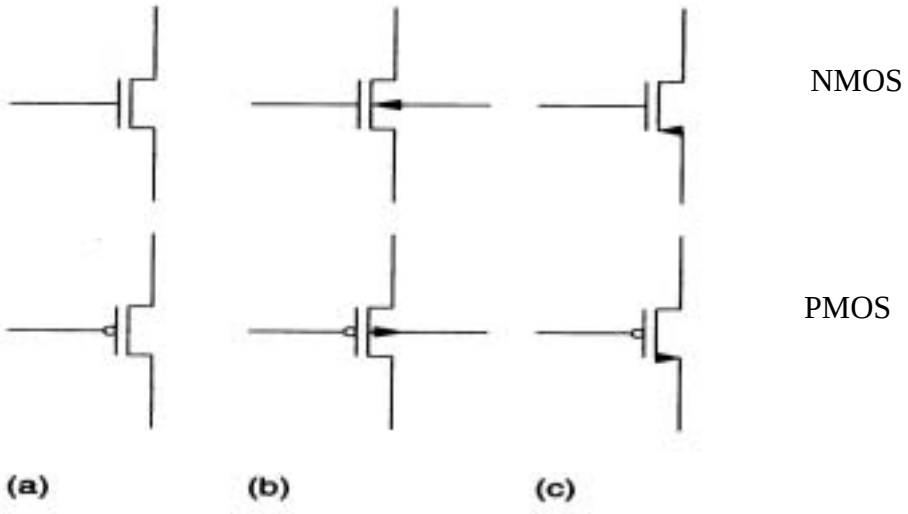
Chapter 2 MOS Transistor theory

2.1 Introduction

- An MOS transistor is a majority-carrier device, in which the current in a conducting channel between the source and the drain is modulated by a voltage applied to the gate.

- Symbols



- NMOS (n-type MOS transistor)

(1) Majority carrier = electrons

(2) A positive voltage applied on the gate with respect to the substrate enhances the number of electrons in the channel and hence increases the conductivity of the channel.

(3) If gate voltage is less than a threshold voltage Vt , the channel is cut-off (very low current between source & drain).

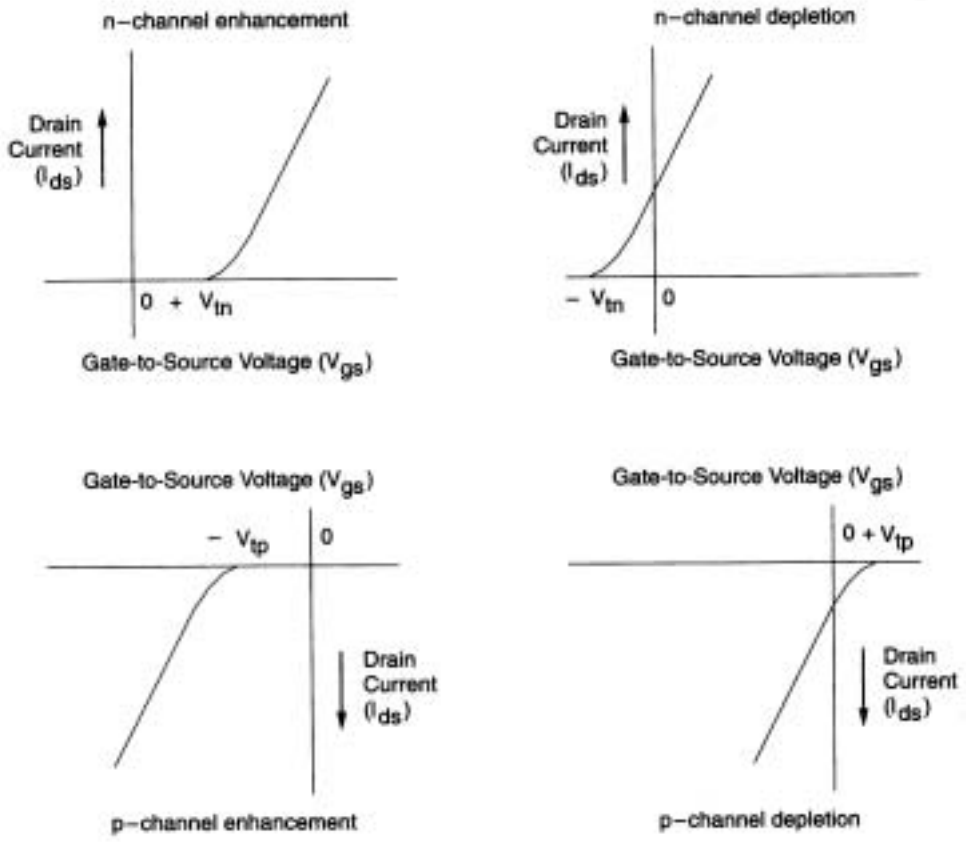
- PMOS (p-type MOS transistor)

(1) Majority carrier = holes

(2) Applied voltage is negative with respect to substrate.

- Threshold voltage (*Vt*):
  The voltage at which an MOS device begins to conduct ("turn on")

- Relationship between **Vgs** (gate-to-source voltage) and the source-to-drain current (Ids) , given a fixed drain-to-source voltage (Vds).

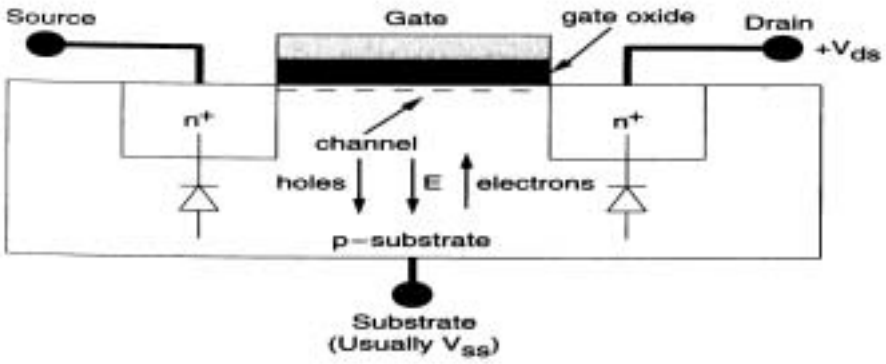


(1) Devices that are normally cut-off with zero gate bias are classified as "enhancement-mode "devices.

(2) Devices that conduct with zero gate bias are called "depletion-mode "devices.

(3) Enhancement-mode devices are more popular in practical use.

2.1.1 NMOS Enhancement Transistor



- Consist of

(1) Moderately doped p-type silicon substrate
(2) Two heavily doped n^+ regions, the source and drain, are diffused.
(3) Channel is covered by a thin insulating layer of silicon dioxide (SiO2) called " Gate Oxide "
(4) Over the oxide is a polycrystalline silicon (polysilicon) electrode, referred to as the "Gate"

- Features

(1) Since the oxide layer is an insulator, the DC current from the gate to channel is essentially zero.

(2) No physical distinction between the drain and source regions.

(3) Since SiO2 has low loss and high dielectric strength, the application of high gate fields is feasible.

- In operation

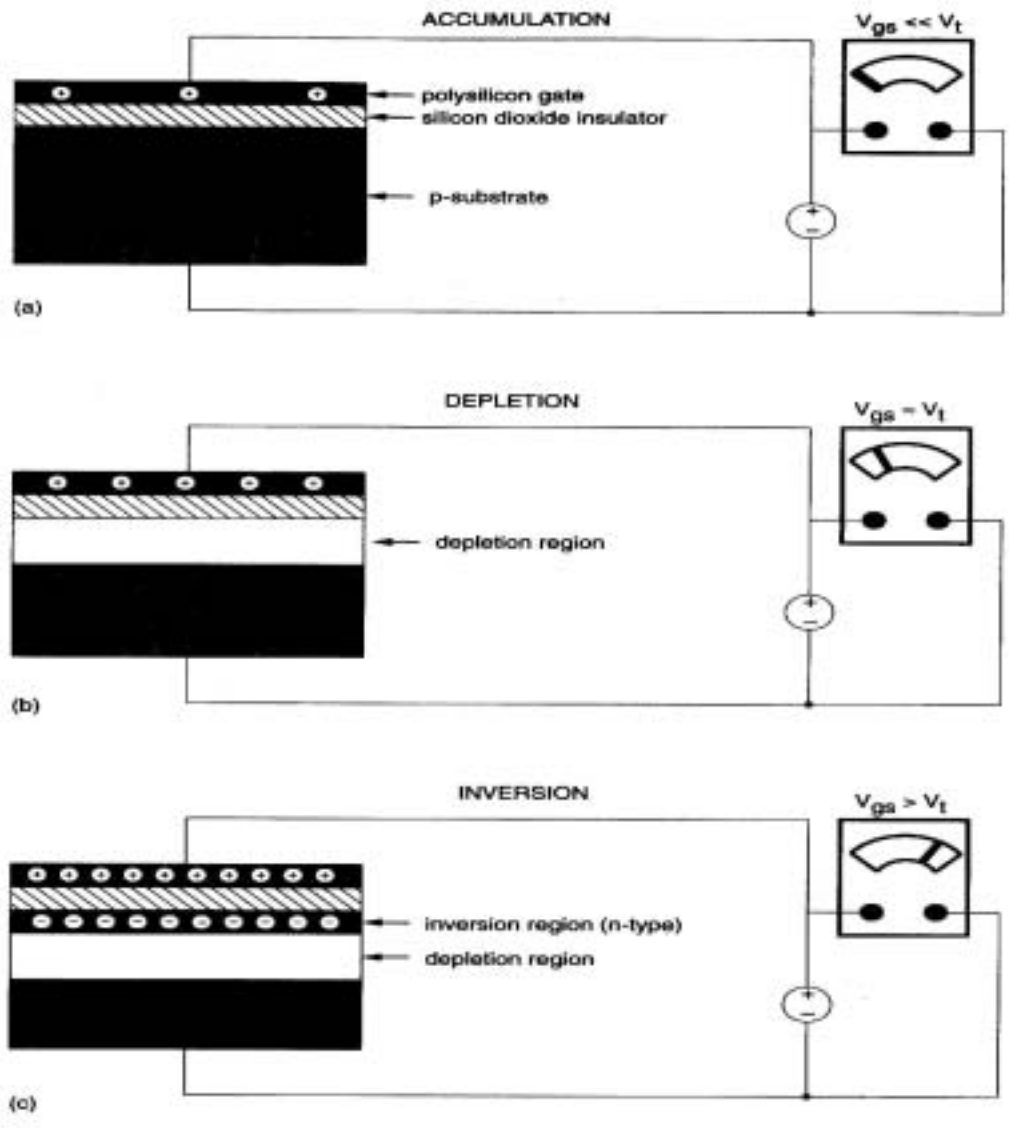
(1) Set Vds > 0 in operation

(2) Vgs =0 → no current flow between source and drain. They are insulated by two reversed-biased PN junctions (see Fig 2.3).

(3) When Vg > 0 , the produced E field attracts electrons toward the gate and repels holes.

(4) If Vg is sufficiently large, the region under the gate changes from p-type to n-type(due to accumulation of attracted elections) and provides a conducting path between source and drain.←→The thin layer of p-type silicon is said to be "inverted".

(5) Three modes (see Fig 2.4)
   a. Accumulation mode (Vgs << Vt)
   b. Depletion mode (Vgs =Vt)
   c. Inversion mode (Vgs > Vt)



- Electrically

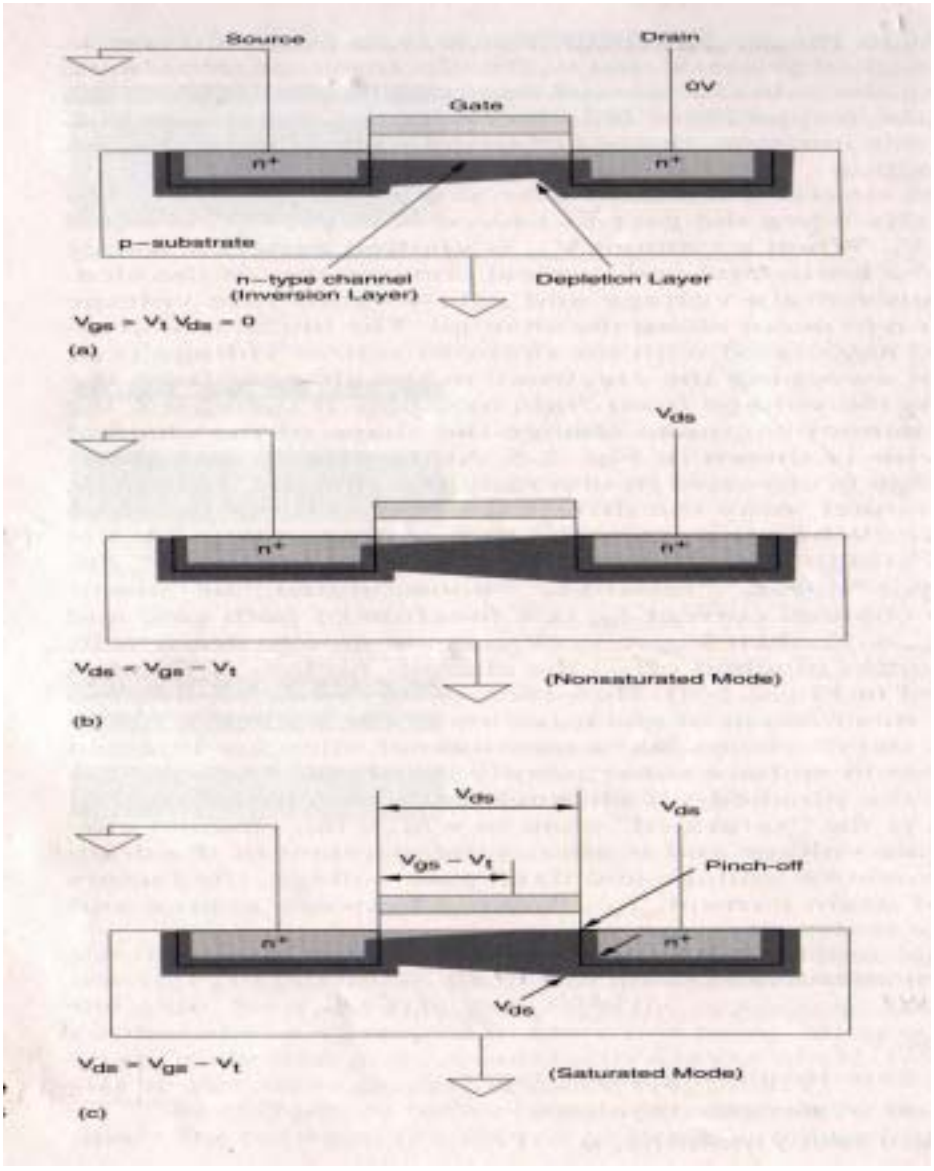
(1) An MOS device can be considered as a voltage-controlled switch that conducts when Vgs >Vt (given Vds>0)

(2) An MOS device can be considered as a voltage-controlled resistor (See Fig 2.5)

- Effective gate voltage (Vgs-Vt)

- At the source end , the full gate voltage is effective in the inverting the channel.

- At the drain end , only the difference between the gate and drain voltage is effective



- Pinch-off

(1) Vds > Vgs-Vt => Vgd < Vt => Vd > Vg –Vt (Vg is not big enough)

(2) The channel no longer reaches the drain. (Fig 2.5 c)

(3) As electrons leave the drain depletion region and are subsequently accelerated toward the drain.

(4) The voltage across the pinched-off region remains at (Vgs-Vt) =>"saturated" state in which the channel current as controlled by Vg , and is independent of Vd

- For fixed Vds and Vg , Ids is function of

(1) Distance between drain & source
(2) Channel width
(3) Vt
(4) Thickness of gate oxide
(5) The dielectric constant of gate oxide
(6) Carrier (hole or electron) mobility , μ .

- Conducting mode

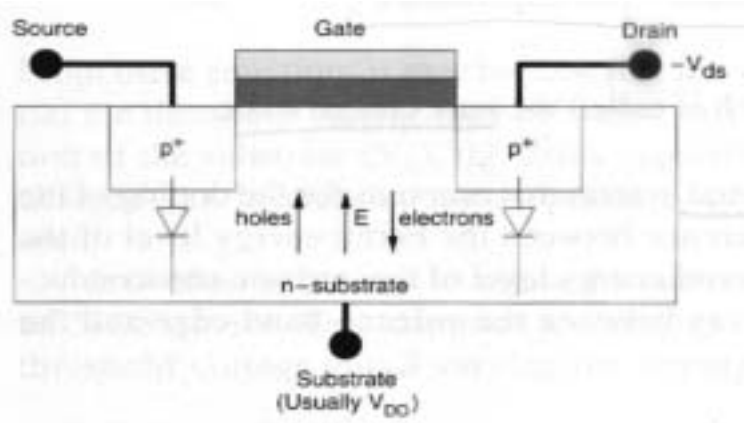
(1) "cut-off " region : Ids $\approx$ 0 , Vgs < Vt
(2) " Nonsaturated" region : weak inversion region, when Ids depends on Vg & Vd
(3) "Saturated" region: channel is strongly inverted and Ids is ideally independent of Vds (pinch-off region)
(4) "Avalanche breakdown" (pinch-through) : very high Vd => gate has no control over Ids

2.1.2 PMOS Enhancement Transistor

(1) Vg < 0

(2) Holes are major carrier

(3) Vd < 0 , which sweeps holes from the source through the channel to the drain .



2.1.3 Threshold voltage

- A function of

(1) Gate conductor material
(2) Gate insulator material
(3) Gate insulator thickness
(4) Impurity at the silicon-insulator interface
(5) Voltage between the source and the substrate Vsb
(6) Temperature
    a. -4 mV/'C – high substrate doping
    b. -2 mV/'C – low substrate doping

2.2 MOS equations

2.2.1 Basic DC equations

- Three MOS operating regions

(1) Cutoff or subthreshold region

$I_{ds}=0$, $V_{gs} \le V_t$

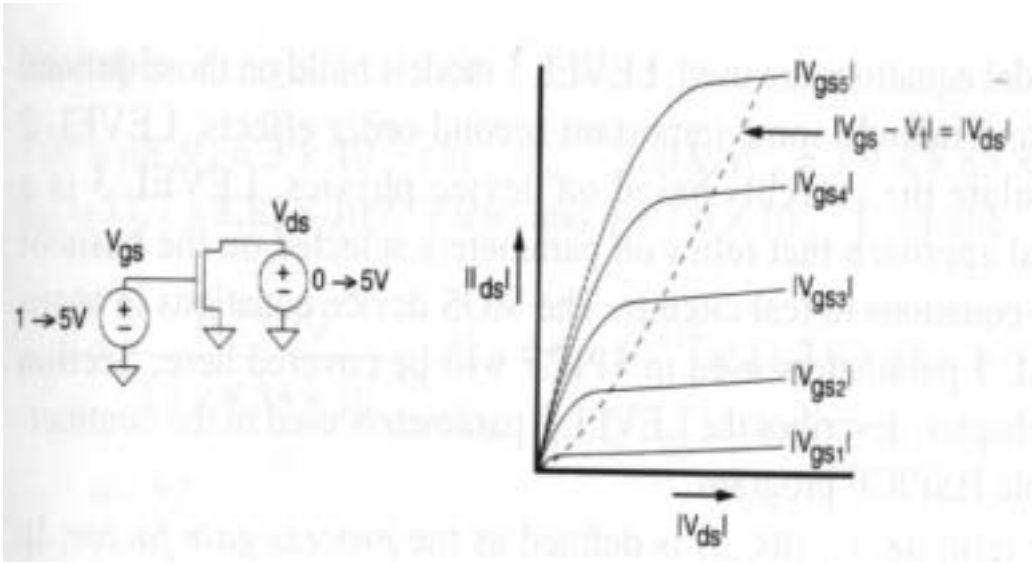
(2) Nonsaturation, linear or triode region

$$I_{ds} = \beta\left[\left(V_{gs} - V_t\right)V_{ds} - \frac{V_{ds}^{\ 2}}{2}\right] \quad 0<Vgs<Vgs-Vt$$

$$\approx \beta\left[V_{gs} - V_t\right]V_{ds} \quad \text{When Vds << Vgs-Vt}$$

(3) Saturation region

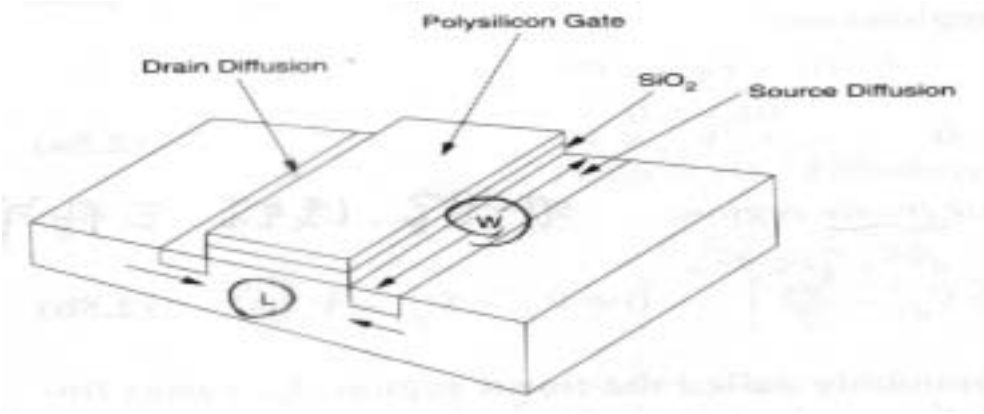
$$I_{ds} = \beta\frac{\left(V_{gs} - V_t\right)^2}{2} \quad , \quad 0< Vgs-Vt<Vds$$



- Vd at which the device becomes saturated is called Vdsat (drain saturation voltage)

- β : MOS transistor gain factor

Function of (1) process parameter (2) device geometry



(1) μ = effective mobility of the carrier in the channel

(2) ε = permittivity of the gate oxide

(3) t_{ox} = thickness of the gate oxide

Note: $\frac{\varepsilon}{t_{ox}} = C_{ox}$ => $\beta = \mu C_{ox}\left(\frac{W}{L}\right)$

- Example

Typical CMOS

◎(~1 μ) process

(1) μ_n=500 cm^2/V-sec

(2) ε =3.9 ε_0 =3.9*8.85*10^{-14} F/cm (permittivity of SiO_2)

(3) t_{ox}=200 $\mathring{A}$

$$\beta_n = \frac{\mu\varepsilon}{t_{ox}}\left(\frac{W}{L}\right) = 88.5\frac{W}{L}\mu A/V^2$$

◎ $\mu_p = 180\, cm^2/V-\sec \Rightarrow \beta_p = 31.9\frac{W}{L}\ \mu A/V^2$

◎ $\frac{\beta_N}{\beta_p} = 2.8$ (2~3 depending on process)

2.2.2. Seven Second-order Effect

- SPICE : Simulation Program with Integrated Circuit Emphasis

- LEVEL: 1,2,3

(1) Basic DC Equations + Some second-order effects
(2) Based on device physics
(3) Add more parameters to match real circuits

e.g., Process gain factor

SPICE : Kp (10-100 μ A/V^2 with 10%-20% variation)

A. Channel-length modulation

- When an MOS device is in saturation.
- $L_{eff} = L - L_{short}$

$$L_{short} = \sqrt{2\frac{\varepsilon_{si}}{qN_A}(Vds-(Vgs-Vt))}$$

=>L ↓ => β ↑ => Ids ↑

$$I_{ds} = \frac{K}{2}\left(\frac{W}{L}\right)(V_{gs}-V_t)^2(1+\lambda V_{ds})$$

With $K = \frac{\mu\varepsilon}{t_{ox}}$: process gain factor

λ :channel length modulation factor ($0.02V^{-1}$ to $0.005\ V^{-1}$)

(In SPICE level 1 : λ =LAMBDA)

B. Drain punchthrough (avalanche breakdown)

V_D is very high , Ids is independent of Vgs
Good for I/O protection circuit.

C. Threshold voltage (Vt) – Body effect (Vsb)

- $Vt = V_{fb} + 2\phi_b + \frac{\sqrt{2\varepsilon_{si} q N_A (2\phi_b + |V_{SB}|)}}{C_{ox}}$

$=> V_t = V_{t0} + \gamma\left[\sqrt{2\phi_b + |V_{SB}|} - \sqrt{2\phi_b}\right]$

(1) Vsb : Substrate bias
(2) Vt0 : Vt at Vsb=0

(3) γ : a constant which describes the substrate bias effect

(range:0.4~1.2) $\gamma = \frac{t_{ox}}{\varepsilon_{ox}}\sqrt{2q\varepsilon_{si}N_A} = \frac{1}{C_{ox}}\sqrt{2q\varepsilon_{si}N_A}$

(4) SPICE

- γ : GAMMA in SPICE model
- Vto : VT0
- N_A : NSUB
- $\psi_s = 2\psi_b$: PHI (the surface potential at the onset of strong inversion)

Subthreshold region

- Cut-off = subthreshold region
- Ids ≈ 0 (Subthreshold region)
- But the finite value of Ids may be used to construct very low power circuits.
- In Level 1 SPICE , subthreshold current is set 0

Others:

- **Mobility variation**
- **Fowler-Nordheim Tunneling**
- **Impact Ionization (Hot electrons effect)**

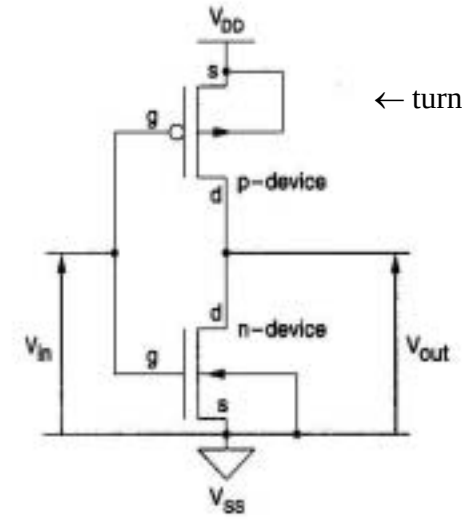
2.2.3 MOS Models

- MOS model = Ideal Equations + Second-order Effects + Additional Curve-fitting parameters

- Many semiconductor vendors expend a lot of effects to model the devices they manufacture.(Standard : Level 3 SPICE)

- Main SPICE DC parameters in level 1,2,3 in 1 μ n-well CMOS process.

2.3 CMOS inverter DC characteristics



$\leftarrow$ turn on $\rightarrow$

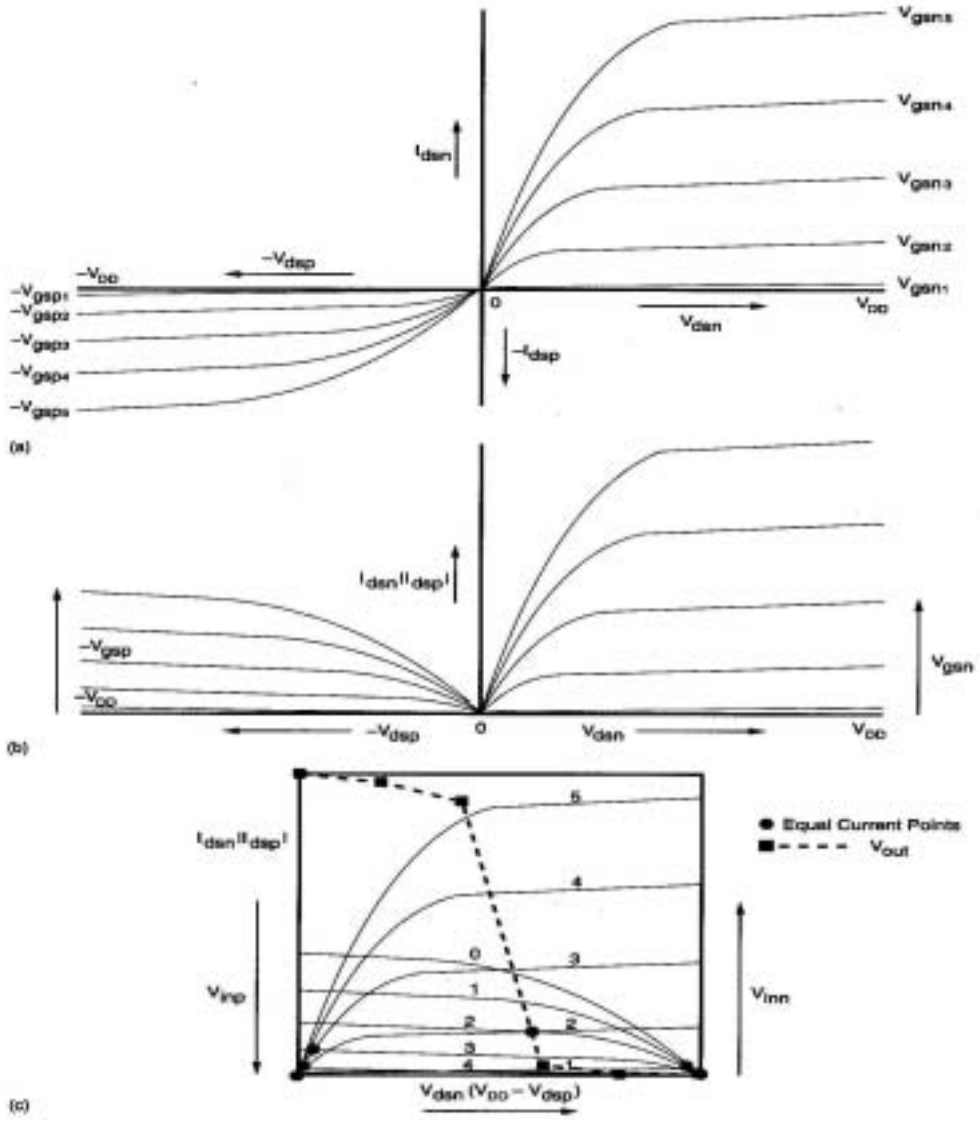
$$V_{gs} = V_g - V_{DD} < -|V_{tp}|$$

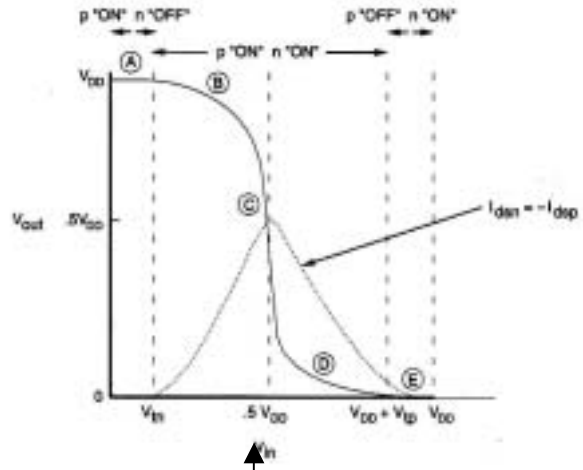
$$\begin{cases} V_{gs} = V_{in} \\ V_{ds} = V_{out} \end{cases}$$

$$\Rightarrow V_g < V_{DD} - |V_{tp}|$$

$$\Rightarrow V_{in} < V_{DD} - |V_{tp}|$$

(check Fig. 2.12)





Both transistors are “on”

$$P = fcv^2 \cdot \alpha$$

(Switching activity)

- Solve for $\begin{cases} I_{dsn} = -I_{dsp} \\ V_{inn} = V_{inp} \end{cases}$

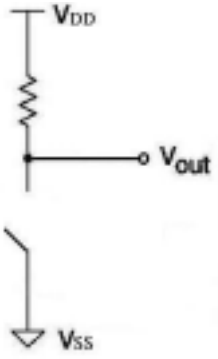
(1) Region A. $0 \le V_{in} \le V_{tn}$

$\begin{cases}$ n-device is ‘ off ’, $I_{dsn} = 0 \ (= -I_{dsp})$

p-device is in ‘linear’ mode

$V_{out} - V_{DD} = \underline{V_{dsp} = 0}$

$\Rightarrow V_{out} = V_{DD}$



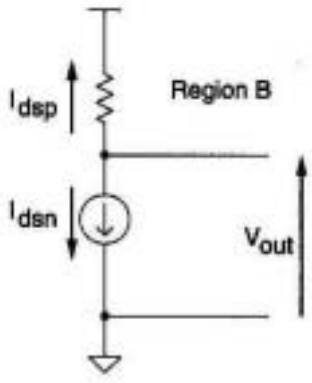
(2) Region B. $V_{tn} \le V_{in} \le V_{DD}/2$

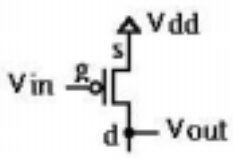
$\begin{cases}$ p-device : linear mode

n-device : saturation mode

n : $I_{dsn} = \beta_n \dfrac{[V_{in} - V_{tn}]^2}{2}$, $\beta_n = \dfrac{\mu_n \varepsilon}{t_{ox}} (\dfrac{W_n}{L_n})$

p : $\begin{cases} V_{gs} = V_{in} - V_{DD} \\ V_{ds} = V_{out} - V_{DD} \end{cases}$





$$\therefore \quad I_{dsp} = -\beta_p [\underbrace{(V_{in} - V_{DD} - V_{tp})}_{V_{gs} - V_{tp}} \underbrace{(V_{out} - V_{DD})}_{V_{ds}} - \frac{(V_{out} - V_{DD})^2}{2}]$$

with $\beta_p = \frac{\mu_p \varepsilon}{t_{ox}} (\frac{W_p}{L_p})$

solve for $I_{dsp} = -I_{dsn}$

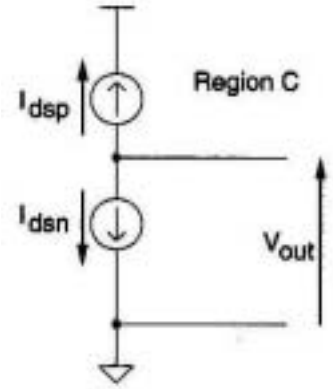
$$\Rightarrow V_{out} = (V_{in} - V_{tp}) + \sqrt{(V_{in} - V_{tp})^2 - 2(V_{in} - \frac{V_{DD}}{2} - V_{tp})V_{DD} - \frac{\beta_n}{\beta_p}(V_{in} - V_{tn})^2}$$

(3) Region C. PMOS, NMOS : saturation

$$I_{dsp} = -\frac{\beta_p}{2}(V_{in} - V_{DD} - V_{tp})^2$$

$$I_{dsn} = \frac{\beta_n}{2}(V_{in} - V_{tn})^2$$

with $I_{dsp} = -I_{dsn}$



$$\Rightarrow V_{in} = \frac{V_{DD} + V_{tp} + V_{tn}\sqrt{\frac{\beta_n}{\beta_p}}}{1 + \sqrt{\frac{\beta_n}{\beta_p}}}$$

$\Rightarrow$ by setting $\beta_n = \beta_p$ and $V_{tn} = -V_{tp}$

we have $\boxed{V_{in} = \frac{V_{DD}}{2}}$: one value only

possible V_{out}

N-MOS $\begin{cases} V_{in} - V_{tn} < V_{out} \\ V_{gs} - V_{tn} < V_{ds} \end{cases}$

P-MOS $\begin{cases} (V_{DD} - V_{in}) - V_{tp} > (V_{DD} - V_{out}) \\ V_{gs} - V_{tp} > V_{ds} \end{cases}$

$\Rightarrow V_{out} < V_{in} - V_{tp}$

$\Rightarrow$ ☆ $\boxed{V_{in} - V_{tn} < V_{out} < V_{in} - V_{tp}}$

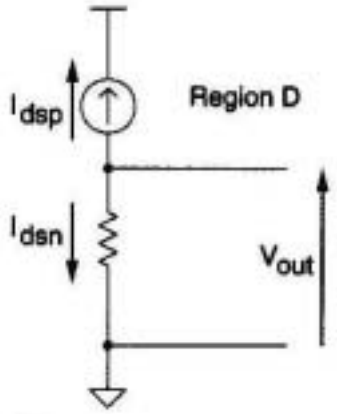
Negative value

$\Rightarrow$ V_{in} is fixed at $\frac{V_{DD}}{2}$, V_{out} varies

$\Rightarrow$ make the o/p transition very steep

(4) Region D. $V_{DD}/2 < V_{in} < V_{DD} + V_{tp}$

$\begin{cases}$ P-MOS : saturation mode
N-MOS : linear mode



$$I_{dsp} = -\frac{1}{2}\beta_p (V_{in} - V_{DD} - V_{tp})^2$$

$$I_{dsn} = \beta_n [(V_{in} - V_{tn})V_{out} - \frac{V_{out}^{\ 2}}{2}]$$

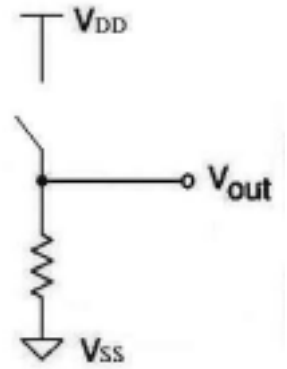
solve $I_{dsn} = -I_{dsp}$

$$\Rightarrow V_{out} = (V_{in} - V_{tn}) - \sqrt{(V_{in} - V_{tn})^2 - \frac{\beta_p}{\beta_n}(V_{in} - V_{DD} - V_{tp})^2}$$

(5) Region E. $V_{in} \geq V_{DD} + V_{tp}$

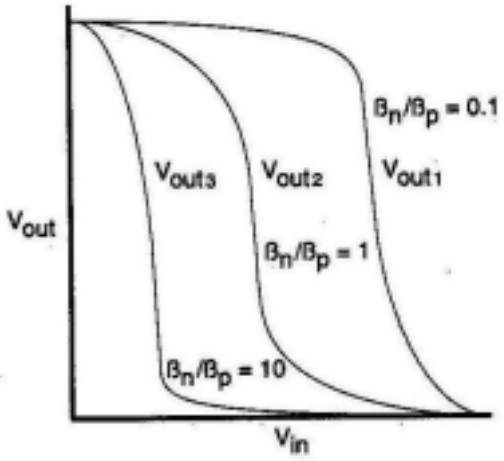
$\rightarrow$ p-device ‘ off ‘ (n-device is in ‘ linear ’ mode)

$\rightarrow$ $I_{dsp} = 0 \Rightarrow I_{dsn} = 0 \Rightarrow V_{out} = 0$



see Table 2.3 for summary

2.3.1 β_n / β_p ratio (watch Eq.(2.24))



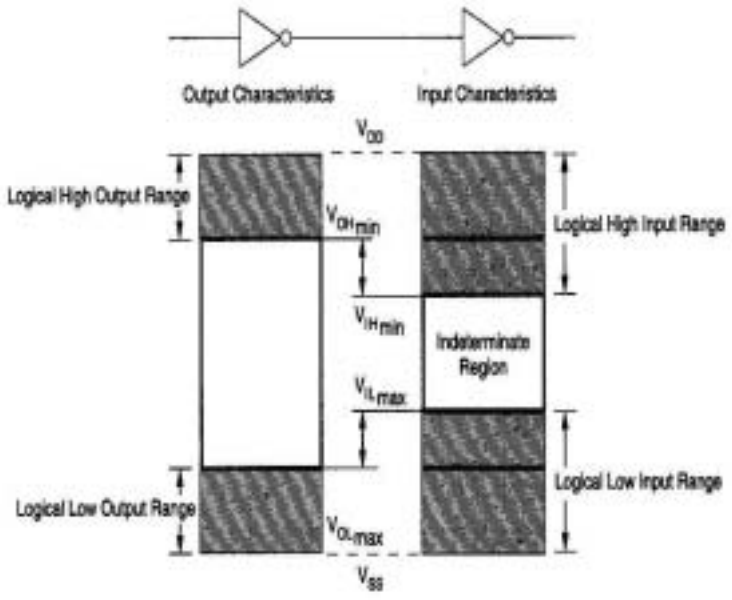
$$\beta_n / \beta_p \leftrightarrow \frac{W_n / L_n}{W_p / L_p}$$

- Note $\beta \propto T^{-1.5}$ ($\because$ T ↑ , μ ↓)

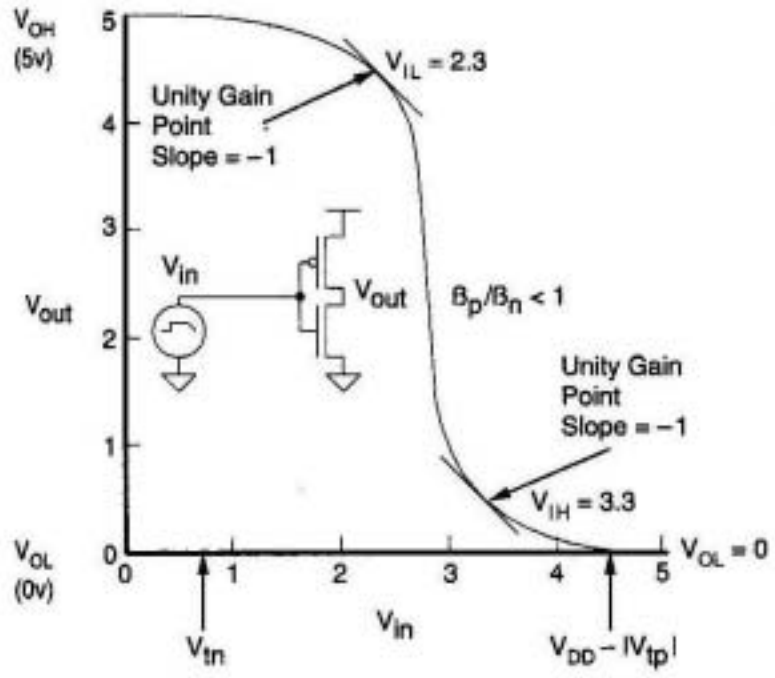
$$\Rightarrow \boxed{I_{ds} \propto T^{-1.5}}$$

2.3.2 Noise Margin

- This parameter allows us to determine the allowable noise voltage on the input of a gate so that the output will not be affected.



$$\begin{cases} NM_H = |V_{OH\min} - V_{IH\min}| \\ NM_L = |V_{IL\max} - V_{OL\max}| \end{cases}$$



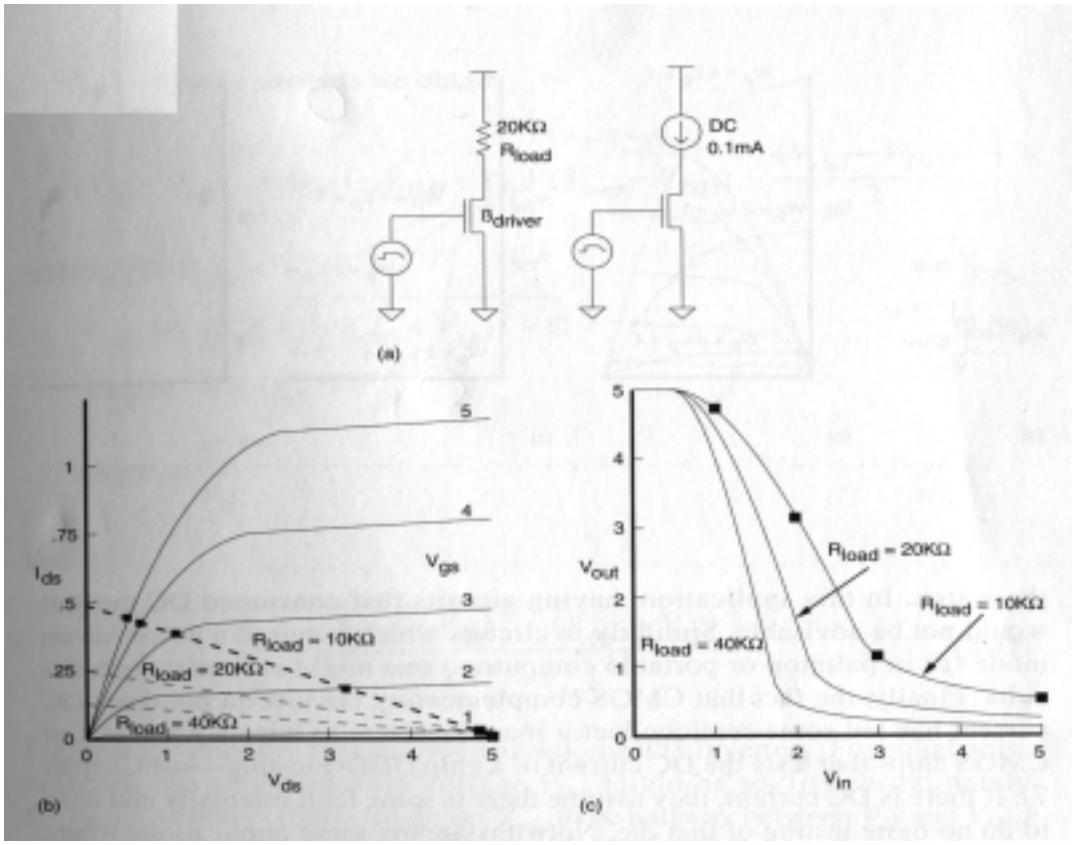
$$\begin{cases} V_{IL} = 2.3 \\ V_{IH} = 3.3 \end{cases}$$

$$\begin{cases} V_{OL} \\ V_{OH} \end{cases}$$ are more difficult (will be discussed later)

(left as your exercise)

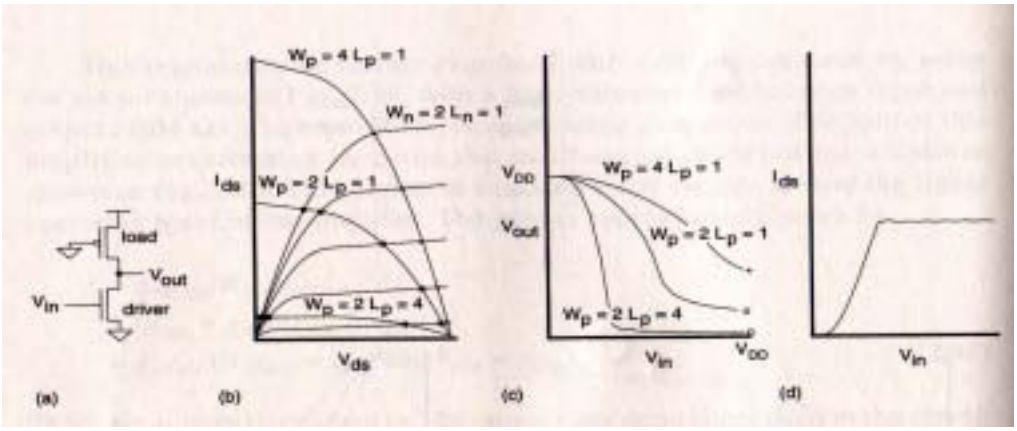
2.4 Static Load MOS inverters

- Resistor-load inverter
  Current-source-load inverter

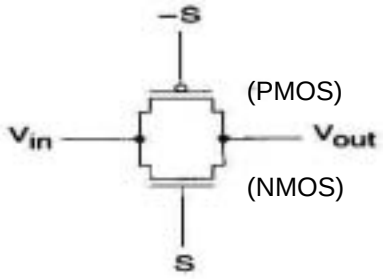


2.4.1 Pseudo-NMOS inverter

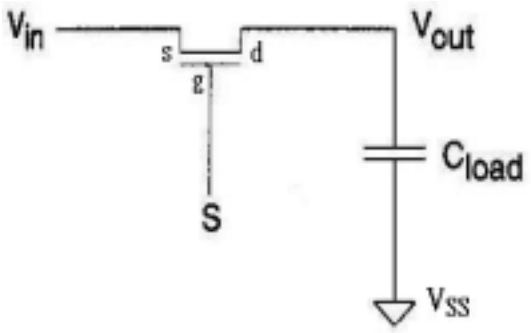
- Fast (constant current)
- power-consuming P↑ ⇒ ∵I↑ ⇒ but speed↑



2.6 The Transmission Gate



- NMOS pass transistor



C_{load} is initially discharged

$V_{out} = V_{SS}$

with S=0 (V_{SS})

$V_{gs} = 0V$

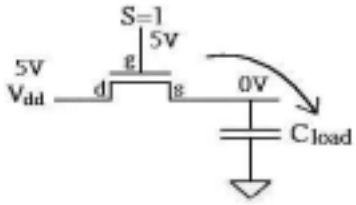
$I_{ds} = 0$

V_{out} remains at V_{SS}

$V_{gs} = 5V > V_t$, NMOS On $\Rightarrow$ $V_{in} = V_{out} = 0V$

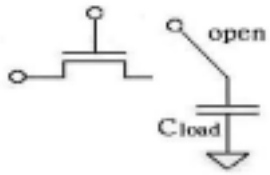
S=1 (V_{dd})

$V_{gs} = V_{DD}$ (initially)



$V_{gs} = 5V > V_t \Rightarrow$ charge

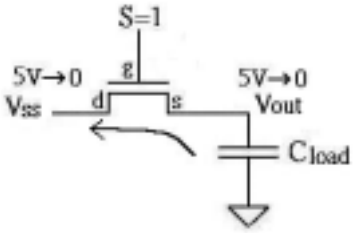
- $\because V_{in} > V_{out}$, current from V_{in} to V_{out}
- As the output voltage approaches $\boxed{V_{DD} - V_{tn}}$, the n-device begins to <u>turn off</u>



- S=0 (open circuit) , V_{out} remains at $V_{DD} - V_{tn}(V_{dd})$,

  where $V_{tn}(V_{dd})$ denotes the V_t at $V_s = V_{dd}$ '(body effect)'

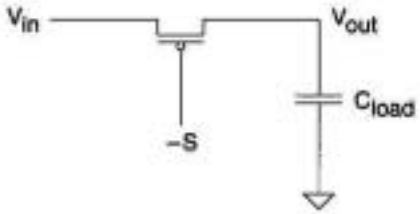
$$\begin{cases} \text{S=1} \\ V_{in} = 0 \\ V_{out} = V_{DD} - V_{tn}(V_{dd}) \end{cases}$$



n-device begin to conduct , and V_{out} fall to V_{ss}

$$\Rightarrow \begin{cases} \text{Transmission of Logic 1 is degraded, } (V_{DD} - V_{tn}) \\ \text{Transmission of Logic 0 is not degraded, } (V_{ss}) \end{cases}$$

- PMOS pass transistor



- S=1 ,(S=0) (open)

  $V_{in} = V_{ss}$

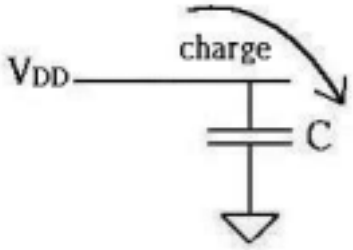
  $V_{out} = V_{ss}$

- S=0 (close)

  $V_{in} = V_{DD}$, current

  V_{out} to V_{DD}

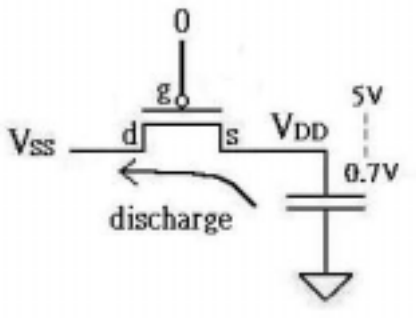
  charge (C_load)



- S=0 (close)

  $V_{in} = V_{ss}$,

  $V_{out} = V_{DD}$

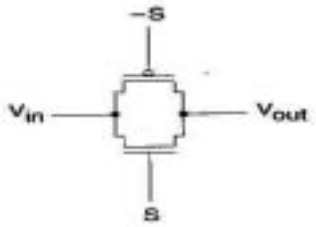


- Discharge C_load
  until through p-device

  until $V_{out} = V_{tp}\ (V_{ss})$, at which point the transistor stops conducting

  $\Rightarrow$ p-MOS passes good ‘1’
  p-MOS passes poor ‘0’

- Transmission gate can pass logic ‘1’ and ‘0’ <u>without</u> degradation !
- overall behavior



(1) S=0 $(\overline{S} = 1)$:

N,P devices are ‘OFF’

$V_{in} = V_{SS}$, $V_{out} = Z$ (high impedance)

$V_{in} = V_{DD}$, $V_{out} = Z$

(2) S=1 $(\overline{S} = 0)$:

N,P devices are ‘ON’

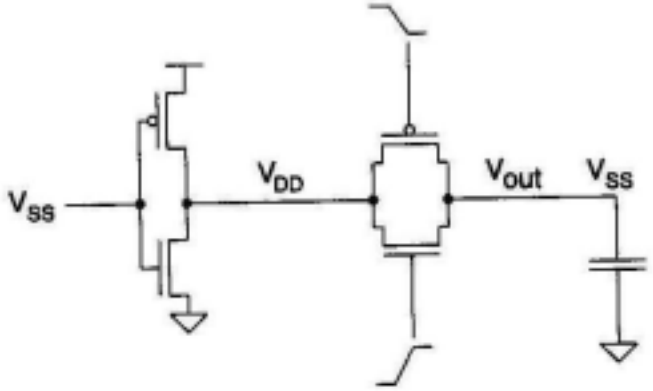
$V_{in} = V_{SS}$, $V_{out} = V_{SS}$, $V_{in} = V_{DD}$, $V_{out} = V_{DD}$

- Used in <u>multiplexing element</u> & <u>latch element</u> act as voltage-controlled <u>resistor</u> connecting the input and output

- Example to analyze a CMOS circuit

Way 1=MAN
Way 2=SPICE

(1) Capacitor loaded circuit



Cload at Vss

* Cload is large

- When S is ON (NMOS , S = 0 1

  PMOS , $\overline{S}$ = 1 0)

- Results: Currents of the pass transistor are monitored

  Vout ↑ (transmission gate) , $V_{gs}(p) = -5$ (constant current)

  (PMOS)
  It starts at 'saturation' ⟶ 'nonsaturation'

  as $\left|V_{gsp} - V_{tp}\right| > \left|V_{dsp}\right|$

  Vout ↑ (transmission gate)

  (NMOS)

  always at 'saturation' , ∵ $V_{dsn} = V_{gsn}$

  $$\Rightarrow V_{gsn} - V_{tn} < V_{dsn}$$

  { Rise

  After $V_{out} \rightarrow V_{DD} - V_{tn}$, NMOS is 'off'

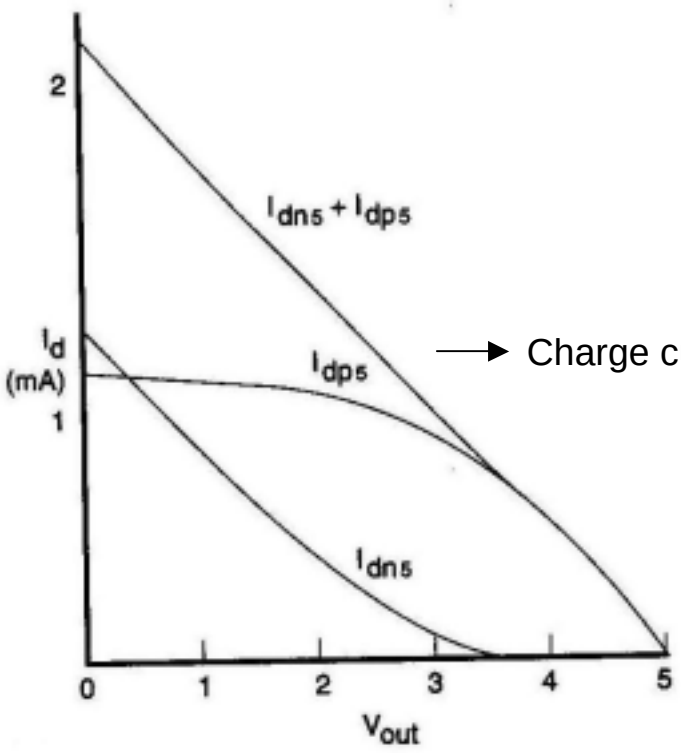
■ Three regions of operation :

A	N saturation	P saturation	$V_{out} < \lvert V_{tp} \rvert$
B	N saturation	P nonsaturation	$\lvert V_{tp} \rvert < V_{out} < V_{DD} - V_{tn}$
C	N off	P nonsaturation	$V_{DD} - V_{tn} < V_{out}$

a. Region A
   - p : constant current source
   - n : current varies inversely with Vout

b. Region B
   both currents vary linearly (inverse) with Vout

c. Region C
   p-current varies inverse linearly with Vout



→ Charge current amount

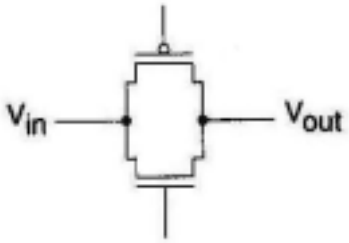
check : $V_{in} = V_{SS}$, $V_{out} = V_{DD} \rightarrow V_{SS}$

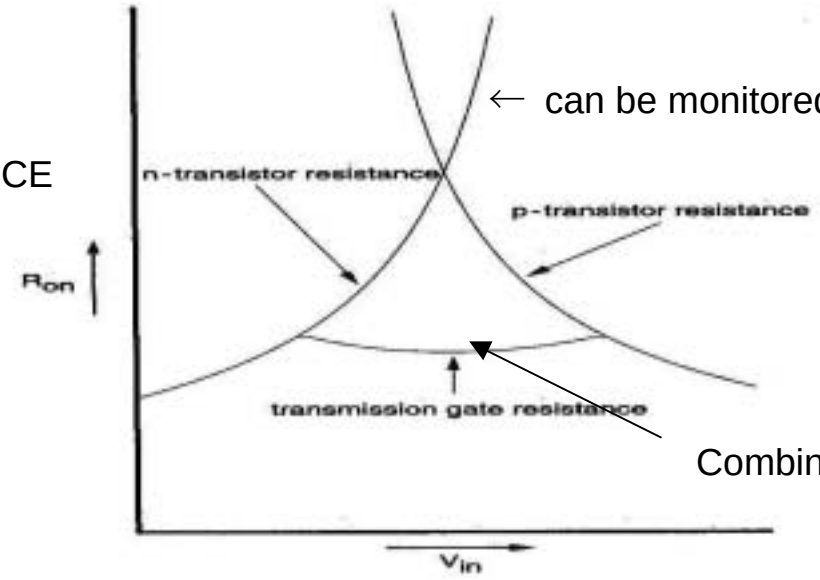
(2) Lightly loaded circuit (Cload is small)

- Vout follows Vin very closely

- Fig 2.35(d) { n-current for $V_{out} - V_{in} = -0.1$; p-current

- Three regions of operation :
  a. n (linear) , p (off)
  b. n (linear) , p (linear)
  c. n (off) , p (linear)





← can be monitored by using SPICE