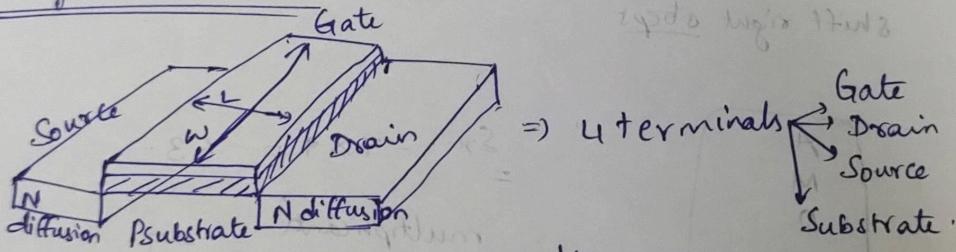


## Unit-3

### \* MOS Transistors:-

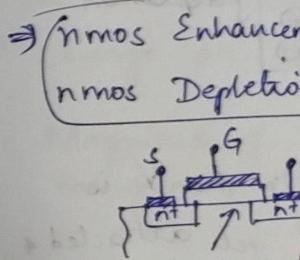
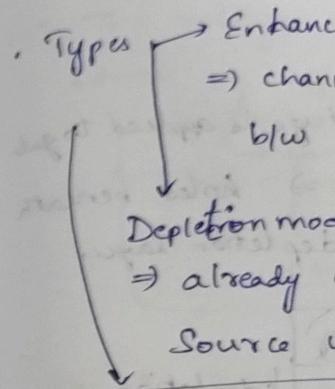
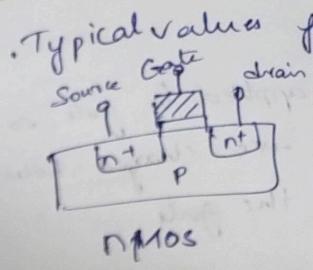
- Extremely useful devices; almost all the electronics like phone, laptop, WiFi & Bluetooth.
- 2 types
  - NMOS  $\Rightarrow$    $\Rightarrow$  switch which connects source to drain  
 $\Rightarrow$  If  $V_{GS} > V_{TH}$  (around 1V), turns on the device.
  - PMOS  $\Rightarrow$    $\Rightarrow$  switch which connects source to drain  
 $\Rightarrow$  If  $V_{GS} < V_{TH}$   $\Rightarrow$  turns on the device  
 $(-ve V_{GS})$

### \* Physical structure:-



normally Source & the substrate are connected together.

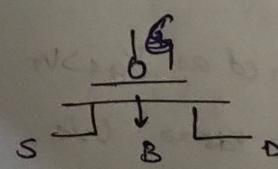
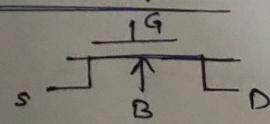
- If the gate is separated from substrate by  $\text{SiO}_2$ , creating similar structure of capacitor plates
- $\Rightarrow$  If +ve voltage applied to gate  $\Rightarrow$  -ve charges are induced on substrate surface & they create (inversion layer) conduction path b/w drain & source.
- $\Rightarrow$  min. voltage needed to create inversion layer  $\Rightarrow$  threshold voltage.
- If  $V_{GS} < V_T \Rightarrow I_{DS} = 0$ .



No channel when  $V_G = 0$

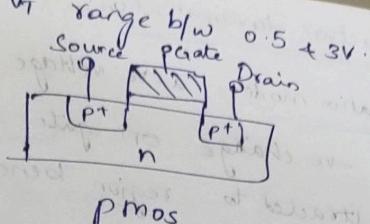
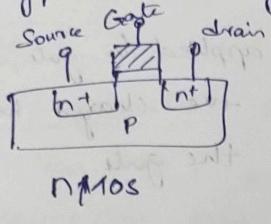
Conduction b/w drain & source by applying  $V_G$   
 [Enhancement]

### Circuit symbols:



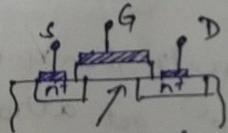
$$\overline{I_D} = \frac{(0.9 - 0.4)^2}{(1.4 - 0.4)L}$$

Typical values for  $V_t$  range b/w 0.5 to 3V.



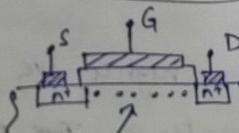
- . Types
  - Enhancement mode
    - ⇒ channel width is not existing is created b/w drain & source with apply of  $V_{GS}$ .
  - Depletion mode
    - ⇒ already existing channel width b/w drain + source is reduced with appln of  $V_{GS}$ .

⇒ NMOS Enhancement      PMOS Enhancement  
NMOS Depletion      PMOS Depletion.



No channel when  $V_G = 0$

Conduction b/w source & drain regions is enhanced by applying gate voltage [Enhancement]

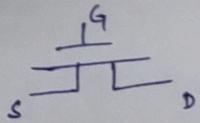
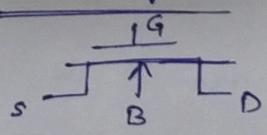


channel when  $V_G = 0$

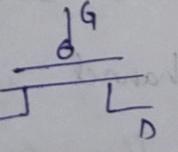
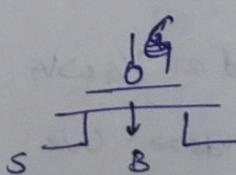
A gate voltage must be applied to deplete the region in order to turn off the transistor.

[Depletion]

Circuit symbols:-



⇒ NMOS

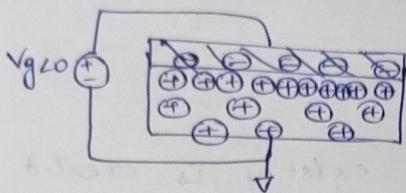


⇒ PMOS

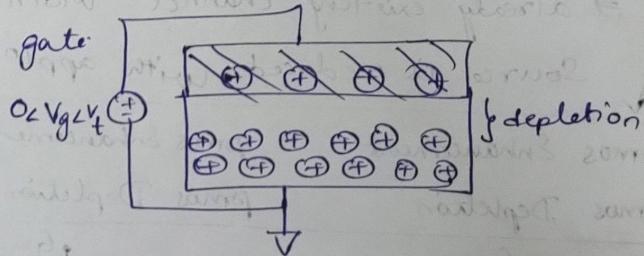
$$(1.4 - 0.4) \mu$$

### \* NMOS structure demonstration:-

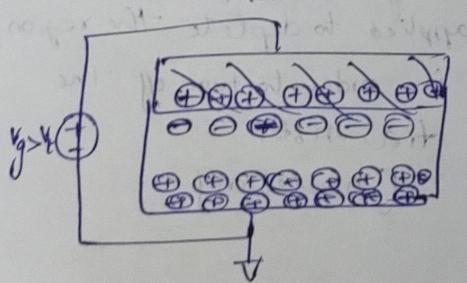
- ① Accumulation mode:- a -ve voltage is applied to gate, so there is -ve charge on gate. The +ve charged holes are attracted to region beneath the gate.



- ② Depletion mode:- A small +ve voltage is applied to gate, resulting in some +ve charge on gate  $\Rightarrow$  holes beneath gate are repelled  $\Rightarrow$  resulting in depletion layer below the gate.



- ③ Inversion:- higher +ve potential greater than  $V_t$  is applied to gate  $\Rightarrow$  some free  $e^-$  get attracted + +ve holes are repelled. The conductive layer of  $e^-$  is called inversion.

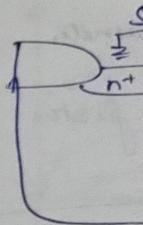


### \* Operating regions:-

- ① Cut-off:-  $\rightarrow V_{GS} > V_t$ ;  $V_{DS} = 0V$ .

• the inversion layer/channel is formed as  $V_{GS} > V_t$  but there is no current flow as  $V_{DS} = 0V$ . We say the transistor is OFF and this is called cutoff.

- convenient
  - $I_D = 0$
- ② Non-saturated
- $V_{GS} > V_t$
  - Assume now
  - The voltage source.



the threshold charge  $P$

- Magnitude voltage

$$\Rightarrow I_{DS} =$$

$$\Rightarrow t =$$

$$\Rightarrow Q_{ch}$$

$$\Rightarrow Q_{ch}$$

$$V_{GS}$$

$$\Rightarrow I_d$$

Applied to gate  
positive charged holes  
in the gate.

Positive charge applied to gate  
holes beneath depletion layer.

Depletion layer

than  $V_t$  is  
attracted.

Layer of  $e^-$

Depletion layer  
is formed

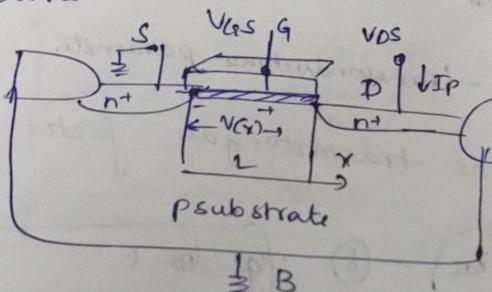
$V_{GS} > V_t$

We  
cutoff

- convenient to say, current through OFF transistor = 0.
- $I_D = 0$ .

## (2) Non-saturation / linear region:

- $V_{GS} > V_t$ ;  $0 < V_{DS} < V_{SAT}$ .
- Assume now that  $V_{GS} > V_t$  & that  $V_{DS}$  is applied.
- The voltage diff. causes  $I_D$  to flow from drain to source.



At a point  $x$ ,  $V(x)$  &  
gate to channel voltage  
 $= V_{GS} - V(x)$ .

Assuming that voltage exceeds

- the threshold voltage all along the channel, induced channel charge per unit area at  $x$  can be  $Q_{channel}$ .

- Magnitude of current  $\uparrow$  linearly with increasing drain voltage till,  $V_{GS} \geq V_t + V_{DS} < V_{SAT} = V_{GS} - V_t$

$$\Rightarrow I_{DS} = \frac{Q_{channel}}{t} \quad t \text{ time taken to drift across the channel}$$

$$\Rightarrow t = \frac{L_{channel}}{\text{Velocity of } e^-} = \frac{L}{\mu_n E} = \frac{L^2}{\mu_n V_{DS}} \quad (2)$$

$$Q_{channel} = C_g V \quad \text{where } C_g = \text{gate capacitance} \\ V = \text{applied } V_{GS} \quad (3)$$

$$\Rightarrow Q_{channel} = C_g \left( V_{GS} - V_t - \frac{1}{2} V_{DS} \right) \quad (4)$$

$V_{GS} - V_t$  = effective voltage.

$$\Rightarrow I_{DS} = \frac{C_g \left( V_{GS} - V_t - \frac{1}{2} V_{DS} \right)}{\frac{L^2}{\mu_n V_{DS}}} \quad (5)$$

$$(1.4 - 0.4)L$$

$$\Rightarrow C_g = \frac{\sum A^2}{d} = \frac{\sum \epsilon_0 \cdot W \cdot L}{t_{ox}} - (6)$$

$$\Rightarrow I_{ds} = \frac{\epsilon_0 \cdot W \cdot L}{t_{ox}} \left( V_{gs} - V_t - \frac{1}{2} V_{ds} \right) \frac{L^2}{\mu_n V_{ds}} - (7)$$

where  $\frac{W}{L}$  = aspect ratio

$$k_n = \frac{\mu_n \epsilon_0}{t_{ox}} = \text{transconductance parameter}$$

$$B_n = k_n \frac{W}{L} = \text{Mos transistor gain factor}$$

$$\Rightarrow I_{ds} = B \left( V_{gs} - V_t - \frac{1}{2} V_{ds} \right) V_{ds} - (8)$$

$$V_{gs} \geq V_{th} +$$

$$V_{ds} < V_{sat} = V_{gs} - V_{th}$$

③ Saturation region: -  $V_{gs} > V_t$ ;  $0 < V_{ds} \geq V_{gs} - V_t$  or

$$0 < V_{ds} \leq V_{sat}$$

$\Rightarrow$  As value of  $V_{ds}$  is ↑, the assumption that the channel voltage is larger than threshold all along the channel all along channel ceases to hold.

$\Rightarrow$  This happens when  $V_{gs} - V(x) < V_t$

$\Rightarrow$  At that pt, induced charge is 0, the conducting channel disappears or is pinched off.

$\Rightarrow$  Beyond a certain pt, the surface conc. is like  $\ln\left(\frac{n_s}{n_B}\right)$  & ∴ stop ↑ & the surface potential is pinched off. when  $n_s \approx N_A$ .

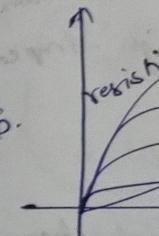
$$\Rightarrow \psi_s = \frac{kT}{q} \ln\left(\frac{n_s}{n_B}\right) \quad (\psi_s = \text{surface potential})$$

$$\Rightarrow V_{gs} > V_t$$

$\Rightarrow$  Subst

$$\Rightarrow I_{ds} =$$

$$I_{ds} =$$



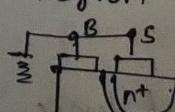
\* Channel

• the current constant

• the effective

modulation

region -



$$V_{ds} = 160$$

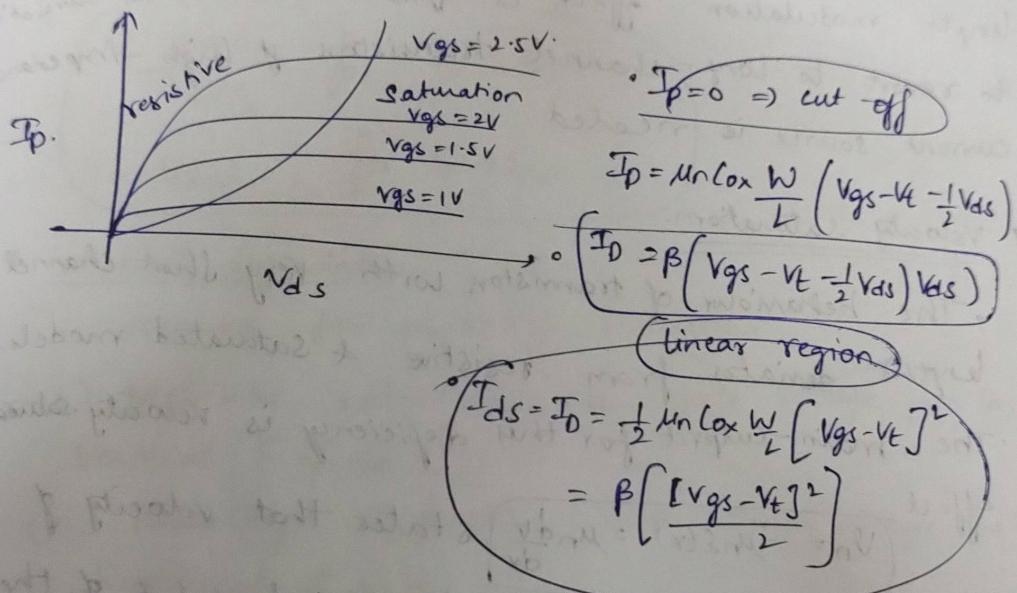
$$\Rightarrow V_{GS} > V_T ; 0V < V_{DS} \geq V_{GS} - V_T \quad (\text{or}) \quad 0V < V_{DS} = V_{SAT}$$

$\Rightarrow$  Substitute  $V_{DS} = V_{GS} - V_T$ . in ⑧

$$\Rightarrow I_{DS} = \beta \left( V_{GS} - V_T - \frac{1}{2}(V_{GS} - V_T) \right) V_{GS} - V_T$$

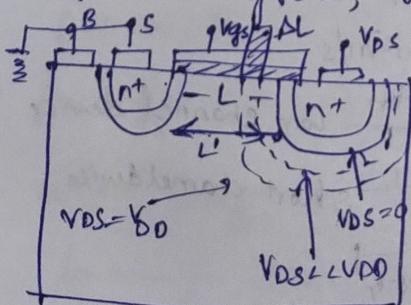
$$\Rightarrow I_{DS} = \beta \left( \frac{[V_{GS} - V_T]^2}{2} \right)$$

$$\boxed{I_{DS} = \beta \left[ \frac{V_{DS}^2}{2} \right]} \quad - ⑨$$



### \* Channel-length modulation:-

- the transistor in saturation mode acts as a perfect current source (or) current bw source & drain is constant. But this is not entirely correct
- The effective length of conductive channel is actually modulated by applied  $V_{DS}$ .  $\uparrow V_{DS}$ , causes depletion region to grow, reducing length of effective channel.



$$L = L' - AL$$

$$\bullet I_{DS} = \beta \left[ \frac{V_{DS}^2}{2} \right]$$

$$I_D = \frac{k_n}{2} \frac{W}{L} (V_{GS} - V_T)^2$$

Current  $\uparrow$  when  $L$  is  $\downarrow$ .

$$I_D = I_{D0} (1 + \alpha V_{DS})$$

$\Rightarrow \alpha$  is an empirical parameter,  
called channel length modulation.

$\Rightarrow$  In shorter transistors, drain-junction depletion region presents a larger fraction of channel & channel length modulation effect is visible.  $\therefore$  it is advisable to resort to long-channel transistor if high-impedance current source is needed.

#### \* Velocity saturation:-

- The behaviour of transistors with very short channel lengths deviates from resistive & saturated models.

- The main-culprit for this deficiency is velocity saturation effect.

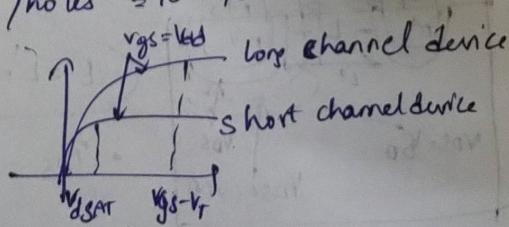
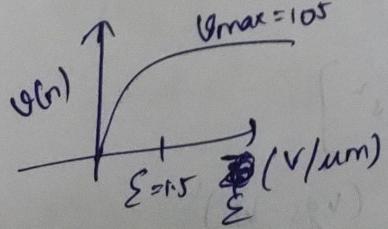
$(V_n = -\mu_n \epsilon(x) = \mu_n \frac{dv}{dx})$  states that velocity of carriers is prop to electrical field, independent of the value of that field.

$\Rightarrow \epsilon$ -field is given by  $\frac{dv}{dx} = \frac{V_{DS}}{L}$ .

$\Rightarrow$  If length &  $dx \downarrow$ ,  $E$ -field  $\uparrow$  & hence saturates the velocity.

- Velocity is saturated beyond  $E_{critical}$ .  $e^\ominus$  encounter more collision & hence don't pick up speed.

$\Rightarrow$  Max. velocity of  $e^\ominus$ /holes  $= 10^5 \text{ m/s}$



In short channel  
 $V_{DS}$  reaches  
 $E_{critical}$   
 $\Rightarrow$  At high field  
linear model  
reaches  $E_{critical}$   
due to saturation

$$V = \frac{\mu_n \epsilon}{1 + \epsilon / E_c}$$

$= V_{sat}$

#### \* Mobility degeneration

- 2 reasons

- (i) Mobility due to vertical field

- (ii) Mobility due to horizontal field

- Mobility due to bulk mobility

$$\Rightarrow \frac{1}{E_{eff} E_{critical}}$$

Parameter

- for large  $E_{eff}$

- When  $E_{eff}$  is lowering

$$\left| \frac{I_{D2}}{I_{D1}} \right| = \frac{25 \times 10^{-1}}{10^{-1}}$$

critical parameters  
length modulation  
depletion  
channel & channel  
it is advised  
of high - impedance

short channel  
aturated models.  
velocity saturation  
velocity of  
erent of the

hence saturates

$e^{\ominus}$  encountered

channel device  
channel device

In short channel, the device enters saturation before  $V_{DS}$  reaches  $V_{GS} - V_T$ .

At high field strengths, carriers fail to follow this linear model. When the E-field along the channel reaches  $E_{critical}$ , velocity of carriers tends to saturate due to scattering effects.

$$\Rightarrow \frac{V = \mu n E}{1 + E/E_c} \text{ for } E_p \leq E_c$$
$$= V_{sat} \text{ for } E_p \geq E_c$$

#### \* Mobility degradation.

- 2 reasons for mobility reduction

- (i) Mobility reduction with gate voltage due to vertical E-field
- (ii) Mobility reduction with drain voltage due to horizontal E-field.
- Mobility degradation reduces surface mobility w.r.t bulk mobility.

$$\frac{M_H}{L_{eff} E_{crit}} = \frac{\mu_0}{1 + \alpha_2 V_{DS}} \quad (12)$$

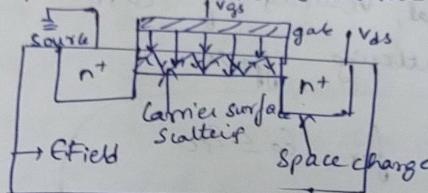
$\Rightarrow \frac{1}{L_{eff} E_{crit}}$  is referred as drain bias mobility reduction parameter & in some texts denoted as  $\alpha_2$ .

- for large transistor,  $\alpha_2$  is  $\ll 1$  thus  $M_H = \mu_0$ .
- when  $E_{eff} \downarrow$ ,  $\alpha_2 \uparrow$  &  $\alpha_2 V_{DS}$  becomes important, lowering the mobility below  $\mu_0$ .

$$I_{D2} = \frac{10^{-1}}{2.5} = 4 \text{ mA}$$
$$\times 5 \times 10^{-2}$$

\* Mobility reduction with gate voltage due to vertical E field:-

- In MOS transistor, current flows very close to silicon surface. As consequence, mobility of current carriers is lower than deep inside the substrate due to random scattering mech.



- A vertical E-field exists in MOSFET due to applied gate voltage, which creates conduction channel.
- When carriers move within channel under horizontal E-field, they feel the effect of gate induced vertical E-field, pushing carriers towards the gate oxide. This provokes carriers to make the collision with oxide channel interface. The oxide - channel interface is rough & imperfect, thus carriers lose mobility. This effect is known as surface scattering.
- The surface mobility depends on how much carriers interact with the interface. We can note that as surface mobility in absence of vertical E-field. The higher the E-field, lower is surface mobility.

$$\Rightarrow \mu_H = \frac{\mu_0}{1 + \alpha_s V_{ds}} = \frac{\mu_0}{1 + \alpha_s V_{ds}}$$

(at zero)

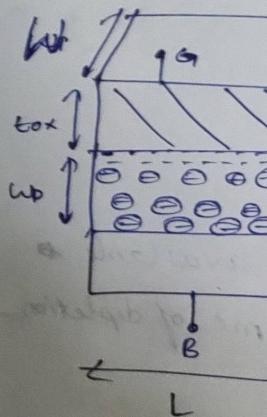
. At low E field linearly with increases above more slowly  
 $E\text{ field} = 10^5 \text{ V/cm}$

\* Derivation of

- Gate potential as threshold

$$\Rightarrow V_{GB} = \phi_0$$

$$\Rightarrow (V_{GB} = \phi_0)$$



$\phi_s$  is pinned

$$\Rightarrow n_s \approx n_i$$

$$\Rightarrow (\phi_s = \phi_i)$$

As per law

$$as n_s \approx n_i$$

$$P = \frac{n_i}{N_i}$$

voltage due to vertical  
very close to silicon  
current carriers  
rate due to random

ET due to applied  
channel.

der horizontal  
te induced vertical  
gate oxide. This  
ion with oxide  
interface is  
nobility. This

ch  $e^-$  interact  
~~as~~ as  
field. The  
bility,

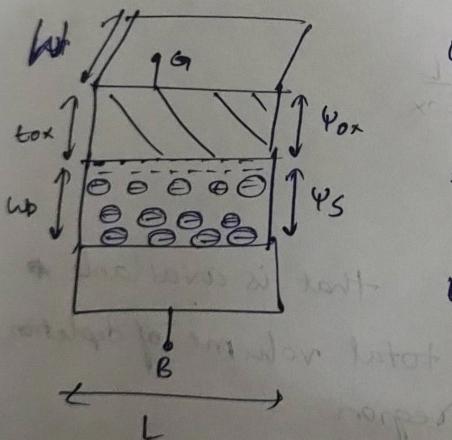
- At low  $E$  field,  $\propto$  drift velocity  $V_d$  in channel varies linearly with  $E$  field intensity. However as  $E$  field increases above  $10^4 \text{ V/cm}$ , the drift velocity tends to  $\propto$  more slowly & approaches  $V_{d(\text{sat})} = 10^7 \text{ cm/s}$  around  $E$  field  $= 10^5 \text{ V/cm}$ .

### \* Derivation of threshold voltage:

- Gate potential needed to invert the surface is called as threshold voltage.

$$\Rightarrow V_{GB} = \text{potential across oxide} + \text{potential across the surface}$$

$$\Rightarrow V_{GB} = \psi_{ox} + \psi_s \quad (1)$$



The  $E$ -field terminates across the immobile  $-ve$  charges & also at free  $e^-$ .

$\Rightarrow$  As  $e^-$  are attracted towards the surface, beyond a certain pt, the surface potential is pinned as small  $\uparrow$  in surface conc. is like 'ln' of that

$$\Rightarrow \ln\left(\frac{n_s}{n_{s0}}\right) = \frac{q\psi_s}{kT} \quad (2)$$

$\psi_s$  is pinned when surface conc  $\approx$  bulk conc.

$$\Rightarrow n_s \approx N_A$$

$$\Rightarrow \psi_s = \frac{kT}{q} \ln\left(\frac{n_s}{n_{s0}}\right) \quad (2)$$

As per law of mass action,  $n_p = n_i e^{-L}$ .

$$\text{as } n_s \approx N_A$$

$$P = \frac{n_i e^{-L}}{N_A}$$

$$\Rightarrow \psi_s = \frac{kT}{q} \ln\left(\frac{\frac{N_A}{n_i e^{-L}}}{N_A}\right)$$

$$N_A = kT/q \ln\left(\frac{N_A}{n_i}\right) \times 10^{-2}$$

$$\Rightarrow \Psi_s = 2 \frac{kT}{q} \ln \left( \frac{N_A}{n_i} \right) - \textcircled{3} \text{a}$$

Strong inversion occurs at  $V = 2V_{\text{Fermi}}$

$$\Rightarrow \Psi_s = 2 \frac{kT}{q} \ln \left( \frac{N_A}{n_i} \right)$$

$= 2$  (Fermi potential)

$$\textcircled{4} \quad \Psi_s = 2 \phi_F$$

$$\Rightarrow \Psi_{ox} = - \frac{(Q_D' + Q_I')}{C_{ox}} - \textcircled{3}$$

$Q_D'$  &  $Q_I'$  represent depletion & inversion charges

$Q_D + Q_I$  represent depletion & inversion charges per unit area.

$$\Rightarrow C_{ox} = \epsilon_r \epsilon_0 \frac{A}{d} = \epsilon_r \epsilon_0 \frac{WL}{t_{ox}}$$

$$\Rightarrow \frac{\epsilon_r \epsilon_0}{t_{ox}} = C_{ox}$$

$Q_D'$  = total depletion charge that is available

~~$Q_D$~~   $Q_D$  = conc. into total volume of depletion region

$$\Rightarrow Q_D' = q N_A (WL W_D) - \textcircled{4}$$

$$\Rightarrow W_D = \sqrt{\frac{2 \epsilon_s | \Psi_s |}{q N_A}} - \textcircled{4(a)}$$

$\Rightarrow$  Using  $\textcircled{4}$  in  $\textcircled{3}$

$$\Rightarrow Q_D' = q N_A \left( WL \sqrt{\frac{2 \epsilon_s | \Psi_s |}{q N_A}} \right) - \textcircled{5}$$

$$\Rightarrow \textcircled{5} \quad \boxed{Q_D' = \left( \sqrt{2 \epsilon_s | \Psi_s | q N_A} \right) WL} - \textcircled{6}$$

$$\Rightarrow V_{gb} = \Psi_{ox}$$

$$\Rightarrow V_{gs} =$$

$\Rightarrow$  Gate P

$$\Rightarrow \boxed{\frac{Q_I}{C_{ox}} =}$$

where  $\boxed{\Psi_s}$

$\Rightarrow$  In general occurs in

$\Rightarrow V_t$  is f

$$\Psi_s = 1 -$$

$$\Rightarrow V_{Th} =$$

$$\Rightarrow V_{I0}$$

$\gamma$  is

the im

$\Rightarrow$  Observed

while

$$= I_o \cdot \frac{(1.4 - 0.4)^2}{(1.4 + 0.4)^2}$$

$$\Rightarrow V_{gb} = \psi_{ox} + \psi_s \quad (Q_D = -\sqrt{2\varepsilon_{si}(\psi_s) N_A q})$$

$$\Rightarrow V_{gs} = \left( \psi_s - \frac{Q_D}{C_{ox}} \right) - \frac{Q_D}{C_{ox}}$$

Gate potential needed to invert

$$\Rightarrow \left( \frac{Q_D}{C_{ox}} = \psi_s = \sqrt{2\varepsilon_{si}(\psi_s) N_A q} \right) = V_{Th} \quad (8)$$

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

- $\Rightarrow$  In general, value of  $V_{gs}$  where strong inversion occurs is called  $V_T$ .
- $\Rightarrow V_T$  is function of
- difference in work-func. b/w gate & substrate
  - oxide thickness
  - Fermi voltage
  - dosage of ions implanted for threshold adjustment

- $\Rightarrow$  In case a substrate bias voltage is applied,
- $$\psi_s = | -2\phi_F + V_{SB} |$$
- $$\Rightarrow V_{Th} = V_{T0} + \gamma \left( \sqrt{| -2\phi_F + V_{SB} |} - \sqrt{| -2\phi_F |} \right) \quad (9)$$
- $V_{T0}$  is threshold voltage for  $V_{SB} = 0$ ,  $\gamma$  is called body effect coefficient, expresses the impact of changes in  $V_{SB}$ .
- $\Rightarrow$  Observe that  $V_{Th}$  has +ve value for typical nmos, while it is -ve for normal PMOS transistor

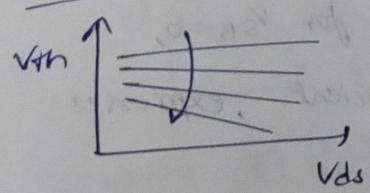
$$\Rightarrow V_{Th} = V_{IO} + \gamma \left( \sqrt{1-2\phi_F} + V_{SB} - \sqrt{1-2\phi_F} \right) - ⑦$$

\* In a long channel device, the channel formation is controlled by gate & substrate. The gate voltage will control essentially all the space charge induced in the channel region. As length  $\downarrow$ , charge control of channel is shared by the four terminals called, Charge Sharing.

- The total charge below gate controlled by gate voltage in short channel device, is correspondingly less than that controlled by gate in long-channel device.

- Consequently, lower gate voltage is req. to attain threshold in a short-channel device. Now as drain voltage inc. the reverse biased space charge region at the drain extends further into channel area & gate will control even less bulk charge.

### Effects of threshold voltage:-



- The deeper depletion region accompanied by larger surface potential, the device can conduct more current.

- This effect can be considered as the reduction of  $V_{Th}$  as drain current is func. of  $(V_{GS} - V_{Th})$ .

- Inc. in  $V_{ds}$  & reduction of channel length will decrease effective  $V_{Th}$ .

### Down-scaling

- Cmos have

disadv  $\rightarrow$  E  
S

### Reducing the E-field

### Types of scaling

#### ① Const. field

pioneering  
from which

- to  $\uparrow$  per

- ① reduce

- ② inc.

device

- const. f  
the power

- However  
Voltage a

#### ② Const.

- does r  
prefere  
'compar

$I_D =$

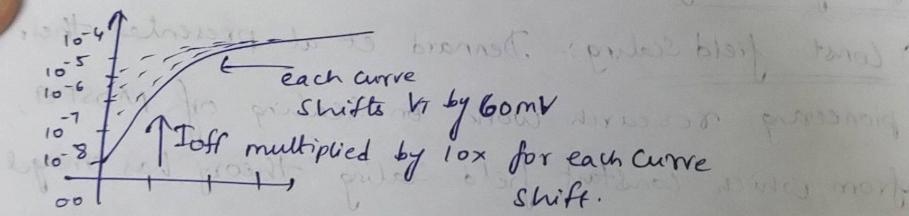
$I_D =$

disadv -  $E$  field  $\uparrow$  as min. feature length is reduced  
 • This leads to velocity saturation, mobility degradation,  $\uparrow$  leakage currents & lower breakdown voltages.

\* Components of leakage power due to scaling of MOSFET:

$$P_{\text{dynamic}} = C_L V_{DD} f \quad P_{\text{static}} = I_{\text{leak}} \cdot V_{DD}$$

- Supply voltage has been scaled down - to keep power consumption under control
- The threshold voltage scaling results in substantial  $\uparrow$  of leakage current



\* Short channel effects:

- Reverse biased diode leakage
- Gate-oxide tunneling
- Gate induced drain leakage (GIDL)
- Subthreshold leakage
- Drain-induced barrier lowering (DIBL)

\* DIBL & Punch through:

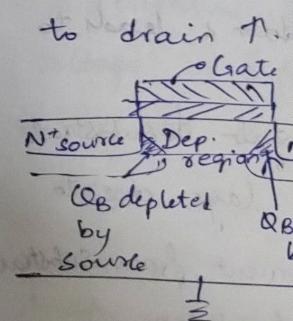
- One of major challenges in transistor Scaling are short channel effects with channel lengths ( $L_G < 100\text{nm}$ )

- DIBL occurs short-channel interaction with potential barrier injection of carriers to reduce gate voltage

① Punch through:

the drain current can be increased for a given drain voltage

② DIBL:- the drain current increases



\* reverse-bias leakage

• the current  $I_{DS}$  is exponential Source-substrate bias is applied

• to limit the n-region area leakage current diode leakage

1) - ⑦  
 nel formation,  
 gate voltage  
 charge  
 length ↓,  
 by the  
 ing  
 kled by  
 is corresponding  
 -channel  
 - attain  
 as drain  
 charge region  
 nnel area  
 e  
 on accompania  
 trial, the  
 re current  
 eduction  
 mbo  
 hn)  
 length

### \* Down-scaling of CMOS

- CMOS have been scaled to achieve
  - higher density
  - higher performance
  - lower power consumption
- disadv → E-field within gate oxide grow larger Short channel effects
- ⇒ Reducing the critical dimensions while keeping the E-field const. yields
  - higher speed
  - reduced power consumption

### \* Types of Scaling :-

- Const. field scaling :- Dennard et al. presented their pioneering research work on Scaling of MOSFET, from which constant field scaling theory has emerged.

- to ↑ performance of a MOSFET, we must
  - reduce linearly the size of transistor
  - inc. doping conc. in a way which keeps E-field device const
- const. field scaling yields largest reduction in the power delay product of single transistor.
- However, it requires a reduction in power supply voltage as one of the min. feature size.

- Const. voltage scaling :-

- does not have this problem & is therefore the preferred scaling method since it provides voltage compatibility with older ckt technologies.

$$\Rightarrow I_D \propto \frac{2\mu C_{ox} W}{L} (V_{GS} - V_{T,12})$$

$$\begin{aligned}
 V_{CE} &= 0V \\
 &= 2k_L \\
 &= 77.2 k_L \\
 &= 77.2 \text{ kV}
 \end{aligned}$$

$V_{GS}$  of a mosfet  
 saturation is 900  
 Neglecting the ch  
 current for an appl

length is reduced  
mobility  
via breakdown

ing of MOSFET.

$V_{DD}$ .

keep power

in substantial

diode leakage

arching

rain leakage

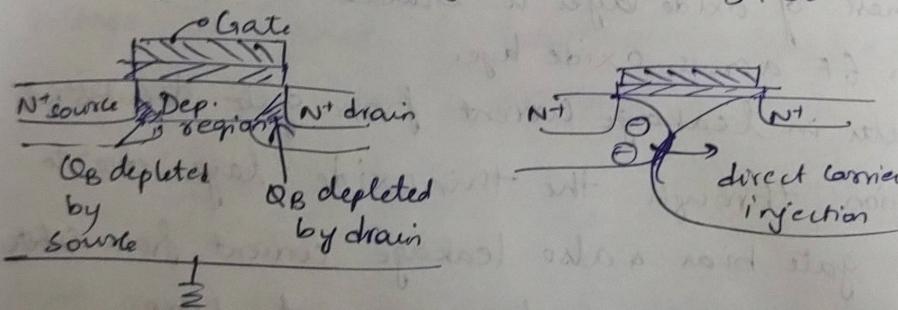
lowering

Scaling  
el lengths

- DIBL occurs when high drain voltage is applied to short-channel devices, the depletion region of drain interact with source depletion region as source potential barrier height is lowered resulting in injection of carriers into channel from the source due to reduced threshold voltage as it is independent of gate voltage.

① Punch through:- For short channel, depletion region from the drain can reach the source side & reduces the barrier for  $e^-$  injection.

② DIBL:- the barrier for  $e^-$  injection from source to drain  $N^+$ . This is known as DIBL.



#### \* reverse-biased diode leakage:-

- the current due to reverse biased drain substrate + source-substrate junction is relatively small, but there is an exponential rise in current when large forward biased is applied to the substrate region.
- to limit the effect due to scaling, the p-region & n-region are heavily doped & results in BTBT leakage current & dominates the reverse biased diode leakage.

$$I_C = \frac{V_{CC}}{R_C} \text{ at } V_{CC} = 20V$$

$$\frac{V_{CC} - V_{BE}}{R_B} = \frac{20}{10mA}$$

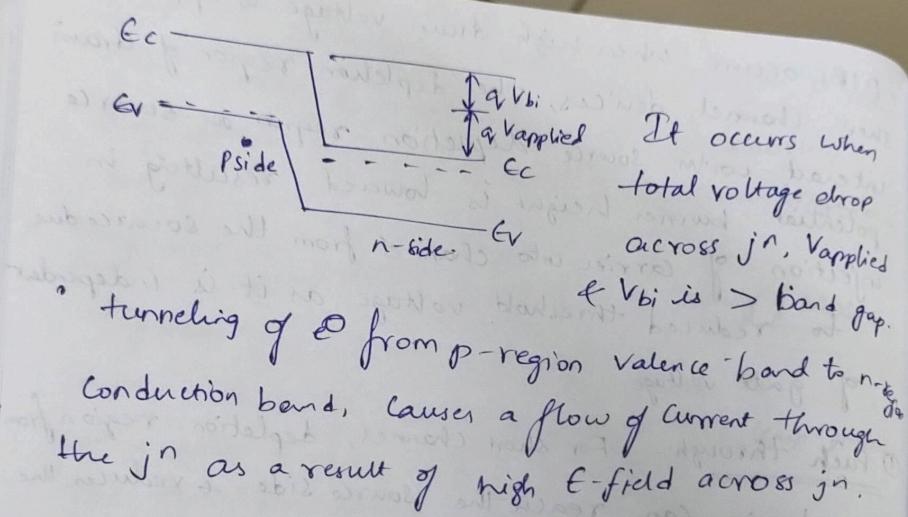
$$\frac{V_{CC} - V_{BE}}{I_B} = \frac{20}{25mA}$$

$$R_C = 2k\Omega, R_B =$$

2 to source  
of 400mV, Work  
is observed to  
effect + assumption  
find the

$$I_D = C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2$$

$$\frac{I_D}{I_{D_s}} = (V_{GS} - V_{TH})^2$$

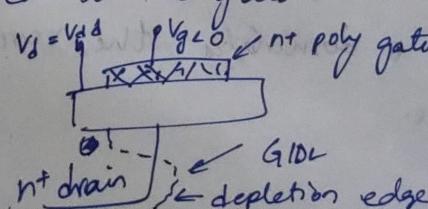


#### \* Gate-oxide Tunneling :-

- With scaling of transistor to nm dimensions the thickness of oxide layer is scaled down & leads to ↑ in EF across oxide layer.
- results in leakage current from gate to substrate region through the thin oxide layer due to -ve gate bias & also leakage current from substrate region to gate due to +ve gate bias.

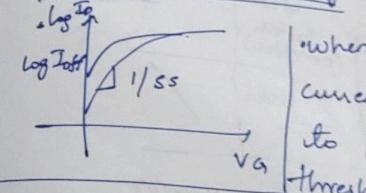
#### \* Gate induced drain leakage:-

- A high field effect in MOS transistor drain  $j^n$  results in GIDL as field is crowded near the surface due to narrowing of depletion layer at the silicon surface when large negative-biased is applied to the gate



nt drain region under even inverted. GIDL drain doping conc. as it gives lower series drive currents  $I_{\text{d}}$ .

#### \* Subthreshold leakage:



estimated using

- A derivative of log voltage is defined as is defined the  $SS$  as

$$SS = \frac{dVg}{d(\log I_d)} = \frac{k_B T}{q \log e}$$

At room temp,  $SS$  Const,  $T$  is room temp Capacitance. per unit

- The presence of  $SS$  60mV voltage reduction decreases with a fa

$SS$ , if  $e_m$  The MOSFET due to in provides the quest computing

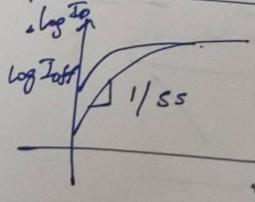
It occurs when total voltage drop across  $j^n$ ,  $V_{\text{applied}}$  is  $>$  band gap. Hence band to current through it across  $j^n$ .

dimensions the own & leads to gate to substrate layer due to ent from substrate bias.

MOSFET drain led near the layer at reverse-biased

$n^+$  drain region under the gate can be depleted & even inverted. GIDL can be minimised by  $\uparrow$  the drain doping conc. and also by abrupt doping as it gives lower series resistance req. for high transistor drive currents ( $I_D$ ).

#### \* Subthreshold leakage:



When  $V_{GS} < V_{Th}$ , weak inversion current flows from source region to drain is defined as sub-threshold current & it can be estimated using subthreshold swing (SS).

- A derivative of  $\log I_D$  of drain current versus gate voltage is defined as subthreshold slope & its inverse is defined the SS & expressed as

$$SS = \frac{dV_G}{d(\log I_D)} = \frac{k_B T}{q \log e} \left( 1 + \frac{C_{ch}}{C_{ox}} \right)$$

- At room temp,  $SS = 60 \text{ mV/decade}$ ,  $k_B = \text{Boltzmann Const}$ ,  $T$  is room temp.,  $C_h + C_{ox}$  = channel + oxide capacitance per unit area.

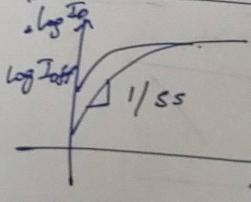
- The presence of SS slope implicates that for every 60mV voltage reduction, weak inversion current decreases with a factor of 10.

- So, if we The limit in subthreshold slope of MOSFET due to increase in OFF state current provides the quest for energy efficiency in computing

age ( $V_{GS}$ ) of a m  
in saturation is  
1mA. Neglecting  
that the Mos  
drain current for

It occurs when total voltage drop across  $j^n$ ,  $V_{\text{applied}}$  is  $>$  band gap. Hence band to n<sup>+</sup> drain region under the gate can be depleted & even inverted. GIDL can be minimised by ↑ the drain doping conc. and also by abrupt doping as it gives lower series resistance req. for high transistor drive currents [Eq].

#### \* Subthreshold leakage:-



When  $V_{\text{GS}} < V_{\text{Th}}$ , weak inversion current flows from source region to drain is defined as sub-threshold current & it can be estimated using subthreshold swing (SS)

- A derivative of  $\log I_D$  of drain current versus gate voltage is defined as subthreshold slope & its inverse is defined the SS & expressed as

$$SS = \frac{dV_G}{d(\log I_D)} = \frac{k_B T}{q \log e} \left( 1 + \frac{C_{ch}}{C_{ox}} \right)$$

- At room temp,  $SS = 60 \text{ mV/decade}$ ,  $k_B = \text{Boltzmann Const}$ ,  $T$  is room temp.,  $C_h + C_{ox}$  = channel + oxide capacitance per unit area.

- The presence of SS slope implicates that for every 60mV voltage reduction, weak inversion current decreases with a factor of 10.

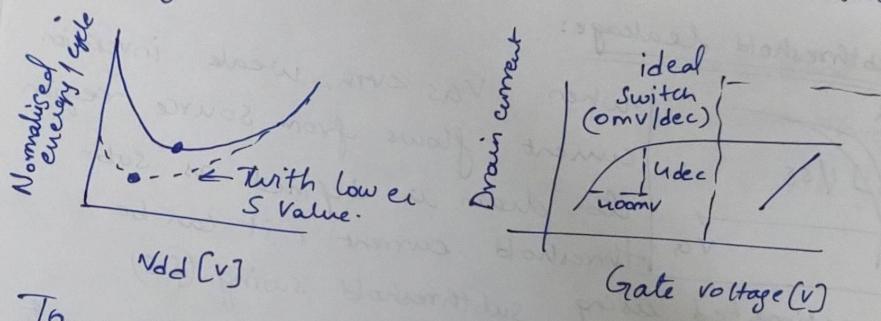
- So, if we The limit in subthreshold slope of MOSFET due to increase in OFF state current provides the quest for energy efficiency in

computing

1)  $L$

### \* CMOS energy efficiency:-

• Dynamic energy reduces quadratically as the supply voltage is scaled down. To maintain certain switching speed, threshold voltage of MOSFET must be scaled down as well, which ↑ the leakage energy.



To overcome CMOS energy efficiency limit, alternative transistor designs which can achieve a steeper sub-threshold swing.

### \* Dynamic behaviour of MOSFET transistor:-

• dynamic response of a MOSFET transistor is a sole function of time it takes to discharge the parasitic capacitances that are intrinsic to the device, & the extra capacitance introduced by the interconnecting lines.

behaviour of intrinsic capacitances originate from

3 sources → bank MOS Structure

→ channel charge

depletion regions of reverse biased pn-junctions of drain & source.

### \* Gate

- The channel unit
- The gate
- $C_g$  is

• Ideally at the & drain oxide

• Hence, than drain

• It all gate & so

(strictly)

$$C_{GS0} =$$

$$C_{gate} =$$

### \* Channel

- The g is both components

ically as the  
naintain certain  
ge of MOSFET  
ich ↑ the leakage

tage (V)

nit, alternative  
steeper sub-

ctor is a  
to discharge  
ntrinsic to  
ntroduced by

te from

biased p<sup>n</sup>

### \* Gate capacitance:-

- The gate of Mosfet is isolated from conducting channel by gate oxide that has a capacitance per unit area equal to  $C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$ .
- The total value of this capacitance is called  $C_g$ .
- $C_g$  is decomposed into 2 elements
  - One part of  $C_g$  contributes to channel charge
  - another part is  $C_{so}$  solely due to the topological structure of transistor.
- Ideally, source + drain diffusion should end right at the edge of gate oxide. In reality, both source & drain tend to extend somewhat below the oxide by an amount  $x_d$ , called lateral diffusion.
- Hence, effective channel length  $L$  becomes shorter than drawn length  $L_d$  by a factor of  $\Delta L = 2x_d$ .
- It also gives rise to parasitic capacitance b/w gate & source that is called overlap capacitance (strictly linear & has fixed value)

$$C_{so} = C_{GDO} = C_{ox} \times d \times W = C_{ox} W$$

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

### \* Channel capacitance:-

- The gate - to - channel capacitance  $C_{gc}$  varies in both magnitude & in its division into 3 components

$$V_{CE} = 0V$$

$$= 2kA$$

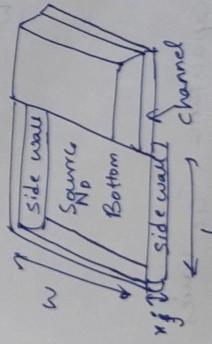
$$= 0.7 = 772$$

$$= 772 kA$$

voltage  $(V_{GS})$ ,  
working in sat  
to be  $1mA$ . N  
ering that -  
the drain c

$$(V_S - V_{TH})^2$$

- ①  $C_{GCS}$   $\Rightarrow$  channel capacitance :-
- ②  $C_{GDP}$
- ③  $C_{GCB}$
- $\Rightarrow$
- \*  $C_{GSW} = C_{JSW}$   
 $= C_{JSU}$   
 $(C_{SW} > C_{JSU})$
- \*  $C_{GSW} = C_{JSU}$
- Cutoff :- no-channel, total capacitance  
 $\Rightarrow C_{ox} W_{eff}$  appears between  
 gate & bulk.
- Triode :- Inversion layer - acts as  
 conductor
- Saturation :- Pinch off :  $C_{GD} \propto 0$ ,  $C_{GB} = 0$   
 $\Rightarrow C_{GS} \Rightarrow$  average  $(\frac{2}{3}) C_{ox} W_{eff}$ .
- \* Junction capacitance :-
  - A final capacitive component is contributed by the reverse biased source-body and drain body pn-junctions.
  - The depletion-region capacitance is non-linear & decreases when reverse bias is raised as discussed earlier.
  - To understand the components of junction capacitance we look at source region + its surroundings
- \*  $C_{GS} = C_{GCS}$   
 $C_{GD} = C_{GCS}$   
 $C_{GB} = C_G$   
 $\Rightarrow C_{SB} = C_{Sdi}$   
 $C_{DB} = C_{di}$
- \* Mos - Parant  
Source - drain
  - Bottom p plate
  - $\Rightarrow C_{bottom} = C_{WLs}$
  - Side wall junctions - formed by source( $N_D$ ) & p channel stop ( $N_A$ )
  - Graded junction ( $m = 1/3$ )
- \* The performance affected by a being the Source region
- \* This effect



$$C_{SW} = C'_{JSW} \times j (W + 2L_S)$$

$$= C_{JSW} / (W + 2L_S)$$

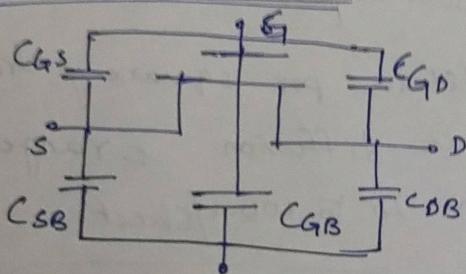
$$C_{SW} = C'_{JSW} \times j \quad j = j^{\text{depth}}$$

$$C_{diff} = C_{bottom} + C_{SW}$$

$$= C_j * \text{Area} + C_{JSW} * \text{Perimeter}$$

$$= C_i L_S W + C_{JSW} (2L_S + W)$$

\* Capacitive device model:-



$$C_{GS} = C_{GCS} + C_{GSO};$$

$$C_{GD} = C_{GCD} + C_{GOO}$$

$$C_{GB} = C_{GCB}$$

$$\Rightarrow C_{SB} = C_{diff}$$

$$C_{DB} = C_{diff}$$

\* MOS- Parasitic & Contact resistance:-

Source-drain resistance:-

The performance of CMOS ckt may further be affected by another set of parasitic elements, being the resistances in series with drain + source regions.

This effect become more pronounced when

$$V_{CE} = 0V$$

$$= 2k\mu$$

$$0 - 0.7 \quad 25 \mu A = 772$$

$$= 772 k\mu$$

voltage  $(V_{GS})$   
working is sat  
to be 1mA. N  
suming that -  
the drain a

$$(V_S - V_{TH})^2$$

transistors are scaled down, as this leads to shallower junctions + smaller contact openings become smaller.

$\Rightarrow$  resistance of drain (source) region,

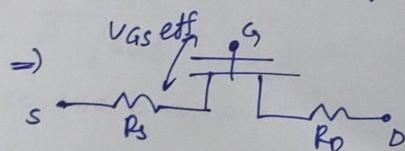
$$R_{S,D} = \frac{L_{S,D}}{W} R_{\text{sheet}} + R_c$$

$R_c$  = contact resistance

$W$  = width of transistor

$L_{S,D}$  = length of source/drain region

$R_{\text{sheet}}$  = sheet resistance per square of drain source diffusion. & ranges from 20 to 100  $\Omega/\text{square}$



$$R_S = (L_S/W) R_{\text{sheet}} + R_c$$

$$R_D = (L_D/W) R_{\text{sheet}} + R_c$$

$\Rightarrow$  The series resistance causes a deterioration in the device performance, as it reduces the drain current for given control voltage.

- Keeping its value as small as possible is thus an important design goal for both the device & circuit engineer.

- One option, popular in most contemporary processes is to cover the drain & source regions with low resistivity material such as titanium / tungsten. This process is called silicidation & effectively reduces sheet resistance to values in range from 1 to 10  $\Omega/\text{square}$ .