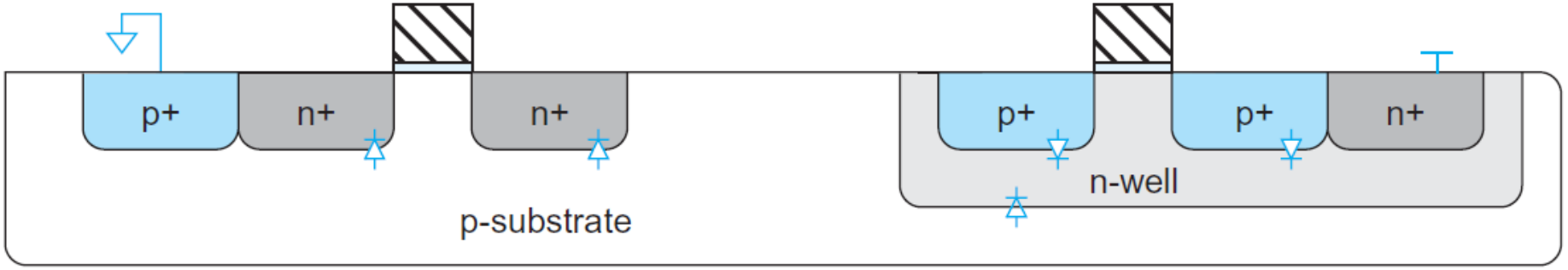


Junction Leakage

- I_S: Reverse saturation current
  - $\sim 0.1 - 0.01 \text{fA}/\mu\text{m}^2$
  - Negligible compared to other leakage mechanisms
- More significantly, heavily doped drains (due to halo implants) are subject to "Junction Tunneling":
  - Band-to-band tunneling (BTBT)
  - Gate-induced drain leakage (GIDL)

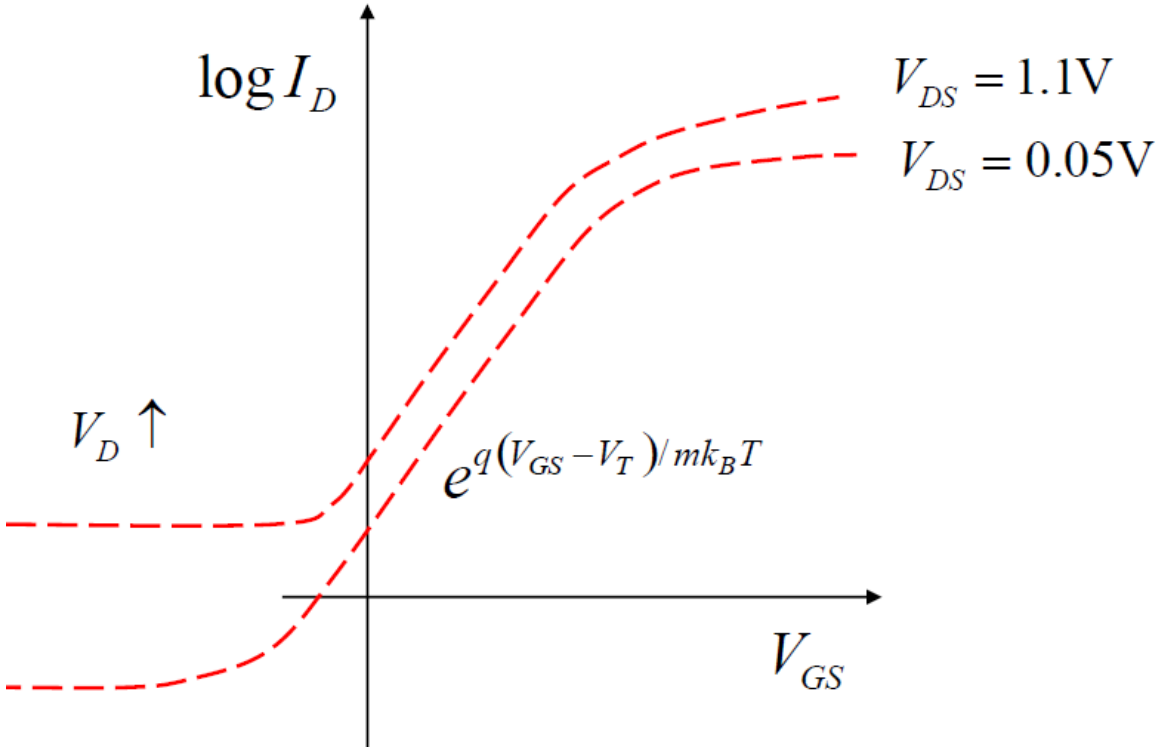


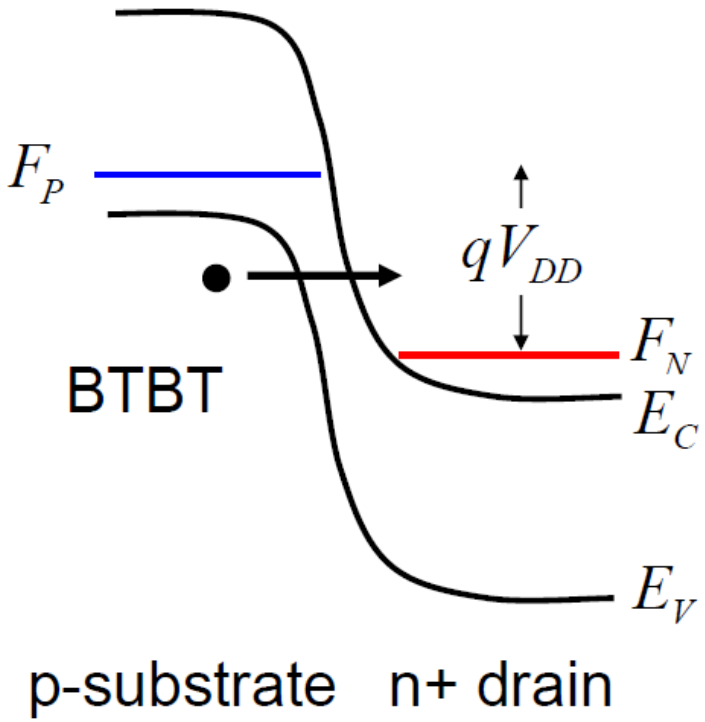
[www.axcelis.com]



BTBT

- Band-to-band tunneling across heavily doped p-n junctions
- In the drain-substrate junction
- Occurs in “ON” state

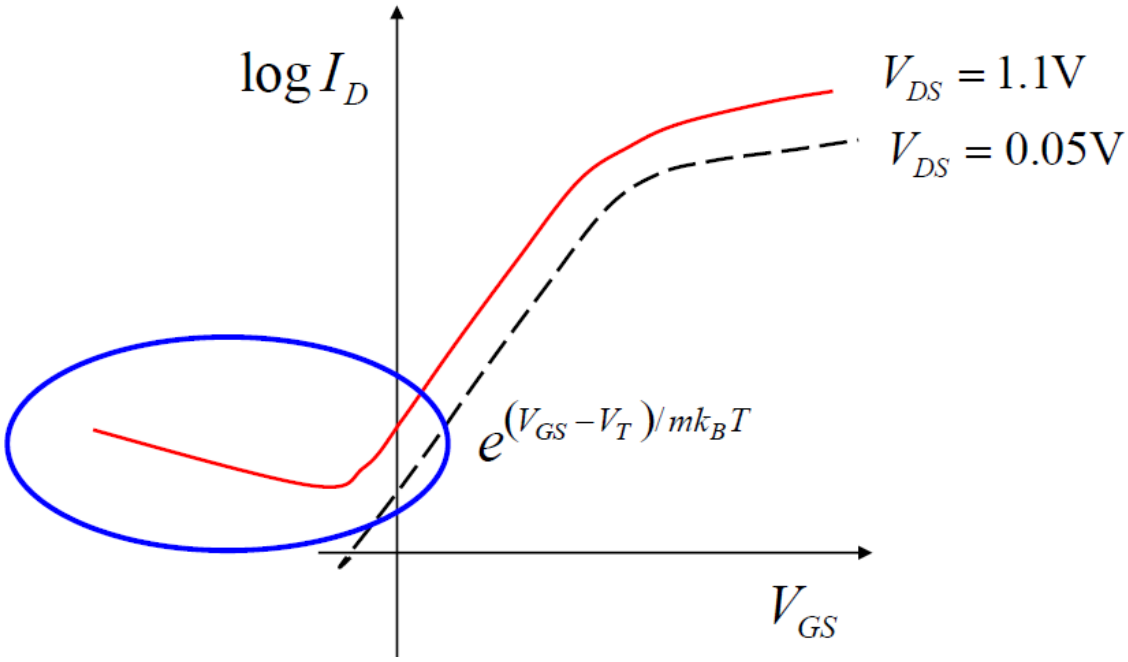


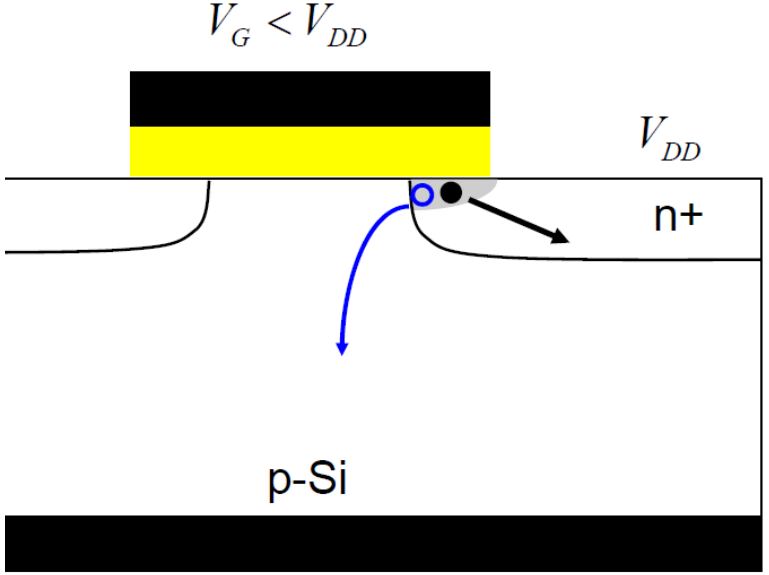


[Lundstrom, Purdue]

GIDL

- Gate induced BTBT at the overlap between gate and drain
- Occurs in "OFF" state (at negative V_{gs})
- Lesson learned: Negative V_{gs} may make leakage worse





[Lundstrom, Purdue]