GOVT. POLYTECHNIC BALASORE

[LECTURE NOTE]

VLSI & EMBEDDED SYSTEM
TH 2

DIPLOMA

5TH SEMESTER, E & TC ENGINEERING

PREPARED BY:-ER. YOGASAKTI YOGAMAYA
(Lect. In E&TC)

DEPT. OF ELECTRONICS &TELECOMMUNICATION ENGINEERING

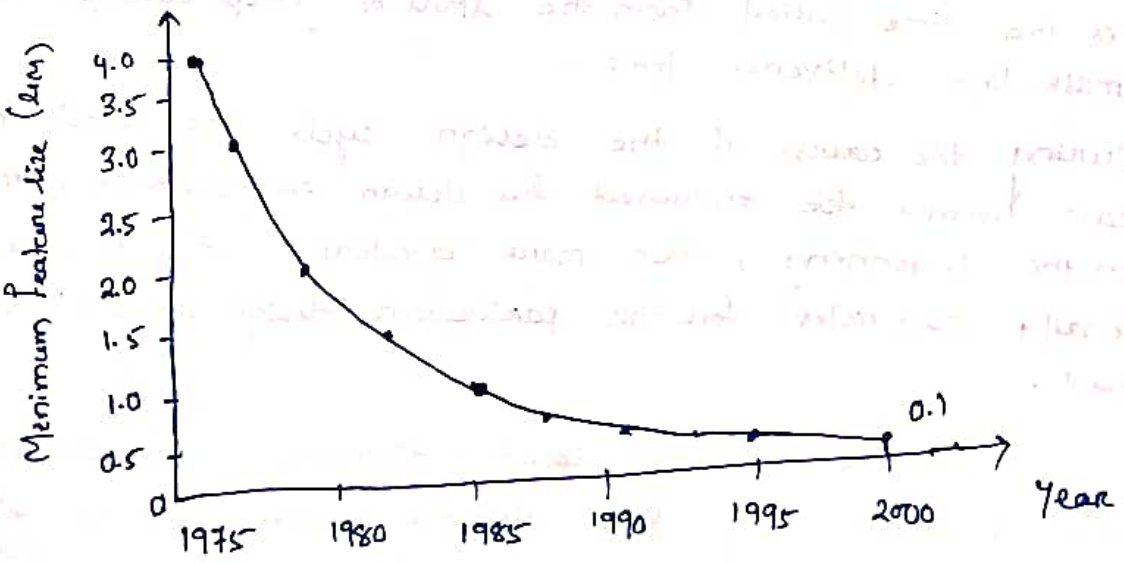
Historical Perspective :

→ As more & more complex functions are required in various data processing & telecommunication devices, the need to integrate these functions in a small package is also increasing.

→ The level of integration as measured by the number of logic gates in a monolithic chip has been raising due to rapid progress in processing technology & interconnect technology.

Era	Year/Date	Complexity (# of logic blocks per chip)
Single Transistor	1958	<1
unit logic (one gate)	1960	1
Multifunction	1962	2-4
Complex function	1964	5-20
Medium Scale Integration (MSI)	1967	20-200
Large Scale Integration (LSI)	1972	200-2000
very large Scale Integration (VLSI)	1978	2000-20,000
ULtra Large Scale Integration (ULSI)	1989	20000 - above

* The monolithic integration of a large number of functions on a single chip usually provides :
  - → less area/volume & therefore compactness
  - → Less power consumption
  - → Less testing requirements at system level
  - → Higher reliability, mainly due to improve on-chip interconnects
  - → Higher speed, due to significantly reduced interconnection length
  - → Significant cost savings



→ Advances in device manufacturing technology allow the steady reduction of minimum feature size (such as minimum channel length of a transistor or an interconnect width realizable on chip).

→ Memory circuits are highly regular, & thus more cells can be integrated with much less area for interconnects.

→ Digital CMOS ICs have been the driving force behind VLSI for high performance computing & other scientific & engineering applications.

→ The demand for digital CMOS ICs will be continuously strong due to salient features such as low power, reliable performance, circuit techniques for high speech using dynamic circuits & ongoing improvements in planning technology.

→ Now the minimum feature size in CMOS IC can decrease to 0.035 μm (35 nm). With such technology, the level of integration in a single chip can be on the order of several tens of billions of transistors for logic chips or even higher in case of memory chip.

VLSI Design Methodologies :

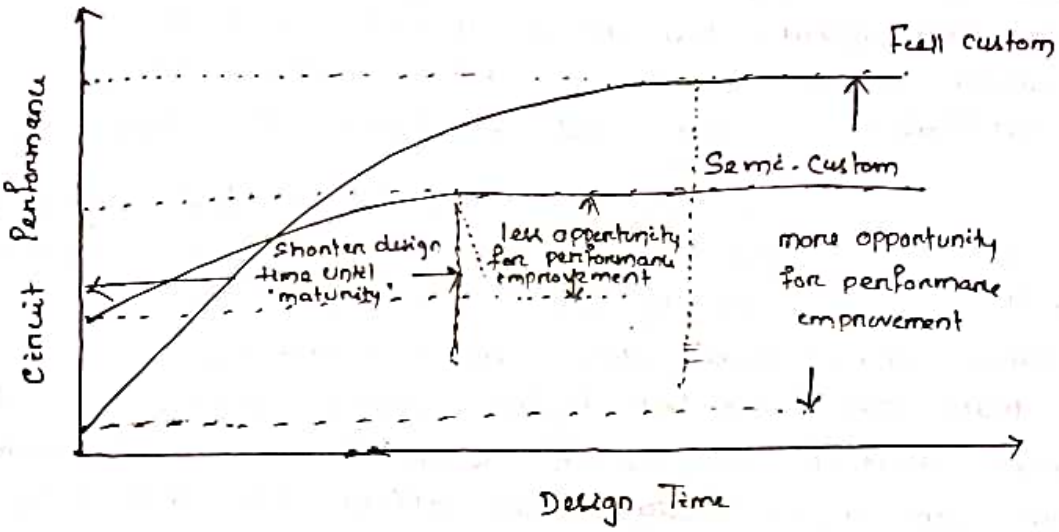
→ The logic chips such as microprocessor chips & Digital signal processing (DSP) chips contains large arrays memories & different functional units. Thus their design complexity is considered much higher than that of memory chips.

→ The design complexity of logic chips increase almost exponentially with the number of transistors to be integrated.

→ This is translated into an increase in the design cycle time, which is the time period from the start of chip development until the mask-tape delivered time.

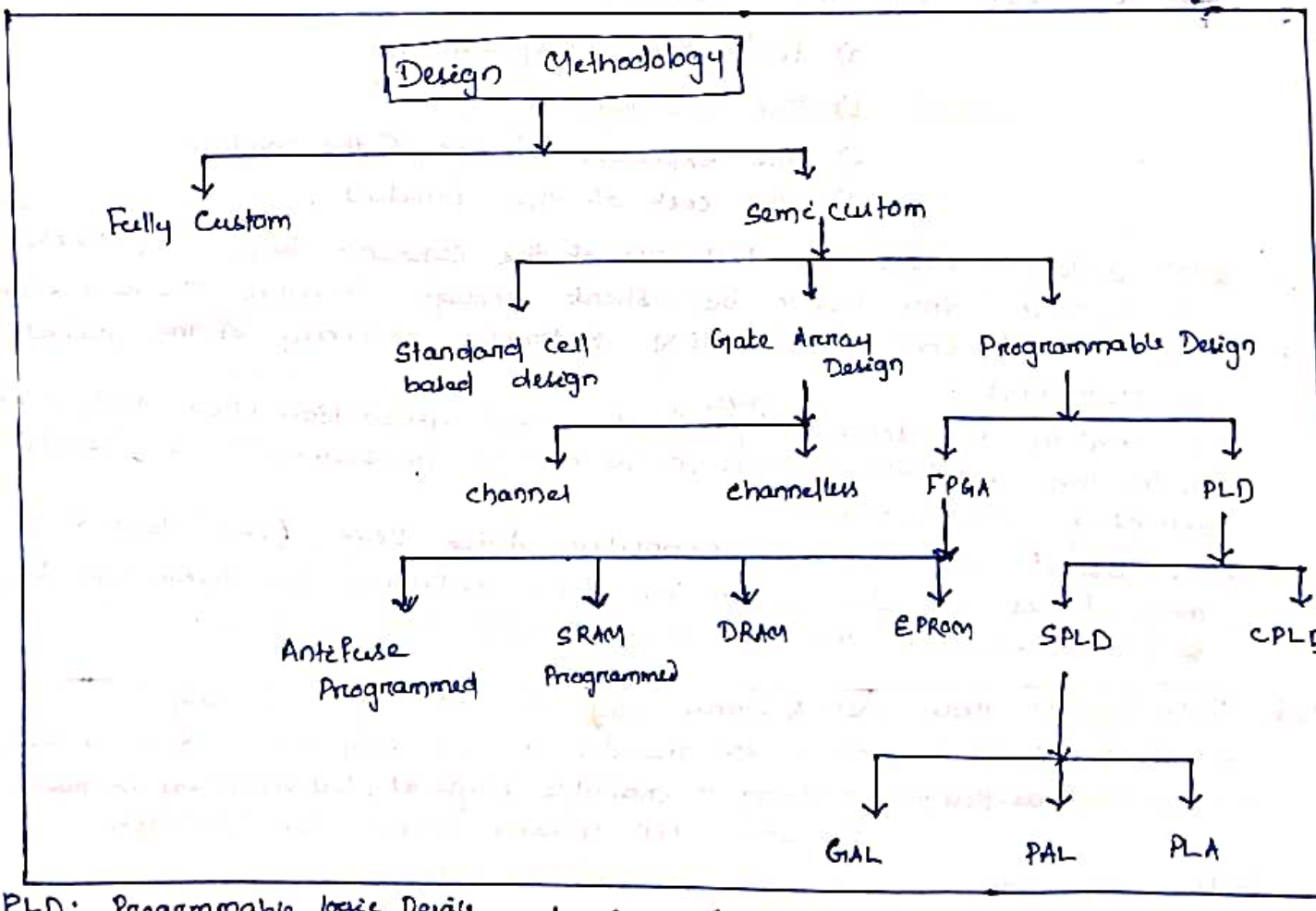
→ During the course of the design cycle the circuit performance can usually be increased by design improvements; more rapidly in the beginning, then more gradually until the performance finally saturates for the particular design style & technology being used.

* The level of circuit performance which can be reached within a certain design time strongly depends on the efficiency of the design methodologies as well as design style.



→ Here two different VLSI design styles are compared for their relative - relative merits in the design of the same product.

* Using the full-custom design style (where the geometry & the placement of every transistor can be optimized individually) requires a longer time until design maturity can be reached, yet the inherent flexibility of adjusting almost every aspect of circuit design allows for more opportunity for circuit performance improvement during the design style.



PLD: Programmable logic Device
SPLD: Simple programmable logic Device
CPLD: Complex " " "

GAL: Gate Array logic
PAL: Programmable array logic

→ The final product typically has a high level of performance (e.g high processing speed, low power dissipation) & the silicon area is relatively small because of better area utilization. But this comes at a larger cost in terms of design time.

* Using a semi custom design style (such as standard-cell based design or FPGA) will allow a shorter design time until design maturity can be achieved.

→ In the early design phase, the circuit performance can be even higher than that of a full-custom design, since some of the components used in semicustom design are already optimized.

→ But the semicustom design style offers less opportunity for performance improvement, & the overall performance of the final product will inevitably be less than that of a full-custom design.

* The choice of the particular design style for a VLSI product depends on the performance requirements the technology being used

* The choice of the particular design style for a VLSI product depends on

a) Performance Requirements
b) The technology being used
c) The expected lifetime of the product
d) The cost of the product.

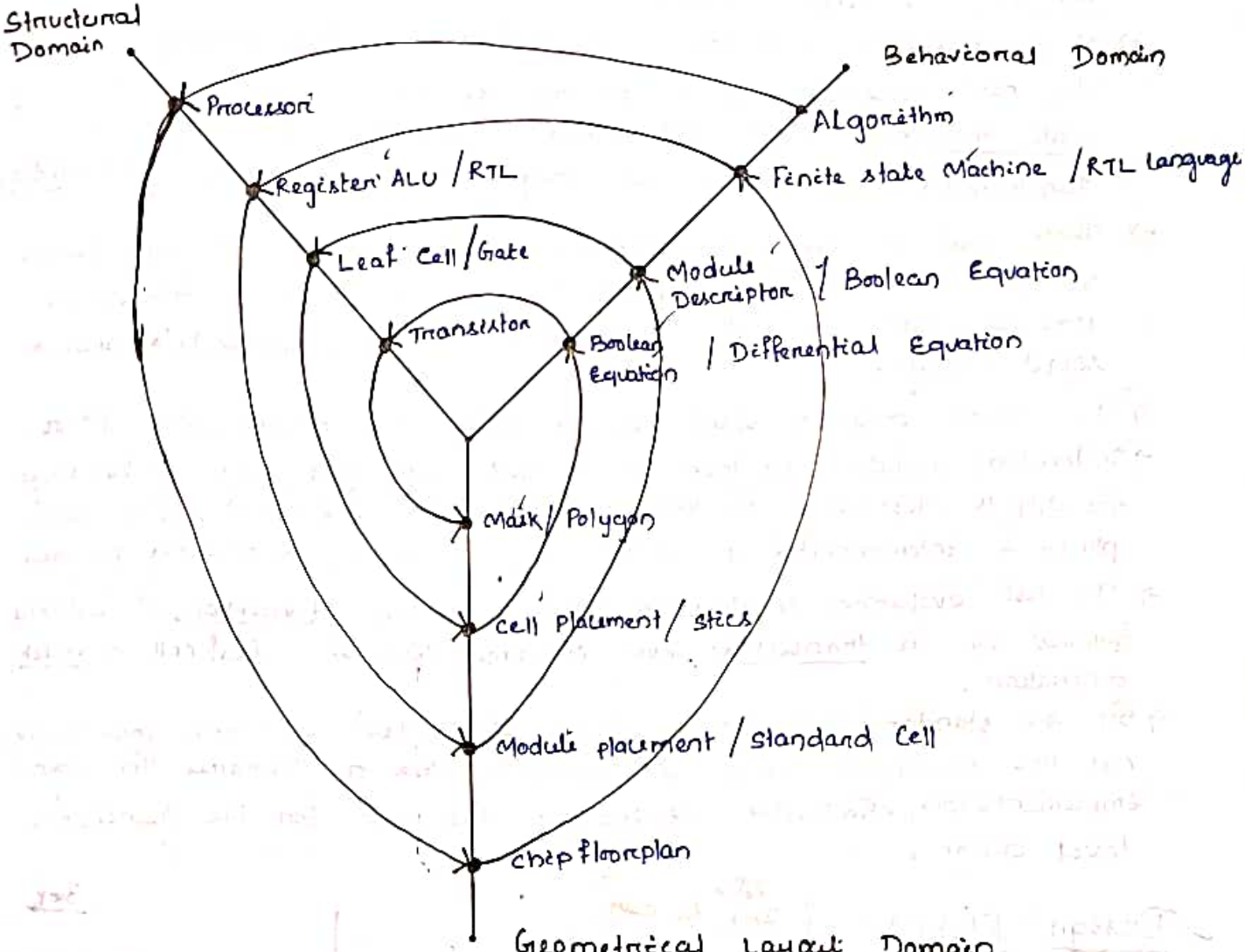
* In order to make the best use of the current technology, the chip development time has to be short enough to allow the manufacture maturing of chip manufacturing & timely delivery of the product to customers.

* In reality, the design cycle of the next generation chip usually overlaps with the production cycle of the current generation chips, thereby assuring continuity.

→ The use of sophisticated computer Aided Design (CAD) tools & methodologies are also essential for reducing the design cycle time & for managing the increasing design complexity.

Note: Fully Custom: Here each & every part of ckt will be design.
Semicustom: All components are available & we'll only connect to get a desire
Standard Cell base Design: Nothing is available physical, but everything in paper.
Ex: Book. Cell library produced by fabrication.
Gate array Design: Something is available physically.
Programmable Design: Everything available in physically & we only write the program for operation.

VLSI Design Flow:



[Fig: Simplified VLSI design flow in three domains (Y-chart Representation)]

→ The design process, at various levels, is usually evolutionary in nature.
→ It starts with a given set of requirements. Initial design is developed & tested against the requirements.
→ When requirements are not met, the design has to be improved. If such improvement is either not possible or too costly, then a revision of requirements & an impact analysis must be considered.

* The Y chart introduced by D. Gajski illustrates a simplified design flow for most logic chips, using design activities on three different axes (domains) which resembles the letter Y.

→ The Y-chart consists of three domains of ~~representation~~ representation namely

i) Behavioral domain or Functional domain
ii) Structural domain
iii) Geometrical layout domain

→ The design flow starts from the algorithm that describes the behavior of the target chip. The corresponding architecture of the processor is first defined.

→ It is mapped onto the chip surface by floorplanning.

→ The next design evolution in the behavioral domain defines finite state machines (FSM) which are structurally implemented with functionally modules such as registers & arithmetic logic Unit (ALUs).

→ These modules are then geometrically placed onto the chip surface using CAD tools for automatic module placement followed by routing, with a goal of minimizing the interconnects' area & signal delays.

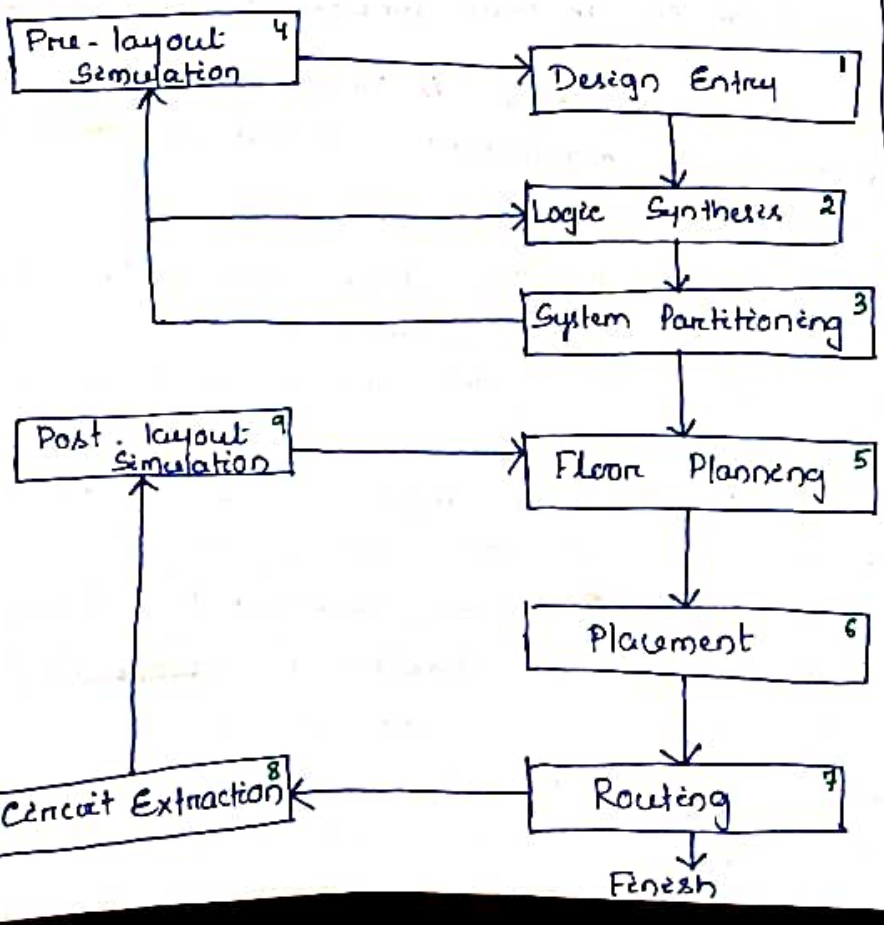
→ The third evolution starts with a behavioral module description.

→ Individual modules are then implemented with leaf cells. At this stage the chip is described in terms of logic gates (leaf cells), which can be placed & interconnected by using a cell placement & routing program.

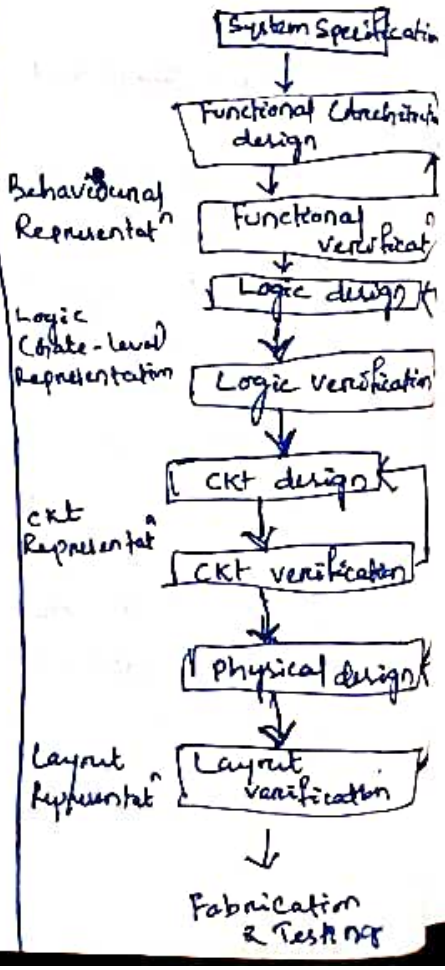
→ The last evolution involves a detailed Boolean description of leaf cells followed by a transistor level implementation of leaf cell & mask generation.

→ In the standard-cell based design style, leaf cells are pre-designed (at the transistor level) & stored in ~~the~~ a library for logic implementation, effectively eliminating the need for the transistor level design.

Design Flow: (Refer Book for easy)



Book



1. Design Entry : Enter the design into an Asic (Application specific Ic) design system either using a HDL or schematic table.

2. Logic Synthesis : Use an HDL & logic synthesis tool to produce a net list i.e description of logic cell & their connection.

3. System Partitioning : Divide a large system into Asic size pieces.

4. Prelayout Simulation : check to see the design function correctly.

5. Floor Planning : Arrange the blocks of the net list on the chip. or, it describes the interconnection of the blocks (RAM, ROM, ALU etc) logic cells (NAND, NOR, FF etc) within the blocks & the logic cell connections.

6. Placement : Decide the location of cell on the block.

7. Routing : Make connection between cell s & blocks.

8. Circuit Extraction: Determine the resistance & capacitance of the interconnets.

9. Post layout Simulation: check to see the design still works with the added loads of the interconnects.

* Step 1-4 are parts of logic design & step 5-9 are parts of physical design.

Design Hierarchy :

→ The use of the hierarchy or "divide & conquer" technique involves dividing a module into sub-modules & then repeating this operation on the submodules until the complexity of the smaller parts becomes manageable.

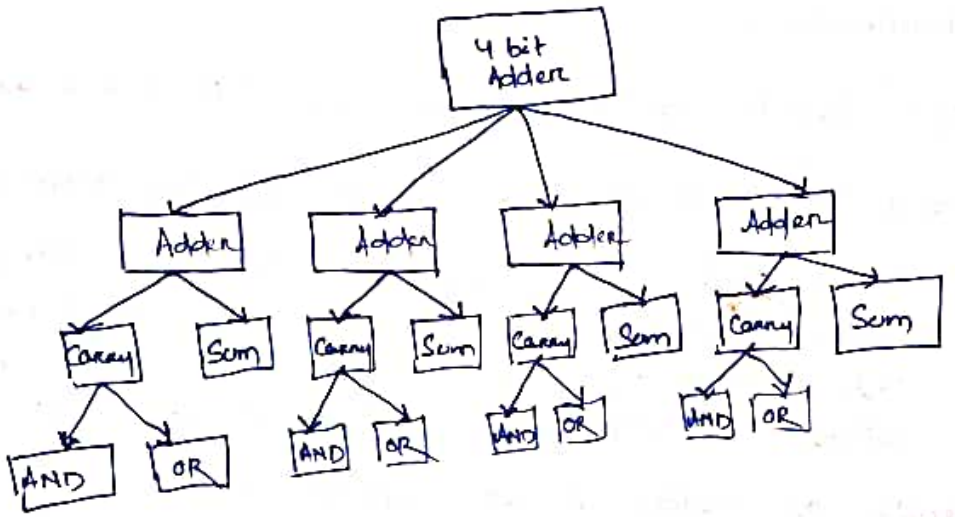
→ This approach is very similar to software development where large programs are split into smaller & smaller section until simple subroutines, with well-defined functions & interfaces.

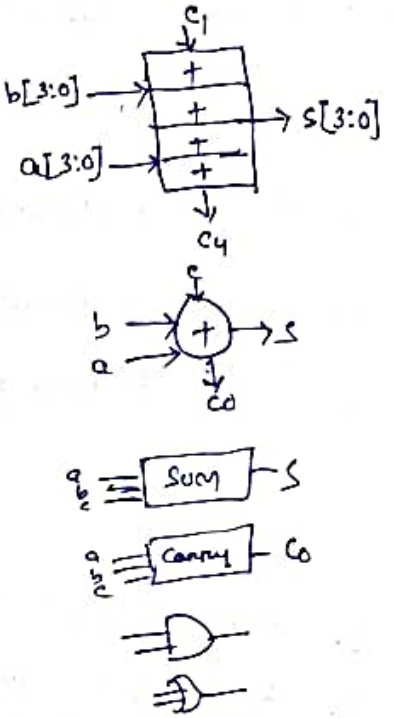
→ The design of a VLSI chip can be represented in three domains. Correspondingly, a hierarchy structure can be described in each domain separately.

→ Example of structural hierarchy which shows the structural decomposition of a CMOS 4 bit adder into its components.

→ The adder can be decomposed progressively into 1 bit adders, separate carry & sum circuits & finally into individual logic gates.

→ At this lower level of hierarchy, the design of a simple circuit realizing a well defined Boolean function is much easier to handle than at the higher levels of the hierarchy.





(Structural decomposition of a 4 bit adder, showing the levels of hierarchy)

* On physical domain, partitioning a complex system into its various functional blocks will provide a valuable guide for the actual realization of these block on the chip. Obviously, the approximate shape & size (area) of each sub-module should be estimated in order to provide a useful floorplan.

Concepts of Regularity, Modularity & Locality:

* Regularity means that the hierarchical decomposition of a large system should result in not only simple, but also similar blocks as much as possible.

→ Ex: An example of Regularity is the design of array structures consisting of identical cells - such as a parallel multiplication array.

→ Regularity can exist at all levels of abstraction. For example, at the transistor level uniformly sized transistors simplify the design & at the logic level, identical gate structures can be used.

* Modularity in design means that the various functional blocks which make up the larger system must have well-defined functions & interfaces.

→ Modularity allows that each block or module can be designed relatively independently from each other, since there is no ambiguity about

the function & the signal interface of these blocks.

→ All of the blocks can be obtained combined with ease at the end of the design process, to form the large system.

→ The concept of modularity enables the parallelization of the design process.

→ By defining well-characterized interface for each module in the system, we effectively ensure that the internals of each module become unimportant to the exterior modules. Internal details remain at the local level.

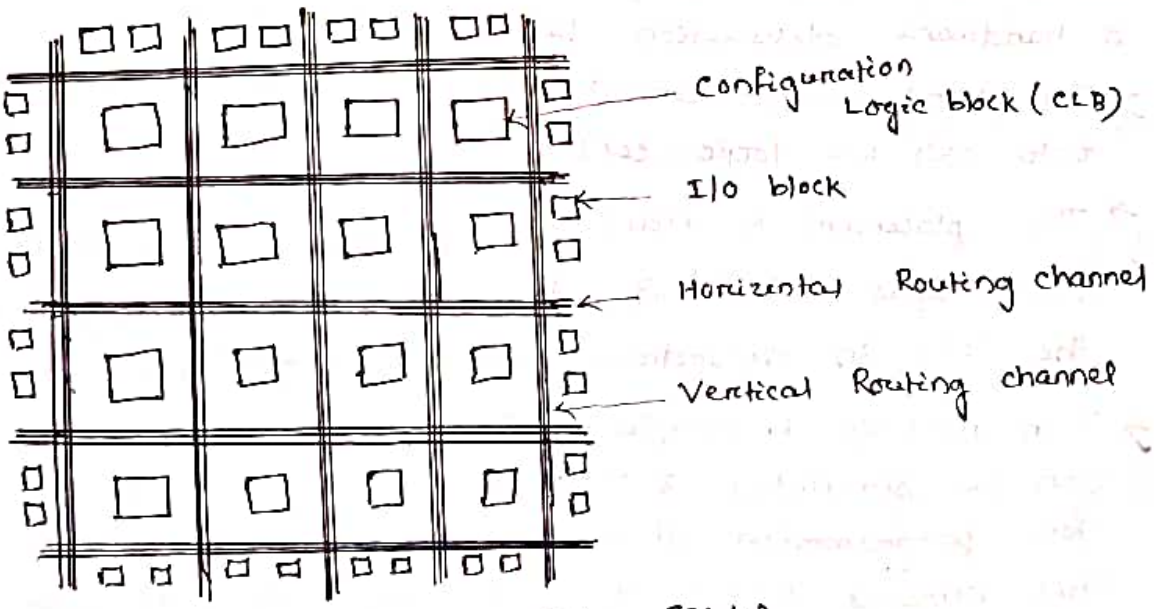
* The concept of locality also ensures that connections are mostly between neighboring modules, avoiding long distance connections as much as possible.

→ This last point is extremely important for avoiding long interconnect delays.

Saikam

VLSI Design Styles:

* Field Programmable Gate Array (FPGA):



(General architecture of Xilinx FPGAs)

→ Fully fabricated FPGA chips containing tens to hundreds of thousands, or even more, of logic gates with programmable interconnects, are available to users for their custom hardware programming to realize desired functionality.

→ This design style provides a means of for fast prototyping & also for cost-effective chip design, especially for low volume applications.

→ A typical field programming gate array (FPGA) chip consists of IIO buffers, an array of configurable logic blocks (CLBs), & programmable interconnect structures.

→ The programming of the interconnects is accomplished by programming of RAM cells whose output terminals are connected to the gates of MOS pass transistors.

→ Thus, the signal routing between the CLBs & the IIO block is accomplished by setting the configurable switch matrices accordingly

* The complexity of a FPGA chip is typically determined by the number of CLBs it contains.

→ State of the art FPGA chip can support system clock frequency upto hundreds of MHz.

→ With use of dedicated CAD tools, the gate utilization rate can exceeds 90%.

→ The typical design flow of an FPGA chip starts with the behavioral description of its functionality, using a hardware description language as VHDL.

→ The synchronized architecture is then technology mapped (or partitioned) into ckts or logic cells.

→ The placement & routing step assigns individual logic cells to FPGA sites (CLBs) & determines the routing patterns among the cells in accordance with the net list.

→ After routing is completed; the on chip performance of the design can be simulated & verified before down loading the design for programming of the FPGA chip. The programming of the chip remains valid as long as the chip is powered on, or until it is re programmed.

* The largest advantage of FPGA based design is the very short turn around time, i.e the time required from the start of the design process until a functional chip is available.

Gate Array Design:

→ Gate array implementation requires a two-step manufacturing process.

→ The first phase, which is based on generic (standard) masks, results in an array of uncommitted transistors on each GA chip.

→ A corner of a gate array chip contains bonding pads on its left & bottom edges, diodes for I/O protections nmos & pmos transistors for chip output driver circuits adjacent to bonding pads, arrays of nmos, pmos transistors, underpass wire segments, & power & ground buses along with contact windows.

→ Typical GA platforms allow dedicated areas, called channels. for interconnect routing between rows & columns of MOS transistors.

→ The availability of these routing channels simplifies the interconnections.

→ Interconnection patterns that perform basic logic gates can be stored in a library, which can then be used to customize rows of uncommitted transistor according to the netlist.

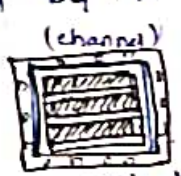
→ Some of the platform offer dedicated memory (RAM) arrays to allow higher density.

* In modern GAs, multiple metal layers are used for channel routing. With the use of multiple interconnect layers, the routing can also be achieved over the active cell areas, thus the routing channels can be removed as in Sea-of-Gates (SOG) chips.

→ Here the entire chip surface is covered with uncommitted nmos & pmos transistors.

→ For intercell routing, some of the uncommitted transistors must be sacrificed.

* The GA chip utilization factor, as measured by the used chip area divided by the total chip area.



Standard-Cells Based Design:

→ The standard cells based design is the most prevelent full-custom design styles which requires development of a full custom mask set.

→ ~~It is also~~ The standard cell is also called the polycell.

→ In this design style, all of the commonly used logic cells are developed, characterized & stored in a standard cell library.

* Standard-Cells Based Design:

→ The standard cell is also called polycell.

→ In this design style, all of the commonly used logic cells are developed, characterized, & stored in a standard cell Library.

→ A typical Library may contain a few hundred cells including inverters, NAND Gates, NOR gates, complex AOI, OAI gates, D latches, & flipflops.

→ Each gate type can be implemented in several versions to provide adequate driving capability for different fan-outs.

→ Each cell is characterized according to several different characteri-zation categories, including

* Delay time Versus load capacitance
* Circuit simulation model
* Timing simulation model
* Fault simulation model
* Cell data for place & route
* Mask data.

→ To enable automated placement of cells & routing of inter-cell connections, each cell layout is designed with a fixed height, so that a number of cells can be abutted side-by-side to form rows.

→ The power & ground rails typically run parallel to the upper & lower boundaries of the cell, thus neighbouring cells share a common power bus & common ground bus.

→ The input & output pins are located on the upper & lower boundaries of the cell.

→ The nmos transistors are located closer to the ground rail while the pmos transistors are placed closer to power rail.

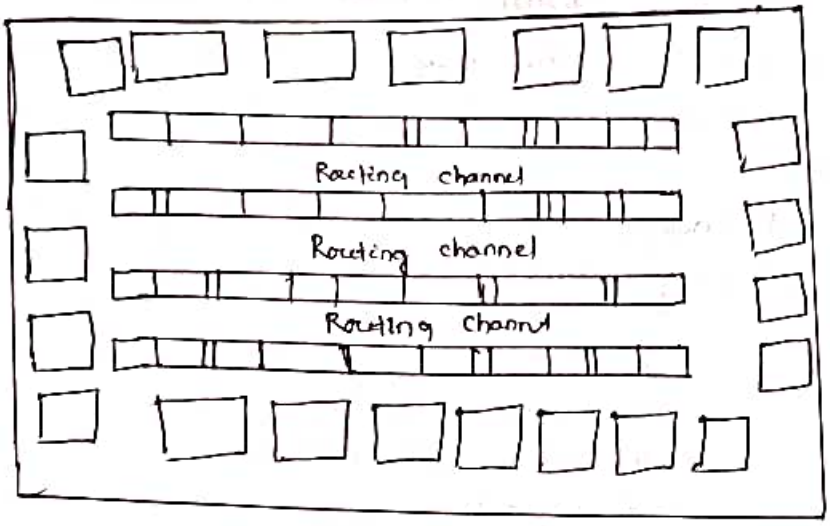
→ The chip area contains rows & columns of standard cells. Between cell rows are channels for dedicated interval routing.

→ The signal delay, noise margins, & power consumption of each cell should be also optimized with proper sizing of transistors using ckt simulation.

→ If a number of cells must share the same input and/or output signals, a common signal bus structure can also be incorporated into the standard-cells based chip layout.

→ Within the cell block, the separation betn neighboring rows depend on the number of wires in the routing channel betwee the cell rows.

→ If a high interconnect density can be achieved in the routing channel, the standard cell rows can be placed closer to each other, resulting in a smaller chip area.



(A simplified floorplan of standard cells based design)

Full Custom Design:

→ In a truly fully-custom design, the entire mask design is done anew without use of any library.

→ The development cost of such a design style is becoming prohibitively his. Thus the concept of design reuse is becoming popular in order to reduce design cycle time & development cost.

→ The most rigorous full custom design can be the design of memory cell, be it static or dynamic.

→ For logic chip design, a good compromise can be achieved by using a combination of different design styles on the same chip, such as standard cells, data-path cells, & programmable logic arrays (PLAs).

→ In real full custom layout in which the geometry, orientation, & placement of every transistor is done individually by the designer, designer productivity is usually very low - typically a few tens of transistors per day, per designer.

→ In digital CMOS VLSI, full-custom design is rarely used due to the high labor cost.

→ Ex of hybrid full custom design is Intel Pentium µp chip.

* Computer Aided Design Technology:

→ Computer Aided design (CAD) tools are essential for timely development of integrated ckt.

→ Although CAD tools cannot replace the creative & inventive parts of the design activities, the majority of time-consuming & computation intensive mechanistic parts of the design can be executed by using CAD tools.

→ The CAD technology for VLSI chip design can be catagories into the following areas:

* High level Synthesis
* Logic Synthesis
* ckt optimization
* Layout
* Simulation
* Design rules checking
* Formal Verification

Synthesis Tools:

→ The high level synthesis tools using hardware description languages (HDLs). such as VHDL or Verilog, address the automation of the design phase in the top level of the design hierarchy.

→ With an accurate estimation of lower level design features, such as chip area & signal delay, it can very effectively determine the types & quantities of modules to be included in the chip design.

→ Many tools are customized for particular design needs, especially for area minimization, low power, high speed, or their weighted combination.

Layout Tools:

→ The tools for ckt optimization are concern with transistor sizing for minimization of delays & with process variations, noise & reliability hazards.

→ ~~The layout CAD tools are goal driven & include some degree of optimization functions.~~

→ The layout cAD tools include floor planning, place & routine, & module generation.

→ To achieve an optimum or near optimum placement of all standard cells on chip, the signal routing bet^n the cells can be accomplished with minimum interconnect area & minimum delay.

→ Once the physical locations of all the cells in a design are determined, an automatic routing tool is used to create the metal interconnections between the cell terminals, based on the gate netlist.

Simulation & Verification Tools:

The simulation category, which is the most mature area of VLSI CAD, includes many tools ranging from ckt level simulation, timing level simulation, logic level simulation & behavioral simulation.

→ The aim of all simulation CAD tools is to determine if the designed ckt meets the required specifications, at all stages of the design process.

→ Logic simulation is performed mainly to verify the functionality of the ckt, i.e to determine if the designed ckt actually has the desired logic behavior.

→ A number of test vectors (inputs) are applied to the ckt during logic simulation, & the outputs are compared with expected output patterns.

* ckt level or electrical simulation tools are routinely used to determine nominal & worst case gate delays to identify delay-critical signal paths or elements, & to predict the influence of parasitic effects upon ckt behavior.

* The design rules checking CAD category included the tools for layout rules checking, electrical rules checking, & reliability rules checking.

Saikat

Fabrication of MOSFET

Introduction:

→ A well-established CMOS fabrication technology requires both n-channel (nmos) & p-channel (pmos) transistors be built on the same chip substrate.

→ To accomodate both nmos & pmos devices, special region must be created in which the semiconductor type is opposite to the substrate type. These regions are called Wells or tubs.

→ A pwell is created in an n type substrate or vice-versa.

→ In twin-tub CMOS technology, additional tubes of the same type as the substrate can also be created for device optimization.

Fabrication Process Flow:

Basic steps:

~~NMOS Fabrication~~:

Notes:

→ The integrated ckt may be viewed as a set of patterned layers of doped silicon, polysilicon, metal & insulating silicon dioxide.

→ In general, a layer must patterned before the next layer of material is applied on the chip.

* The process used to transfer a pattern to a layer on the chip is called "Lithography".

* The sequence starts with the thermal oxidation of the silicon surface by which an oxide layer of about 1mm thickness (Fig b)

→ The entire oxide surface is then covered with a layer of photoresist, which is essentially a light sensitive, acid-resistant organic polymer, initially insoluble in the developing solution. (Fig c)

→ If the photoresist material is exposed to ultraviolet (UV) light, the exposed areas becomes soluble so that they are no longer resistant to etching solvents. (Fig d)

→ To selectively expose the photoresist, we have to cover some of the areas on the surface with a mask during exposure.

→ When the structure with the mask on the top is exposed to UV light, areas which are covered by the opaque features on the mask are shielded.

→ On the areas where the UV light can pass through, the photoresist is exposed & becomes soluble. (fig d)

* The type of photo resist which is initially insoluble & becomes soluble after exposure to UV light is called positive photoresists

→ The photo-resistive which is initially soluble & becomes insoluble after exposure to UV light, called negative photoresist.

→ Negative photoresists are more sensitive to light, but their photolithographic resolution is not as high as that of positive photoresists. Therefore negative photoresists are used less commonly in the manufacturing of high density integrated ckt.

* Following the UV exposure step, the unexposed portion of the photoresist can be removed by a solvent.

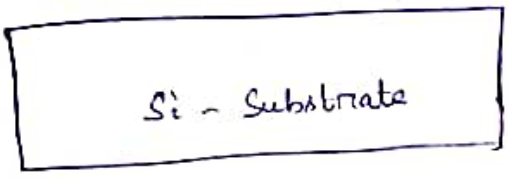
→ The silicon dioxide regions which are not covered by hardened photoresist can be etched away either by using a chemical solvent (HF acid) or by using a dry etch (plasma etch) process. (fig e)

→ The remaining photoresist can now be stripped from the silicon dioxide surface by using another solvent, leaving the patterned silicon dioxide features on the surface (fig g)

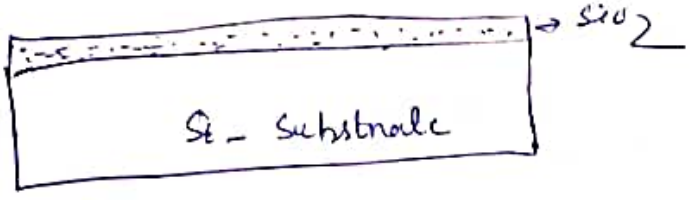
→ For accurate generation of high density patterns requires in sub-micron devices, electron beam (E-beam) Lithography is used instead of optical lithography.

Fig:

(a)

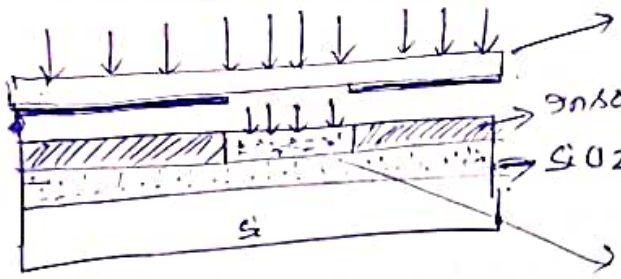


(b)

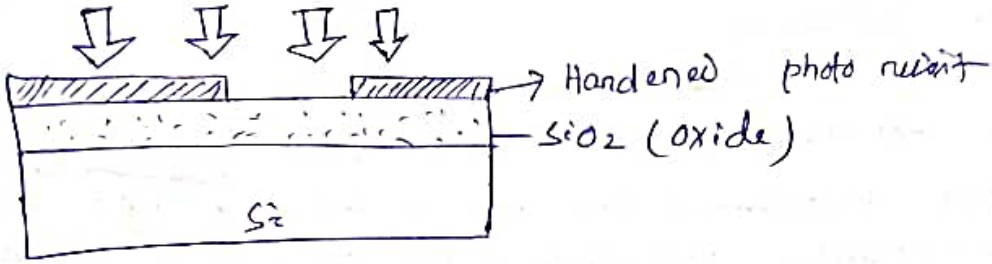


(c)

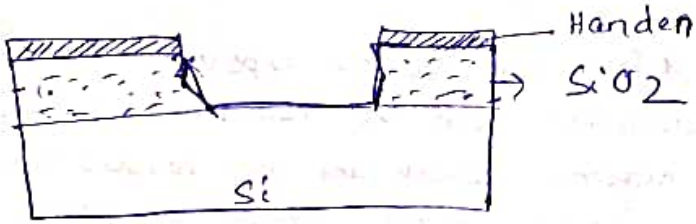
(d)



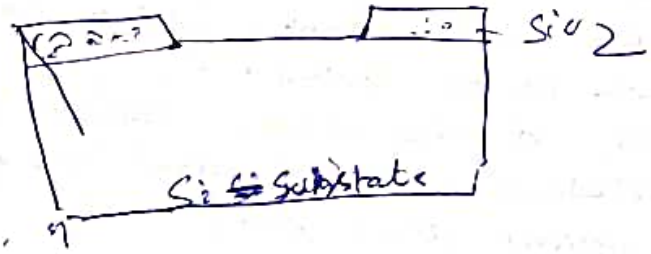
(e) chemical etch (HF acid) or dry etch (Plasma)



(f)



(g)



Pucknell

nMOS Fabrication:

1. Processing is carried out on a thin wafer cut from a single crystal of silicon of high purity into which the required P-impurities are introduced as the crystal is grown.

2. A layer of silicon dioxide (SiO_2) typically 1 μm thick is grown all over the surface of the wafer to protect the surface, act as a barrier to dopants during processing, & provide a generally insulating substrate onto which other layer may be deposited & patterned.

3. The surface is now covered with a photoresist which is deposited onto the wafer & spun to achieve an even distribution of the required thickness.

4. The photoresist layer is then exposed to ultraviolet light through a mask which defines those regions into which diffusion is to take place together with transistor channels.

5. These areas are subsequently readily etched away together with the underlying silicon dioxide so that the wafer surface is exposed in the window defined by the mask.

6. The remaining photoresist is removed & a thin layer of SiO_2 is grown over the entire chip surface & then the polysilicon is deposited on top of this to form the gate structure. The polysilicon layer consists of heavily doped polysilicon deposited by chemical vapour deposition (CVD). In the fabrication of the fine pattern device, precise control of thickness, impurity concentration & resistivity is necessary.

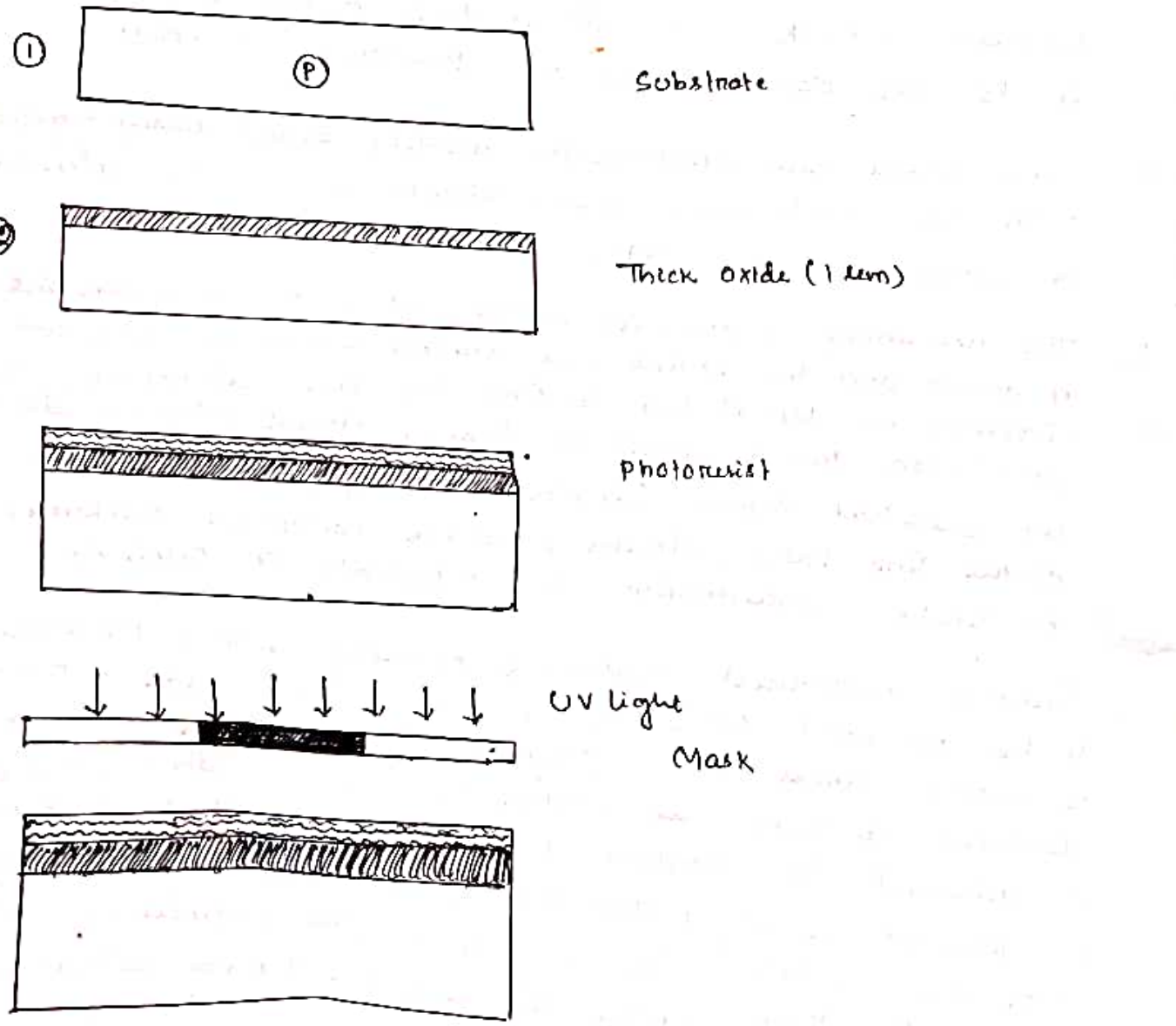
7. Further photoresist coating & masking allows the polysilicon to be patterned (as step 6), & then thin oxide is removed to expose areas into which n type impurities are to be diffused to form the source & drain as shown. Diffusion is achieved by heating the wafer to a high temperature & passing a gas containing the desired n type impurity over the surface. The polysilicon with underlying thin oxide & the thick oxide act as masks during diffusion — the process is self-aligning.

8. Thick Oxide (SiO_2) is grown over all again & is then masked with photoresist & etched to exposed selected areas of the Polysilicon gate & the drain & source areas where connections are to be made.

9. The whole chip then has metal (aluminium) deposited over its surface to a thickness typically of 1 μm. This metal layer is then masked & etched to form the required interconnection pattern.

The process revolves around the formation of deposition & patterning of three layers, separated by silicon dioxide insulation. The layers are diffusion within the substrate, polysilicon on oxide on the substrate, & metal insulated again by oxide.

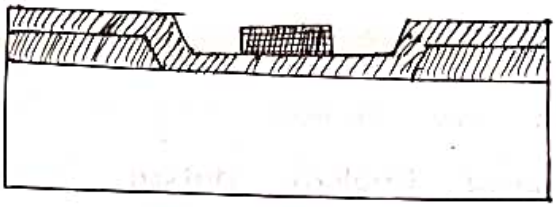
Figure:



⑤

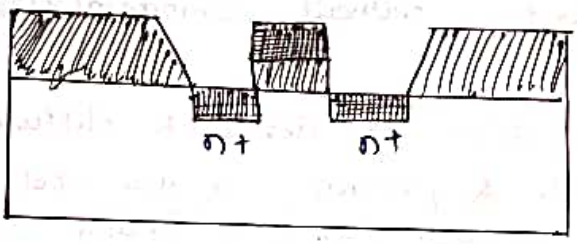
Window in oxide

⑥



Patterned poly (1-2 μm) on thin oxide (800 - 1000 Å)

⑦

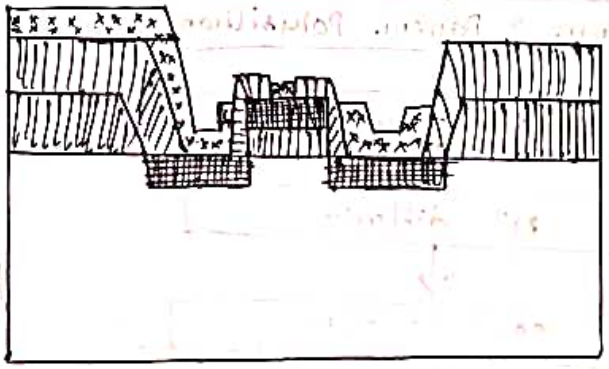


n^+ diffusion (1 μm deep)

⑧

Contact holes (cuts)

⑨



patterned metallization (aluminum) (1 μm)

CMOS Fabrication:

The n-Well Process:

→ N-well cmos ckts are superior to p-well because of the lower substrate bias effects on transistors threshold voltage & inherently lower parasitic capacitances associated with source & drain region.

● Typical n-well fabrication steps are:

→ The 1st mask defines the n-well regions. This is followed by a low dose phosphorus implant driven in by a high temperature diffusion step to form the n-wells.

→ The well depth is optimized to ensure against p substrate to p^+ diffusion breakdown without compromising the n well to n^+ mask separation.

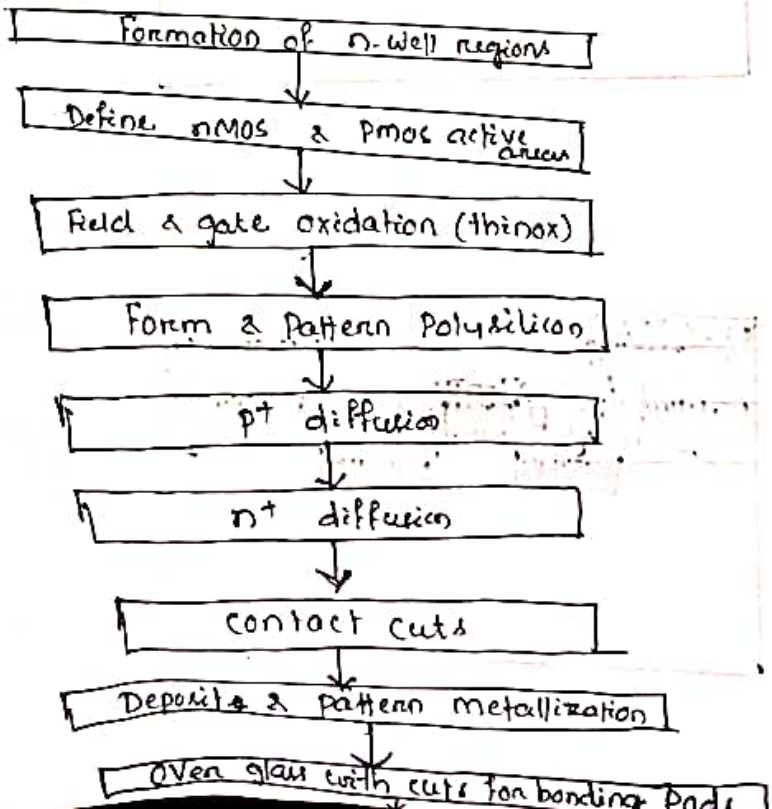
→ The next steps are define the devices & diffusion paths, grow field oxide, deposite & pattern by the polysilicon, carryout the diffusions, make contact cuts, & finally metallize as below.

* An n^+ mask & its complement may be used to define the n^- & p- diffusion regions respectively. These same masks also include the VDD & Vss contacts respectively.

→ n well process creates non-optimum p channel characteristics.

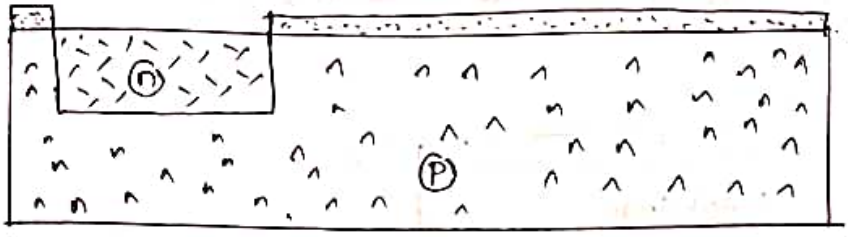
Steps:

***

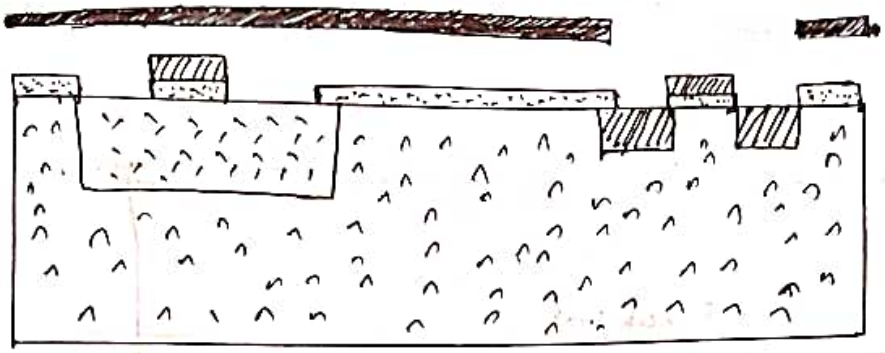


By Rukaiya

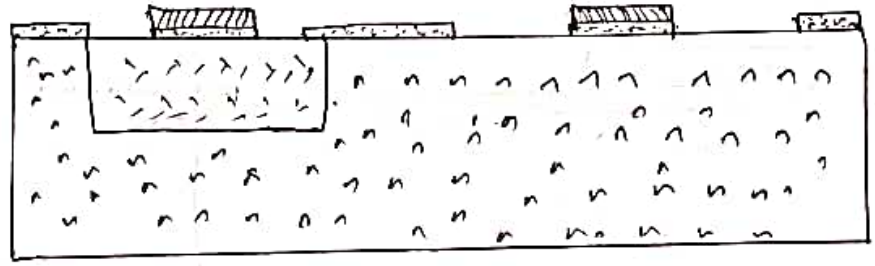
①



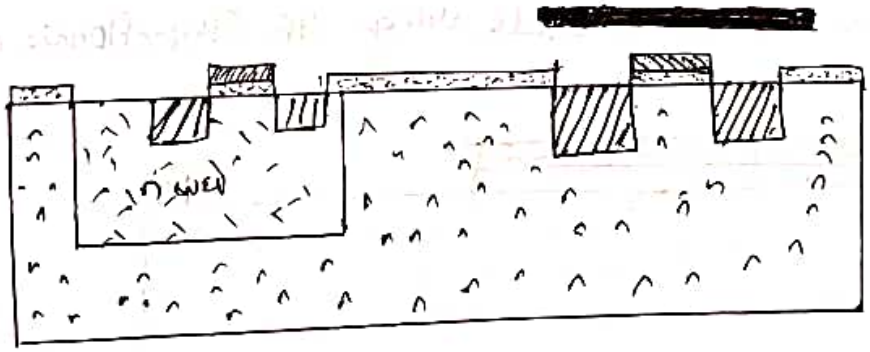
③

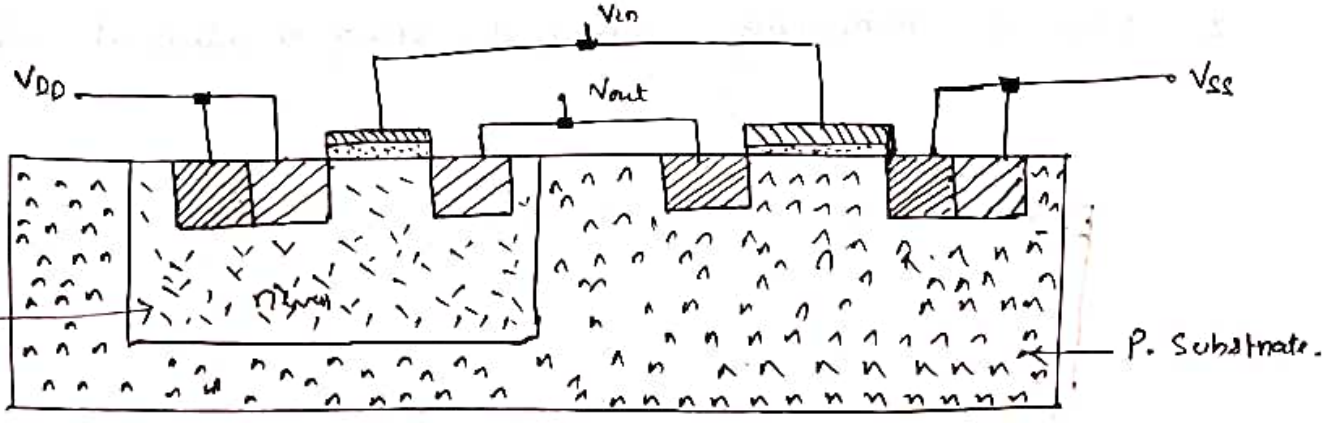


②



④



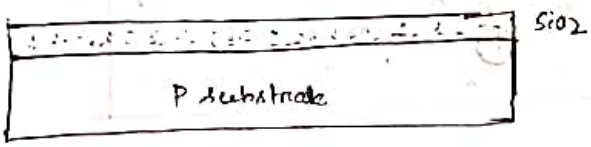


(Cross Sectional View of n-well CMOS inverter)

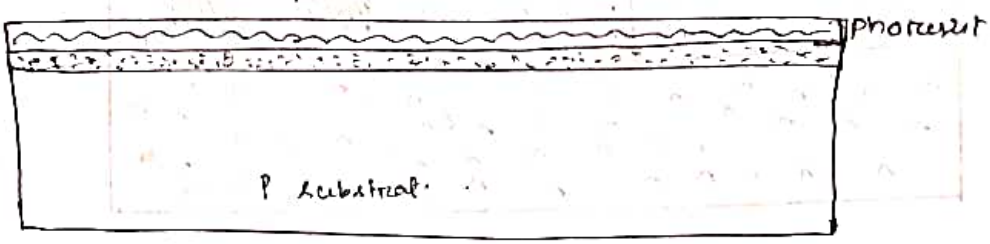
Internet

Steps :

1. Blank wafer covered with a layer of SiO_2 using Oxidation.

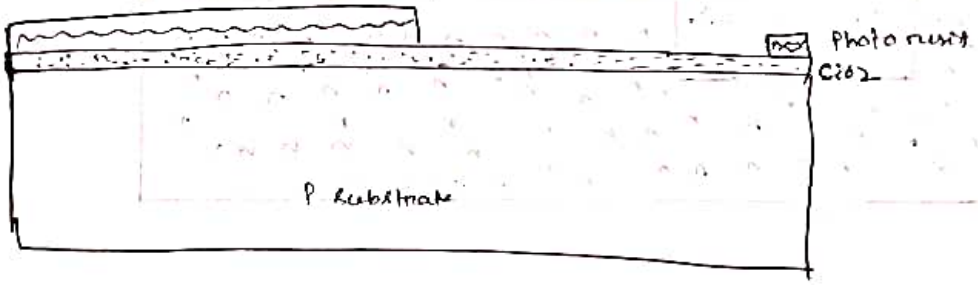


2. Spin on the photoresist. Exposed to UV light using the n-well mask. (Photolithography)

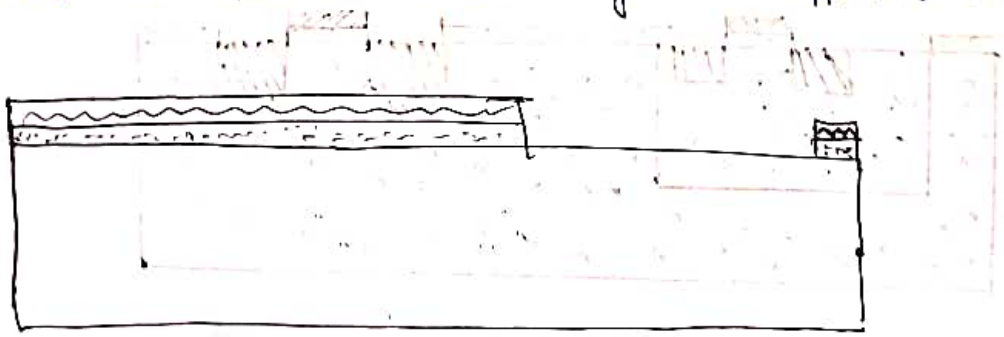


~~N well Process~~

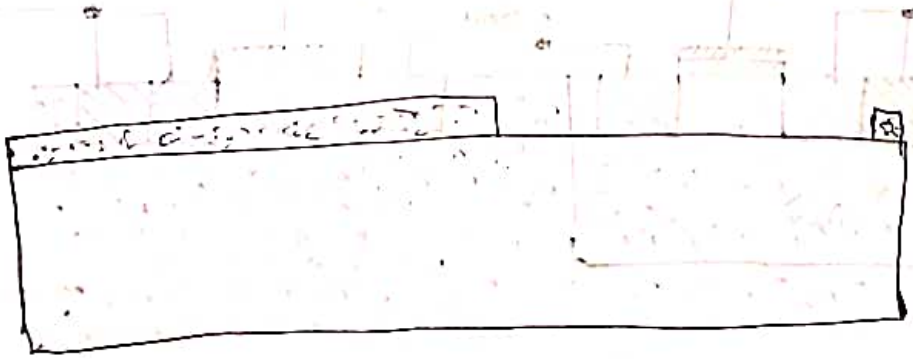
3. Strip off the exposed photoresist using organic Solvents



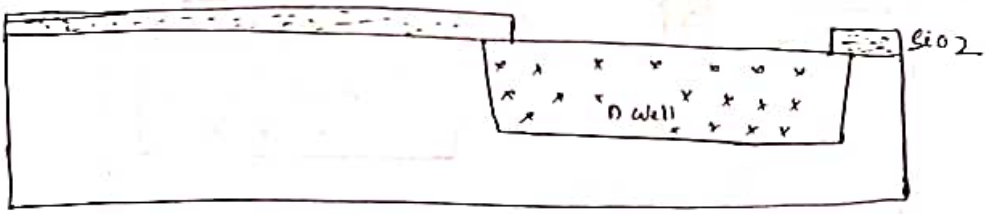
4. Etched the uncovered oxide using HF (Hydrofluoric acid)



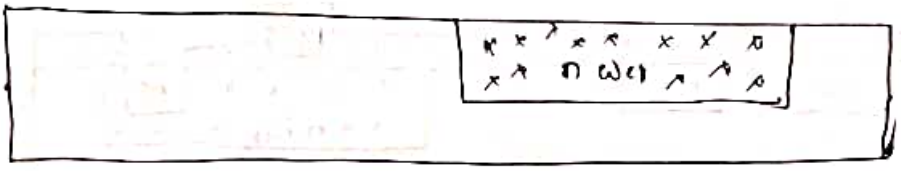
5. Etch the remaining photoresist using a mixer of acids



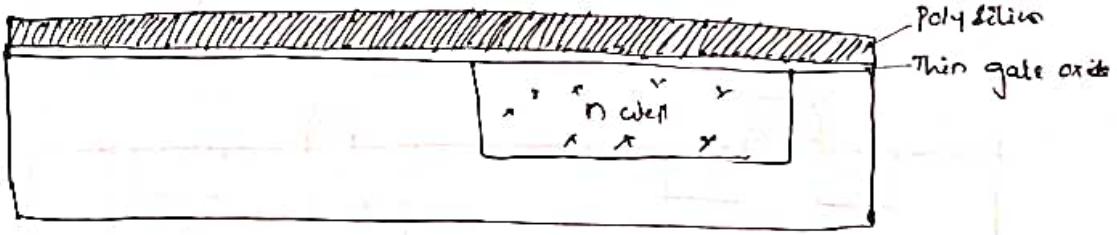
6. n-well is formed using diffusion or ion implantation.



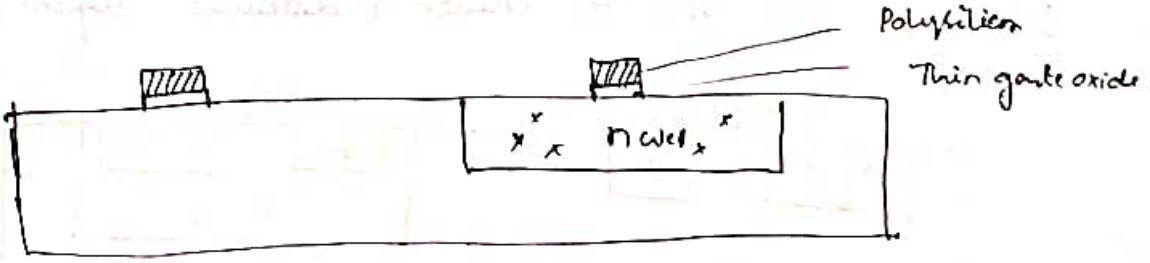
7. Strip off remaining oxide using HF. Subsequent step use the same photo lithography process.



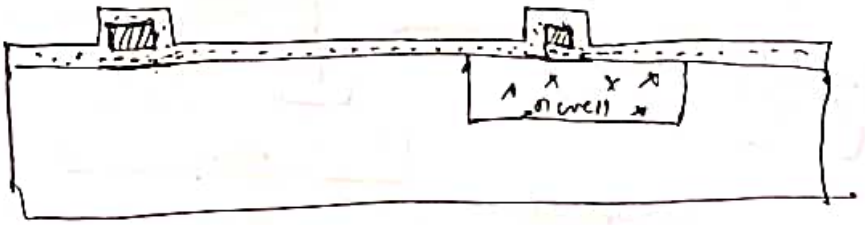
8. Deposite thin layer of oxide. Use CVD to form poly & dope heavily to increase conductivity.



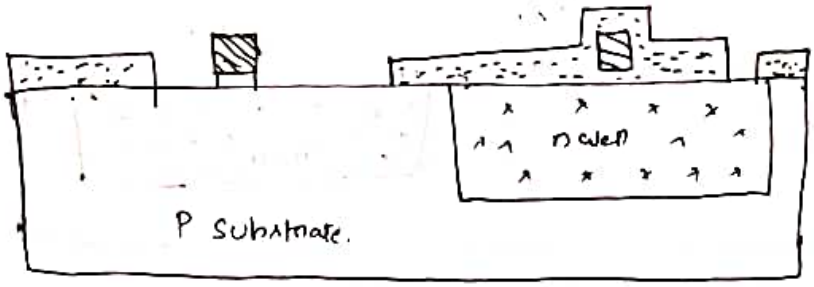
9. Pattern poly using the previously discussed photolithography process.



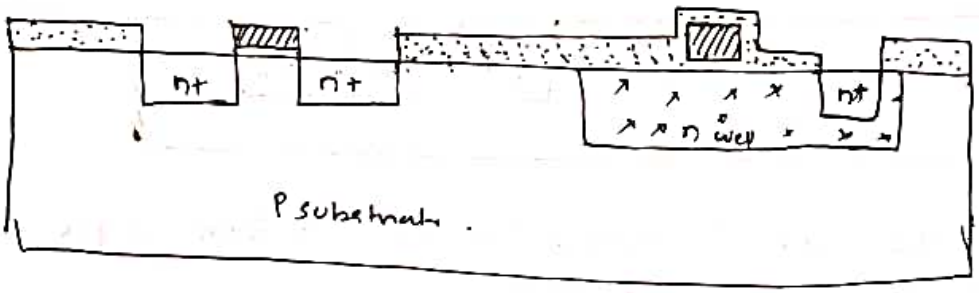
10. Cover with oxide to define n diffusion regions.



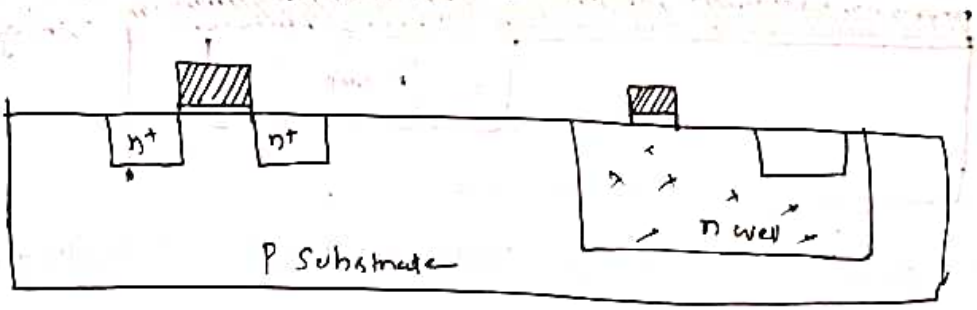
11. Pattern oxide using n+ active mask to define n diffusion region.



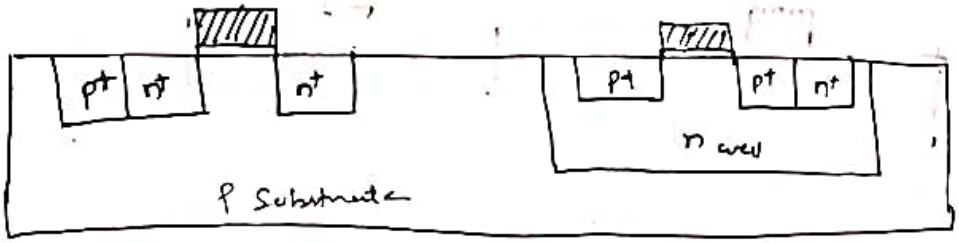
12. Diffusion or ion implantation used to create n diffusion regions.



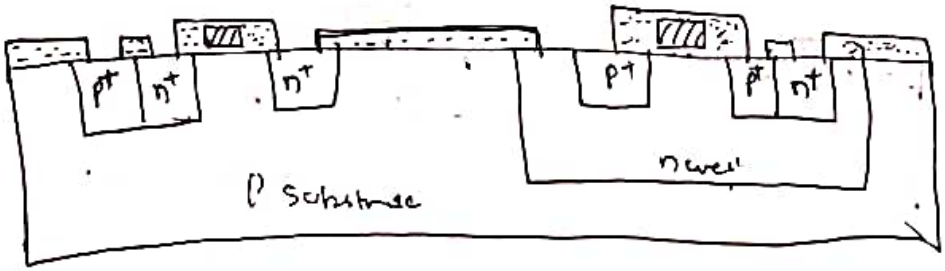
13. Strip off the oxide to complete patterning step.



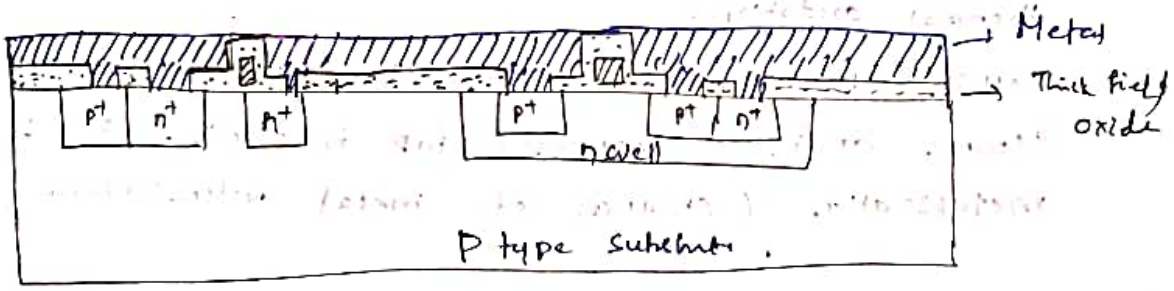
14. Similar steps used to create P diffusion regions.



15. Cover chip with thick field oxide & etch oxide where contact cuts are needed.



15. Remove Excess metal leaving wires



CMOS n-well Fabrication Process flow:

- A well established CMOS fabrication technology requires both n-channel (n-MOS) & p-channel (p-MOS) transistors be built on the same chip substrate.

→ To accomodate both n MOS & P MOS devices, special region must be created in which the semiconductor type is opposite to the substrate type. These regions are called wells or tubs.

→ A p-well is created in an n type substrate or vice-versa.

→ A twin-tub CMOS technology, additional tubes of the same type as the substrate can also be created for device optimization.

* The simplified process sequence for the fabrication starts with the creation of the n-well region for pMOS transistor, by impurity implantation into the substrate.

→ Then a thick oxide is grown in the region surrounding the nMOS & pMOS active region.

→ The thin oxide is subsequently grown on the surface through thermal oxidation.

→ These steps are followed by the creation of n+ & p+ regions (Source, drain & channel-stop implants) & by final metallization - (creation of metal interconnects).

Steps:

***

Saikam

MOS Transistor

Chapter: 4

General Concept:

The Fermi potential ϕ_f, which is the function of temperature & doping, denotes the difference between the intrinsic fermilevel E_i & fermilevel E_F.

$$\phi_f = \frac{E_F - E_i}{q}$$

For p type semiconductor, the fermi potential can be approximate by:

$$\phi_{fp} = \frac{KT}{q} \ln \frac{n_i}{N_A}$$

For n type semiconductor, Fermi potential can be given by

$$\phi_{fn} = \frac{KT}{q} \ln \frac{N_D}{n_i}$$

Where K → Boltzmann constant
q → The Unit (electron) Charge
n_i → Intrinsic carrier concentration
T → Temperature.
N_A → Acceptor (Boron) Concentration
N_D → Donor concentration

* The electron affinity of silicon, which is the potential difference (free space) between the conduction band level & vaccume level, is denoted by $q\chi$.

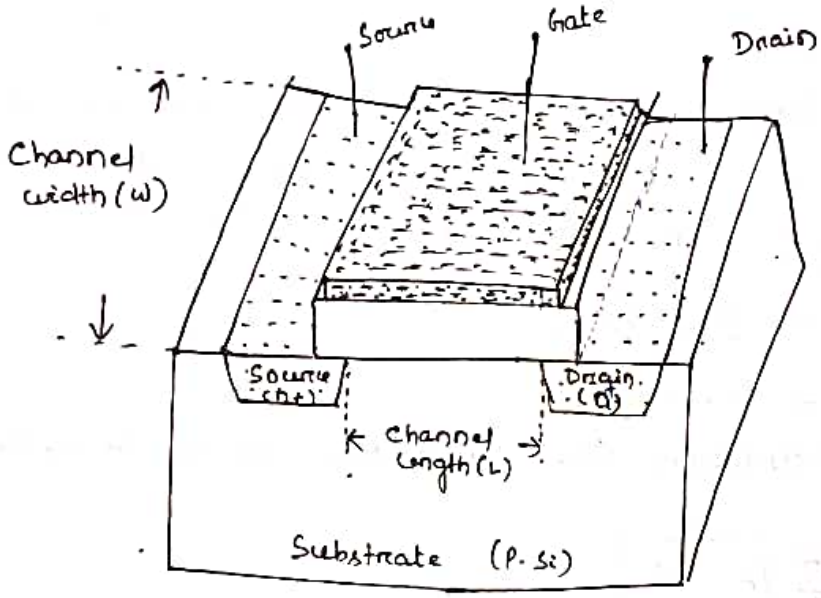
* The energy Required for an electron to move from the Fermi level into free space is called work function $q\phi_s$.

$$q\phi_s = q\chi + E_c - E_f$$

Where E_c = conduction band level
E_F = Fermi level.

* Fermi Level: The fermilevel (E_F) is the maximum energy level which is occupied by an electron at absolute Zero (0K) temperature.
- The fermi level lies at the middle of the forbidden energy gap.

Structure & Operation of Mos Transistor (MOSFET):



(Fig: The Physical structure of an n-channel enhancement type MOSFET)

→ The basic structure of an n-channel MOSFET is shown above figure.

Structure:

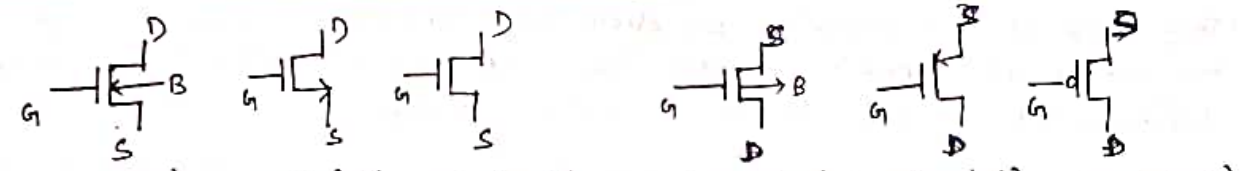
* It is a 4 terminal device consisting of a p type substrate, in which two n^+ diffusion region, the drain & the source are formed. The surface of the substrate region bet^n the drain & source is ~~formed~~ covered with a thin oxide layer, & the metal (or polysilicon) gate is deposited on the top of this gate dielectric. The two n+ regions will be the current-conducting terminals of the device.

* The distance bet^n drain & source diffusion region is the channel length 'L' & the lateral extent of the channel is the channel width 'W'. The thickness of the oxide layer covering the channel region is "tox".

* A mos transistor which has no conducting channel region at zero bias is called "Enhancement type MOSFET." If a conducting channel already exists at zero bias gate bias, the device is called a 'depletion type MOSFET'. In a MOSFET with P-type substrate & with n+ source & drain region, the channel region to be formed on the surface is n type. Thus such ~~type~~ a device with P type Substrate is called 'n-channel MOSFET'. (For this versa p-channel MOSFET) In an n channel MOSFET, source is defined as the n+ region which has a lower potential than the other n+ region, the drain.

V_{GS} → Gate to Source voltage
V_{DS} → Drain to " "
$-V_{BS}$ → Substrate to " "

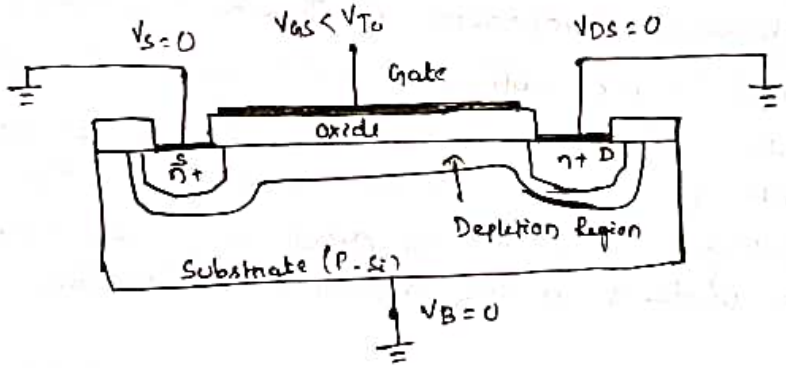
(∵ B → Substrate)



(4 terminal) (simplified) (Simplified) (4terminal) (simplified) (Simplified)

n channel MOSFET (P channel MOSFET)

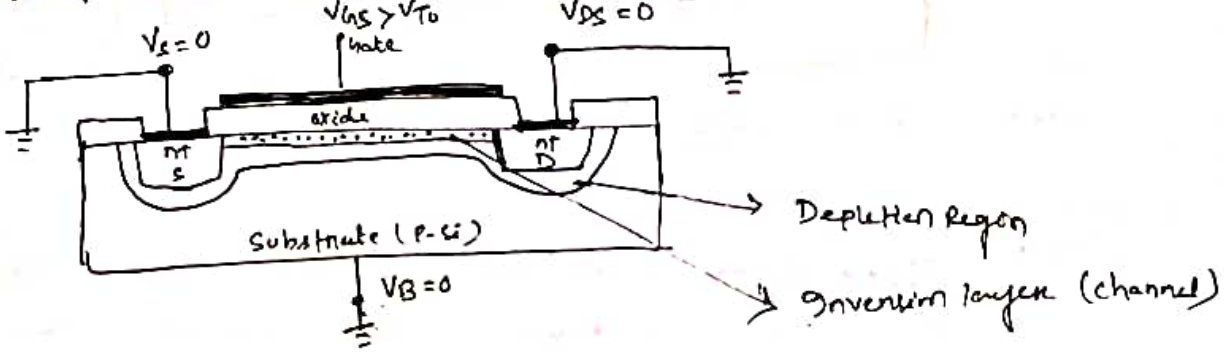
The simple operation principle of this device is : control the current conduction between the source & the drain using the electric field generated by gate voltage as a control variable.



(Formation of a depletion region in an n channel enhancement type MOSFET).

When bias condition is applied across an enhancement type MOSFET, the source, the drain & the substrate terminals are connected to ground. A positive voltage V_{GS} is applied to gate to create the conducting channel underneath the gate. For small gate voltage, the majority carriers (holes) are repelled back into the substrate, & the surface of P type substrate is depleted. Since the surface is devoid of any mobile carriers, current conduction bet^n the source & drain is not possible. (inversion layer is created)

Assume the V_{GS} voltage is further increased. As soon as the surface potential in the channel region reaches $-\phi_{FP}$, surface inversion will established & a conducting n type layer will form bet^n source & drain diffusion region as shown below fig.



(Formation of an inversion layer (channel) in an n-channel enhancement-type MOSFET)

→ This channel now provides an electrical connection bet^n two n+ regions & it allows current flow, as long as there is a potential difference the source & drain terminal voltage.

* The value of V_{GS} voltage needed to cause surface inversion is called "Threshold voltage" V_{TO}.

→ Thus the MOSFET can conduct no current bet^n source & drain terminals unless $V_{GS} > V_{TO}$.

The Threshold voltage:

* The value of Gate to Source Voltage V_{GS} needed to cause surface inversion is called "Threshold voltage" denoted as V_{TO}.

→ The four physical component of Threshold voltage:

i) The work function difference bet^n Gate & channel → ϕ_{GC}
ii) The Gate voltage component to change the surface potential
iii) The Gate voltage component to offset the depletion region charge
iv) The voltage component to offset the fixed charges in the gate oxide & in the silicon oxide interface.

For zero substrate bias, the threshold voltage V_{TO} is

$$V_{TO} = \phi_{GC} - 2\phi_F - \frac{Q_{BO}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}}$$

For non zero substrate bias,

$$V_T = \phi_{GC} - 2\phi_F - \frac{Q_B}{C_{ox}} - \frac{Q_{ox}}{C_{ox}}$$

The General form of Threshold voltage

$$V_T = V_{TO} - \frac{Q_B - Q_{BO}}{C_{ox}}$$

The most general Expression for Threshold voltage, V_T

$$V_T = V_{TO} \mp \gamma \left(\sqrt{|1 - 2\phi_f + V_{SB}|} - \sqrt{|2\phi_f|} \right)$$

where $\gamma = \frac{\sqrt{2q \cdot N_A \cdot \varepsilon_{si}}}{C_{ox}}$ is the substrate bias (body effect) coefficient.

Where ϕ_{GC} → Work function difference bet^n Gate & Channel

Q_{BO} → Depletion Region Charge density at surface inversion

$$= -\sqrt{2q \cdot N_A \cdot \varepsilon_{si} \cdot (-2\phi_F)}$$

Work function: The min^n energy required to extract an electron from a metal surface.

Q_B = The depletion region charge density as a function of V_{SB}.

$$Q_B = -\sqrt{2q \cdot N_A \cdot E_{si} \left| -2\phi_F + V_{SB} \right|}$$

C_{ox} = Gate oxide capacitance = $\frac{\varepsilon_{ox}}{t_{ox}}$

N_A = doping concentration of p type material.

ε_{ox} =

t_{ox} = Gate oxide thickness. , q = Electron charge = 1.602×10^{-19}

ϕ_F = Fermi potential = $\frac{KT}{q} \ln\left(\frac{N_{sub}}{n_i}\right)$ T = Room Temperature

ε_{si} = dielectric constant of silicon = 12 $K = 1.38 \times 10^{-23}$

n_i = Intrinsic carrier concentration = 1.5×10^{10} , N_{sub} = doping concentration of substrate.

MOSFET Operation: A Qualitative View:

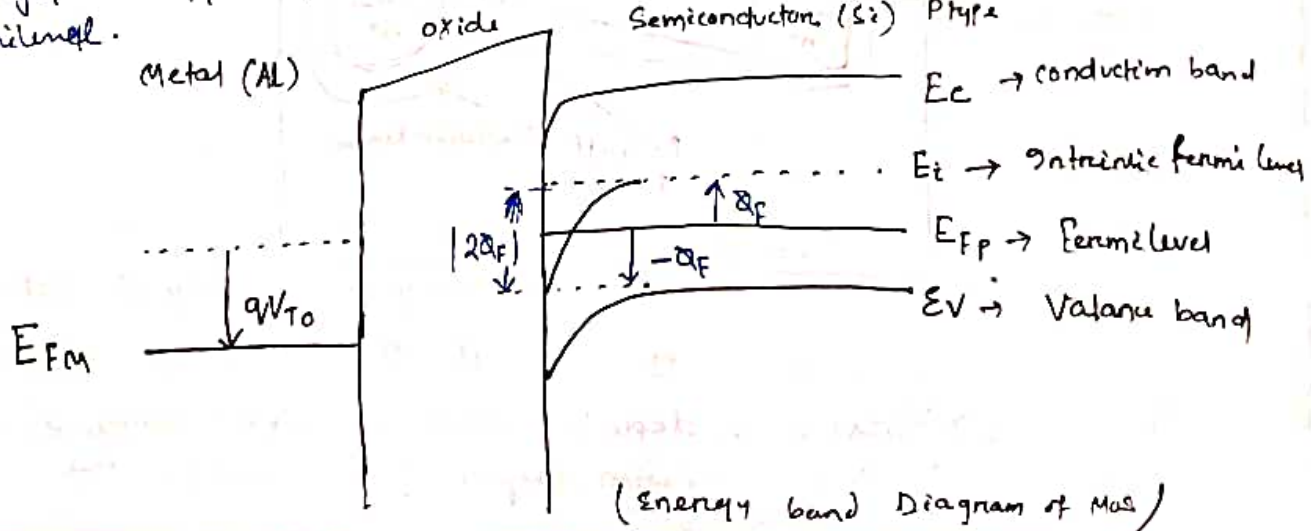
→ The MOSFET consists of a MOS capacitor with two pn junction placed immediately adjacent to the channel region that is controlled by the MOS Gate.

→ The carrier i.e electrons in an nMOS transistor, enter the structure through Source & leave through Drain & are subjected to the control of Gate voltage.

→ To ensure that both p-n junctions are reverse biased initially, the substrate potential is kept lower than the other three terminal potential.

(1) * When $0 < V_{GS} < V_{TO}$, the gated region bet^n the Source & drain is depleted, no carrier flow through the channel.

(2) → As gate voltage increased beyond the threshold voltage ($V_{GS} > V_{TO}$). the mid gap energy level E_i at the surface is pulled below the Fermi level.



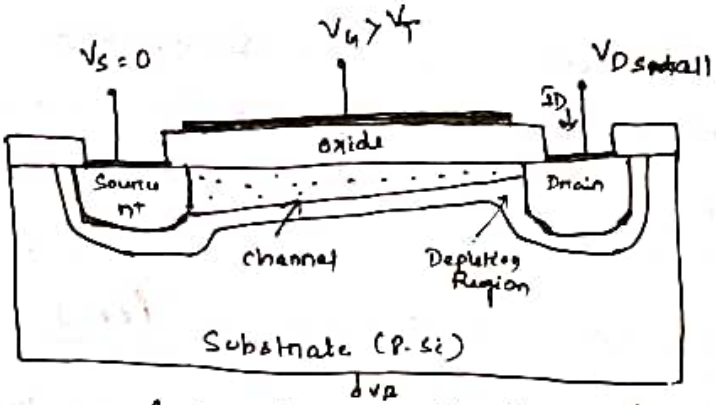
(Energy band Diagram of MOS)

causing the surface potential ϕ_s to turn positive & invert the surface.

→ Once the inversion layer is established on the surface, an n-type conducting channel forms bet^n the S & D, which is capable of carrying drain current.

③ When $V_{GS} > V_{TO}$ & $V_{DS} = 0$, thermal equi. - brium exists in the inverted channel region & I_D is equal to zero.

Fig:



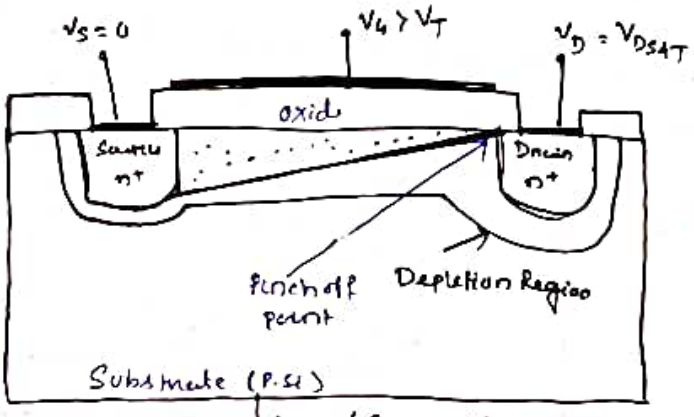
(Operating in the Linear Region)

→ If a small drain voltage $V_{DS} > 0$ is applied, a drain current I_D proportional to V_{DS} will flow from S to D through conducting channel. The inversion layer i.e the channel forms a continuous current path from S to D. This operation mode is called "Linear mode" or "Linear Region." Thus in linear region operation, the channel region acts as a voltage-controlled resistor.

* As the drain voltage is increased, the inversion layer charge charge & the channel depth at the drain ends start to decrease.

* For $V_{DS} = V_{DSAT}$, the inversion ~~layer~~ charge at the drain is reduced to zero, which is called "Pinch off point".

Fig:



(Operating at the Edge of Saturation)

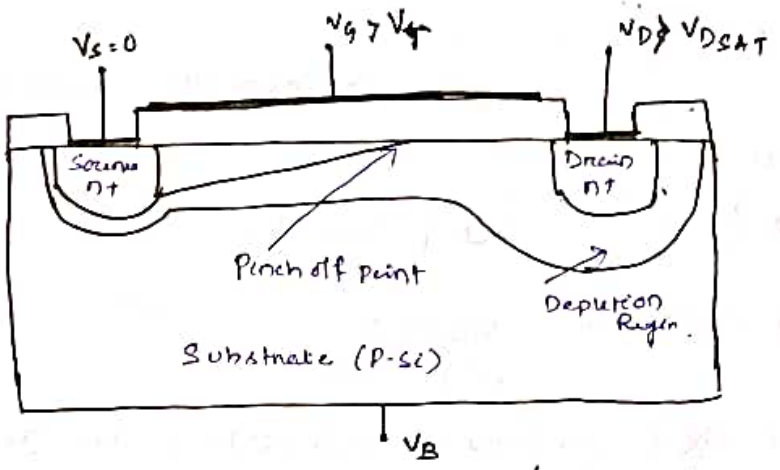
④ Beyond the pinch off point i.e for $V_{DS} > V_{DSAT}$, a depleted surface region forms adjacent to the drain & this depletion region grows towards the S with increasing drain voltage.

→ This operation mode of MOSFET is called "Saturation mode" or "Saturation region".

→ on this region effective channel length is reduced as the inversion layer near the drain vanishes.

* The pinch off (depleted) section of the channel absorbs most of the excess voltage drop ($V_{DS} - V_{DSAT}$) & a high field region forms bet^n the channel end & the drain boundary.

Fig:



(operating beyond saturation)

*

Razavi

MOSFET Current - Voltage characteristics:

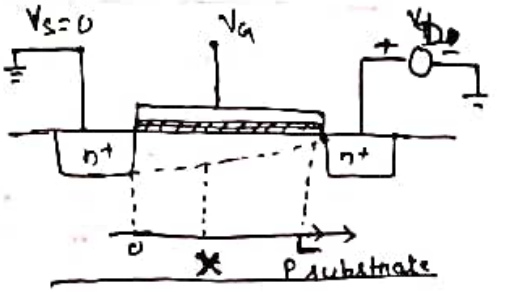
Consider a semiconductor bar carrying a current I. If the charge density along the direction of current is Qd coulombs per meter & the velocity of the charge is v meter per second, then

$$I = Q_d \cdot v$$

We know that

$$Q = CV$$

$$\text{i.e } Q_d = W C_{ox} (V_{GS} - V_{TH})$$



C_{ox} is multiplied by W to represent the total capacitance per unit length. C_{ox} → Gate oxide capacitance per unit area, W → channel width.

→ Let the drain voltage is greater than zero. So the local voltage difference bet^n gate & channel varies from V_G to $V_G - V_D$

Thus the channel density at a point x along the channel can be written as

$$Q_d(x) = W C_{ox} [V_{GS} - V(x) - V_{TH}]$$

Where $V(x)$ is the channel potential at x.

Now current is given by

$$I_D = -W C_{ox} [V_{GS} - V(x) - V_{TO}] v$$

where the -ve sign is intended because the charge carriers are negative & v denotes the velocity of the current in the channel

We know that $\boxed{v = \mu E}$ where μ is the mobility of charge carriers & E is the Electric field.

$$E(x) = -\frac{dv}{dx}$$ (the mobility of electron = μ_n)

By substituting this

$$I_D = W C_{ox} [V_{GS} - V(x) - V_{TO}] \mu_n \frac{dv}{dx}$$

Due to boundary condition $V(0) = 0$

$V(L) = V_{DS}$

Multiply both side dx & performing integration we get

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

Since I_D is constant along the channel

$$I_D L = W \mu_n C_{ox} \int_0^{V_{DS}} [V_{GS} - V(x) - V_{TO}] \, dV(x)$$

$$\Rightarrow I_D = \mu_n C_{ox} \frac{W}{L} \left[(V_{GS} - V_{TO}) V_{DS} - \frac{1}{2} V_{DS}^2 \right] \quad \text{— (1)}$$

$$\Rightarrow \boxed{I_D = \frac{\mu_n C_{ox}}{2} \frac{W}{L} \left[2(V_{GS} - V_{TO}) V_{DS} - V_{DS}^2 \right]} \quad \text{— (2)}$$

This eqn can also be written as

$$\boxed{I_D = \frac{K'}{2} \frac{W}{L} \left[2(V_{GS} - V_{TO}) V_{DS} - V_{DS}^2 \right]}$$

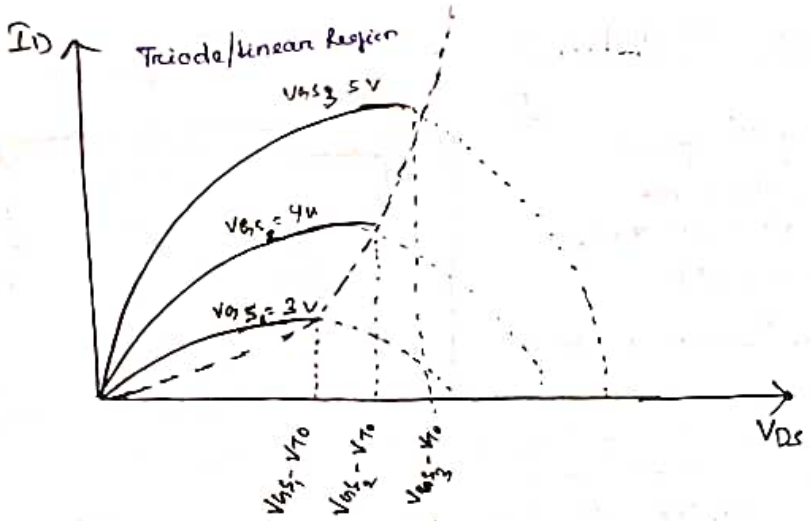
$$\boxed{I_D = \frac{K}{2} \left[2(V_{GS} - V_{TO}) V_{DS} - V_{DS}^2 \right]}$$

Where $K' = \mu_n C_{ox}$ & $K = K' \frac{W}{L}$

$\frac{W}{L}$ → Aspect Ratio

$V_{GS} - V_{TO}$ = Overdrive voltage or effective voltage.

* If $\boxed{V_{DS} \leq V_{GS} - V_{TO}}$, the device operates in "Triode Region" or Linear region.



(Drain current verses drain source voltage in linear Region)

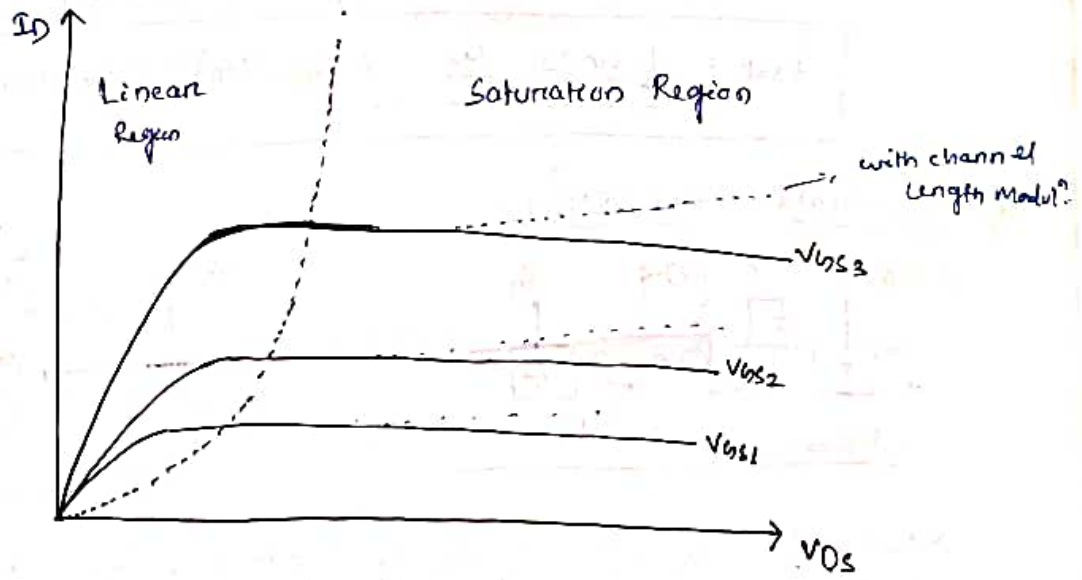
Beyond the linear region on saturation region

$$\boxed{V_{DS} \geq V_{DSAT} = V_{GS} - V_{TO}}$$

$$I_{Dsat} = \frac{\mu_n C_{ox}}{2} \cdot \frac{W}{L} \cdot \left[2 (V_{GS} - V_{TO})(V_{GS} - V_{TO}) - (V_{GS} - V_{TO})^2 \right]$$

$$\boxed{I_{Dsat} = \frac{\mu_n C_{ox}}{2} \frac{W}{L} (V_{GS} - V_{TO})^2} \quad \text{——— (3)}$$

Thus, the drain current I_D becomes a function only of the gate-source voltage V_{GS}, beyond the saturation boundary.

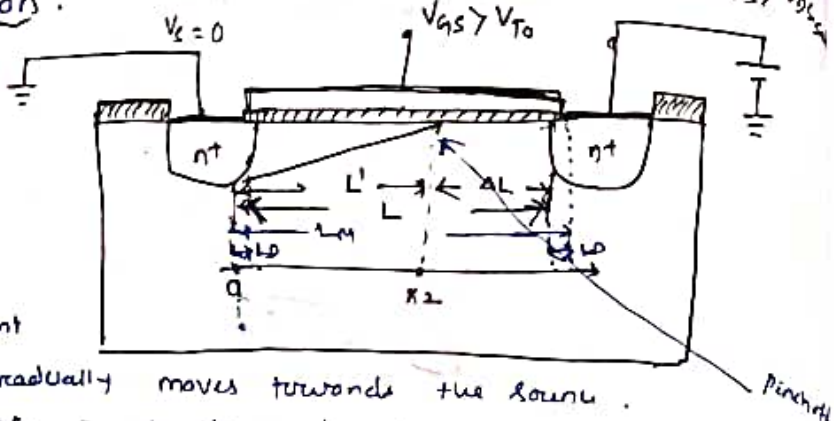


(Saturation of drain current.

* Gradual Channel Approximation (GCA):
  - → Substrate & source are connected to same potential, $V_{SB} = 0$.
  - → V_T is constant through out the channel & the electric field E_x is more prominent than E_y.
  - → The channel length is sub micron range i.e nanometer.

Parameters :

i) Channel Length Modulation :



→ If V_{DS} is slightly greater than $V_{GS} - V_{TH}$, then the inversion layer stops at $x \leq L$, & the channel pinch off.

→ As V_{DS} increases further, the point at which Q_d equal to zero gradually moves towards the source.

So $x = 0$ to $x = L'$ & $V(x) = 0$ to $V(x) = V_{GS} - V_{TO}$

L_M → Mask Length
L_D → Gate source & gate drain overlap
$L = L_M - 2L_D$

$$I_{D_{sat}} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L'} (V_{GS} - V_{TO})^2 \quad \text{—— (4)}$$

* The actual length of the inverted channel gradually decreases as the potential difference betⁿ gate & drain increases. So L' is a function of V_{DS}. This effect is called "channel length modulation".

$$L' = L - \Delta L$$

$$\Rightarrow \frac{1}{L'} = \frac{1}{L - \Delta L} = \frac{(1 + \Delta L)}{(L - \Delta L)(L + \Delta L)} = \frac{1 + \Delta L}{L^2 - (\Delta L)^2} \quad (\because \Delta L^2 \ll L^2 \text{ so negligible})$$

$$\Rightarrow \frac{1}{L'} = \frac{1 + \Delta L}{L^2} = \left(1 + \frac{\Delta L}{L}\right)/L \Rightarrow \boxed{\frac{1}{L'} = \frac{(1 + \lambda V_{DS})}{L}}$$

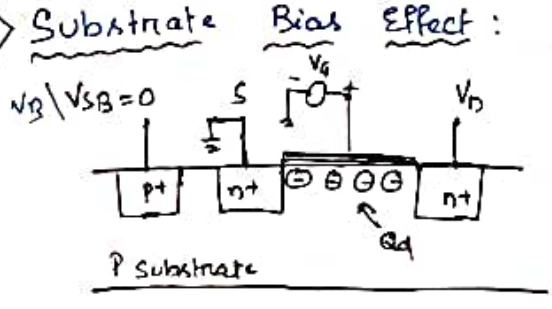
Let $\frac{\Delta L}{L} = \lambda V_{DS}$

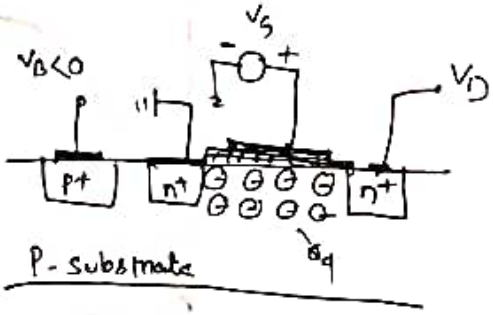
where λ is channel length modulation coefficient. For longer channel, λ is smaller.

By substituting L' value in eqⁿ (4) we get

$$\boxed{I_{Dsat} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TO})^2 (1 + \lambda V_{DS})} \quad \text{—— (5)}$$

ii) Substrate Bias Effect :





→ Let $V_S = V_D = 0$, & V_G is less than V_{TH}, so a depletion region is formed under the gate but no inversion layer exist.

→ As V_B becomes more negative, more holes are attracted to the substrate connection, leaving a larger negative charge behind, the depletion region becomes wider.

* The threshold voltage is a function of the total charge in the depletion region because the gate charge must mirror Q_d before

an inversion layer is formed.

* Thus, as V_B drops, V_{td} increases, V_{TH} also increases. This is called "body effect" or the "backgate effect".

→ With body effect

$$V_{T(V_{SB})} = V_{TO} + \gamma \left(\sqrt{|2\phi_F| + V_{SB}} - \sqrt{|2\phi_F|} \right)$$

Where $\gamma = \sqrt{2q\varepsilon_{si} N_{sub}} / C_{ox}$ denotes the body effect coefficient

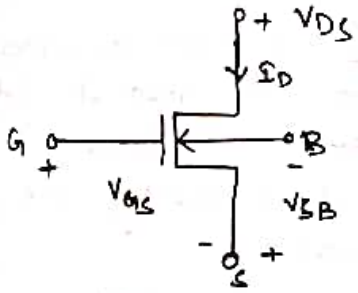
γ typically lies in the range of 0.3 to 0.4 $V^{1/2}$.

So

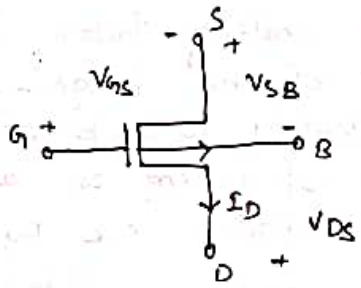
$$I_{D(lin)} = \frac{\mu_n C_{ox}}{2} \cdot \frac{W}{L} \cdot \left[2 (V_{GS} - V_{T(V_{SB})}) V_{DS} - V_{DS}^2 \right]$$

$$I_{D(sat)} = \frac{\mu_n C_{ox}}{2} \frac{W}{L} \cdot (V_{GS} - V_{T(V_{SB})}) (1 + \lambda . V_{DS})$$

Terminal Voltage & current of the nMOS & PMOS transistor:



(n-channel MOSFET)



(P-channel MOSFET)

Current-Voltage equation of the P-channel MOSFET:

$I_D = 0$ for $V_{GS} > V_T$

$$I_{D(lin)} = \frac{\mu_p C_{ox}}{2} \cdot \frac{W}{L} \left[2 (V_{GS} - V_T) V_{DS} - V_{DS}^2 \right]$$ for $V_{GS} \leq V_T$ & $V_{DS} > V_{GS} - V_T$

$$I_{D(sat)} = \frac{\mu_p C_{ox}}{2} \cdot \frac{W}{L} (V_{GS} - V_T)^2 (1 + \lambda . V_{DS})$$ for $V_{GS} \leq V_T$ & $V_{DS} \leqslant V_{GS} - V_T$

For n-channel MOSFET:

$I_D = 0$ for $V_{GS} < V_T$

$$I_{D(lin)} = \frac{\mu_n C_{ox}}{2} \cdot \frac{W}{L} \cdot \left[2 (V_{GS} - V_T) V_{DS} - V_{DS}^2 \right]$$ for $V_{GS} \geqslant V_T$ & $V_{DS} < V_{GS} - V_T$.

$$I_{D(sat)} = \frac{\mu_n C_{ox}}{2} \cdot \frac{W}{L} (V_{GS} - V_T)^2 (1 + \lambda . V_{DS})$$ for $V_{GS} \geqslant V_T$ & $V_{DS} \geqslant V_{GS} - V_T$.

MOSFET Scaling & Small-Geometry Effects:

→ The design of high density chips in MOS VLSI technology requires that the packing density of MOSFETs used in the ckts is as high as possible, consequently that the size of transistors are as small as possible.

→ The reduction of the size, i.e the dimensions of MOSFETs, is commonly refered to as scaling. It is expected that the operational characteristics of the MOS transistor will change with reduction of its dimension.

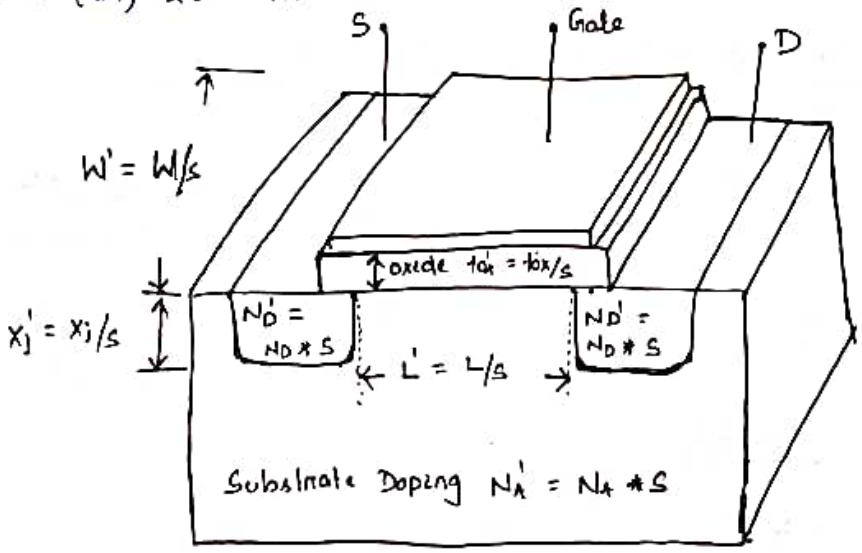
→ Scaling of MOS transistor is concerned with systemic reduction of overall dimensions of the devices as allowed by available technology, while preserving the geometric ratios found in the larger devices.

→ The proportional scaling of all devices in a ckt would certainly results in a reduction of the total silicon area occupied by the ckt, thereby increasing the overall functional density of the chip.

→ A constant scaling factor $S > 1$. All the horizontal & vertical dimensions of the large-size transistor are then divided by this scaling factor to obtain the ~~scaling~~ scaled device.

→ The extent of scaling is achievable is determined by the fabrication technology i.e by minimum feature size.

year	1985	1987	1989	1991	1993	1995	1997	1999
Feature size (μm)	2.5	1.7	1.2	1.0	0.8	0.5	0.35	0.25



(Scaling of MOSFET by a scaling factor of S).

→ The scaling of all dimensions by a factor of $S > 1$ leads to the reduction of the area occupied by the transistor by a factor of S^2.

→ There are two basic types of size-reduction strategies: Full Scaling (also called constant field scaling) & Constant-Voltage Scaling.

i) Full Scaling (Constant field Scaling):

→ This scaling option attempts to preserve the magnitude of internal electric fields in the MOSFETs, while the dimensions are scaled down by a factor s. To achieve this, all potentials must be scaled down proportionally by the same scaling factor.

→ Poisson eqn describe that the charge density must be increased by a factor s in order to maintain the field conditions.

Influence of full scaling:

Quantity	Before scaling	After Scaling
i) channel length	L	$L' = L/s$
ii) channel width	W	$W' = W/s$
iii) Gate Oxide thickness	t_{ox}	$t_{ox}' = t_{ox}/s$
iv) Junction depth	x_j	$x_j' = x_j/s$
v) Power supply voltage	V_{DD}	$V_{DD}' = V_{DD}/s$
vi) Threshold voltage	V_{To}	$V_{To}' = V_{To}/s$
vii) Doping densities	N_D N_A	$N_D' = s.N_D$ $N_A' = s.N_A$
viii) Oxide capacitance	C_{ox}	$C_{ox}' = s.C_{ox}$
ix) Drain Current	I_D	$I_D' = I_D/s$
x) Power dissipation	P	$P' = P/s^2$
xi) Power density	$P/Area$	$P'/Area = P/Area$.

→ The surface mobility μ_n, aspect ration W/L will remain unchanged under scaling.

* The power dissipation is an attractive features of full scaling. With the device area ~~dissipation~~ reduction by s^2, the power density per unit area remaining virtually unchanged for the scaled device.

ii) Constant - Voltage Scaling:

→ In full scaling the power supply voltage & all terminal voltages be scaled down proportionally with the device dimensions. This

→ In particular, the peripheral & interface circuitary may require certain voltage levels for all input & output voltages, which in turn would necessitate multiple power supply voltages & complicated level-shifter arrangements. For these reasons constant-voltage scaling is usually preferred over full scaling.

* In constant-voltage scaling, all dimensions of the MOSFET are reduced by a factor S. The power supply voltage & the terminal voltages, on the other hand, remain unchanged.

→ The doping densities must be increased by a factor of S^2 in order to preserve the charge-field regions.

* Quantity	Before Scaling	After scaling
i) Dimensions	W, L, tox, xj	Reduced by S ($W' = W/S$, $t_{ox}' = t_{ox}/S$, $L' = L/S$, $x_j' = x_j/S$)
ii) Voltages	VDD, VT	Remain unchanged
iii) Doping densities	NA, ND	Increased by S^2 ($N_A' = S^2 N_A$, $N_D' = S^2 N_D$)
iv) Oxide capacitance	Cox	$C_{ox}' = S \cdot C_{ox}$
v) Drain current	I_D	$I_D' = S \cdot I_D$
vi) Power Dissipation	P	$P' = S \cdot P$
vii) Power density	P/Area	$P'/Area' = S^3 \cdot \frac{P}{Area}$

* This scaling may preferred over full scaling in many practical cases because of the external voltage level constraints.

* This scaling increases the drain current & power density by a factor S^3 which may cause serious reliability problems for the scaled transistor such as electromigration, hot carrier degradation, oxide breakdown & electrical over stress.

iii) Short-channel Effect:

→ A MOSFET can be defined as a short-channel device if the effective channel length 'Leff' is approximately equal to the source & drain junction depth 'xj'.

→ The short-channel effect that arises in this case are attributed to two physical phenomena i) The limitation imposed on electron drift characteristics in the channel ii) The modification of the threshold voltage due to shortening channel length.

iv) Narrow-channel Effect:

→ MOS transistors that have channel width 'W' on the same order of magnitude as the maximum depletion region thickness X_{dm} are defined as narrow channel device.

→ The most significant narrow-channel effect is that the actual threshold voltage of such device is larger than that predicted by the conventional threshold voltage formula.

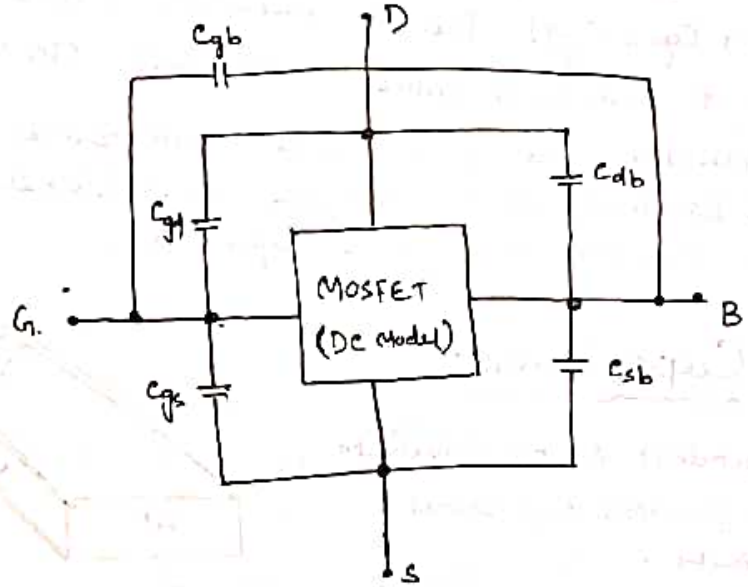
$$V_{T0}(\text{narrow channel}) = V_{T0} + \Delta V_{T0}.$$

MOSFET Capacitances:

* The onchip capacitances found in MOS ckts are not lumped but distributed.

* Based on their physical origins, the parasitic device capacitances can be classified into two major groups:
  i) Oxide-related capacitances
  ii) Junction capacitances.

i) Oxide-Related Capacitances:



→ The gate electrode overlaps both source region & the drain region at the edge. The two overlap capacitances that arise as a result of this structural arrangement are called $C_{GD(overlap)}$ & $C_{GS(overlap)}$ respectively.

$$C_{GS(overlap)} = C_{ox} \cdot W \cdot L_D$$
$$C_{GD(overlap)} = C_{ox} \cdot W \cdot L_D$$

Both source & drain have same W & C_{ox}.

with $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$

Both of these capacitances are voltage independent.

C_{gs} = Gate to Source capacitance
C_{gd} = " " Drain "
C_{gb} = " " Substrate "

} These are voltage dependent.

① In "cut off mode" surface is not inverted ∴ there is no conducting channel.

So $C_{gs} = C_{gd} = 0$

& $C_{gb} = C_{ox} \cdot W \cdot L$

② In "Linear mode" operation, the inverted channel extends across MOSFET bet^n source & drain. Due to conducting channel

$C_{gb} = 0$ & $C_{gs} \cong C_{gd} \cong \frac{1}{2} \cdot C_{ox} \cdot W \cdot L$

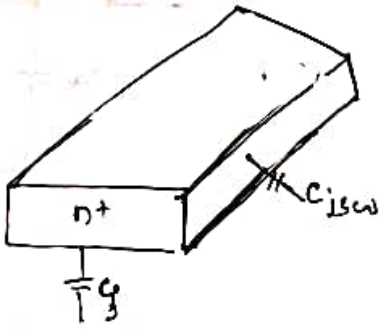
③ In "Saturation mode", the inversion layer on the surface doesn't extend to the drain, but it is pinch off.

$C_{gb} = 0$ & $C_{gd} = 0$ $\quad C_{gs} \cong \frac{2}{3} \cdot C_{ox} \cdot W \cdot L$

* The sum of all 3 voltage-dependent (distributed) gate-oxide capacitances ($C_{gb} + C_{gs} + C_{gd}$) has a minimum value $0.66\, C_{ox} W L$ (in saturation) & maximum value of $C_{ox} \cdot W \cdot L$ (in cut off & linear)

→ For simple calculation, all 3 capacitances are considered as connected parallel & a constant worst case value ($C_{ox} \cdot W(L + 2L_D)$) can be used for sum of MOSFET gate oxide capacitance.

ii) Junction Capacitance:



→ The voltage-dependent source-substrate & drain-substrate junction capacitances, C_{sb} & C_{db} respectively.

→ Both these capacitances are due to the depletion charge surrounding the respective source or drain diffusion regions embedded in the substrate.

→ Both these junctions are reverse biased under normal operating conditions of the MOSFET & that the amount of junction capacitance is a function of applied terminal voltage.

* The junction capacitance expression is

$$C_j(V) = A \cdot \sqrt{\frac{\varepsilon_{si} \cdot q}{2} \left(\frac{N_A \cdot N_D}{N_A + N_D}\right)} \cdot \frac{1}{\phi_0 - V}$$

In General form of junction capacitance

$$C_j(V) = \frac{A \cdot C_{j0}}{\left(1 - \frac{V}{\phi_0}\right)^m} \qquad \text{——— (1)}$$

where A = junction area.

ϕ_0 = Built-in junction potential = $\frac{KT}{q} \ln\left(\frac{N_A \cdot N_D}{n_i^2}\right)$

m = Grading coefficient, m = $\frac{1}{2}$ for abrupt junction
= $\frac{1}{3}$ for linearly graded junction

C_{j0} = The zero bias junction capacitance per unit area

$$\rightarrow \boxed{C_{j0} = \sqrt{\frac{\varepsilon_{si} \cdot q}{2}\left(\frac{N_A \cdot N_D}{N_A + N_D}\right) \cdot \frac{1}{\phi_0}}}$$

V = Reverse bias voltage (negative).

→ The equivalent large signal capacitance → (i.e change in bias condition)

$$\boxed{C_{eq} = -\frac{A \cdot C_{j0} \cdot \phi_0}{(V_2 - V_1)(1-m)}\left[\left(1-\frac{V_2}{\phi_0}\right)^{1-m} - \left(1-\frac{V_1}{\phi_0}\right)^{1-m}\right]}$$

C_{eq} is always calculated for a transition bet^n two known voltage levels. (Let the reverse voltage changes V_1 to V_2)

on $\boxed{C_{eq} = \frac{1}{V_2 - V_1}\int_{V_1}^{V_2} C_j(V)\, dV}$

* Assume that the side wall doping density is given by $N_{A(SW)}$, the zero bias capacitance per unit area can be

$$\boxed{C_{josw} = \sqrt{\frac{\varepsilon_{si} \cdot q}{2}\left(\frac{N_{A(SW)} \cdot N_D}{N_{A(SW)} + N_D}\right) \cdot \frac{1}{\phi_{osw}}}} \quad \text{—(2)}$$

Where ϕ_{osw} is the built in potential of side wall junction.

Modeling of Mos Transistor Using Spice:

* Spice (Simulation program with integrated ckt emphasis) is a general purpose ckt simulator which is used very widely both in the microelectronic industry & educational institution for ckt design.

→ The SPICE software had 3 built in MOSFET models:

i) LEVEL1 (MOS1) is described by a square law current voltage characteristics.

ii) LEVEL 2 (MOS2) is a detailed analytical MOSFET model

iii) LEVEL 3 (MOS3) is a semi empirical model.

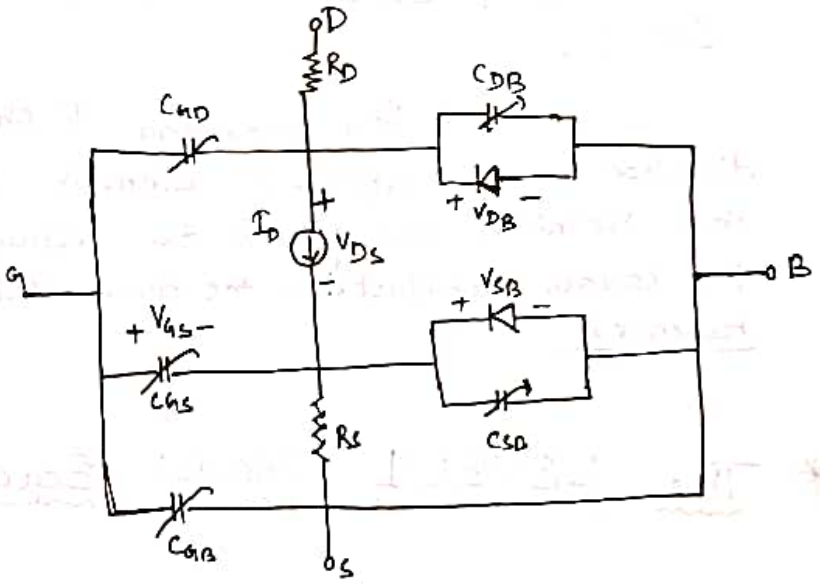
→ Both MOS2 & MOS3 include 2nd order effects such as short channel threshold voltage, subthreshold conduction, Charge control capacitance scattering limited velocity saturation etc.

* Basic Concept:

CKT description:

→ The equivalent ckt structure of NMOS LEVEL 1 model, which is the default MOSFET model in SPICE is shown.

→ The Basic structure is also typical for the LEVEL 2 & LEVEL 3 models.



(Equivalent ckt Structure of the LEVEL1 MOSFET model in SPICE).

→ The voltage-controlled current source I_D determines the steady state current-voltage behaviour of the device.

→ The voltage-controlled (nonlinear) capacitors connected bet the terminals represent the parasitic oxide-related & junction capacitance.

→ The source-substrate & the drain-substrate junctions, which are reverse biased under normal operating conditions are represented by ideal diodes in this equivalent ckt.

→ Finally, the parasitic source & drain resistances are represented the resistor R_D & R_S respectively, connected between the drain current source & the respective terminals.

* The basic geometry of an nMOS transistor can be described by specifying the nominal channel (gate) length L & the channel width W, i, t, both of which are indicated on the element description line.

→ The channel width W is, by definition, the width of the area covered by the thin gate oxide.

→ The effective channel length Leff is defined as the distance on the surface bet^n the two (source & drain) diffusion regions.

→ Thus in order to find the effective channel length, the gate-source overlap distance & gate-drain overlap distance must be subtracted from the nominal (mask) gate length specified on the device description line.

→ The amount of gate overlap over source & drain can be specified by using the lateral diffusion coefficient LD in SPICE.

* For modeling P channel mos transistors, the direction of the dependent current source, the polarities of the terminal voltages, & the direction of two diodes representing the source-substrate & the drain-substrate junction must be reversed.

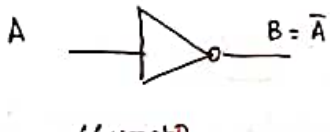
* The LEVEL 1 Model Equations:

Saikam

Chapter: 5

Mos Inverters:

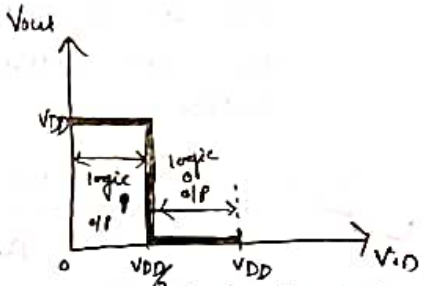
Introduction :



(Symbol)

A	B
0	1
1	0

(Truth table)



(VTC of ideal inverter)

→ In MOS inverter, both input variable A & o/p variable B are represented by node voltage, referred to ground potential.

→ Using positive logic convention, the Boolean (logic) value of '1' can be represented by a VDD & logic '0' represented by low voltage 0.

→ The DC voltage transfer cls (VTC) of ideal inverter is shown in the fig.

→ The voltage V_{TH} is called inverter Threshold voltage. For any i/p voltage bet^n 0 & $V_{TH} = V_{DD}/2$, the o/p voltage equal to VDD (logic 1)

→ The output switches from VDD to 0 when i/p equal to V_{TH}.

→ For Any i/p voltage bet^n V_{TH} & VDD o/p assumes a value 0.

* Thus an input voltage $0 \leq V_{in} < V_{TH}$ is interpreted by this ideal inverter as logic 0 & input voltage $V_{TH} < V_{in} \leq V_{DD}$ is interpreted ~~but~~ as logic 1.

NMOS Inverters:

→ The input voltage of the inverter ckt is also the gate-to-source voltage of the nMOS transistor ($V_{in} = V_{GS}$), while the output voltage of the ckt is equal to the drain to source voltage ($V_{out} = V_{DS}$).

→ The source & the substrate terminals of the nMOS transistor, also called the driven transistor, are connected to ground potential, hence source to substrate voltage is $V_{SB} = 0$.

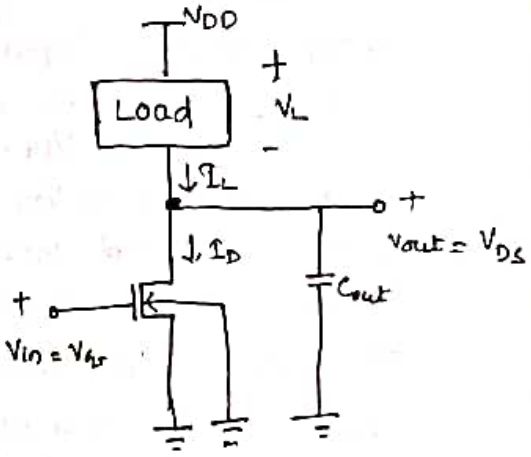


Fig: (General ckt structure of an nMos inverter)

→ In this generalized representation, the load device is represented as a two-terminal ckt element with terminal current I_L & terminal voltage V_L (2_L).

→ One terminal of the load device is connected to the drain of the n-channel MOSFET, while the other terminal is connected to VDD, the power supply voltage.

→ The characteristics of the inverter ckt actually depend very strongly upon the type & cls of the load device.

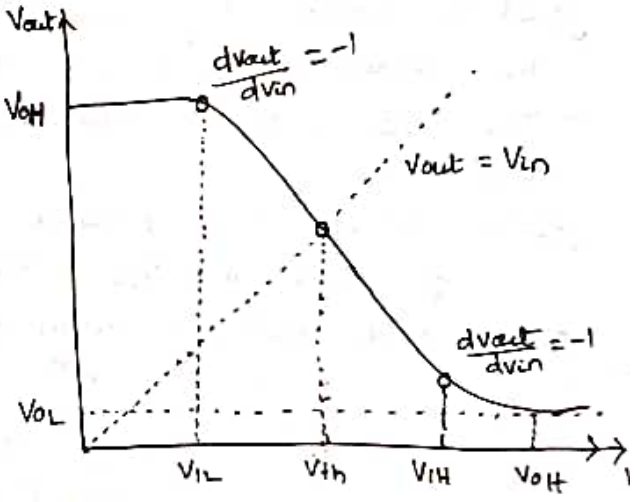
→ Since the DC gate current of an MOS transistor is negligible for all practical purposes, there will no current flow into or out of the input & output terminal of the inverter in DC steady state.

Voltage Transfer characteristics (VTC):

By applying Kirchoff's current Law (KCL) to ckt (fig 1), the load current is equal to nMOS drain current.

$$I_D(V_{in}, V_{out}) = I_L(V_L)$$

→ The VTC describes V_{out} as a function of V_{in} under DC condition.



→ For very low input voltage levels, the output voltage V_{out} is equal to the high value of V_{OH} (output high voltage)

→ In this case, the driver nMOS transistor is in cut off, hence does not conduct any current.

→ As V_{in} increases, the driver transistor starts conducting a certain drain current & V_{out} starts to decrease.

→ There are two critical points on this curve, where slope of the $V_{out}(V_{in})$ (VTC of nMOS Transistor) cls becomes equal to -1.

i.e $$\boxed{\frac{dV_{out}}{dV_{in}} = -1}$$

→ The smaller input voltage satisfy this condition is called input low voltage V_{IL} & the larger V_{in} satisfy this cond^n is called input high voltage V_{IH}.

→ When $V_{in} = V_{OH}$, the V_{out} drops reach a value V_{OL}.

→ The inverted threshold voltage V_{th} is considered as the transition voltage & defined as the point where $V_{in} = V_{out}$ on the VTC.

The Functional definitions are:

V_{OH} : Max^m Output Voltage when the output level is logic 1

V_{OL} : Min^m " " " " " " " " " 0

V_{IL} : Max^m Input " which can be interpreted as logic 0'

V_{IH} : Min^m " " " " " " " " 1.

Noise Margins:

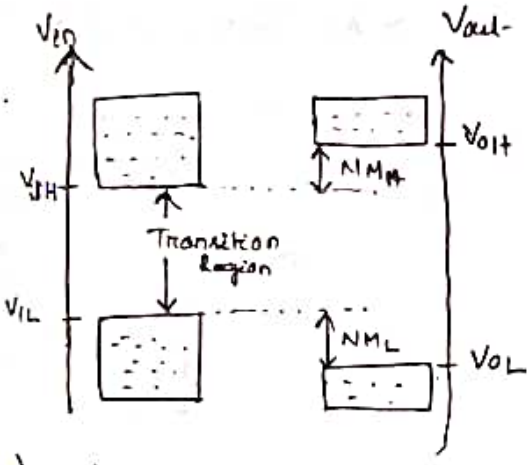
→ The ckt noise is the unwanted signal.

→ The noise tolerance for digital ckt called noise margin, denoted by NM.

→ The noise immunity of the ckt increases with NM.

→ Two noise margin will be defined: the noise margin for low signal level (NM_L) & the noise margin for high signal level (NM_H)

$$N_{ML} = V_{IL} - V_{OL}$$
$$N_{MH} = V_{OH} - V_{IH}$$



(The noise margin of NM_L & NM_H. The shaded regions indicate valid high & low levels of i/p & o/p signal.

→ Generally $V_{out} = f(V_{in})$

when noise is added to input

$$V_{out}' = f(V_{in} + \Delta V_{noise})$$

By Taylor series

$$V_{out}' = f(V_{in}) + \frac{dV_{out}}{dV_{in}} \cdot \Delta V_{noise} + \text{higher order terms neglect.}$$

So Perturbed output = Nominal output + Gain × External Perturbation.

Resistive Load Inverter:

→ Here an enhancement-type nMOS transistor act as driver ckt & a simple linear resistor R_L act as load.

→ In DC steady state operation

$$I_D = I_R.$$

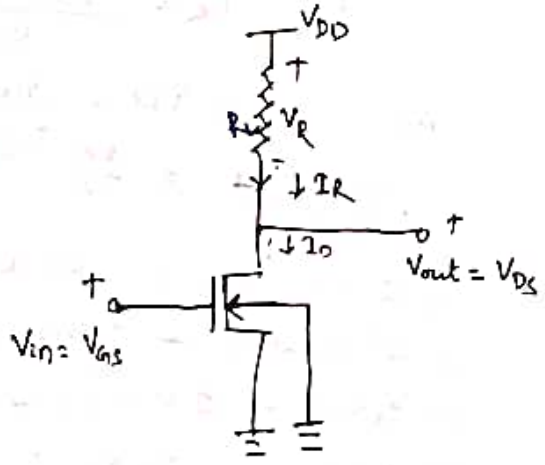
→ $V_{SB} = 0$ as source & substrate is grounded.

→ If $V_{in} < V_{TO}$, transistor is cut off hence doesn't conduct any current.

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

$$\Rightarrow \boxed{V_{out} = V_{DD}} \quad (\because I_R = I_D = 0)$$

→ As V_{in} increase beyond V_{TO}, transistor start conducting & non zero drain current.



Resistive-load inverter ckt.

→ As $(V_{DS} = V_{out}) > (V_{in} - V_{TO})$, MOSFET operates in saturation.

Thus $\boxed{I_R = \frac{k_n}{2}(V_{in} - V_{TO})^2}$

→ As V_{in} increase I_D increase hence V_{out} decrease. When $V_{in} > V_{out} + V_{TO}$, .. transistor enters linear mode.

$$\boxed{I_R = \frac{k_n}{2}\left[2(V_{in} - V_{TO}) \cdot V_{out} - V_{out}^2\right]}$$

Calculation:

V_{OH} :

We know that $V_{out} = V_{DD} - R_L \cdot I_R$

When V_{in} is low i.e less than V_{TO}, $I_D = I_R = 0$

So $V_{out} = V_{DD}$

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

V_{OL} :

Let us assume $V_{in} = V_{OH} = V_{DD}$. , $V_{out} = V_{DS} = V_{OL}$

As $V_{in} - V_{TO} > V_{out}$, transistor operates in linear region.

So $I_R = \frac{V_{DD} - V_{out}}{R_L}$.

As $I_R = I_D$, then

$$\frac{V_{DD} - V_{out}}{R_L} = \frac{k_n}{2}\left[2(V_{DD} - V_{TO})V_{OL} - V_{OL}^2\right]$$

$$\Rightarrow 2(V_{DD} - V_{OL}) = k_n \cdot R_L \left[2(V_{DD} - V_{TO})V_{OL} - V_{OL}^2\right]$$

$$\Rightarrow 2(V_{DD} - V_{OL}) - 2k_n R_L (V_{DD} - V_{TO})V_{OL} = -k_n \cdot R_L V_{OL}^2$$

$$\Rightarrow 2V_{DD} - 2V_{OL}\left[1 + k_n R_L (V_{DD} - V_{TO})\right] = -k_n \cdot R_L V_{OL}^2$$

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

$$\rightarrow V_{OL}^2 - 2\left(V_{DD} - V_{TO} + \frac{1}{k_n \cdot R_L}\right)V_{OL} + \frac{2}{k_n R_L} \cdot V_{DD} = 0$$

By solving

$$\boxed{V_{OL} = V_{DD} - V_{TO} + \frac{1}{k_n R_L} - \sqrt{\left(V_{DD} - V_{TO} + \frac{1}{k_n R_L}\right)^2 - \frac{2V_{DD}}{k_n R_L}}}$$

V_{IL} :

When $V_{out} > V_{in} - V_{TO}$, transistor operates in saturation.

$$I_R = I_D$$

$$\Rightarrow \frac{V_{DD} - V_{out}}{R_L} = \frac{K_n}{2}(V_{in} - V_{TO})^2$$

differentiate both side w.r.t V_{in} ~~is equal to (-1) at V_{IL}~~

$$-\frac{1}{R_L}\frac{\partial V_{out}}{\partial V_{in}} = K_n.(V_{in} - V_{TO})$$

since the derivative of output voltage w.r.t input voltage equal to (-1) at V_{IL}, By substituting this

$$-\frac{1}{R_L}(-1) = K_n(V_{IL} - V_{TO})$$

$$\Rightarrow \boxed{V_{IL} = V_{TO} + \frac{1}{K_n R_L}}$$

So V_{out} is, $V_{out}(V_{in} = V_{IL}) = V_{DD} - \frac{K_n R_L}{2}\left(V_{TO} + \frac{1}{K_n R_L} - V_{TO}\right)^2$

$$\Rightarrow V_{out}(V_{in} = V_{IL}) = V_{DD} - \frac{1}{2K_n R_L}$$

V_{IH} :

When $V_{out} < V_{in} - V_{TH}$, transistor operates in Linear Region.

By KCL, $I_R = I_D$

$$\Rightarrow \frac{V_{DD} - V_{out}}{R_L} = \frac{K_n}{2}\left[2(V_{in} - V_{TO})V_{out} - V_{out}^2\right]$$

differentiate both side w.r.t V_{in}

$$-\frac{1}{R_L}\frac{\partial V_{out}}{\partial V_{in}} = \frac{K_n}{2}\left[2(V_{in} - V_{TO})\frac{dV_{out}}{dV_{in}} + 2V_{out} - 2V_{out}.\frac{dV_{out}}{dV_{in}}\right]$$

When $V_{in} = V_{IH}$, $\frac{dV_{out}}{dV_{in}} = -1$, by substituting this

$$-\frac{1}{R_L}(-1) = K_n\left[(V_{IH} - V_{TO})(-1) + 2V_{out}\right]$$

$$\Rightarrow \boxed{V_{IH} = V_{TO} + 2V_{out} - \frac{1}{K_n R_L}}$$

By substituting the value.

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

By solving ~~Value for~~ $V_{out}(V_{in} = V_{IH}) = \sqrt{\frac{2}{3}\frac{V_{DD}}{K_n.R_L}}$

Again by substituting this $\boxed{V_{IH} = V_{TO} + \sqrt{\frac{8}{3}\frac{V_{DD}}{K_n R_L}} - \frac{1}{K_n R_L}}$.

Inverters With n-type MOSFET Load:

→ As the silicon area occupied by the transistor is smaller than resistive load, so MOSFET is used as load device.

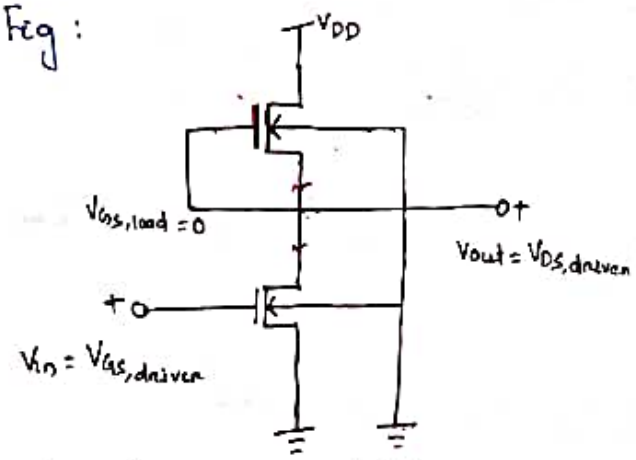
→ Enhancement-load nmos inverters are not used in large scale digital application as it suffers from relatively high stand by (DC) power dissipation.

Depletion-Load nMOS Inverter:

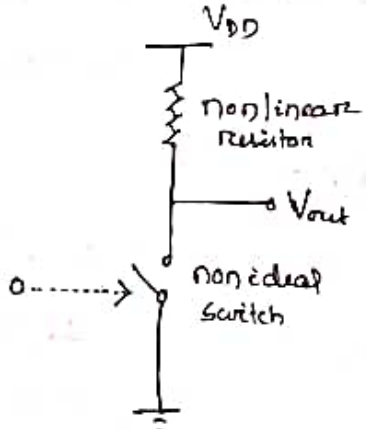
Advantages of depletion-load inverters

i) Sharp VTC transition & better noise margin
ii) Single power supply
iii) Smaller overall layout area.

Fig:



(Inverter ckt with depletion type n-Mos load)



(simplified equivalent ckt consisting a nonlinear load resistor & nonideal switch controlled by the input)

→ The driver device is an enhancement type n-mos transistor, with $V_{To,driver} > 0$, whereas the load is a depletion-type nmos transistor with $V_{To,load} < 0$.

→ The gate & source node of the load transistor are connected so $V_{GS,Load} = 0$ always.

→ As threshold voltage of depletion-type load is negative, $V_{GS_{load}} > V_{T,load}$ is satisfied & load has a conducting channel regardless of input, output voltages.

→ Both the transistors are built on the same p-type substrate which is connected to ground. The load device is subject to the substrate-bias effect, so that its ~~substrate~~ threshold voltage is a function of its source to substrate voltage).

$V_{SB\,load} = V_{out}$.

$$V_{T,load} = V_{T0,load} + \gamma\left(\sqrt{|2\phi_F| + V_{out}} - \sqrt{|2\phi_F|}\right)$$

→ When $V_{out} < V_{DD} + V_{T,load}$, the load transistor is in saturation.
i.e $V_{DS,load} > V_{GS,load} - V_{T,load}$

then

$$I_{D,load} = \frac{k_{n,load}}{2}\left[-V_{T,load}(V_{out})\right]^2 = \frac{k_{n,load}}{2}\cdot\left|V_{T,load}(V_{out})\right|^2$$

For $V_{out} > V_{DD} + V_{T,load}$, load transistor is in linear region.

$$I_{D,load} = \frac{k_{n,load}}{2}\left[2\left|V_{T,load}(V_{out})\right|(V_{DD} - V_{out}) - (V_{DD} - V_{out})^2\right]$$

The VTC of this inverter can be constructed by setting,

$I_{D,driven} = I_{D,load}$
$V_{GS,driver} = V_{in}$
$V_{DS,driver} = V_{out}$.

Calculation:

V_{OH}:

When $V_{in} < V_{T0,driver}$, then driven is in cutoff & not conduct current. But load transistor operates in linear region has also zero drain current. Let $I_{D,load} = 0$ & substituting $V_{OH} = V_{out}$

$$I_{D,load} = \frac{k_{n,load}}{2}\left[2\left|V_{T,load}(V_{OH})\right|(V_{DD} - V_{OH}) - (V_{DD} - V_{OH})^2\right] = 0$$

The only valid solution in linear region is, $\boxed{V_{OH} = V_{DD}}$

V_{OL} =

Assume $V_{in} = V_{OH} < V_{DD}$. So driver operates in knear & load in saturation.

$$I_{D,load} = I_{D,driven}$$

$$\frac{k_{driven}}{2}\left[2(V_{OH} - V_{T0}).V_{OL} - V_{OL}^2\right] = \frac{k_{load}}{2}\left[-V_{T,load}(V_{OL})\right]^2$$

V_{OL} can be solve by temporarily neglecting the dependence of $V_{T,load}$ on V_{OL}. So

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

V_{IL} :

→ The slope of V_{TC} is (−1) i.e $\frac{dV_{out}}{dV_{in}} = -1$ when $V_{in} = V_{IL}$

Here driver operates in saturation while load operates in linear region.

By KCL $I_{D\,driver} = I_{D\,load}$

$$\frac{K_{driver}}{2}(V_{in} - V_{TO})^2 = \frac{K_{load}}{2}\left[2|V_{T,load}(V_{out})|(V_{DD} - V_{out}) - (V_{DD} - V_{out})^2\right]$$

differentiate both sides w.r.t V_{in}

$$K_{driver}(V_{in} - V_{TO}) = \frac{K_{load}}{2}\left[2|V_{T,load}(V_{out})|\left(-\frac{dV_{out}}{dV_{in}}\right) + 2(V_{DD} - V_{out})\left(-\frac{dV_{T,load}}{dV_{in}}\right) - 2(V_{DD} - V_{out})\left(-\frac{dV_{out}}{dV_{in}}\right)\right]$$

Here $\frac{dV_{T,load}}{dV_{in}}$ is negligible w.r.t others & substitute V_{IL} for V_{in} &

put $\frac{dV_{out}}{dV_{in}} = -1$,

$$\boxed{V_{IL} = V_{TO} + \frac{K_{load}}{K_{driver}}\left[V_{out} - V_{DD} + |V_{T\,load}(V_{out})|\right]}$$

V_{IH} :

Here driver operates in linear & load operates in saturation.

$$I_{D,driver} = I_{D,load}$$

$$\Rightarrow \frac{K_{driver}}{2}\left[2(V_{in} - V_{TO}).V_{out} - V_{out}^2\right] = \frac{K_{load}}{2}\left[-V_{T,load}(V_{out})\right]^2$$

Differentiating both sides w.r.t V_{in}

$$K_{driver}\left[V_{out} + (V_{in} - V_{TO})\frac{dV_{out}}{dV_{in}} - V_{out}.\left(\frac{dV_{out}}{dV_{in}}\right)\right] = K_{load}\left[-V_{T,load}(V_{out})\right].\left(\frac{dV_{T\,load}}{dV_{out}}\right).\frac{dV_{out}}{dV_{in}}$$

Substitute $dV_{out}/dV_{in} = -1$ & $V_{in} = V_{IH}$.

$$\boxed{V_{IH} = V_{TO} + 2V_{out} + \left(\frac{K_{load}}{K_{driver}}\right).\left[-V_{T,load}(V_{out})\right].\left(\frac{dV_{T,load}}{dV_{out}}\right)}$$

Here $\frac{dV_{T,load}}{dV_{out}} = \frac{\gamma}{2\sqrt{|2\phi_F| + V_{out}}}$ is not neglected.

CMOS Inverter :-

→ An inverter which is consists of an enhancement - type nmos transistor & an enhancement - type pmos transistor, operating in complementary mode is called complementary MOS (CMOS).

→ The ckt topology is complementary push-pull in the sense that for high input, the nmos transistor drives (pulls down) the output node while the pmos transistor act as load, & for low input, the pmos transistor drives (pulls up) the output node while nmos transistor act as load.

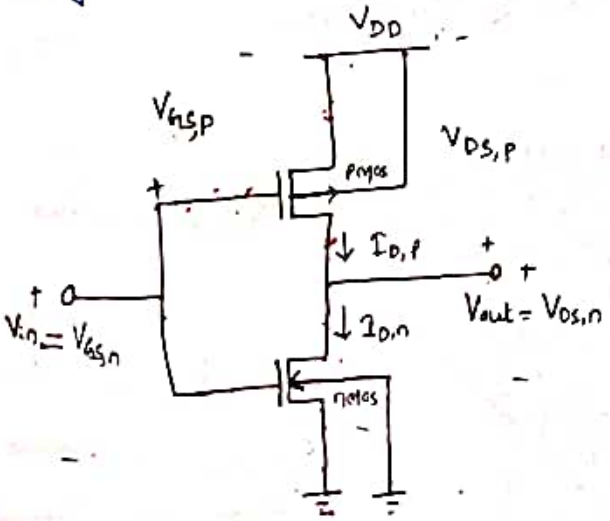
Advantage :

i) The steady state power dissipation is virtually negligible except for small power dissipation due to leakage current.

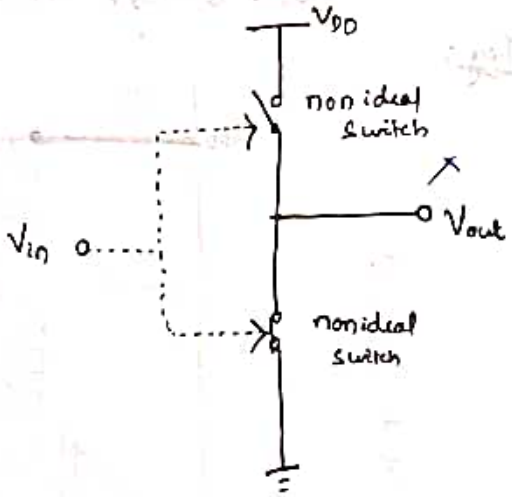
ii) The VTC exhibits a full output voltage swing bet^n 0V & V_{DD}, & that the VTC transition is usually very sharp. Thus VTC of the CMOS inverter resembles that of an ideal inverter.

Circuit Operation :-

Fig :



(CMOS Inverter Circuit)



(Simplified View of the CMOS inverter, consisting of two complementary nonideal switches.

→ The substrate of the nmos transistor is connected to the ground, while the substrate of the pmos transistor is connected to the power supply voltage, V_{DD}, in order to reverse-bias the source & drain junction.

From the ckt:

$$\left.\begin{aligned} V_{GS,n} &= V_{in} \\ V_{DS,n} &= V_{out} \end{aligned}\right\} \text{ — eqn (1)}$$

also
$$\left.\begin{aligned} V_{GS,p} &= -(V_{DD} - V_{in}) \\ V_{DS,p} &= -(V_{DD} - V_{out}) \end{aligned}\right\} \text{ — (2)}$$

→ When $V_{in} < V_{TO,n}$, the nmos transistor is cut off, pmos transistor operating in linear region.

→ Since $I_{D,n} = I_{D,p} = 0$, $V_{DS,p}$ of pmos is also equal to zero.

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

→ When V_{in} exceeds $(V_{DD} + V_{TO,p})$, pmos turn off & nmos operates in linear region, but $V_{DS,n} = 0$ due to $I_{Dp} = I_{Dn} = 0$.

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

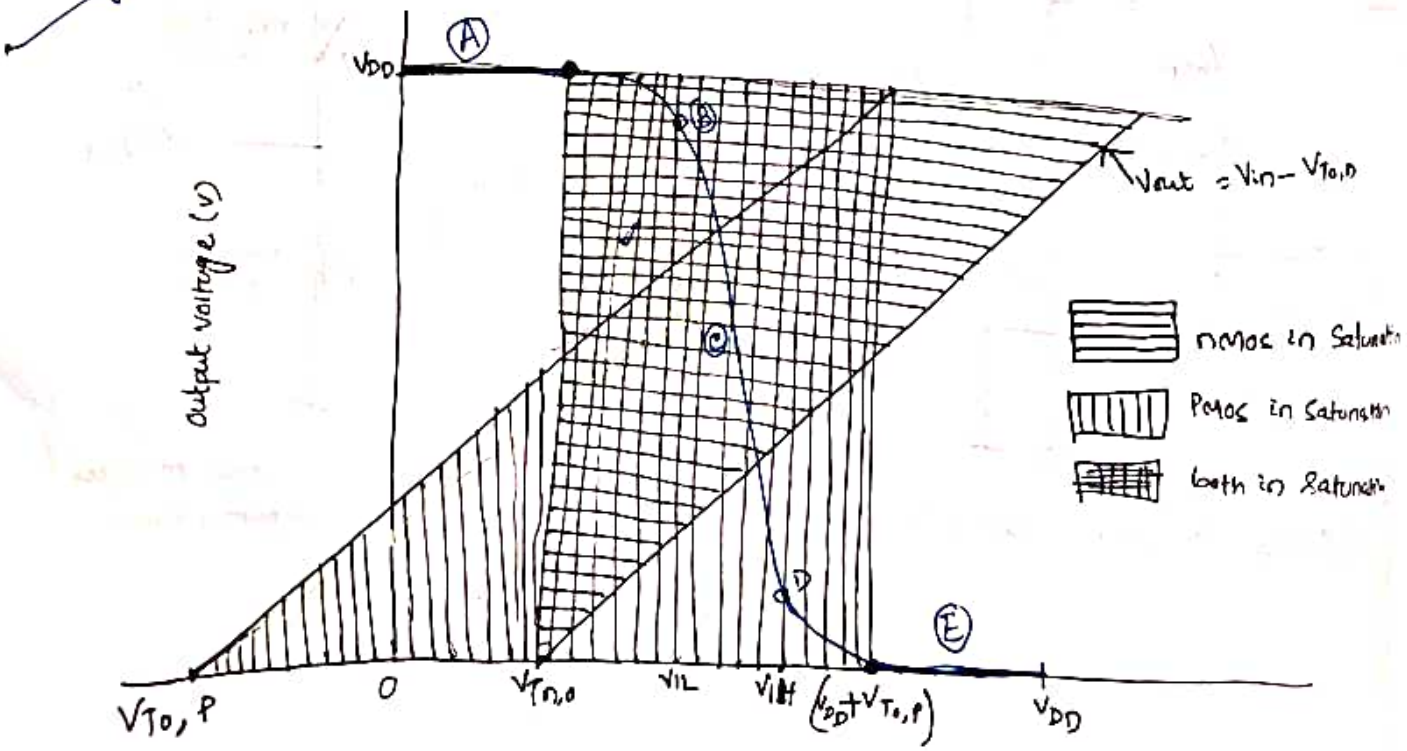
→ The nmos transistor operates in saturation if $V_{in} > V_{TO,n}$ & if the following condition is satisfied.

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

Pmos transistor operates in saturation if $V_{in} < (V_{DD} + V_{TO,p})$ & if

$$V_{DS,p} \leq V_{GS,p} - V_{TO,p} \Leftrightarrow V_{out} \leq V_{in} - V_{TO,p}.$$

Fig:



Operating regions of the nmos & Pmos Transistor

Calculation of V_{IL} :

When $V_{in} = V_{IL}$, the nmos operates in saturation & pmos operates in linear region.

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

$$\Rightarrow \frac{k_n}{2}(V_{GS,n} - V_{To,n})^2 = \frac{k_p}{2}\left[2(V_{GS,p} - V_{To,p})V_{DS,p} - V_{DS,p}^2\right]$$

Using eqn (1) & (2)

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

Differentiating both side w.r.t V_{in}

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

Substituting $V_{in} = V_{IL}$ & $\frac{\partial V_{out}}{\partial V_{in}} = -1$

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

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

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

V_{IH} :

When V_{in} is V_{IH} nmos operates in linear region & pmos operates in saturation region.

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

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

Using eqn (1) & (2)

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

Differentiating both sides w.r.t to V_{in}

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

Substituting $V_{in} = V_{IH}$ & $\frac{\partial V_{out}}{\partial V_{in}} = -1$

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

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

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

V_{TH} :

When $V_{in} = V_{th} = V_{out}$

both nmos & pmos operate in the saturation region.

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

$\Rightarrow (V_{in} - V_{TO,n}) = \sqrt{\frac{k_p}{k_n}} \cdot (V_{in} - V_{DD} - V_{TO,p})$

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

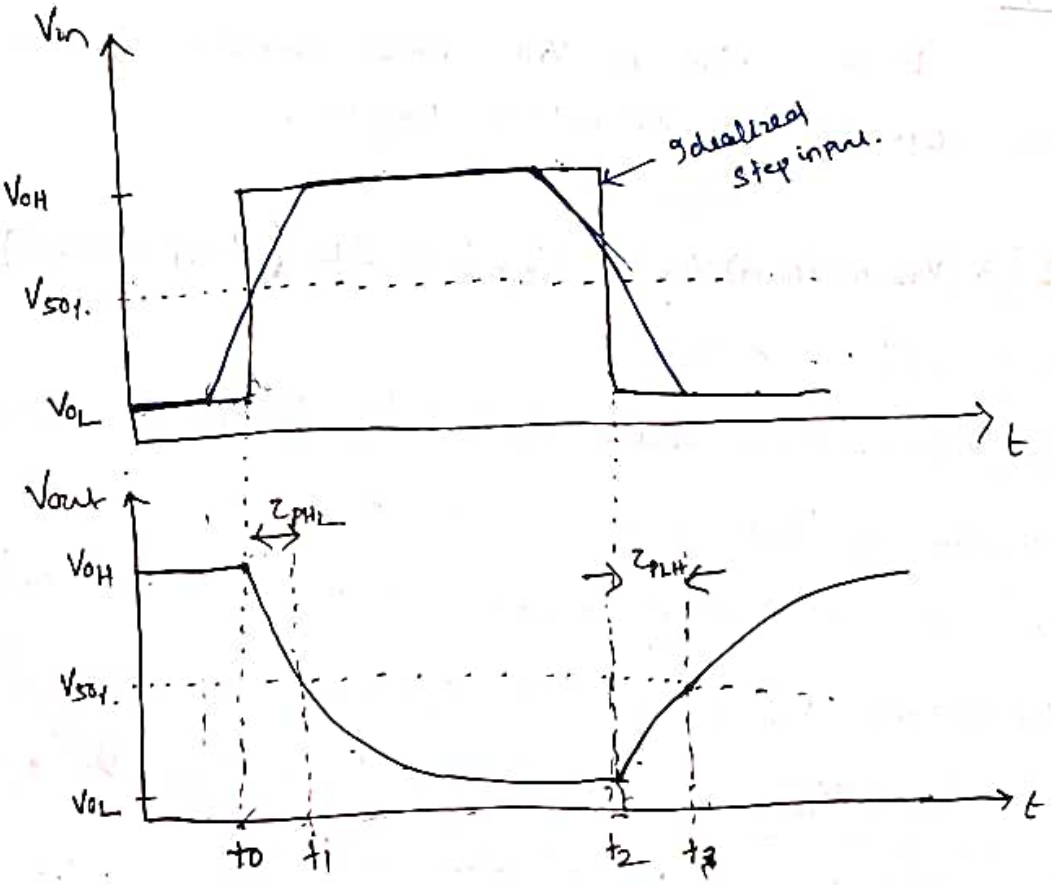
Correct solution for V_{in}

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

$$V_{th} = \frac{V_{TO,n} + \frac{1}{K_R}(V_{DD} + V_{TO,p})}{1 + \sqrt{\frac{1}{K_R}}}$$

Delay Time Definitions :

Fig:



(Input & output voltage waveforms of a typical inverter & the definitions of propagation delay times. The input voltage is idealized as a step pulse for simplicity)

The combined capacitance at the output node will be called the load capacitance.

$$C_{Load} = C_{gd,n} + C_{gd,p} + C_{db,p} + C_{db,n} + C_{int} + C_g$$

where C_{int} = Lumped interconnect capacitance
C_g = Capacitance due to thin oxide

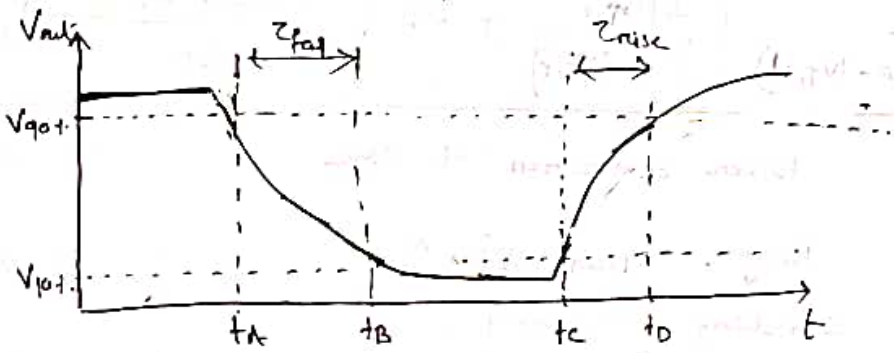
τ_{PHL} : is the time delay between the $V_{50\%}$ transition of the rising input voltage & $V_{50\%}$ transition of the falling output voltage.

τ_{PLH} : is the time delay bet^n the $V_{50\%}$ transition of the falling input voltage & $V_{50\%}$ transition of the rising output voltage.

τ_p : Average propagation delay

$$\tau_p = \frac{\tau_{PHL} + \tau_{PLH}}{2}$$

Fig:



(Output voltage rise & fall times)

* The "rise time" τ_{rise} is defined as the time required for the output voltage to rise from $V_{10\%}$ level to $V_{90\%}$ level.

* The "Fall Time" τ_{fall} is defined as the time required for the output voltage to drop from $V_{90\%}$ level to $V_{10\%}$ level.

$$V_{10\%} = V_{OL} + 0.1\,(V_{OH} - V_{OL})$$
$$V_{90\%} = V_{OL} + 0.9\,(V_{OH} - V_{OL})$$

Calculation of Delay times:

$$\tau_{PHL} = \frac{C_{load} \cdot \Delta V_{HL}}{I_{avg,HL}} = \frac{C_{load}\,(V_{OH} - V_{50\%})}{I_{avg,HL}} \quad \text{— (1)}$$

$$= \frac{C_{load}}{k_n (V_{OH} - V_{T,n})} \left[\frac{2V_{T,n}}{V_{OH} - V_{T,n}} + \ln\left(4\,\frac{(V_{OH} - V_{T,n})}{V_{OH} + V_{OL}} - 1 \right) \right]$$

For CMOS $V_{OH} = V_{DD}$, & $V_{OL} = 0$

So
$$\tau_{PHL} = \frac{C_{load}}{k_n(V_{DD} - V_{T,n})}\left[\frac{2V_{T,n}}{V_{DD} - V_{T,n}} + \ln\left(\frac{4(V_{DD} - V_{T,n})}{V_{DD}} - 1\right)\right]$$

Similarly

$$\tau_{PLH} = \frac{C_{load} \cdot \Delta V_{LH}}{I_{avg,LH}} = \frac{C_{load}(V_{SOH} - V_{OL})}{I_{avg,LH}}$$

$$= \frac{C_{Load}}{k_p(V_{OH} - V_{OL} - |V_{T,P}|)}\left[\frac{2|V_{T,P}|}{V_{OH} - V_{OL} - |V_{T,P}|} + \ln\left(\frac{2(V_{OH} - V_{OL} - |V_{T,P}|)}{V_{OH} - V_{SOH}} - 1\right)\right]$$

$$\tau_{PLH} = \frac{C_{load}}{k_p(V_{DD} - |V_{T,P}|)}\left[\frac{2|V_{T,P}|}{V_{DD} - |V_{T,P}|} + \ln\left(\frac{4(V_{DD} - |V_{T,P}|)}{V_{DD}} - 1\right)\right]$$

Switching Power dissipation of CMOS

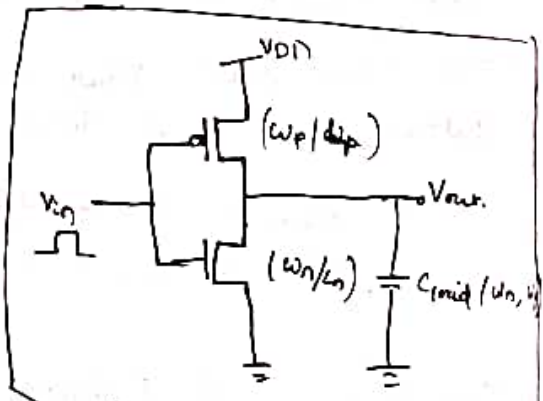
$$P_{avg} = C_{Load} \times V_{DD}^2 \times f$$

Where f = Switching frequency.

Inverter Design with Delay Constraints :

The combined load capacitance of the inverter is a function of transistor size.

→ For a given required (target) delay value of τ_{PHL}, the (W/L) ratio of the nmos transistor is



(General ckt structure considered in the inverter design problem)

$$\left(\frac{W_n}{L_n}\right) = \frac{C_{Load}}{\tau_{PHL} k_n C_{ox}(V_{DD} - V_{T,n})}\left[\frac{2V_{T,n}}{V_{DD} - V_{T,n}} + \ln\left(\frac{4(V_{DD} - V_{T,n})}{V_{DD}} - 1\right)\right]^{-1}$$

Similarly, the (W/L) ratio of PMOS transistor to satisfy a given value of τ_{PLH} is

$$\left(\frac{W_P}{L_P}\right) = \frac{C_{load}}{\tau_{PLH} k_p C_{ox}(V_{DD} - |V_{T,P}|)}\left[\frac{2|V_{T,P}|}{V_{DD} - |V_{T,P}|} + \ln\left(\frac{4(V_{DD} - |V_{T,P}|)}{V_{DD}} - 1\right)\right]^{-1}$$

The transistor dimension should be chosen so that all delay times remain below the required target value.

$$C_{load} = C_{gd,n}(W_n) + C_{gd,p}(W_p) + C_{db,n}(W_n) + C_{db,p}(W_p) + C_{int} + C_g$$
$$= f(W_n, W_p)$$

where C_g → Fan out capacitance. (thin oxide capacitance over gate area)
C_{int} = lumped interconnect capacitance (parasitic capacitance contribution of metal or polysilicon connection between inverters)

The total capacitive load of the inverter

$$\boxed{C_{load} = d_0 + d_n W_n + d_p W_p}$$

where $d_0 = 2D_{drain}(C_{jsw,n} K_{eq,n} + C_{jsw,p} K_{eq,p}) + C_{int} + C_g$

$$d_n = K_{eq,n}(C_{j0,n} D_{drain} + 2C_{jsw,n})$$
$$d_p = K_{eq,p}(C_{j0,p} D_{drain} + 2C_{jsw,p})$$

When $K_{eq,n}, K_{eq,p}$ → Voltage equivalence factor.
$C_{j0,n}, C_{j0,p}$ → Zero bias junction capacitance
$C_{jsw,n}, C_{jsw,p}$ → Side wall junction capacitance.

The transistor aspect ratio, $\boxed{R \equiv \frac{W_p}{W_n}}$

Now the delay

$$\tau_{PHL} = \tau_n \left(\frac{d_0 + (d_n + R d_p) W_n}{W_n} \right)$$

& $$\tau_{PLH} = \tau_p \left(\frac{d_0 + \left(\frac{d_n}{R} + d_p\right) W_p}{W_p} \right)$$

Where $$\tau_n = \frac{L_n}{\mu_n C_{ox}(V_{DD} - V_{T,n})} \times \left[\frac{2V_{T,n}}{V_{DD} - V_{T,n}} + \ln\left(\frac{4(V_{DD} - V_{T,n})}{V_{DD}} - 1 \right) \right]$$

$$\tau_p = \frac{L_p}{\mu_p C_{ox}(V_{DD} - |V_{T,p}|)} \times \left[\frac{2|V_{T,p}|}{V_{DD} - |V_{T,p}|} + \ln\left(\frac{4(V_{DD} - |V_{T,p}|)}{V_{DD}} - 1 \right) \right]$$

→ An inherent limitation to switching speed in CMOS inverters due to drain parasitic capacitance.

→ Increasing W_n & W_p to reduce the propagation delay times will have a diminishing influence upon delay beyond certain values & the delay values will asymptotically approach a limit value for large W_n & W_p

So $\tau_{PHL}^{limit} = \Gamma_n (d_n + R d_p)$

$\tau_{PLH}^{limit} = \Gamma_p \left(\frac{d_n}{R} + d_p \right)$

→ The propagation delay time of a CMOS inverter cannot be reduce beyond these limit value.

→ This limit is independent of Cint & Cg.

***

Estimation of Interconnect Parasitics:

* The switching speed of a logic gate is determined by, taking assumption that loads are mainly capacitive & Lumped.

→ The conventional delay estimation approaches seek to classify 3 main components of the output load, all of which are assumed to be purely capacitive as

1) Internal parasitic capacitances of the transistors
2) Interconnect (Line) Capacitances
3) Input capacitances of the fan-out gates.

→ The capacitive/inductive coupling & the signal interference bet^n neighbouring lines should also be taken into consideration for accurate estimation of delay.

→ The gate delay due to capacitive load components dominates the line delay.

→ For sub-micron technologies, the interconnect delay starts to dominate the gate delay. In order to deal with the implications & to optimize a system for speed, chip designer must have reliable & efficient means for

i) estimating the ~~interconnect~~ & interconnect parasitics in a large chip
ii) Simulating the transient effects.

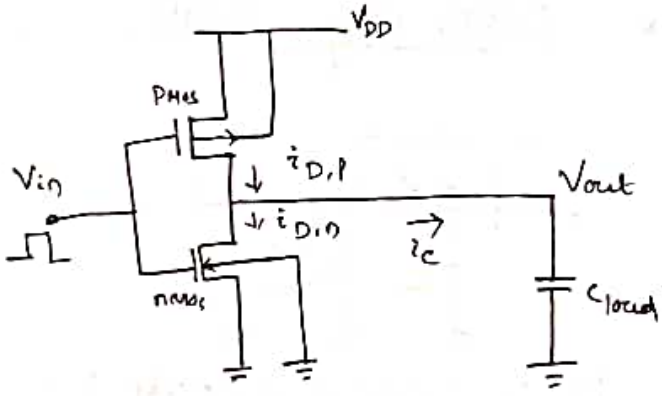
Switching Power Dissipation of CMOS Inverter:

→ The static power dissipation of the CMOS inverter is quite negligible.

→ But during switching event where output load capacitance is alternatingly charged up & changed down, for which the CMOS inverter inevitably dissipates power.

→ Consider a simple CMOS inverter ckt which input voltage is an ideal step waveform with negligible rise & fall time.

→ When the input voltage switches from low to high, the pmos transistor turns off & nmos transistor start conducting.



(Cmos inverter used in the dynamic power-dissipation analysis)

→ During this phase, the output load capacitance C_{load} is being discharged through nmos transistor.

→ Thus the capacitor current equals to the instantaneous drain current of the nmos transistor.

→ Similarly when the Vin switches from high to low, the nmos transistor is turned off & pmos transistor starts conducting. So C_{load} is being charging. Therefore, the capacitor current i_c equals to the instantaneous drain current of pmos transistor.

The average power dissipated

$$P_{avg} = \frac{1}{T}\int_0^T V(t)\cdot i(t)\,dt$$

Since during switching, the nmos & pmos transistor in a CMOS inverter conduct current for one-half period each, so

$$P_{avg} = \frac{1}{T}\left[\int_0^{T/2} V_{out}\left(-C_{load}\frac{dV_{out}}{dt}\right)dt + \int_{T/2}^{T}(V_{DD}-V_{out})\left(C_{load}\frac{dV_{out}}{dt}\right)dt\right]$$

$$= \frac{1}{T}\left[\left(-C_{load}\frac{V_{out}^2}{2}\right)\Big|_0^{T/2} + \left(V_{DD}\cdot V_{out}\cdot C_{load} - \frac{1}{2}C_{load}V_{out}^2\right)\Big|_{T/2}^{T}\right)$$

$$\Rightarrow \boxed{P_{avg} = \frac{1}{T}C_{load}V_{DD}^2}$$

$$\Rightarrow \boxed{P_{avg} = C_{load}V_{DD}^2\cdot f} \quad \text{where } T = \frac{1}{f} \text{ or } f = \frac{1}{T}$$

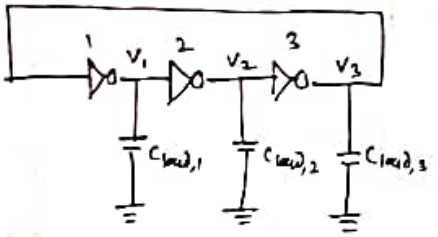
→ The average power dissipation is independent of all transistor characteristics & transistor sizes.

PDP (Power-delay Product): It is the average energy required for a gate to switch its o/p voltage from low to high & high to low.

$$\boxed{PDP = C_{load}\cdot V_{DD}^2}$$

→ (Refer Book)

CMOS Ring Oscillator :



(Fig:1, 3 stage ring osc ckt consisting of identical inverter)

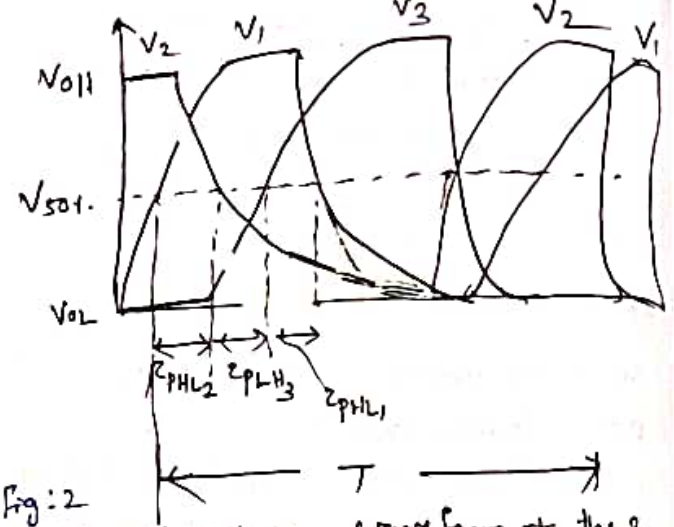


Fig:2 Typical voltage waveform of the 3 inverters,

→ Consider the cascade connection of 3 identical CMOS inverter where o/p node of 3rd inverter is connected to the i/p node of 1st inverter.

→ The 3 inverters form a voltage feedback loop.

→ The ckt doesn't have a stable operating point. The only dc point at which the input-output of all inverters are equal to V_{th} are unstable & any disturbance in node voltages would make the ckt drift away from the dc operating point.

→ A closed loop cascade connection of any odd number of inverter would display astable behavior.

→ Such ckt will oscillate once any of the inverter input output voltages deviates from unstable operating point (V_{th}). Therefore the ckt is called Ring Oscillator.

→ Fig 2 shows typical o/p voltage waveforms of 3 inverters during oscillation. As the o/p voltage V_1 of the 1st inverter rises from V_{OL} to V_{OH} it triggers the 2nd inverter output V_2 to fall from V_{OH} to V_{OL}.

→ The difference bet^n $V_{50\%}$ Crossing times of V_1 & V_2 is the signal propagation delay τ_{pHL_2} of 2nd inverter.

→ As the o/p voltage of V_2 of the 2nd inverter falls, it triggers the o/p voltage V_3 of the 3rd inverter to rise from V_{OL} to V_{OH}

→ Again the difference bet^n $V_{50\%}$ crossing times of V_2 & V_3 is called τ_{pLH_3}

→ In this three ckt the oscillation period can be expressed as sum of the six propagation delay time.

→ Since three inverters are assumed to be identical

$$C_{load_1} = C_{load_2} = C_{load_3}$$

$$T = \tau_{pHL_1} + \tau_{pLH_1} + \tau_{pHL_2} + \tau_{pLH_2} + \tau_{pHL_3} + \tau_{pLH_3}$$

$$= 2\tau_p + 2\tau_p + 2\tau_p = 3 \times 2.\tau_p = 6\tau_p$$

$$f = \frac{1}{T} = \frac{1}{2 \times 3 \tau_p}$$ where f = frequency of oscillation

Generally for n no of cascade inverter

$$f = \frac{1}{2.n.\tau_p}$$

Determination of pull-up to pull down Ratio's :

i) NMOS Inverter driven by another nmos Inverter :

(nmos inverter driven by another inverter)

→ An inverter is driven from the output of another similar inverter

→ consider the depletion mode transistor for which $V_{gs} = 0$ under all conditions . The requirement

$$V_{in} = V_{out} = V_{inv}$$

For equal margins around the inverter threshold , we set $V_{in} = 0.5 V_{DD}$

At this point both transistor are in saturation &

$$I_{ds} = k \frac{W}{L} \frac{(V_{GS} - V_{T_0})^2}{2}$$

on the depletion mode

$$I_{ds} = k \cdot \frac{W_{p.u}}{L_{p.u}} \frac{(-V_{td})^2}{2} \quad \text{since } V_{GS} = 0$$

& in the enhancement mode

$$I_{ds} = k \cdot \frac{W_{p.d}}{L_{p.d}} \frac{(V_{in} - V_{to})^2}{2} \quad \text{Since } V_{GS} = V_{in}$$

Equating (since currents are the same) we have

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

where $W_{p.d}$, $L_{p.d}$, $W_{p.u}$ & $L_{p.u}$ are the widths & Lengths of the pull-down & pull-up transistor respectively .

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

We have $$\frac{1}{Z_{p.d}} (V_{in} - V_{to})^2 = \frac{1}{Z_{p.u}} (-V_{td})^2$$

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

By substituting typical values

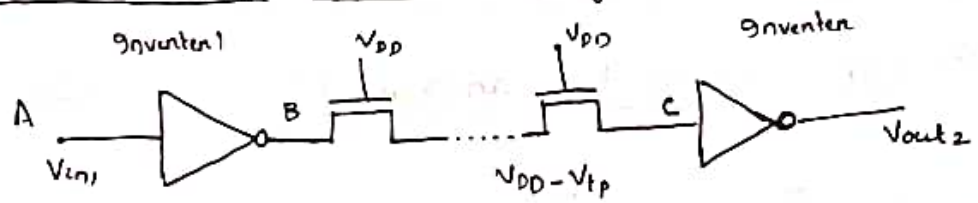
$V_{to} = 0.2 V_{DD}$; $V_{td} = -0.6 V_{DD}$, $V_{in} = 0.5 V_{DD}$ (for equal margins)

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

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

Thus $$\boxed{\frac{Z_{p.u}}{Z_{p.d}} = \frac{4}{1}}$$

ii) NMOS Inverter driven through one or more pass transistor:



(Pull-up to pull-down ratios for inverting logic coupled by pass transistor)

→ Here the input to inverter 2 comes from the output of inverter 1 but passes through one more nMOS transistors used as switches in series (called pass transistors).

→ The connection of pass transistors in series will degrade the logic 1 level into inverter 2 so that the output will not be a proper logic 0 level.

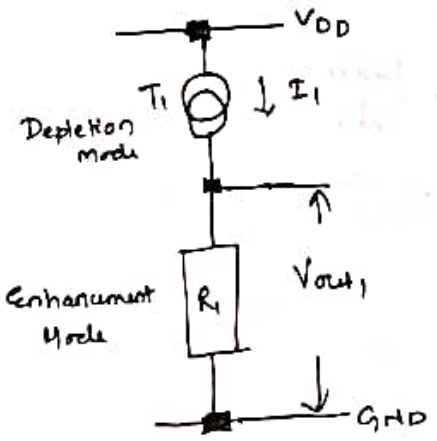
→ The critical condition is when point A is at 0 volts & B is thus at V_{DD}, but the voltage into inverter 2 at point C is now reduced from V_{DD} by the threshold voltage of the series pass transistors.

→ With all pass transistor gates connected to V_{DD}, there is a loss of V_{tp}, since no static current flows through them & there can be no voltage drop in the channels.

→ Therefore input voltage to inverter 2 is

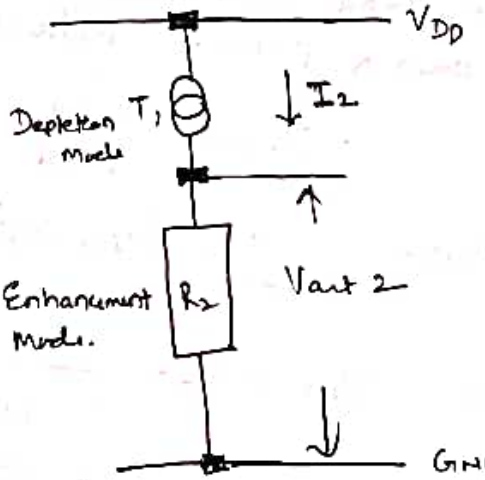
$$V_{in2} = V_{DD} - V_{top}$$

where V_{top} = threshold voltage for a pass transistor.



(Inverter 1 with input = VDD)

Fig (1)



(Inverter 2 with input = $V_{DD}-V_{tp}$

(Equivalent Ckt of Inverter 1 & 2) Fig (2)

Consider inverter 1 with input = V_{DD}. If the input is at V_{DD}, then the p-d transistor T_2 is conducting but with a low voltage across it. Therefore, it is in its resistive region represented by R_1 in fig (1) & the p.u transistor T_1 is in saturation & is represented as a current source.

For the p.d transistor

$$I_{ds} = K \frac{W_{p.d1}}{L_{p.d1}} \left[(V_{DD} - V_{t_0}) V_{DS1} - \frac{V_{DS1}^2}{2} \right]$$

Therefore $$R_1 = \frac{V_{DS1}}{I_{DS}} = \frac{1}{K} \frac{L_{p.d1}}{W_{p.d1}} \left[\frac{1}{V_{DD} - V_{t_0} - \frac{V_{DS1}}{2}} \right]$$

Note that V_{DS1} is small & $\frac{V_{DS1}}{2}$ may be ignored.

Thus

$$R_1 \doteq \frac{1}{K} Z_{p.d1} \left(\frac{1}{V_{DD} - V_{t_0}} \right)$$

For depletion mode p.u in saturation with $V_{GS} = 0$

$$I_1 = I_{DS} = K \frac{W_{PU1}}{L_{p.u1}} \frac{(-V_{tod})^2}{2}$$

The product

$$I_1 R_1 = V_{out1}$$

Thus $$V_{out1} = I_1 R_1 = \frac{Z_{p.d1}}{Z_{pU1}} \left(\frac{1}{V_{DD} - V_{t0}} \right) \frac{(V_{tod})^2}{2}$$

Consider inverter 2 (Fig(2)) when input = $V_{DD} - V_{top}$. As for inverter 1

$$R_2 \doteq \frac{1}{K} Z_{pd2} \frac{1}{((V_{DD} - V_{top}) - V_{t_0})}$$

$$I_2 = K \frac{1}{Z_{p.u2}} \frac{(-V_{tod})^2}{2}$$

Hence

$$V_{out2} = I_2 R_2 = \frac{Z_{p.d2}}{Z_{p.u2}} \left(\frac{1}{V_{DD} - V_{top} - V_t} \right) \frac{(-V_{td})^2}{2}$$

If the inverter 2 is to have the same voltage under these conditions then $V_{out1} = V_{out2}$.

$$I_1 R_1 = I_2 R_2$$

Therefore

$$\frac{Z_{p.u2}}{Z_{p.d2}} = \frac{Z_{p.u1}}{Z_{p.d1}} \quad \frac{(V_{DD} - V_{to})}{(V_{DD} - V_{top} - V_t)}$$

Taking typical values:

$$V_{to} = 0.2\, V_{DD}$$

$$V_{top} = 0.3\, V_{DD}$$

$$\frac{Z_{p.u2}}{Z_{p.d2}} = \frac{Z_{p.u1}}{Z_{p.d1}} \quad \frac{0.8}{0.5}$$

Therefore

$$\frac{Z_{p.u2}}{Z_{p.d.2}} \doteq 2\frac{Z_{p.u.1}}{Z_{p.d1}} = \frac{8}{1}$$

Layout Design Rules:

→ The physical mask layout of any ckt to be manufactured using a particular process must conform to a set of geometric constraints, or rules which are generally called "Layout design Rules".

→ These rules specify the minimum allowable line widths for the physical objects on chip such as metal, polysilicon interconnects, min^m feature dimensions & min^m allowable separation bet^n two such features.

→ If a metal line width is made too small, it is possible for the line to break during fabrication process or afterwards resulting in an open ckt.

→ If two lines are placed too close to each other then they may form an unwanted short ckt by merging during or after fabricat^n process.

→ The main objective is to achieve a high overall yield & reliability while using smallest possible area.

The design rules are usually described in two ways:

i) Micron Rule, in which the layout constraints such as min^m feature size & min^m allowable feature separation are in terms of absolute dimension in micrometer.

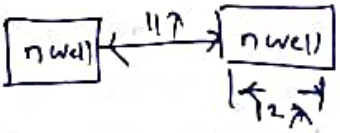
ii) Lambda Rule: which specify the layout constraints in terms of a single parameter (λ) & thus allow linear, proportional scaling of all geometrical constraints.

Ex: The Silicon CMOS (SCMOS) design Rule.

N-well

Min^m well size → 12λ

well to well spacing → 11λ



Polysilicon -1

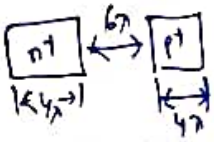
min^m polysilicon width → 2λ

Polysilicon Gate on diffusion → 2λ

n diffusion & p Diffusion:

Min^m n+ & p+ diffusion width → 4λ

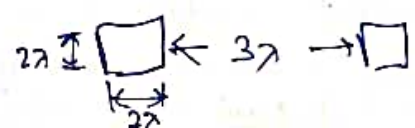
" spacing bet^n p+ & n+ diffusion → 6λ



Contacts:

Min^m contact size - 2λ x 2λ

Spacing bet^n 2 contacts → 3λ



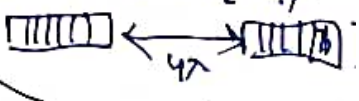
Active area Rule:

Min^m active area width =

Metal

Min^m metal width → 3λ

Bet^n 2 metals → 4λ



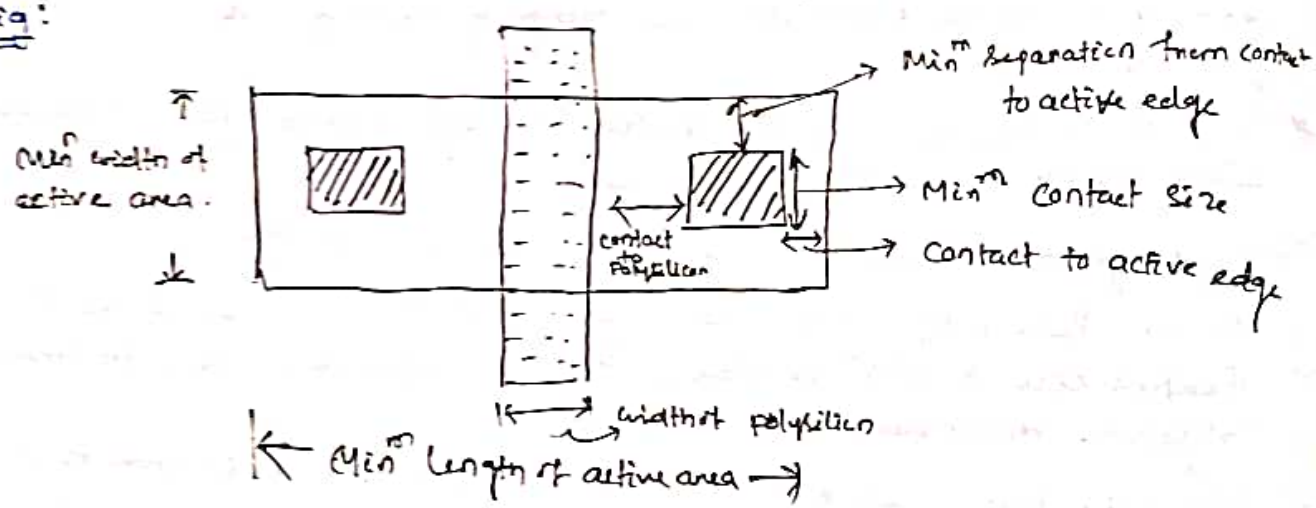
Full Custom Mask Layout Design:

CMOS Inverter Layout design:

1) First we need to create individual transistors according to the design Rules. Assume that we attempt to design the inverter with min^m size transistors.

2) Then the width of the active area is determined by min^m diffusion contact size & min^m separation from diffusion contact both active area edges. The width of the polysilicon line over the active area is typically taken as the min^m poly width. Then the min^m overall length of the active area = (Min^m polysilicon width) + 2(min^m Poly-to-contact spacing) + 2(min^m contact size) + 2(min^m spacing from contact to active area edge).

Fig:



3) The pmos transistor must be placed in an nwell region & the min^m size of the n-well is decided by the PMOS active area & min^m n-well overlap over nt.

4) The distance bet^n nmos & p mos transistor is determined by the min^m separation bet^n the nt active area & the nwell.

5) The polysilicon gates of nmos & PMOS transistors are usually aligned so that gate connections can be made with a single polysilicon size of least possible length.

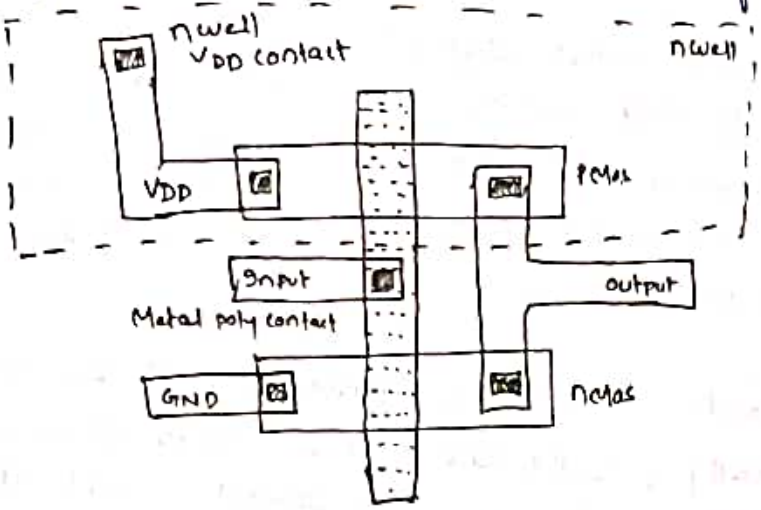
6) The reason for avoiding long polysilicon connections is, that the large parasitic capacitance & parasitic resistance of polysilicon may result in significant Rc delays.

7) Even local signal connections are preferable made with metal lines as small as possible.

8) The final step is the local interconnections in metal for the output node, VDD & GND contacts.

→ The dimensions of metal lines in a mask layout are usually dictated by min^m metal width & min^m metal separation.

This fig: shows the complete mask layout of the CMOS inverter.



(Mask layout)

Stick Diagram:

→ The design process are aided by simple concepts such as stick & symbolic diagram, but the key element is the set of design rules.

→ Design rules are comm link bet designer specific requirements & the fabricator who materialised them.

→ Stick diagram may be used to convey layer information through the use of colour code as well as monochrome hatches.

→ Mask layout information which is also colour coded may also be hatched for monochrome encoding.

Steps in drawing stick diagram:

1) Draw the metal VDD & ground rails in parallel, drawing enough space bet them.

2) Thinox path may be drawn bet the rails for inverters & inverter based logic & also make appropriate contacts.

3) Remember that polysilicon cross thinox wherever transistors are required

4) For depletion mode transistors 9mplants are to be done.

5) Each transistor right length, width (L:W) ratios are important in nmos & nmos lsi.

6) Signal path may also be switched by pass transistor & long signal path may often required metal buses.

7) The stick diagram may well represent only a small section ckt which will be replicated many times.

6) Therefore power rails & buses are run in parallel in metal

x The control signals are propagated at right angle on polysilicon. At this stage of design leaf cell boundaries are conveniently shown on the stick diagram & these are placed so that replicated cell may be directly interconnected by direct abutment on side by side & / or top to bottom basis.

Leaf cell: The layout of I.C i.e floor plan of Ic can be represented hierachically, cells are built from other cells, except for those cells that are at lowest level of hierarchy. These lowest level of the cell are called "leaf cells"

CIF: Caltech Intermediate format.

Table:

Colour	Stick Encoding	layers	Mask layout Encoding	CIF layer
Green		n-diffusion (or active Thing)		ND
Red		Polysilicon		NP
Blue		Metal 1		NM
Black		contact cut		NC
Grey	Not applicable	overglass		NG
Mmas only Yellow		gmplant		NI
Mmas only Brown		Buried contact		NB

Demarcation line

n-type (red over Green) transistor (Green transistor)

P type (Red over yellow) transistor (yellow transistor)

Demarkation line

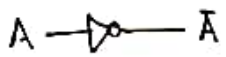
(n type & P type transistor in CMOS)

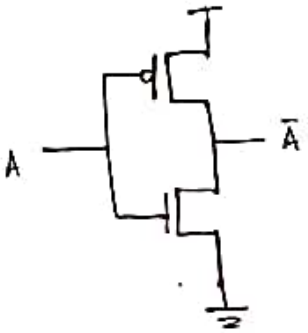
Examples:

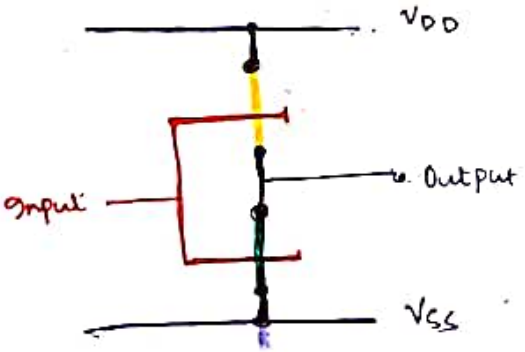
Notes:

PMOS →	Product	→	Parallel
nMOS →	Product	→	Series
PMOS →	Add	→	Series
nMOS →	Add	→	Parallel

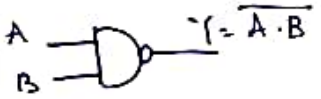
NOT

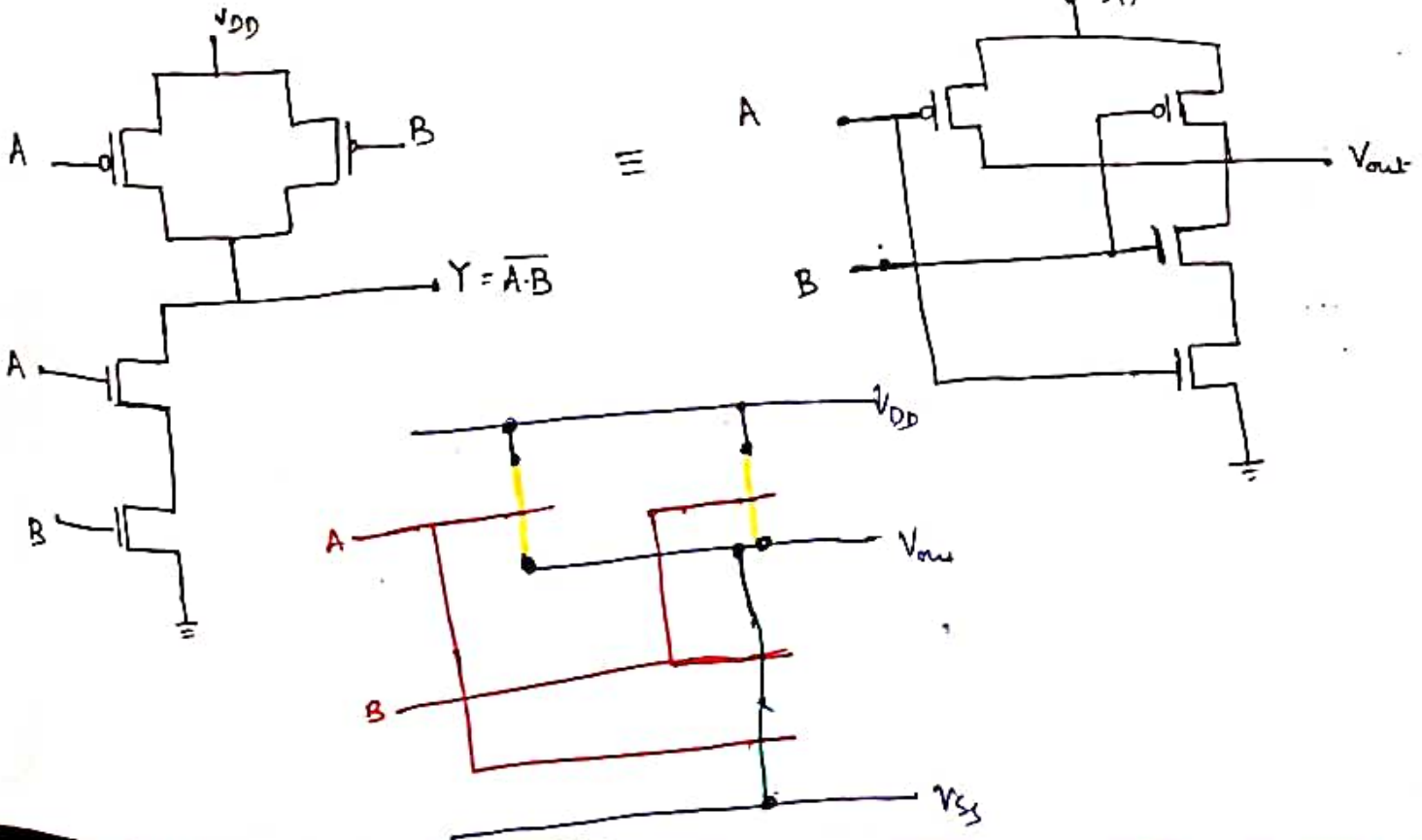




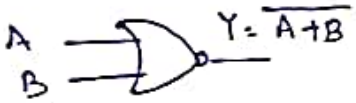


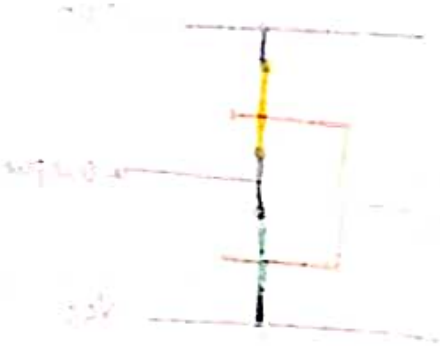
NAND:





NOR :





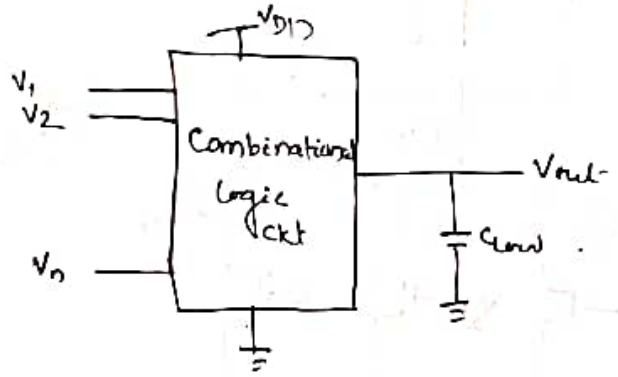
: Combinational Mos Logic Circuits:

Introduction:

→ combinational logic ckts or gates which perform Boolean operations on multiple i/p variables & determine the outputs as Boolean fn's of the inputs, are the basic building blocks of all digital system.

→ All the input variables are represented by node voltage, referenced to ground potentials.

→ Using positive logic convention, the boolean value of '1' can be represented by a high voltage of V_{DD} & '0' can be represented by a low voltage of 0.

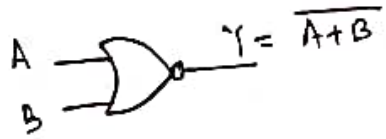


formula

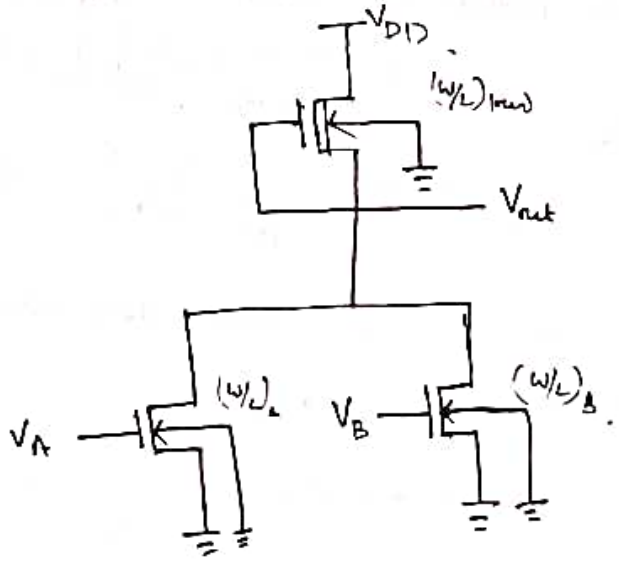
* PPP *

Mos Logic ckt with depletion nmos Loads:

i) Two-input NOR Gate:



V_A	V_B	V_{out}
low	low	high
low	High	low
High	low	low
High	High	lower



→ The Boolean OR operation is performed by the parallel connection of the two enhancement-type nmos driver transistors.

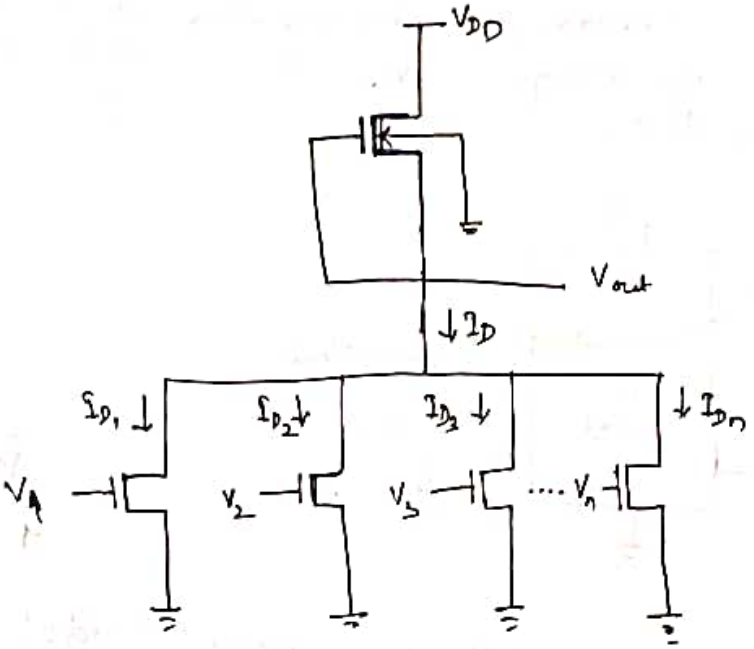
→ If the i/p voltage V_A or V_B is equal to ~~high~~ logic-high level, the corresponding driver transistor turn on & provides a

conducting path bet^n the o/p node & the ground. Hence the voltage become low.

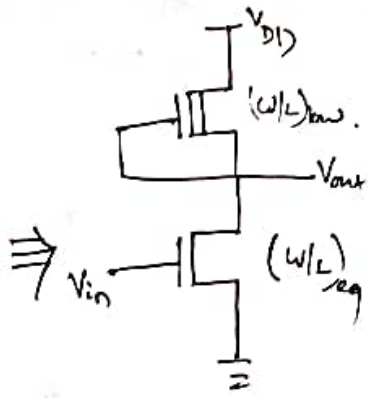
→ Similar result is achieved when both V_A & V_B high.

→ When both V_A & V_B are low, both driver transistor remain cut off. The output node voltage is pulled to a logic high level by the depletion-type nMOS load transistor.

ii) Generalized NOR Structure with Multiple Inputs:



(n input NOR Gate)



(Equivalent ckt)

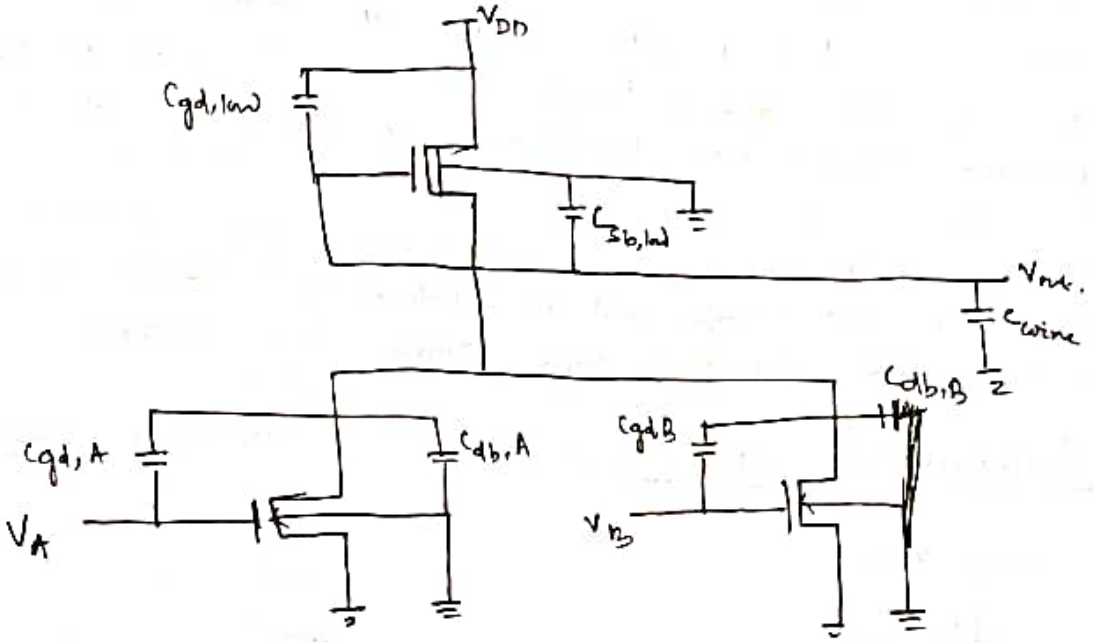
The pull down current expression can be written as

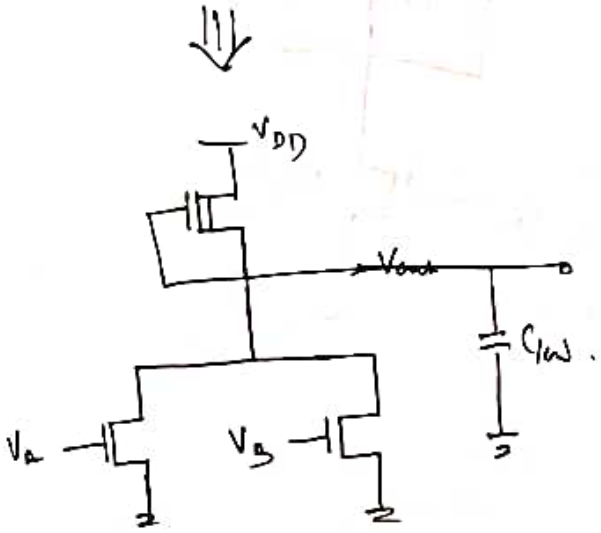
$$I_D = \begin{cases} \frac{\mu_n C_{ox}}{2} \left(\sum_{K(on)} \left(\frac{W}{L}\right)_K \right) \left[2(V_{GS} - V_{T0}) V_{out} - V_{out}^2 \right] & \rightarrow \text{Linear} \\ \frac{\mu_n C_{ox}}{2} \left(\sum_{K(on)} \left(\frac{W}{L}\right)_K \right) \left(V_{GS} - V_{T0} \right)^2 & \rightarrow \text{Saturation.} \end{cases}$$

Thus the multiple-input nor gate can be also be reduced to an equivalent inverter.

$$\left(\frac{W}{L}\right)_{equivalent} = \sum_{K(on)} \left(\frac{W}{L}\right)_K$$

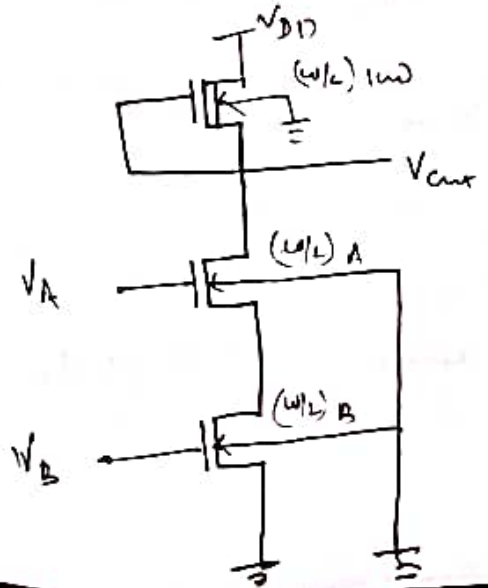
iii) Transient Analysis of NOR Gate:





$$C_{low} = C_{gd,A} + C_{gd,B} + C_{gd,low} + C_{db,A} + C_{db,B} + C_{sb,low} + C_{wire}$$

iv) Two input NAND Gate:



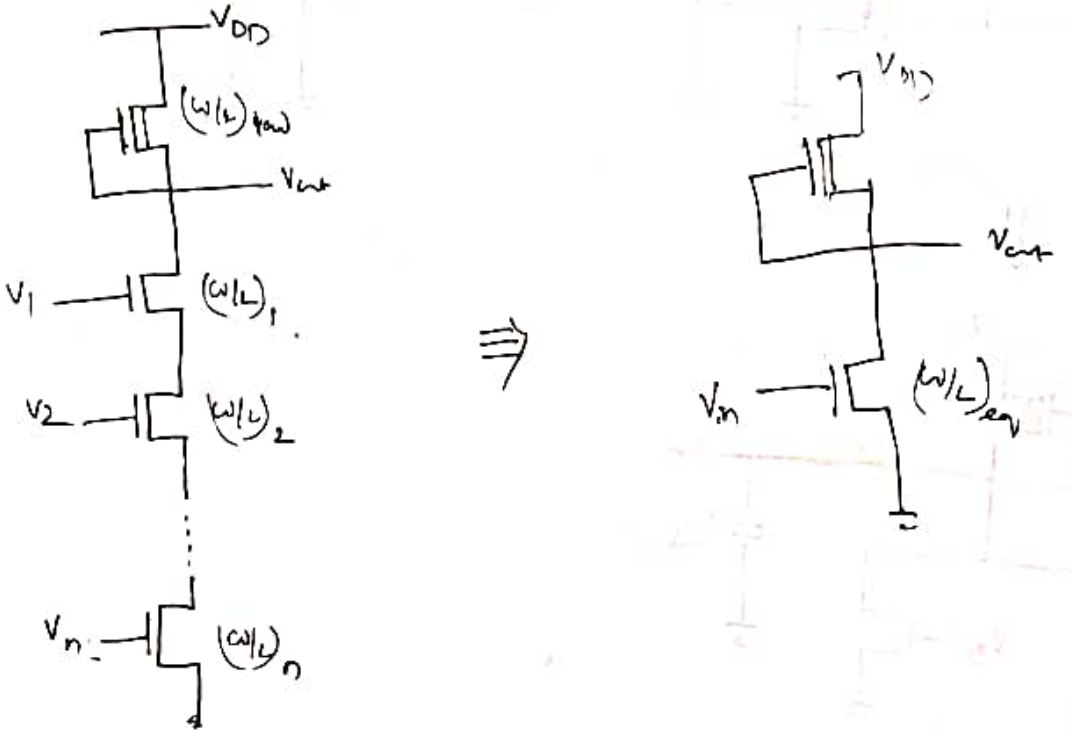
V_A	V_B	V_{out}
low	low	high
low	high	high
high	low	high
high	high	low

→ The boolean AND operation is performed by series connection of the two enhancement type nmos driver transistor.

→ There is a conducting path betⁿ the o/p node & ground only if V_A & V_B are equal to high i.e only if both the series connected drivers are turned on. In this case o/p voltage will be low.

→ Otherwise, either one or both the driver transistors will be off & the o/p voltage will be pulled to a high logic high level by the depletion-type nmos load transistor.

v) Generalized NAND structure with multiple Input:



$$I_D = \frac{\mu_n C_{ox}}{2}\left(\frac{1}{\sum_{K(on)}\frac{1}{(\frac{W}{L})_K}}\right)\cdot\begin{cases}\left[2(V_{in}-V_{T0})V_{out}-V_{out}^2\right] \rightarrow \text{linear}\\ (V_{in}-V_{T0})^2 \rightarrow \text{Saturn}^n\end{cases}$$

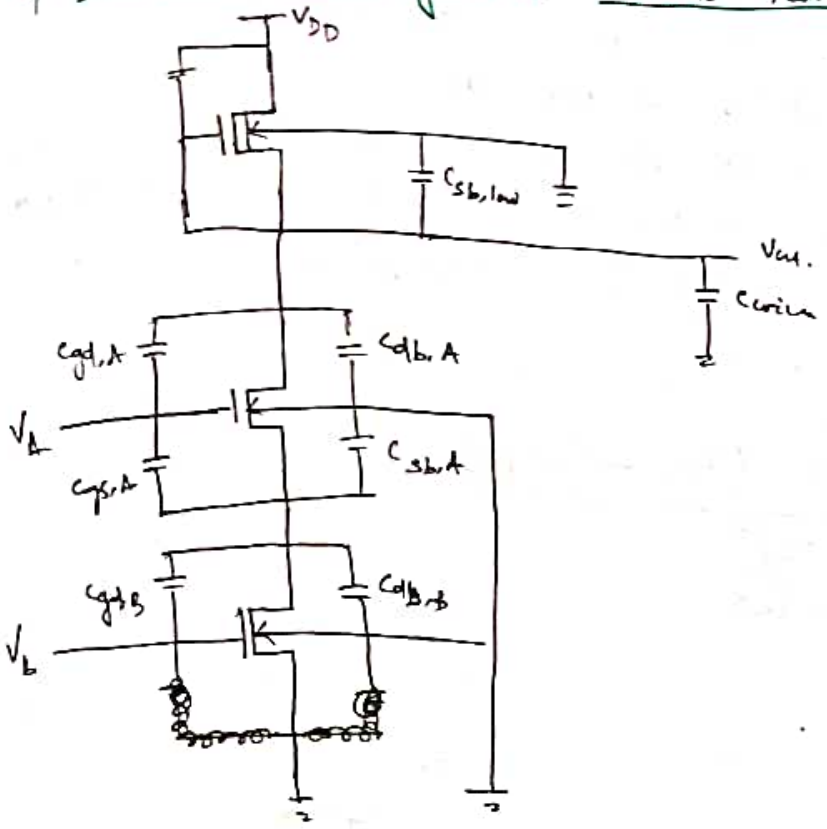
The (W/L) ratio of the equivalent driver transistor is

$$\left(\frac{W}{L}\right)_{equivalent} = \frac{1}{\sum_{K(on)}\frac{1}{(W/L)_K}}$$

If the series connected transistors are identical i.e $(W/L)_1 = (W/L)_2 = \dots = (W/L)_n$ then $(W/L)_{eqv}$ becomes

$$\left(W/L\right)_{equivalent} = \frac{1}{n}\left(\frac{W}{L}\right)$$

vi) Transient Analysis of NAND Gate:



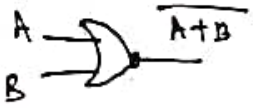
$C_{load} = C_{gd,low} + C_{gd,A} + C_{gd,B} + C_{gs,A} + C_{db,A} + C_{db,B} + C_{sb,A} + C_{sb,low} + C_{wire}$

The lumped output capacitance is

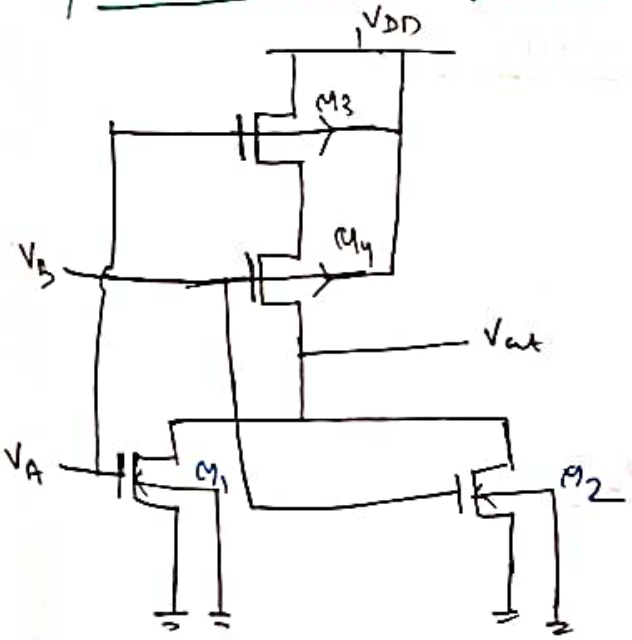
$C_{load} = C_{gd,low} + C_{gd,A} + C_{db,A} + C_{sb,low} + C_{wire}$.

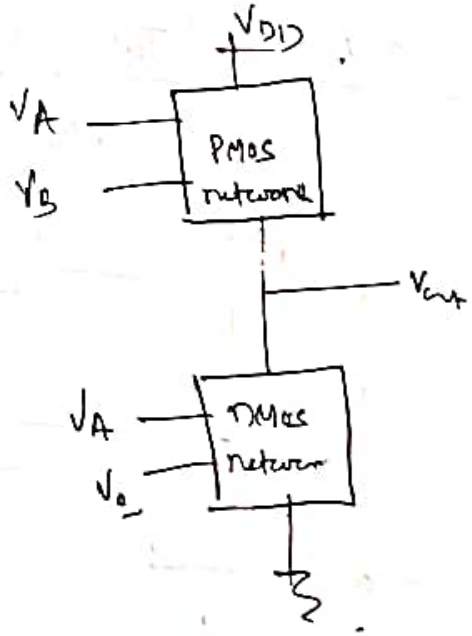
CMOS Logic Ckt:

i) CMOS NOR2 (Two-Input NOR) Gate:



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





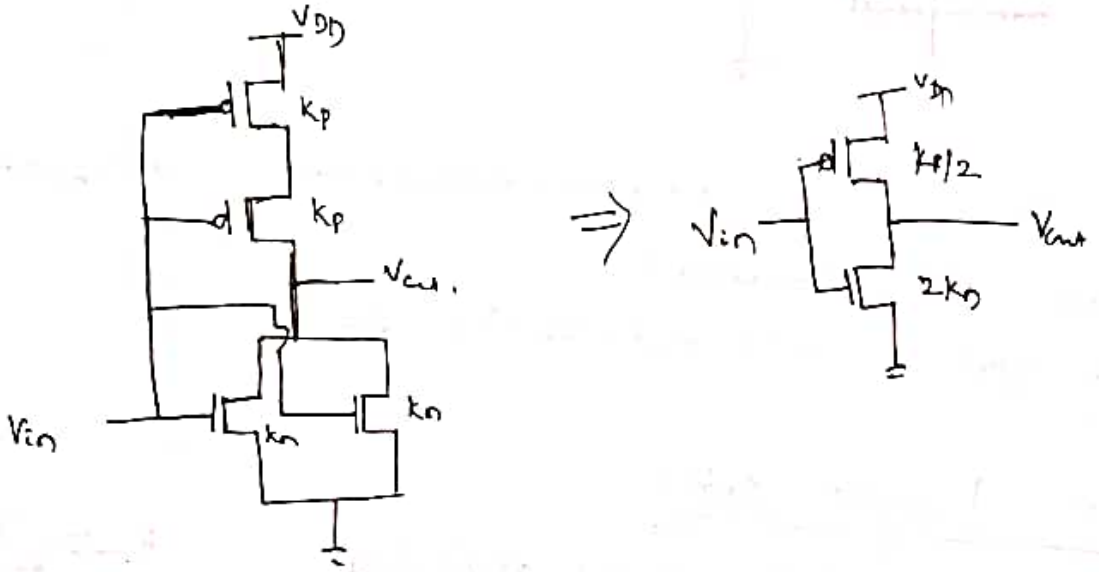
→ When either one or both inputs are high, i.e when the n-net creates a conducting path bet^n the o/p node & the ground, the p-net is cut off. So $V_{out} = 0V$.

→ On the other hand, if both input voltages are low, i.e the n-net is cut off, then the p-net creates a conducting path bet^n the o/p nodes & the supply voltage V_{DD}. So the

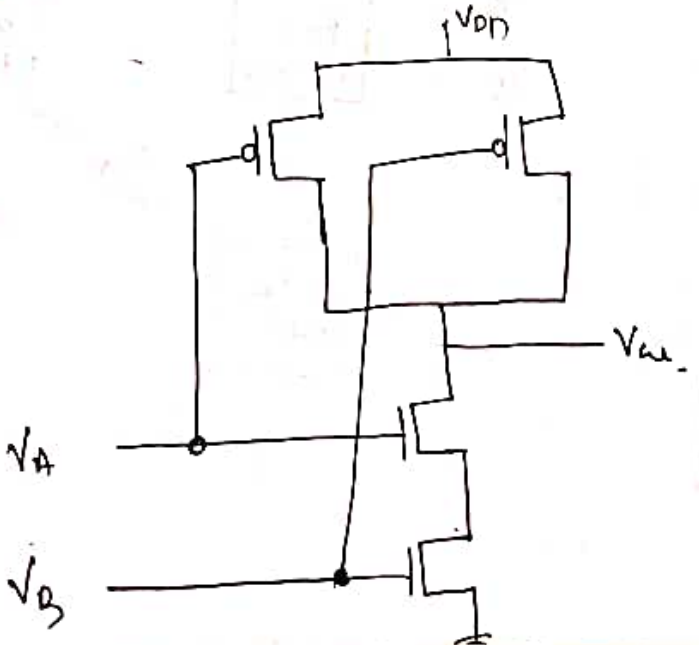
$$V_{out} = V_{DD}.$$

$$V_{th} = \frac{V_{T,n} + \sqrt{\frac{k_p}{4k_n}}\,(V_{DD} - |V_{T,p}|)}{1 + \sqrt{\frac{k_p}{4k_n}}}$$

or



CMOS NAND 2 (Two-Input NAND) Gate:



→ When either V_A or V_B or both V_A & V_B are low then n. net is cut off & at the same time either or both PMOS form a conducting path for V_{out} to V_{DD}. So $V_{out} = V_{DD}$

→ If both V_A & V_B are high then P. net is cut off & n. net forms a conducting path for V_{out} to ground. So $V_{out} = 0V$.

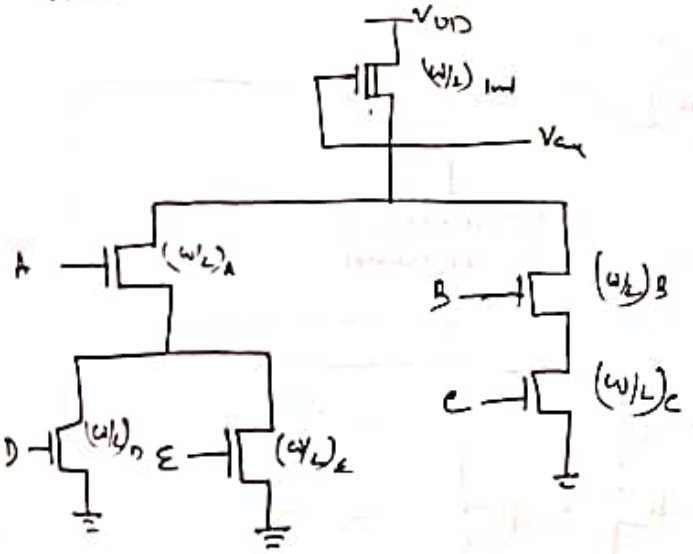
$$V_{th} = \frac{V_{T,n} + 2\sqrt{\frac{K_p}{K_n}}(V_{DD} - |V_{T,p}|)}{1 + 2\sqrt{\frac{K_p}{K_n}}}$$

Complex Logic ckt :

Consider the following Boolean function as an example

$$Z = \overline{A(D+E) + BC}$$

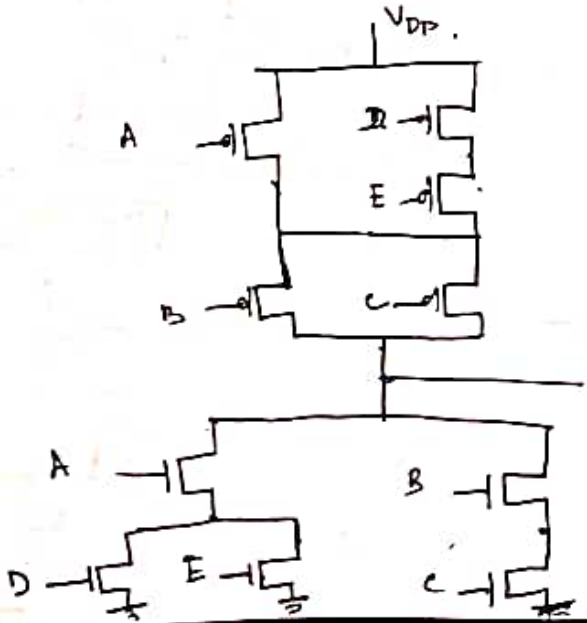
Using nMOS depletion load complex logic gate the ckt is as shown below:

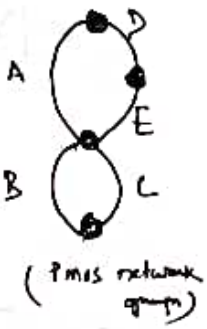


Hence $(W/L)_{equivalent}$

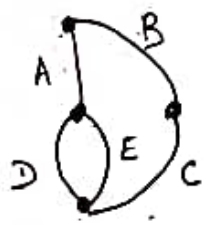
$$\left(\frac{W}{L}\right)_{eqv} = \frac{1}{\frac{1}{(W/L)_A} + \frac{1}{(W/L)_C}} + \frac{1}{\frac{1}{(W/L)_A} + \frac{1}{(W/L)_D} + \left(\frac{W}{L}\right)}$$

Using CMOS Logic:



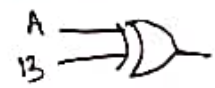


(PMOS network graph)

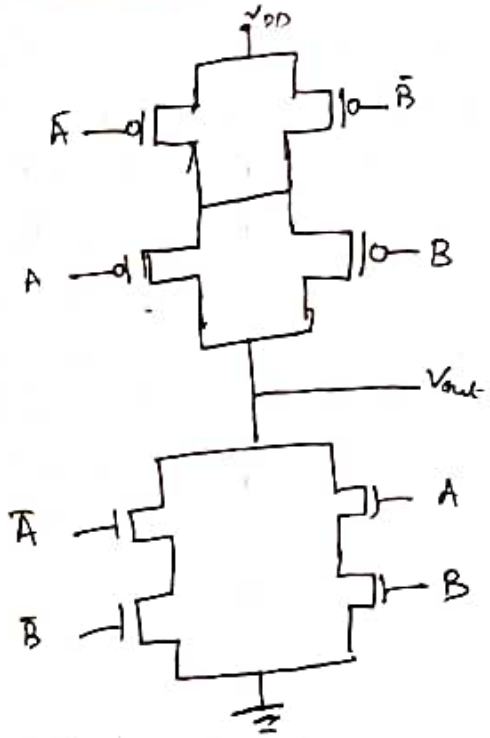


NMOS network

X-OR



$A \oplus B = \bar{A}B + A\bar{B}$

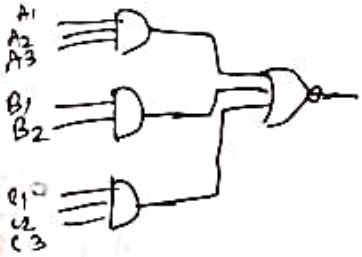


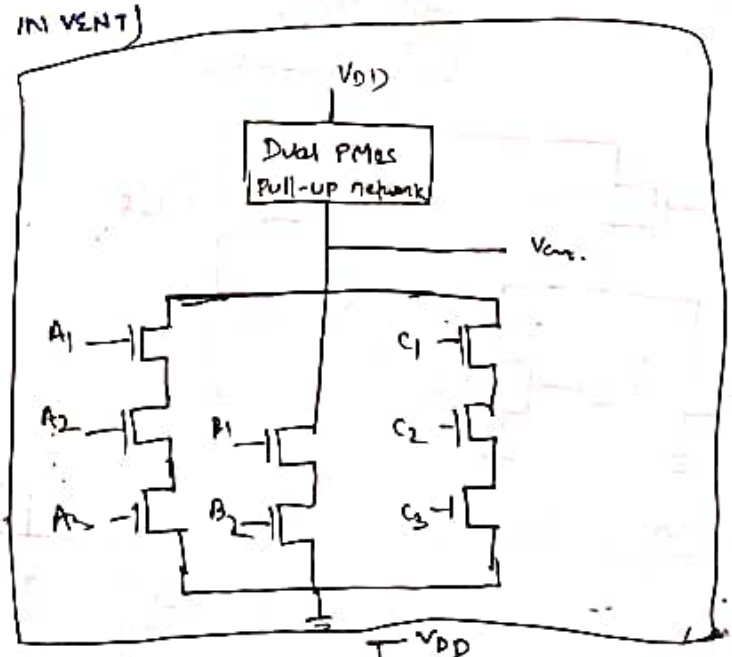
Section AOI & OAI Gates:

→ There are two important ckt catagories as subset of the general complex CMOS gate topology.

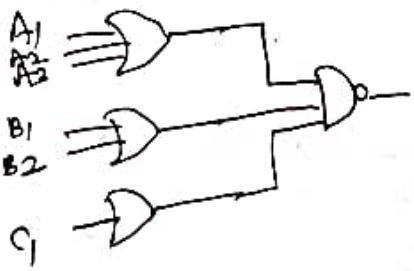
i) AOI (AND - OR - INVERT)
ii) OAI (OR - AND - INVENT)

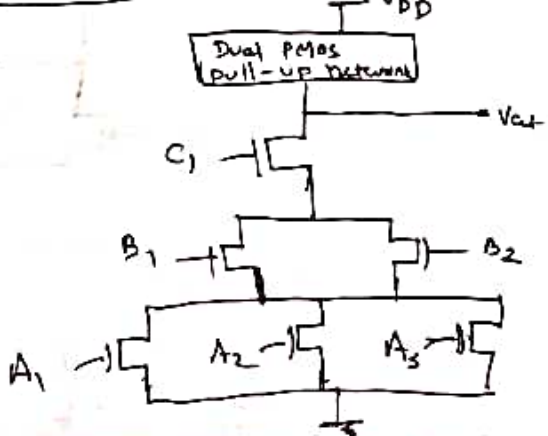
AOI





OAI



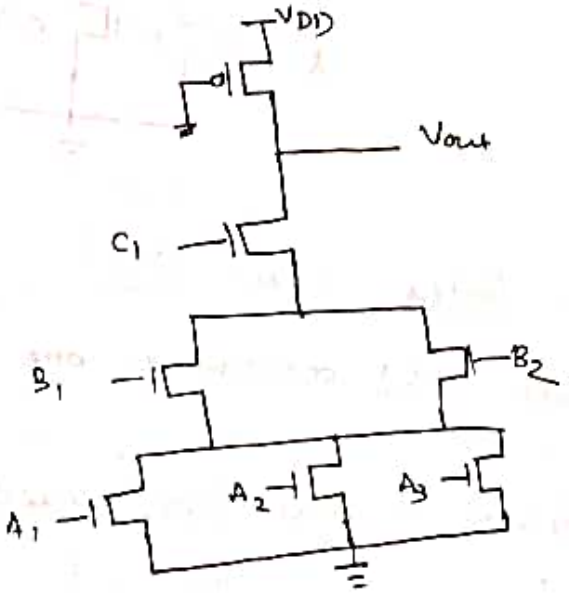


Pseudo- nMOS Gates:

→ The large area requirements of complex CMOS gates present a problem in high density designs. Since two complementary transistors, one nMOS & one pMOS are needed for every input.

→ One possible approach to reduce the no of transistors is to use a single pMOS transistor, with its gate terminal connected to ground, as the load device.

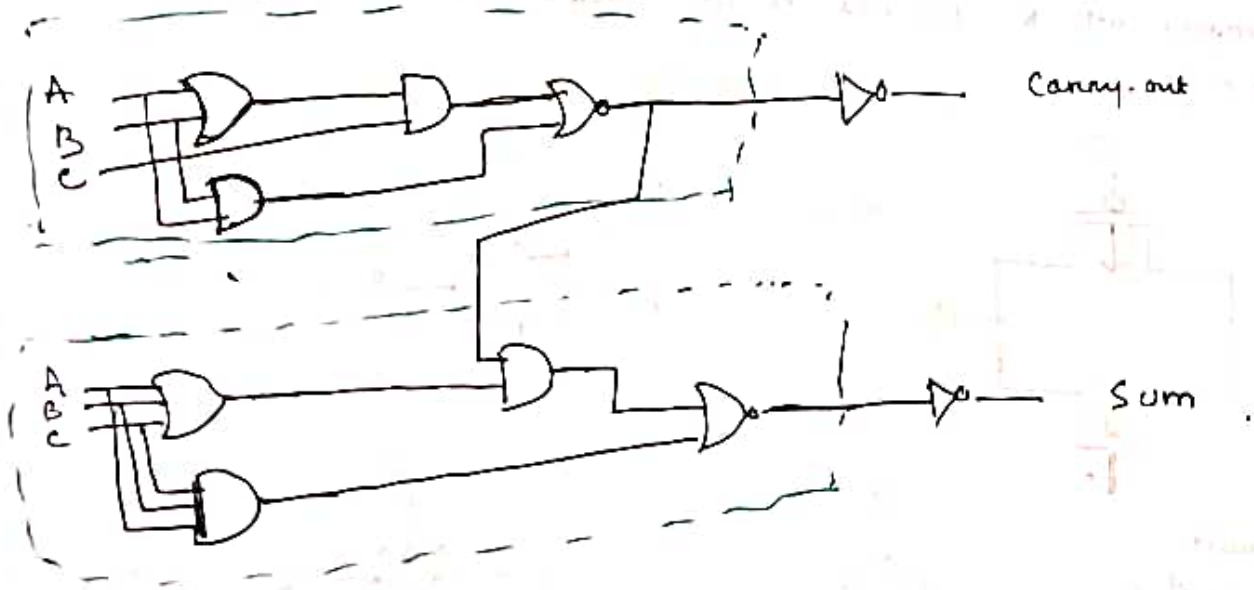
→ The most significant disadvantage of using a pseudo-nMOS gate instead of a full-CMOS gate is the nonzero static power dissipation, since the always-on pMOS load device conducts a steady state current when the o/p voltage is lower than VDD.

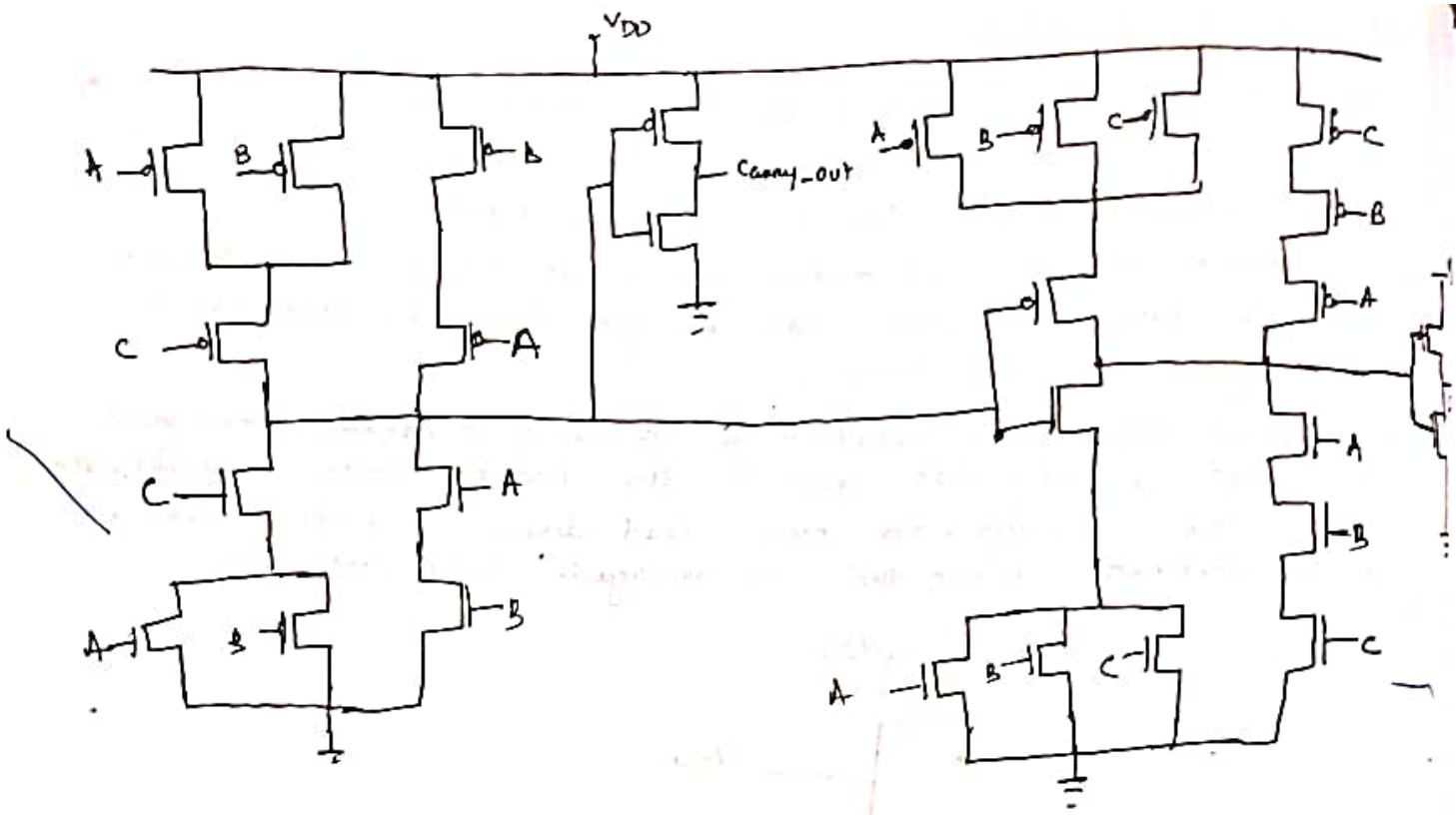


CMOS Full Adder: (one bit full adder)

Sum_out = $A \oplus B \oplus C = ABC + A\bar{B}\bar{C} + \bar{A}B\bar{C} + \bar{A}\bar{C}B$

Carry-out = $AB + BC + CA$.





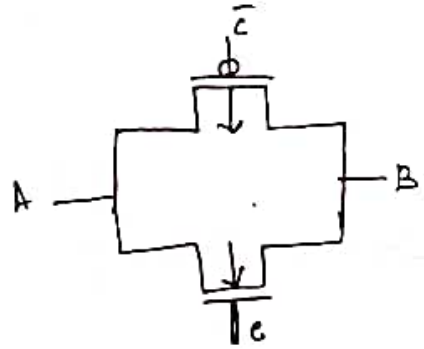
CMOS Transmission Gates (Pass Gates)

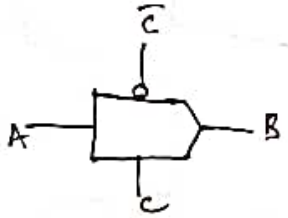
→ CMOS transmission Gate (TG) consists of one nmos & one pmos connected in parallel.

→ The gate voltages applied to these two transistors are complementary to each other.

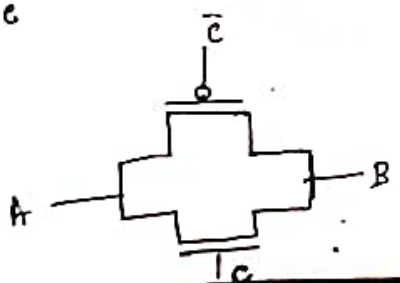
→ As such CMOS TG operates as a bidirectional switch bet^n nodes A & B which is controlled by C.

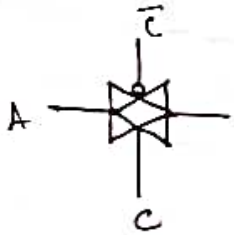
→ If control signal 'C' is high then both transistors are turned ON & provide a low resistance connecting path bet^n A & B.





(Four different Representation of the CMOS TG)



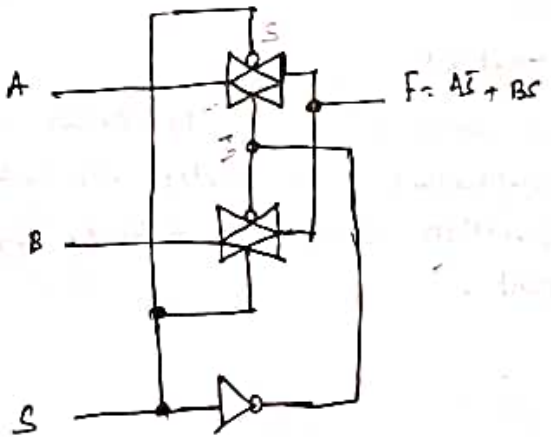


B = A if C = 1

if C = 0, B = 0

→ If control signal 'C' is low then both transistors will be off & the path bet^n A & B will be open circuited. This is called 'high impedance state'.

Two input Mux:

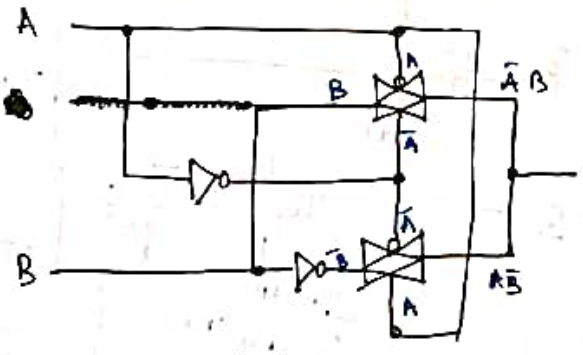


→ If control input S is logic-high, then bottom TG will conduct & the o/p will equal to i/p B.

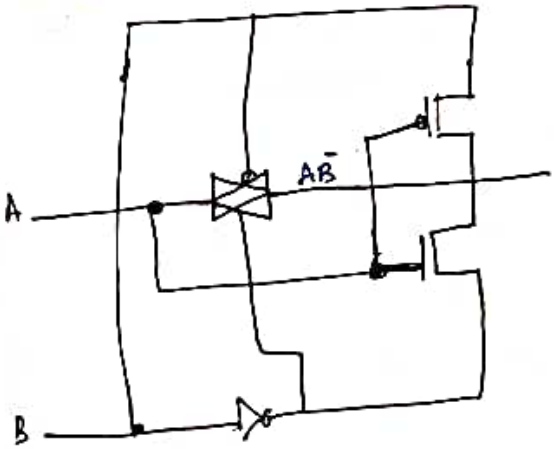
→ If S is eq low, the bottom TG will turn off & Top TG will connect th input A to the o/p node.

(Two i/p Mux using CMOS TG)

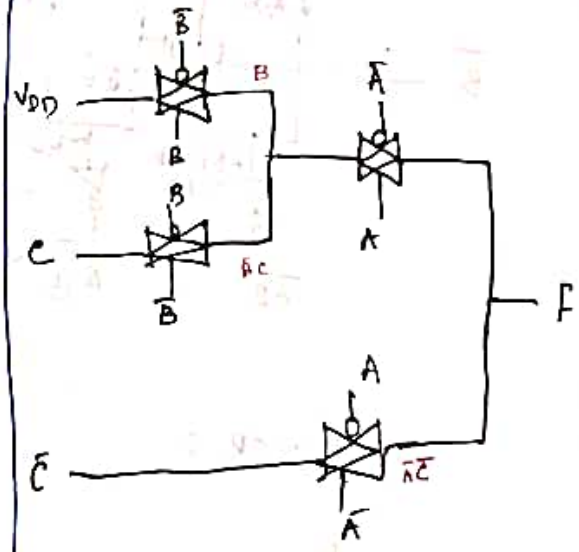
CMOS TG XOR (8 transistor):



CMOS TG XOR (6 transistor):



eg) F = AB + ĀC̄ + AB̄C



Complementary Pass Transistor Logic (CPL):

→ The main purpose of using CPL is to use a purely nmos pass transistor network for all logic operation instead of CMOS Th network.

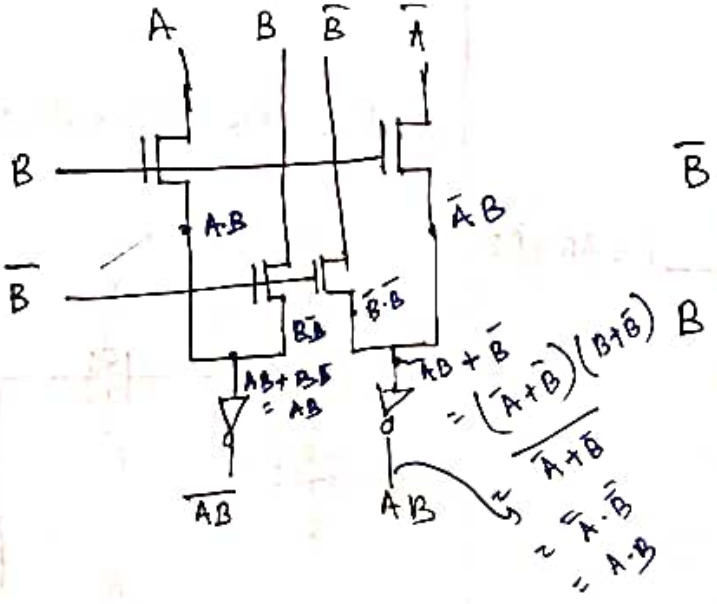
→ All inputs are applied in complementary form i.e every input signal & its inverse must be provided.

→ The ckt also provide complementary output.

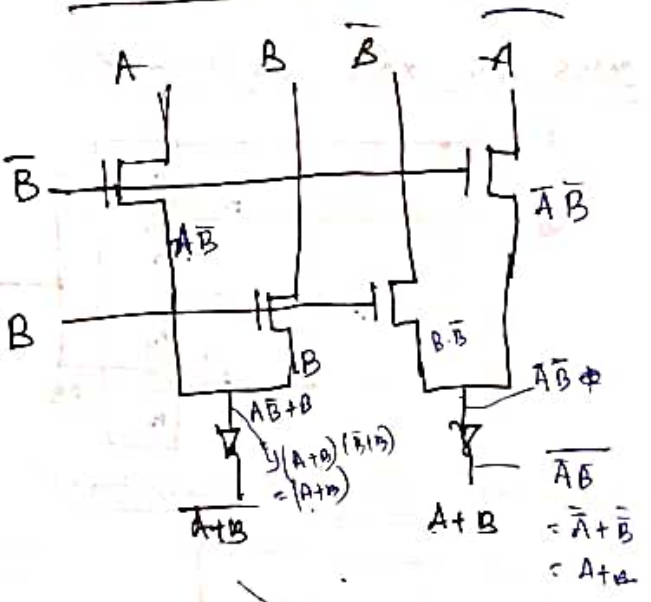
Advantage: (1) The elimination of PMOS transistors from the pass-gate network significantly reduces the parasitic capacitances associated with each node in the ckt, thus the operation speed is typically higher compared to a Full-CMOS counterpart.

2) Reduce overall noise immunity.

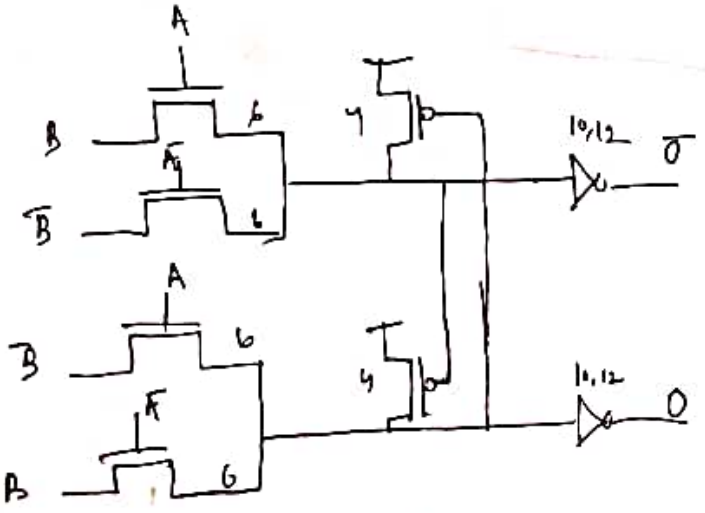
CPL NAND Gate:



CPL NOR Gate:



CPL X-OR:



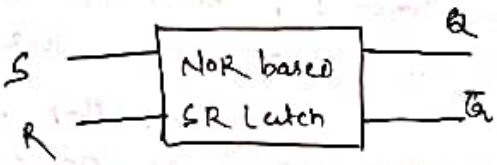
Sequencial MOS Logic Ckt

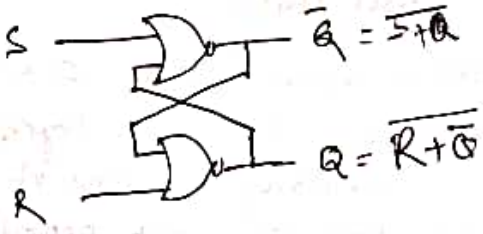
→ In sequential ckt, the o/p is determined by the current i/p as well as previously applied i/p variables.

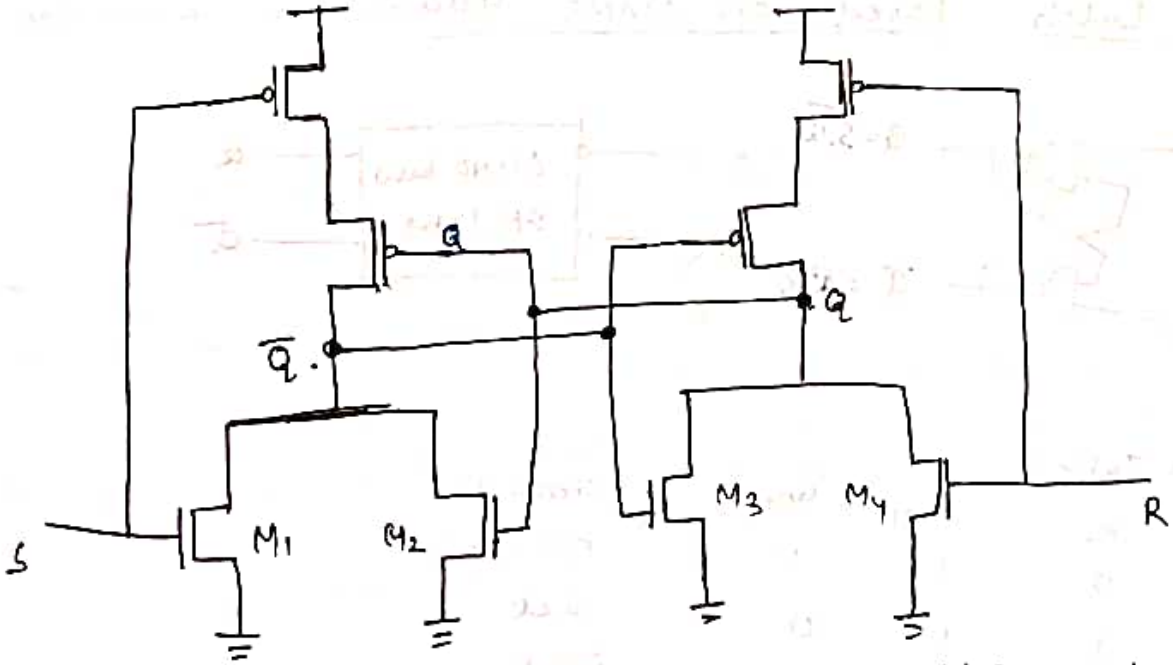
→ It consists of a combinational ckt & a memory block in feedback loop.

SR Latch ckt:

→ The bistable element consisting ~~of~~ of two cross coupled inverters (i.e NAND gate or NOR Gate) that two stable operating modes or states.







(CMOS SR Latch ckt based on NOR2 gate)

Truth Table: (NOR based SR Latch)

S	R	Q_{n+1}	$\overline{Q_{n+1}}$	Operation
0	0	Q_n	$\bar{Q}_n$	hold
1	0	1	0	Set
0	1	0	1	Reset
1	1	0	0	not allowed.

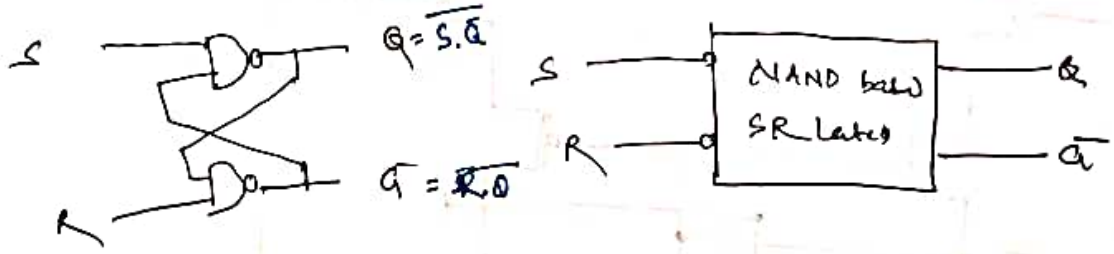
→ When both input signals are equal to logic '0', it will preserve (hold) either one of its two stable operating points (states) as determined by the previous inputs.

→ If the set input (S) is equal to logic '1' & the reset input is equal to logic '0' then output node Q will be forced to logic '1' while the output node $\bar{Q}$ is forced to logic '0'. It means SR latch will set regardless of its previous state.

→ If S is equal to '0' & R equal to '1' then the o/p node 'Q' will be forced to logic '0' & $\bar{Q}$ is forced to logic 1. Thus with this input combination the latch is reset regardless of past states.

→ Finally when both S & R are logic 1, both o/p nodes will be forced to logic 0 which contradicts the complementary nature of Q & $\bar{Q}$. Therefore this input combination is not permitted & considered as not allowed condition.

SR Latch based on NAND gate:

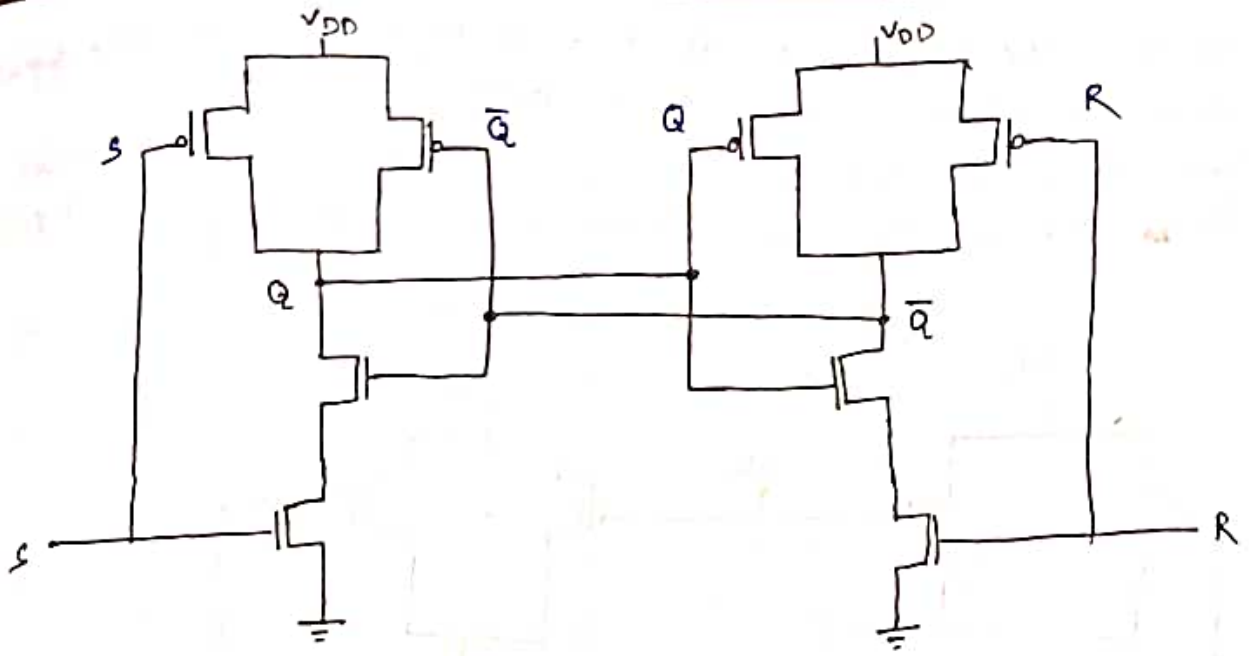


Truth Table:

S	R	Q_{n+1}	$\bar{Q}_{n+1}$	Operation
0	0	1	1	not allowed
0	1	1	0	Set
1	0	0	1	Reset
1	1	Q_n	$\bar{Q}_n$	hold

→ When both S & R are logic at logic 1 the SR latch preserves its last state.

→ If S = 0 & R = 1 then Q is forced to logic '1' while $\bar{Q}$ is forced to logic 0. So latch is Set.



→ If S = 1 & R = 0 then Q will be forced to logic '0' & 'Q̄' is forced to logic '1'. So latch is reset.

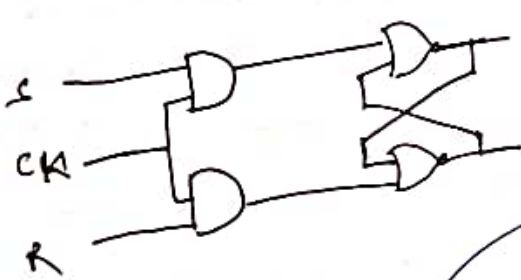
→ When both S & R are 0 then both Q & Q̄ are forced to logic 1 which contradicts the complementary nature of Q & Q̄. So this i/p combination is called 'not allowed' ~~Combination~~ Combination.

Clocked Latch & Flip flop circuits:

Clocked SR Latch: (SR F/F)

→ All the SR latches ~~are~~ ckts are Asynchronous.

→ To provide synchronous operation, the ckt response can be controlled simply by adding a gating clk signal to the ckt, so that the o/p's will respond to the input levels only during the active period of a CLK pulse.



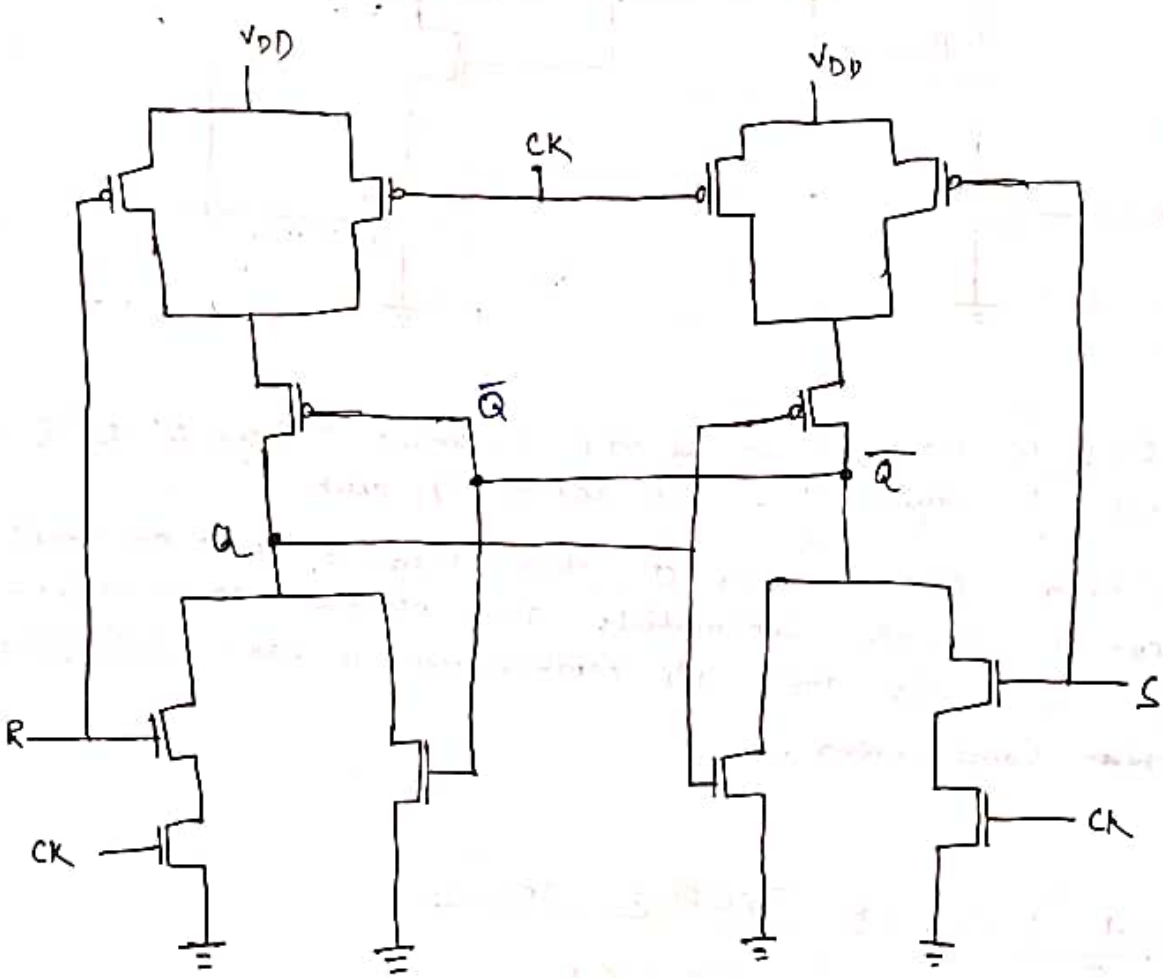
$\bar{Q} = \overline{S.CLK + Q}$

$Q = \overline{R.CLK + \bar{Q}}$

CLK	S	R	Q̄	Q	
0	0	—	Q̄	Q	- hold
0	1	—	1	0	- Reset
1	0	—	0	1	- Set
1	1	—	0	0	- not allow

→ If the ~~clock~~ clock (CK) is equal to logic '0', the input signal have no influence upon the ckt response.

→ The o/p of the two AND gates will remain at logic '0', which forces the SR Latch to hold its current state regardless of S & R.



→ To active the ckt the clock signal is necessary.

→ When CK=1 then only S,R inputs can affect the ckt output.

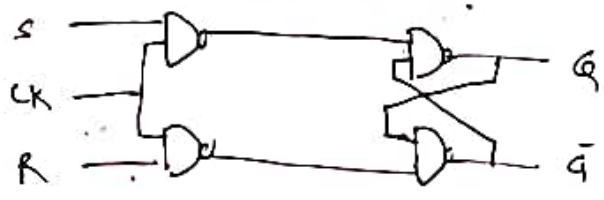
→ When CK=1, S=0 & R=0 then the FlF holds the previous o/p.

→ When CK=1 & S=1, R=0. then the FlF is set i.e Q=1 & Q̄=0.

→ When CK=1 & S=0, R=1.. then FlF is Reset i.e Q=0 & Q̄=1.

→ When all S=1, R=1 & CK=1 then Q & Q̄ all becomes low. Thus contradicting the complementary nature of Q & Q̄. So this combination is called not allowed combination.

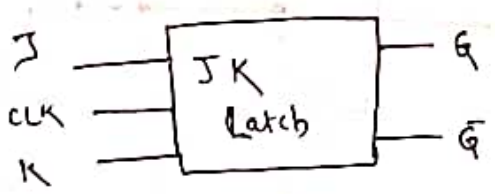
OR

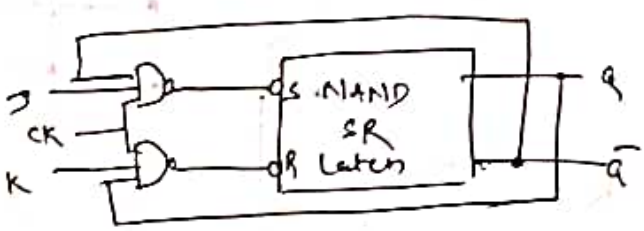


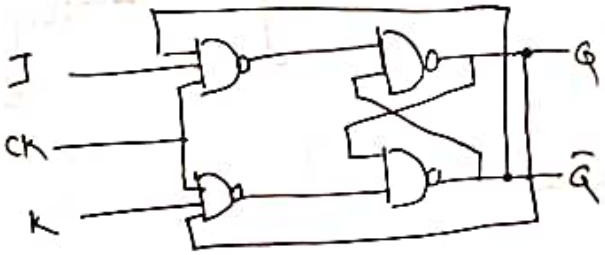
clocked JK Latch or JK Flip Flop:

→ SR Latch suffers a problem that when all inputs are high & then it's output state is indetermined or not allowed condition.

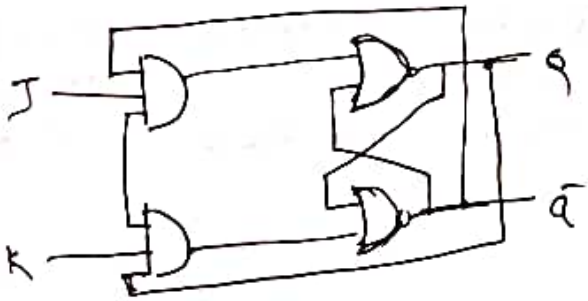
→ To overcome this problem add two feedback line from o/p to input which becomes a JK FF.







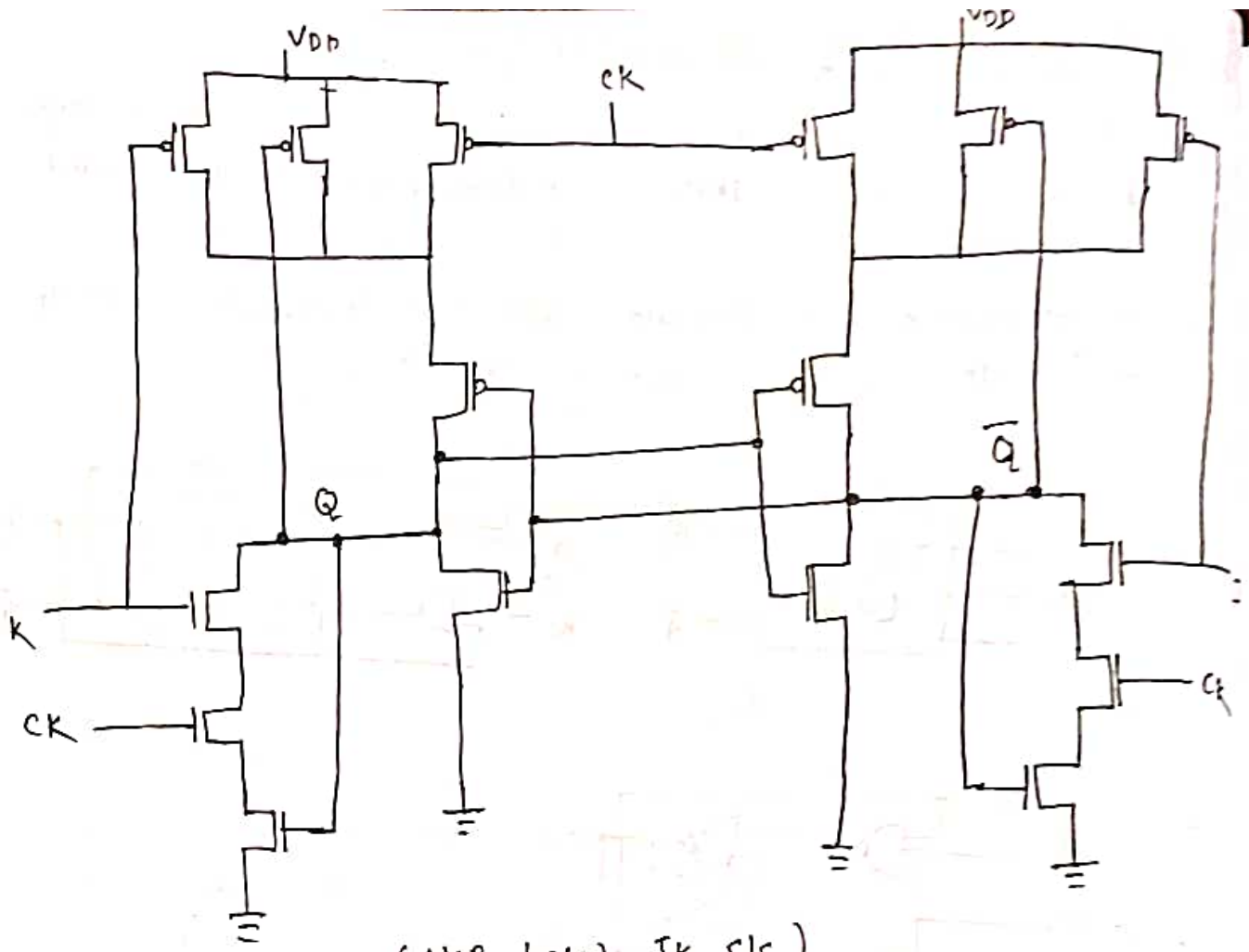
(JK Using NAND Gate)



(JK Using NOR Gate)

Truth Table:

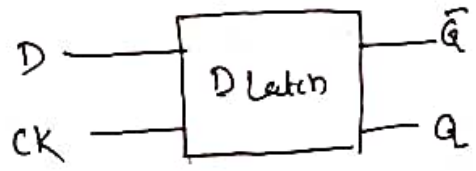
CK	J	K	Q_{n+1}	$\overline{Q_{n+1}}$	Operation
1	0	0	Q_n	$\overline{Q_n}$	hold
1	0	1	0	1	reset
1	1	0	1	0	Set
1	1	1	$\overline{Q_n}$	Q_n	toggle

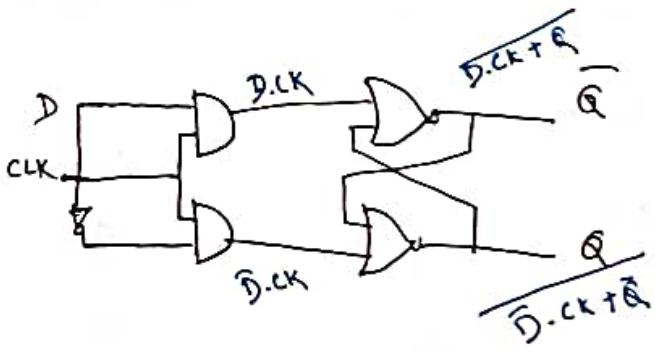


(NOR based JK FlF)

→ When CLK is made high then only change on input can affect
→ When CK=1, J=0 & K=0, then FlF holds the past state.
→ When CK=1, J=1 & K=0, then FlF is set i.e Q becomes 1 & $\bar{Q}$ becomes 0.
→ When CK=1, J=0 & K=1 " " " Reset i.e " Q " 0
→ When CK=1, J=1 & K=1 then the o/p becomes the complement of previous state.

D Flip flop:

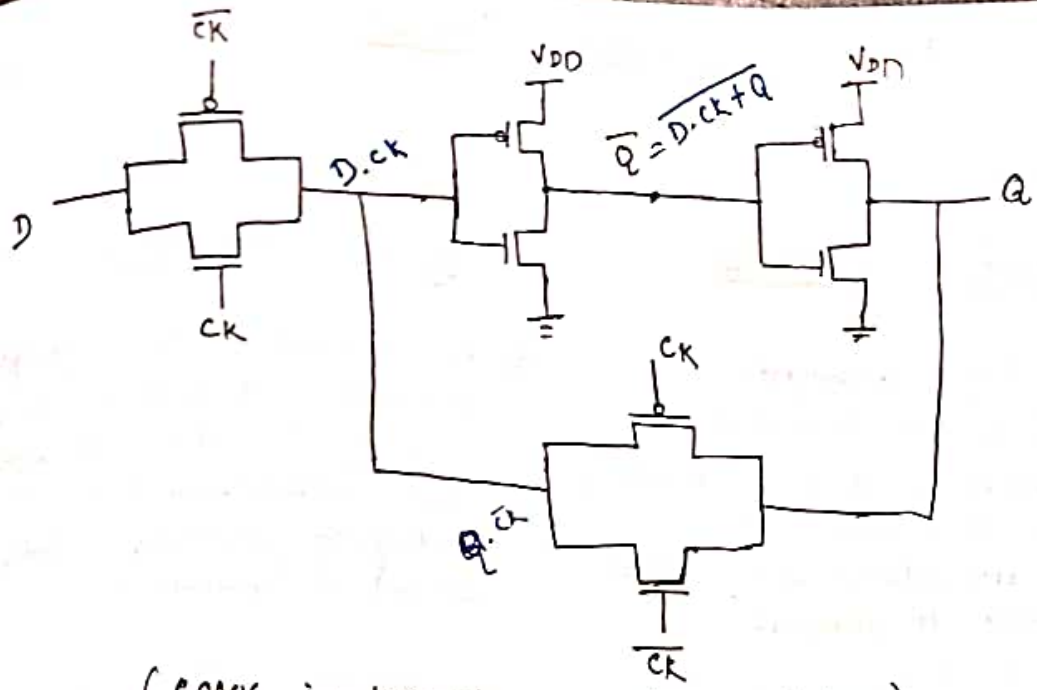




Truth Table:

CK	D	Q_{n+1}	$\bar{Q}_{n+1}$
	0	0	1
1			
	1	1	0
1			

Here when the clock is high then output Q becomes same as input after some time delay.



(CMOS implementation of D latch)

→ CK input acts as an enable signal which allows data to be accepted into the D latch.

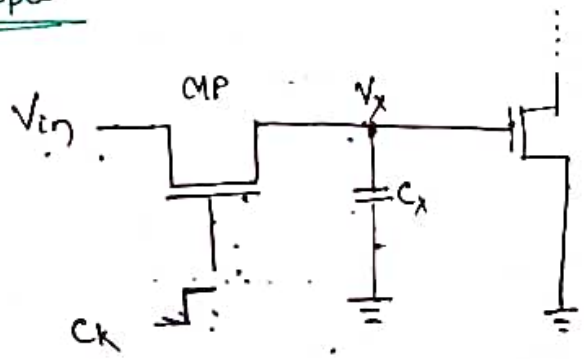
→ It is used for temporary storage of data or as a delay element..

Dynamic Logic Circuits

Introduction:

Static Logic Circuit	Dynamic Logic circuit
i) A static logic ~~circuit~~ gate generates its output corresponding to the applied input voltage after a certain time delay & it can preserve its output level as long as power supply is provided.	i) Operation of dynamic logic gate depends on temporary (transient) storage of charge in parasitic node capacitance instead of relaying on steady state circuit behaviour.
ii) Periodic updating of internal node voltages are not required.	ii) Periodic updating of internal node voltages required since stored charge in a capacitor can not be retained indefinitely.
iii) It does not require clock.	iii) It requires periodic clock for charge refreshing.
iv) It requires large number of transistors to implement a function.	iv) It requires less no of transistors to implement a function.
v) Requires more silicon area.	v) Requires less silicon area.
vi) Power consumption more	vi) Less power consumption.
vii) Synchronisation not possible	vii) Synchronisation possible.
viii) More time delay	viii) Less time delay.

Principle:



→ The fundamental building block of nMOS logic ckts, consisting an nMOS pass transistor driving the gate of another nmos transistor.

→ The parasitic input capacitance C_x plays an important role in dynamic operation of this ckt.

→ The input pass transistor M_P is being driven by the external periodic CLK signal & act as an access switch to either charge up or charge down the C_x depending upon i/p signal V_{in}.

→ Thus the two possible operations:

1) When the clock is high (CK=1), the pass transistor turns on. The capacitor C_x is charge up (if $V_{in}=1$) or charge down (if $V_{in}=0$) through pass transistor M_P, depending on the input voltage level V_{in}. The output of the depletion-load nMOS inverter obviously assumes a logic-low or logic-high level, depending on the voltage V_x.

2) When the clock is low (CK=0) the pass transistor M_P turns off, & the capacitor C_x is isolated from input voltage V_{in}. Since there is no current path from the intermediate node x to either V_{DD} or ground, the amount of charge stored in C_x during previous cycle determines the output voltage level a.

→ The "hold" operation during the inactive clock cycle is accomplished by temporarily storing charge in the parasitic capacitance C_x.

→ Correct operation of the ckt critically depends on how long a sufficient amount of charge can be retained at node x, before the o/p state change due to charge leakage.

→ Therefore, the capacitive intermediate node x is also called a "soft node". The nature of soft node makes the dynamic ckt more vulnerable to the so called single event upsets (SEUs) caused by a particle or cosmic ray hits on integrated ckt.

Note: The pass transistor M_P provides the only current path to intermediate capacitive node X (soft node).

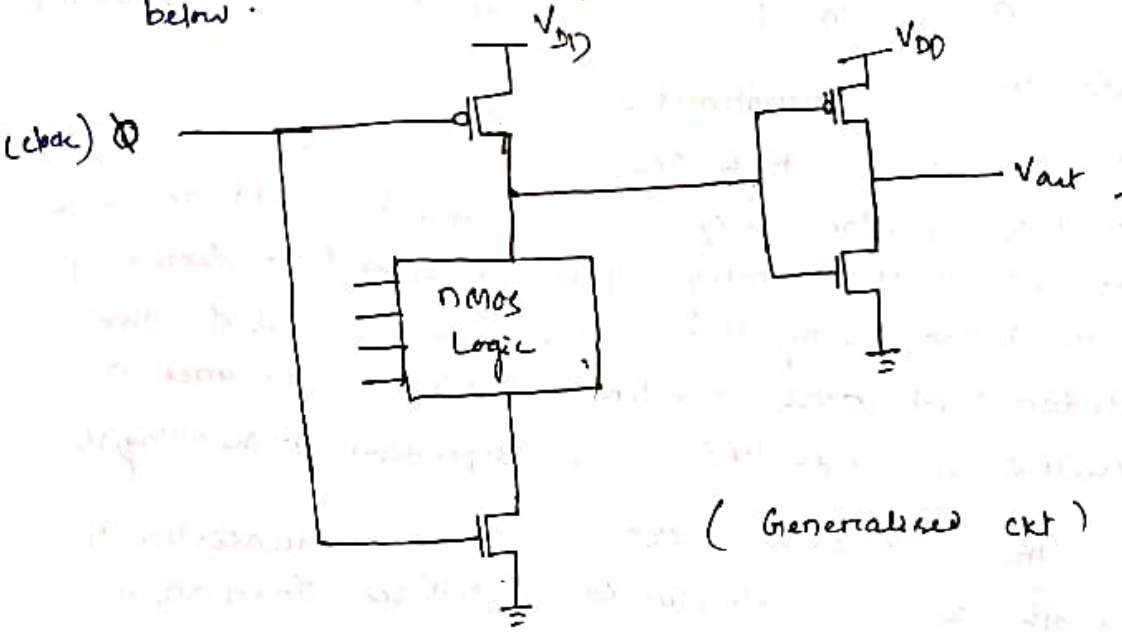
any input is low clock output is high & when all inputs are high

High- Performance Dynamic CMOS Circuits :

The goal of using high performance dynamic CMOS ckt is to achieve reliable, high speed, compact ckts using the least complicated clocking scheme possible.

Domino CMOS Logic :

→ Generalized ckt diagram of a domino CMOS logic gate is shown below.



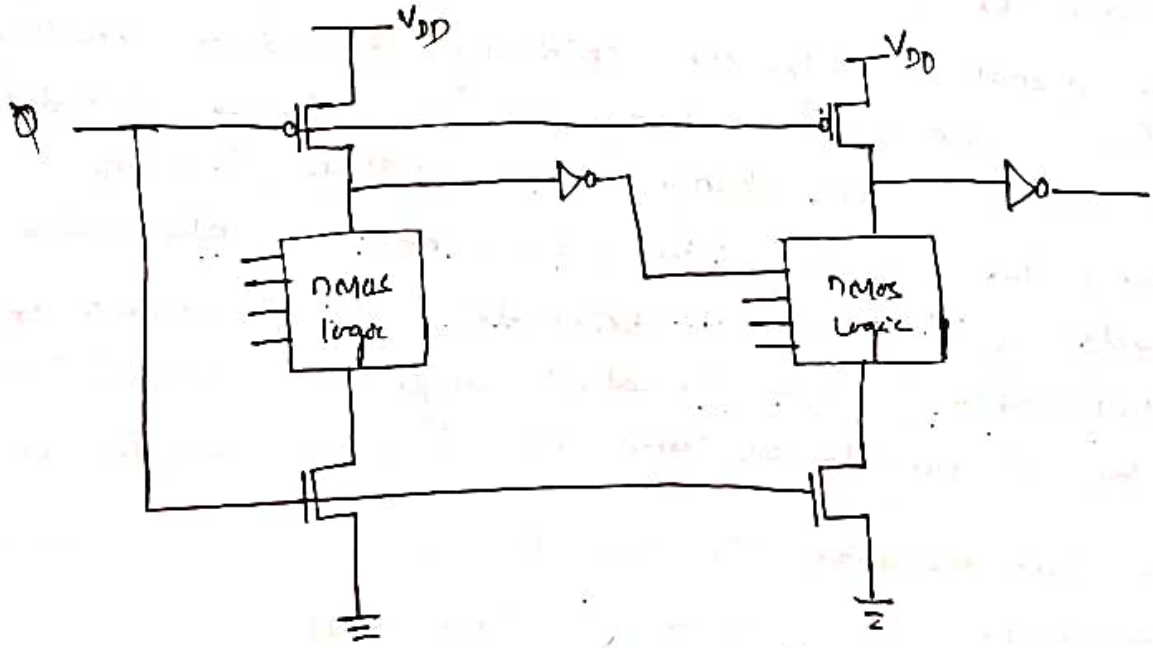
(Generalised ckt)

Φ → clock

→ It indicates that a dynamic CMOS logic stage is cascaded with a static CMOS inverter stage.

→ The addition of the inverter allows us to operate a number of such structures in cascade.

⇒ Cascaded domino CMOS Logic Gate :



(Cascaded domino CMOS Logic)

→ During the precharge phase when Φ = low, the output node of dynamic CMOS stage is precharged to logic high level & the output of CMOS inverter becomes low.

→ At the beginning of evaluation phase when clock rises again there are two possibilities.

i) The output node of dynamic CMOS stage is discharged to logic low level.

or ii) The output node remains at logic high. Consequently the consequently the inverter can make at most one transition from 0 to 1.

→ So when we build a system by cascading domino CMOS logic gate, all the input transistors in the subsequent block will be off because inverter output = 0 during precharge phase.

→ During the evaluation phase each inverter output can make at most one ~~transistor~~ transition (0 to 1) & hence each input of subsequent logic block can also make at most one transition (0 to 1).

Advantages:

i) Domino CMOS logic gates allow a significant reduction in number of transistor required.
ii) Less area Required.
iii) Less Power dissipation.
iv) More Speed.

Limitation of domino CMOS Logic:

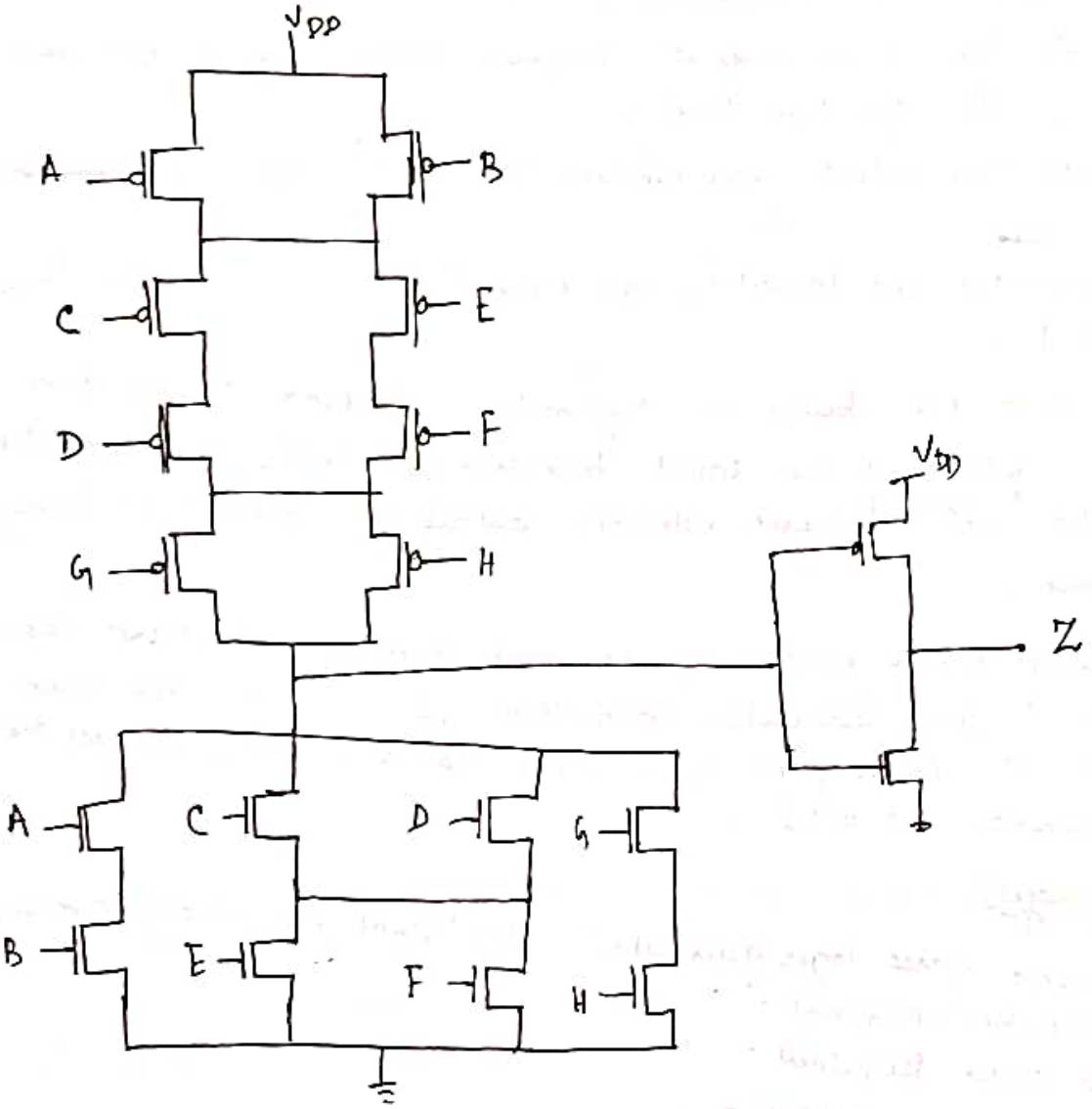
i) Only non inverting structures can be implemented using domino CMOS, so if necessary, inversion must be carried out by using conventional CMOS Logic

ii) Charge sharing problem.

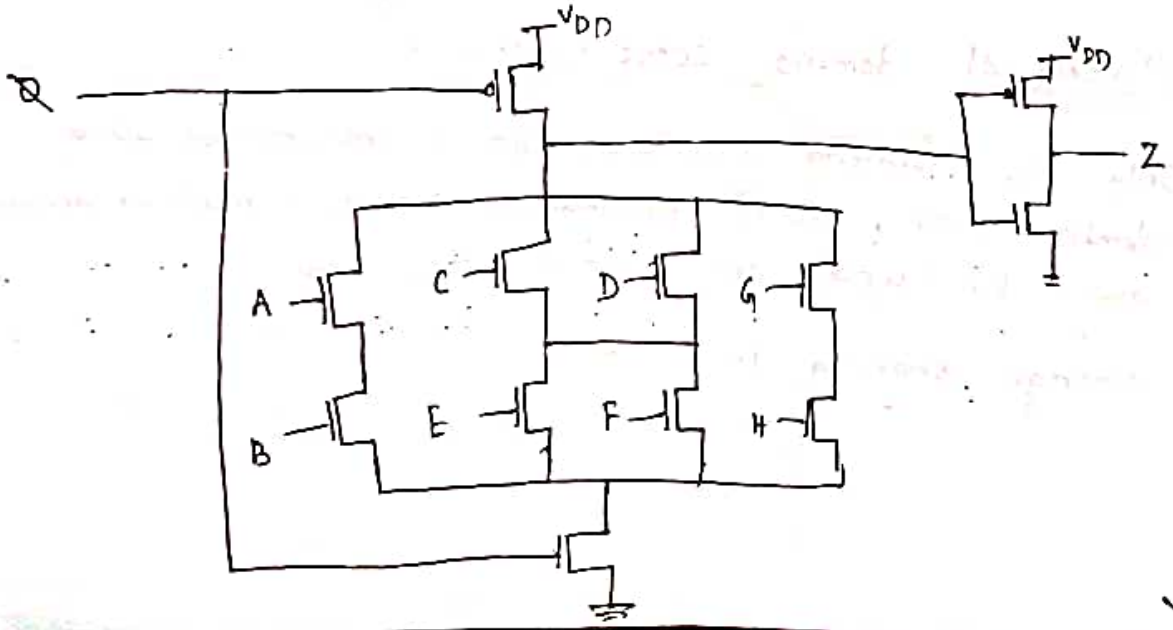
ii) Examples: $Z = AB + (C+D)(E+F) + GH$

Realize & implement this by CMOS static & domino CMOS technology.

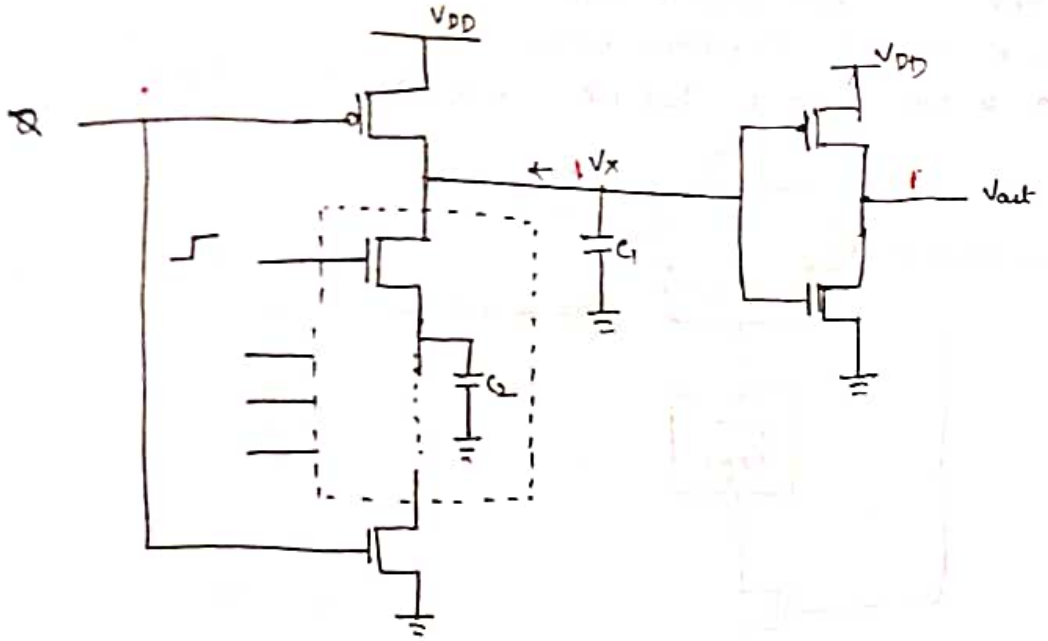
Conventional CMOS Logic:



domino CMOS:



Charge Sharing Problem:



→ Consider the domino CMOS Logic gate as shown above where intermediate node capacitance C_2 is comparible in size to C_1.

→ Assume that all inputs are low initially & intermediate node voltage across C_2 has 0V.

→ During the precharge phase when A = low, the output node capacitance C_1 is charged up to logic high (i.e VDD)

→ During evaluate phase when A = high, (if) let the input signal of uppermost nmos transistor become high so the charge initially stored in output node capacitance C_1 will be changed by C_2 which called charge sharing problem.

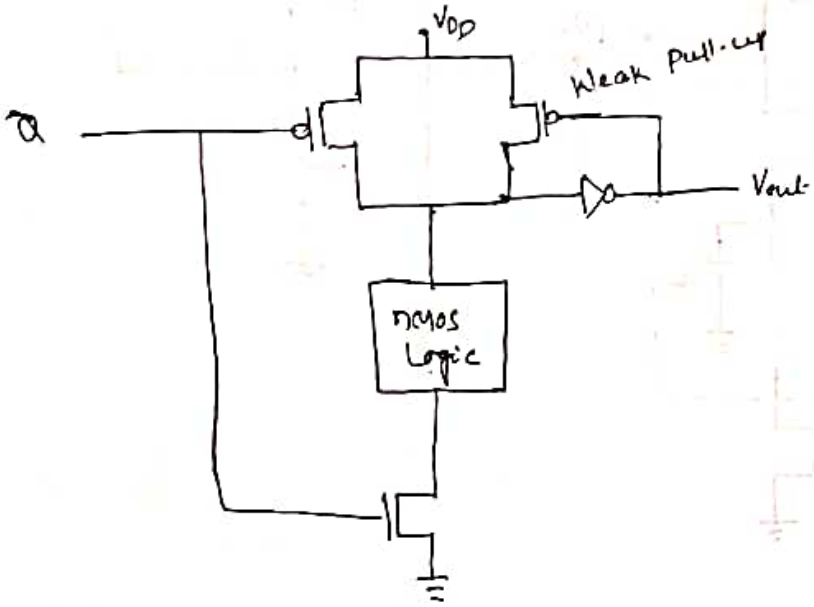
→ The output node voltage after charge sharing will be $\frac{V_{DD} \cdot C_1}{C_1 + C_2}$. Now if $C_1 = C_2$ so the o/p node voltage will be $\frac{V_{DD}}{2}$.

→ Now if the Vth of the inverter is more than $V_{DD}/2$ the output voltage of the following inverter will be high which is a logic error.

...high & when all inputs are high, out is low.

Solution of charge sharing problem:

- Make C_2 much smaller than C_1.
- Make V_{th} of inverter smaller.
- Add a weak pMOS pull up device as shown below.

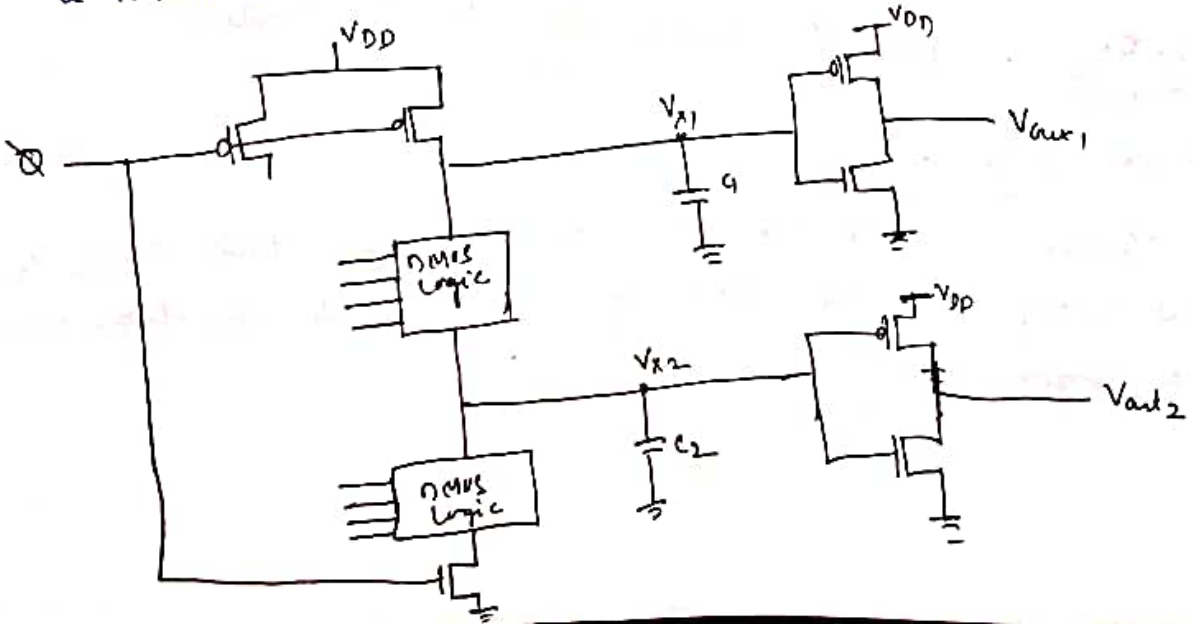


→ This weak pMOS pull up device essentially forces high output unless there is a strong pull down path bet^n the output & ground.

→ The weak pMOS transistor will be turned ON only when the precharge node voltage is kept high. Otherwise it will be turned off when Vout is high.

→ Another solution is to use separate pMOS transistor to precharge all intermediate nodes in nMOS pull down tree which have large parasitic capacitances.

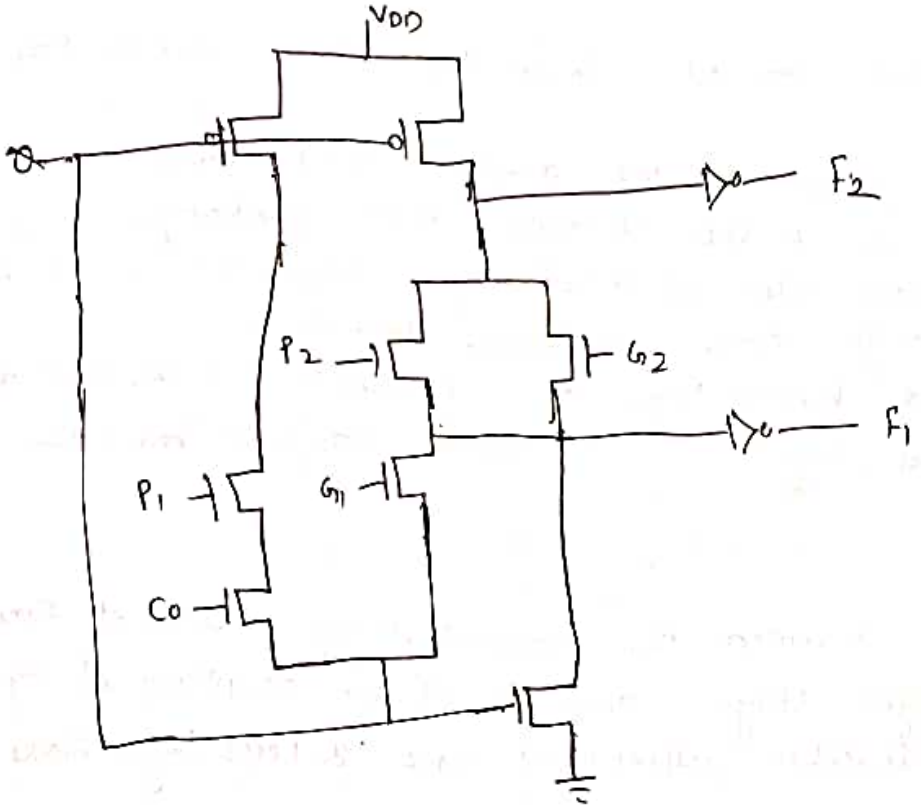
→ Advantage of using multiple precharge transistor enables us to use precharged intermediate nodes as a resource of additional output as shown below.



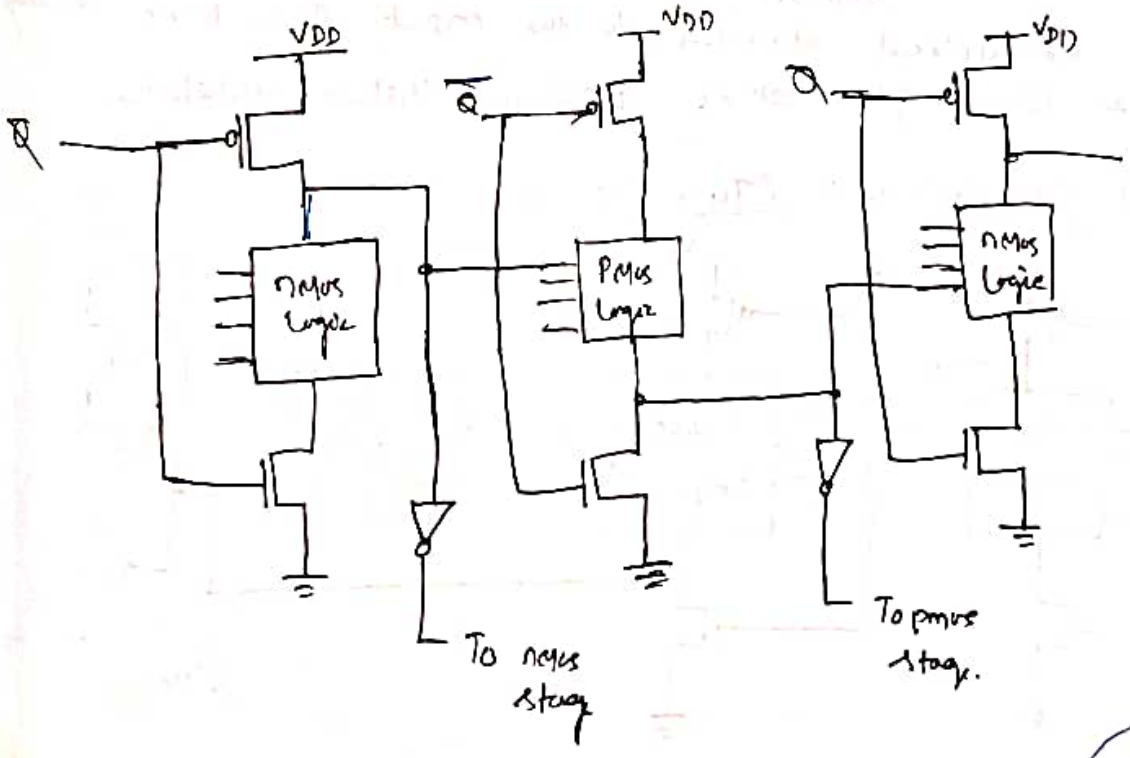
Q. Implement $F_1 = G_1 + P_1 C_0$

$F_2 = G_2 + P_2 G_1 + P_2 P_1 C_0 = G_2 + P_2 (G_1 + P_1 C_0)$

~~Simulate~~ using domino CMOS Logic Gate



NORA CMOS Logic (NP- Domino Logic):



output is high & when all inputs are high

→ In domino CMOS logic gates all logic operations are performed by nMOS transistors acting as pull down networks, while the role of pMOS transistors is only to precharge the o/p nodes.

→ In NORA CMOS or NP domino logic both nMOS & PMOS are used.

→ When clock signal is low the output nodes of nMOS logic blocks are precharged to VDD through PMOS precharge transistors where as output nodes of PMOS logic blocks are pre discharged to 0V through nMOS discharge transistor.

→ When clock Φ becomes high Φ $\overline{\Phi}$ becomes low & all cascaded nMOS & PMOS logic blocks evaluate one after the other.

Advantages:

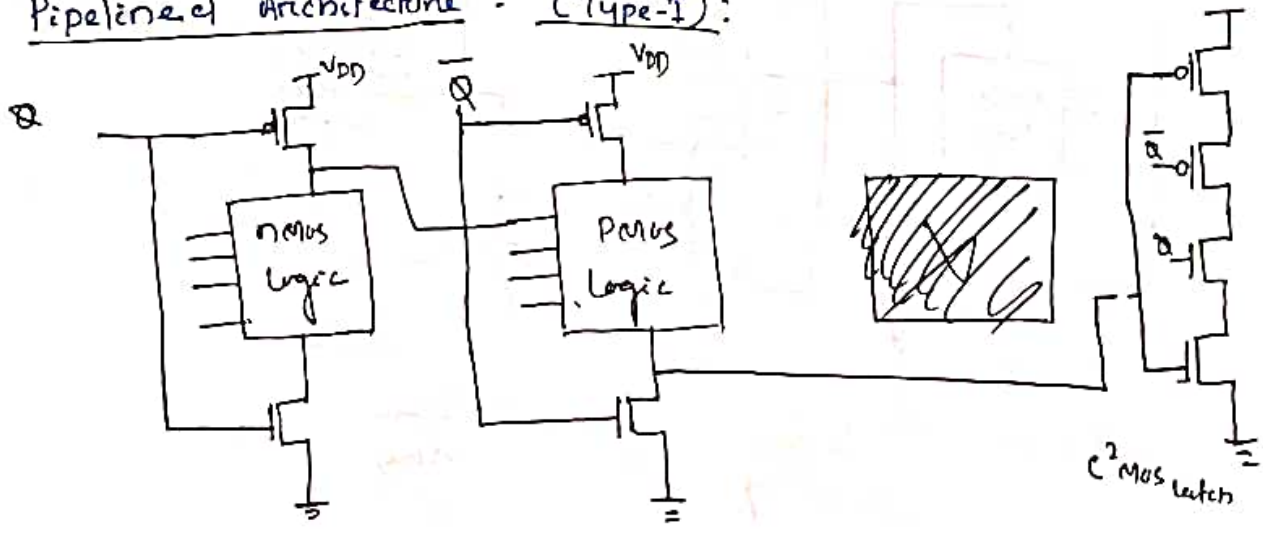
1) Static CMOS inverter not required at the output of every dynamic logic stage. Instead direct coupling of logic blocks is feasible alternating nMOS & PMOS logic blocks.

2) NORA logic is compatible with domino CMOS logic. That is outputs of NORA nMOS logic blocks can be inverted & then & then applied to the i/p of Domino CMOS block which also driven by clock signal Φ.

Similarly the buffered o/p of a domino CMOS stage can be applied directly to the input of a NORA CMOS stage

3) NORA CMOS Logic allows pipelined system architecture.

Pipelined Architecture : (Type-I):

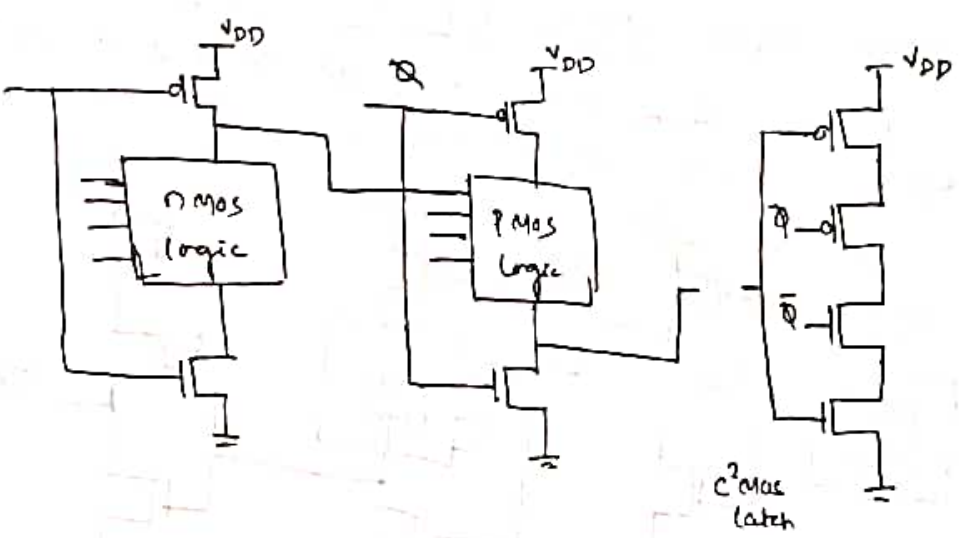


(Pipelined System means Cascading alternating Φ & $\overline{\Phi}$ section)

→ It consists of nMOS & PMOS logic log blocks & clocked CMOS (C^2MOS) output buffer.

→ All stages of this ckt perform pre-change - discharge operation levels when the clock is low & all stages of the ckt evaluate output a $\overline{Q}$ section meaning the evaluation occurs during active $\overline{Q}$.

Type II :

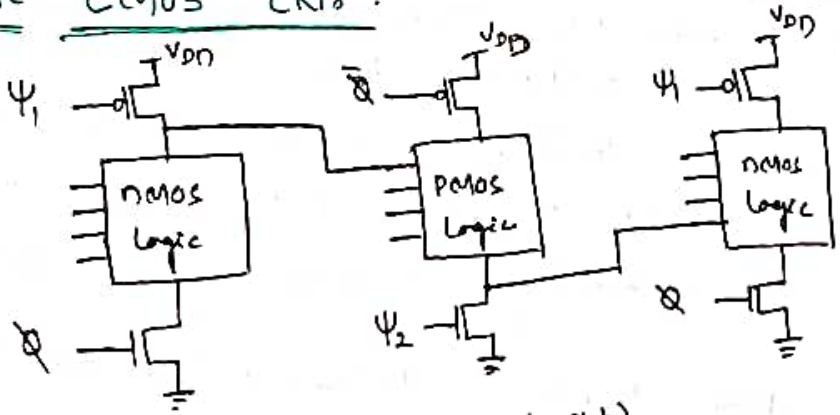


→ In this case all logic stages perform pre-charge & discharge operations when clock is high & all stages evaluate when clock is low. Therefore this ckt is called a $\overline{Q}$ section meaning the evaluation occurs during active $\overline{Q}$.

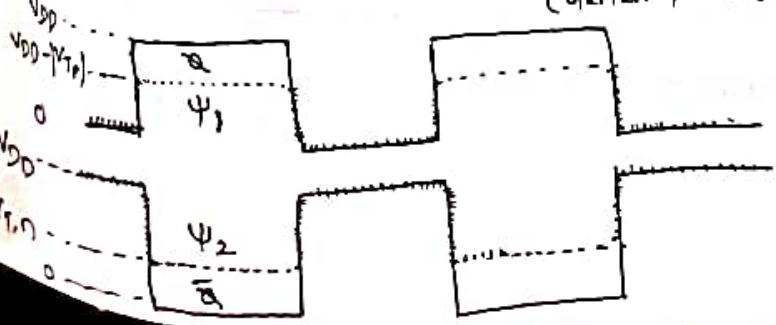
Disadvantage of NORA CMOS Logic:

NORA CMOS logic gates suffer from charge sharing & leakage problem.

Zipper CMOS CKts :



(General ckt).



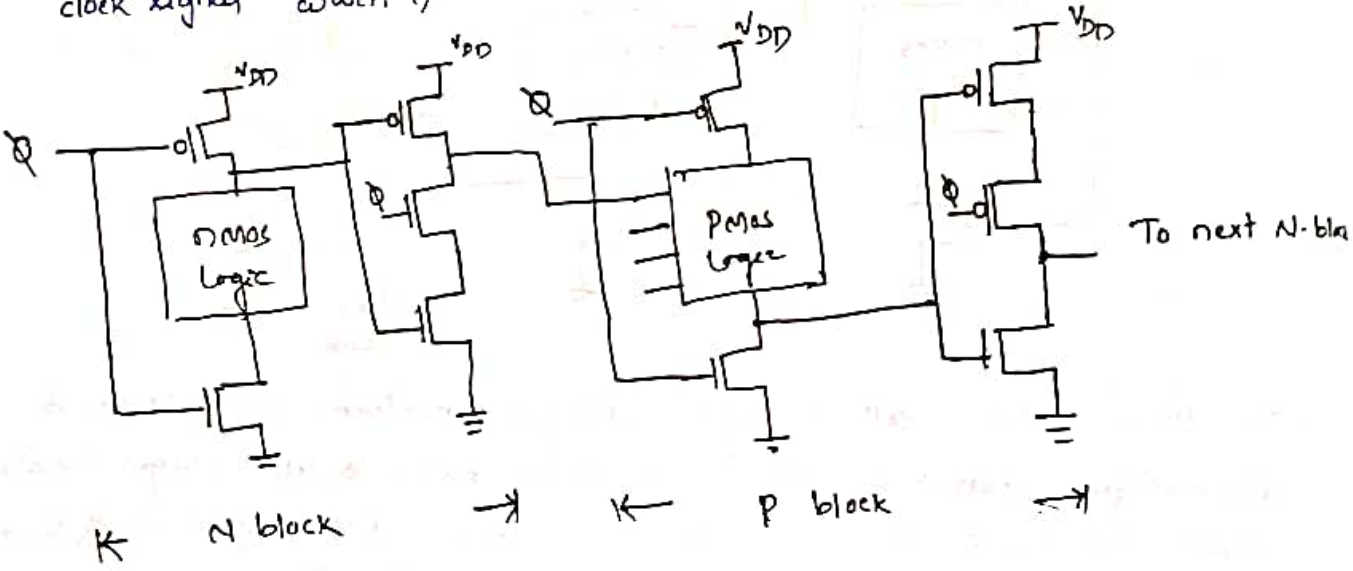
(clock signal of Zipper CMOS)

→ The basic ckt architecture of Zipper CMOS essentially identical to NORA CMOS except clock signals.

→ In particular, the clock signals which drive PMOS precharge & NMOS discharge transistors allow these transistors to remain in weak conduction or near cut-off during the evaluation phase thus compensating the charge sharing or charge leakage problems.

True Single-Phase clock (TSPC) Dynamic CMOS:

→ It is different from NORA CMOS ckt in that it uses single clock signal which is never inverted.



(A pipelined true single-phase clock CMOS System)

→ The ckt consists of alternating blocks called nblock & Pblock.

→ Each block is driven by same clock signal Φ.

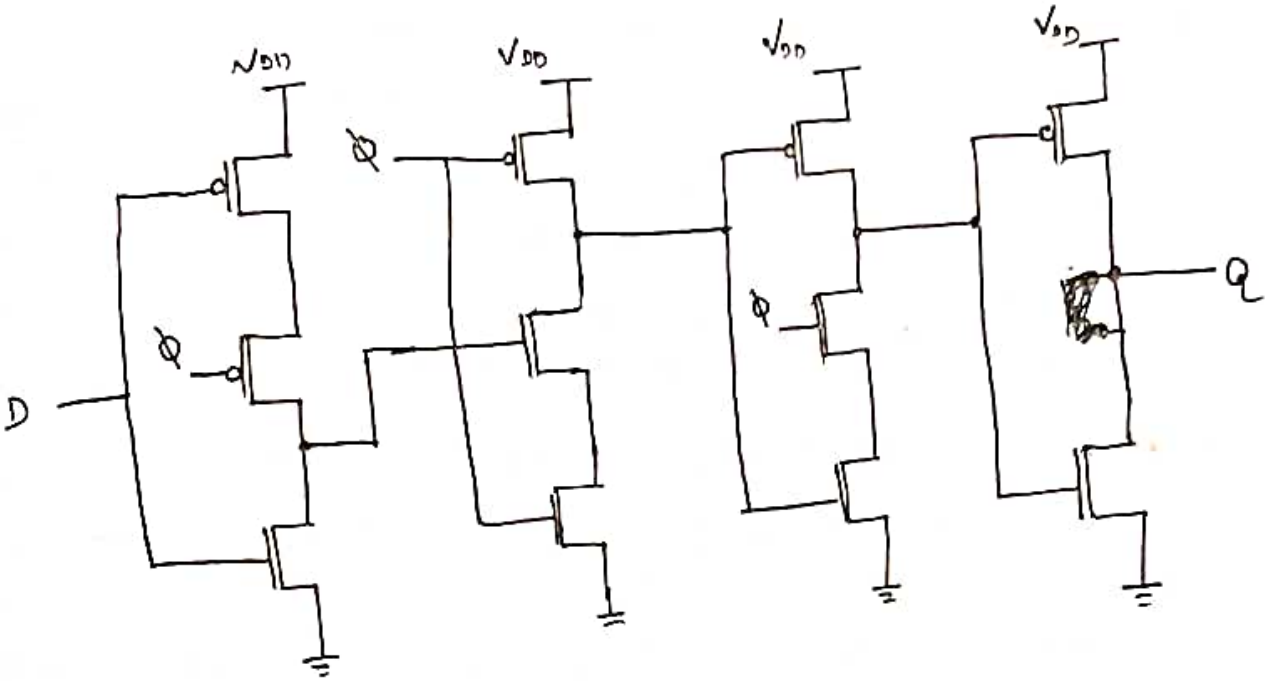
→ An n block is constructed by cascading a dynamic NMOS stage & a dynamic latch while a P-block is constructed by a dynamic PMOS & a dynamic latch.

→ When the clock signal is low the output node of nblock is precharged to VDD by PMOS precharge transistor.

→ When clock becomes ~~two~~ high th nmos logic stage output is evaluated.

→ On the otherhand Pblock predischarges when the clock is high & evaluates when clock is low. This means ~~that~~ it allows pipelined operation using single clock signal.

Rising edge triggered or Positive edge Triggered D Flip-flop using TSRC CMOS Logic :



→ When the clock signal is low the first stage acts as a transparent ~~transistor~~ latch to recive the i/p signal, while the o/p node of the 2nd stage is precharged.

→ During this time, the 3rd & 4th stage simply keep the previous output state.

→ When clock signal switches from low to high, the first stage ceases to be transparent & the 2nd stage starts evaluation.

→ At the same time, the 3rd stage becomes transparent & transmits the sampled value to the output.

→ The final stage is only used to obtain non-inverted output level.

Processor Technology:

* Technology is defined as a manner of accomplishing a task, especially using technical process, methods or knowledge.

→ The three types of technologies that are central to embedded system design:

i) Processor technologies
ii) Ic technologies
iii) Design technologies

- Processor technology relates to the architecture of the computation engine used to implement a system's desired functionality.
- Processor is usually associated with programmable software processors are non programmable digital systems.

General Purpose Processor - Software:

- The designer of a general purpose processor or Microprocessor, builds a programmable device that is suitable for a variety of applecation.
- One feature of such a processor is a program memory where program can't be built into the digital ckt, & program is done according to the requirement.
- Another feature is general datapath: The data path must be general enough to handle a variety of computation.
- An embedded system designer simply uses a general purpose processor, by programming the processor memory to carry out the required functionality.
- Using general purpose processor in an embedded system, several design benifits are there.

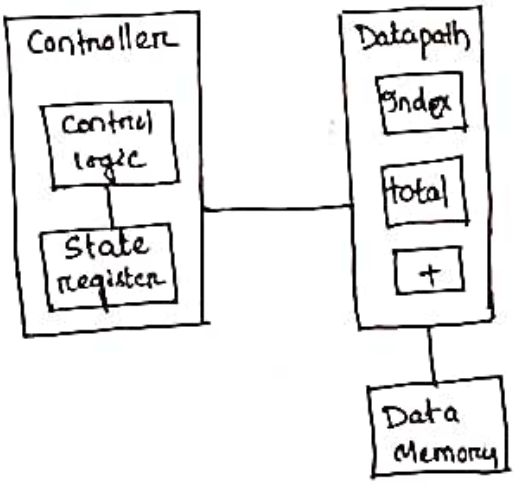
i) Time to Market & NRE cost low (non reccurring engineering cost)
ii) High flexibility
iii) Unit cost low
iv) Fast performance.

Single - Purpose Processor - Hardware:

- A single purpose processor is a digital ckt design to execute exactly one program. Ex: Digital camera.
- The JPEG coder, execute a single program that compress & decompress video frames.
- An embedded system designer may create a single purpose processor by designing a custom digital ckt.
- Benifits of single - purpose processor:
  - i) Performance may fast
  - ii) Size & power small
  - iii) Unit cost low for large quantities

Drawbacks:
- i) Design time & NRE cost high
- ii) Low flexibility
- iii) Unit cost high for small quantities

Architecture:



Application - Specific Processors:

→ An application specific instruction set processor (ASIP) is a programmable processor optimized for a particular class of applications having having common characteristics, such as embedded control, digital signal processing or Telecommunication.

→ Using an ASIP in an embedded system can provide the benifit of flexibility, good performance, power & size.

→ But it requires a large NRE cost to built the processor.

→ Two well known types of ASIPs are i) Microcontrollers ii) DSP

* Microcontroller: It is a microprocessor that has been optimized for embedded control applications.
- Microcontroller tends to have simple datapath that excel at bit level operations & at reading and writing of external bits.
- Incorporation of peripherals (ADC, PWM, timer, Counter) enables single chip implementations & hence smaller & lower cost products.

* Digital Signal Processors (DSPs)
- It is another type of ASIP.
- A DSP is a microprocessor design to perform common operations on digital signals.
- These operations carry out common signal processing tasks like signal filtering, transformation or a combination.
- Such operations are usually math-intensive.
- To support such operations, a special purpose datapath components are required.

IC Technology:

→ Ic technology involves the manner in which we map a digital (gate level) implementation onto an IC.

→ An Ic often called a chip is a semiconductor device consisting of a set of connected transistors & other devices.

→ Ic technologies are differ by customization of Ic for a particular design.

→ Ic technology is independent of from processor technology.

→ Semiconductors consists of numerous layers. The bottom layers form transistors, middle layer form logic components & top layers connects these components with layers.

→ One way to create these layers is by depositing photo-sensitive chemicals on the chip surface & then shining light through mask to change regions of chemicals.

→ A set of mask is often called layout.

→ For each Ic technology all layer must eventually be built to get a working Ic.

Full - Custom/VLSI :

→ On full-custom IC technology, we optimize all layers for a particular embedded system's digital implementation.

- Such optimization includes placing the transistor to minimize interconnection lengths, sizing the transistor to optimize signal transmissions & routing wires among the transistors.
- ~~Full custom~~ Once we complete all masks, we send the mask specifications to a fabrication plant that builds actual ICs.
- Full custom Ic design, often referred to as VLSI design has a very high NRE cost & long turnaround times, but having excellent performance with small size & power.

Semiconductor ASIC (Gate array & standard cell)

- An application specific Ic (ASIC) technology, the lower layers are fully or partially built, leaving us to finish the upper layers.
- On gate array ASIC technology, the masks for the transistor & gate levels are already built (Ic already consists of array of gates).

→ The remaining task is to connect these gates to achieve our particular implementation.

→ On a standard-cell ASIC technology, logic cells such as AND gate or an AND-OR-INVERT combination, the mask portions are predesigned.

→ Thus the remaining task is to arrange these portions into complete masks for the gate level, & then to connect the cells.

→ ASICs are the most popular Ic technology as they provide for good performance & size with less NRE cost than full-custom Ic.

PLD (Programmable Logic Device):

→ On PLD technology, all layers already exists.

→ The layers implement a programmable ckt, where programming has a lower-level meaning than a software program.

→ The programming that takes place may consists of creating or destroying connection between wires that connect gates, either by blowing a fuse or setting a bit in a programmable switch.

→ Small devices called programmers, connect to a desktop computer & typically perform such programming.

→ PLD is two types, ① Simple & complex.

→ Simple PLD is a programmable logic array (PLA) which consists of programmable array of AND & OR gates.

→ Another type ② is a PAL (Programmable array logic) which uses just one programmable array to reduce the number of expensive programmable components.

* One type of Complex PLD is FPGA (Field programmable gate array).

→ FPGA offers more general connectivity logic blocks.

- PLDs offer very low NRE cost & almost instant Ic availability.

→ Typically bigger than ASICs, higher unit cost, consume more power, slower