



ECE-2221

VLSI Design

Dr. Arjun Sunil Rao
arjun.rao@manipal.edu



Course Outcome

CO1: Explain the fundamental principles and technologies used in MOS digital integrated circuit design.

CO2: Demonstrate the application of MOSFET to create logic gates and circuits and explain delay models.

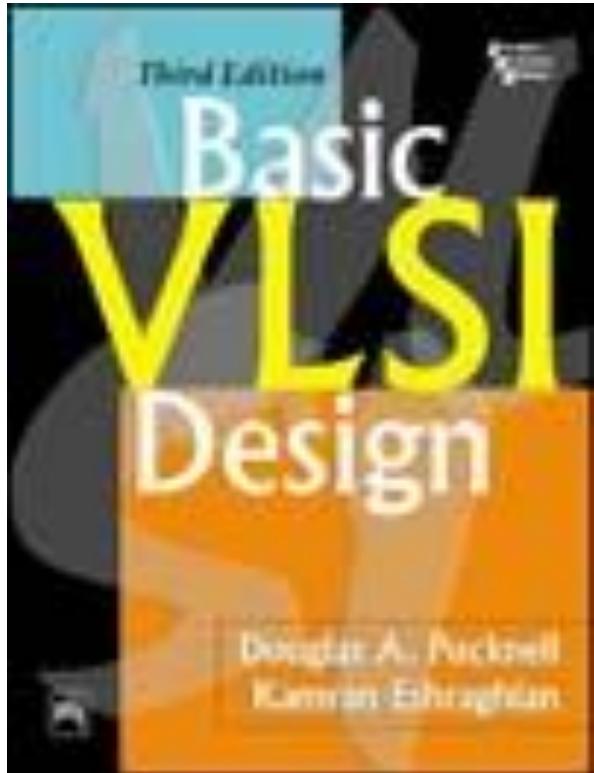
CO3: Illustrate various IC fabrication techniques with stick diagrams and layouts.

CO4: Analyze and design MOS based subsystem design.

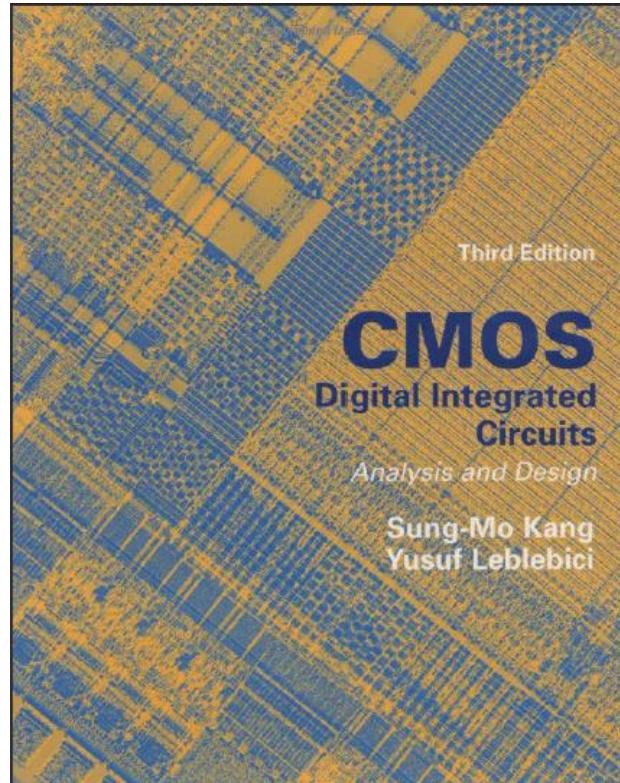
CO5: Analyze and design various memory structures.



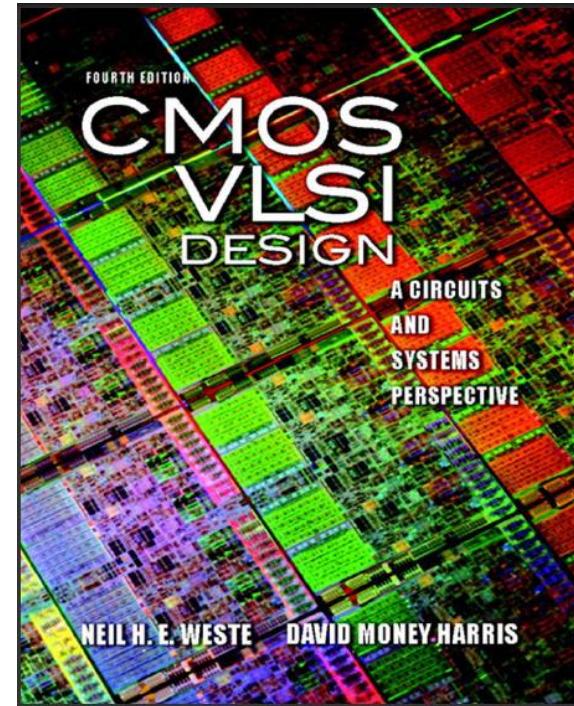
Reference Books



"Basic VLSI Design", 3rd Ed., PHI
By: PUCKNELL DOUGLAS
A.ESHRAGHIAN,
KAMRAN



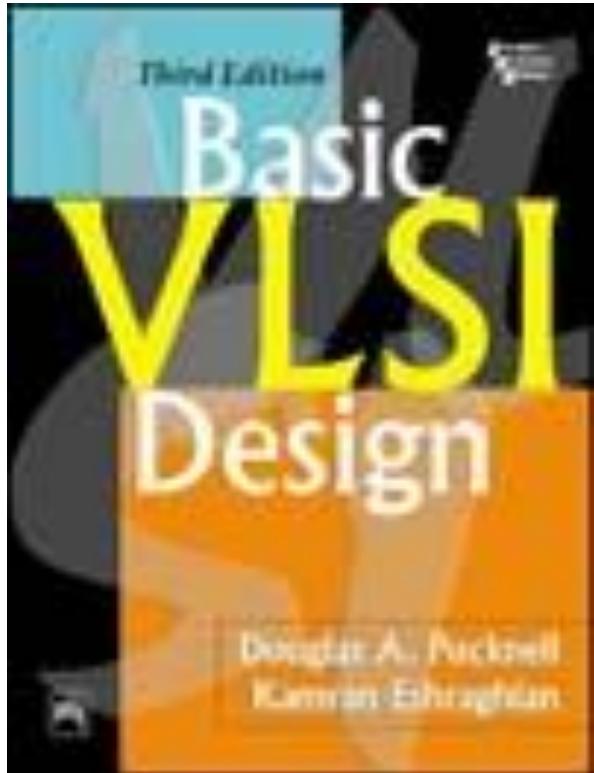
"CMOS DIGITAL INTEGRATED CIRCUITS:Analysis and Design"
3rd Ed. McGraw-Hill.
By:SUNG-MO (STEVE) KANG



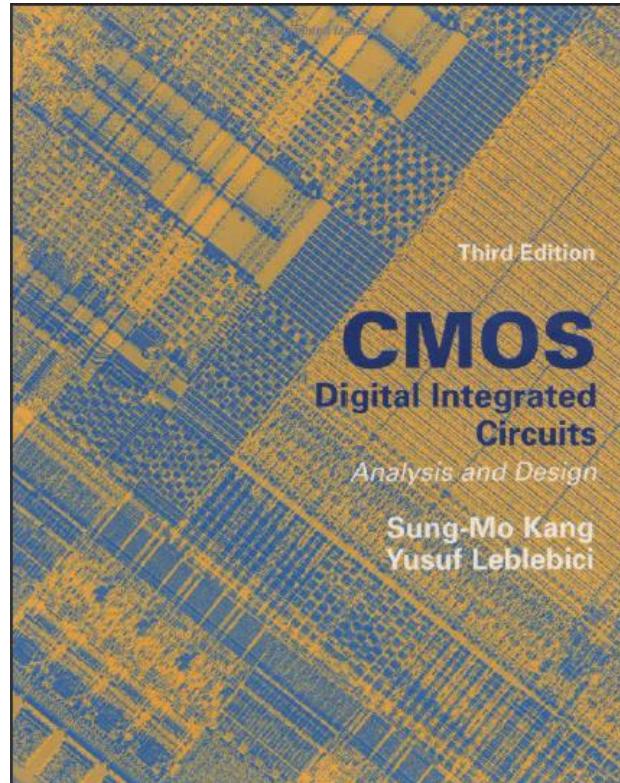
"CMOS VLSI Design:A Circuits and Systems Perspective" 4th ed. Addison-Wesley
By: Neil H. E. Weste, David Money Harris



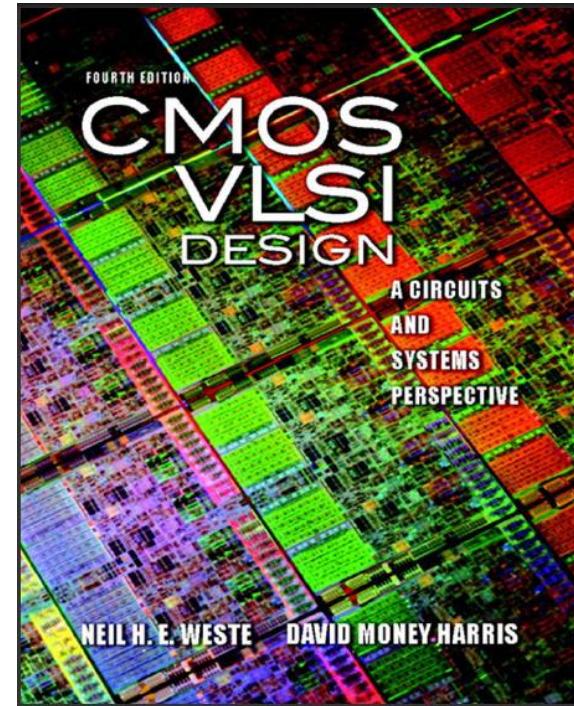
Reference Books



"Basic VLSI Design", 3rd Ed., PHI
By: PUCKNELL DOUGLAS
A.ESHRAGHIAN,
KAMRAN



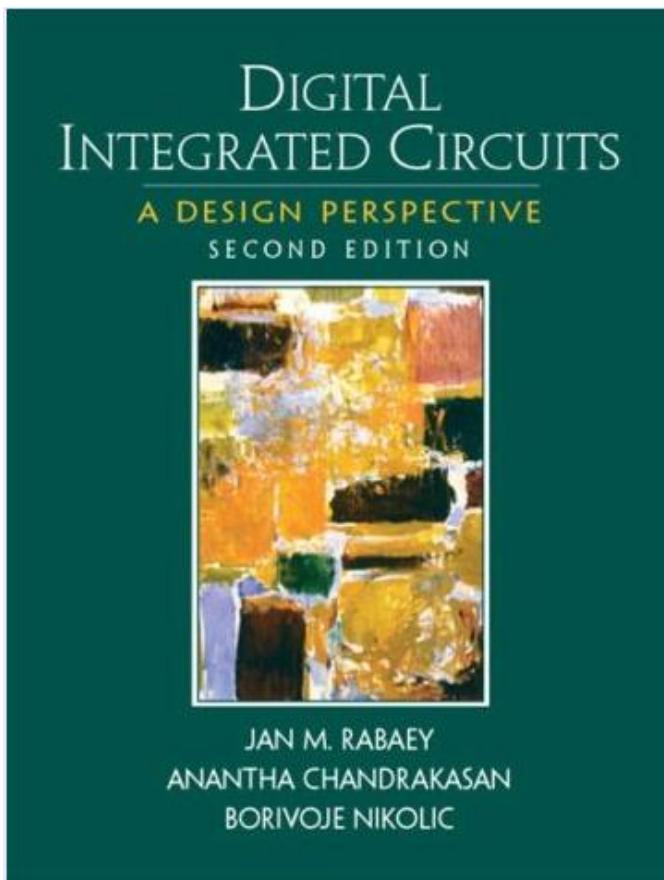
"CMOS DIGITAL INTEGRATED CIRCUITS:Analysis and Design"
3rd Ed. McGraw-Hill.
By:SUNG-MO (STEVE) KANG



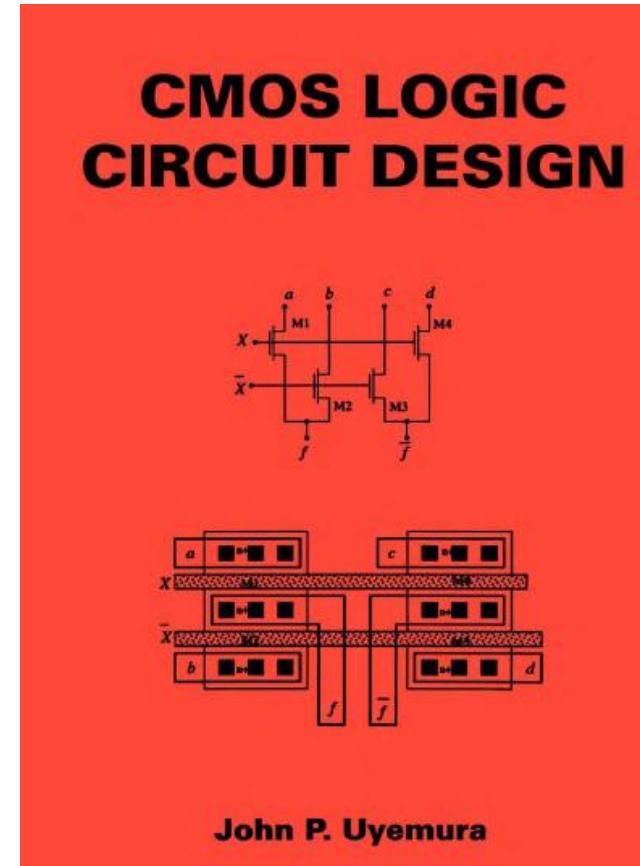
"CMOS VLSI Design:A Circuits and Systems Perspective" 4th ed. Addison-Wesley
By: Neil H. E. Weste, David Money Harris



Reference Books



**Digital Integrated Circuits – A Design Perspective”,
2nd ed. by J. Rabaey, A. Chandrakasan, B. Nikolic**



**“CMOS LOGIC CIRCUIT DESIGN”
KLUWER ACADEMIC PUBLISHERS
By: John P. Uyemura**



Introduction

What is expected ?

Complexity → ∞

Intel 4004 Processor: 2300 Transistor → 10µm Technology
Intel i7 Processor : 3,200,000,000 Transistors → 14nm Technology

Cost → Min

Intel 4004 Processor
Intel i7 Processor

Power → Min

Intel 4004 Processor
Intel i7 Processor

Size → Min

Intel 4004 Processor
Intel i7 Processor

Delay → Min

Intel 4004 Processor: 740KHz
Intel i7 Processor : 3.6 GHz



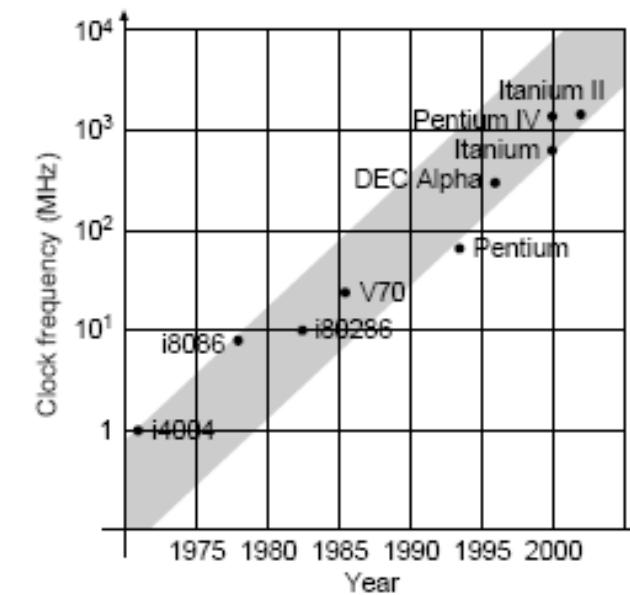
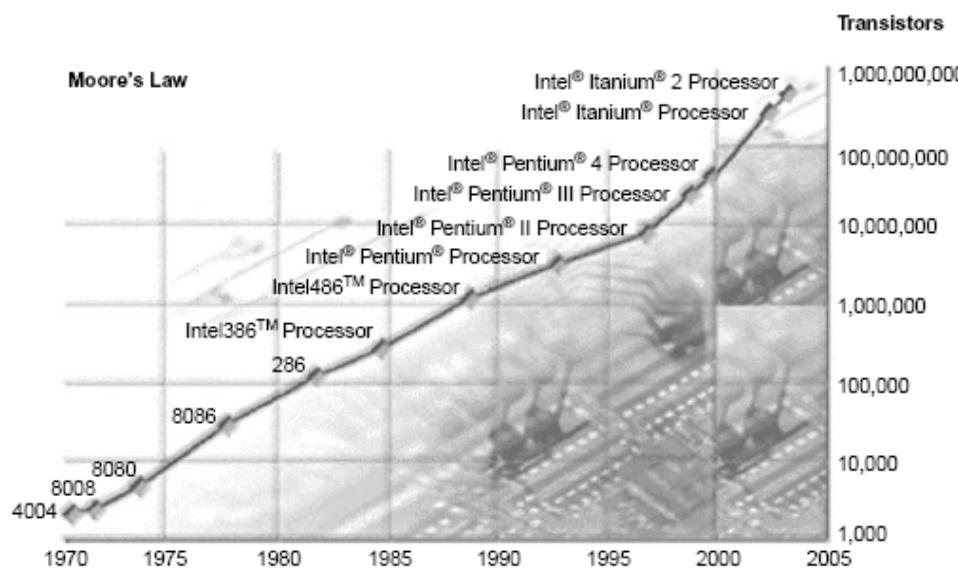
Technology generations

Integration level	Year	No. of transistors
SSI	1950s	Less than 10^2
MSI	1960s	$10^2 \approx 10^3$
LSI	1970s	$10^3 \approx 10^5$
VLSI	1980s	$10^5 \approx 10^7$
ULSI	1990s	$10^7 \approx 10^9$
SLSI	2000s	Over 10^9



Moore's Law

The evolution of MOS technology has followed the famous Moore's law that predicts a steady decrease in gate length. As predicted by Gordon Moore in the 1960s, integrated circuit (IC) **densities have been doubling approximately every 18 months**, and this doubling in size has been accompanied by a similar exponential increase in circuit speed (or more precisely, clock frequency).



On-chip transistor count increase for the Intel processors (Source: Intel).



Moore's Law and our expectation

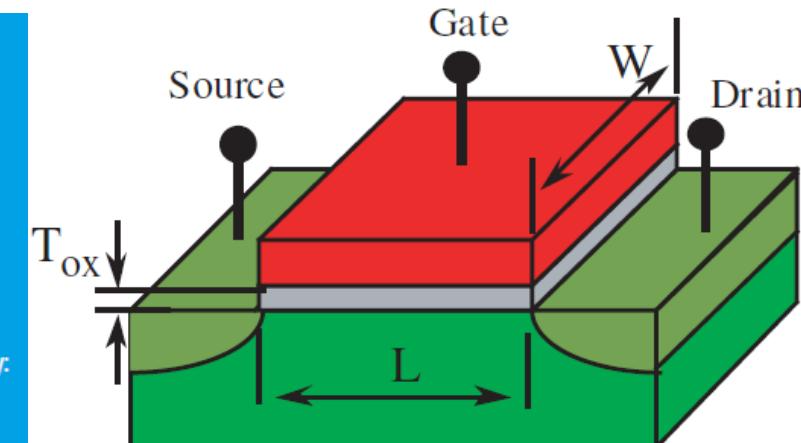
1
1971
Intel® 4004 processor
Initial clock speed: 108KHz
Transistors: 2,300
Manufacturing technology: 10 micron

2
1972
Intel® 8008 processor
Initial clock speed: 800KHz
Transistors: 3,500
Manufacturing technology: 10 micron

3
1974
Intel® 8080 processor
Initial clock speed: 2MHz
Transistors: 4,500
Manufacturing technology: 6 micron

4
1978
Intel® 8086 processor
Initial clock speed: 5MHz
Transistors: 29,000
Manufacturing technology: 3 micron

5
1982
Intel® 286™ processor
Initial clock speed: 6MHz
Transistors: 134,000
Manufacturing technology: 1.5 micron



Assume initial $L=1, W=1$ and $T_{ox}=1$

$$W=0.7, L=0.7, T_{ox}=0.7$$

=> Lateral and vertical dimensions reduce 30 %

$$\text{Area Cap} = C = \frac{0.7 \times 0.7}{0.7} = 0.7$$

- **Area reduced by 50%**
- **Capacitance reduced by 30%**

Assume initial $V_{dd}=1, V_t=1$ and $T_{ox}=1$

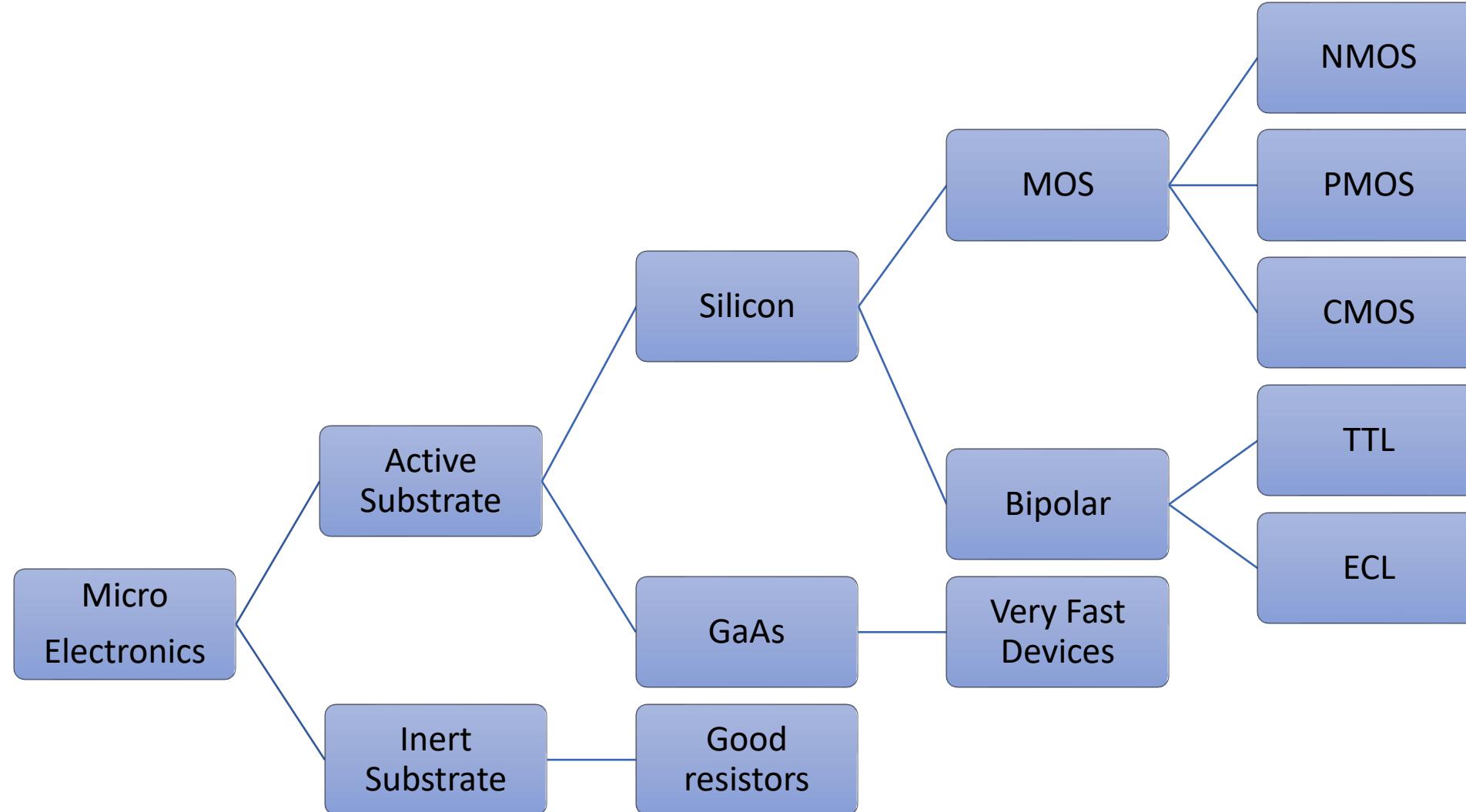
$$V_{dd}=0.7, V_t=0.7, T_{ox}=0.7,$$

$$T = \frac{C \times V_{dd}}{I} = 0.7, \text{ Power} = CV^2f = \frac{0.7 \times 0.7^2}{0.7} = 0.7^2$$

=> **Delay reduces by 30 % and Power reduces by 50 %**



Microelectronics Technology





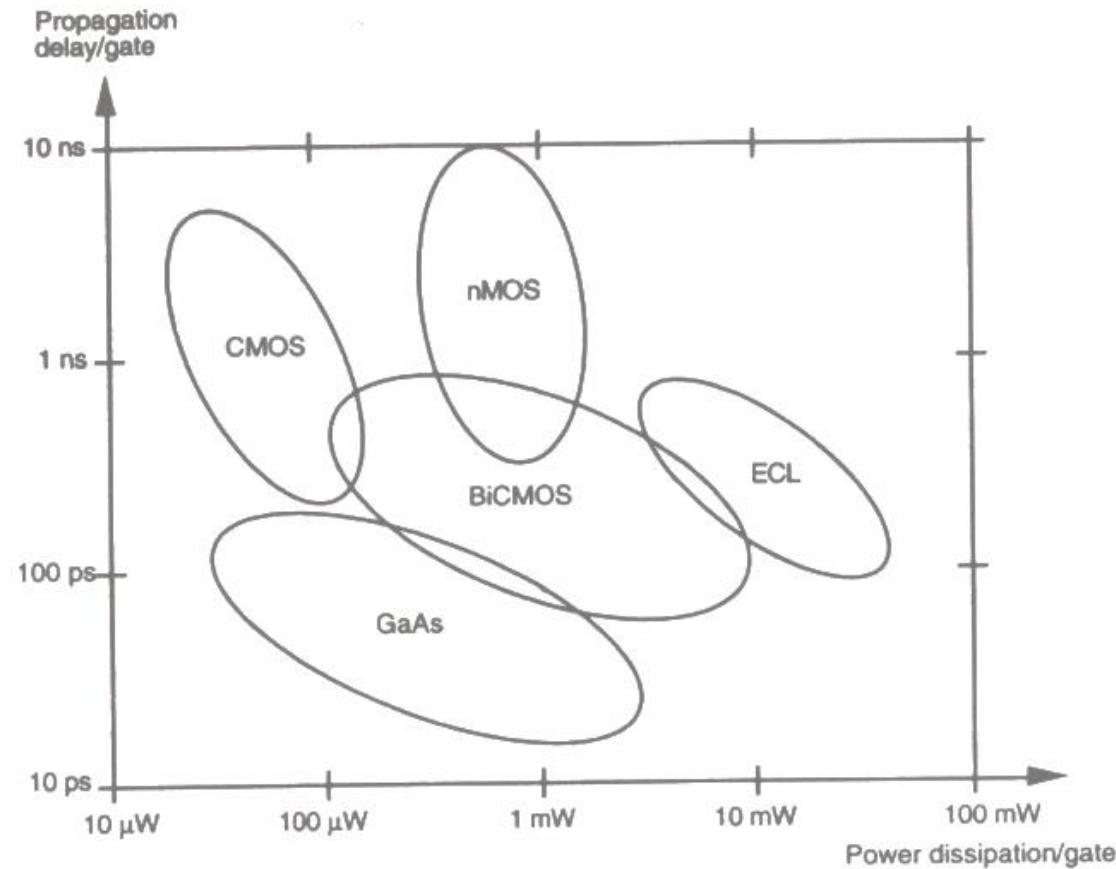
MOS vs BJT

Factors	CMOS	Bipolar
Static Power Dissipation	Low	High
Input Impedance	High	Low
Noise Margin	High	Low
Packing Density	High	Low
Fan-out	Low	High
Direction	Bidirectional	Unidirectional

CMOS is superior!



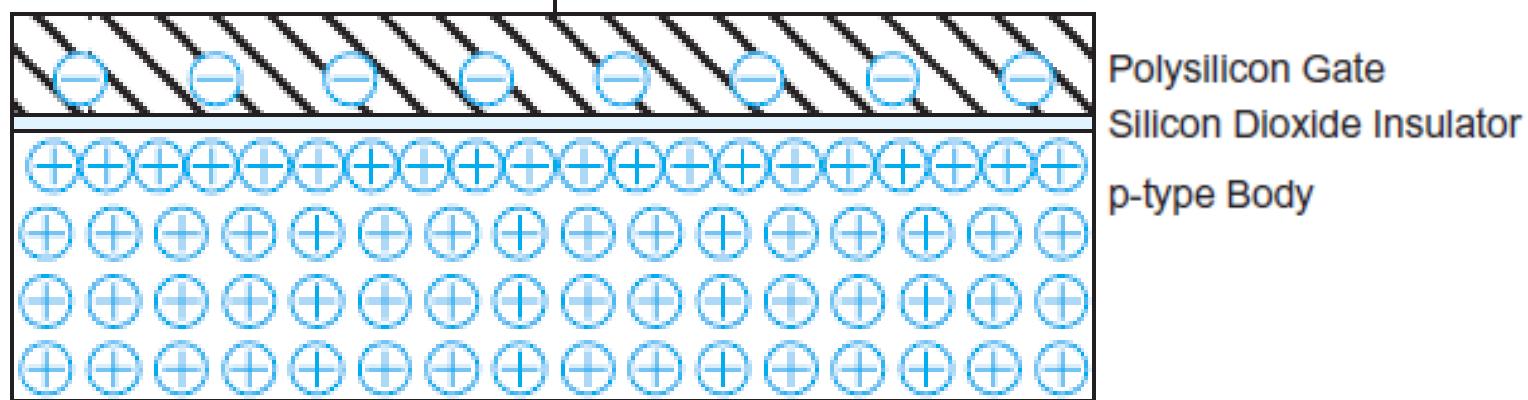
Power Dissipation Vs. Delay



CMOS offers low powers dissipation with large delay



MOS Capacitor (MOSCAP)

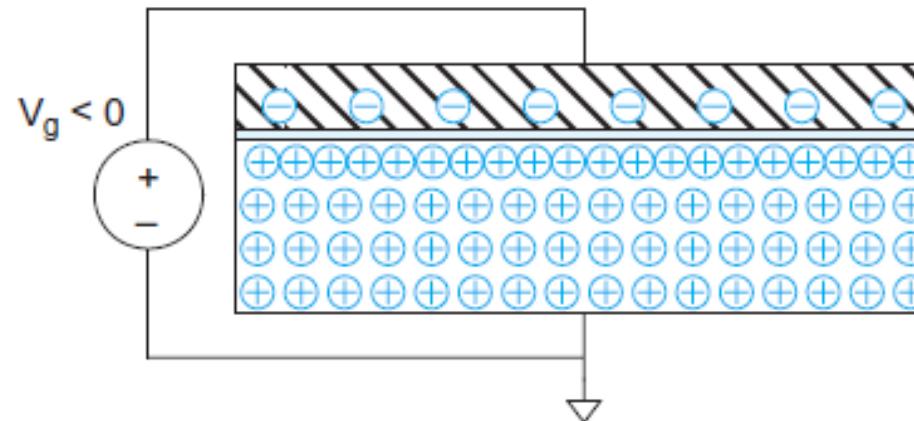


Oxide thickness, Threshold voltage, and Doping levels, depend on the fabrication process, and cannot be changed by design; they are technology parameters.



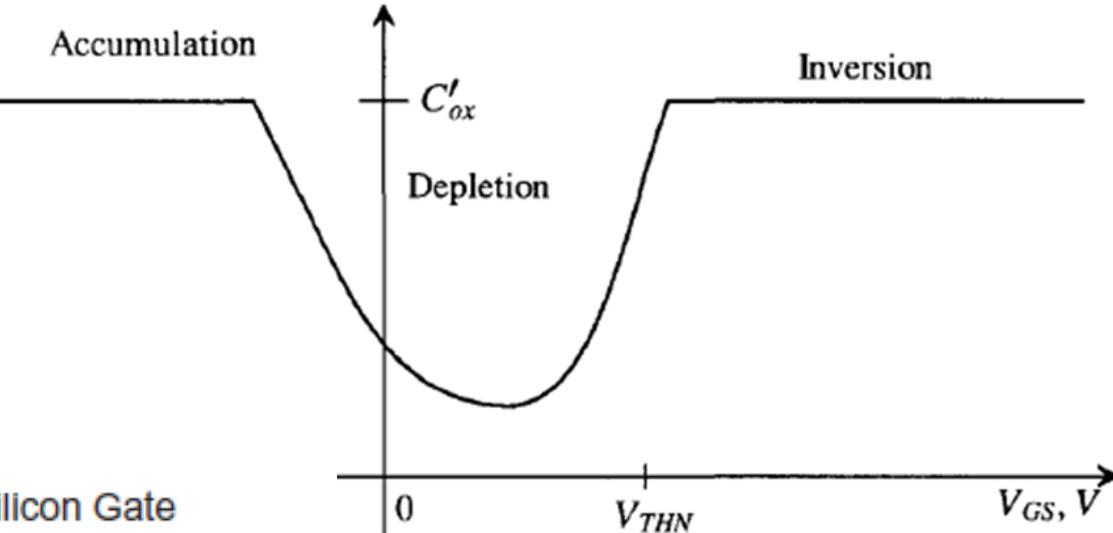
MOSCAP

i. Accumulation



Polysilicon Gate
Silicon Dioxide Insulator
p-type Body

$$C_g = \frac{\epsilon_{ox}}{t_{ox}}$$

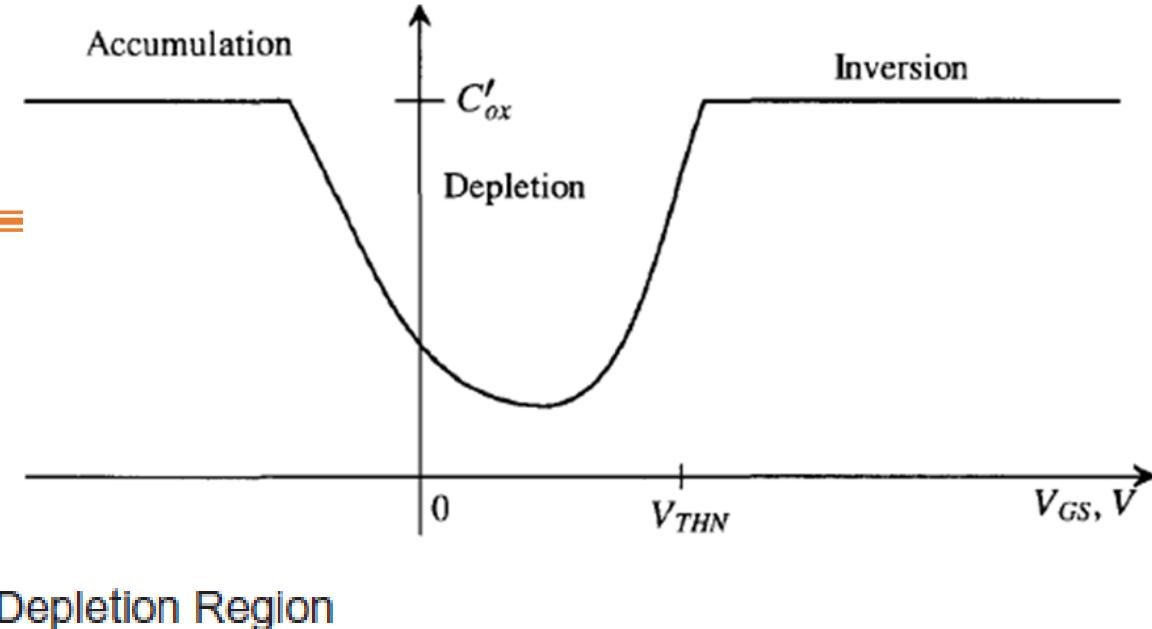
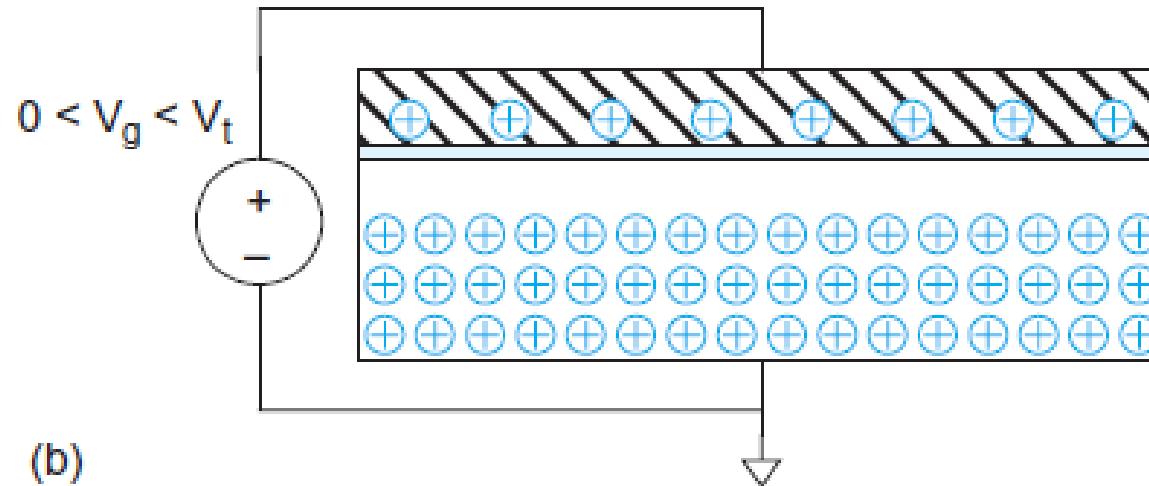


A negative voltage is applied to the gate, so there is negative charge on the gate. The mobile positively charged holes are attracted to the region beneath the gate. This is called the *accumulation* mode



MOSCAP

ii. Depletion



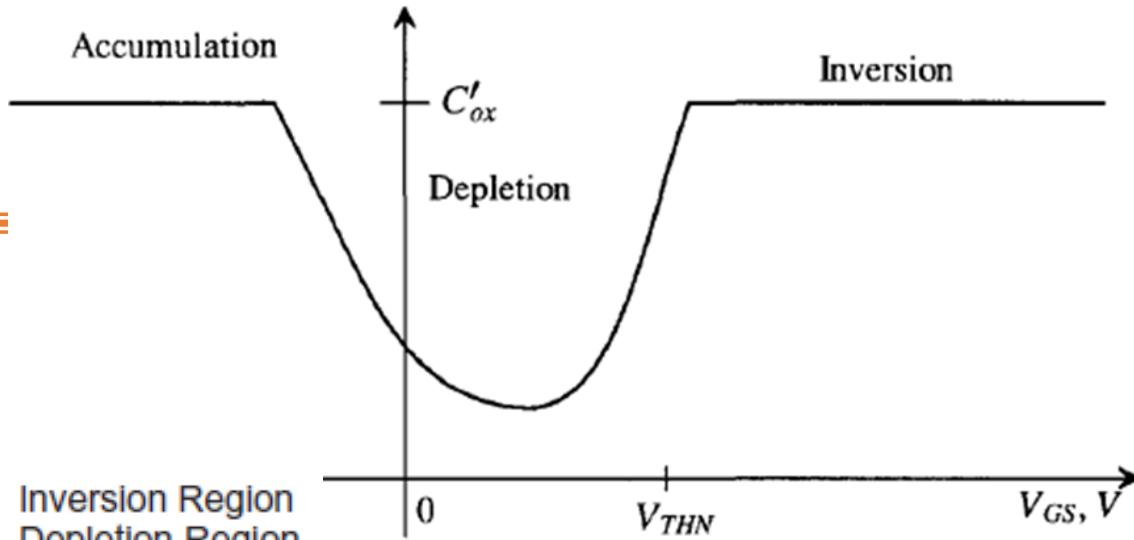
Depletion Region

$$C_g = \frac{\epsilon_{ox}}{t_{ox}}$$

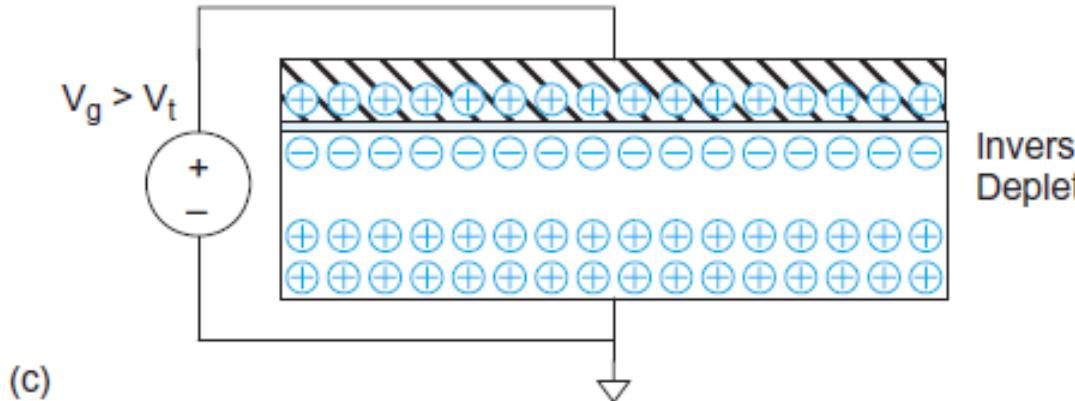
A small positive voltage is applied to the gate, resulting in some positive charge on the gate. The holes in the body are repelled from the region directly beneath the gate, resulting in a *depletion* region forming below the gate.



MOSCAP



iii. *inversion*

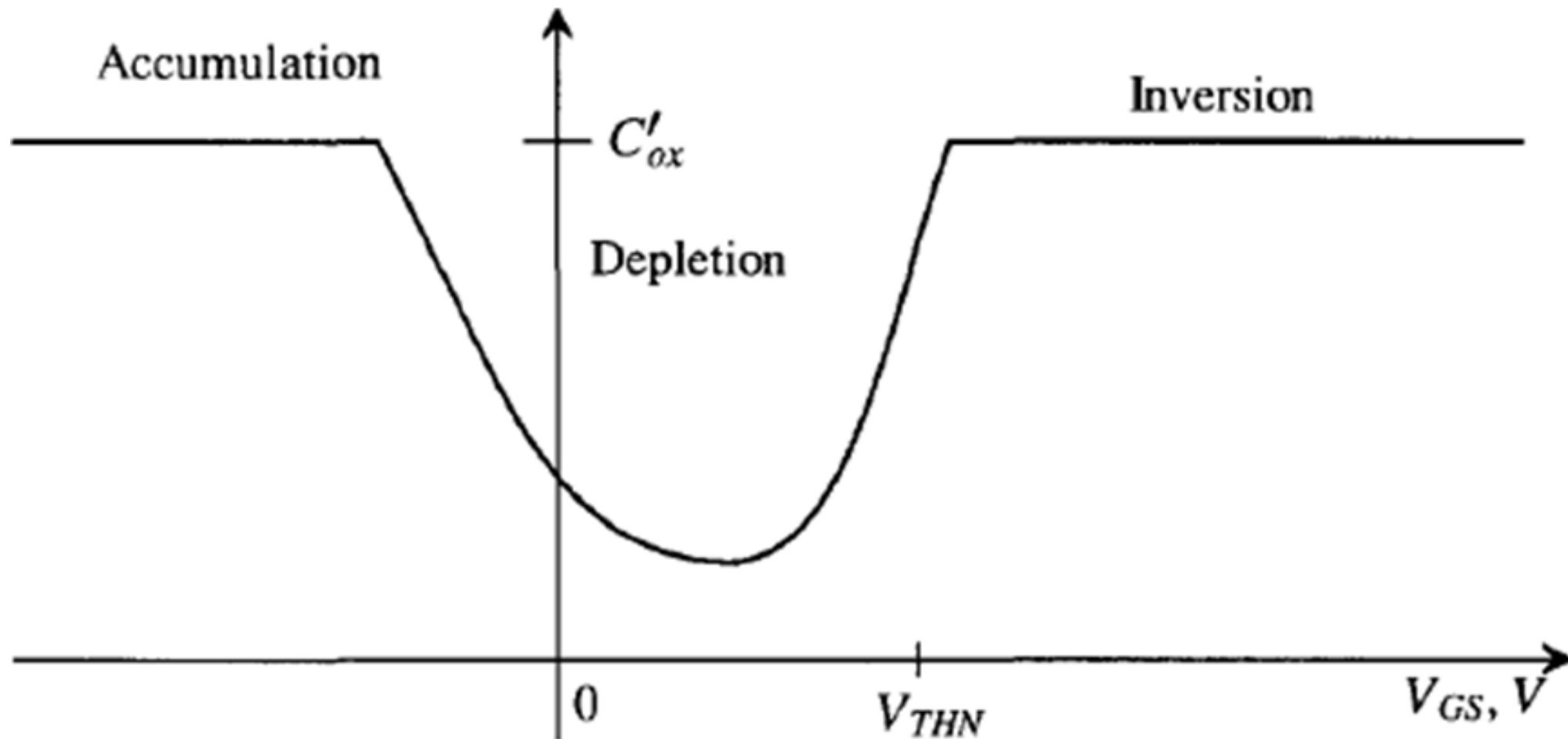


$$C_g = \frac{\epsilon_{ox}}{t_{ox}}$$

A higher positive potential exceeding a critical threshold voltage V_t is applied, attracting more positive charge to the gate. The holes are repelled further and some free electrons in the body are attracted to the region beneath the gate. This conductive layer of electrons in the p-type body is called the *inversion* layer



MOS Capacitor (MOSCAP)





MOSFET

Enhancement type MOSFET

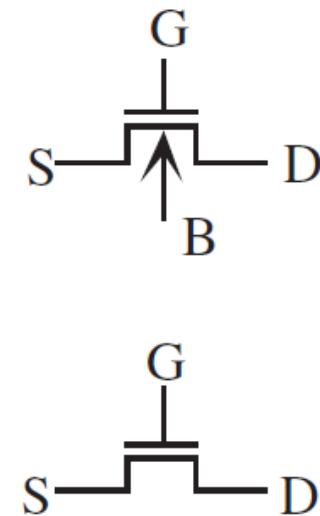
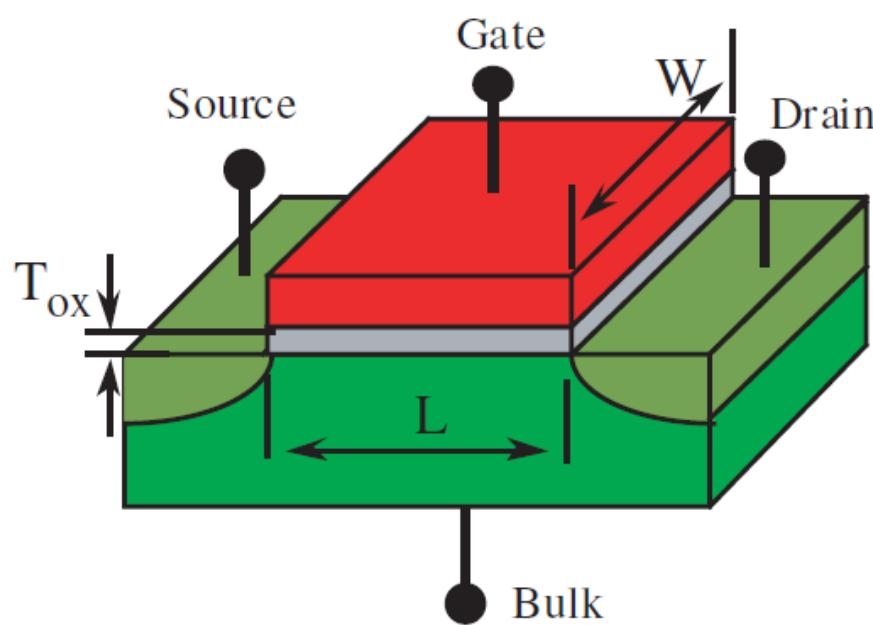
1. N-Channel MOSFET (NMOS)
2. P-Channel MOSFET (PMOS)

Depletion type MOSFET

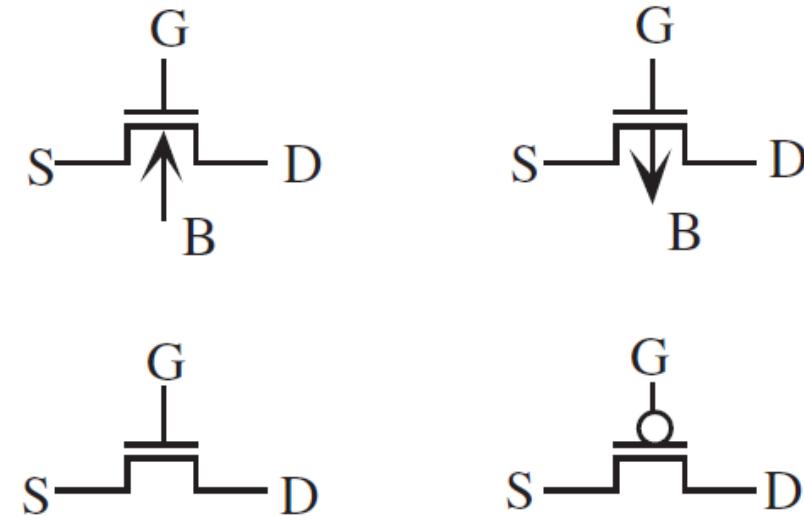
1. N-Channel MOSFET (NMOS)
2. P-Channel MOSFET (PMOS)



Enhancement MOSFET



nMOS



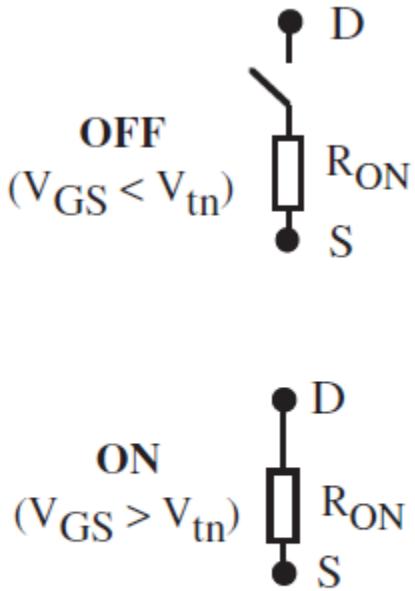
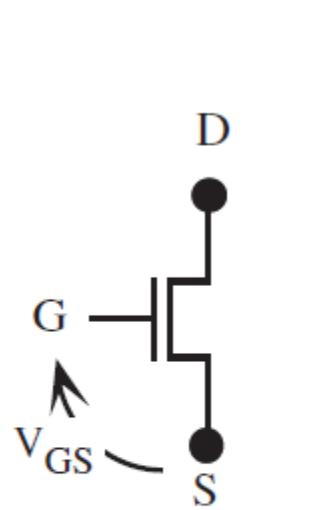
pMOS

$T_{ox} = 15\text{\AA}$ to 100\AA (Diameter of SiO_2 molecule is about 3.2\AA)

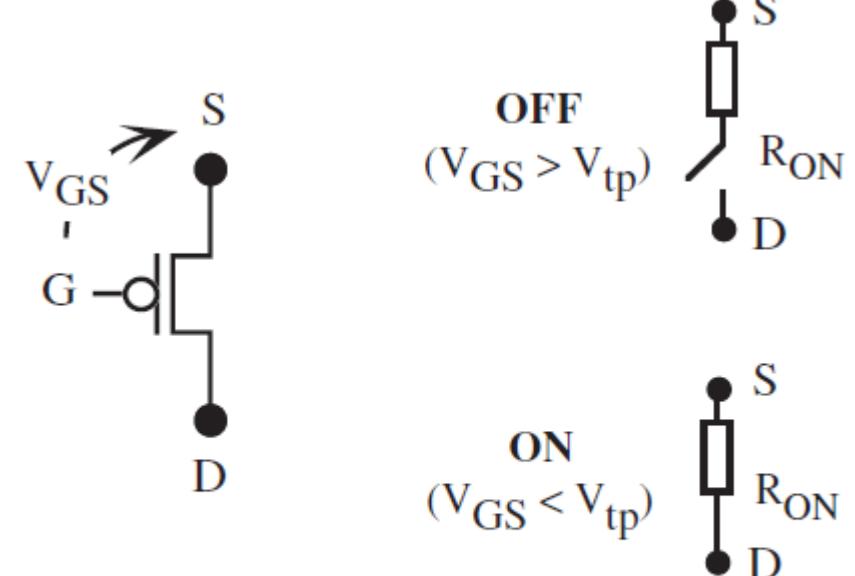
Arrows always point from P to N, so an NMOS (N-channel in P-well or P-substrate) has the arrow pointing in (from the bulk to the channel)



Enhancement MOSFET



nMOS ($V_{tn} > 0, V_{DS} \geq 0, V_{GS} \geq 0$)



pMOS ($V_{tp} < 0, V_{DS} \leq 0, V_{GS} \leq 0$)

V_{Th} is fixed for NMOS and PMOS devices for given fabrication process



Enhancement nMOS



Operation of nMOS

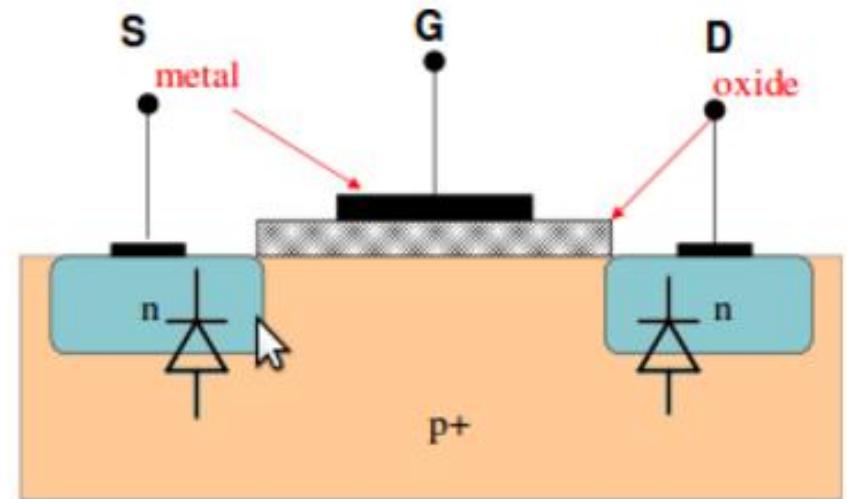
Depending on the relative voltages of the source, drain and gate, the NMOS transistor may operate in any of three regions viz :

- Cut_off : Current flow is essentially zero (also called accumulation region)
- Linear : (Non saturated region)-It is weak inversion region drain current depends on gate and drain voltage.
- Saturation : It is strong inversion region where drain current is independent of drain-source voltage.



Operation of nMOS

Cut-off Region



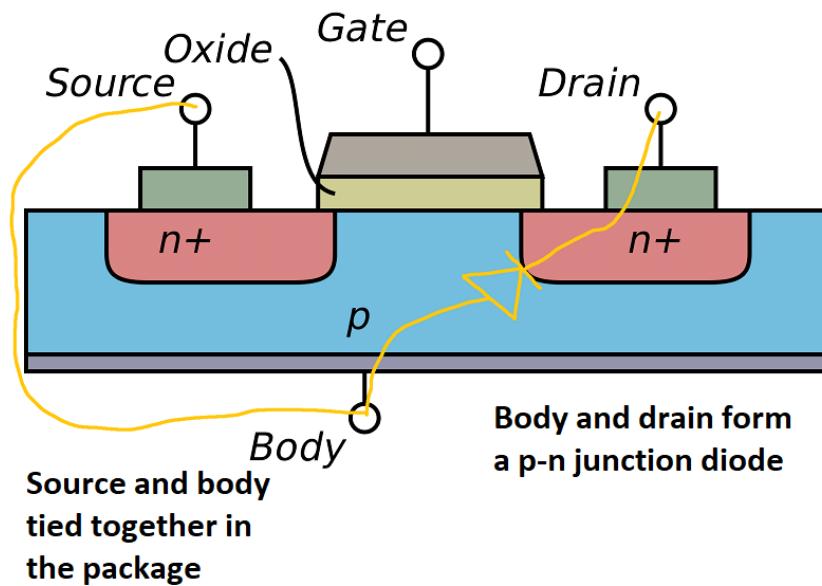
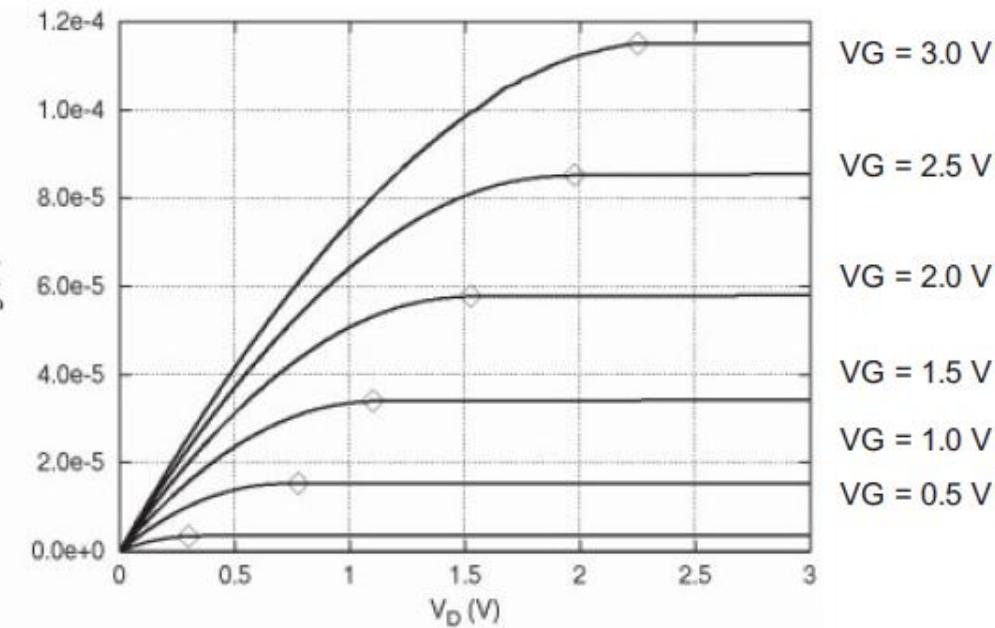
Source (S)

$V_{GS}=0$

$V_{DS}=0$

n+

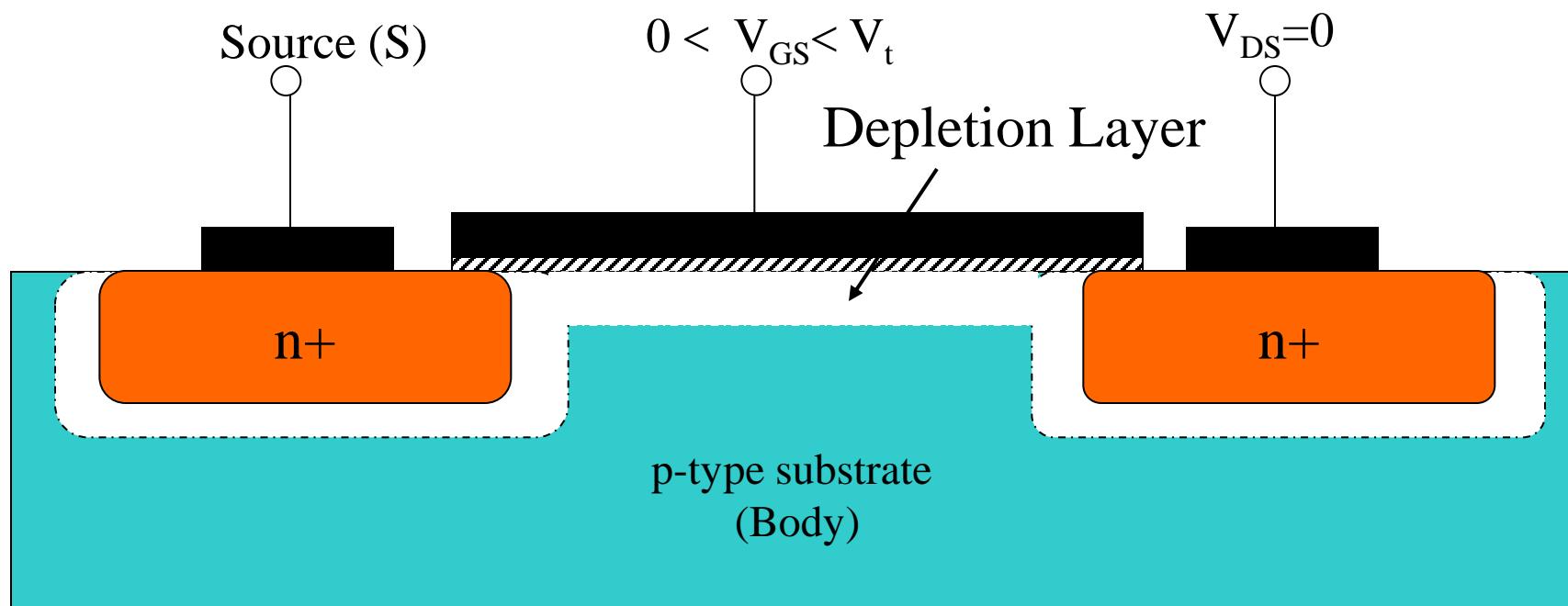
p-type substrate
(Body)





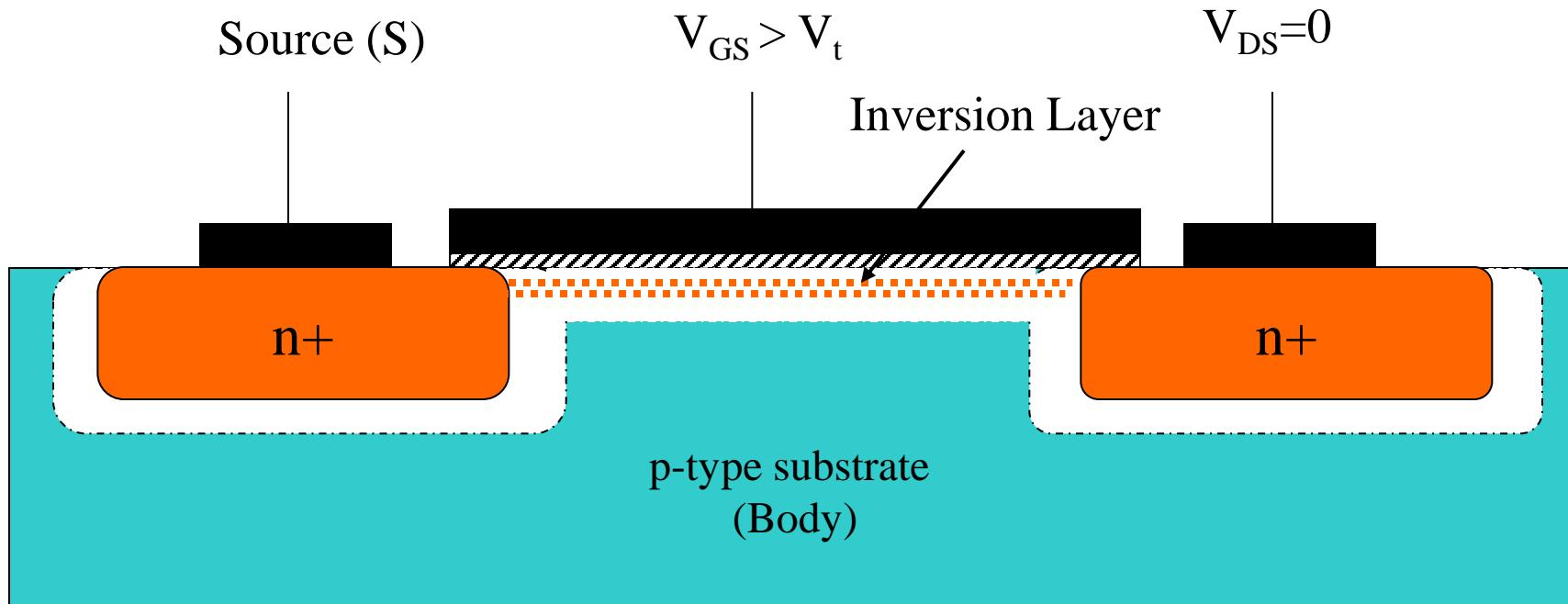
Operation of nMOS

Cut-off Region





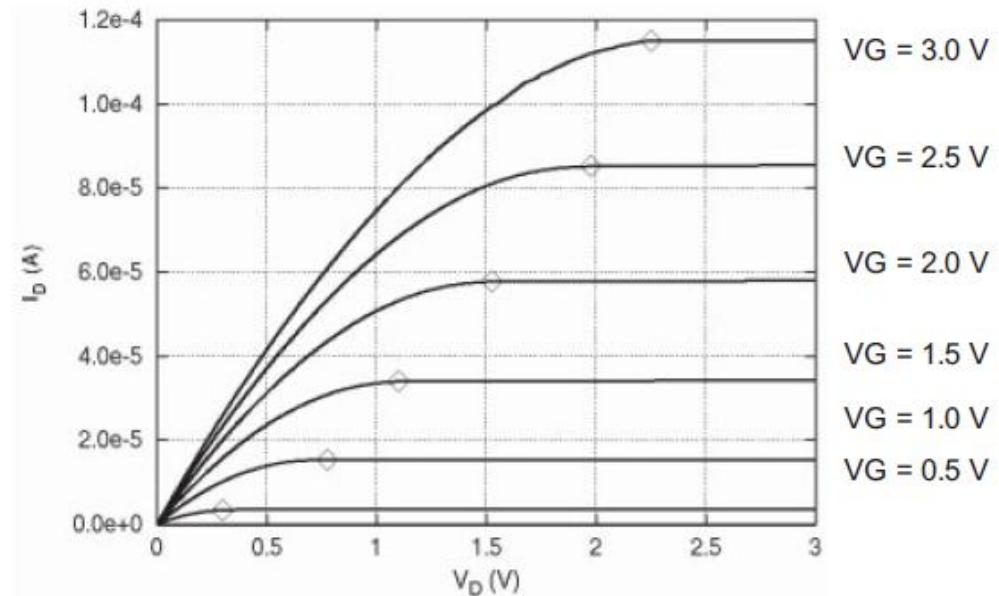
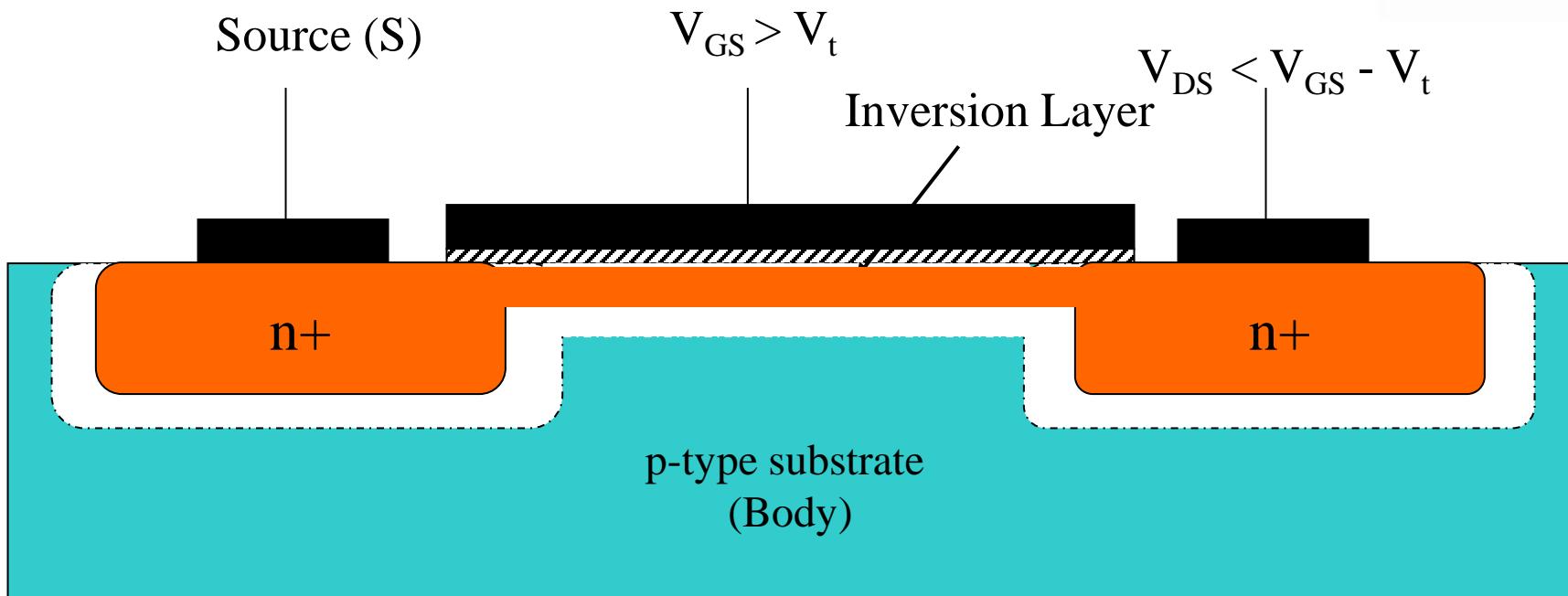
Operation of nMOS





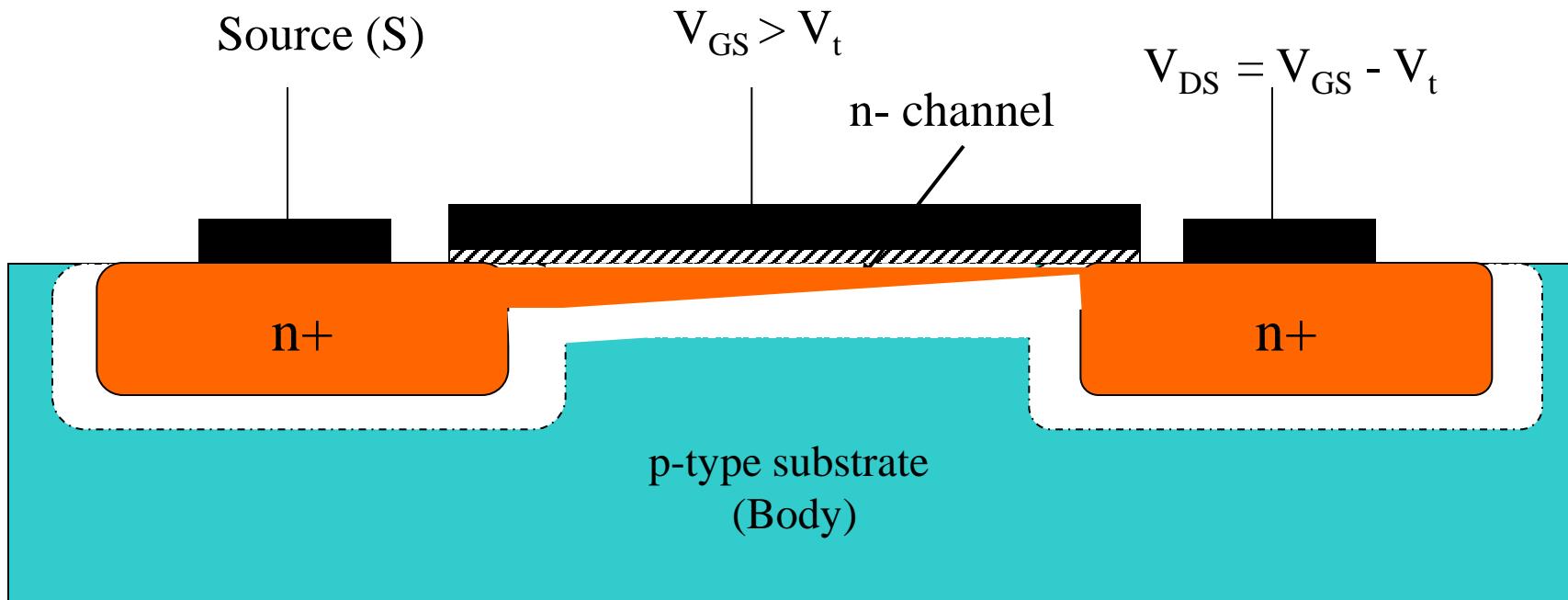
Operation of nMOS

Linear region





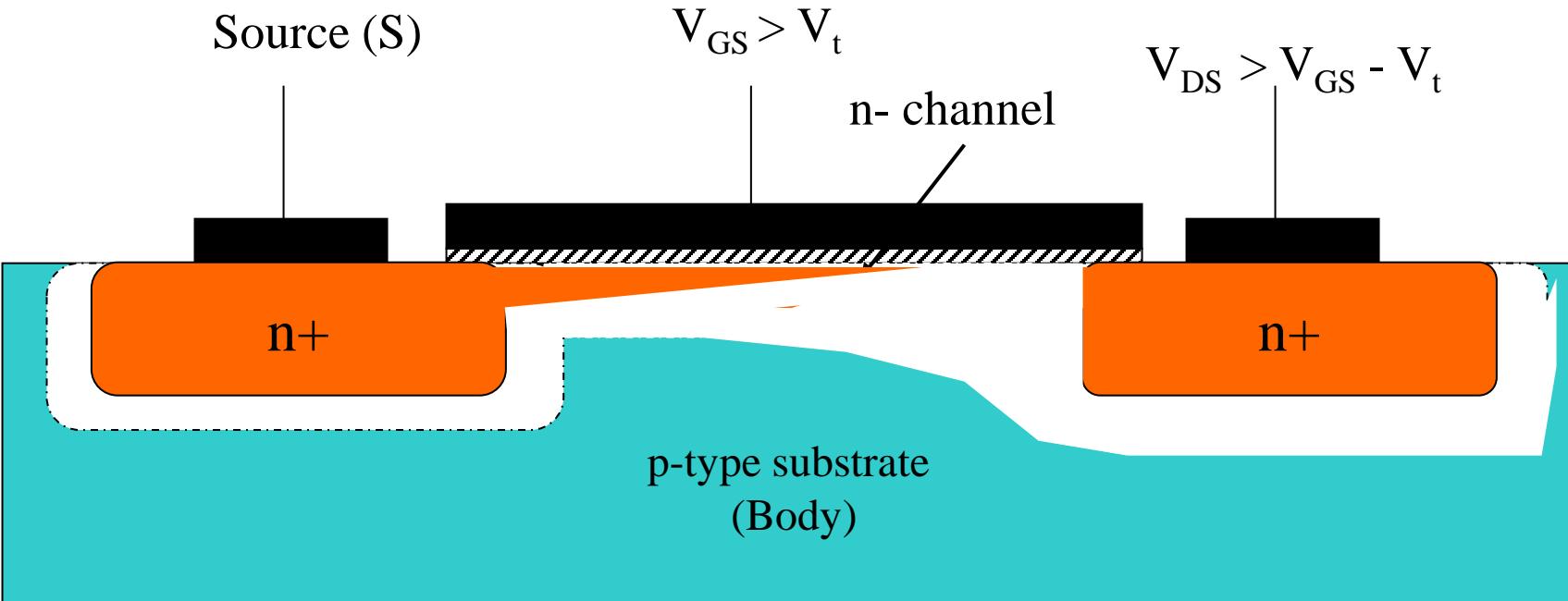
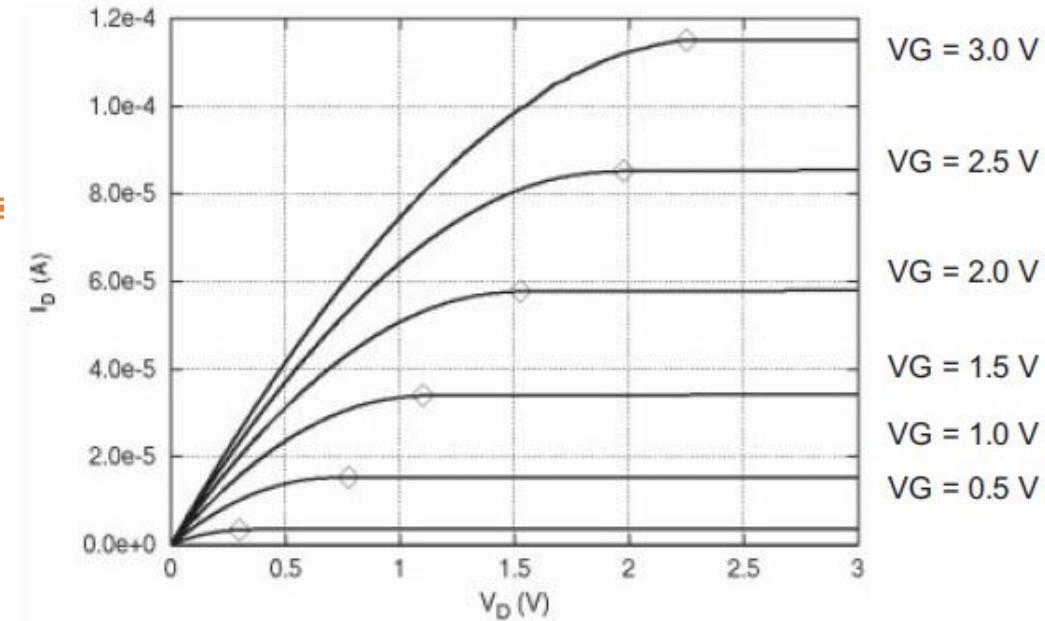
Operation of nMOS





Operation of nMOS

Saturation Region



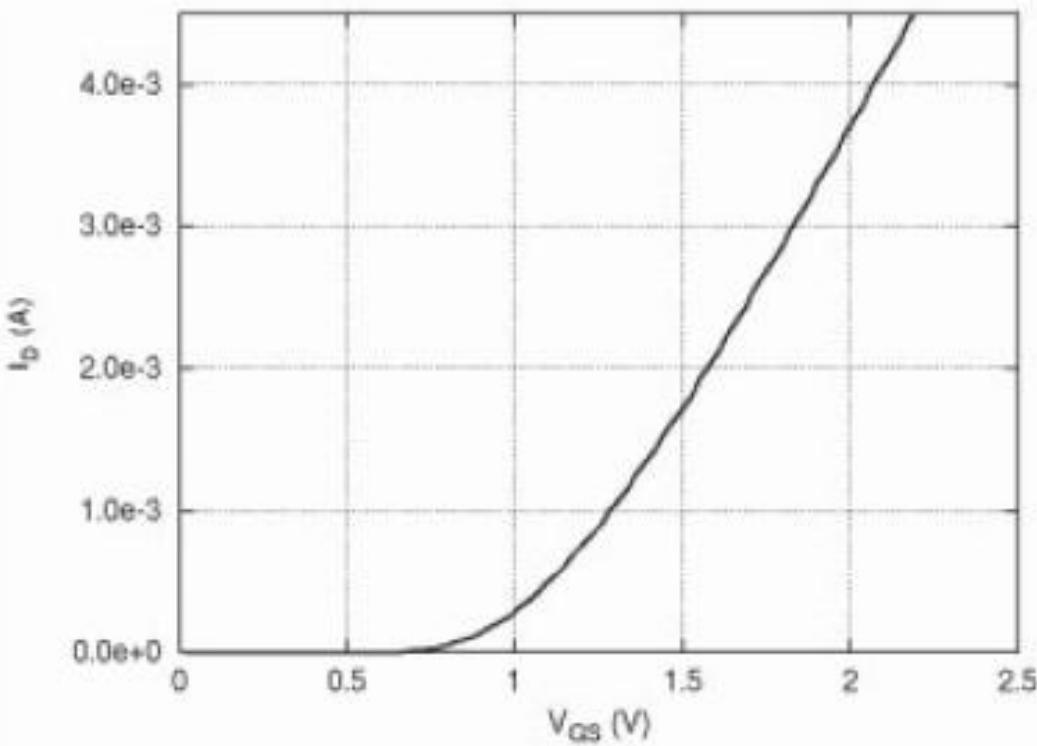


Important observations

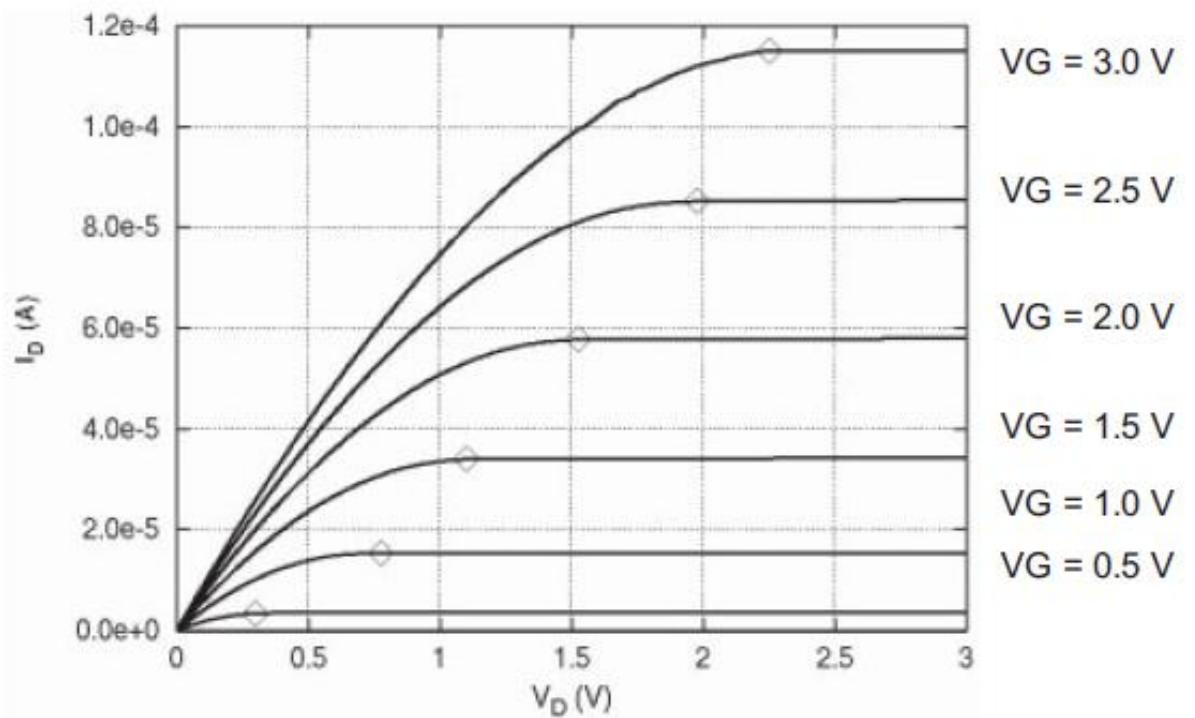
1. The bulk or substrate of *n*MOS transistors must always be connected to the lower voltage that is the reference terminal.
2. The positive convention current in an *n*MOS device is from the drain to the source, and is referred to as I_{DS} or just I_D , since drain and source current are equal.
3. When a positive voltage is applied to the drain terminal, the drain current depends on the voltage applied to the gate control terminal.



Input and output characteristics of nMOS



input characteristics (I_D vs. V_{GS}) for an nMOS,



nMOS transistor output characteristics (I_D vs. V_{DS})

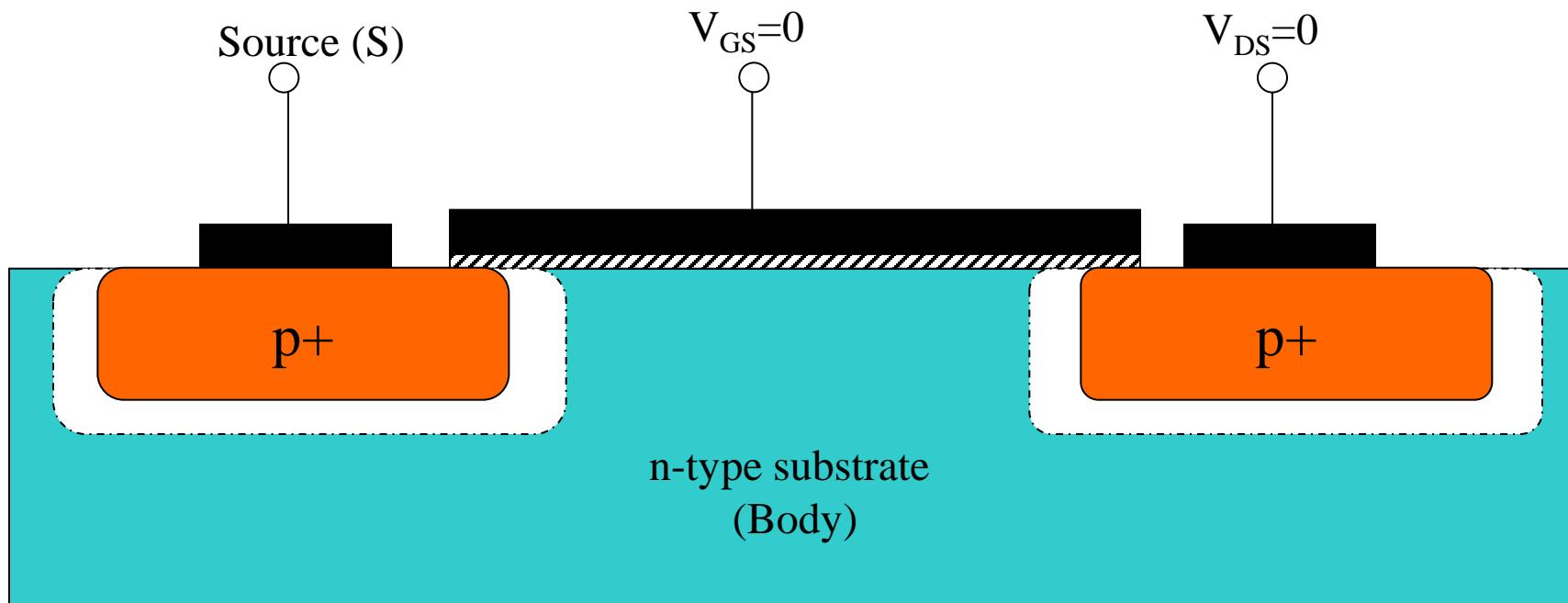


Enhancement pMOS



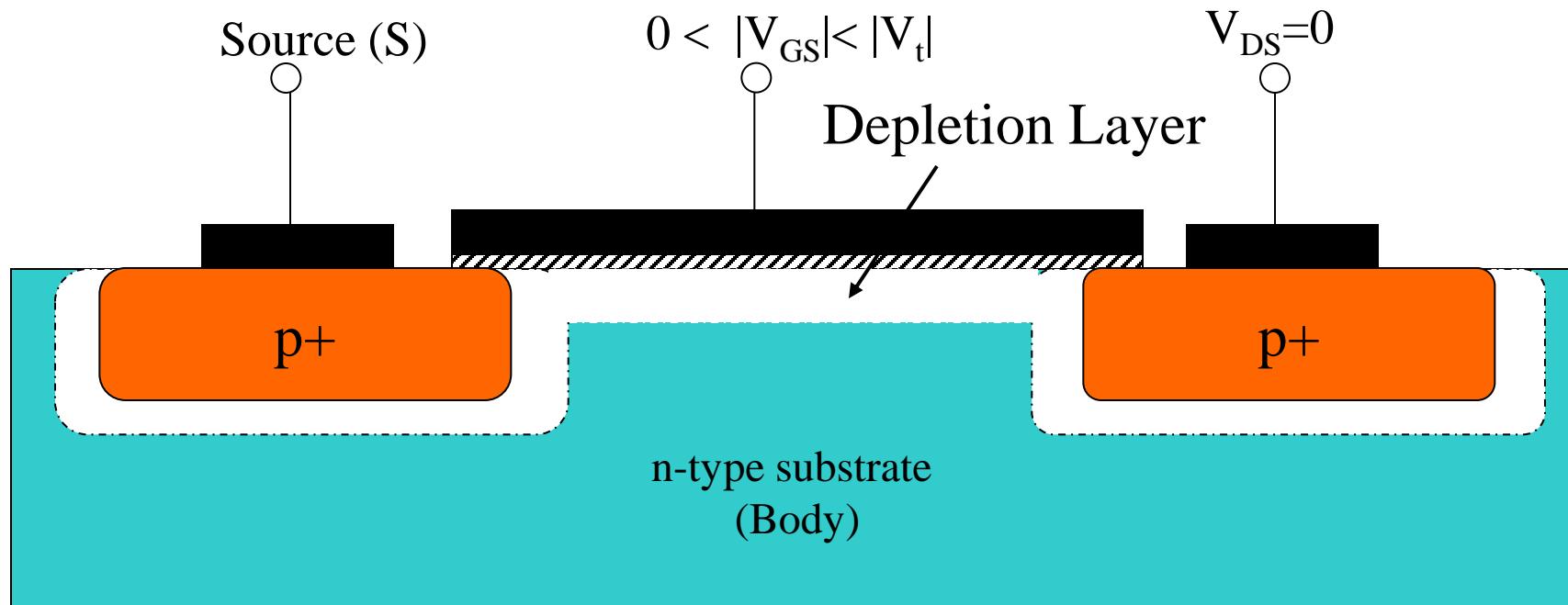
Enhancement pMOS

Cut-off Region



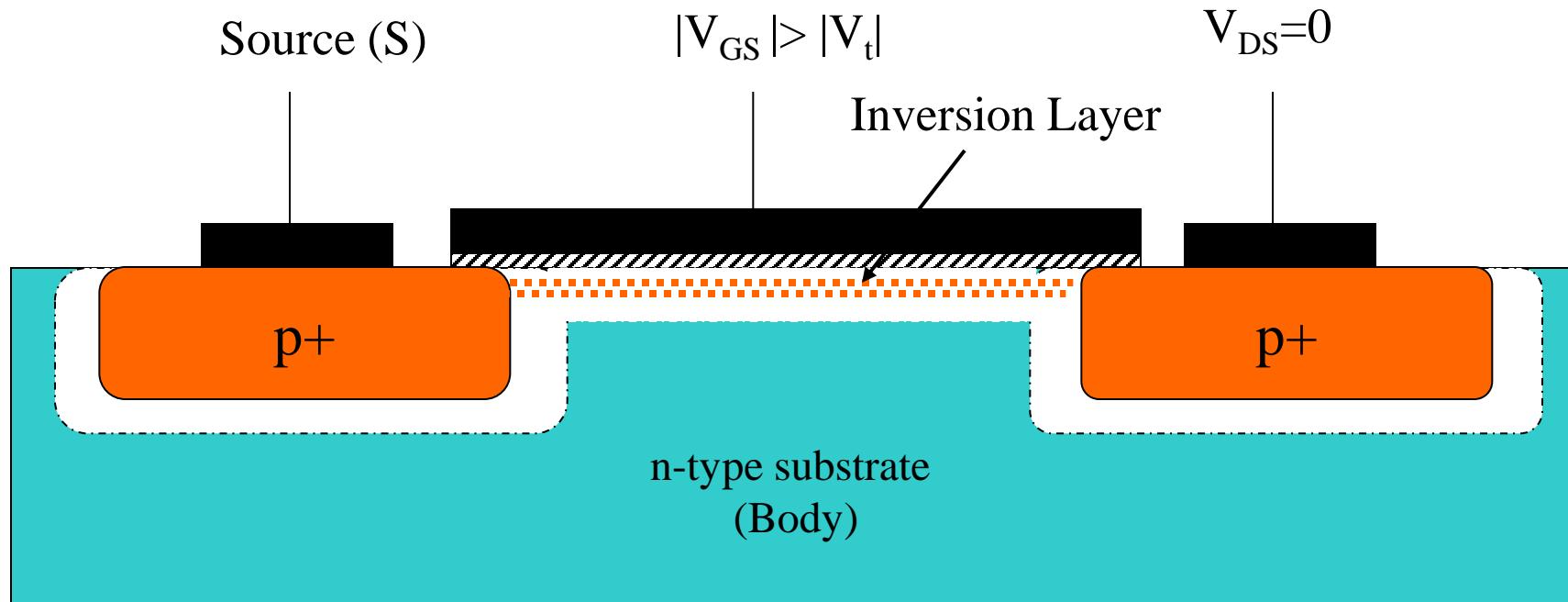


Enhancement pMOS





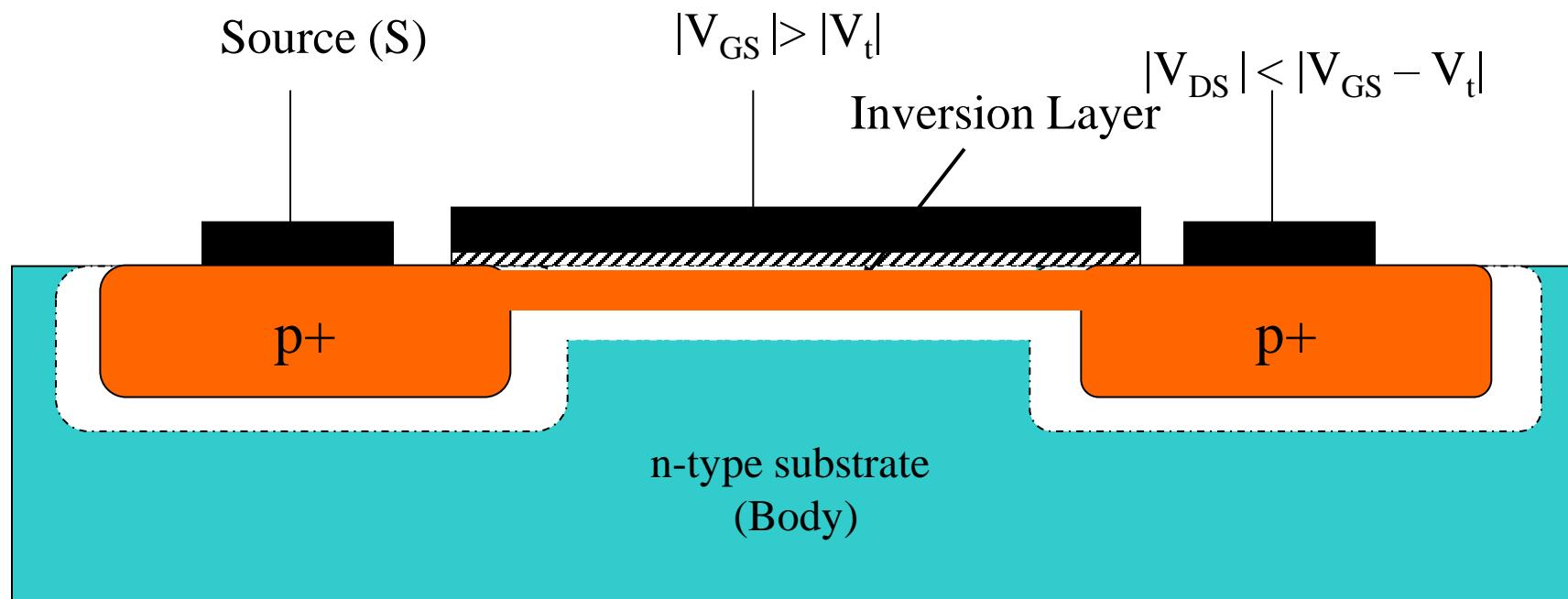
Enhancement pMOS





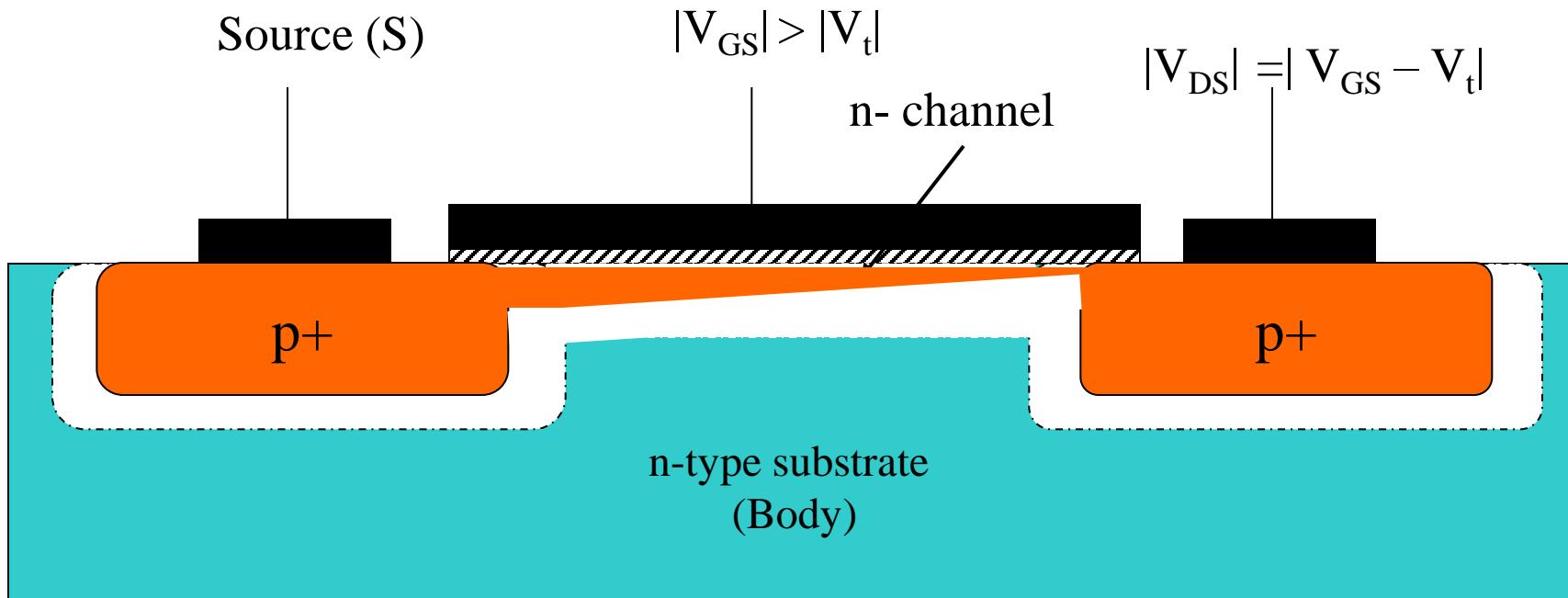
Enhancement pMOS

Linear region





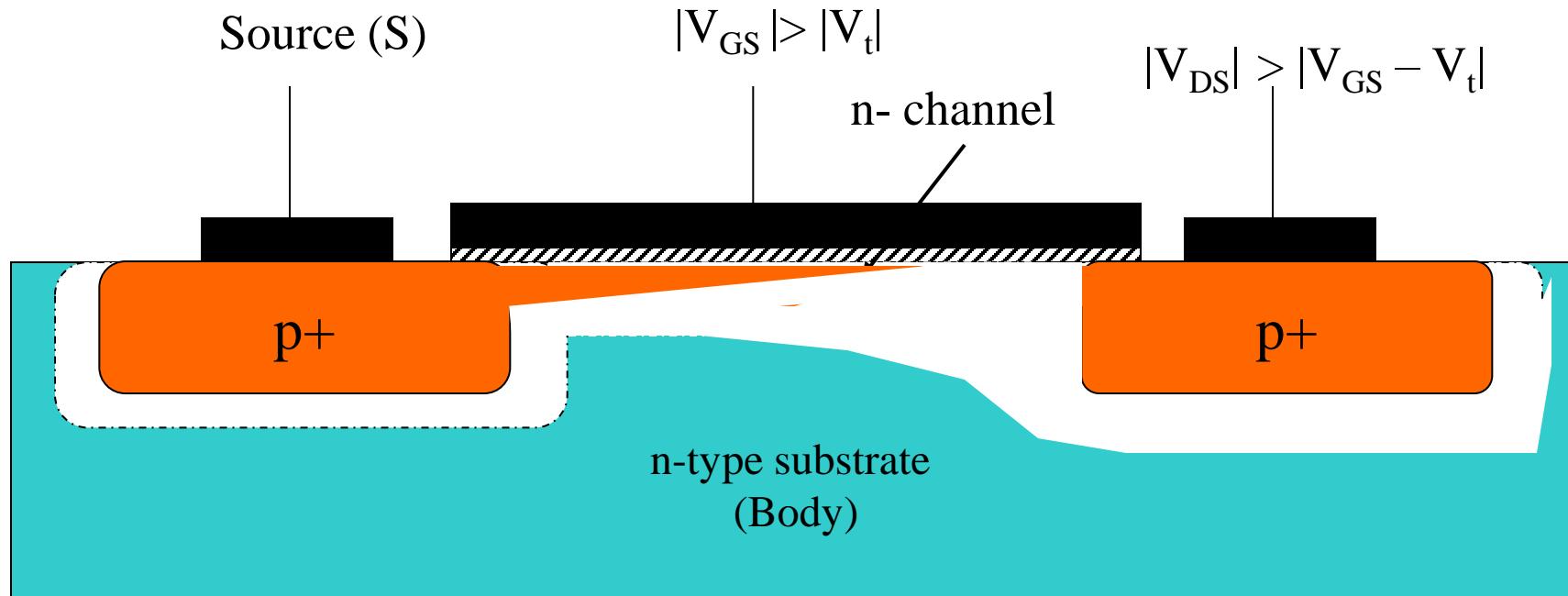
Enhancement pMOS

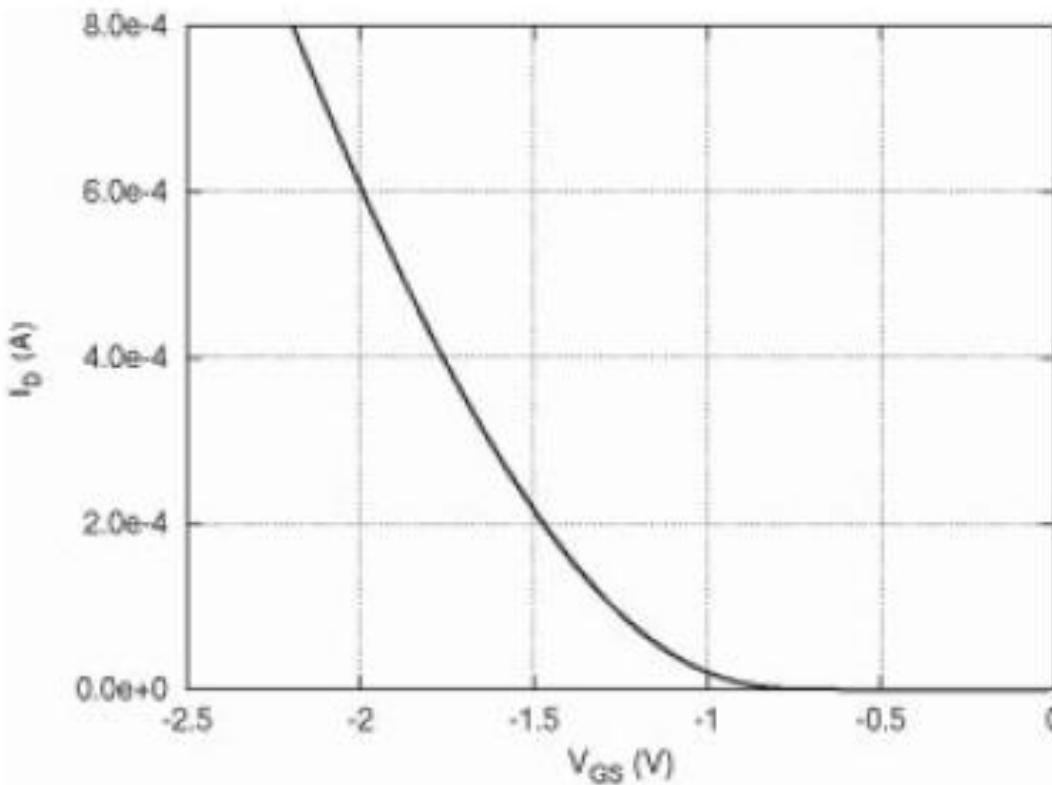




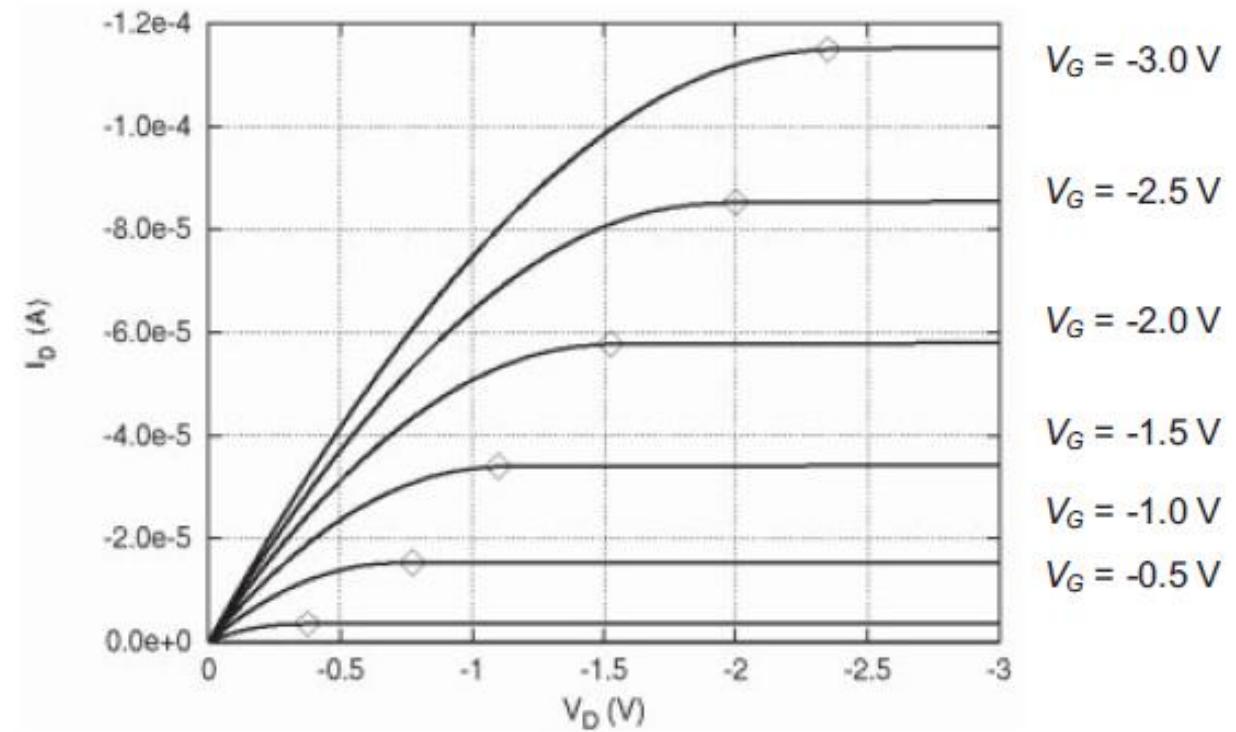
Enhancement pMOS

Saturation Region





input characteristics (I_D vs. V_{GS}) for an *p*MOS,

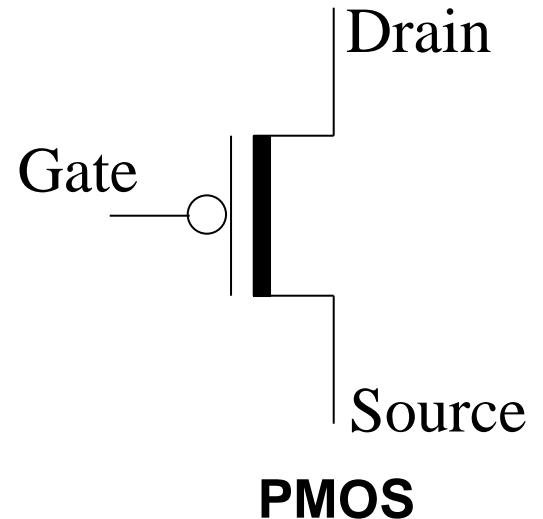
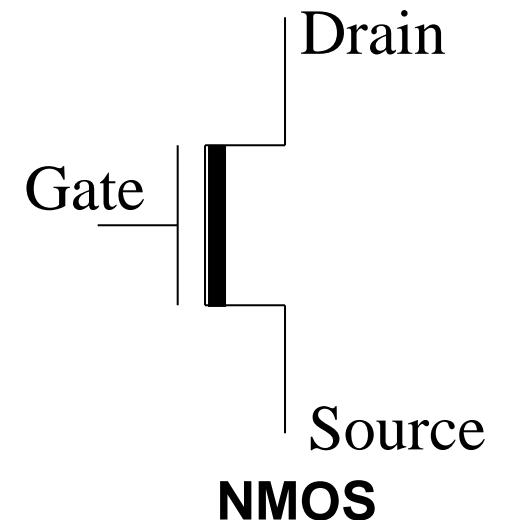
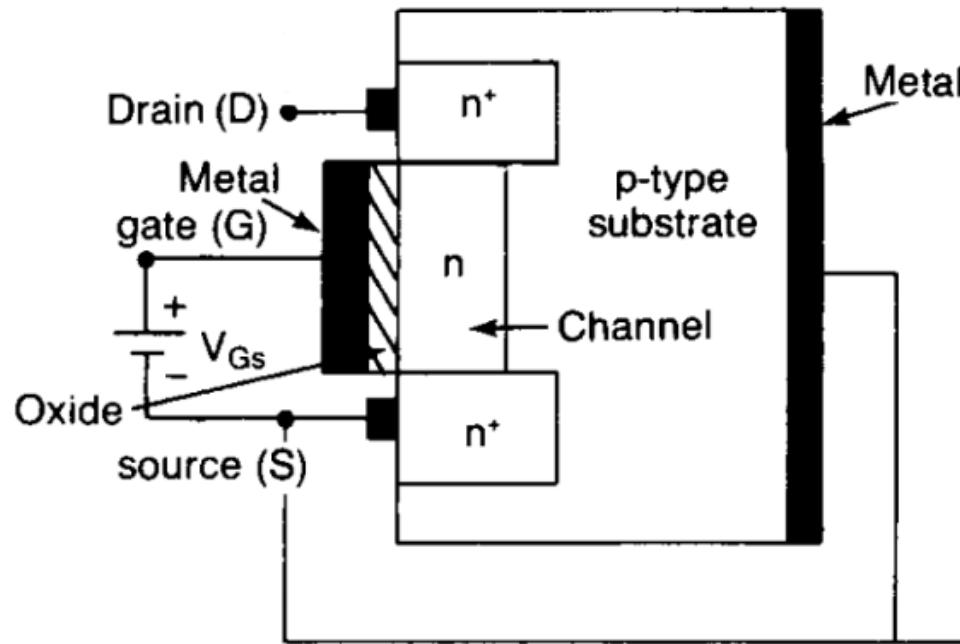


*p*MOS transistor output characteristics



Depletion nMOS

Depletion nMOS



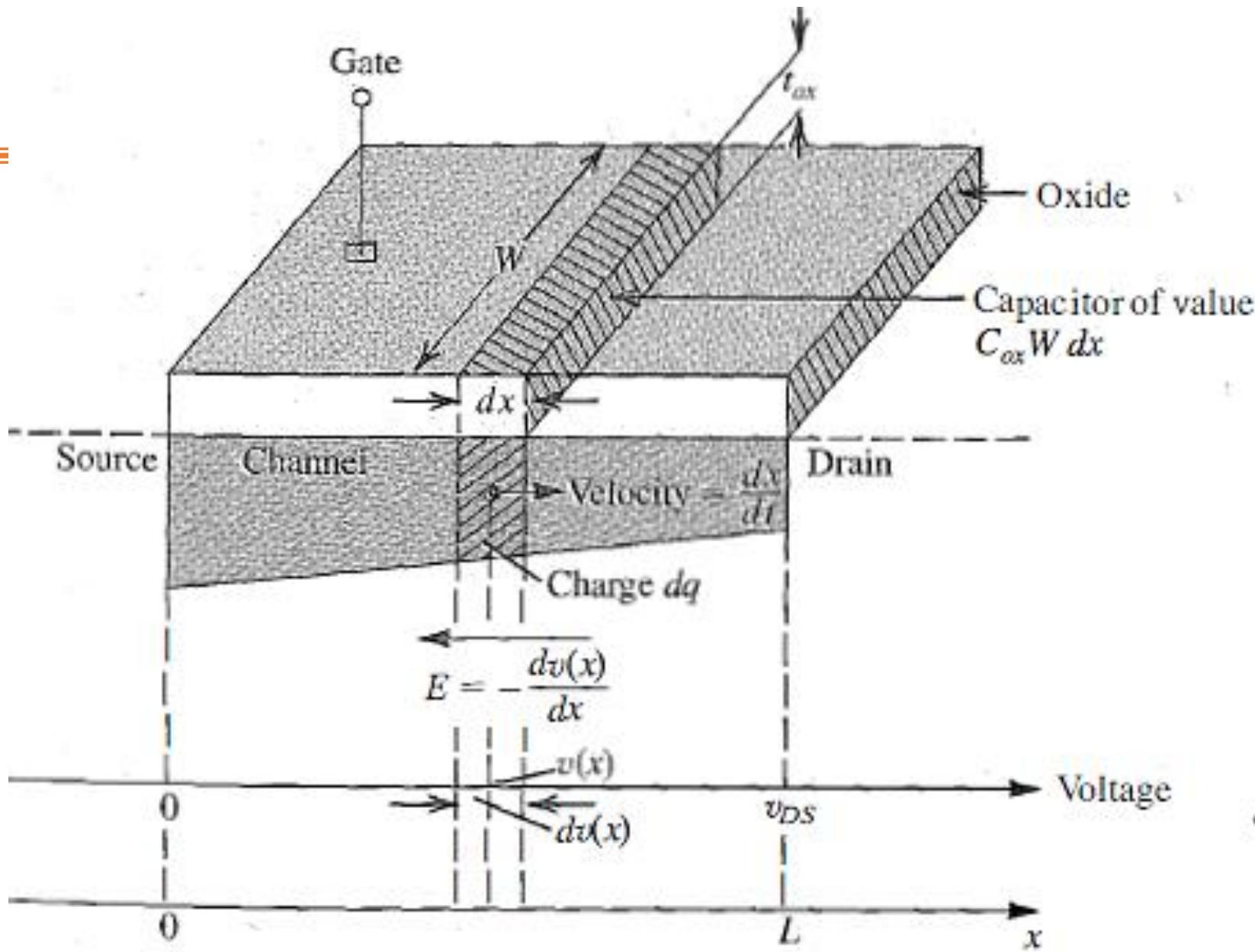
- In Depletion MOS structure, the source & drain are diffused on P- substrate as shown above.
- Positive voltages enhances number of electrons from source to drain.
- Negative voltage applied to gate reduces the drain current
- This is called as ' normally ON ' MOS.



Drain Current (I_{DS}) Derivation For Enhancement NMOS



Drain to Source Current I_{DS}





The gate and the channel region form a parallel plate capacitor for which the oxide layer serves as a dielectric.

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

$$\epsilon_{ox} = 3.9\epsilon_0 = 3.9 \times 8.854 \times 10^{-12} = 3.45 \times 10^{-11} \text{ F/m}$$

Consider the infinitesimal strip of the gate at distance x from the source. The capacitance of this strip is

$$C_{ox} W dx$$

To find the charge stored on this infinitesimal strip of the gate capacitance, we multiply the capacitance by the effective voltage between the gate and the channel at point x ,

$$V_{GS} - V(x) - V_t$$



Charge dq in the infinitesimal portion of the channel at point x is

$$dq = -C_{ox}Wdx[V_{GS} - V(x) - V_t]$$

$$\frac{dq}{dx} = -C_{ox}Wdx[V_{GS} - V(x) - V_t]$$

Negative sign accounts for the fact that dq is a negative charge

The voltage V_{DS} produces an electric field along the channel in the negative x direction

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

The electric field $E(x)$ causes the electron charge dq to drift toward the drain with a velocity $\frac{dx}{dt}$



$$Velocity \frac{dx}{dt} = -\mu_n E(x) = \mu_n \frac{dV(x)}{dx}$$

Where μ_n is the mobility of electrons in the channel (called surface mobility).

The resulting drift current i

$$i = \frac{dq}{dt} = \frac{dq}{dx} \frac{dx}{dt}$$

$$i = -\mu_n C_{ox} W [V_{GS} - V(x) - V_t] \frac{dV(x)}{dx}$$



Thus, i must be equal to the source-to-drain current. Since we are interested in the drain-to-source current I_{DS} , we can find it as

$$I_{DS} = \mu_n C_{ox} W [V_{GS} - V(x) - V_t] \frac{dV(x)}{dx}$$

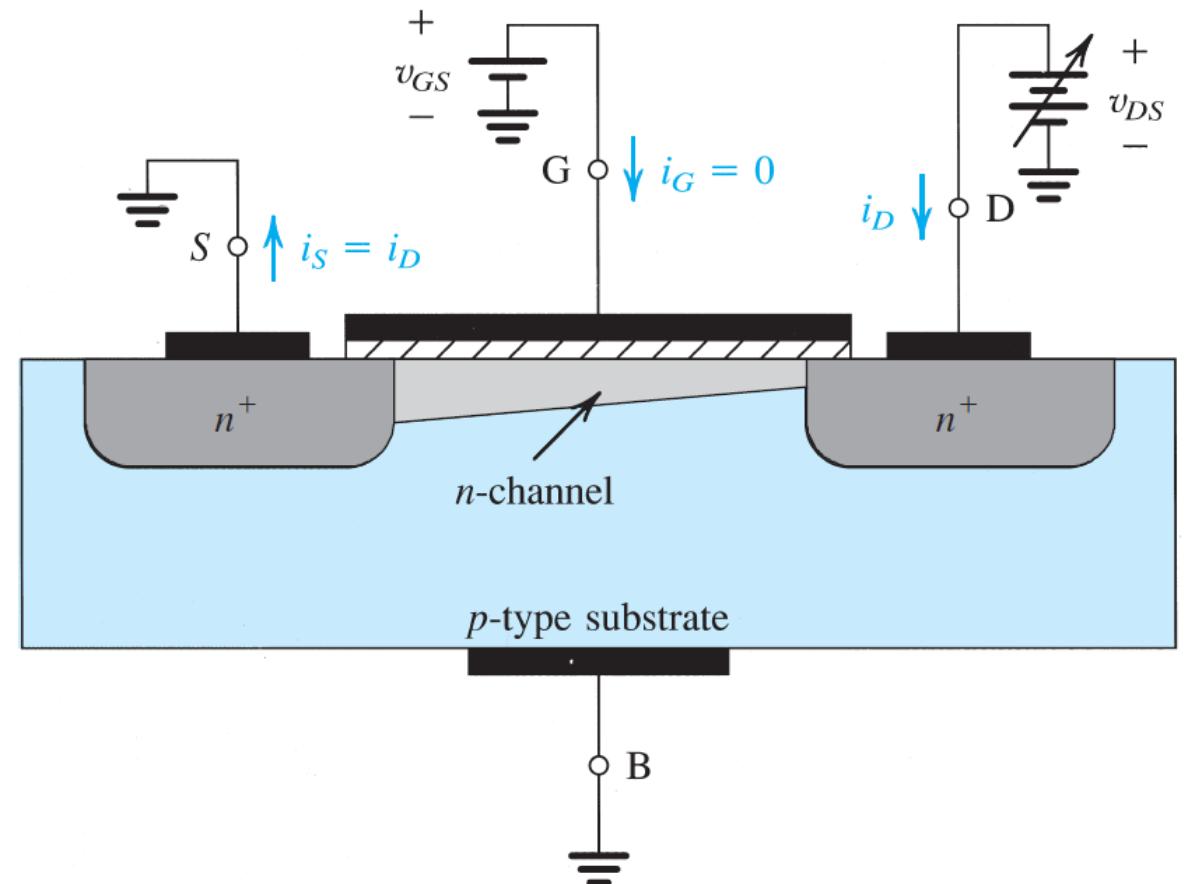
$$I_{DS} dx = \mu_n C_{ox} W [V_{GS} - V(x) - V_t] dV(x)$$

Integrating both sides of this equation from $x = 0$ to $x = L$ and, correspondingly, for $V(0) = 0$ to $V(L) = V_{DS}$,

$$\int_0^L I_{DS} dx = \int_0^{V_{DS}} \mu_n C_{ox} W [V_{GS} - V(x) - V_t] dV(x)$$



Linear region: $I_{DS} = \mu_n C_{ox} \frac{W}{L} [(V_{GS} - V_t)V_{DS} - \frac{V_{DS}^2}{2}]$





Saturation Region $V_{DS} \geq (V_{GS} - V_T)$

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} [(V_{GS} - V_t)(V_{GS} - V_t) - \frac{(V_{GS} - V_t)^2}{2}]$$

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} [(V_{GS} - V_t)^2 - \frac{(V_{GS} - V_t)^2}{2}]$$

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

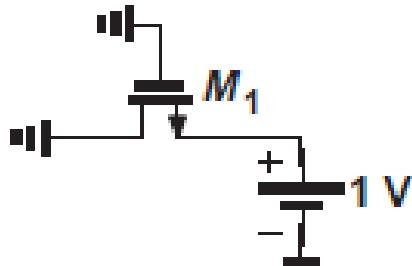


Cut-off region: $I_D = 0$

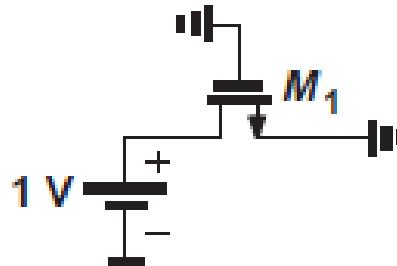
Linear region: $I_D = \mu_n C_{ox} \frac{W}{L} [(V_{GS} - V_t)V_{DS} - \frac{V_{DS}^2}{2}]$

Saturation region: $I_D = \mu_n C_{ox} \frac{W}{2L} (V_{GS} - V_t)^2$

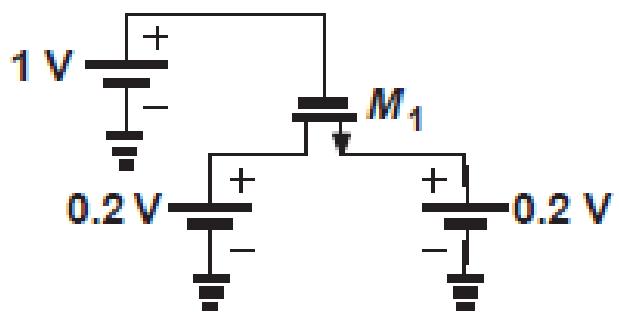
Determine the region of operation of M_1 in each of the circuits shown in Fig.



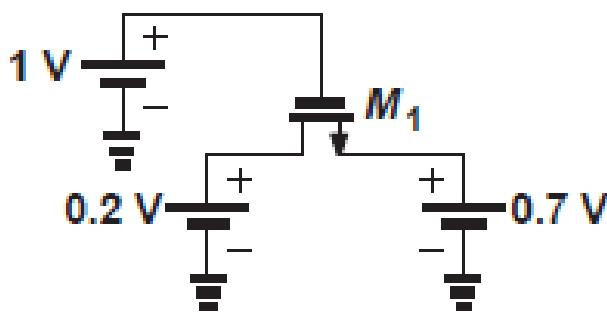
(a)



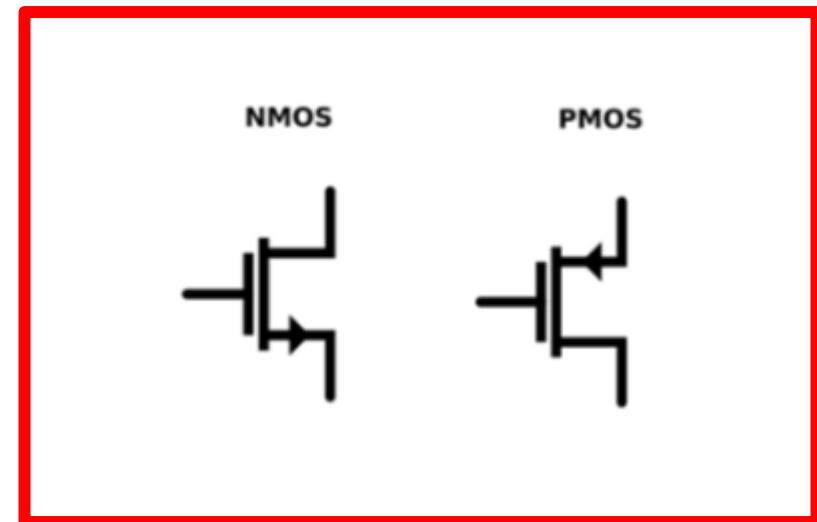
(b)



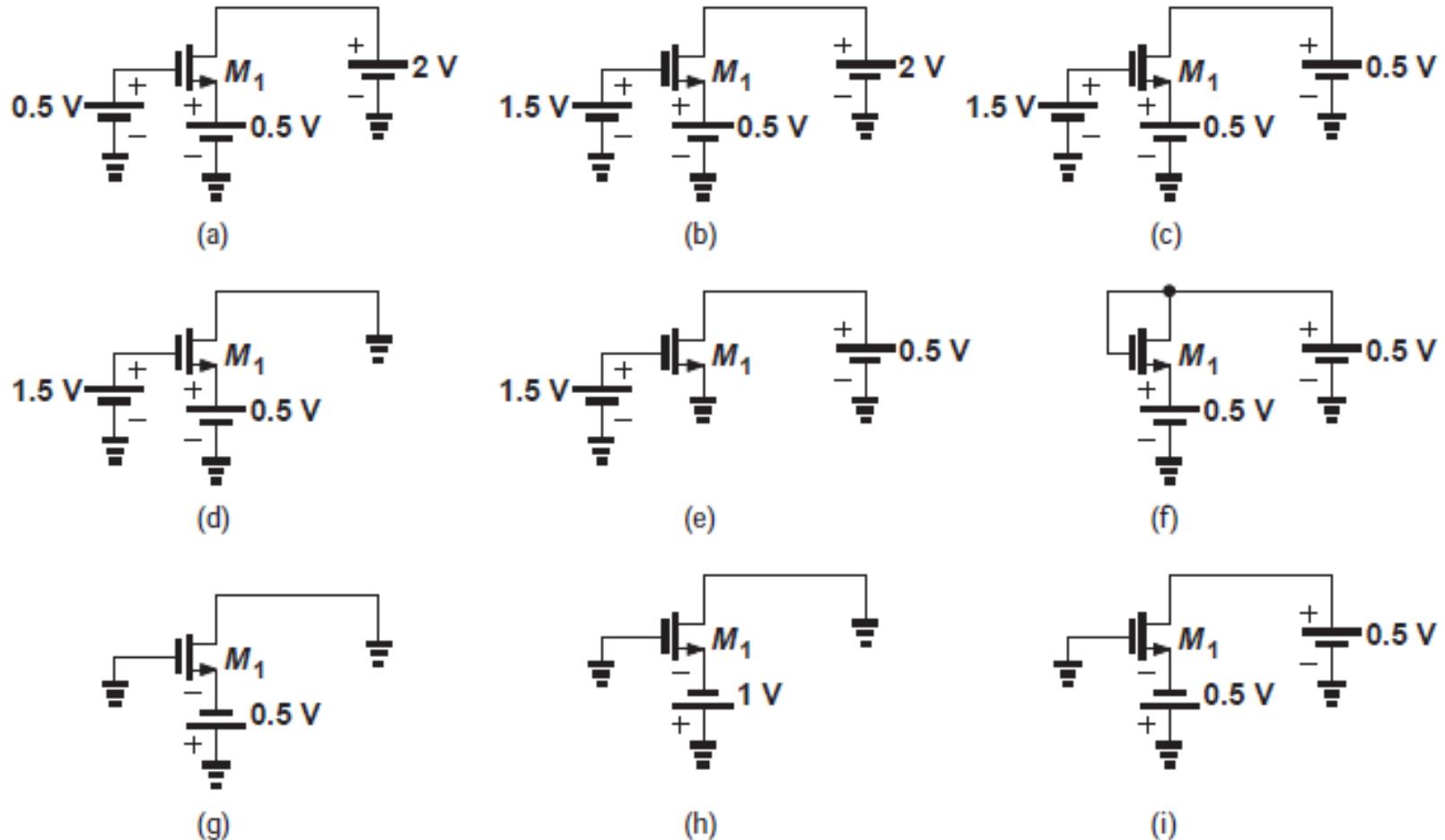
(c)



(d)



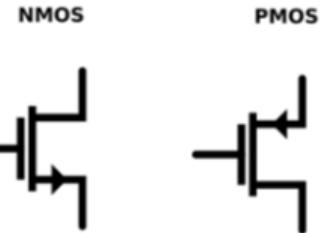
$$V_{TH} = 0.4 \text{ V for NMOS devices and } -0.4 \text{ V for PMOS devices.}$$



$V_{TH} = 0.4$ V for NMOS devices and -0.4 V for PMOS devices.



Determine the region of operation of M_1 in each of the circuits shown in Fig.

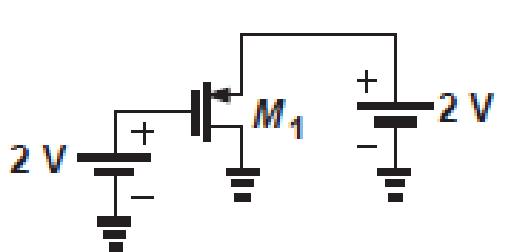


$V_{SG} < |V_{th}| \rightarrow \text{Cut-off}$

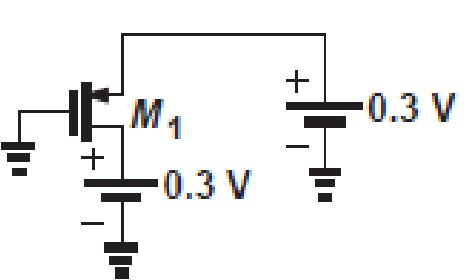
$V_{SG} > |V_{th}| \& V_{SD} < (V_{SG} - |V_{th}|) \rightarrow \text{Triode region}$

$V_{SG} > |V_{th}| \& V_{SD} = (V_{SG} - |V_{th}|) \rightarrow \text{Edge of saturation region}$

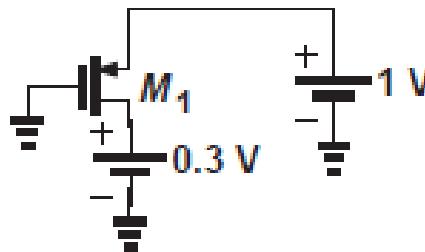
$V_{SG} > |V_{th}| \& V_{SD} > (V_{SG} - |V_{th}|) \rightarrow \text{Saturation region}$



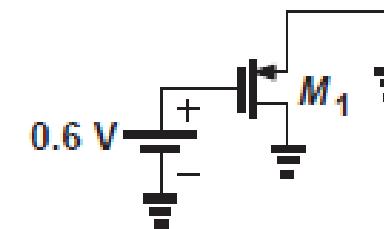
(a)



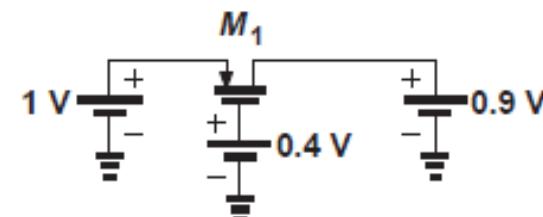
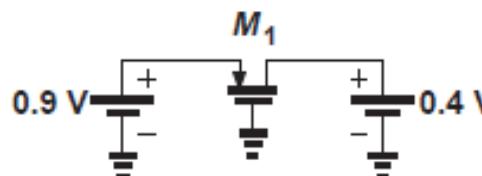
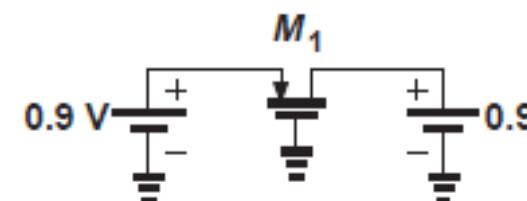
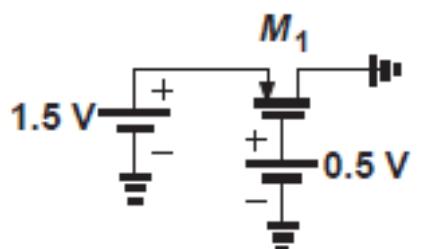
(b)



(c)



(d)



$$V_{th} = 0.4 \text{ V for NMOS devices and } -0.4 \text{ V for PMOS devices.}$$



1. A 0.18- μm fabrication process is specified to have $t_{\text{ox}} = 4\text{nm}$, $\mu_n = 450\text{cm}^2/\text{Vs}$ and $V_{\text{Th}} = 0.5\text{V}$. Find the value of $\mu_n C_{\text{ox}}$ (Also known as k_n' process transconductance) For a MOSFET with minimum length fabricated in this process, find the required value of W so that the device exhibits a channel resistance r_{DS} of 1K at $V_{\text{GS}} = 1\text{V}$. **Given** $\epsilon_{\text{ox}} = 3.45 \times 10^{-11} \text{ F/m}$.

$$K_n' = 388 \mu\text{A/V}^2$$

$$W = 0.93 \mu\text{m}$$



2. Consider a process technology for which $L_{min} = 0.4 \mu m$, $t_{ox} = 8 nm$, $\mu_n = 450 \text{ cm}^2/V \cdot s$, and $V_{th} = 0.7 V$.

(a) Find C_{ox} and K_n' . $C_{ox} = 4.31 \times 10^{-3} F/m^2$ $K_n' = 194.1 \mu A/V^2$

(b) For a MOSFET with $W/L = 8 \mu m / 0.8 \mu m$ calculate the values of V_{OV} , V_{GS} , and V_{DSmin} needed to operate the transistor in the saturation region with a dc current $I_D = 100 \mu A$. $V_{DSmin} = V_{OV} = 0.32 V$ $V_{GS} = 1.015 V$

(c) For the device in (b), find the values of V_{OV} and V_{GS} required to cause the device to operate as a 1000-resistor for very small v_{Ds} $V_{OV} = 0.51 V$ $V_{GS} = 1.215 V$



3. For a 0.8- μm process technology for which $t_{\text{ox}} = 15 \text{ nm}$ and $\mu_n = 550 \text{ cm}^2/\text{V}\cdot\text{s}$, find C_{ox} , K'_n , and the overdrive voltage V_{OV} required to operate a transistor having $W/L=20$ in saturation with $I_D = 0.2 \text{ mA}$. What is the minimum value of V_{DS} needed?

$$C_{\text{ox}} = 2.301 \times 10^{-3} \text{ F/m}^2 \quad K'_n = 126.5 \mu\text{A/V}^2 \quad V_{DS\min} = V_{\text{OV}} = .397 \text{ V} \approx 0.4 \text{ V}$$

4. A circuit designer intending to operate a MOSFET in saturation is considering the effect of changing the device dimensions and operating voltages on the drain current I_{DS} . Specifically, by what factor does I_{DS} change in each of the following cases?

- (a) The channel length is doubled.
- (b) The channel width is doubled.
- (c) The overdrive voltage is doubled.
- (d) The drain-to-source voltage is doubled.

Ans. 0.5; 2; 4; no change



5. An enhancement type NMOS transistor with $V_{TH} = 0.7V$ has its source terminal grounded and a 1.5 V DC is applied to the gate. In what region does the device operate :for (a) $V_D = +0.5 V$ (b) $V_D = +0.9 V$ (c) $V_D = +3 V$

If the NMOS device in $\mu_n C_{ox} = 100 \mu\text{A/V}^2$, $W = 10 \mu\text{m}$, and $L = 1 \mu\text{m}$, find the value of drain current that results in each of the three cases (a), (b), and (c).

- (a) Non Satⁿ→275 μA
- (b) Satⁿ→320 μA
- (c) Satⁿ→320 μA



Second-order Effects (Non-ideal I-V effects)

1. Velocity saturation and mobility degradation.
2. Channel length modulation
3. Body effect
4. Subthreshold conduction
5. Junction leakage.
6. Tunnelling
7. Temperature dependence
8. Geometry dependence



1. Velocity saturation and mobility degradation

At high lateral field strengths (V_{ds}/L), carrier velocity ceases to increase linearly with field strength.

This is called **velocity saturation**.

$$v = \begin{cases} \frac{\mu_{\text{eff}} E}{1 + \frac{E}{E_c}} & E < E_c \\ v_{\text{sat}} & E \geq E_c \end{cases}$$

Where E is electric field between the drain and source.
The **critical electric** field is

$$E_c = \frac{2v_{\text{sat}}}{\mu_{\text{eff}}}$$

At strong vertical electric fields resulting from large V_{gs} causes the carrier to scatter against the surface and also reduce the carrier mobility.

This effect is called **mobility degradation**.

$$\mu_{\text{eff}-n} = \frac{540 \frac{\text{cm}^2}{\text{V} \cdot \text{s}}}{1 + \left(\frac{V_{gs} + V_t}{0.54 \frac{\text{V}}{\text{nm}} t_{\text{ox}}} \right)^{1.85}}$$

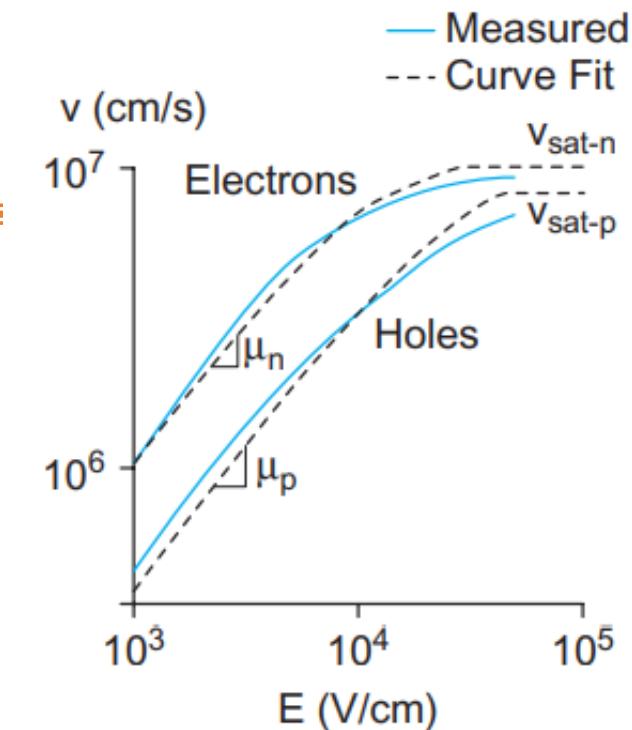
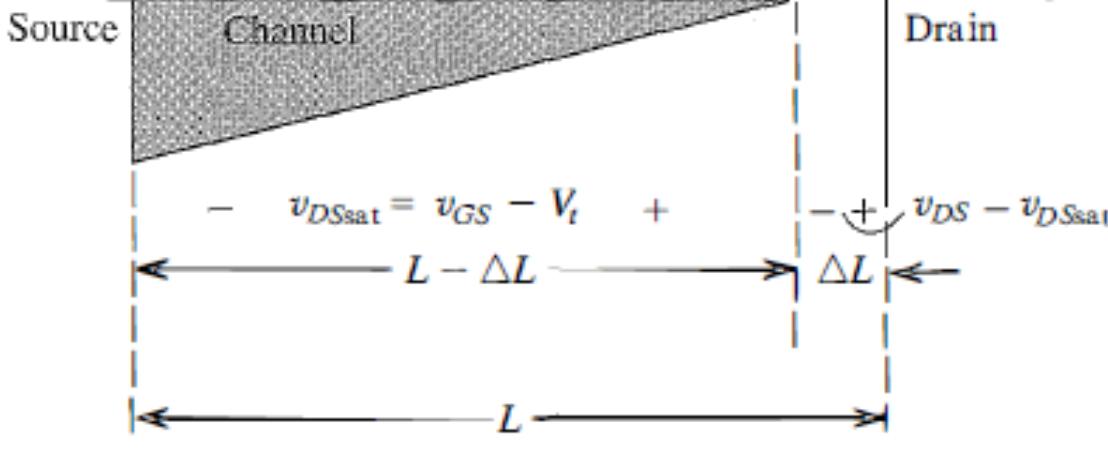


FIGURE 2.15 Carrier velocity vs. electric field at 300 K, adapted from [Jacoboni77]. Velocity saturates at high fields.

$$\mu_{\text{eff}-p} = \frac{185 \frac{\text{cm}^2}{\text{V} \cdot \text{s}}}{1 + \left| \frac{V_{gs} + 1.5V_t}{0.338 \frac{\text{V}}{\text{nm}} t_{\text{ox}}} \right|^{1.85}}$$



2. Channel length modulation

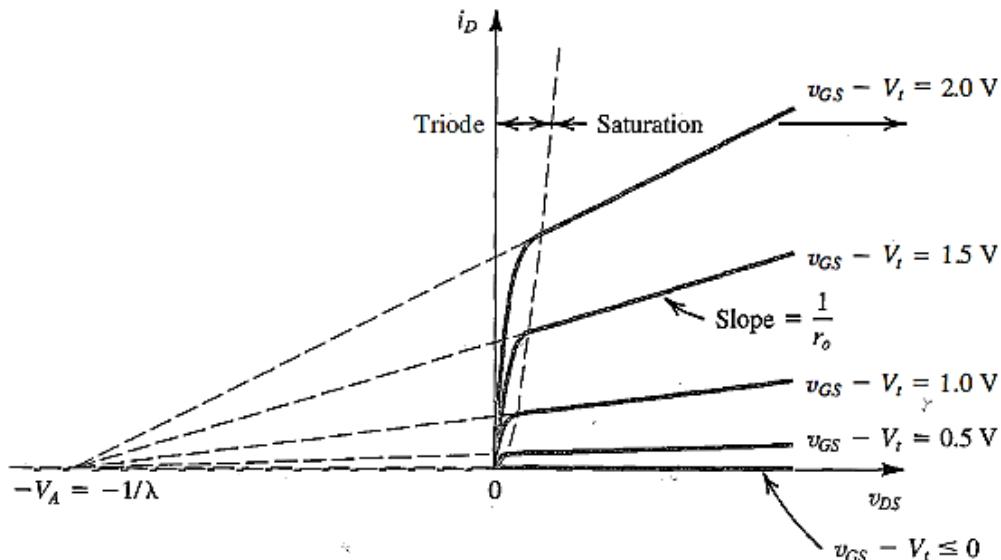
Increasing V_{ds} decreases the effective channel length. Shorter channel length results in higher current; thus, I_{ds} increases with V_{ds} in saturation

$$\begin{aligned} i_D &= \frac{1}{2} k'_n \frac{W}{L - \Delta L} (v_{GS} - V_t)^2 \\ &\equiv \frac{1}{2} k'_n \frac{W}{L} \frac{1}{1 - (\Delta L/L)} (v_{GS} - V_t)^2 \\ &\cong \frac{1}{2} k'_n \frac{W}{L} \left(1 + \frac{\Delta L}{L}\right) (v_{GS} - V_t)^2 \\ \Delta L &= \lambda' v_{DS} \end{aligned}$$

$$\begin{aligned} i_D &= \frac{1}{2} k'_n \frac{W}{L} \left(1 + \frac{\lambda'}{L} v_{DS}\right) (v_{GS} - V_t)^2 \\ \lambda &= \frac{\lambda'}{L} \end{aligned}$$

λ is a process-technology parameter with the dimensions of V^{-1}

$$i_D = \frac{1}{2} k'_n \frac{W}{L} (v_{GS} - V_t)^2 (1 + \lambda v_{DS})$$



$$i_D = \frac{1}{2} k'_n \frac{W}{L} (v_{GS} - V_t)^2 (1 + \lambda v_{DS})$$

$$I_{ds} = \frac{\beta}{2} V_{GT}^2 \left(1 + \frac{V_{ds}}{V_A} \right)$$

where V_A is called the Early voltage.
 β is the current drive capability ratio

$$\beta = \mu C_{ox} \frac{W}{L}$$

The term $V_{gs} - V_t$ arises so often that it is convenient to abbreviate it as V_{GT}



The enhancement type NMOS transistor is biased in the saturation region. The measured drain current is 2mA when drain to source voltage is 4V. When drain to source voltage is increased to 5V, while keeping the gate to source voltage same, the drain current is increased to 2.05mA. Calculate the channel length modulation parameter (λ)

Ans: $\lambda=0.0277 \text{ V}^{-1}$



3. Body effect

- When a voltage V_{sb} is applied between the source and body, it increases the amount of charge required to invert the channel, hence, it increases the threshold voltage. The potential difference between source and body V_{sb} affects the threshold voltage.

The threshold voltage is given by,

$$\phi_s = 2v_T \ln \frac{N_A}{n_i} \quad (2.31)$$

$$V_t = V_{t0} + \gamma \left(\sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s} \right) \quad (2.30)$$

$$\gamma = \frac{t_{ox}}{\epsilon_{ox}} \sqrt{2q\epsilon_{si} N_A} = \frac{\sqrt{2q\epsilon_{si} N_A}}{C_{ox}} \quad (2.32)$$

Where V_{t0} is threshold voltage when S is at body potential

ϕ_s is the surface potential at threshold. γ is the *body effect coefficient*, typically in the range 0.4 to 1 V^{1/2}.

N_A is doping level of the substrate, n_i is the doping level of the wells

At room temperature the thermal voltage $v_T = kT/q = 26$ mV :



Consider the nMOS transistor in a 65 nm process with a nominal threshold voltage of 0.3 V and a doping level of $8 \times 10^{17} \text{ cm}^{-3}$. The body is tied to ground with a substrate contact. How much does the threshold change at room temperature if the source is at 0.6 V instead of 0? Given $n_i = 1.45 \times 10^{10} \text{ cm}^{-3}$, $t_{ox} = 10.5 \times 10^{-8} \text{ cm}$,

SOLUTION: At room temperature, the thermal voltage $v_T = kT/q = 26 \text{ mV}$:

$$\phi_s = 2v_T \ln \frac{N_A}{n_i}$$

$$\gamma = \frac{t_{ox}}{\epsilon_{ox}} \sqrt{2q\epsilon_{si}N_A} = \frac{\sqrt{2q\epsilon_{si}N_A}}{C_{ox}}$$

$$V_t = V_{t0} + \gamma \left(\sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s} \right)$$

$$\phi_s = 2(0.026 \text{ V}) \ln \frac{8 \times 10^{17} \text{ cm}^{-3}}{1.45 \times 10^{10} \text{ cm}^{-3}} = 0.93 \text{ V}$$

$$\gamma = \frac{10.5 \times 10^{-8} \text{ cm}}{3.9 \times 8.85 \times 10^{-14} \frac{\text{F}}{\text{cm}}} \sqrt{2(1.6 \times 10^{-19} \text{ C})(11.7 \times 8.85 \times 10^{-14} \frac{\text{F}}{\text{cm}})(8 \times 10^{17} \text{ cm}^{-3})} = 0.16$$

$$V_t = 0.3 + \gamma \left(\sqrt{\phi_s + 0.6 \text{ V}} - \sqrt{\phi_s} \right) = 0.34 \text{ V}$$



4. Sub threshold conduction

Ideally nMOS will not allow the current to flow through it when $V_{gs} < V_t$.

But in real transistors, the current doesn't cut off below threshold.

This conduction is also known as leakage and flows even though the MOS is OFF.

$$I_{ds} = I_{ds0} e^{\frac{V_{gs} - V_{t0} + \eta V_{ds} - k\gamma V_{sb}}{nv_T}} \left(1 - e^{\frac{-V_{ds}}{v_T}} \right)$$
$$I_{ds0} = \beta v_T^2 e^{1.8} \quad \beta = \mu C_{ox} \frac{W}{L}$$

- I_{ds} is Subthreshold leakage current
- I_{ds0} is the current at threshold
- n is a process-dependent term affected by the depletion region characteristics typically in the range of 1.3–1.7 for CMOS processes
- The final term indicates that leakage is 0 if $V_{ds} = 0$,
- v_T is thermal voltage

5. Junction leakage

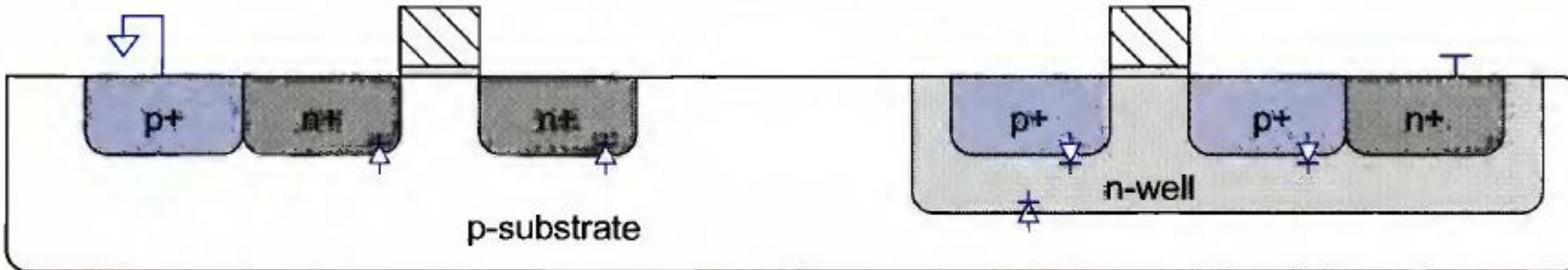


FIG 2.19 Reverse-biased diodes in CMOS circuits

diodes still conduct a small amount of current I_D .

$$I_D = I_s \left(e^{\frac{V_D}{v_T}} - 1 \right)$$

V_D is the diode voltage (e.g., $-V_{sb}$ or $-V_{db}$)

I_s depends on doping levels and on the area and perimeter of the diffusion region



6. Tunnelling

- Two physical mechanisms for gate tunneling are called **Fowler-Nordheim (FN) tunneling** and **direct tunneling**.
- FN tunneling is most important at high voltage and moderate oxide thickness.
- Direct tunneling is most important at lower voltage with thin oxides and is the dominant leakage component. It is given by the formula,

$$I_{\text{gate}} = WA \left(\frac{V_{DD}}{t_{\text{ox}}} \right)^2 e^{-B \frac{t_{\text{ox}}}{V_{DD}}}$$

where A and B are technology constants.

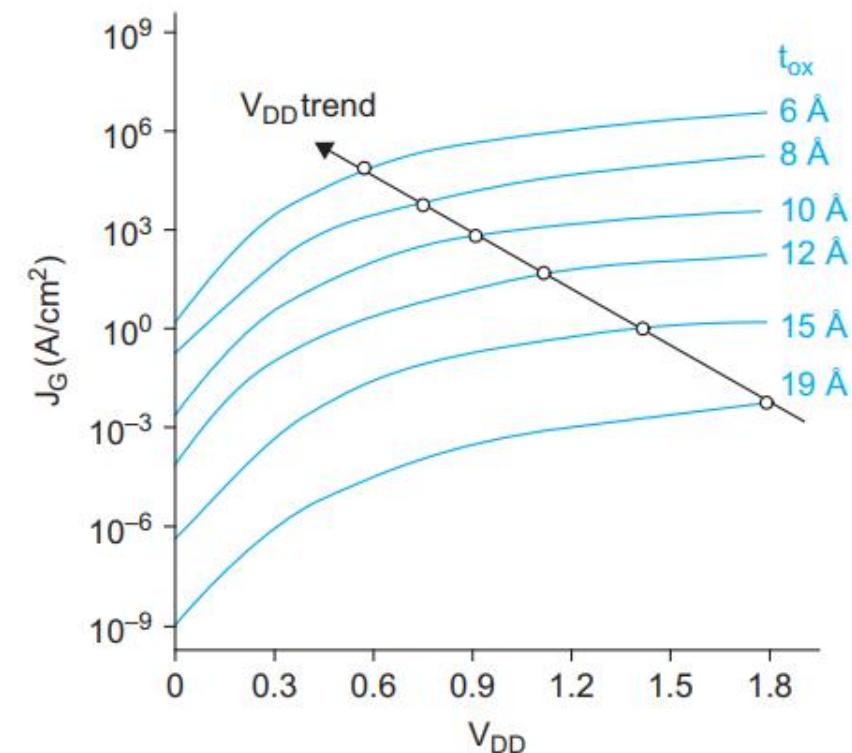


FIGURE 2.21 Gate leakage current from [Song01]

7. Temperature dependence

At low V_{gs} , the current has a positive temperature coefficient.

At high V_{gs} , the current has a negative temperature coefficient; i.e., it decreases with temperature.

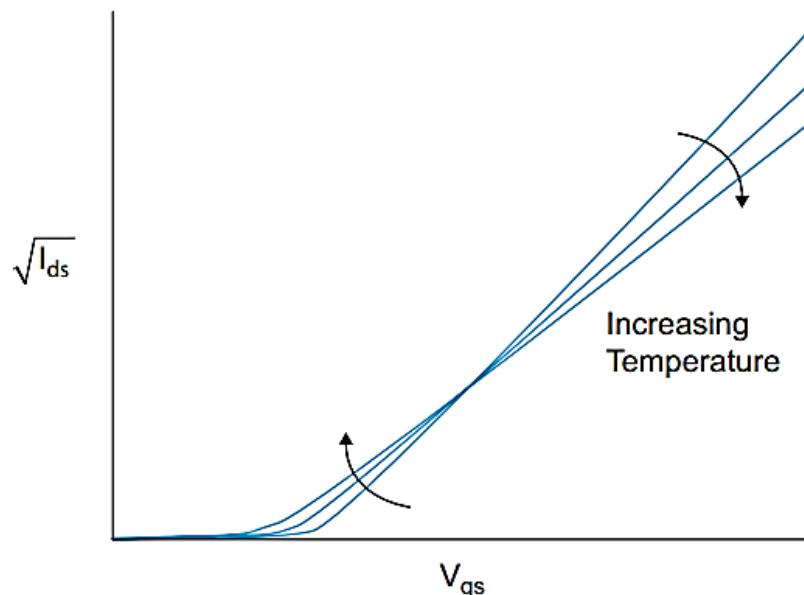


FIGURE 2.23 I-V characteristics of nMOS transistor in saturation at various temperatures

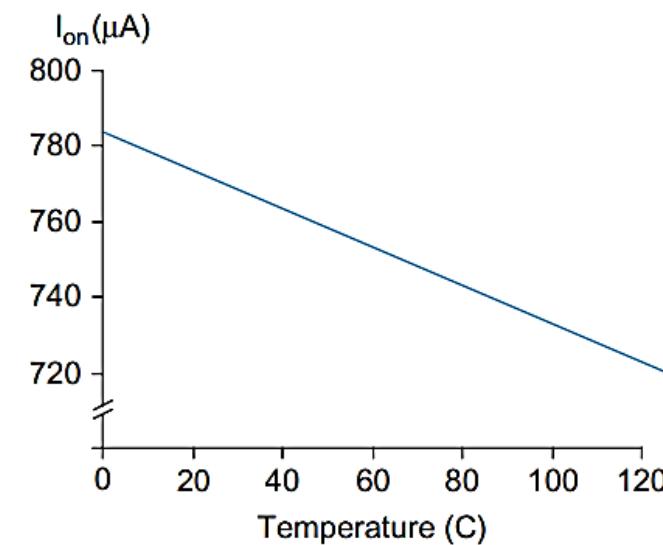


FIGURE 2.24 I_{dsat} vs. temperature



8. Geometry dependence

- The layout designer draws transistors with width and length W_{drawn} and L_{drawn} .
- The actual gate dimensions may differ by some factors X_W and X_L .
- Moreover, the source and drain tend to diffuse laterally under the gate by L_D , producing a shorter effective channel length.
- Similarly, W_D accounts for other effects that shrink the transistor width.
- Putting these factors together, we can compute effective transistor lengths and widths, as:

$$L_{\text{eff}} = L_{\text{drawn}} + X_L - 2L_D$$

$$W_{\text{eff}} = W_{\text{drawn}} + X_W - 2W_D$$

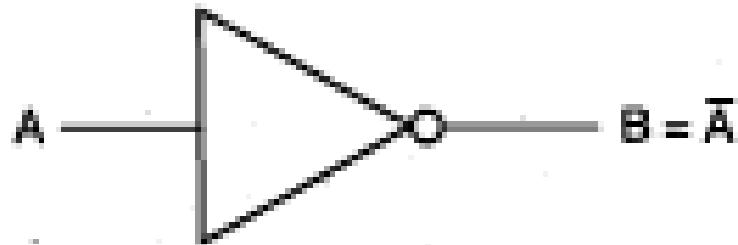


Inverters

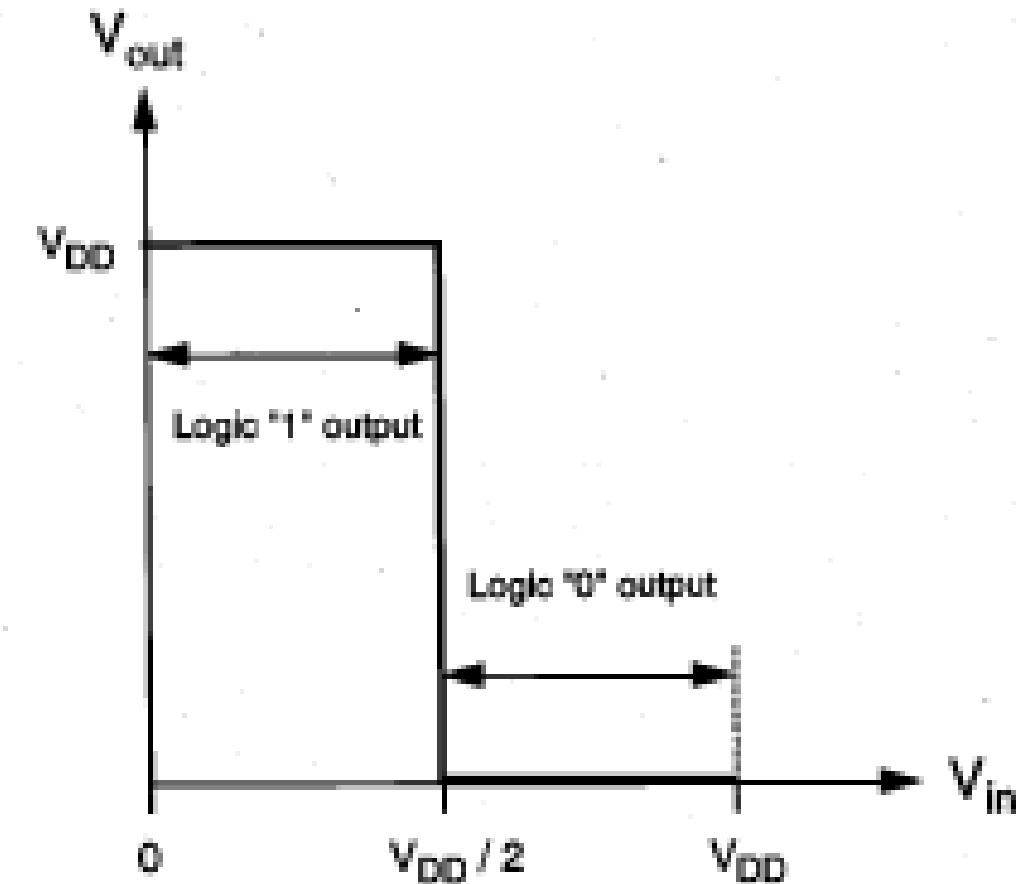
1. Ideal inverter
2. Resistive load inverter
3. n-type MOSFET load
 - i. Enhancement load inverter
 - ii. Depletion load inverter
4. CMOS inverter



1. Ideal inverter



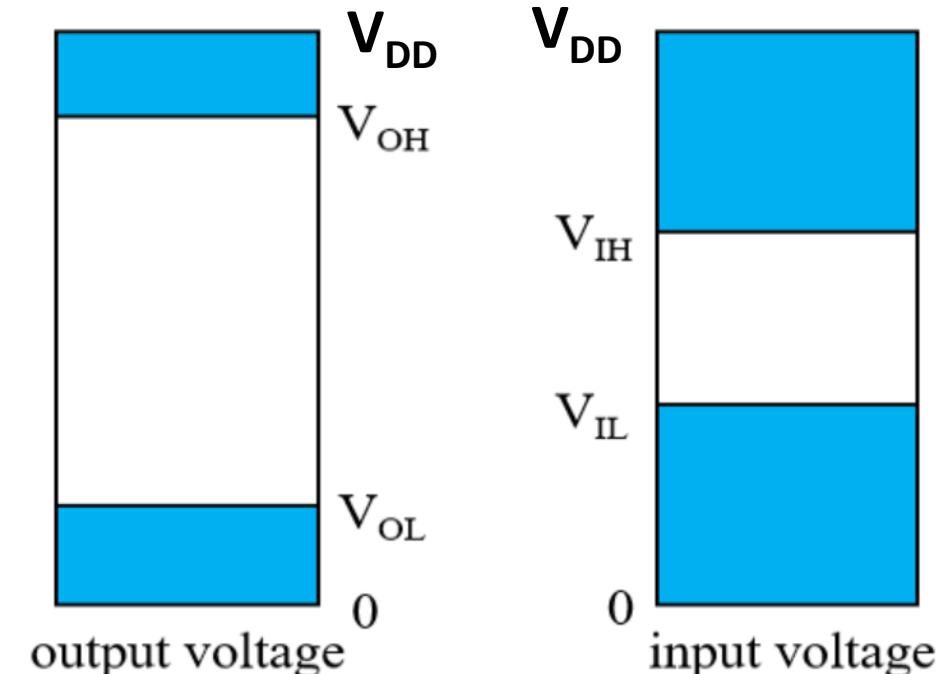
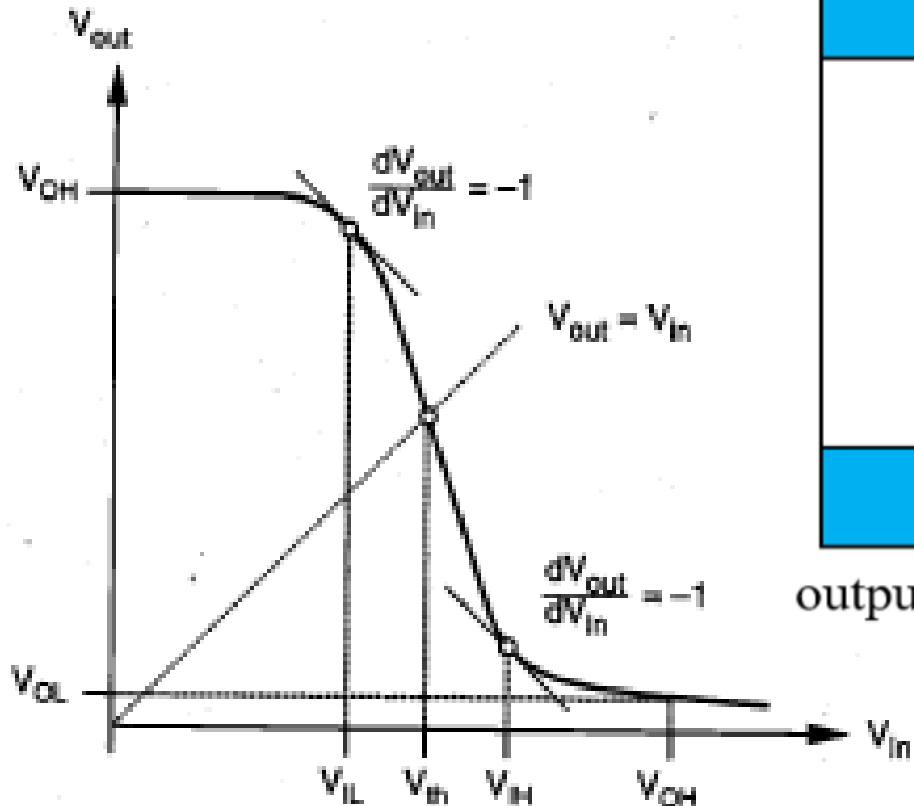
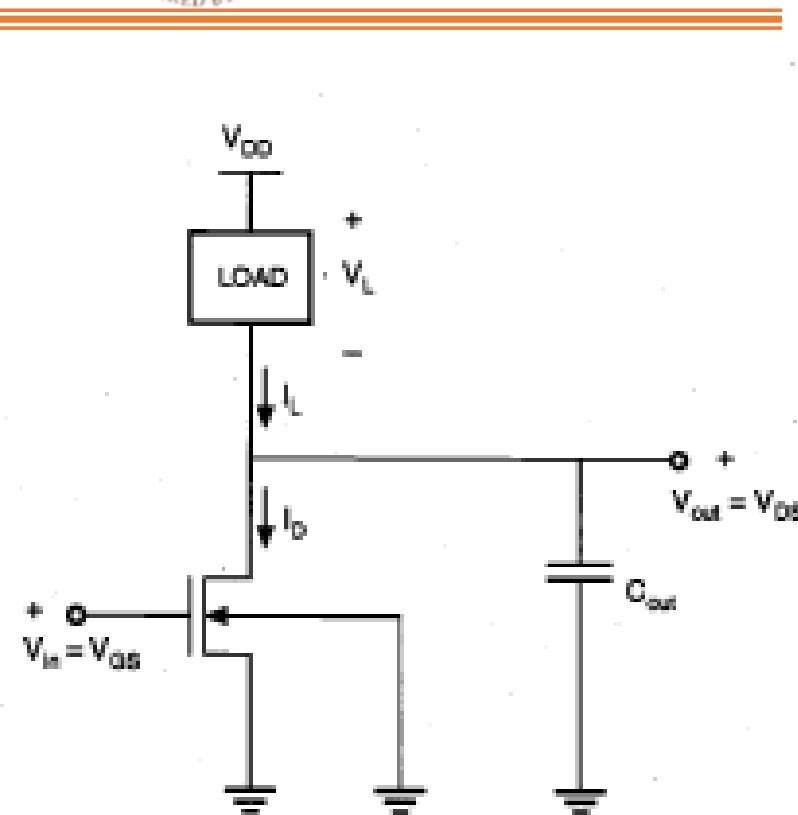
A	B
0	1
1	0



DC Voltage transfer characteristics (VTC)
of an ideal inverter



General circuit of an nMOS inverter



V_{OH} : Minimum output voltage when the output level is logic "1"

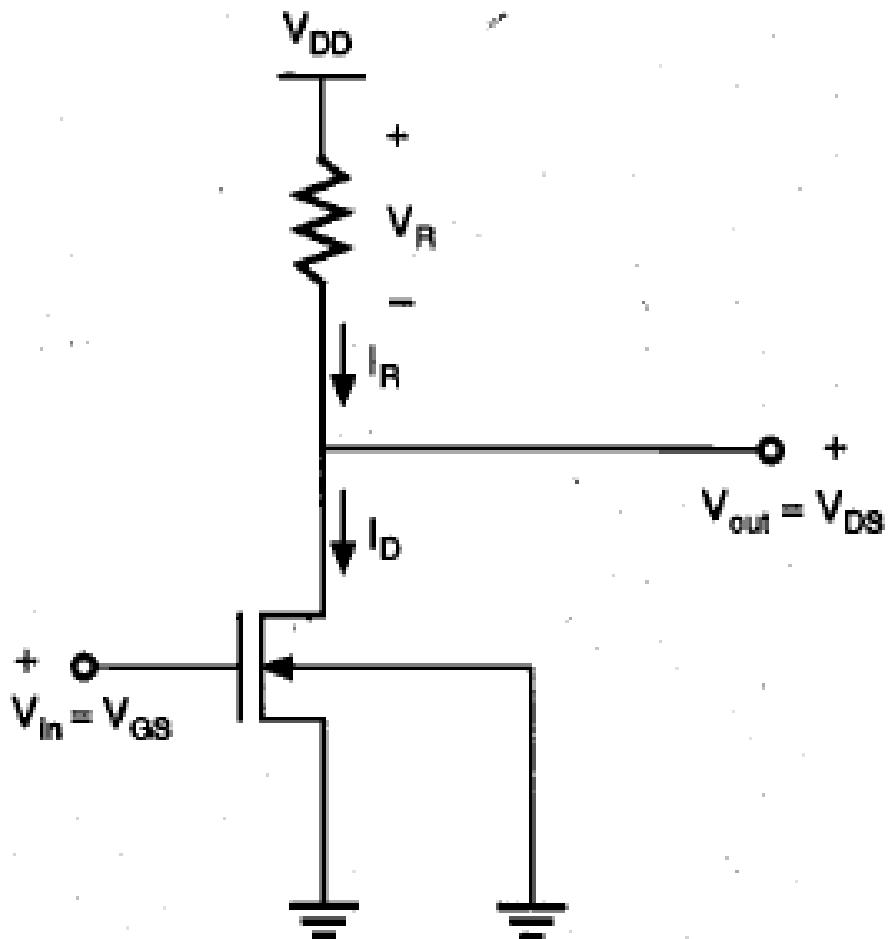
V_{OL} : Maximum output voltage when the output level is logic "0"

V_{IH} : Minimum input voltage which can be *interpreted* as logic "1"

V_{IL} : Maximum input voltage which can be *interpreted* as logic "0"



2. Resistive load Inverter



output voltage V_{out} is

$$V_{out} = V_{DD} - I_R R$$

$$V_{out} = V_{DD} - I_D R$$

$$I_D = \frac{V_{DD} - V_{out}}{R}$$

So, Drain current equation will be

- if MOSFET is there in saturation region then

$$V_{in} - V_{To} < V_{out} \text{ and } I_D = \frac{K}{2}(V_{GS} - V_{TO})^2$$

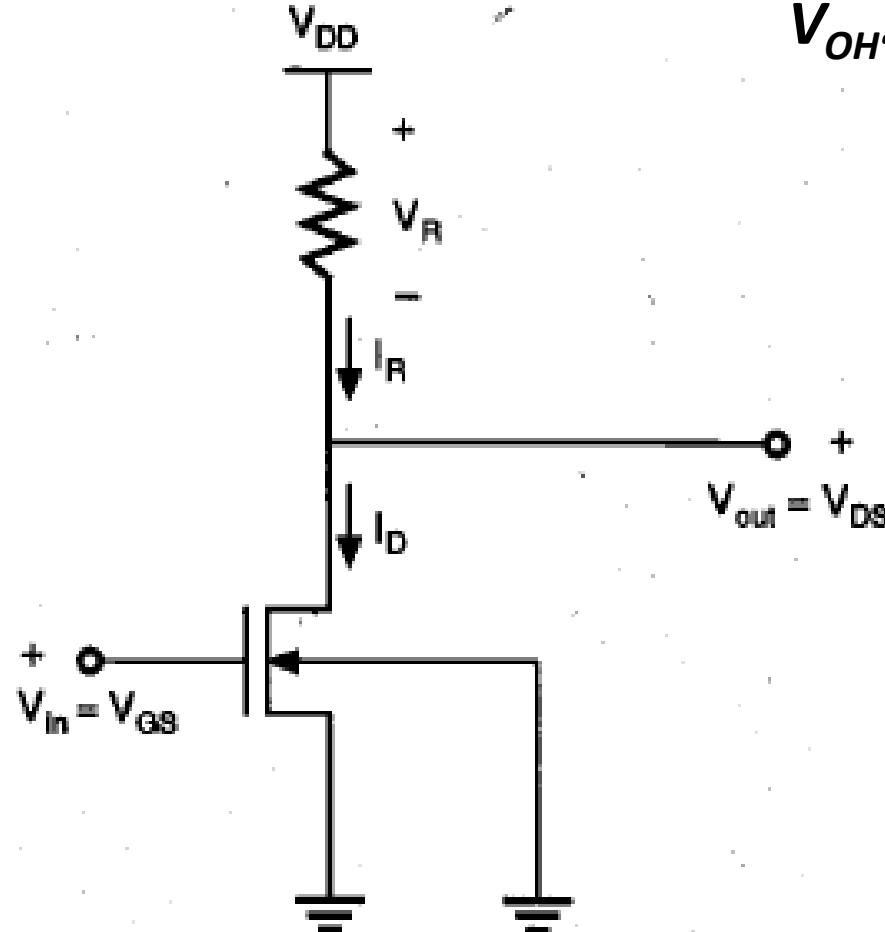
- If MOSFET is there in linear region then

$$V_{in} - V_{To} > V_{out} \text{ and } I_D = \frac{K}{2}[2(V_{GS} - V_{TO})V_{DS} - V_{DS}^2]$$

Input Voltage Range	Operating Mode
$V_{in} < V_{To}$	cut-off
$V_{To} \leq V_{in} < V_{out} + V_{To}$	saturation
$V_{in} \geq V_{out} + V_{To}$	linear



Calculation of V_{OH} , V_{OL} , V_{IL} , V_{IH}



V_{OH} : Minimum output voltage when the output level is logic " 1"

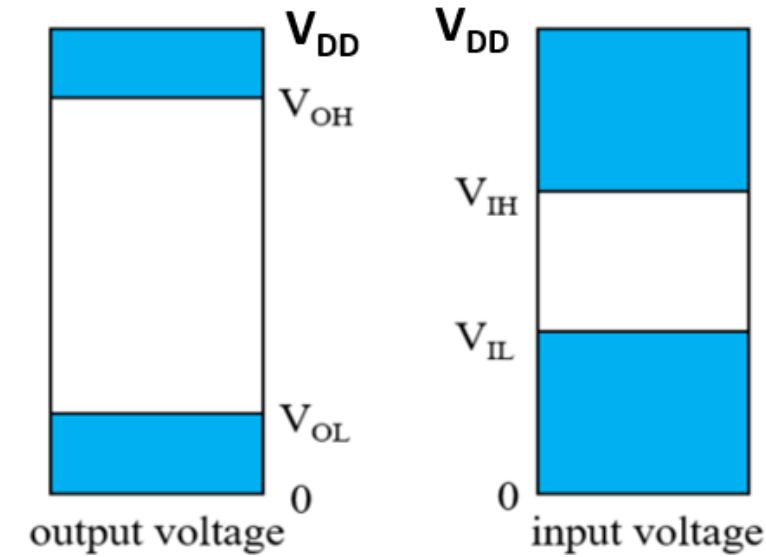
V_{OH} -

- Output voltage V_{out} is

$$V_{out} = V_{DD} - I_D R$$

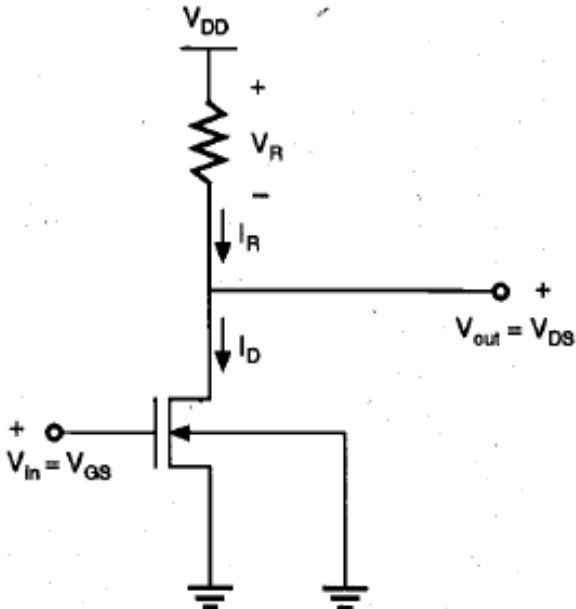
$$\Rightarrow V_{out} = V_{DD}$$

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





Calculation of V_{OH} , V_{OL} , V_{IL} , V_{IH}



K' is intrinsic gain parameter given by $k' = \mu C_{ox}$

K is current parameter given by $k = k'W/L$

V_{OL} : Maximum output voltage when the output level is logic "0"

- When $V_{in} - V_{TO} > V_{out}$, MOSFET is there in linear region, so, drain current will be

$$I_D = \frac{K}{2} [2(V_{in} - V_{TO})V_{out} - V_{out}^2]$$

- According to kirchoff's law in the drain current is

$$I_D = \frac{V_{DD} - V_{out}}{R}$$

- If we compare these two equations

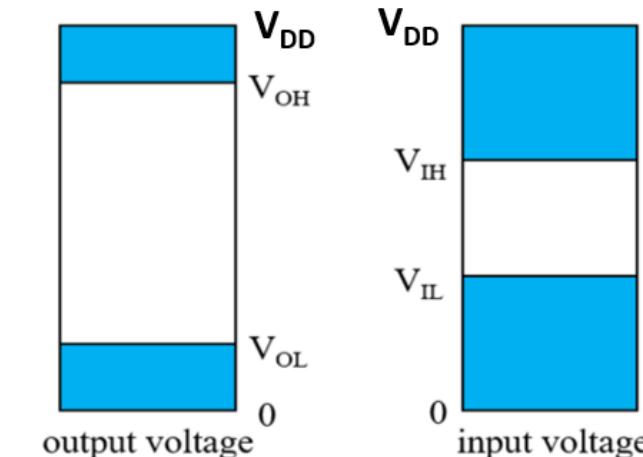
$$\frac{V_{DD} - V_{out}}{R} = \frac{K}{2} [2(V_{in} - V_{TO})V_{out} - V_{out}^2]$$

$$\Rightarrow \frac{V_{DD} - V_{OL}}{R} = \frac{K}{2} [2(V_{DD} - V_{TO})V_{OL} - V_{OL}^2]$$

$$\Rightarrow V_{OL}^2 - 2(V_{DD} - V_{TO} + \frac{1}{KR})V_{OL} + \frac{2}{KR}V_{DD} = 0$$

- If we solve the above equation, we get

$$V_{OL} = V_{DD} - V_{TO} + \frac{1}{KR} - \sqrt{(V_{DD} - V_{TO} + \frac{1}{KR})^2 - \frac{2V_{DD}}{KR}}$$



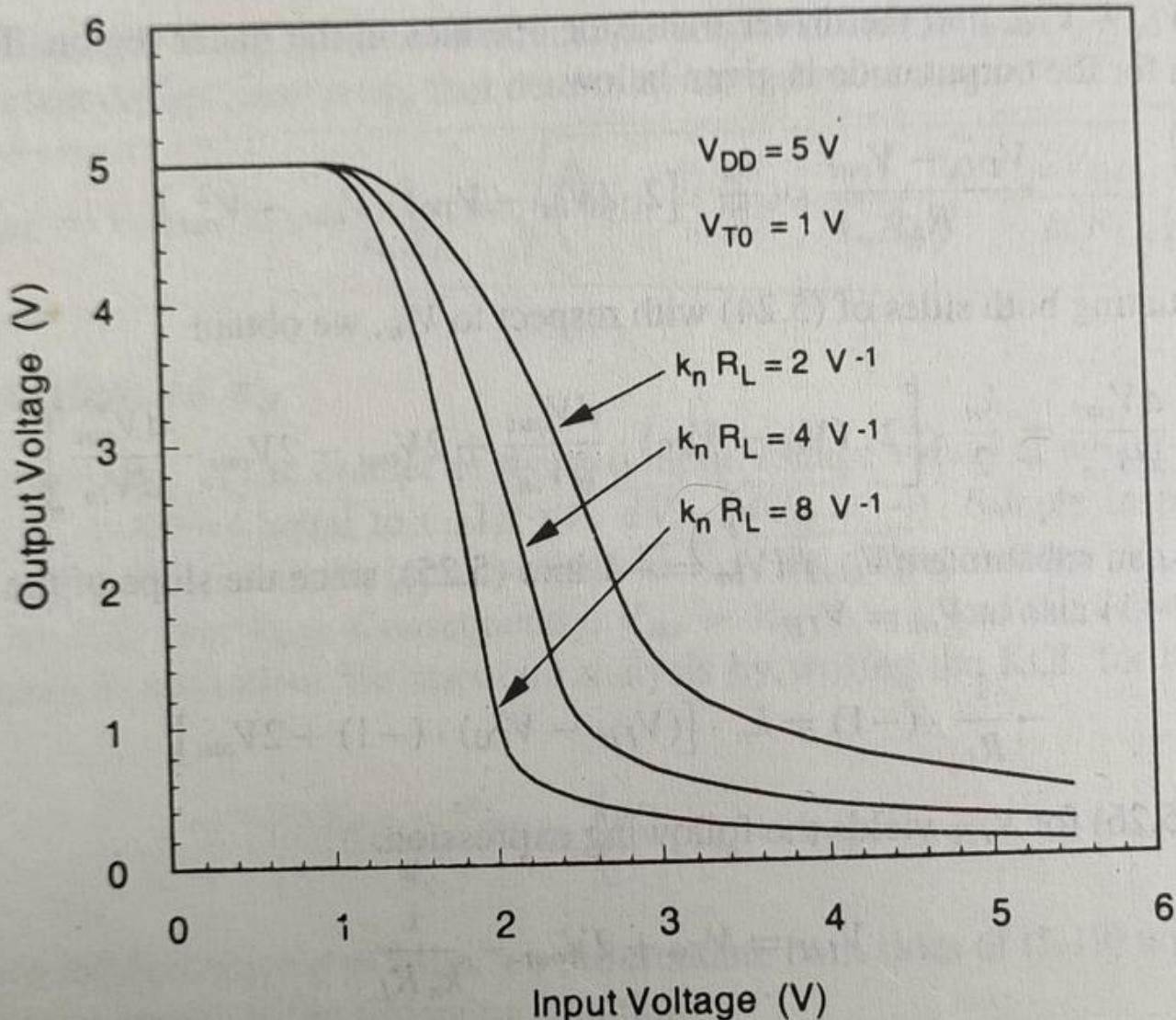
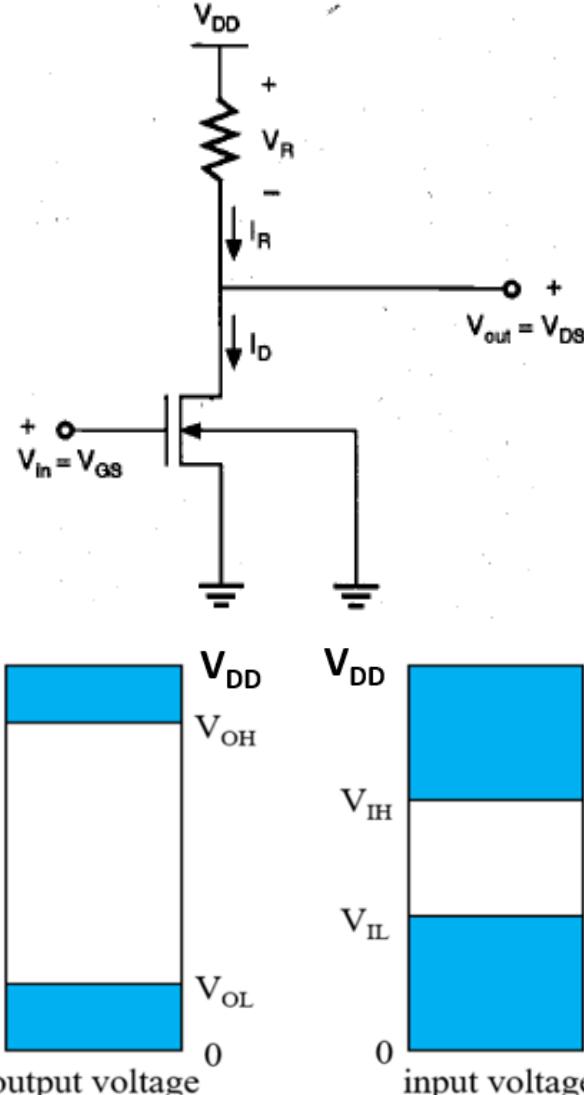


Figure 5.9 Voltage transfer characteristics of the resistive-load inverter, for different values of the parameter ($k_n R_L$).



Calculation of V_{OH} , V_{OL} , V_{IL} , V_{IH}



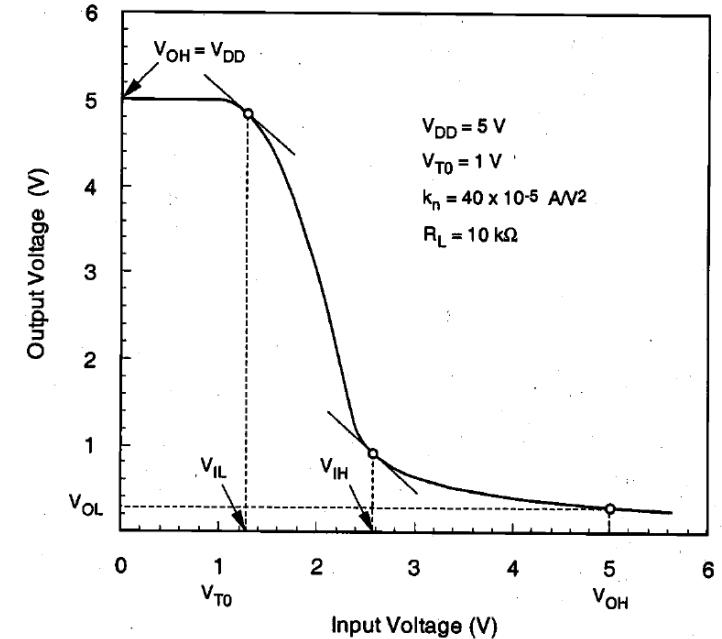
V_{IL} : Maximum input voltage which can be *interpreted* as logic "0"

$$\frac{V_{DD} - V_{out}}{R_L} = \frac{k_n}{2} \cdot (V_{in} - V_{T0})^2$$

$$-\frac{1}{R_L} \cdot \frac{dV_{out}}{dV_{in}} = k_n \cdot (V_{in} - V_{T0})$$

$$-\frac{1}{R_L} \cdot (-1) = k_n \cdot (V_{IL} - V_{T0})$$

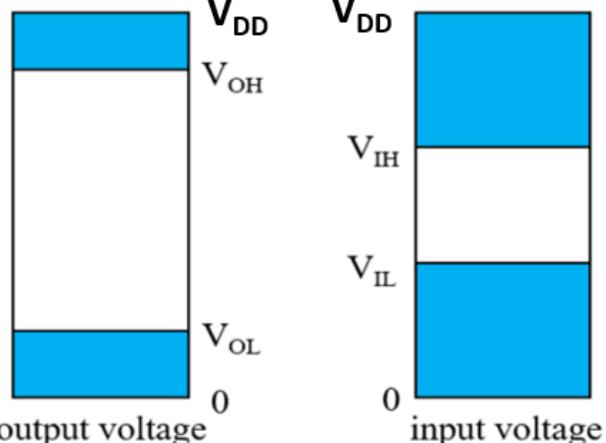
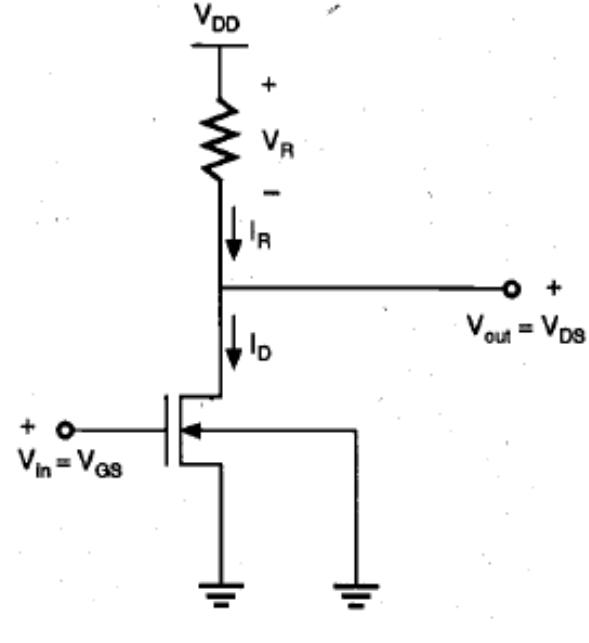
$$V_{IL} = V_{T0} + \frac{1}{k_n R_L}$$



Input Voltage Range	Operating Mode
$V_{in} < V_{T0}$	cut-off
$V_{T0} \leq V_{in} < V_{out} + V_{T0}$	saturation
$V_{in} \geq V_{out} + V_{T0}$	linear



Calculation of V_{OH} , V_{OL} , V_{IL} , V_{IH}



V_{IH} : Minimum input voltage which can be *interpreted* as logic "1"

$$\frac{V_{DD} - V_{out}}{R_L} = \frac{k_n}{2} \cdot [2 \cdot (V_{in} - V_{T0}) \cdot V_{out} - V_{out}^2] \quad \text{----(1)}$$

$$-\frac{1}{R_L} \cdot \frac{dV_{out}}{dV_{in}} = \frac{k_n}{2} \cdot \left[2 \cdot (V_{in} - V_{T0}) \cdot \frac{dV_{out}}{dV_{in}} + 2V_{out} - 2V_{out} \cdot \frac{dV_{out}}{dV_{in}} \right]$$

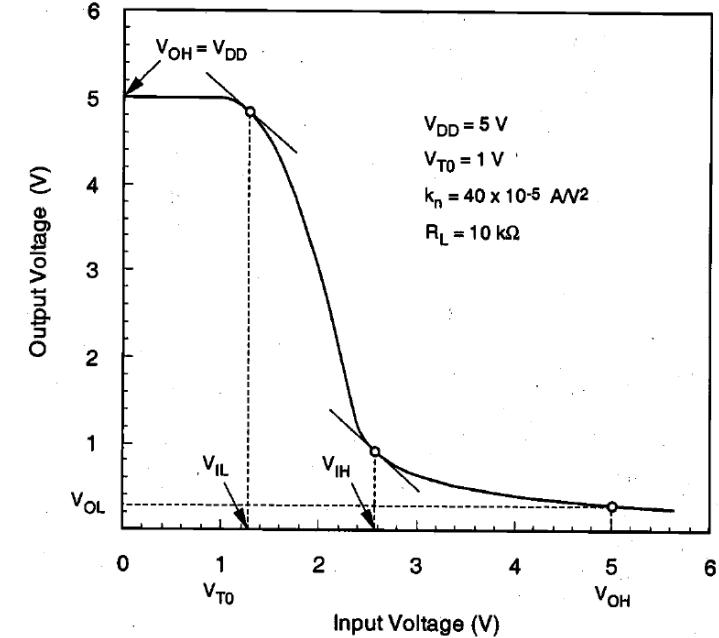
$$-\frac{1}{R_L} \cdot (-1) = k_n \cdot [(V_{IH} - V_{T0}) \cdot (-1) + 2V_{out}]$$

$$V_{IH} = V_{T0} + 2V_{out} - \frac{1}{k_n R_L} \quad \text{----(2)}$$

$$\frac{V_{DD} - V_{out}}{R_L} = \frac{k_n}{2} \cdot \left[2 \cdot \left(V_{T0} + 2V_{out} - \frac{1}{k_n R_L} - V_{T0} \right) \cdot V_{out} - V_{out}^2 \right]$$

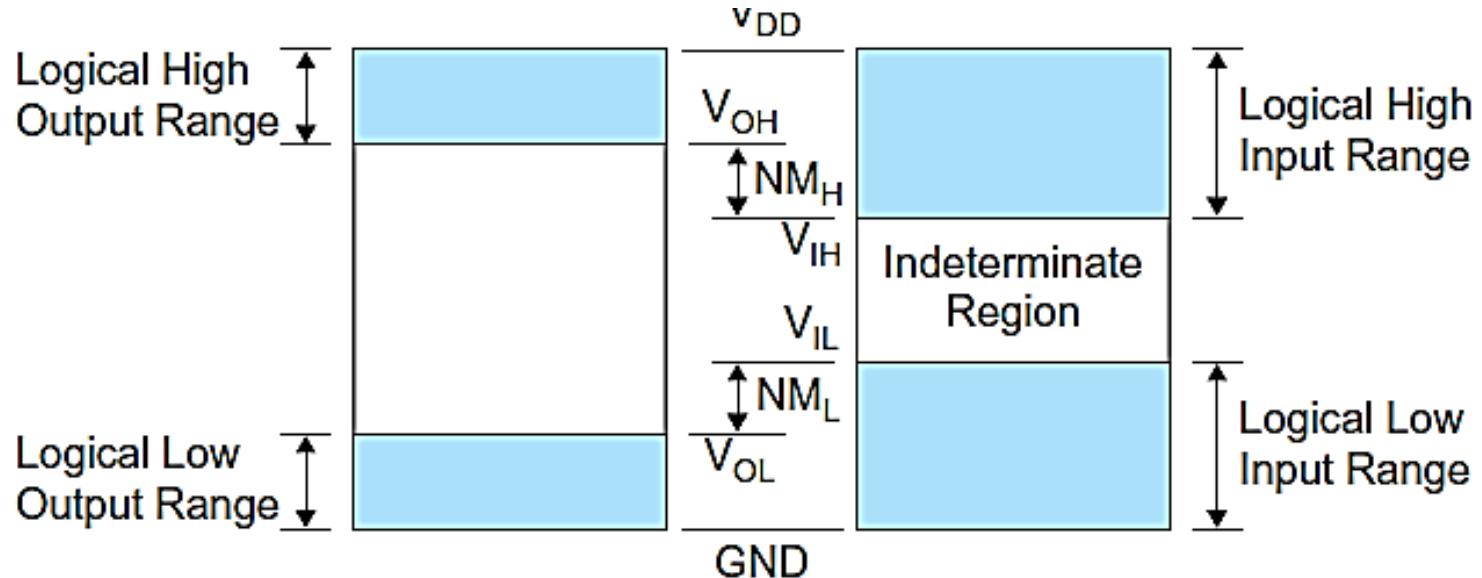
$$V_{out} (V_{in} = V_{IH}) = \sqrt{\frac{2}{3} \cdot \frac{V_{DD}}{k_n R_L}}$$

$$V_{IH} = V_{T0} + \sqrt{\frac{8}{3} \cdot \frac{V_{DD}}{k_n R_L} - \frac{1}{k_n R_L}}$$



Input Voltage Range	Operating Mode
$V_{in} < V_{T0}$	cut-off
$V_{T0} \leq V_{in} < V_{out} + V_{T0}$	saturation
$V_{in} \geq V_{out} + V_{T0}$	linear

Noise margin (NM_H and NM_L)



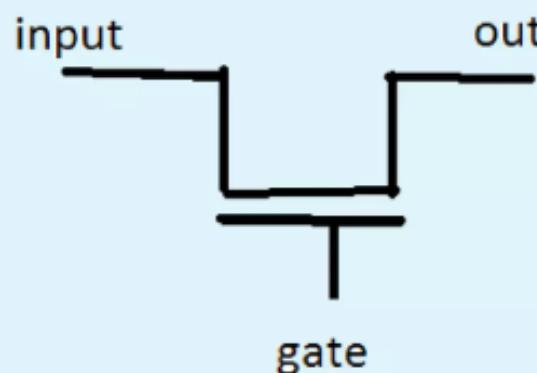
- Noise margin allows to determine the allowable noise voltage on the input of a gate so that the output will not be corrupted.
- The specification most commonly used to describe noise margin (or *noise immunity*) uses two parameters: the *LOW* noise margin, NM_L , and the *HIGH* noise margin, NM_H .

$$NM_L = V_{IL} - V_{OL}$$

$$NM_H = V_{OH} - V_{IH}$$

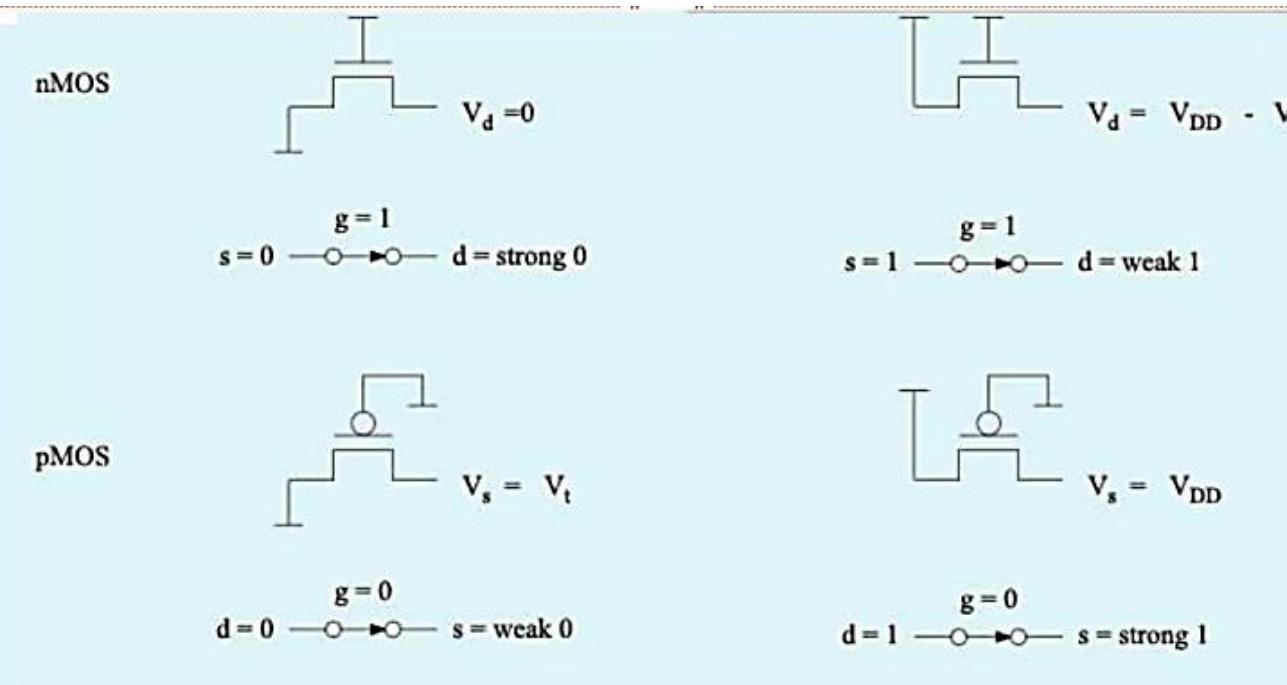
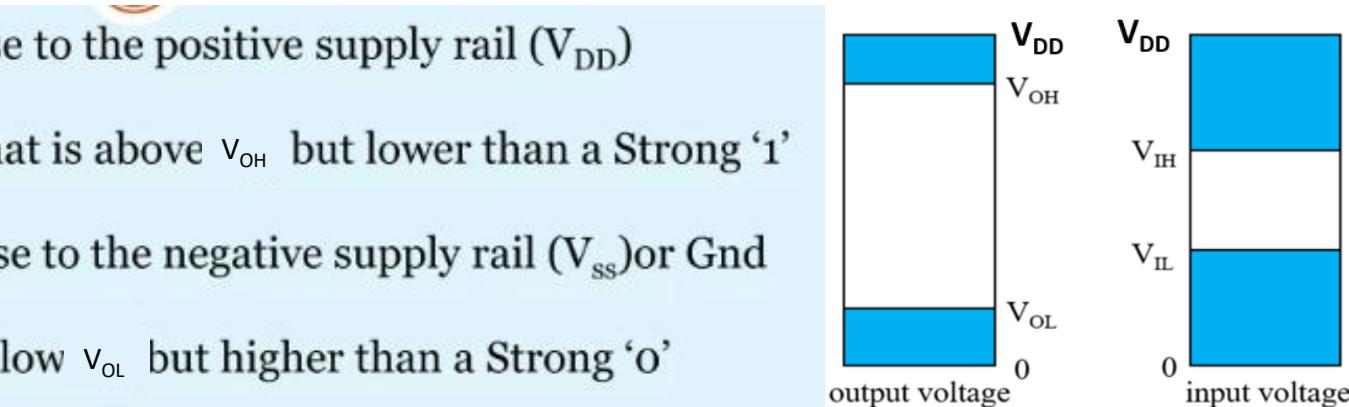
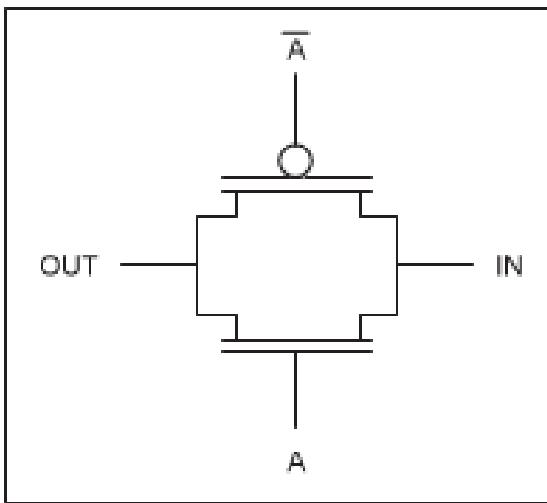


Pass transistor logic and Transmission gate logic



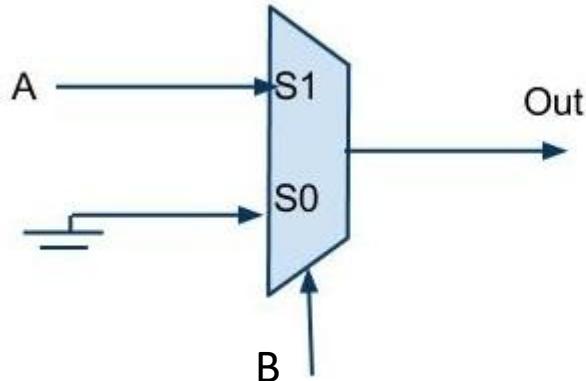
- Strong '1': An output very close to the positive supply rail (V_{DD})
- Weak '1': An output voltage that is above V_{OH} but lower than a Strong '1'
- Strong '0': An output very close to the negative supply rail (V_{ss}) or Gnd
- Weak '0': An output that is below V_{OL} but higher than a Strong '0'

nMOS passes a **STRONG 0** but **WEAK 1**
pMOS passes a **STRONG 1** but **WEAK 0**

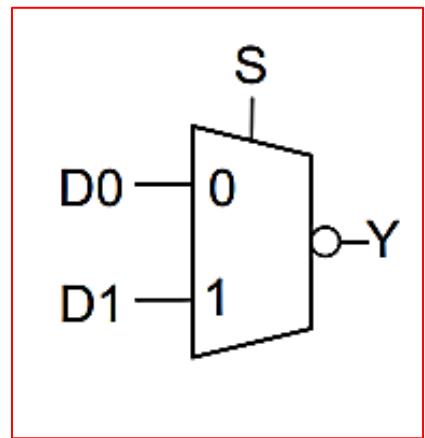
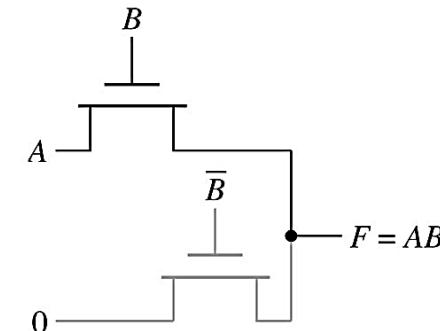




PTL and TGL - Example

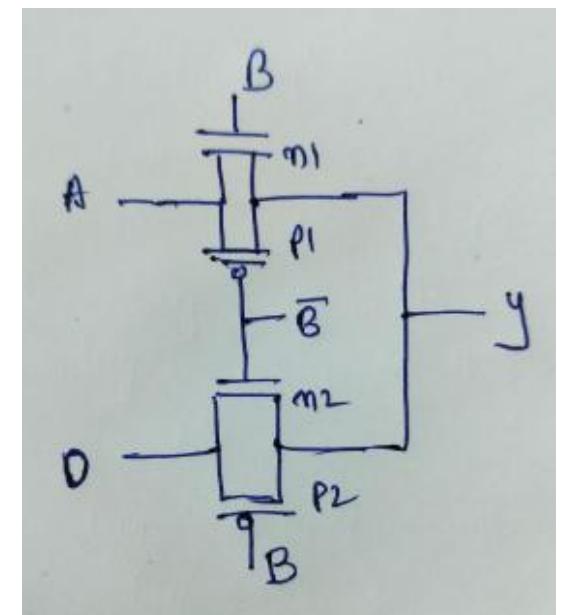


B	A	Out
0	0	0
0	1	0
1	0	0
1	1	1



$$Y = \bar{S} \cdot D0 + S \cdot D1.$$

	$V_{GS}=0$	$V_{GS}=1$
nMOS	OFF	ON
PMOS	ON	OFF

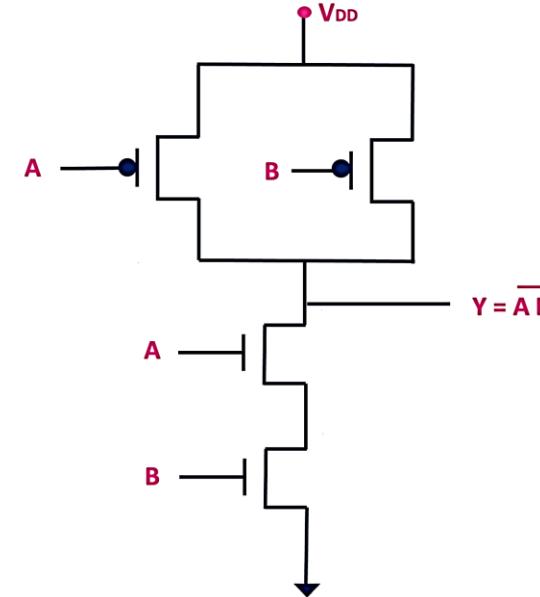
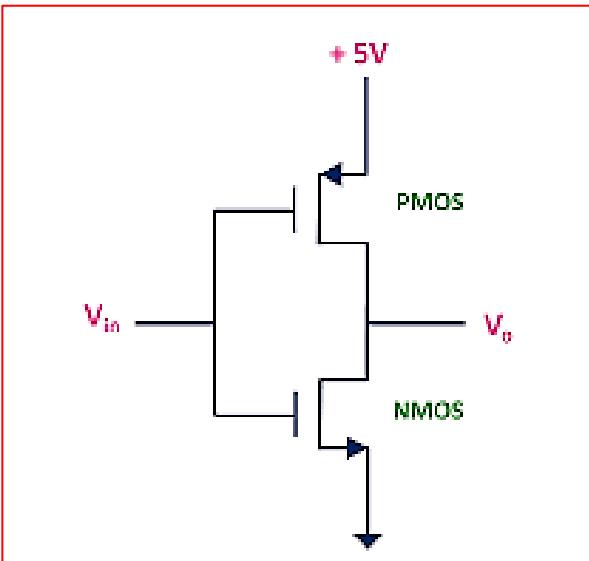




Basics of CMOS circuits

	$V_{GS}=0$	$V_{GS}=1$
nMOS	OFF	ON
PMOS	ON	OFF

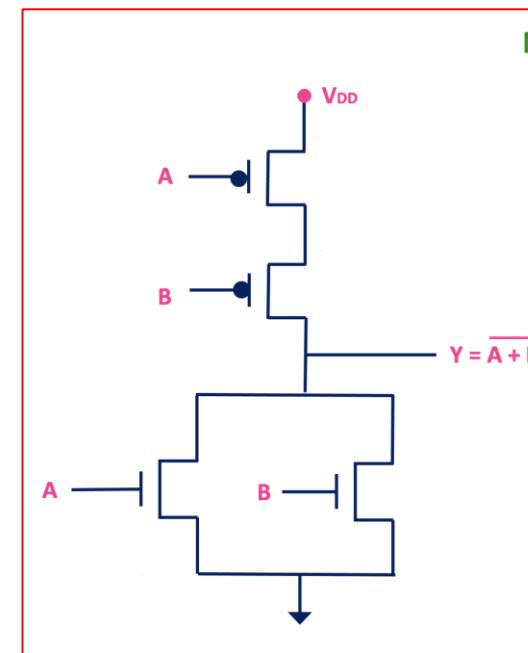
Operation	AND	OR
nMOS	Series	Parallel
PMOS	Parallel	Series



Truth Table

A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

NOR Gate

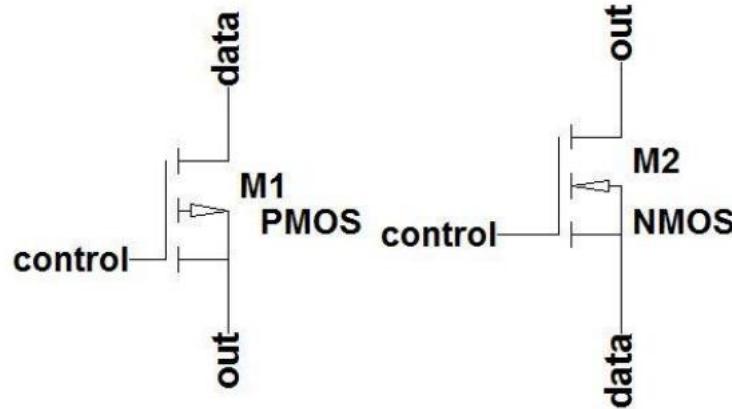


Truth Table

A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

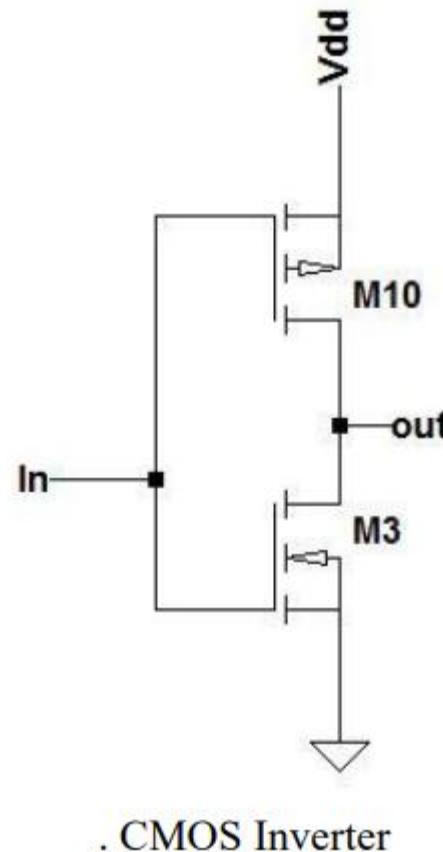


Structural Verilog code of MOSFETs



Syntax:

```
nmos n1(out, data, control) ;  
pmos p1(out, data, control) ;
```



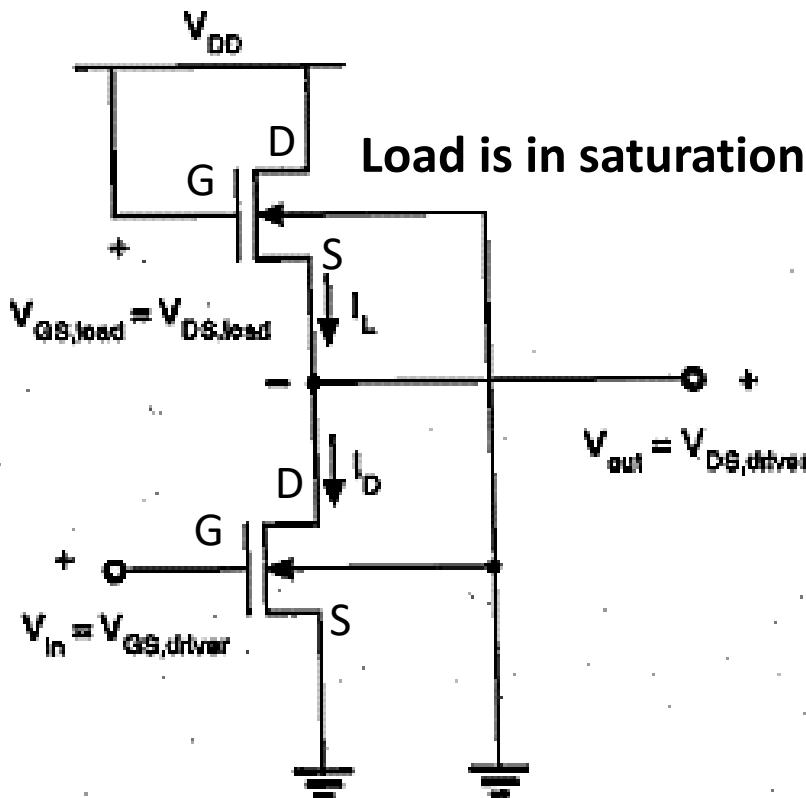
Verilog Code:

```
module cmos1(out,in);  
output out;  
input in;  
supply1 vdd;  
supply0 gnd;  
wire out;  
  
pmos (out,vdd,in);  
nmos (out,gnd,in);  
endmodule
```

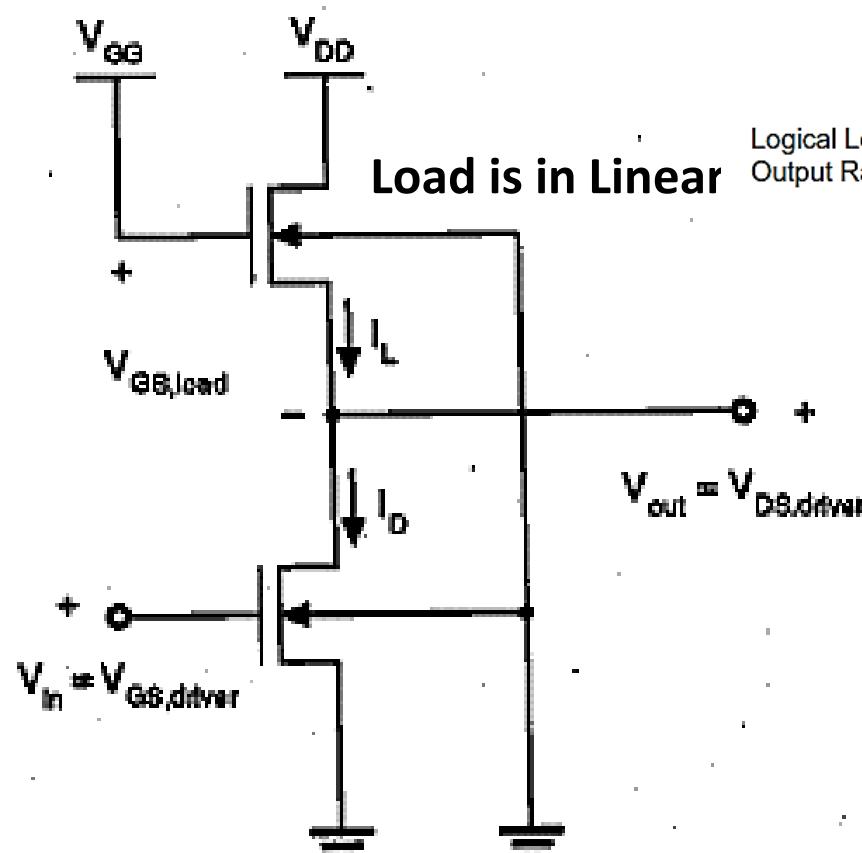


3. n-type MOSFET load

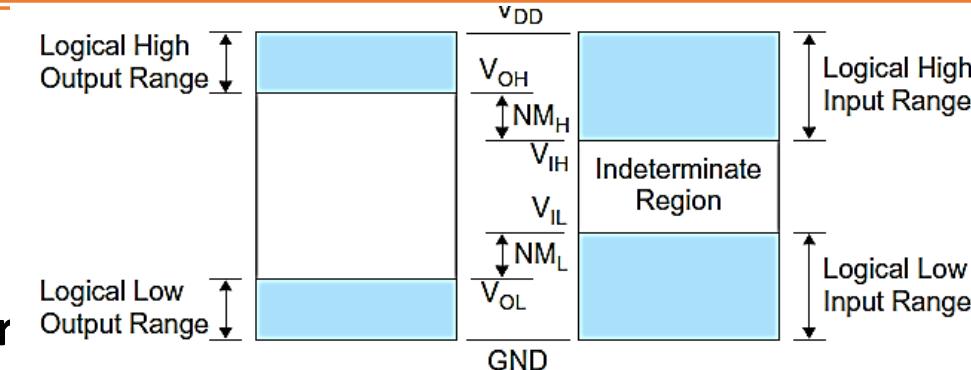
i. Enhancement load nMOS inverter



$$V_{OH} = V_{DD} - V_{T,load}$$



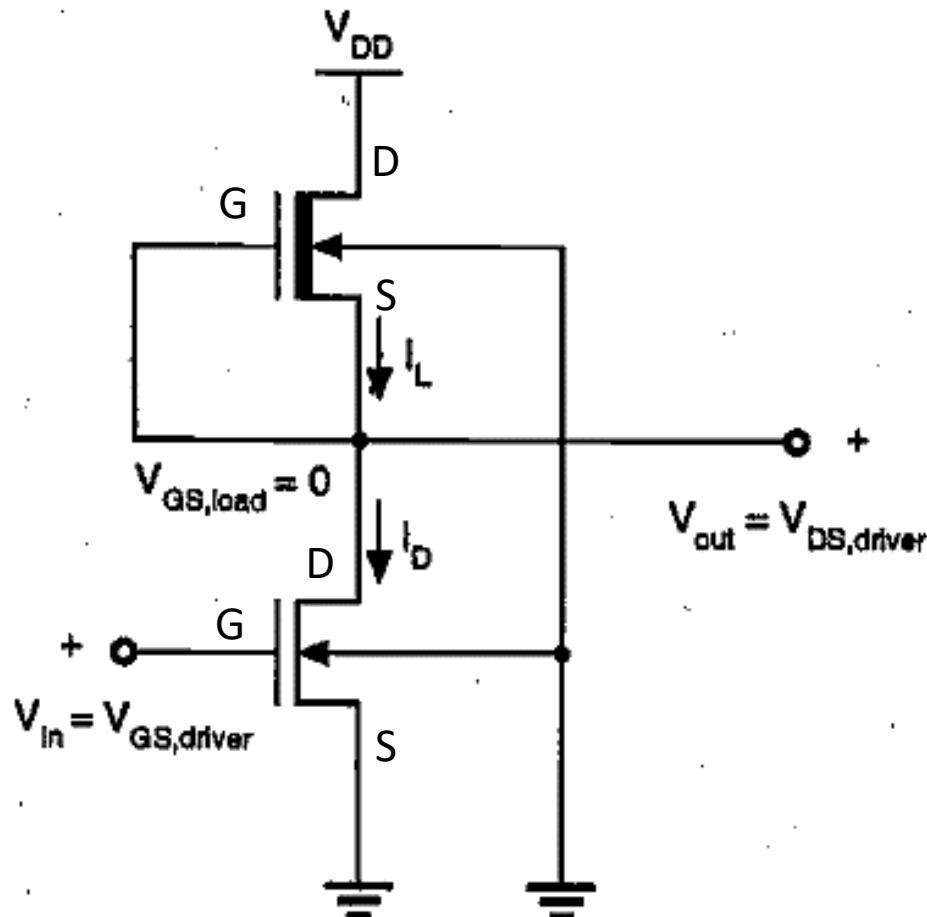
$$V_{OH} = V_{DD}$$





3. n-type MOSFET load

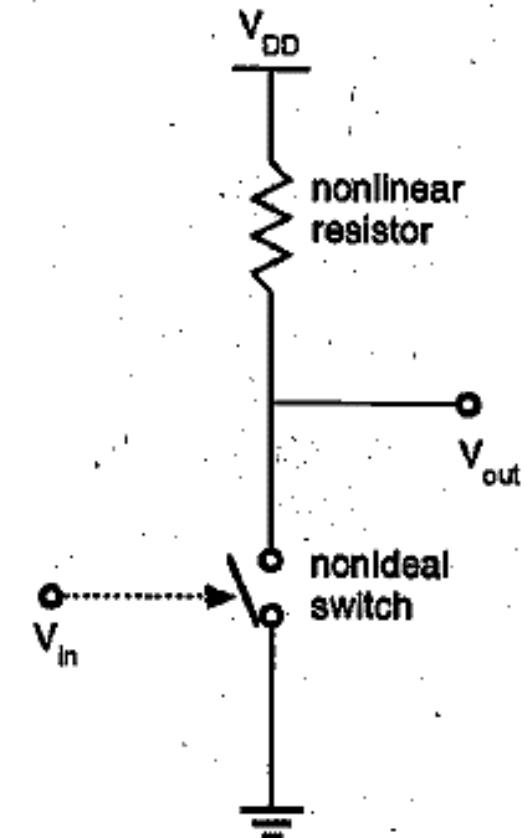
ii. Depletion load nMOS inverter



$$V_{GS(\text{Load})} = 0$$

$$V_{T0,\text{load}} < 0$$

$$V_{GS,load} > V_{T,load}$$

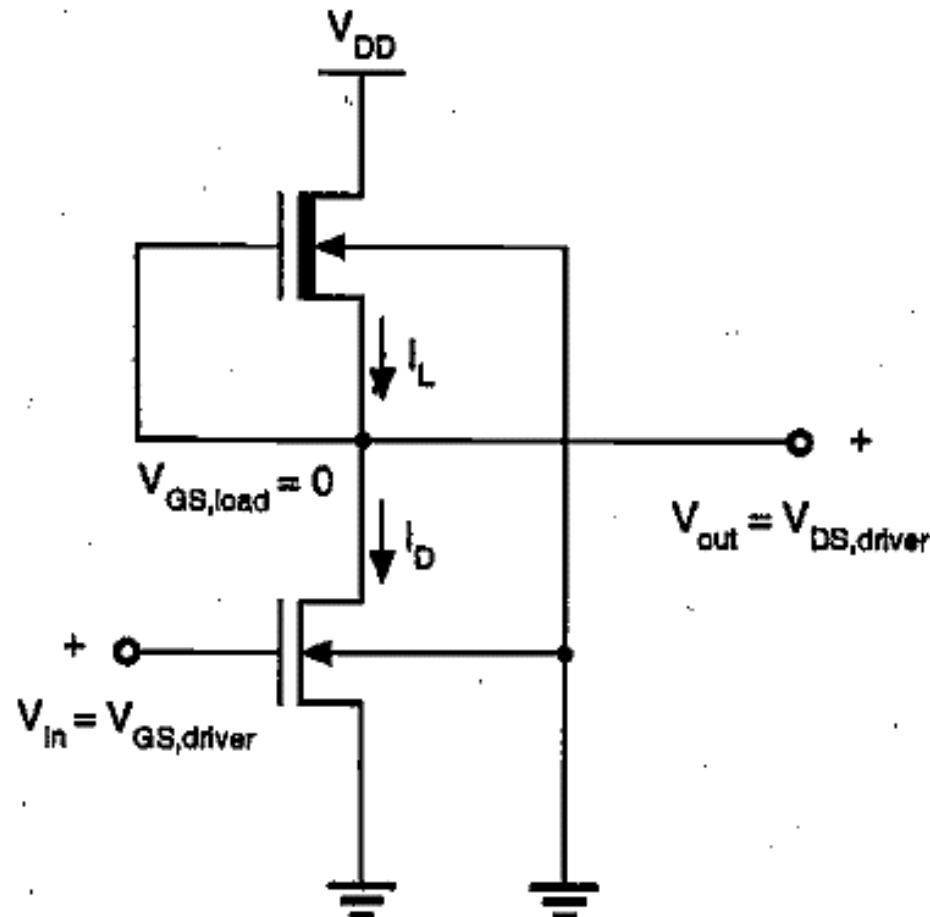


At this condition, the load will always be in ON state.



3. n-type MOSFET load

ii. Depletion load nMOS inverter



$$V_{OH} = V_{DD}$$

$$V_{OL} = V_{OH} - V_{T0} - \sqrt{(V_{OH} - V_{T0})^2 - \left(\frac{k_{load}}{k_{driver}}\right) \cdot |V_{T,load}(V_{OL})|^2}$$

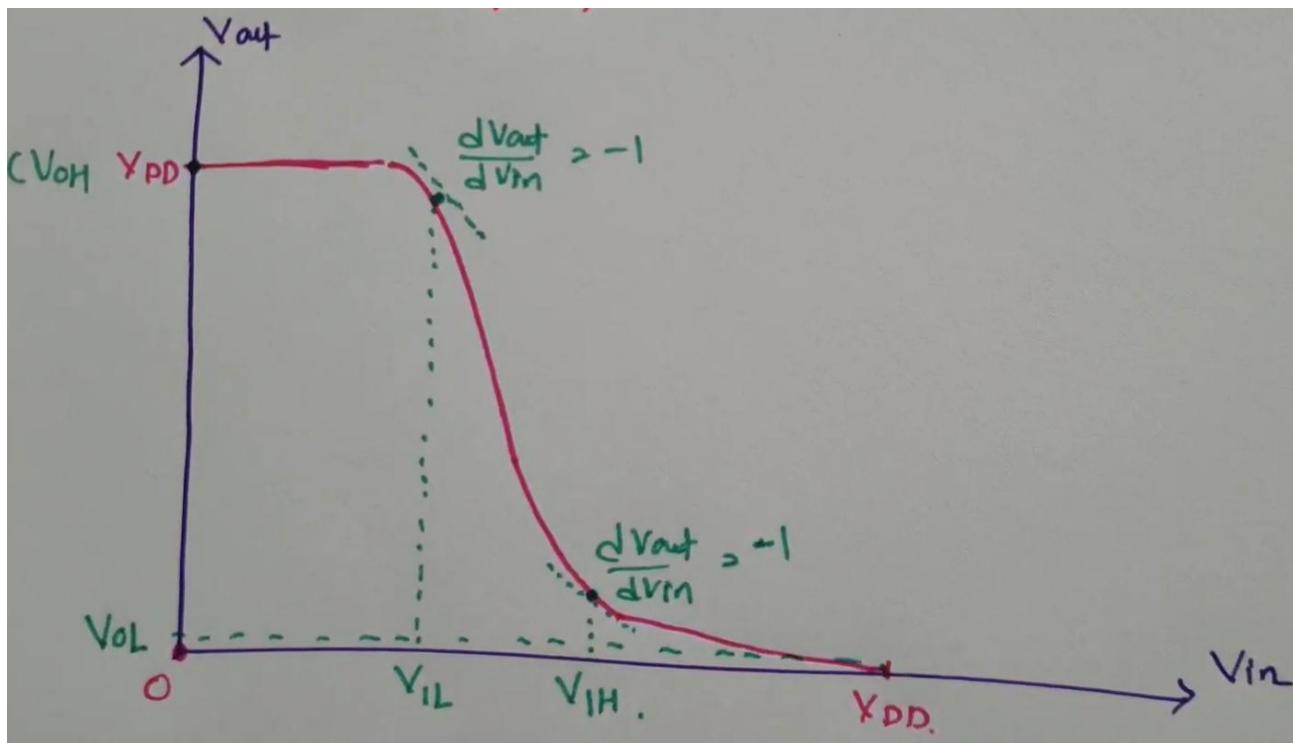
$$V_{IL} = V_{T0} + \left(\frac{k_{load}}{k_{driver}}\right) \cdot [V_{out} - V_{DD} + |V_{T,load}(V_{out})|]$$

$$V_{IH} = V_{T0} + 2V_{out} + \left(\frac{k_{load}}{k_{driver}}\right) \cdot [-V_{T,load}(V_{out})] \cdot \left(\frac{dV_{T,load}}{dV_{out}}\right)$$



3. n-type MOSFET load

ii. Depletion load nMOS inverter



V_{in}	V_{out}	Driver operating region	Load operating region
V_{OL}	V_{OH}	cut-off	linear
V_{IL}	$\approx V_{OH}$	saturation	linear
V_{OH}	V_{OL}	linear	saturation



NMOS inverter with derivation of Z_{p.u}/Z_{p.d} ratio

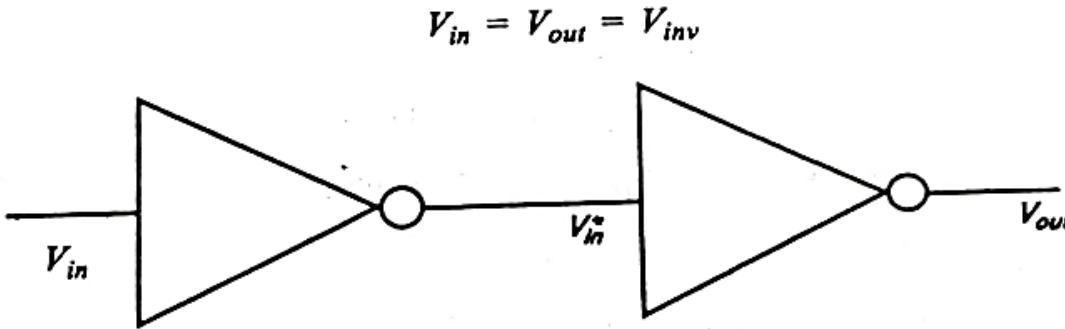
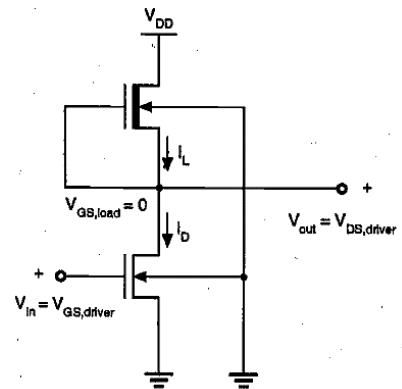


FIGURE 2.8 nMOS inverter driven directly by another inverter.

we set $V_{inv} = 0.5V_{DD}$

At this point both transistors are in saturation and



$$I_{ds} = K \frac{W}{L} \frac{(V_{gs} - V_t)^2}{2}$$

In the depletion mode

$$I_{ds} = K \frac{W_{p.u.}}{L_{p.u.}} \frac{(-V_{td})^2}{2} \text{ since } V_{gs} = 0$$

and in the enhancement mode

$$I_{ds} = K \frac{W_{p.d.}}{L_{p.d.}} \frac{(V_{inv} - V_t)^2}{2} \text{ since } V_{gs} = V_{inv}$$

Equating (since currents are the same) we have

$$\frac{W_{p.d.}}{L_{p.d.}} (V_{inv} - V_t)^2 = \frac{W_{p.u.}}{L_{p.u.}} (-V_{td})^2$$

Now write

$$Z_{p.d.} = \frac{L_{p.d.}}{W_{p.d.}}; Z_{p.u.} = \frac{L_{p.u.}}{W_{p.u.}}$$



we have

$$\frac{1}{Z_{p.d.}} (V_{inv} - V_t)^2 = \frac{1}{Z_{p.u.}} (-V_{td})^2$$

$$V_{inv} = V_t - \frac{V_{td}}{\sqrt{Z_{p.u.}/Z_{p.d.}}}$$

Now we can substitute typical values as follows:

$$V_t = 0.2V_{DD}; V_{td} = -0.6V_{DD}$$
$$V_{inv} = 0.5V_{DD} \text{ (for equal margins)}$$

thus, from equation (2.9)

whence

$$0.5 = 0.2 + \frac{0.6}{\sqrt{Z_{p.u.}/Z_{p.d.}}}$$

and thus

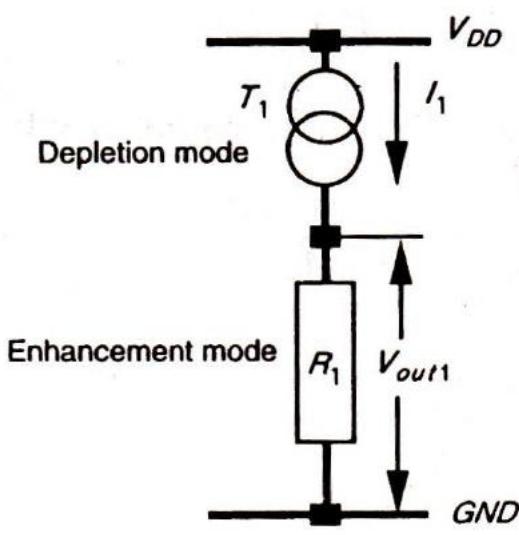
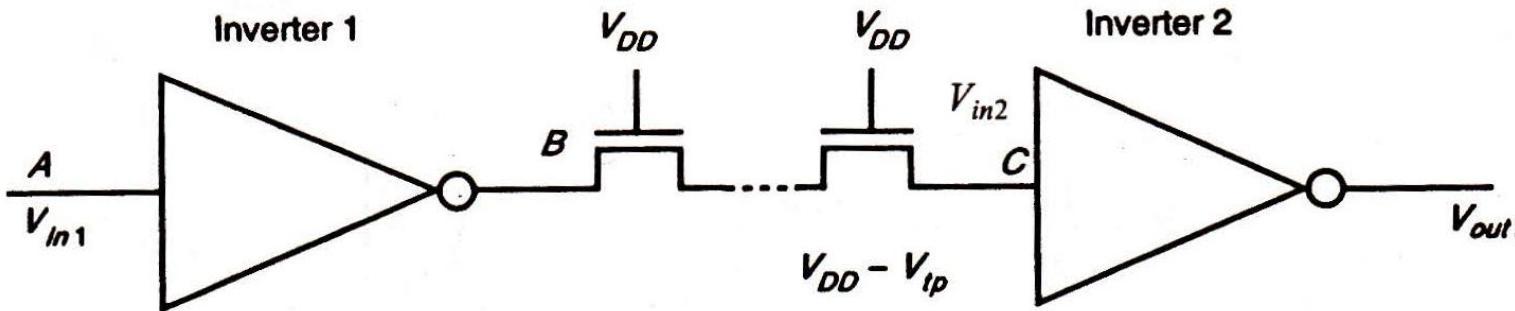
$$\sqrt{Z_{p.u.}/Z_{p.d.}} = 2$$

for an inverter directly driven by an inverter.

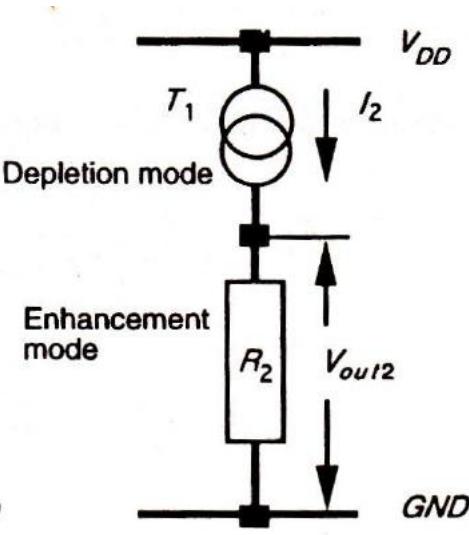
$$Z_{p.u.}/Z_{p.d.} = 4/1$$



NMOS inverter with derivation of Z_{pu}/Z_{pd} ratio (degraded input)



(a) Inverter 1 with input = V_{DD}



(b) Inverter 2 with input = $V_{DD} - V_{ip}$

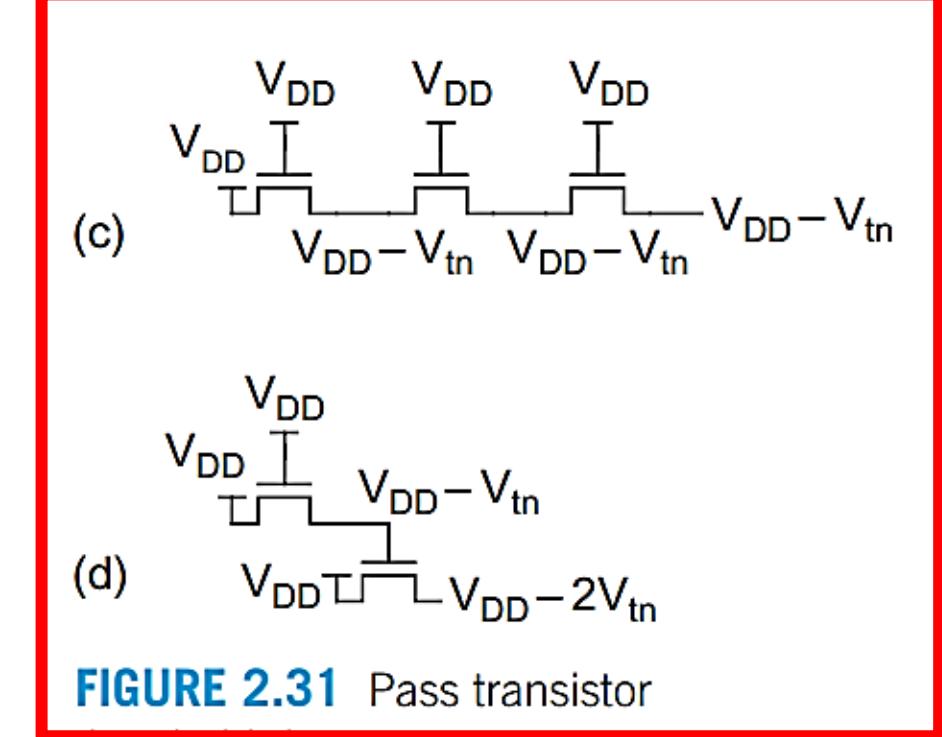


FIGURE 2.31 Pass transistor



NMOS inverter with derivation of Z_{pu}/Z_{pd} ratio (degraded input)

$$R_1 = \frac{1}{K} Z_{p.d.1} \left(\frac{1}{V_{DD} - V_t} \right)$$

$$I_1 = I_{ds} = K \frac{W_{p.u.1}}{L_{p.u.1}} \frac{(-V_{td})^2}{2}$$

$$V_{out1} = I_1 R_1 = \frac{Z_{p.d.1}}{Z_{p.u.1}} \left(\frac{1}{V_{DD} - V_t} \right) \frac{(-V_{td})^2}{2}$$

Inverter 1

$$R_2 = \frac{1}{K} Z_{p.d.2} \frac{1}{((V_{DD} - V_{tp}) - V_t)}$$

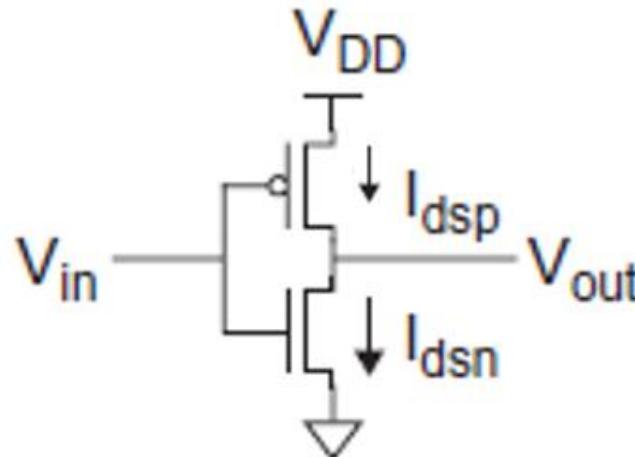
$$I_2 = K \frac{1}{Z_{p.u.2}} \frac{(-V_{td})^2}{2}$$

$$V_{out2} = I_2 R_2 = \frac{Z_{p.d.2}}{Z_{p.u.2}} \left(\frac{1}{V_{DD} - V_{tp} - V_t} \right) \frac{(-V_{td})^2}{2}$$

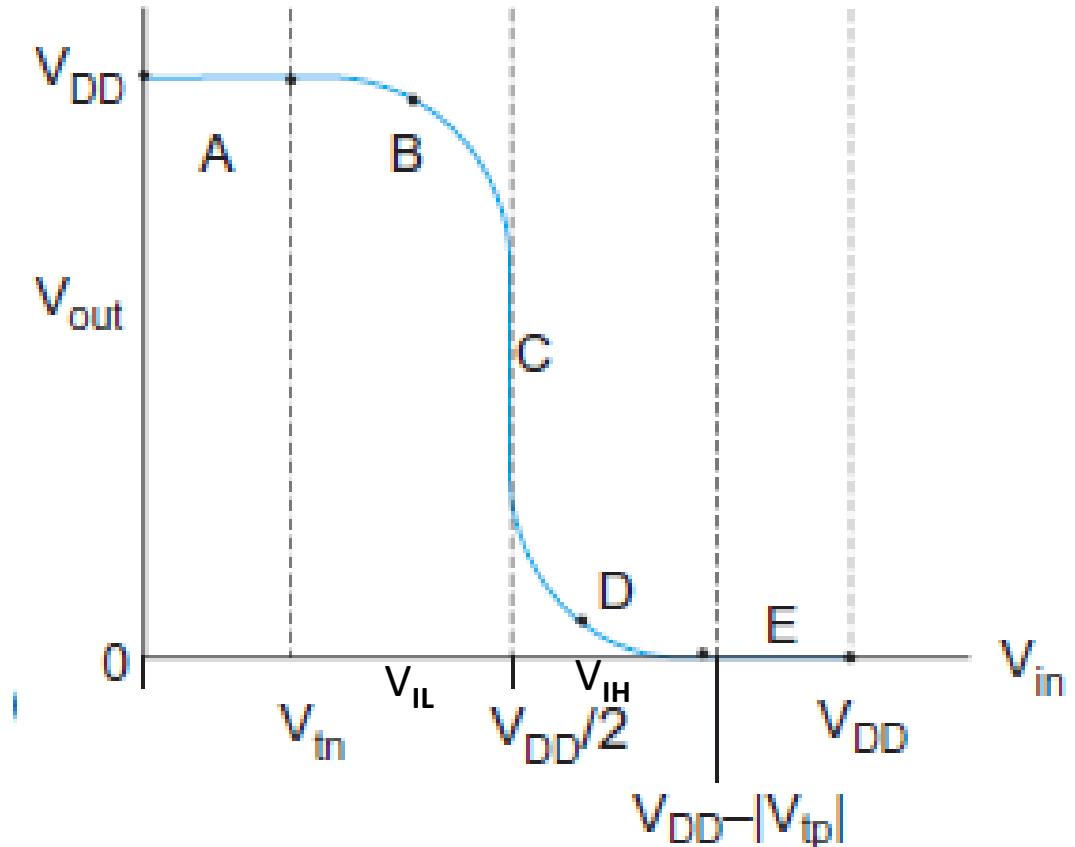
Inverter 2



4. CMOS inverter



Region	nMOS	pMOS
A	cut-off	linear
B	saturation	linear
C	saturation	saturation
D	linear	saturation
E	linear	cut-off





4. CMOS inverter

$$V_{GS,n} = V_{in}$$

$$V_{DS,n} = V_{OUT}$$

$$V_{GS,P} = V_{in} - V_{DD}$$

$$V_{DS,P} = V_{out} - V_{DD}$$

The nMOS operates in the saturation region if $V_{in} > V_{TO}$ and if following conditions are satisfied.

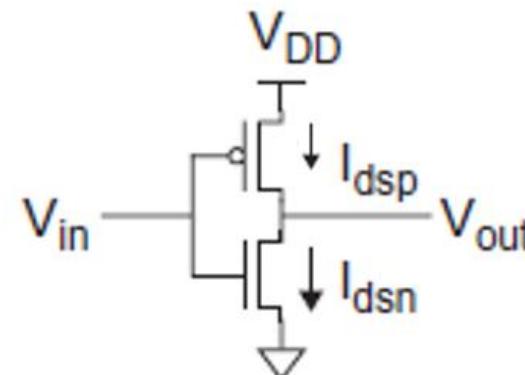
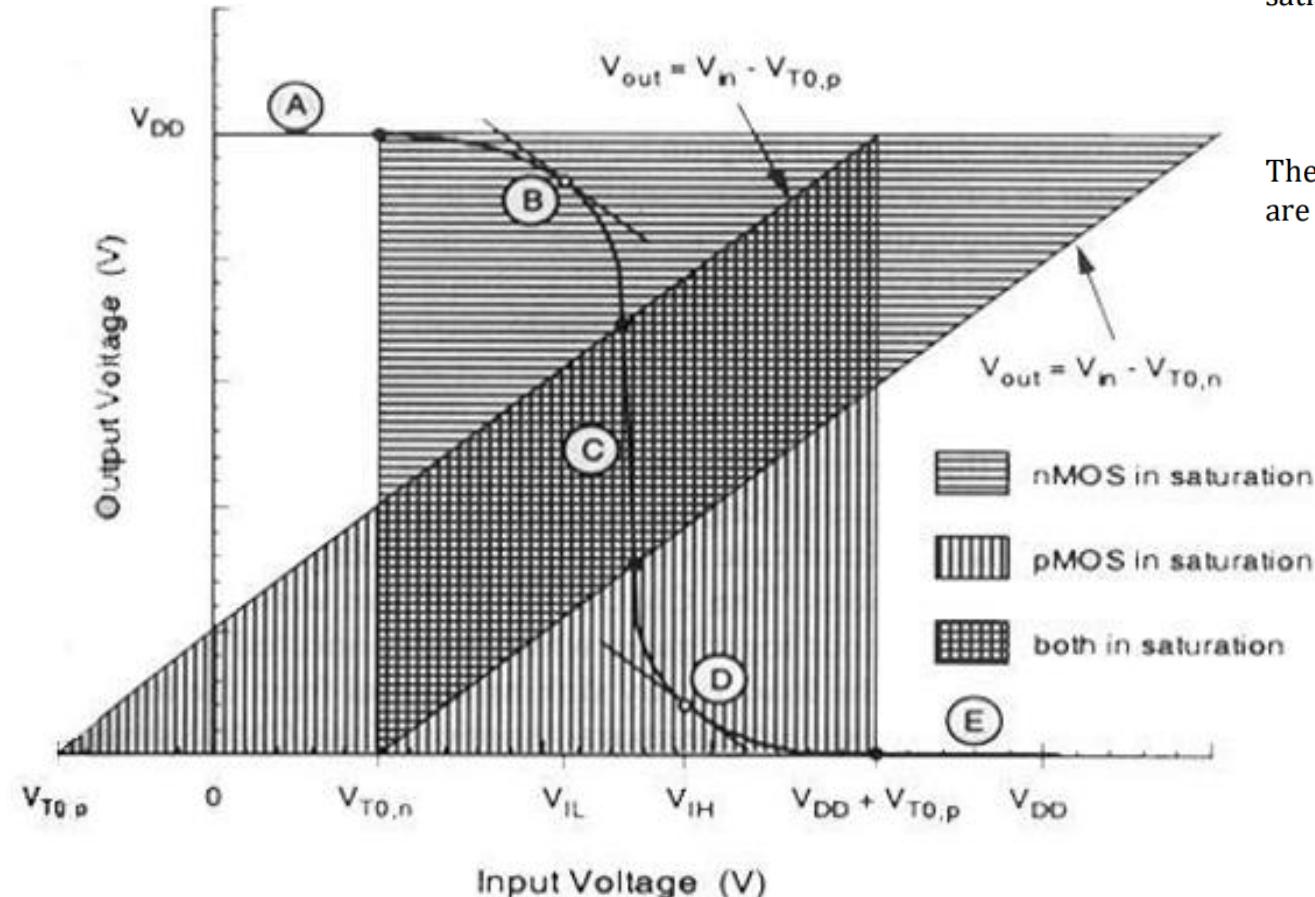
$$V_{DS,n} \geq V_{GS,n} - V_{TO,n}$$

$$V_{out} \geq V_{in} - V_{TO,n}$$

The pMOS operates in the saturation region if $V_{in} < V_{DD} + V_{TO,p}$ and if following conditions are satisfied.

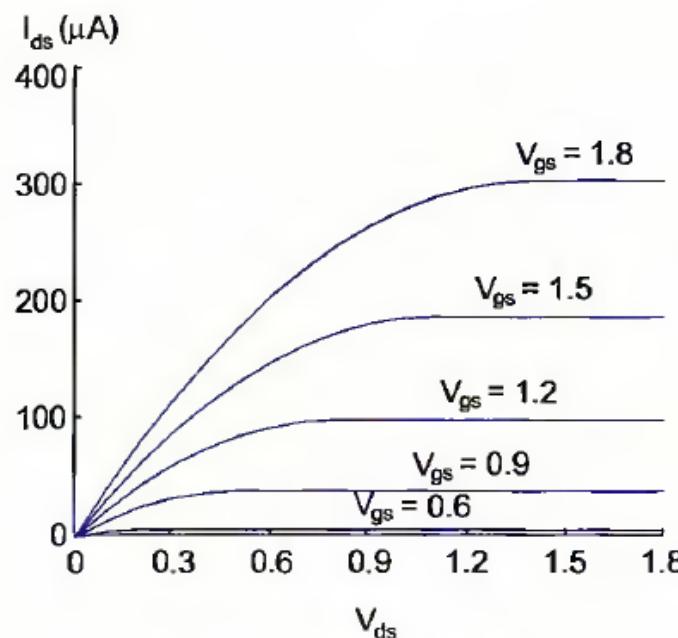
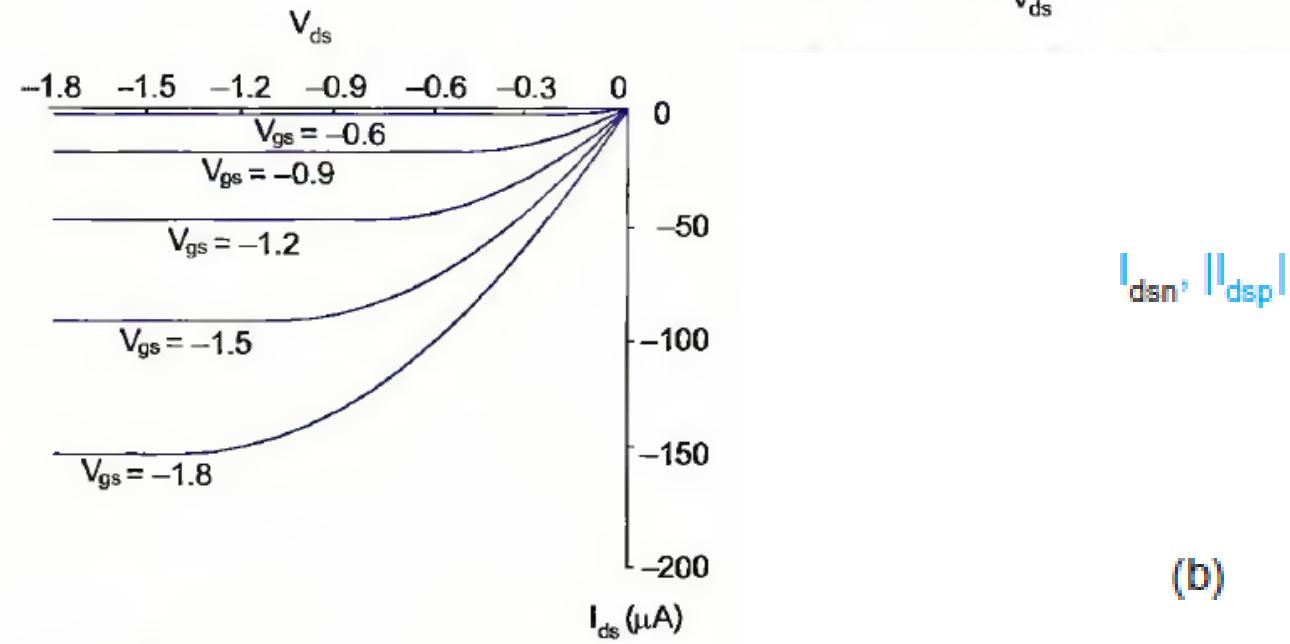
$$V_{DS,P} \leq V_{GS,P} - V_{TO,P}$$

$$V_{out} \leq V_{in} - V_{TO,P}$$



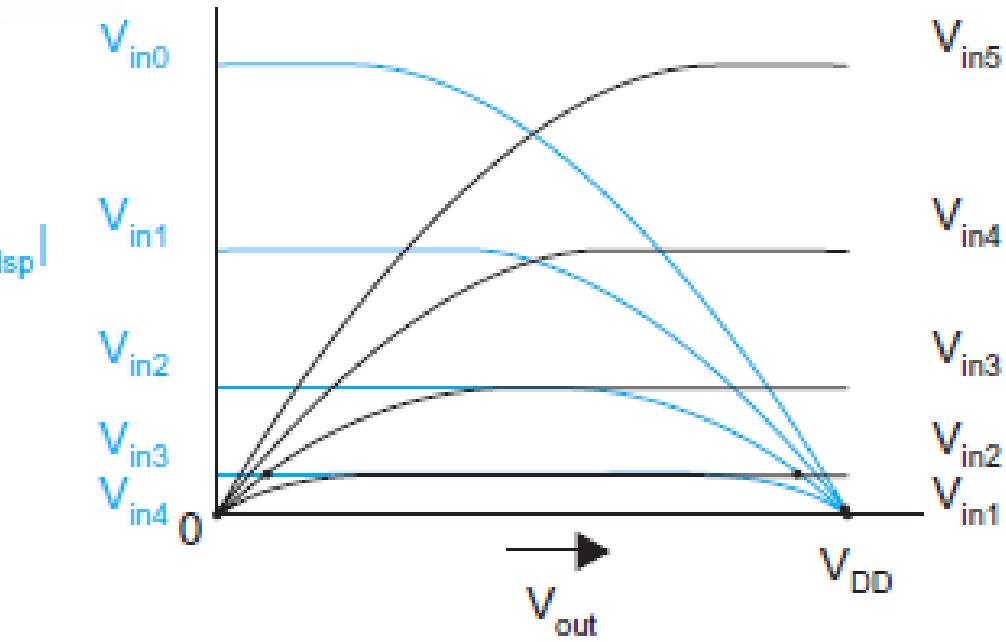


4. CMOS inverter



(b)

$|I_{dsn}|, |I_{dsp}|$





4. CMOS inverter

Governing Conditions

	Cutoff	Linear	Saturated
nMOS	$V_{gsn} < V_{tn}$	$V_{gsn} > V_{tn}$	$V_{gsn} > V_{tn}$
	$V_{in} < V_{tn}$	$V_{in} > V_{tn}$	$V_{in} > V_{tn}$
		$V_{dsn} < V_{gsn} - V_{tn}$	$V_{dsn} > V_{gsn} - V_{tn}$
		$V_{out} < V_{in} - V_{tn}$	$V_{out} > V_{in} - V_{tn}$
pMOS	$V_{gsp} > V_{tp}$	$V_{gsp} < V_{tp}$	$V_{gsp} < V_{tp}$
	$V_{in} > V_{tp} + V_{DD}$	$V_{in} < V_{tp} + V_{DD}$	$V_{in} < V_{tp} + V_{DD}$
		$V_{dsp} > V_{gsp} - V_{tp}$	$V_{dsp} < V_{gsp} - V_{tp}$
		$V_{out} > V_{in} - V_{tp}$	$V_{out} < V_{in} - V_{tp}$



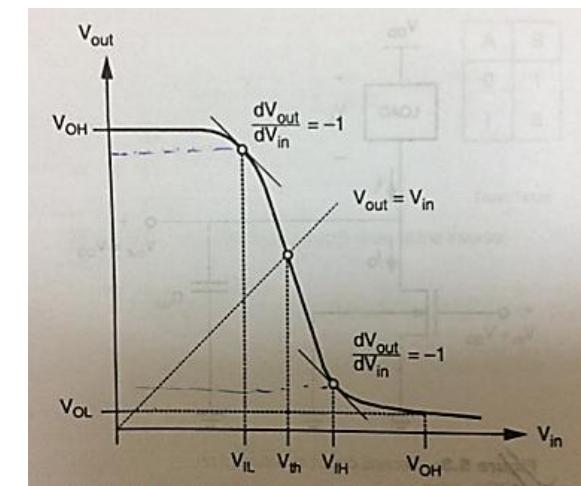
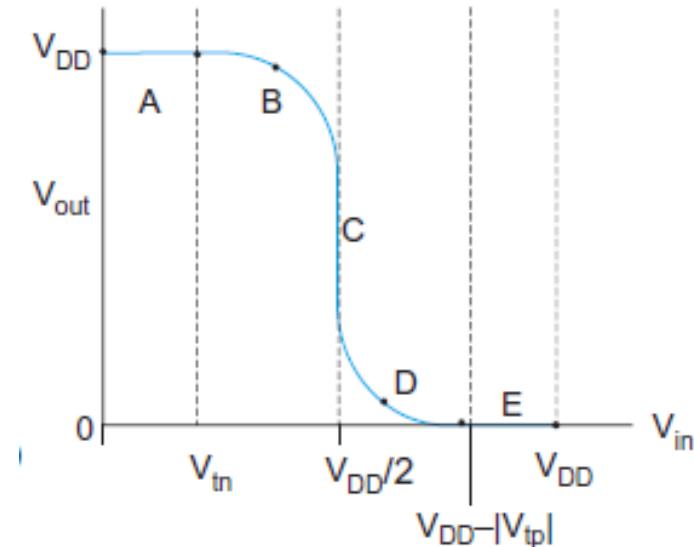
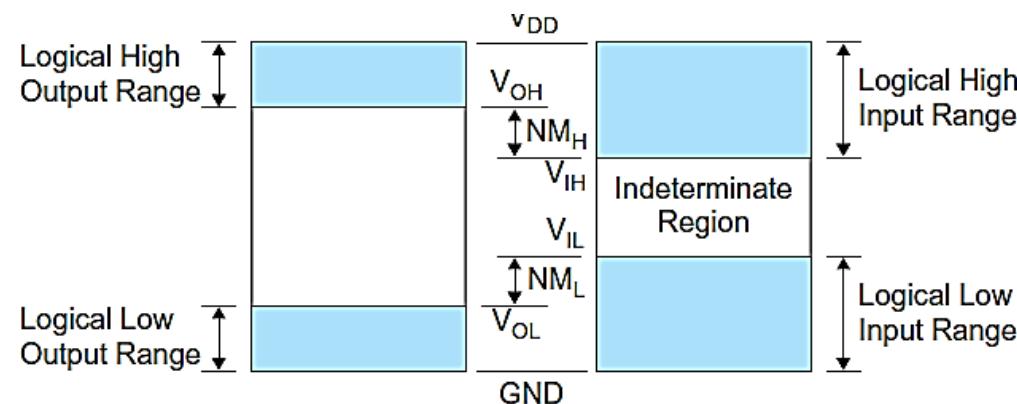
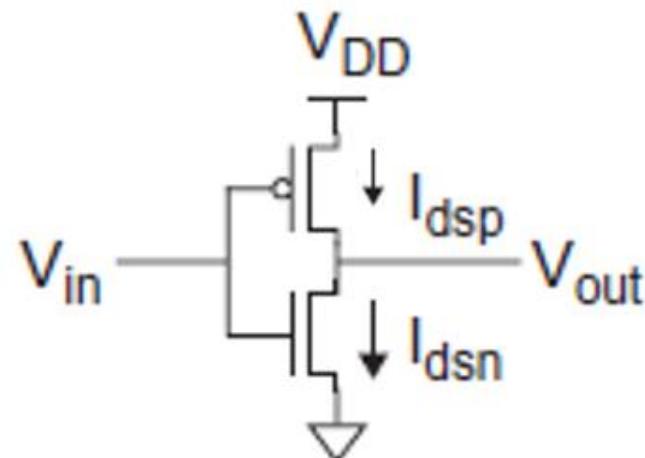
Region A \rightarrow V_{OH}

PMOS – Linear Region
NMOS - Cutoff

$$V_{out} = V_{DD} - I_{DSP} R_{PMOS}$$

$$V_{out} = V_{OH} = V_{DD}$$

Region	nMOS	pMOS
A	cut-off	linear
B	saturation	linear
C	saturation	saturation
D	linear	saturation
E	linear	cut-off





Region B \rightarrow V_{IL}

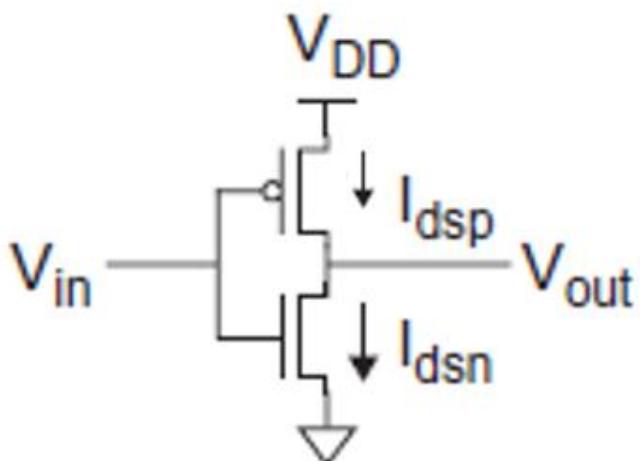
$$\frac{k_n}{2} \cdot (V_{GS,n} - V_{T0,n})^2 = \frac{k_p}{2} \cdot [2 \cdot (V_{GS,p} - V_{T0,p}) \cdot V_{DS,p} - V_{DS,p}^2]$$

$$\frac{k_n}{2} \cdot (V_{in} - V_{T0,n})^2 = \frac{k_p}{2} \cdot [2 \cdot (V_{in} - V_{DD} - V_{T0,p})$$

$$\cdot (V_{out} - V_{DD}) - (V_{out} - V_{DD})^2]$$

The slope of the VTC is equal to (-1), when the input voltage is $V_{in} = V_{IL}$.

Region	nMOS	pMOS
A	cut-off	linear
B	saturation	linear
C	saturation	saturation
D	linear	saturation
E	linear	cut-off



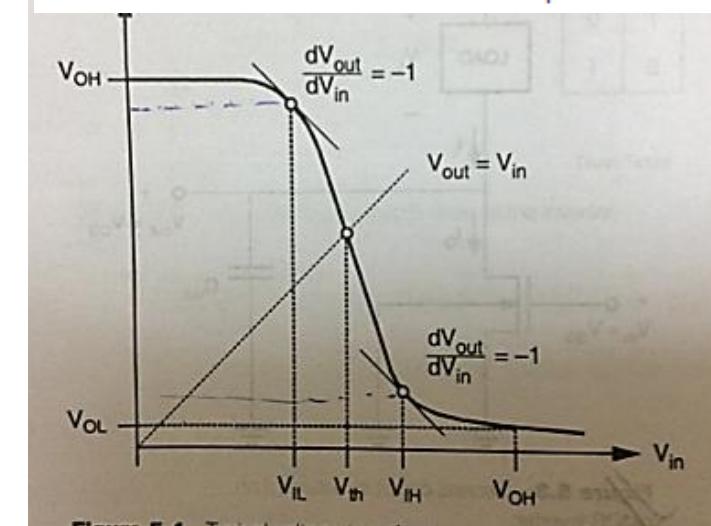
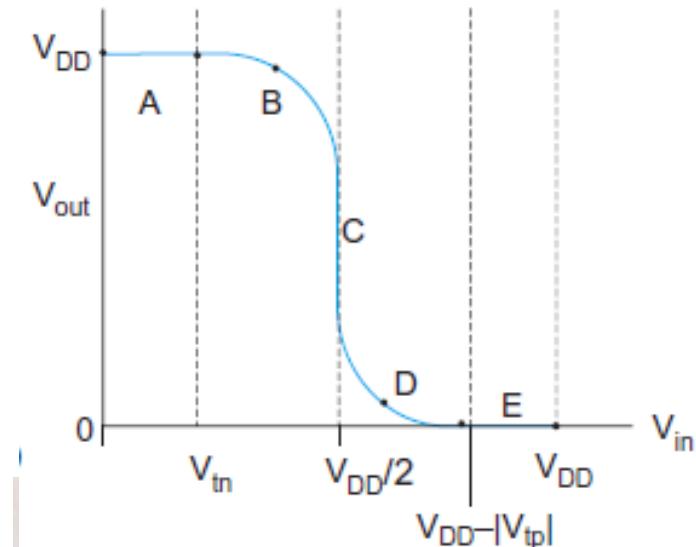
$$V_{GS,n} = V_{in}$$

$$V_{DS,n} = V_{OUT}$$

$$V_{GS,p} = V_{in} - V_{DD}$$

$$V_{DS,p} = V_{out} - V_{DD}$$

PMOS – Linear Region
NMOS – Saturation Region





Region B $\rightarrow V_{IL}$

PMOS – Linear Region
NMOS – Saturation Region

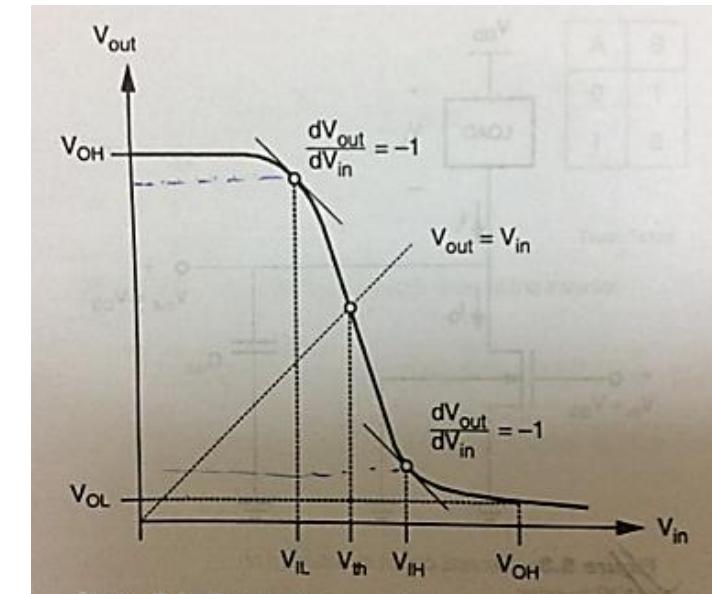
To satisfy the derivative condition at V_{IL} we differentiate both sides with respect to V_{in} .

$$k_n \cdot (V_{in} - V_{T0,n}) = k_p \cdot \left[(V_{in} - V_{DD} - V_{T0,p}) \cdot \left(\frac{dV_{out}}{dV_{in}} \right) + (V_{out} - V_{DD}) \cdot \left(\frac{dV_{out}}{dV_{in}} \right) \right]$$

Substituting $V_{in} = V_{IL}$ and $(dV_{out}/dV_{in}) = -1$ in (5.60), we obtain

$$k_n \cdot (V_{IL} - V_{T0,n}) = k_p \cdot (2V_{out} - V_{IL} + V_{T0,p} - V_{DD})$$

$$V_{IL} = \frac{2V_{out} + V_{T0,p} - V_{DD} + k_R V_{T0,n}}{1 + k_R}$$



Where

$$k_R = \frac{k_n}{k_p}$$

Region C $\rightarrow V_{th}$

PMOS – Saturation Region
NMOS – Saturation Region

$$\frac{k_n}{2} \cdot (V_{GS,n} - V_{T0,n})^2 = \frac{k_p}{2} \cdot (V_{GS,p} - V_{T0,p})^2$$

Replacing $V_{GS,n}$ and $V_{GS,p}$ in (5.68) according to (5.51) and (5.52), we obtain

$$\frac{k_n}{2} \cdot (V_{in} - V_{T0,n})^2 = \frac{k_p}{2} \cdot (V_{in} - V_{DD} - V_{T0,p})^2$$

The correct solution for V_{in} for this equation is

$$V_{in} \cdot \left(1 + \sqrt{\frac{k_p}{k_n}}\right) = V_{T0,n} + \sqrt{\frac{k_p}{k_n} \cdot (V_{DD} + V_{T0,p})}$$

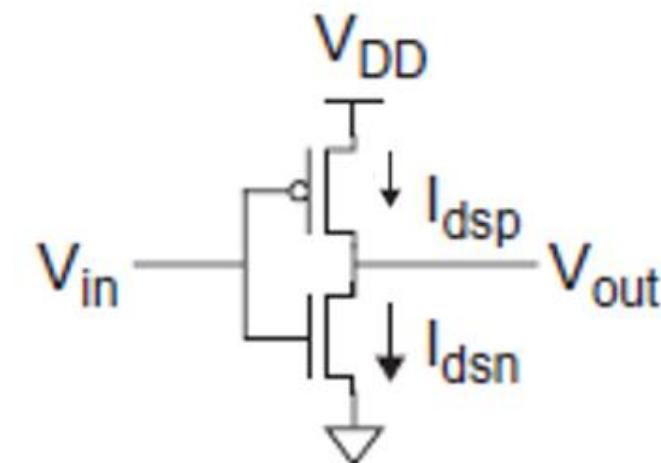
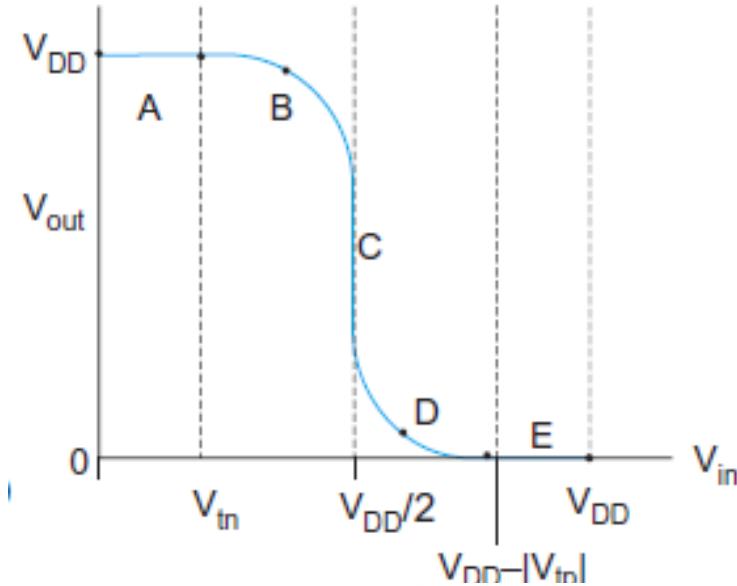
$$V_{GS,n} = V_{in}$$

$$V_{DS,n} = V_{out}$$

$$V_{GS,p} = V_{in} - V_{DD}$$

$$V_{DS,p} = V_{out} - V_{DD}$$

Region	nMOS	pMOS
A	cut-off	linear
B	saturation	linear
C	saturation	saturation
D	linear	saturation
E	linear	cut-off





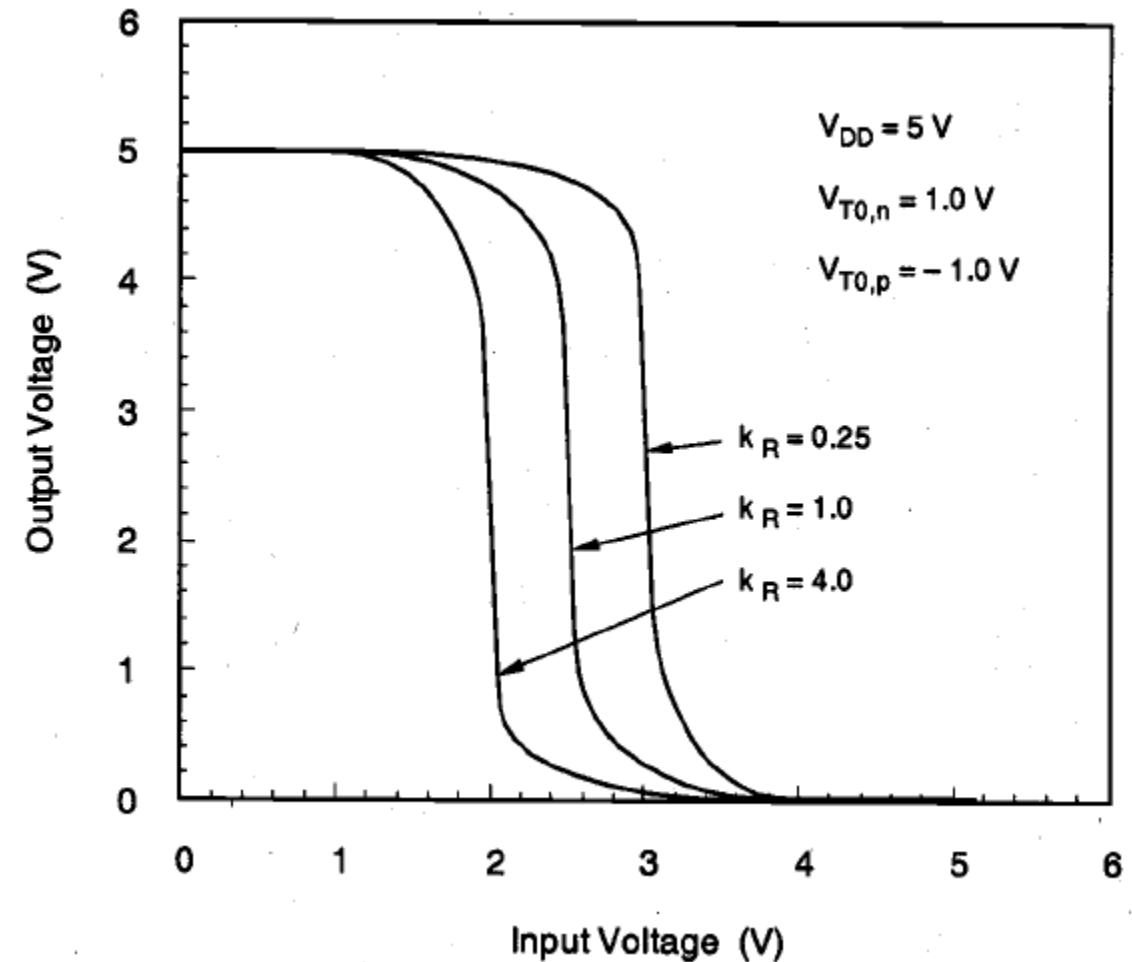
Region C $\rightarrow V_{th}$

$$V_{th} = \frac{V_{T0,n} + \sqrt{\frac{1}{k_R} \cdot (V_{DD} + V_{T0,p})}}{1 + \sqrt{\frac{1}{k_R}}}$$

Where

$$k_R = \frac{k_n}{k_p}$$

PMOS – Saturation Region
NMOS – Saturation Region



Region D $\rightarrow V_{IH}$

PMOS – Saturation Region
NMOS – Linear Region

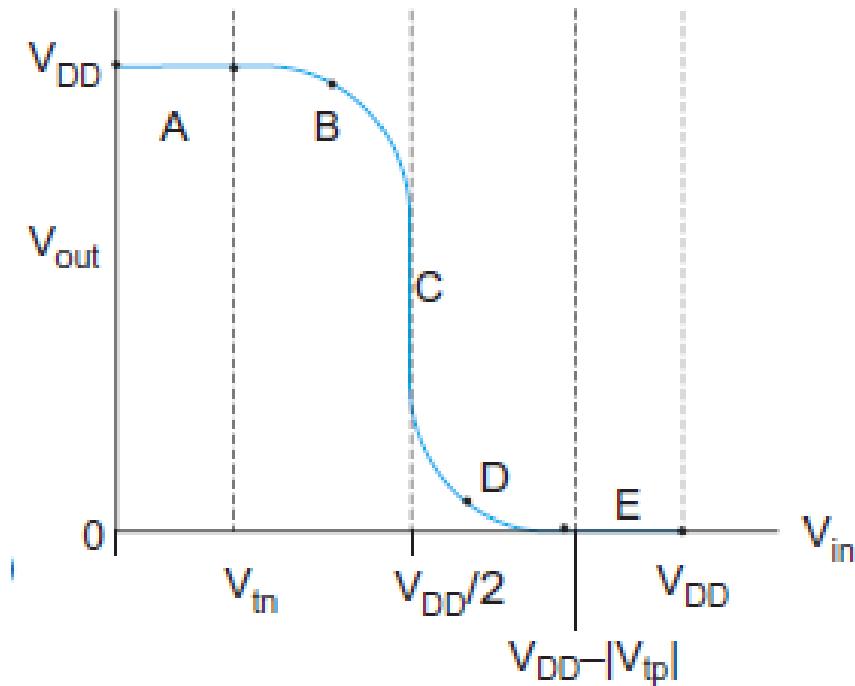
$$\frac{k_n}{2} \cdot [2 \cdot (V_{GS,n} - V_{T0,n}) \cdot V_{DS,n} - V_{DS,n}^2] = \frac{k_p}{2} \cdot (V_{GS,p} - V_{T0,p})^2$$

$$\frac{k_n}{2} \cdot [2 \cdot (V_{in} - V_{T0,n}) \cdot V_{out} - V_{out}^2] = \frac{k_p}{2} \cdot (V_{in} - V_{DD} - V_{T0,p})^2$$

Now, differentiate both sides of (5.64) with respect to V_{in} .

$$k_n \cdot \left[(V_{in} - V_{T0,n}) \cdot \left(\frac{dV_{out}}{dV_{in}} \right) + V_{out} - V_{out} \cdot \left(\frac{dV_{out}}{dV_{in}} \right) \right] \\ = k_p \cdot (V_{in} - V_{DD} - V_{T0,p})$$

Region	nMOS	pMOS
A	cut-off	linear
B	saturation	linear
C	saturation	saturation
D	linear	saturation
E	linear	cut-off



$$V_{GS,n} = V_{in}$$

$$V_{DS,n} = V_{OUT}$$

$$V_{GS,P} = V_{in} - V_{DD}$$

$$V_{DS,P} = V_{out} - V_{DD}$$



Region D $\rightarrow V_{IH}$

PMOS – Saturation Region
NMOS – Linear Region

Substituting $V_{in} = V_{IH}$ and $(dV_{out}/dV_{in}) = -1$

$$k_n \cdot (-V_{IH} + V_{T0,n} + 2V_{out}) = k_p \cdot (V_{IH} - V_{DD} - V_{T0,p})$$

$$V_{IH} = \frac{V_{DD} + V_{T0,p} + k_R \cdot (2V_{out} + V_{T0,n})}{1 + k_R}$$

Where

$$k_R = \frac{k_n}{k_p}$$



Region E $\rightarrow V_{OL}$

PMOS – Cut-off Region
NMOS – Linear Region

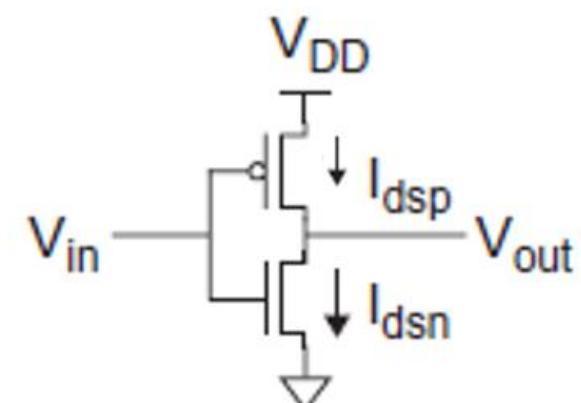
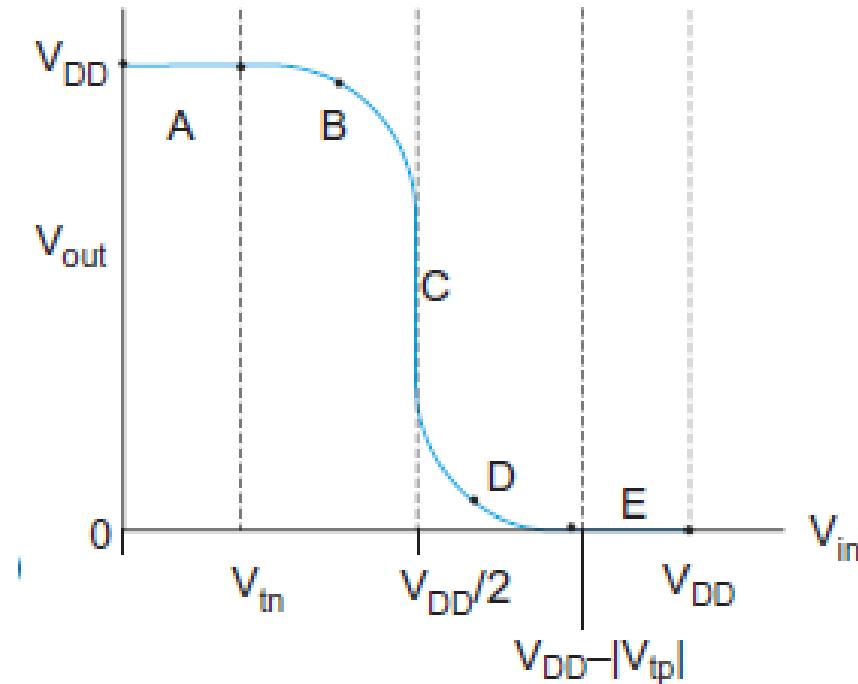
when the input voltage exceeds VDD the pMOS transistor is turned off. In this case, the nMOS transistor is operating in the linear region, but its drain to- source voltage is equal to zero because

$$I_{D,n} = I_{D,p} = 0$$

The output voltage of the circuit is

$$V_{out} = V_{OL} = 0$$

Region	nMOS	pMOS
A	cut-off	linear
B	saturation	linear
C	saturation	saturation
D	linear	saturation
E	linear	cut-off





CMOS Logic

1. The Inverter
2. The NAND Gate
3. CMOS Logic Gates
4. The NOR Gate
5. Compound Gates
6. Pass Transistors and Transmission Gates
7. Tristates
8. Multiplexers
9. Sequential Circuits



1. NOT gate (Inverter)

TABLE 1.1 Inverter truth table

A	Y
0	1
1	0

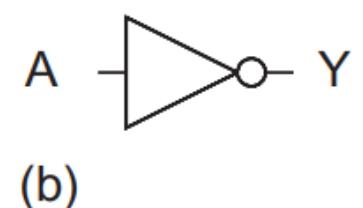
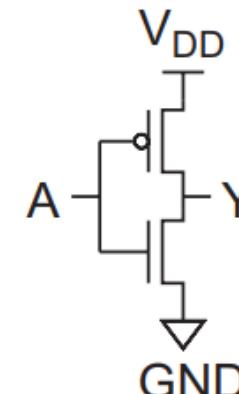


FIGURE 1.11
Inverter schematic
(a) and symbol
(b) $Y = \bar{A}$



2. NAND gate

	Gate=0	Gate=1
nMOS	OFF	ON
PMOS	ON	OFF

Operation	AND	OR
nMOS	Series	Parallel
PMOS	Parallel	Series

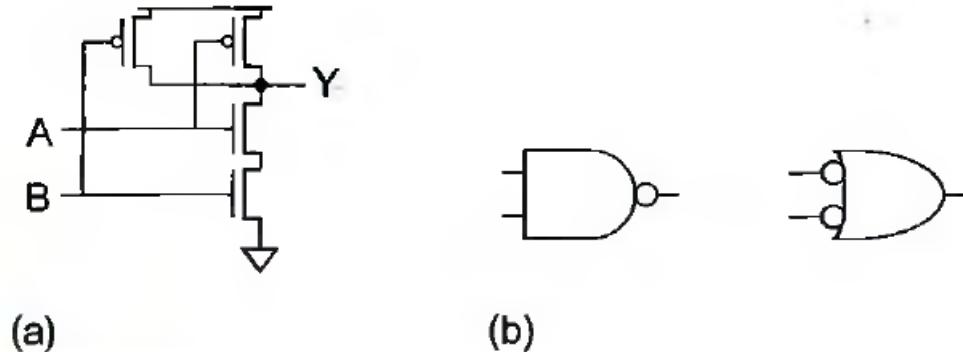


FIG 1.11 2-input NAND gate schematic (a) and symbol (b) $Y = \overline{A \bullet B}$

Table 1.2 NAND gate truth table

A	B	pull-down network	pull-up network	Y
0	0	OFF	ON	1
0	1	OFF	ON	1
1	0	OFF	ON	1
1	1	ON	OFF	0



3. CMOS Logic gates

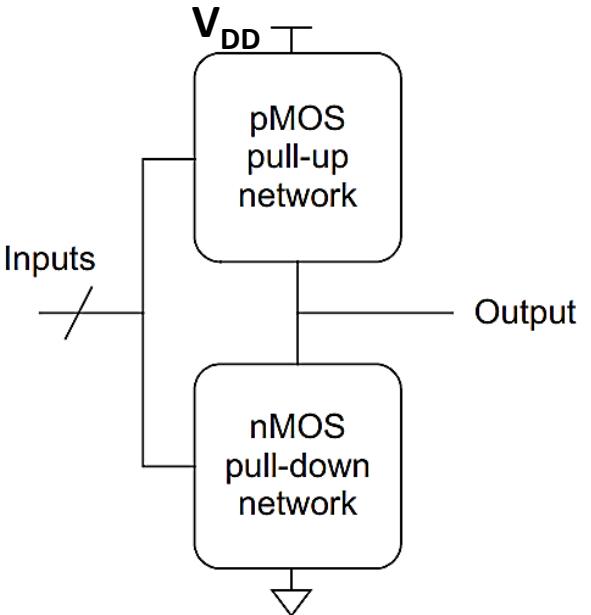


FIGURE 1.14 General logic gate using pull-up and pull-down networks

TABLE 1.3 Output states of CMOS logic gates

	pull-up OFF	pull-up ON
pull-down OFF	Z	1
pull-down ON	0	crowbarred (X)

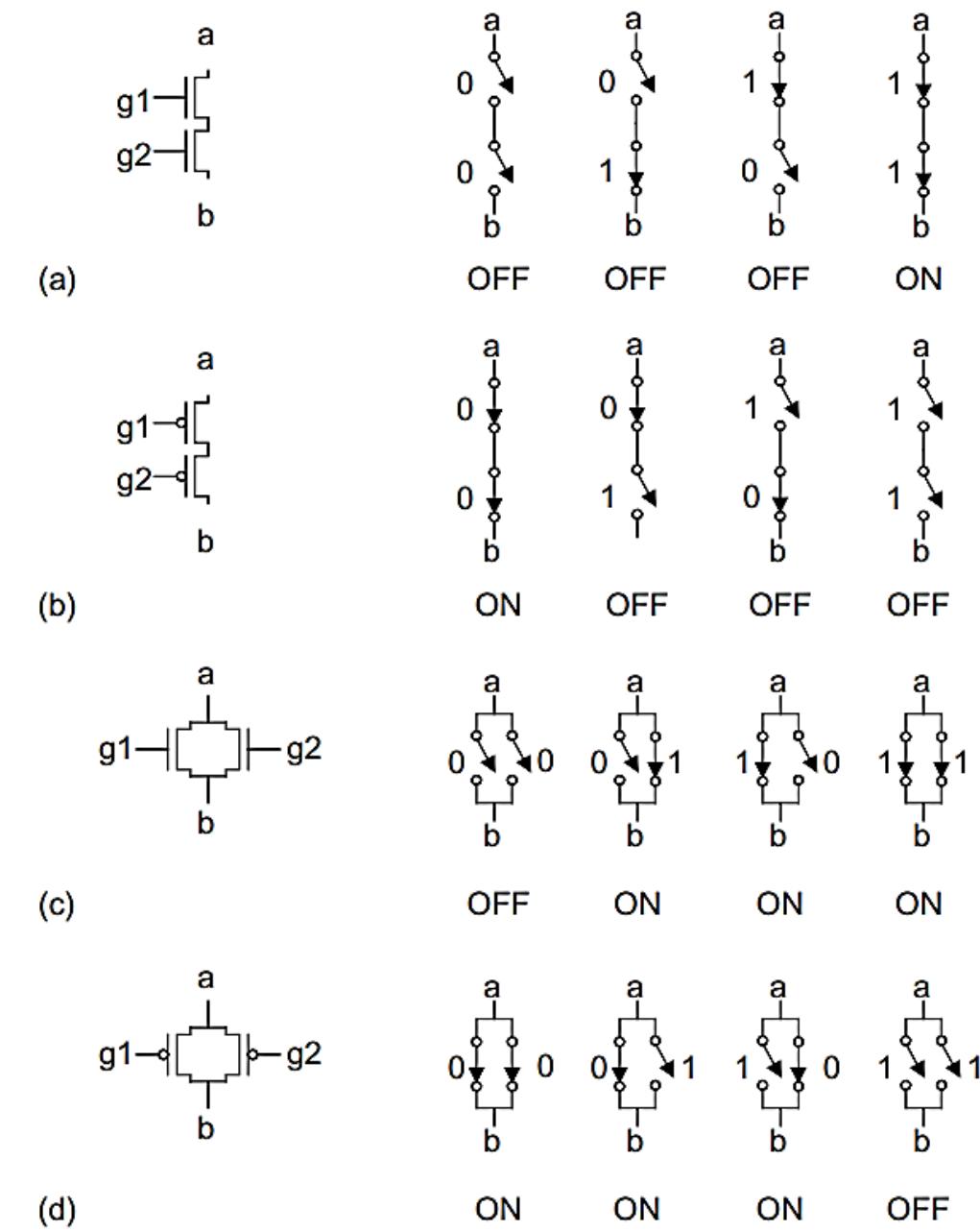


FIGURE 1.15 Connection and behavior of series and parallel transistors



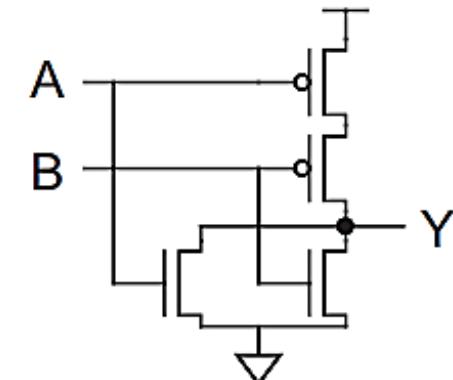
4. NOR gate

Operation	AND	OR
nMOS	Series	Parallel
PMOS	Parallel	Series

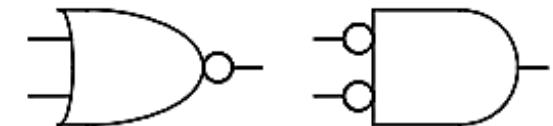
	Gate=0	Gate=1
nMOS	OFF	ON
PMOS	ON	OFF

TABLE 1.4 NOR gate truth table

A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0



(a)



(b)

FIGURE 1.16 2-input NOR gate schematic (a) and symbol (b) $Y = \overline{A + B}$

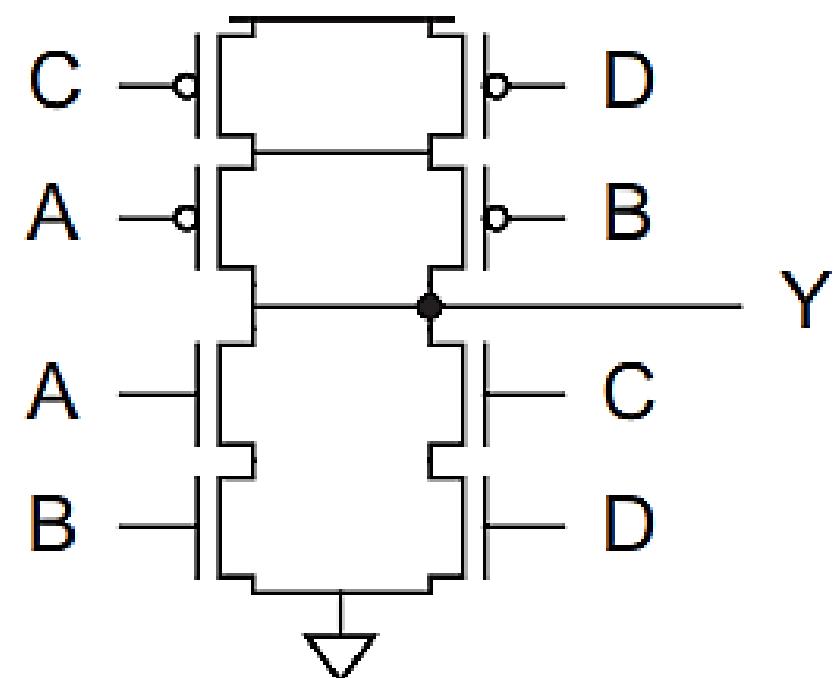
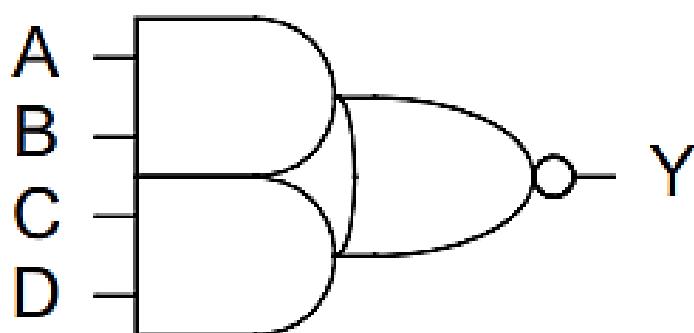


5. Compound gates

$$Y = \overline{(A \cdot B) + (C \cdot D)}$$

This function is sometimes called AND-OR-INVERT-22, or AOI22

Operation	AND	OR
nMOS	Series	Parallel
PMOS	Parallel	Series



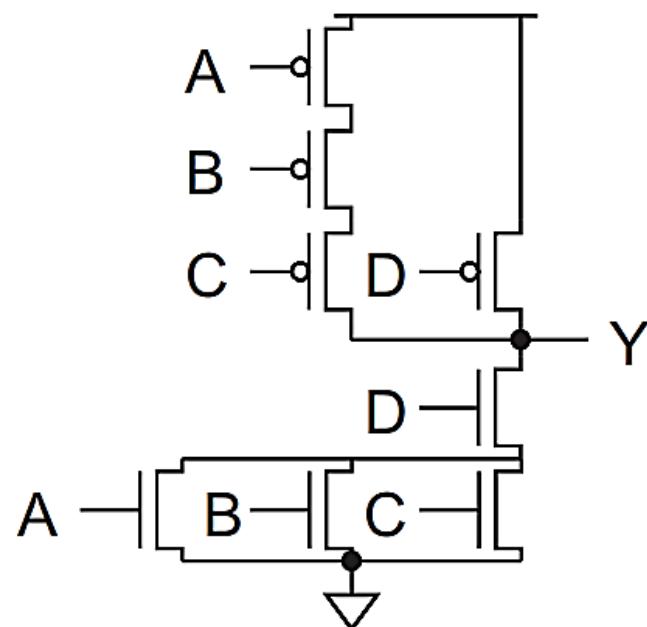


AOI (AND OR INVERT)

- $\overline{AB} + C$ is known as a 2-1 AOI gate.
- $\overline{AB} + \overline{CD}$ is known as a 2-2 AOI gate.
- $\overline{ABC} + \overline{DEF}$ is known as a 3-3 AOI gate.
- $\overline{ABCD} + \overline{EFGH}$ is known as a 4-4 AOI gate.
- $\overline{ABCDE} + \overline{FGH} + \overline{JK}$ is known as a 4-3-2 AOI gate.
- and other variations.

Sketch a static CMOS gate computing $Y = \overline{(A + B + C) \cdot D}$.

Figure 1.19 shows such an OR-AND-INVERT-3-1 (OAI31) gate.



Operation	AND	OR
nMOS	Series	Parallel
PMOS	Parallel	Series



6. Pass transistors and Transmission gates

- > V_{dd} is the rich source of “strong 1”.
- Gnd is the rich source of “strong 0”.
- An nMOS transistor is an almost perfect switch when passing a '0' and thus we say it passes a strong "0".
- However, the nMOS transistor is imperfect at passing a '1'. It passes “weak 1”.
- A pMOS transistor is an almost perfect switch when passing a '1' and thus we say it passes a strong “1”.
- However, the pMOS transistor is imperfect at passing a '0. It passes “weak 0”.

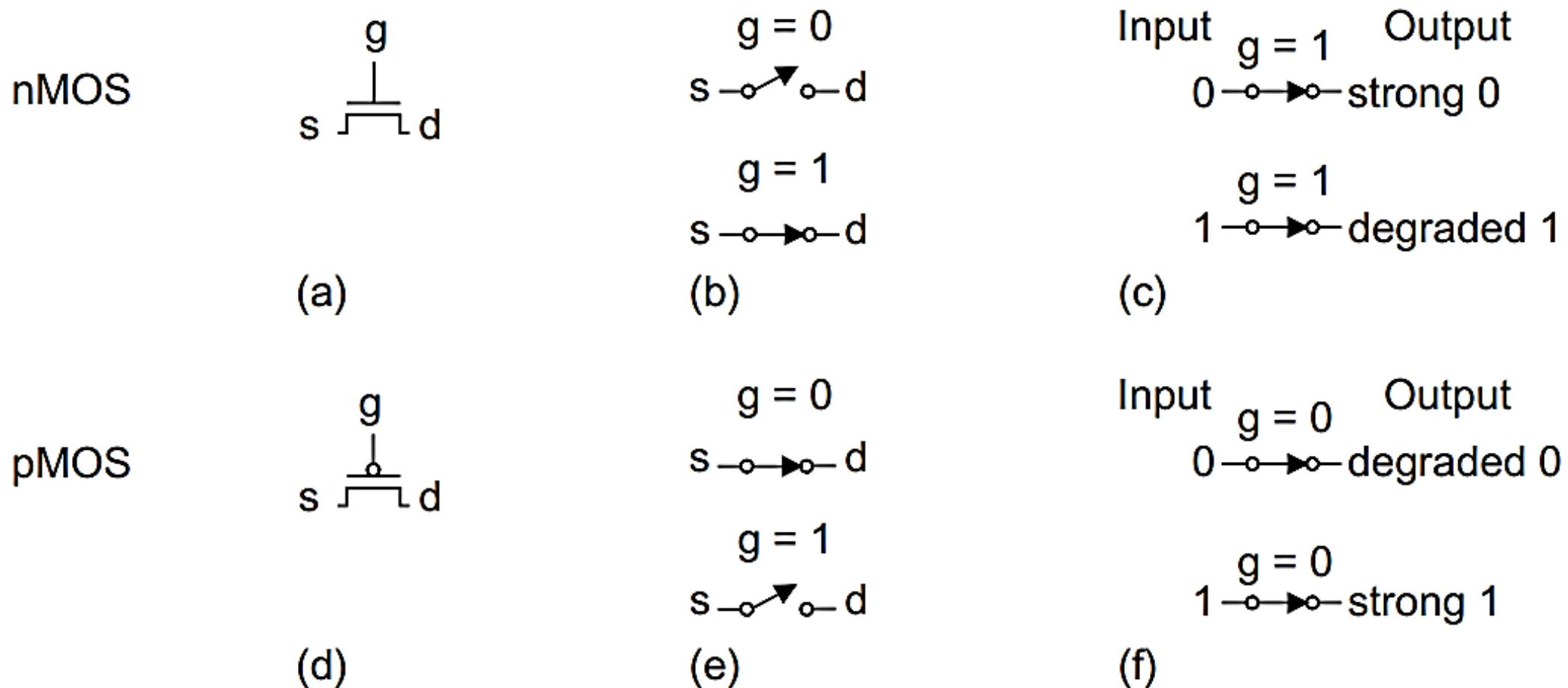


FIGURE 1.20 Pass transistor strong and degraded outputs

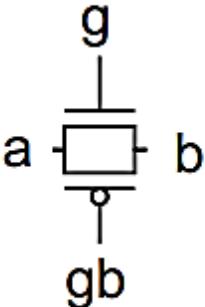
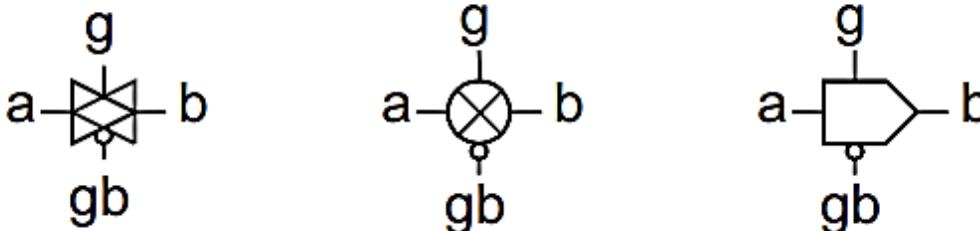
	Input	Output
(a)		$g = 0, gb = 1$ $a \rightarrowtail b$
(b)	$g = 1, gb = 0$ $a \rightarrowtail b$	$g = 1, gb = 0$ $0 \rightarrowtail \text{strong } 0$
(c)		$g = 1, gb = 0$ $1 \rightarrowtail \text{strong } 1$
(d)		

FIGURE 1.21 Transmission gate



7. Tristates

a) Tristate buffer

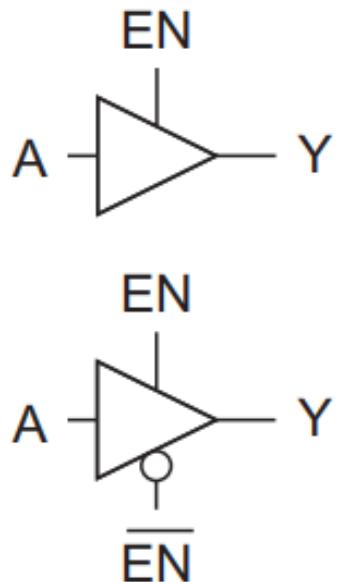


FIGURE 1.25
Tristate buffer
symbol

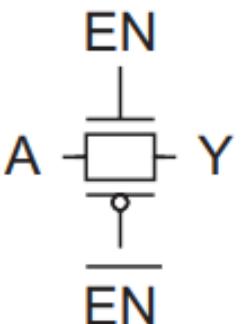


FIGURE 1.26
Transmission gate

TABLE 1.5 Truth table for tristate

EN / \overline{EN}	A	Y
0 / 1	0	Z
0 / 1	1	Z
1 / 0	0	0
1 / 0	1	1



7. Tristates

b) Tristate inverter

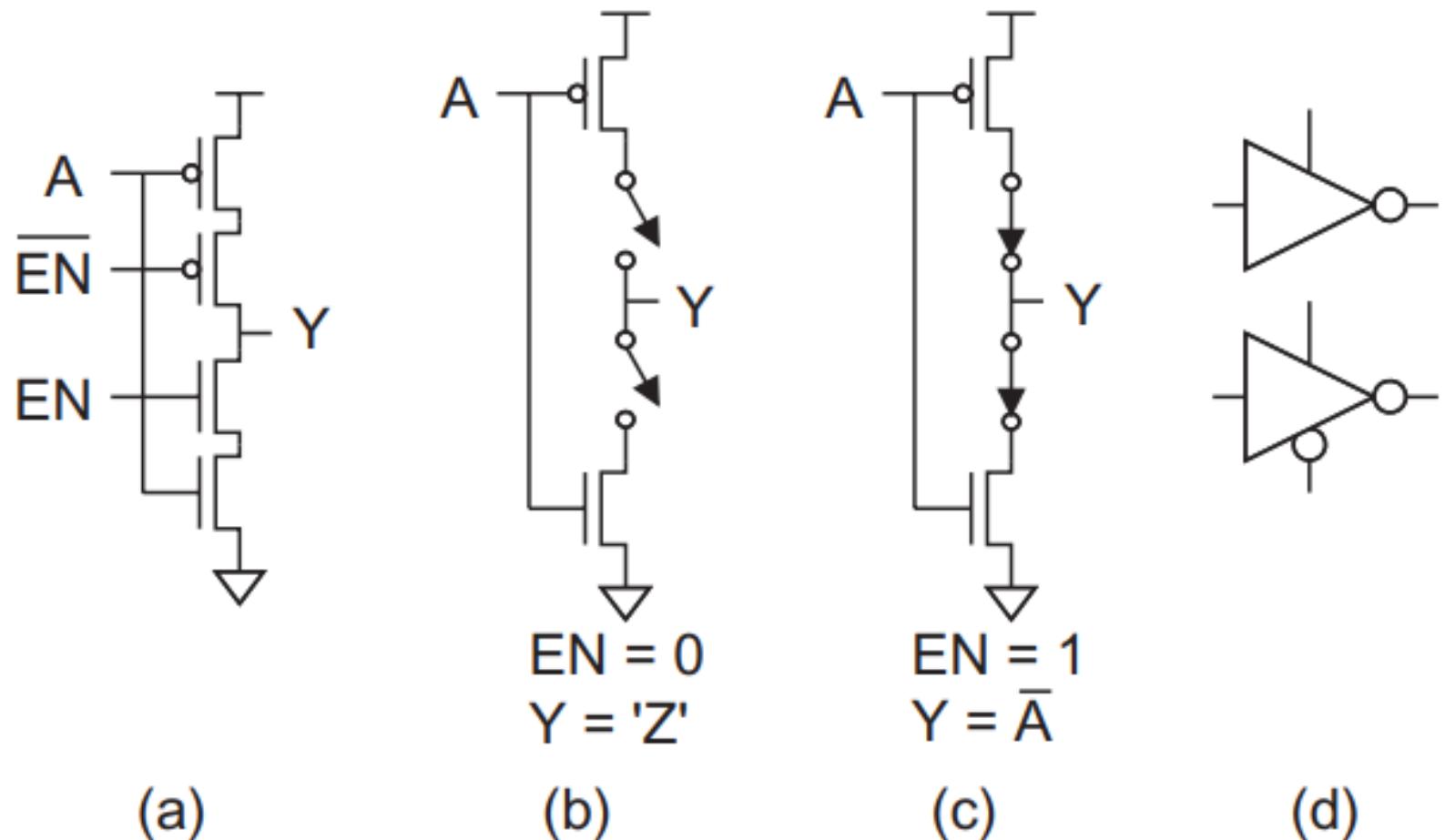


FIGURE 1.27 Tristate Inverter



8. Multiplexers

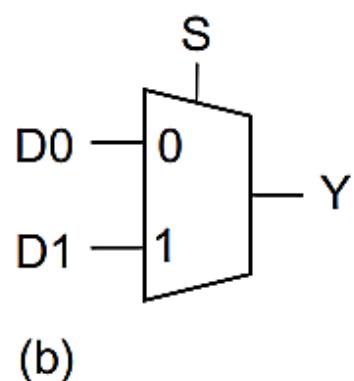
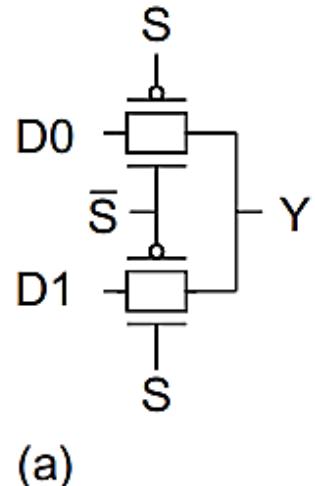
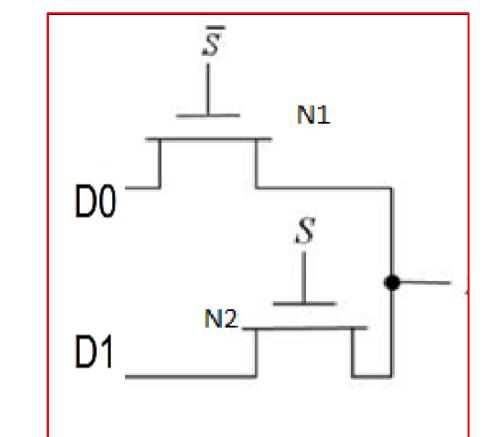


FIGURE 1.28 Transmission gate multiplexer



$$Y = \bar{S} \cdot D0 + S \cdot D1.$$

TABLE 1.6 Multiplexer truth table

S / \bar{S}	$D1$	$D0$	Y
0 / 1	X	0	0
0 / 1	X	1	1
1 / 0	0	X	0
1 / 0	1	X	1

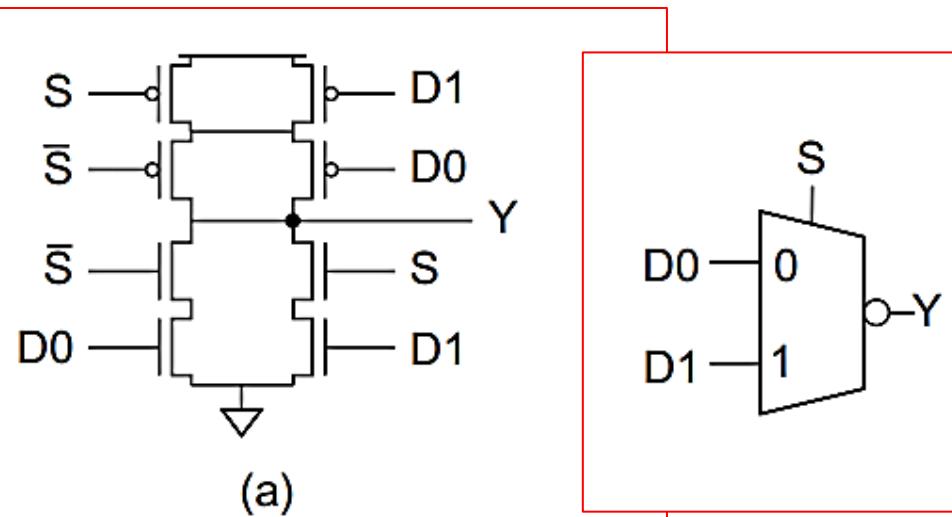


FIGURE 1.29 Inverting multiplexer



9. Latches

When $\text{CLK}=1$, latch is transparent
When $\text{CLK}=0$, latch is opaque

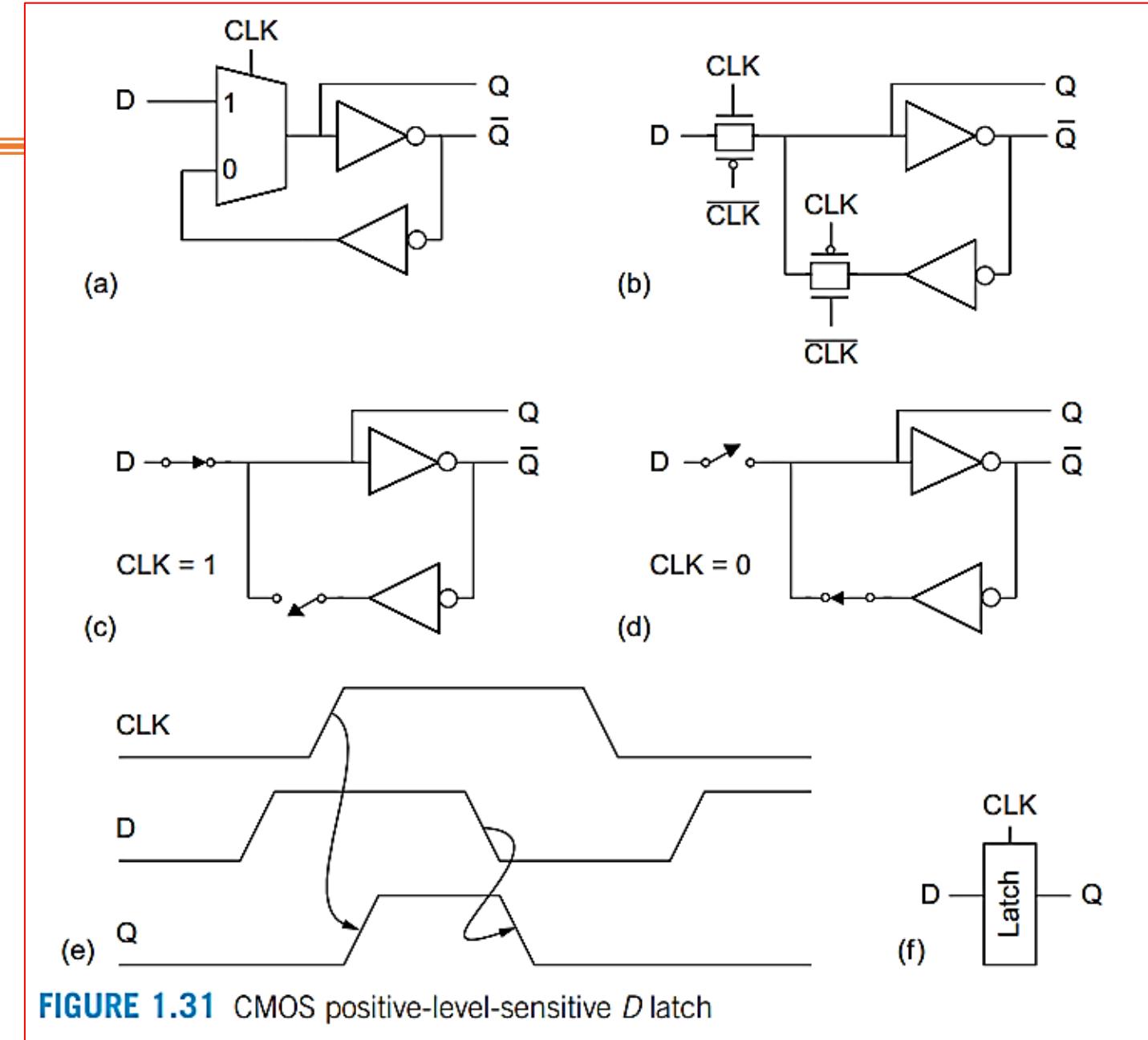
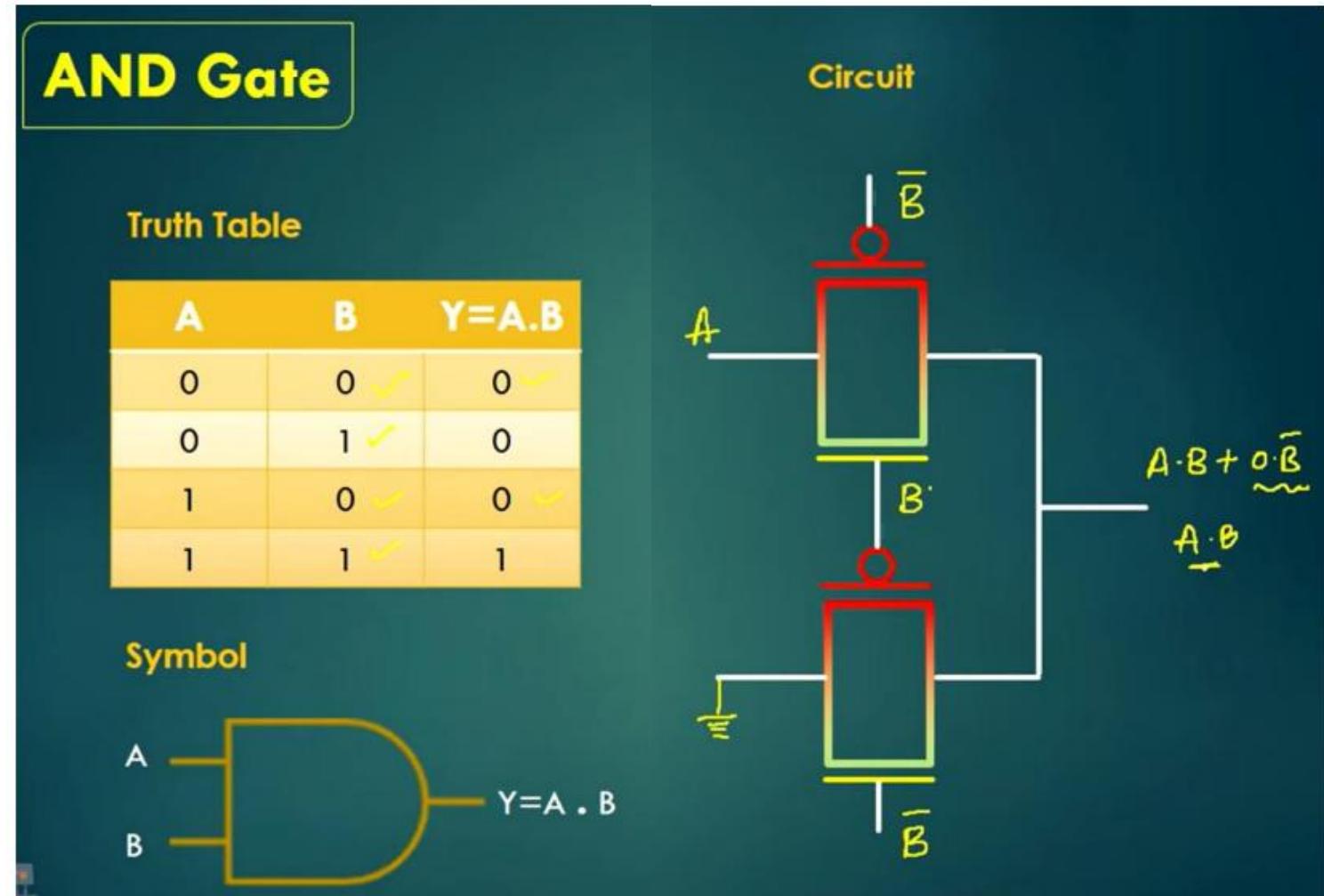
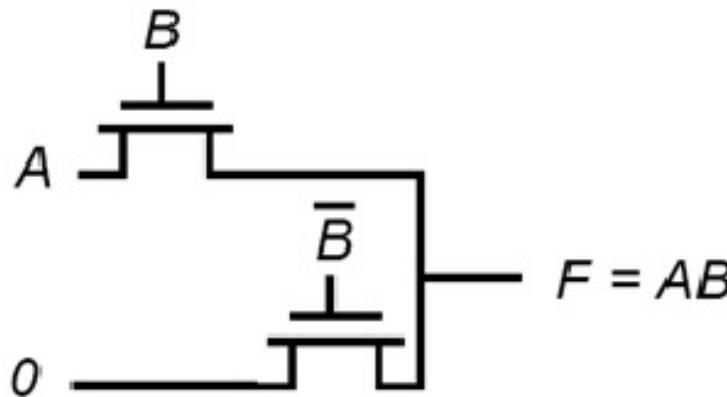
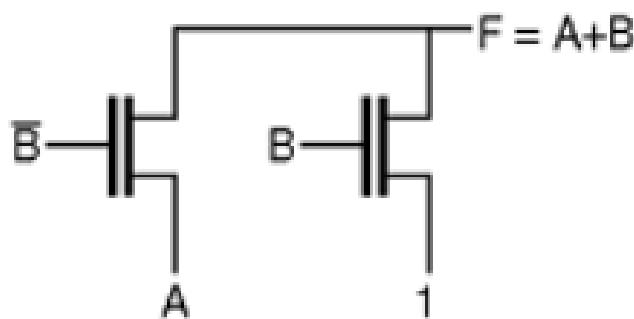


FIGURE 1.31 CMOS positive-level-sensitive D latch



Implementing Boolean expression using PT and TG logic





OR Gate

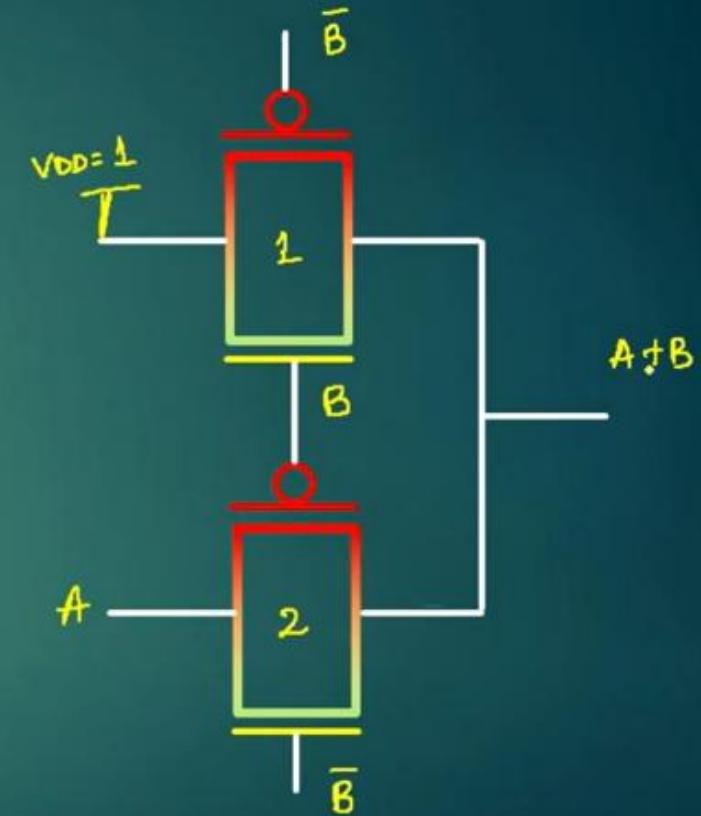
Truth Table

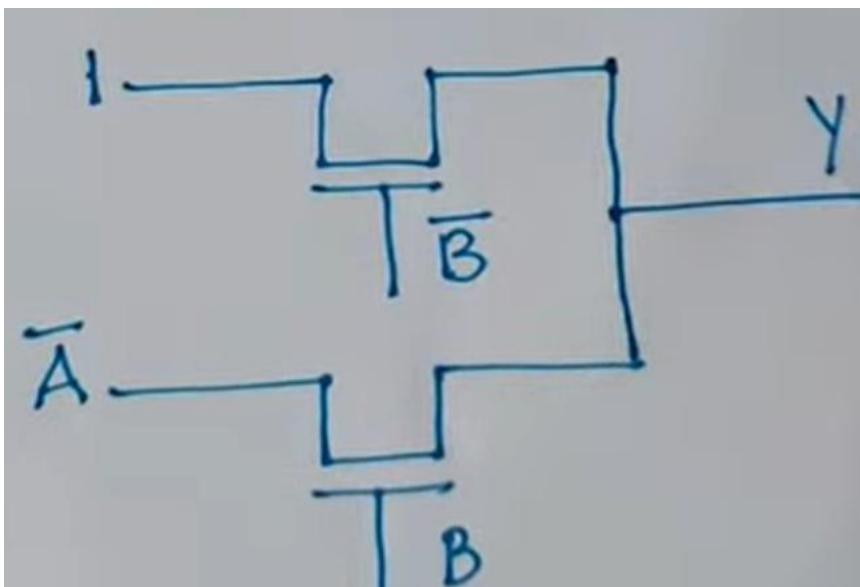
A	B	Y=A+B
0 ✓	0 ✓	0 ✓
0	1 ✓	1 ✓
1 ✓	0 ✓	1 ✓
1	1 ✓	1 ✓

Symbol



Circuit





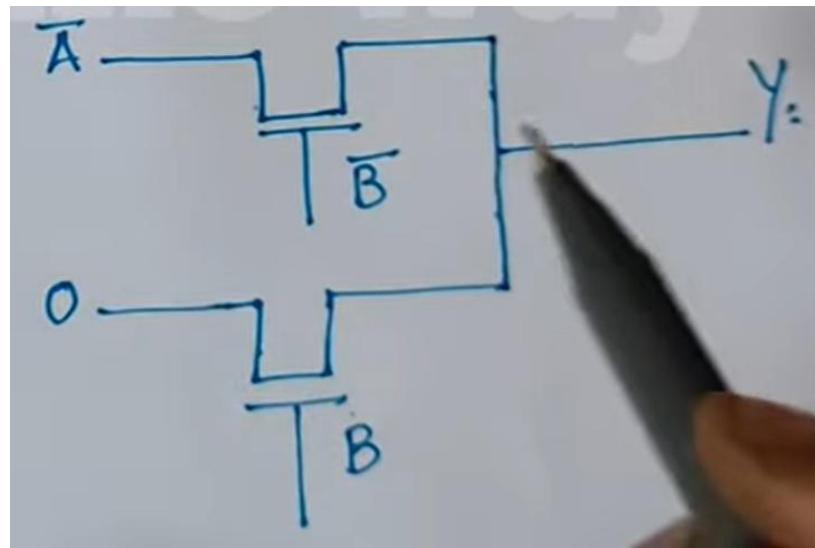
NAND Gate

Truth Table

A	B	$Y = \overline{A \cdot B}$
0	0	1
0	1	1
1	0	1
1	1	0

Symbol

Circuit

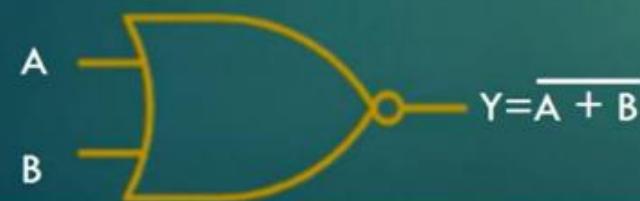


NOR Gate

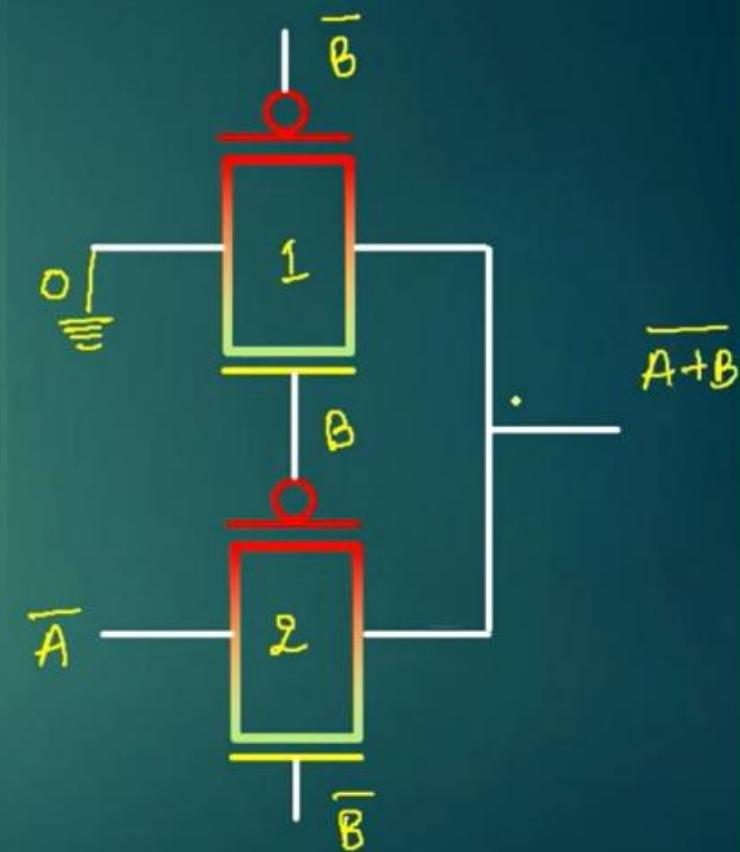
Truth Table

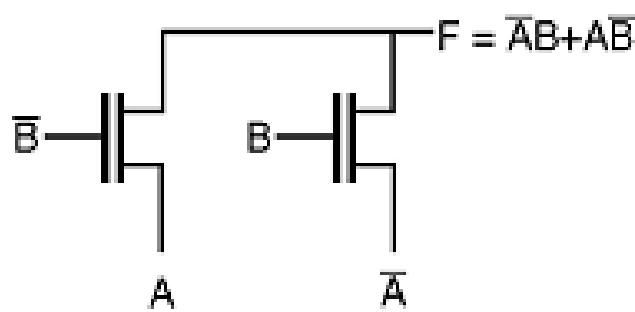
A	B	$Y = \overline{A+B}$
0	0	1
0	1	0
1	0	0
1	1	0

Symbol



Circuit





XOR Gate

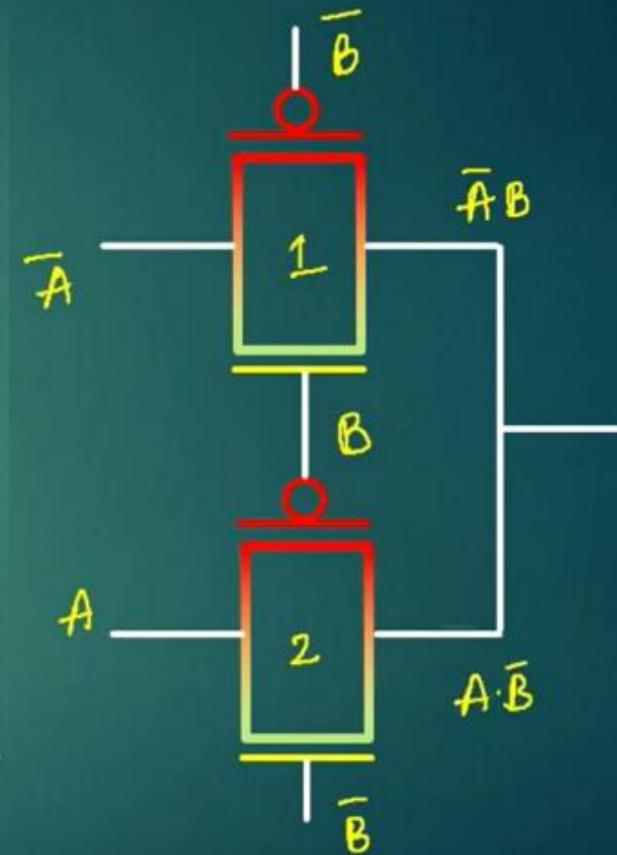
Truth Table

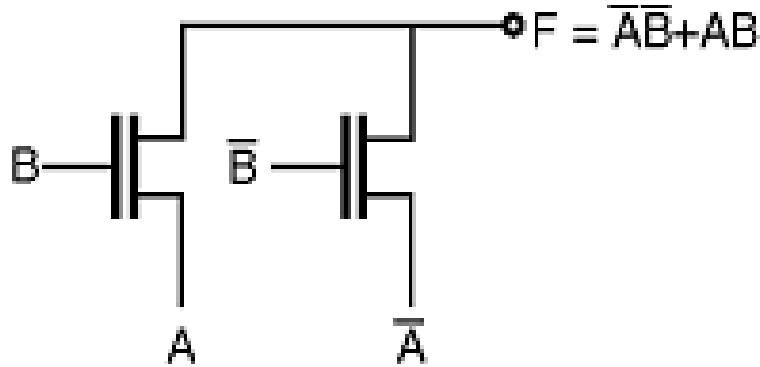
A	B	$Y = A \oplus B$
0 ✓	0 ✓	0 ✓
0 ✓	1 ✓	1 ✓
1 ✓	0 ✓	1 ✓
1 ✓	1 ✓	0 ✓

Symbol



Circuit





XNOR Gate

Truth Table

A	B	$Y = \overline{A \oplus B}$
0	0	1
0	1	0
1	0	0
1	1	1

Symbol

$Y = AB + \overline{A}\overline{B}$

$Y = \overline{A \oplus B}$

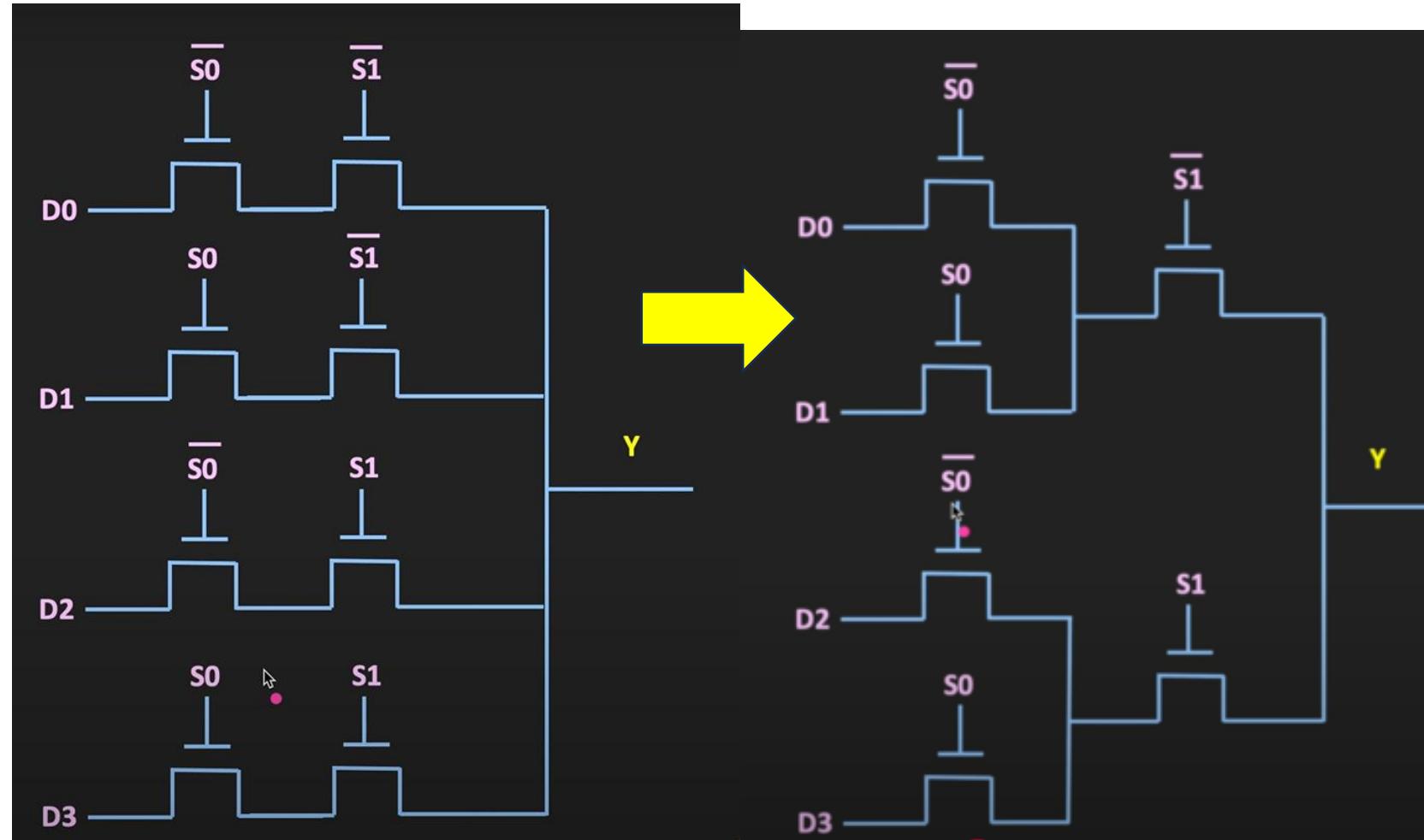
Circuit

A logic circuit diagram showing an XNOR gate. It consists of two AND gates followed by an OR gate. The inputs are labeled A and B. The output is labeled $\overline{A \oplus B}$. The equation for the output is given as $Y = \overline{A \oplus B}$.

4 x 1 MUX



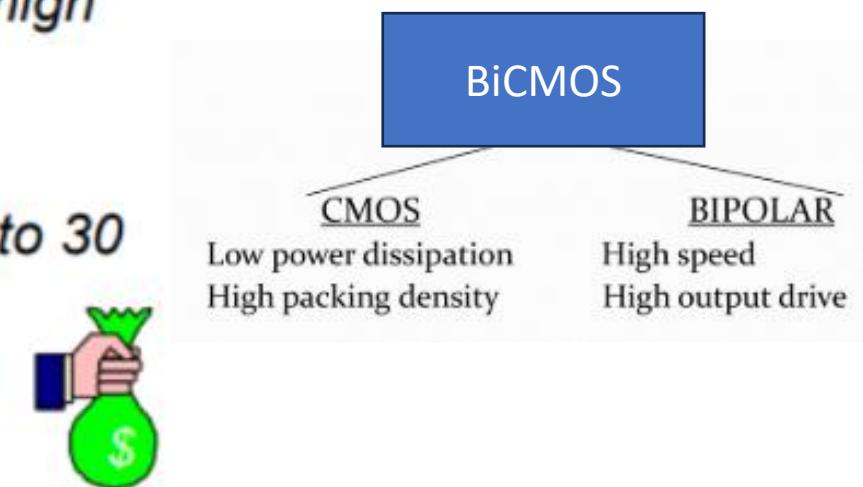
$$Y = \overline{S_1} \overline{S_0} D_0 + \overline{S_1} S_0 D_1 + S_1 \overline{S_0} D_2 + S_1 S_0 D_3$$





BiCMOS –Best of Both Worlds?

- CMOS circuitry exhibits very low power dissipation, but
- Bipolar logic achieves higher speed and current drive capability.
- BiCMOS achieves low standby dissipation like CMOS, but high speed and current drive capability like TTL and ECL.
- The disadvantage of BiCMOS is fabrication complexity (up to 30 masking steps, compared to about 20 for bipolar logic or CMOS). This translates into higher cost and longer design cycles.
- Notable examples of the BiCMOS technology are the Intel P6 (a.k.a. Pentium Pro) which appeared in 1996, and its successor the P7.





Comparison between CMOS and Bipolar

Factors	CMOS	Bipolar
Static Power Dissipation	Low	High
Input Impedance	High	Low
Noise Margin	High	Low
Packing Density	High	Low
Fan-out	Low	High
Direction	Bidirectional	Unidirectional

BiCMOS inverter

- When $V_{in} = 0$, T_3 is off so that T_1 will be non-conducting. But T_4 is ON and supplies current to the base of T_2 which will conduct and act as a current source to charge C_L towards 5V. Output of Inverter will be 5V minus V_{BE} of T2
- With $V_{in} = 5V$, T_4 is OFF so that T_2 will be non-conducting. But T_3 will be ON and will supply current to base of T_1 which will conduct and act as a current sink to C_L discharging it to 0. Output of inverter will fall to 0 volts plus V_{BEsat} of T1.
- Owing to the presence of direct path from V_{DD} to Gnd through T_3 and T_1 , this is not good arrangement to implement since there is significant **static current** flow whenever V_{in} is logic 1.

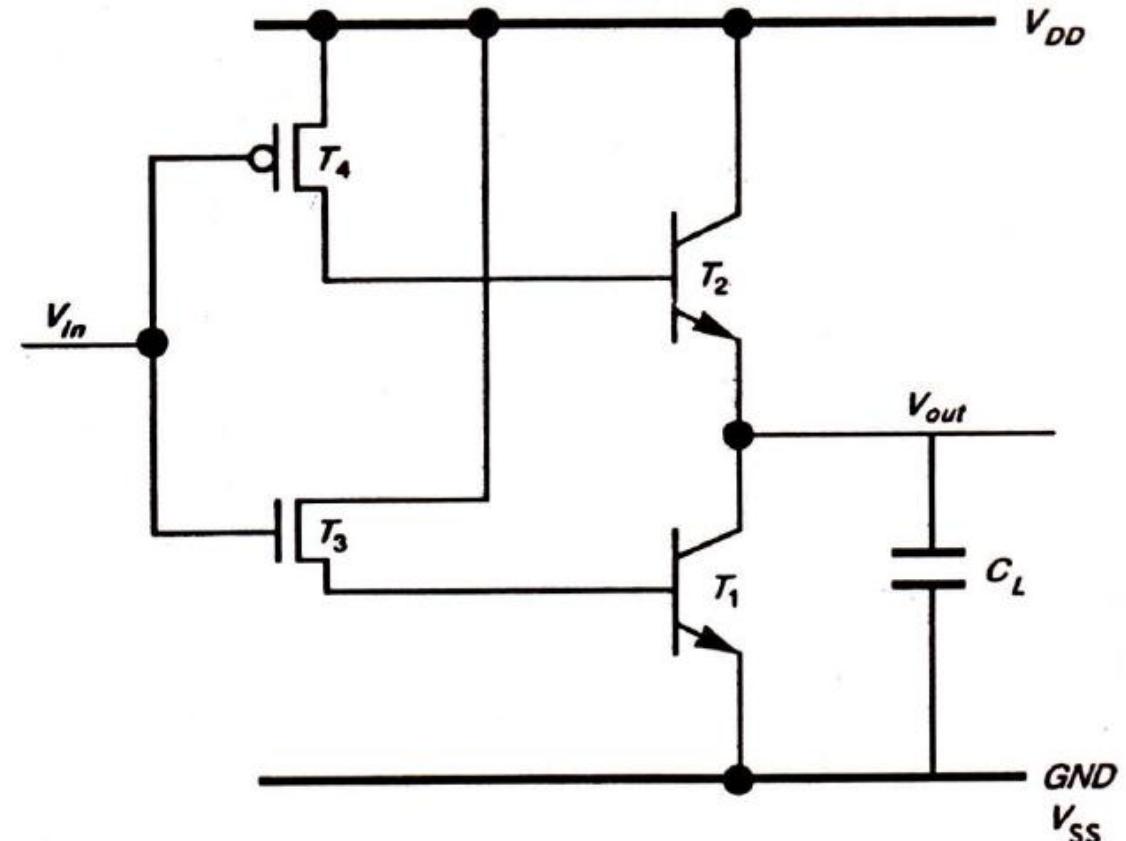


FIGURE 2.17 A simple BiCMOS inverter.

BiCMOS inverter and logic gates

- An improved version of BiCMOS inverter is shown in Figure, in which DC path through T_3 and T_1 is eliminated.
- In both above circuits, there is no discharge path for the current from the base of BJT when it is being turned off.

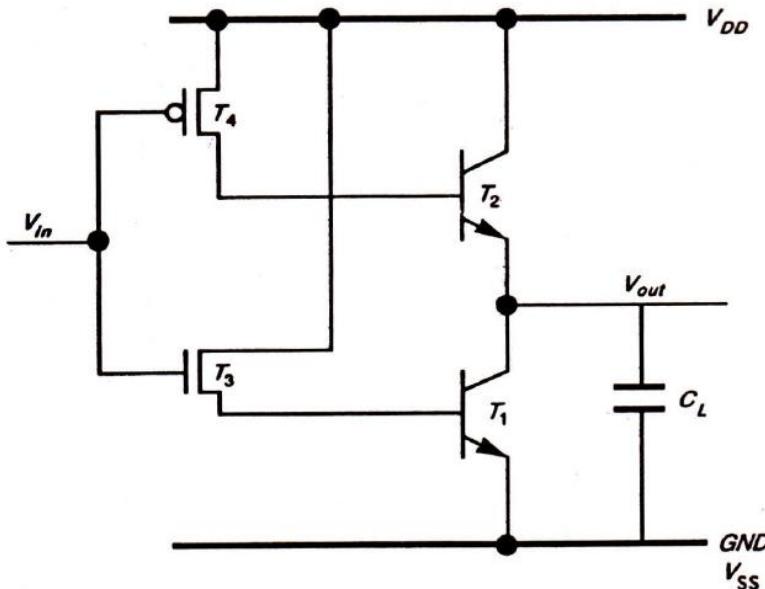


FIGURE 2.17 A simple BiCMOS inverter.

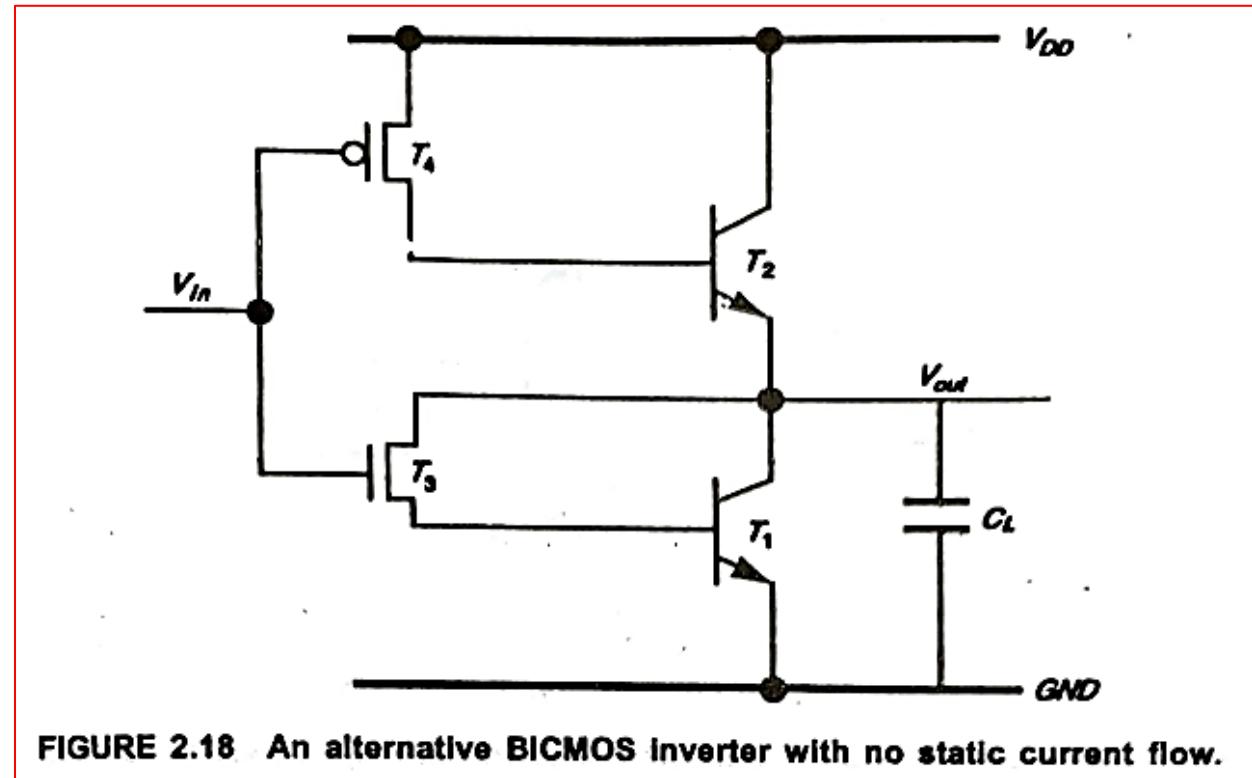


FIGURE 2.18 An alternative BiCMOS inverter with no static current flow.



BiCMOS inverter and logic gates

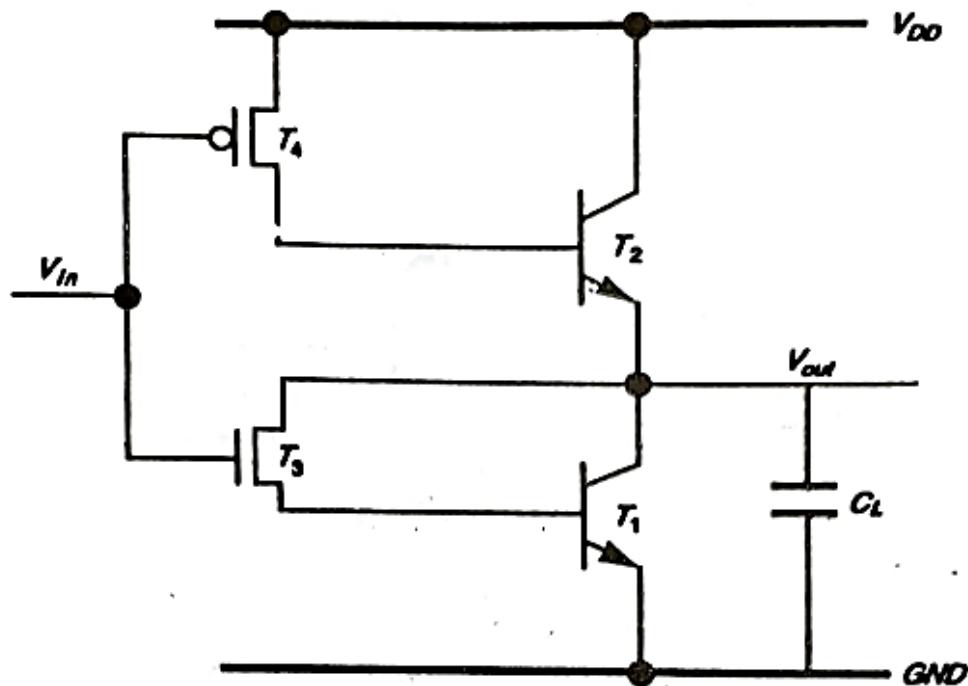


FIGURE 2.18 An alternative BiCMOS inverter with no static current flow.

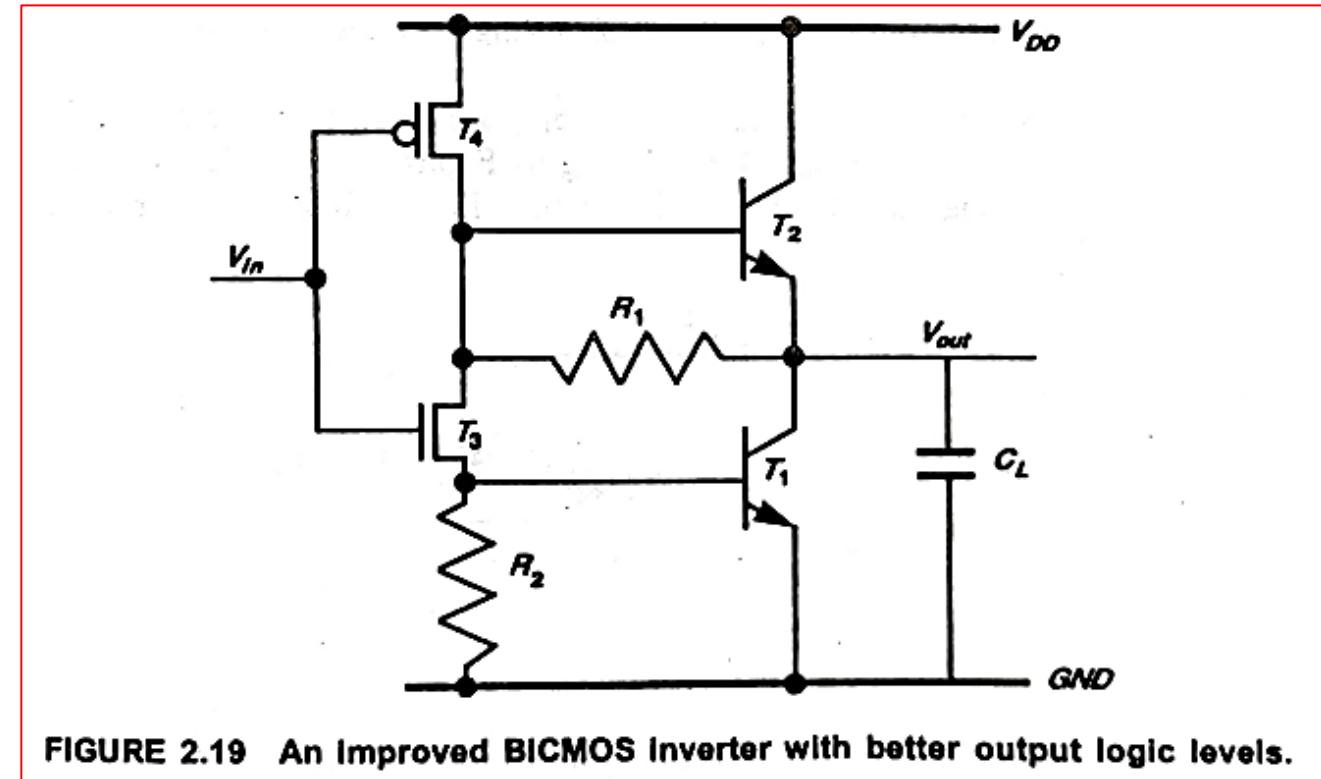


FIGURE 2.19 An Improved BiCMOS Inverter with better output logic levels.

BiCMOS inverter

- An improved inverter arrangement is shown in Figure.
- This provides improved output voltage swing when each BJT is OFF., also provide discharge path for base current during turn-off.

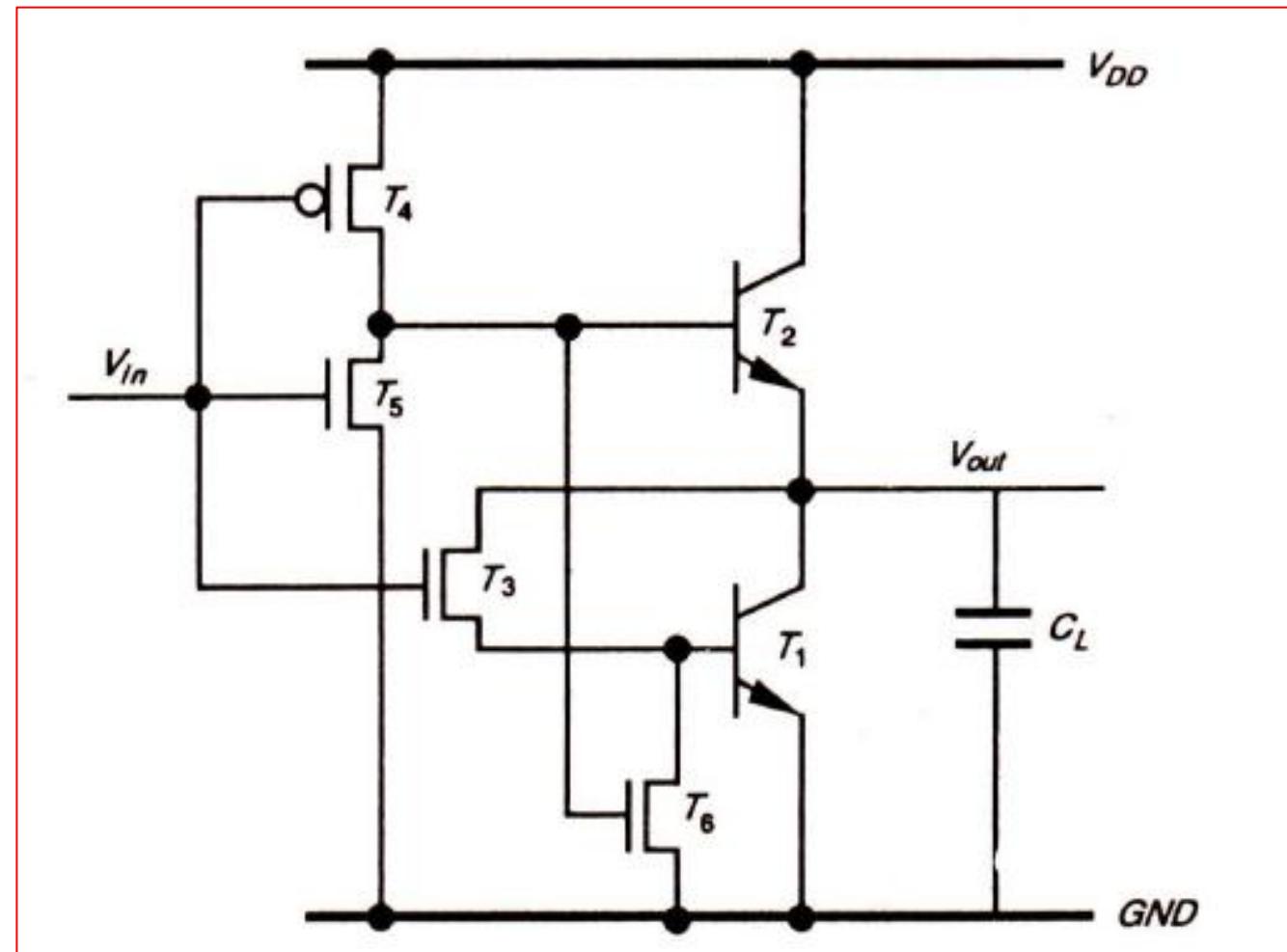
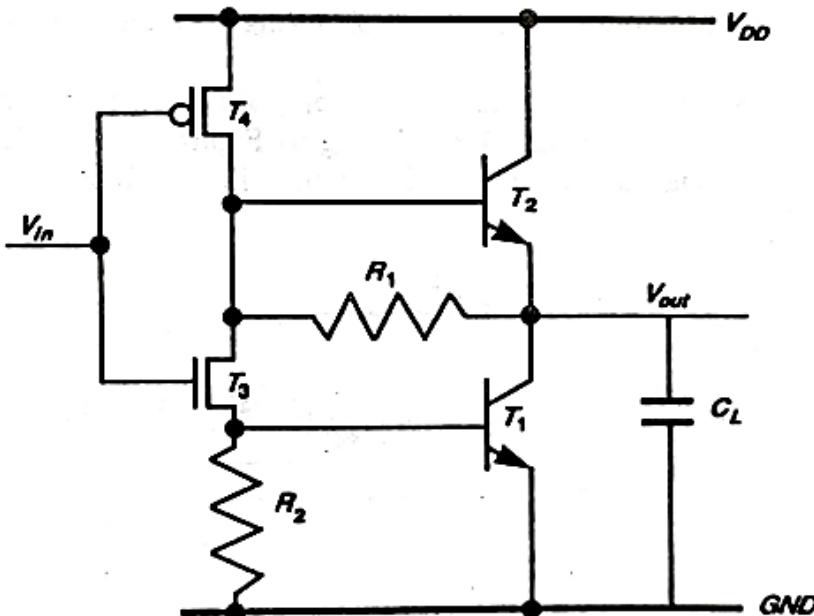
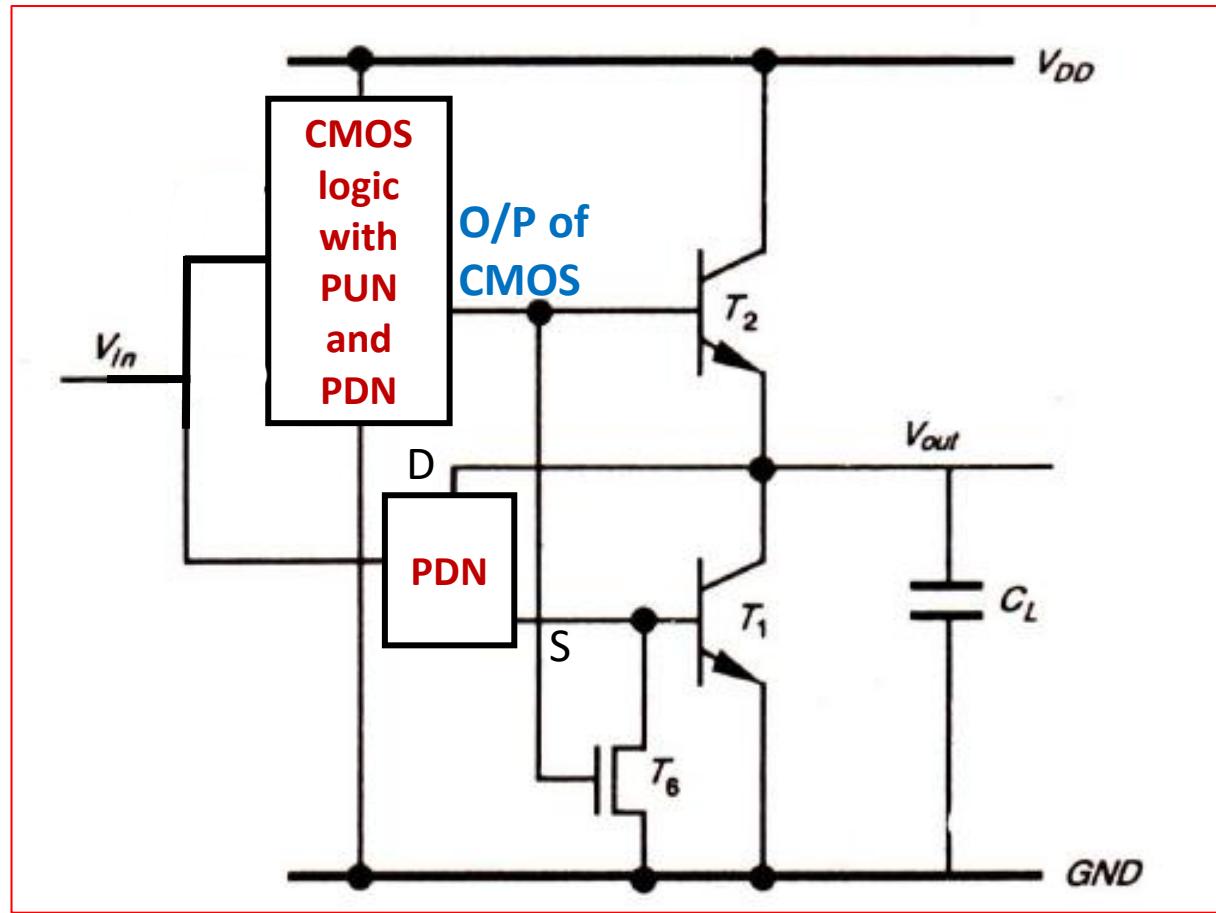


FIGURE 2.19 An Improved BiCMOS Inverter with better output logic levels.

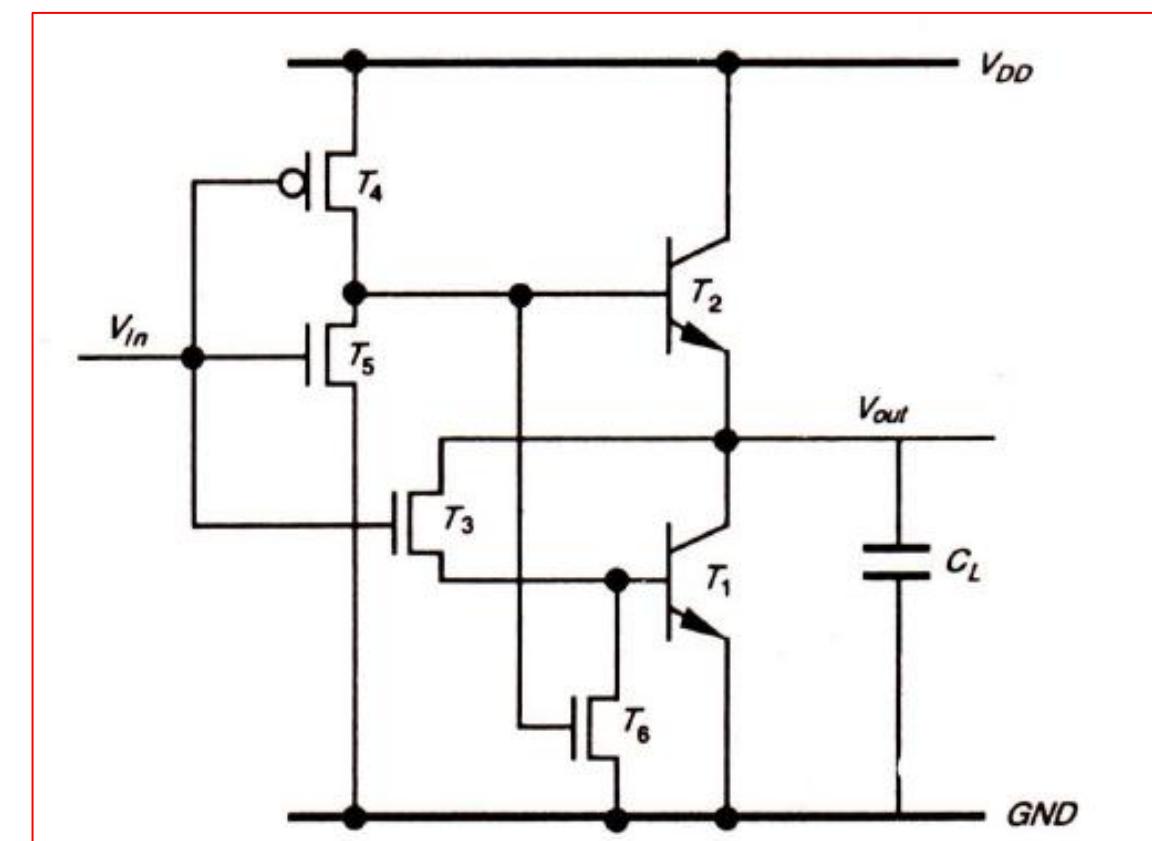


BiCMOS logic gates

General circuit



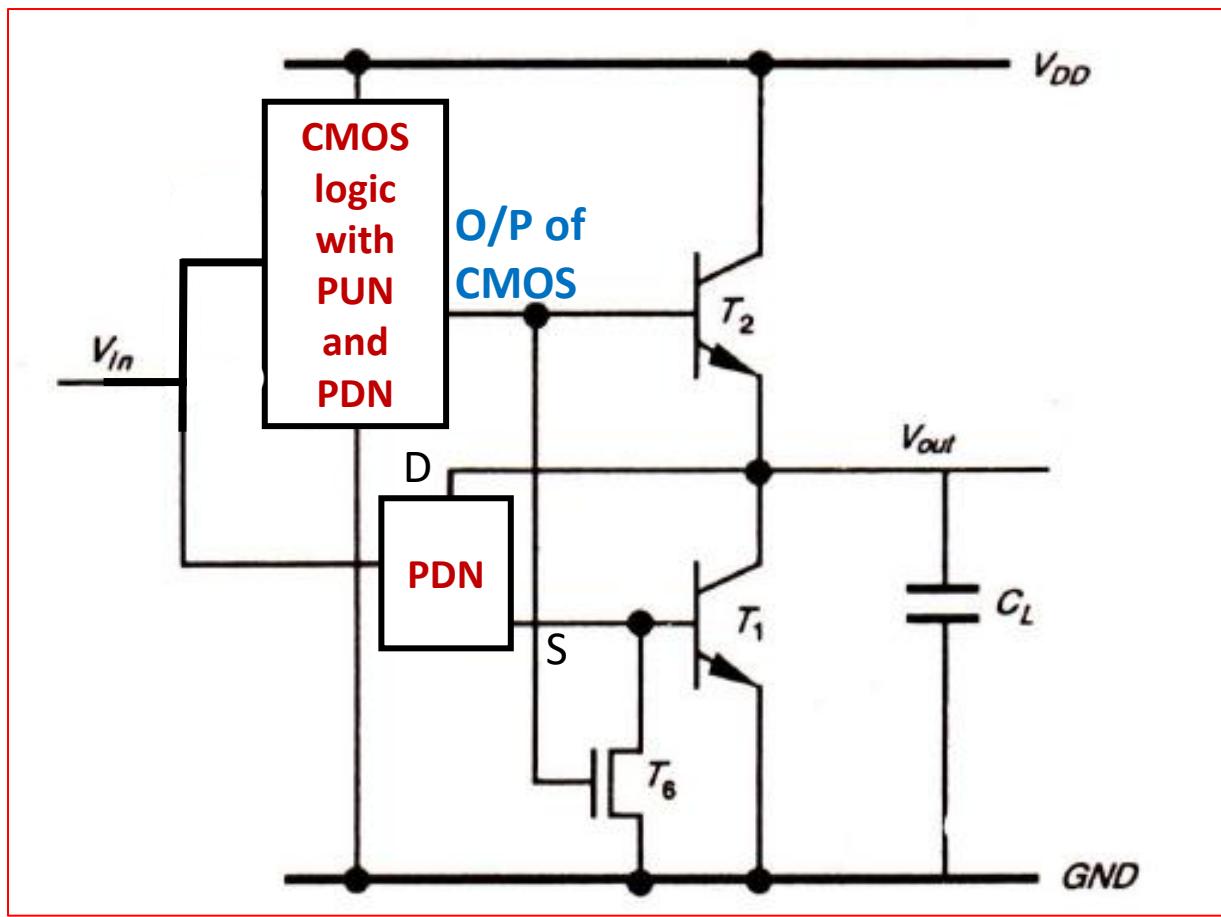
BiCMOS Inverter



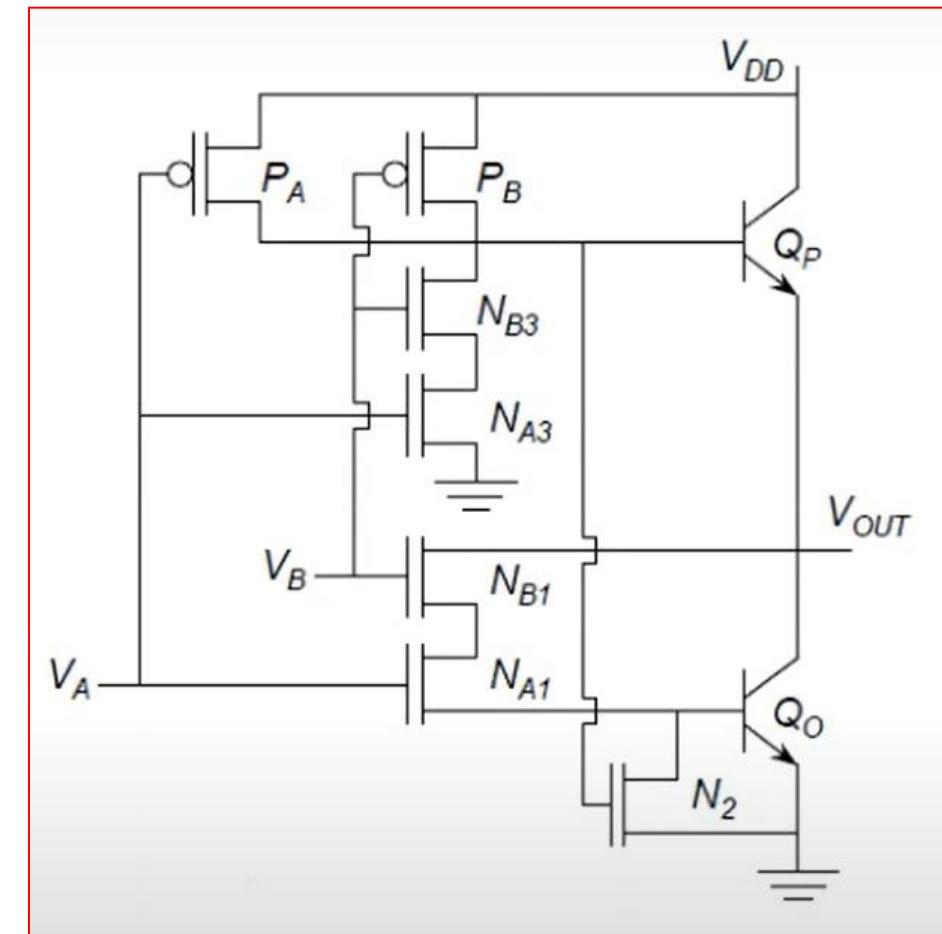


BiCMOS logic gates

General circuit



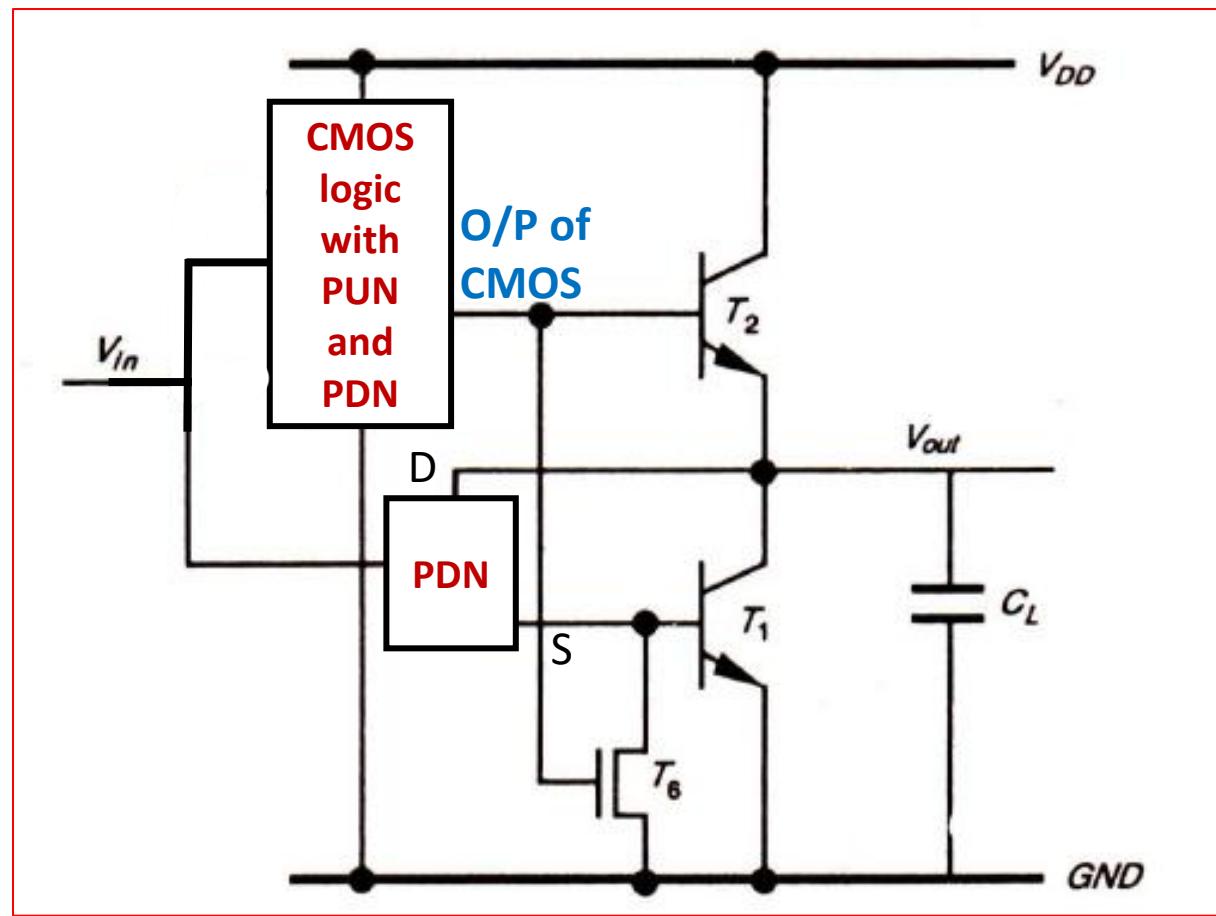
BiCMOS NAND gate



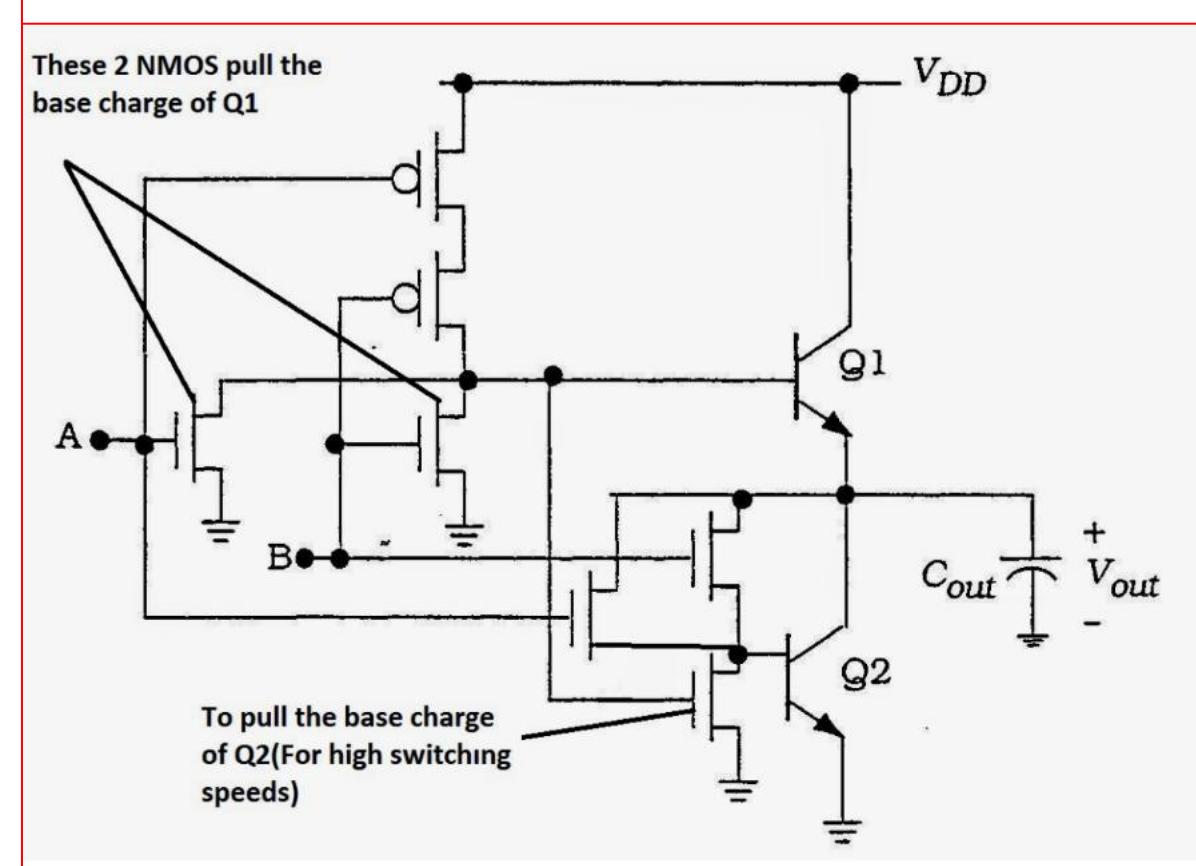


BiCMOS logic gates

General circuit



BiCMOS NOR gate

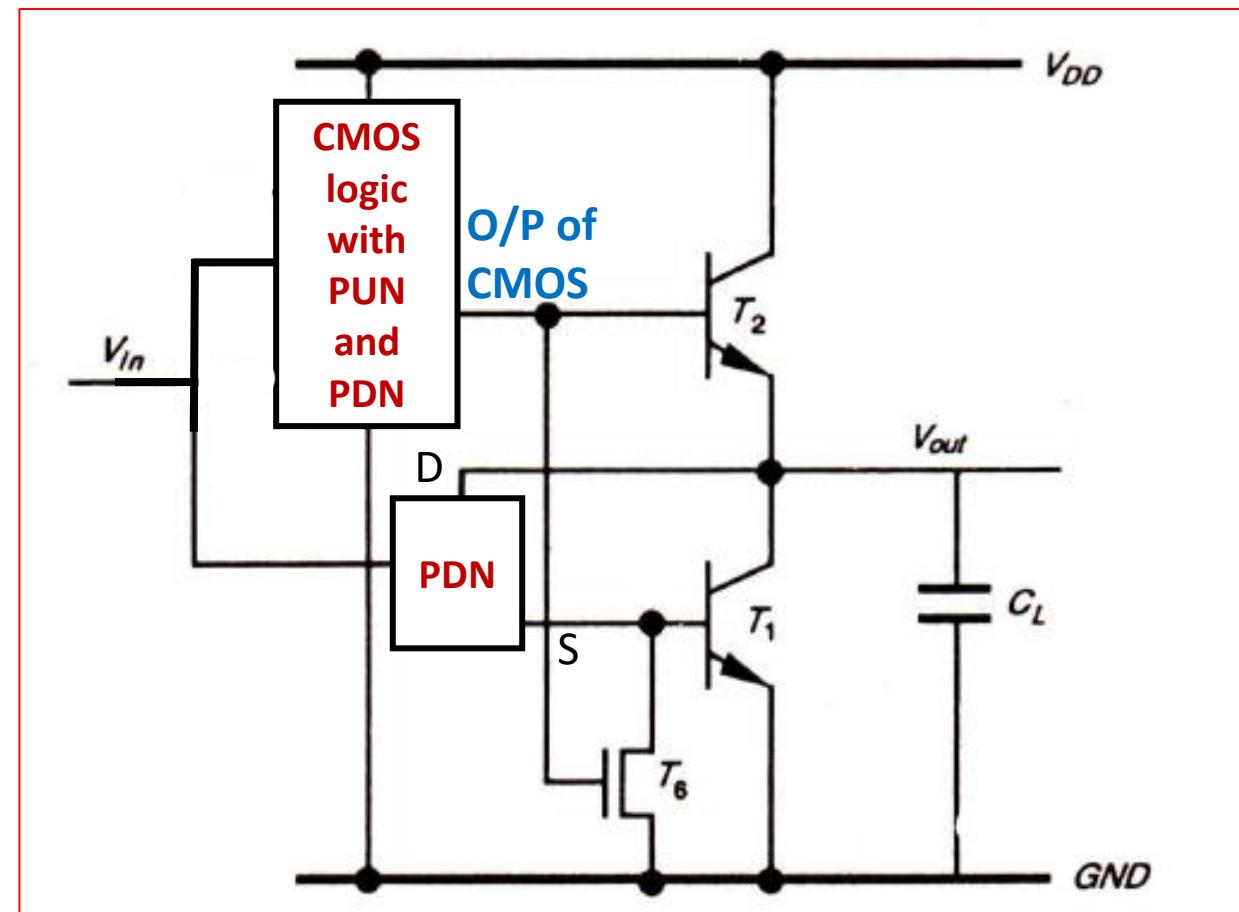




BiCMOS logic gates

Implement the following using BiCMOS logic

1. $Y = (AB+C)'$
2. $Y=(AB+CD)'$
3. AOI 43
4. OAI 32





Advantages of BiCMOS technology

- It follows that BiCMOS technology goes some way towards combining the virtues of both CMOS and Bipolar technologies
- Improved speed over purely-CMOS technology
- Lower power dissipation than purely-bipolar technology(Lower power consumption than bipolar)
- Flexible I/Os for high performance
- Improved current drive over CMOS
- Improved packing density over bipolar
- High input impedance
- Low output impedance
- High Gain and low noise



Other forms of CMOS logic

1. Pseudo nMOS logic
2. Dynamic CMOS logic
3. CMOS domino logic
4. Clocked CMOS (C^2MOS) logic



1. Pseudo nMOS inverter with derivation of Z_{pu}/Z_{pd} ratio

Clearly, if we replace the depletion mode pull-up transistor of the standard nMOS circuits with a p-transistor with gate connected to V_{ss} (Gnd), we have a structure similar to the nMOS equivalent.

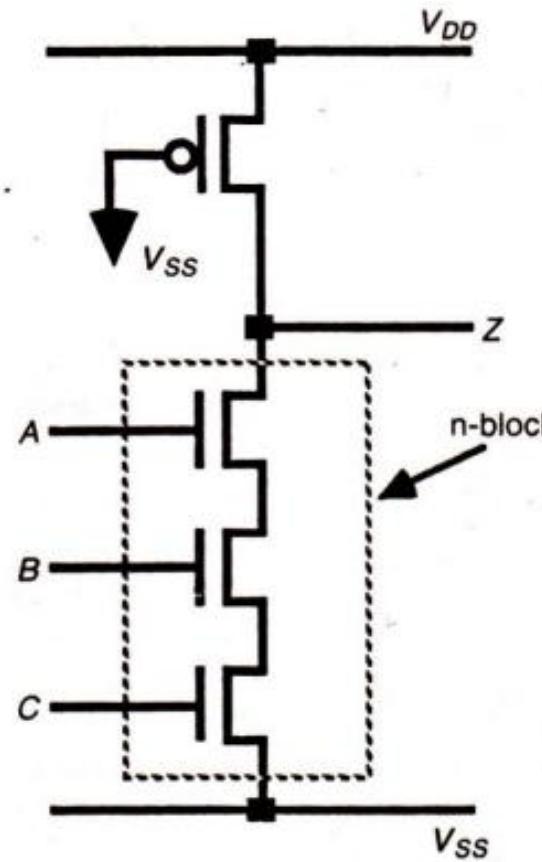
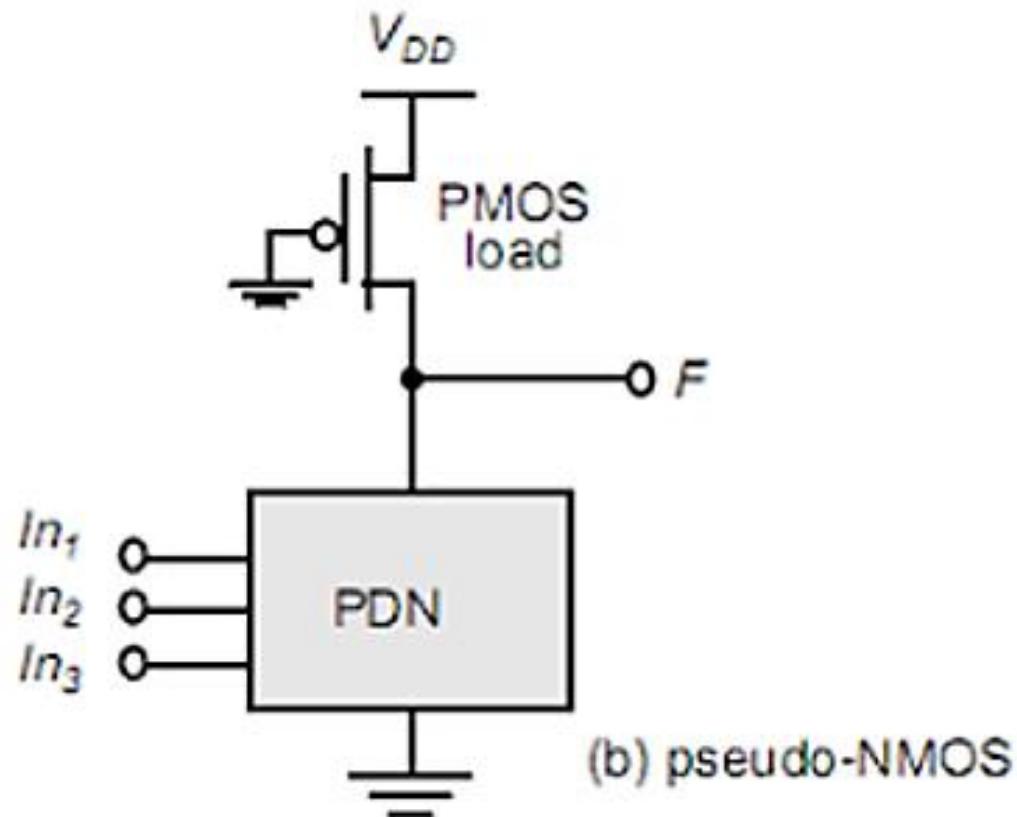


FIGURE 6.9 Pseudo-nMOS Nand gate.

- Consider a situation where on pseudo-nMOS inverter is being driven by another similar inverter.
- As for the nMOS analysis, we consider the conditions for which,

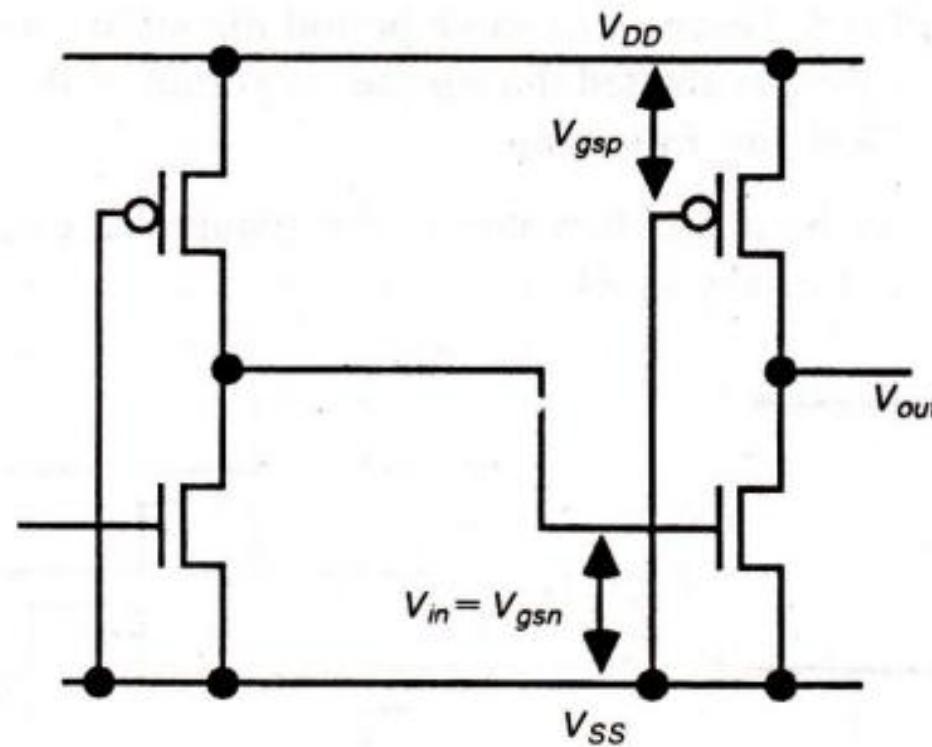
$$V_{inv} = V_{DD}/2$$

- At this point n-device is in saturation,

$$0 < V_{gsn} - V_{tn} < V_{dsn}$$

- p-device is in linear,

$$0 < V_{dsp} < V_{gsp} - V_{tp}$$



Pseudo-nMOS inverter when driven from a similar inverter.



Equating currents of the n-transistor and the p-transistor, and by suitable rearrangement of the resultant expression, we obtain,

$$V_{inv} = V_{tn} + \frac{(2\mu_p/\mu_n)^{1/2} [(-V_{DD} - V_{tn})V_{dsp} - V_{dsp}^2]^{1/2}}{(Z_{p.u.}/Z_{p.d.})^{1/2}}$$

Where, $Z_{p.u.} = L_p/W_p$ $Z_{p.d.} = L_n/W_n$

With, $V_{inv} = 0.5V_{DD}$

$$V_{tn} = |V_{tp}| = 0.2V_{DD}$$

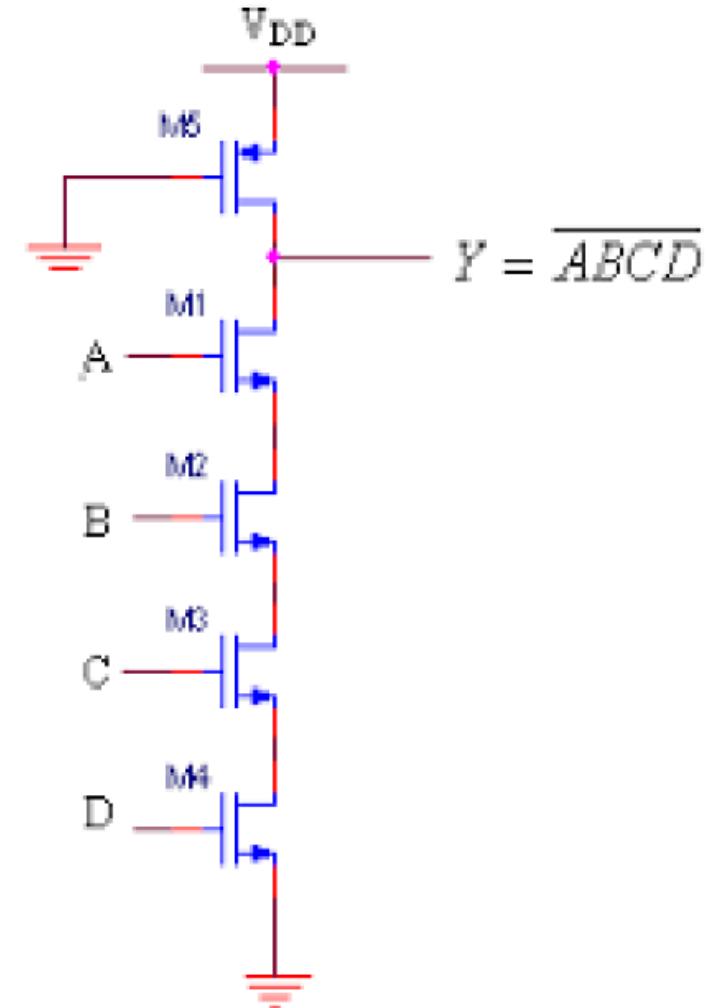
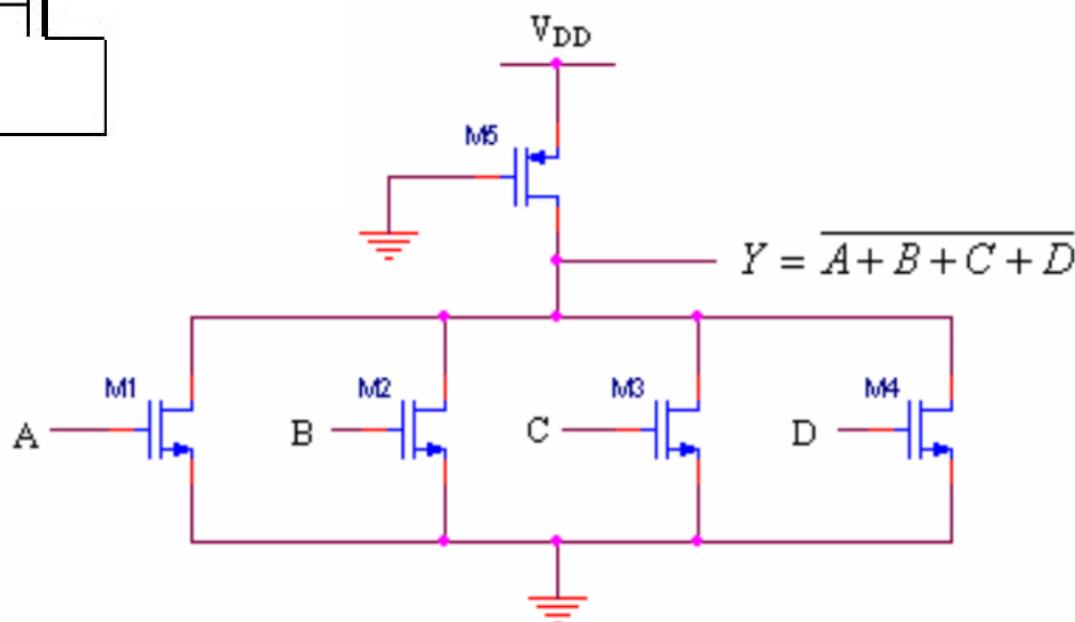
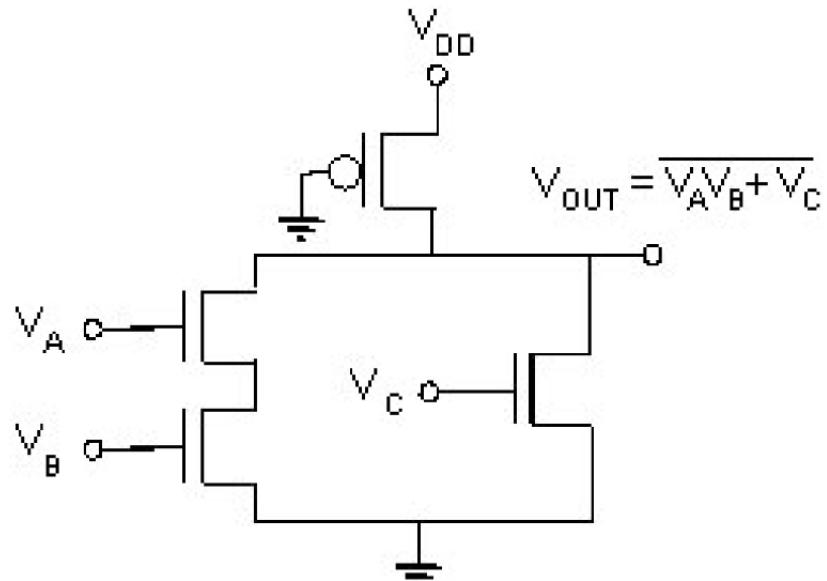
$$V_{DD} = 5 \text{ V}$$

$$\mu_n = 2.5 \mu_p$$

We obtain, $\frac{Z_{p.u.}}{Z_{p.d.}} = \frac{3}{1}$



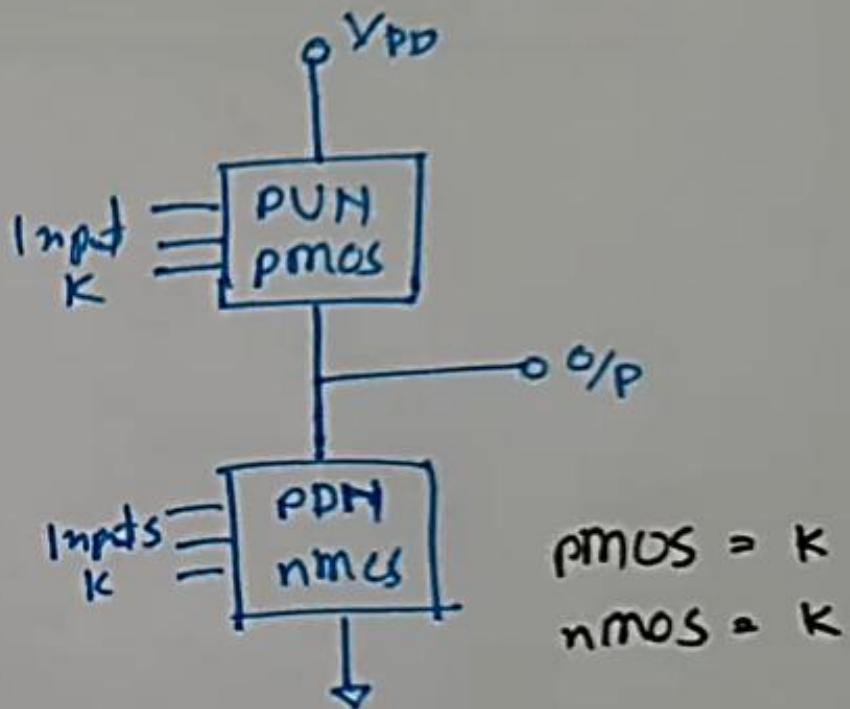
Implement the following Boolean expressions using pseudo nMOS logic



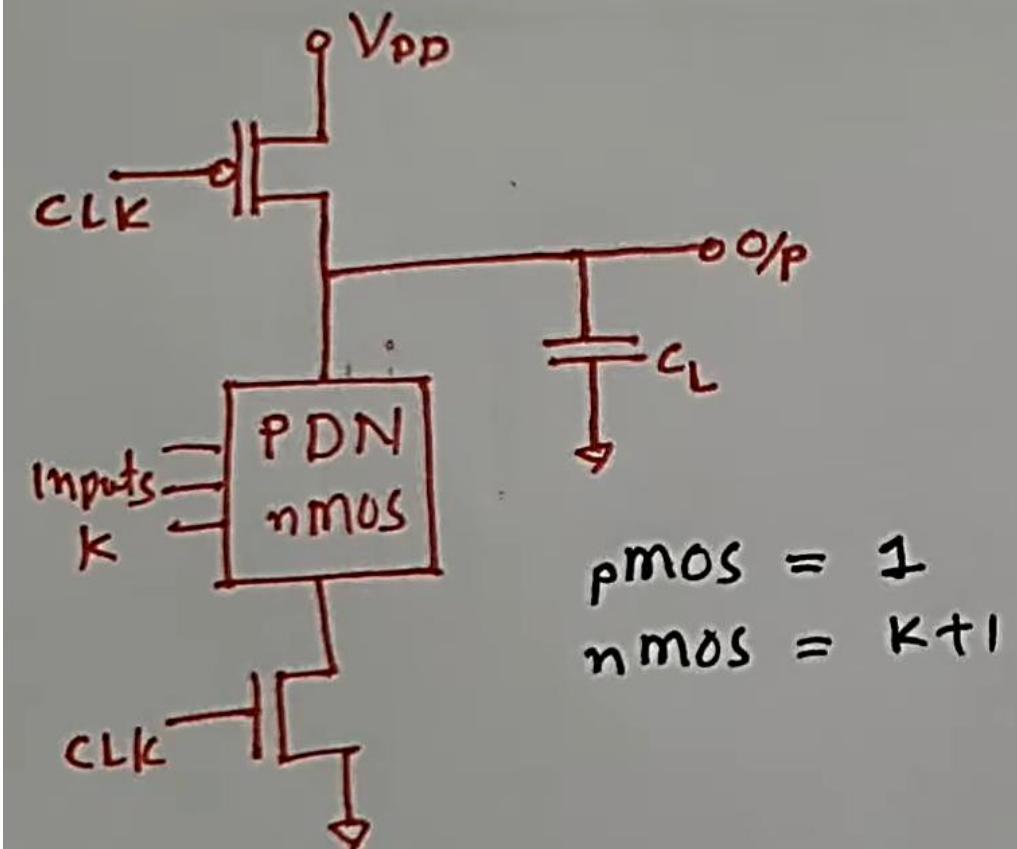


2. Dynamic CMOS inverter

Static CMOS

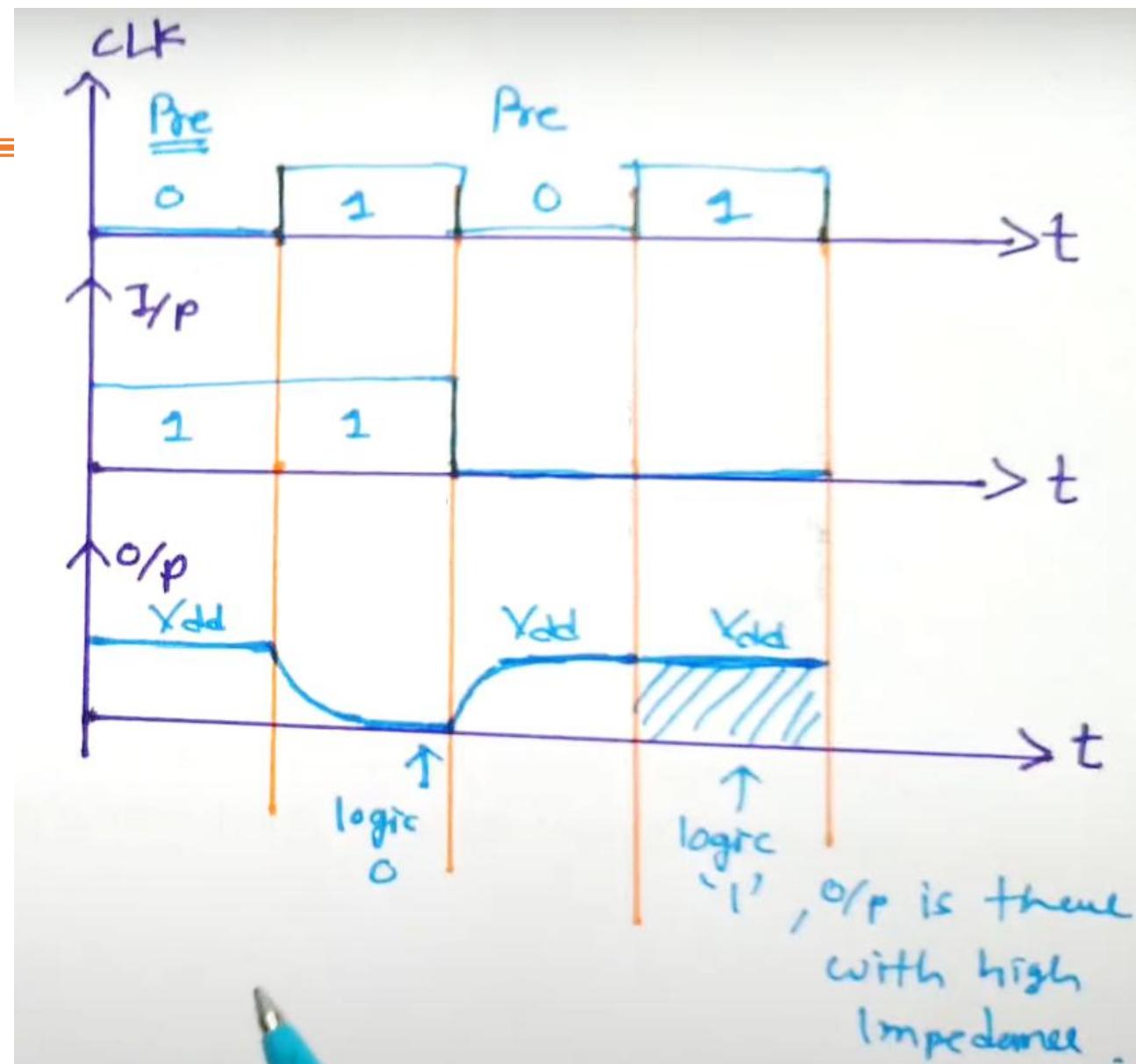
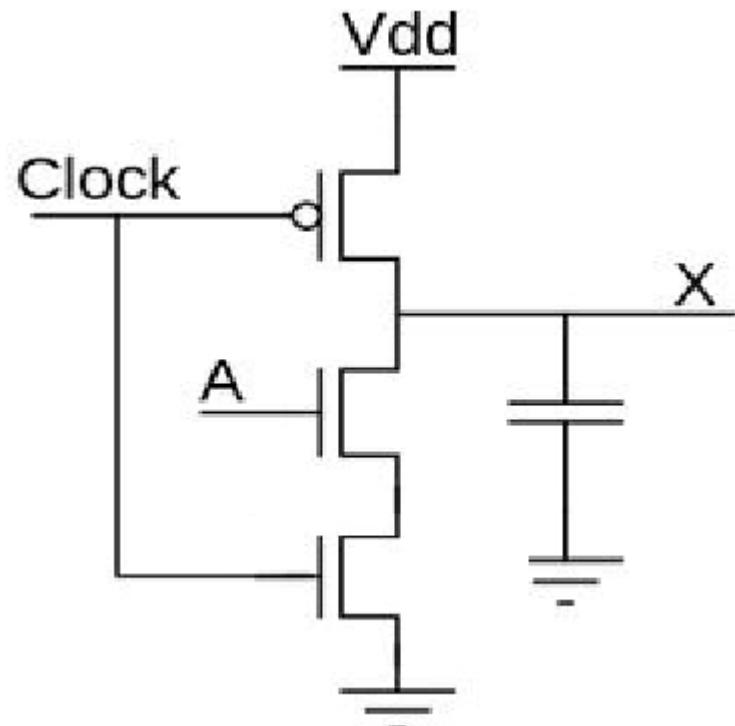


Dynamic CMOS



Circuit works in 2 modes:

1. Pre-charge mode (occurs when clock = 0)
2. Evaluation mode (occurs when clock = 1)

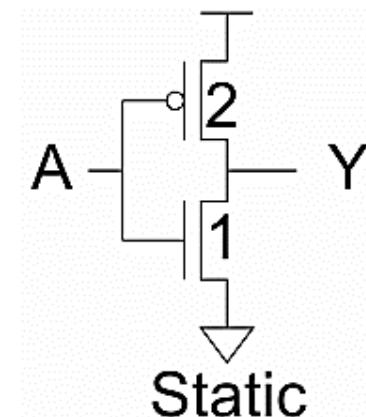


Disadvantages

1. When output is 1 it will be at high impedance mode which is more affected by noise.
2. It requires pre-charging.
3. Cascading issue and Race around condition.
4. Less capacitance loading.

Advantages:

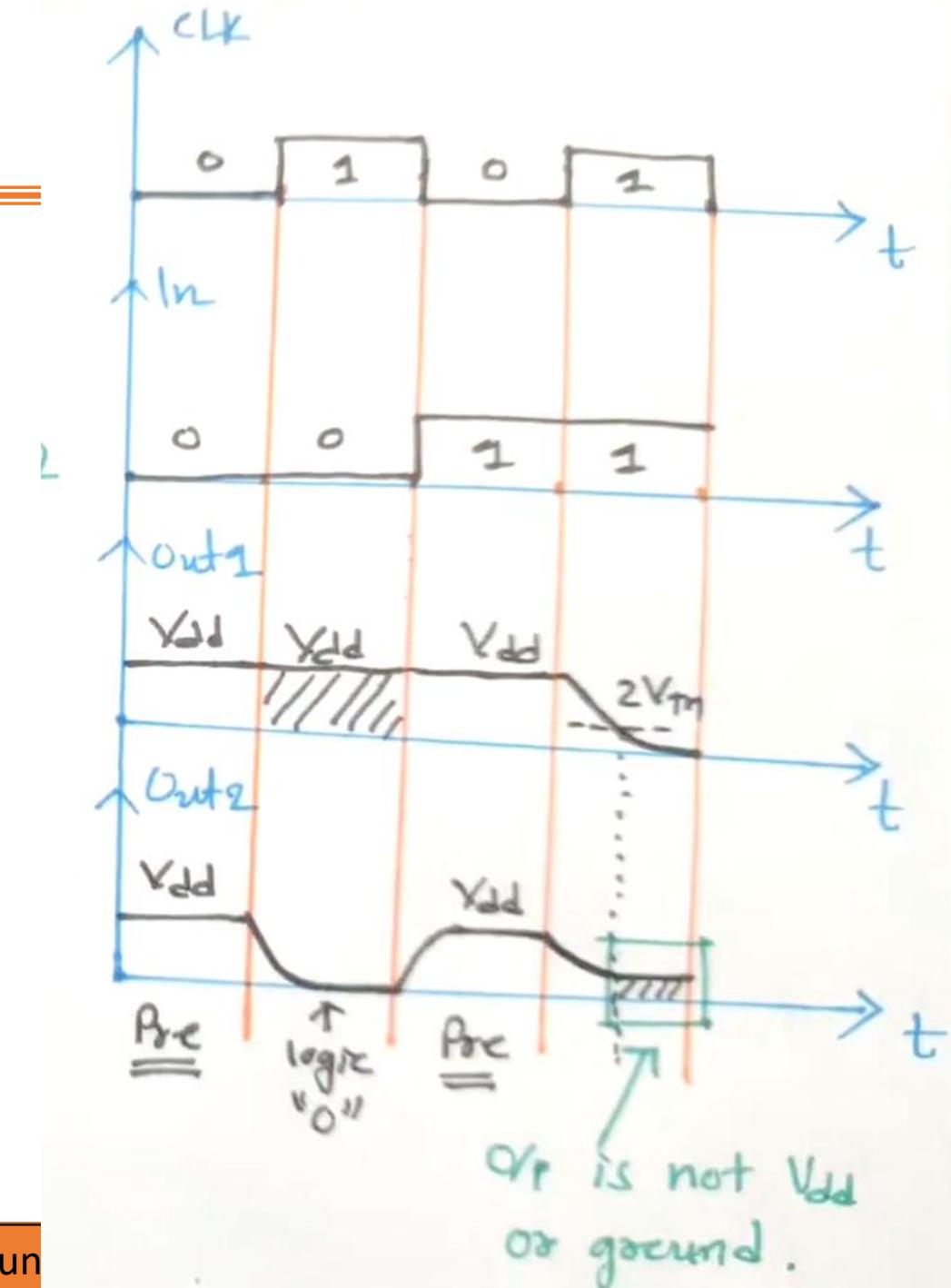
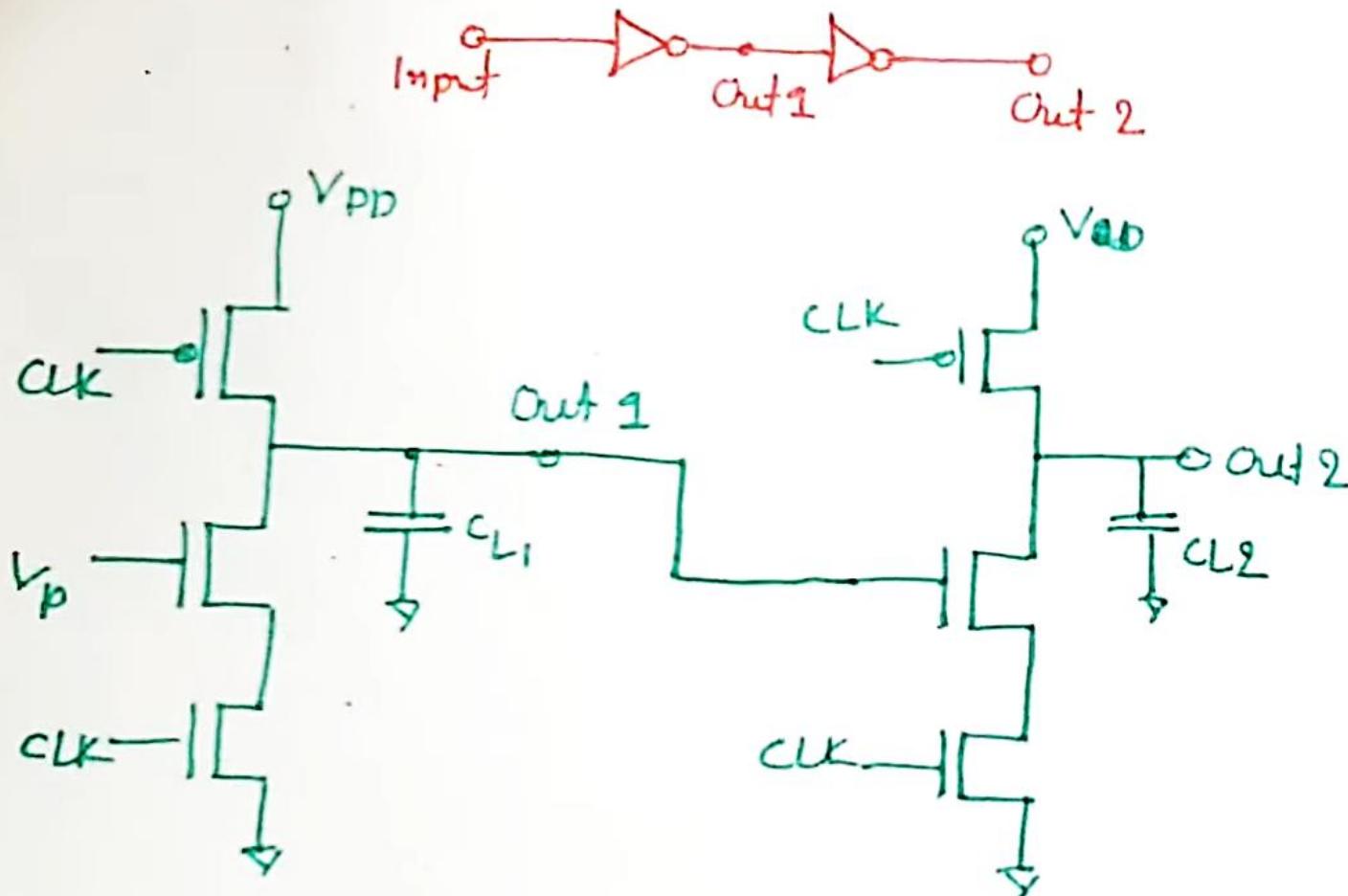
1. $V_{OH} = V_{DD}$ and $V_{OL}=0$. Maximum output swing is achieved.
2. Steady state current is zero.
3. Number of PMOS is only one. Hence its switching speed.
4. Circuit size is small.





Cascading problem in Dynamic CMOS

→ Case





3. CMOS Domino logic

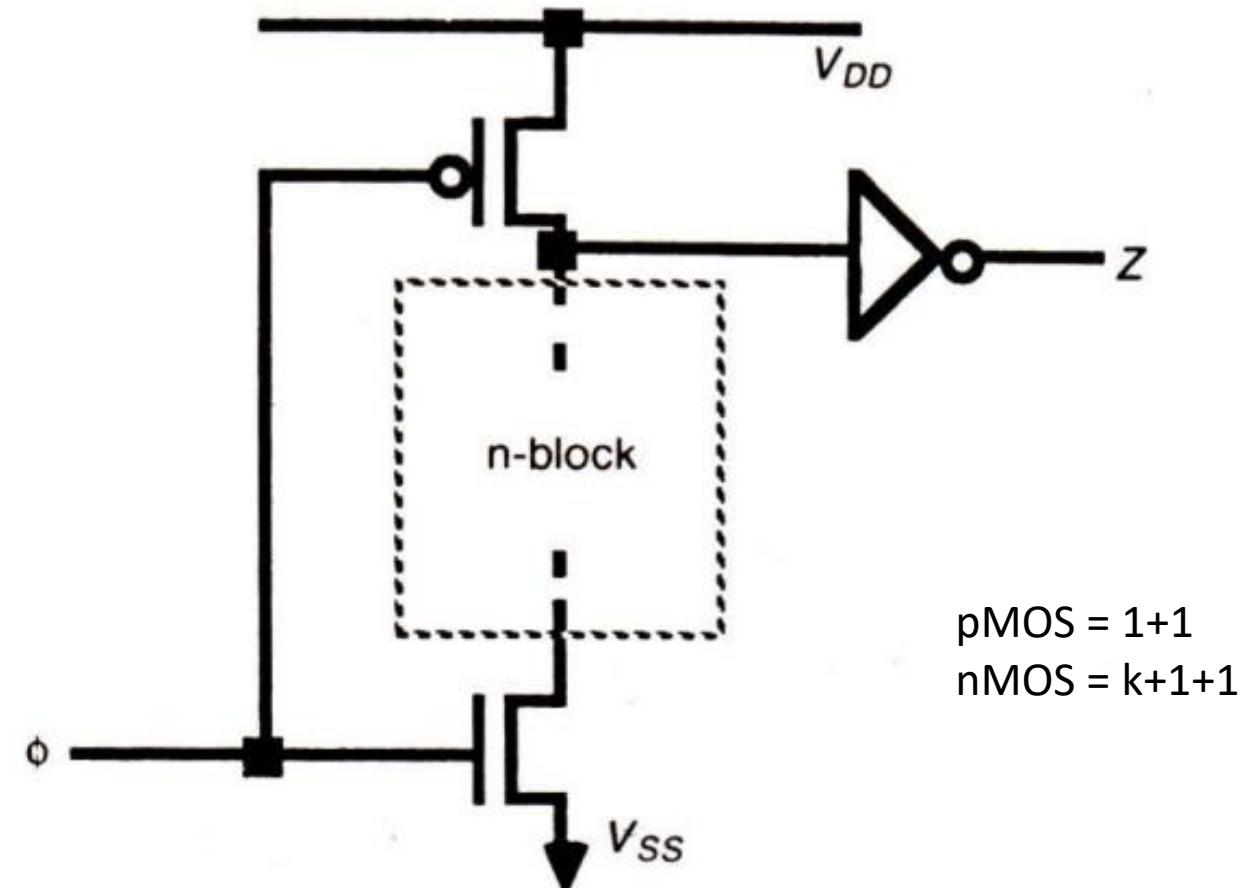
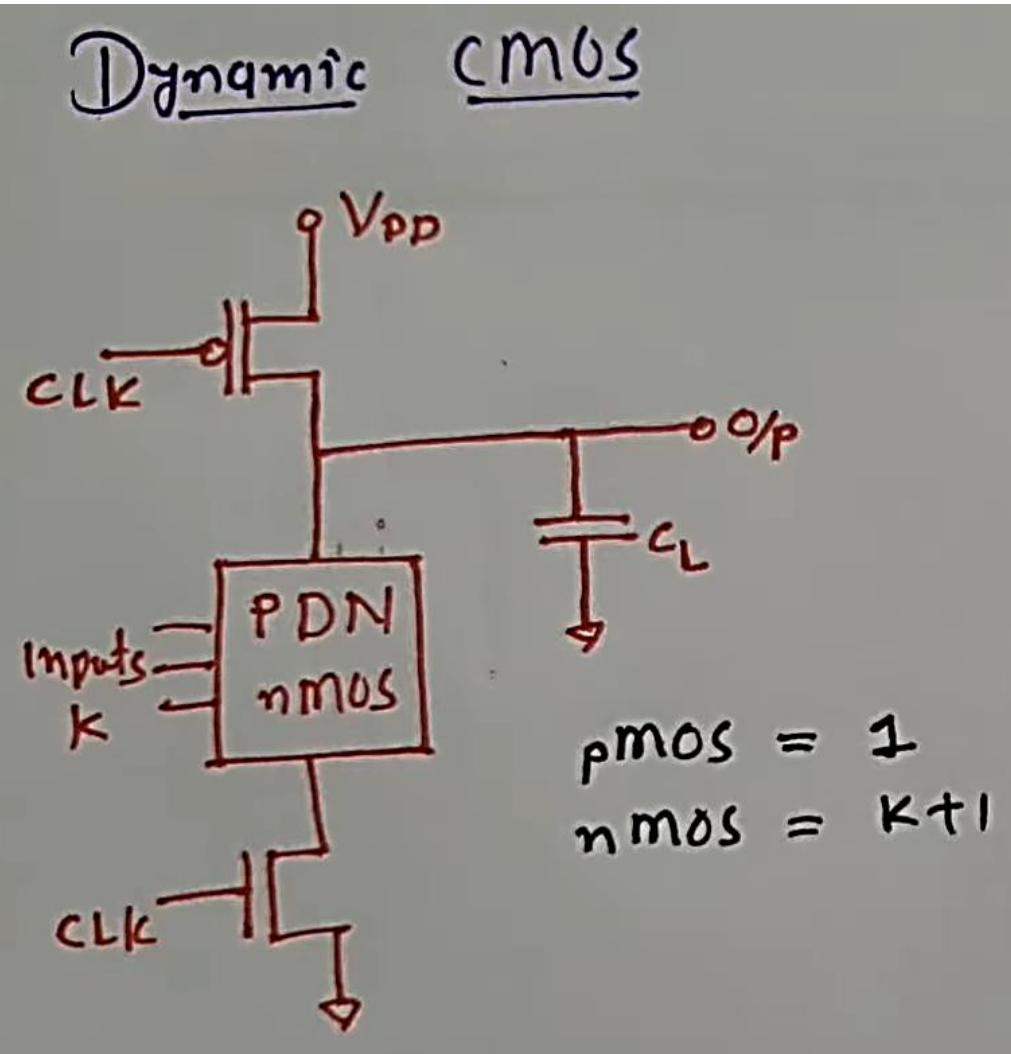
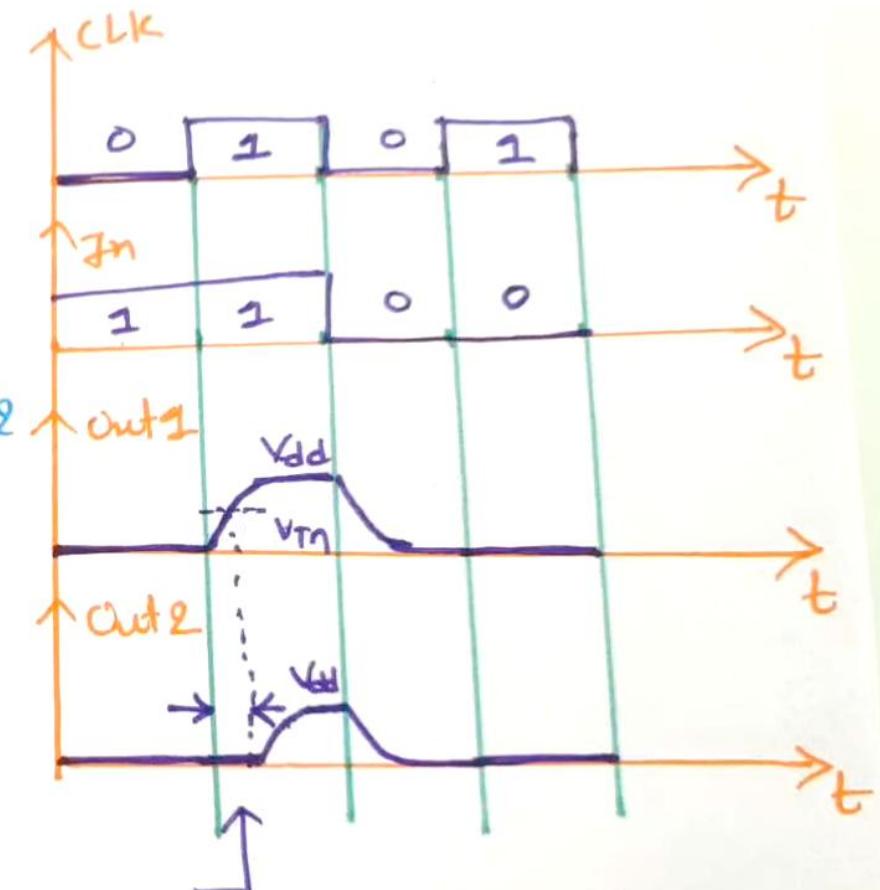
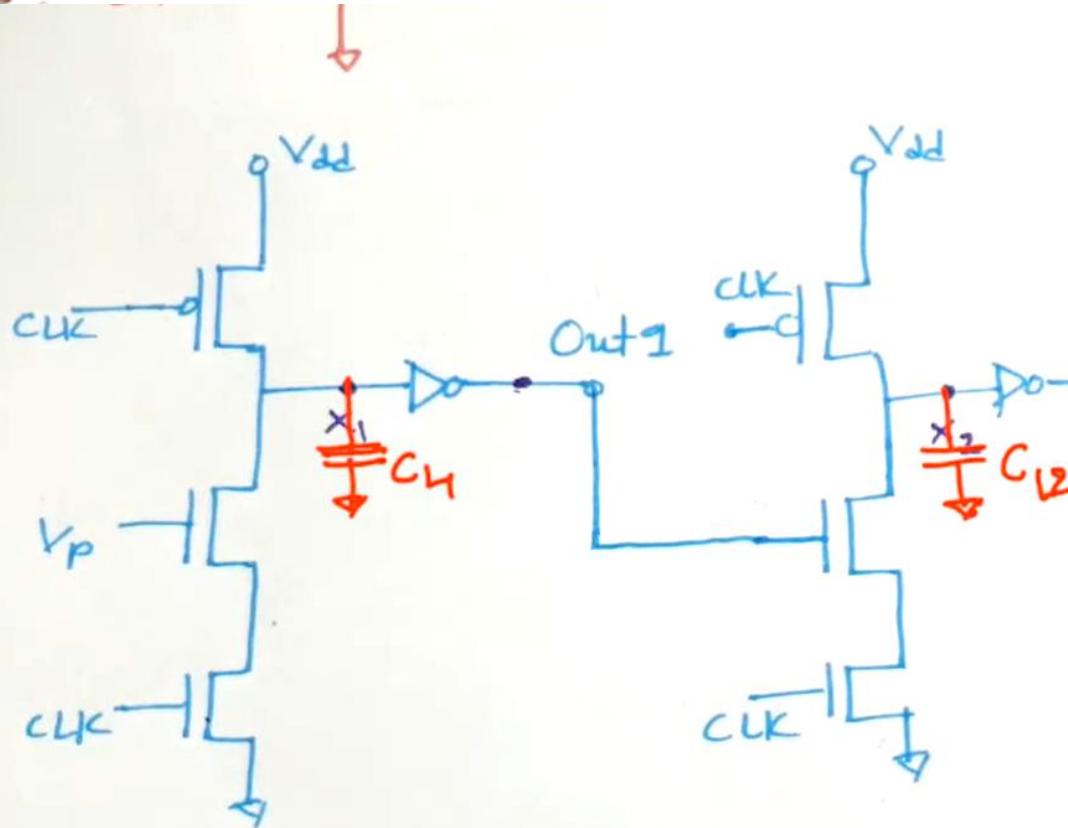


FIGURE 6.13 CMOS domino logic.



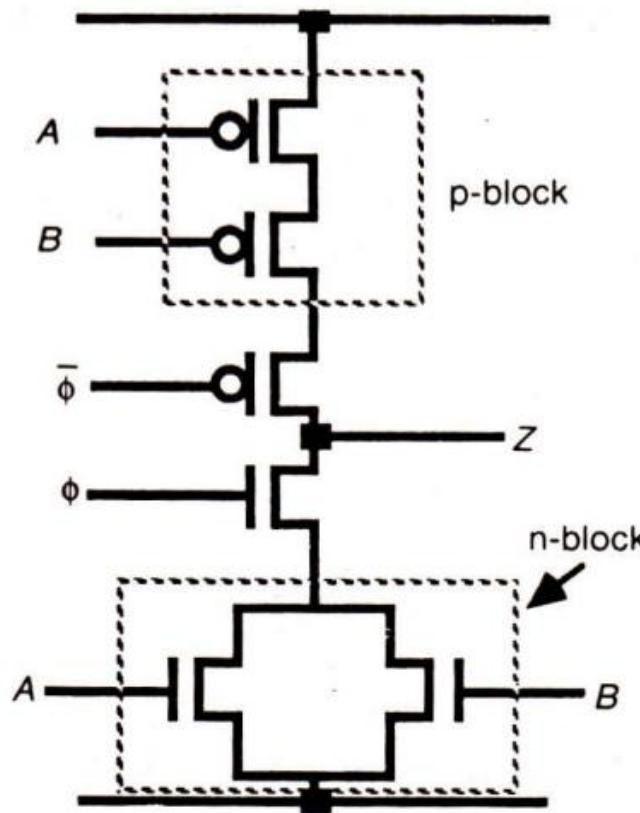
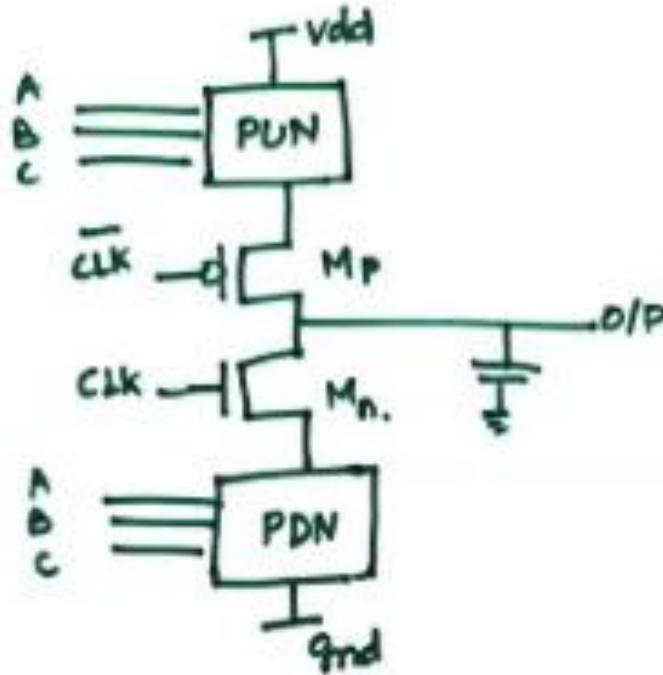
Cascading of CMOS Domino logic



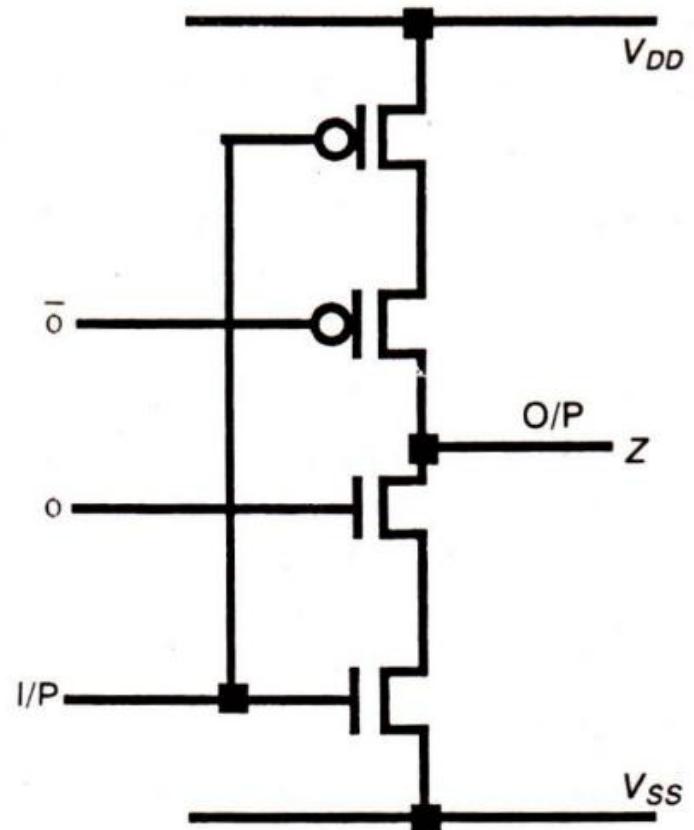
Domino logic is solving issue of Cascading but it falls in a problem RACE.



4. Clocked CMOS (C^2MOS) logic



(a) 2 1/P Nor gate



(b) Inverter

FIGURE 6.12 Clocked CMOS (C^2MOS) logic.



Implement the following using C²MOS logic

1.NAND2

2.Out= $(a(b+c)+xy)'$



Fabrication of MOSFETs

Major steps involved in fabrication are:

1. Oxidation of silicon wafers
2. Diffusion of silicon wafers
3. Photo-lithography
4. Metallization and packaging.

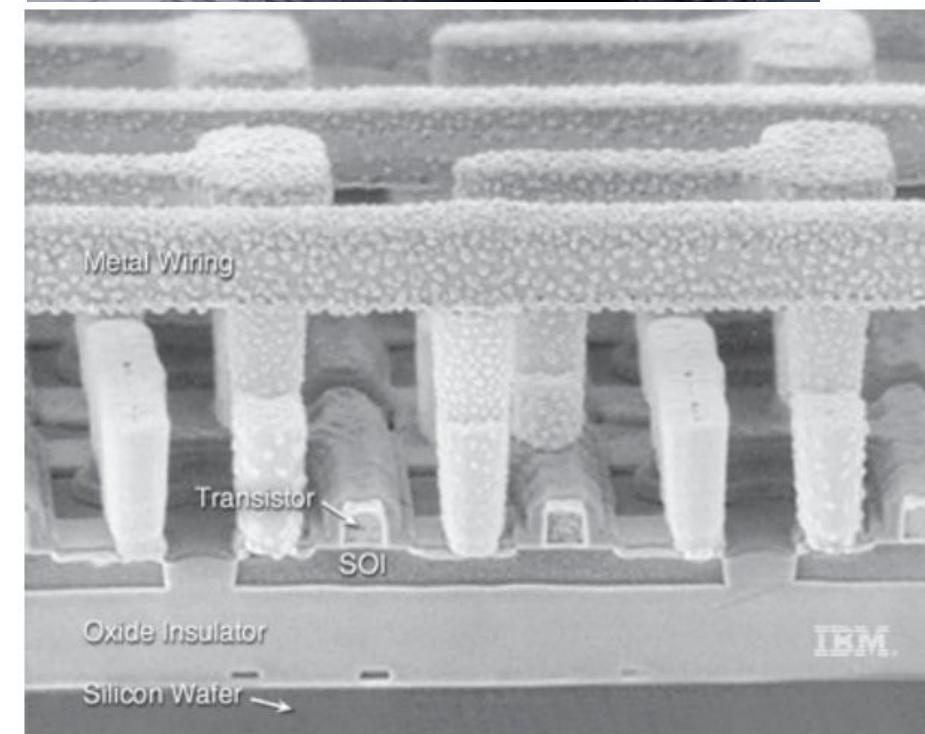
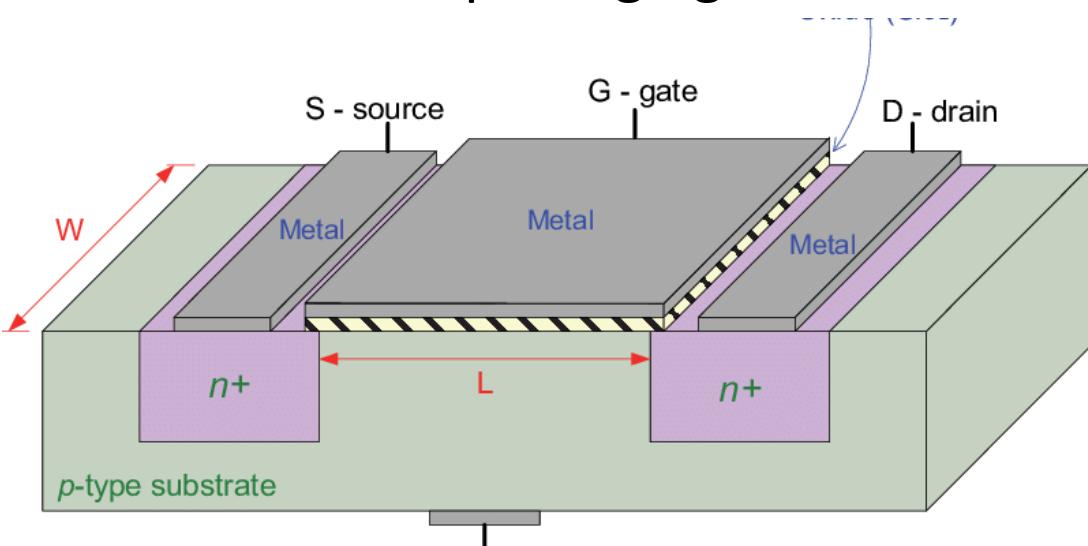


FIGURE 9.62 IBM SOI process electron micrograph
(Courtesy of International Business Machines Corporation.
Unauthorized use not permitted.)



Fabrication of MOSFETs

Mask

- Common material used for masks are Photoresist, Polysilicon, Silicon dioxide, Silicon nitride.
- To create mask:
 - (a) deposit mask material over entire surface
 - (b) cut windows in the mask to create exposed areas
 - (c) deposit dopant
 - (d) remove un-required mask material
- Masks plays important role in process called selective diffusions.
- The selective diffusion involves
 1. Patterning windows in a mask material on the surface of the wafer.
 2. Subjecting the exposed areas to a dopant source.
 3. Removing any un-required mask material.



Photolithography

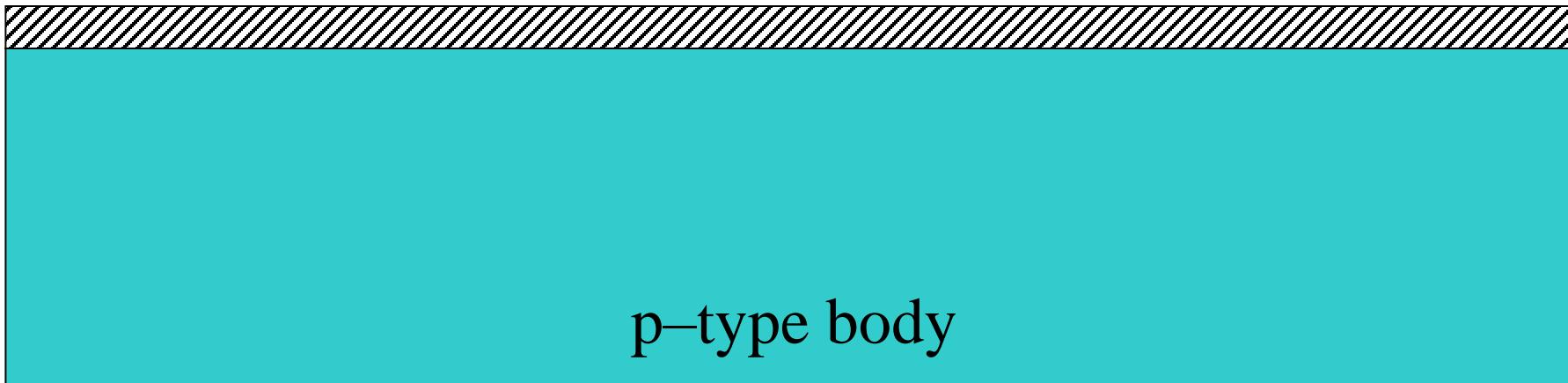
- The Process of using an optical image and a photosensitive film to produce a pattern on a substrate is **photolithography**
- Photolithography depends on a photosensitive film called a photo-resist.

Types of resist

- **Positive resist**, a resist that become soluble when exposed and forms a positive image of the plate.
- **Negative resist**, a resist that lose solubility when illuminated forms a negative image of the plate.



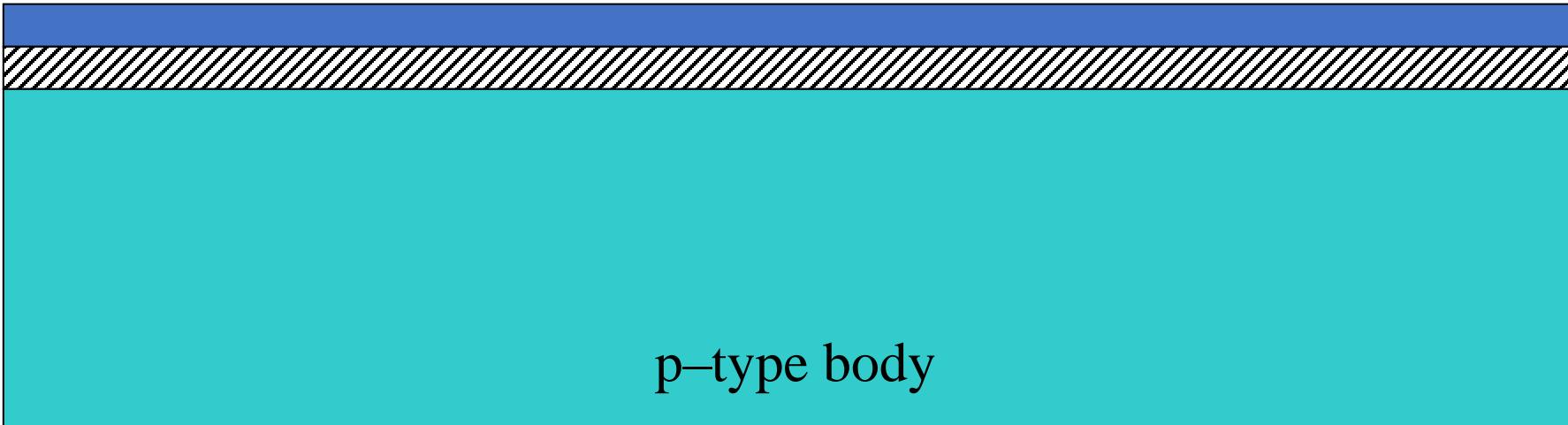
Photolithography



Substrate



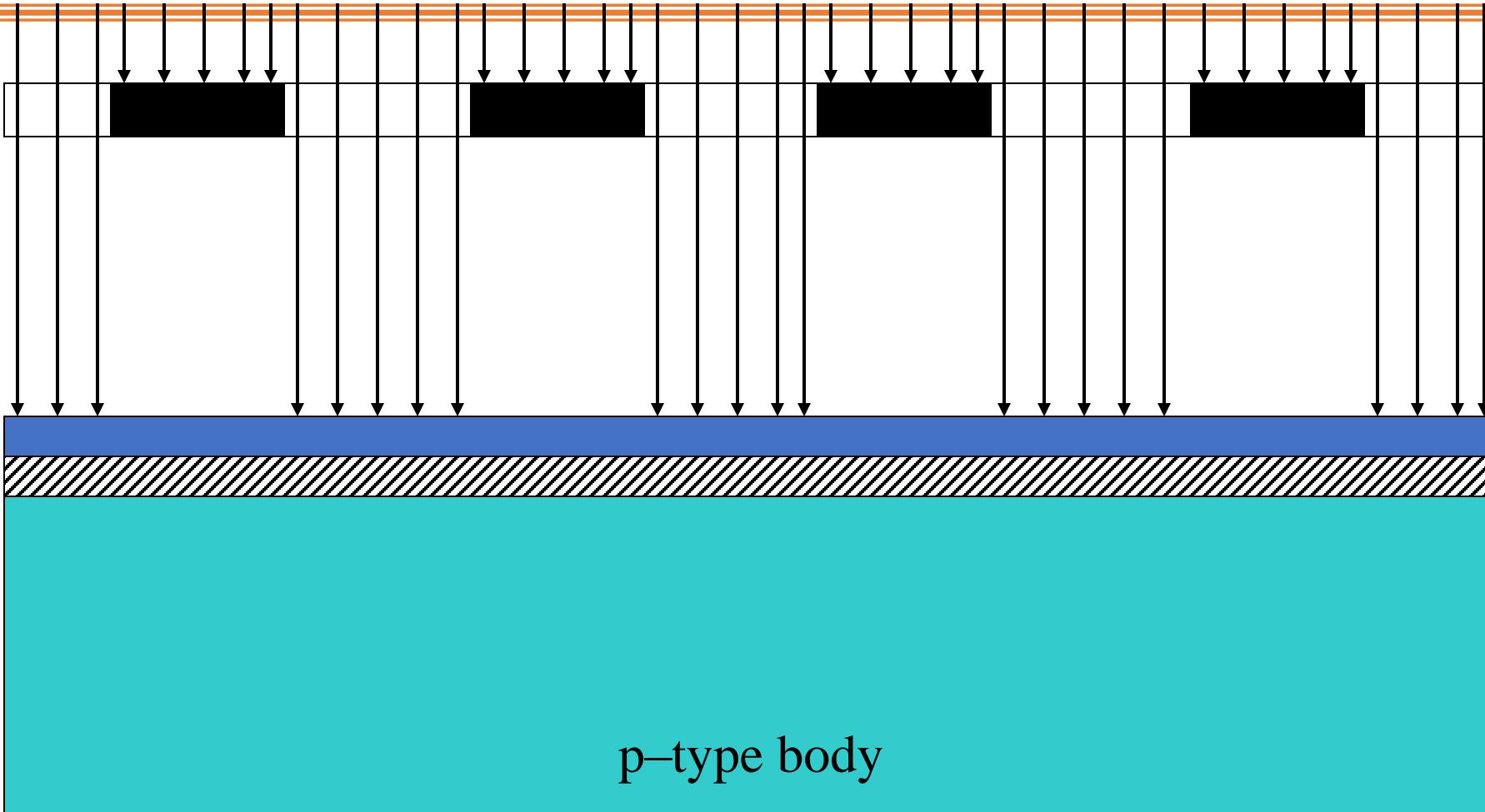
Photolithography



Resist application



Photolithography



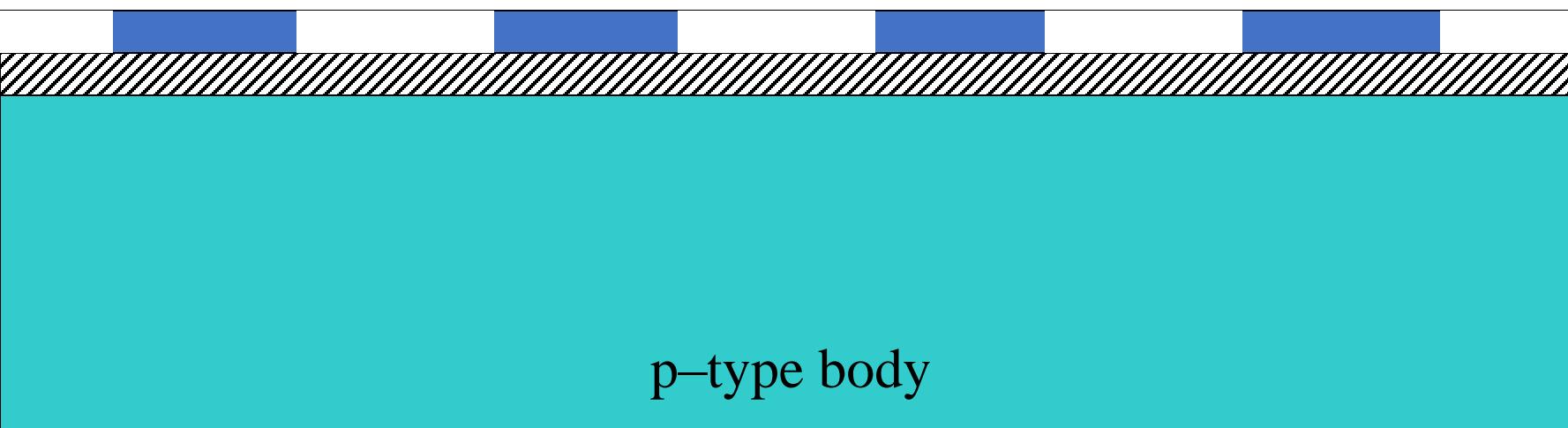
UV Exposure



Photolithography



Etching



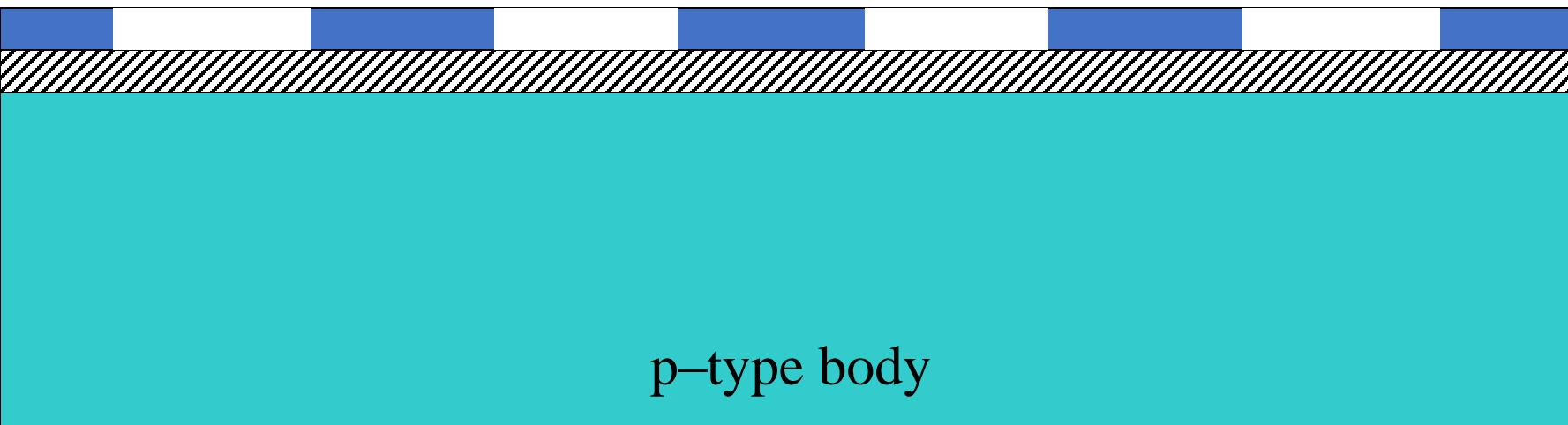
Positive Resist



Photolithography



Etching

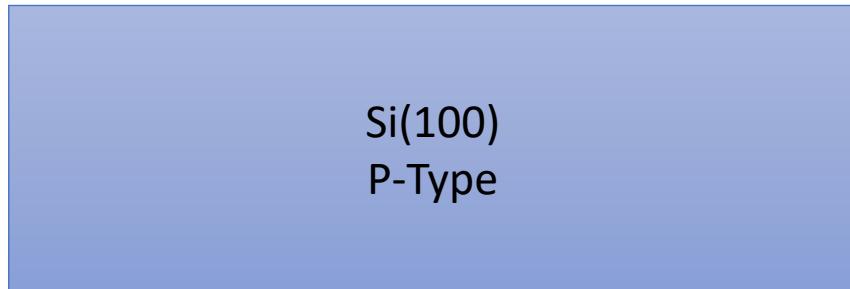


Negative Resist



nMOS fabrication steps

1. Selection of Substrate

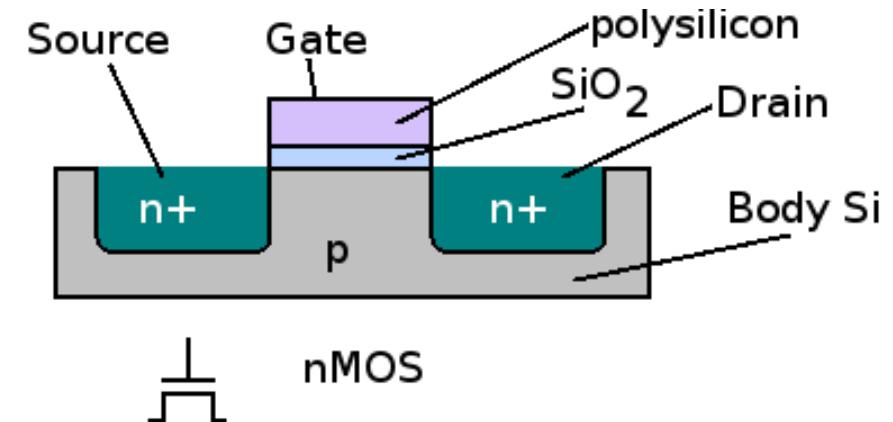
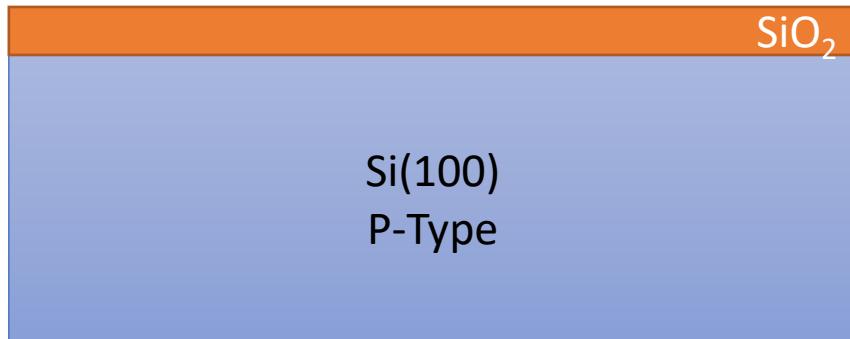


2. Cleaning

Dry oxidation: $\text{Si} + \text{O}_2 \rightarrow \text{SiO}_2$ (good quality)

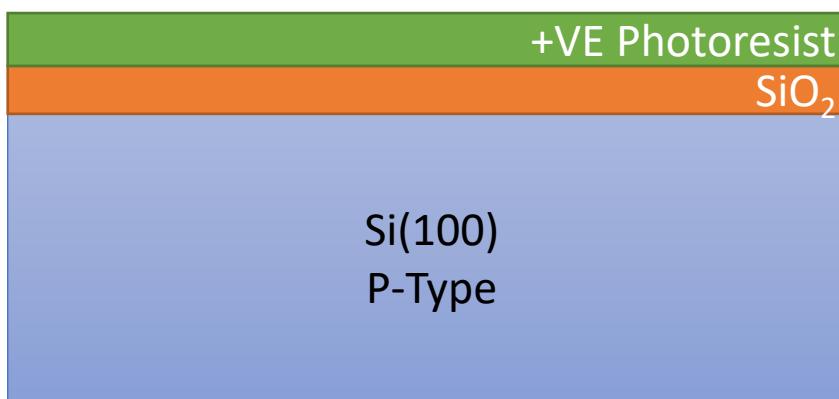
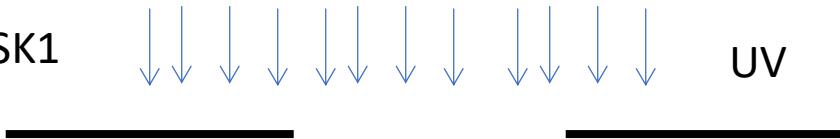
Wet Oxidation: $\text{Si} + 2\text{H}_2\text{O} \rightarrow \text{SiO}_2 + 2\text{H}_2$ (poor quality)

3. Oxidation

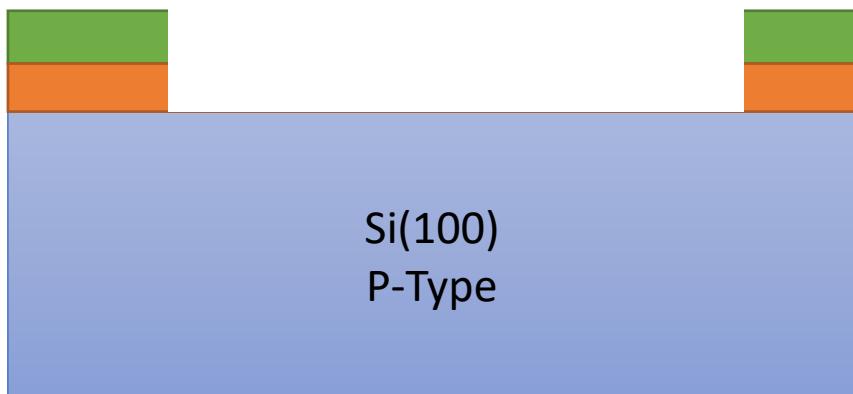




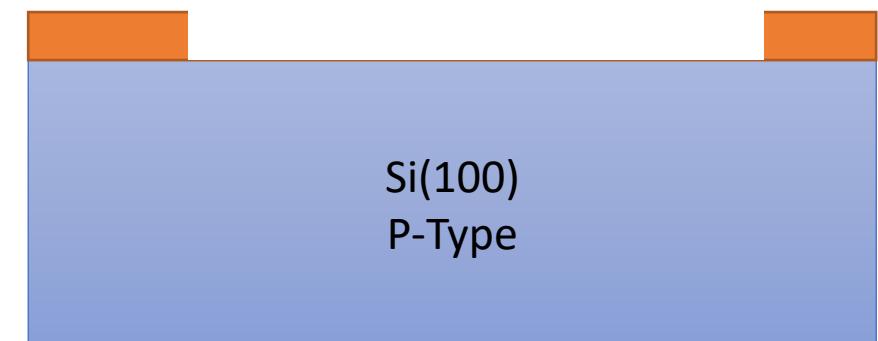
4. Lithography with MASK1



Photoresist development and Oxide Etching

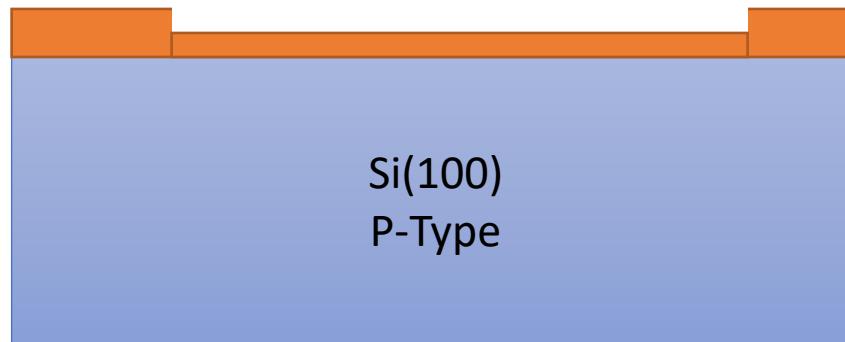


Photoresist Etching





Gate Oxidation

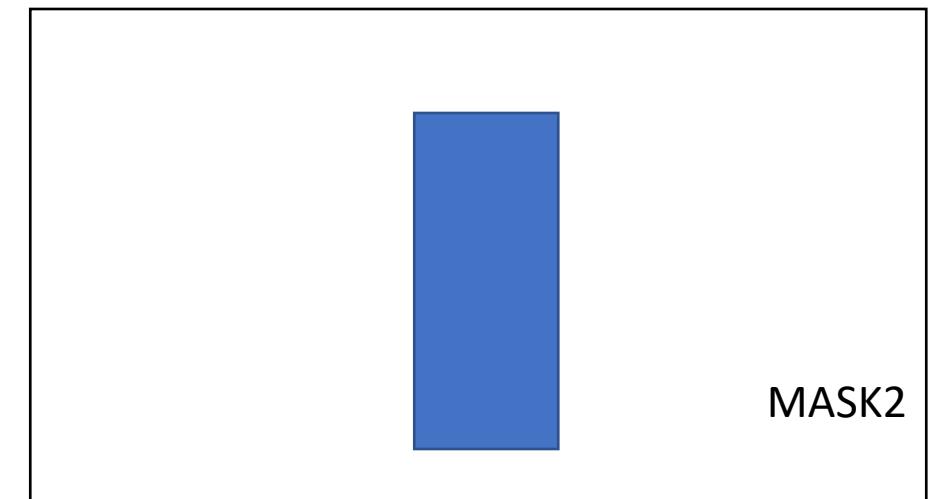
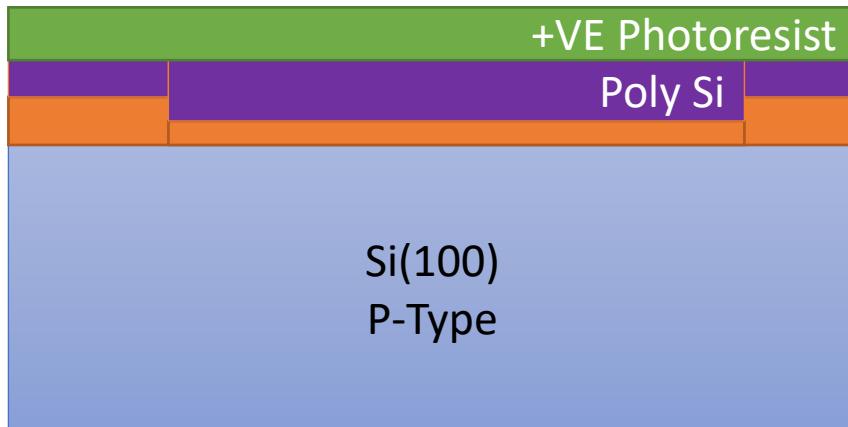
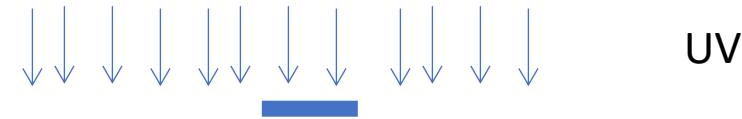


Poly Silicon Deposition



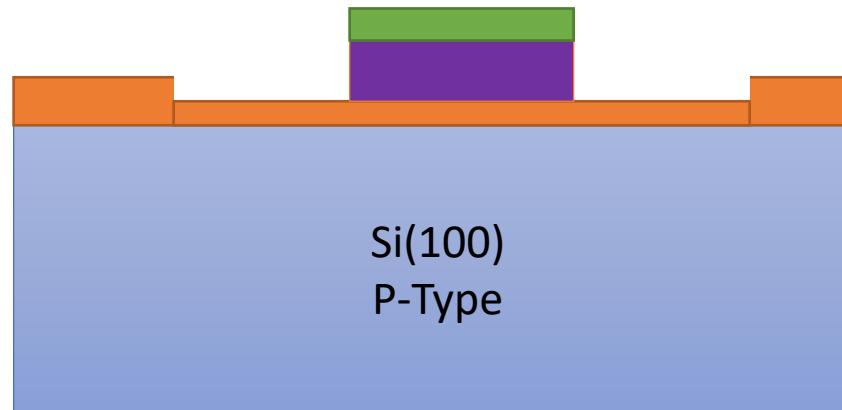


Lithography for Gate Electrode

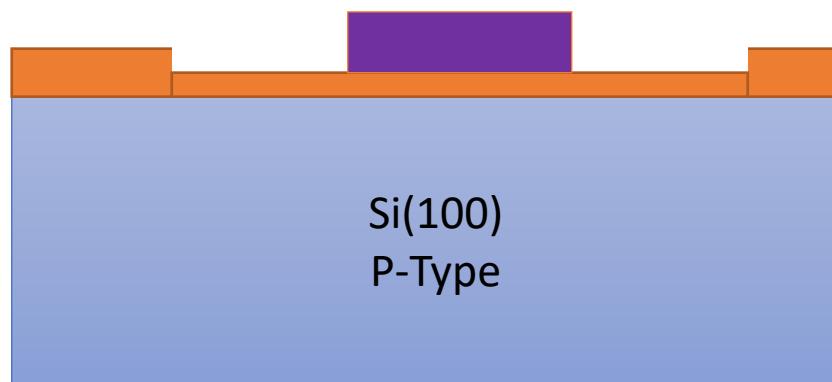




Poly Patterning

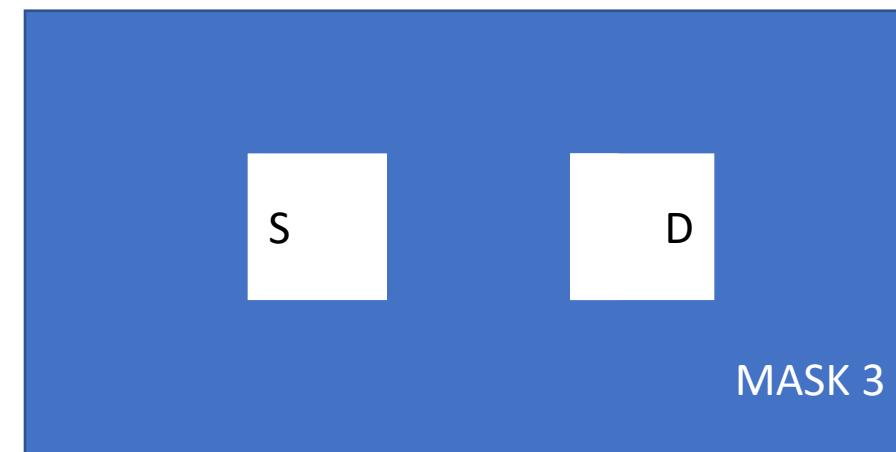
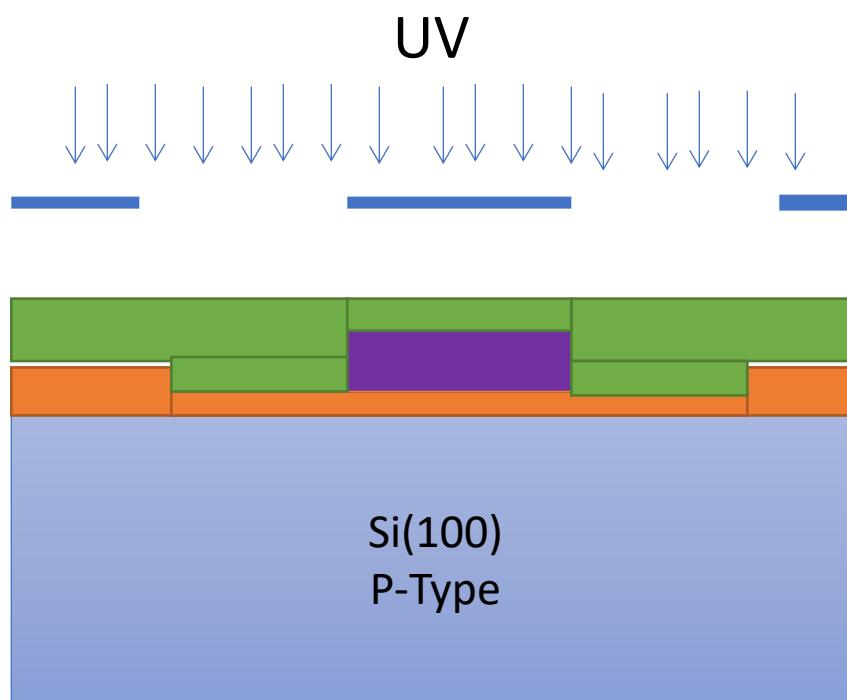


Photoresist Cleaning



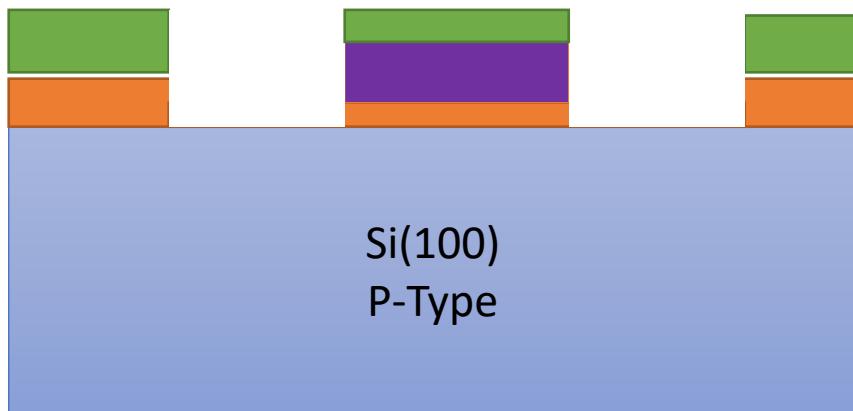


Lithography for Source and Drain region

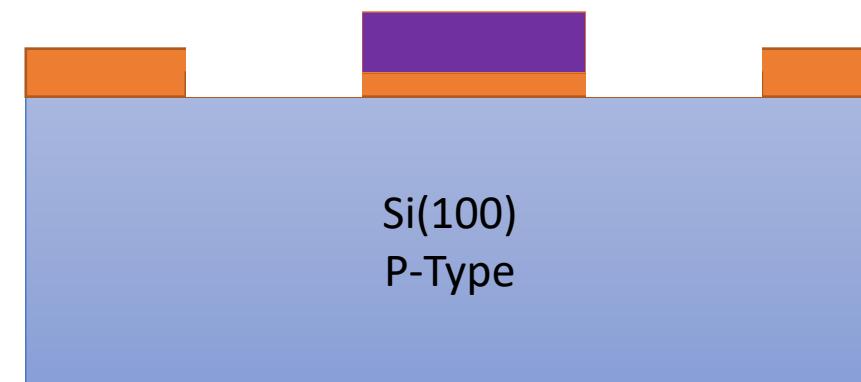




Oxide etching (HF Cleaning)

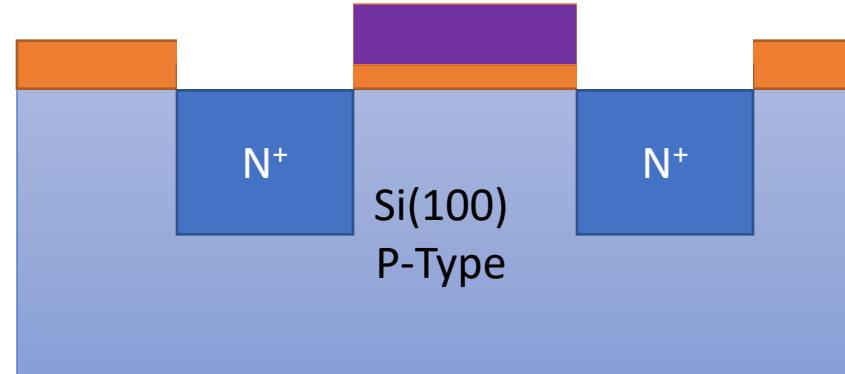


Photoresist cleaning

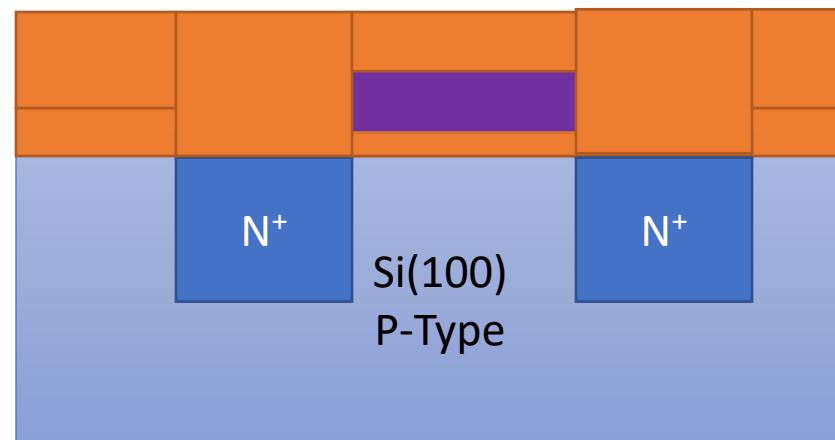




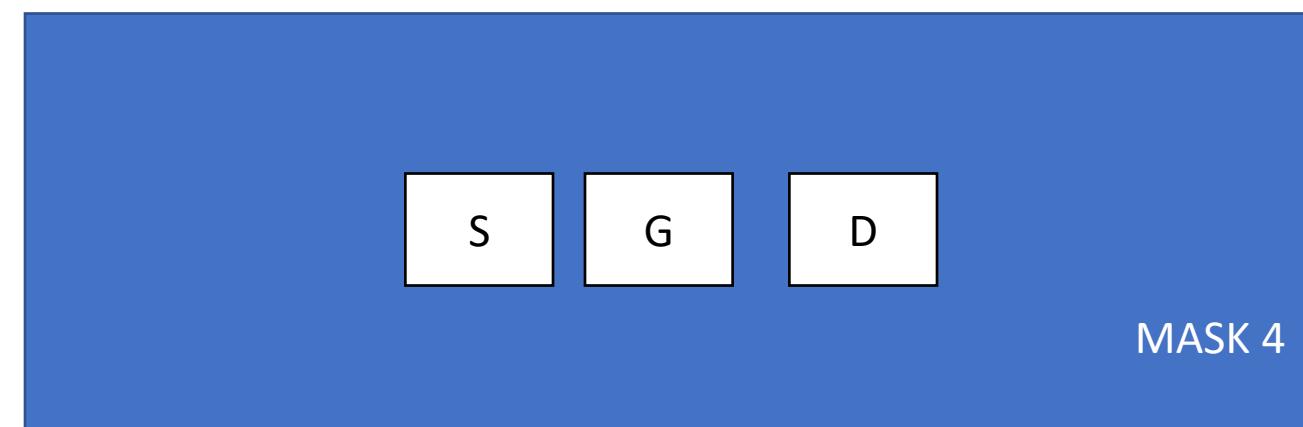
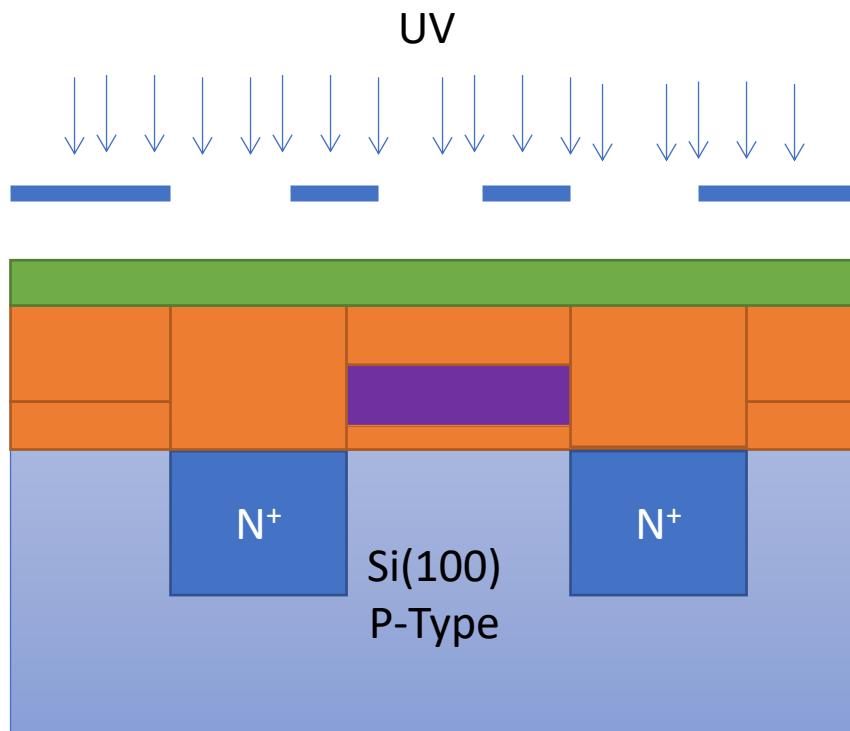
Ion Implantation



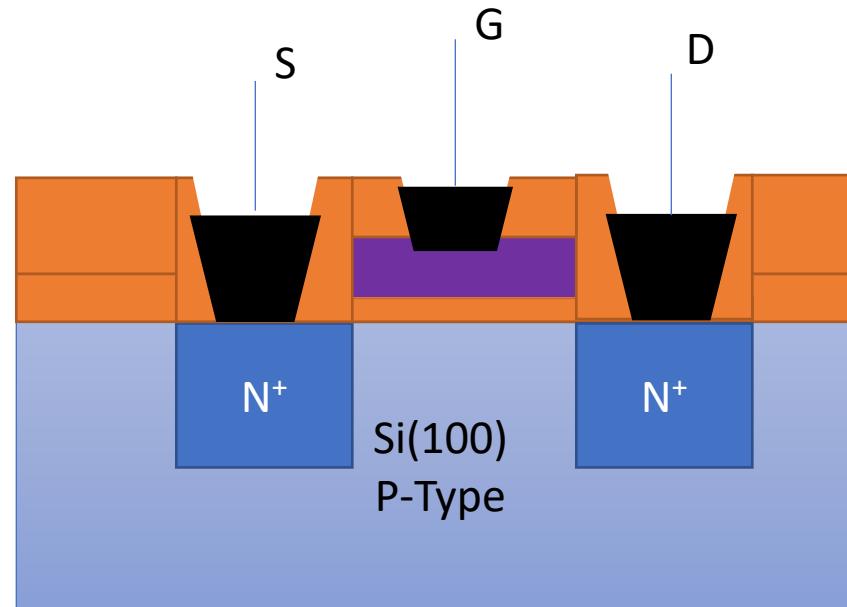
Thick Oxide Deposition



Lithography and Contact Opening



Metallization and Packaging



Body terminal is not shown.....



CMOS fabrication steps

Technologies used for CMOS fabrications include

- **N-well process**
- **P-well process**
- **Twin-tub process**
- **Silicon on insulator (SOI).**



P-wells and n-wells

- In order to have both types of transistors on the same substrate, the substrate is divided into “well” regions (*Shaded region in the standard cells*)
- Two types of wells are available - n- well and p- well
- In a p- substrate, an n- well is used to create a local region of n type substrate, wherein the designer can create p- transistors
- In a n- substrate, a p- well creates a local p- type substrate region, to accommodate the n- transistors.
- Hence, every p- device is surrounded by an n- well, that must be connected to V_{DD} via a V_{DD} substrate contact.
- Similarly, n- devices are surrounded by p- well connected to GND using a GND substrate contact.



Cross-section of CMOS inverter

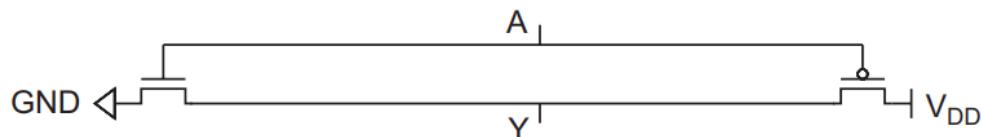
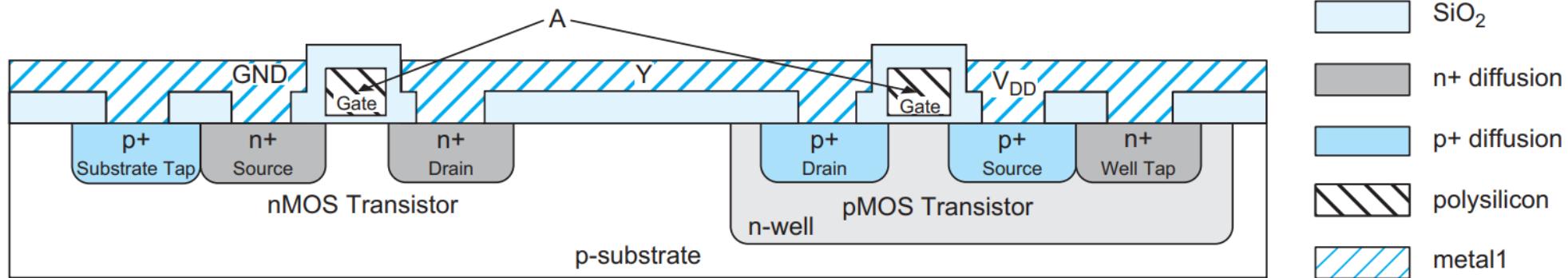
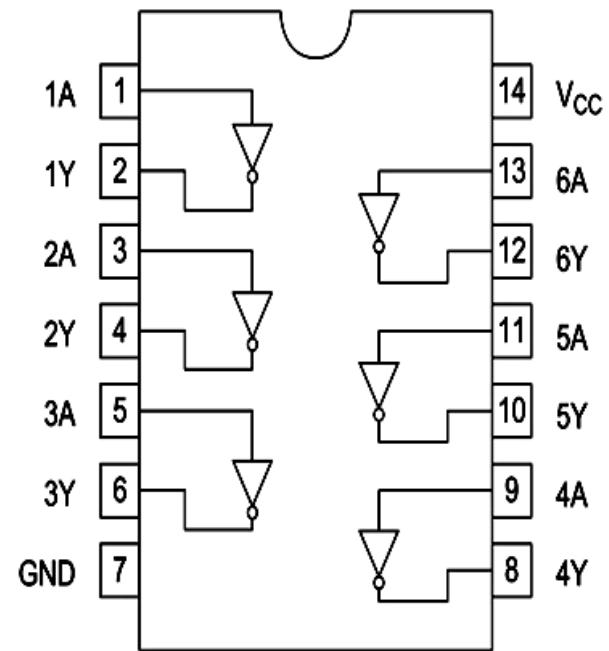
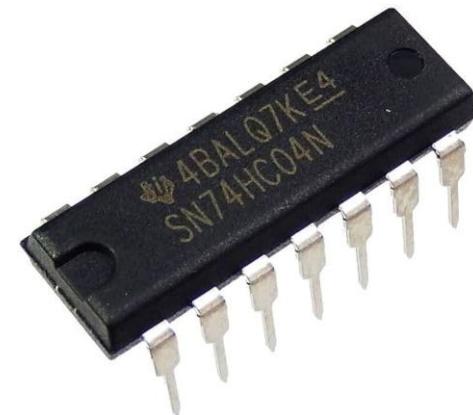
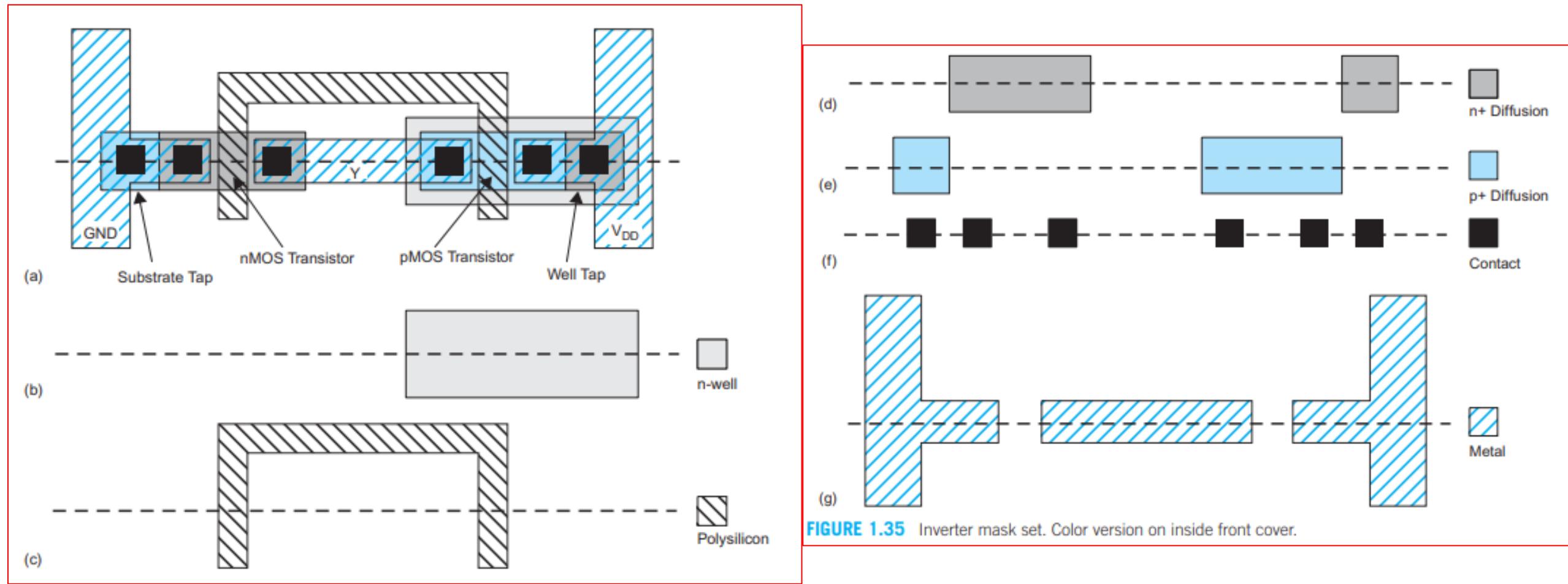


FIGURE 1.34 Inverter cross-section with well and substrate contacts. Color version on inside front cover.



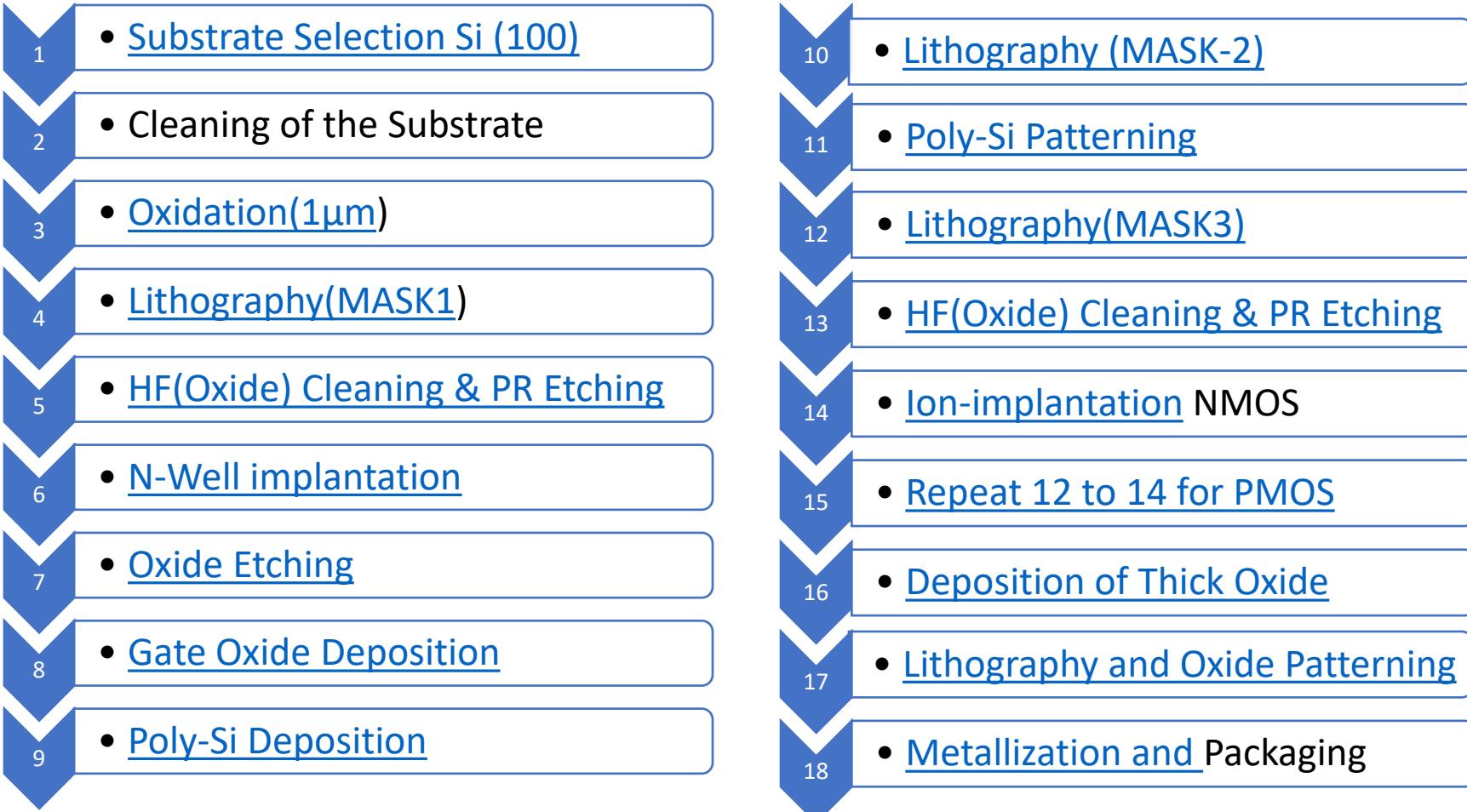


MASKs for CMOS Inverter





1. N-well CMOS inverter





1. Selection of Substrate

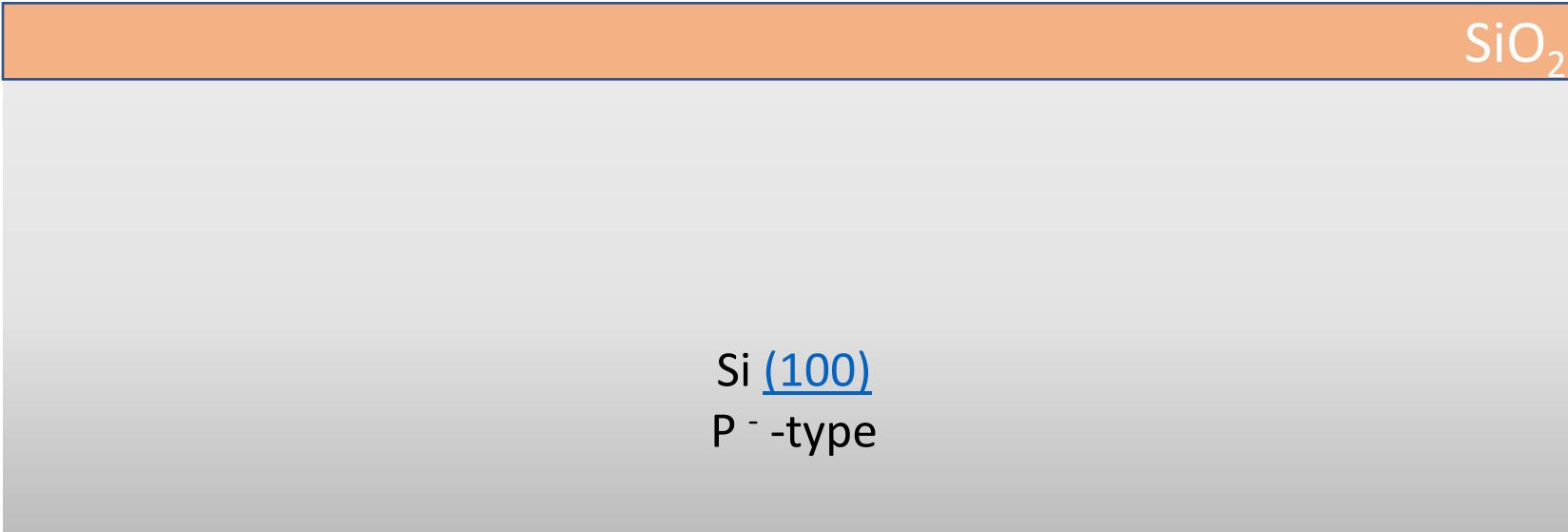
Si [\(100\)](#)
P -type

The n-well CMOS process starts with a moderately doped (with impurity concentration typically less than 10^{15} cm^{-3}) P-type silicon substrate.

2. Cleaning of Substrate

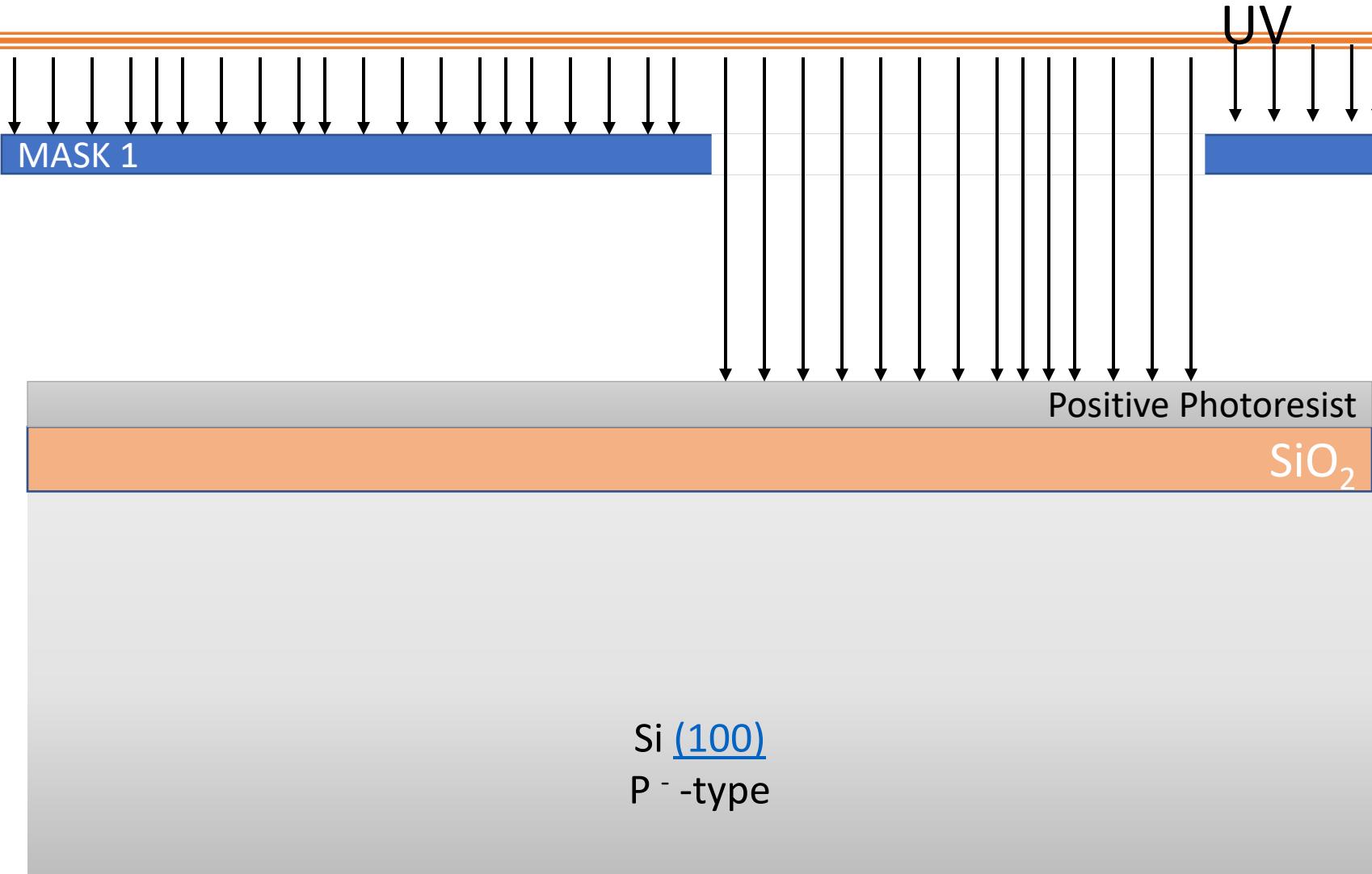


3. Oxidation($1\mu\text{m}$)



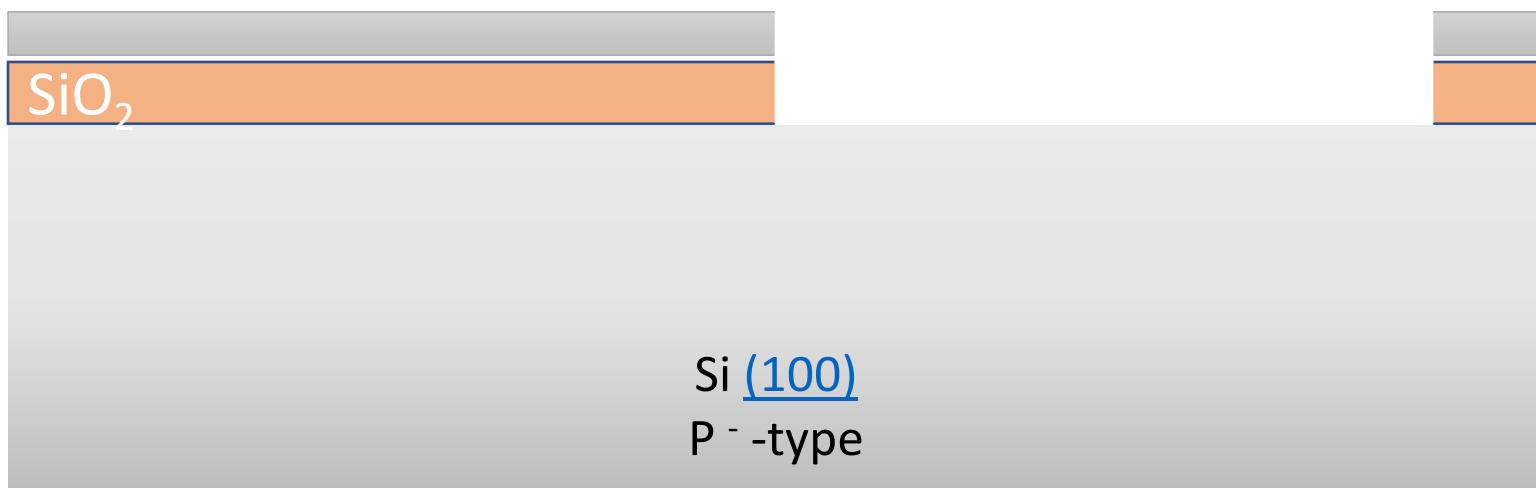
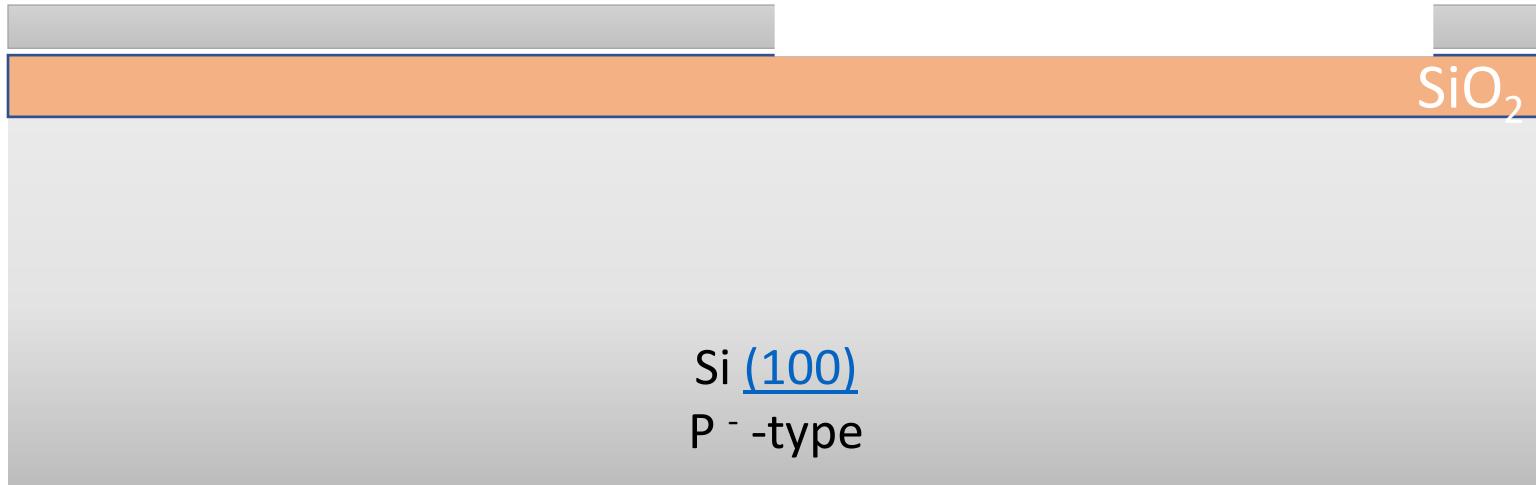


4. Lithography(MASK1)



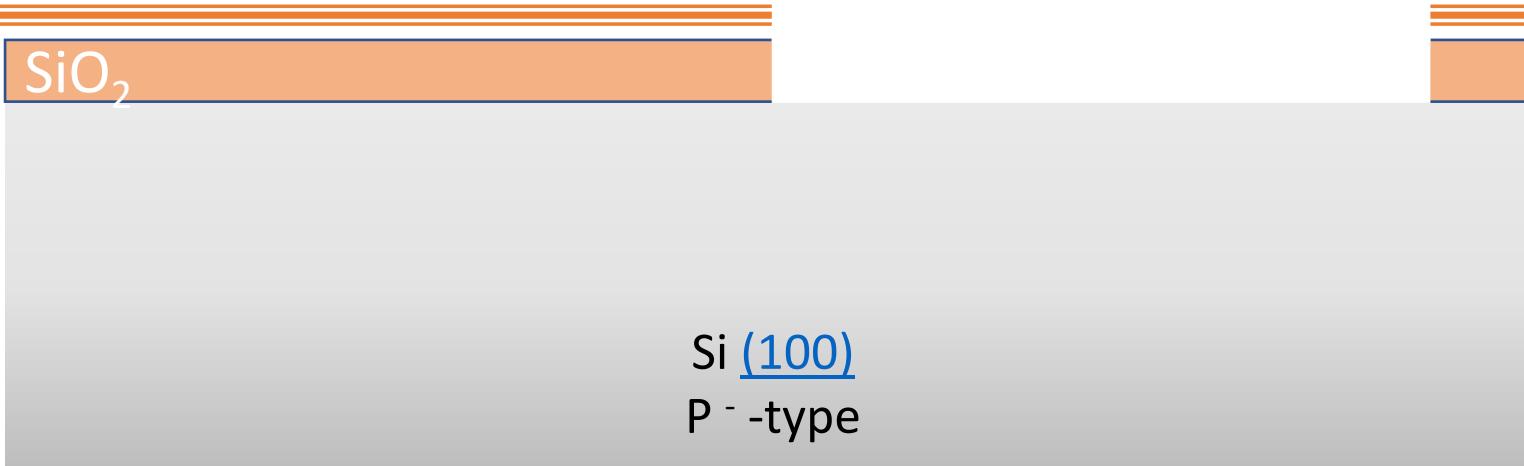


5. HF(Oxide) Cleaning & PR Etching

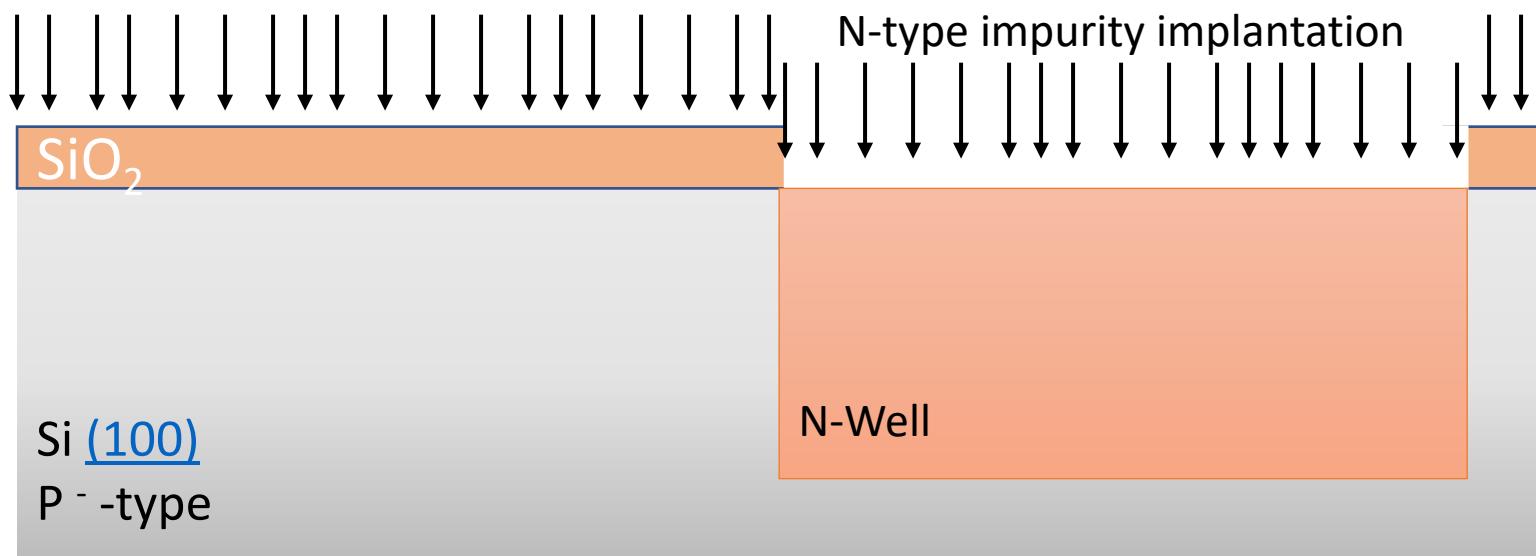




HF(Oxide) Cleaning & PR Etching



6. N-Well Implantation





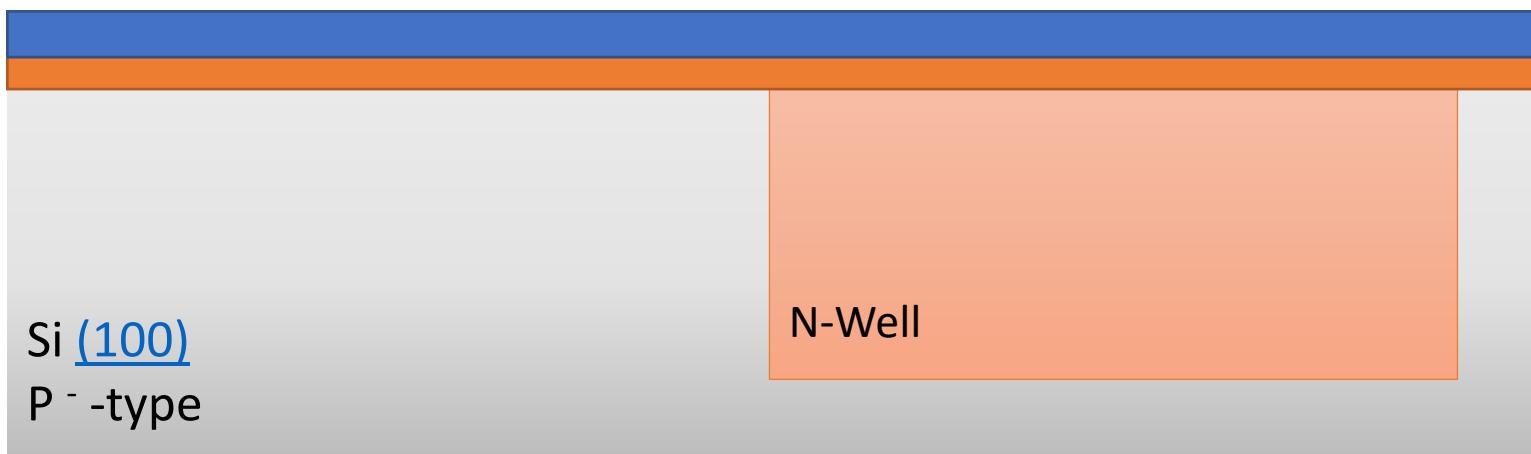
7. Oxide etching



8. Gate Oxide deposition

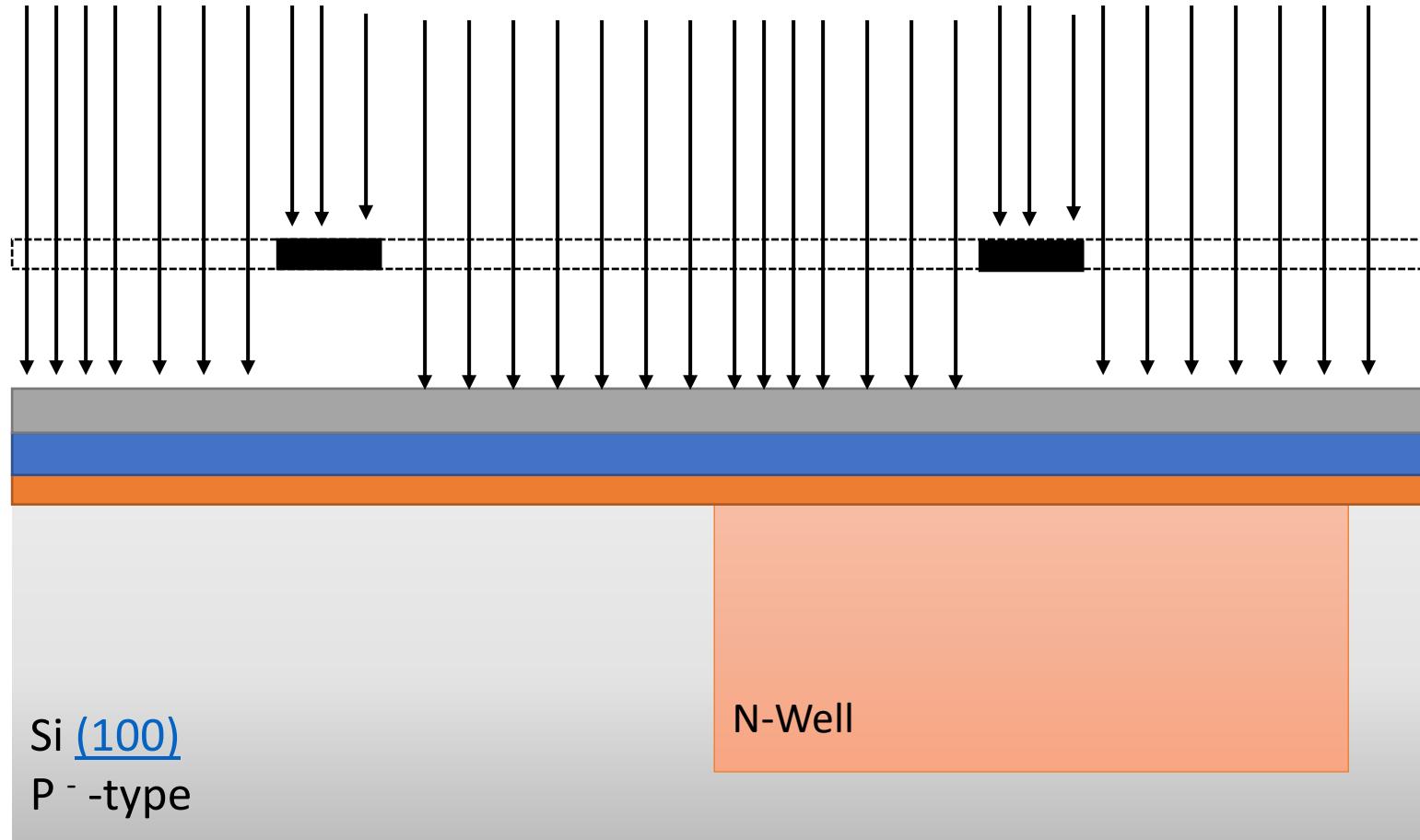


9. Poly-Silicon Deposition





10. Lithography(MASK2)



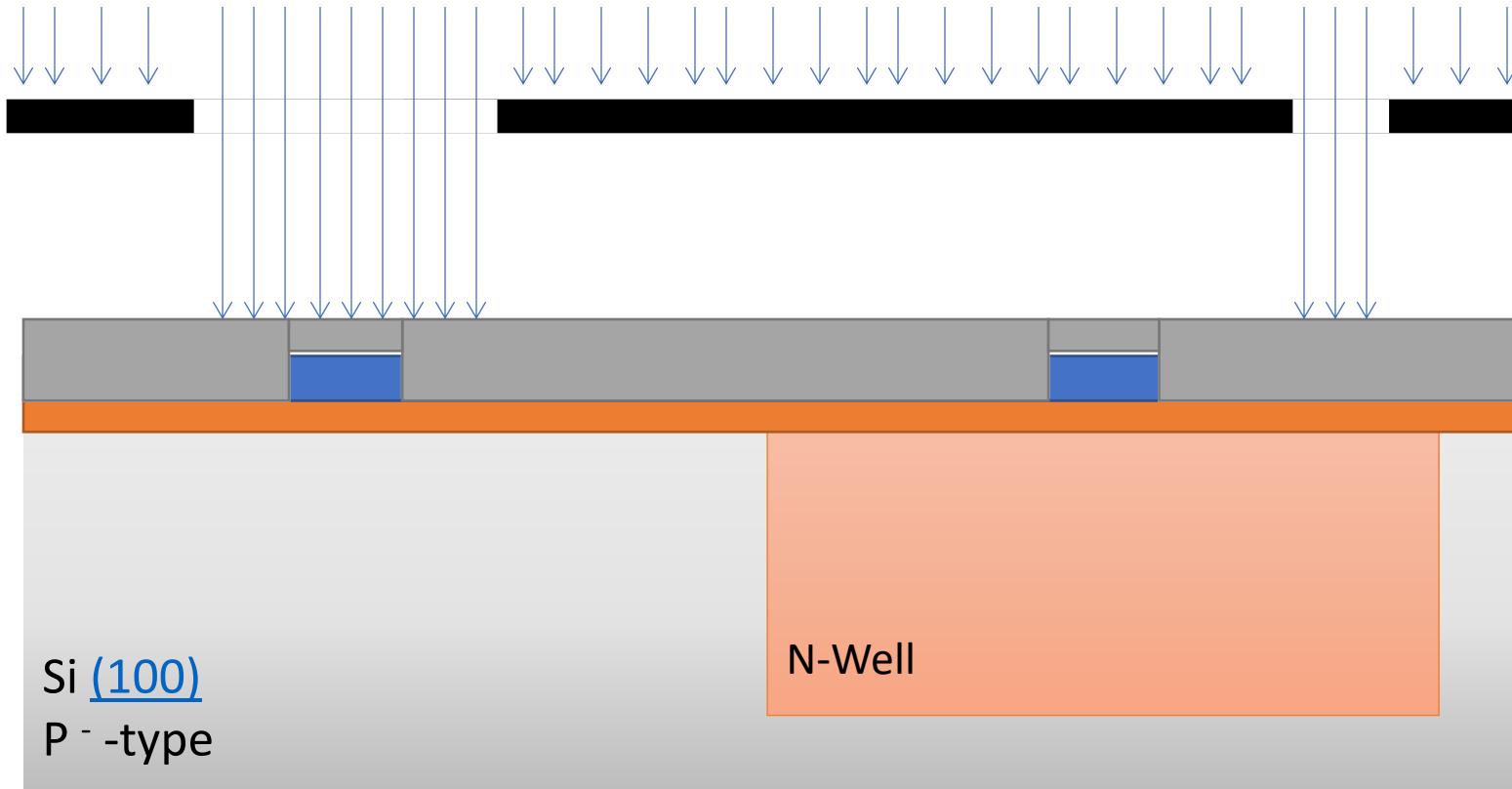


11. Poly-Si patterning



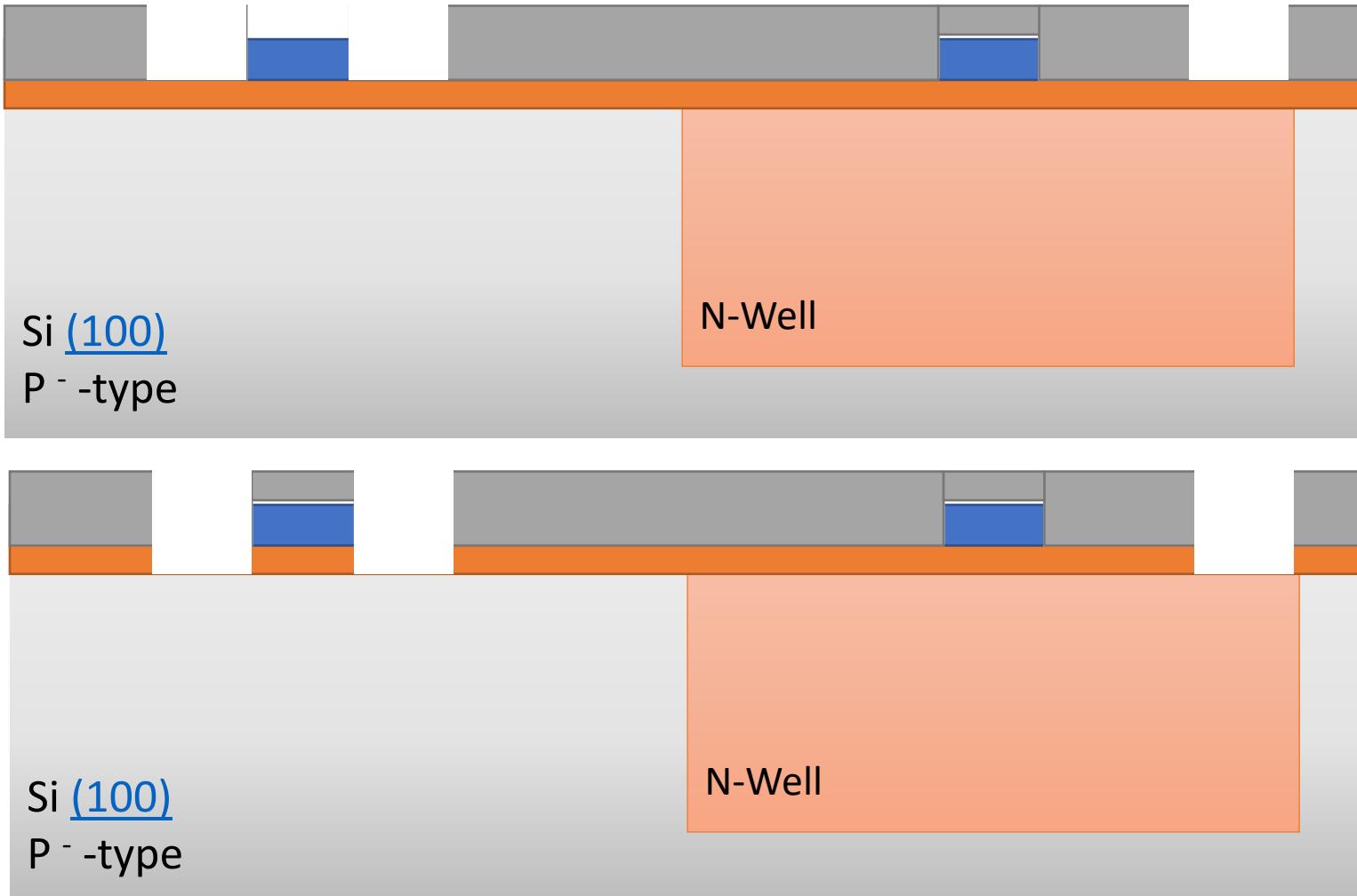


12. Lithography(MASK3) for n+

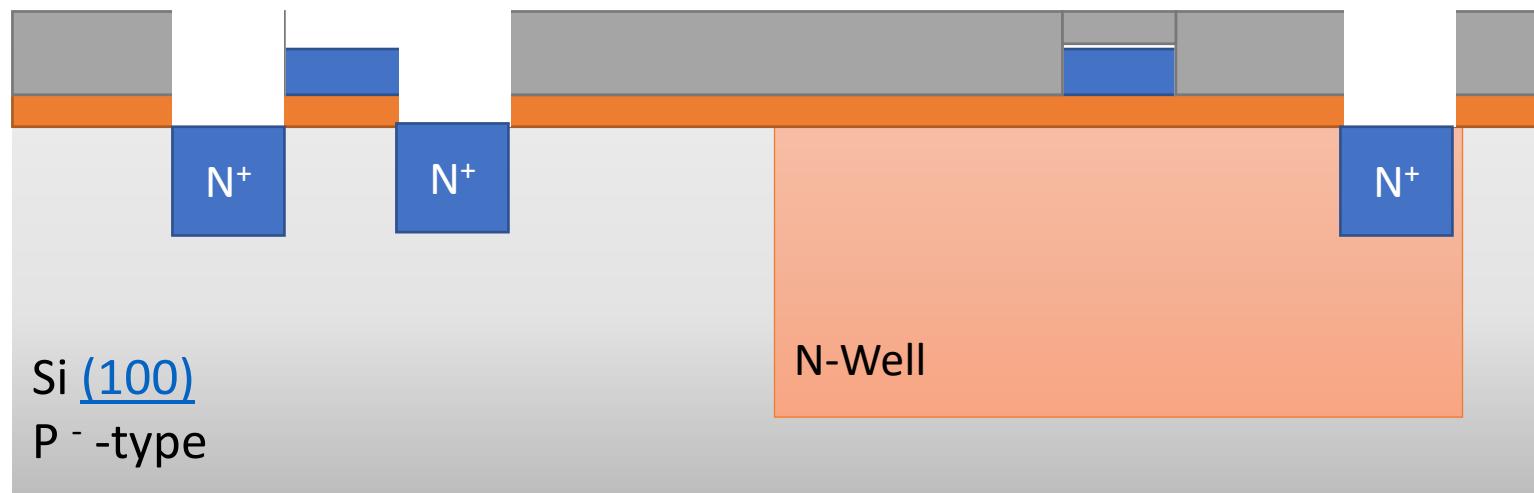




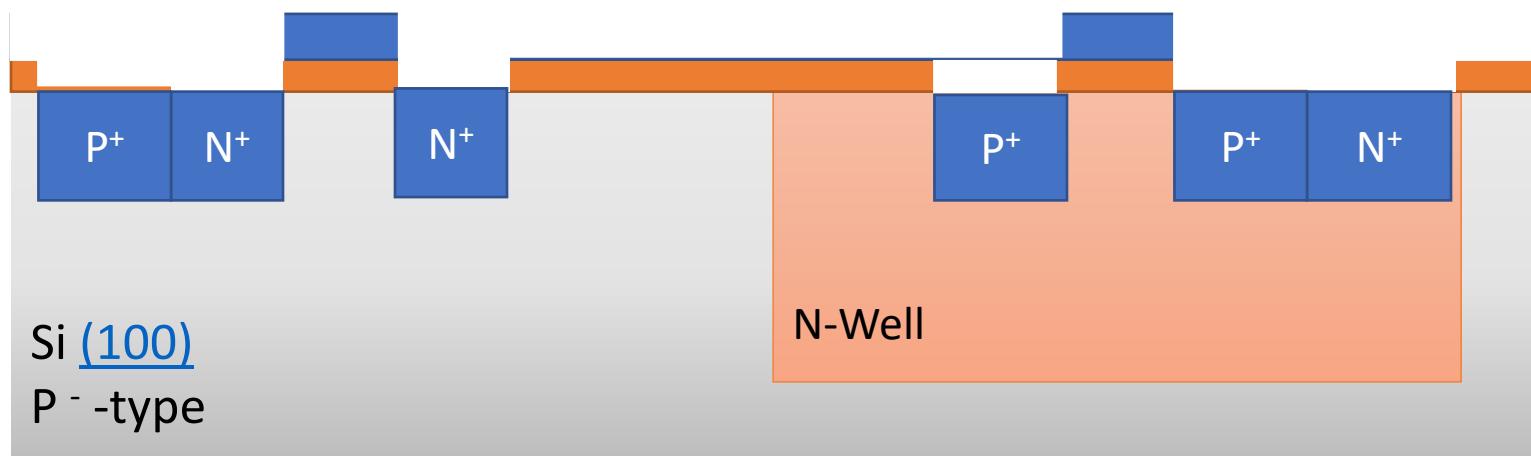
13. HF(Oxide) Cleaning & PR Etching



14. Ion Implantation



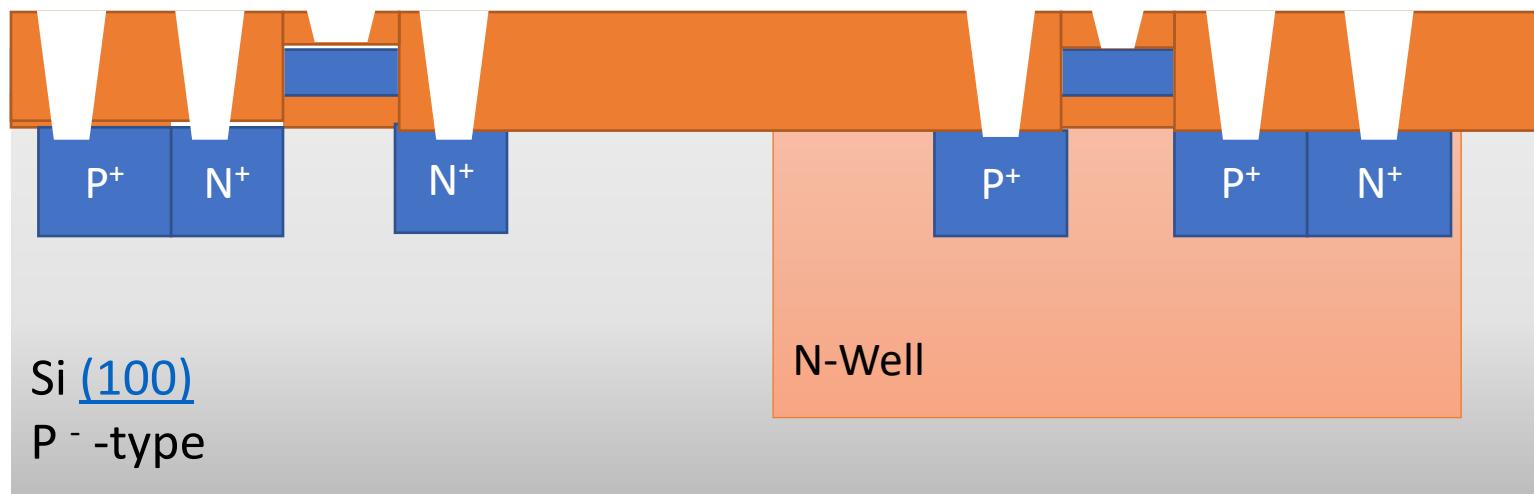
15. Repeat 12 to 14 for PMOS



16. Deposition of Thick Oxide

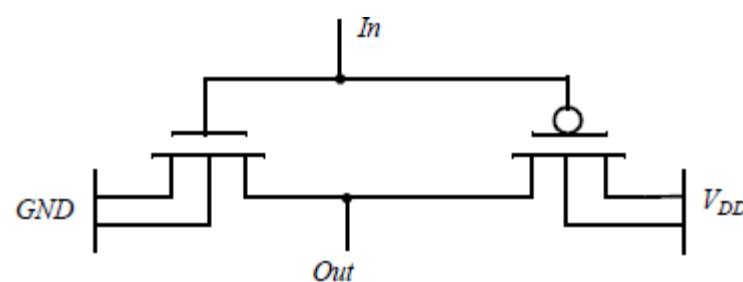
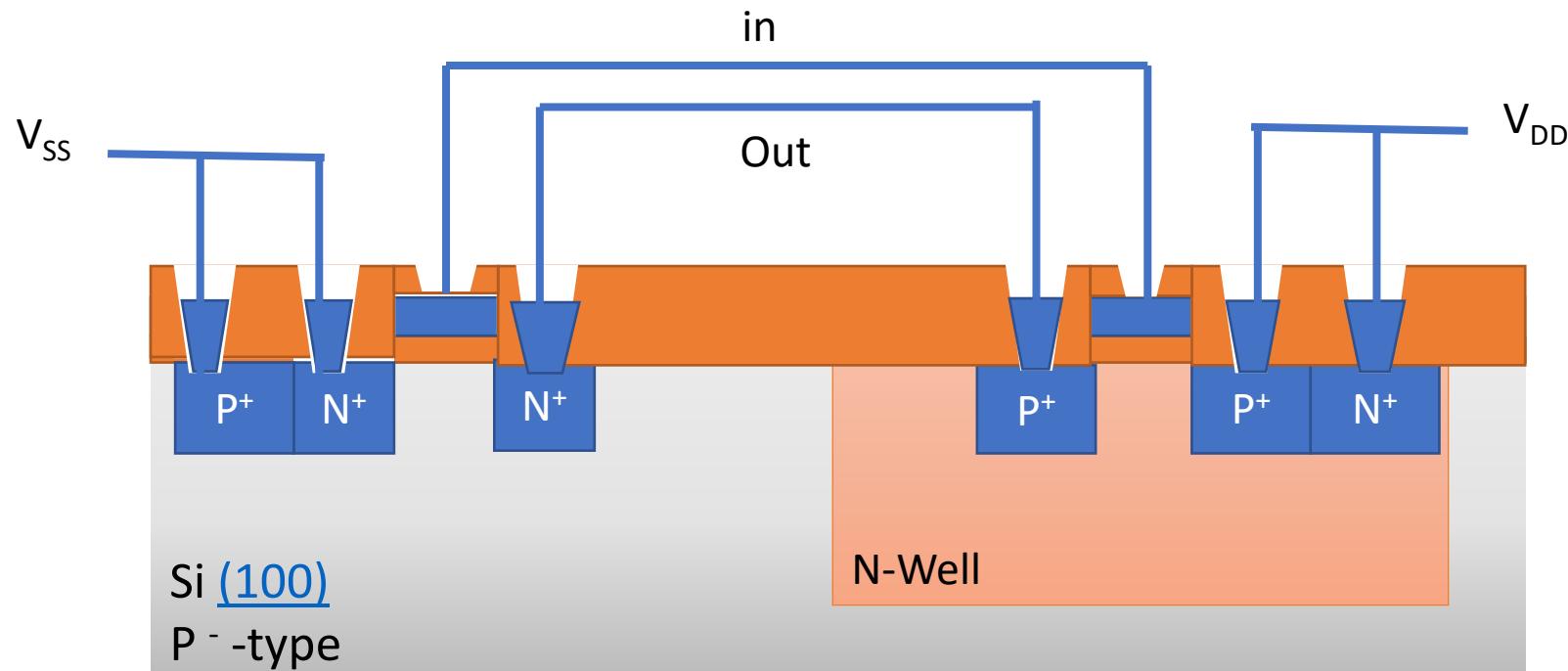


17. Lithography and Oxide Patterning





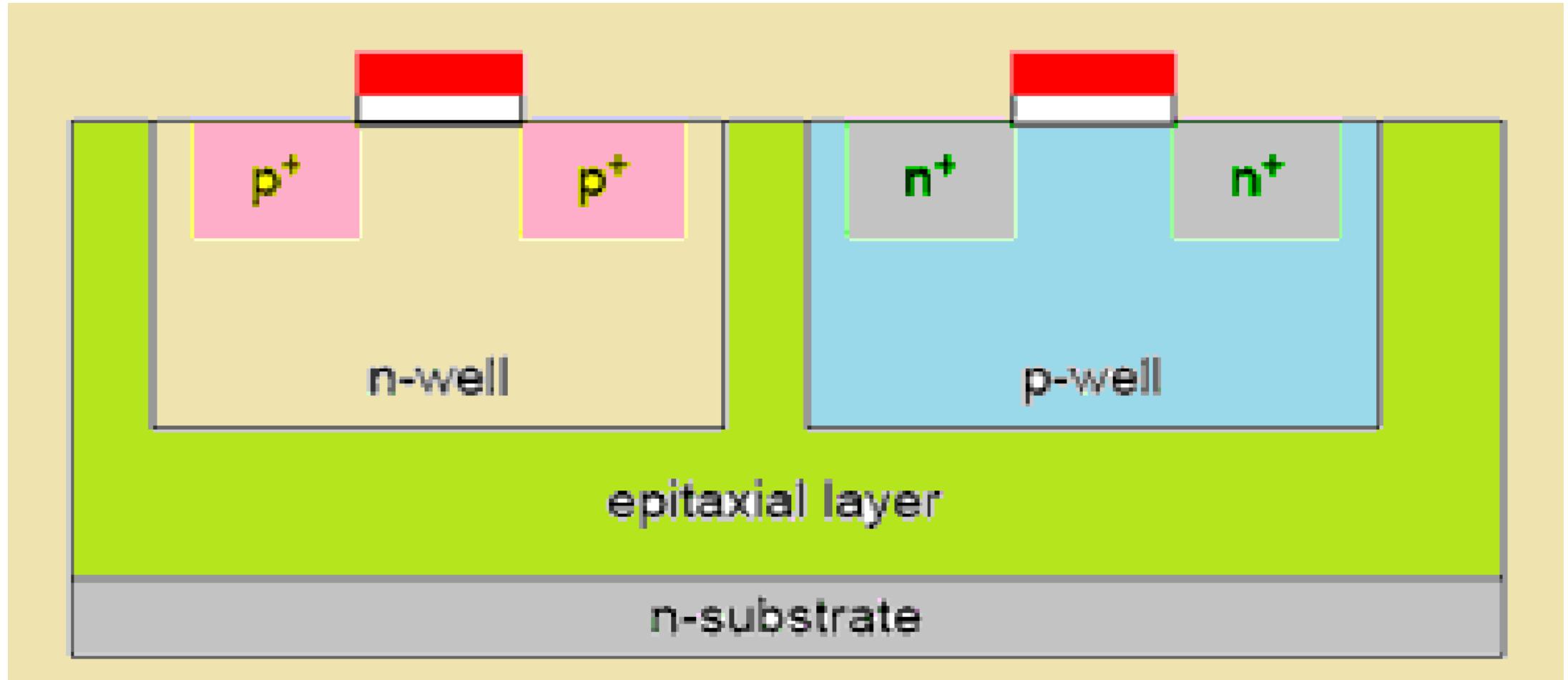
18. Metallization and packaging





2. p-well fabrication

3. Twin-tub fabrication





1. Selection of Substrate

Si [\(100\)](#)

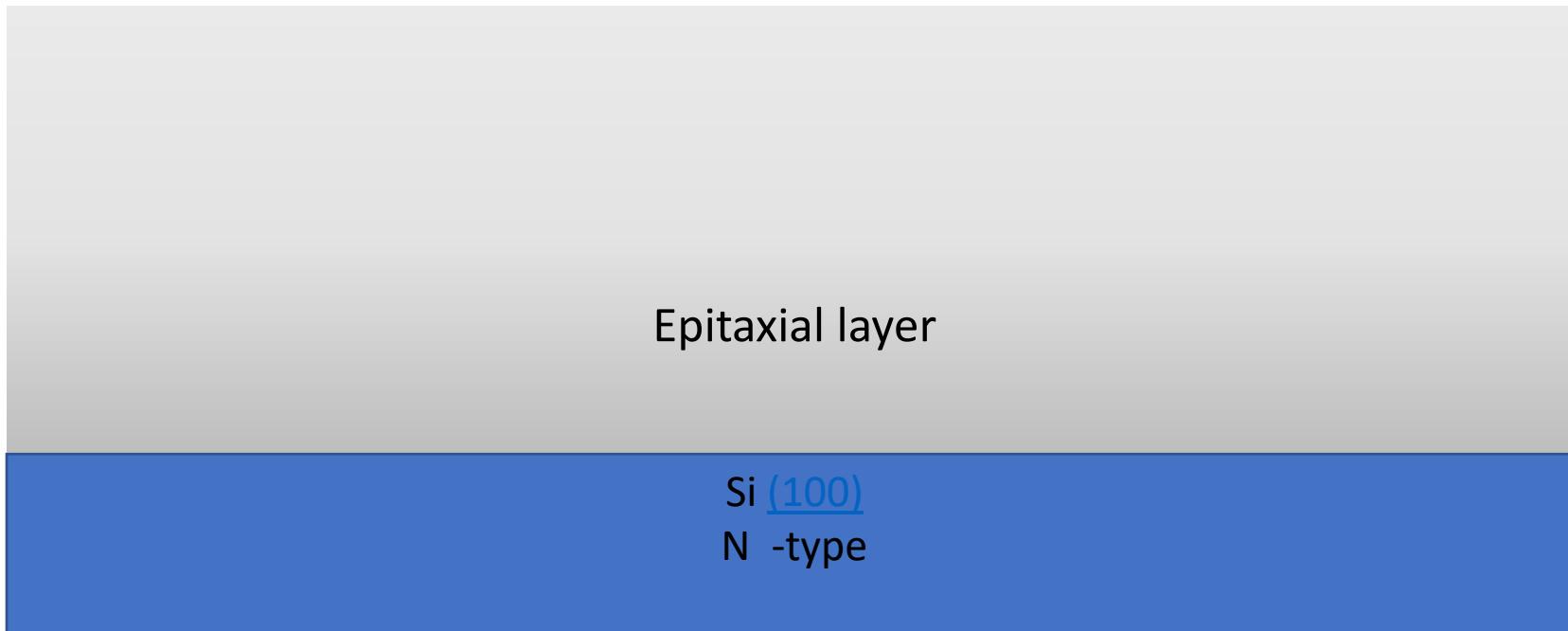
N -type

The twin-tub CMOS process starts with a high resistive n-type (100) silicon substrate.

2. Cleaning of Substrate

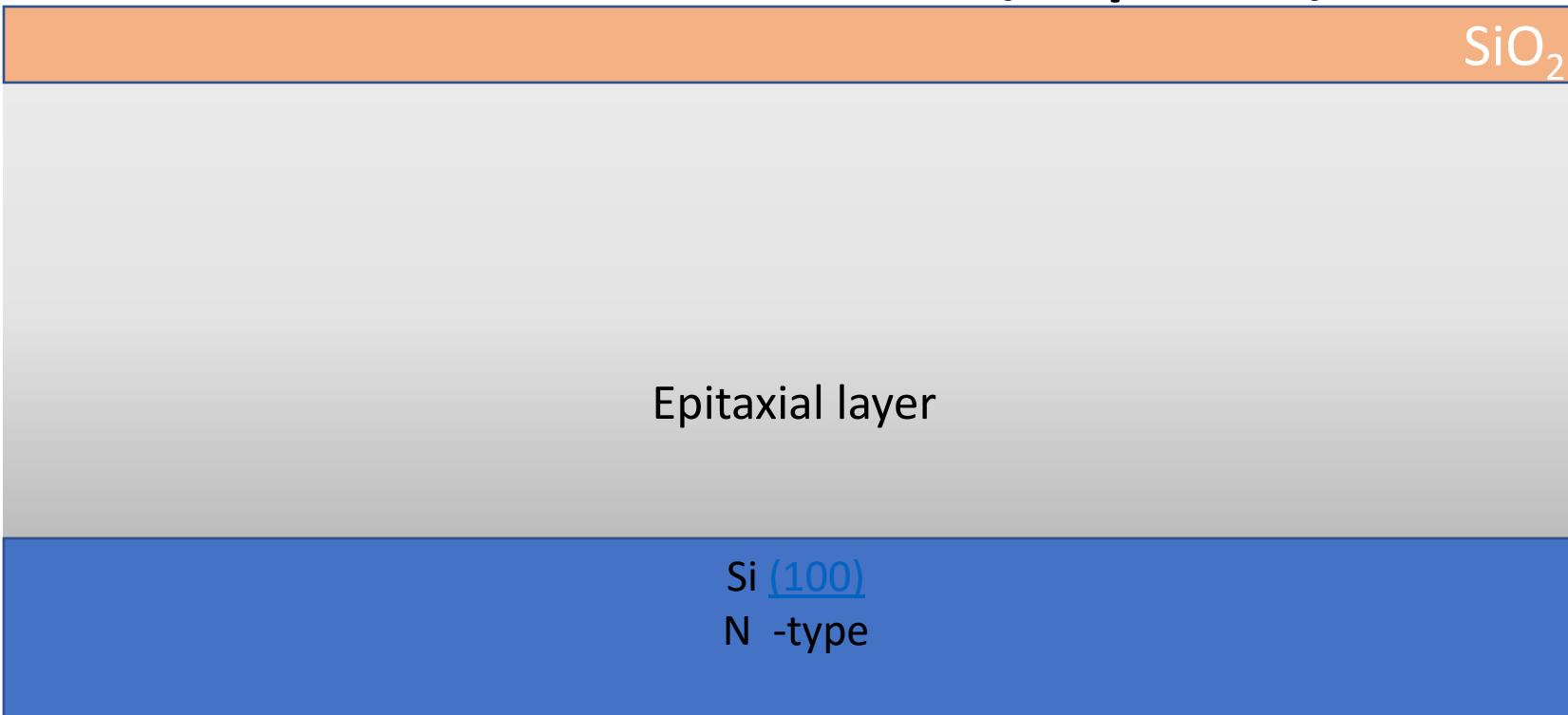


Epitaxial Layer Deposition



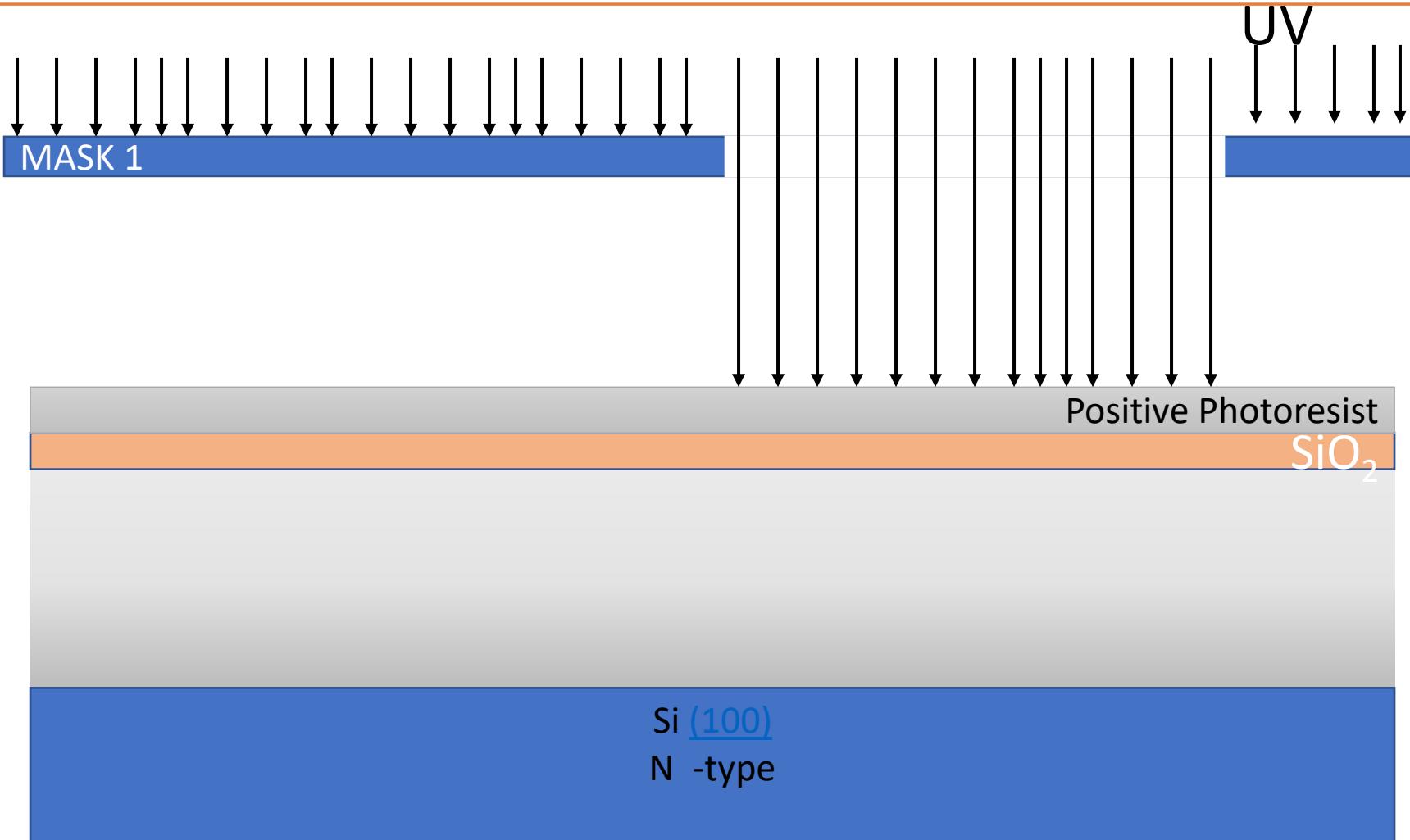


3. Oxidation($1\mu\text{m}$)

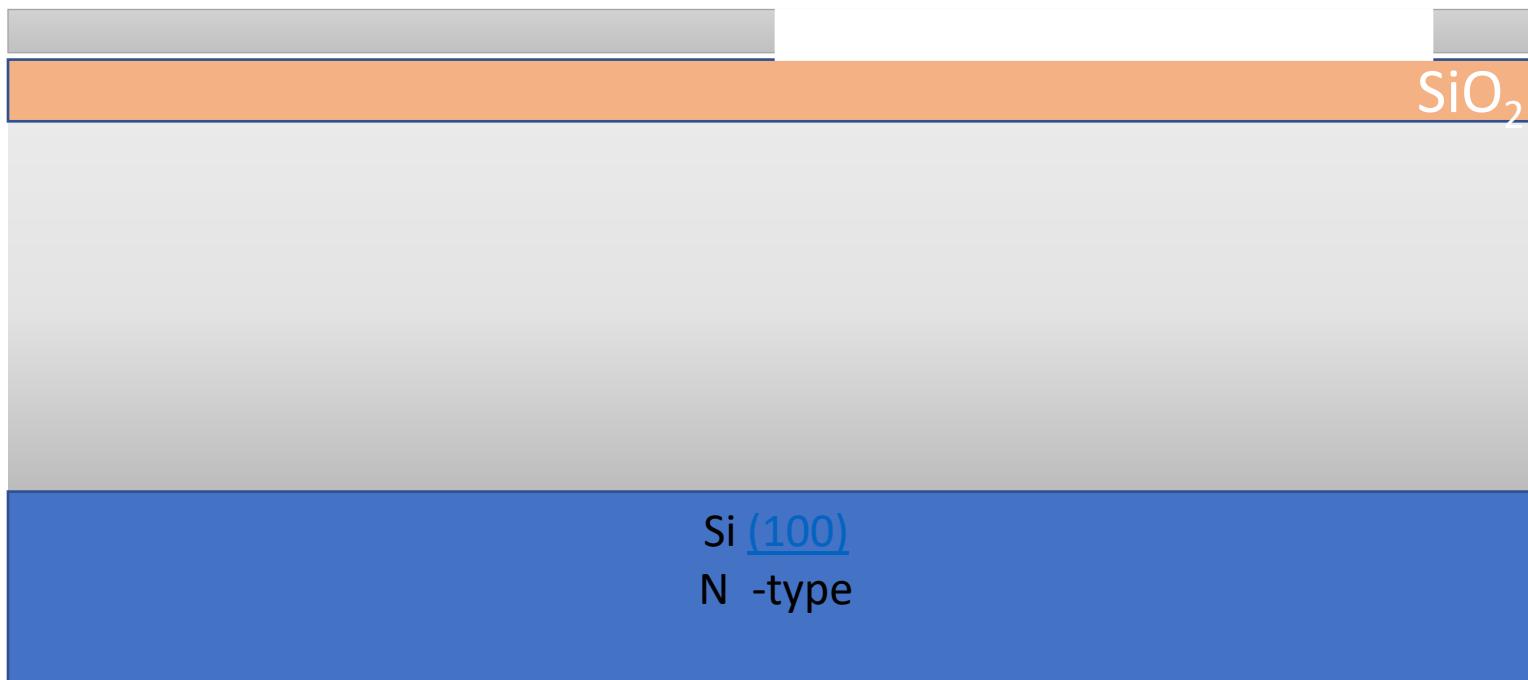


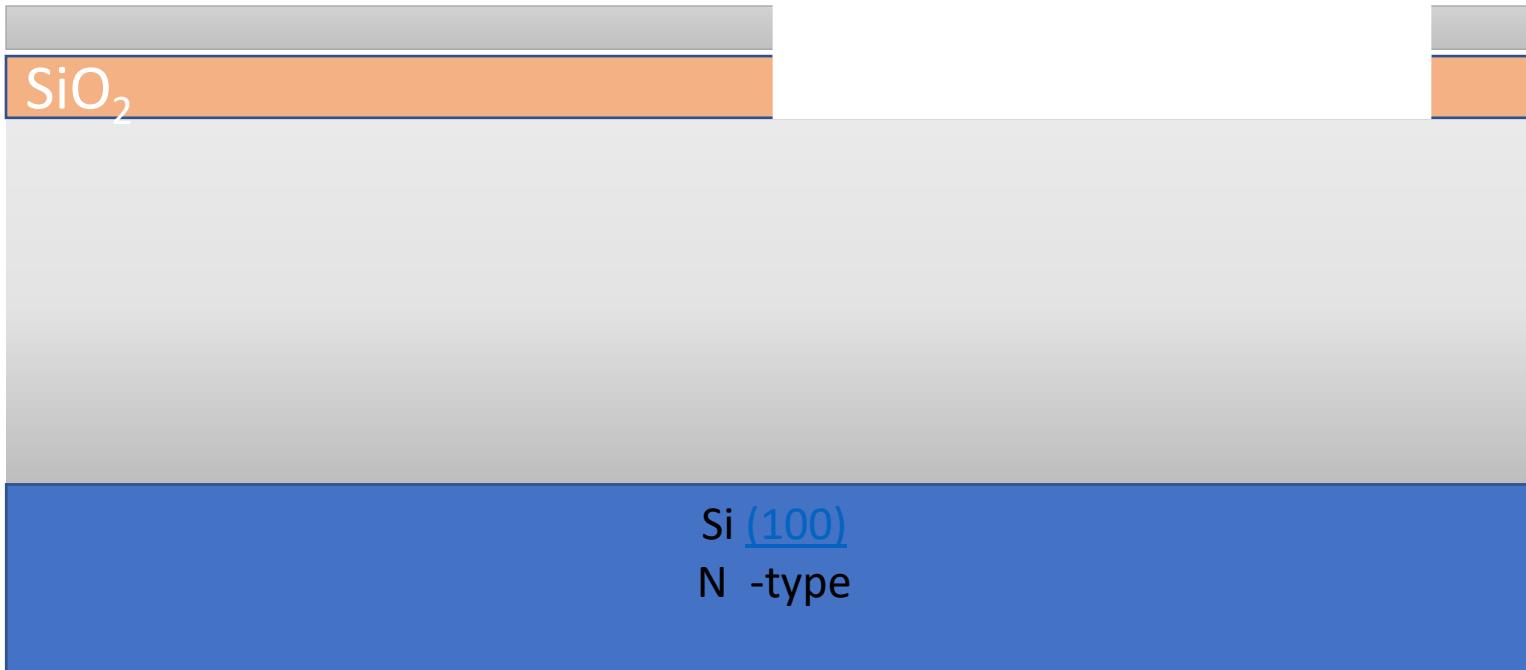


Lithography(MASK1)



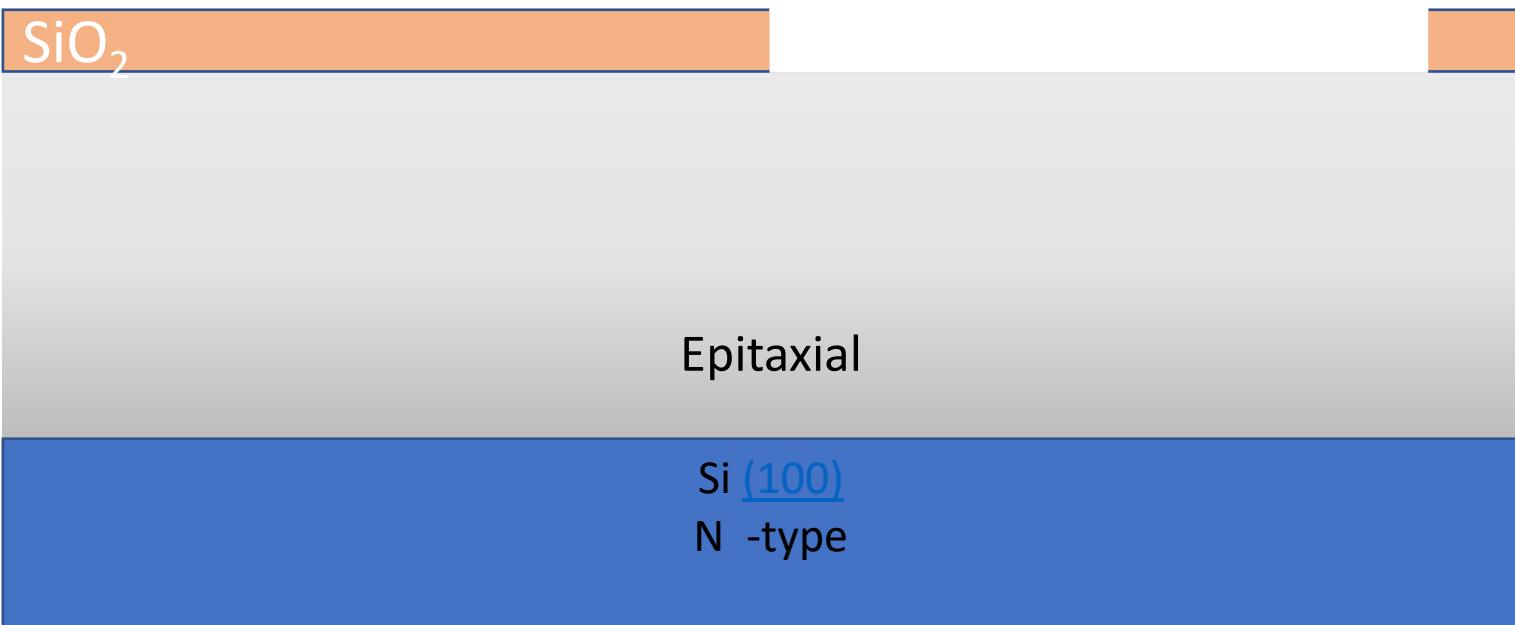
HF(Oxide) Cleaning & PR Etching

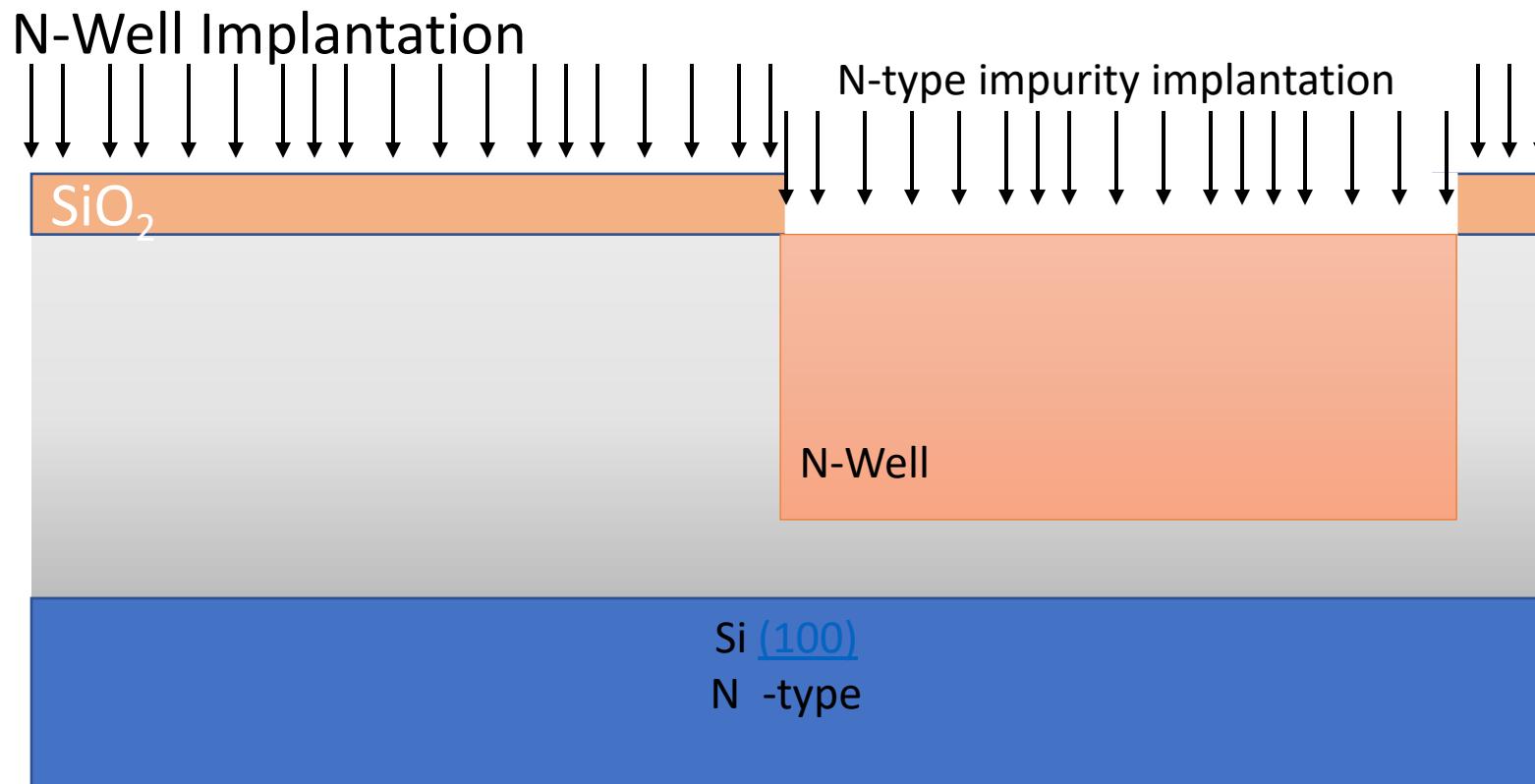




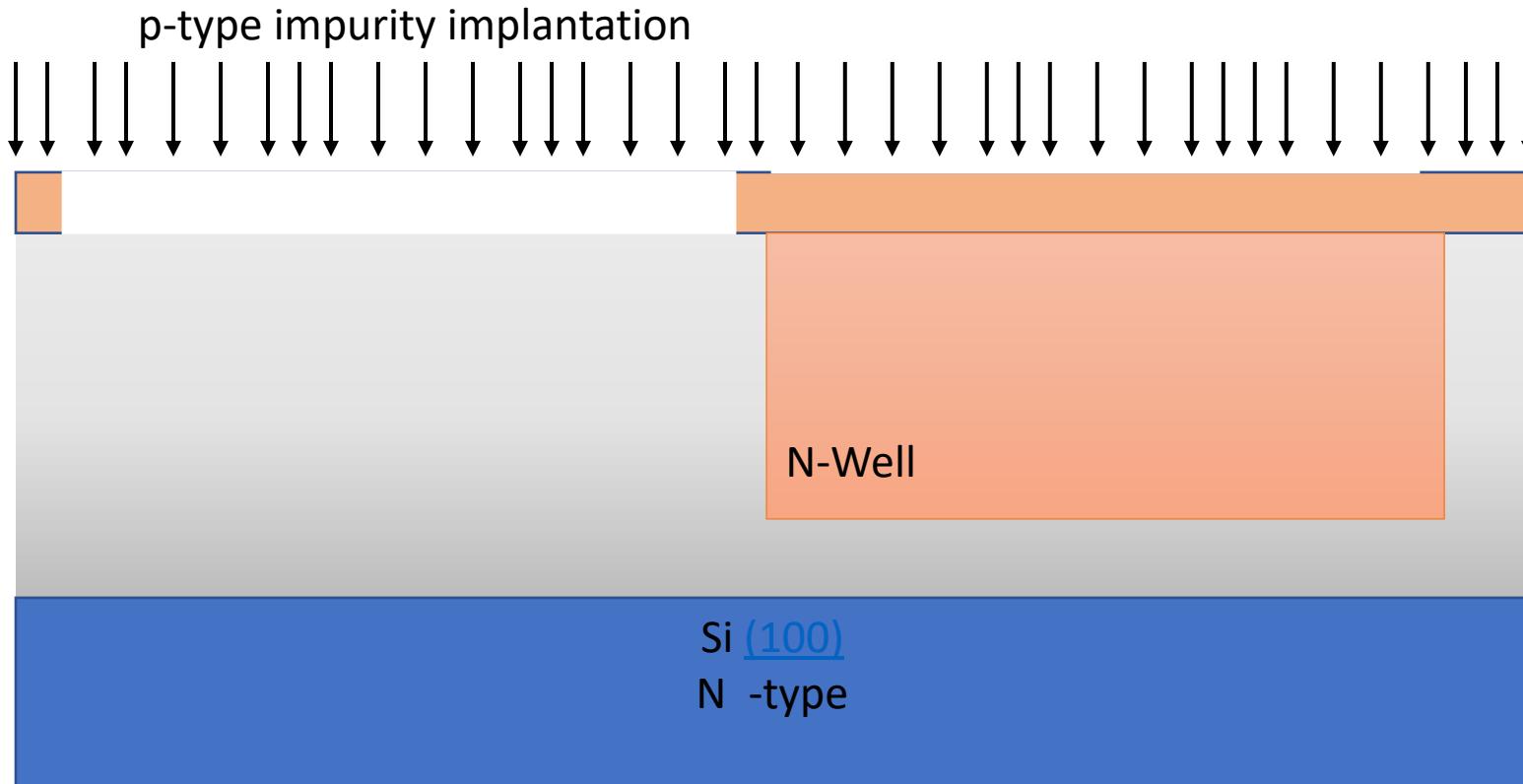


HF(Oxide) Cleaning & PR Etching

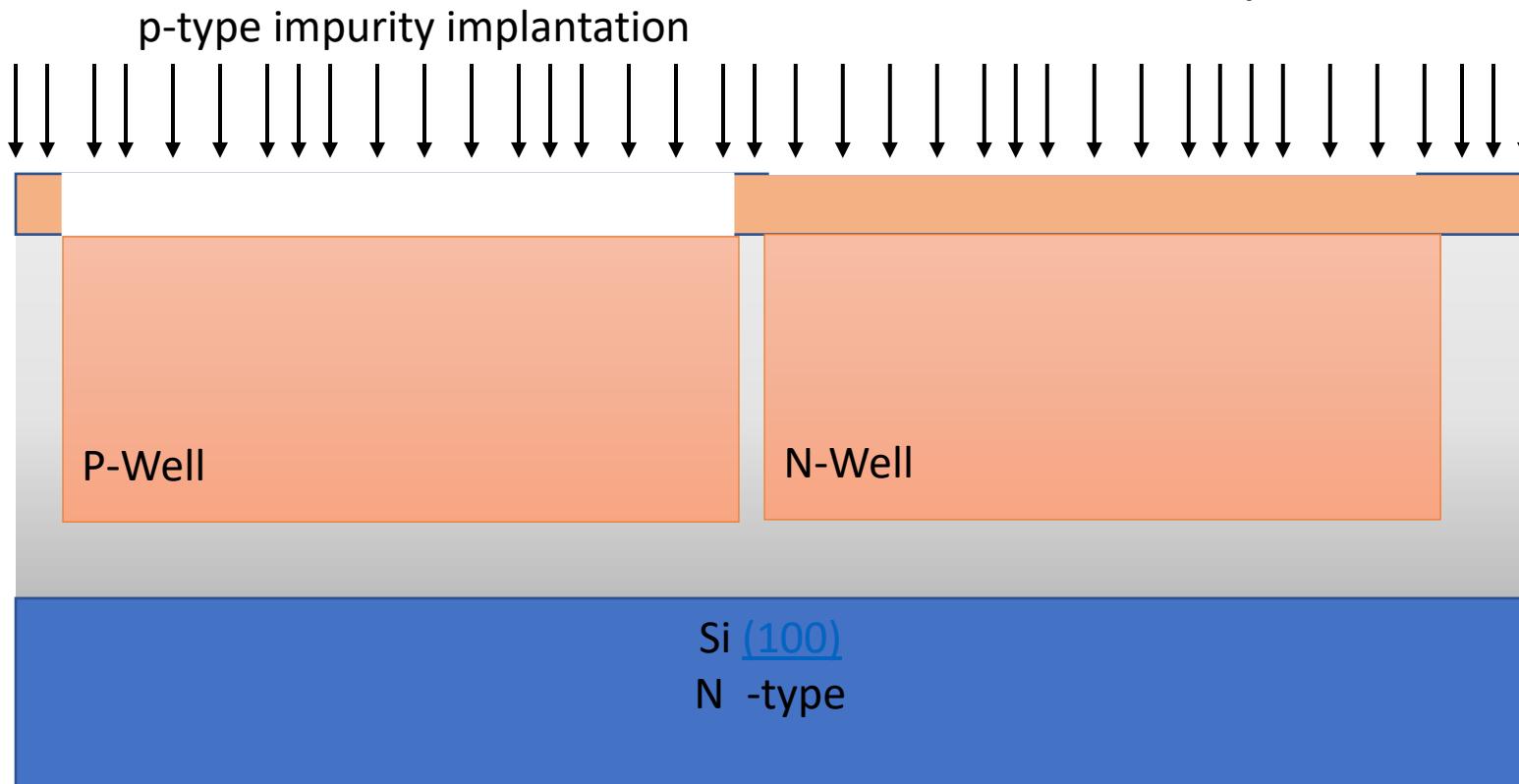




P-Well Implantation

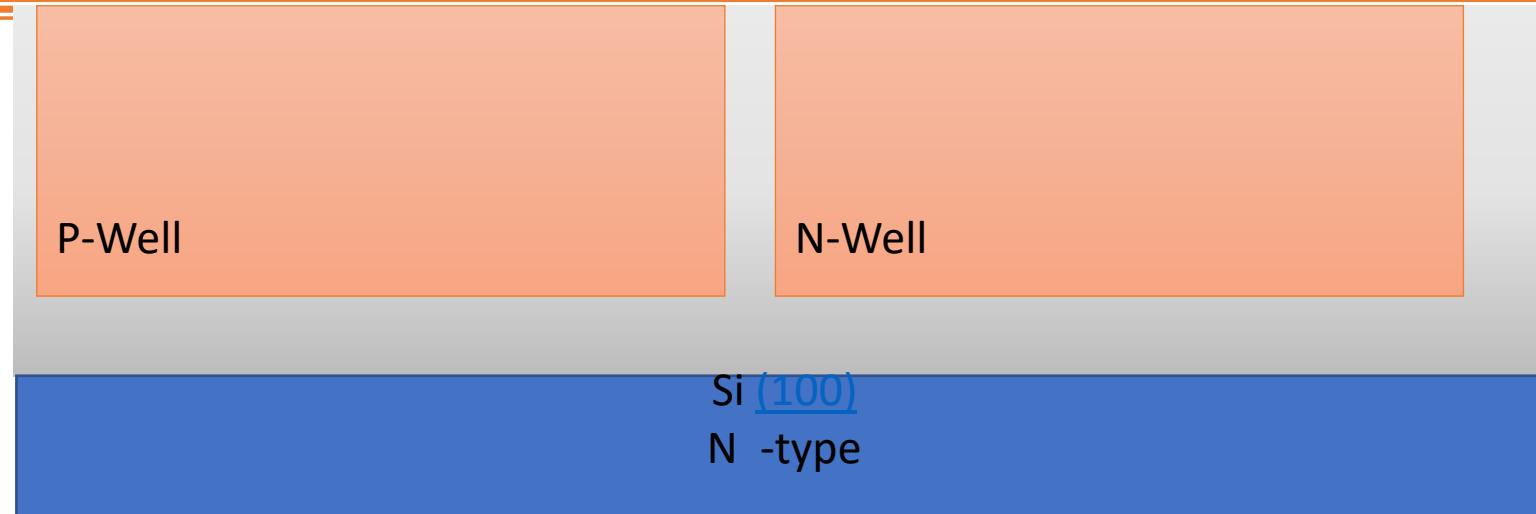


P-Well Implantation

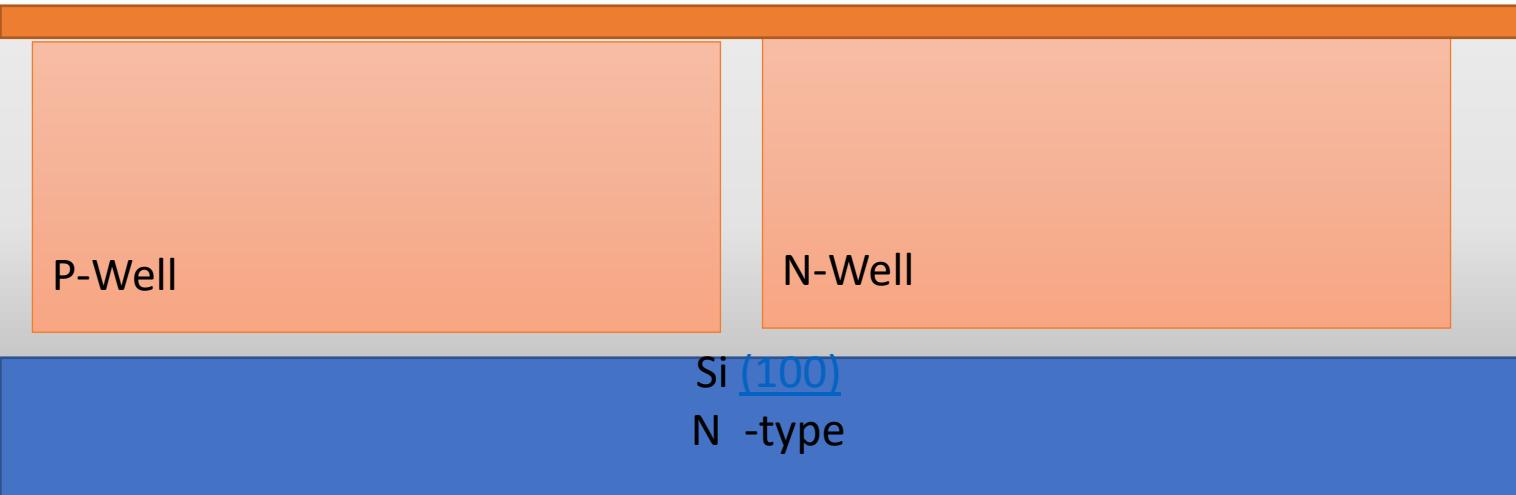




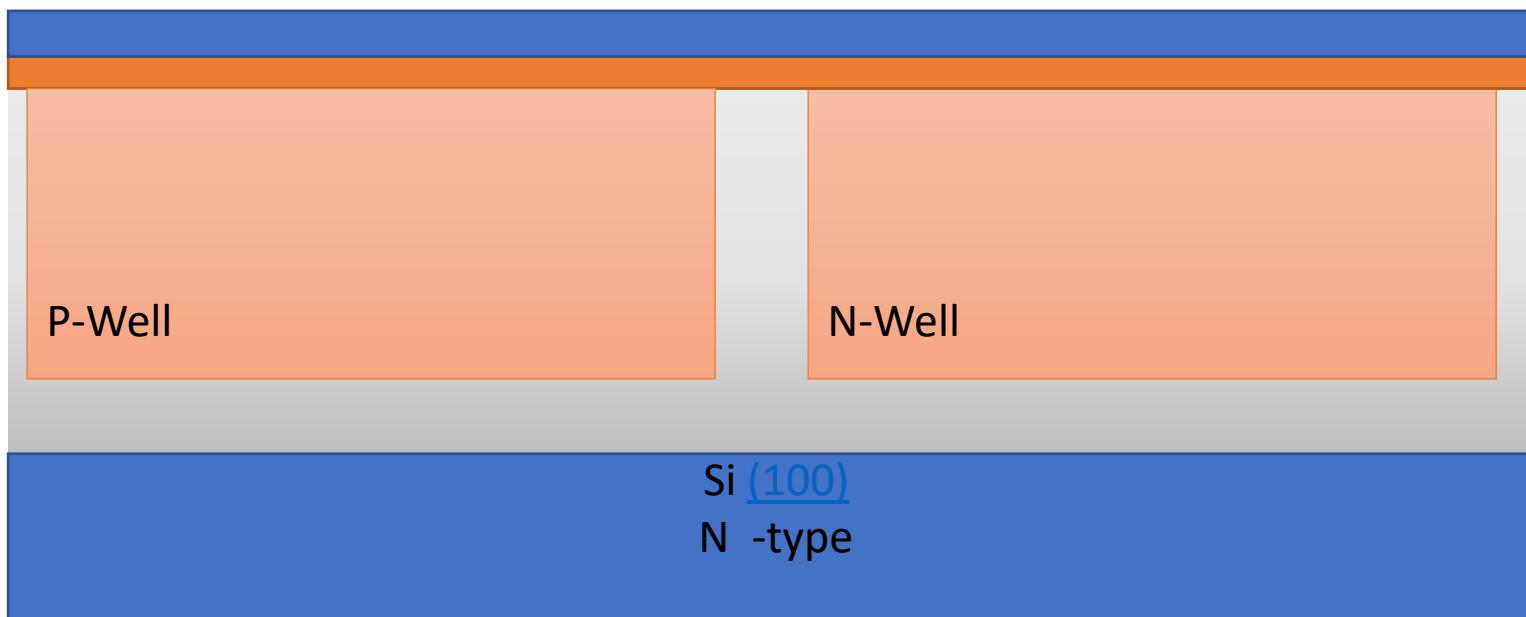
Oxide Cleaning



Gate Oxidation

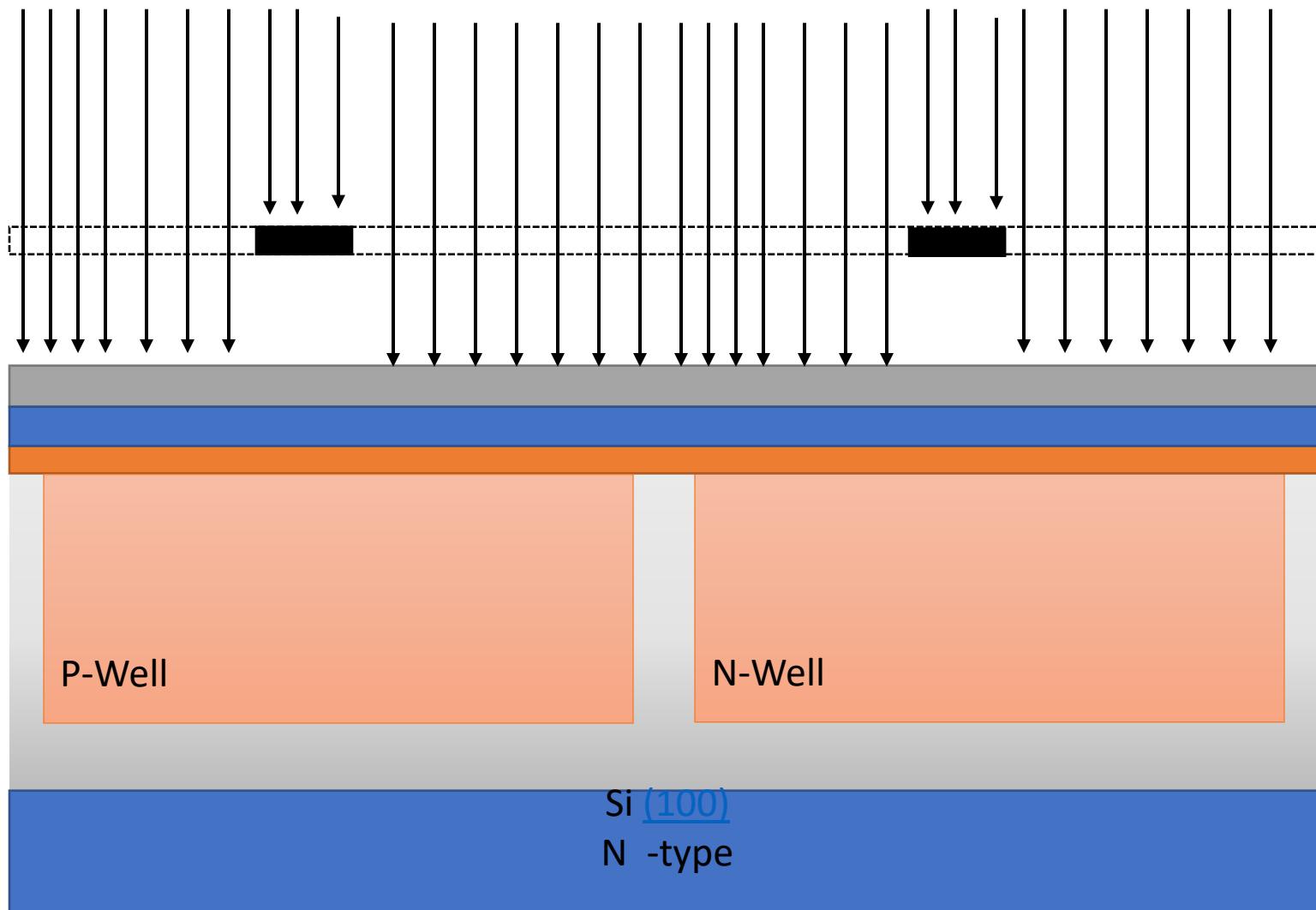


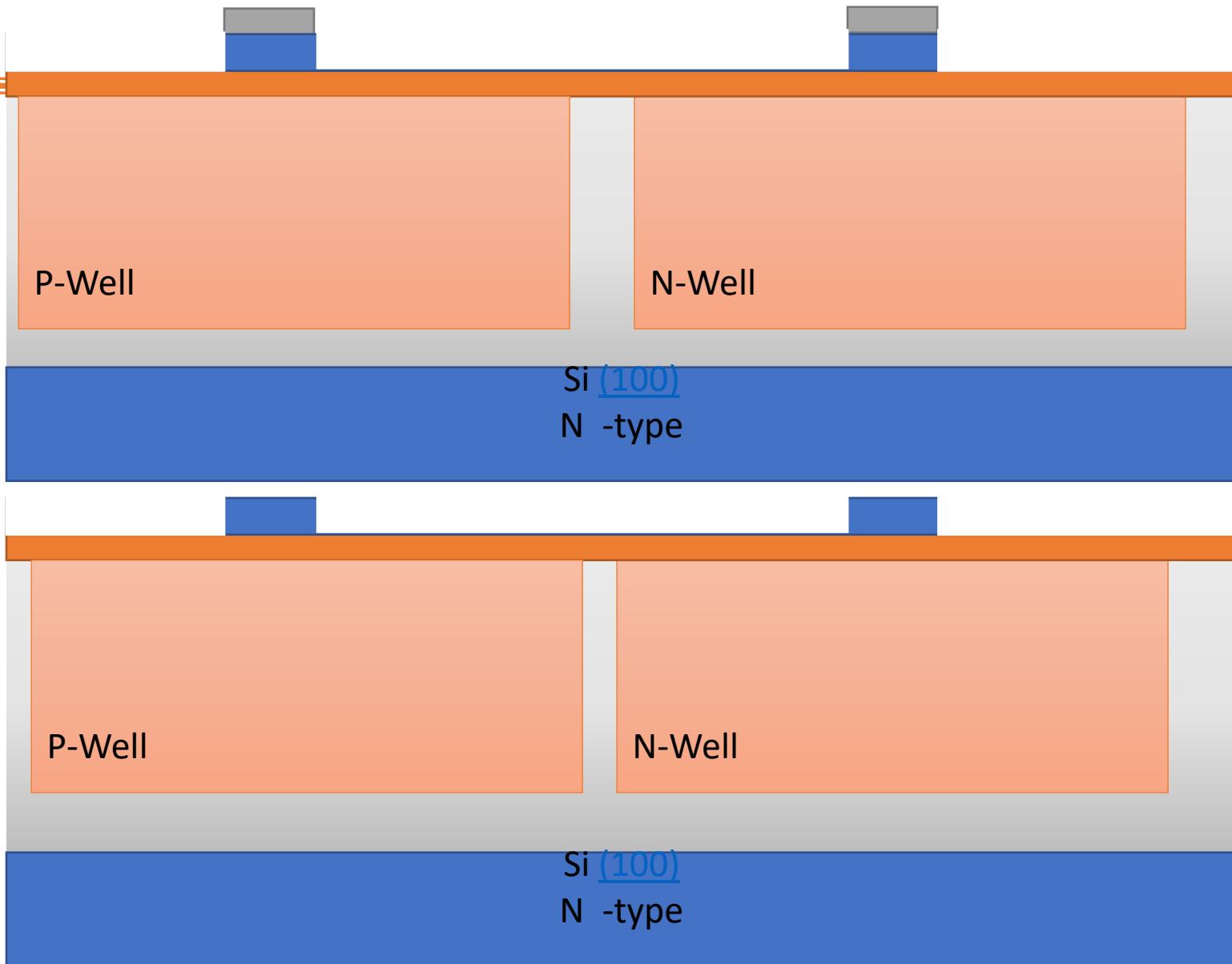
Poly-Silicon Deposition





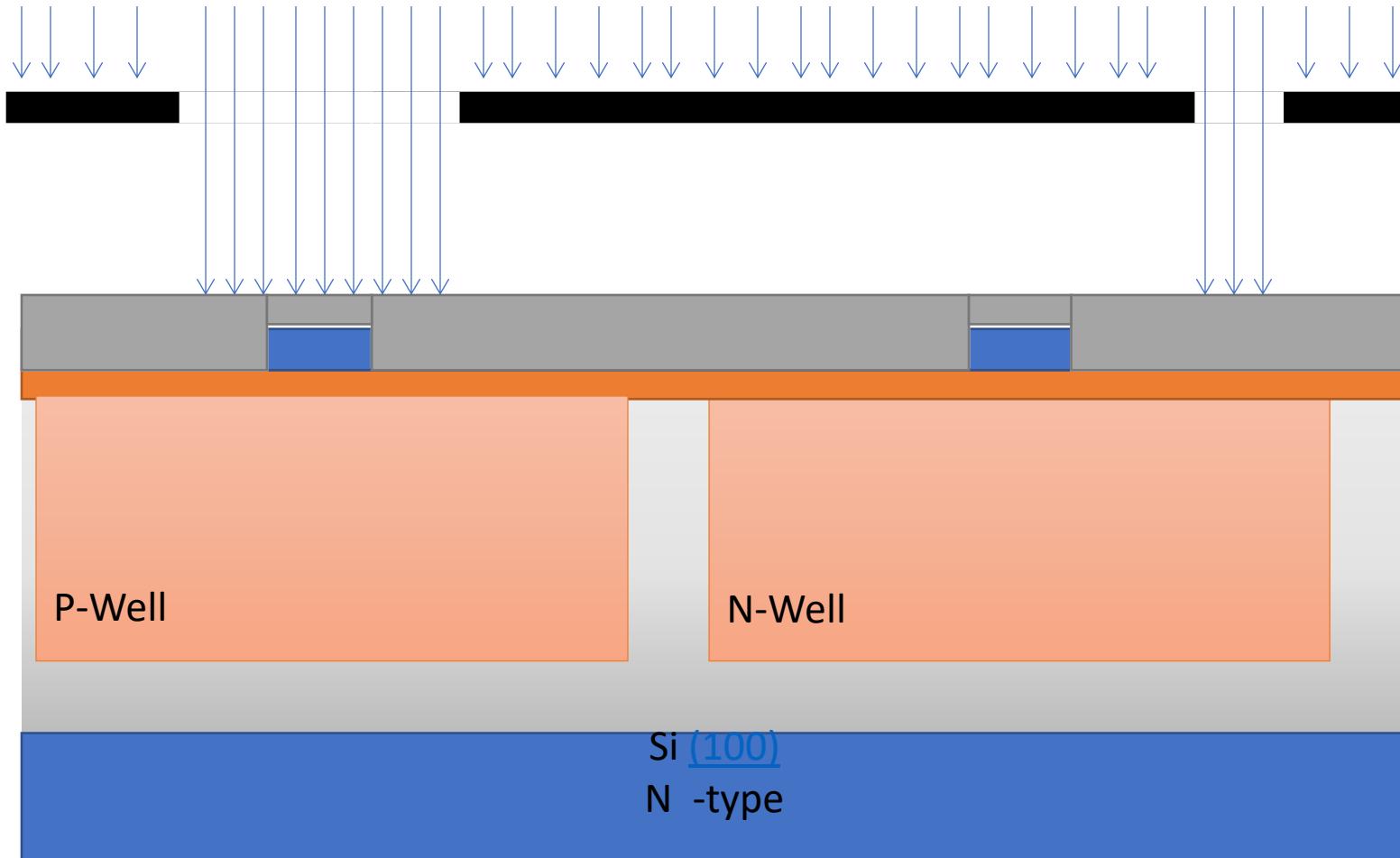
Lithography(MASK2)

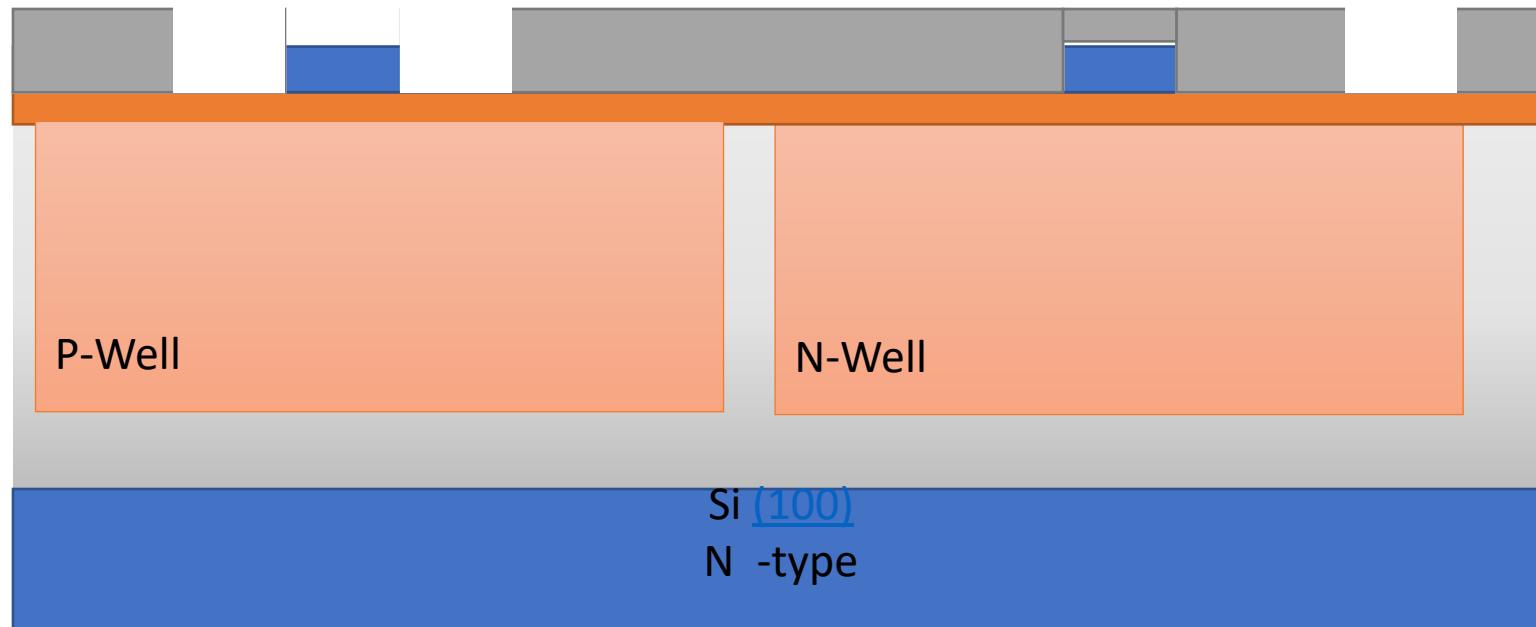




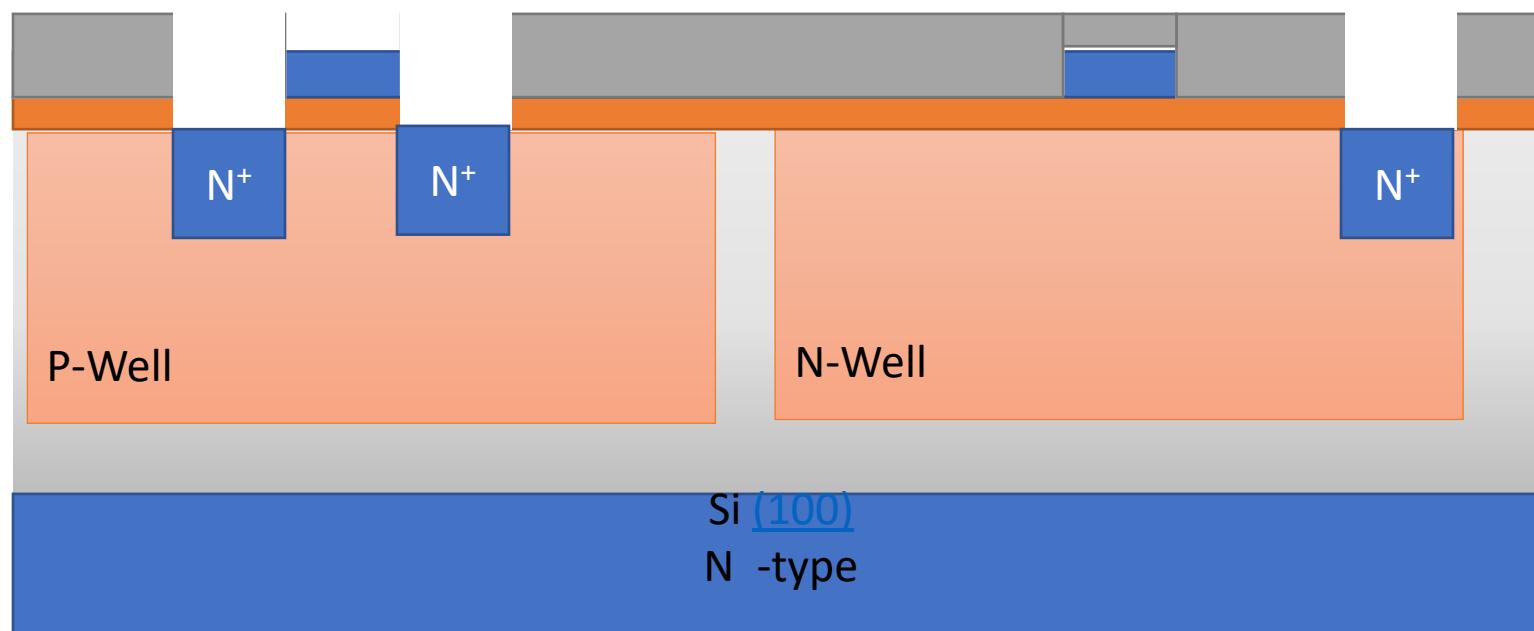


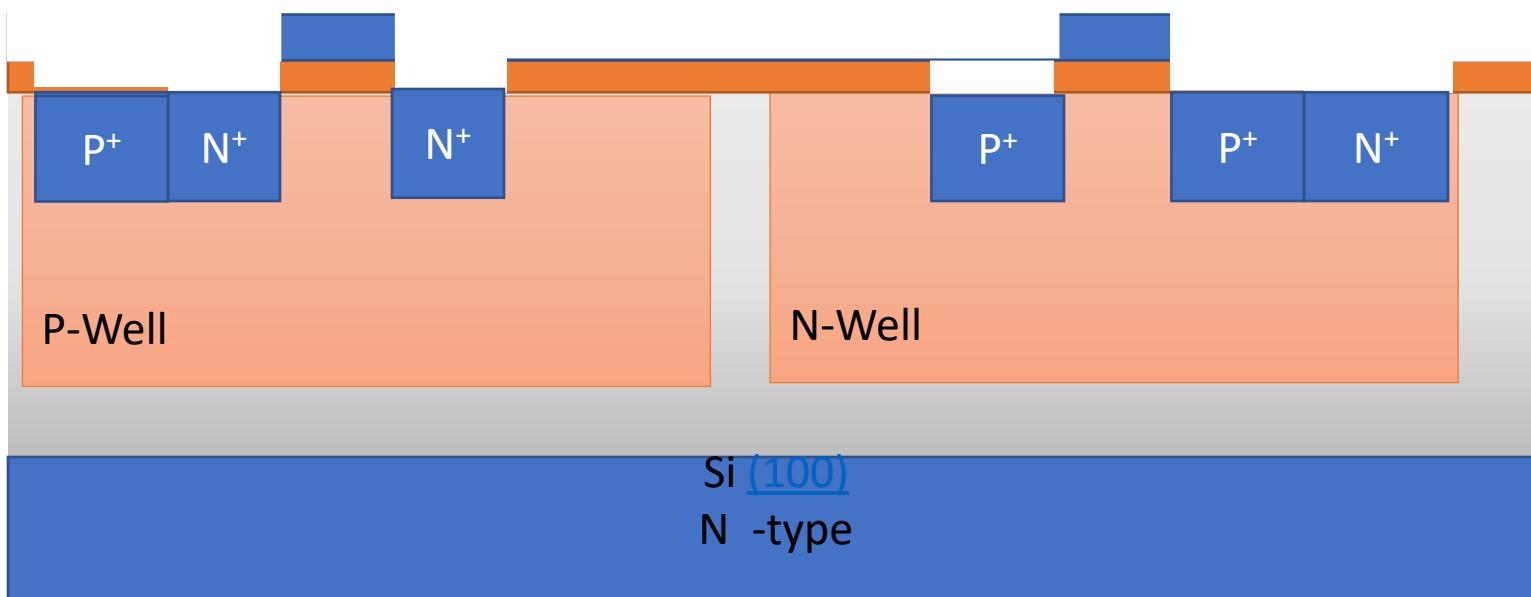
Lithography(MASK3)



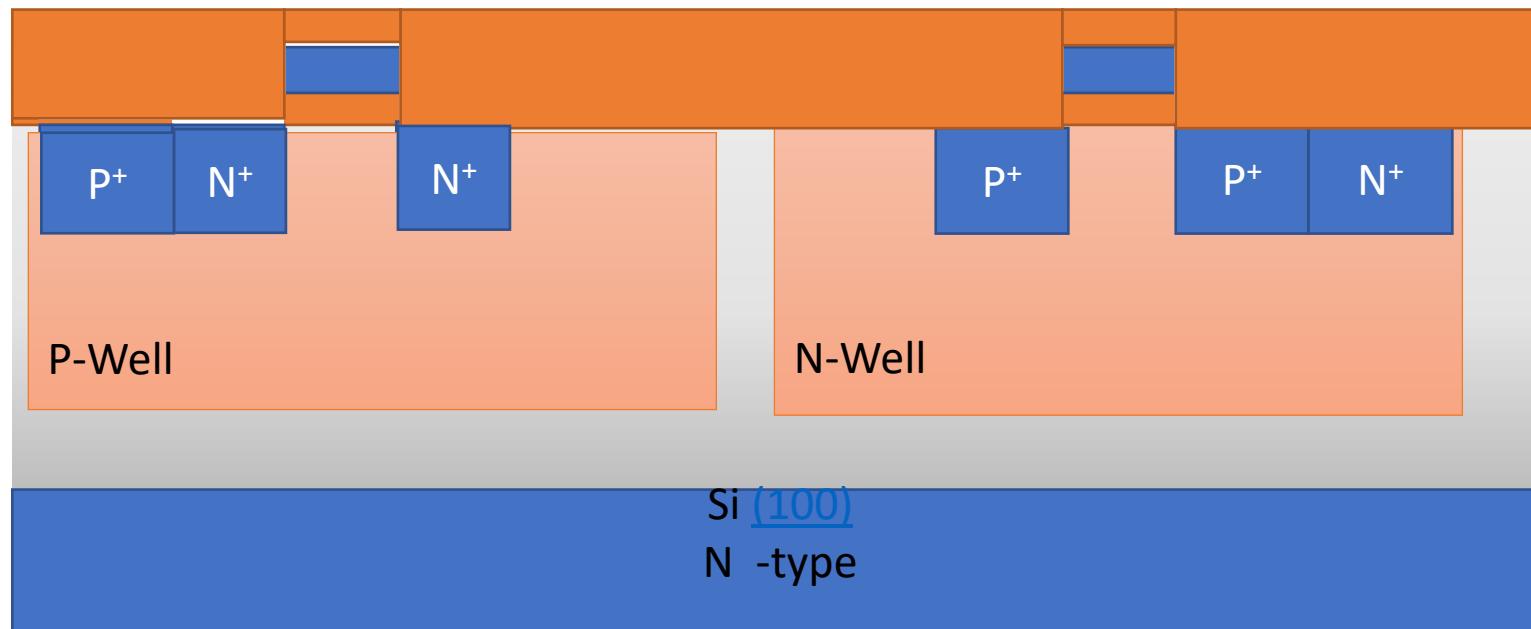


Ion Implantation

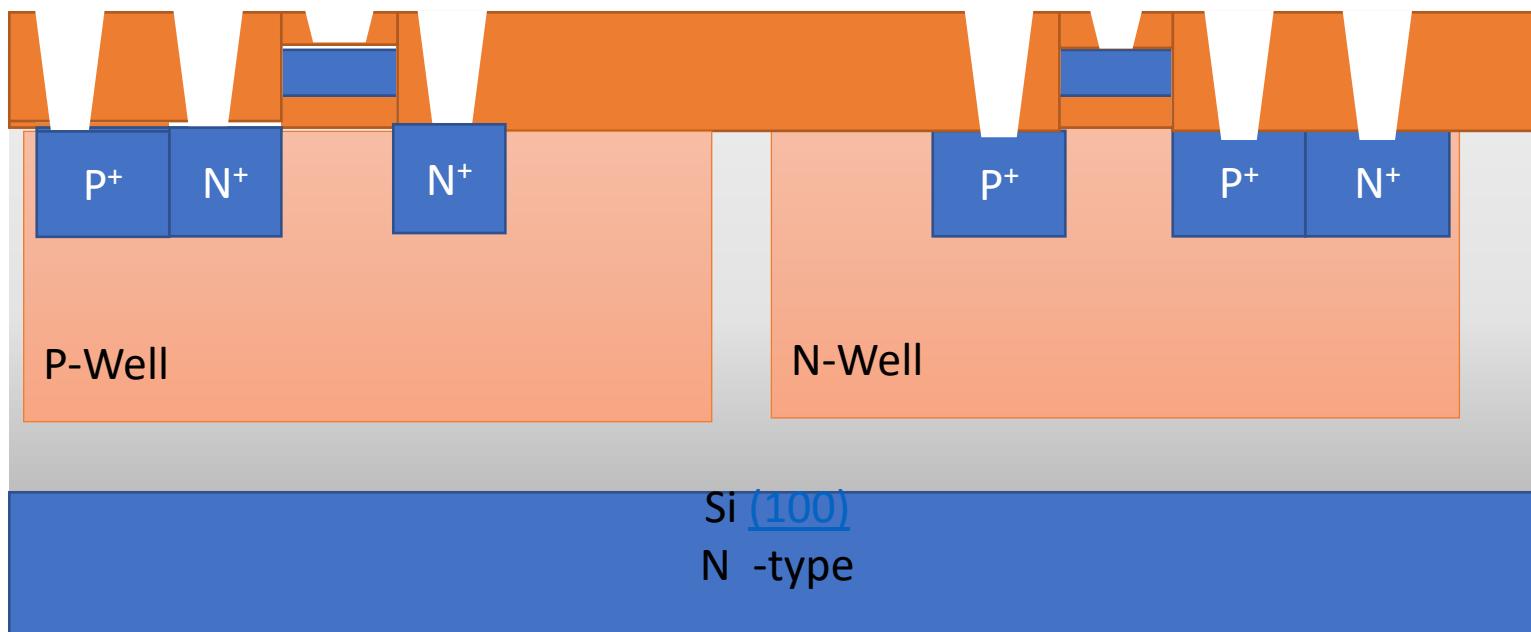




Thick Oxide Deposition

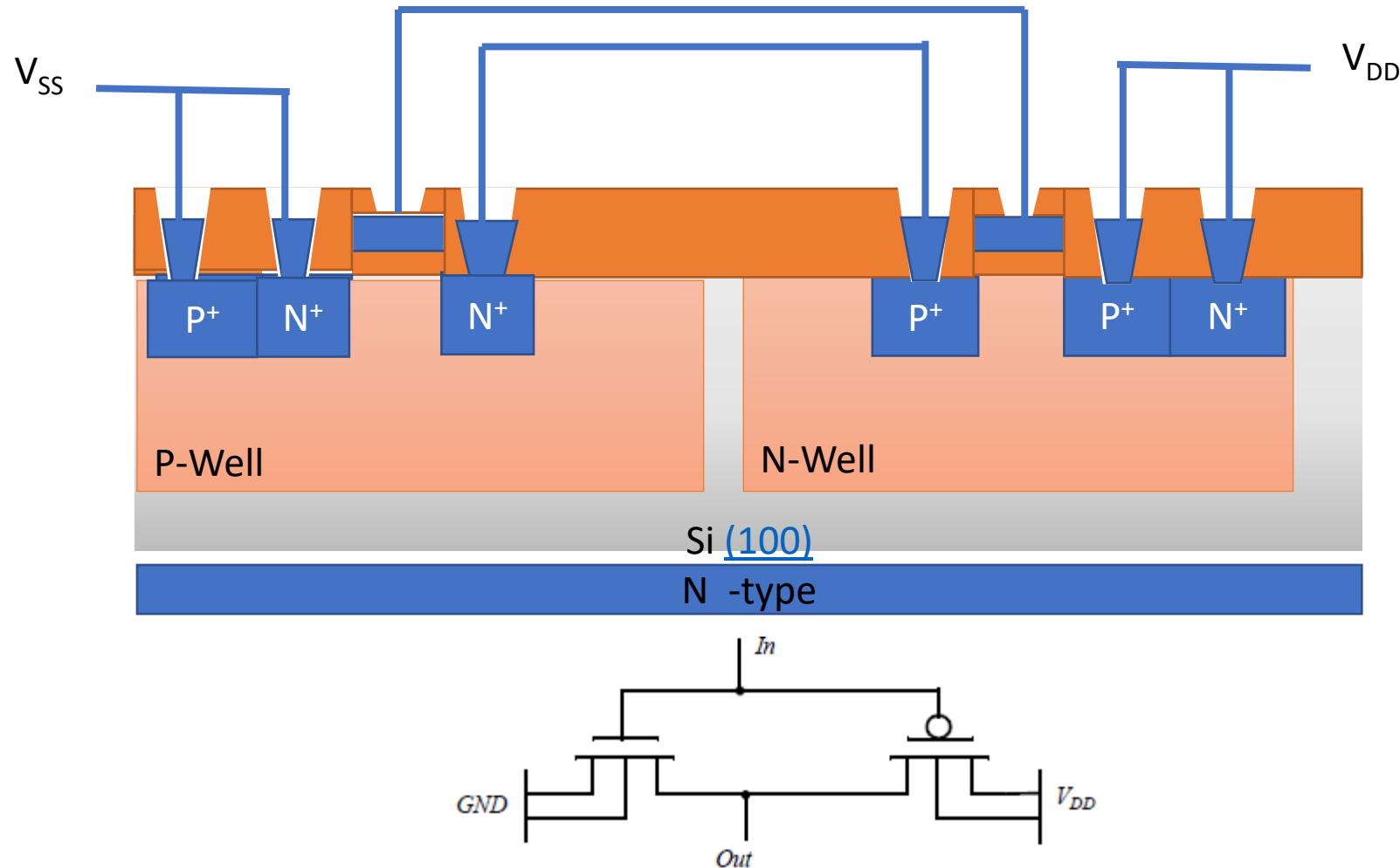


Lithography and Oxide Patterning



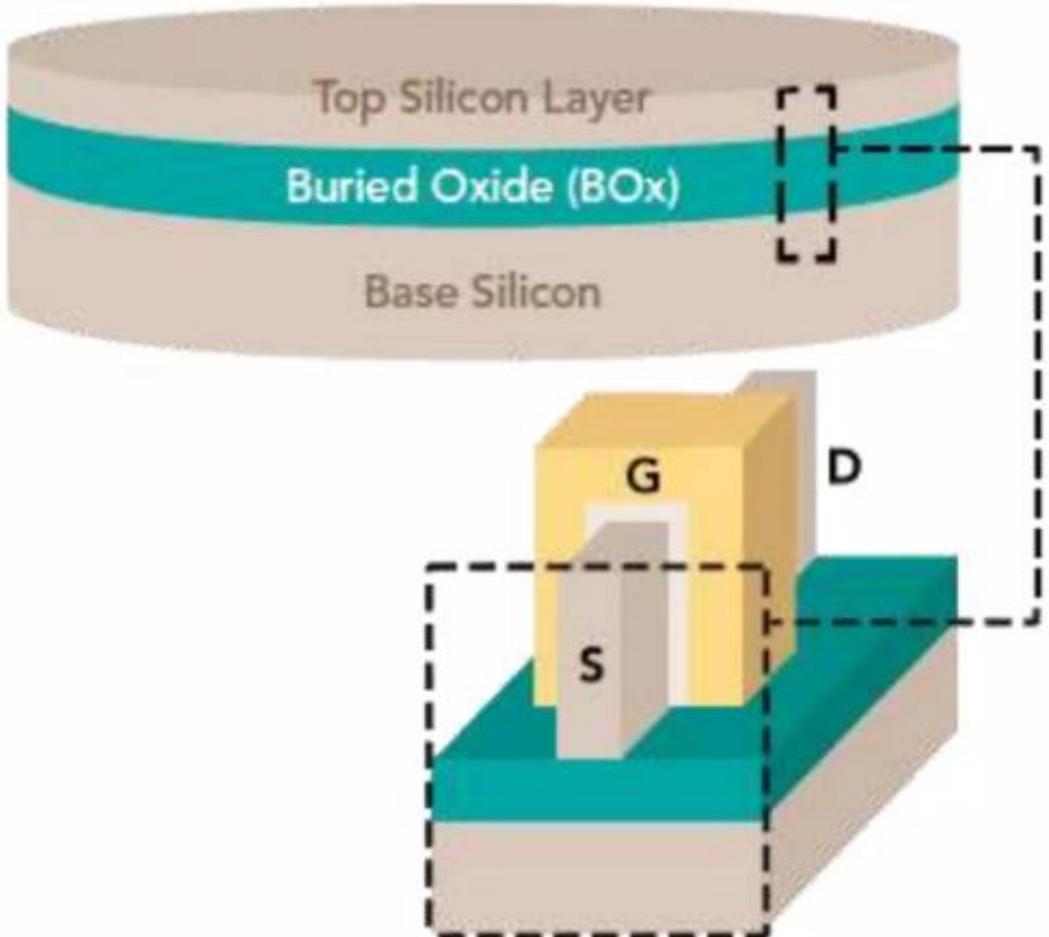


Metallization and Patterning

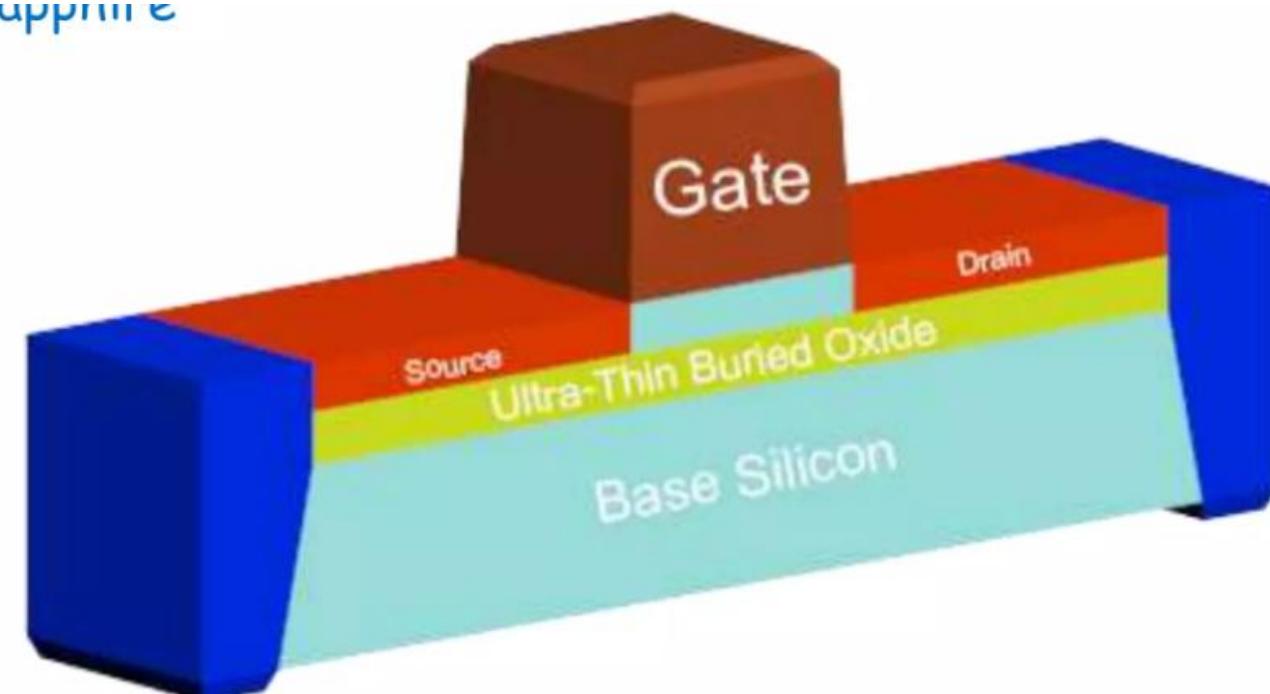


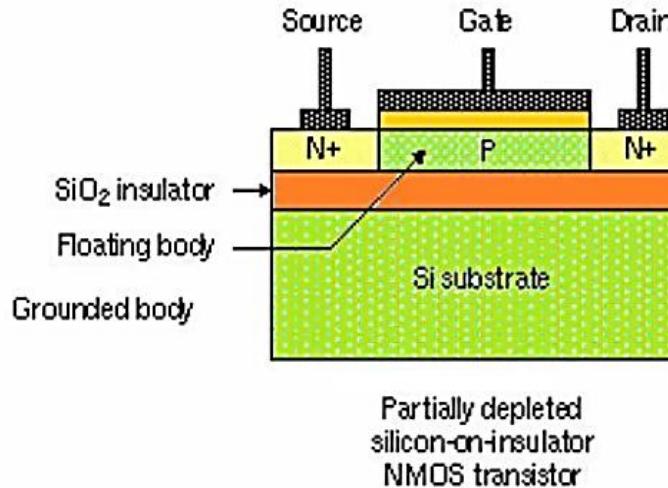
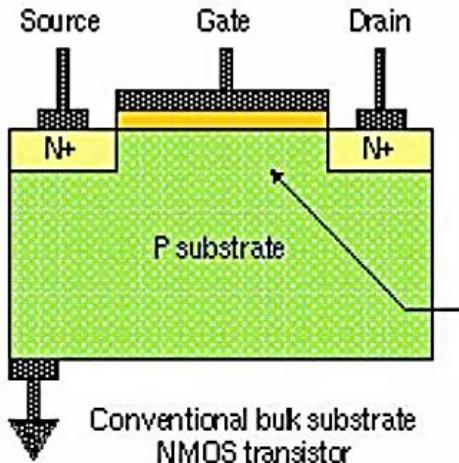


4. Silicon on Insulator (SOI)

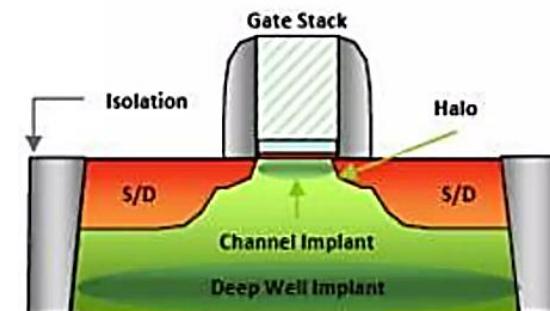


uprise

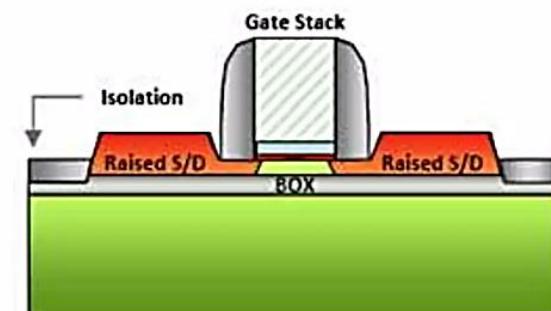




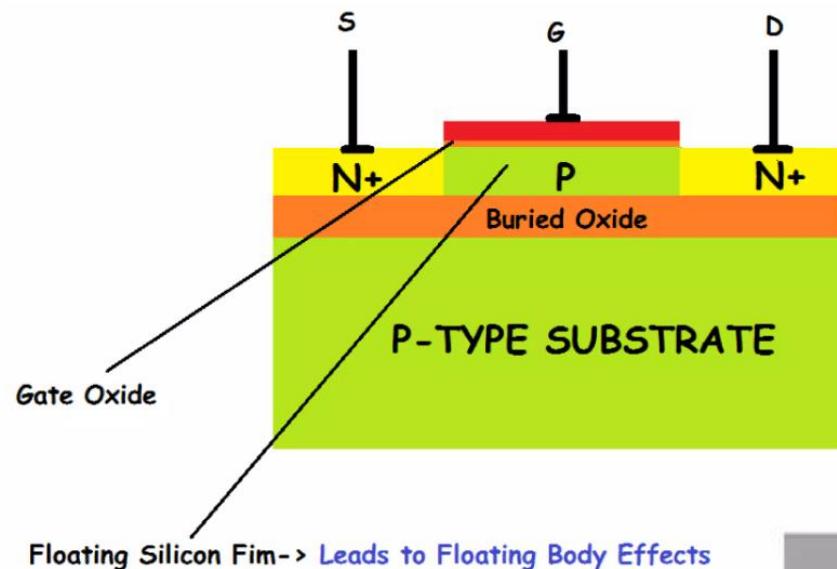
- **BOX: Buried Oxide**



Bulk Device



FD-SOI Device





SOI fabrication steps

1. Substrate selection

Silicon substrate

2. Deposition of insulator

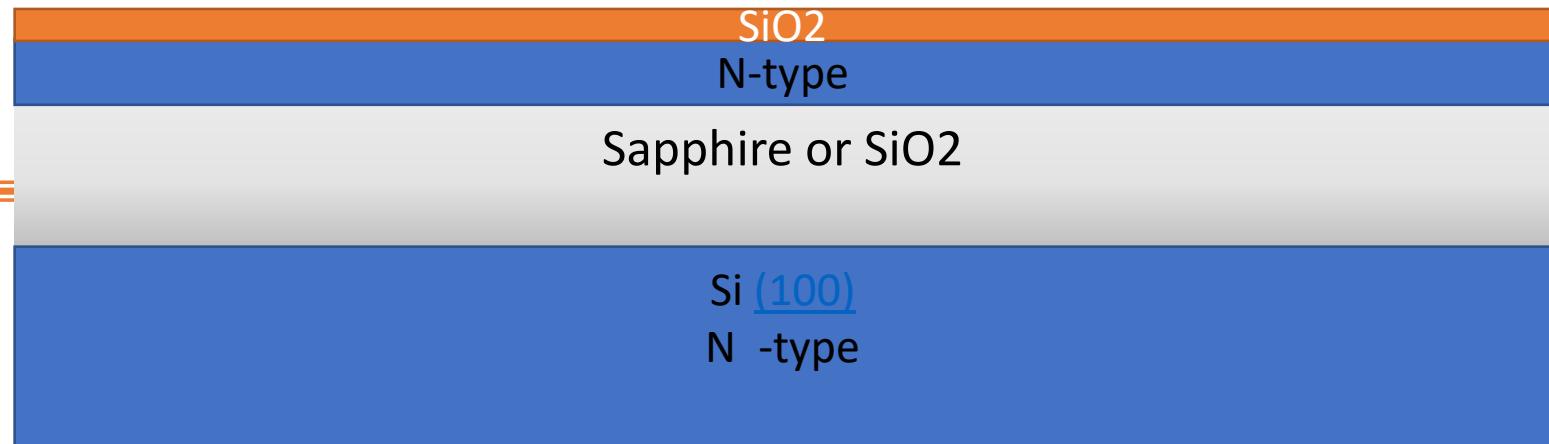
Si [\(100\)](#)
N -type

3. Deposition of Silicon

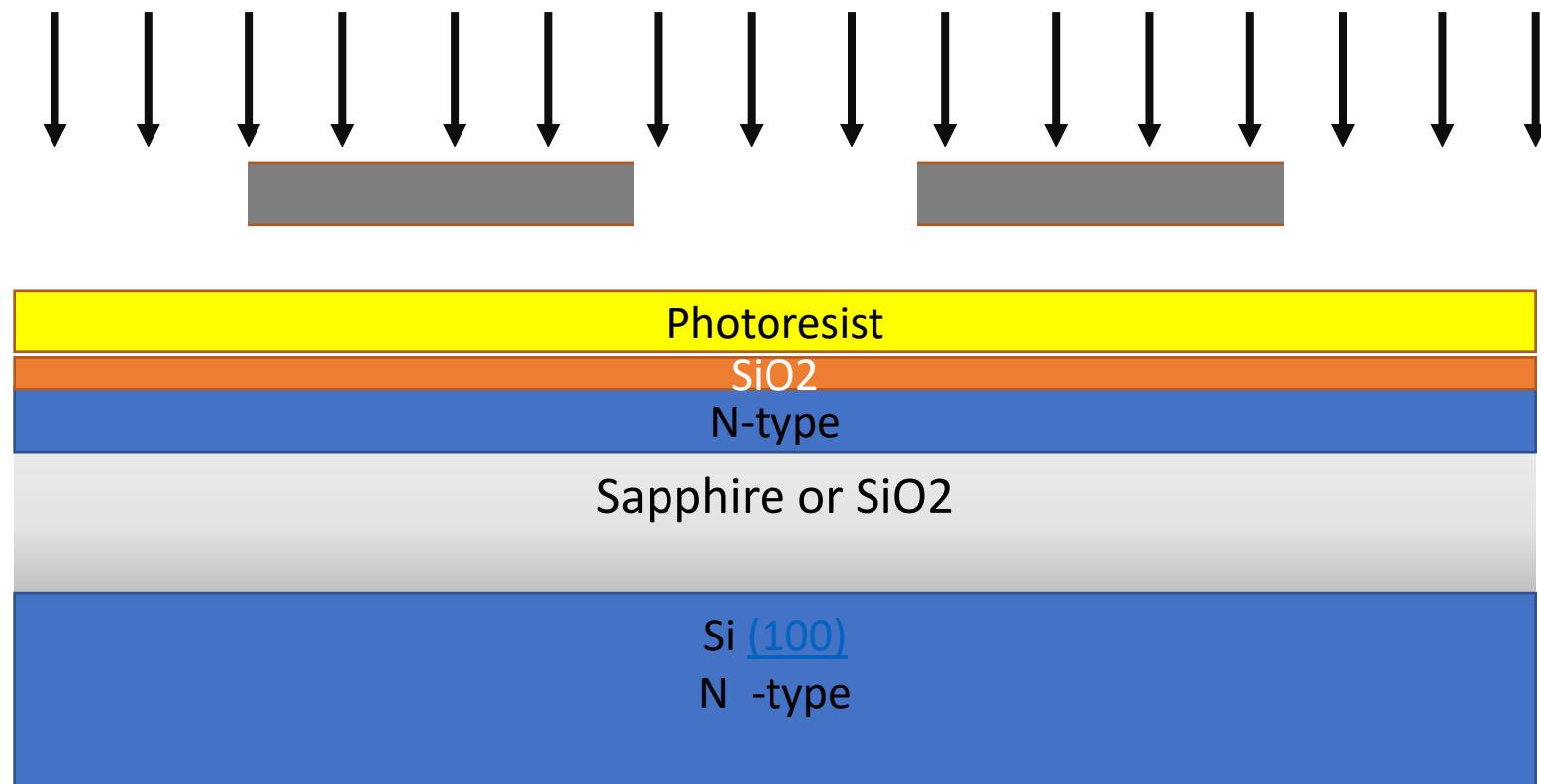
N-type
Sapphire or SiO₂
Si [\(100\)](#)
N -type

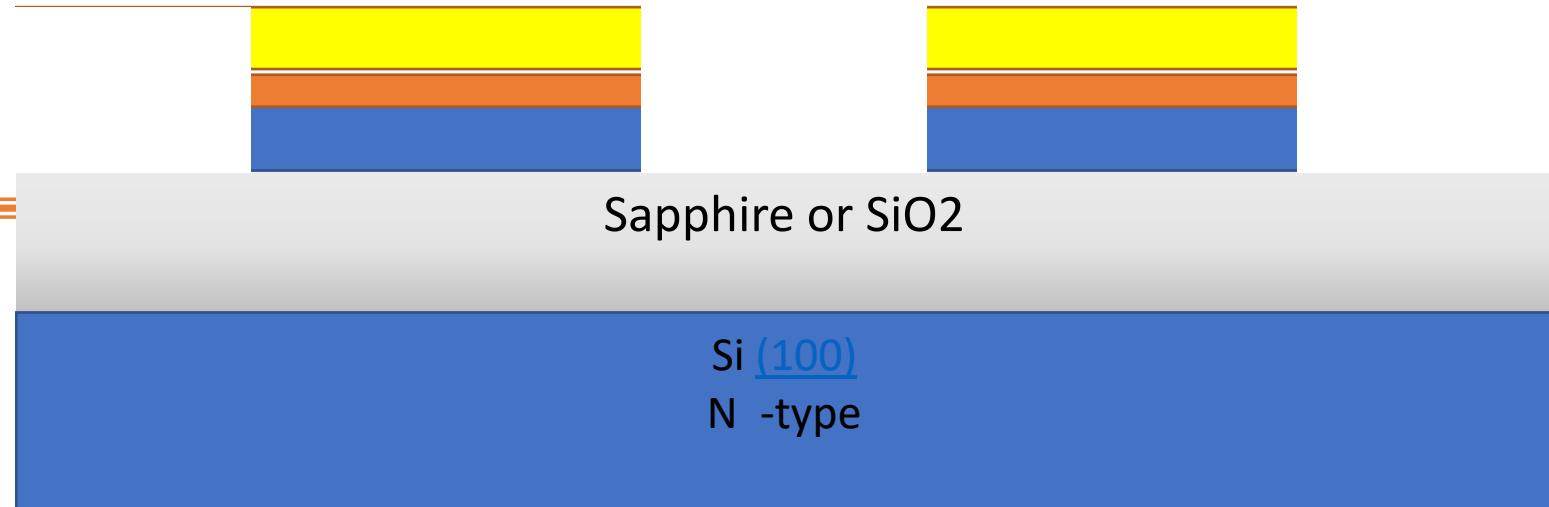


4. Oxide

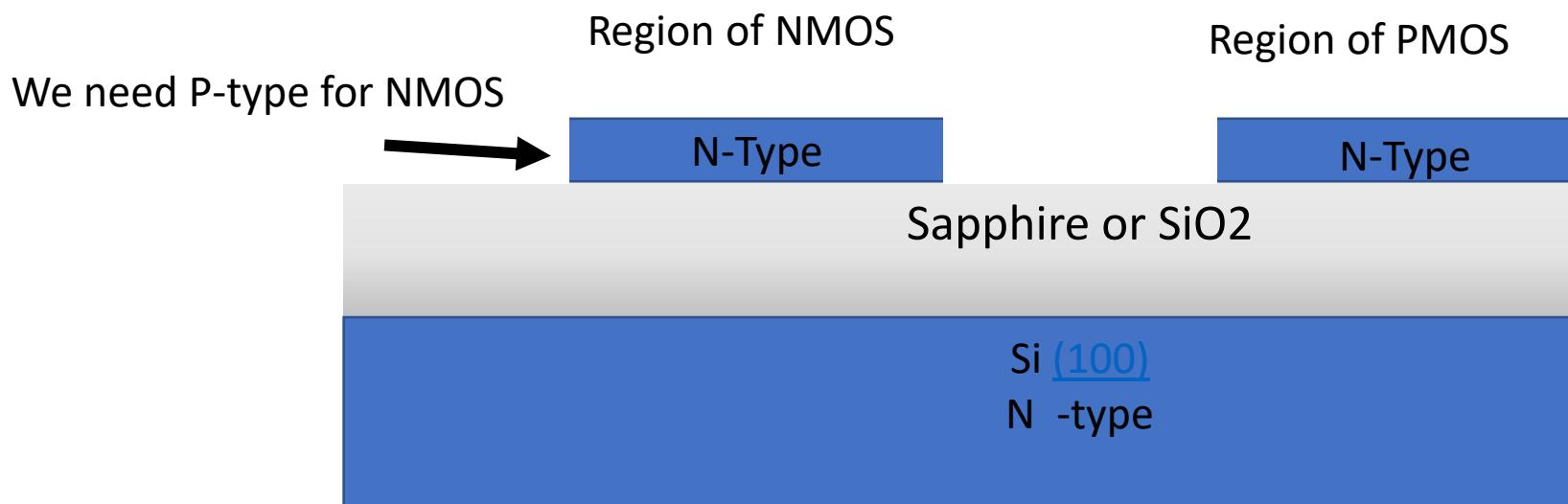


5. Lithography



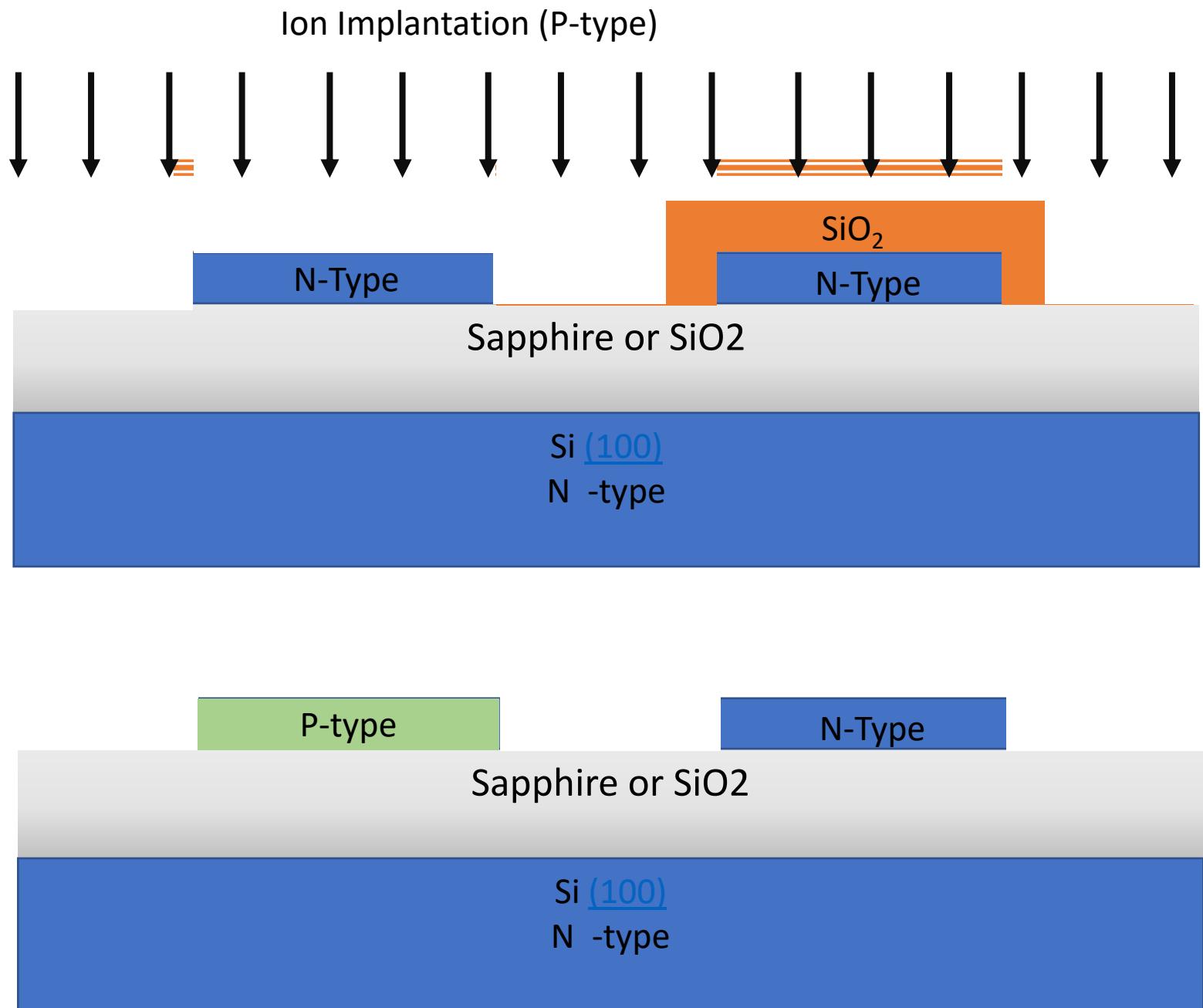


6. Patterning



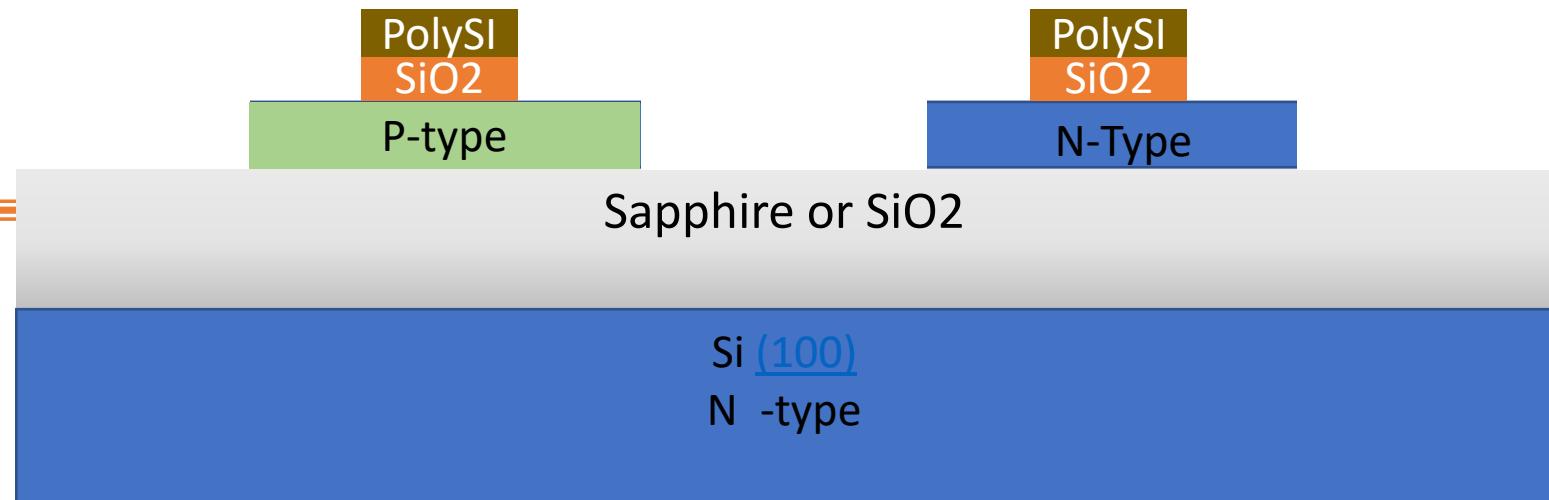


7. Ion implantation for P type impurities

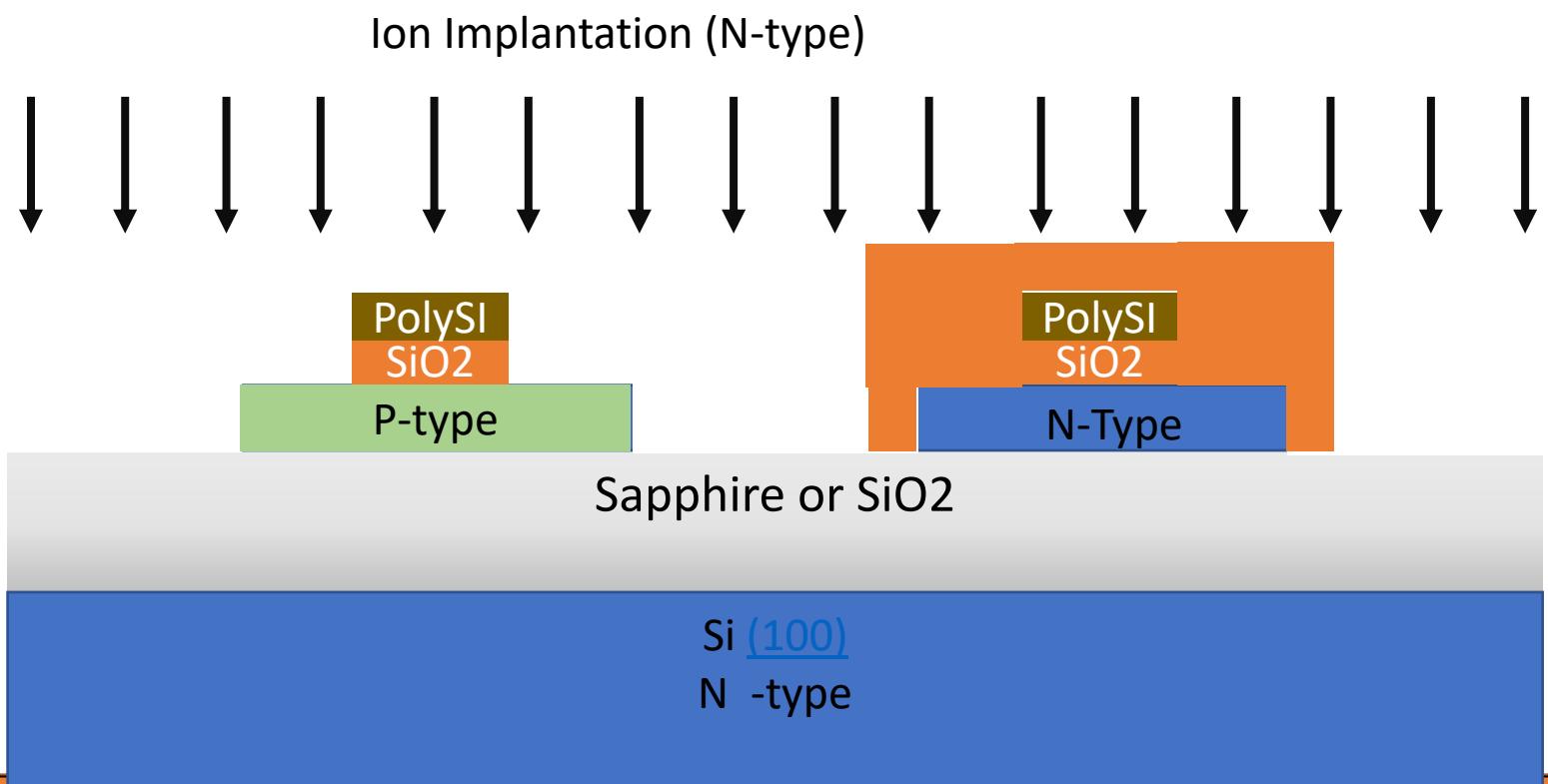




8. Gate terminal

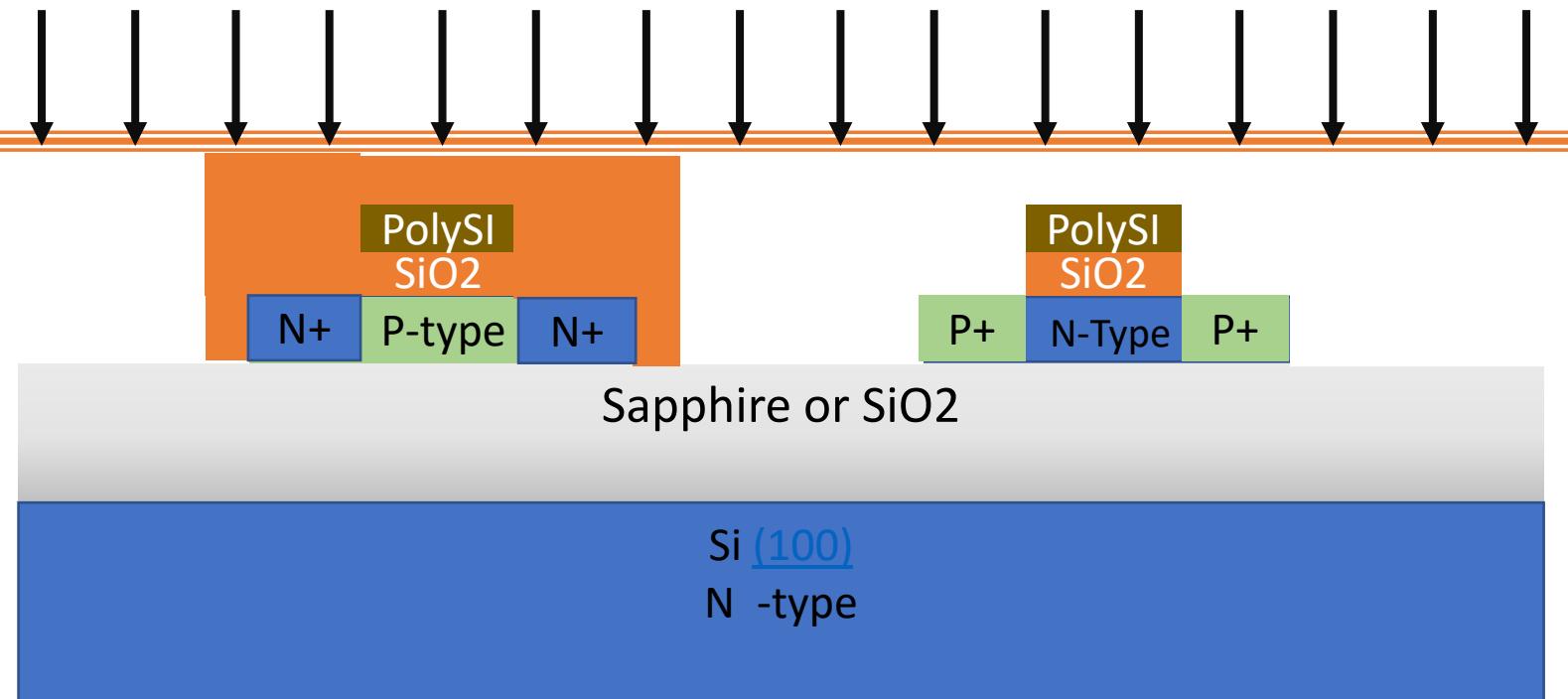


9. Ion Implantation for N+ diffusion region

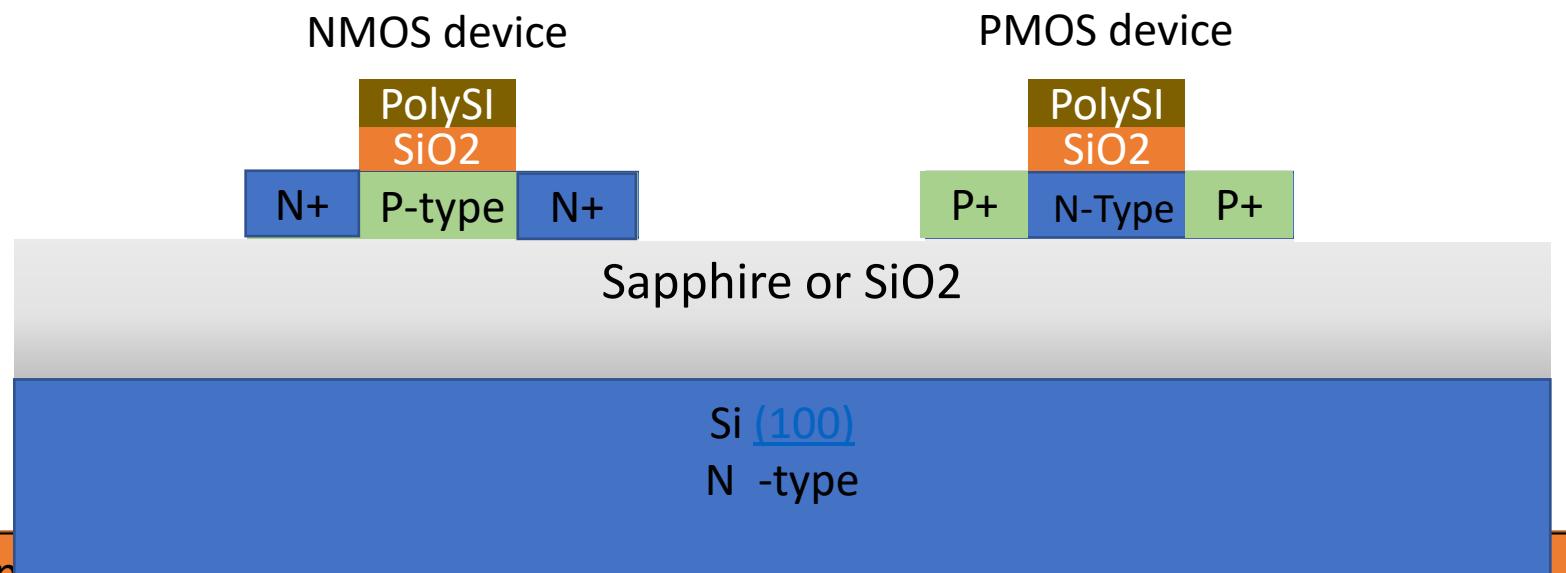




Ion Implantation (P-type)



10. Ion Implantation for P+ diffusion region





11. Metallization and packaging

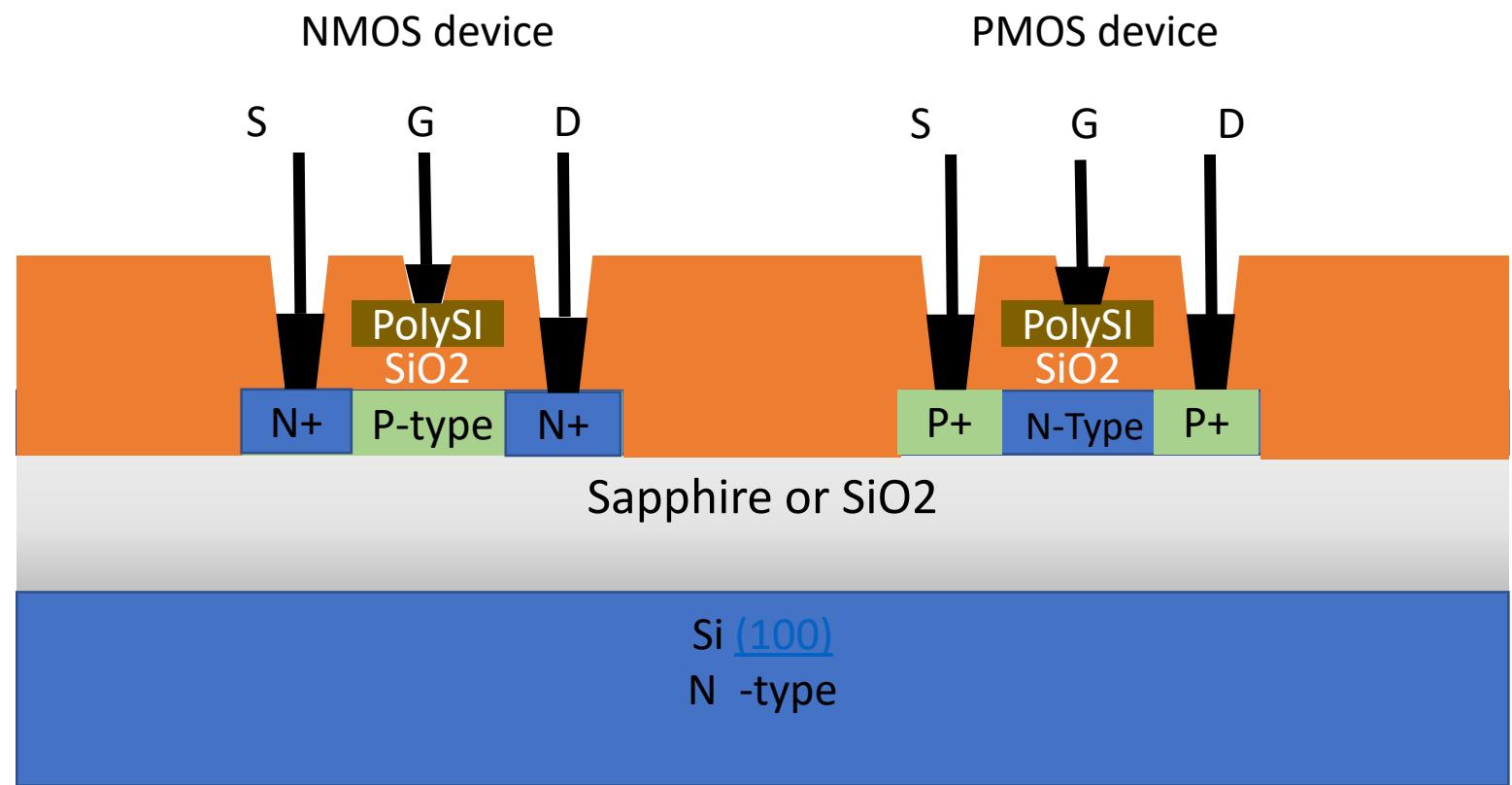
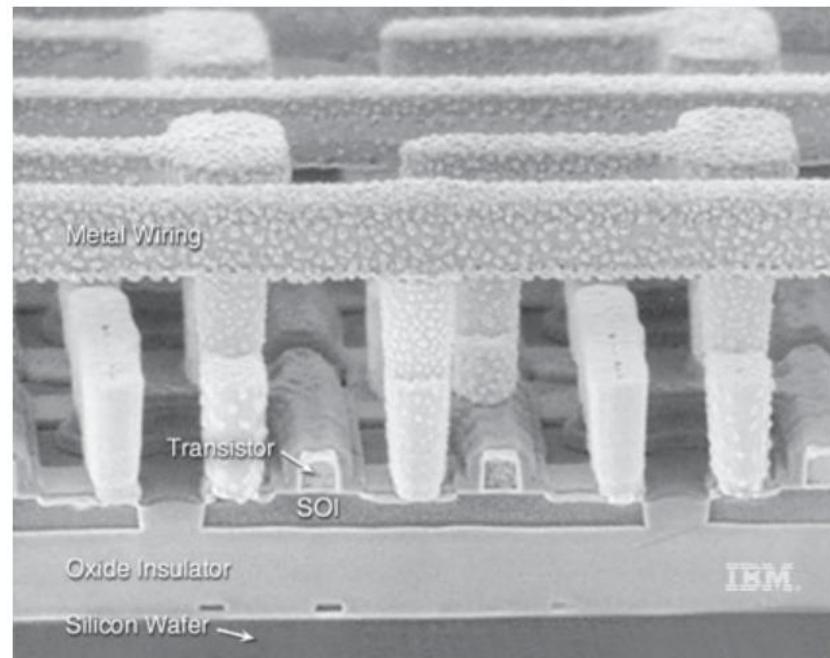


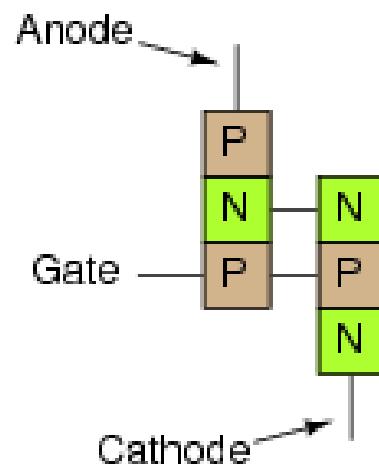
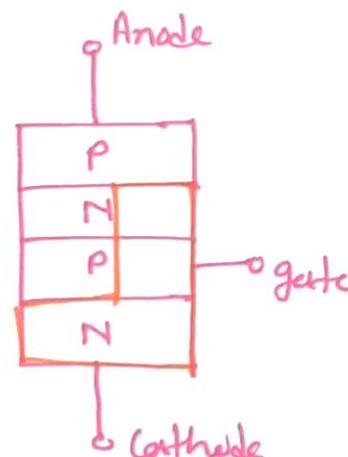
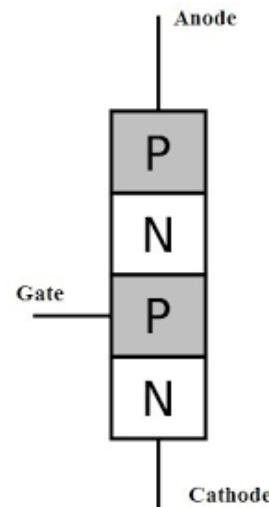
FIGURE 9.62 IBM SOI process electron micrograph
(Courtesy of International Business Machines Corporation.
Unauthorized use not permitted.)

Latch up in CMOS and its prevention

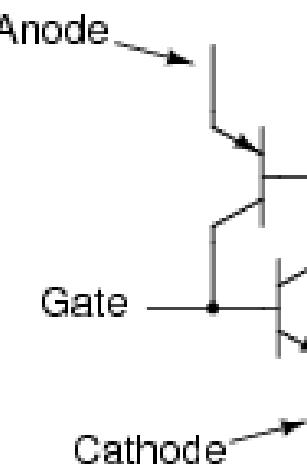
- Tendency of CMOS chips to develop low-resistance paths between V_{DD} and V_{SS} called Latch-up.
- This is possible due to parasitic components.
- This leads to static current flow from V_{DD} to V_{SS}

Silicon Controlled Rectifier (SCR): Applying positive voltage at gate will turn SCR ON.

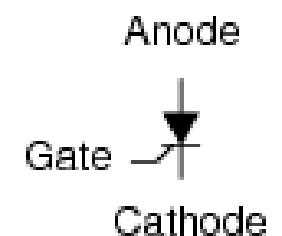
Removing gate voltage doesn't turn SCR OFF. Apply a reverse voltage across the SCR will turn it OFF



Physical diagram

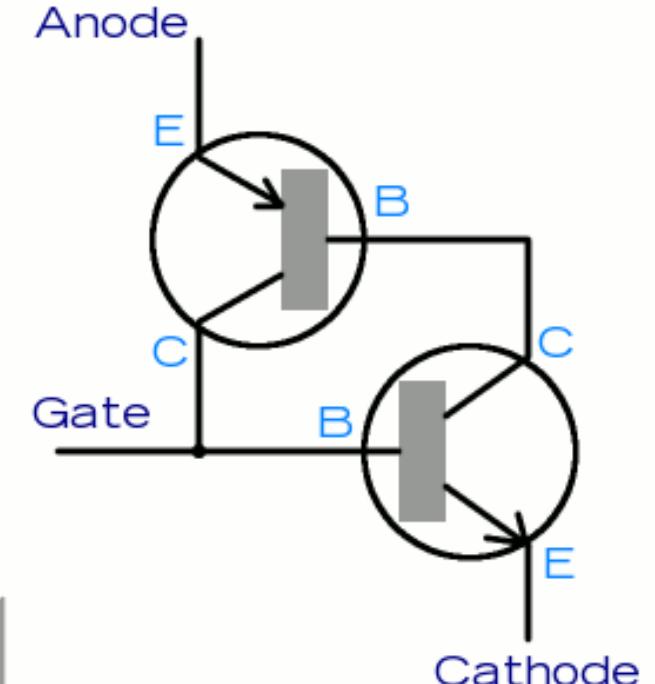


Equivalent schematic



Schematic symbol

1. With no current flowing into the gate, the thyristor is switched off and no current flows between the anode and the cathode.
2. When a current flows into the gate, it effectively flows into the base (input) of the lower (n-p-n) transistor, turning it on.
3. Once the lower transistor is switched on, current can flow through it, activating the base (input) of the upper (p-n-p) transistor, turning that on as well.
4. Once both transistors are turned on completely ("saturated"), current can flow all the way through both of them—through the entire thyristor from the anode to the cathode.
5. Since the two transistors keep one another switched on, the thyristor stays on—"latches"—even if the gate current is removed.



1

- These BJTs form a silicon-controlled rectifier (SCR) with positive feedback and virtually short circuit the power rail to-ground, thus causing excessive current flows and even permanent device damage.
- PNP transistor whose base is formed by the n-well with its base-to-collector current gain (β_1) as high as several hundreds.
- NPN transistor with its base formed by the p-type substrate. The base-to-collector current gain β_2 of this lateral transistor may range from a few tenths to tens

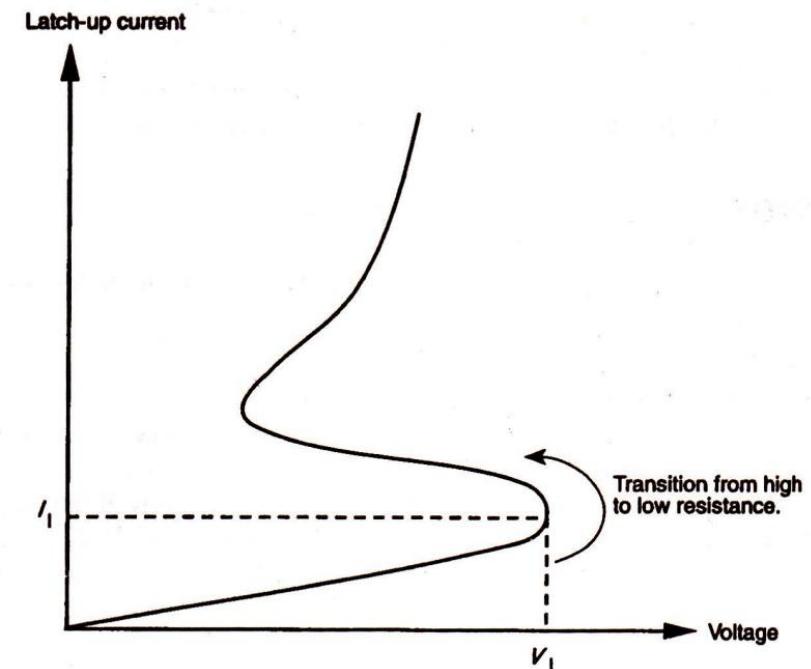
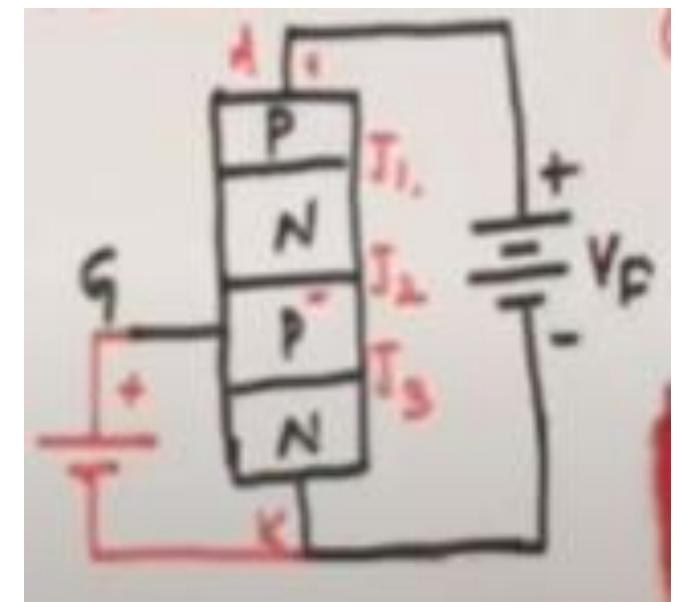
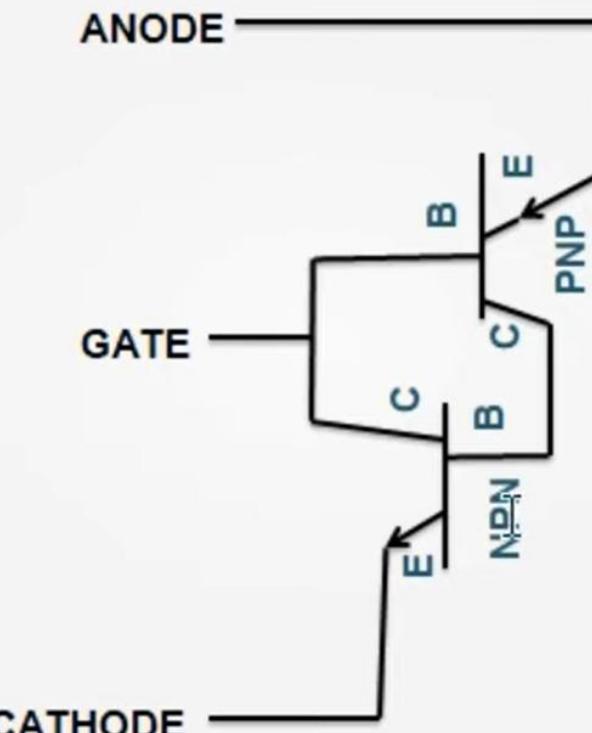
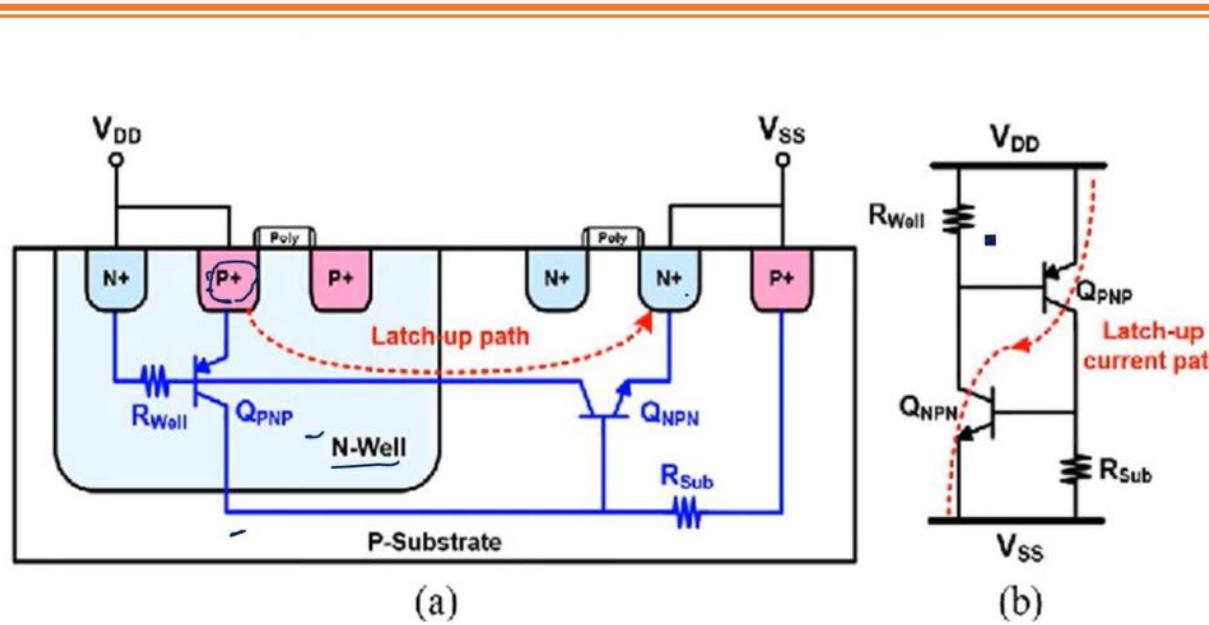
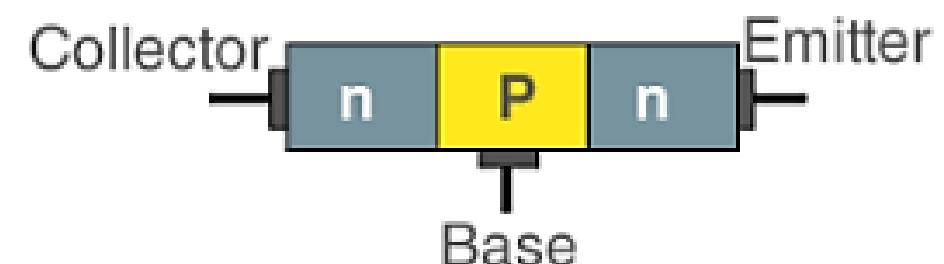
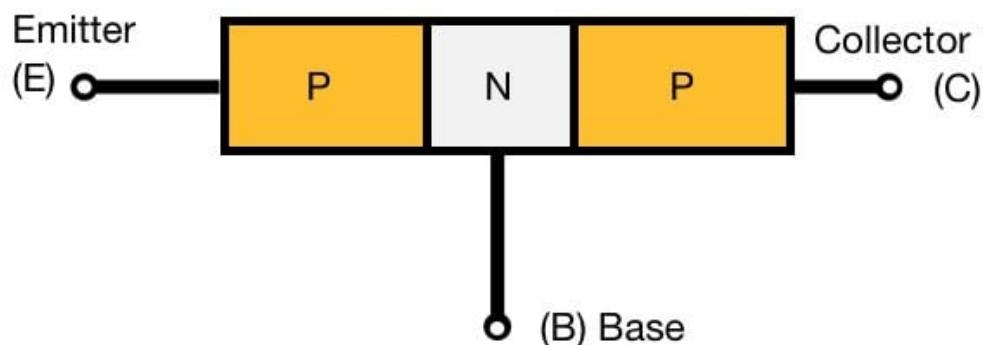


FIGURE 2.23 Latch-up current versus voltage.





Silicon Controlled Rectifier





Latch-up

- R_{well} represents the parasitic resistance in the n-well structure with its value ranging from $1\text{ k}\Omega$ to $20\text{ k}\Omega$.
- R_{sub} can be as high as several hundred ohms.
- Unless the SCR is triggered by an external disturbance, the collector currents of both transistors consist of the reverse leakage currents of the collector-base junctions and therefore, their current gains are very low.
- If the collector current of one of the transistors is temporarily increased by an external disturbance, however, the resulting feedback loop causes this current perturbation to be multiplied by $(\beta_1 \beta_2)$. This event is called the *triggering* of the SCR.

Techniques to overcome Latch-up

- Use p+ guard-band rings connected to ground around nMOS transistors and n+ guard rings connected to VDD around pMOS transistors to reduce R_{well} and R_{sub} and to capture injected minority carriers before they reach the base of the parasitic BJTs.
- Every well should have at least one tap.
- Having small values of R_{well} and R_{sub} .
- Place substrate and well contacts as close as possible to the source connections of MOS transistors to reduce the values of R_{well} and R_{sub} .

SOI Devices

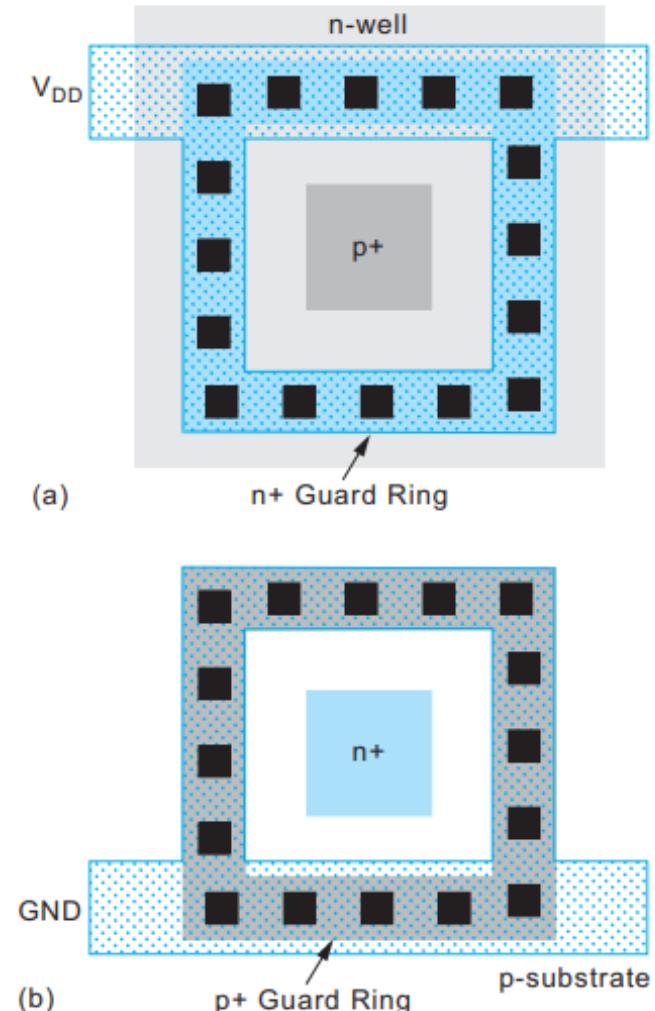


FIGURE 7.13 Guard rings



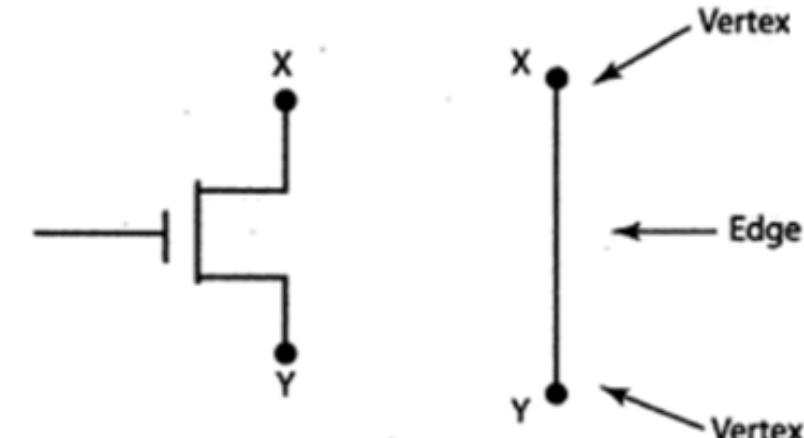
Euler's path, Stick diagrams and Layouts



Euler's path

- An uninterrupted diffusion strip is possible only if there exists a Euler path in the logic graph

- Diffusion without break.
- Common Polysilicon running between P and N diffusion layers.
- Less Number of Contacts.



Euler path: A path through all nodes in the graph such that each edge is visited once and only once, Nodes can be repeated.



Stick diagrams

- We have seen that MOS circuits are formed on four basic layers-**n-diffusion**, **p-diffusion**, **polysilicon**, and **metal**, which are isolated from one another by thick or thin (*thinox*) silicon dioxide insulating layers.
- Polysilicon and *thinox* regions interact so that a transistor is formed where they cross the diffusion layer.
- In some processes, there may be a second metal layer and also, in some processes, a second polysilicon layer.
- Layers may deliberately joined together where contacts are formed.



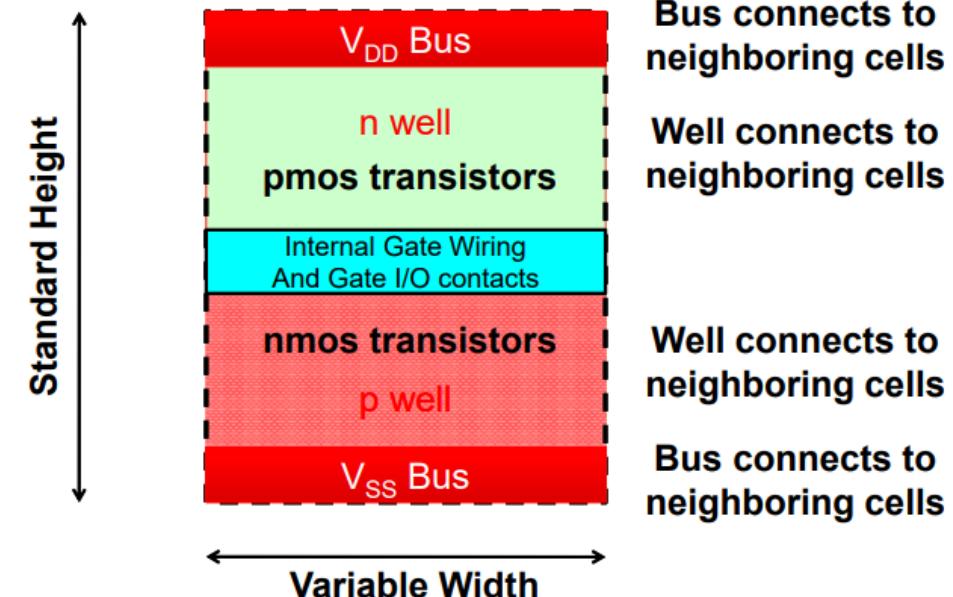
Stick diagrams

Stick diagrams may be used to convey layer information through the use of a color code

For example, in the case of nMOS design, green for n-diffusion, yellow for p-diffusion, red for polysilicon, blue for metal, yellow for implant, and black for contact areas.

- VLSI design aims to translate circuit concepts onto silicon.
- stick diagrams are a means of capturing topography and layer information using simple diagrams.
- Stick diagrams convey layer information through colour codes (or monochrome encoding).
- Acts as an interface between symbolic circuit and the actual layout.

Standard Cell Layout

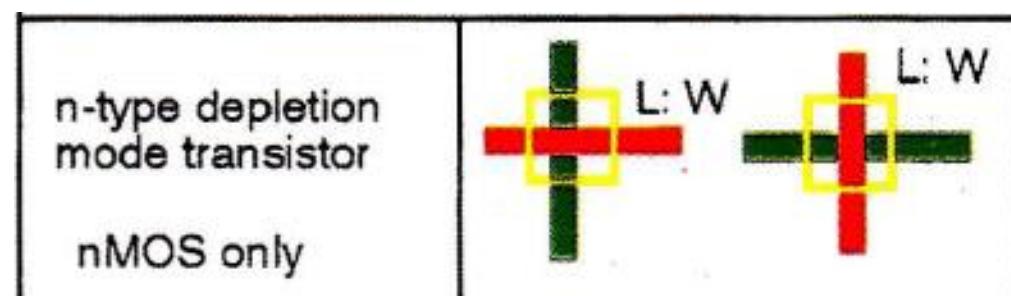
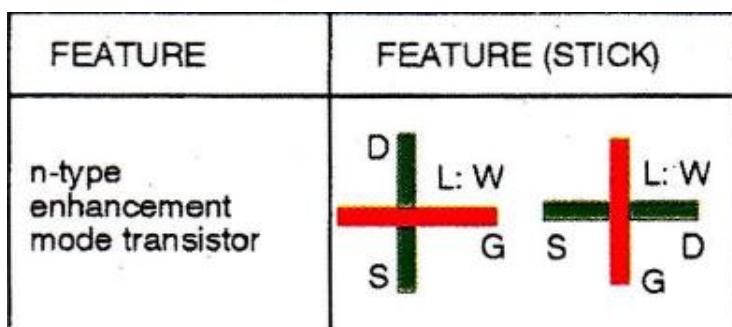




Stick diagrams

COLOR	STICK ENCODING	LAYERS
GREEN		n-diffusion (n+ active) Thinox*
RED		Polysilicon
BLUE		Metal 1
BLACK		Contact cut
GRAY	NOT APPLICABLE	Overglass

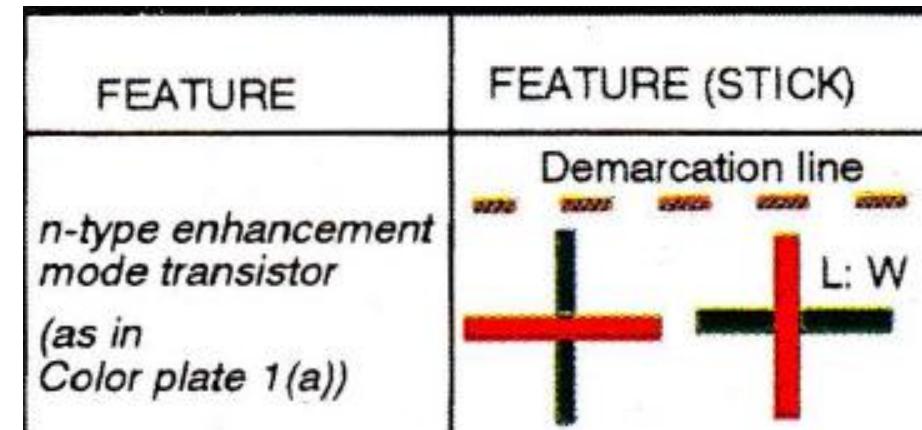
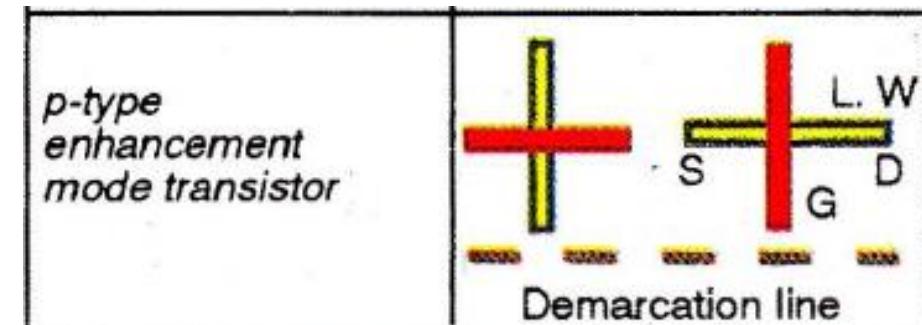
nMOS ONLY YELLOW		Implant
nMOS ONLY BROWN		Buried contact





Stick diagrams

YELLOW (STICK)		p-diffusion (p+ active)
YELLOW		p+ mask
DARK BLUE OR PURPLE		Metal 2
BLACK		VIA
BROWN		p-well
BLACK		V_{DD} or V_{SS} contact

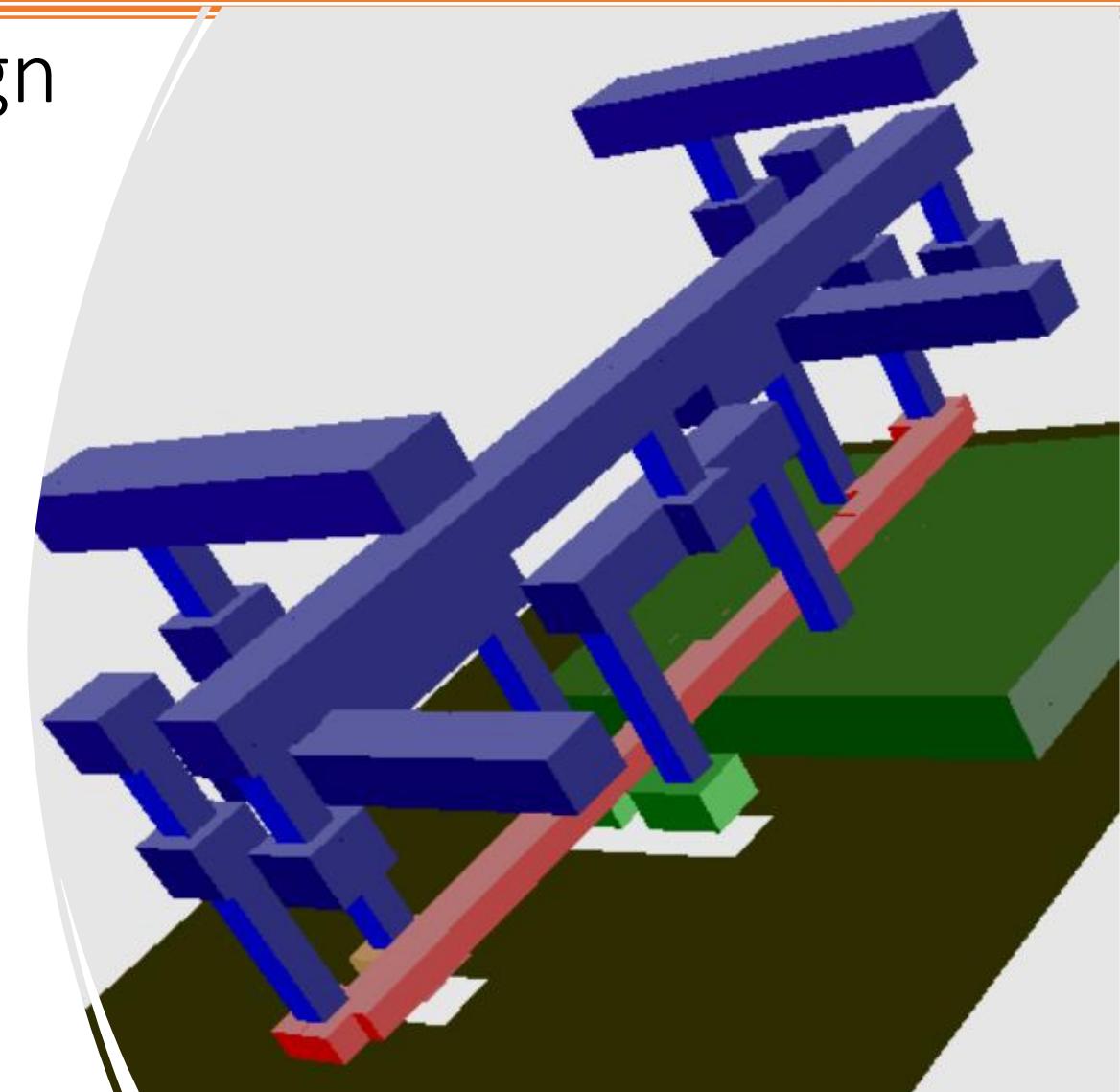




Layouts and Lambda based design rules

Importance of Lambda based design rules

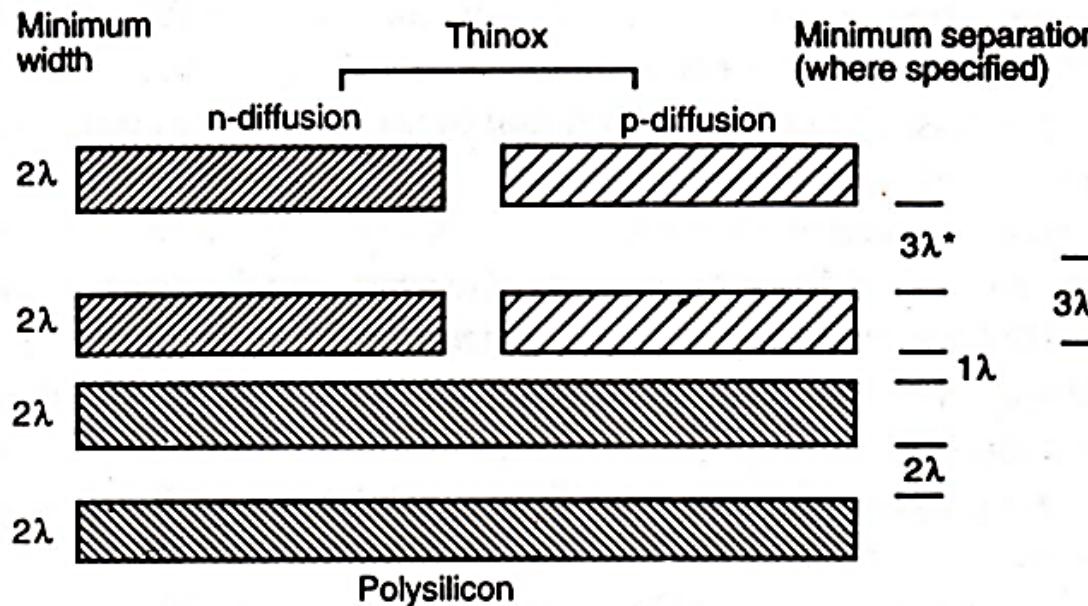
- To allow for shape contraction.
- To ensure adequate continuity of the intervening materials.
- To avoid the possibility of metal-metal or polySi-polySi regions overlapping and conducting currents.
- To prevent the lines overlapping to form unwanted capacitor.
- Feature size refers to minimum transistor length, so lambda is half the feature size.





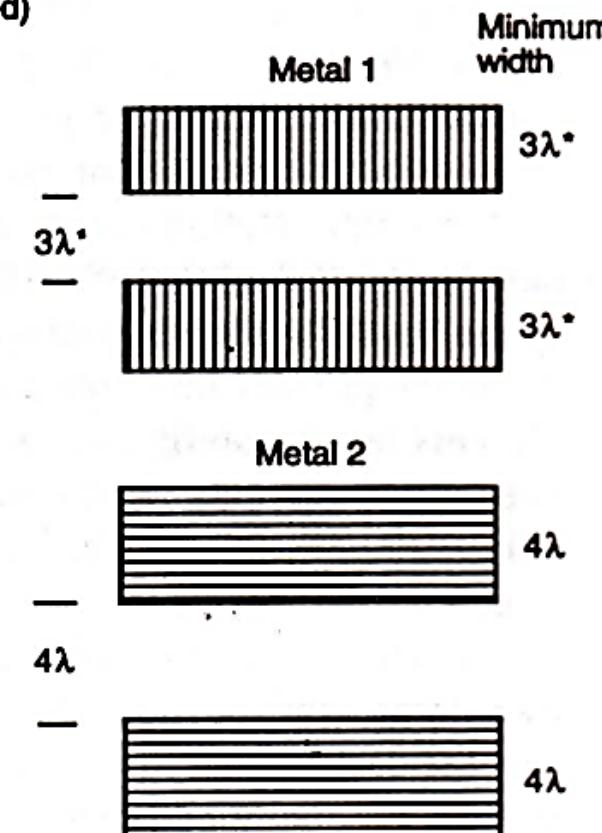
1. Design rules for wires (nMOS and CMOS)

Key: Polysilicon n-diffusion p-diffusion Transistor channel (polysilicon over thinox)



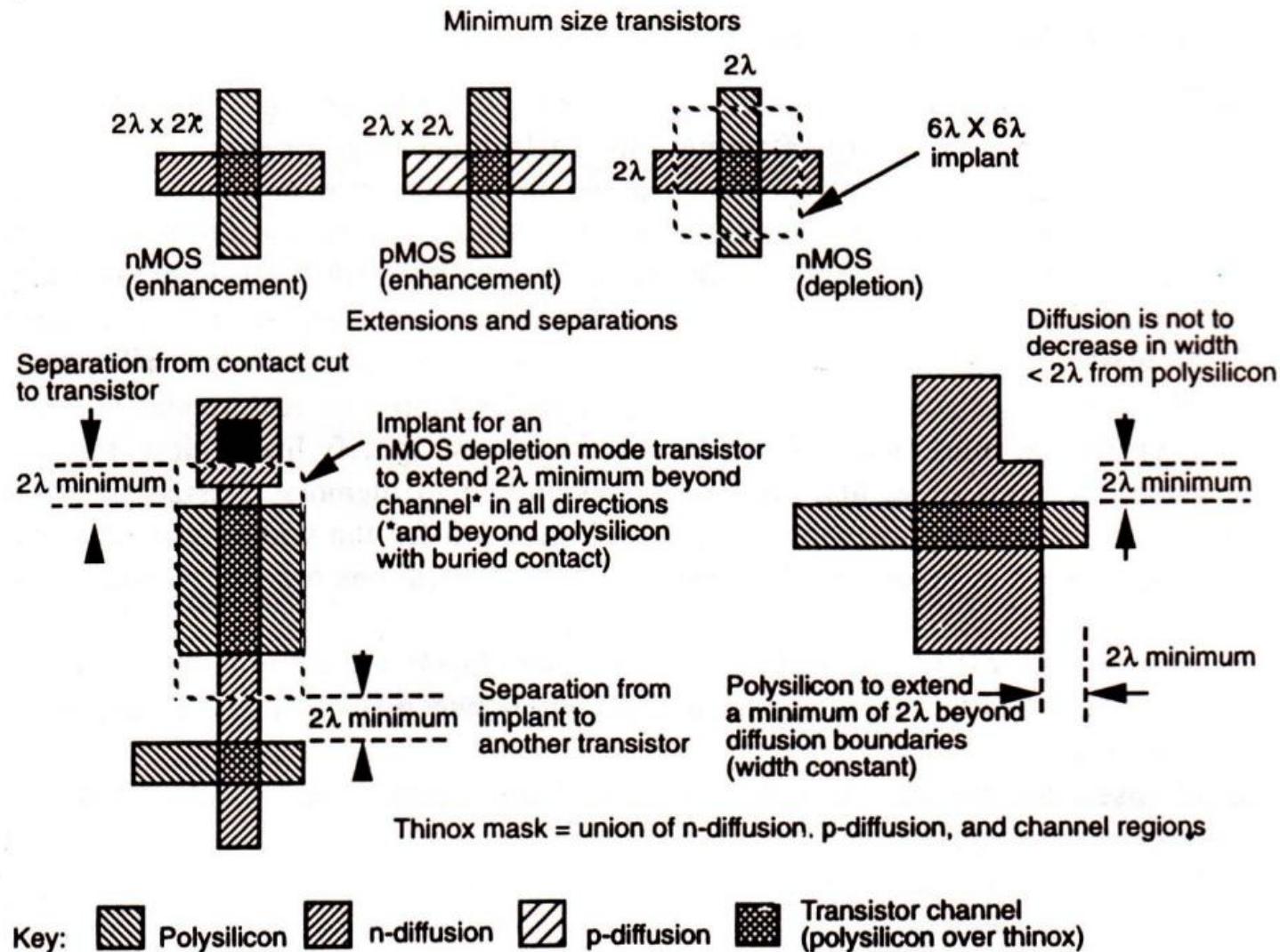
Where no separation is specified, wires may overlap or cross (e.g. metal is not constrained by any other layer). For p-well CMOS, note that n-diffusion wires can only exist inside and p-diffusion wires outside the p-well.

*Note: Many fabrication houses now accept 2λ diffusion to diffusion separation and 2λ metal 1 width and separation.



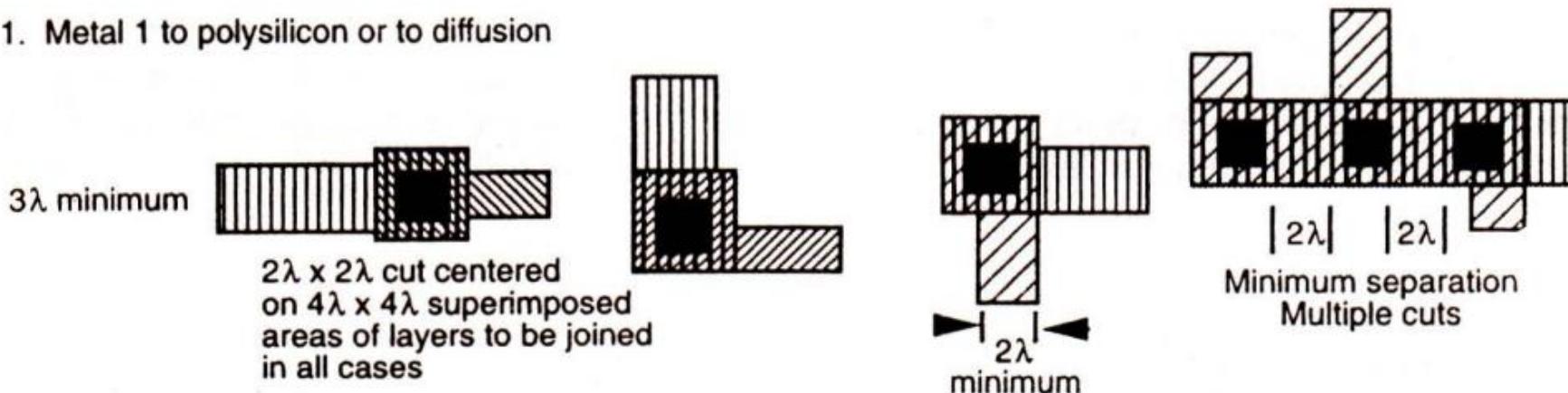


2. Transistor design rules (nMOS, pMOS and CMOS).

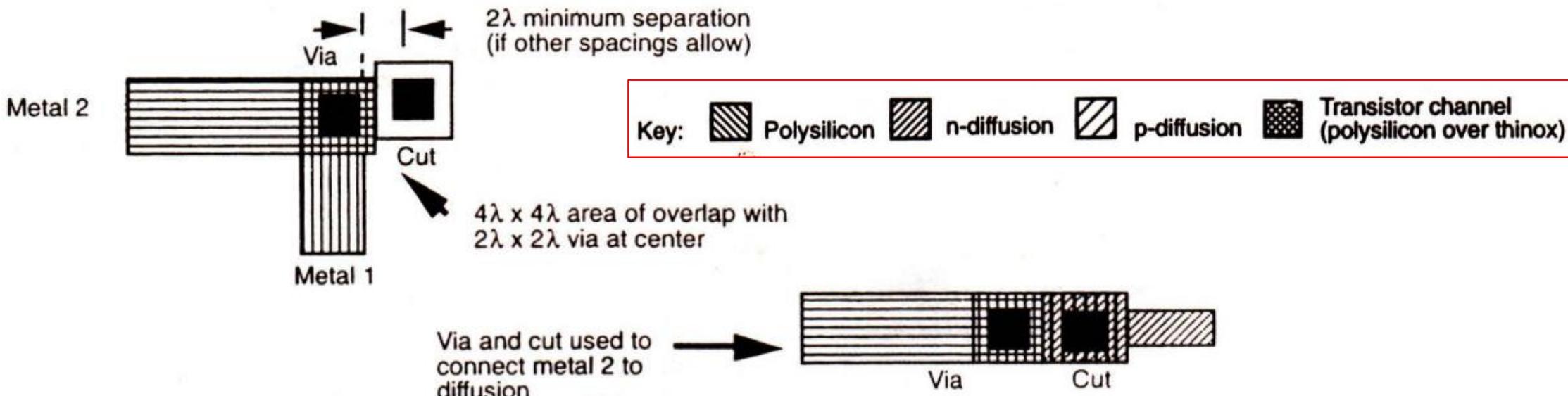


3. Contacts (nMOS and CMOS).

1. Metal 1 to polysilicon or to diffusion

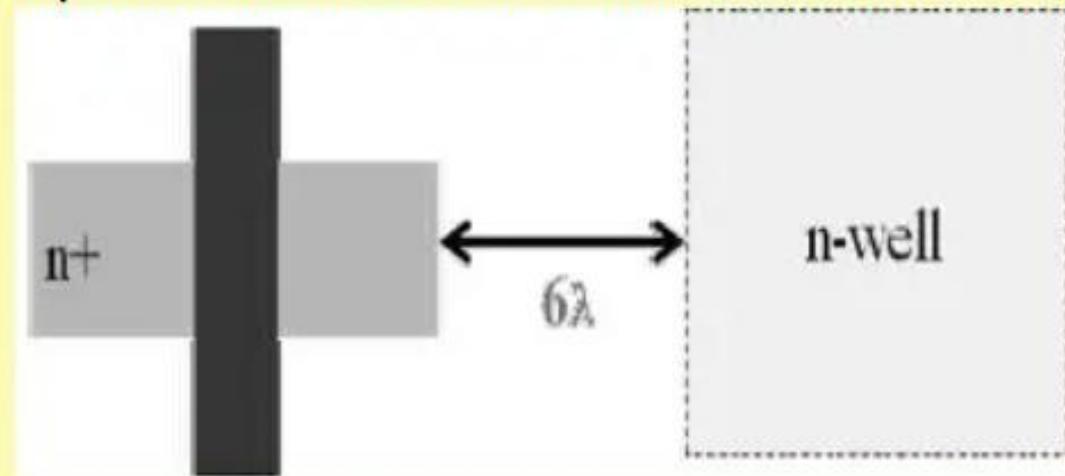


2. Via (contact from metal 2 to metal 1 and thence to other layers)



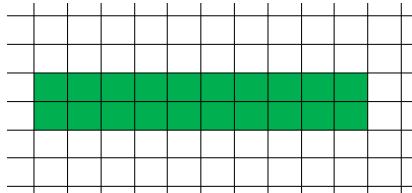
4. N-well rule

- i. To ensure the separation of the PMOS and NMOS devices, n-well supporting PMOS is 6λ away from the active area of NMOS transistor.
- ii. N-well must completely surround the PMOS device active area by 2λ
- iii. The threshold implant mask covers all n-well and surrounds the n-well by λ

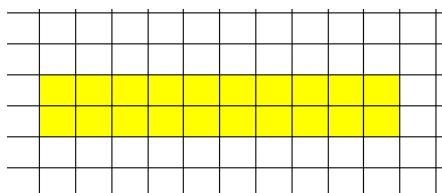




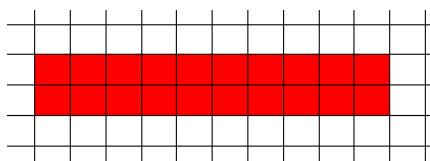
Recommendation: Use graphical page to draw layout



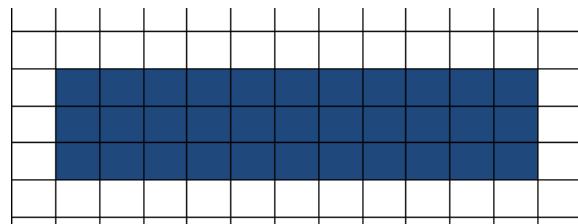
N⁺ Diffusion



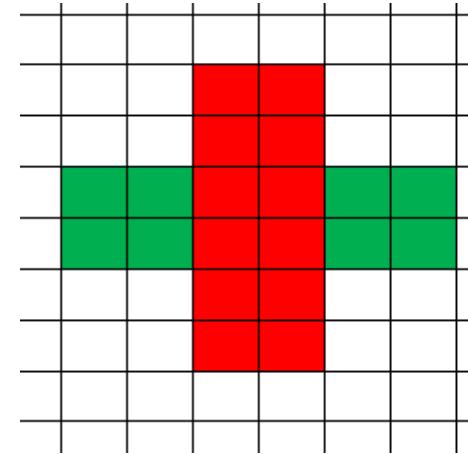
P⁺ Diffusion



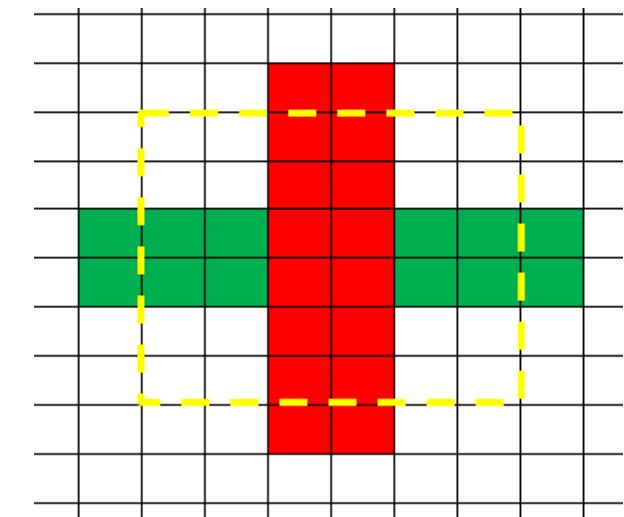
Polysilicon



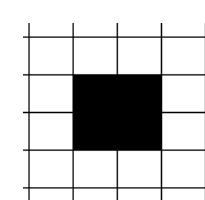
Metal



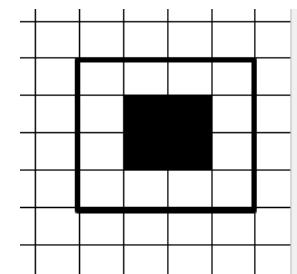
nMOS
(Enhancement) FET



nMOS (Depletion) FET



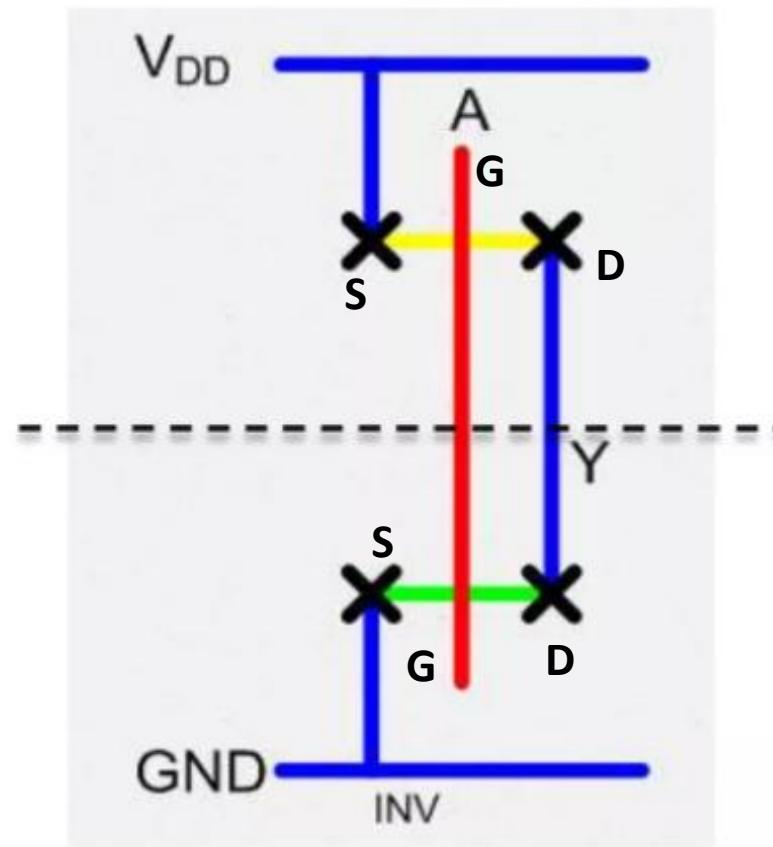
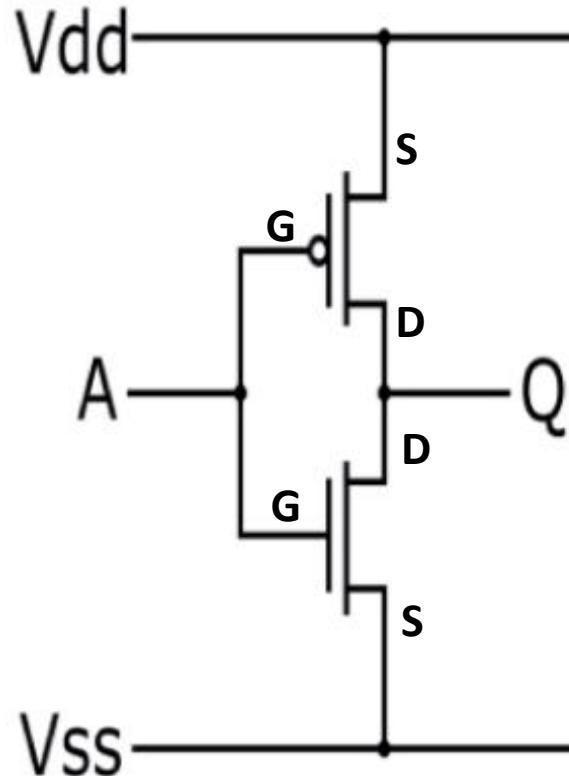
Contact cut



Contact pad

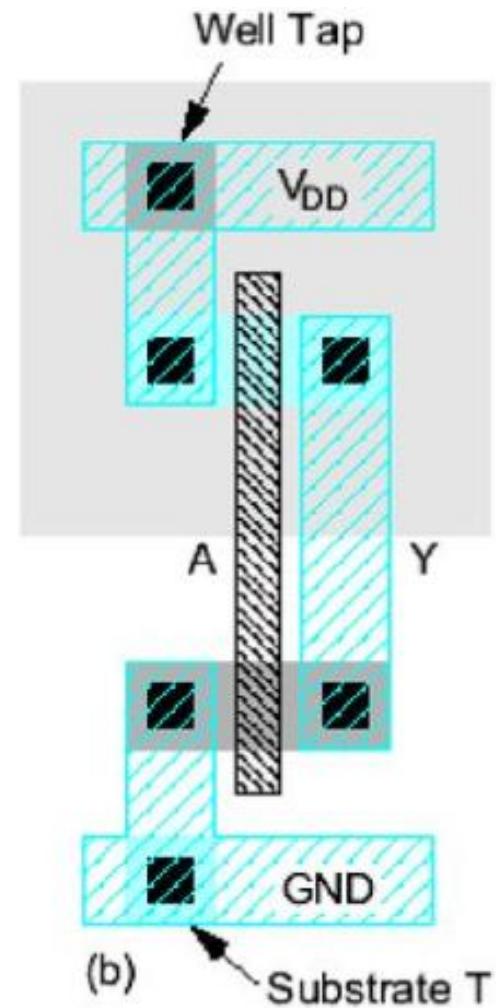


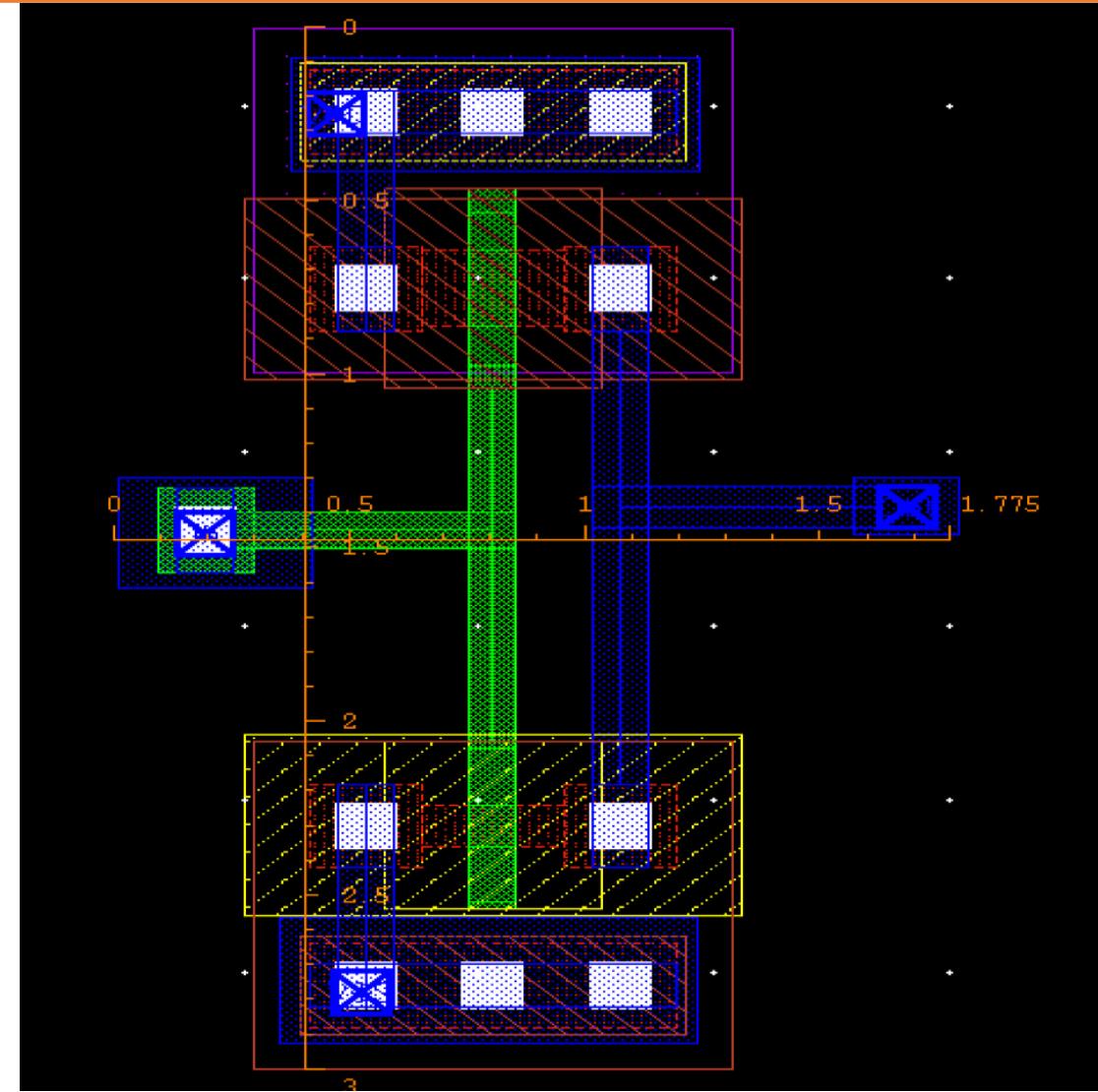
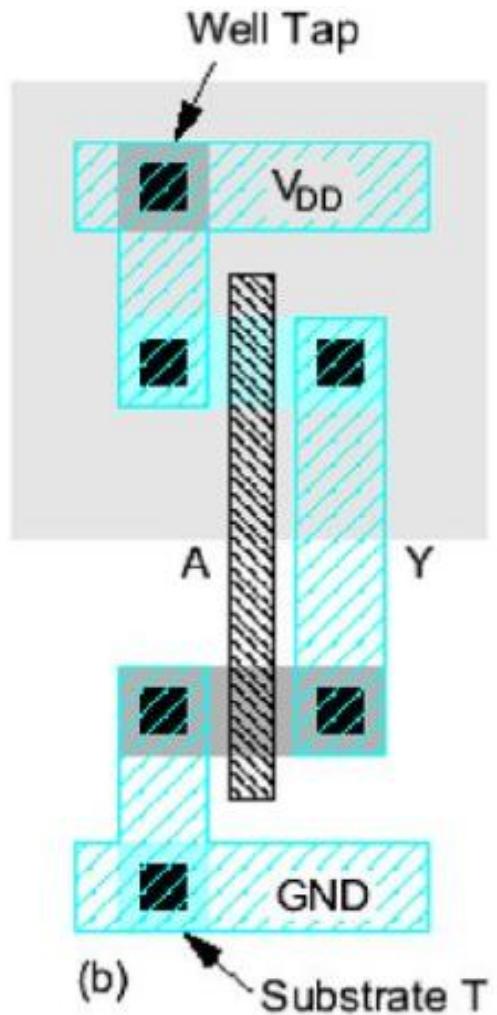
Draw the stick diagram and layout of CMOS inverter.

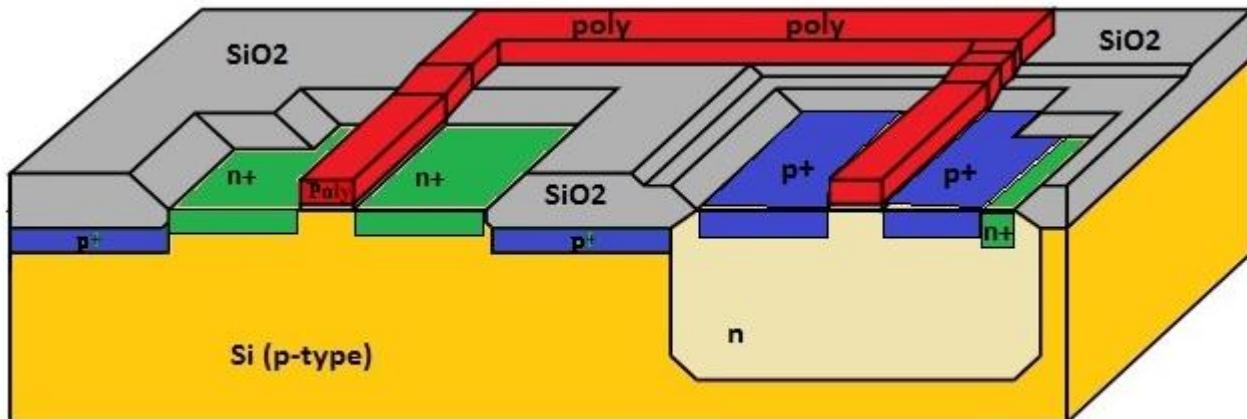


□ Metal
□ Polysilicon
□ Metal Contact

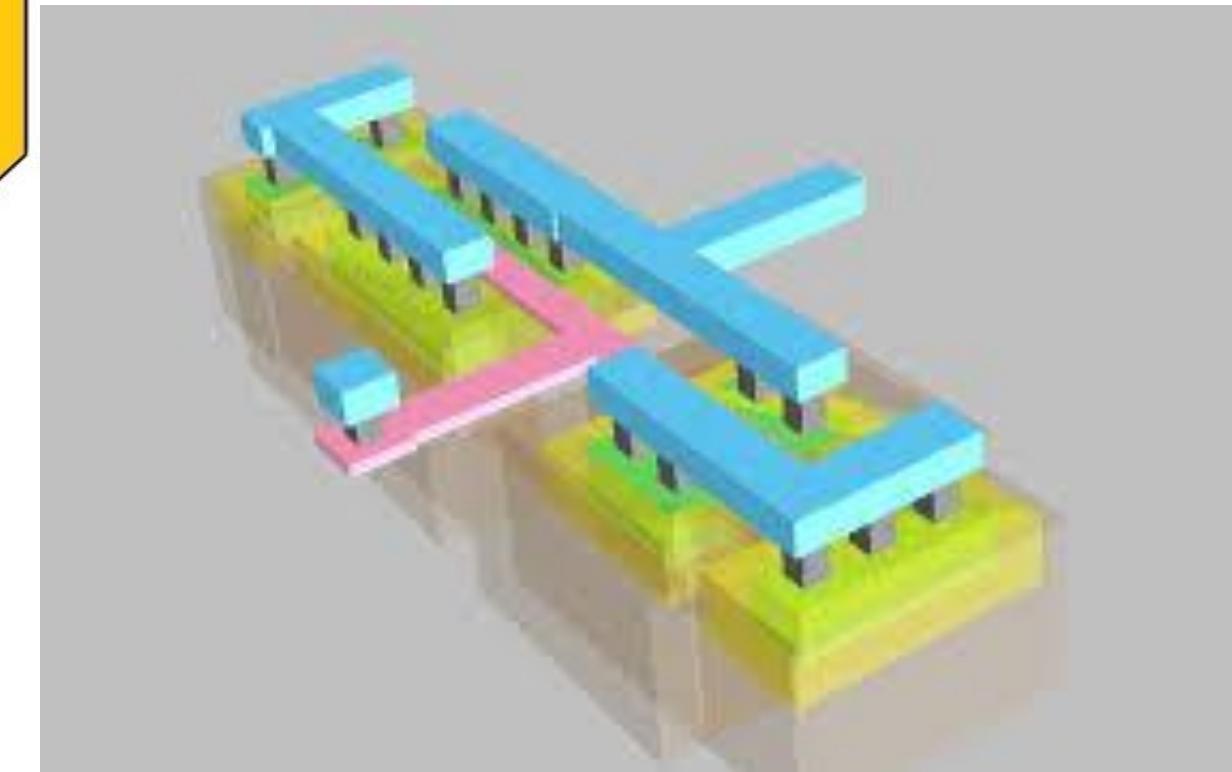
□ P-Doping
□ N-Doping
Demarcation line





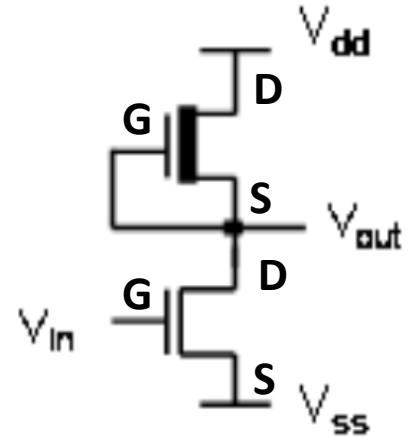


3D view of silicon wafer. Now we have PMOS and NMOS Device.

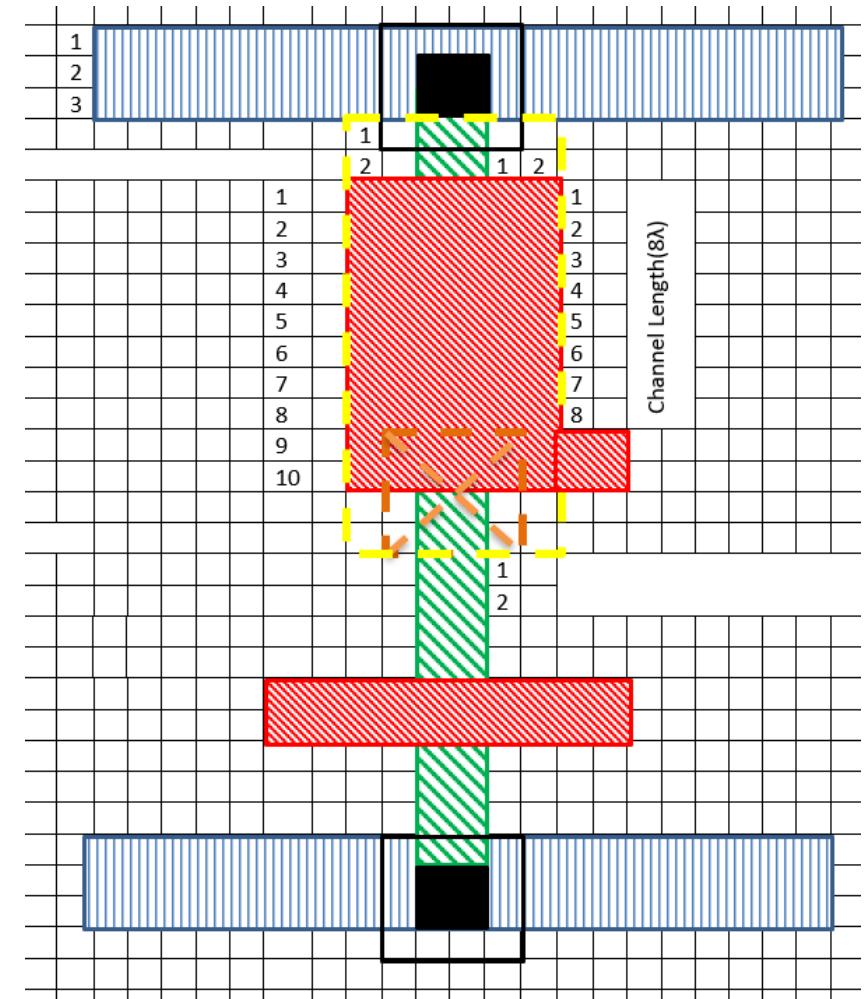
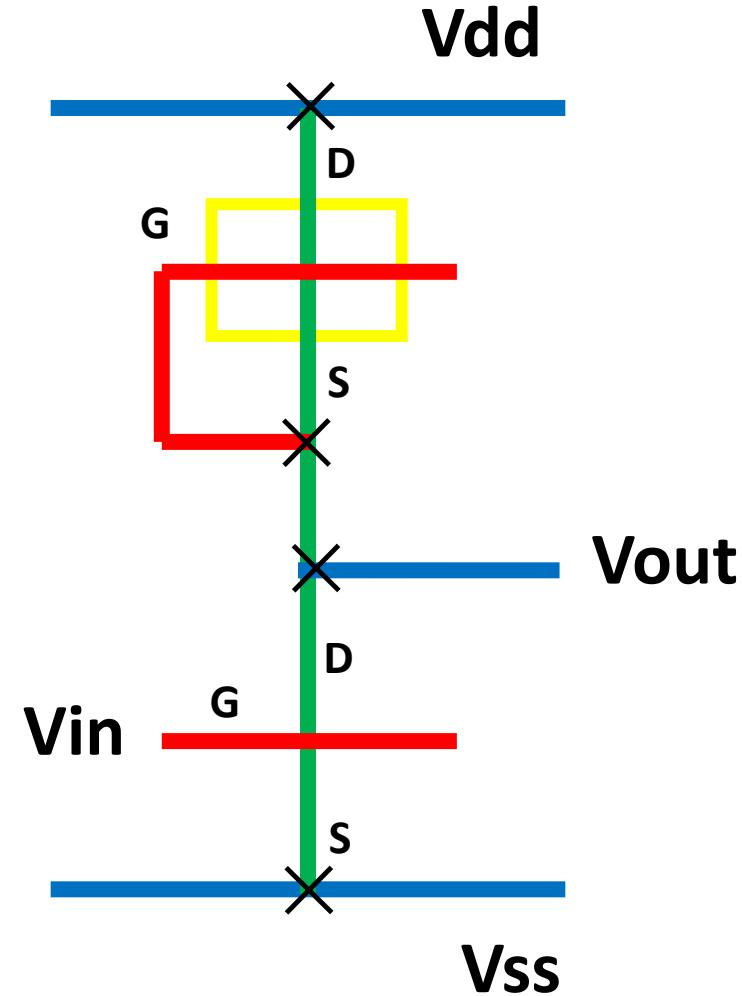




Draw the stick diagram and layout of nMOS inverter with depletion load.

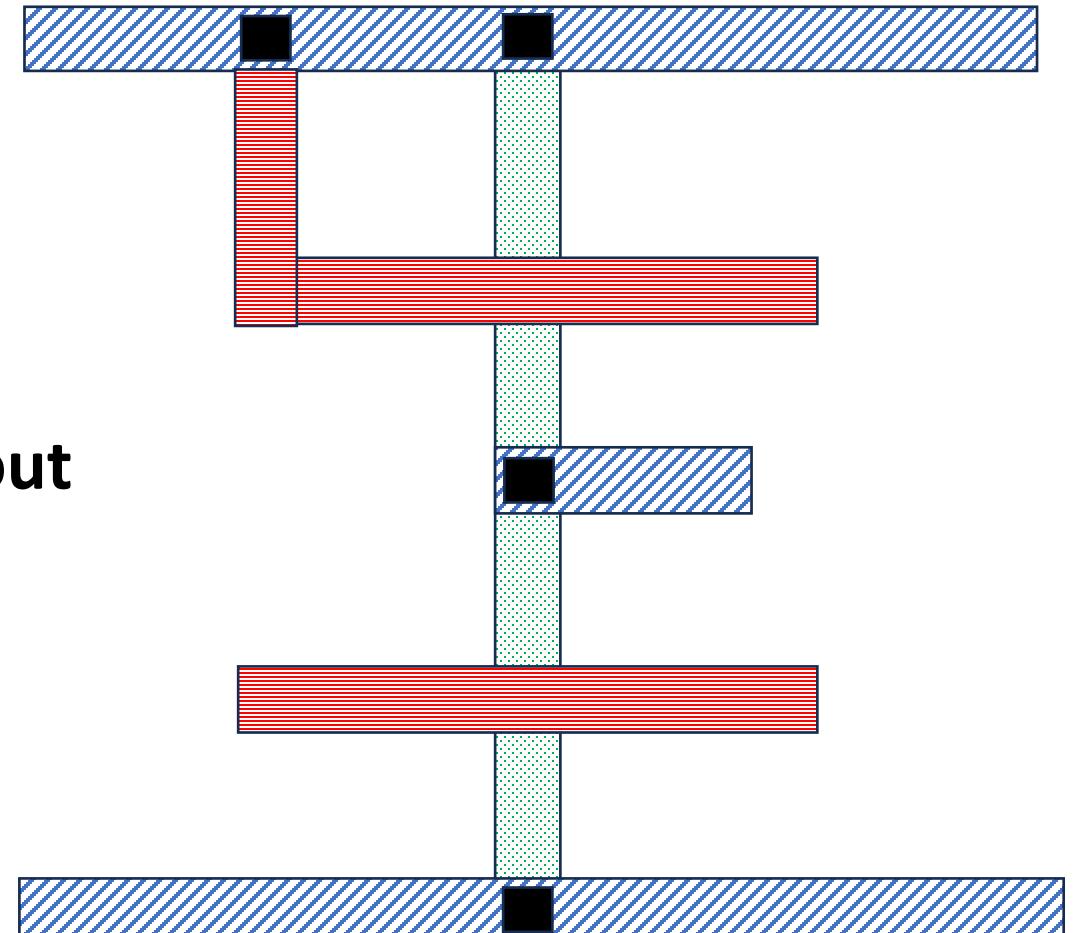
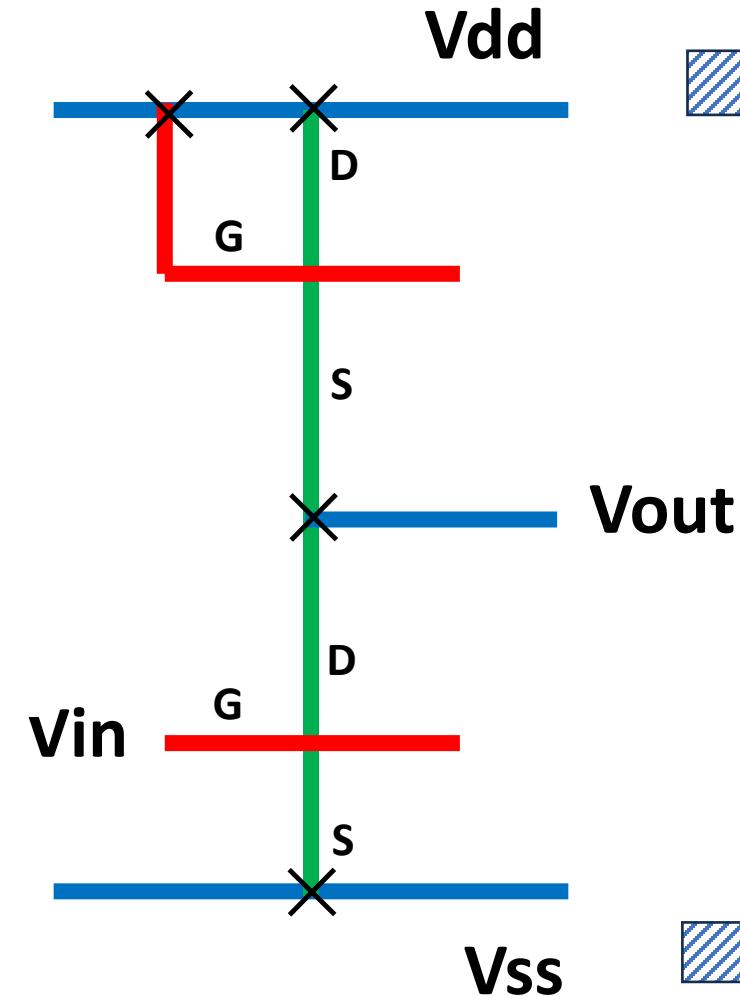
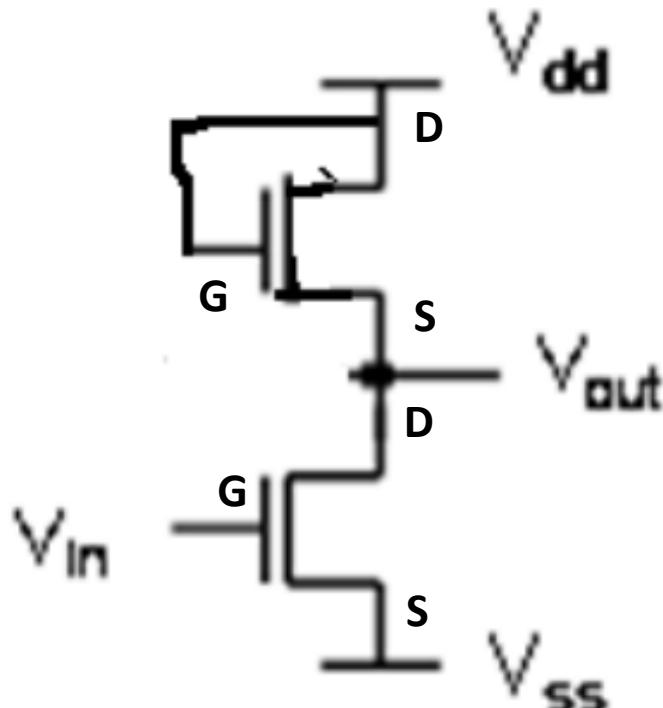


- Metal
- Polysilicon
- Metal Contact
- P-Doping
- N-Doping





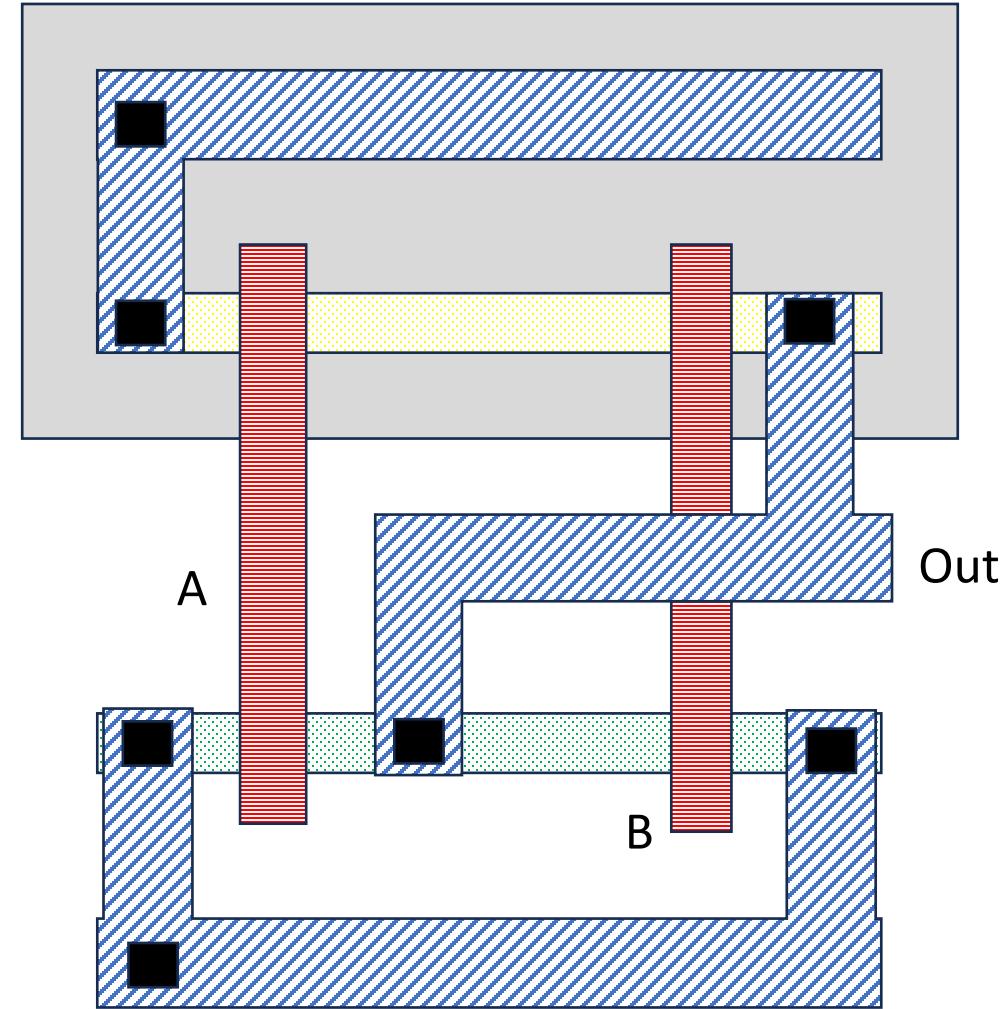
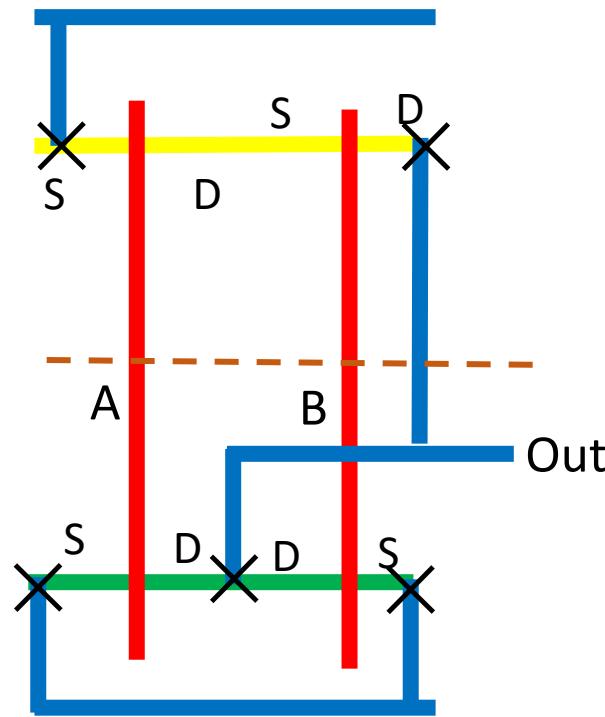
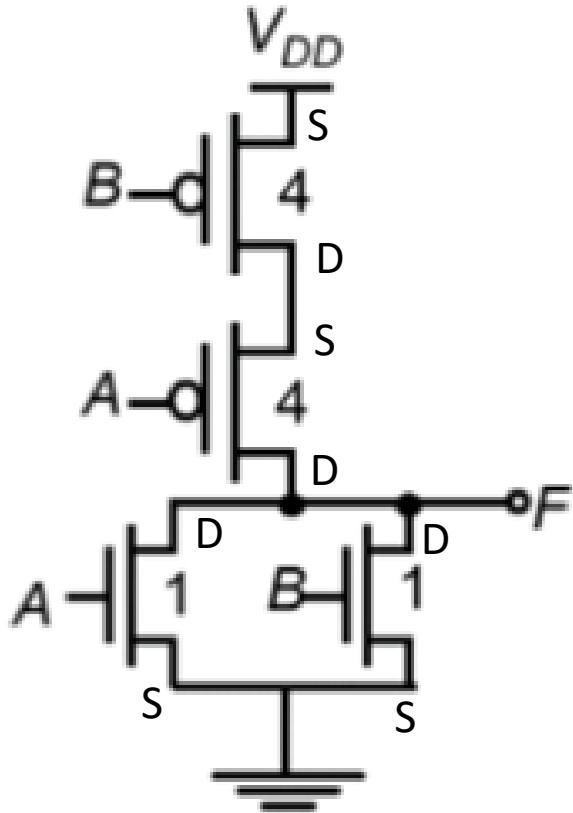
Draw the stick diagram and layout of nMOS inverter with enhancement load.

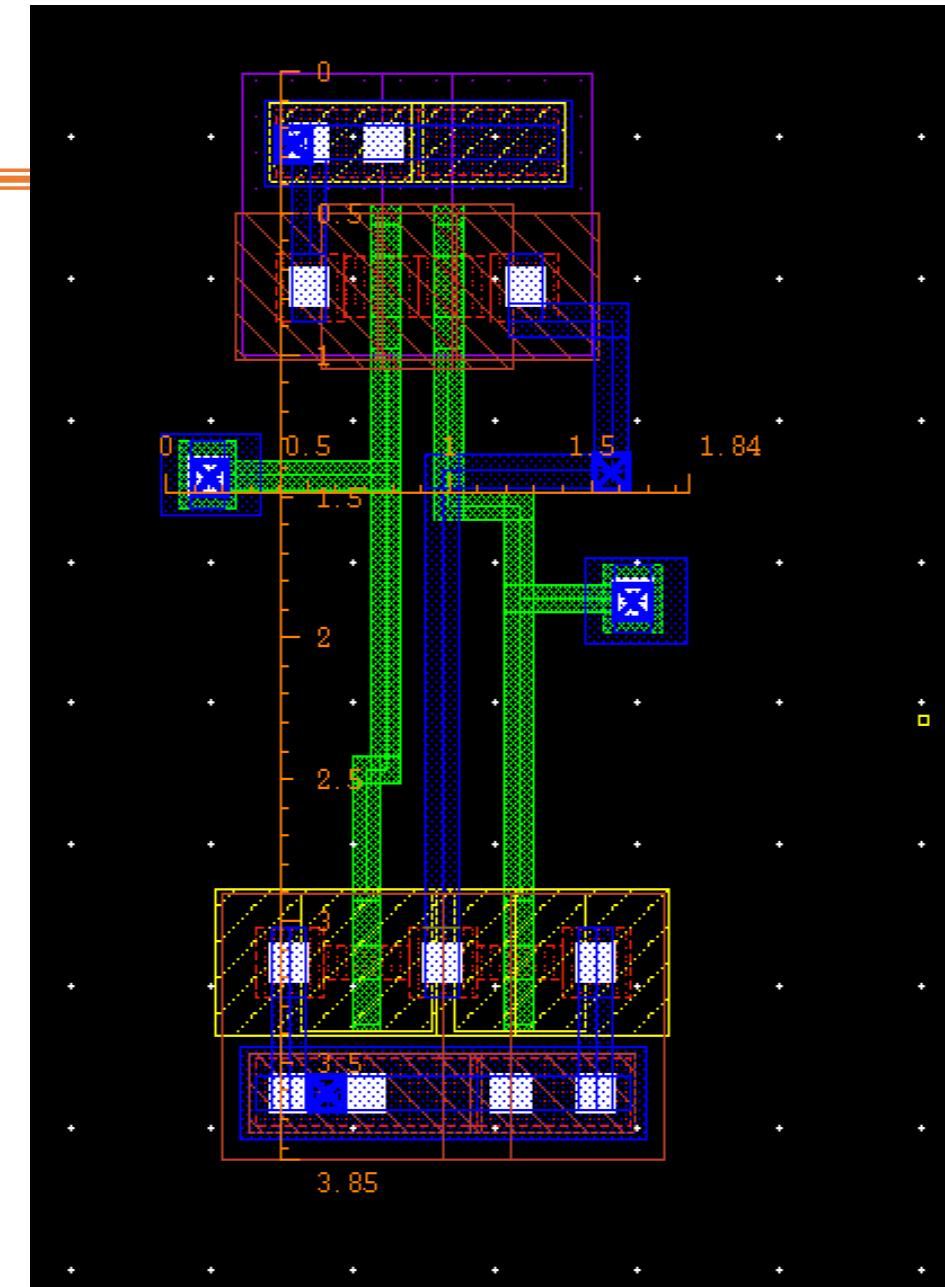
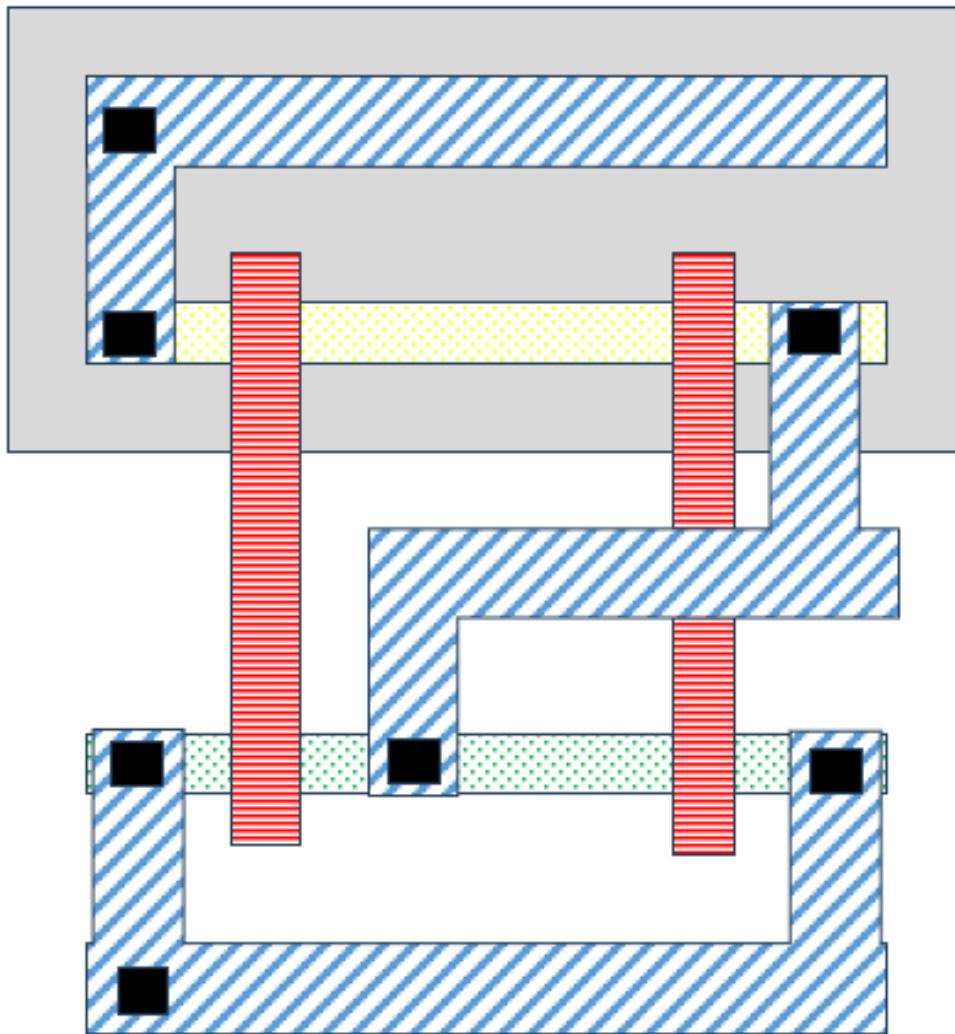


- Metal
- Polysilicon
- Metal Contact
- P-Doping
- N-Doping



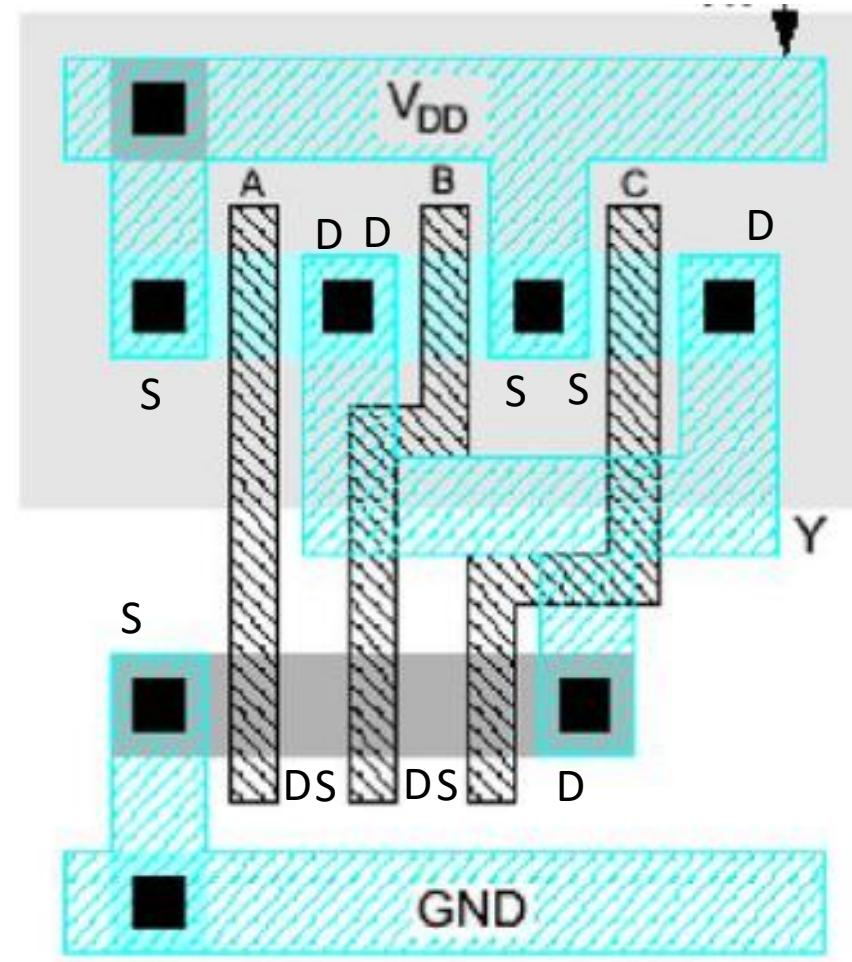
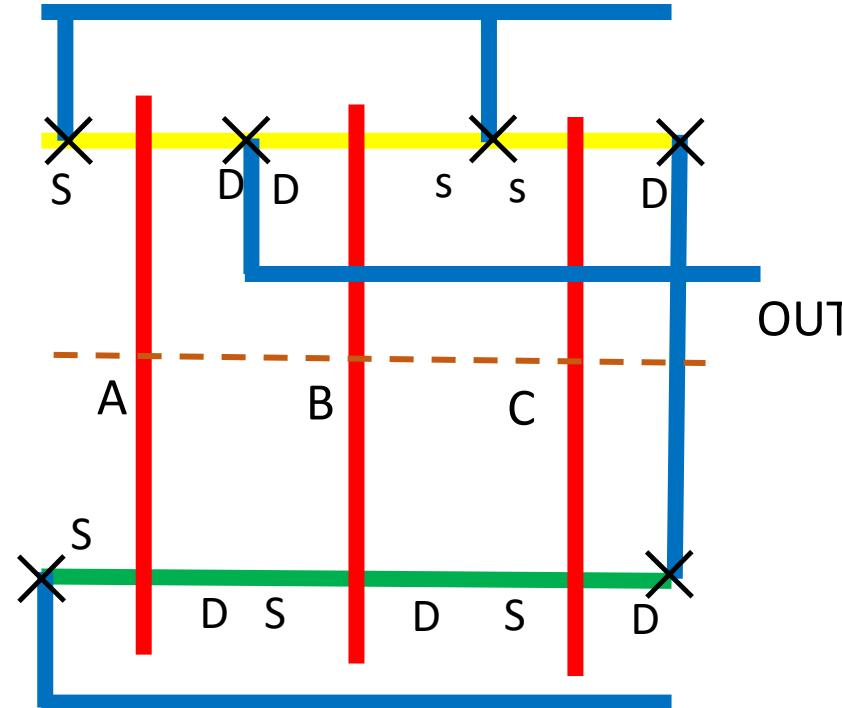
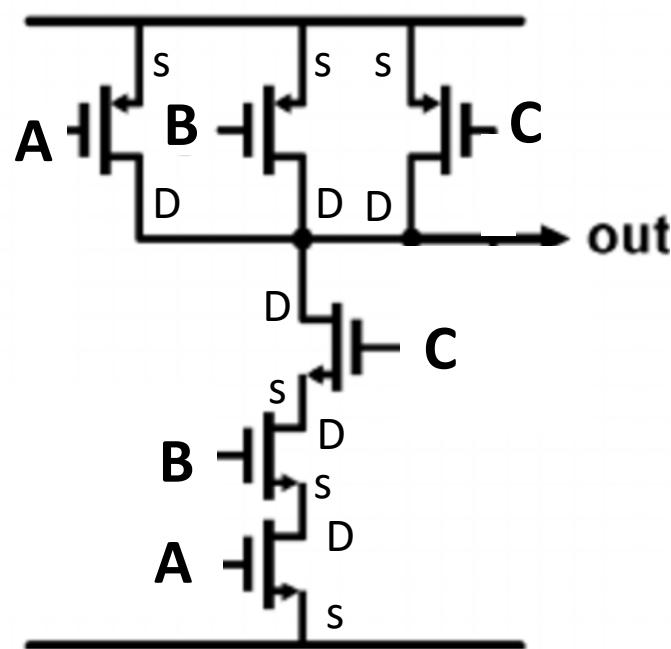
Draw the stick diagram and layout of CMOS NOR2







Draw the stick diagram and layout of CMOS NAND3



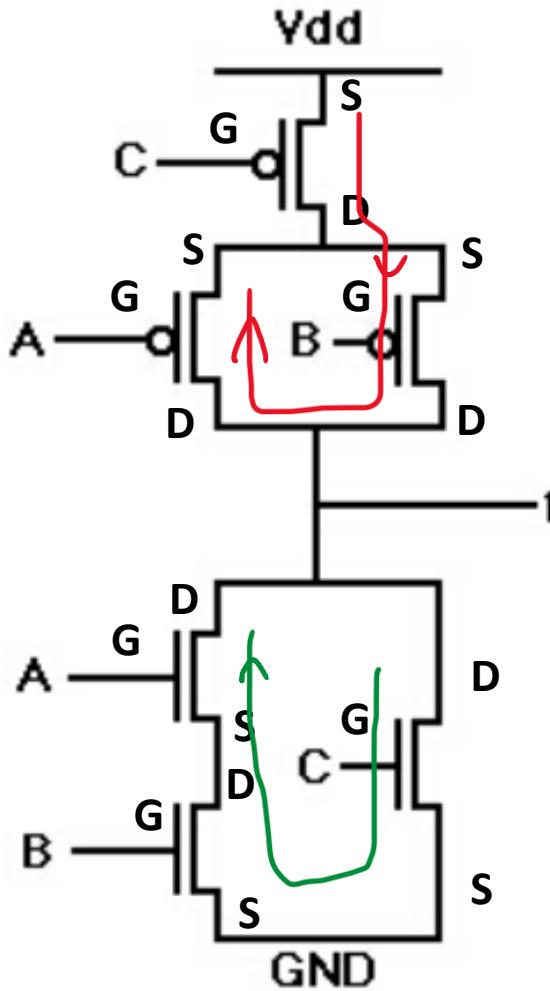


Draw the stick diagram and layout of CMOS NAND2

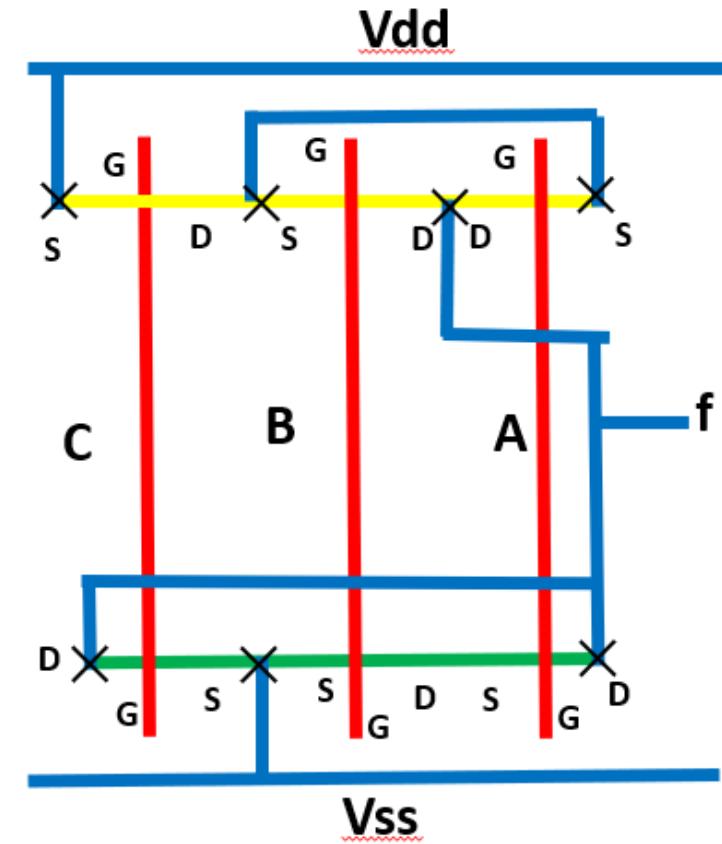
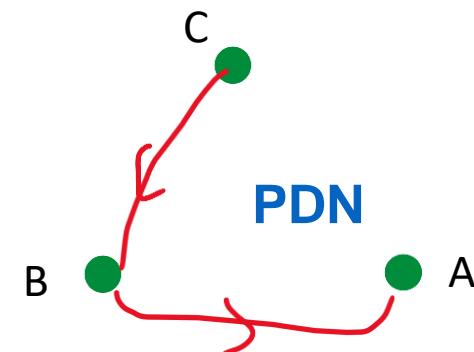
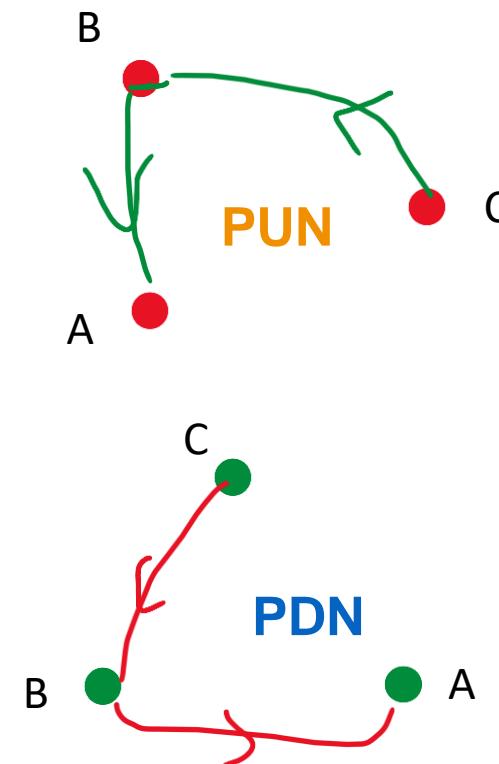
Draw the stick diagram and layout of CMOS NOR3

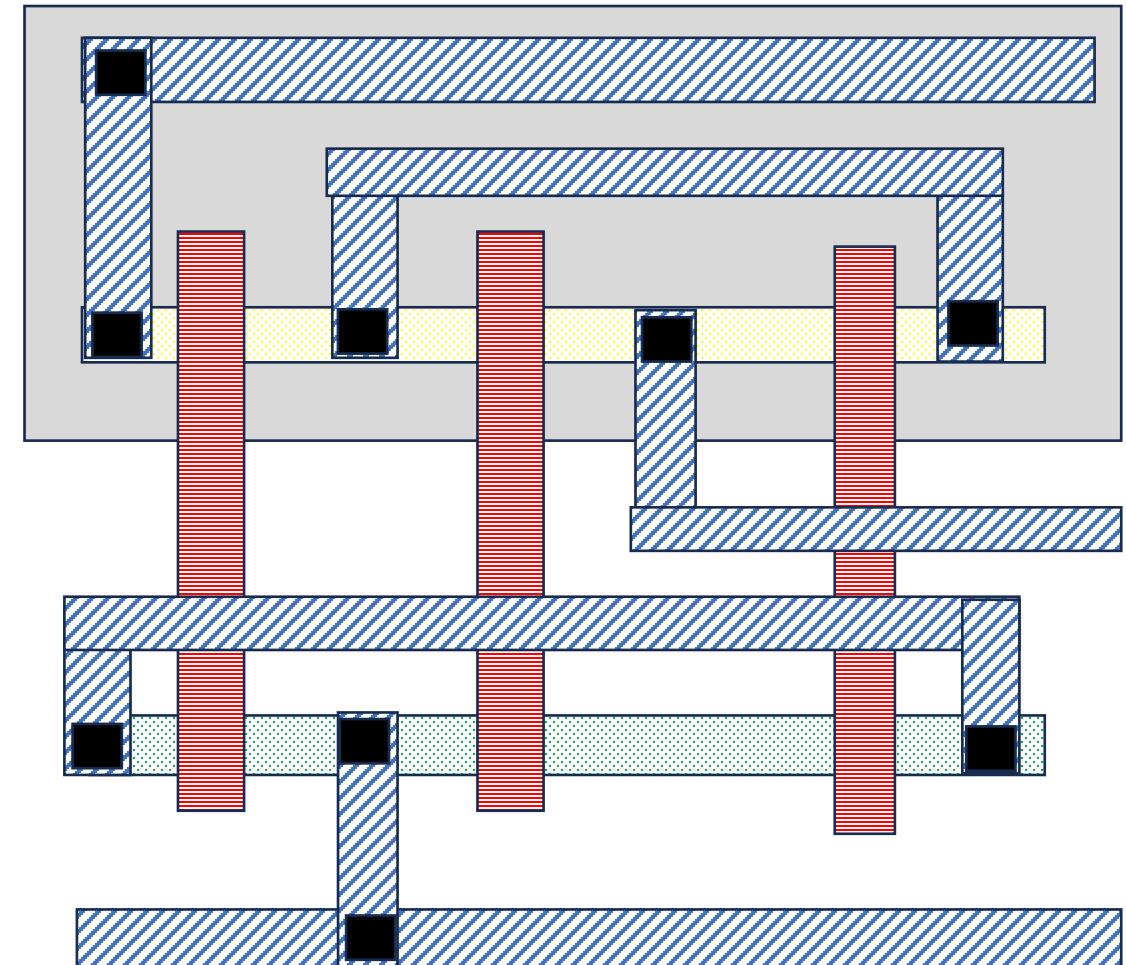
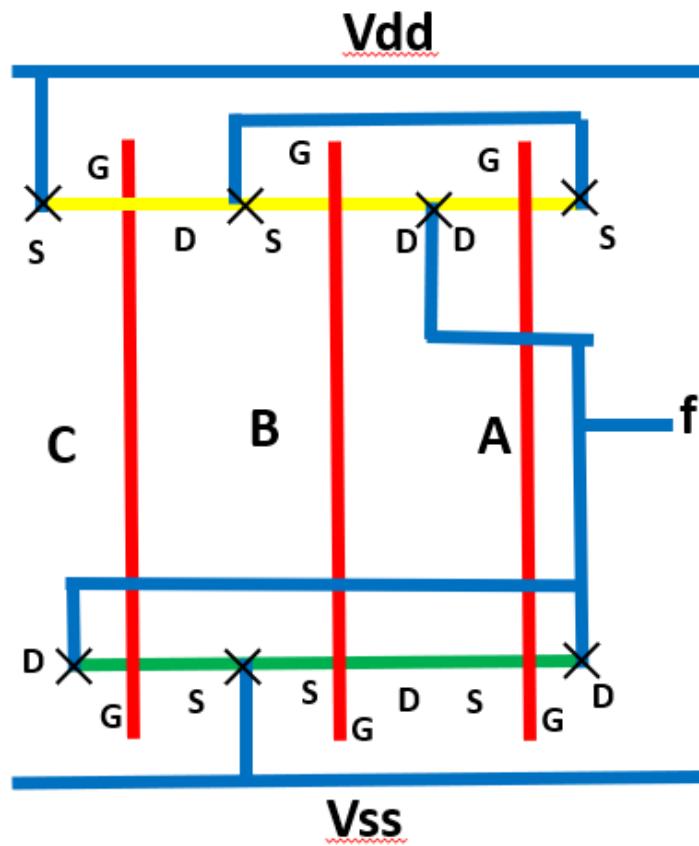


Draw the stick diagram and layout of graph of $f = \overline{(A \bullet B)} + C$



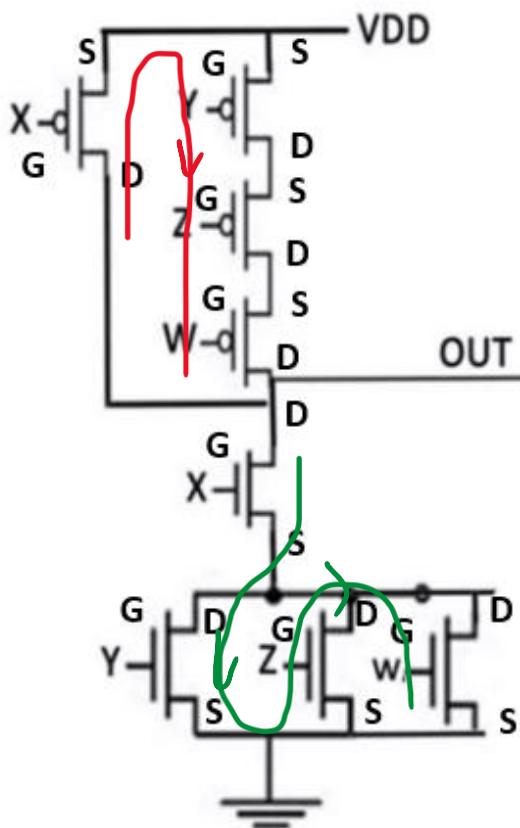
Euler's path is C → B → A



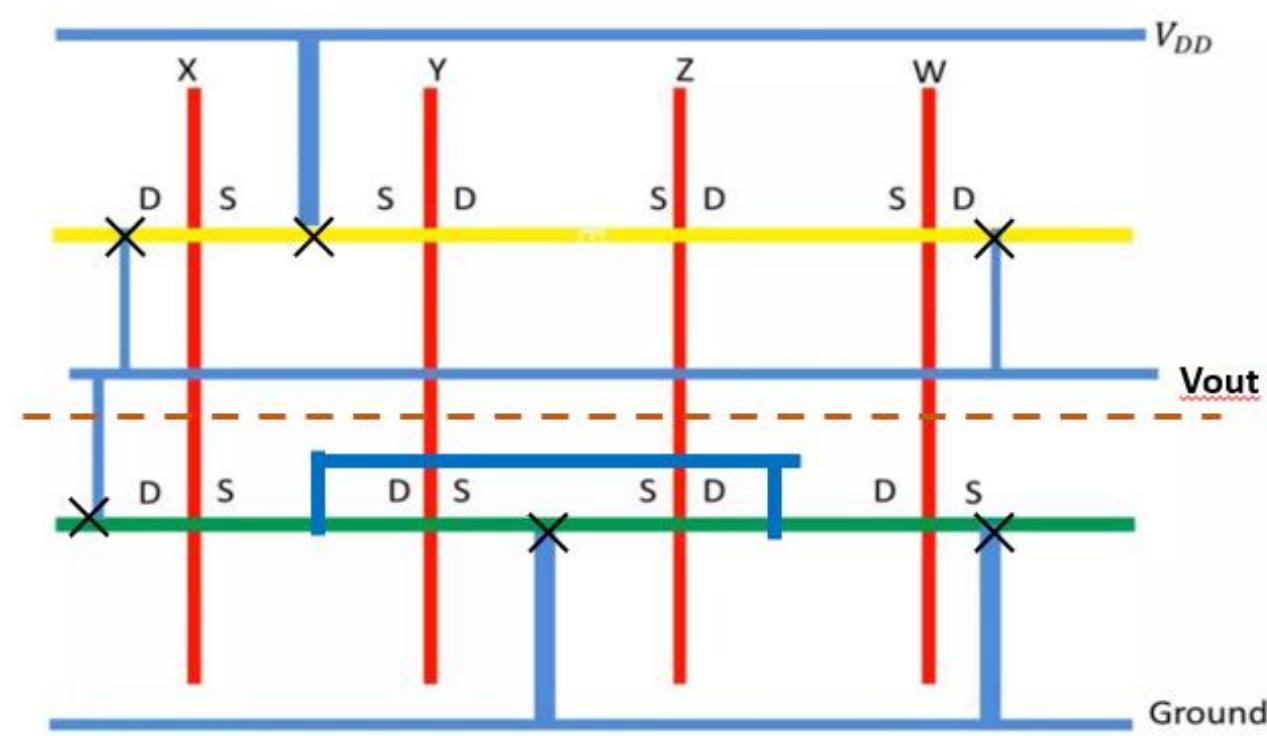
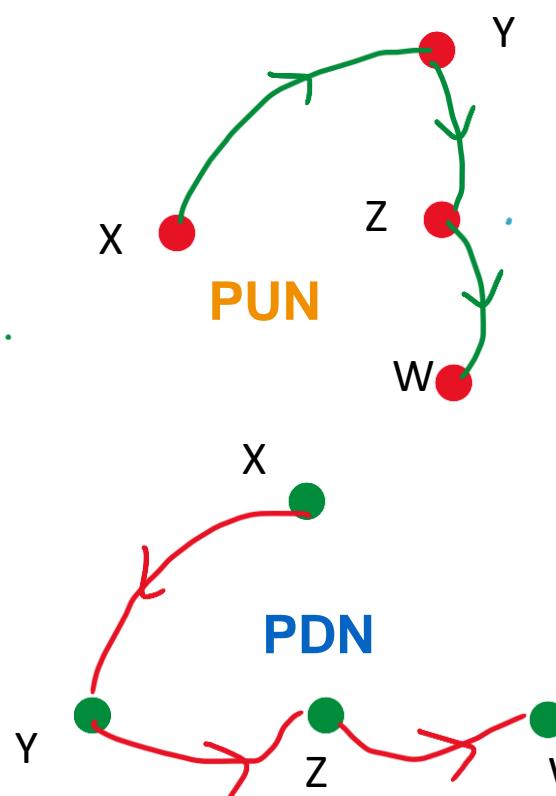


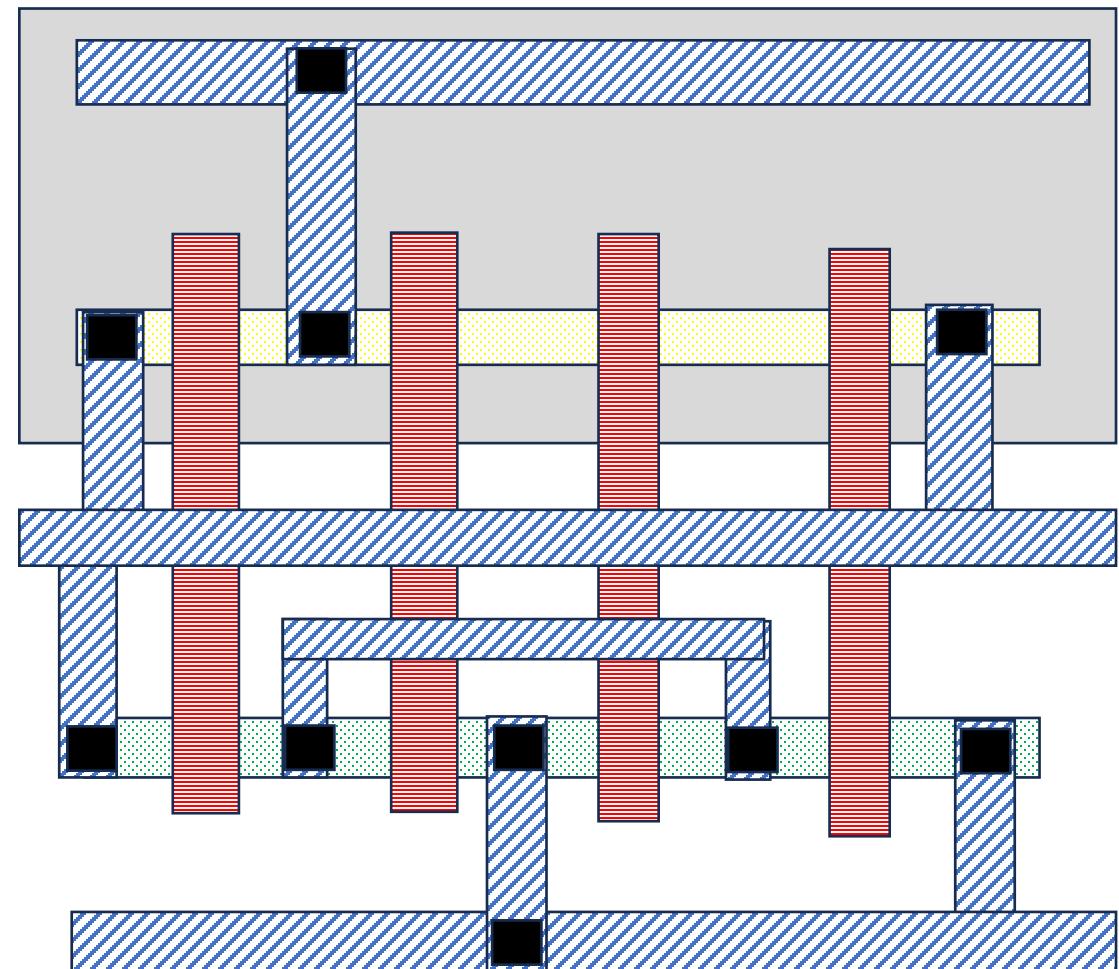
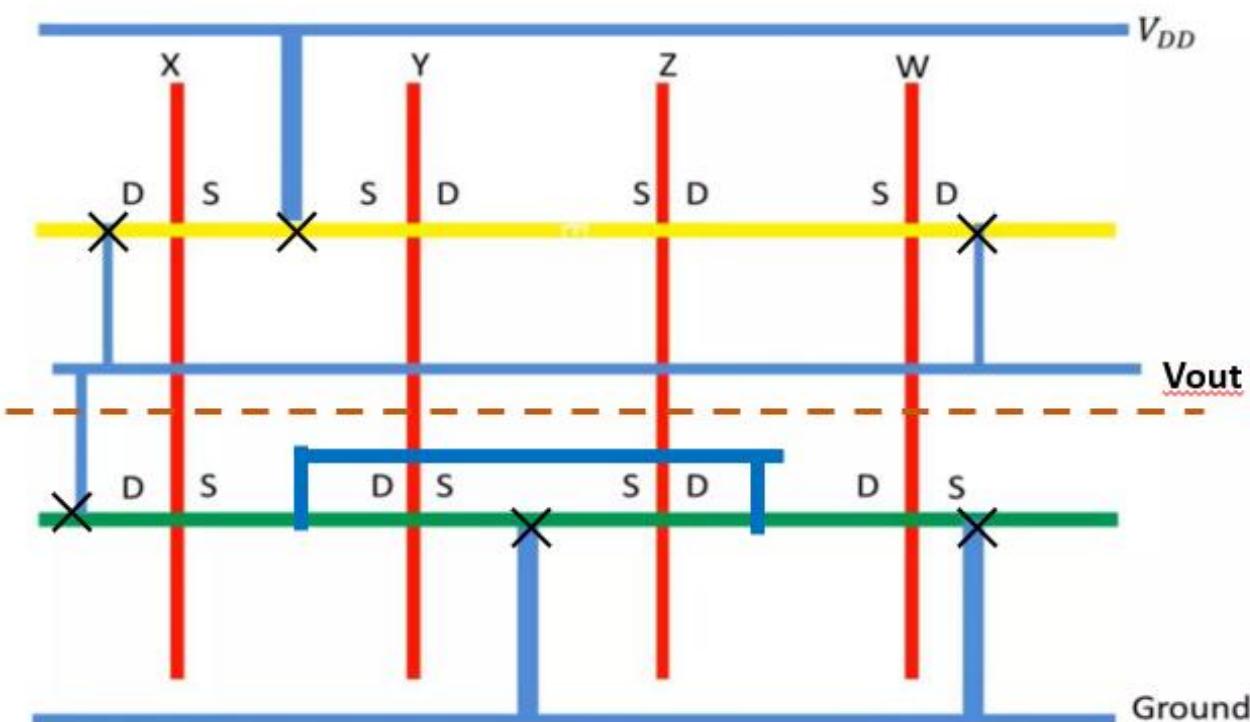


Draw the stick diagram and layout of graph of $F=((y.z.w)+x)'$



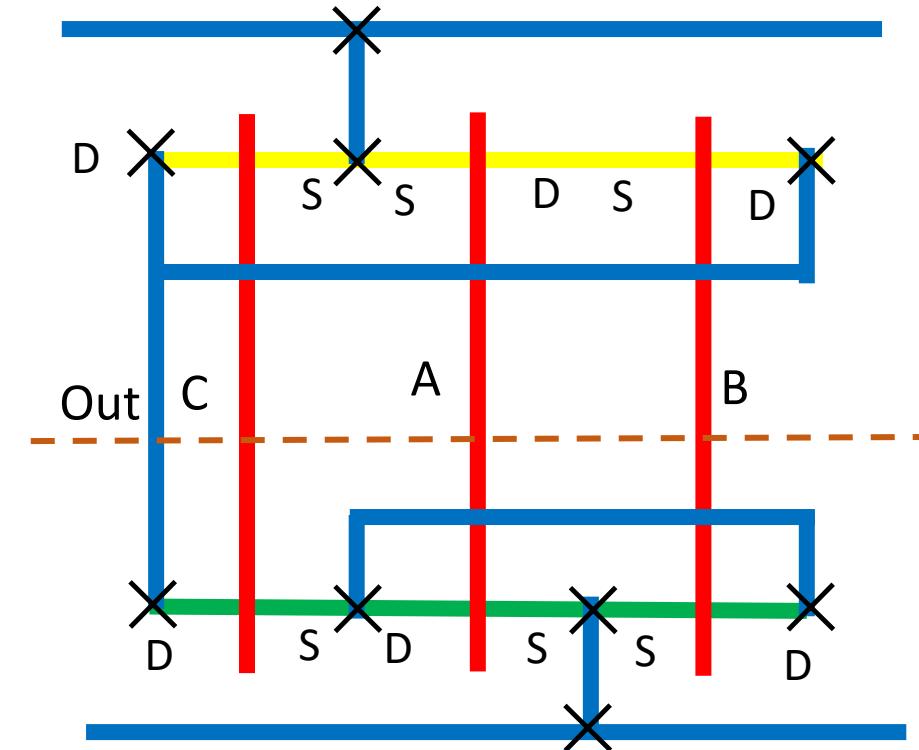
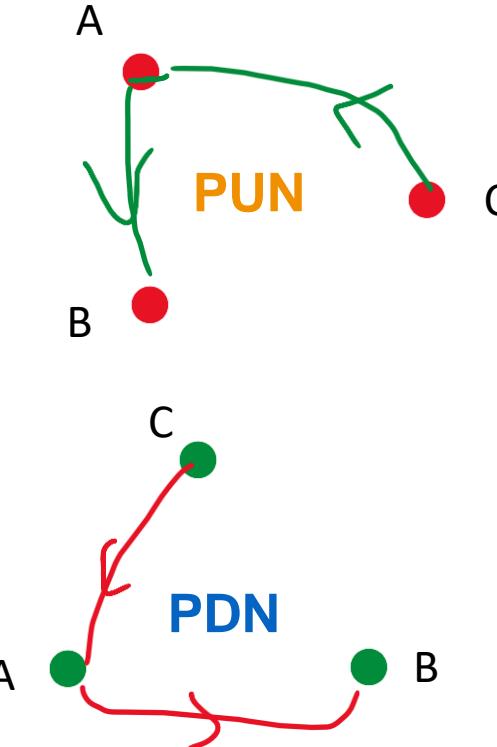
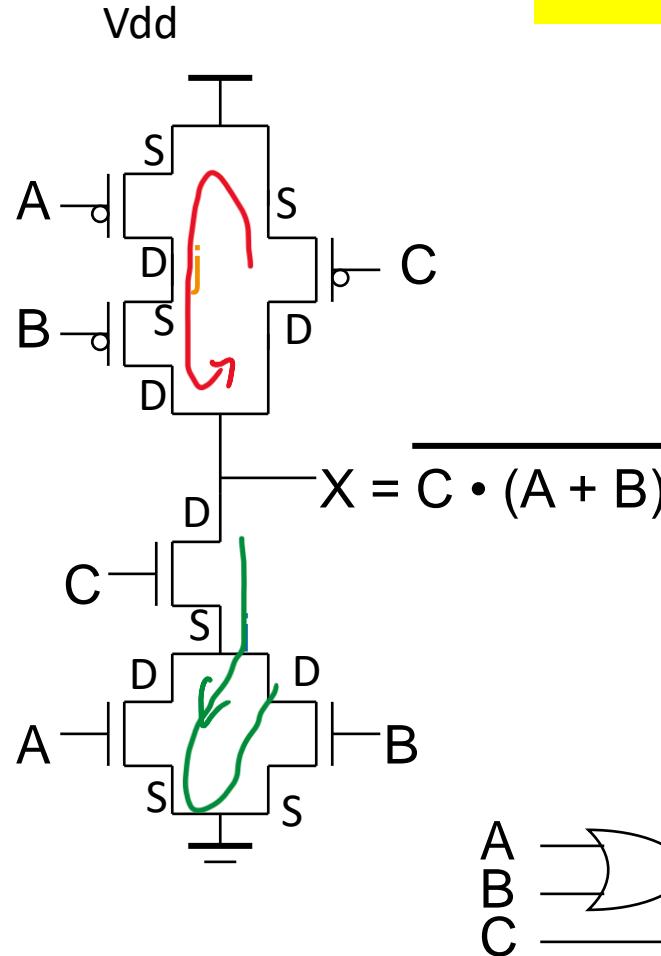
Euler's path is $X \rightarrow Y \rightarrow Z \rightarrow W$

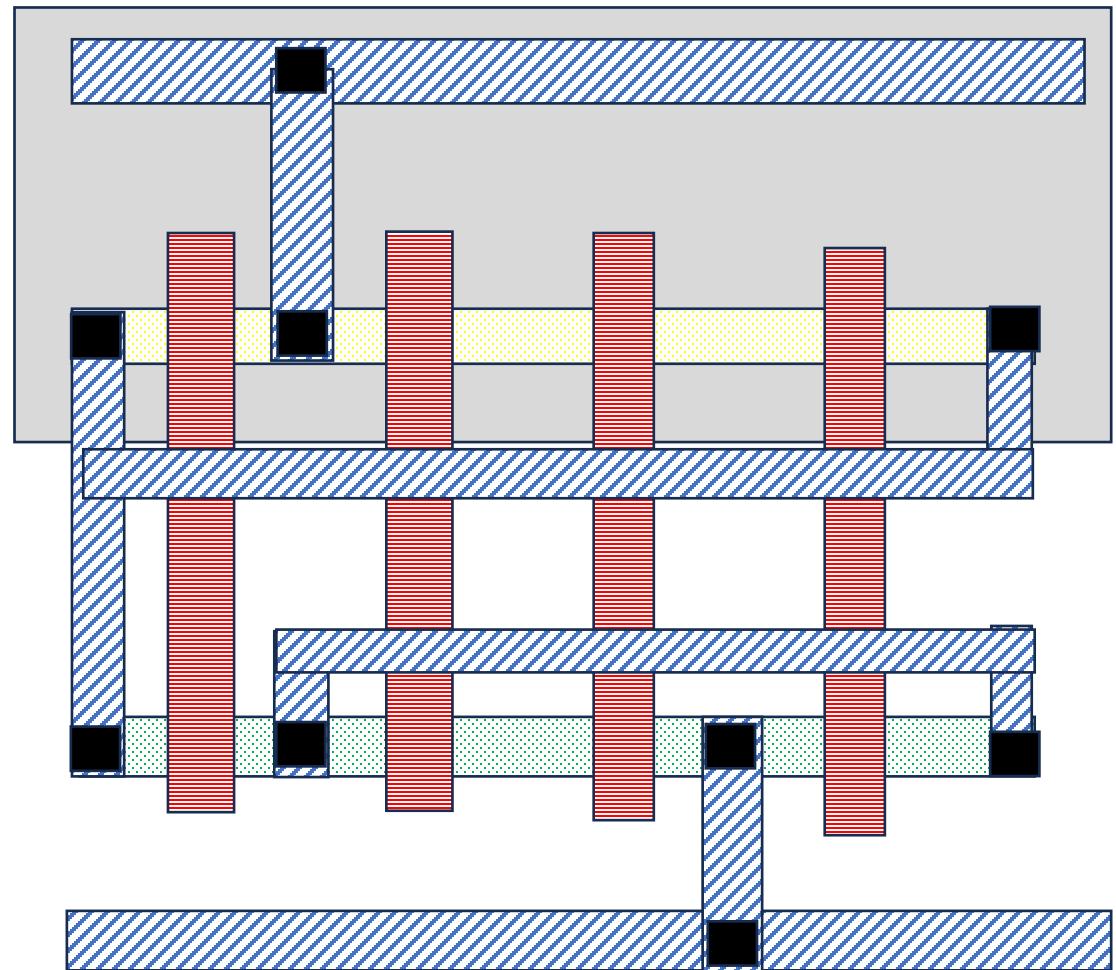
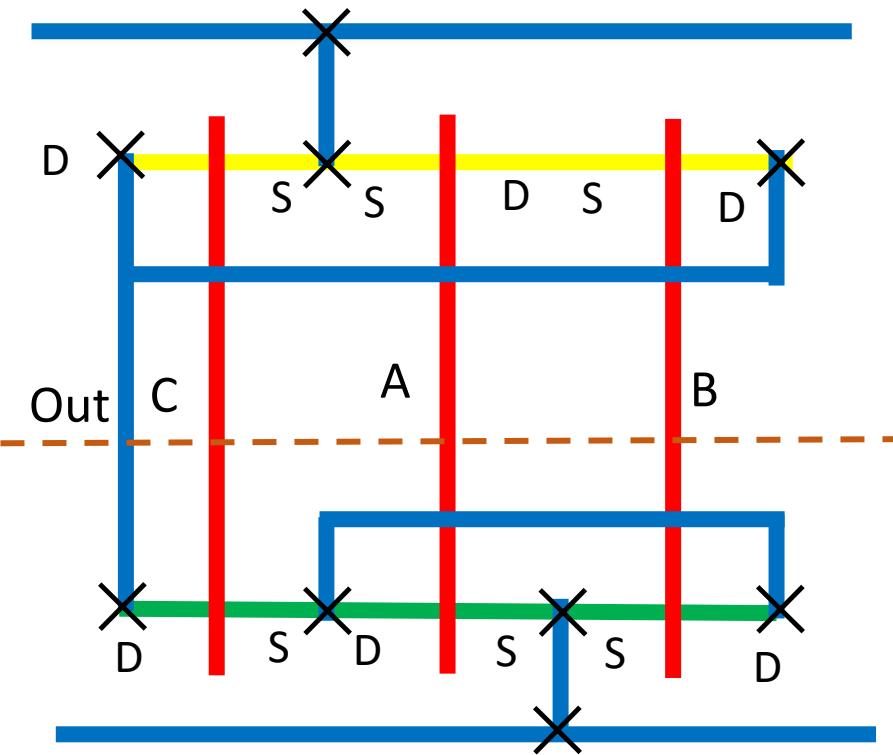






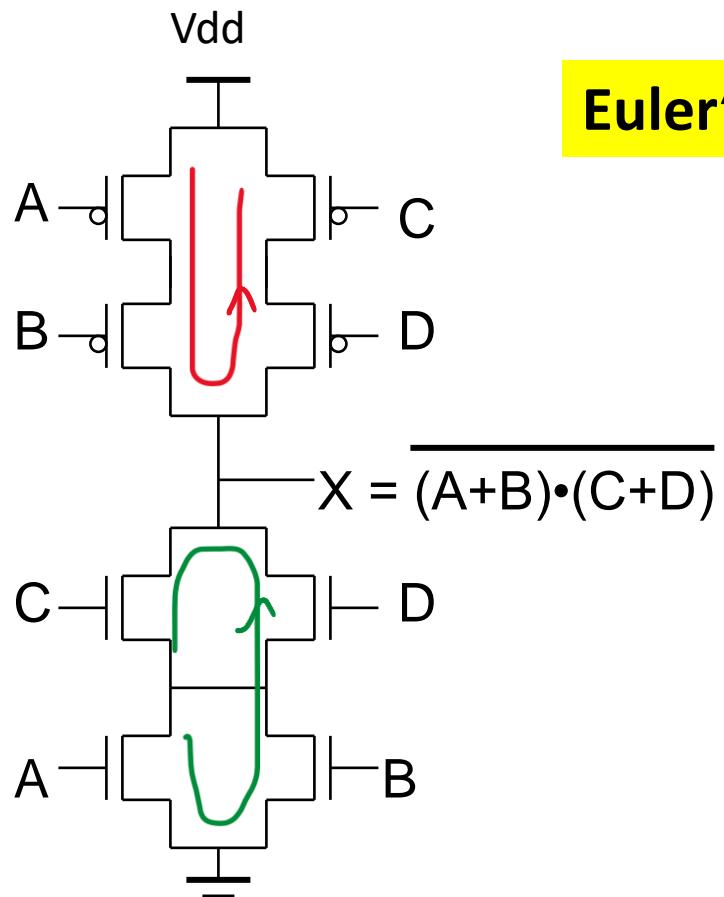
Draw the stick diagram and layout graph of $X = \overline{C} \cdot (\overline{A} + \overline{B})$



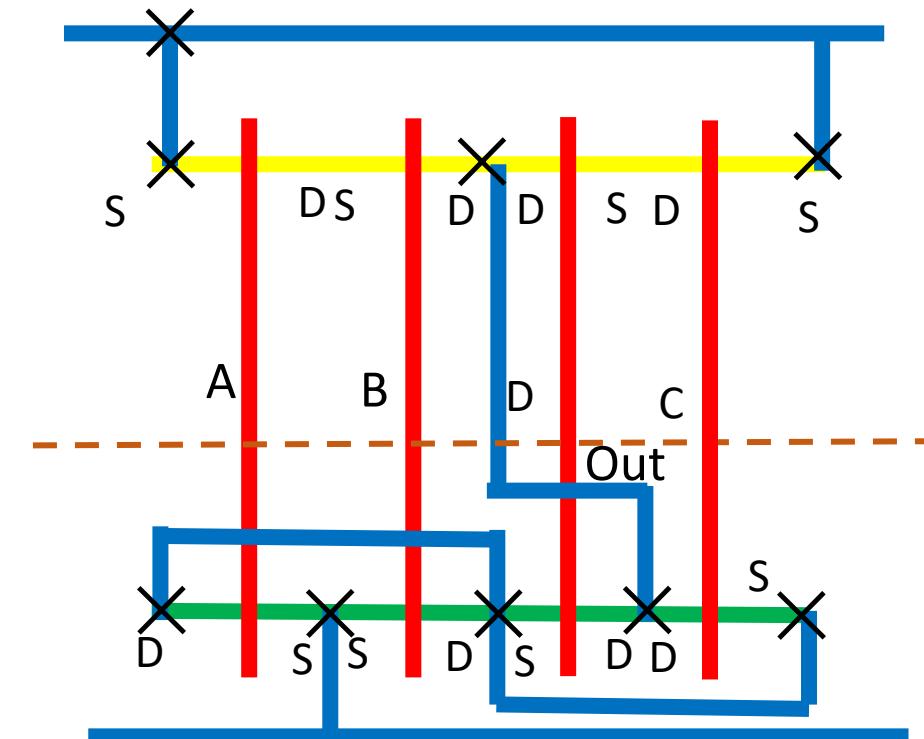
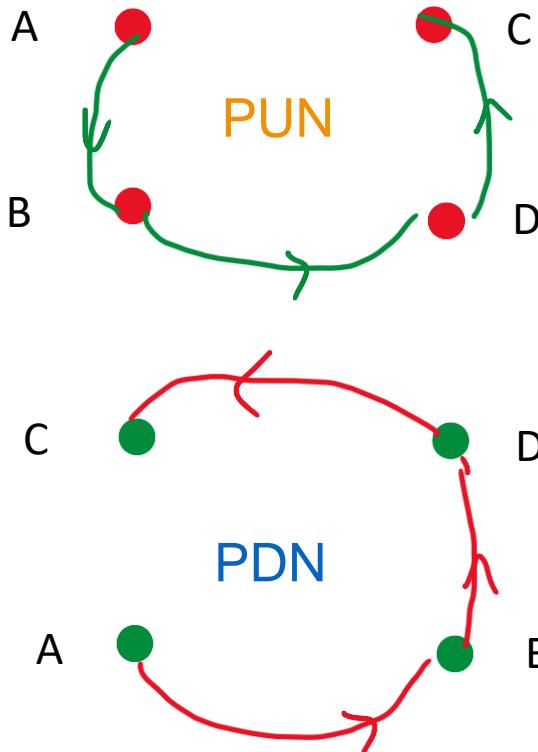


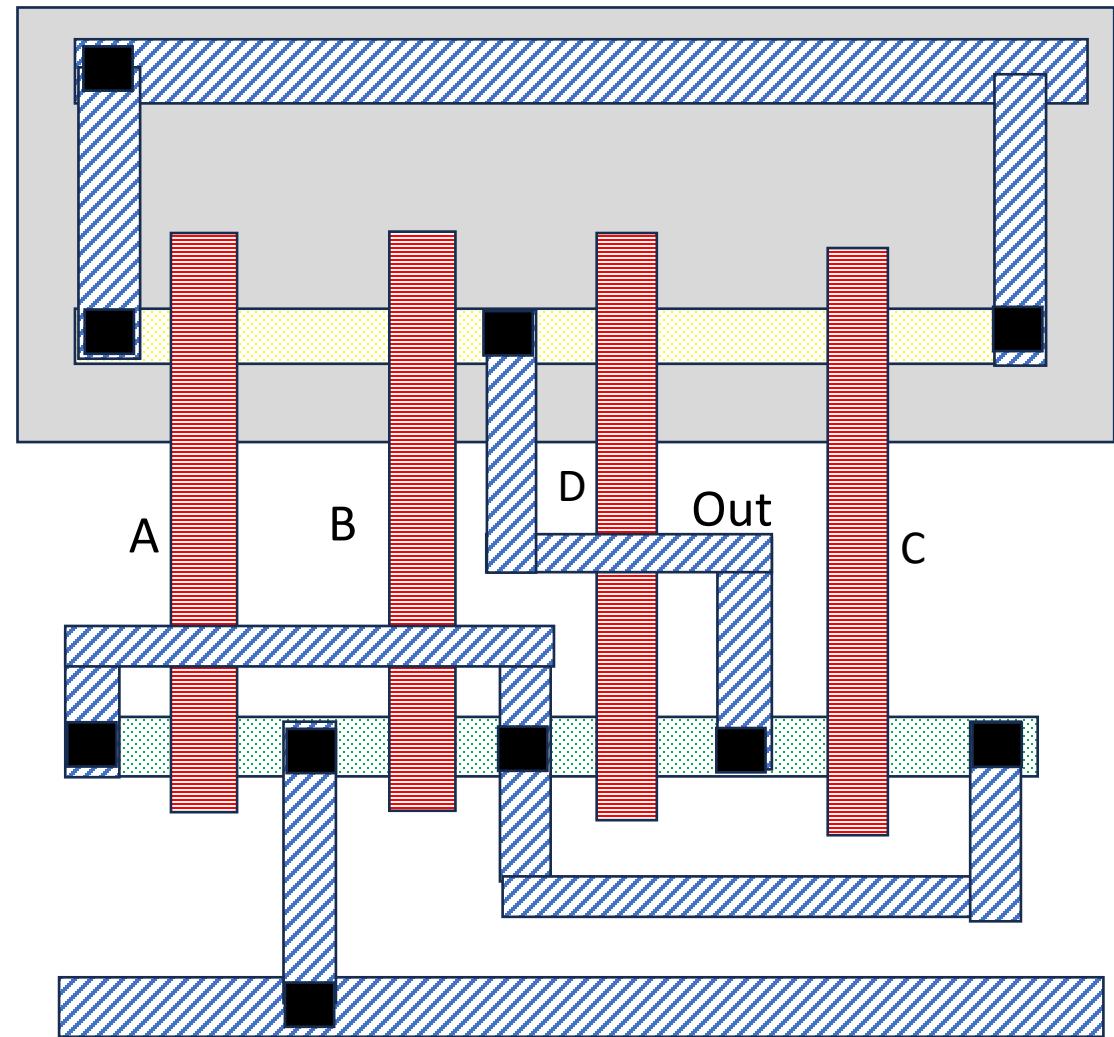
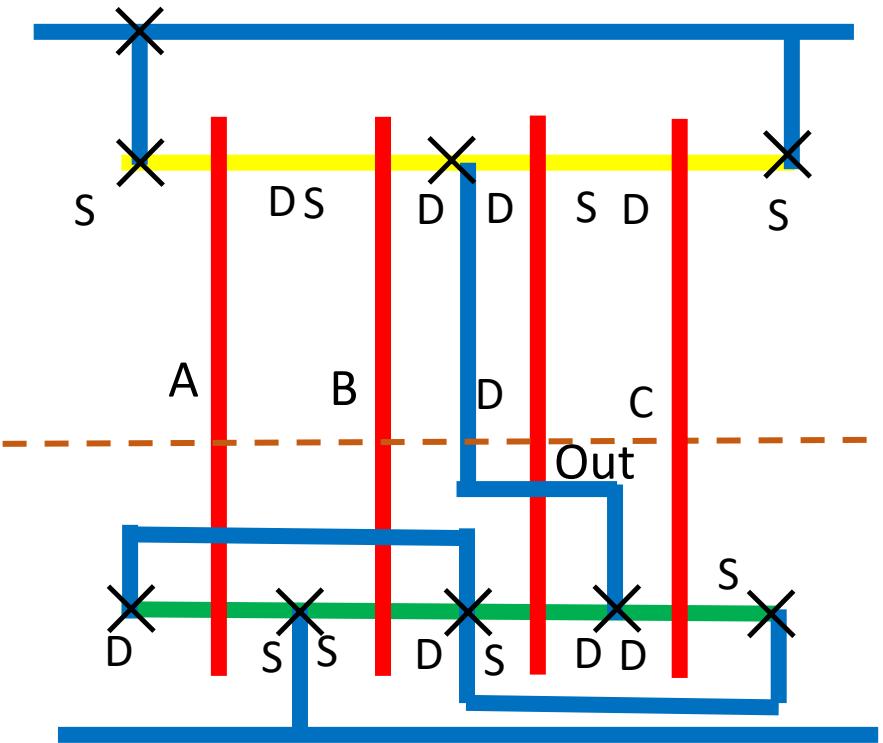


Sketch the stick diagram and layout of OAI22



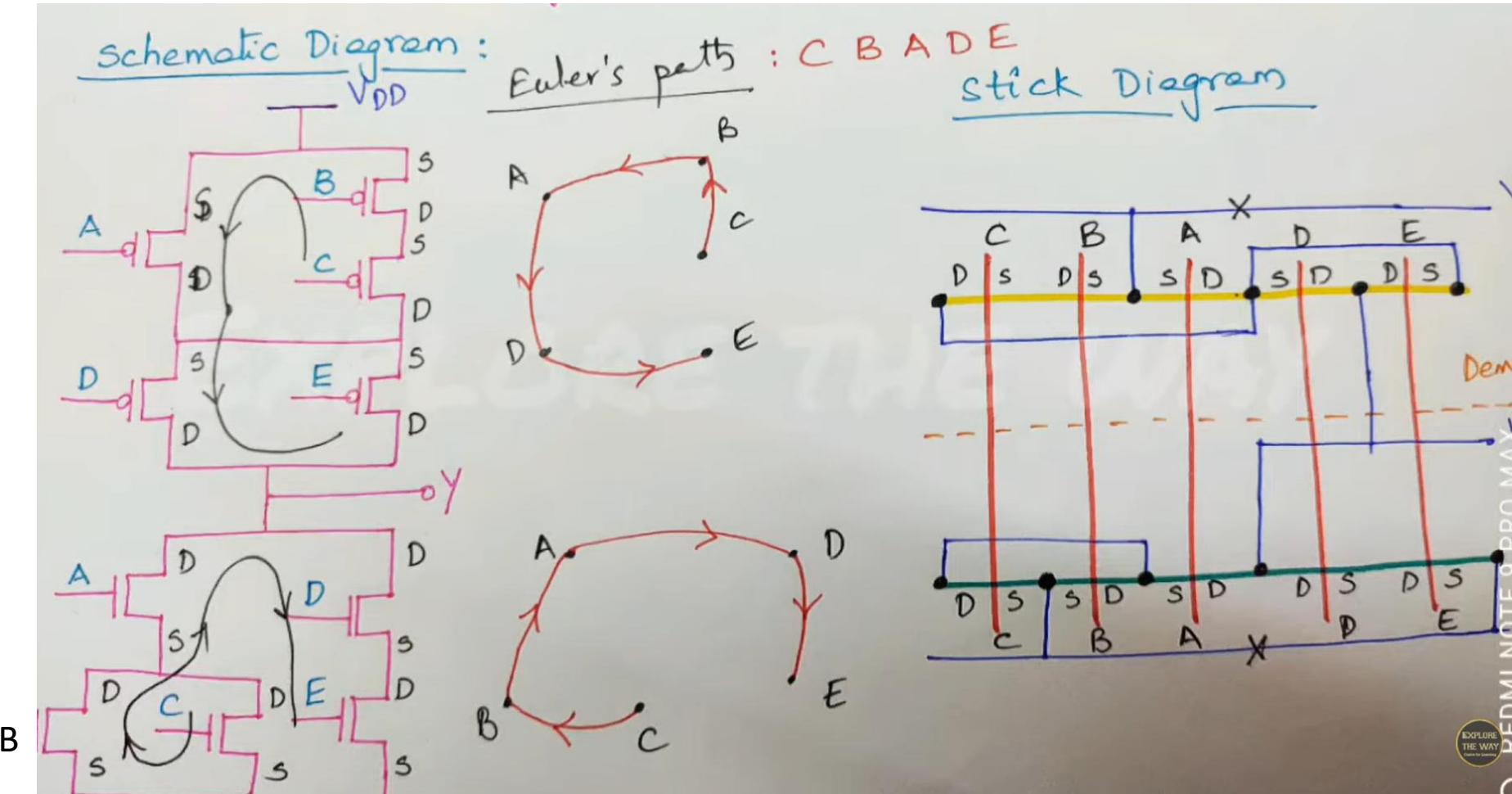
Euler's path is $A \rightarrow B \rightarrow D \rightarrow C$

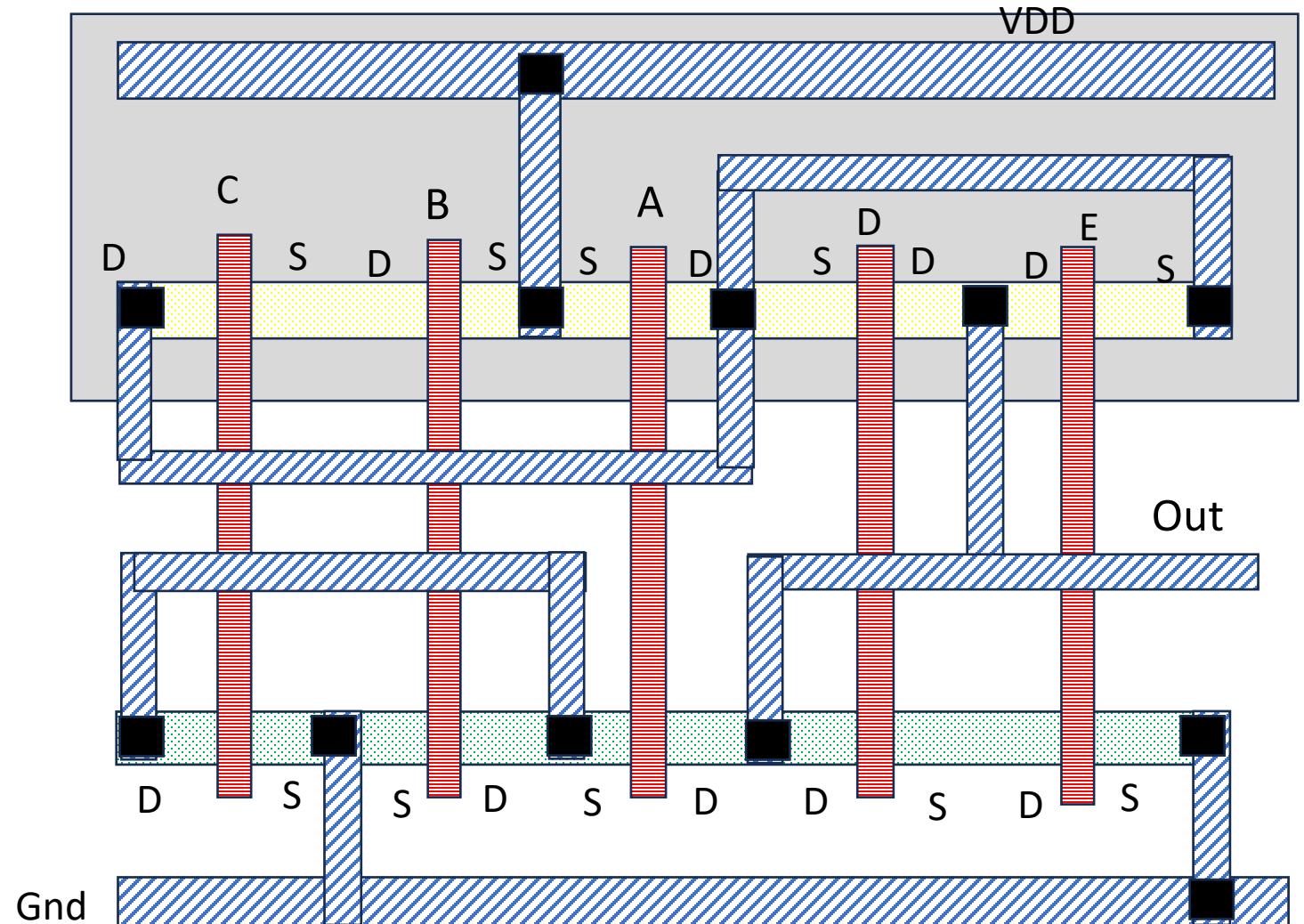
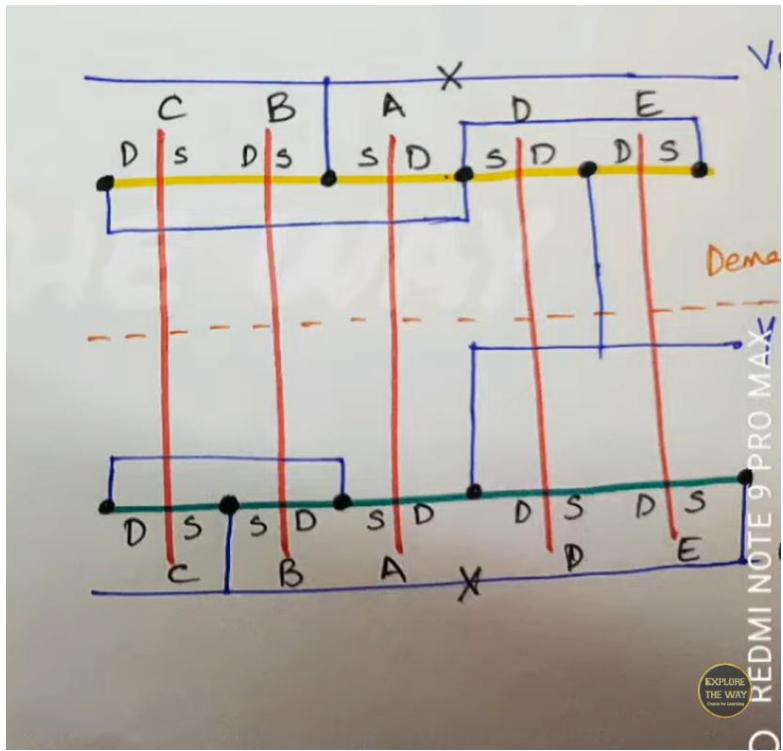






Draw stick diagram and layout of $y=(A(B+C)+DE)'$



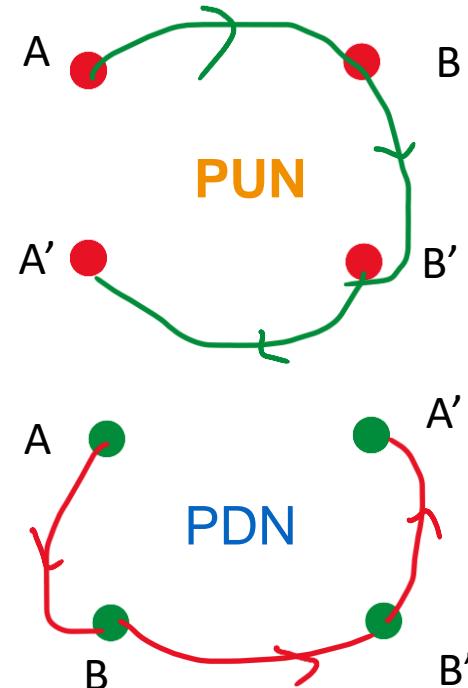
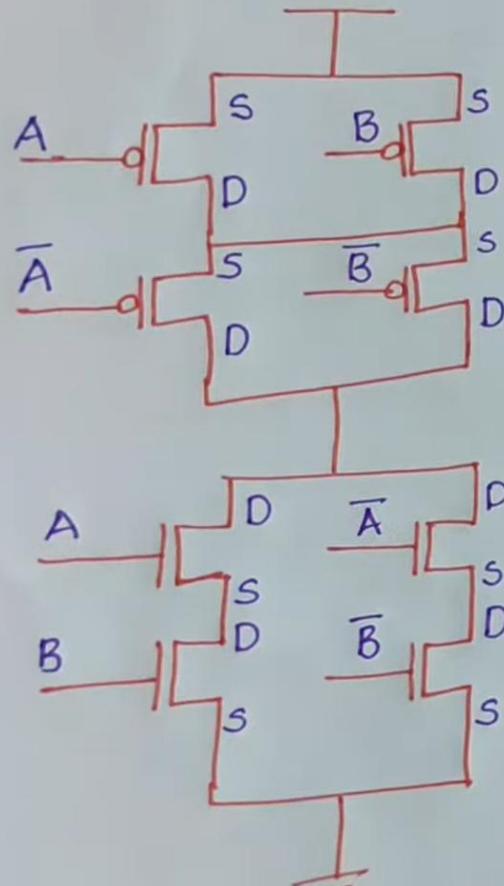




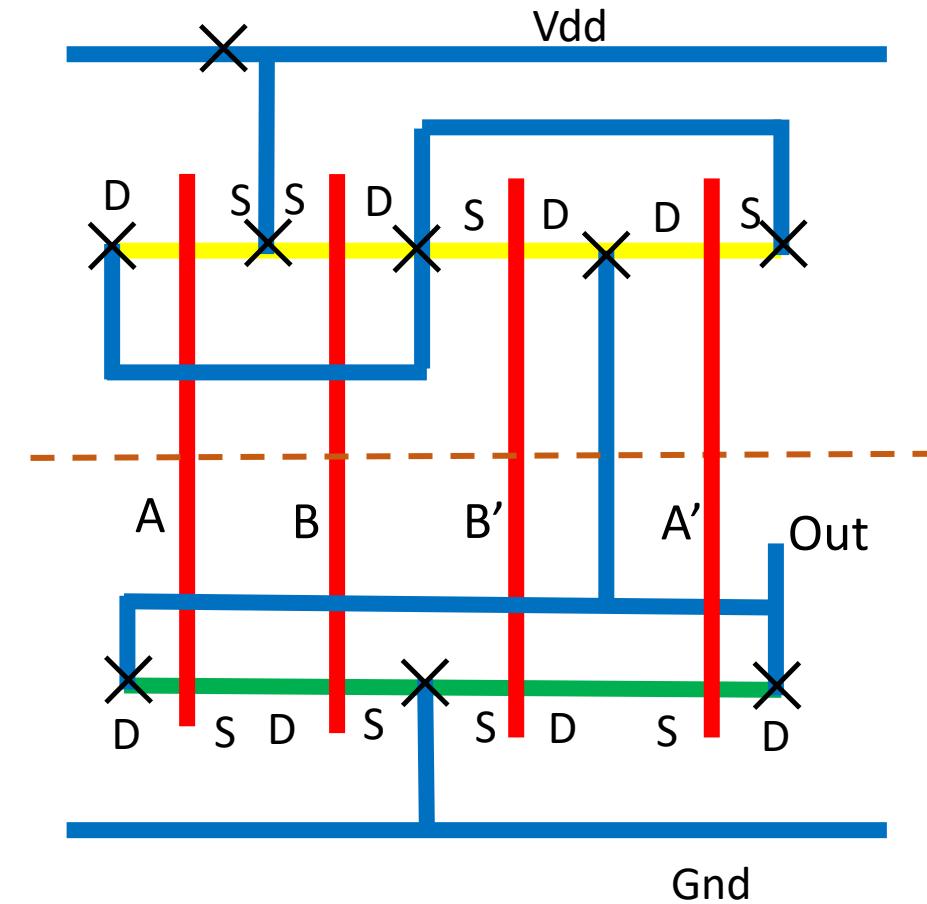
Draw stick diagram and layout of two input EXOR gate

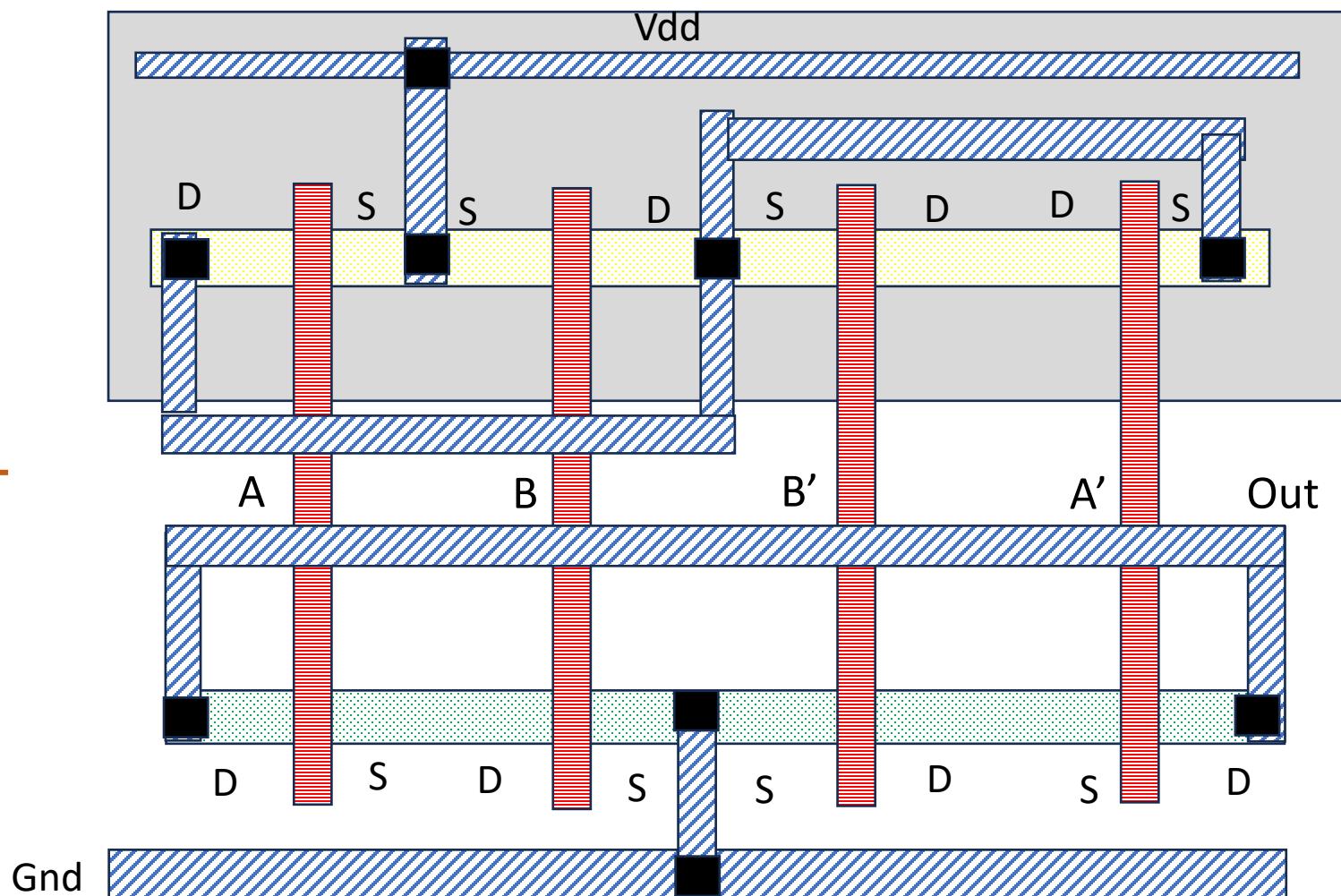
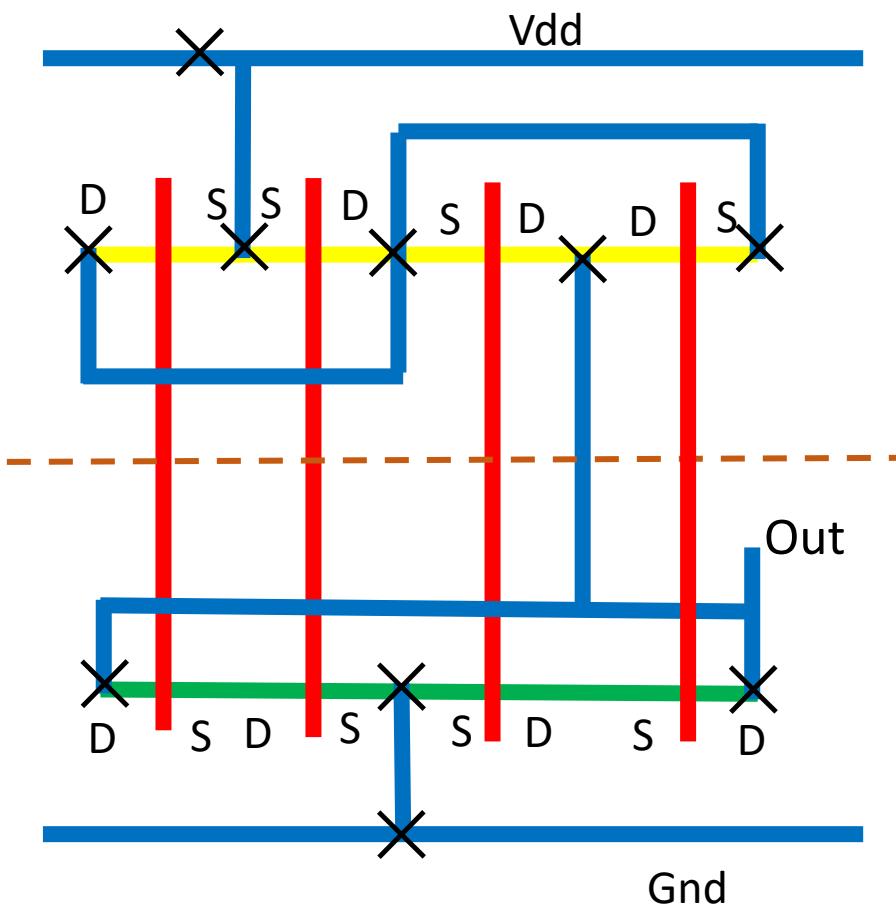
Logic function $Y = \overline{A} \cdot B + A \overline{B} = \overline{AB + \overline{AB}}$

Schematic Diagram :



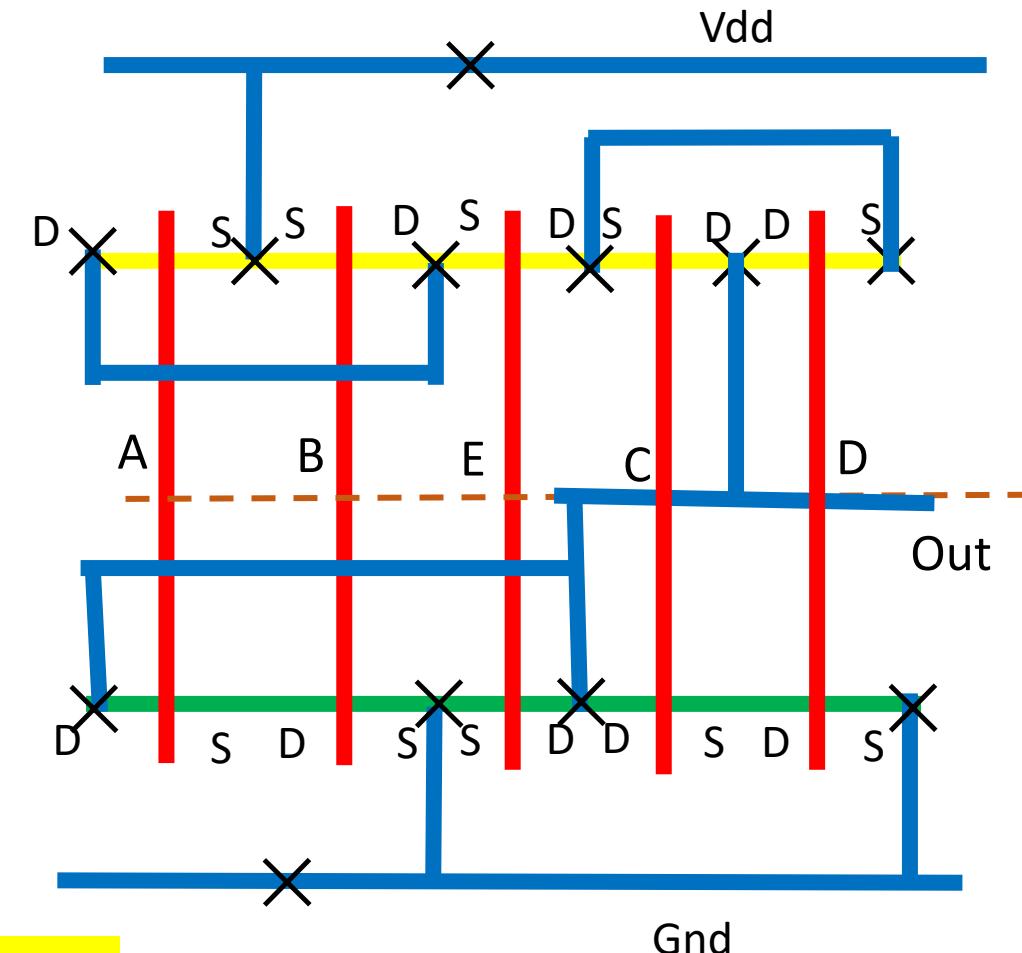
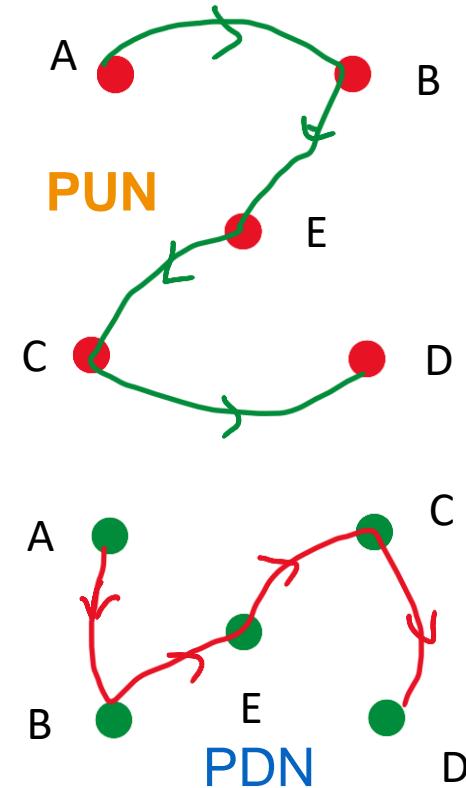
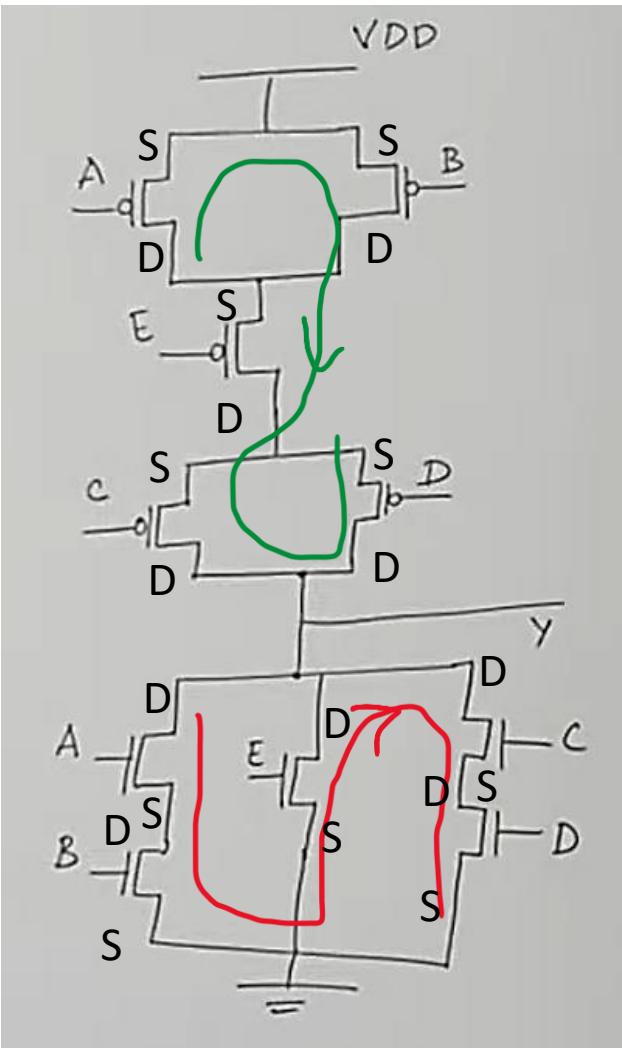
Euler's path is $A \rightarrow B \rightarrow B' \rightarrow A'$



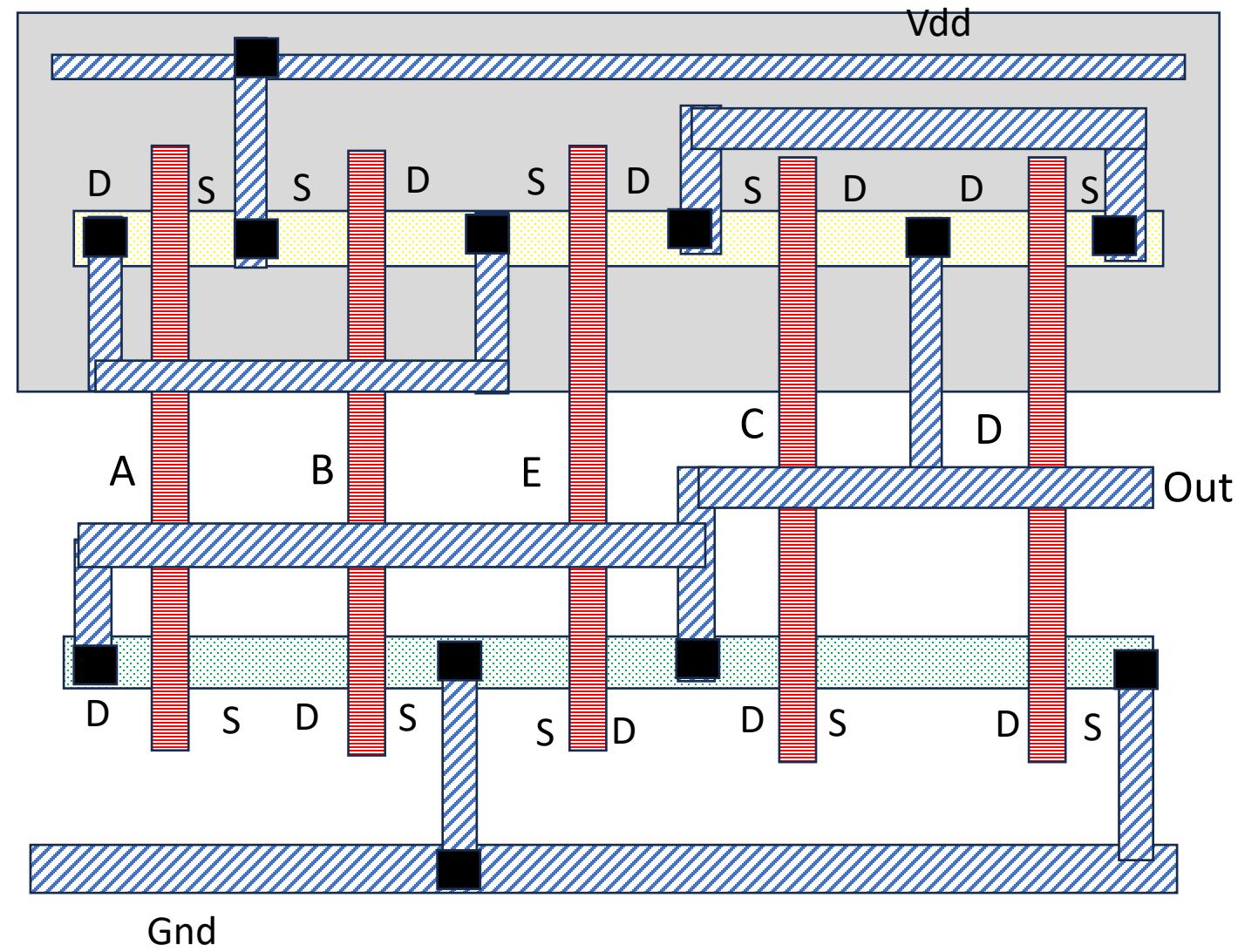
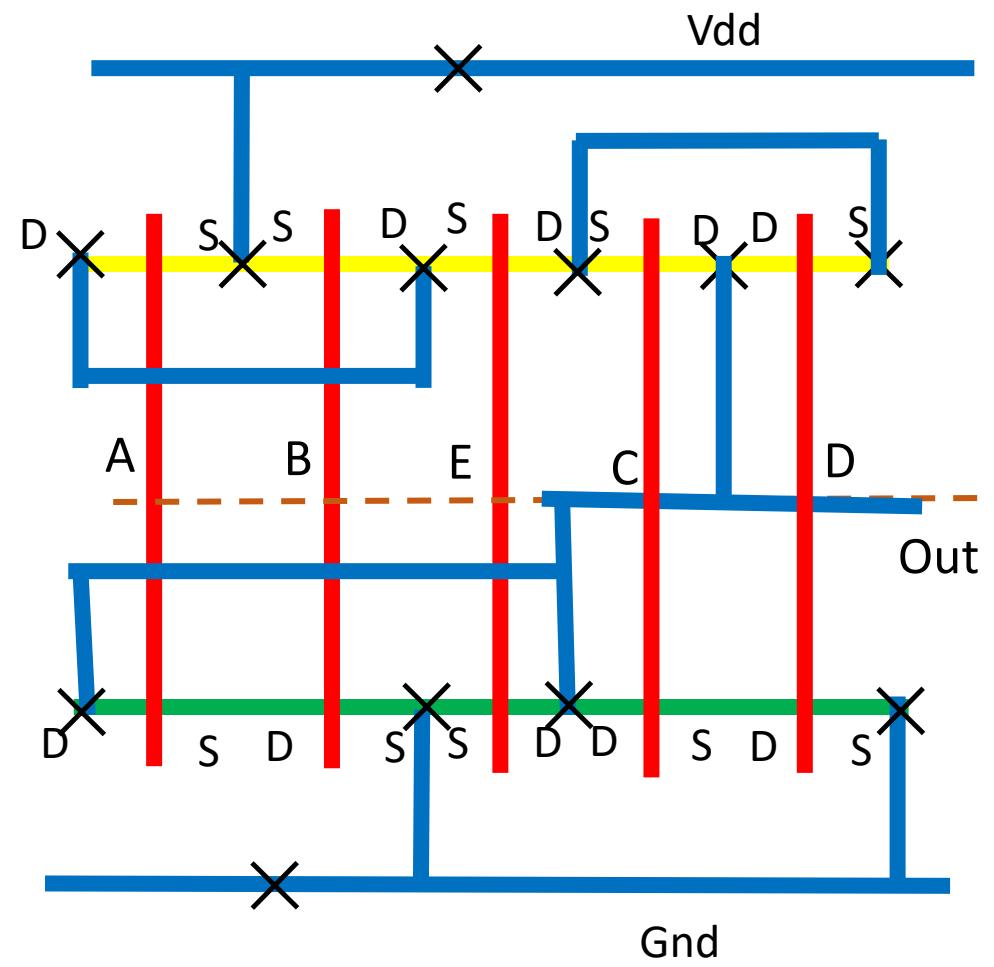




Draw stick diagram and layout of $y=(AB+E+CD)'$

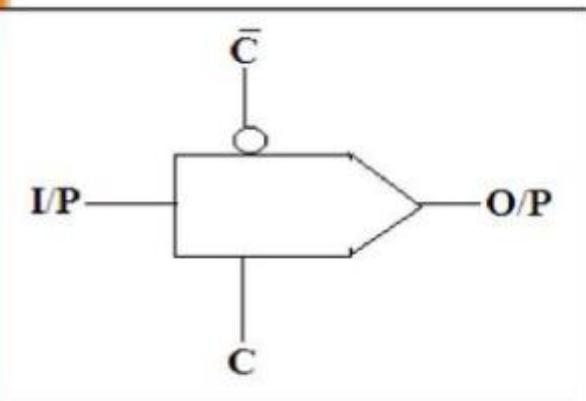


Euler's path is $A \rightarrow B \rightarrow E \rightarrow C \rightarrow D$

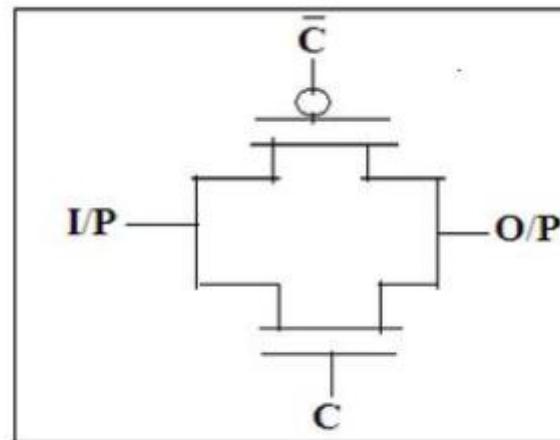




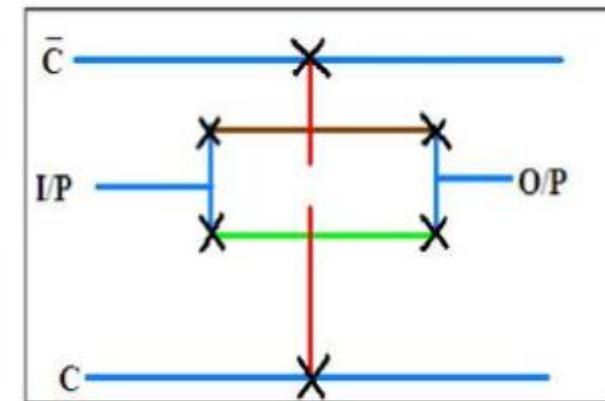
Draw stick diagram and layout of transmission gate



Symbol

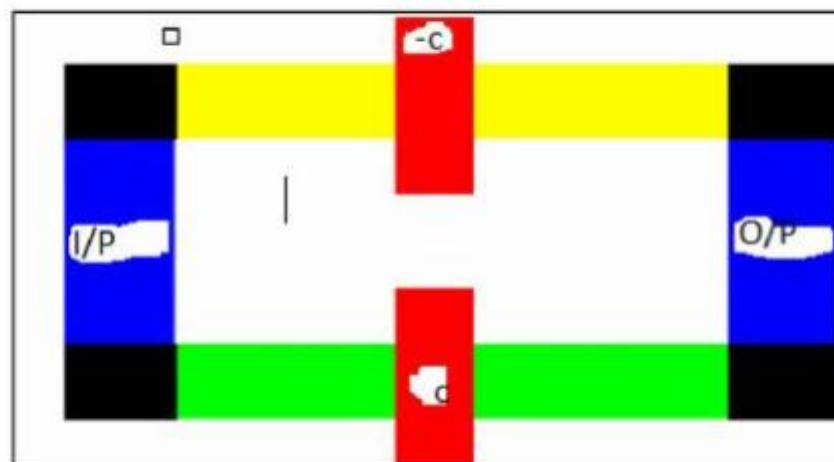


schematic



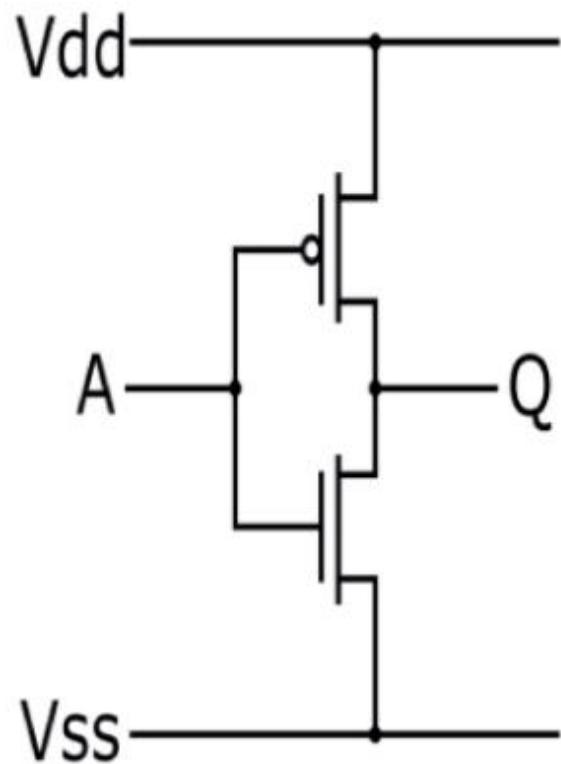
stick diagram

layout

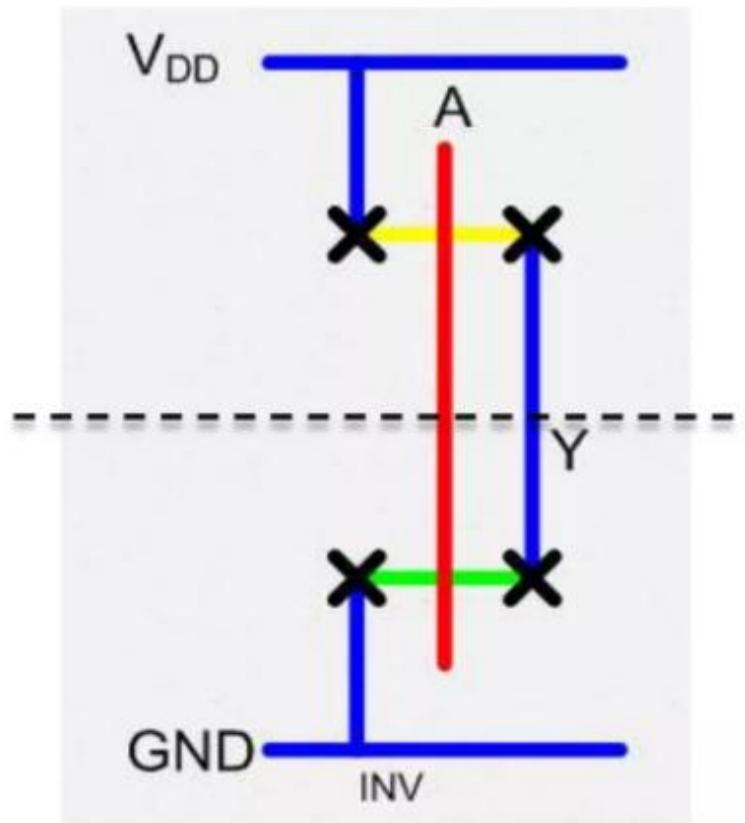




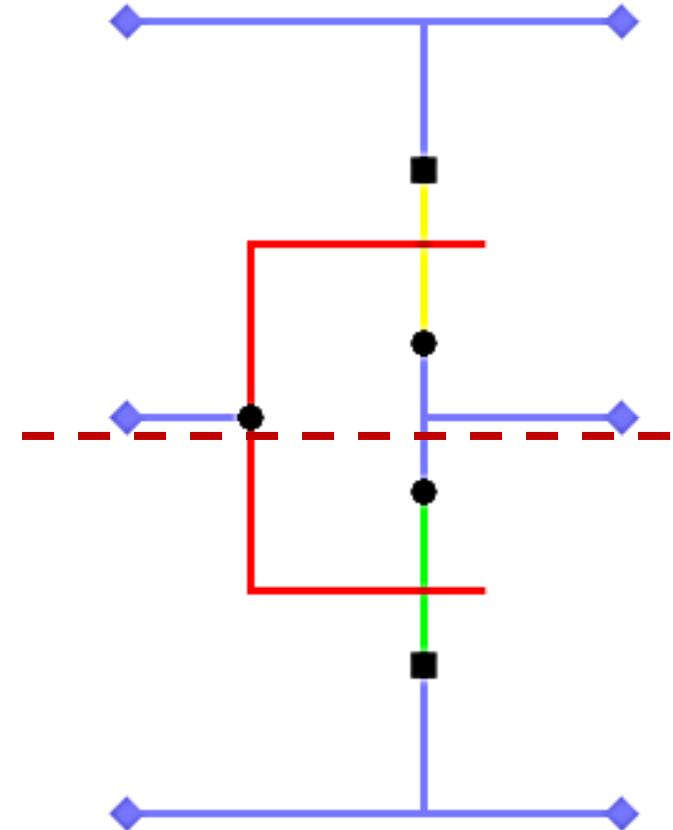
FYI



Horizontal Placement



Vertical Placement





Draw stick diagram and layout of nMOS depletion load NOR2

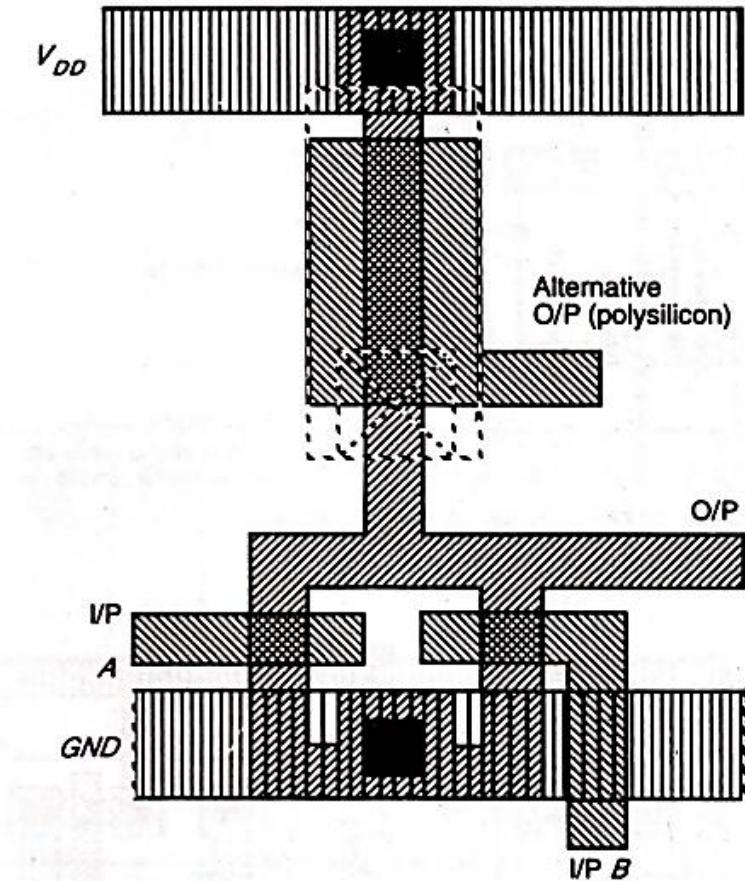
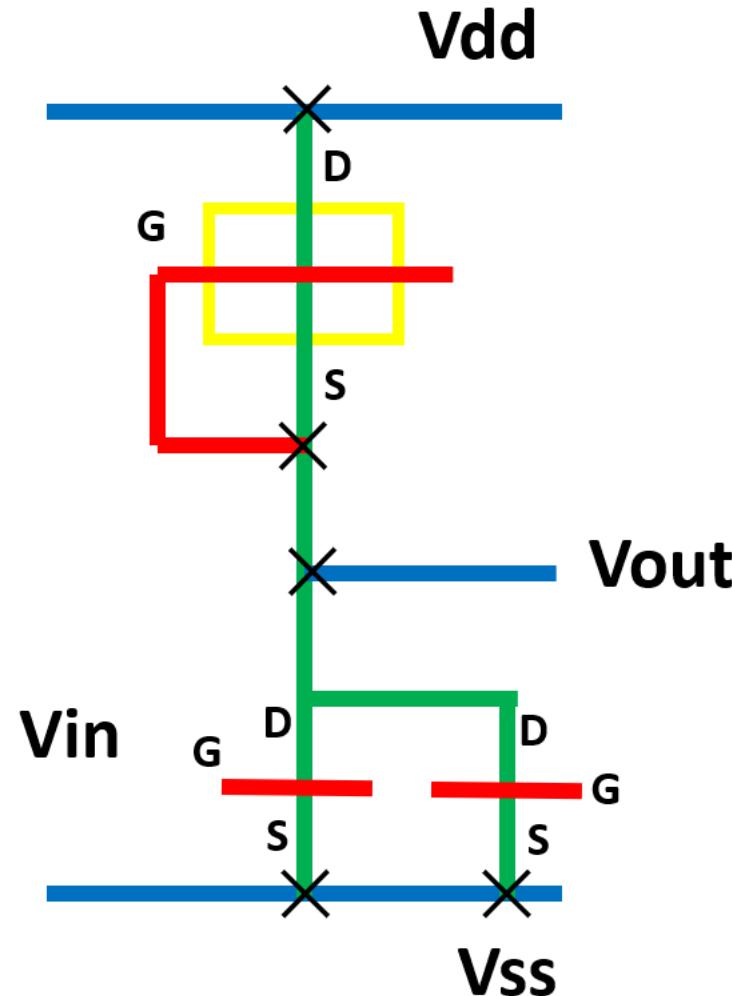
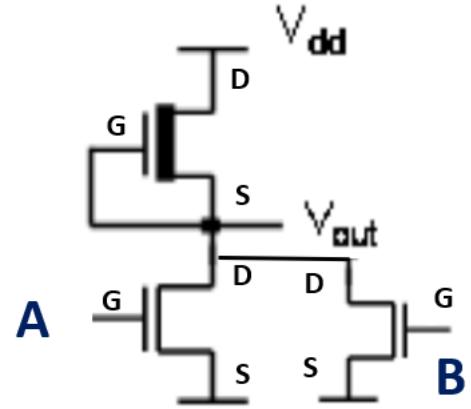
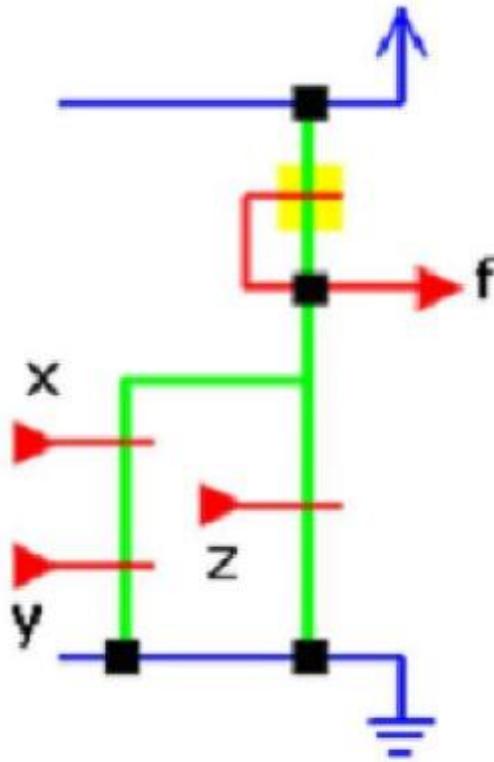


FIGURE 3.16 Two I/P nMOS Nor gate.



Draw stick diagram and layout of nMOS depletion load $F=(XY+Z)'$



Draw layout

figure 7: stick diagram of a given function f.



Sheet resistance (R_s)

Consider a uniform slab of conducting material of resistivity ρ , width W , thickness t , and length between faces L .

$$R_{AB} = \frac{\rho L}{A} \text{ ohm}$$

$$R_{AB} = \frac{\rho L}{tW} \text{ ohm}$$

Now, consider the case in which $L = W$, that is, a square of resistive material, then

$$R_{AB} = \frac{\rho}{t} = R_s$$

where

R_s = ohm per square or sheet resistance

Thus

$$R_s = \frac{\rho}{t} \text{ ohm per square}$$

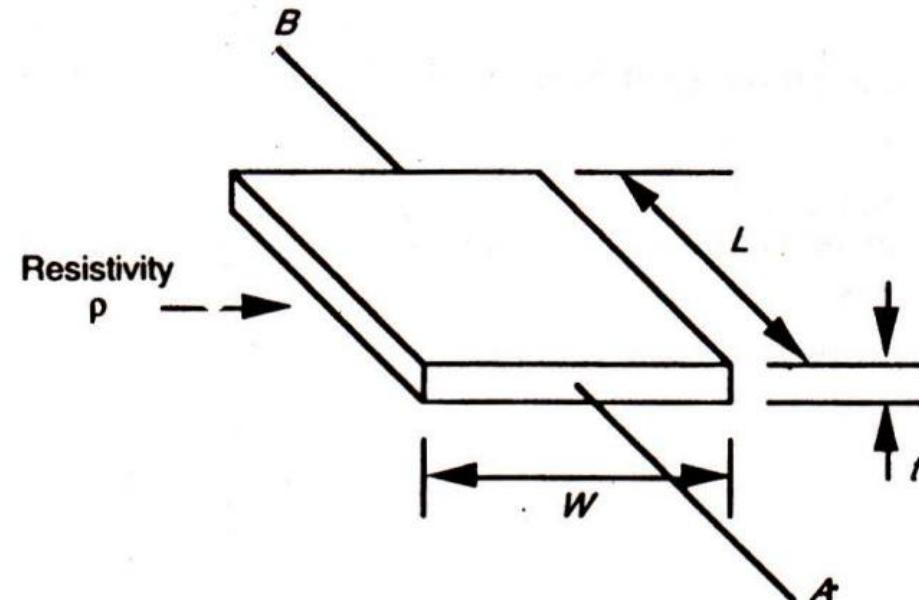


FIGURE 4.1 Sheet resistance model.

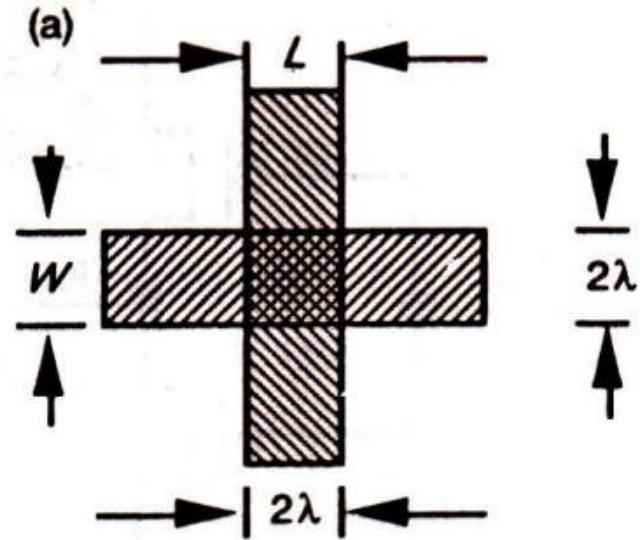
Layer	R_s ohm per square		
	5 μm	2 μm Orbit	Orbit 1.2 μm
Metal	0.03	0.04	0.04
Diffusion (or active)**	10→50	20→45	20→45
Silicide	2→4	—	—
Polysilicon	15→100	15→30	15→30
n-transistor channel	10^4 †	2×10^4 †	2×10^4 †
p-transistor channel	2.5×10^4 †	4.5×10^4 †	4.5×10^4 †



Sheet resistance concept applied to enhancement MOSFETs

The simple n-type pass transistor of Figure has a channel length $L = 2\lambda$ and a channel width $W = 2\lambda$. The channel is, therefore, square and channel resistance (with or without implant).

$$R = 1 \text{ square} \times R_s \frac{\text{ohm}}{\text{square}} = R_s = 10^4 \text{ ohm}^*$$



The length to width (L/W) ratio is 1:1 in this case.

Layer	R _s ohm per square		
	5 μm	2μm Orbit	Orbit 1.2 μm
Metal	0.03	0.04	0.04
Diffusion (or active)**	10→50	20→45	20→45
Silicide	2→4	—	—
Polysilicon	15→100	15→30	15→30
n-transistor channel	10 ^{4†}	2 × 10 ^{4†}	2 × 10 ^{4†}
p-transistor channel	2.5 × 10 ^{4†}	4.5 × 10 ^{4†}	4.5 × 10 ^{4†}



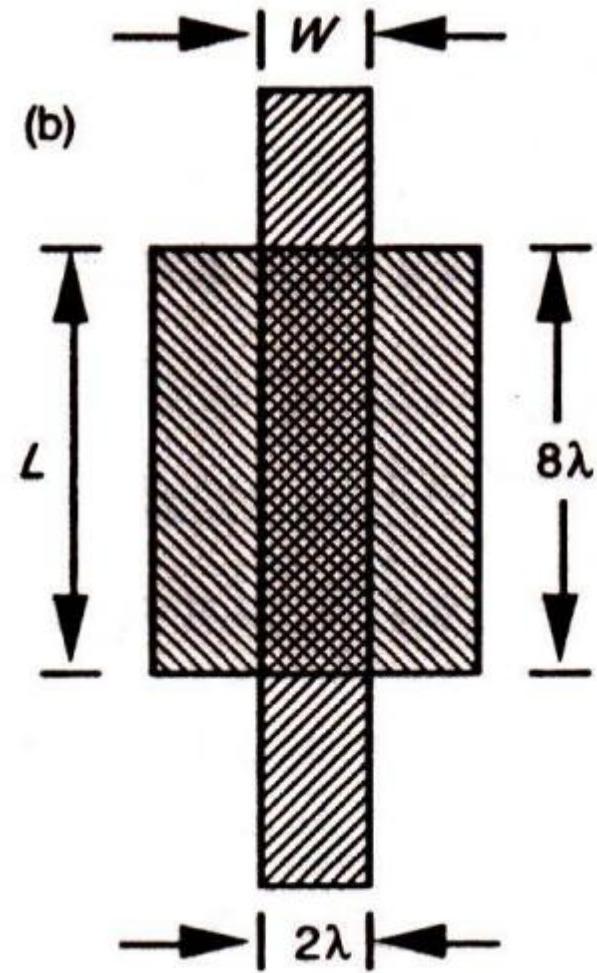
Sheet resistance concept applied to depletion MOSFETs

The transistor structure of Figure 4.2(b) has a channel length $L = 8\lambda$. and width $W = 2\lambda$. Therefore,

$$Z = \frac{L}{W} = 4$$

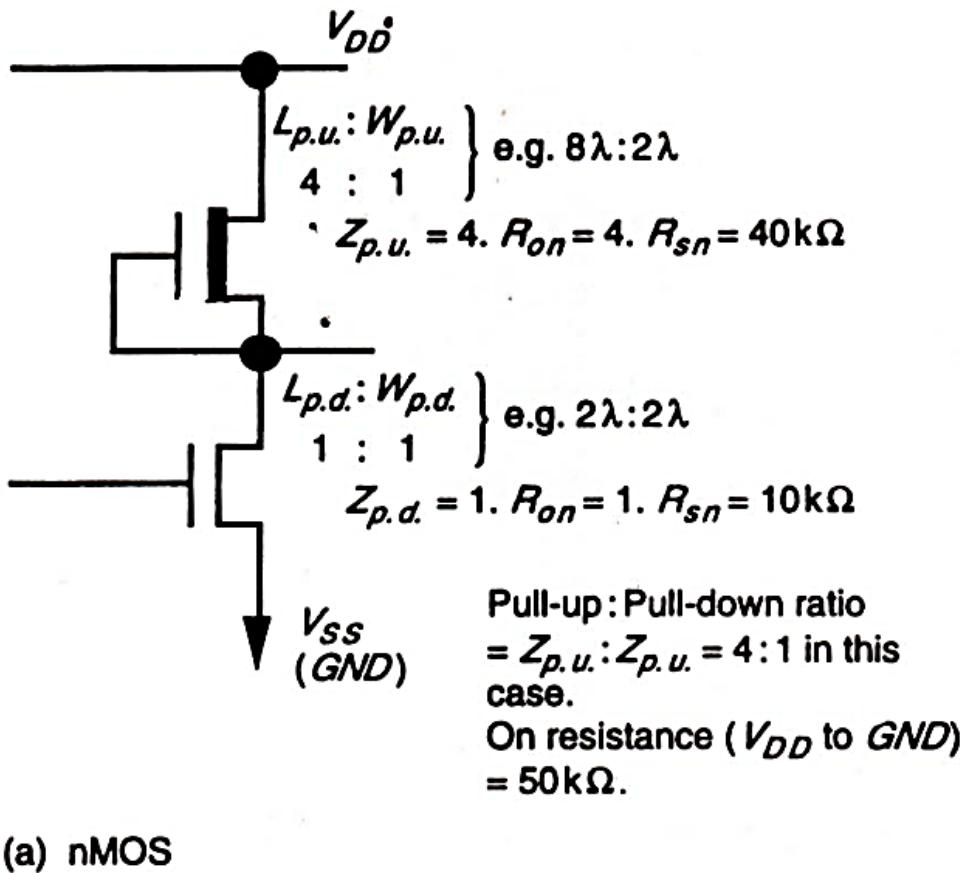
$$R = ZR_s = 4 \times 10^4 \text{ ohm}$$

Layer	<i>R_s ohm per square</i>		
	5 μm	2μm Orbit	Orbit 1.2 μm
Metal	0.03	0.04	0.04
Diffusion (or active)**	10→50	20→45	20→45
Silicide	2→4	—	—
Polysilicon	15→100	15→30	15→30
n-transistor channel	10 ^{4†}	2 × 10 ^{4†}	2 × 10 ^{4†}
p-transistor channel	2.5 × 10 ^{4†}	4.5 × 10 ^{4†}	4.5 × 10 ^{4†}

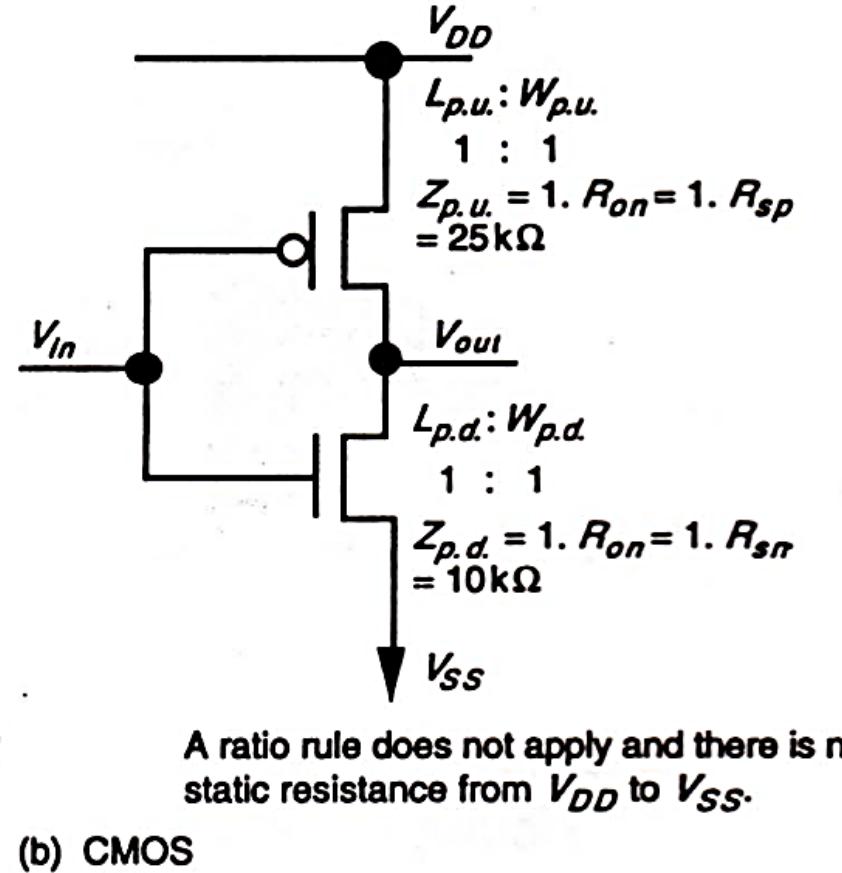




Sheet resistance concept applied to inverters



(a) nMOS



(b) CMOS

Note: R_{on} = 'on' resistance; R_{sn} = n-channel sheet resistance; R_{sp} = p-channel sheet resistance.

FIGURE 4.3 Inverter resistance calculation.



Area Capacitances of layers

For any layer, knowing the dielectric (silicon dioxide) thickness, we can calculate area capacitance as follows:

$$C = \frac{\epsilon_0 \epsilon_{ins} A}{D} \text{ farads}$$

A normal approach is to give layer area capacitances in pF/ μm^2

TABLE 4.2 Typical area capacitance values for MOS circuits

<i>Capacitance</i>	<i>Value in pF $\times 10^{-4}/\mu\text{m}^2$ (Relative values in brackets)</i>		
	<i>5 μm</i>	<i>2 μm</i>	<i>1.2 μm</i>
Gate to channel	4 (1.0)	8 (1.0)	16 (1.0)
Diffusion (active)	1 (0.25)	1.75 (0.22)	3.75 (0.23)
Polysilicon* to substrate	0.4 (0.1)	0.6 (0.075)	0.6 (0.038)
Metal 1 to substrate	0.3 (0.075)	0.33 (0.04)	0.33 (0.02)
Metal 2 to substrate	0.2 (0.05)	0.17 (0.02)	0.17 (0.01)
Metal 2 to metal 1	0.4 (0.1)	0.5 (0.06)	0.5 (0.03)
Metal 2 to polysilicon	0.3 (0.075)	0.3 (0.038)	0.3 (0.018)



STANDARD UNIT OF CAPACITANCE $\square C_g$

The unit is denoted $\square C_g$ and is defined the gate-to-channel capacitance of a MOS transistor having $W = L =$ feature size, that is, a 'standard' or 'feature size' square

$\square C_g$ may be evaluated for any MOS process. For example, for 5 μm MOS circuits:

$$\text{Area/standard square} = 5 \mu\text{m} \times 5 \mu\text{m} = 25 \mu\text{m}^2 \text{ (= area of minimum size transistor)}$$

$$\text{Capacitance value (from Table 4.2)} = 4 \times 10^{-4} \text{ pF}/\mu\text{m}^2$$

$$\text{Thus, standard value } \square C_g = 25 \mu\text{m}^2 \times 4 \times 10^{-4} \text{ pF}/\mu\text{m}^2 = .01 \text{ pF}$$

or, for 2 μm MOS circuits (Orbit):

$$\text{Area/standard square} = 2 \mu\text{m} \times 2 \mu\text{m} = 4 \mu\text{m}^2$$

$$\text{Gate capacitance value (from Table 4.2)} = 8 \times 10^{-4} \text{ pF}/\mu\text{m}^2$$

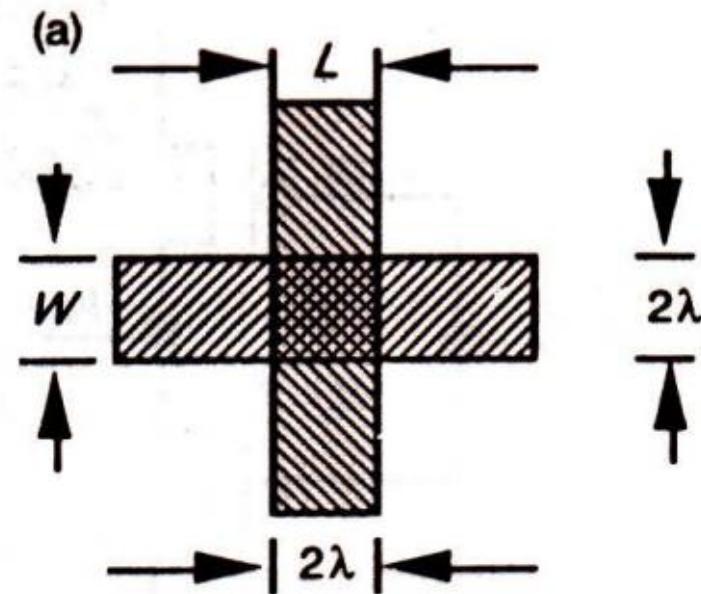
$$\text{Thus, standard value } \square C_g = 4 \mu\text{m}^2 \times 8 \times 10^{-4} \text{ pF}/\mu\text{m}^2 = .0032 \text{ pF}$$

and, for 1.2 μm MOS circuits (Orbit):

$$\text{Area/standard square} = 1.2 \mu\text{m} \times 1.2 \mu\text{m} = 1.44 \mu\text{m}^2$$

$$\text{Gate capacitance value (from Table 4.2)} = 16 \times 10^{-4} \text{ pF}/\mu\text{m}^2$$

$$\text{Thus, standard value } \square C_g = 1.44 \mu\text{m}^2 \times 16 \times 10^{-4} \text{ pF}/\mu\text{m}^2 = .0023 \text{ pF}$$





Calculation of Area Capacitance

- The calculation of capacitance values may now be undertaken by establishing the ratio between the area of interest and the area of standard (feature size square) gate ($2\lambda \times 2\lambda$) and multiplying this ratio by the appropriate relative C value from Table 4.2.
- The product will give the required capacitance in $\square\text{Cg}$ units.

Consider the area defined in Figure 4.4. First, we must calculate the area relative to that of a standard gate.

$$\text{Relative area} = \frac{20\lambda \times 3\lambda}{2\lambda \times 2\lambda} = 15$$

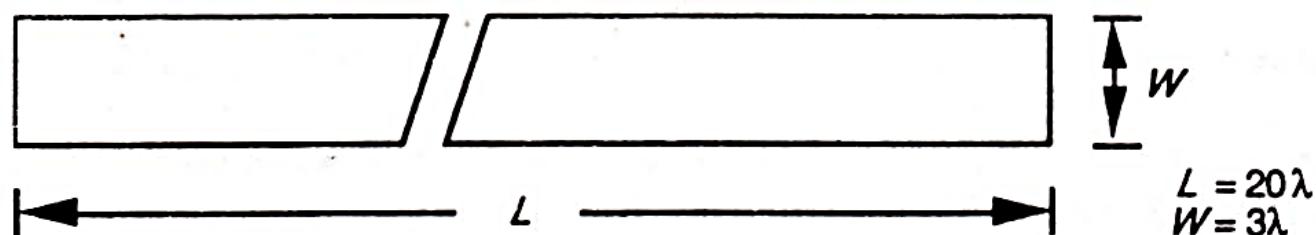


FIGURE 4.4 Simple area for capacitance calculation.



TABLE 4.2 Typical area capacitance values for MOS circuits

Capacitance	Value in $pF \times 10^{-4}/\mu m^2$ (Relative values in brackets)		
	5 μm	2 μm	1.2 μm
Gate to channel	4 (1.0)	8 (1.0)	16 (1.0)
Diffusion (active)	1 (0.25)	1.75 (0.22)	3.75 (0.23)
Polysilicon* to substrate	0.4 (0.1)	0.6 (0.075)	0.6 (0.038)
Metal 1 to substrate	0.3 (0.075)	0.33 (0.04)	0.33 (0.02)
Metal 2 to substrate	0.2 (0.05)	0.17 (0.02)	0.17 (0.01)
Metal 2 to metal 1	0.4 (0.1)	0.5 (0.06)	0.5 (0.03)
Metal 2 to polysilicon	0.3 (0.075)	0.3 (0.038)	0.3 (0.018)

Now:

1. Consider the area in metal 1.

Capacitance to substrate = relative area \times relative C value

$$= 15 \times 0.0750 \square C_g$$

$$= 1.125 \square C_g$$

That is, the defined area in metal has a capacitance to substrate 1.125 times that of a feature size square gate area.

2. Consider the same area in polysilicon.

$$\text{Capacitance to substrate} = 15 \times 0.1 \square C_g$$

$$= 1.5 \square C_g$$

3. Consider the same area in n-type diffusion.

$$\text{Capacitance to substrate} = 15 \times 0.25 \square C_g$$

$$= 3.75 \square C_g^*$$

$$\text{Relative area} = \frac{20\lambda \times 3\lambda}{2\lambda \times 2\lambda} = 15$$

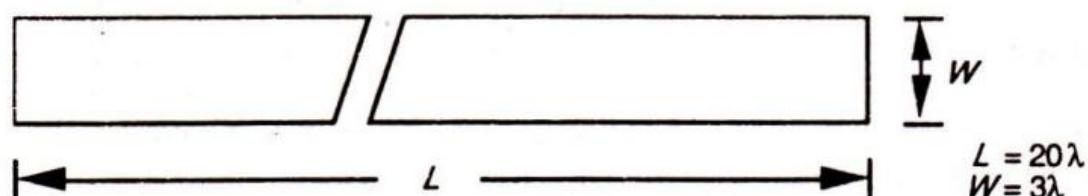


FIGURE 4.4 Simple area for capacitance calculation.

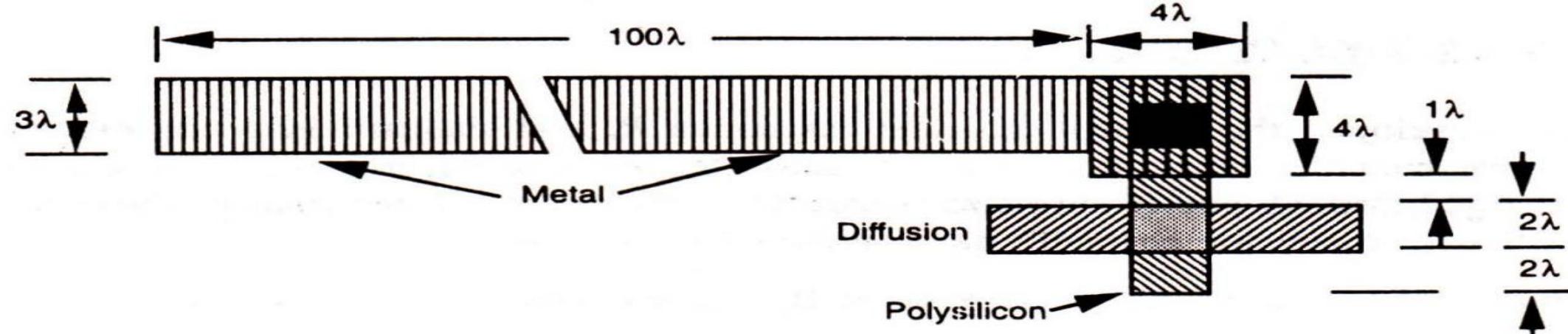


FIGURE 4.5 Capacitance calculation (multilayer).

Calculations of area capacitance values associated with structures occupying more than one layer, as in Figure 4.5,

Consider the metal area (less the contact region where the metal is connected to polysilicon and shielded from the substrate)

$$\text{Ratio} = \frac{\text{Metal area}}{\text{Standard gate area}} = \frac{100\lambda \times 3\lambda}{4\lambda^2} = 75$$

$$\text{Metal capacitance } C_m = 75 \times 0.075 = 5.625 \square C_g$$

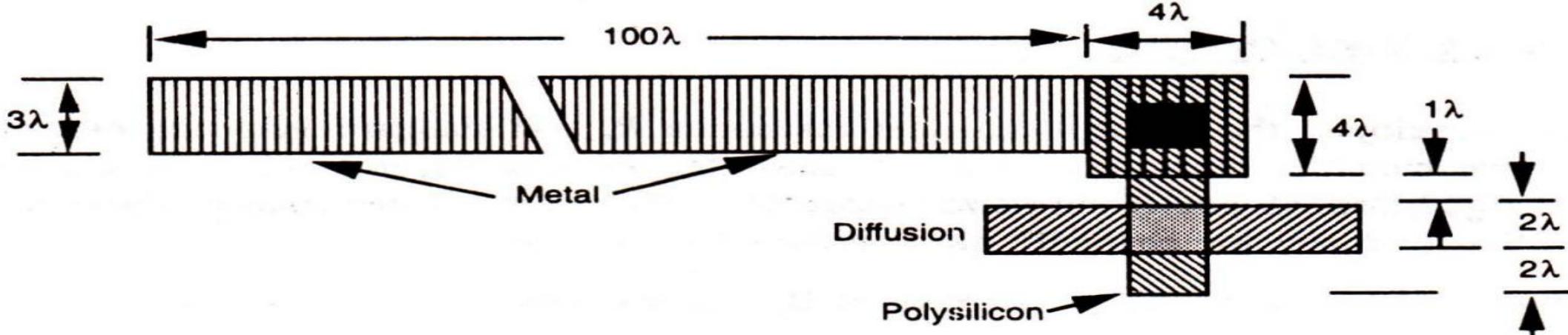


FIGURE 4.5 Capacitance calculation (multilayer).

Consider the polysilicon area (excluding the gate region)

$$\text{Polysilicon area} = 4\lambda \times 4\lambda + 3\lambda \times 2\lambda = 22\lambda^2$$

Therefore

$$\text{Polysilicon capacitance } C_p = \frac{22}{4} \times 0.1 = .55 \square C_g$$

For the transistor,

$$\text{Gate capacitance } C_g = 1 \square C_g$$

$$\boxed{\text{Total capacitance } C_T = C_m + C_p + C_g \doteq 7.20 \square C_g}$$



The Delay Unit τ

- We have developed the concept of sheet resistance R_s and standard gate capacitance unit $\square C_g$
- **Time constant:** It is time required to charge one standard (feature size square) gate area capacitance through one feature size square of n channel resistance (that is, through R_s for an nMOS pass transistor channel),

Time constant $\tau = (1R_s \text{ (n channel)} \times 1\square C_g)$ seconds

If we consider the case of one standard (feature size square) gate area capacitance being charged through one feature size square of n channel resistance (that is, through R_s for an nMOS pass transistor channel), as in Figure 4.6, we have:

This can be evaluated for any technology and for 5 μm technology,

$$\tau = 10^4 \text{ ohm} \times 0.01 \text{ pF} = 0.1 \text{ nsec}$$

and for 2 μm (Orbit) technology,

$$\tau = 2 \times 10^4 \text{ ohm} \times 0.0032 \text{ pF} = 0.064 \text{ nsec}$$

and for 1.2 μm (Orbit) technology,

$$\tau = 2 \times 10^4 \text{ ohm} \times 0.0023 \text{ pF} = 0.046 \text{ nsec}$$

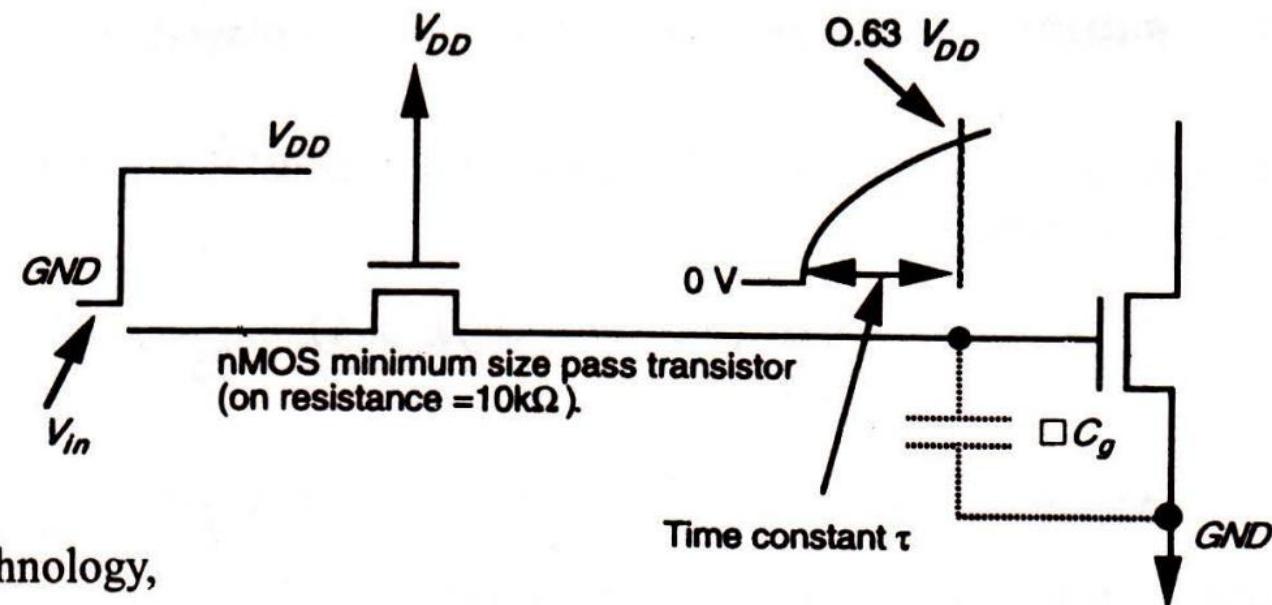


FIGURE 4.6 Model for derivation of τ .



Note that τ thus obtained is not much different from transit time τ_{sd} calculated from equation (2.2).

$$\tau_{sd} = \frac{L^2}{\mu_n V_{ds}}$$

Note that V_{ds} varies as C_g charges from 0 volts to 63% of V_{DD} in period τ in Figure 4.6, so that an appropriate value for V_{ds} is the average value = 3 volts. For 5 μm technology, then,

$$\begin{aligned}\tau_{sd} &= \frac{25 \mu\text{m}^2 \text{ V sec}}{650 \text{ cm}^2 \text{ 3 V}} \times \frac{10^9 \text{ nsec cm}^2}{10^8 \mu\text{m}^2} \\ &= 0.13 \text{ nsec}\end{aligned}$$

This is very close to the theoretical time constant τ calculated above.



Estimation of delay in NMOS and CMOS inverters



nMOS inverter delays

INV-1

When $V_{in} = 1$,
enh - ON

$$\tau_1 = \frac{R_s}{2} \times 1 \square C_g$$

$$\tau_1 = R_s \square C_g$$

$$\tau_1 = 1 \tau$$

INV-2

When $V_{in} = 0$ v. then enhancement trans. is in OFF state.

then depletion mode comes into action.

$$\tau_2 = z \cdot R_s \times 1 \square C_g$$

$$\approx 4 R_s \times 1 \square C_g$$

$$= 4 \cdot R_s \square C_g$$

$$= 4 \tau$$

In general, the delay through a pair of similar nMOS inverters is

$$T_d = (1 + Z_{p.u.}/Z_{p.d.})\tau$$

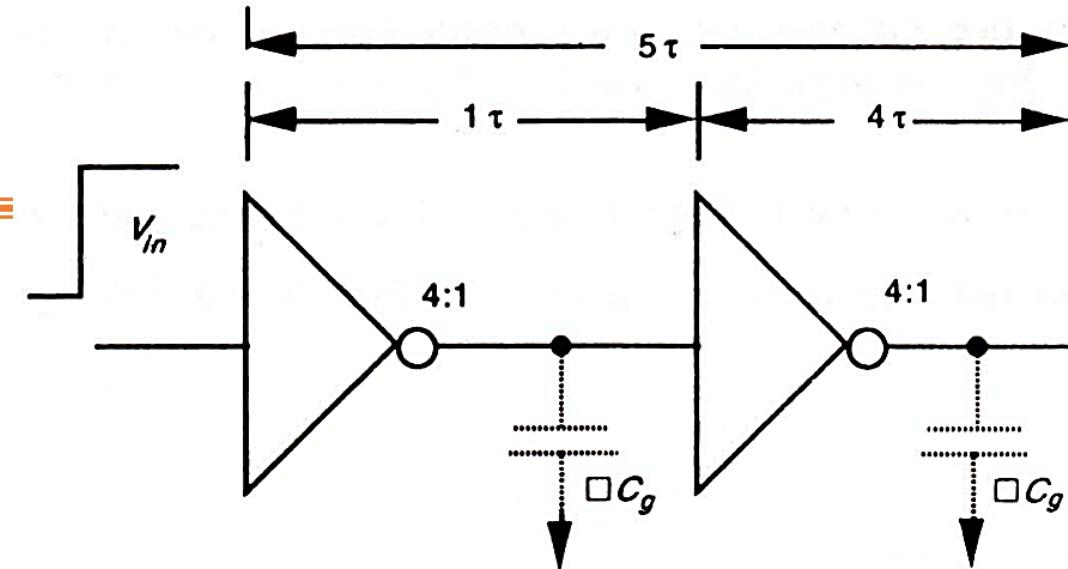
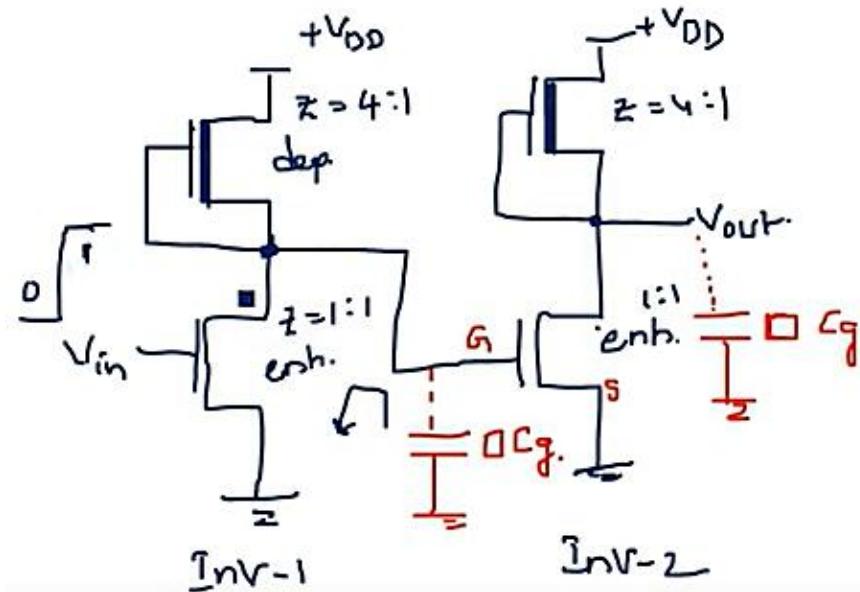


FIGURE 4.7 nMOS inverter pair delay.





CMOS inverter delays

1. For asymmetric CMOS inverter, assuming the width of nMOS and pMOS to be equal, the input gate capacitance will be $2\square C_g$
2. For symmetric CMOS inverter, i.e. for $\beta_n = \beta_p$, the input capacitance is,

$$1\square C_g \text{ (n-device)} + 2.5\square C_g \text{ (p-device)} = 3.5\square C_g$$

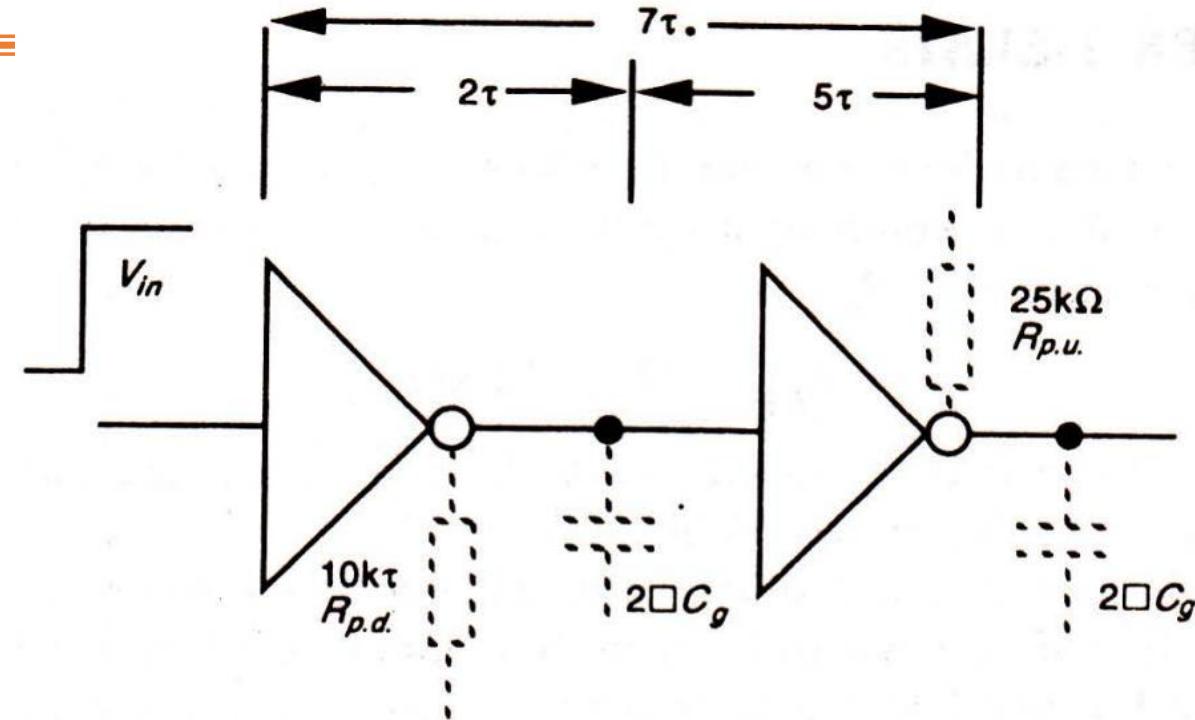


FIGURE 4.8 Minimum size CMOS inverter pair delay.



A more formal estimation of CMOS Inverter delay

A CMOS inverter, in general, either charges or discharges a capacitive load C_L and rise-time τ_r or fall-time τ_f can be estimated from the following simple analysis.

$$\beta = \frac{\mu C_{ox} W}{L}$$

1. Rise-time estimation

- In this analysis we assume that the p-device stays in saturation for the entire charging period of the load capacitor C_L . The circuit may then be modeled as in Figure 4.9.
- The saturation current for the p-transistor is given by

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

This current charges C_L and, since its magnitude is approximately constant, we have

$$V_{out} = \frac{I_{dsp} t}{C_L}$$

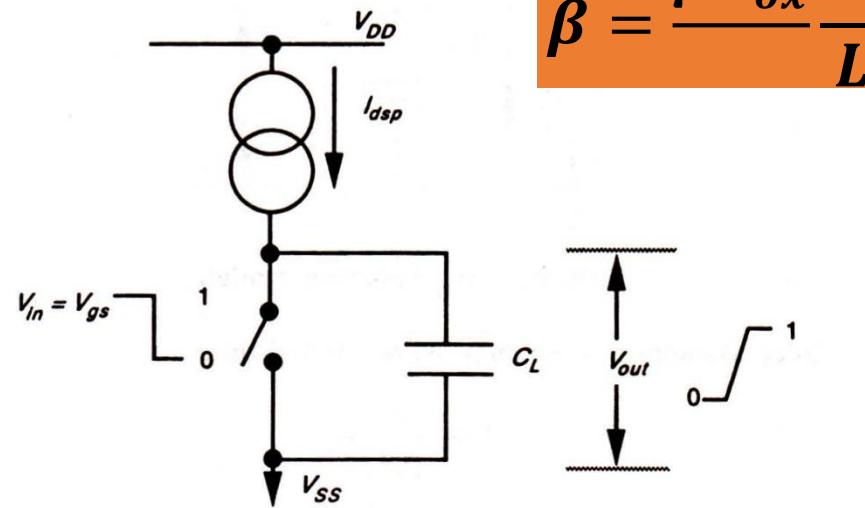
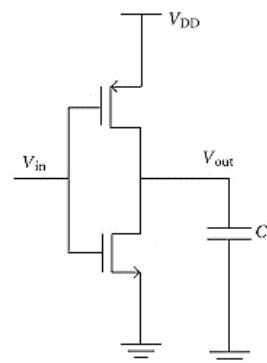


FIGURE 4.9 Rise-time model.





Substituting for I_{dsp} and rearranging we have

$$\beta = \frac{\mu C_{ox}}{L} W$$

$$t = \frac{2C_L V_{out}}{\beta_p (V_{gs} - |V_{tp}|)^2}$$

We now assume that $t = \tau_r$ when $V_{out} = +V_{DD}$, so that

$$\tau_r = \frac{2V_{DD}C_L}{\beta_p (V_{DD} - |V_{tp}|)^2}$$

with $|V_{tp}| = 0.2V_{DD}$, then

$$\tau_r \doteq \frac{3C_L}{\beta_p V_{DD}}$$

2. Fall time estimation

Similar reasoning can be applied to the discharge of C_L through the n-transistor. The circuit model in this case is given as Figure 4.10.

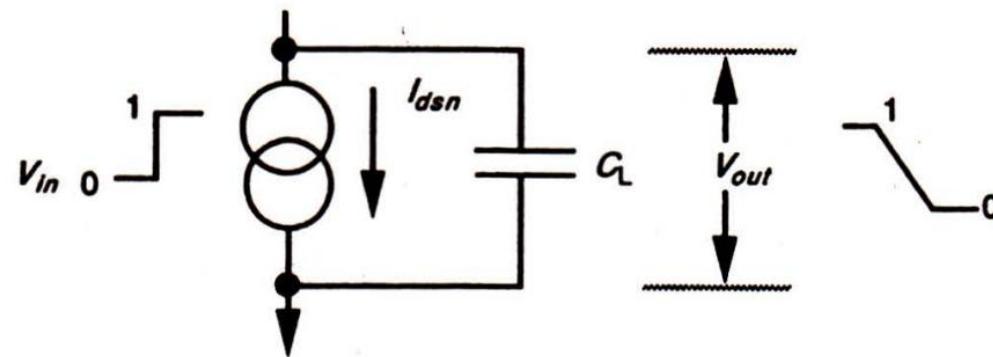


FIGURE 4.10 Fall-time model.

Making similar assumptions we may write for fall-time:

$$\tau_f \doteq \frac{3C_L}{\beta_n V_{DD}}$$



4.7.1.3 Summary of CMOS rise and fall factors

Using these expressions we may deduce that:

$$\beta = \frac{\mu C_{ox} W}{L}$$

$$\frac{\tau_r}{\tau_f} = \frac{\beta_n}{\beta_p}$$

But $\mu_n = 2.5 \mu_p$ and hence $\beta_n \approx 2.5\beta_p$, so that the rise-time is slower by a factor of 2.5 when using minimum size devices for both 'n' and 'p'.

In order to achieve symmetrical operation using minimum channel length, we would need to make $W_p = 2.5W_n$ and for minimum size lambda-based geometries this would result in the inverter having an input capacitance of $1\square C_g$ (n-device) + $2.5\square C_g$ (p-device) = $3.5\square C_g$ in total.



Driving Large Capacitive loads

- The problem of driving comparatively large capacitive loads arises when signals must be propagated from the on-chip to off-chip destinations.
- Generally, typical off chip capacitances may be several orders higher than on-chip $\square C_g$ values.
- If the off-chip load is denoted C_L then, $C_L \geq 10^4 \square C_g$ (typically)
- Clearly capacitances of this order must be driven through low resistances, otherwise excessively long delays will occur.
- Three methods to derive large capacitance,
 1. Cascaded Inverters as Drivers
 2. Super buffers
 3. Bi-CMOS drivers



1. Cascaded Inverters as Drivers

- For MOS circuits, low resistance values for $Z_{p.d.}$ and $Z_{p.u.}$ imply low L: W ratios; in other words, channels must be made very wide to reduce resistance value and, in consequence, an inverter to meet this need occupies a large area.
- Clearly, as the width factor increases, so the capacitive load presented at the inverter input increases, and the area occupied increases also.
- Equally clearly, the rate at which the width increases (that is, the value of f) will influence the number N of stages which must be cascaded to drive a particular value of C_L .

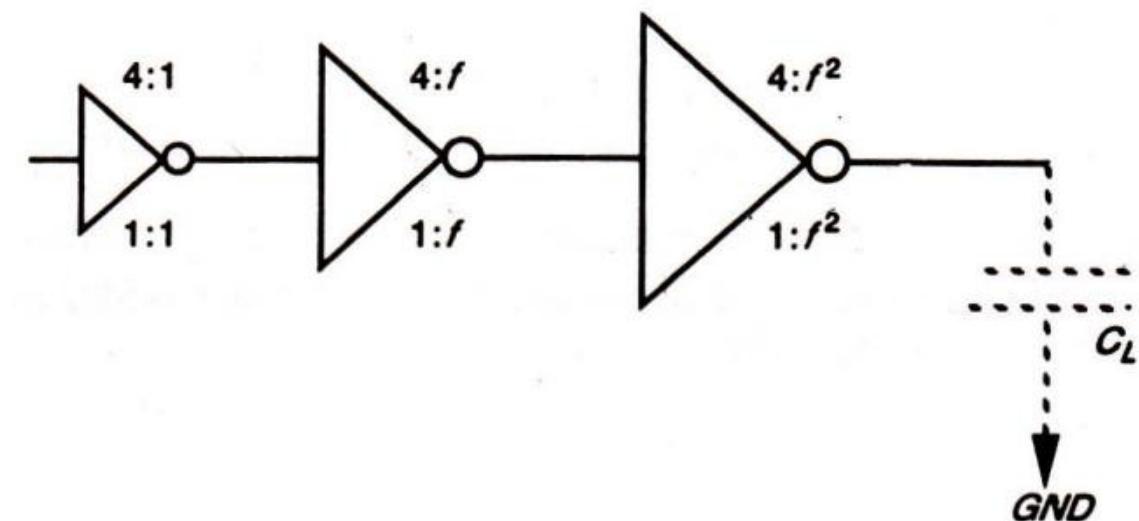


FIGURE 4.11 Driving large capacitive loads.



With large f , N decreases but delay per stage increases. For 4:1 nMOS inverters

$$\left. \begin{array}{l} \text{delay per stage} = f\tau \text{ for } \Delta V_{in} \\ \text{or} = 4f\tau \text{ for } \nabla V_{in} \end{array} \right\} \quad \text{where } \Delta V_{in} \text{ indicates logic 0 to 1 transition and } \nabla V_{in} \text{ indicates logic 1 to 0 transition of } V_{in}$$

Therefore, total delay per nMOS pair = $5f\tau$. A similar treatment yields delay per CMOS pair = $7f\tau$. Now let

$$y = \frac{C_L}{\square C_g} = f^N$$

so that the choice of f and N are interdependent.

$$\ln(y) = N \ln(f)$$

$$N = \frac{\ln(y)}{\ln(f)}$$



Thus, for N even

$$\text{total delay} = \frac{N}{2} 5f\tau = 2.5 Nf\tau \text{ (nMOS)}$$

$$\text{or} = \frac{N}{2} 7f\tau = 3.5 Nf\tau \text{ (CMOS)}$$

Thus, in all cases

$$\text{delay} \propto Nf\tau = \frac{\ln(y)}{\ln(f)} f\tau$$

Thus, assuming that $f = e$, we have

$$\text{Number of stages } N = \ln(y)$$

and overall delay t_d

$$N \text{ even: } t_d = 2.5eN \tau \text{ (nMOS)}$$

$$\text{or } t_d = 3.5eN \tau \text{ (CMOS)}$$

$$\left. \begin{array}{l} N \text{ odd: } t_d = [2.5(N - 1) + 1]e\tau \text{ (nMOS)} \\ \text{or } t_d = [3.5(N - 1) + 2]e\tau \text{ (CMOS)} \end{array} \right\} \text{for } \Delta V_{in}$$

or

$$\left. \begin{array}{l} t_d = [2.5(N - 1) + 4]e\tau \text{ (nMOS)} \\ \text{or } t_d = [3.5(N - 1) + 5]e\tau \text{ (CMOS)} \end{array} \right\} \text{for } \nabla V_{in}$$

It can be shown that total delay is minimized if f assumes the value e (base of natural logarithms); that is, each stage should be approximately 2.7 times wider than its predecessor. This applies to CMOS as well as nMOS inverters.



2. Super buffers

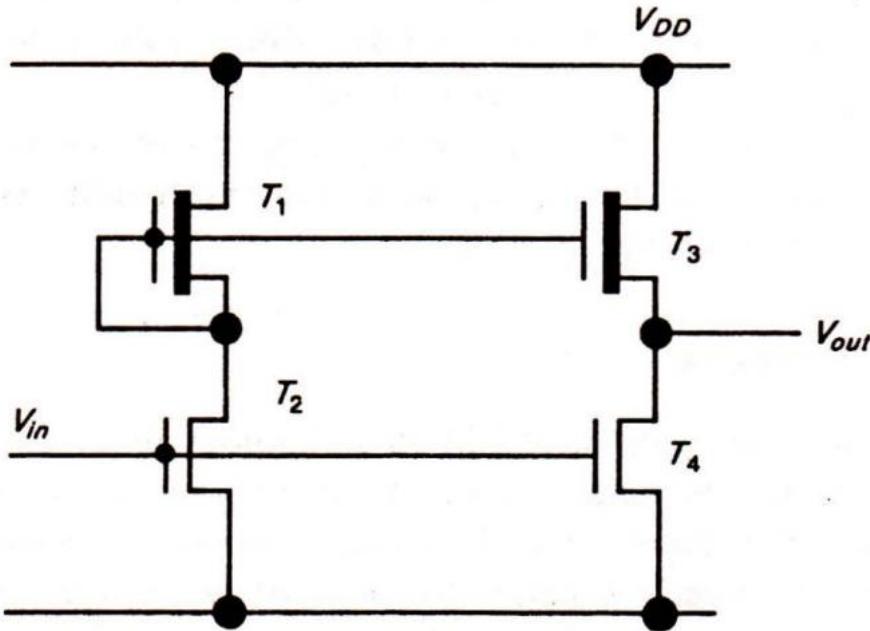


FIGURE 4.12 Inverting type nMOS super buffer.

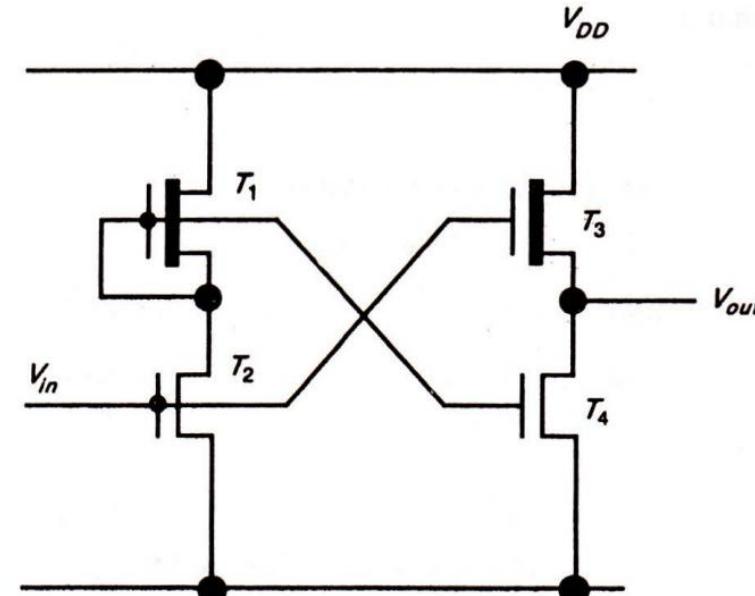


FIGURE 4.13 Non-inverting type nMOS super buffer.

- Since $I_{ds} \propto V_{gs}$ then, doubling the effective V_{gs} will increase the current and thus reduce the delay in charging any capacitance on the output, so that more symmetrical transitions are achieved.
- The structures shown when realized in 5 μm technology are capable of driving loads of 2 pF with 5 nsec rise-time.



3. Bi-CMOS drivers

- BJTs have better transconductance g_m and current/area (I/A) than MOSFETs. This indicates high current drive capabilities for small areas in silicon.
- BJTs have better swinging performance as compared to the MOSFETs.

It may be shown that the time Δt necessary to change the output voltage V_{out} by an amount equal to the input voltage V_{in} is given by

$$\Delta t = \frac{C_L}{g_m}$$

where g_m is the transconductance of the bipolar transistor.

Clearly, since the bipolar transistor has a relatively high transconductance, the value of Δt is small.

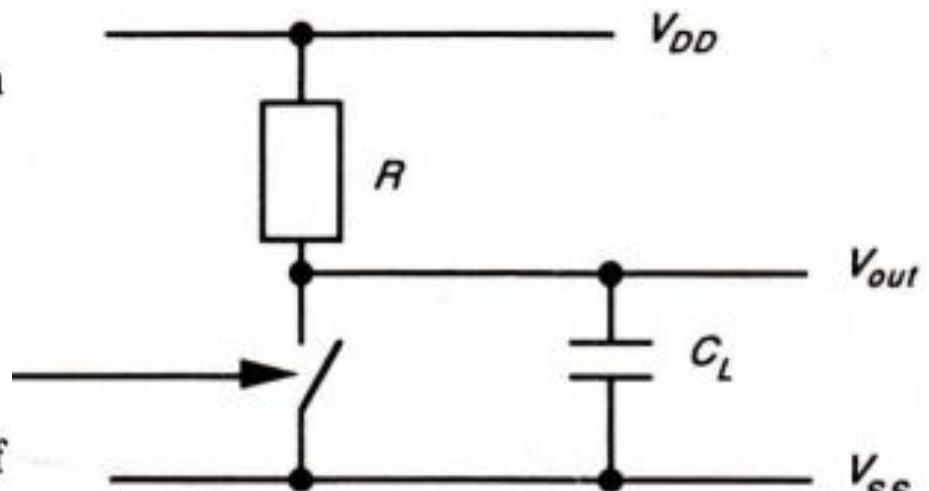


FIGURE 4.14 Driving ability of bipolar transistor.

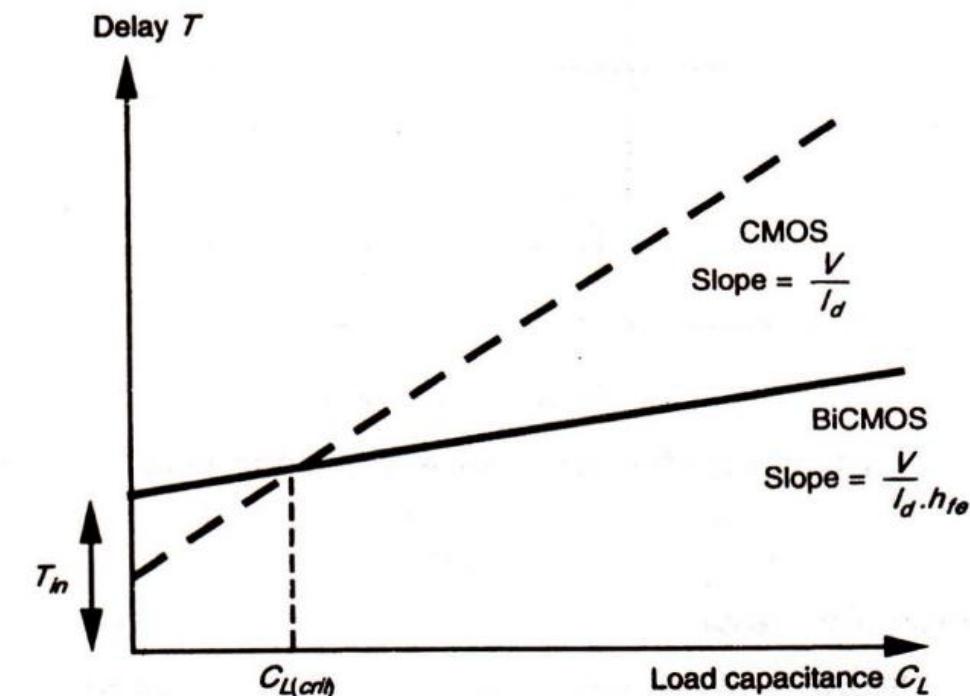
T_{in} - an initial time necessary to charge the base emitter junction of the bipolar (npn) transistor.

For BJT, $T_{in} = 2\text{ns}$

For MOSFET $T_{in} = 1\text{ns}$

T_L - the time taken to charge the output load capacitance C_L

- It will be seen that there is a critical value of load capacitance $C_{L(\text{cric})}$ below which the BiCMOS driver is slower than a comparable CMOS driver.



- Delay of BiCMOS inverter can be described by

$$T = T_{in} + (V/I_d) (1/h_{fe}) C_L$$

where

T_{in} = time to charge up base/emitter junction

h_{fe} = transistor current gain (common emitter)

- Delay for BiCMOS inverter is reduced by a factor of h_{fe} compared with a CMOS inverter.

FIGURE 4.15 Delay estimation.

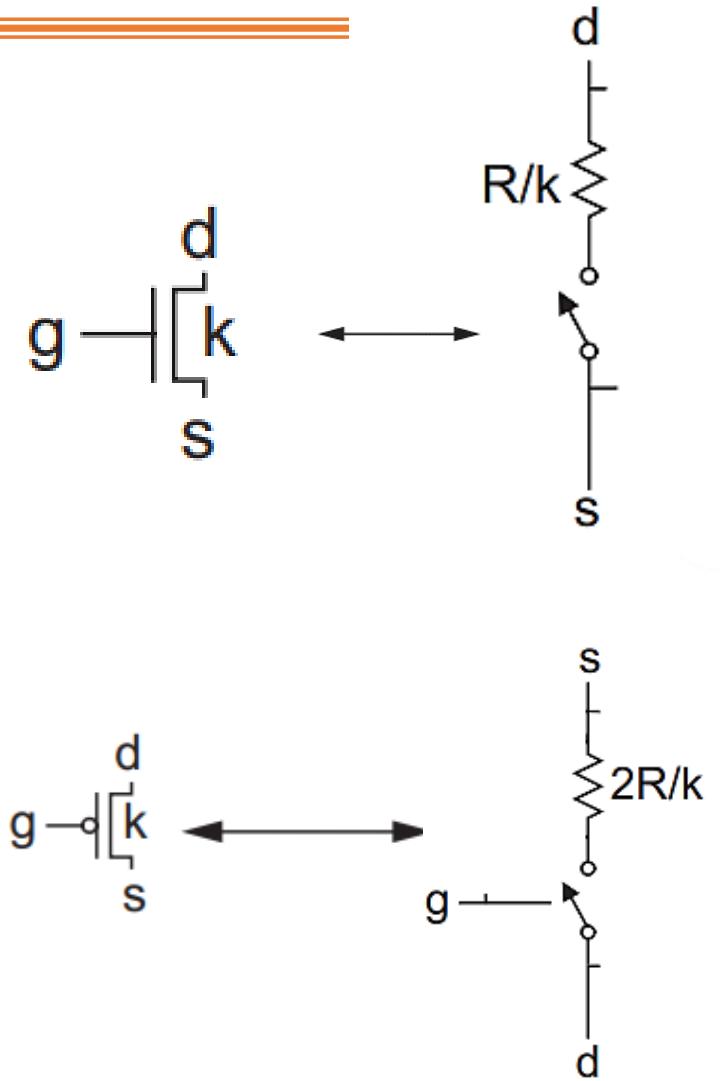


RC delay model

RC delay models approximate the nonlinear transistor I-V and C-V characteristics with an average resistance and capacitance over the switching range of the gate.

1. Effective resistance:

- The RC delay model treats a transistor as a switch in series with a resistor.
- A unit nMOS transistor is defined to have effective resistance R .
- An nMOS transistor of k times unit width has resistance R/k .
- A unit pMOS transistor has greater resistance, generally in the range of $2R$.
- A pMOS transistor of k times unit width has resistance $2R/k$.





2. Gate and diffusion capacitances

- Each transistor has gate and diffusion capacitance.
- We define C to be the gate capacitance of a unit transistor of either flavor.
- A transistor of k times unit width has capacitance **kC** .
- Wider transistors have proportionally greater diffusion capacitance.
- Increasing channel length increases gate capacitance proportionally but does not affect diffusion capacitance.

	Resistance	Capacitance (both gate and diffusion capacitances)
nMOS	R/k	kC
pMOS	$2R/k$	kC

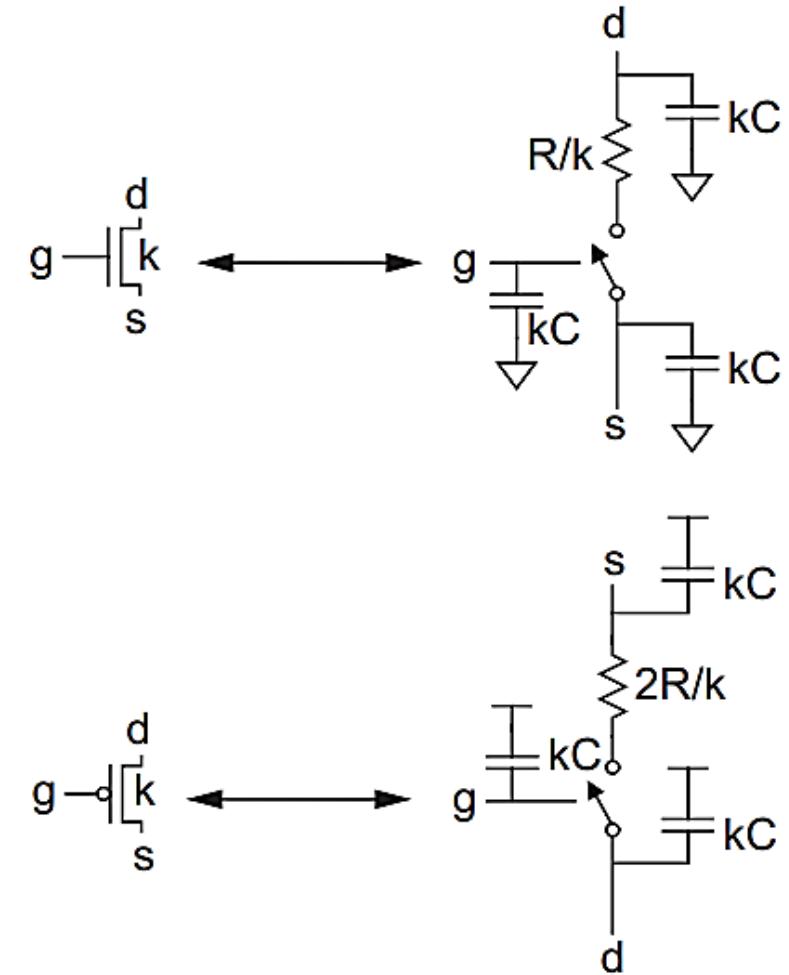


FIGURE 4.5
Equivalent circuits for transistors



3. Equivalent RC circuits

- Figure 4.5 shows the RC equivalent circuit of nMOS and pMOS.
- Figure 4.6 shows the equivalent circuit for a fanout-of-1 inverter with negligible wire capacitance

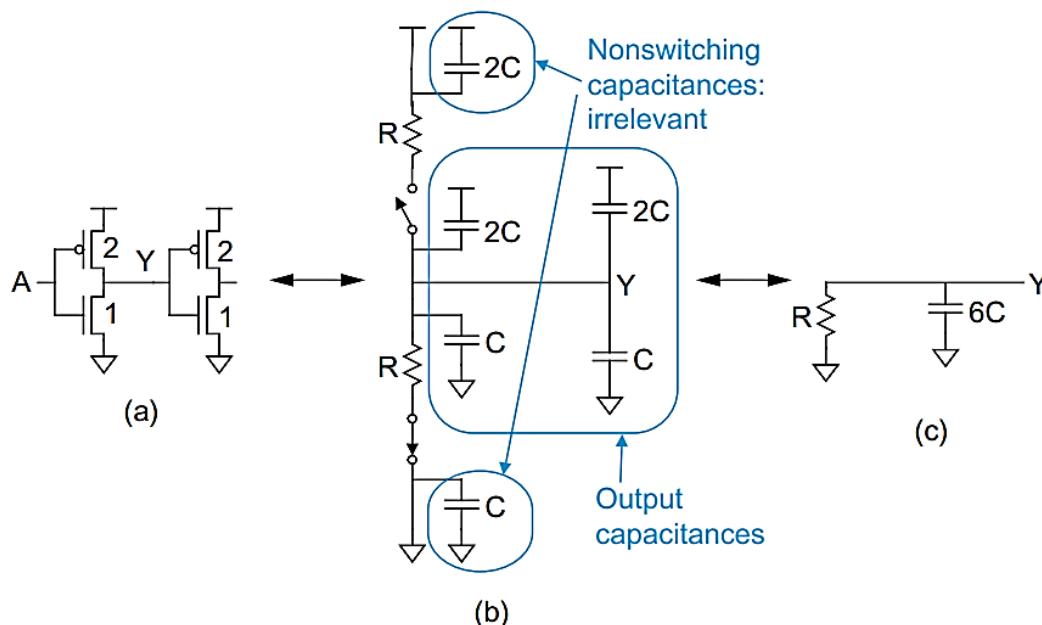


FIGURE 4.6 Equivalent circuit for an inverter

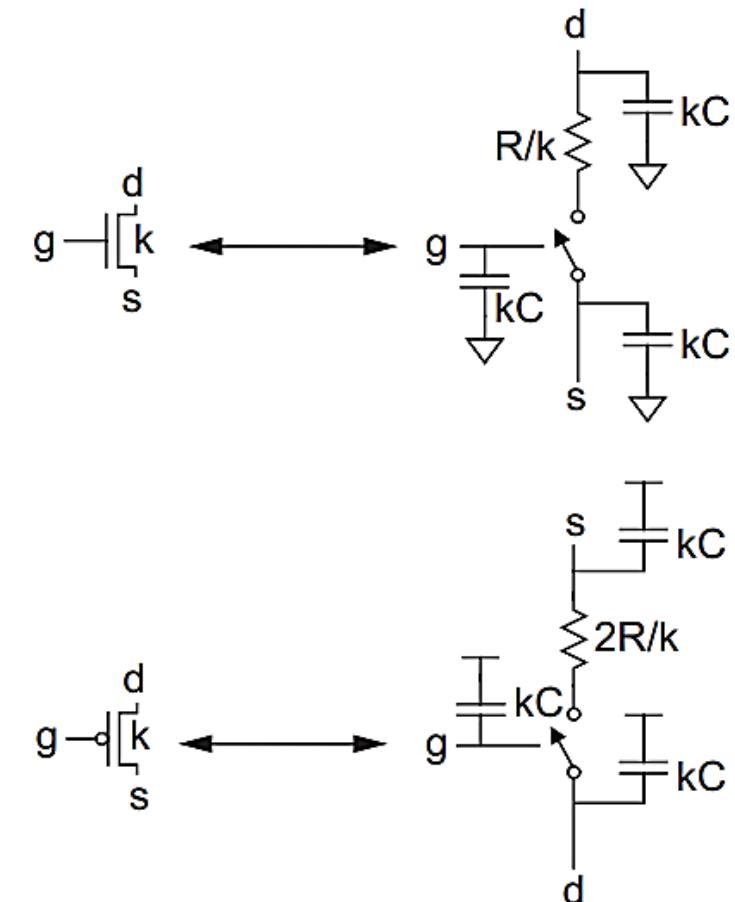
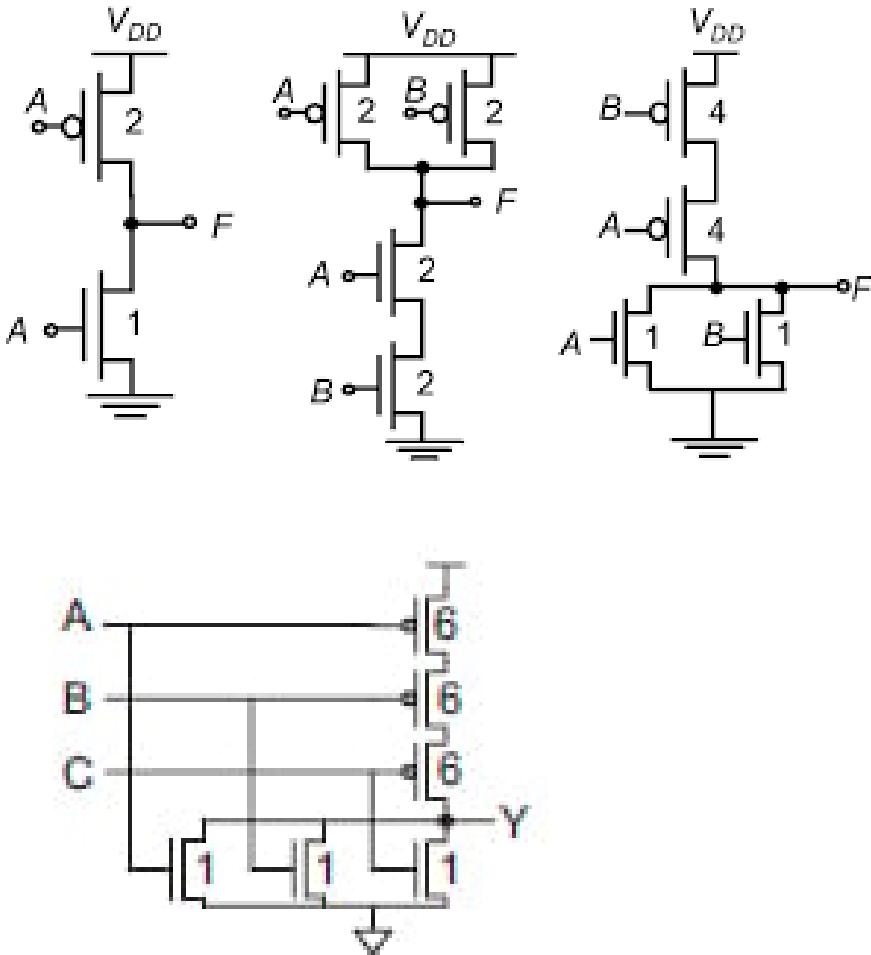


FIGURE 4.5
Equivalent circuits for transistors



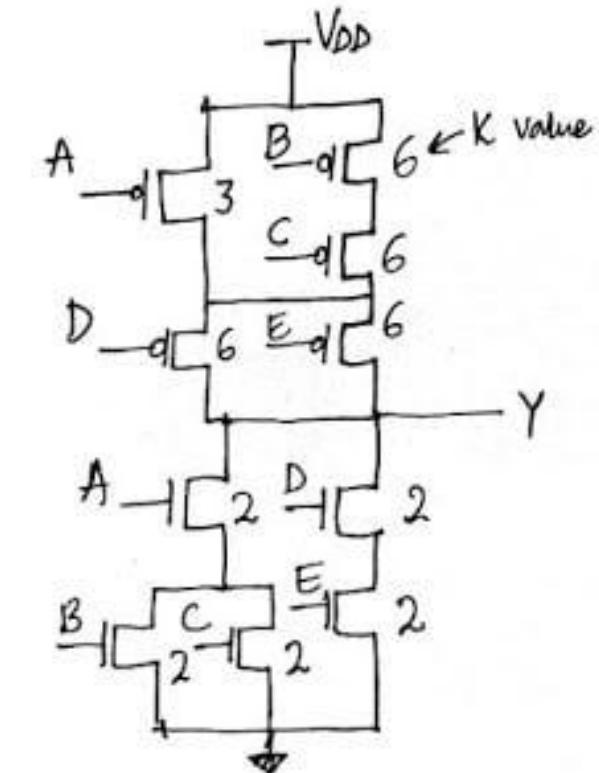
Transistor sizing



- for symmetrical response (dc, ac)
- for performance

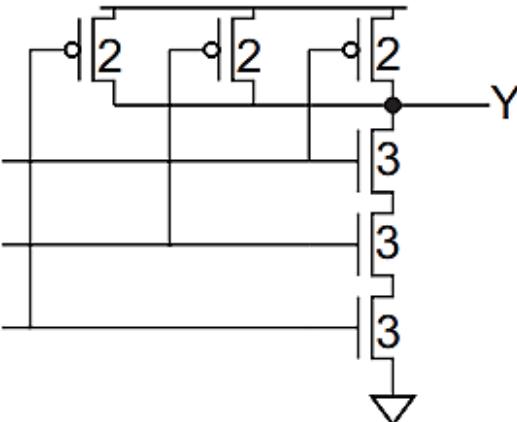
Input Dependent
Focus on worst-case

Numbers indicate
transistor sizing
with minimum size
equal to 1

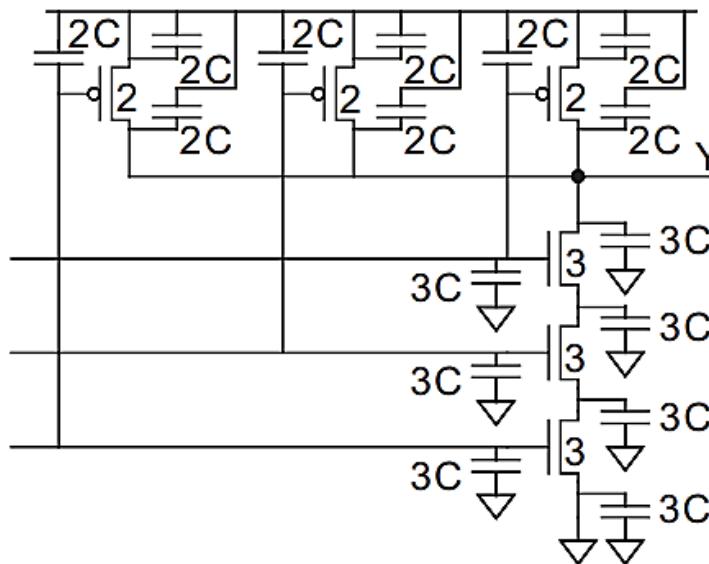




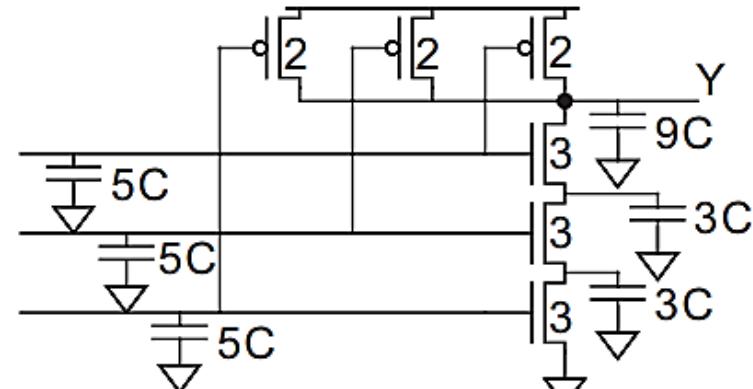
Sketch a 3-input NAND gate with transistor widths chosen to achieve effective rise and fall resistance equal to that of a unit inverter (R). Annotate the gate with its gate and diffusion capacitances. Assume all diffusion nodes are contacted. Then sketch equivalent circuits for the falling output transition and for the worst-case rising output transition.



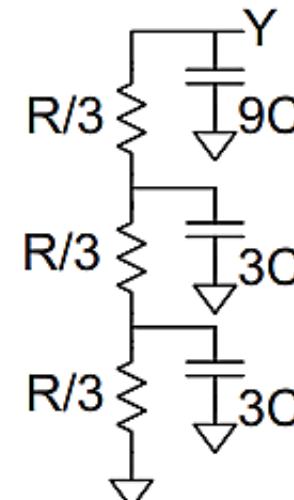
(a)



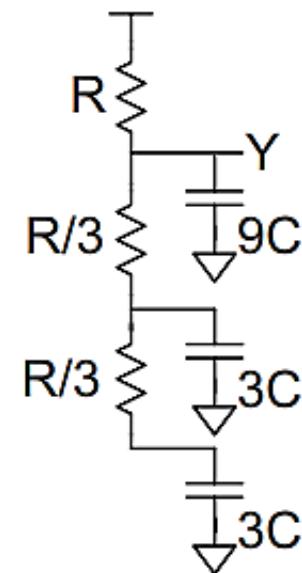
(b)



(c)



Falling



Rising

(d)

(e)



4. Elmore delay

- In general, most circuits of interest can be represented as an RC tree, i.e., an RC circuit with no loops.
- The root of the tree is the voltage source, and the leaves are the capacitors at the ends of the branches.
- The Elmore delay model estimates the delay from a source switching to one of the leaf nodes changing as the sum over each node i of the capacitance C_i on the node, multiplied by the effective resistance R_{is} on the shared path from the source to the node and the leaf.

$$t_{pd} = \sum_i R_{is} C_i$$

1. Compute the Elmore delay for V_{out} in the 2nd order RC system from Figure 4.10.

- The circuit has a source and two nodes.
- At node n_1 , the capacitance is C_1 and the resistance to the source is R_1 .
- At node V_{out} , the capacitance is C_2 and the resistance to the source is $(R_1 + R_2)$.
- Hence, the Elmore delay is $t_{pd} = R_1 C_1 + (R_1 + R_2) C_2$

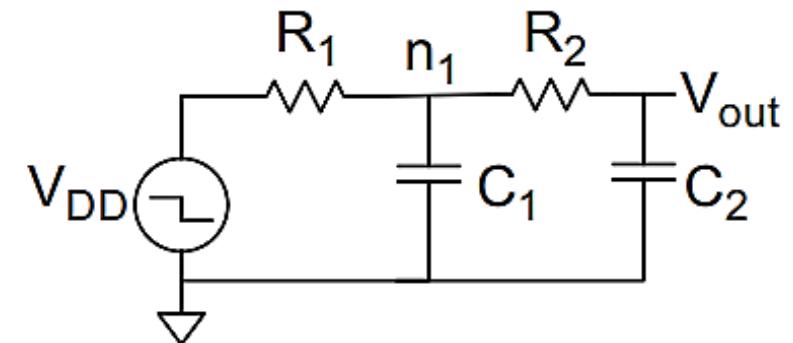


FIGURE 4.10 Second-order RC system



2. Estimate t_{pd} for a unit inverter driving m identical unit inverters.

- Figure 4.12 shows an equivalent circuit for the falling transition.
- Each load inverter presents $3C$ units of gate capacitance, for a total of $3mC$.
- Elmore delay is $t_{pd} = (3 + 3m)RC$

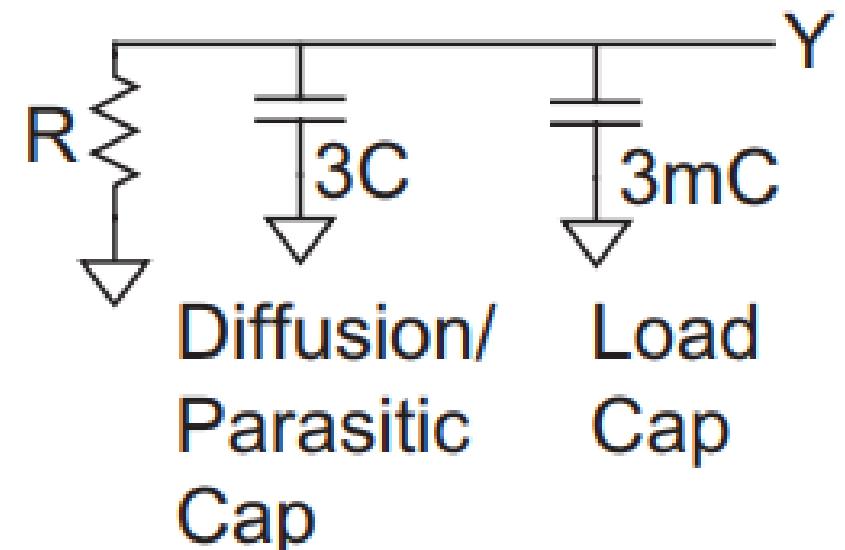


FIGURE 4.12 Equivalent circuit for inverter

3. Repeat Example 2 if the driver is w times unit size.

- The driver transistors are w times as wide, so the effective resistance decreases by a factor of w .
- The diffusion capacitance increases by a factor of w .
- The Elmore delay is $t_{pd} = ((3w + 3m)C)(R/w)$

$$t_{pd} = (3 + 3m/w)RC.$$

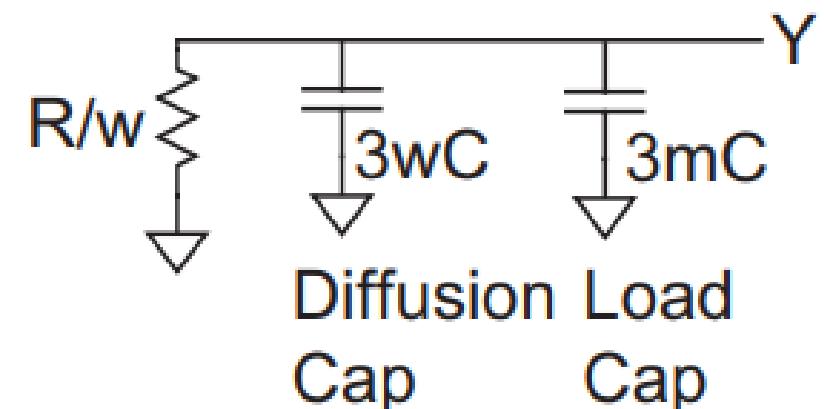


FIGURE 4.13 Equivalent circuit for wider inverter

4. If a unit transistor has $R = 10 \text{ k ohm}$ and $C = 0.1 \text{ fF}$ in a 65 nm process, compute the delay, in picoseconds, of the inverter in Figure 4.14 with a fanout of $m = 4$.

- The RC product in the 65 nm process is $(10 \text{ k ohm})(0.1 \text{ fF}) = 1 \text{ ps}$.
- For $m=4$, the delay is $(3+3m)(1 \text{ ps}) = 15 \text{ ps}$
- The inverter can switch about 66 billion times per second.

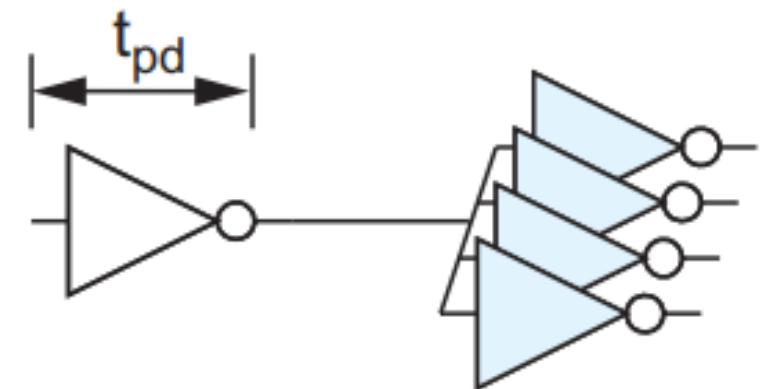


FIGURE 4.14 Fanout-of-4 (FO4) inverter



Linear delay model

- The RC delay model showed that delay is a linear function of the fanout of a gate.
- The normalized delay of a gate can be expressed in units of τ , as,

$$d = f + p$$

- p is the **parasitic delay** inherent to the gate when no load is attached. f is the **effort delay** or **stage effort** that depends on the complexity and fanout of the gate:

$$f = gh$$

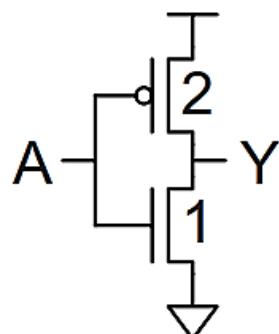
- g is the logical effort. A gate driving h identical copies of itself is said to have a fanout or electrical effort of h .

$$h = \frac{C_{\text{out}}}{C_{\text{in}}}$$

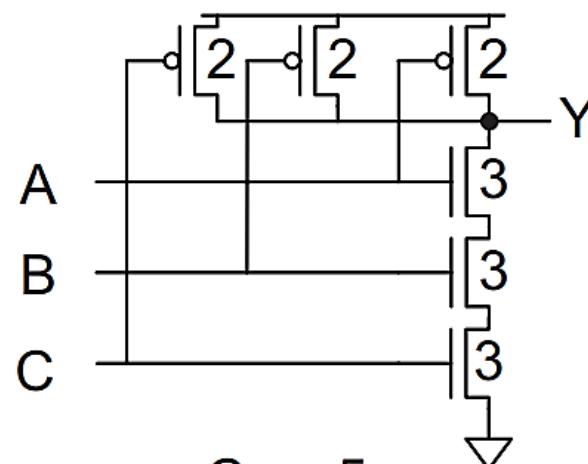
- An inverter is defined to have a logical effort of 1.

1. Logical effort (g)

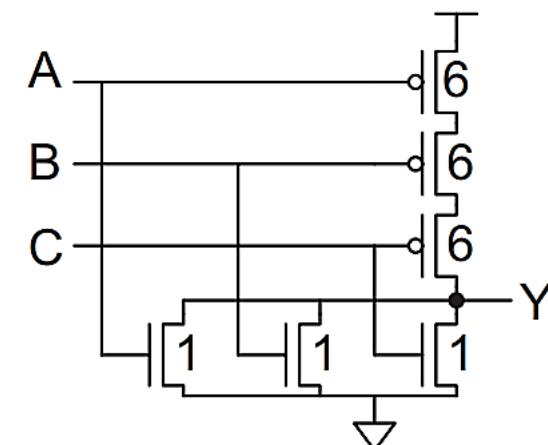
- Logical effort of a gate is defined as ***the ratio of the input capacitance of the gate to the input capacitance of an inverter*** that can deliver the same output current.
- Equivalently, logical effort indicates how much worse a gate is at producing output current as compared to an inverter, given that each input of the gate may only present as much input capacitance as the inverter.



(a) $C_{in} = 3$
 $g = 3/3$



(b) $C_{in} = 5$
 $g = 5/3$



(c) $C_{in} = 7$
 $g = 7/3$



TABLE 4.2 Logical effort of common gates

Gate Type	Number of Inputs				
	1	2	3	4	n
inverter	1				
NAND		4/3	5/3	6/3	$(n + 2)/3$
NOR		5/3	7/3	9/3	$(2n + 1)/3$
tristate, multiplexer	2	2	2	2	2



2. Parasitic delay (P)

- The parasitic delay of a gate is the delay of the gate when it drives zero load.
- It can be estimated with RC delay models.
- A crude method good for hand calculations is to count only diffusion capacitance on the output node.
- For example, consider the gates in Figure 4.22, transistor widths were chosen to give a resistance of R in each gate.
- The inverter has three units of diffusion capacitance on the output, so the parasitic delay is $3RC = \tau$
- This is then normalized parasitic delay equal to 1. This is generally denoted as P_{inv}
- P_{inv} is the ratio of diffusion capacitance to gate capacitance.**

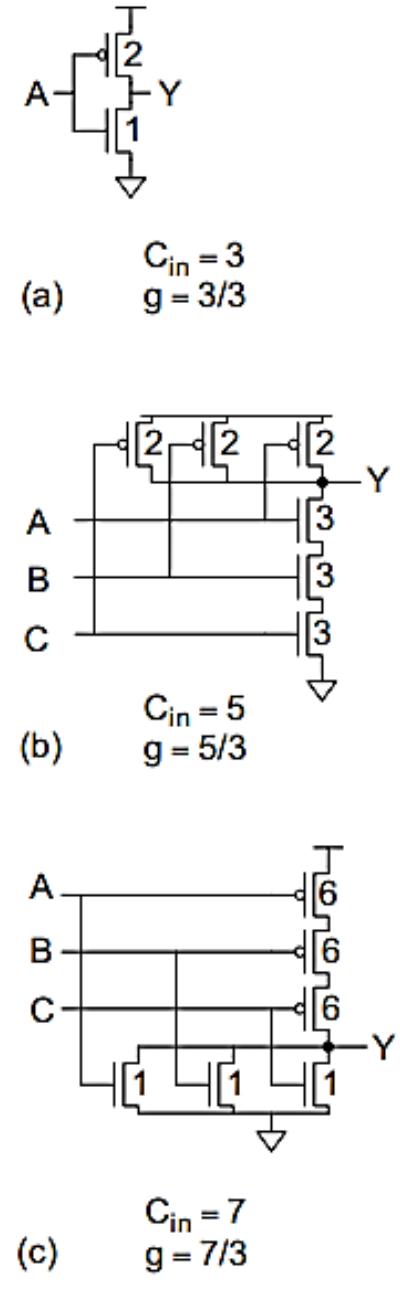


FIGURE 4.22 Logic gates sized for unit resistance



- The 3-input NAND and NOR each have 9 units of diffusion capacitance on the output, so the parasitic delay is three times as great ($3P_{inv}$, or simply 3).

TABLE 4.3 Parasitic delay of common gates

Gate Type	Number of Inputs				
	1	2	3	4	n
inverter	1				
NAND		2	3	4	n
NOR		2	3	4	n
tristate, multiplexer	2	4	6	8	$2n$



3. Delay in logic gates

Example 4.10

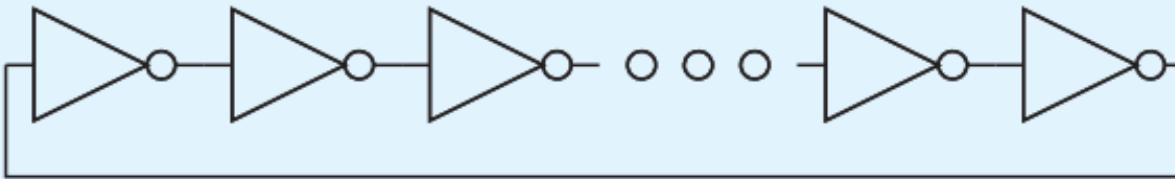
Use the linear delay model to estimate the delay of the fanout-of-4 (FO4) inverter from Example 4.6. Assume the inverter is constructed in a 65 nm process with $\tau = 3 \text{ ps}$.

- The logical effort of the inverter is $g = 1$, by definition.
- The electrical effort, $h = 4$ because the load is four gates of equal size.
- The parasitic delay of an inverter is $p_{inv} = 1$.
- The total delay is $d = gh + p = 1 \times 4 + 1 = 5$ in normalized terms, or $t_{pd} = 15 \text{ ps}$ in absolute terms.



Example 4.11

A ring oscillator is constructed from an odd number of inverters, as shown in Figure 4.24. Estimate the frequency of an N -stage ring oscillator.



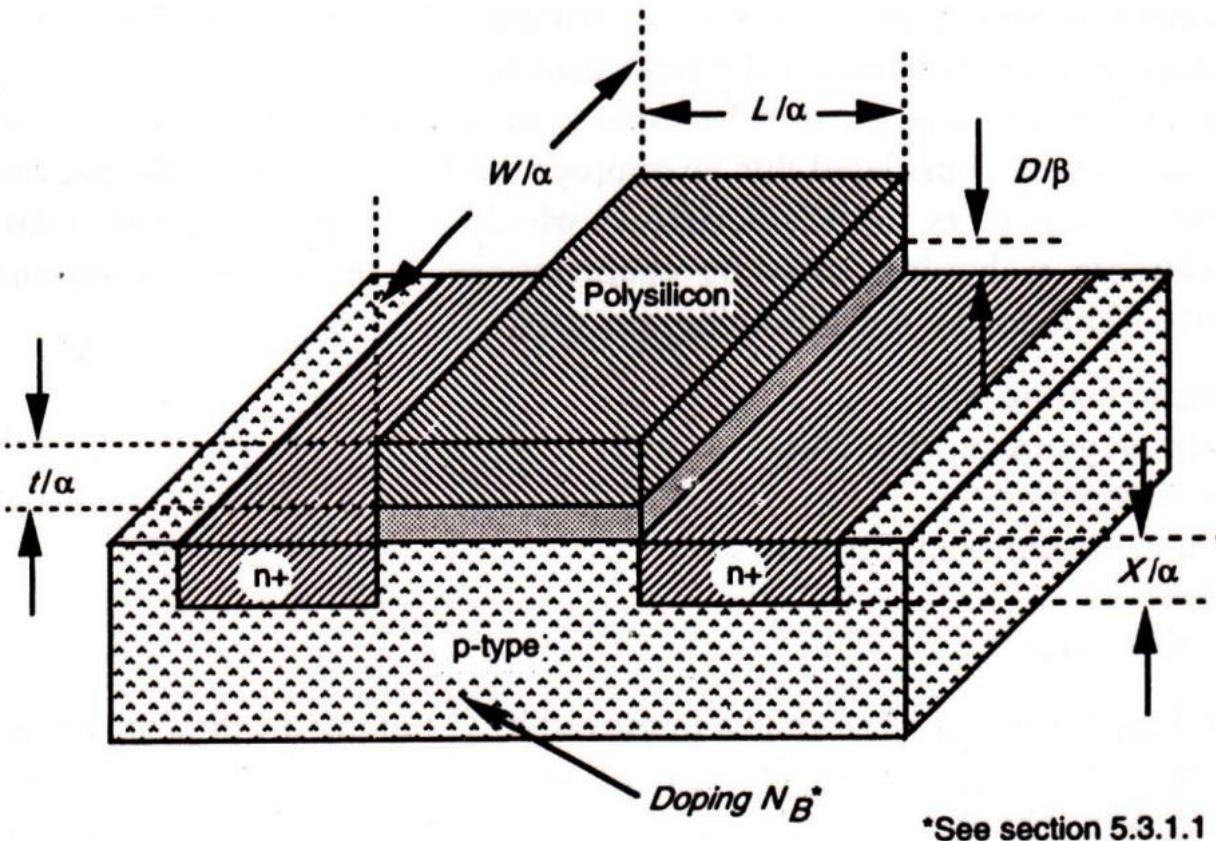
- The logical effort of the inverter is $g = 1$, by definition.
- The electrical effort h of each inverter is also 1 because it drives a single identical load.
- The parasitic delay p is also 1.
- The delay of each stage is $d = gh + p = 1 \times 1 + 1 = 2$.
- An N -stage ring oscillator has a period of $2N$ stage delays because a value must propagate twice around the ring to regain the original polarity. Therefore, the period is $T = 2 \times 2N$. The frequency is the reciprocal of the period, $1/4N$.
- A 31-stage ring oscillator in a 65 nm process has a frequency of $1/(4 \times 31 \times 3 \text{ ps}) = 2.7 \text{ GHz}$



Scaling of MOS circuits

- VLSI fabrication technology is still in the process of evolution which is leading to smaller line widths and feature size and to higher packing density of circuitry on a chip.
- The scaling down of feature size generally leads to improved performance.
- VLSI technology may be characterized in terms of several indicators like:
 1. Minimum feature size
 2. Number of gates on one chip
 3. Power dissipation
 4. Maximum operational frequency
 5. Die size
 6. Production cost
- It is essential for the designer to understand the implementation and the effects of scaling.

Scaling models and scaling factors



Scaling can be done using 2 factors, i.e., **α** and **β** .

β is for scaling oxide thickness D and voltages.

α is of all other parameters like L, A, W, t etc.

FIGURE 5.1 Scaled nMOS transistor (pMOS similar).



Scaling factors for device parameters

1. **Gate Area A_g :** $A_g = L \cdot W$

where L and W are the channel length and width, respectively. Both are scaled by $1/\alpha$. Thus A_g is scaled by $1/\alpha^2$

2. **Gate capacitance Per Unit Area C_0 or C_{ox}** $C_0 = \frac{\epsilon_{ox}}{D}$

where ϵ_{ox} is the permittivity of the gate oxide (thinox) [= $\epsilon_{ins} \cdot \epsilon_0$]

D is the gate oxide thickness which is scaled by $1/\beta$

Thus C_0 is scaled by $\frac{1}{1/\beta} = \beta$

3. **Gate capacitance C_g**

$$C_g = C_0 L \cdot W$$

Thus C_g is scaled by $\beta \frac{1}{\alpha^2} = \frac{\beta}{\alpha^2}$



4. Parasitic Capacitance C_x

C_x is proportional to $\frac{A_x}{d}$

- where d is the depletion width around source or drain which is scaled by $1/\alpha$,
- A_x is the area of the depletion region around source or drain which is scaled by $1/\alpha^2$.

Thus C_x is scaled by $\frac{1}{\alpha^2} \cdot \frac{1}{1/\alpha} = \frac{1}{\alpha}$

5. Carrier Density In Channel Q_{on}

$$Q_{on} = C_0 \cdot V_{gs}$$

- where Q_{on} is the average charge per unit area in the channel in the 'on' state.
- Note that C_0 is scaled by β and V_{gs} is scaled by $1/\beta$.
- Thus- Q_{on} is scaled by 1

6. Channel Resistance R_{on}

$$R_{on} = \frac{L}{W} \frac{1}{Q_{on}\mu}$$

where μ is the carrier mobility in the channel and is assumed constant.

Thus R_{on} is scaled by $\frac{1}{\alpha} \frac{1}{1/\alpha} 1 = 1$



7. Gate Delay T_d

T_d is proportional to $R_{on} \cdot C_g$

Thus T_d is scaled by $\frac{1}{\alpha^2}$ $\frac{\beta}{\alpha^2}$

8. Maximum Operating Frequency f_0

$$f_0 = \frac{W}{L} \frac{\mu C_0 V_{DD}}{C_g}$$

or, f_0 is inversely proportional to delay T_d .

Thus f_0 is scaled by $\frac{1}{\beta/\alpha^2} = \frac{\alpha^2}{\beta}$

9. Saturation Current I_{dss}

$$I_{dss} = \frac{C_0 \mu}{2} \frac{W}{L} (V_{gs} - V_t)^2$$

noting that both V_{gs} and V_t are scaled by $1/\beta$, we have

I_{dss} is scaled by $\beta(1/\beta)^2 = 1/\beta$



10. Current Density J

$$J = \frac{I_{dss}}{A}$$

where A is the cross-sectional area of the channel in the ‘on’ state which is scaled by $1/\alpha^2$

So, J is scaled by $\frac{1/\beta}{1/\alpha^2} = \frac{\alpha^2}{\beta}$

11. Switching Energy Per Gate E_g

$$E_g = \frac{1}{2} C_g (V_{DD})^2$$

So, E_g is scaled by $\frac{\beta}{\alpha^2} \cdot \frac{1}{\beta^2} = \frac{1}{\alpha^2 \beta}$



12. Power Dissipation Per Gate P_g

P_g comprises two components such that

$$P_g = P_{gs} + P_{gd}$$

where the static component

$$P_{gs} = \frac{(V_{DD})^2}{R_{on}}$$

and the dynamic component

$$P_{gd} = E_g f_0$$

It will be seen that both P_{gs} and P_{gd} are scaled by $1/\beta^2$

So, P_g is scaled by $1/\beta^2$



13. Power Dissipation Per Unit Area P_a

$$P_a = \frac{P_g}{A_g}$$

So, P_a is scaled by $\frac{1/\beta^2}{1/\alpha^2} = \alpha^2/\beta^2$

14. Power-speed Product P_T

$$P_T = P_g \cdot T_d$$

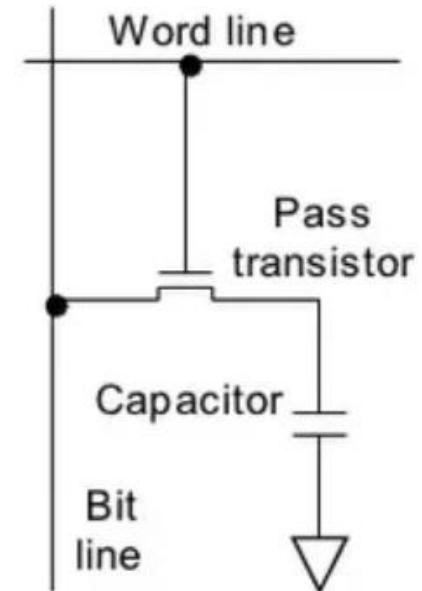
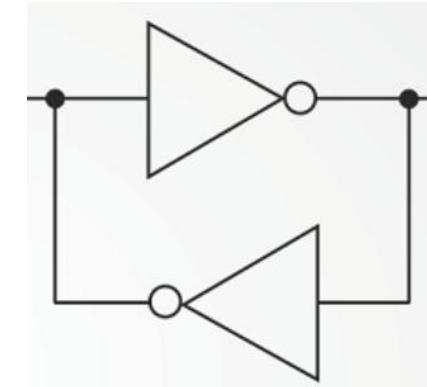
So, P_T is scaled by $\frac{1}{\beta^2} \cdot \frac{\beta}{\alpha^2} = \frac{1}{\alpha^2\beta}$



Random Access Memory (RAM)

Two types:

1. **Static RAM (SRAM):** Stored information depends on the state of transistor, and it requires continuous power supply.
2. **Dynamic RAM (DRAM):** Stored information depends on charge in capacitor and doesn't require continuous power supply but the charge in the capacitor must be refreshed after a particular interval of time.



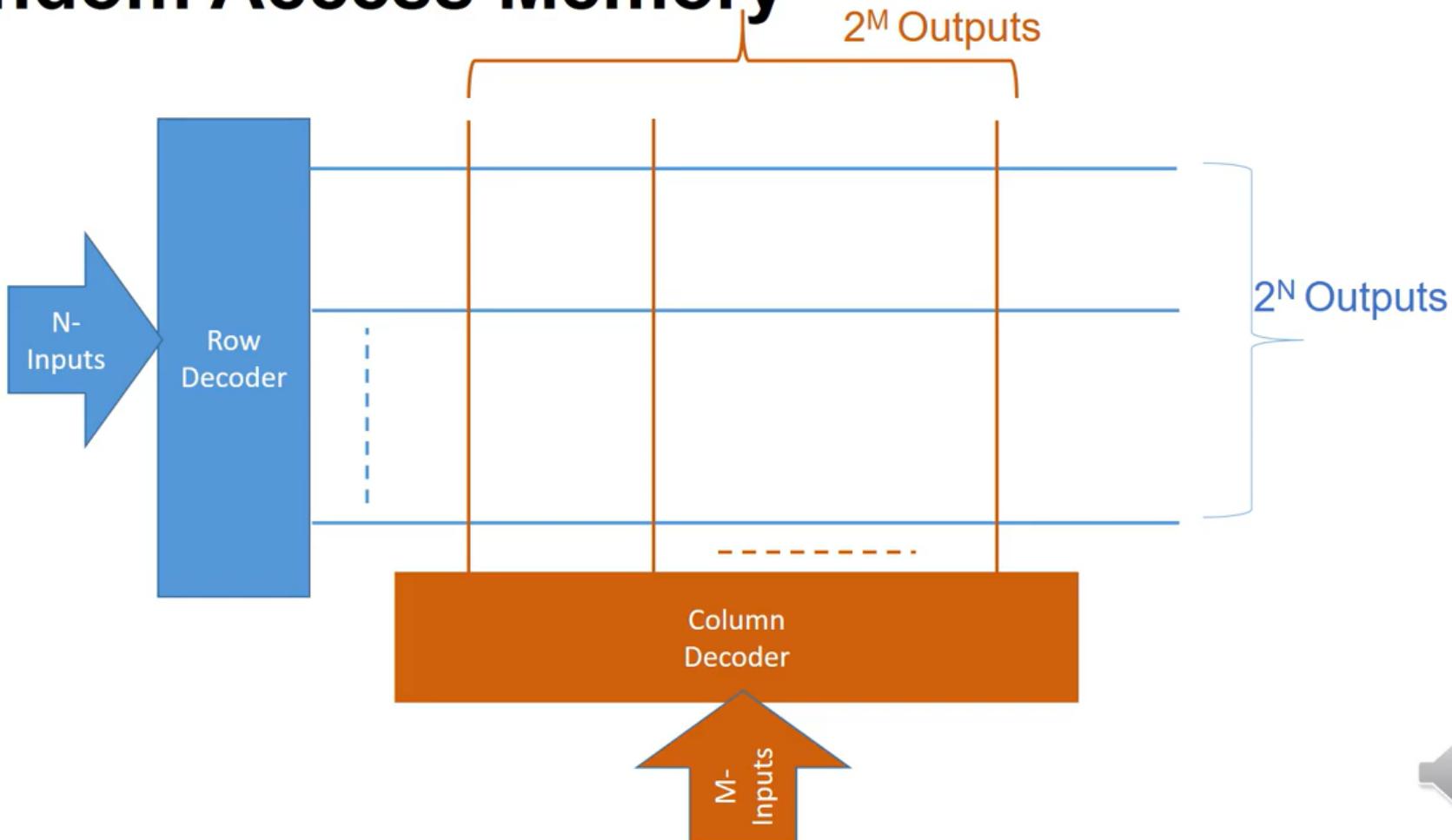


Random Access Memory



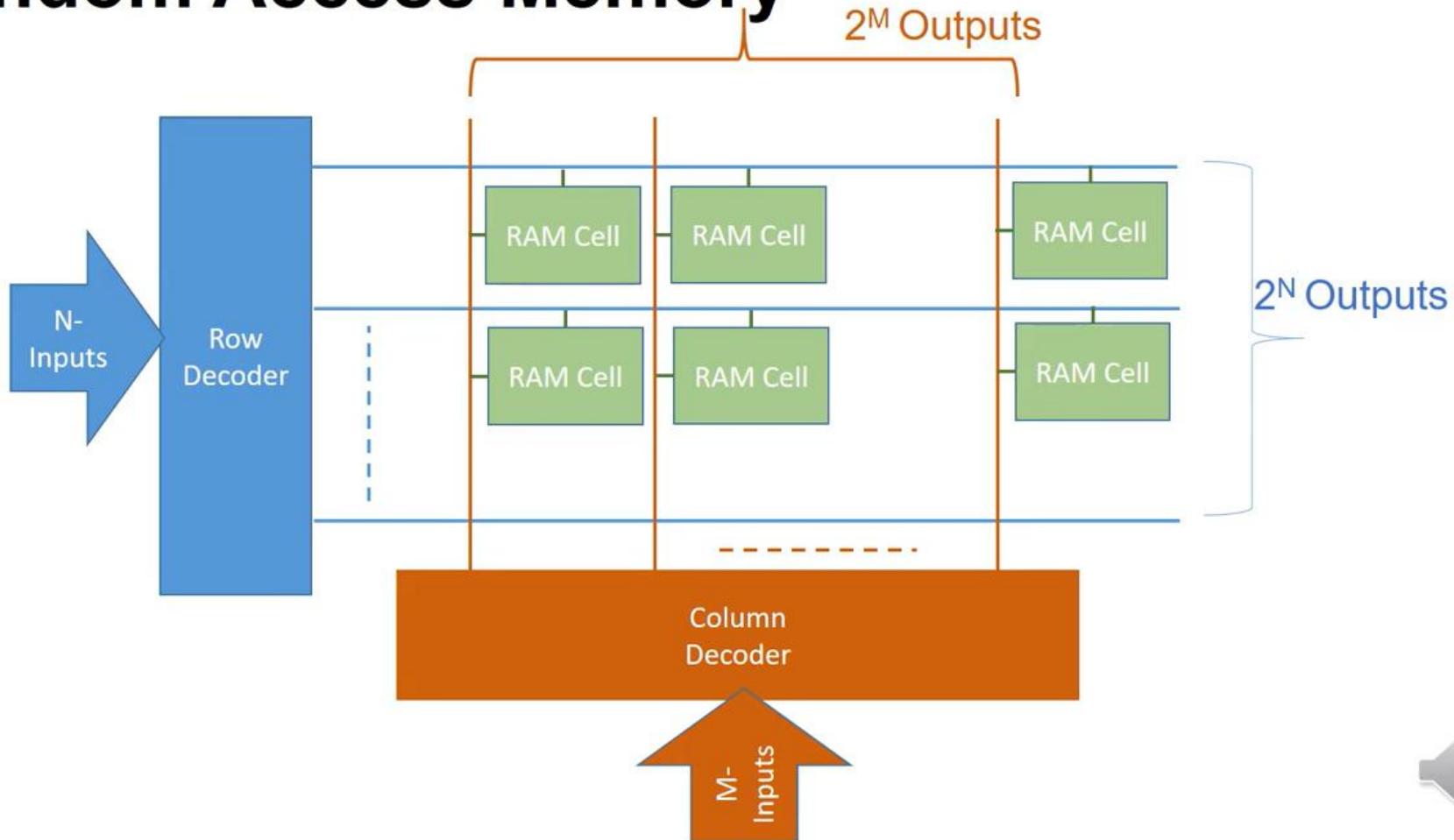


Random Access Memory





Random Access Memory



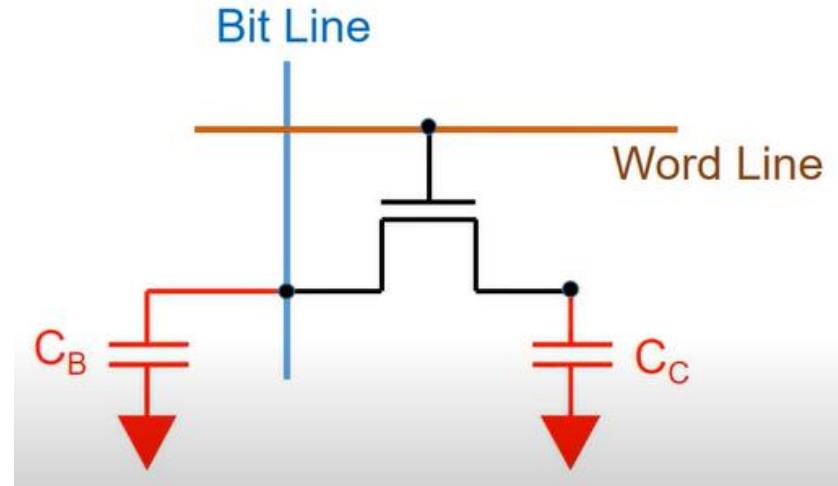


=

1 transistor- Dynamic RAM (1T-DRAM)



1 transistor- Dynamic RAM (1T-DRAM)



- **Word line:** It is the output of row decoder. If you want to read or write data then word line should be **high**.
- **Bit line:** It is the output of column decoder. To write any data, the data is supplied through bit line. To read any data, the data is read through bit line.





$$V_B = \frac{Q_B}{C_B}$$

$$V_C = \frac{Q_C}{C_C}$$



Final voltage

$$V_F = \frac{Q_B + Q_C}{C_B + C_C}$$

Difference in bitline voltage

$$\Delta V_B = V_B - V_F$$

$$\Delta V_B = \frac{Q_B}{C_B} - \frac{Q_B + Q_C}{C_B + C_C}$$

$$\Delta V_B = \frac{Q_B(C_B + C_C) - C_B(Q_B + Q_C)}{C_B(C_B + C_C)}$$

$$\Delta V_B = \frac{Q_B \left(1 + \frac{C_C}{C_B}\right) - (Q_B + Q_C)}{(C_B + C_C)}$$

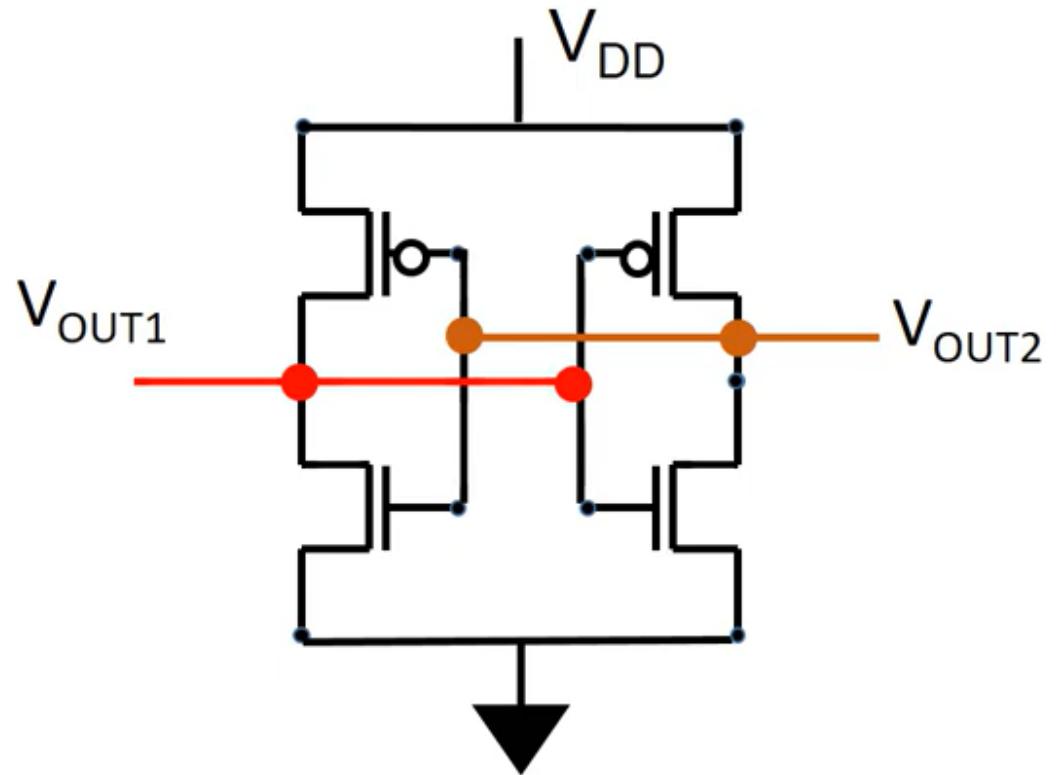
$$\Delta V_B = \frac{C_C(V_B - V_C)}{(C_B + C_C)}$$

Assume $V_B=2.5V$, $V_C=5V$, $C_B=1pF$, and $C_C=50fF$,

$$\Delta V_B = -119mV$$

$V_F = 2.381 V$ or $2.619 V$. Hence, the data is destroyed





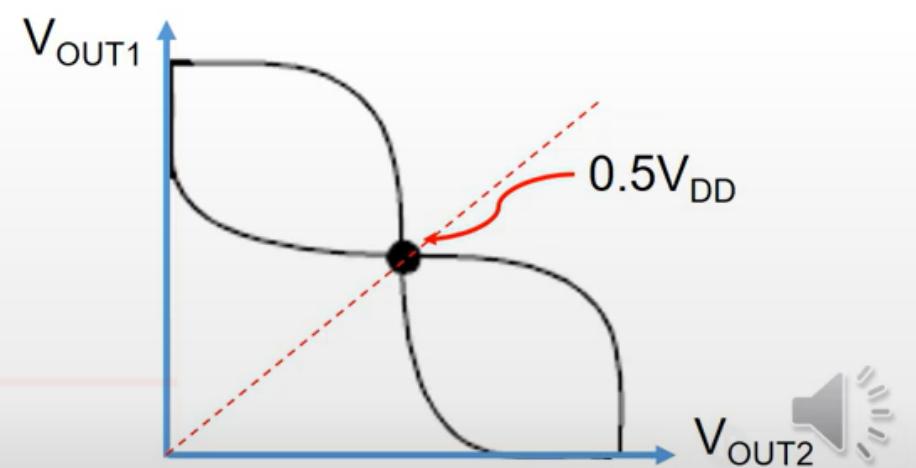
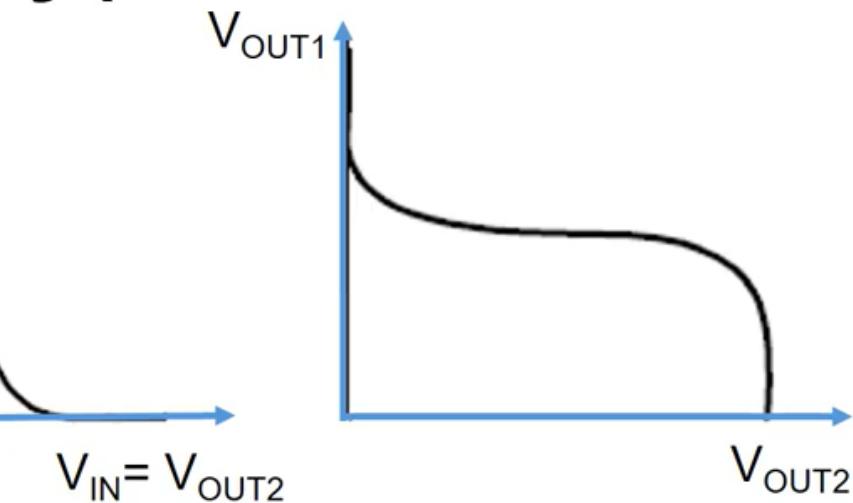
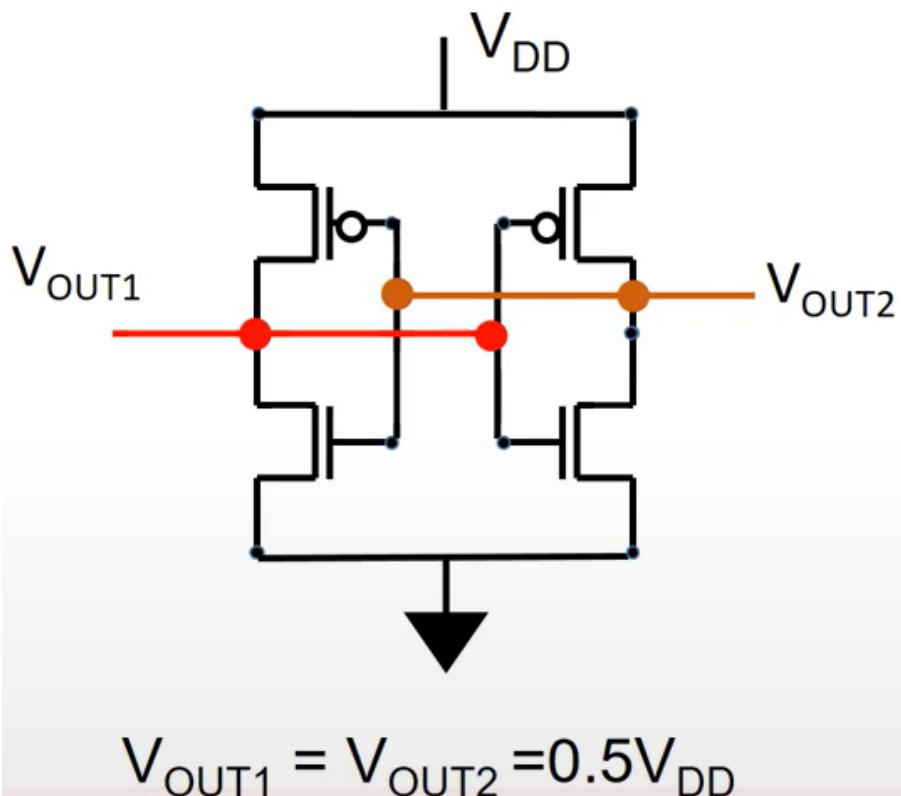
Sense Amplifier

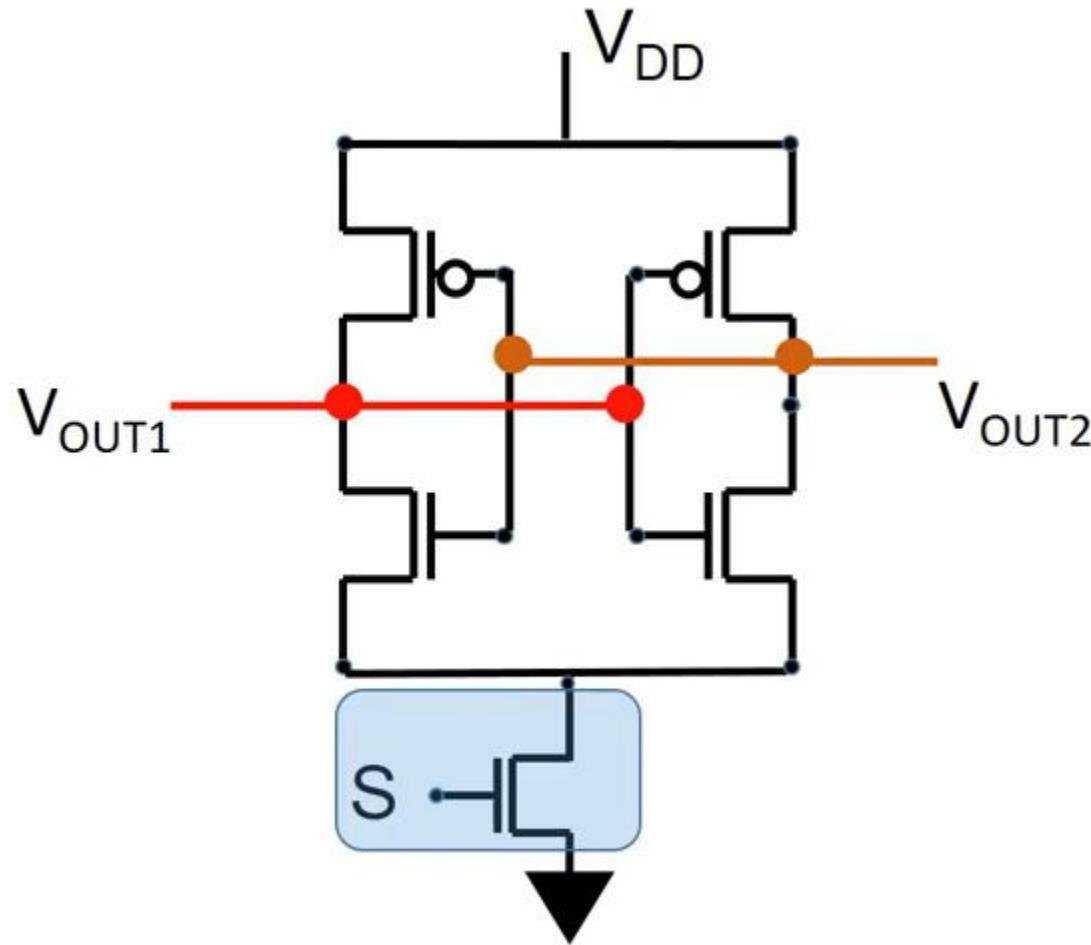
$$V_{OUT1} = V_{OUT2} = 0.5V_{DD}$$

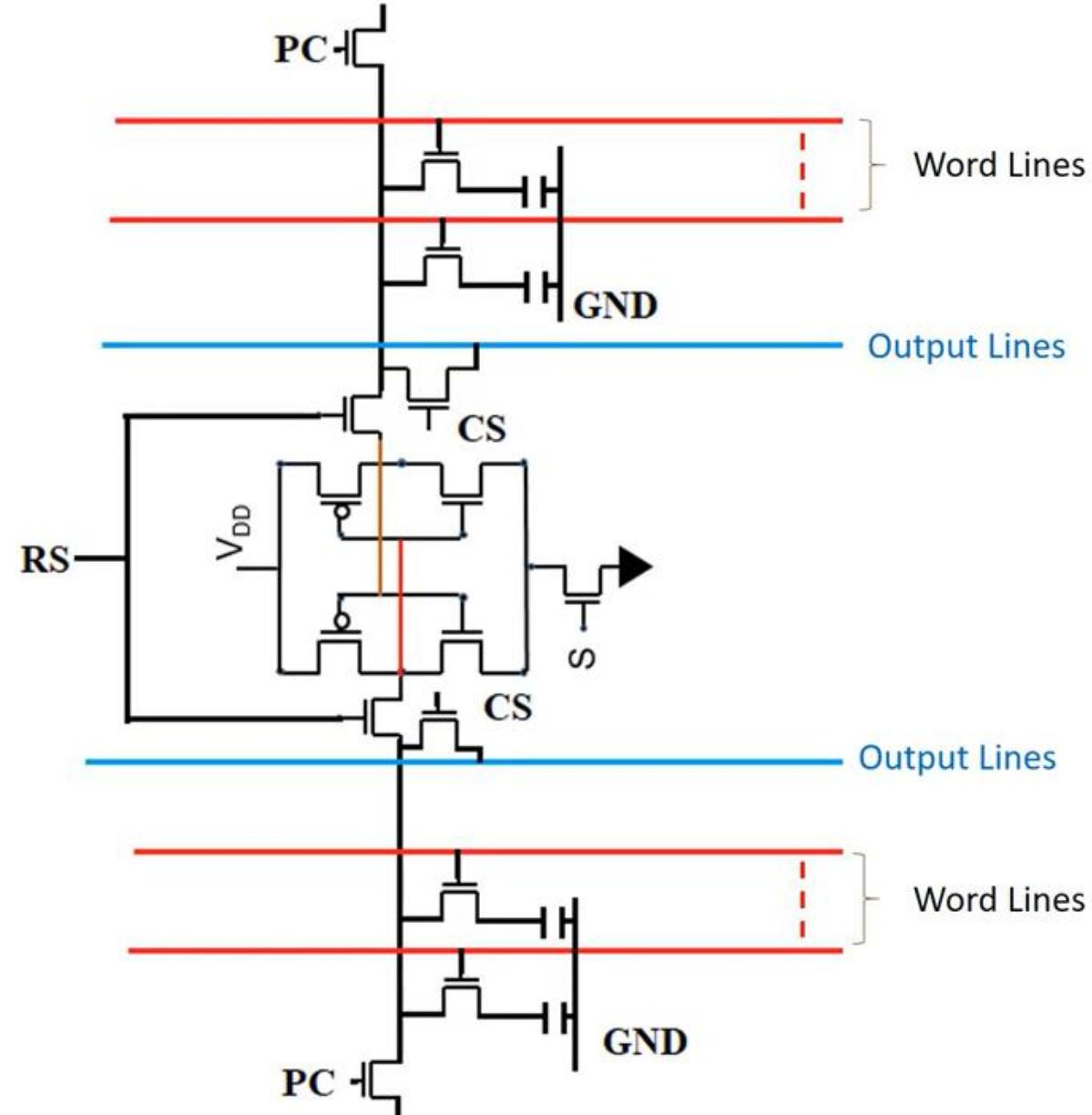


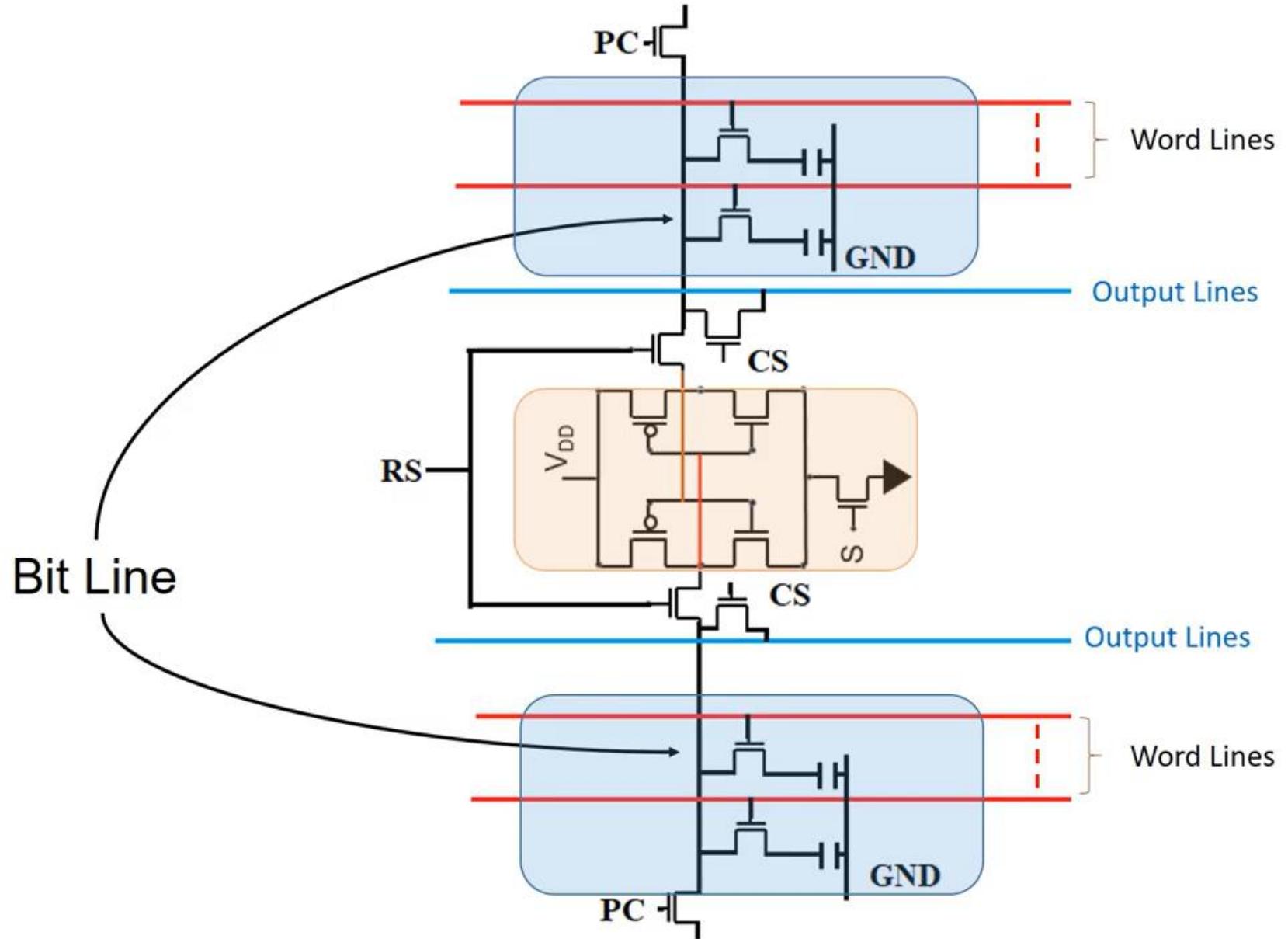


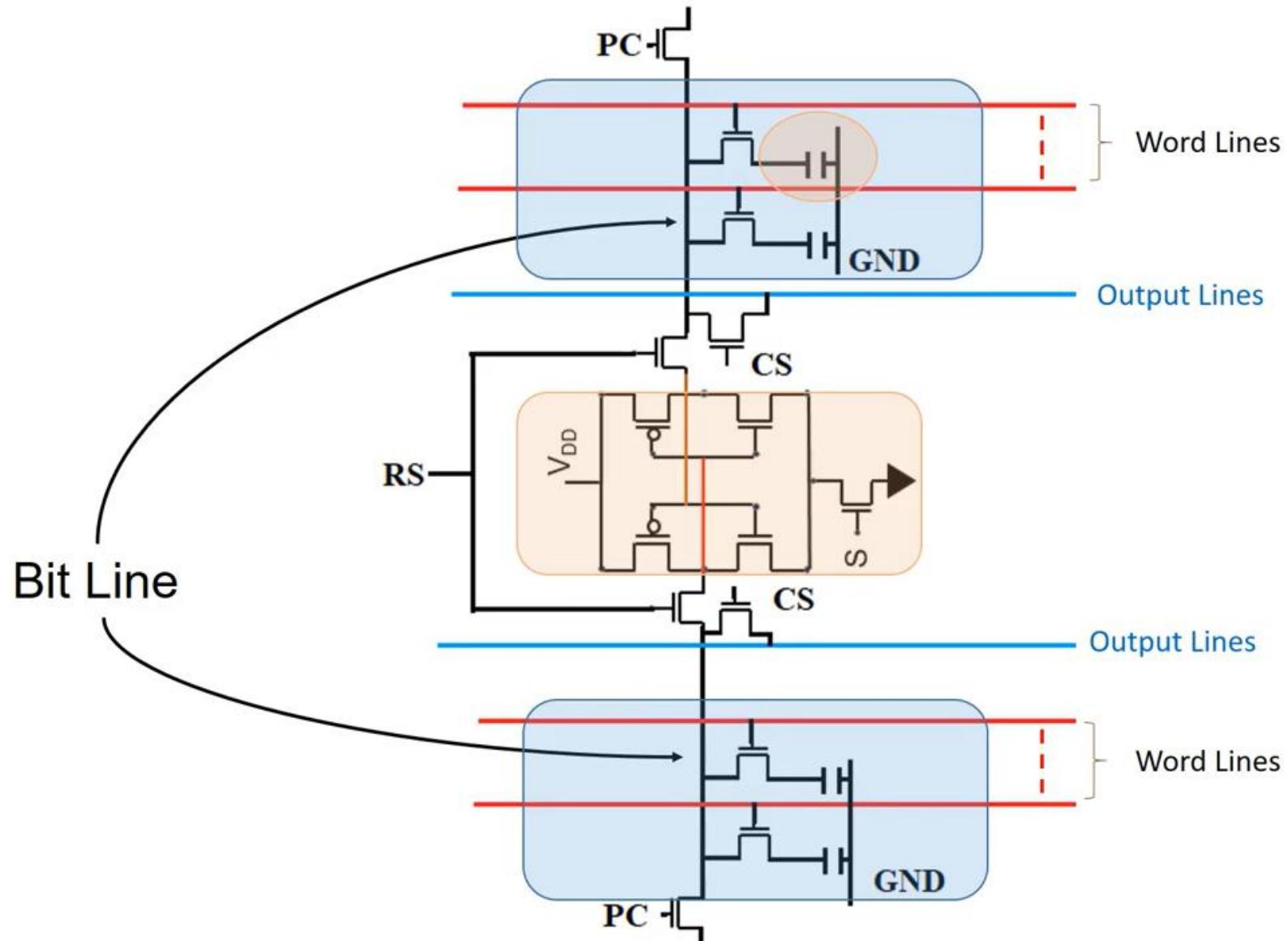
Sense Amplifier













Random Access Memory(1-T DRAM)

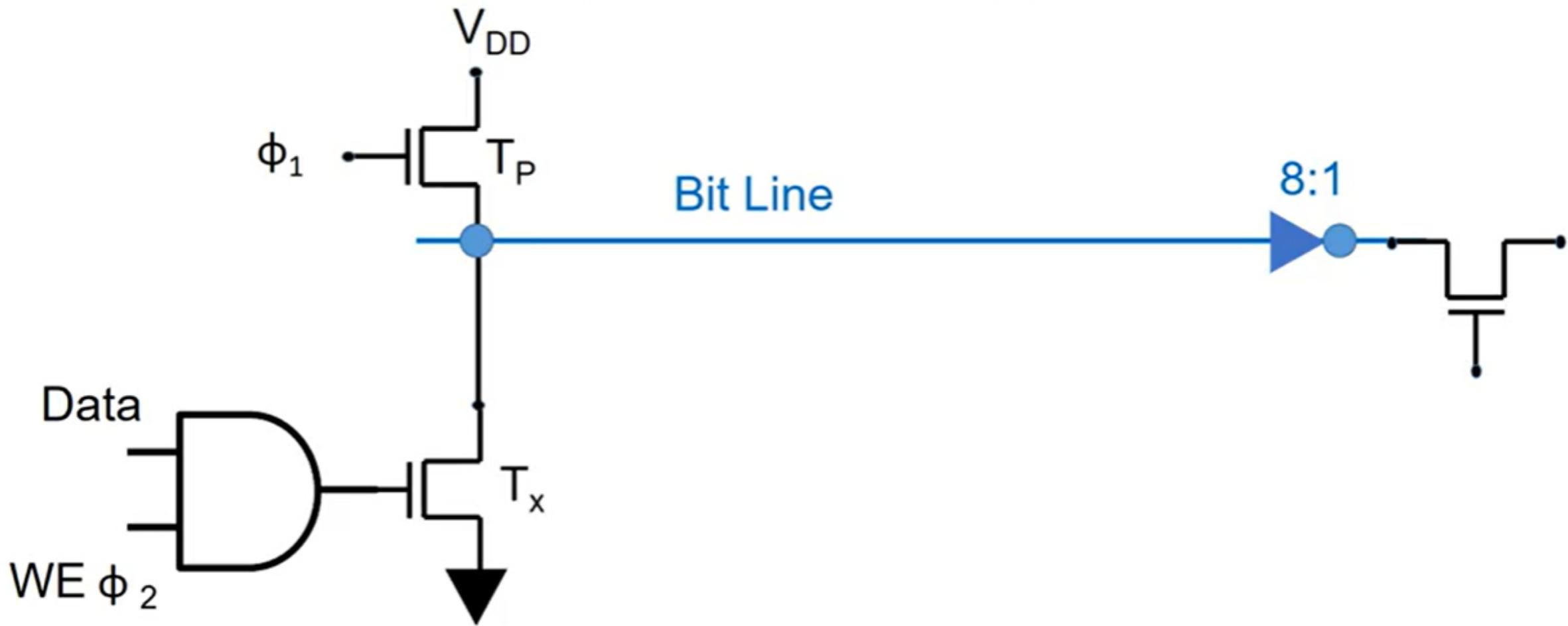
- **Write operation** : 1 T dynamic memory cell consists of a capacitor C_c which can be **charged** from **read/write line(Bit line)**, provided that the **Word Line** is high
- **Read operation** : The state of the charge across C_c can be read via the same read/write line with a high **Word Line** . Hence, *during read operation, a sense amplifier is used to differentiate between a stored 0 and a stored 1.*

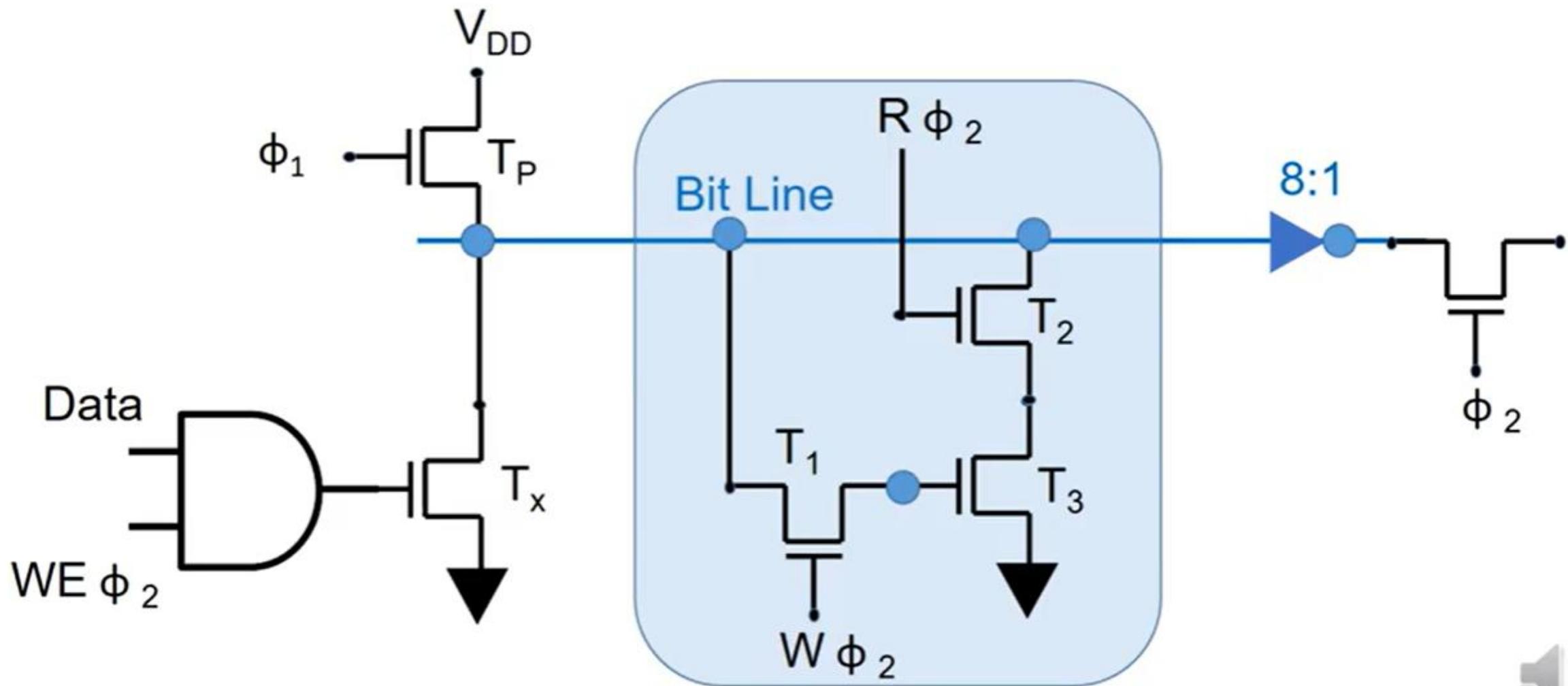


3T-DRAM

3T DRAM uses 2 Phase Clock Signal

- Two clocks cannot be high at the same time.
- Two clocks can be low at the same time.

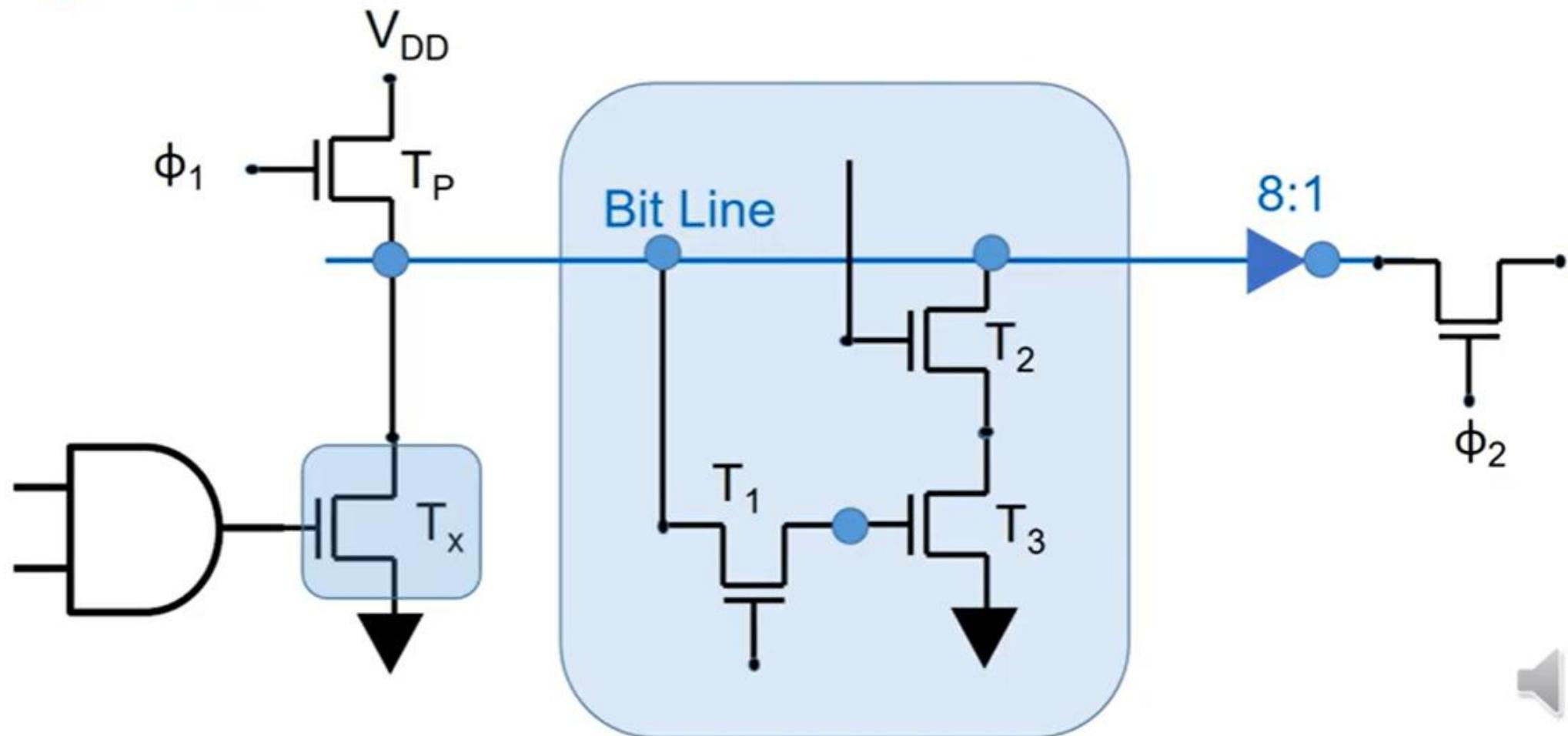






Write 1

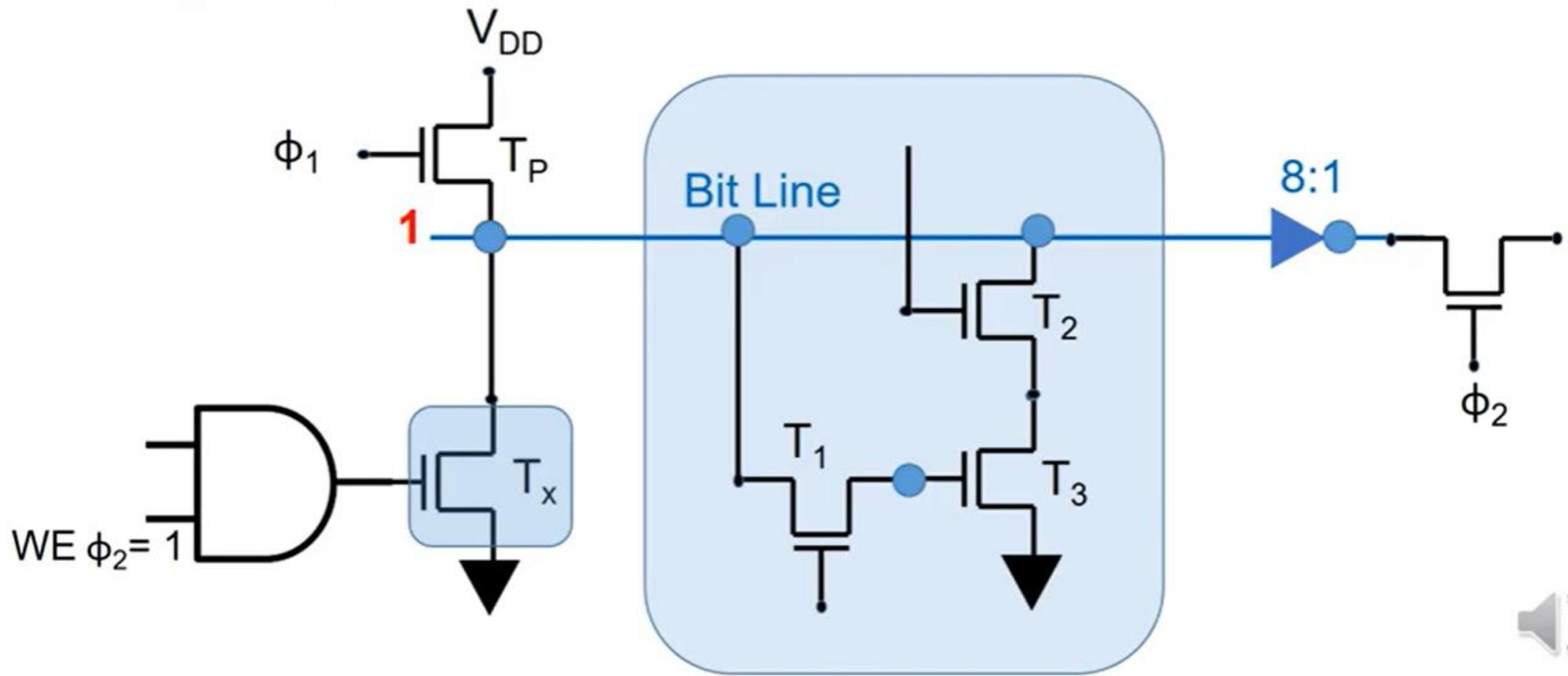
Write Operation





Write 1

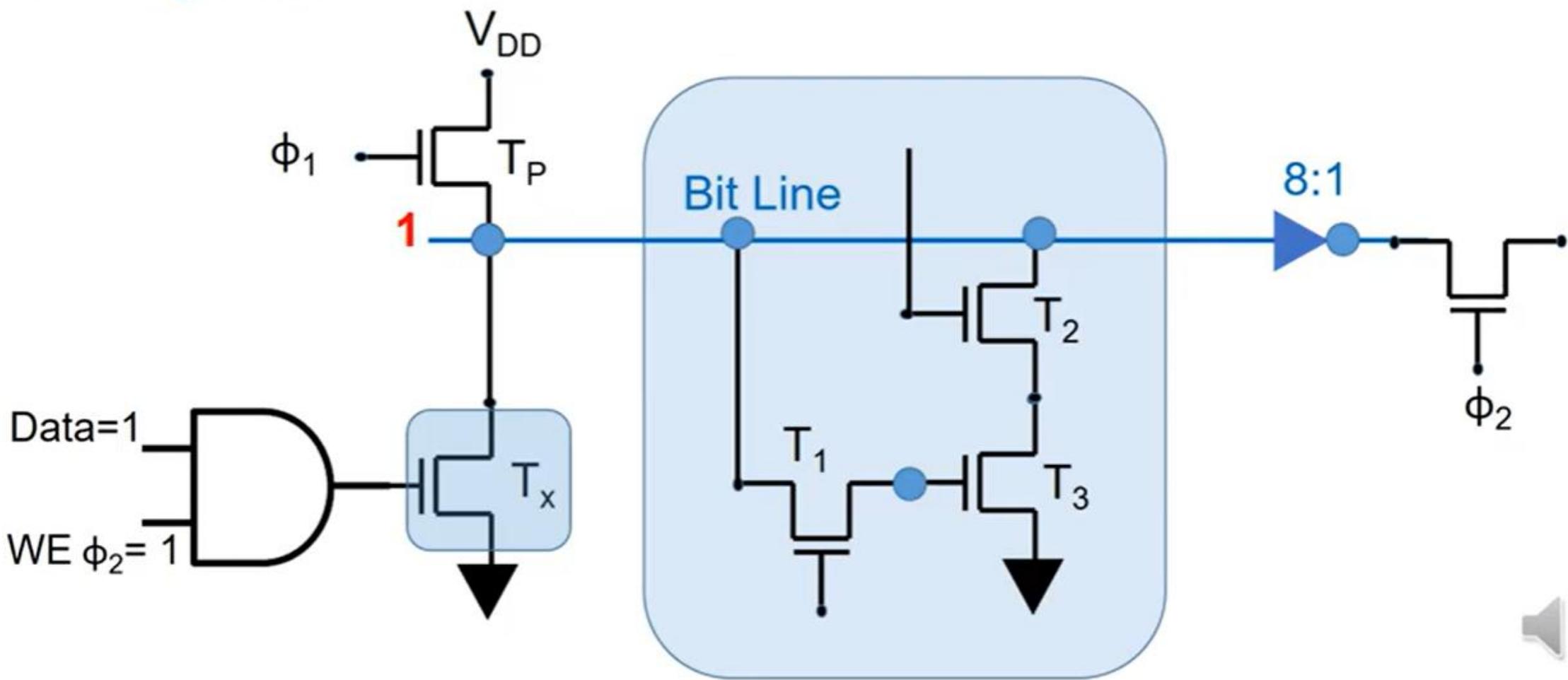
Write Operation





Write 1

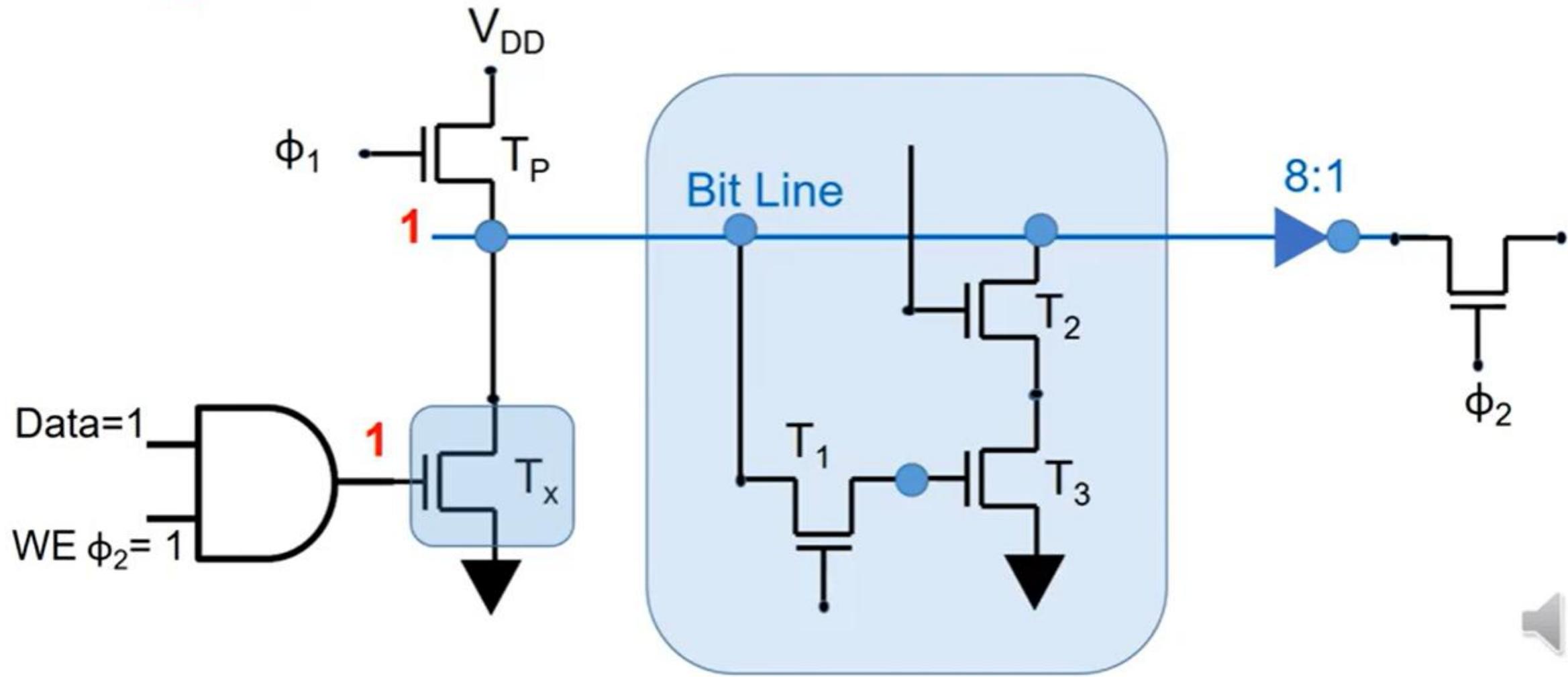
Write Operation





Write 1

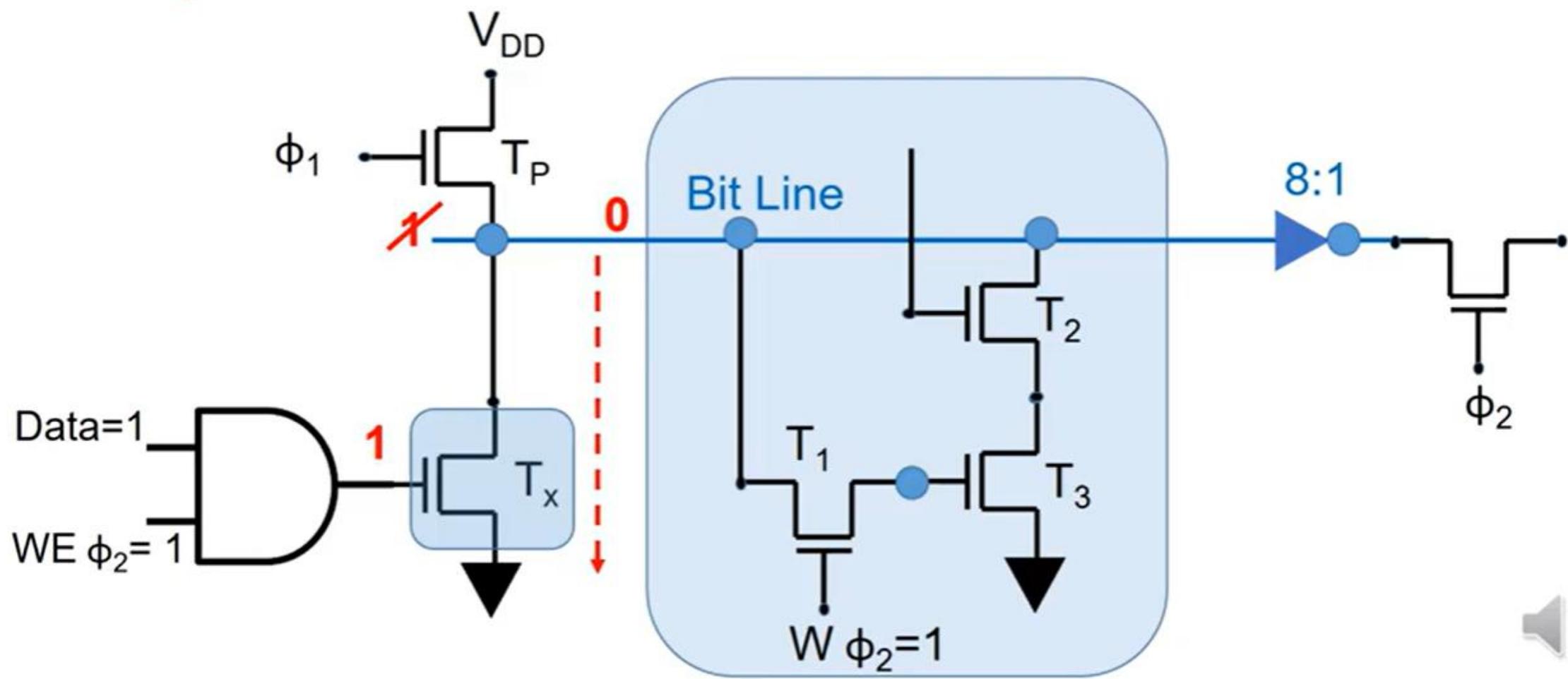
Write Operation





Write 1

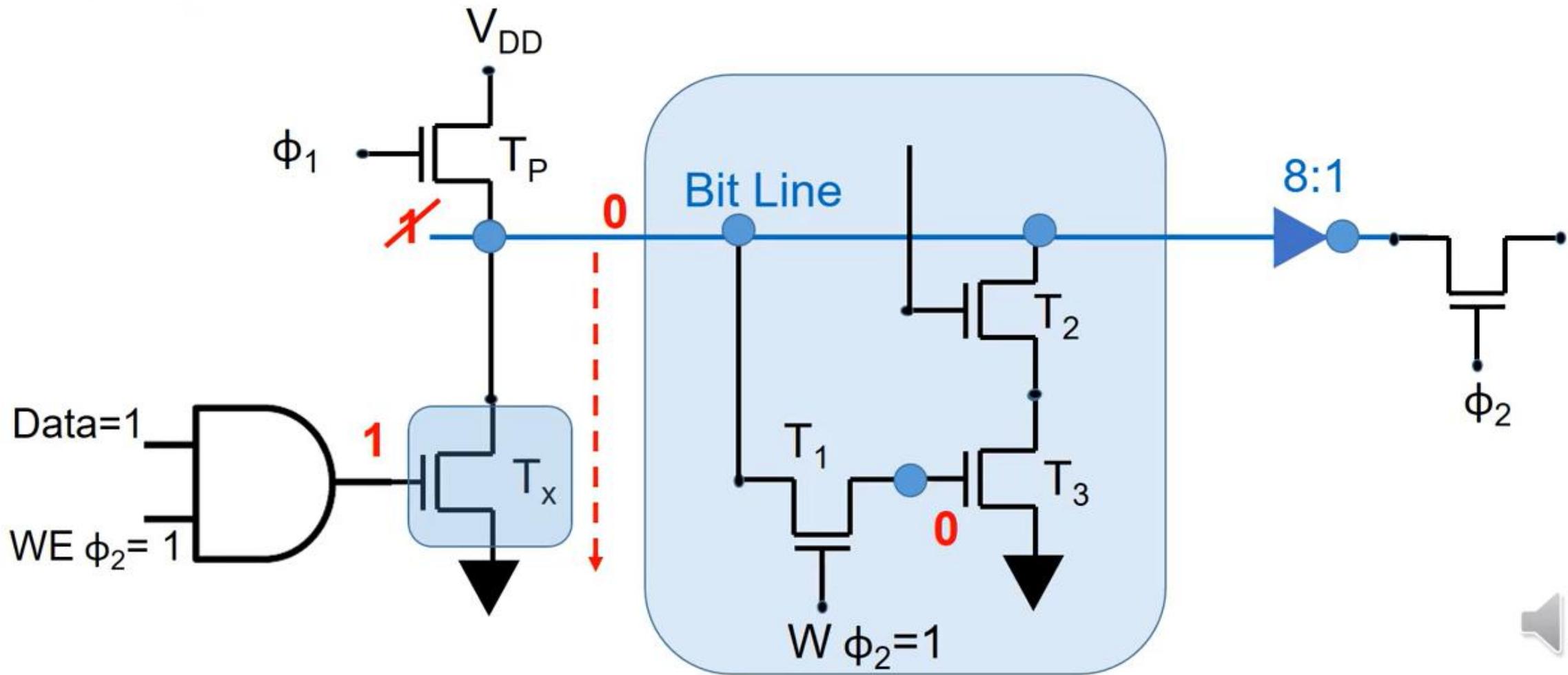
Write Operation





Write 1

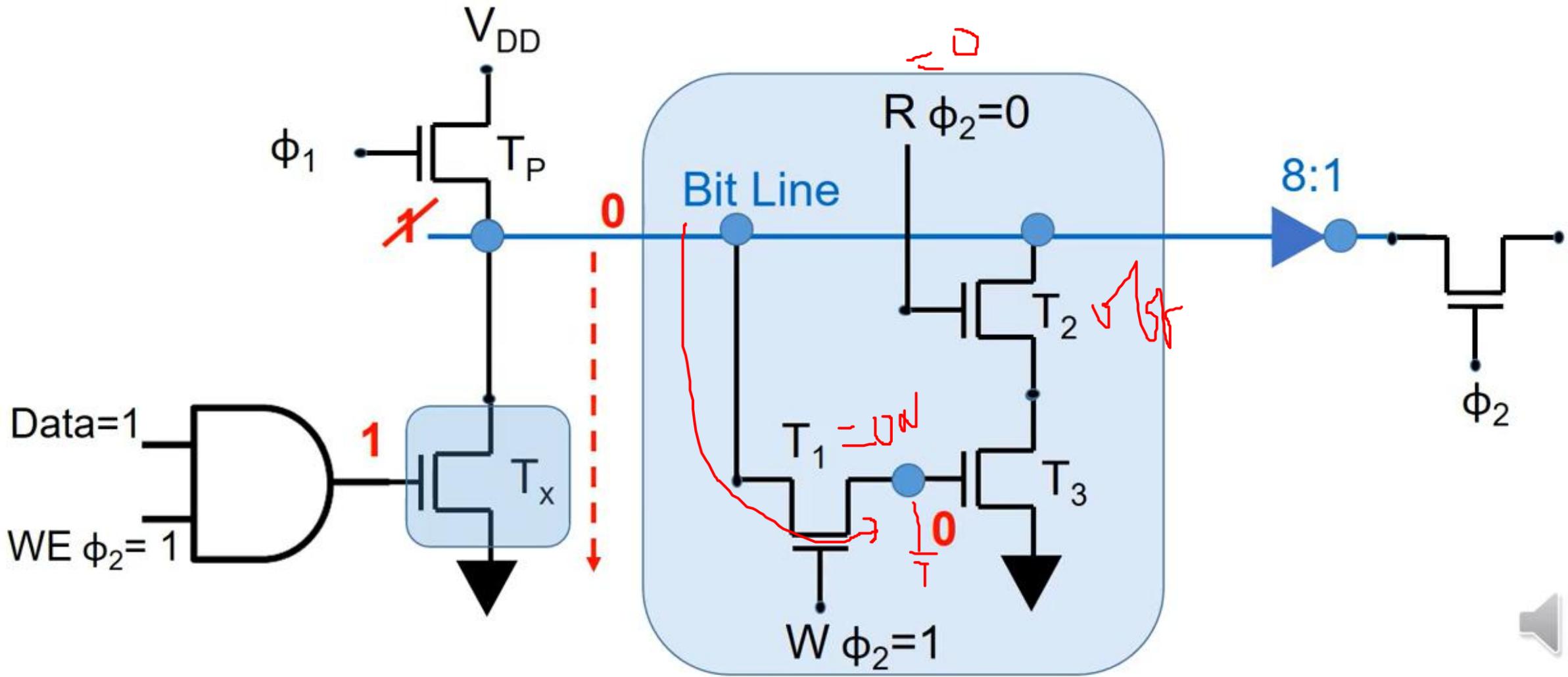
Write Operation





Write 1

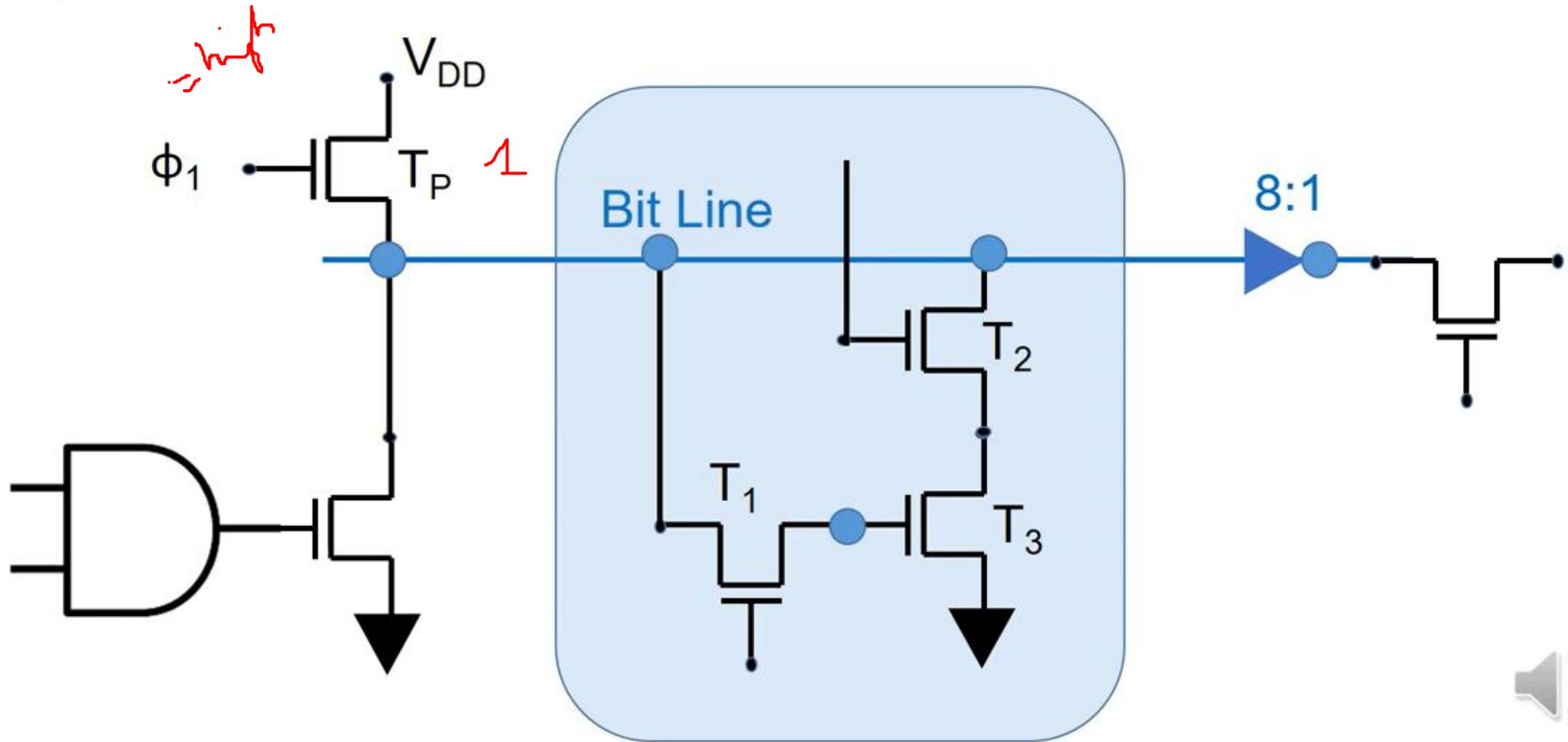
Write Operation





Read 0

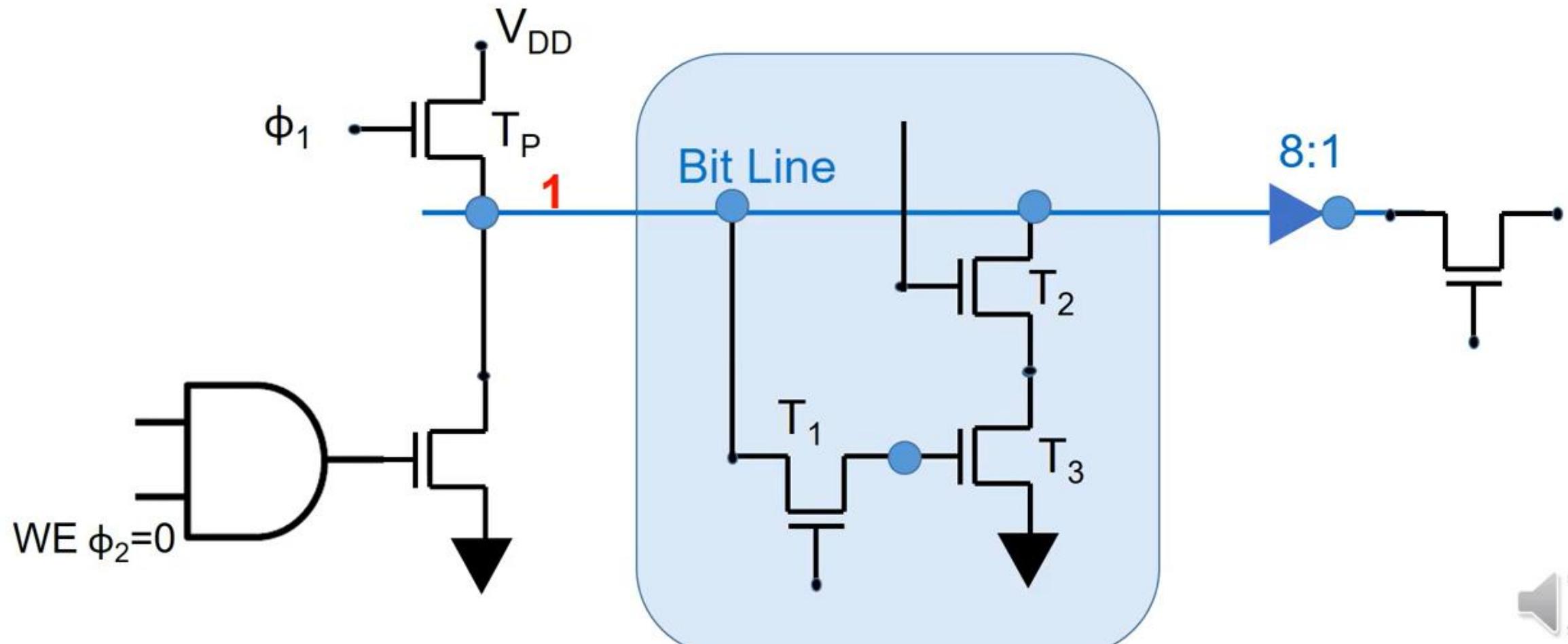
Read Operation





Read 0

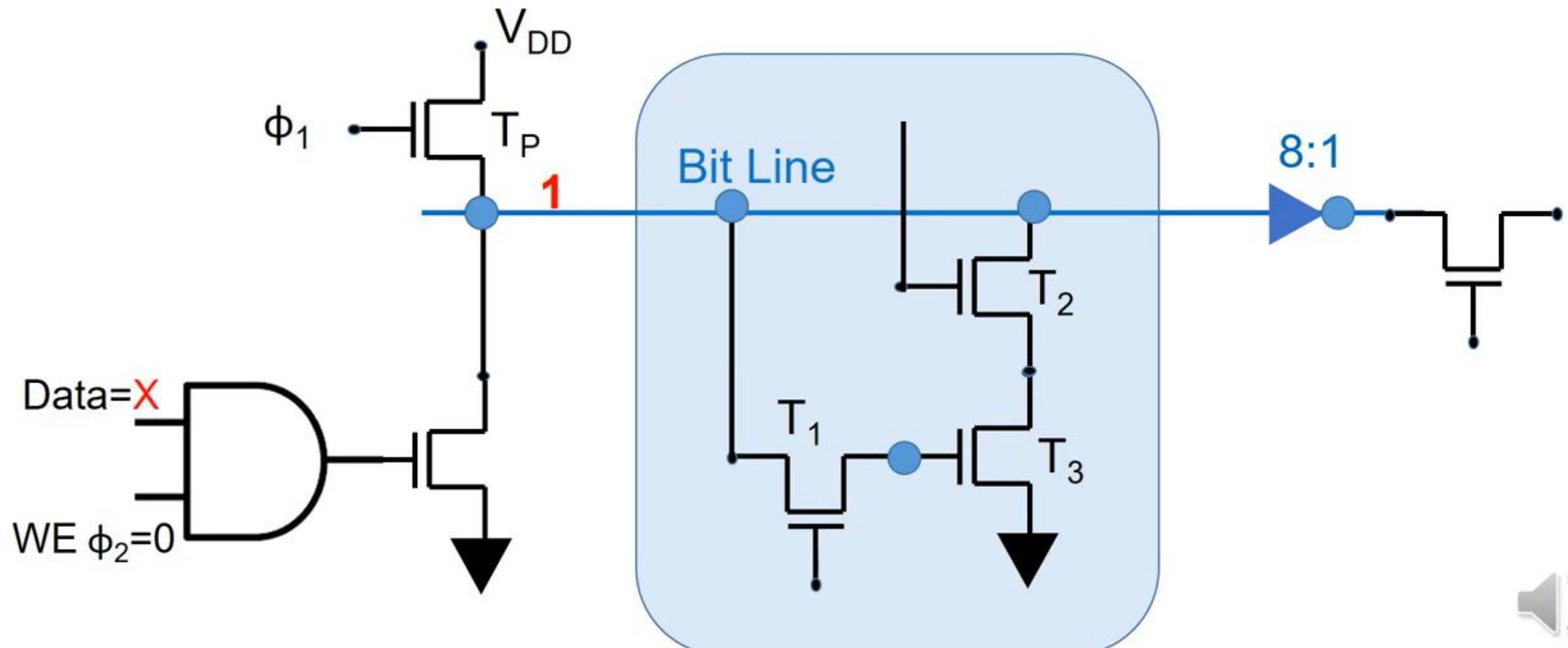
Read Operation





Read 0

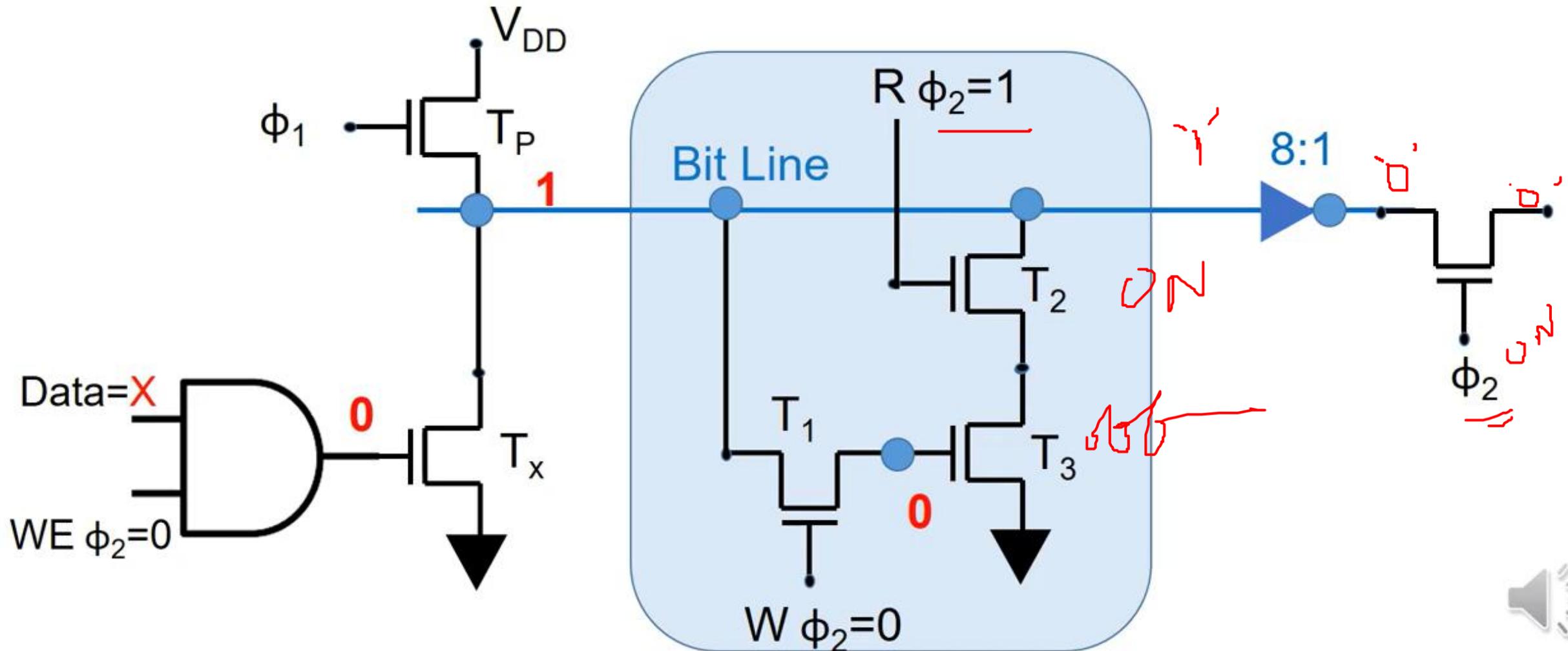
Read Operation





Read Operation

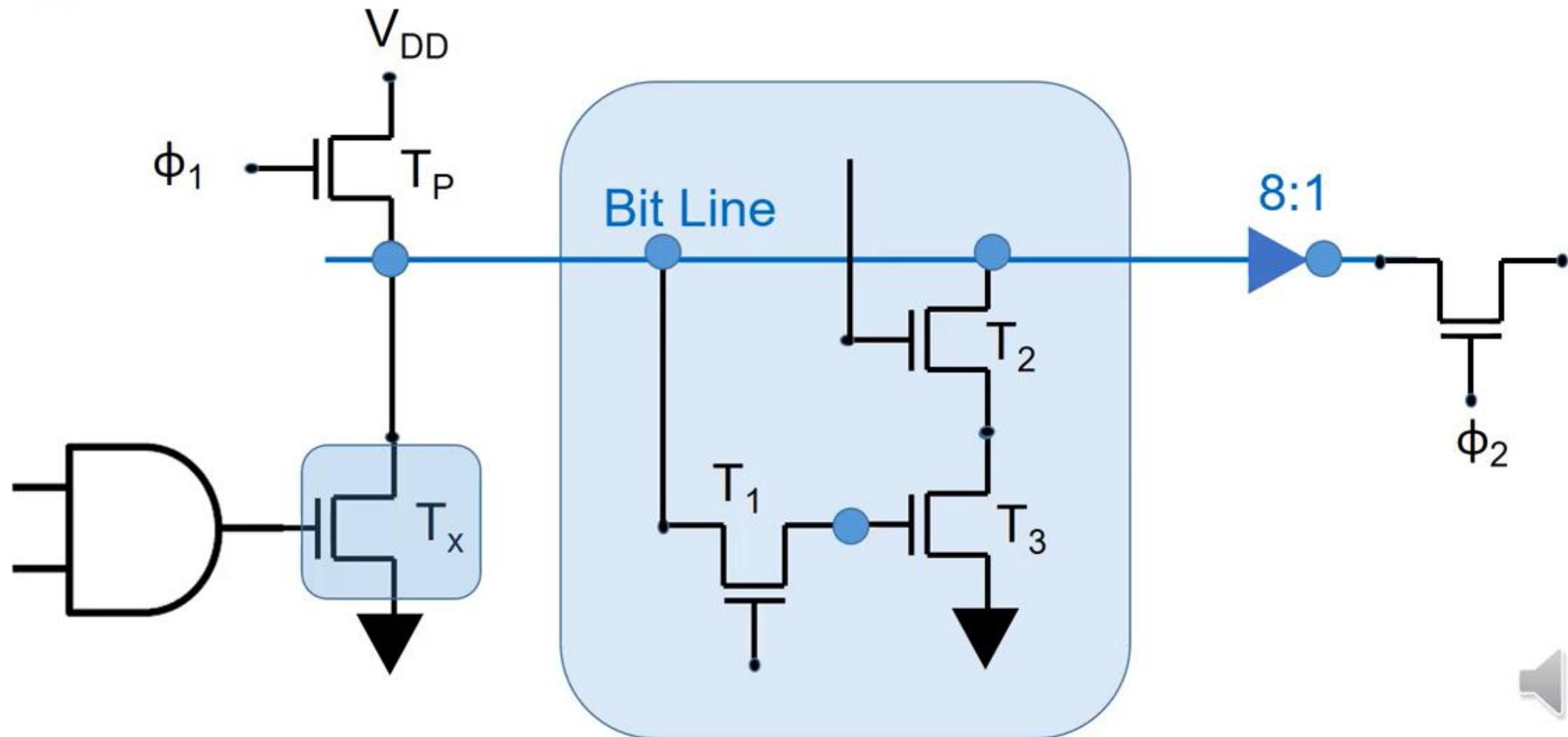
Read 0





Write 0

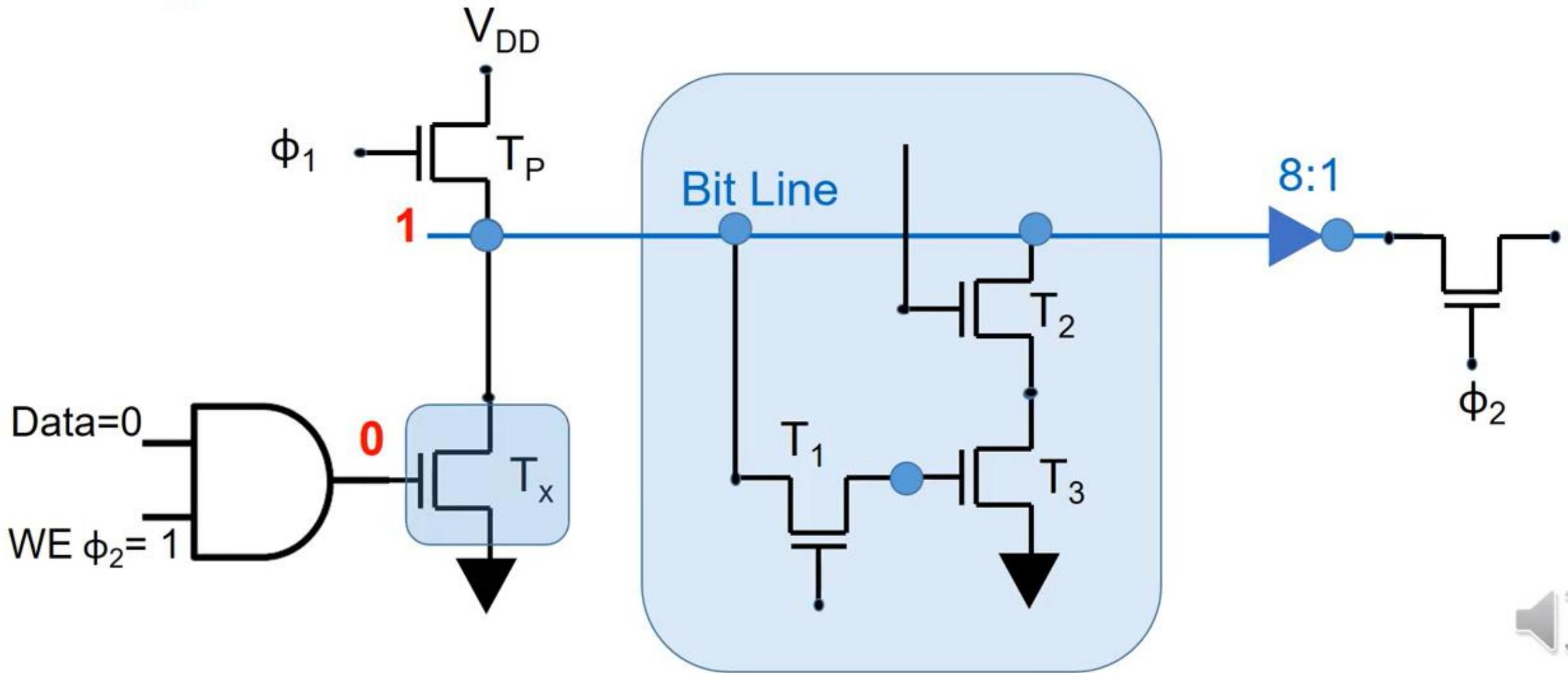
Write Operation





Write 0

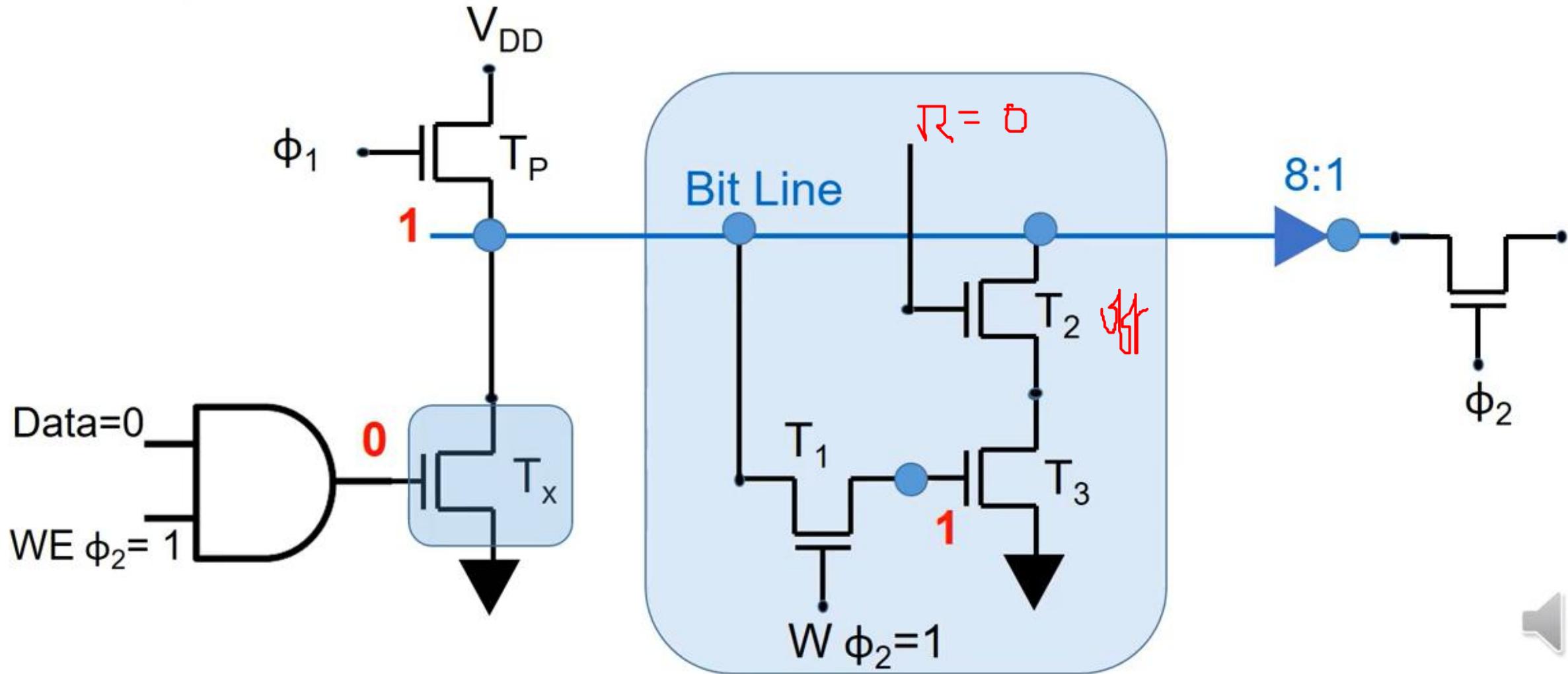
Write Operation





Write 0

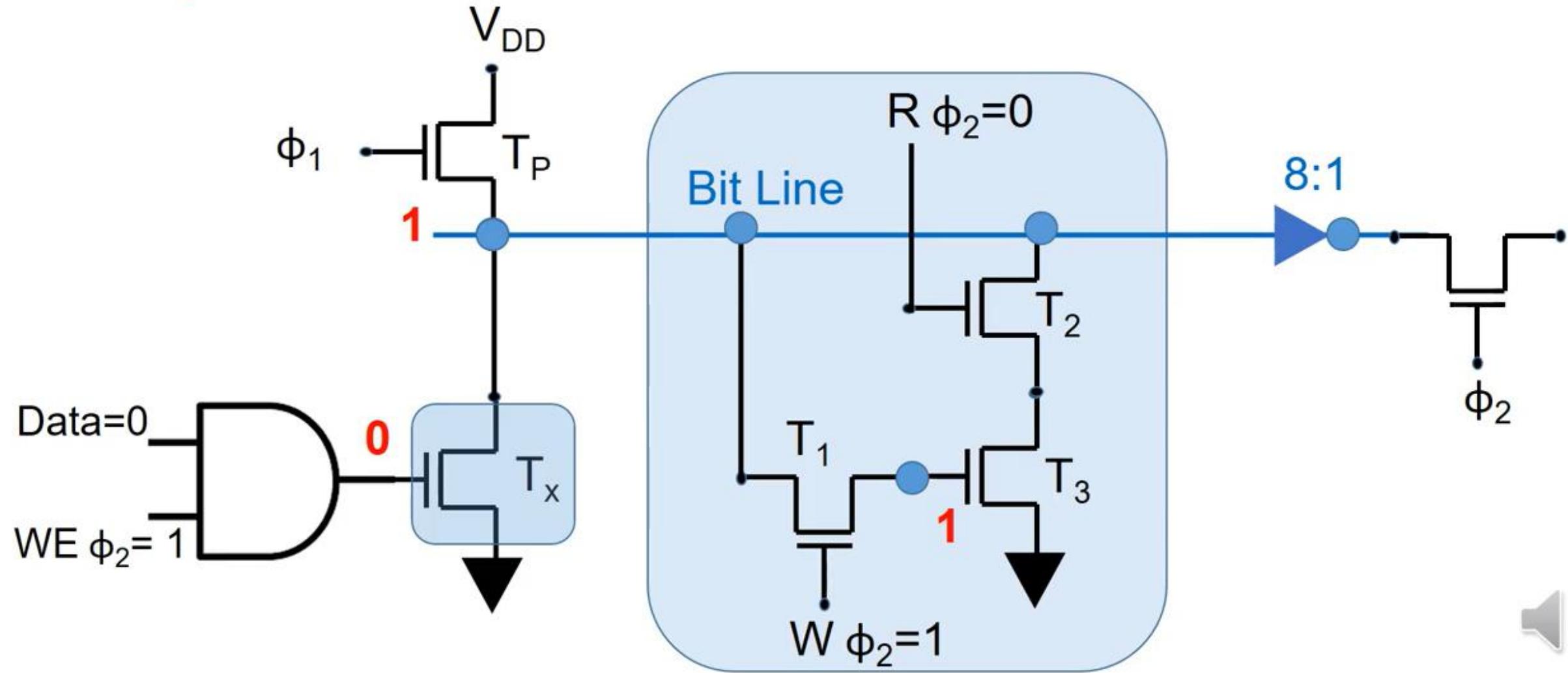
Write Operation





Write 0

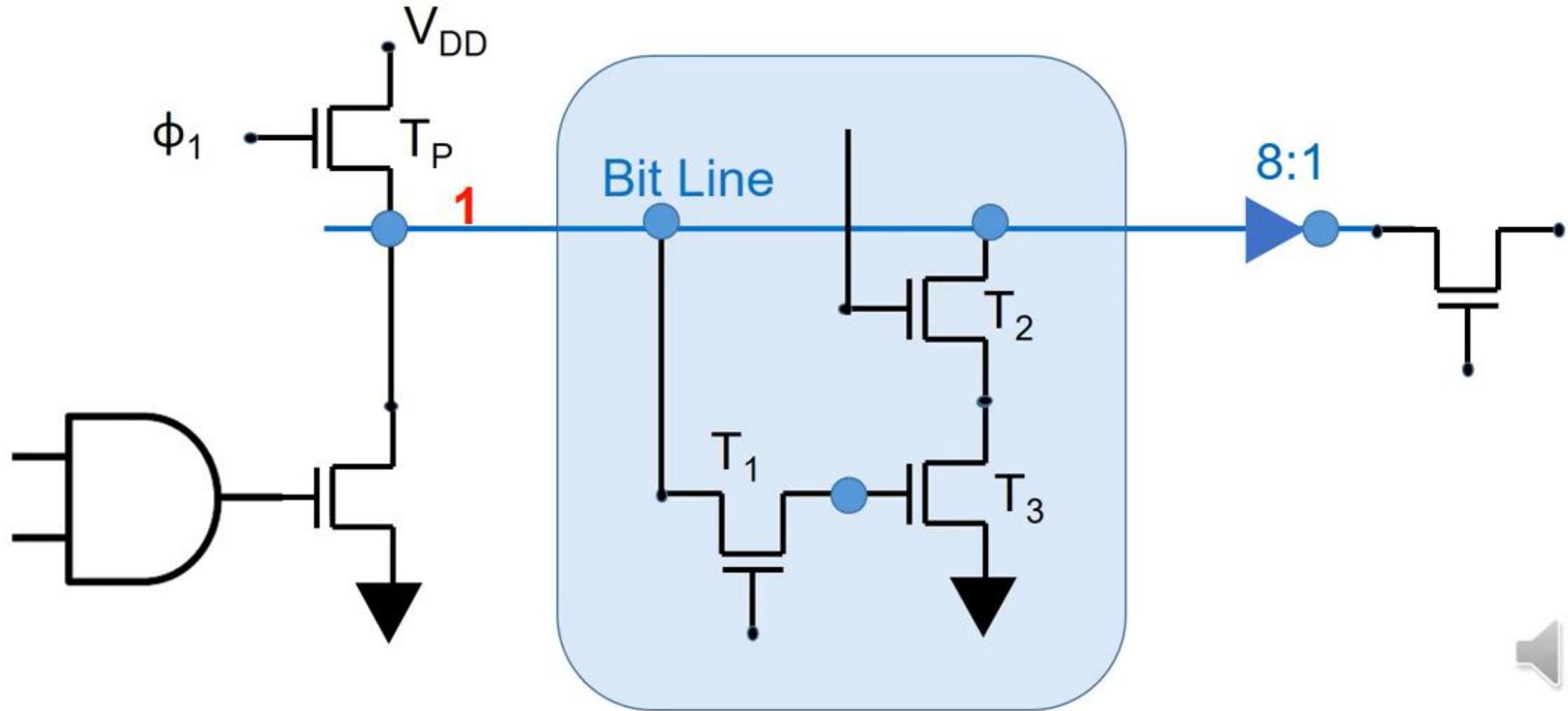
Write Operation





Read 1

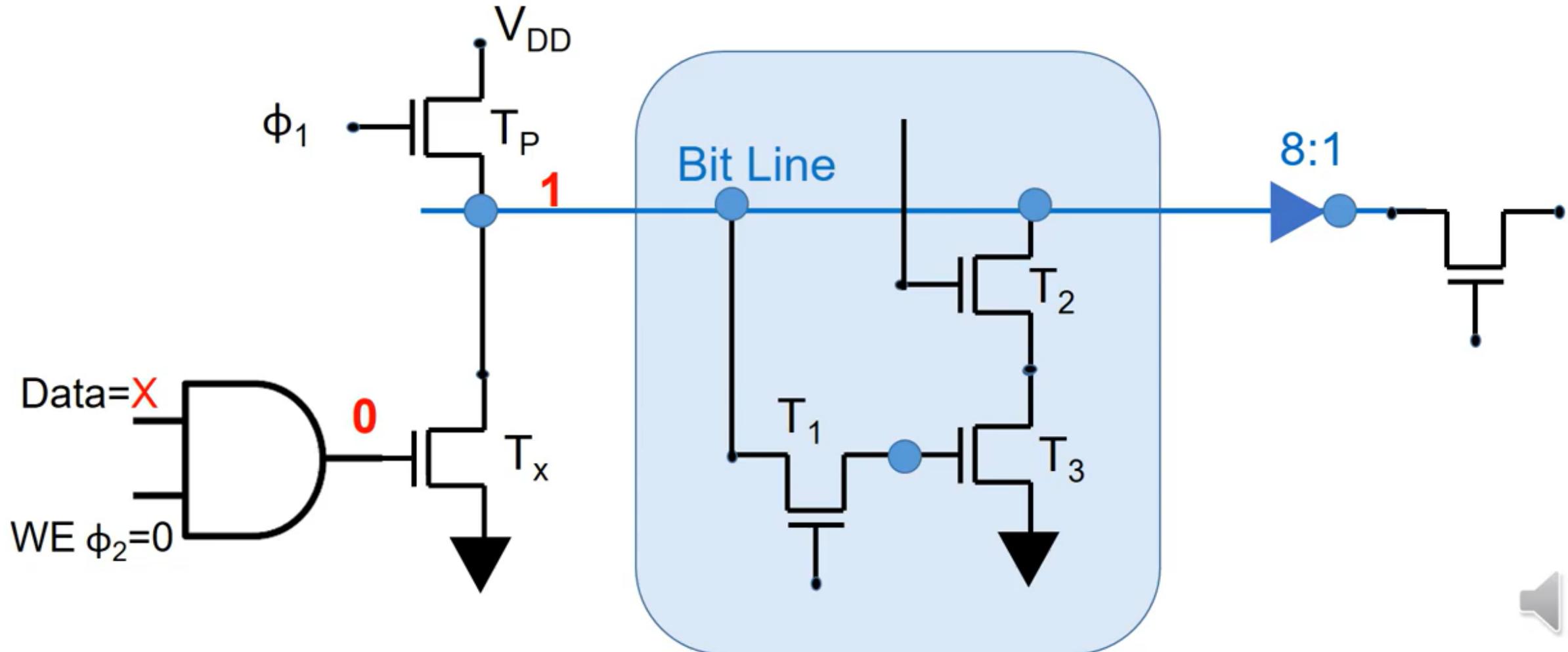
Read Operation





Read 1

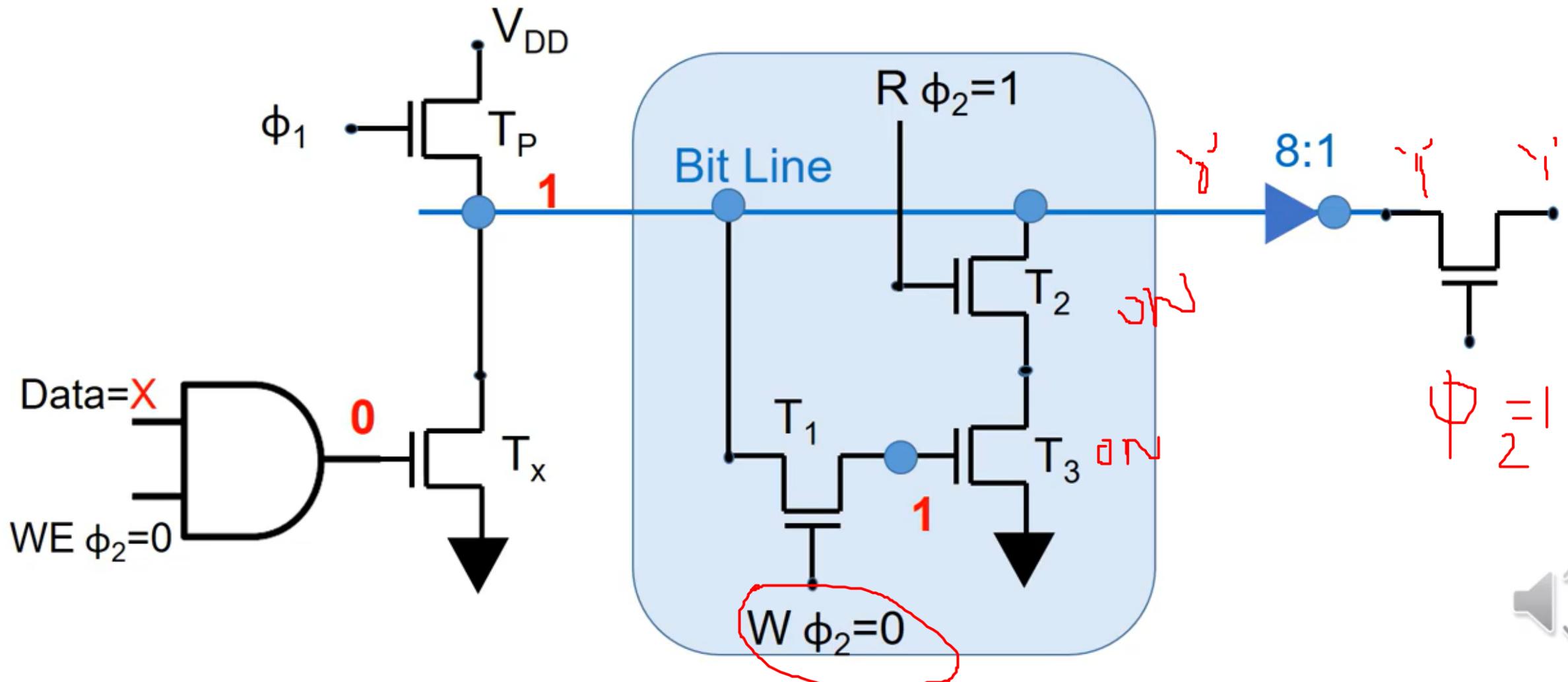
Read Operation



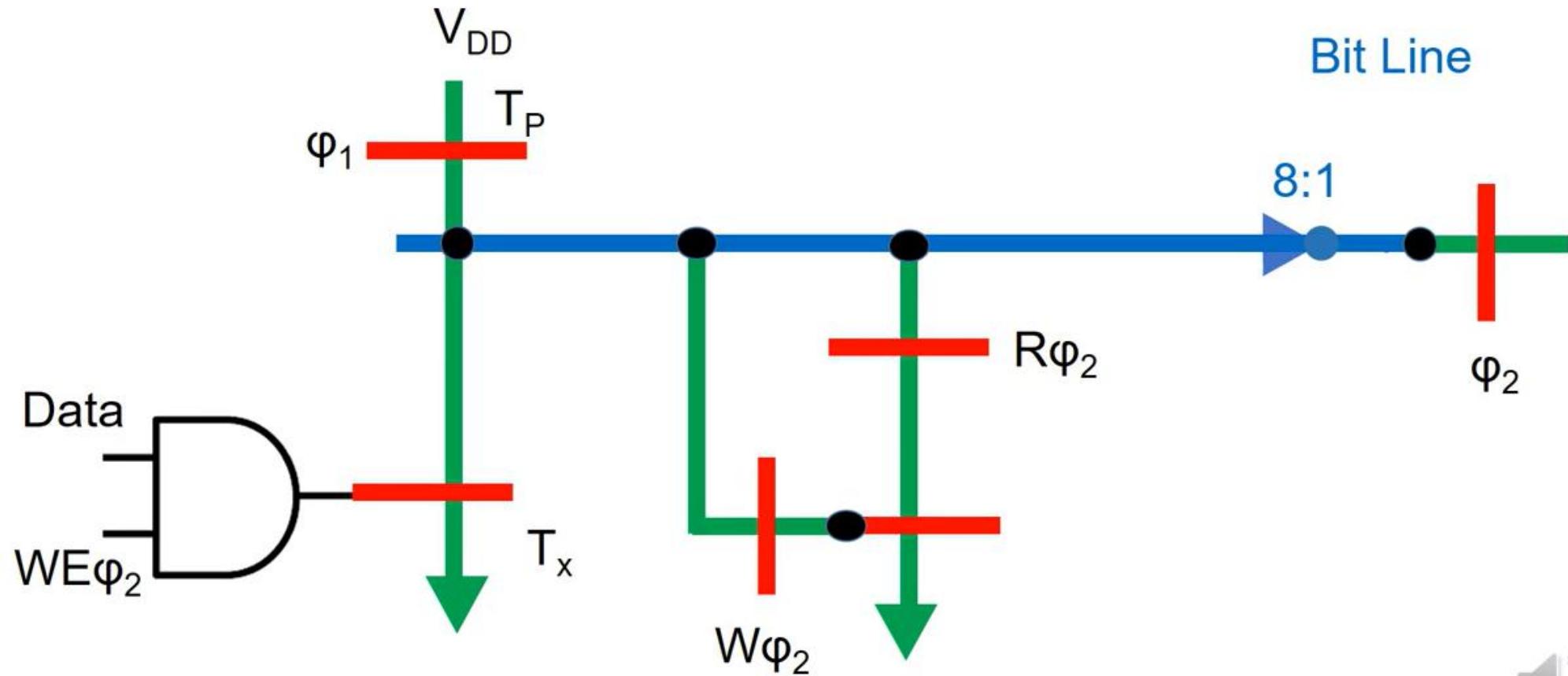


Read 1

Read Operation



Random Access Memory (3T-DRAM)

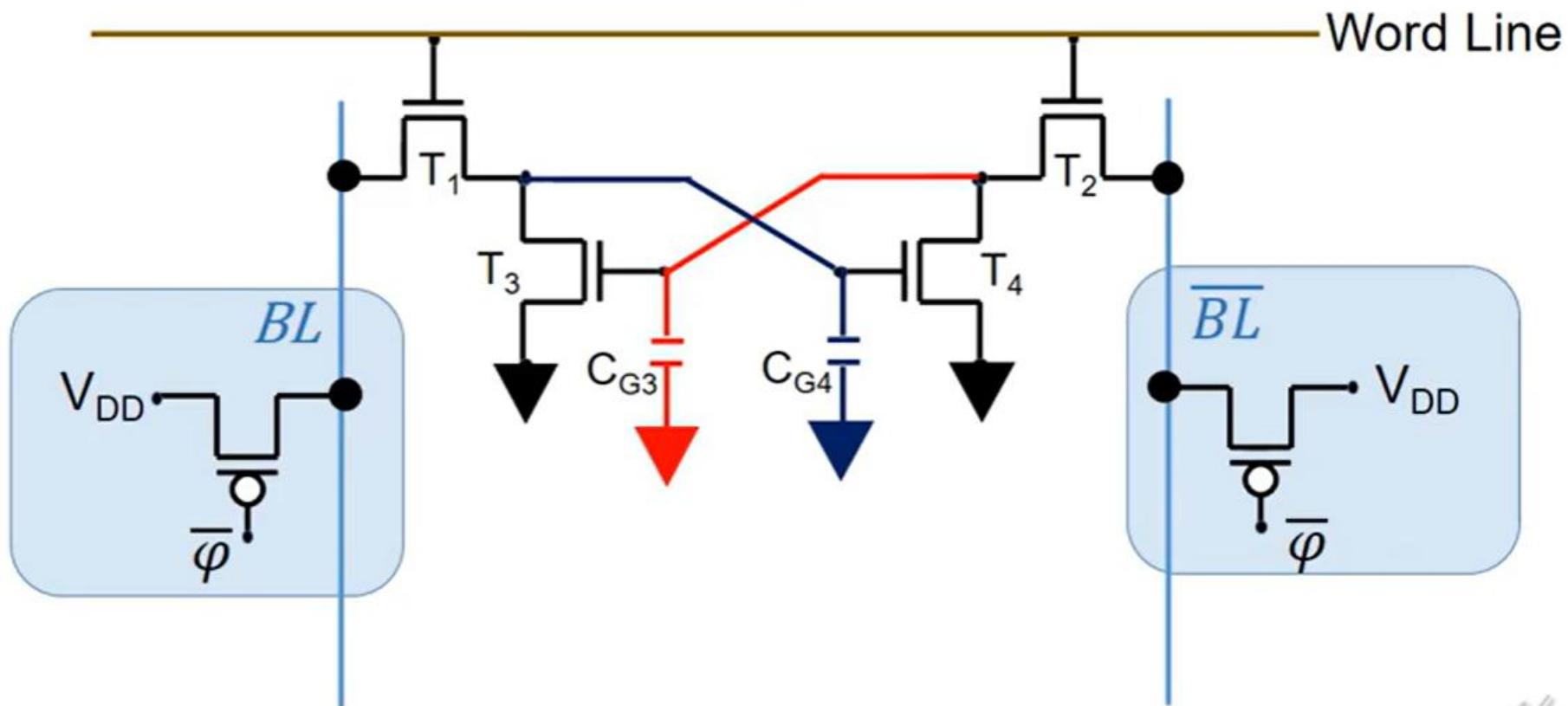




4T-DRAM (4 Transistor – Dynamic Random Access Memory)

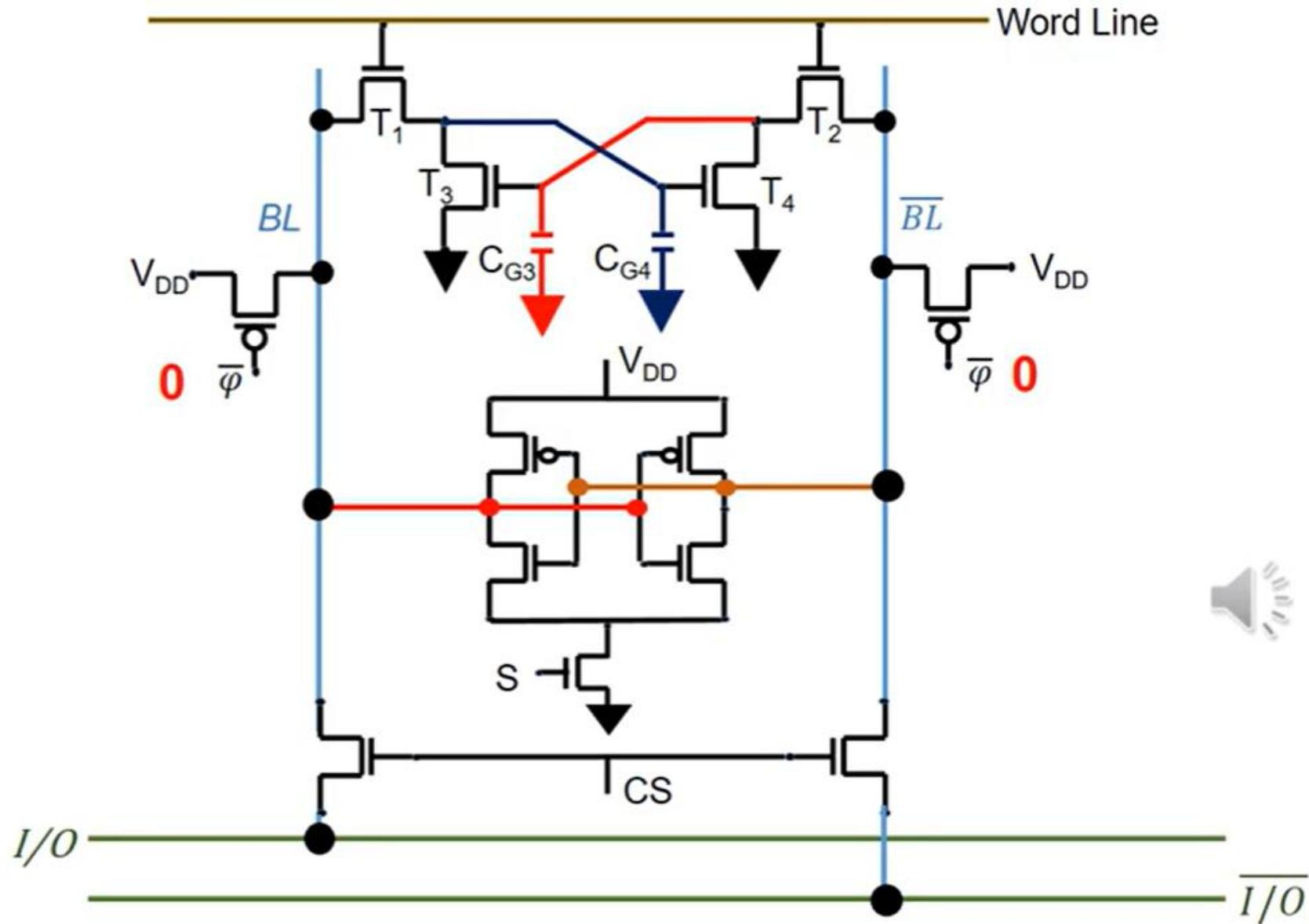


Random Access Memory (4T-DRAM)



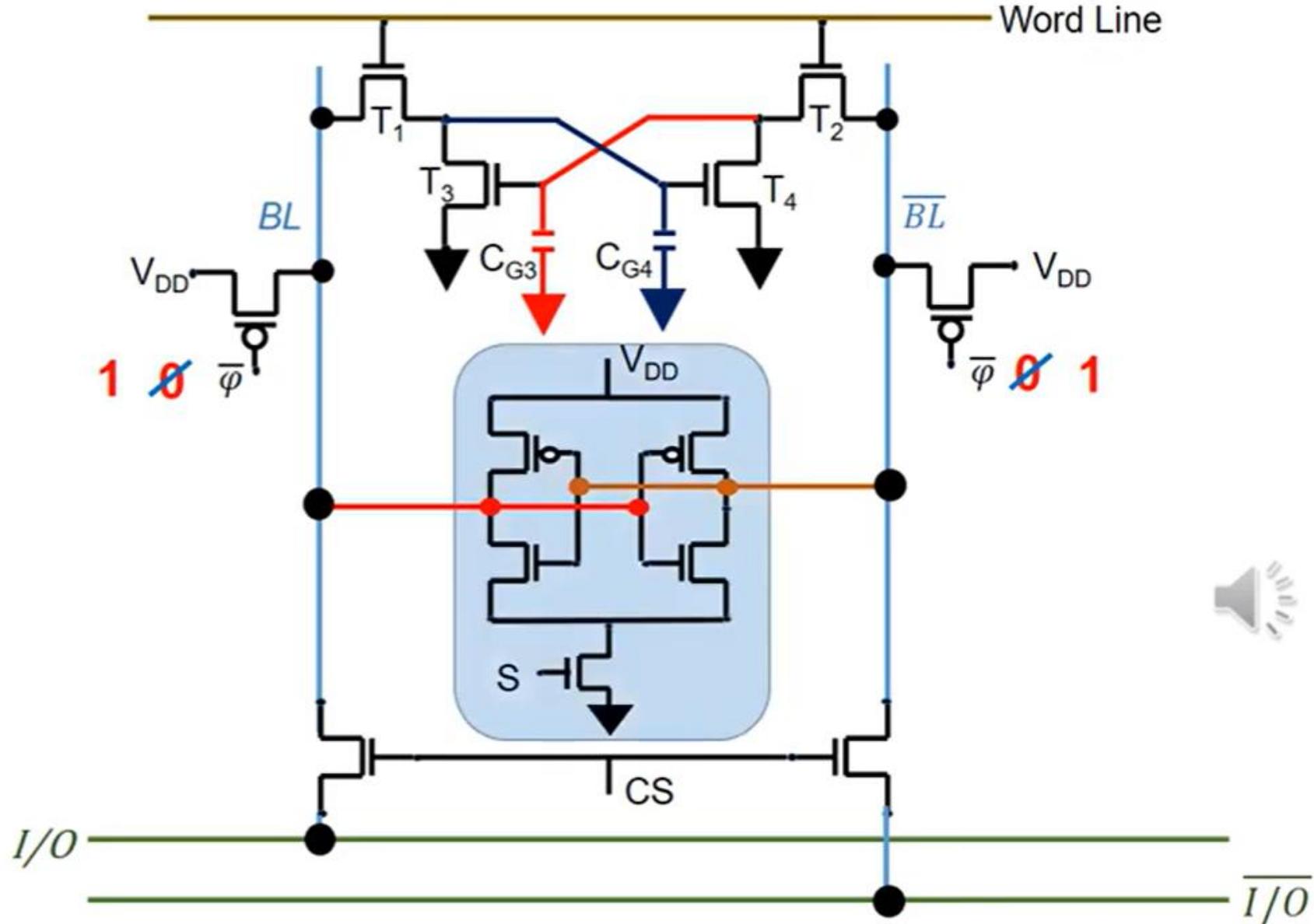


4T-DRAM Write Operation



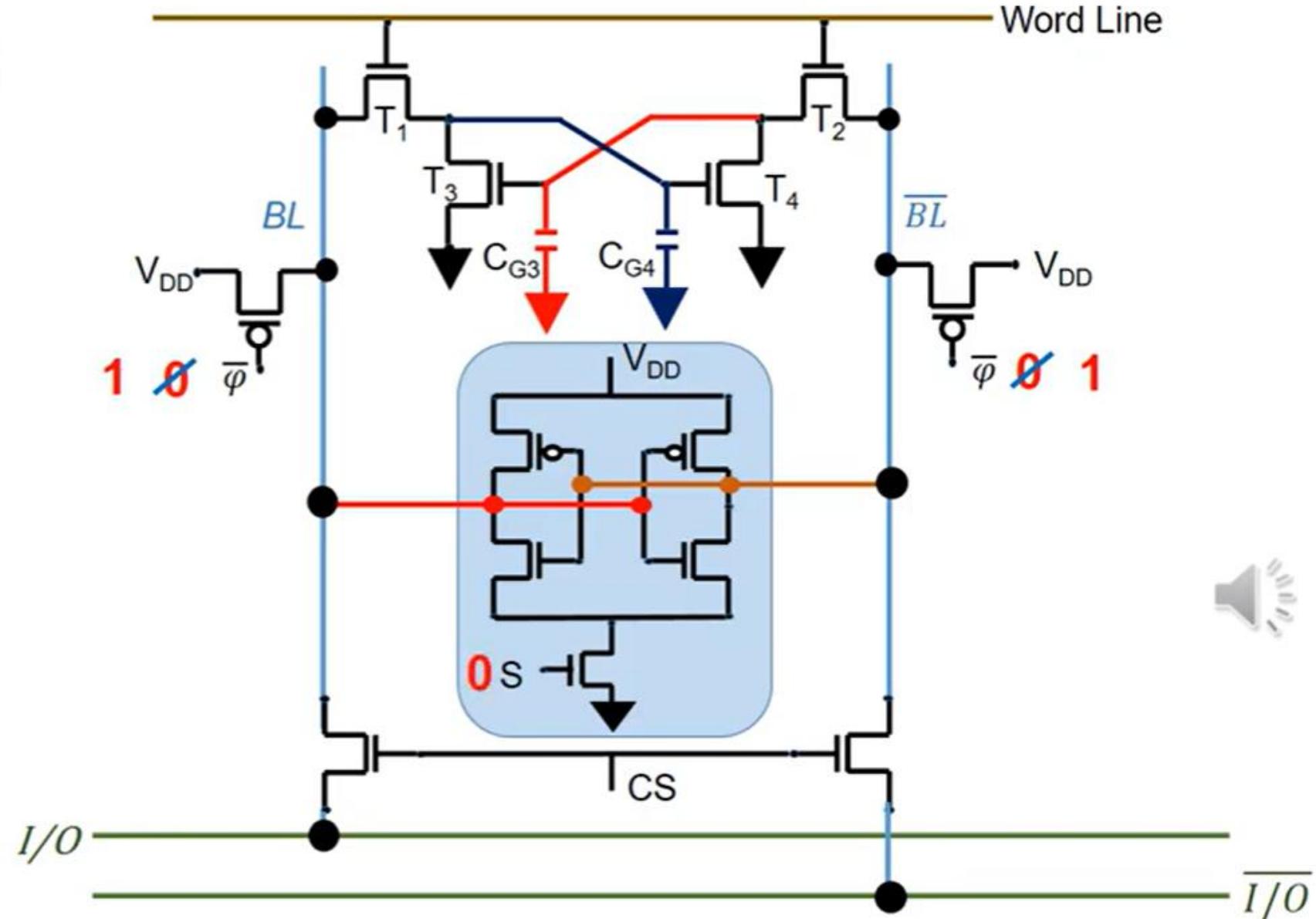


4T-DRAM Write Operation



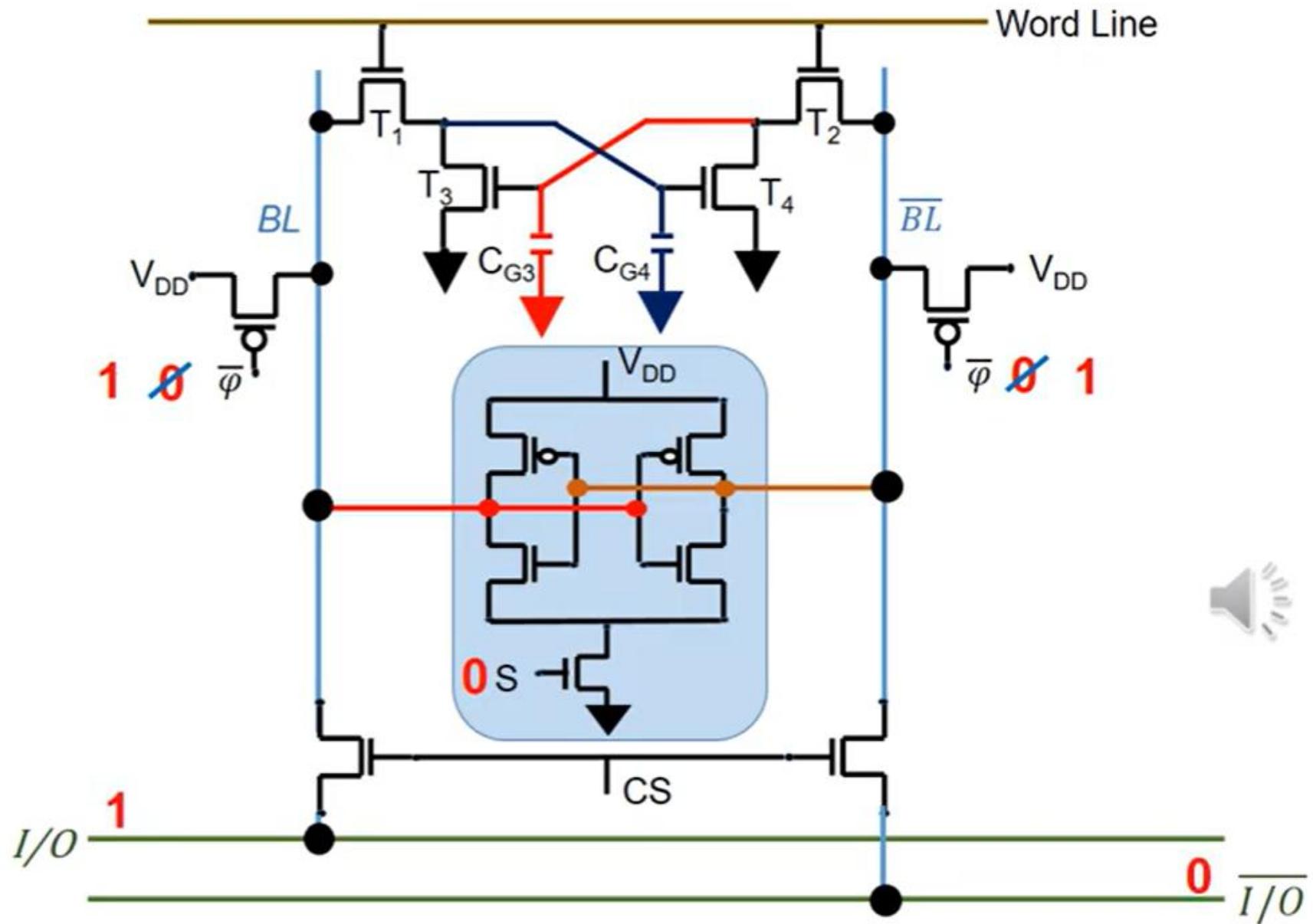


4T-DRAM Write Operation



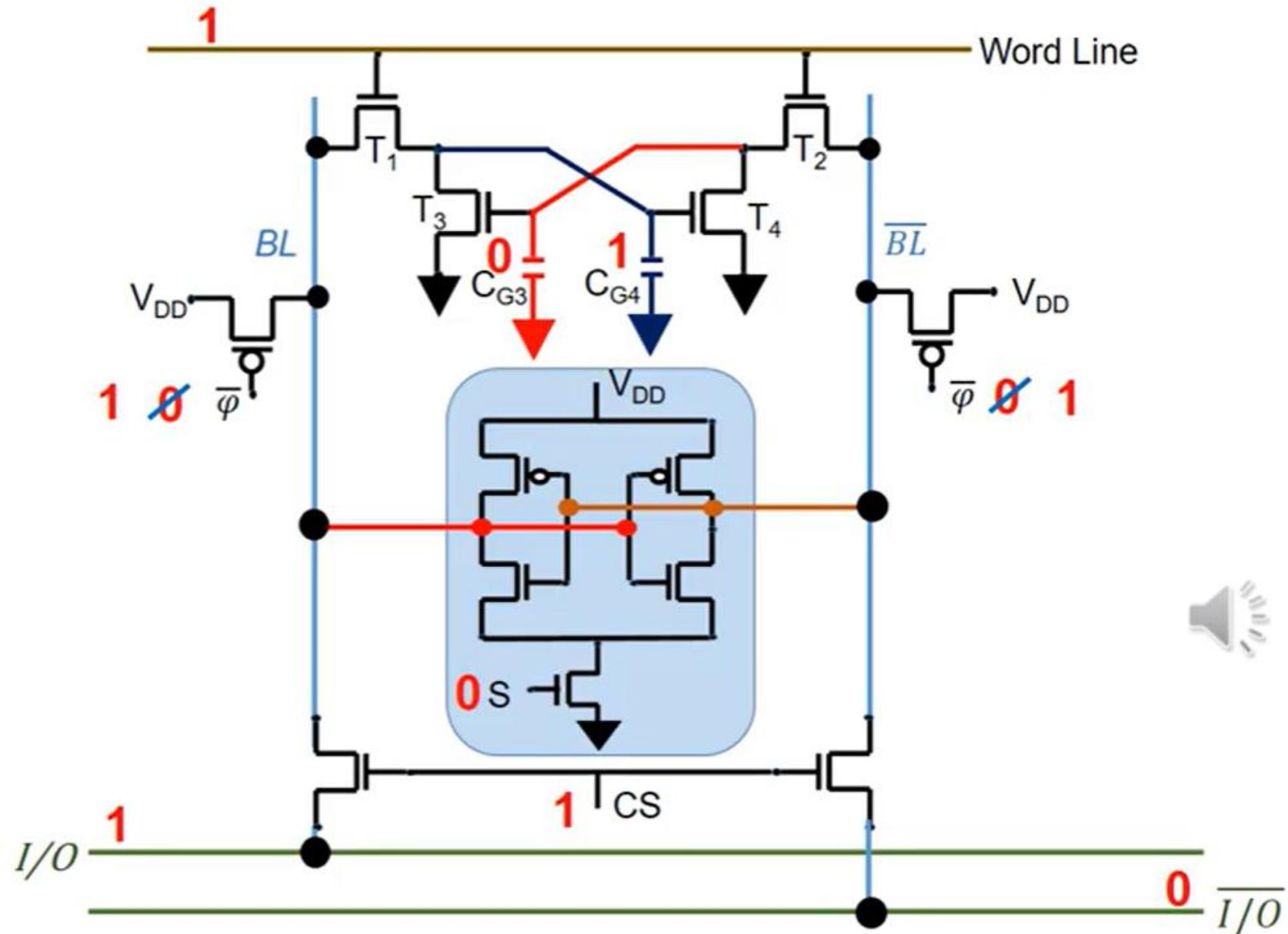


4T-DRAM Write Operation



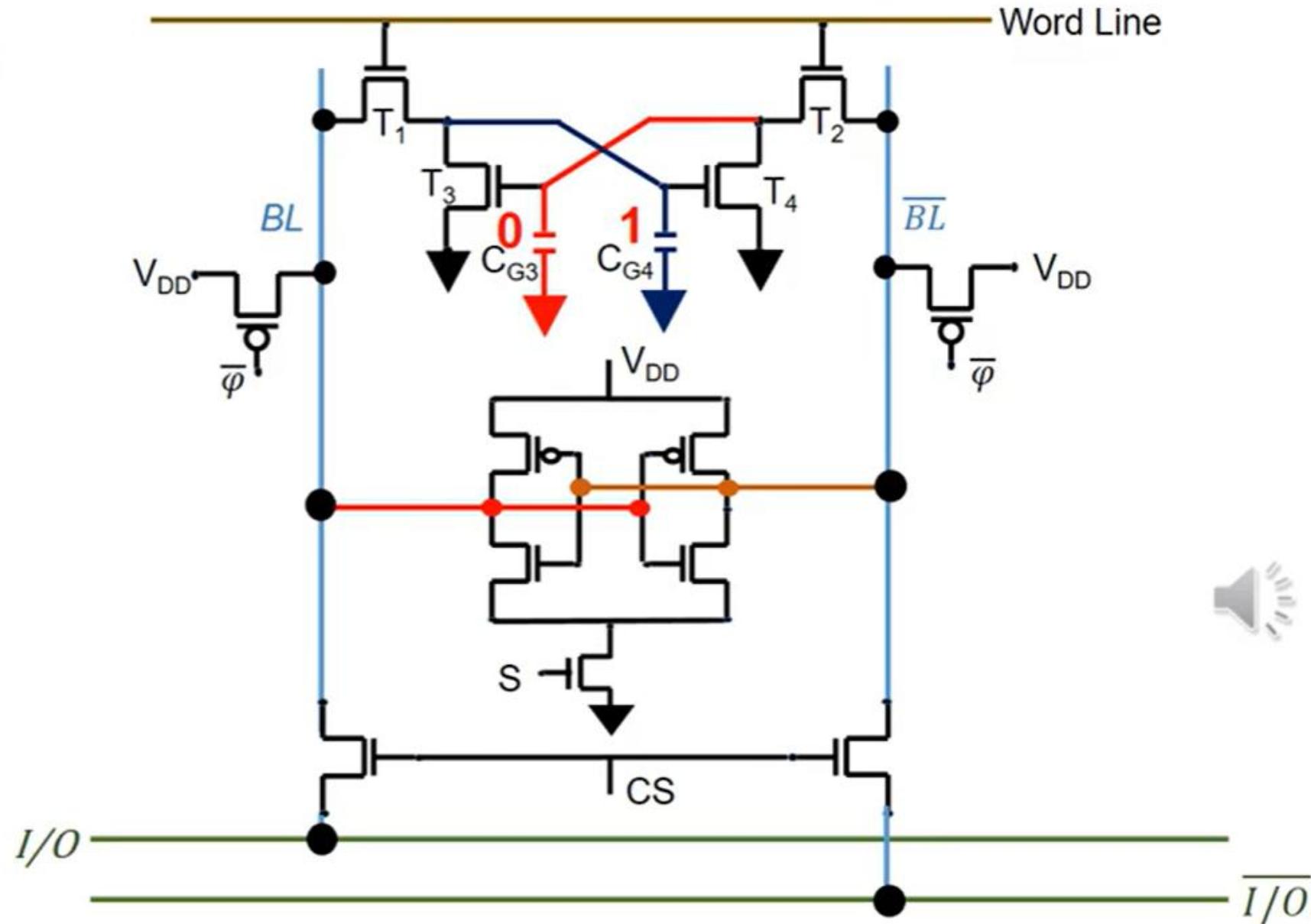


4T-DRAM Write Operation



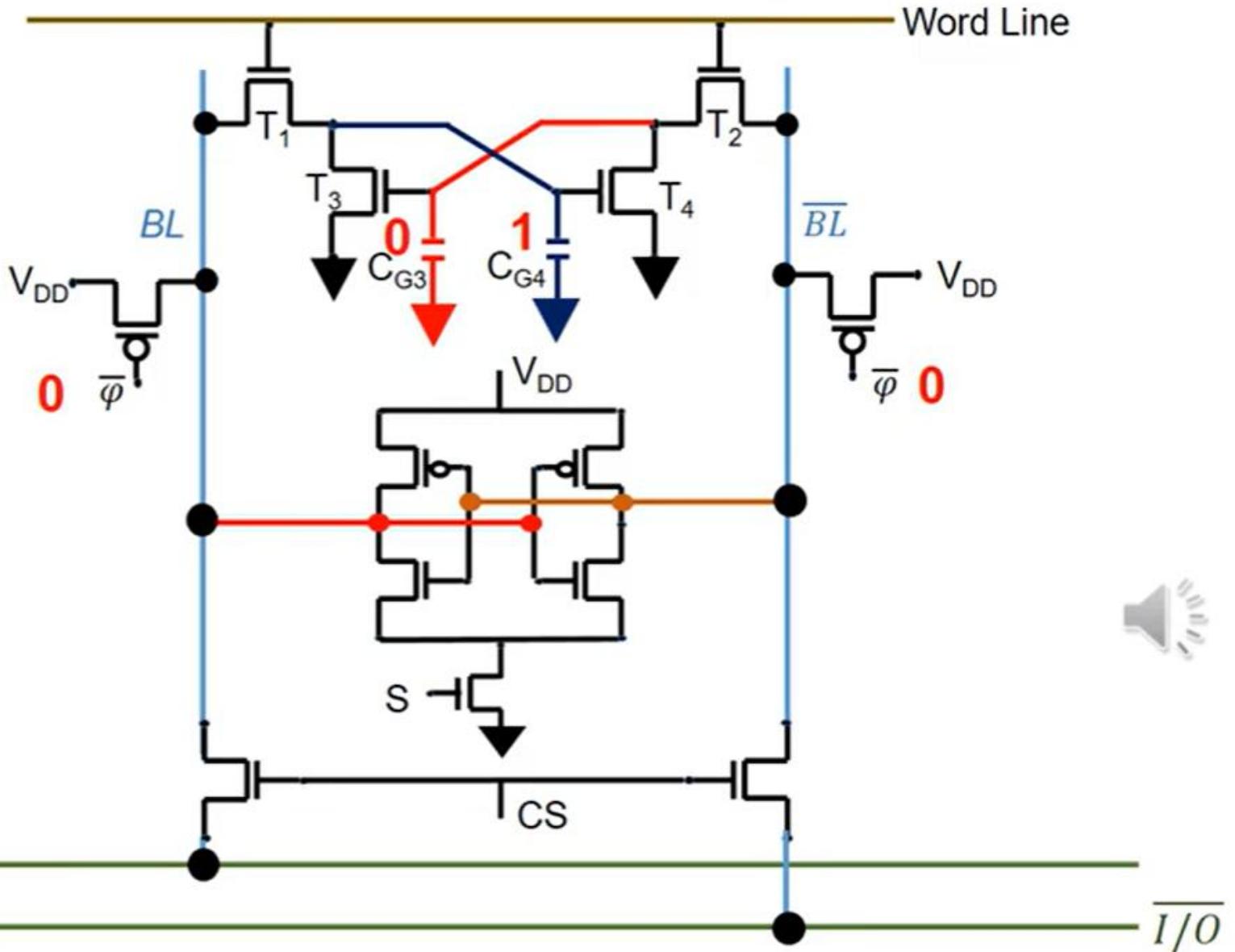


4T-DRAM Read Operation



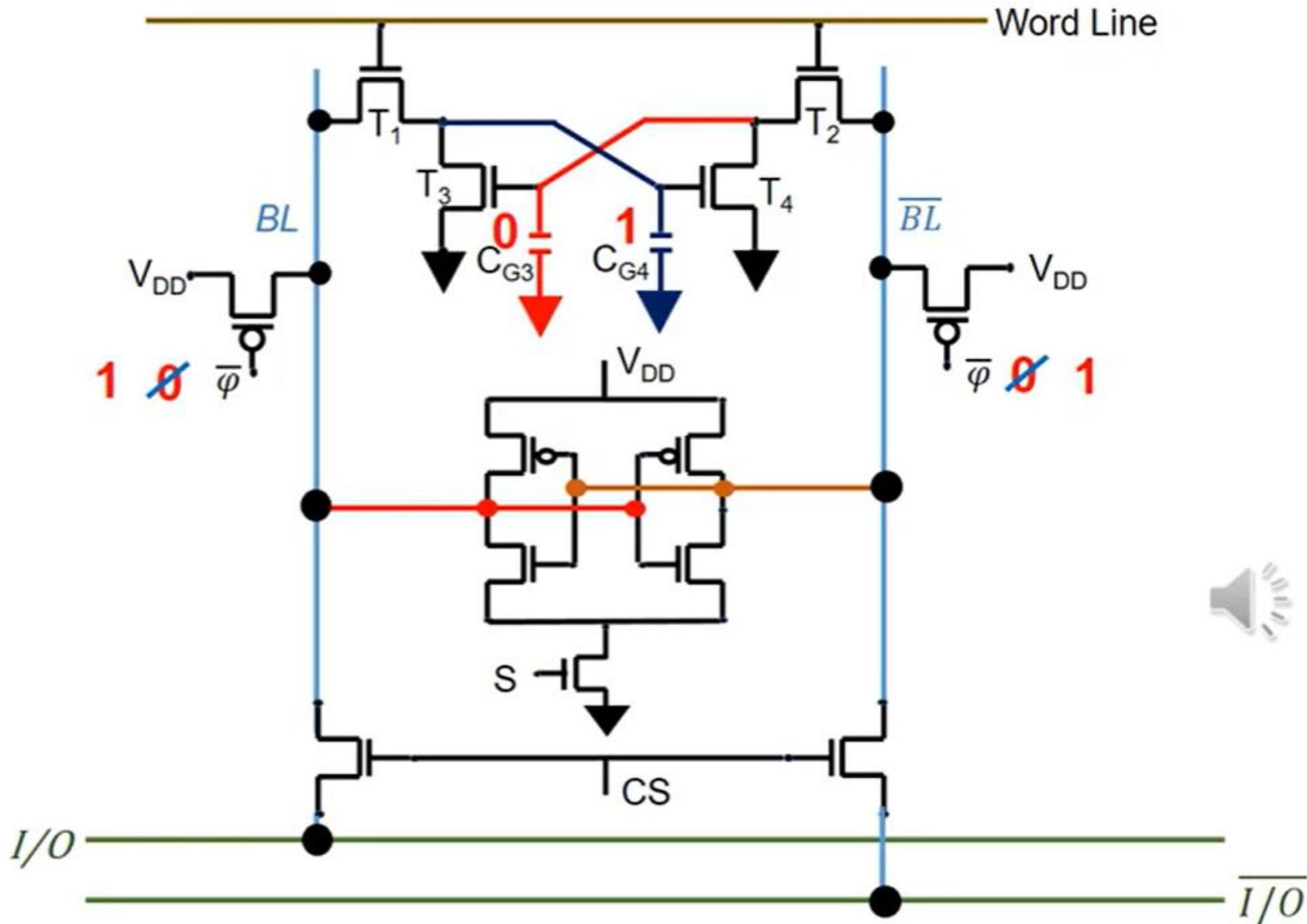


4T-DRAM Read Operation



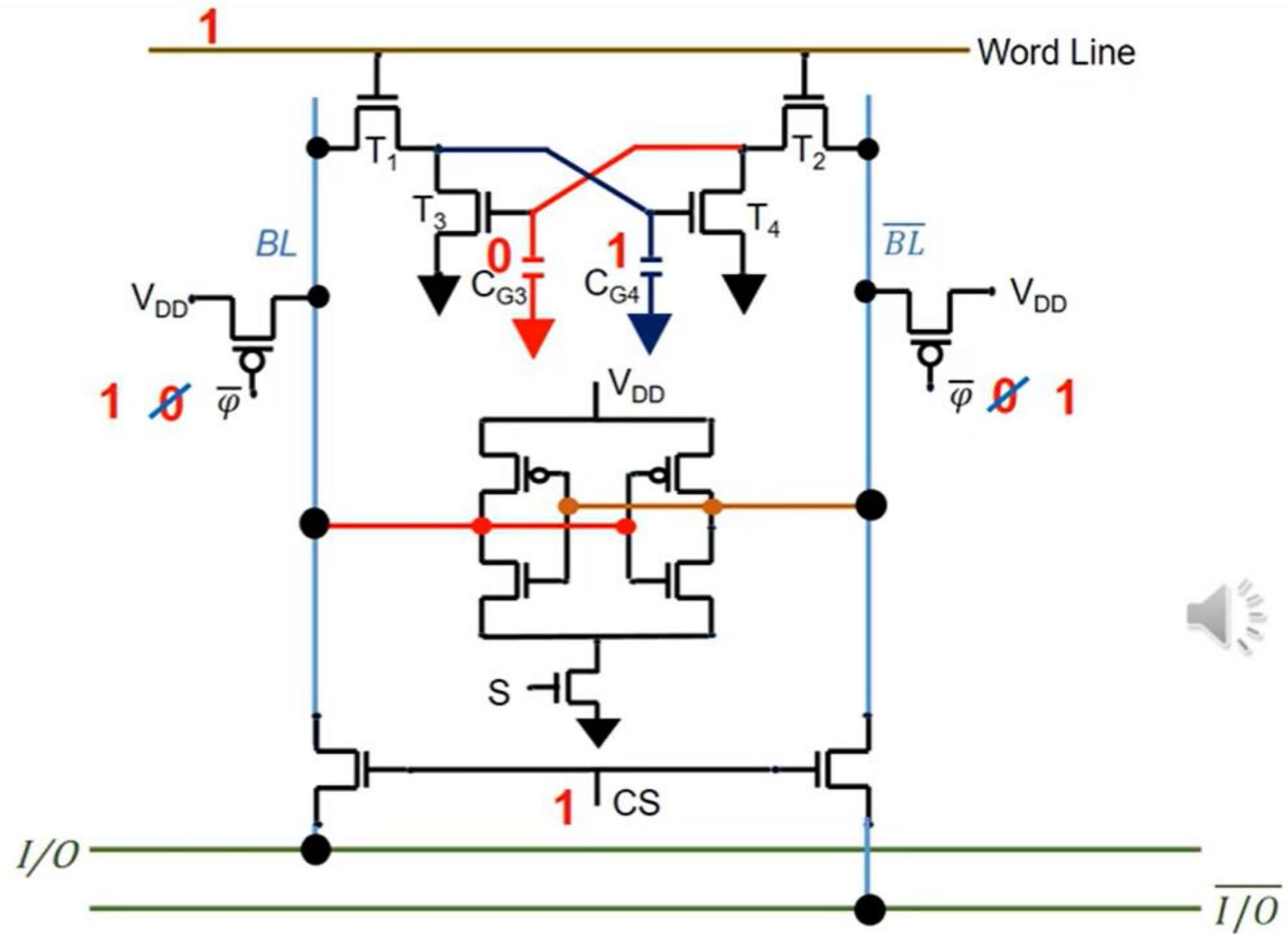


4T-DRAM Read Operation



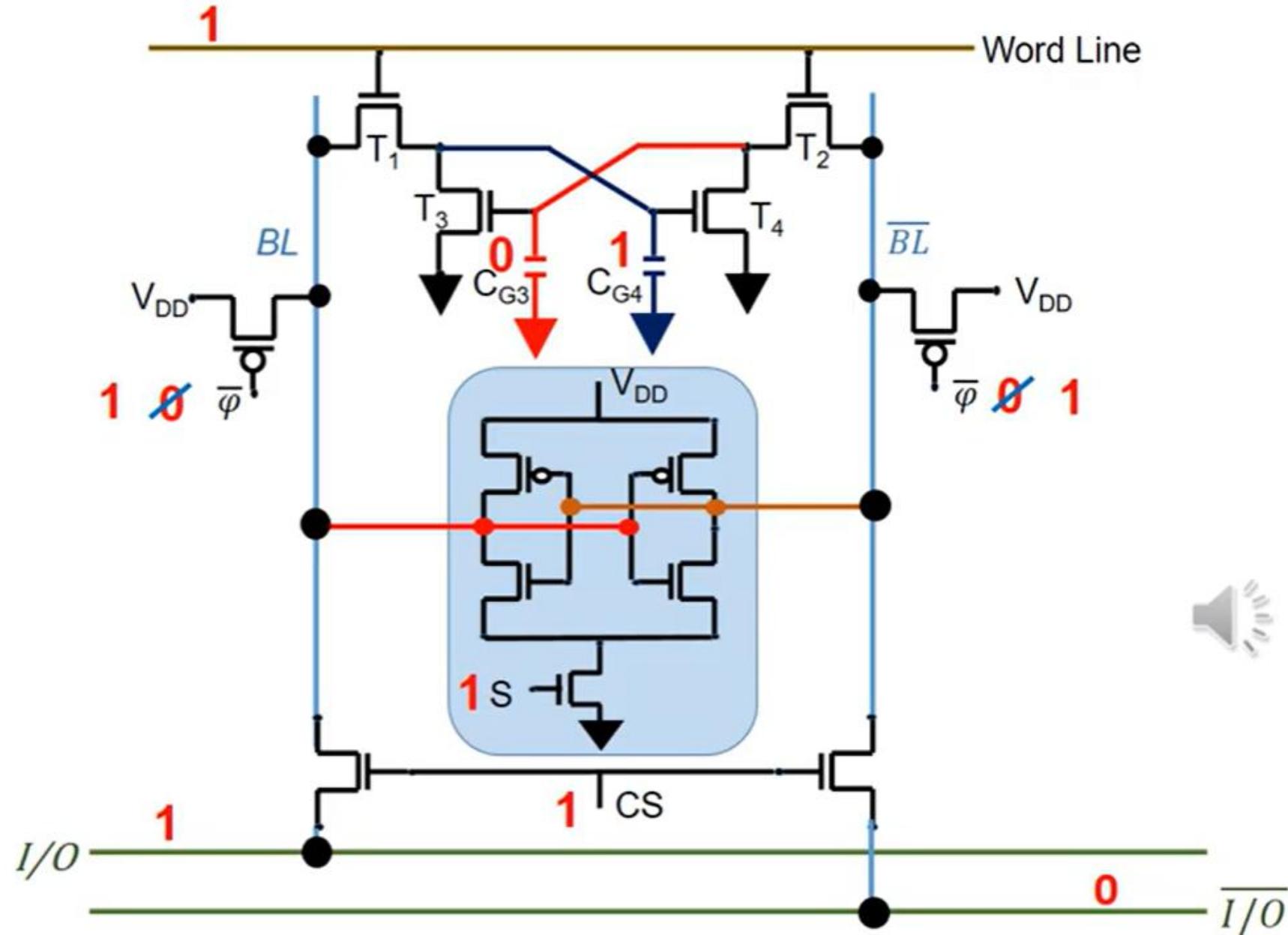


4T-DRAM Read Operation





4T-DRAM Read Operation

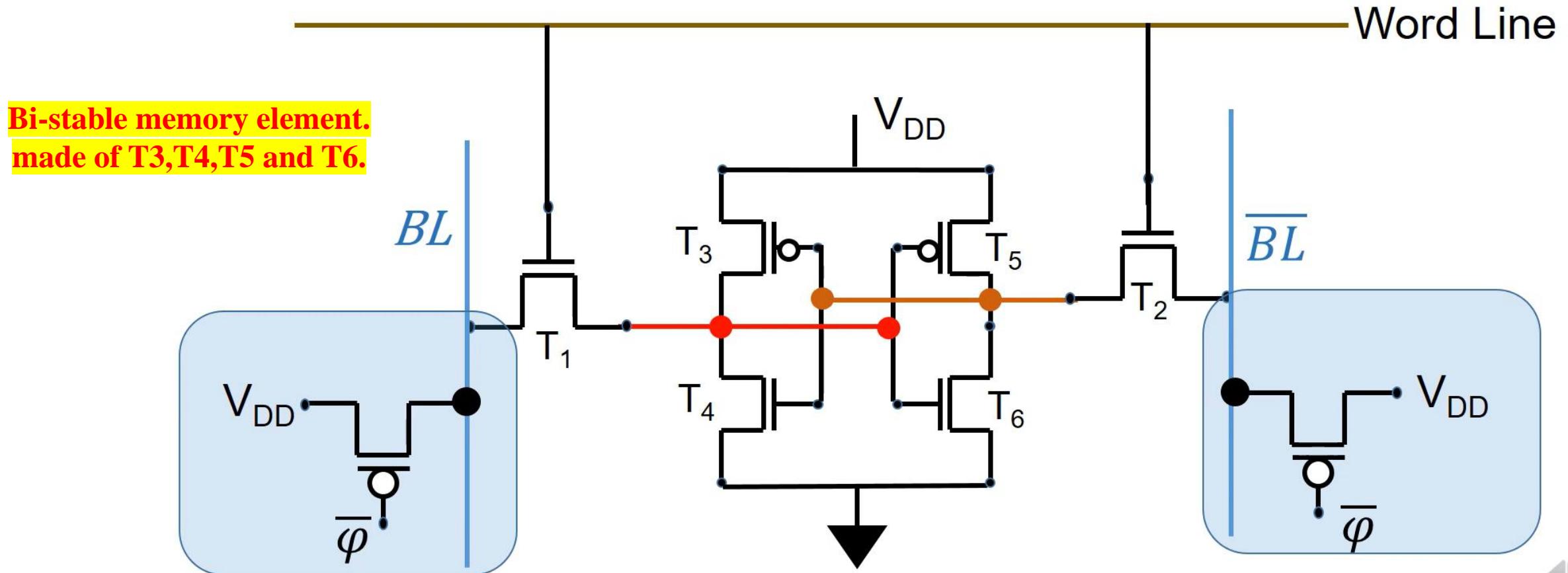




6T-SRAM (6 Transistor – Static Random Access Memory)



Random Access Memory (6T-SRAM)

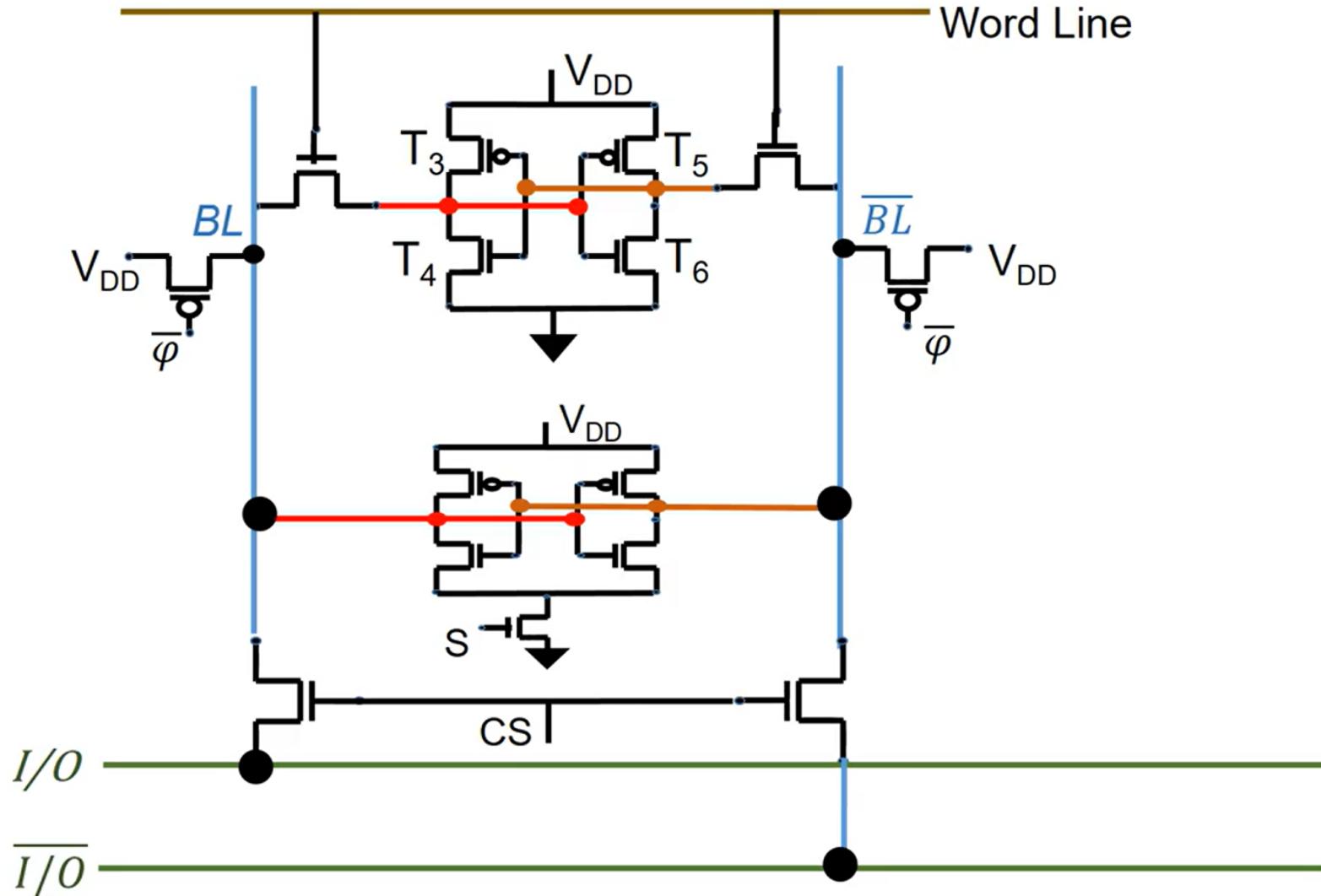


Uses Pre-charge logic, before each Read and Write we need to precharge BL and \overline{BL} '





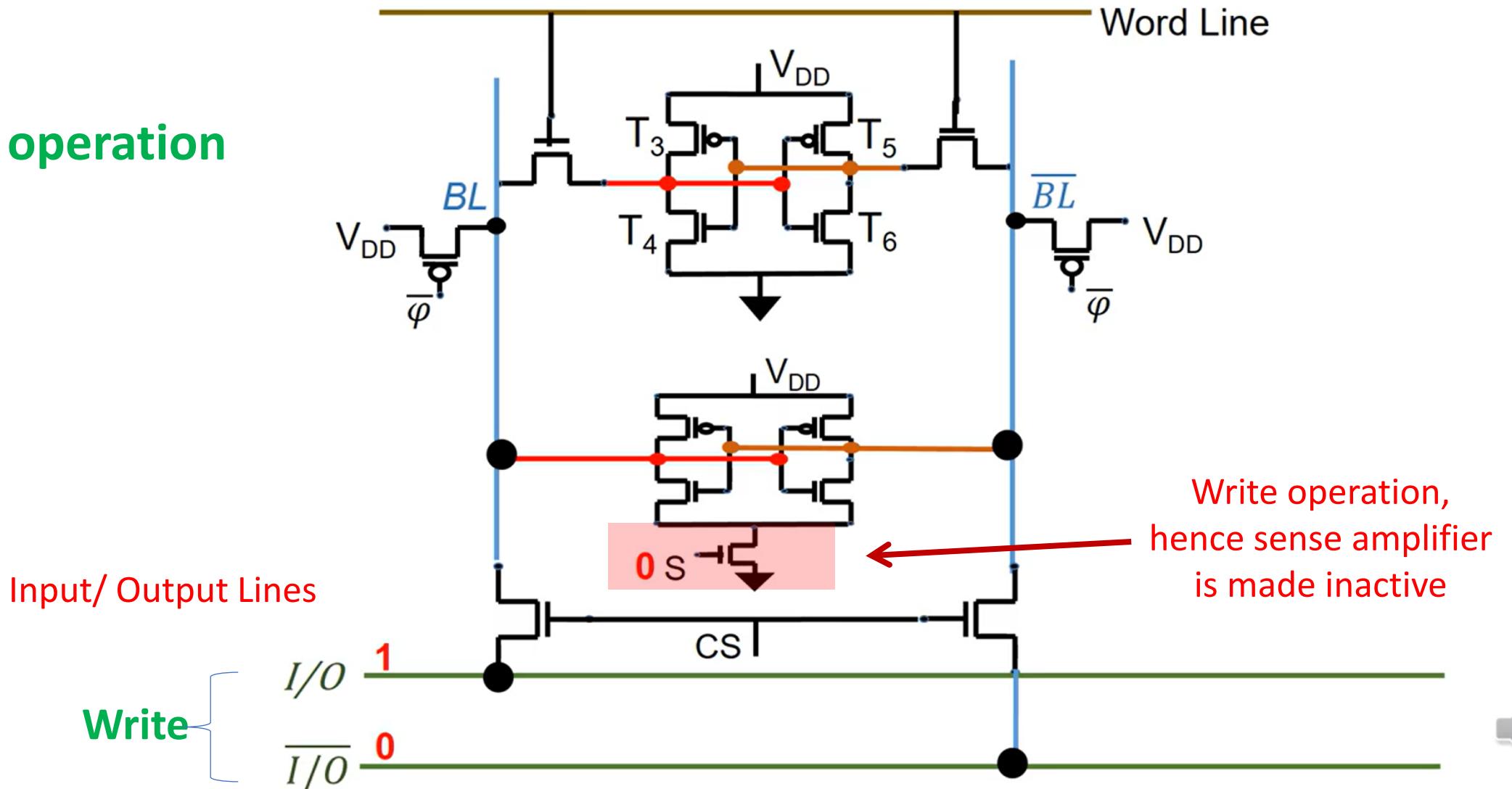
Random Access Memory (6T-SRAM)





Random Access Memory (6T-SRAM)

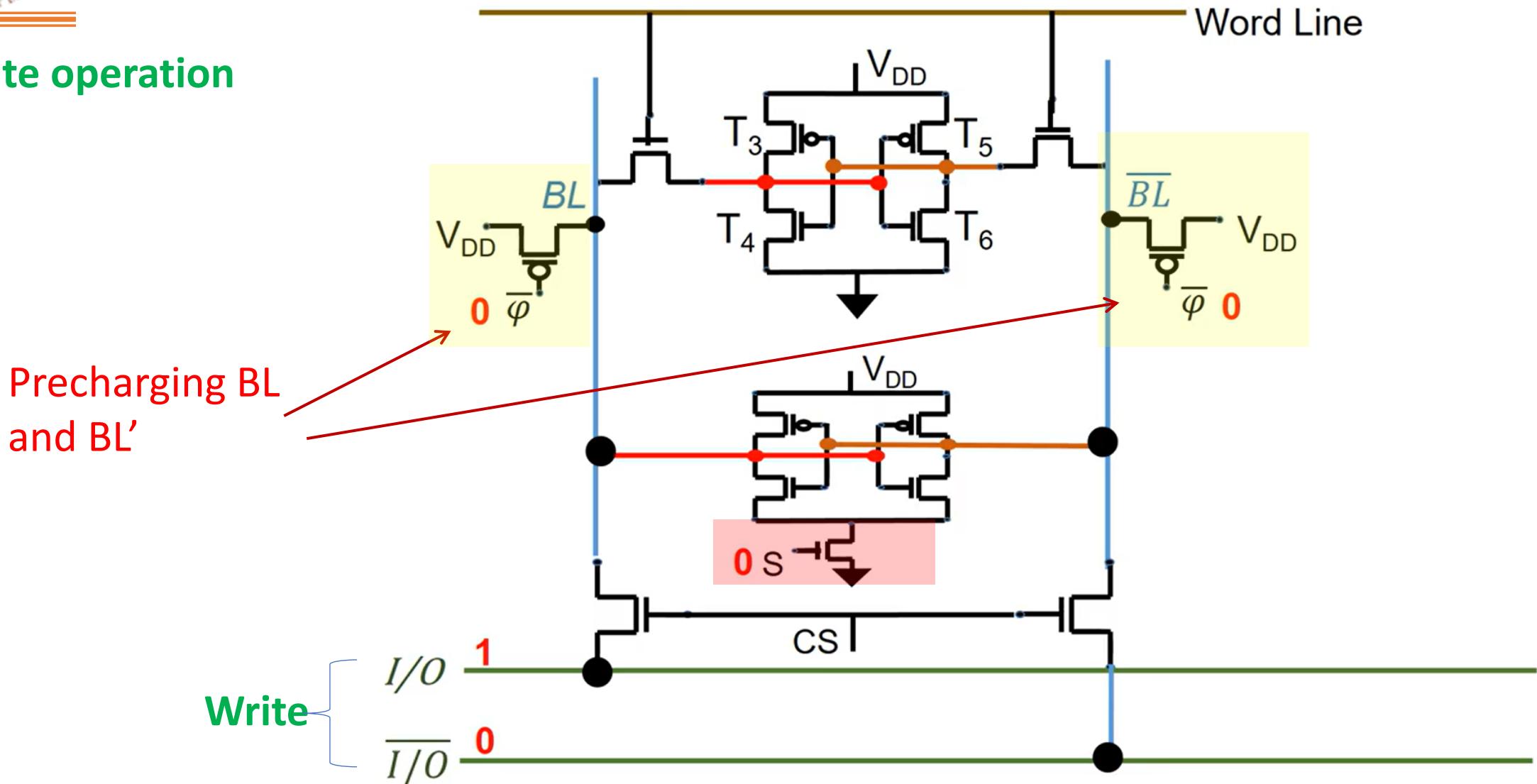
Write operation





Random Access Memory (6T-SRAM)

Write operation





Random Access Memory (6T-SRAM)

Write operation

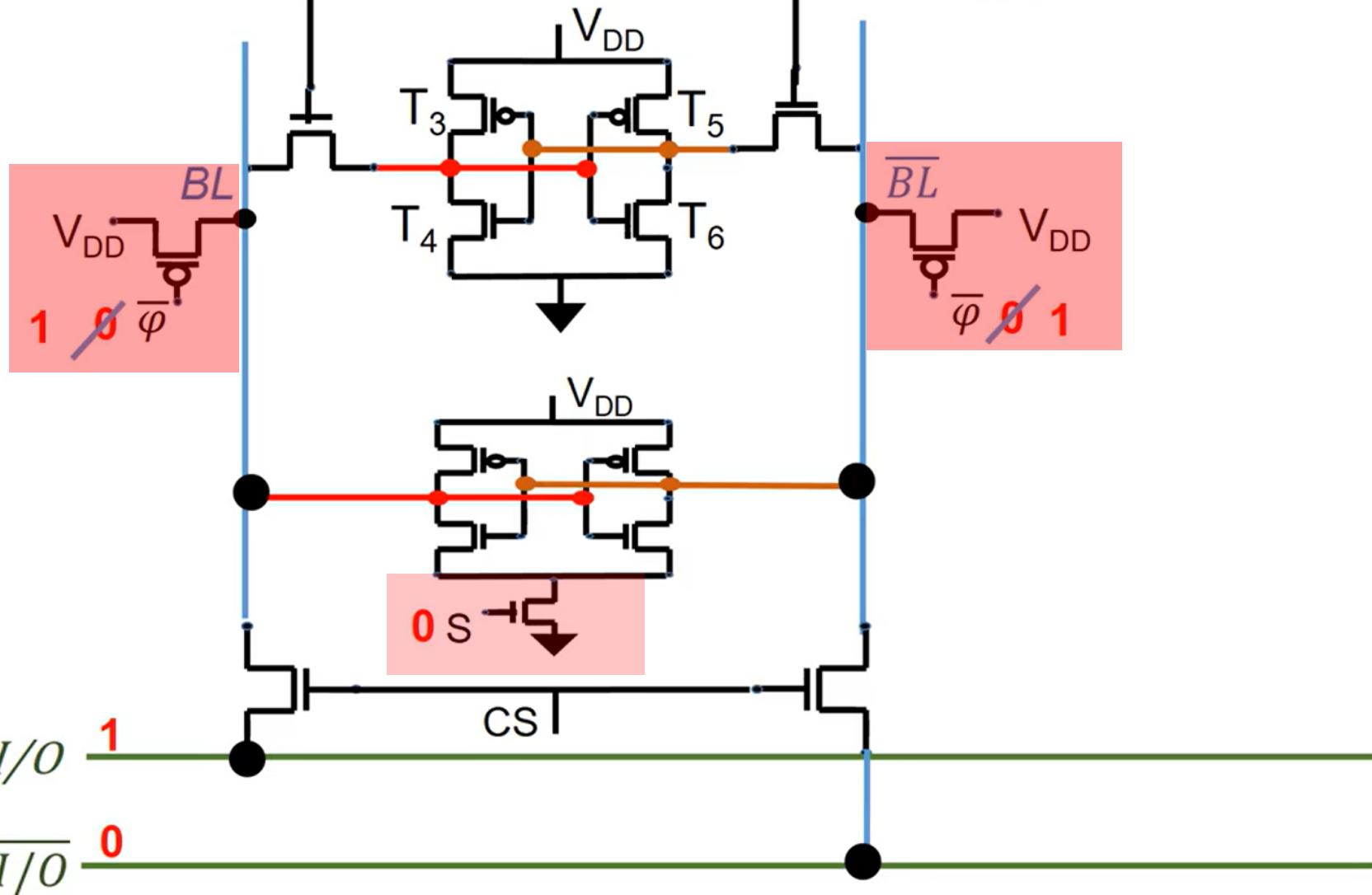
Precharging
phase ends

Write

I/O 1
 $\overline{I/O}$ 0

0

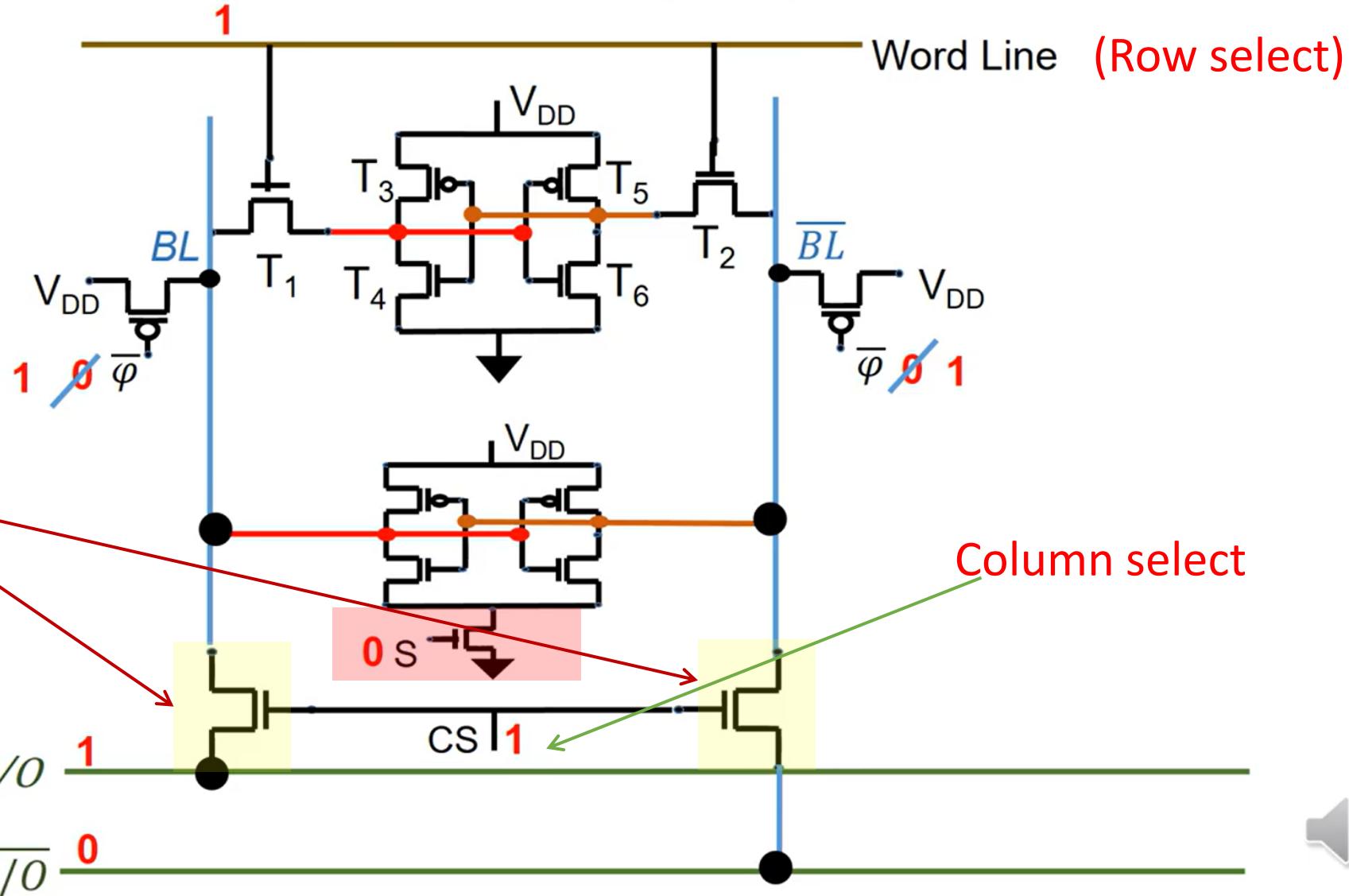
Word Line



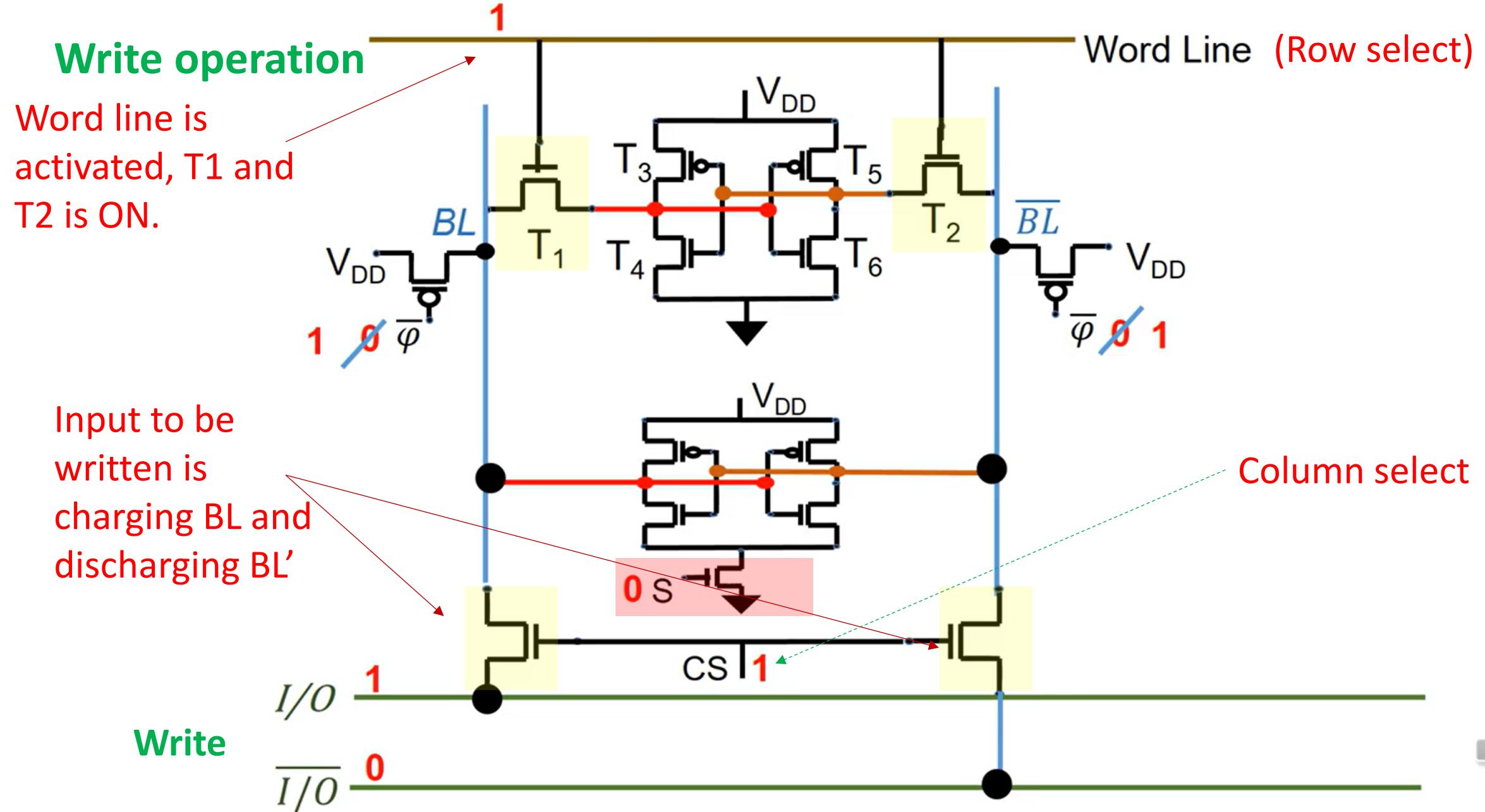


Random Access Memory (6T-SRAM)

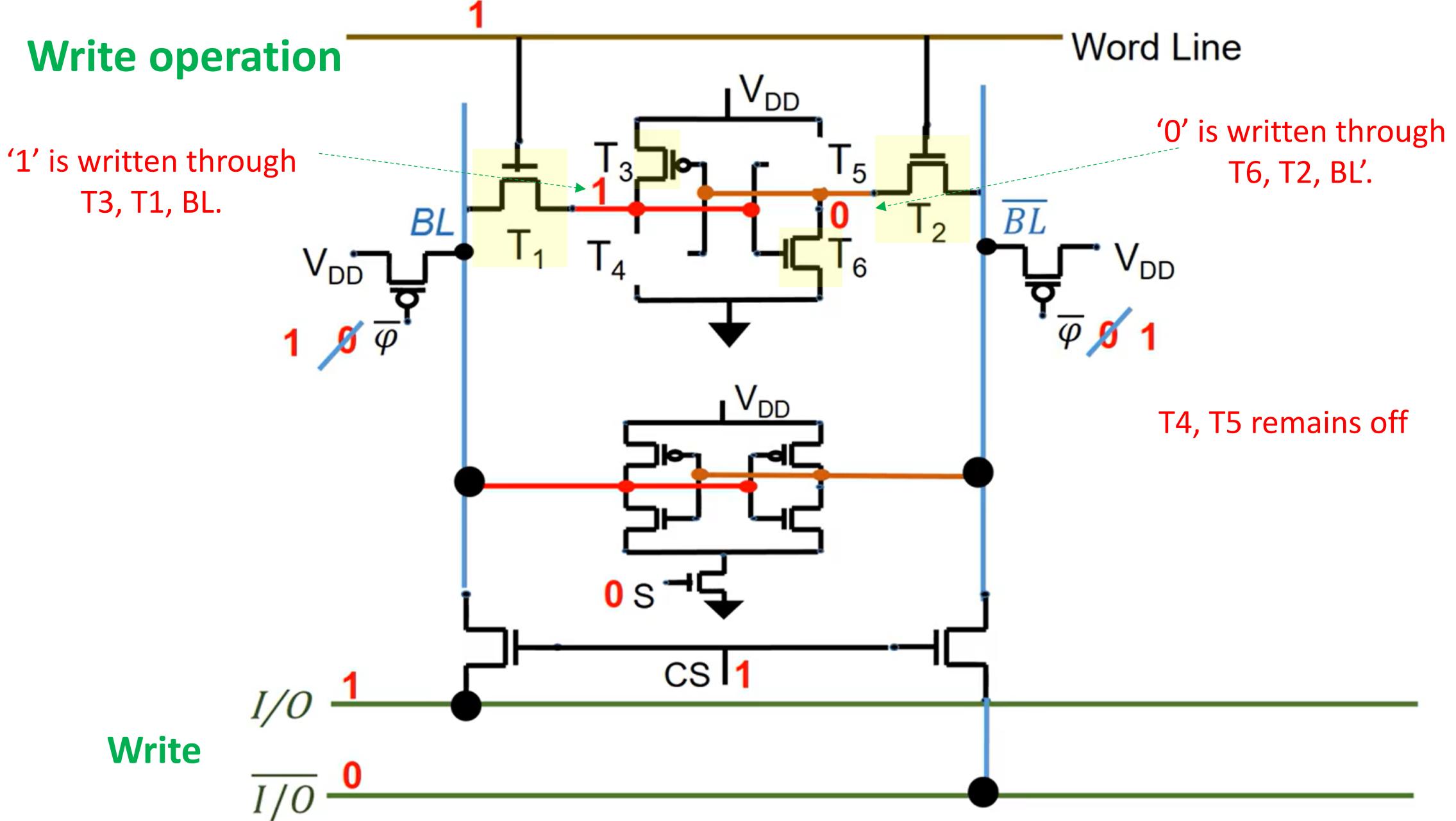
Write operation



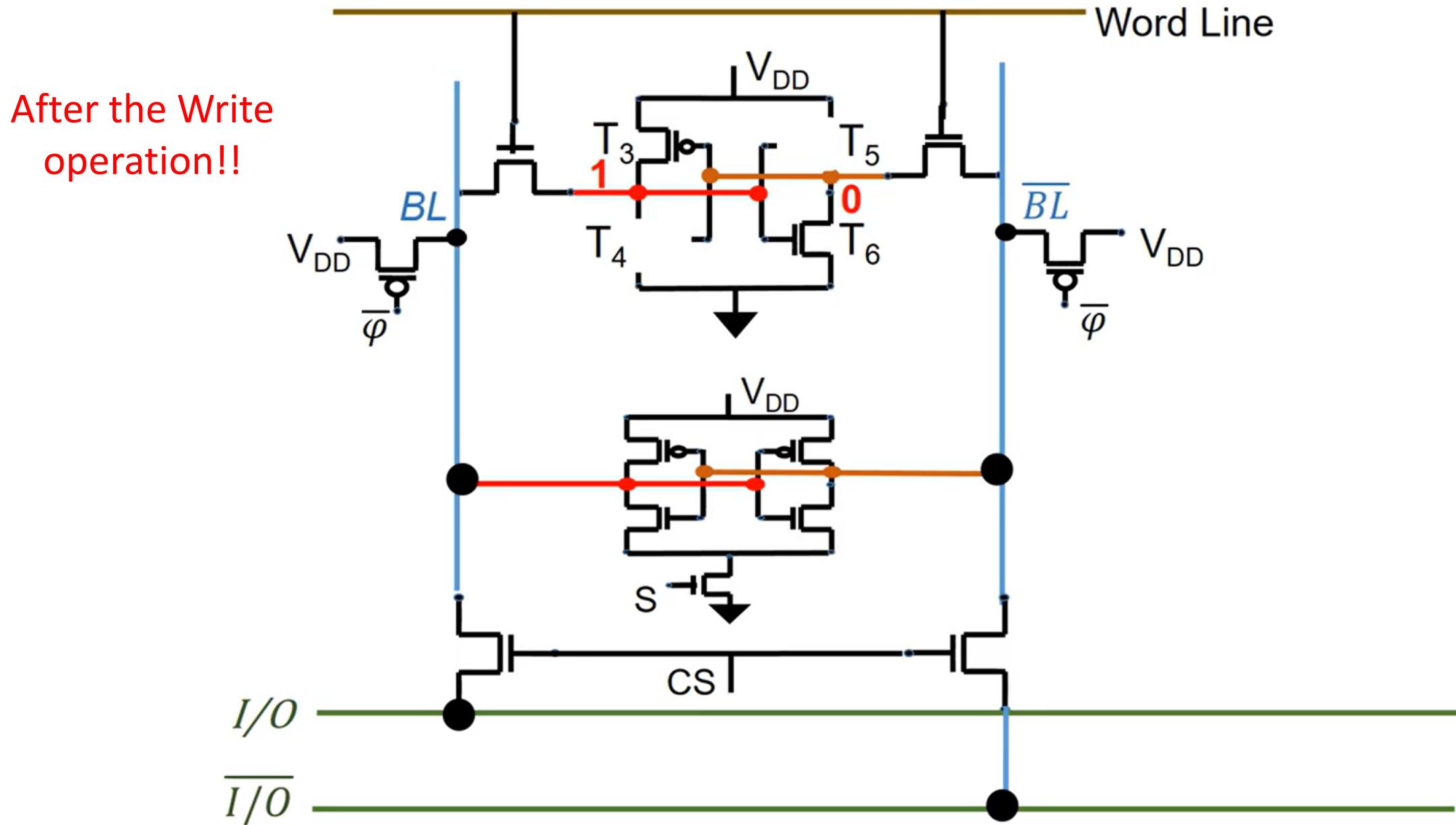
Random Access Memory (6T-SRAM)



Random Access Memory (6T-SRAM)

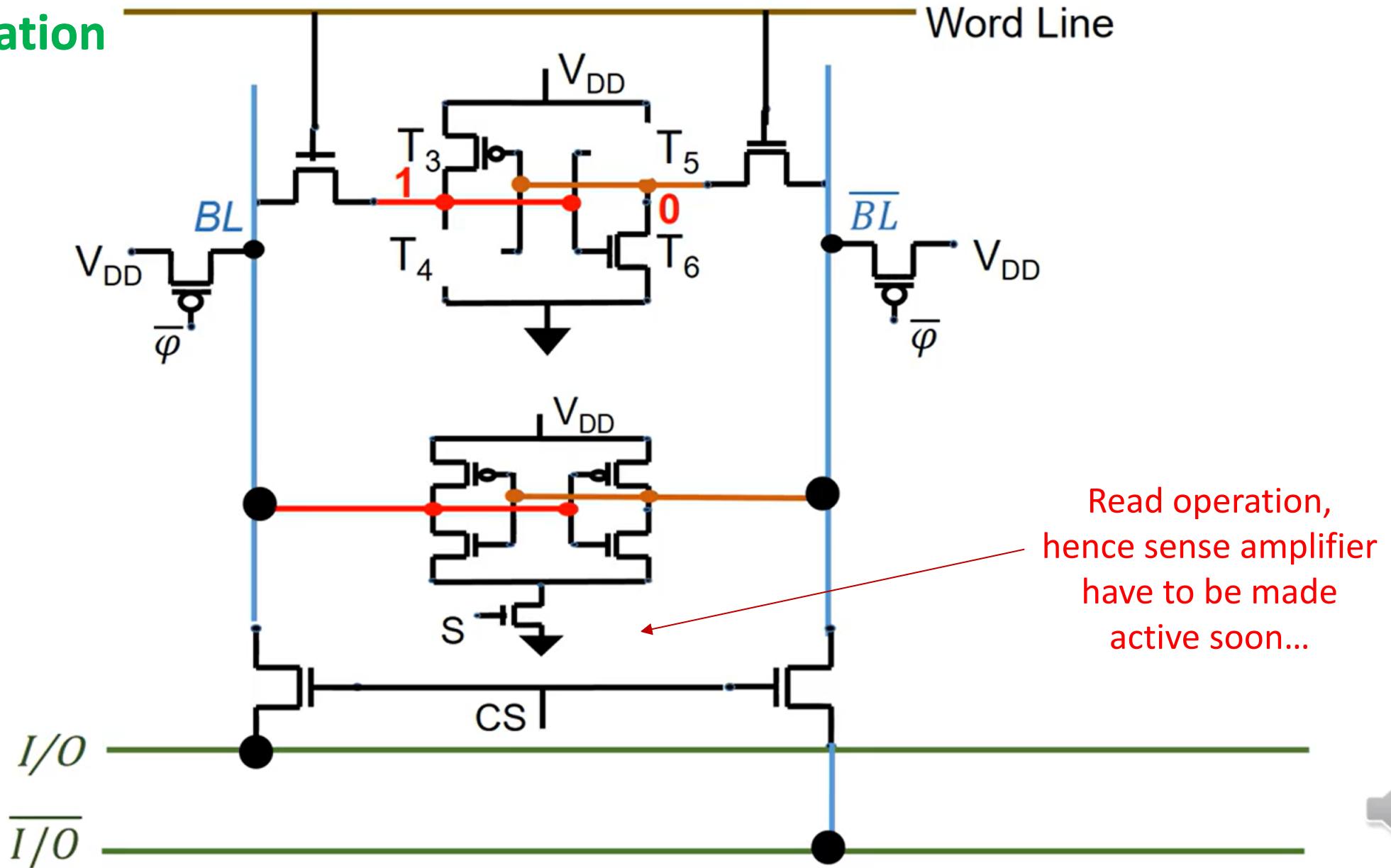


Random Access Memory (6T-SRAM)



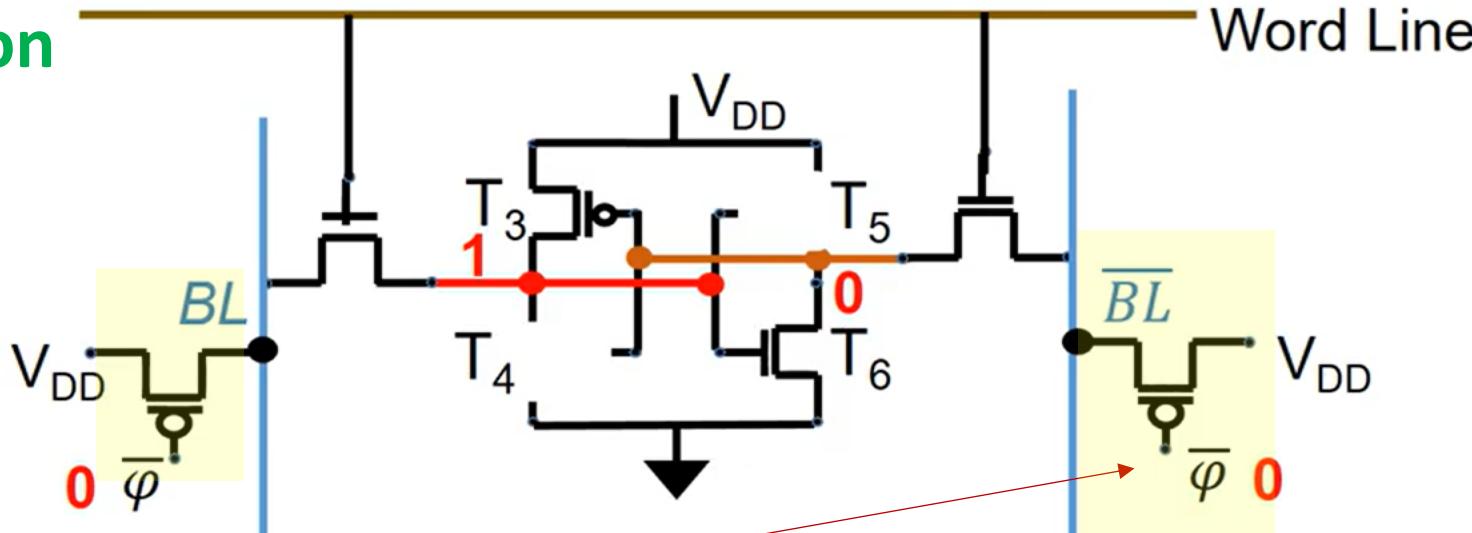
Random Access Memory (6T-SRAM)

Read operation

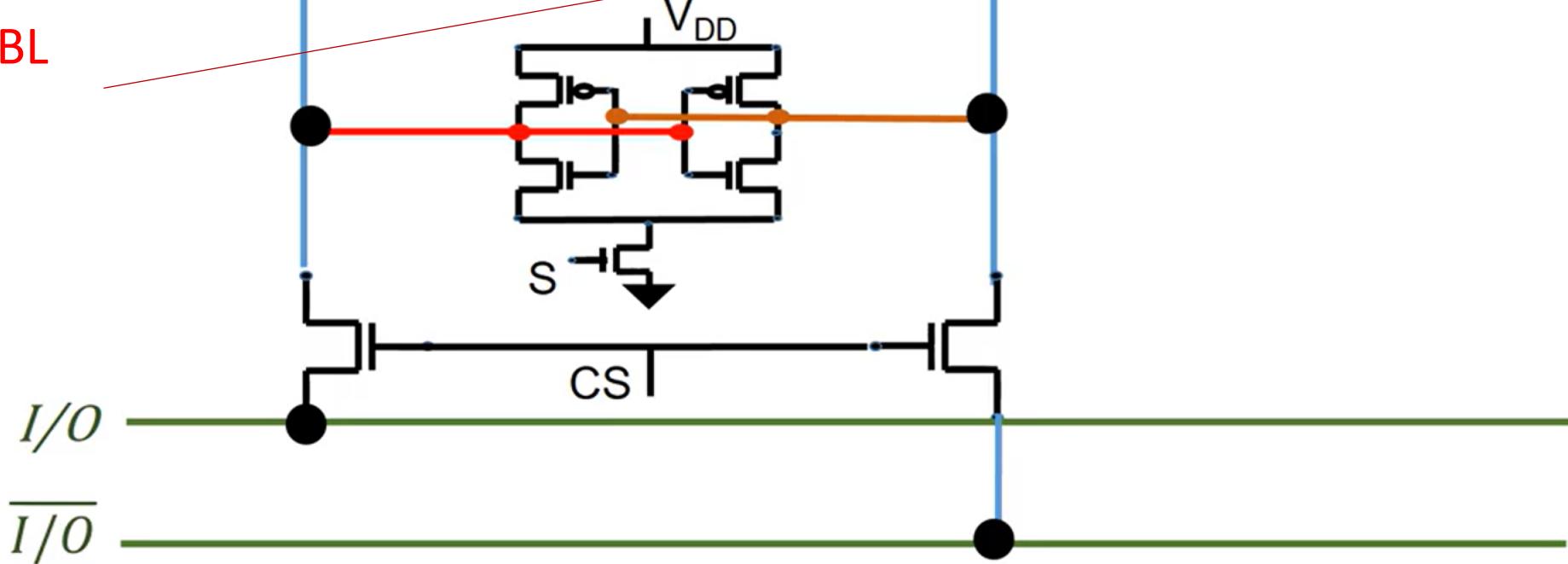


Random Access Memory (6T-SRAM)

Read operation

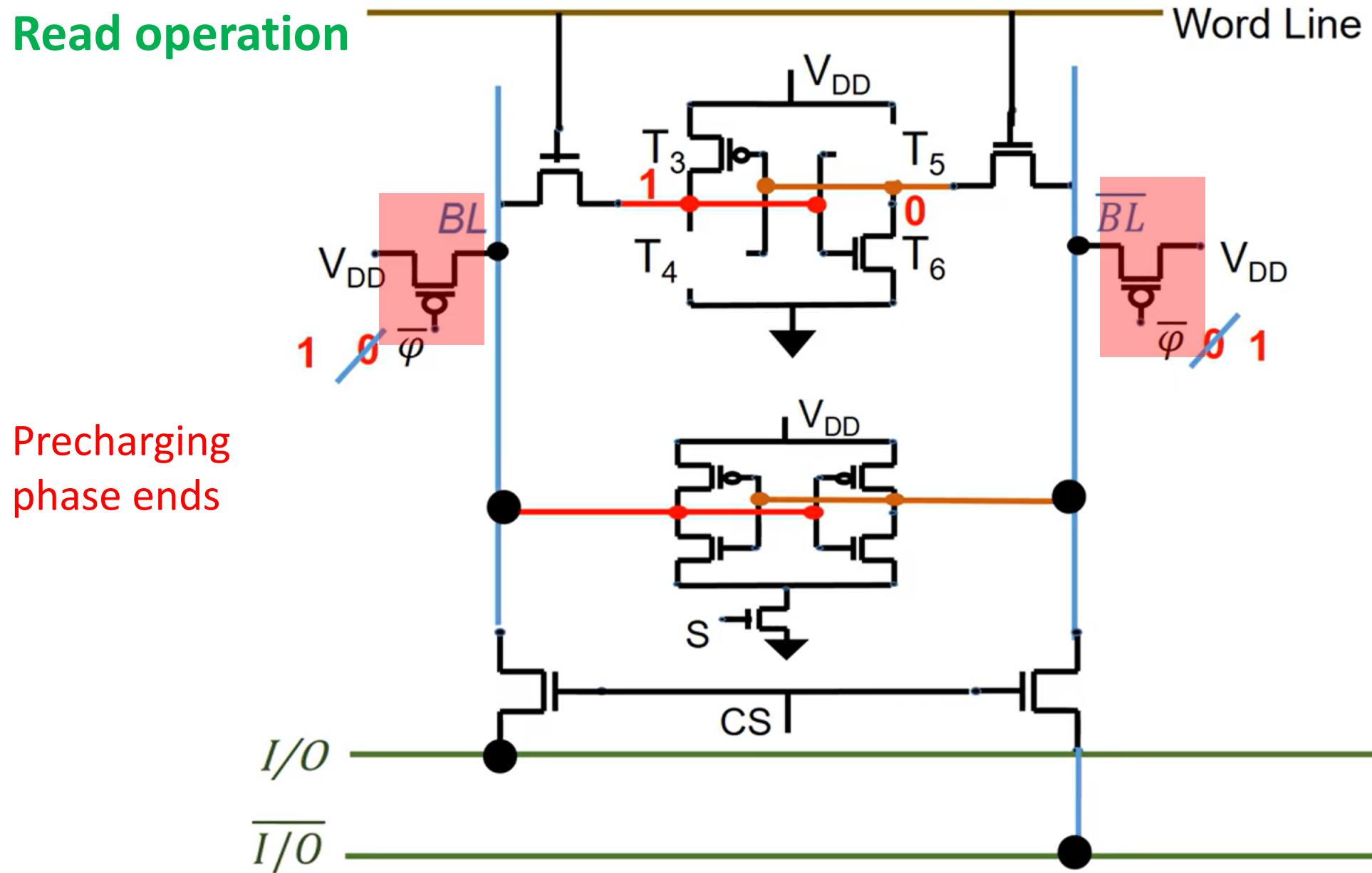


Precharging BL
and BL'

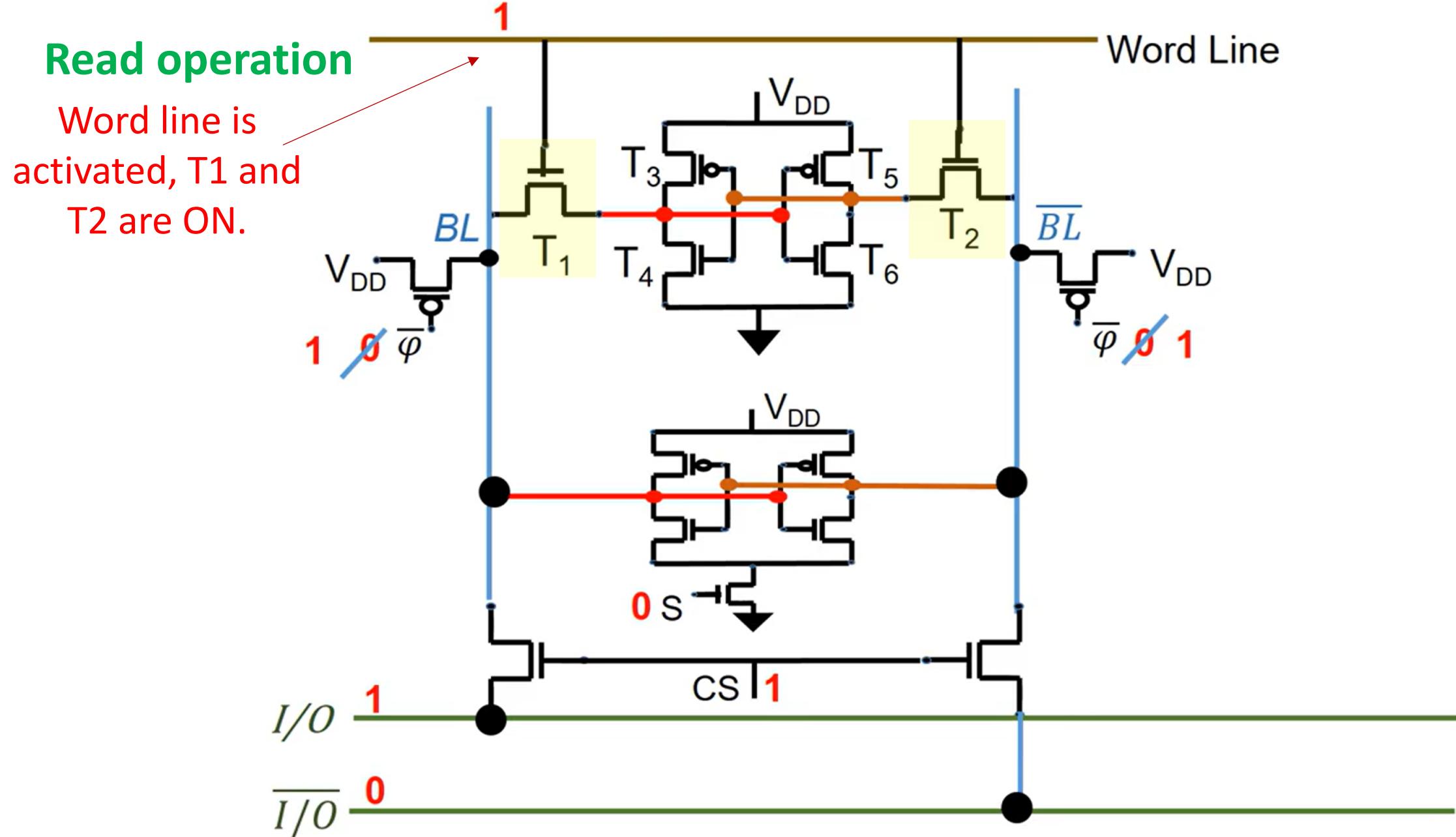


Random Access Memory (6T-SRAM)

Read operation

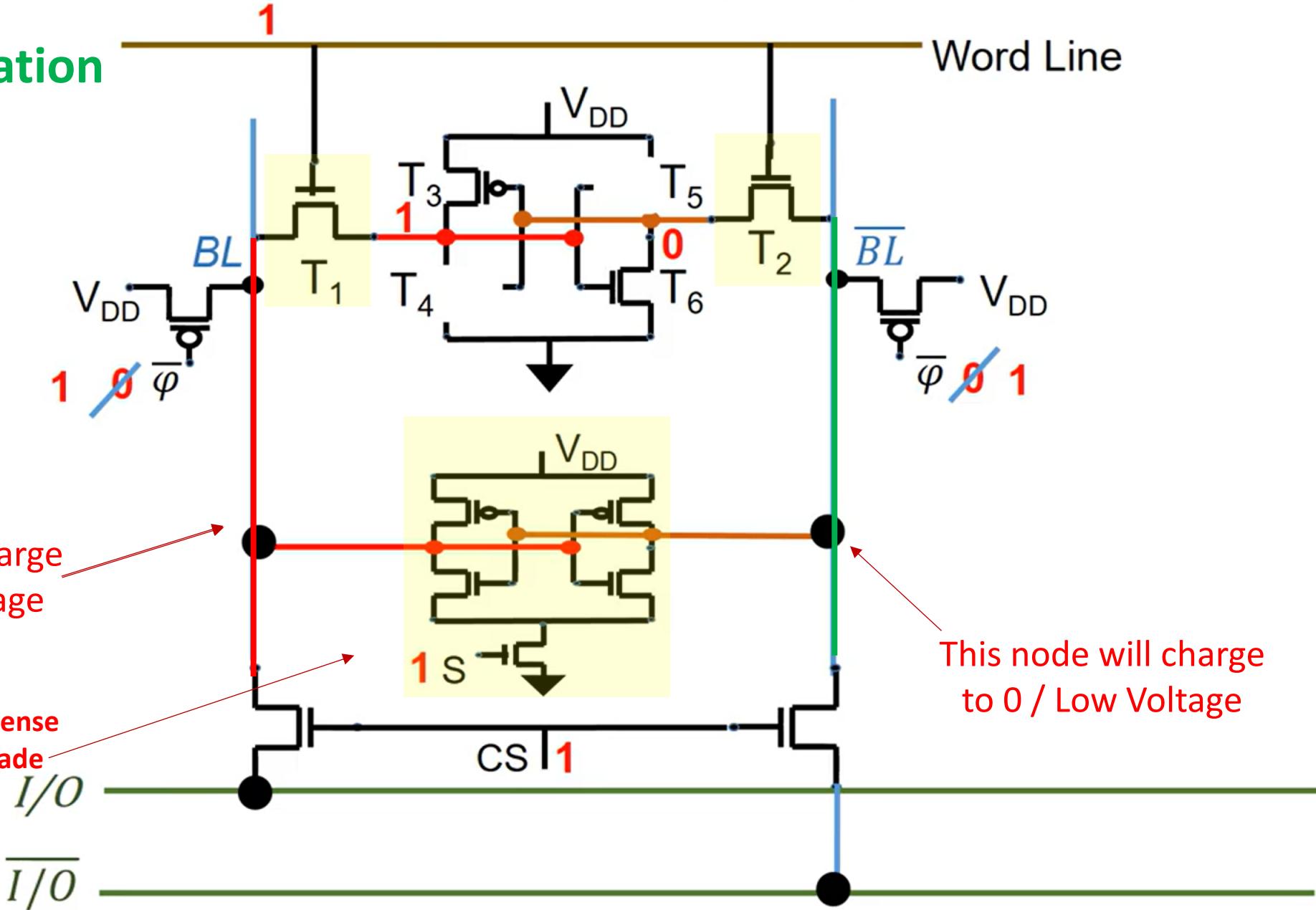


Random Access Memory (6T-SRAM)



Random Access Memory (6T-SRAM)

Read operation

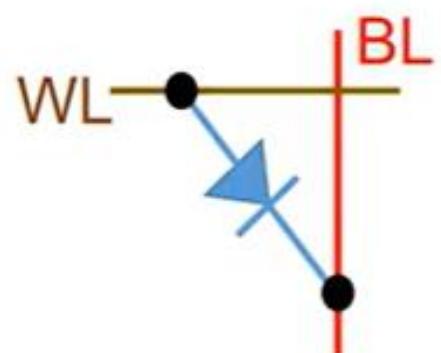
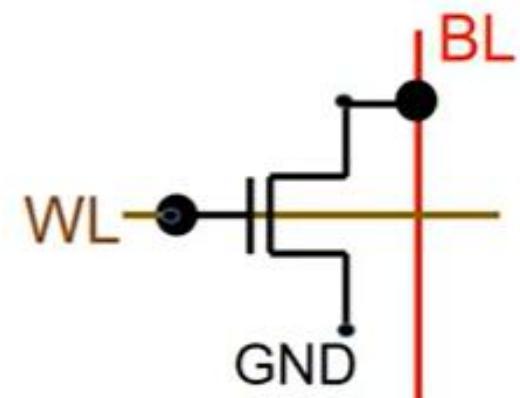




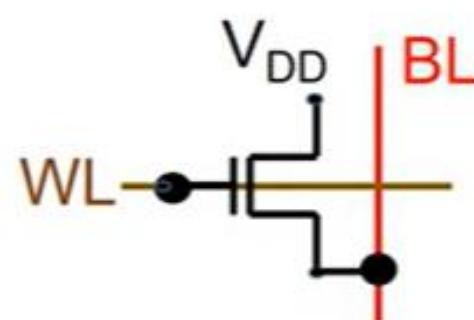
Read Only Memory (ROM)



Read-Only Memory Cells



Diode ROM



MOS ROM-1



MOS ROM-2



Three types of ROM

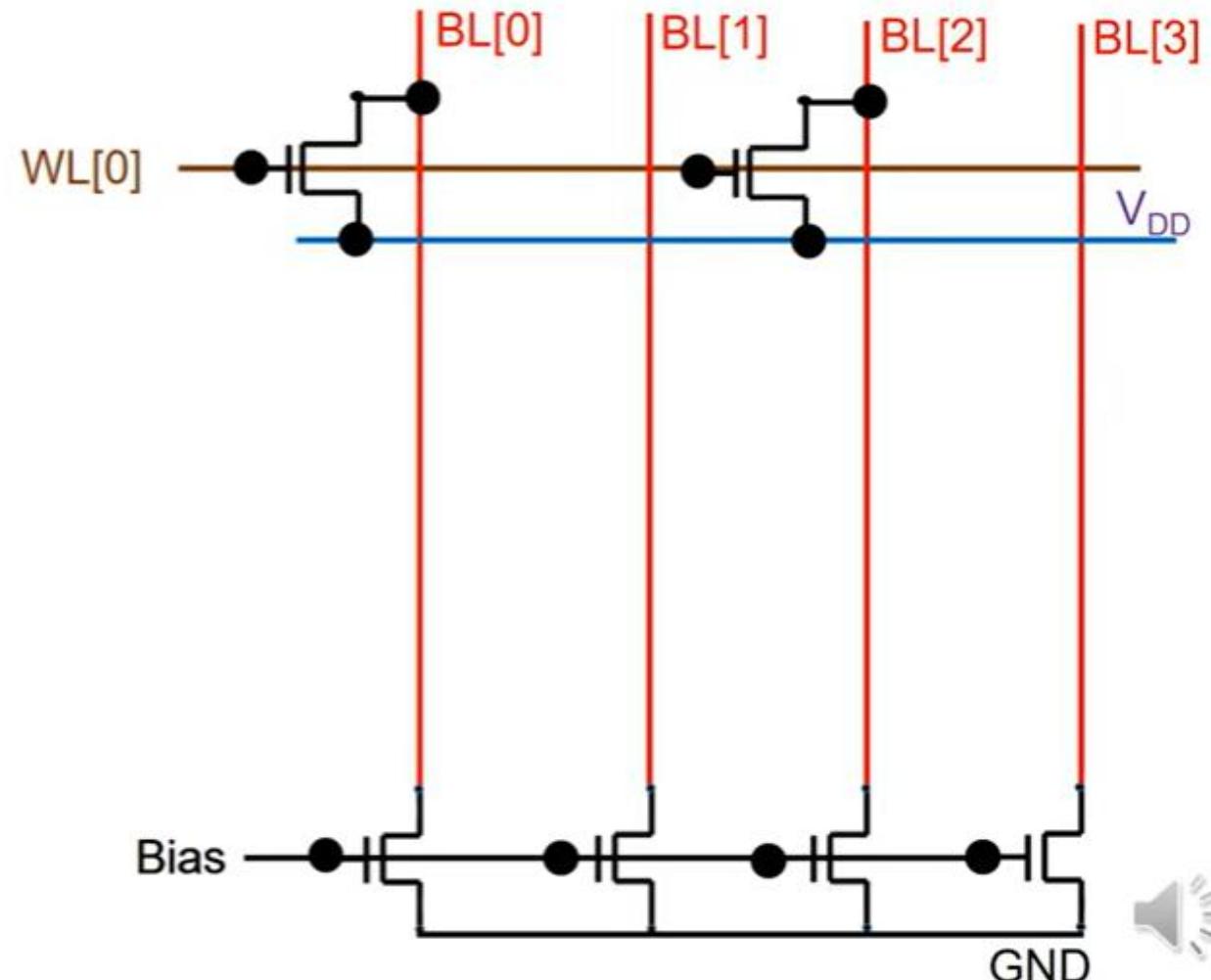
1. OR ROM
2. NOR ROM
3. NAND ROM



MOS OR ROM

WL[0] = 1 0 1 0
WL[1] = 0 1 1 0
WL[2] = 0 0 0 1
WL[3] = 0 1 1 0

Active high decoder

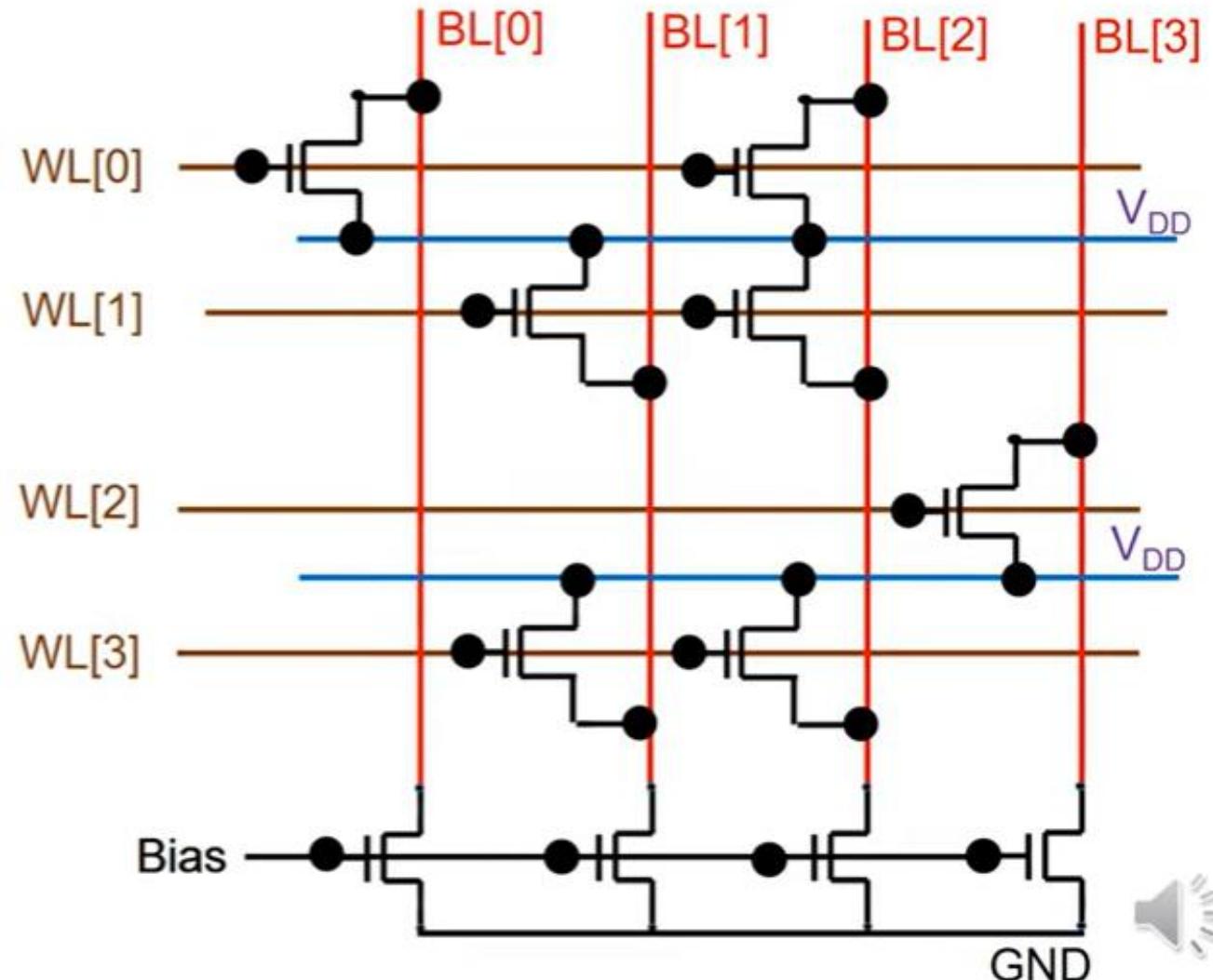




MOS OR ROM

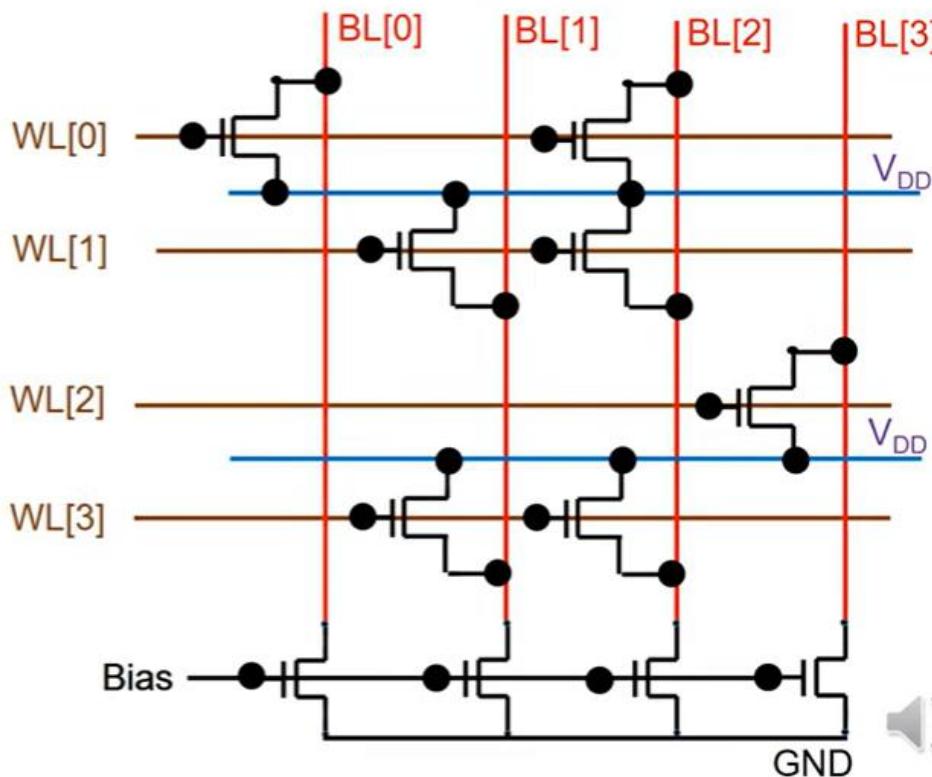
WL[0]= 1 0 1 0
WL[1]= 0 1 1 0
WL[2]= 0 0 0 1
WL[3]= 0 1 1 0

Active high decoder



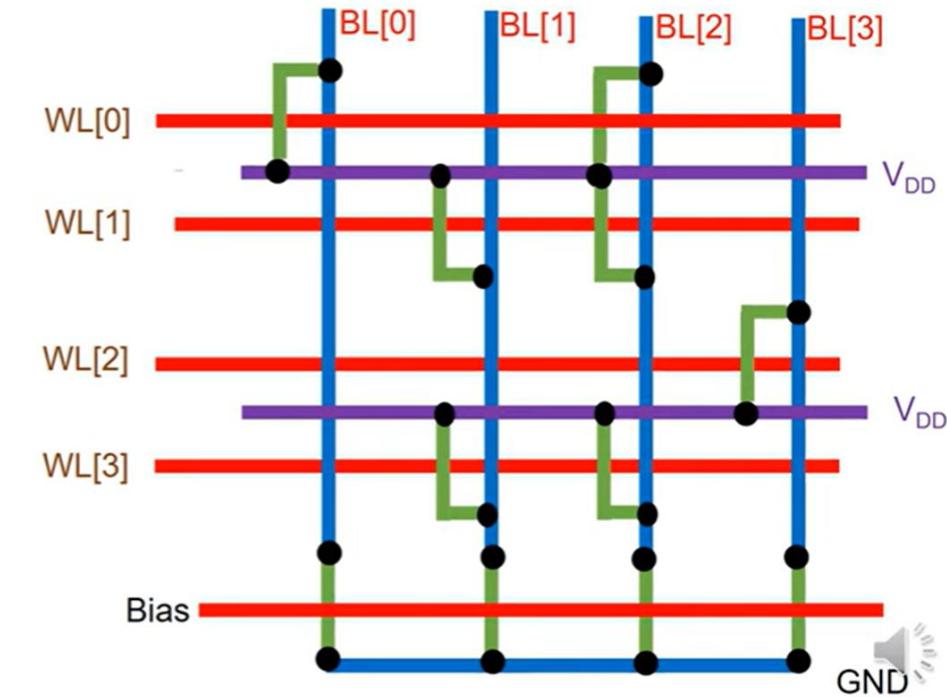


MOS OR ROM



$WL[0]=1010$
 $WL[1]=0110$
 $WL[2]=0001$
 $WL[3]=0110$

Active high decoder

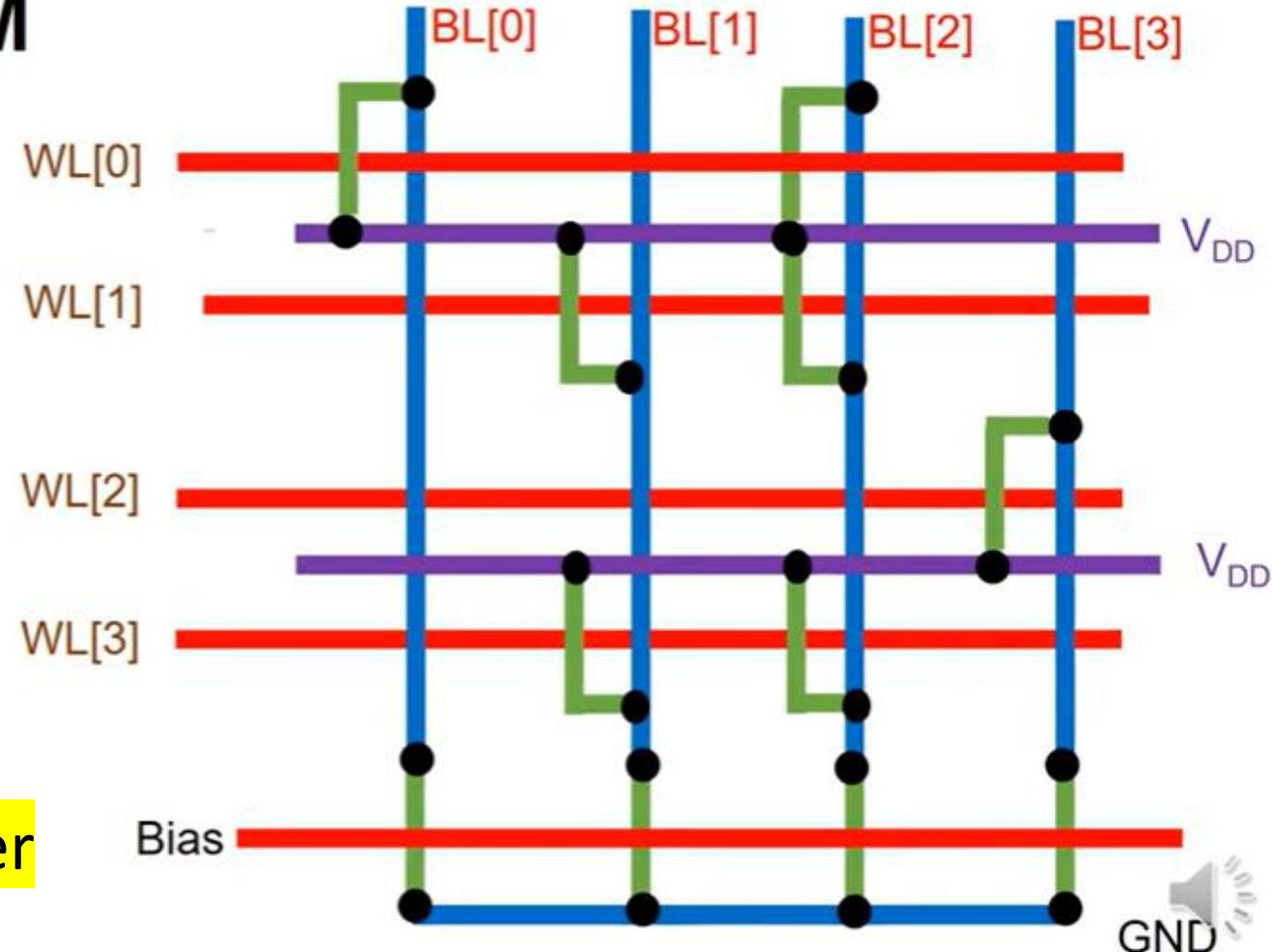




MOS OR ROM

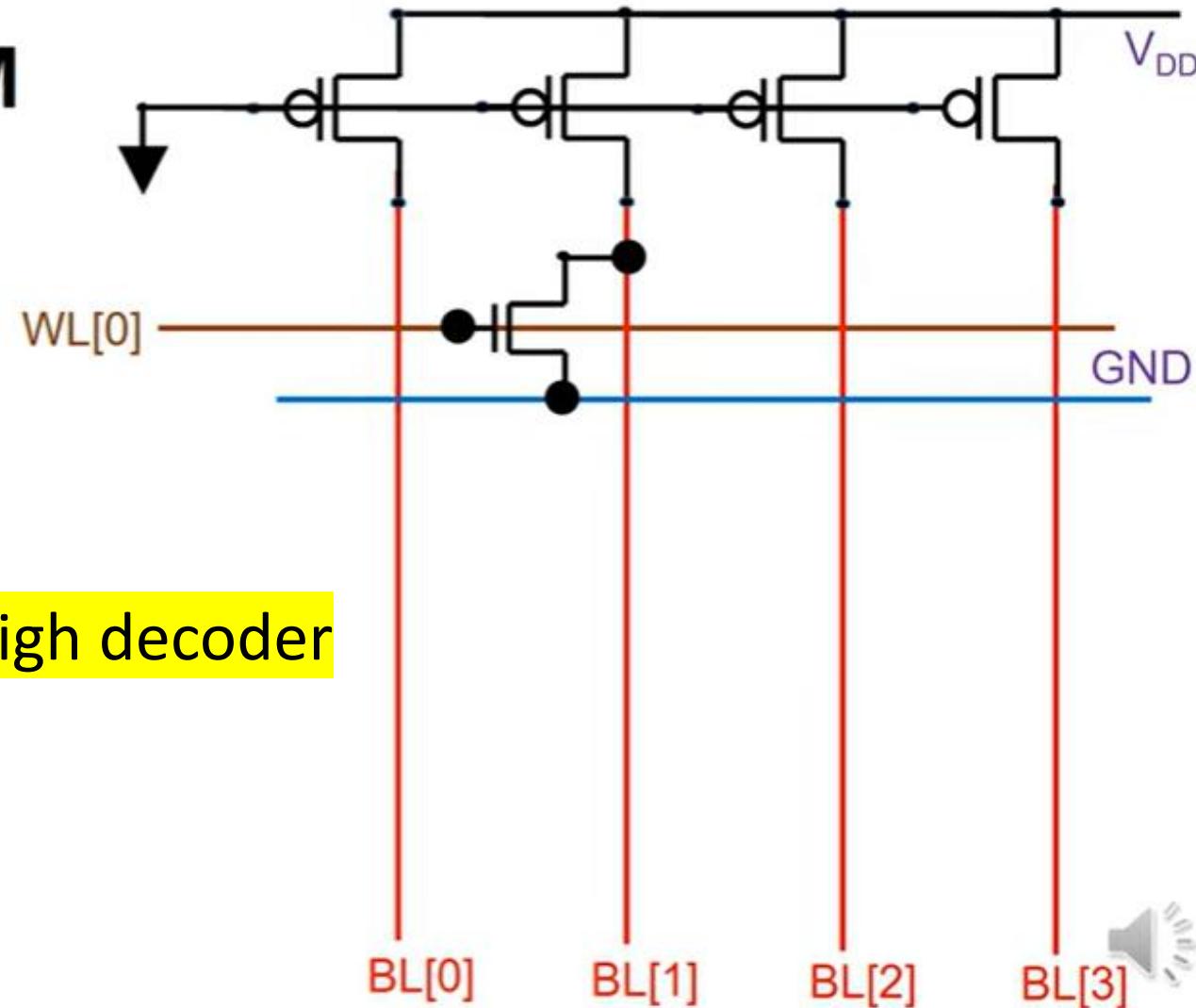
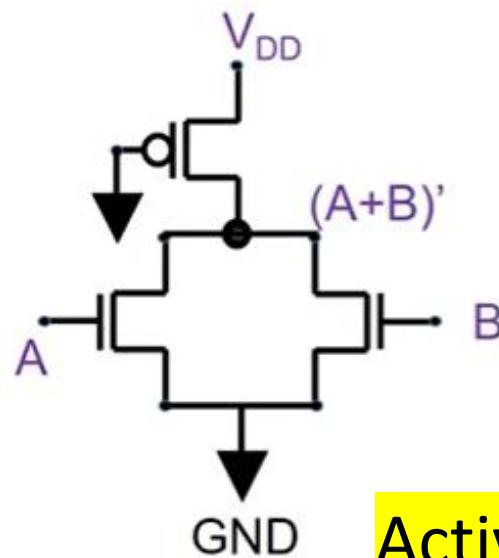
WL[0] = 1 0 1 0
WL[1] = 0 1 1 0
WL[2] = 0 0 0 1
WL[3] = 0 1 1 0

Active high decoder





MOS NOR ROM



$WL[0]=1010$

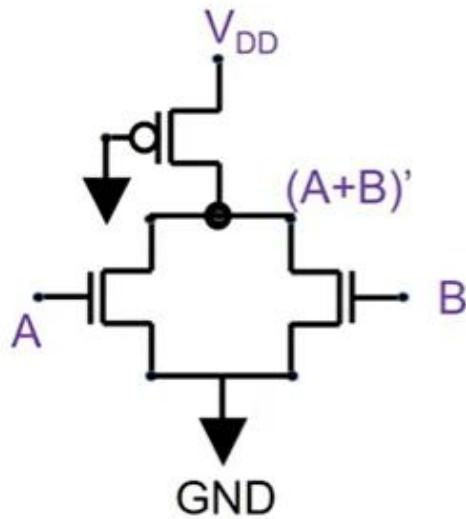
$WL[1]=0110$

$WL[2]=0001$

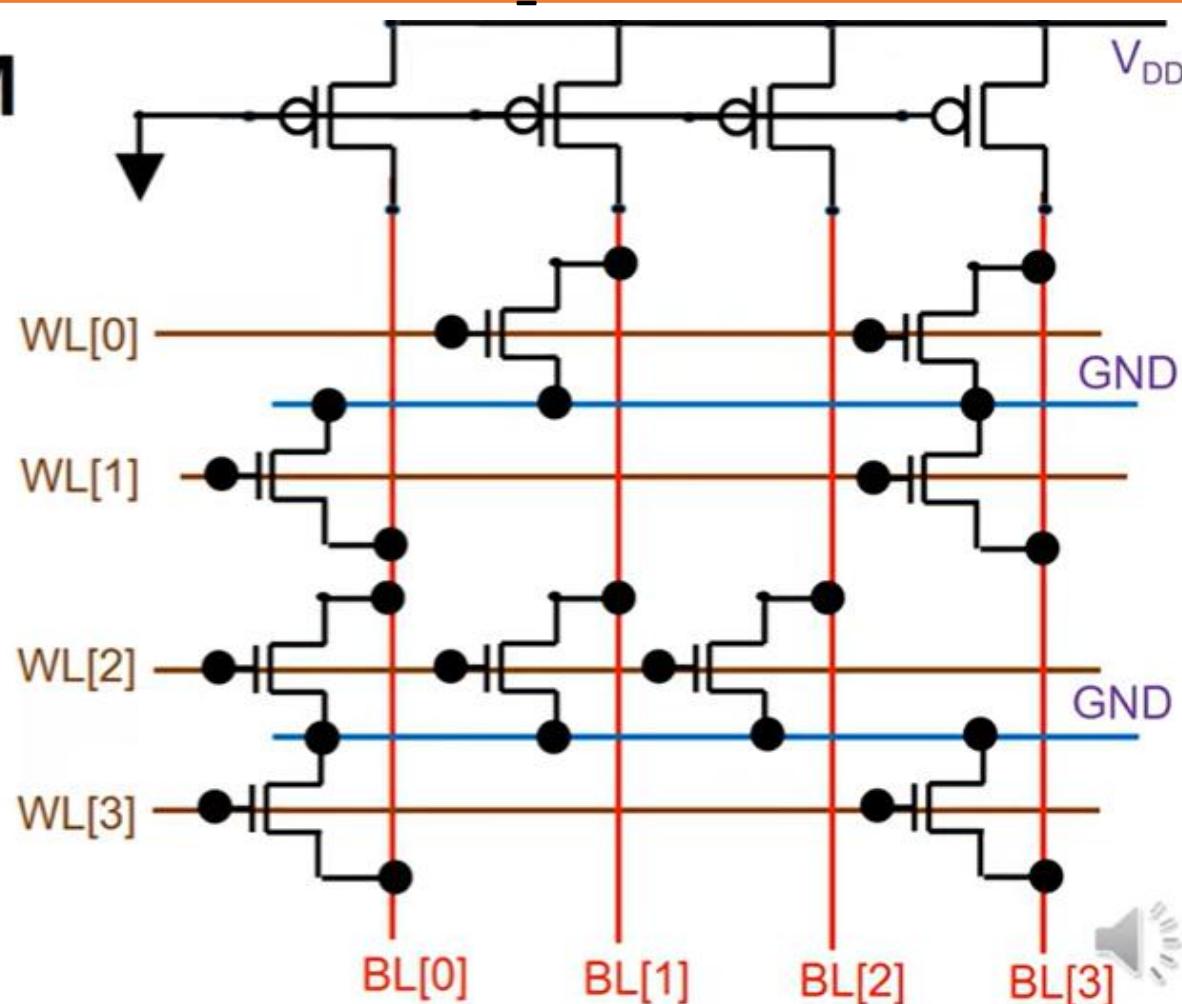
$WL[3]=0110$



MOS NOR ROM



$WL[0] = 1010$
 $WL[1] = 0110$
 $WL[2] = 0001$
 $WL[3] = 0110$

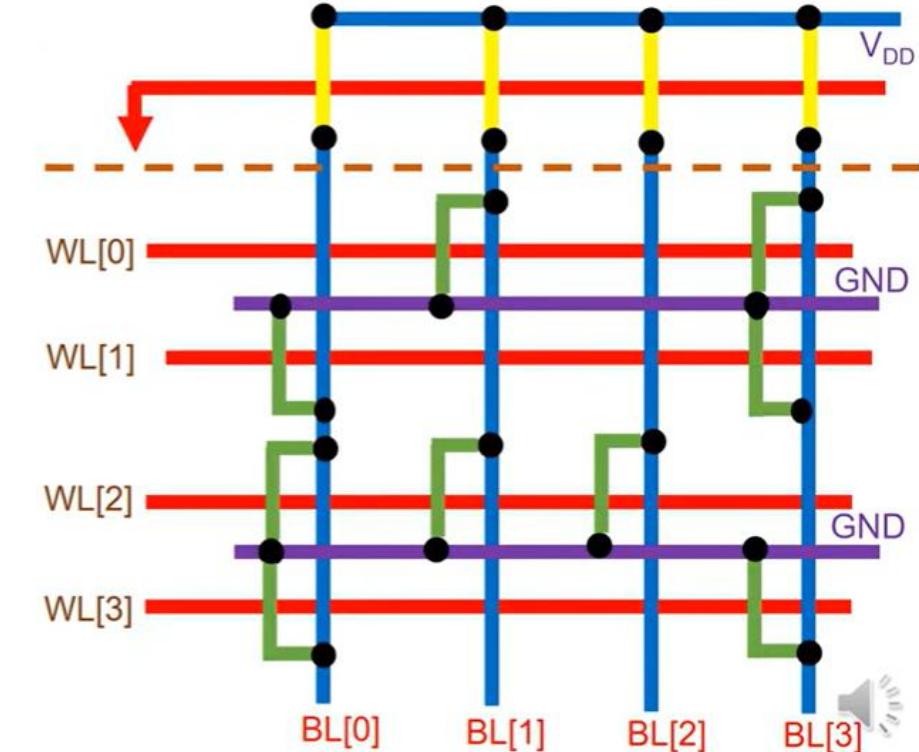
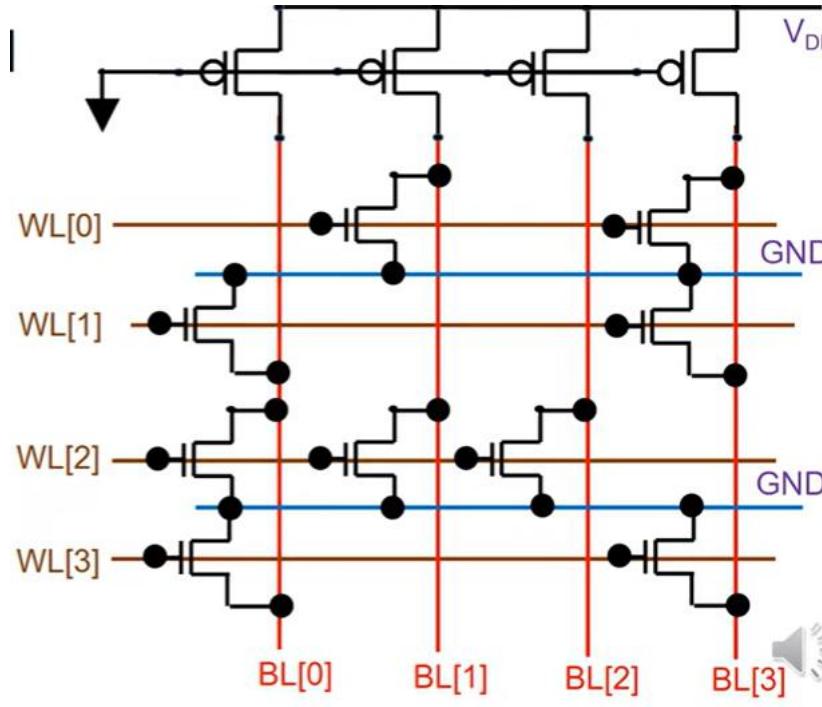


Active high decoder



MOS NOR ROM

WL[0]= 1 0 1 0
WL[1]= 0 1 1 0
WL[2]= 0 0 0 1
WL[3]= 0 1 1 0



Active high decoder



MOS NOR ROM

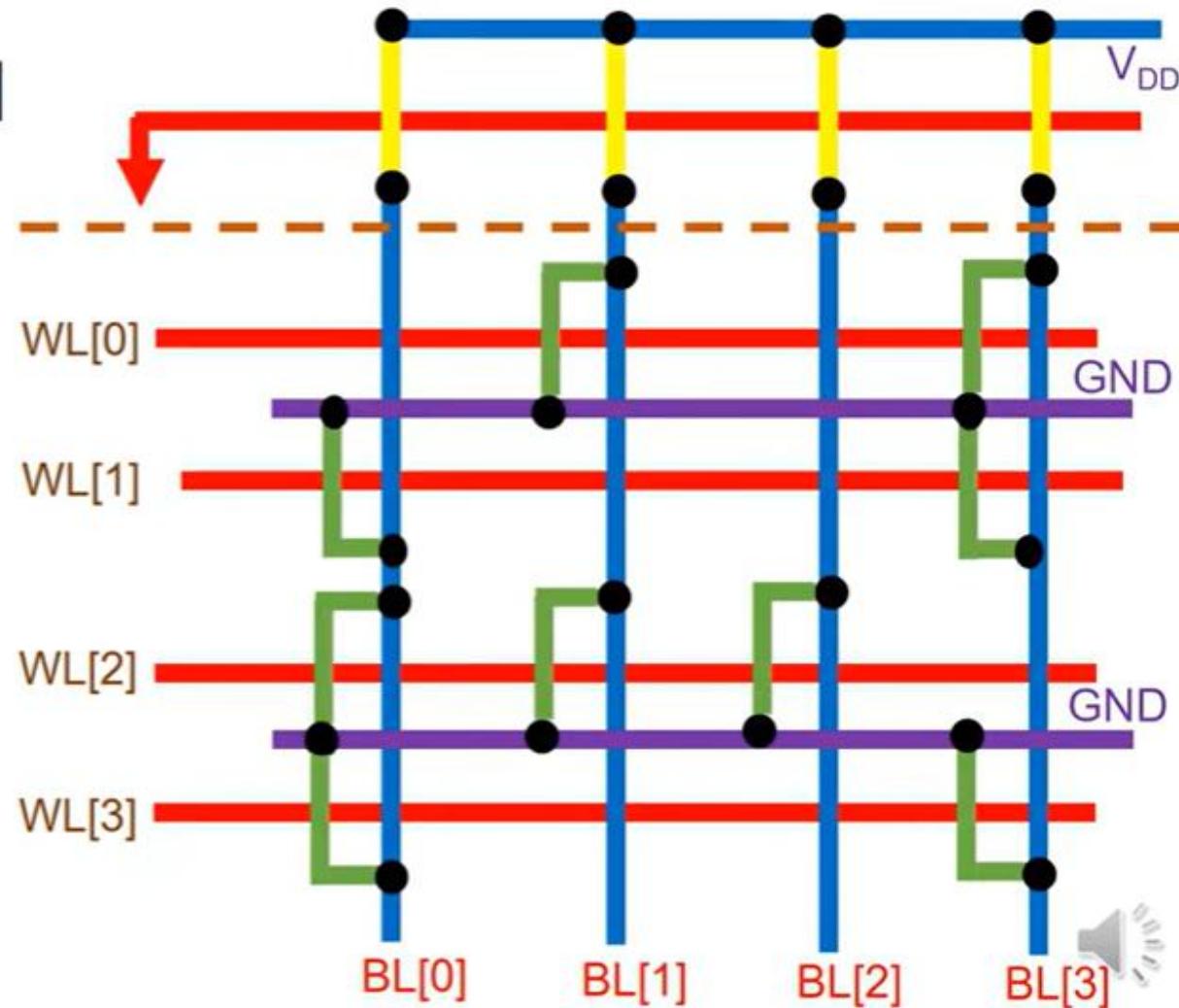
WL[0]= 1 0 1 0

WL[1]= 0 1 1 0

WL[2]= 0 0 0 1

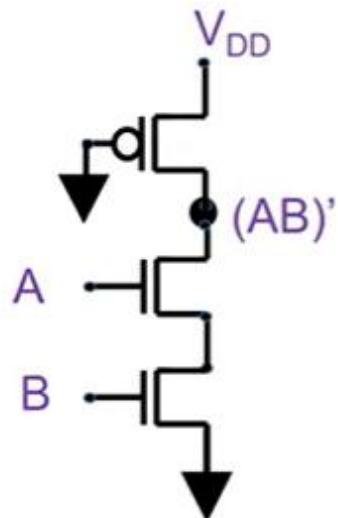
WL[3]= 0 1 1 0

Active high decoder

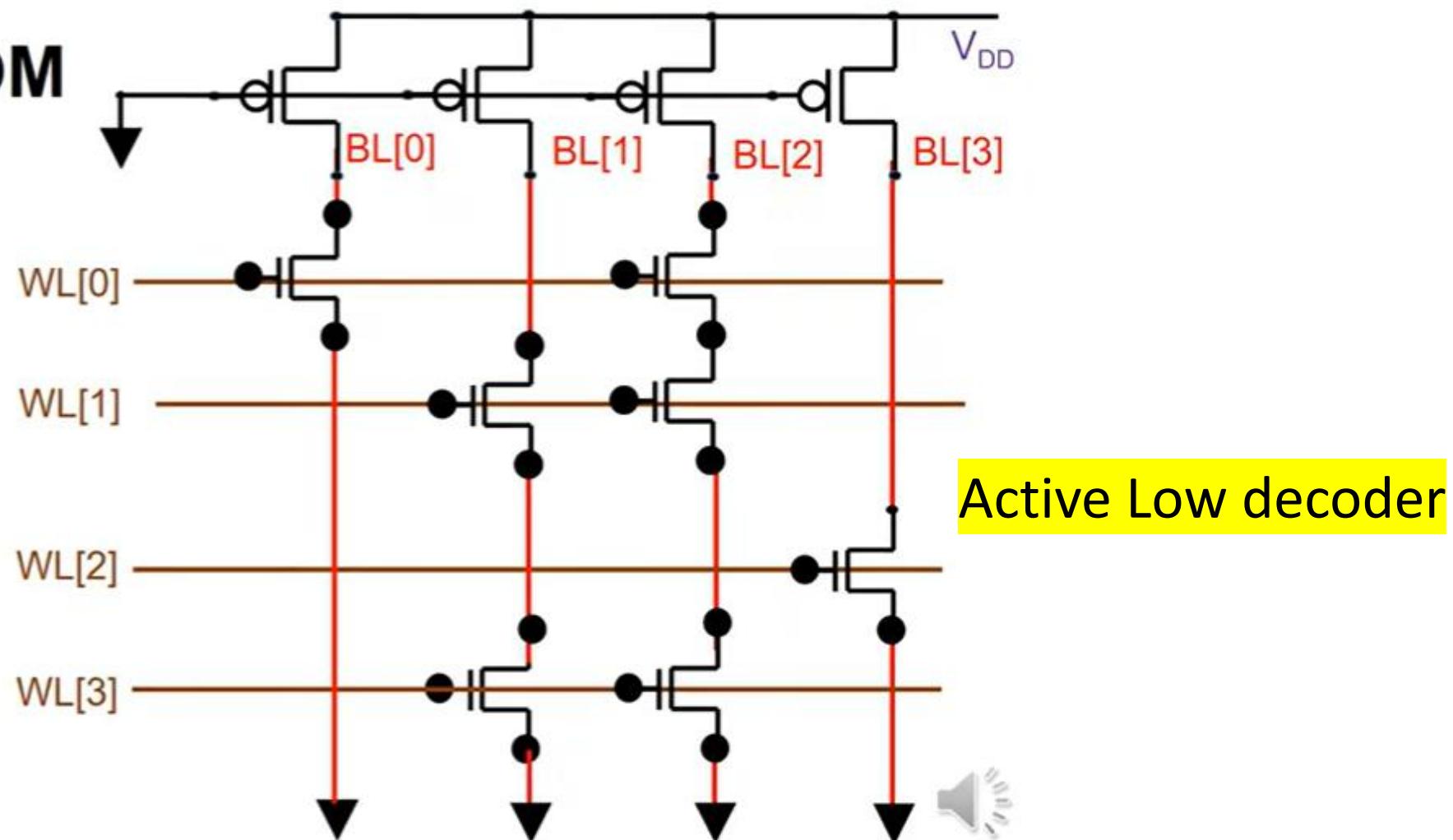




MOS NAND ROM

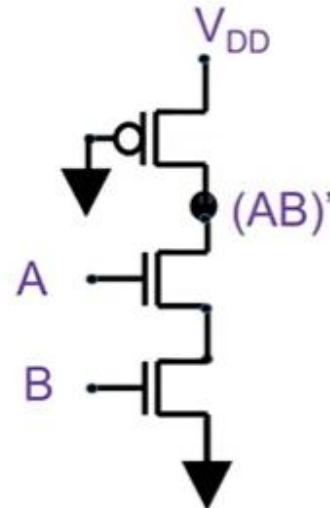


WL[0]= 1 0 1 0
WL[1]= 0 1 1 0
WL[2]= 0 0 0 1
WL[3]= 0 1 1 0





MOS NAND ROM

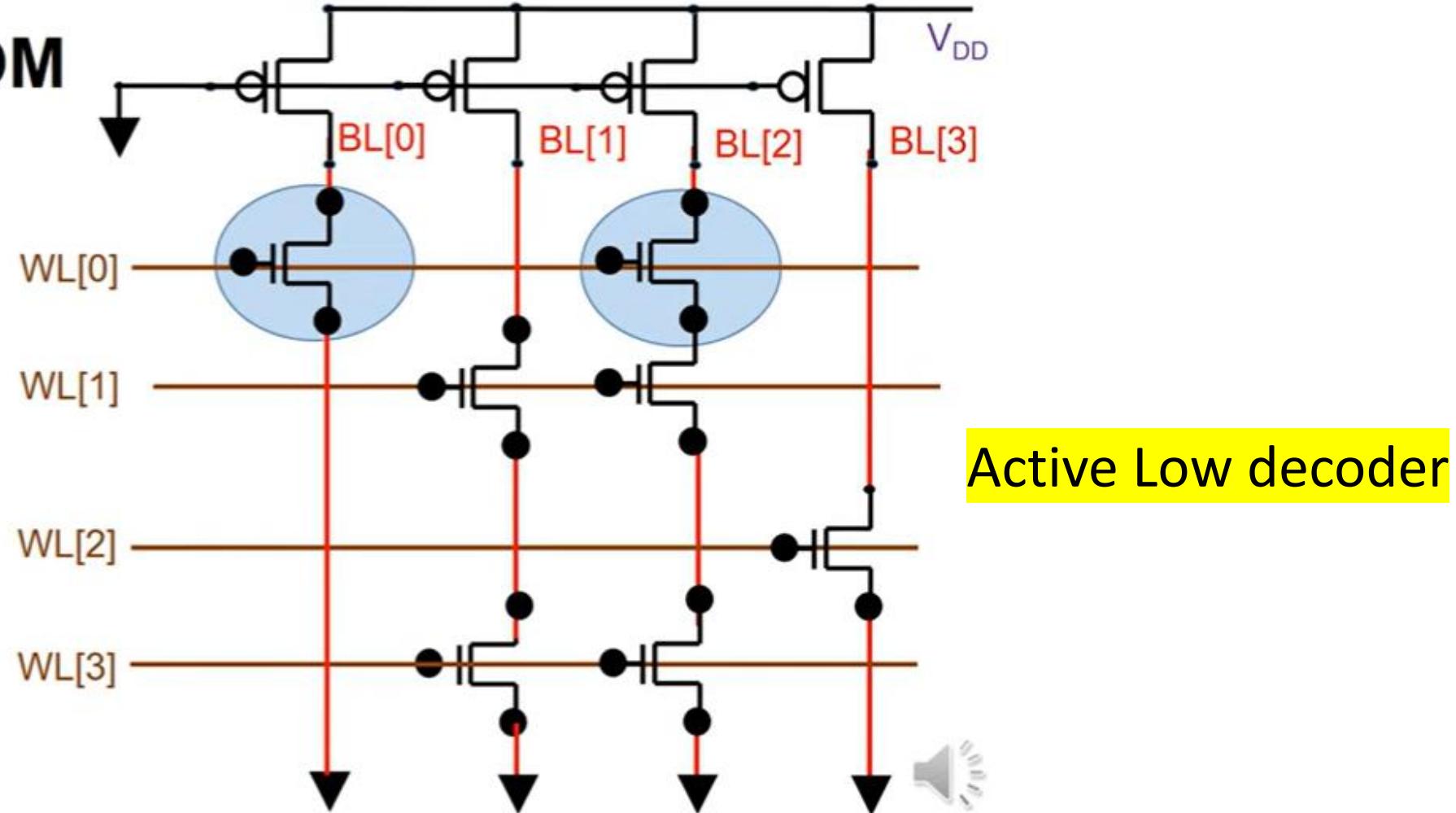


WL[0]= 1 0 1 0

WL[1]= 0 1 1 0

WL[2]= 0 0 0 1

WL[3]= 0 1 1 0

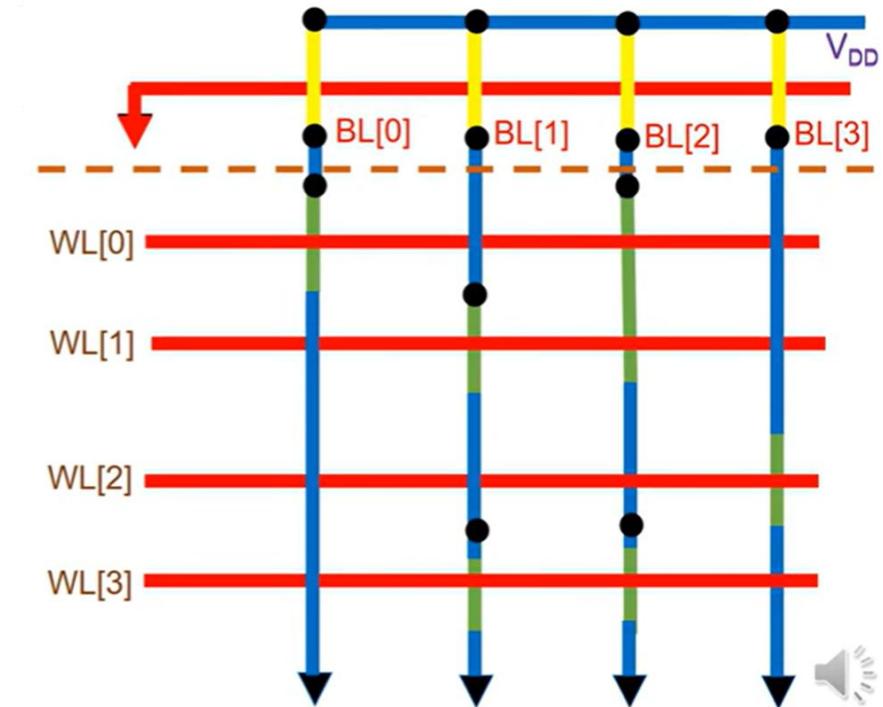
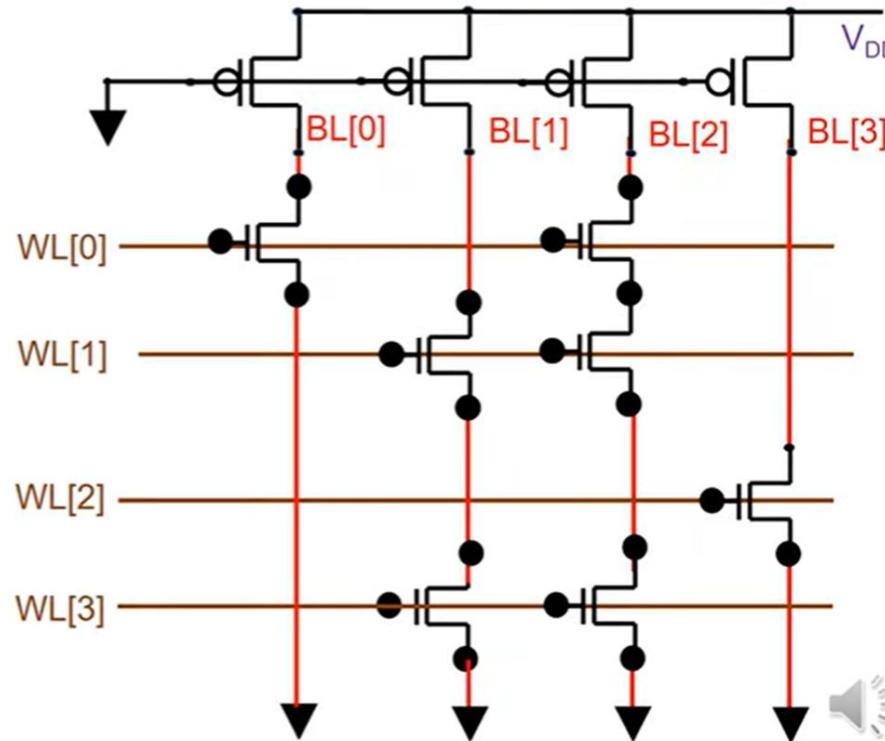


Active Low decoder



MOS NAND ROM

WL[0]= 1 0 1 0
WL[1]= 0 1 1 0
WL[2]= 0 0 0 1
WL[3]= 0 1 1 0



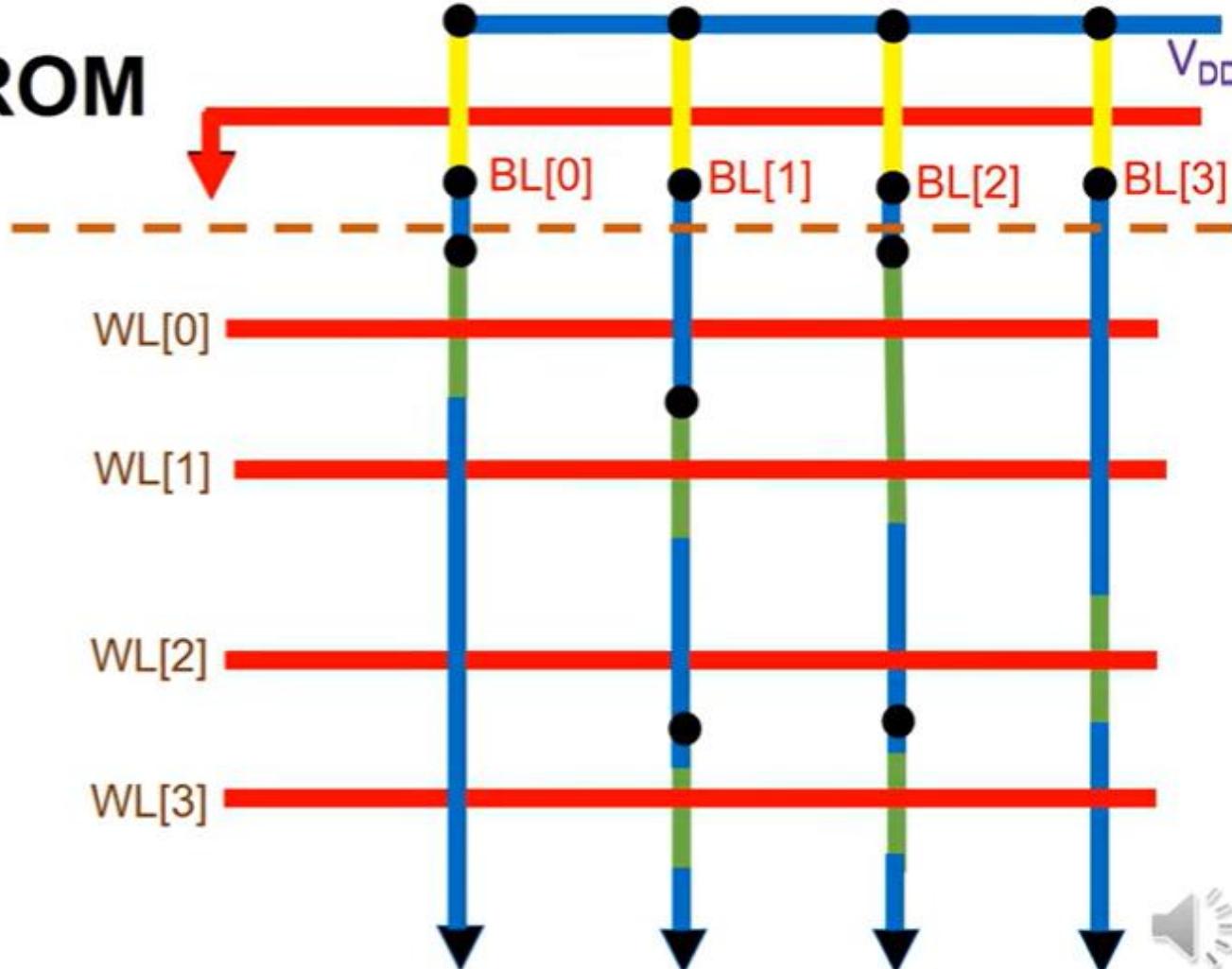
Active Low decoder



MOS NAND ROM

WL[0]= 1 0 1 0
WL[1]= 0 1 1 0
WL[2]= 0 0 0 1
WL[3]= 0 1 1 0

Active Low decoder





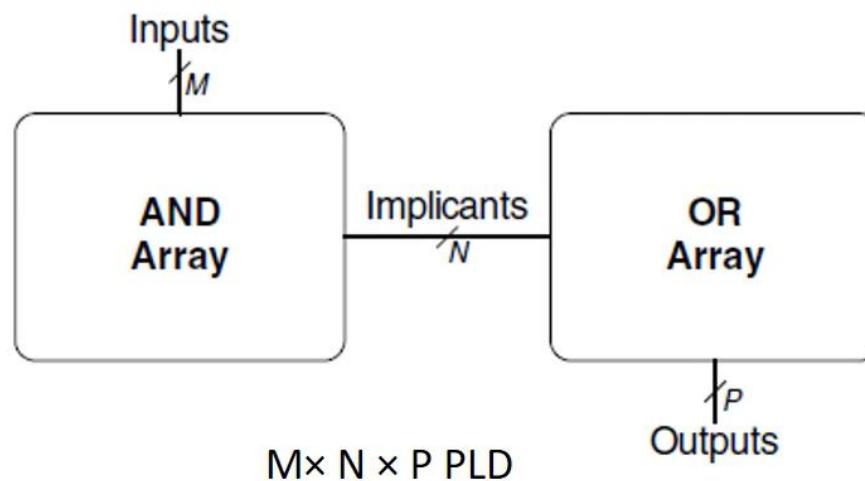
=

Q.1: Implement following 4-bit words using NOR and NAND ROM structure. W[0]: 0101; W[1]: 0011; W[2]: 1001; W[3]: 0110.



Programmable Logic Device (PLD)

Definition: A *Programmable Logic Device* (PLD) is a chip that is manufactured with a programmable configuration, enabling it to serve in many arbitrary applications.



Programmable Read Only Memory (PROMs) implement two-level **combinational logic** in sum-of-products (SOP) form. To implement the sequential logic we need additional memory element i.e. Flip-Flops.



Programmable Logic Devices(PLD)

Types of the AND–OR structured programmable logic devices are

1. Programmable read-only memory (PROM)
2. Programmable logic array (PLA)
3. Programmable array logic (PAL)

PLD	AND Plane	OR Plane
PROM	Fixed	Programmable
PLA	Programmable	Programmable
PAL	Programmable	Fixed



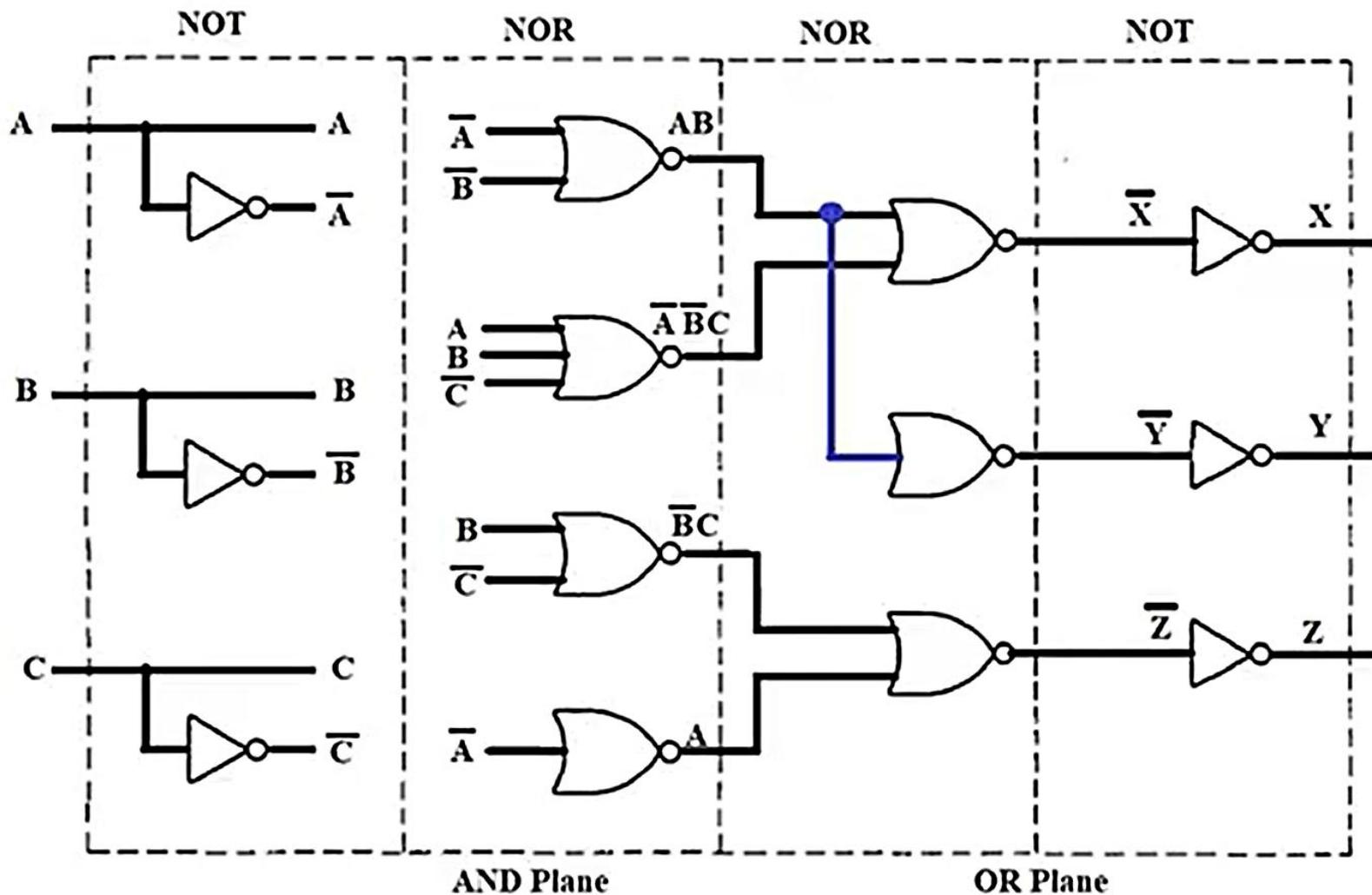
Programmable Logic Array (PLA)

$$X = AB + \bar{A}\bar{B}C$$

$$Y = AB$$

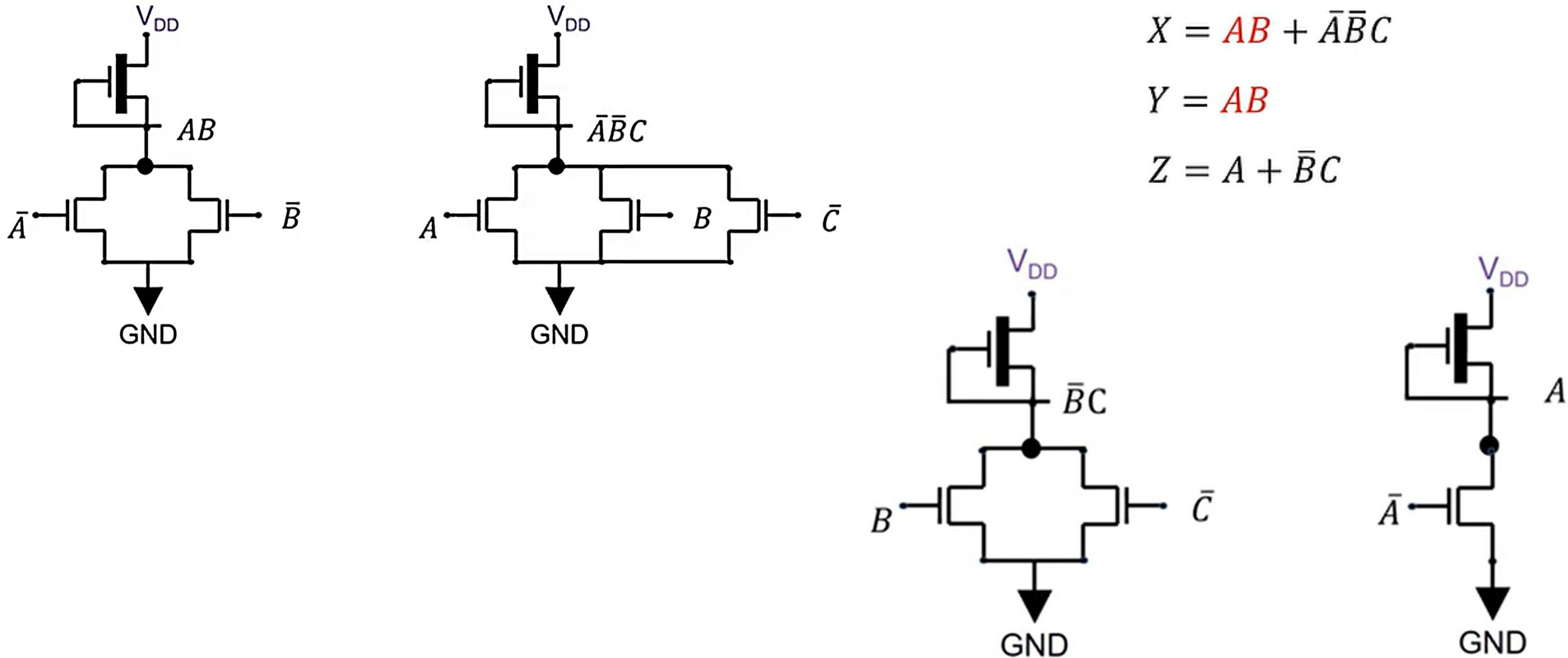
$$Z = A + \bar{B}C$$

1. nMOS PLA
2. CMOS PLA





1. nMOS PLA



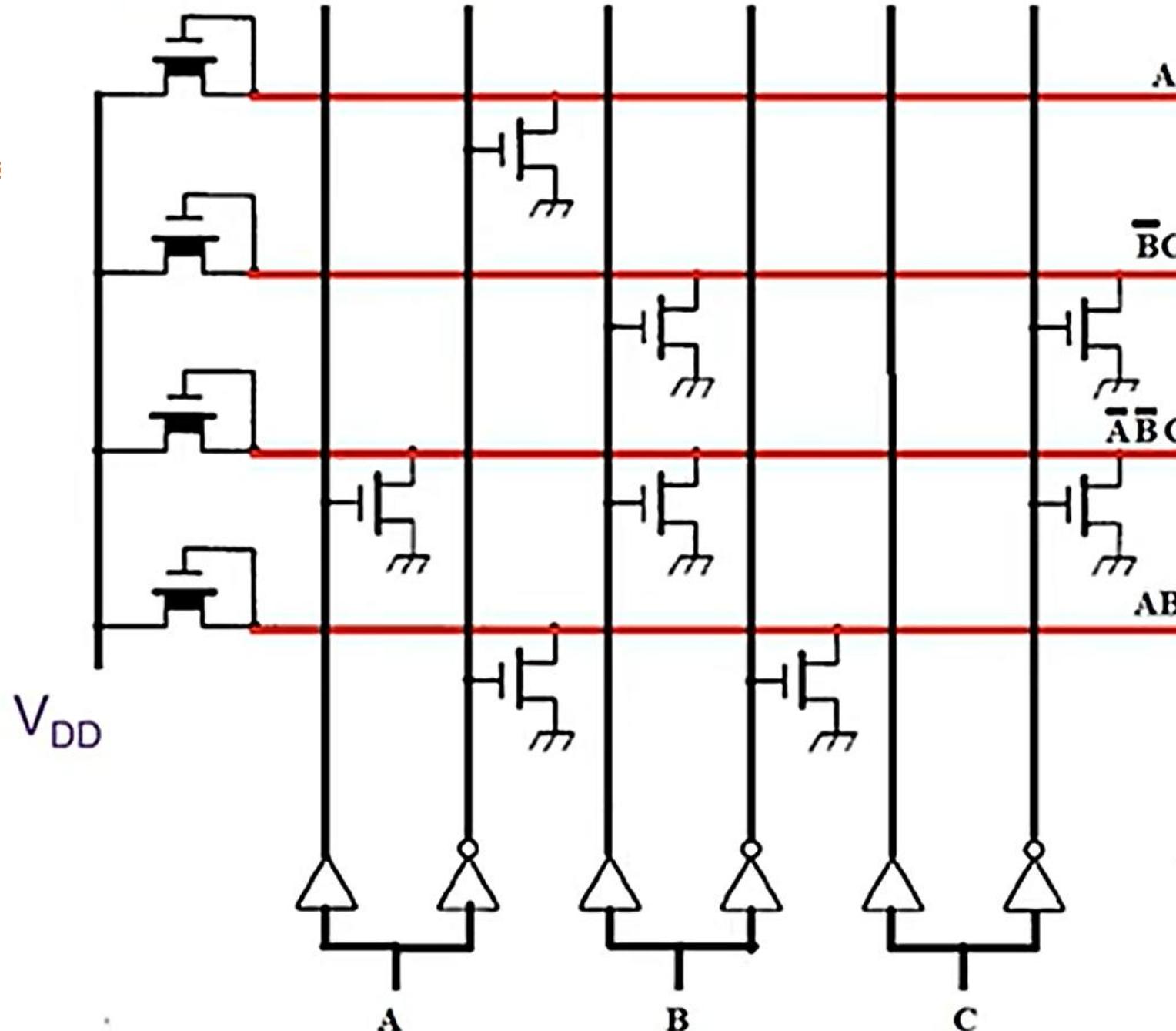


Stage 1: AND array

$$X = AB + \bar{A}\bar{B}C$$

$$Y = AB$$

$$Z = A + \bar{B}C$$



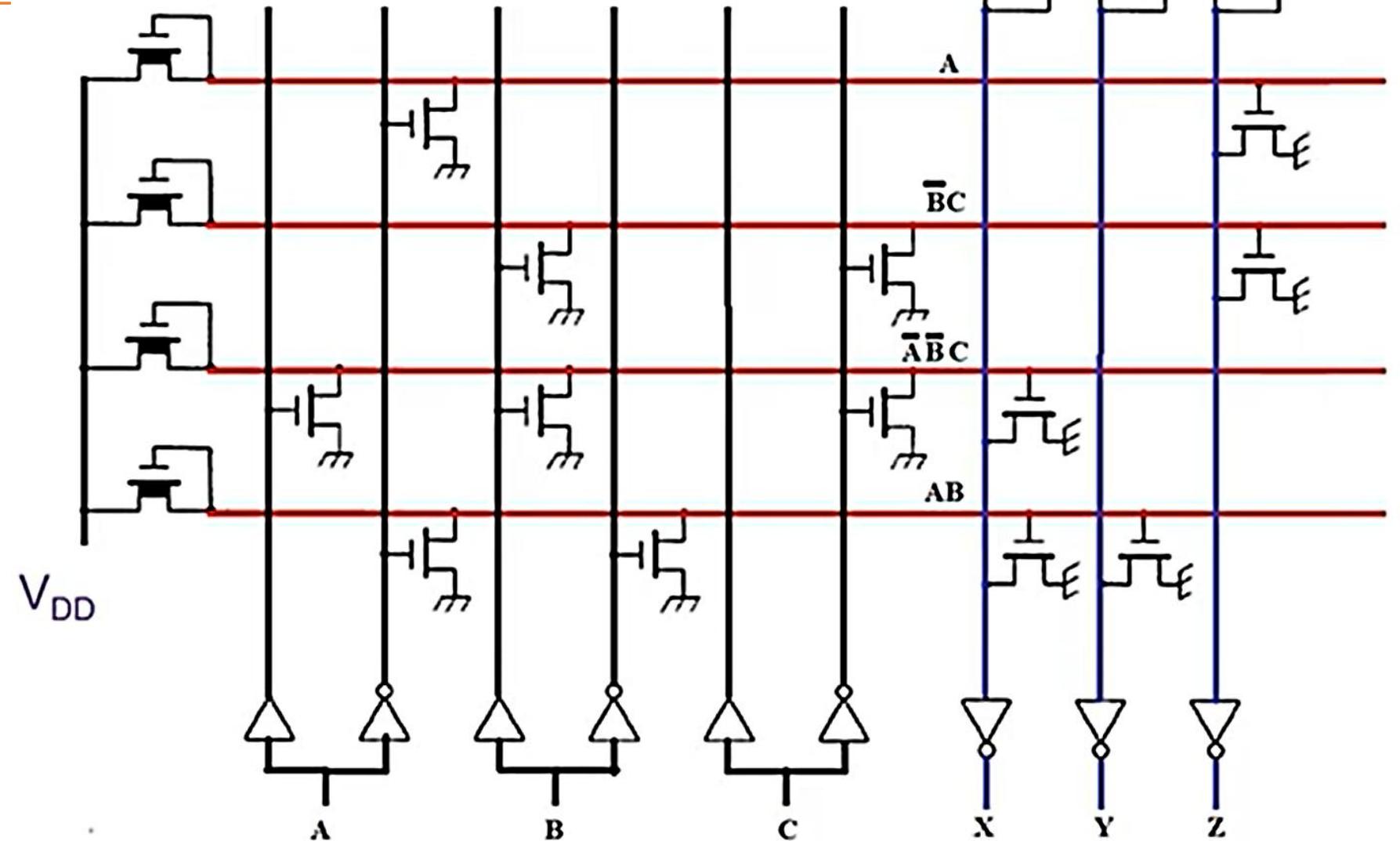


$$X = AB + \bar{A}\bar{B}C$$

$$Y = AB$$

$$Z = A + \bar{B}C$$

Stage 1: AND array

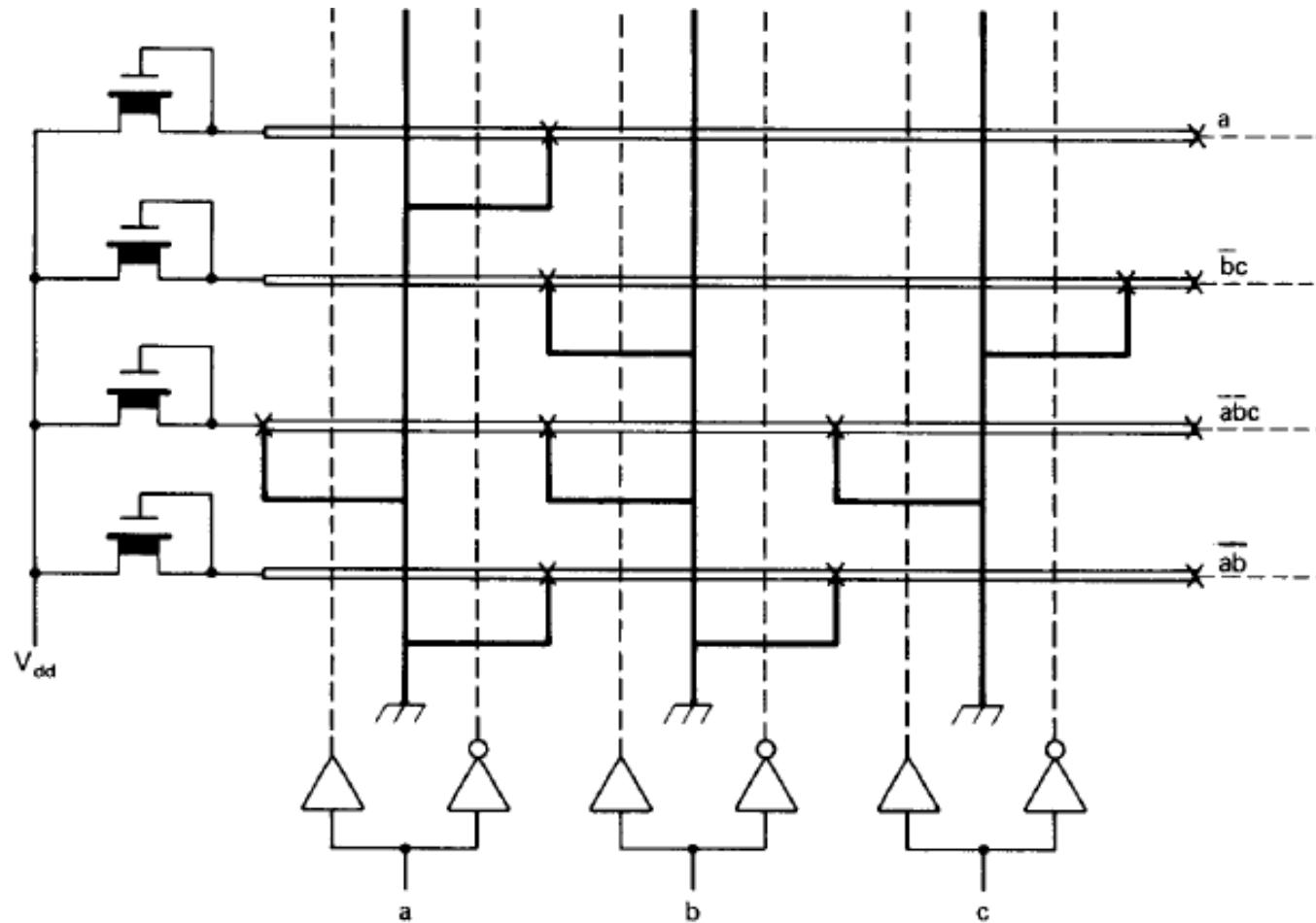


Stage 2: OR array



STICK DIAGRAM OF PLA

Stage 1: AND array

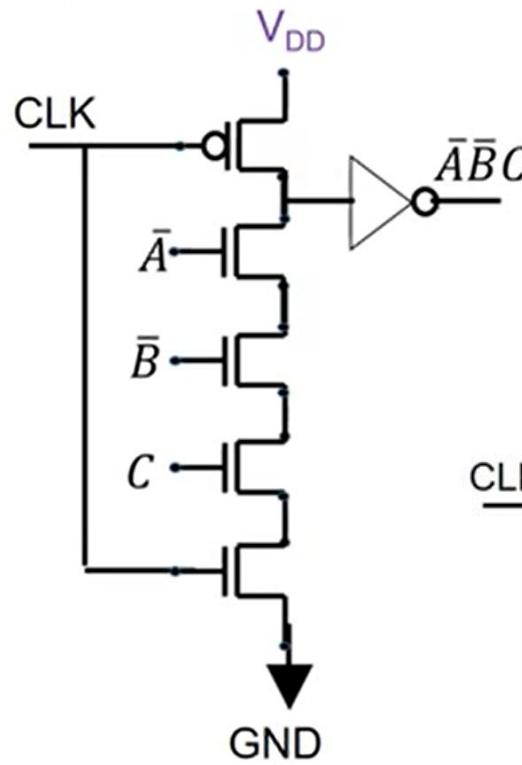
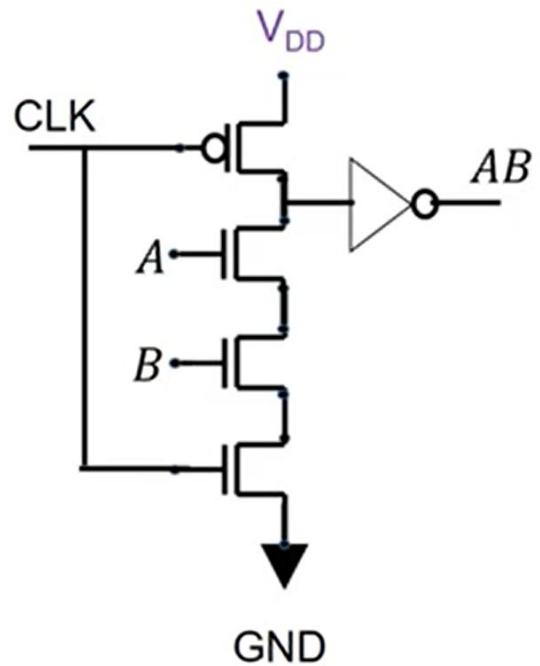


Stage 2: OR array

Do it for stage 2 also



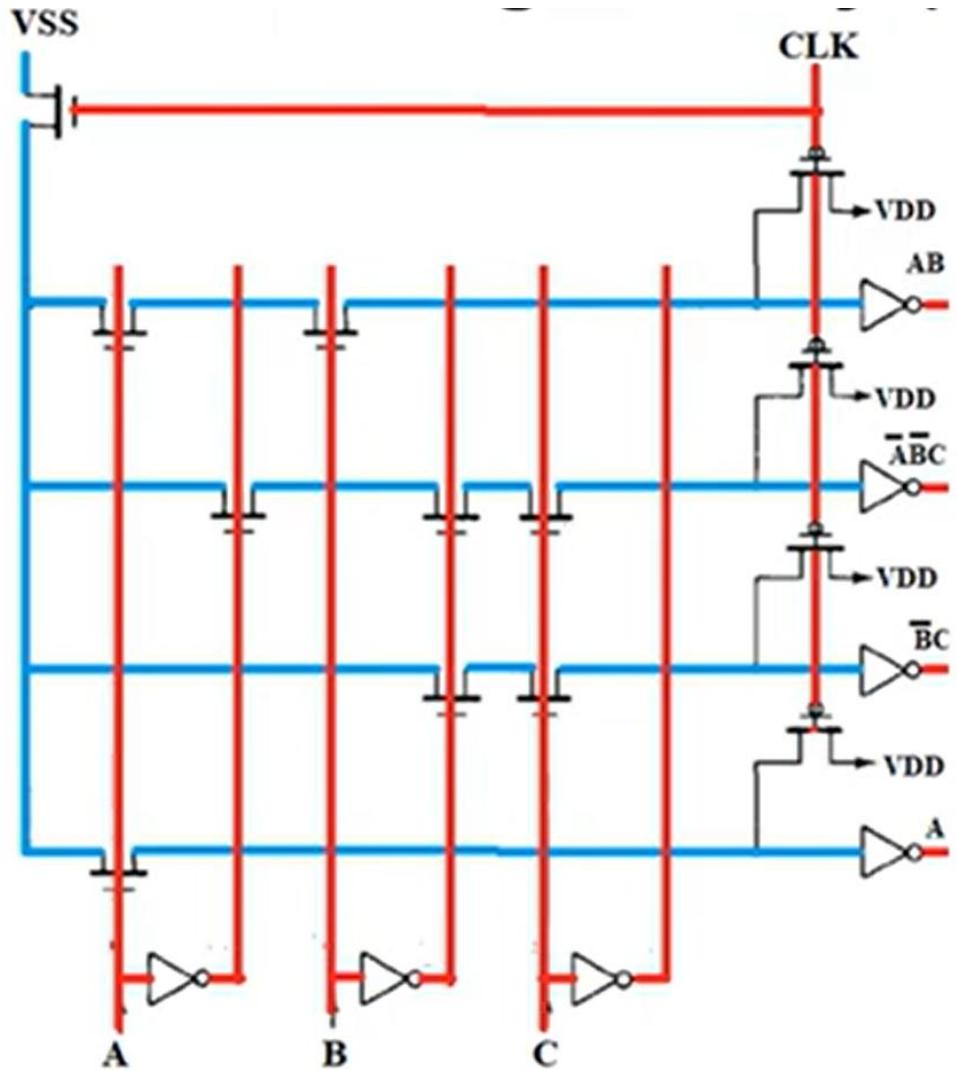
2. CMOS PLA



$$X = AB + \bar{A}\bar{B}C$$

$$Y = AB$$

$$Z = A + \bar{B}C$$



Stage 1: AND array

$$X = AB + \bar{A}\bar{B}\bar{C}$$

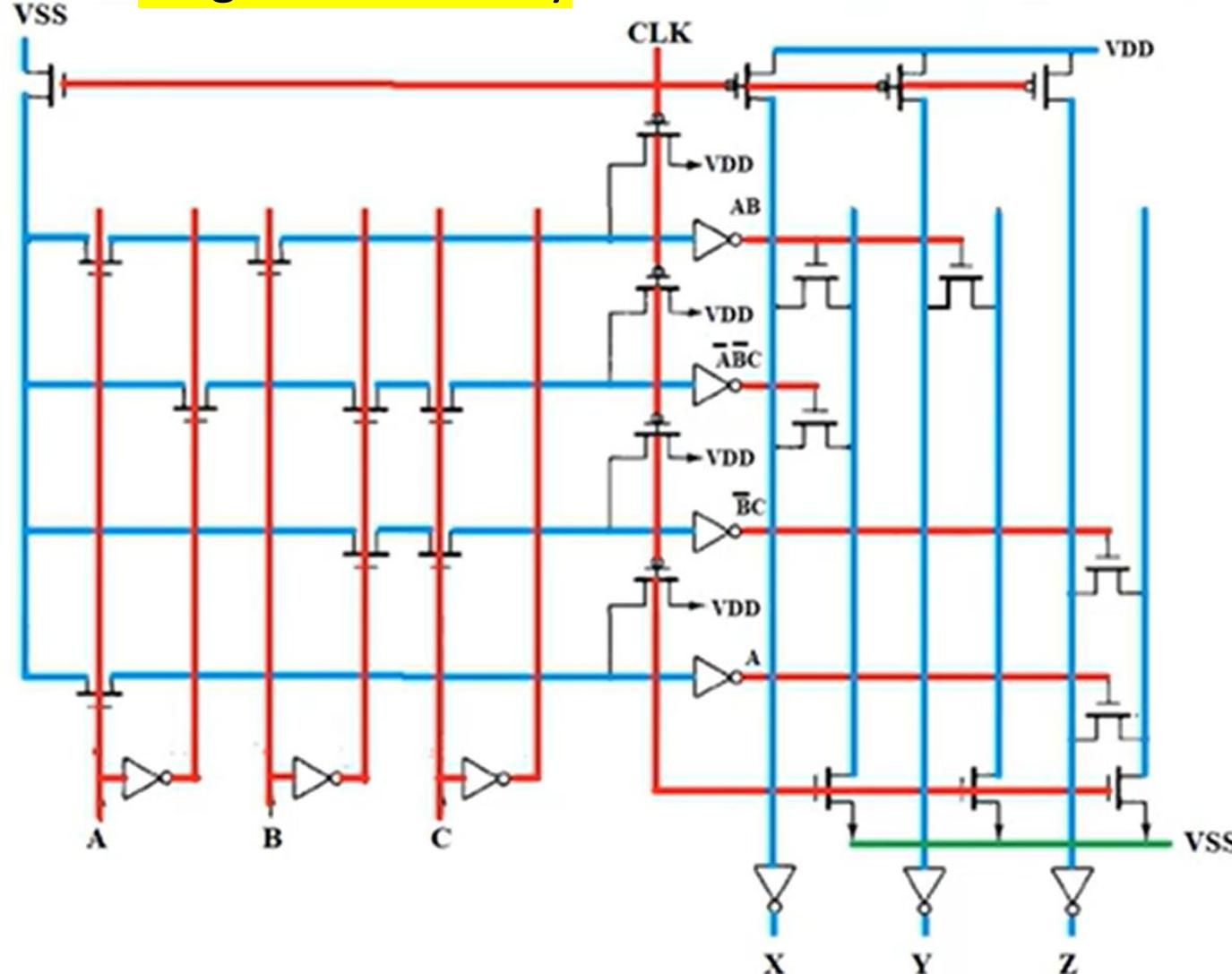
$$Y = AB$$

$$Z = A + \bar{B}C$$



Stage 1: AND array

Stage 2: OR array



$$X = AB + \bar{A}\bar{B}C$$

$$Y = AB$$

$$Z = A + \bar{B}C$$

Draw the stick diagram
of this circuit



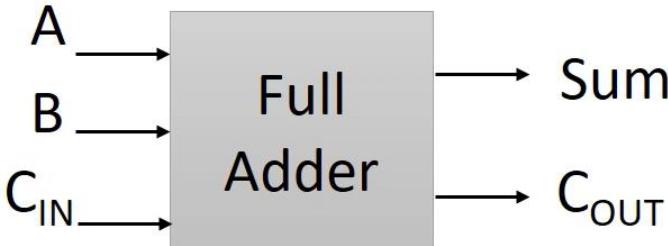
VLSI Design: Subsystem Design

- ❖ ALU
- ❖ Parity Generator
- ❖ Bus arbitration logic and shifter
- ❖ Crossbar switch
- ❖ Barrel Shifter



=

Arithmetic Logic Unit (ALU)



A	B	C _{IN}	Sum	C _{OUT}
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

BC_{IN}

	00	01	11	10
0	0	1	0	1
1	1	0	1	0

$$\text{Sum} = (\text{A XOR B}) \text{ XOR } \text{C}_{\text{IN}}$$

BC_{IN}

	00	01	11	10
0	0	0	1	0
1	0	1	1	1

$$C_{\text{OUT}} = AB \text{ OR } [(A'B \text{ OR } AB') \text{ AND } \text{C}_{\text{IN}}]$$



When Cin = 0,
Then Sum= A xor B
When Cin = 1,
Then Sum= A xnor B

When Cin = 0,
Then Cout = A and B
When Cin = 1,
Then Cout = A or B



$$\text{Sum} = (A \text{ XOR } B) \text{ XOR } C_{IN}$$

If $C_{IN}=0$

$$\text{Sum} = (A \text{ XOR } B) \text{ XOR } 0$$

$\text{Sum} = (A \text{ XOR } B) \rightarrow \text{XOR Function}$

If $C_{IN}=1$

$$\text{Sum} = (A \text{ XOR } B) \text{ XOR } 1$$

$\text{Sum} = (A \text{ XNOR } B) \rightarrow \text{XNOR Function}$

$$C_{OUT} = AB \text{ OR } [(A'B \text{ OR } AB') \text{ AND } C_{IN}]$$

If $C_{IN}=0$

$$C_{OUT} = AB \text{ OR } [(A'B \text{ OR } AB') \text{ AND } 0]$$

$C_{OUT} = AB \rightarrow \text{AND Function}$

If $C_{IN}=1$

$$C_{OUT} = AB \text{ OR } [(A'B \text{ OR } AB') \text{ AND } 1]$$

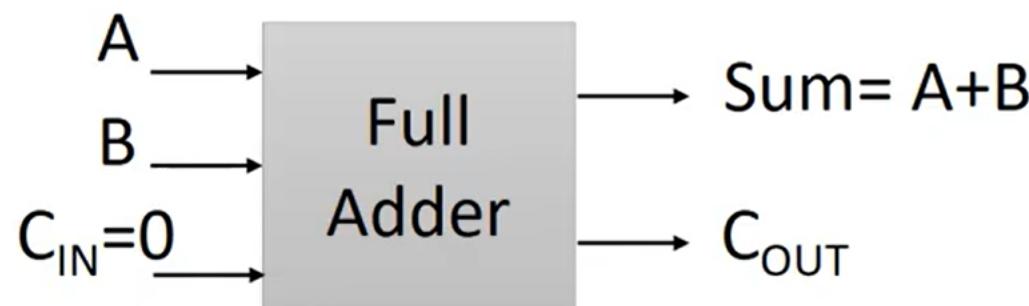
$C_{OUT} = AB \text{ OR } A'B \text{ OR } AB'$

$C_{OUT} = A \text{ OR } B \rightarrow \text{OR Function}$

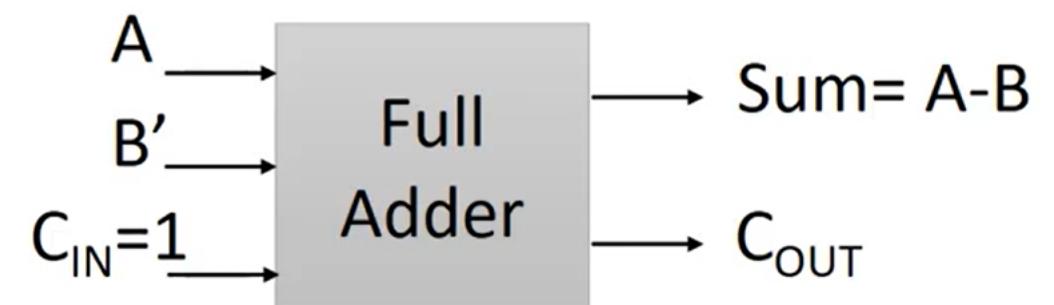




Addition



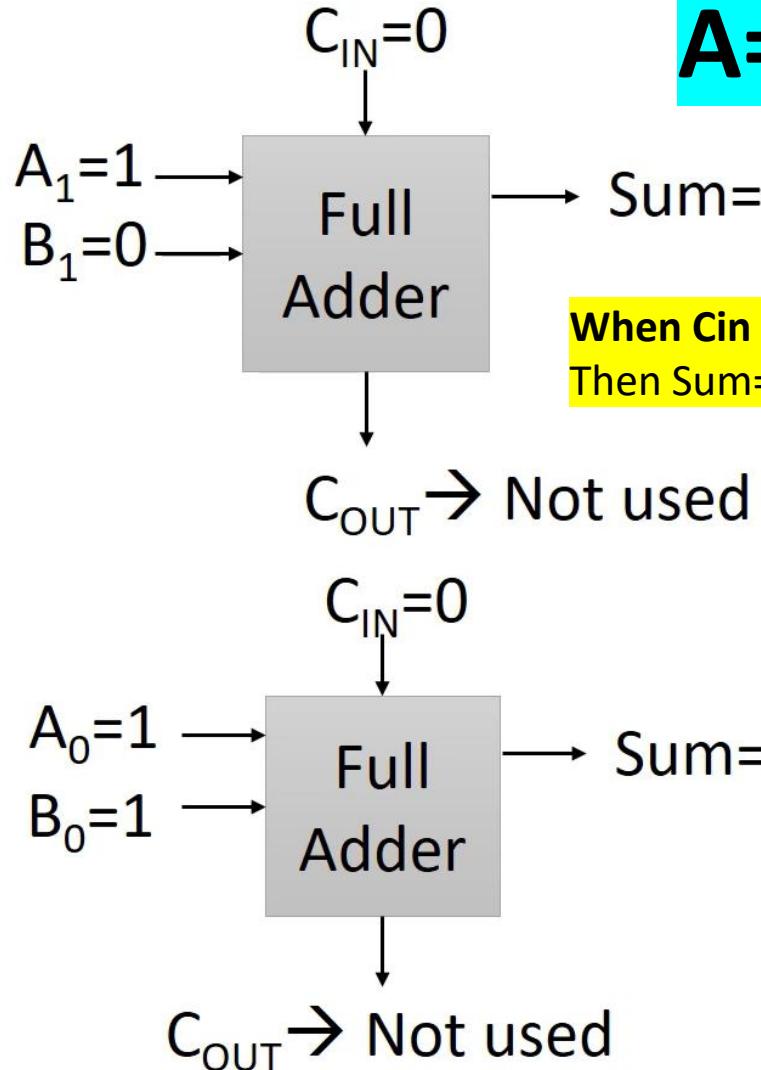
Subtraction



$$A-B = A+(B'+1)$$

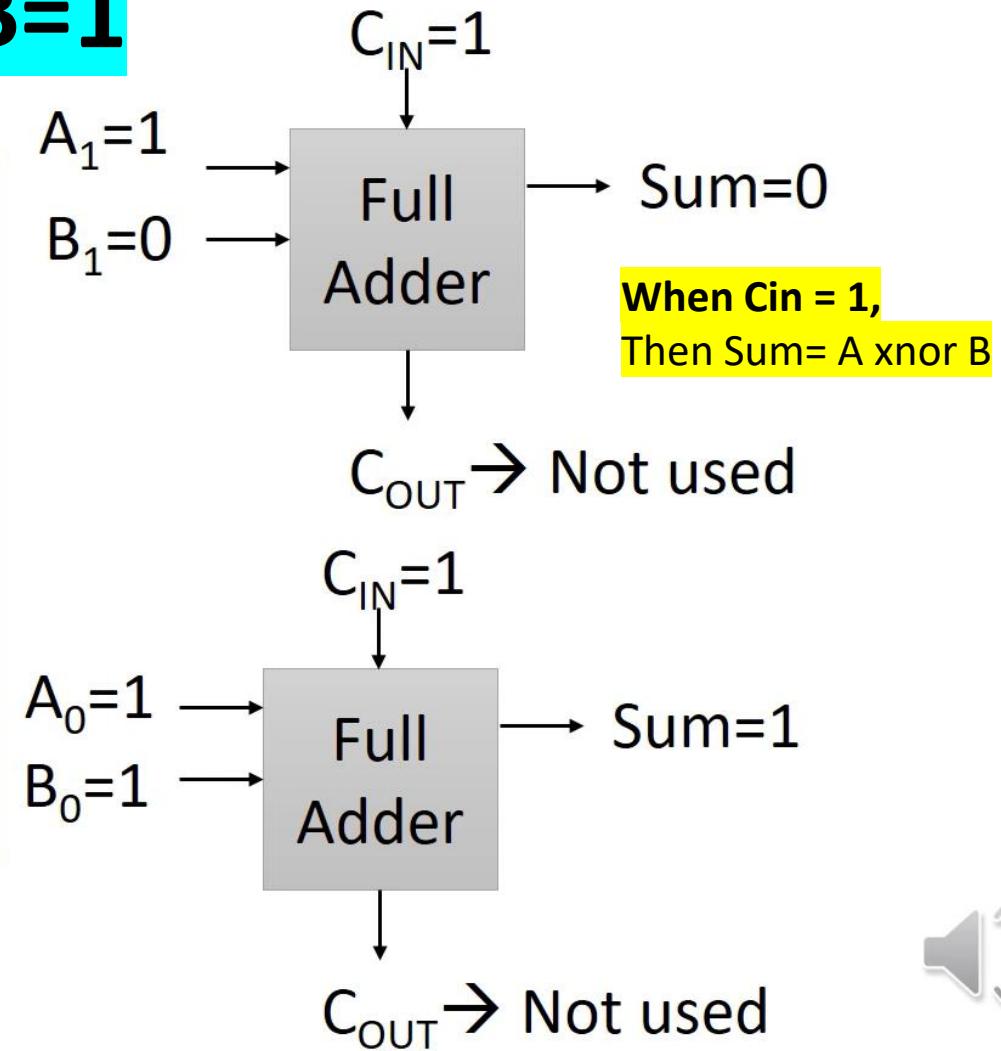


Sum = (A XOR B) XOR C_{IN}

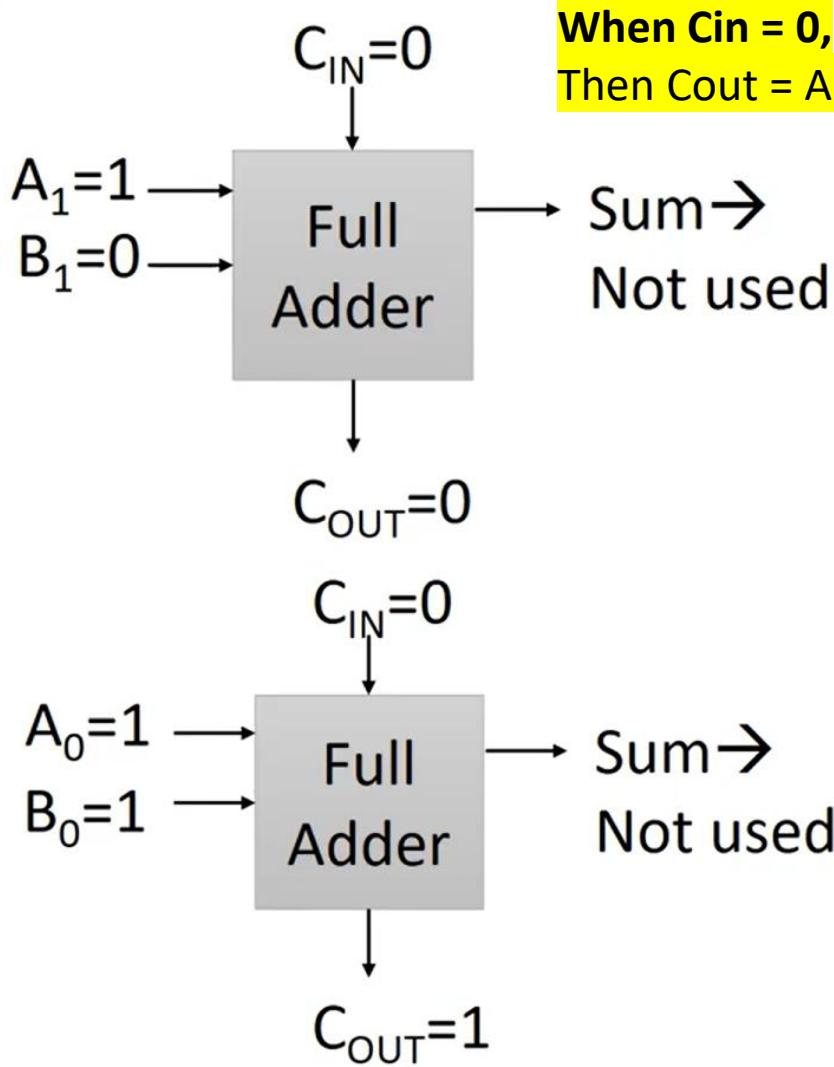


A=3 and B=1

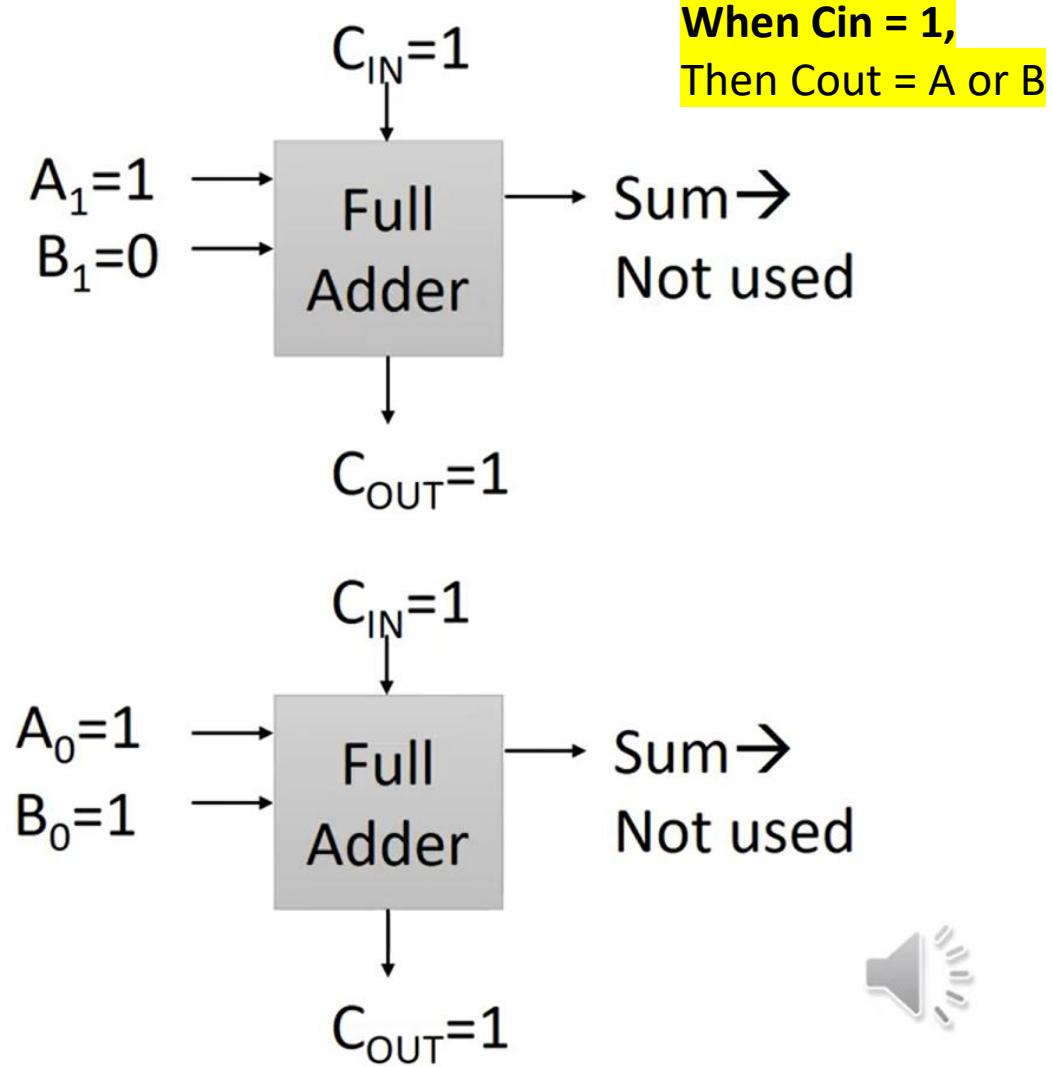
Sum = (A XOR B) XOR C_{IN}



$$C_{OUT} = AB \text{ OR } [(A'B \text{ OR } AB') \text{ AND } C_{IN}]$$



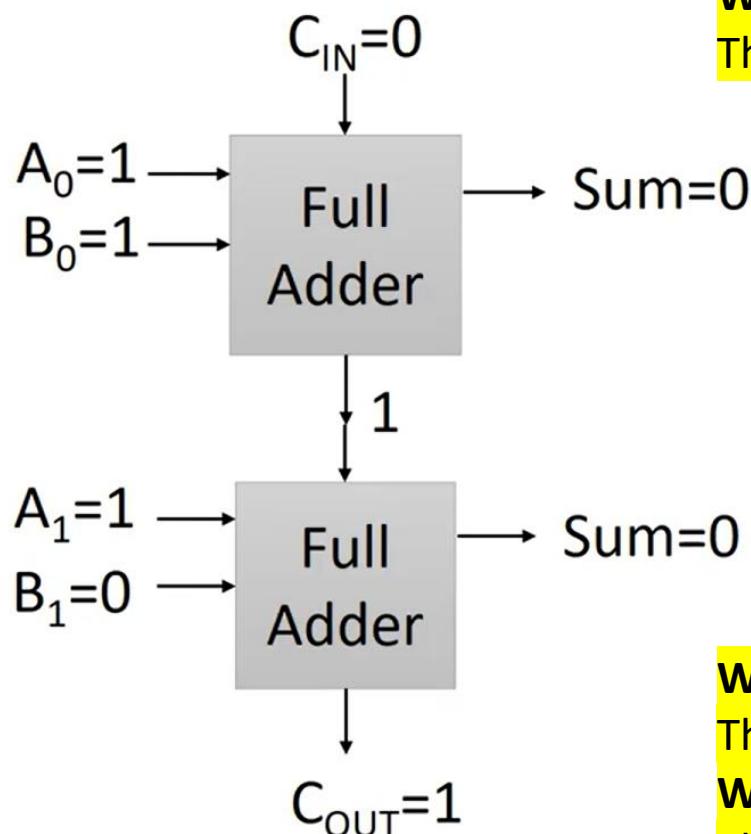
$$C_{OUT} = AB \text{ OR } [(A'B \text{ OR } AB') \text{ AND } C_{IN}]$$





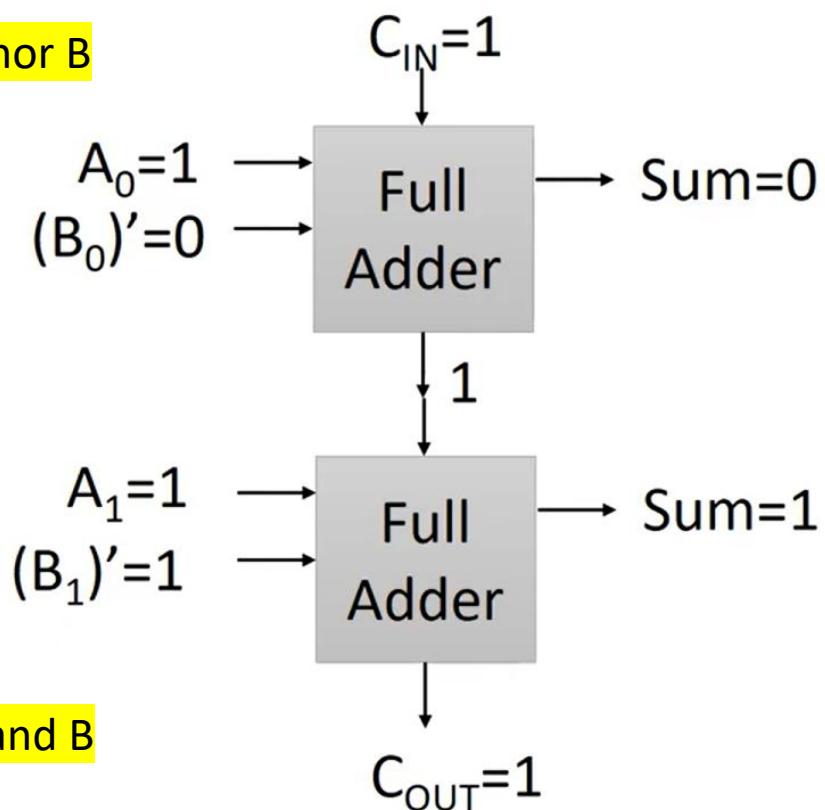
A=3 and B=1

Addition



When $Cin = 0$,
Then $Sum = A \text{ xor } B$
When $Cin = 1$,
Then $Sum = A \text{ xnor } B$

Subtraction

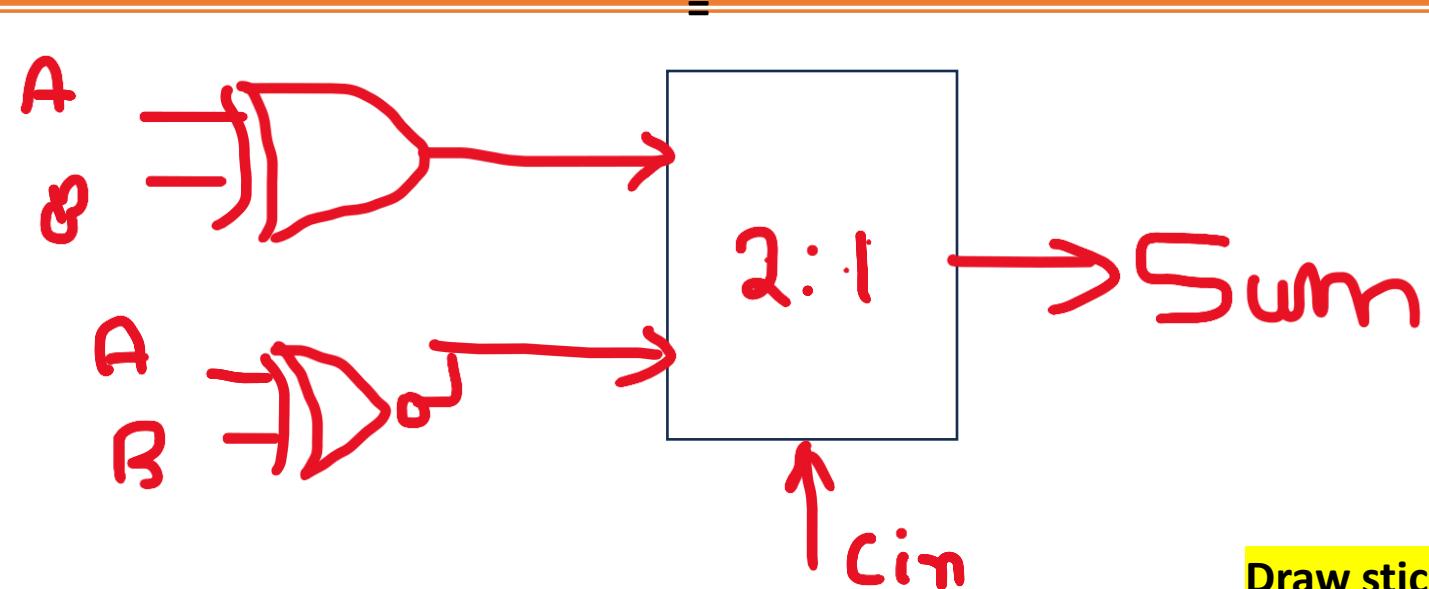


When $Cin = 0$,
Then $Cout = A \text{ and } B$
When $Cin = 1$,
Then $Cout = A \text{ or } B$



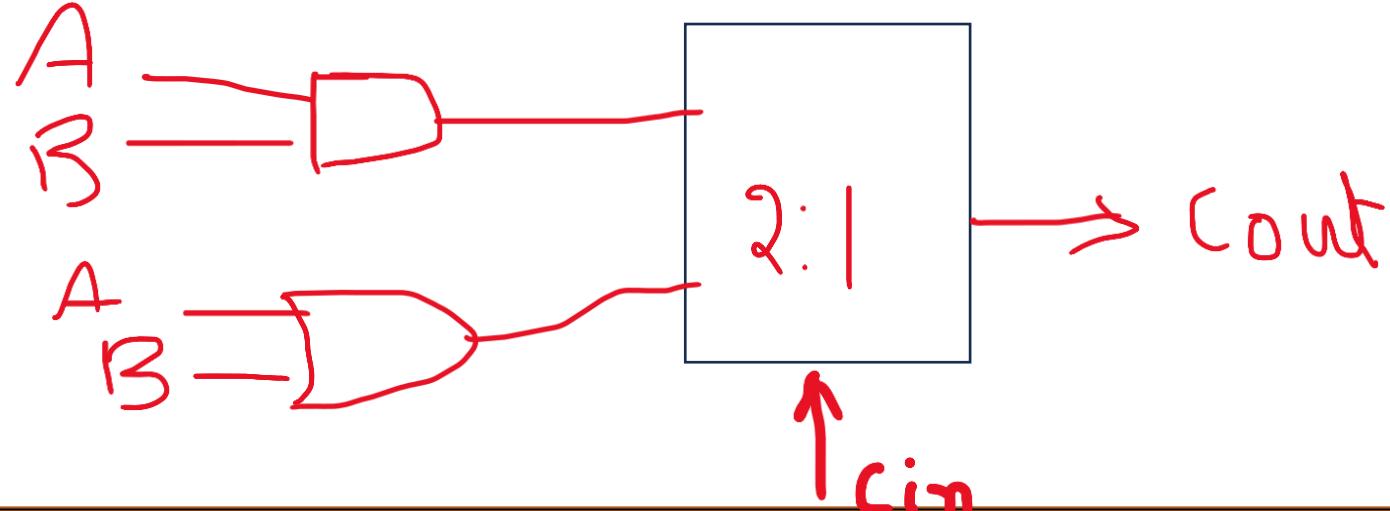


When $Cin = 0$,
Then $Sum = A \oplus B$
When $Cin = 1$,
Then $Sum = A \oplus B$



When $Cin = 0$,
Then $Cout = A \oplus B$
When $Cin = 1$,
Then $Cout = A \oplus B$

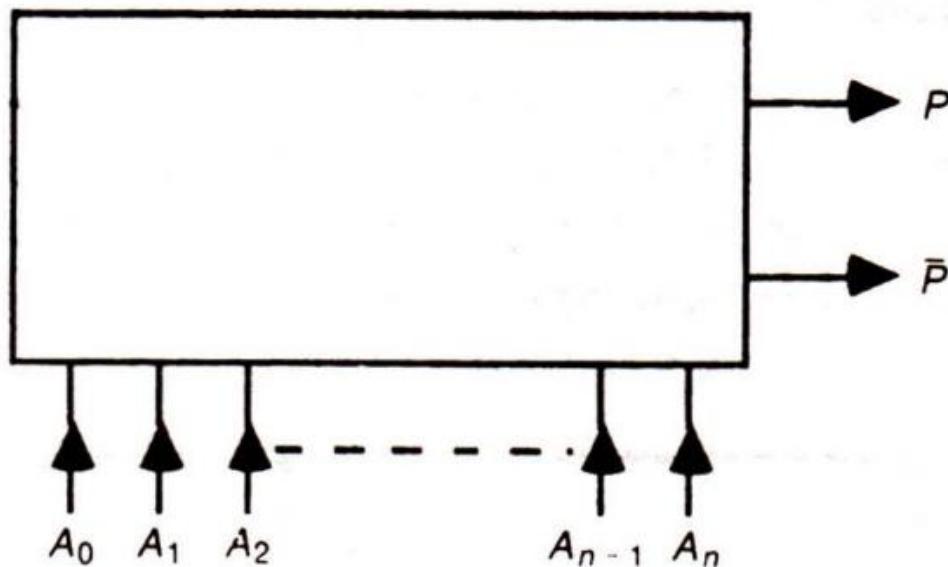
Draw stick diagram





Parity generator

- A circuit is to be designed to indicate the parity of a binary number or word.
- The requirement is indicated in Figure 6.15 for an $(n + 1)$ -bit input.

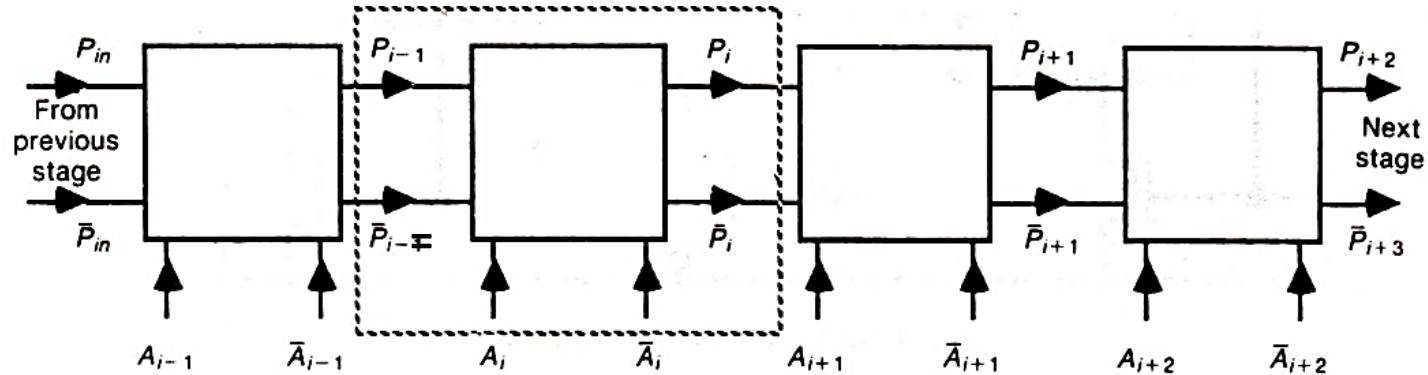


Note: $P = \begin{cases} 1 & \text{Even number of 1s at input} \\ 0 & \text{Odd number of 1s at input} \end{cases}$

FIGURE 6.15 Parity generator basic block diagram.



- Since the number of bits is undefined, we must find a general solution on a cascadable bit-wise basis so that “n” can have any value.
- A suitably regular structure is set out in Figure 6.16.
- From this, we may recognize _ a standard or basic one-bit cell from which an n-bit parity generator may be formed. Such a cell is shown in Figure 6.17.



Note: Parity requirements are set at the left-most cell where $P_{in} = 1$ sets even and $P_{in} = 0$ sets odd parity.

FIGURE 6.16 Parity generator—structured design approach.

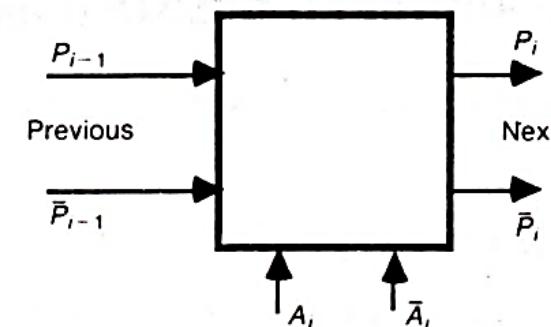
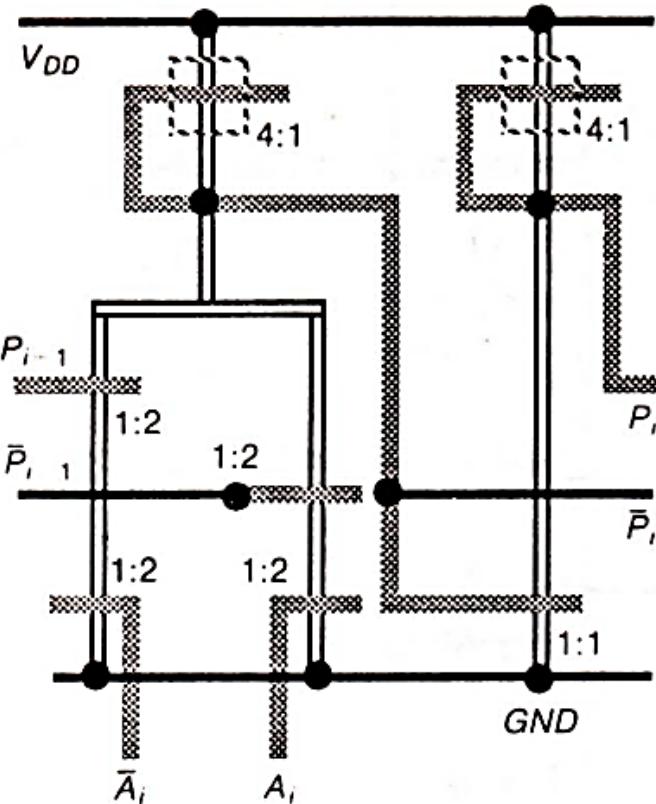


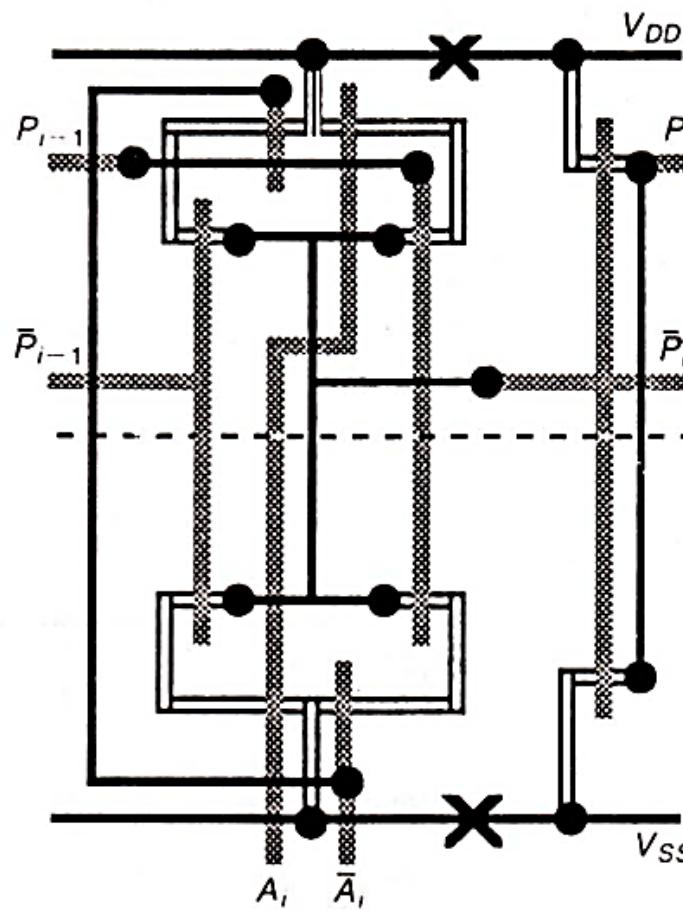
FIGURE 6.17 Parity generator—basic one-bit cell.

$A_i \neq 1$ parity is changed, $P_i = \bar{P}_{i-1}$
 $A_i = 0$ parity is unchanged, $P_i = P_{i-1}$

$$P_i = \bar{P}_{i-1} \cdot A_i + P_{i-1} \cdot \bar{A}_i$$



(a) nMOS



(b) CMOS

$$P_i = \bar{P}_{i-1} \cdot A_i + P_{i-1} \cdot \bar{A}_i$$

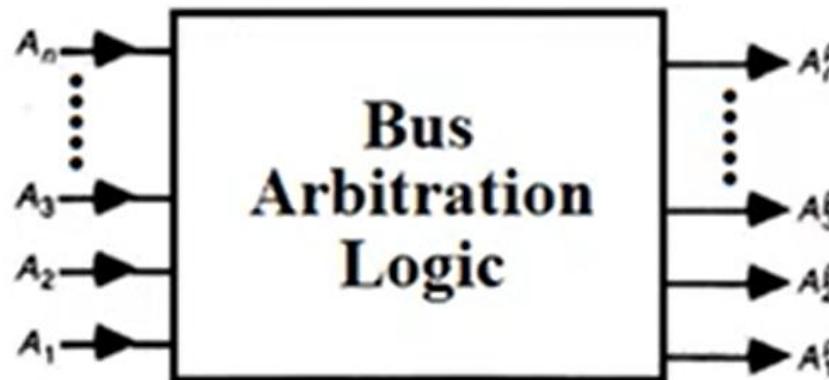
FIGURE 6.18 Stick diagrams (parity generator).



Bus Arbitration logic

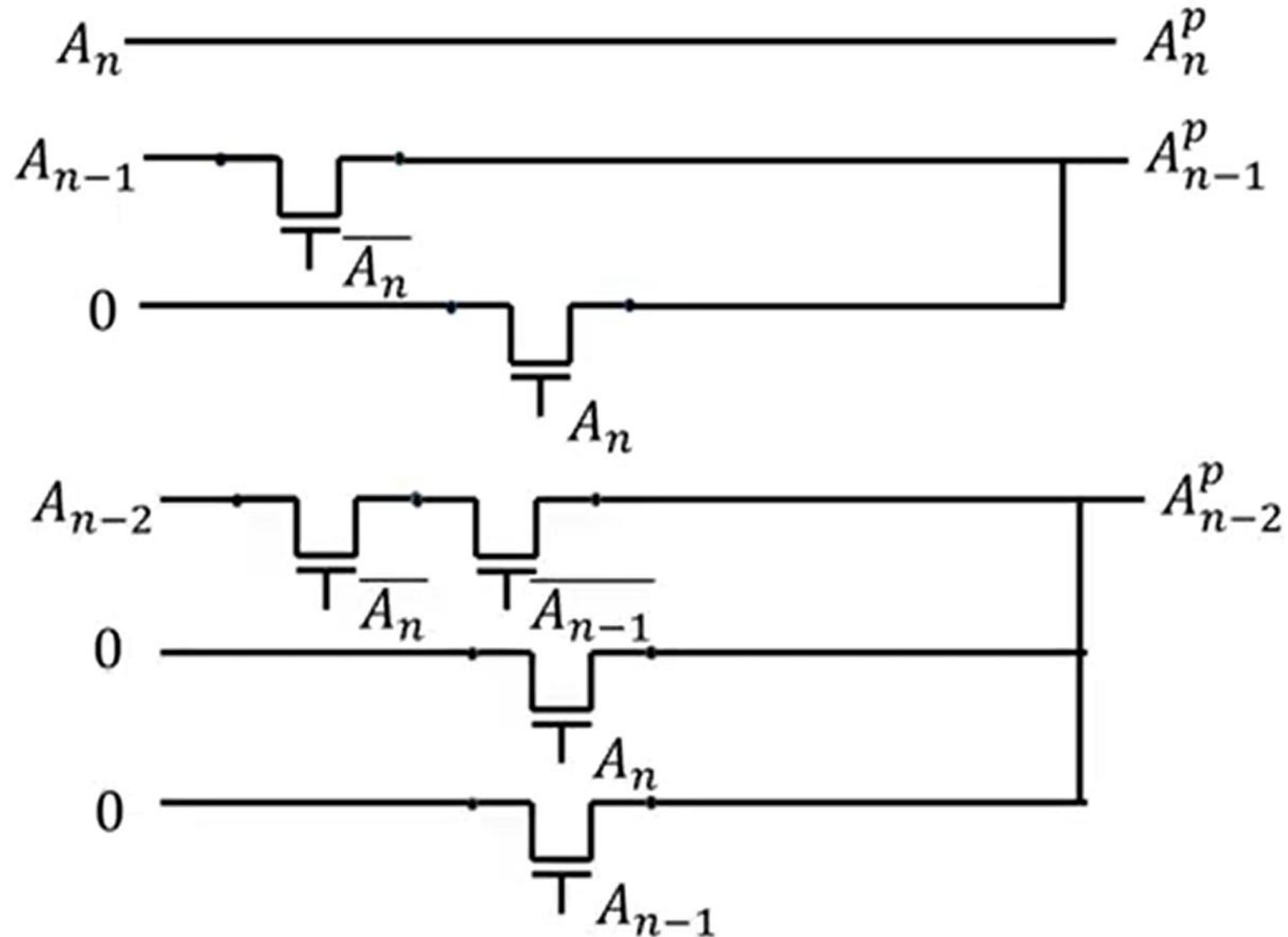


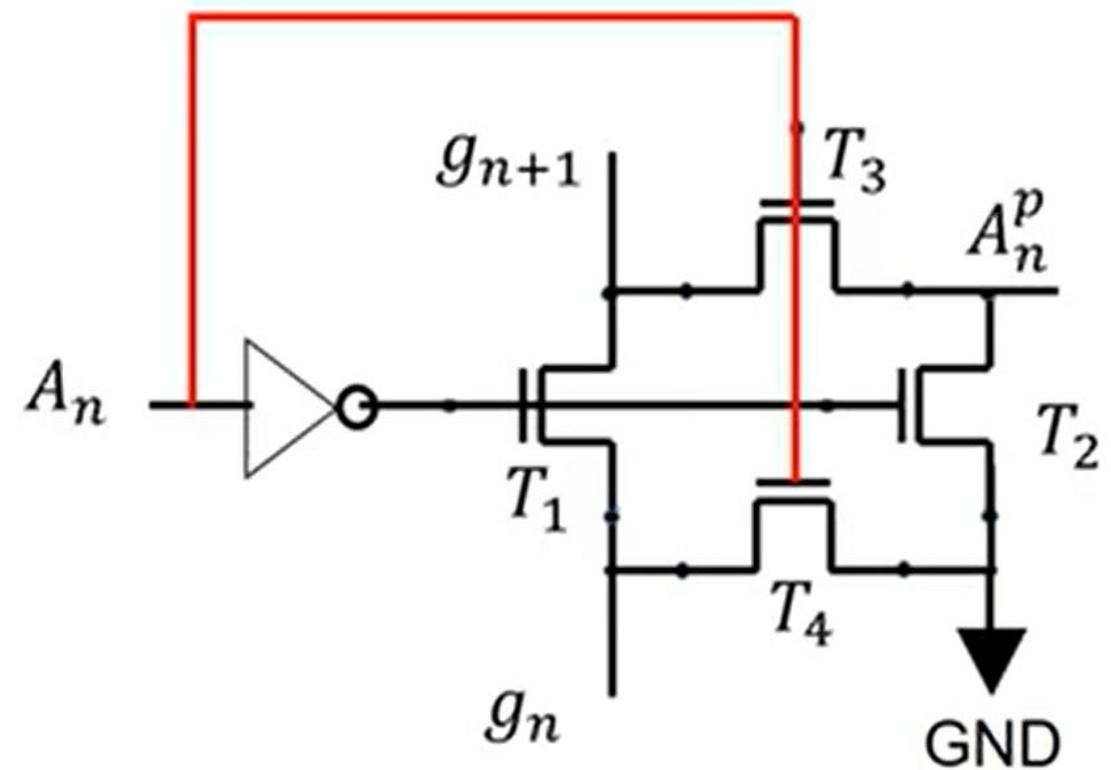
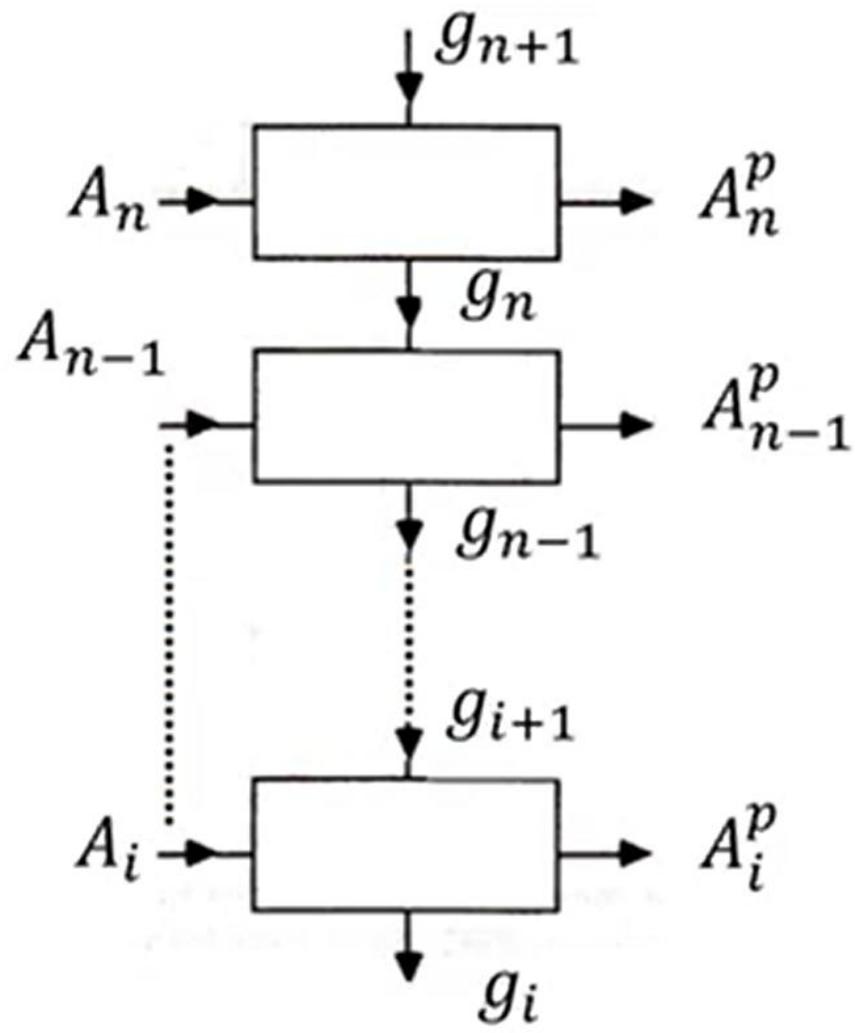
Highest priority



Lowest priority

A_n	A_3	A_2	A_1	A_n^P	A_3^P	A_2^P	A_1^P
0		0	0	0	0		0	0	0
0		0	0	1	0		0	0	1
0		0	1	X	0		0	1	0
0		1	X	X	0		1	0	0
.....
1		X	X	X	1		0	0	0

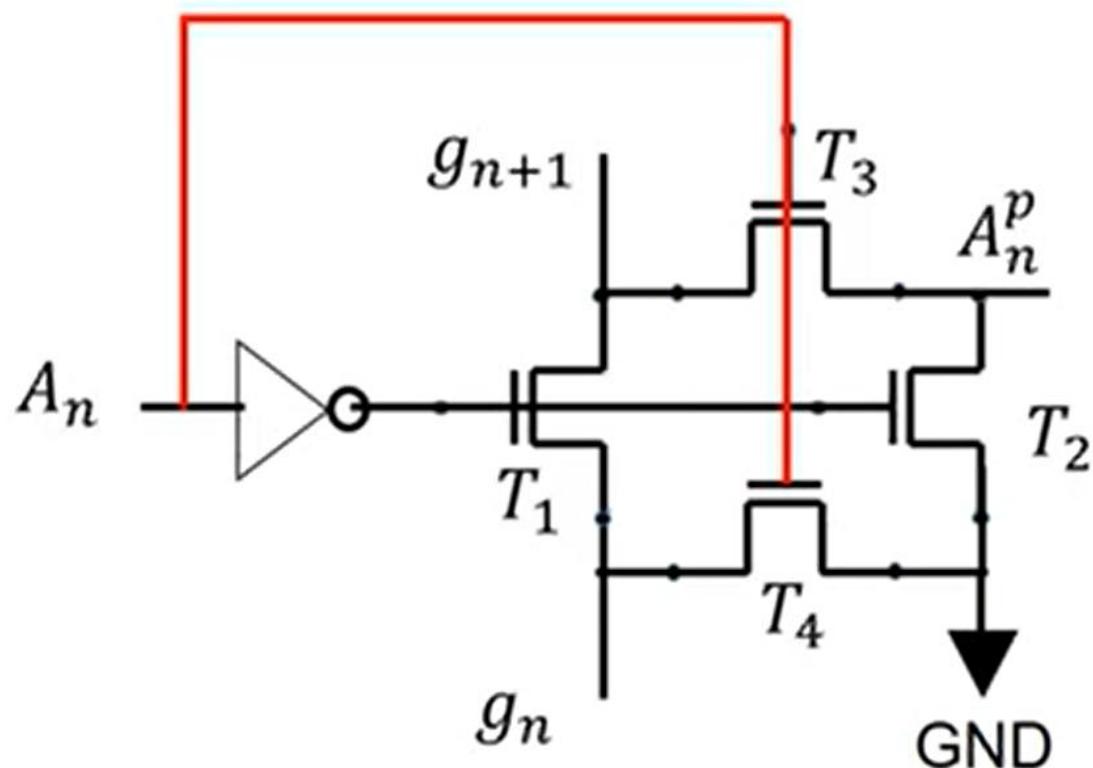




$$A_n^p = A_n g_{n+1}$$

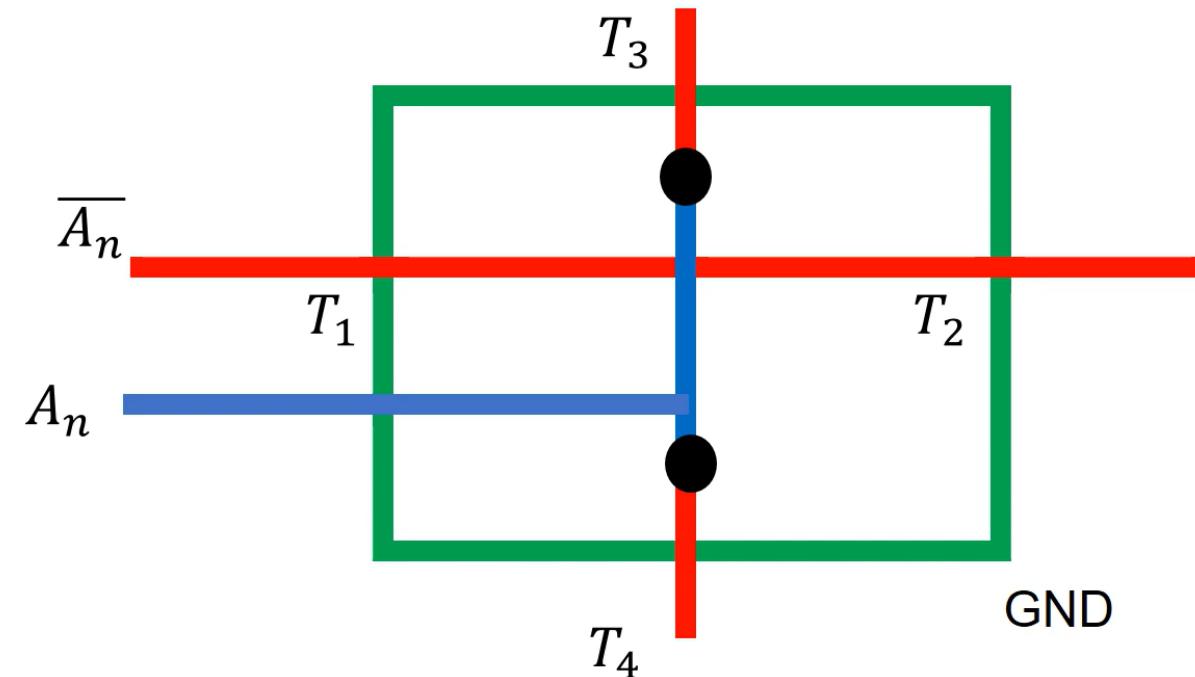
$$g_n = g_{n+1} \overline{A_n}$$





$$A_n^p = A_n g_{n+1} \bar{A}_n$$

$$g_n = g_{n+1} \bar{A}_n$$





Shifter



=

- Any general purpose n-bit shifter should be able to shift incoming data by up to $n - 1$ places in a right-shift or left-shift direction.

The shifter must have:

- input from a four-line parallel data bus;
- four output lines for the shifted data;
- means of transferring input data to output lines with any shift from zero to three bits inclusive.

1. CROSSBAR SWITCH
2. BARREL SHIFTER



=

1. Crossbar switch



- The attempt to design shift register using cross bar switch (like MOS pass transistor and transmission gate etc.)
- We must have the strategy in which **data flow horizontally** and control signals are kept vertically.
- **Switch and relay contact** based switching networks is called -the *crossbar switch*.

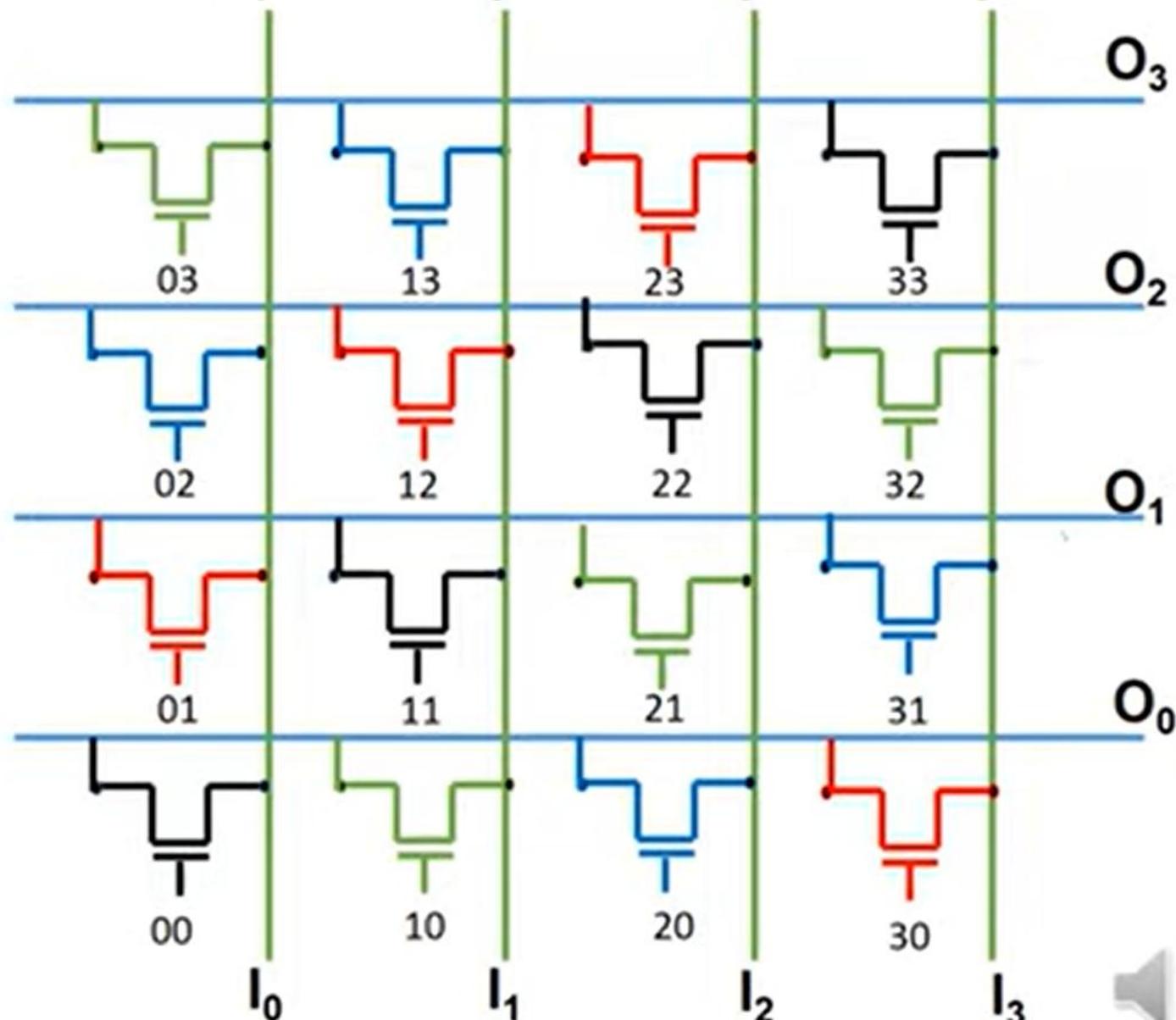


MOS switch implementation of a 4×4 crossbar switch

Shift	O_0	O_1	O_2	O_3
0	I_0	I_1	I_2	I_3
1	I_3	I_0	I_1	I_2
2	I_2	I_3	I_0	I_1
3	I_1	I_2	I_3	I_0

CROSSBAR SWITCH

If all switches are ON then all input are connected to output and short circuit is possible. We may take the combination of switches to shifts of zero, one, two and three bits.





2. Barrel shifter

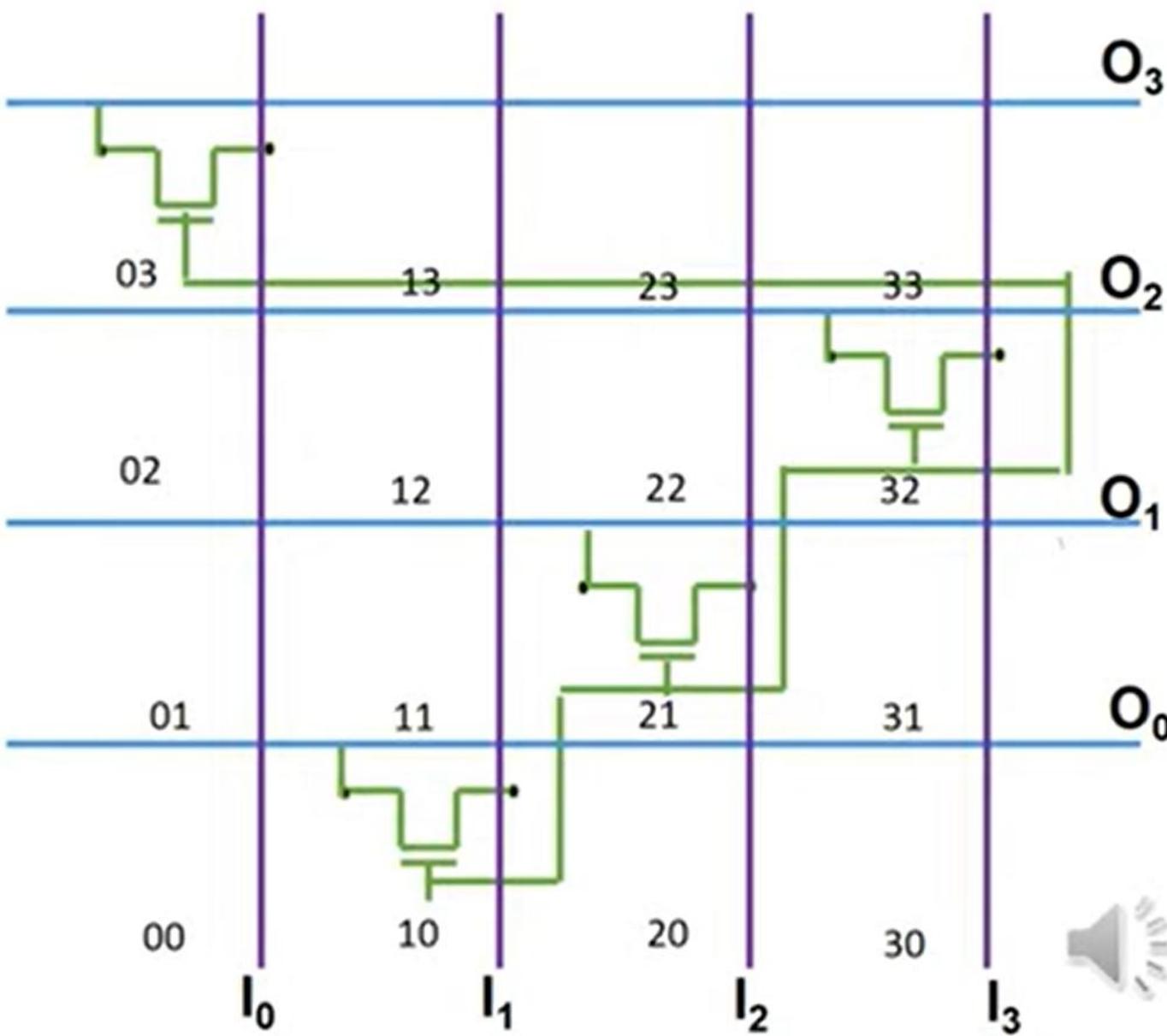
- A *barrel shifter* is a digital circuit that can shift a data word by a specified number of bits in each clock cycle.
- The resulting arrangement is known as a *barrel shifter* as shown below (next slide).

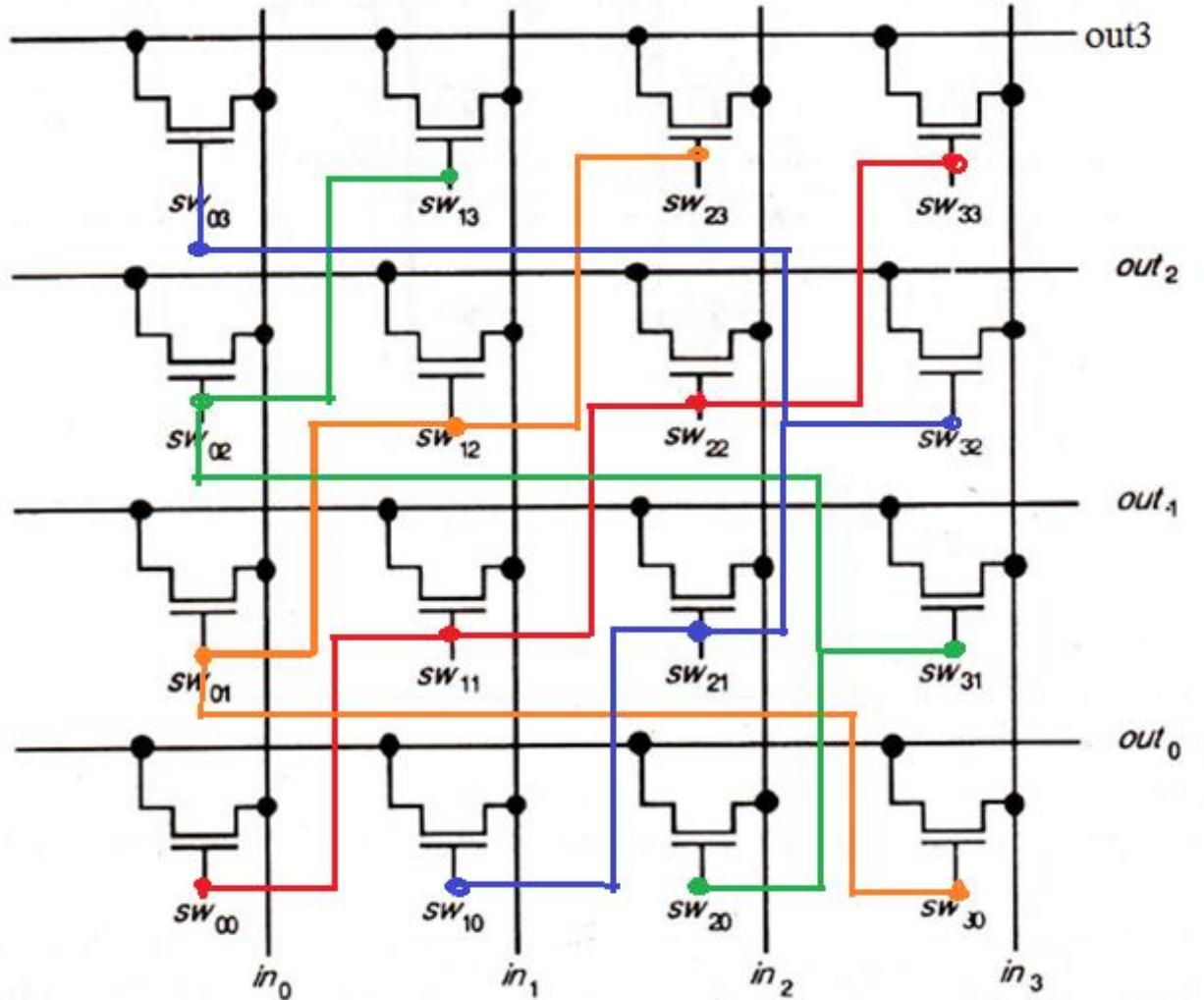
Out 0	Out 1	Out 2	Out 3	
In 0	In1	In2	In3	Shift 0 (eg. 1001)
In3	In0	In1	In2	Shift1 (eg. 1100)
In2	In3	In0	In1	Shift2 (eg. (0110)
In1	In2	In3	In0	Shift3 (eg. (0011)



Shift	O_0	O_1	O_2	O_3
0	I_0	I_1	I_2	I_3
1	I_3	I_0	I_1	I_2
2	I_2	I_3	I_0	I_1
3	I_1	I_2	I_3	I_0

Shift	O_0	O_1	O_2	O_3
0	1	0	0	0
1	0	1	0	0
2	0	0	1	0
3	0	0	0	1





Legend:
SW₁₀ — Output
Input
Shift 0
Shift 1
Shift 2
Shift 3

No of rows = word length of data
No of column= maximum shift in data

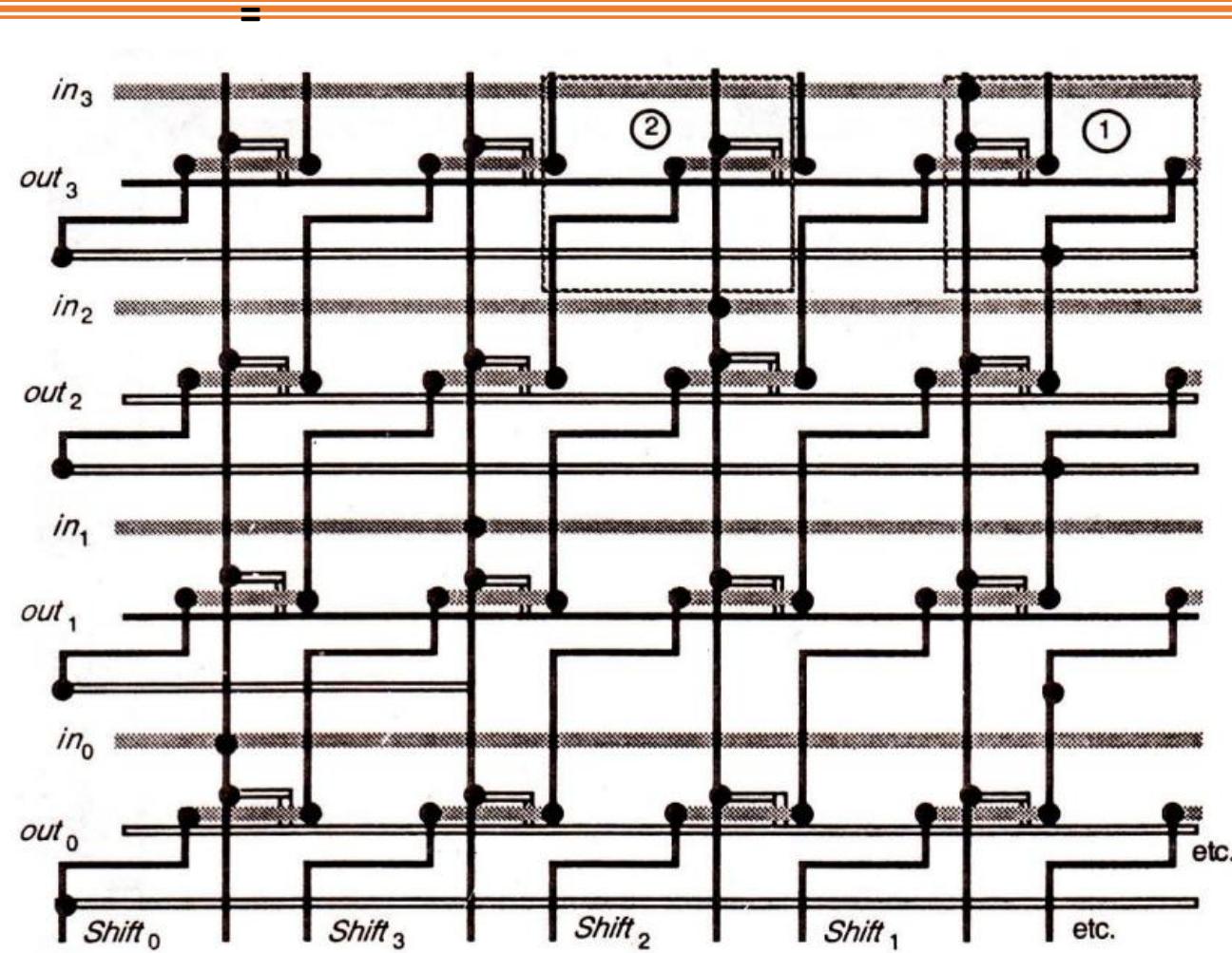
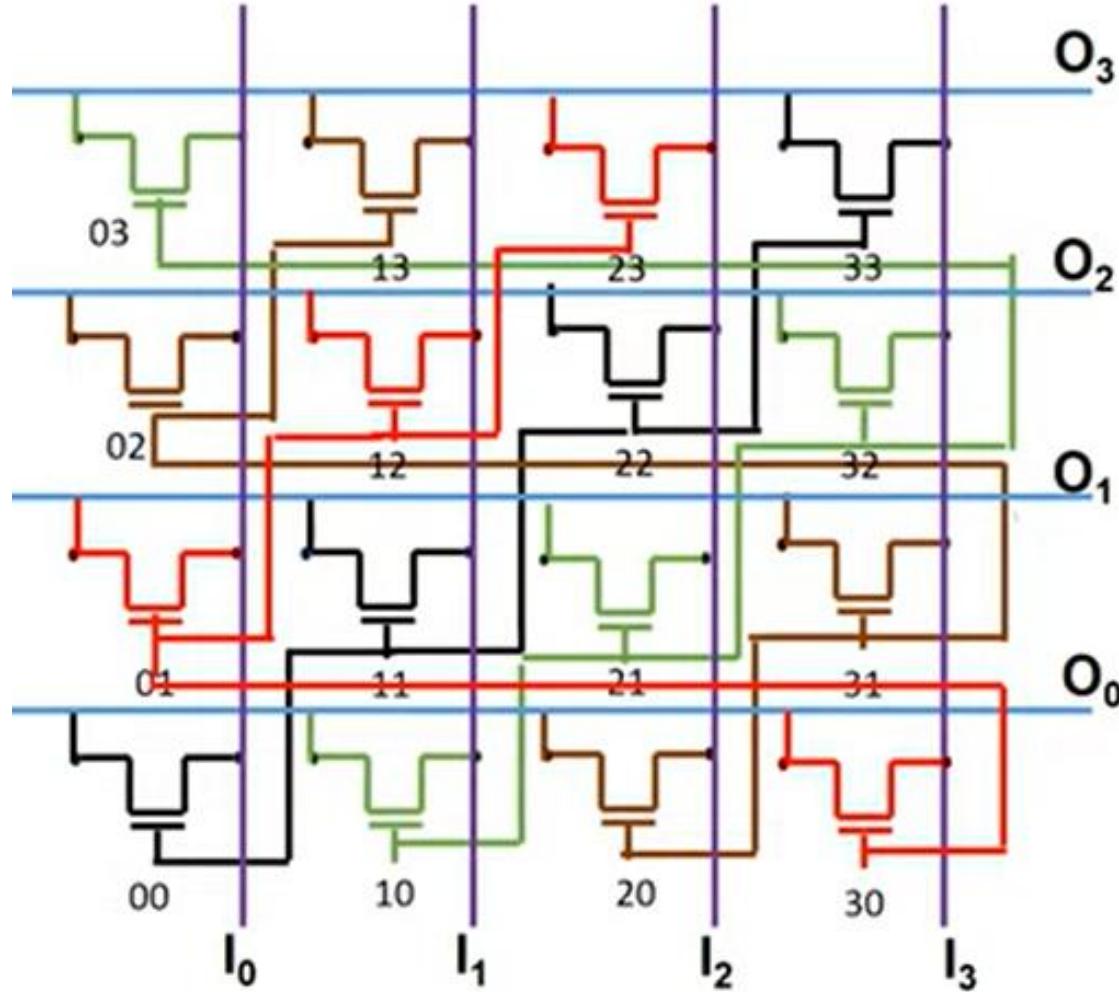
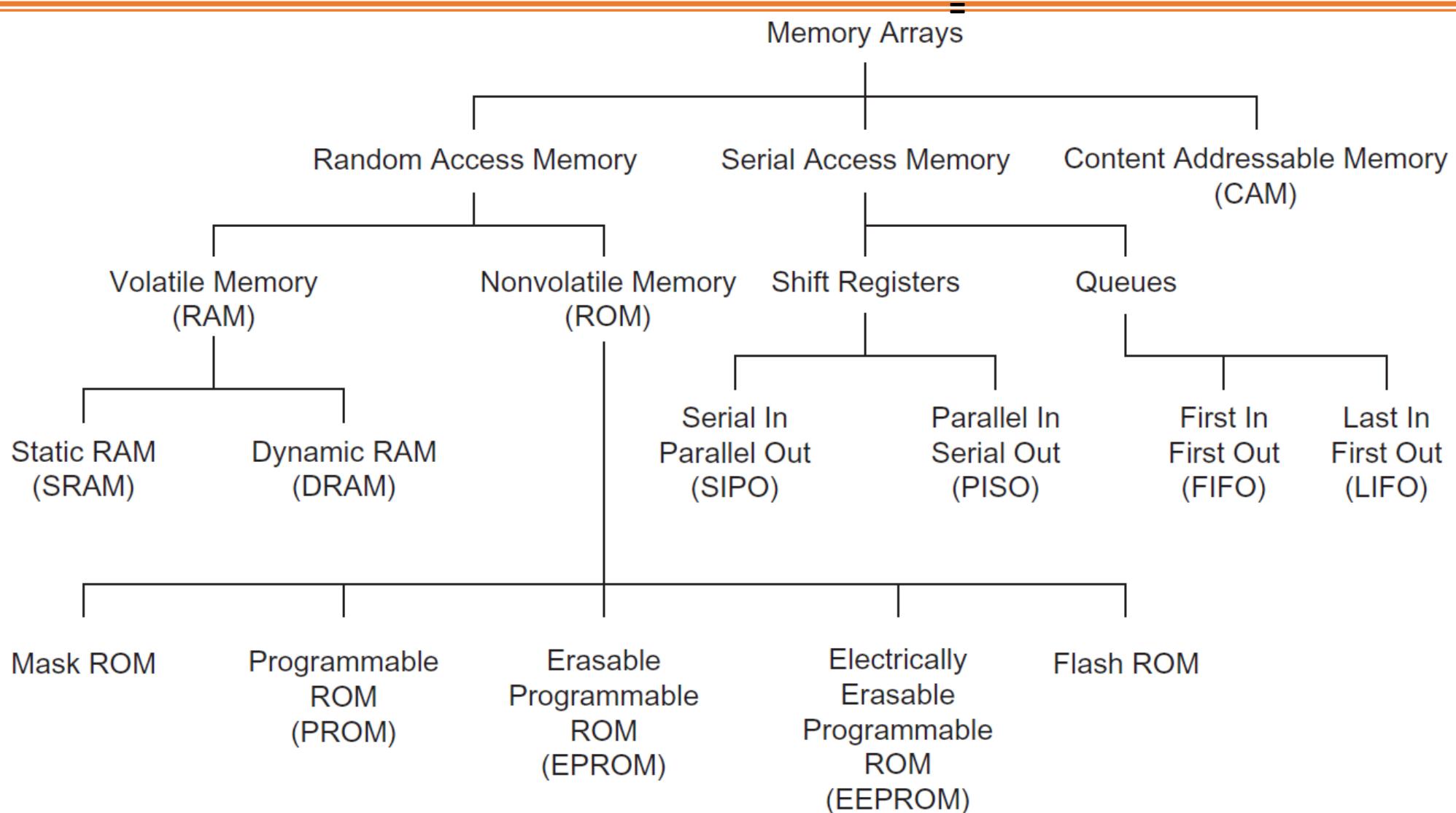


FIGURE 7.8 One possible stick diagram for a 4×4 barrel shifter.



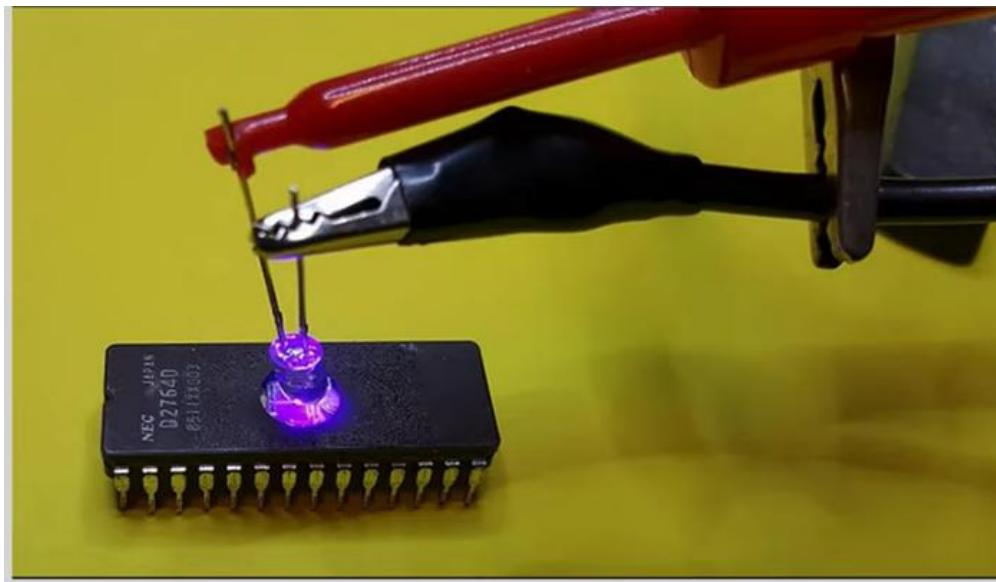
EPROM, EEPROM and flash memory

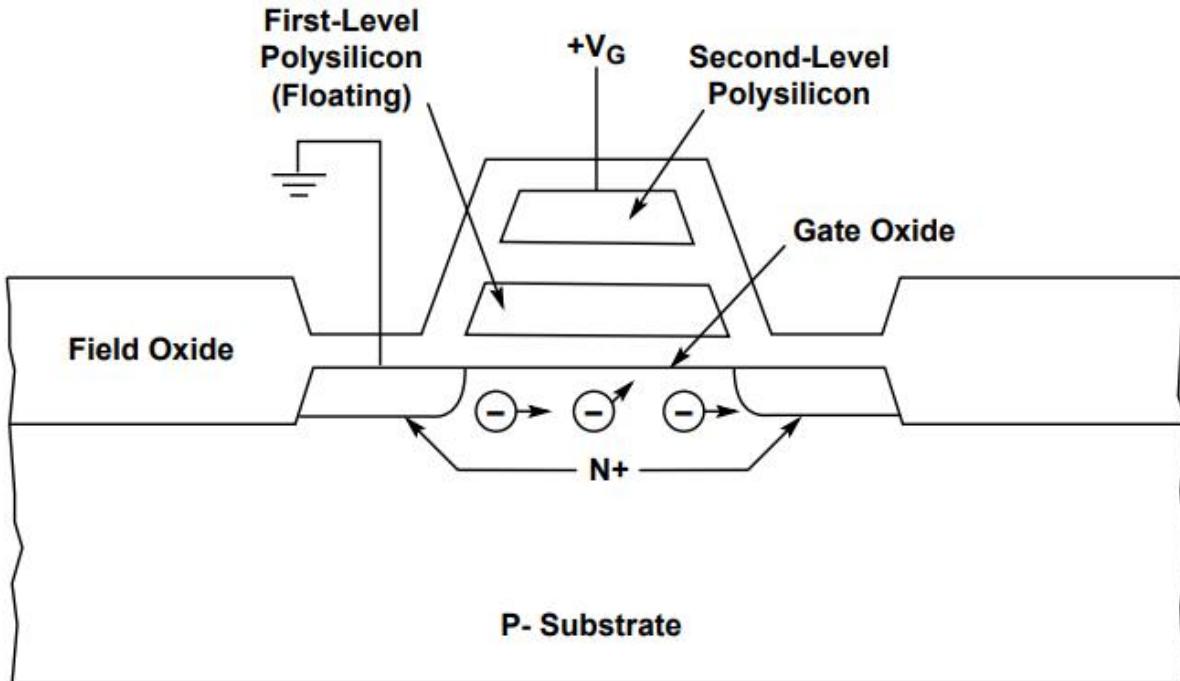




EPROM

- Erasable Programmable Read Only Memory.
- Data can be erased by using UV light.
- Switch is disabled by injecting charge on the gate using high voltage between gate and drain.
- The charge is removed UV light. Reprogramming through exposure through UV light.
- Non-volatile.
- Slower programming than RAM.





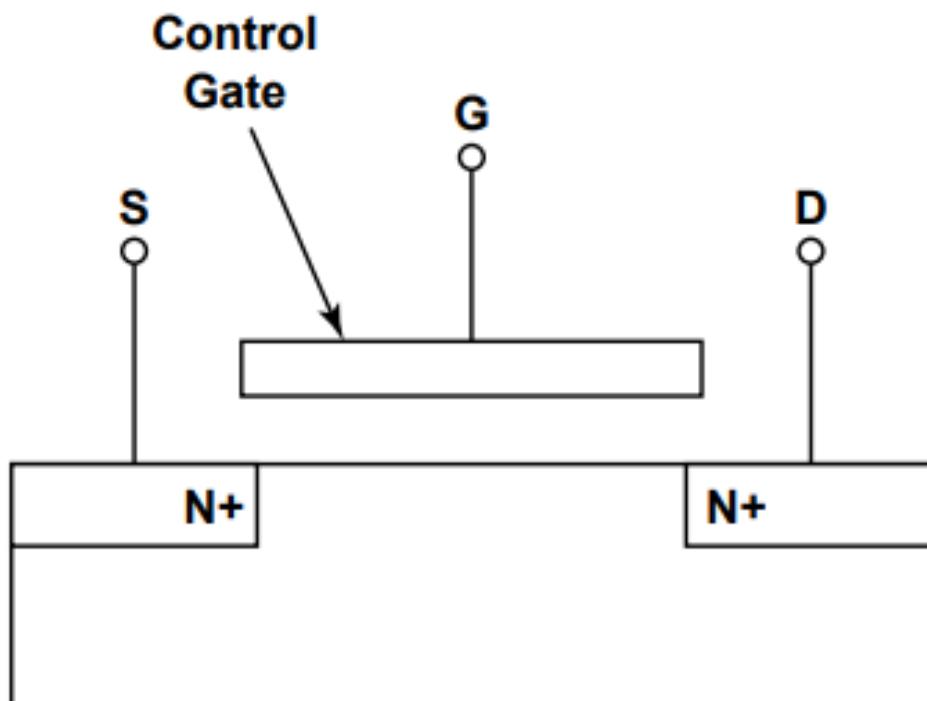
Source: Intel/ICE, "Memory 1997"

18474

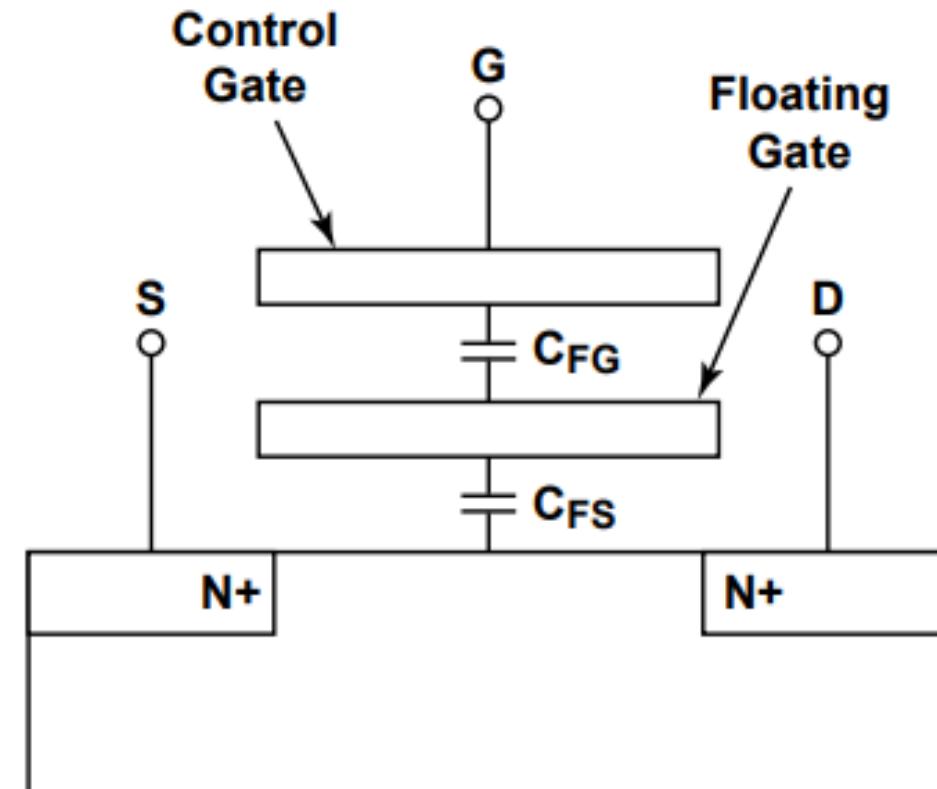
Programming by hot electron injection

- High voltages are applied to the select gate and drain connections of the cell transistor.
- The select gate of the transistor is pulsed “on” causing a large drain current to flow.
- The large bias voltage on the gate connection attracts electrons that penetrate the thin gate oxide and are stored on the floating gate.

1. The EPROM device is programmed by forcing an electrical charge on a small piece of polysilicon material (called the floating gate) located in the memory cell.
2. When this charge is present on this gate, the cell is “programmed,” usually a logic “0,” and when this charge is not present, it is a logic “1.”



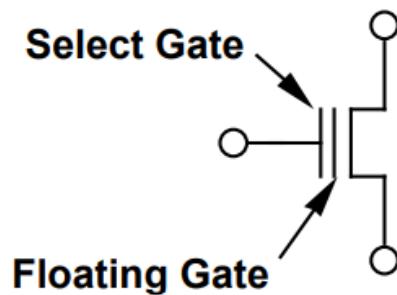
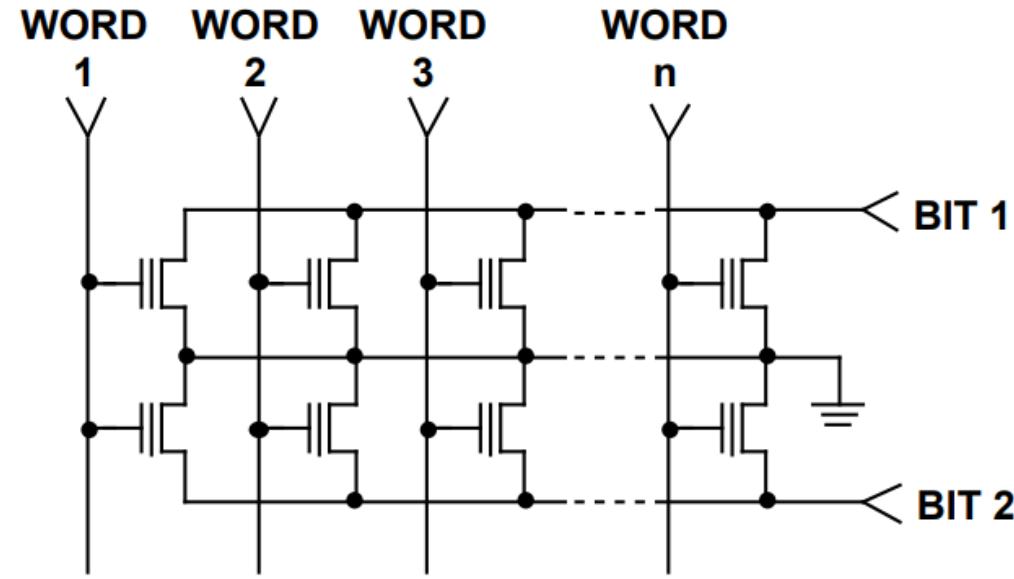
(a) Conventional MOS



(b) Floating-Gate MOS

Source: ICE, "Memory 1997"

20846





=

Advantage:

- No external permanent memory is needed to program it at power up condition.

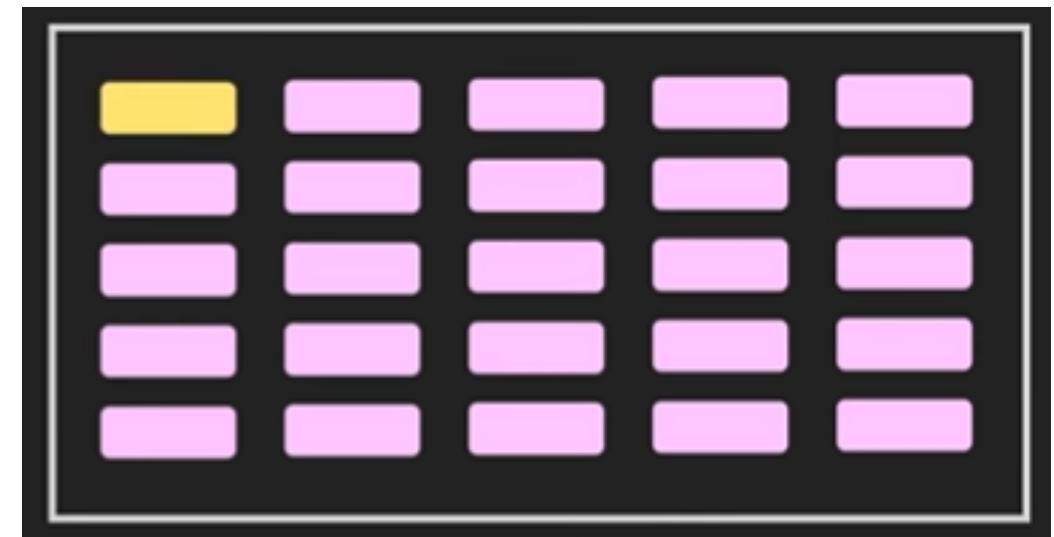
Disadvantages:

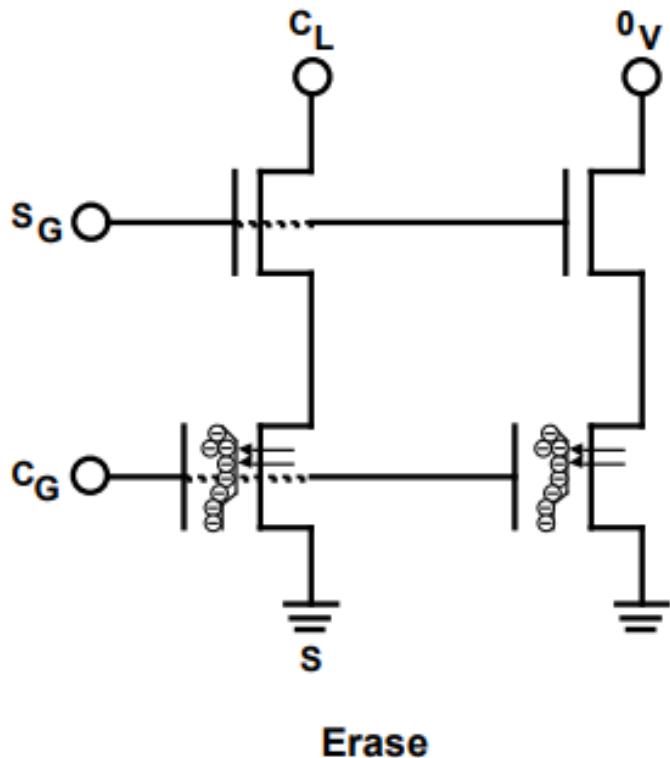
- Extra processing steps are needed.
- Static power loss due to pull up resistor which is a high resistance.



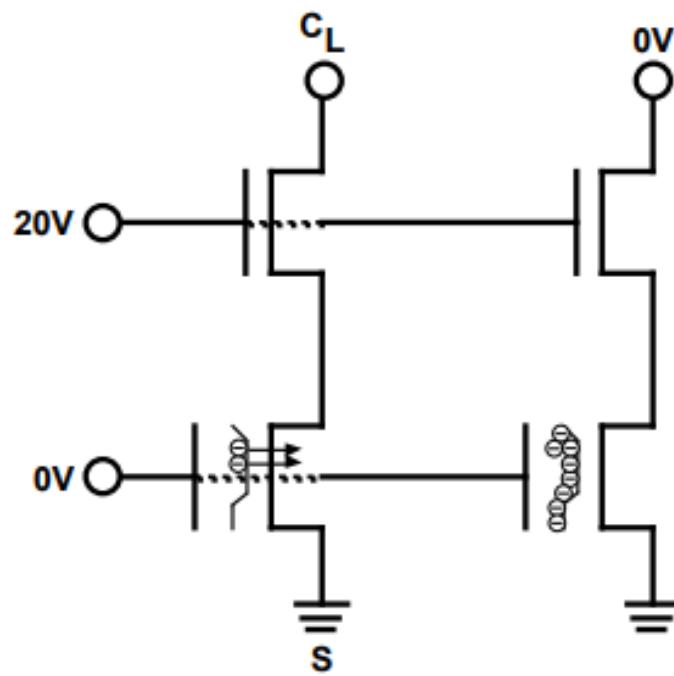
EEPROM

- Electrically Erasable Programmable Read Only Memory.
- Data can be erased electrically.
- Data can be erased multiple times.
- Write cycle-10k to 1L.
- Size- few KBs (stores small data)
- It allows byte level write and erase operations.





Erase



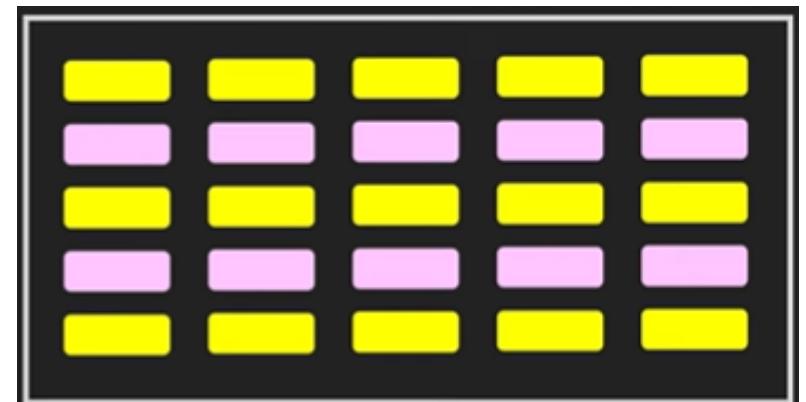
Program

	s_G	c_L	c_G	S
Erase	V_{PP}	0	V_{PP}	0
Program	V_{PP}	V_{PP}	0	0



Flash memory

- Similar to EEPROM.
- Data can be erased electrically.
- Supports block level erase and write of data.
- Data can be erased multiple times.
- Write cycle-10k per block..
- Size- MBs to GBs.
- Used in storage devices like USB, SSD etc.



NAND Flash



Good Write Speed but slower Read Speed

NOR Flash



Slower Write Speed but Faster Read Speed



ROM applications

It is used to store the permanent instructions and data which is critical for device operation

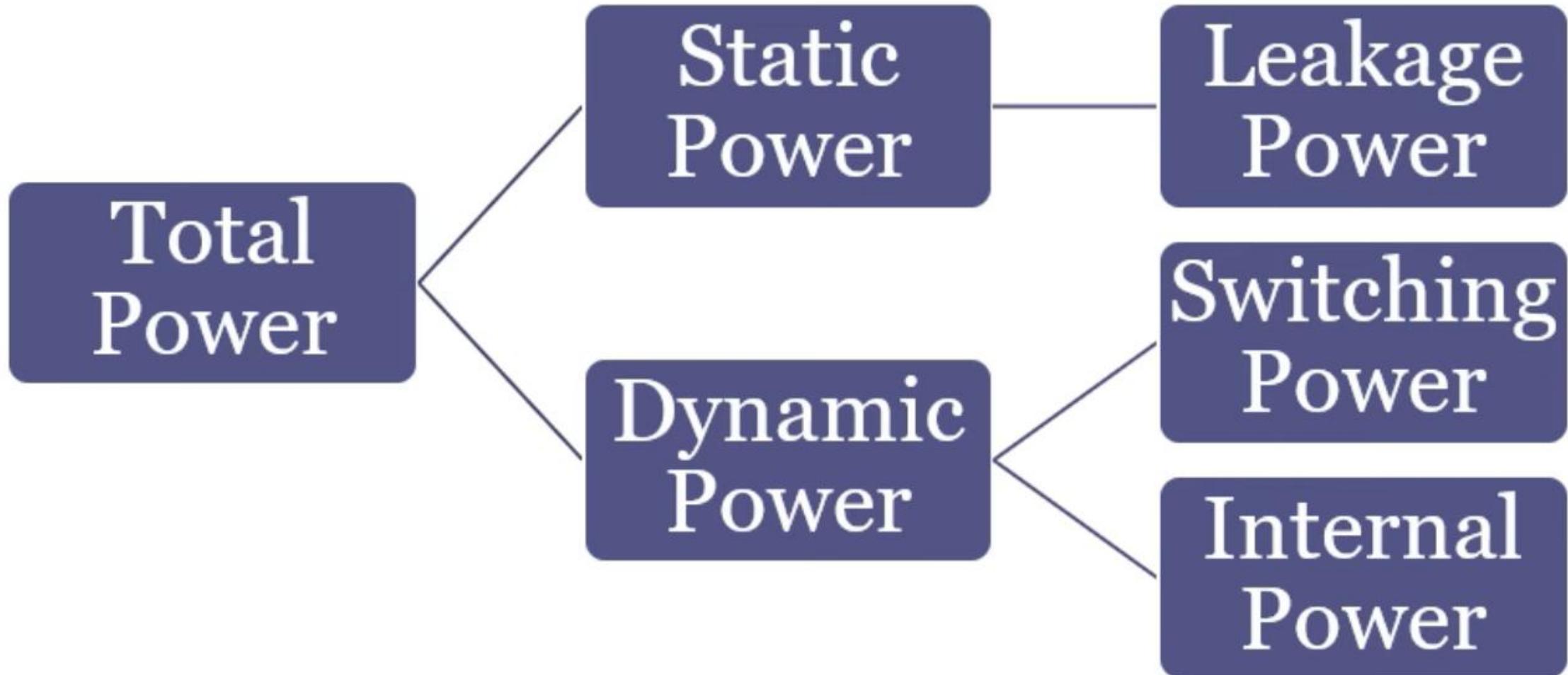
In embedded systems to store the configuration parameters, calibration settings etc.

BIOS –Basic Input Output System

Printers, keyboards, and other peripherals often use ROM to store instructions for basic functionality and communication with the main system.



Power dissipation in CMOS circuits





1. Static Power dissipation

Static Power is the power dissipated by a transistor when it is not switching, i.e. when it is static.

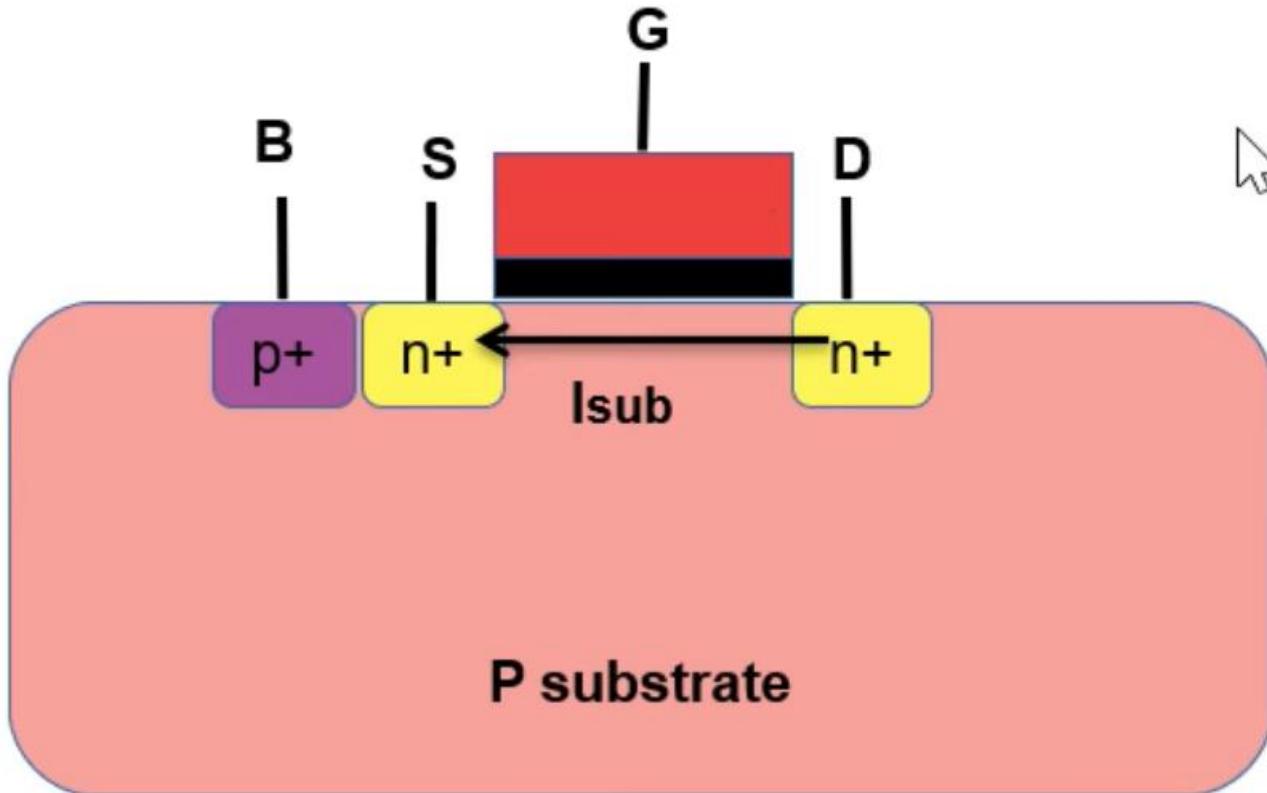
Occurs mainly due to leakage.

1. Subthreshold leakage
2. Gate leakage
3. Junction leakage

Also known as leakage power.



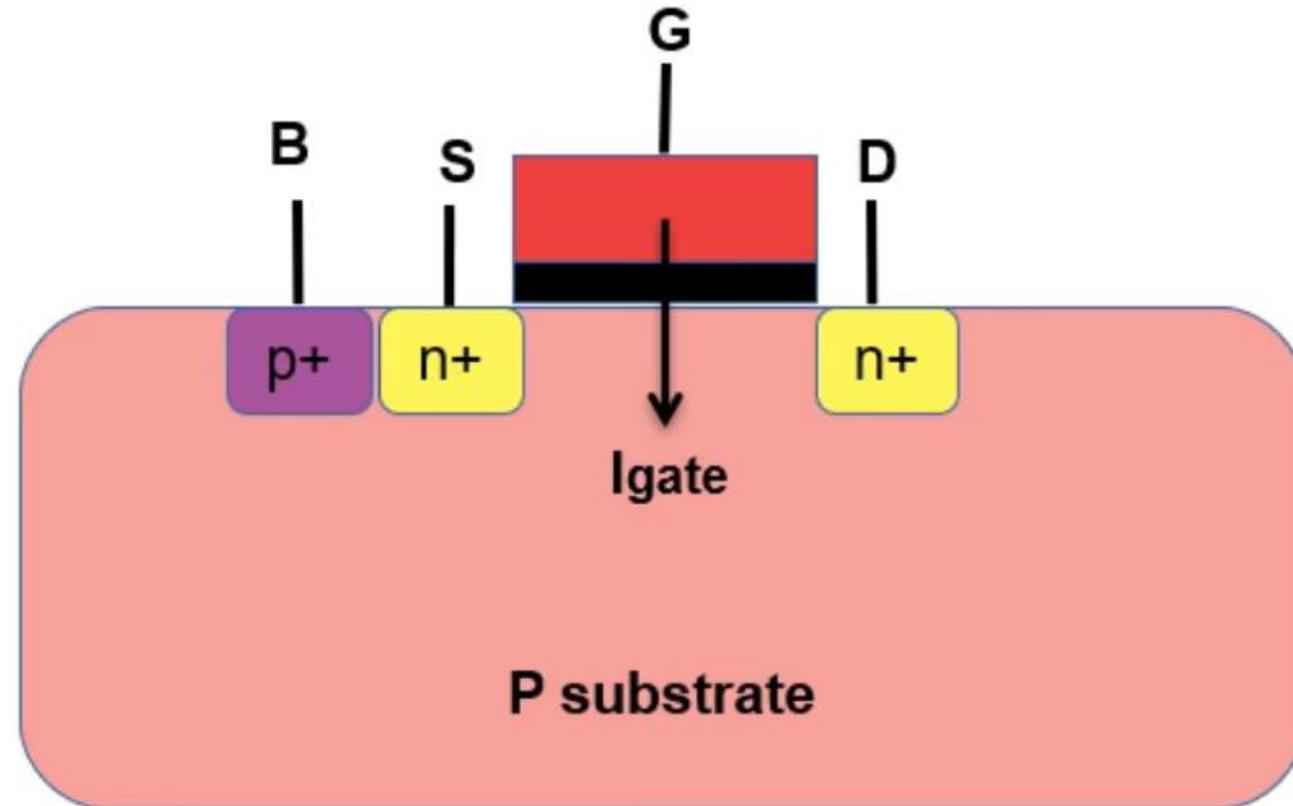
Subthreshold Leakage



Sub-threshold Leakage Current is the current which flows between drain and source when $V_{gs} < V_t$ (for nmos) or $V_{sg} < |V_t|$ (pmos)



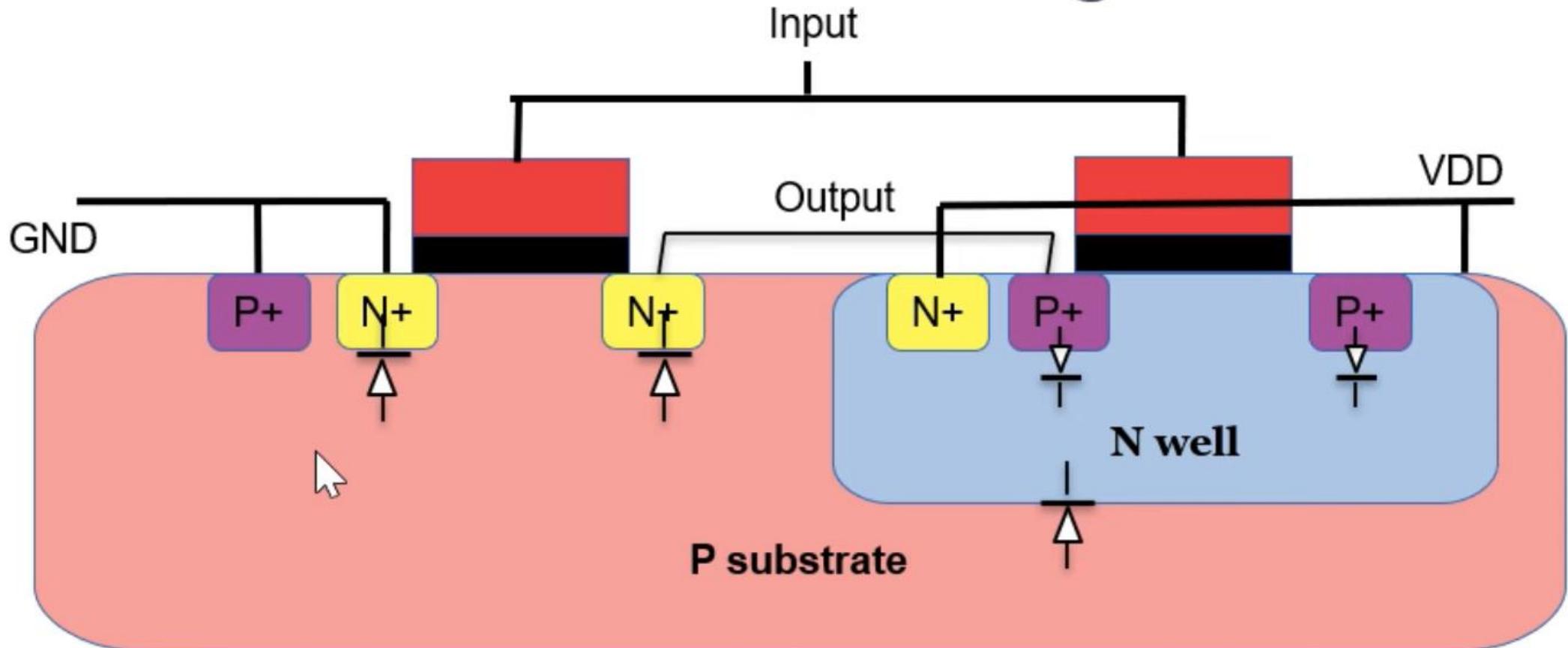
Gate Leakage



Leakage through gate is observed when gate oxide is very thin and it allows current to pass through it.



Junction Leakage



P-N junctions in a transistor form diodes, these diodes are reverse biased. However, reversed biased diodes still conduct a small amount of current , which causes junction leakage current.



DYNAMIC POWER

Dynamic Power is the power which is dissipated when the circuit is active, and a circuit is active anytime voltage on input net changes due to some stimulus applied.

Note: Change in the voltage of input net may or may not lead to change in the logic state of output. But in both the cases dynamic power will be dissipated.



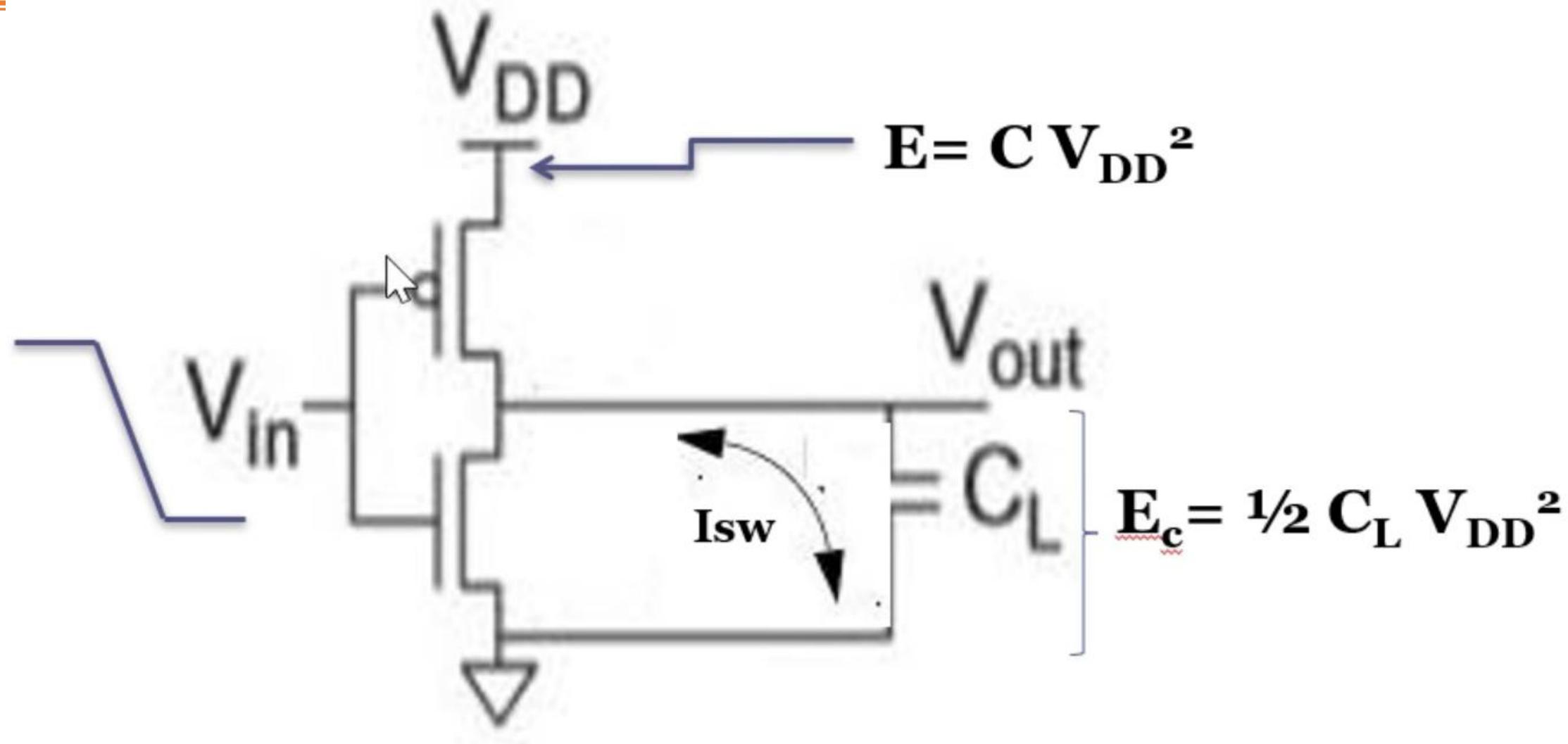


Switching Power

Switching power is the power dissipated while charging and discharging the load capacitor at the output of the cell.

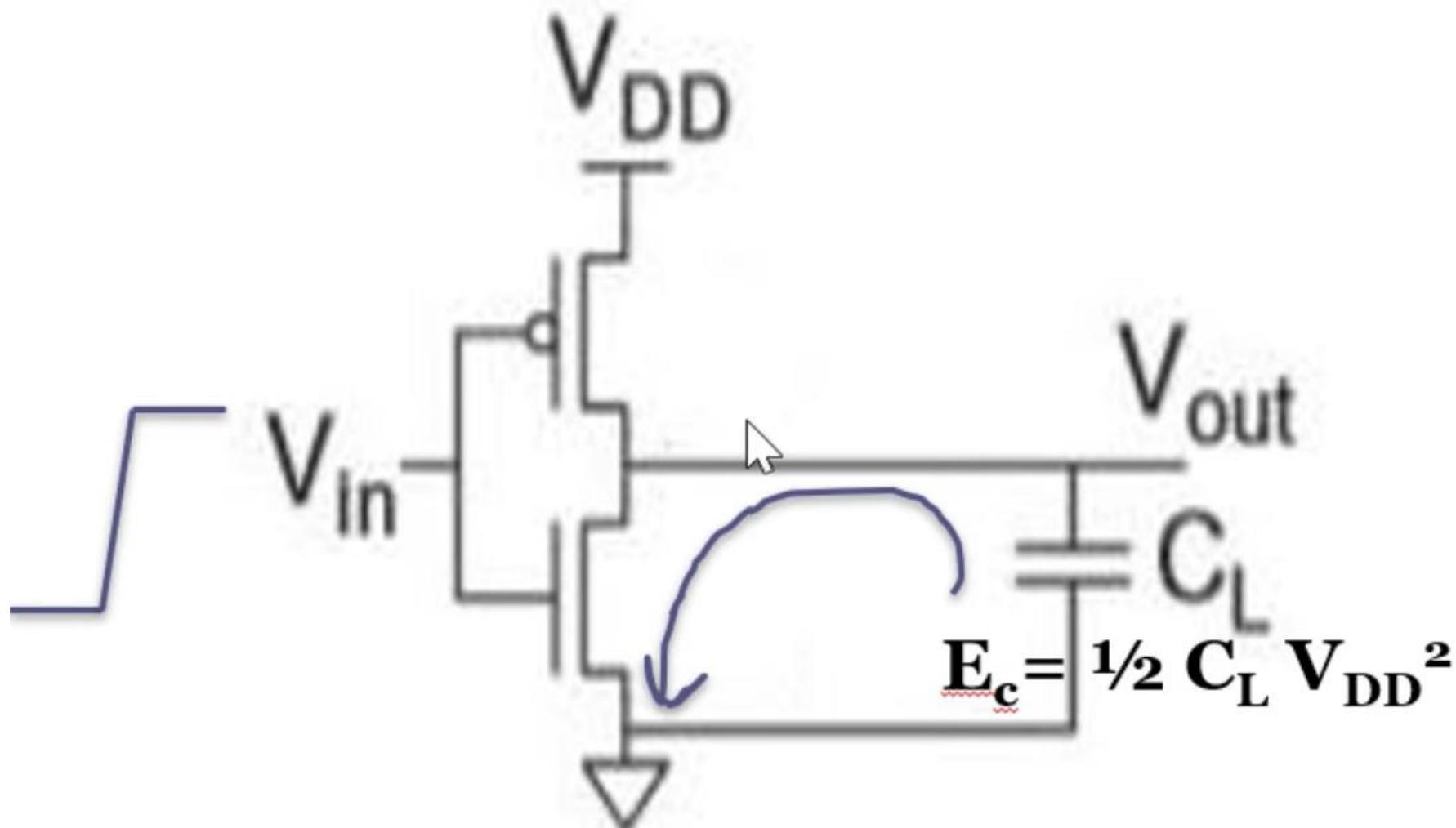


When input switches from 1 to 0





When input switches from 0 back to 1





Switching Power Equation

When the input switches from 1 to 0, the pMOS transistor turns ON and charges the load to VDD.

At that time, the energy delivered by the power supply is :

$$E = C V_{DD}^2$$

Now, the energy stored in the capacitor, i.e. the energy used to charge the capacitor

$$\underline{E_c} = \frac{1}{2} C_L V_{DD}^2$$

It is to be noted here that only half of the energy is used to charge the load capacitor and the other half is dissipated as heat in the pmos transistor.

Now, When the input switches from 0 back to 1, the pMOS transistor turns OFF and the nMOS transistor turns ON, discharging the capacitor.



Switching Power Equation

Now, if the input switches ' α ' times in a cycle, and if the input transition causes output transition as well, we would say that switching activity factor is ' α '.

Thus, $P_{\text{switching}} = \frac{1}{2} \alpha C_L V_{DD}^2 f$,

Where, f = frequency

C_L = Load Capacitance

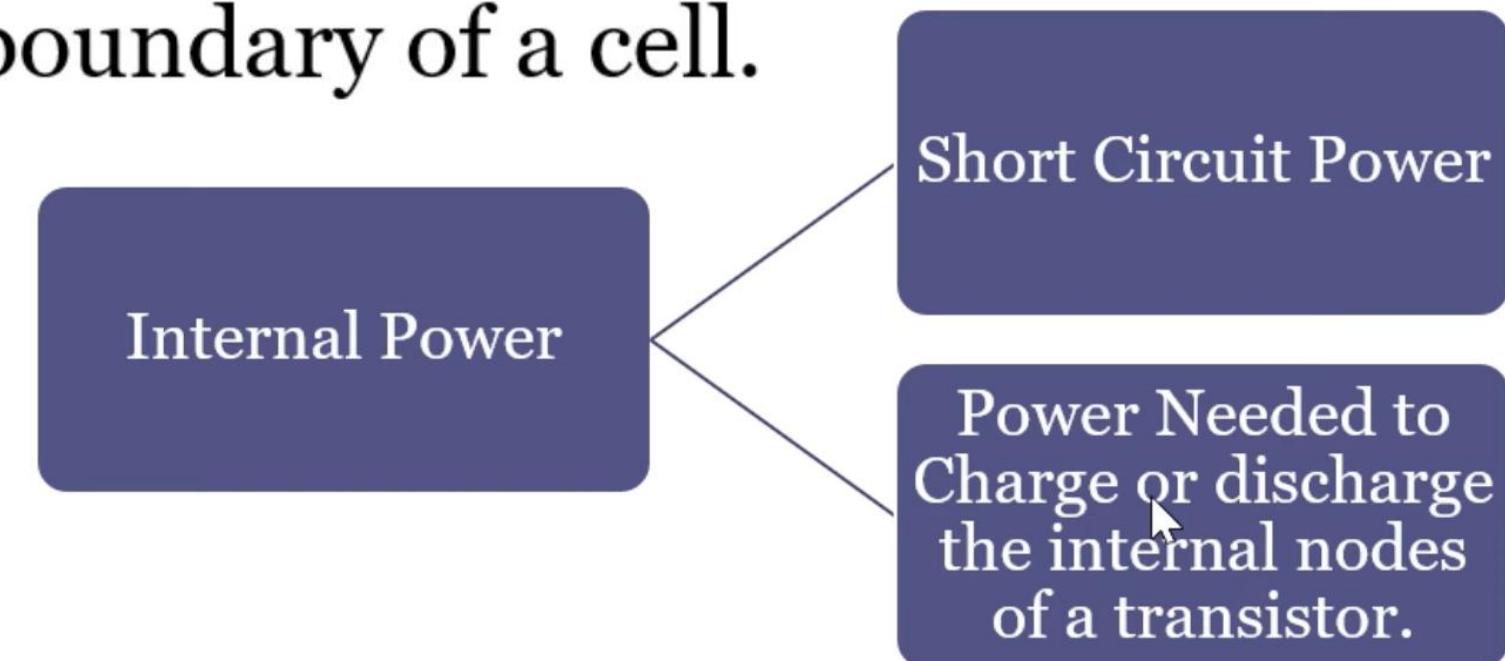
V_{DD} = Supply voltage

Example : Switching Activity Factor for clock (α) = 2



Internal Power

Internal power is any power which is dissipated within the boundary of a cell.





=

Short Circuit Power

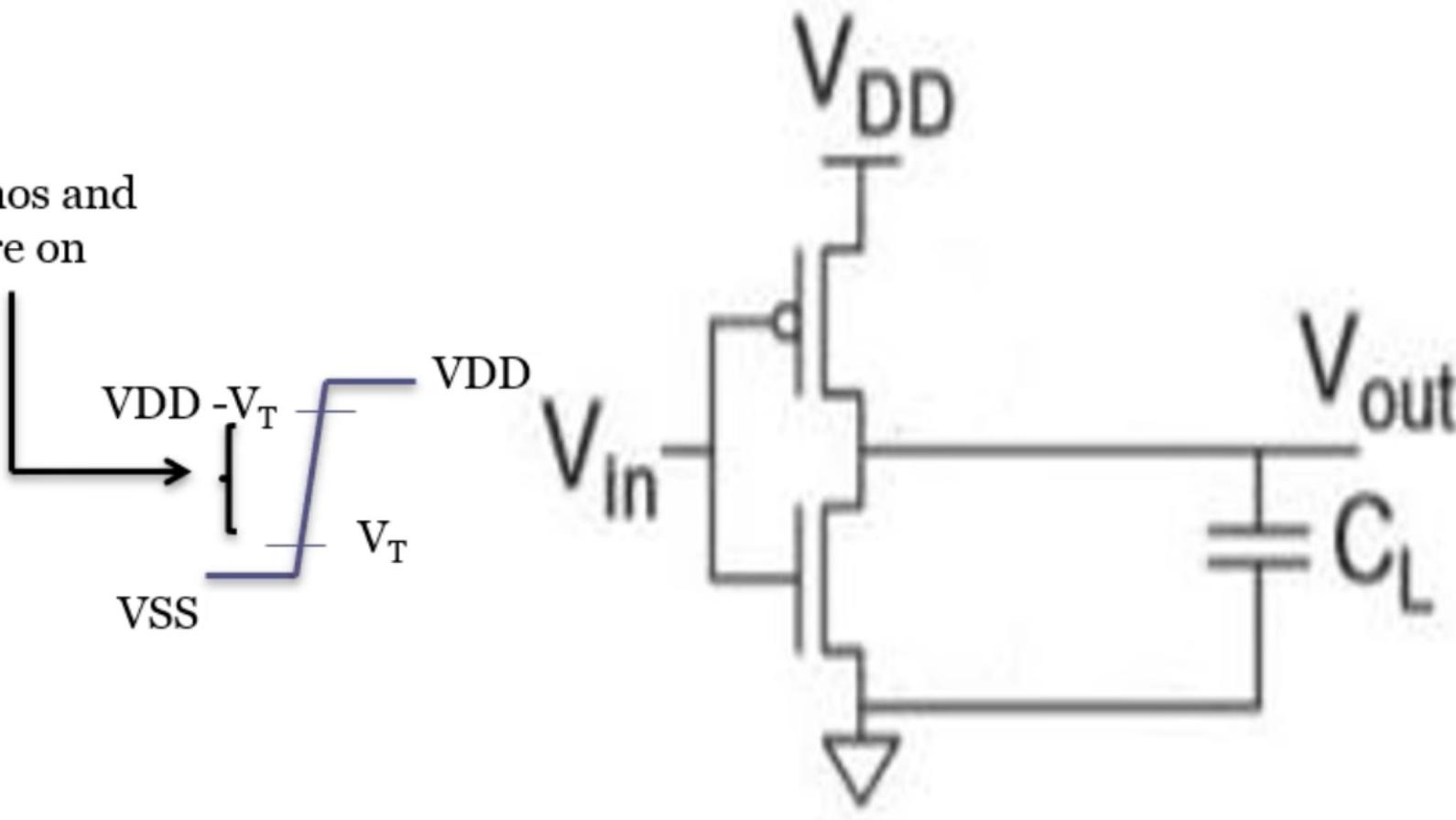
Short Circuit power is the power dissipated when both pmos and nmos are on simultaneously during input transition.





Short Circuit Power

Both Pmos and
Nmos are on



Pmos is on when : $V_{SS} \leq V_i \leq V_{DD} - V_T$

Nmos is on when : $V_T \leq V_i \leq V_{DD}$

Therefore, both Pmos and Nmos are on for : $V_T \leq V_i \leq V_{DD} - V_T$



THANK YOU



ALL THE BEST