MOSFET (I): Fundamentals

Material developed by Prof. C. Z. Zhao

OUTLINE

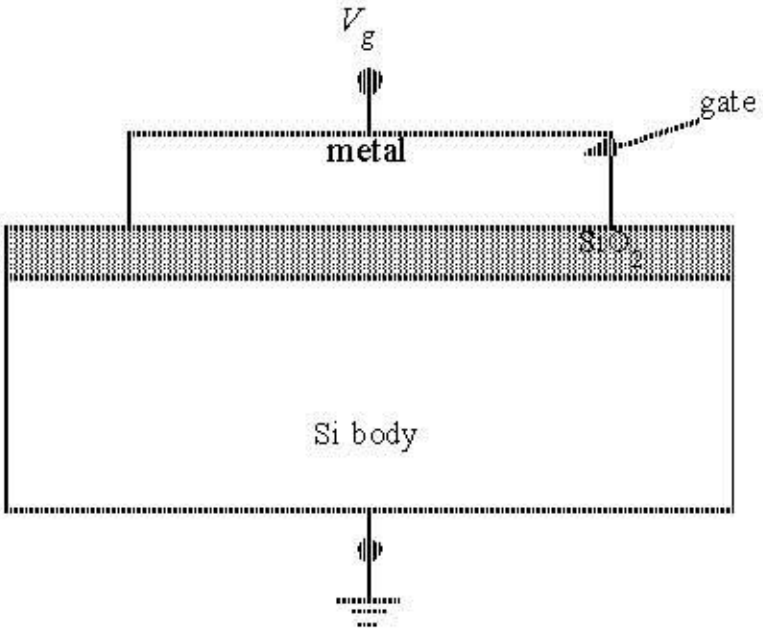
- I - V Characteristics
  - Cutoff Region
  - Linear Region
  - Saturation Region (pinch-off region)
- Switch model of nMOSFETs

Reading: Chapter 3.3

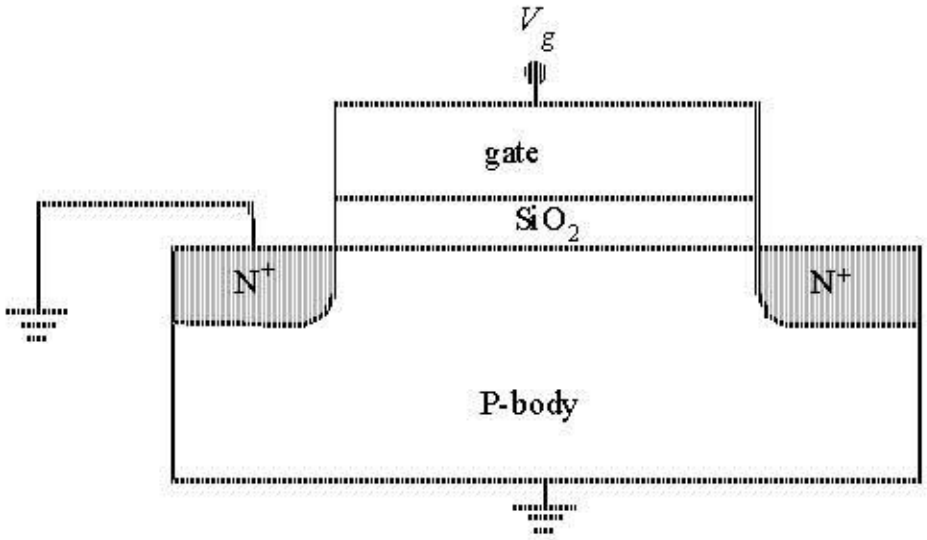
MOS Capacitors

PL

MOS: Metal-Oxide-Semiconductor



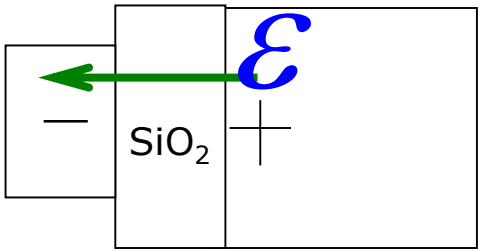
MOS capacitor

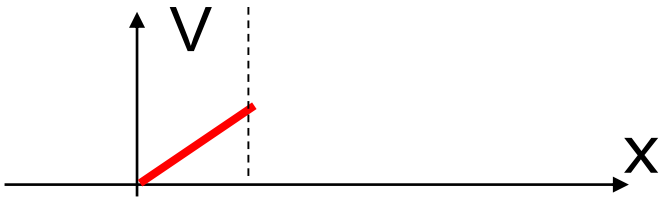


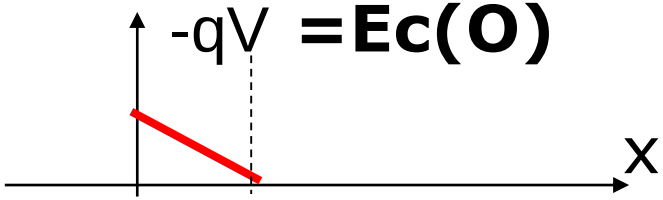
MOS transistor

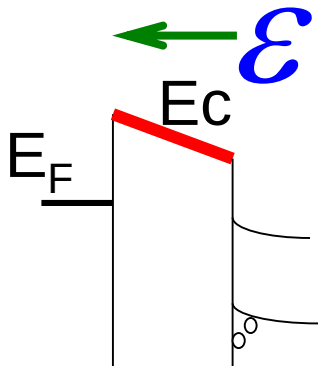
Ec(O) and electric field direction

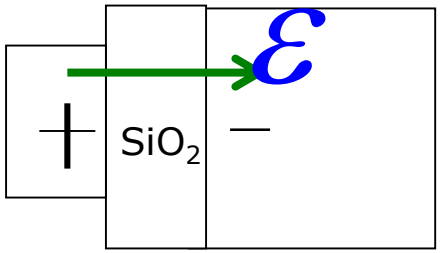
PL

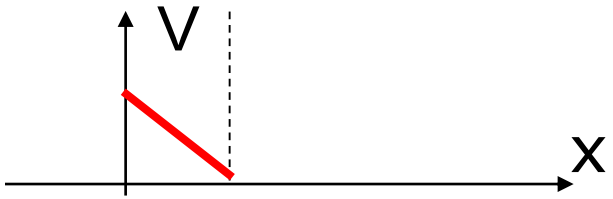


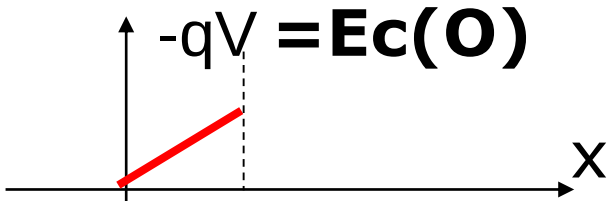


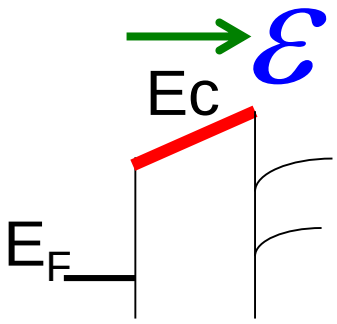










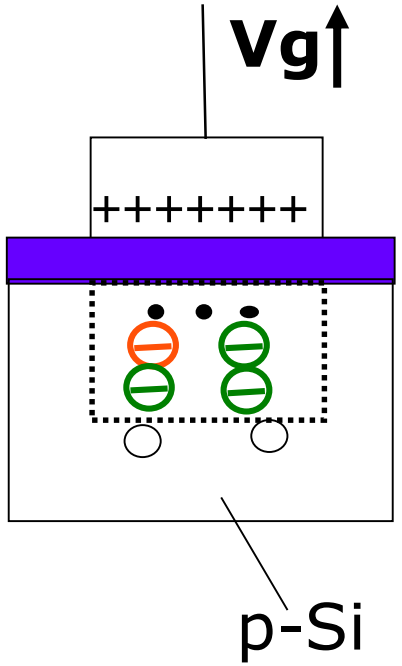


4. Energy band diagram: $Vg>V_m$

PL

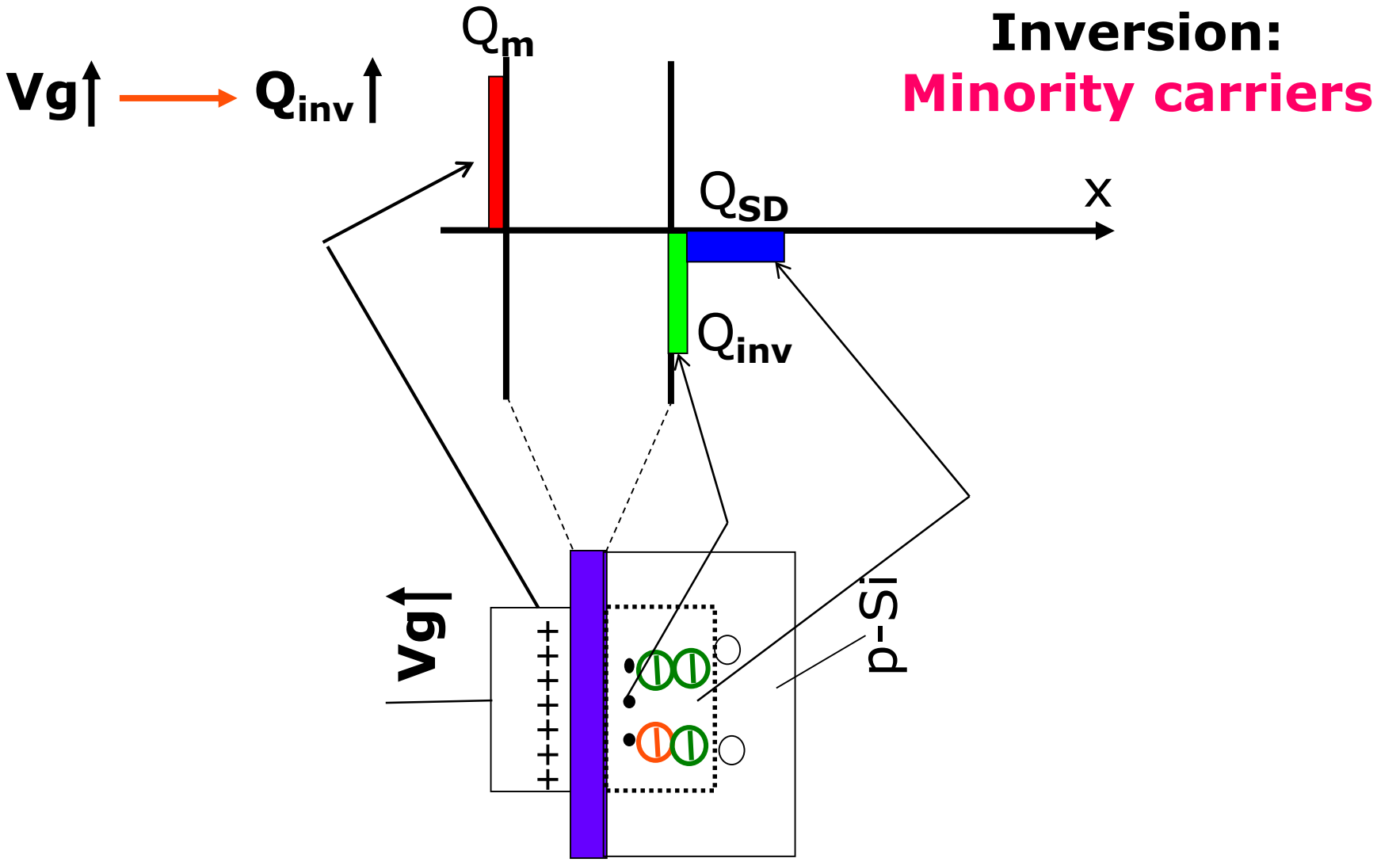
Vg↑ ⟶ Q_{inv}↑

Inversion:
Minority carriers



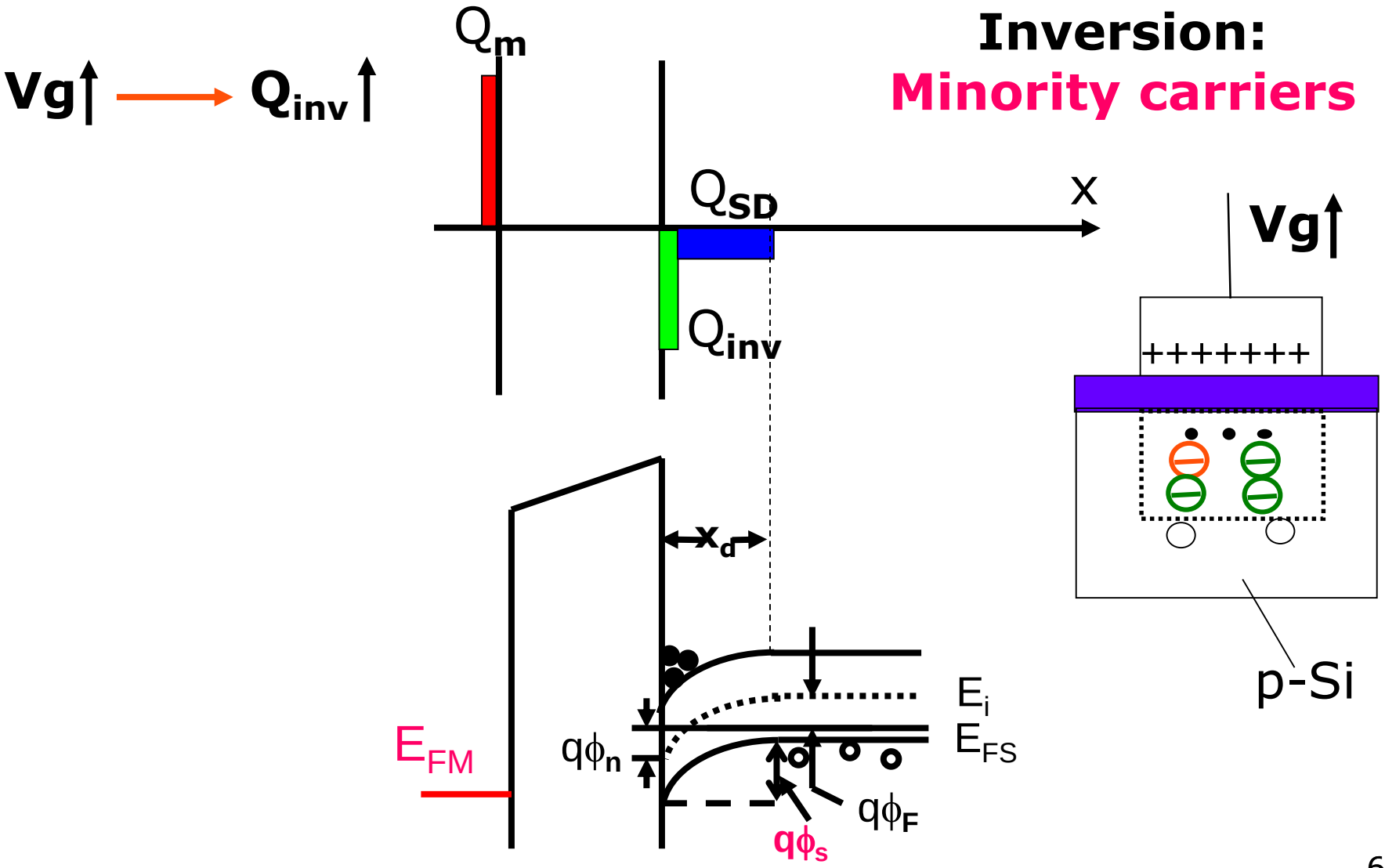
4. Energy band diagram: $Vg>V_m$

PL



4. Energy band diagram: $Vg>V_m$

PL

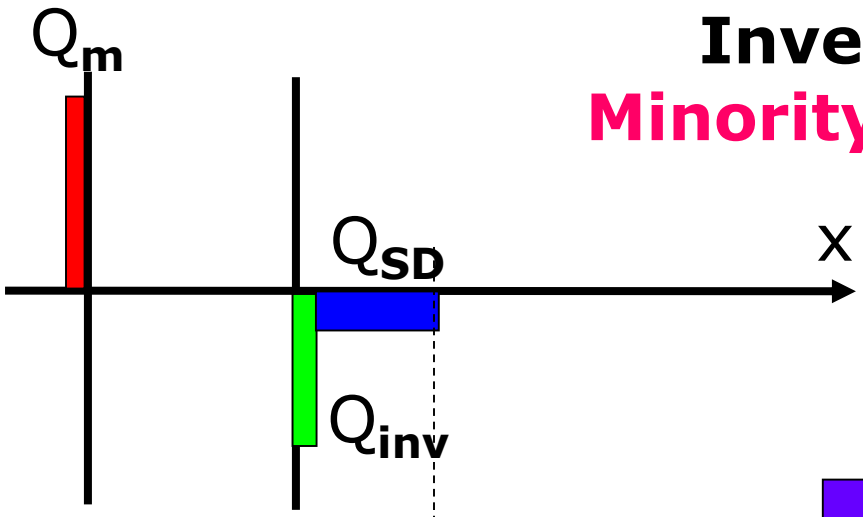


4. Energy band diagram: Vg>V_m

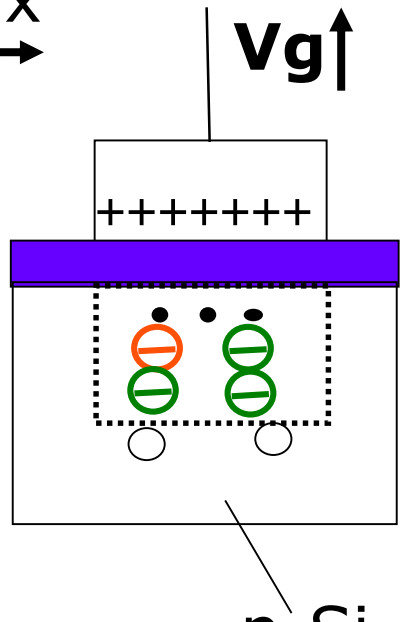
PL

Vg↑ → Q_{inv}↑

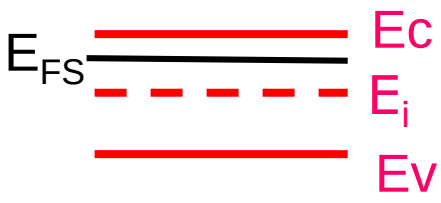
As Vg>Vm, E_{FS} is at the up-half of the bandgap

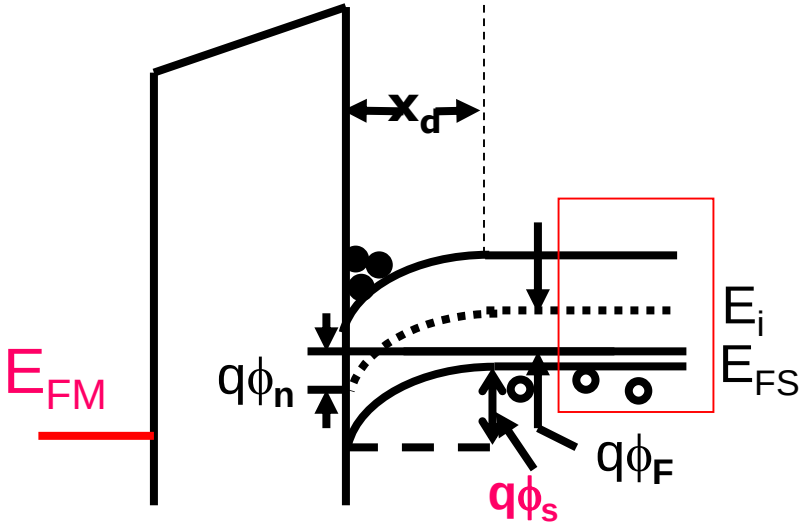


Inversion:
Minority carriers



At the interface





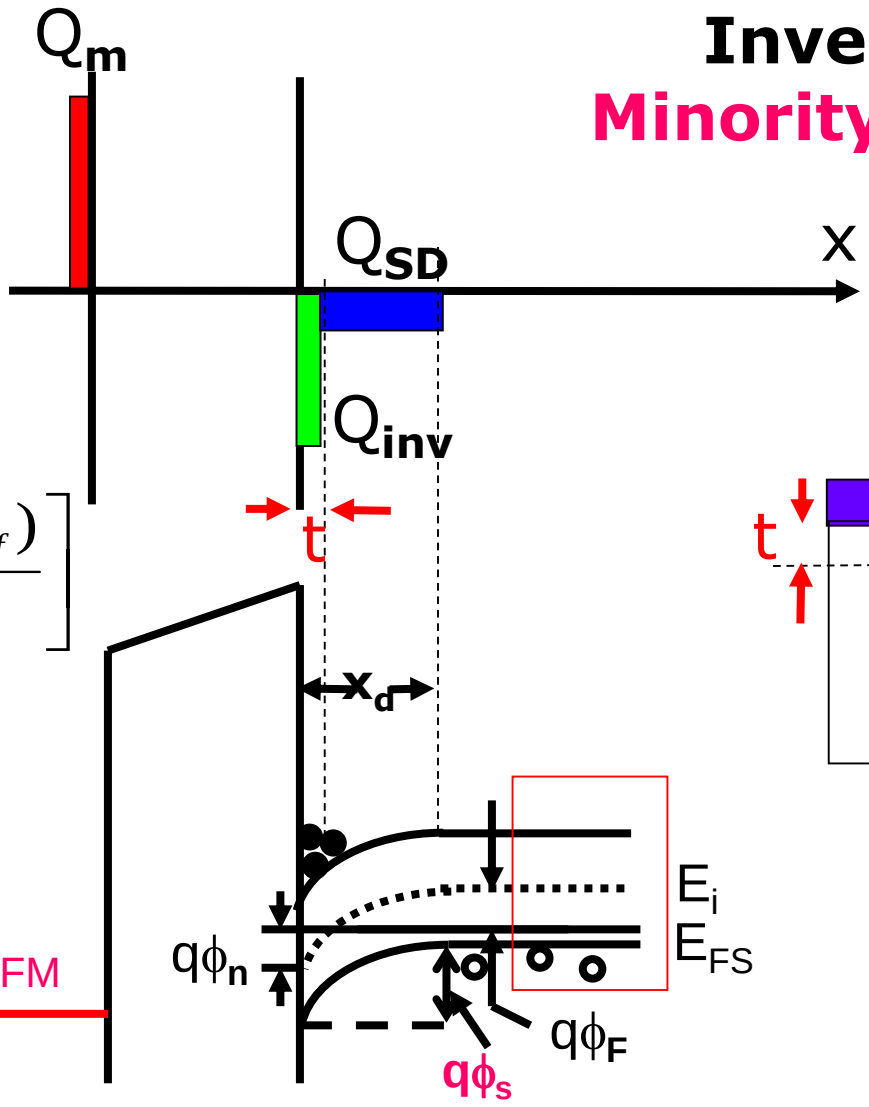
4. Energy band diagram: Vg>V_m

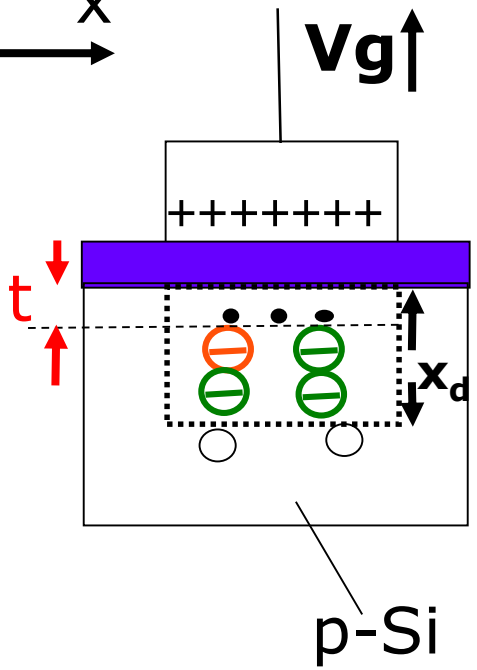
PL

Vg↑ → Q_{inv}↑

C/cm²

Inversion:
Minority carriers





$$n_s = n_i \exp\left[\frac{(E_F - E_{i,surf})}{kT}\right]$$

electrons/cm³

$Q_{inv} = q\, n_s\, t$

4. Inversion

PL

- **Weak Inversion**: **$0<\phi_n<\phi_F$** **($\phi_F < \phi_S < 2\phi_F$)**
- **Strong Inversion: $\phi_n \geq \phi_F$** ($\phi_S \geq 2\phi_F$), electron density at the interface ≥ hole density in Si bulk.
- Vg for strong inversion: **V_T** 'threshold voltage'.

$$\mathbf{Vg = V_T \rightarrow \phi_s = 2\phi_F \rightarrow \phi_n = \phi_F \rightarrow n_s = p_b}$$

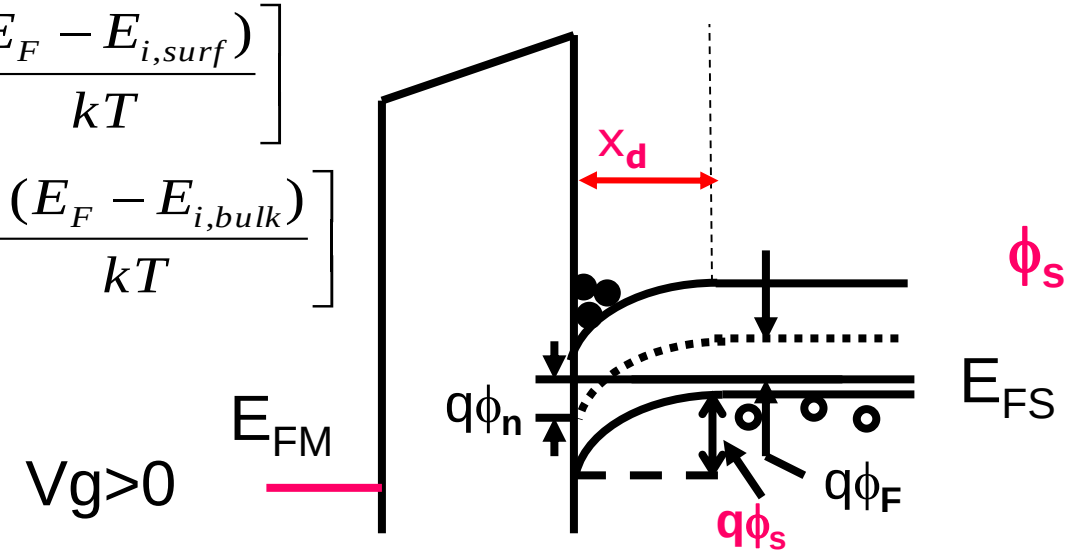
$$n_s = n_i \exp\left[\frac{(E_F - E_{i,surf})}{kT}\right]$$

$$p_b = n_i \exp\left[\frac{-(E_F - E_{i,bulk})}{kT}\right]$$

$$= N_A$$

$$p_s \approx n_i^2 / n_s$$

$$n_b = n_i^2 / p_b$$



$$\phi_s = \phi_F + \phi_n$$

Voltage drops in a MOS system PL

$$V_G = V_{FB} + V_{ox} + \phi_s$$

$V_g = V_T \rightarrow$
$\phi_s = 2\phi_F$

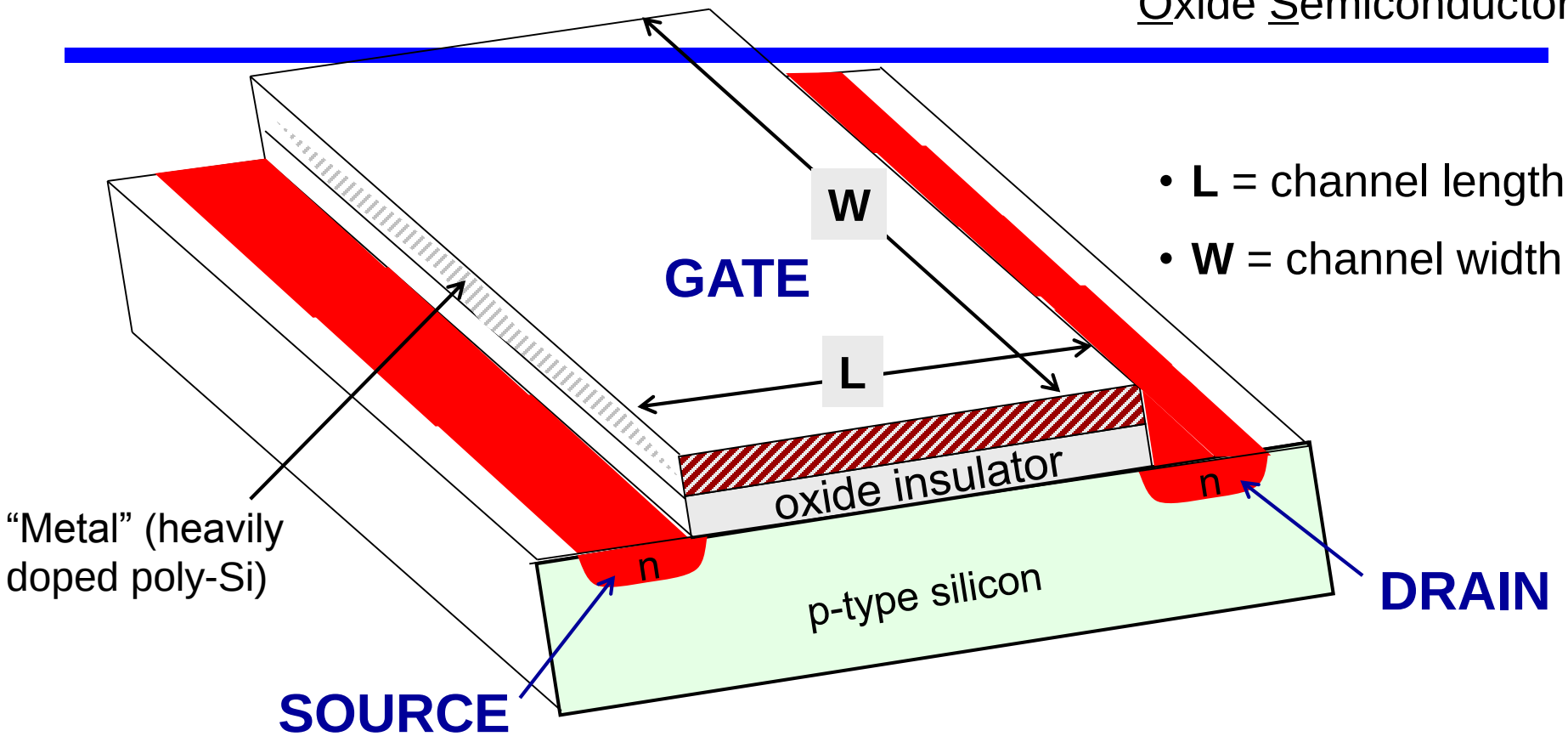
$$V_T = V_{FB} + V_{ox} + 2\phi_F$$

$$V_{ox} = +\frac{\sqrt{2qN_A\varepsilon_{Si}(2\phi_F)}}{C_{ox}}$$ for p-Si sub.

$$V_{ox} = -\frac{\sqrt{2qN_D\varepsilon_{Si}|2\phi_F|}}{C_{ox}}$$ for n-Si sub.

An nMOSFET

- **NMOS:** N-channel Metal Oxide Semiconductor

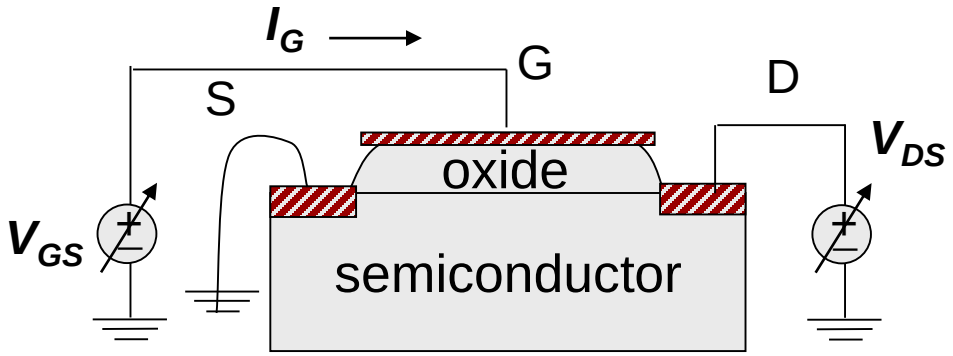


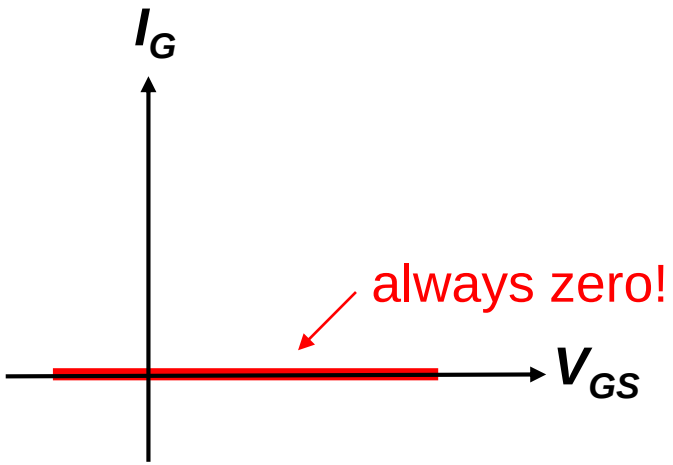
- **L** = channel length
- **W** = channel width

- A GATE electrode is placed above (electrically insulated from) the silicon surface, and is used to control the resistance between the SOURCE and DRAIN regions

Gate oxide

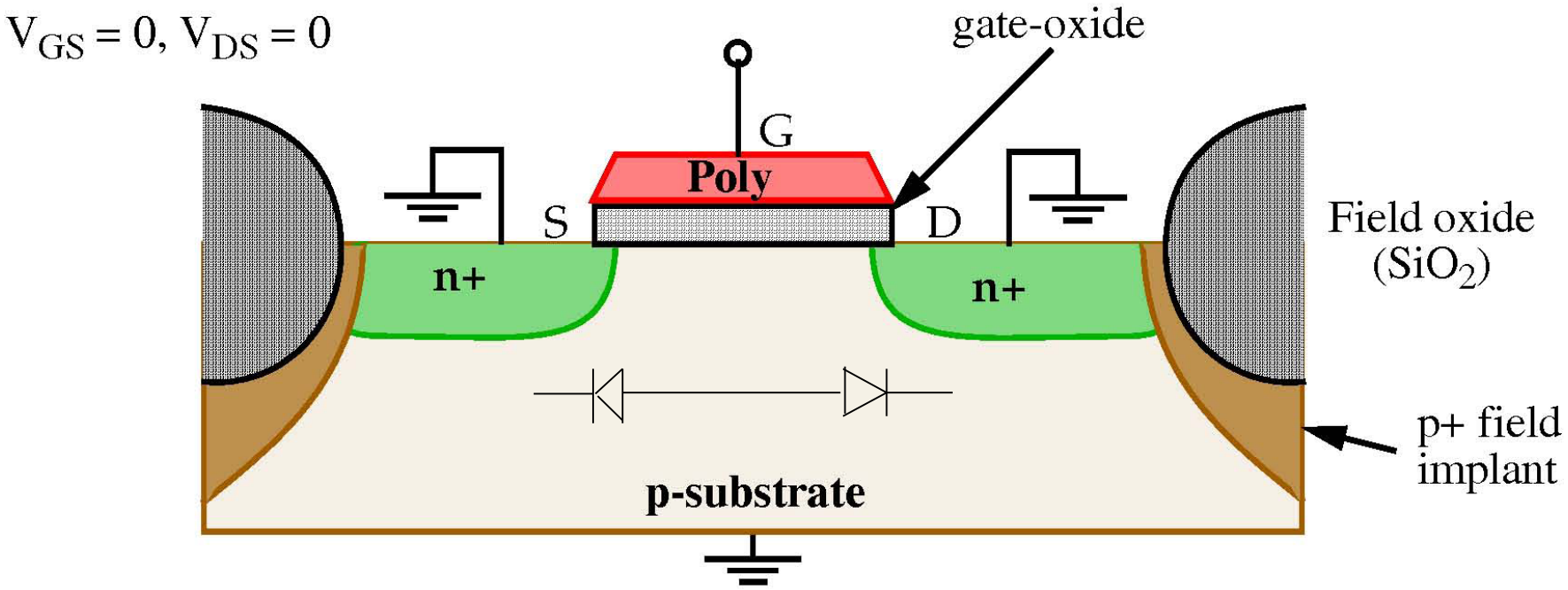
Consider the current I_G (flowing into **G**) versus V_{GS} :





The gate is insulated from the semiconductor, so there is no significant (steady) gate current.

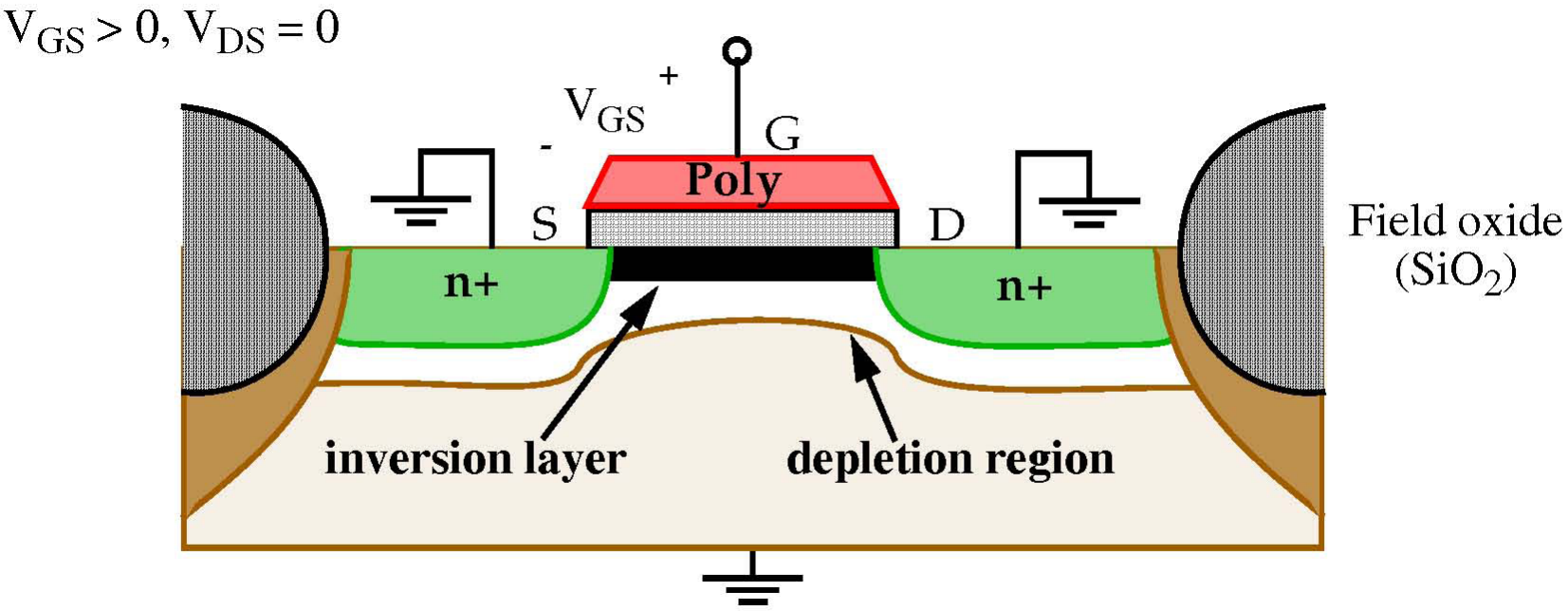
nMOSFET: $V_{GS}<V_T$



Under zero bias, two back-to-back *pn*-junctions create a very high resistive path between source and drain.

Appling a **positive bias** (V_{GS}) to the gate, creates a depletion region under the gate (repells mobile holes). The depletion region is similar to the one occurring in a *pn*-junction.

nMOSFET: $V_{GS}>V_T$



Inversion layer expressions are (the surface potential ϕ_s is $2\phi_F$):

Inversion layer charge $Q_{inv} = -C_{ox}(V_G - V_T)$

$V_G \uparrow \rightarrow Q_{inv} \uparrow$

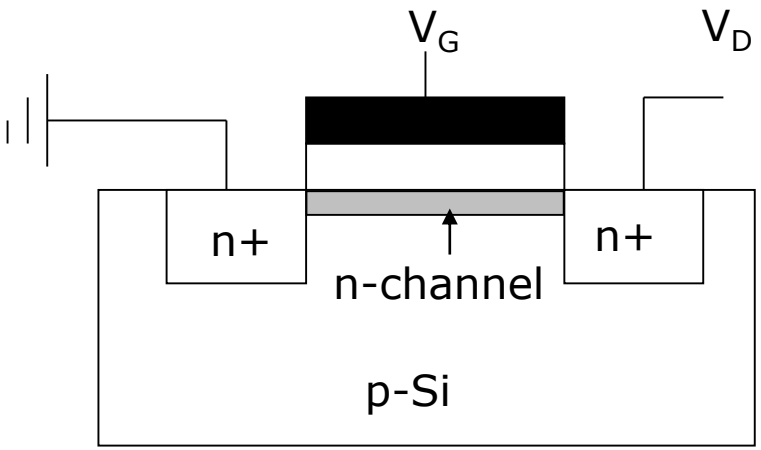
Threshold voltage $V_T = V_{FB} + 2\phi_F + \dfrac{\sqrt{2qN_A\varepsilon_{Si}(2\phi_F)}}{C_{ox}}$

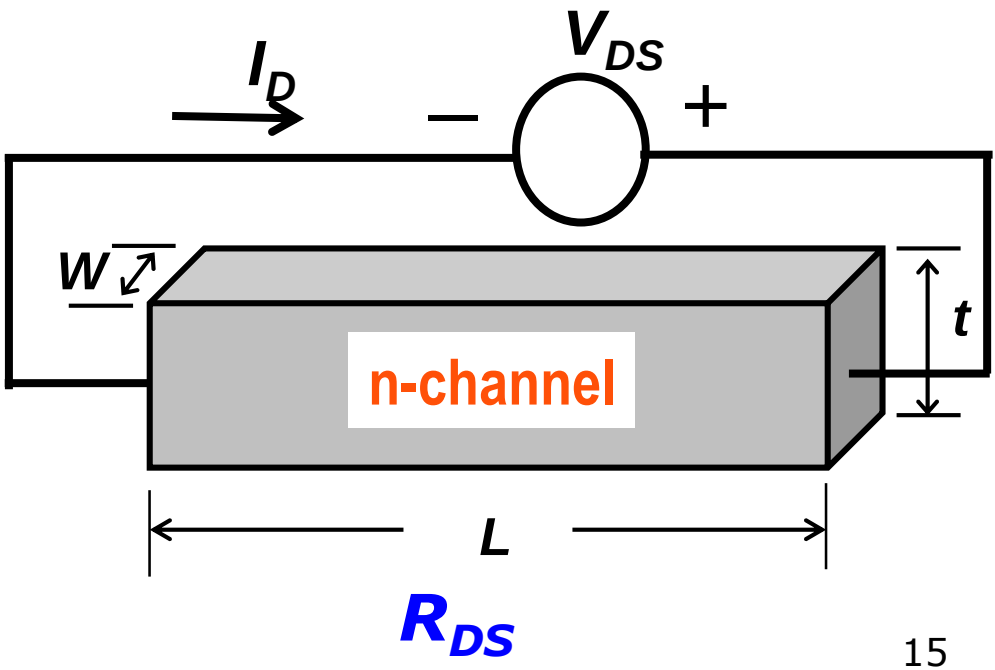
Inversion layer

$$Q_{inv} \times W \times L = q \times n \times t \times W \times L$$
$$\rightarrow Q_{inv} = q\,n\,t$$

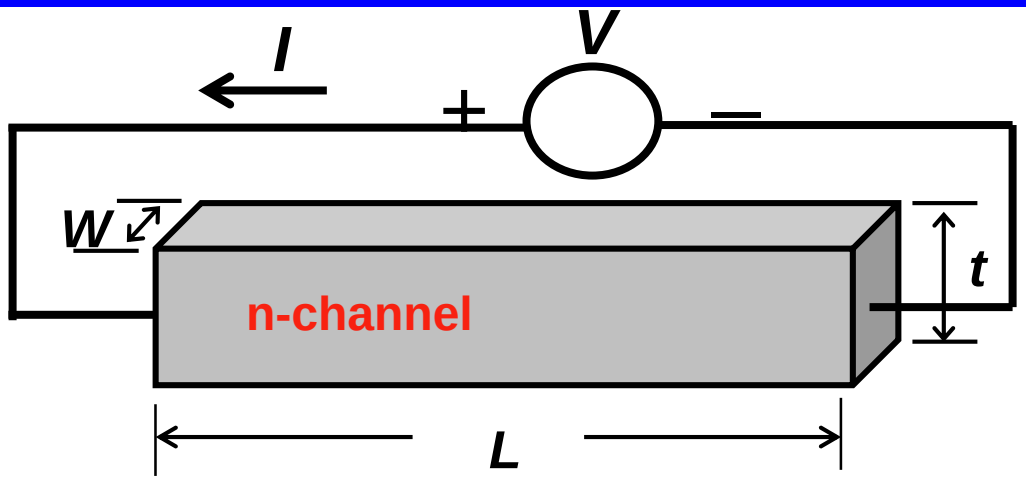
- Without gate bias, MOSFET is off because two diodes are "back-to-back". One of them will be reversely biased. To switch on, the interfacial region is inverted by applying a gate bias.

- Above a certain gate-to-source voltage (***threshold voltage*** V_T), a conducting layer of mobile electrons is formed at the Si surface beneath the oxide. These electrons can carry current between the source and drain.





Electrical Resistance



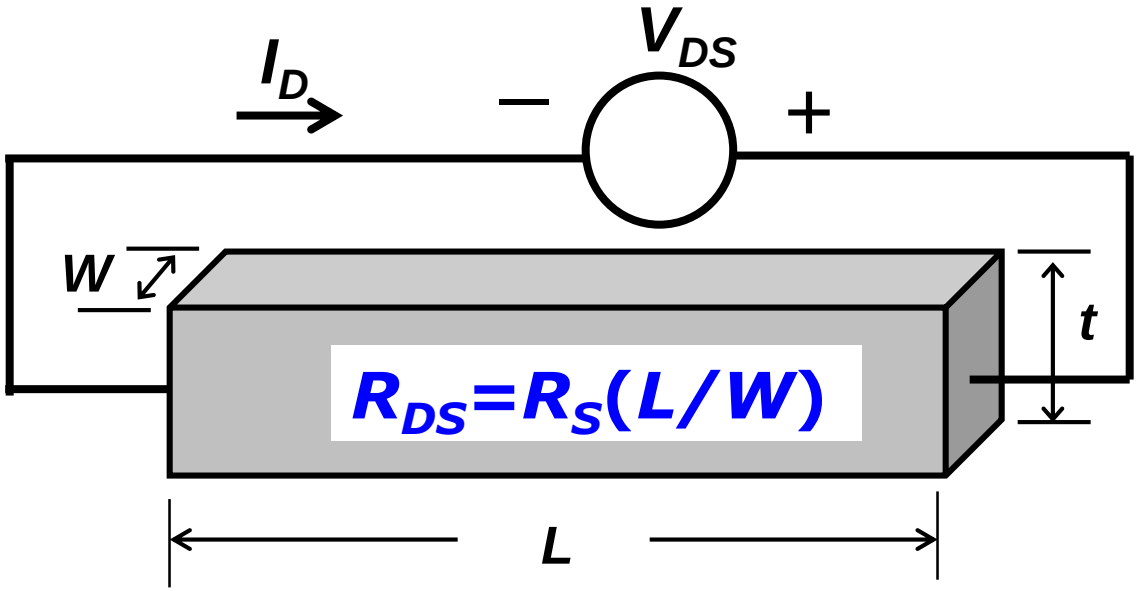
Resistance $R \equiv \frac{V}{I} = \rho \frac{L}{A} = \frac{\rho L}{tW} = \left(\frac{\rho}{t}\right)\left(\frac{L}{W}\right)$ (Unit: ohms)

where ρ is the resistivity ($\Omega \cdot \mathrm{cm}$)

R_S

Inversion layer as a resistor

Consider an n-channel:



$$R_s = \frac{\rho}{t} = \frac{1}{\sigma t} = \frac{1}{q\mu_n n_s t} = \frac{1}{\mu_n Q_{inv}}$$

$V_G \uparrow \rightarrow Q_{inv} \uparrow$, $R_s \downarrow$

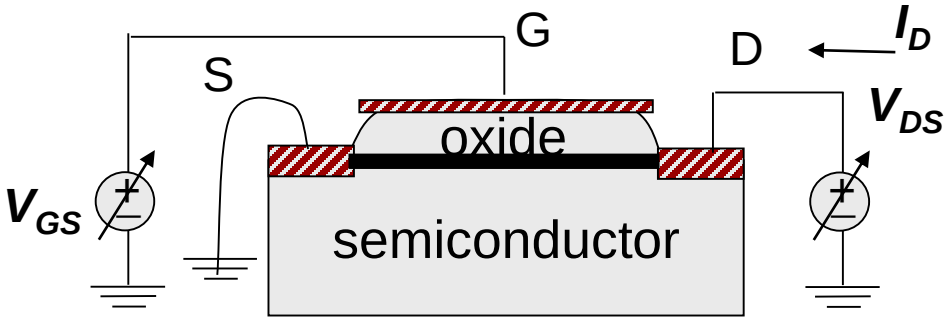
where Q_{inv} is the charge per unit area.

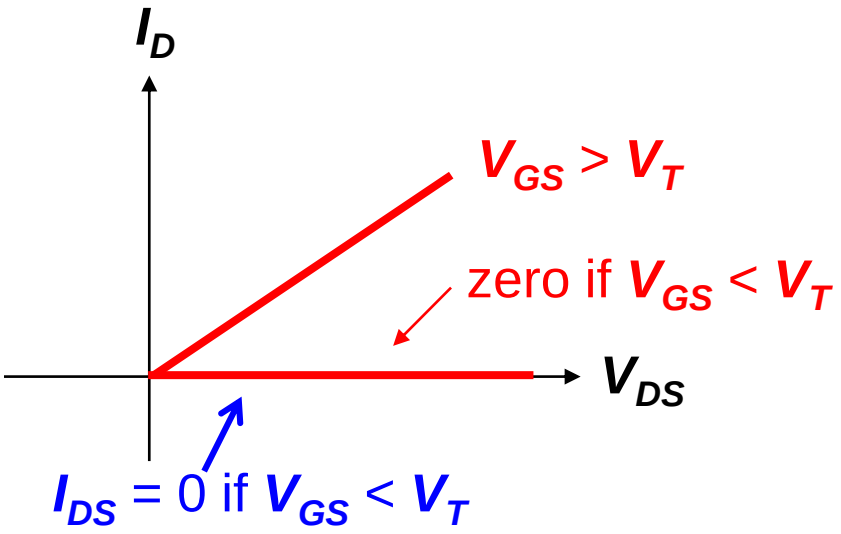
$$Q_{inv} = q\, n_s\, t$$

$$\rho \equiv \frac{1}{\sigma} = \frac{1}{qn\mu_n + qp\mu_p}$$

nMOSFET I_D vs. V_{DS} Characteristics

Next consider I_D (flowing into **D**) versus V_{DS}, as V_{GS} is varied:





"Cutoff" region: $V_{GS} < V_T$

Above threshold ($V_{GS} > V_T$):
"inversion layer" of electrons appears, so conduction between **S** and **D** is possible

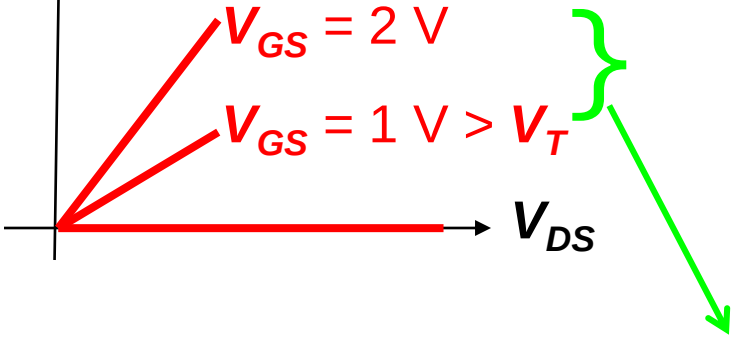
Below "threshold" ($V_{GS} < V_T$):
no charge → no conduction

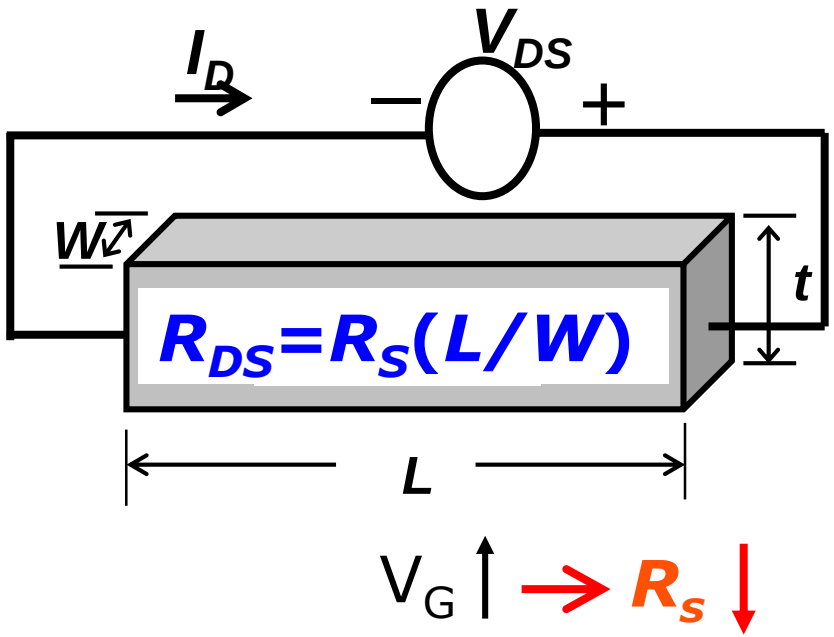
The MOSFET as a Controlled Resistor

- The MOSFET behaves as a resistor when V_{DS} is low:
  - Drain current I_D increases linearly with V_{DS}
  - Resistance R_{DS} between SOURCE & DRAIN depends on V_{GS}
    - R_{DS} is lowered as V_{GS} increases above V_T

NMOSFET Example:

$$I_D = \frac{V_{DS}}{R_{DS}}$$





Linear or Resistive or ohmic or "Triode" Region: $0 < V_{DS} < V_{GS} - V_T$

MOSFET as a Controlled Resistor (cont'd)

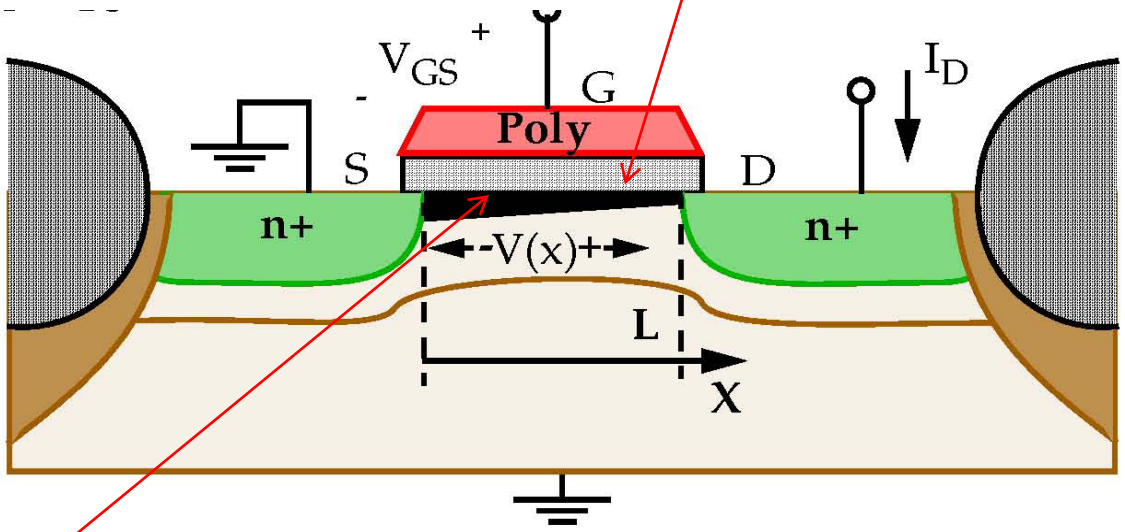
$Q_{inv} = q\ n\ t$, where t is the thickness of the inversion layer.

$$I_D = qn\ v\ Wt = qn\ \mu E\ Wt = -\mu C_{ox} W\ [(V_{GS} - V_T) - V(x)]\ E$$

where, $E = -dV(x)/dx$.

$$\int I_D dx = \int \mu C_{ox} W\ [(V_{GS} - V_T) - V(x)]\ dV(x)$$

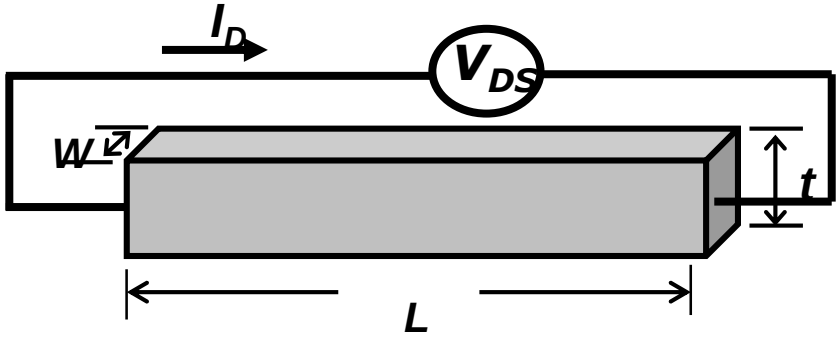
Since I_D is independent of the distance down the channel it can be taken outside the integral sign and the integral is then equal to the channel length. The limits between x=0 and x=L correspond to V(x)=0 and V(x)=V_D.



Inversion charge density $Q_{inv}(x) = -C_{ox}[V_{GS} - V_T - V(x)]$
where $C_{ox} \equiv \varepsilon_{ox} / t_{ox}$

MOSFET as a Controlled Resistor (cont'd)

Let's deduce I_D from R_{DS}



$$I_D = \frac{V_{DS}}{R_{DS}} \quad \& \quad R_{DS} = R_s(L/W) = \frac{L/W}{\mu_n Q_{inv}} = \frac{L/W}{\mu_n C_{ox}(V_{GS} - V_T - \frac{V_{DS}}{2})}$$

average value of $V(x)$

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

We can make R_{DS} low by

- applying a large "gate drive" ($V_{GS} - V_T$)
- making W large and/or L small

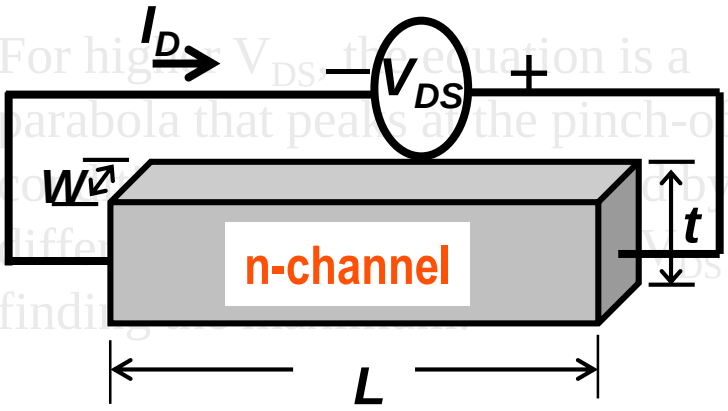
$$R_s = \frac{1}{\mu_n Q_{inv}}$$

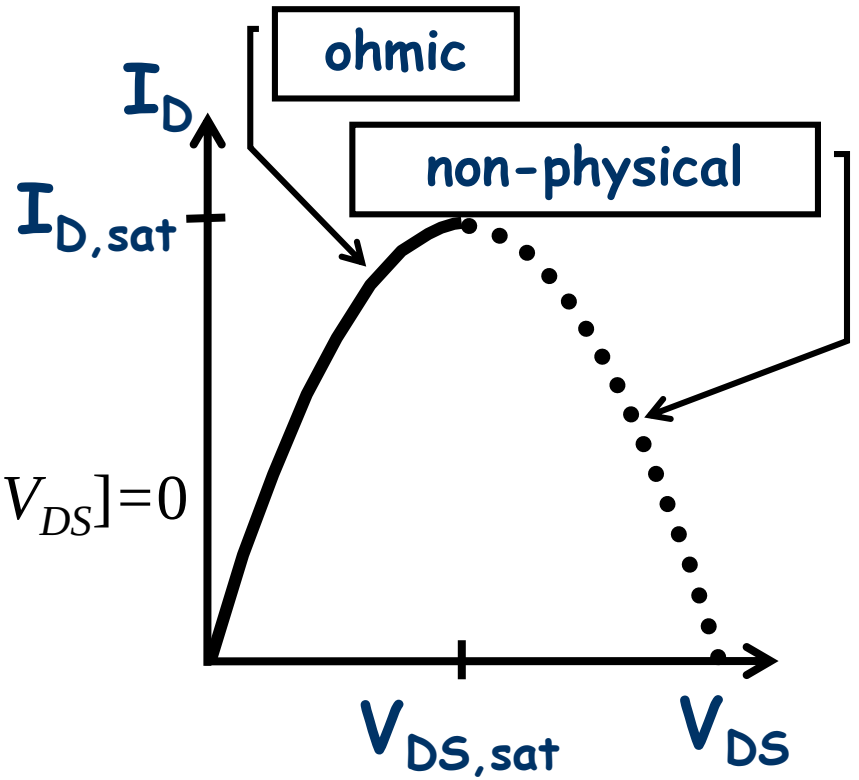
MOSFET as a Controlled Resistor (cont'd)

$$I_D = \mu C_{ox}(W/L)[(V_{GS} - V_T)V_{DS} - V_{DS}^2/2] \qquad (1)$$

for a given V_{GS}

it applies only for the condition $V_{DS} < (V_{GS} - V_T)$ (This is called the 'below pinch-off condition.')





$$dI_D/dV_{DS} = \mu C_{ox}(W/L)[(V_{GS} - V_T) - V_{DS}] = 0$$

so that $V_{DS,sat} = V_{GS} - V_T$

$$I_{D,sat} = \frac{\mu_n C_{Ox}}{2}\frac{W}{L}(V_{GS} - V_T)^2 \qquad (2)$$

MOSFET as a Controlled Resistor (cont'd)

$$I_D = \frac{\mu_n C_{Ox}}{2}\frac{W}{L}\left(2(V_G - V_T)V_{DS} - V_{DS}^2\right) \qquad (1)$$

for a given V_{GS}

Note for $V_{DS} \geq V_G - V_T = V_{DS,sat}$, result is non-physical.

At $V_{DS,sat}$, $n_{s,x=L} = 0$

assume that channel curr.

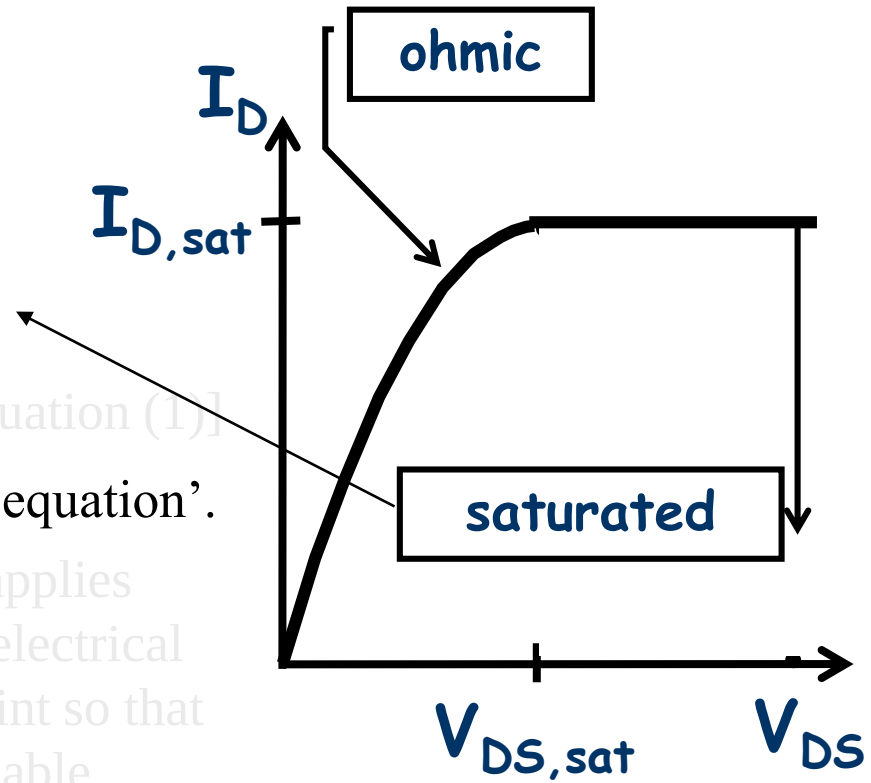
is const for $V_{DS} \geq V_{DS,sat}$

$$I_{D,sat} = \frac{\mu_n C_{Ox}}{2}\frac{W}{L}(V_{GS} - V_T)^2 \qquad (2)$$

[after substituted $V_{DS,sat}$ back into equation (1)]

which is called the 'above pinch-off equation'.

Strictly speaking this equation only applies for V_{GS}-V_T=V_{DS}, but practically the electrical characteristic saturates above this point so that the equation is more generally applicable.



Depletion Region Width W_{dep}

PL

At V_D=0

$$W_{dep} = x_{po} + x_{no} = \sqrt{\frac{2\varepsilon_{Si}}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)\phi_B}$$

At V_D<0

$$W_{dep} = x_p + x_n = \sqrt{\frac{2\varepsilon_{Si}}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)(\phi_B - V_D)}$$

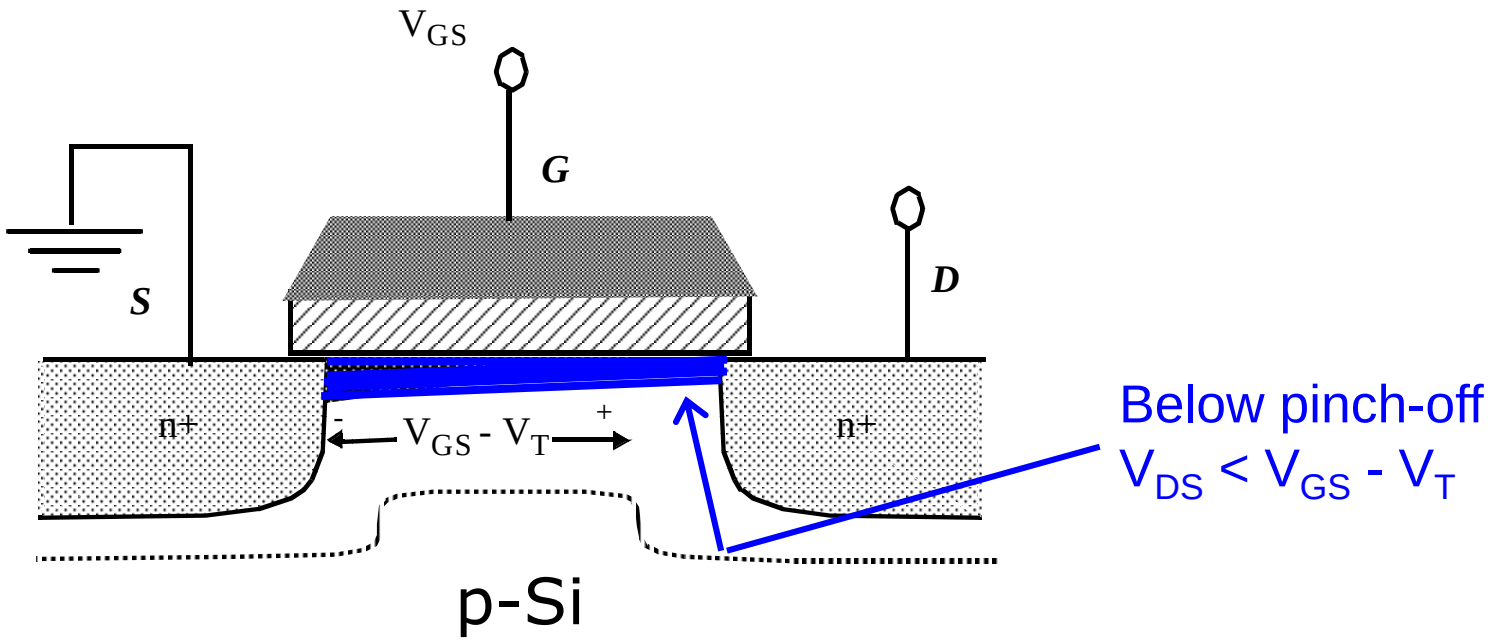
- The width of the depletion region is a function of the bias voltage, and is dependent on $\boldsymbol{N_A}$ and $\boldsymbol{N_D}$.
- If one side is much more heavily doped than the other (which is commonly the case), then this can be simplified:

$$W_{dep} \cong \sqrt{\frac{2\varepsilon_{Si}}{qN}(\phi_B - V_D)}$$

where *N* is the doping concentration on the more lightly doped side.

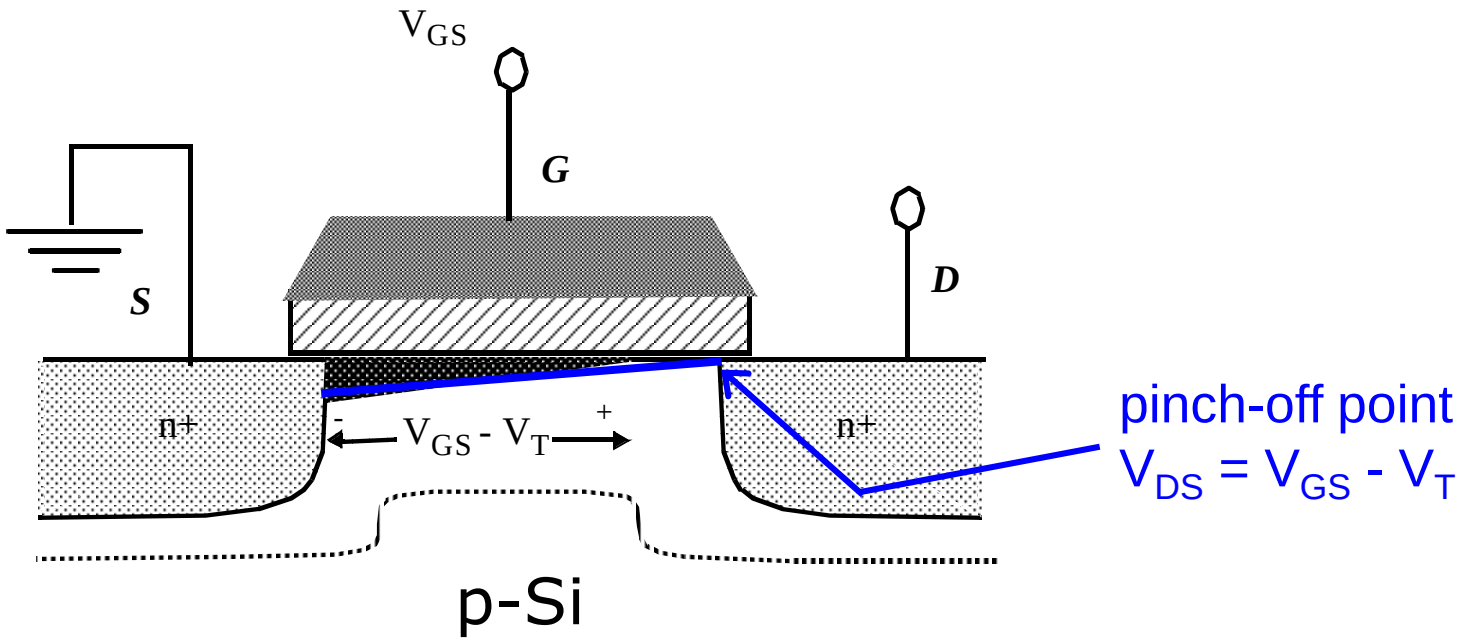
What is “Pinch off” ? 夹断

- For a given V_{GS}, you have a maximum amount of current you can flow through the channel (regardless of V_{DS}).
- When $V_{DS} = V_{GS} - V_T$ the channel is flowing as much current as it can so we call it pinched off (even though current continues to flow).



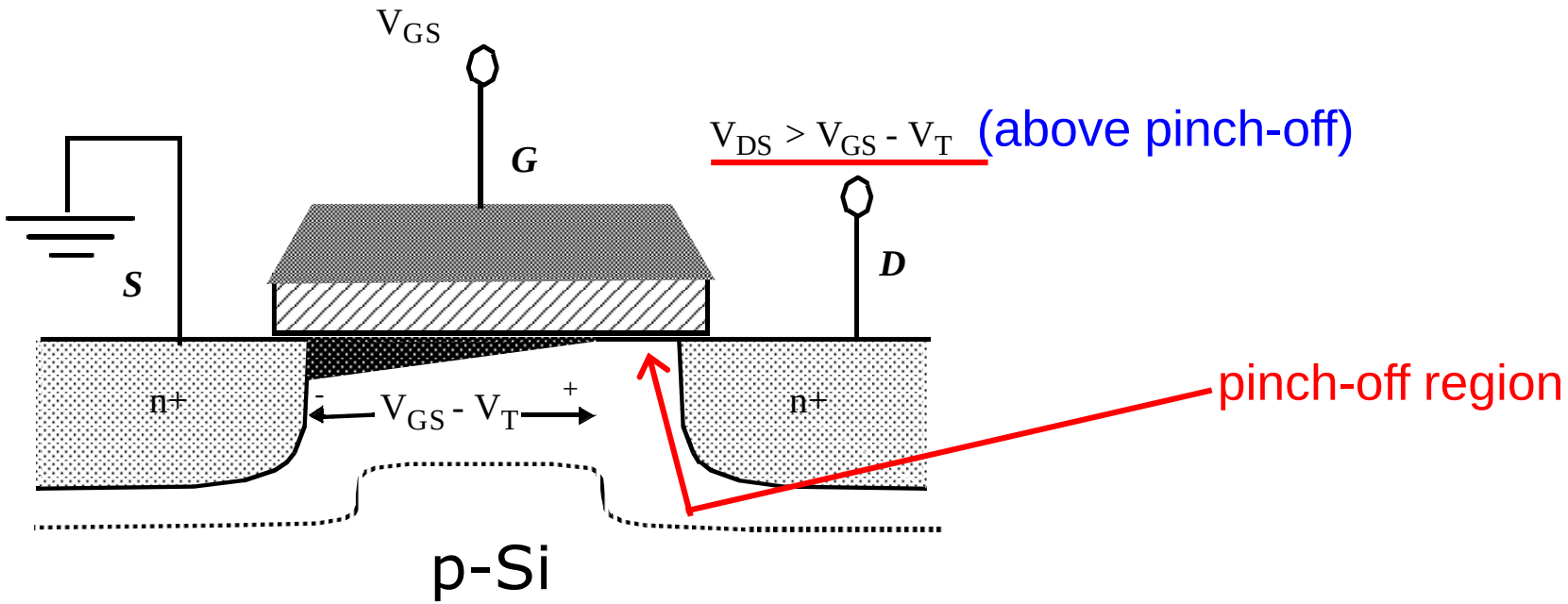
What is “Pinch off” ? 夹断

- For a given V_{GS}, you have a maximum amount of current you can flow through the channel (regardless of V_{DS}).
- When $V_{DS} = V_{GS} - V_T$ the channel is flowing as much current as it can so we call it pinched off (even though current continues to flow).



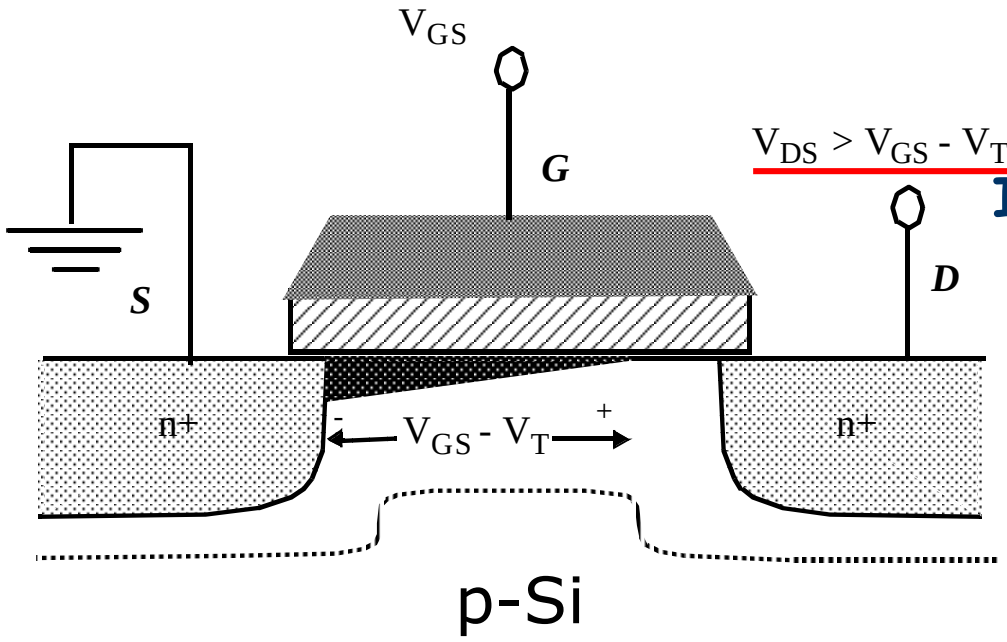
What is “Pinch off” ? 夹断

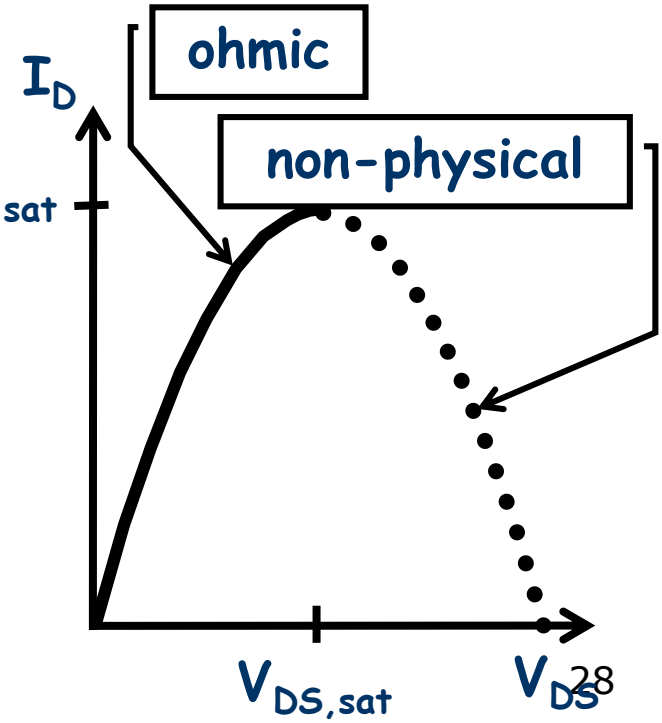
- For a given V_{GS}, you have a maximum amount of current you can flow through the channel (regardless of V_{DS}).
- When $V_{DS} = V_{GS} - V_T$ the channel is flowing as much current as it can so we call it pinched off (even though current continues to flow).



What is “Pinch off” ?

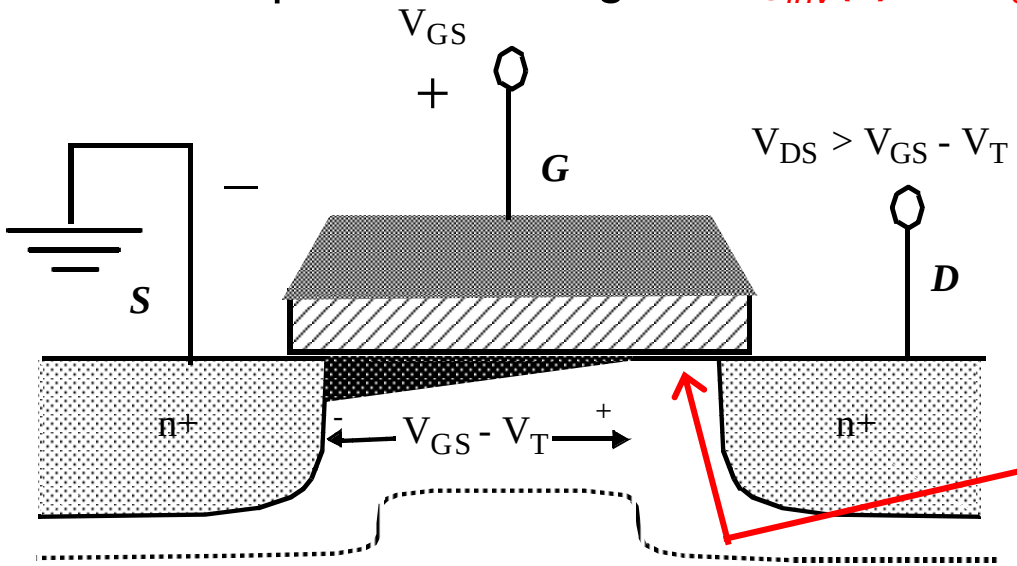
- For a given V_{GS}, you have a maximum amount of current you can flow through the channel (regardless of V_{DS}).
- When $V_{DS} = V_{GS} - V_T$ the channel is flowing as much current as it can so we call it pinched off (even though current continues to flow).





Why “Pinch off” ?

- As V_{DS} increases, the inversion-layer charge density at the drain end of the channel is reduced; therefore, I_D does not increase linearly with V_{DS}.
- When V_{DS} reaches $V_{GS} - V_T$, the channel is “pinched off” at the drain end, and I_D saturates (*i.e.* it does not increase with further increases in V_{DS}).
- In the pinched-off region: $Q_{inv}(x) = -C_{ox}[V_{GS} - V_T - V_{DS,sat}] = 0$



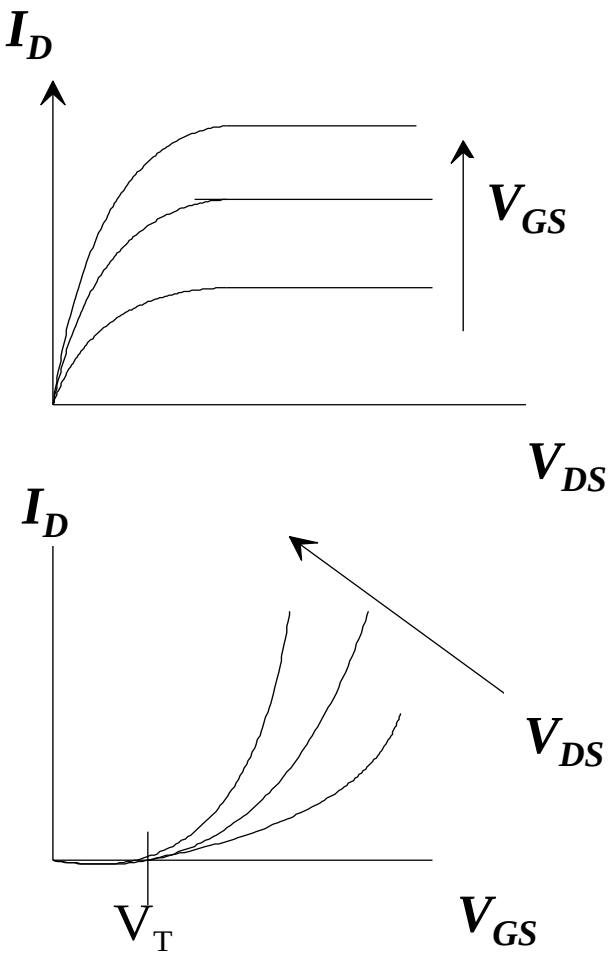
$$I_{D,sat} = \mu_n C_{ox} \frac{W}{2L}(V_{GS} - V_T)^2$$

I_D vs. V_{DS} or V_{GS} Characteristics

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

For $V_{DS} \geq V_{DS,sat}$

$$I_{D,sat} = \frac{\mu_n C_{Ox}}{2}\frac{W}{L}\left(V_{GS}-V_T\right)^2 \qquad (2)$$



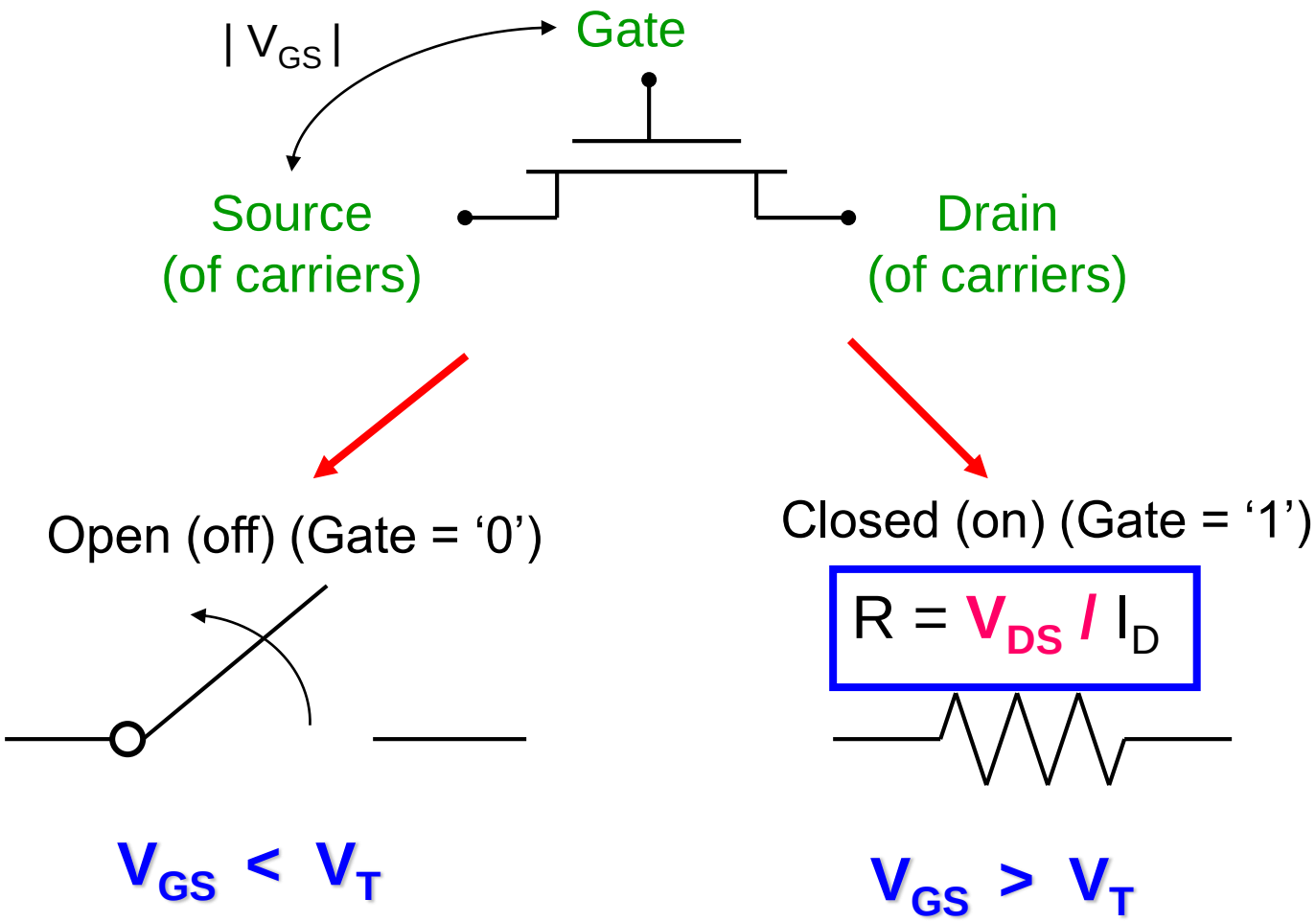
The graphs shows ideal charactersitics. The top graph is the **output characteristic,** the lower one is the **transfer characteristic**. Equations (1) and (2) are the simple form of the design equations.

The quantity μC_{ox} (W/L) = β is the device constant. The designer can only vary W/L so as to change β. Other values are fixed during the process development.

OUTLINE

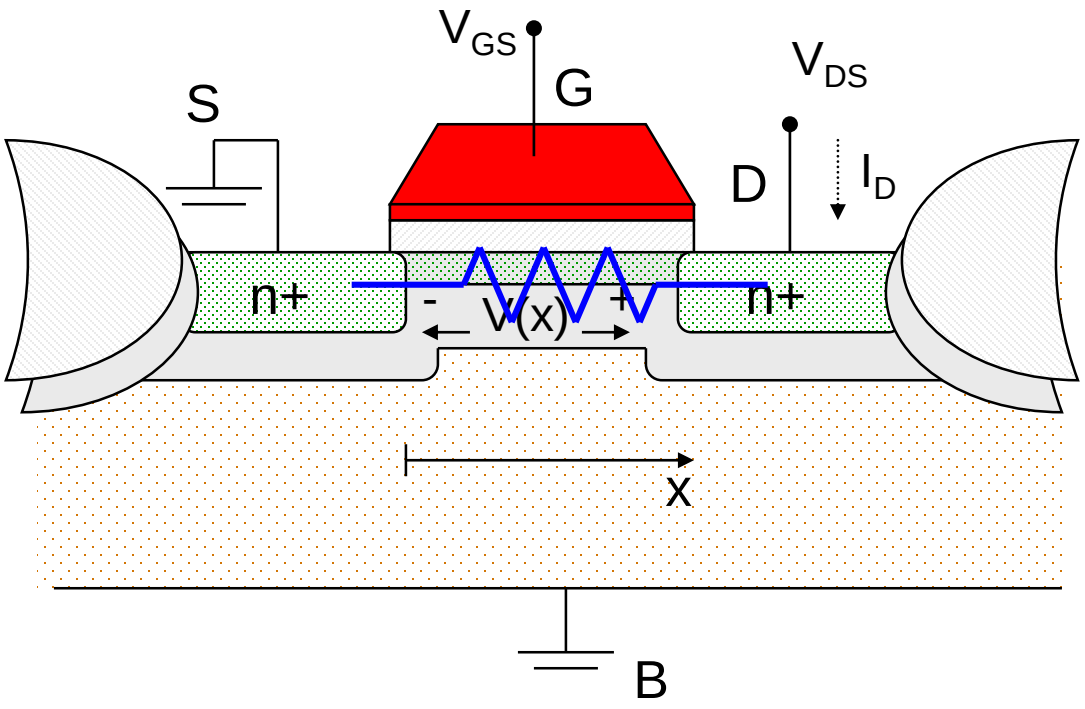
- I - V Characteristics
  - Cutoff Region
  - Linear Region
  - Saturation Region (pinch-off region)
- **Switch model of nMOSFETs**

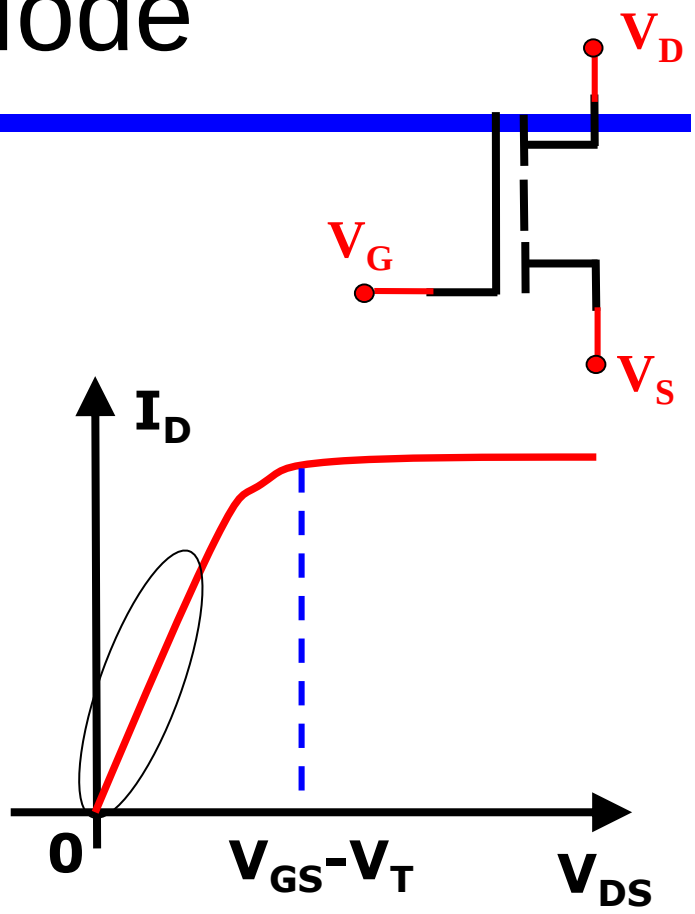
Switch Model of nMOS Transistor



Transistor in Linear Mode

Assuming $V_{GS} > V_T$





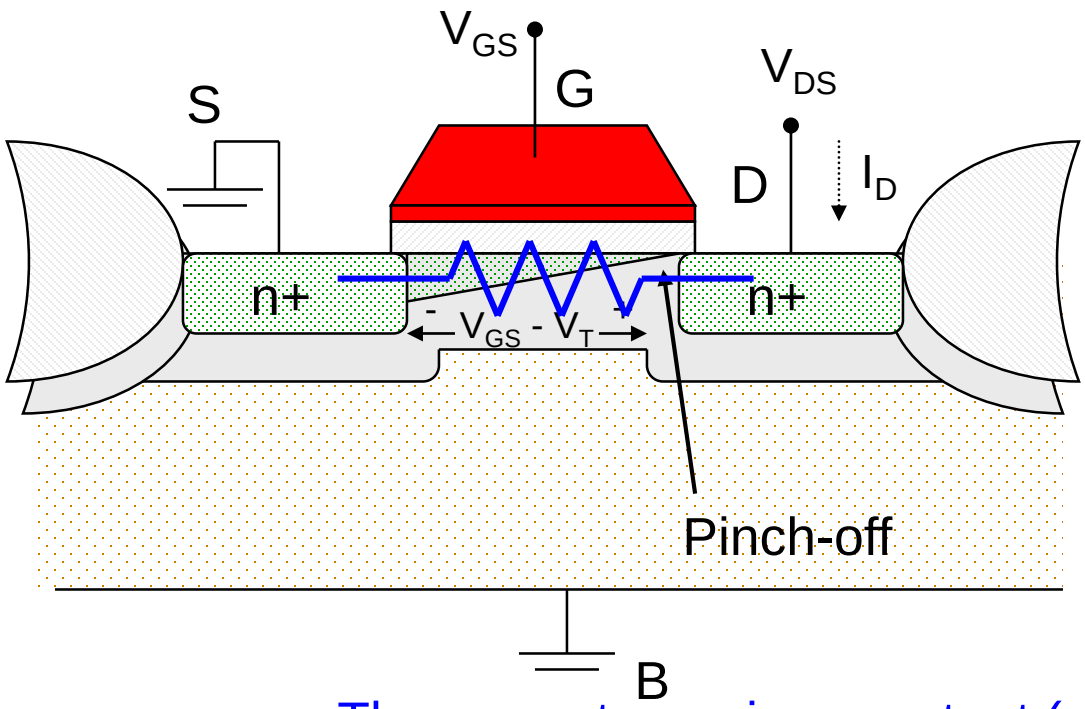
When $V_{DS} \leq V_{GS} - V_T$: $I_D = \beta_0 W/L [(V_{GS} - V_T)V_{DS} - V_{DS}^2/2]$

$\beta_0 = \mu_n C_{ox}$

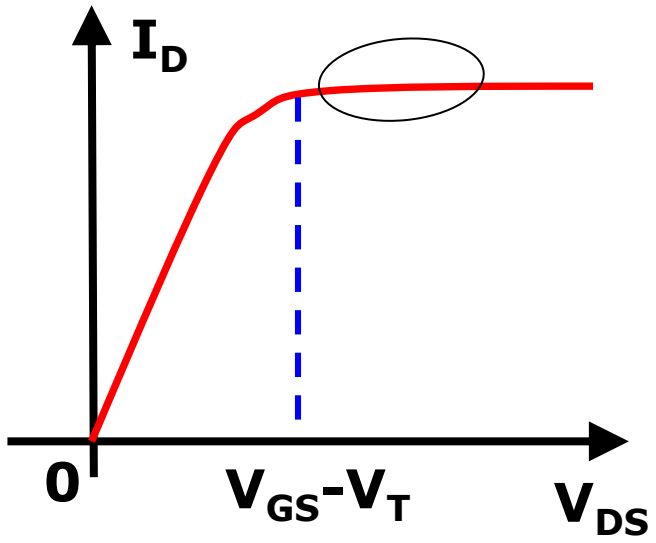
$R = V_{DS} / I_D$

Transistor in Saturation Mode

Assuming $V_{GS} > V_T$



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



The current remains constant (saturates).

When $V_{DS} \geq V_{GS} - V_T$: $I_D = (\beta_0/2)\, W/L\, [(V_{GS} - V_T)^2]$

$R = V_{DS} / I_D$

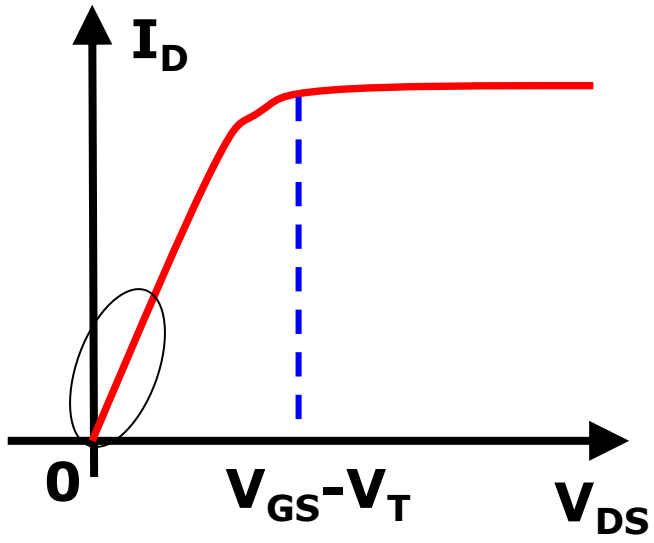
nMOSIC - **The MOST as a linear R_{DS}**

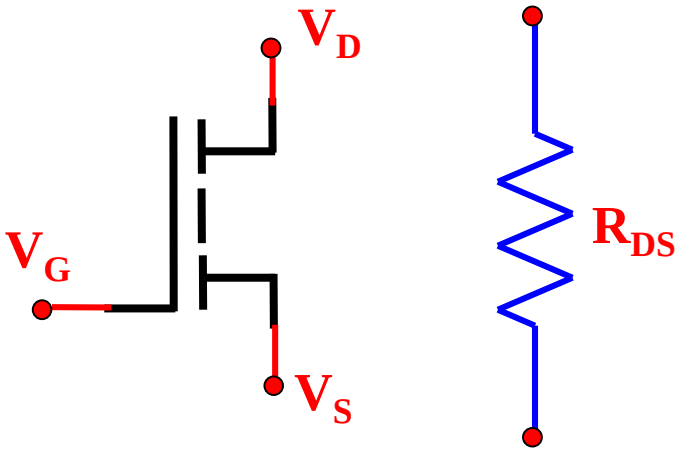
When $V_{DS} \leq V_{GS} - V_T$

$I_D = \beta_0 W/L [(V_{GS} - V_T)V_{DS} - V_{DS}^2/2]$

For **small V_{DS}**, there is a linear dependence between V_{DS} and I_D, hence

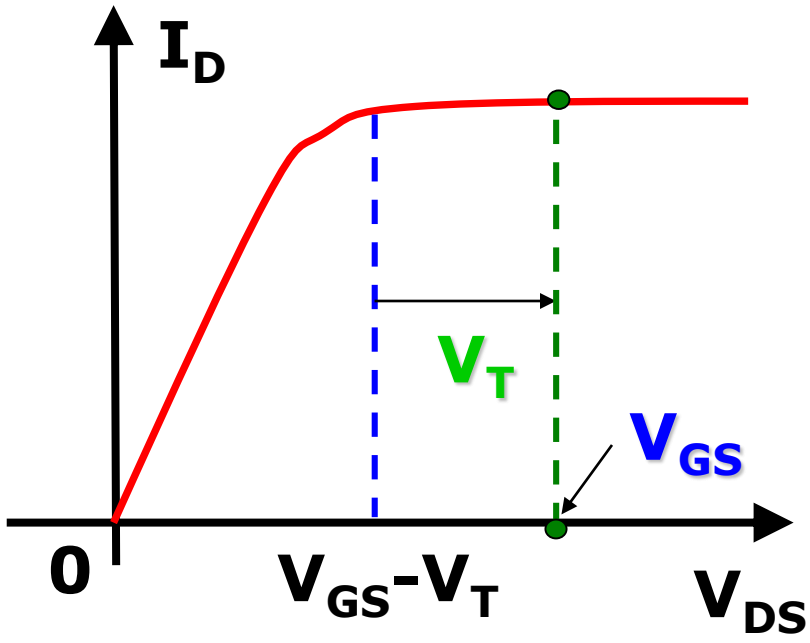
$$1/R_{DS} = I_D/V_{DS}$$
$$= \beta_0 W/L [(V_{GS} - V_T) - V_{DS}/2]$$
$$\approx \beta_0 W/L (V_{GS} - V_T)$$

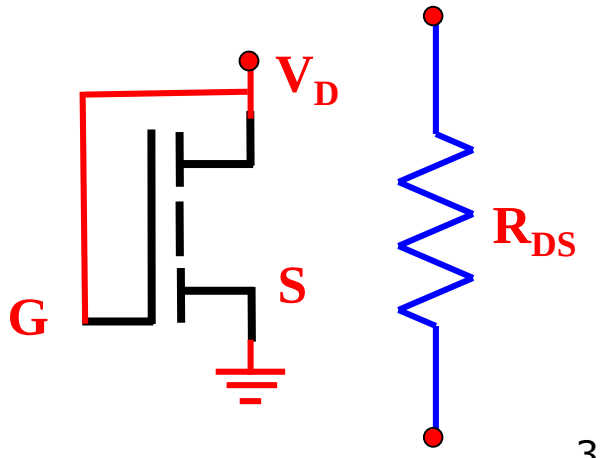




nMOSIC - **The MOST as a Load**

- The use of a load resistor (an implanted layer) leads to rather large structures.
- The need to minimise the area of silicon involved is paramount (see 1st Lecture).
- One method is to use a MOST (frequently called a MOSFET) as a load.
- In this case the gate and the drain are connected together so that $V_{GS}=V_{DS}$. The pinch off point coincides with $V_{GS}-V_T=V_{DS}$.
- The characteristic of the load is shown in the Fig. It extends along the drain axis by an amount V_T.





nMOSIC - **The MOST as a Load**

- **Problem**

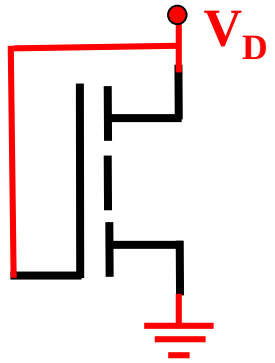
  宽长比

  Calculate the resistance of a load MOST with an **aspect ratio of 1** when the mobility of the electrons is **$1000 cm^2 V^{-1} sec^{-1}$** and the gate capacitance **per unit area** is **$10^{-2} Fm^{-2}$**. The drain voltage is **V_D=5V** and the threshold voltage **V_T= 0.5V**.

- **Solution**

  The drain current is

$$I_D = \mu_n (W/L) C_{ox} (V_G - V_T)^2/2.$$



but V_G=V_D so that

$$R=V_D/I_D=\ldots$$

- **R=100Ω**

nMOSIC - **The MOST as a Load**

- **Problem**

  Calculate the resistance of a load MOST with an aspect ratio of 1 when the mobility of the electrons is $1000 cm^2 V^{-1} sec^{-1}$ and the gate capacitance per unit area is $10^{-2} Fm^{-2}$. The drain voltage is V_D=5V and the threshold voltage V_T= 0.5V.

- **Solution**

  The drain current is the same as at the pinch-off point where

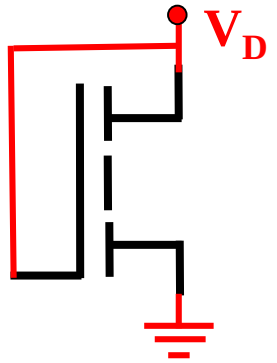
  $$I_D = \mu(W/2L)\, C_{ox}\, (V_G - V_T)^2.$$

  but $V_G = V_D$ so that

  $$I_D = \mu(W/2L)\, C_{ox}\, (V_D - V_T)^2$$

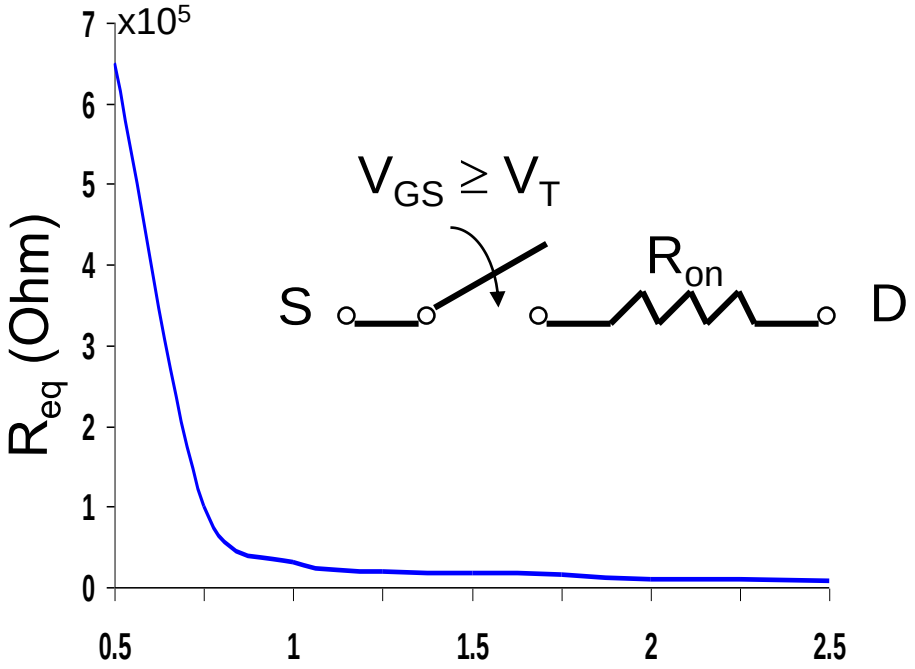
  $$I_D/V_D = 1/R = \mu\, (W/2L)\, C_{ox}\, (V_D - V_T)^2/V_D$$

  $$= 0.1*0.5*10^{-2}*4.5^2 = 1.01*10^{-2},$$



- **R=100Ω**

The Transistor Modeled as a Switch



(for $V_{GS} = V_{DD}$, $V_{DS} = V_{DD} \rightarrow V_{DD}/2$)

Modeled as a switch with infinite off resistance and a finite on resistance, R_{on}

Resistance inversely proportional to W/L (doubling W halves R_{on})

For $V_{DD} >> V_T + V_{DSAT}/2$, R_{on} independent of V_{DD}

Once V_{DD} approaches V_T, R_{on} increases dramatically

V_{DD}(V)	1	1.5	2	2.5
NMOS(kΩ)		19	15	13
PMOS (kΩ)		55	38	31

R_{on} (for W/L = 1)
For larger devices divide R_{eq} by W/L