MOS and CMOS Circuit Design Process:

MOS and CMOS circuit design process involves the concepts such as:

- MOS Layers
- Stick Diagrams
- Lambda based design rules and layout diagrams
- Basic circuit concepts such as: sheet resistance, area capacitance and delay calculation

MOS Layers:

MOS circuits are formed by three layers i.e. diffusion (n or p diffusion layer), polysilicon and metal, which are isolated from one another by thick or thin (thinox) silicon dioxide insulating layers.

- The thin oxide region includes n- diffusion, p- diffusion and transistor channels. Polysilicon and thinox regions interact so that a transistor is formed where they cross one another.
- Layers may be deliberately joined together where contacts are formed.
- The basic MOS transistor properties can be modified by the use of an implant within the thinox region.

The MOS design is aimed at turning a specification into masks for processing silicon to meet the specification.

Stick Diagrams and Layout Diagrams:

Stick diagrams are used to convey layer information and topology through the use of color code and using these stick diagrams mask layouts can be easily designed. The color code for various layers are:

1. Green for n- diffusion
2. Red for polysilicon
3. Blue for metal
4. Yellow for implant or for p- diffusion
5. Black for contact areas

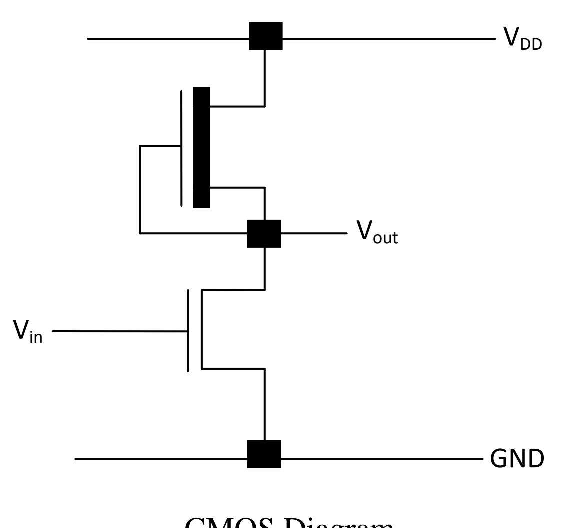
- The layout of stick diagrams faithfully reflects the topology of the actual layout in silicon and the stick diagrams are relatively easily turned into mask layouts.
- As known that the mask layout produced during design will be compatible with the fabrication processes, a set of design rules are set out for layouts so that, if obeyed, the rules will produce layouts which will work in practice.

Mask Layout/ Layout/ Layout Diagram represent an integrated circuit in terms of planar geometric shapes which corresponds to the pattern of the metal, oxide or semiconductor layers that make up the components of the integrated circuit. The dimensions of each layer and the separation between the layers in a layout are parameterized by λ.

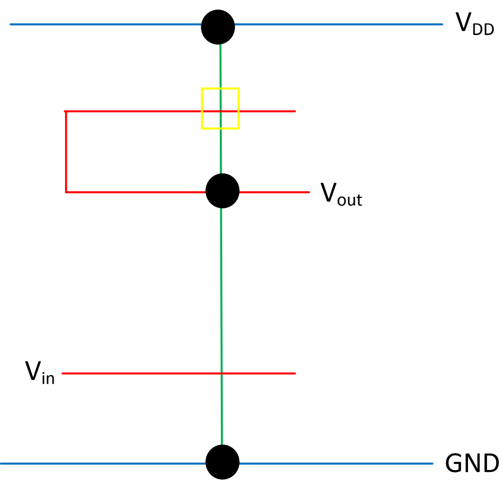
Basic Encoding Concepts for Drawing stick diagrams and mask layout/ Layout Diagram:

Layers and their color	Stick diagram encoding	Mask layout encoding	Monochrome stick encoding	Monochrome mask encoding
n diffusion color- green		Green		
p diffusion color- yellow		Yellow		
Polysilicon color- red		Red		
Implant color- yellow		Yellow		
Metal 1 color- blue		Blue		
Metal 2 color- dark blue		Dark Blue		
contact cut (including buried) color- black		Unburied contact cut Buried contact cut (brown color)		
V_{DD} or V_{SS} contact cut color- black	X			
via cut color- black				
Demarcation line/ pwell color- brown				
nMOS- enhancement mode	G D S			
pMOS- enhancement mode	G S D			
nMOS- depletion mode	G D S			
pMOS- depletion mode	G S D			

1. nMOS inverter:



CMOS Diagram



Stick Diagram

V_{DD}

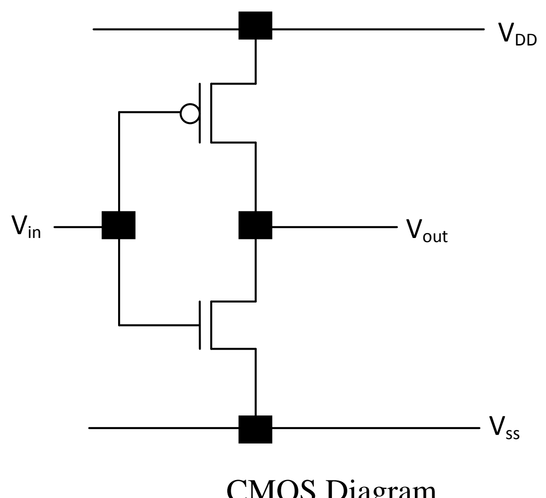
V_{out}

V_{in}

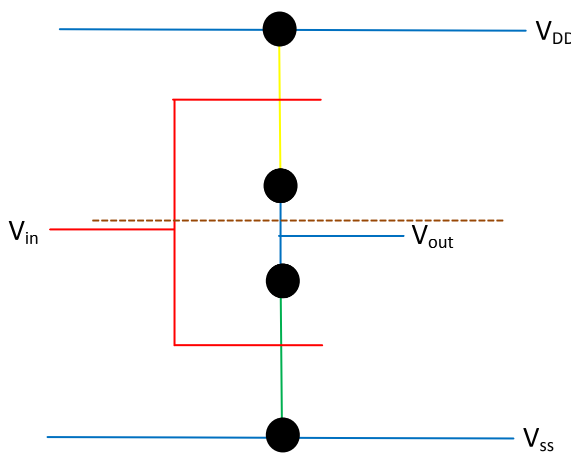
GND

Mask Layout/ Layout Diagram

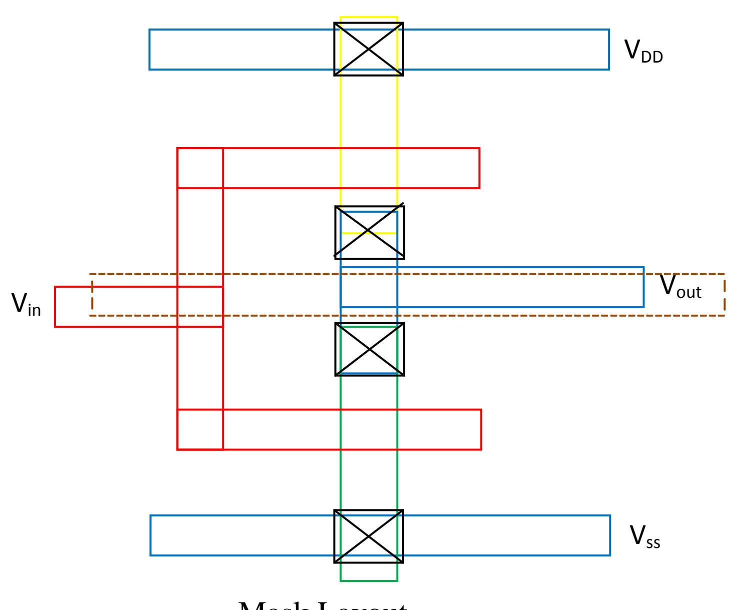
1. CMOS Inverter:



CMOS Diagram



Stick Diagram



Mask Layout

Encodings			
Layers	**Colour**	**Stick diagram**	**Layout diagram**
Metal 1	Blue		
n diffusion	Green		
p diffusion	Yellow		
Polysilicon	Red		
Implant	Yellow		
Contact	Black		

Design Rules:

Design rules provide an effective interface between the circuit/ system designer and the fabrication engineer.

<u>Lambda Based Design rules and layout diagrams</u>:

Lambda based design rules are based on a single parameter lambda λ which leads to a simple set of rules for the designer, providing a process and feature size independent way of setting out mask dimensions to scale. These rules specify line widths, separations, and extensions in terms of λ, and are readily committed to memory.

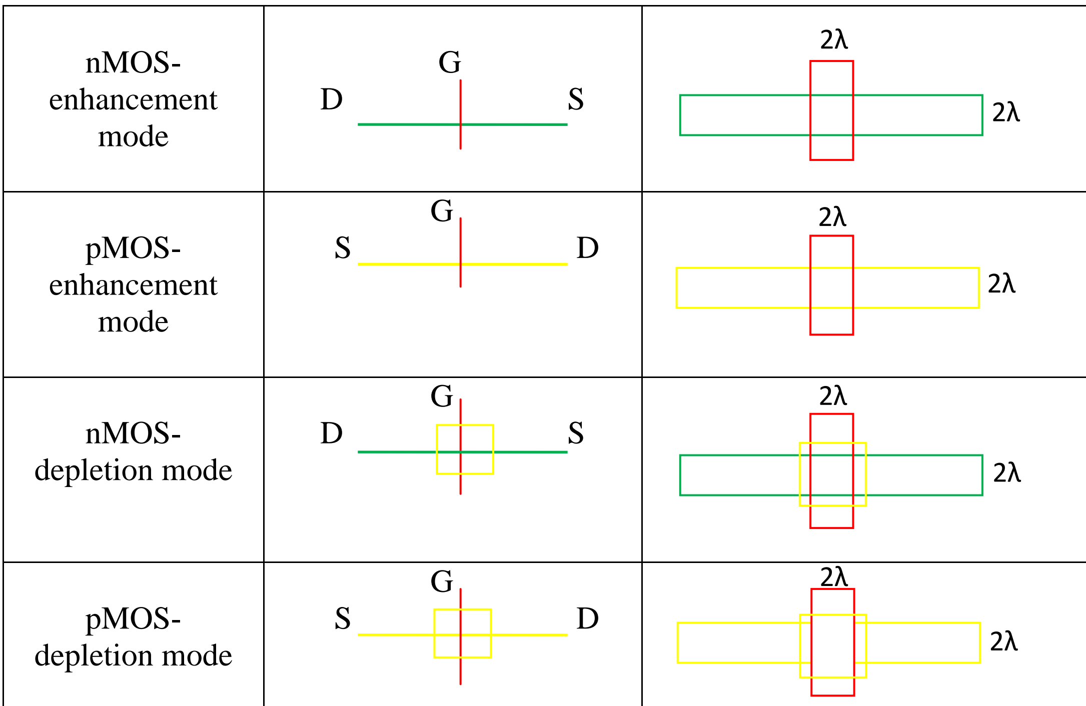
All paths in all the layers will be dimensioned in λ units and subsequently λ can be allocated to an appropriate value compatible with the feature size of the fabrication process i.e. if mask layout obey these rules correctly in the layout, then the mask layout will produce working circuits for a range of values allocated to λ.

Contacts between polysilicon and diffusion in nMOS/ MOS circuits are possible by two approaches:

1. Butting Contact
2. Buried Contact

The latter is generally less space- consuming and is held by many to be the more reliable contact. Therefore consultation to the fabrication work where the designs are to be turned into silicon should be prior.

The **layout diagrams** are drawn on squared paper (say 5mm) where the side of each square is taken to represent λ. The layout diagrams use the design rules using contacts such as butting contact where departures from strict adherence to the rules can take place.



Illustrating λ based rule, using 2λ specification as width for the polysilicon and thinox layers

Lambda based design rules for wires (nMOS and CMOS):

Layer width:

Layer	Minimum Width
n- diffusion	2λ
p- diffusion	2λ
Polysilicon	2λ
Metal 1	3λ
Metal 2	4λ

Separation between the layers:

Layers	Minimum Separation
n- diffusion and n- diffusion	3λ
p- diffusion and p- diffusion	3λ
Polysilicon and polysilicon	2λ
n-diffusion and polysilicon	1λ
p-diffusion and polysilicon	1λ
Metal 1 and metal 1	3λ
Metal 2 and metal 2	4λ

Basic circuit concepts:

Basic circuit concepts help us to calculate the actual resistance, capacitance, delay values associated with the transistors and their circuit wiring and parasitic.

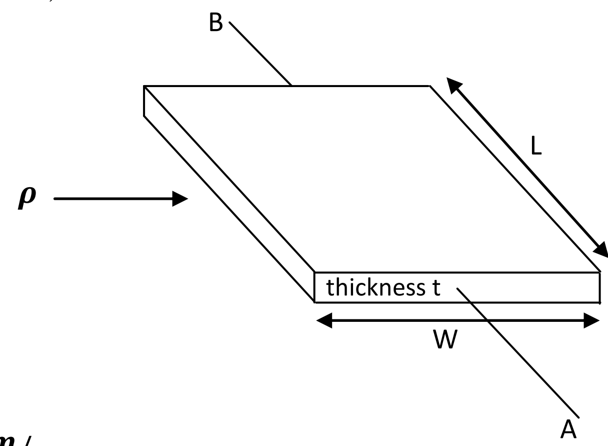
Sheet Resistance R_s:

Sheet resistance is defined as the ratio of resistivity ρ and thickness t for a sheet/ slab.

Consider a uniform slab of conducting material ρ of width W, thickness t, and length L between the faces A and B, then the value of resistance of the slab (sheet) is given as,

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

$$R_{AB} = \frac{\rho L}{t\,W}\ ohm$$



Where:

A = t W = area of cross section of the slab

If L = W, i.e. square of resistive material, then

$$R_{AB} = \frac{\rho}{t} = R_s\ in\ {}^{ohm}/_{square}$$

Where:

R_s = sheet resistance or ohm per square

For a MOSFET transistor

$$R = ZR_s = \frac{L}{W} R_s = 4 \times 10^4 ohm$$

Where:

Z = L/ W

- It is to be noted that R_s is completely independent of the area of the square.

The typical sheet resistances R_s for various MOS layers are (considering different technologies)

Layer	Sheet Resistance R_s		
	5 µm Technology	2 µm Technology	1.2 µm Technology
Metal	0.03	0.04	0.04
n- channel transistor/ pMOS transistor	1×10^4	2×10^4	2×10^4
p- channel transistor/ pMOS transistor	2.5×10^4	4.5×10^4	4.5×10^4
Diffusion	10- 50	10- 50	10- 50
Silicide	2- 4	2- 4	2- 4
Polysilicon	15- 100	15- 100	15- 100

Area Capacitance:

In MOS transistor conducting layers are separated from the substrate and each other by insulating (dielectric) layers, and thus parallel plate capacitive effects are present and are allowed.

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

$$C = \frac{\varepsilon_0 \, \varepsilon_{ins} \, A}{D} = \frac{k \, A}{D} \, farads$$

Where:

D = thickness of silicon dioxide

k = dielectric constant

A = Area of plates

$\varepsilon_{ins} = relative\ permittivity\ of\ silicon\ dioxide$

$\varepsilon_0 = relative\ permittivity\ of\ free\ space = 8.85 \times 10^{-14} F/\,cm$

- Normally area capacitances are given in pF/ μm^2 (where µm = micron = 10^{-6} meter = 10^{-4} cm). The appropriate figure may be calculated as:

$$C\left(\frac{pF}{\mu m^2}\right) = \frac{\varepsilon_0\ \varepsilon_{ins}}{D}\frac{F}{cm^2} \times \frac{10^{12}pF}{F} \times \frac{cm^2}{10^8\mu m^2}$$

The typical area capacitance values for 5µm MOS circuits are:

Capacitance	Value in pF/ µm²	Relative value
Gate to Channel	4×10^{-4}	1
Diffusion to substrate	1×10^{-4}	0.25
Polysilicon to substrate	0.4×10^{-4}	0.1
Metal 1 to substrate	0.3×10^{-4}	0.075
Metal 2 to substrate	0.2×10^{-4}	0.05
Metal 2 to metal 1	0.4×10^{-4}	0.1
Metal 2 to polysilicon	0.3×10^{-4}	0.075

Note: Relative value = Specified value / gate to channel value

- **Standard unit of capacitance "$\square C_g$":**

The standard unit of capacitance is denoted by $\square C_g$ and is defined as the gate- to- channel capacitance of the minimum size (2λ × 2λ) MOS transistor.

- The standard unit of capacitance has provided a convenience to various MOS technologies but which can be used in calculations without associating it with an absolute value.
- $\square C_g$ can be evaluated for any MOS technology.

For example for a 5 µm MOS circuit with λ = 2.5 µm:

Gate area = 5 µm × 5 µm = 25 μm^2

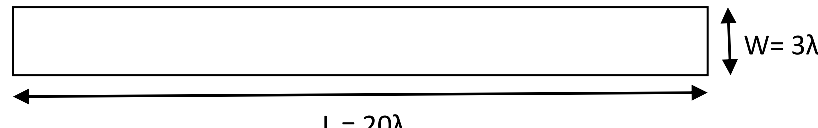
Capacitance value = 4×10^{-4} pF/ μm^2 (using table)

Standard Capacitance $\square C_g$= 25 μm^2 × 4×10^{-4} pF/ μm^2 = .01 pF

- **Some Area Capacitance calculations:**

Here the calculation of capacitance values may now be done by the ratio between the area of interest and the area of the standard gate (2λ × 2λ) and multiplying this ratio by the appropriate relative C value (using the table). The product will give the required capacitance in $\square C_g$ units.

Let's calculate the capacitance of a simple area of length 20λ and width 3λ respectively.



Now we will calculate:

1. Relative Area

$$Relative\ Area = \frac{L \times W}{2\lambda \times 2\lambda} = \frac{20\lambda \times 3\lambda}{2\lambda \times 2\lambda} = 15$$

2. Capacitance to substrate considering the area in metal.

$$Capacitance\ to\ substrate = relative\ area \times relative\ C\ value\ (from\ table)$$
$$Capacitance\ to\ substrate = 15 \times 0.075 \square C_g$$

3. Capacitance to substrate considering the area in polysilicon.

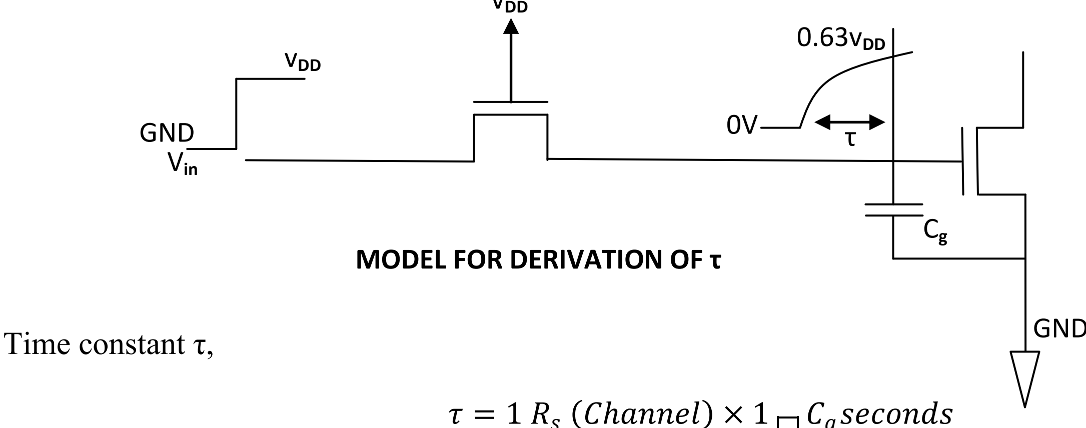
$$Capacitance\ to\ substrate = relative\ area \times relative\ C\ value\ (from\ table)$$
$$Capacitance\ to\ substrate = 15 \times 0.1 \square C_g = 1.5 \square C_g$$

4. Capacitance to substrate considering the area in diffusion.

$$Capacitance\ to\ substrate = relative\ area \times relative\ C\ value\ (from\ table)$$
$$Capacitance\ to\ substrate = 15 \times 0.25 \square C_g = 3.75 \square C_g$$

Delay Calculation/ The delay unit (τ):

Considering the case of one standard gate capacitance being charged through one square of channel resistance (from 2λ by 2λ nMOS pass transistor).



MODEL FOR DERIVATION OF τ

Time constant τ,

$$\tau = 1\ R_s\ (Channel) \times 1 \square C_g seconds$$

The time constant given as above can be evaluated for 5 μm technology so that,

$$Theoretical\ \tau = 10^4 ohm \times 0.01 pF = 0.1\ nsec$$

In practice there are circuit wiring and parasitic capacitances, so τ is increased by a factor 2 or 3 so that for a 5 μm circuit (λ = 2.5 μm),

$$\tau = 0.2 \rightarrow 0.3\ nsec\ is\ typical\ figure$$

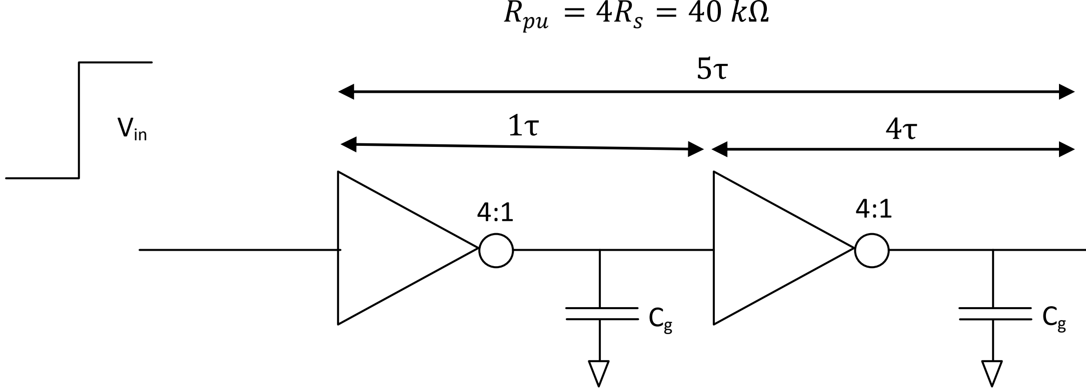
It is to be noted that τ thus obtained is not much different from transit time τ_{sd}, which is given as,

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

Inverter Delays:

Considering a basic 4:1 ratio nMOS inverter in order to achieve the 4:1 Z_{pu} to Z_{pd} ratio, R_{pu} will be 4 R_{pd}, and if R_{pd} is contributed by the minimum size transistor then, clearly, the resistance value associated with R_{pu} is such,

$$R_{pu} = 4R_s = 40\ k\Omega$$



- The R_{pd} value is 1 R_s = 10 kΩ so that the delay associated with the inverter will depend on whether it is being turned on or off and if considering the pair of cascaded inverters, then delay over the pair will be constant irrespective of the sense of the logic level transition of the input to the first. (Assuming $\tau = 0.3$ nsec and making no extra allowances fro wiring capacitance). We have an overall delay of $\tau + 4\tau = 5\tau$.

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

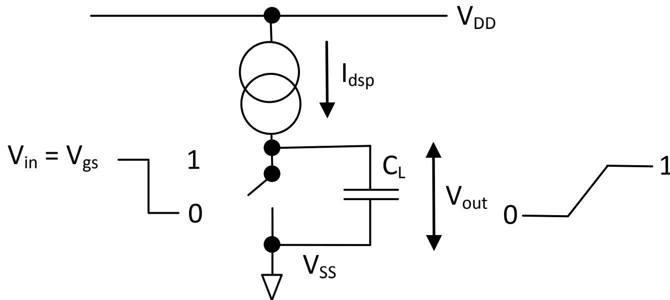
$$T_d = \left(1 + \frac{Z_{pu}}{Z_{pd}}\right)\tau$$

Thus, the inverter pair delay for inverters having 4:1 ratio is 5τ (which should be multiplied by a suitable factor to allow for wiring).

➢ 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 analaysis:

1. Rise time estimation:
   Here, 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 modelled as shown.



The saturation current for the p- transistor is given as,

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

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

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

Substituting $I_{dsp} = \frac{\beta_p\left(V_{gs} - |V_{tp}|\right)^2}{2}$ in $V_{out} = \frac{I_{dsp}\, t}{C_L}$, we get

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

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

Assuming that t = τ_r when $V_{out} = +V_{DD}$, so that

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

With $|V_{tp}| = 0.2\, V_{DD},\ then$

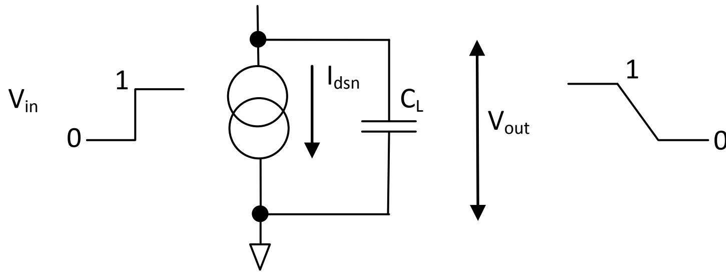
$$\boldsymbol{\tau_r = \frac{3\, C_L}{\beta_p\, V_{DD}}}$$

Algebraically,

$$\boldsymbol{\tau_r = 2.2\tau_p}$$

Therefore, the charging of C_L is divided more correctly into two parts i.e. saturation and the resistive region of the transistor.

2. Fall- time estimation:



Similar reasoning can be applied for the discharge of C_L through the p- transistor. Therefore, Similarly, we can write,

$$\tau_f = \frac{3\, C_L}{\beta_n\, V_{DD}}$$

Algebraically,

$$\tau_f = 2.2\tau_n$$

Therefore, we can summarize the inverter delay as:

$$\frac{\tau_r}{\tau_f} = \frac{\frac{3\, C_L}{\beta_p\, V_{DD}}}{\frac{3\, C_L}{\beta_n\, V_{DD}}} = \frac{\beta_n}{\beta_p}$$

3. Propagation Delay/ propagation time estimation:
   The propagation delay time $\tau_{propagation}$ is often used to estimate the 'reaction' delay time from input to output. When we use step- like input voltages, the propagation delay is defined by the simple average of two time- intervals.

$$\tau_{propagation} = 0.35\,(\tau_n + \tau_p)$$

Factors which affect rise and fall times:

1. τ_r and τ_f are proportional to $\frac{1}{V_{DD}}$.
2. τ_r and τ_f are proportional to C_L.
3. $\tau_r = \tau_f$ for equal n and p transistor geometries.

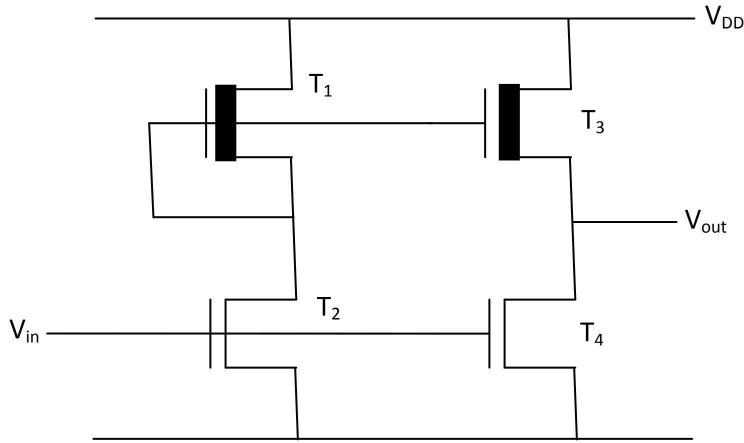
Super buffer:

A super buffer is a common alleviative approach for undesirable rise of delay problems of an conventional inverter/ inverter when it is used to drive more significant capacitive loads.

There are two types of super buffers:

1. Inverting type of super buffer
2. Non inverting type of super buffer

Inverting type super buffer (nMOS):

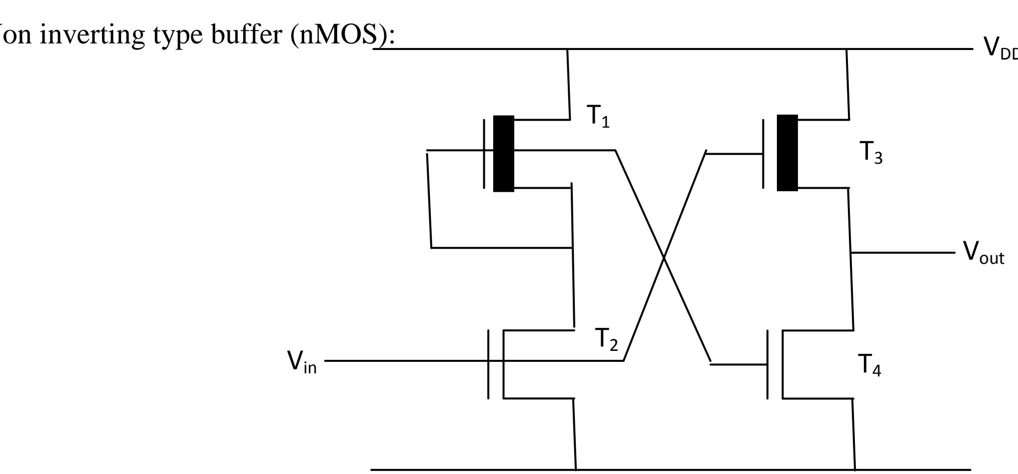


The inverting type as shown above is considered with a positive going logic transition V_{in} at the input, it is seen that the inverter formed by T_1 and T_2 is turned ON and thus the gate T_3 is pulled down toward 0V with a small delay. Thus T_3 is cut off while T_4 (the gate of which is also connected to V_{in}) is turned ON and the output is pulled down quickly.

Now considering the opposite transition, when V_{in} drops to 0V then the gate of T_3 is allowed to rise quickly to V_{DD}. Thus as T_4 is also turned OFF by V_{in}, T_3 is caused to conduct with V_{DD} on its gate, that is, with twice the average voltage which would apply if the gate was tied to the source as in the conventional inverter.

Since I_{ds} is directly proportional to V_{gs}, then it doubles the effective V_{gs} will increase the current and thus reduce the delay in charging any capacitance on the output. Thus more symmetrical transitions are achieved.

Non inverting type buffer (nMOS):



The corresponding non inverting buffer as shown which has perspective structure of driving loads of 2 pF and with 5 nsec risetime.

If the inverting or non inverting buffer is arranged based on the native transistor, then it is known as native super buffer.

Channel Length Modulation and Velocity Saturation:

The voltages exceeding the onset of saturation there is an effective decrease in the channel length of short channel transistor, this is referred as **channel length modulation**.

For example, the change in channel length *ΔL* for a n- transistor is approximated by,

$$\Delta L = \sqrt{\frac{2\ \varepsilon_0\ \varepsilon_{Si}}{q\ N_A}(V_{ds} - V_{th})}$$

And the resultant drain to source current ${I^1}_{ds}$ is approximated by,

$${I^1}_{ds} = I_{ds}\frac{L}{L - \Delta L}$$

Velocity Saturation:

When the drain to source voltage of a short channel transistor exceeds a critical value, the charge carriers reach their maximum scattering limited velocity before pinch off. Thus less current is available from a short channel transistor than from a long channel transistor with similar width to length ratio and processing.

Therefore, channel length modulation and velocity saturation are the two effects important for short channel transistors, i.e. channel lengths ≤ 3 µm, and these effects should be taken into account.

Fan-in and Fan- out:

The number of inputs to a logic gate in an inverter while adding complementary transistor pairs which increases the delay times as the capacitance of the transistor is increased is called **fan- in (FI)** and the number of gates is specified by the **fan- out (FO)** of the circuit. The fan- out gates acts as a load to the driving circuit because of their input capacitance.

Problems:

1. A resistor of value 100 kΩ needs to be made from a resistive layer of thickness 1µm. If the resistivity of the material is 1 Ωcm and the strip of width 5 µm is used, then what should be the length of the strip?

Sol.

Given:

R = 100 kΩ = 1000×10^3 Ω

ρ = 1 Ωcm = 1×10^{-2}

t = 1 µm = 1×10^{-6} m

W = 5 µm = 5×10^{-6} m

To find:

L = ?

WKT,

$$R = \frac{\rho\, L}{t\, W}$$

$$L = \frac{R\, t\, W}{\rho} = \frac{1000 \times 10^3 \times 1 \times 10^{-6} \times 5 \times 10^{-6}}{1 \times 10^{-2}} = 5 \times 10^{-5} m$$

Therefore, the length of the strip is $5 \times 10^{-5} m$ respectively.

2. A layer of MOS circuit has a resistivity of 1 Ωcm, a section of this material is 5 µm thick, 5 µm wide and has a length of 50 µm, calculate the resistance from one of the section to the other using the concept of sheet resistance.

Sol.

Given:

ρ = 1 Ωcm = 1×10^{-2}

t = 5 µm = 5×10^{-6} m

W = 5 µm = 5×10^{-6} m

L = 50 µm = 50×10^{-6} m

To find:

R = ? using R_s, so first finding R_s also

WKT,

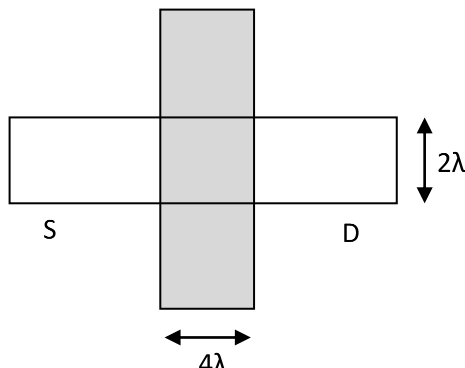
$$R_s = \frac{\rho}{t} = \frac{1 \times 10^{-2}}{5 \times 10^{-6}} = 0.2 \times 10^4 \; {}^{ohm}/_{square}$$

And

$$R = R_s \frac{L}{W} = 0.2 \times 10^4 \times \frac{50 \times 10^{-6}}{5 \times 10^{-6}} = 2 \times 10^4 \Omega$$

Therefore, the value of resistance is 2×10^4 Ω respectively.

3. For the given transistor structure, calculate the channel resistance in 5 µm, 2 µm and 1.2 µm technologies?



Sol.

Given:

L = 4λ

W = 2λ

- For nMOS:
  - In 5 µm technology

  WKT,

$$R = R_s \times \frac{L}{W} = 1 \times 10^4 \times \frac{4\lambda}{2\lambda} = 2 \times 10^4 \Omega$$

- In 2 µm technology

  WKT,

$$R = R_s \times \frac{L}{W} = 2 \times 10^4 \times \frac{4\lambda}{2\lambda} = 40\ k\Omega$$

- In 1.2 µm technology

  WKT,

$$R = R_s \times \frac{L}{W} = 2 \times 10^4 \times \frac{4\lambda}{2\lambda} = 40\ k\Omega$$

- For pMOS:
  - In 5 µm technology

    WKT,

$$R = R_s \times \frac{L}{W} = 2.5 \times 10^4 \times \frac{4\lambda}{2\lambda} = 50\ k\Omega$$

  - In 2 µm technology

    WKT,

$$R = R_s \times \frac{L}{W} = 4.5 \times 10^4 \times \frac{4\lambda}{2\lambda} = 90\ k\Omega$$

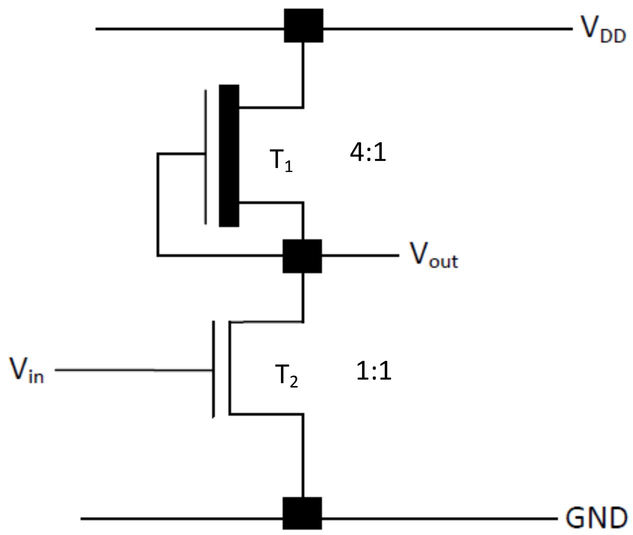
  - In 1.2 µm technology

    WKT,

$$R = R_s \times \frac{L}{W} = 4.5 \times 10^4 \times \frac{4\lambda}{2\lambda} = 90\ k\Omega$$

Therefore, the channel resistance of the given transistor are found.

4. For the given nMOS inverter, calculate the total resistance in 5 µm and 2 µm techmologies.



Sol.

Given:

The inverter has two transistors T_1 with L = 4 and W = 1 and transistor T_2 with L = 1 and W = 1.

- In 5 µm technology

WKT,

$$R = R_s \times \frac{L}{W}$$

$$R_{Total} = R_{T_1} + R_{T_2}$$

$$R_{Total} = \left(1 \times 10^4 \times \frac{4}{1}\right) + \left(1 \times 10^4 \times \frac{1}{1}\right) = 50\ k\Omega$$

- In 2 µm technology

WKT,

$$R = R_s \times \frac{L}{W}$$

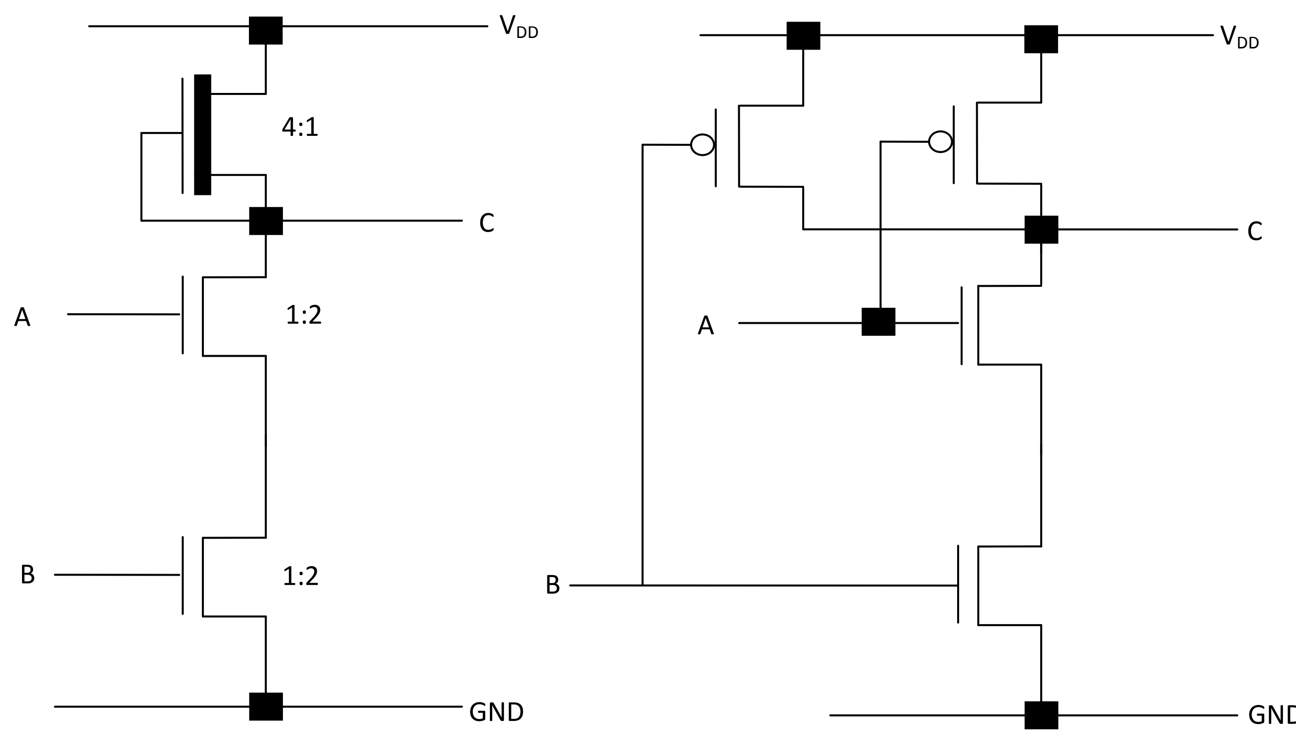
$$R_{Total} = R_{T_1} + R_{T_2}$$

$$R_{Total} = \left(2 \times 10^4 \times \frac{4}{1}\right) + \left(2 \times 10^4 \times \frac{1}{1}\right) = 100\ k\Omega$$

Therefore, the total resistance of the inverter in 5 µm technology is 50 kΩ and in 2µm technology is 100 kΩ respectively.

Assignment: (a) Draw the stick diagrams and mask layouts/ layout diagrams of the following and also find the total channel resistance of the nMOS NAND and NOR gates given:

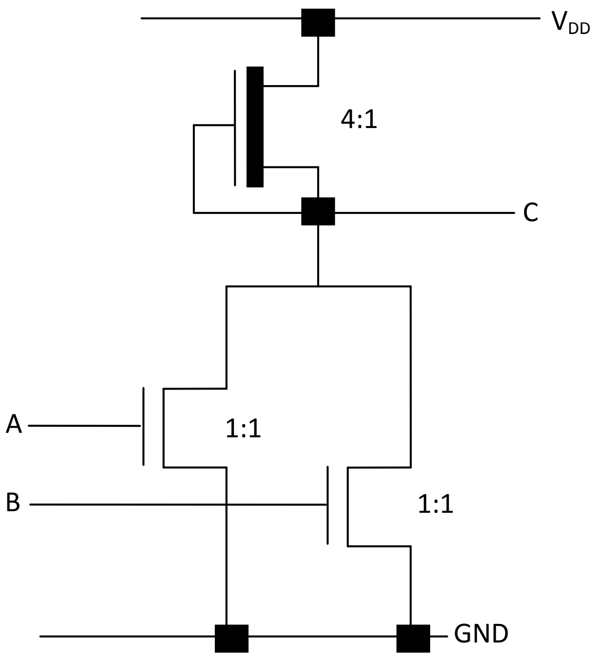
1. nMOS and CMOS NAND Gate:



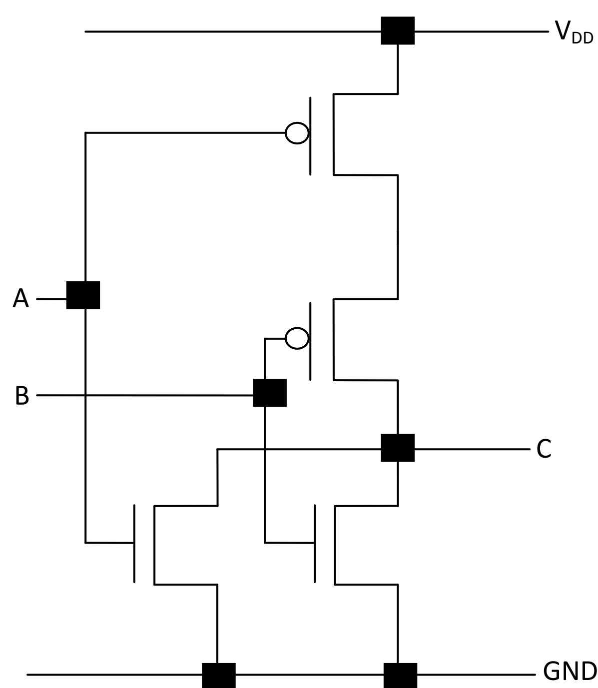
Circuit symbol of nMOS NAND Gate

Circuit symbol of CMOS NAND Gate

2. nMOS and CMOS NOR Gate:

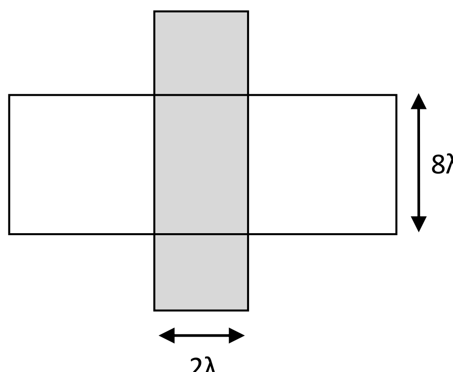


Circuit symbol of nMOS NOR Gate



Circuit symbol of CMOS NOR Gate

(b) For the given transistor structure, calculate the channel resistance in 5 μm, 2 μm and 1.2 μm technologies?



(c) Calculate the total resistance in a CMOS inverter in 5 μm, 2 μm and 1.2 μm technologies?

(Note/ Hint : For CMOS inverter $L:W = 1:1$ $(i.e.)$ $\frac{L}{W} = \frac{1}{1}$)