EC 8095-VLSI Design

Unit-I
INTRODUCTION OF MOS TRANSISTOR

PREPARED BY;
A.SURESH KUMAR,AP/ECE

www.rejinpaul.com

WHY VLSI DESIGN?

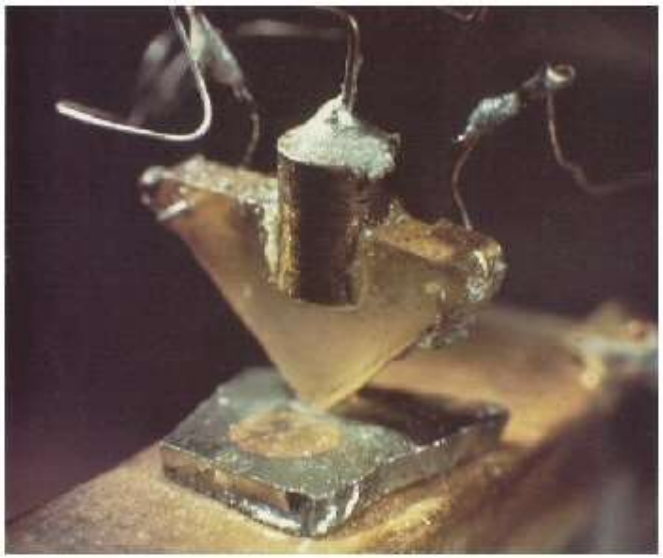
- Integration improves the design
  - Lower power consumption
  - higher speed(reducing delay)
  - Physically smaller(Area)
- Integration reduces manufacturing cost - (almost) no manual assembly.

www.rejinpaul.com

A Brief History
Invention of the Transistor

- Vacuum tubes ruled in first half of 20th century Large, expensive power-hungry, unreliable

- 1947: first point contact transistor (3 terminal devices)
  - Shockley, Bardeen and Brattain at Bell Labs



www.rejinpaul.com

A Brief History, contd..

- 1958: First integrated circuit
  - Flip-flop using two transistors
  - Built by Jack Kilby (Nobel Laureate) at Texas Instruments
  - Robert Noyce (Fairchild) is also considered as a co-inventor

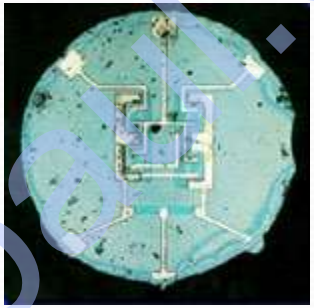
Kilby's IC

smithsonianchips.si.edu/ augarten/

www.rejinpaul.com

A Brief History, contd.

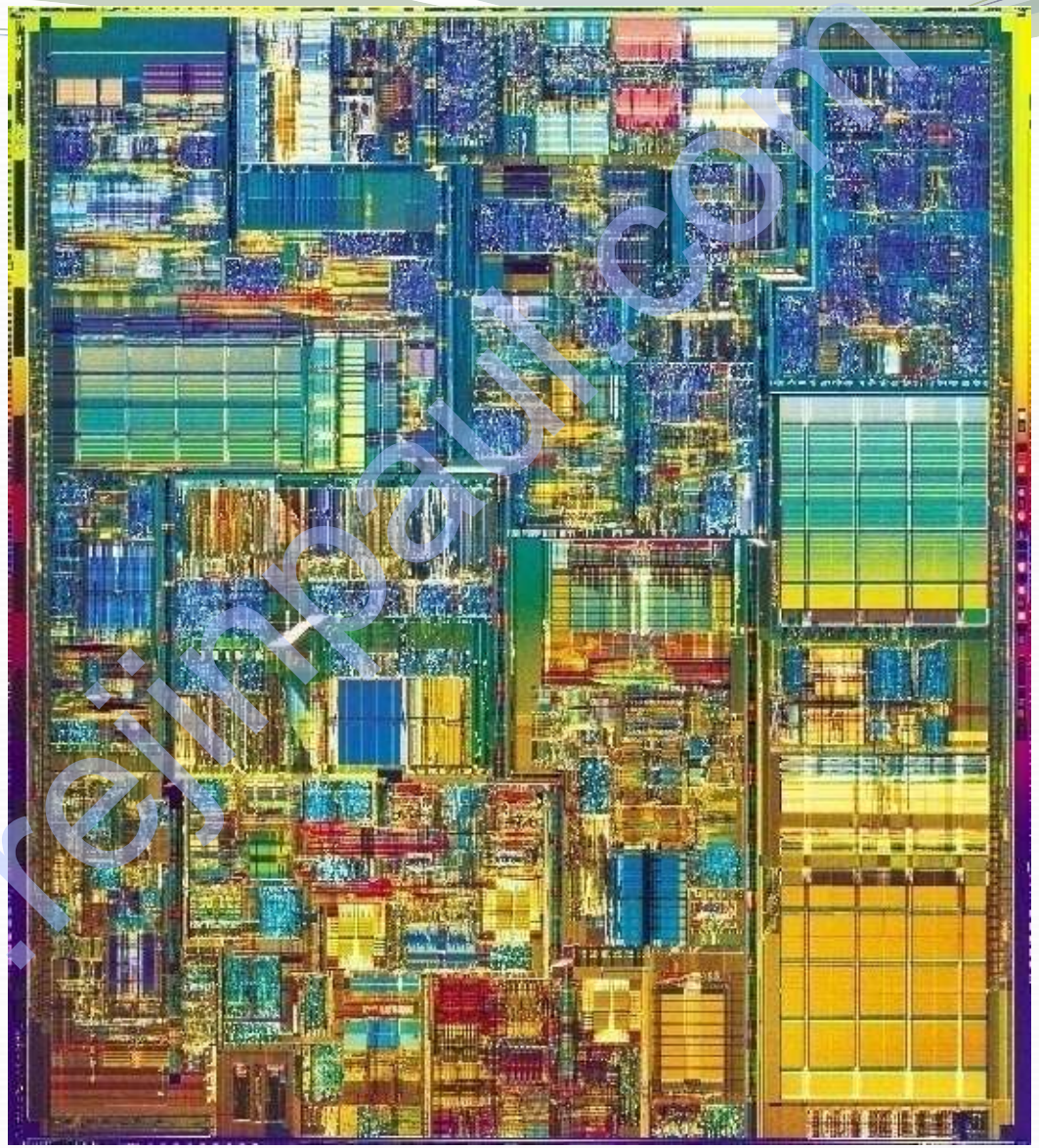
- First Planer IC built in 1961



- 2003
  - Intel Pentium 4 μprocessor (55 million transistors)
  - 512 Mbit DRAM (> 0.5 billion transistors)
- 53% compound annual growth rate over 45 years
  - No other technology has grown so fast so long
- Driven by miniaturization of transistors
  - Smaller is cheaper, faster, lower in power!
  - Revolutionary effects on society

www.rejinpaul.com

Pentium 4 Processor

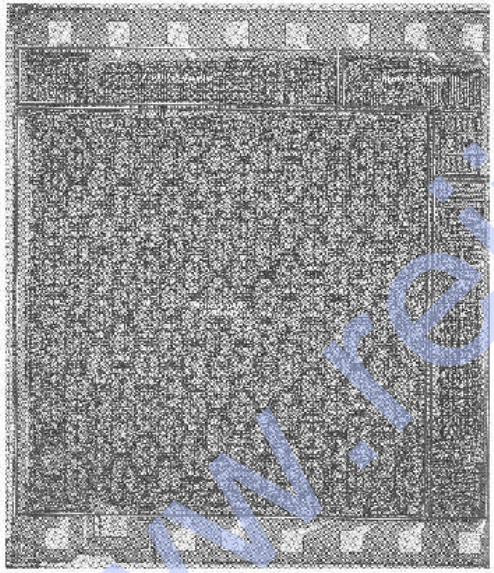


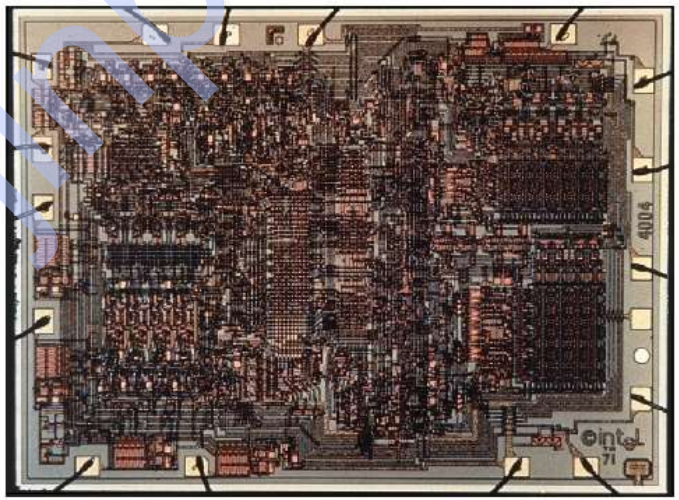
http://www.intel.com/intel/intelis/museum/online/hist_micro/hof/index.htm

www.rejinpaul.com

MOS Integrated Circuits

- 1970's processes usually had only nMOS transistors
  Inexpensive, but consume power while idle
- 1980s-present: CMOS processes for low idle power



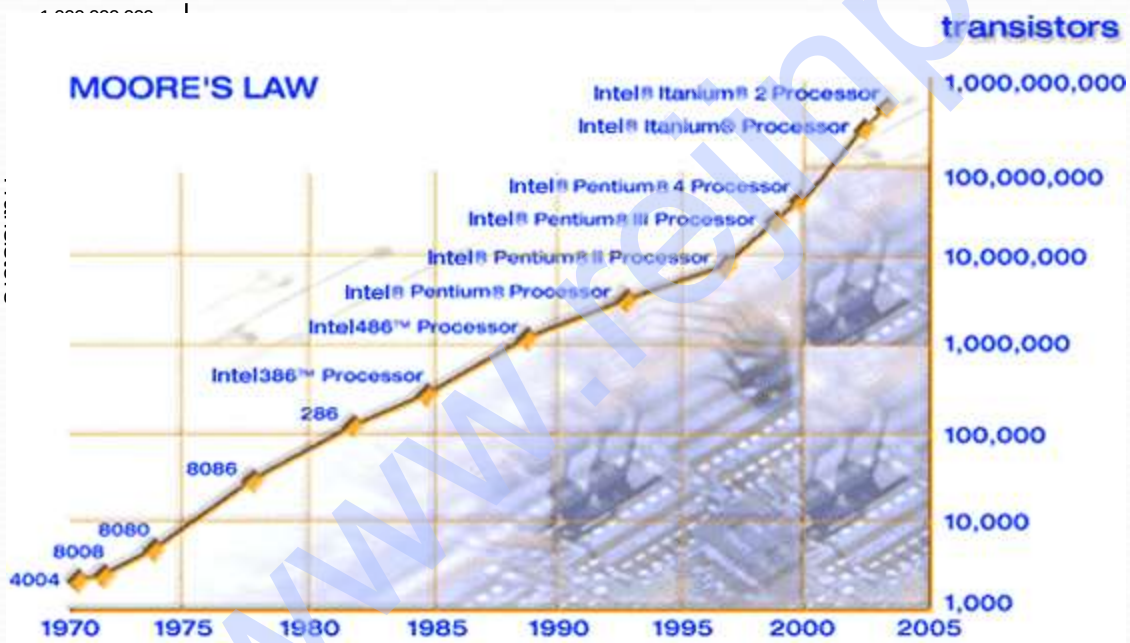


Intel 1101 256-bit SRAM Intel 4004 4-bit μProc

www.rejinpaul.com

Moore's Law

- 1965: Gordon Moore plotted transistor on each chip
  - Number of transistors per chip counts have doubled every18months.



http://www.intel.com/technology/silicon/mooreslaw/

Integration Levels

SSI: 10 gates

MSI: 1000 gates

LSI: 10,000 gates

VLSI: > 10k gates

www.rejinpaul.com

VLSI Applications

- VLSI is an implementation technology for electronic circuitry - analogue or digital
- It is concerned with forming a pattern of interconnected switches and gates on the surface of a crystal of semiconductor
- Microprocessors
  - personal computers
  - microcontrollers
- Memory - DRAM / SRAM
- Special Purpose Processors - ASICS (CD players, DSP applications)
- Optical Switches

www.rejinpaul.com

- Modern transistors are few microns wide and approximately 0.1 micron or less in length
- Human hair is 80-90 microns in diameter

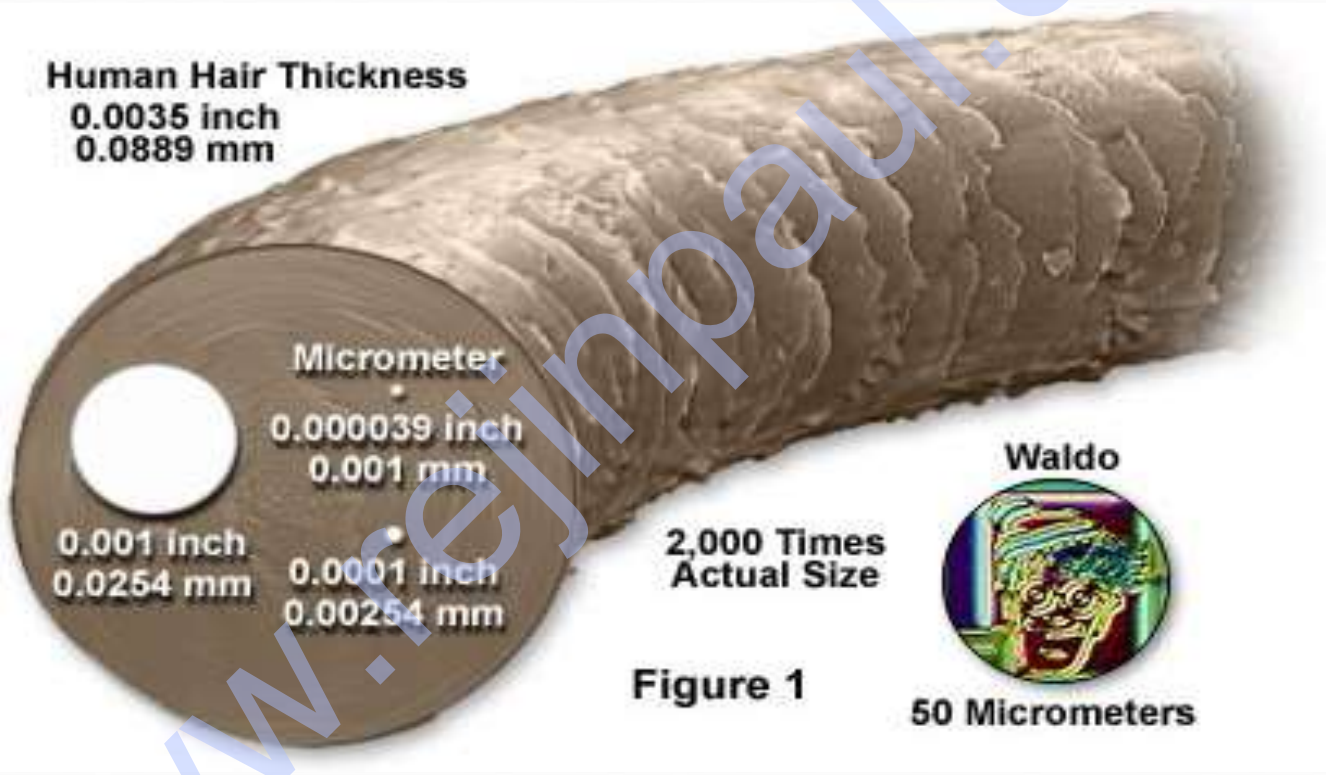


Figure 1

Ref: http://micro.magnet.fsu.edu/creatures/technical/sizematters.html

www.rejinpaul.com

What is a Silicon Chip?

- A pattern of interconnected switches and gates on the surface of a crystal of semiconductor.
- These switches and gates are made of
  - areas of n-type silicon
  - areas of p-type silicon
  - areas of insulator
  - lines of conductor (interconnects) joining areas together
    - Aluminium, Copper, Titanium, Molybdenum, polysilicon, tungsten
- The geometry of these areas is known as the layout of the chip
- Connections from the chip to the outside world are made around the edge of the chip to facilitate connections to other devices

www.rejinpaul.com

Switches

- Digital equipment is largely composed of switches
- Switches can be built from many technologies
  - relays (from which the earliest computers were built)
  - transistors
- The perfect digital switch would have the following:
  - switch instantly
  - use no power
  - have an infinite resistance when off and zero resistance when on
- Real switches are not like this!

www.rejimpaul.com

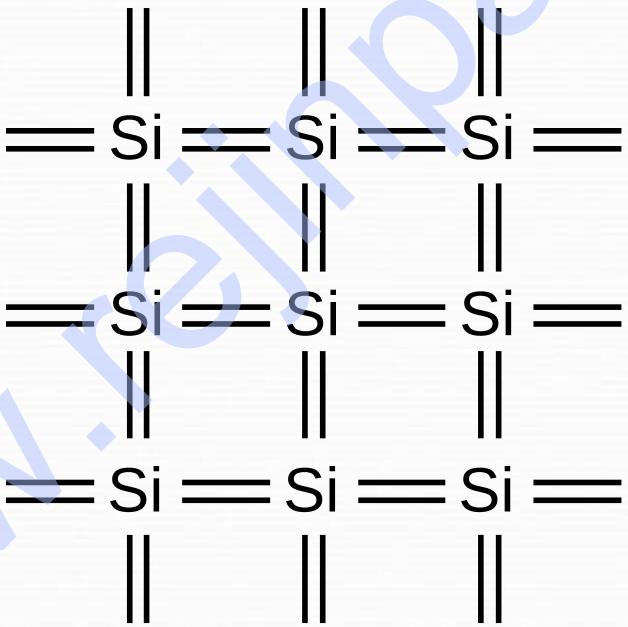
Semiconductors and Doping

- Adding trace amounts of certain materials to semiconductors alters the crystal structure and can change their electrical properties
  - in particular it can change the number of free electrons or holes
- N-Type
  - semiconductor has free electrons
  - dopant is (typically) phosphorus, arsenic, antimony
- P-Type
  - semiconductor has free holes
  - dopant is (typically) boron, indium, gallium

www.reinpaul.com

Silicon Lattice

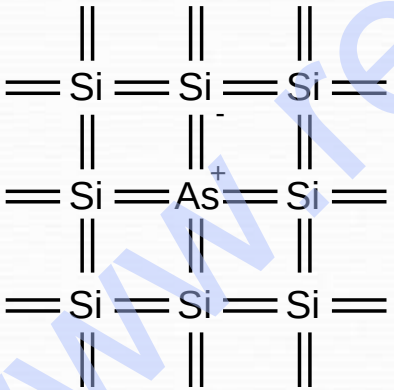
- Transistors are built on a silicon substrate
- Silicon is a Group IV material
- Forms crystal lattice with bonds to four neighbors

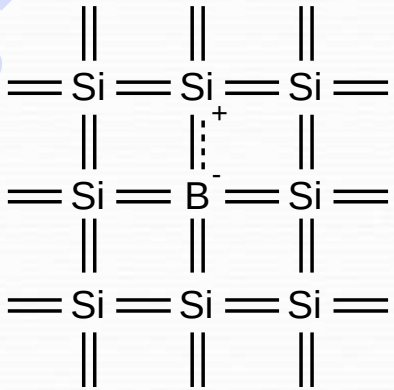


www.jeinpaul.com

Dopants

- Silicon is a semiconductor
- Pure silicon has no free carriers and conducts poorly
- Adding dopants increases the conductivity
- Group V: extra electron (n-type)
- Group III: missing electron, called hole (p-type)

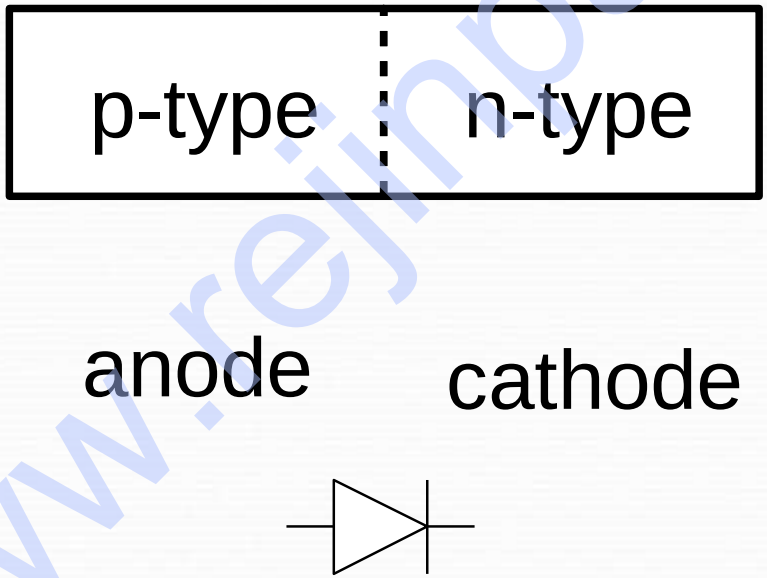




www.rejinpaul.com

p-n Junctions

- A junction between p-type and n-type semiconductor forms a diode.
- Current flows only in one direction



www.rejinpaul.com

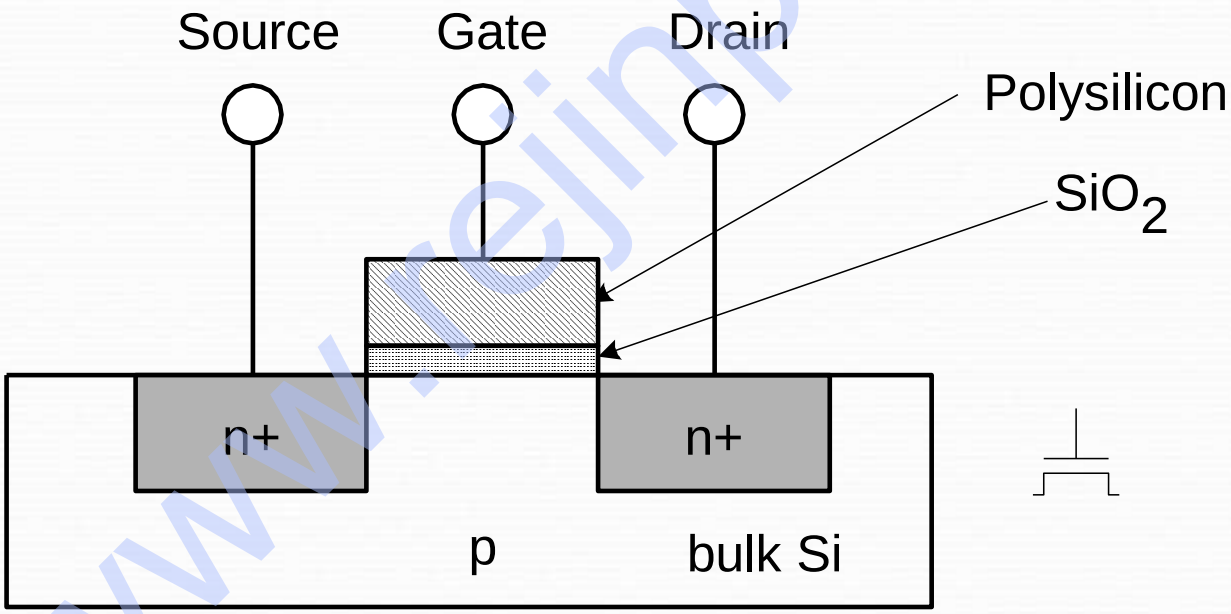
Transistor Types

- Bipolar transistors
  - npn or pnp silicon structure
  - Small current into very thin base layer controls large currents between emitter and collector
  - Base currents limit integration density.
- Metal Oxide Semiconductor Field Effect Transistors
  - nMOS and pMOS MOSFETS
  - Voltage applied to insulated gate controls current between source and drain
  - Low power allows very high integration
  - High speed switching applications
  - Operated by low current level.

www.rejinpaul.com

NMOS Transistor

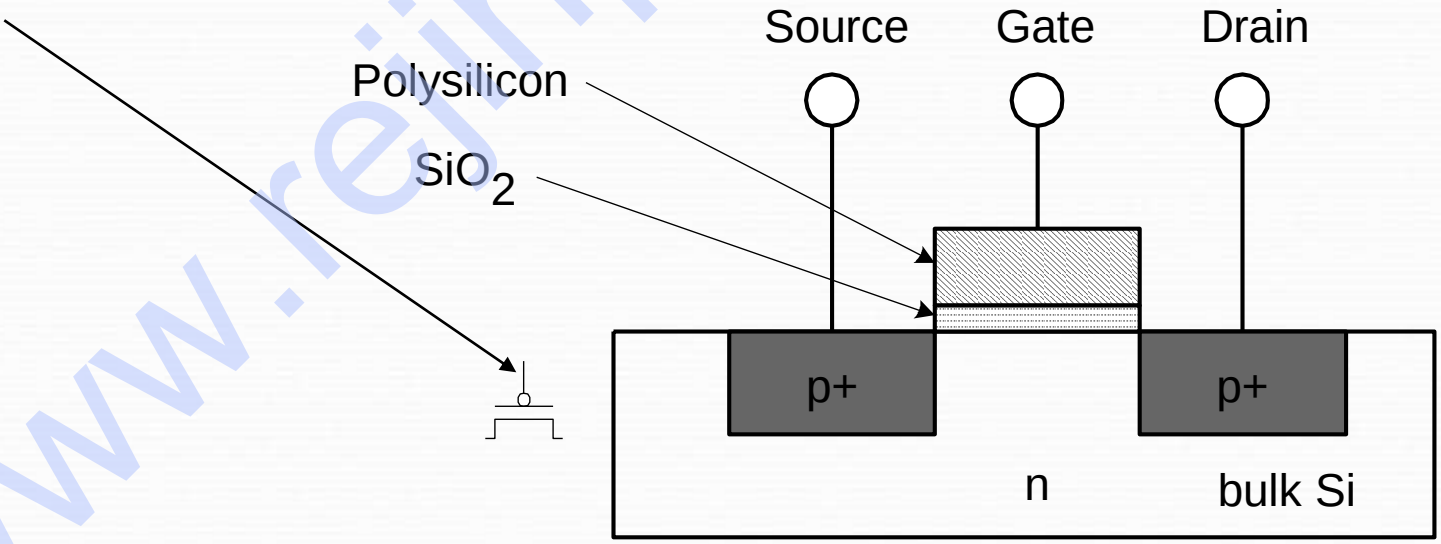
- Four terminal device: gate, source, drain, body
- Gate – oxide – body stack looks like a capacitor
  - Gate and body are conductors (body is also called the substrate)
  - SiO_2 (oxide) is a "good" insulator (separates the gate from the body
  - Enhancement mode→Apply +ve voltage for turn ON
  - Depletion mode→small –ve vol for turn ON the gate



www.rejinpaul.com

PMOS Transistor

- Similar, but doping and voltages reversed
  - Body tied to high voltage (V_{DD})
  - Drain is at a lower voltage than the Source
  - Gate low: transistor ON
  - Gate high: transistor OFF
  - Bubble indicates inverted behaviour.



www.rejinpaul.com

Modes of Operation NMOS

1)Accumulation Mode

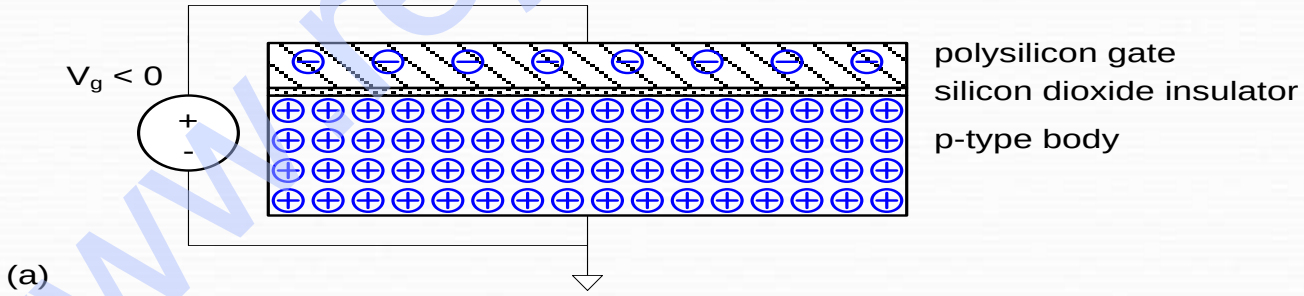
2)Depletion Mode

3)Inversion Mode

www.rejinpaul.com

1)Accumulation Mode

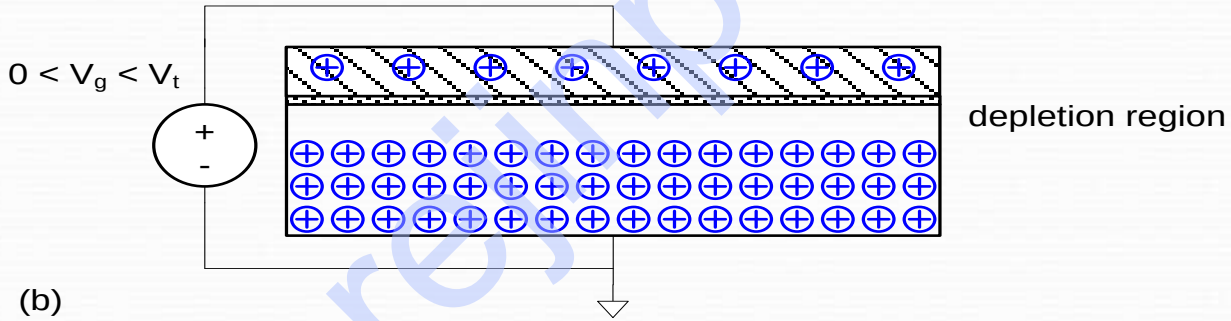
- Body is commonly tied to ground (o V)
- Drain is at a higher voltage than Source
- Negative voltage is applied to gate($V_{gs} << V_t$) .
- Positively charged holes are attracted to the region below the gate.
  - P-type body is at low voltage
  - Source-body and drain-body "diodes" are OFF
  - No current flows, transistor is OFF



www.rejinpaul.com

2) Depletion Mode

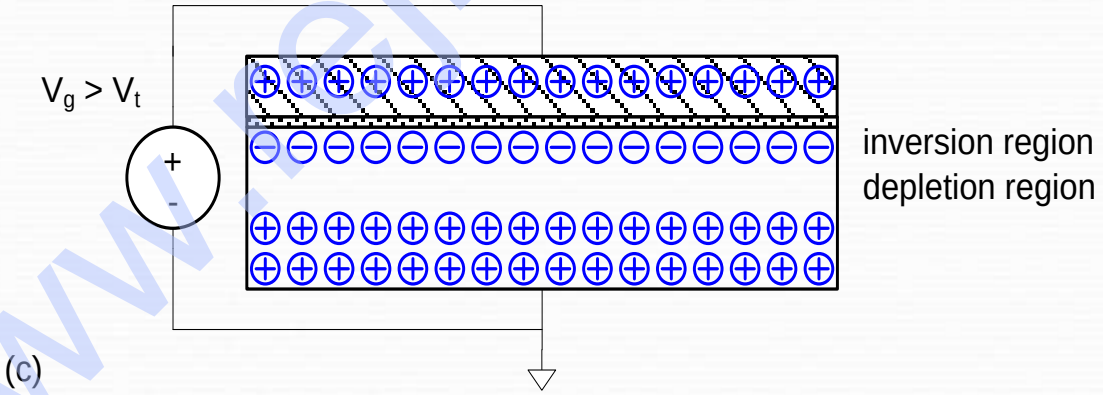
- Low positive voltage is applied to gate,some +ve charge on the gate.
- When $V_{gs}>0$ → Electrons attracted towards gate and holes are repelled.



www.rejinpaul.com

3)Inversion Mode

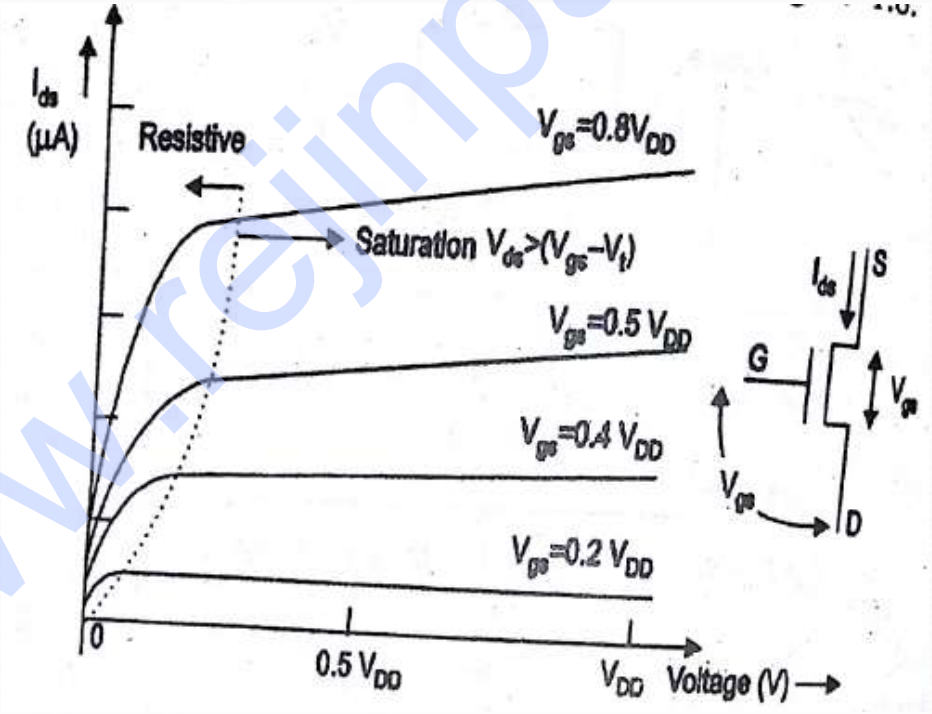
- When the gate is at a high voltage $V_{gs} >> V_t$
  - Negative charge is attracted to body under the gate
  - Inverts a channel under gate to "n-type" (N-channel, hence called the NMOS)
  - Now current can flow through "n-type" silicon from source through channel to drain, transistor is ON



www.rejinpaul.com

VI Characteristics

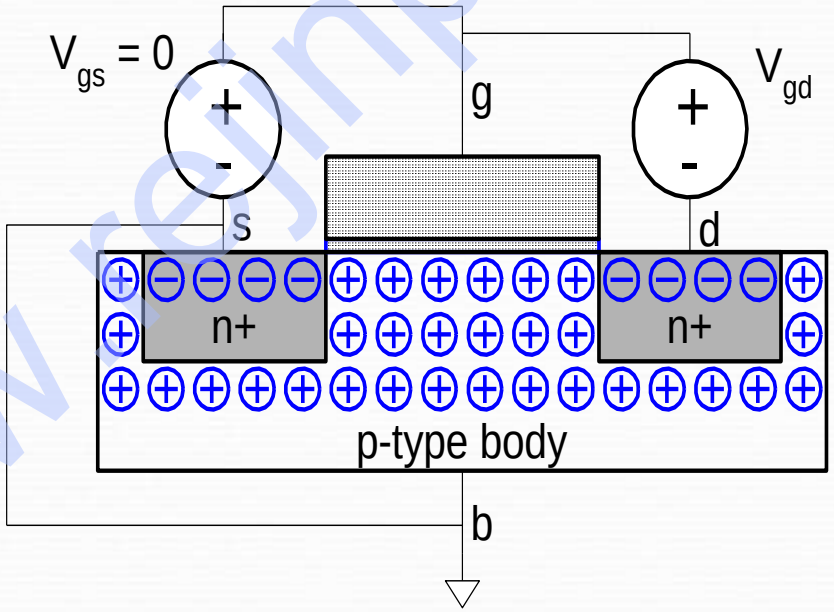
1. Cut-off region
2. Non-saturated region
3. Saturated region



www.rejinpaul.com

1)Cut-off region

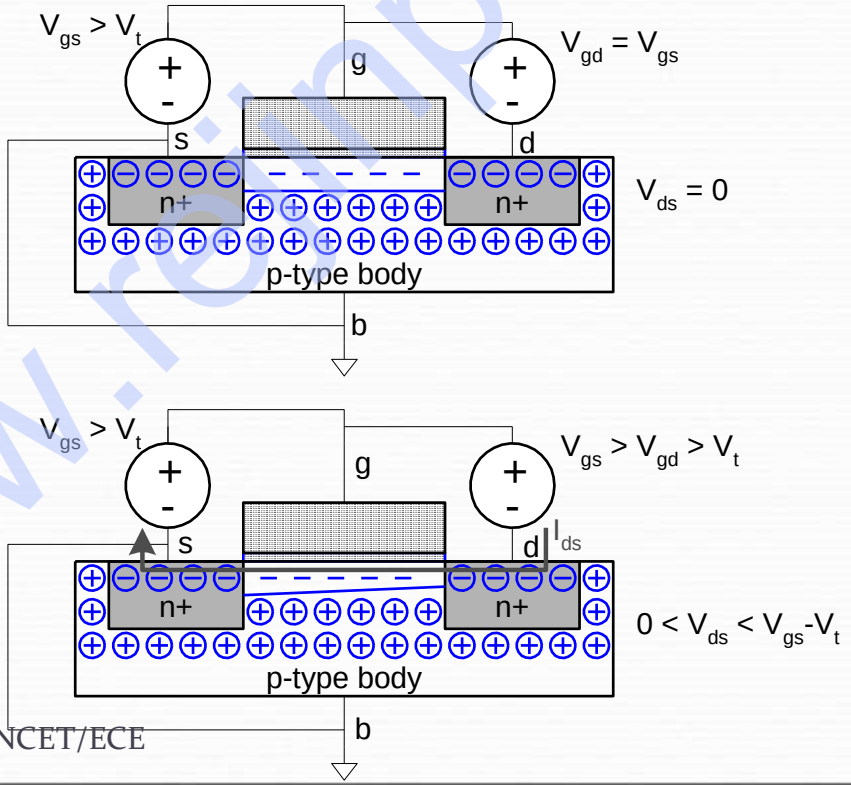
- When $V_{gs}<V_t$,the body has free holes but no free electrons.
- The junction between the body and the source or drain is reverse biased.
- Hence no current will flow
- $I_{ds} \approx 0$



www.rejinpaul.com

2)Non-saturated region

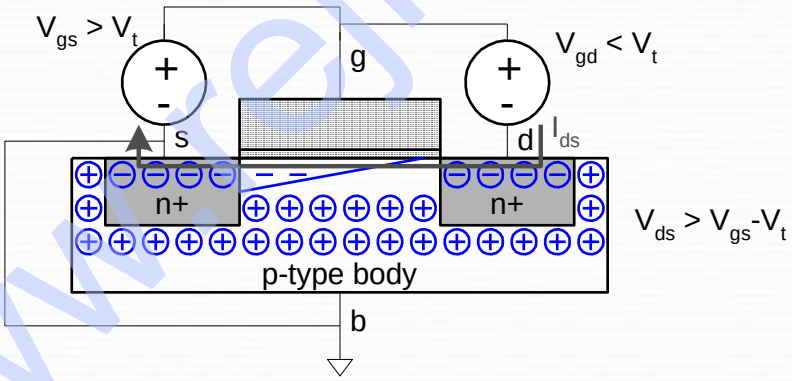
- Case (i) When $V_{gs}>V_{t,}$ there is a conductive path is created between source and drain.
- The number of carriers and conductivity increases with the gate voltage.
- The potential difference between drain and source is $V_{ds}=V_{gs}-V_{gd}$
- Case (ii) Small positive voltage ($o<V_{ds}<V_{gs}-V_t$) is applied I_{ds} flow drain to source.



www.rejinpaul.com

3) Saturated region

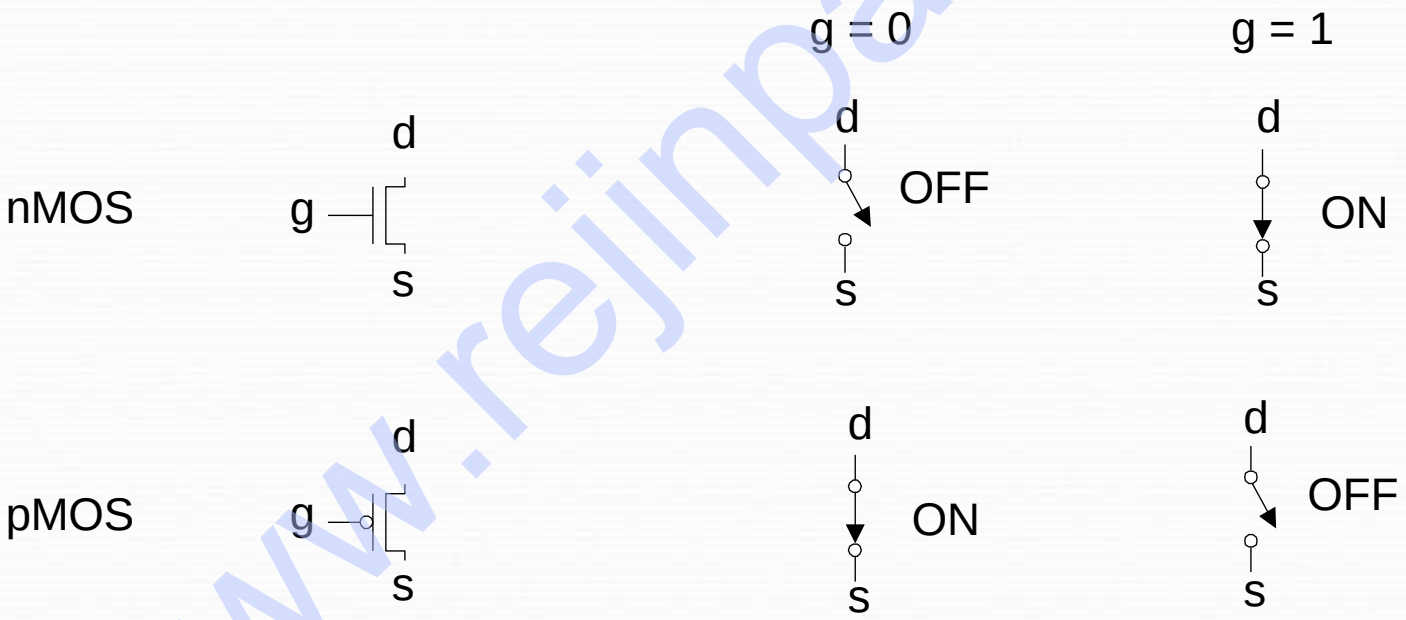
- When $V_{gs}>V_t$ channel is no longer inverted near the drain and becomes pinchedoff.
- Electrons are drifted under the poisitive drain voltage.
- As electrons reach the end of the channel ,they are injected into the depletion region and accelerated towards the drain.



www.rejinpaul.com

Transistors as Switches

- We can view MOS transistors as electrically controlled switches
- Voltage at gate controls path from source to drain



www.rejinpaul.com

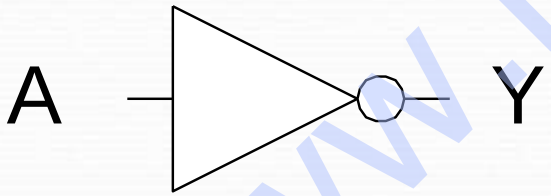
CMOS LOGIC

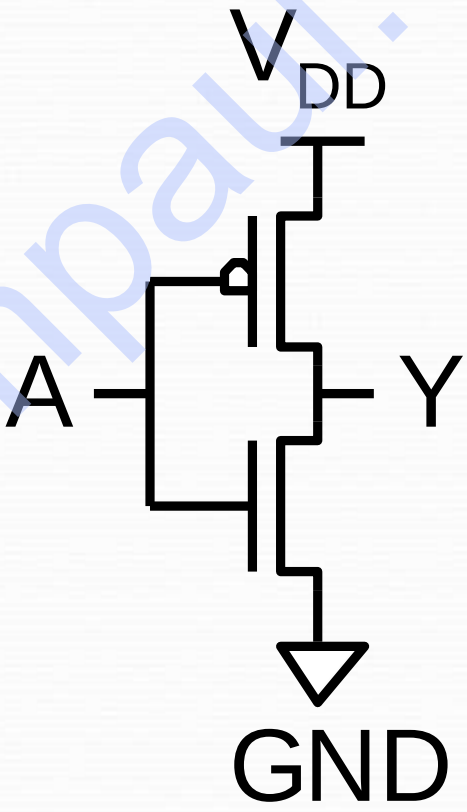
- Inverter
- NAND Gate
- NOR gate

www.rejinpaul.com

1)CMOS Inverter

A	Y
0	
1	

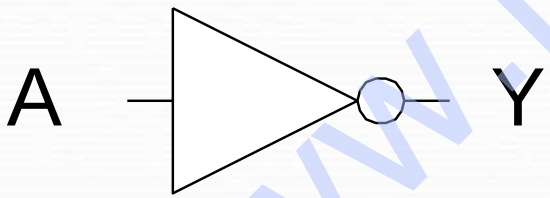


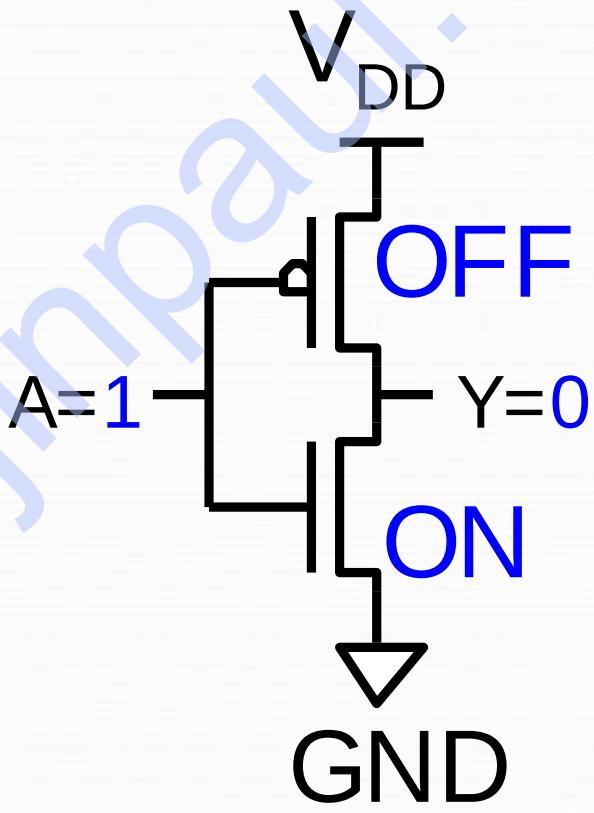


www.rejinpaul.com

1)CMOS Inverter

A	Y
0	
1	**0**



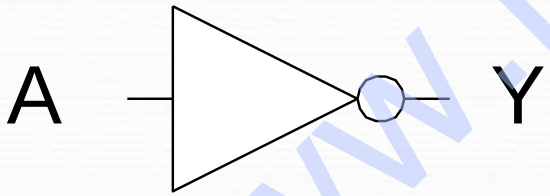


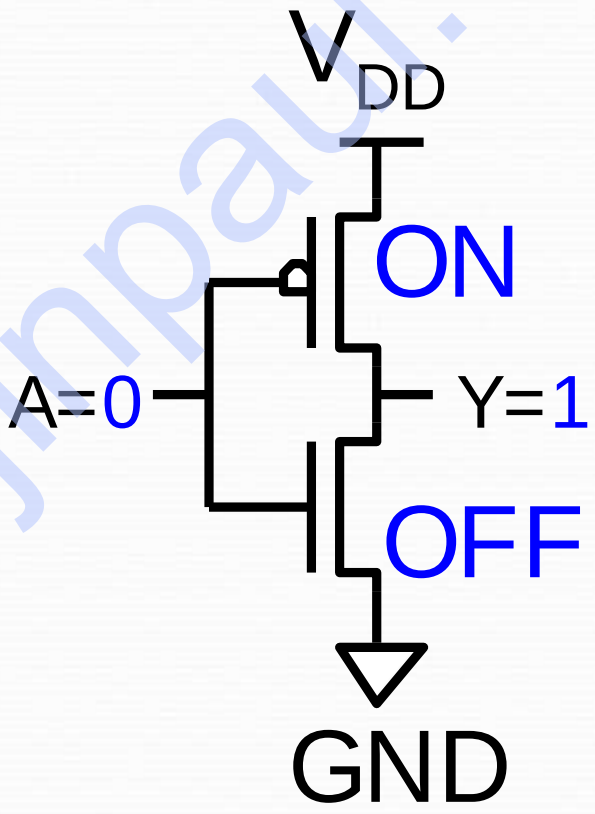
Y is pulled low by the turned on NMOS Device. Hence NMOS is the pull-down device.

www.rejinpaul.com

1)CMOS Inverter

A	Y
0	1
1	0



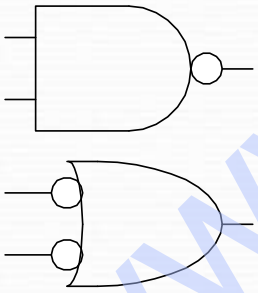


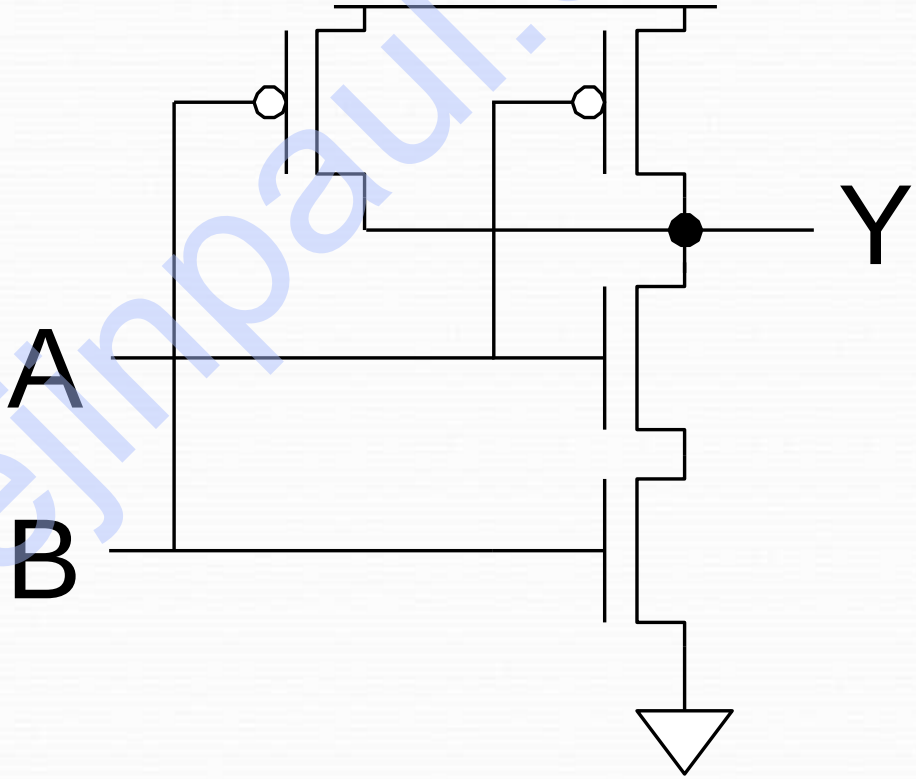
Y is pulled high by the turned on PMOS Device. Hence PMOS is the pull-up device.

www.rejinpaul.com

2) CMOS NAND Gate

A	B	Y
0	0	
0	1	
1	0	
1	1	

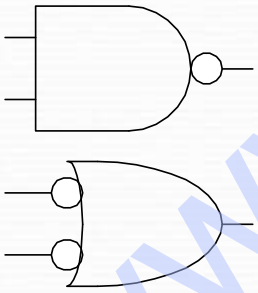


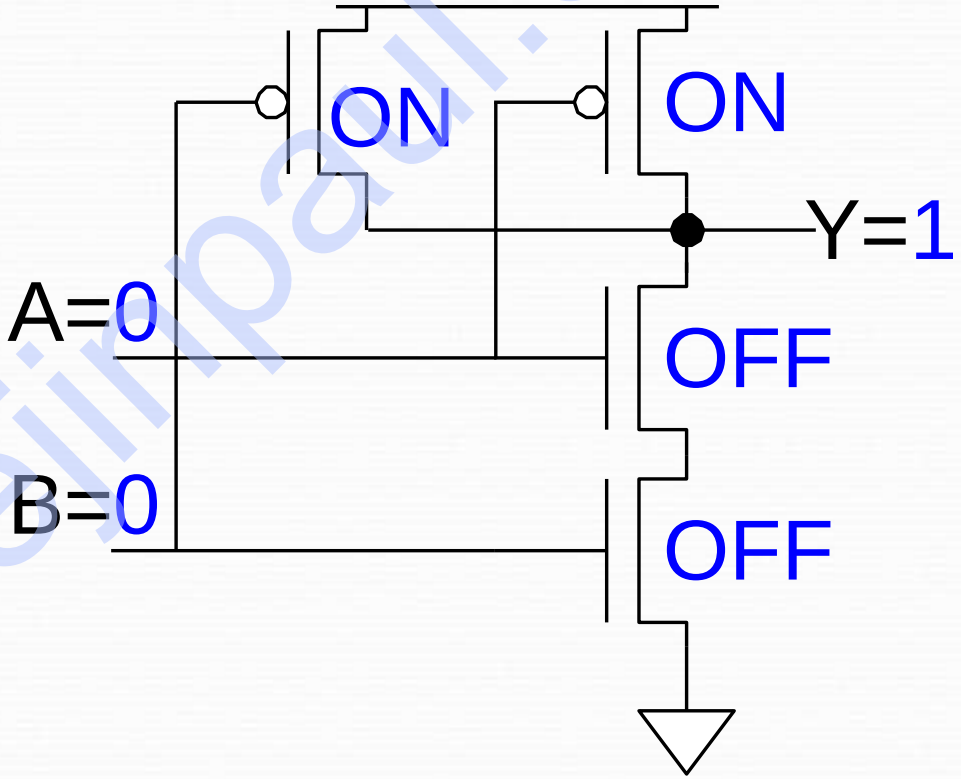


www.rejinpaul.com

2) CMOS NAND Gate

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

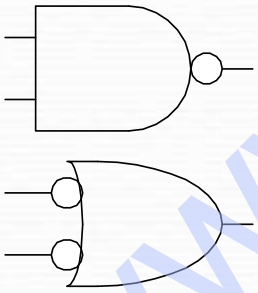


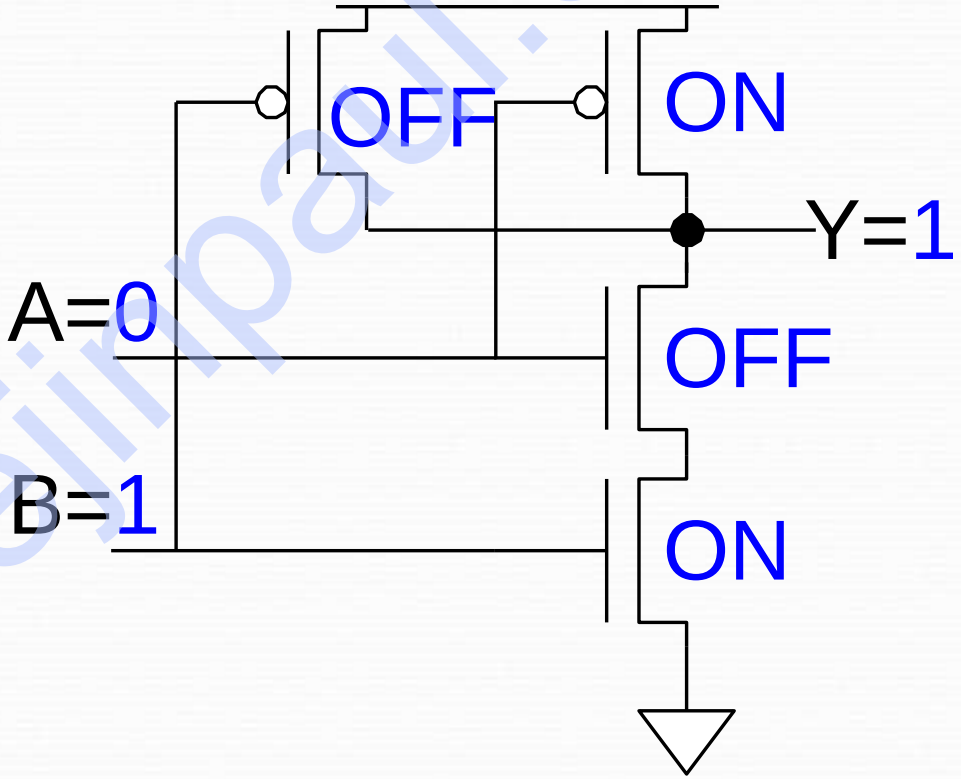


www.reijinpaul.com

2) CMOS NAND Gate

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

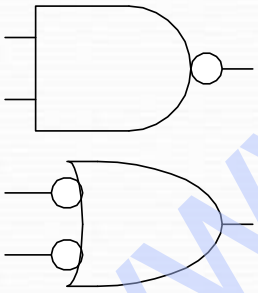


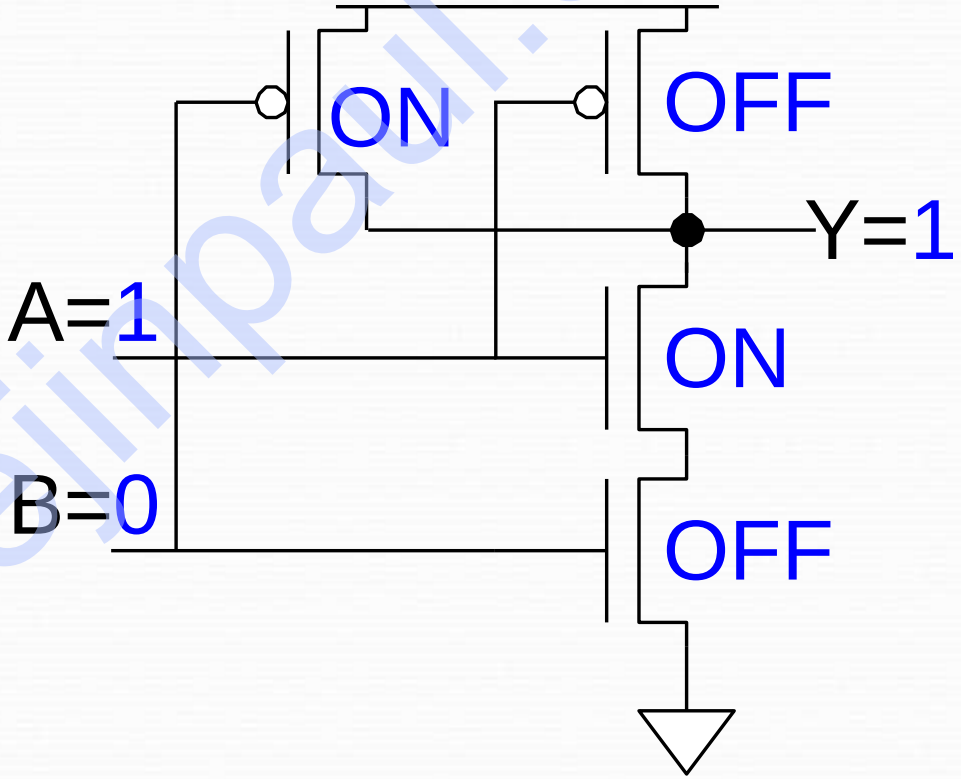


www.reijinpaul.com

2) CMOS NAND Gate

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

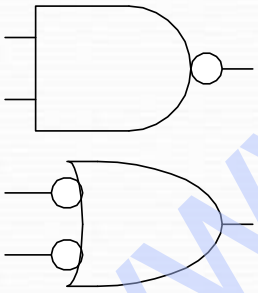


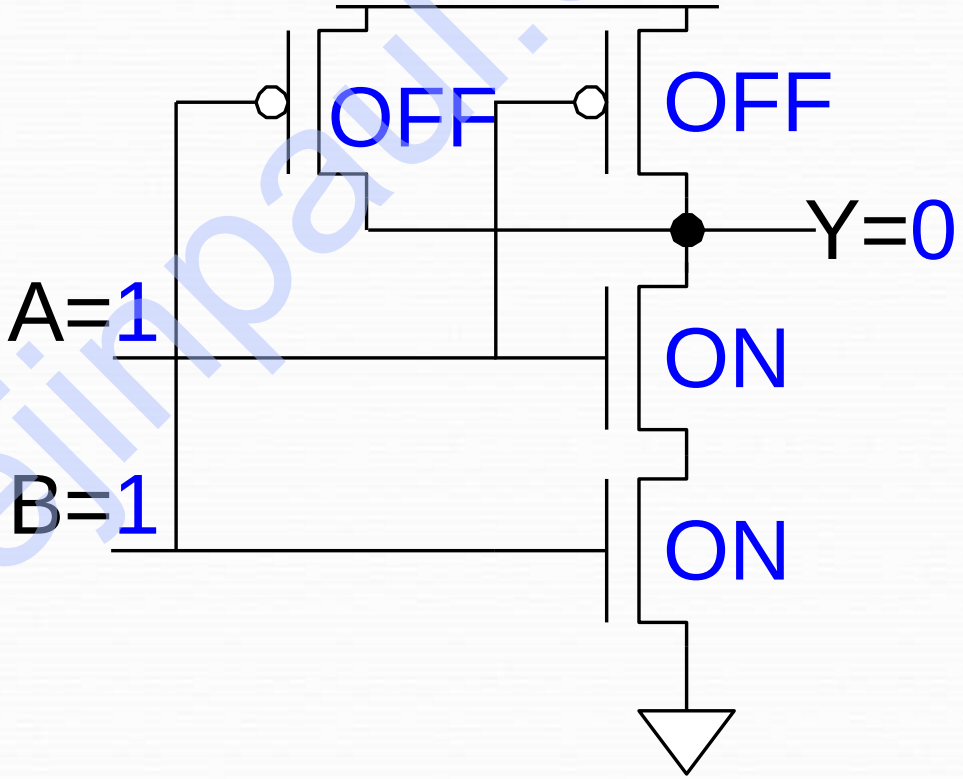


www.reijinpaul.com

2) CMOS NAND Gate

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

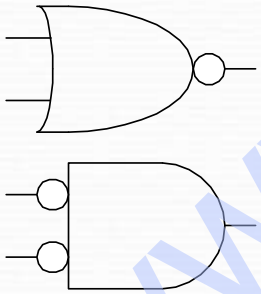


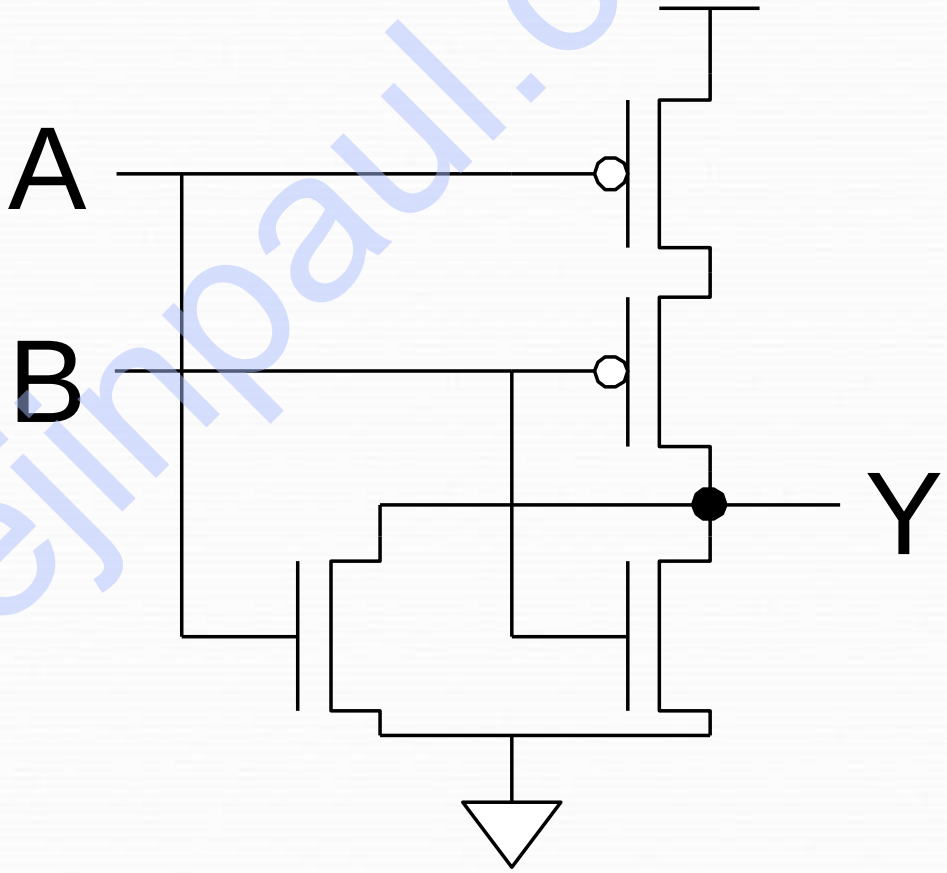


www.reijinpaul.com

3) CMOS NOR Gate

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





www.rejinpaul.com

Electrical Properties of CMOS

1. Ideal I-V characteristics
2. C-V Characteristics
3. Non Ideal Characteristics
4. DC Transfer Characteristics
5. Device Models

www.rejinpaul.com

1)Ideal I-V characteristics

1. Cut-off or Sub-threshold mode→When $V_{gs}<V_{t,}$ I_{ds}=o,then transistor is turned off.No conduction takes place between drain and source.
2. Linear or non saturation mode→When Vgs>Vt and Vds<(Vgs)→Transistor turned on.Channel or conduction path is created which allows current flow from drain-source.
3. Saturation mode→When $V_{gs}>V_t$ and $V_{ds}>(V_{gs}-V_t)$→Switch is turned on.Channel is created which allows current flow between drain and source.If drain voltage is higher than the gate voltage,portion of channel is turned off.Hence the drain current is independant with drain voltage and current is controlled by V_{gs}

www.rejinpaul.com

Ideal Transistor I-V

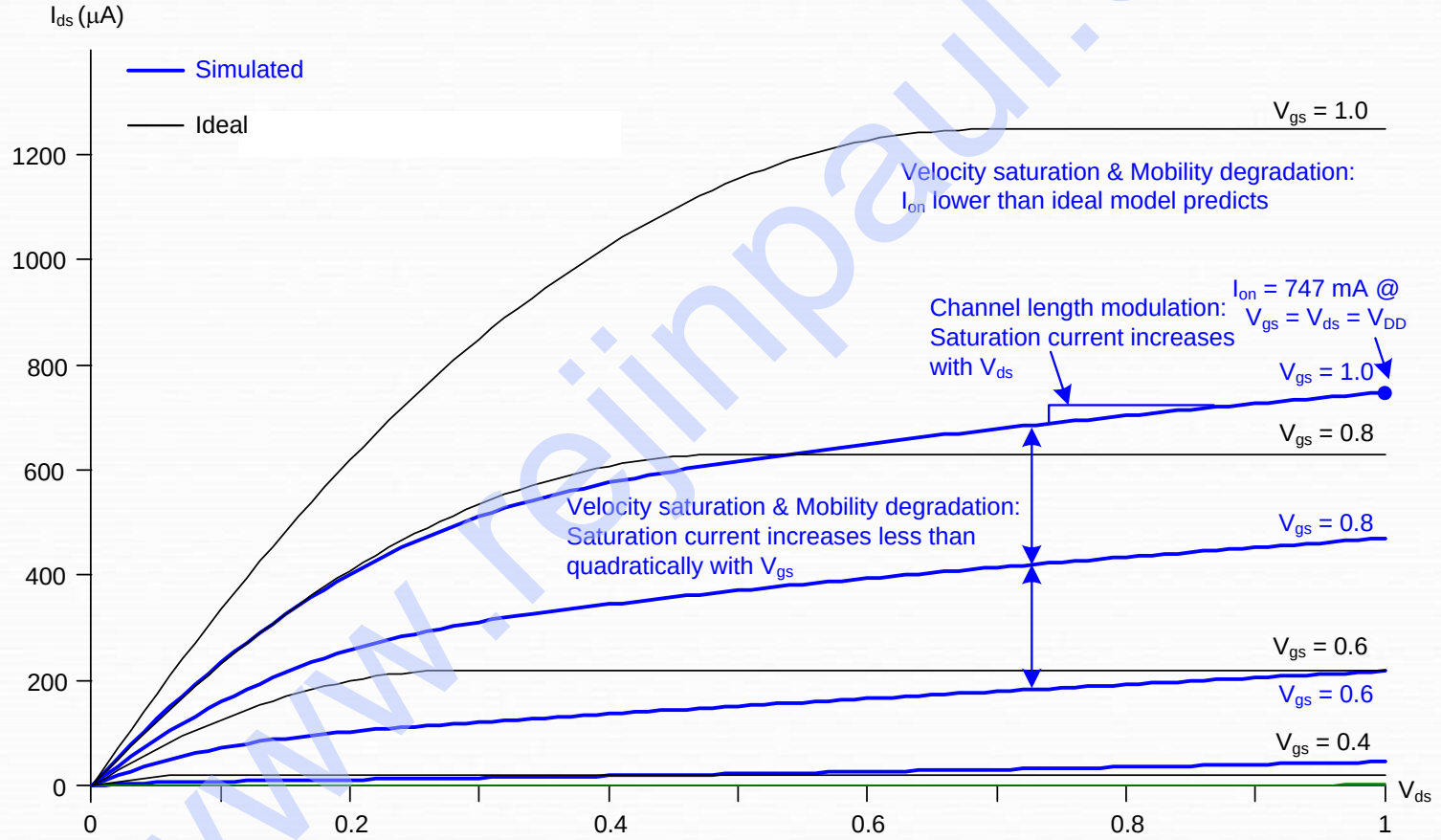
- *Shockley* long-channel transistor models

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{cutoff} \\ \beta\left(V_{gs} - V_t - V_{ds}/2\right)V_{ds} & V_{ds} < V_{dsat} \quad \text{linear} \\ \frac{\beta}{2}\left(V_{gs} - V_t\right)^2 & V_{ds} > V_{dsat} \quad \text{saturation} \end{cases}$$

www.rejinpaul.com

Ideal vs. Simulated nMOS I-V Plot

- 65 nm IBM process, V_{DD} = 1.0 V



www.rejinpaul.com

2)C-V Characteristics

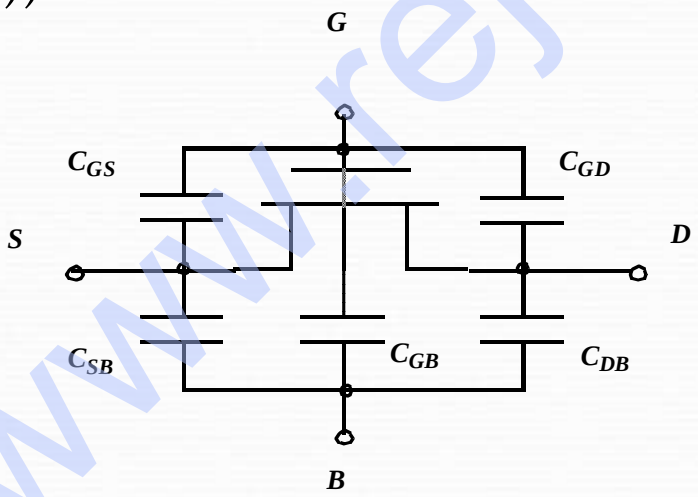
- Switching speed of MOS system depends on

1) Parasitic capacitance(or)Gate capacitance

2) Interconnecting wires

3) Resistance of transistors and wires

- **Total capacitance of MOS system**=(Gate capacitance + diffusion(or)depletion capacitance + routing capacitance(other wires))



www.rejinpaul.com

Simple MOS capacitance Model

- Each terminal of a MOS has capacitance to other terminal.
- Gate of MOS transistor is a good capacitor. Capacitance is needed to attract charge to invert the channel.
- This capacitance has no fundamental operation on the device but it has some impact on ckt performance→parasitic capacitance

$$C_g = \frac{\varepsilon_{ins}\,\varepsilon_0}{t_{ox}} WL$$

$$C_g = C_{ox} WL$$

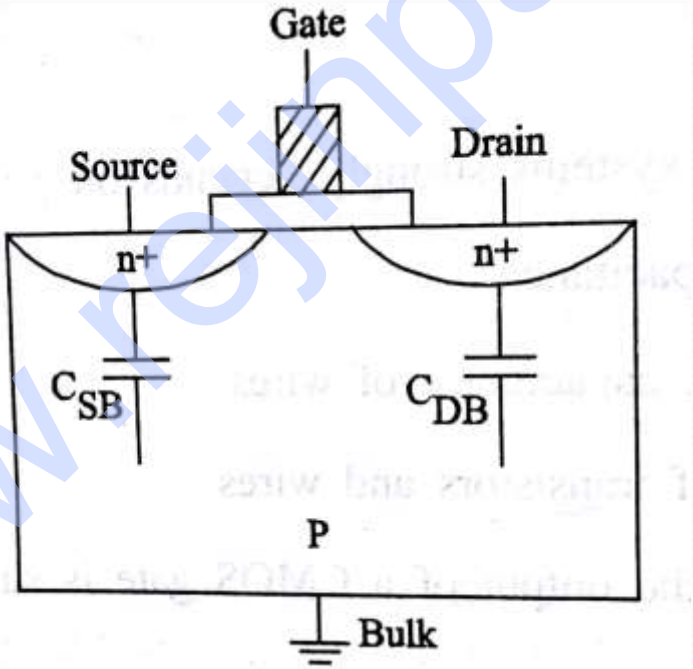
$$C_g = C_{permicron} \cdot W$$

$$C_{permicron} = C_{ox} \times L = \frac{\varepsilon_{ins} \cdot \varepsilon_0}{t_{ox}} L$$

www.rejinpaul.com

Diffusion Capacitance

- In addition to the gate, the source and drain also has capacitance.
- Due to reverse bias PN Junction
- These capacitance are not fundamental to operation of the devices, but do impact circuit performance and hence this is called as parasitic capacitors



www.rejinpaul.com

POSSIBLE LAYOUT SCHEME:

Normally the diffusion capacitors reduced by possible layout scheme

- Isolated Contacted Diffusion
- Shared Contacted Diffusion
- Merged uncontacted Diffusion

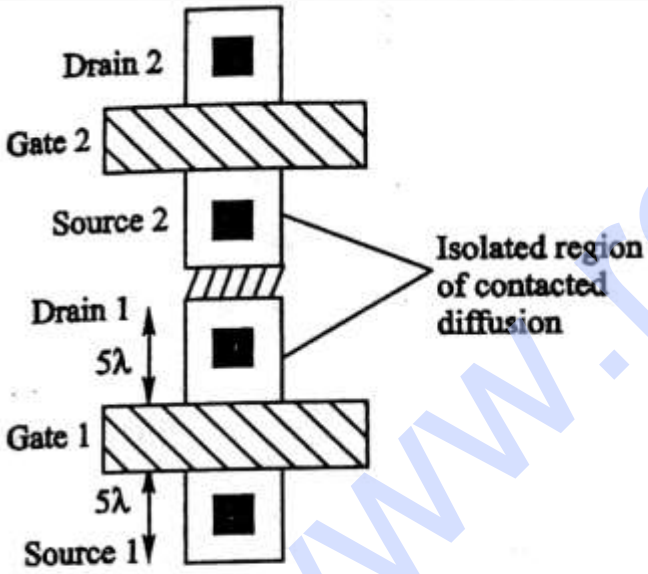


Figure 1.18 Isolated Contacted Diffusion

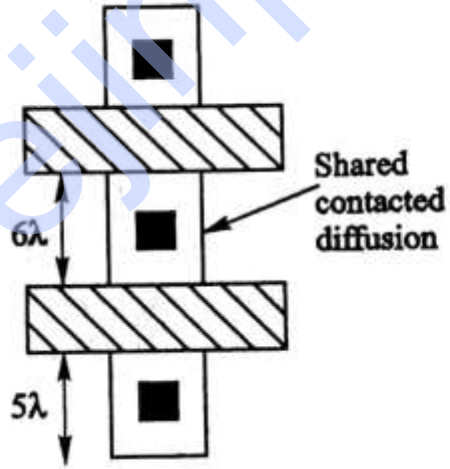


Figure 1.19 Shared Contacted Diffusion

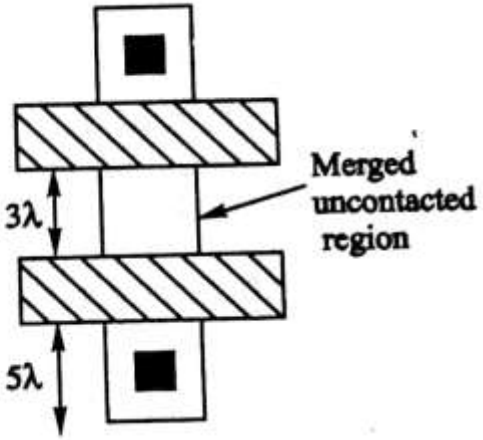


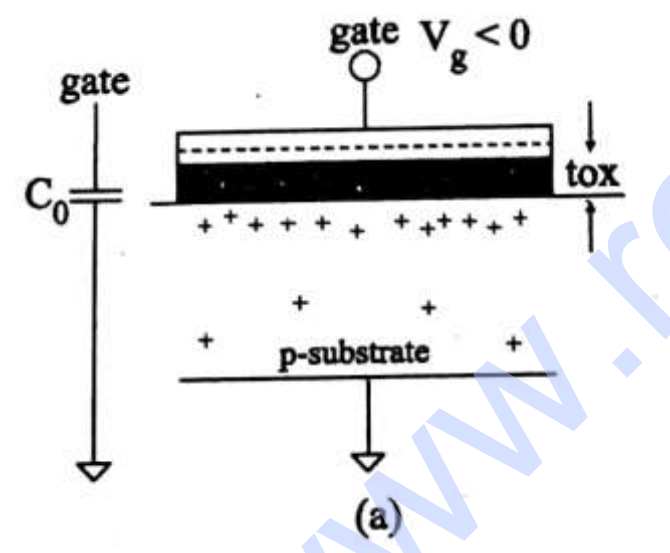
Figure 1.20 Merged Uncontacted Diffusion

www.rejinpaul.com

Detailed MOS Gate Capacitance Model

- Accumulation
- Depletion
- Inversion

1.Accumulation:



$$C_g = C_0 = \left(\frac{\varepsilon_{ins}\, \varepsilon_0}{t_{ox}} \right) \cdot A$$

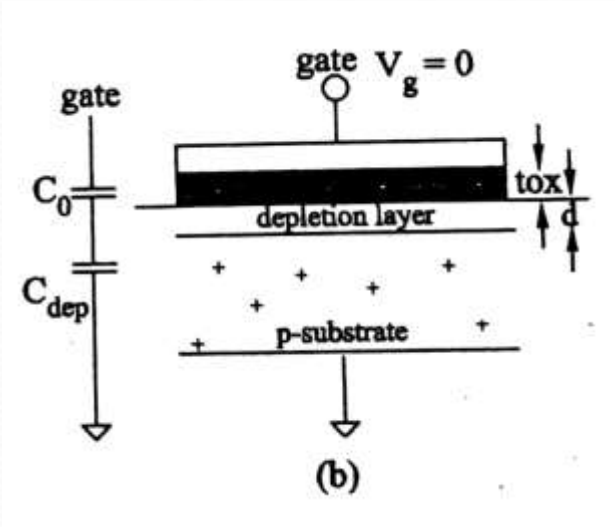
where, A= area of gate

ε_{ins} = dielectric constant of SiO_2 taken as 3.9

ε_0 = permittivity of free space

www.rejinpaul.com

2.Depletion Layer



$$C_{dep} = \left(\frac{\varepsilon_0 \varepsilon_{si}}{d} \right) A$$

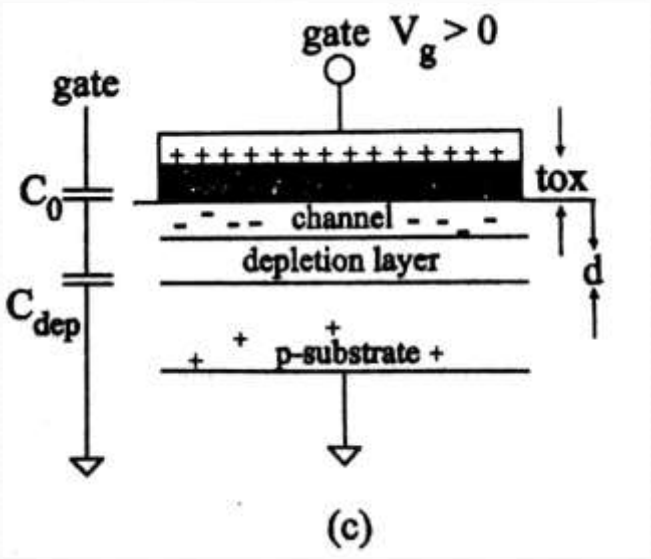
where, d = depletion layer depth

ε_{si} = dielectric constant of silicon, taken as 12

$$C_{gb} = \frac{C_0 \cdot C_{dep}}{C_0 + C_{dep}}$$

www.rejinpaul.com

3.Inversion Layer



$$C_{gb} = C_0 \text{ low frequency}$$

$$C_{gb} = \frac{C_0 C_{dep}}{C_0 + C_{dep}} \text{ high frequency}$$

www.rejinpaul.com

MOS Device Capacitance:

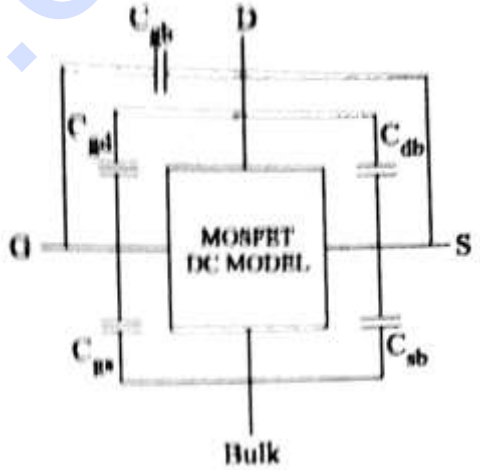
1. Oxide-related capacitances
2. Junction capacitances

The gate-oxide related capacitance are

- C_{gd} = gate-to-drain capacitance
- C_{gs} = gate-to-source capacitance
- C_{gb} = gate-to-substrate capacitance

C_{sb} = source-to-substrate capacitance

C_{db} = drain-to-substrate capacitance

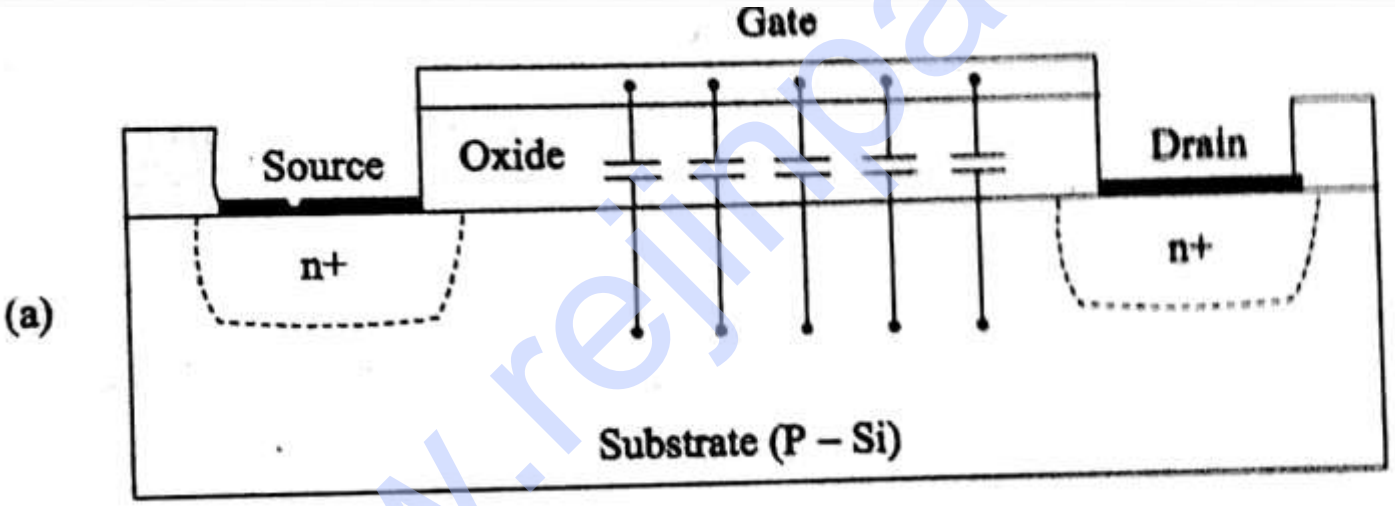


$$C_g = C_{gb} + C_{gs} + C_{gd}$$

www.rejinpaul.com

- The behaviour of the gate capacitance of a MOS device are explained in terms of its three operating regions
- Cut off regions:

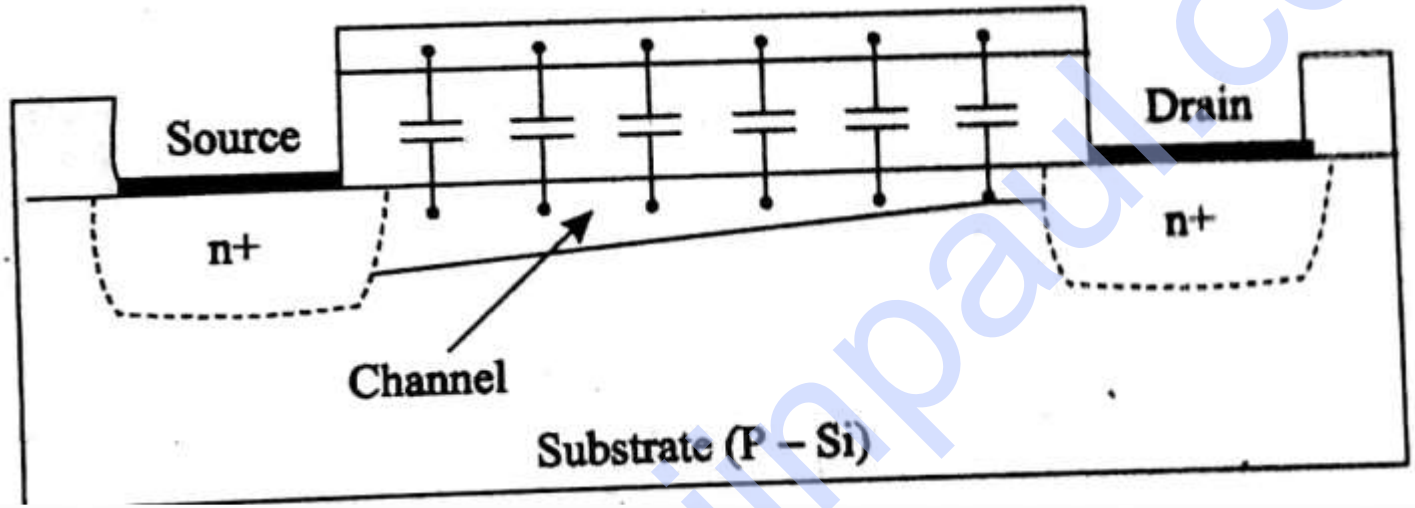
$V_{gs} < V_t$, MOS device is OFF, there is no channel hence $C_{gs} = C_{gd} = 0$.



(a)

www.rejinpaul.com

Linear Region

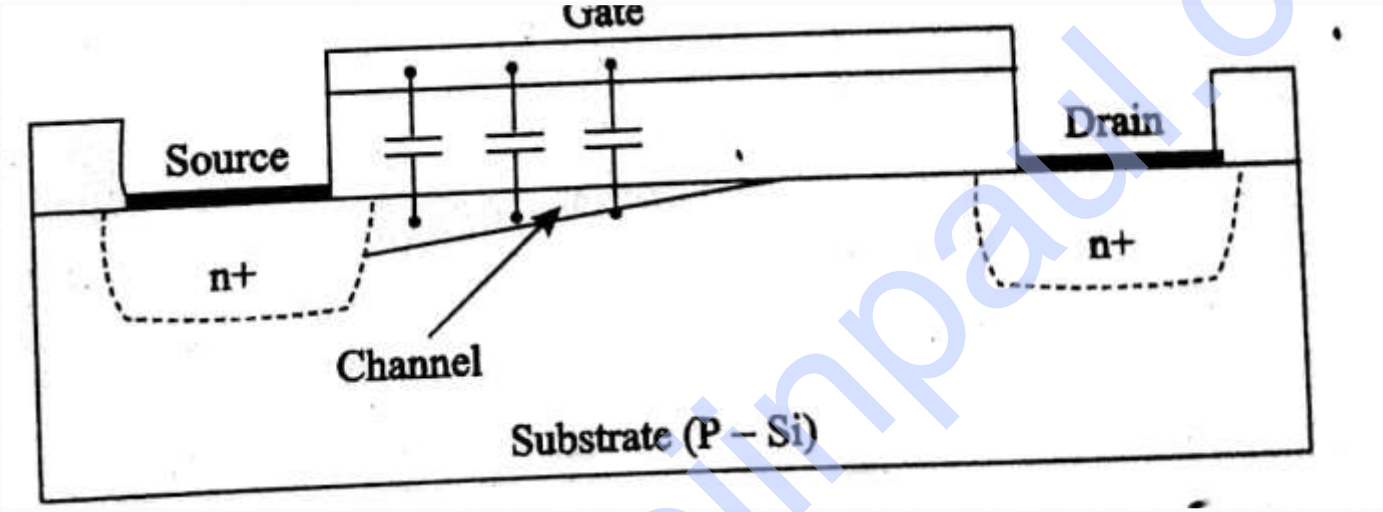


$$0 < V_{ds} < V_{gs} - V_t$$

$$C_{gd} = C_{gs} = \frac{1}{2}\left(\frac{\varepsilon_0 \varepsilon_{ox}}{t_{ox}}\right) \cdot A$$

www.rejinpaul.com

Saturation region



$$0 < V_{gs} - V_t < V_{ds}$$

$$C_{gs} = \frac{2}{3}\left(\frac{\varepsilon_0\, \varepsilon_{ins}}{t_{ox}}\right) A$$

www.rejinpaul.com

Approximation of Intrinsic MOS gate Capacitance

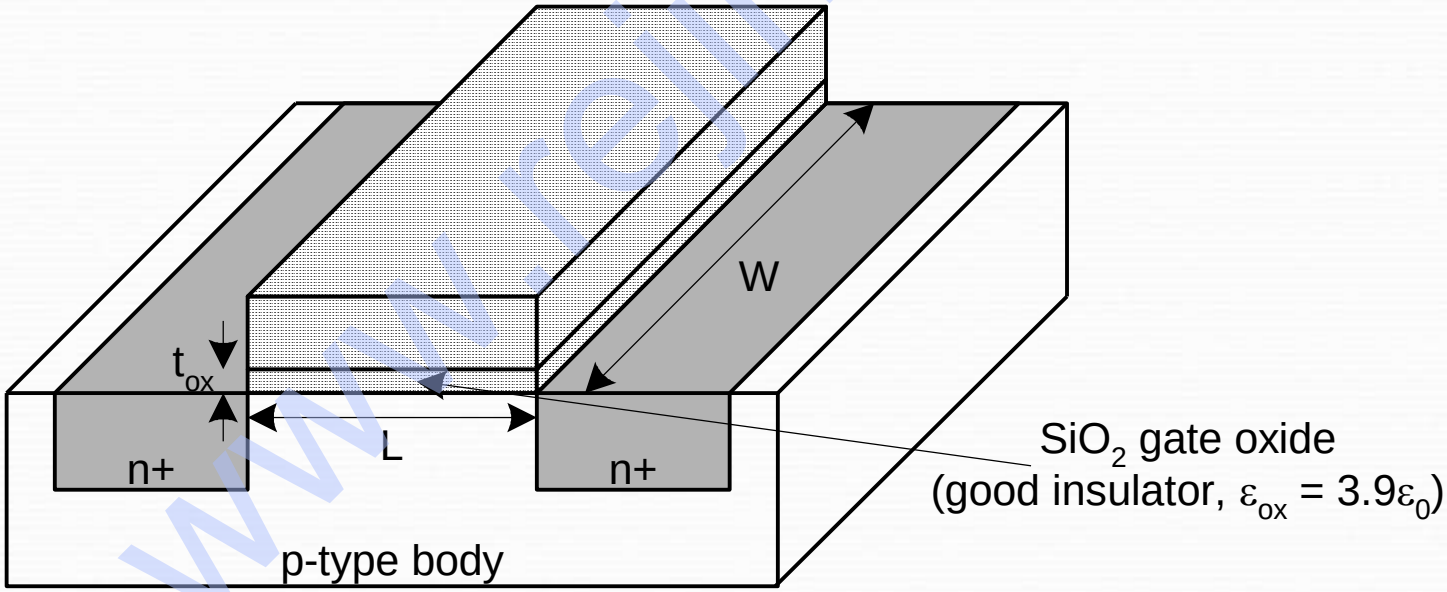
...intrinsic MOS gate capacitance

Capacitance			
Parameter	Cut-off	non-saturated	saturated
C_{gb}	$\frac{\varepsilon A}{t_{ox}}$	0	0
C_{gs}	0	$\frac{\varepsilon A}{2t_{ox}}$	$\frac{2\varepsilon A}{3t_{ox}}$
C_{gd}	0	$\frac{\varepsilon A}{2t_{ox}}$	0
$C_g = C_{gb} + C_{gs} + C_{gd}$	$\frac{\varepsilon A}{t_{ox}}$	$\frac{\varepsilon A}{t_{ox}}$	$\frac{2\varepsilon A}{3t_{ox}}$

www.rejinpaul.com

1)Gate Capacitance

- Gate of MOS transistor is a good capacitor. Capacitance is needed to attract charge to invert the channel.
- This capacitance has no fundamental operation on the device but it has some impact on ckt performance→parasitic capacitance
- Gate capacitance is given by
- $C_{gs} = \varepsilon_{ox}WL/t_{ox} = C_{ox}WL = C_{permicron}W$
- Total Gate capcitance$C_g=C_{gs}+C_{gb}+C_{gd}$



www.rejinpaul.com

Gate Capacitance Behaviour

1)Cut-off region→$V_{gs}<V_T$(device off.no conduction,hence $C_{gs}=C_{gd}=0$)

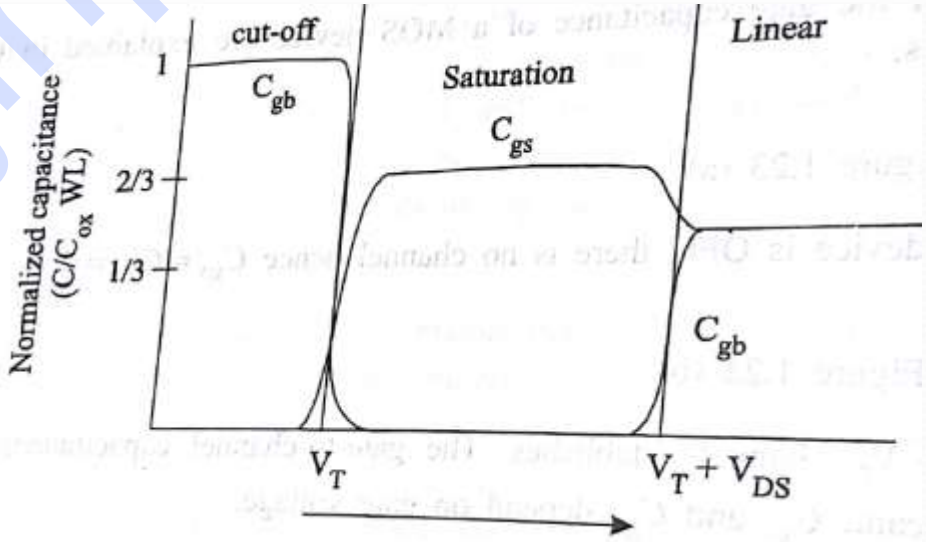
2) Linear region→$0<V_{ds}<V_{gs}-V_T$(Channel created.gate-channel capacitance depends on gate voltage)

$$C_{gd}=C_{gs}=\frac{1}{2}\left(\frac{\varepsilon_0\,\varepsilon_{ins}}{t_{ox}}\right)\cdot A$$

3) Saturated region→$0<V_{gs}-V_T<V_{ds}$

- Channel is inverted.
- Drain channel goes to off.

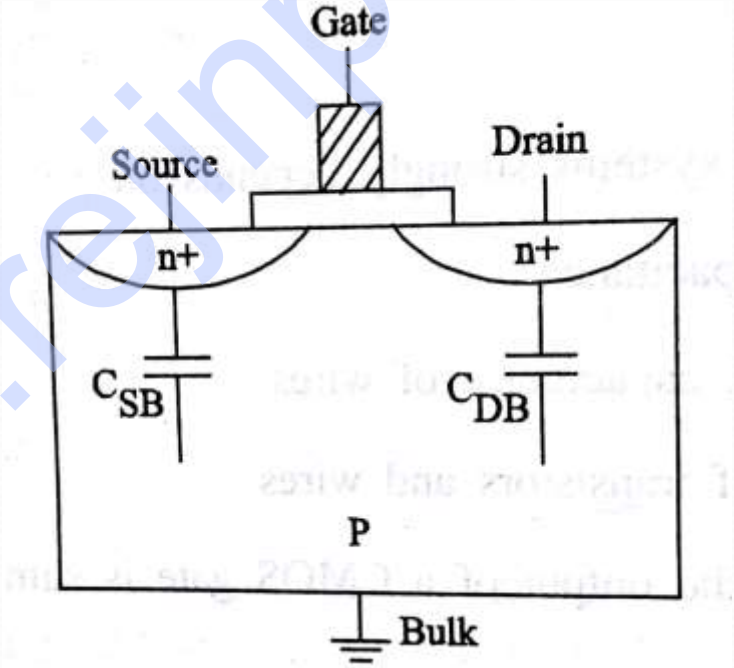
$$C_{gs}=\frac{2}{3}\left(\frac{\varepsilon_0\,\varepsilon_{ins}}{t_{ox}}\right)A$$



www.rejinpaul.com

2)Diffusion Capacitance

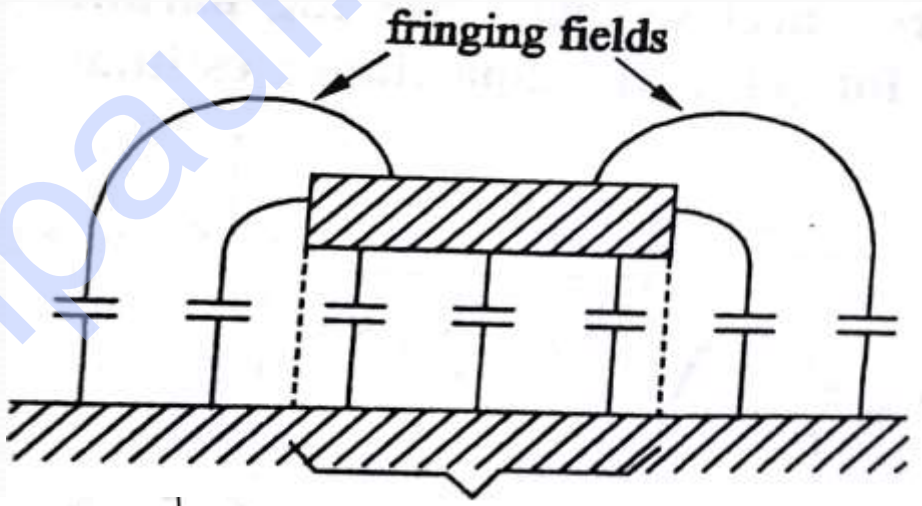
- Source and drain terminal have capacitance
- At the reverse biased condition due to presence of pn-junction between source-drain or body.
- Associated capacitances C_{sb} and C_{db}
- It has high C & low R.



www.reinpaul.com

3)Routing Capacitance

- Routing capacitance between metal and poly layers
- Each interconnection line or wire has shape, thickness and distance from the ground plane



Types

- Metal-metal line
- Metal line-ground
- Metal line-layers

$$C = \varepsilon \left[\frac{\left(w - \frac{t}{2}\right)}{h} + \frac{2\pi}{\ln\left[1 + \frac{2h}{t} + \sqrt{\frac{2h}{t} + \left(\frac{2h}{t} + 2\right)}\right]} \right] \quad \text{for } w \geq t/2$$

www.rejinpaul.com

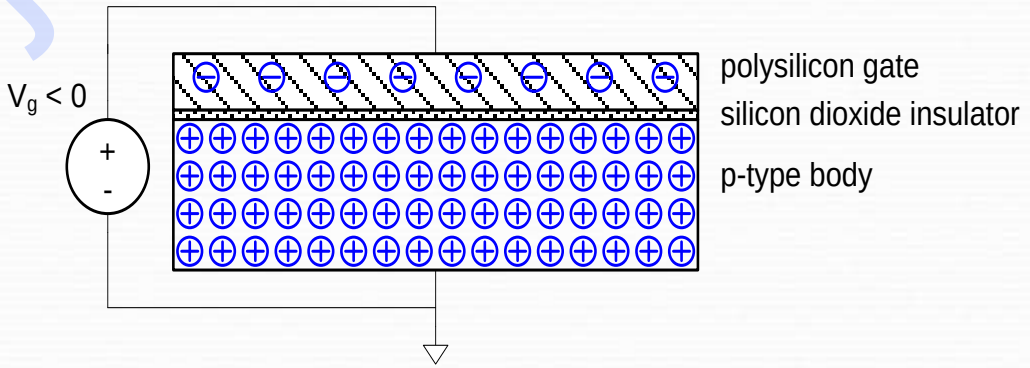
MOS Capacitance model

- C-V characteristics of a MOS capacitor depend on the state of the semiconductor surface.

1)Accumulation layer

- Layer is formed when $\mathbf{V_{gs} < V_t}$
- Gate act as one plate & p-substrate act as second plate.
- Parallel plate gate capacitance formed

$$C_g = C_0 = \left(\frac{\varepsilon_{ins}\, \varepsilon_0}{t_{ox}} \right) \cdot A$$



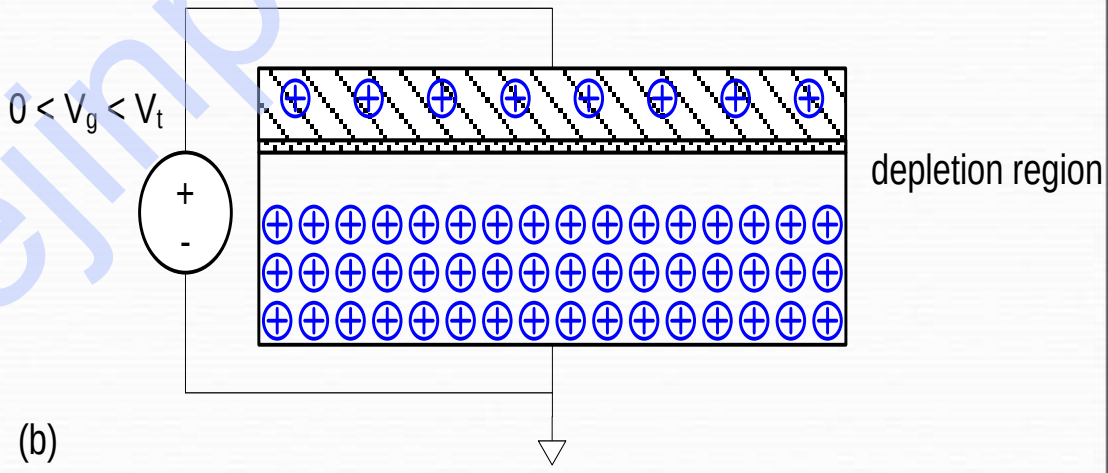
(a)

www.rejinpaul.com

2) Depletion layer

- When $V_{gs} > V_t$ → electrons attracted towards gate while holes are repelled hence depletion layer is formed under the gate.

$$C_{dep} = \left(\frac{\varepsilon_0 \varepsilon_{si}}{d} \right) A$$



(b)

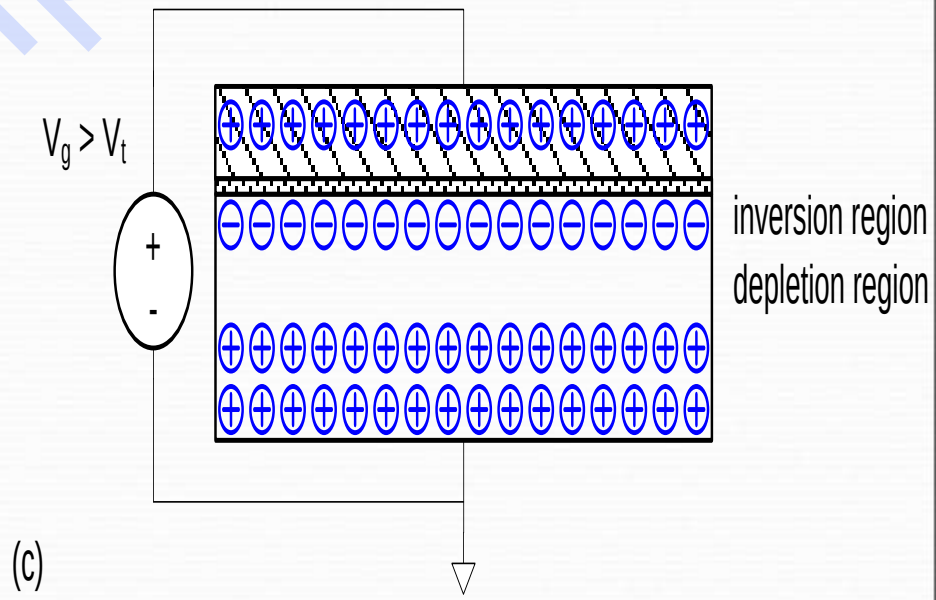
www.rejinpaul.com

3) Inversion Layer

- When $V_{gs} >> V_t$ → electrons are fully attracted towards gate region.
- As a result p-type substrate is fully occupied by n-type.(Inverted)

$$C_{gb} = C_0 \quad \text{low frequency}$$

$$C_{gb} = \frac{C_0 C_{dep}}{C_0 + C_{dep}} \quad \text{high frequency}$$



www.rejinpaul.com

3)Non Ideal I-V Characteristics

1)Mobility Degradation
2)Velocity Saturation
3)Channel Length Modulation
4)Threshold Voltage
5) Body Effect
6)Leakage
- Sub-threshold Leakage
- Gate Leakage
- Junction Leakage

7)Tunneling
8)Temperature variation

www.rejmpaul.com

3.1)Mobility Degradation

- Mobility of charge carriers varies by the following factor

1)High Temperature

2)Type of charge carriers

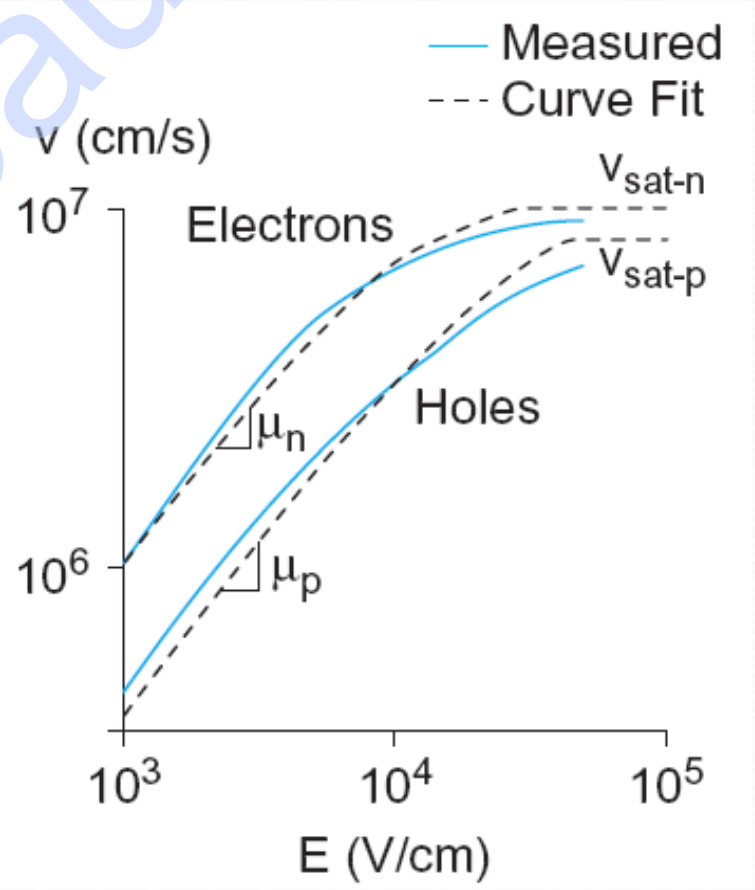
3)doping

www.rejinpaul.com

3.2)Velocity Saturation

- At high lateral field strength(E), carrier velocity Increased.
  - Carriers scatter off atoms in silicon lattice
  - Velocity reaches v_{sat}
    - Electrons: 10^7 cm/s
    - Holes: 8 x 10^6 cm/s

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



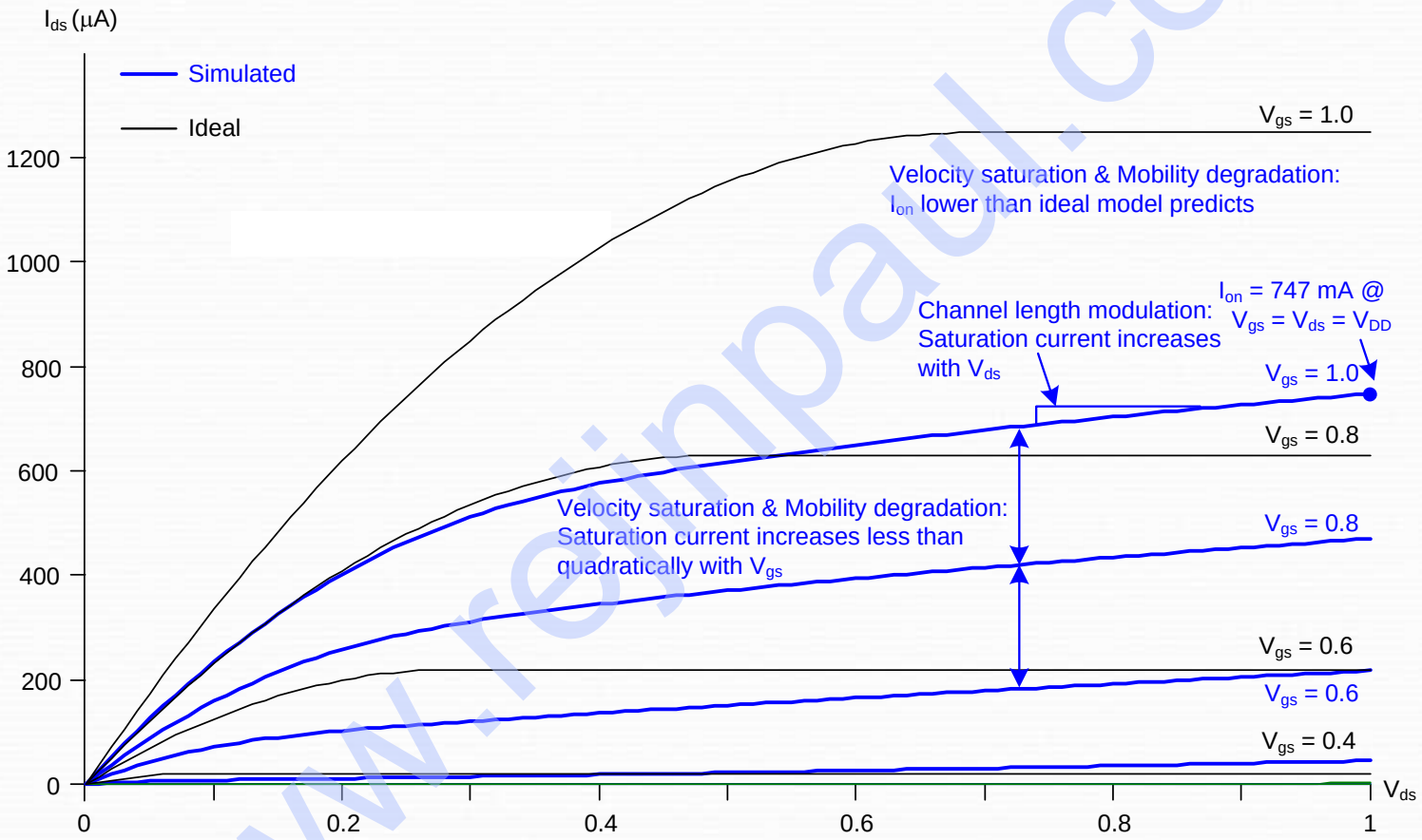
www.rejinpaul.com

- Drain current is dependent on voltage without velocity saturation and linearly dependant when fully velocity saturated.

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{cutoff} \\ \beta\left(V_{gs} - V_t - V_{ds}/2\right)V_{ds} & V_{ds} < V_{dsat} \quad \text{linear} \\ \frac{\beta}{2}\left(V_{gs} - V_t\right)^2 & V_{ds} > V_{dsat} \quad \text{saturation} \end{cases}$$

www.rejinpaul.com

- As channel length becomes shorter ,the lateral E-field Increases and transistors become more velocity saturated and supply voltage is constant



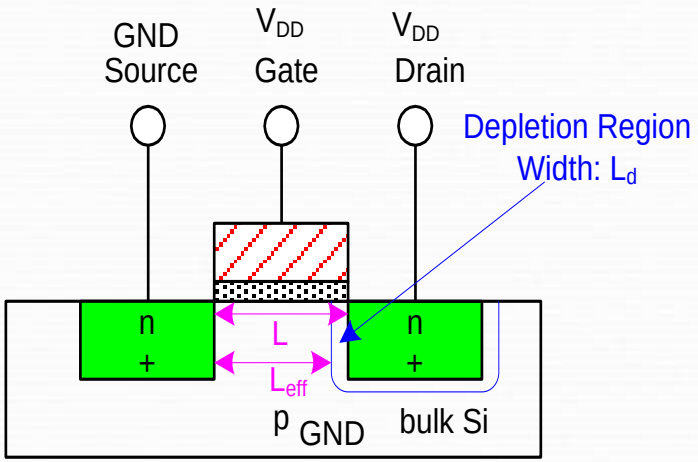
www.rejinpaul.com

3.3)Channel Length Modulation

- Reverse-biased p-n junctions form a *depletion region*
  - Region between n and p with no carriers
  - Width of depletion L_d region grows with reverse bias
  - $L_{eff} = L - L_d$
- Shorter L_{eff} gives more current
  - I_{ds} increases with V_{ds}

$$I_{ds} = \frac{\beta}{2}\left(V_{gs} - V_t\right)^2\left(1 + \lambda V_{ds}\right)$$

λ- channel length modulator factor



www.rejinpaul.com

3.4)Threshold Voltage(V_t)

- It is defined as the voltage applied between the gate and source, which the device starts conduction or turns on.

Parameters define V_t

1)Gate conductor material 2) Gate insulation material

3) Insulator thickness 4) Impurities

5) Voltage between source and substrate

$$V_t = 2\phi_b + \frac{\sqrt{2\,\varepsilon_{si}\,q\,N_A\,2\phi_b}}{C_{ox}} + V_{fb}$$

$2\phi_b$ ↓ bulk potential

$\frac{\sqrt{2\,\varepsilon_{si}\,q\,N_A\,2\phi_b}}{C_{ox}}$ ↓ Ideal threshold voltage

V_{fb} ↓ flat band voltage

$$\phi_b = \frac{KT}{q} \ln\left(\frac{N_A}{N_i}\right) = \text{Bulk potential}$$

www.rejinpaul.com

Body effect:

- The potential difference between the source and body (Vsb) will affect the threshold voltage (Vt) which is called body effect.

www.rejinpaul.com

3.5)Body Effect

- Body is a fourth transistor terminal. The change in threshold voltage of the MOS transistor because of the non zero bias to the body is called Body effect.
- It occurs when body or substrate of transistor is not biased at the same level of source .
- Because of vol difference between source and substrate Vt is not constant.

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

- ϕ_s = *surface potential* at threshold
- γ = *body effect coefficient*

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

  - Vsb→substrate voltage, Φb→bulk potential,NA→doping concentartion

www.reinpaul.com

3.6)Leakage Sources

1. Sub-threshold conduction

   Transistors can't abruptly turn ON or OFF

2. Gate leakage

   Tunneling through ultra thin gate dielectric

3. Junction leakage

   Reverse-biased PN junction diode current

www.rejinpaul.com

3.6.1)Subthreshold Leakage

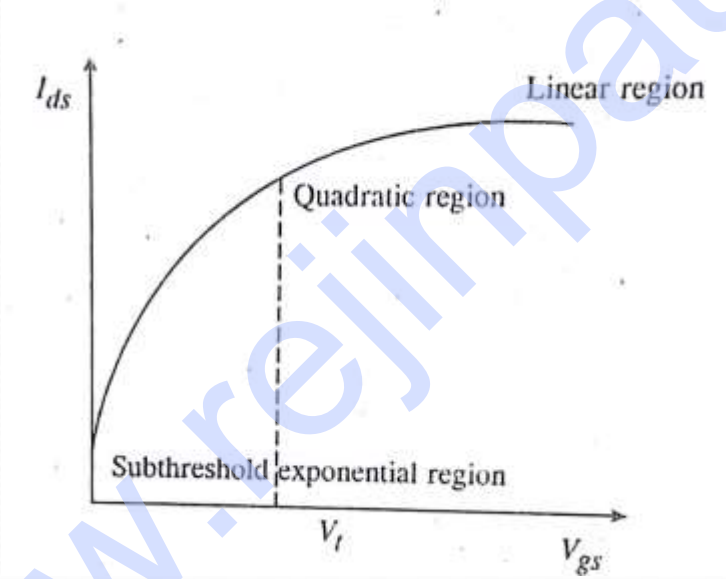
- In ideal transistor if the device turns off abruptly when Vgs<Vt
- But in practically because of inversion layer some current flows from source to drain.
- This conduction is called leakage and the resulting current is leakage current.
- I_{ds}→ current at the threshold and its dependant on process and device geometry

$$I_{ds} = I_{ds0} e^{\frac{V_{gs} - V_{t0} + \eta V_{ds} - k_\gamma V_{sb}}{n v_T}} \left(1 - e^{\frac{-V_{ds}}{v_T}}\right)$$

www.rejinpaul.com

MOS Subthreshold Graph

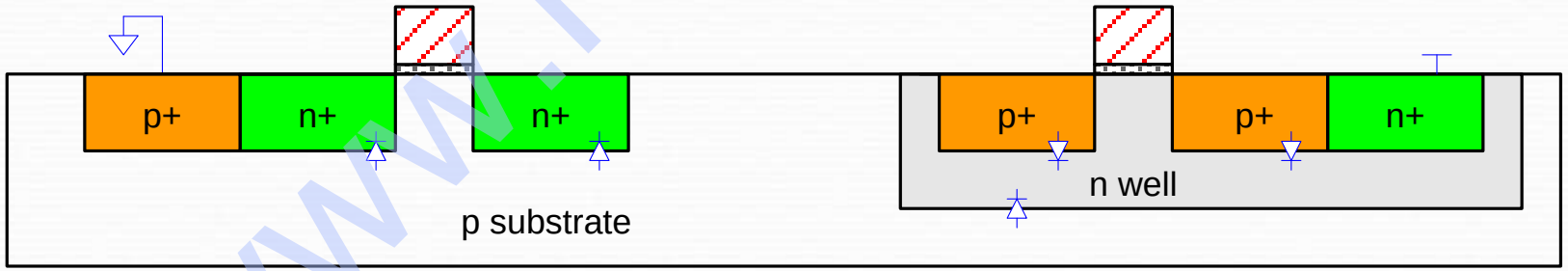
- Graph shows that the graph does not drop to zero immediately for $V_{gs}<V_t$,but decays in an exponential.



www.rejinpaul.com

3.6.3)Junction Leakage

- Reverse-biased p-n junctions have some leakage
  - Ordinary diode leakage
  - Band-to-band tunneling (BTBT)



www.rejinpaul.com

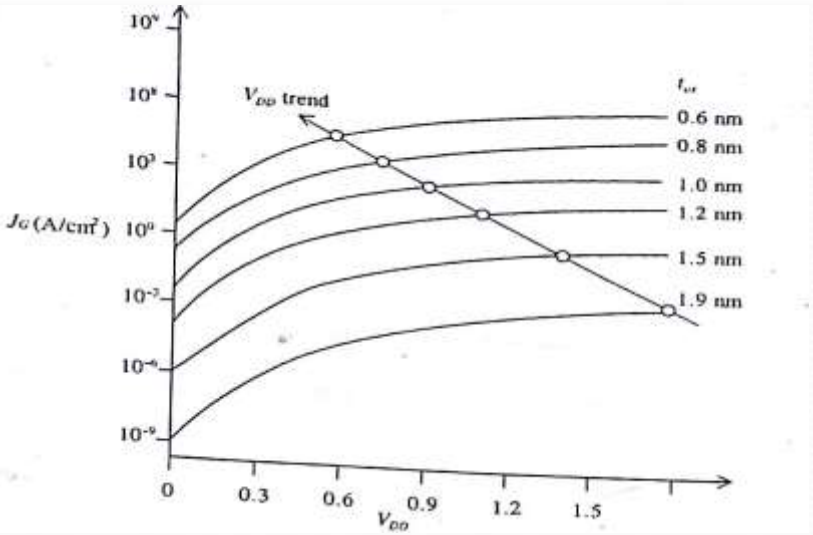
3.7) Tunneling

- Tunneling across heavily doped p-n junctions
  - Especially sidewall between drain & channel
    Increases junction leakage to significant levels

$$I_{BTBT} = WX_j A \frac{E_j}{E_g^{0.5}} V_{dd} \, e^{-B\frac{E_g^{1.5}}{E_j}}$$

$$E_j = \sqrt{\frac{2qN_{halo}N_{sd}}{\varepsilon\left(N_{halo}+N_{sd}\right)}}\left(V_{DD} + v_T \ln\frac{N_{halo}N_{sd}}{n_i^2}\right)$$

  - X_j: sidewall junction depth
  - E_g: bandgap voltage
  - A, B: tech constants



www.rejinpaul.com

3.8)Temperature Sensitivity

- Increasing temperature which Reduces mobility and Reduces V_t
- I_{ON} → decreases with temperature
- I_{OFF} →increases with temperature

$$\mu(T) = \mu(T_r)\left(\frac{T}{T_r}\right)^{-k\mu}$$

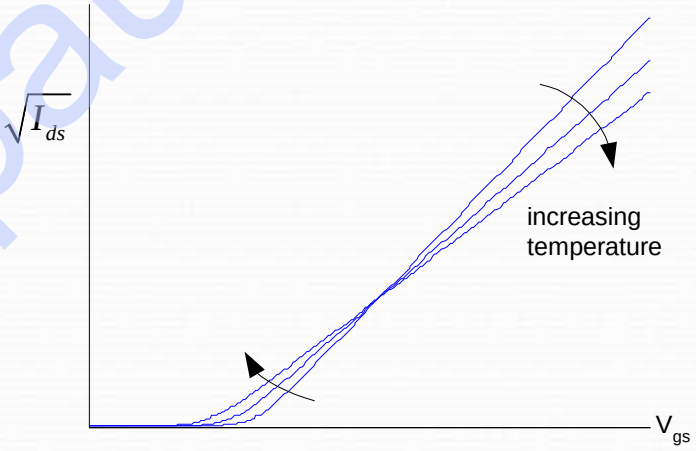
T = absolute temperature

T_r = room temperature

$k\mu$ = fitting parameter in the range from 1.2 to 2.0

The magnitude of the threshold voltage decreases linearly with temperature.

i.e. $V_t(T) = V_t(T_r) - K_u(T - T_r)$



www.rejinpaul.com

4) DC Transfer Characteristics

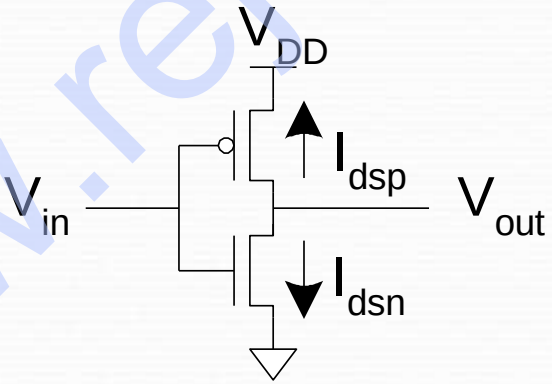
- DC transfer characteristics of the inverter are a function of the output voltage (V_{out}) with respect to the input voltage (V_{in}).
- DC transfer characteristics are

1. CMOS inverter characteristics
2. Beta Ratio Effect
3. Noise Margin
4. Ratioed Inverter
5. Pass transistor logic(unit I,II)
6. Transmission Gate
7. Tristate Inverter.

www.rejinpaul.com

DC Response of CMOS

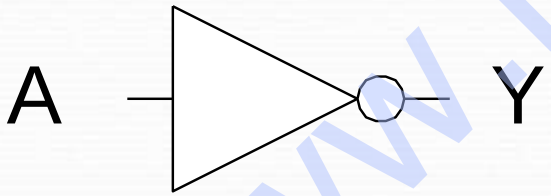
- Ex: Inverter
  - When $V_{in} = 0$ -> $V_{out} = V_{DD}$
  - When $V_{in} = V_{DD}$ -> $V_{out} = 0$

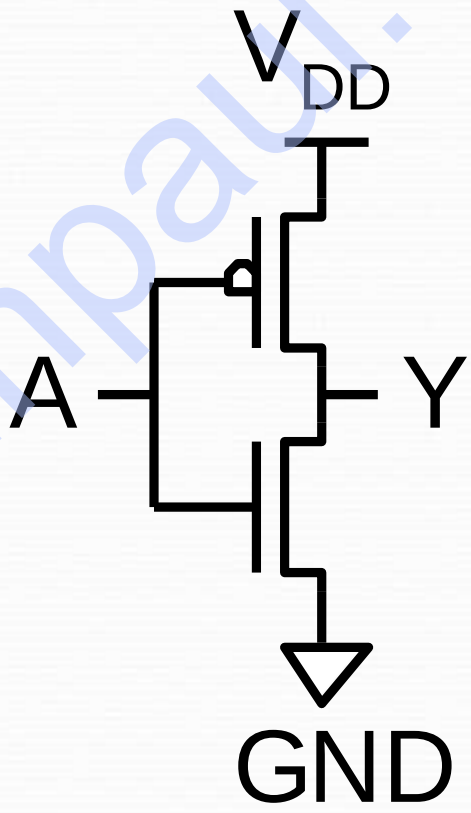


www.reinpaul.com

CMOS Inverter

A	Y
0	
1	

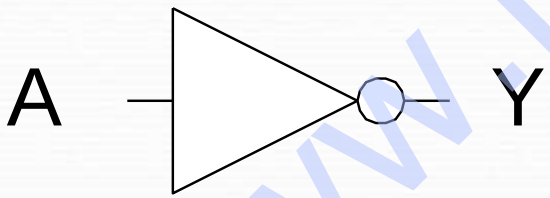


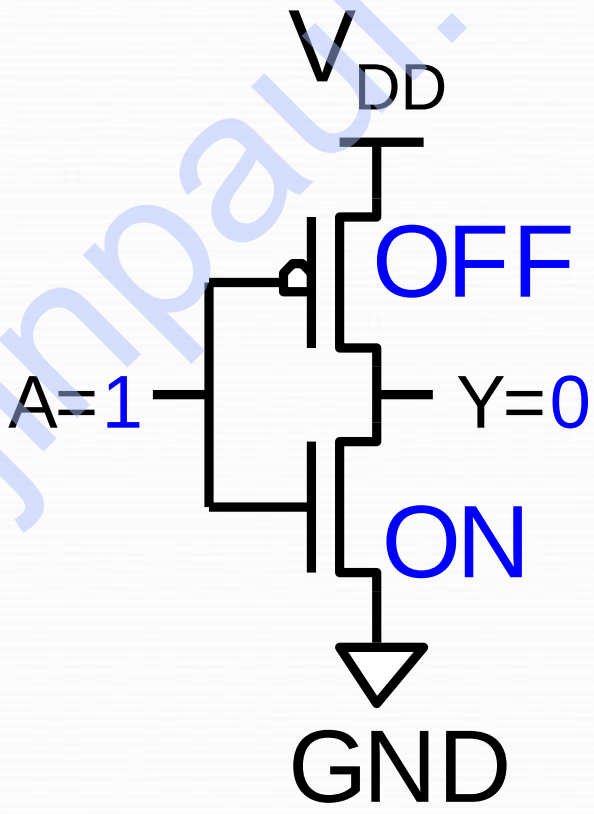


www.rejinpaul.com

CMOS Inverter

A	Y
0	
1	**0**



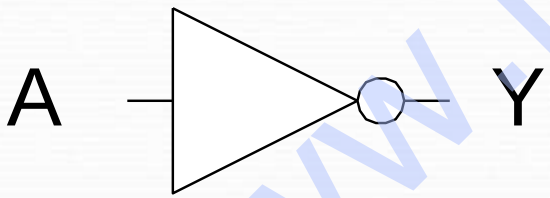


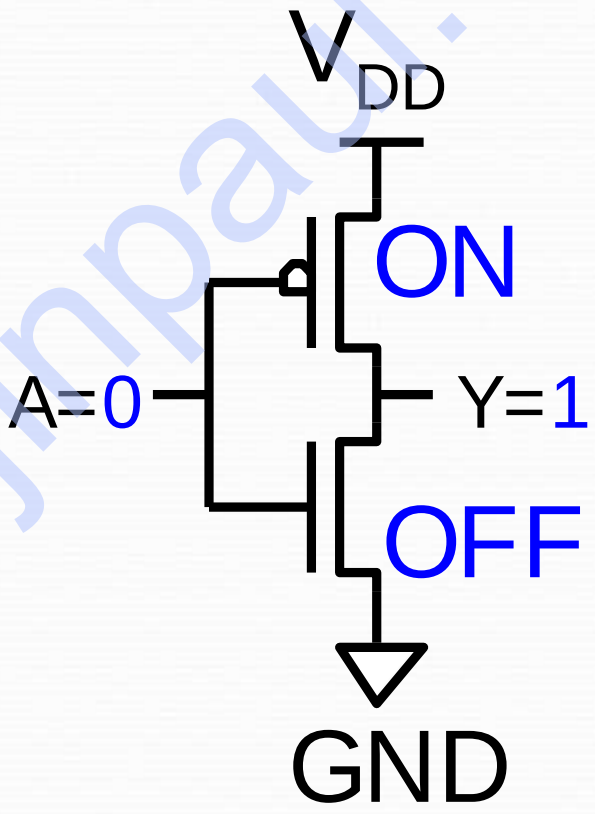
Y is pulled low by the turned on NMOS Device. Hence NMOS is the pull-down device.

www.rejinpaul.com

CMOS Inverter

A	Y
0	1
1	0



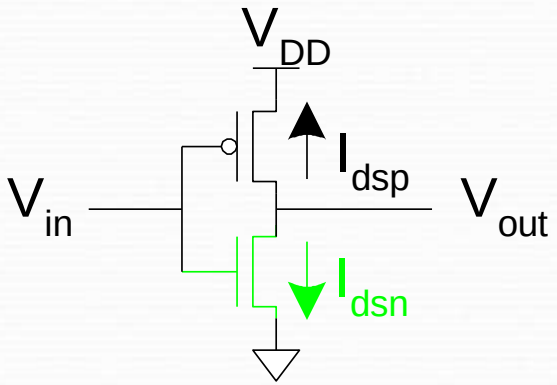


Y is pulled high by the turned on PMOS Device. Hence PMOS is the pull-up device.

www.rejinpaul.com

nMOS Operation

Cutoff	Linear	Saturated
V_{gsn} <	V_{gsn} >	V_{gsn} >
	V_{dsn} <	V_{dsn} >



www.rejinpaul.com

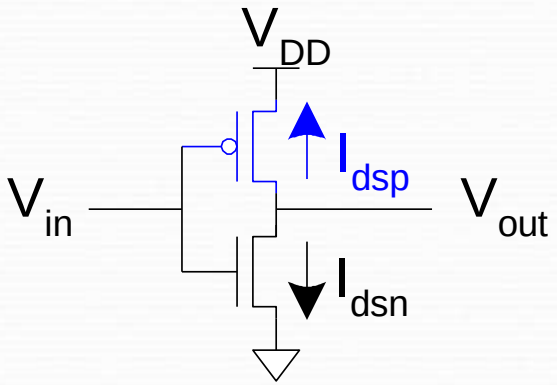
pMOS Operation

Cutoff	Linear	Saturated
$V_{gsp} > V_{tp}$ $V_{in} > V_{DD} + V_{tp}$	$V_{gsp} < V_{tp}$ $V_{in} < V_{DD} + V_{tp}$ $V_{dsp} > V_{gsp} - V_{tp}$ $V_{out} > V_{in} - V_{tp}$	$V_{gsp} < V_{tp}$ $V_{in} < V_{DD} + V_{tp}$ $V_{dsp} < V_{gsp} - V_{tp}$ $V_{out} < V_{in} - V_{tp}$

$V_{gsp} = V_{in} - V_{DD}$

$V_{dsp} = V_{out} - V_{DD}$

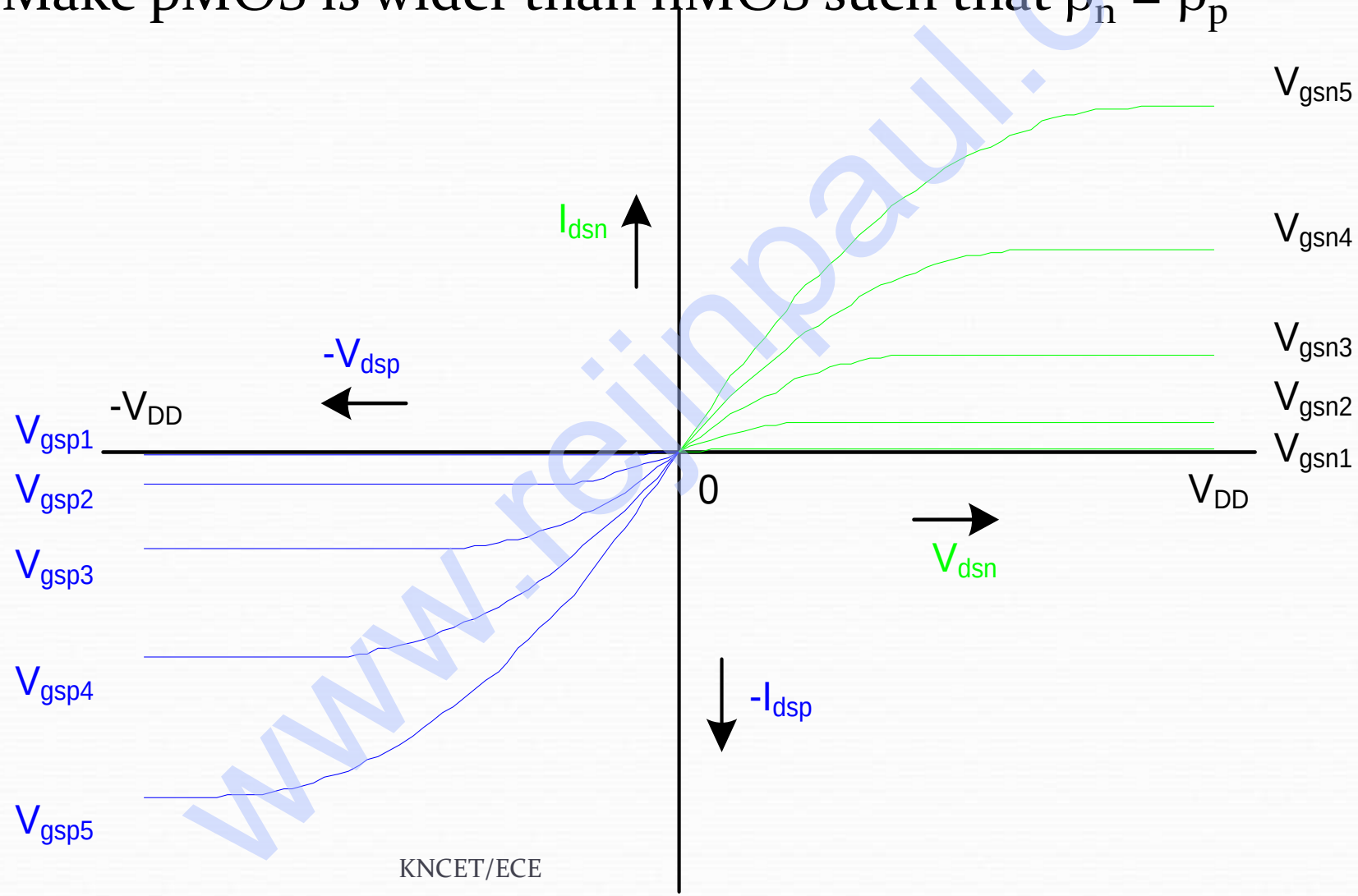
$V_{tp} < 0$



www.rejinpaul.com

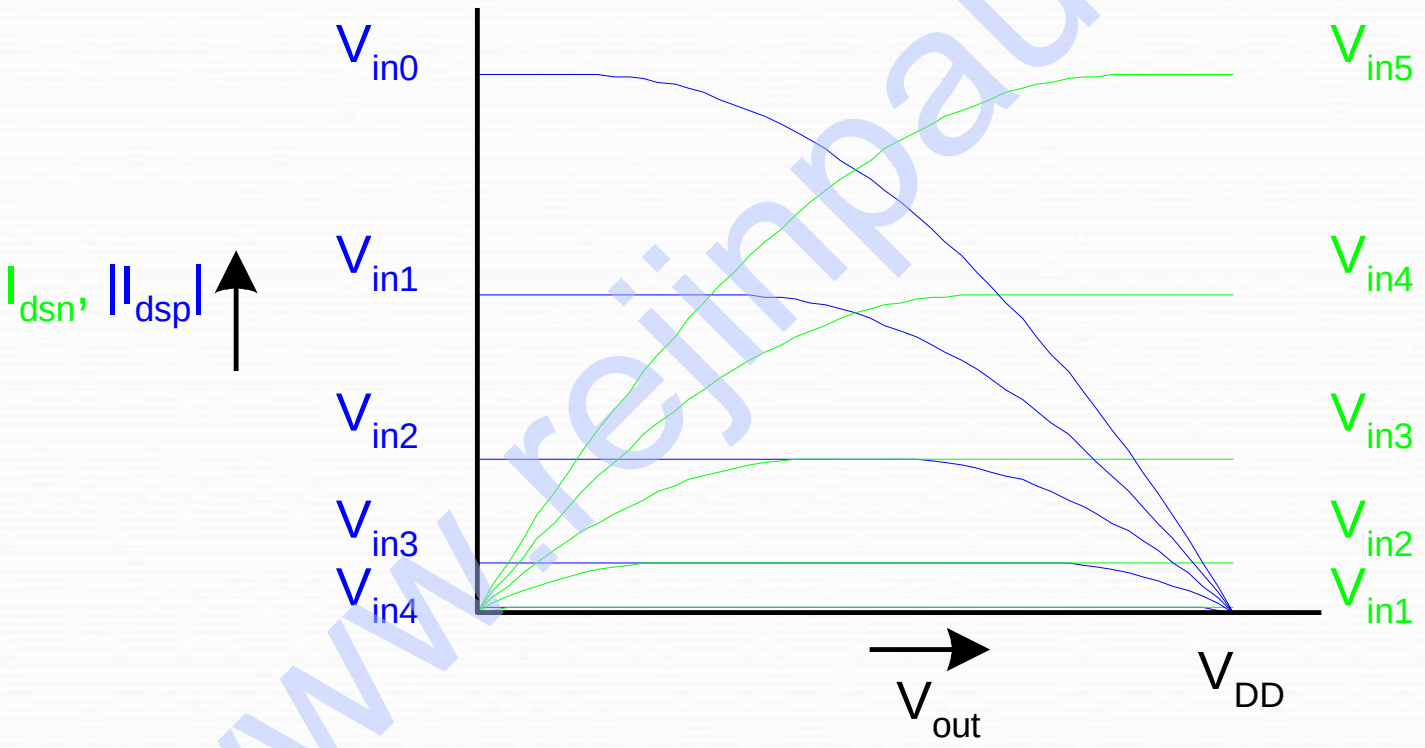
I-V Characteristics

- Make pMOS is wider than nMOS such that $\beta_n = \beta_p$



www.rejinpaul.com

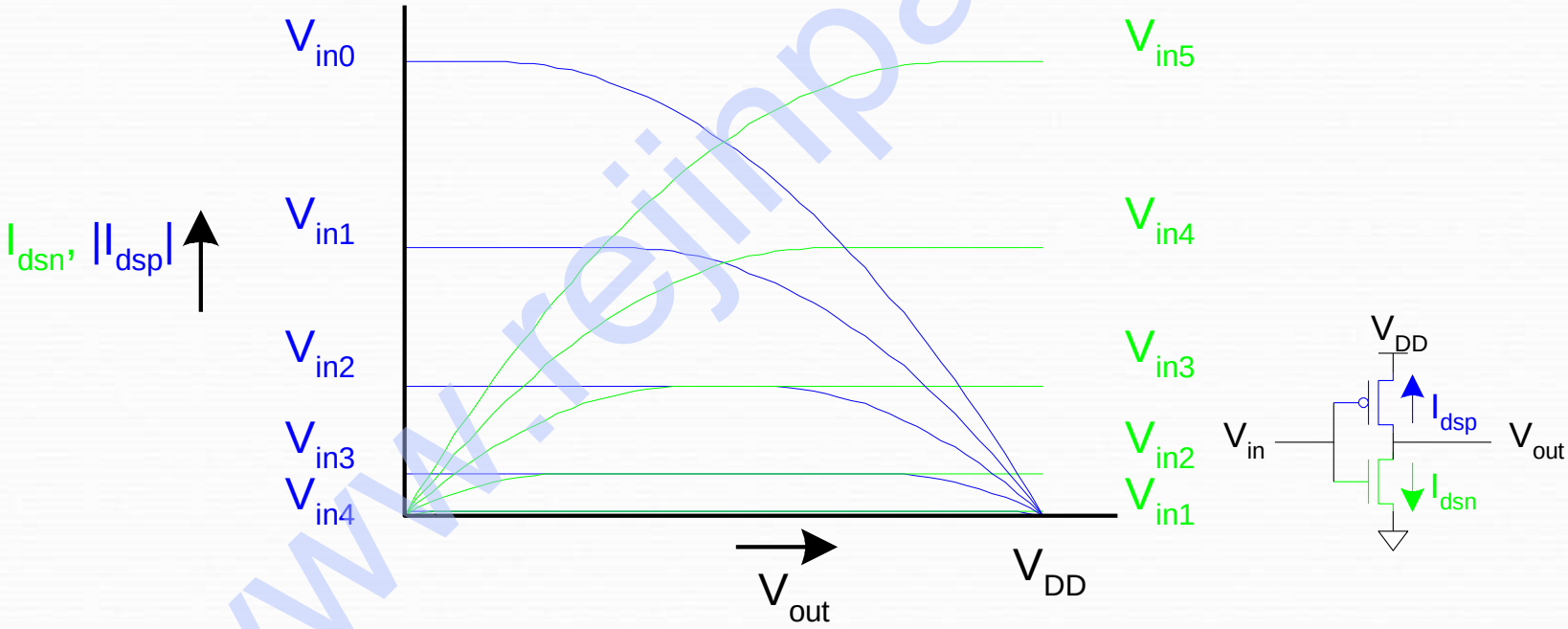
Current vs. V_{out}, V_{in}



www.rejinpaul.com

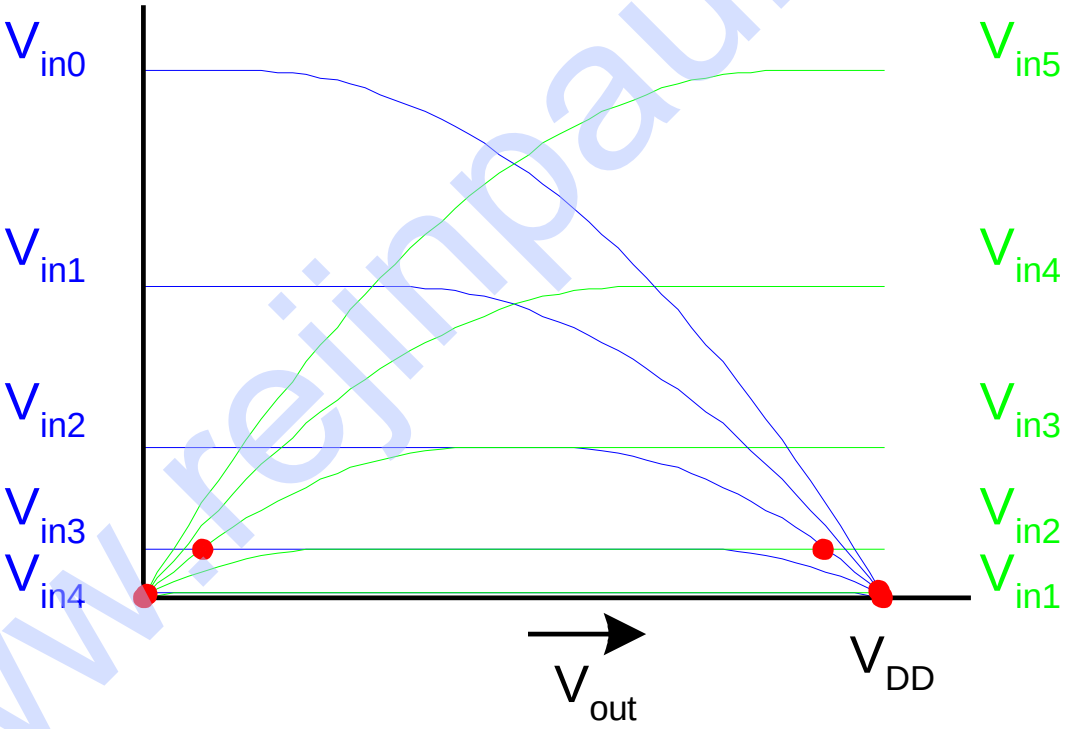
Load Line Analysis

- For a given V_{in}:
  - Plot I_{dsn}, I_{dsp} vs. V_{out}
  - V_{out} must be where |currents| are equal in



www.rejinpaul.com

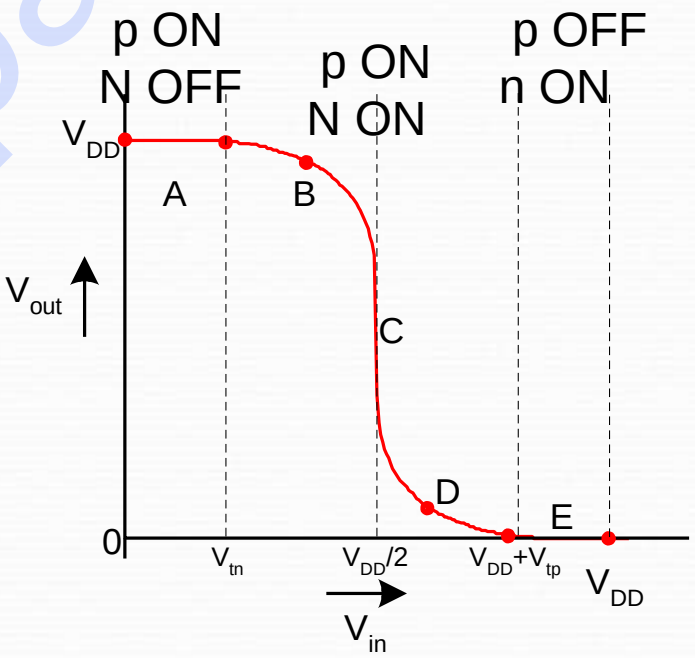
Load Line Analysis



www.rejinpaul.com

Operating Regions

Region	nMOS	pMOS
A		
B		
C		
D		
E		



www.rejinpaul.com

CMOS Inverter Operating Regions

- **Region A** occurs when 0 =V_{in} =$V_{t(n\text{-}type)}$.
  - The n-device is in cut-off ($I_{dsn\ =0}$).
  - p-device is in linear region,
  - I_{dsn} = 0
  - For p-MOSV_{out} = V_{DD}.
- **Region B** occurs when the condition V_{tn} =V_{in} = V_{DD}/2 is met.
  - Here p-device is in its non-saturated region Vds= 0.
  - n-device is in saturation
- Saturation current Idsn is obtained by setting V_{gs} = V_{in} resulting in the equation:

$$I_{dsn} = \frac{\beta_n}{2}\left[V_{un} - V_{tn}\right]^2$$

www.reinpaul.com

CMOS Inverter Operating Regions

- **Region C** has that both n- and p-devices are in saturation.
- Saturation currents for the two devices are:

$$I_{dsp} = -\frac{\beta_p}{2}\left(V_{in} - V_{DD} - V_{tp}\right)^2;\ V_{in} < V_{tp} + V_{DD}$$

AND

$$I_{dsn} = \frac{\beta_n}{2}\left(V_{in} - V_{tn}\right)^2;\ V_{in} > V_{tn}$$

- **Region D** is defined by the inequality

$$\frac{V_{DD}}{2} < V_{in} \le V_{DD} + V_{tp}$$

- p-device is in saturation while n-device is in its non-saturation region.

$$I_{dsp} = -\frac{\beta_p}{2}\left(V_{in} - V_{DD} - V_{tp}\right)^2;\ V_{in} < V_{tp} + V_{DD}$$

AND

$$I_{dsn} = \beta_n\left[\left(V_{in} - V_{tn}\right)V_{out} - \left(\frac{V_{out}}{2}\right)^2\right];\ V_{in} > V_{tn}$$

- Equating the drain currents allows us to solve for Vout. (See supplemental notes for algebraic manipulations).

www.rejinpaul.com

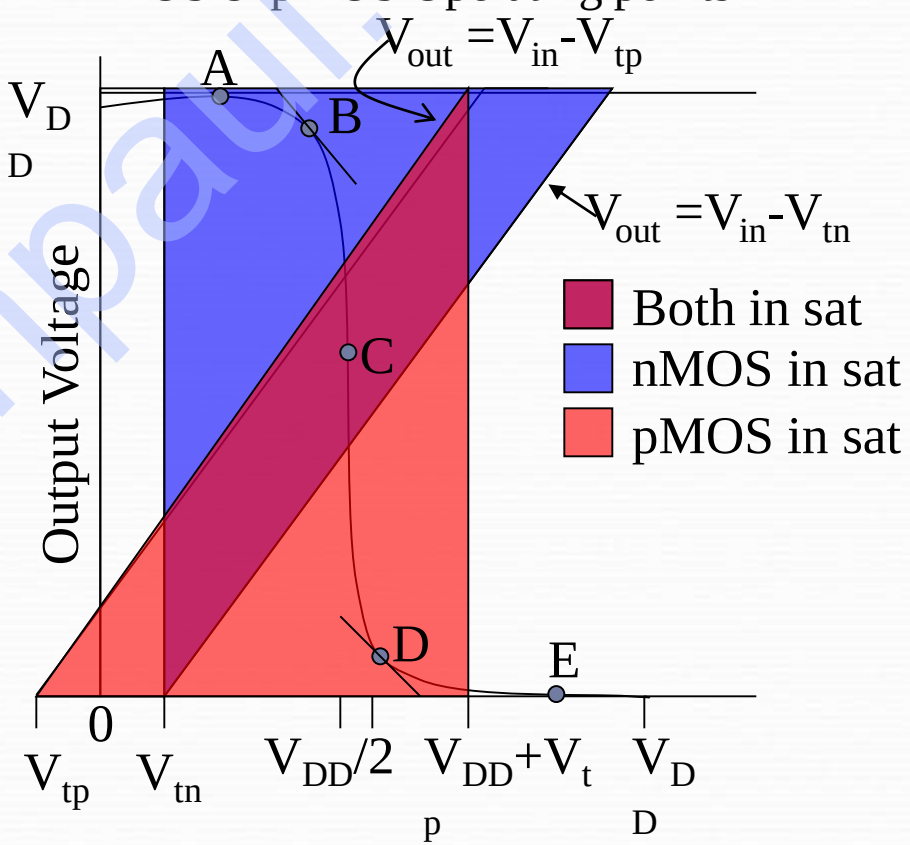
CMOS Inverter Operating Regions

- In **Region E** the input condition satisfies:

$$V_{in} \geq V_{DD} - V_{tp}$$

- The p-type device is in cut-off: $I_{dsp}=0$
- The n-type device is in linear mode
- $V_{gsp} = V_{in} - V_{DD}$ and this is a more positive value compared to V_{tp}.
- $V_{out} = 0$

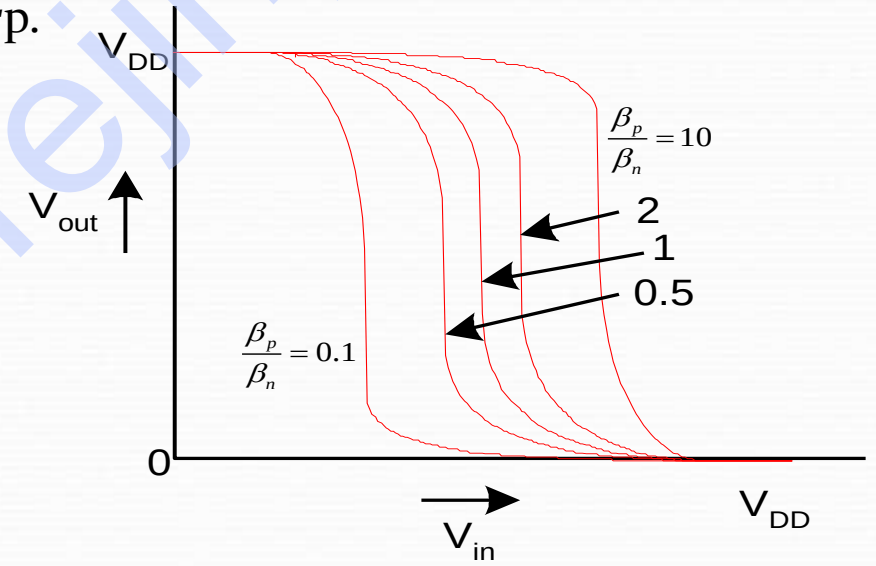
- nMOS & pMOS Operating points



www.rejinpaul.com

2)Beta Ratio

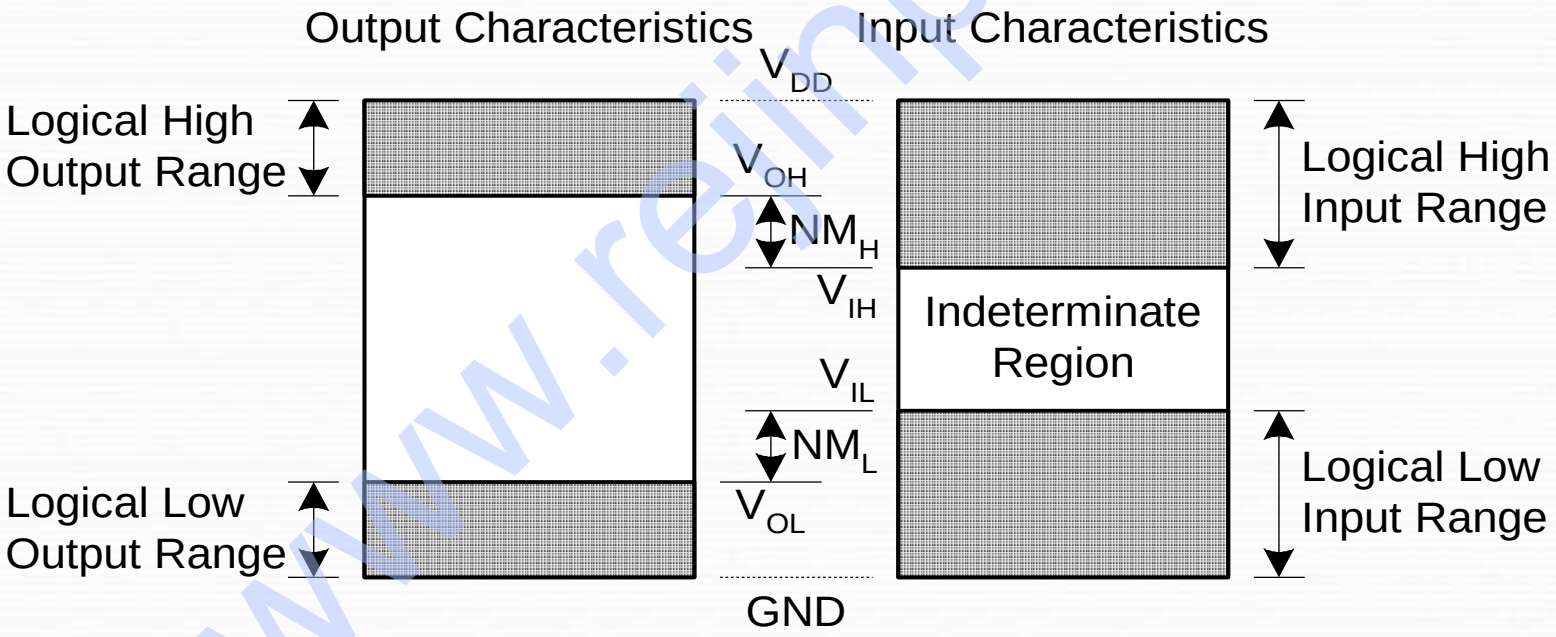
- If $\beta_p/\beta_n = 1$ Called normal or un-*skewed* inverter,where the inverter threshold voltage V_{in} is $V_{DD}/2$.
- If $\beta_p/\beta_n > 1$ the inverter is HI-skewed Inverter.it has a stronger p-MOS transistor.If the input $V_{DD}/2$ then the output is greater than $V_{DD}/2$.
- If $\beta_p/\beta_n < 1$ the inverter is LO-skewed Inverter.It has a weaker p-MOS and has low switching threshold.
- If $\beta_p/\beta_n \neq 1$, switching point will move from $V_{DD}/2$
- From the graph,when beta ratio is changed,the swiching threshold moves,but the o/p voltage remains sharp.



www.rejinpaul.com

3)Noise Margins

- Noise margin indicates, how well the gate will perform under noisy conditions.
- It is used to determine the allowable noise voltage or rate for gate so that the output not affected.



www.rejinpaul.com

- Low Noise Margin→Difference in magnitude between Max low input voltage to max low output voltage produced by the driving gate.

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

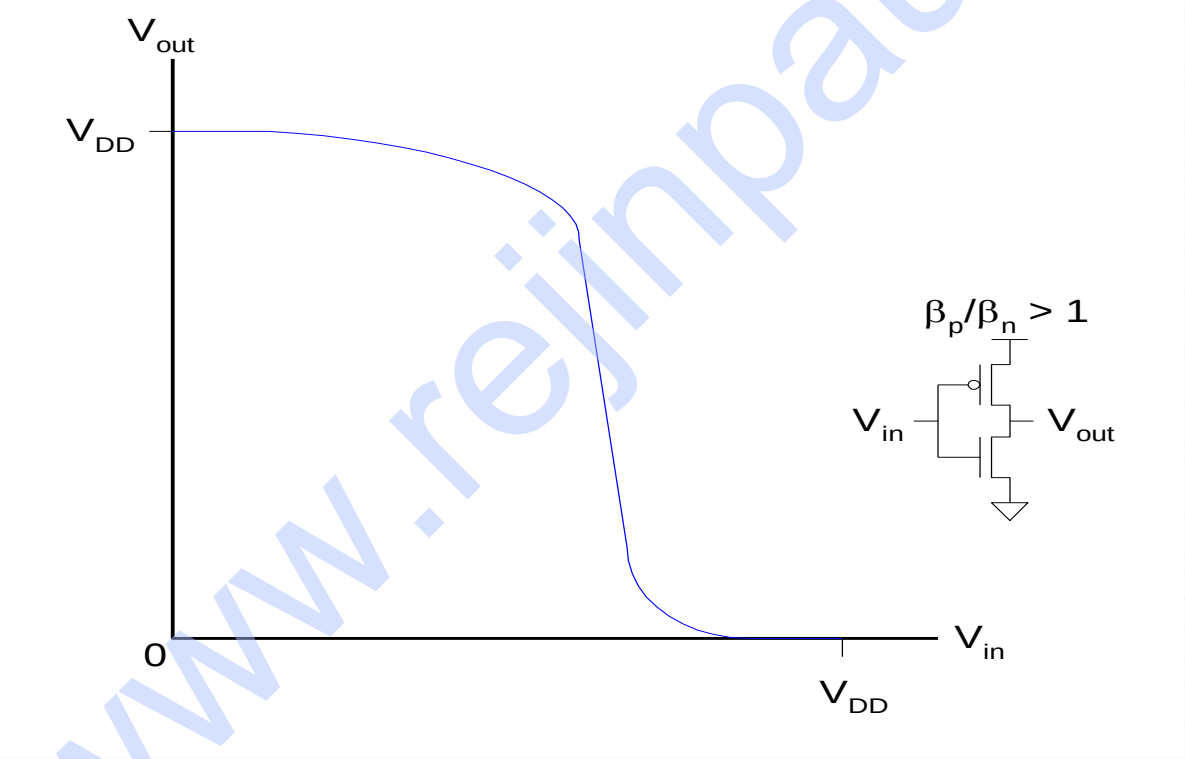
- High Noise Margin→ Difference in magnitude between Min high output voltage to min high input voltage produced by the driving gate.

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

www.rejinpaul.com

CMOS Inverter Noise margin

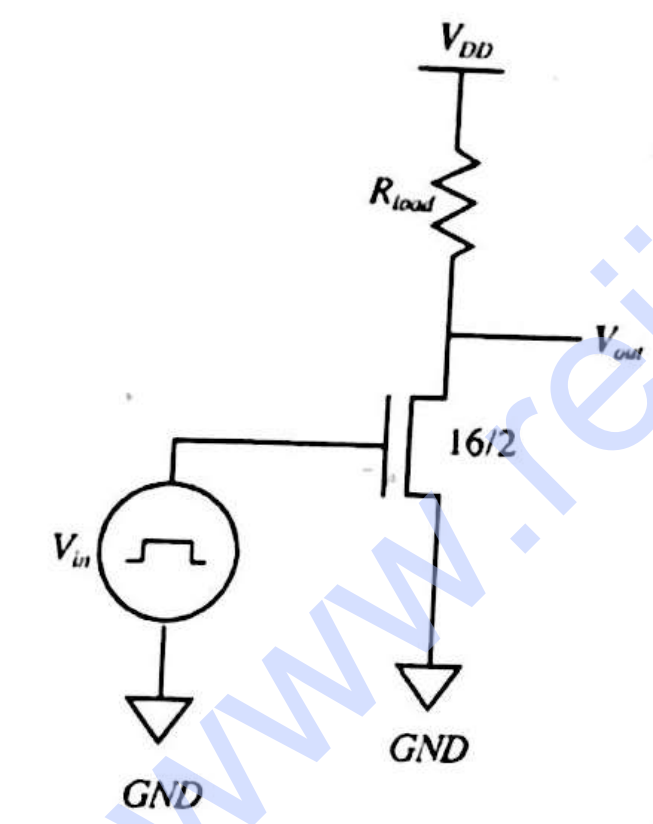
- To maximize noise margins, select logic levels at unity gain point of DC transfer characteristics.

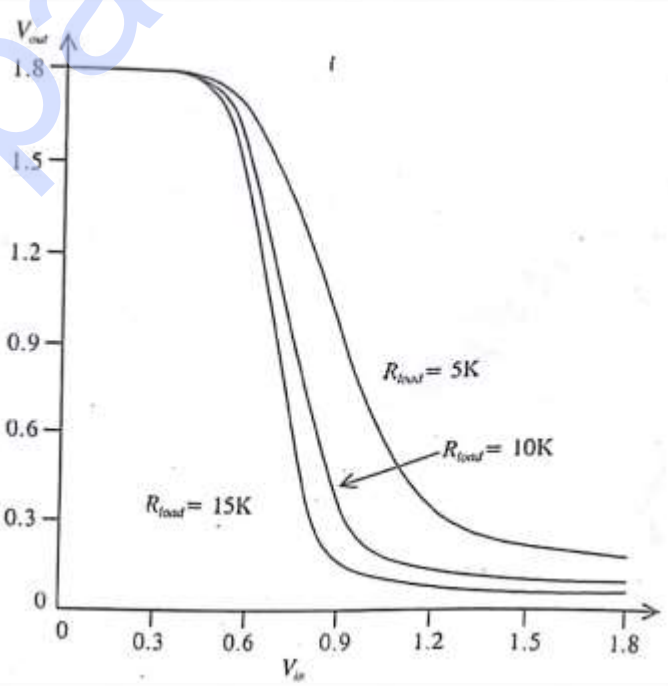


www.rejinpaul.com

4)Ratioed Inverter

- nMOS inverter uses resistive load.
- If $V_{in}=V_{DD}$,then $V_{out}=o$
- If $V_{in}=o$,then $V_{out}=V_{DD}$

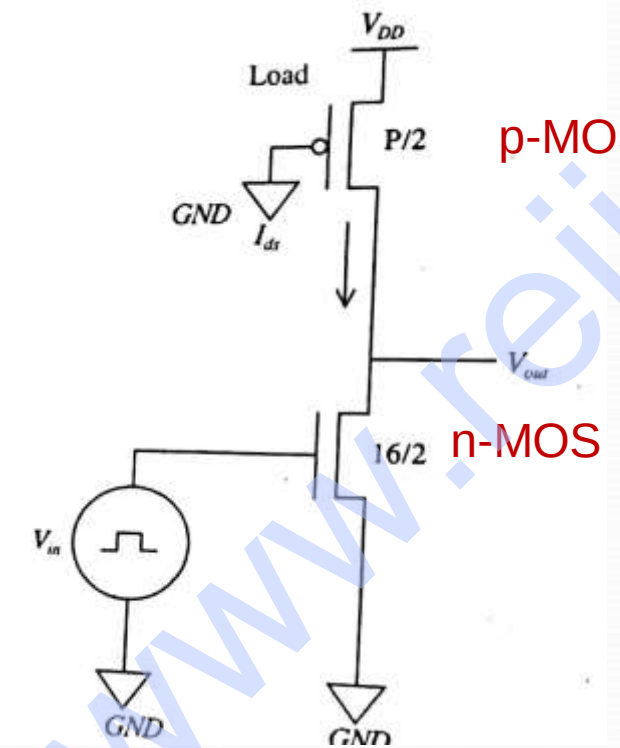


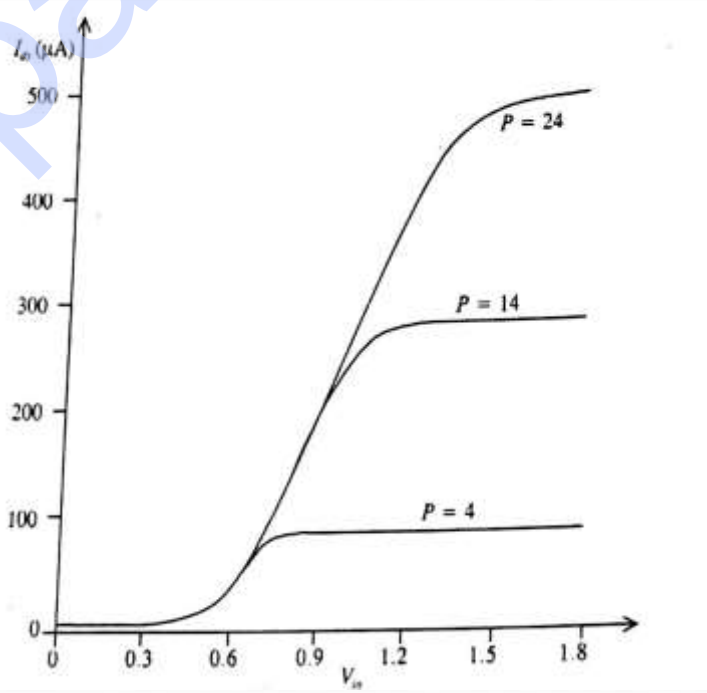


www.rejinpaul.com

Pseudo nMOS Inverter

- It use pMOS transistor pull-up load with gate permanently grounded.
- pMOS transistor offers faster rise time but less sharp transfer characteristics.
- When nMOS transistor is turned ON→constant DC current flows in the circuit.

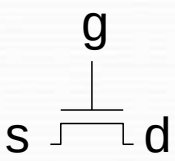


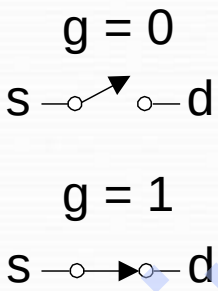


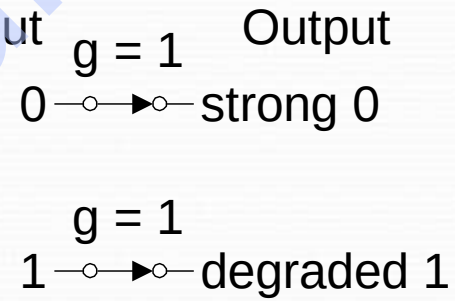
www.rejinpaul.com

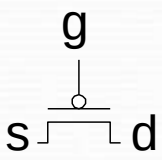
5)Pass Transistors

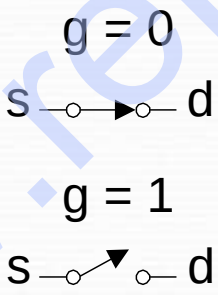
- Transistors can be used as switches

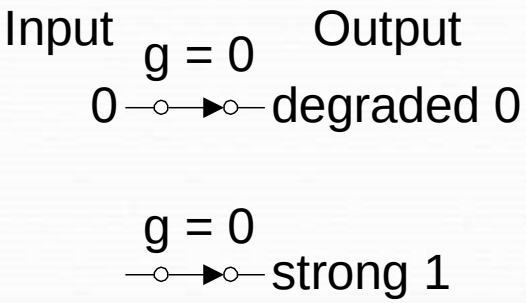










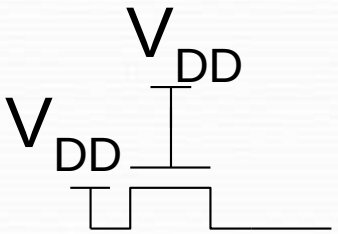


For nMOS→when gate is high then input passed to output and when gate is low the output is high impedance or floating(Z).

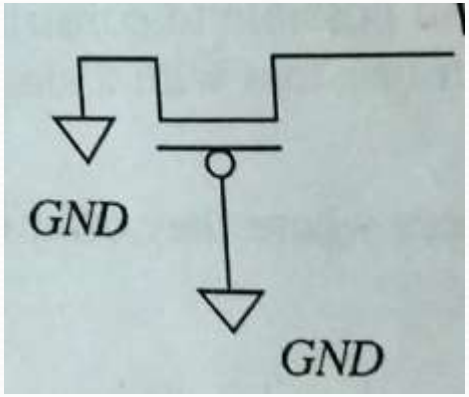
www.rejinpaul.com

nMOS-pMOS Transistors

- $V_g = V_{DD}$
  - If $V_s > V_{DD}-V_t$, $V_{gs} < V_t$
  - Hence transistor would turn itself off
- nMOS pass transistors pull no higher than $V_{DD}-V_{tn}$
  - Called a degraded "1"
  - Approach degraded value slowly (low I_{ds})

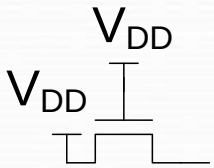


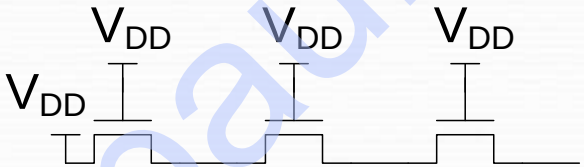
- pMOS pass transistors pull no lower than V_{tp}
- Transmission gates are needed to pass both 0 and 1

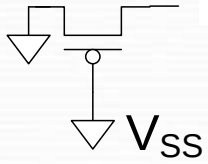


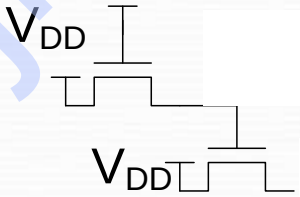
www.rejinpaul.com

Pass Transistor Ckts









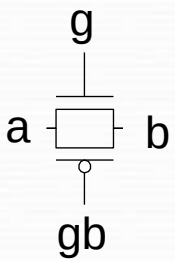
www.rejinpaul.com

6)Transmission gate

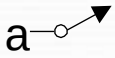
- Pass transistors produce degraded outputs but *Transmission gates* pass both 0 and 1 well as a strong output.
- Transmission gate logic consists of an nMOS and a pMOS transistor in parallel with gates controlled by complementary signals.
- At least one of the two transistor is ON and the output voltage is either strong 0 or strong 1.
- It is superior than CMOS ckts interms of layout density, ckt delay and power consumption.
- Gate is driven by control signal.
- Source is connected with pass variable.

www.rejinpaul.com

Transmission Gates

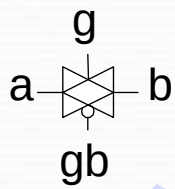


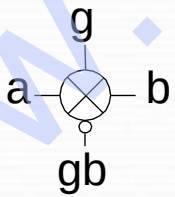
g = 0, gb = 1

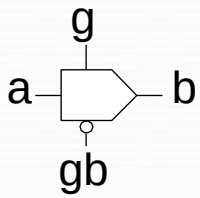


g = 1, gb = 0

Input	Output
g = 1, gb = 0 0	strong 0
g = 1, gb = 0 1	strong 1



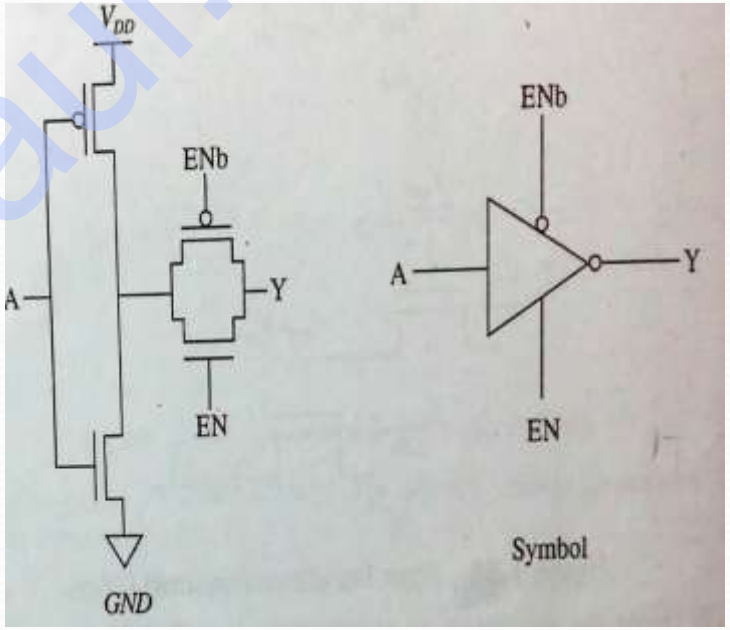




www.rejinpaul.com

6)Tristate Inverter

- It is constructed by cascading a transmission gate with an inverter.
- When EN=0 →output Y is in tristate.
- When EN=1 →output Y is the complement of input A.
- *Tristate buffer* produces Z when not enabled



EN	A	Y
0	0	Z
0	1	Z
1	0	1
1	1	0

www.rejinpaul.com

Device Models

- SPICE software developed by UC Berkeley in 1970
- Level 1 Models→C-V characteristics MOS
- Level 2Models→Analysis of MOS
- Level 3 Models→Analysis of short channel MOS.
- Level 4 (or)BSIM Models→short channel IGFET model
- Diffusion Capacitance Models →Analyze parasitic capacitance of S-G

www.rejinpaul.com

1)LEVEL-1 Model

- The dc characteristics of the MOS LEVEL-1 model are defined by the device electrical parameters VTO,KP,LAMBDA,PHI,GAMMA

NMOS Model			
LEVEL = 1	VTO = 0.7	GAMMA = 0.45	PHI = 0.9
NSUB = 9e + 14	LD = 0.08e − 6	UO = 350	LAMBDA = 0.1
TOX = 9e − 9	PB = 0.9	CJ = 0.56e − 3	CJSW = 0.35e − 11
MJ = 0.45	MJSW = 0.2	CGDO = 0.4e − 9	JS = 1.0e − 8
PMOS Model			
LEVEL = 1	VTO = − 0.8	GAMMA = 0.4	PHI = 0.8
NSUB = 5e + 14	LD = 0.96e − 6	UO = 100	LAMBDA = 0.2
TOX = 9e − 9	PB = 0.9	CJ = 0.94e − 3	CJSW = 0.32e − 11
MJ = 0.5	MJSW = 0.3	CGDO = 0.3e − 9	JS = 0.5e − 8

www.rejinpaul.com

$$\text{PHI} = 2 \times \frac{kT}{q} \ln\left(\frac{\text{NSUB}}{n_i}\right)$$

$$C_{ox} = \frac{\varepsilon_{ox}}{T_{OX}}$$

$$\text{GAMMA} = \sqrt{\frac{2\varepsilon_{si} q \times \text{NSUB}}{C_{ox}}}$$

$$V_T = VTO + \text{GAMMA}\left(\sqrt{\text{PHI} - V_{BS}} - \sqrt{\text{PHI}}\right)$$

$$KP = U0 \times C_{ox}$$

VTO: threshold voltage with zero V_{SB} (unit: V)

GAMMA: body effect coefficient (unit: $V^{1/2}$)

PHI: $2\Phi_F$ (unit: V)

TOX: gate oxide thickness (unit: m)

www.rejinpaul.com

NSUB:	substrate doping (unit: cm)
LD:	source/drain side diffusion (unit: m)
UO:	channel mobility (unit: $cm^2/V/s$)
LAMBDA:	channel-length modulation co-efficient (unit: V^{-1})
CJ:	source/drain bottom-plate junction capacitance per unit area (unit: F/
CJSW:	source/drain sidewall junction capacitance per unit length (unit: F/m)
PB:	source/drain junction built-in potential (unit: V)
MJ:	exponent in CJ equation (unitless)
MJSW:	exponent in CJSW equation (unitless)
CGDO:	gate-drain overlap capacitance per unit width (unit: F/m)
CGSO:	gate-source overlap capacitance per unit length (unit: F/m)
JS:	source/drain leakage current per unit area (unit: A/m^2)

www.reinpaul.com

- Drain current

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{cutoff} \\ KP \dfrac{W_{eff}}{L_{eff}} (1 + \text{LAMBDA} \cdot V_{ds}) \left(V_{gs} - V_t - \dfrac{V_{ds}}{2} \right) V_{ds} & V_{ds} < V_{gs} - V_t \\ \dfrac{KP}{2} \dfrac{W_{eff}}{L_{eff}} (1 + \text{LAMBDA} \cdot V_{ds}) (V_{gs} - V_t)^2 & V_{ds} > V_{gs} - V_{gt} \quad \text{saturation} \end{cases}$$

www.rejinpaul.com

2)LEVEL-2 Model

- More accurate model for drain current
- Simplified version of model-1

$$I_{DS} = \frac{KP}{(1 - \text{LAMBDA} \cdot V_{ds})} \cdot \frac{W_{\text{eff}}}{L_{\text{eff}}} \left\{ \left(V_{gs} - V_{FB} - |2\phi_F| - \frac{V_{ds}}{2} \right) \cdot V_{ds} \right.$$

$$\left. - \frac{2}{3} \text{GAMMA} \left[(V_{ds} - V_{BS} + |2\phi_F|)^{3/2} - (-V_{BS} + |2\phi_F|)^{3/2} \right] \right\}$$

www.rejinpaul.com

3)LEVEL-3 Model

- Simulation of short channel MOS transistors
- Channel length down to 2µm
- Improve efficiency and limit the complexity
- Simulation time reduced

$$I_{ds} = \mu_s \cdot C_{ox} \cdot \frac{W_{eff}}{L_{eff}} \left(V_{gs} - V_t - \frac{1+FB}{2} \cdot V_{ds} \right) \cdot V_{ds}$$

www.rejinpaul.com

4)BSIM Model

- **B**erkely **S**hort-Channel **I**GFET **M**odel.
- Simple method and use small number of parameters for analysis.
- Accuracy and efficiency.
- Used in industries and silicon foundaries

Features

- Analyze I-V characteristics across **Subthreshold,linear and saturation regions**
- Detailed analysis of **body effect,Short channel effect,mobility degradation.**
- It has diffusion,gate and resistance models

www.rejinpaul.com

5)Diffusion Capacitance Models

- Analyze parasitic capacitance of **Source and drain** regions with simple pn-junction model.
- 2 separate models for **bottom area depletion capacitance and side wall depletion capacitance**

$$C_{Sb} = AS \cdot CJ \cdot \left[1 + \frac{V_{BS}}{PB} \right]^{-MJ} + PS \cdot CJSW \cdot \left[1 + \frac{V_{BS}}{PHP} \right]^{-MJSW}$$

$$C_{sb} = AS \cdot CJ \cdot \left(1 + \frac{V_{BS}}{PB} \right)^{-MJ} + (PS - W) \cdot CJSW \cdot \left(1 + \frac{V_{BS}}{PBSW} \right)^{-MJSW}$$

$$+ W \cdot CJSWG \cdot \left(1 + \frac{V_{BS}}{PBSWG} \right)^{-MJSWG}$$

www.rejinpaul.com

Propagation delays

- Delay Estimation→ Circuit simulators used to compute how fast a circuit operates /switching.(waveforms and delay).
- To design critical path ,quick delay estimation is required.
- Delay is dependant on the layout.
- **Types**

1)RC Delay Model

2)Elmore Delay Model

3)Logical Delay Model

4)Parasitic Delay model

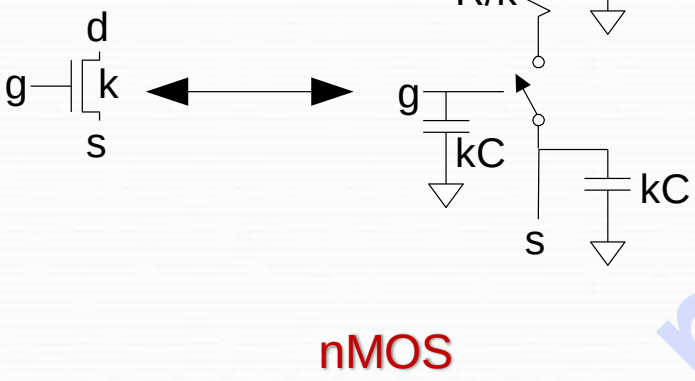
www.rejinpaul.com

1)RC Delay Model

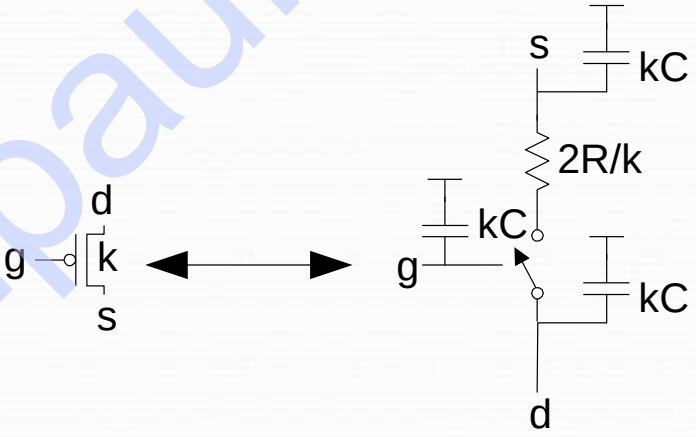
- The delay of logic gate is computed as the product of RC.
- Delay of logic gate is depends upon the transistor width in the gate & the capacitance of the load.
- Use equivalent circuits for MOS transistors
  - Unit nMOS has resistance R, capacitance C
  - Unit pMOS has resistance 2R, capacitance C
- Capacitance proportional to transistor width(k)
- Resistance inversely proportional to transistor width(k)
- If multiple transistor in series→R_{eff}=sum of individual resistance.
- If multiple transistor in parallel→R_{eff}=Resistance of the individual transistor

www.rejinpaul.com

RC Delay model for nMOS&pMOS



nMOS



pMOS

www.rejinpaul.com

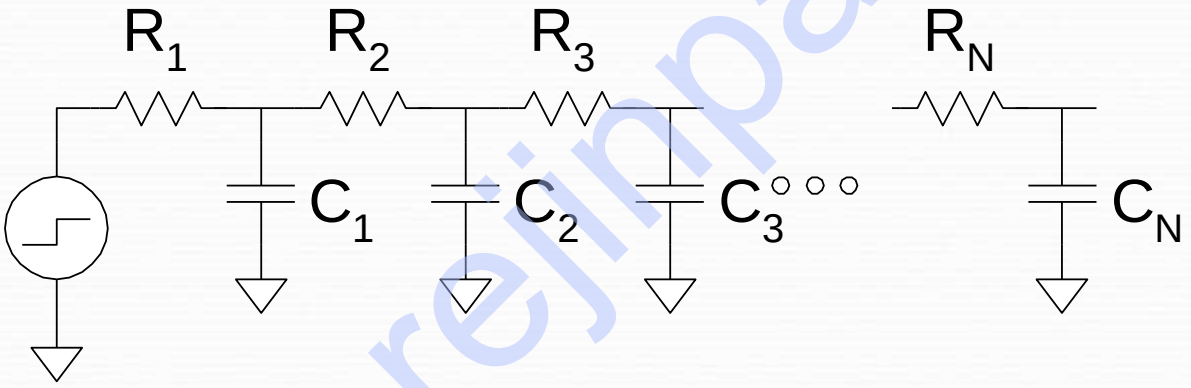
RC Values

- Capacitance
  - $C = C_g = C_s = C_d$ = 2 fF/μm of gate width in 0.6 μm process and upto 1 fF/μm in 65 nm process.
- Resistance
  - R ≈ 10 KΩ in 0.6 μm process
  - 1.25 KΩ in 65 nm process
- Unit transistors
  - May refer to minimum contacted device (4/2 λ)
  - Or maybe 1 μm wide device

www.rejinpaul.com

2)Elmore Delay

- Used to estimate delay value in RC ladder circuit
- ON transistors look like resistors.
- Pullup or pulldown network modeled as *RC ladder*
- Delay is equal to the sum over each node in the ladder of the resistance R_{n-i} between the node and is multiplied by the capacitance on the node.



$$t_{pd} \approx \sum_{\text{nodes } i} R_{i-to-source} C_i$$

$$= R_1 C_1 + (R_1 + R_2) C_2 + \ldots + (R_1 + R_2 + \ldots + R_N) C_N$$

www.rejinpaul.com

3)Logical Effort delay

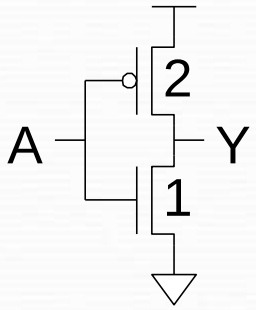
- Logical Effort
- Delay in a Logic Gate
- Multistage Logic Networks
- Choosing the Best Number of Stages

www.rejinpaul.com

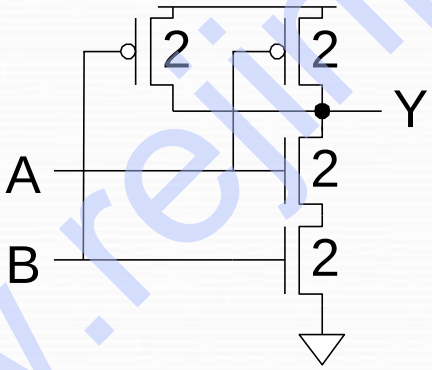
Computing Logical Effort

- Logical effort is the ratio of the **input capacitance of a gate to the input capacitance of an inverter** delivering the same output current.
- Inverter has 3 units of input capacitance(g=n)
- NAND gate has 4 units of input capacitance(g=[(n+2)/3]
- NOR gate has 5 units of input capacitance(g=[(2n+1)/3]

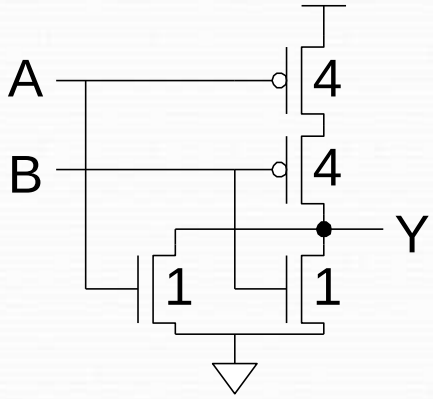
No of input



C_{in} = 3
g = 3/3



C_{in} = 4
g = 4/3



C_{in} = 5
g = 5/3

www.rejinpaul.com

Logical effort of common of Gates

Gate type	Number of inputs				
	1	2	3	4	n
Inverter	1				
NAND		4/3	5/3	6/3	(n+2)/3
NOR		5/3	7/3	9/3	(2n+1)/3
Tristate / mux	2	2	2	2	2
XOR, XNOR		4, 4	6, 12, 6	8, 16, 16, 8	

www.rejinpaul.com

Parasitic delay of common of Gates

- In multiples of p_{inv} (≈ 1)

Gate type	Number of inputs				
	1	2	3	4	n
Inverter	1				
NAND		2	3	4	n
NOR		2	3	4	n
Tristate / mux	2	4	6	8	2n
XOR, XNOR		4	6	8	

www.rejinpaul.com

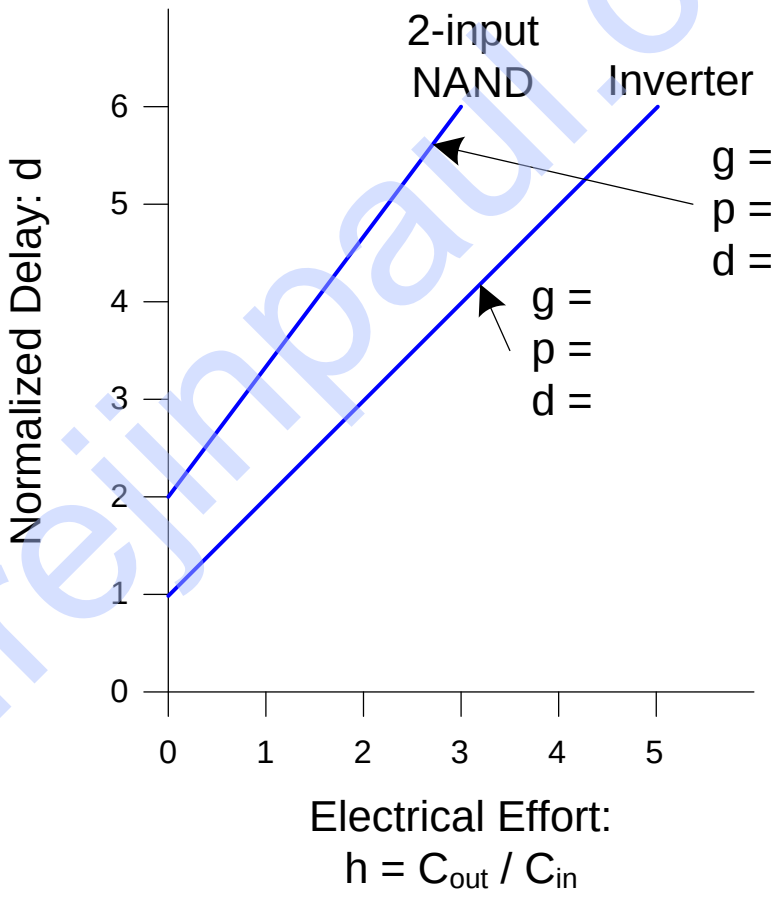
Delay in a Logic Gate

- Propagation delay of a gate is $d = f + p = gh + p$
- f: *effort delay* = gh (depends on the complexity and fan out of the gate)
- g: *logical effort*
  - Measures relative ability of gate to deliver current
  - $g \equiv 1$ for inverter
- h: *electrical effort* = C_{out} / C_{in}
  - Ratio of output to input capacitance
  - Sometimes called fanout(maximum number of digital inputs that the output of a single logic gate can feed)
- p: parasitic delay
  - Represents delay of gate driving no load
  - Set by internal parasitic capacitance

www.rejinpaul.com

Delay Plots

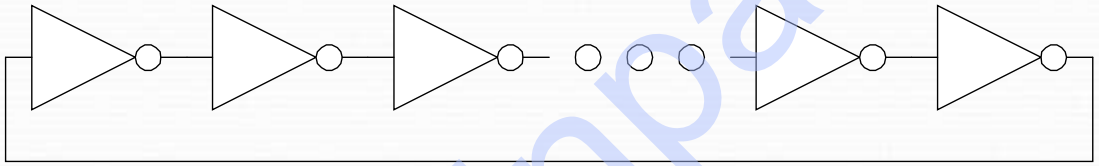
$$d = f + p$$
$$= gh + p$$



www.rejinpaul.com

Example:1 Ring Oscillator

- Estimate the frequency of an N-stage ring oscillator



Logical Effort: g = 1

Electrical Effort: h = 1 (C_{out}/C_{in})

Parasitic Delay: p = 1

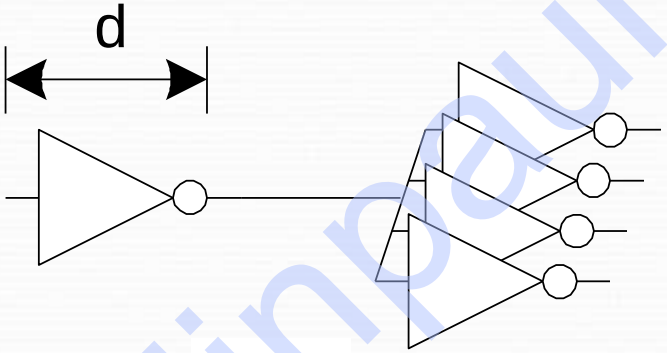
Stage Delay: d = 2

Frequency: $f_{osc} = 1/(2*N*d)$

www.rejinpaul.com

Example:2 FO4 Inverter

- Estimate the delay of a fanout-of-4 (FO_4) inverter



Logical Effort: g =

Electrical Effort: h =

Parasitic Delay: p =

Stage Delay: d =

www.rejinpaul.com

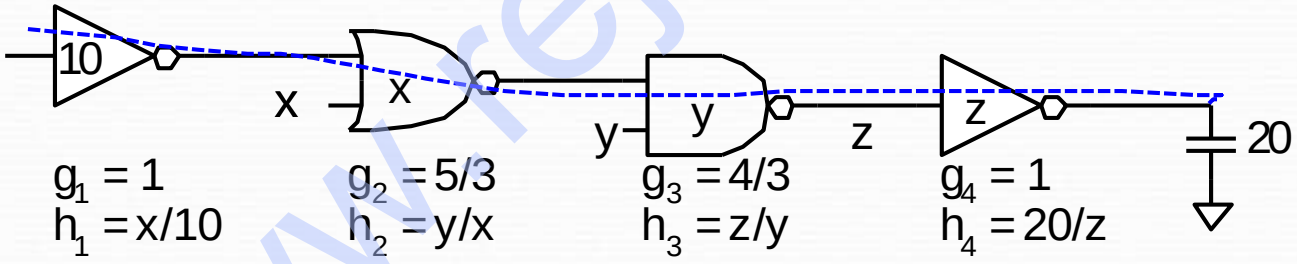
Multistage Logic Networks

- *Path Logical Effort → Mutiplying the logical efforts of all logic gates*

$$G = \prod g_i$$

- *Path Electrical Effort → It is the ratio of output capacitance of logic gate to input capacitance of first gate.*

$$H = \frac{C_{\text{out-path}}}{C_{\text{in-path}}}$$



www.rejinpaul.com

- Branching Effort→total capacitance by a capacitance in the path or accounts for **branching** between stages in path

$$b = \frac{C_{\text{on path}} + C_{\text{off path}}}{C_{\text{on path}}}$$

Note:

$$\prod h_i = BH$$

$$B = \prod b_i$$

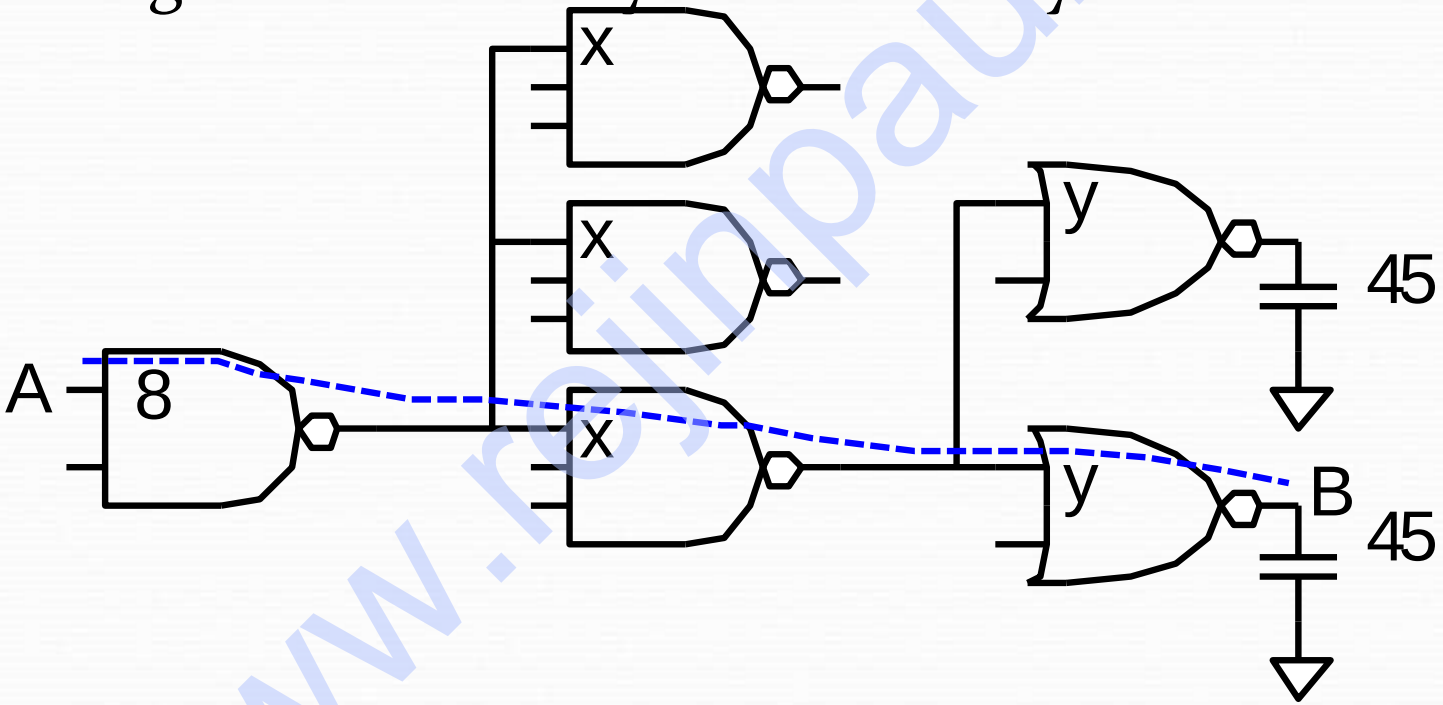
- Path effort→product of logical,electrical ,branching effort
  - F = GBH
  - Path Delay→ sum of delays in each stage

$$D = \sum d_i = D_F + P$$

www.rejinpaul.com

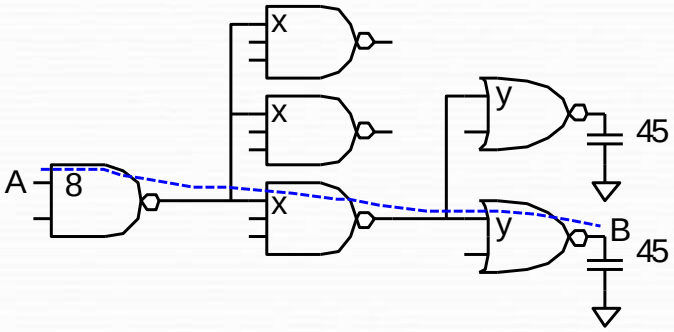
Example: 3-stage path

- Select gate sizes x and y for least delay from A to B



www.rojinpaul.com

Example: 3-stage path



Logical Effort G = (4

Electrical Effort H = 4

Branching Effort B = 3

Path Effort F = G

Best Stage Effort $\hat{f} = \sqrt[3]{}$

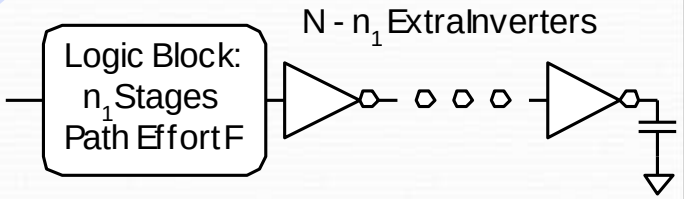
Parasitic Delay P = 2 + 3 + 2 =

Delay D = (gh+p)=3*5 + 7 = 22 = 4.4 FO4

www.rejinpaul.com

Choosing Best Number of Stages

- F→ path effort
- N→ Number of inverters
- Delay of new path is given by



$$D = NF^{\frac{1}{N}} + \sum_{i=1}^{n_1} p_i + (N - n_1)\, p_{inv}$$

$$\frac{\partial D}{\partial N} = -F^{\frac{1}{N}} \ln F^{\frac{1}{N}} + F^{\frac{1}{N}} + p_{inv} = 0$$

$$\rho = F^{\frac{1}{N}}$$

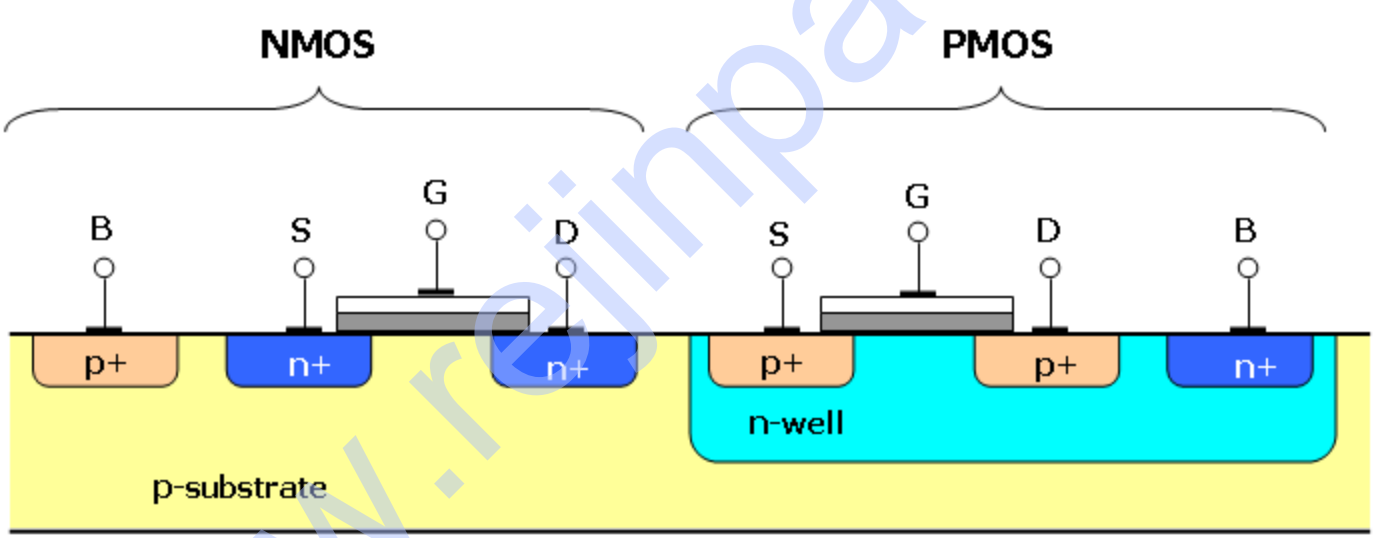
$$p_{inv} + \rho(1 - \ln \rho) = 0$$

ρ = 2.718

www.rejinpaul.com

CMOS Fabrication Technologies

- Complementary MOS technology uses both nmos and pmos transistor



www.rejinpaul.com

- CMOS can be fabricated using different process such as;
- N-Well process for CMOS fabrication
- P-Well process
- Twin tup process

www.rejinpaul.com

- CMOS can be obtained by integrating both the NMOS and PMOS transistor on same chip substrate.
- For integration of these PMOS and NMOS device on the same chip, special region called as wells or tubs are required.
- P-Well has to be created on N-Substrate or n-Well has to be created on P-Substrate

www.rejinpaul.com

Step 1: Substrate

- Primarily start with a P-Substrate

www.rejinpaul.com

Step2:Oxidation

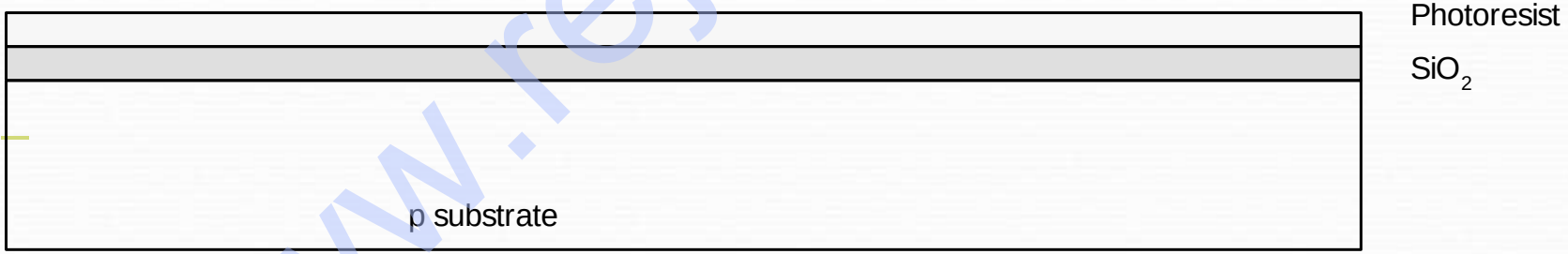
- The oxidation process is done by using high purity oxygen and hydrogen, which are exposed in an oxidation furnace approximately at 1000 degree centigrade.

www.rejinpaul.com

Step3:Photoresist

- **Spin on photoresist**
  - Photoresist is a light-sensitive organic polymer
  - Property changes where exposed to light

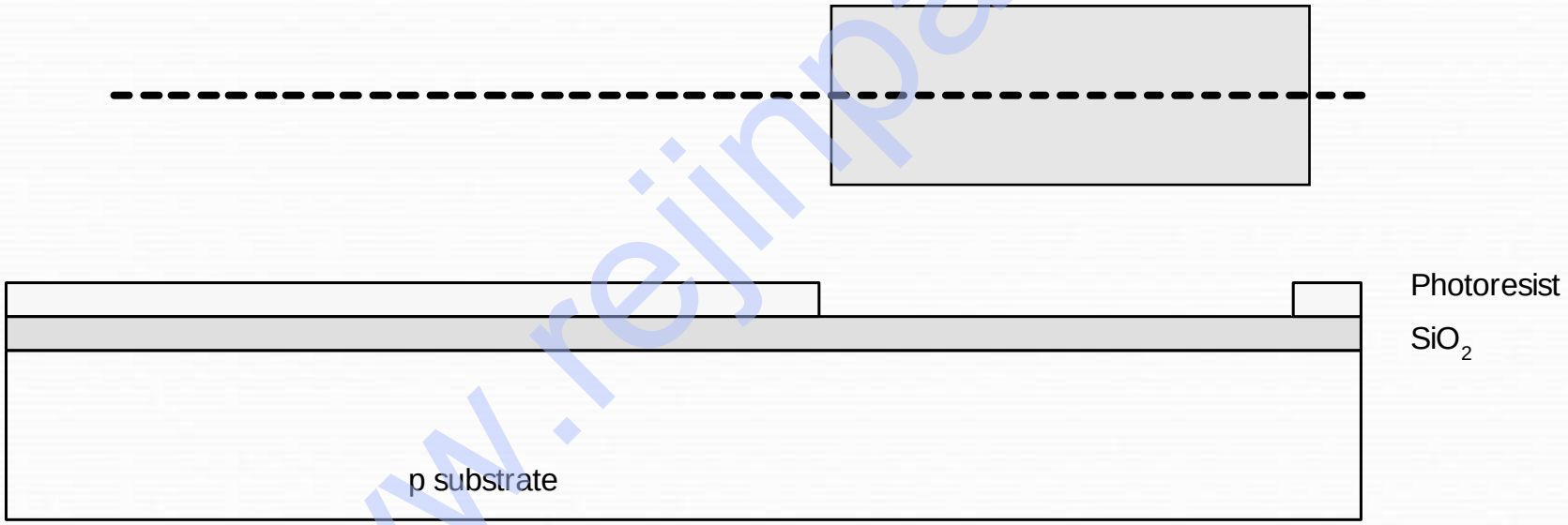
- Two types of photoresists (positive or negative)
  - **Positive resists** can be removed if exposed to UV light
  - **Negative resists** cannot be removed if exposed to UV light



www.rejinpaul.com

Step4: Masking

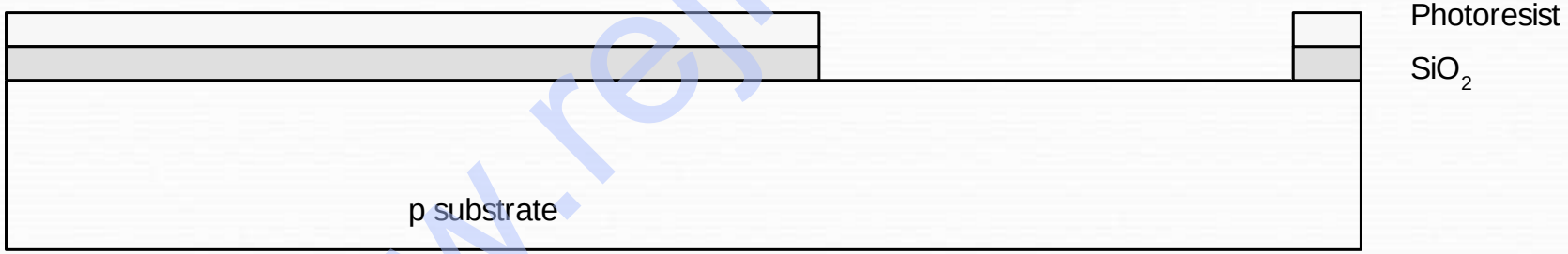
The photoresist is exposed to Ultra-violate (UV) light through the n-well mask



www.rejinpaul.com

Step5:Photoresist Removal(Etching)

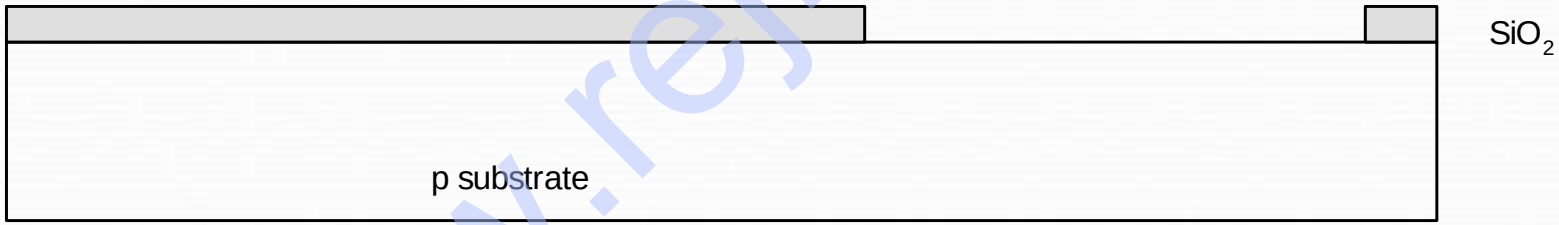
- A photo resist and SiO2 is removed by treating a wafer with the basic hydrofluoric acid solution



www.rejinpaul.com

Step6:Strip Photoresist

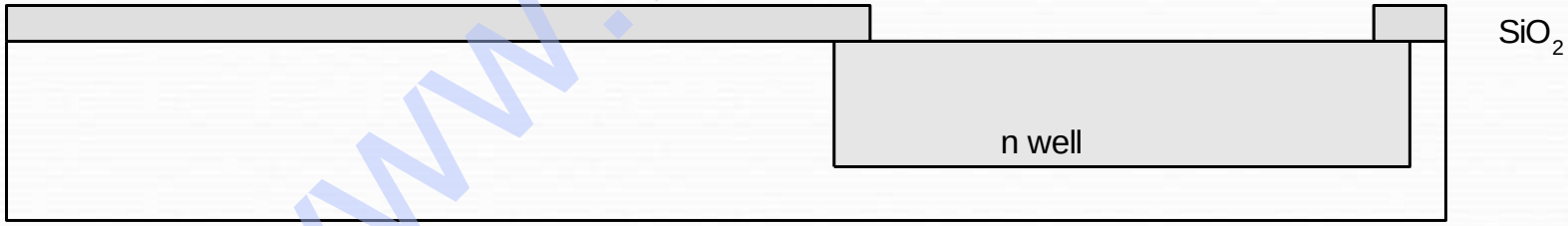
- The entire photoresist layer is Strip off



www.rejinpaul.com

Step7: formation of n-well

- **Diffusion**
  - Place wafer in furnace with arsenic-rich gas
  - Heat until As atoms diffuse into exposed Si
- **Ion Implanatation**
  - Introducing impurity atoms into the semiconductor crystal.



www.rejinpaul.com

Step8:Removal of SiO2 layer

- Using hydrofluoric acid the remaining SiO2 layer is removed

www.rejinpaul.com

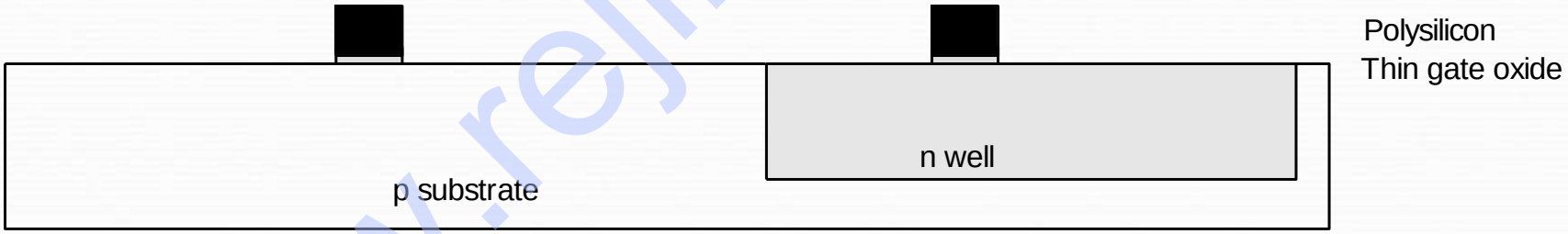
Step9: Deposition of Polysilicon

- **Deposit very thin layer of gate oxide**
  - < 20 Å (6-7 atomic layers)
- Chemical Vapour Deposition (CVD) process is used to deposit a very thin layer of gate oxide

www.rejinpaul.com

Step10:Polysilicon Patterning

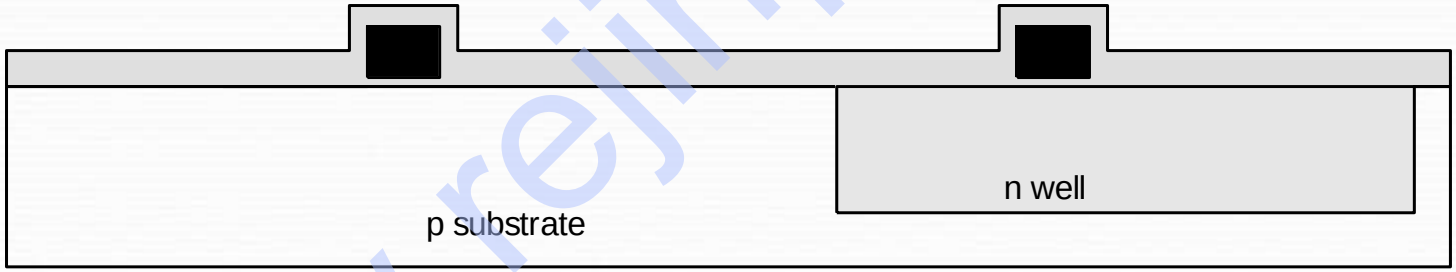
- Except a two small regions is required for forming the gates of NMOS and PMOS, the remaining layer is stripped off.



www.rejinpaul.com

Step11:Oxidation Process

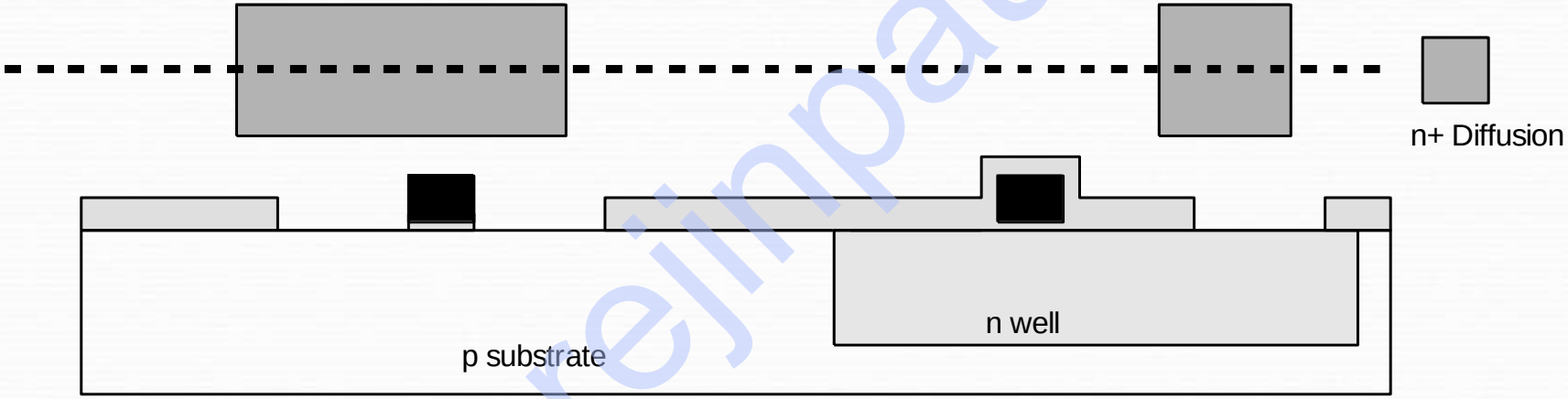
- Next, an oxidation layer is formed on this layer with two small regions for a formation of gate terminals of NMOS and PMOS

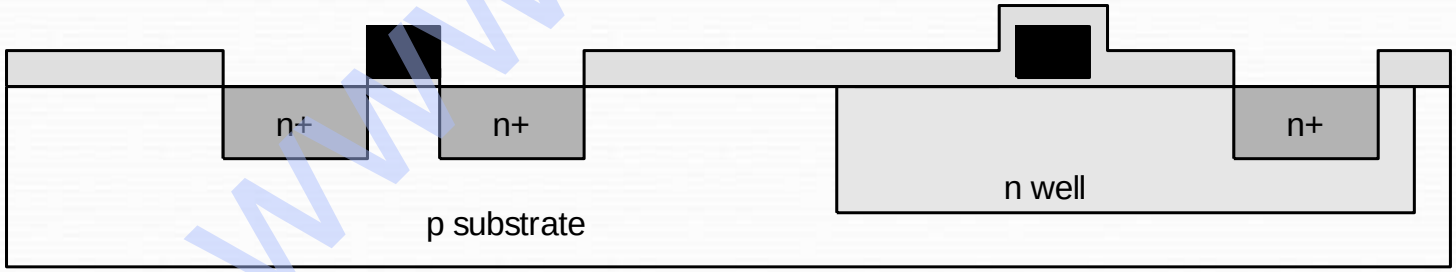


www.rejinpaul.com

Step12:Masking and N-diffusion

- By using masking process small gaps are made for the purpose of N-diffusion
- The n-type (n+) dopants are diffused or ion implanted, and the three n+ are formed for the formation of the terminals of NMOS

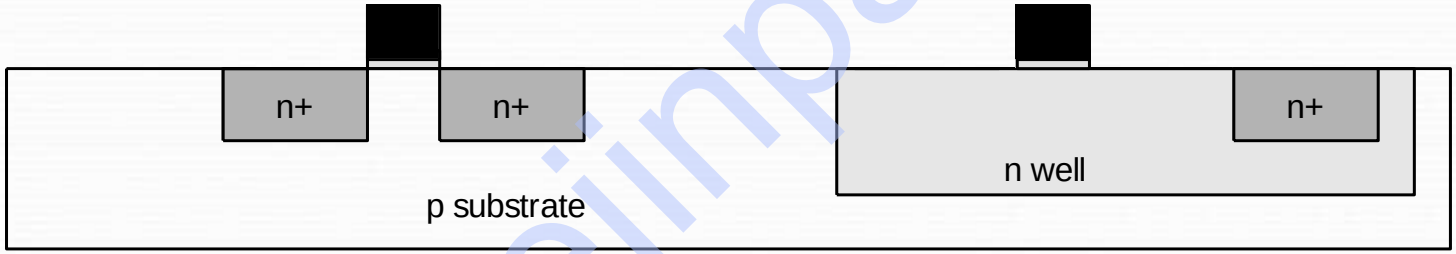




www.rejinpaul.com

Step13:Oxide Stripping

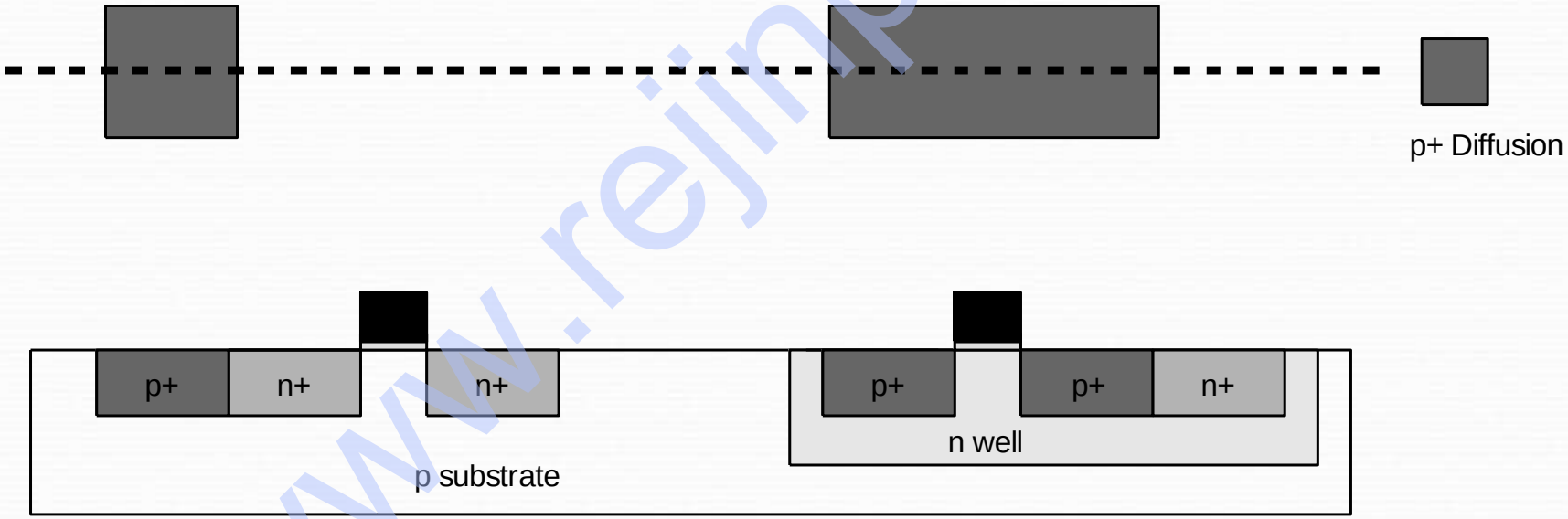
- The remaining oxidation layer is stripped off



www.rejinpaul.com

Step14:P-Diffusion/implantation

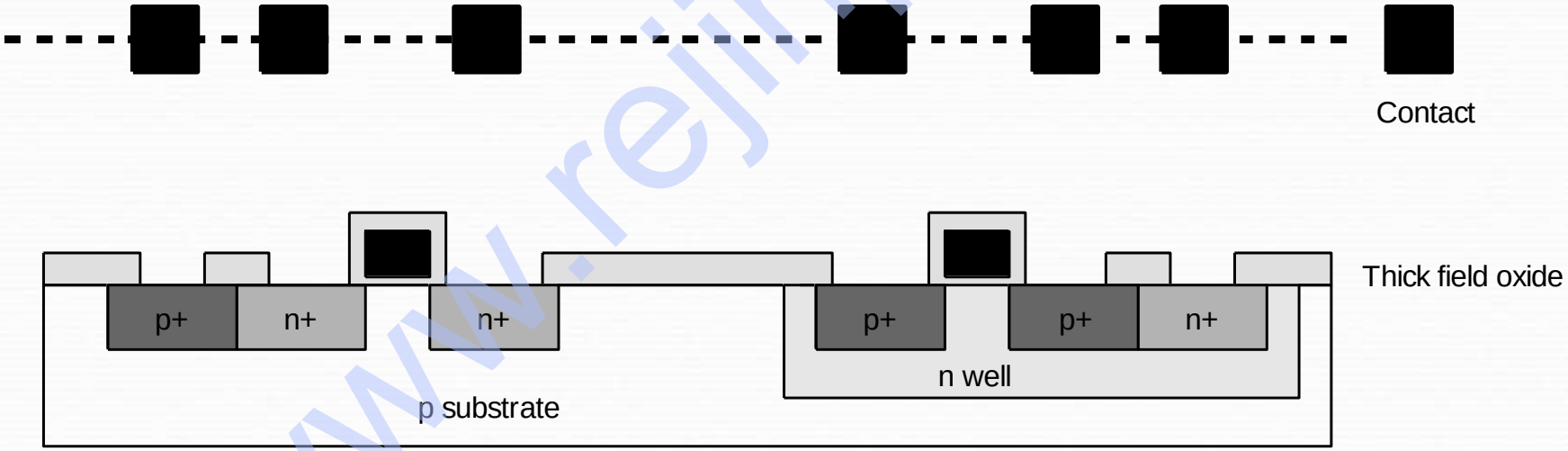
- Similar to n diffusion process , the P- diffusions regions are diffused to form the terminals of the PMOS



www.rejinpaul.com

Step15:Thick Field oxide

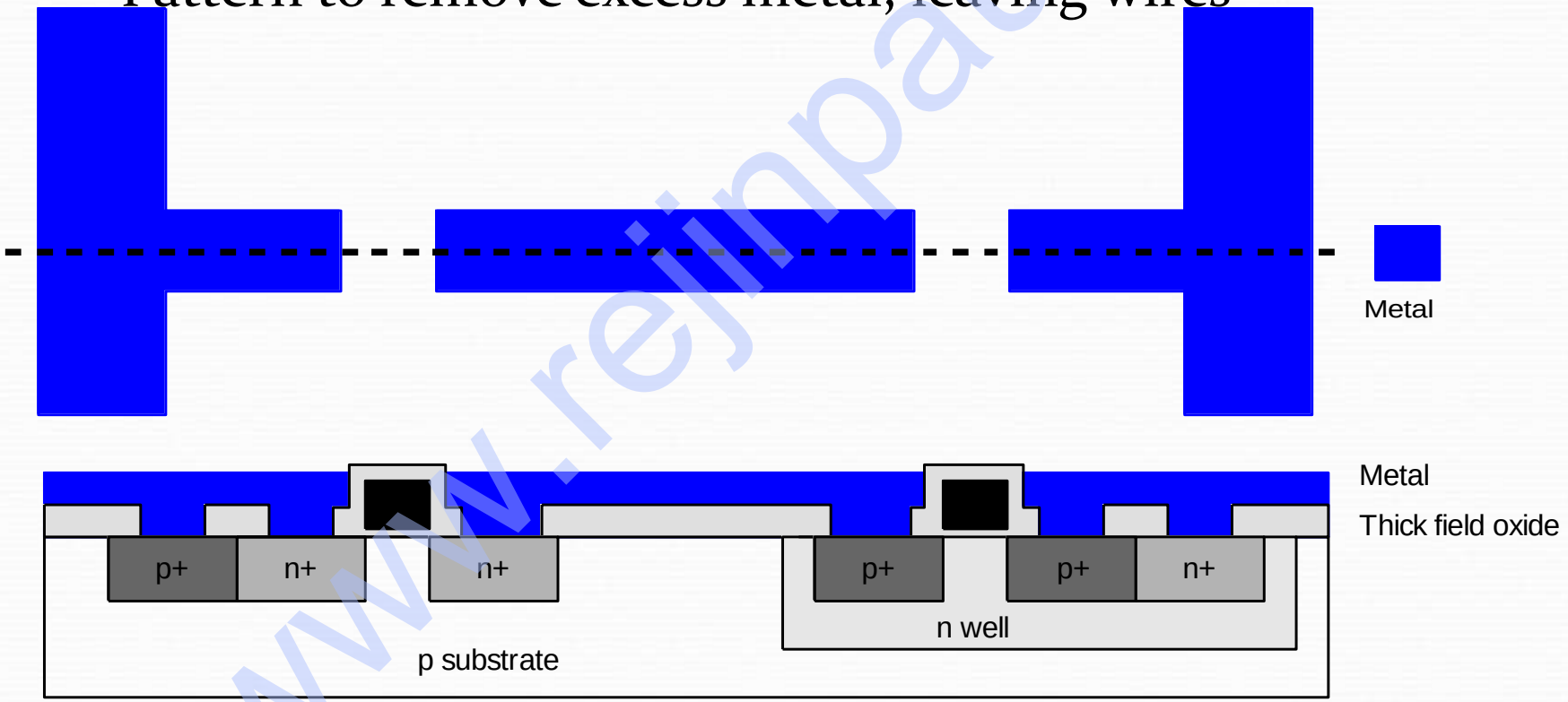
- A thick field oxide is formed an all regions except the terminals of the PMOS and NMOS



www.rejinpaul.com

Step16:Metalization

- Sputter on aluminum over whole wafer
- Copper is used in newer technology
- Pattern to remove excess metal, leaving wires

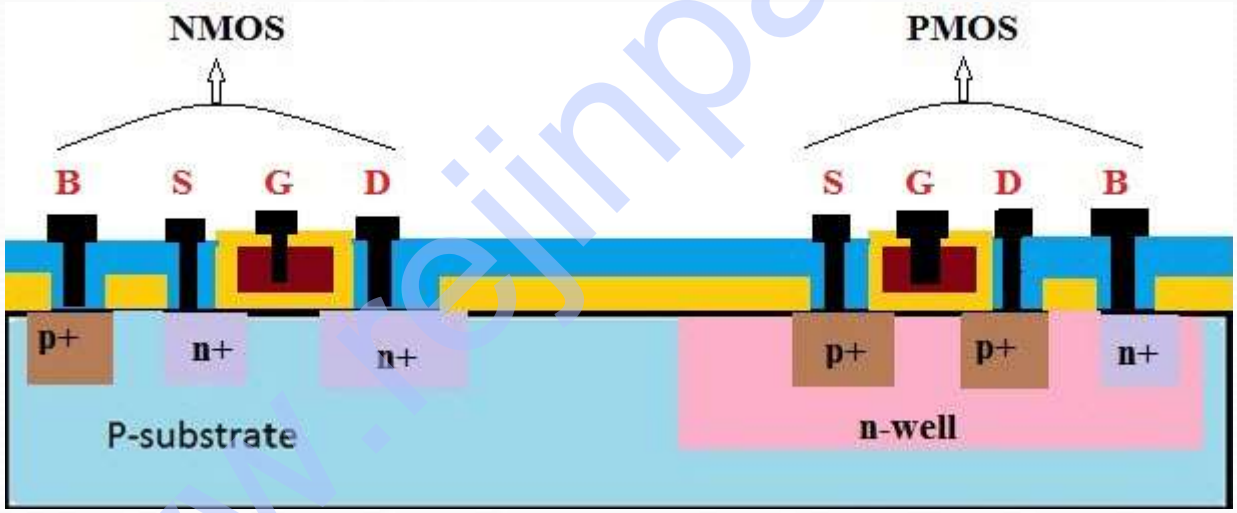


www.rejinpaul.com

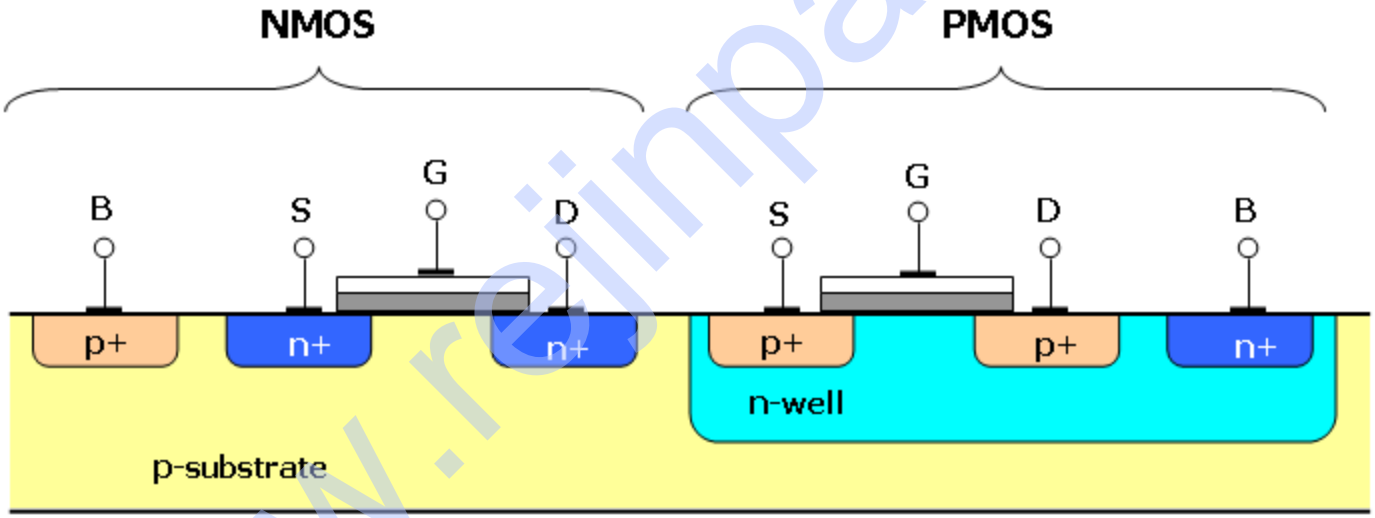
Step17:Terminals

www.rejinpaul.com

Assigning the Names

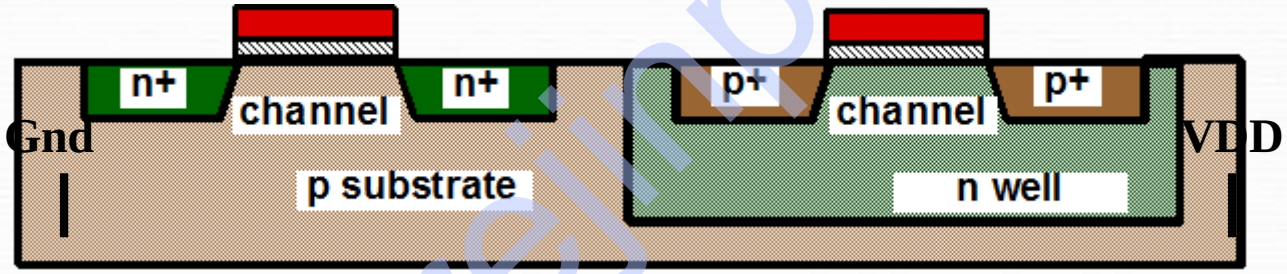


www.rejinpaul.com



www.rejinpaul.com

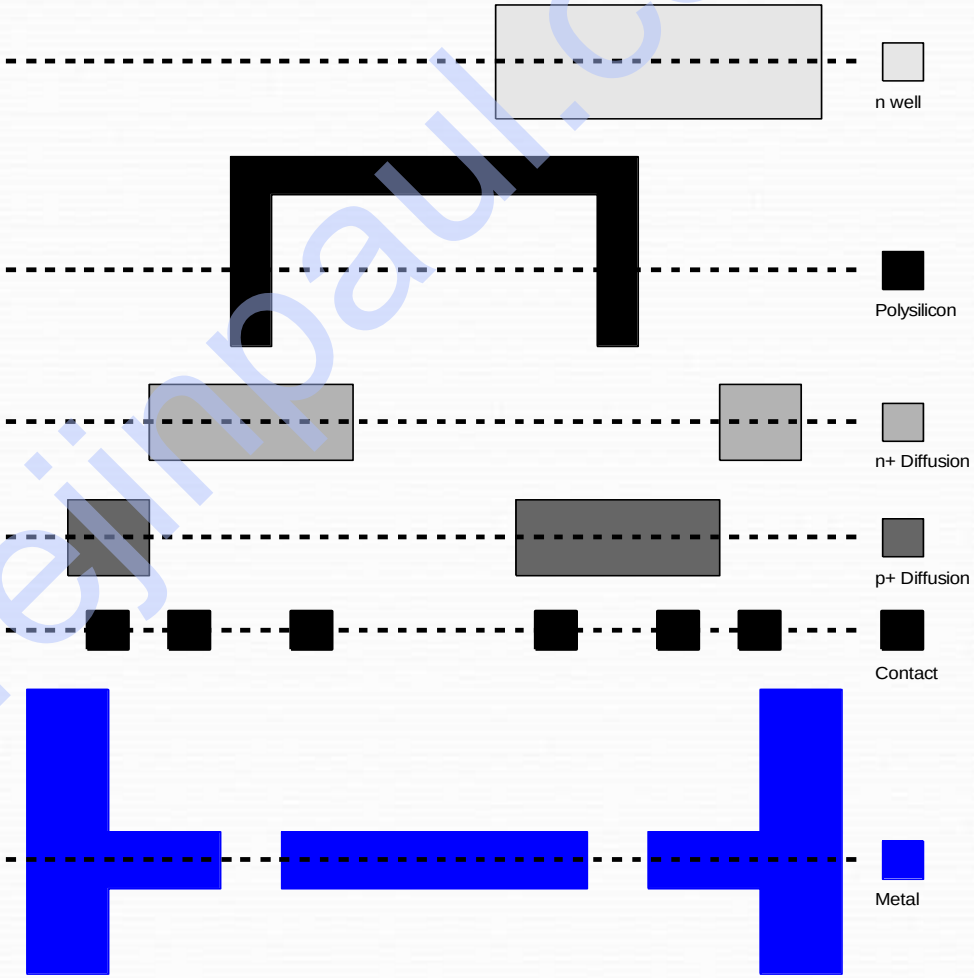
Detailed View - N-Well Process



www.rejinpaul.com

Detailed Mask Views

- n-well
- Polysilicon
- n+ diffusion
- p+ diffusion
- Contact
- Metal



In reality >40 masks may be needed

www.rejinpaul.com

Scaling Principles and fundamental limits

- Process of reduction of dimension of transistors and interconnection between them.
- 1970→gate length 10μm
- 2000→gate length 0.15 μm
- 13% reduction in gate length per year.

Advantages of Scaling:

1. Chip density increases
2. Chip performance increases
3. Parasitic capacitance reduced
4. Interconnect delay reduced
5. Chip cost is reduced

www.rejinpaul.com

1) Constant field scaling

- All device dimensions like channel length,width,supply voltage and threshold voltage reduced by a factor 1/s.
- Voltage V_{DD} is kept constant.
- Substrate doping increased by S.

www.rejimpaul.com

2) Lateral scaling

- Only the gate length is scaled
- Channel length reduced by the factor 1/S.
- If the feature size shrinks from 6m to 1m by maintaining a 5V constant supply volatge→constant voltage scaling.
- This will improve the delay performance and reduce the cost

www.rejinpaul.com

3)Interconnect Scaling

- Scale all dimensions
- Keep the wire height constant.
- The ratio of interconnect to gate delay increase with scaling.
-

www.rejinpaul.com

CMOS inverter scaling

- In CMOS inverter scaling has great impact with the following.

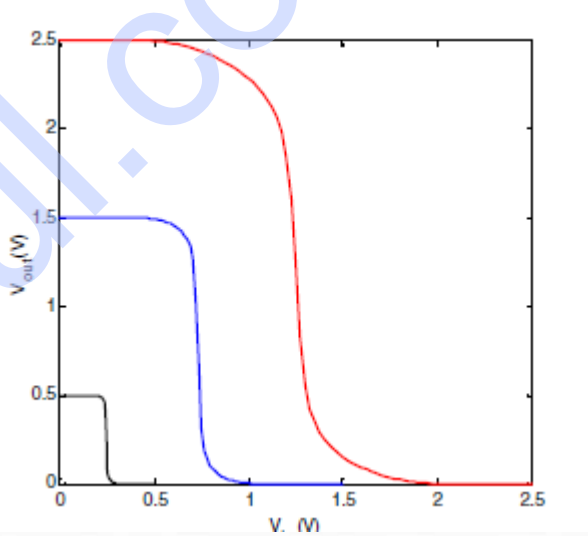
1. Supply voltage

- Total power dissipation of inverter depends on the reduction of V_{DD}
- Static characteristics of the CMOS inverter significant variation of the supply voltage without affecting the functionality of the inverter.
- If the power supply voltage is reduced below the threshold voltages, the voltage transfer characteristics where none of the transistors is conducting.

1. Power and Area Considerations

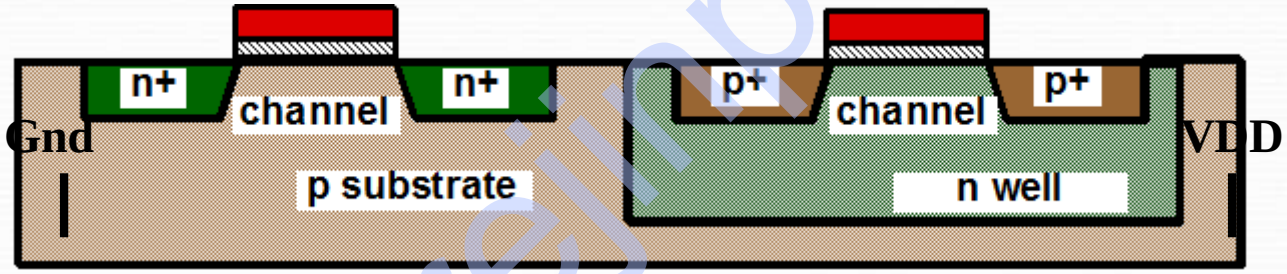
- DC power dissipation in CMOS inverter is almost negligible.
- Overall power consumption is low over other circuits.

V_{tp}=0.5v



www.rejinpaul.com

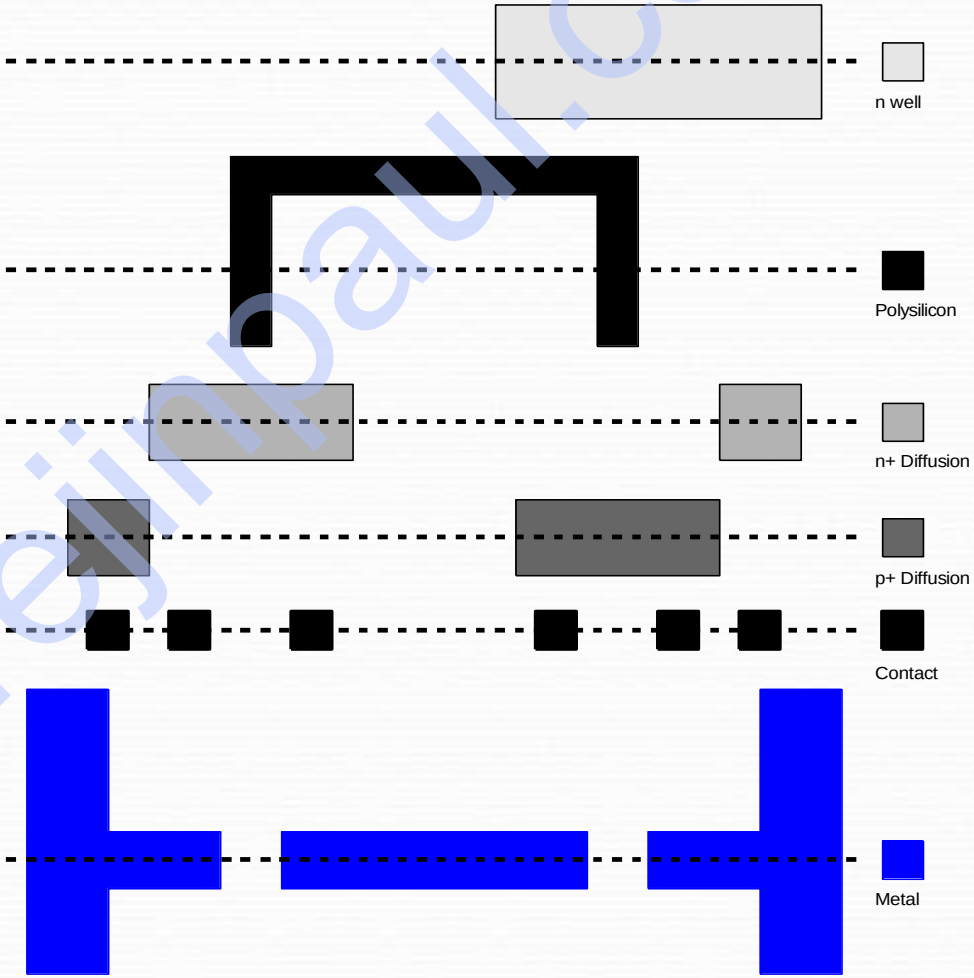
Detailed View - N-Well Process



www.rejinpaul.com

Detailed Mask Views

- n-well
- Polysilicon
- n+ diffusion
- p+ diffusion
- Contact
- Metal



In reality >40 masks may be needed

www.rejinpaul.com

Layout Diagram

- Layout can be very time consuming
  - Design gates to fit together nicely
  - Build a library of standard cells
- Standard cell design methodology
  - V_{DD} and GND should abut (standard height)
  - Adjacent gates should satisfy design rules
  - nMOS at bottom and pMOS at top
  - All gates include well and substrate contacts

www.rejinpaul.com

1)Micron Rules

- All min sizes and spacing specified by microns
- Standard in industry.
- Don't have the multiples of λ.

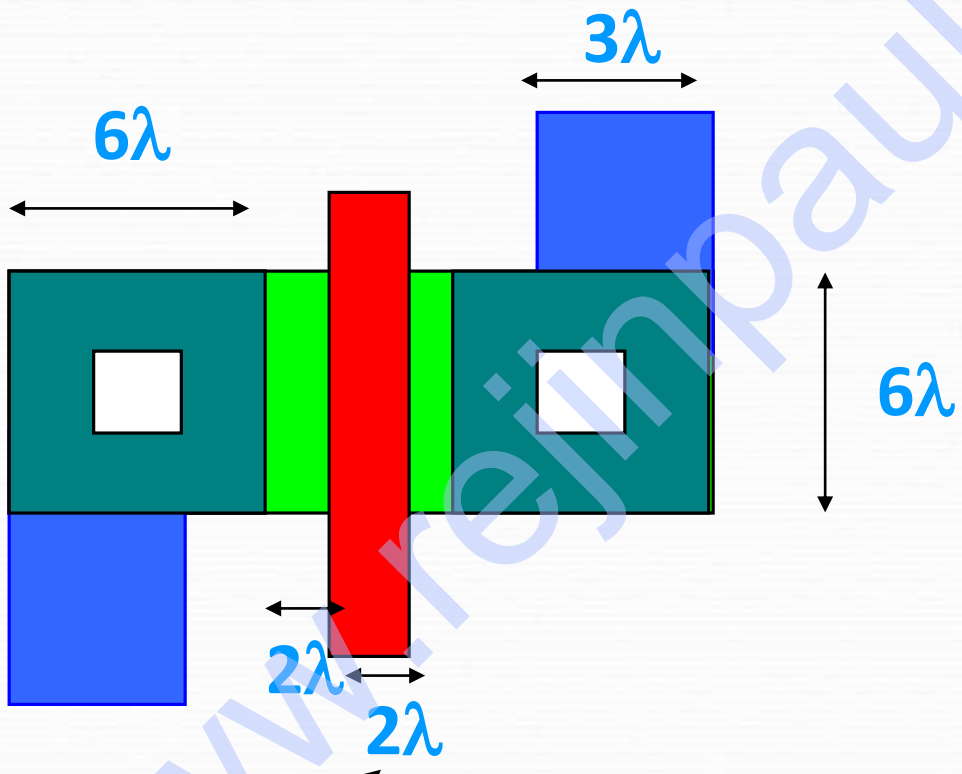
www.rejinpaul.com

2)Lambda-based Design Rules

- Design based on parameter λ.
- Simple designing procedure.
- Min feature size is 2 λ.
- If design is accepted, mask produced on the working ckt.

www.rejinpaul.com

Layout Design rules & Lambda (λ)



All device mask dimensions are based on multiples of λ, e.g., polysilicon minimum width = 2λ. Minimum metal to metal spacing = 3λ

www.rejinpaul.com

3)Transistor Rules

- P-diffusion and n-diffusion should be 3.
- Spacing between active should be 3
- Source and drain surrounded by 5.
- Substrate-well contact 3
- Width of polysilicon 2
- Spacing to poly over field oxide is 2

www.rejimpaul.com

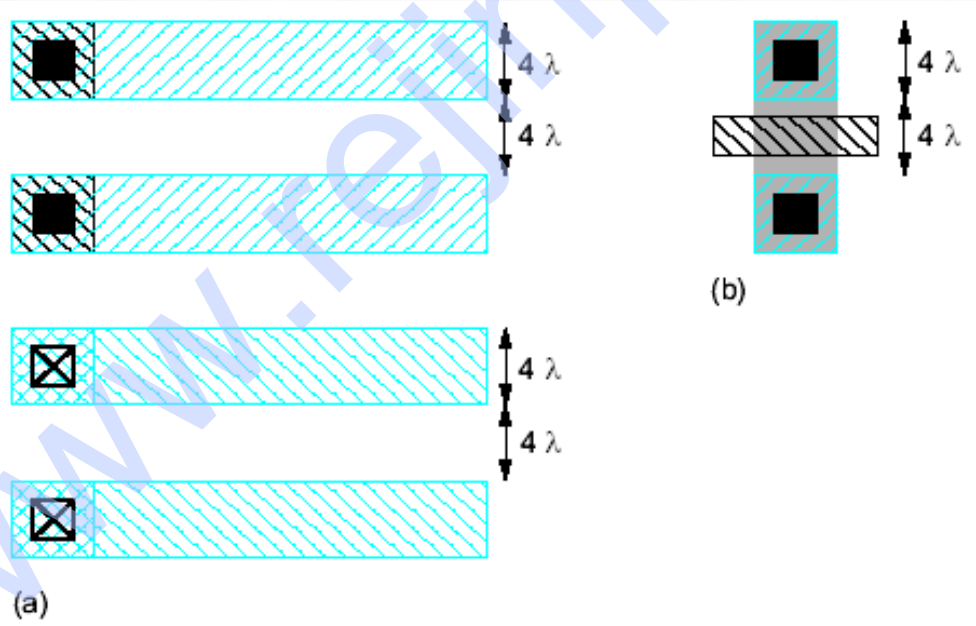
4)Contact Rules

- Contact size is 2x2
- Spacing to contact is 2
- Spacing to gate 2

www.rejinpaul.com

5)Wiring Tracks

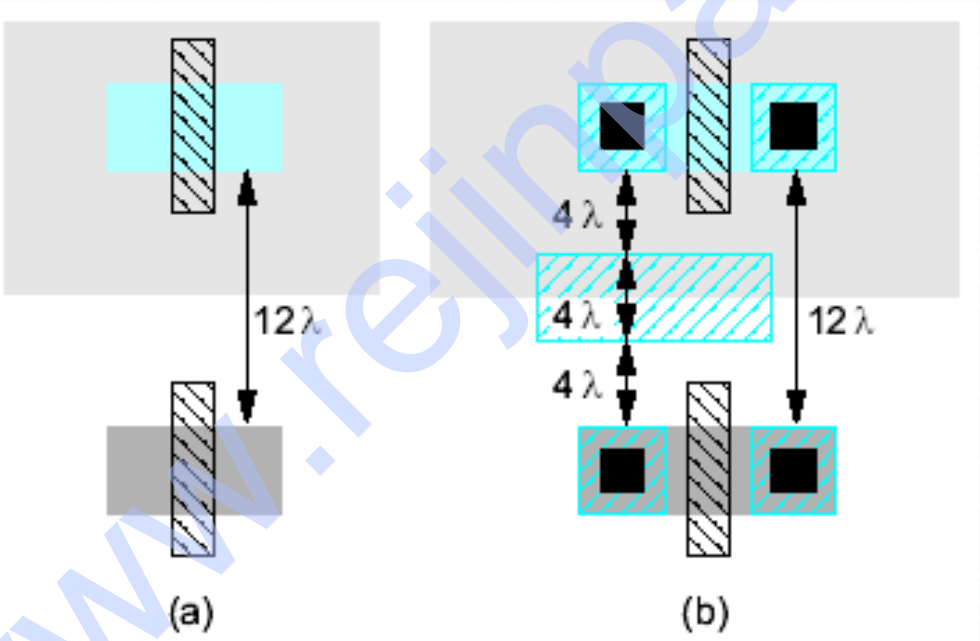
- A *wiring track* is the space required for a wire
  - 4 λ width, 4 λ spacing from neighbor = 8 λ pitch
- Transistors also consume one wiring track



www.rejinpaul.com

6)Well spacing

- Wells must surround transistors by 6 λ
  - Implies 12 λ between opposite transistor flavors
  - Leaves room for one wire track



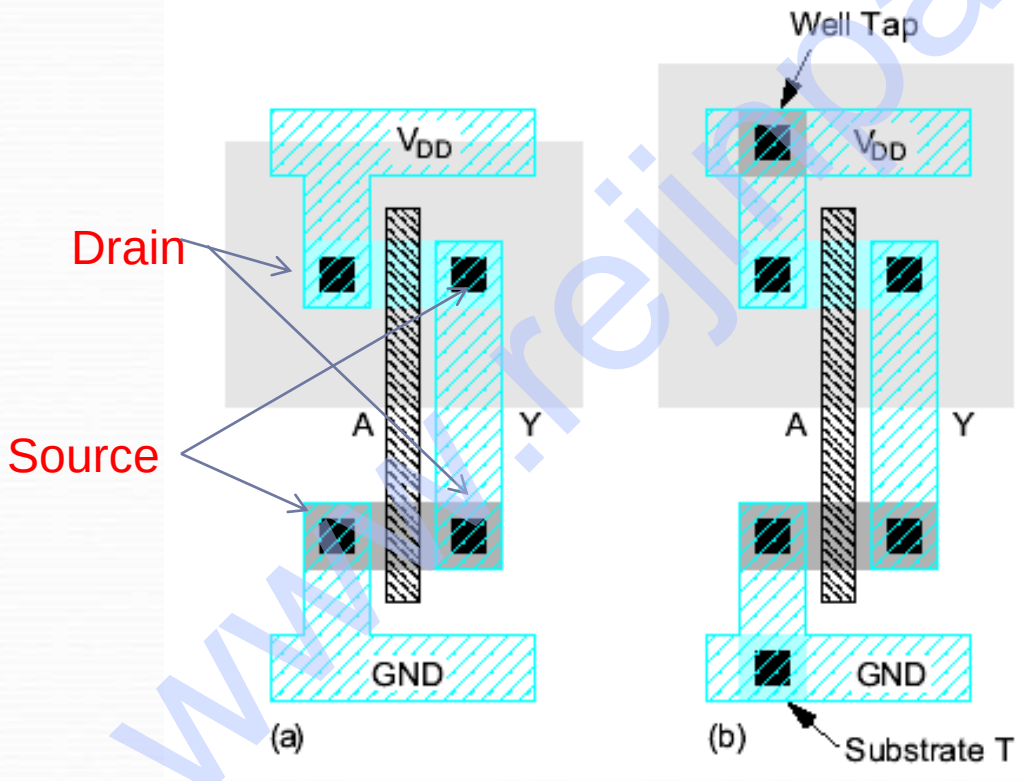
www.rejimpaul.com

Example: Inverter Layout

- The input for the inverter A is connected from the top,botom or left in poly silicon.
- The output for the inverter Y is obtained at the right side of the cell in metal.
- The p-substrate and n-well is tied to ground and V_{DD} respectively.

Inverter Layout

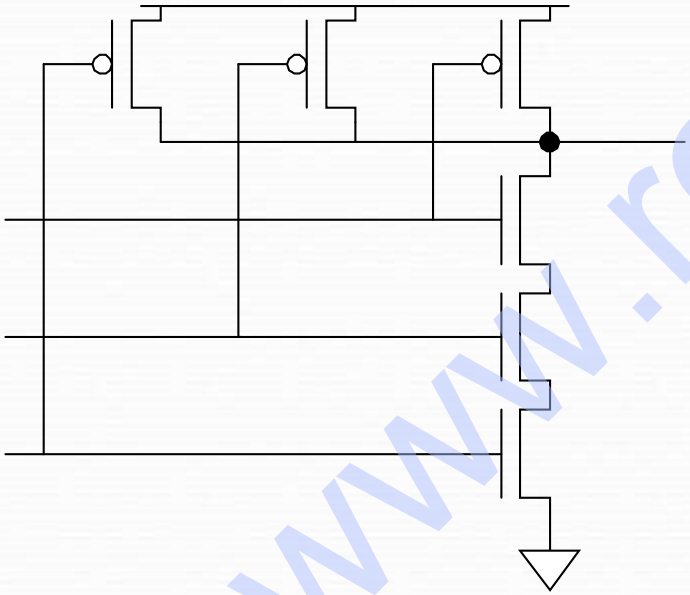
Inverter Layout with well and substrate taps placed under V_{DD}

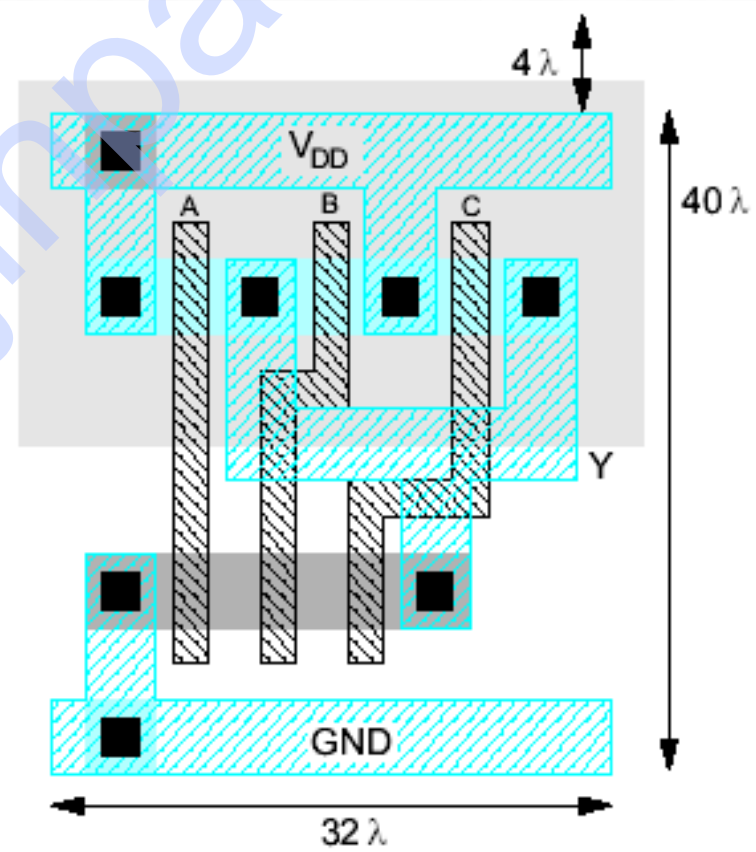


www.rejinpaul.com

3-input NAND-Layout

- n-MOS connected in series.
- p-MOS in parallelHorizontal N-diffusion and p-diffusion strips
- Vertical polysilicon gates
- Metalı V_{DD} rail at top
- Metalı GND rail at bottom
- 32 λ by 40 λ





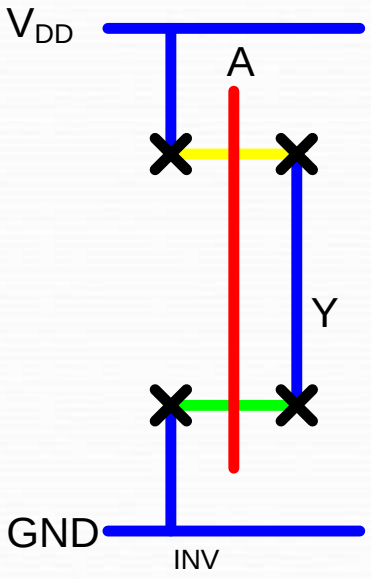
www.rejinpaul.com

Stick Diagrams

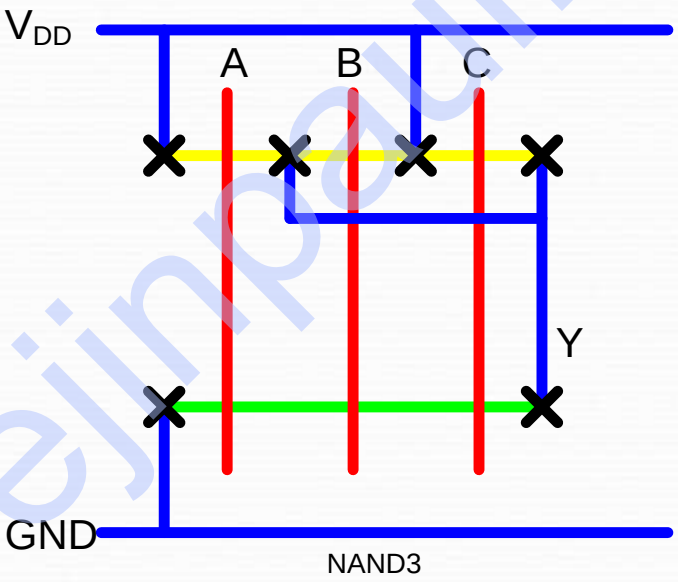
- VLSI design aims to translate circuit concepts onto silicon
- stick diagrams are a means of capturing topography and layer information - simple diagrams
- Stick diagrams convey layer information through color codes (or monochrome encoding)

www.rejinpaul.com

Stick Diagrams



Inverter



3-I/P NAND

www.rejinpaul.com

Example:3

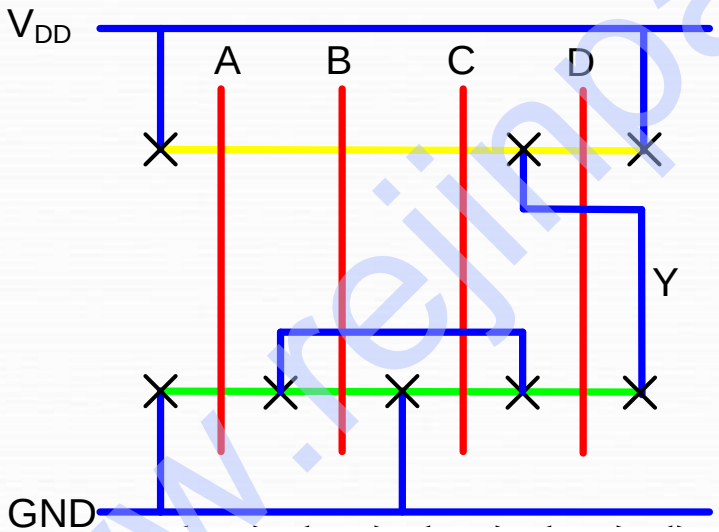
- Sketch a stick diagram for the following

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

www.rejinpaul.com

Example: O3AI

- Sketch a stick diagram for O3AI and estimate area
  - $Y = \overline{(A + B + C) \cdot D}$



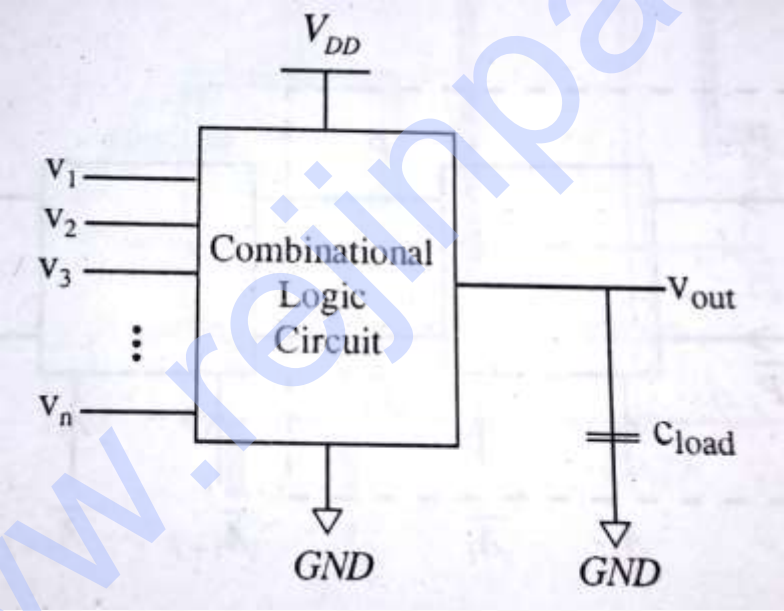
www.rejinpaul.com

Unit-II
COMBINATIONAL LOGIC CIRCUITS

www.rejinpaul.com

Combinational Logic

- Time-Independent logic
- No memory Unit
- The output of the combinational circuit is a function of current or Present Input



www.rejinpaul.com

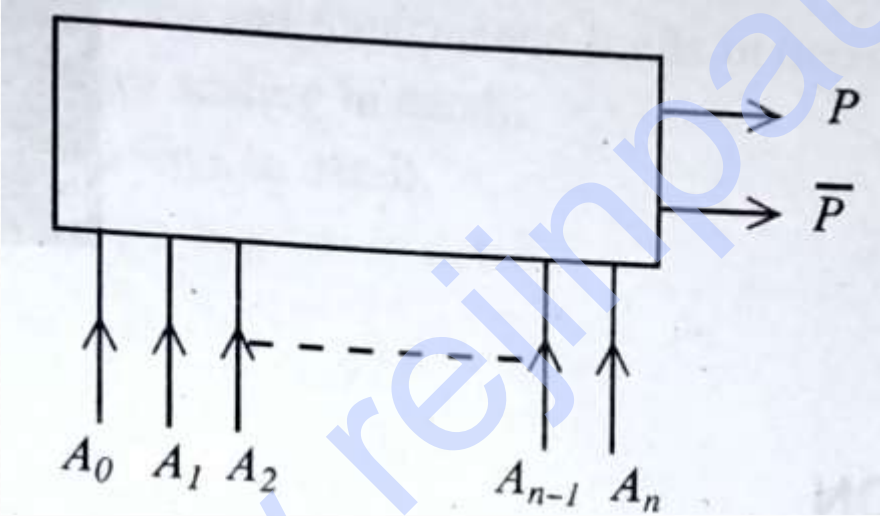
Examples of Combinational Logic Design

1. Parity Generator
2. Multiplexer
3. 4-Line gray code to binary converter

www.rejinpaul.com

1)Parity Generator

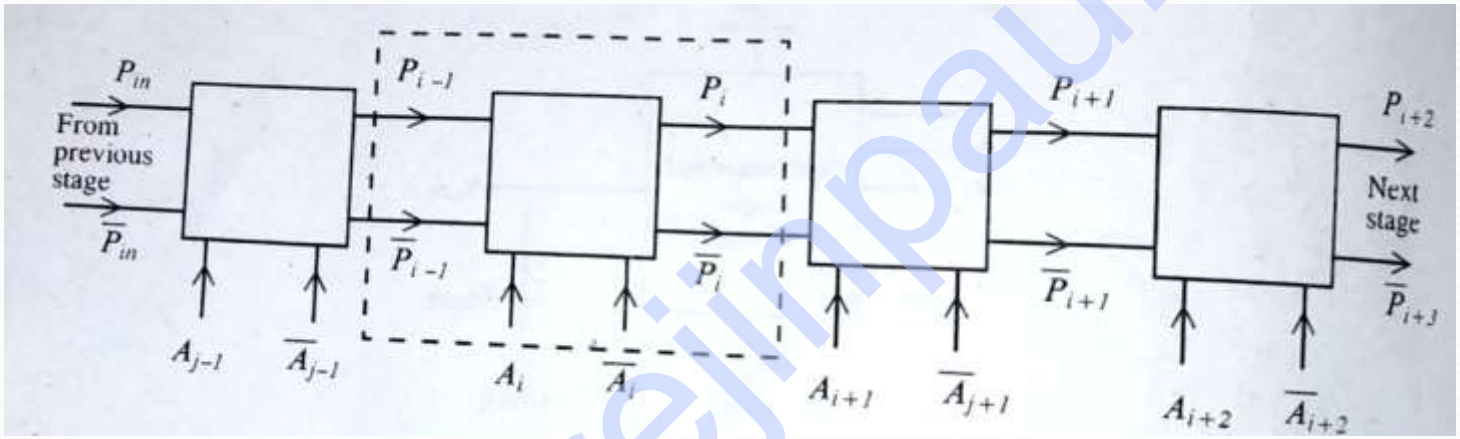
- Parity generator indicates the parity of a binary number or word.
- The value of P is 1 ,when the input has even number of 1's.
- The value of P is 0 ,when the input has odd number of 1's.



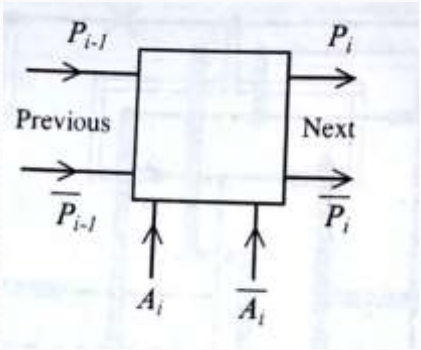
www.reimpaul.com

Structured design approach Parity Generator

- The parity information is passed from one cell to the next and is changed not by a cell.
- The change is made based on the state of the input lines Ai and $\bar{A}i$



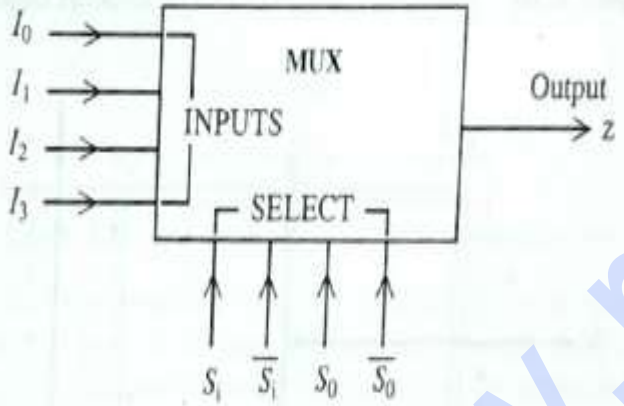
- If A is 1 ,the parity is changed $p_i=\overline{p}_{i-1}$
- If A is 0 ,the parity is un-changed $p_i=p_{i-1}$



www.rejinpaul.com

2)Multiplexer

- It is also called as data selectors.
- Multiplexers are designed using TTL and other logic families.
- It has one output,n inputs and 2^n control lines.
- Output $Z = I_0.\bar{S}_1.\bar{S}_0 + I_1.\bar{S}_1.S_0 + I_2.S_1.\bar{S}_0 + I_3.S_1.S_0$

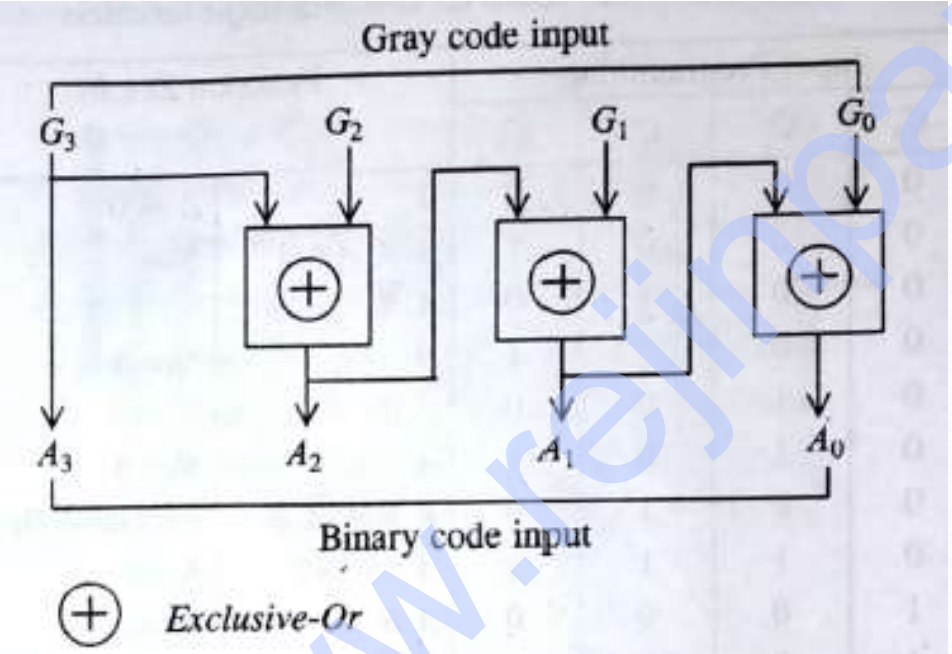


S_1	S_0	Z
0	0	I_0
0	1	I_1
1	0	I_2
1	1	I_3

www.rejinpaul.com

3) 4-Line gray code to binary converter

- G3,G2,G1,Go →Gray inputs
- A3,A2,A1,Ao→Binary code output



$$A_0 = \overline{G}_0.A_1 + G_0.\overline{A}_1$$
$$A_1 = \overline{G}_1.A_2 + G_1.\overline{A}_2$$
$$A_2 = \overline{G}_2.A_3 + G_2.\overline{A}_3$$
$$A_3 = G_3$$

www.rejinpaul.com

Truth Table

Gray code				Binary code			
G_3	G_2	G_1	G_0	A_3	A_2	A_1	A_0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	1
0	0	1	1	0	0	1	0
0	0	1	0	0	0	1	1
0	1	1	0	0	1	0	0
0	1	1	1	0	1	0	1
0	1	0	1	0	1	1	0
0	1	0	0	0	1	1	1
1	1	0	0	1	0	0	0
1	1	0	1	1	0	0	1
1	1	1	1	1	0	1	0
1	1	1	0	1	0	1	1
1	0	1	0	1	1	0	0
1	0	1	1	1	1	0	1
1	0	0	1	1	1	1	0
1	0	0	0	1	1	1	1

www.reinpaul.com

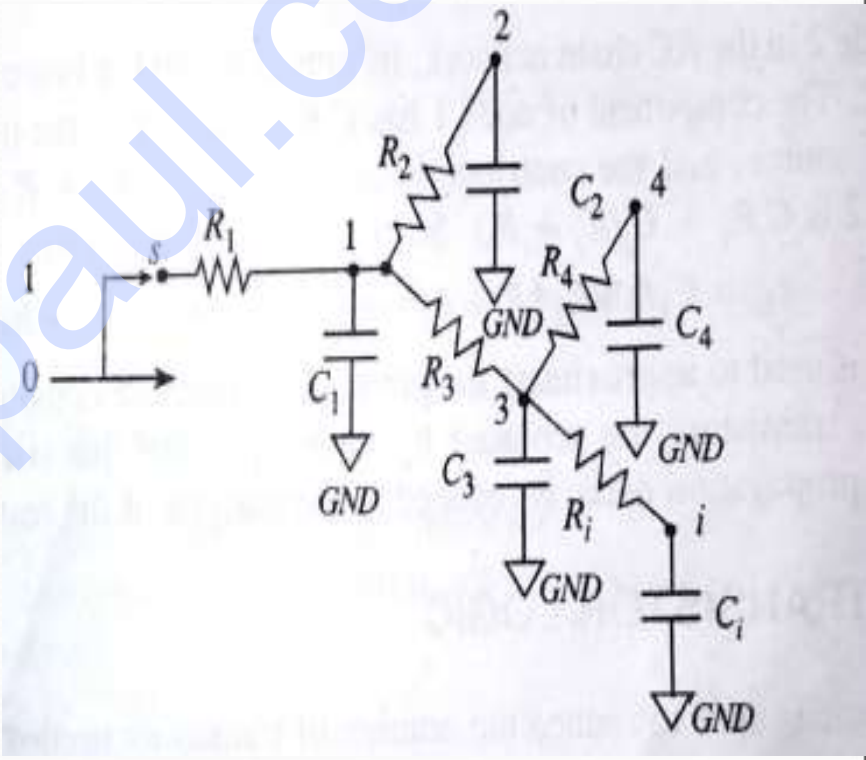
Elmore's delay-Tree N/W

RC tree has a following properties

1. N/W has a single input node
2. All capacitors are between a node and ground
3. N/W doesn't have any resistive loops

- Path resistance between source node s and 4 is given by $R_{44}=R_1+R_3+R_4$
- Step input is applied at node s at time t=o,Elmore delay at node i is given by

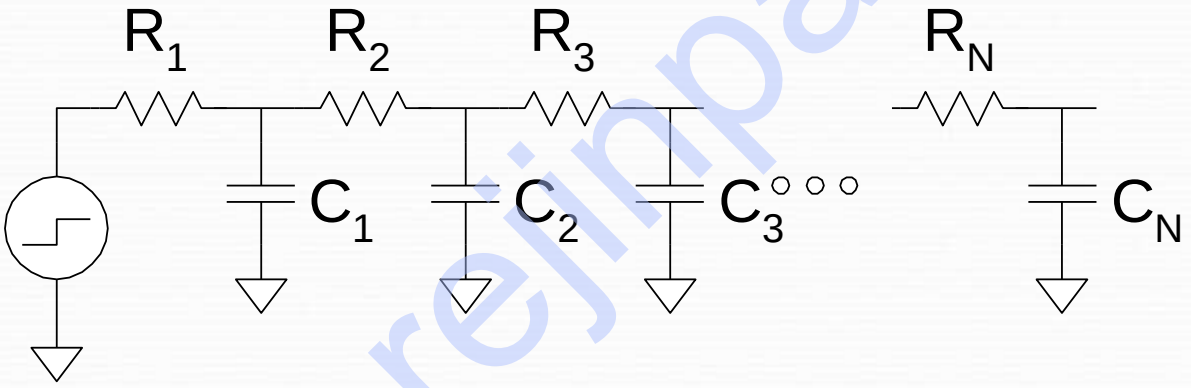
$$\tau_{Di} = \sum_{k=1} C_k R_{ik}$$



www.rejinpaul.com

Elmore's Constant-Rc ladder N/W

- Used to estimate delay value in RC ladder circuit
- ON transistors look like resistors.
- Pullup or pulldown network modeled as *RC ladder*
- Delay is equal to the sum over each node in the ladder of the resistance R_{n-i} between the node and is multiplied by the capacitance on the node.



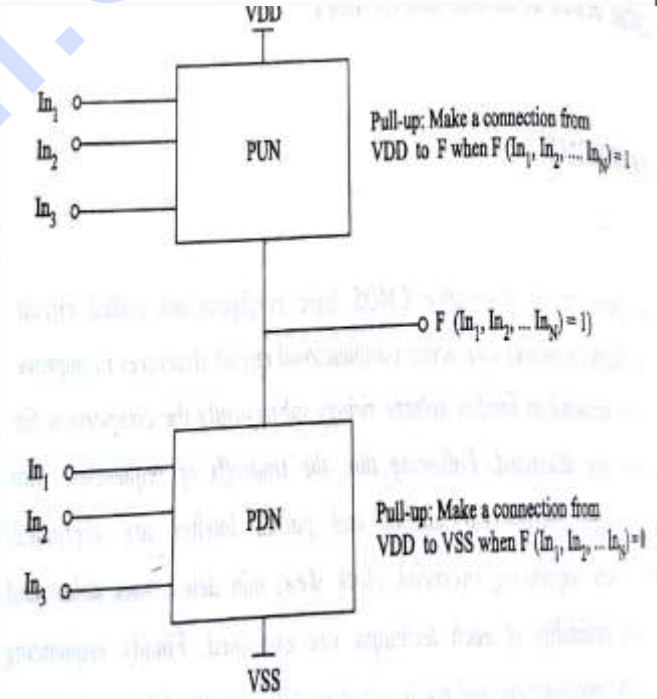
$$t_{pd} \approx \sum_{\text{nodes } i} R_{i-to-source} C_i$$

$$= R_1 C_1 + (R_1 + R_2) C_2 + \ldots + (R_1 + R_2 + \ldots + R_N) C_N$$

www.rejinpaul.com

Static CMOS Design

- The static CMOS style is really an extension of the static CMOS inverter to multiple inputs.
- CMOS structure is low sensitivity to noise, good performance, and low power consumption with no static power dissipation.
- It is a combination of pull-up network(PMOS) & pull down network(NMOS).
- Each gate output is connected to either V_{DD} or V_{SS}
- The PUN and PDN networks are constructed in a mutually (*one and only one of the networks is conducting in steady state).*
- Path always exists between V_{DD} *and the output F, realizing a high output ("one"),* or, alternatively, between V_{SS} *and F for a low output ("zero").*



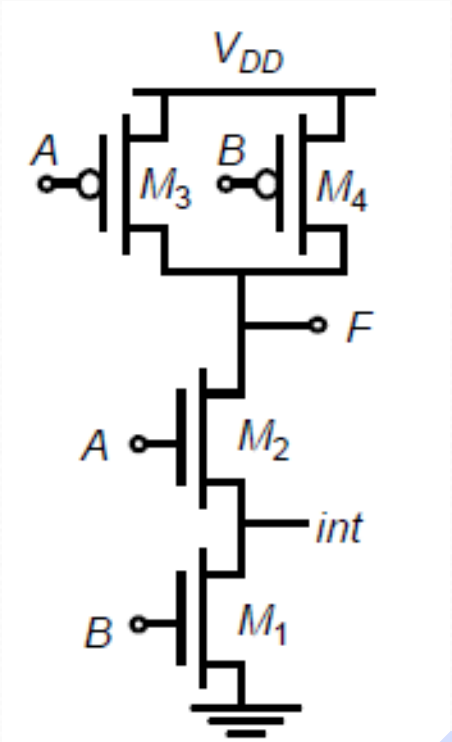
www.reinpaul.com

Important Properties of static CMOS Design(PDN-PUN)

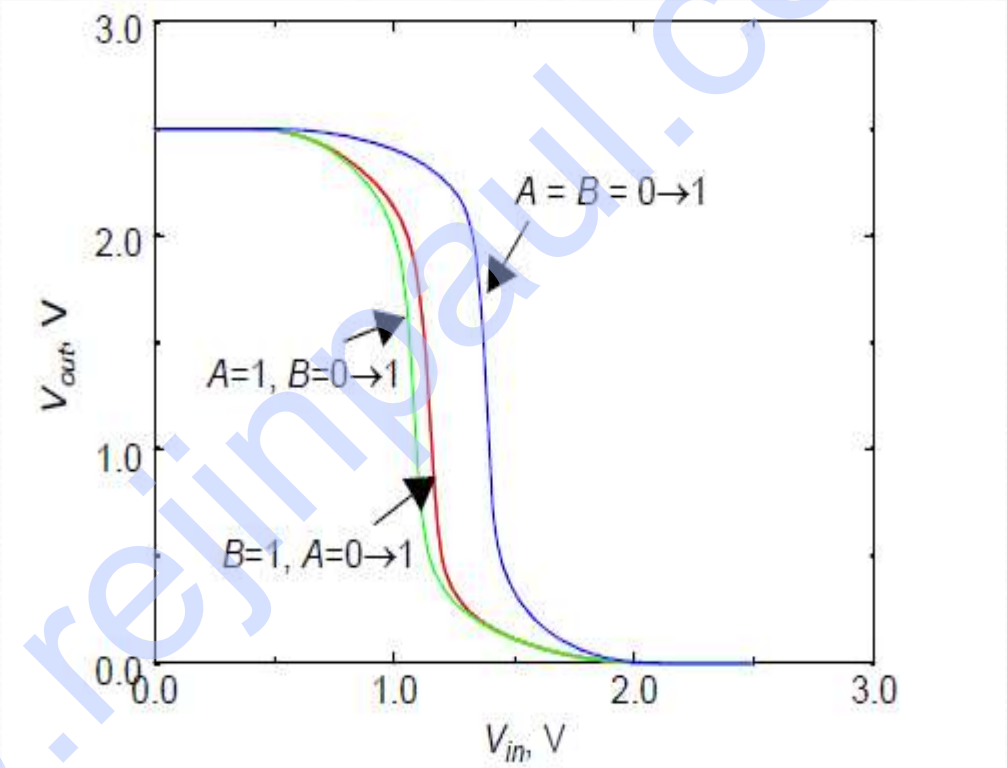
1. A transistor can be a switch controlled by its gate signal. An NMOS switch is *on when the controlling signal is high and is off when the controlling signal* is low.
2. A PMOS transistor acts as an inverse switch that is *on when the controlling* signal is low and *off when the controlling signal is high*.
3. PDN is constructed using NMOS , PUN is constructed using PMOS devices and NMOS transistors produce "strong zeros," and PMOS devices generate "strong ones.
4. NMOS devices connected in series for AND function .(all the inputs high, the series combination conducts the value at one end is transferred to the other end)
5. NMOS transistors connected in parallel for OR function.(A conducting path exists between the output and at least one of the inputs is high.)
6. PMOS devices connected in series for NOR function (both inputs are low, conduct output)
7. PMOS devices connected in parallel for NAND function (one input is low, conduct output)
8. The complementary gate is naturally *inverting, implementing only functions such as* NAND, NOR, and XNOR
9. The number of transistors required to implement an *N-input logic gate is 2N.*

www.rejinpaul.com

Static Properties of Complementary CMOS (NAND)



2-INPUT NAND Gate



VT Curve

- Three possible input combinations switch the output of the gate from high-to-low:(a) $A = B = 0 \rightarrow 1$, (b) $A = 1, B = 0 \rightarrow 1$, and (c) $B = 1, A = 0 \rightarrow 1$.

www.rejinpaul.com

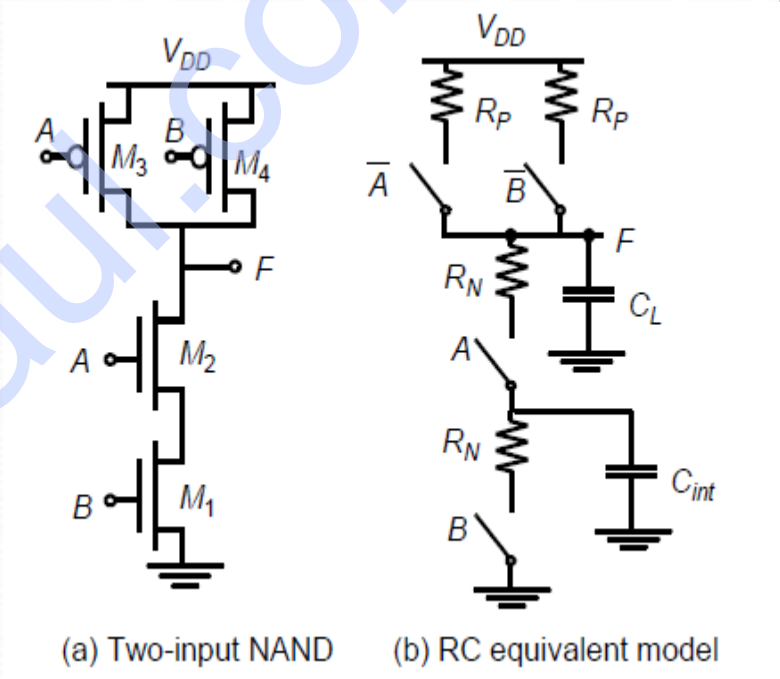
Voltage Transfer Curve(VTC)

- case (a)→both transistors (M3,M4) in the PUN are on simultaneously for *A=B=0, representing a strong pull-up.*
- case (b&c)→only one of the pullup devices is on. The VTC is shifted to the left as a result of the weaker PUN.
- case (b)→ *M3 is turned off, and the gate voltage of M2 is set to VDD. M2 may be considered as a resistor in series with M1. Since the drive on M2 is large,* this resistance is small and has only a small effect on the voltage transfer characteristics.
- The threshold voltage of transistor *M2> M1. (VGS2 = VA - VDS1)*
- case (c)→ transistor M1 acts as a resistor, causing body effect in *M2. VGS1 = VB*

www.rejinpaul.com

Propagation Delay of Complementary CMOS Gates

- The computation of propagation delay similar to the static inverter.
- For delay analysis, each transistor is modeled RC switch level model.
- propagation delay depends upon the input patterns.
- If both inputs are driven low, the two PMOS devices are on in parallel. The delay is 0.69 x(*Rp/2) xCL*
- when only one device turns on, delay is 0.69 x *Rp xCL.*
- For the pull-down path, the output is discharged only if both *A and B are switched* high, and the delay is given by 0.69 x(2*RN) xCL*



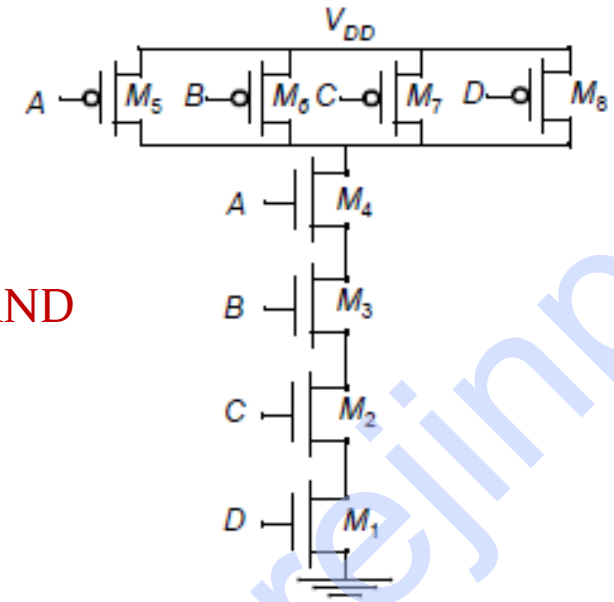
(a) Two-input NAND (b) RC equivalent model

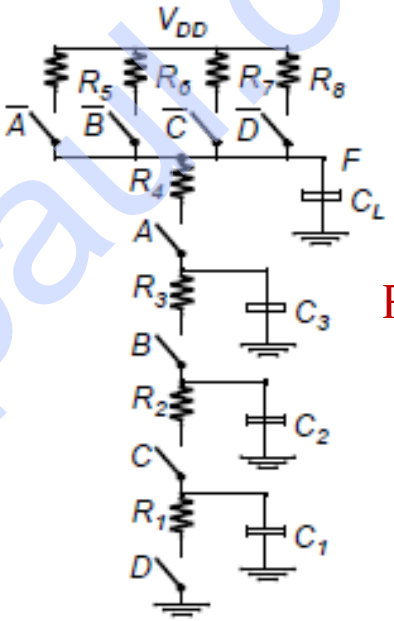
www.rejinpaul.com

Four input NAND gate and its RC model

- Consider the pull-down delay of the circuit. The output is discharged when all inputs are driven high.

4-INPUT NAND





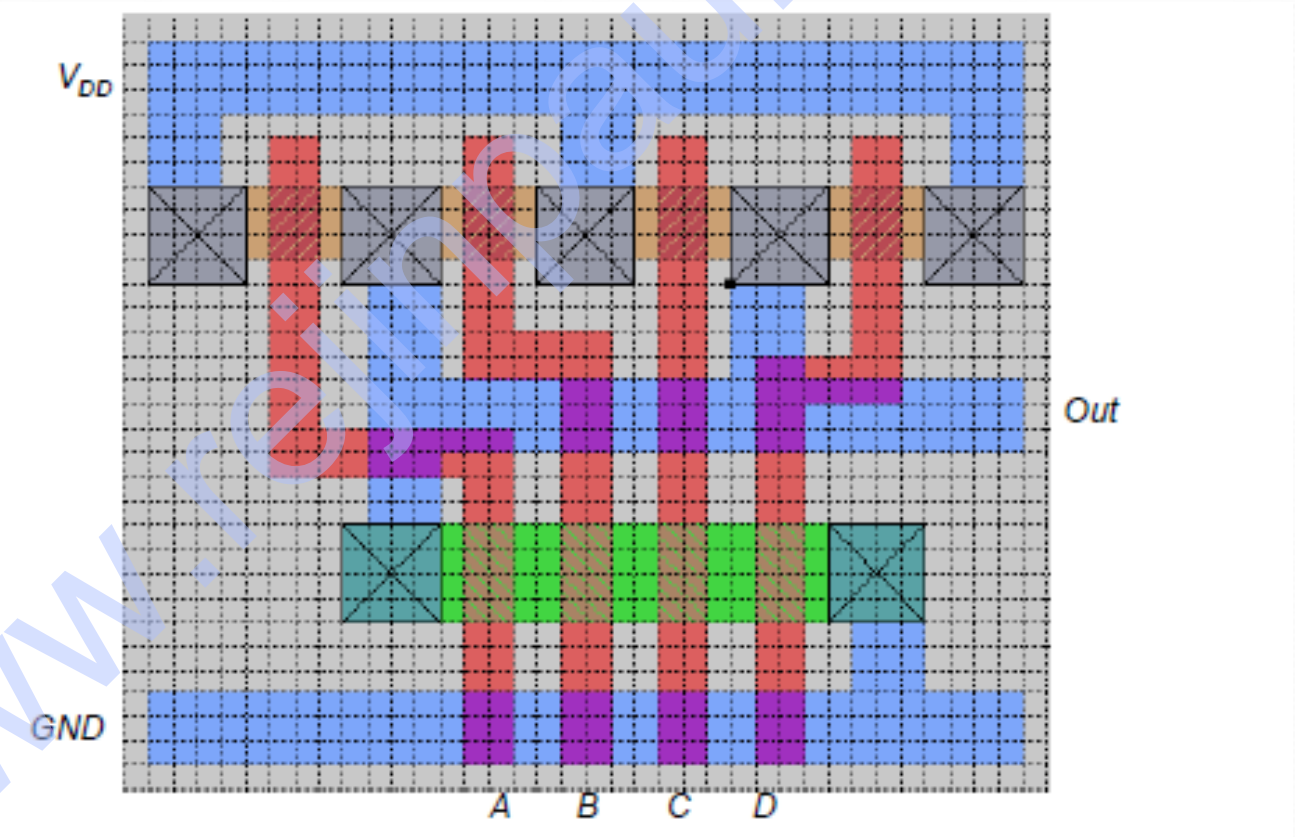
RC-delay model

- $tpHL = 0.69(R1 \times C1 + (R1 + R2) \times C2 + (R1 + R2 + R3) \times C3 + (R1 + R2 + R3 + R4) \times CL)$
- The resistance of *M1 appears in all the terms, which makes this device* especially important when attempting to minimize delay.
- Assume all NMOS devices have an equal size, $tpHL = 0.69RN(C1 + 2 \times C2 + 3 \times C3 + 4 \times CL)$

www.rejinpaul.com

Layoyt 4-input NAND gate

- Assume that all NMOS devices have a W/L of 0.5mm/0.25mm, and all PMOS devices have a device size of 0.375mm/0.25mm.
- PMOS devices share the drain terminal in order to reduce the overall parasitic contribution to the output



www.rejinpaul.com

Design Techniques for Large Fan-in

- Several approaches may be used to reduce delays in large fan-in circuits

1. Transistor Sizing
2. Progressive Transistor Sizing
3. Input Re-Ordering
4. Logic Restructuring

www.rejinpaul.com

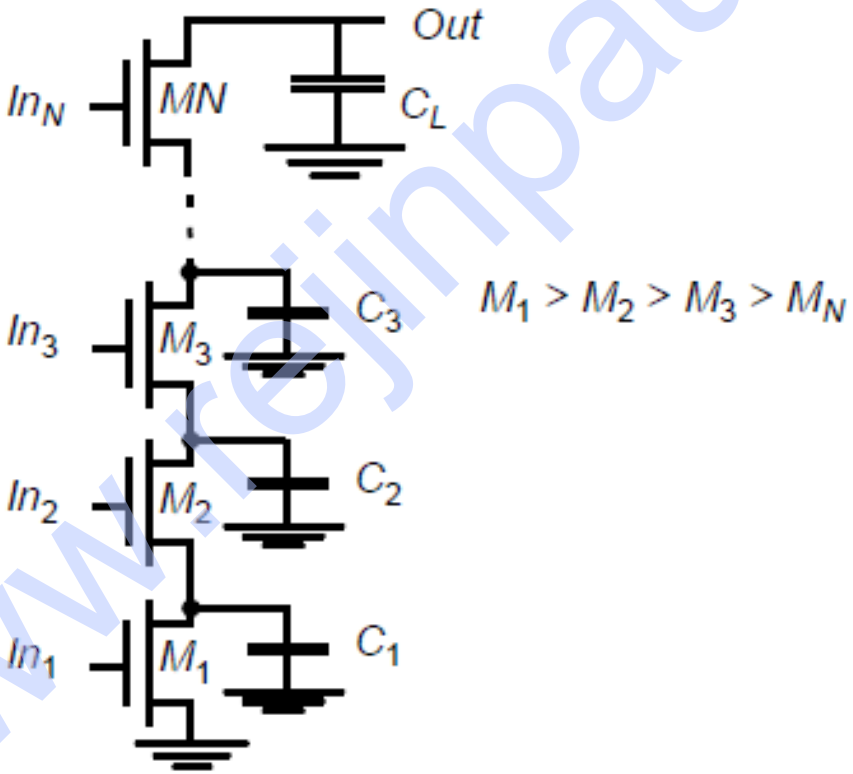
1. Transistor Sizing

- Increasing the transistor size, results in larger parasitic capacitance which will affect the *propagation delay*.
- It will create larger load to the preceding gate.
- Sizing is only effective when the load is dominated by the fan out.

www.rejinpaul.com

2) Progressive Transistor Sizing

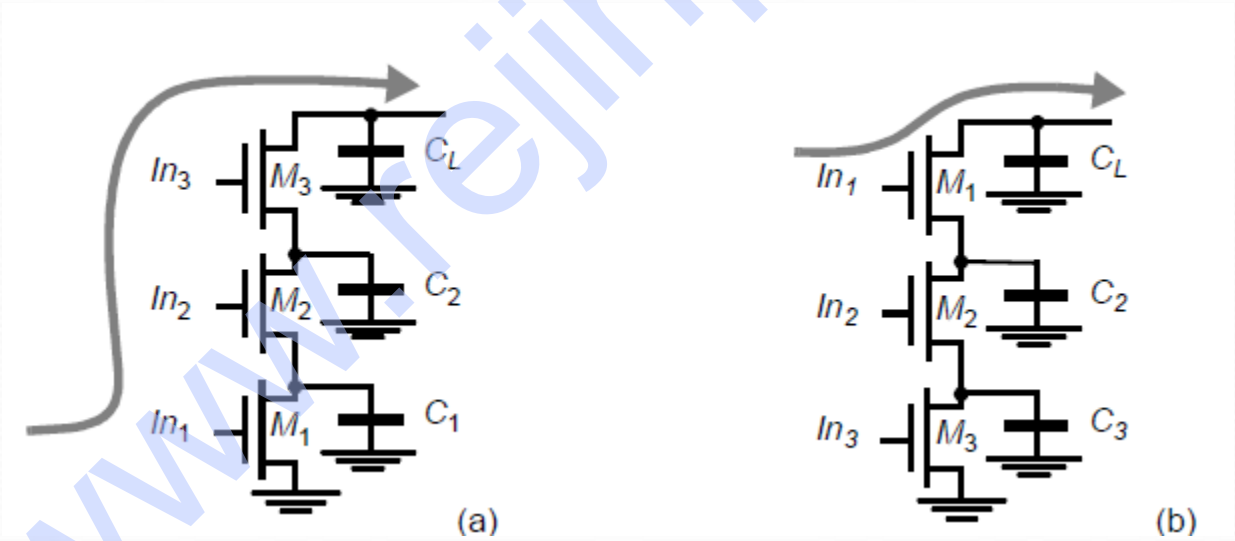
- An alternate approach to uniform sizing →Each transistor is scaled up uniformly.(*M1 > M2> M3 > MN)*
- In this approach, resistance is reduced while reducing capacitance



www.rejinpaul.com

3)Input Re-Ordering

- Some signals in complex combinational logic blocks might be more critical than others.
- Not all inputs of a gate arrive at the same time (due to propagation delays of the preceding logical gates)
- An input signal to a gate is called *critical if it is the last signal of* all inputs to assume a stable value.
- The path through the logic which determines the ultimate speed of the structure is called the *critical path.*
- For speeding up→ Putting the critical-path transistors closer to the output.



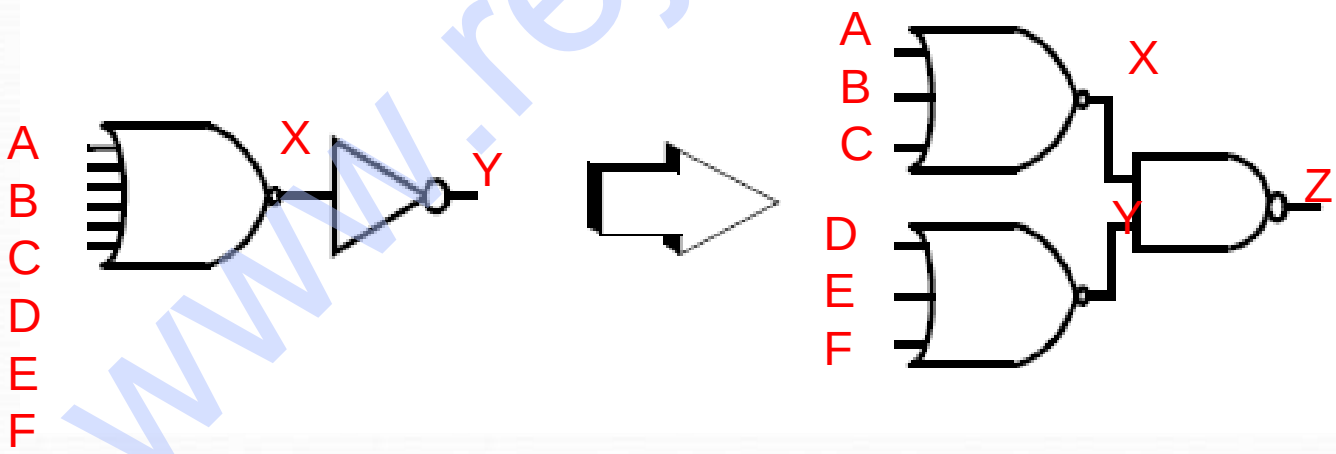
www.rejinpaul.com

- Signal In1 is assumed to be a critical signal. Suppose further that In2 and In3 are high and that In1 undergoes a 0→1 transition. Assume also that CL is initially charged high
- case (a)→ No path to GND exists until M1 is turned on, which is unfortunately the last event to happen. The delay between the arrival of In1 and the output is therefore determined by the time it takes to discharge CL, C1 and C2.
- Case(b)→C1 and C2 are already discharged when In1 changes. Only CL still has to be discharged, resulting in a smaller delay.

www.rejinpaul.com

4)Logic Restructuring

- Logic restructuring can reduce the gate fan-in.
- Changing the topology of a logic network may reduce its power dissipation
- Manipulating the logic equations can reduce the fan-in requirements and hence reduce the gate delay.
- The gate delay on *fan-in* makes the 6-input NOR gate extremely slow.
- Partitioning the NOR-gate into two 3 input gates →improve the speed-up, which offsets by far the extra delay incurred by turning the inverter into a two-input NAND gate.



www.rejinpaul.com

Power Consumption in CMOS Logic Gates

- The power dissipation is a strong function of transistor sizing ,input and output (rise/fall times),device temperature and switching activity.

1. Logic Function

- The transition activity is a strong function of the logic function being implemented.
- For static CMOS gates with statistically independent inputs.
- probability p_o → output will be in the zero , multiplied by the probability p_1 output will be in the one .

$$\alpha_{0\to 1} = p_0 \bullet p_1 = p_0 \bullet (1-p_0)$$

- Ex→ 2-input NOR Gate ($F=\overline{A+B}$) →inputs are independent and uniformly distributed, any *N-input static* gate has a transition probability is given by

$$\alpha_{0\to 1} = \frac{N_0 \bullet \left(2^N - N_0\right)}{2^{2N}} = \frac{3 \bullet \left(2^2 - 3\right)}{2^{2\bullet 2}} = \frac{3}{16}$$

- *No is the number of zero entries and N is the number of input*

www.rejinpaul.com

2)Signal Statistics

- The switching activity of a logic gate is a strong function of the input signal statistics.
- Ex→ 2-input static NOR gate, and let p_a *and* p_b *be the* probabilities that the inputs *A and B are one.*
- The probability that the output node equals *one is given by* $p_1 = (1-p_a)(1-p_b)$ the probability of a transition from 0 to 1 is

$$\alpha_{0->1} = p_0 p_1 = (1-(1-p_a)(1-p_b))(1-p_a)(1-p_b)$$

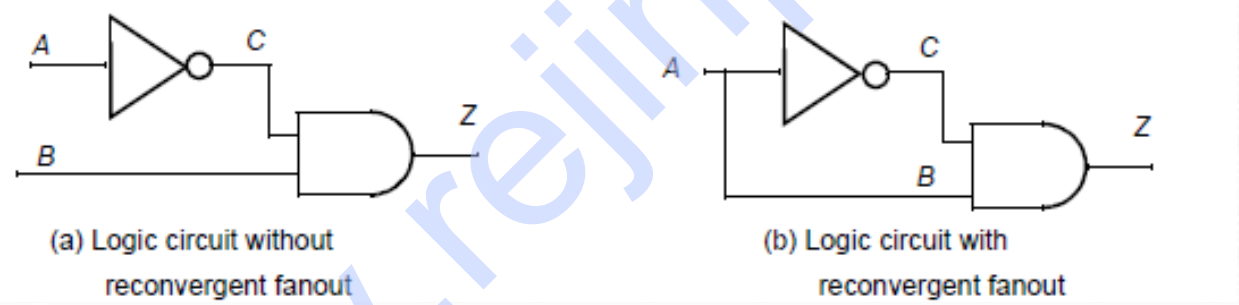
Output transition probabilities for static logic gates

	$\alpha_{0\to1}$
AND	$(1-p_A p_B)p_A p_B$
OR	$(1-p_A)(1-p_B)[1-(1-p_A)(1-p_B)]$
XOR	$[1-(p_A+p_B-2p_A p_B)](p_A+p_B-2p_A p_B)$

www.rejinpaul.com

3) Inter-signal Correlations

- The evaluation of the switching activity is further complicated that signals exhibit correlation in space and time.
- If the primary inputs to a logic network are uncorrelated, the signals become correlated or as they propagate through the logic network.
- Case(a)→ primary inputs, A and B, are uncorrelated and uniformly distributed. Node C has a 1 (o)probability of ½.
- The probability that the node *Z* undergoes a power consuming transition is then determined by AND-gate expression. $p_{0-->1} = (1- p_a\, p_b)\, p_a\, p_b = (1-1/2 \cdot 1/2)\, 1/2 \cdot 1/2 = 3/16$

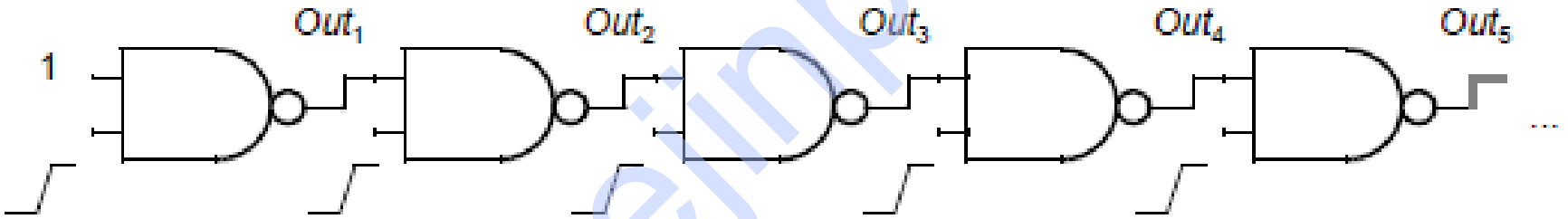


- Case(b)→*C and B are inter-dependent as both are a function of A.*
- Network can be reduced to $Z = C \cdot B = A \cdot \overline{A} = 0$, *and no transition* will ever take place.
- For an AND gate, *Z equals 1 if and only if B and C are equal to 1.* $p_Z = p(Z=1) = p(B=1, C=1)$

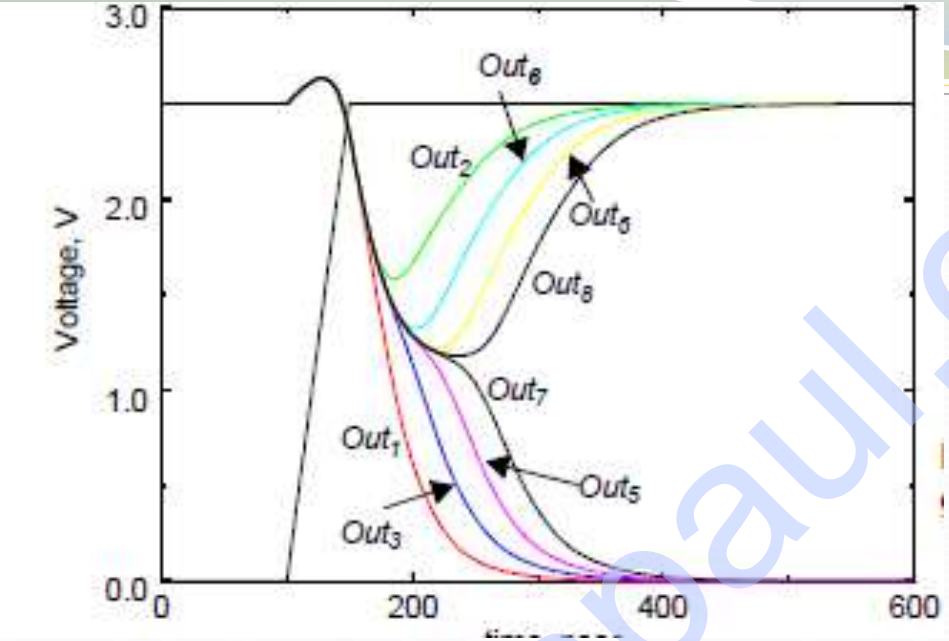
www.rejinpaul.com

4) Dynamic or Glitching Transitions

- Glitching is due to the mismatch of path length in the network.
- Propagation delay from one logic block to the next can cause spurious transitions, called *glitches, critical races, or dynamic hazards.*
- Node can exhibit multiple transitions in a single clock cycle before settling to the correct logic level.



www.rejinpaul.com



- Initially, all the outputs are 1 since one of the inputs was 0. For this particular transition, all the odd bits must transition to 0 while the even bits remain at the value of 1.
- Due to the finite propagation delay, the higher order even outputs start to discharge and the voltage drops.
- When the correct input ripples through the network, the output goes high.
- The glitch on the even bits causes extra power dissipation beyond what is required to strictly implement the logic function.

www.rejinpaul.com

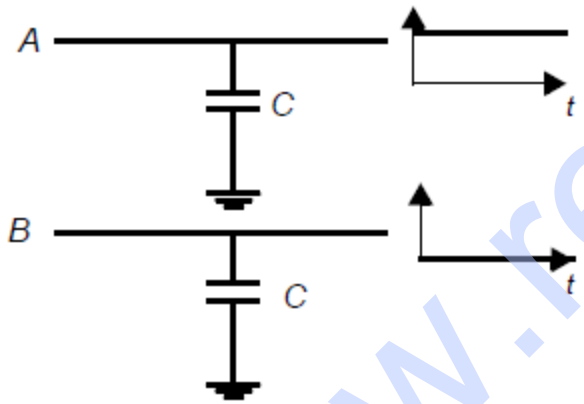
Design Techniques to Reduce Switching Activity

1. Logic Restructuring chain → Chain implementation will have an overall lower switching activity than the tree implementation .
2. Input ordering → Putting the critical-path transistors closer to the output.
3. Time-multiplexing resources
4. Glitch Reduction by balancing signal paths

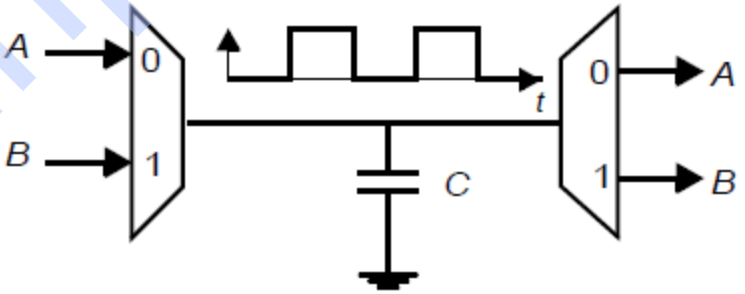
www.rejinpaul.com

3)Time-multiplexing resources

- Time-multiplexing used technique to minimize the implementation area.
- Minimum area solution does not always result in the lowest switching activity.
- In the parallel solution, the switched capacitance is very low since there are very few transitions on the data bits.(*A* is always 1 and *B is 0.)*
- Parallel data transmission requires more area than serial data transmission.
- Serial data transmission needs more switching because of the multiplexer.



(a) parallel data transmission

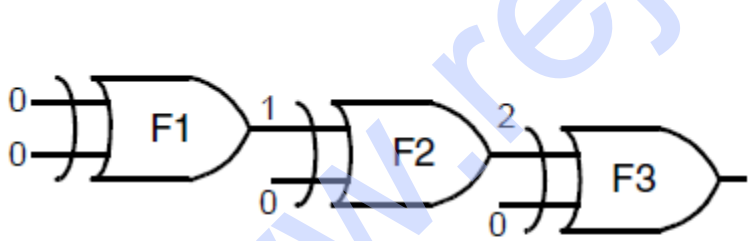


(b) serial data transmission

www.rejinpaul.com

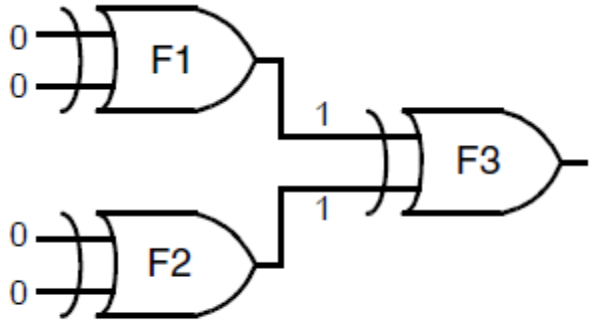
4) Glitch Reduction by balancing signal paths

- If all input signals of a gate change simultaneously, no glitching occurs.
- If input signals change at different times, a dynamic hazard might develop.
- Due to different path lengths, there is a mismatching signal occurs inputs of the network.
- Case (a) suffers from glitching because of different arrival times of the input signals for a gate. For example, for gate *F3, one input settles at* time 0, while the second one only arrives at time 2
- Case(b)→arrival times are identical and will reduce the no of superfluous transistion.



(a) Network sensitive to glitching

XOR Gate



(b) Glitch-free network

www.rejinpaul.com

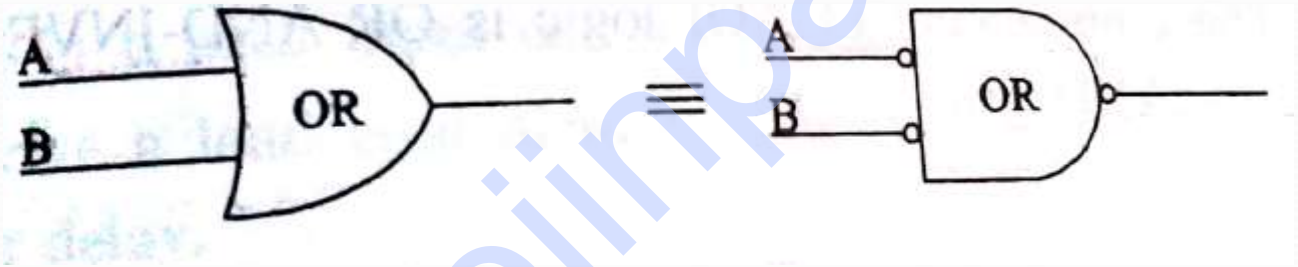
Techniques to improve static CMOS circuits

1. Bubble pushing
2. Compound gates
3. Input ordering Delay effect
4. Asymmetric Gates
5. Skewed gates
6. P/N ratios
7. Multiple Threshold voltage

www.rejinpaul.com

1)Bubble Pushing

- By introducing bubble at the input terminal.
- Static CMOS circuits are dual networks,obtained by Demorgans's theorem



- OR gate is equivalent to a NAND gate with bubbles at its input.
- Similarly(AND=NOR,NAND=NOR,NOR=OR)

www.rejinpaul.com

2)Compound Gates(AOI-OAI Logic)

- AOI→AND-OR-INVERT is a two level logic functions designed by one or more AND gates followed by a OR gate.
- OAI→OR-AND-INVERT is a two level logic functions designed by one or more OR gates followed by a AND gate.

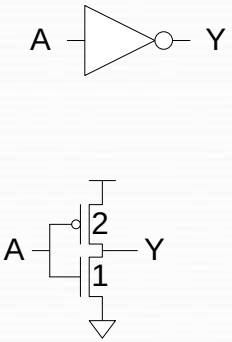
- INVERTER→g=n=1
- NAND→g=(n+2)/3
- NOR→g=(2n+1)/3
- Parasitic delay=(summing size of transisitor attached to the output/3)

www.rejinpaul.com

Logical effort of common of Gates

unit inverter

$$Y = \overline{A}$$

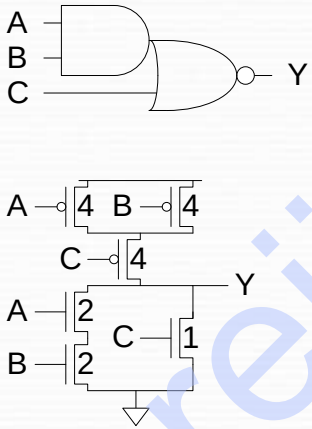


$g_A = 3/3$

$p = 3/3$

AOI21

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



$g_A = 6/3$

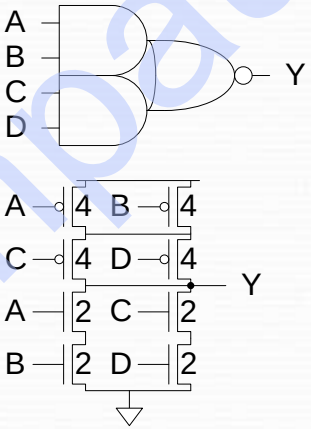
$g_B = 6/3$

$g_C = 5/3$

$p = 7/3$

AOI22

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



$g_A = 6/3$

$g_B = 6/3$

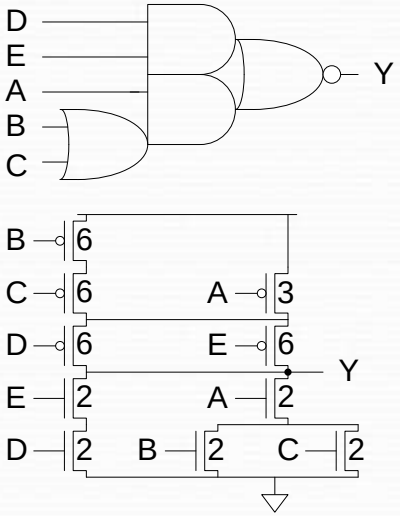
$g_C = 6/3$

$g_D = 6/3$

$p = 12/3$

Complex AOI

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



$g_A = 5/3$

$g_B = 8/3$

$g_C = 8/3$

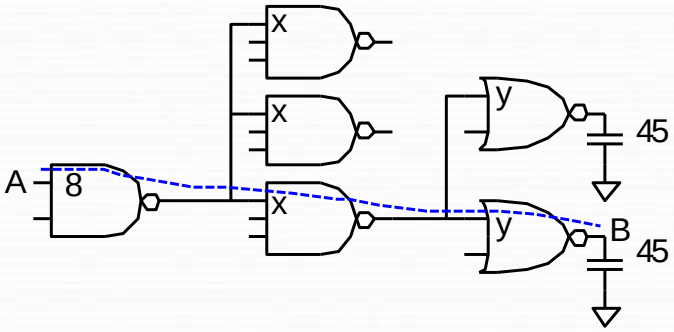
$g_D = 8/3$

$g_E = 8/3$

$p = 16/3$

www.rejinpaul.com

Example: 3-stage path



Logical Effort G = (4

Electrical Effort H = 4

Branching Effort B = 3

Path Effort F = G

Best Stage Effort $\hat{f} = \sqrt[3]{}$

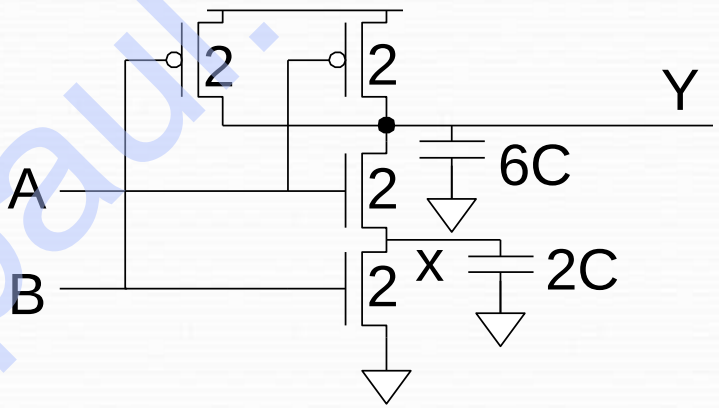
Parasitic Delay P = 2 + 3 + 2 =

Delay D = (gh+p)=3*5 + 7 = 22 = 4.4 FO4

www.rejinpaul.com

3)Input ordering Delay effect

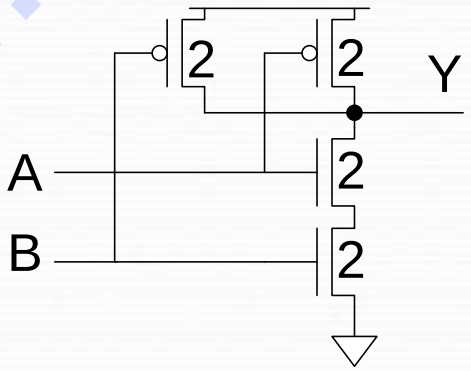
- Consider 2-input NAND gate with different parasitic capacitance.
- For falling output transition occuring,when one input is stable 1 and other rises from 0 to 1 .
- If B raises last,node x will be at V_{DD},because node was pulled up through the nMOS on input A. hence Elmore delay 2.33τ.
- If A raises last,node x will be at 0V,because node discharged through the nMOS on input B.hence Elmore delay 2τ.



www.reimpaul.com

Inner & Outer Inputs

- *The outer input is defined as the* input is input is closest to rail (B)
- *The inner input is defined as the* input is closest to output (A)



- The parasitic delay is small when the inner input switches last because the intermediate nodes have already been discharged.
- Therefore ,if one signal is known to arrive later than the others, the gate is fastest when the signal is connected to the inner input.

www.rejinpaul.com

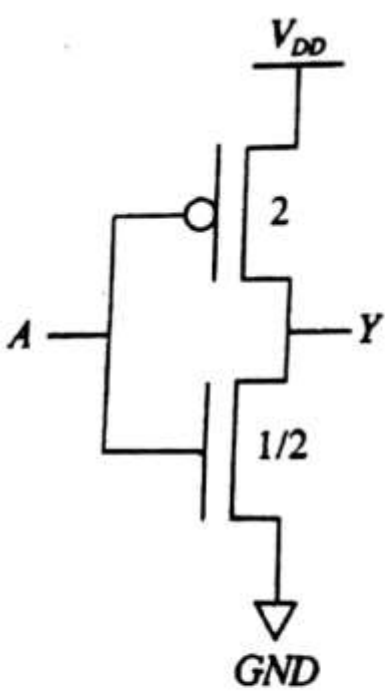
5)Skewed Gates

- Inverters with different β ratios called→skewed inverters.
- If $\beta_p = \beta_n$ →the inverter threshold voltage is $(V_{DD})/2$.
- If $\beta_p > \beta_n$ →the inverter is HI-skew→stronger pMOS
- If $\beta_p < \beta_n$ →the inverter is LO-skew→weaker pMOS.
- Rise time→3 Fall time→3/2
- Logical effort for HI-skew gates is given by

$$\text{The logical effort } g_u = \frac{\text{input capacitance of HI – skew gate}}{\text{input capacitance of unskewed gate (equal rise resistance)}}$$

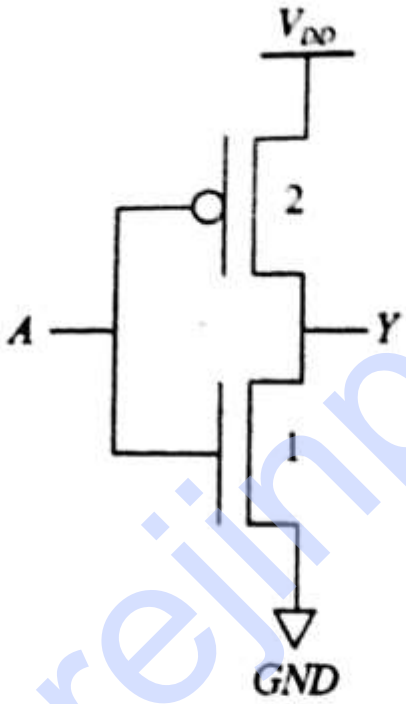
$$\text{The logical effort } g_d = \frac{\text{input capacitance of HI – skew gate}}{\text{input capacitance of unskewed gate (equal fall resistance)}}$$

www.rejinpaul.com



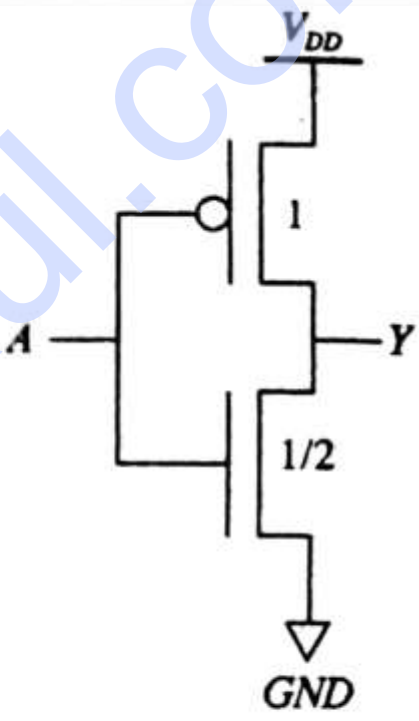
Hi-Skew Inverter

(a)



Unskewed Inverter (Equal Rise Resistance)

(b)

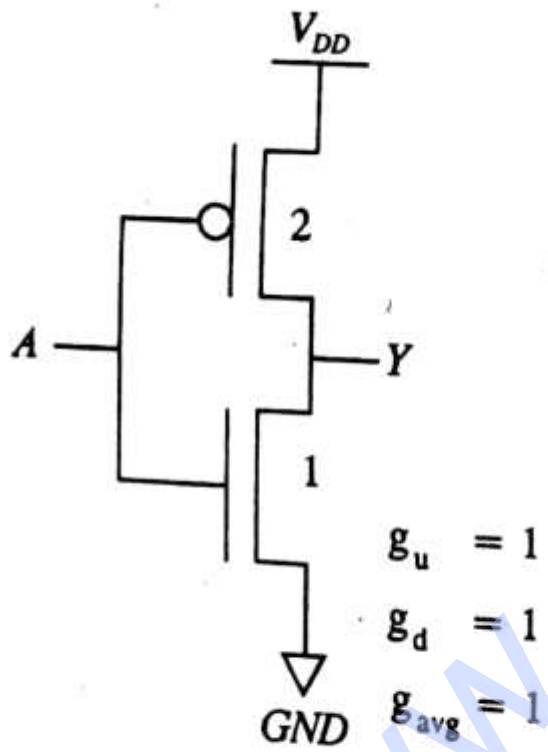


Unskewed Inverter (Equal Fall Resistance)

(c)

www.rejinpaul.com

Normal Skew

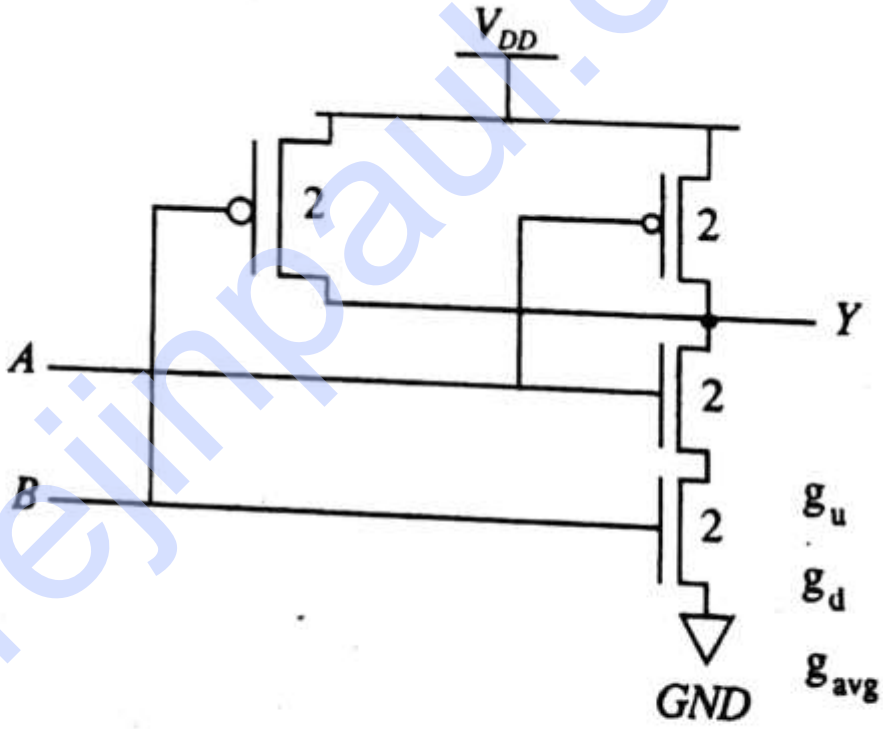


$g_u = 1$

$g_d = 1$

$g_{avg} = 1$

Unskewed Inverter



$g_u = 4/3$

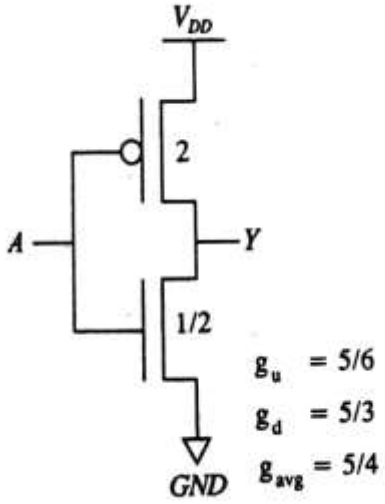
$g_d = 4/3$

$g_{avg} = 4/3$

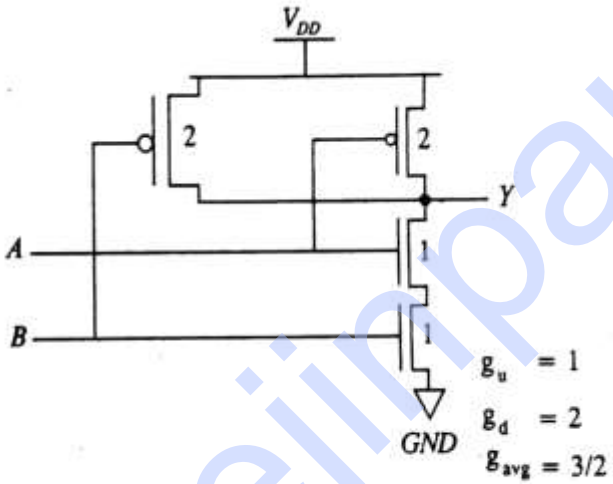
Unskewed NAND

www.rejinpaul.com

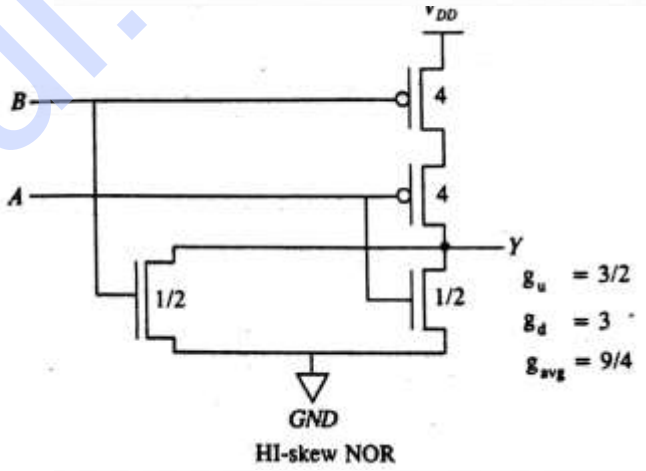
High Skewed Gates



HI-Skew Inverter



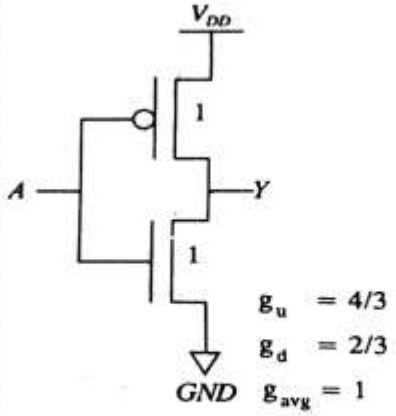
HI-Skew NAND



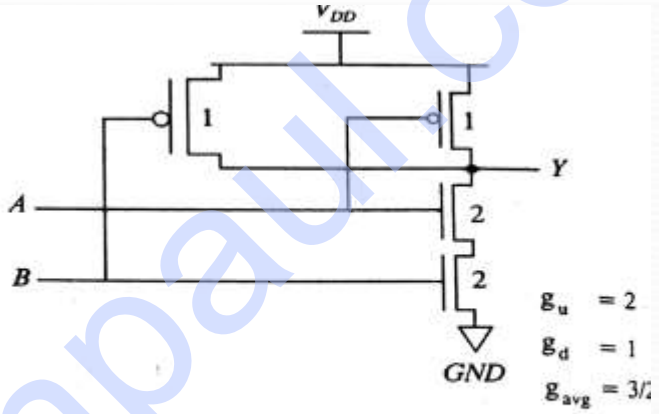
HI-skew NOR

www.rejinpaul.com

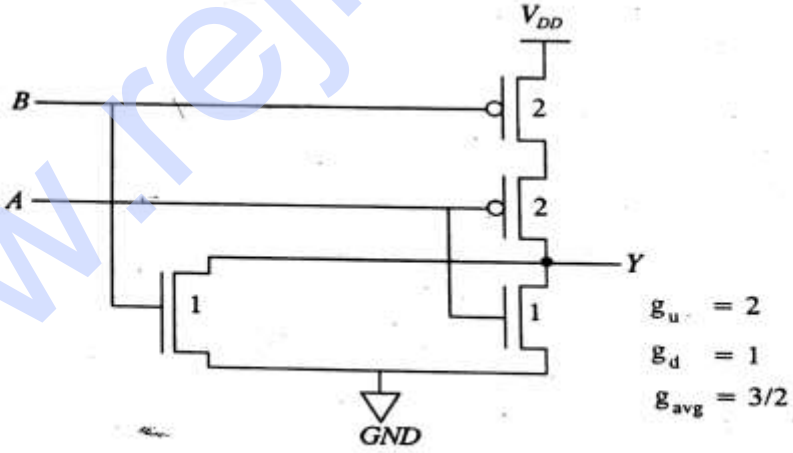
Low Skewed Gates



LO-Skew Inverter



LO-Skew NAND

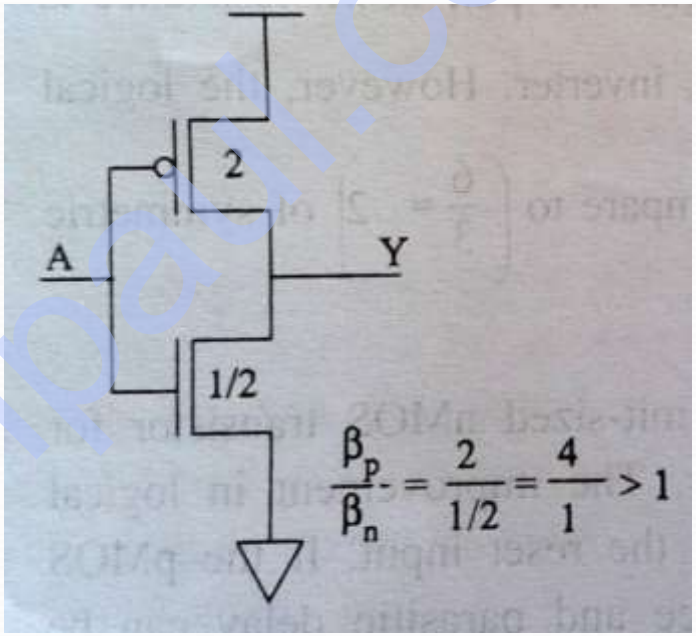


LO-Skew NOR

www.rejinpaul.com

Logical effort calculation of HI-skew gates

- $g_u = \frac{2+1/2}{3}$
- $g_d = \frac{2+1/2}{3/2}$
- $g_{avg} = \frac{g_u + g_d}{2}$
- $g_{avg} = 5/4$



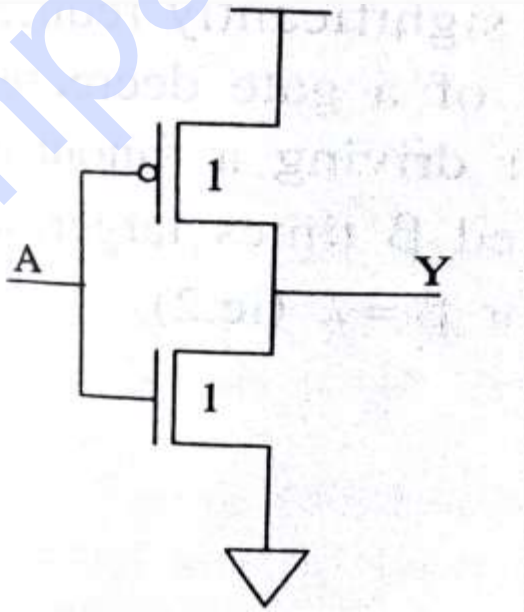
www.rejinpaul.com

Logical effort calculation of LO-skew gates

The logical effort $g_u = \dfrac{\text{input capacitance of LO – skew gate}}{\text{input capacitance of unskewed gate (equal fall resistance)}}$

The logical effort $g_d = \dfrac{\text{input capacitance of LO – skew gate}}{\text{input capacitance of unskewed gate (equal rise resistance)}}$

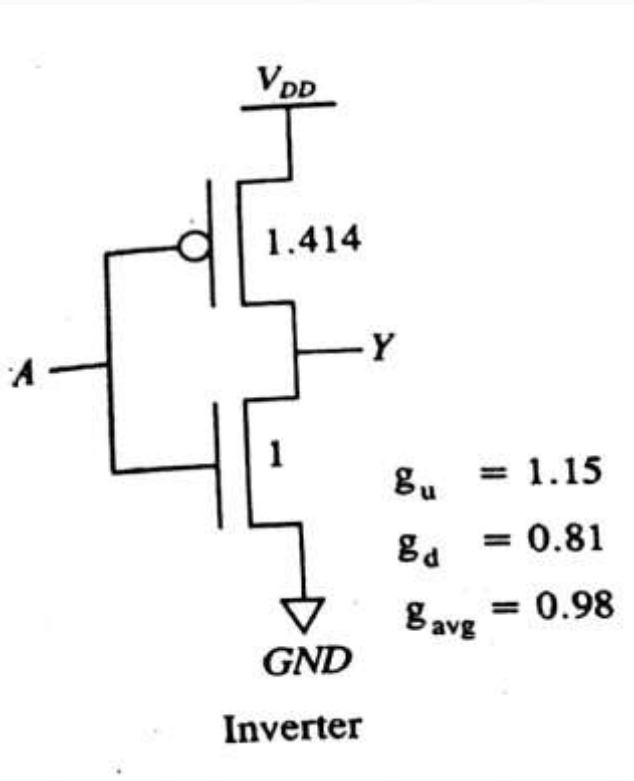
- $g_u = \dfrac{1+1}{3/2}$
- $g_d = \dfrac{1+1}{3}$
- $g_{avg} = \dfrac{g_u + g_d}{2}$
- $g_{avg} = 1$



www.rejinpaul.com

6)P/N Ratios

- The ratio of pMOS to nMOS transistor width is called P/N ratio
- Processes with mobility ratio of un/up=2 gives the best ratio with least delay
- Falling delay→ f(1+ β)λ
- Rising delay→(k/ β) f(1+ β)λ.
- Average delay→(faaling+rising delay)/2



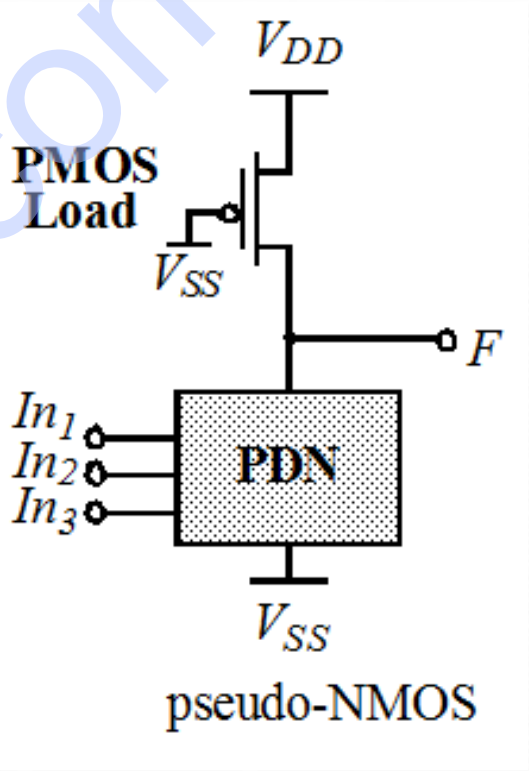
www.rejinpaul.com

Ratioed Circuits

- Ratioed logic used to reduced the number of transistors required to implement a logic function.
- Reduce the cost and power dissipation.
- It use Pseudo-nMOS logic to reduce the number pMOS transistor.
- It is similar to static logic,but replace pull-up stack by single pMOS transistor.
- nMOS transistor can pull-up the output to near gnd.
- Pull-down transistor are scaled 4times as strong as pull-up

Ratioed circuits can be deigned by the following logics

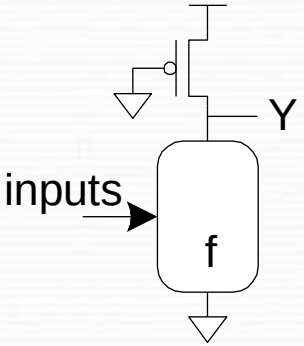
1. Pseudo-nMOS Logic
2. Ganged CMOS Logic
3. Cascade Voltage Switch Logic



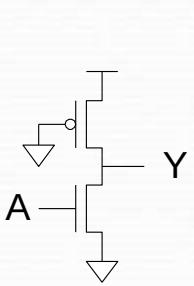
www.rejinpaul.com

1)Pseudo-nMOS Logic

- Design for unit current on output to compare with unit inverter.
- pMOS fights nMOS
- $I_{out} = 4I/3 - I/3$

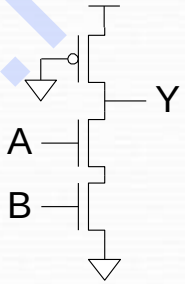


Inverter



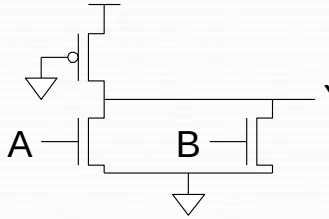
g_u =
g_d =
g_{avg} =
p_u =
p_d =
p_{avg} =

NAND2



g_u =
g_d =
g_{avg} =
p_u =
p_d =
p_{avg} =

NOR2

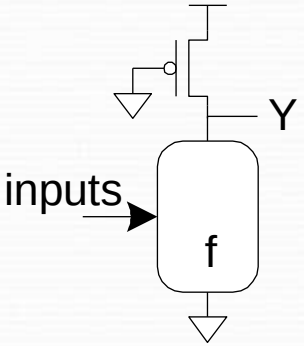


g_u =
g_d =
g_{avg} =
p_u =
p_d =
p_{avg} =

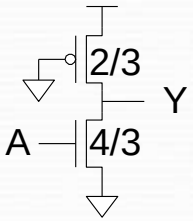
www.rejinpaul.com

Pseudo-nMOS Gates

- Design for unit current on output to compare with unit inverter.
- pMOS fights nMOS

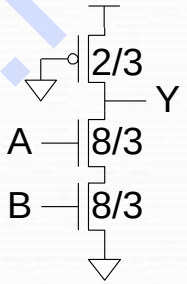


Inverter



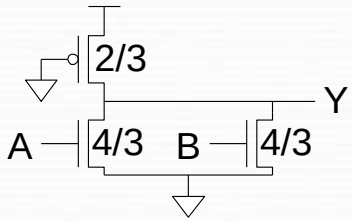
$g_u = 4/3$
$g_d = 4/9$
$g_{avg} = 8/9$
$p_u = 6/3$
$p_d = 6/9$
$p_{avg} = 12/9$

NAND2



$g_u = 8/3$
$g_d = 8/9$
$g_{avg} = 16/9$
$p_u = 10/3$
$p_d = 10/9$
$p_{avg} = 20/9$

NOR2

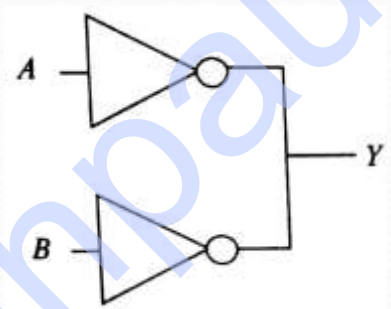


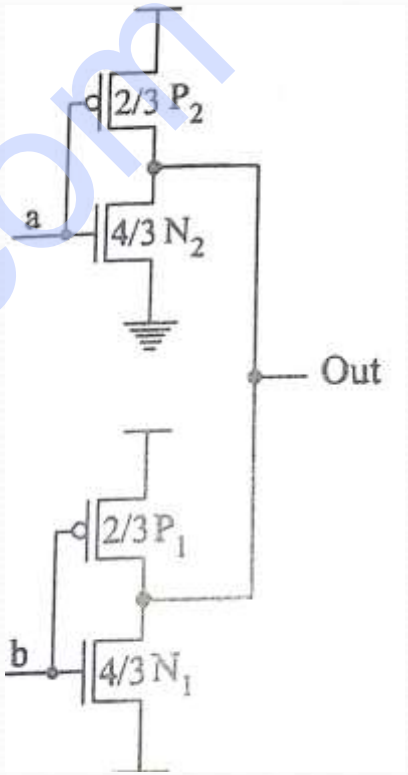
$g_u = 4/3$
$g_d = 4/9$
$g_{avg} = 8/9$
$p_u = 10/3$
$p_d = 10/9$
$p_{avg} = 20/9$

www.rejinpaul.com

2)Ganged CMOS Logic

- Constructed by pair of inverters.
- When output rise,both inverters pull in parallel→more current
- Input capacitance 2/3+4/3=2



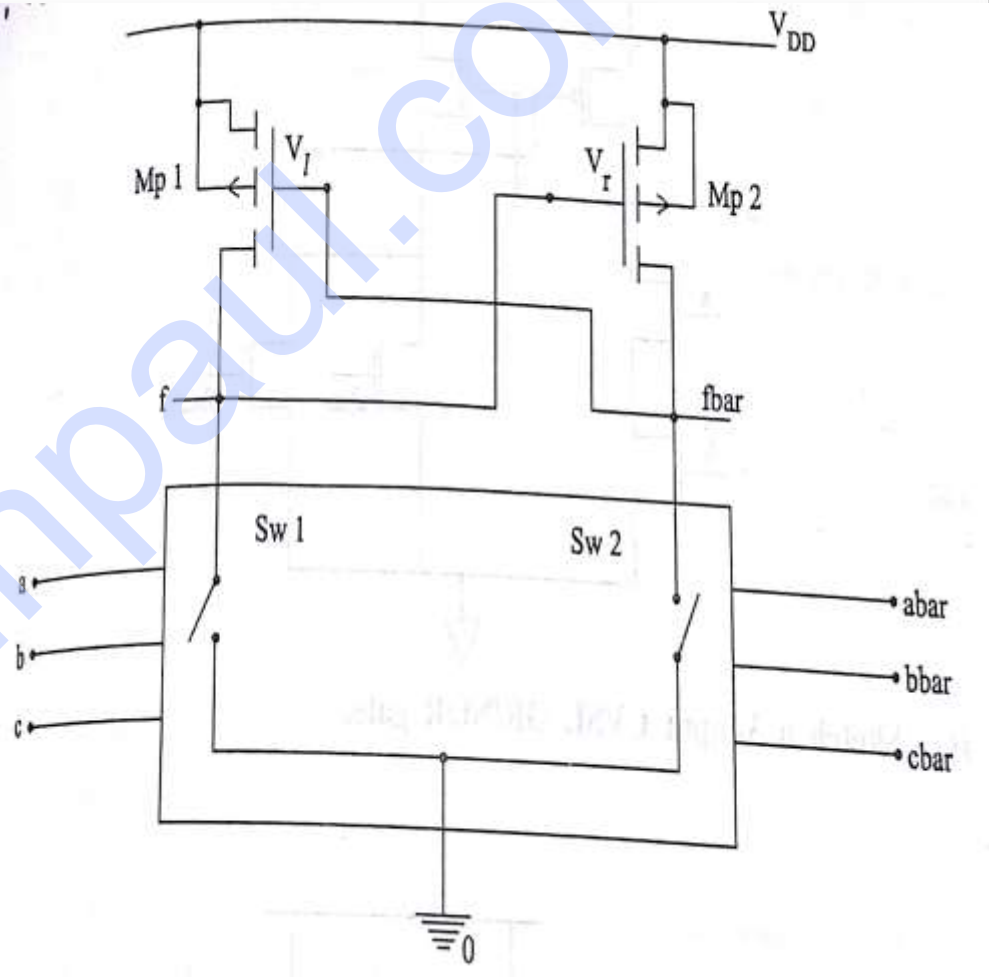


A	B	N1	P1	N2	P2	Y
0	0	OFF	ON	OFF	ON	1
0	1	ON	OFF	OFF	ON	~0
1	0	OFF	ON	ON	OFF	~0
1	1	ON	OFF	ON	OFF	0

www.rejinpaul.com

3)Cascade Voltage Switch Logic-CVSL

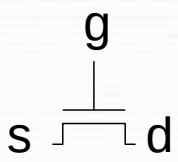
- Dual n-block and a pair of cross-coupled pMOS transistors.
- Compute logic function and its complement.
- High i/p is applied at left PDN and complement is applied right PDN.
- If sw1 closed→ then f=1,fbar=0
- If sw2 closed→ then f=0,fbar=1
- If $V_l = V_{DD}$ →Mp1-ON then f=1 state.

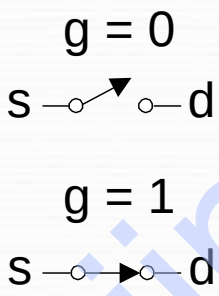


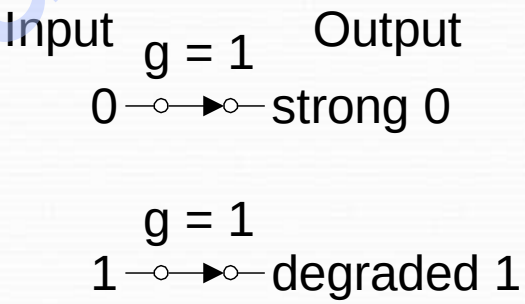
www.rejinpaul.com

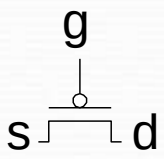
Pass Transistors logic

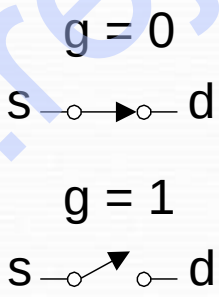
- A popular and widely-used alternative to complementary CMOS is pass-transistor logic.
- which attempts to reduce the number of transistors required to implement logic

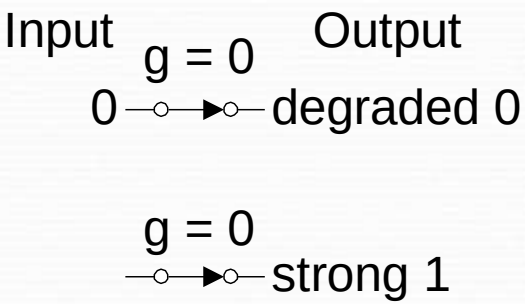








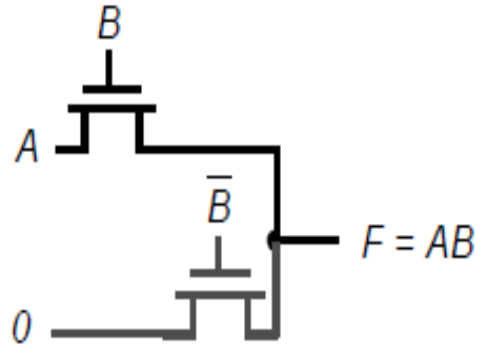




www.rejinpaul.com

Pass-transistor implementation of an AND gate

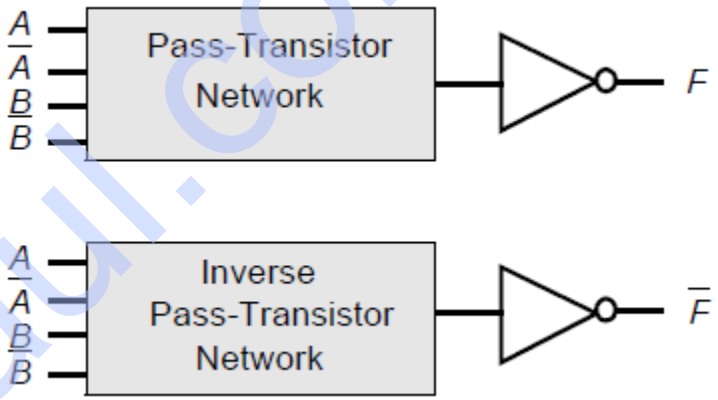
- Implementation of the AND function using only NMOS transistors.
- If the B input is high, the top transistor is ON and copies the input A to the output F.
- When B is low, the bottom pass transistor is turned on and passes a o.
- Implementation of the AND gate requires 4 transistors(including the inverter required to invert *B).*
- *But complementary CMOS implementation* would require 6 transistors.



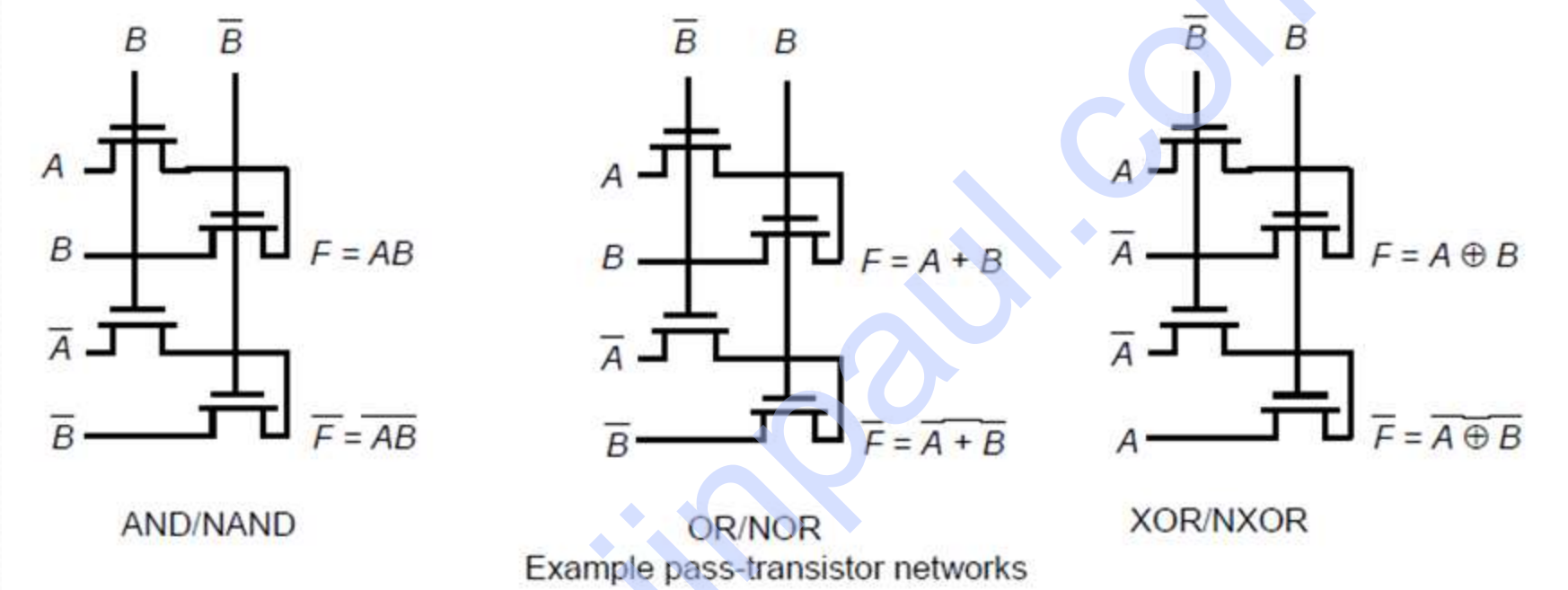
www.rejinpaul.com

Complementary Pass Transistor Logic

- For high performance design, a differential pass-transistor logic family, called CPL or DPL, is commonly used.
- It accept true and complementary inputs and produce true and complementary outputs.
- CPL used to resolve the problem of CVSL.
- CPL uses one half of the gate pull up, while the other half's pull down.
- Logic function is implemented by using only nMOS network.
- It use two set of complementary inputs and CMOS inverters.
- CPL belongs to the class of static gates,because the outputs nodes are always connected to either V_{DD} or V_{SS}.



www.rejinpaul.com



Example pass-transistor networks

Advantages:

Low capacitance.

High speed operation.

Less power consumption.

$Pd=V_{dd}.V_{swing}.C_{node}.f$

$Vswing=V_{dd}-V_{th}$

www.rejinpaul.com

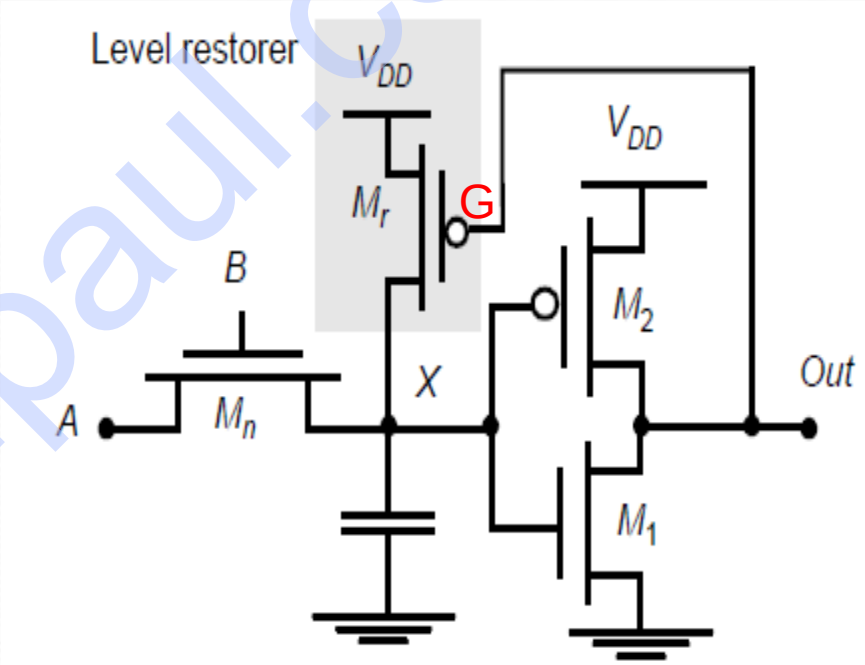
Robust and Efficient Pass-Transistor Design

- DPL- logic, like single-ended pass-transistor logic, suffers from static power dissipation and reduced noise margins.
- Voltage drop problem of DPL or CPL can be solved by the following methods
- ***Solution 1: Level Restoration***
- ***Solution 2: Multiple-Threshold Transistors***
- ***Solution 3: Transmission Gate Logic***

www.rejinpaul.com

1)Level Restoration

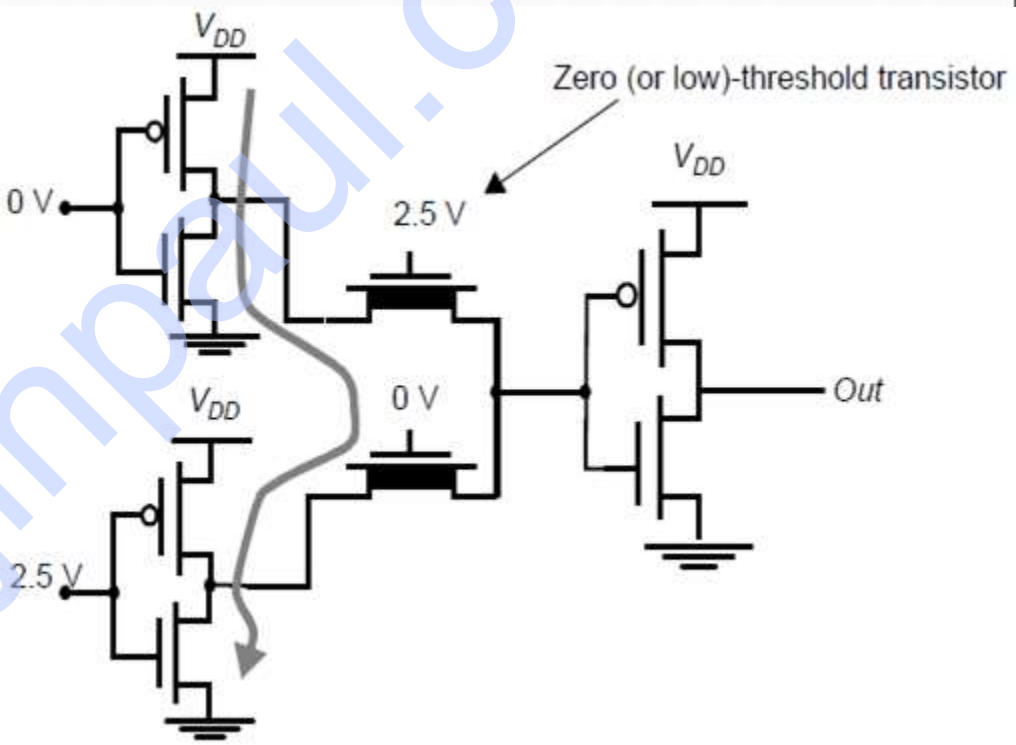
- A common solution to the voltage drop problem is the use single PMOS in a feedback path.
- The gate of the PMOS device is connected to the output of the inverter, its drain connected to the input of the inverter and the source to $V_{DD.}$
- If node X is at oV and out is at $V_{DD.}$ and the Mr is turned off ,with B = $V_{DD.}$ and A = o.
- If input A makes a o to $V_{DD.}$ transition, Mn only charges up node X to $V_{DD.}$ -$_{VTn.}$
- Hence inverter low, turning on the feedback device M_r and pulling node X =V_{DD}.
- This eliminates any static power dissipation in the inverter.
- No static current path can exist through the level restorer and the pass-transistor, since the restorer is only active when *A is high*.



www.rejinpaul.com

2)Multiple Threshold Transistors

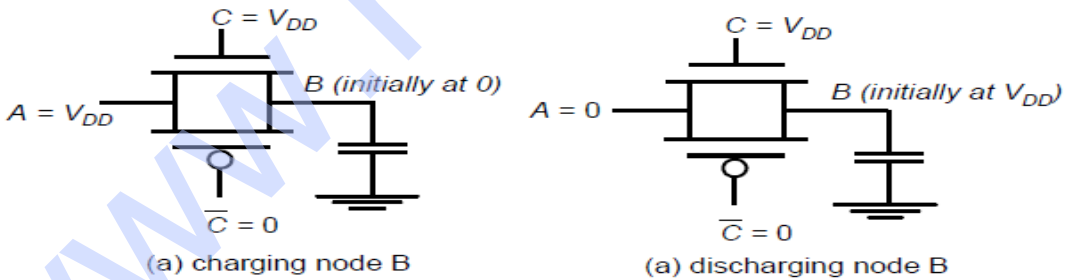
- It uses zero threshold devices for the NMOS pass-transistors eliminates most of the threshold drop, and passes a signal close to V_{DD}.
- All devices other than the pass transistors are implemented using standard high-threshold devices.
- The use of zero-threshold transistors can be dangerous due to the subthreshold currents
- that can flow through the pass-transistors, even if *VGS is slightly below VT.*
- While these leakage paths are not critical when the device is switching constantly, they do pose a significant energy-overhead when the circuit is in the idle state.



www.rejinpaul.com

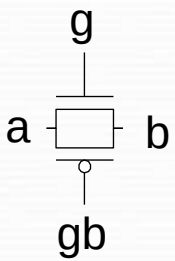
3)Transmission gate(TG)

- Pass transistors produce degraded outputs but *Transmission gates* pass both 0 and 1 well as a strong output.
- Transmission gate logic consists of an nMOS and a pMOS transistor in parallel with gates controlled by complementary signals.
- At least one of the two transistor is ON and the output voltage is either strong 0 or strong 1.
- It is superior than CMOS ckts interms of layout density, ckt delay and power consumption.
- Gate is driven by control signal.
- Source is connected with pass variable.

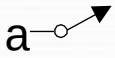


www.rejinpaul.com

Transmission Gates

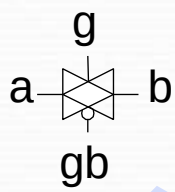


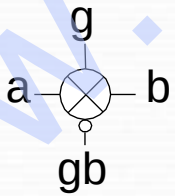
g = 0, gb = 1

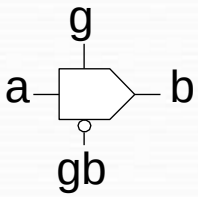


g = 1, gb = 0

Input	Output
g = 1, gb = 0 0	strong 0
g = 1, gb = 0 1	strong 1



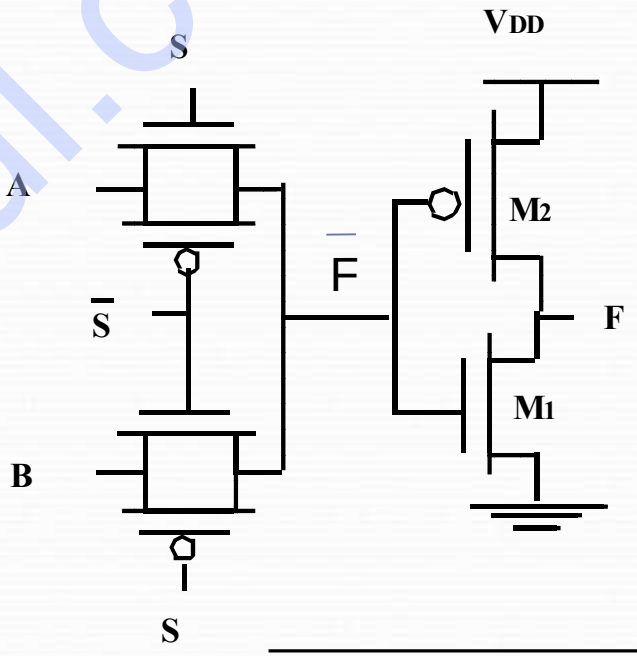




www.rejinpaul.com

Transmission gate Multiplexer

- The gate either selects input A or B on the basis of the value of the control signal S.
- TG output is given by $F=\bar{A}.S+B.\hat{S}$.
- A complementary implementation of this gate requires 8 transistors, but the transmission gate logic needs only 6 transistors.

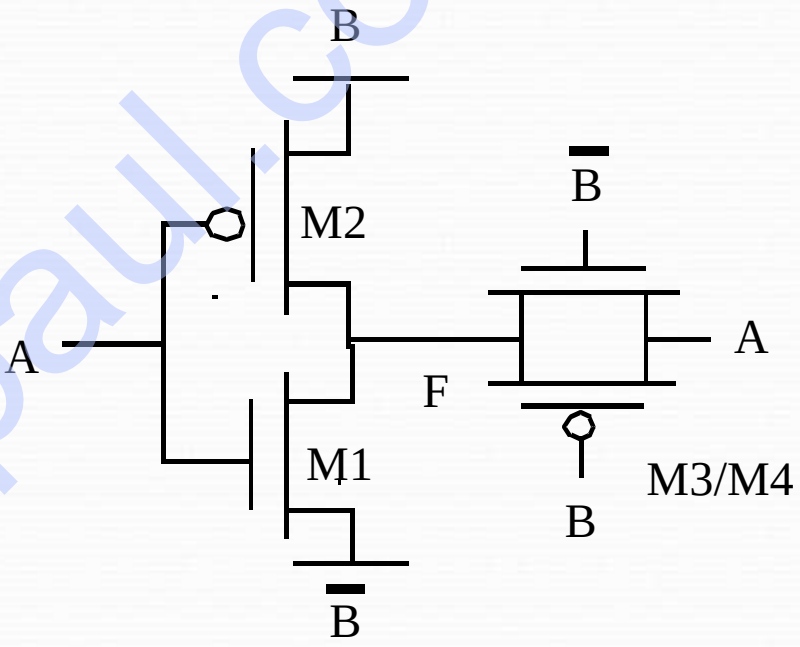


$$F = \overline{(S \cdot A + \bar{S} \cdot B)}$$

www.rejinpaul.com

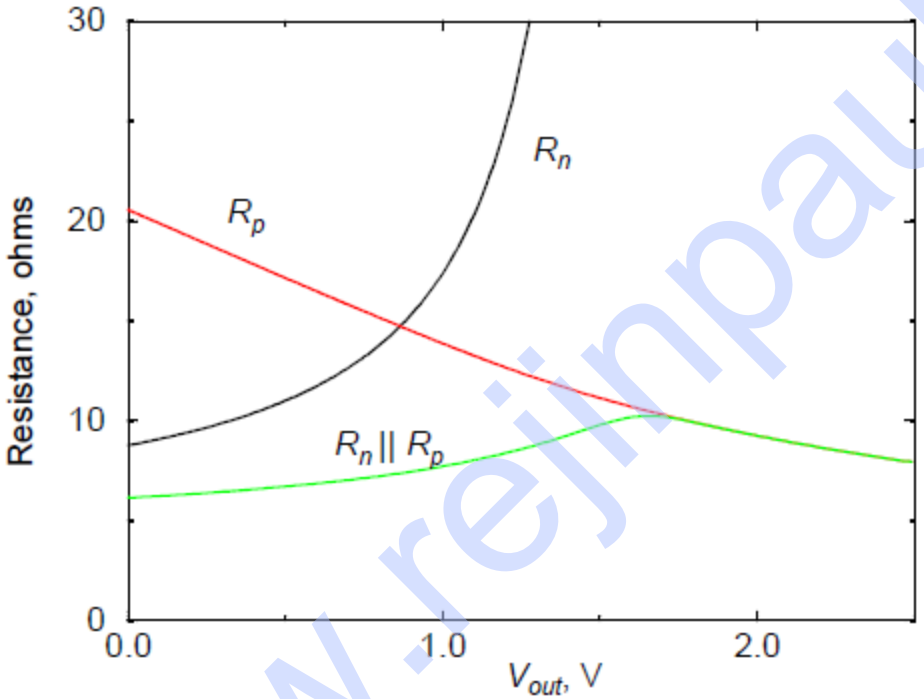
Transmission gate XOR

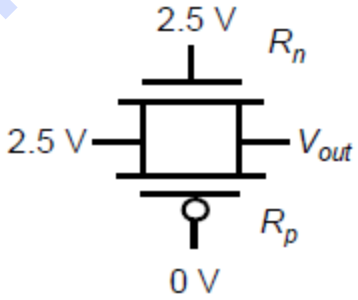
- For implementation TG-XOR Complementary CMOS requires 12 transistor but it requires 6.
- For *B = 1, transistors M1 and M2 act as an inverter while the transmission gate M3/M4 is off; hence F = $\overline{A}B$.*
- For *B = 0,M1 and M2 are disabled, F =$A\overline{B}$.*
- For the values of *A and B, node F always has a connection to either VDD or GND*
- When designing static-pass transistor networks, it is essential to adhere to the low-impedance rule under all circumstances.



www.rejinpaul.com

Performance of Transmission gates



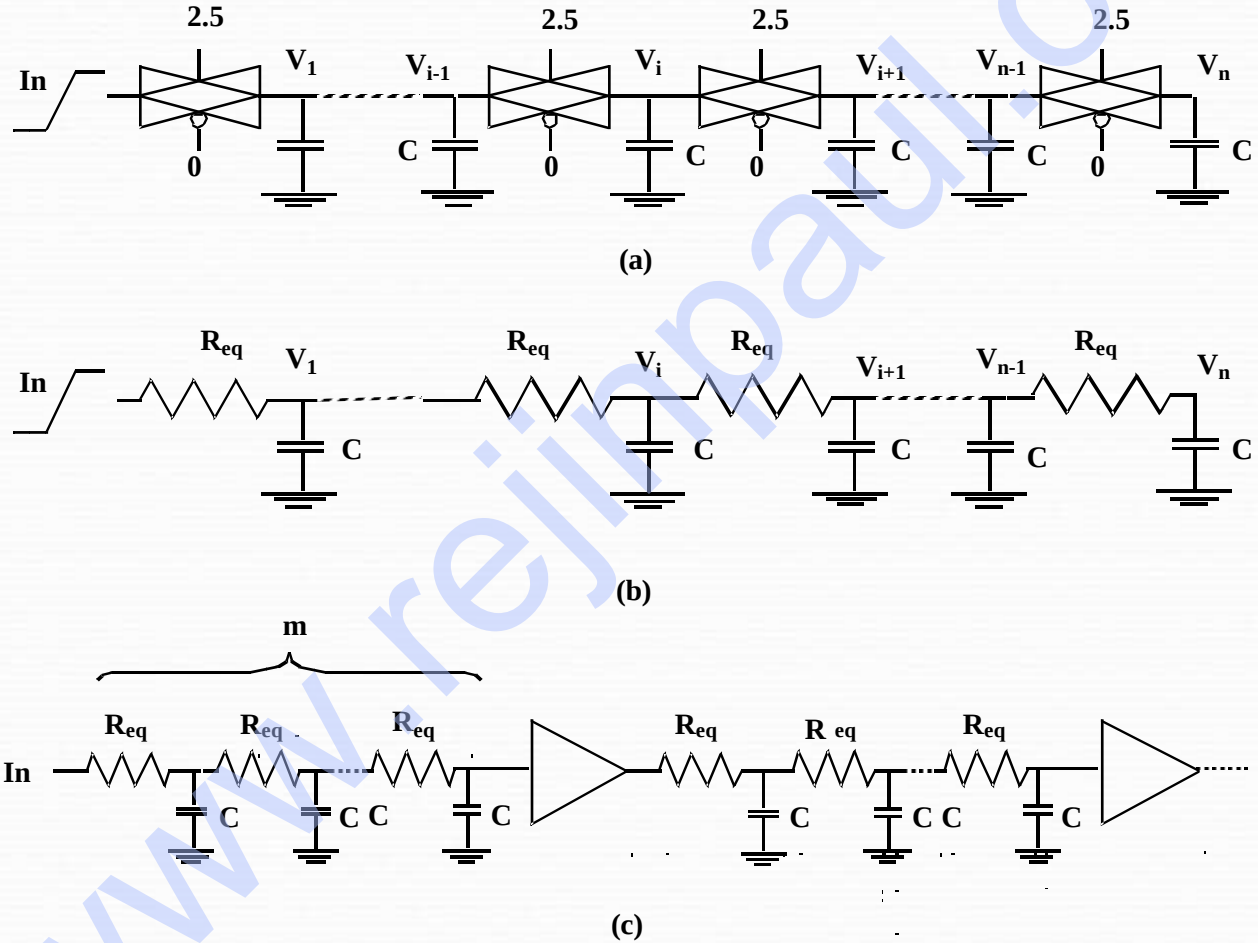


$R_n=(V_{DD}-V_{out})/I_{Dn}$

$R_p=(V_{DD}-V_{out})/-I_{Dp}$

www.rejinpaul.com

Delay in Transmission Gate Networks



www.rejinpaul.com

Delay Optimization

- **Delay of RC chain**

$$t_p = 0.69 \sum_{k=0}^{n} CR_{eq} k = 0.69 CR_{eq} \frac{n(n+1)}{2}$$

- **Delay of Buffered Chain**

$$t_p = 0.69 \left\lfloor \frac{n}{m} CR_{eq} \frac{m(m+1)}{2} \right\rfloor + \left(\frac{n}{m} - 1 \right) t_{buf}$$

$$= 0.69 \left[CR_{eq} \frac{n(m+1)}{2} \right] + \left(\frac{n}{m} - 1 \right) t_{buf}$$

$$m_{opt} = 1.7 \sqrt{\frac{t_{pbuf}}{CR_{eq}}}$$

Buffer can be

(a) Inverter pairs
(b) One inverter + Final correcting inverter

www.reinpaul.com

Dynamic CMOS

- In static circuits at every point in time (except when switching) the output is connected to either GND or VDD via a low resistance path.
  - fan-in of n requires 2n (n N-type + n P-type) devices

- Dynamic circuits rely on the temporary storage of signal values on the capacitance of high impedance nodes.
  - requires on n + 2 (n+1 N-type + 1 P-type) transistors

www.reimpaul.com

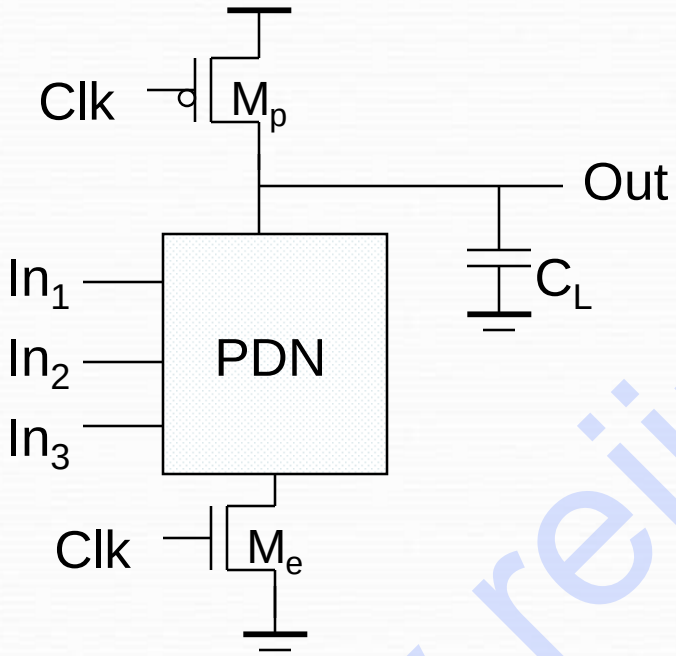
Dynamic CMOS circuits

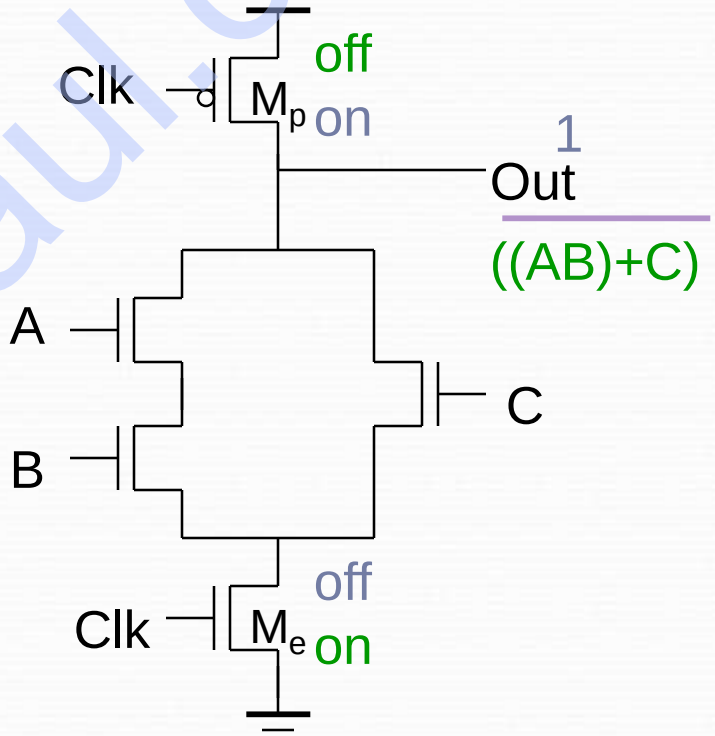
Properties

- Requires only N+2 transistors,less than static CMOS(2N).
- Logic function is implemented by the PDN only.
- The operation of this circuit is divided into two major phases: *precharge and evaluation, with the* mode of operation determined by the *clock signal CLK*.
- Full swing output (V_{OL}=GND and V_{OH}=V_{DD})
- Faster switching speed.
- Small fan-out.(low no transistor,low no of capacitor)
- Power dissipation is good.(no short ckt power)
- PDN starts when input →$\mathbf{V_{th}}$

www.rejinpaul.com

Dynamic Gate





Two phase operation

Precharge (Clk = 0)

Evaluate (Clk = 1)

www.rejinpaul.com

Modes of Operation

1. **Precharge**

- When *CLK = 0, the output node Out is precharged to VDD by the PMOS transistor Mp.*
- During that time, the evaluate NMOS transistor *Me is off,* so that the pull-down path is disabled.

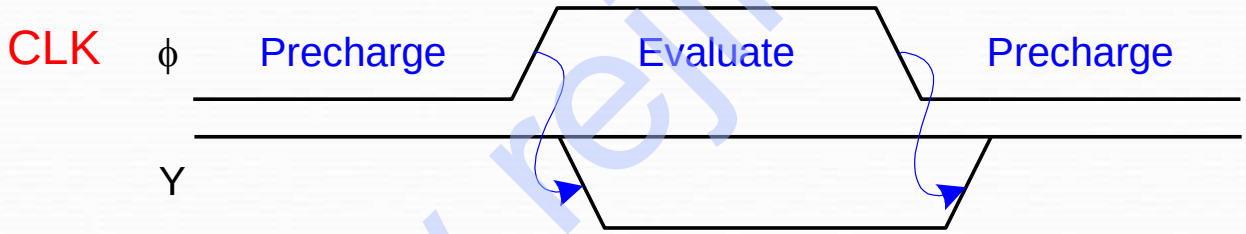
1. **Evaluation**

- For *CLK = 1, the precharge transistor Mp is off, and the evaluation transistor Me is turned*on.
- The output is conditionally discharged based on the input values and the pull-downtopology.
- If the inputs are such that the PDN conducts, then a low resistance path exists between *Out and GND and the output is discharged to GND.*
- *If the PDN is turned off, the* precharged value remains stored on the output capacitance.
- Once the output of a dynamic gate is discharged, it cannot be charged again until the next precharge operation.

www.rejinpaul.com

Conditions on Output

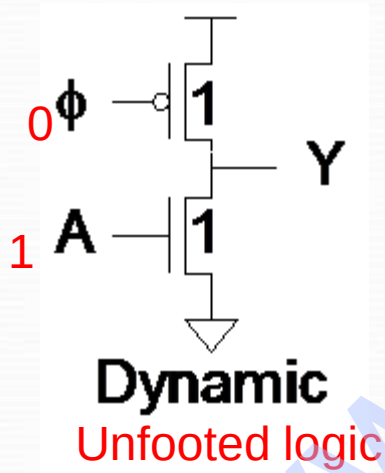
- During the prechargephase (*CLK=0), the output is precharged to VDD regardless of the input values since the* evaluation device is turned off.
- During evaluation (*CLK=1), a conducting path is created* between *Out and GND if (and only if) A·B+C is TRUE.*
- $Out = CLK + (A \times B + C) \times CLK$

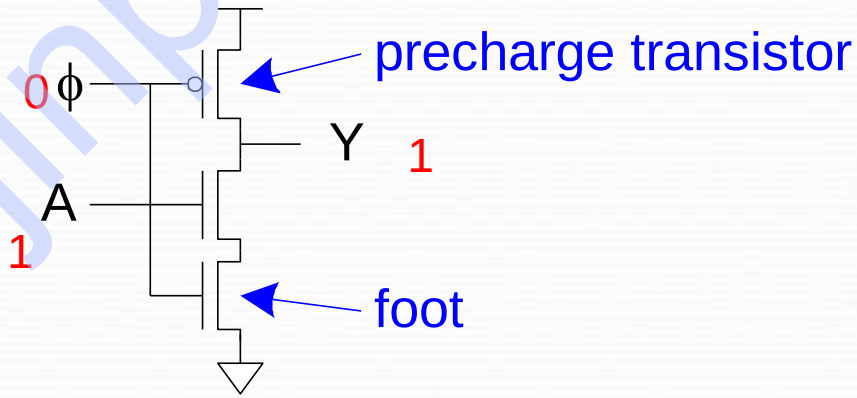


www.rejinpaul.com

1)Foot-Problem

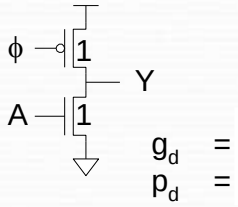
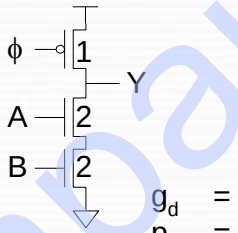
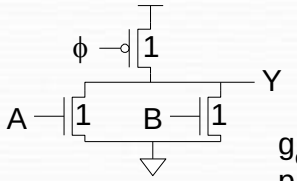
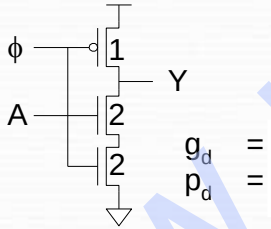
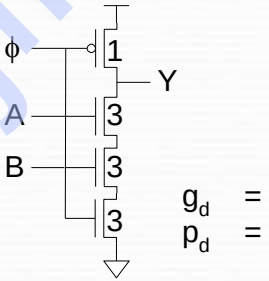
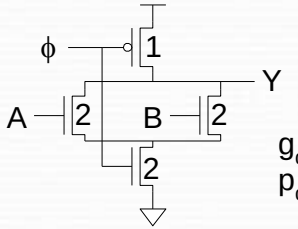
- When CLK=0 and A=1→ pulldown network is ON during precharge.Henc e VDD connected with GND
- Use series nMOS transistor to prevent fight.





www.rejinpaul.com

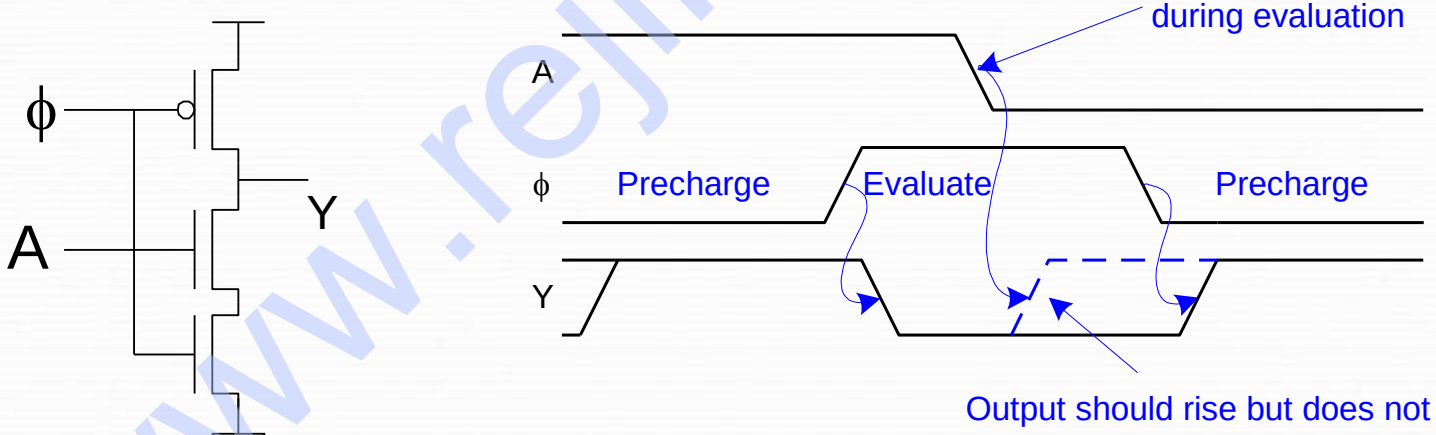
Calculation of Logical Effort

	Inverter	NAND2	NOR2
unfooted	 g_d = p_d =	 g_d = p_d =	 g_d = p_d =
footed	 g_d = p_d =	 g_d = p_d =	 g_d = p_d =

www.rejinpaul.com

2)Monotonicity Problem

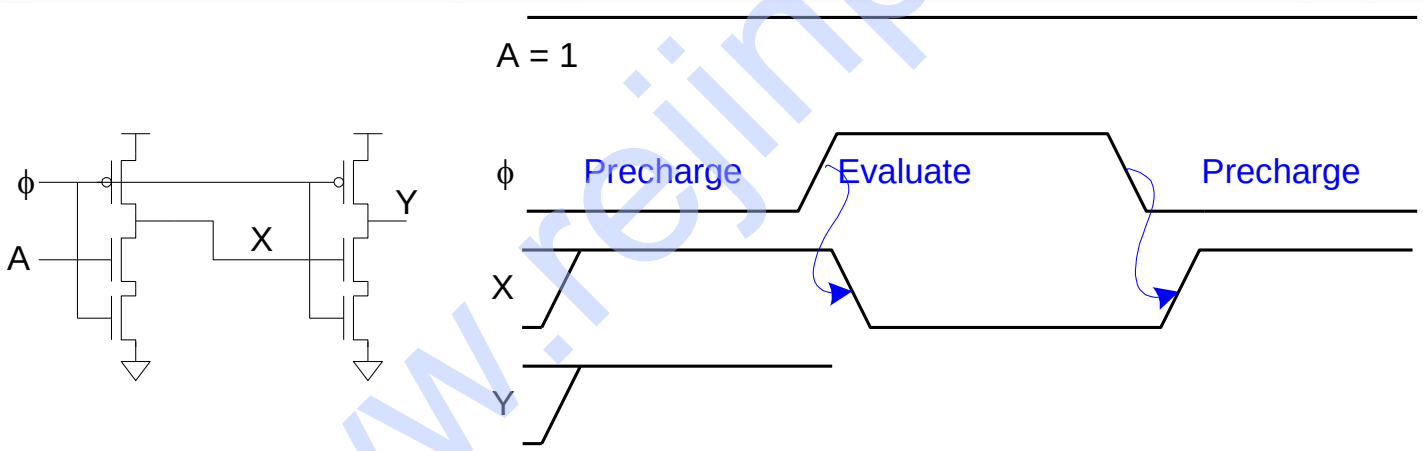
- While a dynamic gate is in evaluation,the inputs must be *monotonically rising.*
- *Input* can start LOW and remain LOW, start LOW and rise HIGH,start HIGH and remain HIGH, but not start HIGH and fall LOW.
- During precharge, the output is pulled HIGH.
- When the clock rises, the input is HIGH,hence output is discharged through PDN.
- The input later falls LOW, turning off the pulldown network. However, the precharge transistor is also OFF so the output floats.
- The output will remain low until the next precharge step



www.rejinpaul.com

Solution for Monotonicity problem-

- But dynamic gates produce monotonically falling outputs during evaluation
- Illegal for one dynamic gate to drive another!



Incorrect connection of dynamic gates

www.rejinpaul.com

CMOS Domino Logic

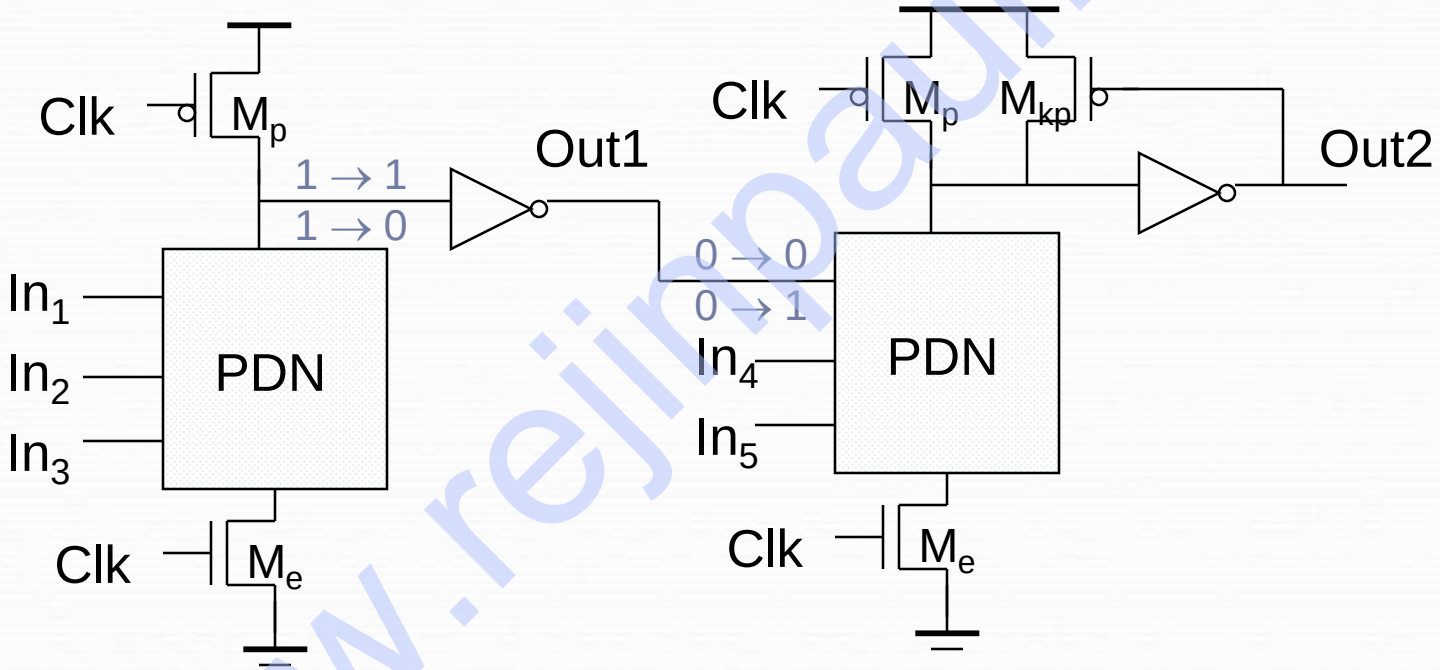
- Monotonically problem can be solved by placing a static CMOS inverter between dynamic gates→domino gates.
- The basic problem of dynamic CMOS logic→output node to be disconnected from V_{DD} or from GND,due to parasitic capacitance still leak the charge which leads logic error.
- Hence a finite time needed to charge for escaping from logic error.(faster clock speed)

Properties

- Only-one inverting logic implemented
- Very high speed
- Easy to reorganize the logic and differential logic(dual rail)
- Use np-CMOS(Zipper)

www.rejinpaul.com

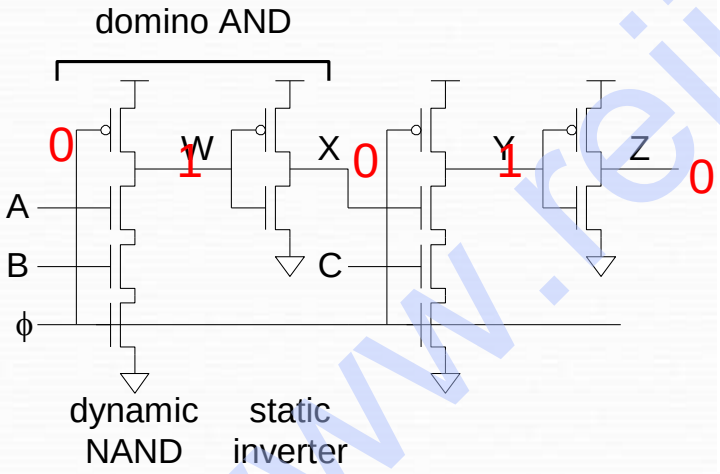
Generalized diagram of Domino Logic

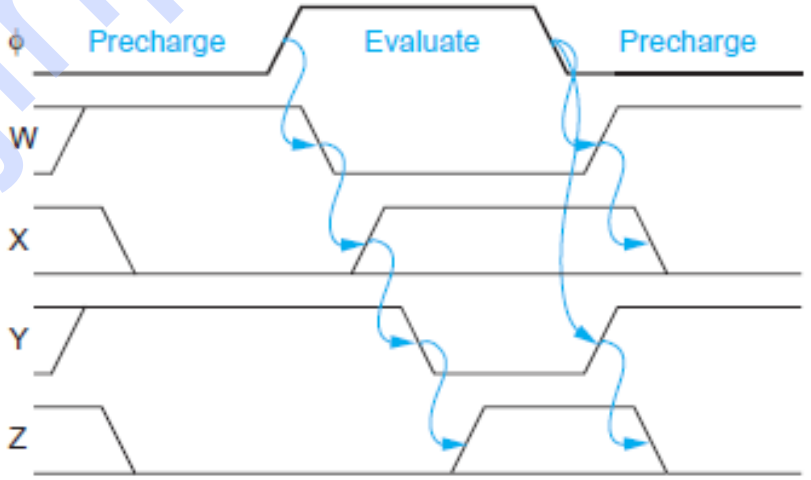


www.rejinpaul.com

1)Domino Gates

- It converts the monotonically falling output into a monotonically rising signal suitable for the next gate.
- A single clock can be used to pre-charge and evaluate all the logic gates within the chain.
- The dynamic output is monotonically falling during evaluation, so the static inverter output is monotonically rising.
- Precharge occurs in parallel, but evaluation occurs sequentially.

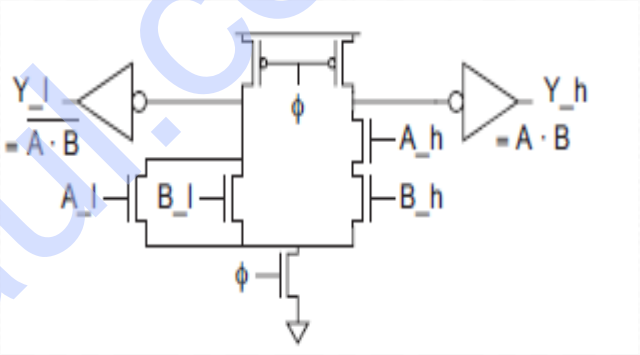


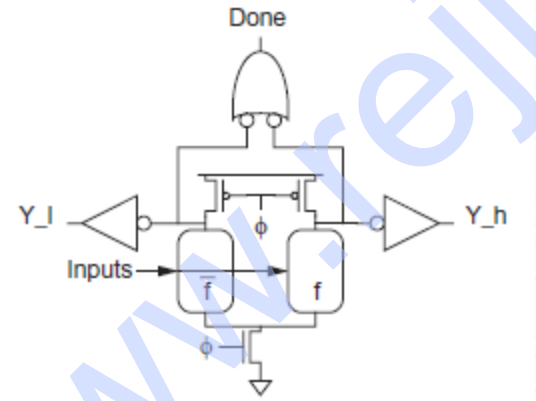


www.rejinpaul.com

2)Dual-Rail Domino Logic

- Domino only performs non-inverting functions(AND, OR but not NAND, NOR, or XOR)
- Dual-rail domino gates accept both true and
- complementary inputs and compute both true and complementary outputs.
- Before computation completes, both rails are precharged.
- When the computation completes, one rail will be asserted.
- NAND gate can be used for completion detection



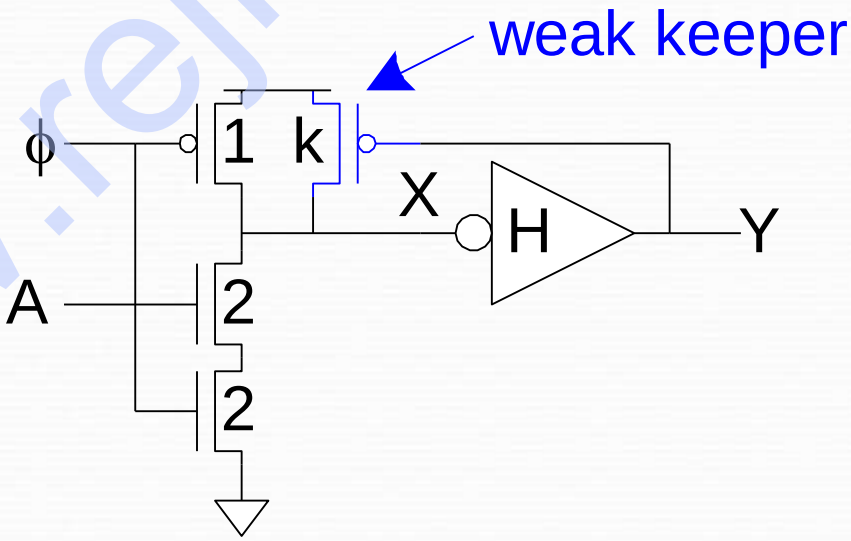


sig_h	sig_l	Meaning
0	0	Precharged
0	1	'0'
1	0	'1'
1	1	invalid

www.rejinpaul.com

Keepers

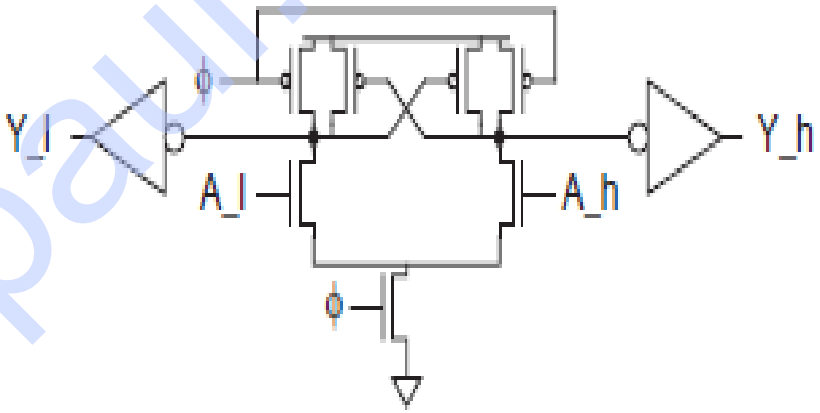
- Dynamic circuits also suffer from charge leakage on the dynamic node. If a dynamic node is pre-charged and then left floating.
  - Due to capacitor leakage.
- If the input rises above Vt ,while the gate is in evaluation, the input transistors will turn on weakly and can incorrectly discharge the output.
- The keeper is a weak transistor that holds the output at the correct level .
- When the dynamic node X is high, the output Y is low and the keeper is ON to prevent X from floating.
- When X falls, eventually Y rises, turning the keeper OFF and avoiding static power dissipation.



www.rejinpaul.com

Differential keeper

- When the gate is precharged, both keeper transistors are OFF and the dynamic outputs float.
- As soon as one of the rails evaluates low, the opposite keeper turns ON.
- The differential keeper is fast because it does not oppose the falling rail.
- As long as one of the rails is guaranteed to fall promptly, the keeper on the other rail will turn on before excessive leakage or noise causes failure.



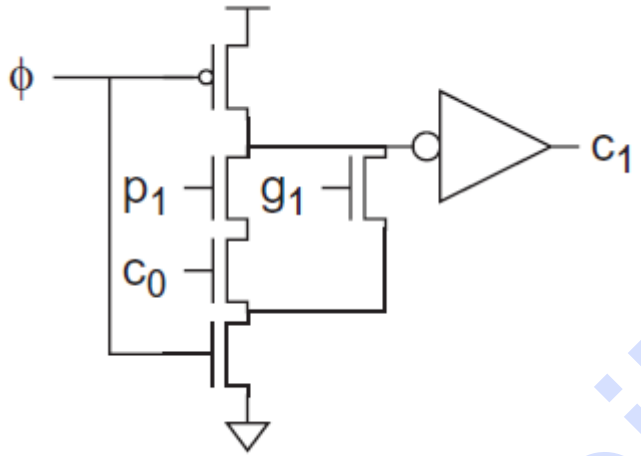
www.rejinpaul.com

Multiple-Output Domino Logic (MODL)

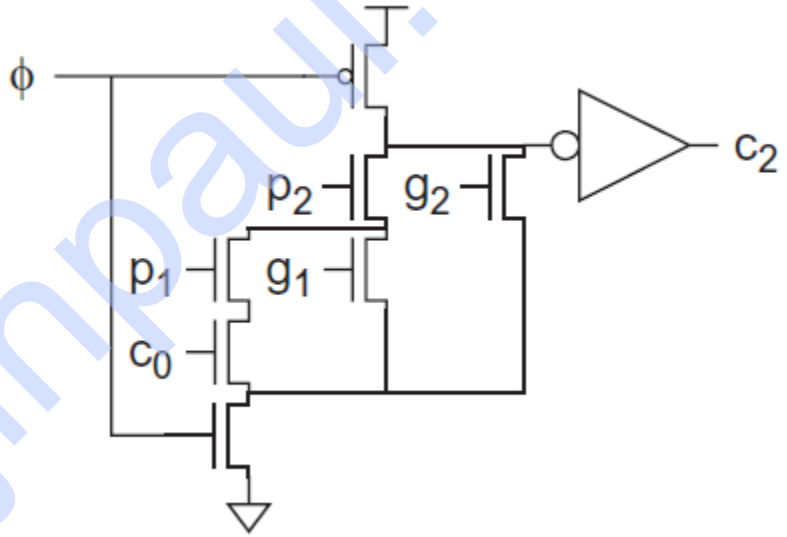
- Multiple-output domino logic (MODL) saves area by combining all of the computations into a multiple-output ga

$$c_1 = g_1 + p_1 c_0$$
$$c_2 = g_2 + p_2(g_1 + p_1 c_0)$$
$$c_3 = g_3 + p_3(g_2 + p_2(g_1 + p_1 c_0))$$
$$c_4 = g_4 + p_4(g_3 + p_3(g_2 + p_2(g_1 + p_1 c_0)))$$

www.rejinpaul.com

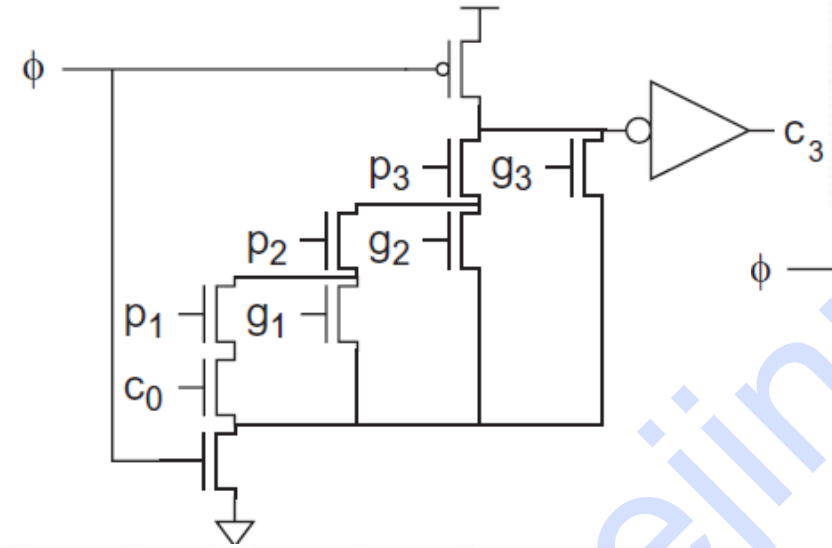


$$c_1 = g_1 + p_1 c_0$$

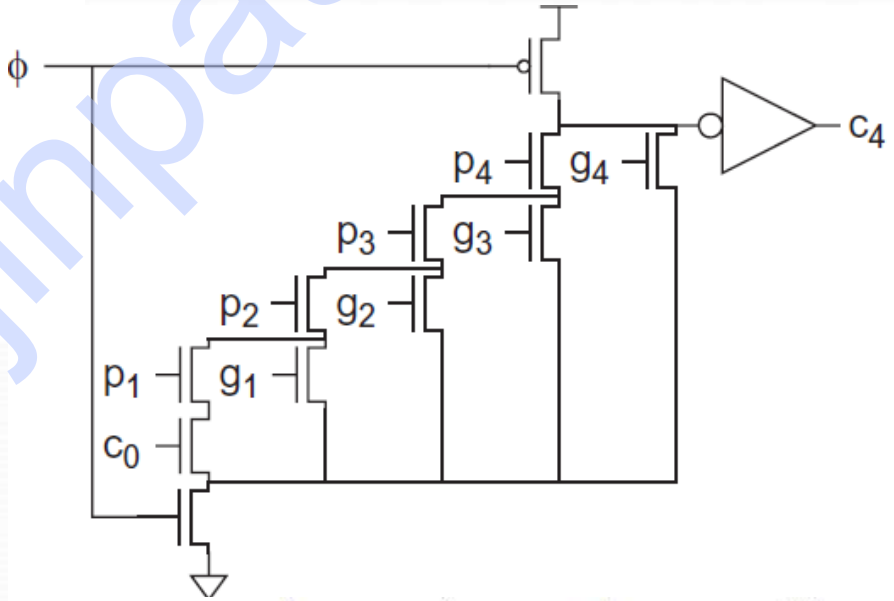


$$c_2 = g_2 + p_2(g_1 + p_1 c_0)$$

www.rejinpaul.com



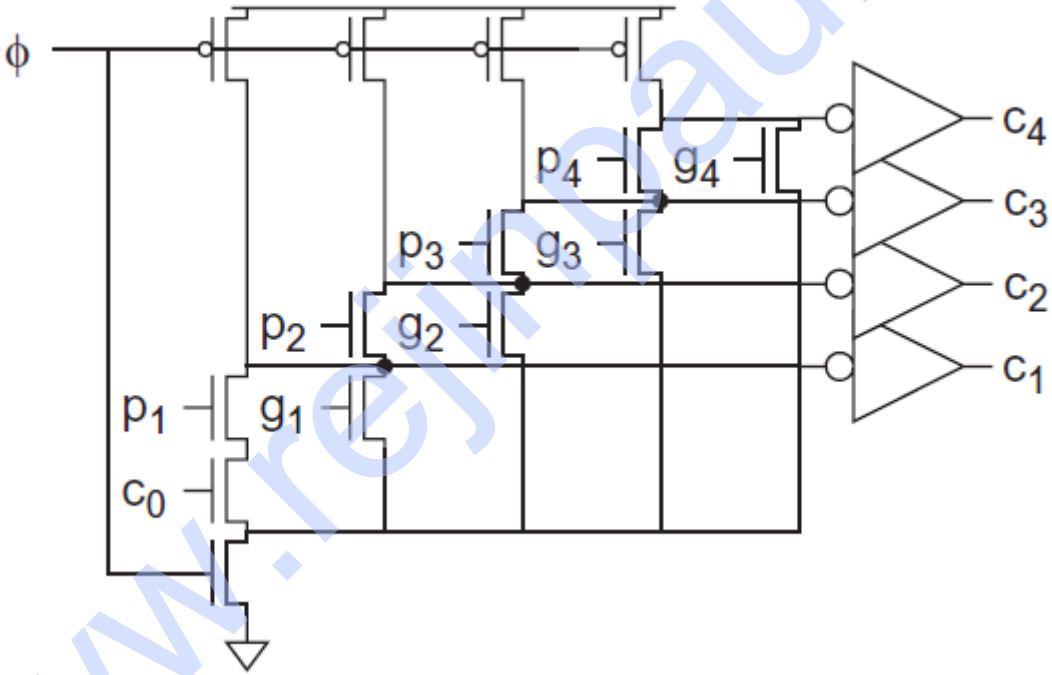
$$c_3 = g_3 + p_3\left(g_2 + p_2\left(g_1 + p_1 c_0\right)\right)$$



$$c_4 = g_4 + p_4\left(g_3 + p_3\left(g_2 + p_2\left(g_1 + p_1 c_0\right)\right)\right)$$

www.rejinpaul.com

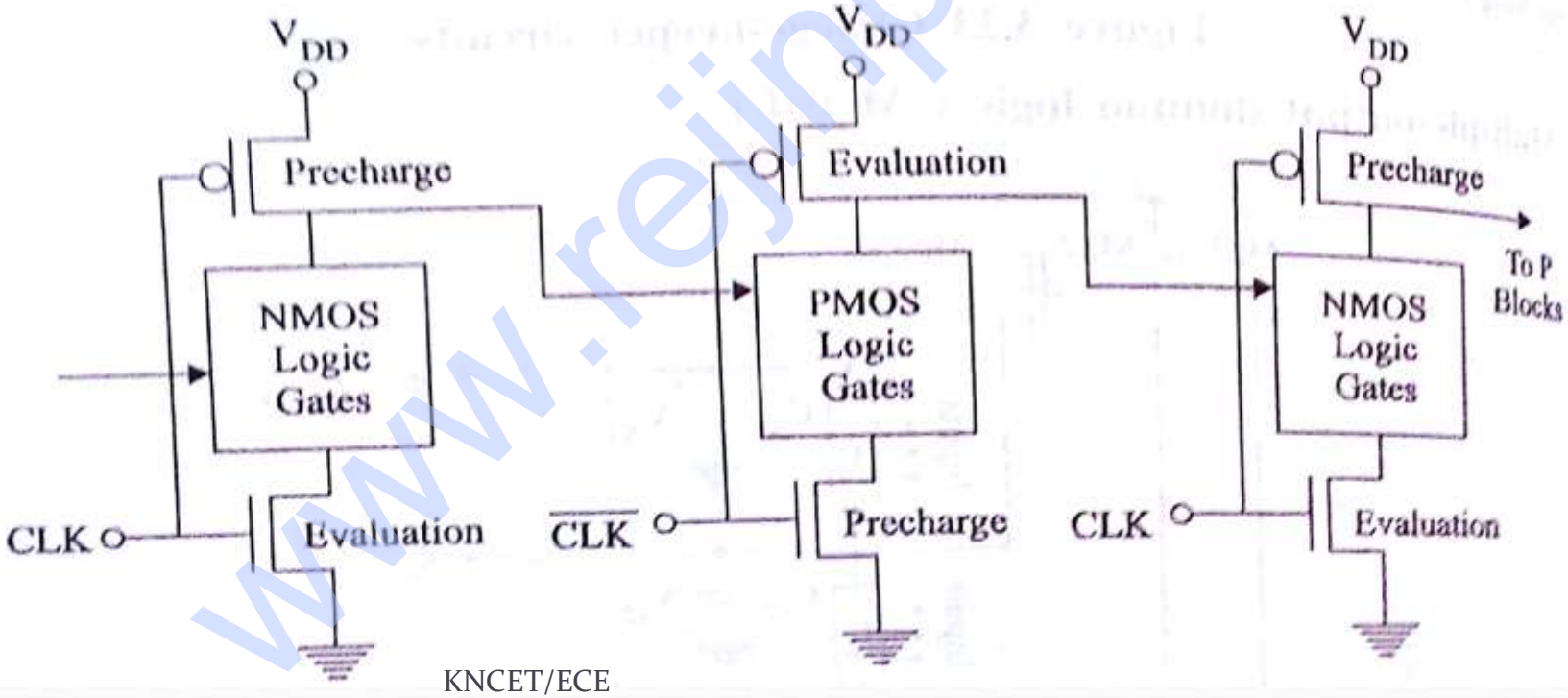
Conventional and MODL carry chains



www.rejinpaul.com

NP-Zipper CMOS

- NORA→(No –Race)→elliminate racing problem.
- The NP-Domino or NORA logic is very susceptible to noise and leakage.
- Zipper Domino has the same structure, but the precharge transistors are left slightly ON during evaluation.



www.rejinpaul.com

Comparison of Logic families

	Ease of Design	Robust-ness	Area	Speed	Power
Static	Very good	Very good	Bad	Bad	Good
Pseudo-nMOS	Average	Average	Good	Good	Very bad
Pass transistor	Difficult	Average	Good (for specific circuits)	Good (for specific circuits)	average
Dynamic logic	Very difficult	Very bad	Good	Good	average

www.rejinpaul.com

Power dissipation

- Static Power dissipation→ due to leakage current when ckt is in idle mode.
- Dynamic Power dissipation→ when the ckt is working due to charging and discharging of capacitor.
- Short circuit Power dissipation→ during switching

www.rejinpaul.com

- Instantaneous Power:

The instantaneous power drawn from the power supply is proportional to the supply current $i_{DD(t)}$ and the supply voltage V_{DD}.

$$P_{VDD}(t) = I_{DD}(t) V_{DD}$$

Energy: the energy consumed over the time interval T is the integral of the instantaneous power

$$E = \int_0^T P(t)dt$$

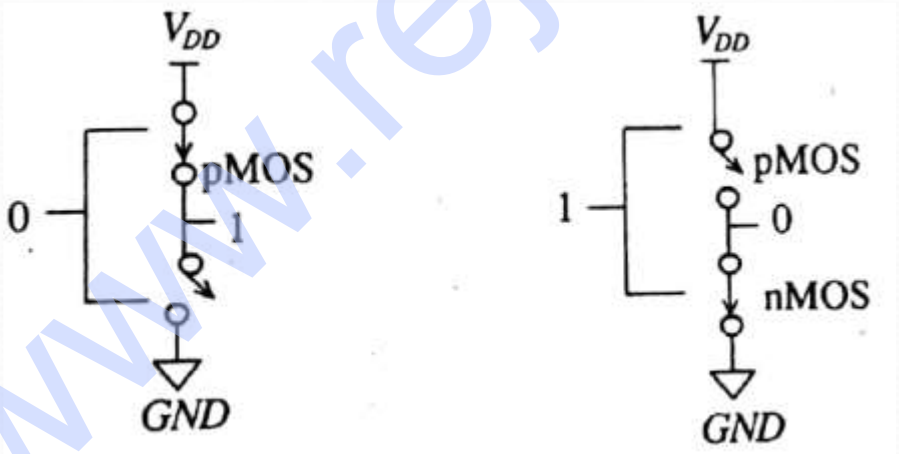
Avg Power: the average power is given by

$$P_{avg} = \frac{E}{T} = \frac{1}{T}\int_0^T P(t)dt$$

www.rejinpaul.com

Static Power Dissipation:

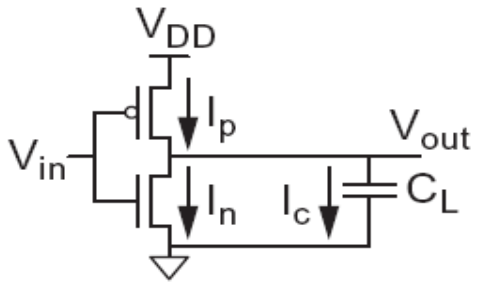
- Reason
  - Subthreshold leakage
  - Gate leakage
  - Junction leakage
  - Contention current
  - $P_{static} = (I_{sub} + I_{gate} + I_{junct} + I_{contention})V_{DD}$



www.rejinpaul.com

Dynamic Power Dissipation

- Reason
  - Charging and Discharging of load capacitance
  - Short circuit current while both pMOS and nMOS networks are partially ON
  - $P_{dynamic} = P_{switching} + P_{shortcircuit}$
  - The avg power dissipation is:



$$P_{dynamic} = \frac{1}{T}\int_0^T i_{DD}(t)V_{DD}dt = V_{DD}/T\int_0^T i_{DD}(t)dt$$

  - Taking integral of the current over the interval T,we get

$$P_{dynamic} = \frac{V_{DD}}{T}\left[Tf_{sw}\ CV_{DD}\right] = CV_{DD}^2 f_{sw}$$

  - fsw = α times the clock frequency f

$$P_{dynamic} = \alpha CV_{DD}^2 f$$

www.rejinpaul.com

Dynamic Power reduction:

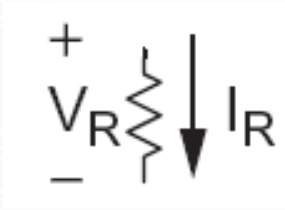
- Activity Factor
- Switching Capacitance
- Power Supply
- Operating Frequency

www.rejinpaul.com

Power in Circuit Elements

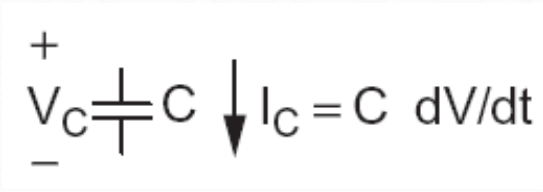
$$P_{VDD}(t) = I_{DD}(t) V_{DD}$$

$$P_R(t) = \frac{V_R^2(t)}{R} = I_R^2(t) R$$



$$E_C = \int_0^{\infty} I(t) V(t)\, dt = \int_0^{\infty} C \frac{dV}{dt} V(t)\, dt$$

$$= C \int_0^{V_C} V(t)\, dV = \tfrac{1}{2} C V_C^2$$



www.rejinpaul.com

1)Static Power

- Static power is consumed even when chip is off.(millions of transistor produce leakage current when temp increases)
- Estimate static power consumption.
  - **Subthreshold leakage**
  - **Gate leakage** 5 nA/μm
  - **Junction leakage**
  - $P_s = \Xi^{n}_{i=1}(I_{static} \times \text{Supply voltage})$

www.rejinpaul.com

2) Dynamic Power dissipation

- When the gate output rises
  - Energy stored in capacitor is

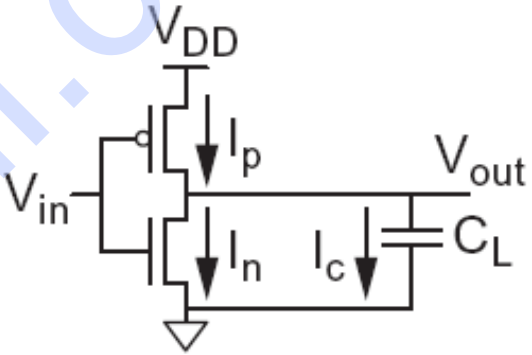
$$E_C = \frac{1}{2} C_L V_{DD}^2$$

  - But energy drawn from the supply is

$$E_{VDD} = \int_0^\infty I(t) V_{DD} dt = \int_0^\infty C_L \frac{dV}{dt} V_{DD} dt$$

$$= C_L V_{DD} \int_0^{V_{DD}} dV = C_L V_{DD}^2$$

  - Half the energy from V_{DD} is dissipated in the pMOS transistor as heat, other half stored in capacitor
- When the gate output falls
  - Energy in capacitor is dumped to GND
  - Dissipated as heat in the nMOS transistor



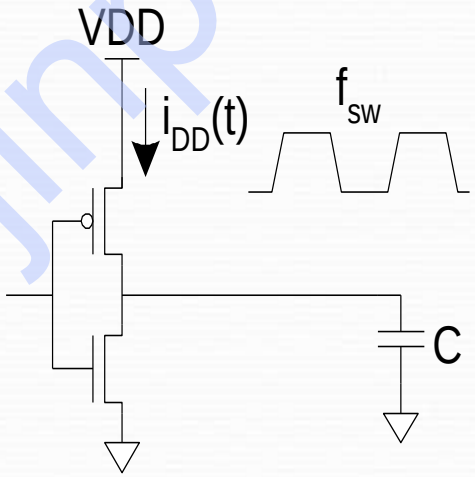
$$P_{\text{switching}} = \alpha C V_{DD}^2 f$$

www.rejinpaul.com

3)Short Circuit Current

- During switching,both nMOS and pMOS will conduct and provide a direct path between V_{DD} for short circuit power dissipation.

- $P_{short} = V_{DD} X I_{short}$



www.rejimpaul.com

Total Power Dissipation

- Static power→ $P_{static} = (I_{sub} + I_{gate} + I_{junct} + I_{contention})V_{DD}$
- Dynamic power→ $P_{dynamic} = P_{switching} + P_{shortcircuit}$
- Short Circuit Power→ $P_{short} = V_{DD} \text{ X } I_{short}$
- Total Power→ $P_{total} = P_{dynamic} + P_{static} + P_{short}$

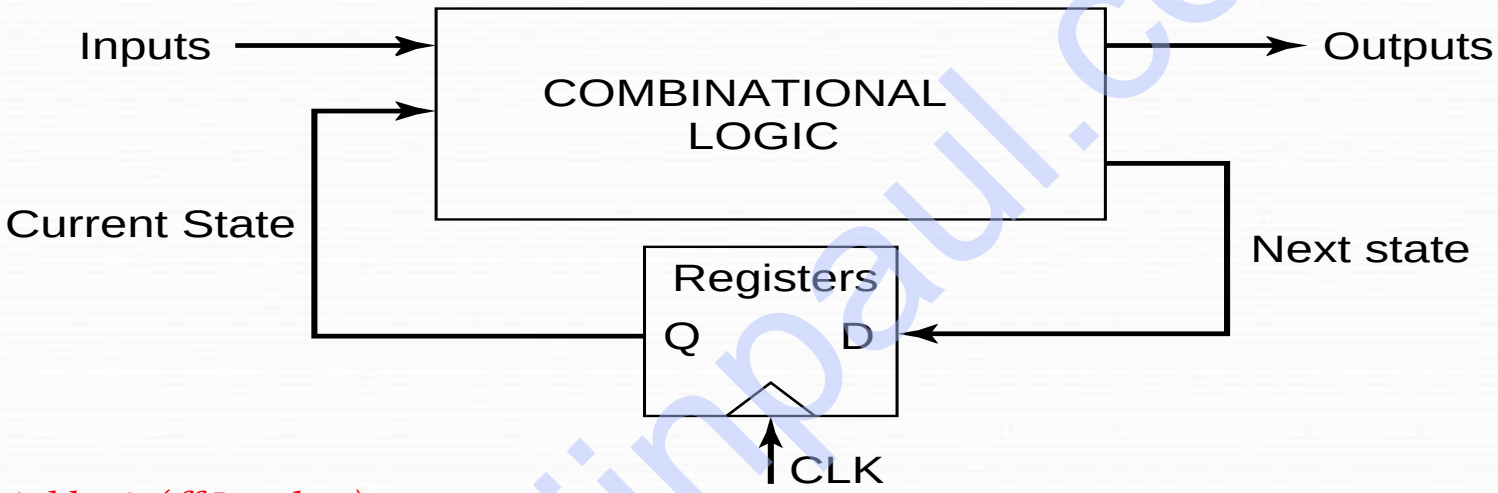
www.reinpaul.com

Unit-III

SEQUENTIAL LOGIC CIRCUITS

www.rejinpaul.com

Introduction



Sequential logic(ff,Latches)

- output not only depends upon the *current values of the inputs, but also upon preceding input values.*
- Eg:1→finite state machine (FSM) that consists of combinational logic and registers, which hold the system state.
- On the rising edge of the clock, the Next State bits are copied to the outputs of the registers (after some propagation delay), and a new cycle begins.

www.rejinpaul.com

Static versus Dynamic Memory

- Static memories →preserve the state as long as the power is turned on. Static memories are built using *positive feedback or regeneration.*
- Dynamic memories →store state for a short period of time—on the order of milliseconds.
- Based on the principle of temporary *charge storage on parasitic capacitors* associated with MOS devices.
- It has higher performance and lower power dissipation than static.

www.rejinpaul.com

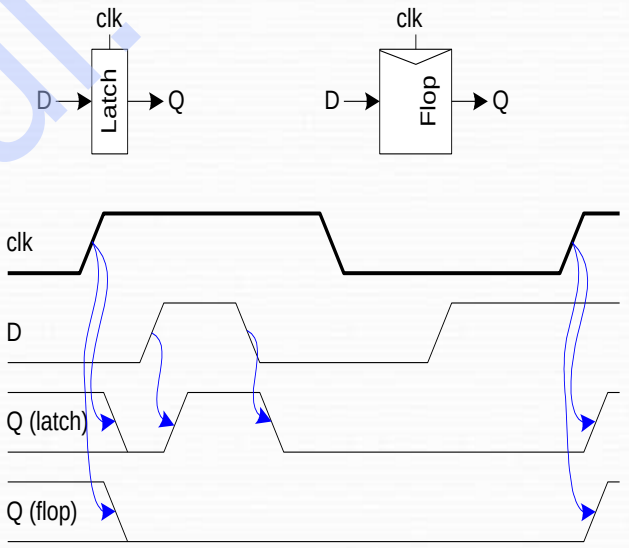
Latches versus Registers

Latch

- CLK is high→ Input is transferred from D to Q
- Transparent→Output follows changes in the data input as long as CLK is high.
- CLK is low→ input is transferred to output Q and is maintained by same until the CLK goes to high

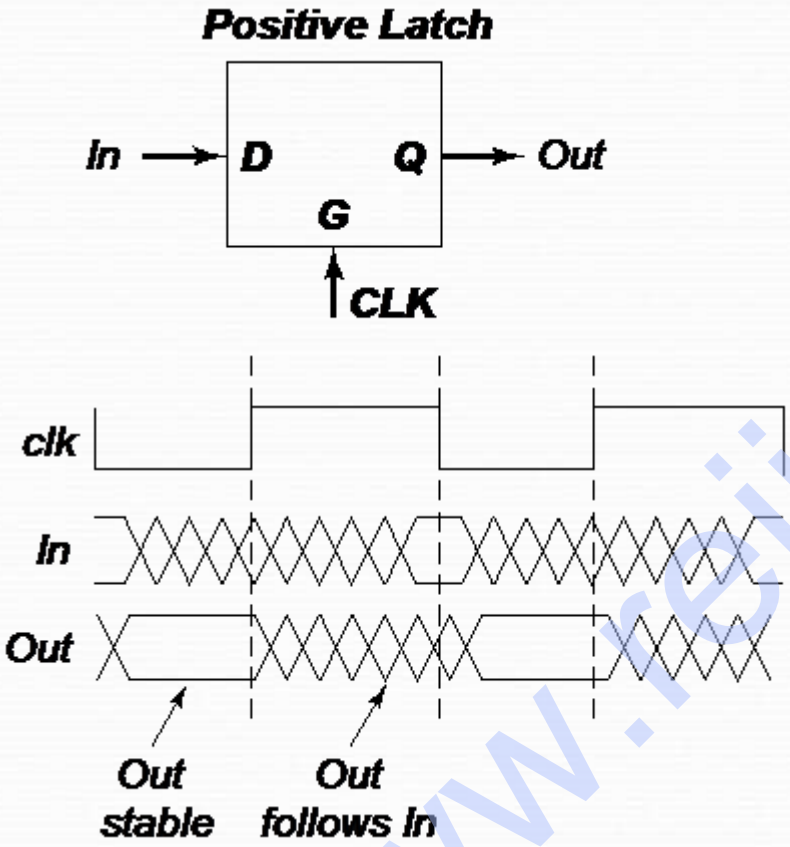
Flip-flop

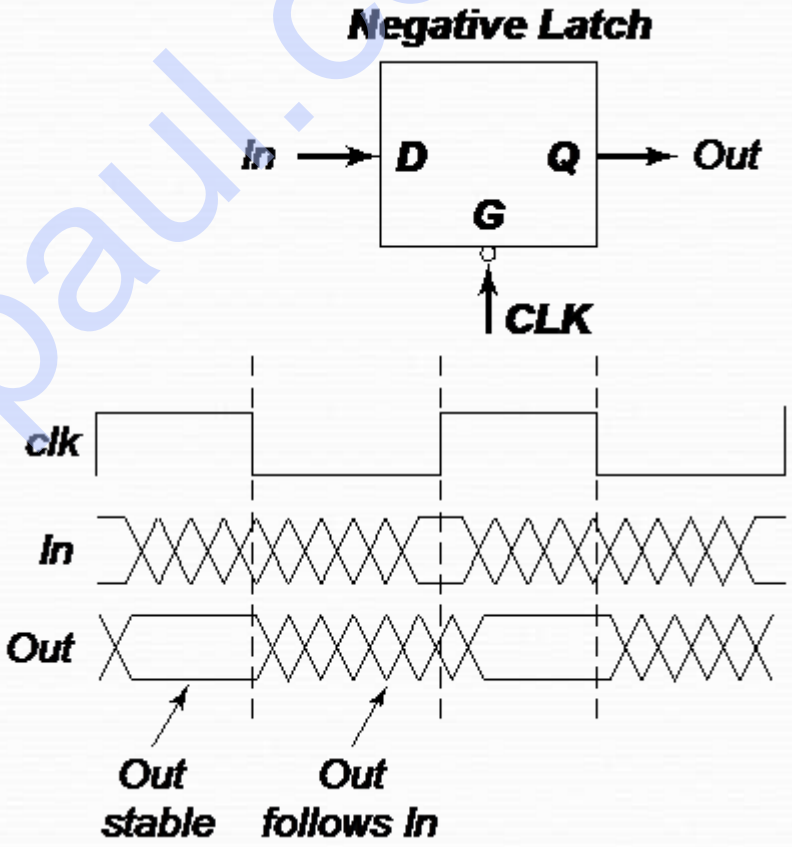
- CLK rising edge→Input D copied to Q.
- Ignore D all other times.



www.rejinpaul.com

Latches

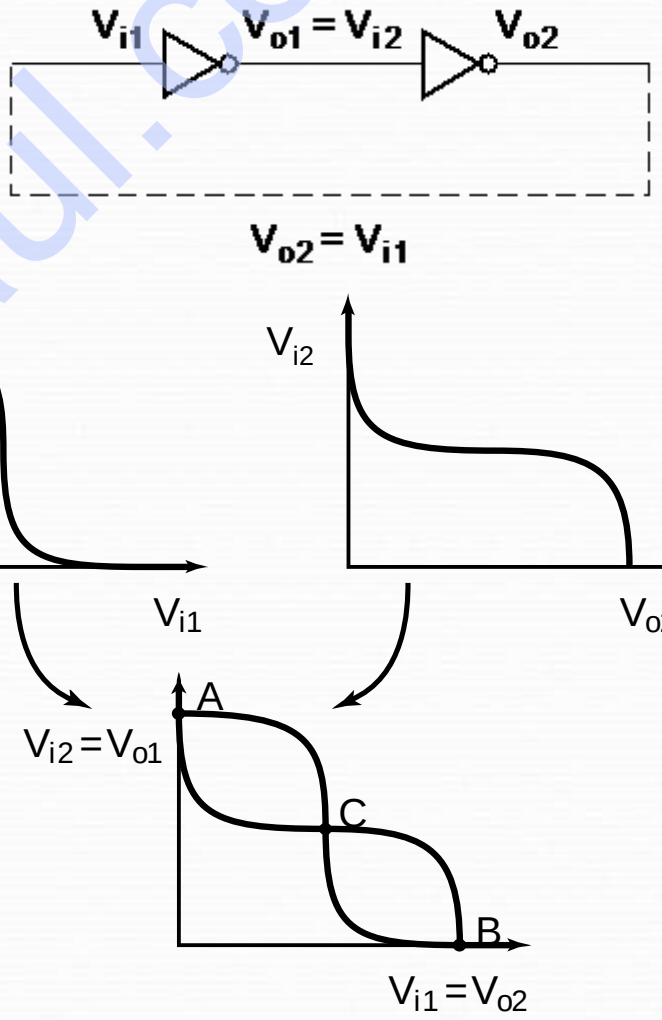




www.rejinpaul.com

Static Latches and Registers

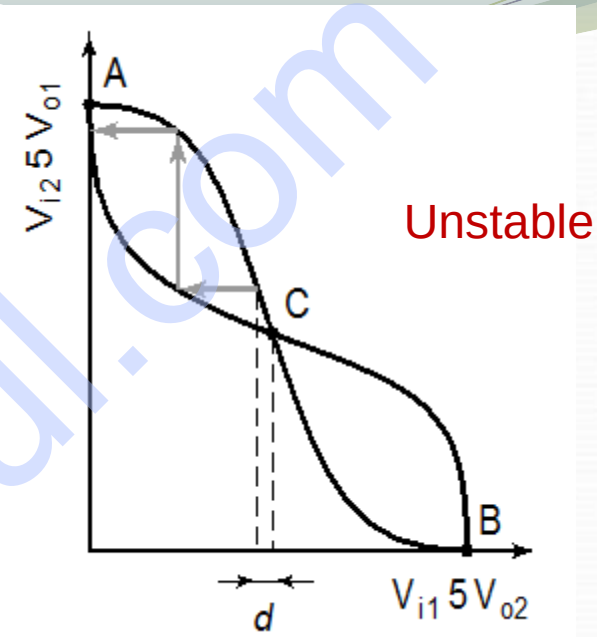
- Bistability Principle
- Static memories use positive feedback to create a *bistable circuit — having two* stable states that represent 0 and 1.
- Two inverters connected in cascade.
- Plot V_{01} → Voltage Transfer Curve (VTC) of the first inverter
- Plot v_{02} → VTC of the second inverter
- Plot $V_{01}=v_{02}$ → VTC (output of the second inverter
- *Vo2 is connected to the input of the first Vi1)*



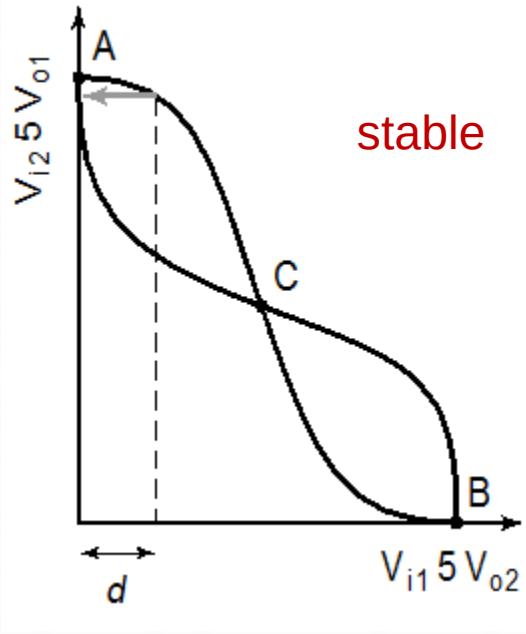
www.rejinpaul.com

Meta-Stability Problem

- The gain around the loop being larger than 1.
- A small deviation d is applied to Vi1 (biased in C). This deviation is amplified by the gain of the inverter.
- The enlarged divergence is applied to the second inverter and amplified once more.
- The bias point moves away from C until one of the operation points A or B is reached. In conclusion, C is an unstable operation point.



- A and B are stable operation points, in these points, the loop gain is much smaller than unity
- cross-coupling of two inverters results in a bi-stable circuit, with two stable states(0 or 1)
- The circuit serves as a memory, storing either a 1 or a 0



www.rejinpaul.com

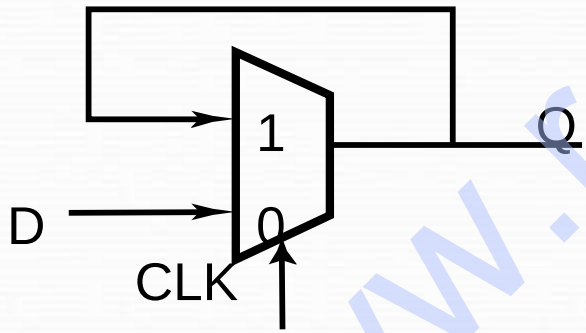
1)Mux-Based Latches

Negative latch
(transparent when CLK= 0)

•When the clock signal is Low, the input 0 of the multiplexer, which connects to the output (0) of the latch, is selected.

•when clock is high, the output(1) is received from feedback.

•The feedback holds the output stable while the clock signal is low.



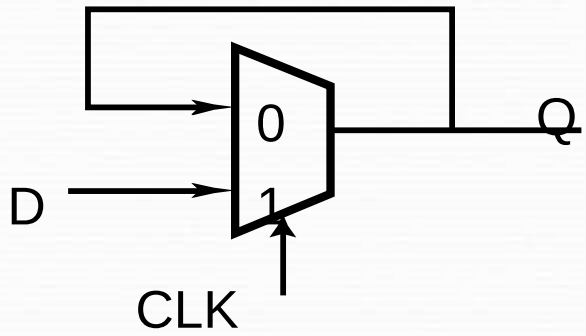
$$Q = \overline{Clk} \cdot Q + Clk \cdot In$$

Positive latch
(transparent when CLK= 1)

•When the clock signal is high, the input 1 of the multiplexer, which connects to the output (1) of the latch, is selected.

•when clock is low, the output(0) is received from feedback.

•The feedback holds the output stable while the clock signal is high.



$$Q = Clk \cdot Q + \overline{Clk} \cdot In$$

Negative and positive latches based on multiplexers.

www.rejinpaul.com

Positive latch using transmission gates.

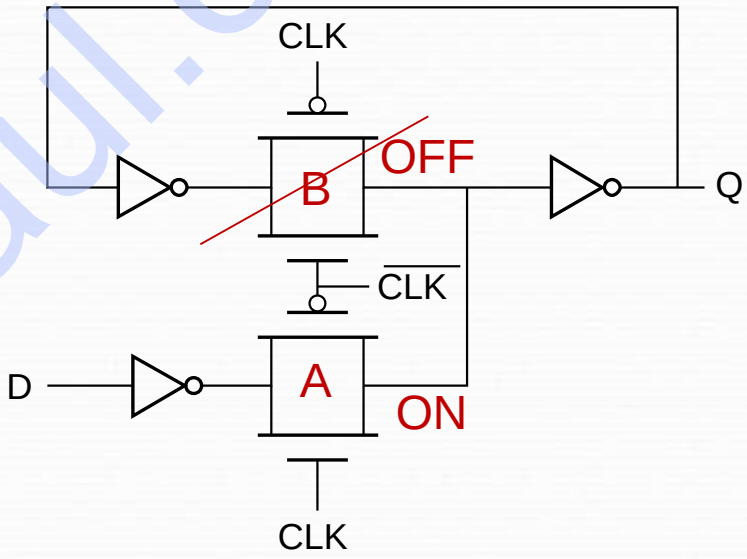
•When CLK is high, the bottom transmission gate (A) is on and D input is copied to the Q output.

•During this phase, the feedback loop is open since the top transmission gate(B) is off.

•The number of transistors that the clock touches is important since it has an activity factor of 1.

•This particular latch implementation is not particularly efficient from this metric as it presents a load of 4 transistors to the CLK signal.

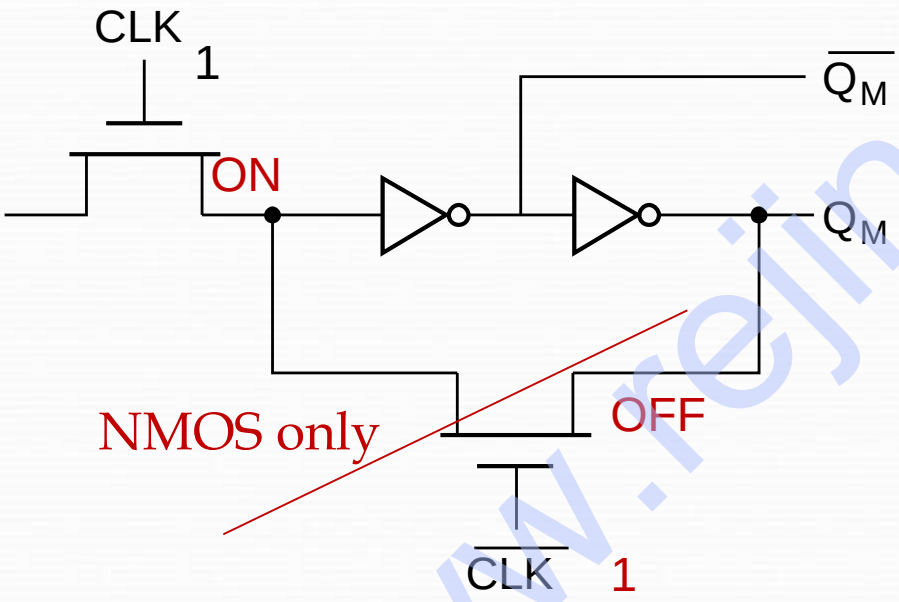
•It is possible to reduce the clock load to two transistors by using implement multiplexers using NMOS only pass transistor

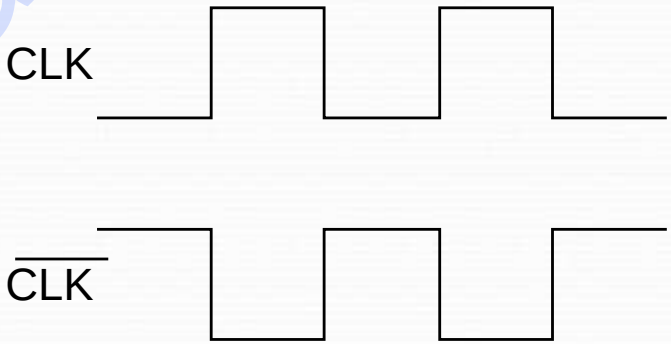


www.rejinpaul.com

Mux-Based Latch-NMOS only

- When CLK is high, the top nMOS is ON and bottom nMOS is off .
- The latch samples the D input to the output(Q_M)



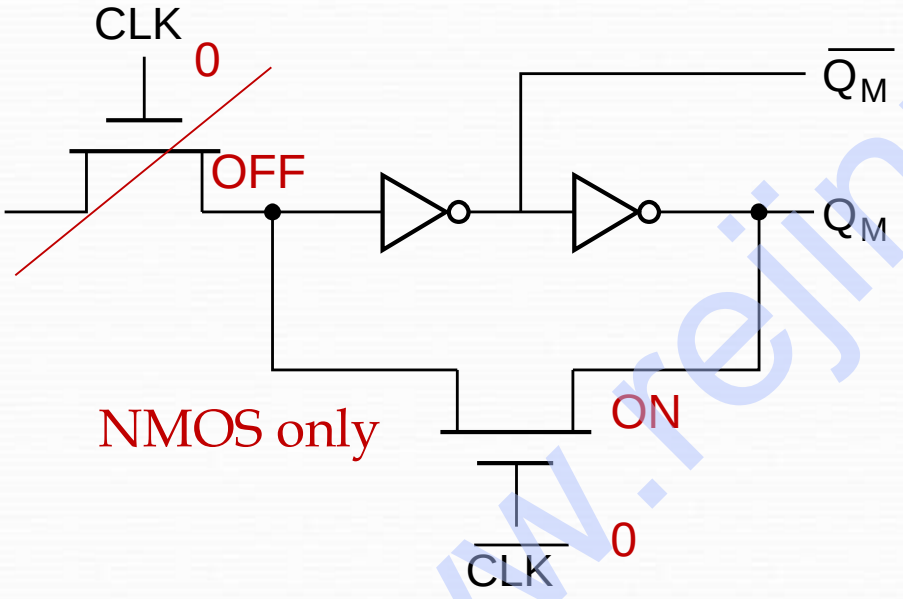


Non-overlapping clocks

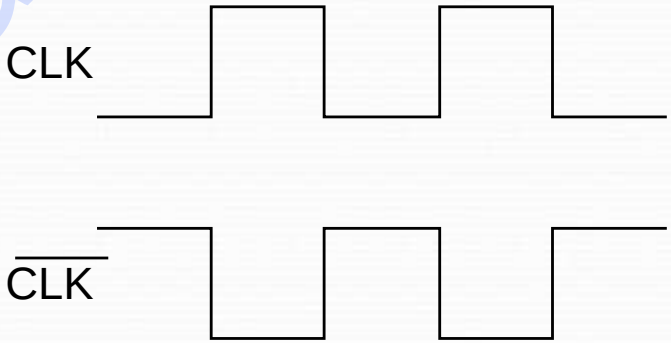
www.rejinpaul.com

Mux-Based Latch-NMOS only

- while a low clock-signal →top nMOS is OFF and bottom nMOS is ON.
- Hence the the feedback-loop enabled , and puts the latch in the hold mode.
- NMOS pass transistors passing a degraded high voltage of VDD-VTn to the input of the first inverter.
- This impacts both noise margin and the switching performance



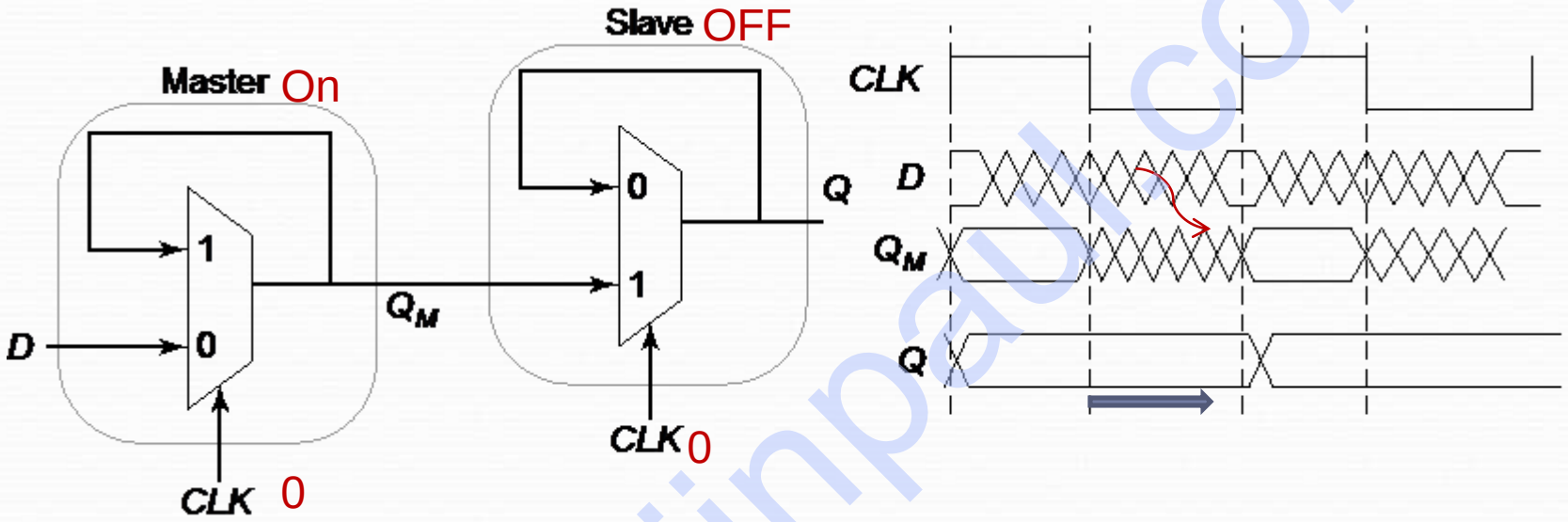
NMOS only



Non-overlapping clocks

www.rejinpaul.com

2) Master-Slave Edge-Triggered Register

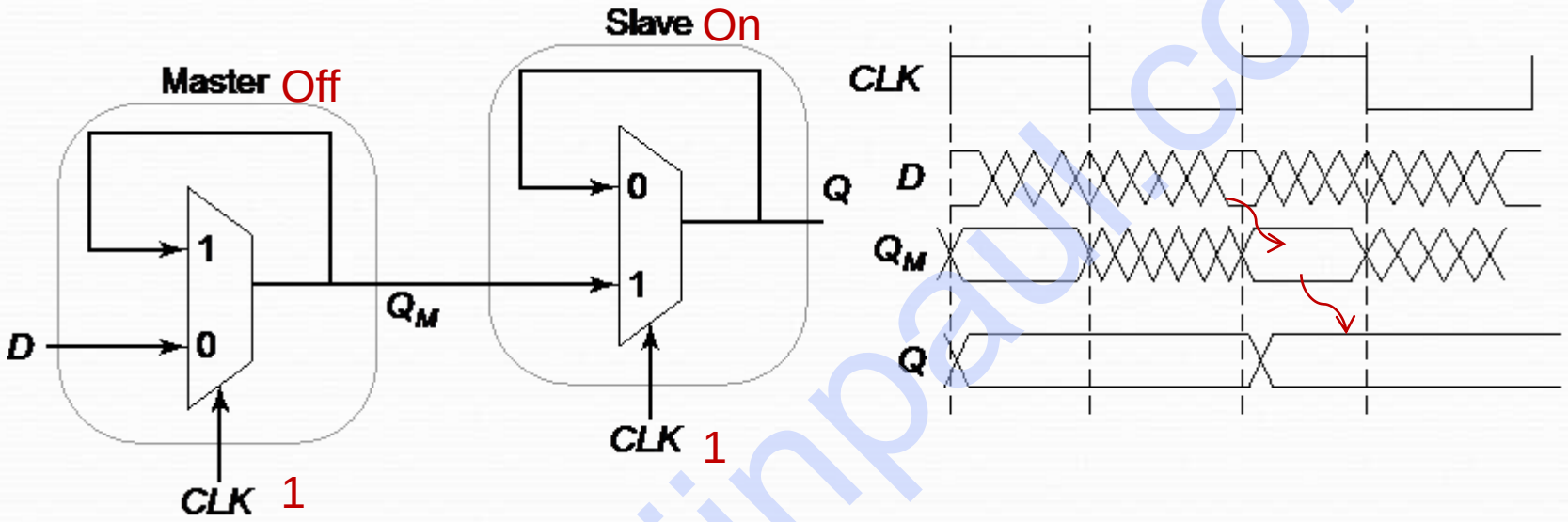


•The register consists of cascading a negative latch (master stage) with a positive latch (slave stage).

• If clock is Low, the master stage is ON and the D input is passed to the master stage output QM. During this period, the slave stage is OFF or hold mode, keeping its previous value using feedback.

www.rejinpaul.com

Master-Slave Edge-Triggered Register

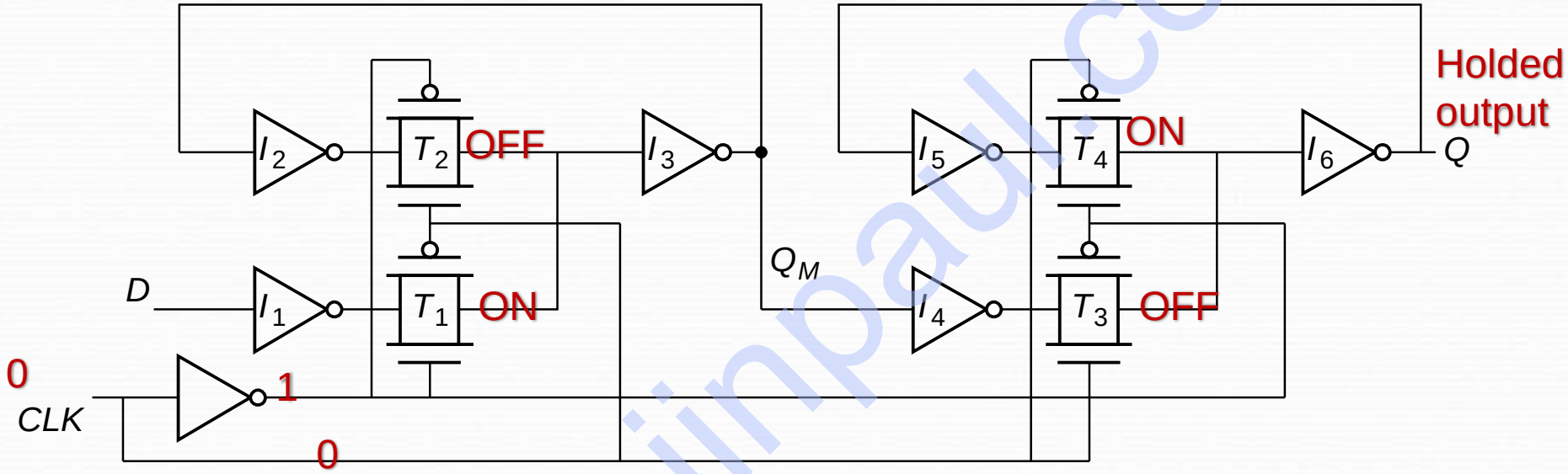


•On the rising edge of the clock,the slave stage is ON and copy the Q_M value to Q , while the master stage is off and remains in a hold mode.

•Since QM is constant during the high phase of the clock, the output Q makes only one transition per cycle.

www.rejinpaul.com

Transistor level Implementation (Master-Slave Positive edge-triggered Register)



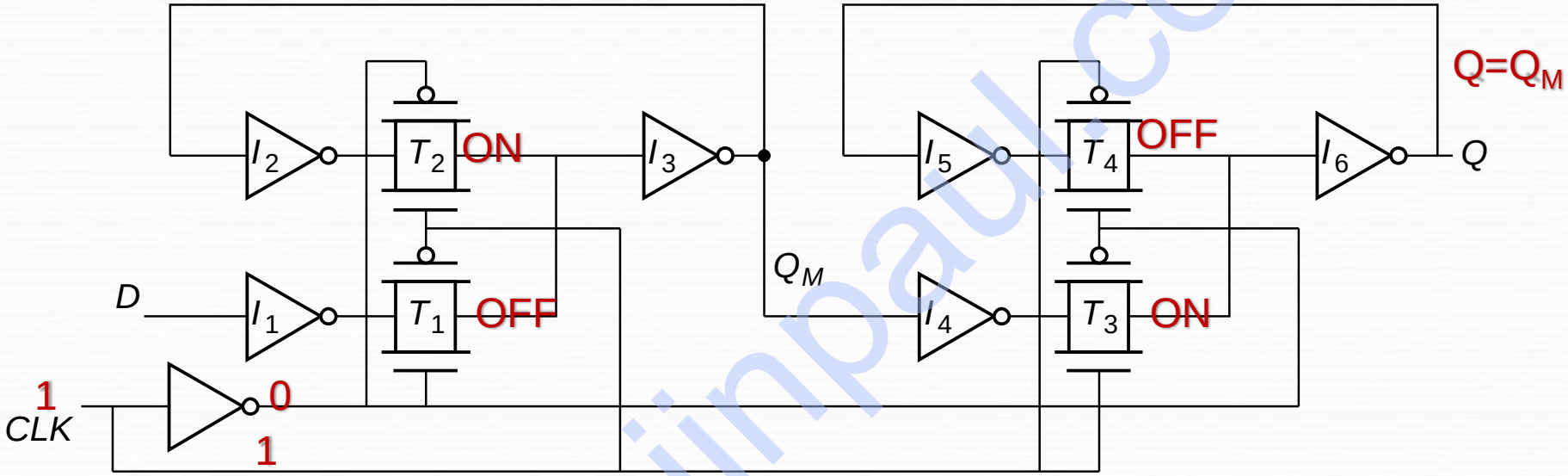
T1,T2→Master stage

T3,T4→Slave stage

•When the clock is low , T1 is on and T2 is off, and the D input is copied to QM via I3. During this period, T3 is off and T4 is on and the cross-coupled inverters (I5, I6) holds the state of the slave latch.

www.rejinpaul.com

Transistor level Implementation (Master-Slave Positive edge-triggered Register)



T1,T2→Master stage

T3,T4→Slave stage

•When the clock goes high, T1 is off and T2 is on, and the cross coupled inverters I3 and I4 holds the state of QM. Also, T3 is on and T4 is off, and QM is copied to the output Q.

www.rejinpaul.com

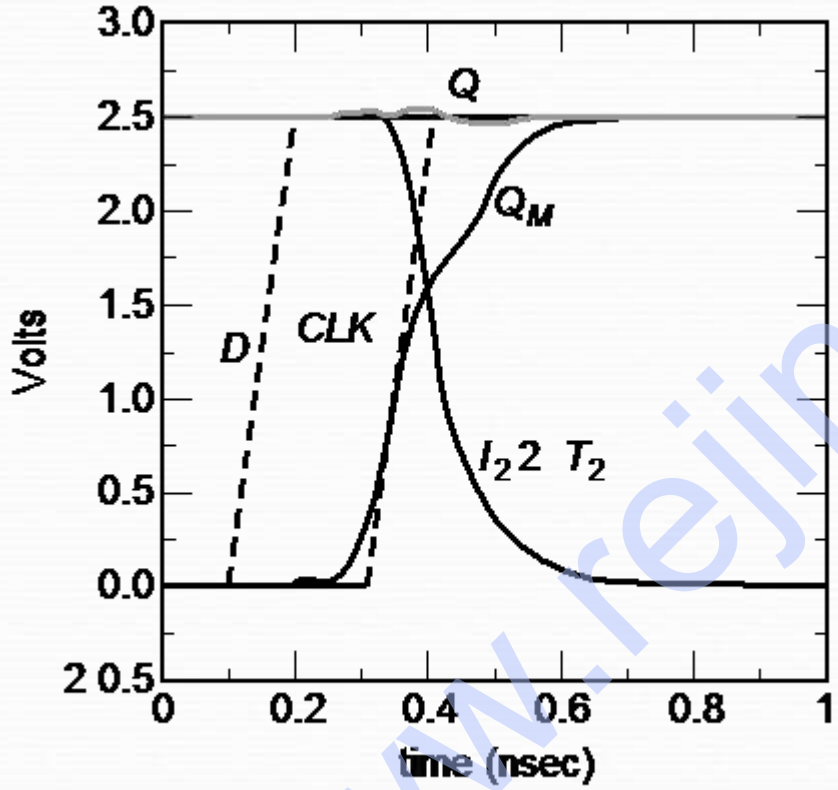
Timing Properties of Multiplexer-based Master-Slave Registers

•Registers are characterized by three important timing parameters.

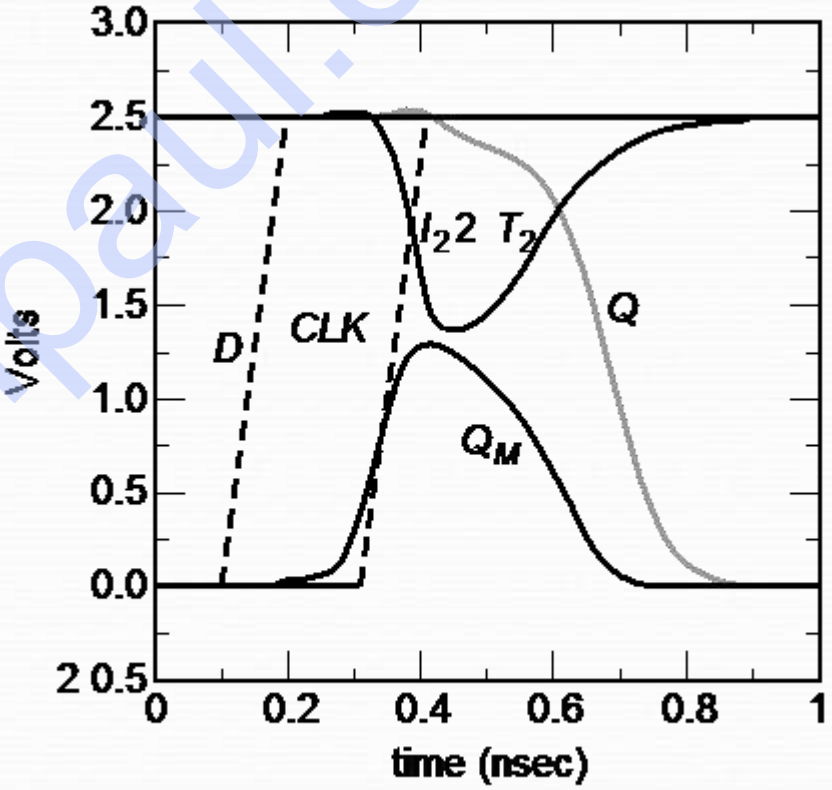
- set-up time →is the time before the rising edge of the clock that the input data D must become valid.
- propagation delay → time taken for the value of QM to propagate to the output Q.
- hold time→ represents the time that the input must be held stable after the rising edge of the clock.

www.rejinpaul.com

Setup Time



(a) T_{setup} 5 0.21 nsec



(b) T_{setup} 5 0.20 nsec

www.rejinpaul.com

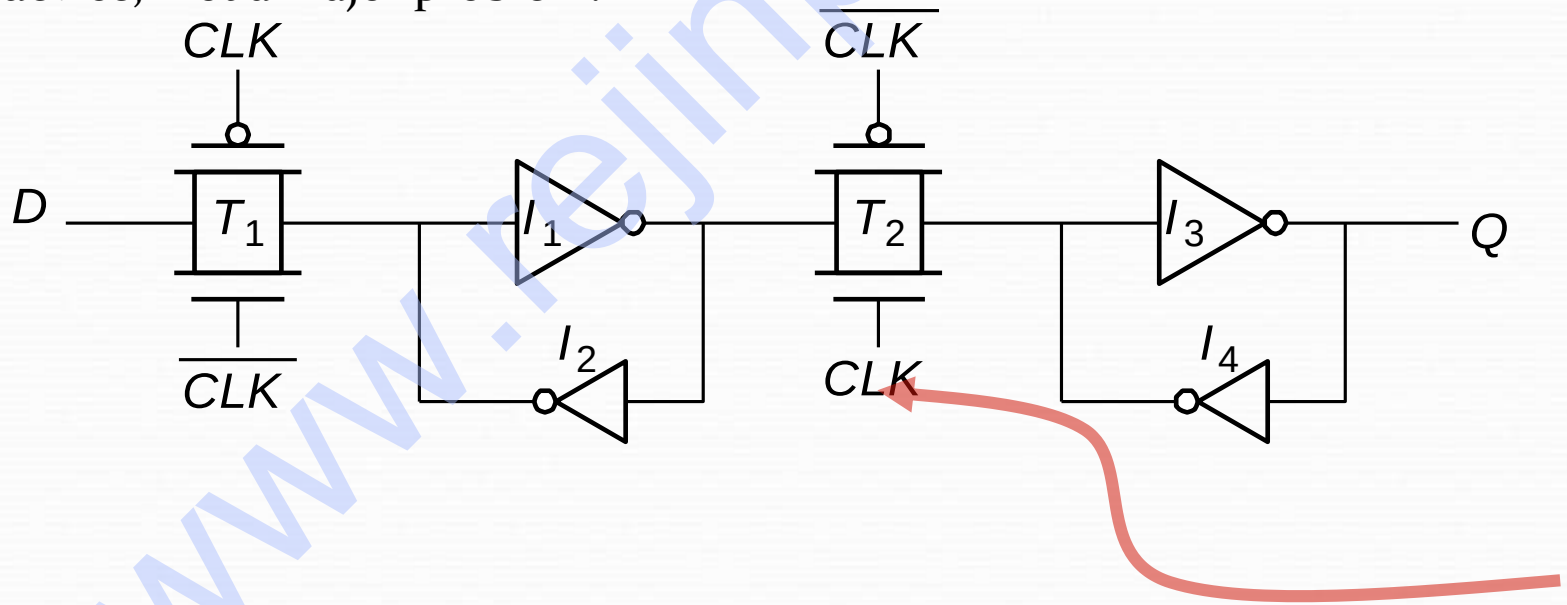
Drawback of the transmission gate register

- Due to high capacitive load presented to the clock signal. The clock load per register is important, since it directly impacts the power dissipation of the clock network.

www.rejinpaul.com

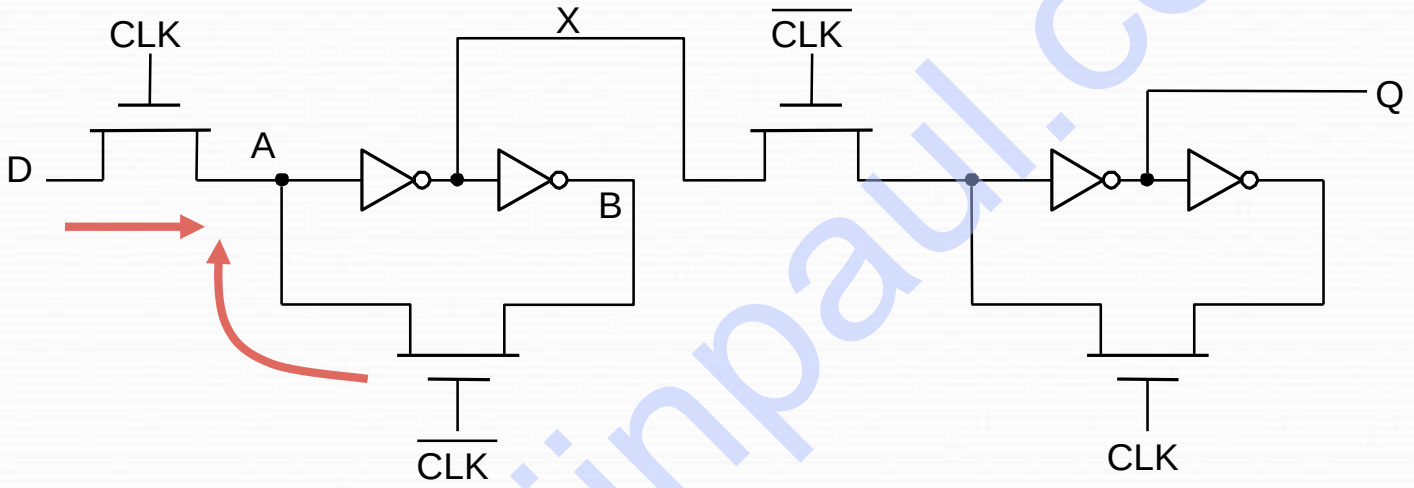
Reduced Clock Load Master-Slave Register

- The transmission gate (T1) and its source driver must overpower the feedback inverter (I2) to switch the state of the cross-coupled inverter.
- The input to the inverter *I1 kept below its switching threshold in order to make a transition.*
- It is essential that the transistors of inverter I2 should be made even weaker.
- The 2nd stage can affect the state of the Ist latch. When the slave stage is on , T2 and I4 to influence the data stored in I1-I2 latch. As long as I4 is a weak device, not a major problem.

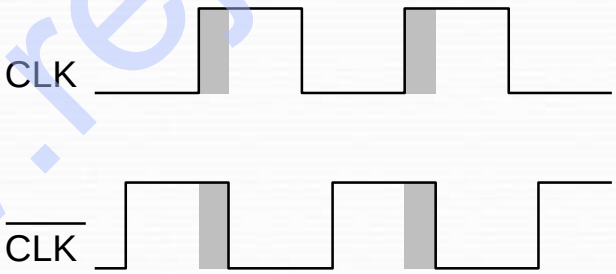


www.rejinpaul.com

Avoiding Clock Overlap



(a) Schematic diagram



(b) Overlapping clock pairs

www.rejinpaul.com

- When the clock goes high, the slave stage off and the master stage ON.
- If CLK and $\overline{CLK}$ are both high for a short period of time (the overlap period), both sampling pass transistors conduct.
- As a result, data at the output can change on the rising edge of the clock.
- This is undesired for a negative edge triggered register. The is know as a race condition
- Q is a function of whether the input D arrives at node X before or after the falling edge of CLK.
- If node X is sampled in the meta-stable state, the output will switch to a value determined by noise in the system.
- • The primary advantage of the multiplexer-based register is the feedback loop is open during the sampling period.
- If there is clock overlap between CLK and CLK, node A can be driven by both D and B, resulting in an undefined state.

www.rejinpaul.com

Dynamic Latches and Registers

Introduction

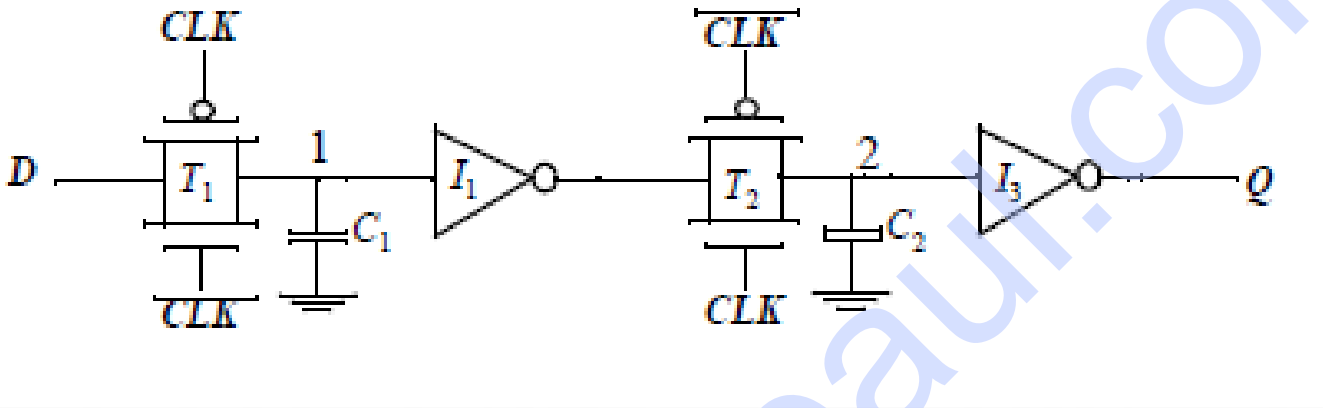
- Dynamic circuits based on temporary storage of charge on parasitic capacitors.
- Charge stored on a capacitor can be used to represent a logic signal.
- The absence of charge denotes a 0, while its presence stands for a stored 1.
- In capacitor charge leakage is always present. So the stored value can hence only be kept for a limited amount of time(milliseconds).

Types

1. Dynamic Transmission-Gate Edge-triggered Registers
2. C^2MOS—A Clock-Skew Insensitive Approach
3. True Single-Phase Clocked Register (TSPCR)

www.reinpaul.com

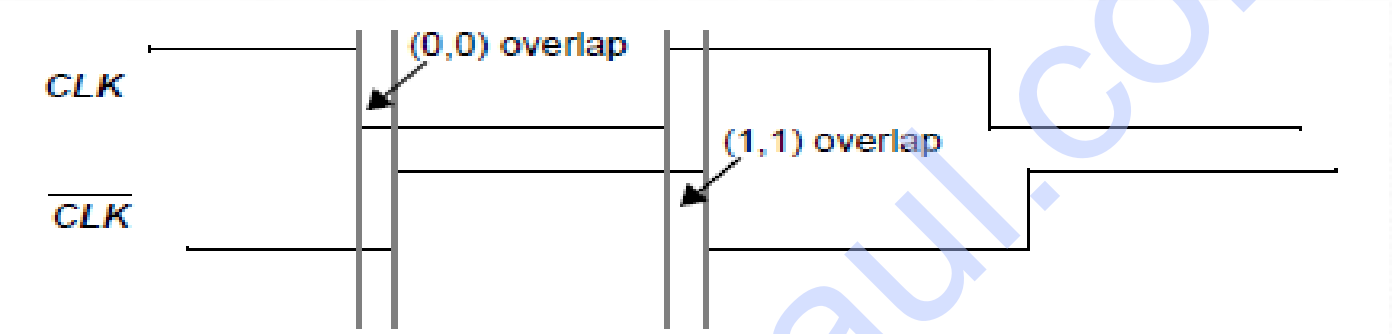
1)Dynamic Transmission-Gate Edge-triggered Registers



- When CLK = 0, the transmission gate T1 turns on, which has capacitance C1 and the junction capacitance of T1, and the overlap gate capacitance of T1.
- During this period, the slave stage(T2) is in a hold mode, with node 2 in a high-impedance (floating) state.
- When CLK = 1, the transmission gate T2 turns on, and the node1 value propagates to the output Q (node 1 is stable during the high phase of the clock when T1 is off).
- Node 2 now stores the inverted version of node 1.

www.rejinpaul.com

Impact of non-overlapping clocks

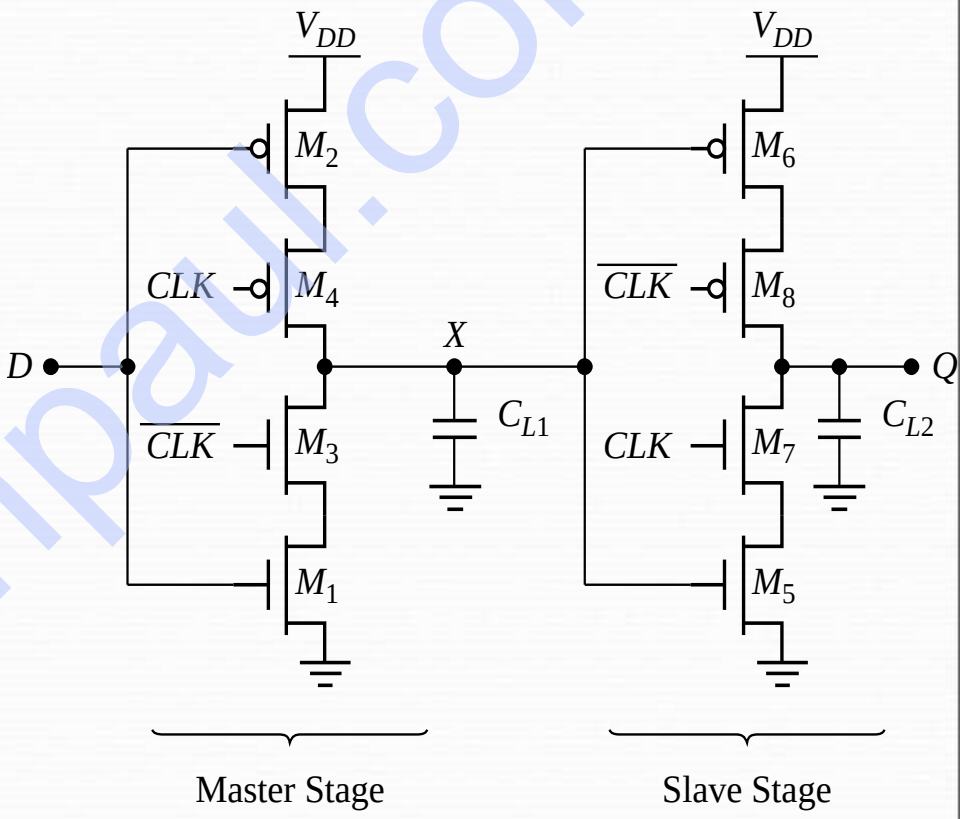


- During the 0-0 overlap period, the NMOS of T1 and the PMOS of T2 are simultaneously ON, creating a direct path for data to flow from the D input of the register to the Q output. This is known as a race condition.
- The output *Q can change* on the falling edge if the overlap period is large.
- During 1-1 overlap region, where an input-output path exists through the PMOS of *T1 and the NMOS of T2. The latter case is* taken care off by enforcing a *hold time constraint. That is, the data must be stable during* the high-high overlap period.
- constraint for the 0-0 overlap is given as: $t_{overlap0-0} < t_{T1} + t_{I1} + t_{T2}$
- constraint for the 1-1 overlap is given as: $t_{hold} > t_{overlap1-1}$

www.reinpaul.com

2.1)C²MOS master-slave positive edge-triggered register

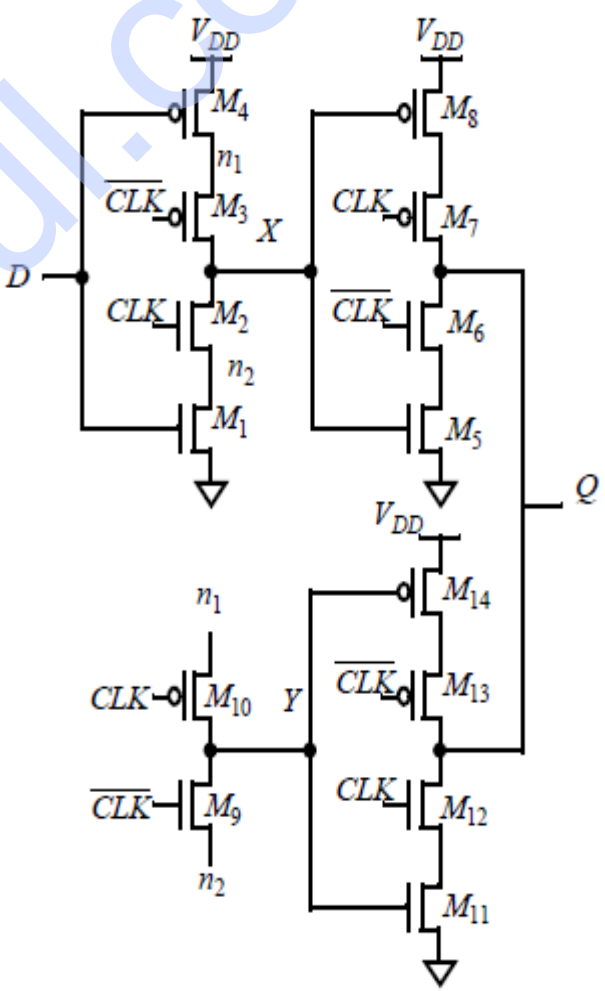
- *If CLK = 0 → The first tri-state driver is ON, and the master stage acts* as an inverter .
- *The master* stage(M3-M4) is in the *evaluation mode and the slave stage is in* a *hold mode.*
- *Both transistors M7 and M8 are off, decoupling the output* from the input. The output *Q retains its previous value stored in CL2.*
- *If CLK = 1 → The master stage is in hold mode (M3-M4 off), slave stage evaluates (M7-M8 on).*
- *The value stored on CL1* propagates to the output node through the slave stage .



www.rejinpaul.com

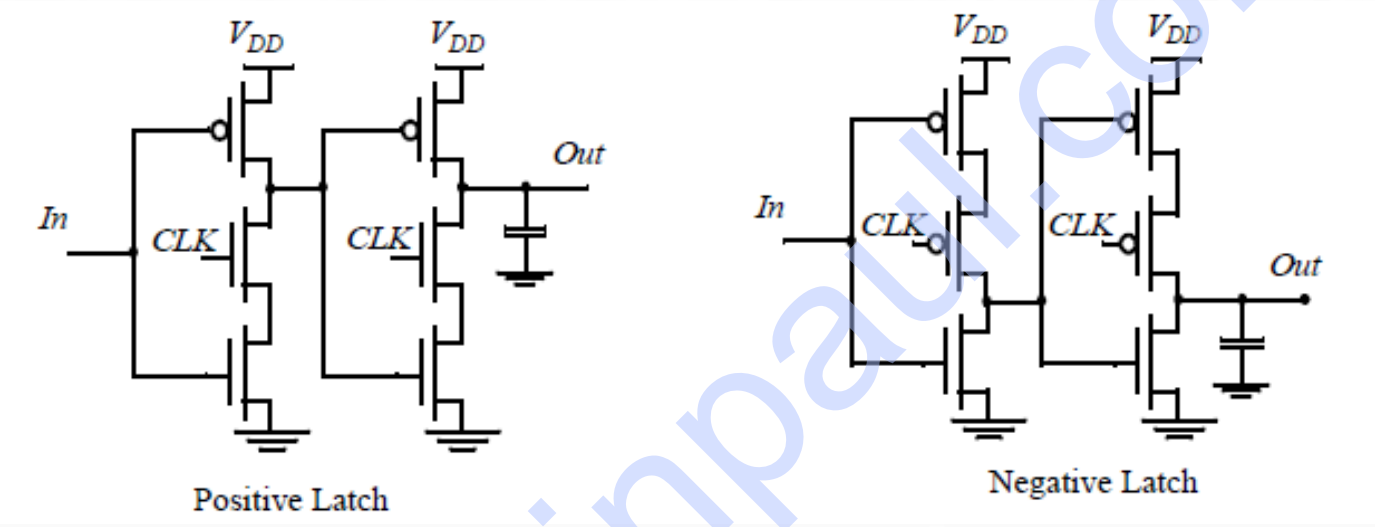
2.2) C²MOS based Dual-edge Registers

- This sequential circuit sampling the input on both clock edges.
- It consists of two parallel master slave based edge-triggered registers, whose outputs are multiplexed using the tri-state drivers.
- When clock is high, the positive latch transistors *M1-M4 is sampling* the inverted *D input on node X. Node Y is held stable, since devices M9 and M10 are turned off.*
- When clock is low, *the top slave latch M5-M8 on, and drives the* inverted value of *X to the Q output.*
- When clock is low, *transistors (M9, M10) is turned on, sampling the inverted D input on node Y.*
- *On the rising edge, the bottom* slave latch conducts, and drives the inverted version of *Y on node Q. Data hence changes* on both edges.



www.rejinpaul.com

3) True Single-Phase Clocked Register (TSPCR)



- It uses a single phase clock.
- For the positive latch, when *CLK is high, the latch is in the transparent mode.*
- The latch is non-inverting state , and propagates the input to the output.
- when *CLK* = 0, *both inverters are disabled, and the latch* is in *hold-mode.*
- *pull-up networks are activated, and the pull-down circuits are* deactivated.
- As a result , no signal can ever propagate from the input of the latch to the output in this mode.

www.rejinpaul.com

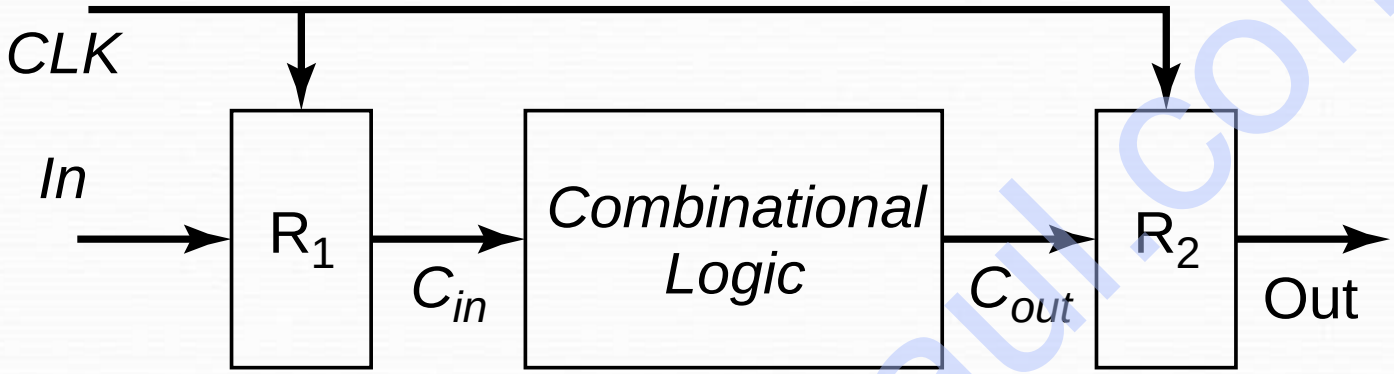
Timing Issues

- In *synchronous system approach, in which all memory elements in the system are* simultaneously updated using a globally distributed periodic synchronization signal is called global clock.
- The analysis of the impact of spatial variations of the clock signal, called *clock skew, and temporal variations* of the clock signal, called *clock jitter.*
- **Classification of Digital Systems**
- In digital systems, signals can be classified depending on a local clock.
- Signals that transition only at predetermined periods in time.
- It can be classified as follows

1. Synchronous
2. Mesochronous
3. Plesiochronous

www.rejinpaul.com

1)Synchronous Timing



•A synchronous signal has the same frequency, and fixed phase offset with respect to the local clock.

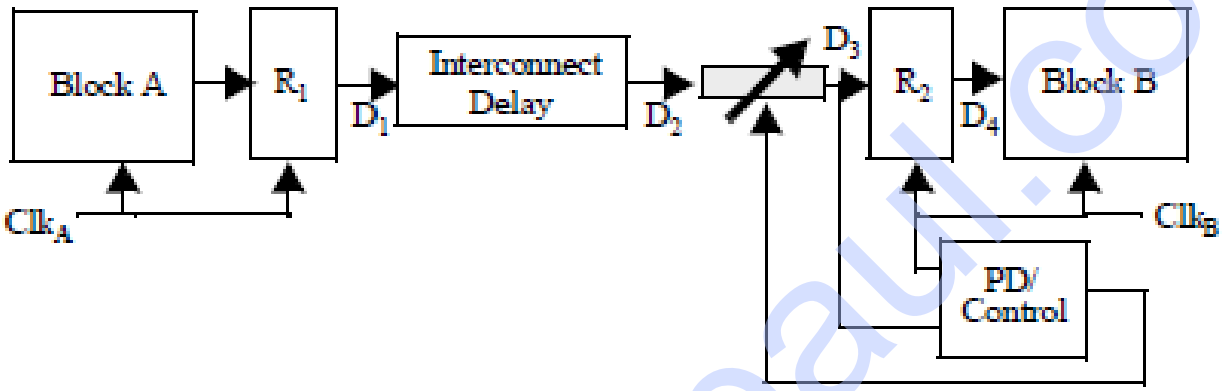
•The signal is “synchronized "with the clock, and the data can be sampled directly without any uncertainty.

•The input data signal In is sampled with register R1 to give signal C_{in}, which is synchronous with the system clock and then passed to the combinational logic block.

•After a setting period, the output C_{out} is obtained from R2 which synchronizes the output with the clock.

www.rejinpaul.com

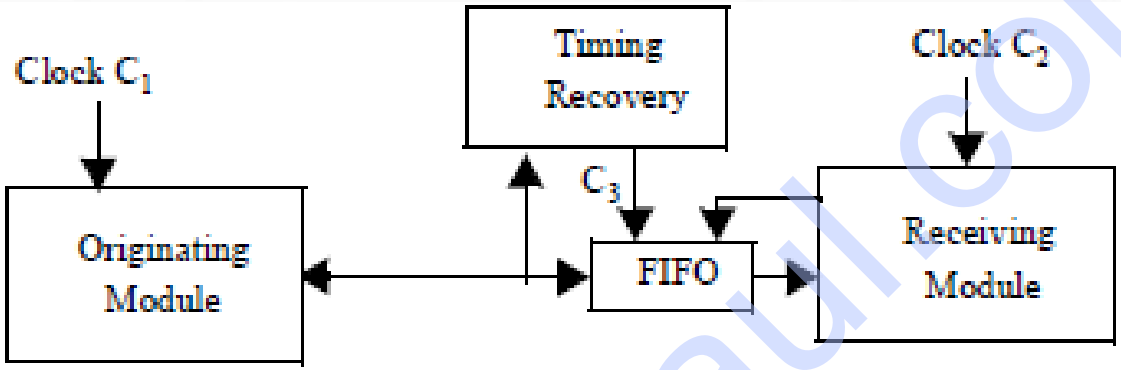
2)Mesochronous interconnect



•A mesochronous signal is one that has the same frequency but an unknown /different phase offset with respect to the local clock.
•In such a system, it is not possible to directly sample the output at the receiving module because of the uncertainty in the phase offset.
•The signal D1 is synchronous with respect to ClkA.
•Data D1 and D2 are middle with ClkB because of the unknown phase difference between Clk_A and Clk_B
•The synchronizer is to adjust the variable delay line such that the data signal D3 is aligned properly with the system clock of block B.
•Register R2 samples the incoming data during the certainty period, then signal D4 becomes synchronous with ClkB.

www.rejinpaul.com

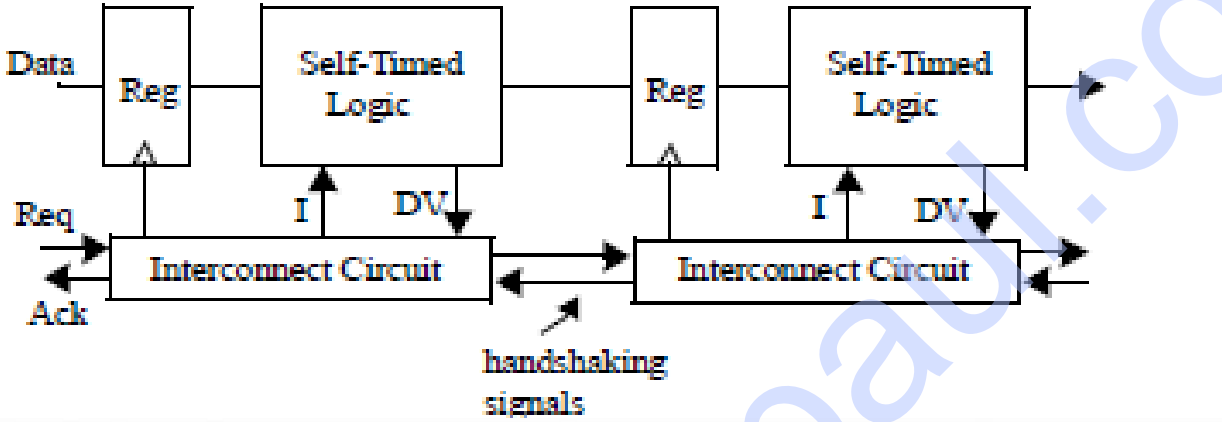
3)Plesiochronous Interconnect



- A plesio-chronous signal has slightly different frequency as the local clock,and also the phase difference drifts in time.
- The transmitted signal can arrive at the receiving module at a different rate than the local clock, so buffering scheme to ensure all data is received.
- In this digital communications ,the originating module issues data at some unknown rate by C1, which is plesiochronous with respect to C2.
- The timing recovery unit is responsible for deriving clock C3 from the data sequence, and buffering the data in a FIFO.
- As a result, C3 will be synchronous with the data at the input of the FIFO and will be meso-chronous with C1.
- If the clock frequencies mismatched between Tx/Rx, data might have to be dropped if the transmit frequency is faster.
- Data can be duplicated if the Tx frequency is slower than the receive frequency.

www.rejinpaul.com

4)Asynchronous Interconnect



•Asynchronous signals can transition at any arbitrary time.(not depends on clock)
•Communication between modules is controlled through a handshaking protocol .
•When a logic block completes an operation, it will generate a signal DV to indicate that output data is valid.
•The handshaking signals then initiate a data transfer to the next block.
•Then latches in the new data and begins a new computation by asserting the initialization signal *I*.

•*Advantages*

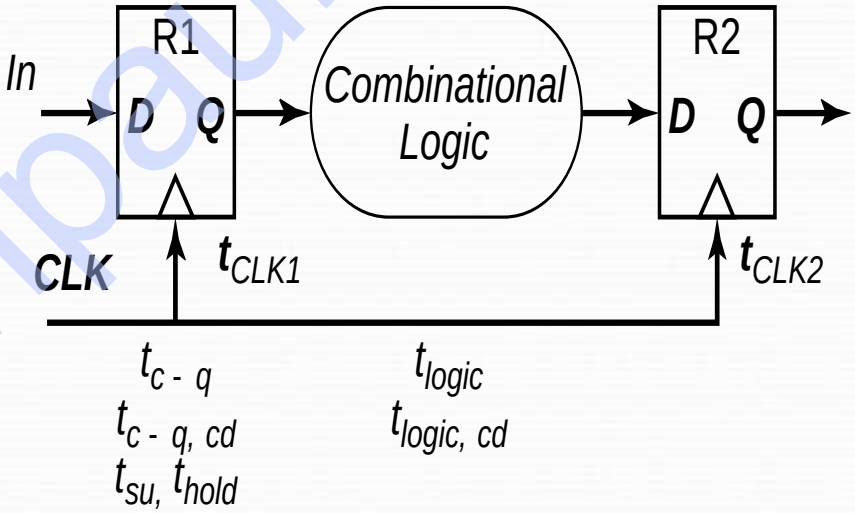
1. No need to manage clock *skew*
2. Computations are performed at the native speed of the logic
3. Block computations occur whenever data becomes available

www.rejinpaul.com

Synchronous Timing Basics

•The clock at registers 1 and 2 have the same clock period and transition at the exact same time. The following timing parameters characterize the timing of the sequential circuit.

$T_{c\text{-}q,cd}$	Contamination or minimum delay of registers
t_{cq}	Maximum propagation of registers
t_{su}	Register Set up time
t_{hold}	Register Hold time
$t_{logic,Cd}$	Contamination or minimum delay of combinational logic
t_{logic}	Maximum propagation of combinational logic



Minimum cycle time:

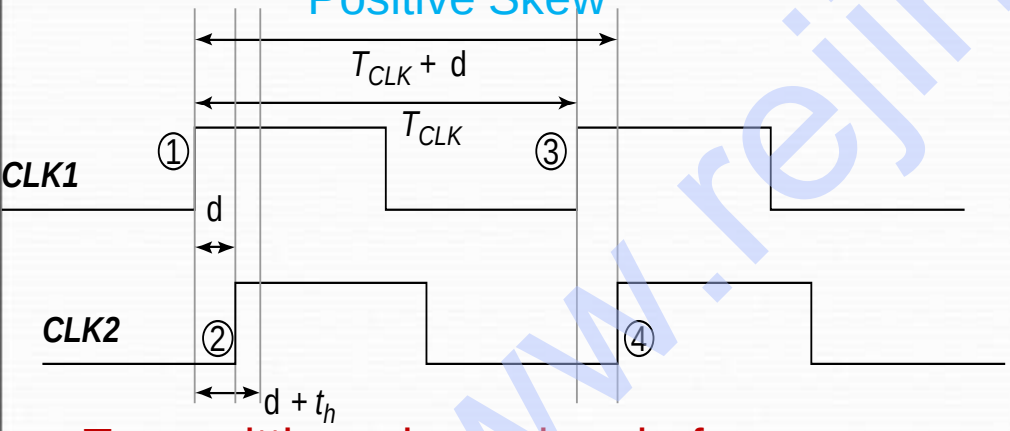
$T - \delta = t_{c\text{-}q} + t_{su} + t_{logic}$

www.rejinpaul.com

Clock Skew

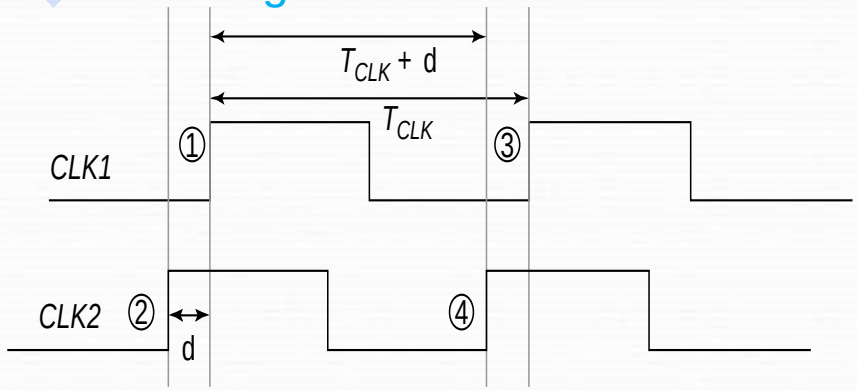
- The spatial variation in arrival time of a clock transition on an integrated circuit is commonly referred to as *clock skew*.
- The clock skew between two points i and j on a IC is given by $\delta_{(i,j)} = t_i - t_j$
- where t_i and t_j are the position of the rising edge of the clock with respect to a reference.
- The clock skew can be positive or negative depending upon the routing direction and position of the clock source.

Positive Skew



Transmitting edge arrives before the receiving edge

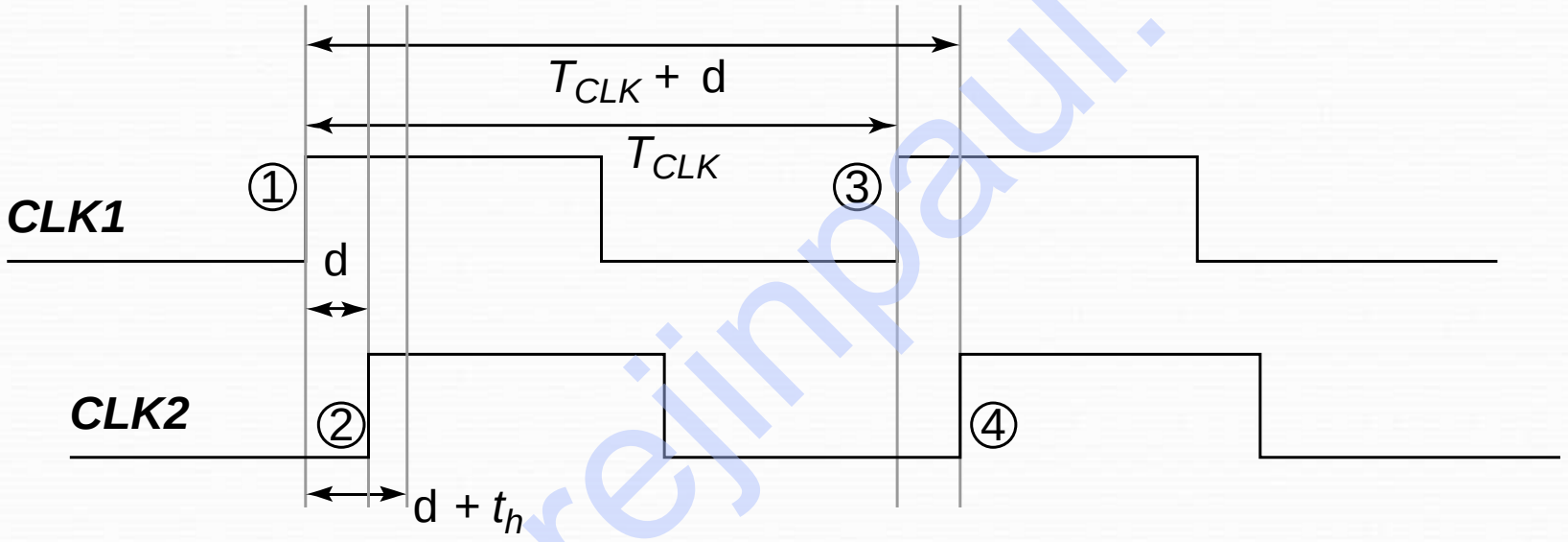
Negative Skew



Receiving edge arrives before the launching edge

www.rejinpaul.com

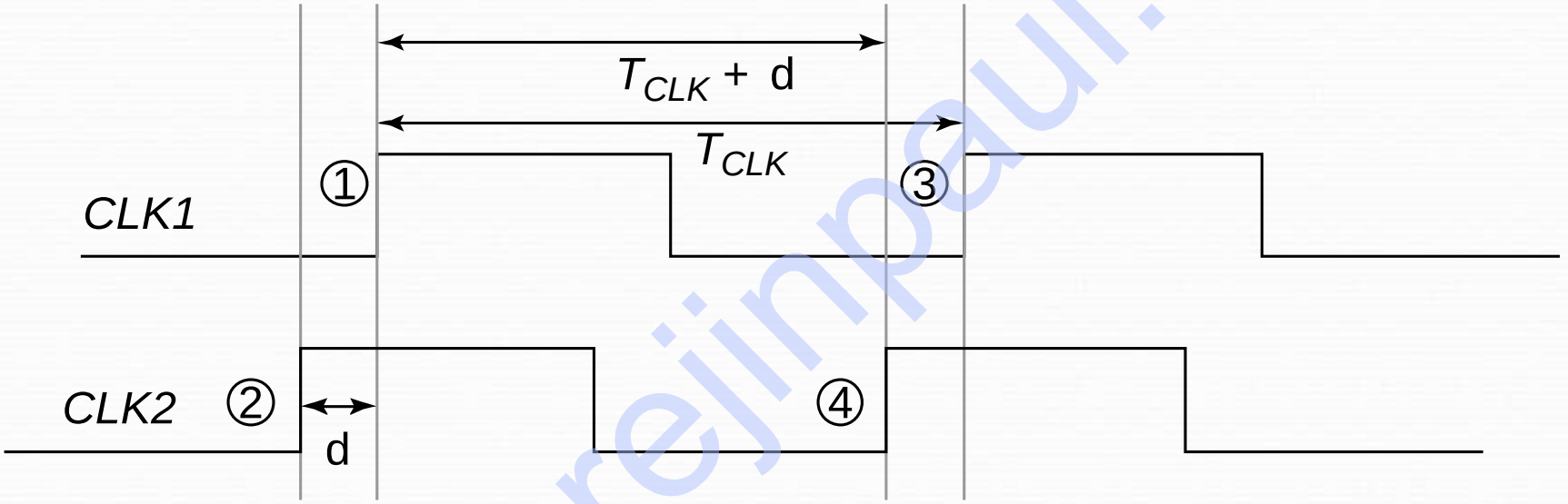
Positive Skew



Launching edge arrives before the receiving edge

www.rejinpaul.com

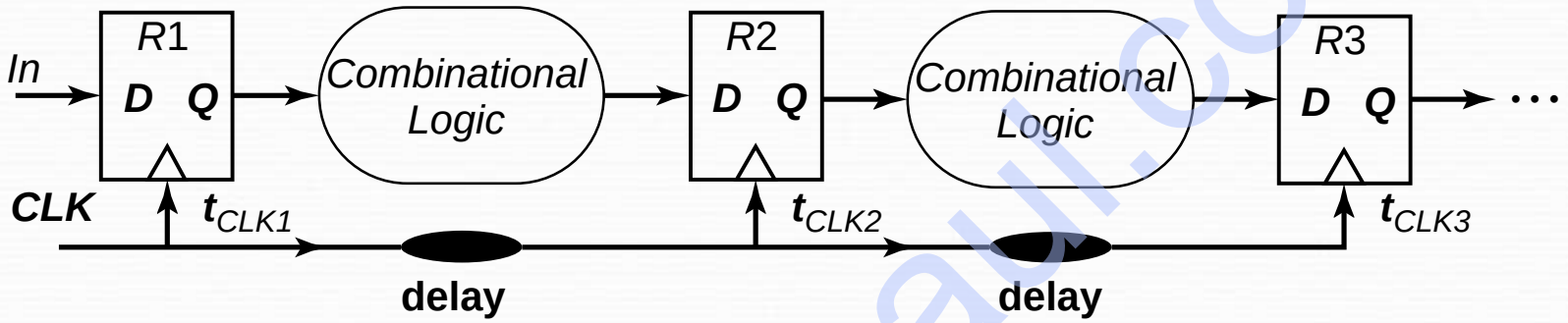
Negative Skew



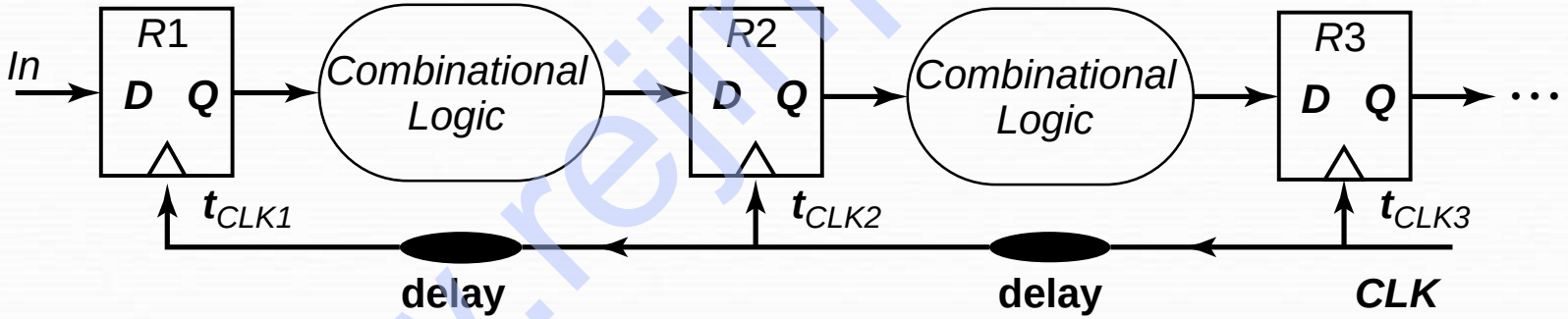
Receiving edge arrives before the launching edge

www.rejinpaul.com

Positive and Negative Skew



(a) Positive skew



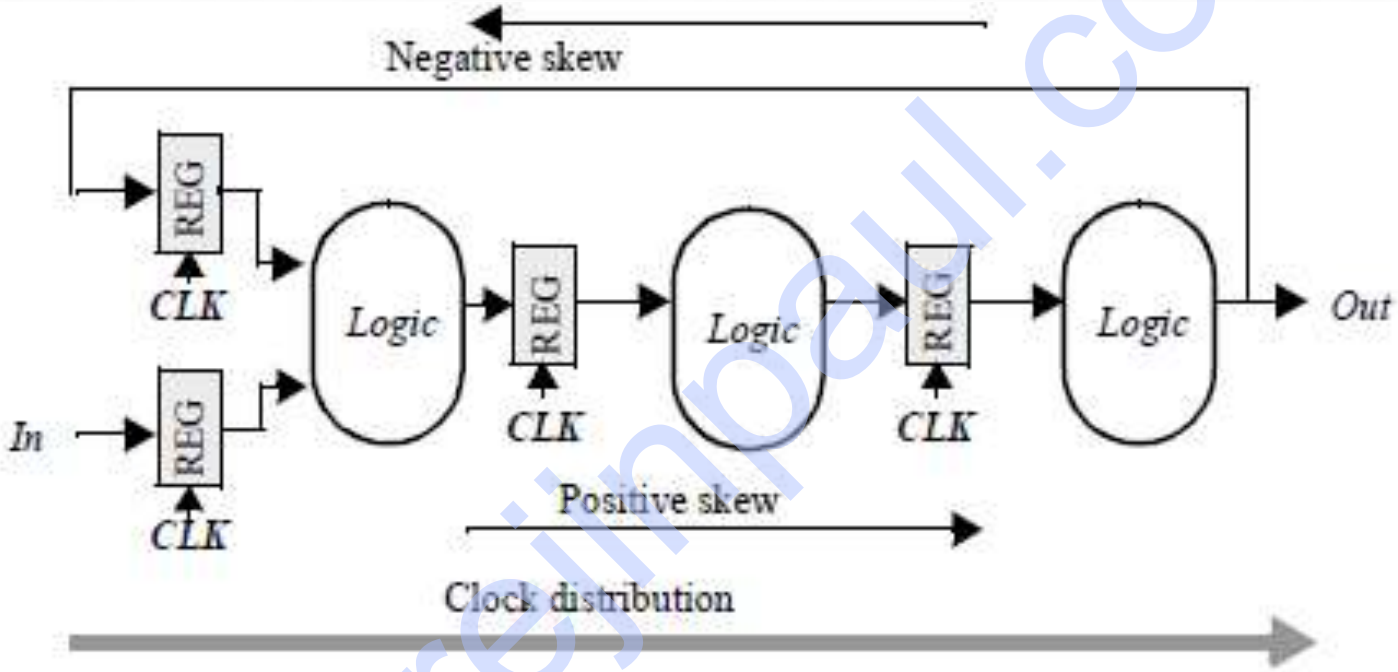
(b) Negative skew

$\delta > 0$ → This corresponds to a clock routed in the same direction as the flow of the data through the pipeline.

$\delta < 0$ → When the clock is routed in the opposite direction of the data , the skew is negative and condition.

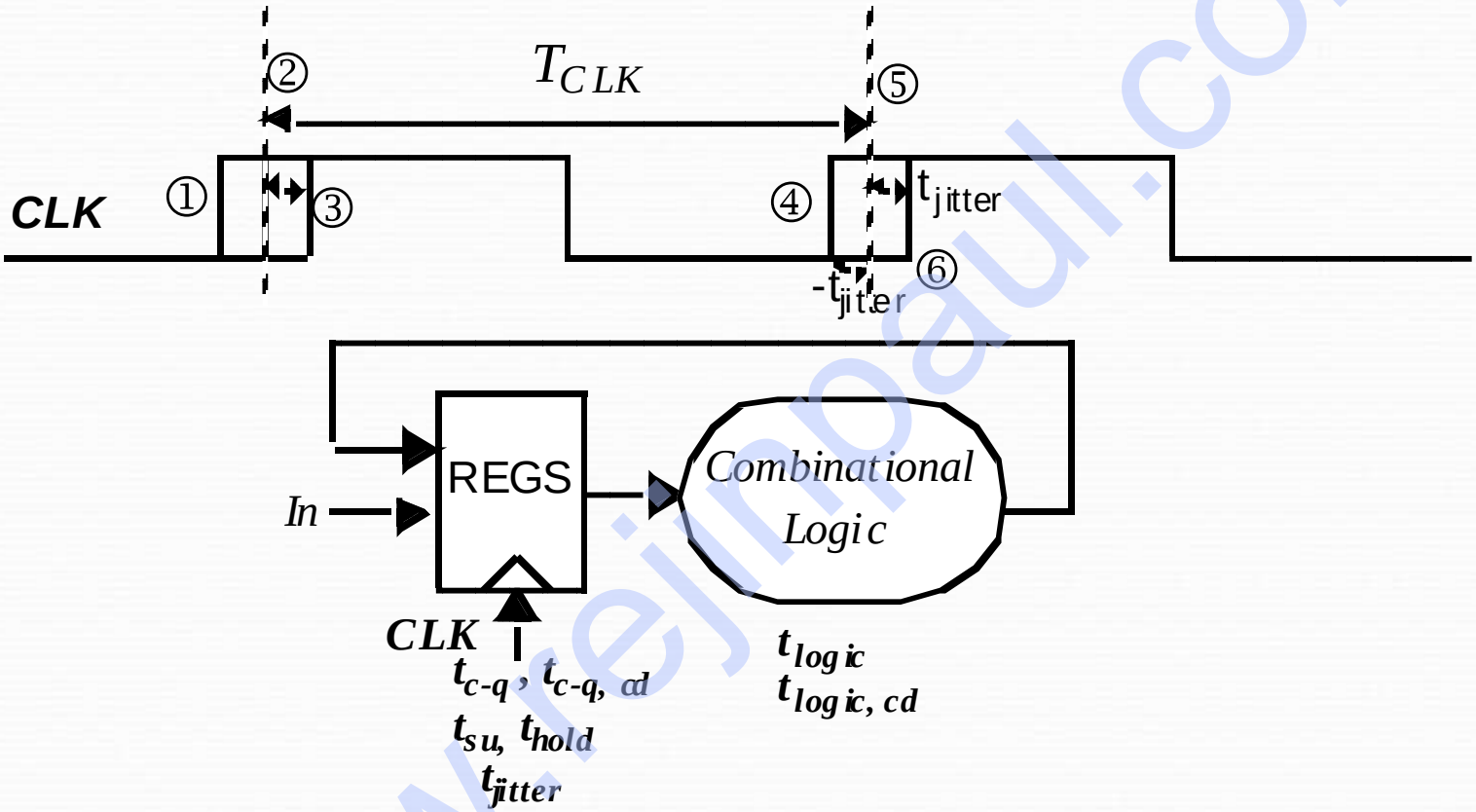
www.reinpaul.com

How to counter Clock Skew?



www.rejinpaul.com

Impact of Jitter

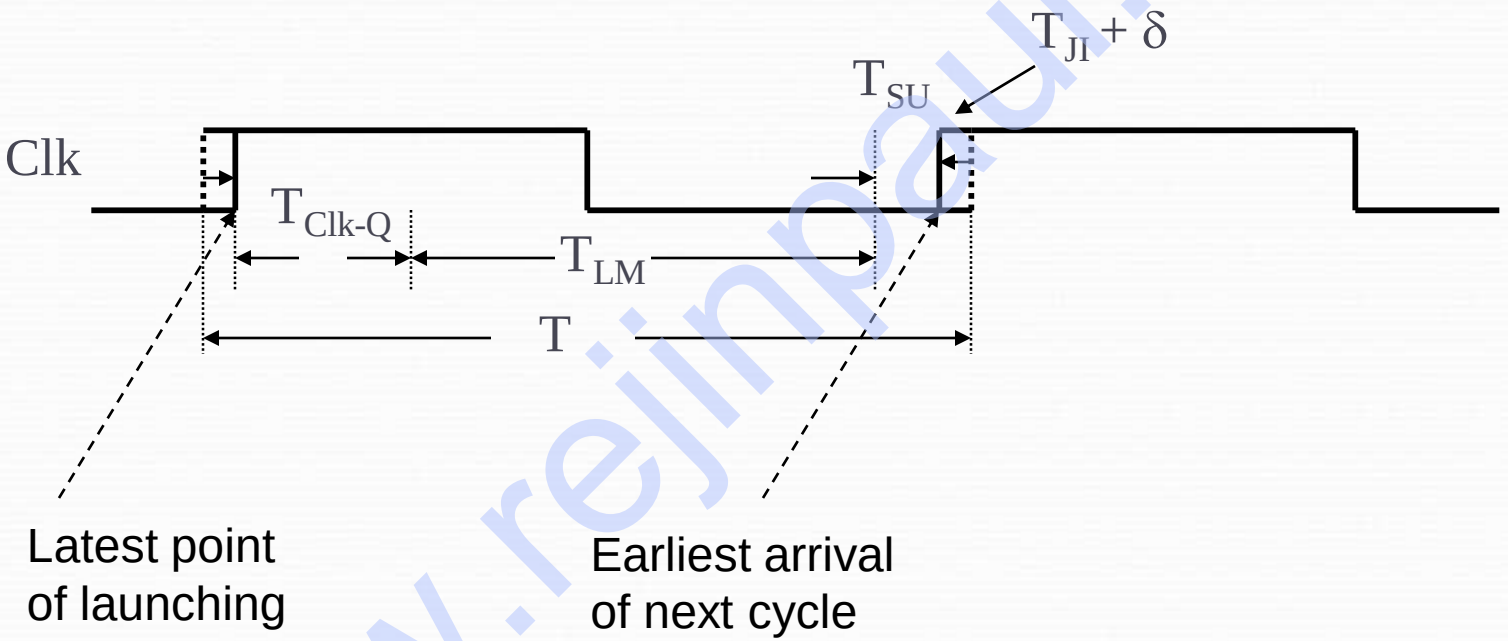


•Clock jitter refers to the temporal variation of the clock period at a given point.

•The clock period can reduce or expand on a cycle-by-cycle basis

www.rejinpaul.com

Longest Logic Path in Edge-Triggered Systems



www.rejinpaul.com

Clock Constraints in Edge-Triggered Systems

If launching edge is late and receiving edge is early, the data will not be too late if:

$$T_{c\text{-}q} + T_{LM} + T_{SU} < T - T_{JI,1} - T_{JI,2} - \delta$$

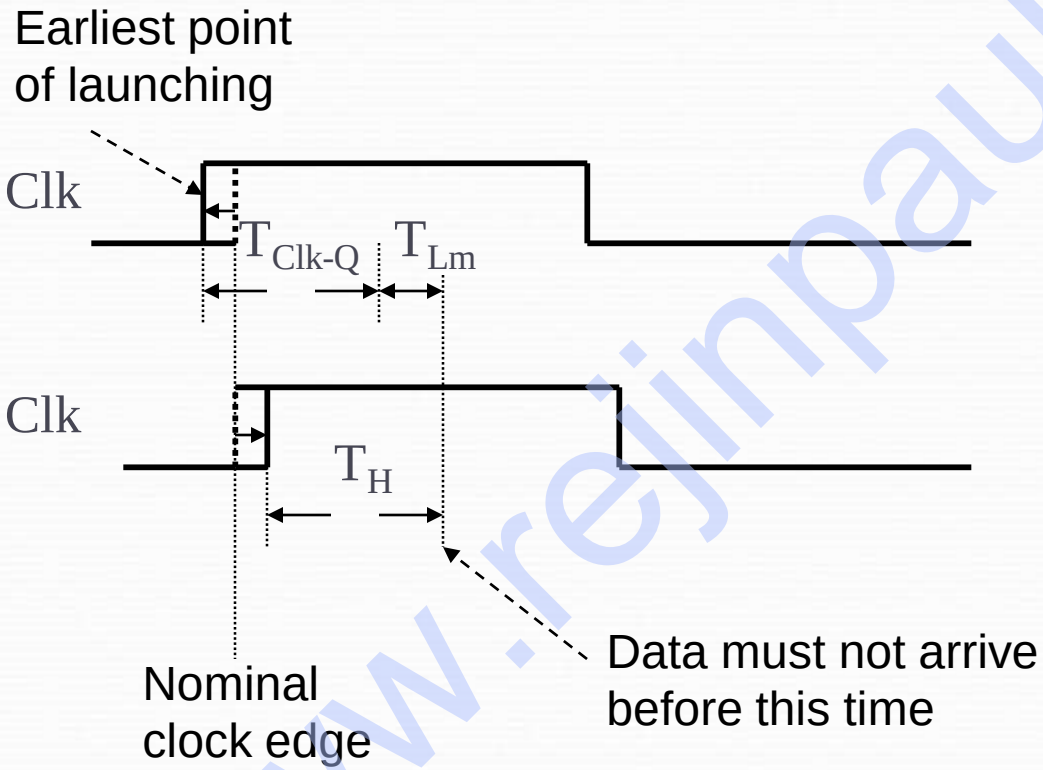
Minimum cycle time is determined by the maximum delays through the logic

$$T_{c\text{-}q} + T_{LM} + T_{SU} + \delta + 2\,T_{JI} < T$$

Skew can be either positive or negative

www.rejinpaul.com

Shortest Path



www.rejinpaul.com

Clock Constraints in Edge-Triggered Systems

If launching edge is early and receiving edge is late:

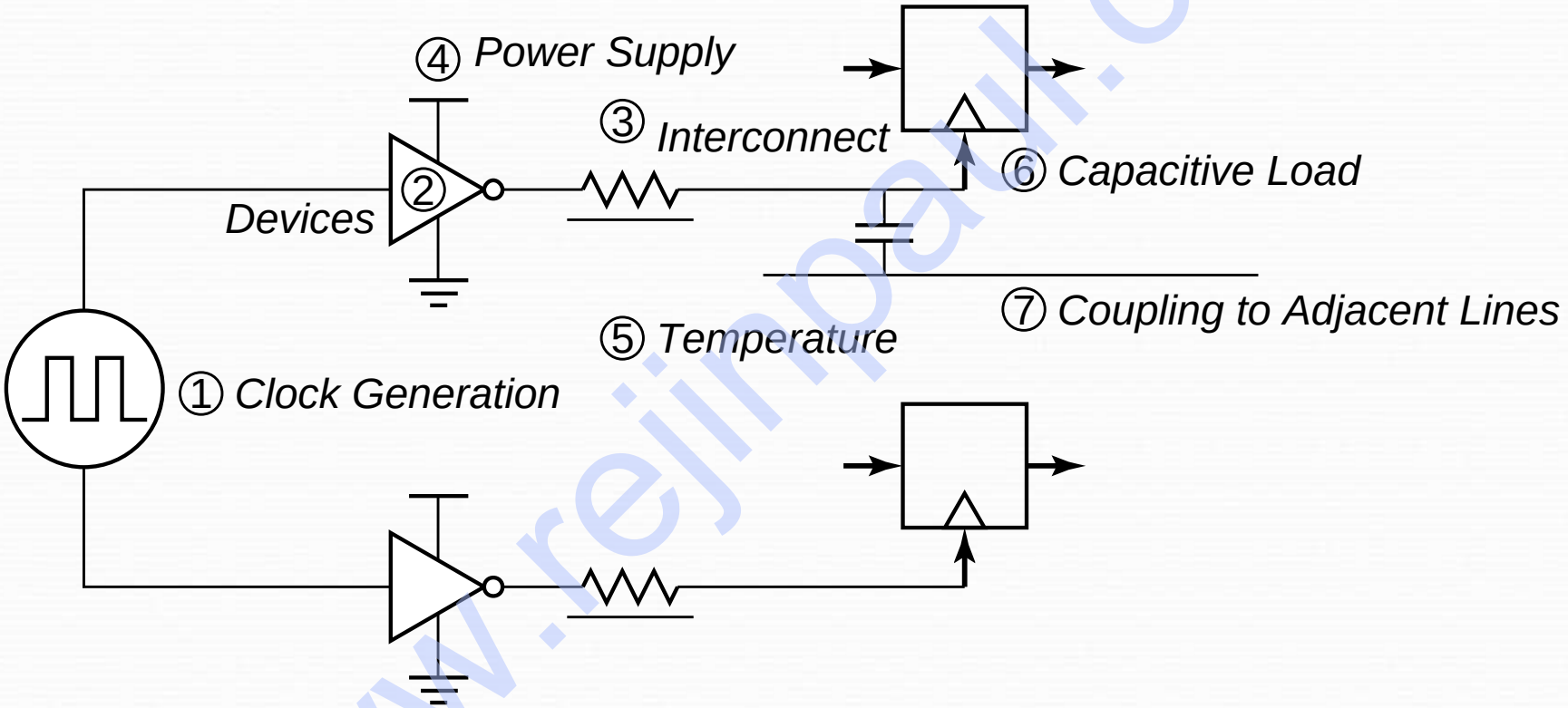
$$T_{c\text{-}q} + T_{LM} - T_{JI,1} < T_H + T_{JI,2} + \delta$$

Minimum logic delay

$$T_{c\text{-}q} + T_{LM} < T_H + 2T_{JI} + \delta$$

www.rejinpaul.com

Sources of Skew and Jitter



www.rejinpaul.com

Clock-Signal Generation→Voltage-Controlled Oscillator is used for clock generation.

- This is an analog circuit, sensitive to intrinsic device noise and power supply variations.

Manufacturing Device Variations→matching of devices in the buffers along multiple clock paths is critical to minimizing timing uncertainty.

- Because of process variations, devices parameters in the buffers vary along different paths, resulting in static skew.

Interconnect Variations→Vertical and lateral dimension variations cause the interconnect capacitance and resistance to vary across a chip.

•variation is static, it causes skew between different paths.

•Other interconnect variations include deviation in the width of the wires and line spacing.

Environmental Variations→The two major sources of environmental variations are temperature and power supply.

•**Temperature** gradients across the chip is a result of variations in power dissipation across the die.

•**Power supply variations** on the hand is the major source of jitter in clock distribution Networks. Static power supply variations may result from fixed currents drawn from various modules

www.reinpaul.com

Capacitive Coupling→coupling between the clock lines and adjacent signal wires and variation in gate capacitance.

•The clock network includes both the interconnect and the gate capacitance of latches and registers.

•Any coupling between the clock wire and adjacent signal results in timing uncertainty.

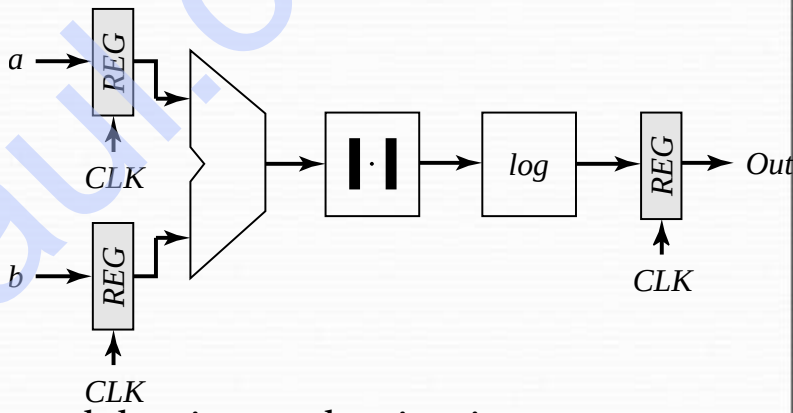
•The adjacent signal can transition in arbitrary directions and at arbitrary times, the exactly coupling to the clock network is not fixed from cycle-to-cycle. This results in clock jitter.

www.rejinpaul.com

Pipelining

- Pipelining is design technique used to accelerate the operation of the datapaths in digital processors.

Example:1 [Reference circuit compute log(|*a* + *b*|)]



- where both *a and b represent* streams of numbers.
- registers are edge-triggered *D registers.*
- The minimal clock period T_{min} *is*

$$T_{min} = t_{pcq} + t_{pd} + t_{setup}$$

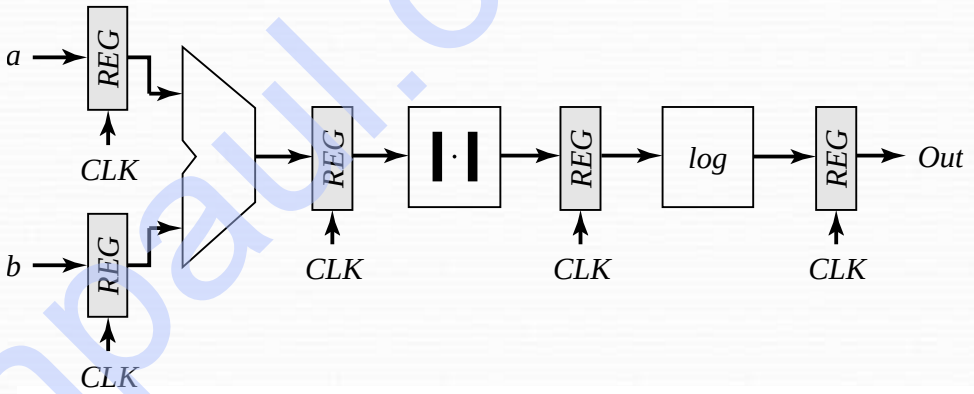
- The delay is generally larger than the delays in registers and dominates the circuit performance.
- If each logic module has an equal propagation delay,then each logic module is then active for only 1/3 of the clock period (if the delay of the register is ignored).
- For example, the adder unit is active during the first third of the period .
- Remains idle—during other 2/3 of the period.

www.rejinpaul.com

Pipelined circuit compute log(|a + b|)]

- Pipelining is a technique to improve the resource utilization, and increase the functional throughput.
- Use registers between the logic blocks.
- The computation for one set of input data to spread over a number of clock periods.
- The result for the data set (a1, b1) only appears at the output after three clock-periods.
- At that time, the circuit has already performed parts of the computations for the next datasets, (a2, b2) and (a3,b3).
- The advantage of pipelined operation is need minimum clock period.
- The combinational circuit block has been partitioned into three sections, each of which has a smaller propagation delay than the original function.
- The minimum allowable clock period is

$T_{min,pipe} = t_{c\text{-}q} + \max(t_{pd,add}, t_{pd,abs}, t_{pd,log})$



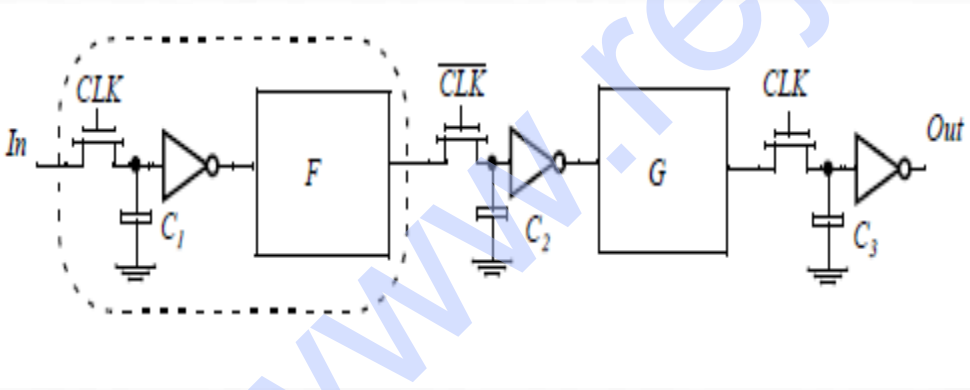
Clock Period	Adder	Absolute Value	Logarithm
1	$a_1 + b_1$		
2	$a_2 + b_2$	$\|a_1 + b_1\|$	
3	$a_3 + b_3$	$\|a_2 + b_2\|$	$\log(\|a_1 + b_1\|)$
4	$a_4 + b_4$	$\|a_3 + b_3\|$	$\log(\|a_2 + b_2\|)$
5	$a_5 + b_5$	$\|a_4 + b_4\|$	$\log(\|a_3 + b_3\|)$

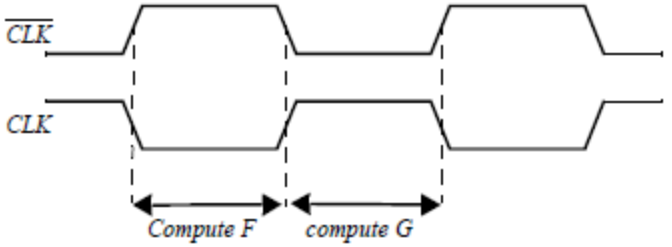
Pipelined Computation

www.rejinpaul.com

Latch- Vs Register-Based Pipelines

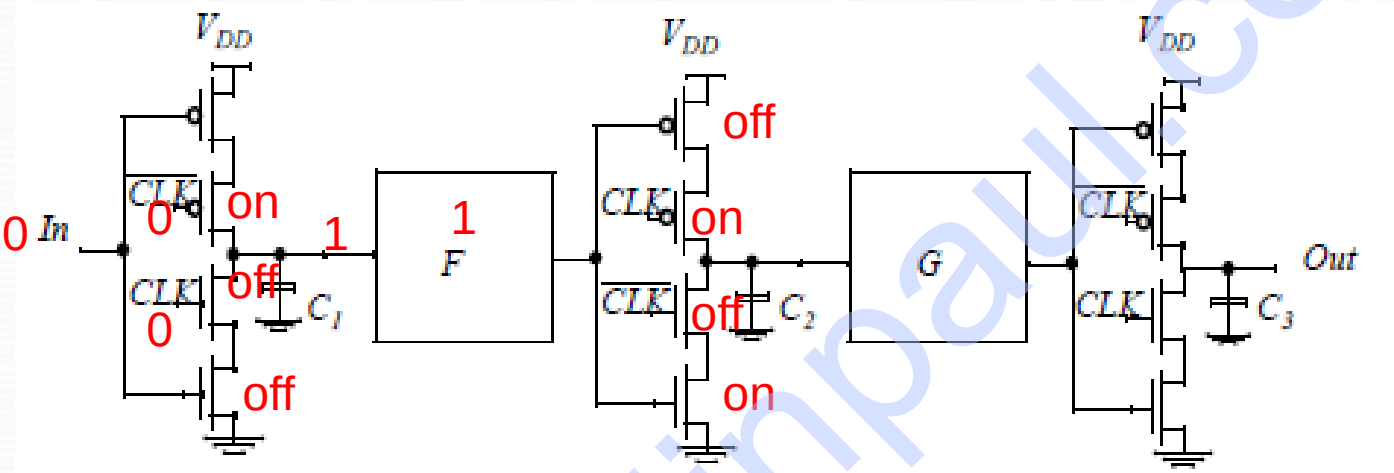
- The pipeline system is implemented based on pass-transistor-based *positive and negative latches instead of edge triggered* registers.
- Logic is introduced between the master and slave latches of a *master-slave system.*
- When the clocks *CLK and* $\overline{CLK}$ *are non-overlapping, correct pipeline operation is* obtained.
- Input data is sampled on *C1 at the Positive edge of CLK and the computation of* logic block *F starts and is stored on C2on the falling edge of CLK,Computation of logic block G starts.*
- The value stored on *C2 at the end of the CLK low phase is the* result of passing the previous input *through the* logic function *F.*
- *When overlap exists between CLK and* $\overline{CLK}$*, the next input is already* being applied to *F, and its effect might propagate to C2 before* $\overline{CLK}$ *goes low*





www.rejinpaul.com

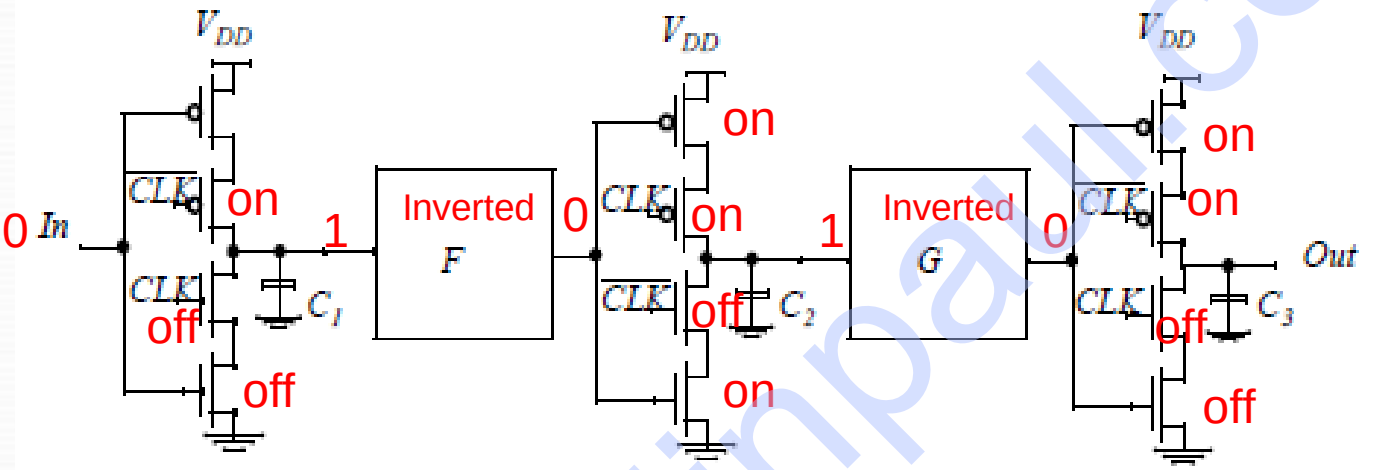
NORA-CMOS Based Pipelined Structures



- A C²MOS-based pipelined circuit is race-free as long as all the logic functions *F* between the latches are non-inverting.
- During a (o-o) overlap between CLK and $\overline{\text{CLK}}$, all C²MOS latches, simplify to pure pull-up networks.
- The only way a signal can race from stage to stage under this condition is when the logic function *F is inverting*.

www.rejinpaul.com

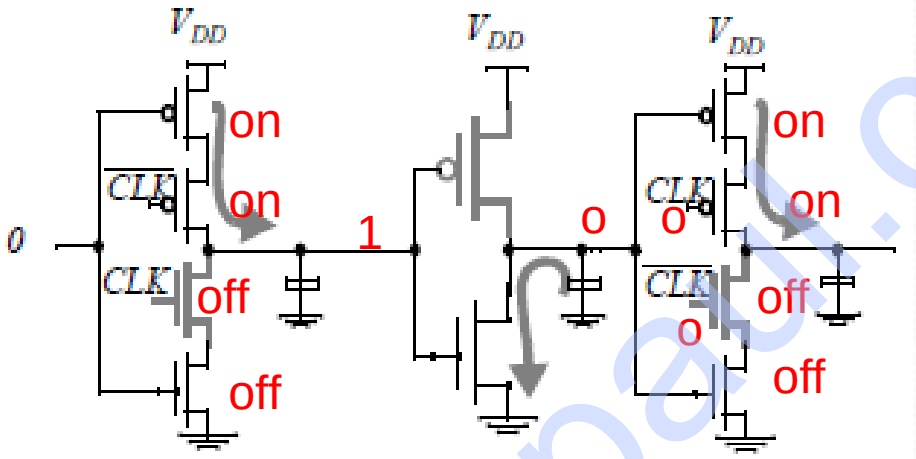
NORA-CMOS Based Pipelined Structures



- A C²MOS-based pipelined circuit is race-free as long as all the logic functions *F* between the latches are non-inverting.
- During a (o-o) overlap between CLK and $\overline{\text{CLK}}$, all C²MOS latches, simplify to pure pull-up networks.
- The only way a signal can race from stage to stage under this condition is when the logic function *F is inverting*.

www.rejinpaul.com

Potential race condition during (0-0) overlap in C²MOS-based design

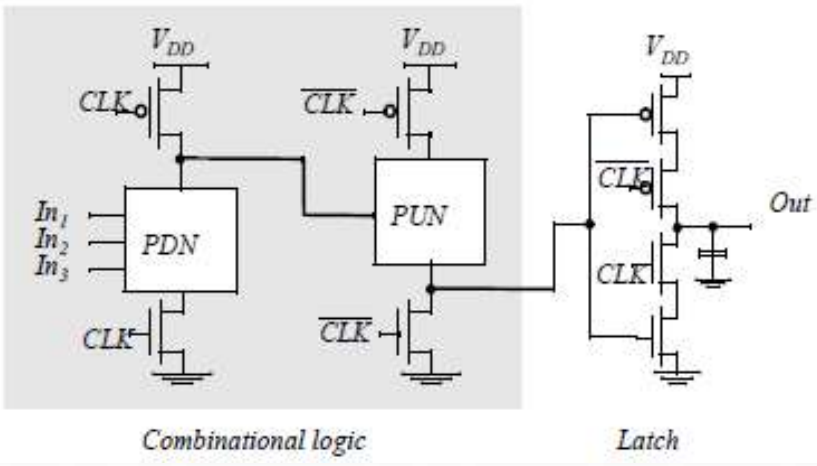


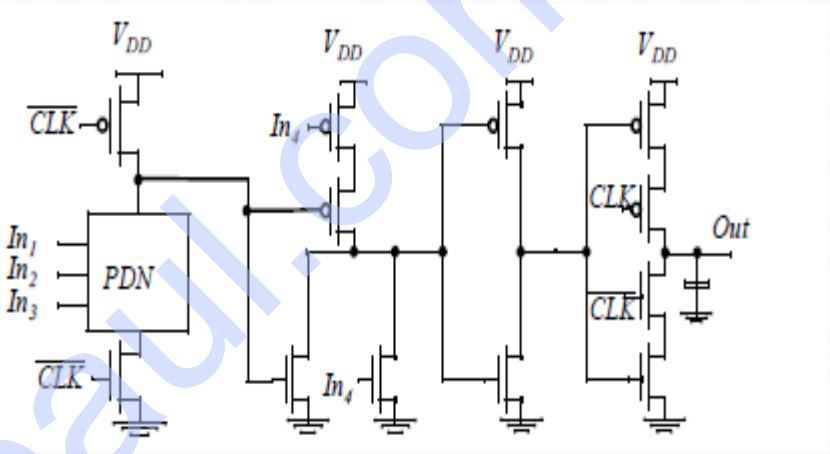
- where F is replaced by a single, static CMOS inverter. Similar considerations are valid for the (1-1) overlap.
- It combines C²MOS pipeline registers and *NORA dynamic logic function* blocks.
- Each module consists of a block of combinational logic that can be a mixture of static and dynamic logic, followed by a C²MOS latch.
- Logic and latch are clocked in such a way that both are simultaneously in either evaluation, or hold (pre-charge) mode.
- A block that is in evaluation during *CLK = 1 is called a CLK-module, while the inverse is called a $\overline{CLK}$-module.*

www.rejinpaul.com

CLK-module

$\overline{CLK}$-module



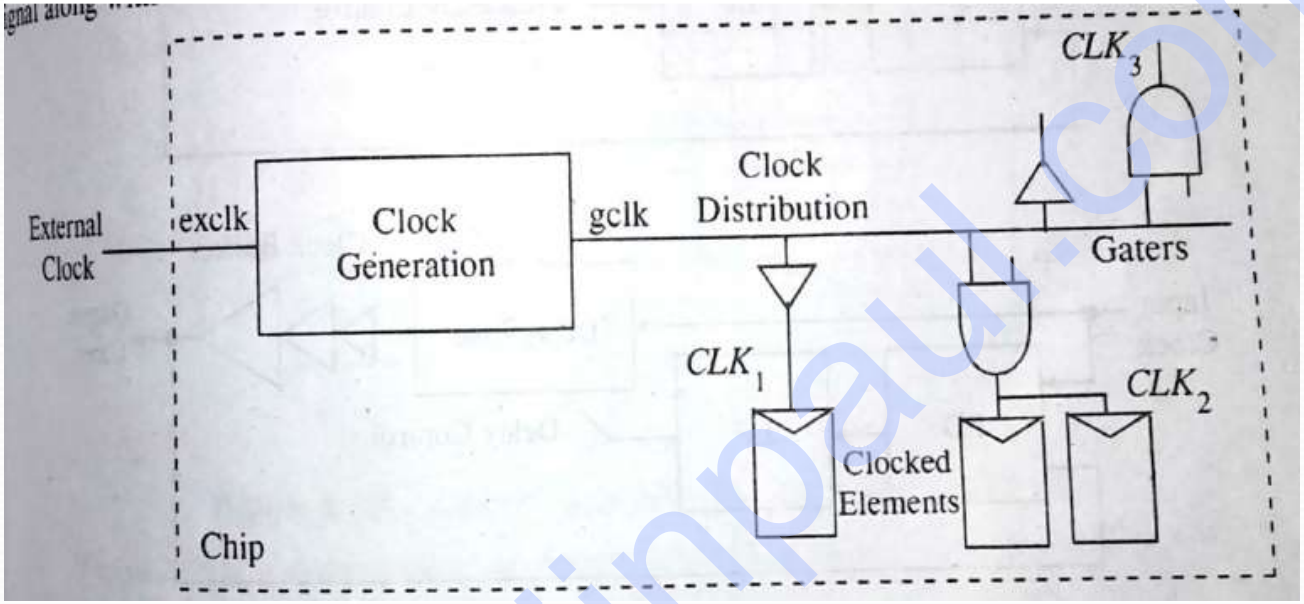


- A NORA datapath consists of a chain of alternating *CLK and* $\overline{CLK}$ *modules. While* one class of modules is pre-charging with its output latch in hold mode, preserving the previous output value, the other class is evaluating.
- Data is passed in a pipelined fashion from module to module.

	CLK block		$\overline{CLK}$ block	
	Logic	Latch	Logic	Latch
CLK = 0	Precharge	Hold	Evaluate	Evaluate
CLK = 1	Evaluate	Evaluate	Precharge	Hold

www.reinpaul.com

Clock System Architecture

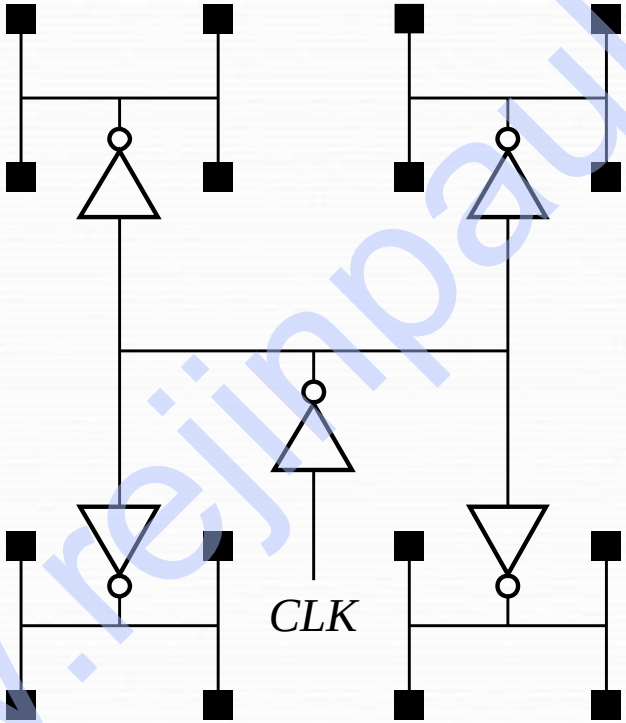


- Choosing the right clocking scheme affects the functionality, speed and power of a circuit.
- The external clock is applied through the I/O pads.
- Clock generation unit consists of either a PLL or delay Locked Loop to adjust the frequency or phase of the global clock.
- The global clock is distributed across all the clocked elements.
- The clock distribution network is designed with min clock skew.
- Local clock gates rx the global clock and derive the physical clock signal along with short wires to small groups of clocked elements

www.rejinpaul.com

Clock Distribution

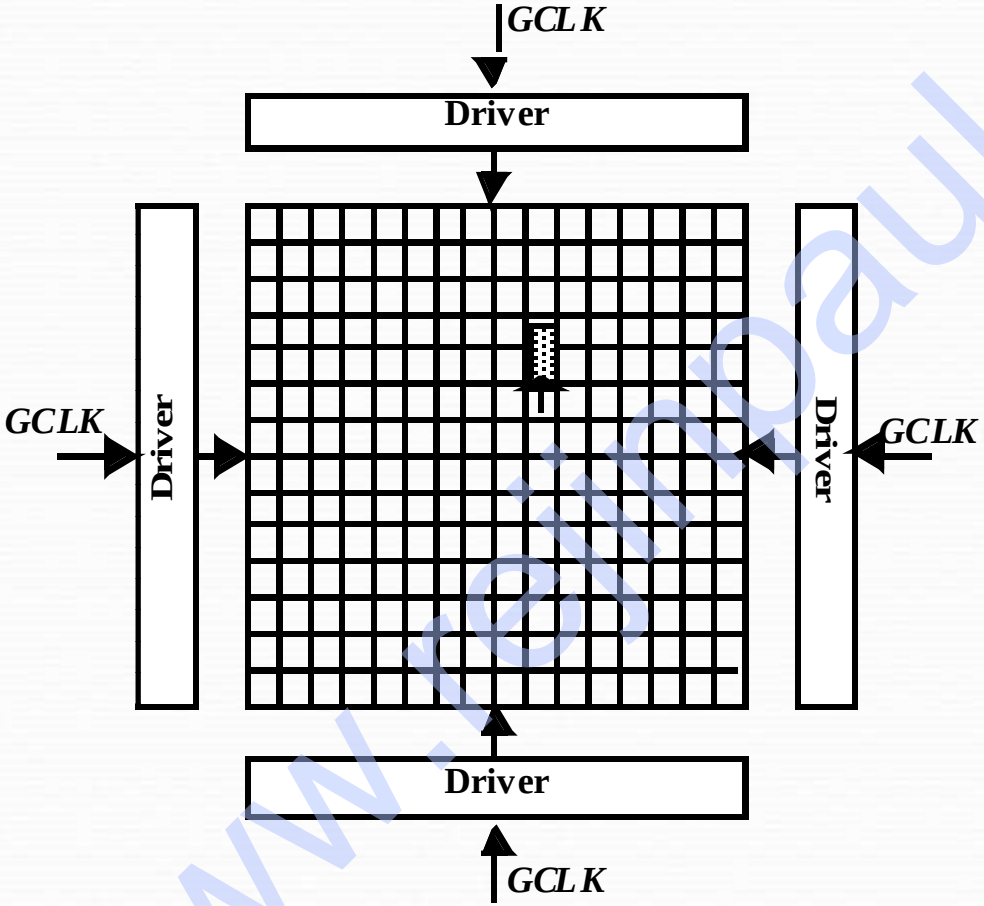
H-tree



Clock is distributed in a tree-like fashion

www.rejinpaul.com

The Grid System



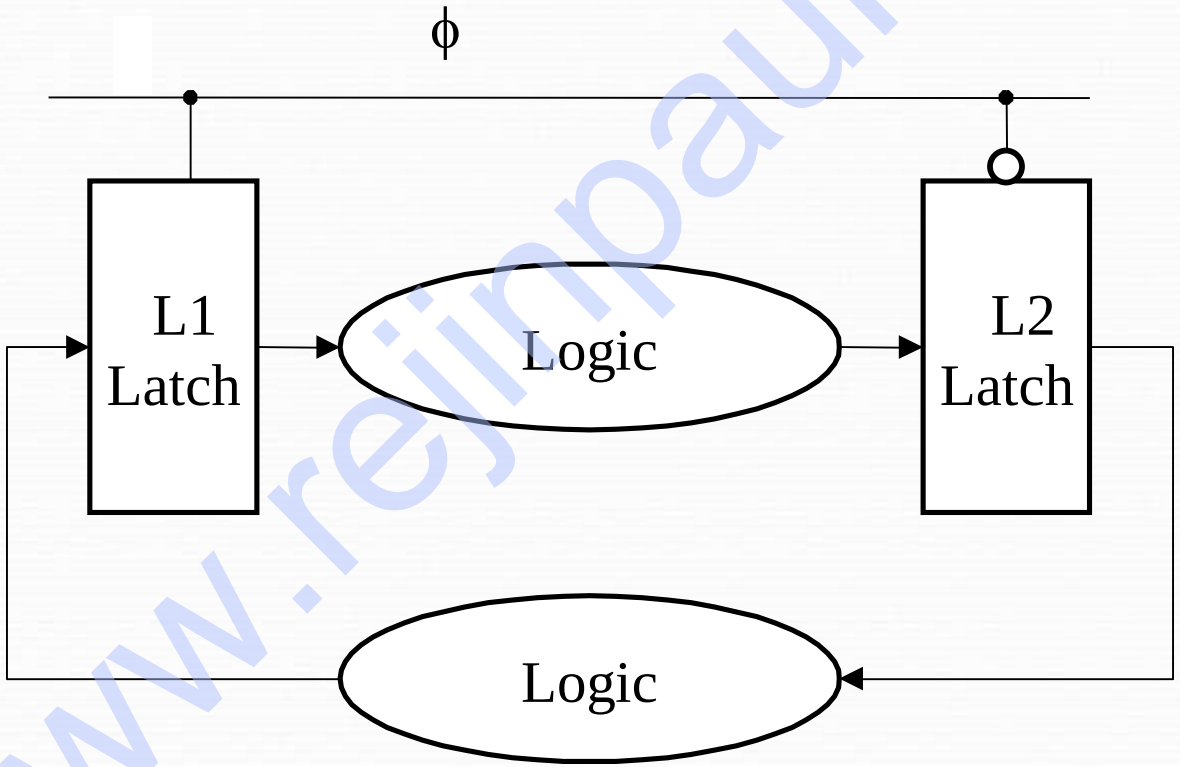
- No rc-matching
- Large power

www.rajnpaul.com

Latch-Based Design

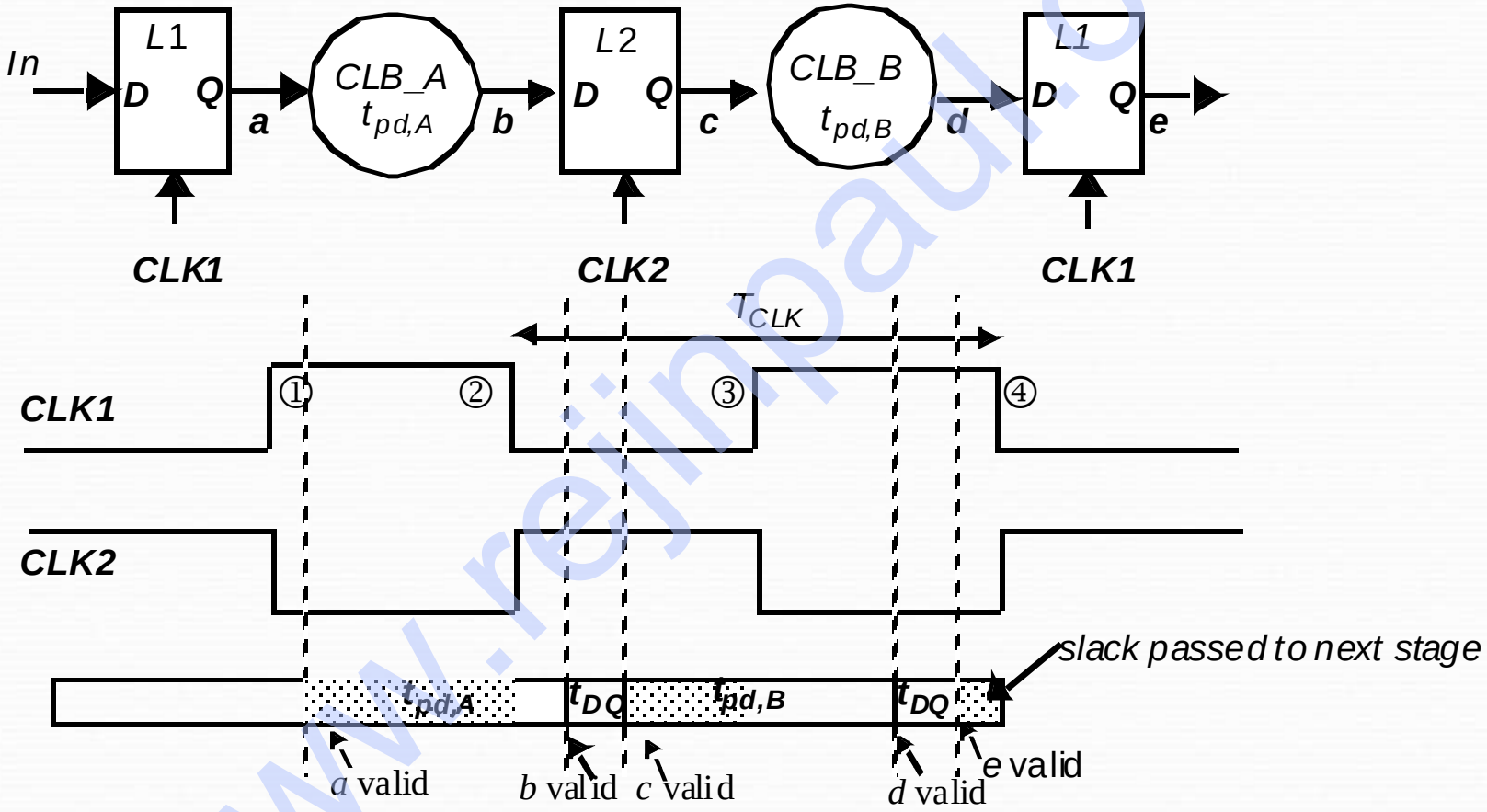
L1 latch is transparent when $\phi = 0$

L2 latch is transparent when $\phi = 1$



www.rejinpaul.com

Slack-borrowing



www.rejinpaul.com

Clock strategies

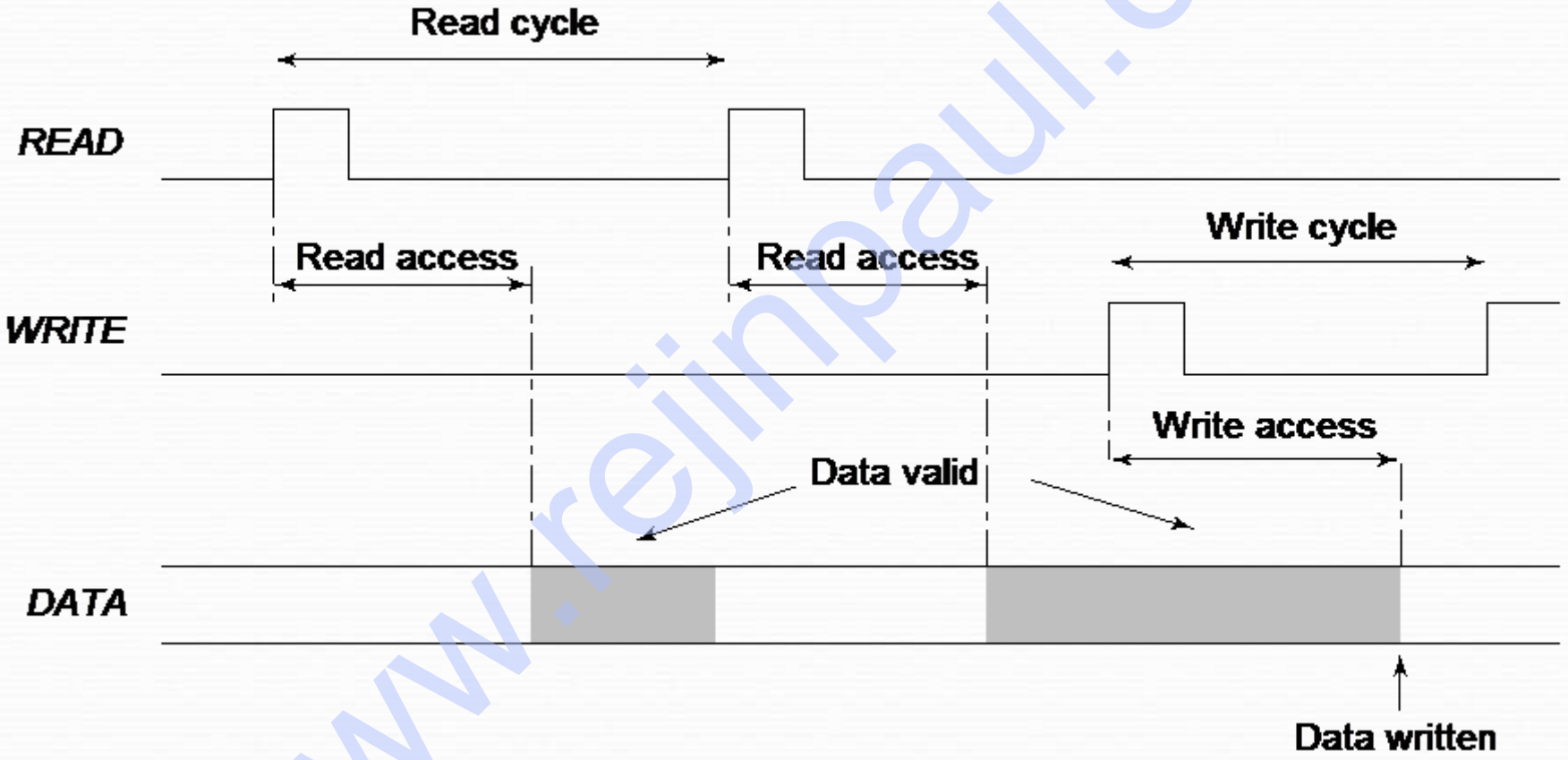
1. Pseudo 2-phase clocking
2. Pseudo 2-phase Memory structure
3. Pseudo 2-phase logic structure
4. 2-phase clocking
5. 2-phase Memory structure
6. 2-phase logic structure
7. 4-phase clocking
8. 4-phase Memory structure
9. 4-phase logic structure
10. Pseudo4-clocking

www.rejinpaul.com

1)Pseudo 2-phase clocking

www.rejinpaul.com

Memory Timing: Definitions



www.rejinpaul.com

Memory Architectures

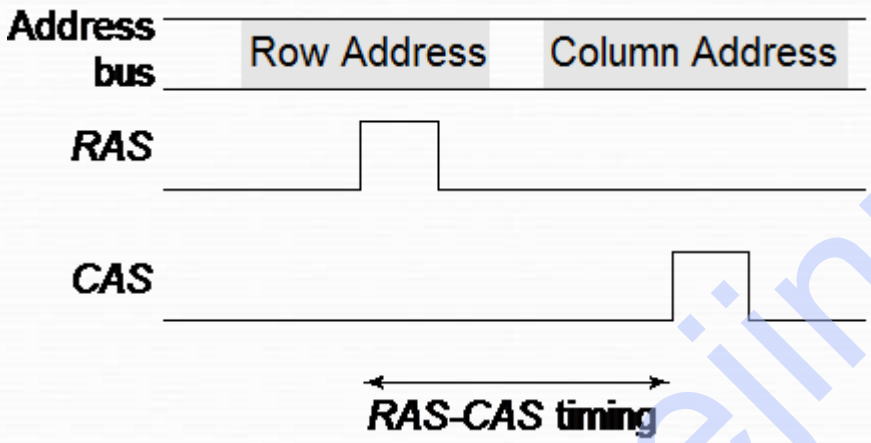
•Large portion of silicon area is dedicated to the storage of data and program instructions.

•Memory is classified as follows

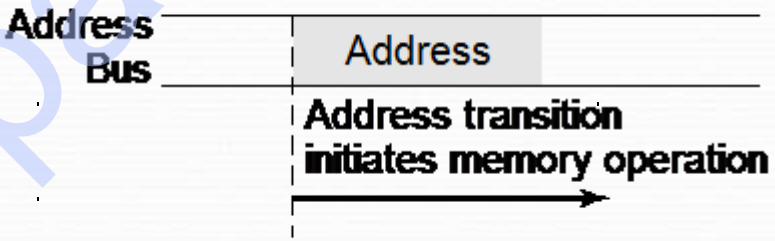
Read-Write Memory		Non-Volatile Read-Write Memory	Read-Only Memory
Random Access	Non-Random Access	EPROM E^2PROM FLASH	Mask-Programmed Programmable (PROM)
SRAM DRAM	FIFO LIFO Shift Register CAM		

www.rejinpaul.com

Memory Timing: Approaches



DRAM Timing
Multiplexed Adressing



SRAM Timing
Self-timed

www.rejinpaul.com

Memory Architecture:

- Memory is classified based on following parameters;
- i. Size
- ii. Timing parameters
- iii. Function
- iv. Access Pattern
- v. input/output architecture
- vi. Application

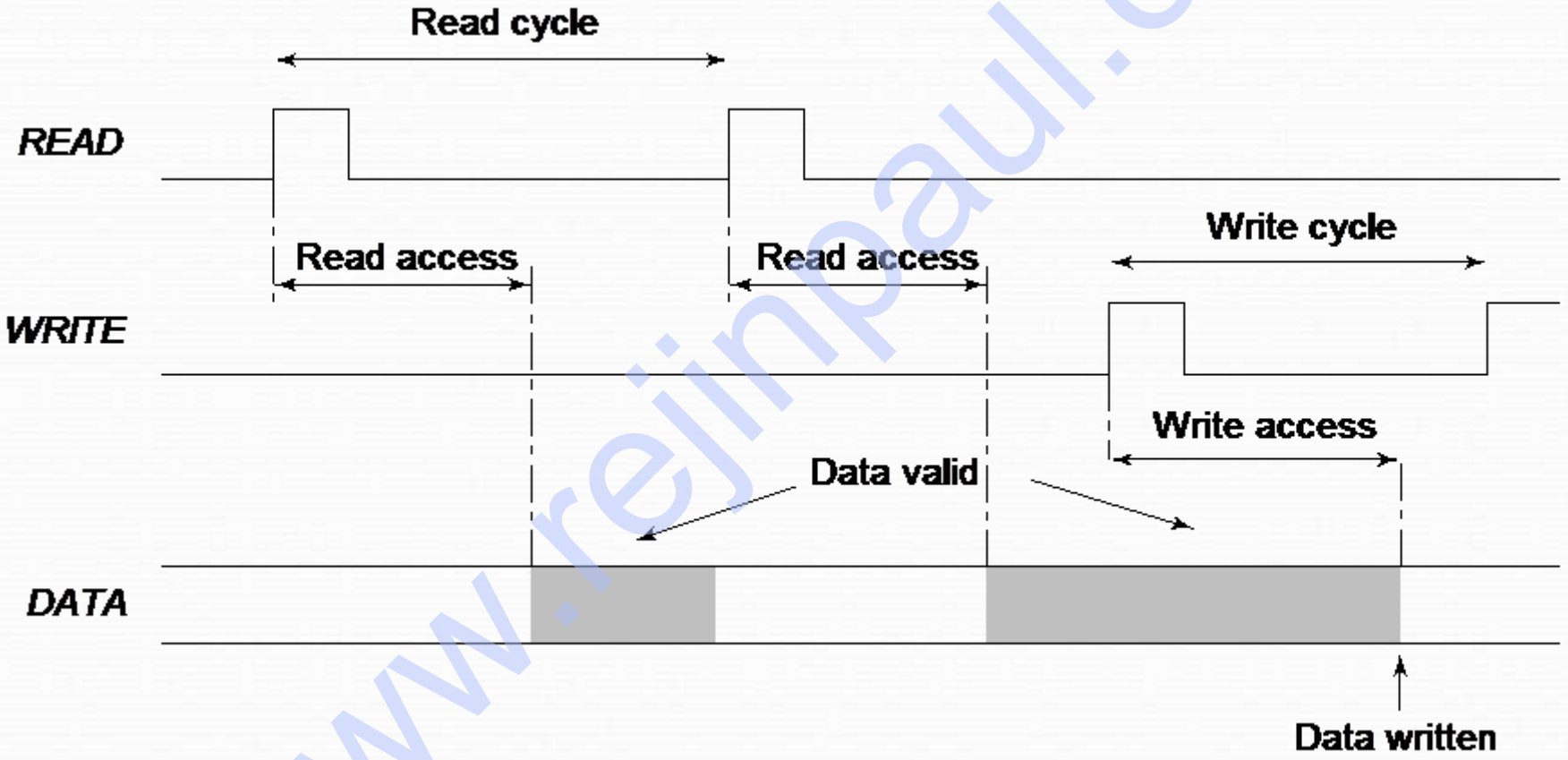
www.rejinpaul.com

Size:

- The Chip designer expresses as size in bytes
- Kb,Gb,Mb
- The system designer expresses the storage requirement in terms of words
- A 32-bit register can store 2^{32} different values

www.rejinpaul.com

Memory Timing: Definitions



www.rejinpaul.com

Function

- ROM
- RWM
- Static Memory
- Dynamic Memory
- Non volatile memory
- Volatile Memory
- EPROM
- EEPROM .etc…

www.rejinpaul.com

Access Pattern

- Random Access Memory
- FIFO
- LIFO
- Content Access Memory(CAM)/Associative memory

www.rejinpaul.com

I/O Architecture:

- Semiconductor memories are classified based on the number of data input and output ports
- Most of the memory single port that is shared for both input and outputs.
- Memories are higher bandwidth requirement have multiple input and output ports
- Ex:RISC Microprocessor

www.reinpaul.com

Application:

1. N-word Memory Architecture
2. Array structured Memory Architecture
3. Hierarchical Memory Architecture
4. Content Addressable Memory Architecture

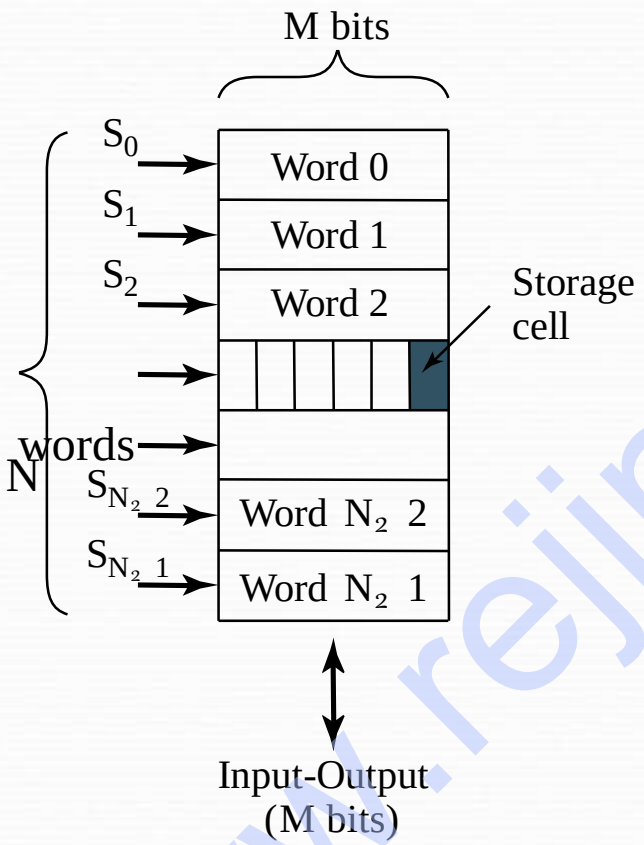
www.rejinpaul.com

Types of Memory Architectures

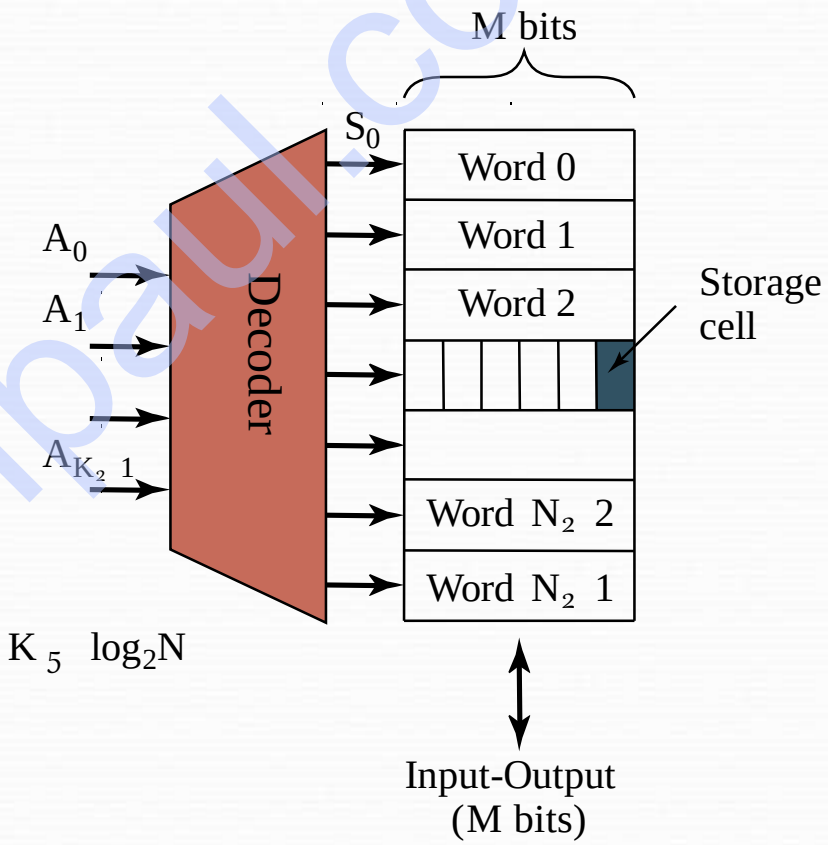
1. N-word Memory Architecture
2. Array structured Memory Architecture
3. Hierarchical Memory Architecture
4. Content Addressable Memory Architecture

www.rejinpaul.com

1) N-word Memory Architecture



N-word architecture for N x M memory
Too many select signals:
N words == N select signals



Decoder reduces the number of select signals

$K = \log_2 N$

www.rejinpaul.com

i)N-word memory Architecture

- In an N-word memory each word is M-bits wide and subsequent memory words are stacked.
- One word at a time is selected for reading or writing with the help of a select bit(S_0 to S_{n-1}).
- If this module is a single port memory, only one signal S_1 can be high at any time.
- If each storage cell is a D-flip flop,then the select signal is used to active or clock the cell.

ii)N-word memory Architecture using Decoder

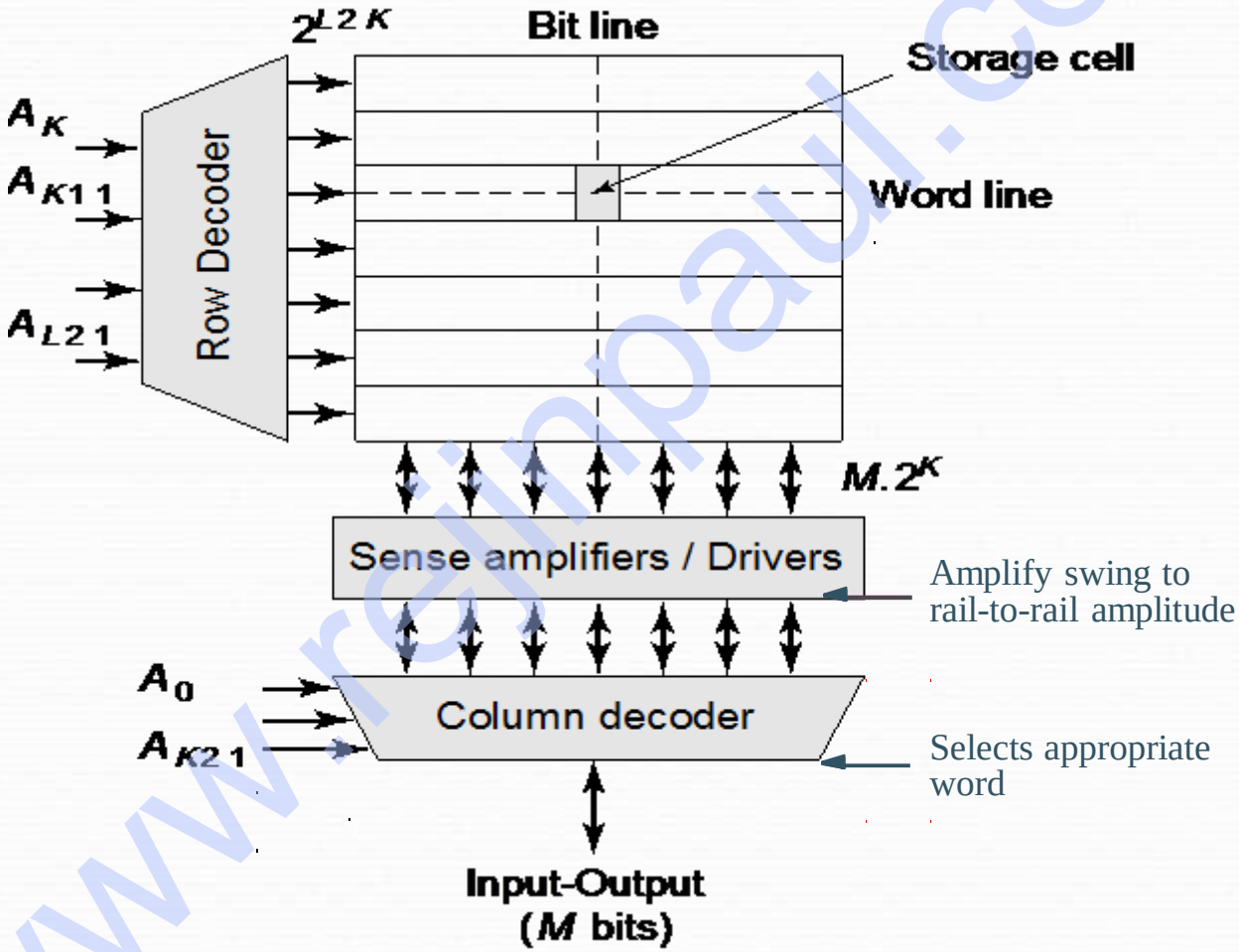
- A decoder is inserted to reduce the number of select signals.
- A memory word is selected by providing a binary encoded address word (A_0 to A_{k-1})
- The decoder translates this address into $N=2^K$ select lines, only one of which is active at a time.

Drawbacks

- Poor memory aspect ratio
- Slow operation
- High delay
- Not easy to implement

www.rejinpaul.com

2)Array-Structured Memory Architecture

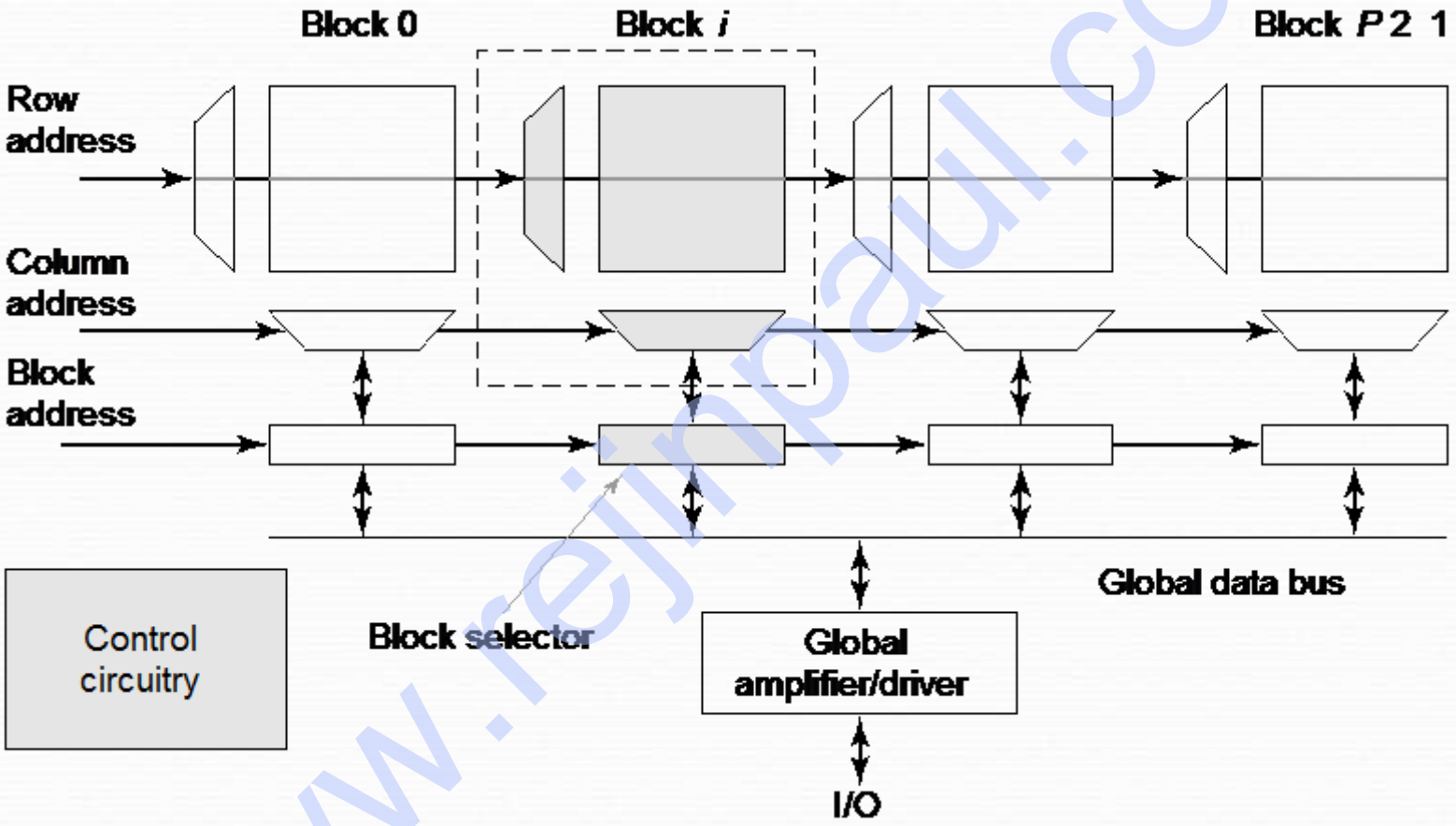


www.rejinpaul.com

- Memory arrays are organized in vertical and horizontal dimensions with same order of magnitude.
- This brings the aspect ratio near to unity.
- Multiple words are stored in a single row and are selected simultaneously.
- To route the correct word to the I/O terminals, an extra circuit called column decoder.
- The address word is partitioned into a column address (A_o to A_{k-1}) and a row address (A_k to A_{L-1}) .
- The row address enables one row of the memory for R/W,while the column address picks one particular word from the selected row.
- The horizontal select line enables a single row of cells called the word line.
- The wire that connects the cells in a single column to the I/O circuit is called the bit line.

www.rejinpaul.com

3)Hierarchical Memory Architecture



www.rejinpaul.com

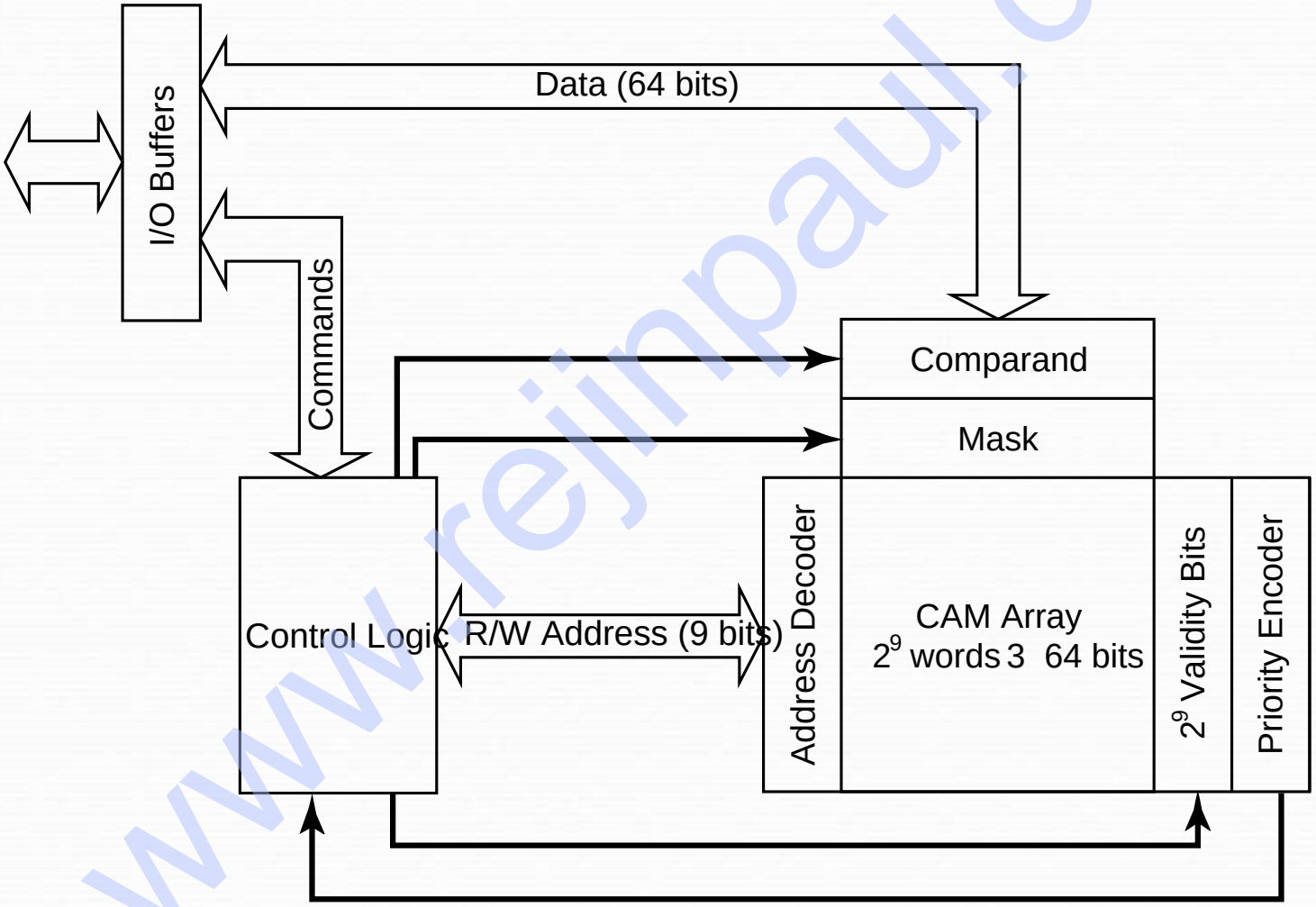
- Array structured memory size up to range 64Kbps to 256Kbps.
- The memory is partitioned into P smaller blocks .
- The composition of each of the individual block is identical.
- A word is selected on the basics of the row and column addresses that are broadcast to all the block.
- An extra address word called the block address, selects one of the P blocks to be read or written.

Advantages:

1. Shorter wires within blocks
2. Fast access times and Low power consumption
3. Block address activates only 1 block => power savings

www.rejinpaul.com

4)Contents-Addressable Memory(CAM)



www.rejinpaul.com

- CAM supports 3 modes of operation read, write and match.
- The read and write modes access and manipulate data in the CAM array as in ordinary memory.
- The match mode is unique to associate memory.
- The comparand block is filled with the data pattern to match and the mask word indicates which bits are significant.
- Every row that matches the pattern is passed to the validity block.
- The valid rows that match are passed to the priority encoder.
- If two or more rows match the pattern, the address of the row in the CAM array is used to break the tie.
- The priority encoder considers all 512 match lines from the CAM array, selects the one with the highest address and encoders it in binary.
- For 512 rows in the CAM array,9 bits are required to indicate the highest row that matched.

www.rejinpaul.com

Memory Control Circuits

1. Address Decoders
2. Sense Amplifiers
3. Input/Output Buffers
4. Control / Timing Circuitry

www.rejinpaul.com

1)Row Decoders

- Row decoder is a Collection of 2^M complex logic gates with M-input.
- It is an 8-bit address decoder.
- Each of the outputs WL_i is a logic function of the 8 input address signals(Ao to A7)
- The row with addresses 0 and 127 are enabled by the following logic functions

(N)AND Decoder

$$WL_0 = A_0 A_1 A_2 A_3 A_4 A_5 A_6 A_7 A_8$$

$$WL_{127} = \bar{A}_0 A_1 A_2 A_3 A_4 A_5 A_6 A_7 A_8$$

NOR Decoder

$$WL_0 = \overline{A_0 + A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8}$$

$$WL_{127} = \overline{A_0 + \bar{A}_1 + \bar{A}_2 + \bar{A}_3 + \bar{A}_4 + \bar{A}_5 + \bar{A}_6 + \bar{A}_7 + \bar{A}_8}$$

www.rejinpaul.com

$$WL_0 = \overline{A}_0\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4\overline{A}_5\overline{A}_6\overline{A}_7$$

$$WL_{127} = \overline{A}_0 A_1 A_2 A_3 A_4 A_5 A_6 A_7$$

$$WL_0 = A_0 + A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7$$

$$WL_{127} = A_0 + \overline{A}_1 + \overline{A}_2 + \overline{A}_3 + \overline{A}_4 + \overline{A}_5 + \overline{A}_6 + \overline{A}_7$$

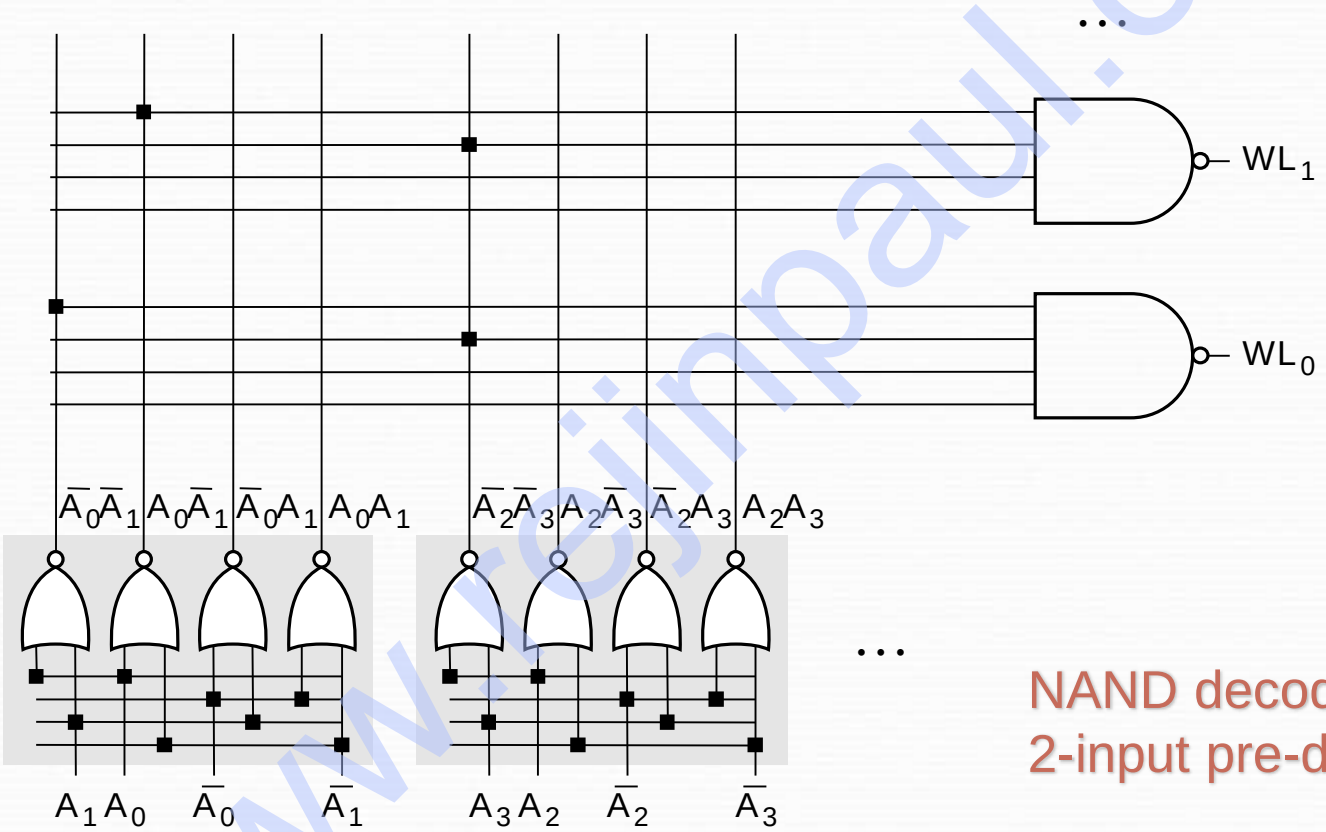
$$WL_0 = \overline{\overline{A}_0\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4\overline{A}_5\overline{A}_6\overline{A}_7}$$

$$= \overline{(A_0 + A_1)(A_2 + A_3)(A_4 + A_5)(A_6 + A_7)}$$

www.reinpaul.com

Row Decoder(Eg:NAND Decoders)

Multi-stage implementation improves performance



NAND decoder using 2-input pre-decoders

www.rejinpaul.com

2-4 DECODER USING NOR

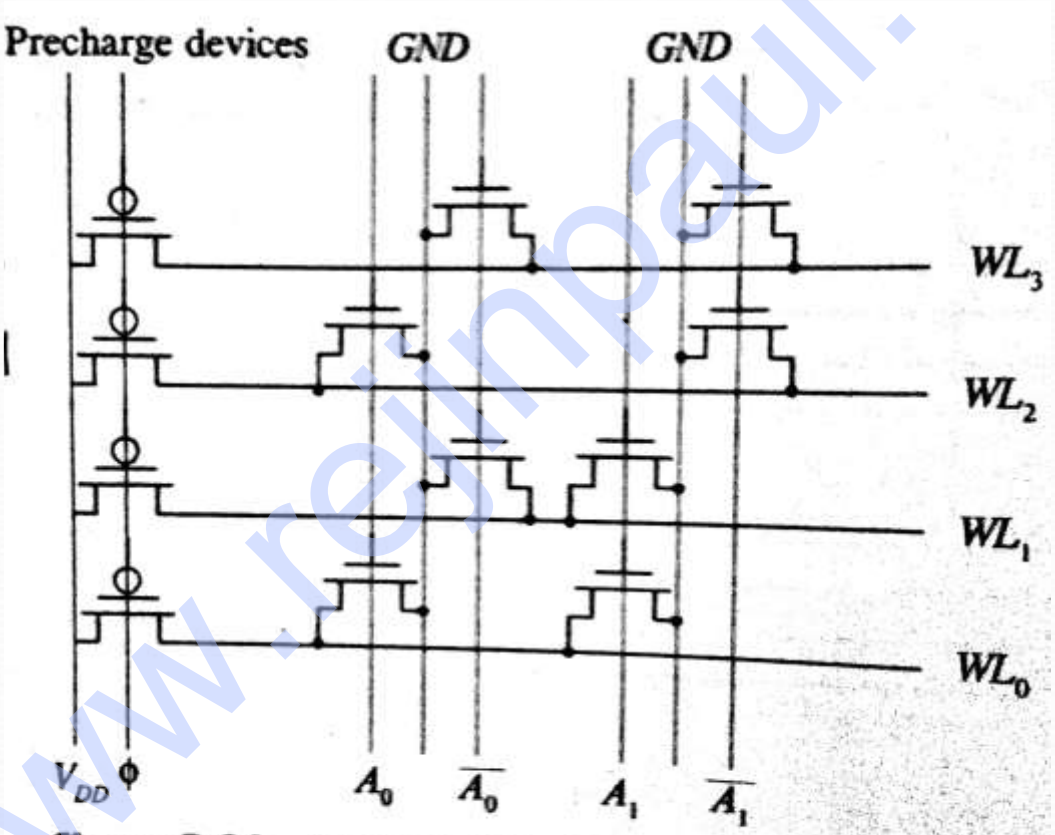
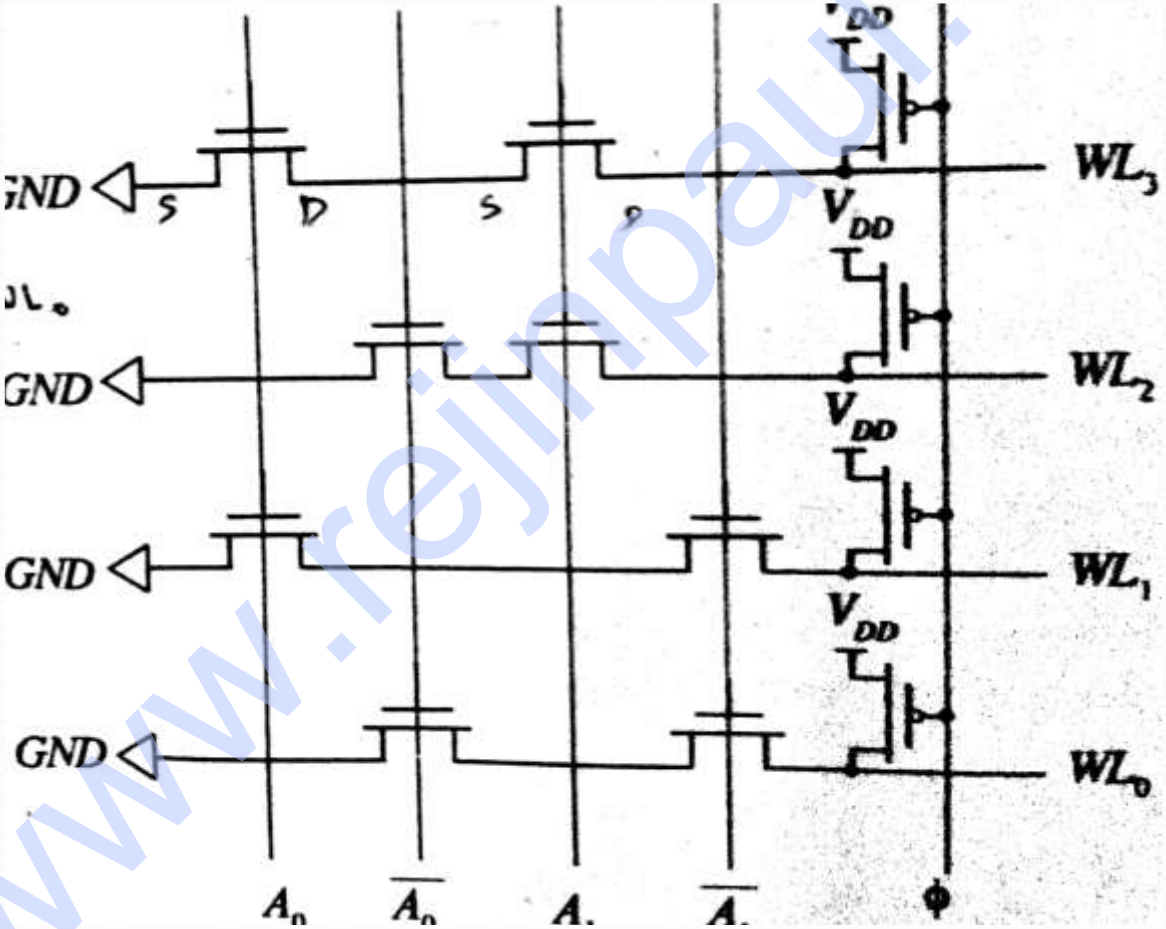


Figure 3.86 Dynamic 2-to-4 Decoder using NOR

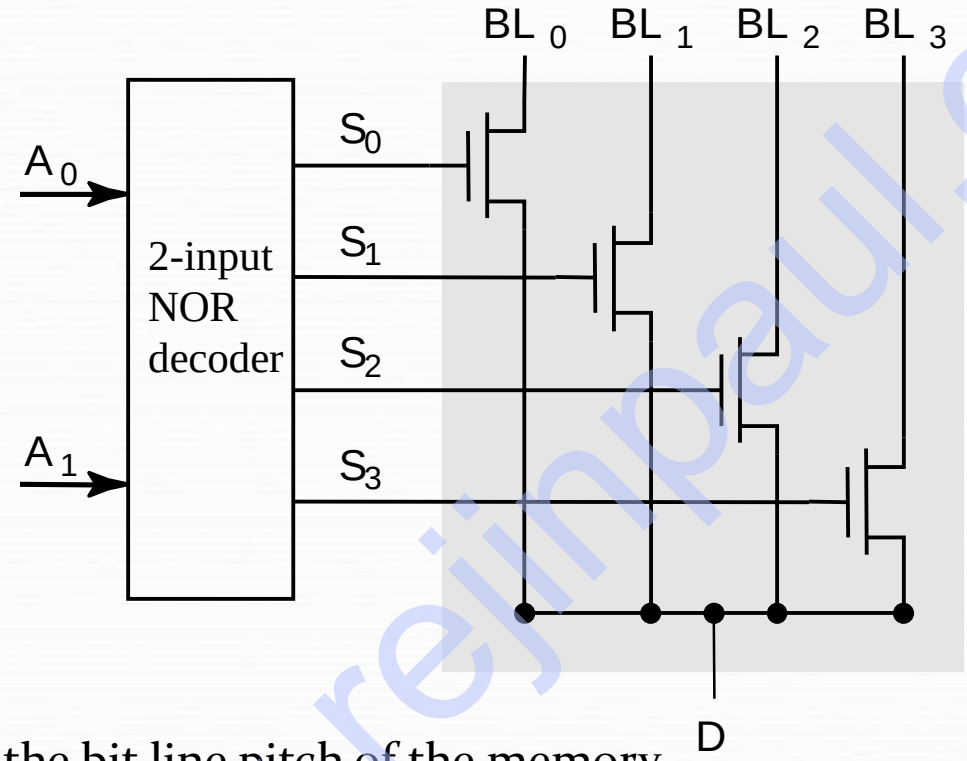
www.rejinpaul.com

2-4 DECODER USING NAND



www.rejinpaul.com

2) 4-to-1 pass-transistor based column decoder



- It matches the bit line pitch of the memory.
- 4 to1 column decoder uses nMOS transistors.
- Complementary transmission gates are used when these mux are shared between the read and write operations in both directions.

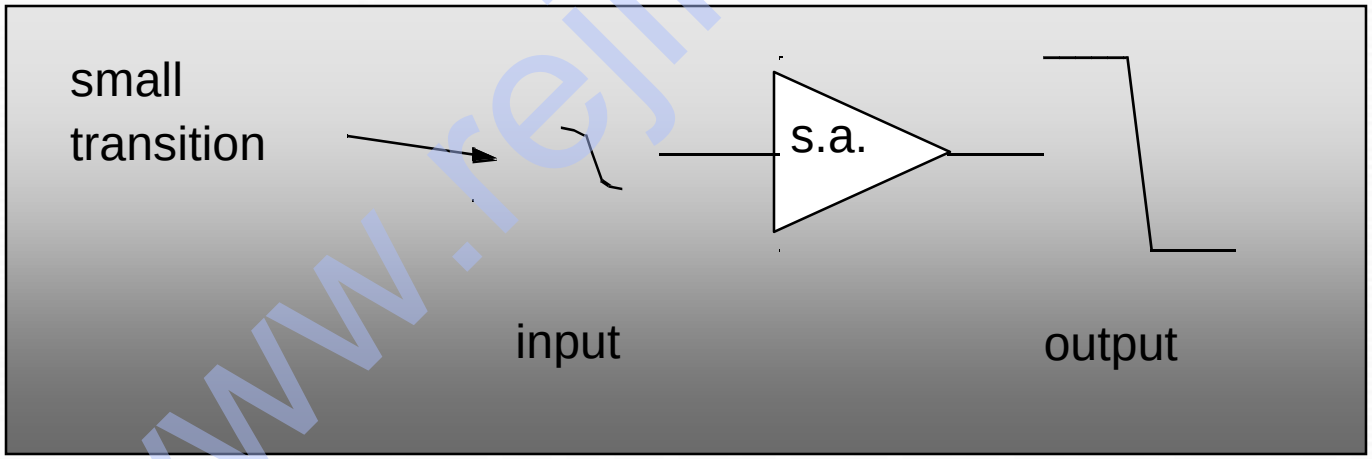
Advantages: High speed operation,propagation delay will not be added.

Disadvantage: Large transistor count

www.rejinpaul.com

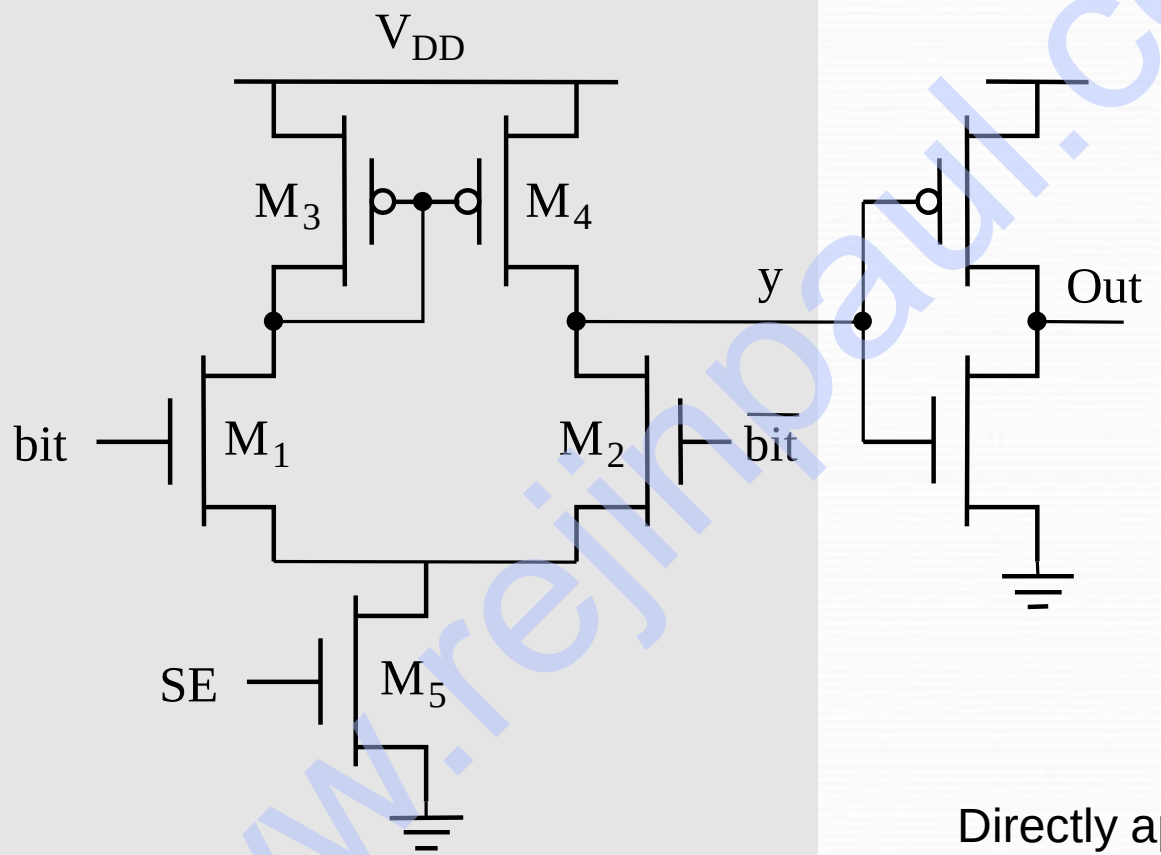
3)Sense Amplifiers

- Sense amplifier play a major role in the functionality,performnace and reliability of memory circuits.
- The important functions of sense amplifier which includes Amplification, Delay reduction, power reduction and signal restoration



www.rejinpaul.com

3.1)Differential Sense Amplifier



Directly applicable to SRAMs

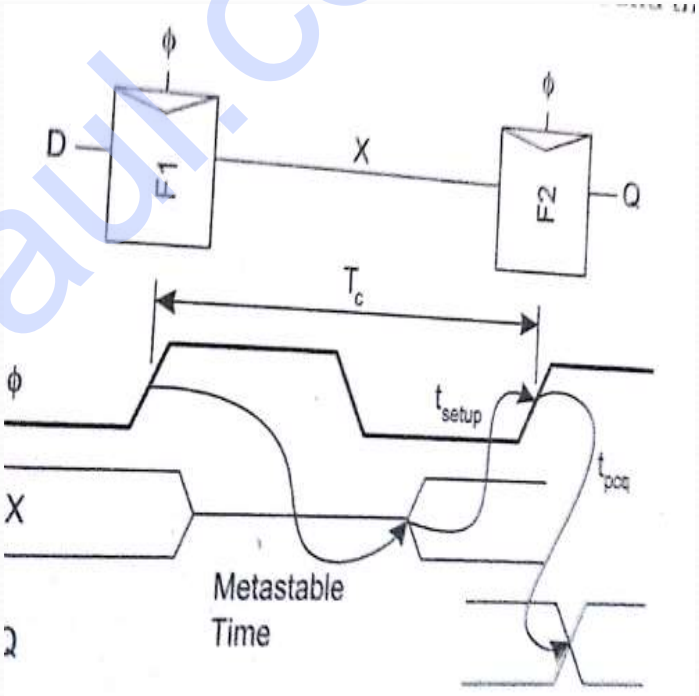
www.rejinpaul.com

- Amplification is based on the current mirroring concept.
- The input signals(bit and $\overline{bit}$)are heavily loaded and driven by SRM memory cell.
- The swing on those line is small as the small memory cell drives a large capacitive load.
- The input is applied M1and M2 and transistor M3 and M4 are an active current mirror load.
- When SE is high the amplifier is conditioned.
- Once the read operation is initiated, one of the bit lines drops.
- SE is enabled when a sufficient differential signal is established and the amplifier evaluates

www.rejinpaul.com

Synchronizer

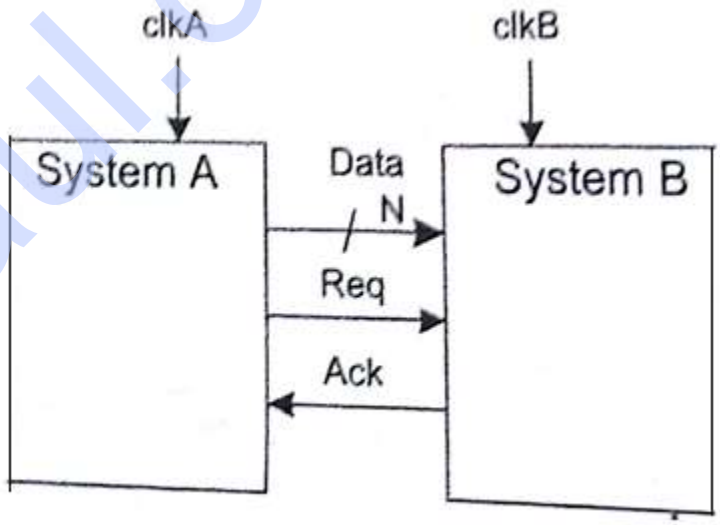
- It is a circuit that accepts an input that can change at arbitrary times and produces an output based on synchronizer clock.
- Synchronizer accepts input D and clk.
- Produces output Q after some delay.
- F1 samples the input D.
- Output X may be meta-stable for some time.
- F2 samples X→produces valid output.
- It can be aligned by clock.
- The synchronizer has a latency of one clock cycle Tc.so X has not settled to valid level.



www.rejinpaul.com

Communication between Asynchronous clock domains

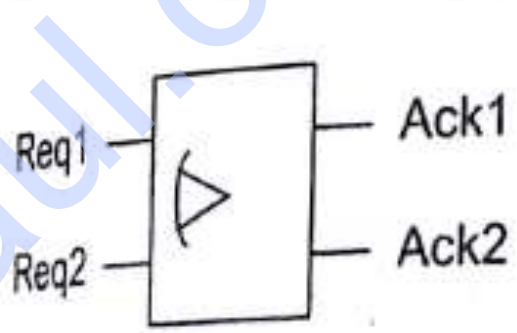
- Not share a common clock
- System A send Data N to system B.
- Each word should be rx by once by B.
- A Send Req→B(To check whether the data is stable in B).
- B send Ack→A.(sampled data is stable and data is received safely)



www.rejinpaul.com

Arbiter

- Same as like synchronizer.
- It determines which of the two inputs arrived as first.
- If the spacing between the inputs exceeds some aperture time, then the first input Tx.
- If the space is small, exactly one of two inputs should be tx.



www.rejinpaul.com

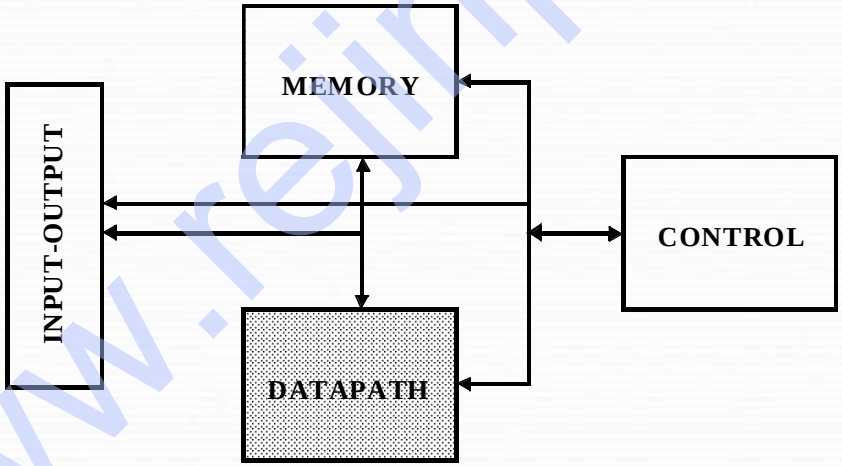
Unit-IV

DESIGNING ARITHMETIC BUILDING BLOCKS

www.rejinpaul.com

Introduction

- The overall system performance dominated by the speed and power consumption of circuits like adders, multipliers and shifters.
- Digital processor has components like datapath,memory,control and I/O blocks.
- Data path is the core of processor, in which all computation are performed.
- A data path consists of interconnection of combinational functions such as arithmetic (addition,multiplication,comparison and shift).



www.rejinpaul.com

Building Blocks for Digital Architectures

Arithmetic unit

- Bit-sliced datapath (adder, multiplier, shifter, comparator, etc.)

Memory

- RAM, ROM, Buffers, Shift registers

Control

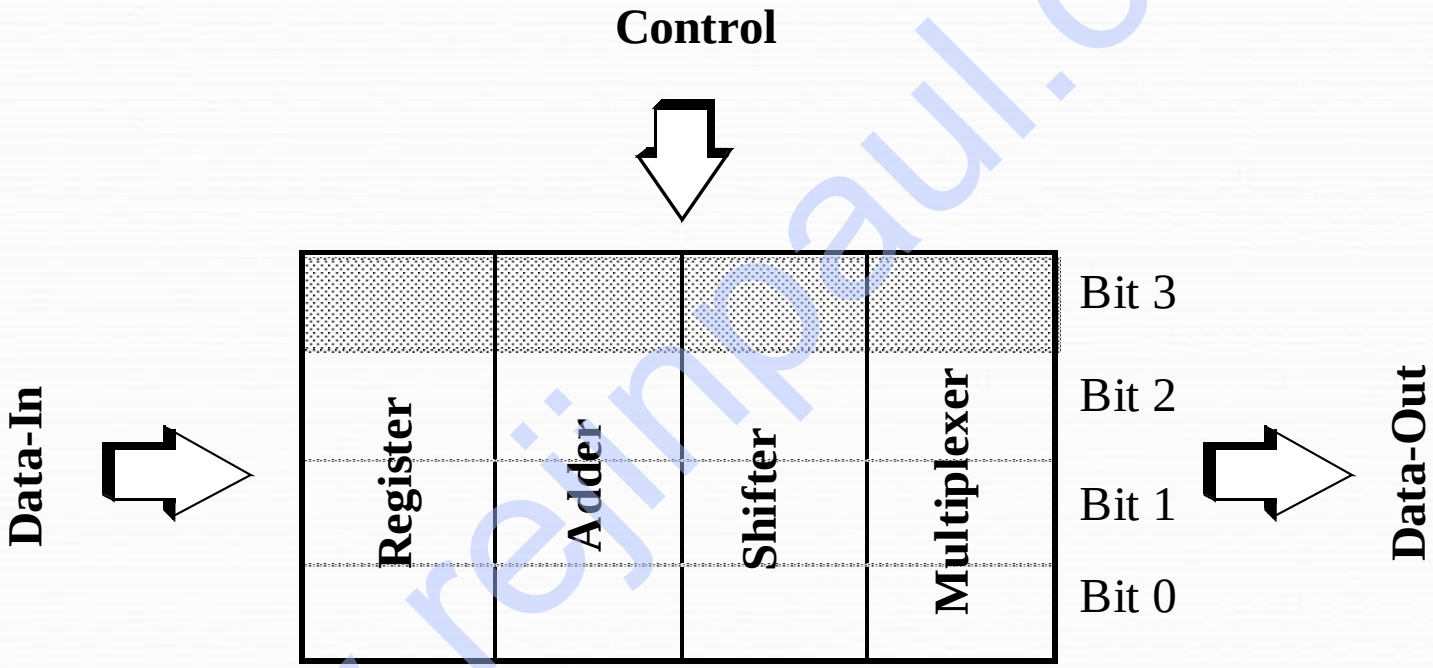
- Finite state machine (PLA, random logic.)
- Counters

Interconnect

- Switches
- Arbiters
- Bus

www.rejinpaul.com

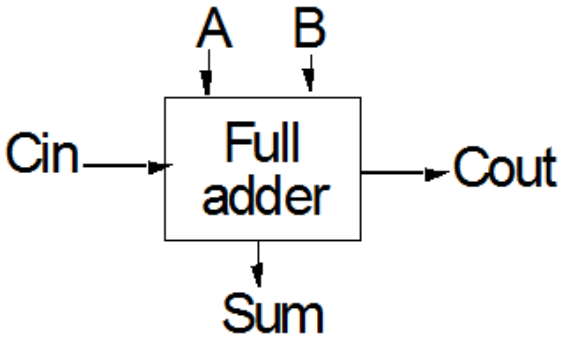
Bit-Sliced Data path Design



Tile identical processing elements

www.reinpaul.com

The Binary Adder



$$S = A \oplus B \oplus C_i$$

$$= A\bar{B}\bar{C}_i + \bar{A}B\bar{C}_i + \bar{A}\bar{B}C_i + ABC$$

$$C_o = AB + BC_i + AC_i$$

A	B	C_i	S	C_o	*Carry status*
0	0	0	0	0	delete
0	0	1	1	0	delete
0	1	0	1	0	propagate
0	1	1	0	1	propagate
1	0	0	1	0	propagate
1	0	1	0	1	propagate
1	1	0	0	1	generate
1	1	1	1	1	generate

The output S and Co are defined as some function of intermediate signals G,D and P.

G=1→Ensures that a carry bit will be generated (1)

D=1 carry bit will be deleted.

P=1 gurantees that an incoming carry C_i will propagate to C_o

www.rejinpaul.com

Express Sum and Carry as a function of P, G, D

Define 3 new variable which ONLY depend on A, B

Generate (G) = AB

Propagate (P) = $A \oplus B$

Delete = $\overline{A}\,\overline{B}$

$$C_o(G,P) = G + PC_i$$

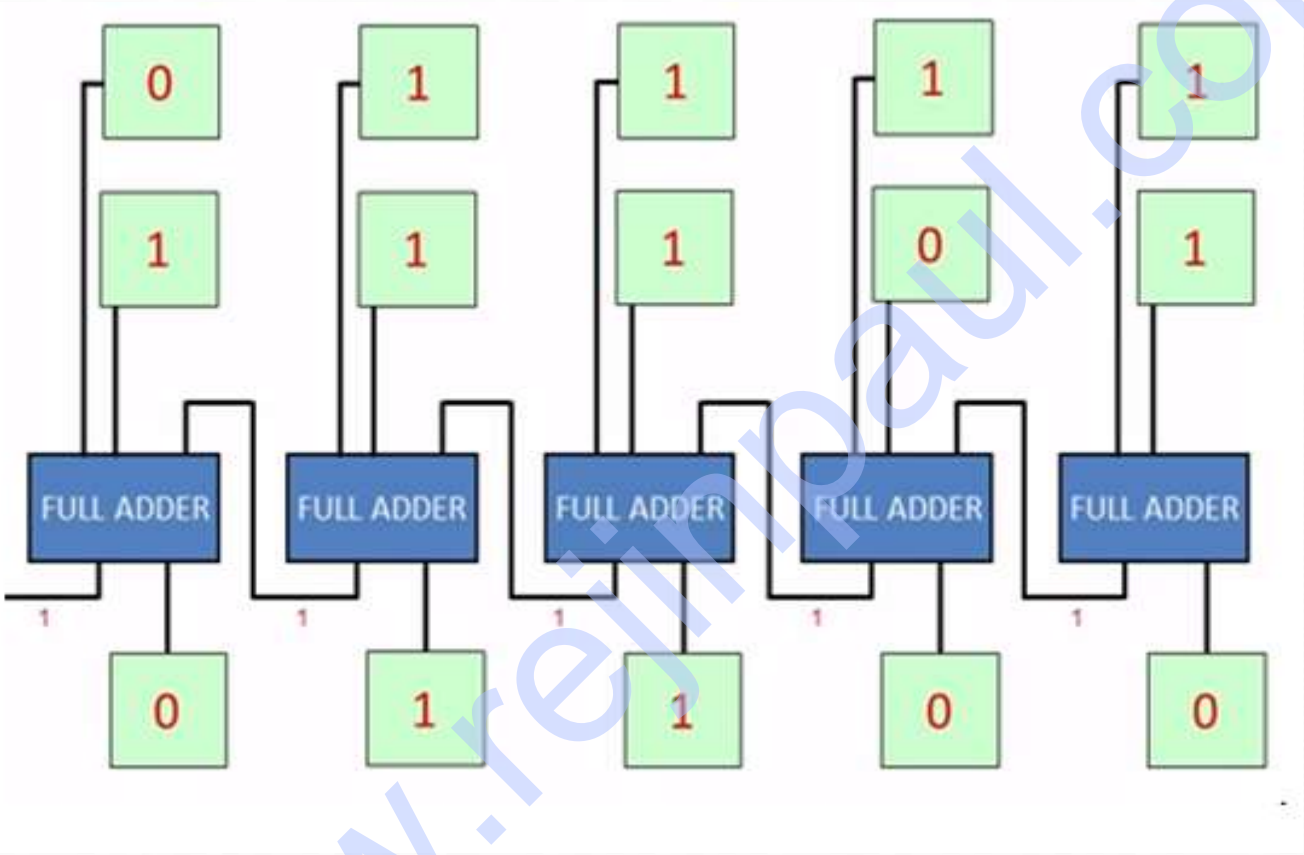
$$S(G,P) = P \oplus C_i$$

Can also derive expressions for S and C_o based on D and P

Note that we will be sometimes using an alternate definition for

Propagate (P) = A + B

www.reinpaul.com



www.reinpaul.com

Advantages

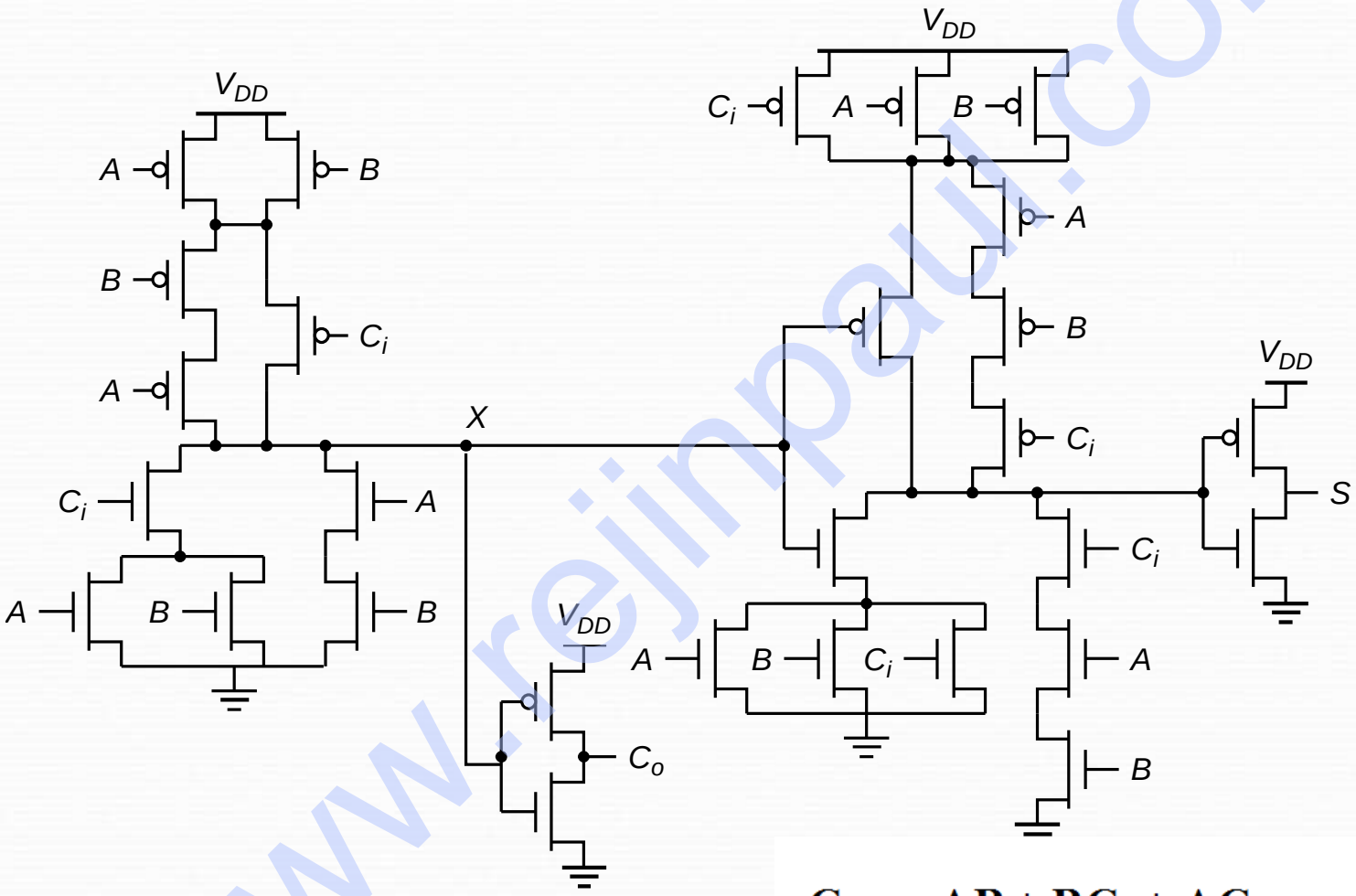
- Consumes less power
- Compact layout giving Smaller chip area

Drawbacks

- Delay is linearly proportional to the number of bits
- Performance decrease with increase in number of bits.
- Worst case delay for N-bit adder is 2N gate delay.

www.reijnpaul.com

2)Complimentary Static CMOS Full Adder



28 Transistors

$$C_o = AB + BC_i + AC_i$$

$$S = ABC_i + C_o\{A + B + C_i\}$$

www.rejinpaul.com

•The complementary static circuitry has less transistor (28) .
•This circuit takes less latency in carry generation.
•This circuit implemented using OR and AND gates instead of XOR gate.

$$C_o = AB + BC_i + AC_i$$

$$S = ABC_i + C_o\{A + B + C_i\}$$

Advantages

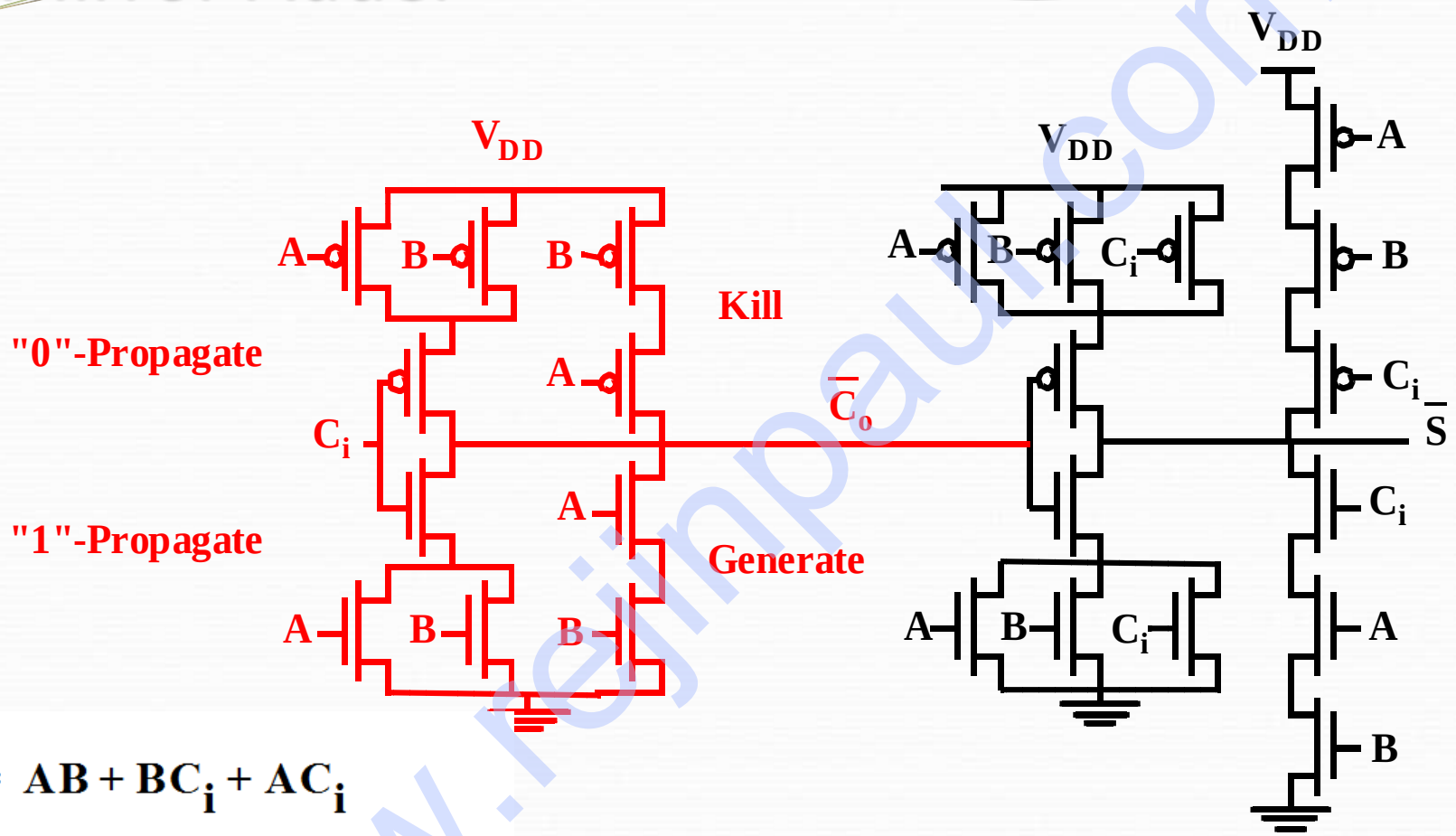
•Reduce logical effort
•Functions at low power supply.
•Provide full swing output and reliability

Drawbacks

•Propagation delay is high
•Consumes large area.
•Large pMOS transistor in pull-up network gives high input capacitance causing extra delay and dynamic power.

www.rejinpaul.com

3) Mirror Adder



$$C_o = AB + BC_i + AC_i$$

$$S = ABC_i + C_o\,(A + B + C_i)$$

24 transistors

www.rejinpaul.com

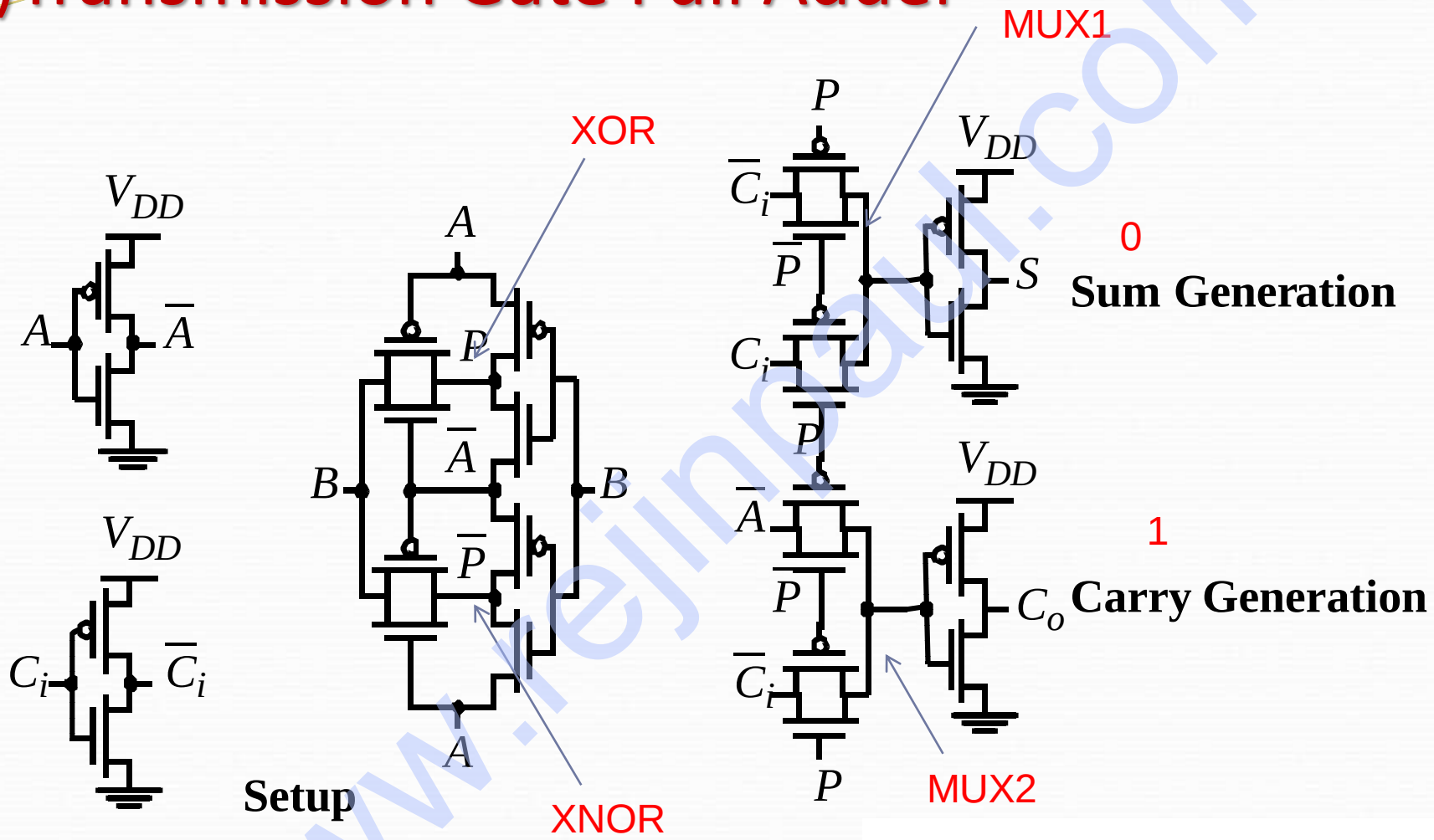
The Mirror Adder

- The NMOS and PMOS chains are completely symmetrical.
  A maximum of two series transistors can be observed in the carry-generation circuitry.
- When laying out the cell, the most critical issue is the minimization of the capacitance at node C_0. The reduction of the diffusion capacitances is particularly important.
- The capacitance at node C_0 is composed of four diffusion capacitances, two internal gate capacitances, and six gate capacitances in the connecting adder cell .
- The transistors connected to C_i are placed closest to the output.
- Only the transistors in the carry stage have to be optimized for optimal speed. All transistors in the sum stage can be minimal size.

Advantages	Drawbacks
Reduction in delay and area	Additional inverters required sum&carry fn
Reduced power consumption	Takes more delay to compute sum & carry
Design procedure is easy	

www.rejinpaul.com

4)Transmission Gate Full Adder



Propagate (P) = A⊕B

$$C_o = AB + BC_i + AC_i$$

$$S = ABC_i + C_o\,(A + B + C_i)$$

www.rejinpaul.com

- It is designed by using Multiplexers and XORs.
- Both are implemented as transmission gates.
- It uses 24 transistors only.
- The outputs P and $\overline{P}$ are fed to the output array of transmission gate to produces sum and carry.
- If the inputs bits are applied simultaneously, then both sum and carry bits are valid at the same time.
- If A =1,B=0 or A=0,B=1→ the output of XOR is high,this enable(MUX1).
- In MUX1 the sum output depends on the complementary input of Ci.
- The output of MUX2 is the complement of $\bar{A}$
  - $C_o(G,P)=G+P\ C_i$
  - $S(G,P)=P\ XOR\ C_i$

Advantages

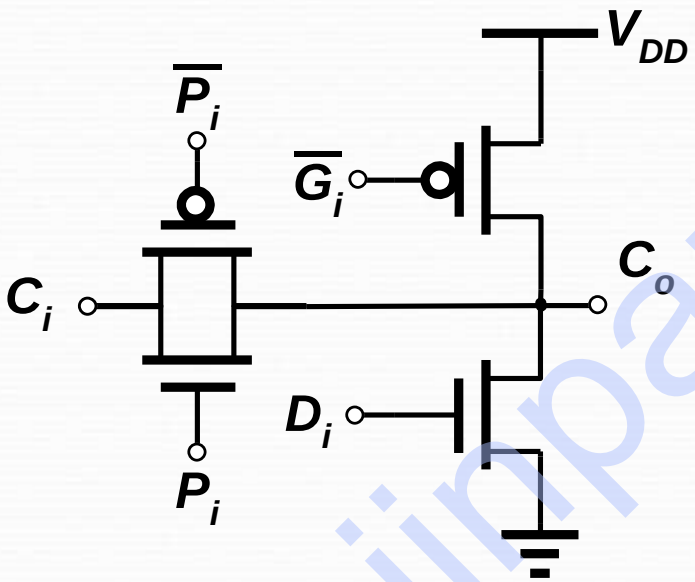
- Path have small resistance
- Both sum and carry validated at the same time.
- Delay of sum and carry outputs are almost equal.

Drawbacks

- High power dissipation
- Poor driving capability
- Having cascading problem

www.rejinpaul.com

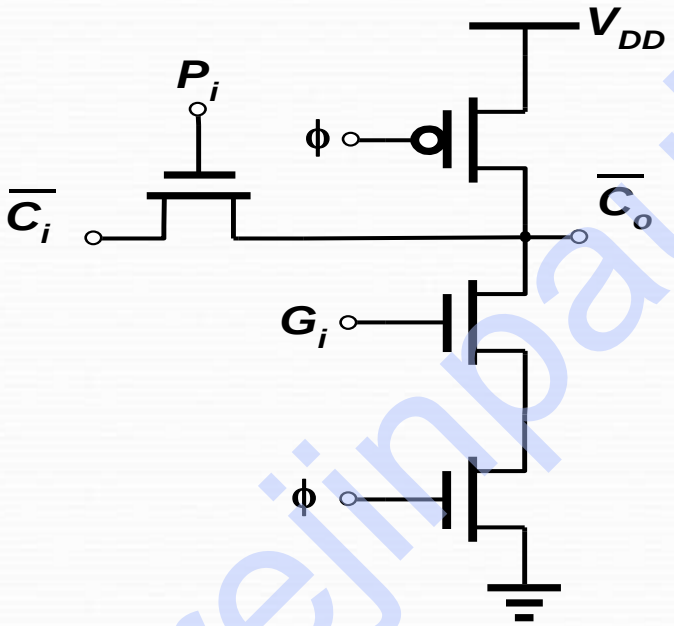
5)Manchester Carry Chain in static logic



- The propagate path is unchanged and C_i passed to C_o output if the propagate signal A_i XOR B_i is true.
- If propagate condition is not satisfied, output is either pulled low by D_i or pulled up by $\overline{G_i}$

www.rejinpaul.com

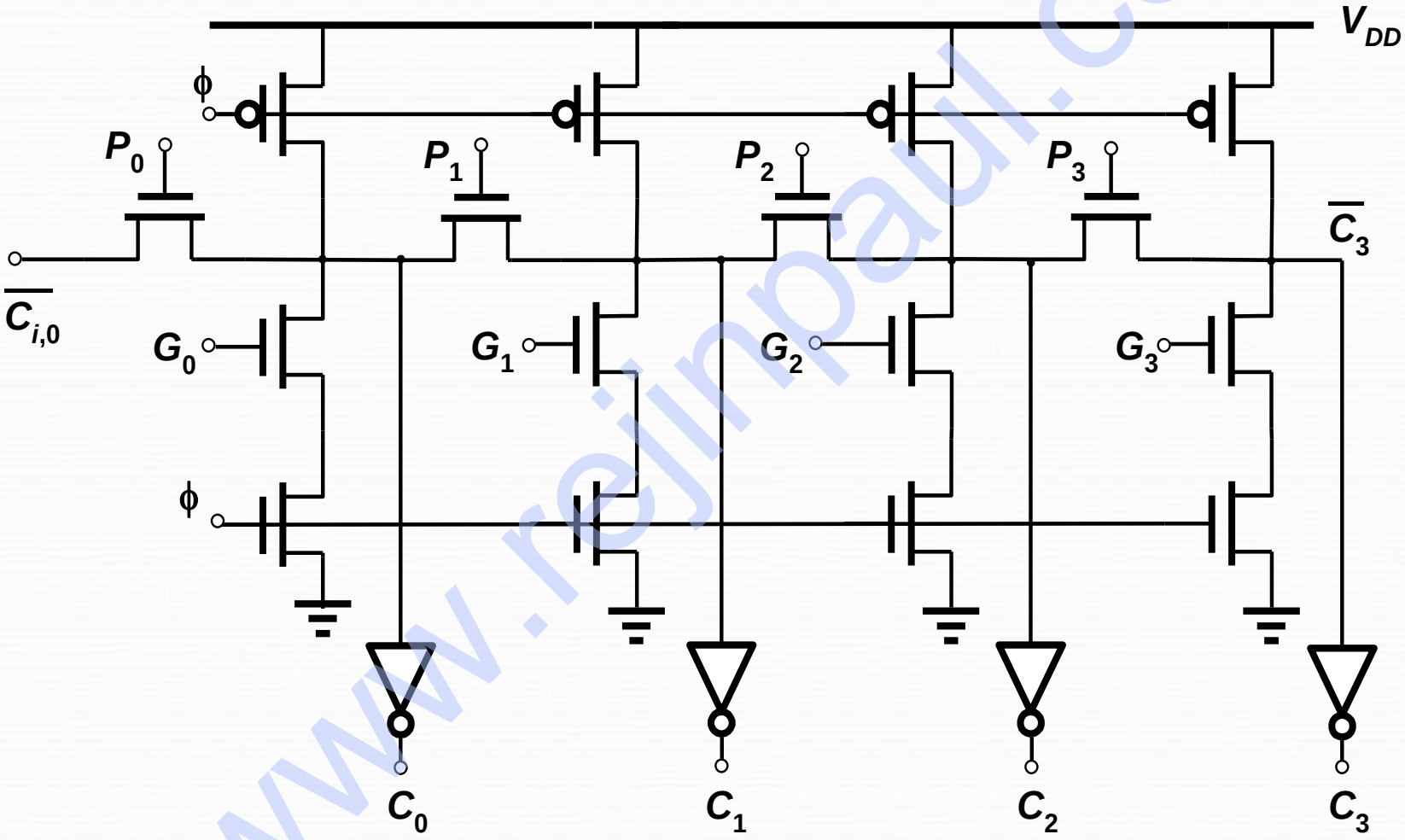
5.1)Manchester Carry Chain in Dynamic logic using propagate and generate



- The transistors in dynamic circuit are monotonic.
- The transmission gate are replaced by nMOS pass transistors .
- Output precharging eliminates the need of kill signal.

www.rejinpaul.com

5.2)Manchester Carry Chain Adder in Dynamic logic



www.rejinpaul.com

•It uses a casacade of pass tansistors to implement the carry chain.
•During precharge ϕ=0,all intermediate nodes of the pass transistor carry chain are precharged to VDD
•During evaluation the node Ai node is discharged when there is an incoming carry and propagate signal Pk is high or when the generate signal for stage k (Gk) is high.
•The worst case delay is given by

$$t_p=0.69\ RC(N(N+1)/2)$$

Advantages
•Has faster carry and propagation output
•It has better performance than static logic
•Implemented using pass transistor which reduces total transistor count.
Drawbacks
•Charge sharing and capacitive loading problem occurs
•It has threshold variation problem so level restoration circuits are needed.

www.rejinpaul.com

Ripple-Carry Adders

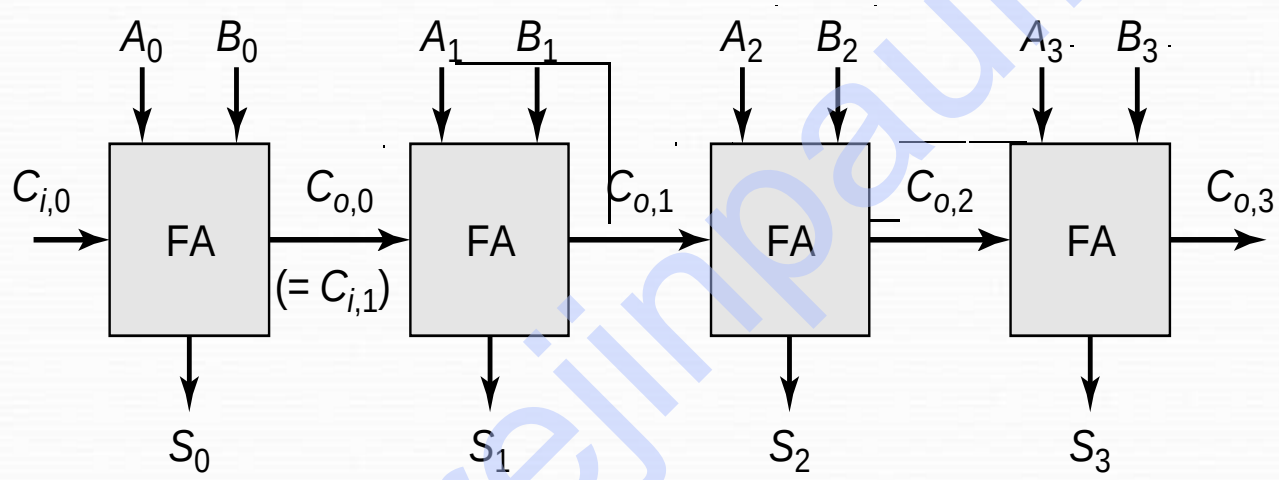
- N-bit Adder can be constructed any cascading N full adder circuits in series , connecting $C_{o,k-1}$ to $C_{i,k}$ for k=1 to N-1 and the first carry in $c_{i,o}$ to 0.
- In ripple carry adder the Carry bit ripples from one stage to the other.
- The circuit delay depends upon the number of logic stages that must traversed and is a function of the applied input signals.
- For some input signals→no ripple occurs. or
- For some input signals→ the carry has to ripple all the way from LSB to MSB.
- The propagation delay(critical path) is the worst case delay over all possible input patterns.
- Delay is proportional to the number of bits in the input words N is given by

$$t_{adder} = (N-1)t_{carry} + t_{sum}$$

- t_{carry}→ propogation delay from C_i to C_o
- t_{sum}→ propogation delay from C_i to S

www.rejinpaul.com

4-Bit Ripple-Carry Adder



case(i) A0A1A2A3=1111;B0B1B2B3=0000;Sum=1111;C in all stages=0
Case(ii) A0A1A2A3=1111;B0B1B2B3=1000; sum=0111;C in last stage =1

www.rejinpaul.com

2)High Speed Adders

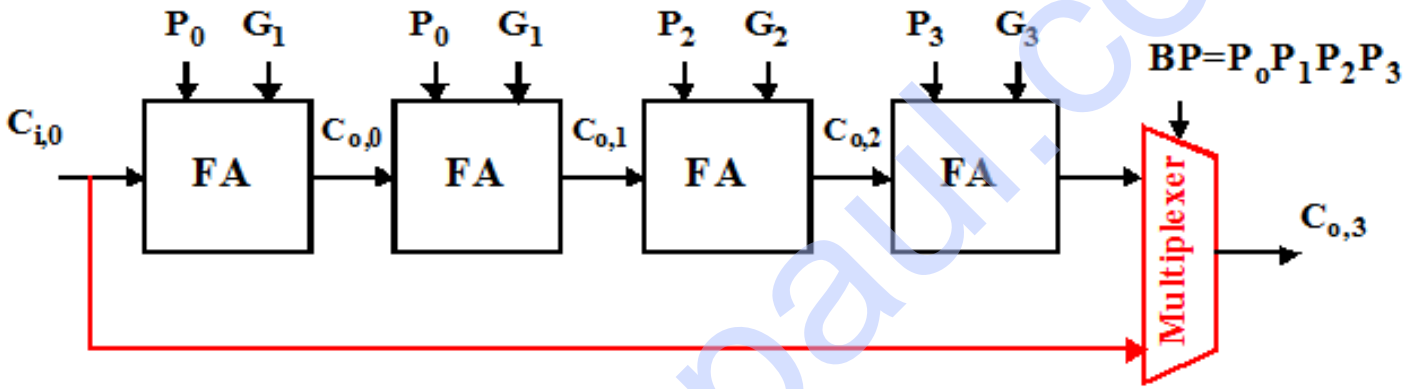
- Ripple carry adders is used for small word length.
- For large word length (multimedia processor requires word length up to (128bits)very fast computation is required.

Types

1. Carry-Bypass Adder
2. Linear Carry-Select Adder
3. Square-Root Carry-Select Adder
4. Carry-Look ahead Adder

www.rejinpaul.com

2.1)Carry-Bypass Adder

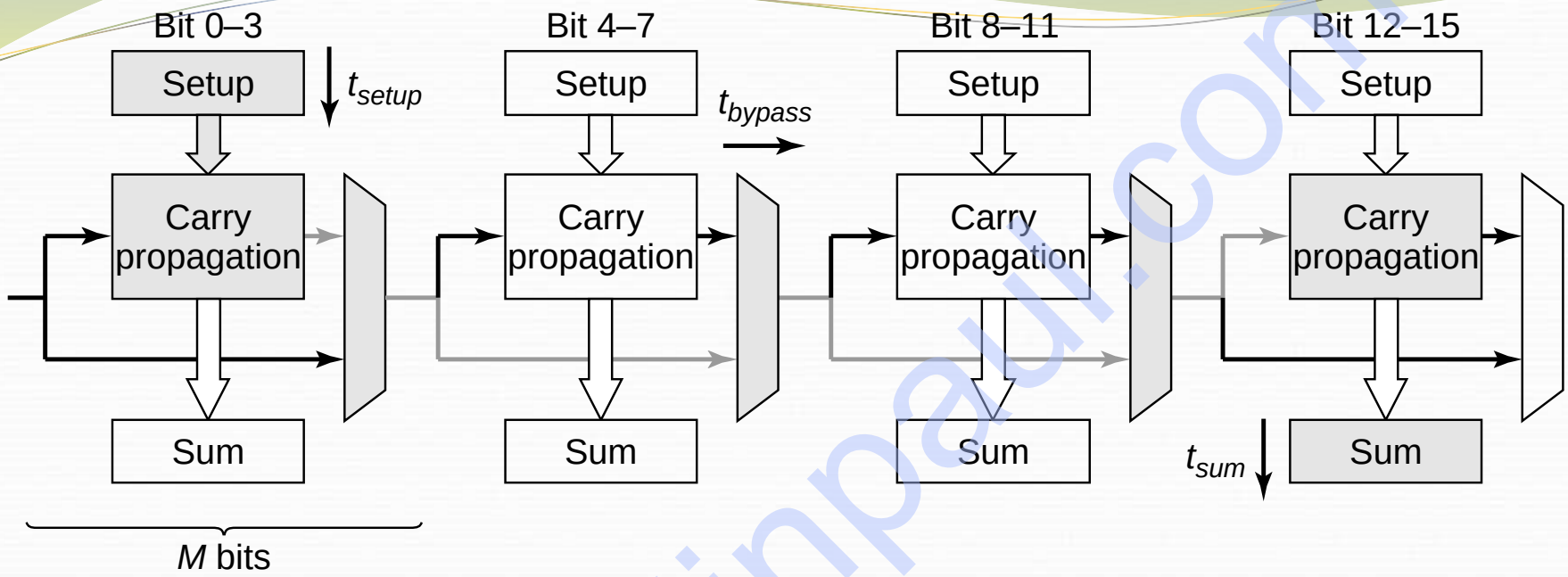


Idea: If (P0 and P1 and P2 and P3 = 1) then $C_{o3} = C_0$, else "kill" or "generate".

•When BP=$P_0P_1P_2P_3$ is 1→the very first incoming carry pass to bypass transistor, otherwise the carry is obtained via the normal route.

www.rejinpaul.com

Carry-Bypass Adder-path delay



- For computing delay of an N-bit adder, the total adder is divided in to N/M equal length with M bits.
- The worst case delay for 16-bit carry bypass adder is given by

$$t_{adder} = t_{setup} + M_{tcarry} + (N/M-1)t_{bypass} + (M-1)t_{carry} + t_{sum}$$

- t_{setup} → fixed overhead time to create the generate and propagate signal
- t_{carry} → propagation delay through single bit
- t_{bypass} → propagation delay through the bypass mux of a single stage
- t_{sum} → time to generate the sum of the final stage

www.rejinpaul.com

Propagation delay(Carry Ripple versus Carry Bypass)

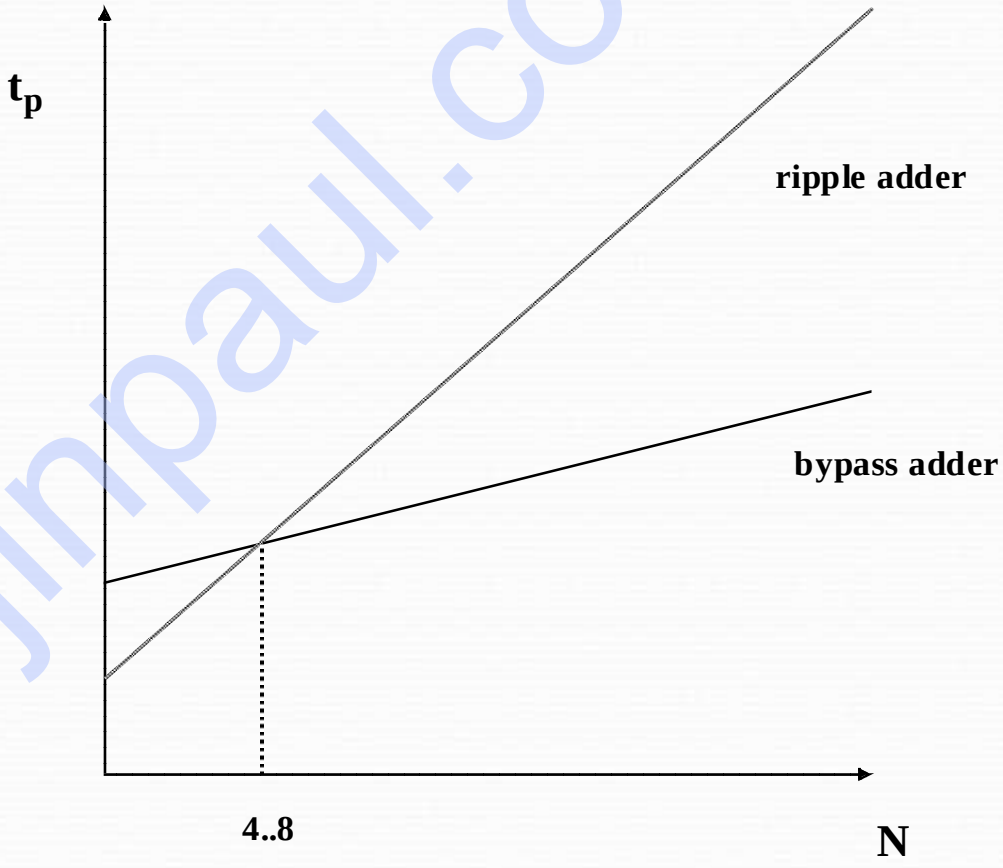
•Delay for the ripple carry adder increases approximately for N-bit.
•Delay is linearly increases for Bypass adder.
•The cross over points depends upon technology considerations and is situated between 4 and 8bits.

Advantages

•Less delay than ripple carry adder
•Allows carry to skip over groups of M-bits

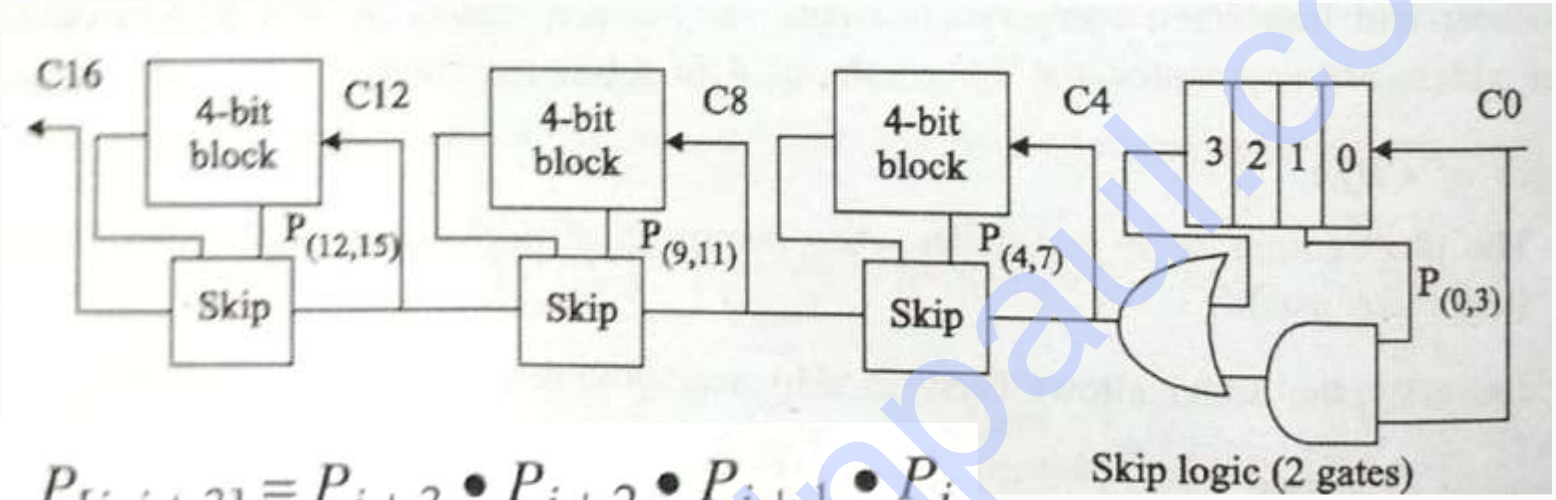
Drawbacks

•Area overhead created by adding the bypass path increases by 10-20%
•Adding the bypass path breaks the regular bit-slice structure.
•For N-bit bypass the delay is linear.



www.rejinpaul.com

Carry Skip Adder



$$P_{[i,\,i+3]} = P_{i+3} \bullet P_{i+2} \bullet P_{i+1} \bullet P_i$$

$$\text{CARRY} = C_{[i+3]} + P_{[i,\,i+3]} \cdots C_i$$

Carry Propagate: $P_i = A_i \oplus B_i$

Sum: $S_i = P_i \oplus C_i$

Carry Out: $C_{i+1} = A_i B_i + P_i C_i$

www.rejinpaul.com

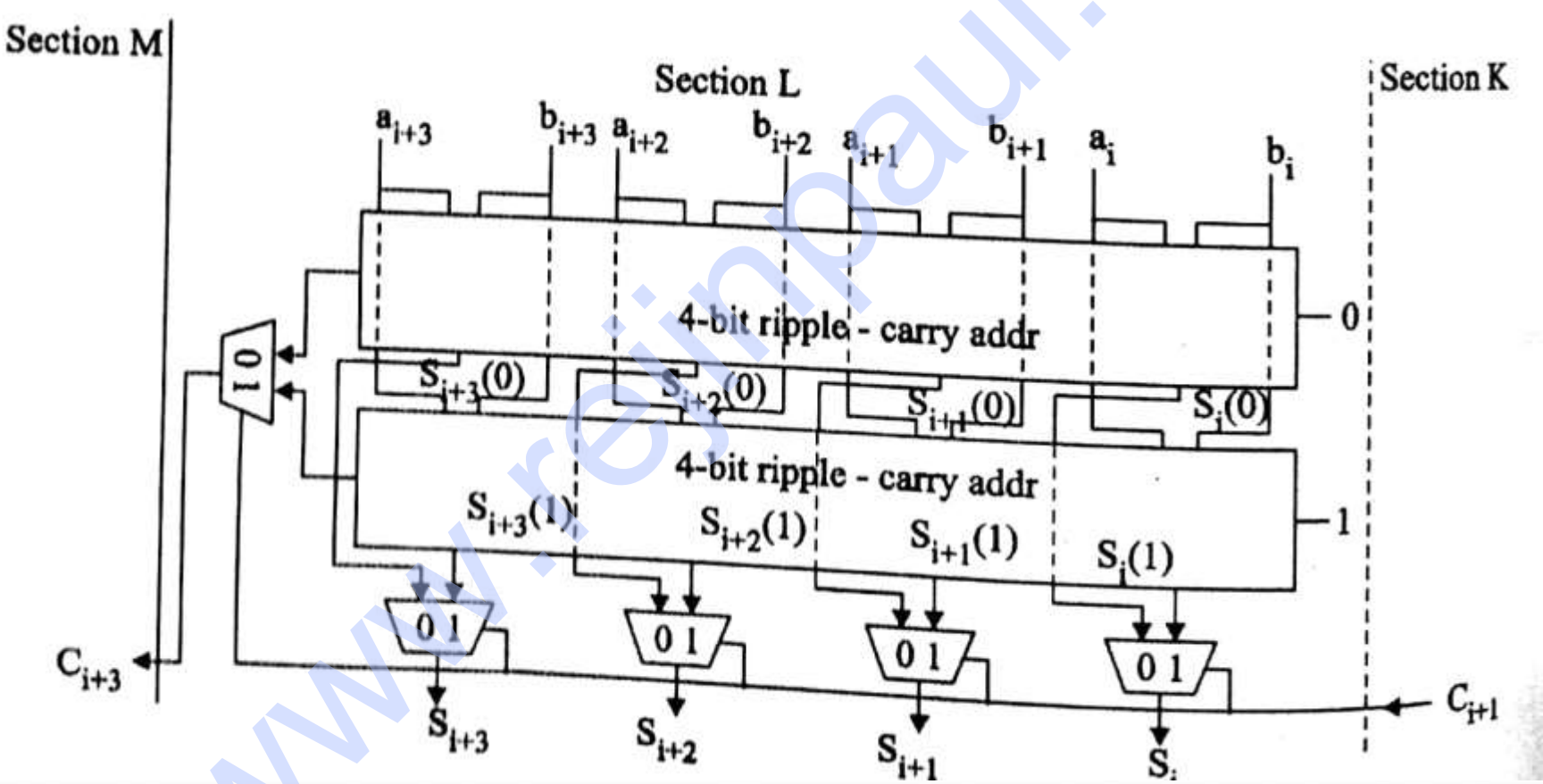
If $A_i = B_i$ then $P_i = 0$, making the carry out, C_{i+1}, depend only on A_i and $B_i \rightarrow C_{i+1} = A_i B_i$

- $C_{i+1} = 0$ if $A_i = B_i = 0$
- $C_{i+1} = 1$ if $A_i = B_i = 1$, $P_i = 1$

Alternatively if $A_i \neq B_i$ then $P_i = 1 \rightarrow C_{i+1} = C_i$

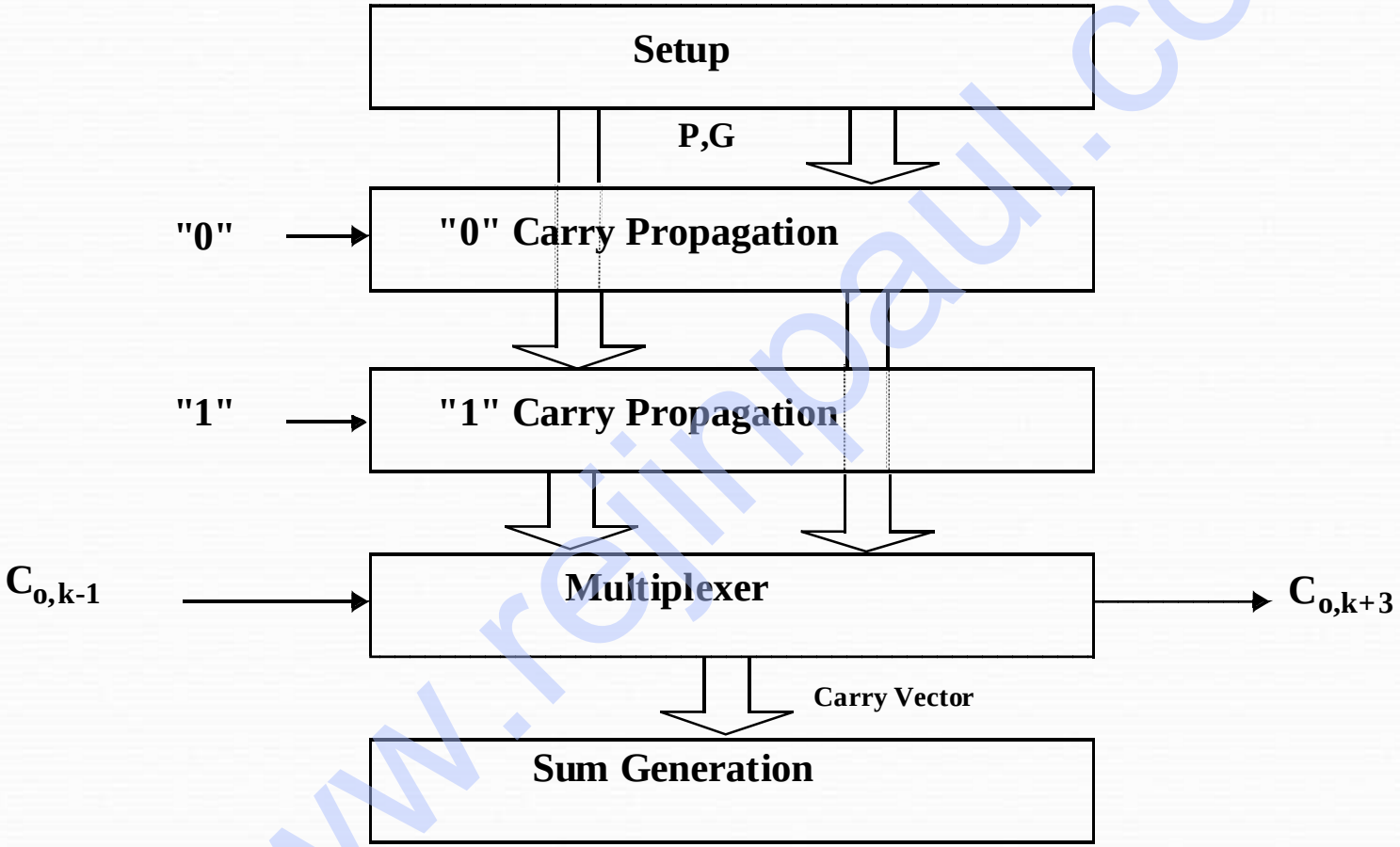
www.rejinpaul.com

Linear Carry-Select Adder



www.rejinpaul.com

2.2)Linear Carry-Select Adder



4-bit carry select module

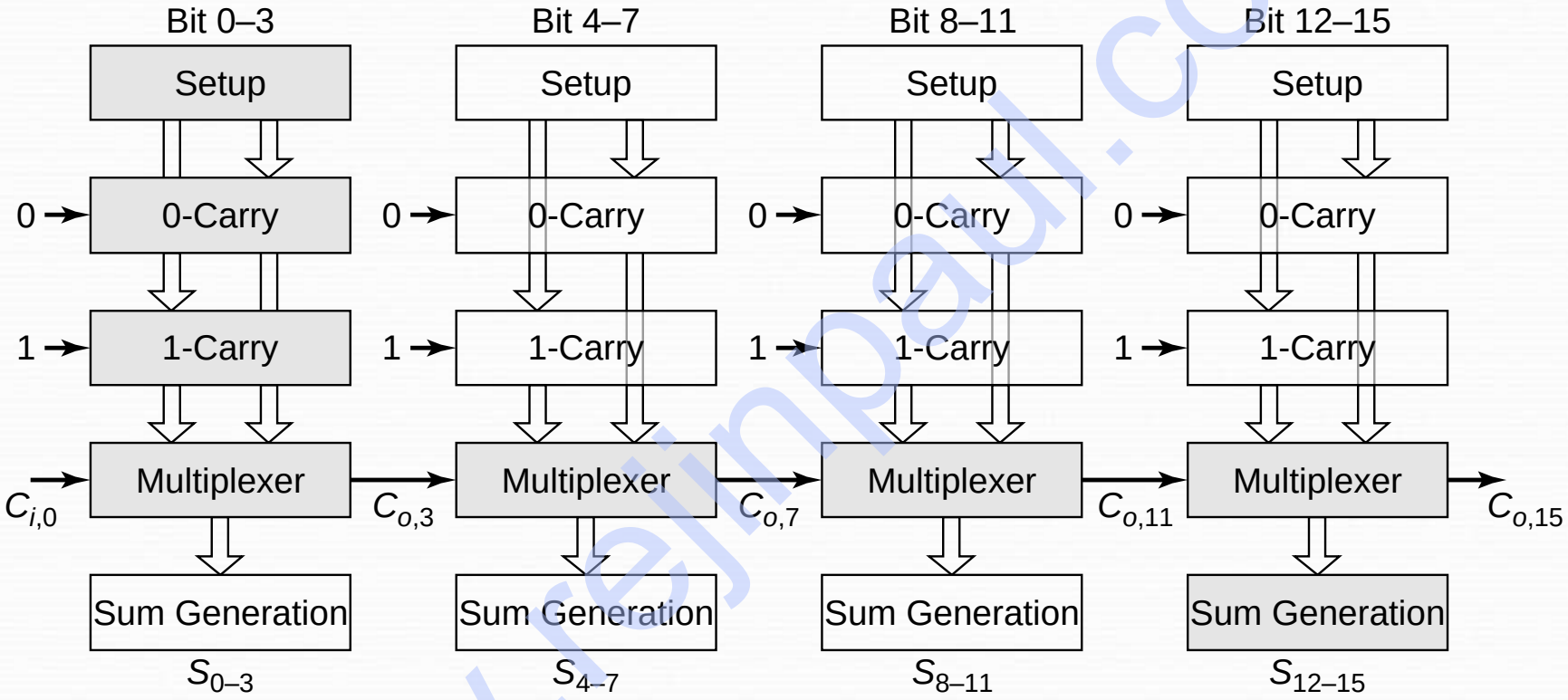
www.reinpaul.com

- In ripple carry adder, every full adder has to wait for the incoming carry.
- In this method, consider both the possible values of carry input(0 or 1)and evaluate the result for both in advance.
- Once the real value of incoming carry is known, the correct result is selected using a multiplexer stage with min delay.
- The worst case propagation delay is given as
- $t_{carry}=t_{setup}+Mt_{carry}+(N/M)t_{mux}+t_{sum}$

- N→Total number of bits
- M→Number of bits per stage respectively
- t_{carry}→delay of carry through a single full adder.
- The carry delay in a single block is proportional to the length of that stage or equals M
- The proportional delay of adder is linearly proportional to N.
- Because the block-select signal selects between 0 and 1 ripple through all stages in the worst case.

www.rejinpaul.com

16-bit Linear Carry Select Adder



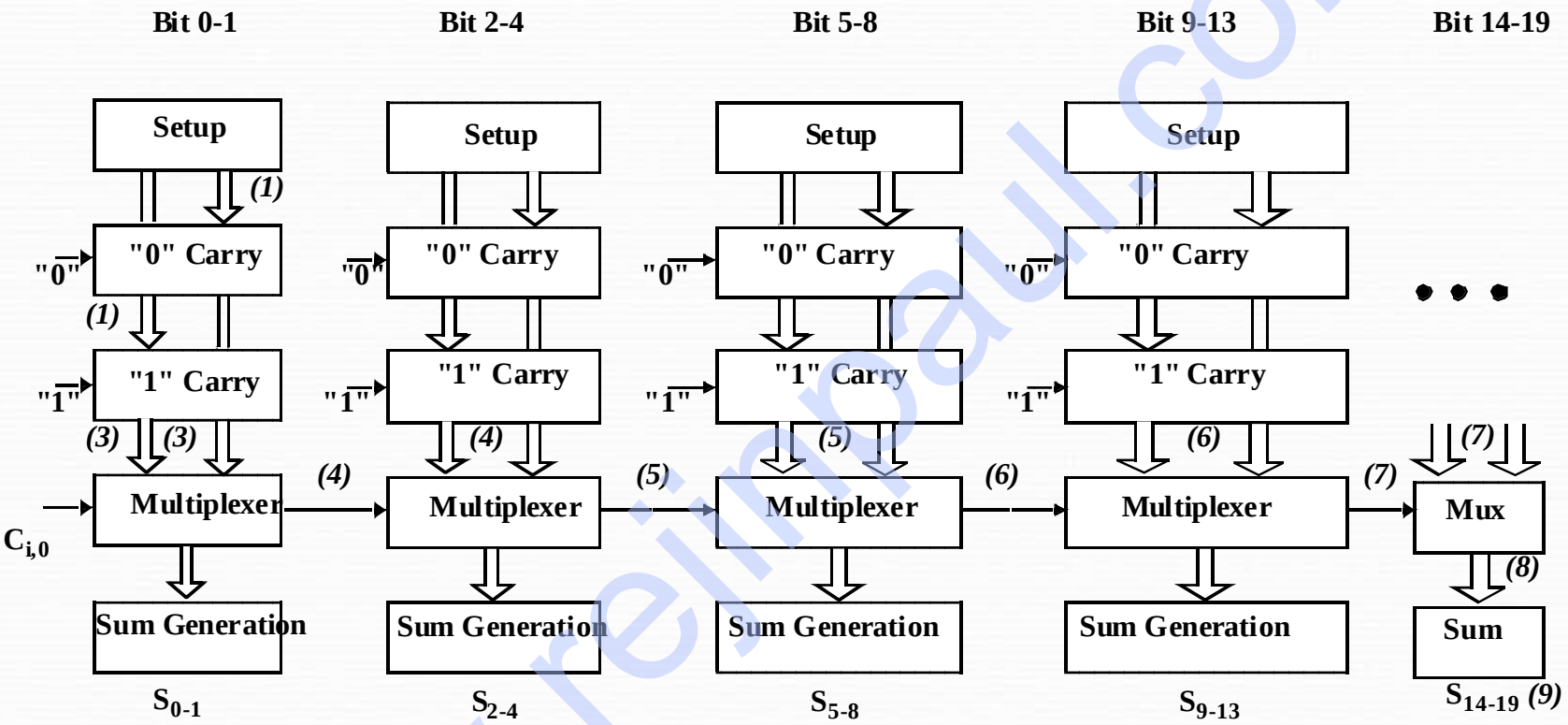
Advantages

- Less propagation delay
- Reduce the computation delay

Drawbacks→High power consumption

www.rejinpaul.com

2.3)Square Root Carry Select-Adder



www.reinpaul.com

$$N = M + (M+1) \ldots\ldots\ldots (M + P - 1)$$

$$= MP + P(P-1) = \frac{P^2}{2} + P(M - \frac{1}{2}).$$

For adders,

Here we have N-bit adder containing P-stages, first stage has M-bits, Second stage has (M+1) bits and so on.

If M<<N(generally true), then

$$N \cong \frac{P^2}{2} \text{ or } P = \sqrt{2N}$$

$$t_{add} = t_{setup} + Mt_{carry} + \sqrt{2N}\, t_{mux} + t_{sum}$$

While in Linear case we had

$$t_{add} = t_{setup} + Mt_{carry} + \frac{N}{M} t_{mux} + t_{sum}$$

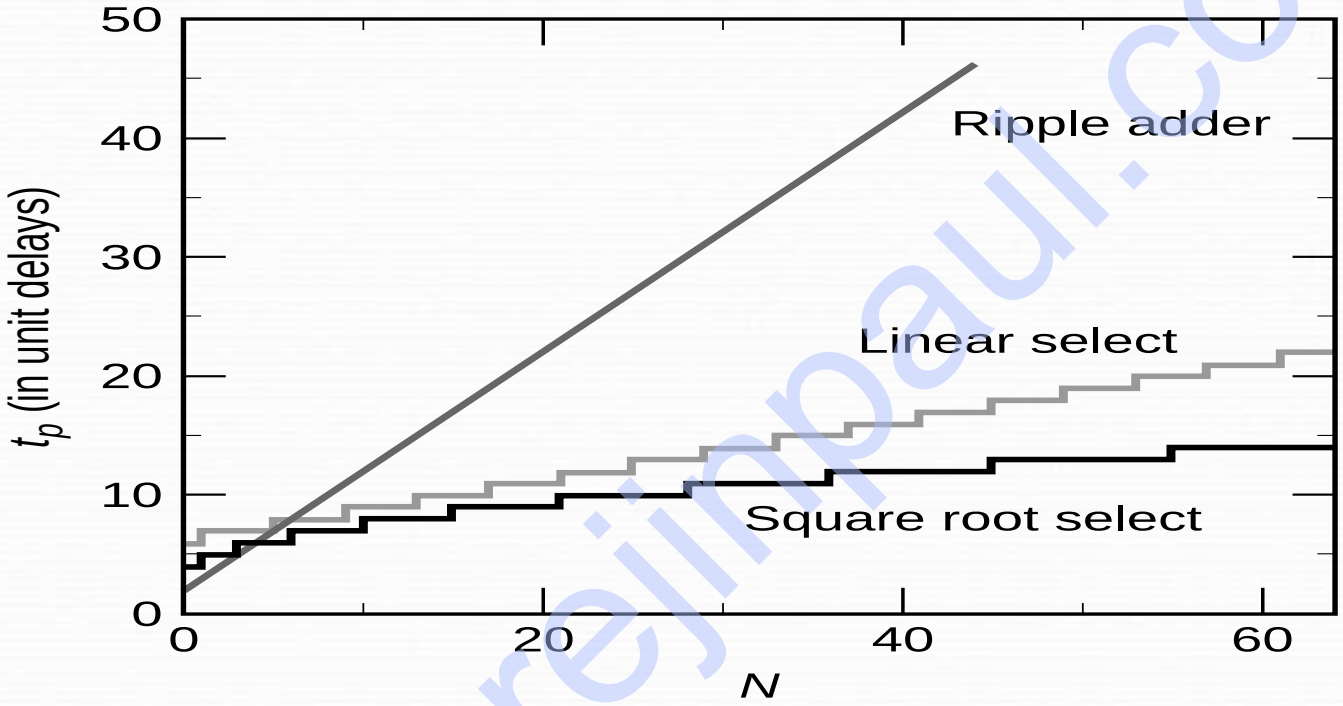
www.rejinpaul.com

•Consider the linear carry –select adder→the full adder and multiplexer have identical propagation delays with a normalized value of 1.
•The worst case arrival times of the signals at the different network nodes with respect to the time the input is applied are marked.
•For the last multiplexer, the inputs are the two-carry chains of the block multiplexer signal from the previous stage.
•The results of the carry chains are stable long before the multiplexer signal arrives .
•So the delay has to be equalized through both paths.
•To overcome this, progressively add more bits to the subsequent stages in the adder, requiring more time for the generation of the carry signals.
•The 1st stage can add 2-bits,the 2nd stage contains 3,the 3rd has 4.
•The 1st stage adds M bits,an additional bit is added to each subsequent stage.
•N=M+(M+1)+(M+2)+(M+3)+.....+(M+P-1)
•The worst case propagation delay

$$t_{add} = t_{setup} + P \cdot t_{carry} + (\sqrt{2N})t_{mux} + t_{sum}$$

www.rejinpaul.com

Adder Delays - Comparison



The delay is proportional to N for large adders.
For large value of N, t_{add} become almost a constant.

www.rejinpaul.com

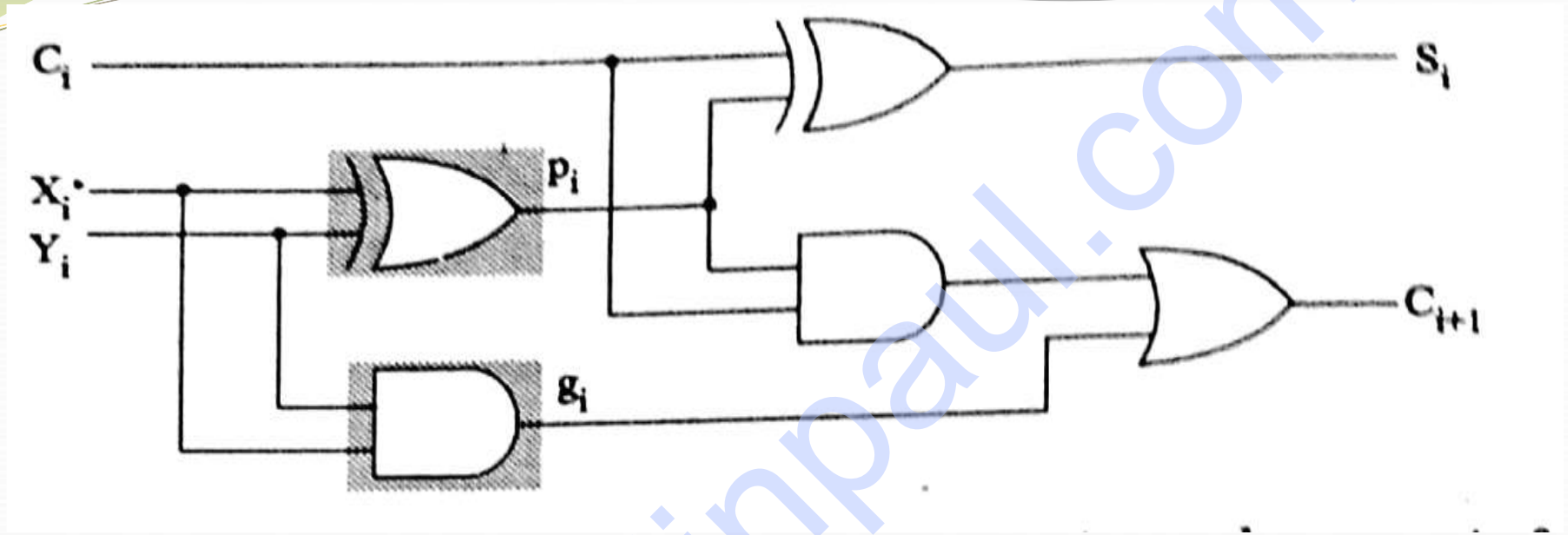
2.4)Carry Lookahead Adders

The following are the types of carry look ahead adders.

1. Monolithic Lookahead Adder
2. Logarithmic Lookahead Adder
3. Logarithmic Lookahead Adder-Alternatives

www.rejinpaul.com

Carry Look Ahead Adder



$$g_i = x_i \bullet y_i$$

$$p_i = x_i \oplus y_i$$

$$s_i = p_i \oplus c_i$$

$$c_{i+1} = g_i + p_i c_i$$

www.rejinpaul.com

As a result, the carry-in of the $(i+1)^{th}$-stage full adder can be generated by using the recursive equation of the c_{i+1}: For example, the first four carry signals are

$$c_1 = g_0 + p_0 c_0$$

$$c_2 = g_1 + p_1 c_0$$

$$= g_1 + p_1 (g_0 + p_0 c_0) = g_1 + p_1 g_0 + p_1 p_0 c_0$$

$$c_3 = g_2 + p_2 c_2$$

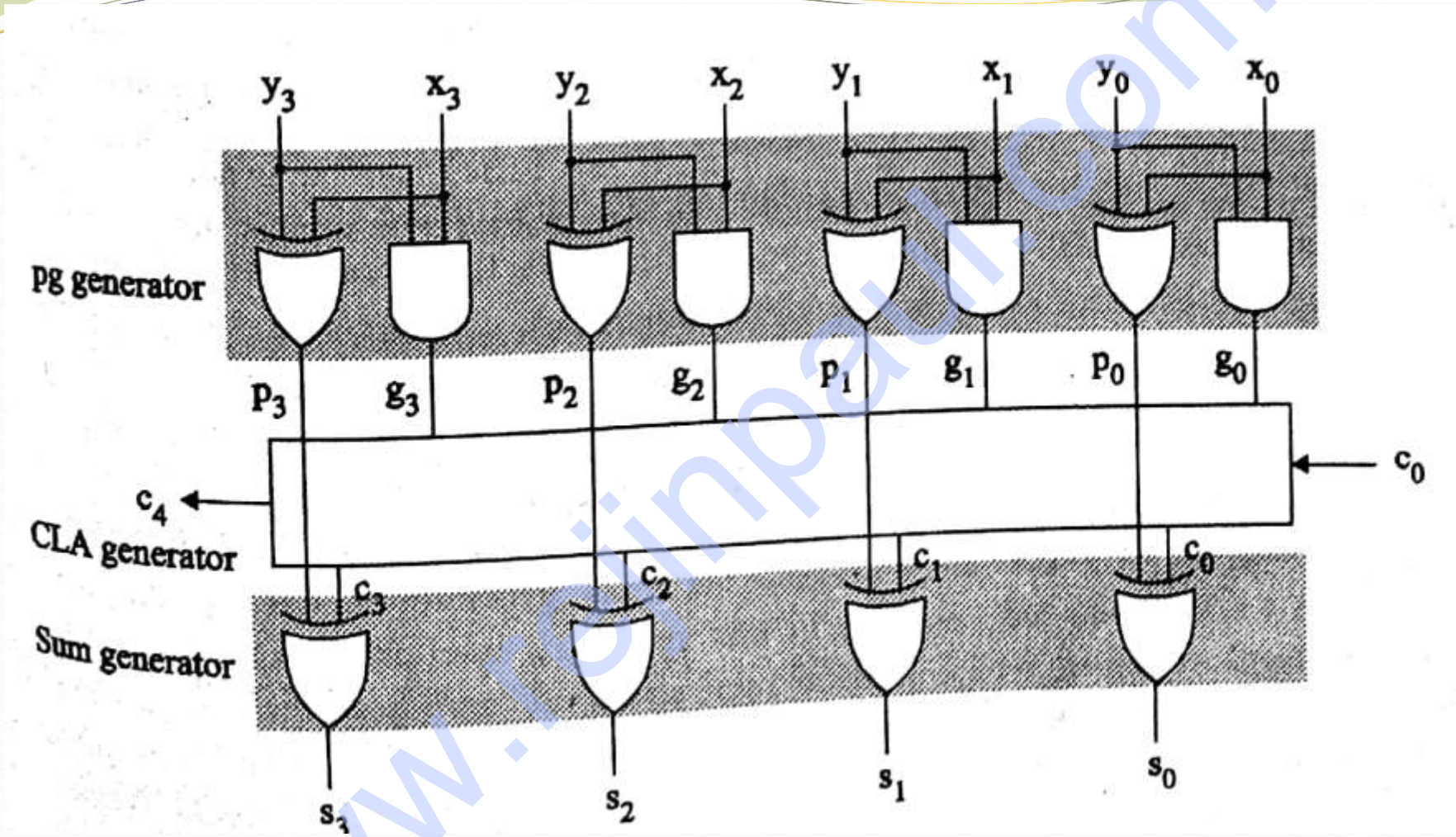
$$= g_2 + p_2 (g_1 + p_1 g_0 + p_1 p_0 c_0) = g_2 + p_2 g_1 + p_2 p_1 g_0 + p_2 p_1 p_0 c_0$$

$$c_4 = g_3 + p_3 c_3$$

$$= g_3 + p_3 (g_2 + p_2 g_1 + p_2 p_1 p_0 c_0) = g_3 + p_3 g_2$$

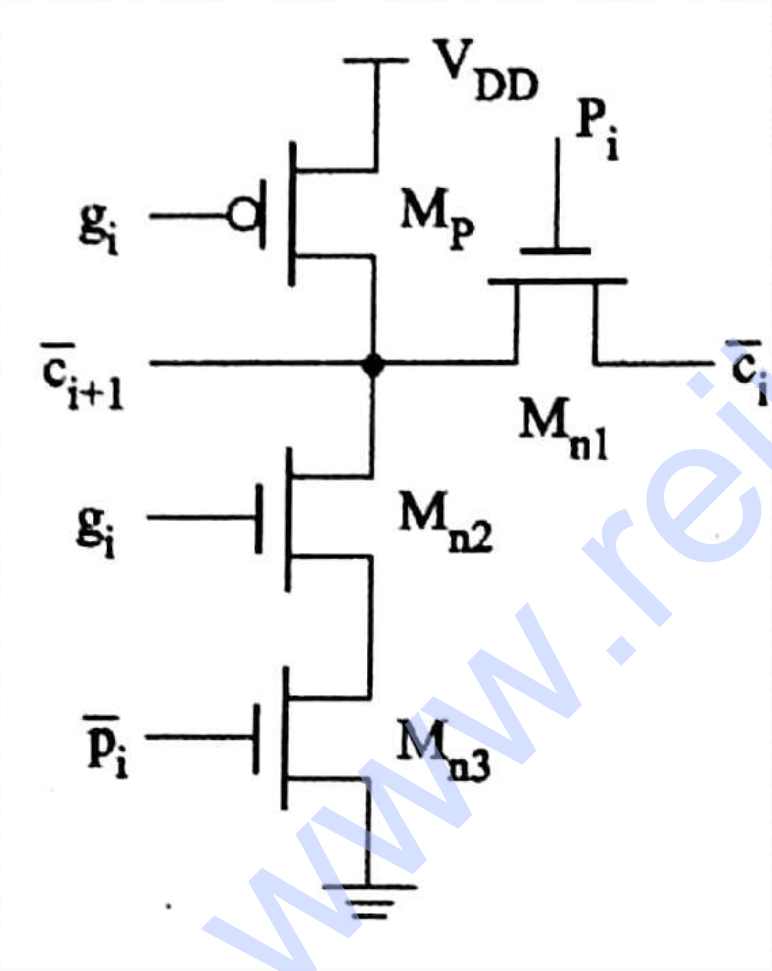
$$+ p_3 p_2 g_1 + p_3 p_2 p_1 g_0 + p_3 p_2 p_1 p_0 c_0$$

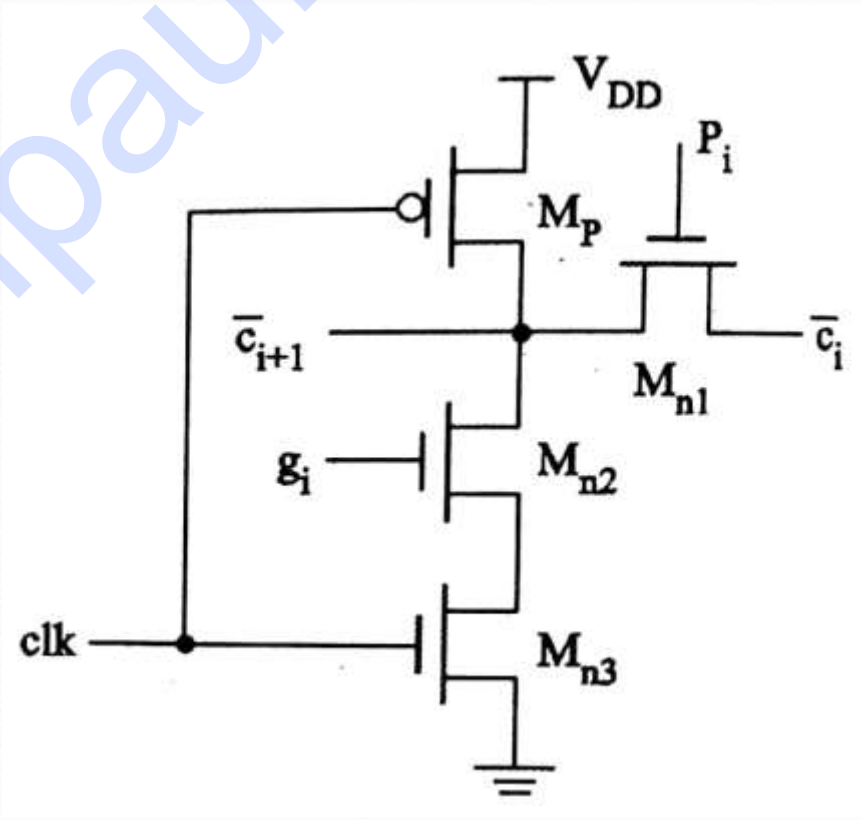
www.rejinpaul.com



www.rajinpau.com

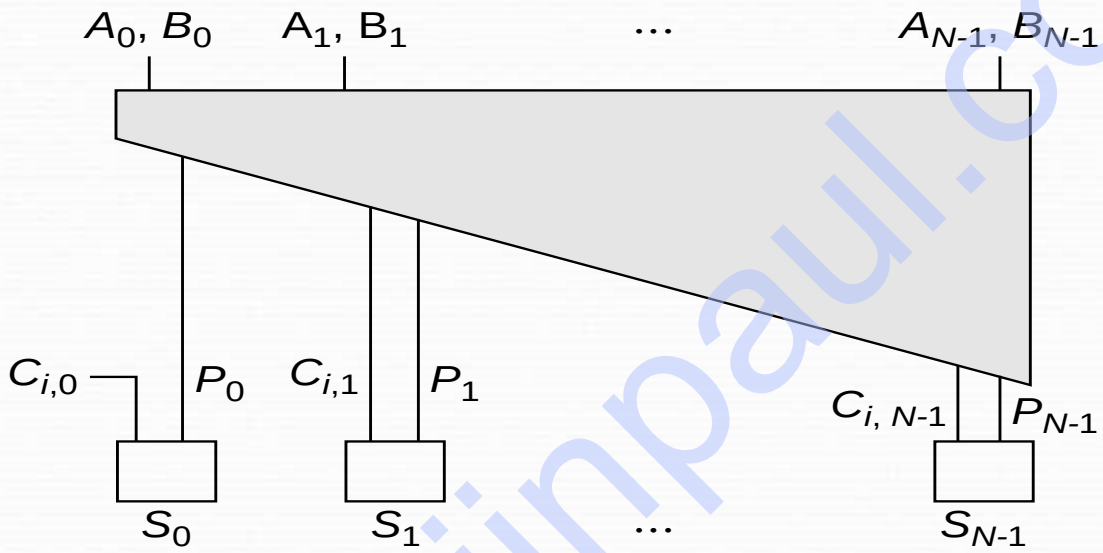
$$C_{i+1} = g_i + p_i c_i$$





www.rejinpaul.com

2.4.1)Monolithic Lookahead - Adder



For fast adder it is essential to get around the rippling effect of the carry. This effect is present in both the carry bypass and carry select adders.

$$C_{o,k} = f(A_k, B_k, C_{o,k-1}) = G_k + P_k C_{o,k-1}$$

The dependency between $C_{o,k}$ to $C_{o,k-1}$ is eliminated by expanding $C_{o,k-1}$

$$C_{o,k} = G_k + P_k(G_{k-1} + P_{k-1} C_{o,k-2})$$

For greater values of N

$$C_{o,k} = G_k + P_k(G_{k-1} + P_{k-1}(\ldots + P_1(G_0 + P_0 C_{i,0})))$$

www.rejinpaul.com

- The value Ci,0 equal to 0.
- For every bit, the carry and sum outputs are independent of the previous bits.
- Thus the ripple effect was eliminated.
- The addition time is independent of the number of bits.

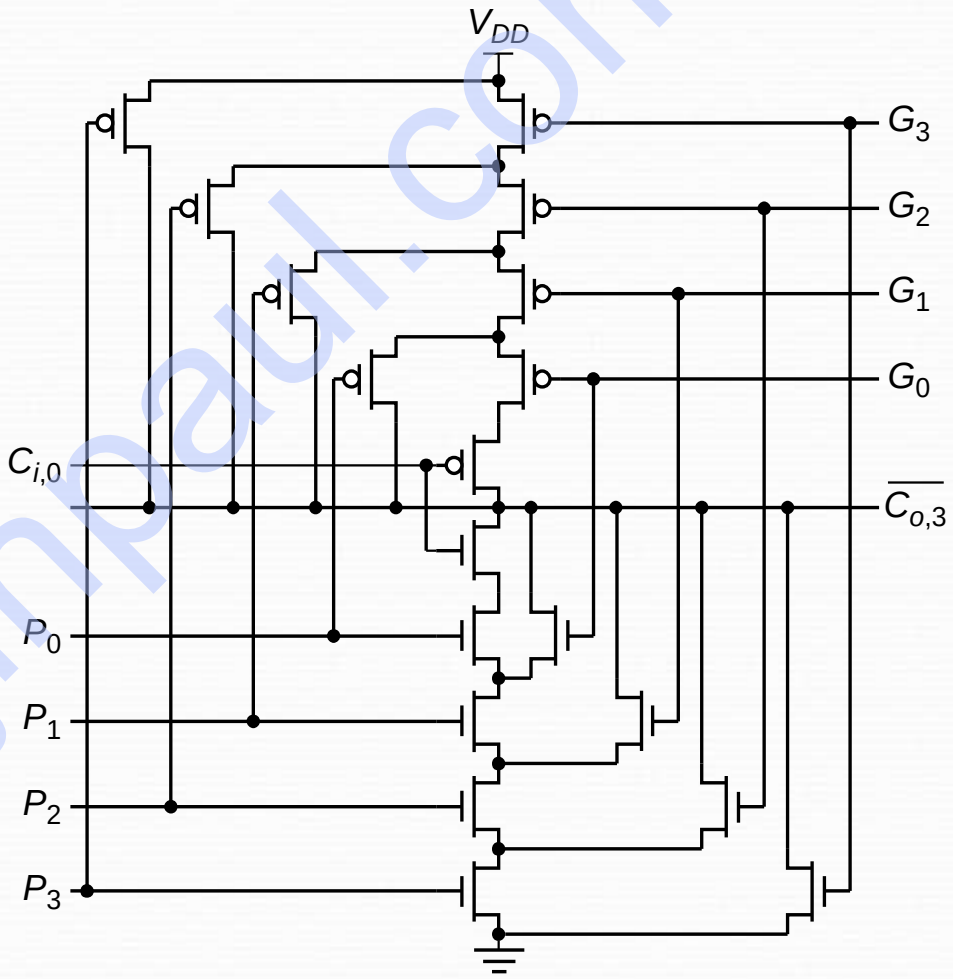
www.rejinpaul.com

Mirror Implementation of 4-bit Look-ahead Adder

- The circuit exploits self duality and the recursively of the carry look ahead adder to build a mirror structure.
- The large fan-in of the circuit makes it prohibitively slow larger values of N.
- Implementing it with smaller gates requires multiple logic levels., hence propagation delay increases.
- The fan-out on some of the signals tends to grow excessively, slowing down the adder even more.
- The area of the implementation grows progressively with N.

Drawbacks

- For Ngroups ,n+1 transistor implementation requires.
- Wide gates and large stacks tends to lead poor performance



www.rejinpaul.com

2.4.2)Logarithmic Look-Ahead Adder

•For high speed adders, the carry propagation and generation are organized into recursive trees.

•For effective implementation→decomposing the carry propagation into subgroups of Nbits.

$$C_{o,0} = G_0 + P_0 C_{i,0}$$

$$C_{o,1} = G_1 + P_1 G_0 + P_1 P_0 C_{i,0}$$

$$C_{o,2} = G_2 + P_2 G_1 + P_2 P_1 G_0 + P_2 P_1 P_0 C_{i,0}$$

$$= (G_2 + P_2 G_1) + (P_2 P_1)(G_0 + P_0 C_{i,0}) = G_{2:1} + P_{2:1} C_{o,0}$$

•Carry propagation process is decomposed into subgroups of 2 bits generate ($G_{i,j}$) and propagate ($P_{i,j}$).

•$G_{i,j}$ is true→if the groups generate carry

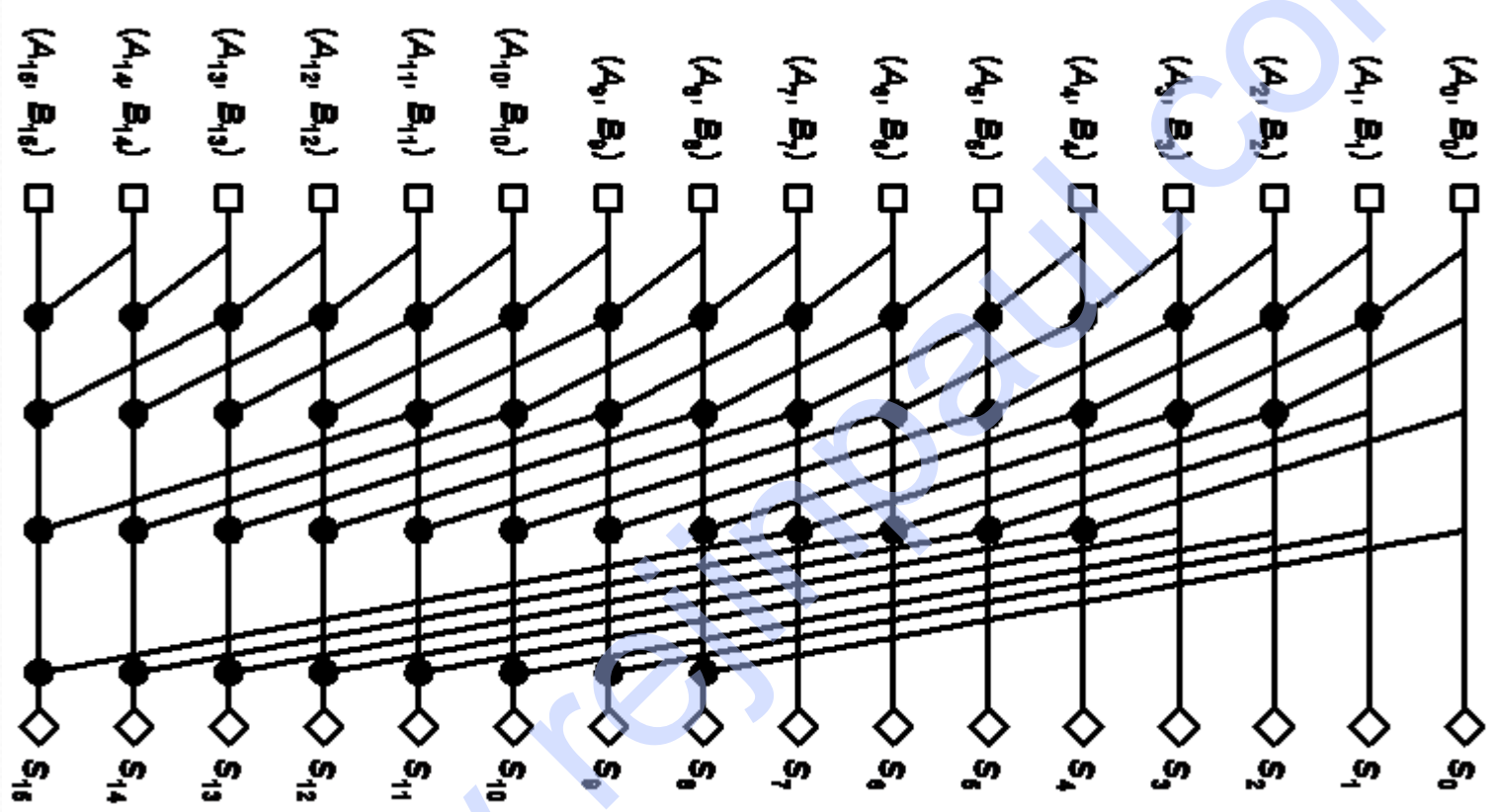
•$P_{i,j}$ is true→if an incoming carry propagates through the complete group.

$$\bullet C_{o,i} = G_i + P_i C_{i,o}$$

www.rejinpaul.com

Tree Adders

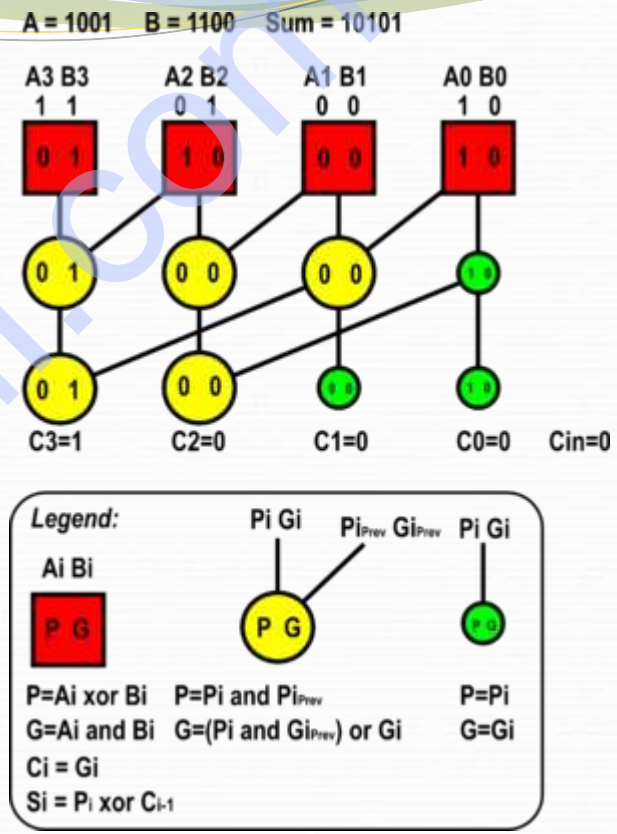
16-bit radix-2 Kogge-Stone tree



The carry at position 15 is computed by combining the results of blocks(0:7) and(8:15).
Each of these in turn is composed hierarchically.
For instance,(4:7) is the composition of (0:3)..
Also (8:15) is the composition of (4:11)..
Also (0:3) consits of (0:1) and (2:3)etc..,

www.resumepal.com

Each vertical stage produces a "propagate" and a "generate" bit, as shown.
The culminating generate bits (the carries) are produced in the last stage (vertically), and these bits are XOR'd with the initial propagate after the input (the red boxes) to produce the sum bits.



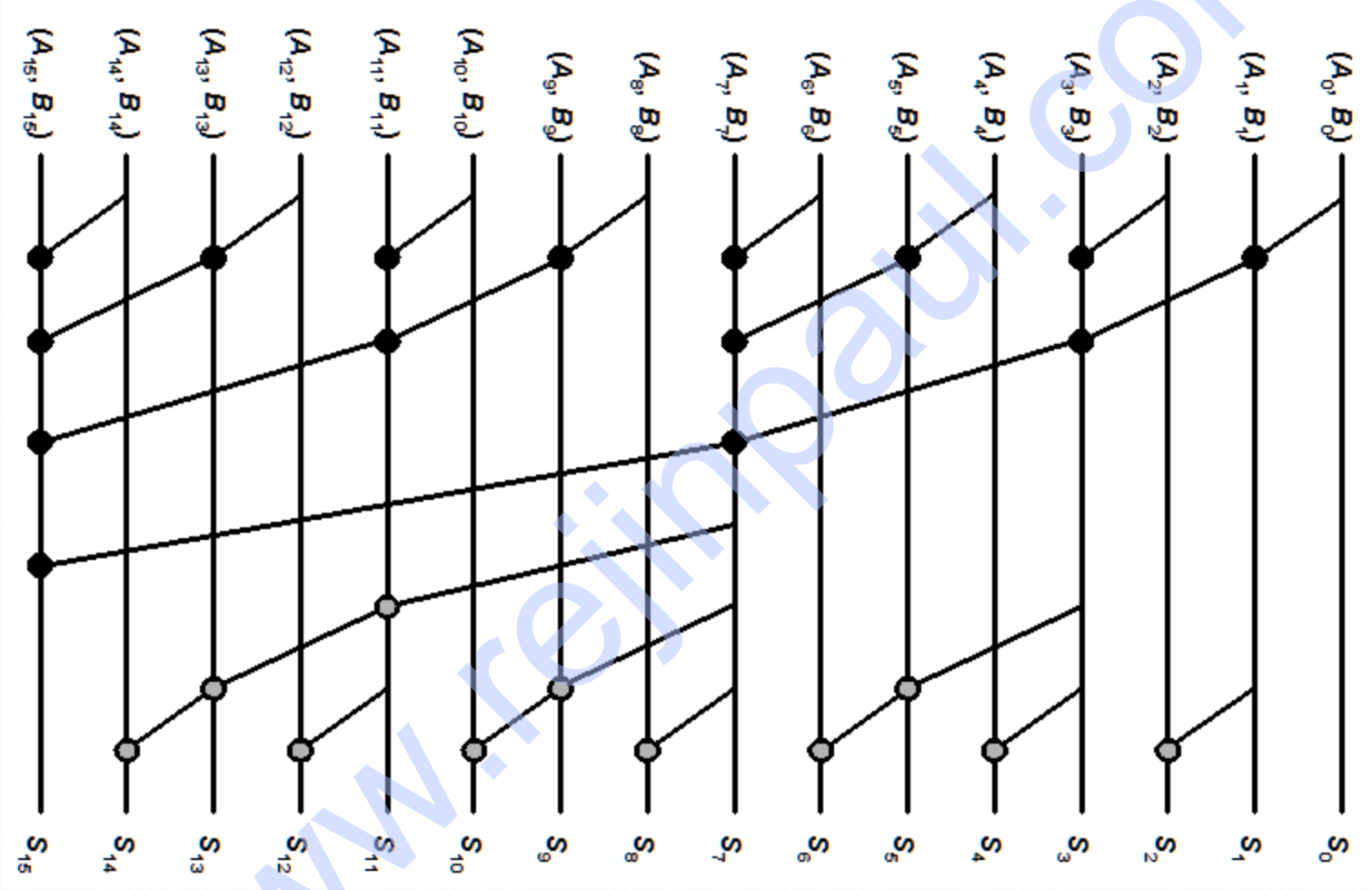
E.g., the first (least-significant) sum bit is calculated by XORing the propagate in the farthest-right red box (a "1") with the carry-in (a "0"), producing a "1".

The second bit is calculated by XORing the propagate in second box from the right (a "0") with Co (a "0"), producing a "0

```
S0 = (A0 XOR B0) XOR Cin
 S1 = (A1 XOR B1) XOR ((A0 AND B0) OR (A0 XOR B0) AND Cin)
S2 = (A2 XOR B2) XOR (((A1 XOR B1) AND ((A0 AND B0) OR (A0 XOR B0)
AND Cin)) OR (A1 AND B1))
S3 = (A3 XOR B3) XOR ((((A2 XOR B2) AND (A1 XOR B1)) AND ((A0 AND
B0) OR (A0 XOR B0) AND Cin)) OR (((A2 XOR B2) AND (A1 AND B1)) OR
(A2 AND B2)))
S4 = (A3 AND B3) OR (A3 XOR B3) AND ((((A2 XOR B2) AND (A1 XOR B1))
AND ((A0 AND B0) OR (A0 XOR B0) AND Cin)) OR (((A2 XOR B2) AND (A1
AND B1)) OR (A2 AND B2)))
```


www.rejinpaul.com

2.4.3) Logarithmic Lookahead Adder-Alternatives



Brent-Kung Tree

www.reinpaul.com

•The forward binary tree structure is not enough to generate the complete set of carry bits.
•An inverse binary tree is needed to realize the other carry bits.
•This structures combines intermediate results to produce the remaining carry bits.
•The adder uses 27 dot gates or half of the 49 needed for a full radix-2 and needs a few wires.
•The wiring structure is less regular .
•Fan out wires from gate to gate making performance optimization more difficult.
•Fan-out of the middle node($C_{0,7}$) has 1sum and 5 dot operations.

Drawbacks

•Not suited for very large adders(>32 adders)

www.rejinpaul.com

The entire architecture is easily understood by dividing the system into three separate stages:

Generate/Propagate Generation

If the inputs to the adder are given by the signals A and B, then the generate and propagate signals are obtained according to the following equations. G = A.B

P = AÅ. B

The Dot (·) Operation =

The basic inputs to this structure are the generate and propagate signals generated in the previous stage.

The · operator is a function that takes in two sets of inputs-- (g, p) and (g', p')-- and generates a set of outputs-- (g + pg', pp').

These building blocks are used for the generation of the carry signals in the structure.

For the generation of the carry signals, the carry for the kth bit from the carry look-ahead concept is given by

Co,k = Gk +Pk(Gk-1 +Pk-1 +P k-1 (…+P1(G0 +P0 Ci,0)))

carry signals as

Co,0 = G0 +P Ci,0 = a (G0,P0) Co,1 = G1 + G0 P1 = a ((G1 , P1)·(G0, P0))

…… C0,k = a ((Gk,Pk)·(Gk-1,Pk-1)·…·(G0,P0))

Sum generation

S= AÅ BÅ C where A and B are the input signals, and C is the carry signal.

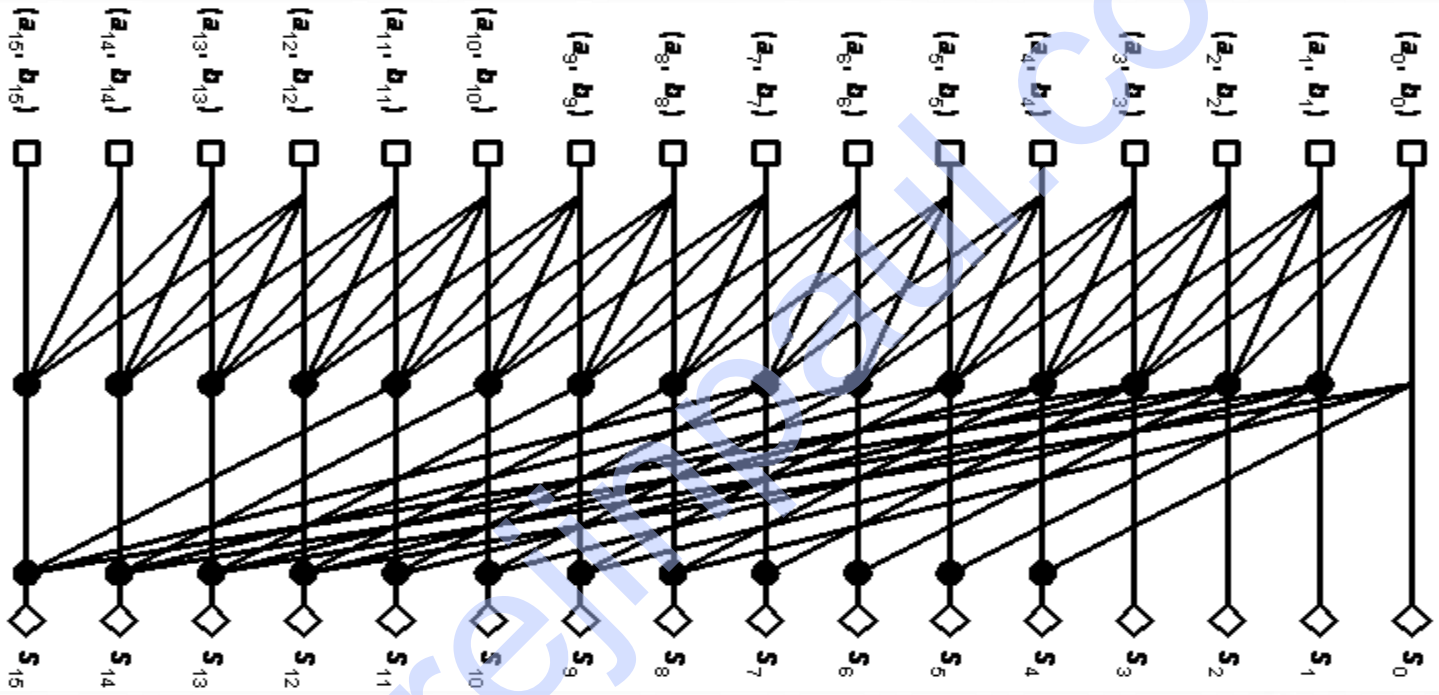
The carry is obtained from the dot operator stage

www.rejinpaul.com

Tree Adders

16-bit radix-4 Kogge-Stone Tree

4:7→0-3
8:11→0-7
12:15→8-11



- It uses building blocks of order 4.
- 16 bit addition needs only 2 stages of carry logic.
- Lookahead adder much larger than a ripple adder.
- Adder is constructed with a propagation delay that is independent of the no of bits.
- Adders requires number of encoding-decoding steps.
- It is embedded in larger structures such as multipliers or high signal processors.

www.rejinpaul.com

2) ACCUMULATORS

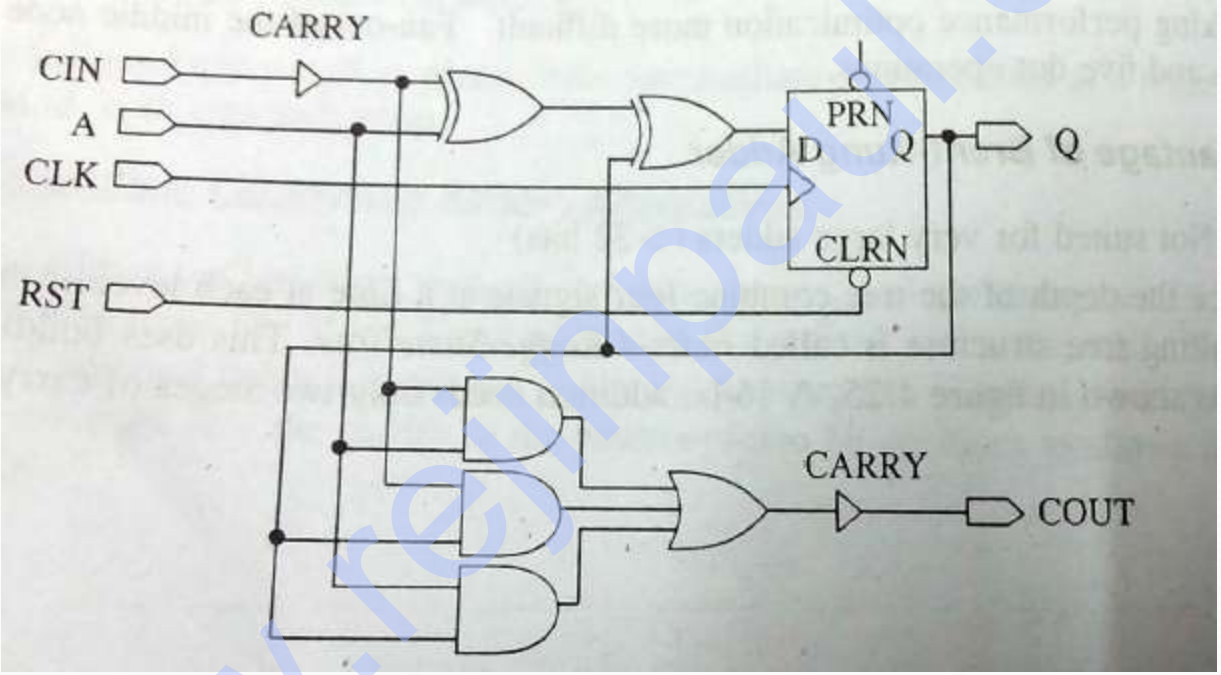
- Accumulator is an adder/subtractor and a register.
- These are combined as a multiplier-accumulator(MAC)

Types

- Single bit Accumulator
- n-bit Accumulator

www.rejinpaul.com

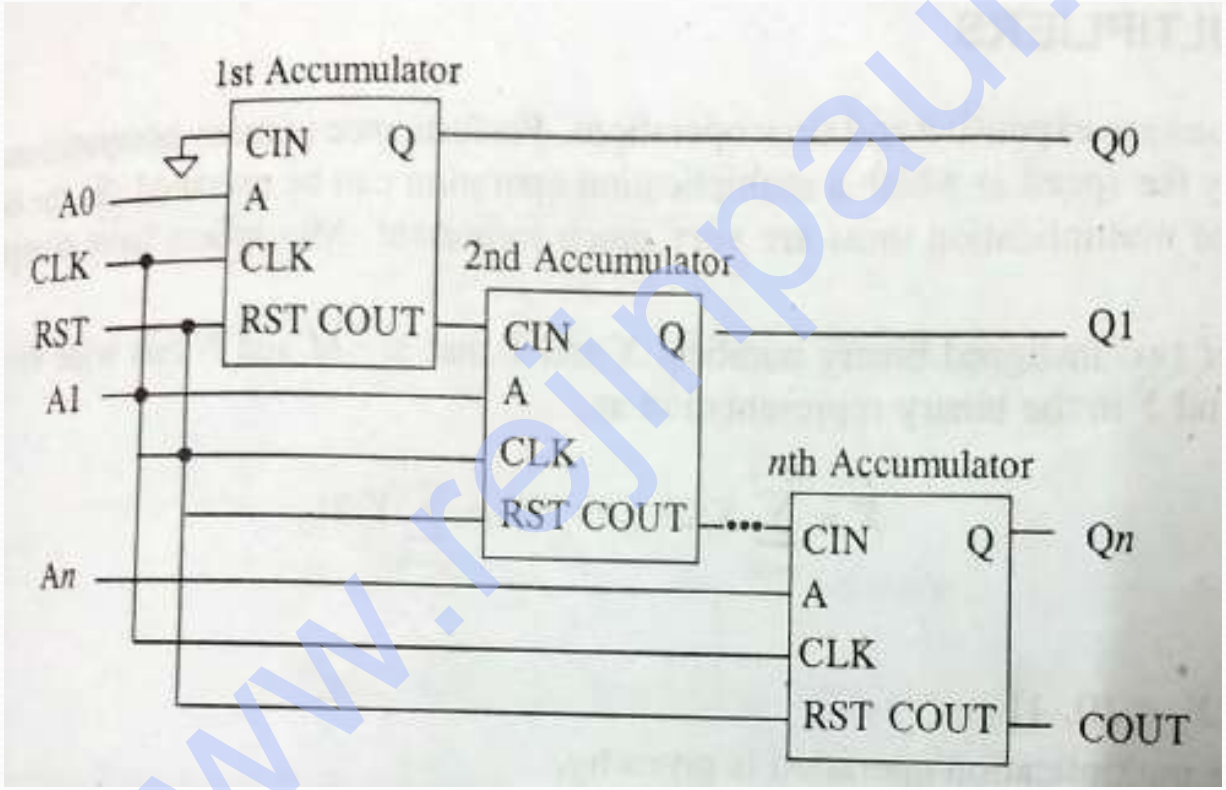
1)Single bit Accumulator



www.reinpaul.com

2) n-bit bit Accumulator

- n-bit accumulator is implemented by cascading the carry out of the n-1 stage to the carry in of the ^{nth} stage to end the single bit accumulator.
-

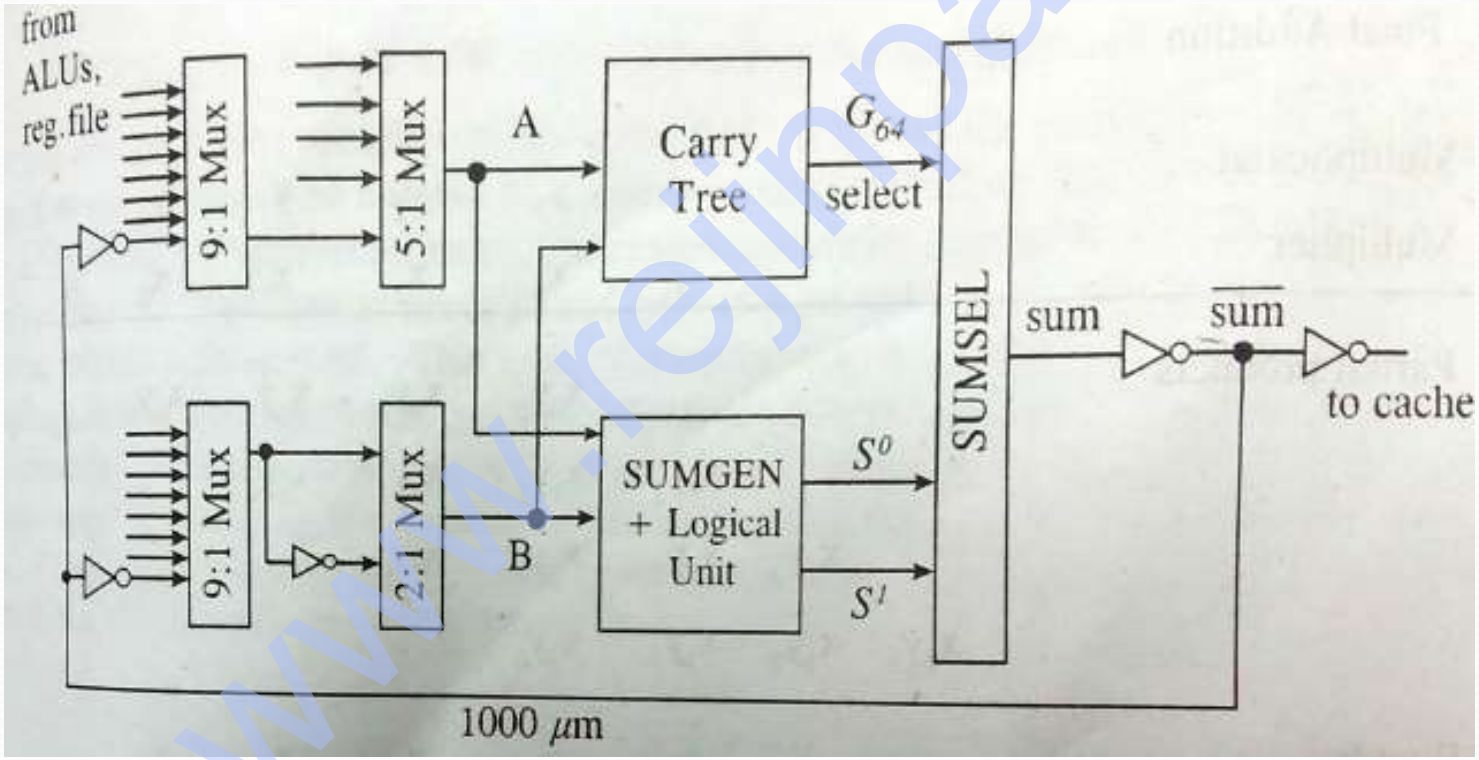


www.rejinpaul.com

Arithmetic Logic unit

- It is a core part of the any microprocessor and microcontroller.
- ALU combines addition, subtraction and other operations such as shift and bitwise logic operations(AND,OR,XOR).

64-bit ALU for High end processor



www.rejinpaul.com

- It consists of two level of wide multiplexers,64-bit adder ,logic unit, an operator merging multiplexer and write bus driver.
- Multiplexer(9:1)→selects 1 from 9 different input sources,3 of which are from register files and caches.
- The other6 are bypass paths from the 6 ALU's .
- One of bypass loops feeds the output of the ALU back to its own input, creating a single cycle loop to define the critical path.
- Multiplexer(5:1)→performs a partial shifting of the operand.
- Multiplexer(2:1)→ inverts the operand to implement subtraction.
- Adder→executes 2's complement addition or subtraction.
- Sum Selection Block→ selects sum selection and also merges the outputs of arithematic and logic units.
- Buffer→ drives the loop-back bus, which presents a large load to ALU.

www.rejinpaul.com

3)Multipliers

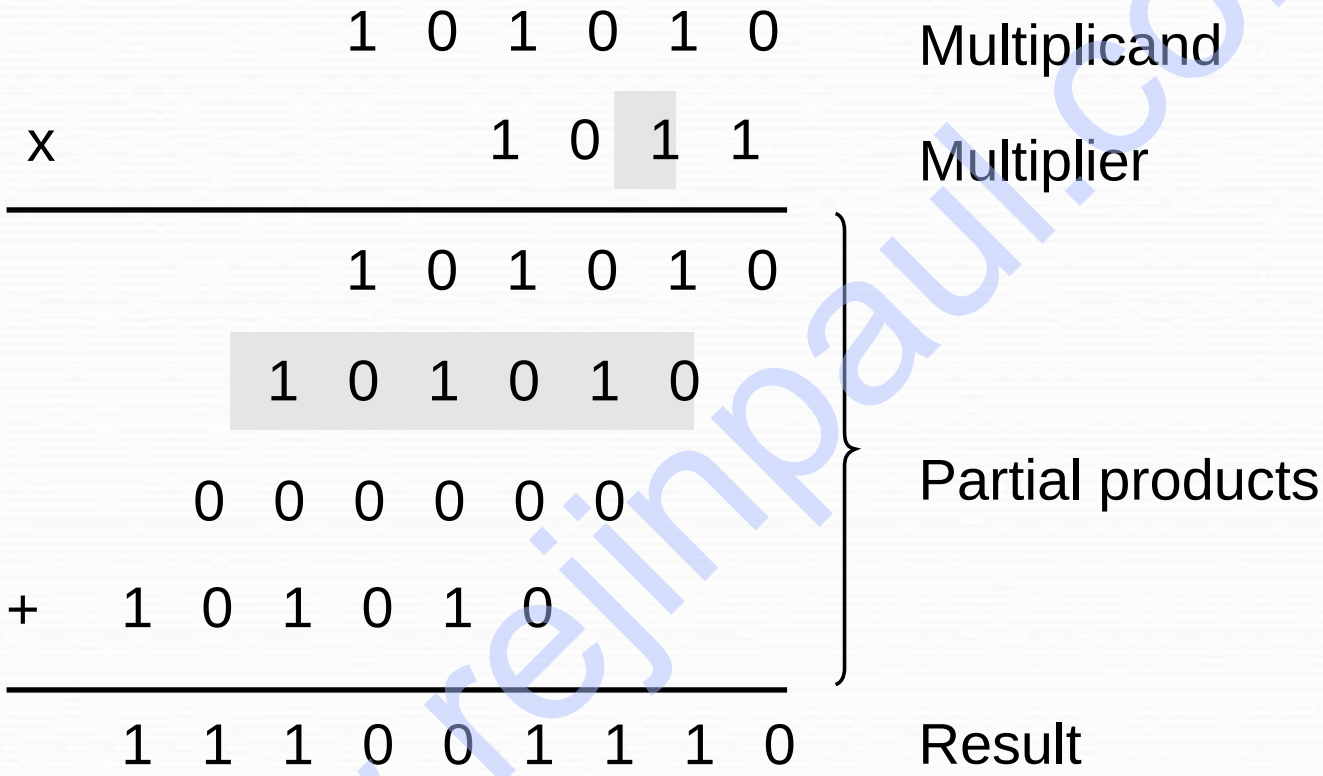
- Multiplications are expensive and slow operations.
- Performance of many computational circuits decided by the speed at which a multiplication can be executed.
- It has many complex adder arrays.
- Consider 2 unsigned binary numbers X and Y that are N bits size,the Multiplication operations is given by

$$Z = X \times Y = \left(\sum_{i=0}^{M-1} X_i 2^i\right)\left(\sum_{j=0}^{N-1} Y_j 2^j\right) = \sum_{k=0}^{M+N-1} Z_k 2^k$$

- 2^i→bit position
- Multiplication is performed by using a single 2-I/P adder.
- Multiplication takes M cycles using N-bit adder.
- The shift and add algorithm for Multiplication adds together M partial products.
- Each partial product is generated by multiplying the multiplicand with a bit of the multiplier (AND) then by shifting the result based on multiplier bits position.

www.rejinpaul.com

Example-Binary Multiplication array structure



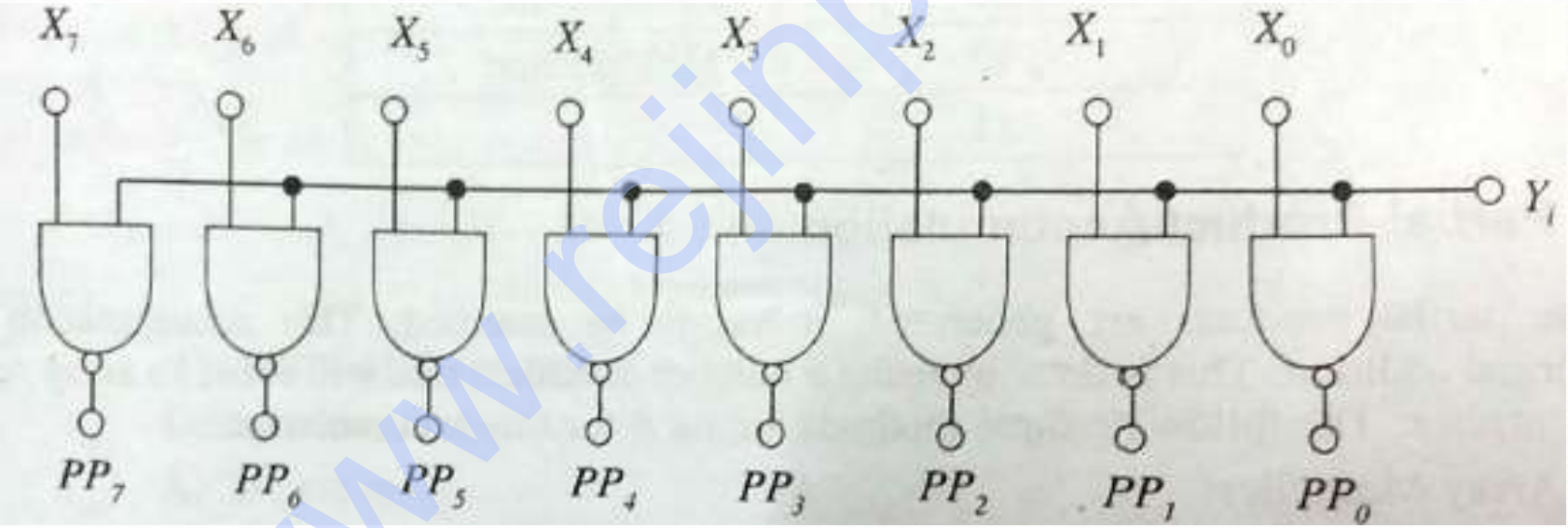
It has the following 3 functions

1. Partial-product Generation
2. Partial-product Accumulation
3. Final Addition

www.rejinpaul.com

3.1) Partial-product Generation

- Partial products result from the logical AND of multiplicand X with a multiplier bit Yi.
- Each row in the parallel –product array is either a copy of the multiplicand or row of zeros.
- Partial-product generation optimization can lead to reduced delay and area.
- But it has many o's that have no impact on the result.
- If multiplier consists of all 1's ,all the partial product exists.
- If multiplier consists of all o's ,all the partial product does not exists.
- This reduces the number of generated partial product by half.



www.rejinpaul.com

Example→ 8-Bit Multiplier

- Consider 8-bit multiplier of the form 011111o,which produces 6 non zeros partial product rows.
- Reduce the number of nonzero rows by reducing this number $(2^7+2^6+2^5+2^4+2^3+2^2)$ into a different format.
- Let 1 is a short notation of for -1.
- This format needs to add only two partial products, but the final adder has to perform subtraction.
- This type of transformation is called Booth's recording.
- Booth's recording reduces the number of partial products to at most one half.
- This ensures that for every two consecutives bits,at most one bit 1 or -1.
- Reducing the number of partial products is equal to reducing the number of additions ,that speedup the operation and reduces area.
- This transformation format is given by
- $$Y = \sum_{j=0}^{N-1} Y_j 4^j$$ with Y=(-2,1,0,1,2)

www.rejinpaul.com

- In modified Booth recording the multiplier is partitioned into 3-bit groups that overlap by one bit.
- The number of partial products equals half of the multiplier width.
- The input bits to the recording process are the two current bits, combined with the upper bit from the next group, moving from MSB to LSB

Multiplier Bits	Recoded Bits
000	0
001	+ Multiplicand
010	+ Multiplicand
011	+ 2 × Multiplicand
100	– 2 × Multiplicand
101	– Multiplicand
110	– Multiplicand
111	0

00→0
10→-1
01→+1
11→0

1st row
000→00→$2^0*0+2^1*0=0$

www.rejinpaul.com

3.2) Partial-product Accumulation

- Once the partial product are generated, it has to be summed.
- This accumulation is a multioperand addition
- This is done by number adders that will form an array →Array multiplier.
- The following 3 methods are used for accumulation

1. Array Multiplier
2. Carry Save Multiplier
3. Tree Multiplier

www.rejinpaul.com

3.2.1) Array Multiplier

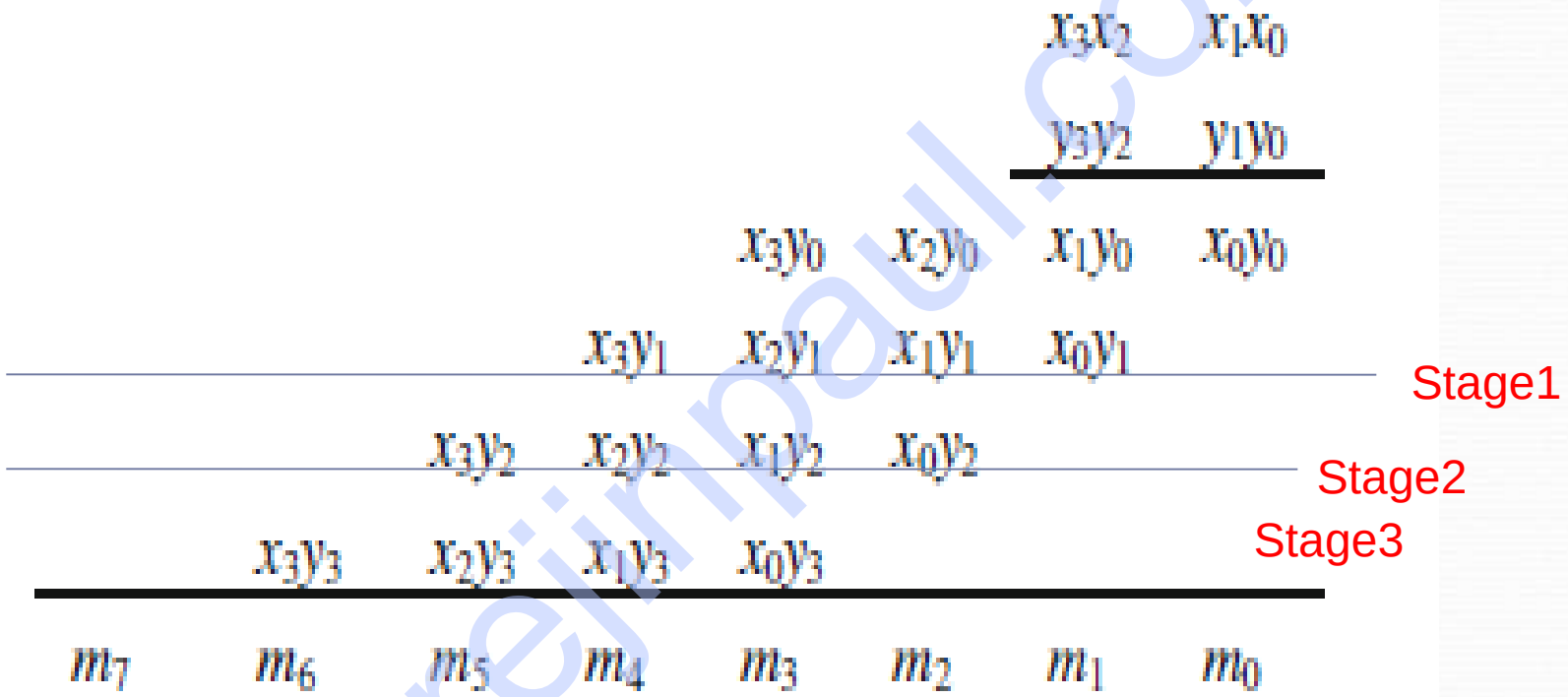
- Generation of N partial products requires NXM 2-bit AND gates.
- Multipliers most area is utilized to add the N partial products.
- This requires N-1 M-bit address.
- Shifting of the partial products for proper alignment is performed by simple routing and does nor require any logic.
- The overall structure compacted into a rectangle to get a very efficient layout.
- Propagation delay is given by

$$t_{mult} \approx [(M-1)+(N-2)]t_{carry}+(N-1)t_{sum}+(N-1)t_{and}$$

- t_{sum} → delay between the input carry and sum bit of the full adder
- t_{carry} → propagation delay between input and output carry.
- t_{and} → AND gate delay
- All critical paths have same length.
- Speeding up one of them by replacing one adder by a fast one such as a carry-select adder does not makes much sense.

www.rejinpaul.com

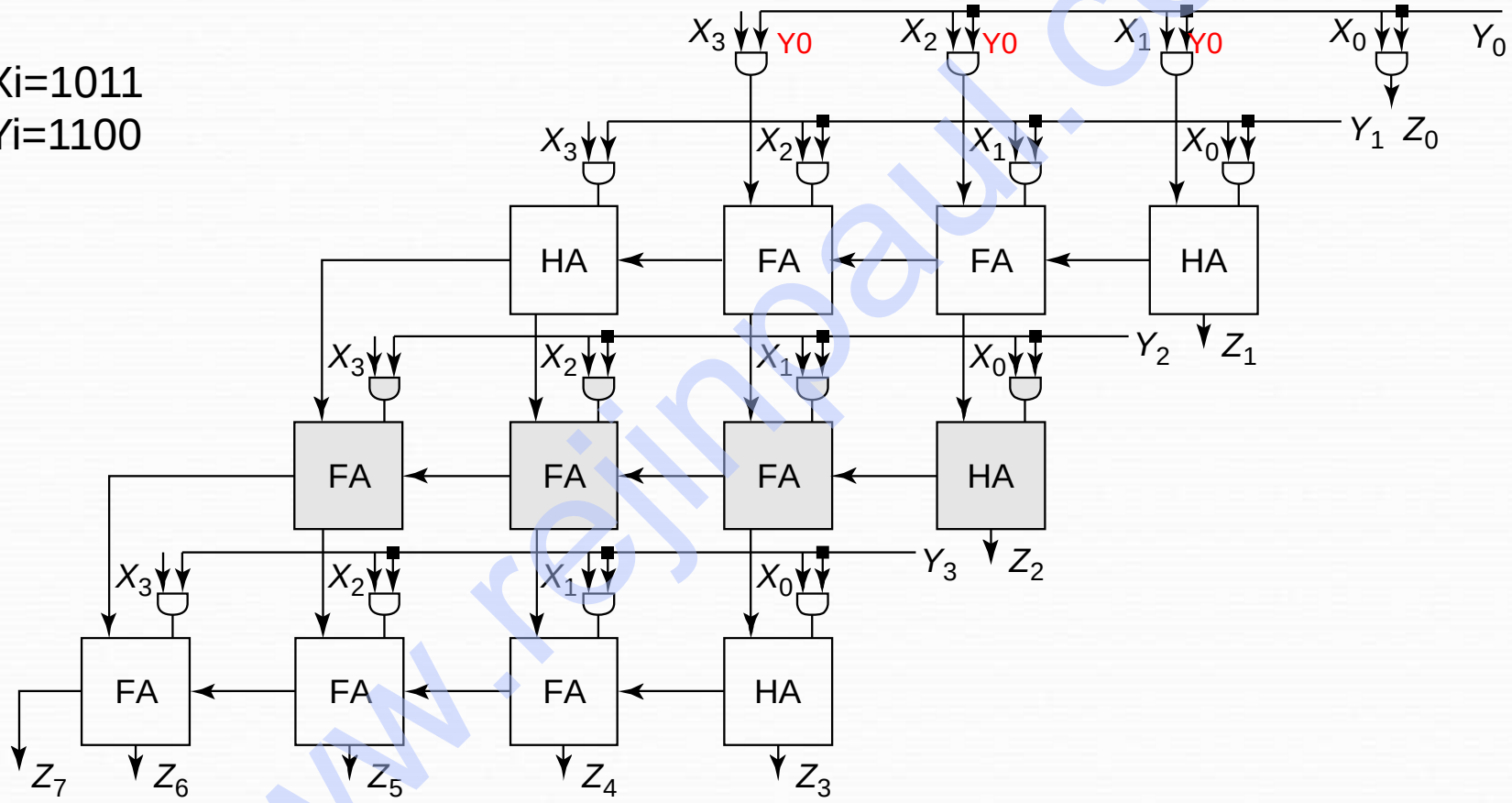
Array Multiplier(4x4)



www.rejinpaul.com

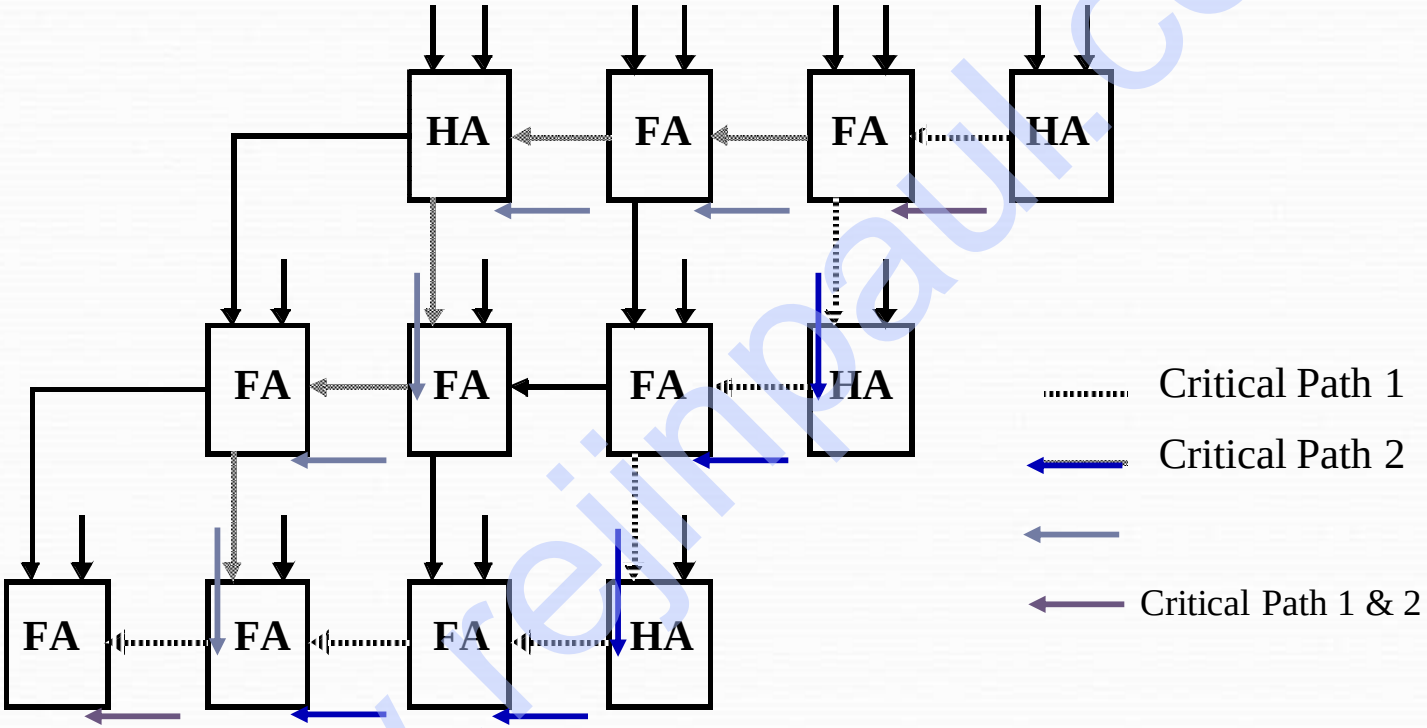
Array Multiplier(4x4)

Xi=1011
Yi=1100



www.rejinpaul.com

Ripple carry Based 4x4 Multiplier



www.rejinpaul.com

3.2.2) Carry Save Multiplier

- An extra adder called vector-merging adder is added to generate the final result.
- The resulting multiplier is called a carry-save multiplier.
- In this carry bits are not immediately added but are saved for the next adder stage.
- In the final stage, carry and sum are merged using a carry look ahead adder.
- The worst case critical path is given by

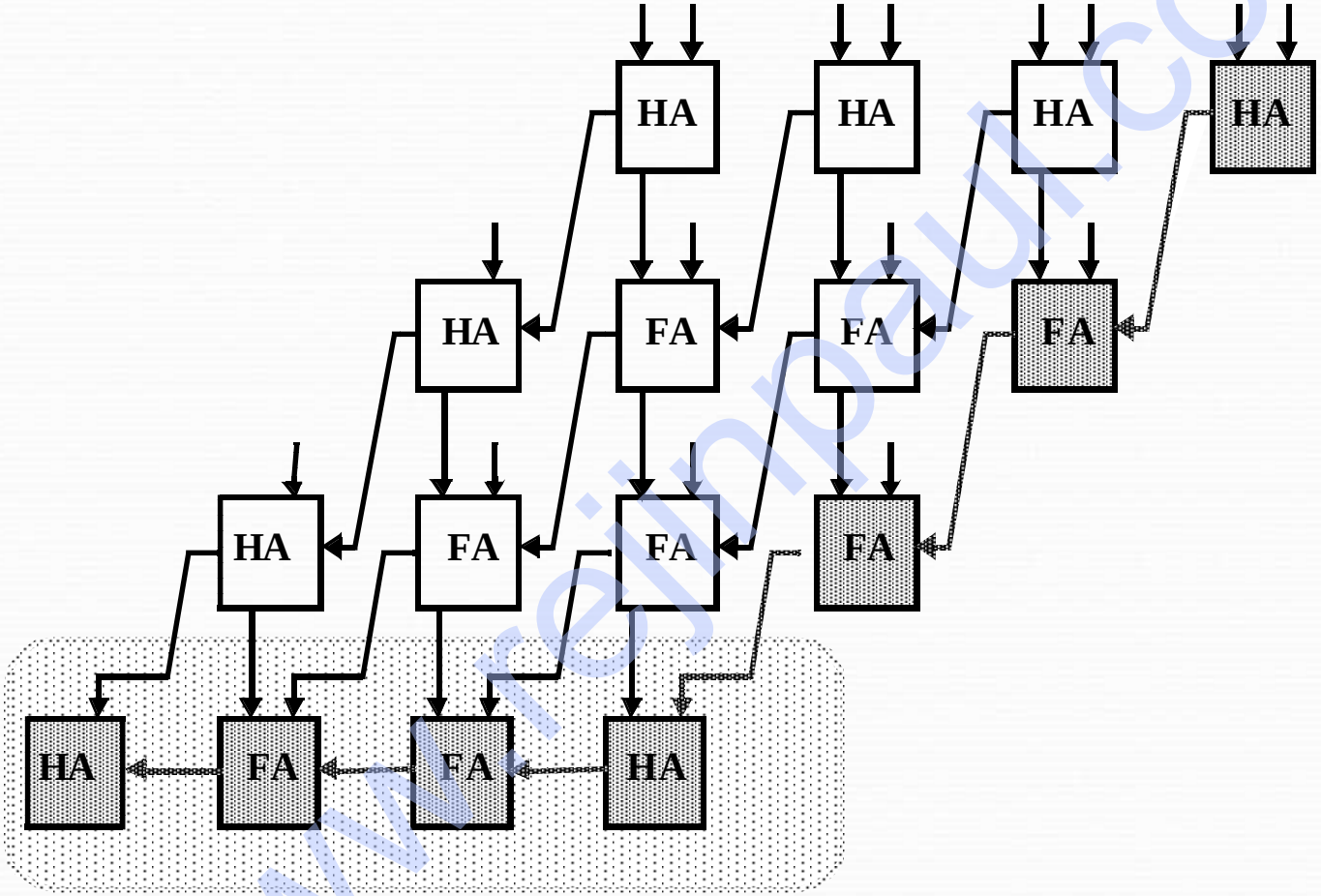
$$t_{mult} = (N-1)t_{carry} + (N-1)t_{and} + t_{merge}]$$

Advantages

- Shorter worst case critical path
- Drawbacks
- Increased area cost (because of extra adder)

www.rejinpaul.com

Carry-Save Multiplier



Vector Merging Adder

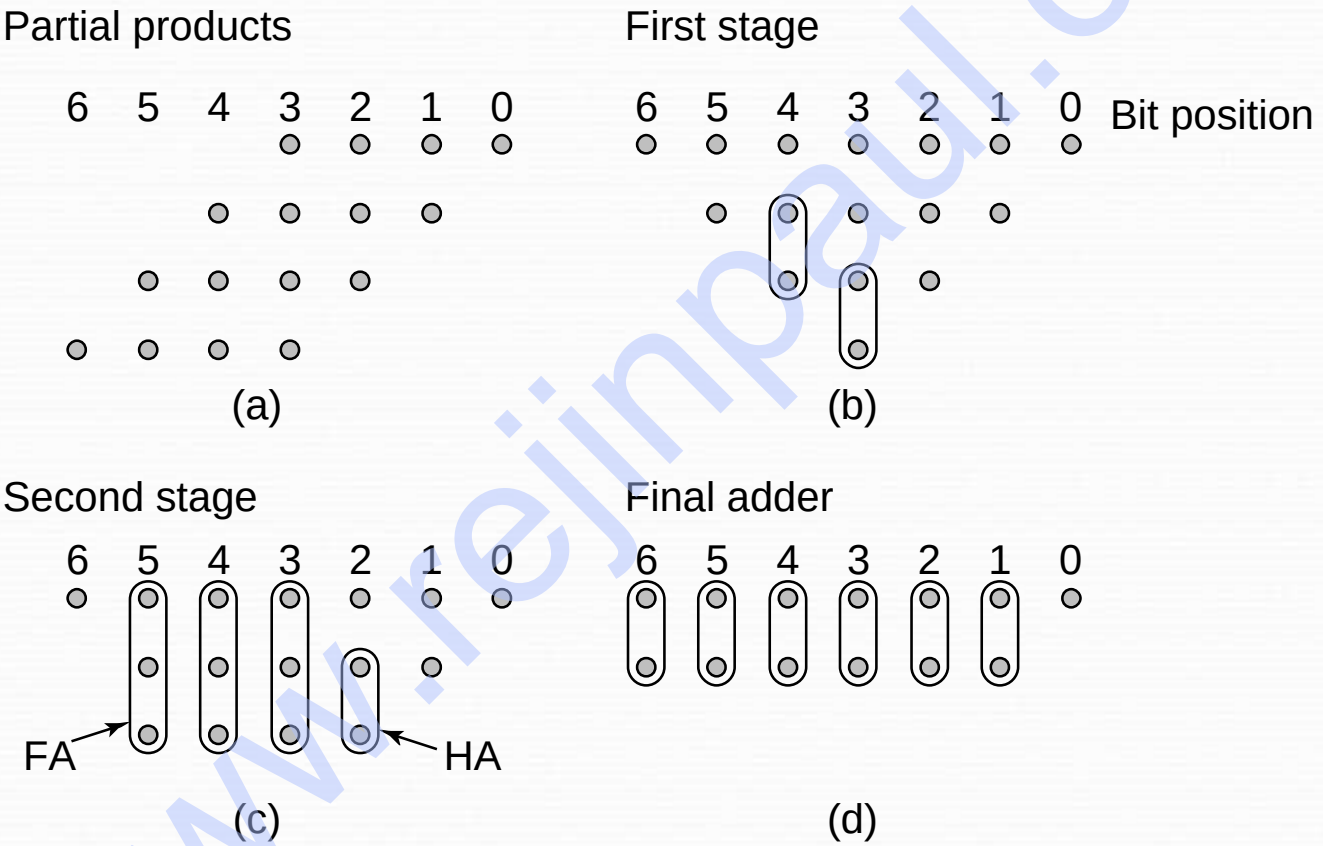
www.rejinpaul.com

3.2.3) Tree Multiplier

- Partial sum adders are arranged tree structure to reduce both the critical path and the number of adders cells.
- exmaple→4 partial products
- The first type of operator used to cover the array is a full adder ,which takes 3 inputs and produces 2 outputs.
- The sum located in the same column and the carry located in the next one.
- Due this ,the full adder is called a 3-2 compressor. it is denoted by a circle covering 3bits.
- The other operator is a half adder which takes 2 inputs bits in a column an produces 2 outputs. it is denoted by a circle covering 2bits.
- To obtain minimal implementation, the tree is covered with FA and HA starting from its densest part.
- Only 3FA&3HA requires compare to carry save multiplier requires 6FA& 6HA.

www.rejinpaul.com

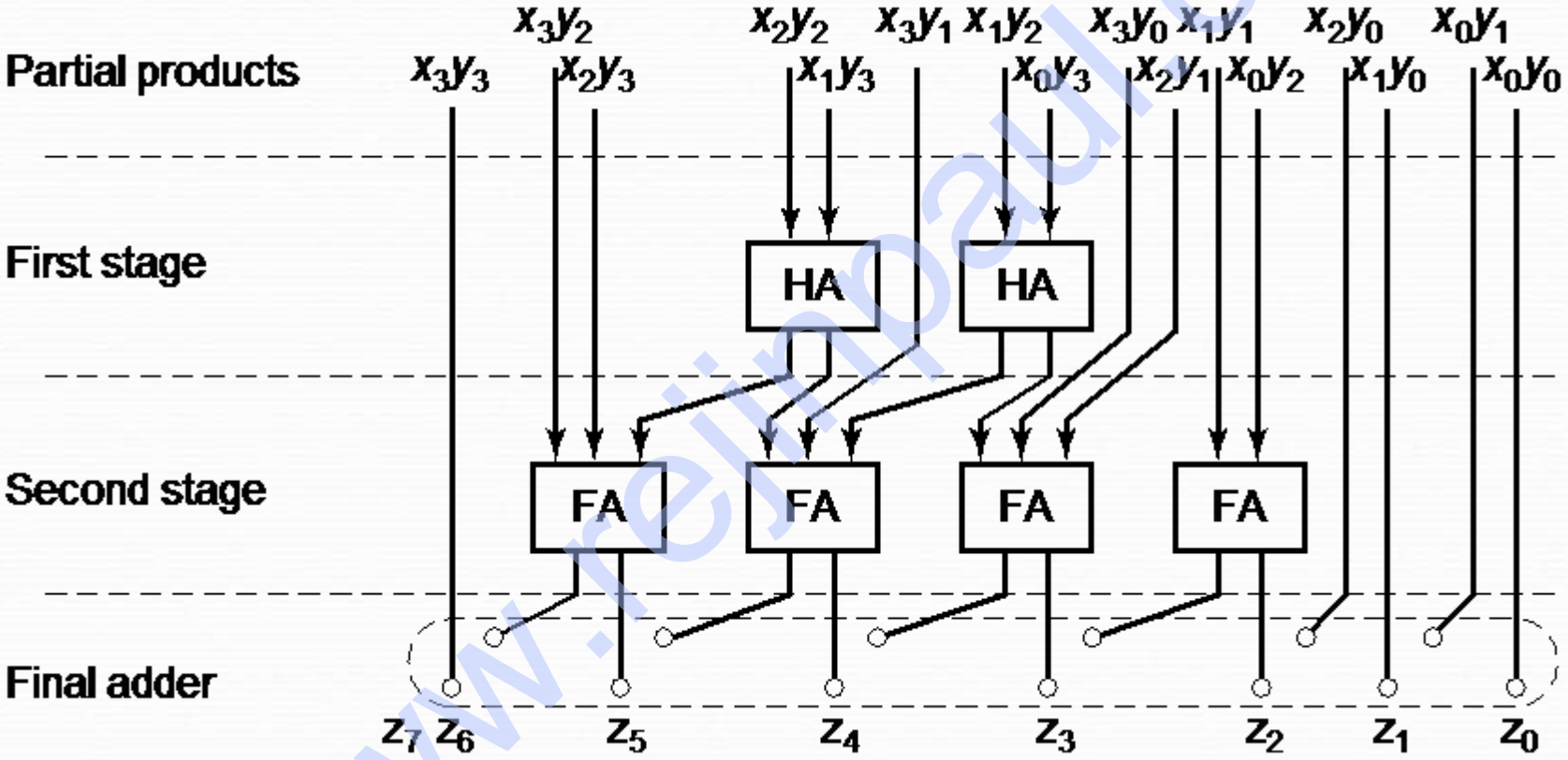
Wallace-Tree Multiplier



www.rejinpaul.com

4-bit Wallace Tree Multiplier

- The tree multiplier realizes substantial hardware savings for larger multipliers.
- The propagation delay is reduced to O(log3/2(N)).
- This structure is substantially faster than the carry save structure for large multiplier word lengths.



www.rejinpaul.com

Advantages

- Minimum propagation delay
- Number of partial products reduced by 2/3 per multiplier stage.
- Less layers

Disadvantages

- Complicated layout
- High power consumption

www.rejinpaul.com

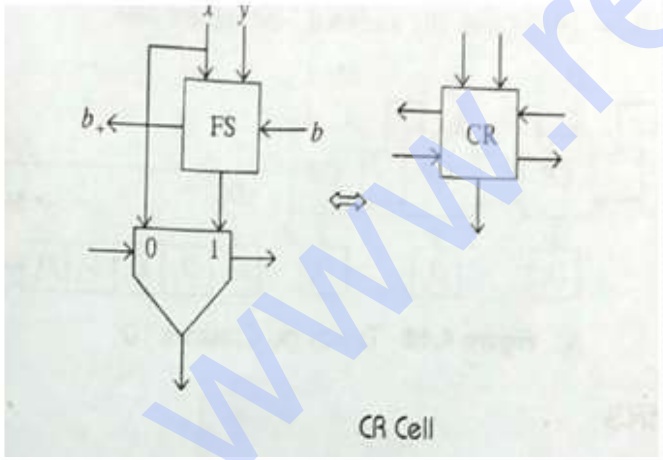
4)Dividers

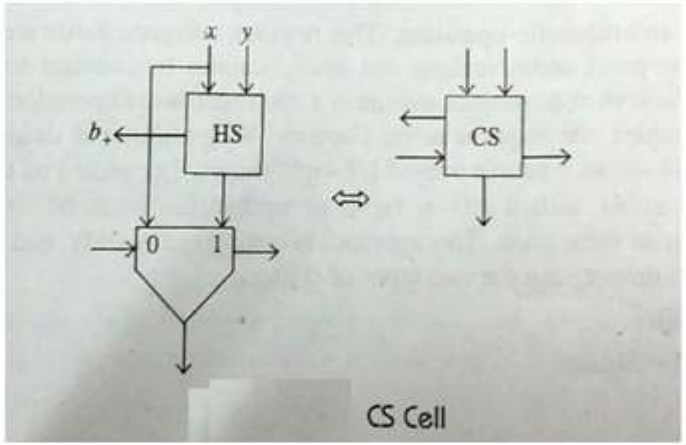
- Divider is performed using subtraction.
- Basic division methods
- Restoring division method→ the quotient bit is set to 1,if the result after subtracting the divisor from the dividend is greater than or equal to 0.
- Otherwise the quotient bit is set tom 0.
- Then the divisor is added back to the dividend in order to restore the original dividend before proceeding to the next bit.
- Non-Restoring division method→the divisor is not added back to XY

www.rejinpaul.com

4.1)Combinational divider(7/4)

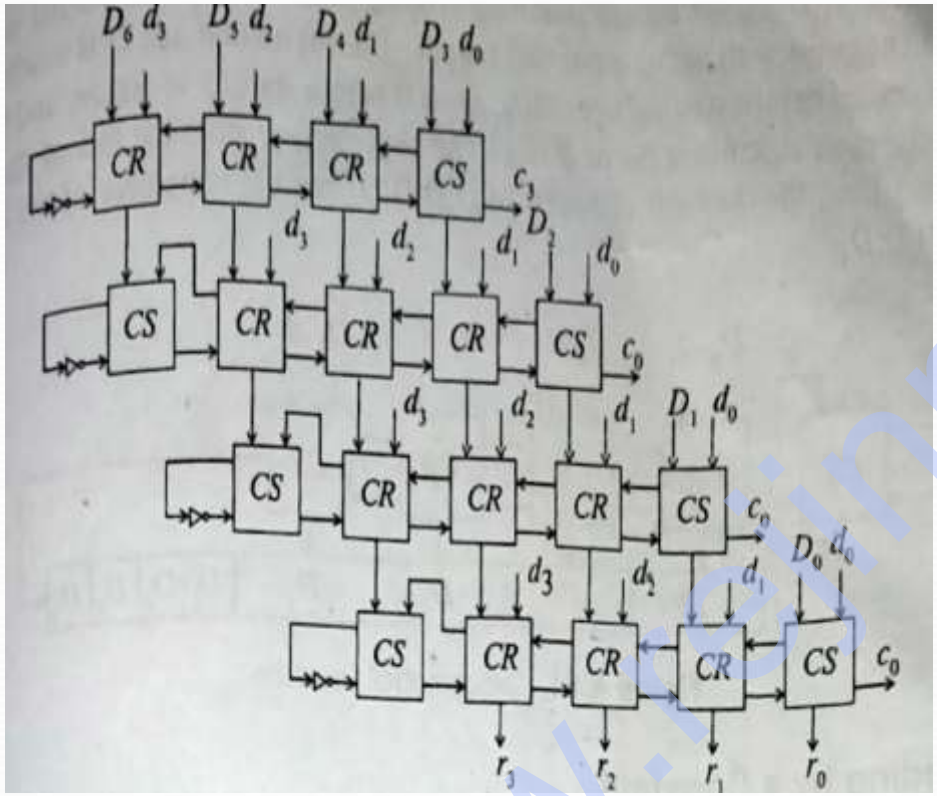
- D→Dividened:d→divisor;C→Quotient;r→remainder
- If r<d→ then D=C*d+r
- The length of the dividend is m=2n-1 and the MSB of the divisor d is 1.
- n→number of bits in the divisor.
- The division is made in n steps,in each of which a bit of the quotient is obtained.
- In the 1st stage d is subtracted from the 4MSB of D($D_6D_5D_4D_3$).
- If the result is positive, the quotient bit is 1 and the difference phases to the next stage as the MSB of modified dividend.
- If the result is negative, the quotient is 0 and the dividend unchanged passes to the next stage.
- The 2nd and subsequent iterations consist on repeating the same as the 1st iteration.
- The remainder $r_3r_2r_1r_0$ is obtained in the 4th iteration.





www.rejinpaul.com

Combinational division(7/4)



let $n = 4$, $D = 0110101$, $d = 1011$.

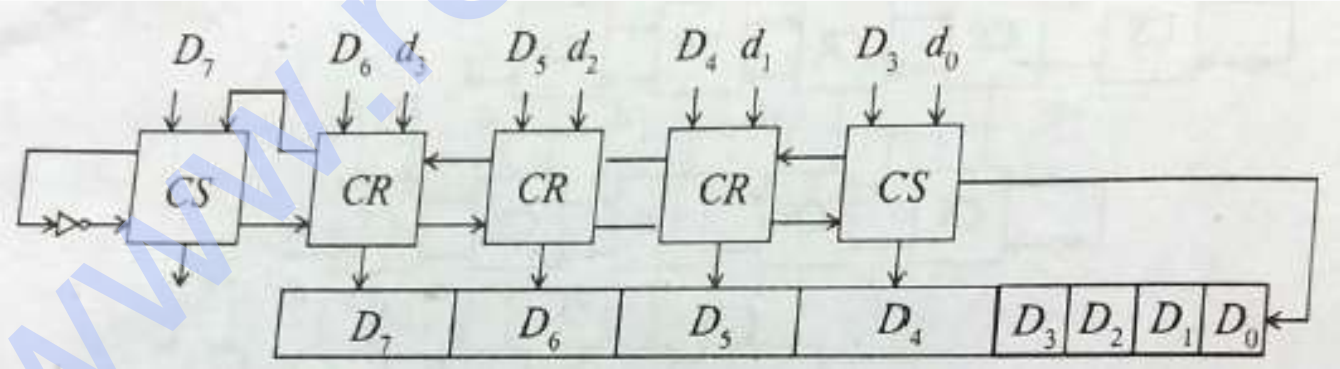
1) $D_6D_5D_4D_3 \rightarrow$	0110	6
	− 1011	11
	Negative	$\rightarrow c_3 = 0$
2) $D_6D_5D_4D_3 \rightarrow$	1101	13
	− 1011	11
	0010	2 $\rightarrow c_2 = 1$
3) $D_6D_5D_4D_3 \rightarrow$	0100	7
	− 1011	11
	Negative	$\rightarrow c_1 = 0$
4) $D_6D_5D_4D_3 \rightarrow$	1001	9
	− 1011	11
	Negative	$\rightarrow c_0 = 0$
	Remainder	$\rightarrow$ 1001

53/11
Q=4
R=9

www.rajinpaul.com

4.2) Sequential divider

- The circuit has three CR cells, two CS cells, one 4-bit latch to store the divisor d and an 8-bit register for the dividend D.
- The D register consists of two 4-bit register consists of two 4-bit register.
- The 1st register ($D_7D_6D_5D_4$) must be simultaneous reading and writing.
- The 2nd ($D_3D_2D_1D_0$) must be a shift register with serial input and output.
- The registers d &D is loaded with the data to be processed before starting operation.
- D_7→is 0 before starting to divide and the divisor is shifted so the MSB of d is 1.
- The shift register store the bits of the quotient.
- After 4 iterations, the quotient is stored in $D_3D_2D_1D_0$ and the remainder of the division is stored in $D_7D_6D_5D_4$.



www.rejinpaul.com

5)Shifters

- Shifters are used in floating point units,scalers, and multiplications by constant numbers.
- Shifting a data word left or right over a constant amount is a hardware operation.

Types of shifters

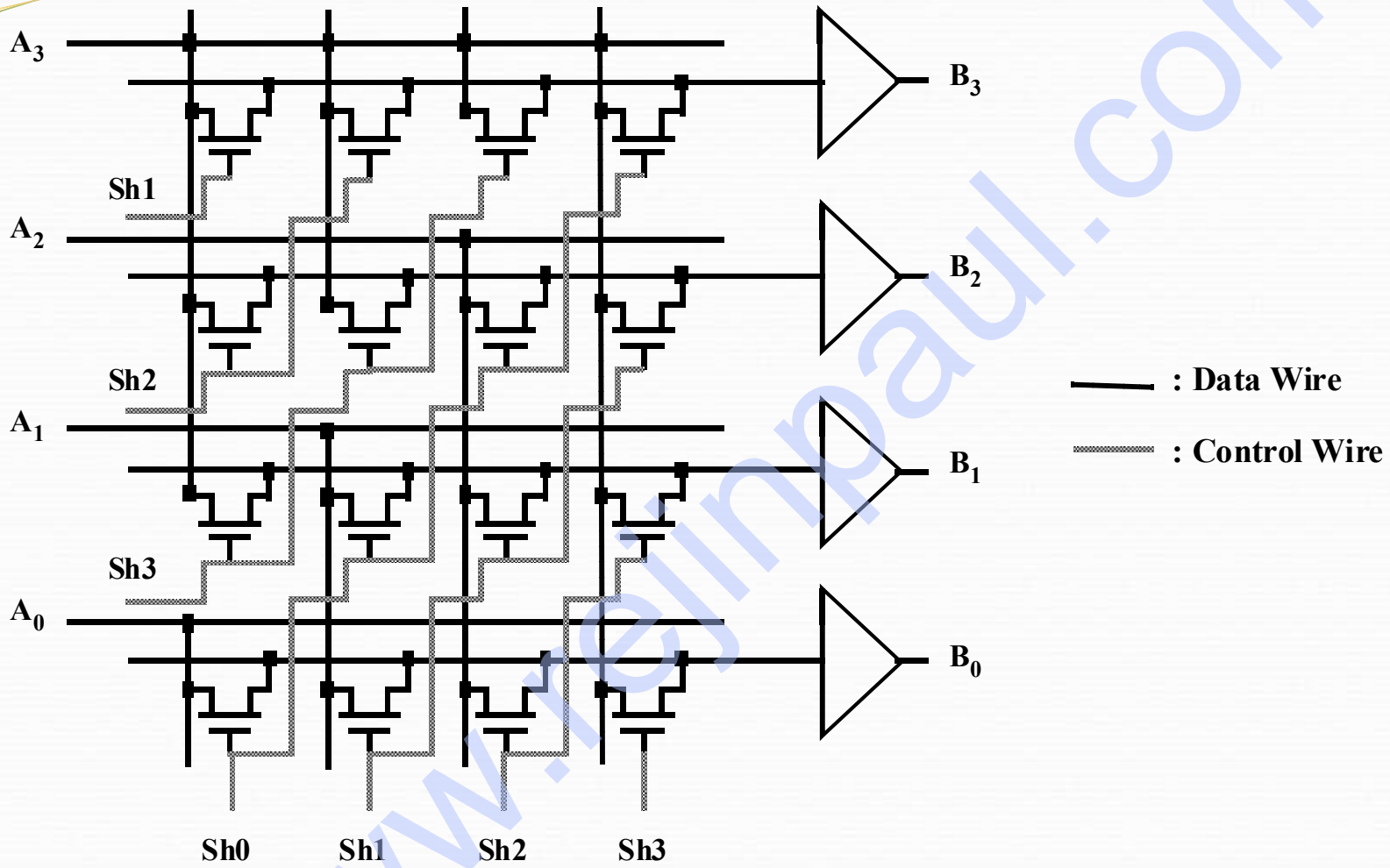
1. Barrel shifter
2. Logarithmic shifter

www.rejinpaul.com

5.1)Barrel shifter

- It consists of an array of transistros,in which the number of rows equals the word length of the data and the number of columns equals the maximum shift width.
- The control wires are routed diagonally through the array.
- Switch gates forms together in groups of fours and also form 4 separate groups corresponding to shifts of 0,1,2, and 3 bits. This arrangement is called barrel shifter.
- Barrel shifters perform 6 operations namely logical right& left shift, arithmetic right shift & left shift, rotate right & rotate left.
- The inter bus switches have gate inputs connected in stair case manner in groups of four and have four shift control inputs that mutually exclusive in the active state.
- CMOS gates are used as switches.
- Barrel shifter connects the input lines representing a word to a group of output lines with required shift determined by control inputs Sh_0, Sh_1, Sh_2,and Sh_3.
- Control inputs decide the direction of the shift.
- If input word has n-bits, shifts from 0 to n-1 bit positions are possible.

www.rejinpaul.com



www.rejinpaul.com

- Layout size in barrel shifter is not dominated by the active transistors, but by the number of wires running through the cell.
- The size of the cell is bounded by the pitch of the metal wires.
- When selecting a shifter, the format in which to present the shift value is provided.
- Barrel shifter needs a control wire for every shift bit.
- A 4-bit shifter needs four control signals.
- In a processor the required shift value comes in an encoded binary format.
- The encoded control word needs two control signals and is represented as 11 for a shift over 3 bits.

Advantages

- Signal has to pass through at most one transmission gate
- Propagation delay is constant and independent of the shift value
- Layout size is not dominated by transistors.

Drawbacks

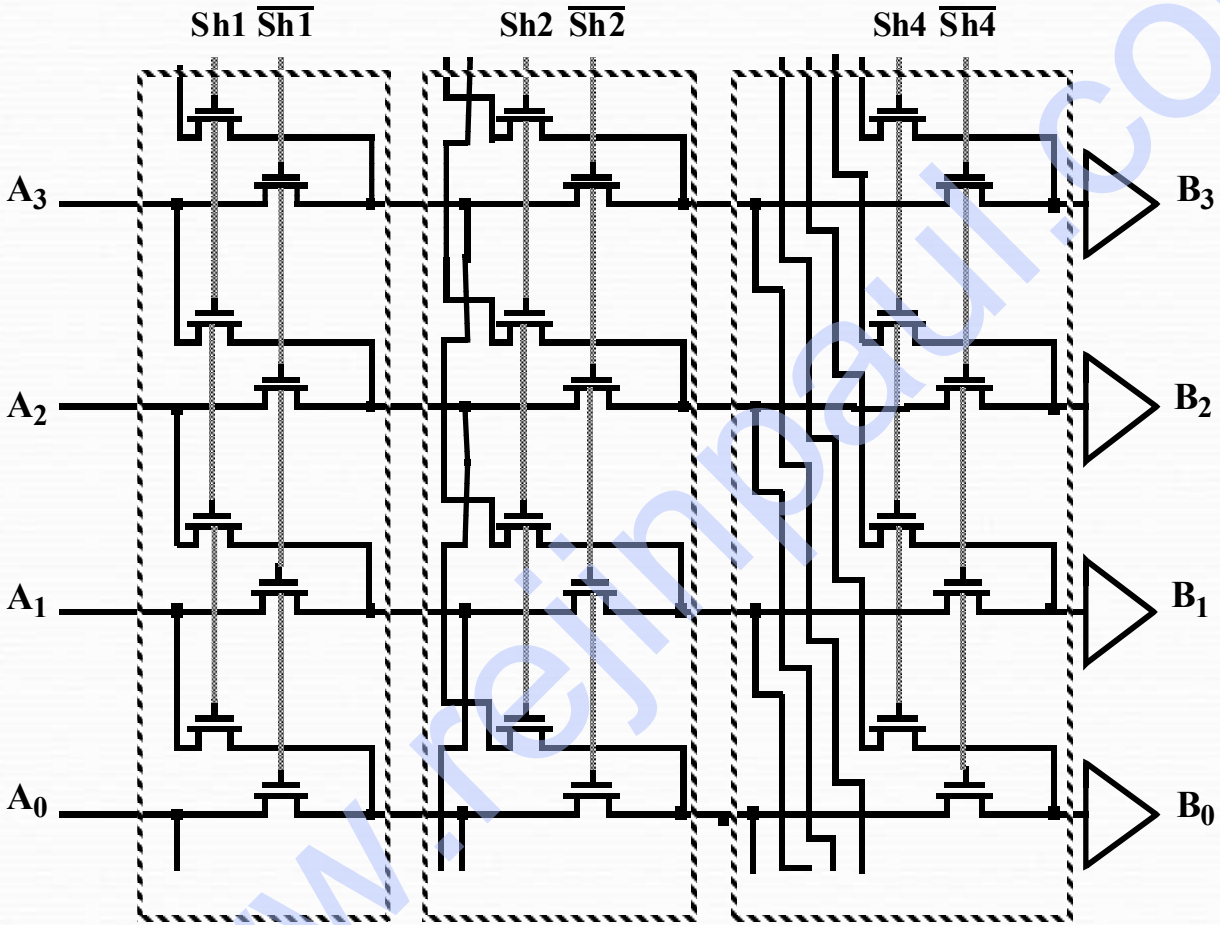
- Capacitance at the input of the buffers rises linearly with the max shift width.
- Used for small shift values
- Separate decoder is needed for decoding control inputs

www.rejinpaul.com

5.2)Logarithmic Shifter(7bits)

- Logarithmic shifter uses staged approach for shifting.
- The total shift value is decomposed into shifts over powers of two.
- A shifter with a maximum shift width of M consists of a $\log_2 M$ stages, where the i^{th} stage either shifts over 2^i or passes data unchanged.
- To shift over 5bits,the 1^{st} stage is set to shift mode, the 2^{nd} to pass mode and the last stage again to shift.
- The control word for this shifter is already encoded and no separate decoder is required.
- Speed of the logarithmic shifter depends on the shift width in a logarithmic way.
- An M-bit shifter requires $\log_2 M$ stages .The series connection of pass transistors slows down the shifter for larger shift values.
- For larger shift values, the logarithmic shifter becomes more effective ,in terms of both area and speed.
- The logarithmic shifter is easily parameterized, allowing for automation generation.

www.rejinpaul.com



www.rejinpaul.com

Speed and Area Tradeoff

- Datapath,speed,area and power can be traded off through the choice of the supply voltage, transistor thresholds and device sizes.
- The following are the techniques to reduce the power.

1. Design time power reduction
2. Run time power management
3. Power reduction in standby or sleep mode
4. Design a trade off.

www.rejinpaul.com

2) Run Time Power Management

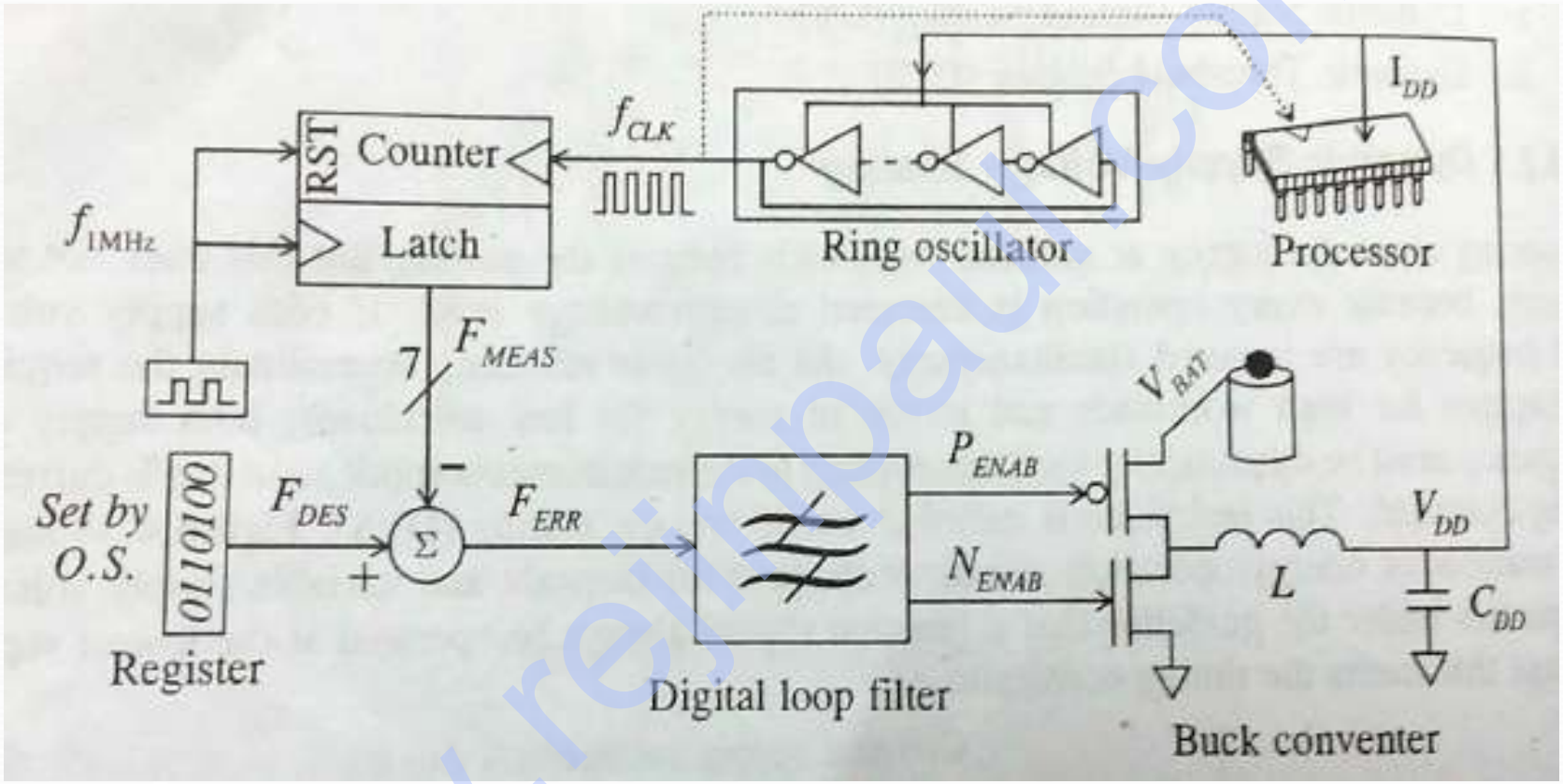
1. Dynamic Supply Voltage Scaling(DVS)
2. Dynamic Threshold Scaling(DTS)

www.rejinpaul.com

Dynamic Supply Voltage Scaling(DVS)

- DVS system is a ring oscillator, whose oscillating frequency matches the microprocessor critical path.
- Inside the power supply control loop, ring oscillator provides the translation between the supply voltage and the clock frequency.
- The operating system sets the desired frequency.
- The current value of ring oscillator frequency is measured and compared with the desired frequency.
- The difference is used as a feedback error (F_{ERR}).
- By adjusting the supply voltage, the supply voltage loop changes the ring oscillator frequency to set the error value to 0.

www.rejinpaul.com

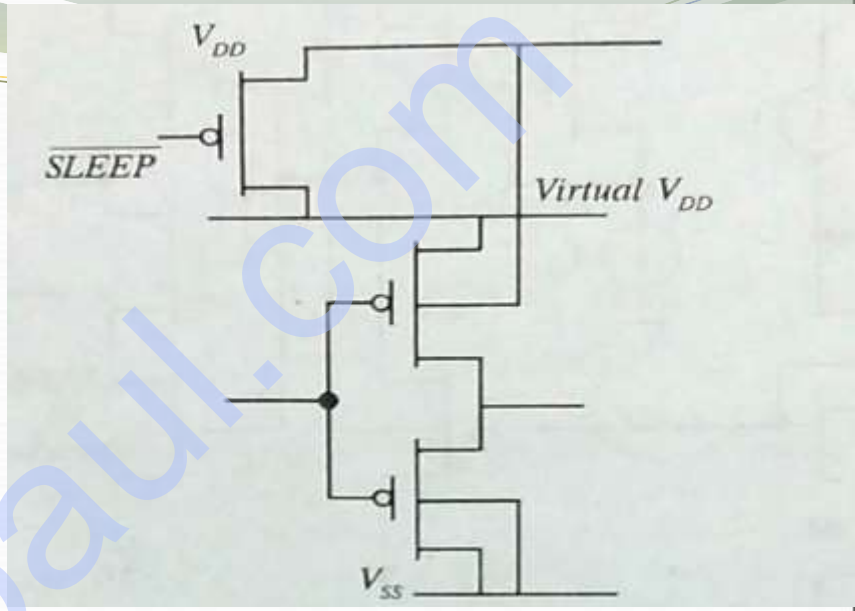


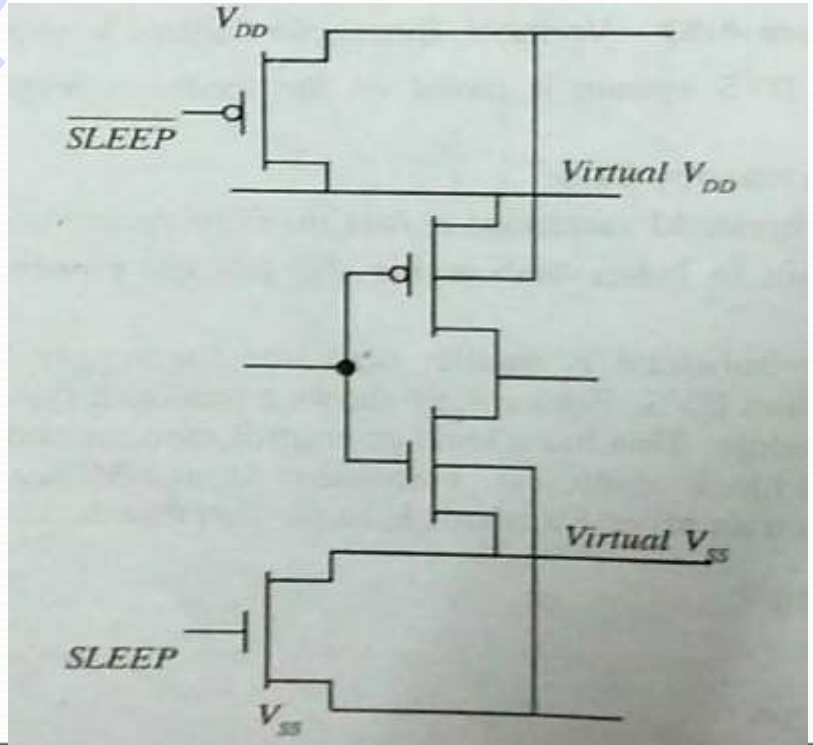
www.rejinpaul.com

Dynamic Threshold Scaling(DTS)

- For large sleep transistors to switch OFF the power supply rails when the circuit is in the sleep mode.
- This approach reduces the leakage current, but increases the design complexity.

- The SLEEP signal is high normal operation. Sleep transistors is small as a resistance as possible.
- Transistors finite resistance results in noise on supply rails.
- To minimize the fluctuations in the supply voltage, the sleep transistors have a very low on-resistance.
- Addition of sleep transistor →leakage is reduced





www.rejinpaul.com

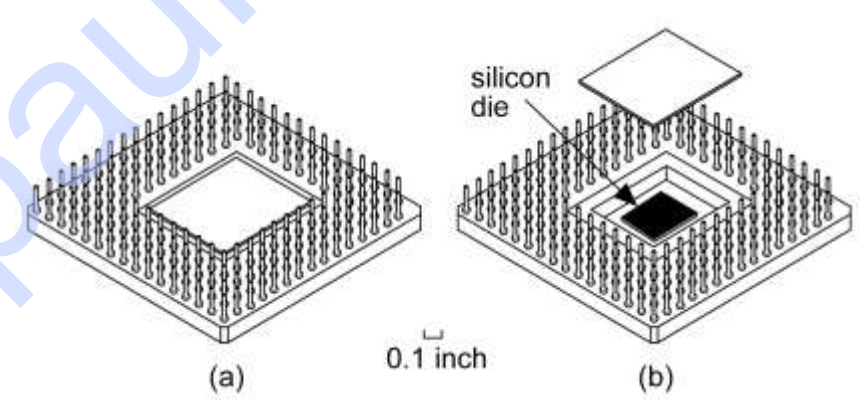
Unit-V

IMPLEMENTATION STRATEGIES

www.rejinpaul.com

INTEGRATED CIRCUITS

- IC→is a set of electronic circuits on a small chip.
- Levels of integration:
  - SSI - Small scale integration
  - MSI - Medium scale integration
  - LSI - Large scale integration
  - VLSI - Very large scale integration
  - USLI - Ultra large scale integration
- Implementation technology
  - TTL
  - ECL
  - MOS - NMOS, CMOS



www.rejinpaul.com

ASIC

- ASIC - Application Specific Integrated Circuit
- In Integrated Circuit (IC) designed to perform a specific function for a specific application.

Advantages

- Used for specific application
- More secure for the design circuit
- Reduction in area and power consumption
- Better performance

Drawbacks

- Difficult to replace
- Difficult to make minor changes in the late stage

www.rejinpaul.com

Types of ASICs

- Full-Custom ASICs
- Standard-Cell–Based ASICs
- Gate-Array–Based ASICs
- Channeled Gate Array
- Channelless Gate Array
- Structured Gate Array
- Programmable Logic Devices
- Field-Programmable Gate Arrays

Semi Custom ASICs

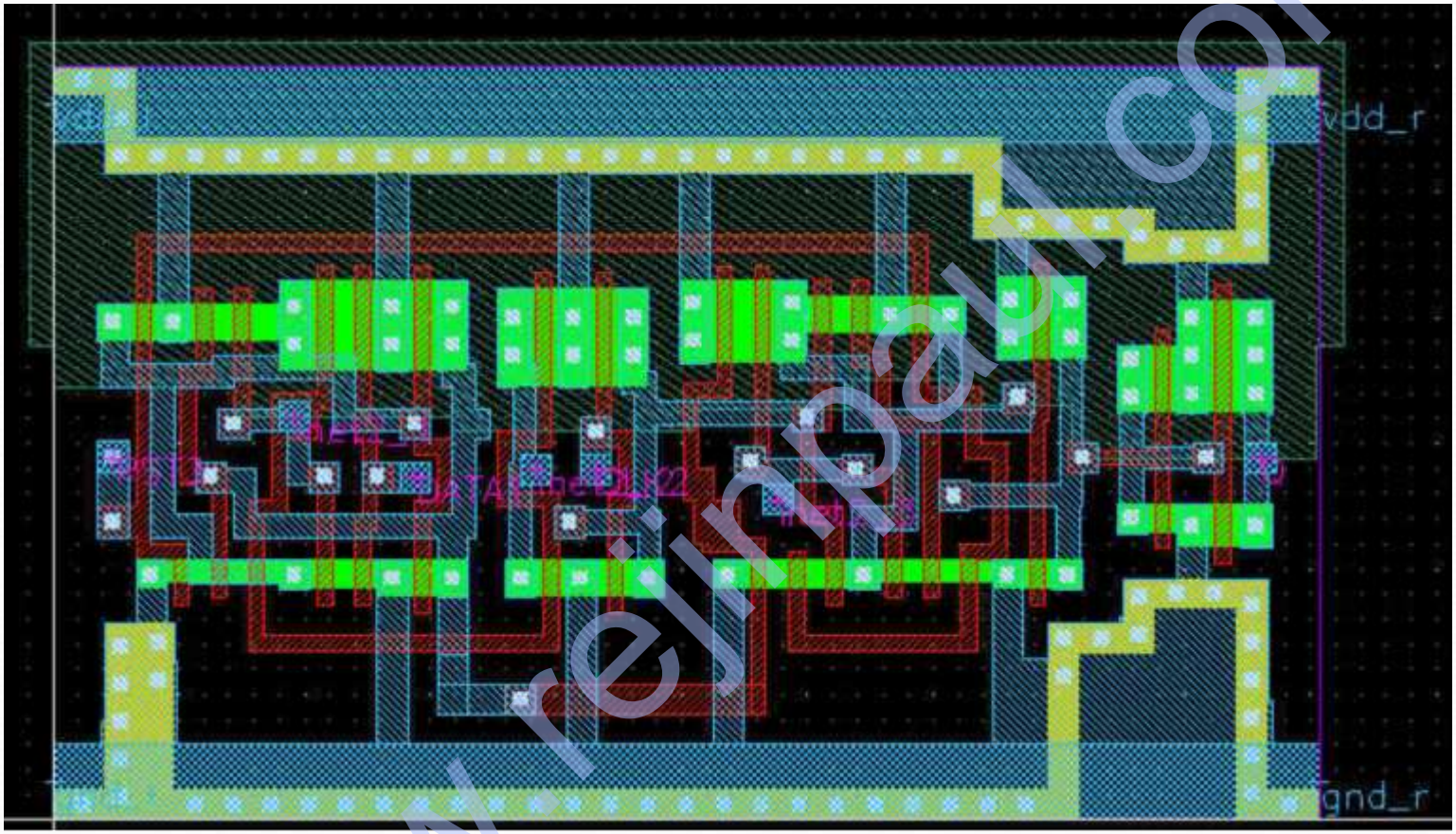
www.rejinpaul.com

1) Full-Custom ASICs

- An engineer (expert) designs some or logic cells ,circuits or layout specifically for one ASIC.
- Designer hand draws geometries which specify transistors and other devices for an integrated circuit.
- Involves the creation of a a completely new chip, which consists of about a dozen masks (for the photolitographic manufacturing process).
- Fabrication time from geometry submission to returned chips is at least 6-8 weeks
- There is no existing cell library available to be used for the entire design.
- Full custom ASIC cannot be modified to suit different applications.
- Full-custom design offers the highest performance Fabrication costs are high
- The disadvantages of full-custom design include increased design time, complexity, design expense, and highest risk.

- Eg→Microprocessors,analog/digital (communications), sensors and actuators, and memory (DRAM).

www.reinpaul.com



www.rejinpaul.com

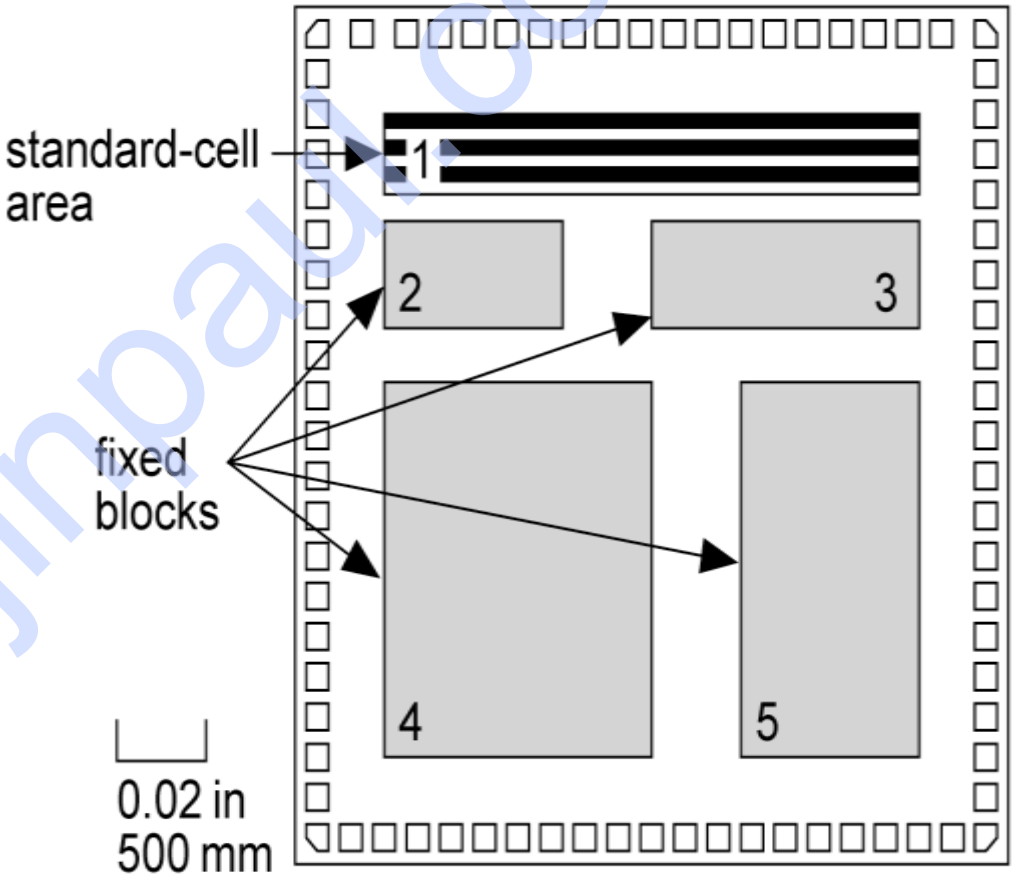
2)Standard-Cell-Based ASICs

- Designer uses a library of predesigned logic cells. (AND gates,OR,MUX).
- Designer does not have to be a VLSI expert.
- Manufacturing lead time is about 2-4 weeks.
- The predesigned logic cells are known as standard cells.
- An automatic place and route tool does the layout.
- Design time can be much faster than full custom because layout is automatically generated.
- The Standard cells areas are called flexible blocks.Flexible blocks in CBIC are built of rows of standard cells.
- This is used in combination with larger predesigned cells,like microcontrollers or microprocessor.These are called megacells.
- Megacells are also called megafunctions , full-custom blocks , system-level macros(SLMs), fixed blocks , cores , or Functional Standard Blocks (FSBs)
- Standard cells can be placed anywhere on the silicon.

www.rejinpaul.com

Cell-Based ASIC (CBIC)

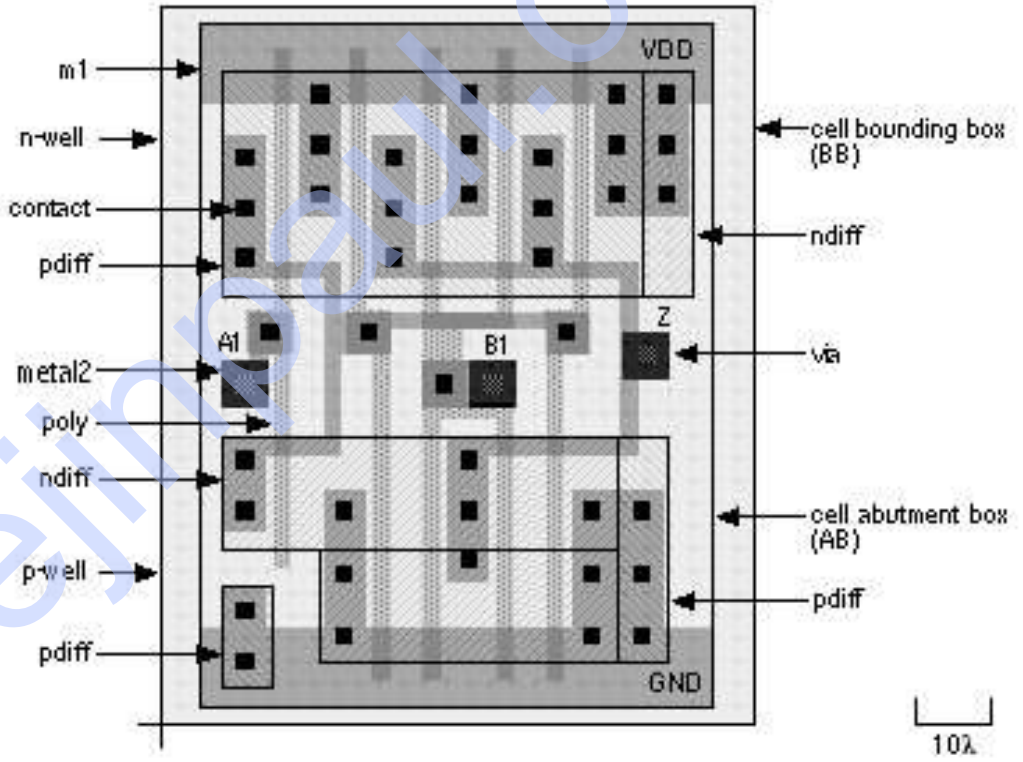
- Standard cells are designed to fit together like bricks in a wall.
- Power and ground buses (V_{DD},V_{SS},GND) run horizontally on metal lines inside the cells.
- Groups of standard cells fit horizontally together to form rows.
- The rows stack vertically to form flexible rectangular blocks.
- Connect a flexible block built from several rows of standard cells to other standard cells blocks .



www.rejinpaul.com

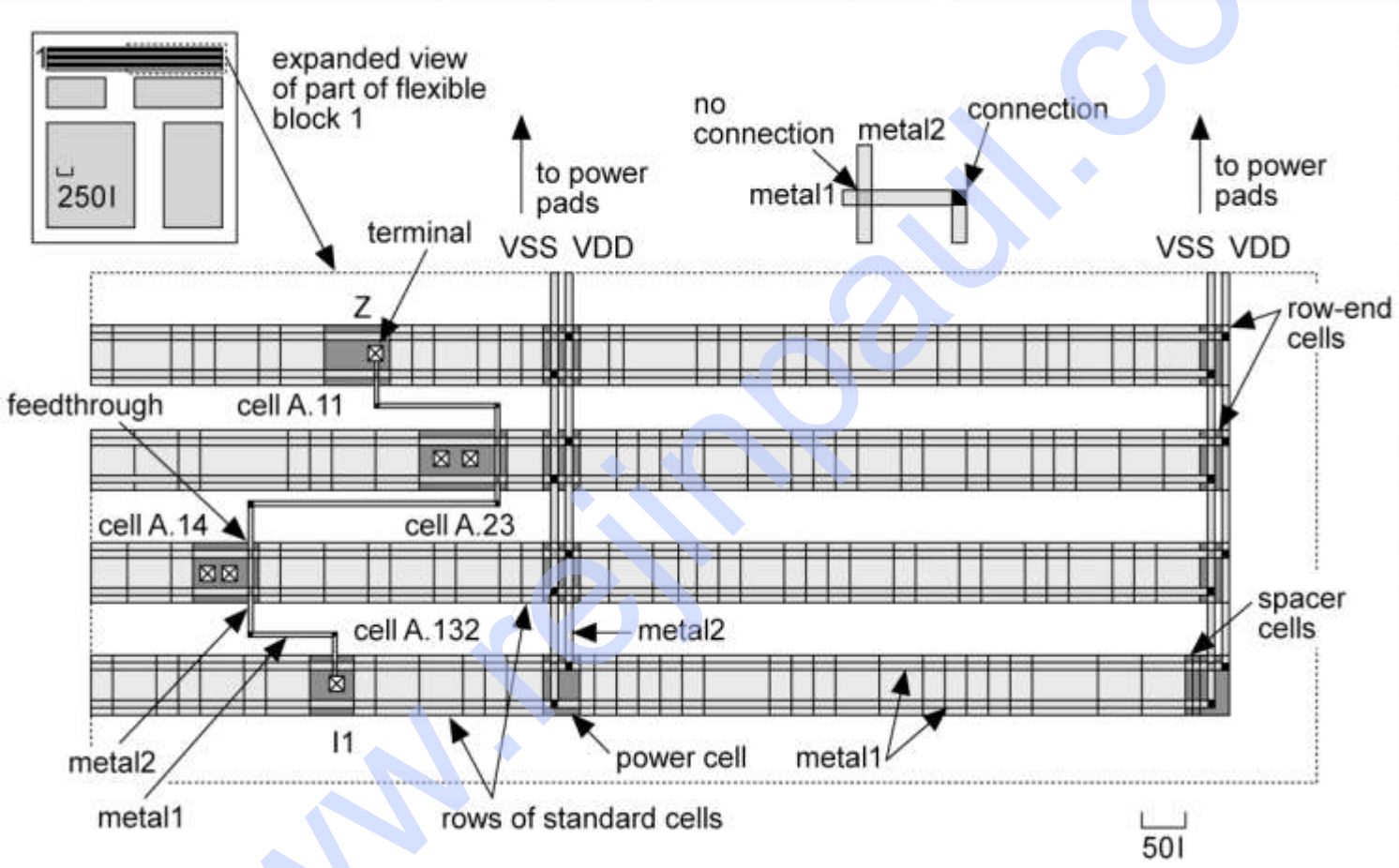
Standard Cell Layout

•This cell would be approximately 25 microns wide on an ASIC with l (lambda) = 0.25 microns (a micron is 10 ^{6} m).
•Standard cells are stacked like bricks in a wall.
•The abutment box (AB) defines the edges of the brick.
• The difference between the bounding box (BB) and the AB is the area of overlap between the bricks.
•Power supplies (labeled VDD and GND) run horizontally inside a standard cell on a metal layer that lies above the transistor layers.
•This standard cell has center connectors (A1, B1, and Z)



www.reinpal.com

Standard Cell ASIC Routing



www.rejinpaul.com

- Modern CMOS ASICs use two, three, or more levels (or layers) of metal for interconnect. This allows wires to cross over different layers.
- In a two-level metal CMOS technology, connections to the standard-cell inputs and outputs are usually made using metal2.
- In a three-level metal technology, connections may be internal to the logic cell.
- In both two-level and three-level metal technology, the power buses (VDD and GND) inside the standard cells normally use the lowest (closest to the transistors) layer of metal (metal1).
- The width of each row of standard cells is adjusted so that they may be aligned using spacer cells .
- The power buses, or rails, are then connected to additional vertical power rails using row-end cells at the aligned ends of each standard-cell block.
- If the rows of standard cells are long, then vertical power rails can also be run in metal2 through the cell rows using special power cells that just connect to VDD and GND.

www.reinpaul.com

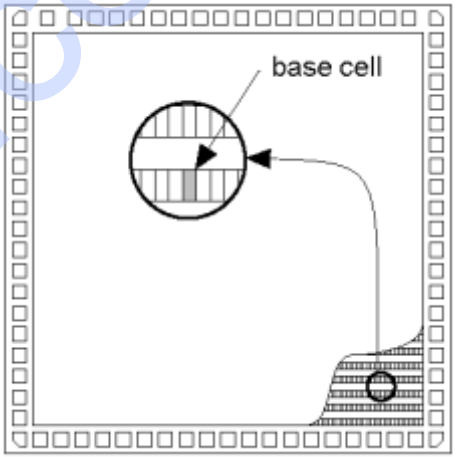
3)Gate-Array-Based ASICs

- In a gate-array-based ASIC, the transistors are predefined on the silicon wafer.
- The predefined pattern of transistors is called the *base array.*
- The smallest element that is replicated to make the base array is called the *base* or *primitive cell*
- To distinguish this type of gate array from other types of gate array, it is often called a masked gate array (MGA).
- The designer chooses from a gate-array library of predesigned and pre-characterized logic cells.
  The logic cells in a gate-array library are often called macros .
- After validation, automatic placement and routing are typically used to convert the macro-based design into a layout on the ASIC using primitive cells
- Types of MGAs:
  1) Channeled Gate Array
  2) Channelless Gate Array
  3) Structured Gate Array
  4) PLD
  5) FPGA

www.rejinpaul.com

3.1)Channeled Gate Array

- Only the interconnect is customized
- The interconnect uses predefined spaces between rows of base cells .
- Manufacturing lead time is between two days and two weeks.
- A channeled gate array is similar to a CBIC.
- Both use rows of cells separated by channels used for interconnect.
- One difference is that the space for interconnect between rows of cells are fixed in height in a channeled gate array, whereas the space between rows of cells may be adjusted in a CBIC

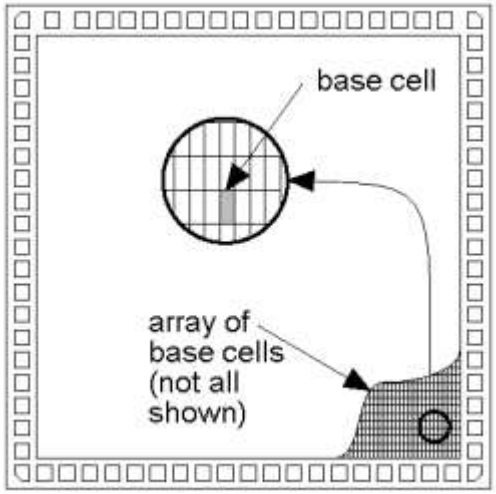


www.rejinpaul.com

3.2)Channelless Gate Array

- Channellness gate array architecture is widely used.
- The routing on a channellness gate array uses rows of unused transistors
- Only some (the top few) mask layers are customized the interconnect.
- Manufacturing lead time is between two days and two weeks.
- The key difference between a channelless gate array and channeled gate array is that there are no predefined areas set aside for routing between cells on a channelless gate array.
- Routing over the top of the gate-array devices

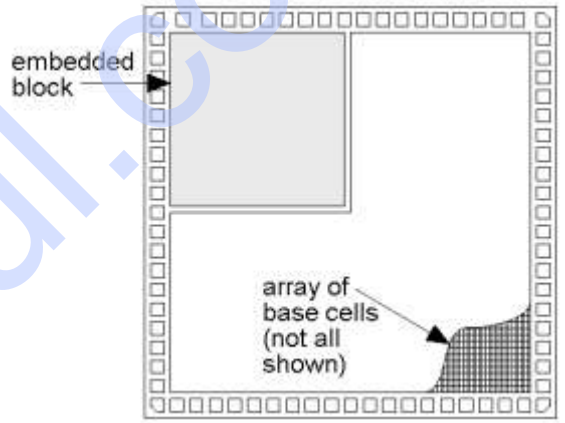
•The logic density of a given silicon area is higher for channelless gate arrays than for channeled gate arrays.
•Customizing the contact layer in a channelless gate array allows us to increase the density of gate-array cells because routing done over the top of unused contact sites.



www.rejinpaul.com

3.3)Structured Gate Array/ Masterslice/ Masterimage

- An embedded gate array or structured gate array combines some of the features of CBICs and MGAs.
- The disadvantages of the MGA is the fixed gate-array base cell.
- This makes the implementation of memory, difficult and inefficient.
- In an embedded gate array we set some of the IC area and dedicate it to a specific function.
- This embedded area either can contain a different base cell that is more suitable for building memory cells, or circuit block of microcontroller.



Gate array die with embedded block

www.reinpaul.com

Features

- Only the interconnect is customized
- Custom blocks can be embedded
- Manufacturing lead time is between two days and two weeks.

Advantages

- An embedded gate array gives the improved area efficiency
- Increased performance of a CBIC but with the lower cost
- Faster turnaround of an MGA

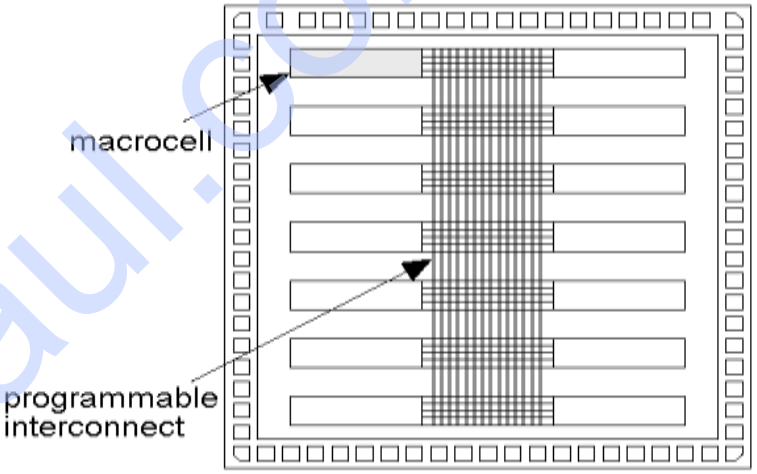
Disadvantages

- Embedded function is fixed. (For example, if an embedded gate array contains an area set aside for a 32 k-bit memory, but we only need a 16 k-bit memory, then we may have to waste half of the embedded memory function)

www.rejinpaul.com

3.4) Programmable Logic Devices-PLD

•Programmable logic devices (PLDs) are standard ICs that are available in standard configurations.
•It sold in very high volume to many different customers.
•PLDs may be configured or programmed. also belong to the family of ASICs.

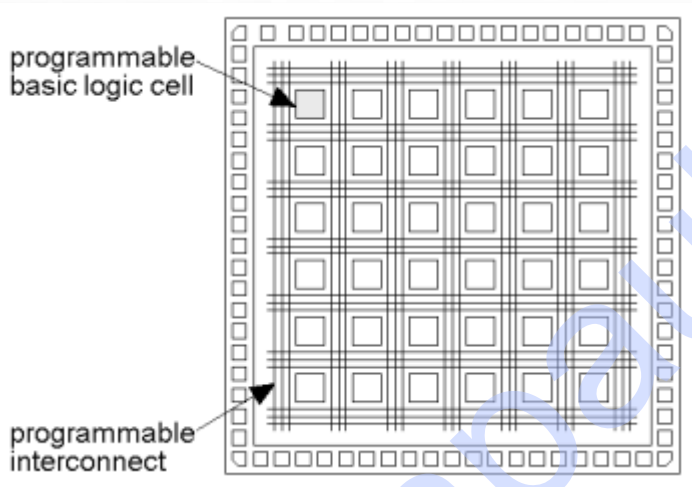


Programmable Logic Device (PLD) die

•Features
- No customized mask layers or logic cells
- Fast design turnaround
- A single large block of programmable interconnect
  - Erasable PLD (EPLD)
  - Mask-programmed PLD
- A matrix of logic macro-cells that usually consist of programmable array logic followed by a flip-flop or latch

www.rejinpaul.com

3.5) Field Programmable Gate Array-FPGA



- Complex PLD's are called FPGA.
- FPGA is larger and complex than a PLA.
- The building blocks of FPGA are Logic cells,I/Oblocks,Programmable interconnect.

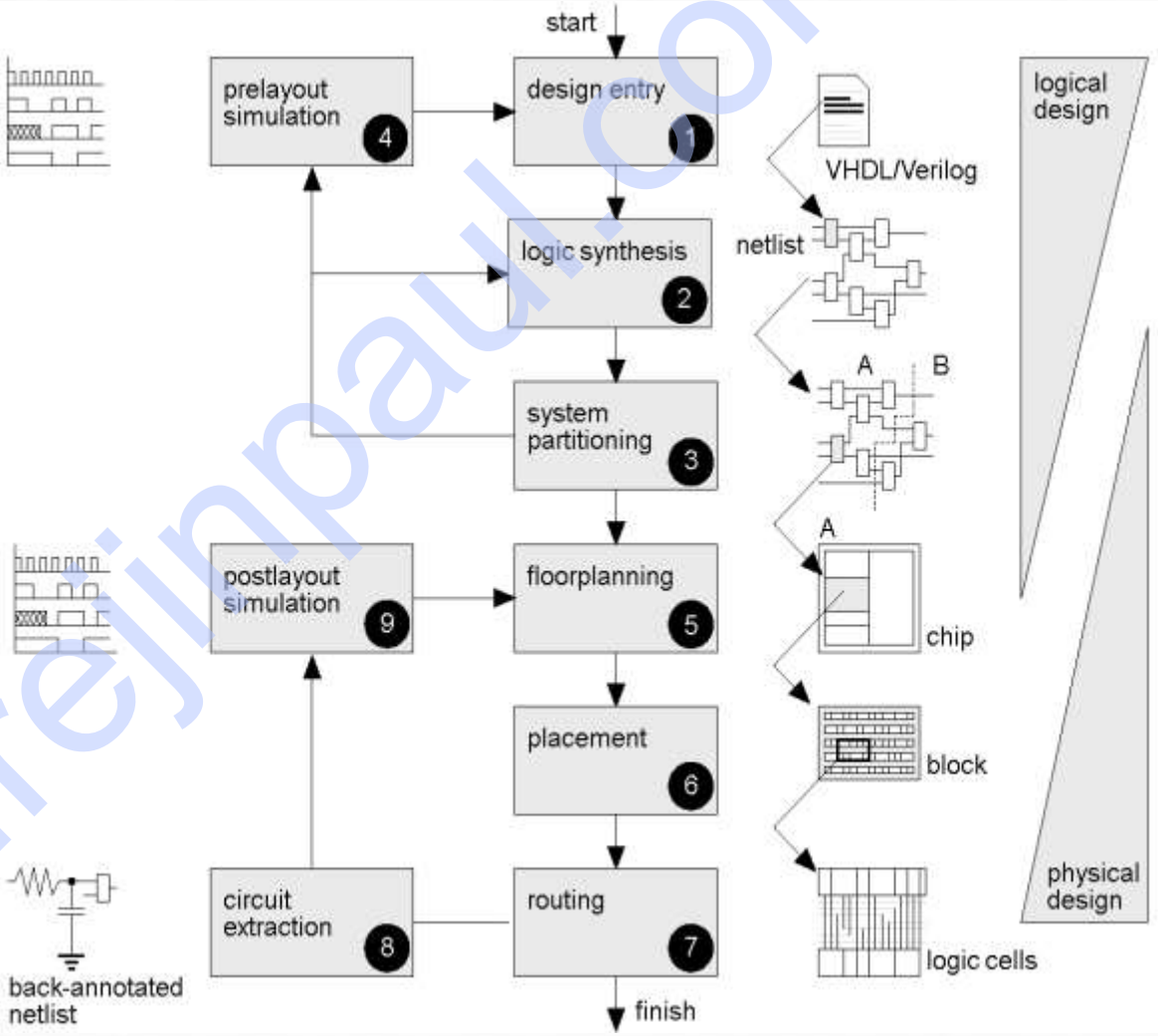
Features

- None of the mask layers are customized
- A method for programming the basic logic cells and the interconnect
- The core is a regular array of programmable basic logic cells that can implement combinational as well as sequential logic (flip-flops)
- A matrix of programmable interconnect surrounds the basic logic cells
- Programmable I/O cells surround the core
- Design turnaround is a few hours

www.rejinpaul.com

ASIC Design Flow

1. **Design entry** - Using a hardware description language (HDL) or schematic entry
2. ***Logic synthesis*** - Produces a netlist - logic cells and their connections
3. ***System partitioning*** - Divide a large system into ASIC-sized pieces
4. **Prelayout simulation** - Check to see if the design functions correctly
5. ***Floorplanning*** - Arrange the blocks of the netlist on the chip
6. ***Placement*** - Decide the locations of cells in a block
7. ***Routing*** - Make the connections between cells and blocks
8. ***Extraction*** - Determine the resistance and capacitance of the interconnect
9. **Postlayout simulation** - Check to see if the design still works with the added loads of the interconnect



www.rejinpaul.com

ASIC Cell Libraries

- A library of cells is used by the designer to design the logic function for an ASIC.
- Options for cell library:
  - (1) Use a design kit from the ASIC vendor
    - Usually requires the use of ASIC vendor approved tools
    - Cells are "phantoms" - empty boxes that get filled in by the vendor when you deliver, or 'hand off" the netlist
    - Vendor may provide more of a "guarantee" that design will work
  - (2) Buy an ASIC-vendor library from a library vendor
    - Library vendor is different from fabricator (foundry)
    - Library may be approved by the foundry (qualified cell library)
    - Allows the designer to own the masks (tooling) for the part when finished
  - (3) You can build your own cell library
    - Difficult and costly

www.reinpaul.com

ASIC Library Development

- A complete ASIC library (suitable for commercial use) must include the following for each cell and macro:
  - A physical layout→It is a cell layout –part of programmable ASIC design.
  - A behavioral model
  - A VHDL or Verilog model
  - A detailed timing model→needed to determine the performance of the ASIC.
  - A test strategy
  - A circuit schematic→describes each cell.
  - A cell icon (symbol)→cell identity with naming information
  - A wire-load model
  - A routing model

www.rejinpaul.com

FPGA BUILDING BLOCK ARCHITECTURE

- All programmable ASIC or FPGAs contain a basic logic cell.
- The basic logic cell is replicated in a regular array across the chip.
- There are the following three different types of basic logic cells: (1) multiplexer based (2) look-up table based (3) programmable array logic.
- The choice among these depends on the programming technology.

Types of Logic cell Configuration

1. Actel ACT (ACT1,ACT 2,ACT3)
2. Xilinx LCA (XC3000, XC4000 , XC5200)
3. Altera FLEX
4. Altera MAX

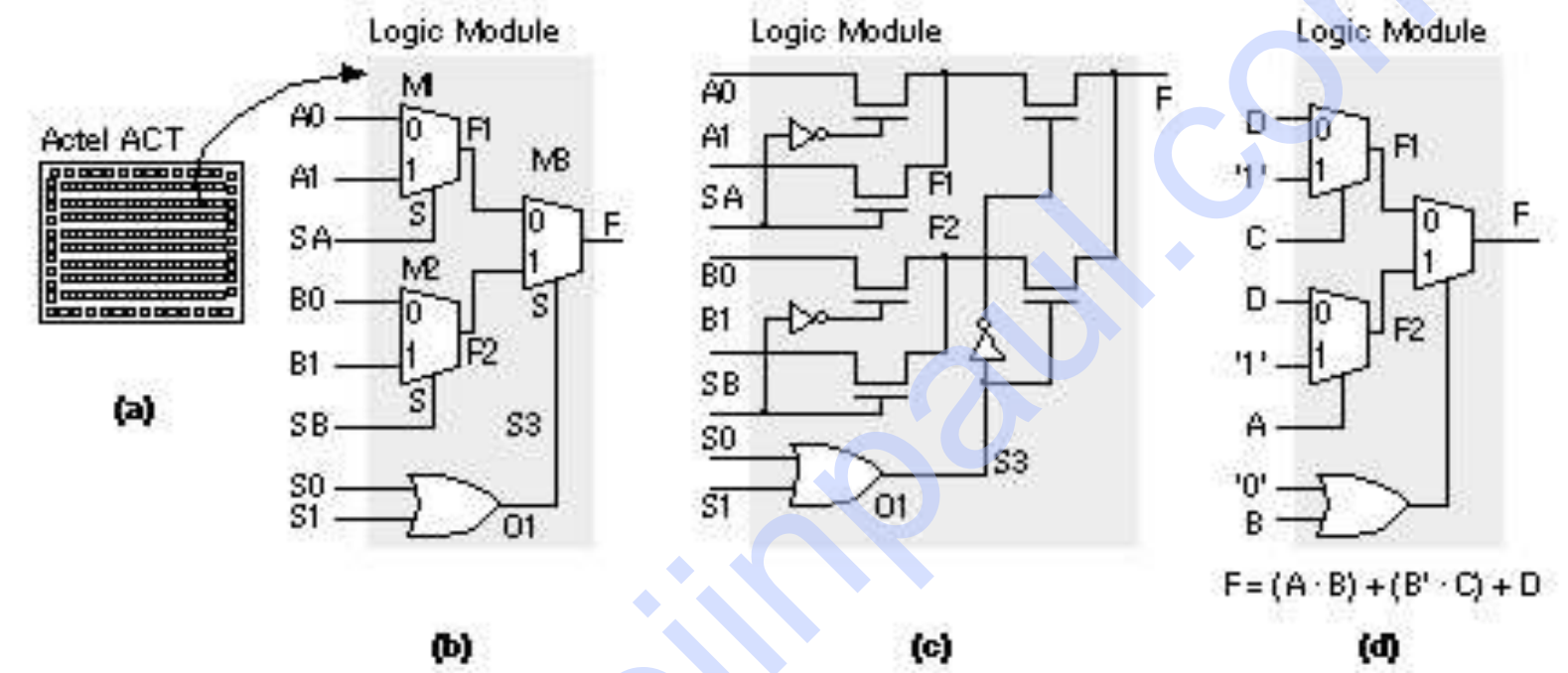
www.reijnpaul.com

1.1)ACTEL FPGA

- ACT1 module has three 2:1 Muxs with AND-OR logic at the select of final MUX and implements all 2 input functions, most 3 input and many 4 input functions.
- Software module generator for ACT1 takes care of all this.
- Apart from variety of combinational logic functions, the ACT1 module can implement sequential logic cells in a flexible and efficient manner. For example an ACT1 module can be used for a transparent Latch or two modules for a flip flop.

www.rejinpaul.com

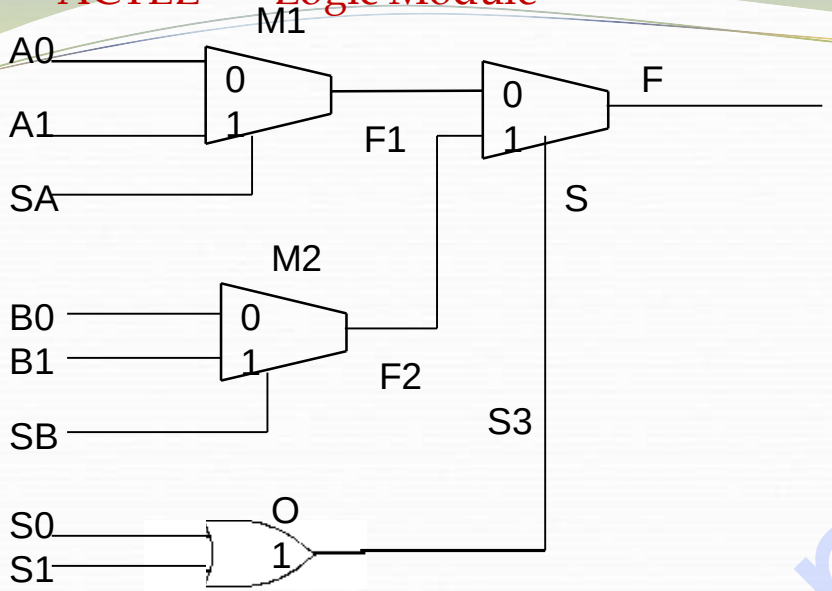
1.1)ACT 1 Logic Module



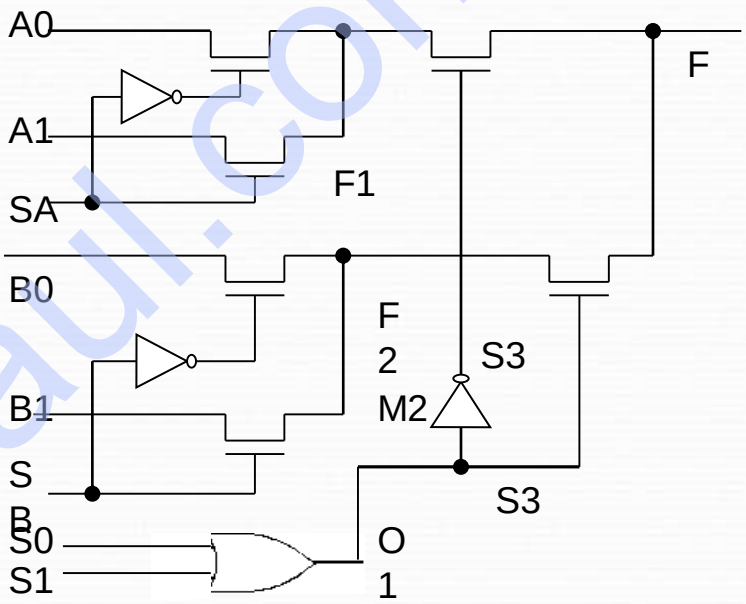
- We can build a logic function using an Actel Logic Module by connecting logic signals to some or all of the Logic Module inputs, and by connecting any remaining Logic Module inputs to VDD or GND.

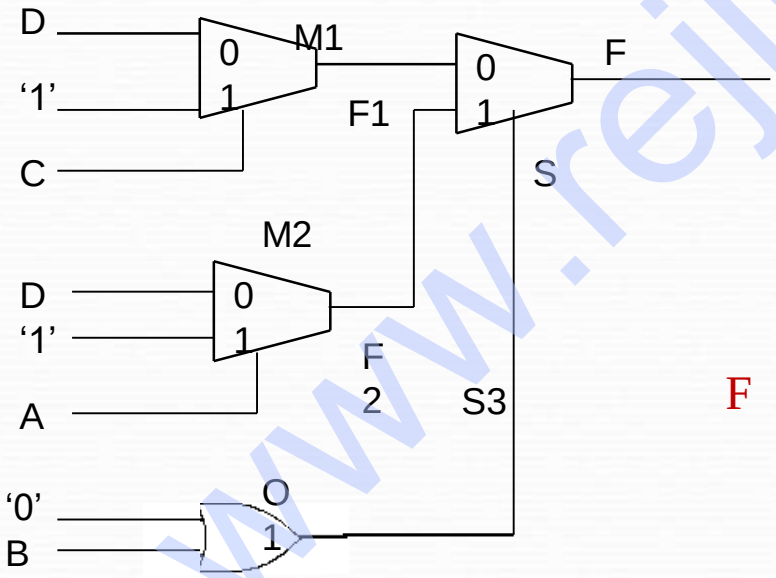
www.rejinpaul.com

ACTEL Logic Module



ACTEL – Implementation using pass transistors





A=1,B=1,C=0,D=0

ACTEL
An example logic macro

$$F = A.B + \overline{B}.C + D$$

$$= B\,[A.B + B.C + D] + \overline{B}[A.B + B.C + D]$$

$$= A.B + B.D + \overline{B}.C + \overline{B}.D$$

$$= B.(A+D) + \overline{B}\,(C+D)$$

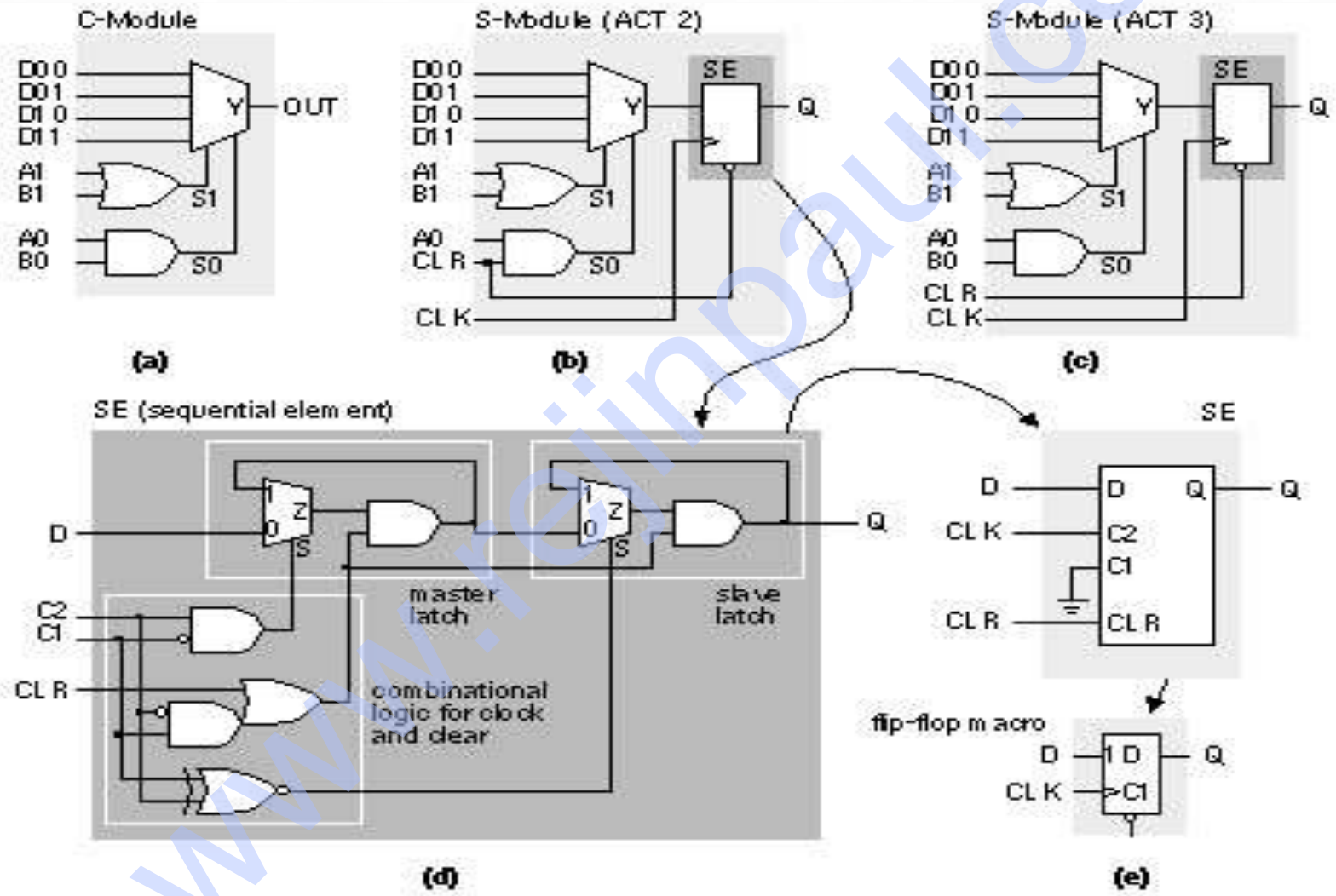
www.rejinpaul.com

Features of ACT1 Logic Module

- Operates at 3.3v(6mA) and 5v(10mA)
- Have up to 2000 gate arrays.
- Have up to 547 programmable logic modules.
- Have up to 273 flip-flops.
- Non-volatile and user programmable
- Fabricated with 1micron CMOS technology

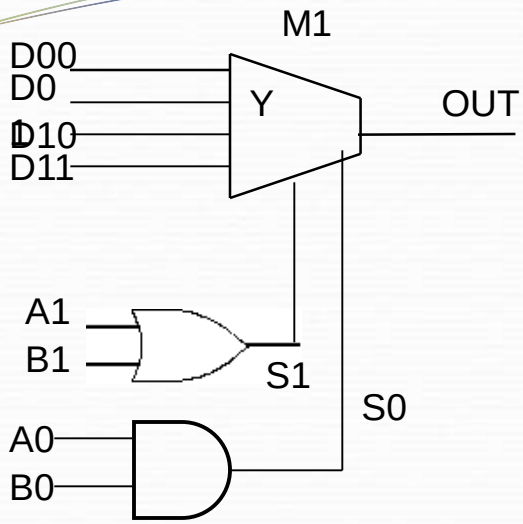
www.rejinpaul.com

1.2)ACT 2 and ACT 3 Logic Modules

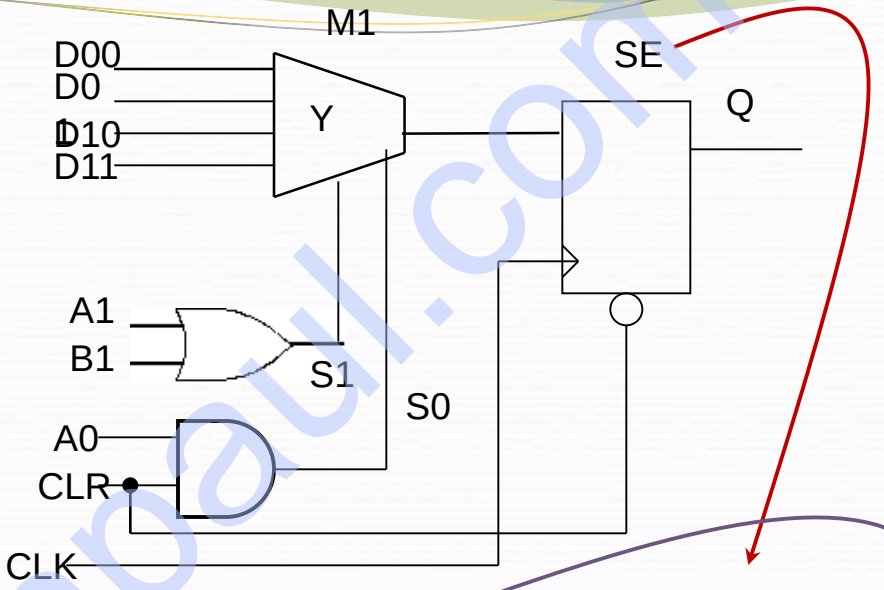


www.rejinpaul.com

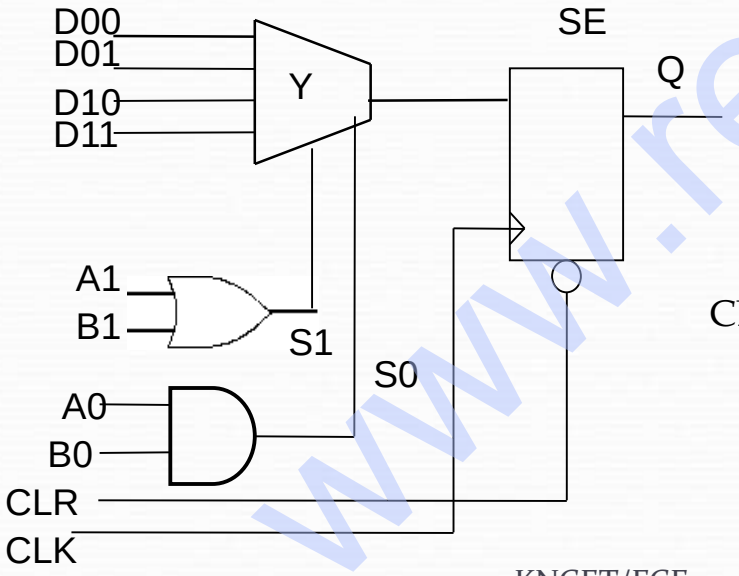
ACTEL ACT C-Module



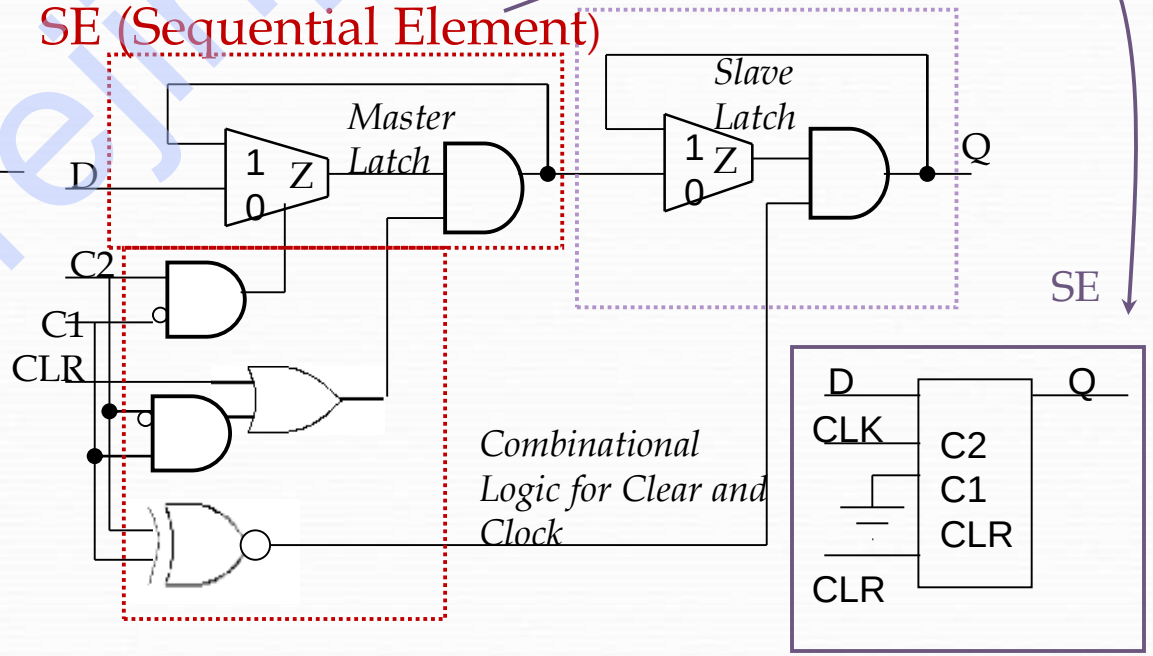
S-Module (ACT 2)



S-Module (ACT 3)



SE (Sequential Element)



www.rejinpaul.com

- The ACT 2 and ACT 3 architectures use two different types of Logic Modules, and one of them does include the equivalent of a D flip-flop.
- The ACT 2 C-Module is similar to ACT 1 but is capable of implementing five-input logic functions.
- The use of MUXes in the Actel Logic Modules (and in other places) can cause
- confusion in using and creating logic macros.
- For the Actel library, setting S = '0' selects input A of a two-input MUX.
- For other libraries setting S = '1' selects input A.
- This can lead to some very hard to find errors when moving schematics between libraries.
- Similar problems arise in flip-flops and latches with MUX inputs.
- A safer way to label the inputs of a two-input MUX is with '0' and '1', corresponding to the input selected when the select input is '1' or '0'.
- This notation can be extended to bigger MUXes

www.rejinpaul.com

Features of ACT2&ACT3Logic Module

- Have up to 8000 gate arrays.
- Have up to 1232 programmable logic modules.
- Have up to 998 flip-flops.
- Non-volatile and user programmable
- Fabricated with 1micron CMOS technology
- Single module sequential functions and wide input combinational functions

www.rejinpaul.com

2)Xilinx LCA

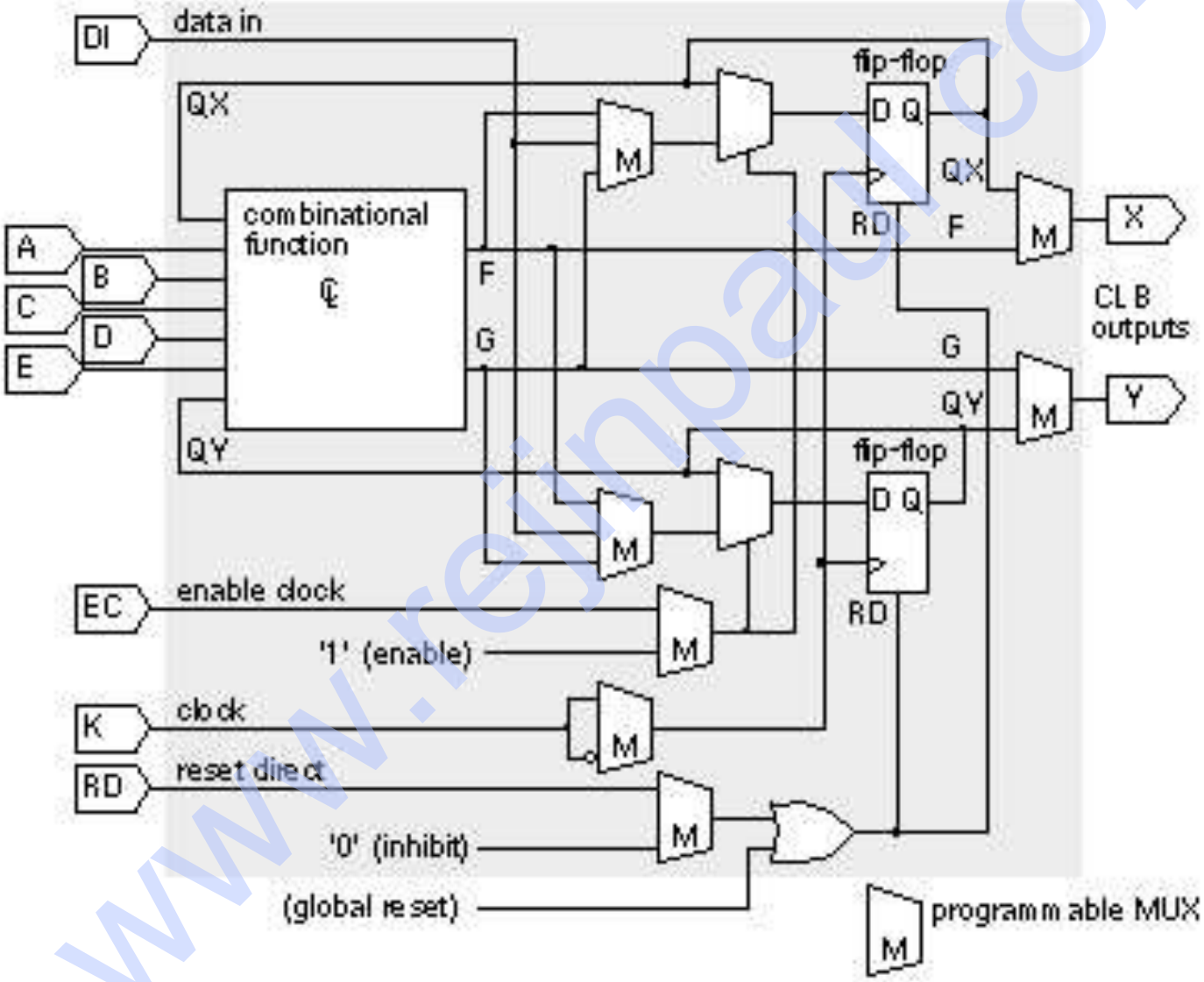
- Xilinx LCA basic logic cells are called the Configurable Logic Blocks or CLBs.
- CLBs are bigger and more complex than the Actel LCA uses coarse grain architecture.
- Xilinx CLBs contain both combinational logic and flip-flops.

Versions

1. XC3000
2. XC4000
3. XC5200

www.rejinpaul.com

2.1)XC3000 CLB



www.reinpaul.com

- The XC3000 CLB, has 5 logic inputs (A E), a common clock input (K), an asynchronous direct-reset input (RD), and an enable (EC).
- Independently connect each of the two CLB outputs (X and Y) to the output of the flip-flops (QX and QY) or to the output of the combinational logic (F and G).
- A 32-bit look-up table (LUT), stored in 32 bits of SRAM, provides the ability to implement combinational logic.
- Eg→To implement the function $F = A \cdot B \cdot C \cdot D \cdot E$
- Set the contents of LUT cell number 31 (with address '11111') in the 32-bit SRAM to a '1'; all the other SRAM cells are set to '0'.
- When you apply the input variables as an address to the 32-bit SRAM, only when ABCDE = '11111' will the output F be a '1'.

Method1

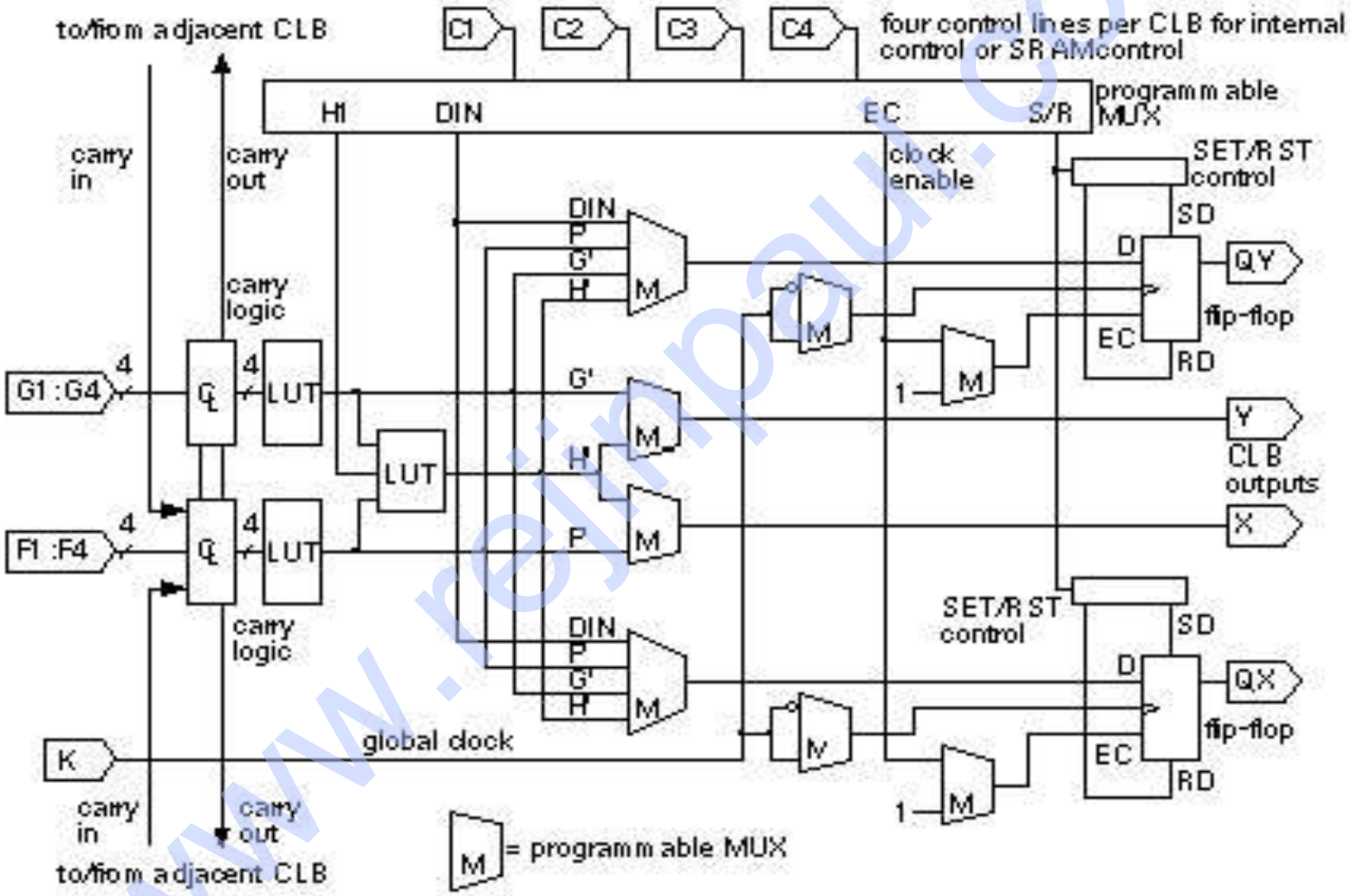
- Split the 32-bit LUT in half to implement two functions of four variables each. choose four input variables from the seven inputs (A E, QX, QY).
- choose two of the inputs from the five CLB inputs (A E); then one function output connects to F and the other output connects to G.

Method2

- Split the 32-bit LUT in half, using one of the seven input variables as a select input to a 2:1 MUX that switches between F and G. This allows you to implement some functions of six and seven variables.

www.rejinpaul.com

2.2)XC4000 Logic Block



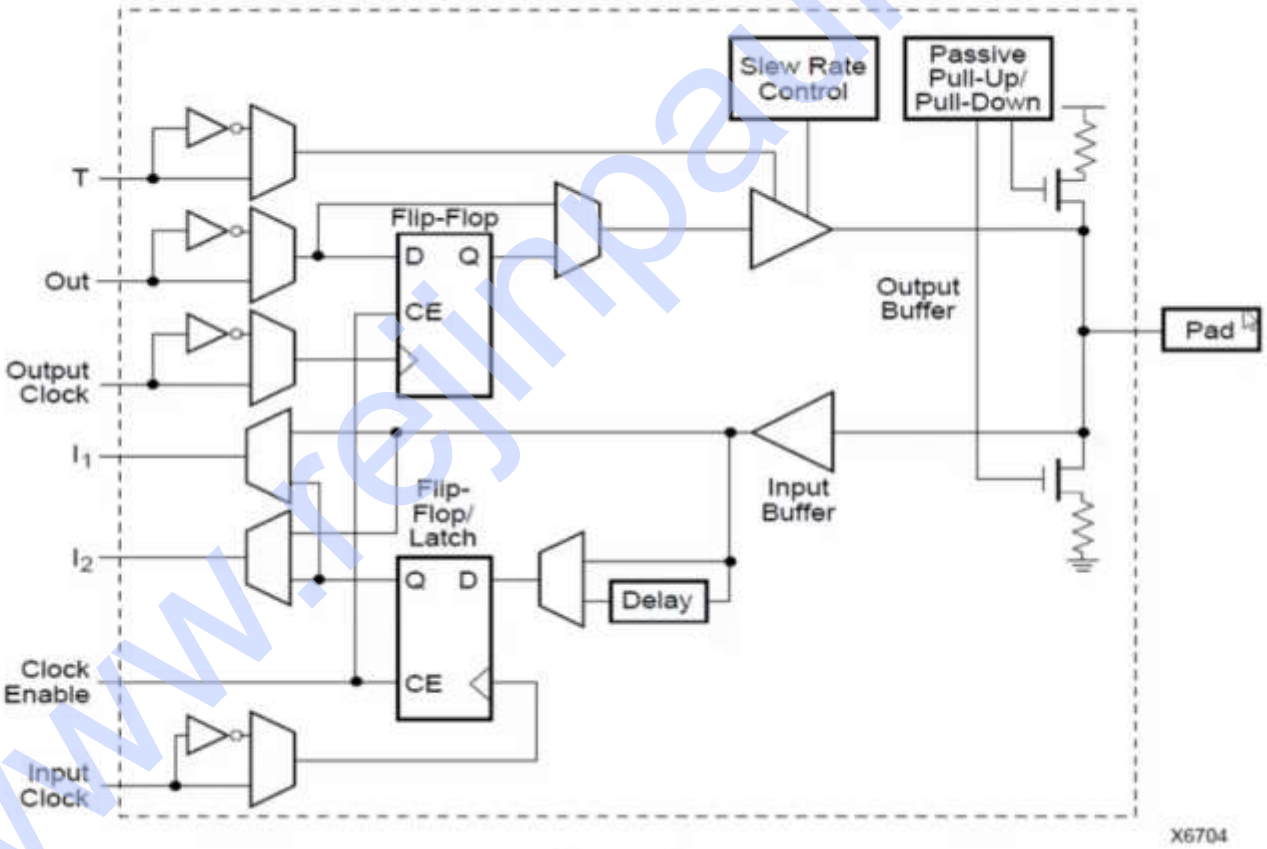
2^8=256bit SRAM

www.rejinpaul.com

- This is a fairly complicated basic logic cell containing 2 four-input LUTs that feed a three-input LUT.
- The XC4000 CLB also has special fast carry logic hard-wired between CLBs. MUX control logic maps four control inputs (C1 C_4) into the four inputs.
- LUT input H1, direct in (DIN), enable clock (EC), and a set / reset control (S/R) for the flip-flops.
- The control inputs (C1 C_4) can also be used to control the use of the F' and G' LUTs as 32 bits of SRAM.

www.rejinpaul.com

INPUT OUTPUTS (IO,s)



www.rejinpaul.com

INTERCONNECT

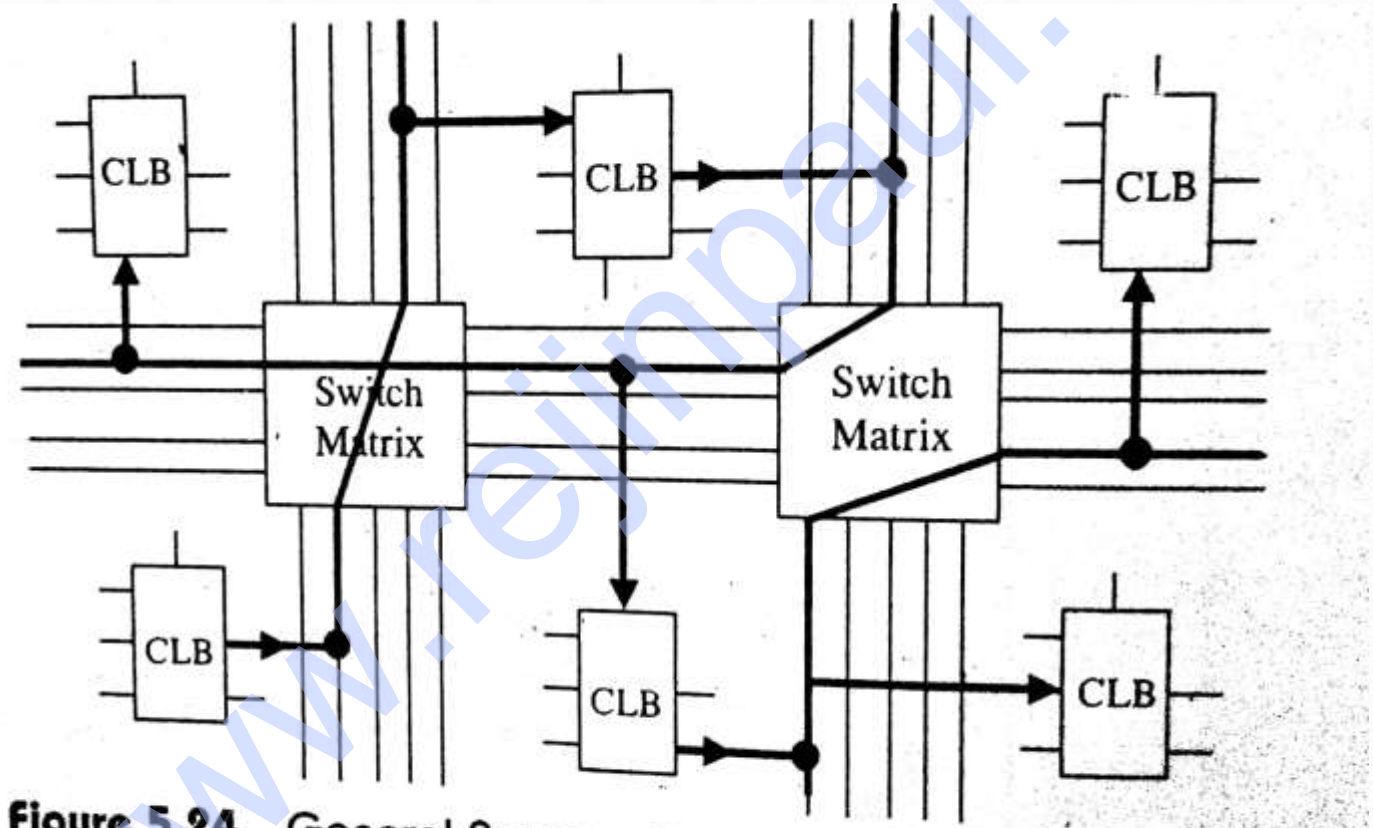
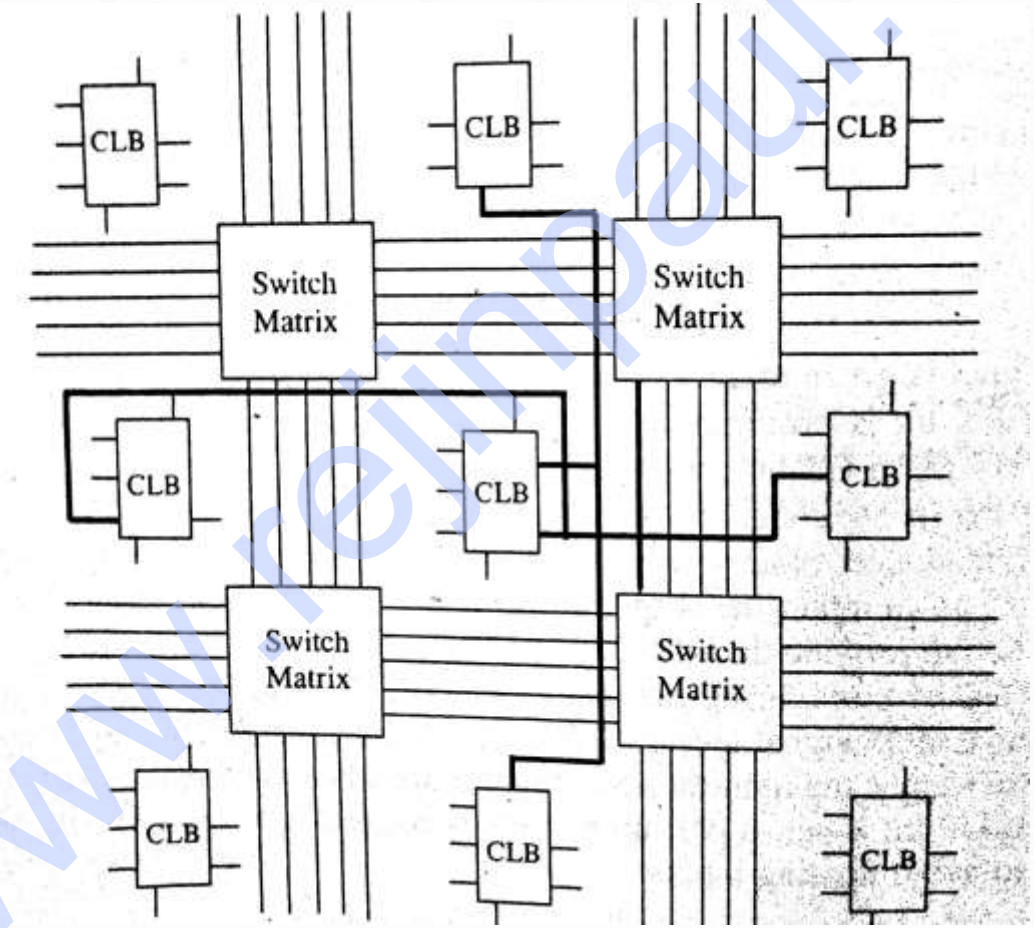


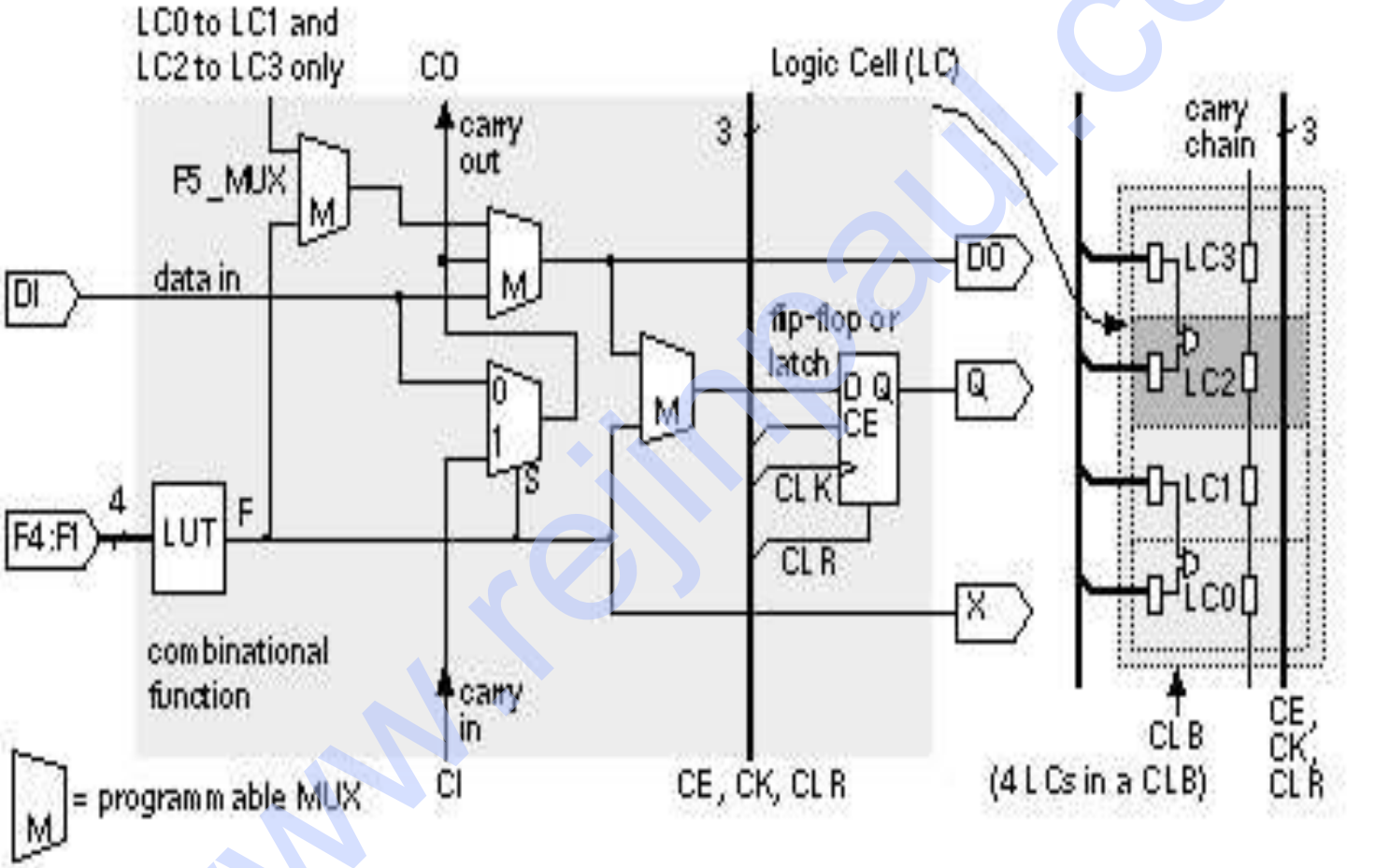
Figure 5.24 General Purpose Programmable Interconnects

www.rejinpaul.com



www.reinpaul.com

2.3)XC5200 Logic Block

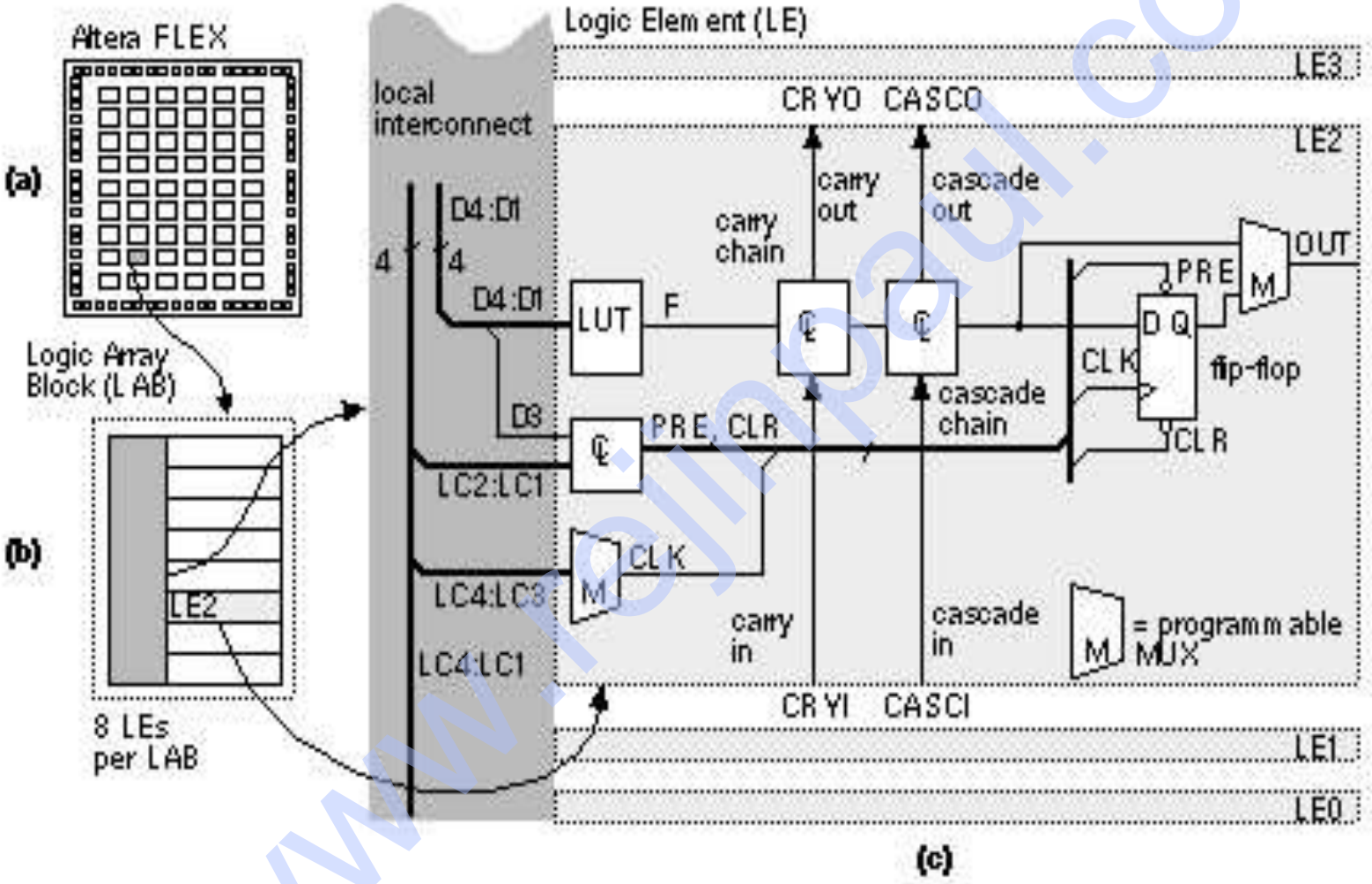


www.rejinpaul.com

- The LC is similar to the CLBs in the XC2000/3000/4000 CLBs, but simpler.
- Xilinx retained the term CLB in the XC5200 to mean a group of four LCs (LCo LC3).
- The XC5200 LC contains a four-input LUT, a flip-flop, and MUXes to handle signal switching.
- The arithmetic carry logic is separate from the LUTs.
- A limited capability to cascade functions is provided to gang two LCs in parallel to provide the equivalent of a five-input LUT.

www.rejinpaul.com

3)Altera FLEX



www.rajnpaul.com

- The basic logic cell of ALTERA FLEX has a Logic Element (LE), that Altera uses in its FLEX 8000 series of FPGAs.
- Apart from the cascade logic (which is slightly simpler in the FLEX LE) the FLEX cell resembles the XC5200 LC architecture.
- Both architectures are based on the same SRAM programming technology.
- The FLEX -LE uses a four-input LUT, a flip-flop, cascade logic, and carry logic.
- Eight LEs are stacked to form a Logic Array Block(LAB)

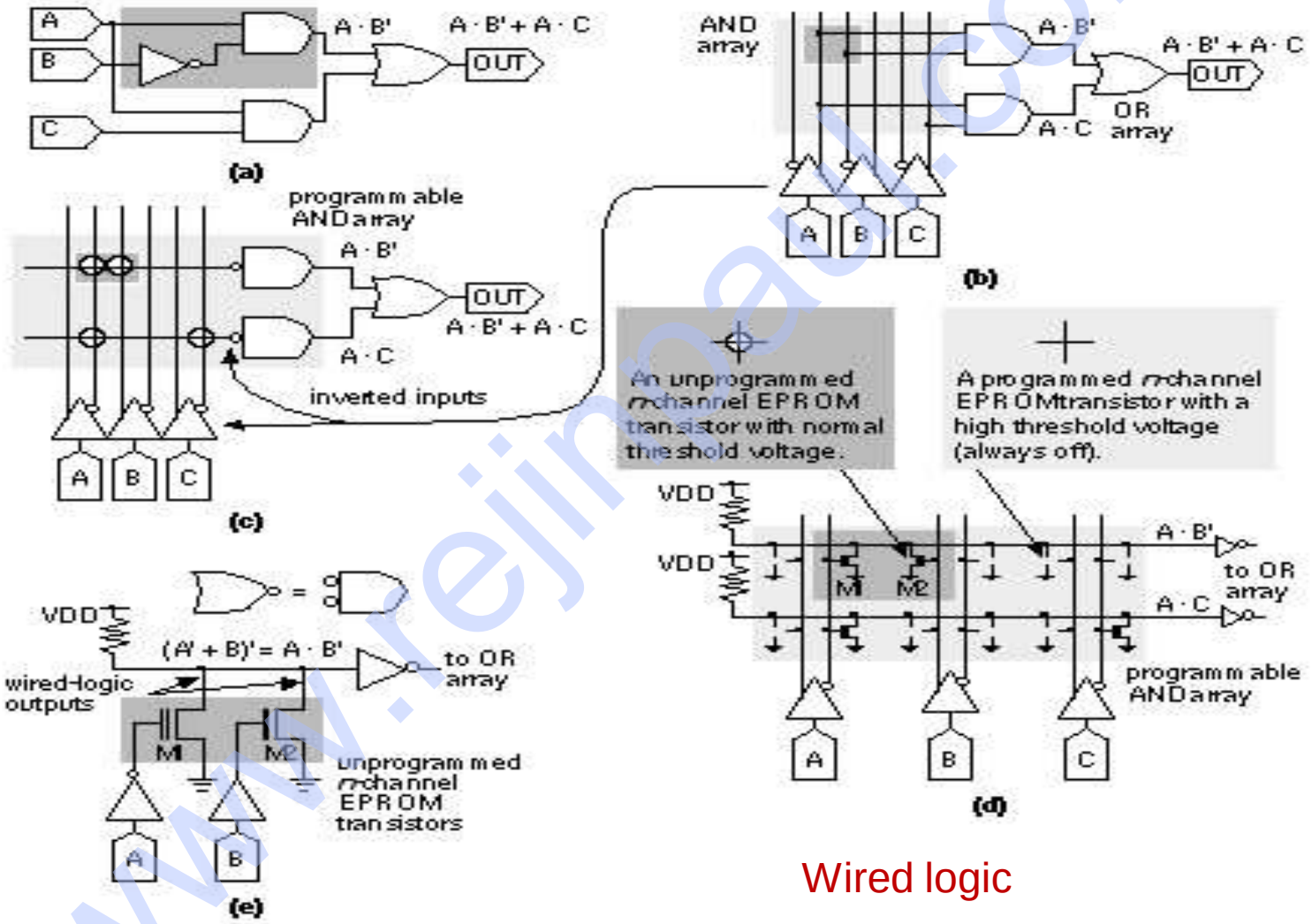
www.rejinpaul.com

4)Altera MAX

Two level logic array

Sumof products

Programmable AND plane



EPROM logic Array

Wired logic

www.rejinpaul.com

- Fig(a)→Two-level logical circuit using a regular structure.
- Which contains a vector of buffers, followed by a vector of AND gates (which construct the product terms) that feed OR gates (which form the sums of the product terms)
- Fig(c)→This representation is simplified by drawing the input lines to a multiple-input AND gate as if they were one horizontal wire, which is known as product-term line.
- The vertical input line is called the word line.
- Fig (d) →The horizontal product-term lines connect to the vertical input lines using the EPROM transistors as pull-downs at each possible connection.
- Applying a '1' to the gate of an unprogrammed EPROM transistor pulls the product-term line low to a '0'.
- A programmed n -channel transistor has a threshold voltage higher than V_{DD} and is therefore always off . Thus a programmed transistor has no effect on the product-term line.
- Fig(e) produces a wired-logic function the output is high only if all of the outputs are high, resulting in a wired-AND function of the outputs.
- The product term line is low when any of the inputs are high.

www.rejinpaul.com

FPGA interconnect routing procedures

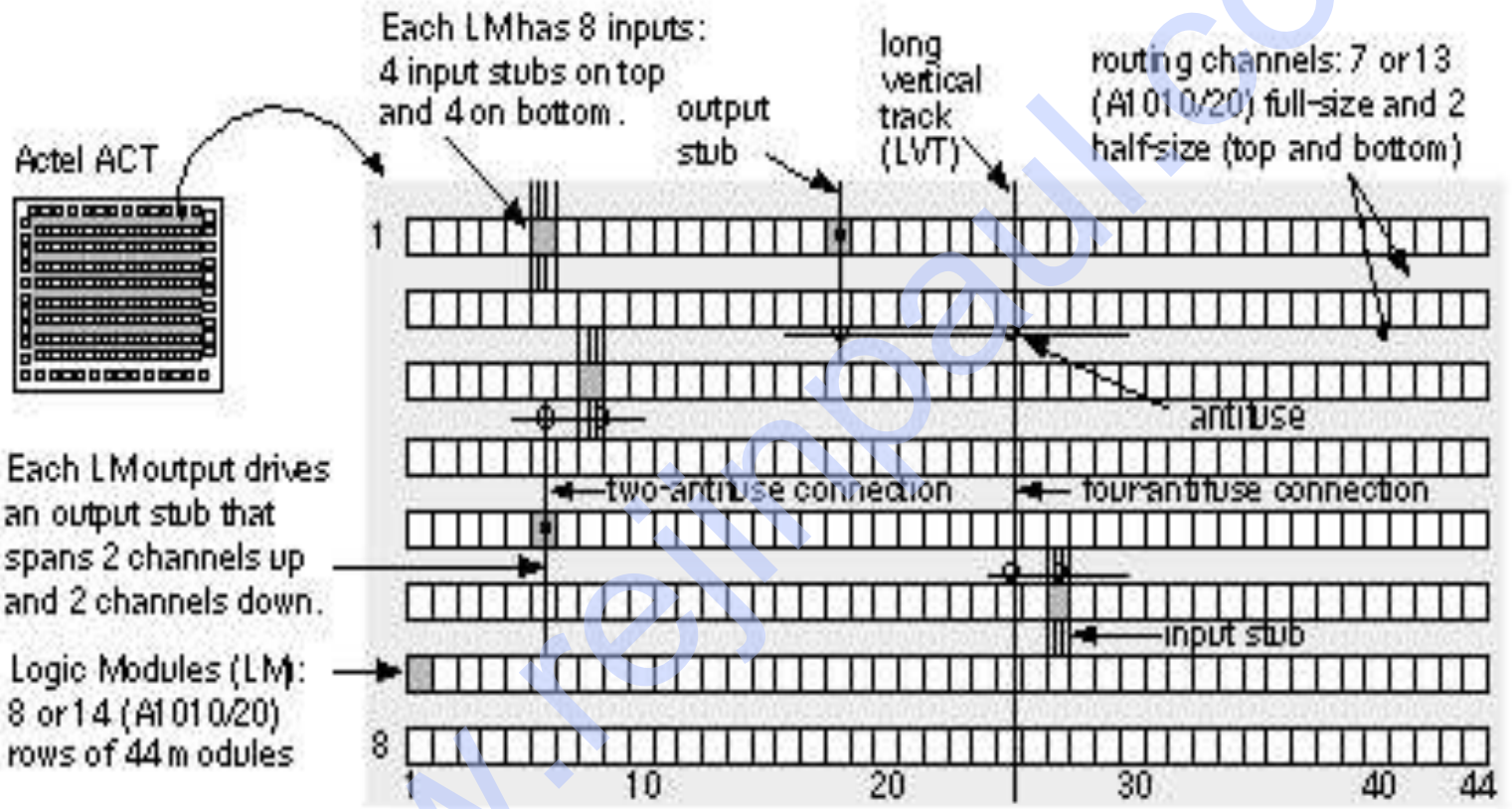
- All FPGAs contain some type of programmable interconnect .
- The structure and complexity of the interconnect is largely determined by the programming technology and the architecture of the basic logic cell.
- The raw material building the interconnect is aluminum-based metallization, which has a sheet resistance of approximately 50 m W /square and a line capacitance of 0.2 pFcm [1] .
- The first programmable ASICs were constructed using two layers of metal.
- Newer programmable ASICs use three or more layers of metal interconnect.

Types of Interconnecting architectures

1. Actel Interconnect
2. Xilinx LCA
3. XILINX EPLD
4. ALTERA MAX(5000, 7000,9000)
5. ALTERA MAX

www.rejimpaul.com

1)Actel ACT

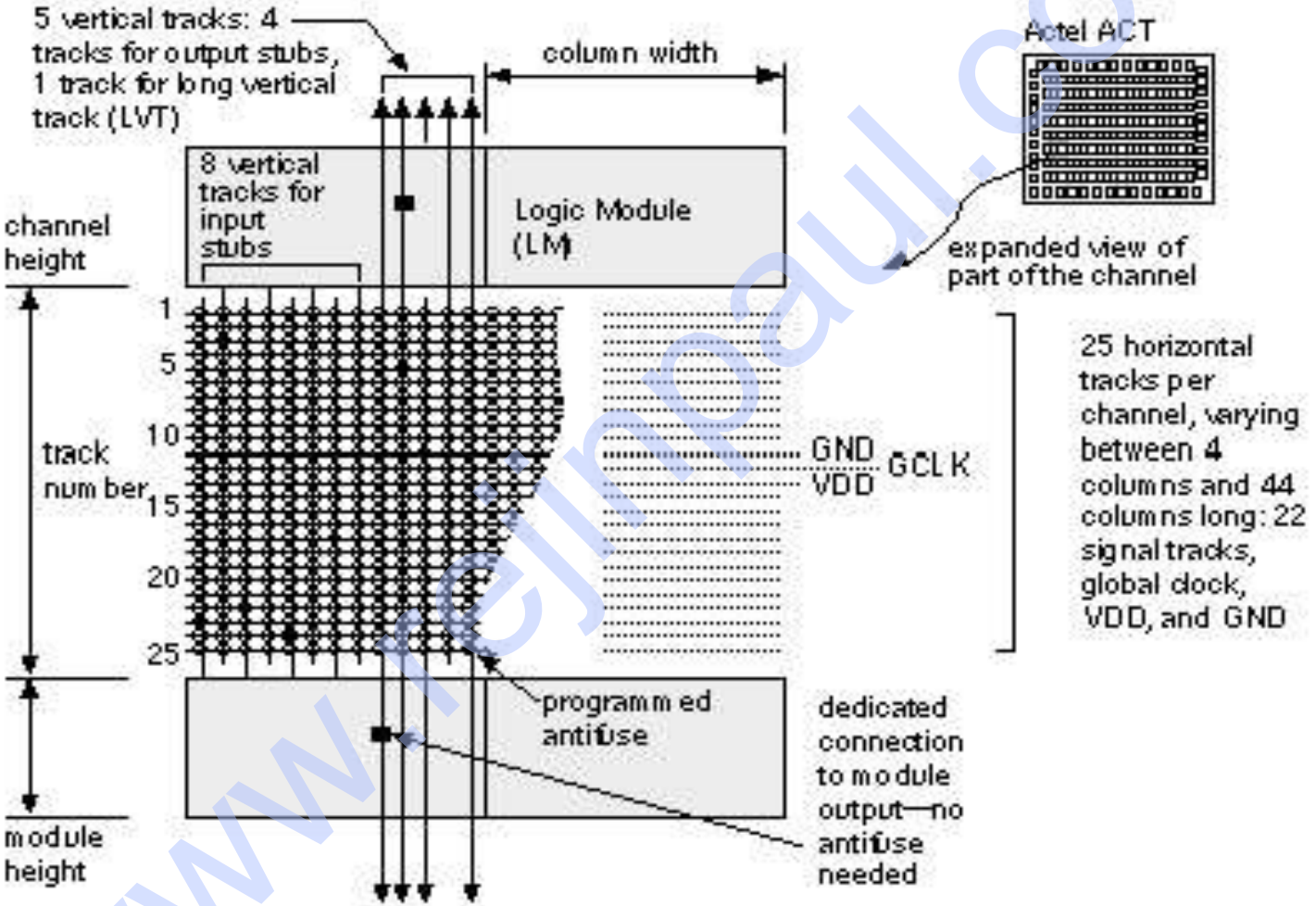


www.learnbaud.com

- The Actel ACT family interconnect is similar to a channeled gate array.
- The channel routing uses dedicated rectangular areas of fixed size within the chip called wiring channels (or just channels).
- The horizontal channels run across the chip in the horizontal direction.
- In the vertical direction there are similar vertical channels that run over the top of the basic logic cells, the Logic Modules. Within the horizontal or vertical channels wires run horizontally or vertically, respectively, within tracks .
- Each track holds one wire. The capacity of a fixed wiring channel is equal to the number of tracks it contains.

www.rejinpaul.com

1.1)ACT 1 Interconnect



www.rejinagul.com

- The ACT 1 interconnection architecture uses 22 horizontal tracks per channel for signal routing with three tracks dedicated to VDD, GND, and the global clock (GCLK), making a total of 25 tracks per channel.
- Horizontal segments vary in length from four columns of Logic Modules to the entire row of modules called long lines
- Four Logic Module inputs are available to the channel below the Logic Module and four inputs to the channel above the Logic Module.
- Eight vertical tracks per Logic Module are available for inputs (four from the Logic Module above the channel and four from the Logic Module below). These are the input of stubs.
- The single Logic Module output connects to a vertical track that extends across the two channels above the module and across the two channels below the module. This is the output stub.
- Thus module outputs use four vertical tracks per module (counting two tracks from the modules below, and two tracks from the modules above each channel).
- One vertical track per column is a long vertical track (LVT) that spans the entire height of the chip (the 1020 contains some segmented LVTs).
- There are thus a total of 13 vertical tracks per column in the ACT 1 architecture (eight for inputs, four for outputs, and one for an LVT).

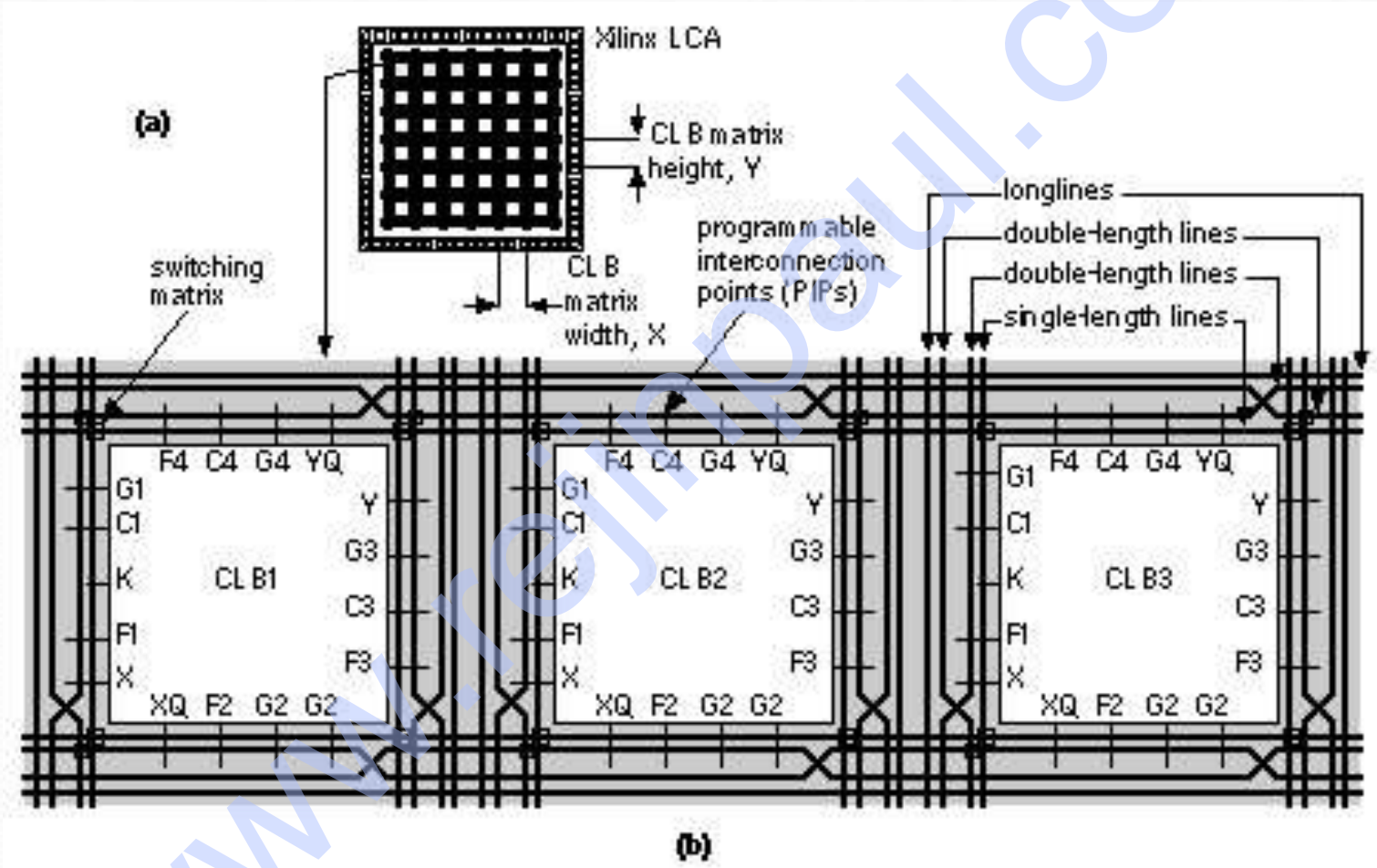
www.rejinpaul.com

1.2)ACT 2 and ACT 3 Interconnect

- The ACT 2 and ACT 3 architectures use increased interconnect resources .
- This reduces further the number of connections that need more than two antifuses. Delay is also reduced by decreasing the population of antifuses in the channels, and by decreasing the antifuse resistance of certain critical antifuses (by increasing the programming current).
- The channel density is the absolute minimum number of tracks needed in a channel to make a given set of connections .
- The ACT 2 devices have 36 horizontal tracks per channel rather than the 22 available in the ACT 1 architecture.
- Horizontal track segments in an ACT 3 device range from a module pair to the full channel length.
- Vertical tracks are input (one up, one down); output (two up, two down); and long Four LVTs are shared by each column pair.
- The ACT 2/3 Logic Modules can accept five inputs, than 4-inputs for ACT1.
- This Logic Modules need an extra two vertical tracks per channel.
- The number of tracks per column thus increases from 13 to 15 in the ACT 2/3 architecture.

www.rejinpaul.com

2)Xilinx LCA



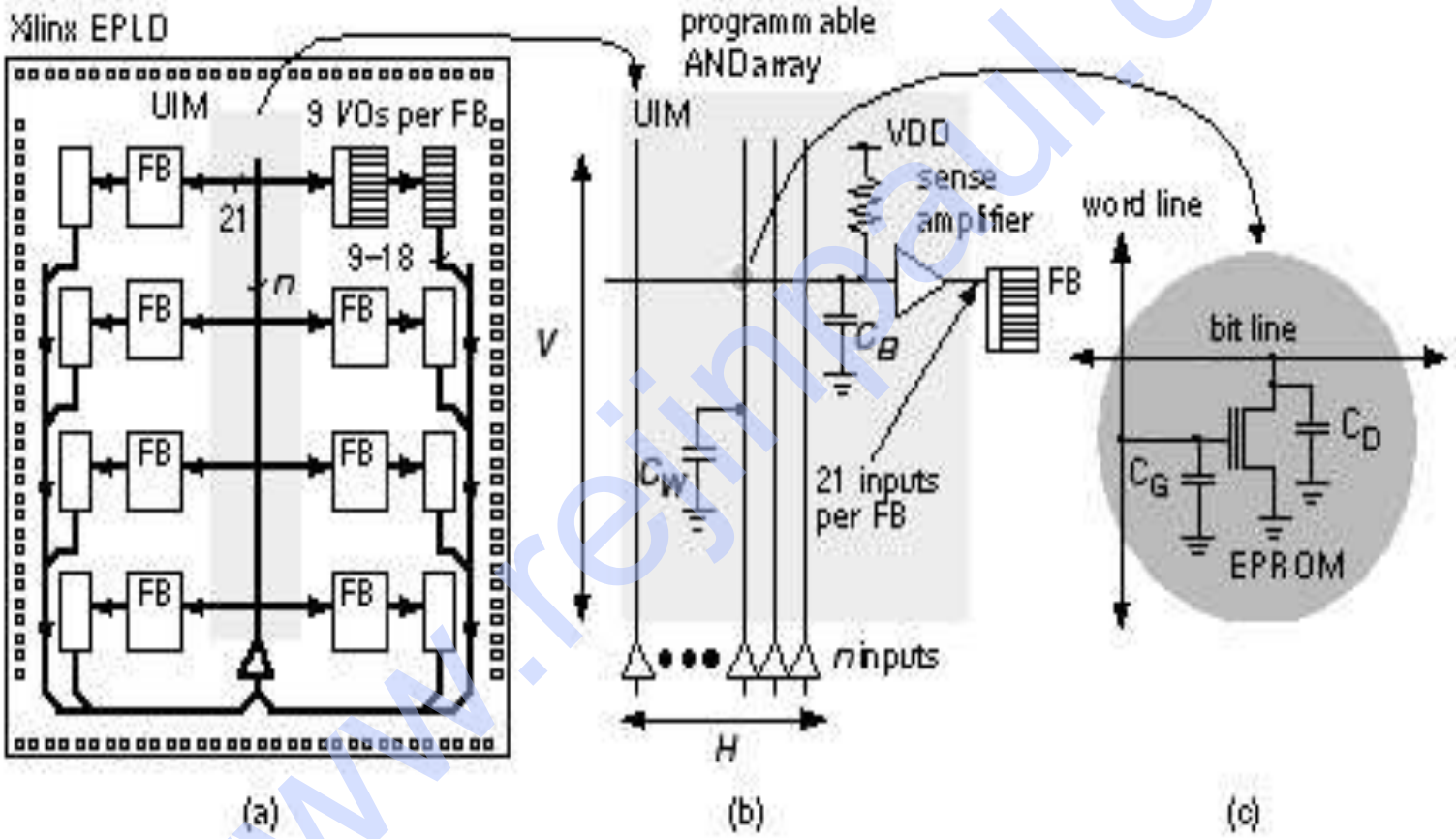
www.rejinpaul.com

- Features

1. The vertical lines and horizontal lines run between CLBs.
2. The general-purpose interconnect joins switch boxes (also known as magic boxes or switching matrices).
3. The long lines run across the entire chip. It is possible to form internal buses using long lines and the three-state buffers that are next to each CLB.
4. The direct connections (not used on the XC4000) bypass the switch matrices and directly connect adjacent CLBs.
5. The Programmable Interconnection Points (PIP s) are programmable pass transistors that connect the CLB inputs and outputs to the routing network.
6. The bidirectional (BIDI) interconnect buffers restore the logic level and logic strength on long interconnect paths.

www.rejinpaul.com

3) Xilinx EPLD

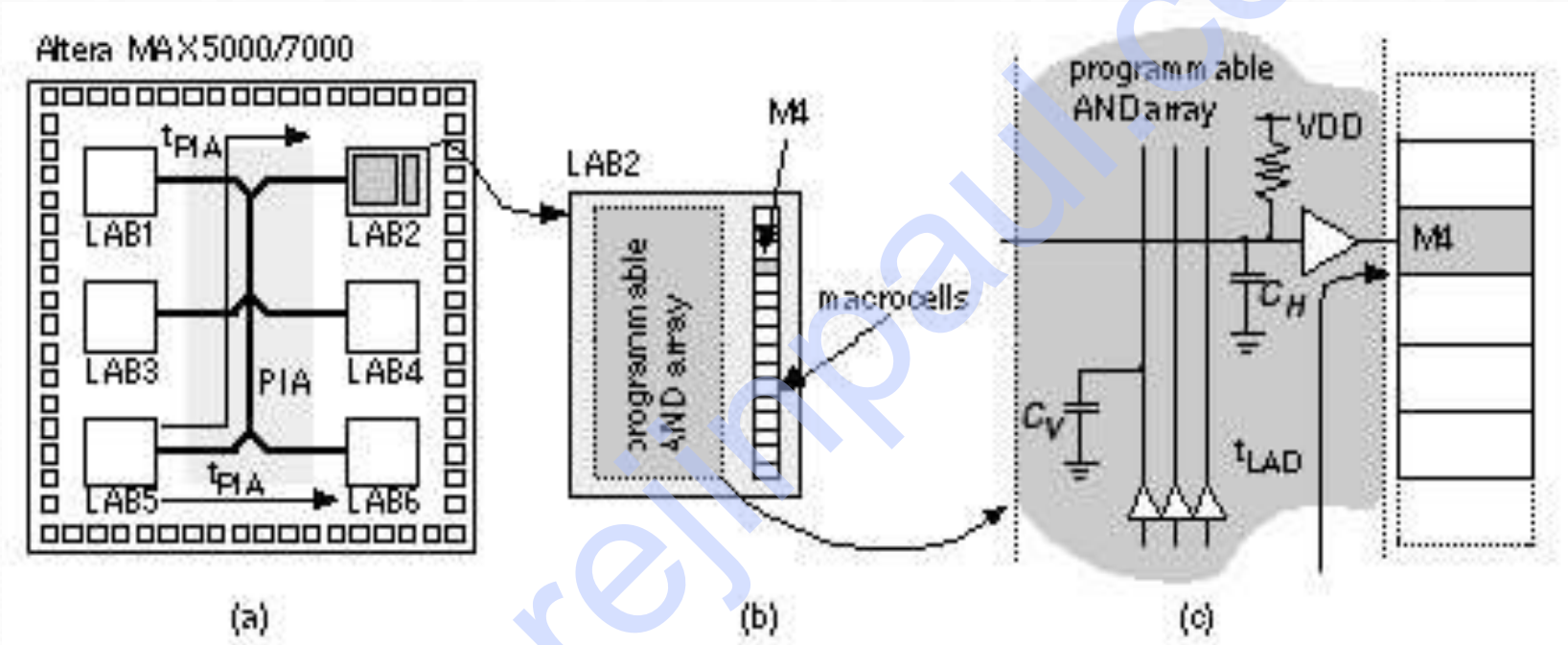


www.reinpaul.com

- The Xilinx EPLD family uses an interconnect bus known as Universal Interconnection Module (UIM) to distribute signals within the FPGA.
- Which uses programmable AND array with constant delay from any input to any output.
- $C_{G-}$is the fixed gate capacitance of the EPROM device.
- C_D-fixed drain parasitic capacitance of the EPROM device.
- C_B-variable horizontal bus (bit line) capacitance.
- C_W-is the variable vertical bus (word line) capacitance.
- UIM has 21 output connections to each FB.
- The XC7272 UIM (with a 4x2 array of eight FBs and has 168 (8 x21) output connections.
- Most of the nine I/O cells attached to each FB have two input connections to the UIM, one from a chip input and one feedback from the macrocell output.

www.reinpaul.com

4)Altera MAX 5000 and 7000



www.rejinpaul.com

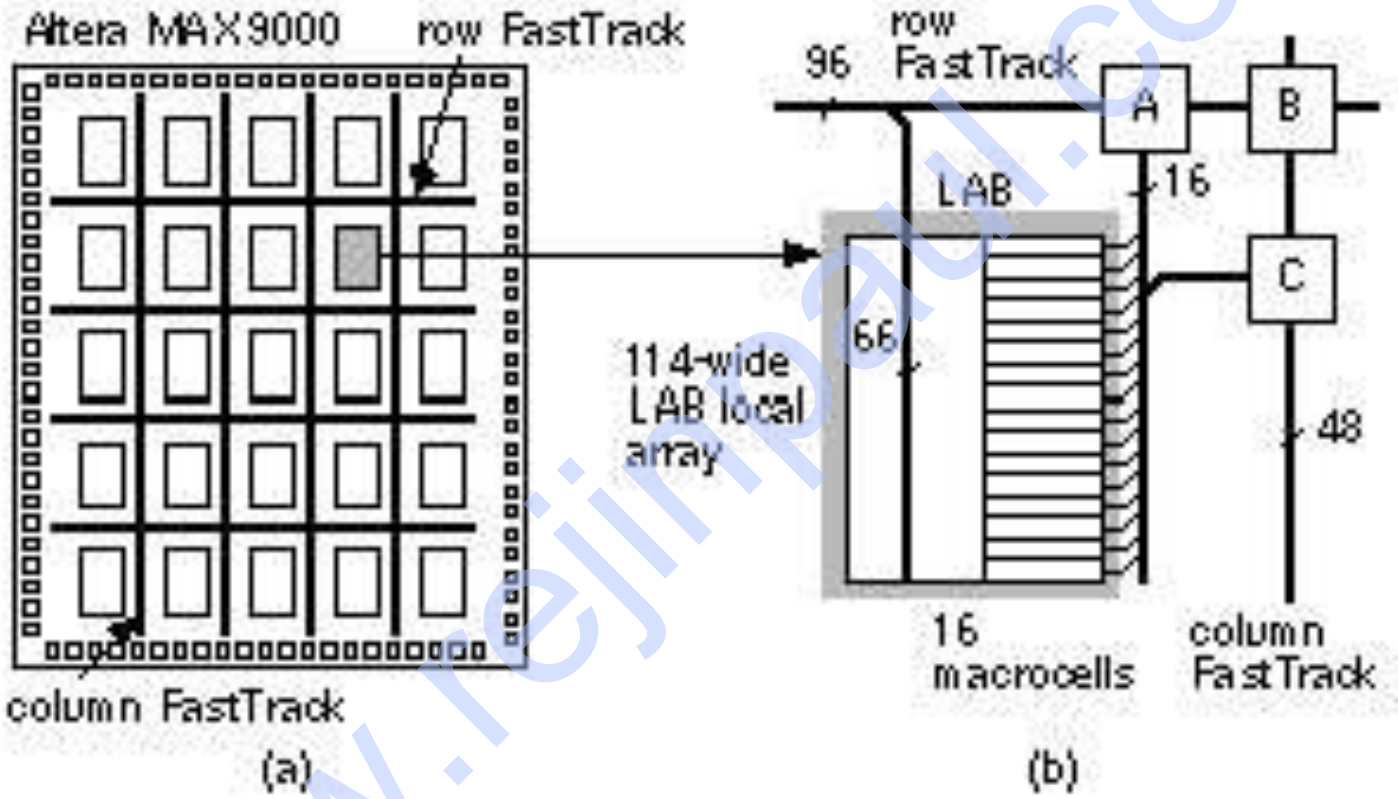
- Altera MAX 5000 devices (except the EPM5032, which has only one LAB) and all MAX 7000 devices use a Programmable Interconnect Array (PIA).

Features

- which uses a fixed number of connections over programmable interconnection schemes
- Fixed routing delay.
- Interconnect structure is Simplified and improved speed of the placement and routing software.
- The delay between LAB1 and LAB2 is the same as the delay between LAB1 and LAB6

www.rejinpaul.com

5)Altera MAX 9000

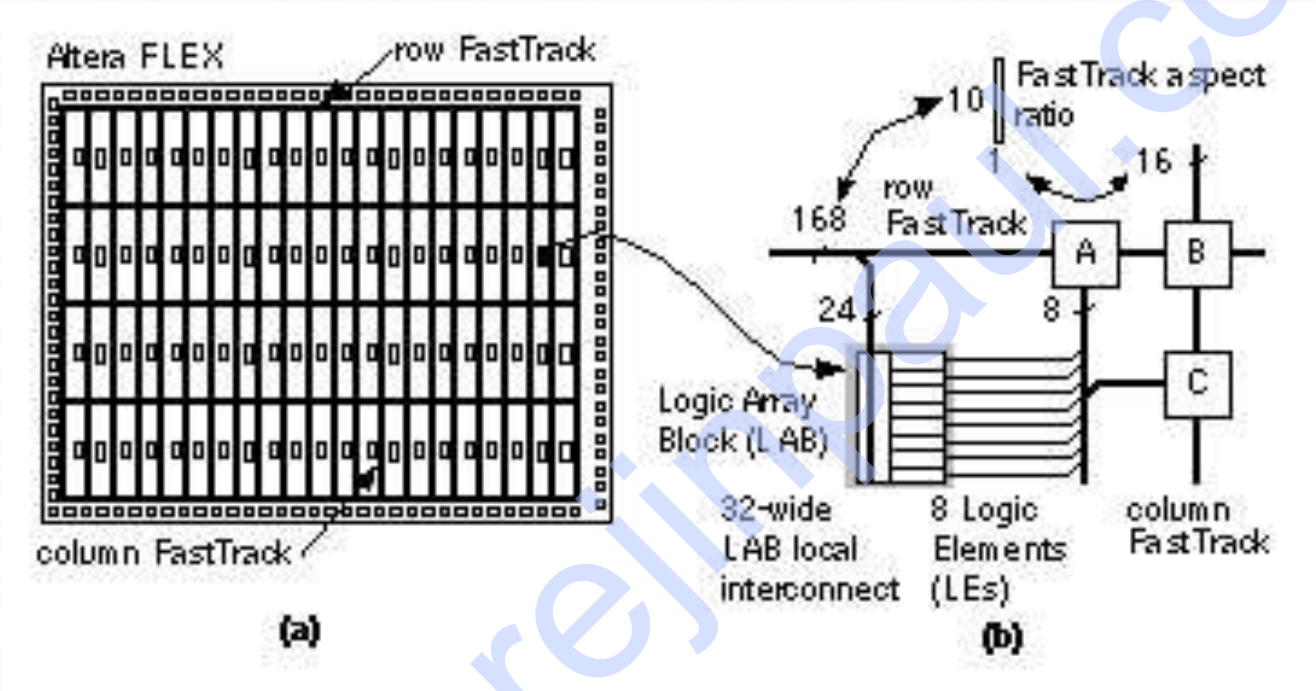


www.rejinpaul.com

- The size of the MAX 9000 LAB arrays varies between 4 x 5 (rows x columns) for the EPM9320 and 7 x5 for the EPM9560.
- The MAX 9000 is an extremely coarse-grained architecture.
- This uses complex PLDs, but the LABs themselves have a finer structure.
- The Boxes A,B and C represent the interconnection between the Fast track buses and the 16 macro cells in each LAB.
- Box A connects a macro cell to one row channel.
- Box B connects three column channels to two row channels.
- Box C connects a macro cell to three column channels.

www.rejinpaul.com

6)Altera FLEX



www.rejinpaul.com

- The interconnect used in the Altera FLEX family of complex PLDs.
- Altera refers to the FLEX interconnect and MAX 9000 interconnect by the same name, FastTrack.
- These two are different because the granularity of the logic cell arrays is different.
- The FLEX architecture is of finer grain than theMAX arrays because of the difference in programming technology.
- The FLEX horizontal interconnect is much denser (at 168 channels per row) than the vertical interconnect (16 channels per column).
- Which creating an aspect ratio for the interconnect of over 10:1 (168:16).
- This imbalance is partly due to the aspect ratio of the die, the array, and the aspect ratio of the basic logic cell, the LAB.
- As an example, the EPF8820 has 4 rows and 21 columns of LABs
- - Horizontal channels = 4 rows x 168 channels/row = 672 channels.
  - Vertical channels = 21 rows x 16 channels/row = 336 channels.
- It appears that there is still approximately twice (672:336) as much interconnect
- Box A connects an LE to two row channels.
- Box B connects two column channels to a row channel.
- Box C connects an LE to two column channels.

www.rejinpaul.com