Lecture 8
Microelectronic Devices

MOS structure – impact of interface states and trapped charge

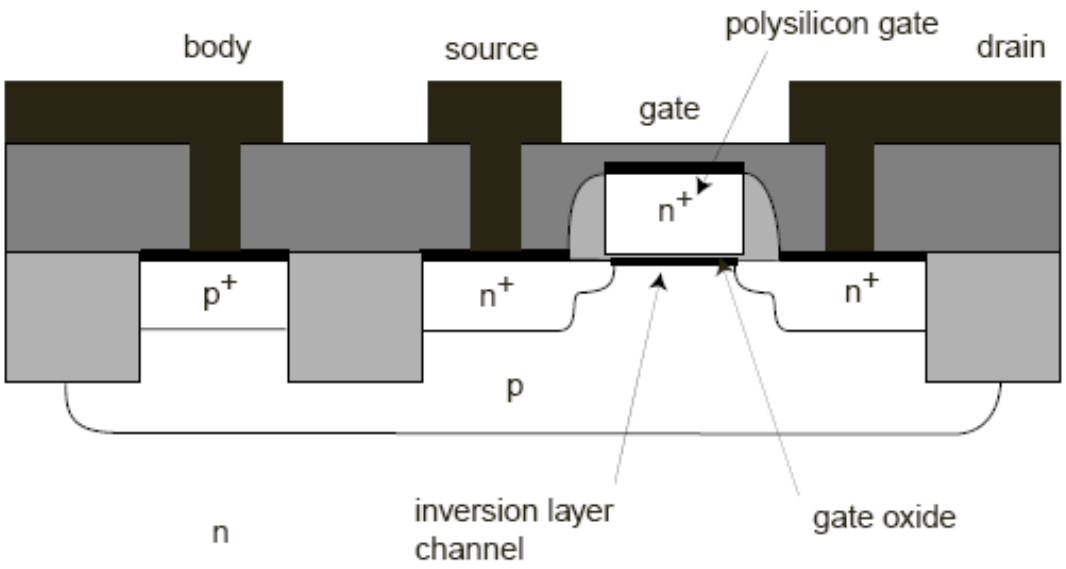
MOSFET operation regimes

- **Linear regime**
- **Saturation regime**

- **Short channel effects**

- **MOSFET scaling and Dennard rule**

MOSFET: layout



In **most MOSFETs** in digital logic circuits **Source and Drain are symmetrical**.

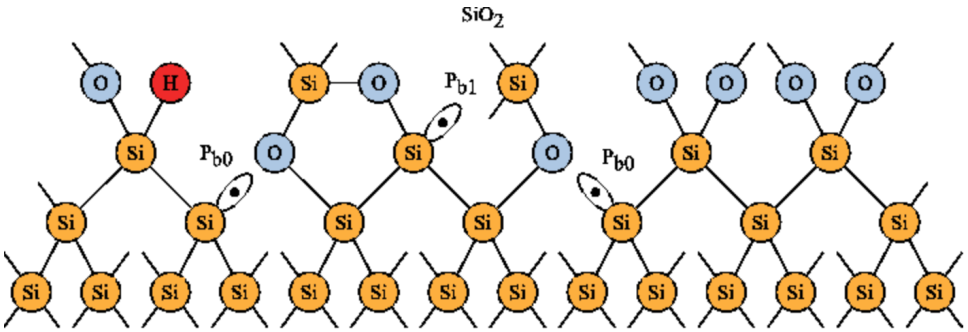
Both are **heavily doped** regions:

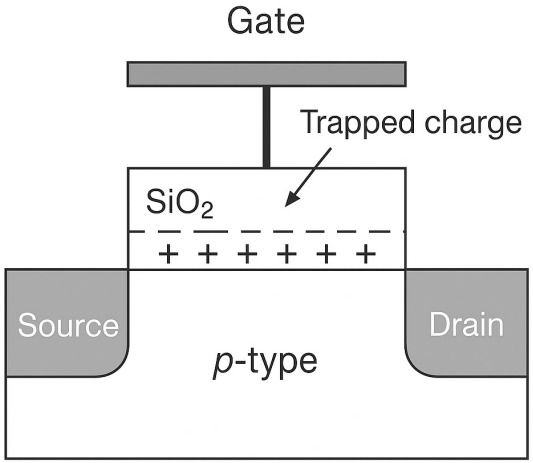
- **n-type (n+)** for **NMOS**
- **p-type (p+)** for **PMOS**

Key elements:

- inversion layer under gate (depending on gate voltage)
- heavily-doped regions in source/drain – formed by implantation/annealing
- inversion layer electrically connects source and drain
- 4-terminal device: body voltage important

Imperfect Si/SiO2 (semiconductor/oxide) interface – trapped charge!





Trapped charges are **unwanted charges** that get "stuck" at the **interface** between the **gate oxide (usually SiO_2)** and the **semiconductor (Si)**.

These charges can arise due to:

- **Manufacturing defects**
- **Radiation exposure**
- **Charge injection during operation**
- **Impurities or interface states**

Trapped charge can be of different nature:

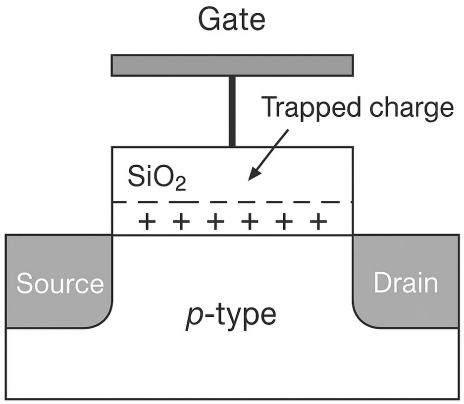
Interface trapped charges at theSi/SiO2 interface - dangling bonds or defects.

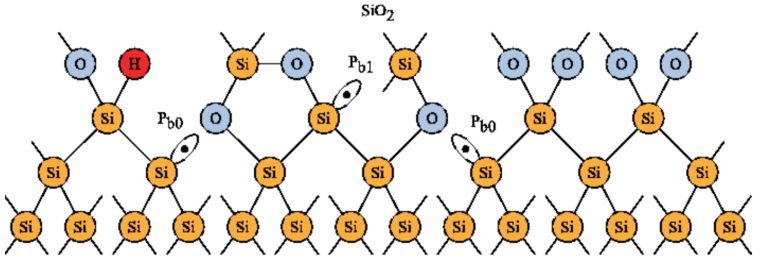
Fixed oxide charges - immobile charges in the oxide, typically positive, near interface

Mobile ionic charges - contaminants that can move under an electric field (positive)

Oxide trapped charges - Charges trapped within the bulk of the oxide

Imperfect Si/SiO2 (semiconductor/oxide) interface – trapped charge!





•The **oxide/semiconductor interface** (especially Si/SiO_2) naturally has **dangling bonds** — unsatisfied chemical bonds that can **trap charges** and act as **interface states**

• The interface states trap electrons or holes Problems: **instability**, **threshold voltage shifts**, and **degraded mobility**

• **Forming gas** (H_2-10%, N_2-90%) reduces density of interface traps in MOSFET fabrication

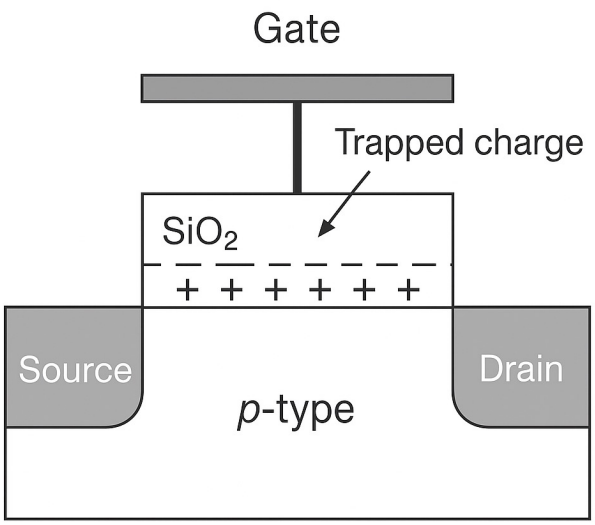
Hydrogen passivation:

•Hydrogen atoms from the forming gas **diffuse into the oxide and interface**

•They **bind to the dangling silicon bonds** (Si• → Si–H)

•This **neutralizes** the defect, removing its ability to trap charge.

Imperfect Si/SiO_2 interface – trapped charge!



Threshold Voltage Shift

- Trapped charges can **increase or decrease the threshold voltage**

Mobility Degradation

•Interface traps can **scatter carriers**, reducing their mobility and thus **degrading drive current**.

Subthreshold Slope Degradation

A high density of interface traps worsens the subthreshold slope, making the MOSFET **less ideal**.

Hysteresis

•In devices like memories, mobile or trapped charges can cause **hysteresis** in the I-V characteristics.

Imperfect Si/SiO_2 interface – trapped charge!

The effect of trapped charge on the threshold voltage depends on the type and polarity of the trapped charge, as well as the type of MOSFET (NMOS or PMOS)

$$\Delta V_T \propto \frac{Q_{\text{trap}}}{C_{ox}}$$

For n-MOSFET:

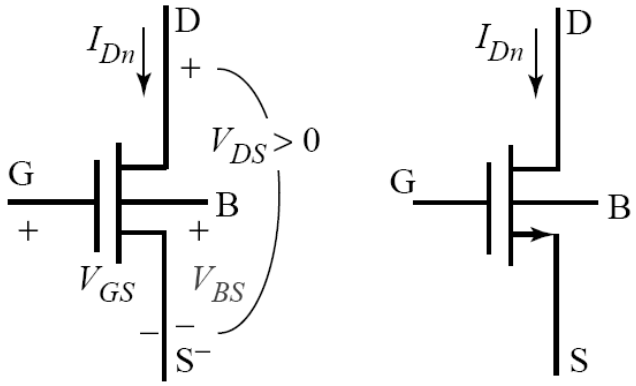
- **Positive trapped charge** :
  - **Increases V_T** → makes it **harder** to turn on the **n-MOSFET**
- **Negative trapped charge**:
  - **Decreases V_T** → makes it **easier** to turn on

For p-MOSFET:

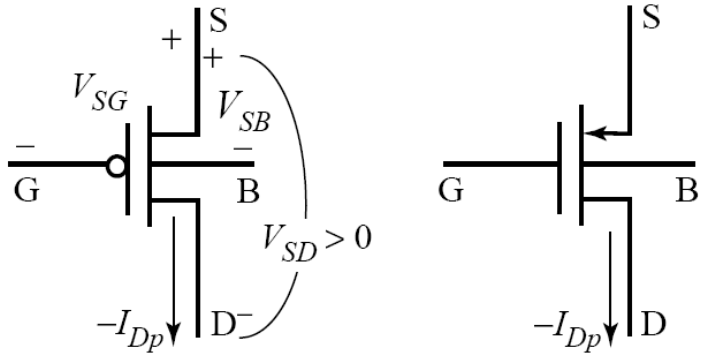
- **Positive trapped charge** :
  - **Decreases V_T** → makes it **easier** to turn on the **p-MOSFET**
- **Negative trapped charge**:
  - **Decreases V_T** → makes it **harder** to turn on

MOSFET: terminology and symbols

n-channel device (n-MOSFET) on p-Si substrate (uses electron inversion layer)



p-channel device (p-MOSFET) on n-Si substrate(uses hole inversion layer)

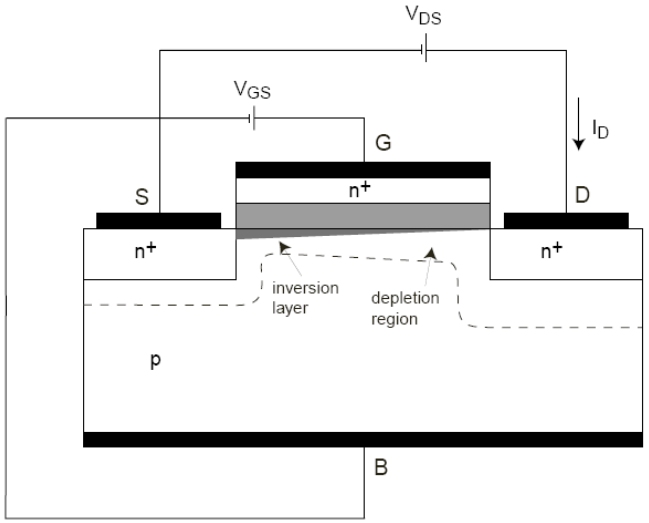


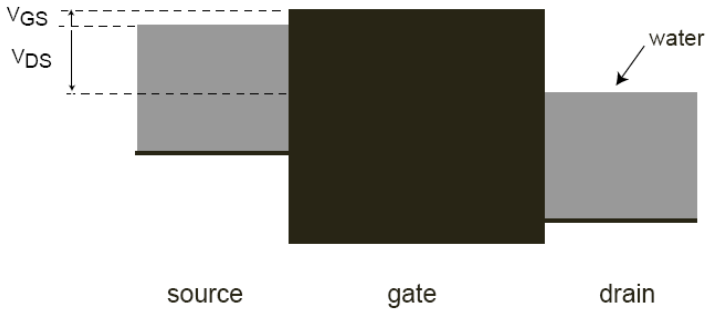
- sometimes the body contact is not figured

MOSFET: qualitative operation (1)

Water analogy of MOSFET:

- Source: water reservoir
- Drain: water reservoir
- Gate: gate between source and drain reservoirs



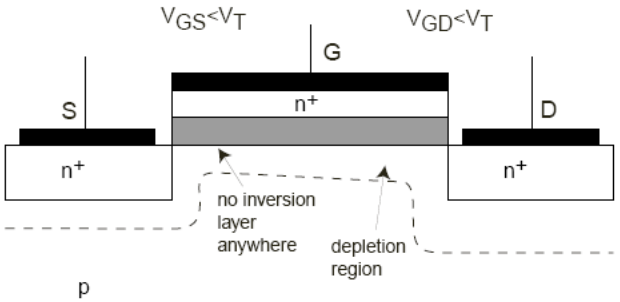


- **gate-to-source voltage** → gate height over source water level
- **drain-to-source voltage** → water level difference between reservoirs

MOSFET: qualitative operation (2)

Regime 1: **Cut-off regime (Off)**

- MOSFET: VGS < VT , VGD < VT with VDS > 0
- Water analogy: gate closed; no water can flow regardless of relative height of source and drain reservoirs.



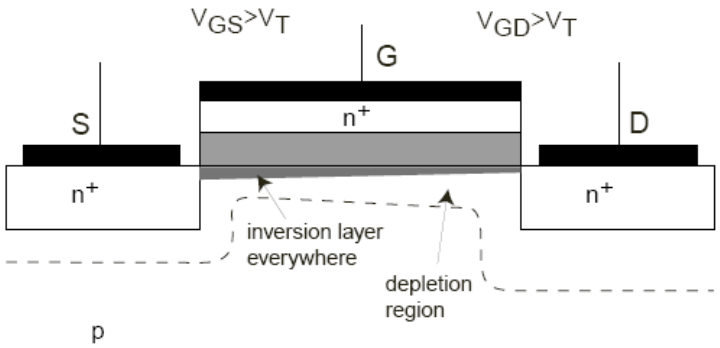
no water flow

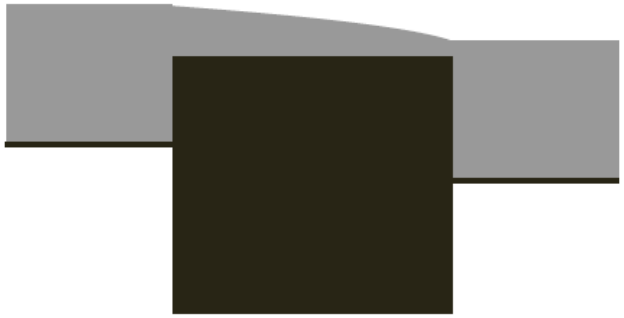
$$I_D = 0$$

MOSFET: qualitative operation (3)

Regime 2: **Linear or Triode regime (On)**

- MOSFET: VGS > VT , VGD > VT, with VDS > 0.
- Water analogy: gate open but small difference in height between source and drain; water flows.





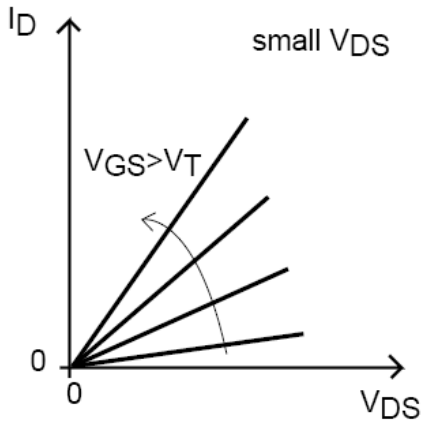
Electrons drift from source to drain: **electrical current, I_D!**

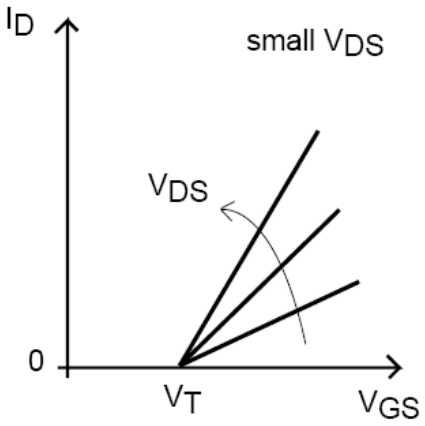
MOSFET: qualitative operation (4)

Regime 2: Linear or Triode regime (On)

Current ~ inversion charge ~ (V_{GS}-V_T) x V_{DS}

- $V_{GS} \uparrow \rightarrow |Q_n| \uparrow \rightarrow I_D \uparrow$
- $V_{DS} \uparrow \rightarrow |E_y| \uparrow \rightarrow I_D \uparrow$

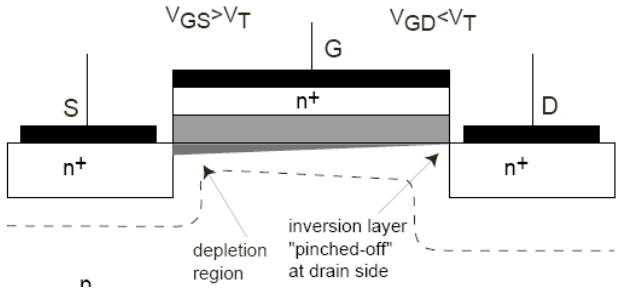


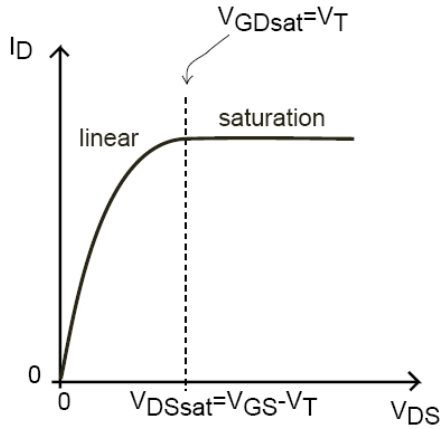


MOSFET: qualitative operation (5)

Regime 3: **Saturation regime (On)**

- MOSFET: $V_{GS} > V_T$, $V_{GD} < V_T$ ($V_{DS} > 0$).
- Water analogy: gate open; water flows from source to drain, but free-drop on drain side) total flow independent of relative reservoir height!



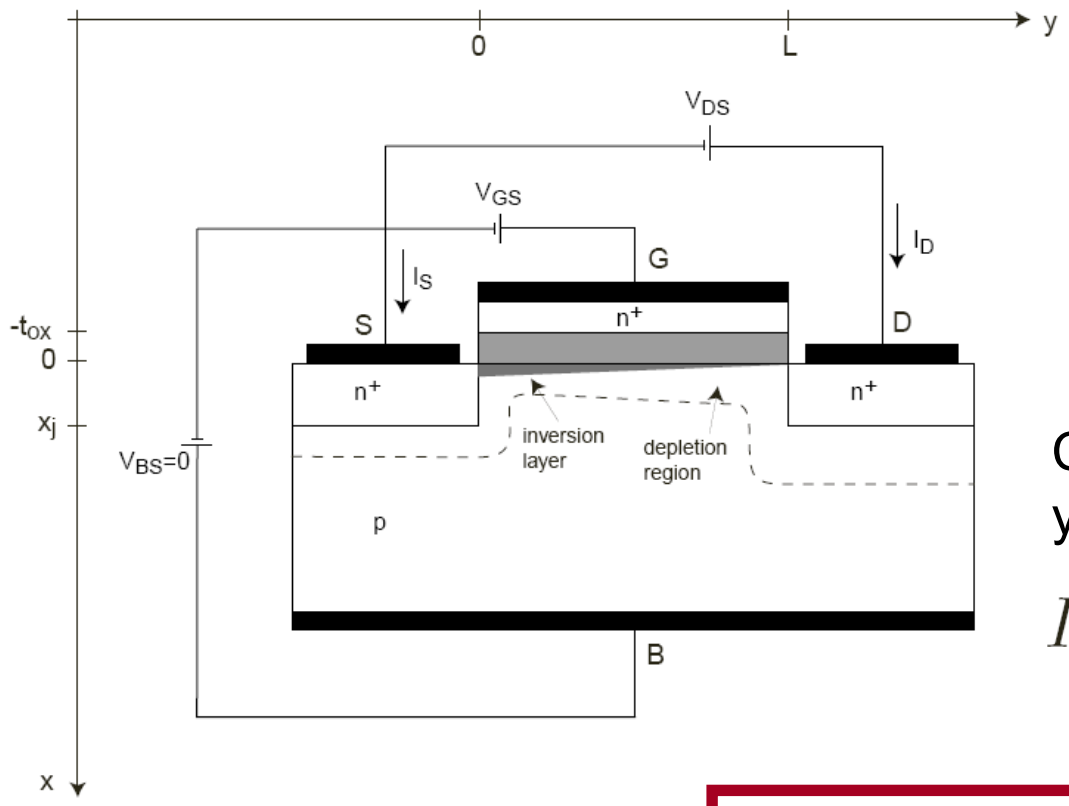


$$I_D = I_{Dsat}$$

Independent of V_{DS}!

MOSFET: basics for I-V characteristics

Geometry of problem (current flows in y-direction only)



Current can only flow in y-direction:

$$I_y = WQ_n(y)v_y(y)$$

Drain terminal current :

$$I_D = -WQ_n(y)v_y(y)$$

MOSFET: the output characteristics

$$I_D = -WQ_n(y)v_y(y)$$

Rewrite in terms of voltage at channel location *y*, $V_c(y)$:

$$v_y(y) \simeq -\mu_n E_y(y) = \mu_n \frac{dV_c(y)}{dy}$$

$$Q_n(y) = -C_{ox}[V_{GS} - V_c(y) - V_T] \qquad V_{GS} - V_c(y) \geq V_T.$$

$$I_D = W\mu_n C_{ox}(V_{GS} - V_c(y) - V_T)\frac{dV_c(y)}{dy}$$

linear first-order differential equation with one unknown, the channel voltage $V_c(y)$, Solved by separating variables

$$I_D dy = W\mu_n C_{ox}(V_{GS} - V_c - V_T)dV_c$$

$$\text{for } y = 0,\ V_c(0) = 0$$

$$\text{for } y = L,\ V_c(L) = V_{DS} \text{ (linear regime)}$$

$$I_D \int_0^L dy = W\mu_n C_{ox} \int_0^{V_{DS}} (V_{GS} - V_c - V_T)dV_c$$

$$\boxed{I_D = \frac{W}{L}\mu_n C_{ox}(V_{GS} - \frac{V_{DS}}{2} - V_T)V_{DS}}$$

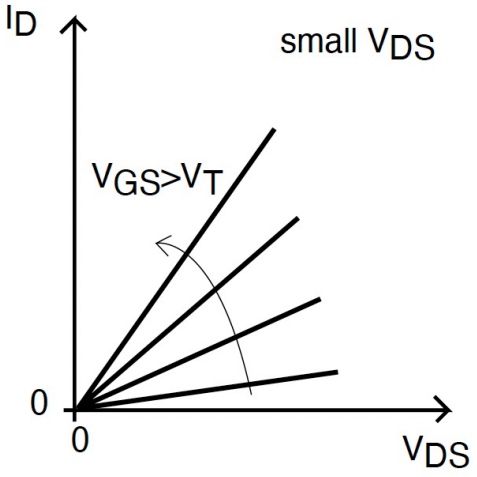
MOSFET: the output characteristics

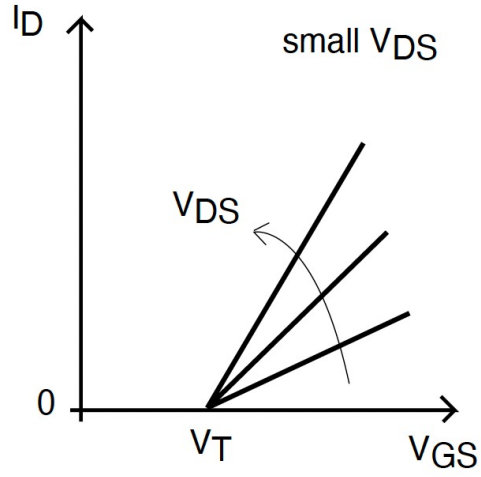
$$I_D = \frac{W}{L}\mu_n C_{ox}(V_{GS} - \frac{V_{DS}}{2} - V_T)V_{DS}$$

For small V_{DS}: $$I_D \simeq \frac{W}{L}\mu_n C_{ox}(V_{GS} - V_T)V_{DS}$$

- $V_{DS} \uparrow \rightarrow I_D \uparrow$ (higher lateral electric field)
- $V_{GS} \uparrow \rightarrow I_D \uparrow$ (higher electron concentratio
- $L \uparrow \rightarrow I_D \downarrow$ (lower lateral electric field)
- $W \uparrow \rightarrow I_D \uparrow$ (wider conduction channel)

Linear (triode) regime



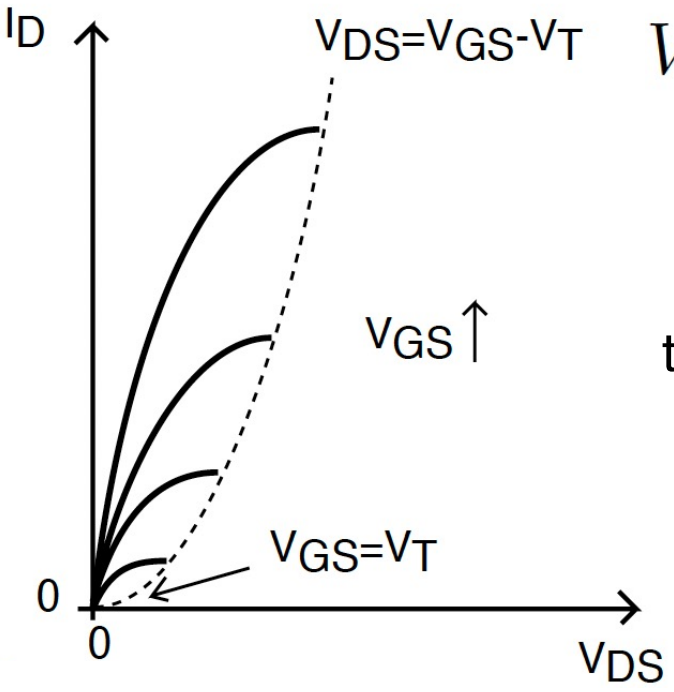


MOSFET: the output characteristics

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

This general equation is valid for $V_{GS} - V_c(y) \geq V_T$ for any y

Worst point is $y = L$, where $V_c(y) = V_{DS}$, hence, equation valid if



$$V_{GS} - V_{DS} \geq V_T \quad \text{or} \quad \boxed{V_{DS} \leq V_{GS} - V_T}$$

term responsible for bend over of I_D: $\frac{V_{DS}}{2}$

Saturation (1)

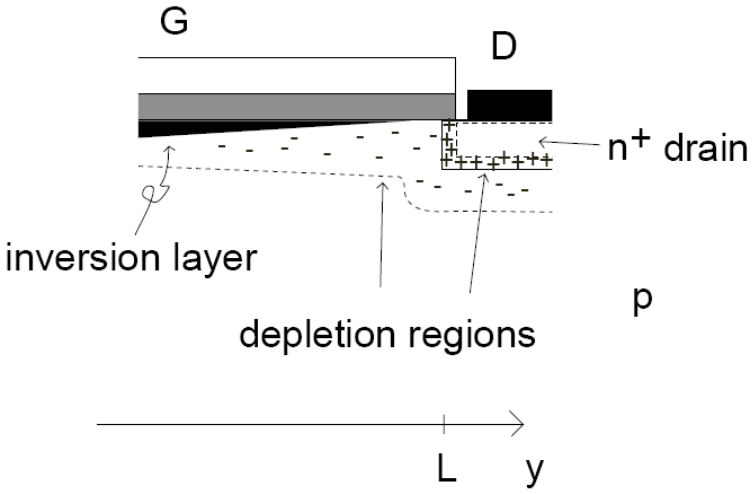
What happens when $V_{DSsat} = V_{GS} - V_T$?

- Charge control relation at drain end of channel:

$$Q_n(L) = -C_{ox}(V_{GS} - V_{DS} - V_T) = 0$$

- No inversion layer at end of channel ⇒ **Pinch-off**

- At pinchoff:
  – electron concentration small but not zero
  – electrons move fast because electric field is very high
  – dominant electrostatic feature: acceptor charge
  – there is no barrier to electron flow (on the contrary!)



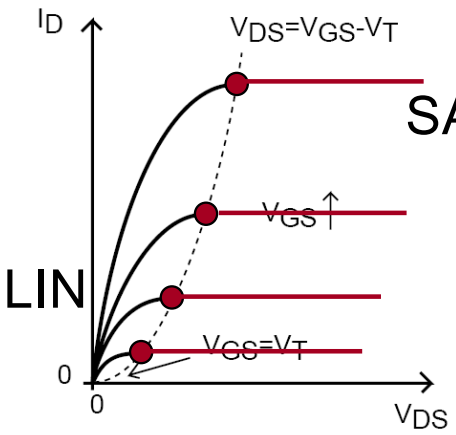
Saturation (2)

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

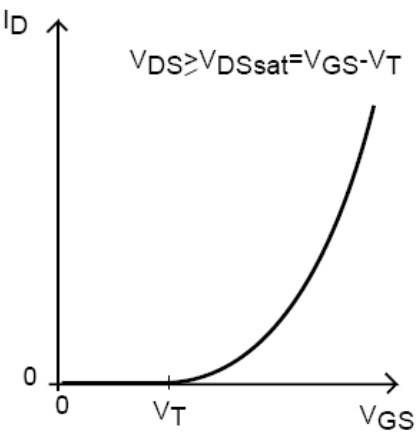
$$V_{DSsat} = V_{GS} - V_T$$

$$I_{Dsat} = \frac{W}{2L}\mu_n C_{ox}(V_{GS} - V_T)^2$$

Output charact.

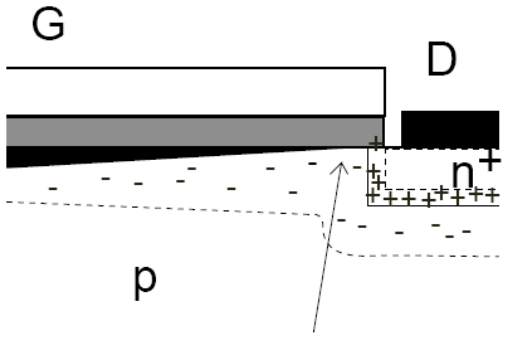


Transfer charact.



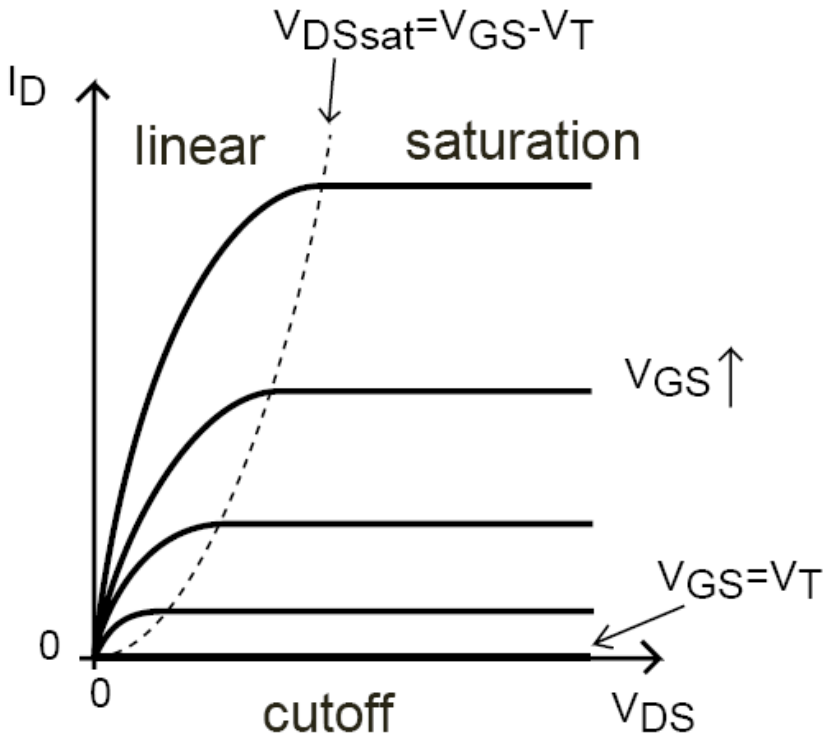
Drain current at pinchoff:

- $I_{Dsat} \propto 1/L$
- $I_{Dsat} \propto W; C_{ox}$



$V_c(L)=V_{DSsat}=V_{GS}-V_T$

MOSFET: regions of operation



• **Cutoff**: $V_{GS} < V_T$ ($V_{GD} < V_T$): no inversion layer anywhere underneath gate

$$I_D = 0$$

• **Linear**: $V_{GS} > V_T$, (with $V_{DS} > 0$): inversion layer everywhere underneath gate

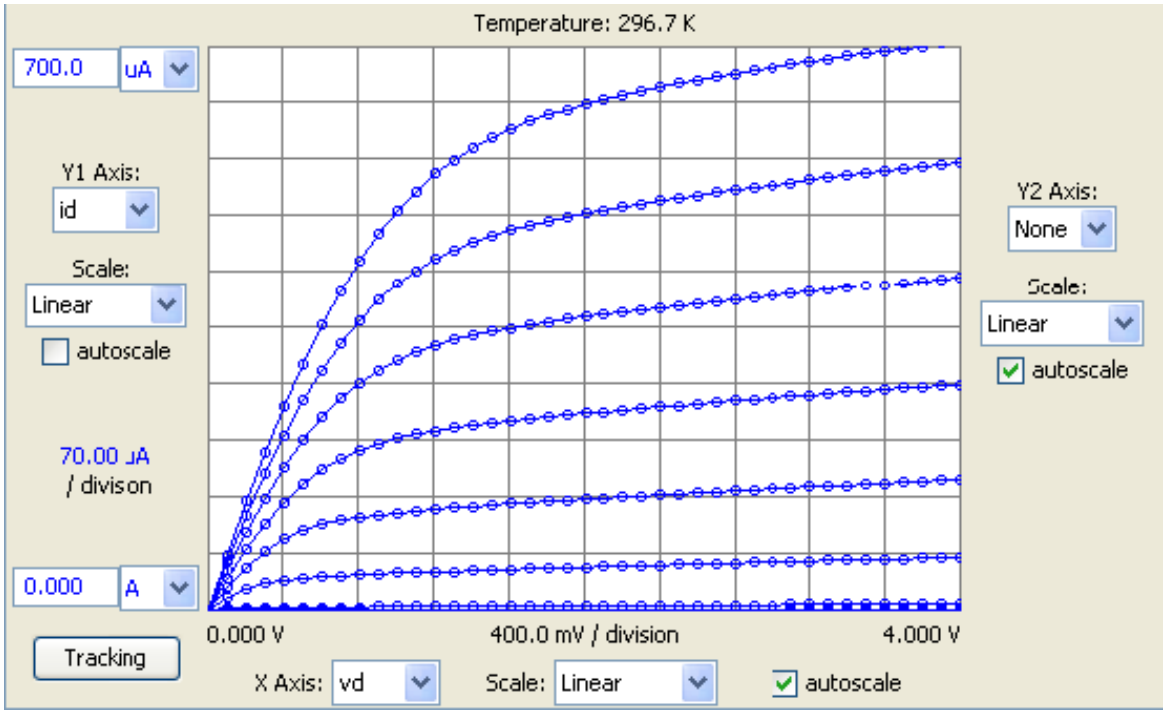
$$I_D = \frac{W}{L}\mu_n C_{ox}(V_{GS} - \frac{V_{DS}}{2} - V_T)V_{DS}$$

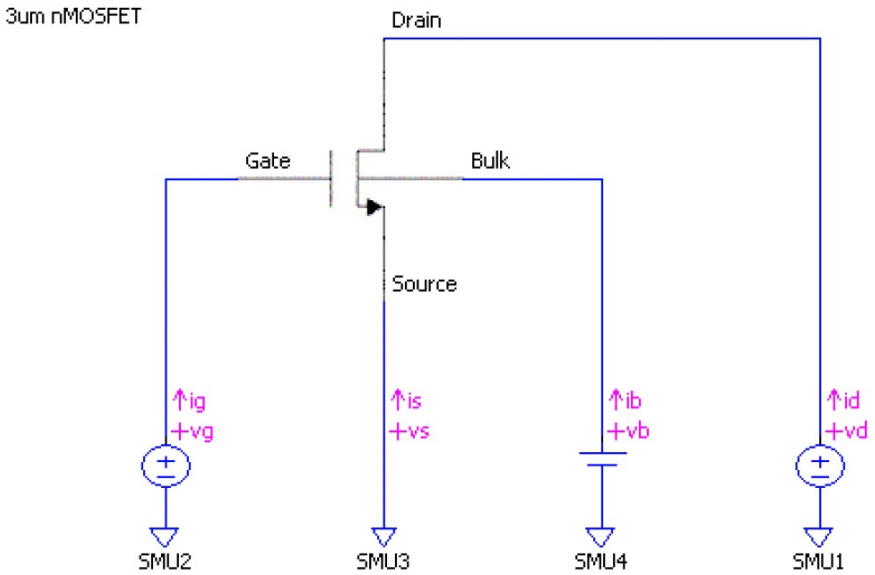
• **Saturation**: $V_{GS} > V_T$, $V_{DS} > V_G - V_T = V_{Dsat}$

$$I_{Dsat} = \frac{W}{2L}\mu_n C_{ox}(V_{GS} - V_T)^2$$

MOSFET: I-V characteristics

Output characteristics: $I_{DS} – V_{DS}$ @ V_{GS} = parameter

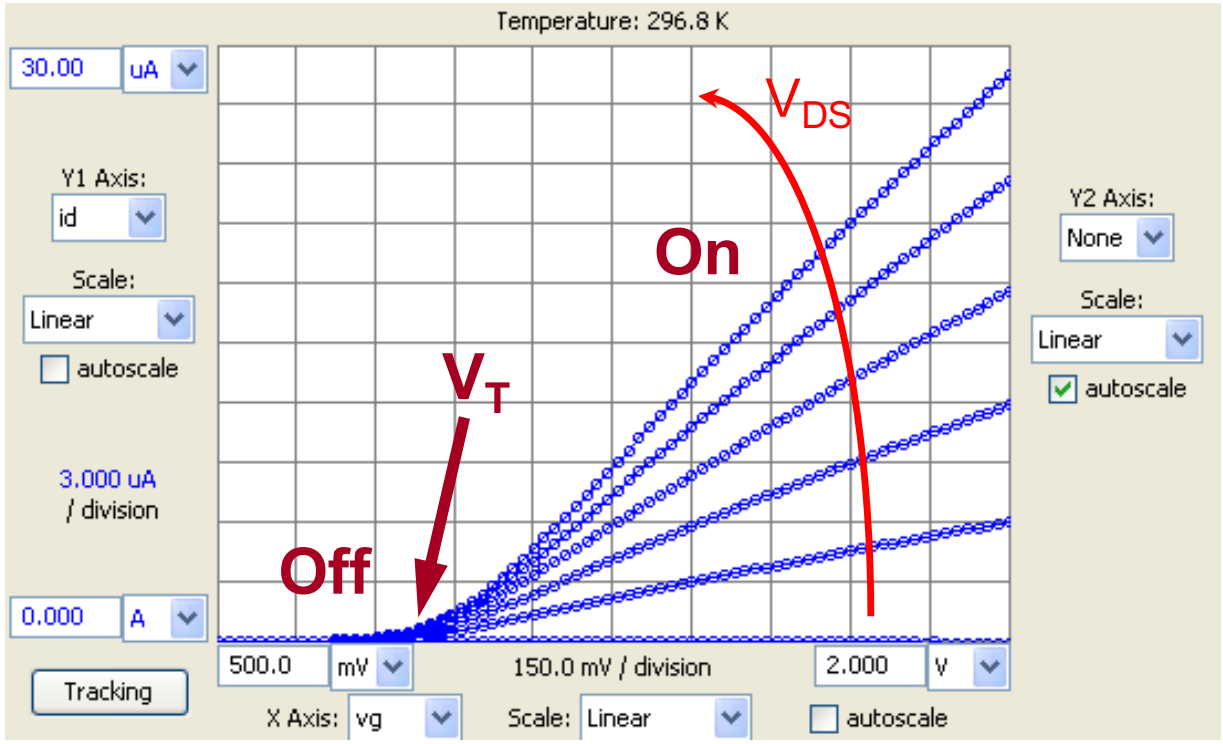




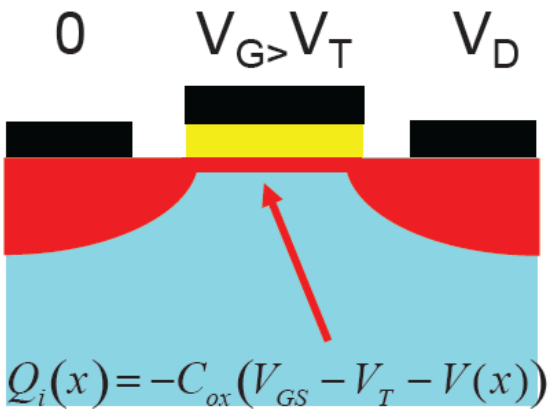
(V_{GS}=0 – 4 V, ΔV_{GS}=0.5 V):

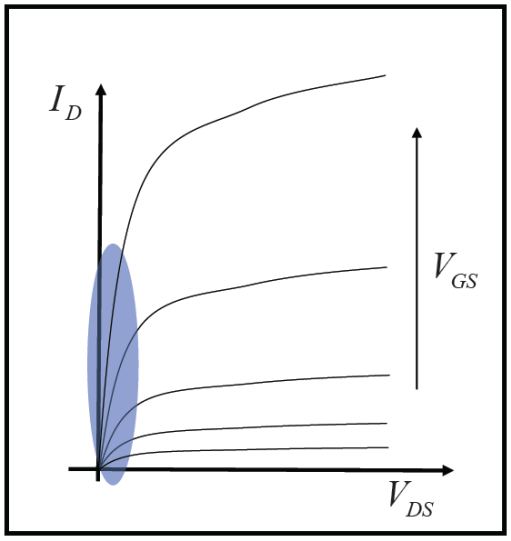
MOSFET: I-V characteristics

Transfer characteristics: $I_{DS} - V_{GS}$
@ V_{DS} = parameter



MOSFET in linear region, strong inversion @ $V_{GS} > V_T$ and low drain voltage, V_{DS}





$$I_D = W\, Q_i(x)\, \upsilon_x(x)$$

$$I_D = W\, C_{ox}(V_{GS} - V_T)\mu_{eff}\mathcal{E}_x$$

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

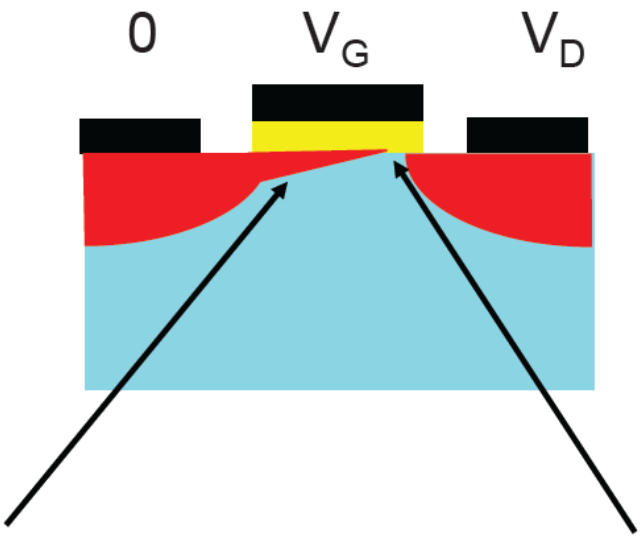
$$I_D = \frac{W}{L}\mu_{eff} C_{ox}(V_{GS} - V_T)V_{DS}$$

Also modeled as resistor:

$$R_{DS} = \frac{V_{DS}}{I_{DS}} = \frac{L}{W\mu_{eff} C_{oxe}(V_G - V_T)}$$

μ_{eff} is the ***effective electron mobility***

MOSFET: pinch-off at high V_{DS}



Cross-over between the linear and saturation regimes

(note that in BJT devices totally different terminology is adopted)

$$Q_i(x) = -C_{ox}\left(V_{GS} - V_T - V(x)\right)$$

$$V(x) = \left(V_{GS} - V_T\right)$$

$$Q_i(x) \approx 0$$

Saturation (1)

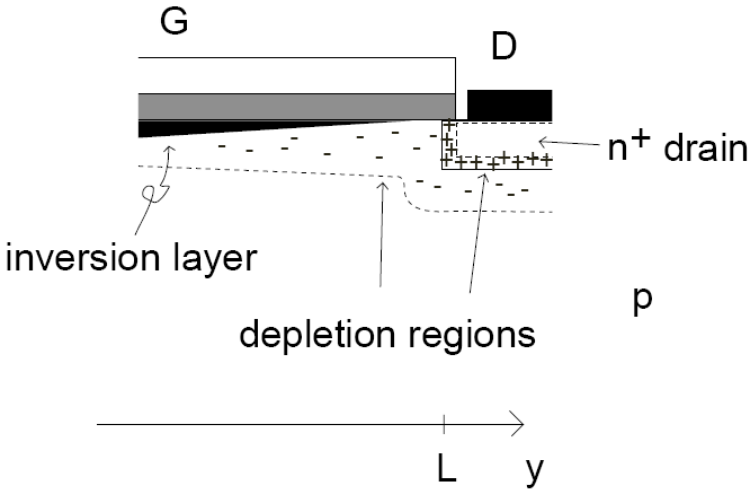
What happens when $V_{DSsat} = V_{GS} - V_T$?

- Charge control relation at drain end of channel:

$$Q_n(L) = -C_{ox}(V_{GS} - V_{DS} - V_T) = 0$$

- No inversion layer at end of channel ⇒ **Pinch-off**

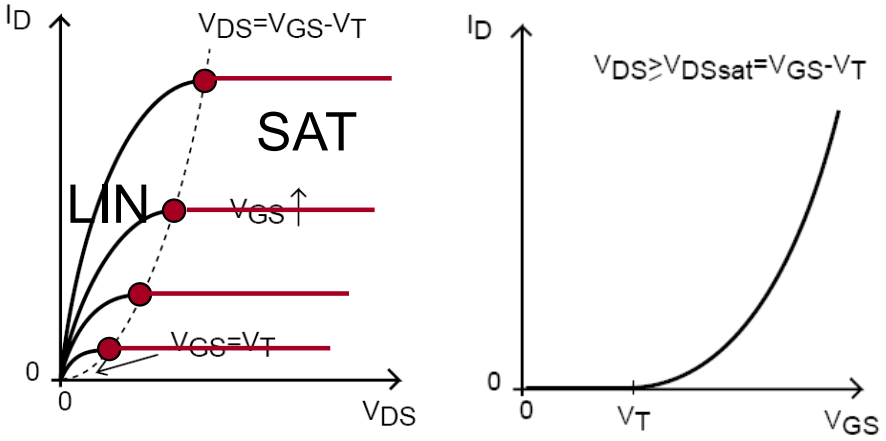
- At pinchoff:
  – electron concentration small but not zero
  – electrons move fast because electric field is very high
  – dominant electrostatic feature: acceptor charge
  – there is no barrier to electron flow (on the contrary!)



Saturation (2)

$$I_{Dsat} = \frac{W}{2L}\mu_n C_{ox}(V_{GS} - V_T)^2$$

Output charact. Transfer charact.



Drain current at pinchoff:

∝ lateral electric field

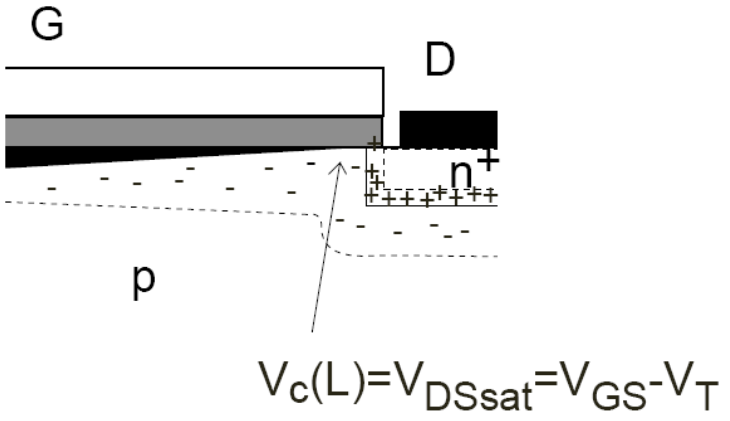
∝ $V_{DSsat} = V_{GS} - V_T$

∝ electron concentration

∝ $V_{GS} - V_T$

⇒ $\mathbf{I_{Dsat} \propto (V_{GS} - V_T)^2}$

- $I_{Dsat} \propto 1/L$

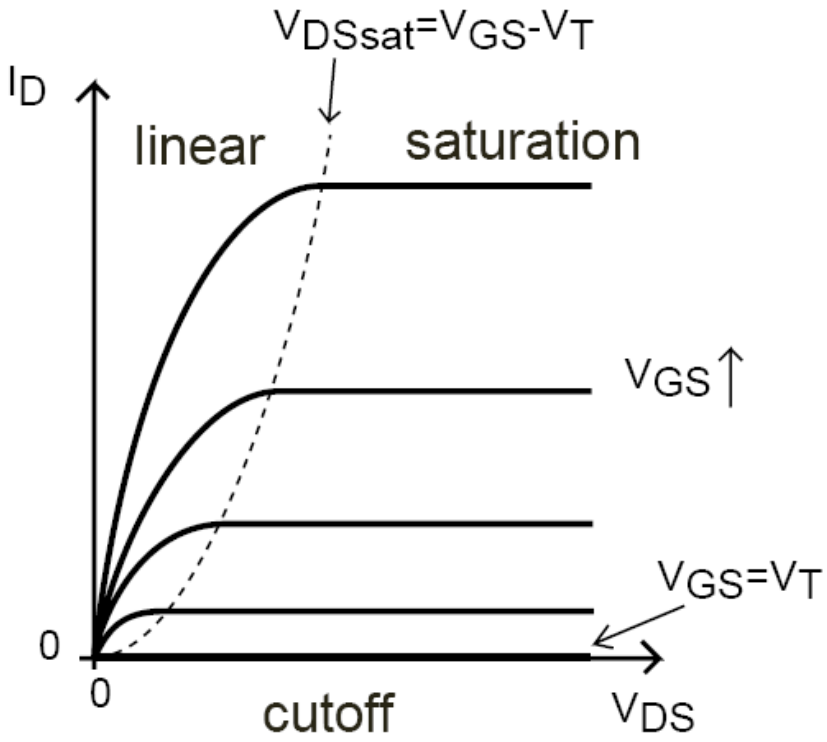


MOSFET Scaling

Key questions

- How does the channel geometry affects impact MOSFET operation in **saturation**? Does the pinch-off point represent a block to the current flow?
- What is the **short channel effect**?
- How should one scale down the MOSFET: what is **constant-field scaling** rule? How the scaling affects the device (technological) parameters?
- What are the **CMOS performance factors**?

MOSFET: regions of operation



• **Cutoff**: $V_{GS} < V_T$ ($V_{GD} < V_T$): no inversion layer anywhere underneath gate

$$I_D = 0$$

• **Linear**: $V_{GS} > V_T$, (with $V_{DS}>0$): inversion layer everywhere underneath gate

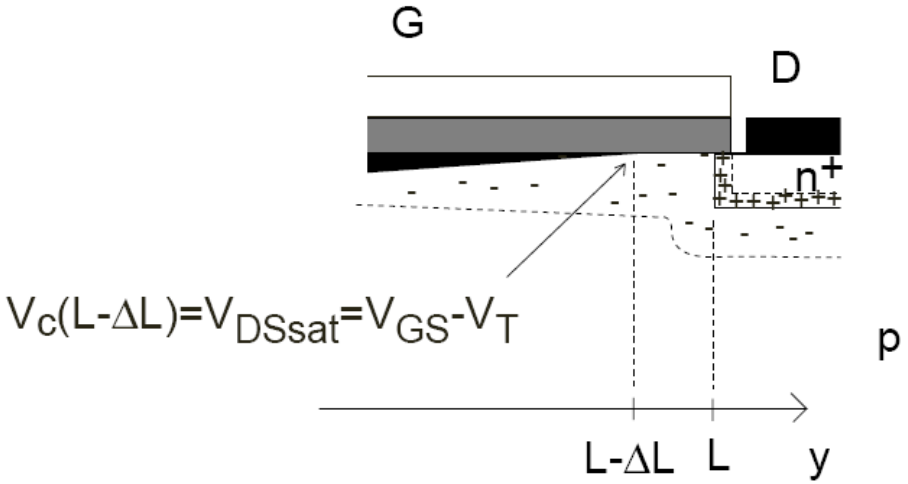
$$I_D = \frac{W}{L}\mu_n C_{ox}(V_{GS} - \frac{V_{DS}}{2} - V_T)V_{DS}$$

• **Saturation**:
$V_{GS} > V_T$, $V_{DS} > V_G - V_T = V_{Dsat}$

$$I_{Dsat} = \frac{W}{2L}\mu_n C_{ox}(V_{GS} - V_T)^2$$

MOSFET: channel length modulation (1)

What happens if $V_{DS} > V_{DSsat} = V_{GS} - V_T$?
Depletion region separating pinch-off point and drain widens (just like in reverse biased pn junction)



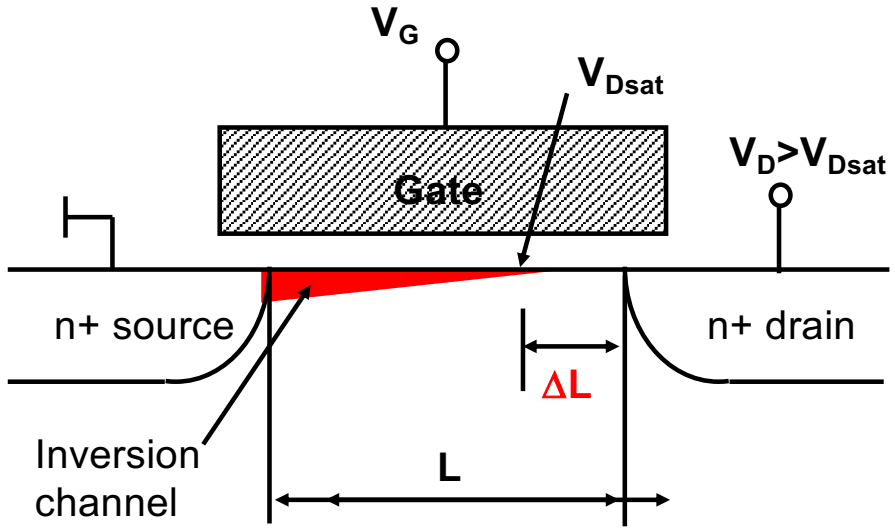
Channel length modulation: $V_{DS}\uparrow \Rightarrow L_{channel}\downarrow \Rightarrow I_D \uparrow$

MOSFET: channel length modulation (2)

Drain current continues to increase after saturation ($V_D > V_{Dsat}$)

- Pinch-off
- Equivalent length: L - ΔL

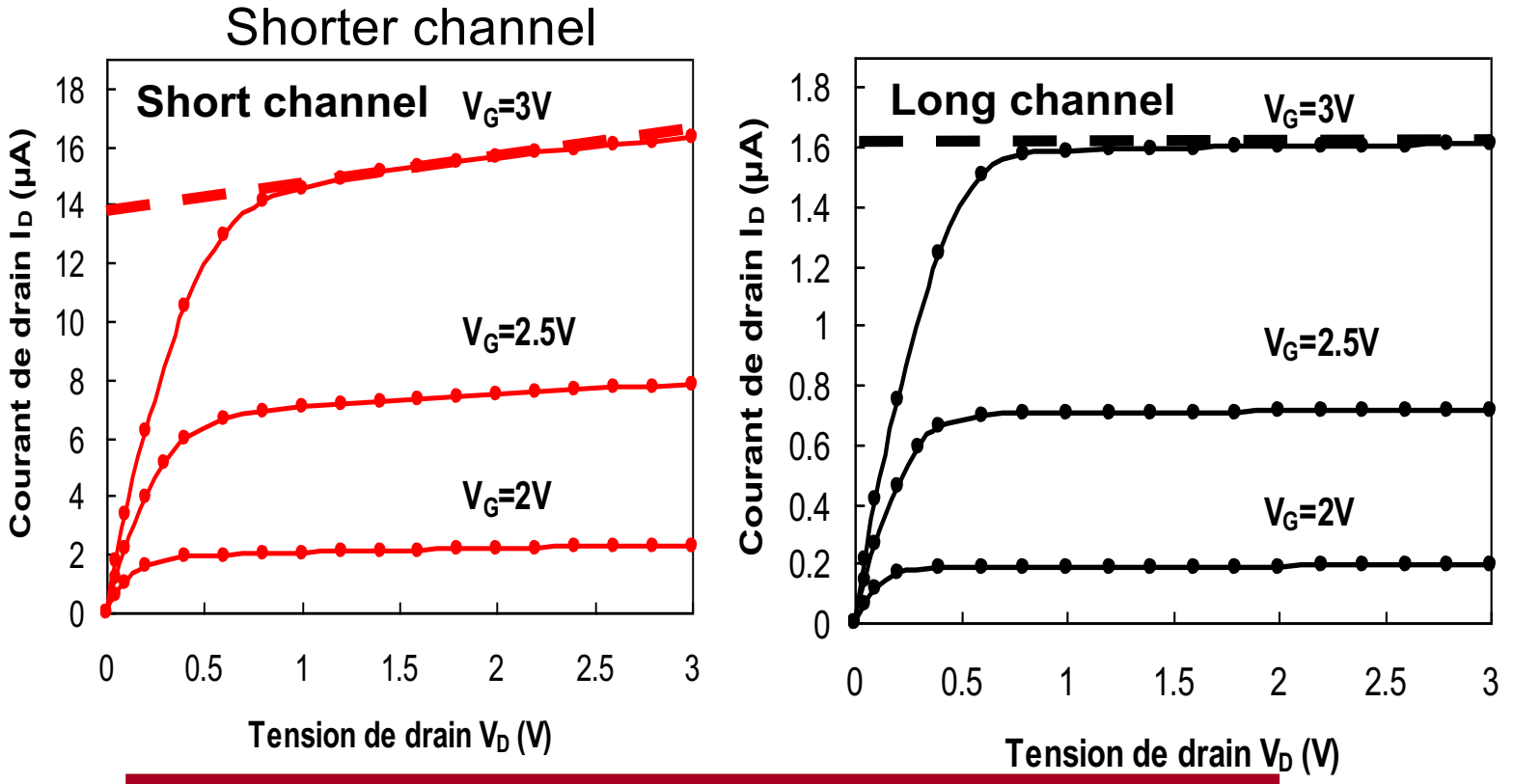
$$I_D = \frac{I_{Dsat}}{1-(\Delta L / L)}$$



Long channel	**Short channel**
$\Delta L << L \rightarrow I_D \approx I_{Dsat}$ = ct. (V_D)	$\Delta L \nearrow$ with $V_D \rightarrow I_D \nearrow$ with V_D

MOSFET: channel length modulation

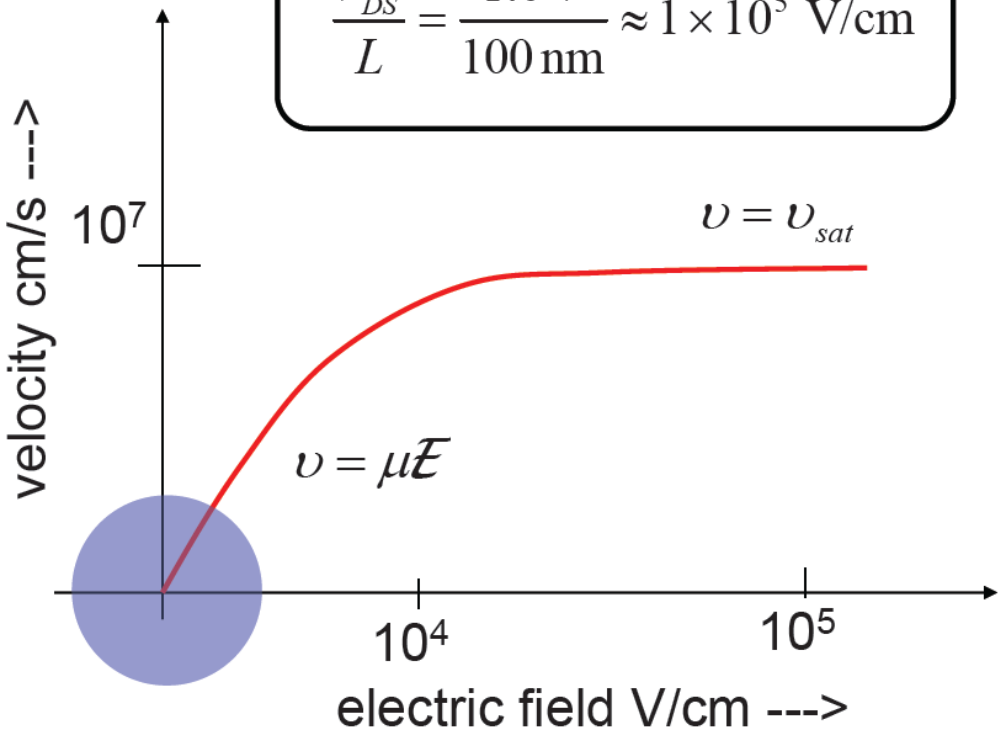
How it looks like and is modeled?



$$I_D = I_{Dsat} = \frac{W}{2L}\mu_n C_{ox}(V_{GS}-V_T)^2[1+\lambda(V_{DS}-V_{DSsat})]$$

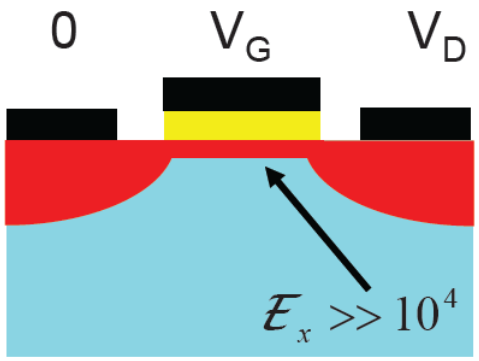
MOSFET: velocity saturation @ small channel length (1)

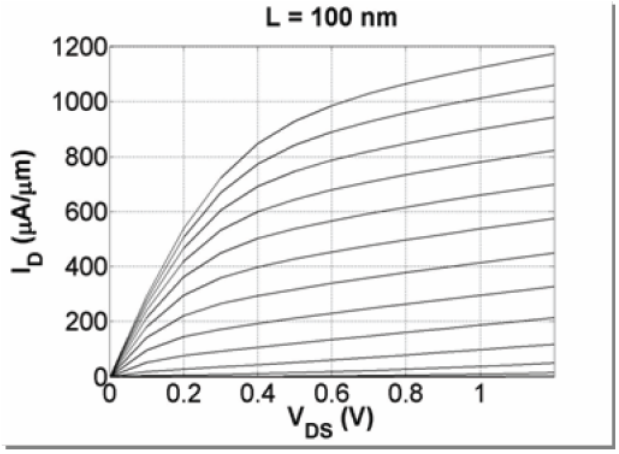
$$\frac{V_{DS}}{L} = \frac{1.0\,\text{V}}{100\,\text{nm}} \approx 1 \times 10^5\ \text{V/cm}$$



In short-channel MOSFETs with velocity saturation, the current saturates **earlier** and becomes **limited by v_{sat}** rather than pinch-off.

MOSFET: velocity saturation @ small channel length (2)





$$I_D = W Q_i(x) \upsilon_x(x)$$

$$I_D = W C_{ox}(V_{GS} - V_T)\upsilon_{sat}$$

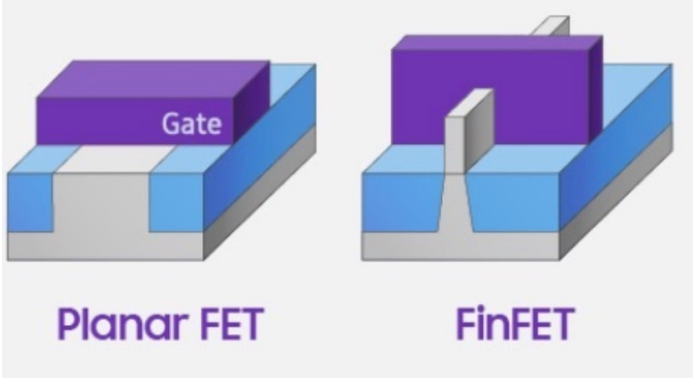
$$I_D = W C_{ox} \upsilon_{sat} (V_{GS} - V_T)$$

saturation means that the **current is linearly dependent on V_{GS}** rather than quadratically, as in the long-channel model

Small channel effects: length scale

• For modern MOSFETs, short-channel effects start to become significant when the channel length drops below roughly 100 nm. The effects increase dramatically as the length approaches 20 nm and below.

In today's tech (5nm/3nm nodes):
• The physical gate length is often below 20 nm, and controlling SCEs becomes extremely difficult.



• That's why planar bulk MOSFETs are largely replaced by FinFETs or even Gate-All-Around FETs (GAAFETs) in cutting-edge nodes.

•These 3D structures provide better electrostatic control over the channel, which helps to better manage SCEs at extreme scaling.

MOSFET scaling: Dennard's rule

*proposed in **1974 by Robert Dennard** and his team at IBM*

*describes how to **scale down** MOSFET dimensions to make smaller, faster, and more power- efficient transistors — the principle that fueled Moore's Law for decades*

• keeping short-channel effects under control by scaling down vertical and horizontal dimensions while also proportionally decreasing the applied voltages and increasing the substrate doping (decreasing the depletion width)

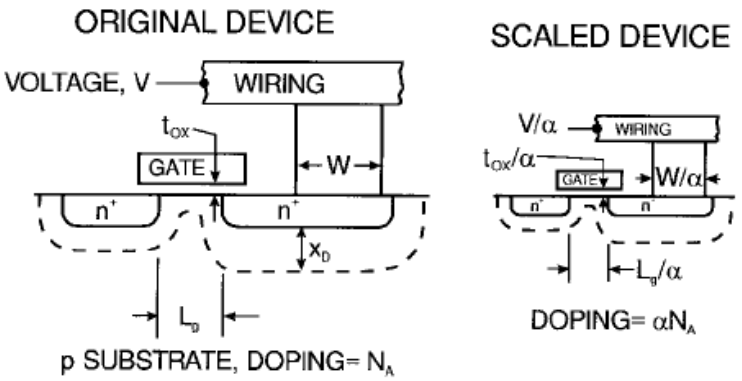
• the principle of constant-field scaling lies in scaling the device voltages and dimensions by the same factor α, so that the electric field remains unchanged

Before scaling:		**After scaling:**
Channel length	= L eff	**Leff / α**
Oxide thickness	= t ox	**tox / α**
Channel doping	= N_B	**N_B x α**
Operating voltage	= V_{DD}	**V_{DD} / α**

• Most important: once the device dimensions and power-supply are scaled, the circuit speeds up by the same factor!

MOSFET scaling: Dennard's rule

Parameter	Scales by	Description
Channel length (L)	$1/\kappa$	Shorter channel
Width (W)	$1/\kappa$	Smaller width
Supply voltage (V)	$1/\kappa$	Lower voltage
Threshold voltage	$1/\kappa$	Also scaled to maintain switching behavior
Capacitance (C)	$1/\kappa$	Smaller gate area
Current (I)	$1/\kappa$	Lower current due to lower voltage
Delay time (t)	$1/\kappa$	Faster switching
Power per circuit	$1/\kappa^2$	Less power per transistor
Power density	≈ constant	Because area and power scale together



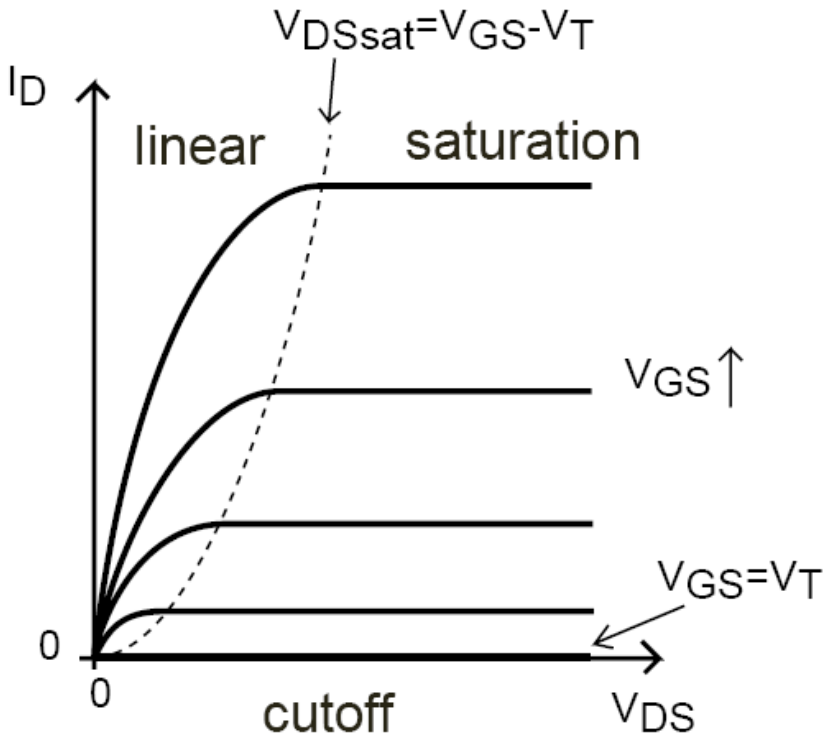
post-Dennard era came in mid-2000s

Subthreshold leakage increased with smaller transistors

Inability to reduce threshold and supply voltages much further (leads to performance and reliability issues)

High power density causing thermal problems

Summary- I: 3 principle regimes of MOSFET operation



• **Cutoff**: $V_{GS} < V_T$ ($V_{GD} < V_T$): no inversion layer anywhere underneath gate

$$I_D = 0$$

• **Linear**: $V_{GS} > V_T$, (with $V_{DS} > 0$): inversion layer everywhere underneath gate

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

• **Saturation**: $V_{GS} > V_T$, $V_{DS} > V_G - V_T = V_{Dsat}$

$$I_{Dsat} = \frac{W}{2L}\mu_n C_{ox}(V_{GS} - V_T)^2$$

Summary- II: scaling MOSFET

- Short-channel effects impact planar FET scaling and must be taken into account

- Channel length modulation and velocity saturation are among the most important factors associated with the short channel

- Dennard scaling rule offer an approach that permits keeping the short channel effects under control

- Modern technology nodes belong to “post-Dennard era” where alternative 3D geometries are required.