DIGITAL INTEGRATED CIRCUITS
(ENCS333)

Dr. Khader Mohammad

T.R : 8:30 - 9:45 : Masri08

DIGITAL INTEGRATED CIRCUITS
(ENCS333)

Dr. Khader Mohammad

T.R : 8:30-9:45 : Masri08

L1: Introduction

R 17-2-2019

- The smallest things that we have in ICs are transistors.
- In the IC design the main transistor we are using is CMOS technology.

* **CMOS**: Complementary metal oxide Silicon / Semiconductor

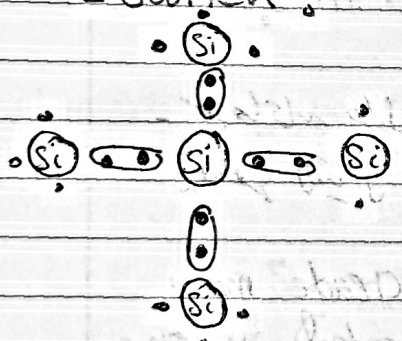
Transistor can use as switch → ON / OFF

Combination with different switches ⇒ different functions.

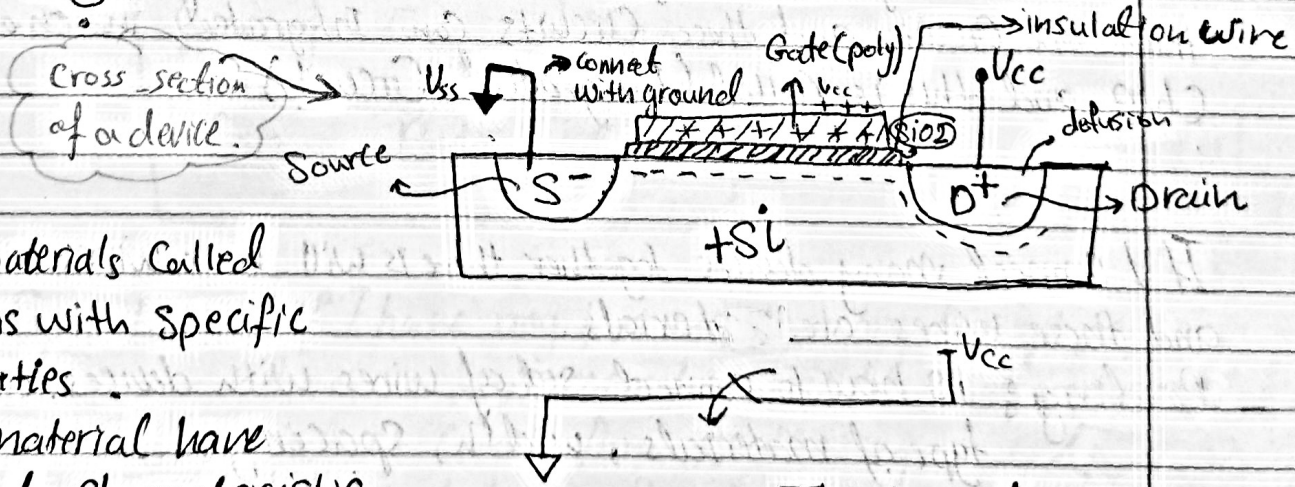
"Serial" if both switches are 'ON', I can connect between A and B.

Sure, using serial is different than using parallel.

- Switch:

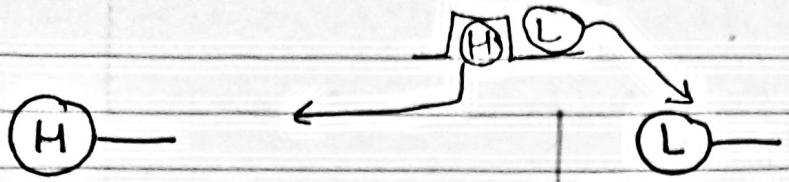


group of Silicon atoms linked together.
the charge of Silicon is positive (Si^+).
we build on it different materials.
materials ⇒ Drain, Source



These materials called defusions with specific properties.
each material have different characteristic.

This type of transistor is n-channel.
current go from Drain to Source.



The device on active high that's mean it works on high phase only (high signal).

If this device will be active-low that's mean it will work on low phase only.

- Some transistors work on high and another work on low. the transistors that work on high called → (NMOST device), and transistors work on low called → (PMOST device).

* How much the voltage is high or low ? ⇒ It depends on process node

process node means the length of transistor on the device (Lenght of Gate).
Because this length controls the size of device and the speed of current.
- كل ما كان طول الـ device أقل، كل ما كان أسرع لأنه المسافة التي يقطعها التيار تكون أقل.

When I have huge number of transistors, I can create more than one circuit and these circuits can integrated in one chip and this is called (Integrated Circuit).

If I connect many devices together there is will be a wires and these wires are materials.
- Routing : is how to connect set of wires with devices. and type of materials, width, spacing....

- placement : is where put the devices (cells).

Each wire or each device contains resistance and capacitance, Inductance.
this is what we call "Parasitic Extraction".
↳ (R, C, L) this's what we need to be able to run Timing.

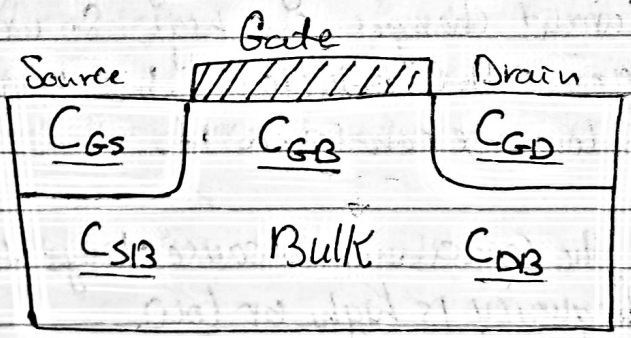
Schematic → translate to → Layout

↓ build the design then Simulate it. we make Simulation to check the time.

↓ actual physical layer

- model for devices contains specific information about all these devices and wire that will used.
If I work on 10 micro, the model of device should be 10 micro. and so on.

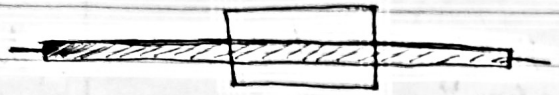
timing : Delay between input and output.
slope → time to be high and time to be low.



* Capacitors : If I have any two wires or any two plates there's a capacitor between them. and it comes from the voltage difference between the plates.

* We connect chips to the outside world ⇒ By [packaging] : make an interface with outside world.

* what we do exactly?
1- design combinational logic from transistors.
2- simulate it to insure that it's functionality correct "modle"
3- Layout
4. Timing Analysis ← Because now we have the actual design



* $V = IR$.
* $R = \frac{\rho . L}{A}$, ρ : resistivity , L : length , A : area of cross section.
* $C = \frac{\epsilon . A}{d}$, ϵ : material between wires
, A : Area between two wires ($L \times W$)
, d : distance between wires.
* $\tau = R . C$

* we ignore the inductors in chip level. but when we talk about top level design we need the inductance.

* $I_C = C \frac{\partial V}{\partial t}$
* $V_L = L \frac{\partial I}{\partial t}$

* The relation between frequency and (Capacitance & Inductance) is: When the frequency is high then the current changes are high so we can not ignore the inductance. but when the freq. is low we can ignore it.

* But we cannot ignore the capacitance because always there is a voltage whether the frequency is high or low.

- DRC (Design Rules) : we take it from people who works on Fab → (manufacture the machine).

▽ = Each device have numbers of transistors, and these transistors are doubles every 2 years.

Smaller devices
- positive : power, Area translate to cost, Speed (timing)
  - ↳ The main IC characteristic.
- negative : Complication in design, Leakage power
  - ↳ Smaller → leakage

* **Wafer**: We can make thousand of chips on the same wafer.
* **Die** : the smallest thing is on the wafer.

- The thickness of wafer depends on the process that it works on.

* number of layers that connect between devices are affect to thickness of wafer. How?

answer

* When the thickness of wafer be small then it will be less expensive → less complexity because the connections between devices are few → Less number of layers.

* When the thickness of wafer be large that is mean the complexity is increased because number of layers will be large to connect the devices together → more expensive.

* Design Abstraction levels.
  - System / Architecture.
  - Module.
  - Gate : build the gates we need.
  - Transistor level
  - device : layout design

- we cut the die from the wafer then we put it in package which has pins to connect with outside world, and those pins should be fast so we care about the material of wires.
[ρ should be very Low] like Gold.
Less resistance

* **Wafer** : A circular piece of pure silicon.
* **Die** : A rectangular piece of silicon that contains one IC design.

L2:
T/12-2-2019

IC Levels of Design

Manual

Specification (Marketing)

↕ Verification

Architecture (RTL)

Verilog (System Verilog)

Logic (Structural / behavioral)

Circuit (Transistor Level)

Layout

Fab → Silicon (Chip) → Testing

Automation (ASIC)

we don't use Automation for Critical designs like memory design, Because we need to make optimisation for placement or Routing the smallest as possible. So we use it in Control logic or in wich [II] differ frequantly. Why? if not we will waste month of work.

[II] RTL from Specification

↓ faster

we use tools like ICC.

- Synthesis
- use library Cells cells for schematic or for layout [Structural Module]

[2] Layout (GDS)

③ fab

④ chip

⑤ Testing

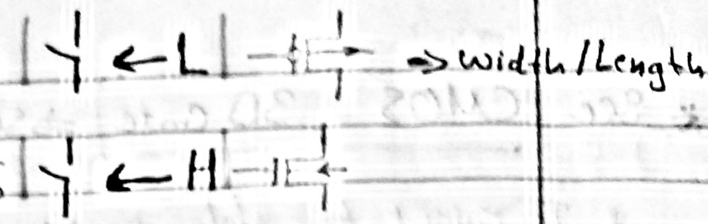
Graphic Data System

Design

** The meaning of these steps are on slides.

ICC ⇒ This is a tool takes (RTL) then do the Reading and placement.

* Transistors (switch) → PMOS, NMOS



* Process → Length → Area under the gate (L)
→ DRC: Design Rule Checker

* The three basic DRC checks: width, Spacing, Enclosure.

- Eeach process have different set of DRC.

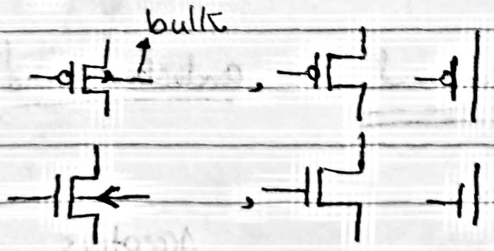
L = 45 nm → easier to make a design than L = 10 nm.

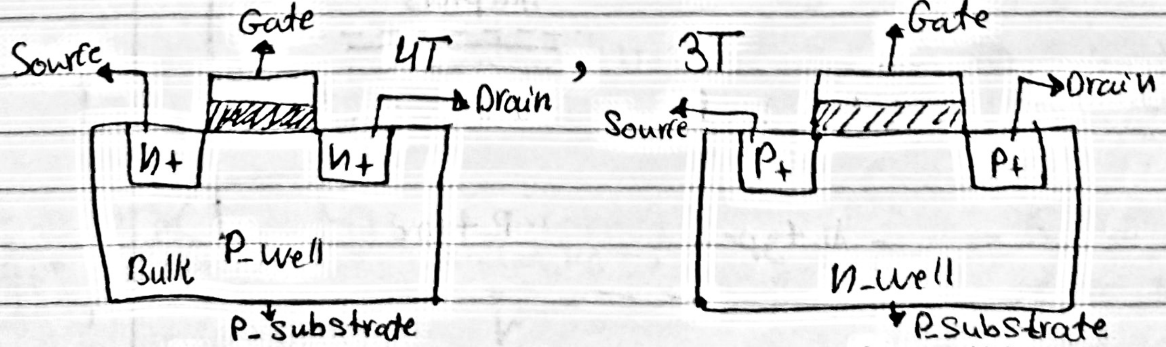
- اختيار ال process مهم جداً لأنها تأثر على كل من :
- Speed - power - Area.

* When we choose L = 10 nm ⇒ - less area
- Speed: faster
- Cost: Less
- Complexity will increase.

* we get the design rule from Fab.

- P-ch :
- N-ch :





n-ch MOS   P-ch MOS

* See CMOS , 3D Gate ⇒ Slides page (26+27) *

* To calculate delay in wire :-

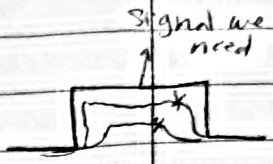
* $\tau = R.C$

* $R = \frac{\rho.L}{A}$, * $C = \frac{\epsilon_0 A}{d}$

$\Rightarrow \tau = \frac{\rho.L}{A} \cdot \frac{\epsilon_0 A}{d}$

$\Rightarrow \tau = \frac{\rho.\epsilon.L}{d}$

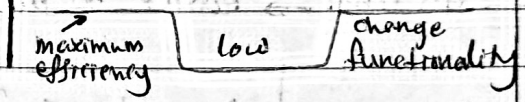
* Reliability :



So we do Validity verification.

→ check if the device still on long time it will function the same [every transistor can switch Complimentary]

In Reliability we use "Bathtub" Curve



describe how the device work

L3 : (Test) Signals

T / 12-2-2019

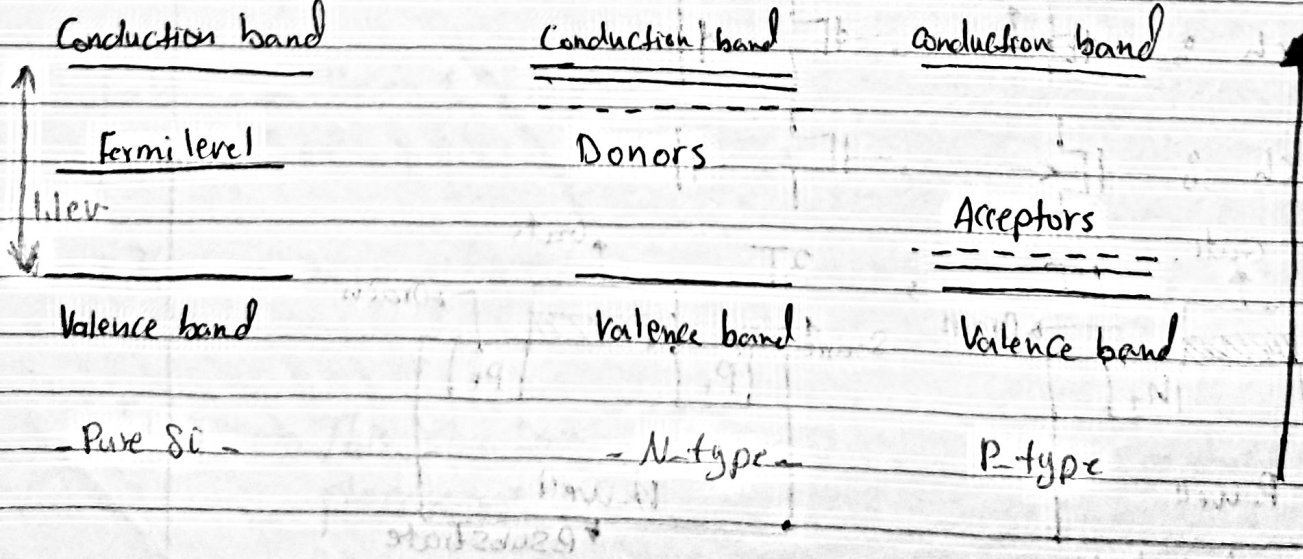
The Semiconductor

NMOS → N-type → free electrons (-ve charge)

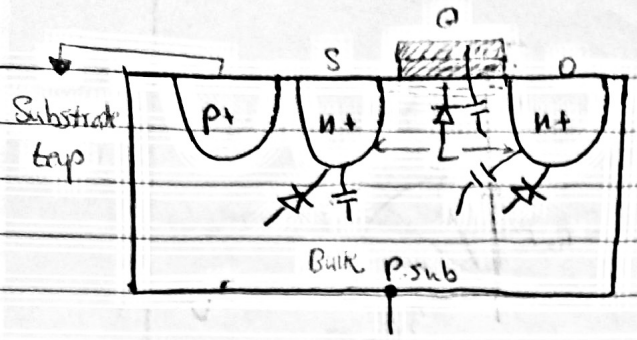
PMOS → P-type → holes (+ve charge)

] after Doping → Slide (5)

* Energy gap : (slide (6) on lect. 2)



The MOS Transistor (Slide 11)



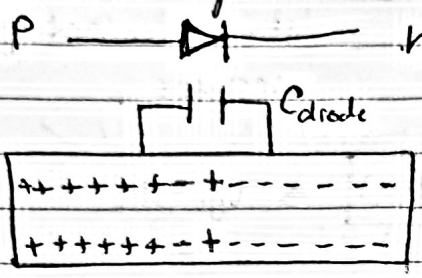
← n-channel device

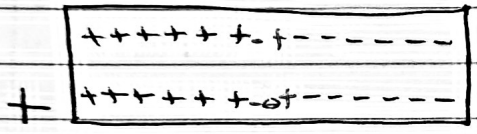
Slide 4

The Semiconductor

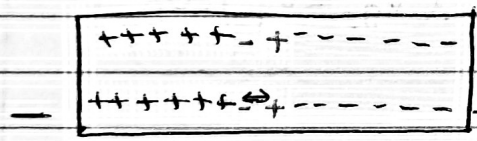
- Silicon Crystal, is composed by atoms with 4 Valance electrons.
- Diamond lattice.
- No free charge.
- Very high resistance.

(Slide 10) **Diode Capacitance**





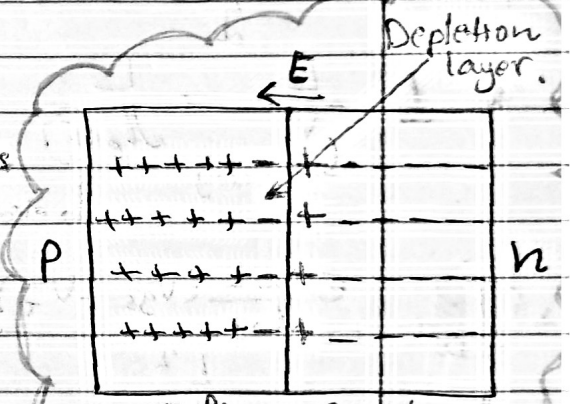
Fowrard bias
C increase



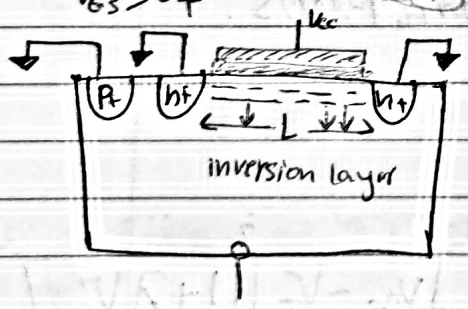
Reverse bias
C decrease

$\downarrow C = \frac{\epsilon_0 A}{d\uparrow}$

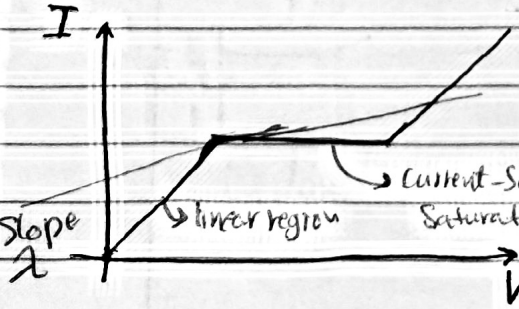
Slide 8



$V_{GS} > V_T$



V_T : minimum Voltage we apply on gate to get inversion layer.



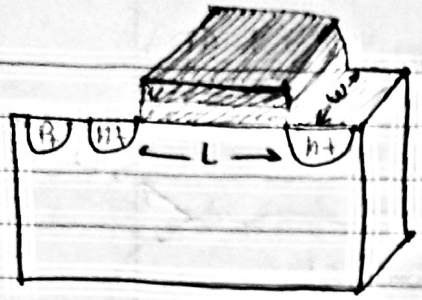
$V = IR$

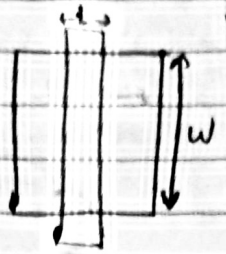
$R = \frac{V}{I}$

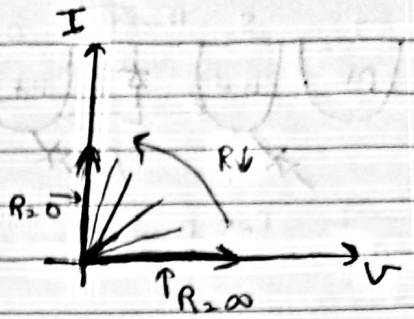
$\Rightarrow W, \mu, t_{ox}, L, \epsilon_0$

W: Width, μ: mobility, t_{ox}: thickness, L: of channel, ϵ_0

these all affect the current in the channel.







→ L increases so R is decreased.

In linear region the device acting as resistor.

time → $i = \frac{dQ}{dt}$, $Q = C_{ox} \cdot V$

- $C_{ox} = \frac{\epsilon (WL)_{area}}{t_{ox}}$, $t_{ox} \rightarrow d$.
- $V = (V_{gs} - V_t) \Rightarrow V_{gs} > V_t$

- $t = \frac{L}{V} = \frac{L}{\mu_e \cdot E} = \frac{L^2}{\mu_e V_{DS}}$

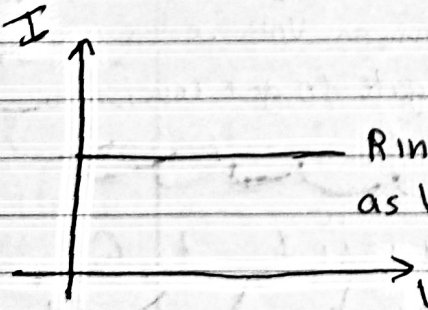
  V → velocity ; E → electric field $= \frac{V_{DS}}{L}$

$$I_{DS} = \frac{\epsilon \frac{WL}{t_{ox}} (V_{gs} - V_T)}{\frac{L^2}{\mu_e V_{DS}}} = \mu_e \frac{\epsilon}{t_{ox}} \frac{W}{L} (V_{gs} - V_T) \cdot V_{DS}$$

→ This is the current in linear region.

- $\mu_n > \mu_p \Rightarrow$ so always NMOS faster than PMOS.

  ↓ more current



- $I_{DS} = \frac{\epsilon W \mu}{2 t_{ox} L} [V_{gs} - V_T]^2 [1 + \lambda V_{DS}]$

in Saturation region.

* Leakage Current: $I_{DS} = I_0 e^{\frac{V_{GS}}{V_T}}$ @ cut-off also @ weak inversion.

[low process] ⇒ Smaller device → I_0 is higher, V_t is lower
So I_{DS} is higher, which affect the power analysis.

* If we need to increase the speed of the device, we usually play with W and V_{GS}.

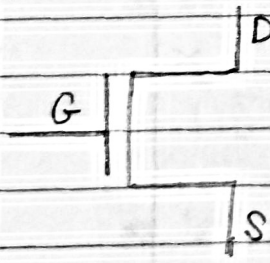
* $V_{GS} > V_{TH}$ ⇒ inversion layer

* $V_T(n) \Rightarrow$ '+' , $V_T(p) \Rightarrow$ '−'

L4:
R | 14-2-2019

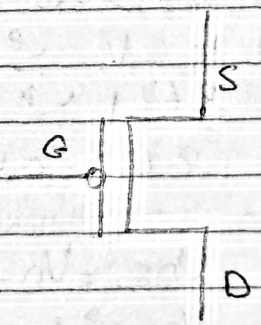
NMOS

I_{DS}
V_{GS}
V_{DS}
V_{Tn}
M_n

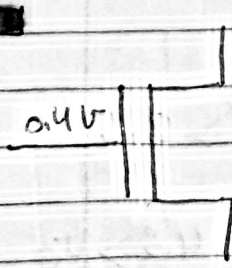


PMOS

I_{SD}
V_{SG}
V_{SD}
V_{TP}
μ_p

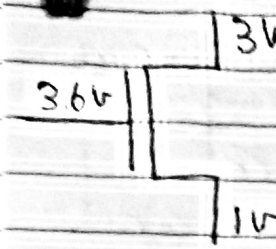


* Examples: [$V_{th} = 0.5$]



⇒ ON/off ?
$V_{GS} \overset{?}{>} V_{Th}$
$V_G - V_S \overset{?}{>} V_{th}$
$0.4 - 0 \overset{?}{>} 0.5$ ⇒ No. So the device off.

1



⇒ ON / off ?

$V_{GS} \overset{?}{>} V_{th}$

$V_G - V_s \overset{?}{>} V_{th}$

$3.6 - 1 \overset{?}{>} 0.5$

$2.6 > 0.5 \Rightarrow$ yes ⇒ So the device ON.

$V_{GS} = 2.6$

$V_{th} = 0.5$

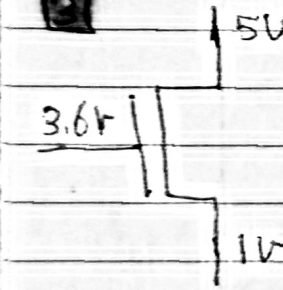
$V_{DS} = V_D - V_S = 3 - 1 = 2.$

$V_{DS} \overset{?}{>} V_{GS} - V_{Th}$

$2 \overset{?}{>} 2.6 - 0.5$

$2 \not> 2.1 \Rightarrow$ No → device in linear mode.

2



⇒ ON / off ?

$V_{GS} \overset{?}{>} V_{th}$

$V_G - V_s \overset{?}{>} V_{th}$

$3.6 - 1 \overset{?}{>} 0.5$

$2.6 > 0.5$. Yes. So the device ON.

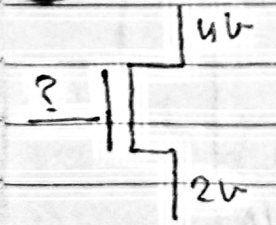
$V_{DS} = V_D - V_s = 5 - 1 = 4$

$V_{DS} \overset{?}{>} V_{GS} - V_{th}$

$4 \overset{?}{>} 2.6 - 0.5$

$4 > 2.1 \rightarrow$ Yes, device in Saturation.

3



what is V_G make the device in Saturation?

$V_G - V_s > V_{th}$

$V_G - 2 > 0.5$ ... ① ⇒ $V_G > 2.5$

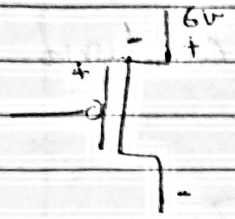
$V_{DS} > V_{GS} - V_t$

$2 > V_G - 2 - 0.5$ ... ② ⇒ $4.5 > V_G$

⇒ $4.5 > V_G > 2.5$

* What is V_G make the device on edge?

$V_G = 4.5$.

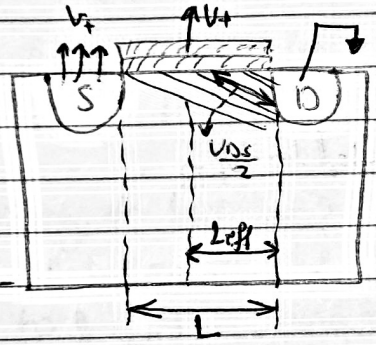


$|V_{SG}| > |V_{thl}|$ check on/off

If on:

Check $V_{SD} > V_{SG} - |V_{thl}|$

Yes → Saturation, No → linear.



"Pinch-off"

channel length modulation

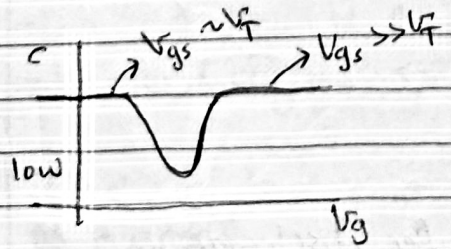
⇒ This is why we put lambda in the equations. [slope]

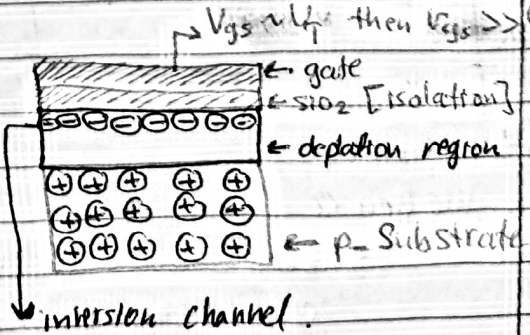
* MOS transistors treated as switches.

or switch with resistance

[Came from linear region since transistor acts as resistor]

- $V_{gs} > V_{th}$ ⇒ inversion.
- Capacitance in depletion region varies due to V_{gs} and its area [channel]





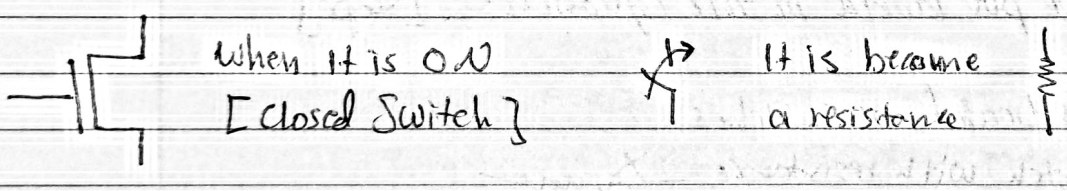
* Why it is limited ?
this limit based on V_{DS}

* V_T :- the minimum voltage you need to apply so you can turn the device ON

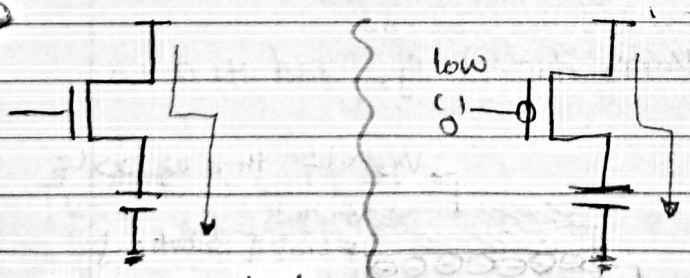
* Depends on :- - Bulk charge [voltage on Substrate].
- Bulk doping - energy [P or N device] - surface
- Charge [SiO_2] - flux.

* How to measure ?
we need : - V_{FB} "forward bias"
- Q_B "charge" → related to V_{FB}
- C_{ox} "capacitance"
- ϕ " of doping"

* What is R_{on} [Rout] ?



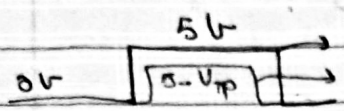
* in Saturation → $R = \infty$ " the device become current source"



the electrons which move never give you a Strong 'one' → so N device always used for discharging and gives Strong 'Zero'.

* What this mean ?

| and we need to charg it .



V_{tp} : V threshold for p device

* AS the device grew in size [higher speed] But [rising time & discharge time differ]

* What is gate leakage ?
Leff will be smaller from the one we draw in layout due to ① the channel ② the overlap with D & S.

* As V_D is higher the drain depletion region increases causing a decrease in $V_T \Rightarrow V_T(V_d)$: V_T function of V_d.

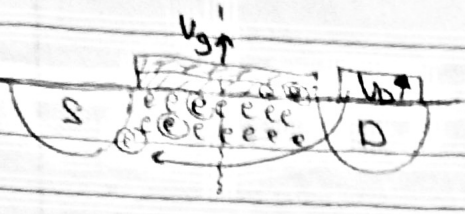
* low $V_T \Rightarrow$ good for Speed .
→ bad for leakage.

* How does the temperature affects the performance of devices?

μ_n, μ_p are functions of temperature T↑ then μ↓ [slower] also T↑ then V_T↑ [more slower] ; we need more voltage to make the device ON, so we need more power & more area.

→ So we need to consider & -
1. ON/off
2. ON → linear, Saturation
3. T, V_T, μ.

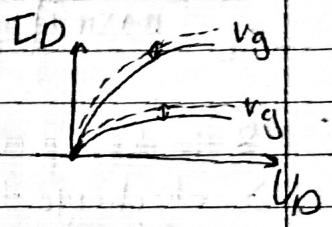
which all affect the design matrices.



"Reliability"

(*) (E) "Hot electrons". It has energy so instead of going to Drain it goes to the oxide. [affect V_t] [slower] affect the Mobility (μ).

(*) Bias Temperature "PchBT"
(*) V_T stability.
(*) gate Stress.

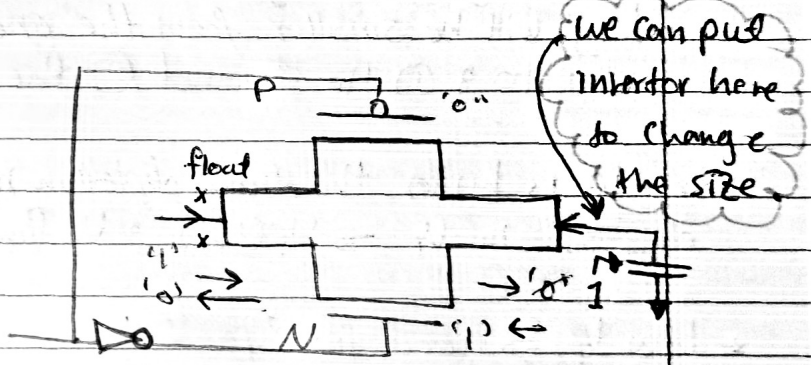


⇒ all these variables make the current less.
⇒ and as V_g higher ⇒ differences is greater.

L5:

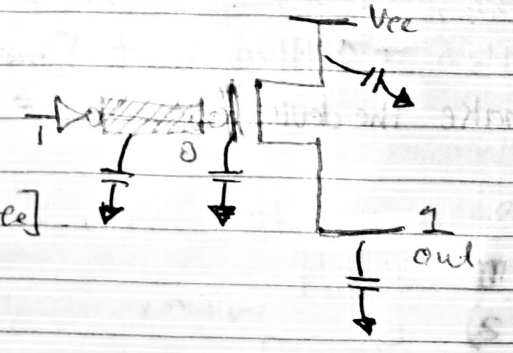
T/ 19-2-2019.

* Pass-gate :-
2-switches PMOS & NMOS to Pass Strong one & Strong Zero to the device.

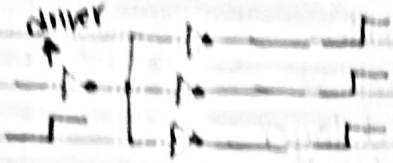


Problems :- ① if I have a float node, It will charge & affect the input. to solve this we can put inverter to prevent the returning data from influencing the input.

* Cload :- These Capacitors Conside My load for the inverter load ⇒ [wire, gate, drain, Source]



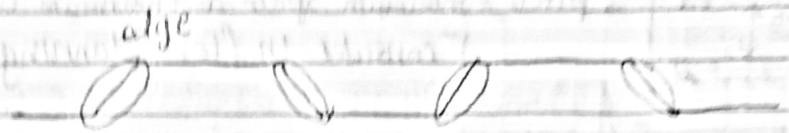
* fanout :



→ It's how many outputs we can connect to the driver, and the outputs with the same slope as the driver.

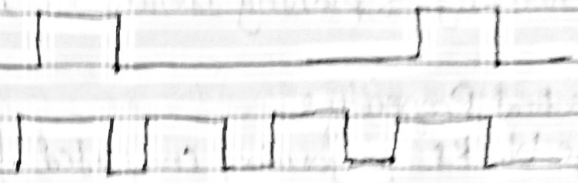
* The driver depends on the load which we need to decrease it

* edge rate :



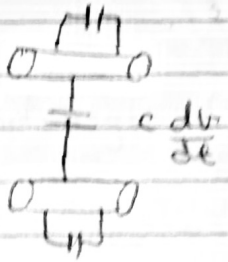
The time it needs to pull-up or pull-down.
* also this time increases when the load is bigger

* AF "activity factor" :

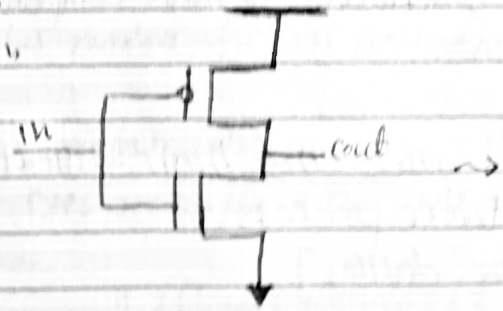


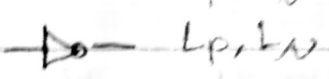
How the signal change with respect to the clk
→ I need to reduce that.

* Coupling Capacitor :
Magnetic field about wires.



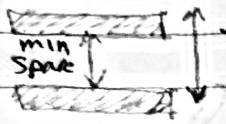
* CMOS :
"inverter"





we need them to find [R, L, C]
→ find load.

- Size of a component depends on W, L.
- as the process is smaller ⇒ more layers.
- Via : to connect metal layers.
- Contact : to connect diffusion with a metal layer.
- as we go up in layer structure, the width of a layer is higher. [less resistance].



→ pitch = minimum space + minimum width
[consider in floor planning]

- Why we do Standardization?
  - Utilization for the area.
  - Routing & planing would be much easier.

- Intellectual Property :
  large blocks performing completed functions.

- PLL ⇒ generate clks.

L6 :
R121-2-2019.

"Productivity"

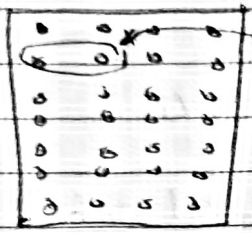
* Technology shrinks by 0.7 / generation. ⇒ more complexity
  ⇝ cost of a function decrease.

* How to design chips with more and more functions ?
  ⇒ we go to lower and lower process
  [flexability to add more devices].

What is foundary ?

Company that do the manufacturing not do the design but some do the both.

- mixed signal ⇒ digital + analog
- photo-lithograph : printed light , they use light to put meterials.
- masks : metal layers.



→ scaled factor (λ) : if I want to change the scale of the design instead of changing all the design . I change lamda only .

- wiering ⇒ micro
- process ⇒ nano
- active meterial ⇒ diffusions.

* non-recurrent cost : cost you once.
* The cost related to wafer and single die [Area].

Yield & Defects

$$Y = \frac{\text{number of good chips per wafer}}{\text{Total number of chips per wafer}} * 100\%$$

$$\text{Die cost} = \frac{\text{wafer cost}}{\text{Dies per wafer} * \text{Die yield}}$$

$$\text{Dies per wafer} = \frac{\pi \times (\text{wafer diameter}/2)^2}{\text{die area}} - \frac{\pi (\text{wafer diameter})}{\sqrt{2 \times \text{die area}}}$$

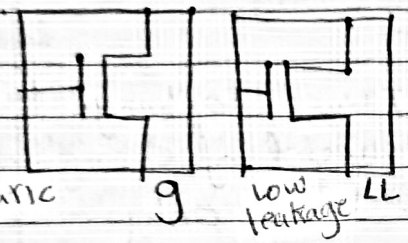
- die yield = $\left[1 + \frac{\text{defects per unit area} \times \text{die area}}{\alpha}\right]^{-\alpha}$

  ⇒ α is approximately 3

- die cost = f(die area)^4

⇒ Circuler wafer ⇒ cost less [lower waste]

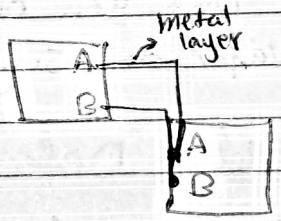
* for same process Ex 90nm



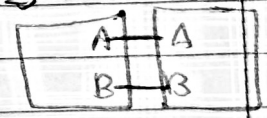
→ lower leakage power
↳ higher V_T
↳ slower [long critical path].

* always in speed path devices, we don't use LL.
* low voltage devices used in memory.

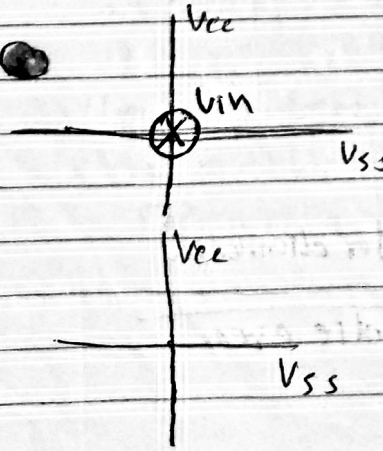
* Power ⇒ ① leakage · ② dynamic · ③ short circuit [static].

- alignment: instead of → 

so we can save metal layers.

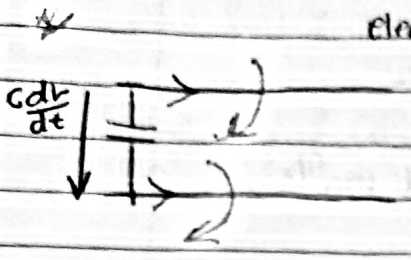
we do ⇒ 

- poly head: It is the metal to poly contact.



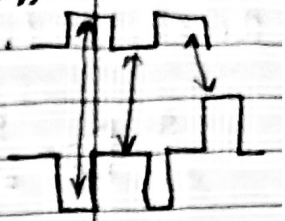
you short Vcc !, you need a hard time to find it.

this called "open", [not connected].



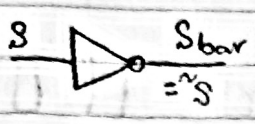
← mutual coupling
= noise
⇒ we put shield [usually for clk signals].

dV ⇒ Voltage difference between wires.



L7: Inverter Design

26-2-2019 / T

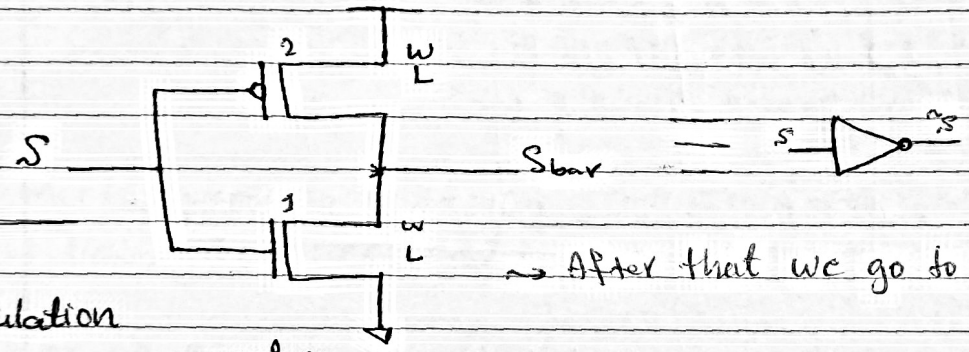


⇒ we need to modle it in Behavioural modle.

[How does the logic behave]

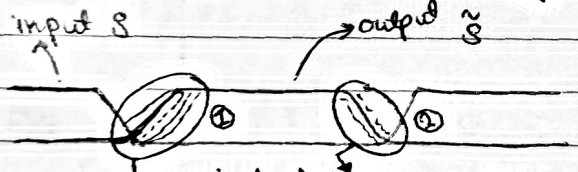
→ RTL done by Architect or Logic Designer then it Converted into Schematic done by Circuit designer.

→ we have transistors inside Schematic level [NMOS + PMOS].



→ After that we go to the layout

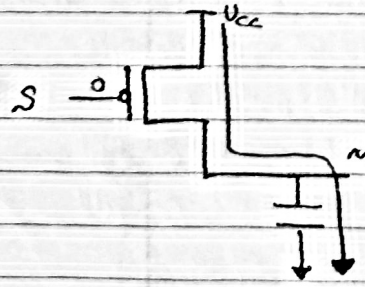
Simulation



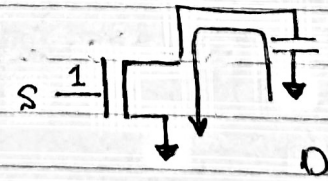
↳ what determine this slope?

N ⇒ take the signal to ground

P ⇒ take the signal to High [it's Connected with V_{cc}]



PMOS : ON
NMOS : OFF



PMOS : OFF
NMOS : ON

$I_o \propto \frac{W}{L}$ → only change this → It's fixed.

→ When S = 0

The Capacitor starts charging and it's charge depends ON W & R of the device.

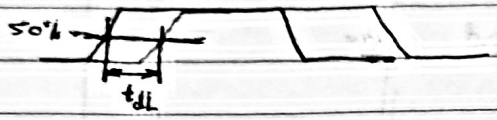
→ When S = 1

The fast of discharging depends on the width of the device.

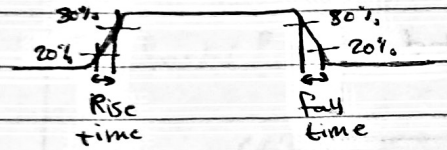
- How fast I can charge the capacitor ?
  Depends on W for PMOS.
- How fast I can discharge the capacitor ?
  Depends on W for NMOS.

* W_{10} for N device take the signal to Low faster than W_5.
  W_{10} for P device take the signal to high faster than W_5.

➡ How do we measure delay between input and output ?

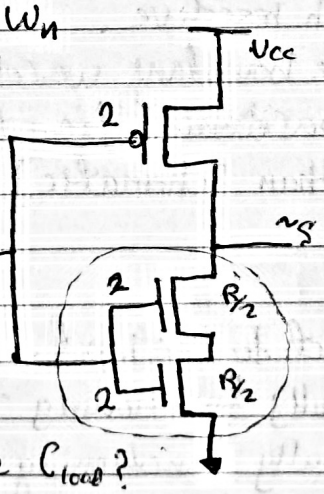


➡ How do we measure slopes ?



* we can take 10% to 90%

* If I want rising slope to be equal to falling slope ??
  → make the width of PMOS larger [NMOS is faster]. So, if the ratio $\frac{M_n}{M_p} = 2 \rightsquigarrow W_p = 2W_n$
  * L is fixed.

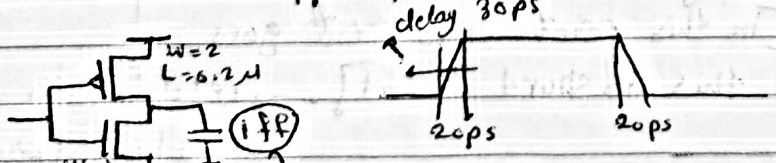


This is what we call Sizing the device.

* we still need to have the same rise and fall slope.

$I \propto W$, $V = RI$

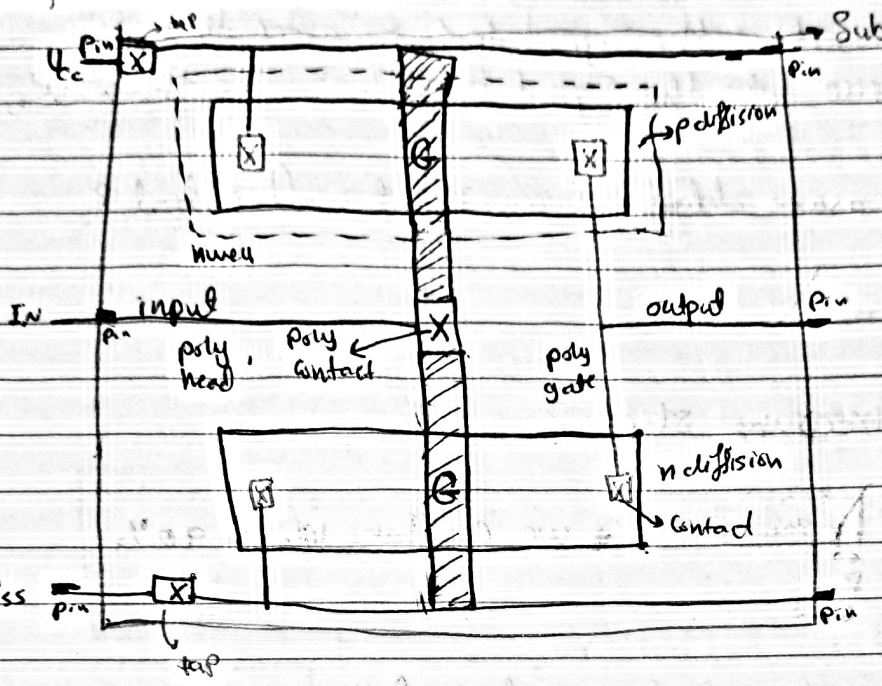
* What will happen if we increase the C_{load} ?



change to 1off ?? ⇒ The delay will increase to 50 ps like example, and slopes could be 30 ps.

⇒ But if we don't want to affect the delay?
→ increase the width of devices [strong driver]
⇒ The area will increase ↑ . delay → constant.

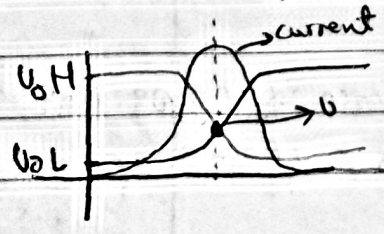
* layout for the inverter :-



* input came from metal layer 1
* Contacts usually high resestive.
- every thing touch the boundary we called it a pin.
* we extract the capacitance from layout.
[more acurate than schematic].

* VTC Curve :-

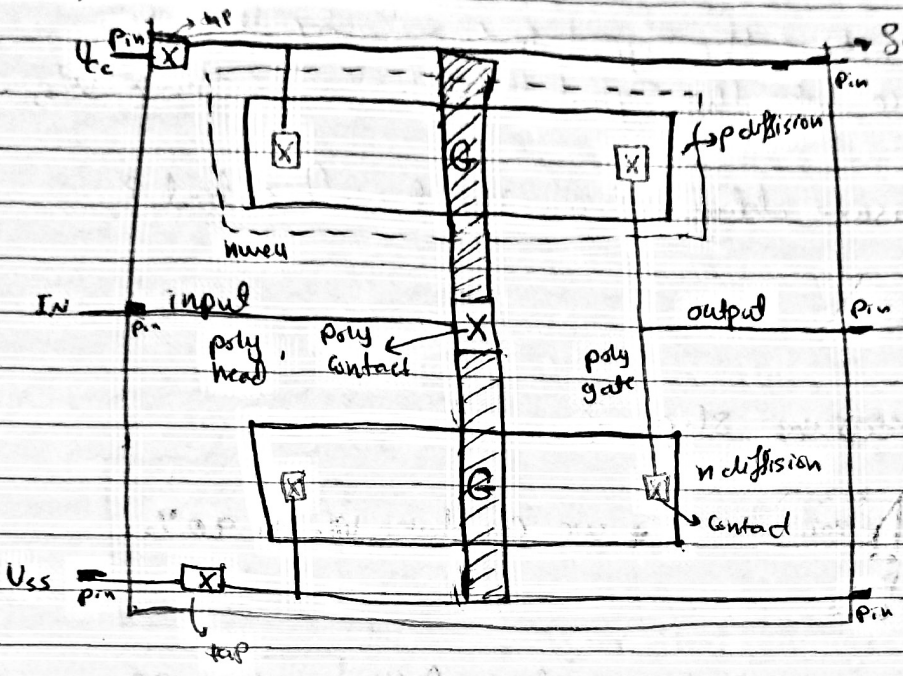
The output with respect to the input..
V_{OH} → when PMOS fully functioning.
V_{OL} → when NMOS fully functioning.
V_M → Both of them are ON. [in this point the current goes from Vcc to Vss and this is short circuit]. → related with Short circuit po



is better → the two devices together were ON for a long time → consumed more power

⇒ But if we don't want to affect the delay?
→ increase the width of devices [strong driver]
⇒ The area will increase ↑ , delay → constant.

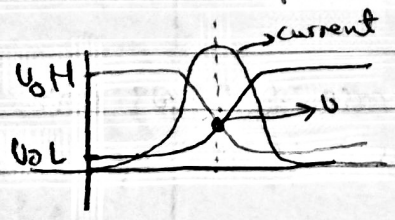
* layout for the inverter :-

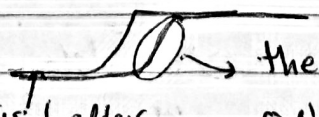


* input came from metal layer 1
* Contacts usually high resestive.
- every thing touch the boundary we called it a pin.
* we extract the capacitance from layout.
[more acurate than schematic].

* VTC Curve :-

The output with respect to the input..
V_{OH} → when PMOS fully functioning.
V_{OL} → when NMOS fully functioning.
V_M → Both of them are ON. [in this point the current goes from Vcc to Vss and this is short circuit]. → related with Short circuit power



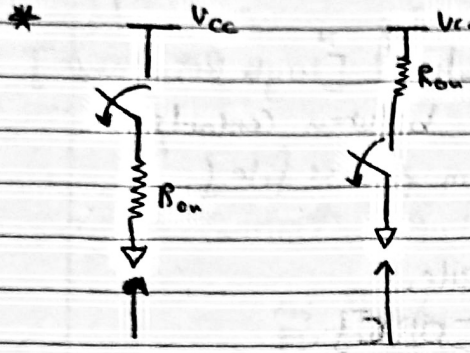


→ the two devices together were ON for a long time → consumed more power
is better

* If I have the current at mid point and I have W_n and need to find W_p ?

→ $I_{D_N}Sat = I_{D_P}Sat$

$W_N \rightleftharpoons W_P$?



$V_{OH} = V_{cc}$

$V_{OL} = 0$

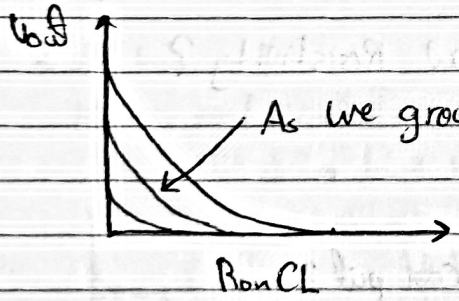
$V_M = f(R_{on_n}, R_{on_p})$

affect full time — affect Rise time

R_{on} ↑ faster full time.

more power (I^2R).

* $t_{PHL} = \ln\frac{1}{2} R_{on} C_L$



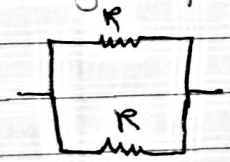
L8: Layout
T / 5-3-2019

* Paint from set of different colors. [based on the circuit].
- The designer try to build a library cell of the inverter.
- make sure that everything does not break [cannot correctly]
* metal 1 → Connection with diffusion or poly
(↑ contact) (↑ poly-head) [high Resistance]

→ from DR we get the dimensions and distances between contacts, in addition of the amount of overlap. [or from process file].

→ Everything I put in the layout it has a rule.
⇒ these rules programmed in the tools [DRC].

- any movement could make a short.
→ why to put two contacts instead of one?



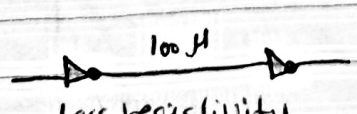
reduce the resistance to R/2, and this is limited to the DR → same to poly-head.

- Why we don't use another material with low Resistivity?
⇒ more expensive.

* In the process file:-
- number of layers: How many metal layer I can put?
- poly Spacing: the distance between two polys.

⇒ The higher I get in metal layer, the resistance decreased
→ So, that why we use higher metal layer to do routing between far blocks.

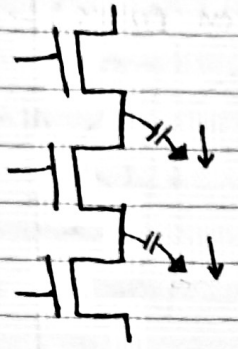
* $\frac{RCL^2}{2}$



- less resistivity

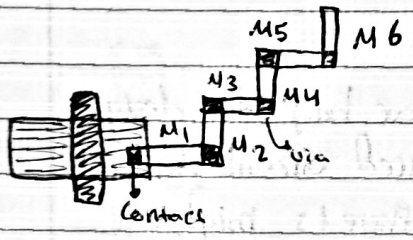
- high resistivity

* Metal layers divided into binary groups to enable vertical and horizontal connection.

* 

I need to reduce this capacitance.
→ So, we share the diffusions.

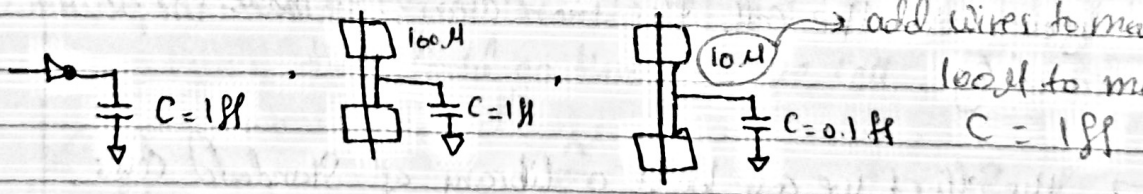
* How to connect diffusion with M6?



* geometry class to generate the tools based on DR.

* Fram view ⇒ to deal with pins not all things inside.
* process scalling → using lambda [scale factor].

* Antenna:
Wires not needed, sometimes we need it [to add capacitance]
⇒ match the delay:



* Density Rule:
amount of metals, poly there is ratio between all the materials
⇒ this is important for the machines that would cut.

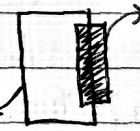
* Why we have two input for inverter?
→ to connect horizontaly or verticaly. [to get the flexibility how to connect]

L9: Design Rules
RI 7-3-2019.

- we do Design Rules to be able to tolerate fabrication errors:
  - mask misalignment
  - dust
  - process parameters
  - Rough Surfaces.

* masks: → metal layers [M1, M2, ...].
* Dust can make short circuit so we make cleaning rooms.

* half design Rules:
anything on edges [excludes pins] must be far half the design rules, because it might have another cell placed around it.



→ outside the cell, because there is might be another cell share it.

Shared metal
or Shared diffusion
↓
make width smaller.

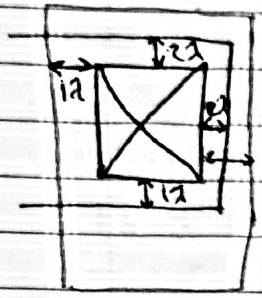
* you should make the size of cell smaller as possible, because the bigger you make the cell, the more areas, the more the power and the more delay.

* to minimize the effort we can build a library of Standard cells.
  - to optimize cells very well.
  - to make the tool make the layout for me.

→ thats what we call thensysics flow.
⇒ we make all the cells with the same hight.

* We should always put PMOS at the top to connect the Vcc in one line.
- one substrate for whole the cell.

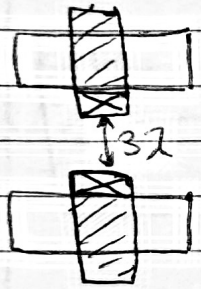
*

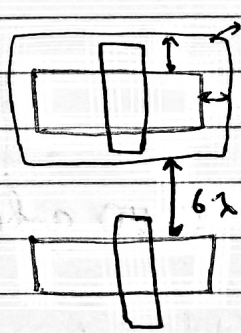


→ the more the area you put, you increase the Capacitance.

* poly uses a width of 2λ → If I have 10nm process, then λ = 5nm.

*

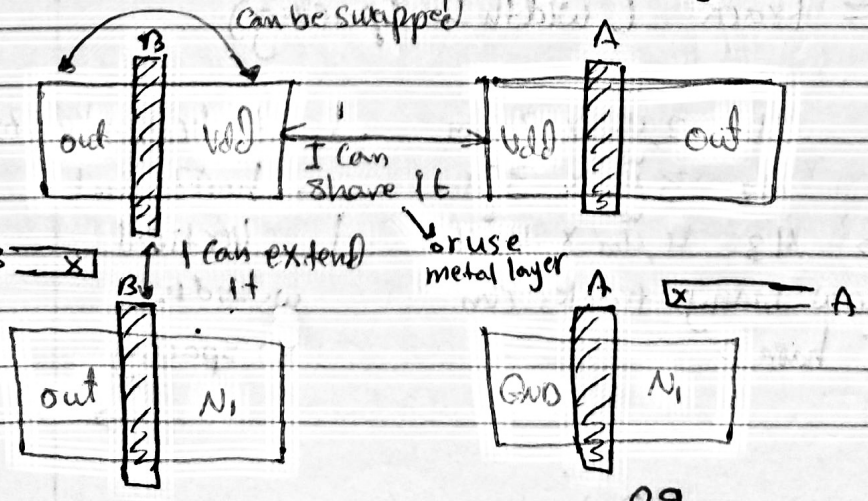


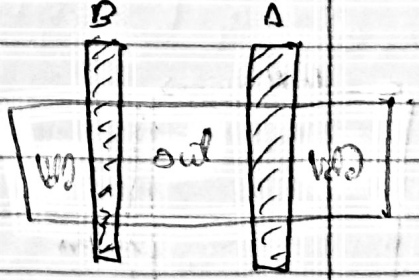


*

Why so long? you can chang the size as the load of the driver can be changed.
→ another reason, to do Standarization.

* Limitation : => ① Device limited. ② Metal limited → metal layers that could go through the cell.



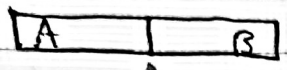


this is the better, because this is Cout → load and I want to minimize it.

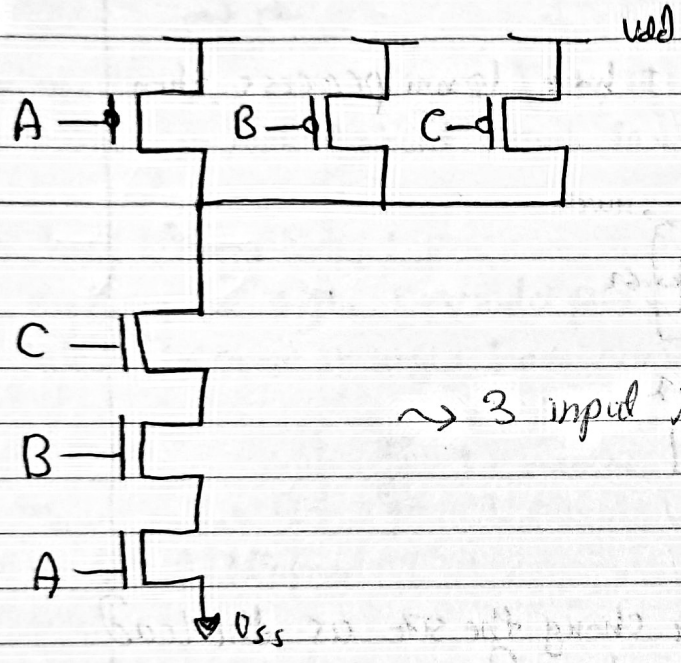
L10:
T/12-3-2019

* What does it mean the width of the devices?
→ As I did the Simulation, this is what will work for me.

*



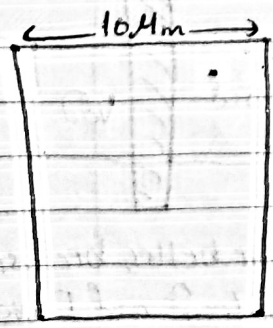
*



→ 3 input NAND gate.

* Wiring Tracks:
Space required for a wire, and it used to calculate hight of the cell [metal limits]
→ depends on the number of tracks and metal layers.
→ wire + distance = track (width + spacing).

*



	w	s	
M1	2 μm	2 μm	Vertical + horizontal
M2	3 μm	3 μm	horizontal.
⇒ M3	4 μm	4 μm	Vertical.

How many tracks can I have?

w: width
s: spacing

→

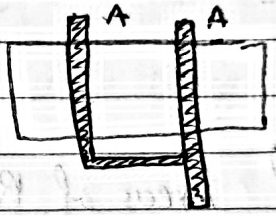
→ each metal layer has a track for its own.

m1: in horizontal ⇒ $\lfloor \frac{20}{4} \rfloor = 5$ tracks and I need to make spacing from the edges (2 μm from above and 2 μm below) → vertical we do 2.

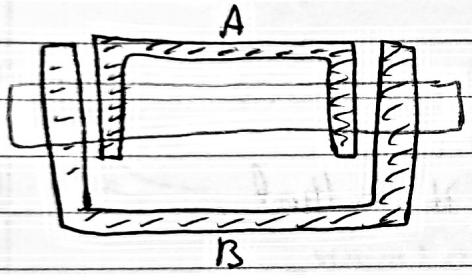
m3: in vertical ⇒ $\lfloor \frac{16}{8} \rfloor = 1$

* Area estimation based on number of tracks.
* Stick diagrams:
  Is a way of sketching the layout [with out doing it].

* legging → high width into two legs but save the performance.



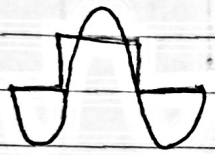
* two devices:



L11:
R14-3-2019

* How to evaluate our design?
→ cost → Reliability → Speed / performance [delay / freq]
→ power.

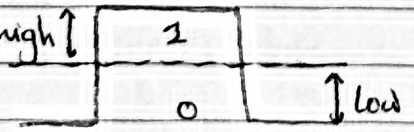
* Reliability :-
- real world is analog and we try to process things into digital. [we care about the signal is high or low].



Areas of Reliability

[* How Signal switch and how frequantly it do that.
* Also Routing
* Things that affects the Reliability :
→ inductive coupling : we have 2 signals, next each other and swith very often $\left[\frac{di}{dt}\uparrow\right]$.
→ capacitive coupling : → depends on $\frac{dv}{dt}$
→ power and ground noise.

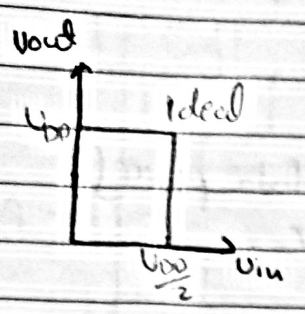
* Noise and Digital Systems :-

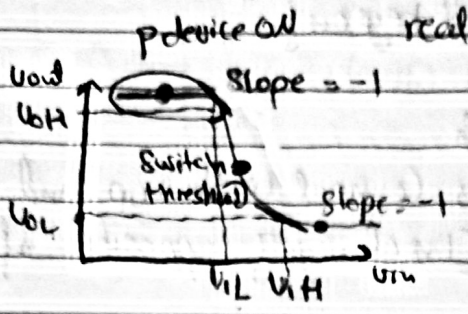


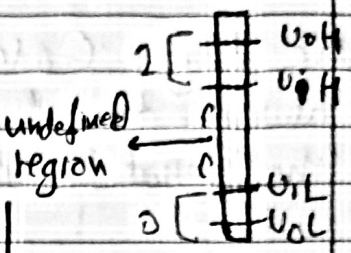
* you need to set up your design Consistance. [where high and where low]
→ Conider the noise.

* digital signal is discreat value of analog signal, but I set a threshold [where Consider high and where low].
→ analog is accurate but we do not take.

* for small amount of noise, out put noise is less than input noise.



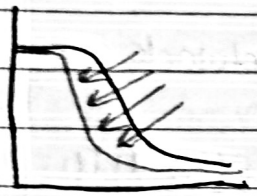




↓ both transistors are ON in this region.

* noise margin
  - high → $NM_H = V_{OH} - V_{IH}$
  - low → $NM_L = V_{IL} - V_{OL}$

→ does the size of the device affect the noise margion?



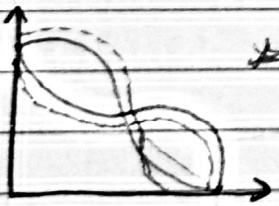
∴ Size affect.
also the slope of the input.

*

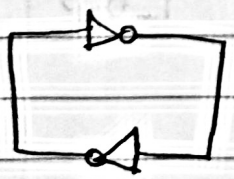
nmos: act normally but pmos not charge to highest rulus

and so on... the noise amplified.

→ the size of them should be proportional to each other.



* I want it to be bigger → the smaller → more noise can't be rounded.



→ keep change the size until I see the eges.

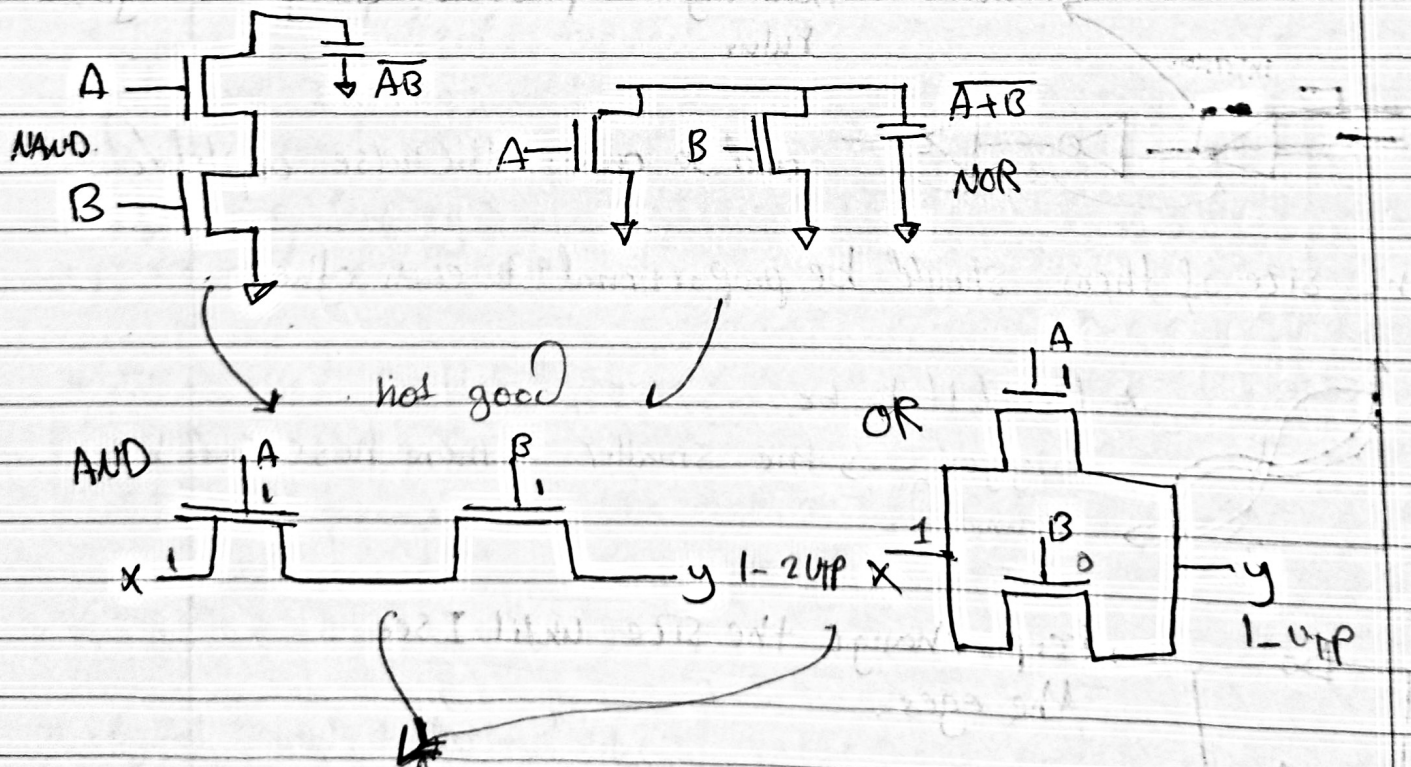
* NM : the bigger the noise margin the better.
[less undefined region]

* Static Logic Gates :
→ always the devices connect with V_{DD} and V_{SS}. [static power]
→ the signal static [unchange] opposite of dynamic.

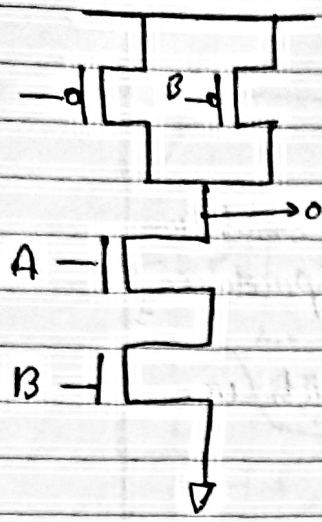
→ CMOS : we have two networks, one is connected with PMOS and the other with NMOS.
high ← Complement to each → low
other

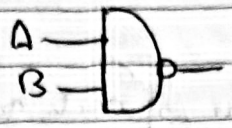
(PDN) pull up pull down (PUN)

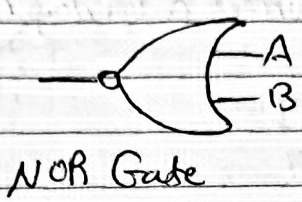
→ So I build the Pnetwork then mirror it to Nnetwork or the opposite.
→ If I use NMOS to pull up → V_{OH} will decrease → less NM

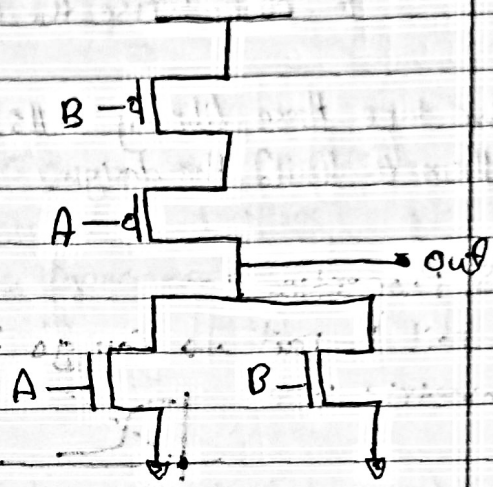


not good → does not give strong one.









* If you have to choose between NAND, NOR what to choose? Why?

- NAND → nmos is faster, so I prefere the nmos to be in Series [long path].

* to design CMOS complex gates → make pull up and pull down networks.

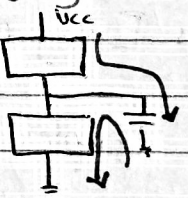
L12 :
T / 19-3-2019

* Library cells : more than one version of each gate.
  → first we do schematic then layout then do simulation to characterise the cell. [Know slopes, delays, Capacitance]

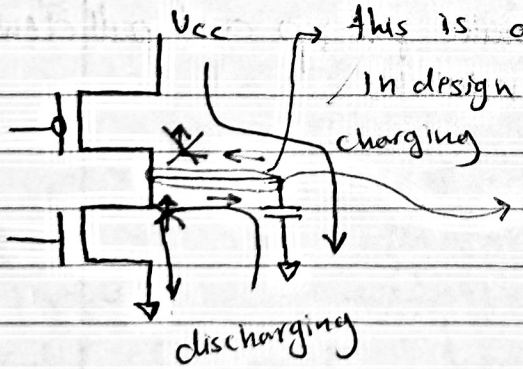
→ we take these library cells, write a verilog code which use them to build the design.

* CMOS Properties :-
  full rail to rail → Can go fully to up or low [1 or 0]
  Symmetric VTC 

Propagation delay is function of load capacitance & resistance.
  ↳ between change in output due to change in input.
NO static power dissipation just charging and discharging [Power Stored]
Direct path current during switching.



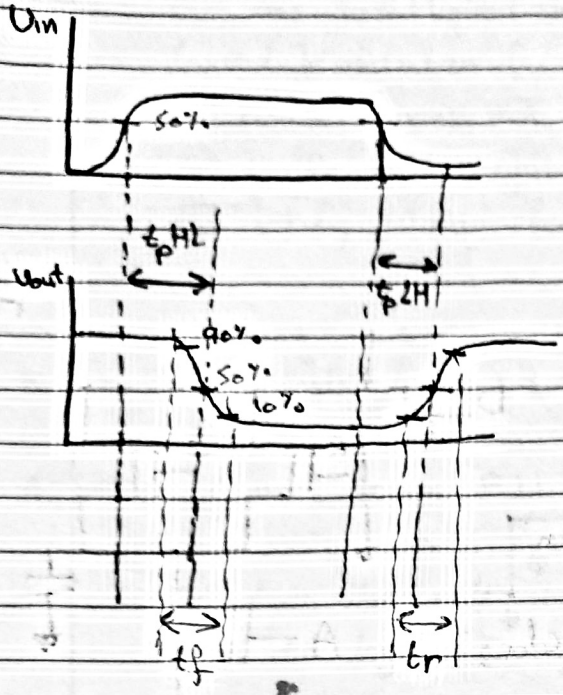
just these ways self heat



this is a problem in design.
charging [self heat] temperature ↑ R ↑ delay ↑
→ the width of this wire must be larger.

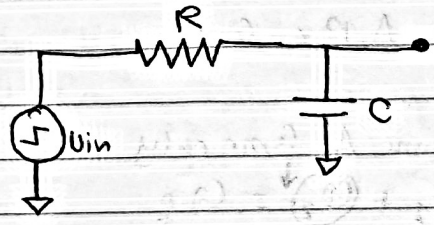
* Example : let $f = \overline{(A+B)\cdot C}$
  pull up = $\overline{((A+B)\cdot C)} = \overline{(A+B)} + \overline{C} = (\overline{A}\cdot\overline{B}) + \overline{C}$
  pull down = $\overline{f} = (A+B)\cdot C$

→



* Delay Definitions :
We use CMOS. we don't using dynamic but in certien pakages

* tr = 10 ps
tr = 5 ps ⇝ this is better
· [larger width at PMOS] or [less capaciter load].



$V_{out} = (1 - e^{-t/\tau})V$

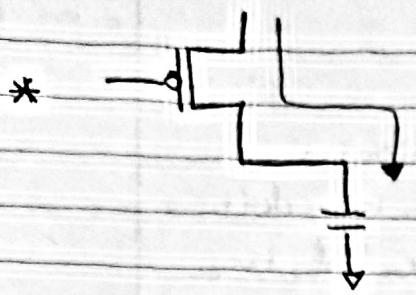
$t_p = \ln(2)\tau = 0.69RC$

* Power disipation :-
- instaneous power : $p(t) = v(t)i(t) = V_{supply} \cdot i(t)$
- Peak Power : $P_{peak} = V_{supply} \cdot i_{peak}$
- Average power : $P_{av} = \frac{1}{T}\int_{t}^{t+T} p(t)\,dt = \frac{V_{supply}}{T}\int_{t}^{t+T} i_{supply}(t)\,dt$.

→ When we change the width we must look at both power + delay
→ Power-Delay ; we do Optimization.

* Energy = Power * time.
→ $Q_{charged} = C_L V_{DD}$ So the $energy = Q V_{DD} = C_L V_{DD}^2$

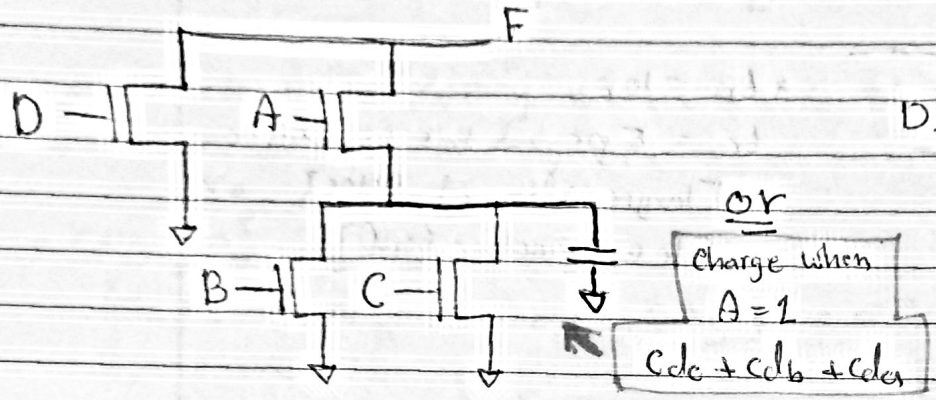
$E_C = \frac{1}{2} C_L V_{DD}^2$; half of the time I do charging and another half do discharging.

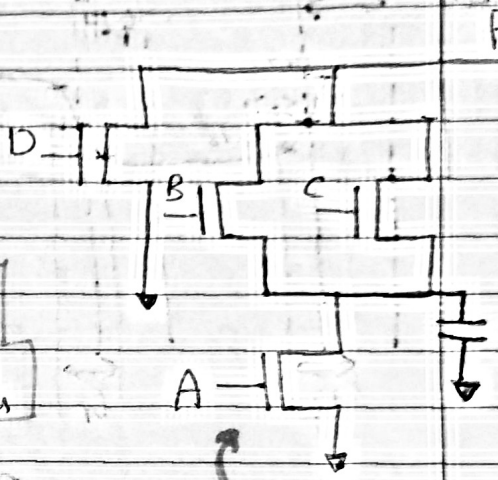


$T_p = 0.69 R_p C$

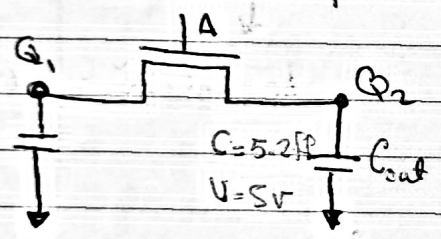
$P = C V_{DD}^2$

change of C affect both





→ So it depends when A, B, C come



Assume A come only

$\uparrow Q_1 + Q_2 \downarrow = Q_f$

$C_1 V_1 + C_2 V_2 = C_f V_f$

$0 + 5.2 ff * 5 = 6.2 ff * V_f$ a big deal.

$V_f = \frac{5.2 ff * 5}{6.2 ff}$ ↓ not good

→ make the signal that will take the out to zero the late.

Charge Sharing Problem in Dynamic it is

L13:

R/ 21-3-2019.

* Sharing metal tracks to minimize its number and so minimize the width of cell.
* Sometimes we go to another level of metal layer instead of add more taks to avoide the increase in the width.

[see stick diagrams on slides]

* LVS: Be sure that layout match the circuit built in schematic.
  logical verification for Schematic.
  1. check Connectivity
  2. Width and Length.
  3. Verify timing delay → not very accurate in Schematic.

  Compare the netlist of each of them

* In Spice netlist ~> m_{t_1} or m_{t_2} or .... : depends of the process.

* EDA: Convert from RTL / Verilog to Layout [Synthesis flow]

  what is the advantage and disadvantage?
  in Control logic [Change very frequantally] — don't do reduction or optimization.

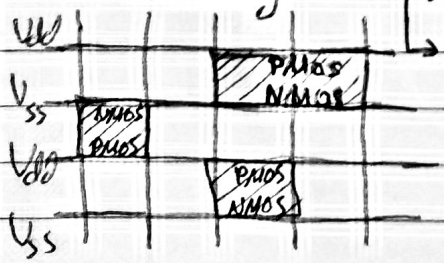
→ each process has its own libraries, and in each library we have different Families

for: Speed — power — Area] depends on what we do optimization for

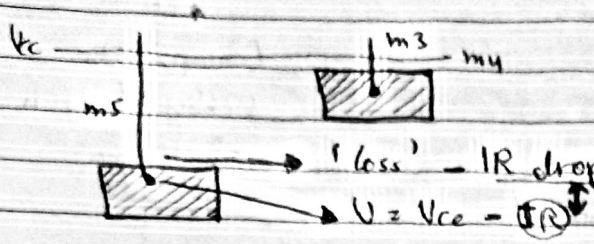
* Test Creation: test the logic not Schematic
  → do different test patterns and in sure that the output match our expectation. Hard work testing.

L 14:
T/26-3-2019

* Physical Synthesis Steps:
  - floorplaning: interface and shape of design + Supply network [V_{dd} & V_{ss}]
  - Placement: minimize the total area and interconnect length.
  - Routing
    → we do it based on how we do the floorplanning.
    → we use grid, 'unit tile' of floorplaning build this grid
      ↳ cells in library designed to be multiple to unit tile.
      → make the design compact more



→ Power grid : make power network [power delivary]



* formal Verification : just test the functionality of the design.
* timing after Pr-layout : STA Static timing Analysis.

→ physical Synthessis : we optimize the cell as required by the designer.

* clock Delay Problems.
  - all clock pins are driven by a single clock.
  → we need it to reach all components without delay.
  → Sometimes we allow some clock skew [no problem]

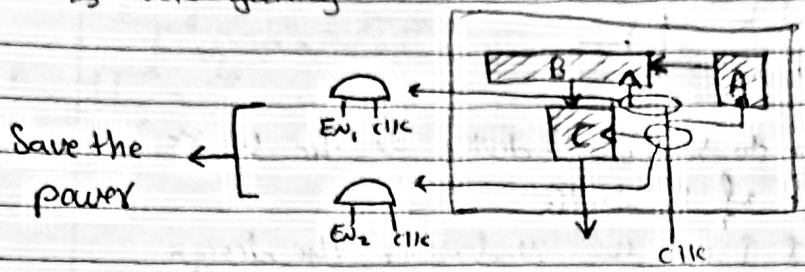
* clock distribution [clock tree].
  - buffer to Compensate the delay
  - how to Connect this tree.
  ] we have special algorithims.

* Digital Standard Cell Library (DSCL)
  AND, OR, NAND, NOR, ...

→ multivoltage and multithreshold libraries ⇒ to do optimization
  high/low
  Ex. low leakage : pick high V_T but slower.

* low power design techniques

↳ clock gating



at first clk, B & C no need to work, so I can turn them off by Enabels.

↳ Static multi-Voltage Same as clock gating but enable on voltage [switching of voltages]

→ put power switch but it will cost you.

L:15

R/28-3-2019

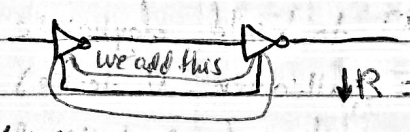
* Design corners :

Design Setup ← at the time we run the charachterization for any cell, we want to check the speed, slope, capacitance ...

[min, Typical, Max] ← Voltage transfer curve reflected to what you choose.

* PD: Power times delay

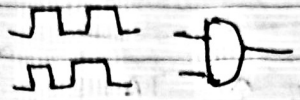
- Specifications
- data sheet
- Circuit design → do simulations
- Layout → check DRC
- extract prasatics
- we generate netlist
- Library characterization. find slopes, delay, power delay, noise margin ...



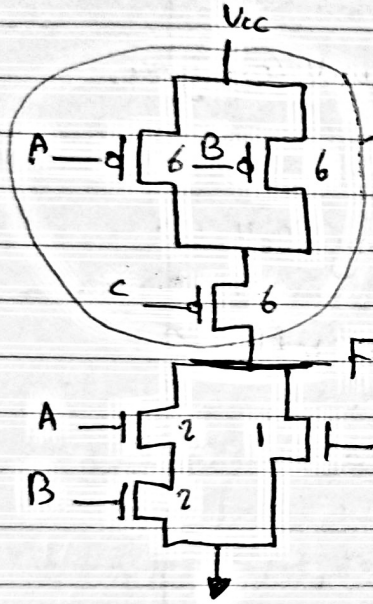
$\downarrow R = \frac{\rho L}{A_\rho}$, C↑, P↑

Slides-L8 ⊛ Dynamic Gates.

⊛

- Static design → 

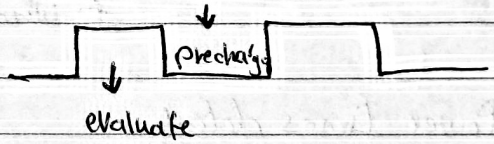
- Dynamic design → we try to make N-device do the worke insted of P-device
So. if I use the N-devices, the design Should be faster and less area.



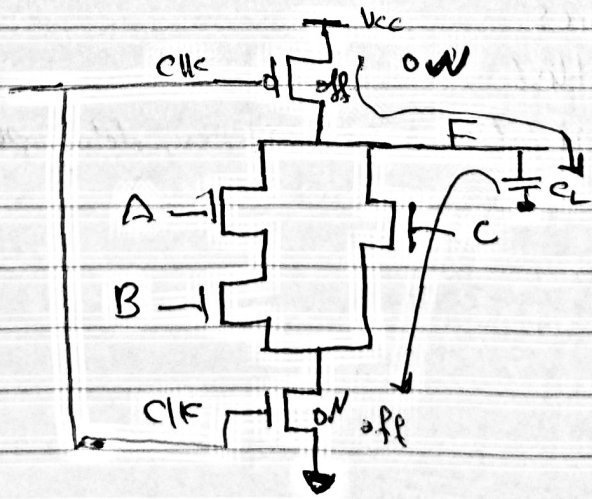
→ we try to eliminate the P-devices

Ex: Sizing → $\frac{N}{P} = \frac{1}{3}$

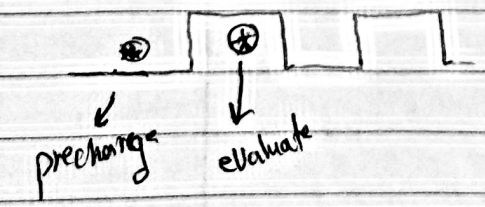
Precharge, evaluate

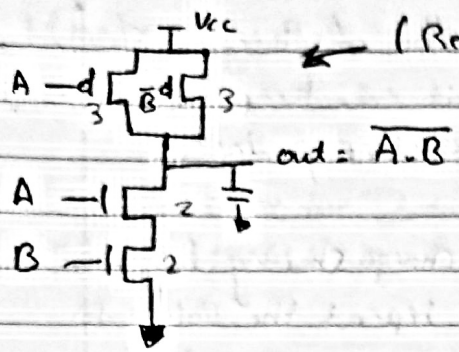


⊛ to build dynamic logic, we use (pull down - N-device) and remove (pull-up-P) and put P-device & N-device with clk.



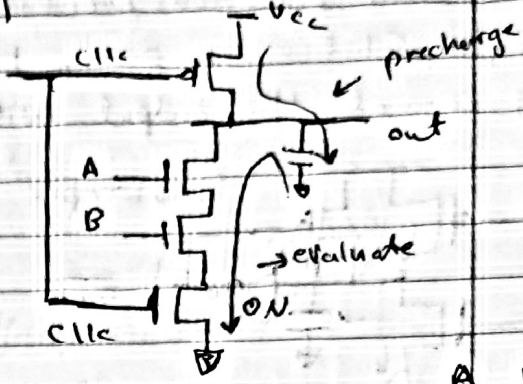
⊛ precharge: charging the capacitor before starting the evaluation.



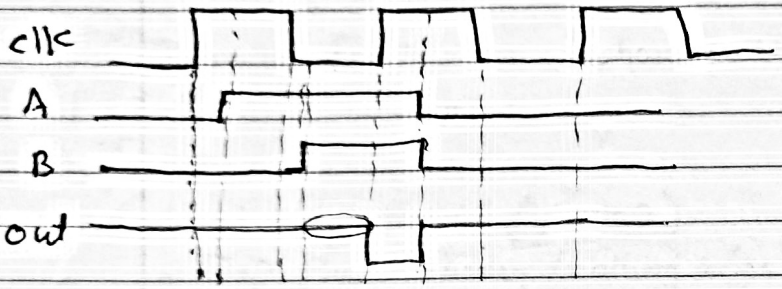


← (Regular NAND-gate)

Build Dynamic logic →

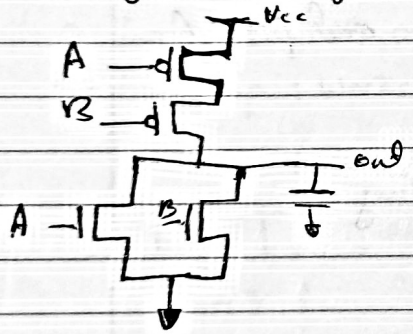


⊛ Timing diagrams.

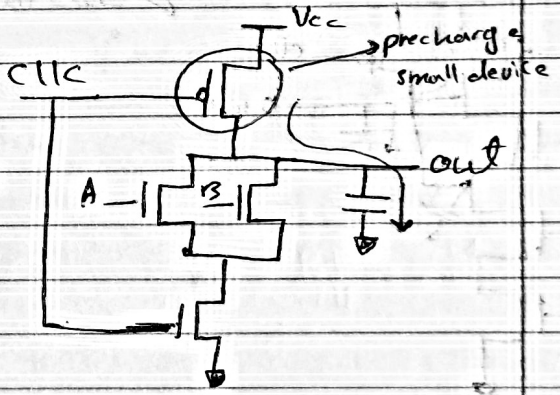


A	B	out
0	0	1
0	1	1
1	0	1
1	1	0

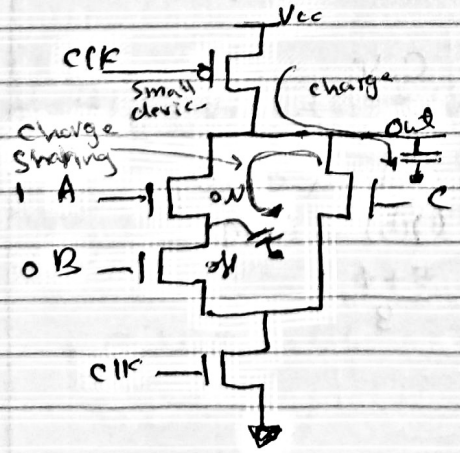
* Regular NOR-gate



Dynamic NOR-gate →



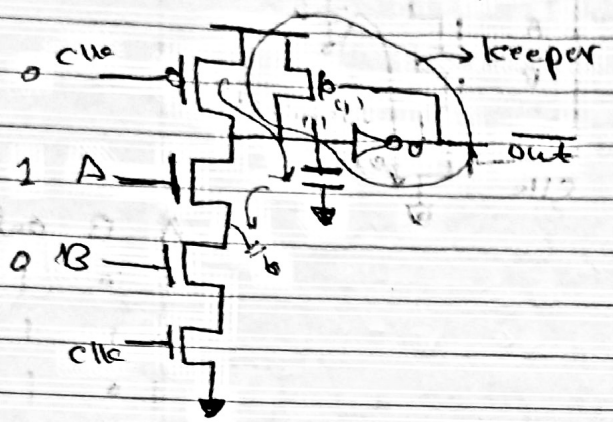
(*) →



← This is happen for each precharge and this is the problem in dynamic logic.

⊛ The problem is ; after each precharge, the charge sharing happens and we loss the charging

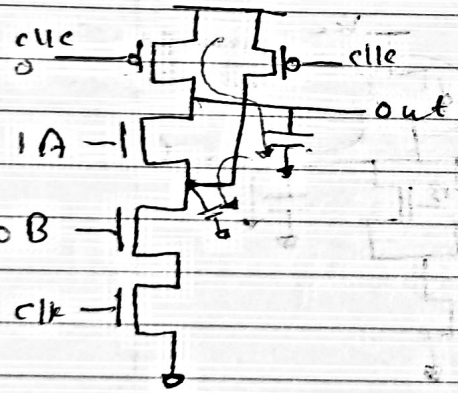
① The solution of previous problem is by making the discharging slow (or prevent it).
So, the solution is put a keeper circuit that can keep the value.
② by connect the output with inverter and another p-device.



'Dynamic NAND'

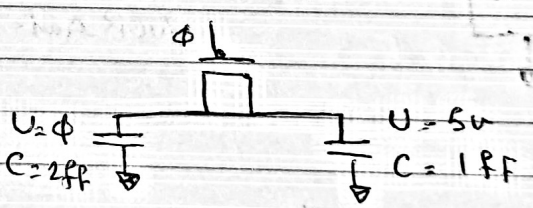
* when the charge sharing happen, the keeper repeat the value to 'one'.

② another way:



← but this increase the area and power

— charge sharing —



$$Q = C.V$$

$$Q_f = Q_1 + Q_2$$

$$(C_1+C_2).V_f = C_1.V_1 + C_2.V_2$$

$$(3)V_f = (1)(5) + (2)\Phi$$

$$V_f = \frac{5+\Phi}{3}$$

L16: Slides_L9

(*) Sequential Gates.

(+) - Use to store the data, because the data is changing, if we don't store it, the data will loss.

(-) - use in finite State machine to represent different States of the machine.

(*) - used in pipelined machine to designate pipestages.

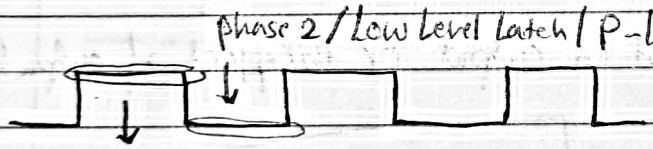
→ Instruction Fetch → Instruction Decode → Execution → STOR ...

* Types:
- Latch (phase 1 / phase 2)
- Flip Flop (Rising / Falling edge).

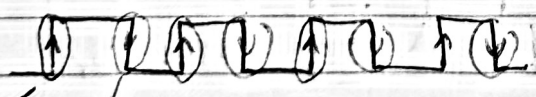
(+) Latch : works on Level

(-) Flip_Flop : Works on edges.

(*) in design we use Flip Flop, if we want a sequence on the edge not in all period.



phase 1
High level latch
N- Latch



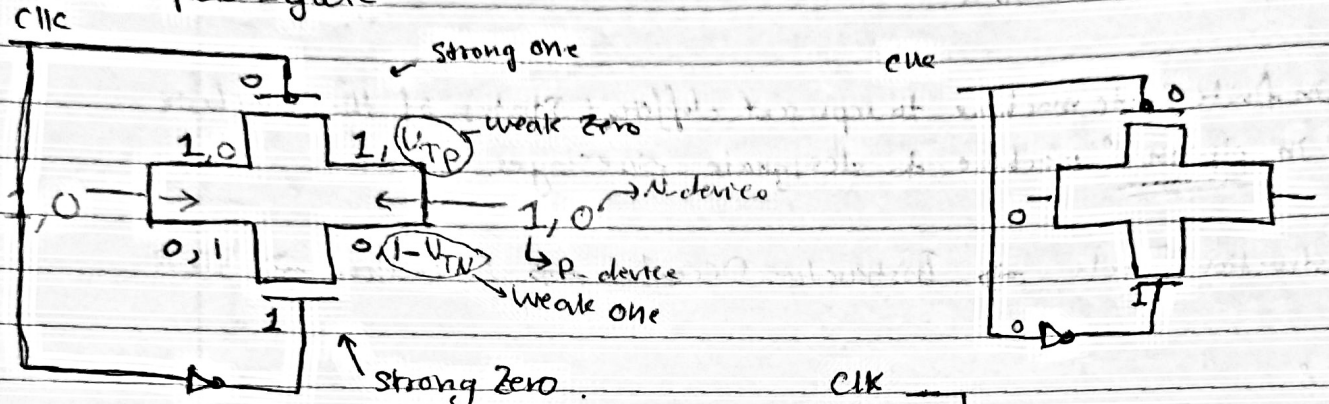
- Works on rise edge.
- Works on low edge.
- Works on two edges (high & low at the same time).

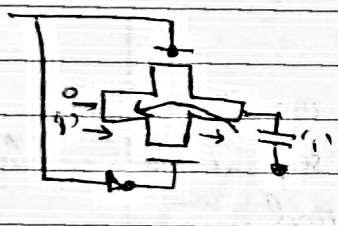
* Enabled Latch → (1) → work
'0' → not work

② Synchronous Set / Reset → based on clk
'1' '0'

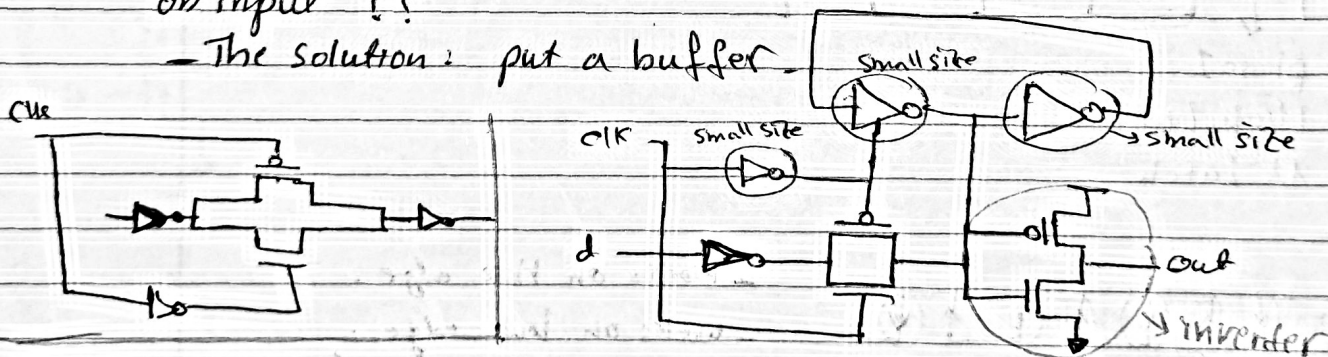
③ Asynchronous Set / Reset → not based on clk.

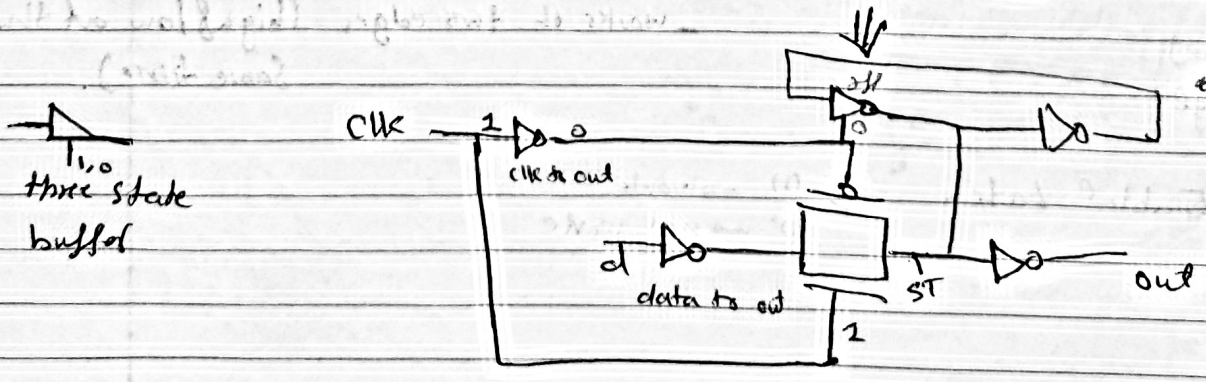
④ pass-gate





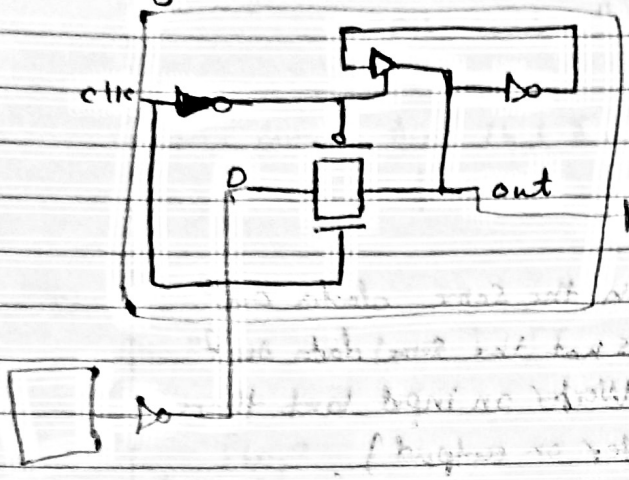
⑤ How I prevente the effect of input on output or the effect of output on input ??
- The solution: put a buffer.





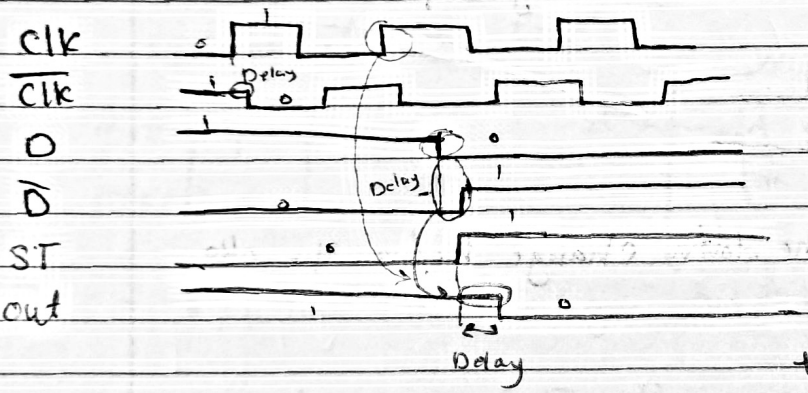
(N-Latch) → clk high — phase 1

* Sometimes, we build the latch without the inverters on input & output.
why?



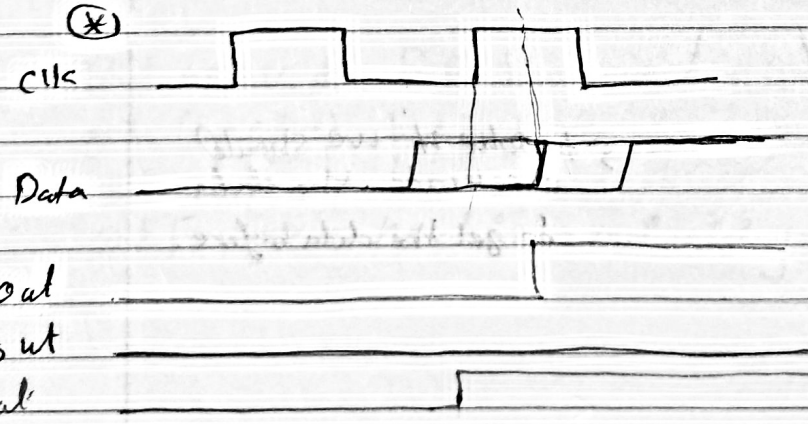
We use this design, when the data comes from another logic, and the output goes to another logic. In this case the inverters are already exist.

* Timing Diagram of N-latch



* hold time
after the latch closed and get the data, I don't want the data change after it. Because the variation (delay) on clk.

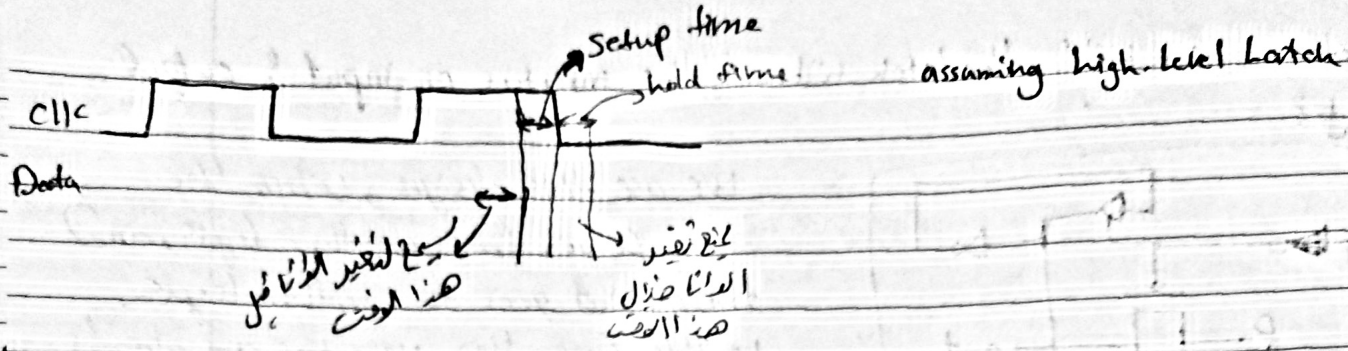
+ How to calculate the delay?
+ clk to out delay
+ Data to out delay.



* Setup time:
the data be available before the clk closed.

phase 1: I take the variation of clock (h-L) or (L-h) at ⎍⎍

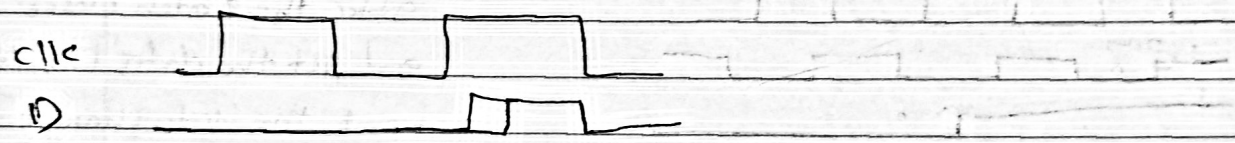
phase 2: I take the variation of clock (L-h) or (h-L) at ⎍⎍



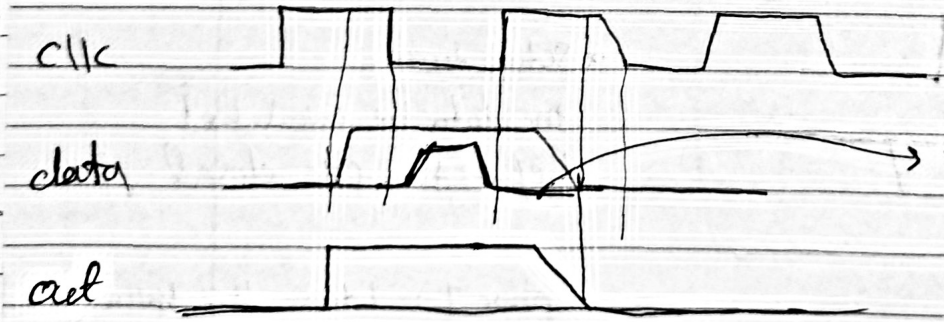
(*) Non-Inverted Latch → The data in the same data out

(*) INverted Latch → The data in is not the same data out
(There is an Werter on input but there is no inverter on output).

(*) Data to out delay → when the data change after clk



(*) clk to out delay → when the data change before the clk



L17:

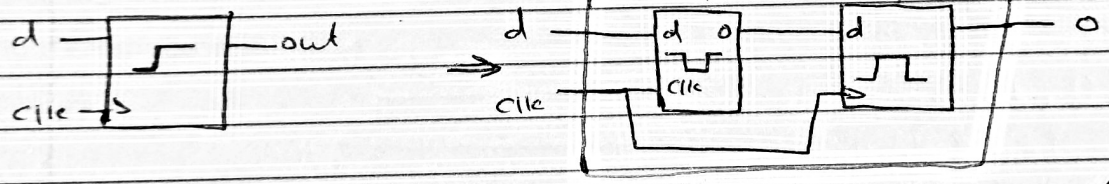
@ Setup time : time that I need data to be ready before the clk is closed in latch.

@ hold_time : time that I need the data stable in a specific period of time.

@ P_Latch
Works on clk Low ⌐‾|_

— Flip Flop :-
two-Latches (master & slave) , one high and another low.

* positive edge Flip Flop



* If last latch on high then the F.F in rising edge.

* The disadvantage of Flip_Flop.
That is used two latches → more area & power

But we use the F.F because we know when the data will get (on edge). but in latch there is a period to get the data.

in F.F there is no data to out delay