

# VLSI External

→ Introduction :-

- Kinds of Computation :-

- 1) Arithmetic : + , - , × , ÷
  - 2) Logical : AND, OR, NAND, NOR, NOT, XOR, XNOR
  - 3) Special Purpose : ex: Convolution in Signal Processing.
- The above computations are considered as one computation / instruction given to the microprocessors.

- Computational Efficiency :-

These are 2 ways to measure

efficiency.

- 1) How fast the computation is ? ex: Server applications
- 2) How much energy is used ? ex: Mobile applications

- The Design of Today's Processors :-

- Example : IBM P9 Chip

Technology - 14 nm FinFET Silicon On Insulator (SOI)

FinFETs started at 1nm to control leakage

current / off current. Before that, MOSFETs

were used.

Area -  $695 \text{ mm}^2$  About the size of a 10₹ coin

No. of Transistors  $\approx$  8 billion Each of these is leaking  
some current, which becomes a huge  
problem

No. of Metal levels - 17 layers of metal above the  
substrate

M1 - M3 pitch (did blur the M levels) - 64 nm

M4 - M5 pitch - 80 nm

M6 - M9 pitch - 128 nm

M10 - M11 pitch - 256 nm

M12 - M15 pitch - 360 nm

M16 - M17 (power & clock) - 2400 nm (widest metal and lowest  
IR drop)

Quadrants - 6

cores - 4 each in a Quadrant

Each quadrant has a shared L2 and L3 cache

Each core has a own L1 cache

L1, L2 - SRAM

} Faster Memory Access than normal RAM

L3 - eDRAM

- The total length of wiring in the above chip is in the order of kilometer.

- The Design Flow :-

- It is heavily automated. Only the critical blocks are automated. Verilog is used in the automation part.

→ P-N Junction :-

- Semiconductor Used : Silicon. Good Mobility and Silicon dioxide is a good insulator

- At room Temp,  $n_i = p_i \approx 10^{10} / \text{cm}^3 \rightarrow$  Not enough to conduct freely.

- Doping using phosphorous (n) or boron (p) increases the electron density at room temp , making it more conductive.
- During conduction, the dopant atom form immovable ions of the opposite charge.
- The electron density is <sup>almost</sup> equal to the n-type dopant concentration, meaning that the contribution of the silicon atoms is negligible
- Law of Mass Action :  $np = n_i^2$   
If  $n = N_D$  ,  $p = \frac{n_i^2}{N_D}$
- Maxwell - Boltzmann Equation :



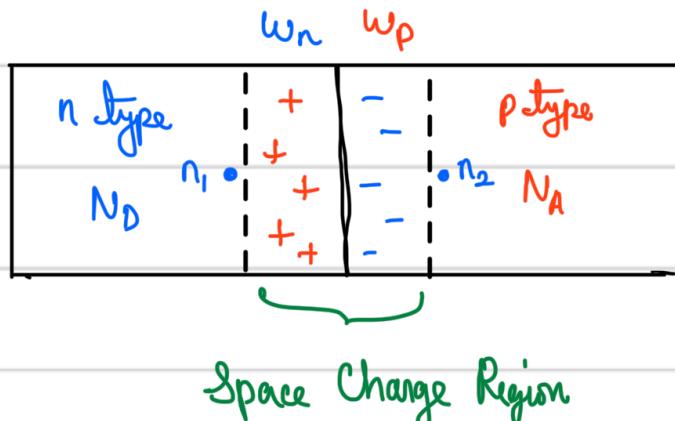
If  $n_1 \neq n_2$ ,

$$V_{12} = \frac{kT}{q} \ln\left(\frac{n_1}{n_2}\right) = \underline{\underline{\Psi_{12}}}$$

$$\frac{kT}{q} @ \text{ Room Temp} = 25 \text{ mV}$$

$$\Rightarrow \frac{n_1}{n_2} = e^{\frac{qV_{bi}}{kT}}$$

- Using this equation in a PN junction



$$q N_D w_n A = q N_A w_p A \quad (\text{To maintain charge neutrality})$$

$$\Rightarrow w_n N_D = w_p N_A \Rightarrow \begin{array}{l} \text{if } N_D \gg N_A \\ \text{then } w_p \gg w_n \end{array}$$

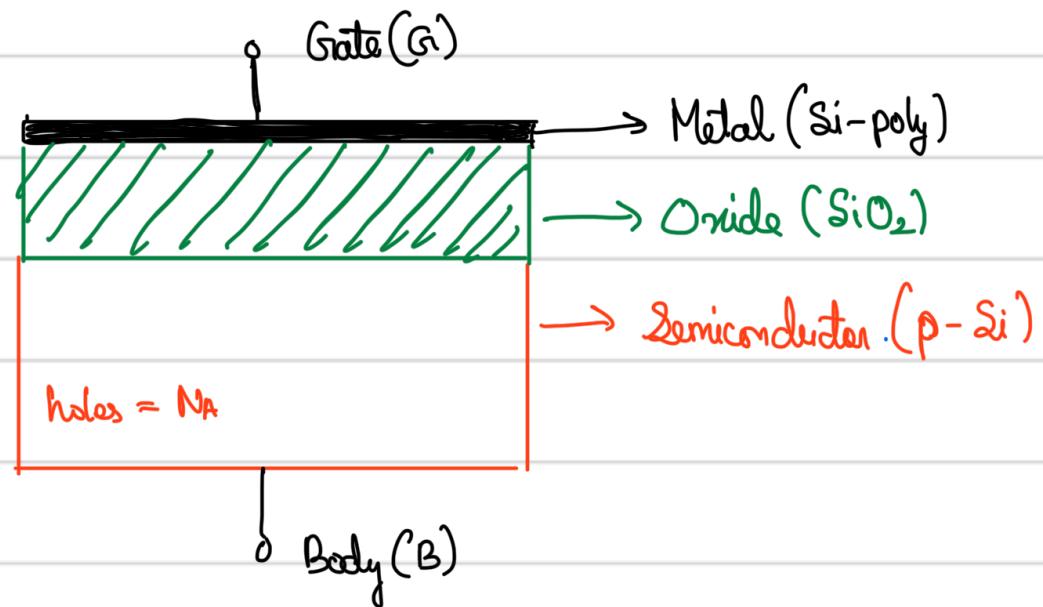
Across the depletion region, due to difference in dopant concentration there will be a voltage difference.

$$n_i = N_D, n_i^2 = \frac{N_D N_A}{N_D + N_A}$$

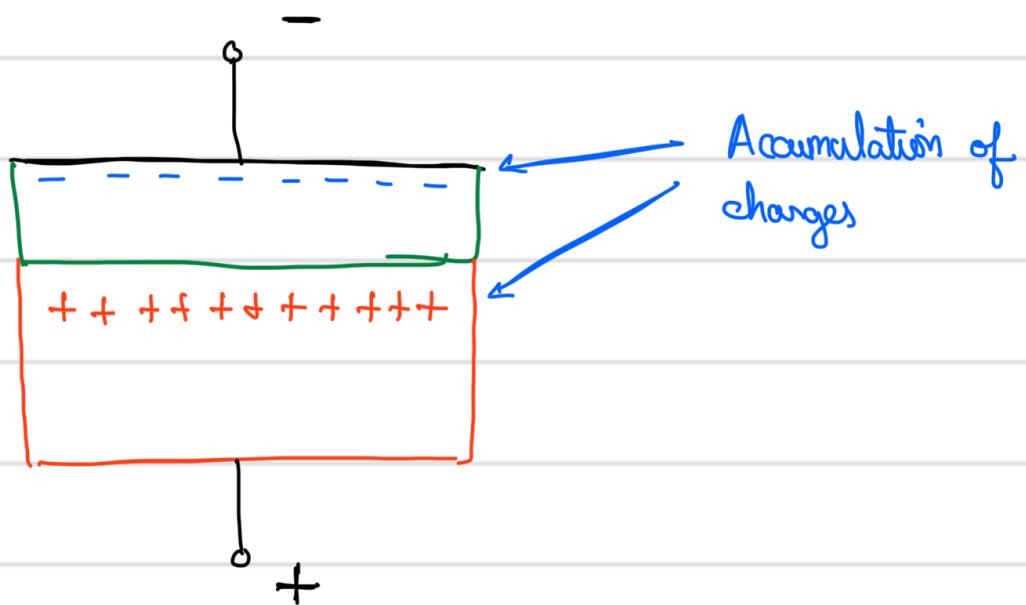
$$\Rightarrow V_{bi} = \frac{kT}{q} \ln \left( \frac{N_A N_D}{n_i^2} \right)$$

→ MOS Capacitor :-

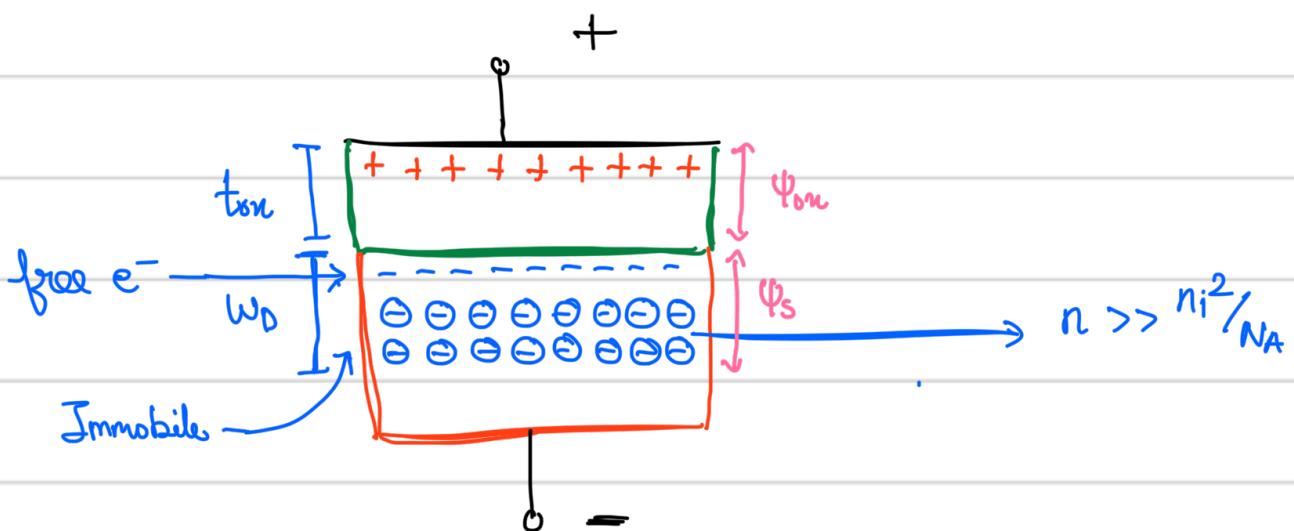
- MOS - Metal Oxide Semiconductor



Case I: If  $V_{GB} < 0$  Accumulation Region of Operation



Case 2:  $V_{AB} > 0$



Now, as per Maxwell Boltzmann law, there will be some voltage difference between the upper and the bottom layer of the Body, due to the attraction of electrons.

Because of this,

$$\psi_s = \frac{kT}{q} \ln \left( \frac{n_s}{n_b} \right)$$

$\psi_s$  will get "pinned" (almost constant) when  $n_s = N_A$ . This is called inversion.

The value of  $V_{AB}$  when  $n_s = N_A$  is called the threshold voltage.

$$\psi_s = \frac{kT}{q} \ln \left( \frac{N_A}{n_i^2 / N_A} \right) =$$

$$\Rightarrow \Psi_s = \frac{2kT}{q} \ln \left( \frac{N_A}{n_i} \right) \rightarrow \text{Surface Potential at inversion}$$

$$V_{TH} = \Psi_{ox} + \Psi_s \quad (\text{KVL})$$

$$\Psi_{ox} = - \left( \frac{Q_D' + Q_I'}{C_{ox}} \right) \quad (Q_D, Q_I - \text{Depletion and Inversion charge per unit area})$$

$$C_{ox}' = \frac{\epsilon_r \epsilon_0 WL}{t_{ox}} \quad C_{ox} = \frac{\epsilon_r \epsilon_0}{t_{ox}} \quad (\text{Capacitance per unit area})$$

$$Q_D' = q \cdot N_A \cdot WL \cdot w_D, \quad w_D = \sqrt{\frac{2\epsilon_{si} |\Psi_s|}{q N_A}} \quad (\text{By Poisson's Eqn})$$

$$Q_D' = \left( \sqrt{2\epsilon_{si} |\Psi_s| q N_A} \right) \cdot WL \Rightarrow Q_D = \sqrt{2\epsilon_{si} |\Psi_s| q N_A}$$

$$Q_I' = C_{ox} (V_{GB} - V_{TH})$$

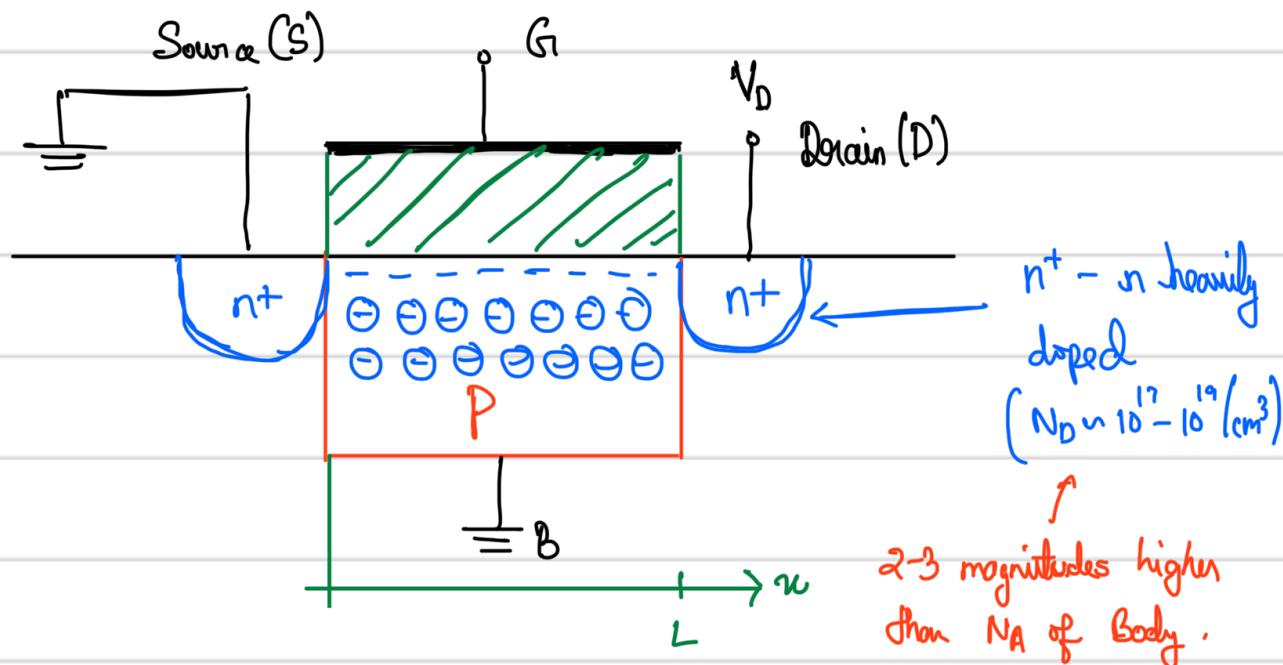
$$\Rightarrow V_{TH} = \Psi_s - \frac{1}{C_{ox}} \sqrt{2\epsilon_{si} |\Psi_s| q N_A}$$

To conclude, if  $V_{GB} > V_{TH}$ , the surface layer of the semiconductor becomes n-doped.

Still the overall behavior of the device does not change, since no current can flow through the gate oxide.

- To make use of the  $n$  type surface layer, we introduce 2 more terminals into the device.

→ MOS Transistors :-



To obtain an expression for current across the transistor across S and D,

$$V(0) = 0, V(L) = V_D$$

The amount of inversion charge in a small length element  $dx$  is,

$$Q_I = -C_{ox} (V_G - V_T - V(x))$$

$$dQ_I = Q_I W dx$$

$$\Rightarrow dQ_I' = -C_{ox} (V_a - V_T - V(x)) \cdot W dx$$

$$\Rightarrow \frac{dQ_I'}{dt} = -C_{ox} (V_a - V_T - V(x)) W \frac{dx}{dt}$$

$$\Rightarrow I_D = -C_{ox} (V_a - V_T - V(x)) W \cdot v_d(x) \quad \xrightarrow{\text{drift velocity at } x}$$

$$v_d(x) = -\mu_n \frac{dV(x)}{dx}$$

$$\Rightarrow I_D = \mu_n C_{ox} W (V_a - V_T - V(x)) \frac{dV(x)}{dx}$$

$$\Rightarrow \int_0^L I_D dx = \int_0^L \mu_n C_{ox} W (V_a - V_T - V(x)) dV(x)$$

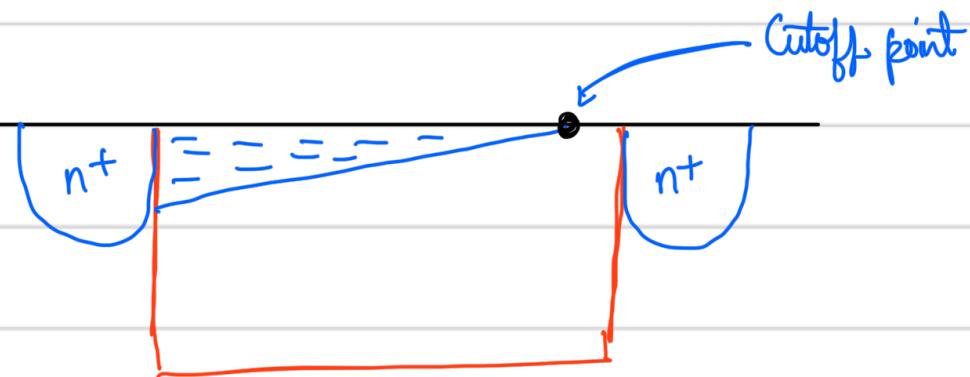
$$\Rightarrow I_D L = \mu_n C_{ox} W \left( V_a - V_T - \frac{V_D^2}{2} \right)$$

$$\Rightarrow I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \left( 2(V_a - V_T) V_D - V_D^2 \right)$$

• However if we keep increasing  $V_{DS}$ ,

The potential near the drain is  $V_a - V_T - V_D$ . If  $V_D > V_a - V_T$ , the inversion layer disappears since the potential goes below threshold.

Therefore at  $V_D = V_A - V_T$  (or  $V_{DS} = V_{AS} - V_T$ ) the channel disappears, and the current becomes saturated.



Beyond the cutoff point, the current is due to injection. So the current does not change because of that region.

$$I_D = \frac{1}{2} \mu n C_{ox} \frac{w}{L} \left( 2(V_{AS} - V_T)(V_{DS} - V_T) - (V_{DS} - V_T)^2 \right)$$

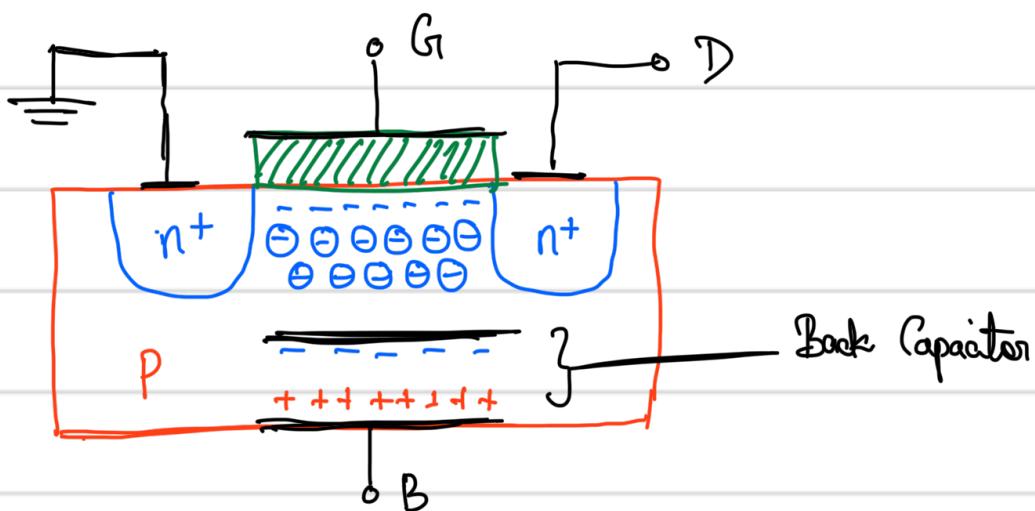
$$I_D = \frac{1}{2} \mu n C_{ox} \frac{w}{L} (V_{AS} - V_T)^2 \rightarrow \text{No dependence of } V_{DS}$$

$$\Rightarrow I_D = \begin{cases} \frac{1}{2} \mu n C_{ox} (2V_{DS}(V_{AS} - V_T) - V_{DS}^2), & V_{DS} \leq V_{AS} - V_T \\ \frac{1}{2} \mu n C_{ox} (V_{AS} - V_T)^2, & V_{DS} > V_{AS} - V_T \end{cases}$$

→ Body Effect :-

$$V_T = \Psi_S + \frac{1}{C_{ox}} \sqrt{2 \sum \epsilon_{Si} q N_A |\Psi_S|}$$

- Earlier it was assumed that the body has the same potential as source (both are grounded), ie,  $V_{BS} = 0$ .
- If  $V_{BS} > 0$ ,



The body terminal will act like a back-capacitor, increasing the negative potential at the surface. Therefore  $V_T$  decreases.

- The empirical equation for  $V_T$  is as follows,

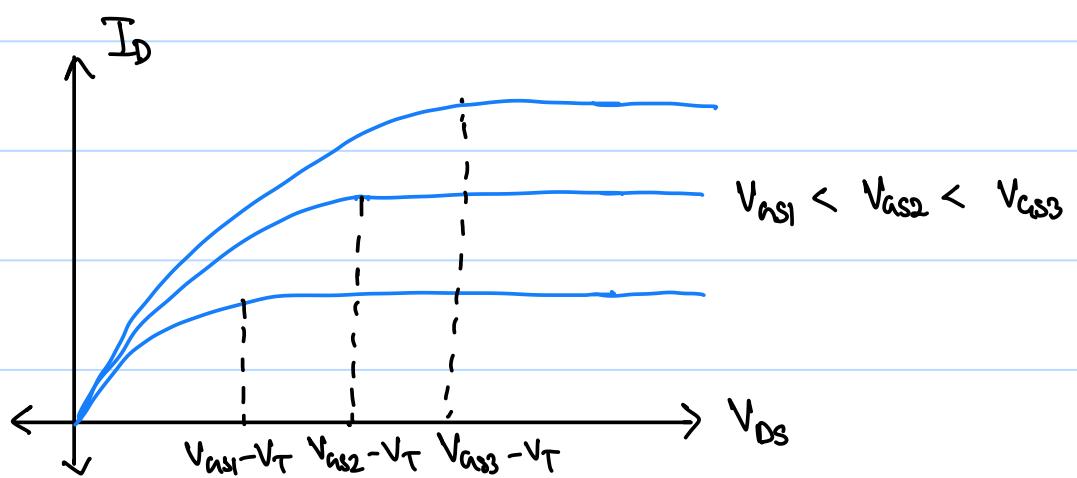
$$V_T = V_{To} + \gamma \left( \sqrt{|\Psi_S + V_{SB}|} - \sqrt{|\Psi_S|} \right)$$

where,  $V_{T0}$  = Threshold voltage at  $V_{SB} = 0$  } Compact modelling  
 $\gamma \rightarrow$  Empirical parameter / Body Effect Cof.

$\gamma$  depends on the other parameters of a transistor.

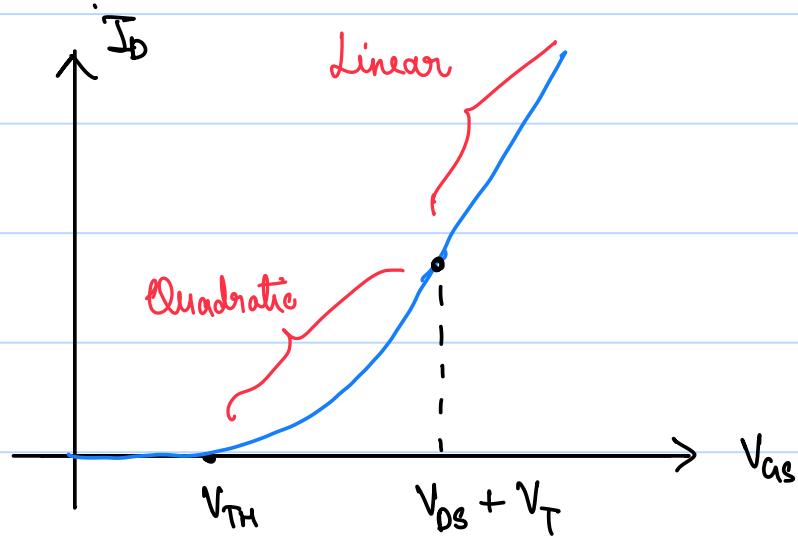
→ IV Characteristics :-

i)  $I_D$  vs  $V_{DS}$  :

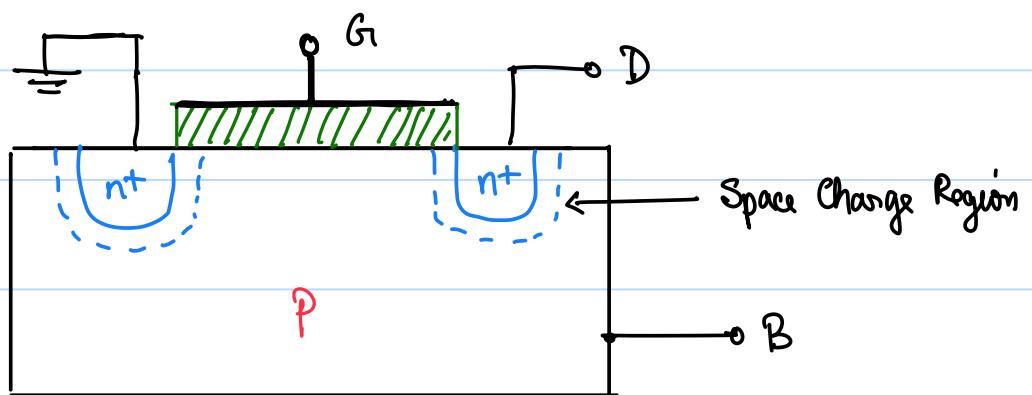


$$I_D = \begin{cases} \mu n C_{ox} \frac{W}{L} V_{DS} \left[ (V_{GS} - V_{TH}) - \frac{V_{DS}}{2} \right], & V_{DS} \leq V_{GS} - V_{TH} \\ \frac{1}{2} \mu n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2, & V_{DS} > V_{GS} - V_{TH} \end{cases}$$

2)  $I_D$  vs  $V_{DS}$ :

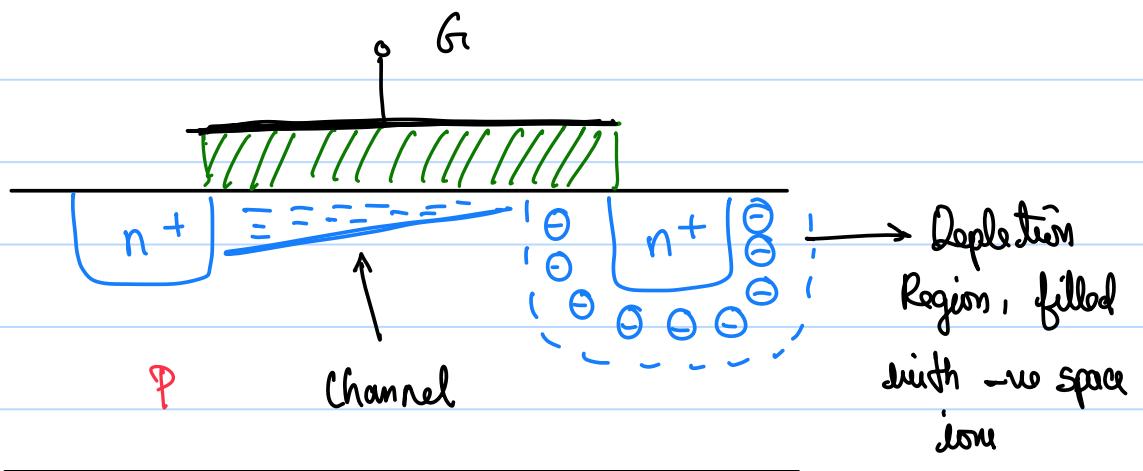


→ Short Channel Transistor :-



- Since the junction between the source (drain) and the body is a PN junction, there will be a space charge region.
- If the channel length is long, this region does not matter, due to its relative short length.

- If the channel length is small, this depletion region will affect the length of the channel.
- The depletion region will extend much further into the p substrate than the source/drain, due to the difference in doping concentration.
- If  $V_D$  is very high, since a large fraction of the channel is depleted,  $V_{TH}$  comes down since the region of inversion is reduced. So, the channel length, and  $V_{TH}$ , are functions of  $V_{DS}$  if drain is small in size.



- In a short channel MOSFET,

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{\omega}{L} (V_{DS} - V_{TH})^2$$

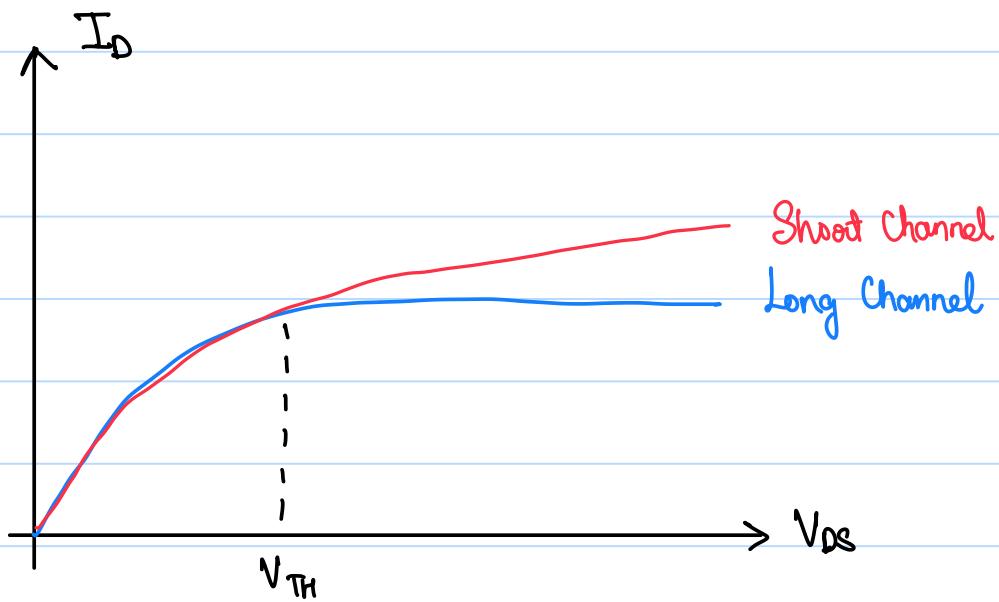
Since in a short channel,  $L \rightarrow L - \Delta L$ ,  $\Delta L = \lambda V_{DS} L$

$$\Rightarrow I_D = \frac{1}{2} \mu_n C_{ox} \frac{\omega}{L - \lambda V_{DS}} (V_{DS} - V_{TH})^2$$

$\lambda$  - Channel Length Modulation Param.

$$\Rightarrow I_D = \frac{1}{2} \mu_n C_{ox} \frac{\omega}{L} (V_{DS} - V_{TH})^2 (1 + \lambda V_{DS})$$

→ Short Channel Equation

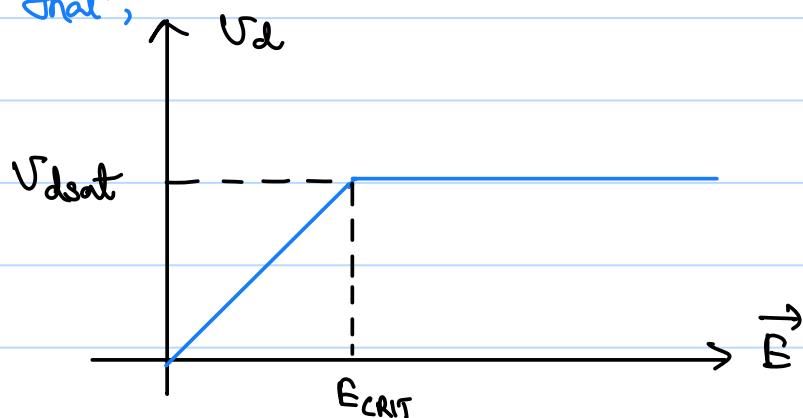


→ Velocity Saturation :-

- The drain current is mainly due to the drift of electrons.
- As transistors become smaller and smaller,  $V_{DD}$  (main drain voltage) becomes smaller and smaller to avoid breakdown.
- By the drift phenomena, we know that,

$$E = \frac{V_d}{\mu}, \quad V_d - \text{drift velocity}$$

Each lattice has a maximum permissible drift velocity, which makes it such that,



The drain voltage at which the drift velocity saturates is given by,

$$\frac{V_{DS}}{L} = \frac{V_{DSAT}}{\mu_n} \quad \mu_n - \text{Mobility of electron}$$

$$\Rightarrow V_{DS} = \frac{V_{DSAT} L}{\mu_n}$$

Because of this, if  $V_{DS}$  is large ( $\text{large } V_{DSAT}$ ), then saturation may be achieved even before  $V_{DS} = V_{DSAT}$ , due to velocity saturation kicking in before channel cutoff.

$$\Rightarrow I_D = \begin{cases} 0 & V_{DS} < V_T \\ \mu_n C_o x \frac{W}{L} V_{DS} [(V_{DS} - V_T) - \frac{V_{DS}}{2}] & V_{DS} \leq V_{DS} - V_T \\ \frac{1}{2} \mu_n C_o x \frac{W}{L} V_{DS} [(V_{DS} - V_T) - \frac{V_{DS}}{2}] & V_{DS} \geq V_{DS} \end{cases}$$

$\& V_{DS} < V_{DS} - V_T$

This can be thought of as the velocity saturation region of operation.

- o Unified Current Model :-

$$I_D = \mu_n C_o x \frac{W}{L} V_{min} \left[ (V_{DS} - V_T) - \frac{V_{min}}{2} \right] (1 + \lambda V_{DS})$$

$$V_{min} = \min (V_{DS} - V_T, V_{DS}, V_{DS})$$

$\downarrow \quad \downarrow \quad \downarrow$   
Saturation threshold velocity saturation

is not a contact though

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

Body Effect

- Device Parameters:  $(K_n^l, V_{DSAT}, \lambda, V_{TH0}, \gamma) \leftarrow$  Level-1 SPICE Model

