UNIT-III

MOS and BiCMOS Circuit Design Processes: MOS layers, stick diagrams, nMOS design style, CMOS design style Design rules and layout & Scaling of MOS Circuits: λ - based design rules, scaling factors for device parameters

****

MOS LAYERS

MOS design is aimed at turning a specification into masks for processing silicon to meet the specification. We have seen that MOS circuits are formed on four basic layers

- N-diffusion
- P-diffusion
- Poly Si
- Metal

Which are isolated from one another by thick or thin (thinox) silicon dioxide insulating layers. The thin oxide (thinox) mask region includes n-diffusion, p-diffusion, and transistor channels. Poly-silicon and thinox regions interact so that a transistor is formed where they cross one another.

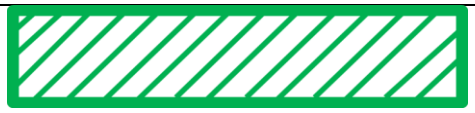
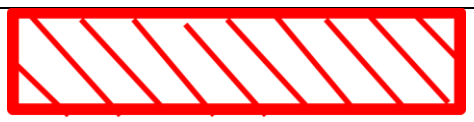
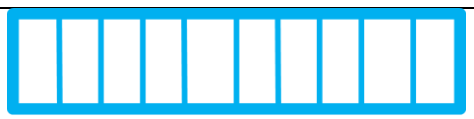
STICK DIAGRAMS

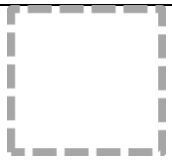
A stick diagram is used to convey layer information using monochrome encoding or color encoding. It is also cartoon of a chip layout. It acts as an interface between symbolic circuit and the actual layout.

Advantages:

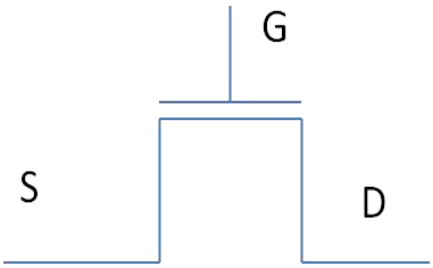
- It can be drawn much easier and faster than a complex layout.
- These are especially important tools for layout built from large cells.

NMOS TECHNOLOGY

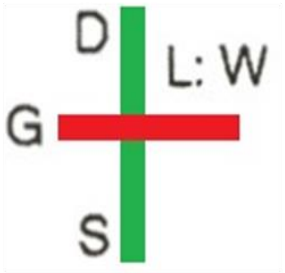
Colour	Stick Encoding	Layer	Mask Layout Encoding	CIF Layer
GREEN		N Diffusion		ND-NMOS Diffusion
RED		Poly-Silicon		NP-NMOS Poly-Silicon
BLUE		Metal-1		NM-NMOS Metal
BLACK		Contact Cut		NC-NMOS Contact Cut

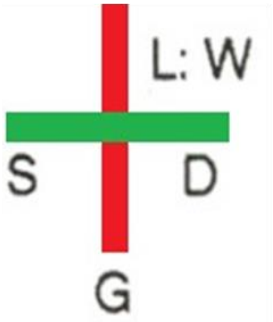
GRAY	-	Over Glass		NG-NMOS Glass
YELLOW (NMOS Only)		Implant		NI-NMOS Implant
BROWN (NMOS Only)		Buried Contact		NB-NMOS Buried Contact

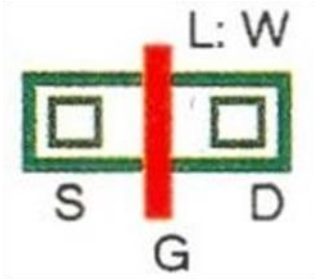
NMOS Enhancement Mode Transistor



STICK DIAGRAM







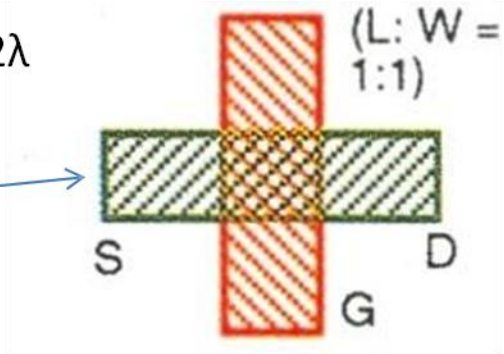
SYMBOL

Poly-Silicon

2λ x 2λ

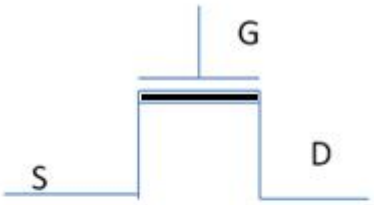
(L: W = 1:1)

N diffusion

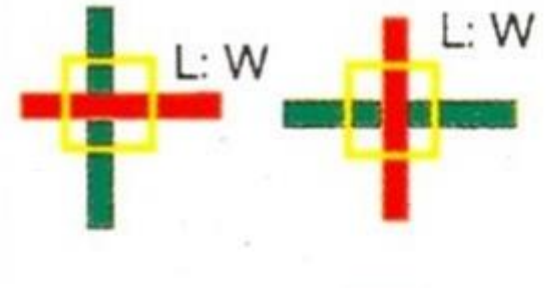


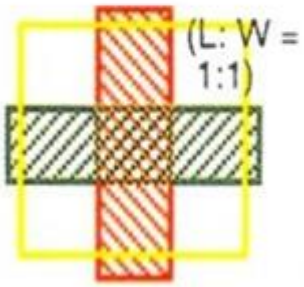
MASK LAYOUT

NMOS Depletion Mode Transistor



SYMBOL





MASK LAYOUT

NMOS Design Style

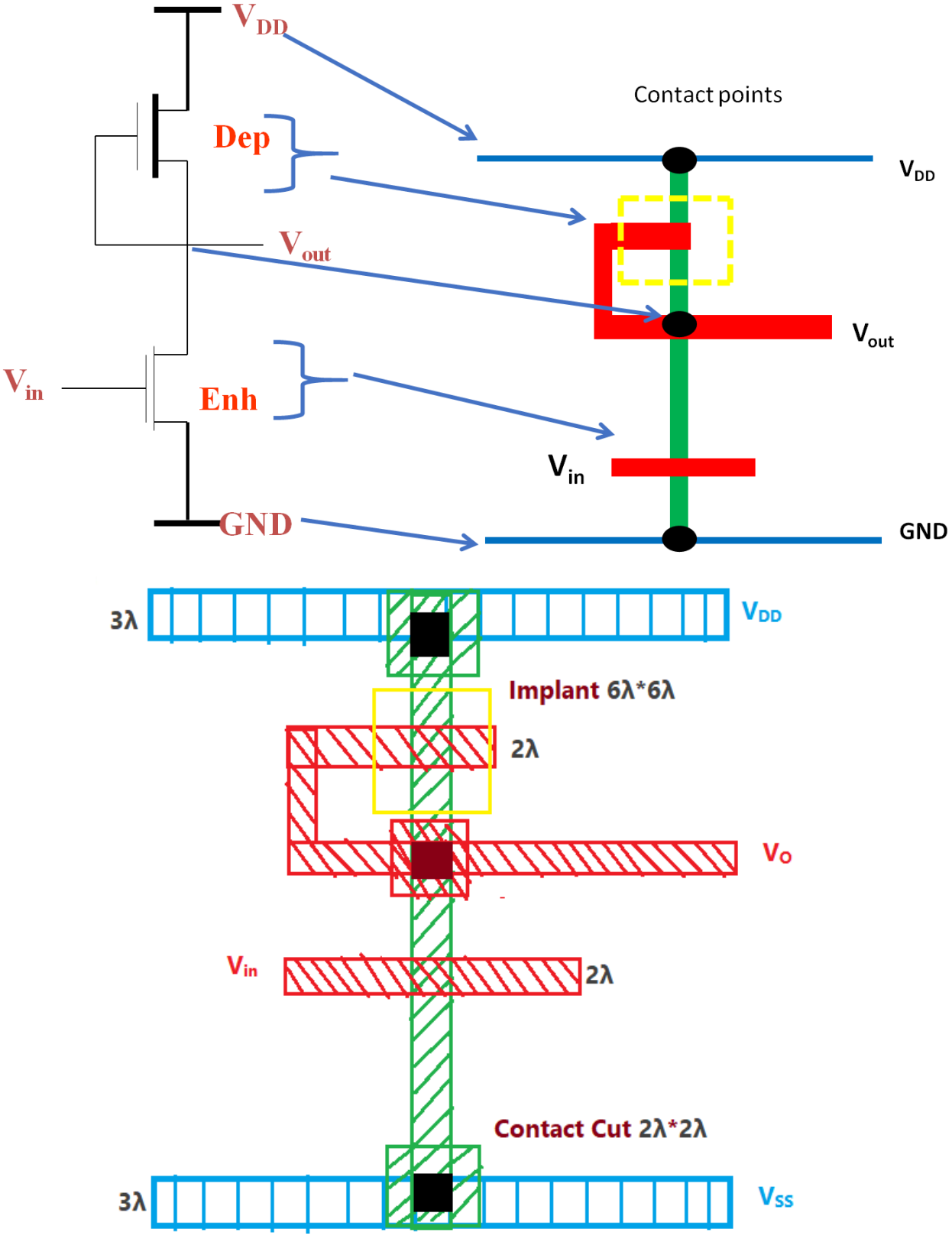
The layout of NMOS involves

- N Diffusion---Green
- Poly-Silicon---Red
- Metal------Blue
- Implant-----Yellow
- Contacts----(Black or Brown [Buried])

A transistor is formed where poly-silicon crosses n diffusion (red over green) and all diffusion wires are n type green.

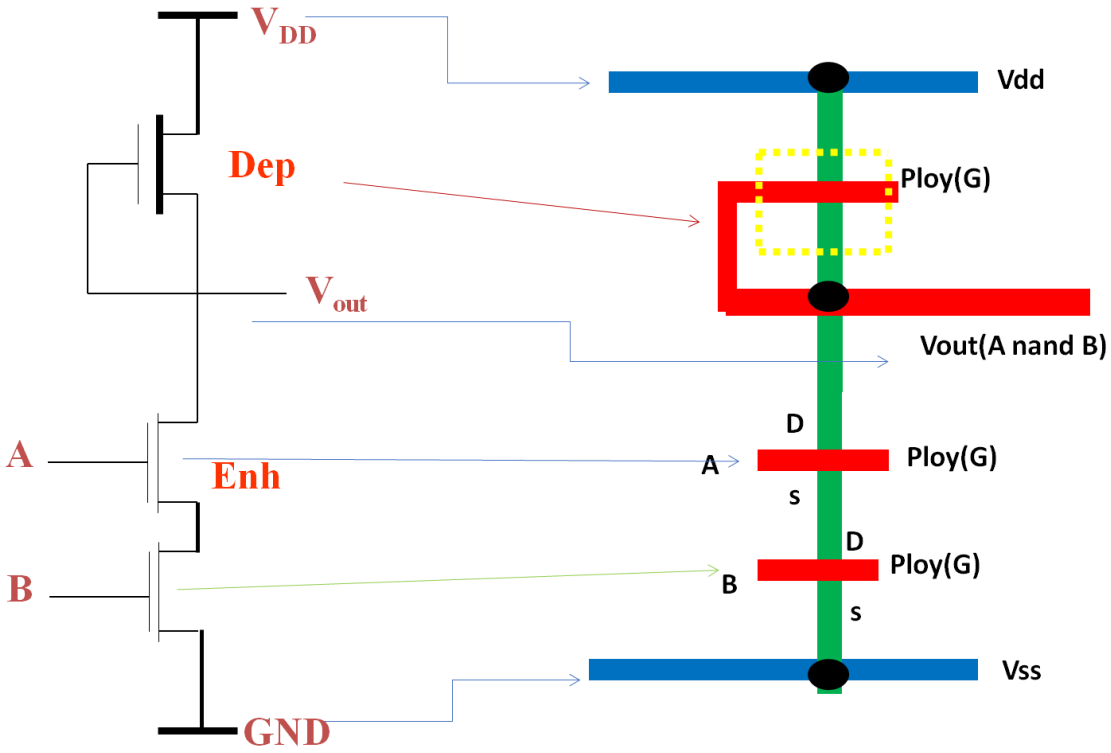
Design a stick diagram and Layout of NMOS inverter using NMOS Technology.

- Draw metal VDD and ground in parallel by keeping enough space between them for other circuit elements.
- Thin-oxide paths may be drawn between the rails for inverts and any inverts based logic.
- Pull up structure usually a depletion mode transistor connected from VDD to the output and enhancement mode transistor which is pull down transistor connected from output to the ground.
- As poly-silicon crosses the thin-oxide where transistor are required to drawn and implant are placed for depletion mode of transistor.

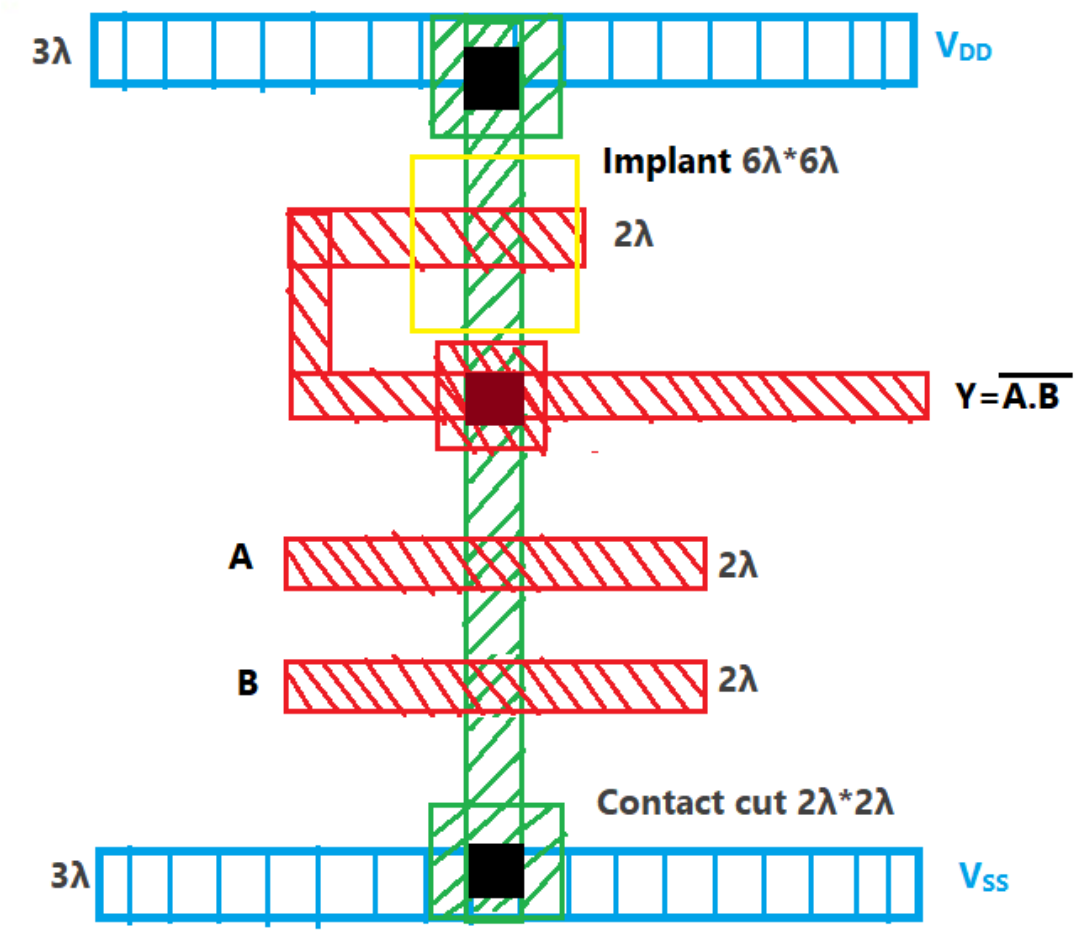


Logic circuit, stick diagram and Layout of NMOS INVERTER

NAND Gate Implementation

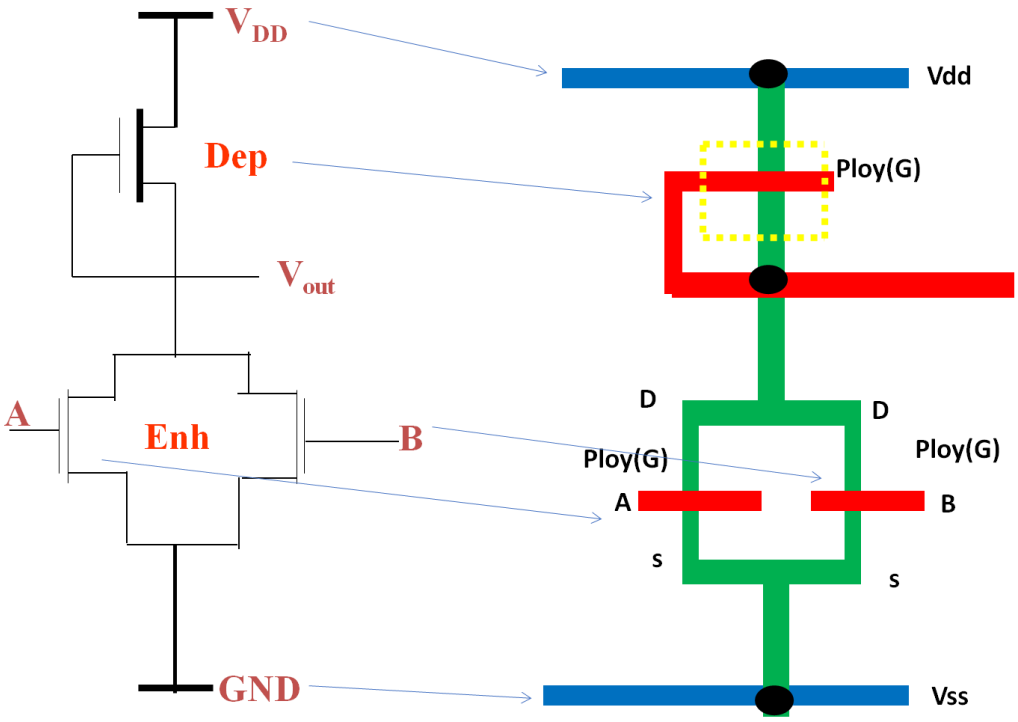


NMOS NAND GATE

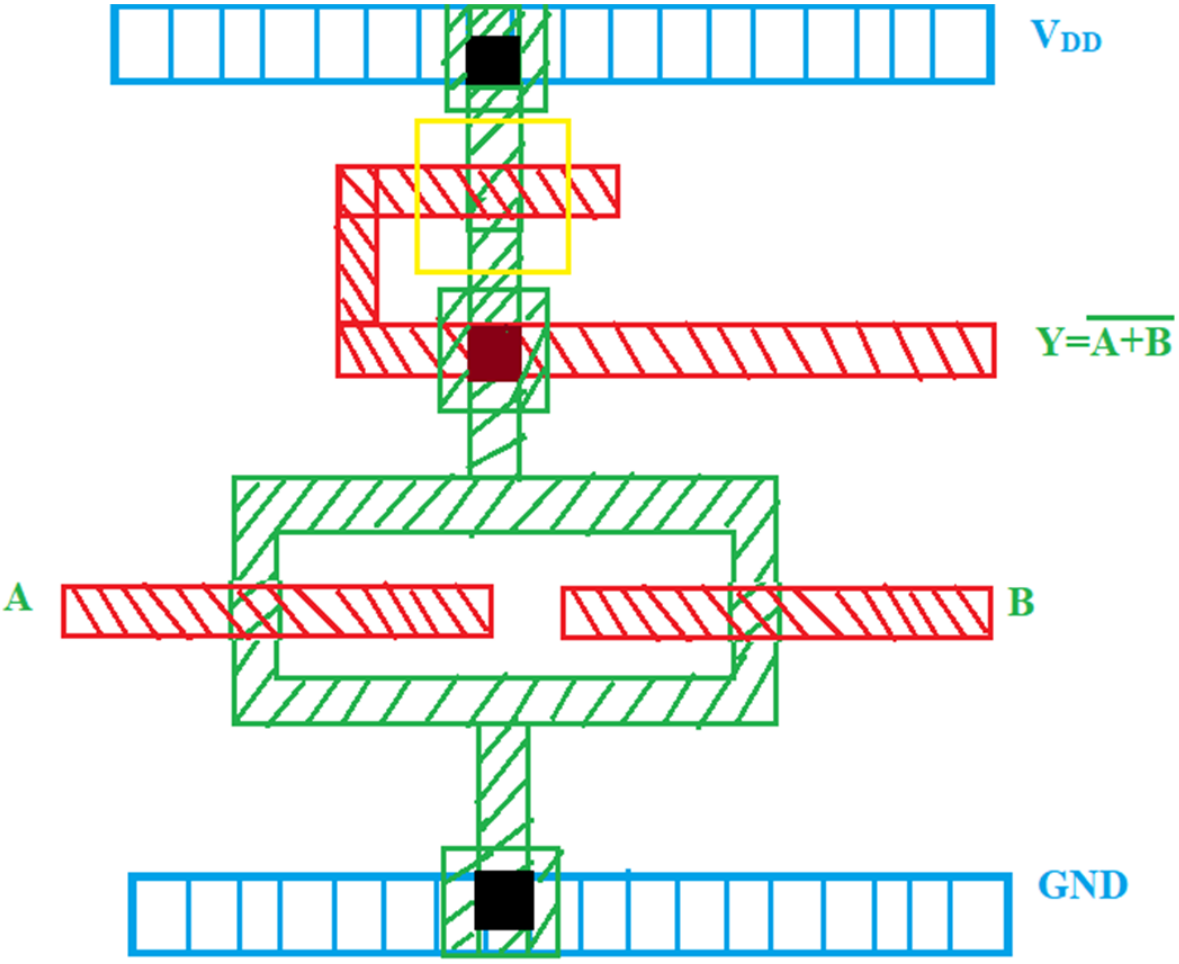


Logic circuit, stick diagram and Layout of NAND Gate

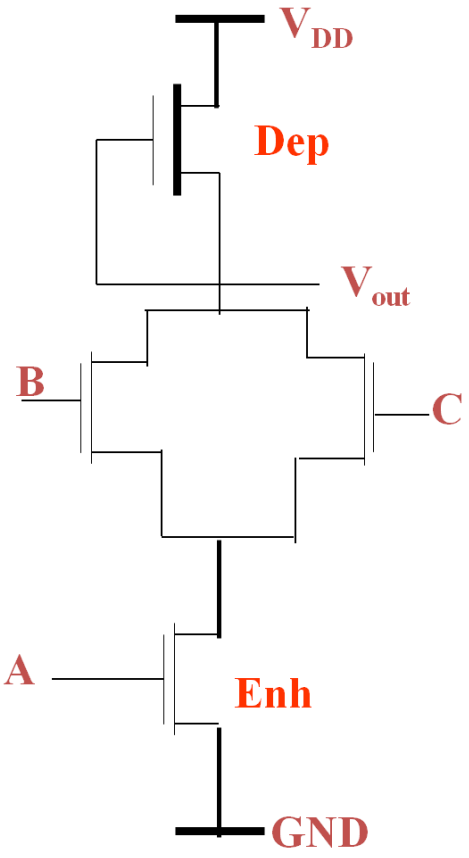
NOR Gate Implementation

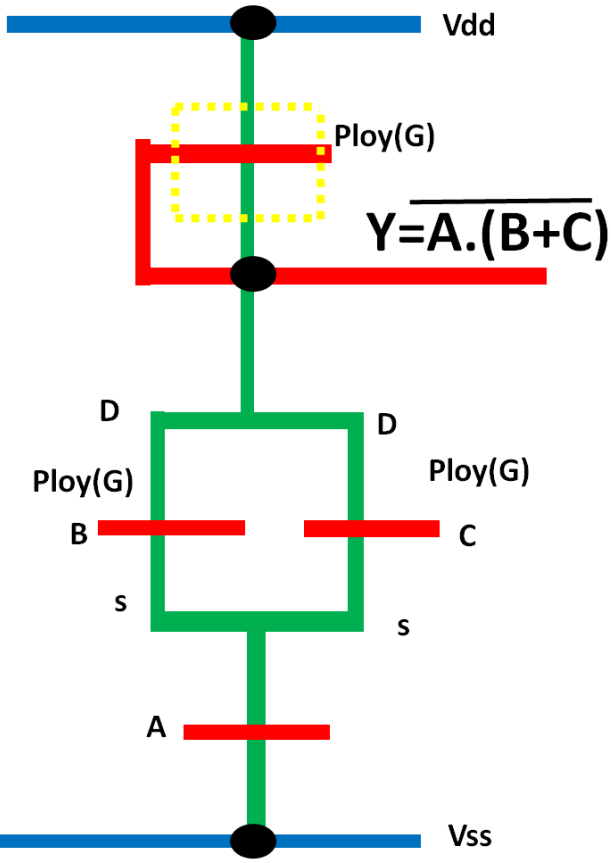


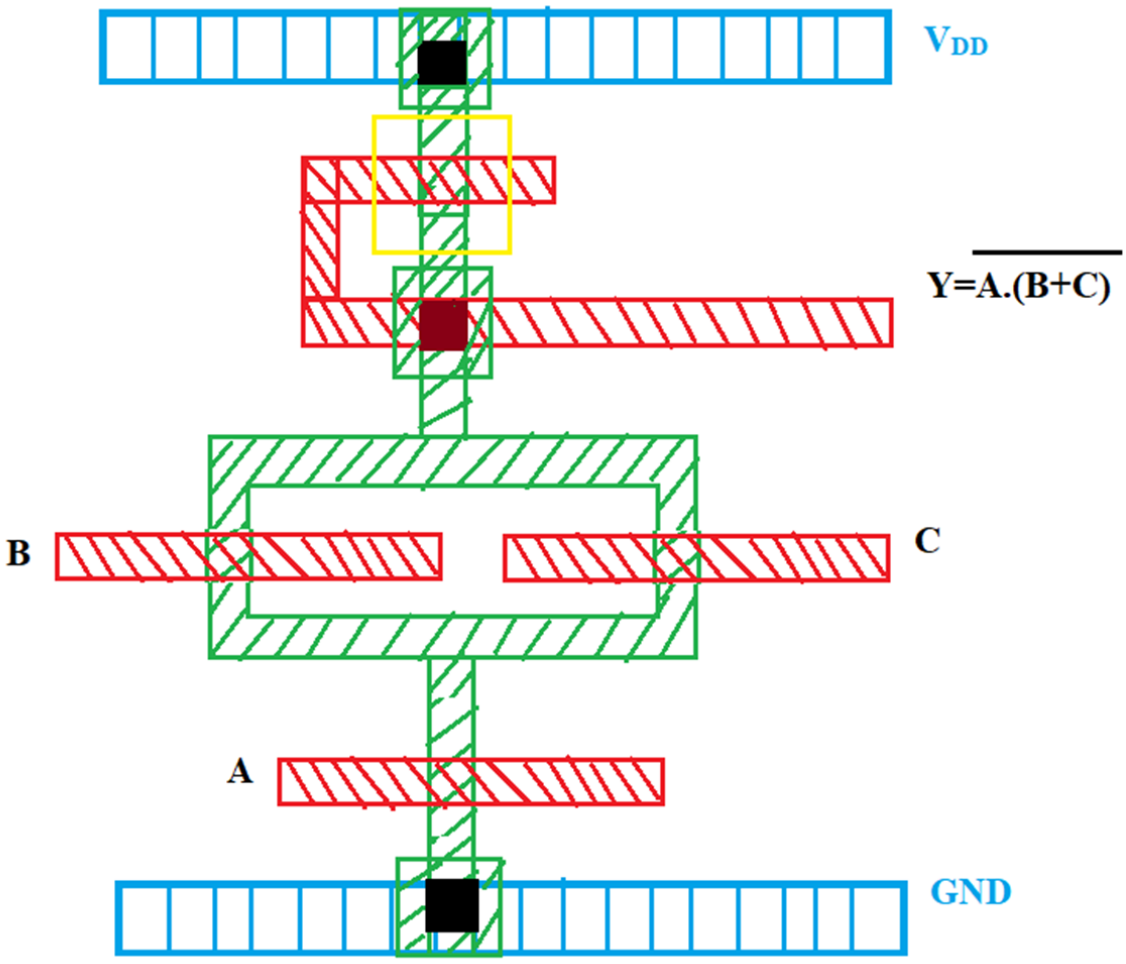
NMOS NOR GATE

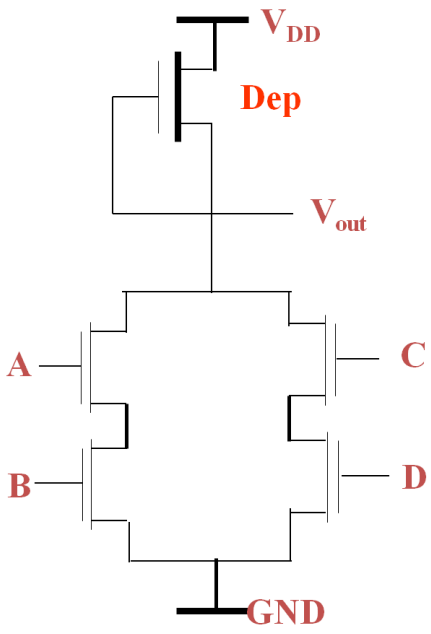


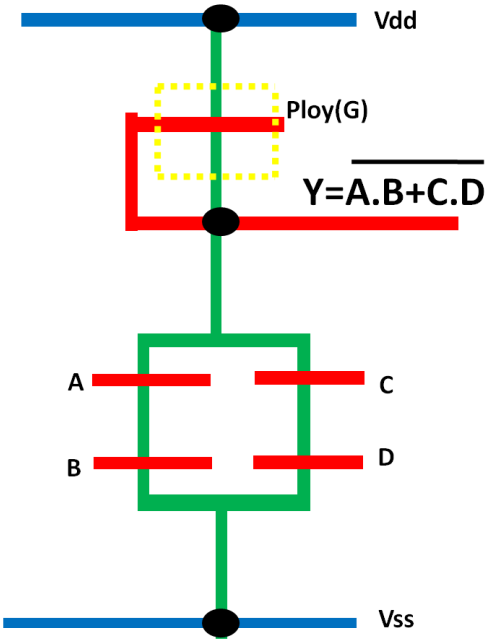
Logic circuit, stick diagram and Layout of NOR Gate

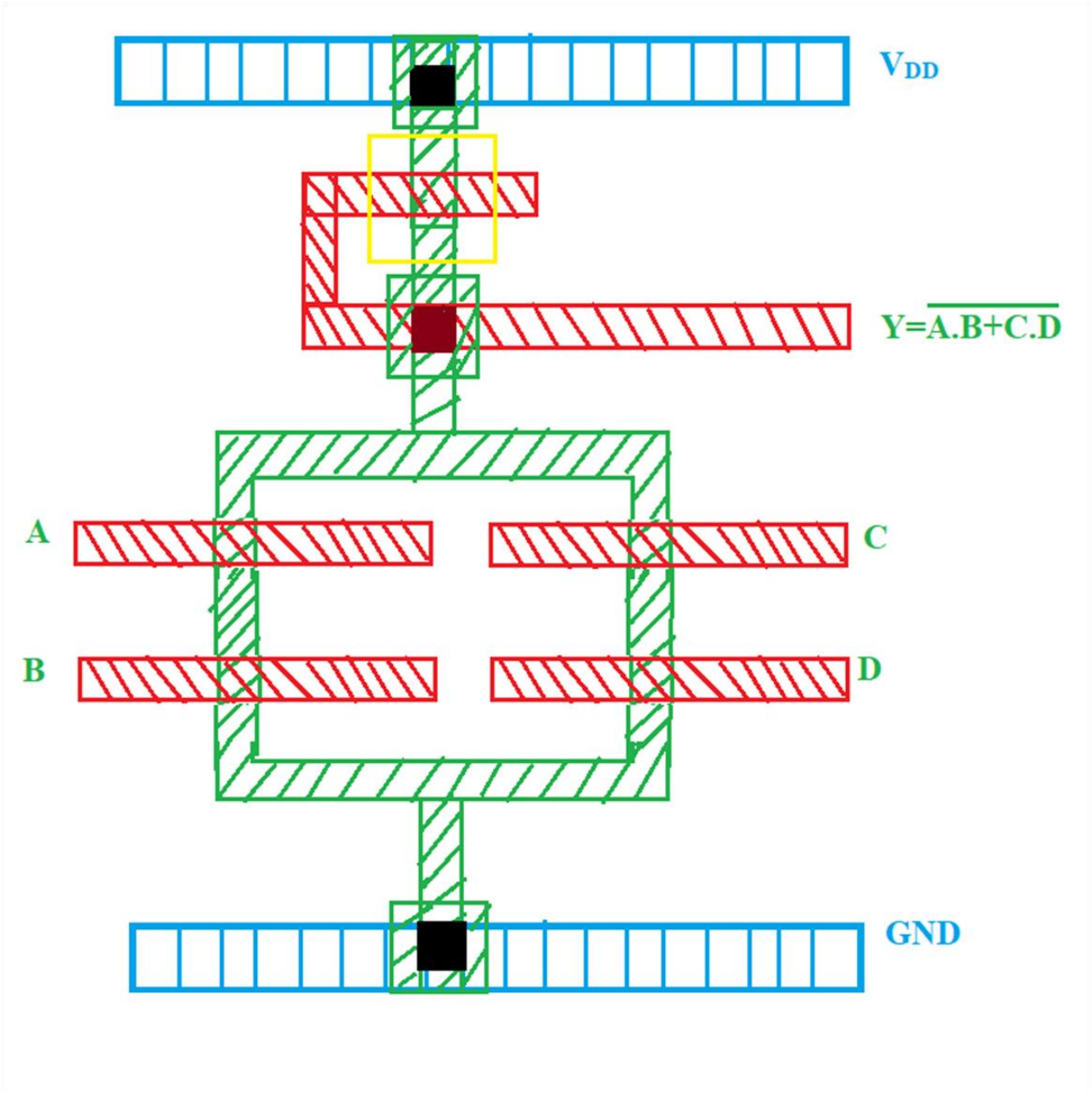

V_{DD}
Dep
V_{out}
B
C
A
Enh
GND


Vdd
Ploy(G)
$Y=\overline{A.(B+C)}$
D
D
Ploy(G)
Ploy(G)
B
C
s
s
A
Vss

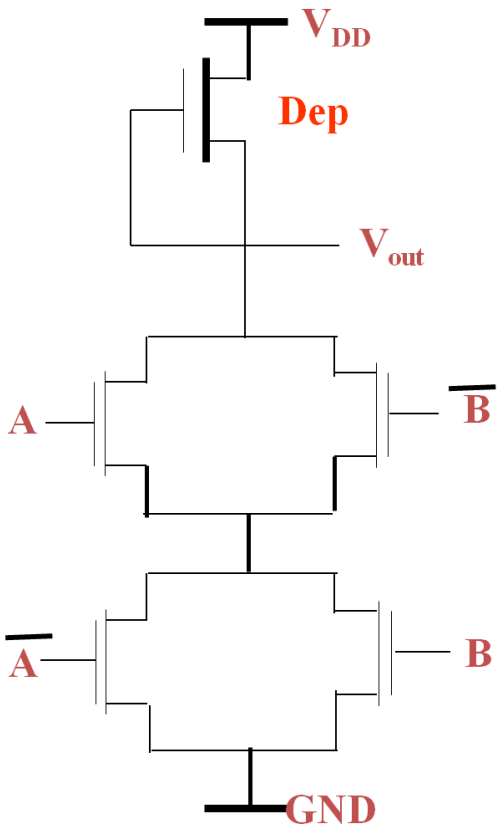

V_{DD}
$Y=\overline{A.(B+C)}$
B
C
A
GND

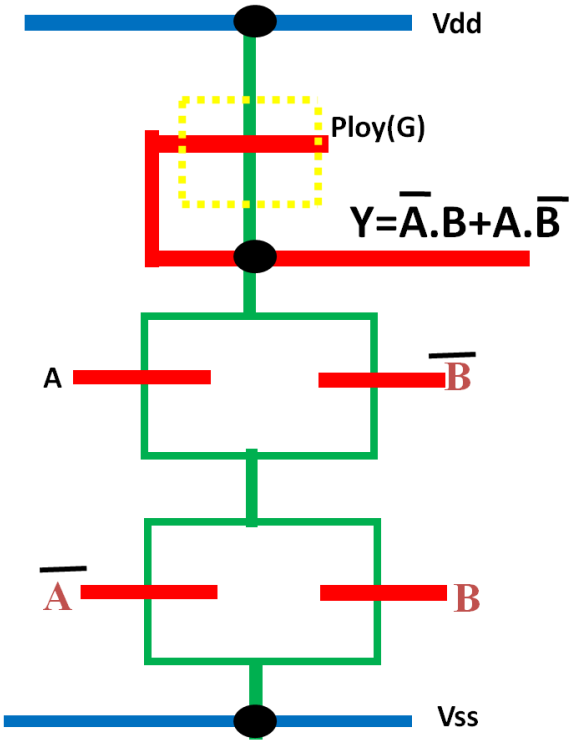

V_{DD}
Dep
V_{out}
A
C
B
D
GND


Vdd
Ploy(G)
Y=A.B+C.D
A
C
B
D
Vss

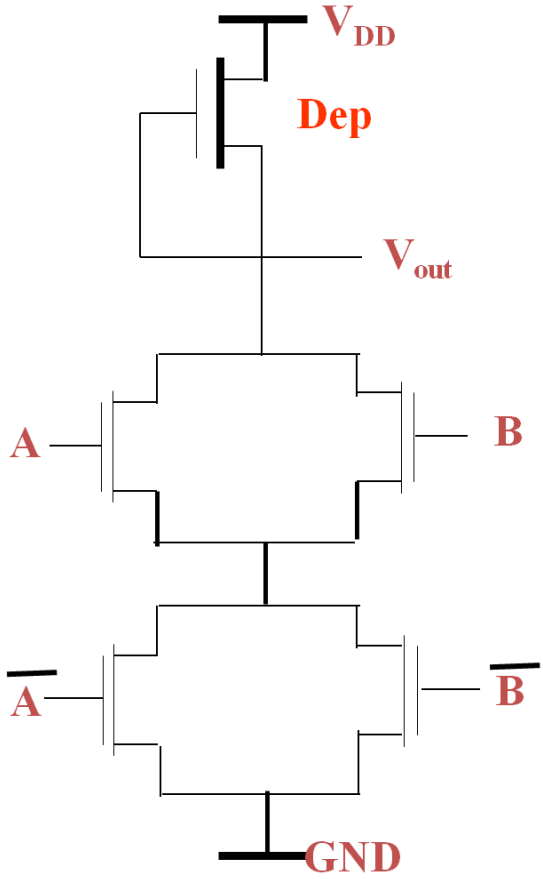

V_{DD}
Y=A.B+C.D
A
C
B
D
GND

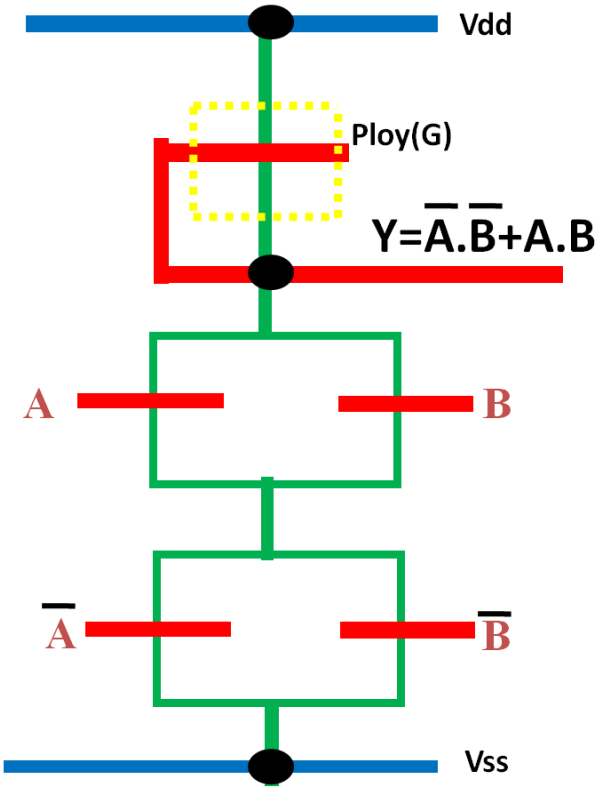
Design a stick diagram for NMOS EX-OR gate.





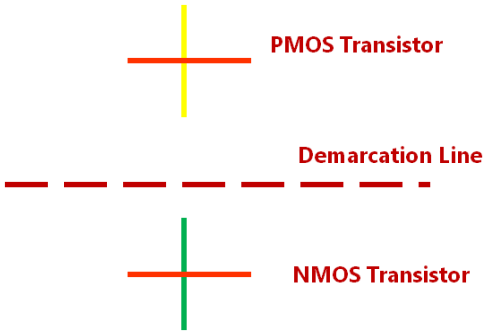
Design a stick diagram for NMOS EX-NOR gate.



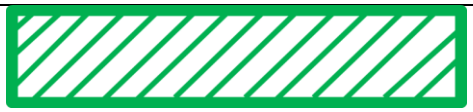
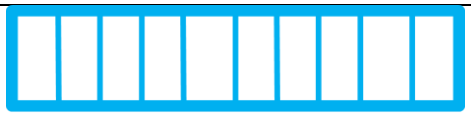
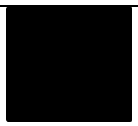
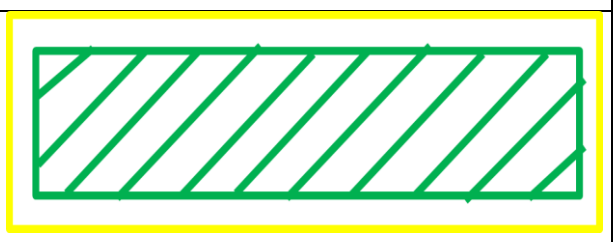


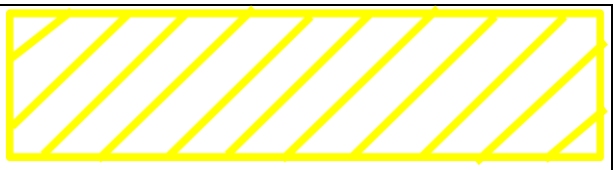
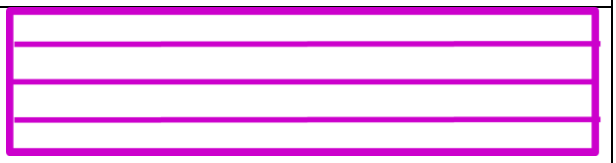
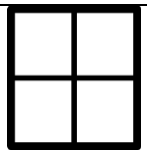
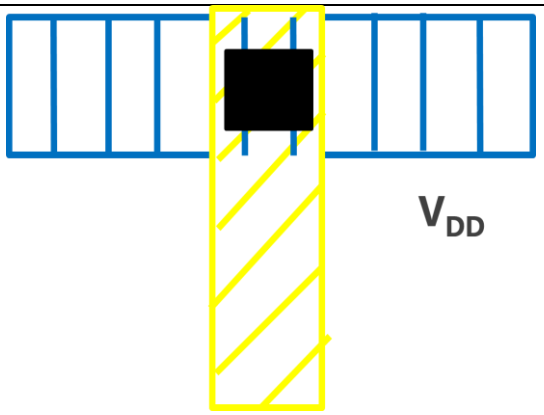
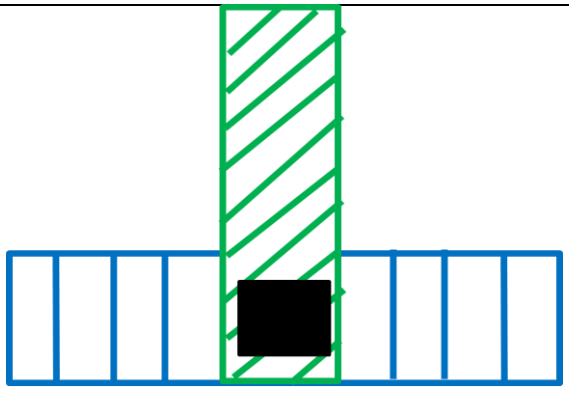
CMOS Design Style

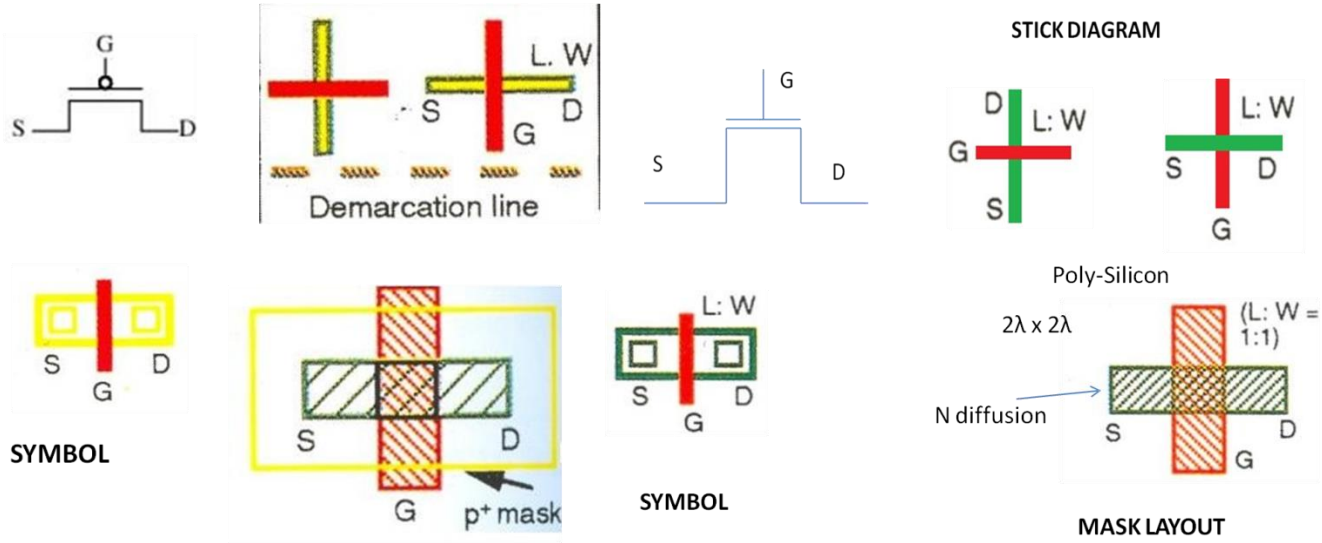
The stick diagram and layout representation for CMOS are logical extension of the NMOS design style. All features and layers defined for NMOS stick diagram with the exception of implant (Yellow) and the buried contact (Brown) are used in CMOS design. Yellow in CMOS design is now used to identify P transistors and wires, as depletion mode devices are not utilized. Two types of transistors used 'P' and 'N' are separated in the stick diagram and layout by demarcation line above which all P transistors are placed and N-devices are placed below the demarcation line.



Diffusion paths must not cross the demarcation line. The 'N' and 'P' features are normally joined by metal where a connection is needed.

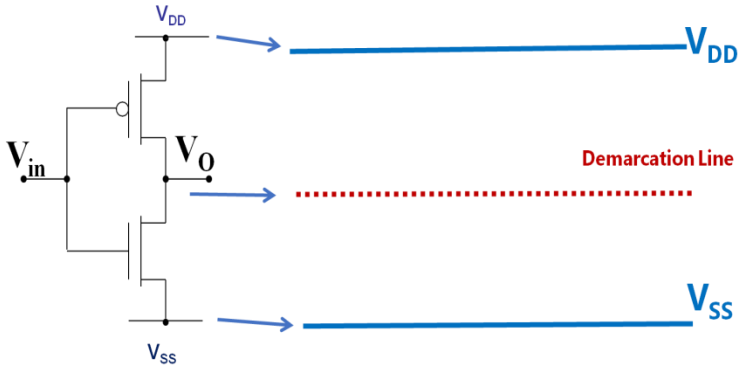
Colour	Stick Encoding	Layer	Mask Layout Encoding
GREEN		N Diffusion	
RED		Poly-Silicon	
BLUE		Metal-1	
BLACK		Contact Cut	
GRAY	-	Over Glass	
YELLOW		P Diffusion	

			
YELLOW		P^+ Mask	
Dark Blue or Purple Colour		Metal-2	
BLACK		Via	
BROWN		P-WELL	
BLACK		V_{DD} or V_{SS} Contact	
			

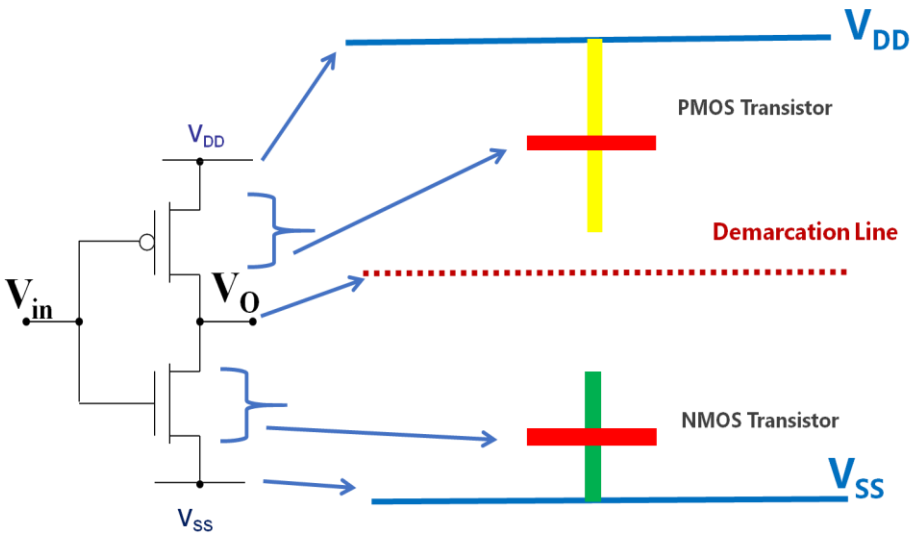


CMOS INVERTER

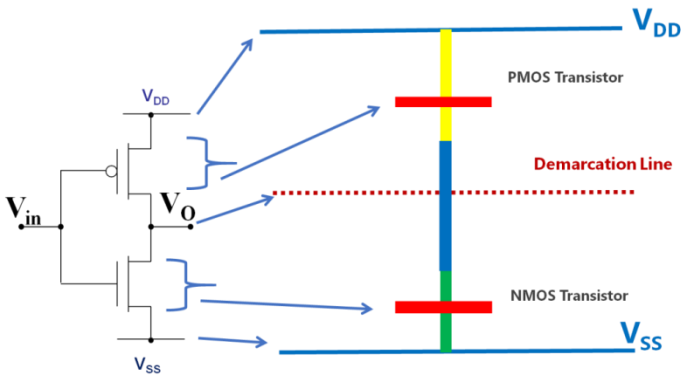
- The design style begins with the drawing V_{DD} and V_{SS}/GND power rails in parallel and creation of demarcation line in between V_{DD} and V_{SS} Rails.



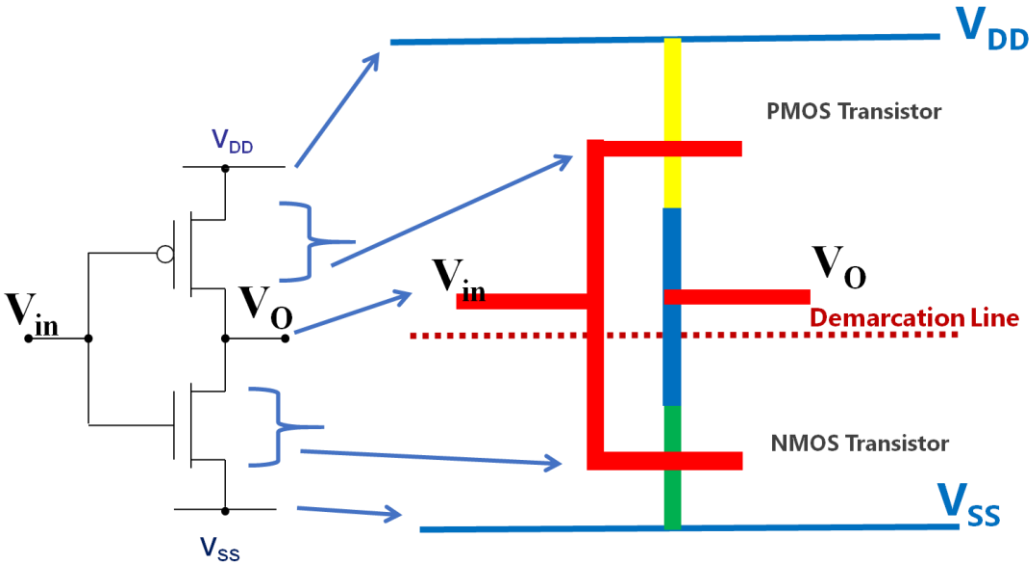
- The N-Transistors are then placed below this line and thus close to VSS, While P transistors are placed above the line and these are conveniently placed with their diffusion paths.

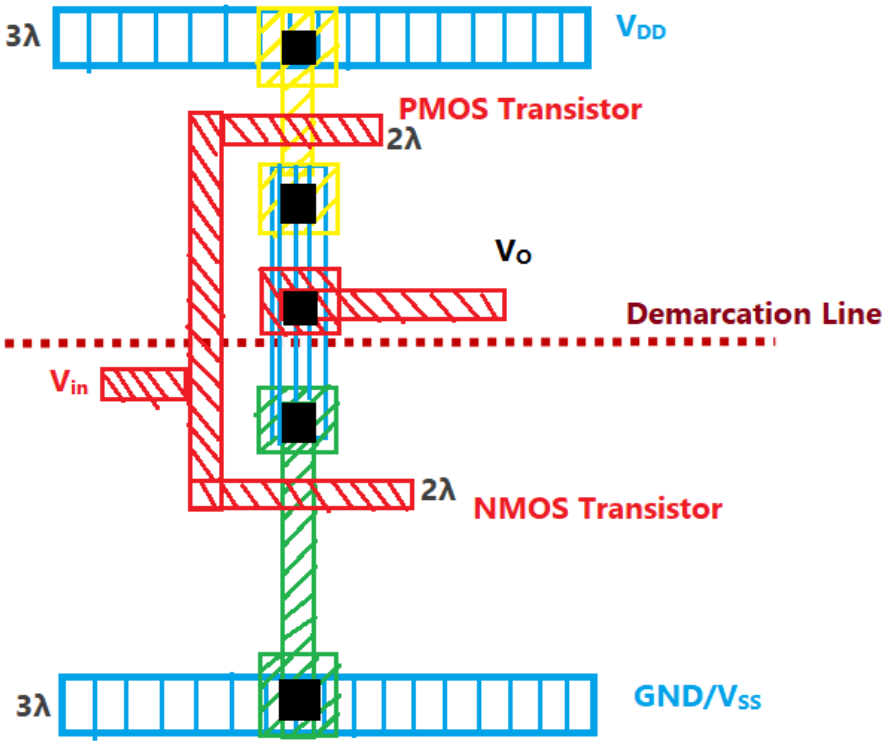


- A sound approach is to now interconnect the N with PMOS transistors are required using metal or poly-silicon and connect to the rails.

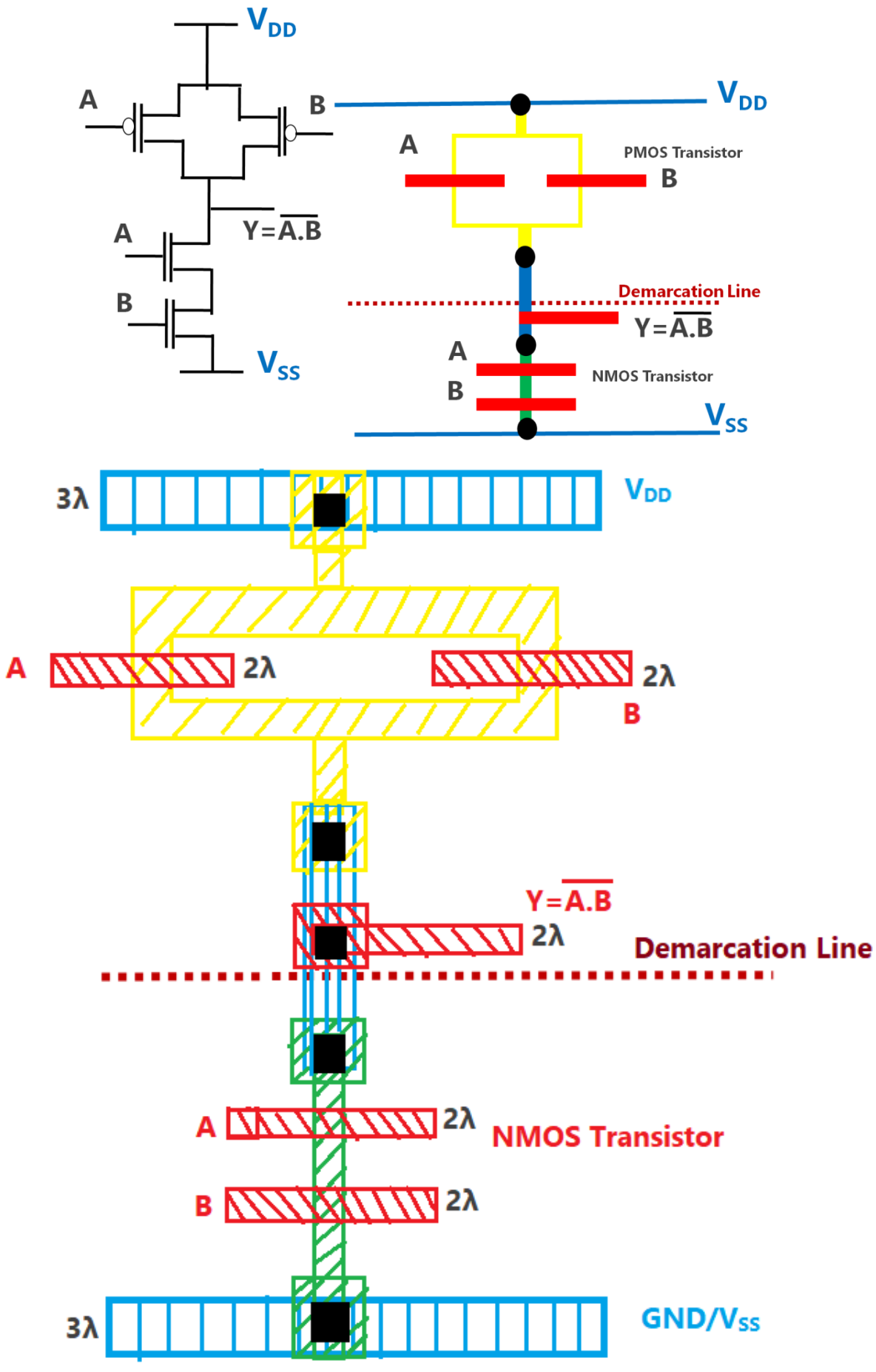


- A sound approach is to now interconnect the N with PMOS transistors are required using metal or poly-silicon and connect to the rails.

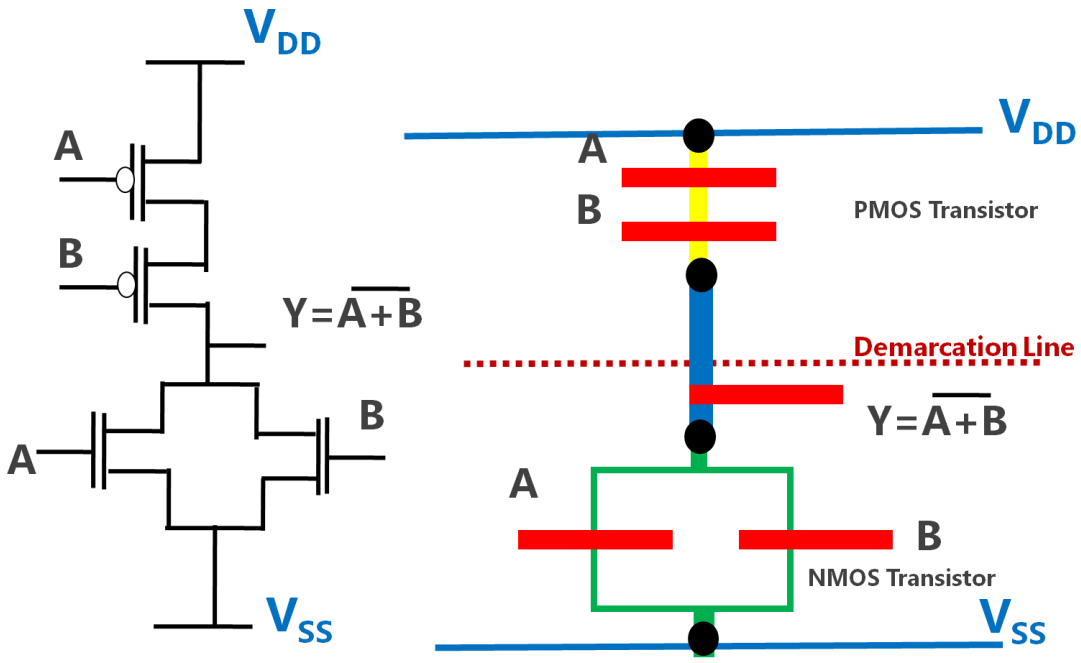


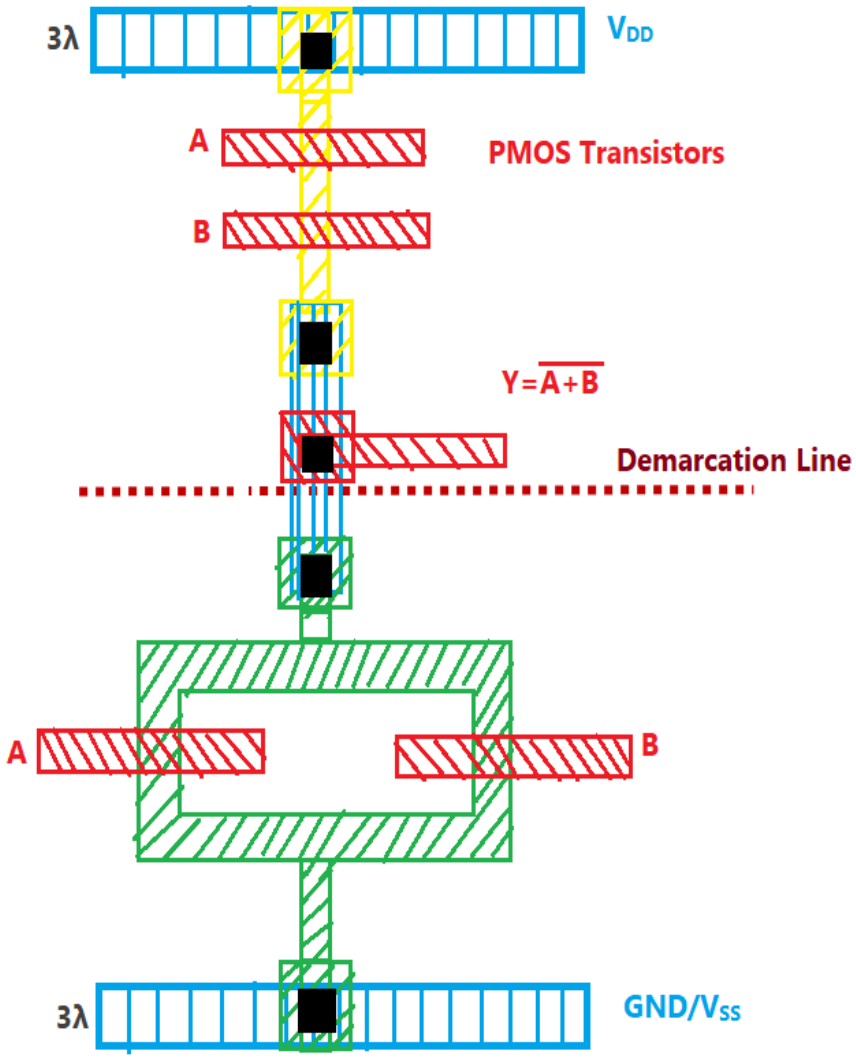


NAND Gate Implementation



NOR Gate Implementation



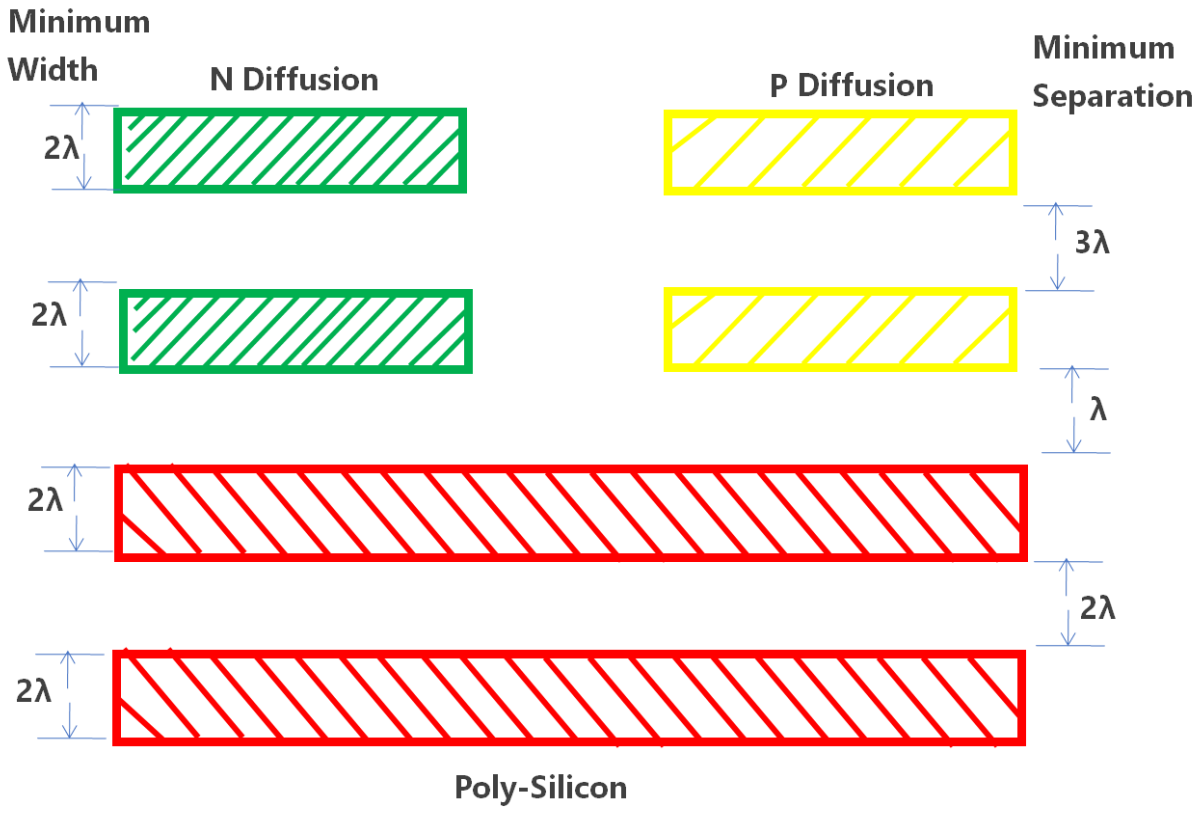


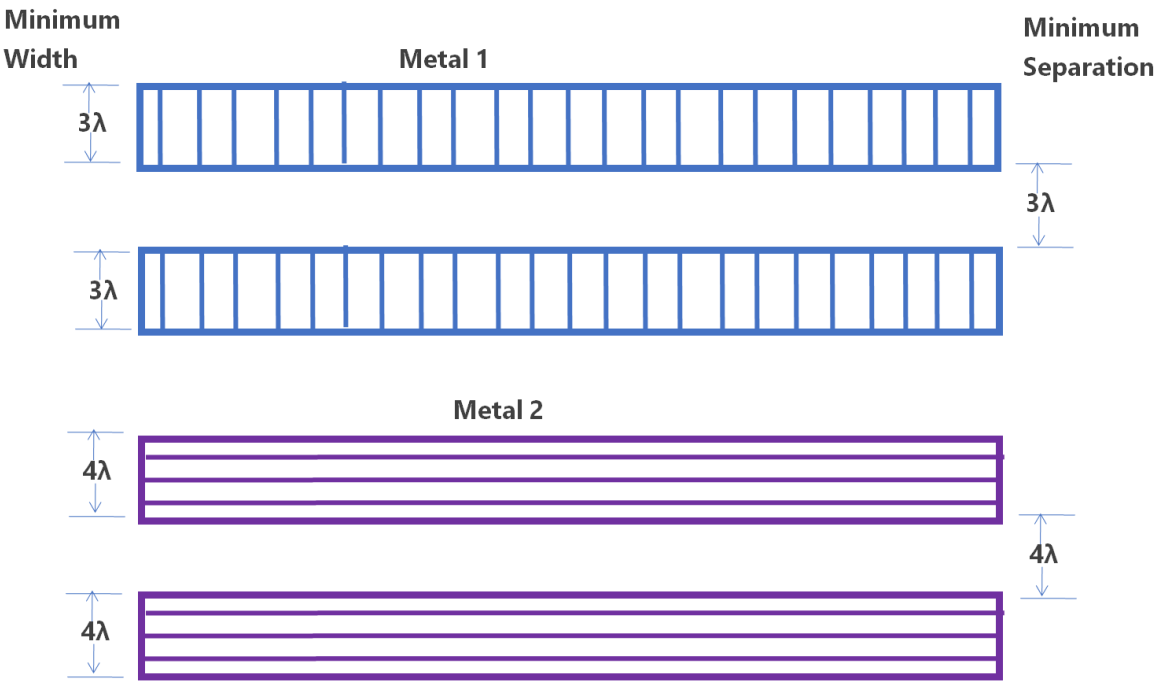
Design Rules and Layouts

- Design rules allow the translation of circuit design concepts, usually in stick diagrams or symbolic form into actual geometry in silicon. Small layouts for improved performance and decreases silicon area.
- Design rules govern the layout of individual components and interactions-spacing and electrical connections between the components.
- Design rules are specific to a particular semiconductor manufacturing process. It determines low level properties of chip design.
- As small a component size as possible is desired to increase the number of functions in the chip.

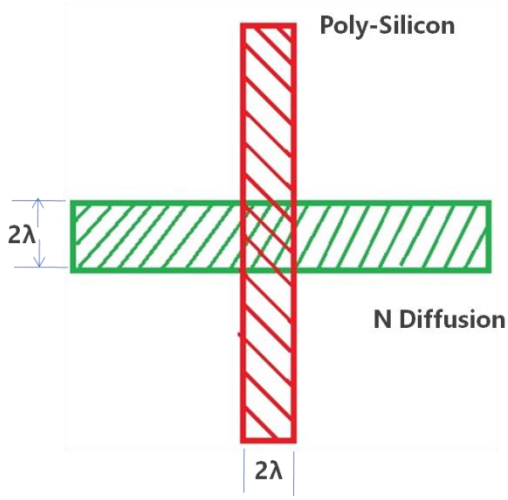
Lambda based design rules

- Design rules can be scaled in terms of 'λ' which is the size of the smallest element in the layout.
- All features can be measured in integral multiplies of 'λ'.
- By choosing a value for 'λ', all dimensions set at a scalable layout.
    - Minimum width of poly-silicon and diffusion line is '2λ'
    - Minimum width of metal 1 is '3λ' and metal 2 is '4λ'
    - Separation between poly-silicon to poly-silicon is '2λ'
    - Separation between metal 1 is '3λ' and metal 2 is '4λ'
    - Minimum separation between diffusion is '3λ'
    - 'λ' space between poly-silicon and metal condition
    - A separation of '2λ' should be between enhancement and depletion.

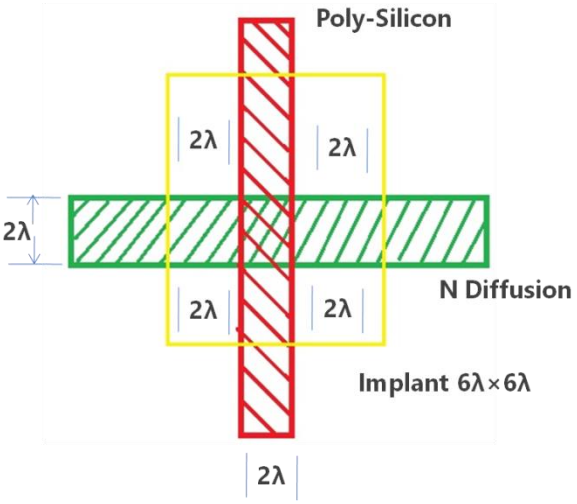




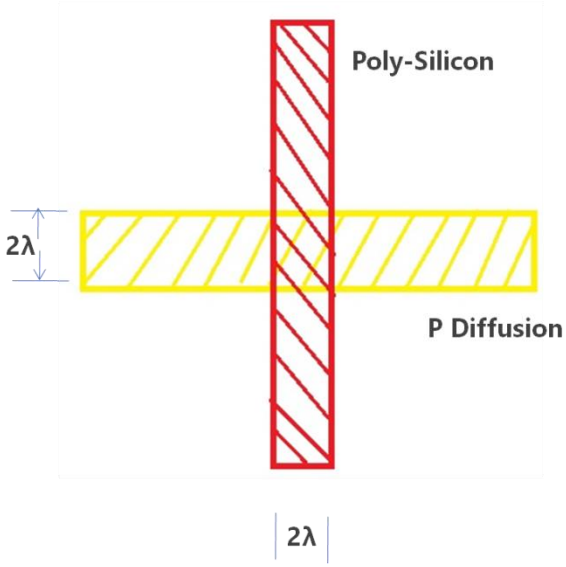
NMOS Enhancement Mode Transistor



NMOS Depletion Mode Transistor



PMOS Enhancement Mode Transistor

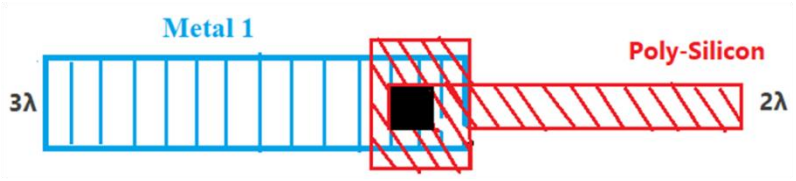


Contact Cuts

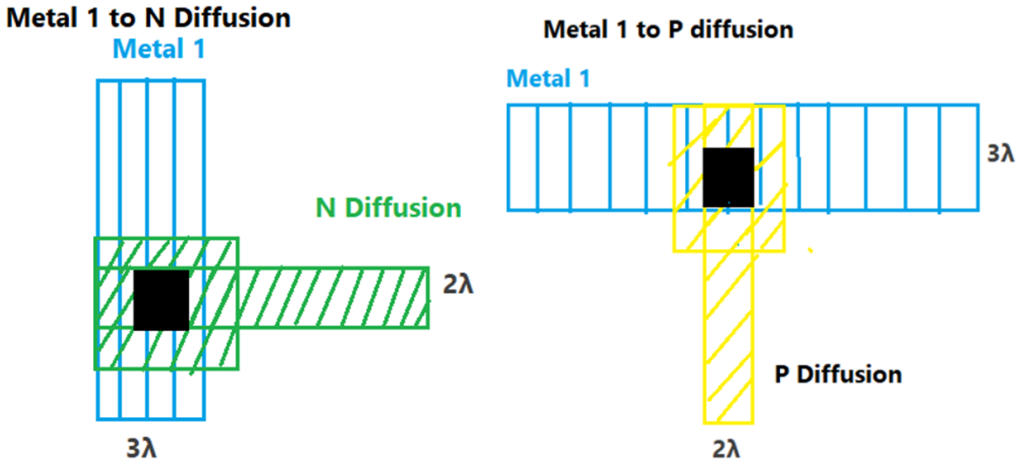
- When making contacts between poly-silicon and diffusion in NMOS circuit, it should be recognized that there are three possible approaches:
  - Poly silicon to metal then metal to diffusion
  - Buried contact (Poly-silicon to diffusion)
  - Butting contact (Poly-silicon to diffusion using metal)

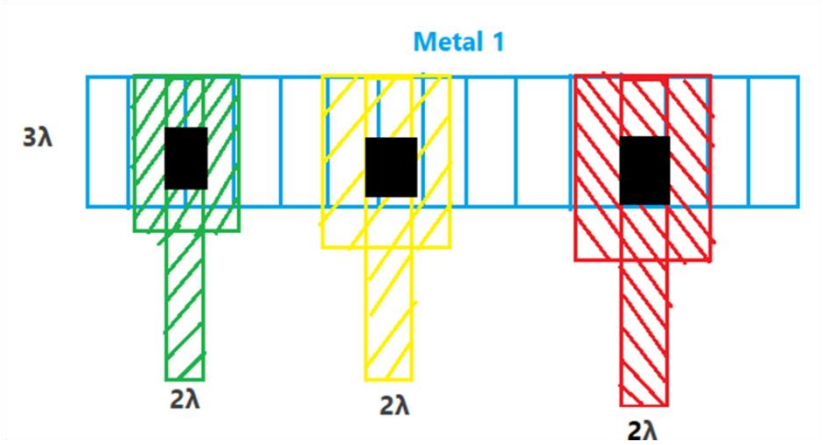
In CMOS design, poly. to diff. contacts are always made via metal. A simple process is followed for making connections between metal and either of the other two layers, the 2λ. x 2λ. Contact cut indicates an area in which the oxide is to be removed down to the underlying poly-silicon or diffusion surface. When deposition of the metal layer takes place the metal is deposited through the contact cut areas onto the underlying area so that contact is made between the layers.

The process is more complex for connecting diffusion to poly-silicon using the butting contact approach, In effect, a 2λ. x 2λ contact cut is made down to each of the layers to be joined. The layers are butted together in such a way that these two contact cuts become contiguous. Since the poly-silicon and diffusion outlines overlap and thin oxide under poly silicon acts as a mask in the diffusion process, the poly-silicon and diffusion layers are also butted together.

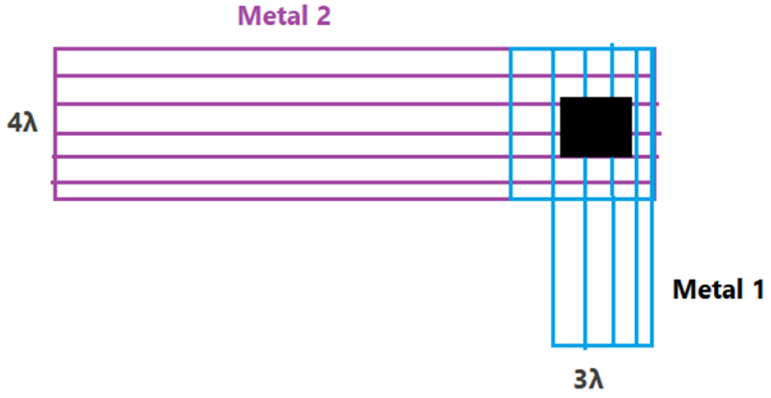


Metal 1 to Poly-Silicon

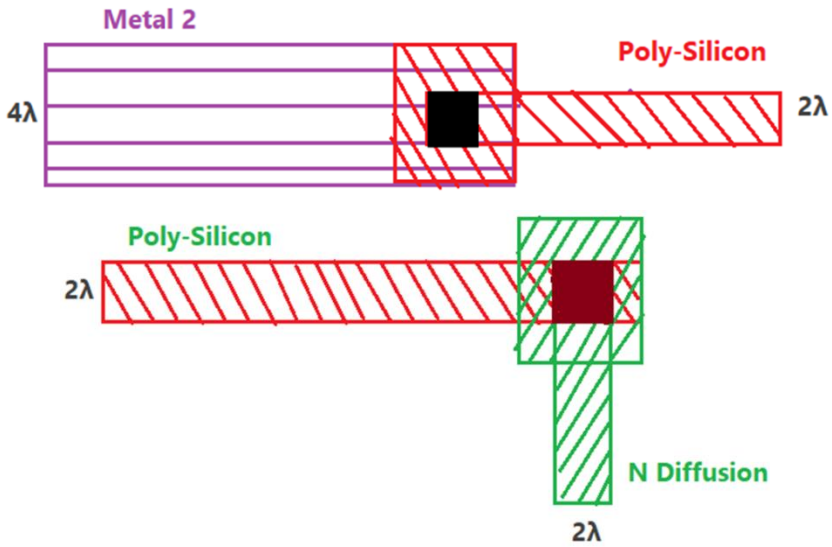


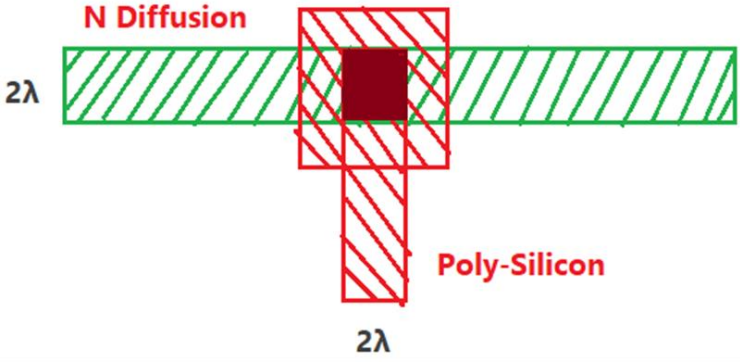


Metal 2 to Metal 1 contact cut



Metal 2 to Poly-Silicon





Scaling of MOS Circuits

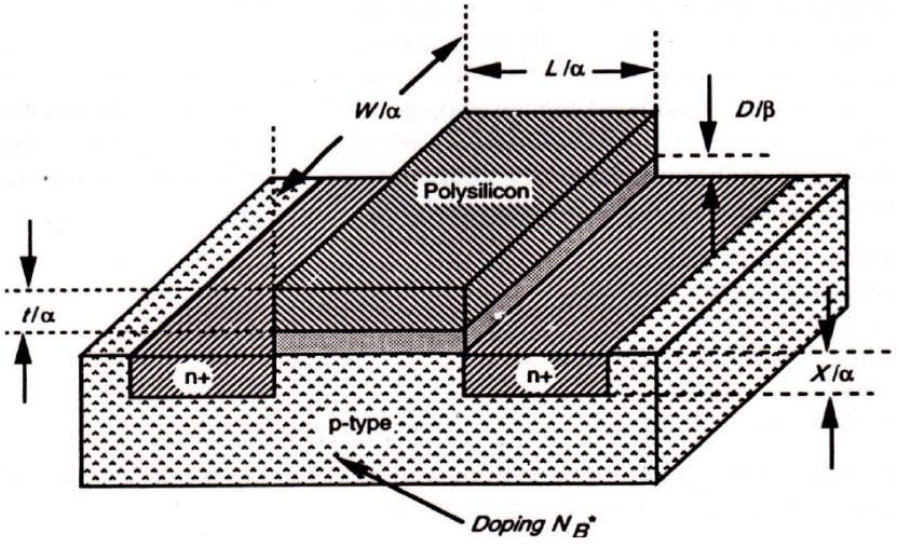
Design of high density chips in MOS VLSI technology requires

- ❖ High Packing density of MOSFETs
- ❖ Small transistor size

The scaling down of feature size generally leads to improved performance.

Microelectronic technology may be characterized in-terms of several indicators or figure of merit.

- ❖ Minimum feature size
- ❖ Number of gates on one chip
- ❖ Power dissipation
- ❖ Maximum operational frequency
- ❖ Die size
- ❖ Production cost



Scaling factors for device parameters

Scaling Models:

The most commonly used models are:

- The constant electric field scaling model
- The constant voltage scaling model
- Combined voltage and dimension scaling model

Device scaling modelled in terms of generic scaling factors: 1/α and 1/β

1/β---> scaling factor for supply voltage V_{DD} and gate oxide thickness.

1/α----> linear dimensions both horizontal and vertical dimensions

For the constant field model and the constant voltage model, β= α and β=1 respectively are applied.

1. Gate Area (Ag)

Gate area Ag=L.W

Where L and W are the length and width of the channel. Both are scaled by 1/α.

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

2. Gate capacitance per unit area (C_O)

Gate capacitance per unit area $C_O=\frac{\varepsilon_{ox}}{D}$

Where ε_{ox} is the permittivity of the gate oxide and D is the gate oxide thickness which is scaled by 1/β.

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

3. Gate Capacitance (C_g)

Gate Capacitance $C_g=C_O.L.W$

Where C_O is scaled by β, Length and width are scaled by 1/α.

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

4. Parasitic Capacitance C_x

Parasitic capacitance C_x is proportional to A_x/d

Where d is the depletion width around source or drain which is scaled by 1/ α, Ax is the area of the depletion region around source and drain which is scaled by $1/\alpha^2$

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

5. Carrier Density in channel Q_{on}

$Q_{on}=C_oV_{gs}$

Where Q_{on} is the average charge per unit area in the channel in the 'ON' state. Note that C_o is scaled by β and V_{gs} is scaled by 1/β.

Thus Q_{on} is scaled by= $\frac{1}{\beta}.\beta = 1$

6. Channel Resistance Ron

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

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

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

7. Gate delay τ_d

Gate Delay $\tau_d=R_{on}C_g$

Where R_{on} is scaled by 1

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

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

8. Maximum operating frequency f_O

f_O is inversely proportional to delay τ_d

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

9. Saturation Current Idss

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

Where C_O is scaled by β

L and W are scaled by 1/α.

V_{gs} and V_t are scaled by 1/ β

I_{dss} is scaled by $\beta . \left(\frac{1}{\beta}\right)^2 = \frac{1}{\beta}$

10. Current Density (J)

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

I_{dss} is scaled by=$\frac{1}{\beta}$

A is scaled by= $\frac{1}{\alpha^2}$

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

11. **Switching energy per gate Eg**

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

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

V_{dd} is scaled by=$\frac{1}{\beta^2}$

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

12. Power dissipation per gate Pg

Pg comprises two components such that

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

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

Dynamic component$P_{gd} = E_g f_O$

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

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

13. Power dissipation per unit area P_a

$$P_a = \frac{P_g}{A_g} = \frac{1/\beta^2}{1/\alpha^2} = \frac{\alpha^2}{\beta^2}$$

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

14. Power Speed Product P_T

PT=Pg. τ_d

So PT is scaled by $\frac{\beta}{\alpha^2} . \frac{1}{\beta^2} = \frac{1}{\alpha^2 \beta}$