INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad -500 043

VLSI
ECE III-II SEM

Prof.V. R. Seshagiri Rao, Professor, ECE,
Dr. V. Vijay, Professor, ECE
Mr. D Khalandar Basha, Associate Professor, ECE
Ms. U. Dhanalakshmi, Assistant Professor, ECE

INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad -500 043

VLSI Design Lecture PPTs

Department	:	ELECTRONICS AND COMMUNICATIONENGINEERING			
Course Code	:	A60432			
Course Title	:	VLSI DESIGN			
Course Faculty	:	Prof. V. R. Seshagiri Rao, Dr. V. Vijay, Professor, Mr. D Khalandar Basha, Assistant Professor and Ms. U. Dhanalakshmi, Assistant Professor			
Course Structure	:	Lectures	Tutorials	Practicals	Credits
		4	1	-	4

Unit I
Introduction to IC technology

Topics

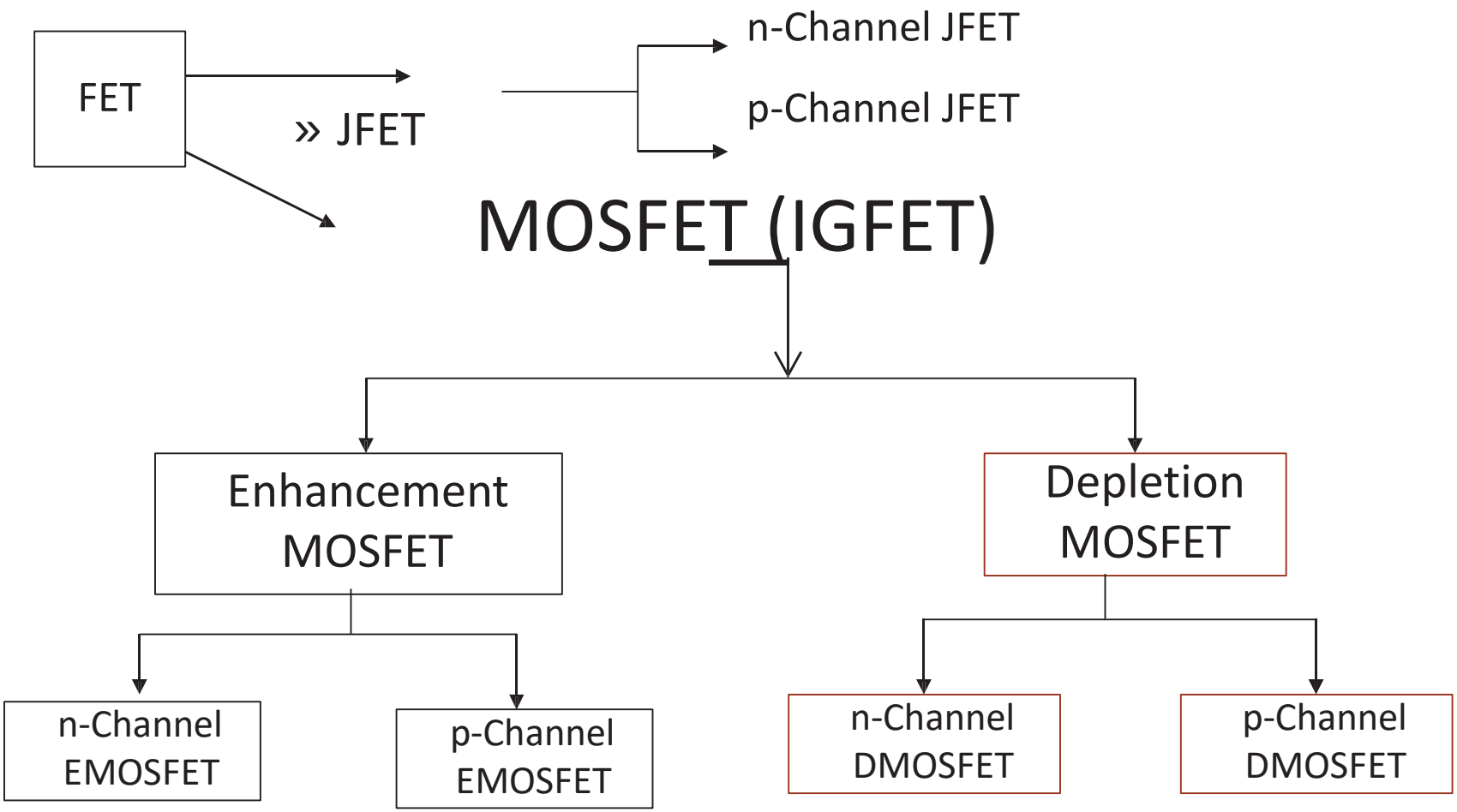
- MOS, PMOS, NMOS, CMOS and BiCMOS Technologies:
- Oxidation
- Lithography
- Diffusion
- Ion implantation
- Metallization
- Encapsulation
- Probe testing

Inegrated Resistors and Capacitors

Acronym of VLSI

- V -> Very
- L -> Large
- S -> Scale
- I -> Integration

Types of Field Effect Transistors (The Classification)



MOSFET (Metal-Oxide Semiconductor Field-Effect Transistor)

Primary component in high-density VLSI chips such as memories and microprocessors

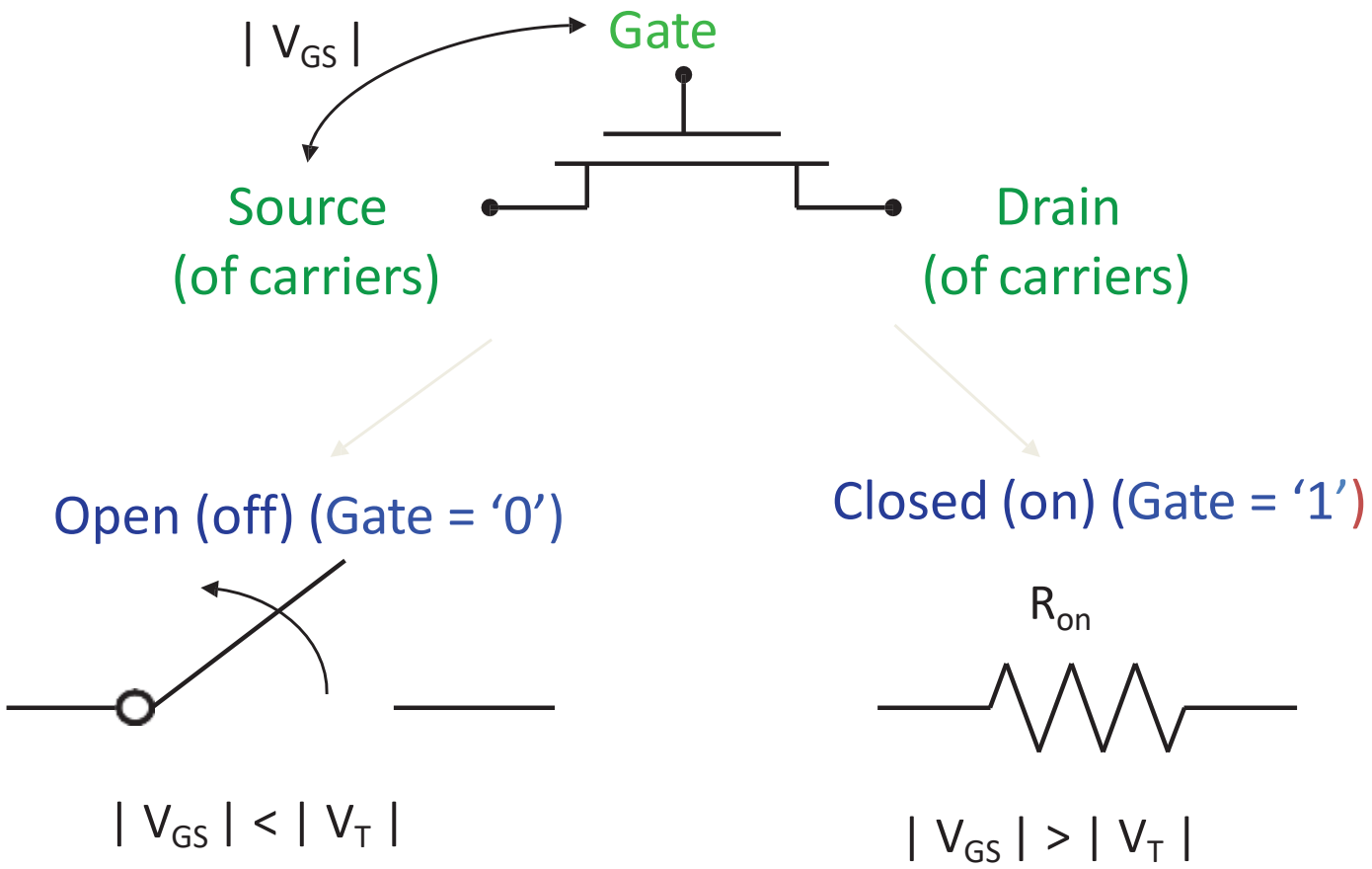
JFET (Junction Field-Effect Transistor)

Finds application especially in analog and RF circuit design

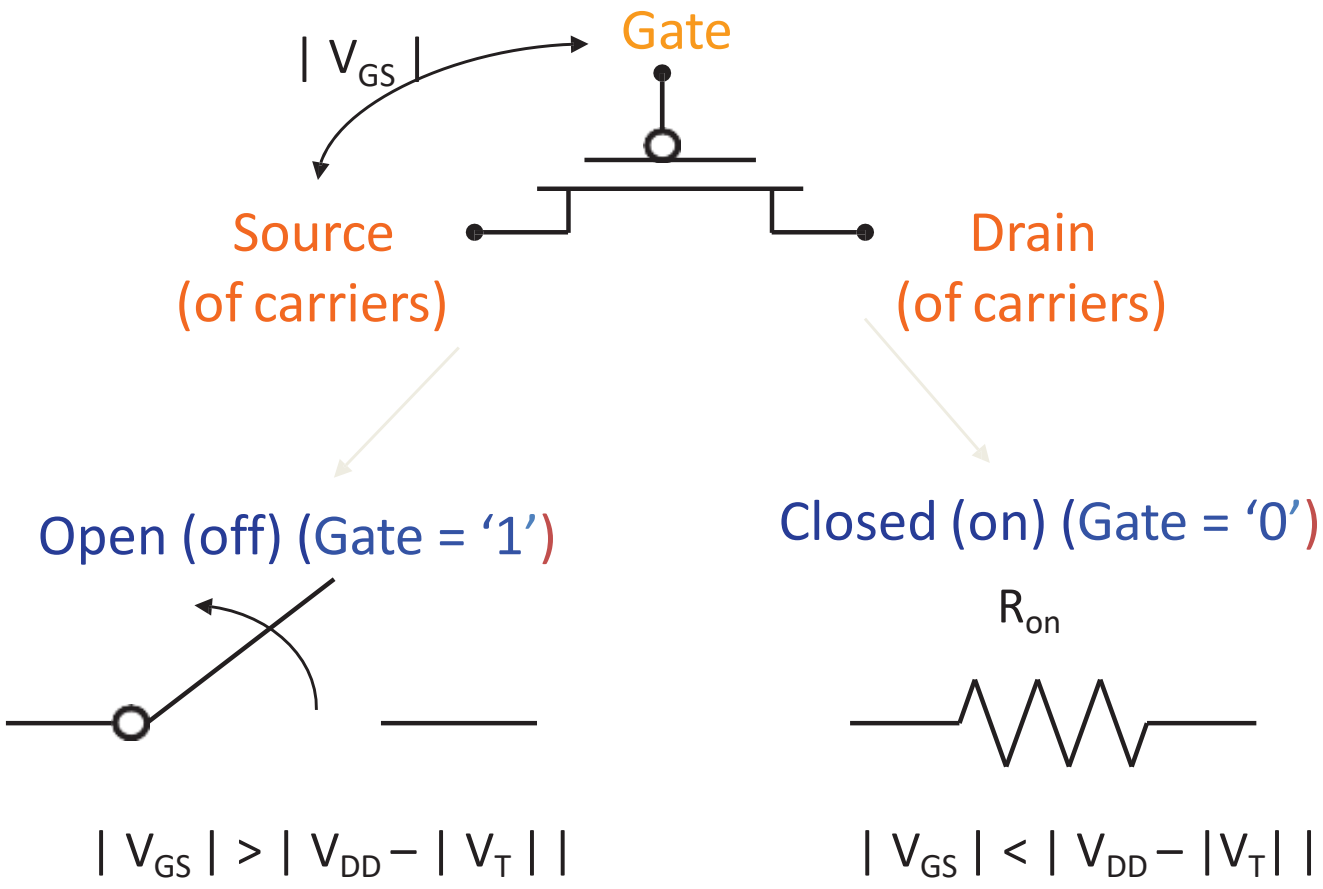
Metal Oxide Semiconductor(MOS)

- Advantages of FET over conventional Transistors
- Unipolar device i. e. operation depends on only one type of charge carriers (h or e)
- Voltage controlled Device (gate voltage controls drain current)
- Very high input impedance (=109-1012 ϕι)
- Source and drain are interchangeable in most Low-frequency applications
- Low Voltage Low Current Operation is possible (Low-power consumption)
- Less Noisy as Compared to BJT
- No minority carrier storage (Turn off is faster)
- Very small in size, occupies very small space in ICs

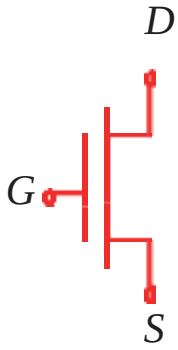
Switch Model of NMOS Transistor



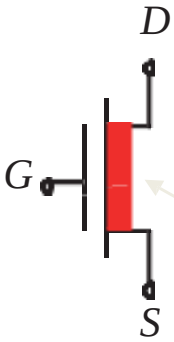
Switch Model of PMOS Transistor



MOS transistors Symbols

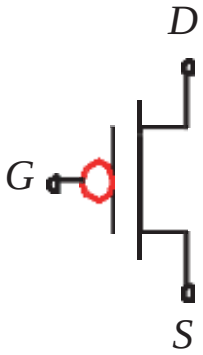


NMOS Enhancement

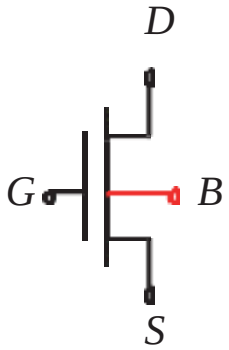


NMOS Depletion

Channel



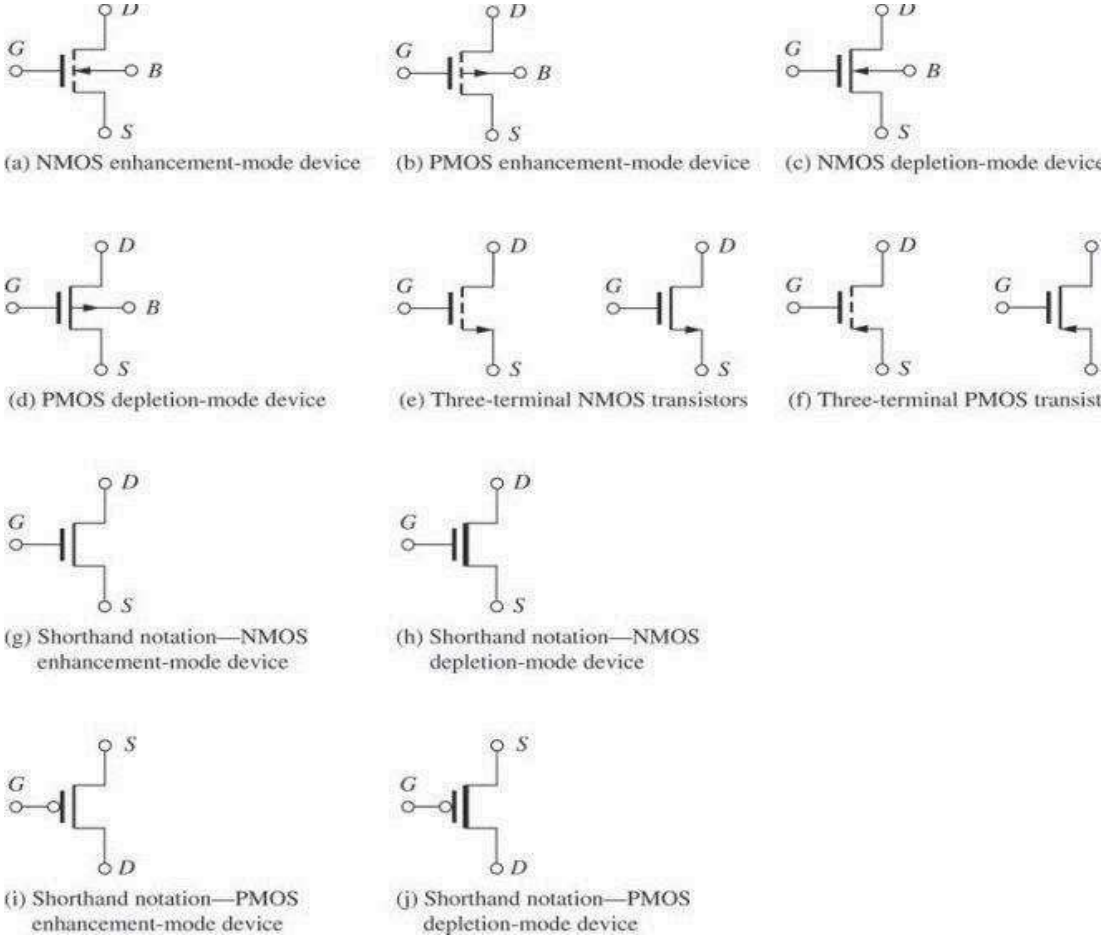
PMOS Enhancement



NMOS with Bulk Contact

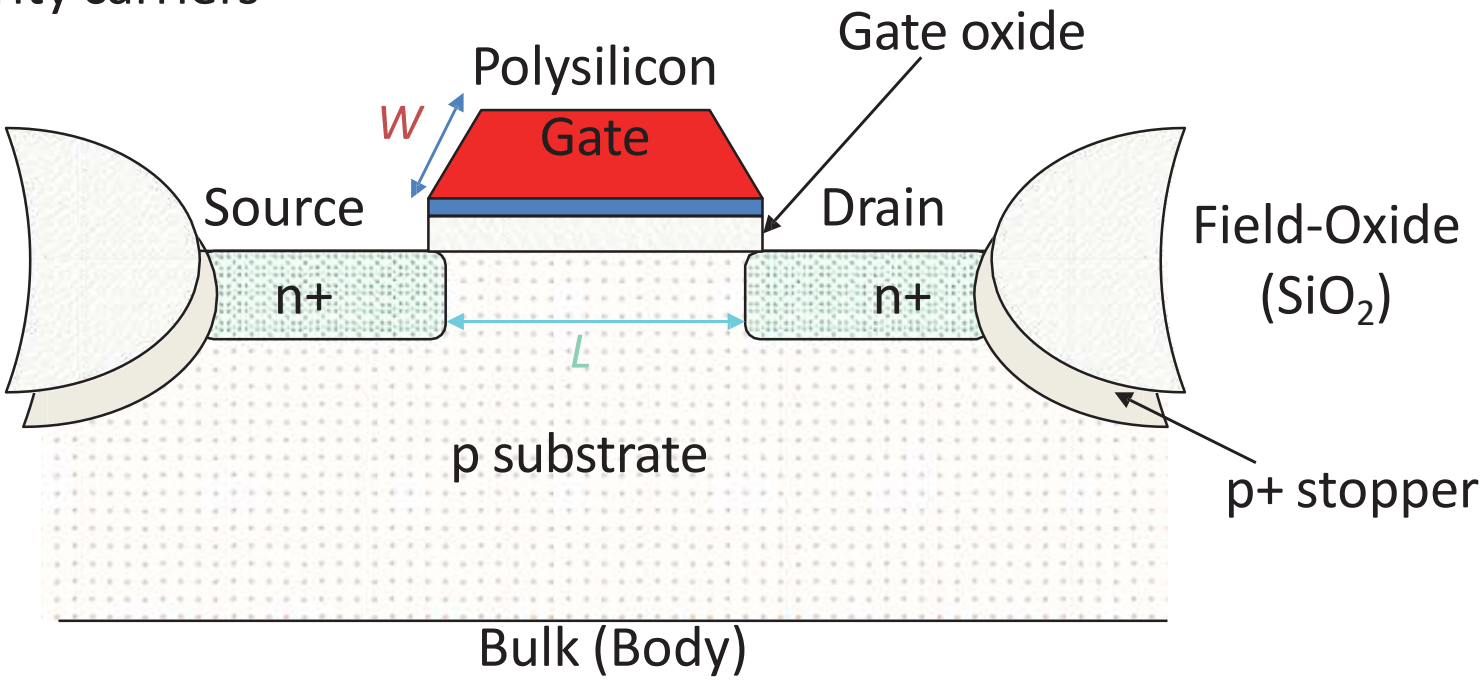
MOSFET Circuit Symbols

- (g) and (i) are the most commonly used symbols in VLSI logic design.
- MOS devices are symmetric.
- In NMOS, n+ region at higher voltage is the drain.
- In PMOS p+ region at lower voltage is the drain



The NMOS Transistor Cross Section

n areas have been doped with donor ions (arsenic) of concentration N_D - electrons are the majority carriers



p areas have been doped with acceptor ions (boron) of concentration N_A - holes are the majority carriers

Carriers and Current

- Carriers always flow from the Source to Drain
- NMOS: Free electrons move from Source to Drain.
  - Current direction is from Drain to Source.
- PMOS: Free holes move from Source to Drain.
  - Current direction is from Source to Drain.

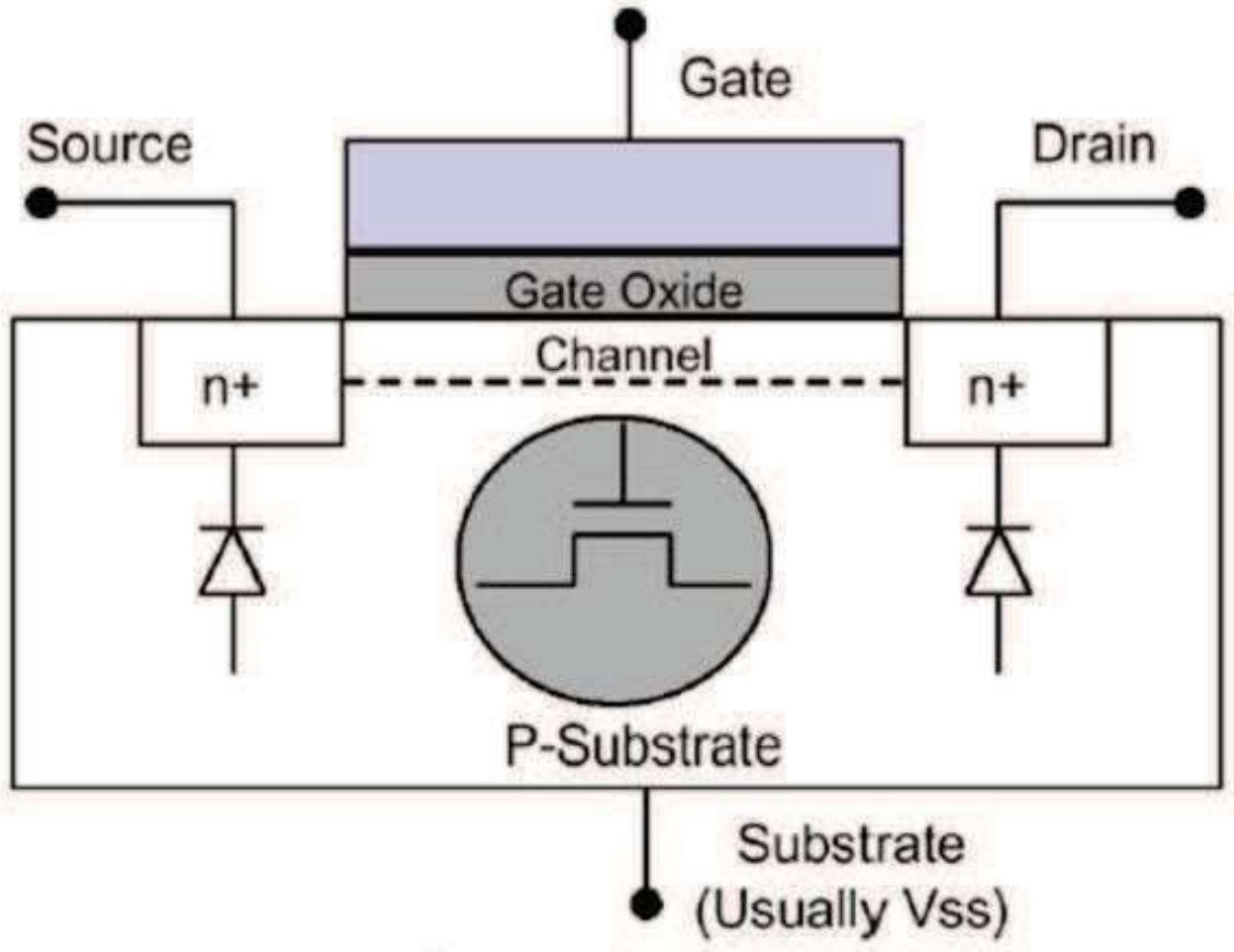
The MOSFET Channel

- Under certain conditions, a thin channel can be formed right underneath the Silicon-Dioxide insulating layer, electrically connecting the Drain to the Source. The depth of the channel (and hence its resistance) can be controlled by the Gate's voltage. The length of the channel (shown in the figures above as L) and the channel's width W, are important design parameters.

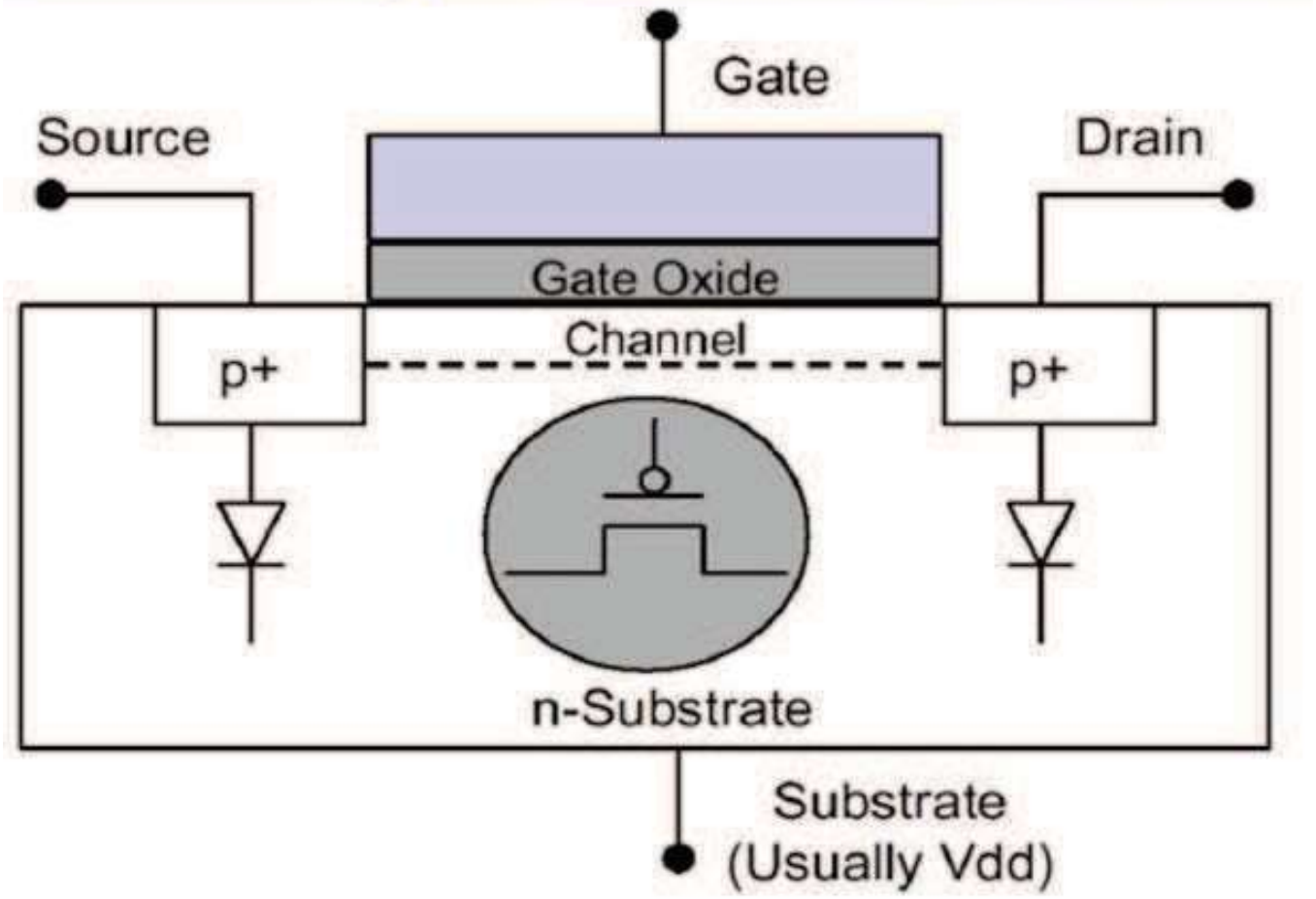
REGION OF OPERATION
CASE-1 (No Gate Voltage)

- Two diodes back to back exist in series.
- One diode is formed by the pn junction between the n+ drain region and the p-type substrate
- Second is formed by the pn junction between the n+ source region and the p-type substrate
- These diodes prevent any flow of the current.
- There exist a very high resistance.

NMos Cut View



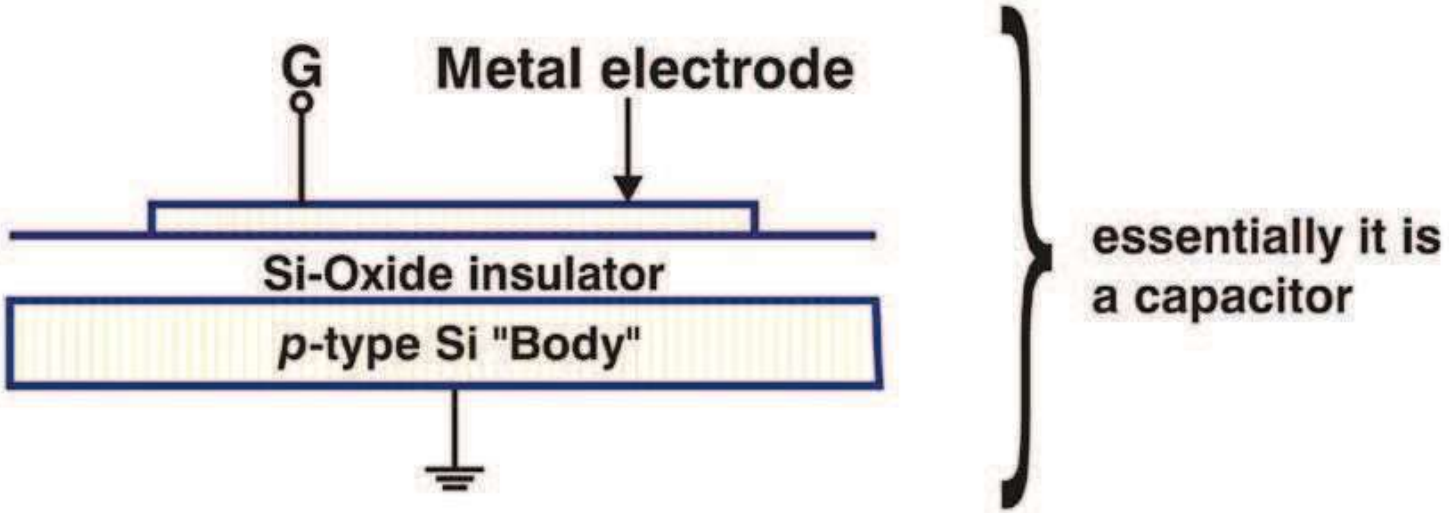
P-type Transistor



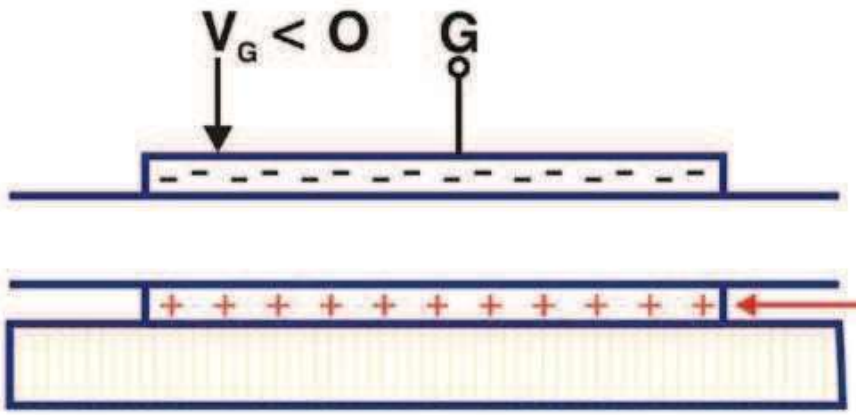
REGION OF OPERATION
Creating a channel

- Apply some positive voltage on the gate terminal.
- This positive voltage pushes the holes downward in the substrate region.
- This causes the electrons to accumulate under the gate terminal.
- At the same time the positive voltage on the gate also attracts the electrons from the n+ region to accumulate under the gate terminal.

The MOS Capacitor

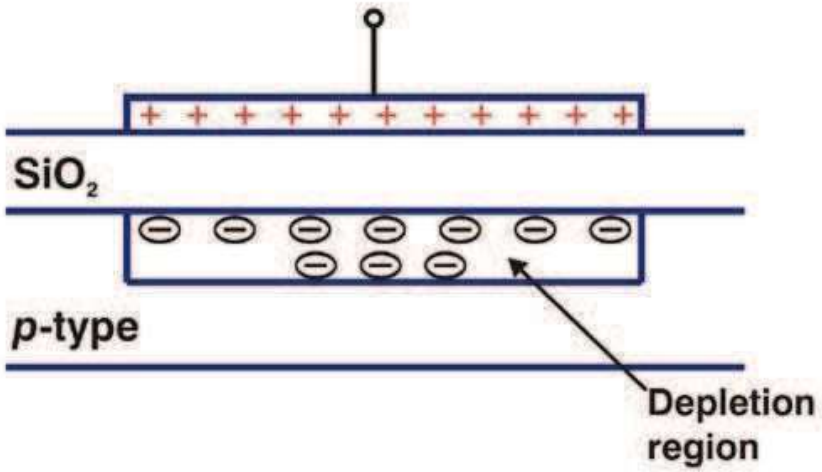


Case 1: Large Negative V_G



Case 2: $V_G < V_t$

(V_G may be small-negative or small-positive)



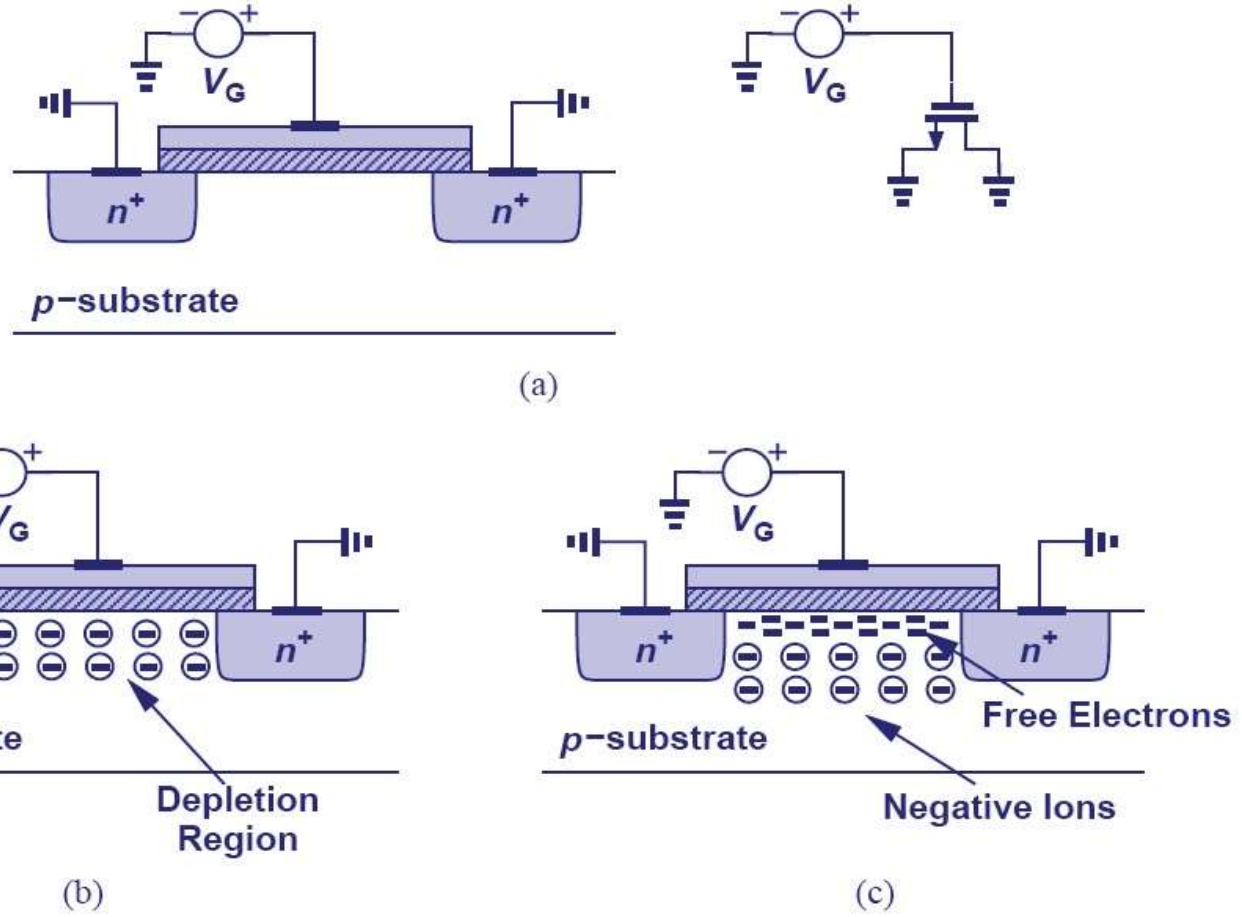
switch from case 1 to case 2 occurs when the hole density near the surface falls below the majority carrier concentration level

REGION OF OPERATION
Creating a channel

- When sufficient electrons are accumulated under the gate an n-region is created, connecting the drain and the source
- This causes the current to flow from the drain to source
- The channel is formed by inverting the substrate surface from p to n, thus induced channel is also called as the inversion layer.
- The voltage between gate and source called vgs at which there are sufficient electron under the gate to form a conducting channel is called threshold voltage Vth.

Formation of Channel

First, the holes arerepelled by the positive gate voltage, leaving behind negative ions and forming a depletion region.Next,elctrons are attracted to the interface, creating a channel("inversion layer").



MOS Transistor Current direction

- The source terminal of an n-channel(p-channel) transistor is defined as whichever of the two terminals has a lower(higher) voltage.
- When a transistor is turned ON, current flows from the drain to source in an n-channel device and from source to drain in a p-channel transistor.
- In both cases, the actual carriers travel from the source to drain.
- The current directions are different because n-channel carriers are negative, whereas p-channel carriers are positive.

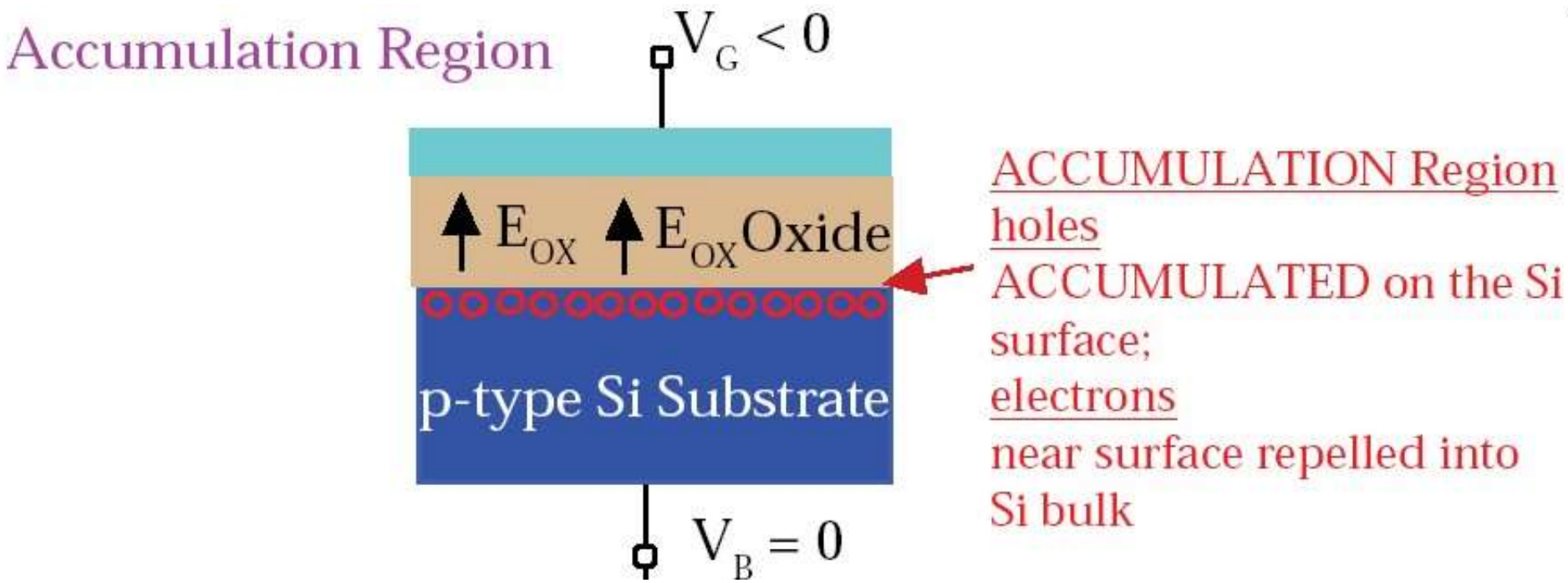
REGION OF OPERATION
Applying small Vds

- Now we applying some small voltage between source and drain
- The voltage Vds causes a current to flow from drain to gate.
- Now as we increase the gate voltage, more current will flow.
- Increasing the gate voltage above the threshold voltage enhances the channel, hence this mode is called as enhancement mode operation.

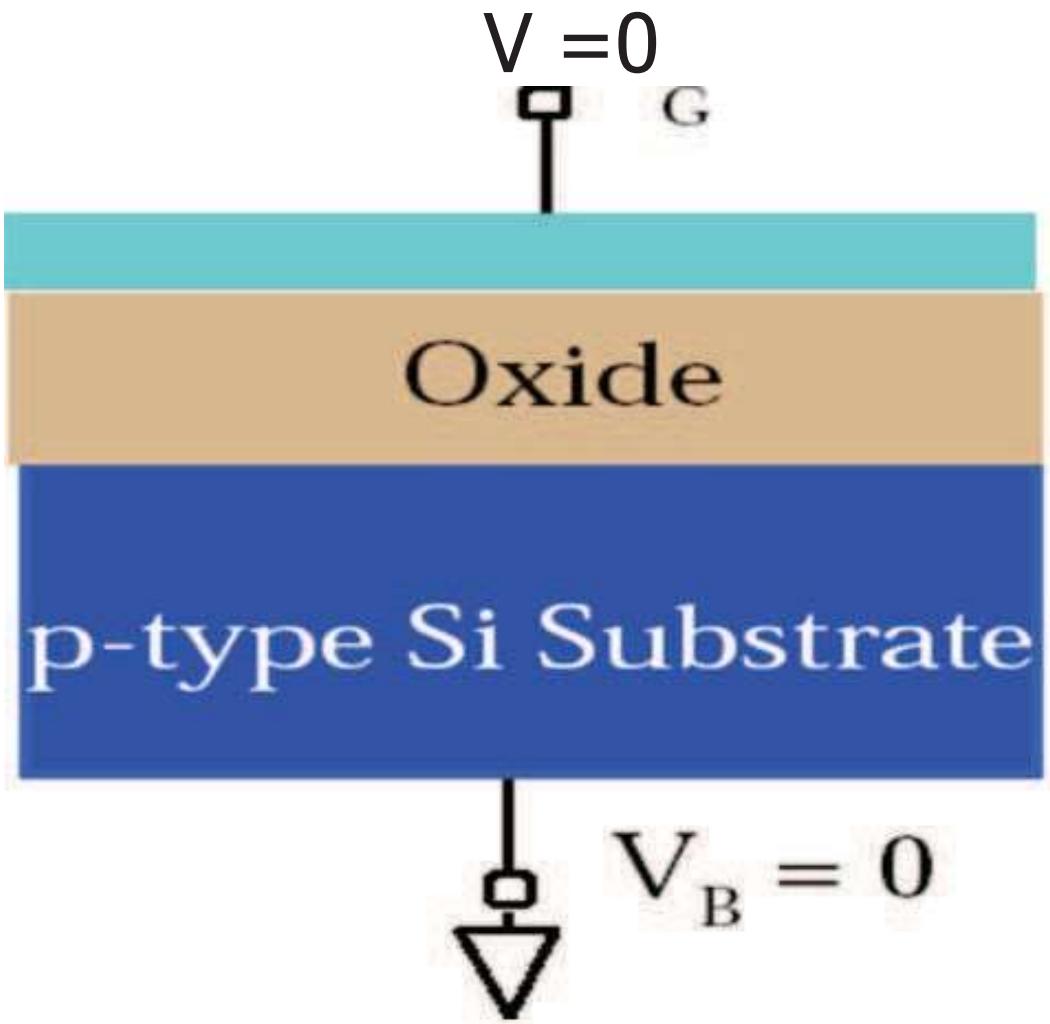
Operation – nMOS Transistor

- Accumulation Mode - If Vgs < 0, then an electric field is established across the substrate.
- Depletion Mode -If 0<Vgs< Vtn, the region under gate will be depleted of charges.
- Inversion Mode – If Vgs > Vtn, the region below the gate will be inverted.

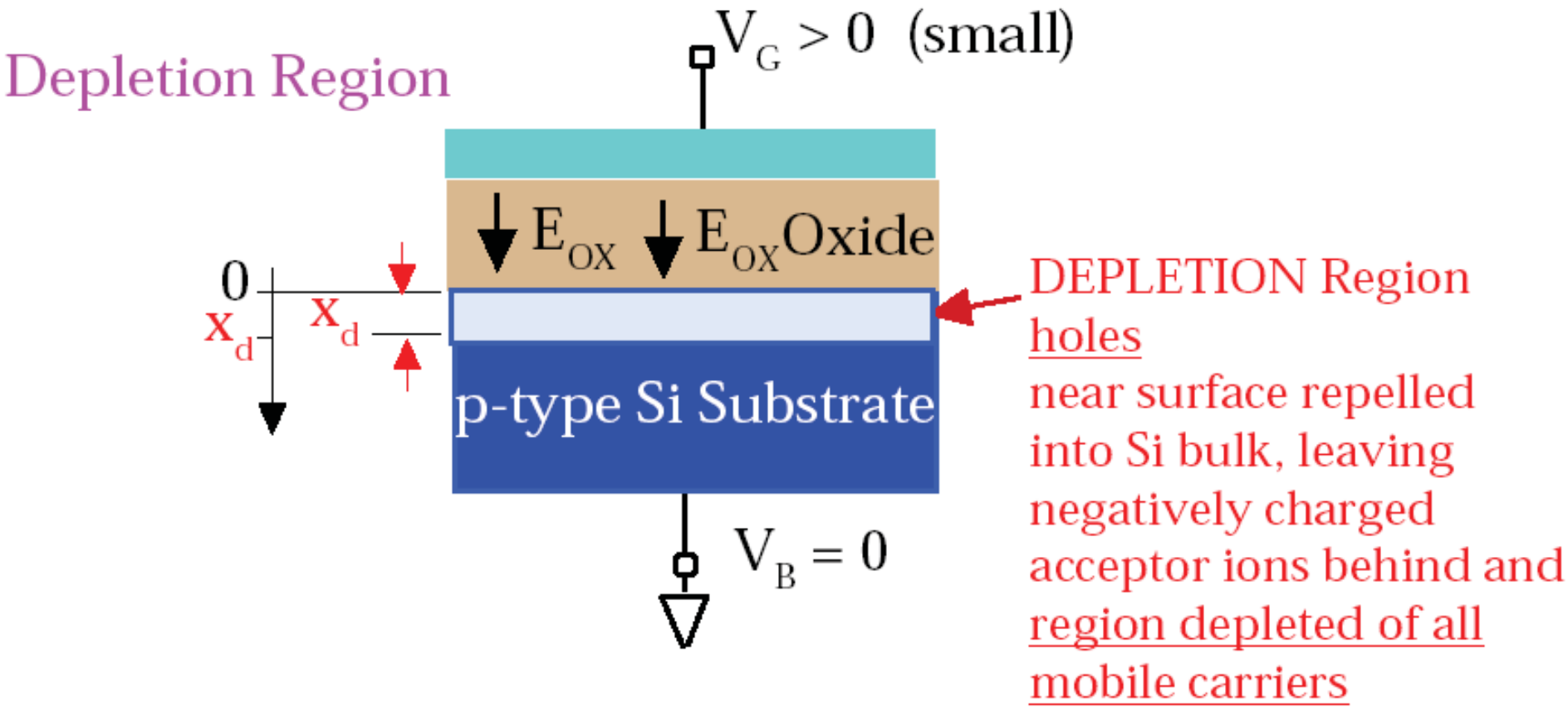
Operation – nMOS Transistor



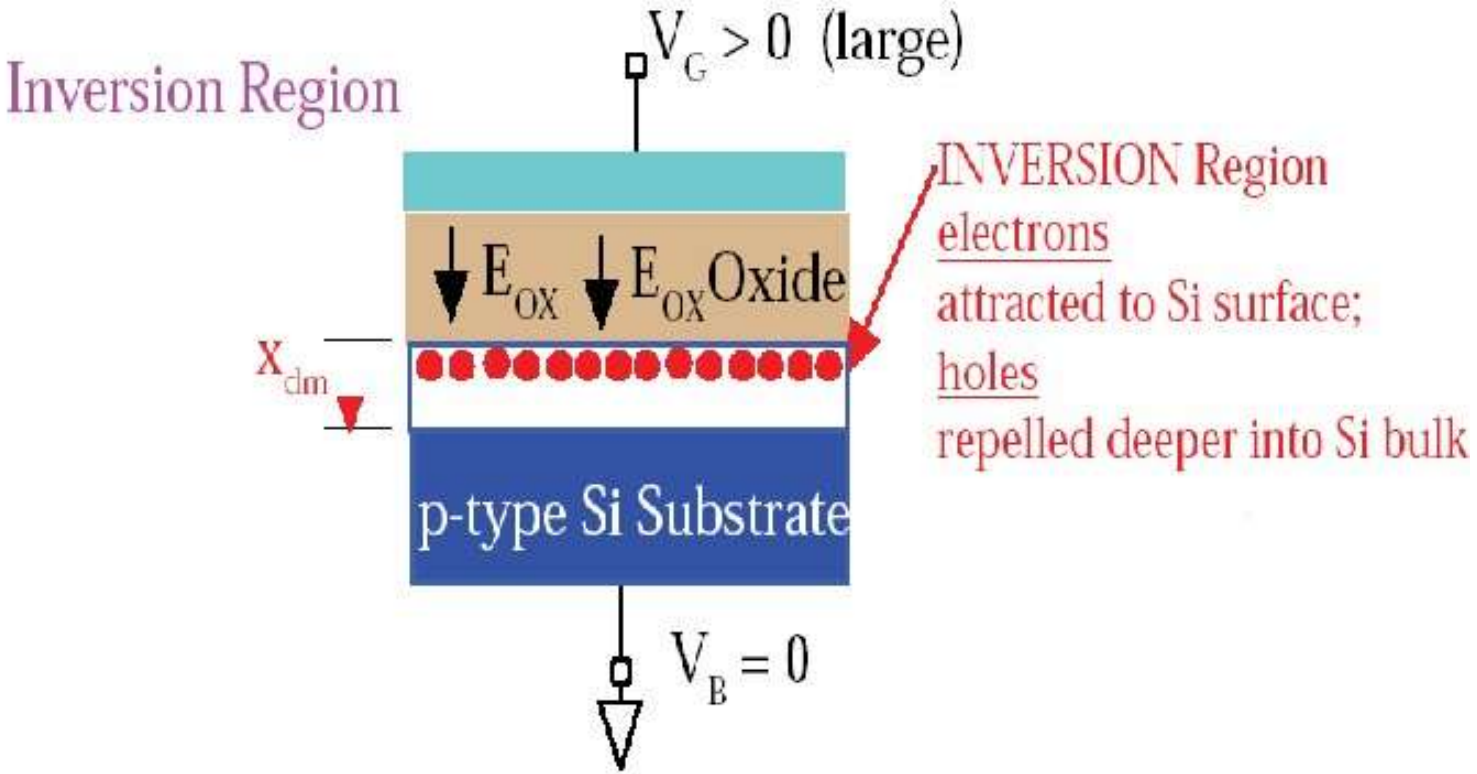
Operation – nMOS Transistor



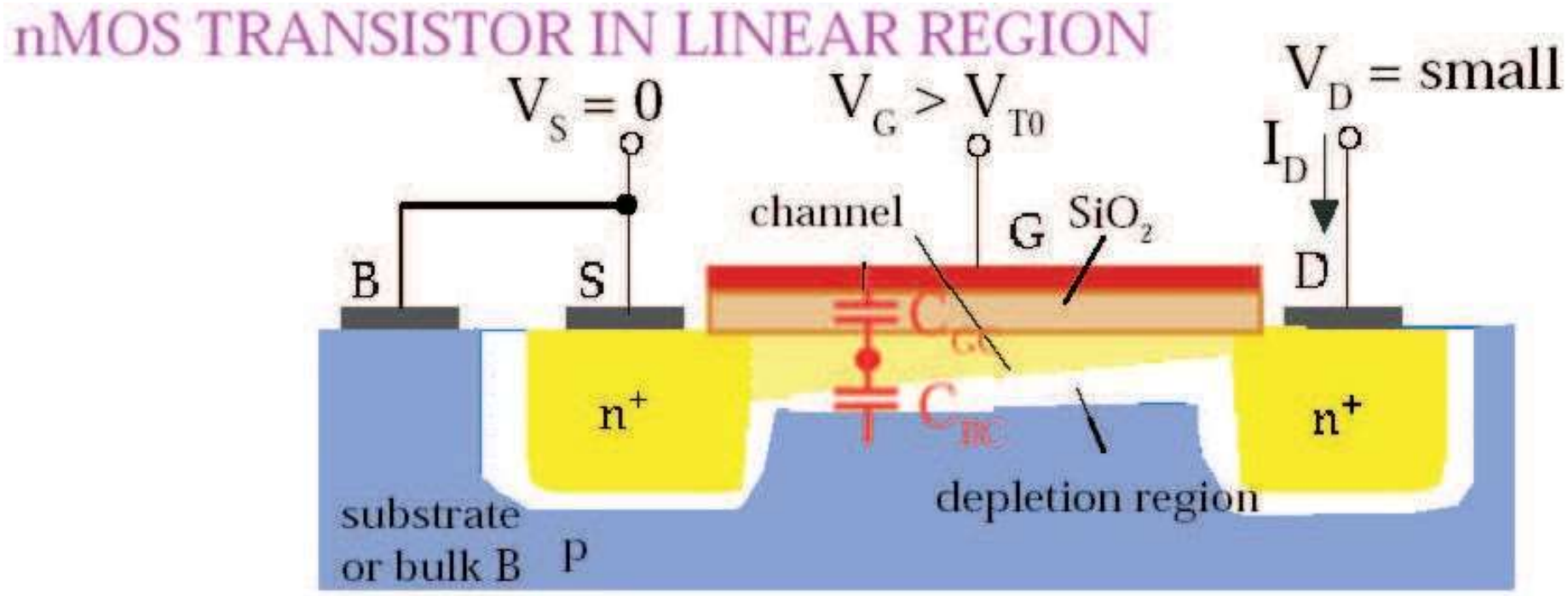
Operation – nMOS Transistor



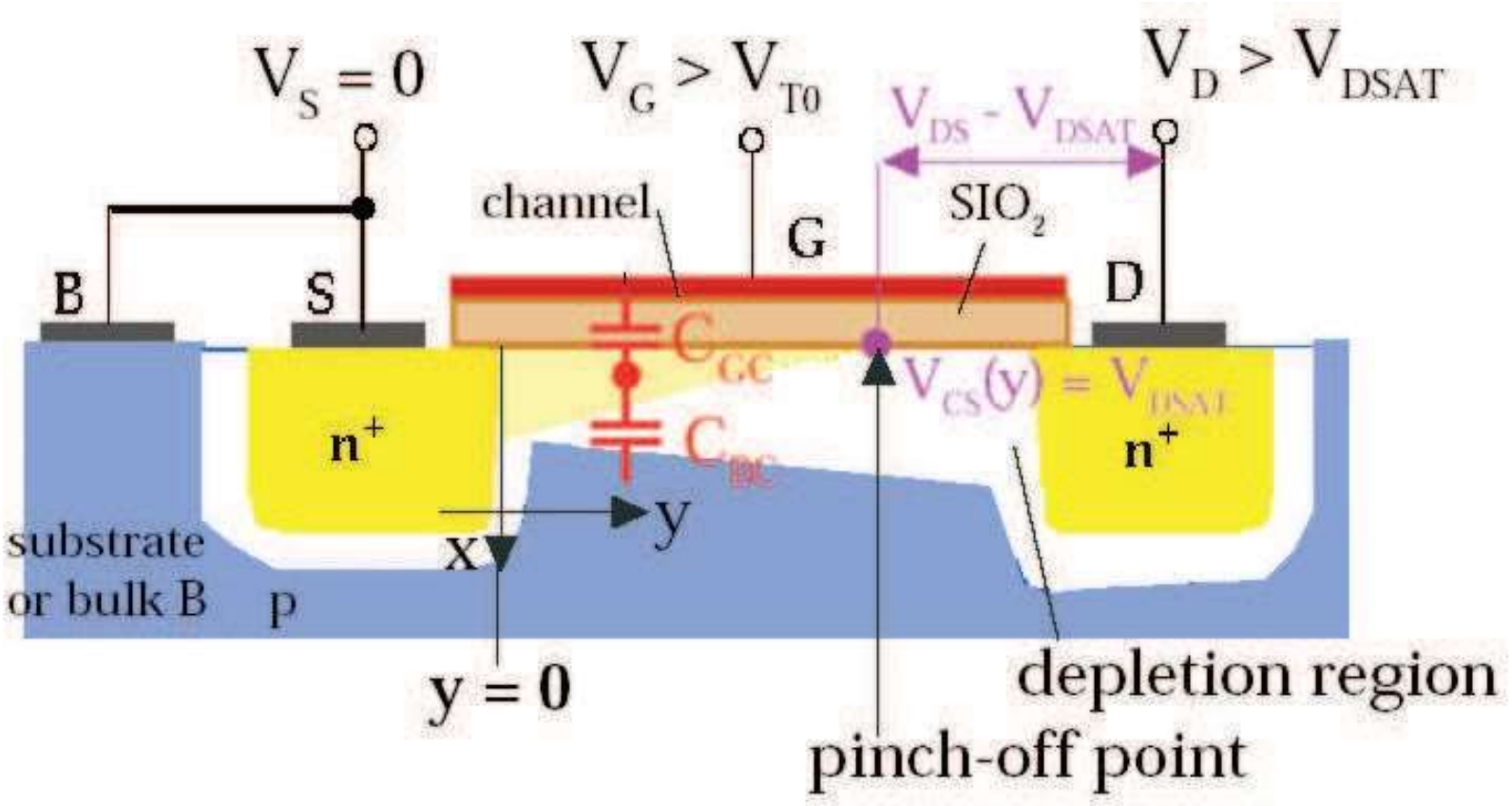
Operation – nMOS Transistor



Operation – nMOS Transistor

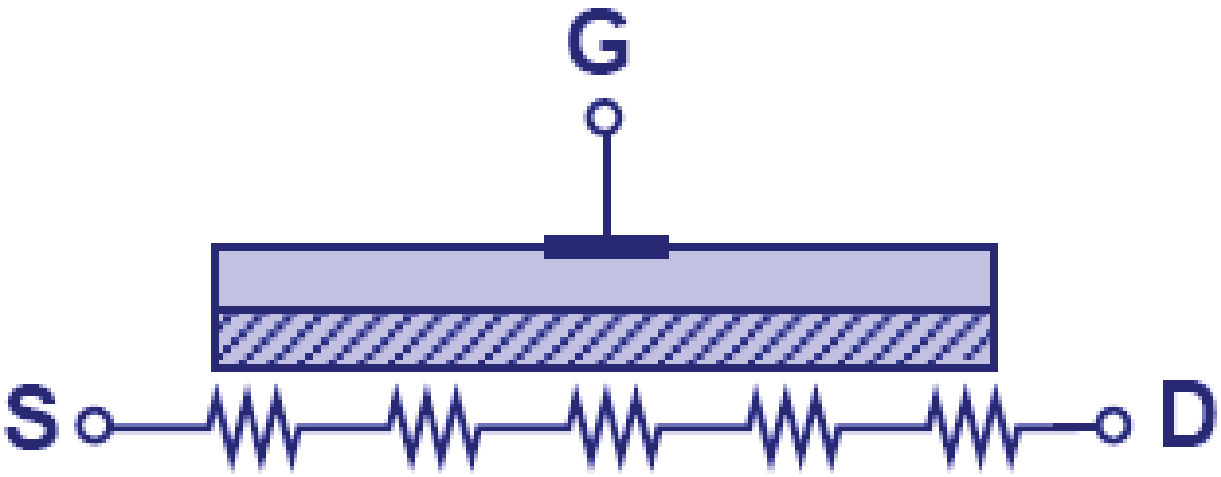


Operation – nMOS Transistor



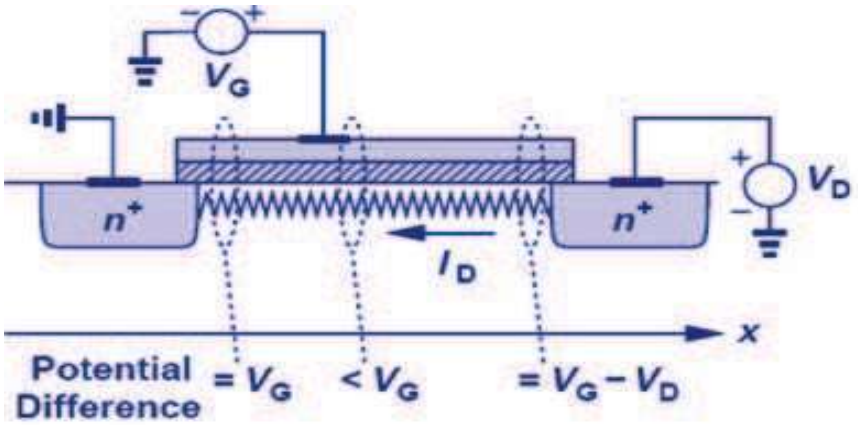
Voltage-Dependent Resistor

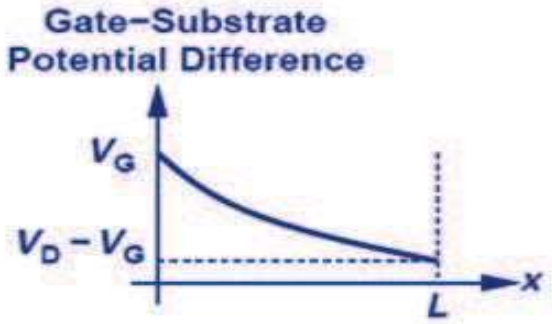
- The inver
of a MOS
seen as a
- Since the
density in
channel depends on
the gate voltage, this
resistance is also
voltage-dependent.



Channel Potential Variation

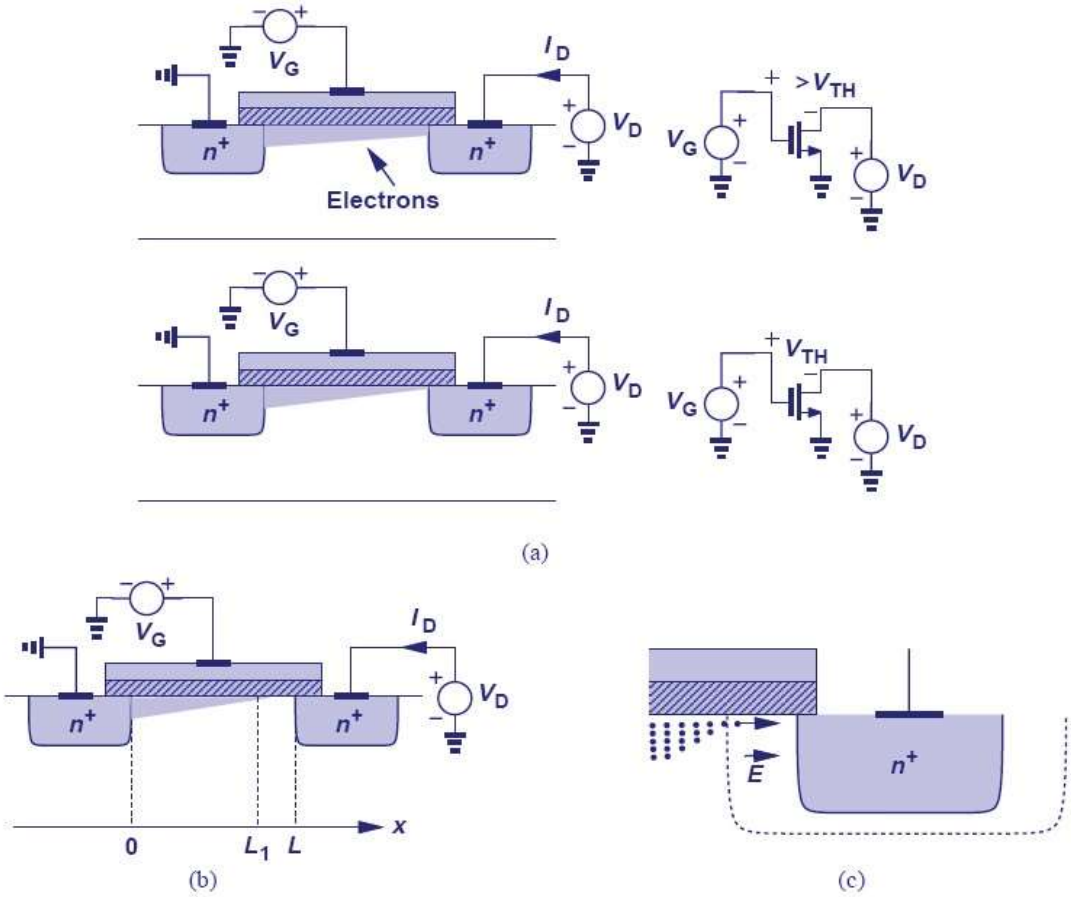
- Since there's a channel resistance between drain and source biase and if drain is higher than the source than the potential between gate and channel will decrease from source to drain.





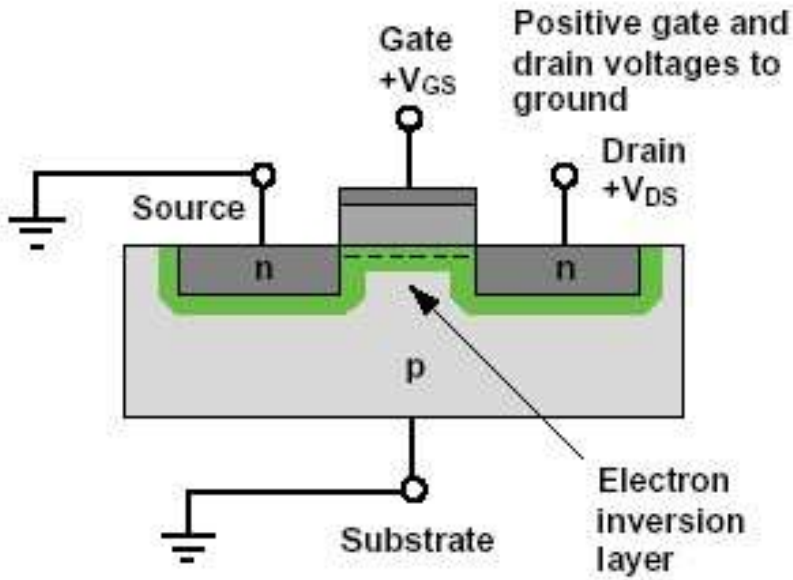
Channel Pinch-Off

- As the poten
difference be
and gate bec
positive, the
layer beneat
interface sta
off around d
- When VD s>
the channel
totally pinch
when VD s< VGs - Vth,
the channel length starts
to decrease.



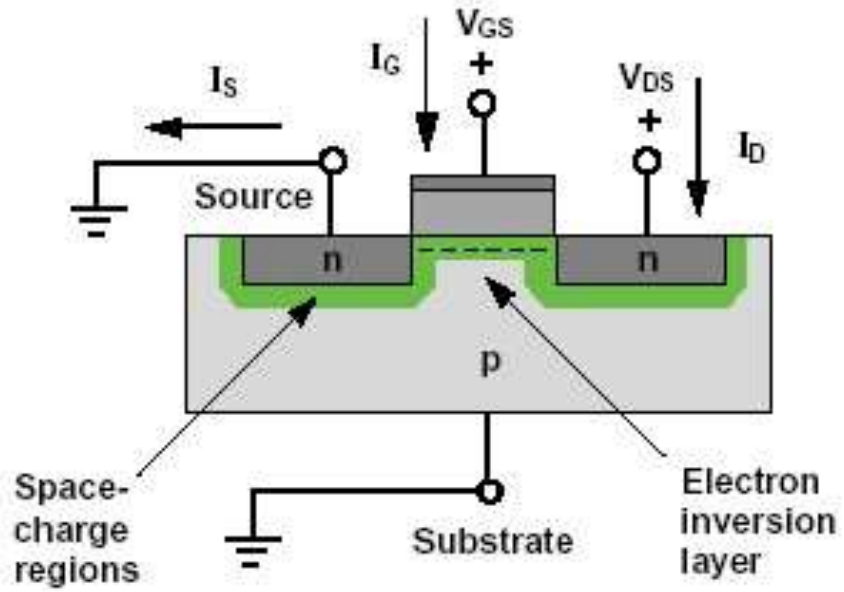
N-Channel MOSFET

- **Enhancement mode** — voltage must be applied to gate to "turn-on" the transistor



N-Channel MOSFET

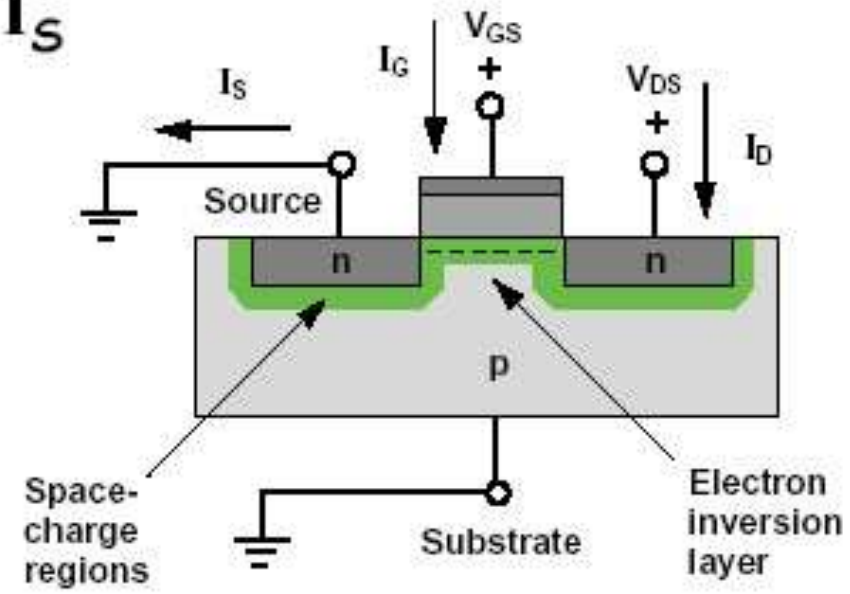
- Threshold (V_{TN}) — The gate voltage required to "turn-on" the transistor



- Minority carriers (electrons) concentrated — connect n drain to n source

N-Channel MOSFET Currents

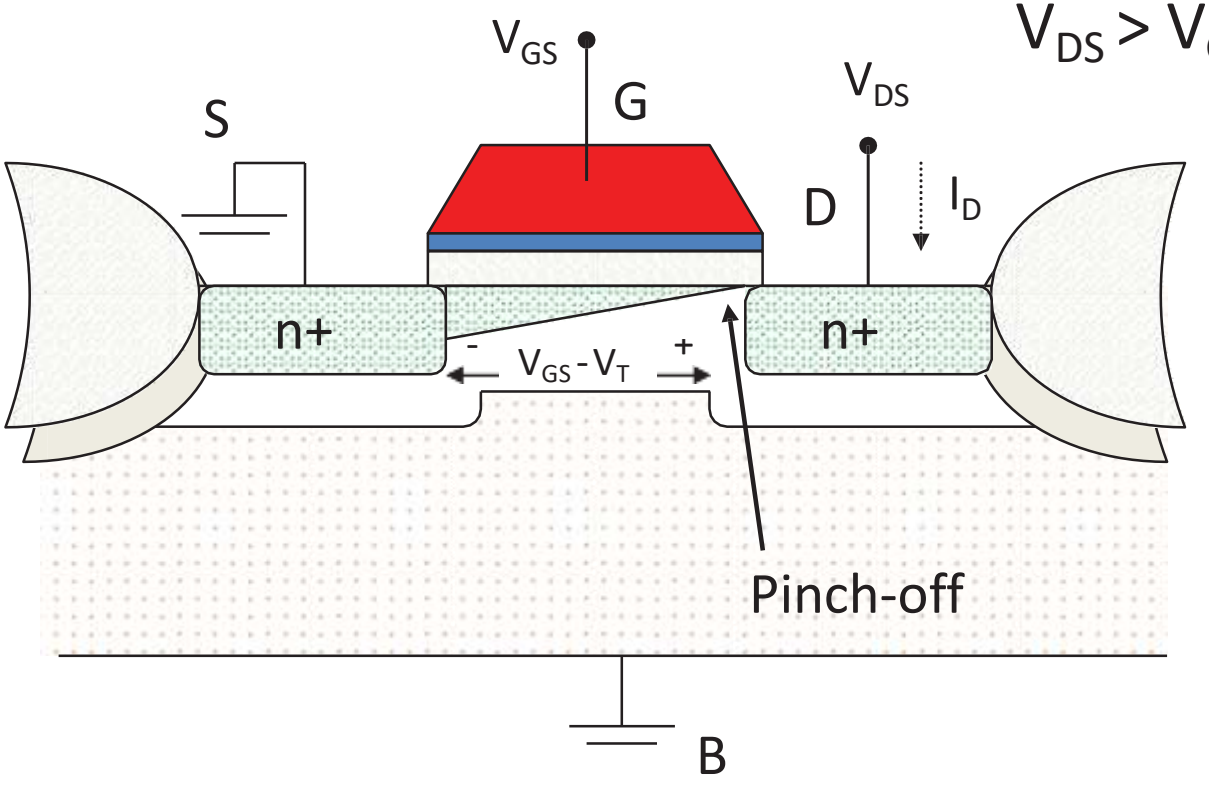
- Gate-to-substrate resistance approaches infinite
- Gate current: $I_G = 0$
- $I_D = I_S$



Transistor in Saturation Mode

Assuming $V_{GS} > V_T$

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



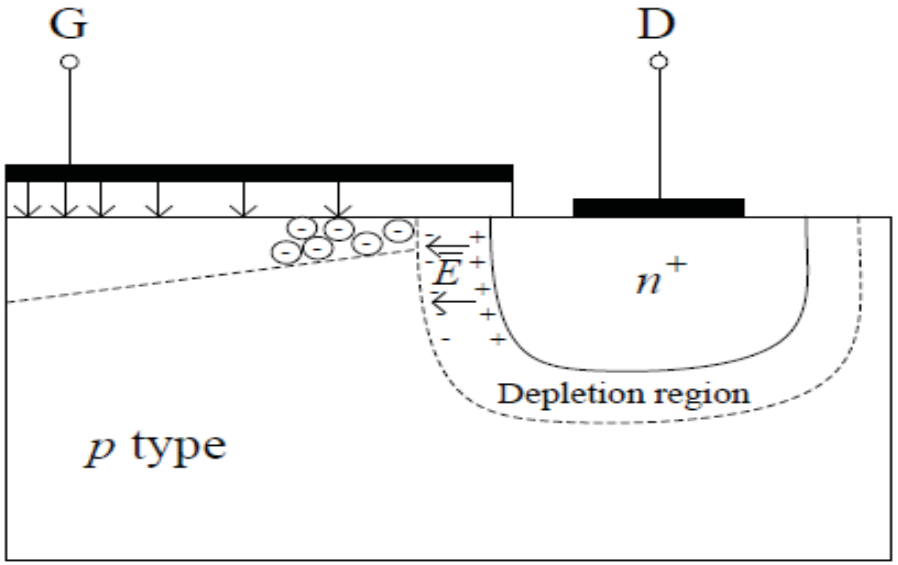
The current remains constant (saturates).

During “pinchoff”

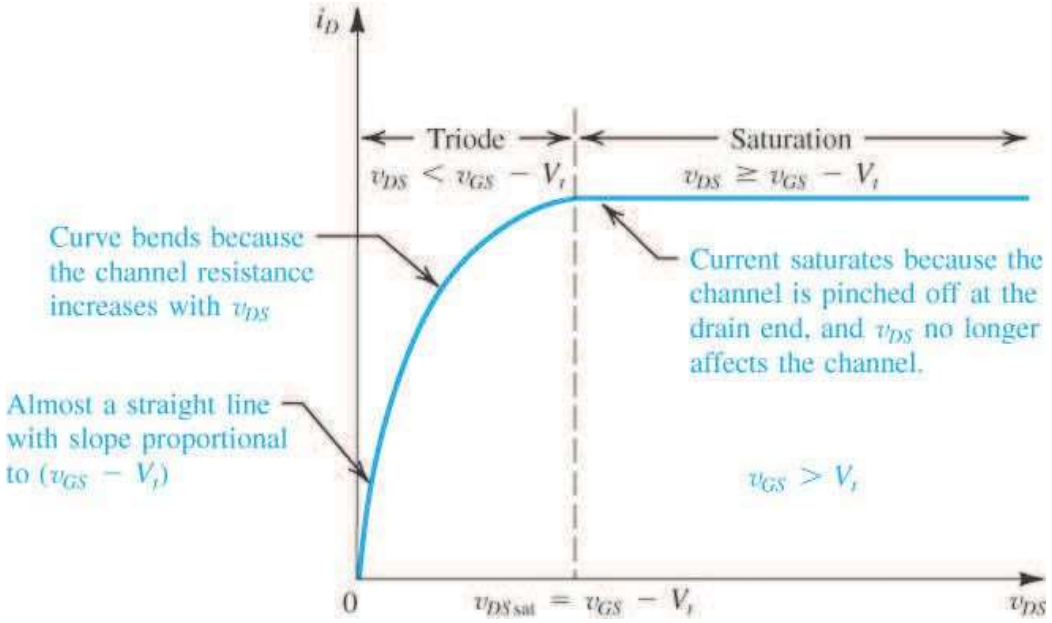
- **Does this mean that the current *i =0 ? Actually, it does not. A* MOSFET that is “pinched off” at the drain end of the channel still conducts current:**

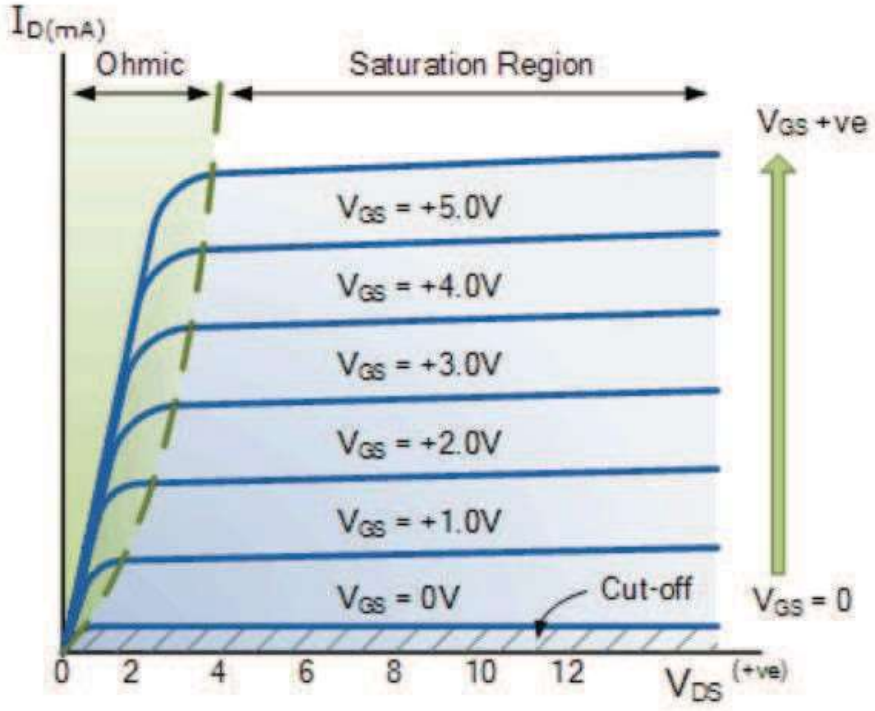
- **The large *E in the depletion region surrounding the drain will* sweep electrons across the end of the pinched off channel to the drain.**

- **This is very similar to the operation of the BJT. For an *npn BJT,* the electric field of the reversed biased CBJ swept electrons from the base to the collector regions.**

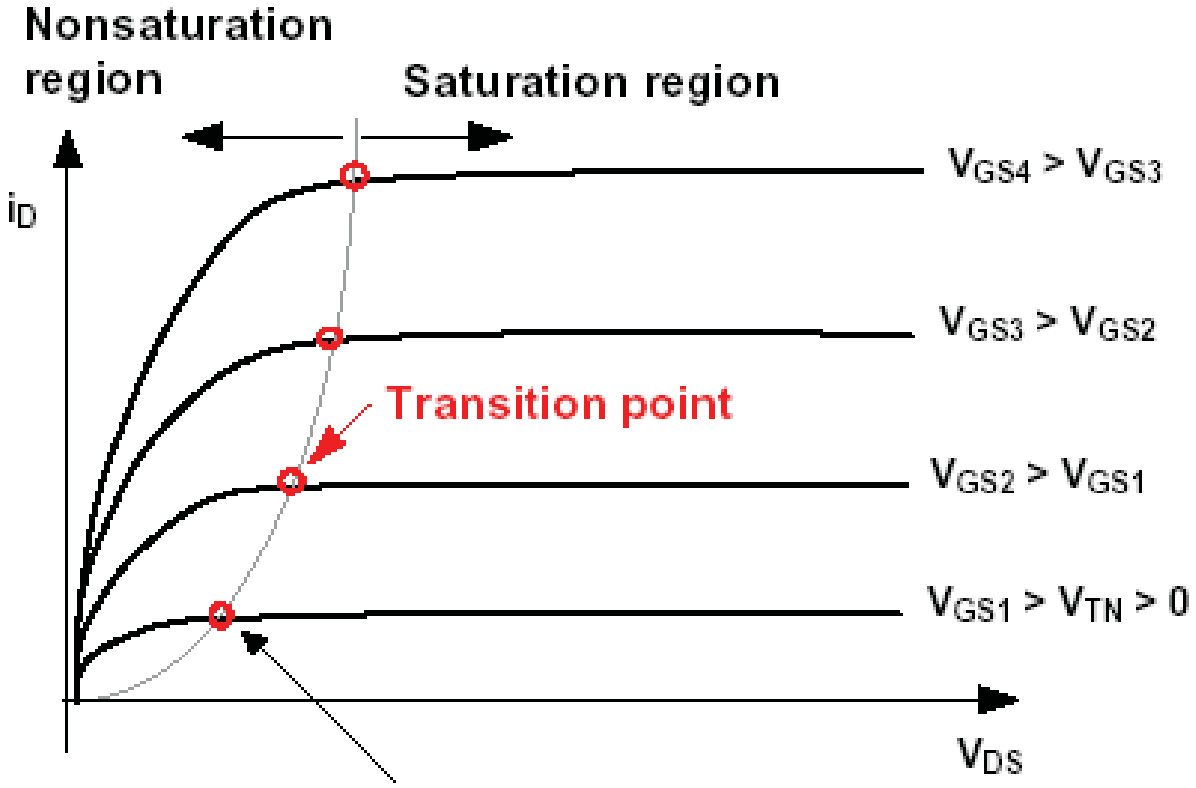


N-Channel MOSFET characteristics





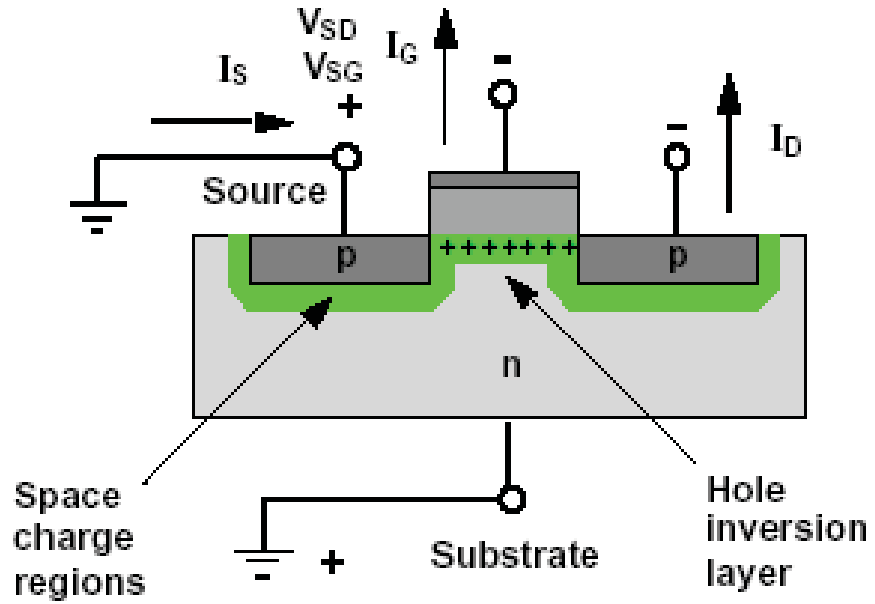
N-Channel MOSFET Characteristics



$$V_{DS(sat)} = V_{GS} - V_{TN}$$

P-Channel MOSFET — PMOS

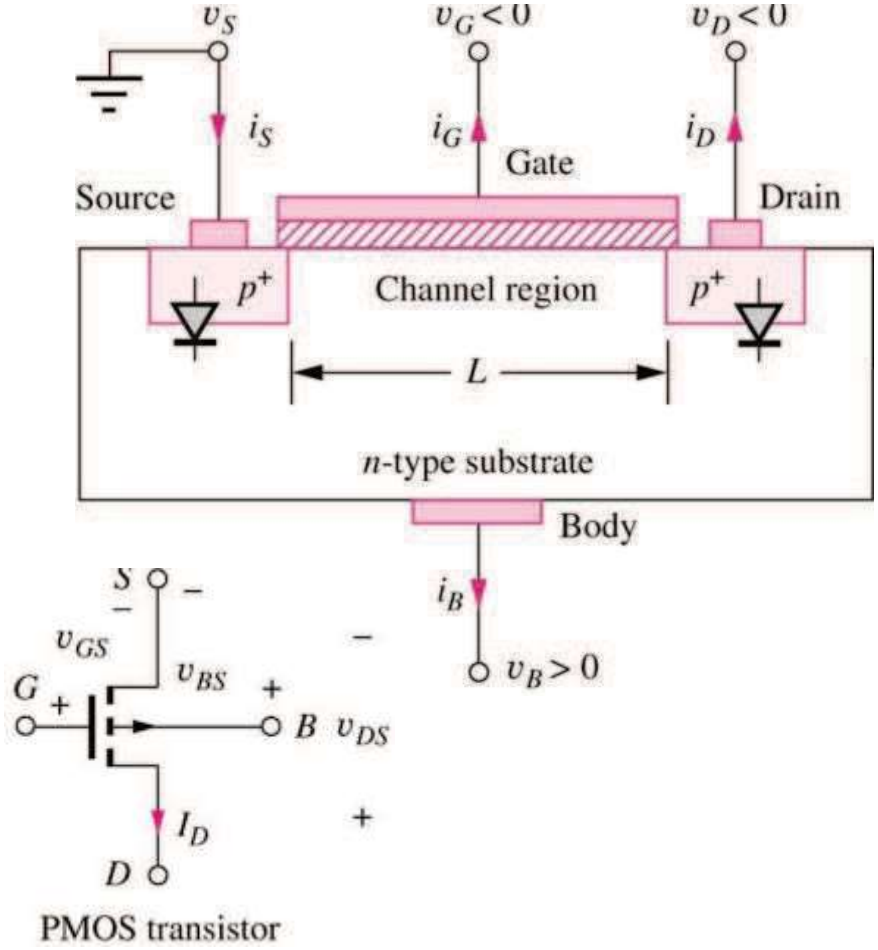
- Threshold (V_{TP}) — The gate voltage required to "turn-on" the transistor



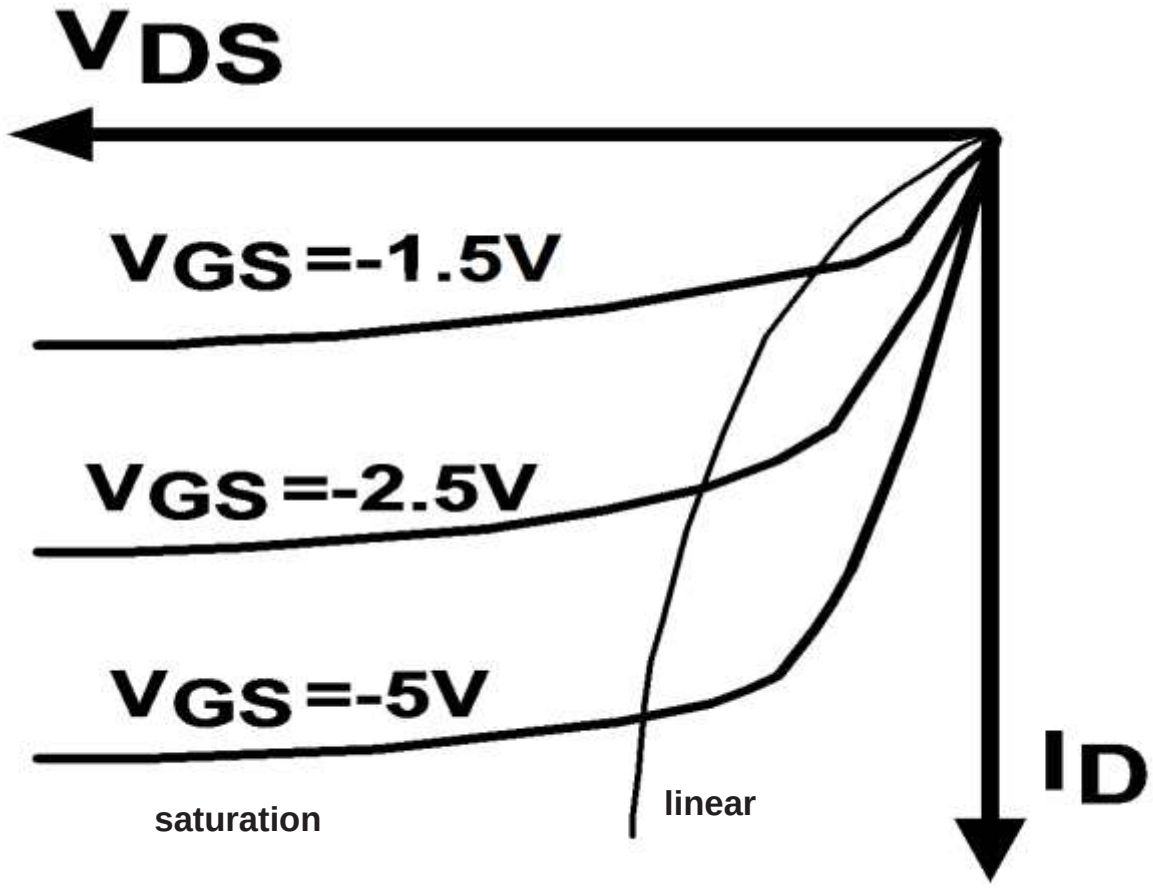
- Minority carriers (holes) created — connect p drain to p source

Enhancement-Mode PMOS Transistors: Structure

- p-type source and drain regions in n-type substrate.
- vGS < 0 required to create p-type inversion layer in channel region
- For current flow, vGS < vTP
- To maintain reverse bias on source-substrate and drain-substrate junctions, vSB < 0 and vDB < 0
- Positive bulk-source potential causes VTP to become more negative



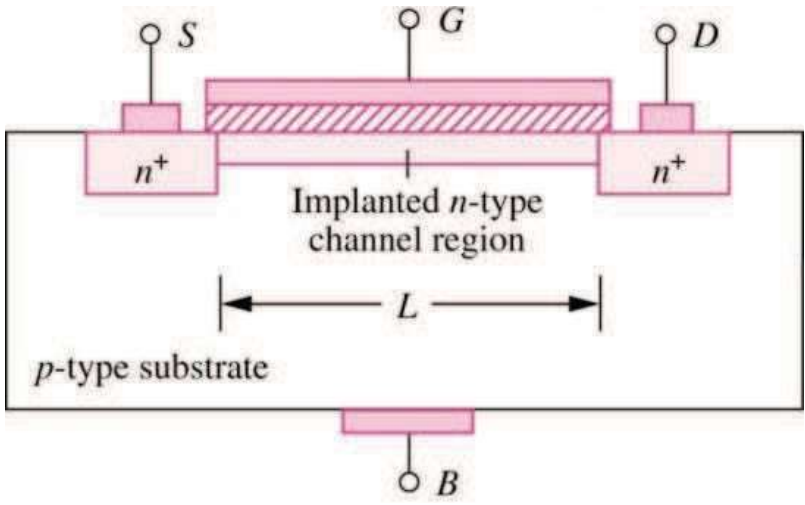
P-channel MOSFET characteristics



p transistor

Depletion-Mode MOSFETS

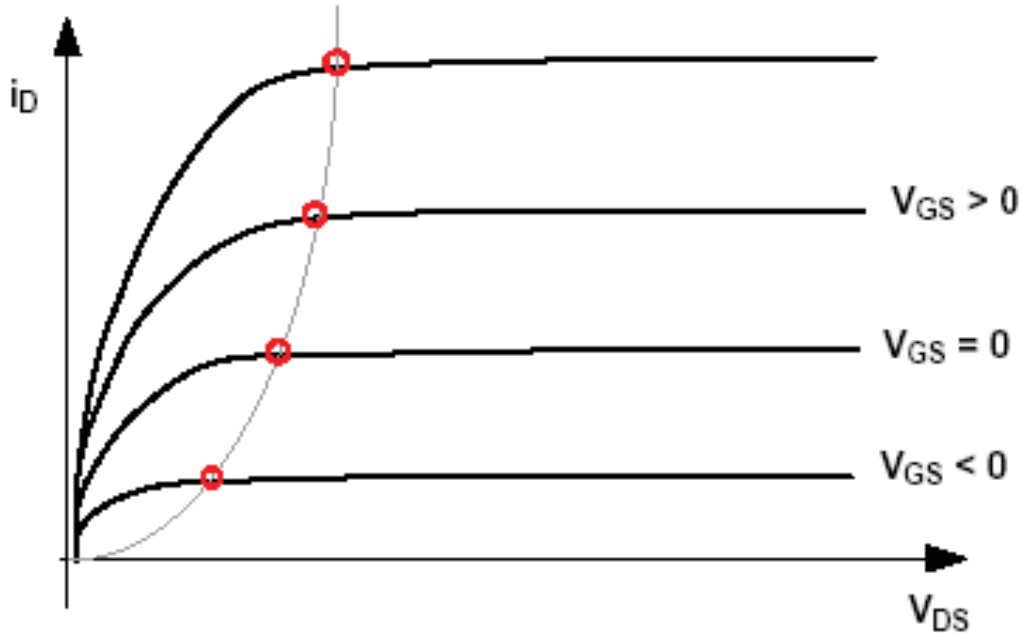
- NMOS transistors with
- Ion implantation process is used to form a built-in n-type channel in the device to connect source and drain by a resistive channel
- Non-zero drain current for vGS = 0; negative vGS required to turn device off. $V_{TN} \square 0$

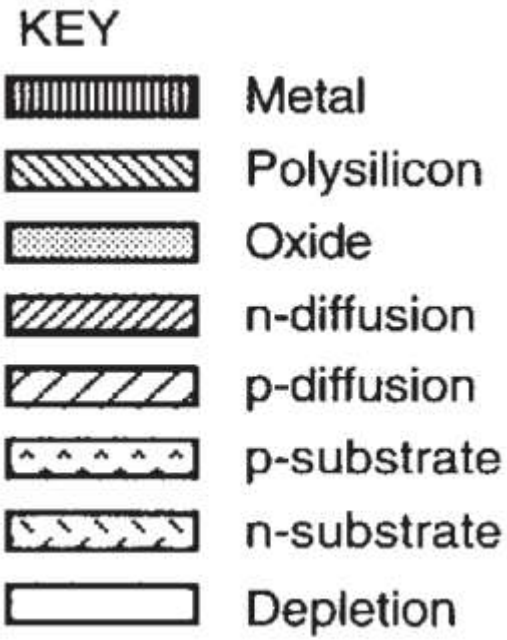


Depletion-Mode MOSFETS

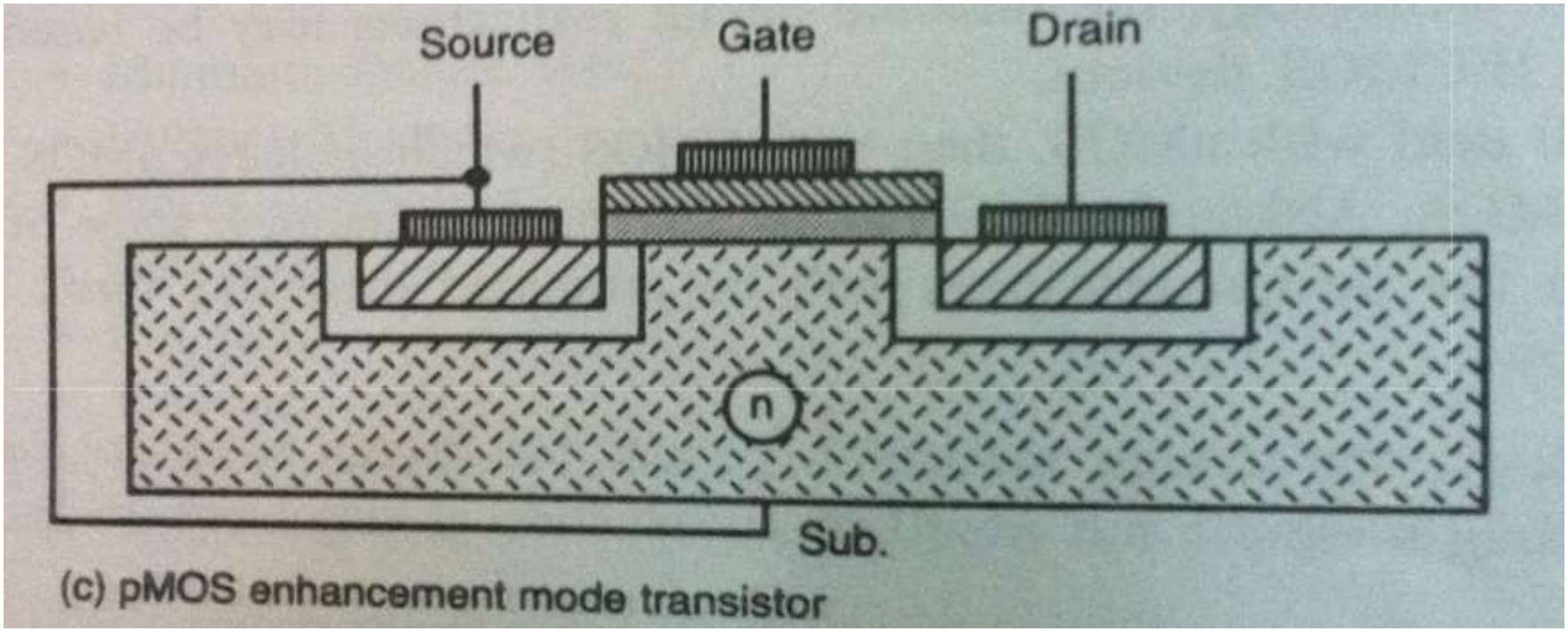
- Channel (inversion layer) exists even at zero gate voltage
- Thus, gate voltage must be applied to "turn-off" the device

N-Channel Depletion-Mode MOSFET





pMOS are 2.5 time slower than nMOS due to electron and hole mobilities



Basic processes involved in fabricating Monolithic ICs

1. Silicon wafer (substrate) preparation
2. Epitaxial growth
3. Oxidation
4. Photolithography
5. Diffusion
6. Ion implantation
7. Metallization
8. Testing
9. Assembly processing & packaging

Oxidation

- Formation of silicon dioxide layer on the surface of Si wafer
  1. protects surface from contaminants
  2. forms insulating layer between conductors
  3. form barrier to dopants during diffusion or ion implantation
  4. grows above and into silicon surface

- Dry oxidation: lower rate and higher quality
- Wet oxidation: higher rate and lower quality

1. SiO_2 is an extremely hard protective coating & is unaffected by almost all reagents except by hydrochloric acid. Thus it stands against any contamination.
2. By selective etching of SiO_2, diffusion of impurities through carefully defined through windows in the SiO_2 can be accomplished to fabricate various components.

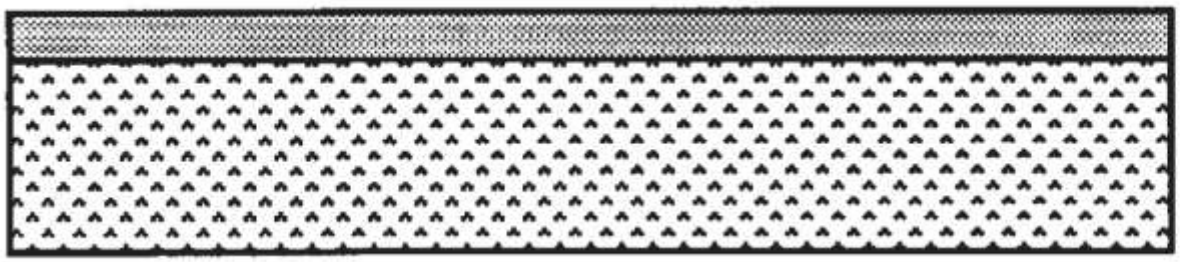
Oxidation

The silicon wafers are stacked up in a quartz boat & then inserted into quartz furnace tube. The Si wafers are raised to a high temperature in the range of 950 to 1150^{o}C & at the same time, exposed to a gas containing O_2 or H_2O or both. The chemical action is

$Si + 2H_2O \text{----------} > Si\ O_2 + 2H_2$ (Wet)

$Si + O_2 \text{----------} > SiO_2$ (Dry)

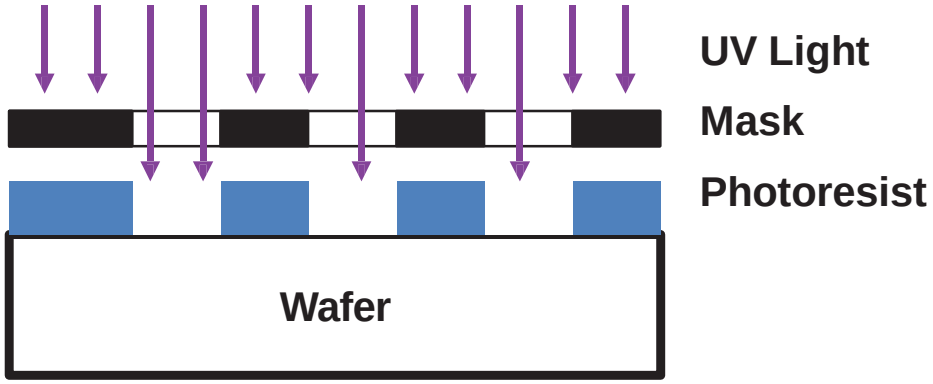
2.



Thick oxide (1 µm)

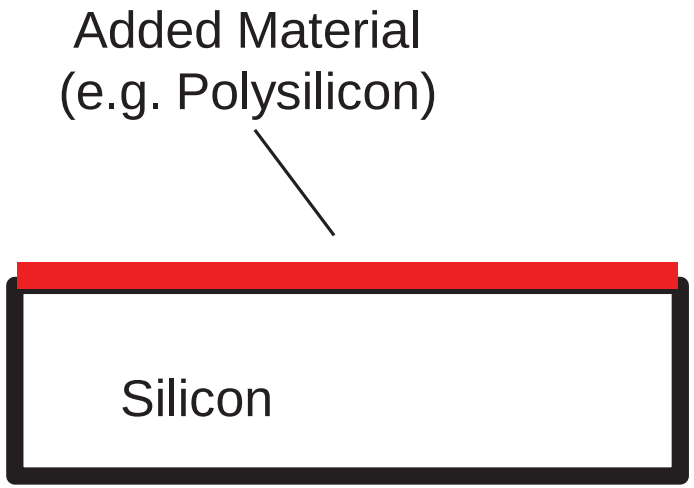
Photolithography

- Coat wafer with photoresist (PR)
- Shine UV light through mask to selectively expose PR
- Use acid to dissolve exposed PR
- Now use exposed areas for
  - Selective doping
  - Selective removal of material under exposed PR



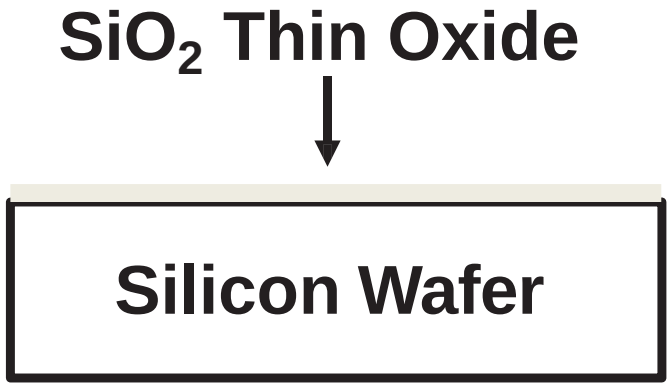
Adding Materials

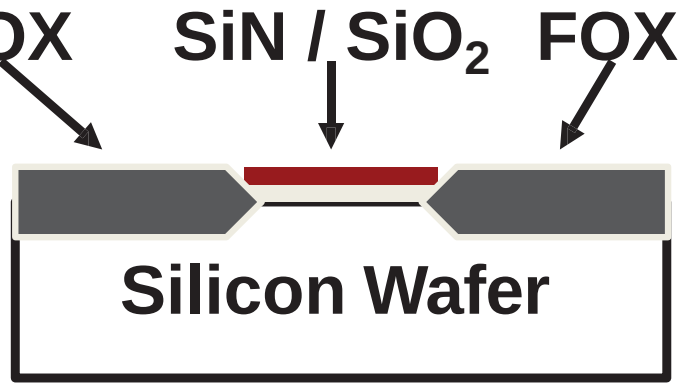
- Add materials on top of silicon
  - Polysilicon
  - Metal
  - Oxide (SiO_2) - Insulator
- Methods
  - Chemical deposition
  - Sputtering (Metal ions)
  - Oxidation



Oxide (SiO_2) - The Key Insulator

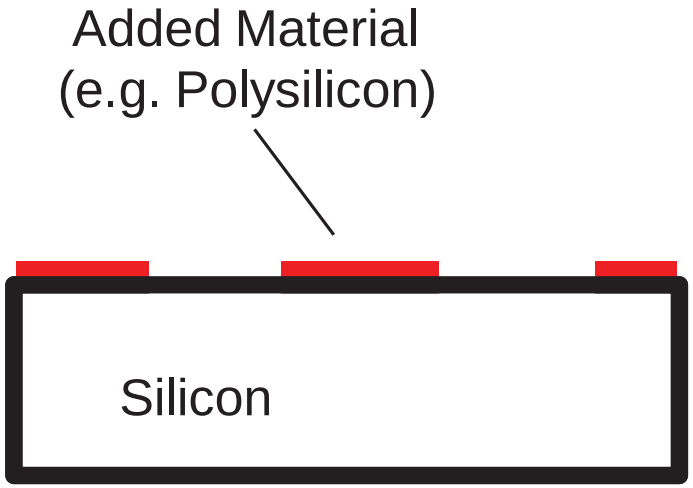
- Thin Oxide
  - Add using chemical deposition
  - Used to form gate insulator & block active areas
- Field Oxide (FOX) - formed by oxidation
  - Wet (H_20 at 900°C - 1000°C) or Dry (O_2 at 1200°C)
  - Used to insulate non-active areas





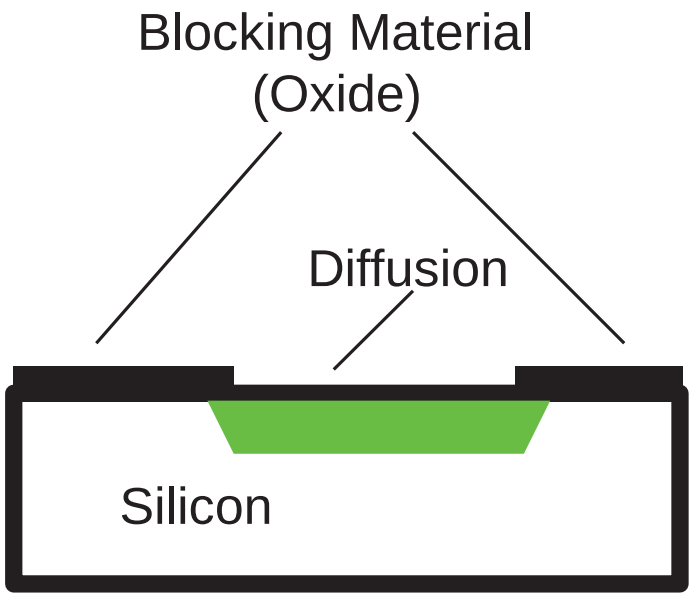
Patterning Materials using Photolithography

- Add material to wafer
- Coat with photoresist
- Selectively remove photoresist
- Remove exposed material
- Remove remaining PR



Diffusion

- Introduce dopant via epitaxy or ion implant e.g. Arsenic (N), Boron (P)
- Allow dopants to diffuse at high temperature
- Block diffusion in selective areas using oxide or PR
- Diffusion spreads both vertically, horizontally



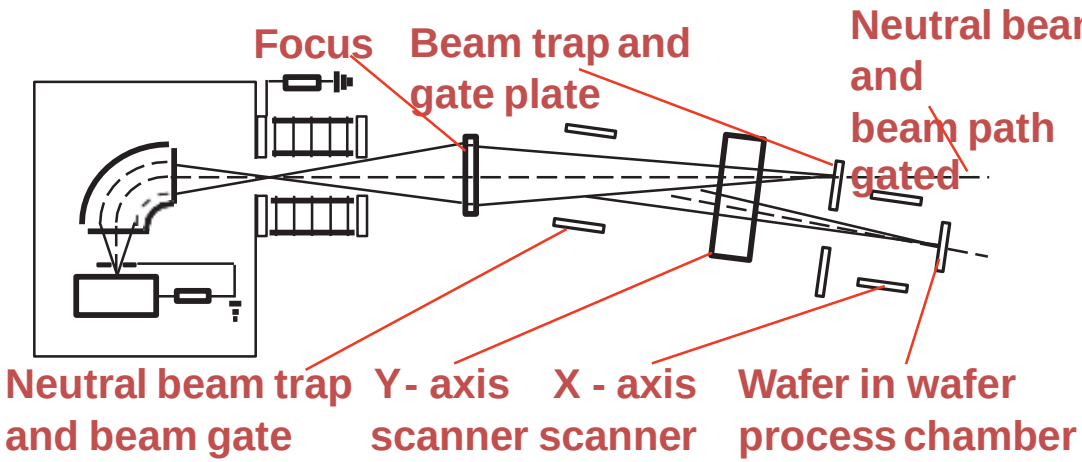
Ion Implantation

Process Conditions

Flow Rate: 5 sccm

Pressure: 10^{-5} Torr

Accelerating Voltage: 5 to 200 keV



Gases

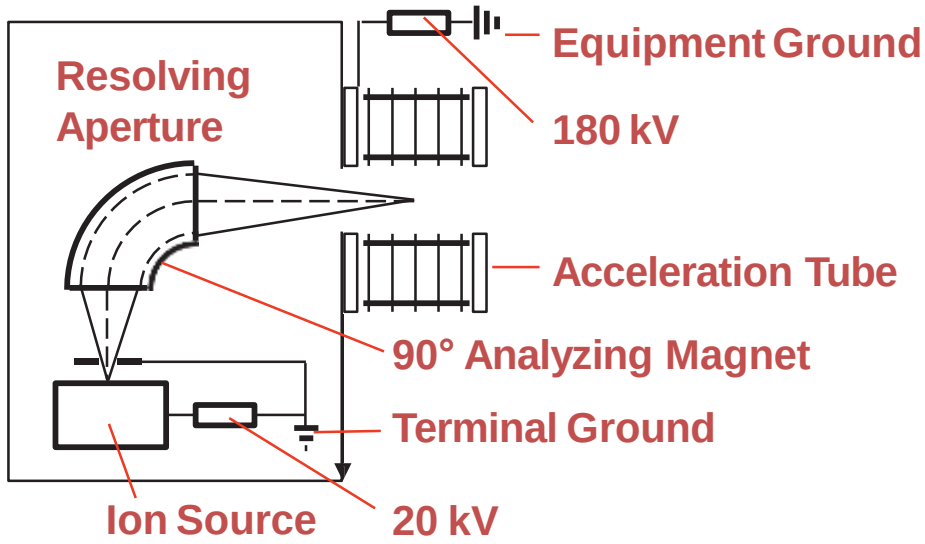
- Ar
- AsH_3
- $B^{11}F_3$ *
- He
- N_2
- PH_3
- SiH_4
- SiF_4
- GeH_4

Solids

- Ga
- In
- Sb

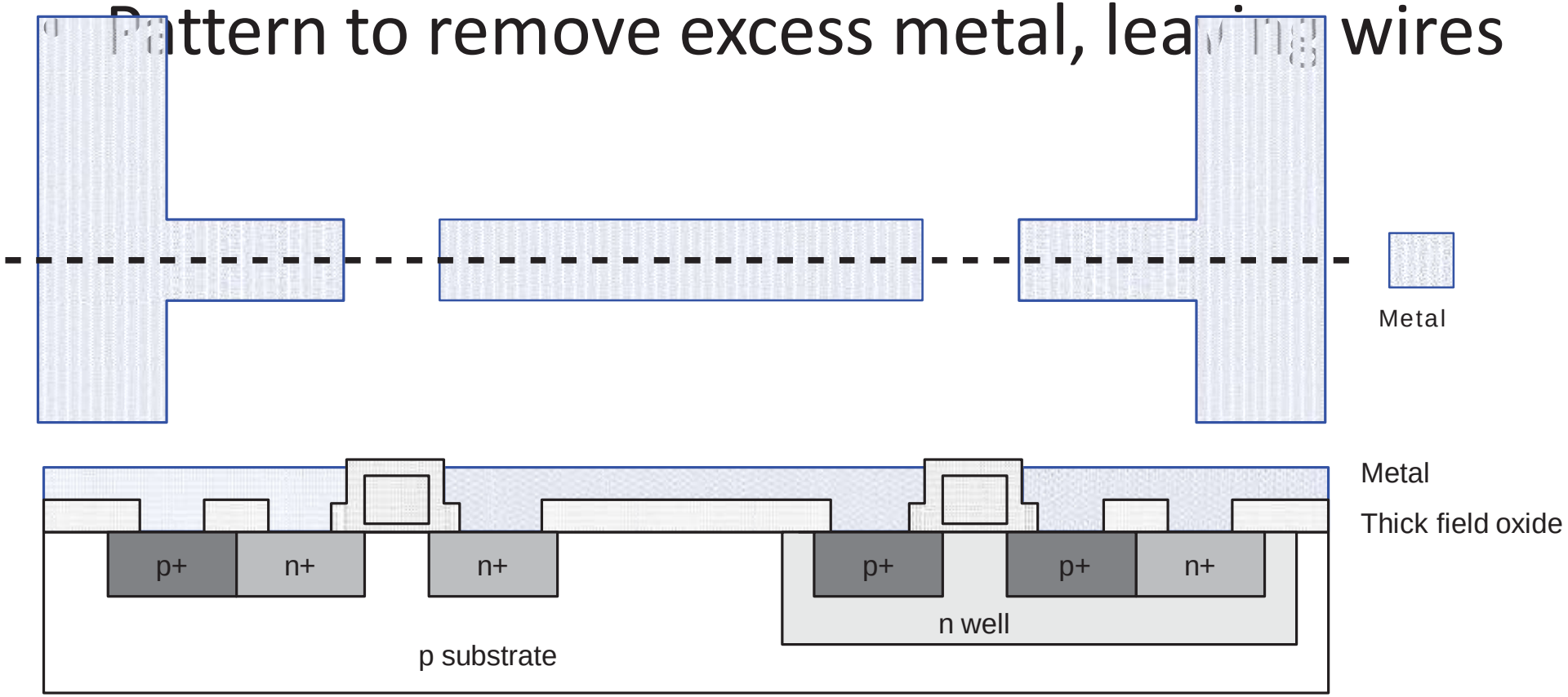
Liquids

- $Al(CH_3)_3$



Metallization

- Sputter on aluminum over whole wafer
- Pattern to remove excess metal, leaving wires



nMOS fabrication steps

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

2.A layer of silicon dioxide (Si02) is grown all over the surface of the wafer to protect the surface, act as a barrier to dopants during processing, and provide a generally insulating substrate on to which other layers may be deposited and patterned.

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

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

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

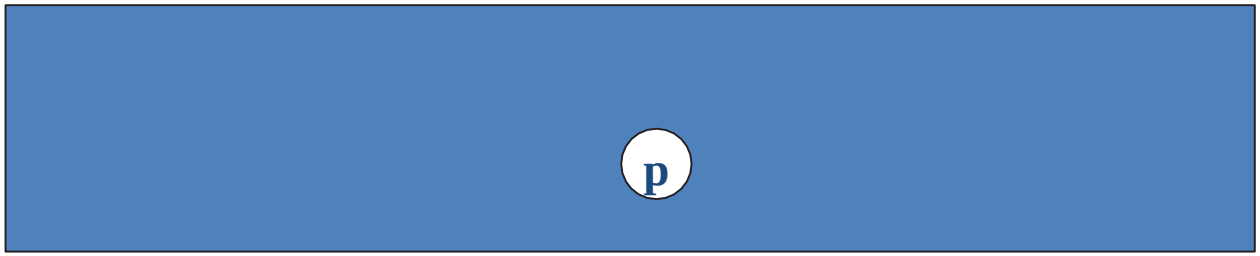
6.The remaining photoresist is removed and a thin layer of Si02 is grown over the entire chip surface and then polysilicon is deposited on top of this to form the gate structure.

7.Further photoresist coating and masking allows the polysilicon to be patterned (as shown in Step 6) and then the thin oxide is removed to expose areas into which

8.Thick oxide (Si02) is grown over all again and is then masked with photoresist and etched to expose selected areas of the polysilicon gate and the drain and source areas where connections (i.e. contact cuts) are to be made.

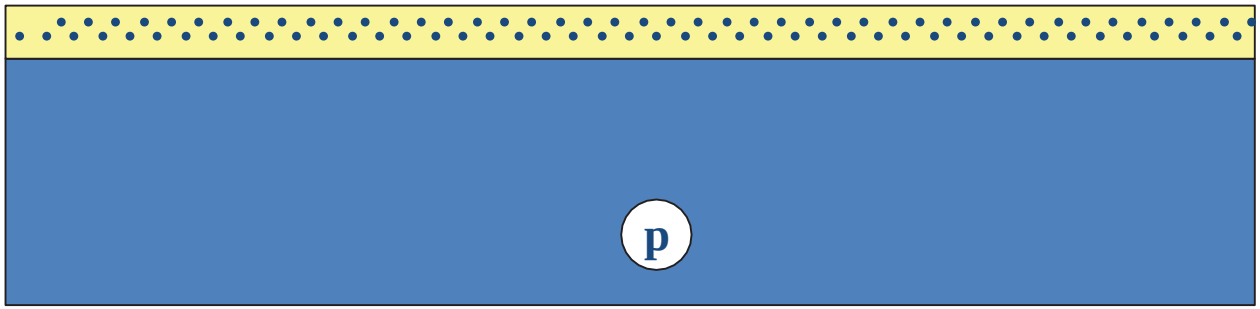
9.The whole chip then has metal (aluminium) deposited over its surface. This metal layer is then masked and etched to form the required interconnection pattern.

1.



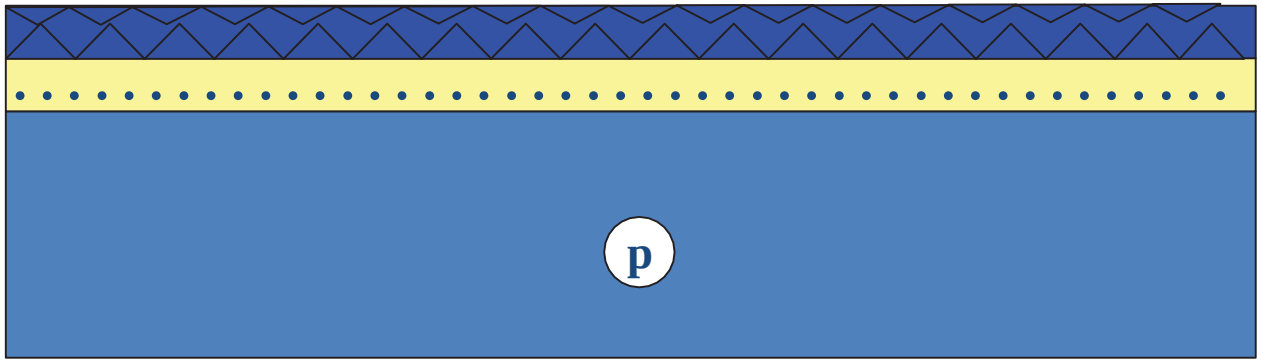
Substrate

2.

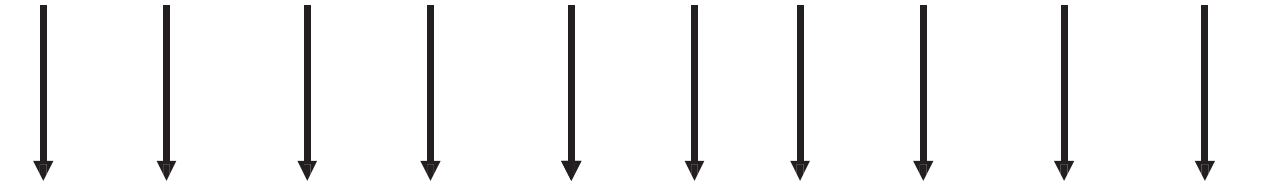


Thick oxide (1μm)

3.



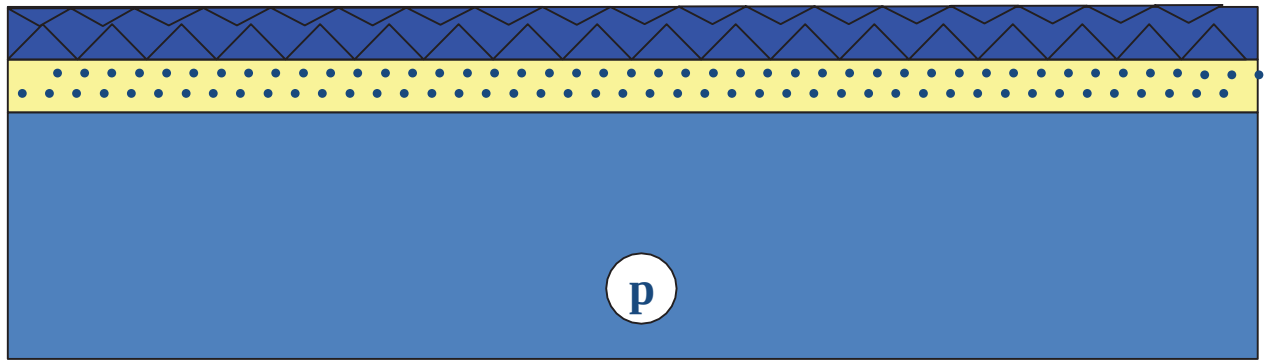
Photoresist



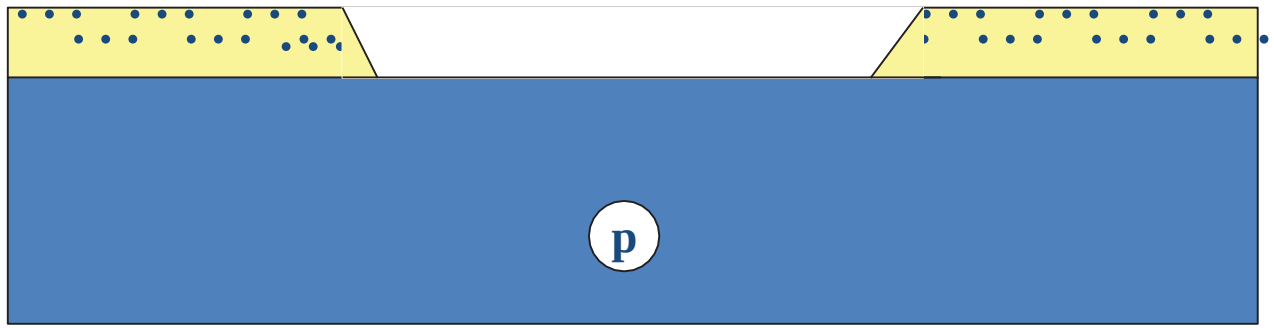
UV light

Mask

4.

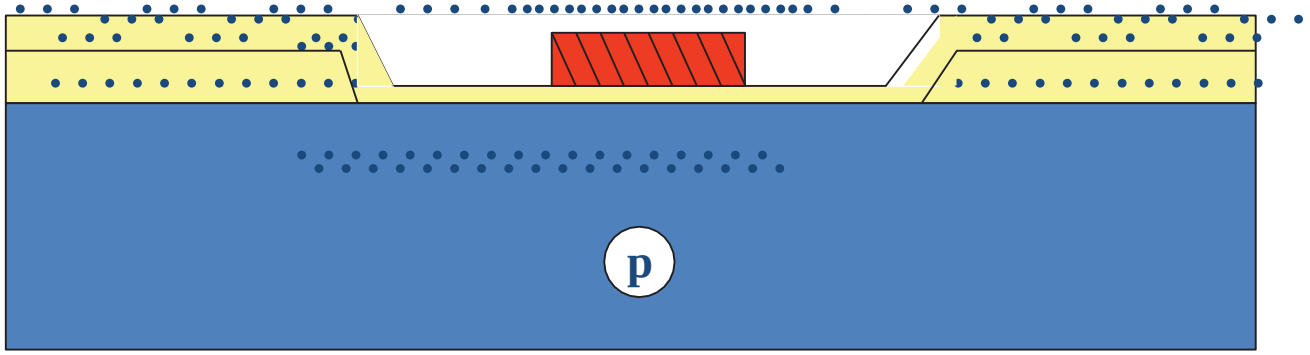


5.



Window in oxide

6.

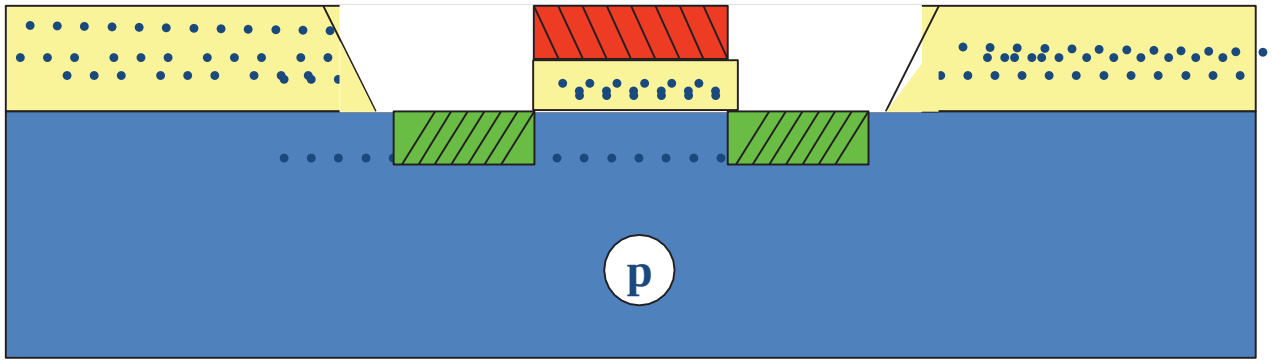


Patterned Poly. (1-2 μm)

On thin oxide

(800-1000A⁰)

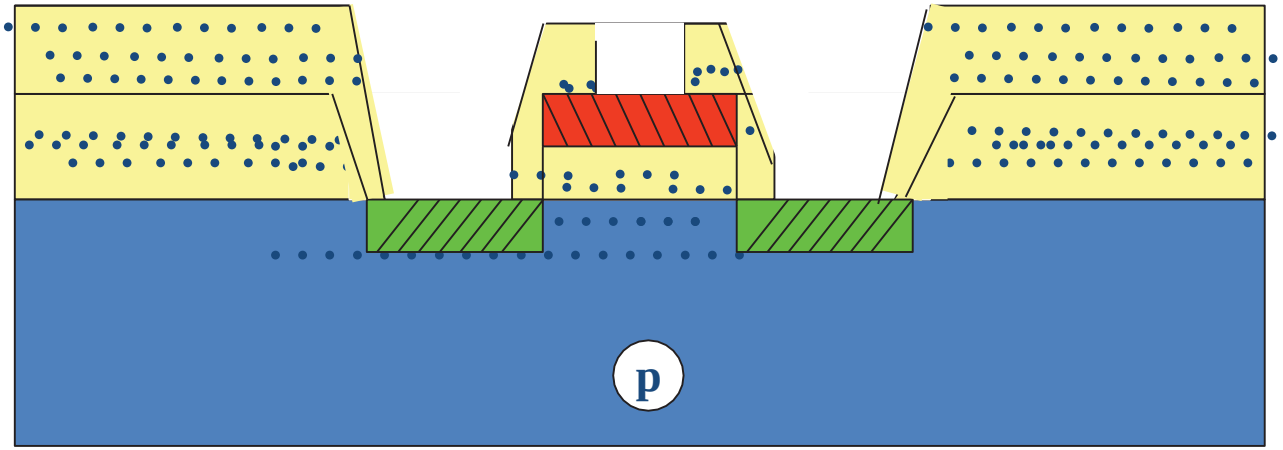
7.



n^+ diffusion

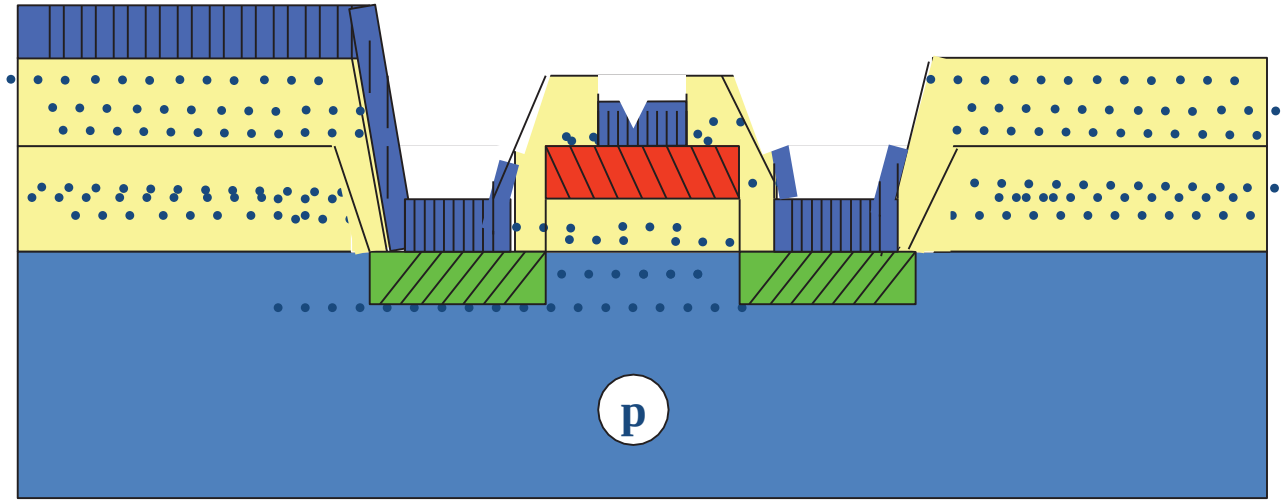
(1 μm deep)

8.



Contact holes (cuts)

9.



Patterned Metallization (aluminum 1 μm)

CMOS FABRICATION

- There are a number of approaches to CMOS fabrication, including the p-well, the n-well, the twin-tub, and the silicon-on-insulator processes.

The p-well Process

In primitive terms, the structure consists of an n-type substrate in which p-devices may

be formed by suitable masking and diffusion and, in order to accommodate n-type devices,

a deep p-well is diffused into the n-type substrate as shown.

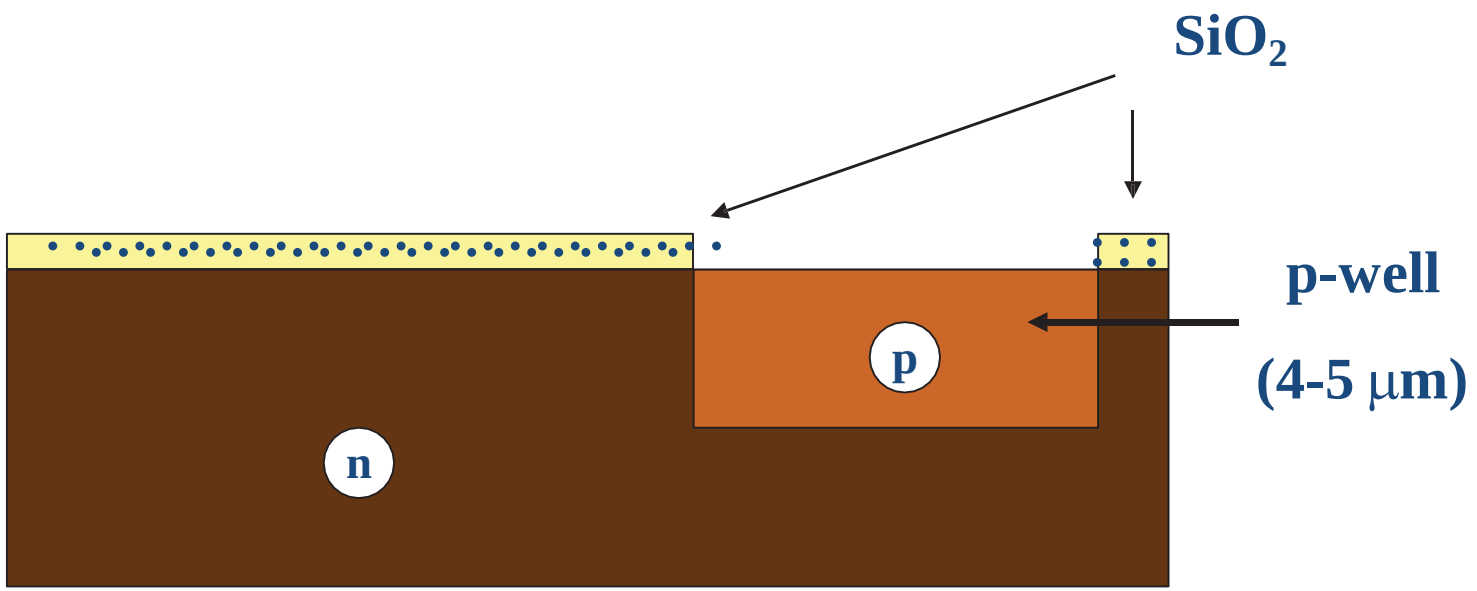
The p-well CMOS fabrication

In all other respects-masking, patterning, and diffusion-the process is similar to nMOS fabrication. In summary, typical processing steps are:

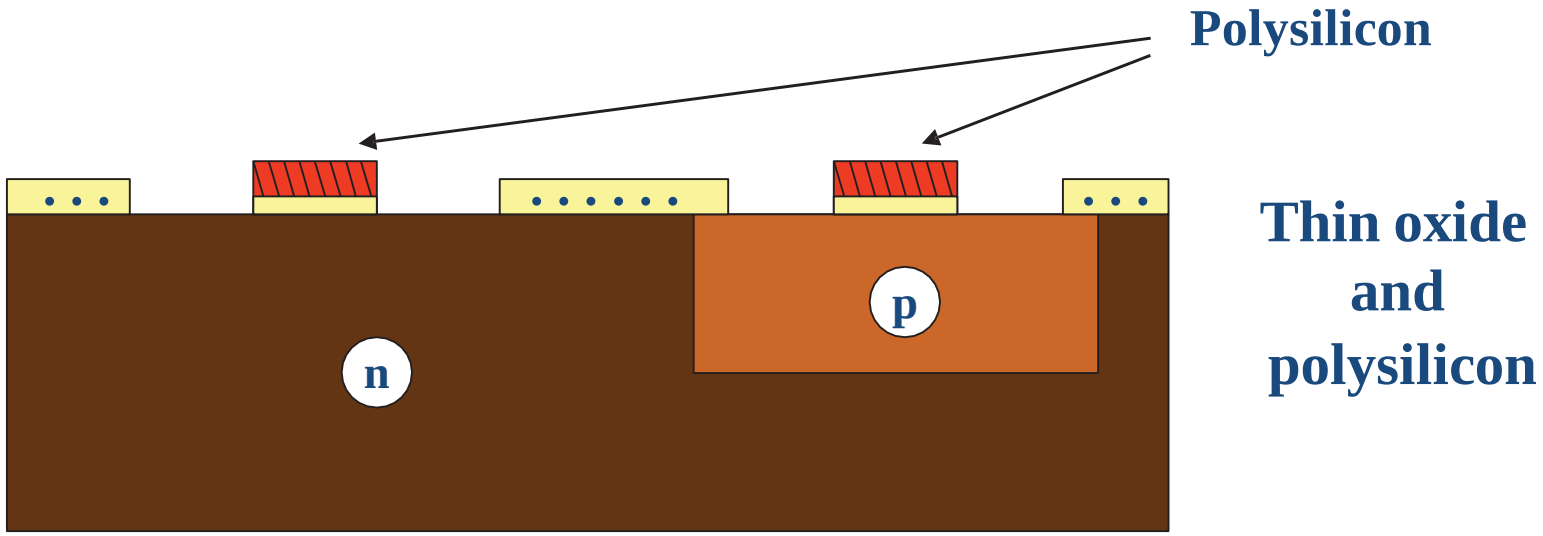
- *Mask* 1 - defines the areas in which the deep p-well diffusions are to take place.
- *Mask 2* - defines the thinox regions, namely those areas where the thick oxide is to be stripped and thin oxide grown to accommodate p- and n-transistors and wires.
- *Mask* 3 - used to pattern the polysilicon layer which is deposited after the thin oxide.
- *Mask* 4 - A p-plus mask is now used (to be in effect "Anded" with Mask 2) to define all areas where p-diffusion is to take place.
- *Mask* 5 - This is usually performed using the negative form of the p-plus mask and defines those areas where n-type diffusion is to take place.
- *Mask* 6 - Contact cuts are now defined.
- *Mask* 7 - The metal layer pattern is defined by this mask.

Mask 8 - An overall passivation (overglass) layer is now applied and Mask 8 ts needed to define the openings for access to bonding pads.

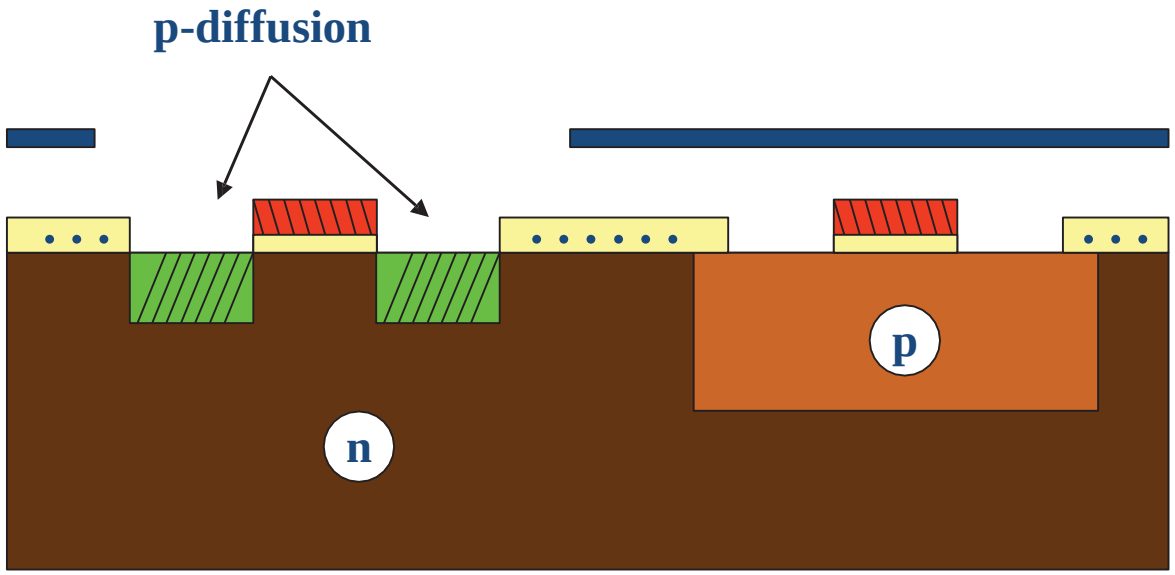
1.


SiO_2
p
n
p-well
(4-5 μm)

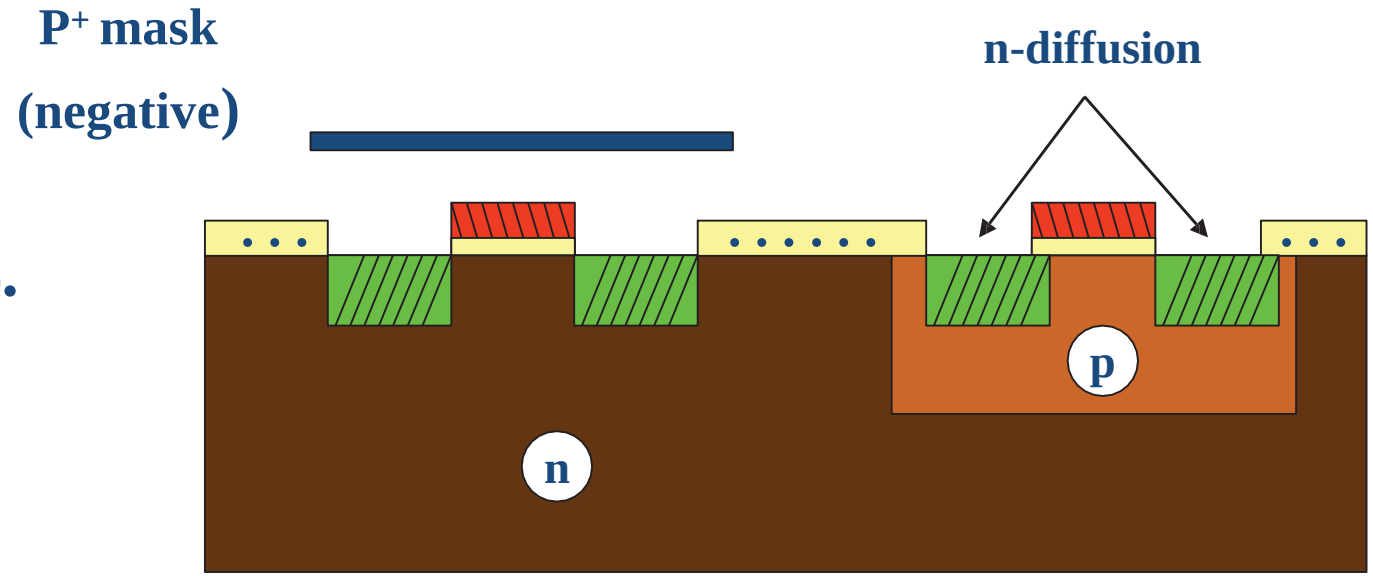
2.

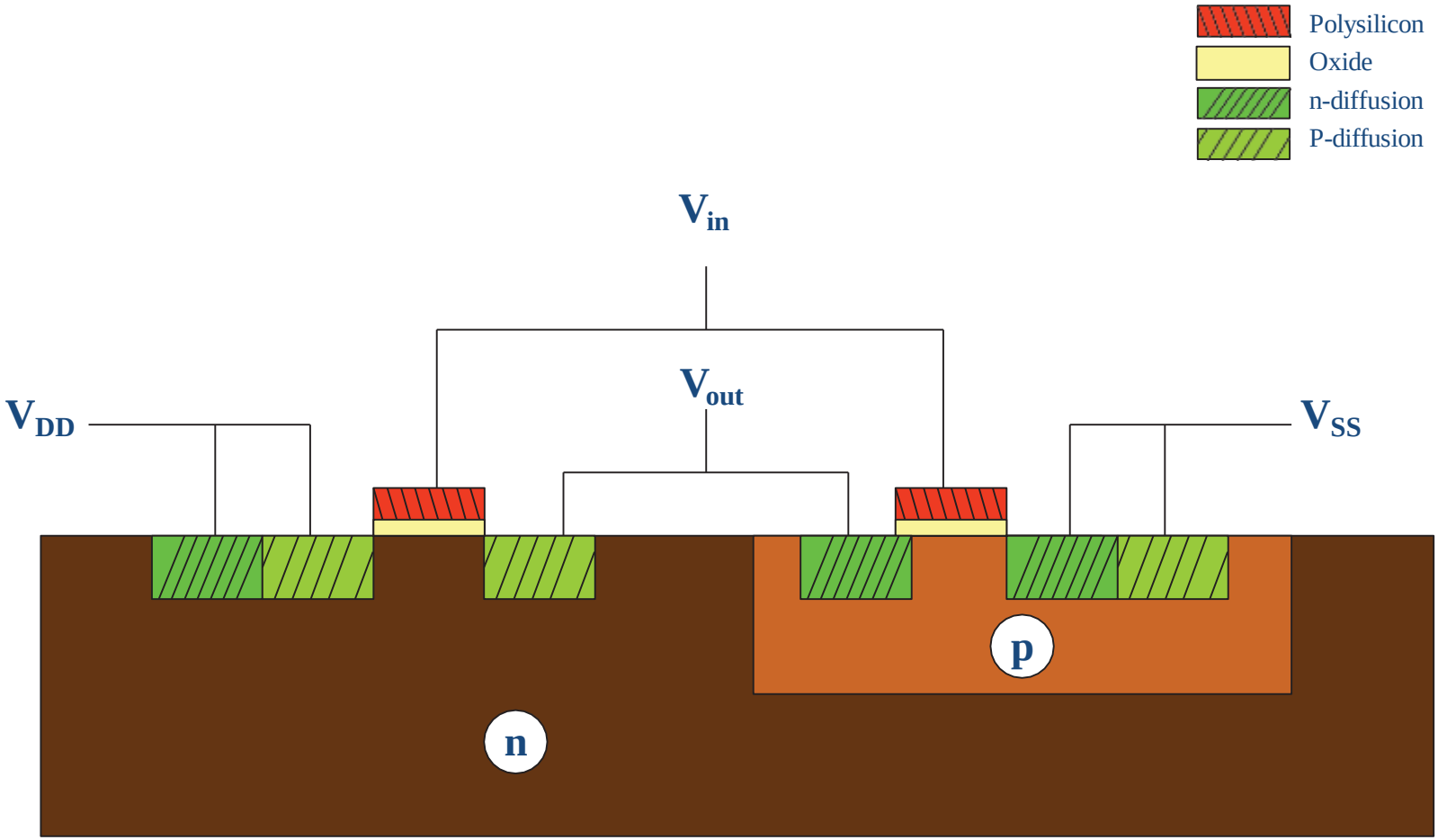

Polysilicon
p
n
Thin oxide
and
polysilicon

3.

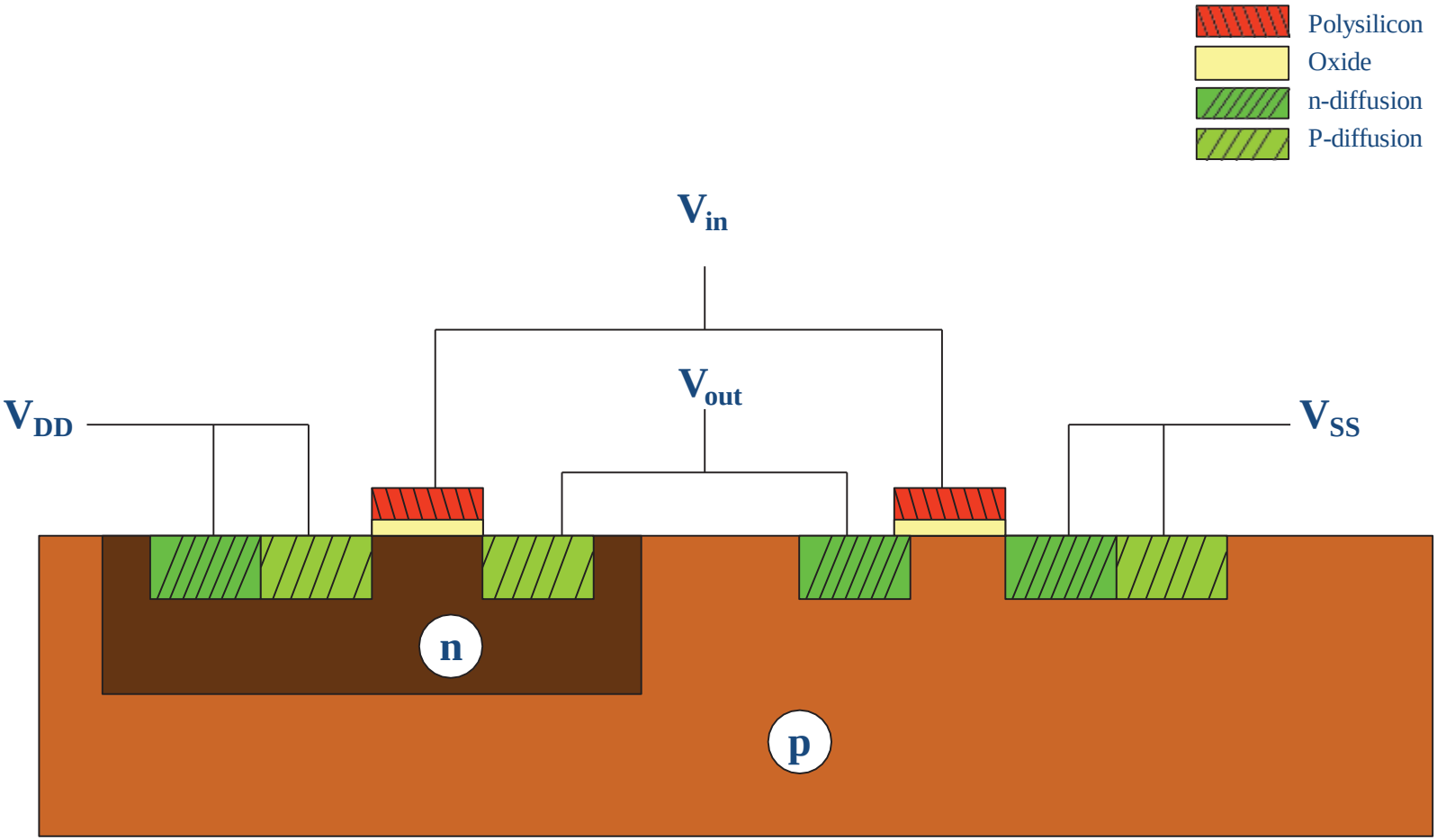

p-diffusion
P+ mask
(positive)
p
n

4.


P+ mask
(negative)
n-diffusion
p
n



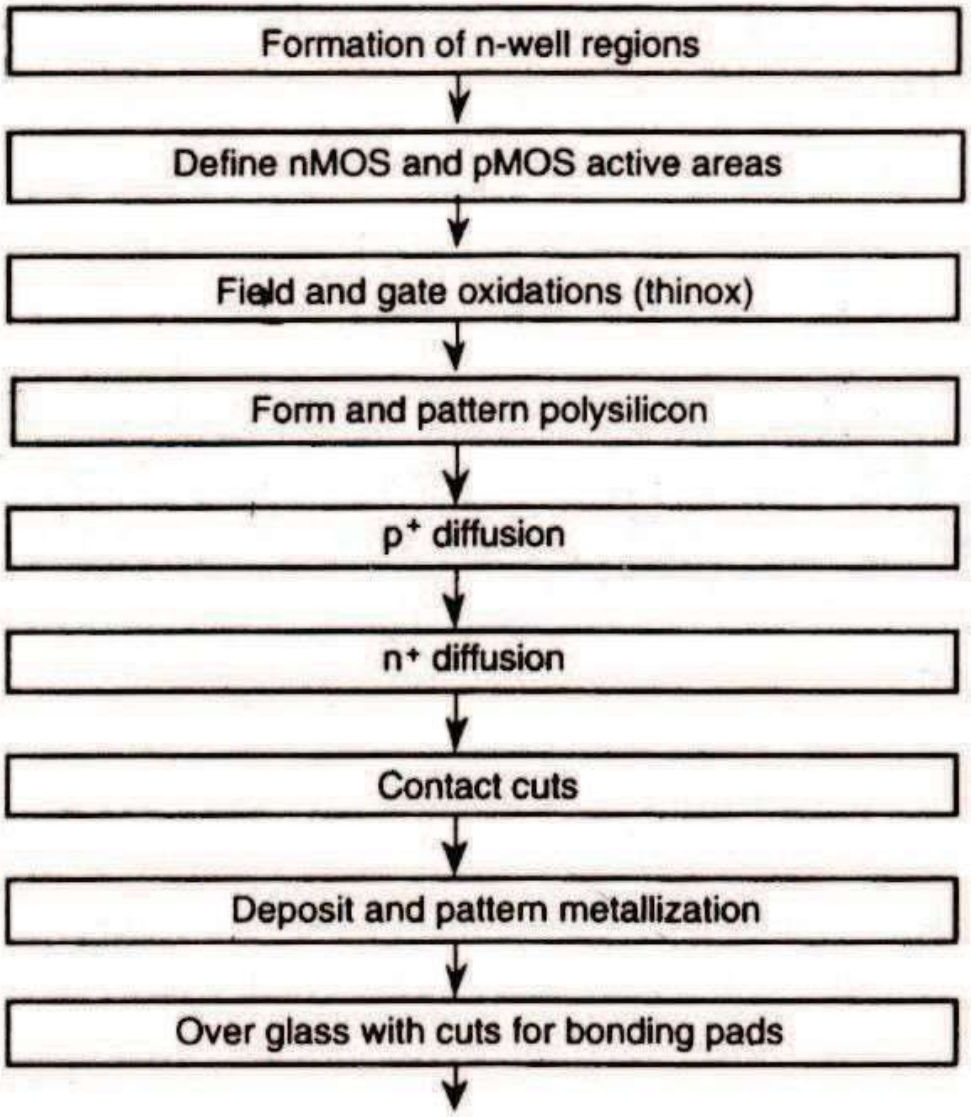
CMOS p-well inverter showing V_{DD} and V_{SS} substrate connections

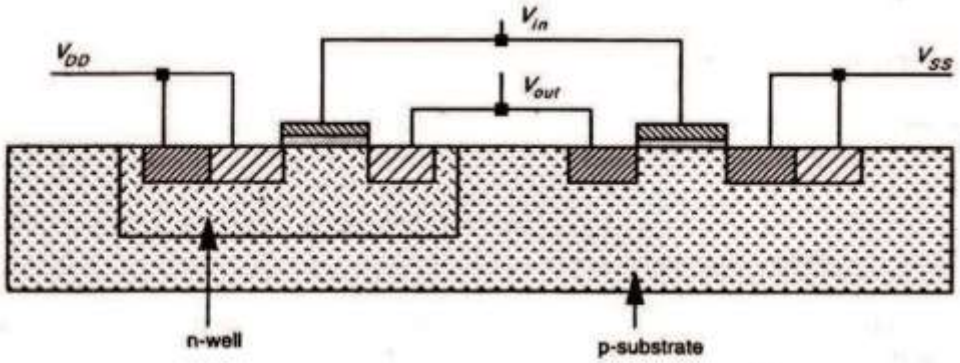


CMOS n-well inverter showing V_{DD} and V_{SS} substrate connections

The n-well Process

- As indicated earlier, although the p-well process is widely used, n-well fabrication has also gained wide acceptance, initially as a retrofit to nMOS lines.





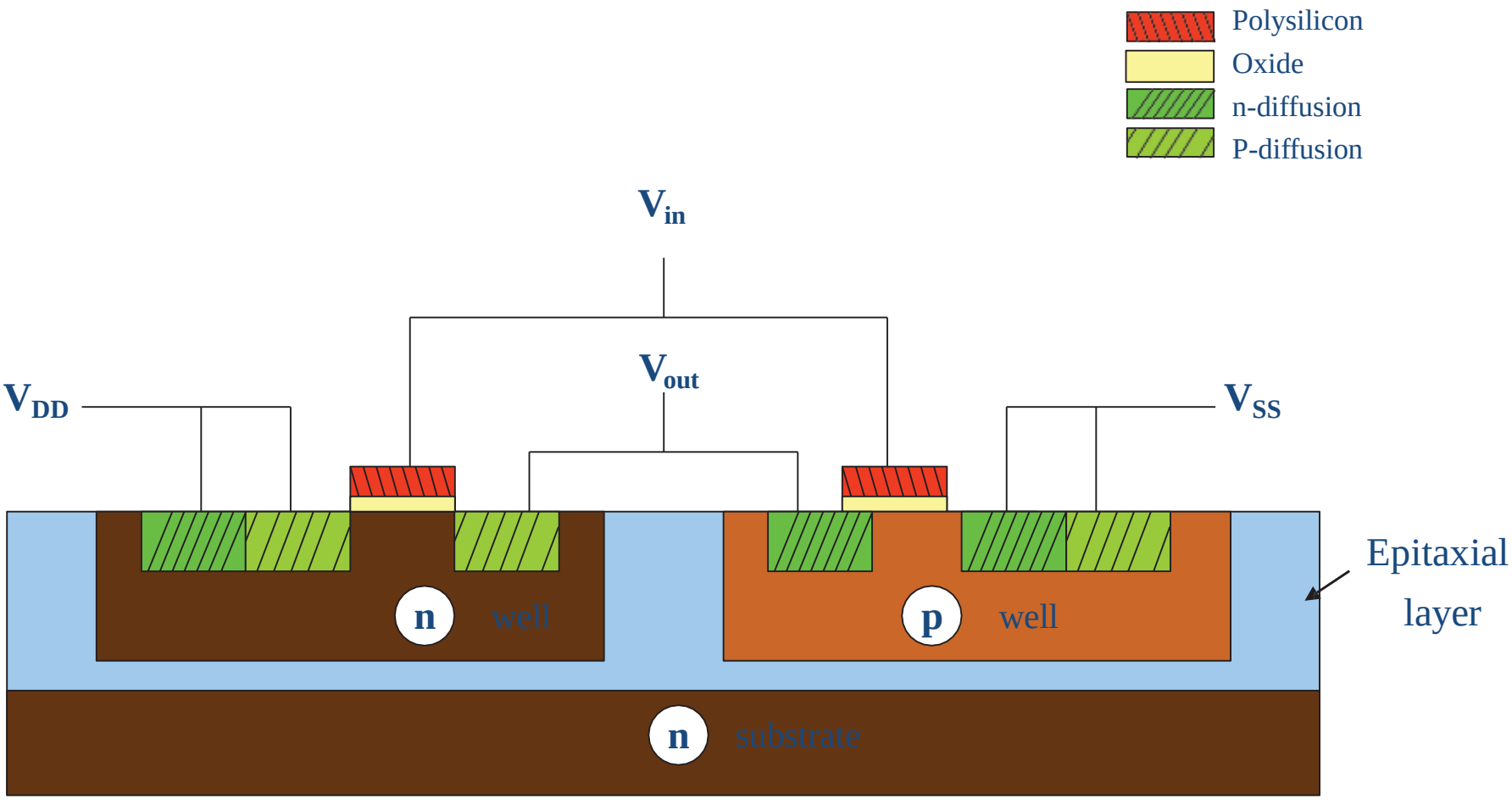
The twin-tub-Tub Process

A logical extension of the p-well and n-well approaches is the twin-tub fabrication process.

Here we start with a substrate of high resistivity n-type material and then create both .. n-well and p-well regions. Through this process it is possible to preserve the performance of n-transistors without compromising the p-transistors. Doping control is more readily achieved

and some relaxation in manufacturing tolerances results. This is particularly important as far

as latch-up is concerned.

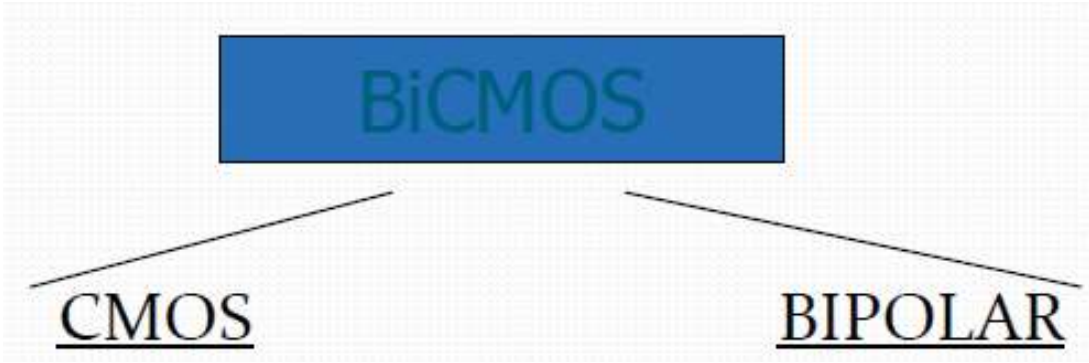


Twin-tub structure

(Alogical extension of the p-well and n-well)

Bi-CMOS

Bipolar compatible CMOS(Bi-CMOS) technology:
Introduced in early 1980s
Combines Bipolar and CMOS logic



Low power dissipation
High packing density
High Noise Margin
High input impedance

High speed
High output drive
High transconductance (g_m)

Features

The objective of the Bi-CMOS is to combine bipolar and CMOS so as to exploit the advantages of both the technologies.

Today Bi-CMOS has become one of the dominant technologies used for high speed, low power and highly functional VLSI circuits.

The process step required for both CMOS and bipolar are almost similar

The primary approach to realize high performance Bi-CMOS devices is the addition of bipolar process steps to a baseline CMOS process.

The Bi-CMOS gates could be used as an effective way of speeding up the VLSI circuits.

The applications of Bi-CMOS are vast.

Advantages of bipolar and CMOS circuits can be retained in Bi-CMOS chips.

Characteristics of Bipolar Technology

Higher switching speed

Higher current drive per unit area, higher gain

Generally better noise performance and better high frequency characteristics

Improved I/O speed (particularly significant with the growing importance of package limitations in high speed systems).

high power dissipation

lower input impedance (high drive current)

low packing density

low delay sensitivity to load

Characteristics of CMOS

Lower static power dissipation

Higher noise margins

Higher packing density

High yield with large integrated complex functions

High input impedance (low drive current)

Scalable threshold voltage

High delay load sensitivity

Low output drive current (issue when driving large capacitive loads)

Bi-directional capability (drain & source are interchangeable)

A near ideal switching device,
Low gain

Bi-CMOS FABRICATION PROCESS

CMOS process

process

1 . N-well

3. PMOS source and drain

4. NMOS source and drain

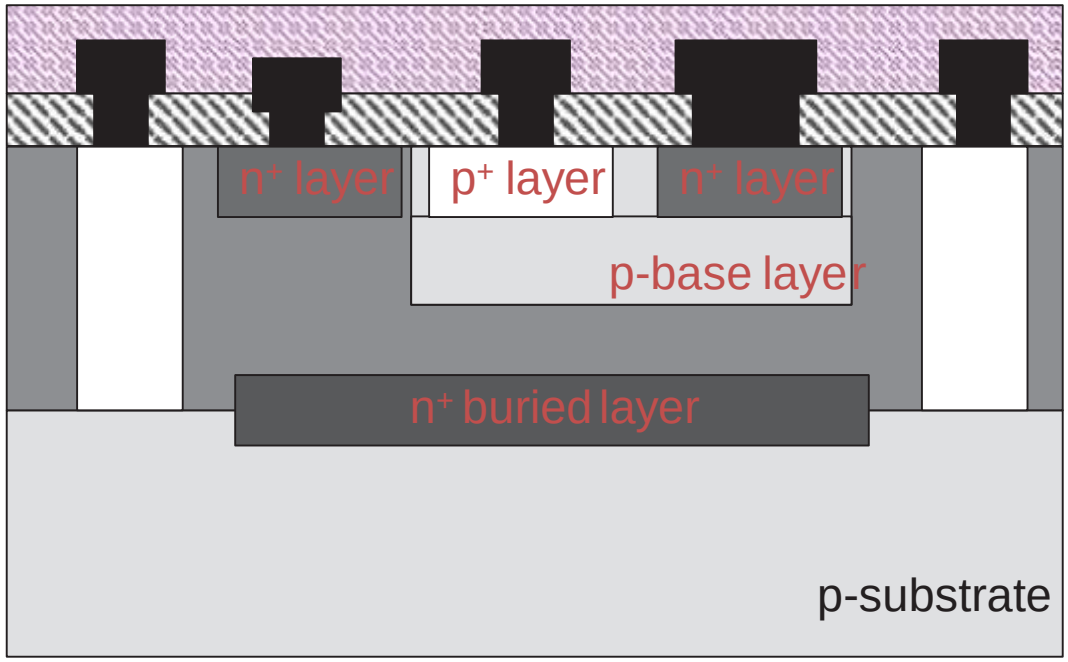
BI-POLAR

1. n+ sub-collector
2. P base doping(extra step)
3. p+ base contact
4. n+ emitter

npn-BJT Fabrication

BJT Processing

1. Implantation of the buried n^+ layer
2. Growth of the epitaxial layer
3. p^+ isolation diffusion
4. Base p-type diffusion
5. Emitter n^+ diffusion
6. p^+ ohmic contact
7. Contact etching
8. Metal deposition and etching
9. Passivation and bond pad opening

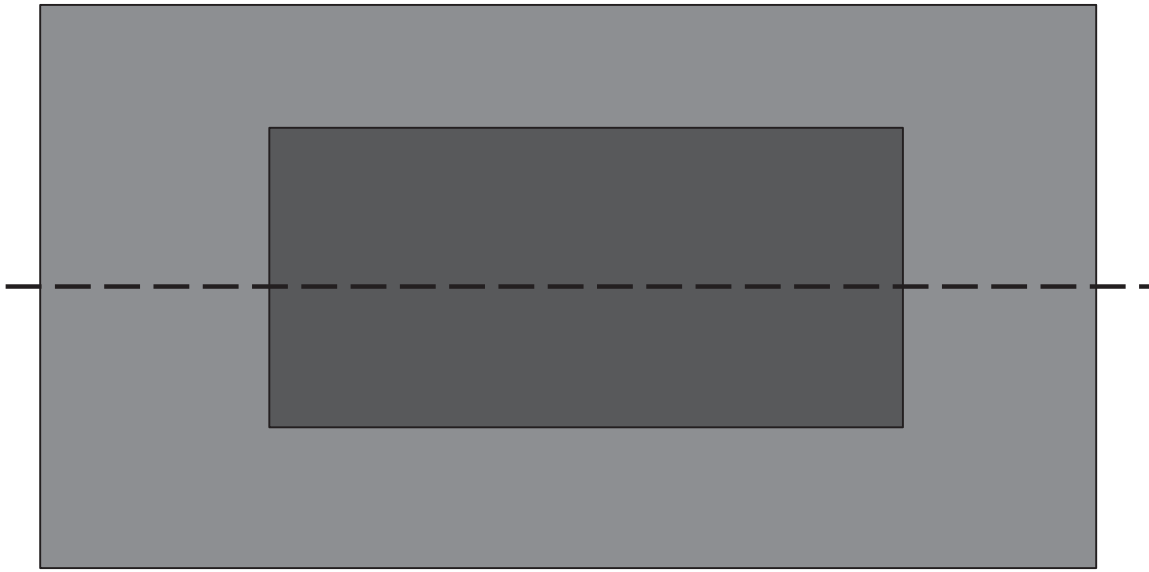


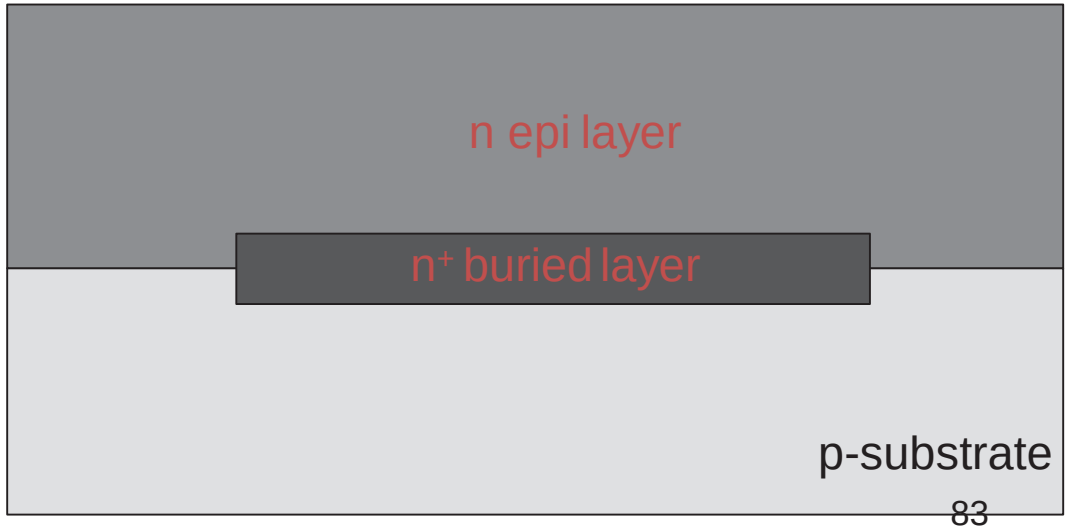
BJT Processing

1. Implantation of the buried n^+ layer
2. Growth of the epitaxial layer
3. p^+ isolation diffusion
4. Base p-type diffusion
5. Emitter n^+ diffusion
6. p^+ ohmic contact
7. Contact etching
8. Metal deposition and etching
9. Passivation and bond pad opening

BJT PROCESSING

1. Implantation of the buried n^+ layer
2. Growth of the epitaxial layer
3. p^+ isolation diffusion
4. Base p-type diffusion
5. Emitter n^+ diffusion
6. p^+ ohmic contact
7. Contact etching
8. Metal deposition and etching
9. Passivation and bond pad opening



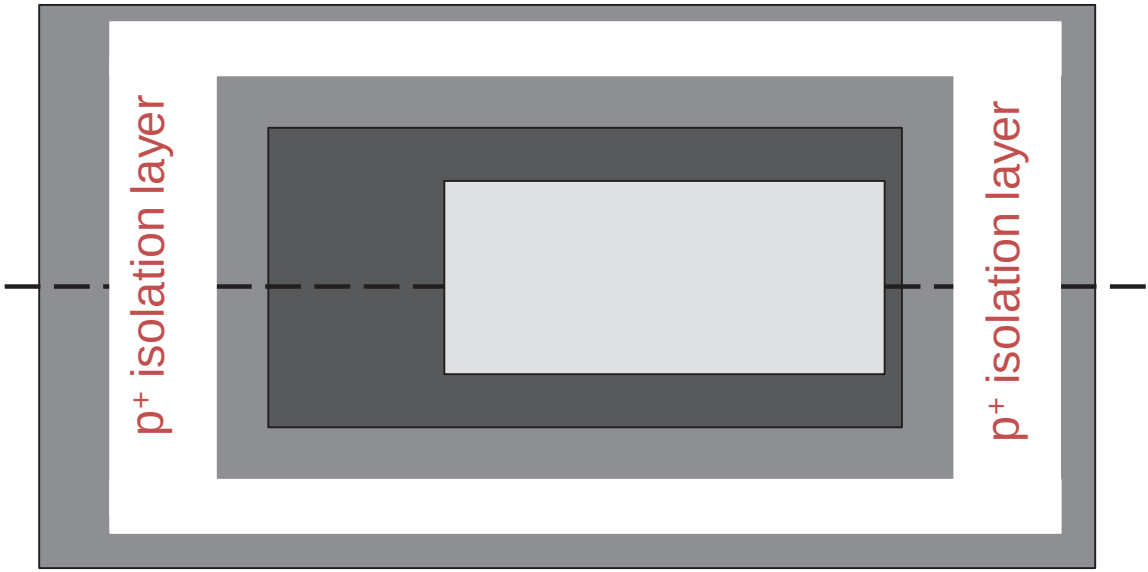


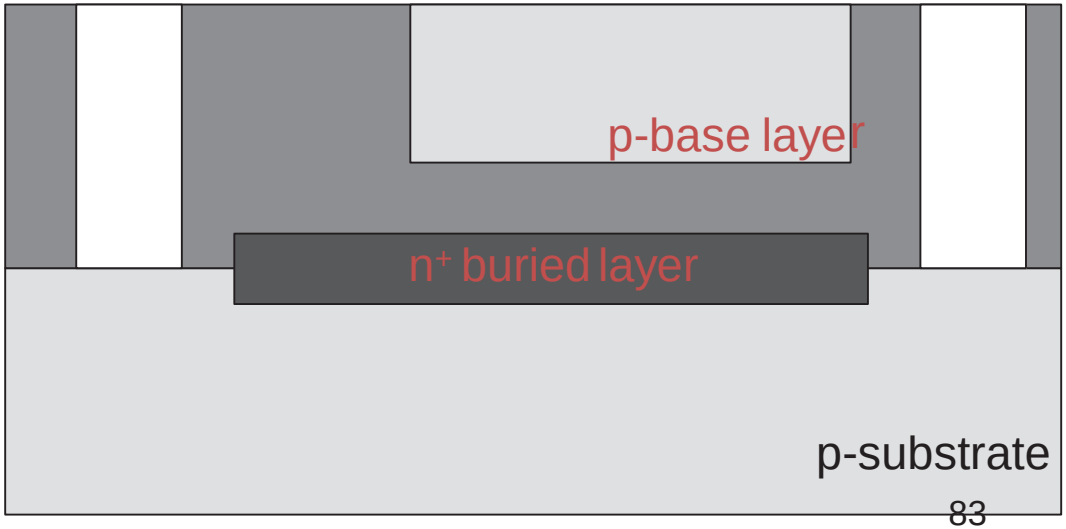
BJT Processing

1. Implantation of the buried n^+ layer
2. Growth of the epitaxial layer
3. p^+ isolation diffusion
4. Base p-type diffusion
5. Emitter n^+ diffusion
6. p^+ ohmic contact
7. Contact etching
8. Metal deposition and etching
9. Passivation and bond pad opening

BJT PROCESSING

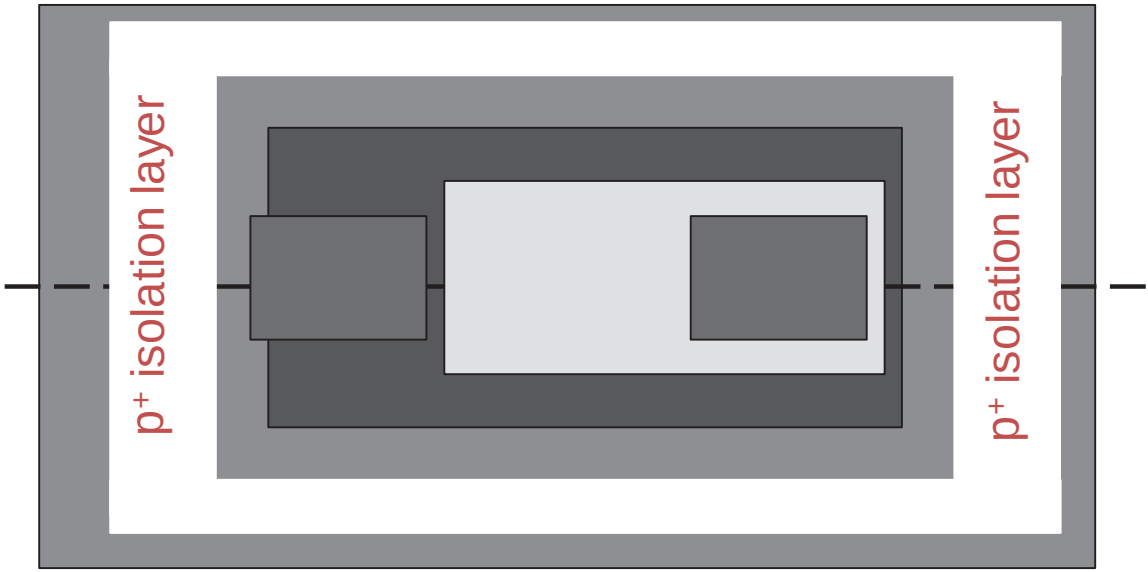
1. Implantation of the buried n^+ layer
2. Growth of the epitaxial layer
3. p^+ isolation diffusion
4. Base p-type diffusion
5. Emitter n^+ diffusion
6. p^+ ohmic contact
7. Contact etching
8. Metal deposition and etching
9. Passivation and bond pad opening

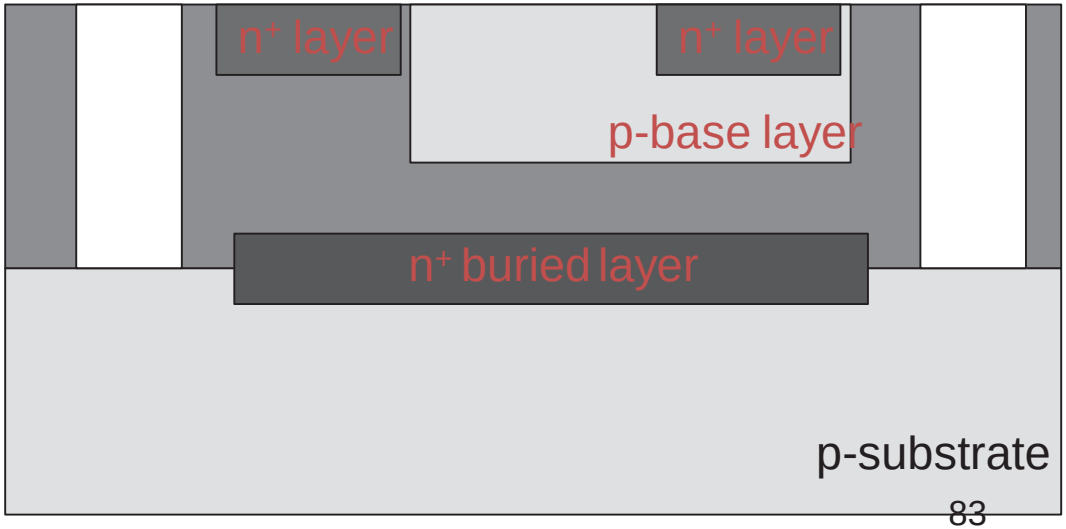




BJT PROCESSING

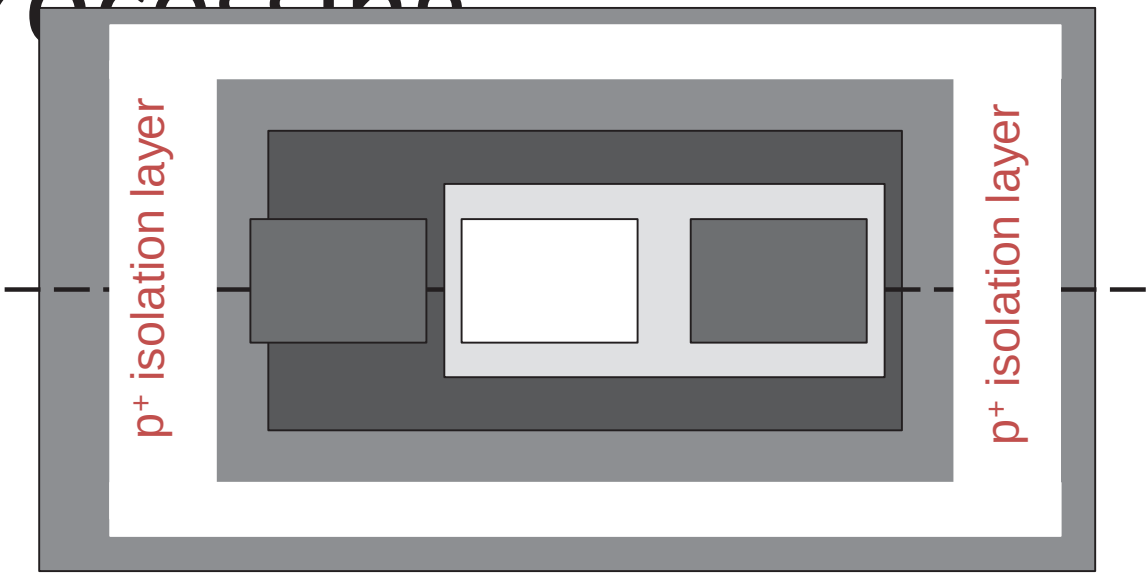
1. Implantation of the buried n^+ layer
2. Growth of the epitaxial layer
3. p^+ isolation diffusion
4. Base p-type diffusion
5. Emitter n^+ diffusion
6. p^+ ohmic contact
7. Contact etching
8. Metal deposition and etching
9. Passivation and bond pad opening

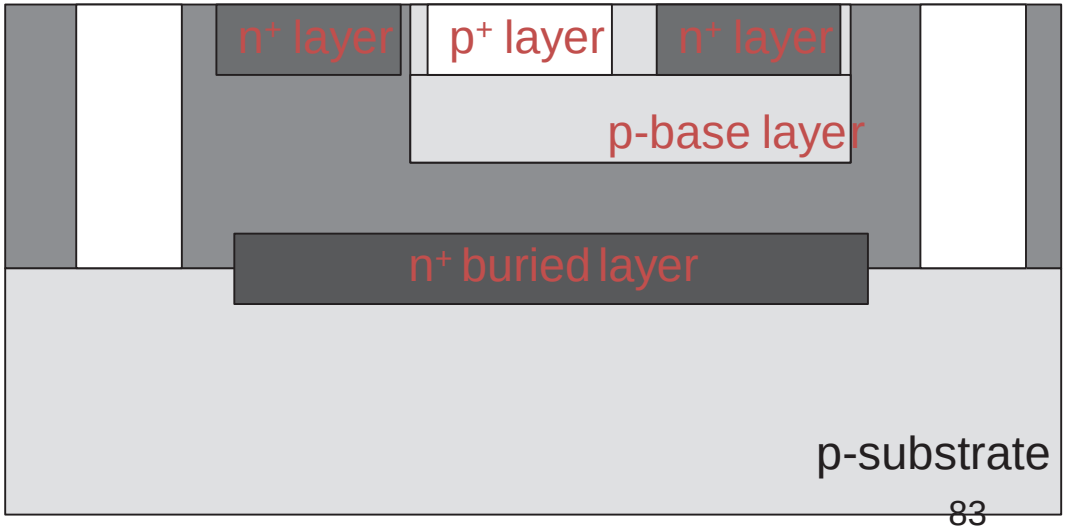




BJT Processing

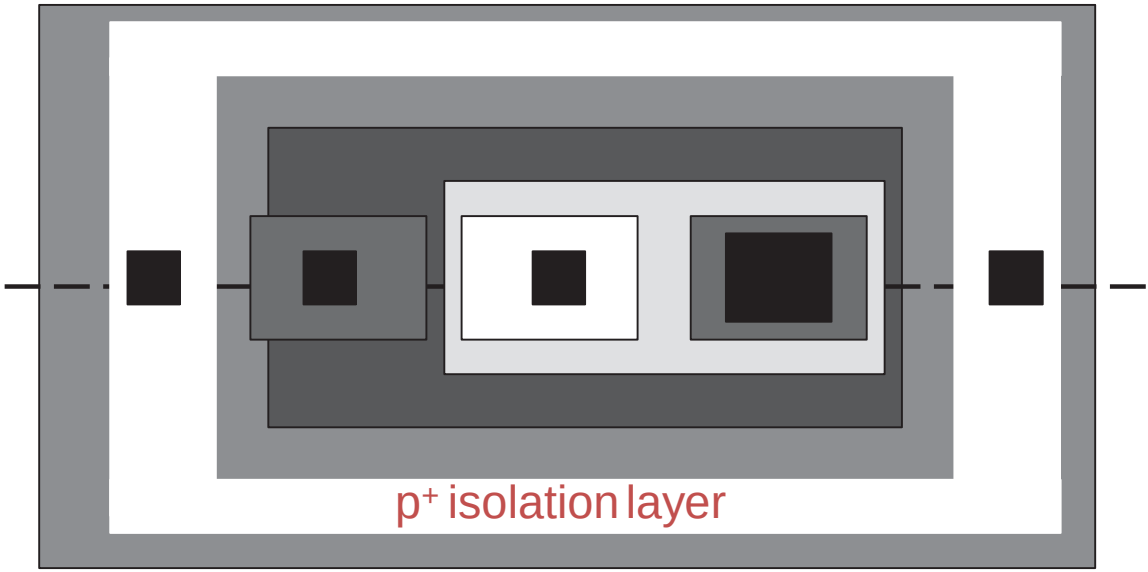
1. Implantation of the buried n^+ layer
2. Growth of the epitaxial layer
3. p^+ isolation diffusion
4. Base p-type diffusion
5. Emitter n^+ diffusion
6. p^+ ohmic contact
7. Contact etching
8. Metal deposition and etching
9. Passivation and bond pad opening

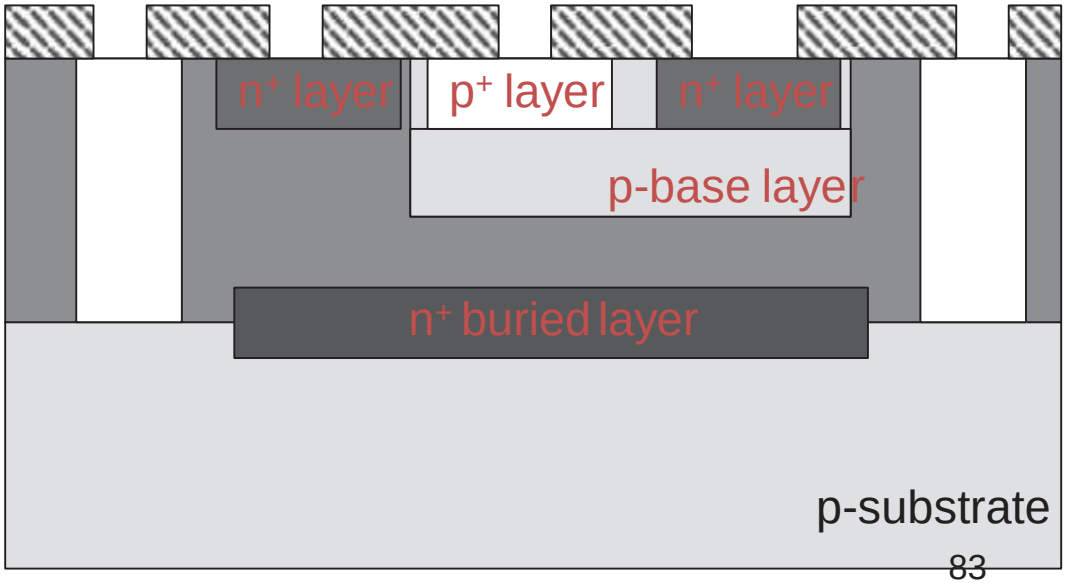




BJT PROCESSING

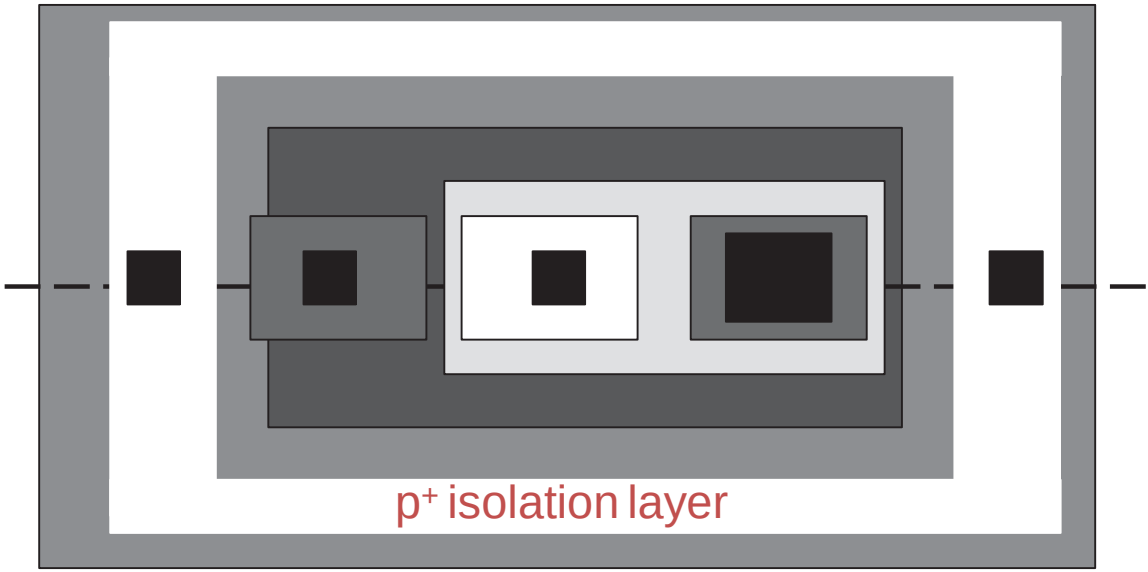
1. Implantation of the buried n^+ layer
2. Growth of the epitaxial layer
3. p^+ isolation diffusion
4. Base p-type diffusion
5. Emitter n^+ diffusion
6. p^+ ohmic contact
7. Contact etching
8. Metal deposition and etching
9. Passivation and bond pad opening

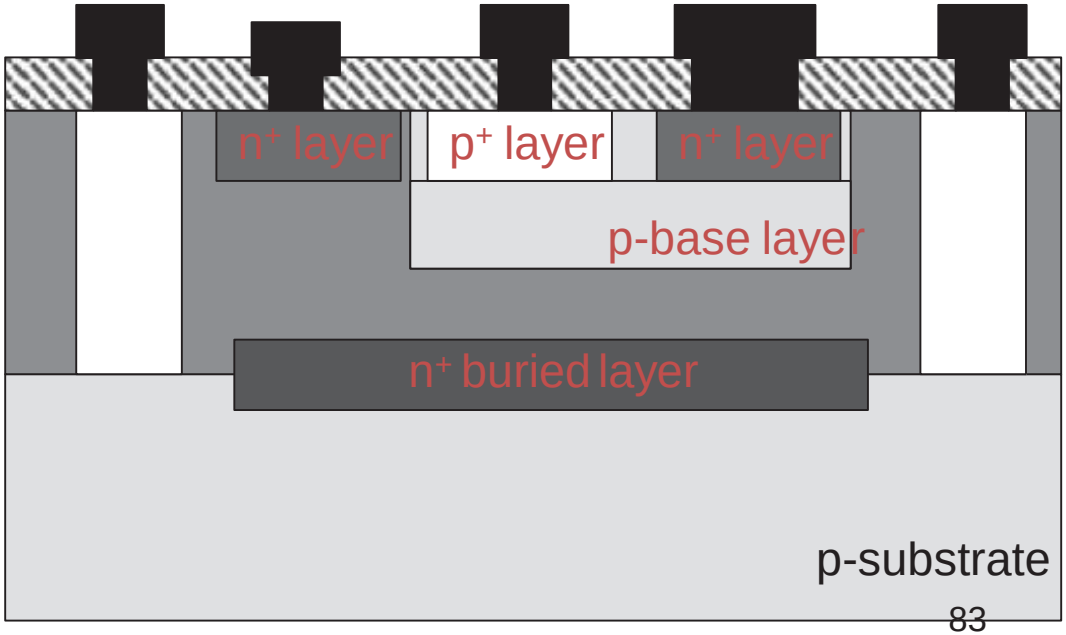




BJT Processing

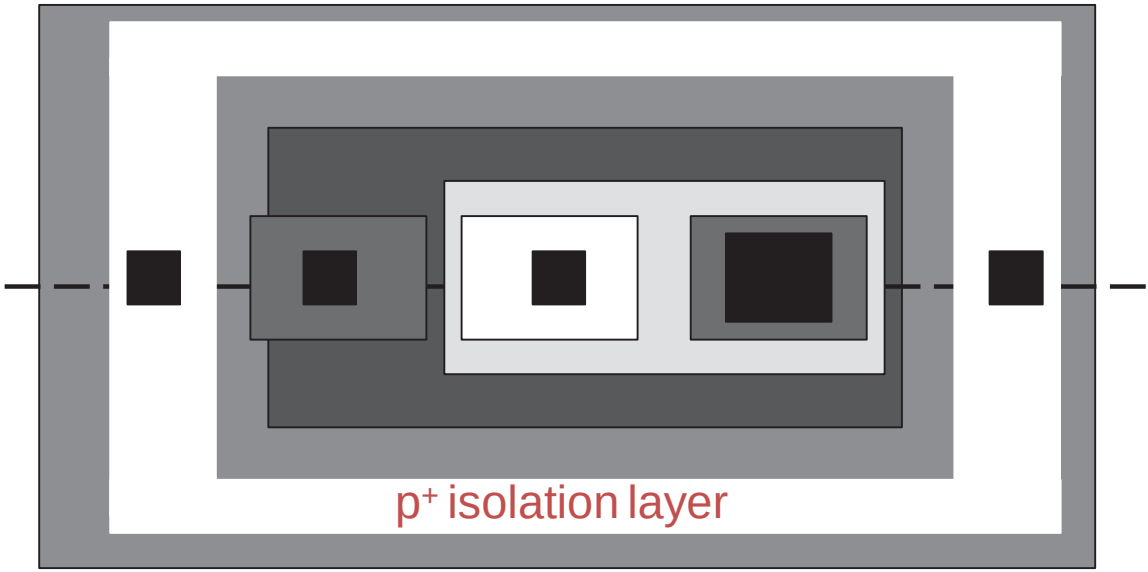
1. Implantation of the buried n^+ layer
2. Growth of the epitaxial layer
3. p^+ isolation diffusion
4. Base p-type diffusion
5. Emitter n^+ diffusion
6. p^+ ohmic contact
7. Contact etching
8. Metal deposition and etching
9. Passivation and bond pad opening

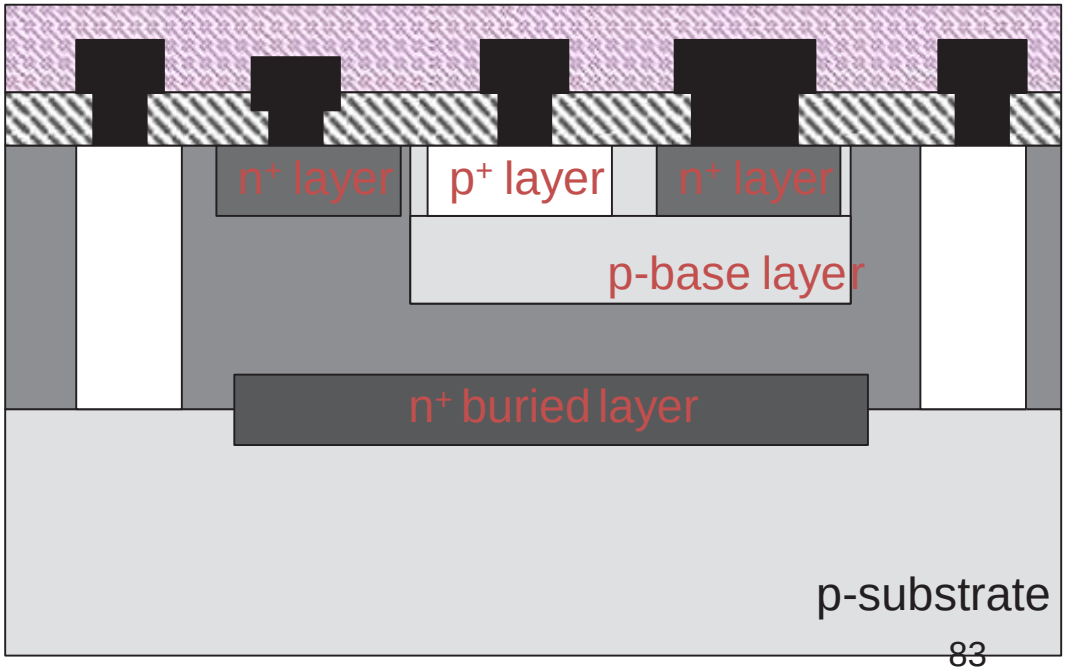




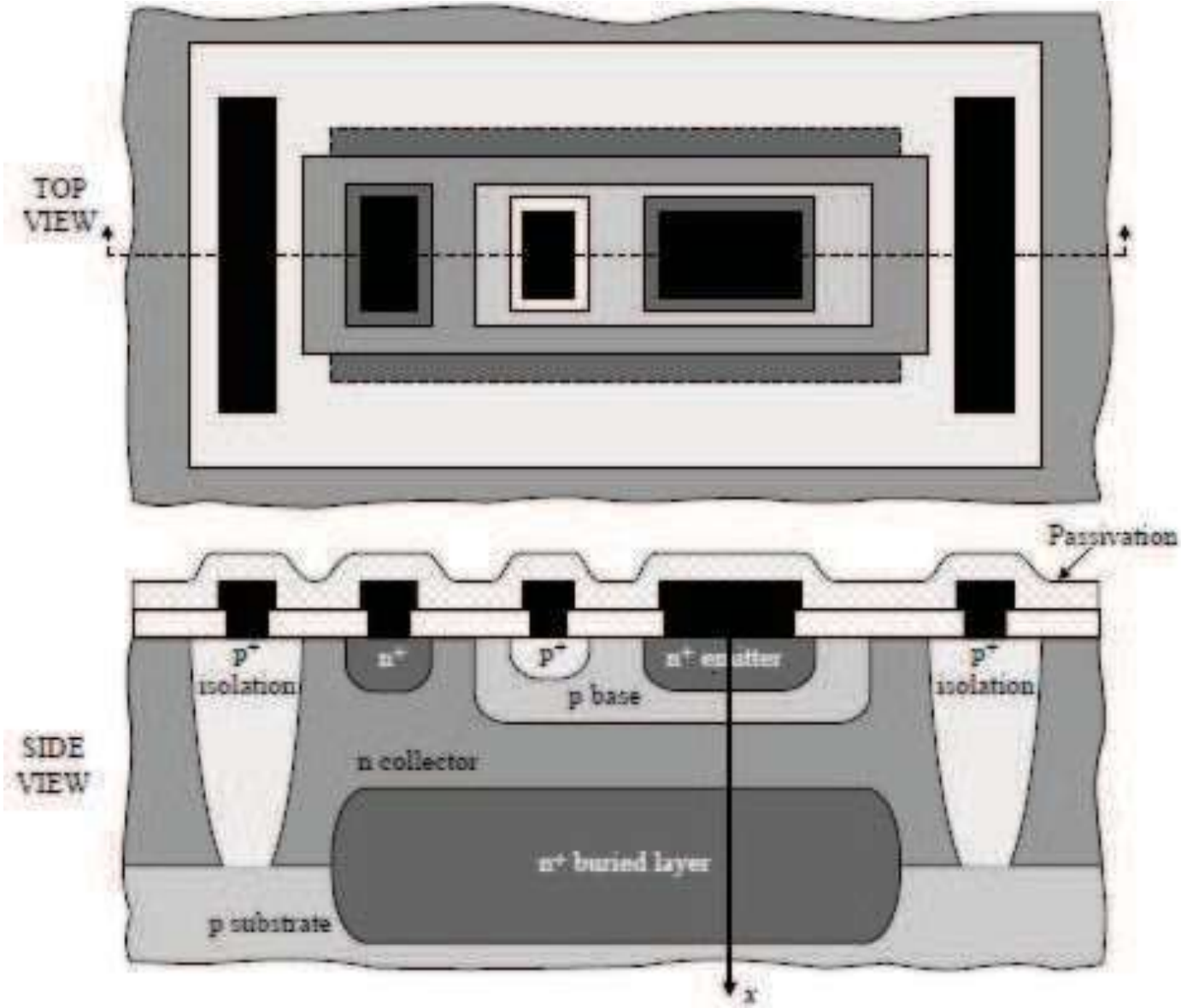
BJT PROCESSING

1. Implantation of the buried n^+ layer
2. Growth of the epitaxial layer
3. p^+ isolation diffusion
4. Base p-type diffusion
5. Emitter n^+ diffusion
6. p^+ ohmic contact
7. Contact etching
8. Metal deposition and etching
9. Passivation and bond pad opening

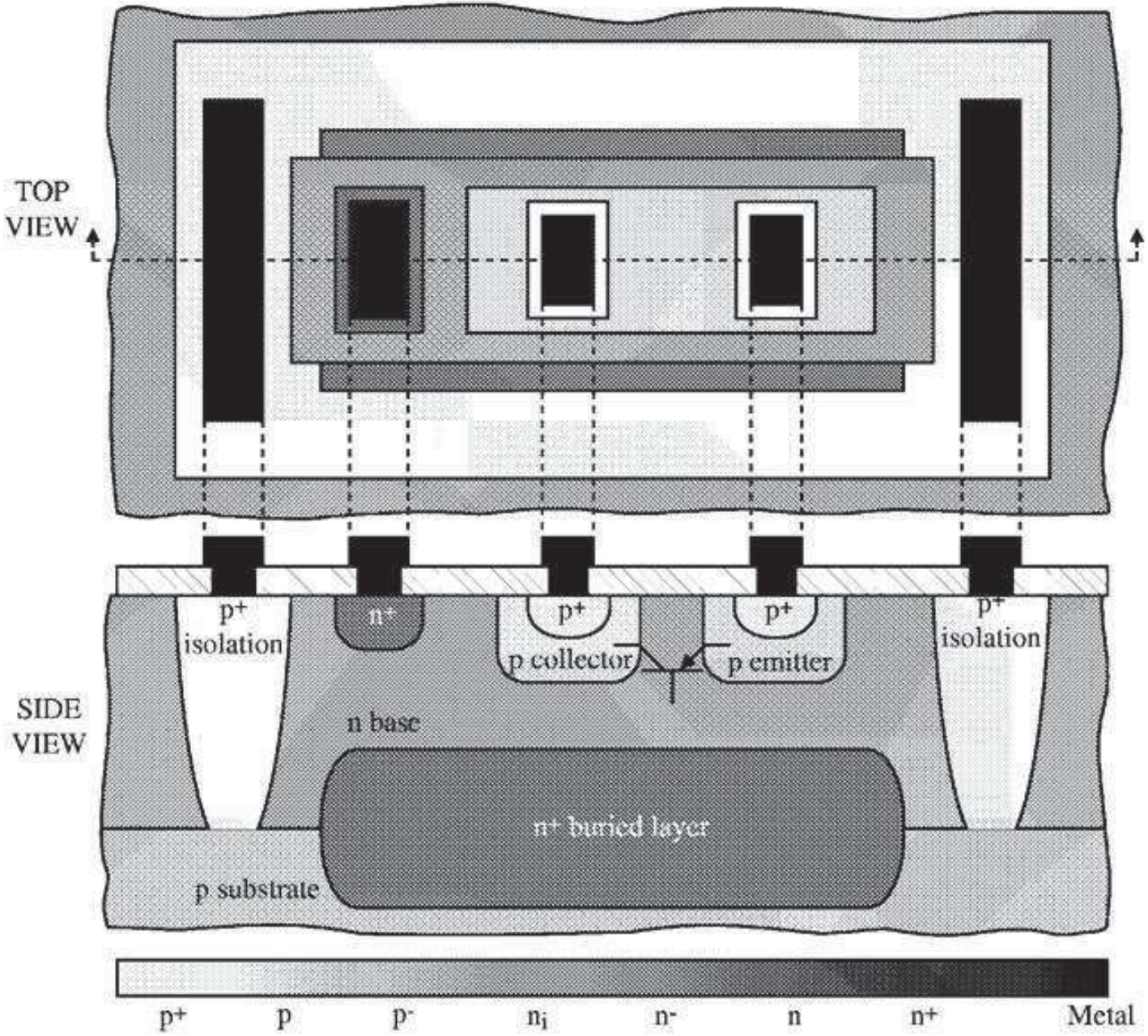




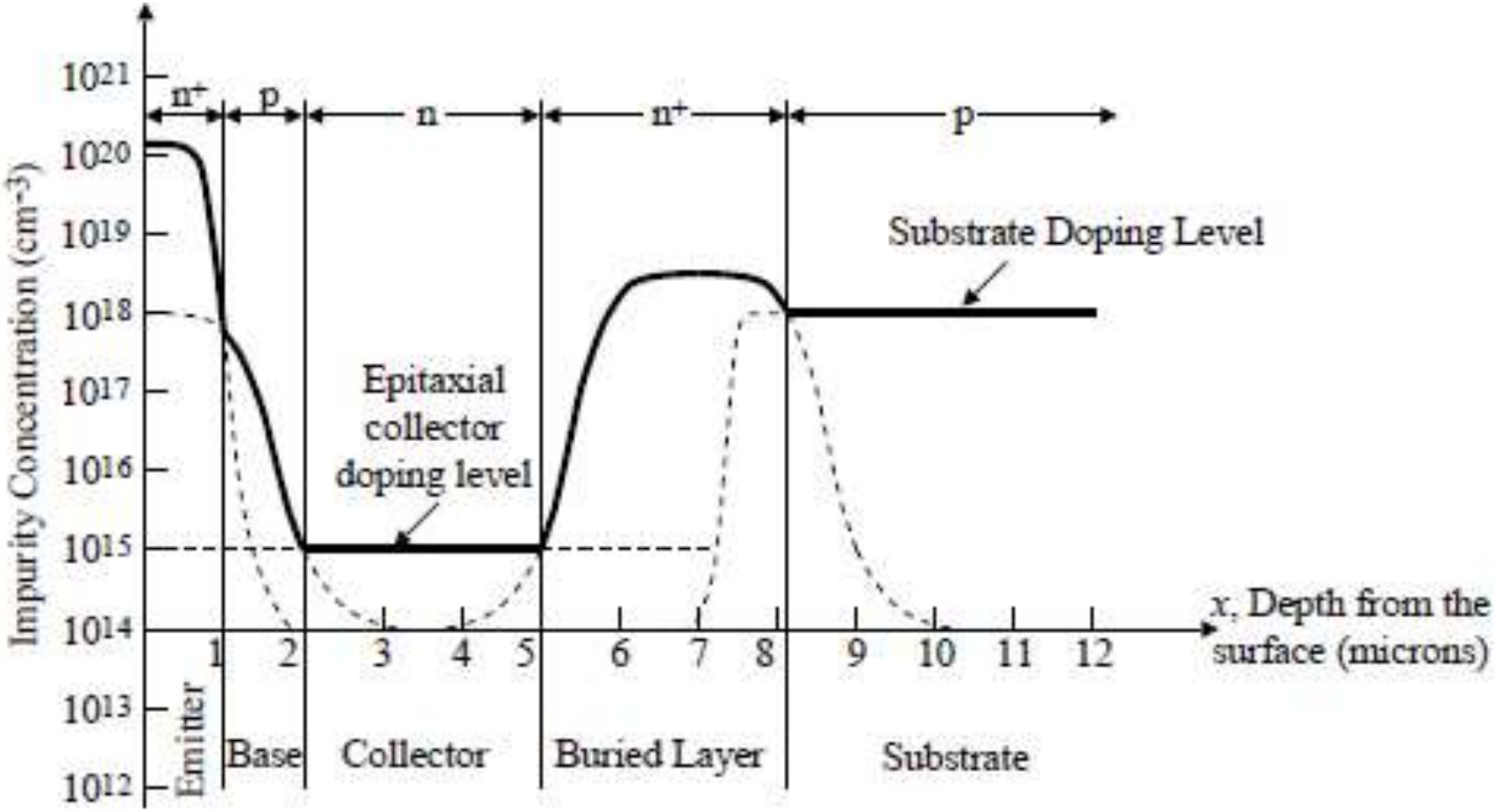
Lateral view of npn BJT



Lateral PNP BJT

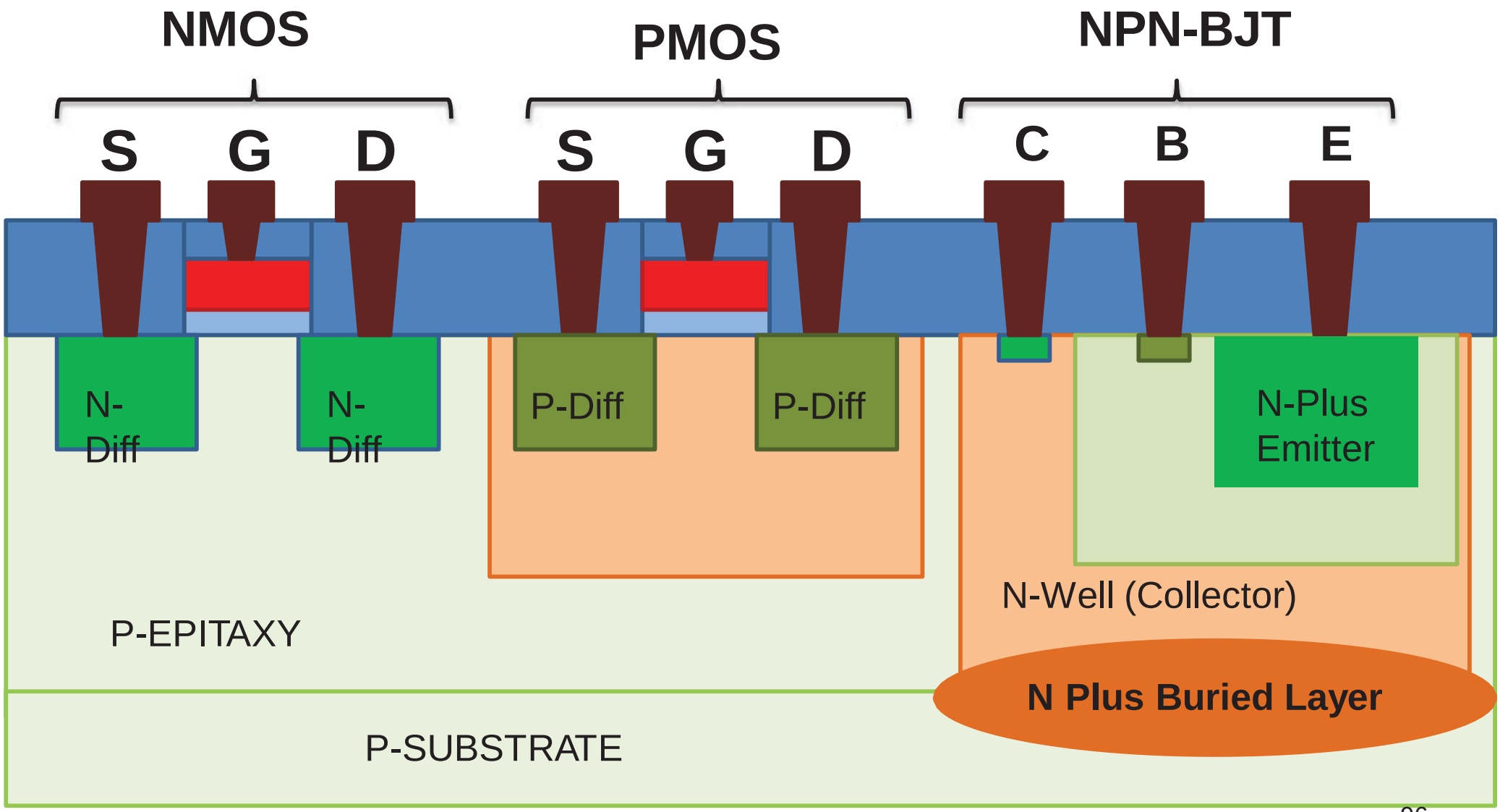


Doping Profiles in a BJT

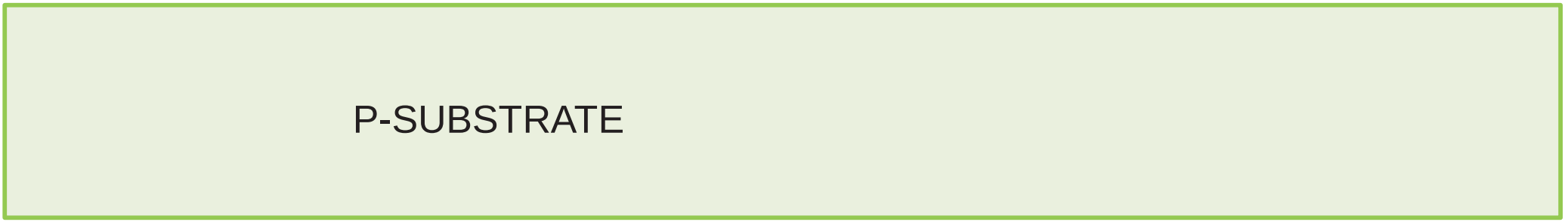


FABRICATION OF BICMOS THROUGH N-WELL

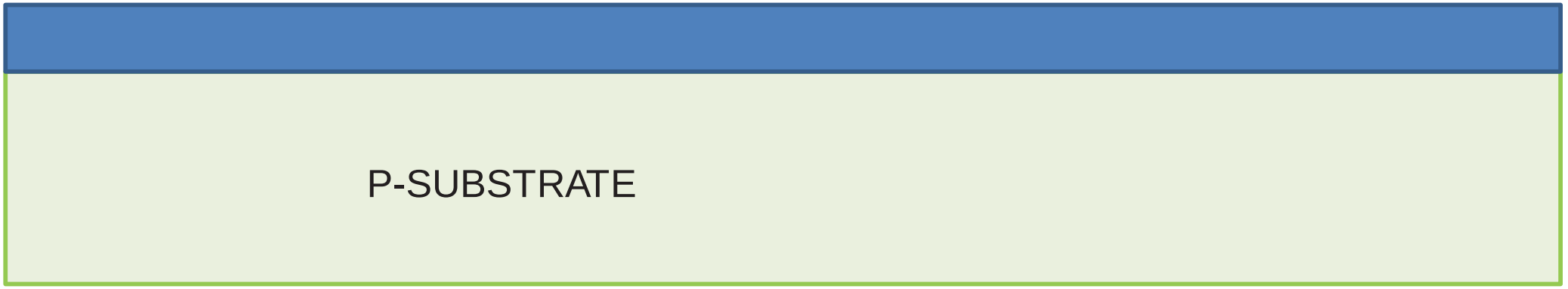
BICMOS STRUCTURE



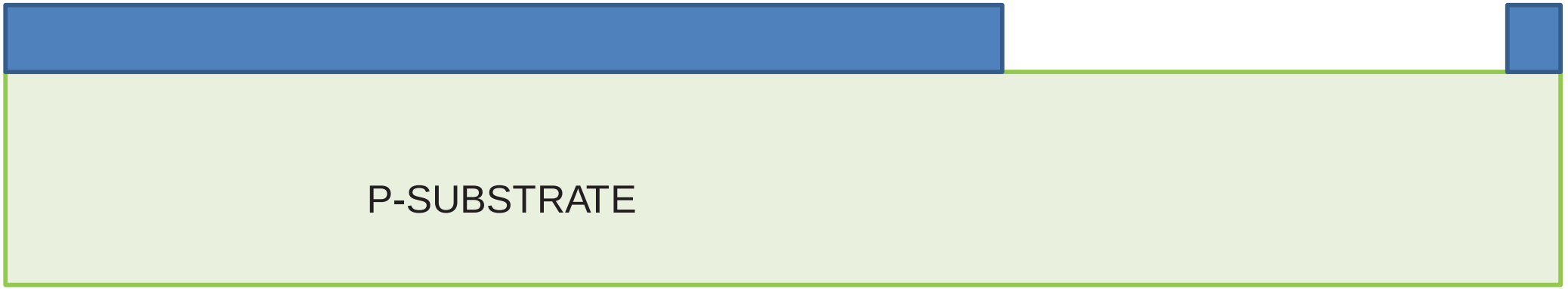
P-SUBSTRATE IS TAKEN



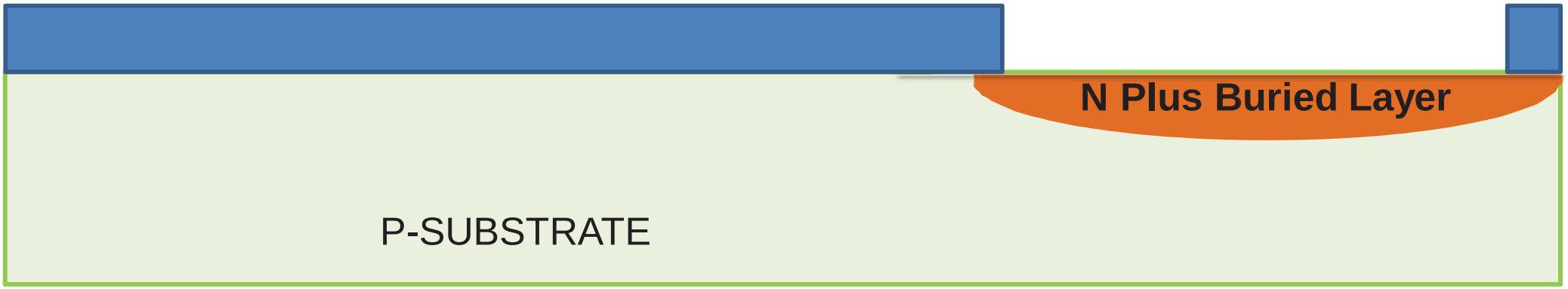
P-TYPE SUBSTRATE IS COVERED WITH OXIDE LAYER



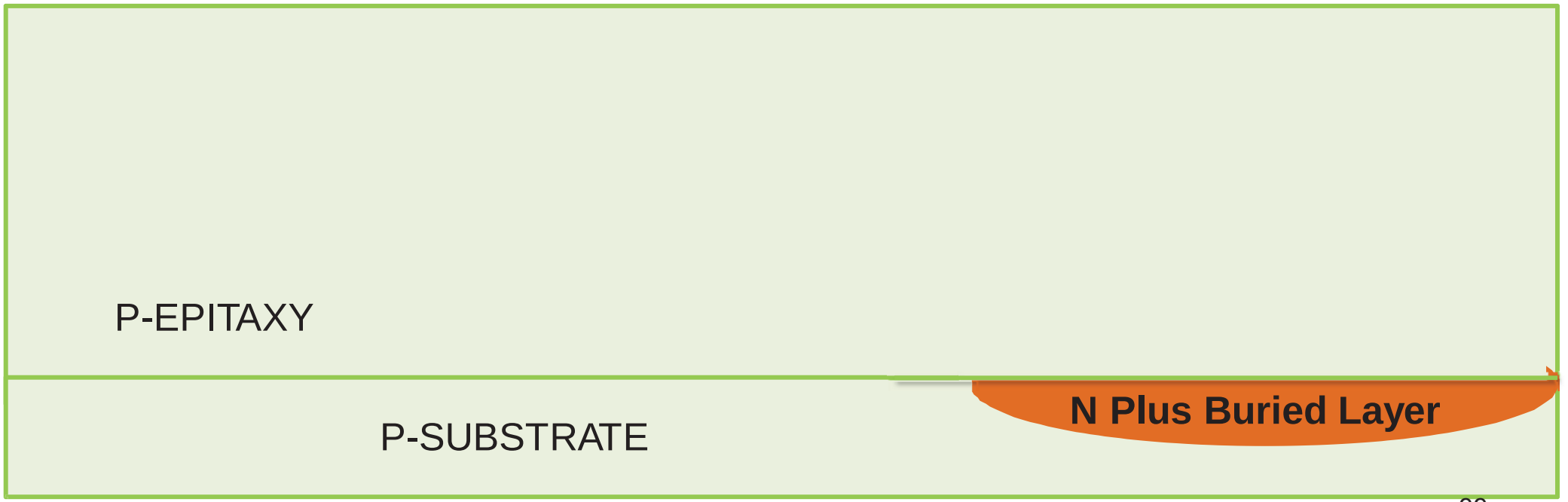
A WINDOW IS OPENED THROUGH OXIDE LAYER



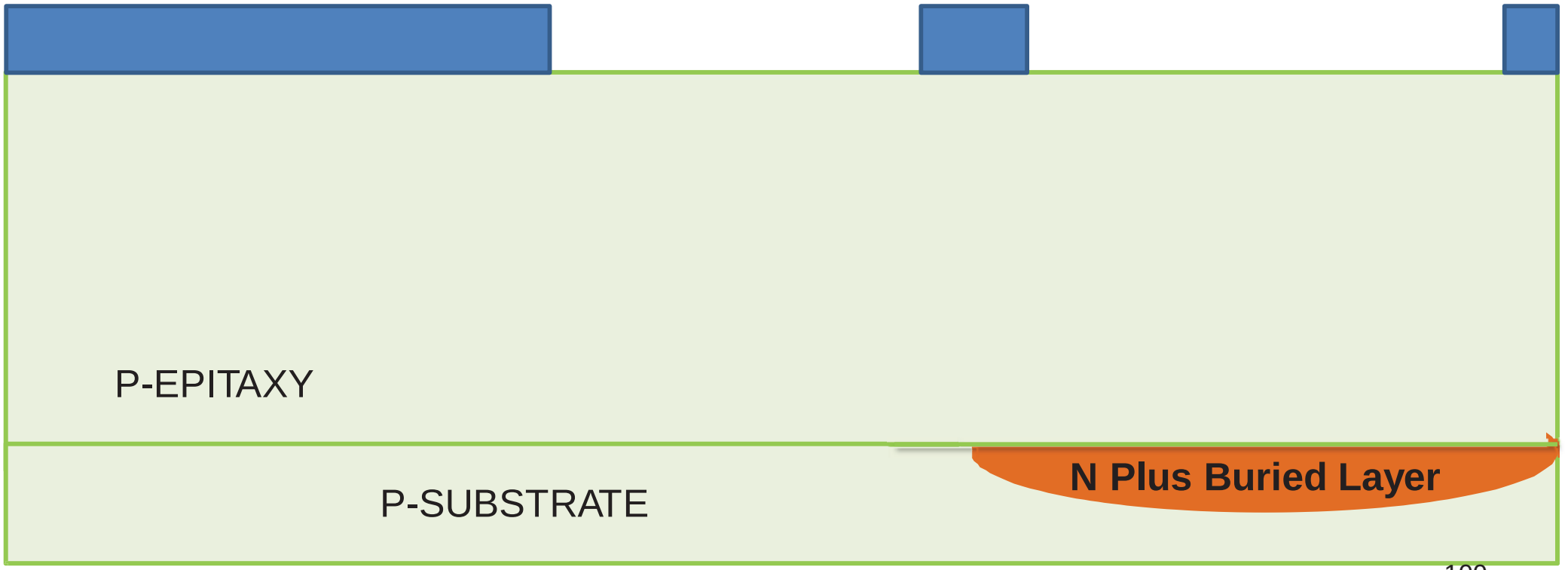
THROUGH THE WINDOW N TYPE IMPURITIES IS HEAVILY DOPED



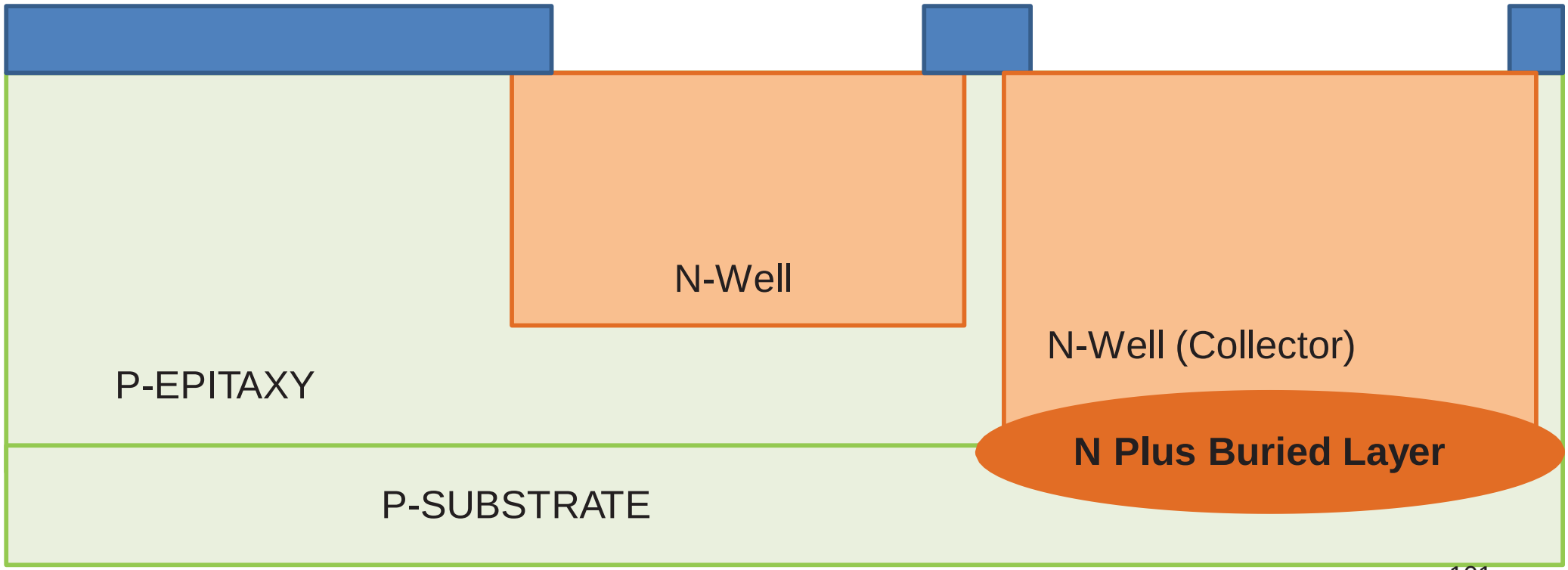
P-EPITAXY LAYER IS GROWN ON THE ENTIRE SURFACE



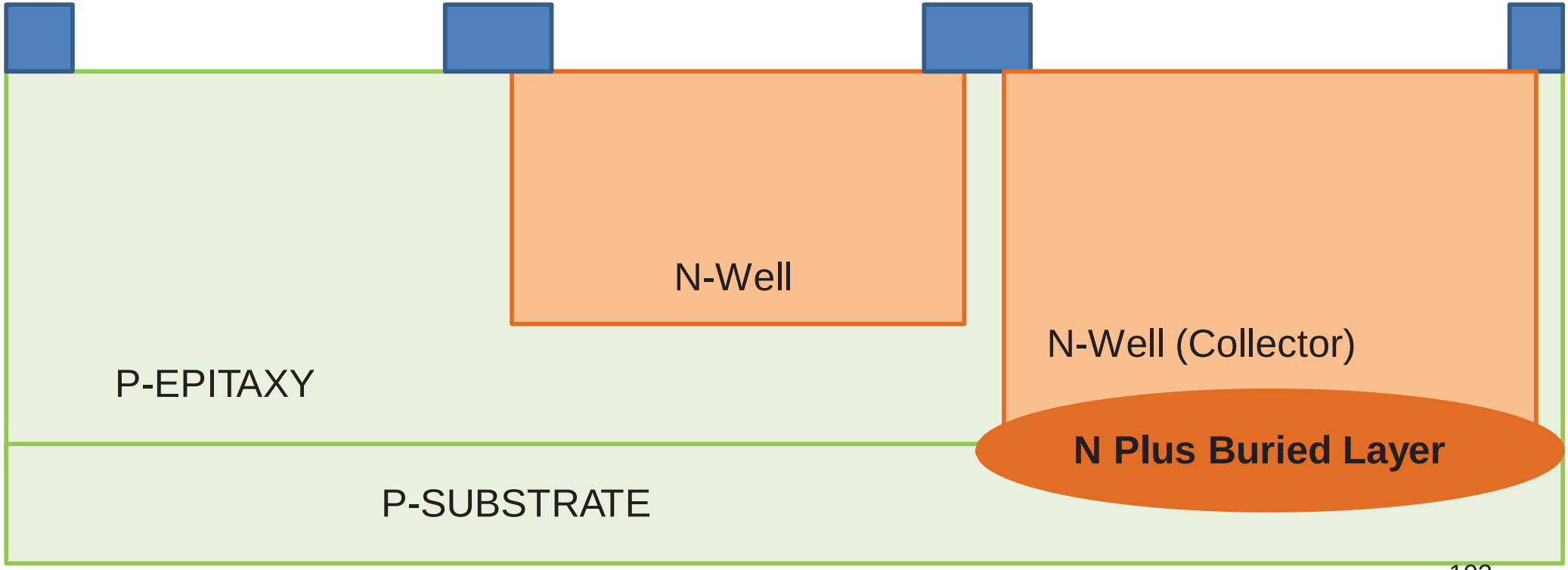
THE ENTIRE SURFACE IS COVERED WITH OXIDE LAYER AND TWO WINDOWS ARE OPENED THROUGH THE OXIDE LAYER



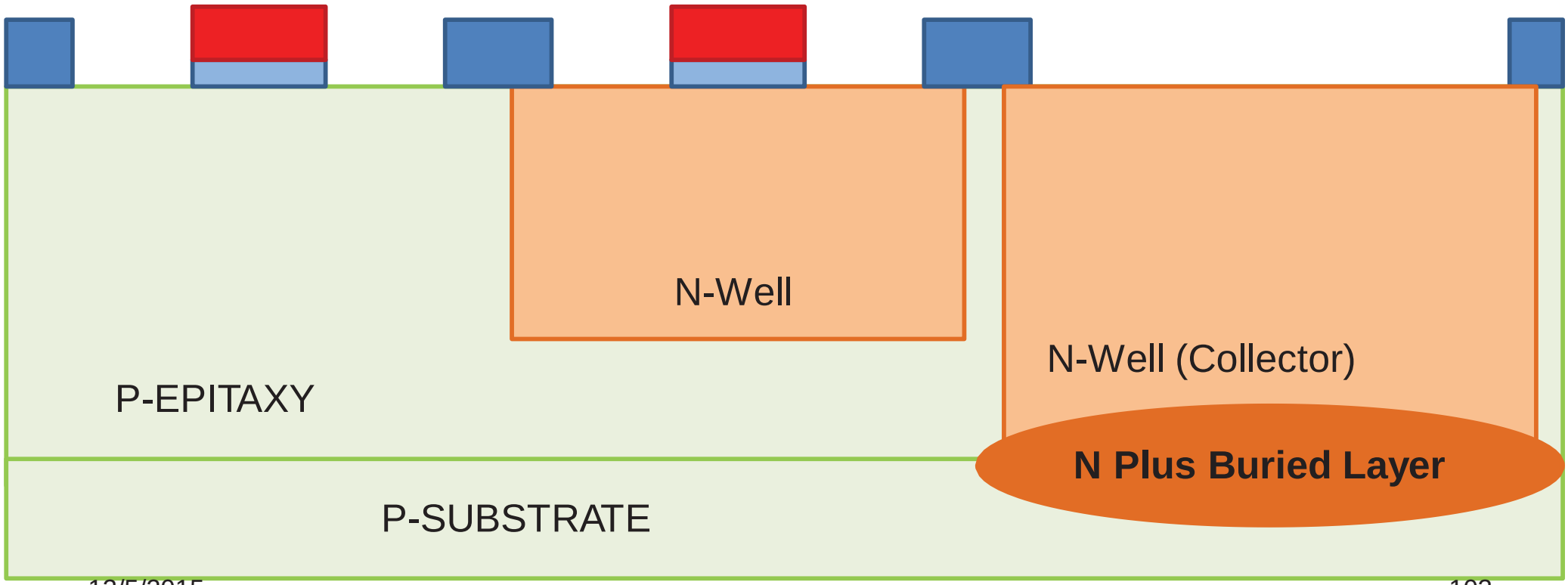
THROUGH THE TWO WINDOWS N-TYPE IMPURITIES ARE DIFFUSED TO FORM N-WELLS



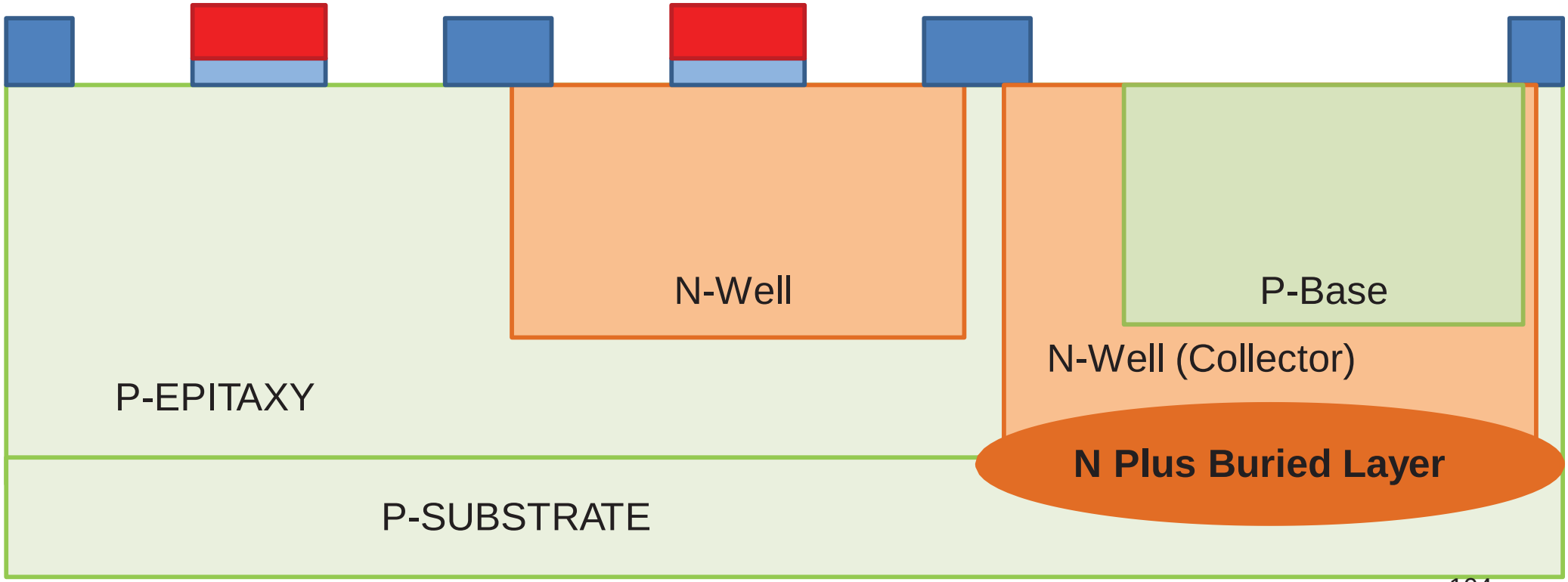
THREE WINDOWS ARE OPENED THROUGH THE OXIDE LAYER , IN THESE THREE WINDOWS THREE ACTIVE DEVICES NMOS,PMOS AND NPN BJT ARE FORMED



THE ENTIRE SURFACE IS COVERED WITH THINOX AND POLYSILICON AND ARE PATTERNED TO FORM THE GATE TERMINALS OF THE NMOS AND PMOS

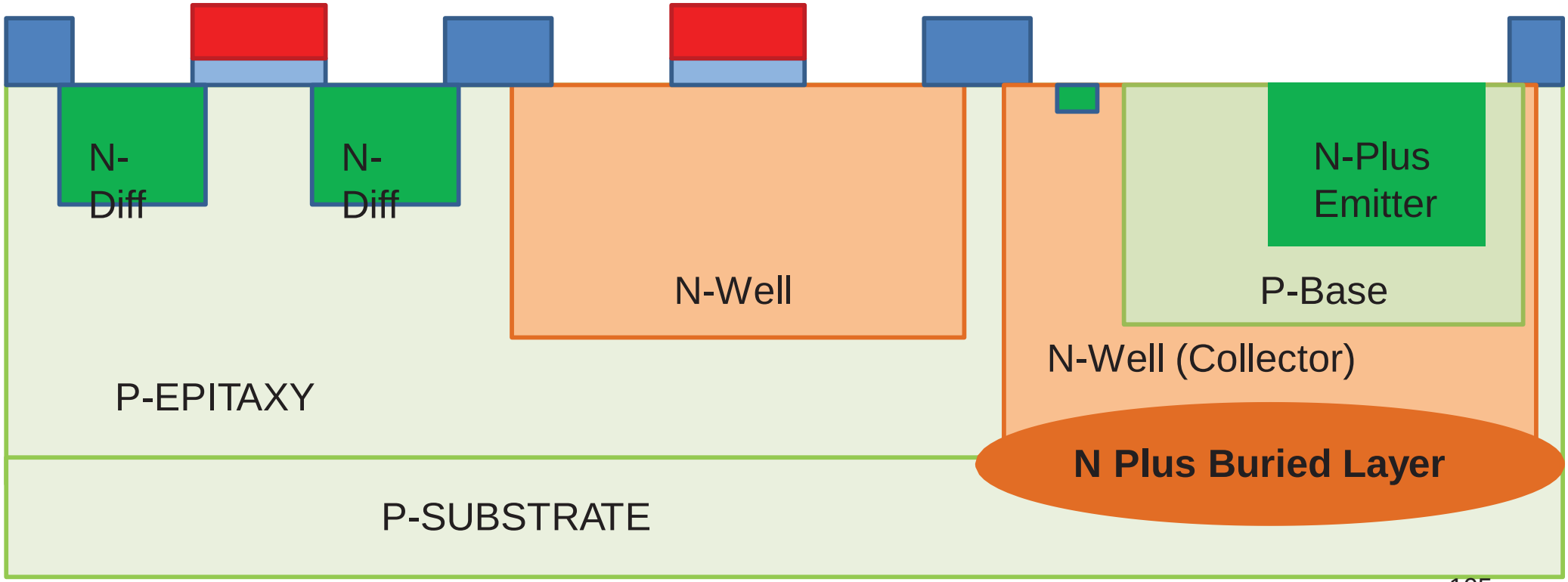


THROUGH THE 3RD WINDOW THE P-IMPURITIES ARE MODERATELY DOPED TO FORM THE BASE TERMINAL OF BJT
N-WELL ACTS LIKE THE COLLECTOR TERMINAL



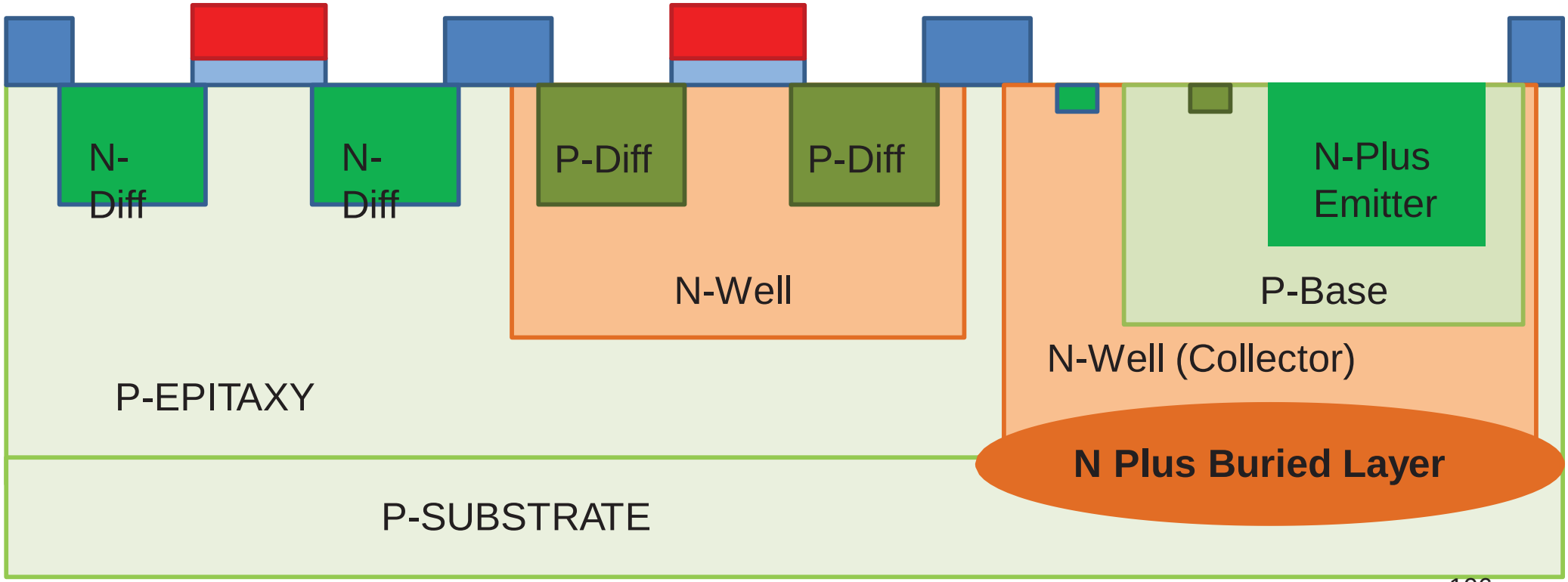
THE N-TYPE IMPURITES ARE HEAVILY DOPED TO FORM

1. SOURCE AND DRAIN REGION OF NMOS
2. EMITTER TERMINAL OF BJT
3. AND INTO NWELL COLLECTOR REGION FOR CONTACT PURPOSE

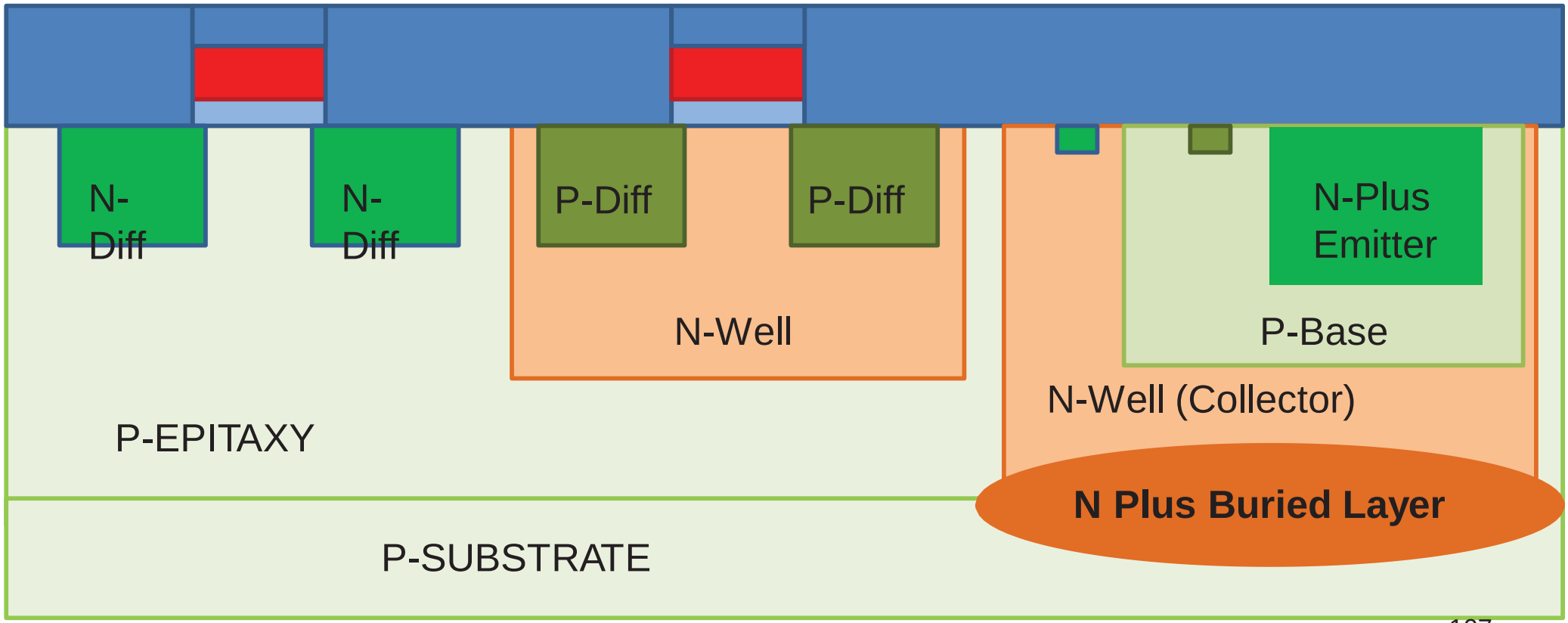


THE P-TYPE IMPURITES ARE HEAVILY DOPED TO FORM

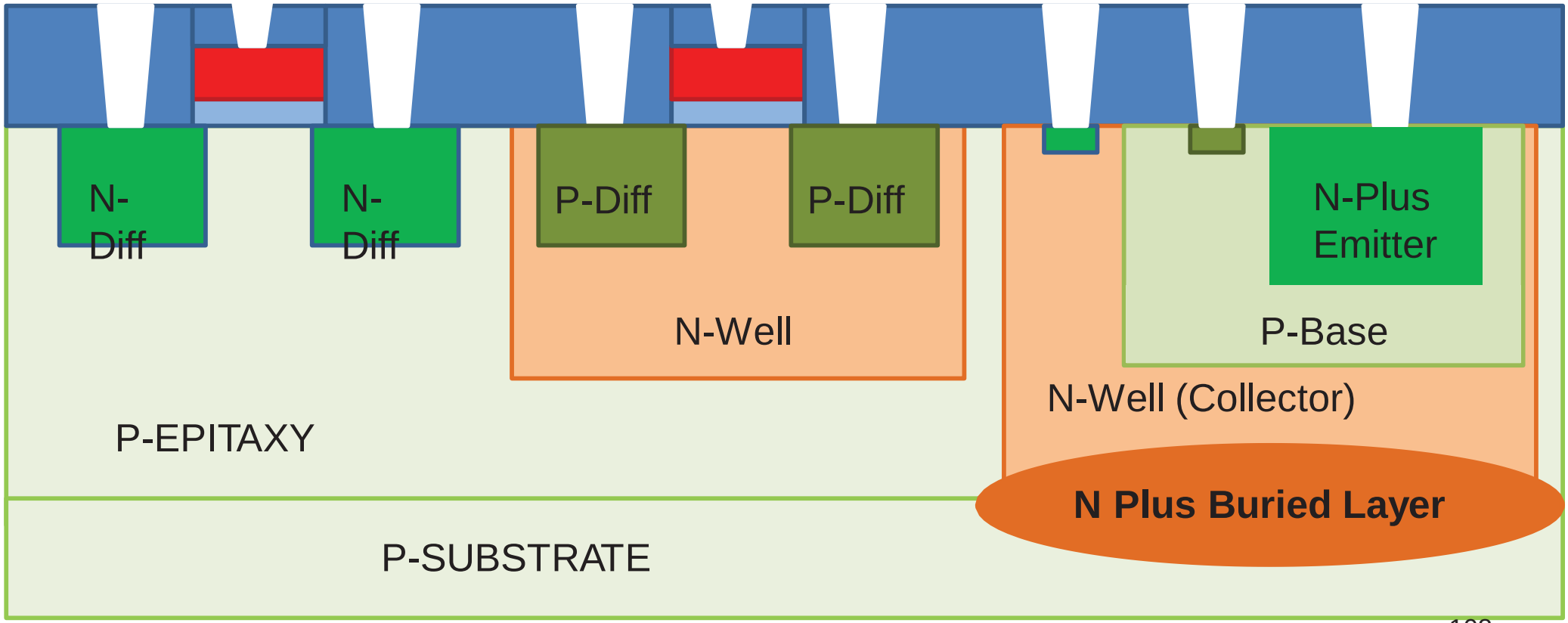
1. SOURCE AND DRAIN REGION OF PMOS
2. AND INTO P-BASE REGION FOR CONTACT PURPOSE



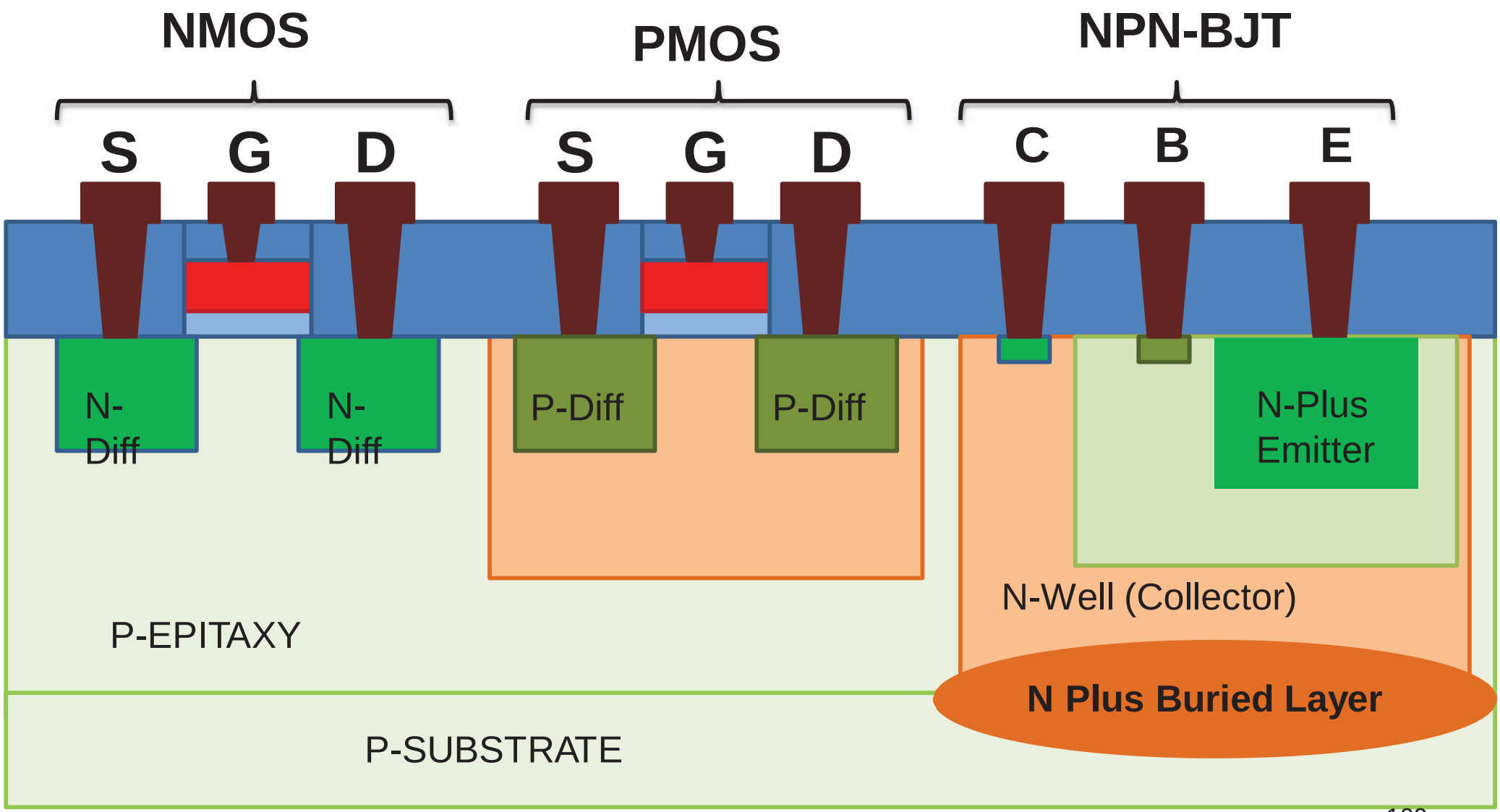
THE ENTIRE SURFACE IS COVERED WITH THICK OXIDE LAYER



THE ENTIRE SURFACE IS COVERED WITH THICK OXIDE LAYER AND IS PATTERNED FOR CONTACT CUTS



METAL CONTACTS ARE FORMED



Resistors & Capacitors fabrication

Ohm's Law

- Current I in terms of Jn
- Voltage V in terms of electric field

$$V = IR$$

$$I = JA = JtW$$

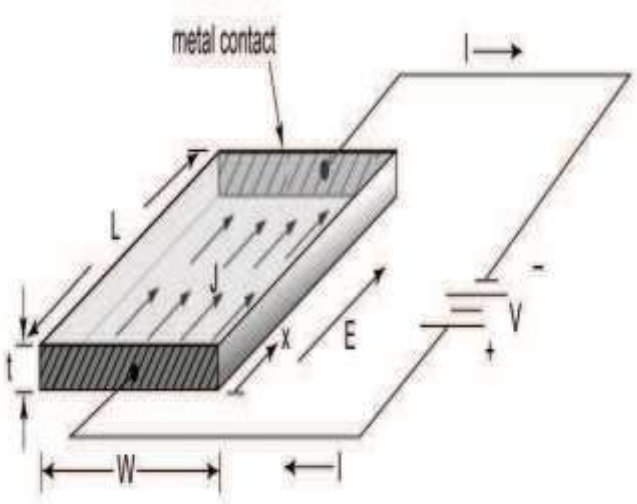
$$I = JA = JtW = \sigma tW E$$

$$E = V / L$$

$$I = JA = JtW = \frac{\sigma tW}{L} V$$

– Result for R

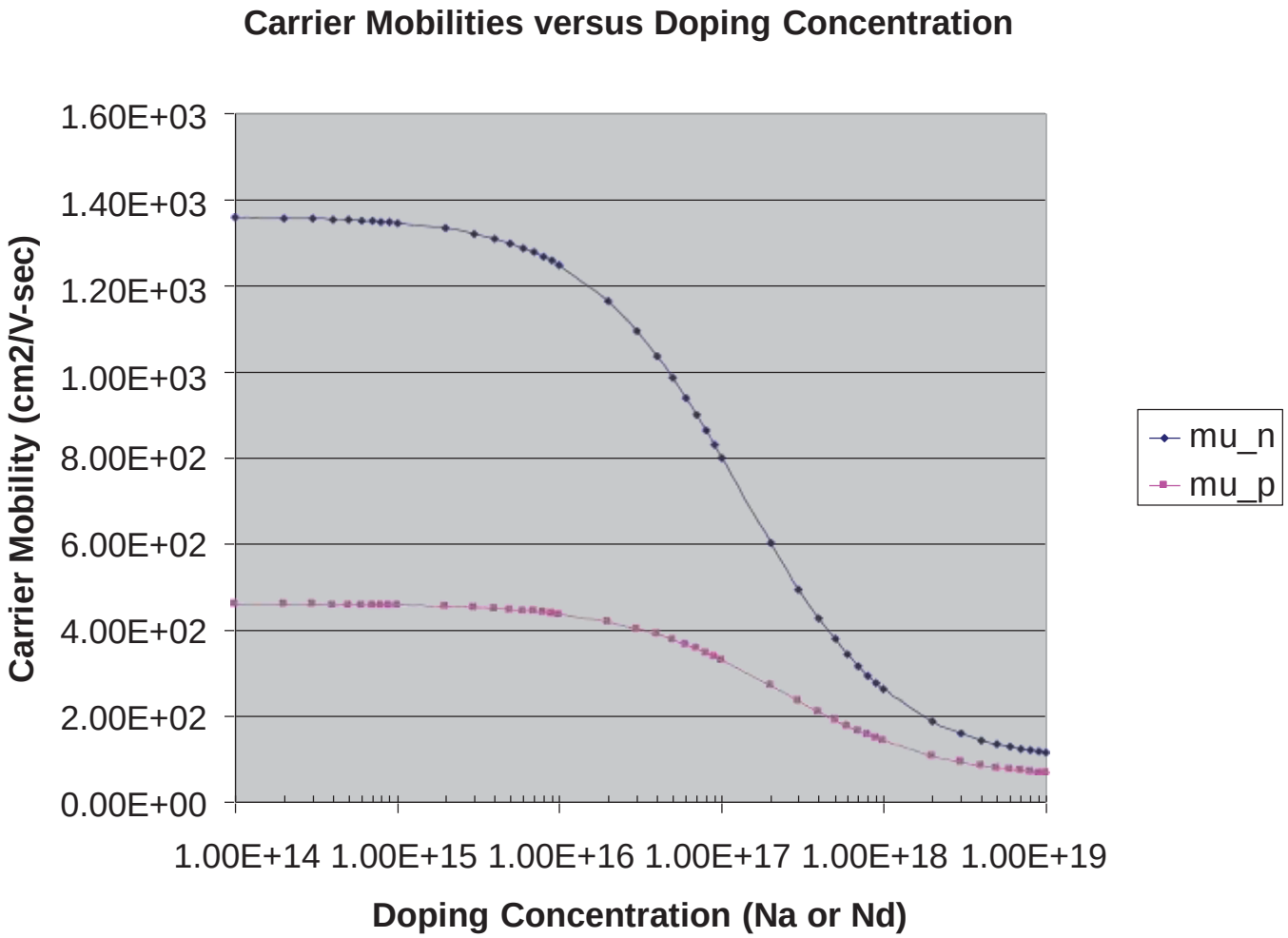
$$R = \frac{L}{W} \frac{1}{\sigma t} \qquad R = \frac{L}{W} \frac{\rho}{t}$$



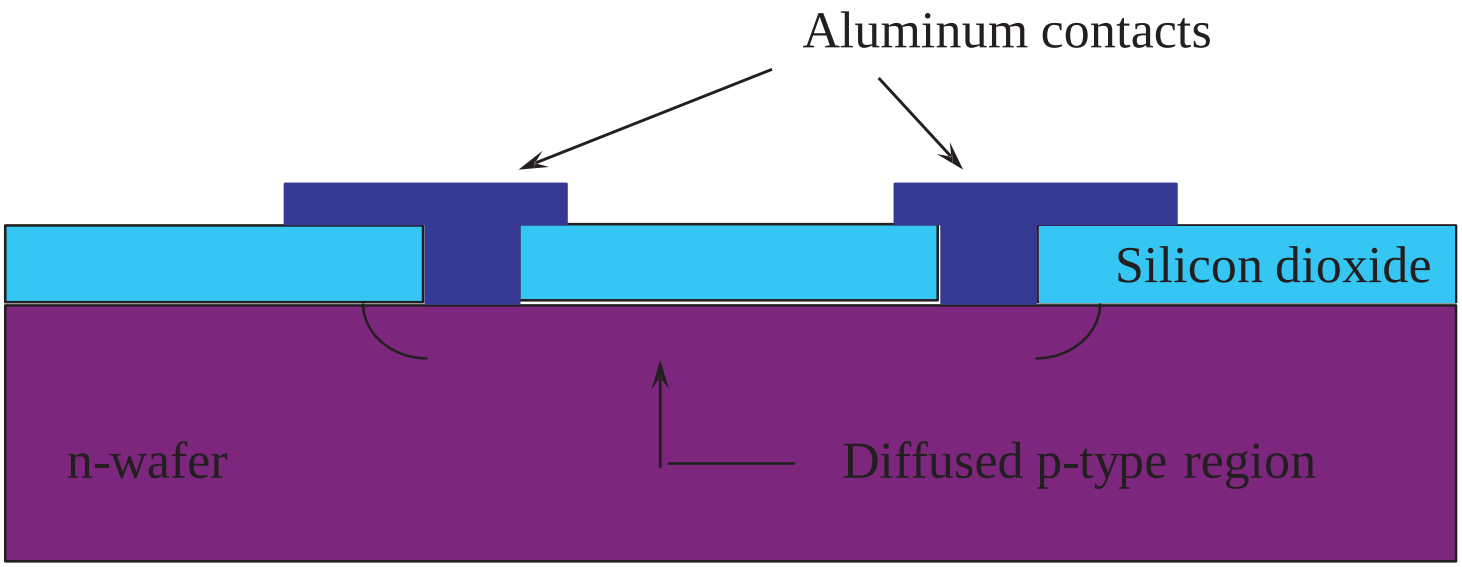
Sheet Resistance (Rs)

- IC resistors have a specified thickness – not under the control of the circuit designer
- Eliminate t by absorbing it into a new parameter: the sheet resistance (Rs)

ELECTRON AND HOLE MOBILITY



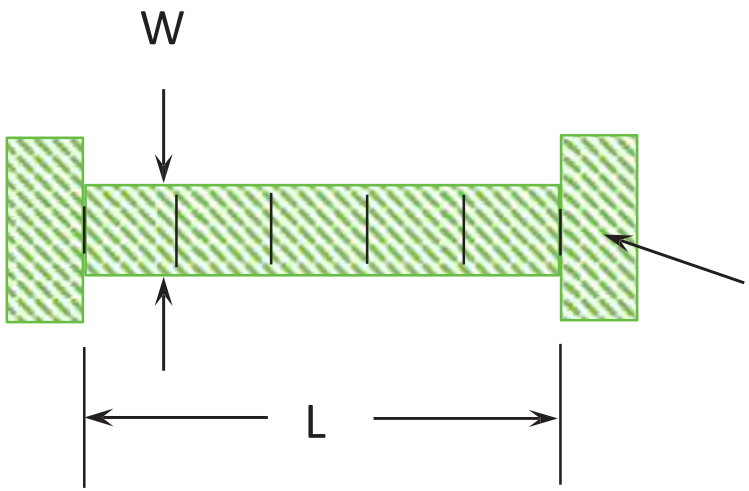
DIFFUSED RESISTOR



The n-type wafer is always biased positive with respect to the p-type diffused region. This ensures that the pn junction that is formed is in reverse bias, and there is no current leaking to the substrate. Current will flow through the diffused resistor from one contact to the other. The I-V characteristic follows Ohm's Law: $I = V/R$

Layout/Mask Layer 1 - Diffusion (green)

Top View



Resistor termination

Side View

N wafer

The resistance, R = rhos (L/W)

The sheet resistance rhos, is the resistance of each square

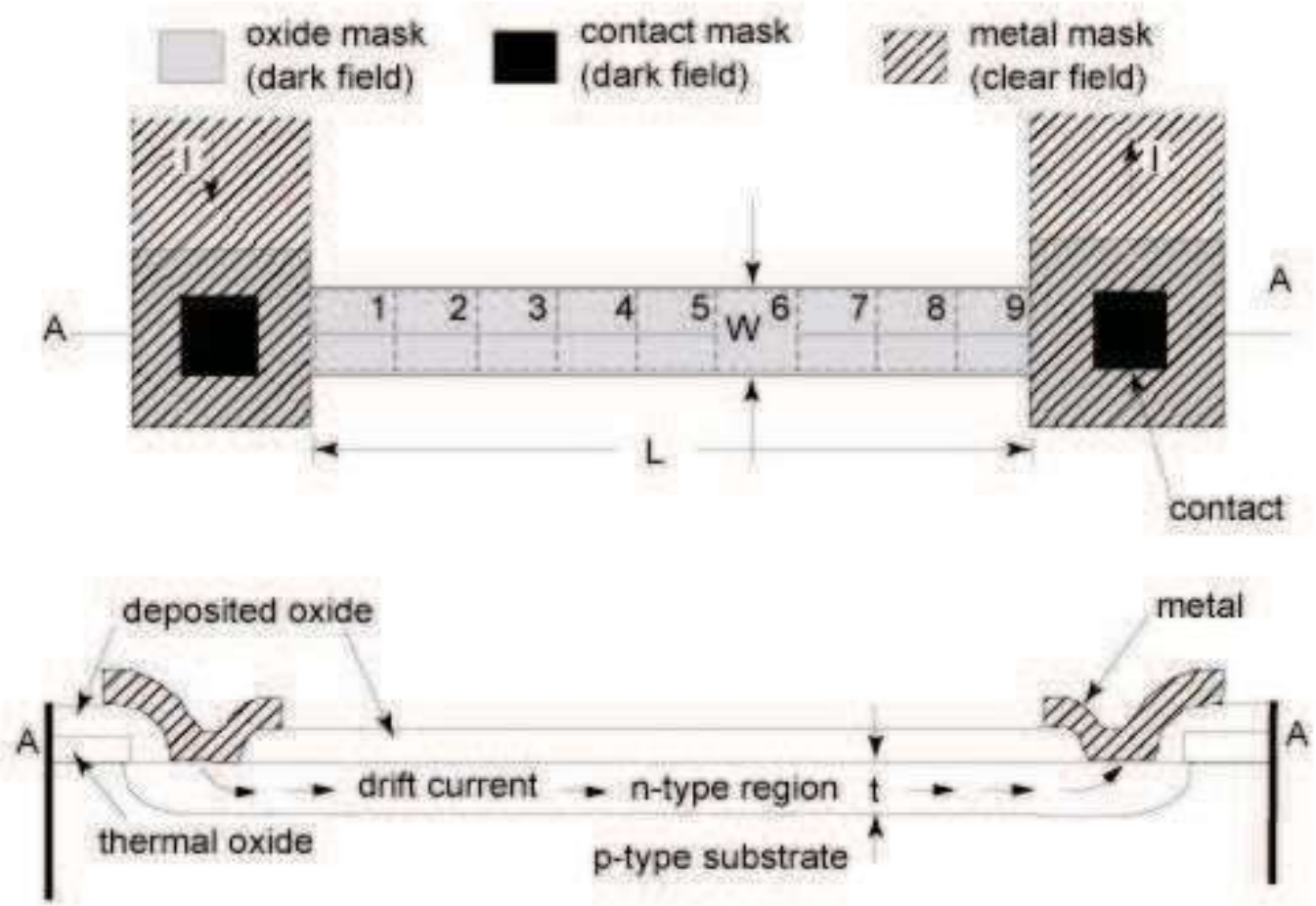
L / W is the number of 'squares' long the resistor is said to be.

5 squares in this case

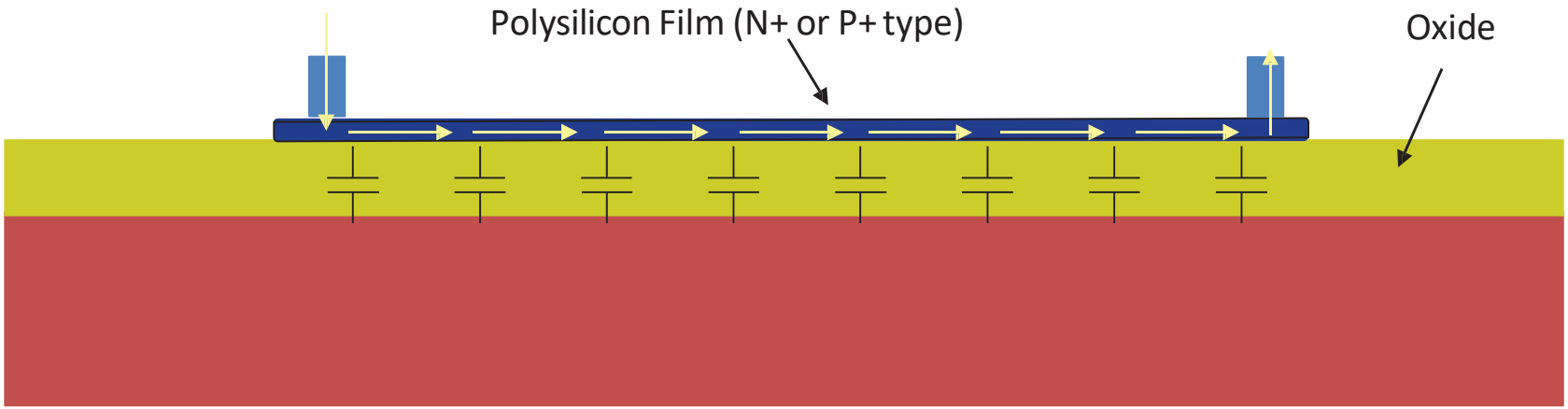
If rhos is 100 ohms per square,
R = 500 ohms

Using Sheet Resistance (Rs)

- Ion-implanted (or “diffused”) IC resistor



Poly Film Resistor



- To lower the capacitive parasitics, we should build the resistor further away from substrate
- We can deposita thin film of “poly” Si (heavily doped) material on top of the oxide
- The poly will have a certain resistance (say 10 Ohms/sq)

Diffused Resistor

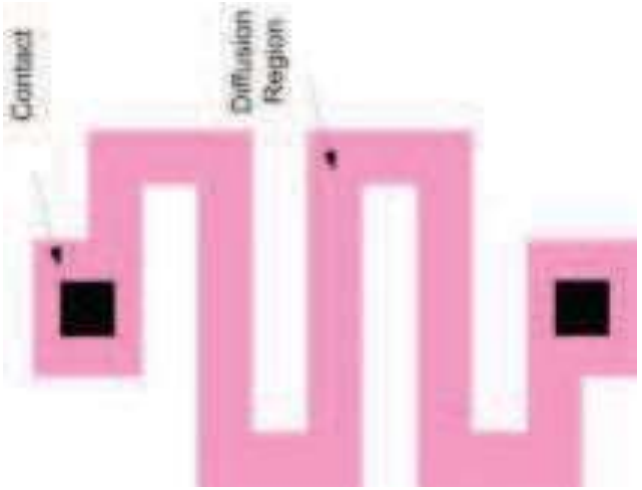
-Dope a region of the silicon (n-type or p-type) to an acceptable NA or ND.

-Then place a contact at each end of the diffusion region.

-The diffusion region will have a given resistivity specified in "Ohms / Square“

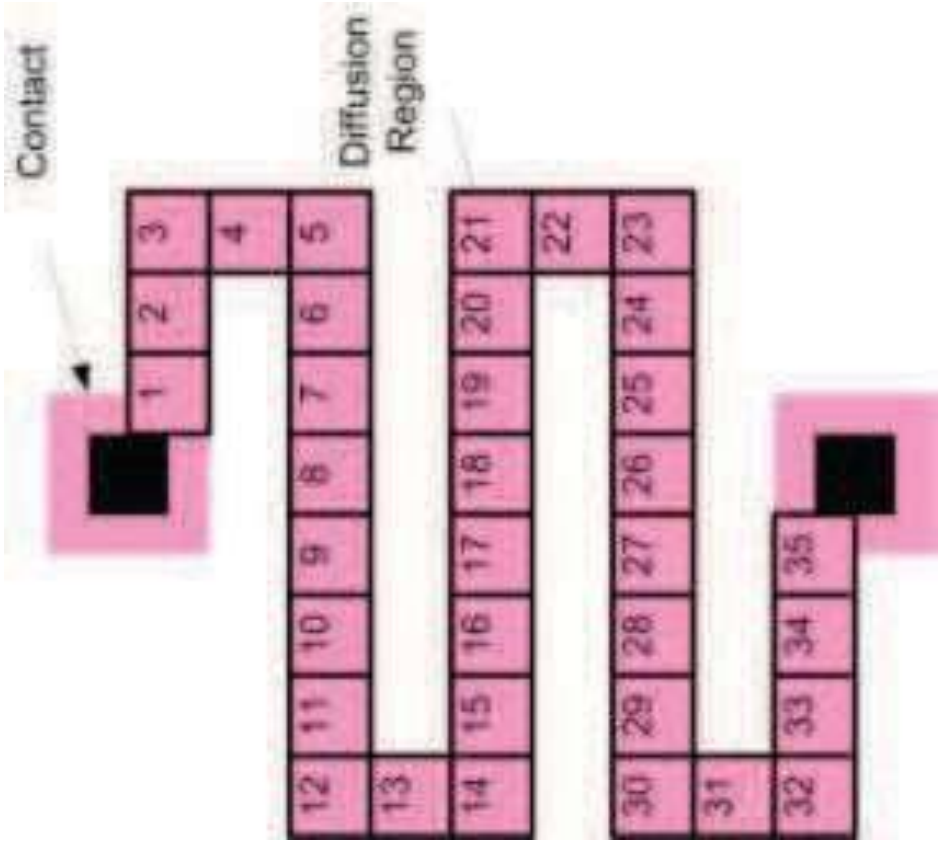
-Then alter the geometry (L/W ration) to get the desired resistance - typically these have a sheet resistance between 100 to 200 ohms/sq - to save space

-These are laid out using a serpentine geometry



-The interesting thing about the l/W ratio is that if l=W, then the shape is a square and R=Rs, this is true no matter how big the square is.
-In fact, the l/W ratio is actually the number of squares in a given trace geometry - We typically just count the squares and use:

R= Rs*(no.of.squares)



Polysilicon Resistor

-Another way of fabricating a resistor is to use Polysilicon.

-Polysilicon has a high resistivity prior to Ion Implantation

-Use undoped Polysilicon to create a high value resistor

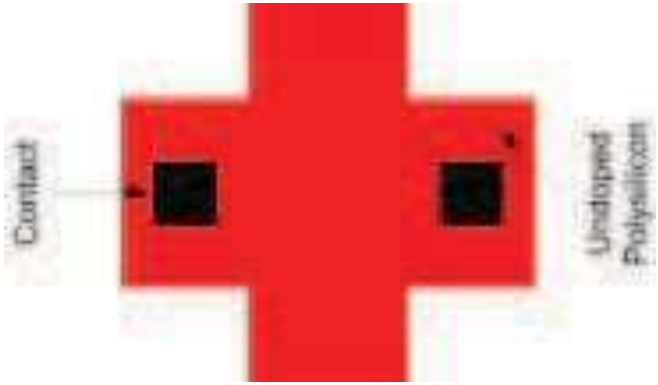
Before Ion Implantation : Rs = 10M Ohms/Square

After Ion Implantation : Rs = 20 to 40 Ohms/Square

-Typically don't even need 1 square to get our resistively so we don't need to do a serpentine layout

-One drawback is that the resistance can vary widely with process when using less than 1 square to get a resistor in the k-Ohms range.

-These are typically used when we just want a BIG resistor and don't care about the exact value

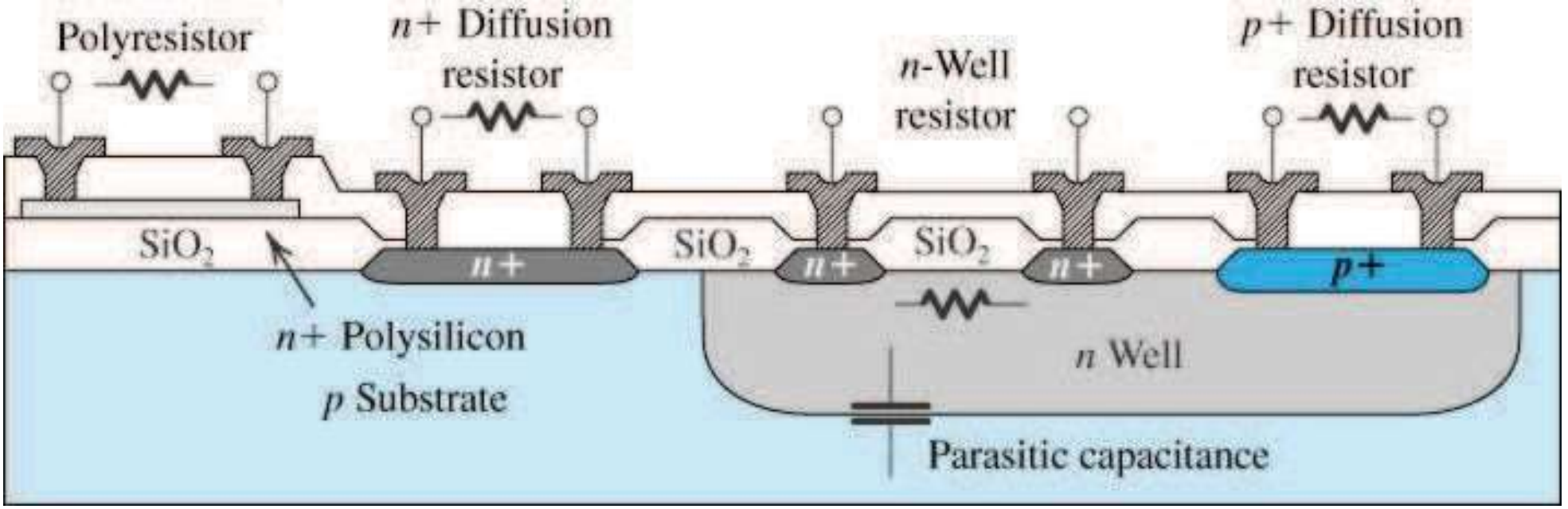


Metal Resistor

-Metal can also be used for very small resistors

-The M1 layer typically has sheet resistance on the order of mohms/sq.

- Use a serpentine layout to get a small resistor (1-10 ohms)



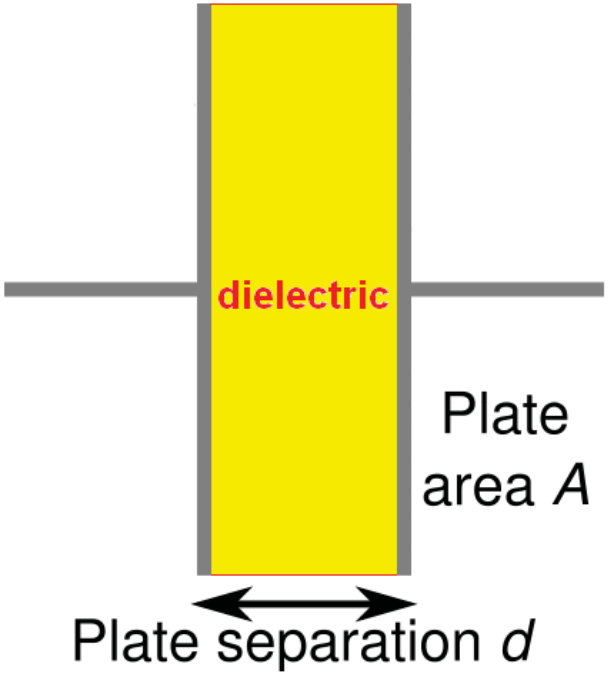
Cross sections of resistors of various types available from a typical *n*-well CMOS process.

-n-well process is used for medium value of resistors, while the n+ and p+ diffusions are useful for low value resistors.

-The resistance value depends on the length and width of the diffused regions, the tolerance of the resistor value is very poor (i.e., 20 to 50 %)

Capacitors

- Composed of two conductive plates separated by an insulator (or dielectric).
  - Commonly illustrated as two parallel metal plates separated by a distance, d.
  - C = e A/d
  - where e = er eo
  - er is the relative dielectric constant
  - eo is the vacuum permittivity



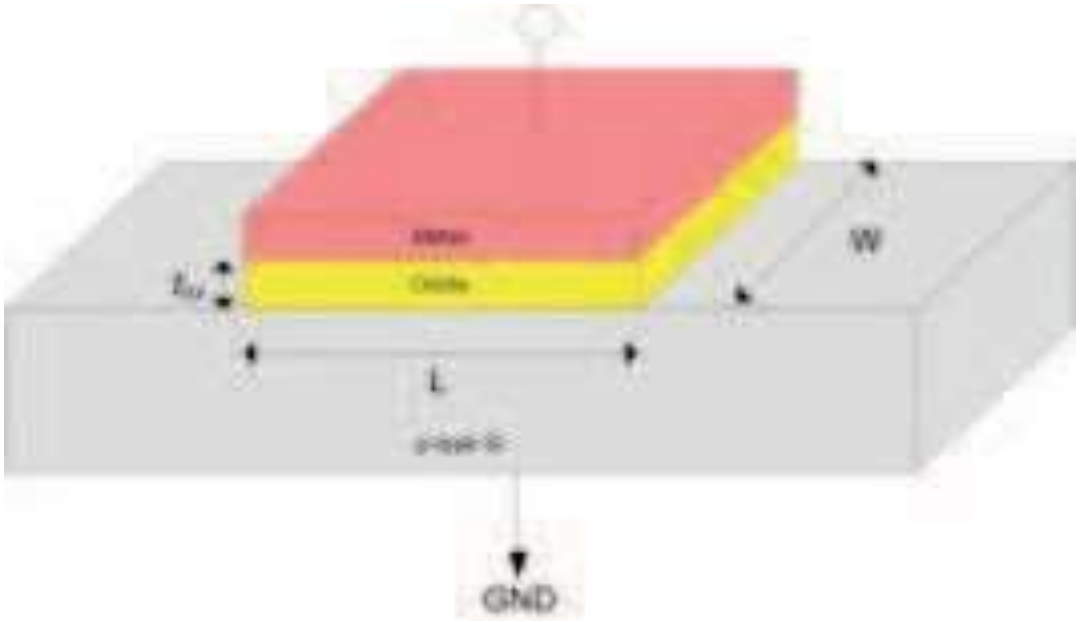
CMOS Capacitors

-There are 3 common ways to make a capacitor

1) MOS Capacitor:

-simply create a MOS structure where the Gate (Metal) terminal is one terminal and the Body (Semiconductor) terminal is Ground

- while this is easy to implement, the capacitance changes with the bias voltage (i.e., VG) due to the depletion and inversion which occurs

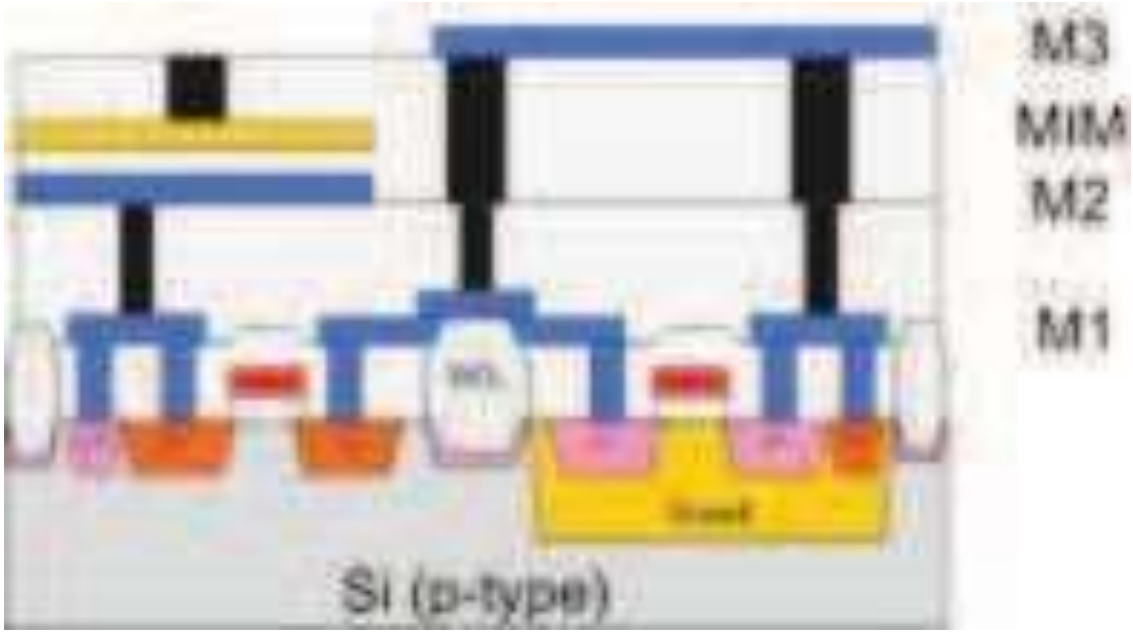


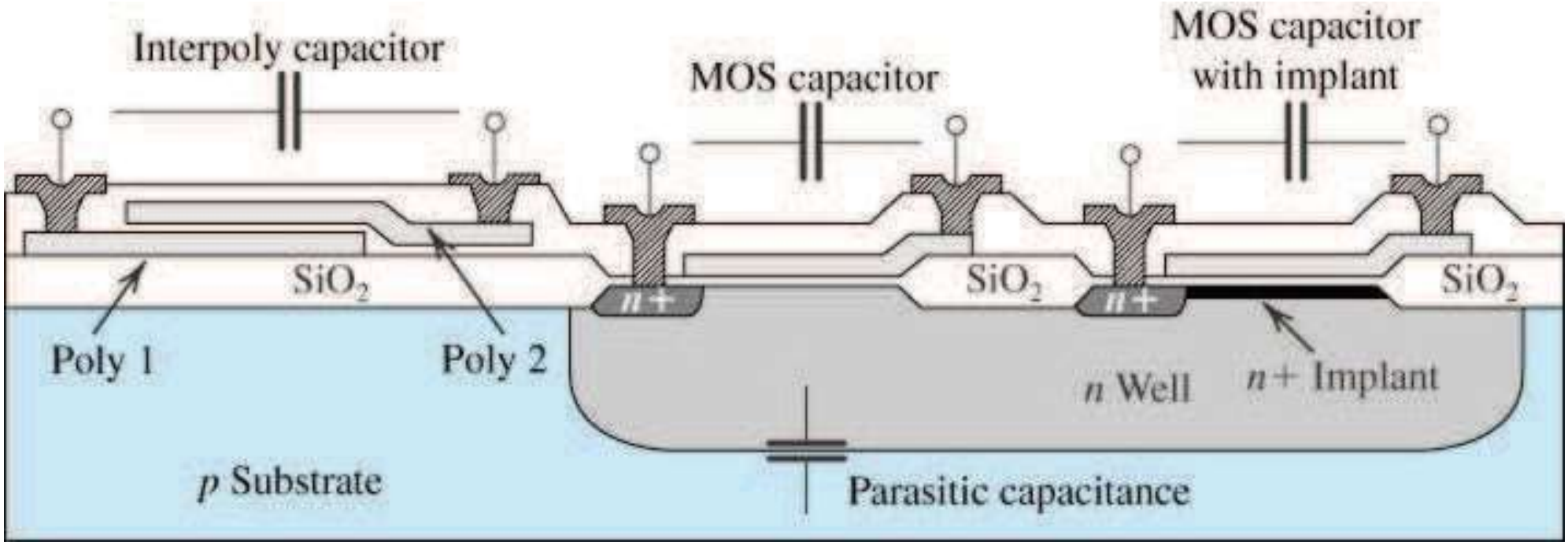
MIM Capacitor

-"Metal Insulator Metal"-this is simply a parallel plate capacitor using two metals and an insulator

-This type of capacitor is created using an extra process step that puts in an additional metal layer that can be very close to one of the other metal layers to get a smaller plate-to-plate separation

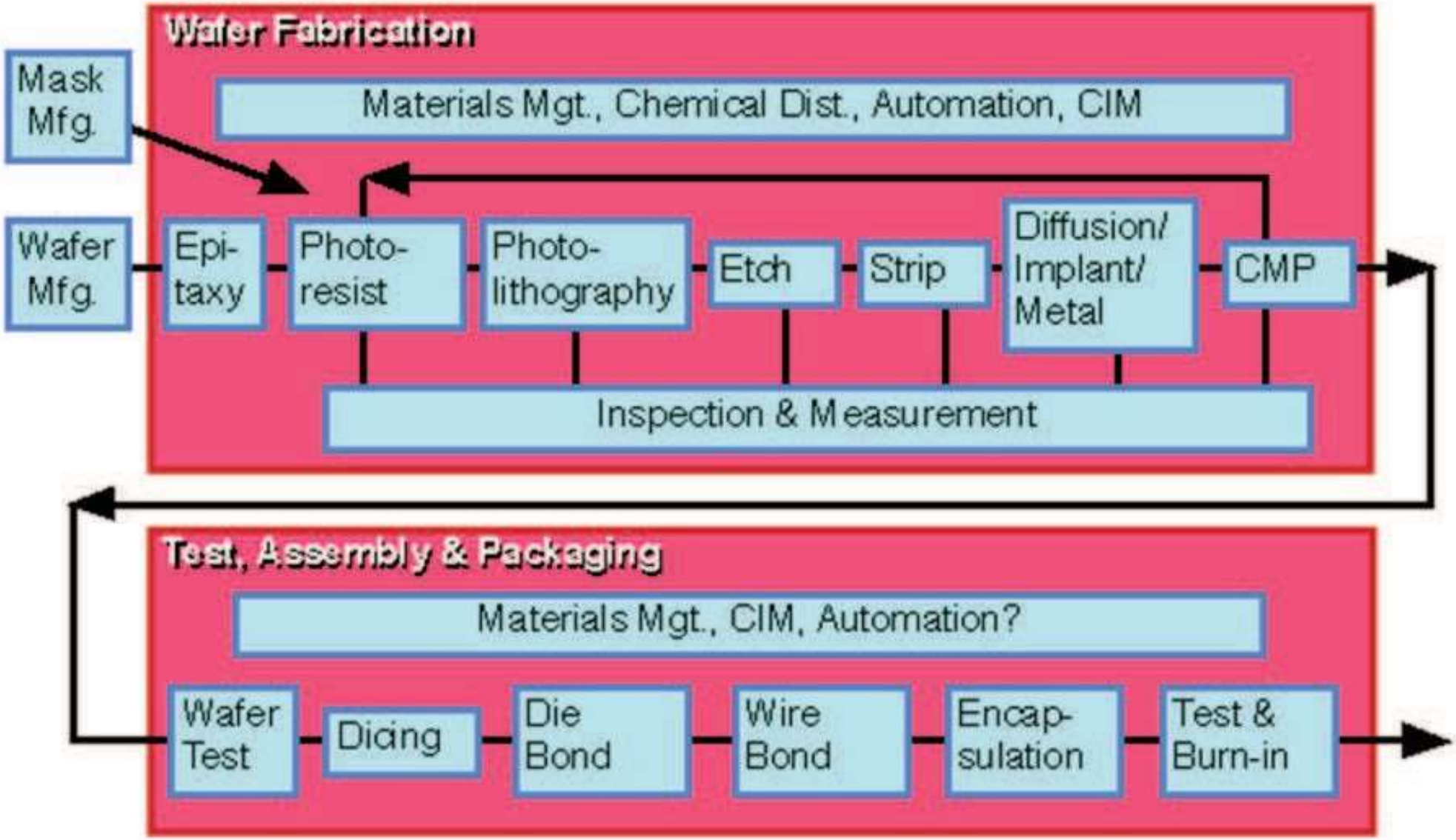
-Since the plates are made of metal, the capacitance doesn't change with bias voltage-these capacitors are not as large as MOS capacitors





Interpoly and MOS capacitors in an *n*-well CMOS process.

The Chip-Making Process



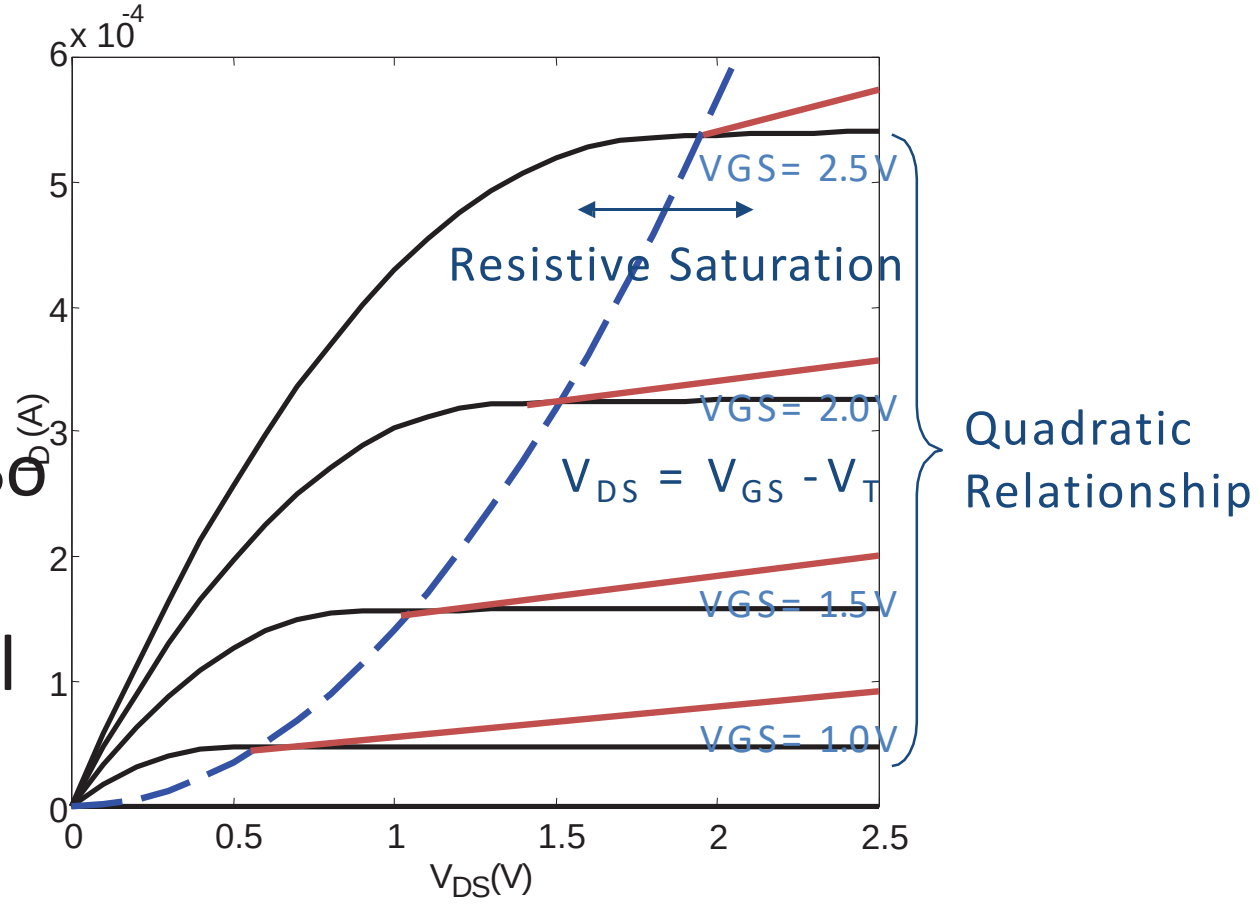
UNIT 2
BASIC ELECTRICAL PROPERTIES

Topics

- Basic electrical properties of MOS and BiCMOS circuits:
- I_{ds}-V_{ds} relationships
- MOS transistor threshold voltage, g_m, g_{ds}
- figure of merit w_o
- pass transistor
- NMOS inverter
- Various pull-ups
- CMOS inverter analysis and design
- BiCMOS inverters

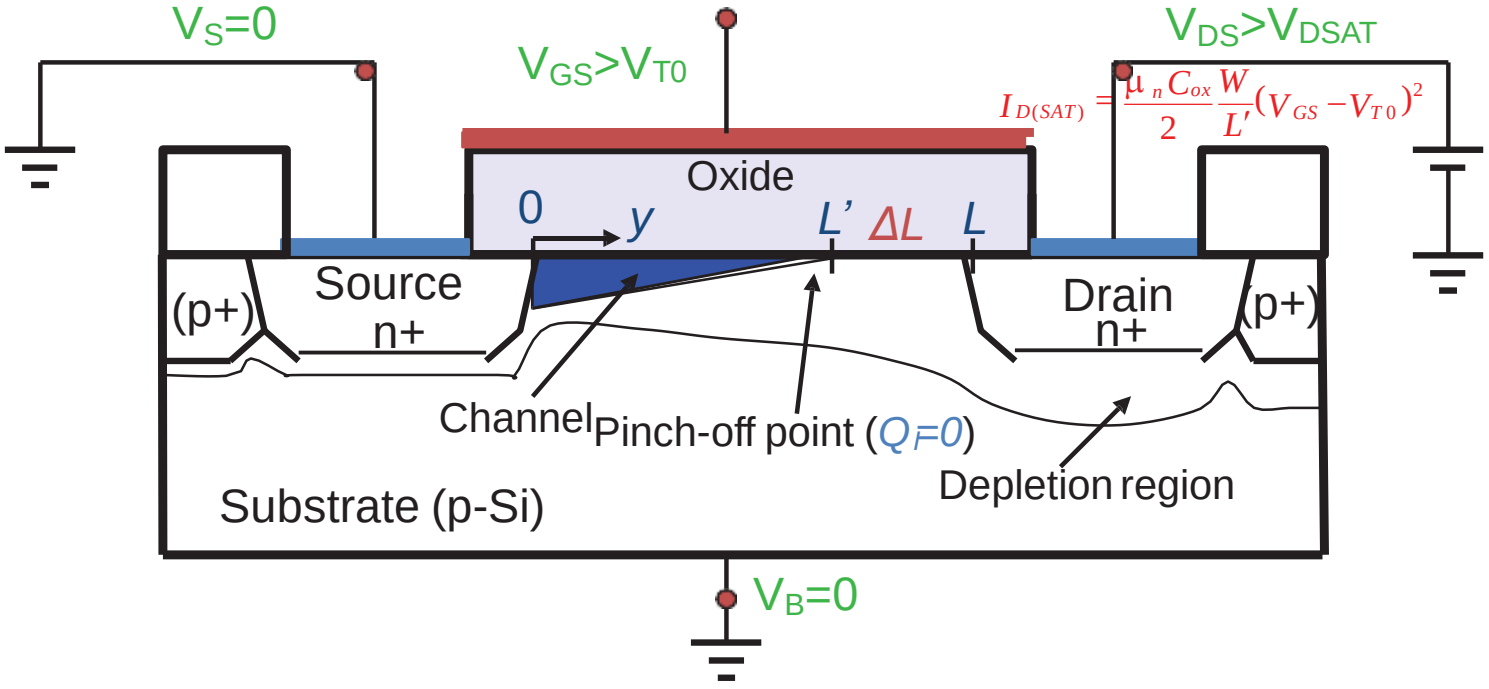
MOSFET I-V Characteristics
I-V Plots, Channel Length Modulation

- Saturation equation yields curves independent of VDS. Not sure! So we consider the effect of channel length modulation.



MOSFET I-V Characteristics Channel Length Modulation

- Channel Length Modulation
  - With pinch-off the channel at the point y such that Vc(y)=VGS - VT0, The effective channel length is equal to L' = L – ΔL
  - ΔL is the length of channel segment over which QI=0.
  - Place L' in the ID(SAT) equation:



MOSFET I-V Characteristics Channel Length Modulation

- ΔL increases with an increase in VDS. We can use

$$\frac{1}{L'} = \frac{1}{L - \Delta L} = \frac{1}{L}\frac{1}{\frac{L-\Delta L}{L}} = \frac{1}{L}\frac{1}{1-\frac{\Delta L}{L}} = \frac{1}{L}\frac{1}{1-ZV_{DS}} = \frac{1}{L}(1+ZV_{DS})$$

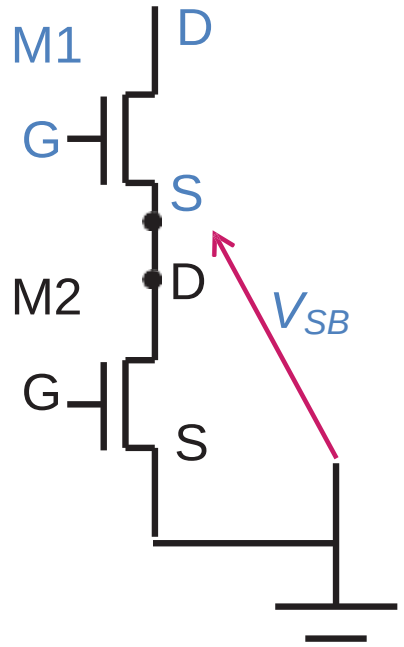
- λ: channel length modulation coefficient
- ID(SAT) can be rewritten as

- The above form produces a discontinuity of current at VDS=VGS-VT0. We can include the term in ID(lin) with little error since λ is typically less than 0.1. We will usually ignore λ in manual calculations.

$$I_{D(SAT)} = \frac{\mu_n C_{ox}}{2}\frac{W}{L}(V_{GS} - V_{T0})^2(1+ZV_{DS})$$

MOSFET I-V Characteristics
Substrate Bias Effect

- So far, VSB=0 and thus VT0 used in the equations.
- Clearly not always true – must consider body effect
- Two MOSFETs in series:



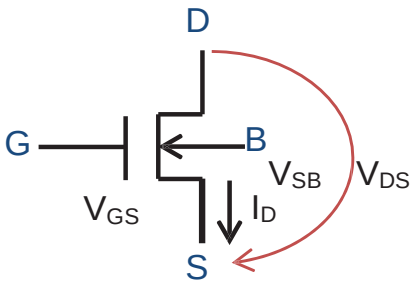
-
-
- VSB(M1) = VDS(M2) ≠ 0. Thus, VT0 in the M1 equation is replaced by VT = VT(VSB) as developed in the threshold voltage section.

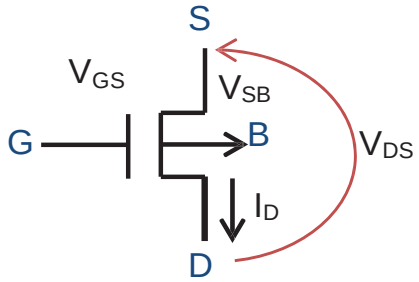
MOSFET I-V Characteristics Substrate Bias Effect (Cont.)

- The general form of ID can be written as
- ID = f (VGS,VDS,VSB)
- which due to the body effect term is non-linear and more difficult to handle in manual calculations

MOSFET I-V Characteristics
Summary of Analytical Equations

– The voltage directions and relationships for the three modes of pMOS are in contrast to those of nMOS.

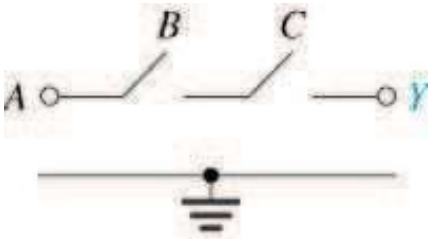


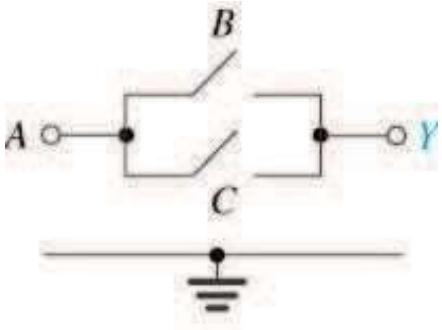


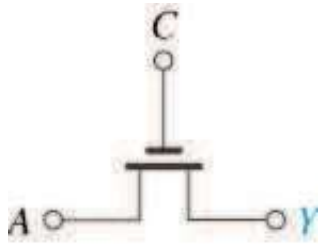
nMOS		
Mode	I_D	**Voltage Range**
Cut-off	0	$V_{GS}<V_T$
Linear	$(\mu_n C_{ox}/2)(W/L)[2(V_{GS}-V_T)V_{DS}-V_{DS}^2]$	$V_{GS}-V_T>V_{DS}<V_{GS}-V_T$
Saturation	$(\mu_n C_{ox}/2)(W/L)(V_{GS}-V_T)^2(1+\lambda V_{DS})$	$V_{GS}-V_T>V_{DS}>V_{GS}-V_T$
pMOS		
Cut-off	0	$V_{GS}>V_T$
Linear	$(\mu_n C_{ox}/2)(W/L)[2(V_{GS}-V_T)V_{DS}-V_{DS}^2]$	$V_{GS}-V_T<V_{DS}>V_{GS}-V_T$
Saturation	$(\mu_n C_{ox}/2)(W/L)(V_{GS}-V_T)^2(1+\lambda V_{DS})$	$V_{GS}-V_T<V_{DS}<V_{GS}-V_T$

Pass-Transistor Logic Circuits (1)

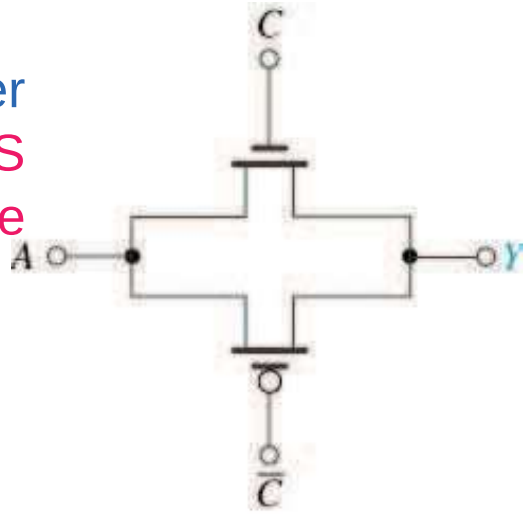
- A simple approach for implementing logic functions utilizes series and parallel combinations of switches that are controlled by input variables to connect the input and output nodes.





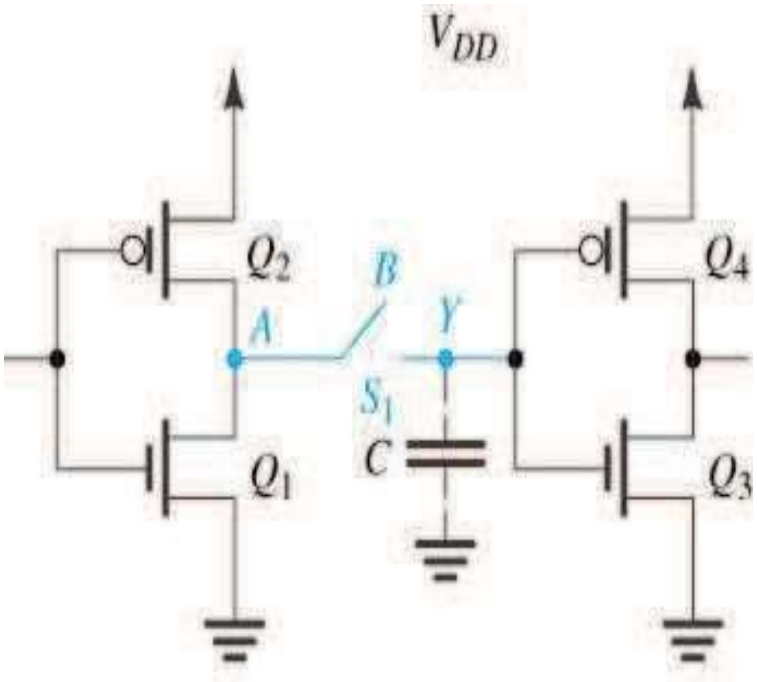


- Each of the switches can be implemented either by a single NMOS transistor or by a pair of CMOS transistors connected in CMOS transmission gate configura t ion.



Pass-Transistor Logic Circuits (2)

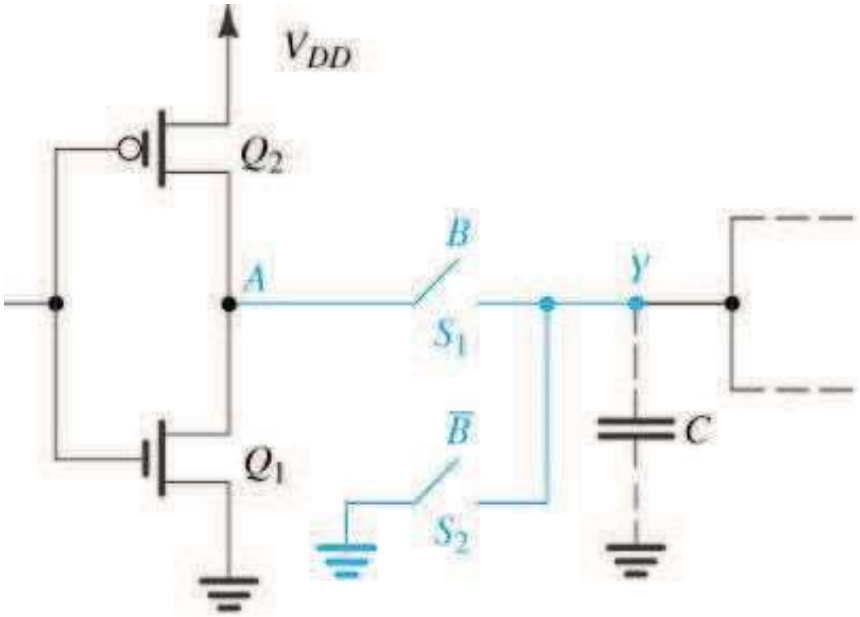
- An essential requirement in the design of pass-transistor logic is ensuring that every circuit node has at all times a low-resistance path to V_{DD} or to ground.



A basic design requirement of PTL circuits is that every node have, at all times, a low resistance path to either ground or V_{DD}. Such a path does not exist in (a) when B is low and S_1 is open. It is provided in
(b) through switch S_2.

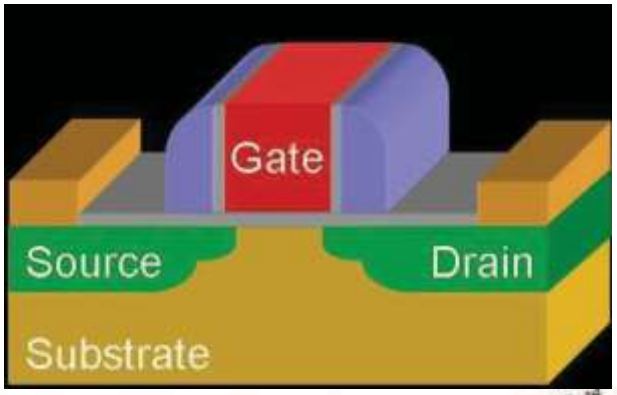
Pass-Transistor Logic Circuits (3)

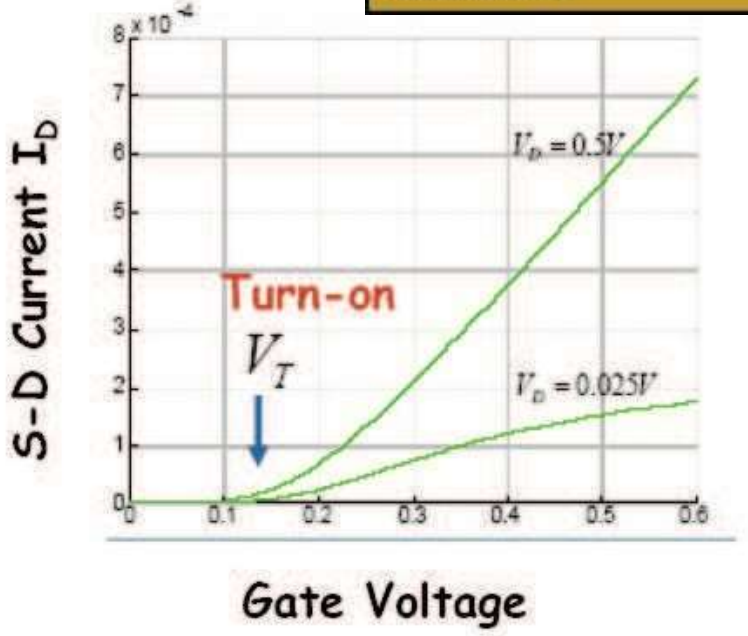
- The problem can be easily solved by establishing for node Y a low-resistance path that is activated when B goes low.

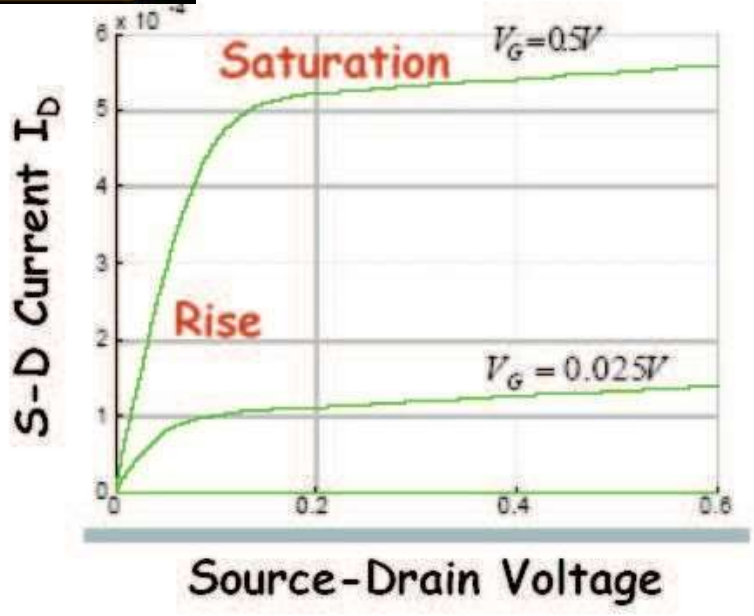


A basic design requirement of PTL circuits is that every node have, at all times, a low resistance path to either ground or V_{DD}. Such a path does not exist in (a) when B is low and S_1 is open. It is provided in (b) through switch S_2.

MOSFET Ids-Vds

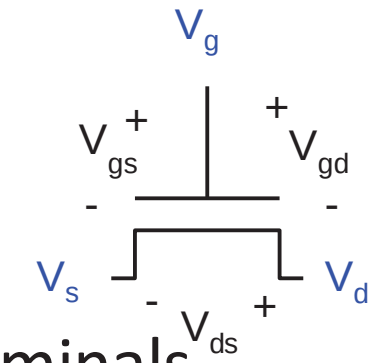






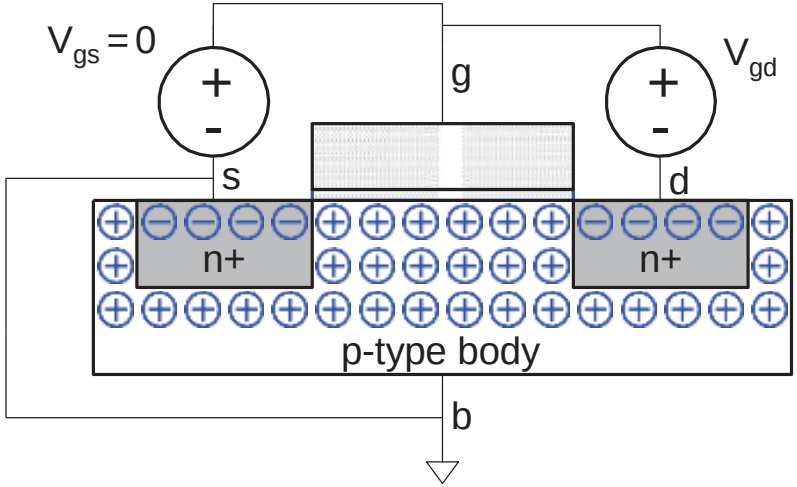
Terminal Voltages

- Mode of operation depends on Vg, Vd, Vs
  - Vgs = Vg – Vs
  - Vgd = Vg – Vd
  - Vds = Vd – Vs = Vgs - Vgd
- Source and drain are symmetric diffusion terminals
  - By convention, source is terminal at lower voltage
  - Hence Vds □ 0
- nMOS body is grounded.
- Three regions of operation
  - Cutoff
  - Linear
  - Saturation



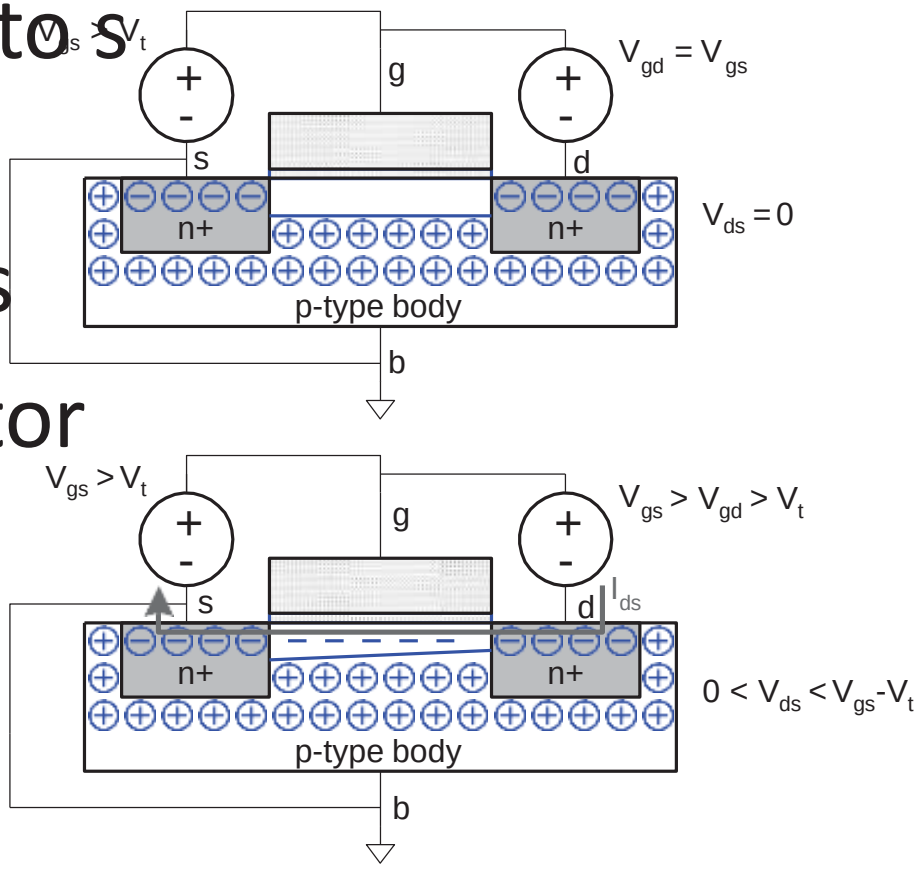
nMOS Cutoff

- No channel
- Ids = 0



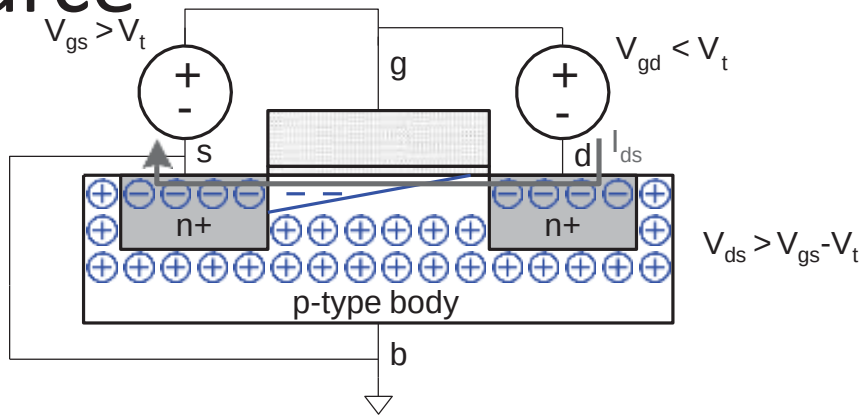
nMOS Linear

- Channel forms
- Current flows from d to s
  - e- from s to d
- Ids increases with Vds
- Similar to linear resistor



nMOS Saturation

- Channel pinches off
- Ids independent of Vds
- We say current saturates
- Similar to current source

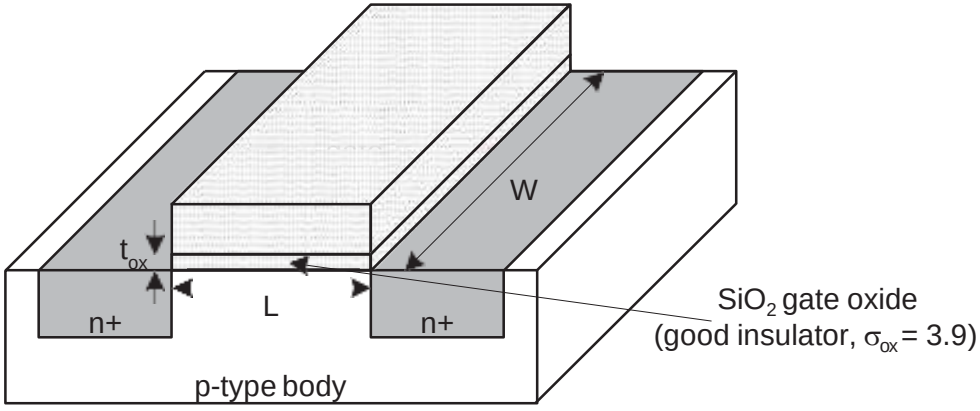


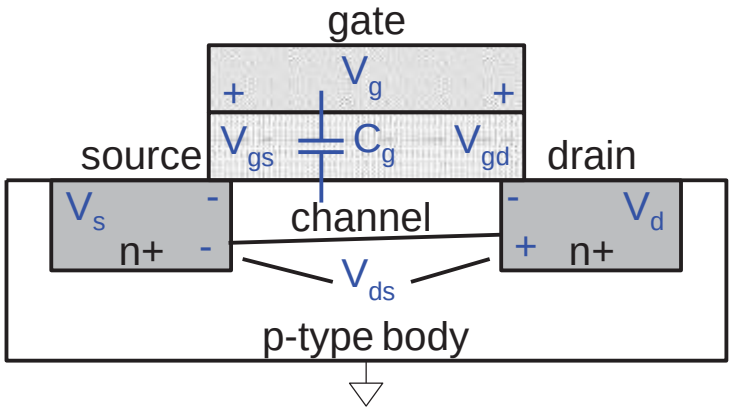
I-V Characteristics

- In Linear region, Ids depends on
  - How much charge is in the channel?
  - How fast is the charge moving?

Channel Charge

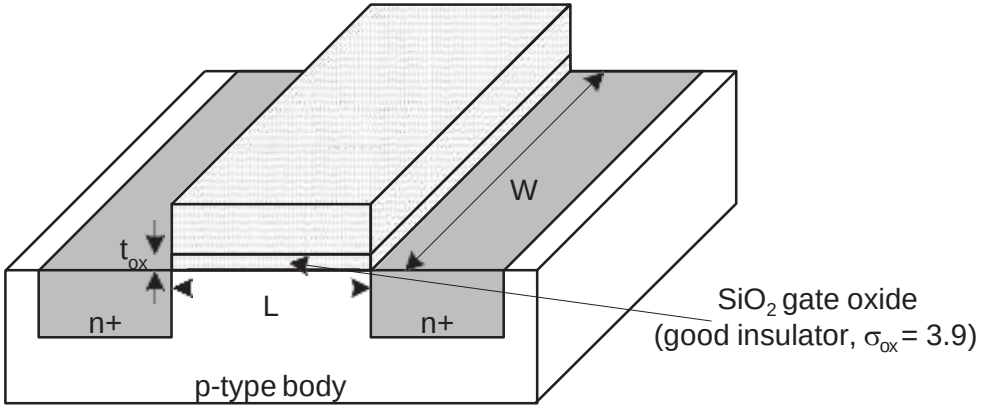
- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- Qchannel =

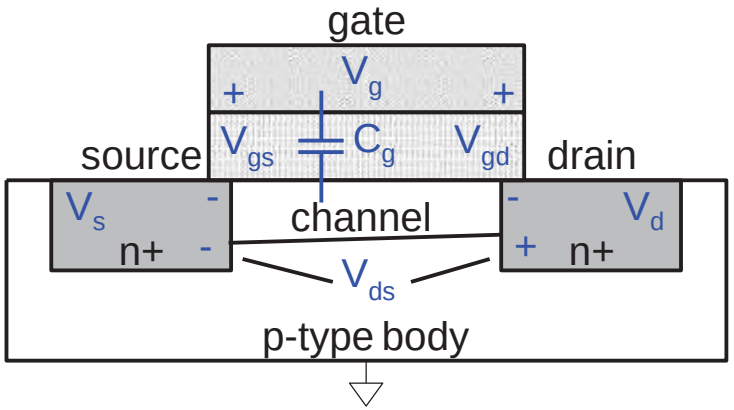




Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- Qchannel = CV
- C =

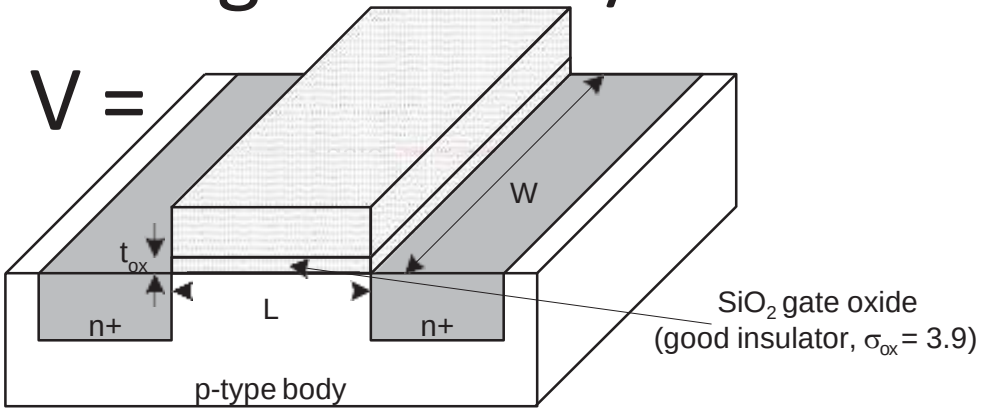


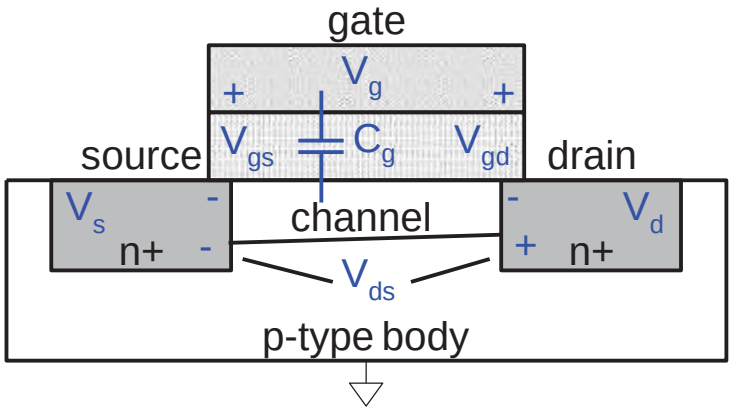


Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- Qchannel = CV
- C = Cg = eoxWL/tox = CoxWL
- V =

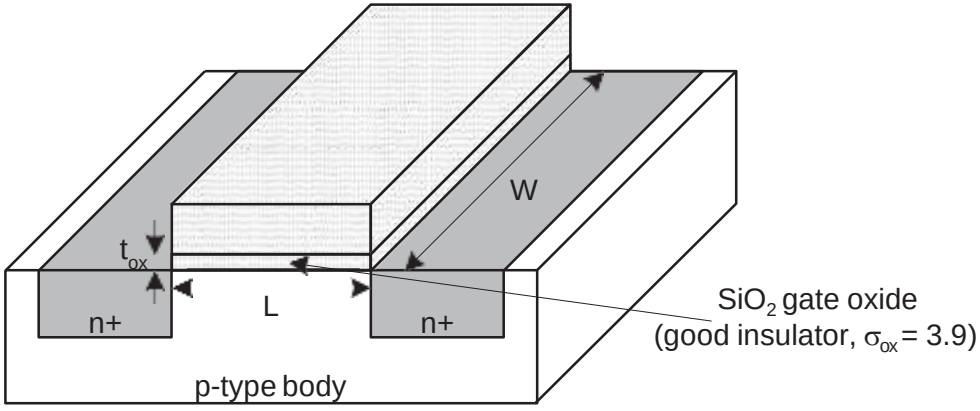
$C_{ox} = \sigma_{ox} / t_{ox}$

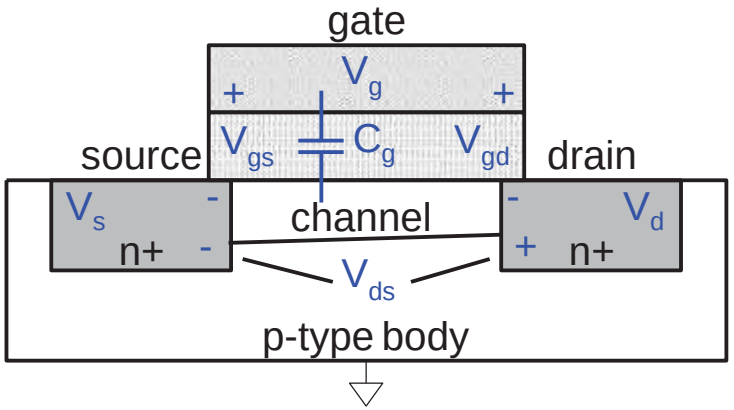




Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- Qchannel = CV
- C = Cg = eoxWL/tox = CoxWL $C_{ox} = \sigma_{ox} / t_{ox}$
- V = Vgc – Vt = (Vgs – Vds/2) – Vt





Carrier velocity

- Charge is carried by e-
- Carrier velocity v proportional to lateral E-field between source and drain

Carrier velocity

- Charge is carried by e-
- Carrier velocity v proportional to lateral E-field between source and drain
- v = mE m called mobility
- E =energy

Carrier velocity

- Charge is carried by e-
- Carrier velocity v proportional to lateral E-field between source and drain
- v = mE m called mobility
- E = Vds/L
- Time for carrier to cross channel:
  - t =

Carrier velocity

- Charge is carried by e-
- Carrier velocity v proportional to lateral E-field between source and drain
- $v = mE$ m called mobility
- $E = V_{ds}/L$
- Time for carrier to cross channel:
  - $t = L / v$

nMOS Linear I-V

- Now we know
  - How much charge Qchannel is in the channel
  - How much time t each carrier takes to cross

$I_{ds} =$

nMOS Linear I-V

- Now we know
  - How much charge Qchannel is in the channel
  - How much time t each carrier takes to cross

$$I_{ds} = \frac{Q_{\text{channel}}}{t}$$

$$=$$

nMOS Linear I-V

- Now we know
  - How much charge Qchannel is in the channel
  - How much time t each carrier takes to cross

$$I_{ds} = \frac{Q_{channel}}{t}$$

$$= \mu C_{ox} \frac{W}{L} \left[(V_{gs} - V_t - V_{ds}/2) V_{ds} \right]$$

$$= \beta \left[(V_{gs} - V_t - V_{ds}/2) V_{ds} \right] \qquad \beta = \mu C_{ox} \frac{W}{L}$$

nMOS Saturation I-V

- If Vgd < Vt, channel pinches off near drain
  - When Vds > Vdsat = Vgs – Vt
- Now drain voltage no longer increases current

$$I_{ds} =$$

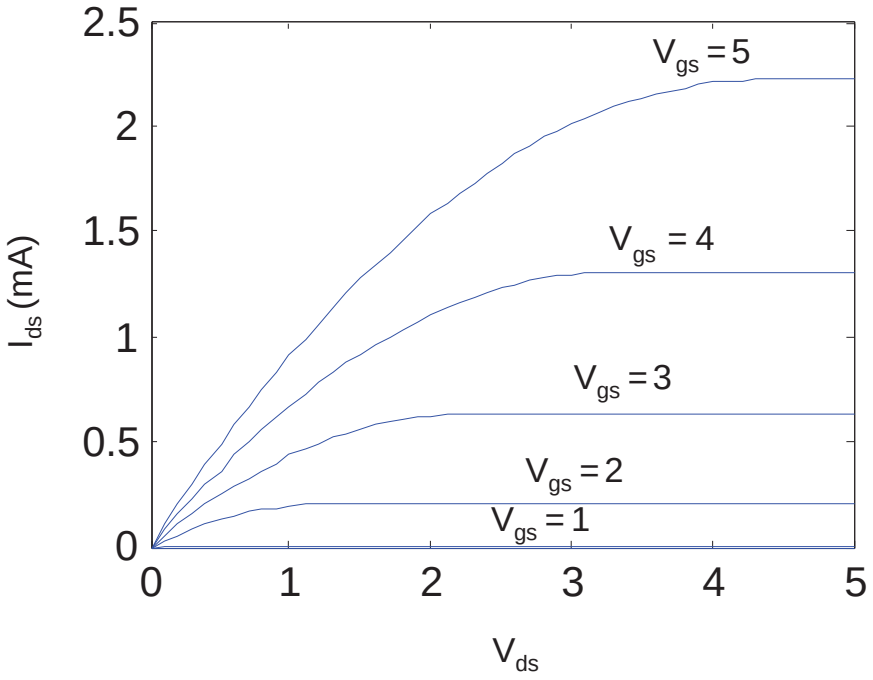
nMOS Saturation I-V

- If Vgd < Vt, channel pinches off near drain
  - When Vds > Vdsat = Vgs – Vt
- Now drain voltage no longer increases current

$$I_{ds} = \beta \left\{ V_{gs} - V_t - V_{dsat}/2 \right\} V_{dsat}$$

Example

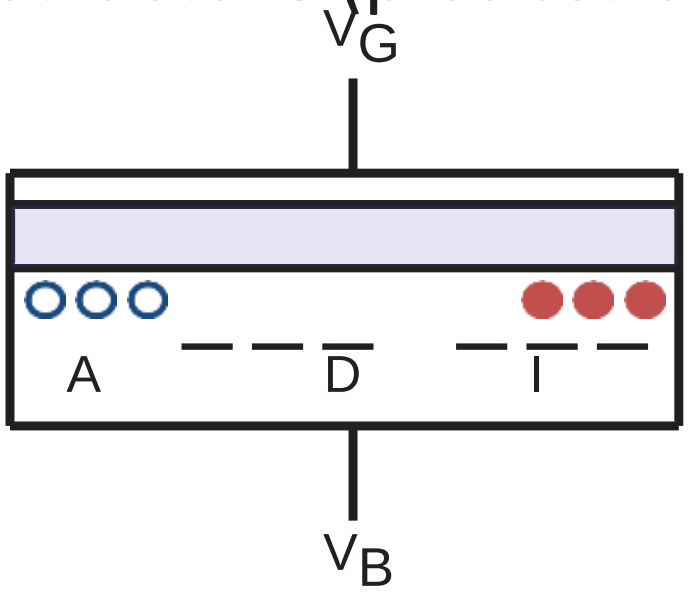
- Example: a 0.6 mm process from AMI semiconductor
  - tox = 100 Å
  - m = 350 cm2/V*s
  - Vt = 0.7 V
- Plot Ids vs. Vds
  - Vgs = 0, 1, 2, 3, 4, 5
  - Use W/L = 4/2 l



$$\beta = \mu C_{ox} \frac{W}{L} = (350)\left\{\frac{3.9 \bullet 8.85 \bullet 10^{-14}}{100 \bullet 10^{-8}}\right\}\left\{\frac{W}{L}\right\} = 120\frac{W}{L}\ \mu A/V^2$$

MOS Transistor Basics
Two Terminal Structure

- Two terminal structure (p-substrate): The MOS capacitor



- Important derived parameters. With VG = VB = 0:
  - F – Buck Fermi Potential (Substrate)
  - S – Surface Potential (Substrate)

MOS Transistor Basics
Two Terminal Structure (Continued)

  - VSB – Flat Band Voltage (applied external voltage to G-B to flatten bands of substrate – equal to built-in potential difference of MOS – equal to work function difference ∃GB between the substrate (channel) and gate.
- Operation
  - With VG<0, VB=0, Accumulation – Holes accumulate at substrate-oxide interface due to attraction of negative bias
  - With VG>0, but small, VB=0, Depletion – Holes repelled from substrate-oxide interface due to positive bias leaving negatively charged fixed acceptors ions behind. The result is a region below the interface that is depleted of mobile carriers.
- Depletion region thickness

$$x_d = \sqrt{\frac{2\sigma_{Si}\left|\exists_s - \exists_F\right|}{q N_A}}$$

MOS Transistor Basics
Two Terminal Structure (Continued)

- Depletion region charge density

  - Note that this density is per unit of area.
  - With VG>0 and larger, VB=0, Inversion – A n-type inversion layer forms, a condition known as surface inversion. The surface is inverted when the density of electrons at the surface equals the density of holes in the bulk. This implies that ∃s has the same magnitude but opposite sign to ∃F. At the point depletion depth fixed and the maximum depletion region depth is at ∃s = -∃F. This depth is:

$$x_{dm} = \sqrt{\frac{2\sigma_{Si}|2\exists_F|}{q N_A}}$$

$$Q = -q N_A x_d = -\sqrt{2q N_A \sigma_{Si} \sharp_s - \exists_F |}$$

MOS Transistor Basics
Two Terminal Structure (Continued)

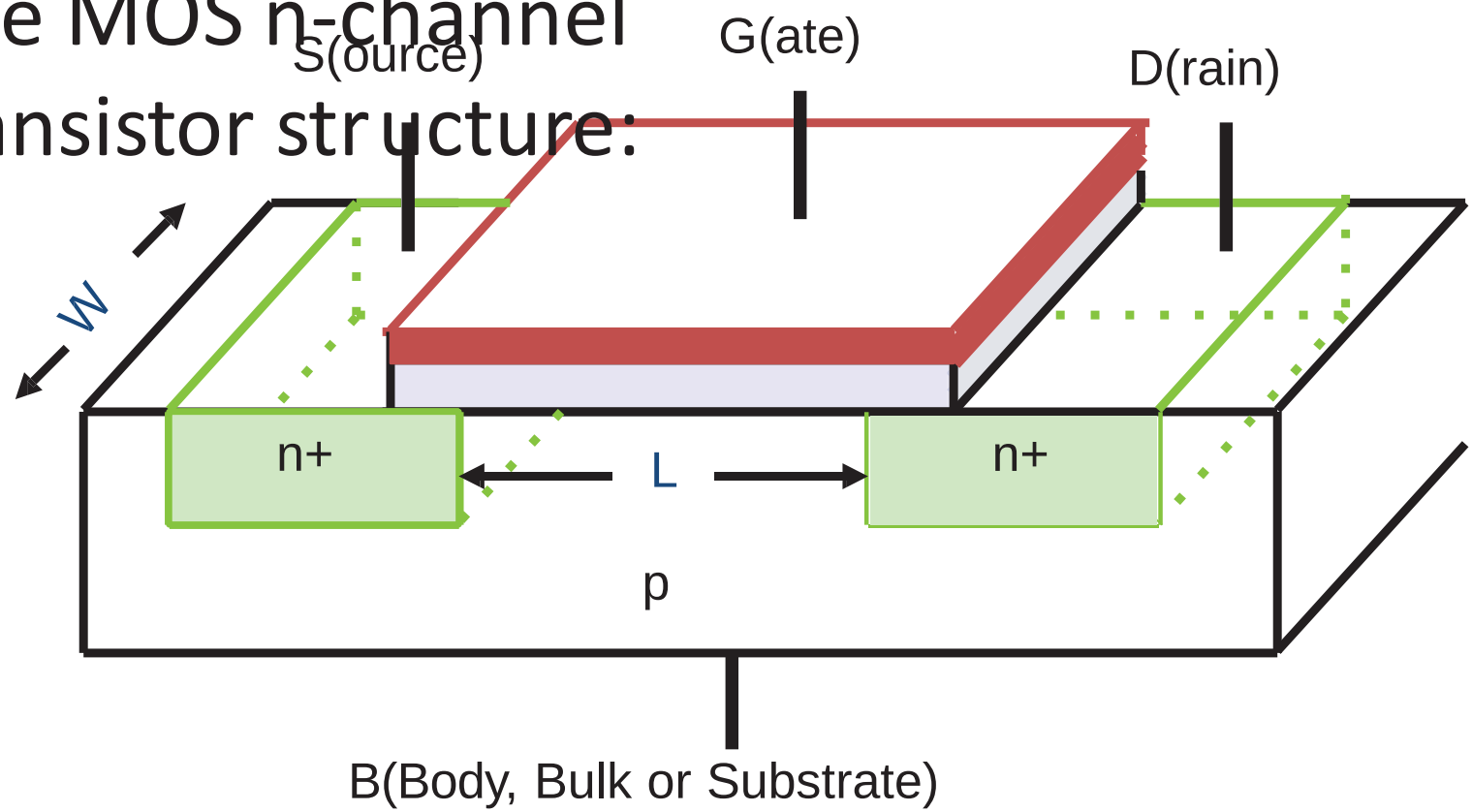
- The corresponding depletion charge density (per unit area) at surface inversion is

- The inversion phenomena is the mechanism that forms the n-channel. The depletion depth and the depletion region charge are critical in determining properties of MOSFET.

$$Q_0 = -q\,N_A x_d = -\sqrt{2q\,N_A \sigma_{Si} \left| 2\exists_F \right|}$$

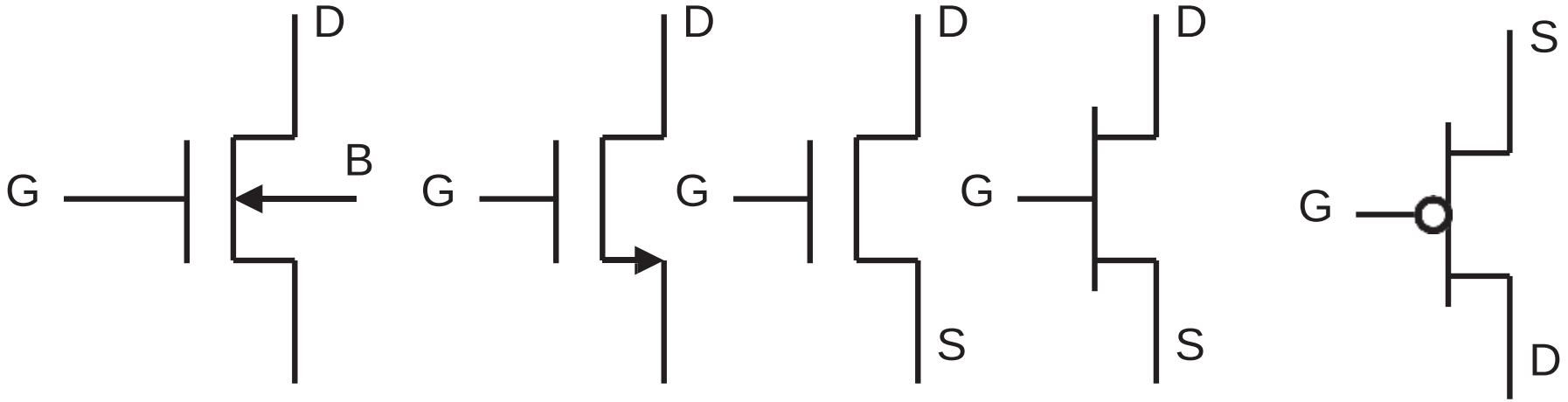
MOS Transistor Basics
Four Terminal Structure

- p-Substrate
- The MOS n-channel transistor structure:



MOS Transistor Basics
Four Terminal Structure (Continued)

- Symbols: n-channel - p-substrate; p-channel – n-substrate

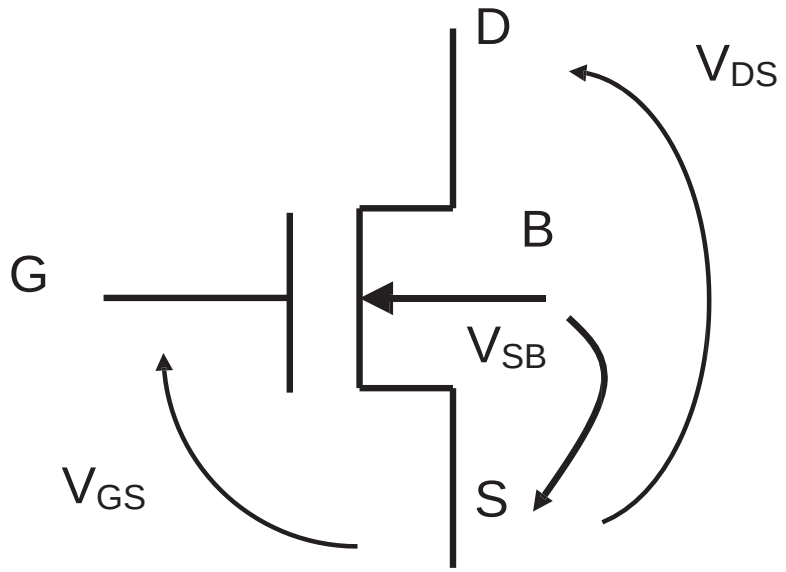


P-

- channel
- Enhancement mode: no conducting channel exists at VGS = 0
- Depletion mode: a conducting channel exists at VGS = 0

MOS Transistor Basics
Four Terminal Structure (Continued)

- Source and drain identification



Threshold Voltage Components

- Consider the prior 3-D drawing: Set VS=0, VDS=0, and VSB=0.
  - Increase VGS until the channel is inverted. Then a conducting channel is formed and the depletion region thickness (depth) is maximum as is the surface potential.
  - The value of VGS needed to cause surface inversion (channel creation) is the threshold voltage VT0. The 0 refers to VSB=0.
  - VGS< VT0: no channel implies no current flow possible. With VGS> VT0, existence the channel implies possible current flow.

Threshold Voltage Components (Cont.)

- 8GC work function difference between gate and channel material which is the built-in voltage that must be offset by voltage applied to flatten the bands at the surface.
- Apply voltage to achieve surface inversion -2∃F
- Additional voltage must be applied to offset the depletion region charge due to the acceptor ions. At inversion, this charge with VSB=0 is QB0= Q0.
- For VSB non-zero,

$$Q = -\sqrt{2q N_A \sigma_{Si} \left| -2\exists_F + V_{SB} \right|}$$

- The voltage required to offset the depletion region charge is defined by –QB/Cox where Cox = εox/tox with tox, the oxide thickness, and Cox, the gate oxide capacitance per unit area.

- 4) The final component is a fixed positive charge density that appears at the interface between the oxide and the substrate, Qox. The voltage to offset this charge is:

$$\frac{-Q_o}{C_{ox}}$$

Threshold Voltage Components (Cont.)

- These components together give:

$$V_T = 8_{GC} - 2\exists_F - \frac{Q_B}{C_{ox}} - \frac{Q_{ox}}{C_{ox}}$$

- For VSB=0, VT0 has QB replaced by QB0. This gives a relationship between VT and VT0 which is:

$$V_T = V_{T0} - \frac{Q_B - Q_{B0}}{C_{ox}}$$

- Thus the actual threshold voltage VT differs from VT0 by the term given. Going back to the definition of QB, this term is equal to:

- In which γ is the substrate-bias (or body effect) coefficient.

$$+ \psi\left(\sqrt{+ 2\exists_F + V_{SB} +} - \sqrt{|2\exists_F|}\right)$$

Threshold Voltage Components (Cont.)

- The final expression for VT0 and VT are

$$\psi = \frac{\sqrt{2q N_A \sigma_{Si}}}{C_{ox}}$$

- and

$$V_{T0} = 8_{GC} - 2\exists_F - \frac{Q_{B0}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}}$$

  - The threshold voltage depends on the source-to-bulk voltage which is clearly separated out. The component is referred to as body effect. If the source to body voltage VSB is non-zero, the corrective term must be applied to VT0.

$$V_T = V_{T0} + \psi\left(\sqrt{\left|+2\exists_F + \ V_{SB}\right|} - \sqrt{\left|2\exists_F\right|}\right)$$

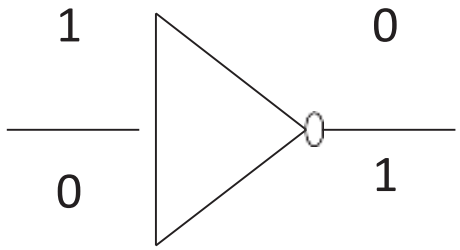
Threshold Voltage Components (Cont.)

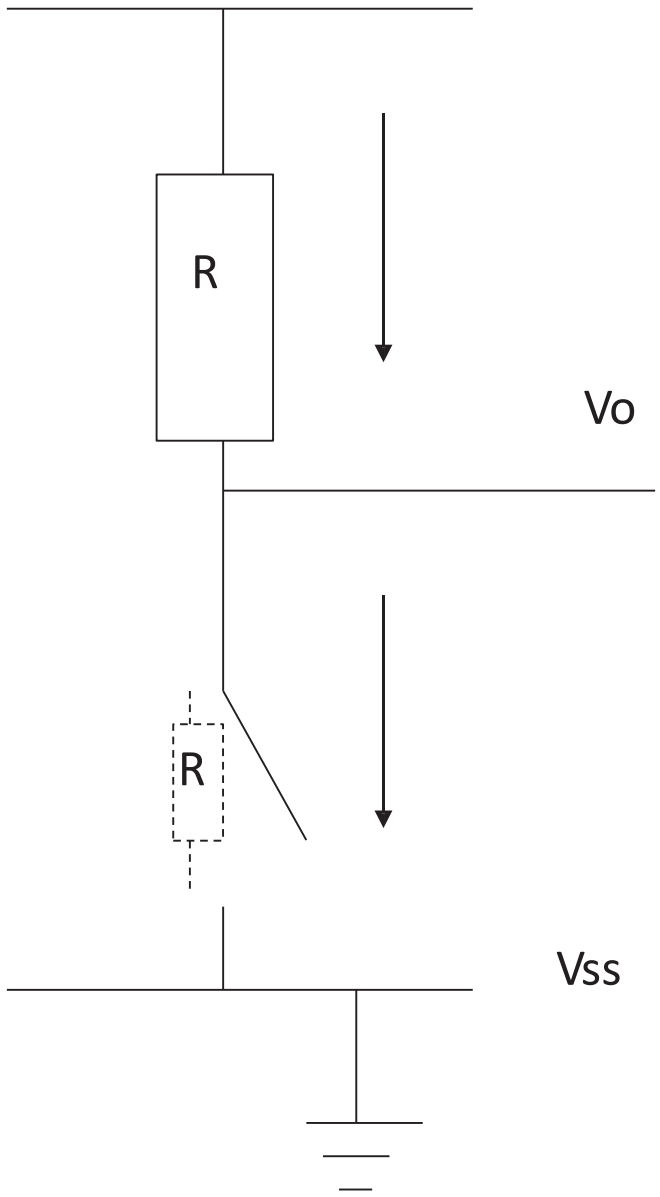
-
- Those parameters in the VT equation are signed. The following table gives their signs for nMOS and pMOS transistor.

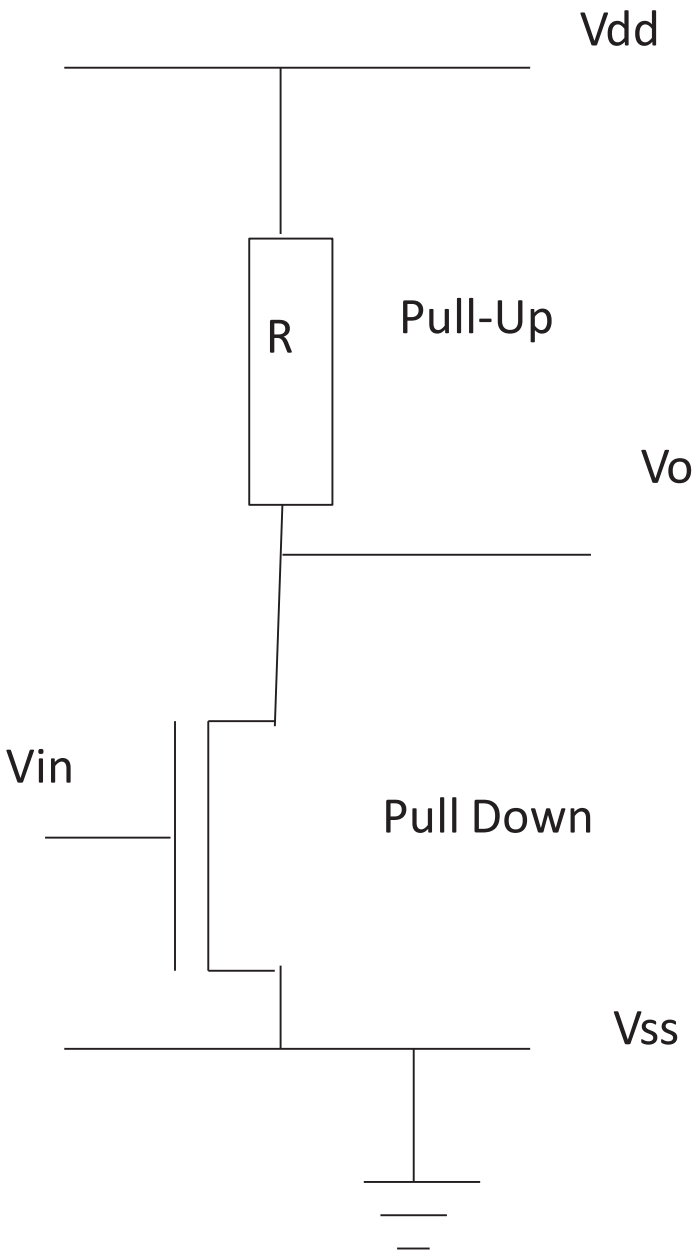
Parameter	nMOS	pMOS
$\exists_F$	−	+
Q_B, Q_{B0}	−	+
γ	+	−
V_{SB}	+	−

- For real designs, the threshold voltage, due to variation in oxide thickness, impurity concentrations, etc., VT0 and γ should be measured from the actual process.

- Inverter : basic requirement for producing a complete range of Logic circuits







Basic Inverter: Transistor with source connected to ground and a load resistor connected from the drain to the positive Supply rail

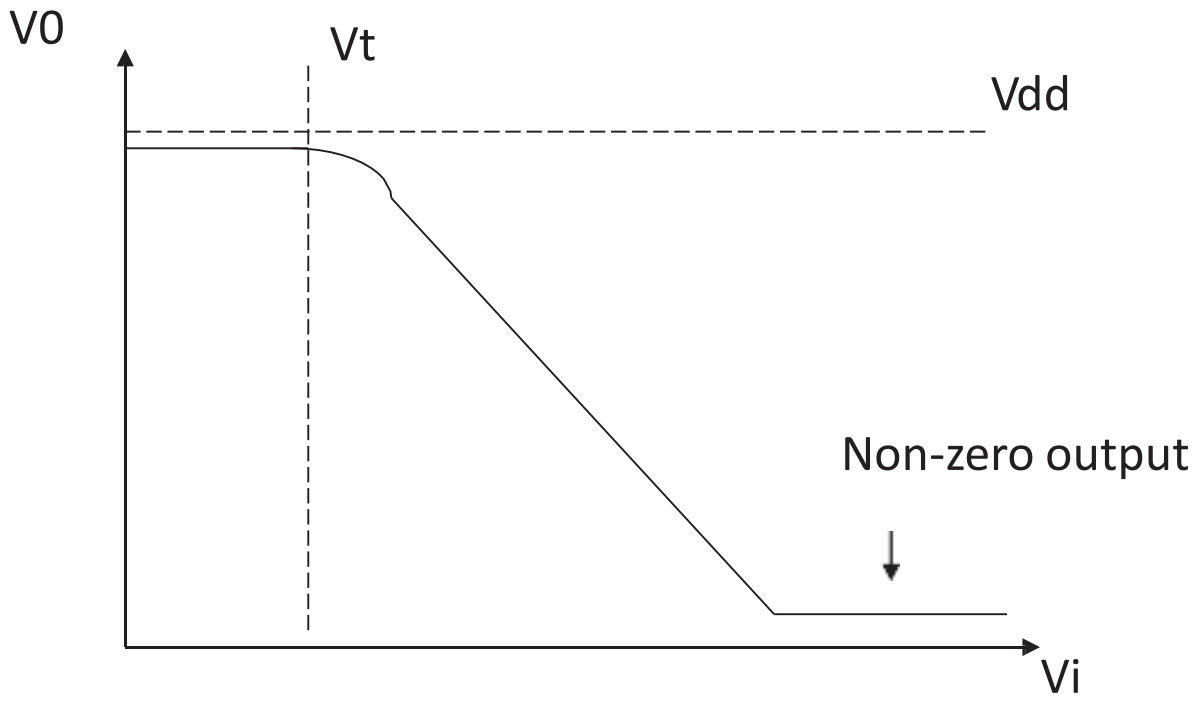
Output is taken from the drain and control input connected between gate and ground

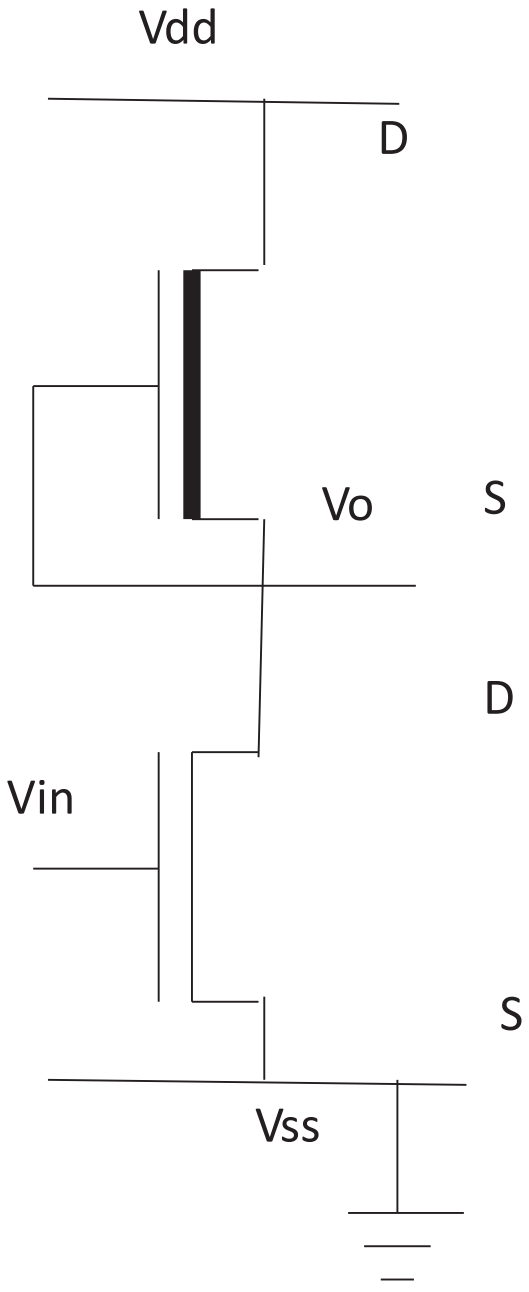
Resistors are not easily formed in silicon - they occupy too much area

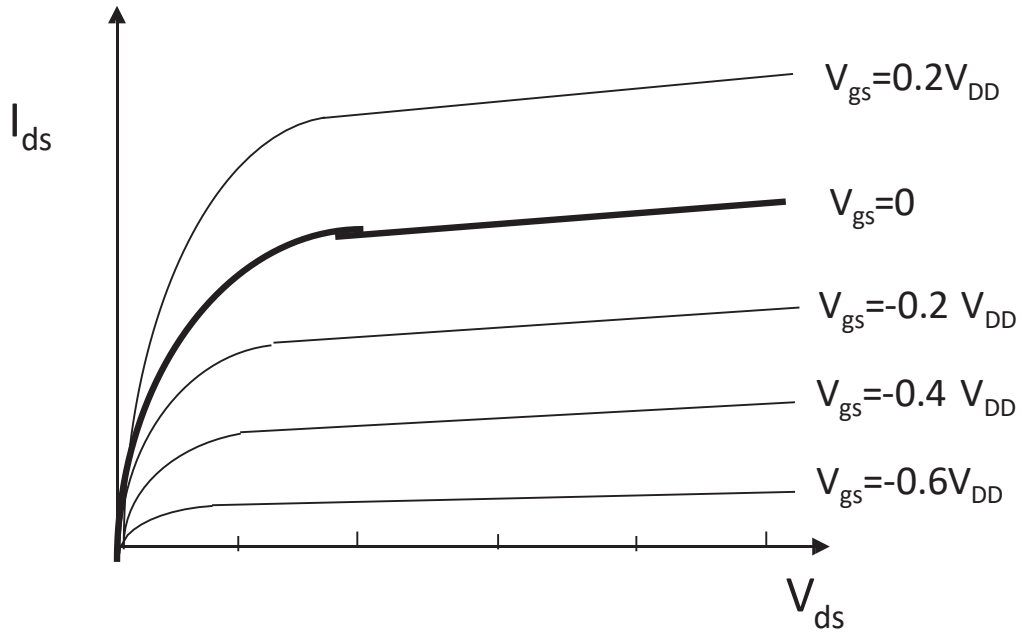
Transistors can be used as the pull-up device

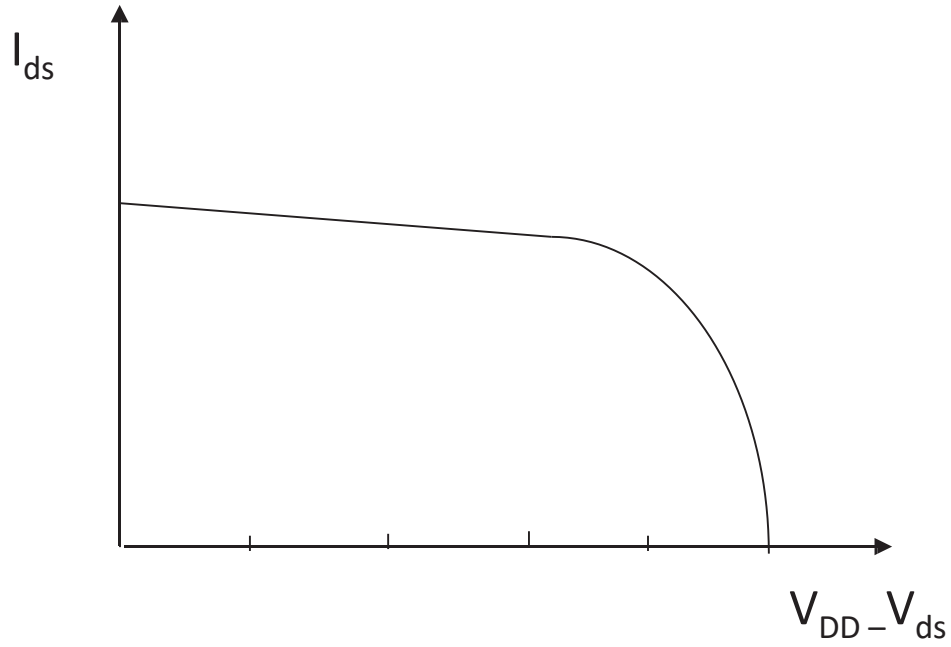
NMOS Depletion Mode Transistor Pull - Up

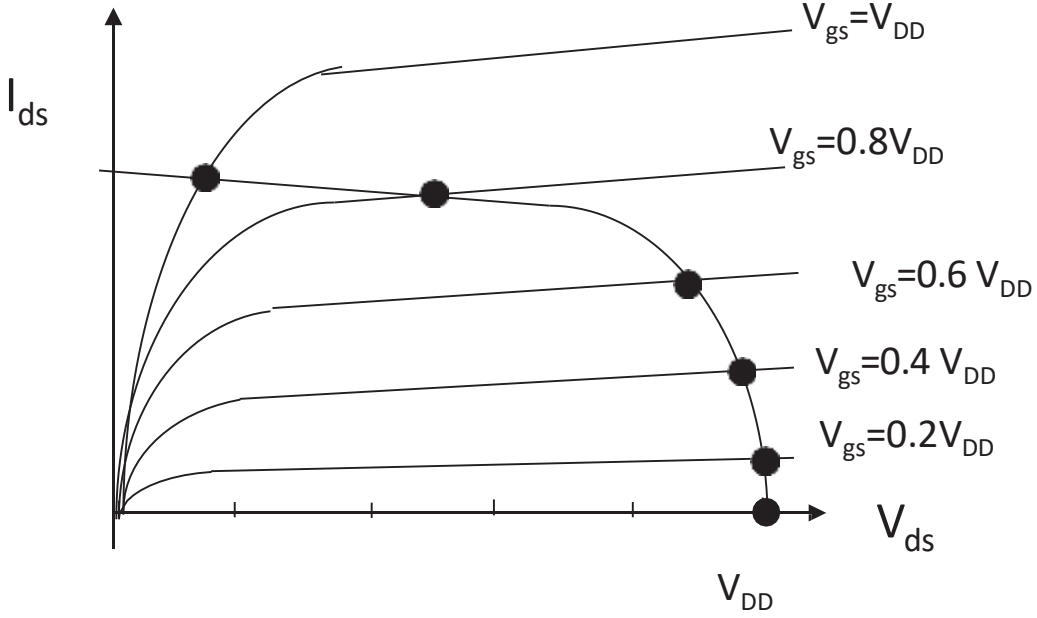
- Pull-Up is always on – Vgs = 0; depletion
- Pull-Down turns on when Vin > Vt
- With no current drawn from outputs, Ids for both transistors is equal

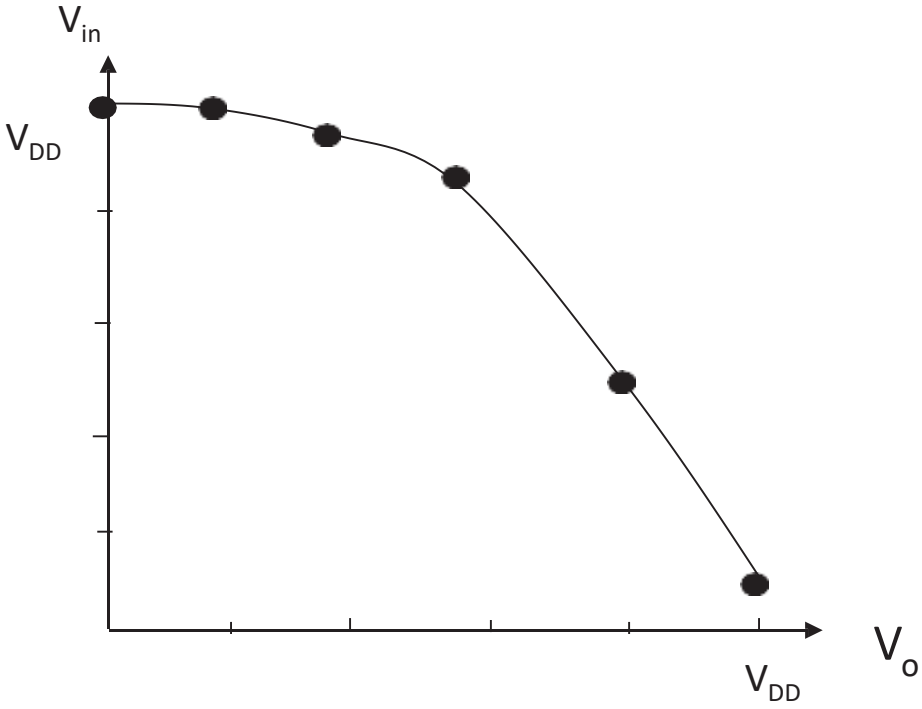


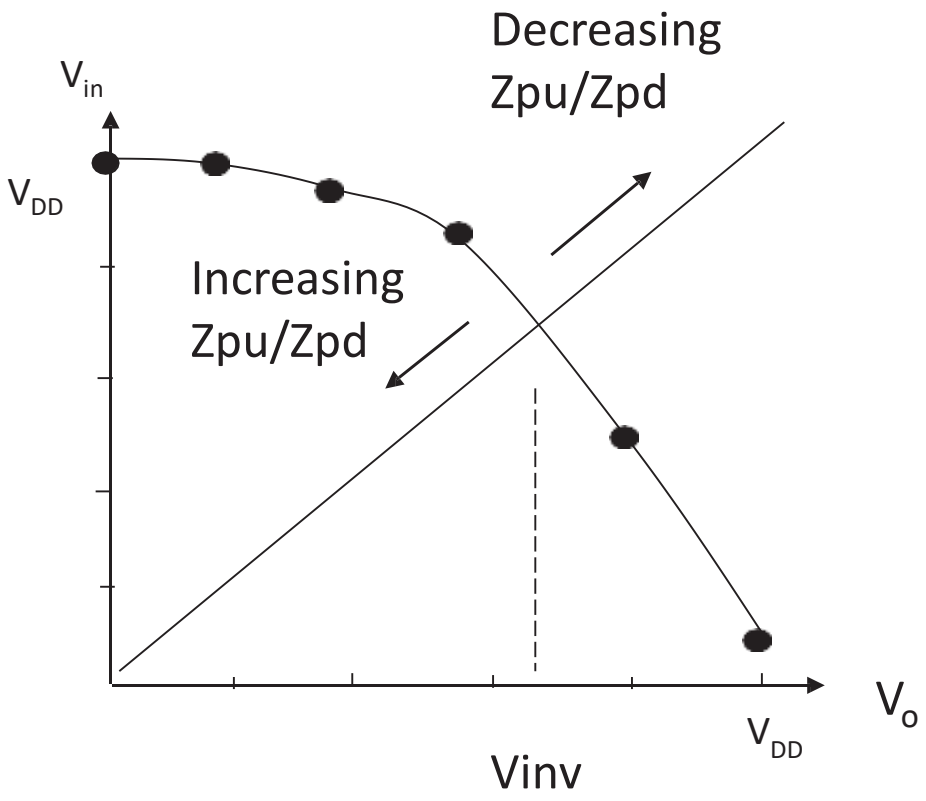



I_{ds}
$V_{gs}=0.2V_{DD}$
$V_{gs}=0$
$V_{gs}=-0.2\ V_{DD}$
$V_{gs}=-0.4\ V_{DD}$
$V_{gs}=-0.6V_{DD}$
V_{ds}


I_{ds}
$V_{DD}-V_{ds}$


I_{ds}
$V_{gs}=V_{DD}$
$V_{gs}=0.8V_{DD}$
$V_{gs}=0.6\ V_{DD}$
$V_{gs}=0.4\ V_{DD}$
$V_{gs}=0.2V_{DD}$
V_{ds}
V_{DD}


V_{in}
V_{DD}
V_{DD}
V_o



- Point where Vo = Vin is called Vinv

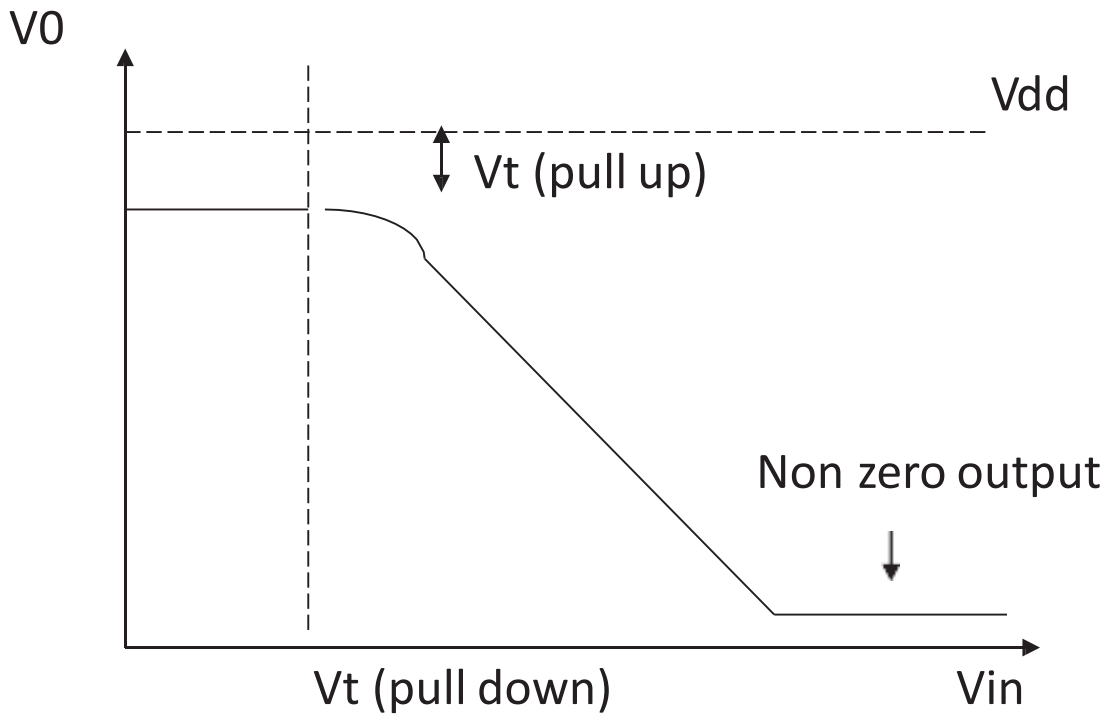
- Transfer Characteristics and Vinv can be shifted by altering ratio of pull-up to Pull down impedances

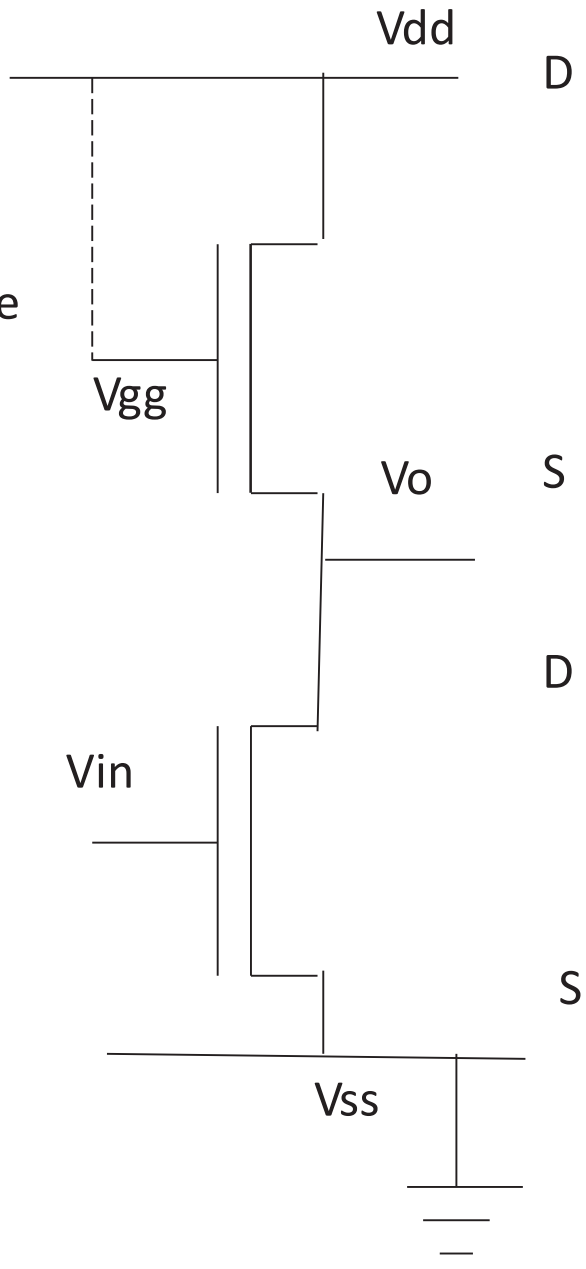
NMOS Depletion Mode Inverter Characteristics

- Dissipation is high since rail to rail current flows when Vin = Logical 1
- Switching of Output from 1 to 0 begins when Vin exceeds Vt of pull down device
- When switching the output from 1 to 0, the pull up device is non-saturated initially and this presents a lower resistance through which to charge capacitors ($V_{ds} < V_{gs} - V_t$)

NMOS Enhancement Mode Transistor Pull - Up

- Dissipation is high since current flows when Vin = 1
- Vout can never reach Vdd (effect of channel)
- Vgg can be derived from a switching source (i.e. one phase of a clock, so that dissipation can be significantly reduced
- If Vgg is higher than Vdd, and extra supply rail is required

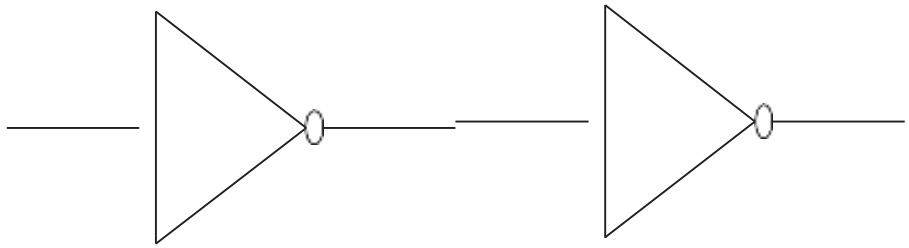




Cascading NMOS Inverters

When cascading logic devices care must be taken to preserve integrity of logic levels

i.e. design circuit so that Vin = Vout = Vinv



Determine pull – up to pull-down ratio for driven inverter

Assume equal margins around inverter; Vinv = 0.5 Vdd

Assume both transistors in saturation, therefore: $I_{ds} = K\,(W/L)\,(V_{gs} - V_t)^2/2$

Depletion mode transistor has gate connected to source, i.e. $V_{gs} = 0$

$I_{ds} = K\,(W_{pu}/L_{pu})\,(-V_{td})^2/2$

Enhancement mode device Vgs = Vinv, therefore

$I_{ds} = K\,(W_{pd}/L_{pd})\,(V_{inv} - V_t)^2/2$

Assume currents are equal through both channels (no current drawn by load)

$(W_{pd}/L_{pd})\,(V_{inv} - V_t)^2 = (W_{pu}/L_{pu})\,(-V_{td})^2$

Convention Z = L/W

$V_{inv} = V_t - V_{td} \,/\, (Z_{pu}/Z_{pd})^{1/2}$

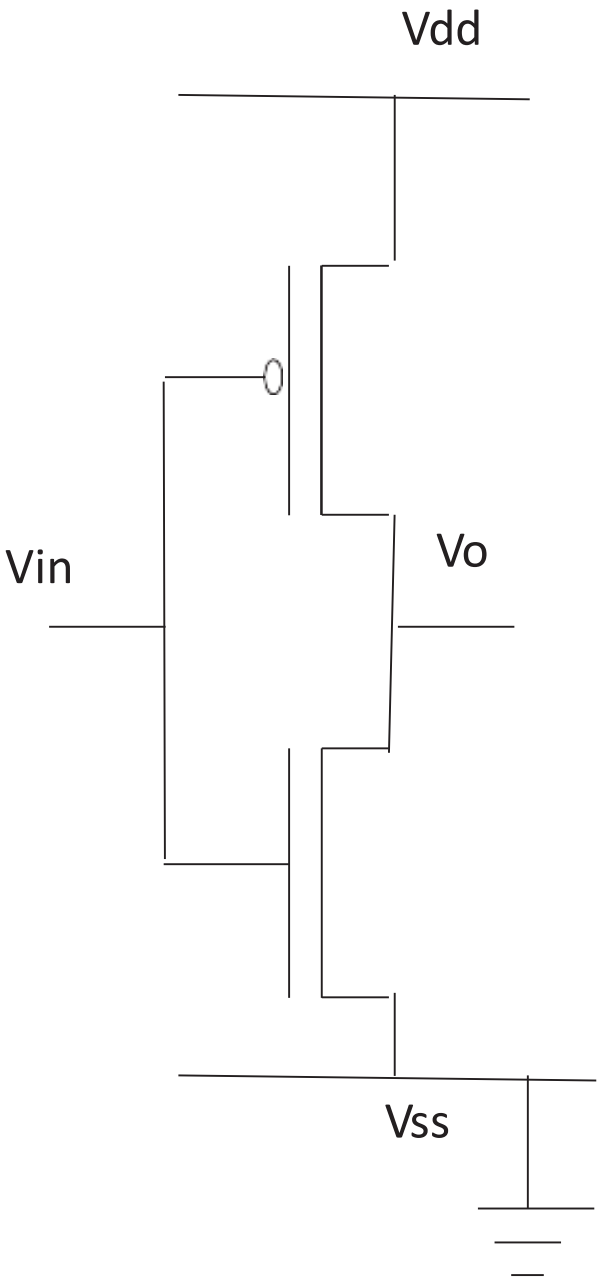
Substitute in typical values $V_t = 0.2\ V_{dd}$; $V_{td} = -0.6\ V_{dd}$; $Vinv = 0.5\ V_{dd}$

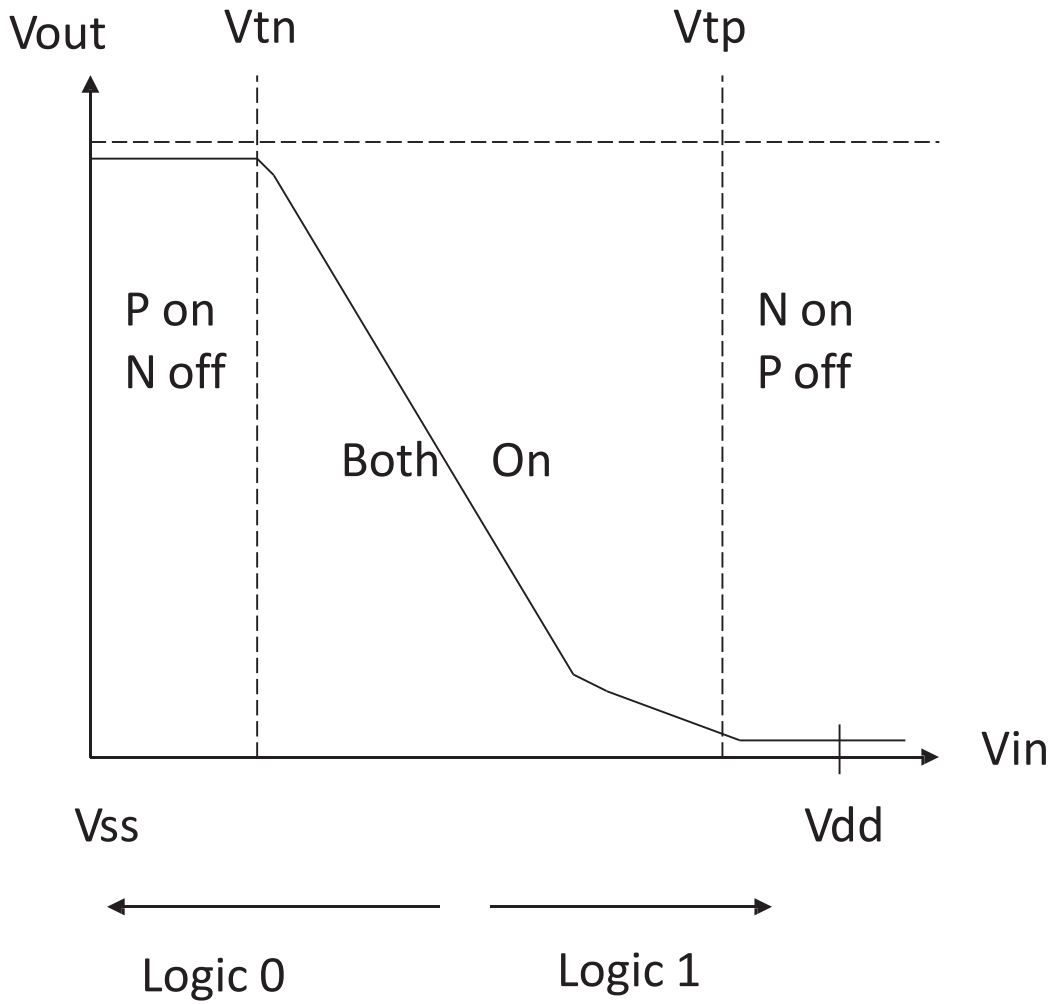
This gives Zpu / Zpd = 4:1 for an nmos inverter directly driven by another inverter

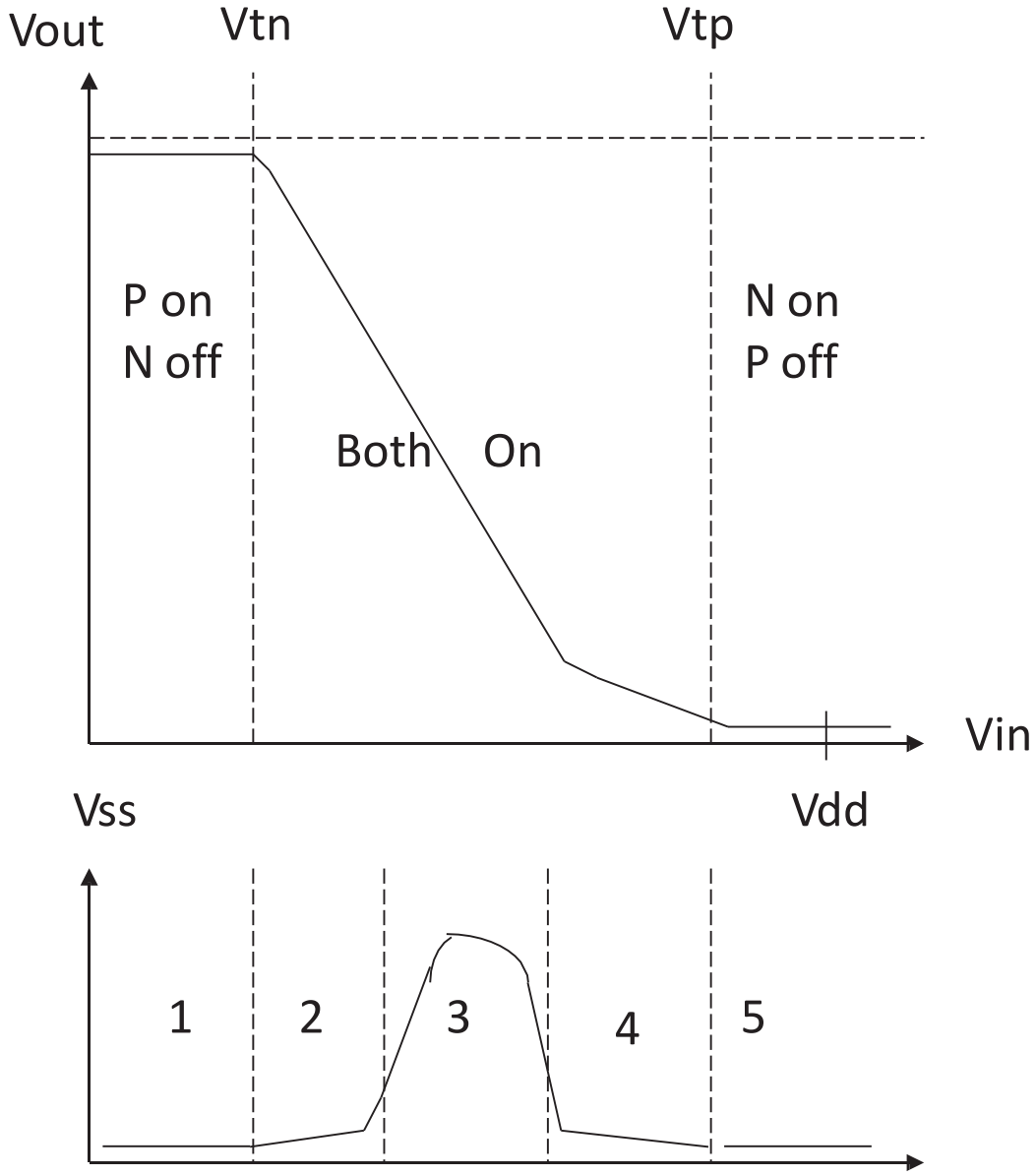
Pull-Up to Pull-Down Ratio for an nMOS inverter driven through 1 or more pass transistors

It is often the case that two inverters are connected via a series of switches (Pass Transistors) We are concerned that connection of transistors in series will degrade the logic levels into Inverter 2. The driven inverter can be designed to deal with this. (Zpu/Zpd >= 8/1)

Complimentary Transistor Pull – Up (CMOS)







1: Logic 0 : p on ; n off

5: Logic 1: p off ; n on

2: Vin > Vtn.
Vdsn large – n in saturation
Vdsp small – p in resistive
Small current from Vdd to Vss

4: same as 2 except reversed p and n

3: Both transistors are in saturation
Large instantaneous current flows

CMOS INVERTER CHARACTERISTICS

Current through n-channel pull-down transistor

$$I_n = \frac{\beta_n}{2}(V_{in} - V_{tn})^2$$

Current through p-channel pull-up transistor

$$I_p = \frac{\beta_p}{2}\left(-(V_{in} - V_{DD}) + V_{tp}\right)^2$$

At logic threshold, $I_n = I_p$

$$\frac{\beta_n}{2}(V_{in} - V_{tn})^2 = \frac{\beta_p}{2}\left(-(V_{in} - V_{DD}) + V_{tp}\right)^2$$

$$\sqrt{\frac{\beta_n}{2}}(V_{in} - V_{tn}) = \sqrt{\frac{\beta_p}{2}}\left(-(V_{in} - V_{DD}) + V_{tp}\right)$$

$$\sqrt{\frac{\beta_n}{\beta_p}}(V_{in} - V_{tn}) = -V_{in} + V_{DD} + V_{tp}$$

$$V_{in}\left\{1 + \sqrt{\frac{\beta_n}{\beta_p}}\right\} = \sqrt{\frac{\beta_n}{\beta_p}}V_{tn} + V_{DD} + V_{tp}$$

$$V_{in} = \frac{V_{DD} + V_{tp} + V_{tn}\sqrt{\frac{\beta_n}{\beta_p}}}{1 + \sqrt{\frac{\beta_n}{\beta_p}}}$$

If $\beta_n = \beta_p$ *and* $V_{tp} = -V_{tn}$

$$V_{in} = \frac{V_{DD}}{2}$$

$$\frac{\mu_p W_p}{L_p} = \frac{\mu_n W_n}{L_n}$$

Mobilities are unequal : $\mu_n = 2.5\ \mu_p$

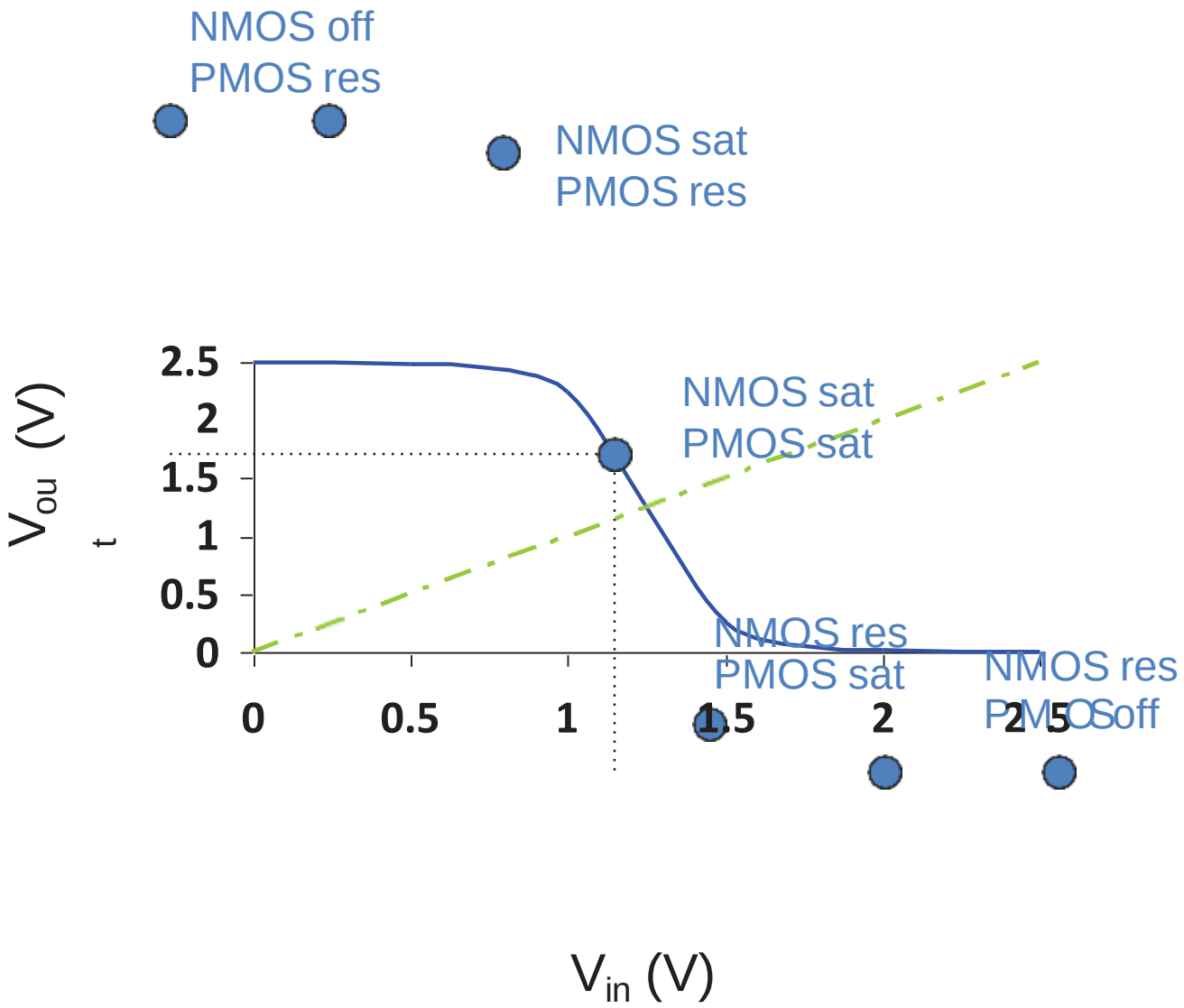
Z = L/W

Z_{pu}/Z_{pd} = 2.5:1 for a symmetrical CMOS inverter

CMOS Inverter Characteristics

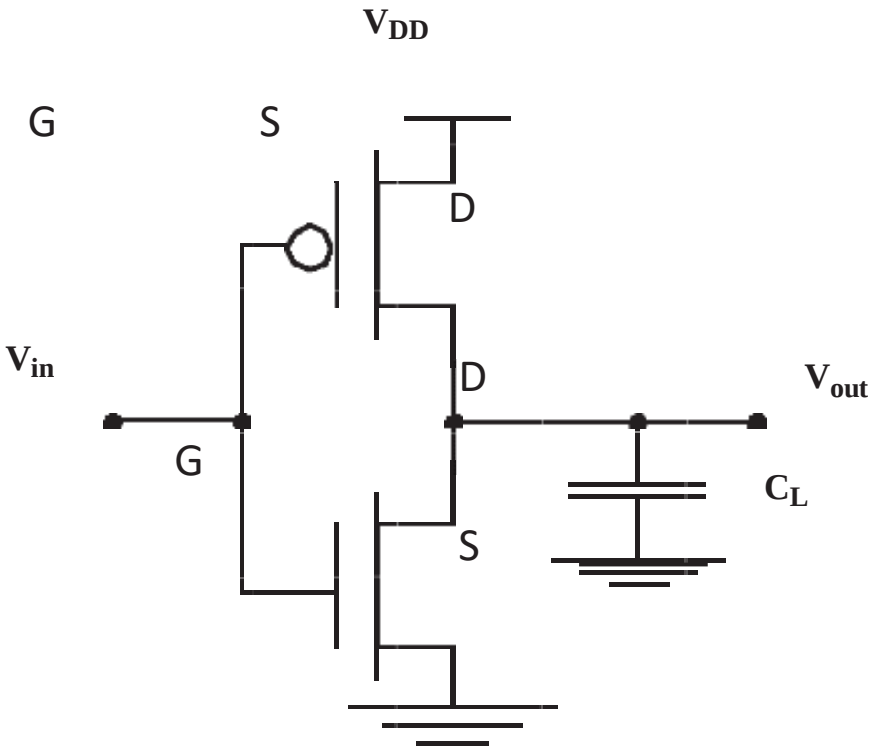
- No current flow for either logical 1 or logical 0 inputs
- Full logical 1 and 0 levels are presented at the output
- For devices of similar dimensions the p – channel is slower than the n – channel device

CMOS Inverter VTC

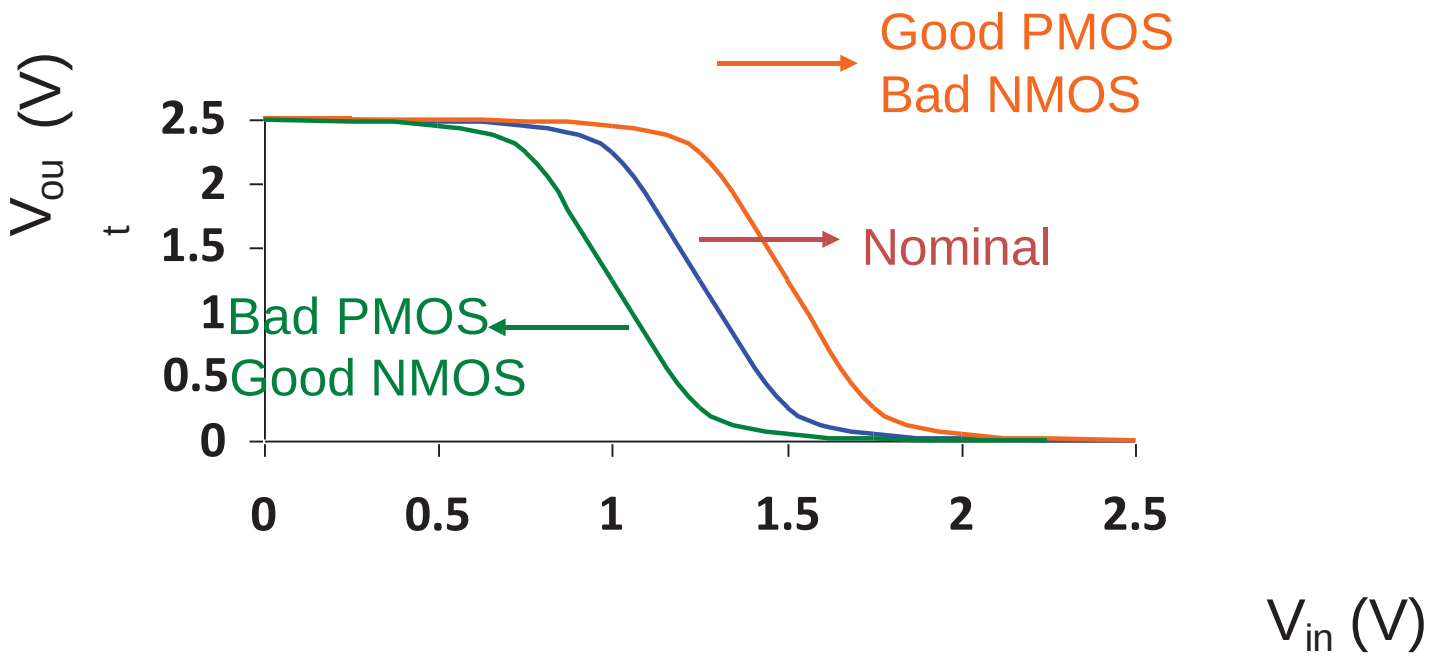


	Cutoff	Linear	Saturation
pMOS	$V_{in} - V_{DD} = V_{GS} > V_T$	$V_{in} - V_{DD} = V_{GS} < V_T$ $V_{in} - V_{out} = V_{GD} < V_T$	$V_{in} - V_{DD} = V_{GS} > V_T$ $V_{in} - V_{out} = V_{GD} > V_T$
nMOS	$V_{in} = V_{GS} < V_T$	$V_{in} = V_{GS} > V_T$ $V_{in} - V_{out} = V_{GD} > V_T$	$V_{in} = V_{GS} > V_T$ $V_{in} - V_{out} = V_{GD} < V_T$

Regions of operations
For nMOS and pMOS
In CMOS inverter



Impact of Process Variation

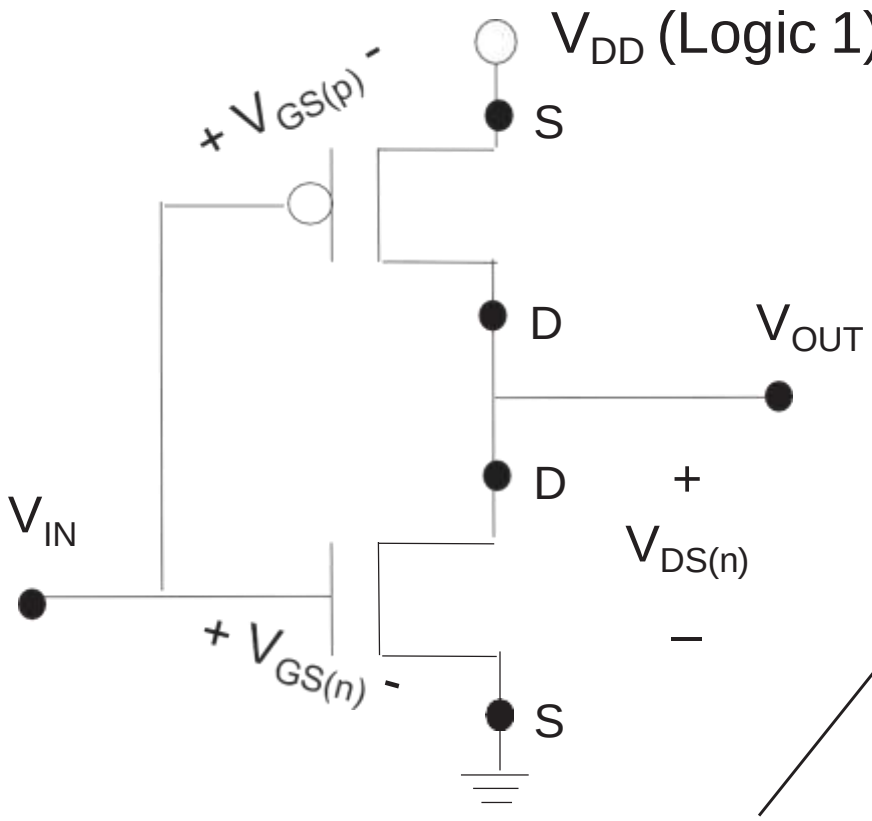


- Pprocess variations (mostly) cause a shift in the switching threshold

Cmos Inverter

- Look at why our NMOS and PMOS inverters might not be the best inverter designs
- Introduce the CMOS inverter
- Analyze how the CMOS inverter works

Linear KVL and KCL Equations



$V_{GS(n)} = V_{IN}$

$V_{GS(p)} = V_{IN} – V_{DD}$

$V_{GS(p)} = V_{DS(n)} - V_{DD}$

$I_{D(p)} = -I_{D(n)}$

$V_{DS(n)} = V_{OUT}$

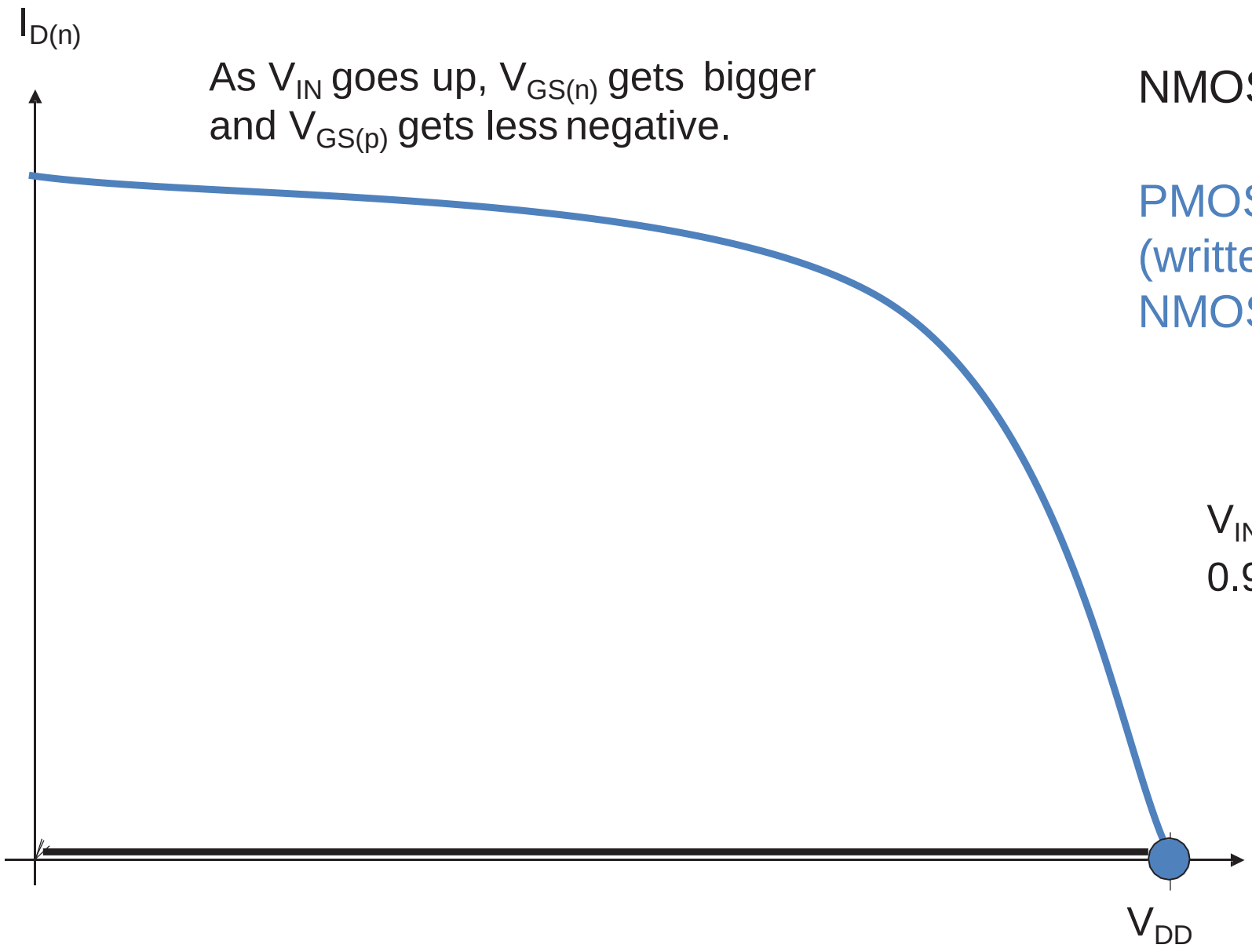
$V_{DS(p)} = V_{OUT} – V_{DD}$

$V_{DS(p)} = V_{DS(n)} - V_{DD}$

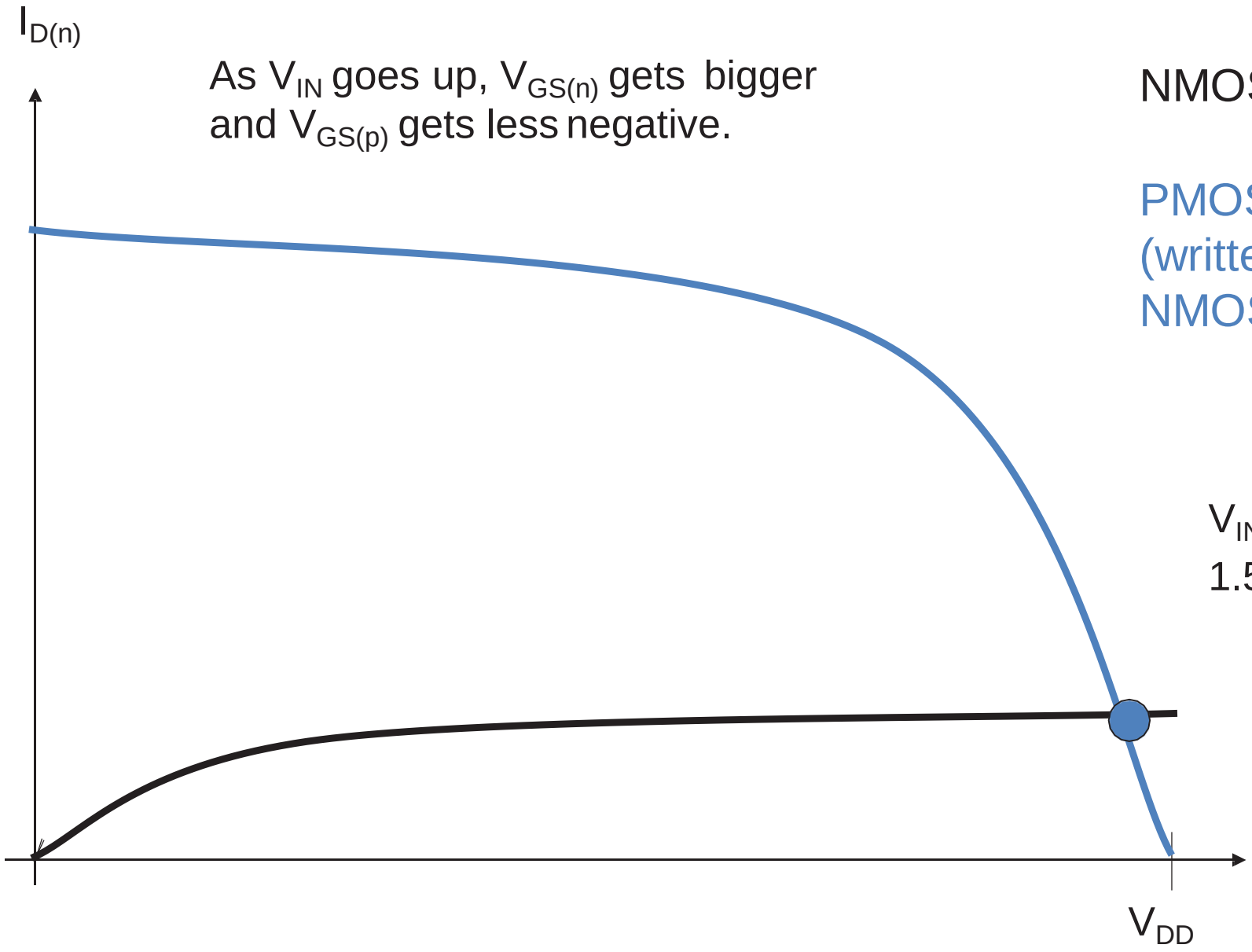
Use these equations

to write both I-V equations in terms of $V_{DS(n)}$ and $I_{D(n)}$

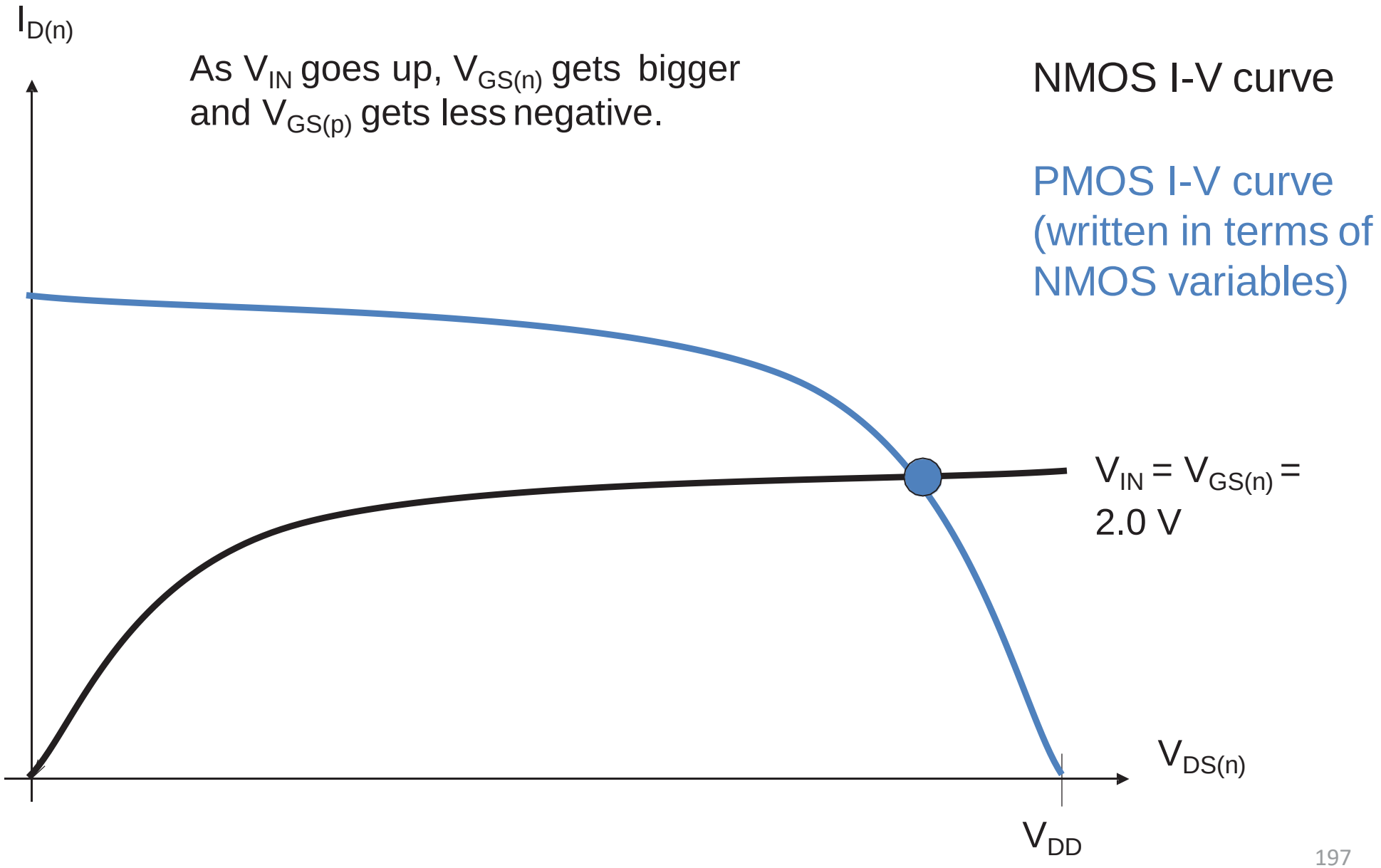
CMOS Analysis



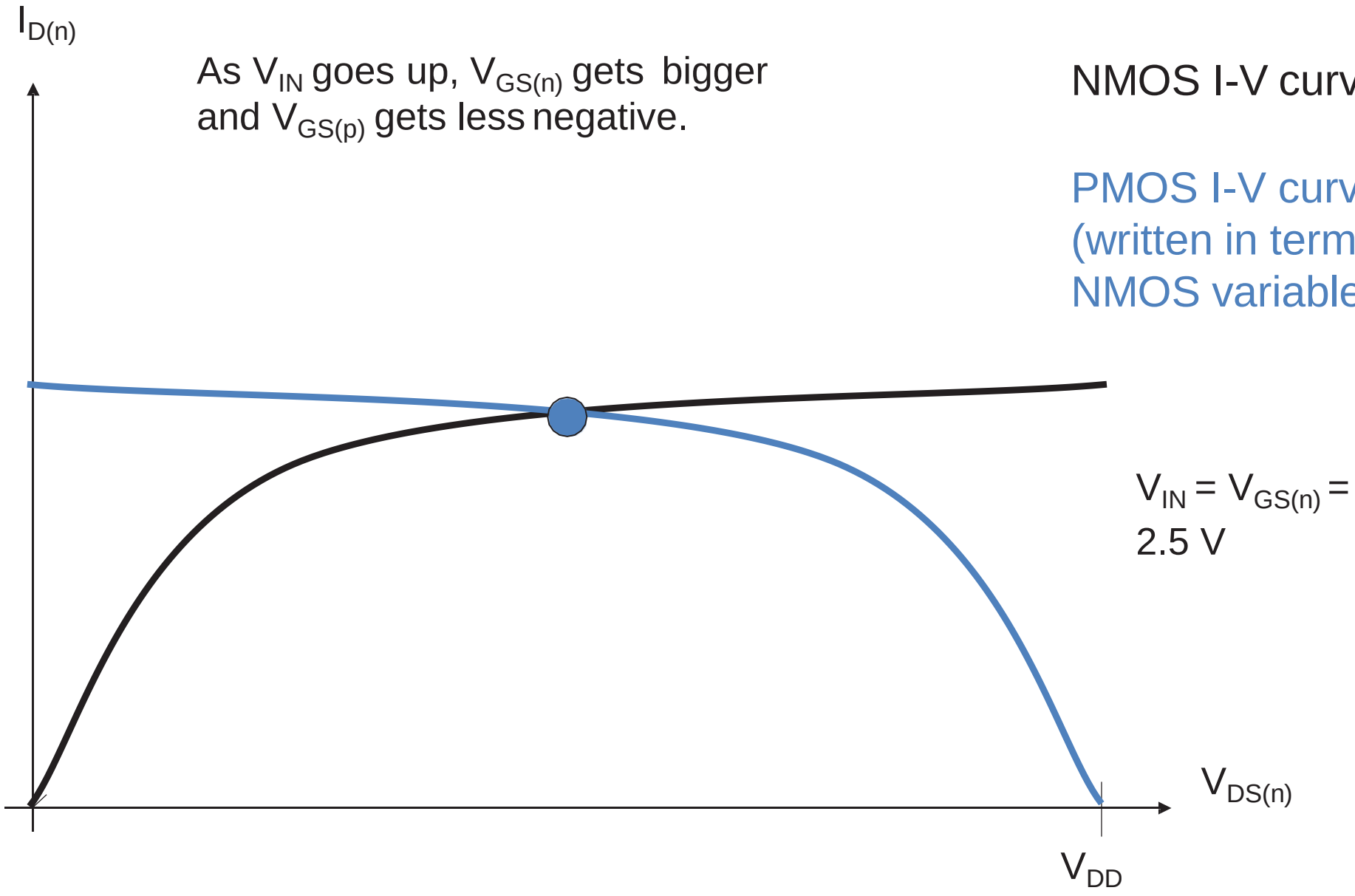
CMOS Analysis



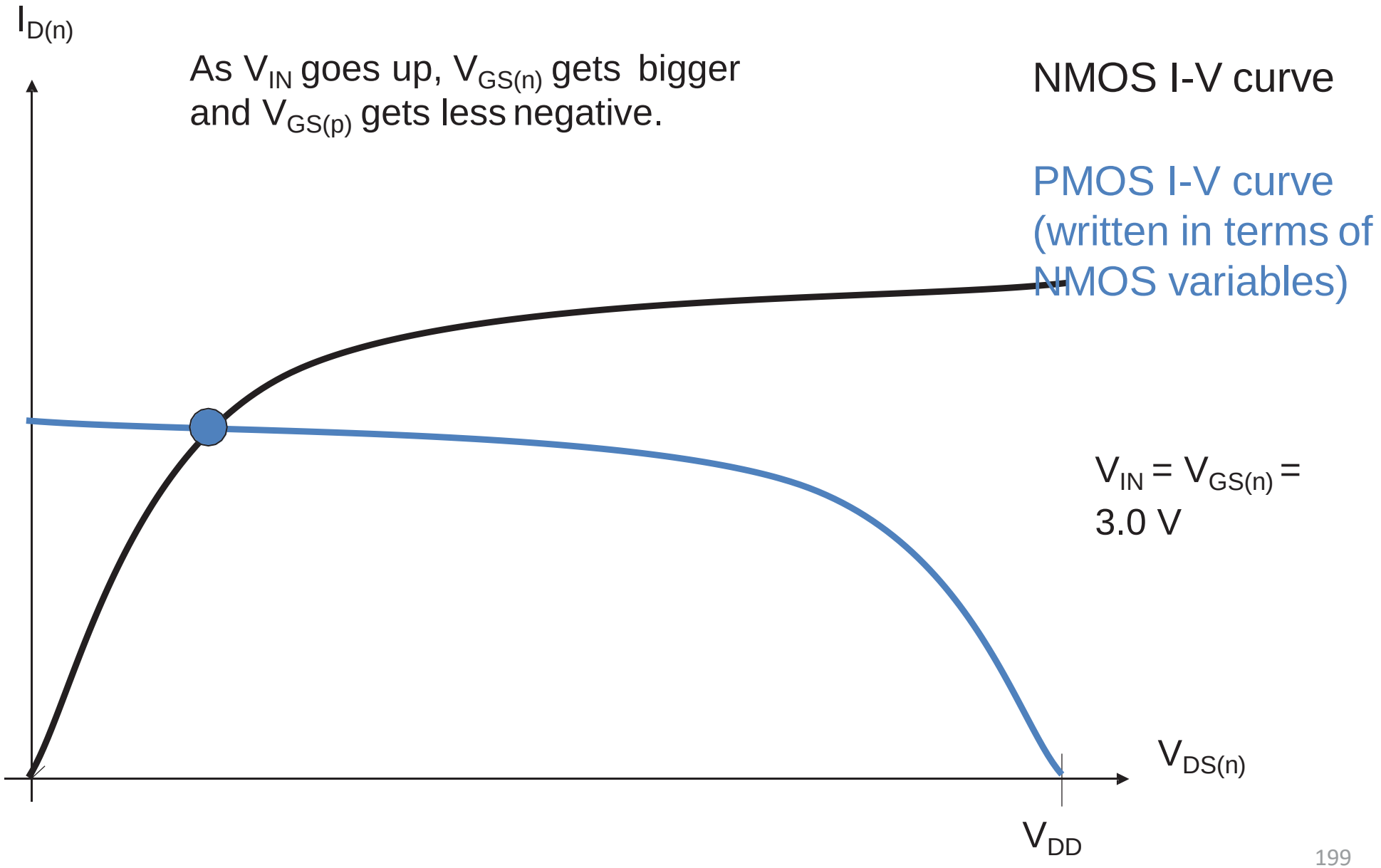
CMOS Analysis



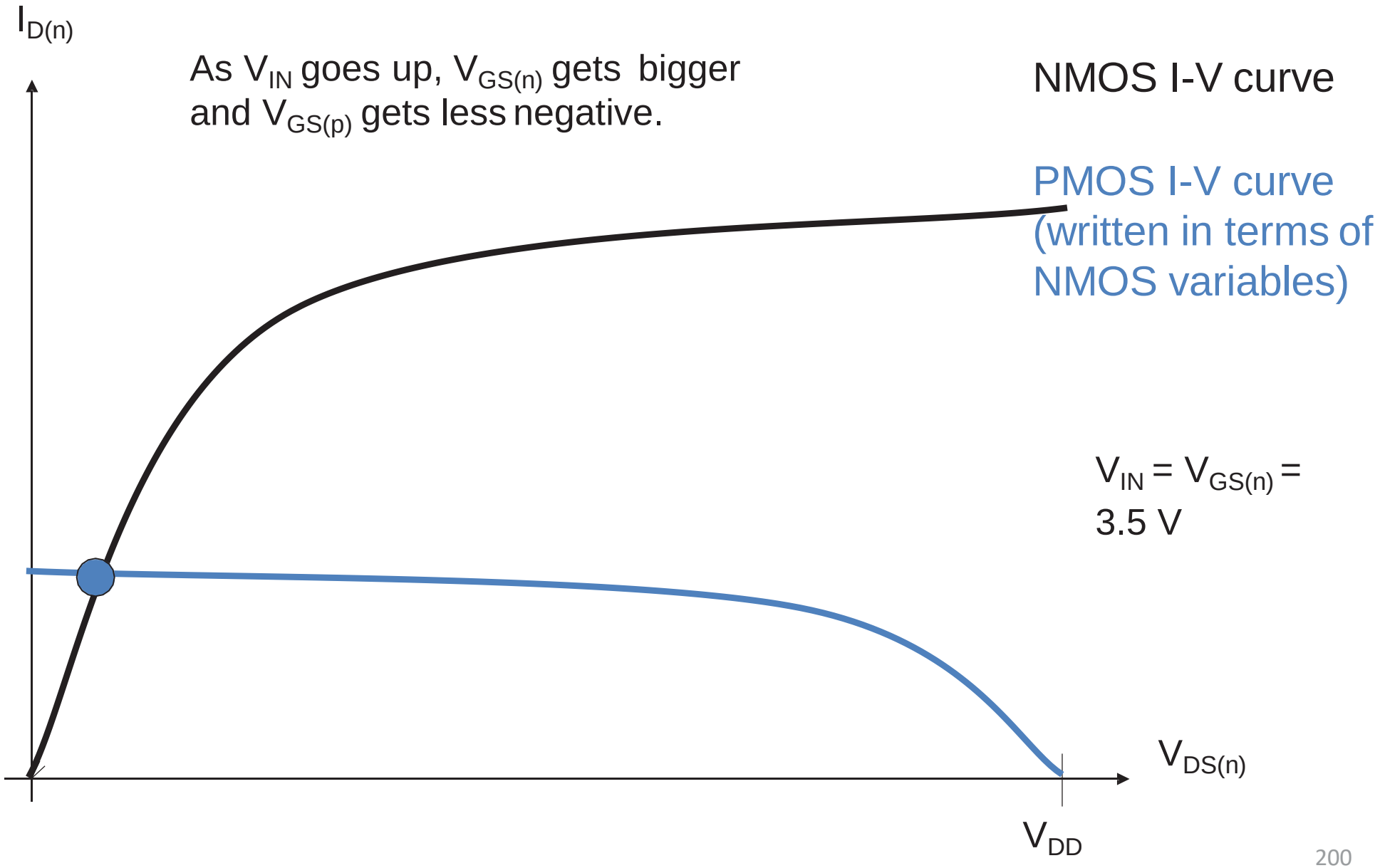
CMOS Analysis



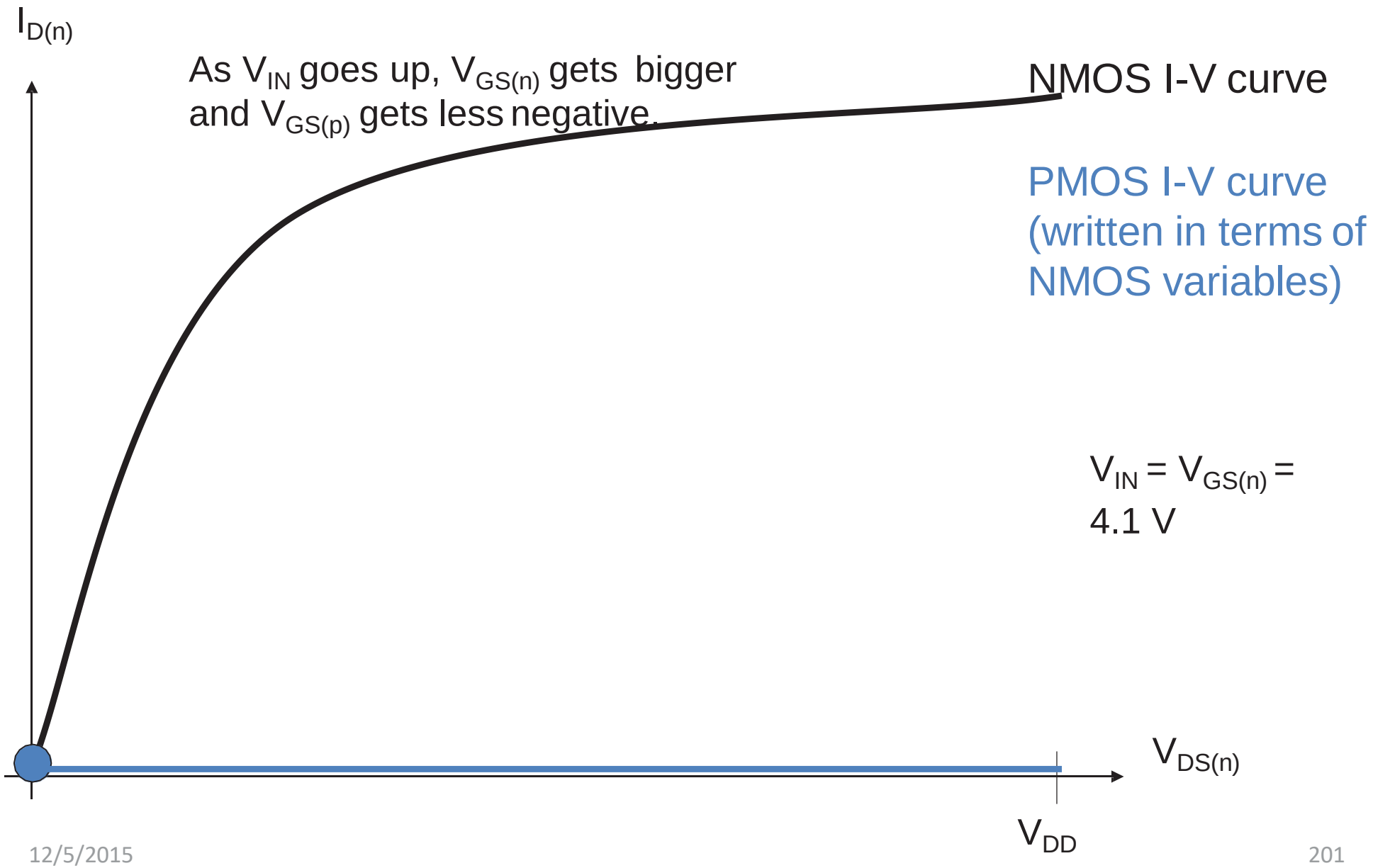
CMOS Analysis



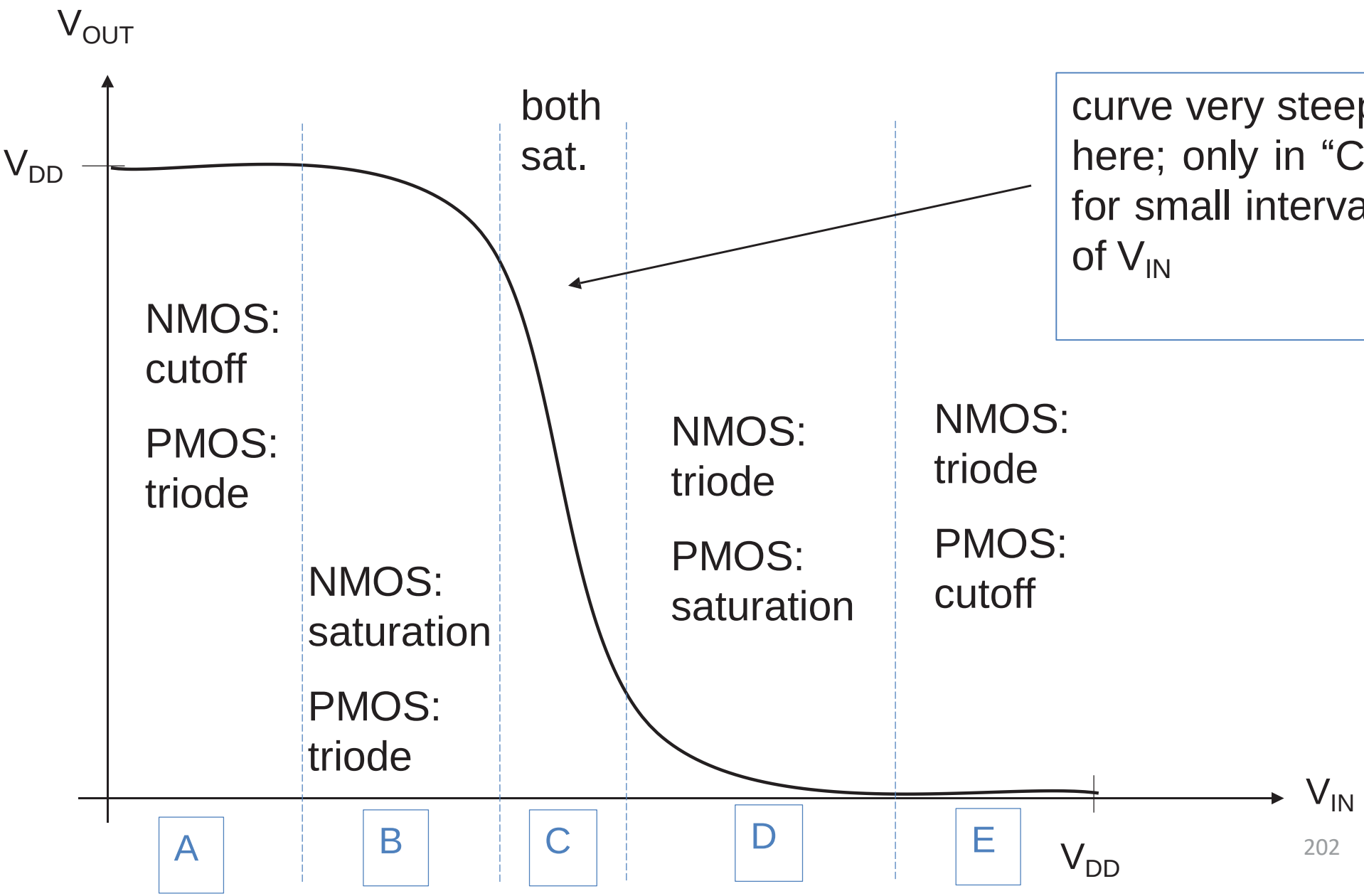
CMOS Analysis



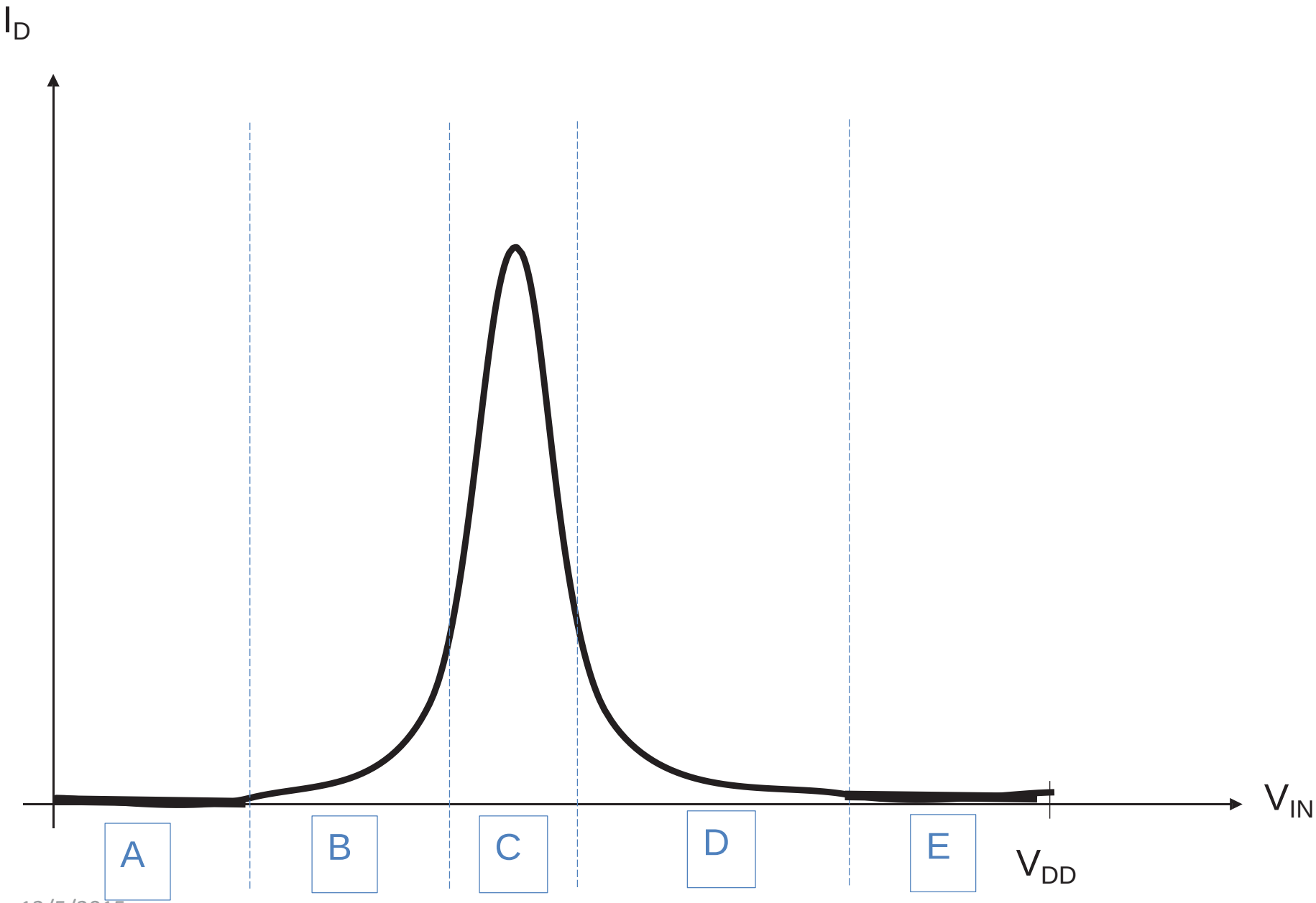
CMOS Analysis



CMOS Inverter VOUT vs. VIN

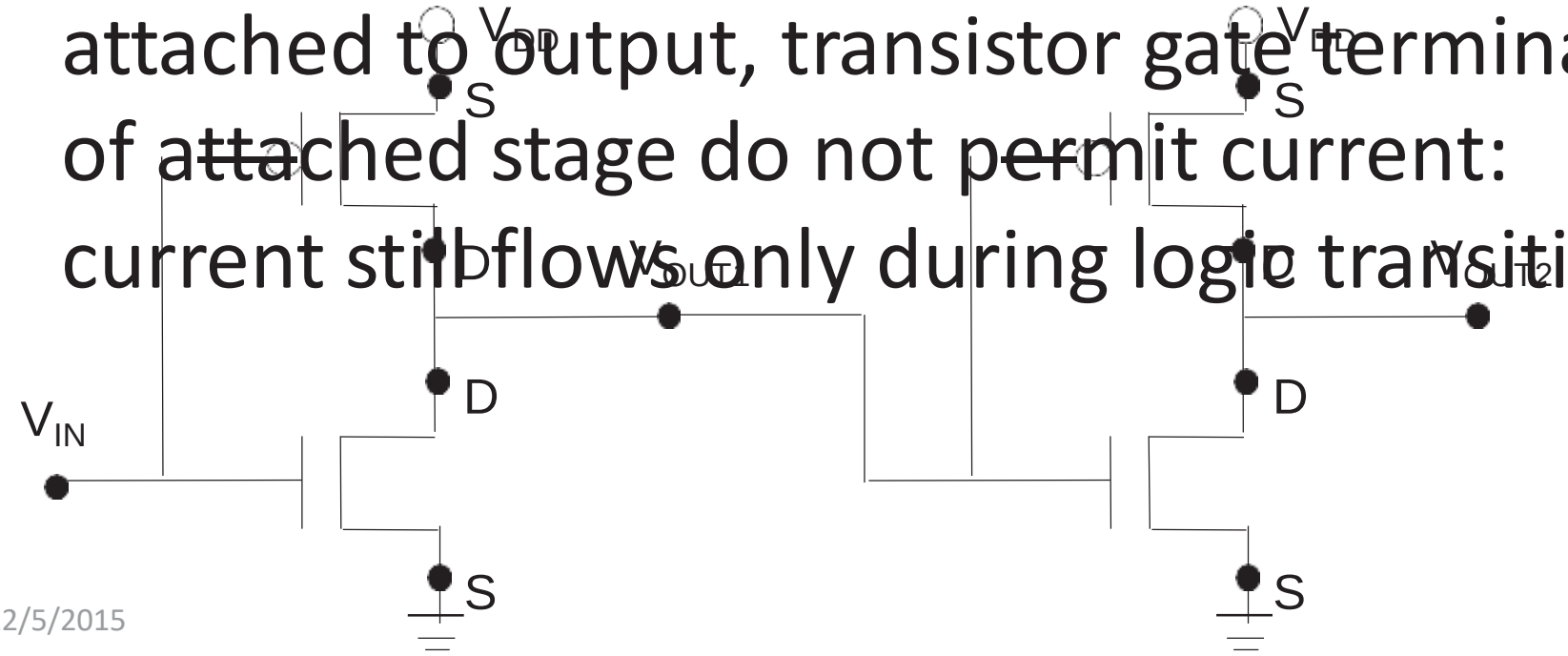


CMOS Inverter ID

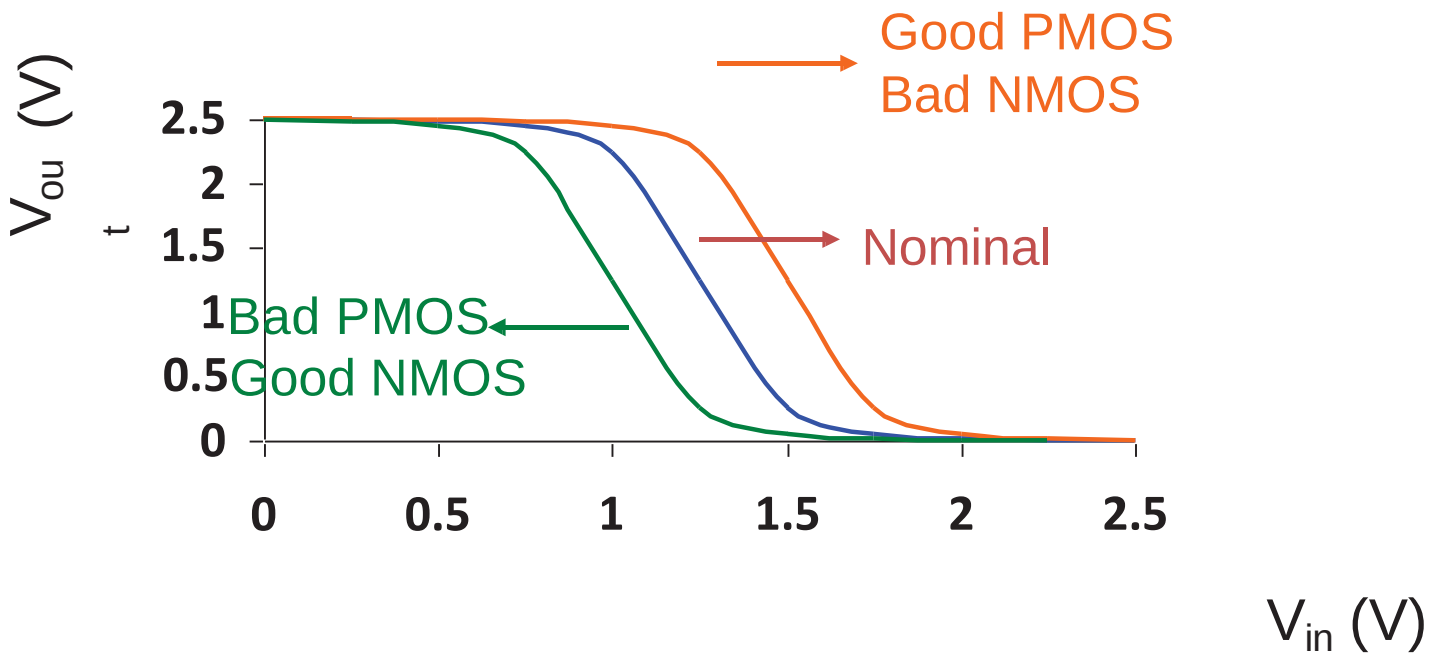


Important Points

- No ID current flow in Regions A and E if nothing attached to output; current flows only during logic transition
- If another inverter (or other CMOS logic) attached to output, transistor gate terminals of attached stage do not permit current: current still flows only during logic transition



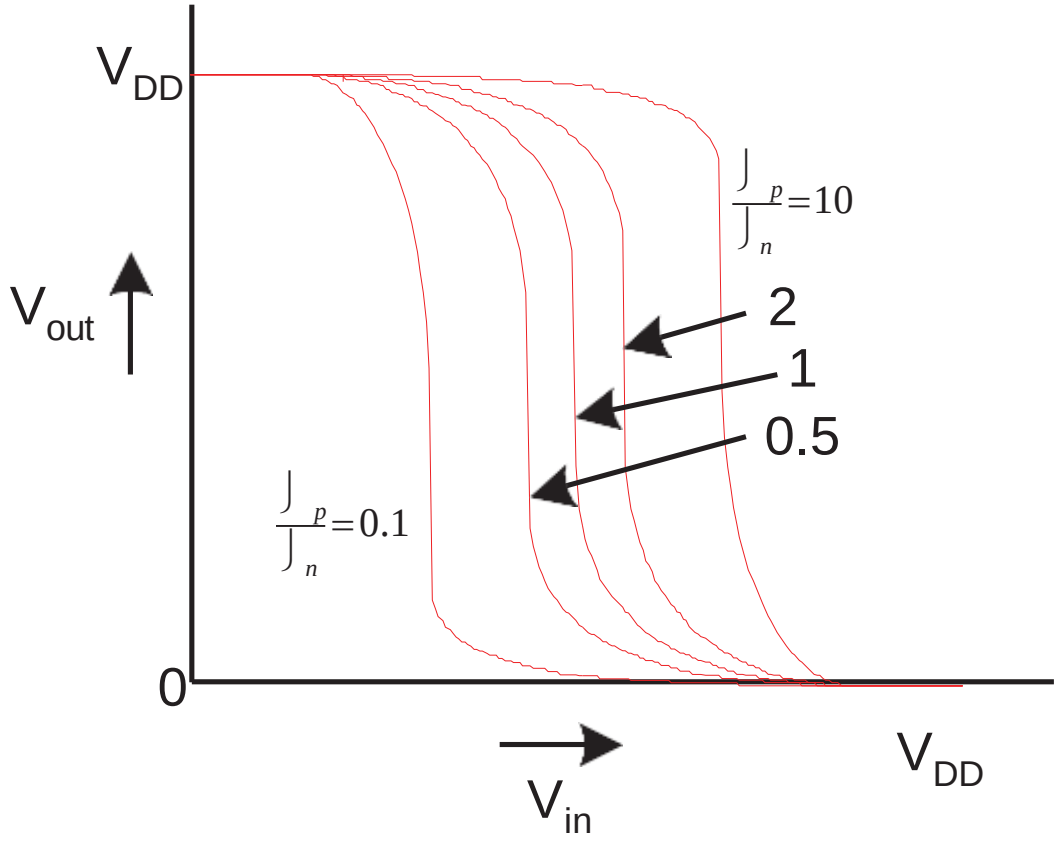
Impact of Process Variation



- Pprocess variations (mostly) cause a shift in the switching threshold

Beta Ratio

- If bp / bn σ 1, switching point will move from VDD/2

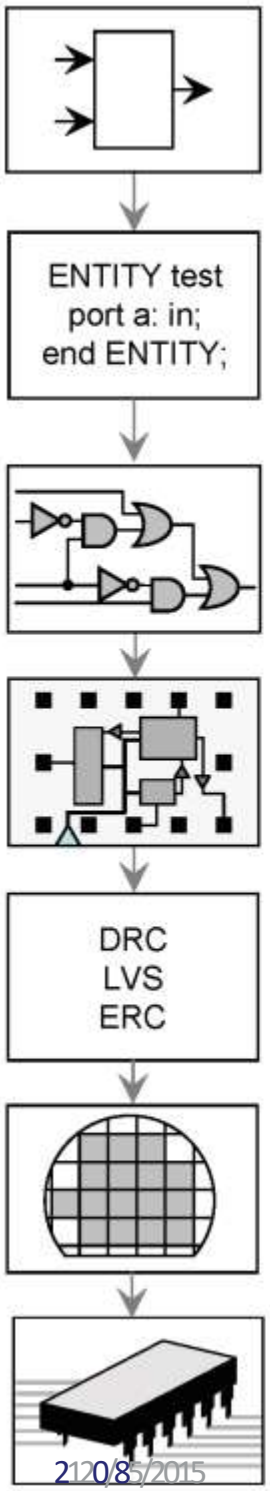


Unit III
VLSI CIRCUIT DESIGN PROCESSES

Topics

- VLSI design flow
- MOS layers
- Stick diagrams
- Design Rules and Layout
- 2 um CMOS design rules for wires
- Contacts and Transistors
- Layout diagrams for NMOS and
- CMOS inverters and gates, Scaling of MOS circuits

VLSI Design of approach of IC



System Specification

Architectural Design

Functional Design and Logic Design

Circuit Design

Physical Design

Physical Verification and Signoff

Fabrication

Packaging and Testing

Chip

- Partitioning
- Floorplanning
- Placement
- Clock Tree Synthesis
- Signal Routing
- Timing Closure

Layer Types

- p-substrate
- n-well
- n+
- p+
- Gate oxide (thin oxide)
- Gate (polycilicon)
- Field Oxide
  - Insulated glass
  - Provide electrical isolation

Stick diagram

Encodings for a simple single metal nMOS process

COLOR	STICK ENCODING MONOCROME	LAYERS	MASK LAYOUT ENCODING MONOCROME	CIF LAYER
GREEN		n diffusion n+active		N D
RED		Thniox Polysilicon		NP
BLUE		Metal 1		N M
BLACK		Contact cut		N C
GRAY	NOT APPLICABLE	Overglass		N G
nMOS ONLY YELLO		Implant		NI
nWMOS ONLY BROWN		Buried contact		NB

Stick Diagrams

Metal

poly

ndiff

pdiff

Can also draw in shades of gray/line style.

Buried Contact

Contact Cut

Stick Diagrams

- VLSI design aims to translate circuit concepts onto silicon.
- Stick diagrams are a means of capturing topography and layer information using simple diagrams.
- Stick diagrams convey layer information through colour codes (or monochrome encoding).
- Acts as an interface between symbolic circuit and the actual layout.

Stick Diagrams

- Does show all components/vias.
- It shows relative placement of components.
- Goes one step closer to the layout
- Helps plan the layout and routing

Stick Diagrams

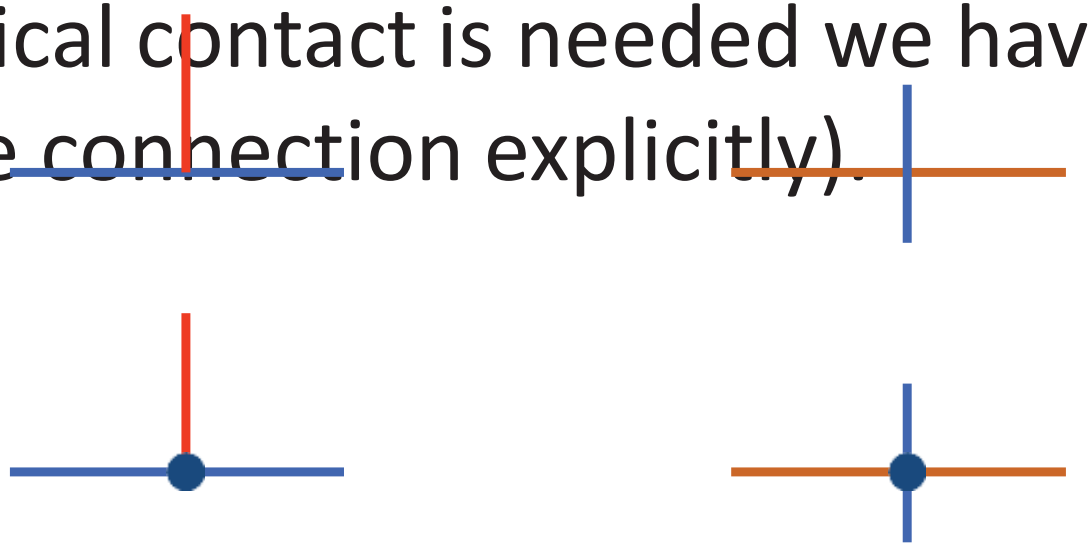
- Does ***not*** show
  - Exact placement of components
  - Transistor sizes
  - Wire lengths, wire widths, tub boundaries.
  - Any other low level details such as parasitics..

Stick Diagrams – Some rules

- Rule 1.
- When two or more ‘sticks’ of the same type cross or touch each other that represents electrical contact.

Stick Diagrams – Some rules

- Rule 2.
- When two or more ‘sticks’ of different type cross or touch each other there is no electrical contact.
  (If electrical contact is needed we have to show the connection explicitly).



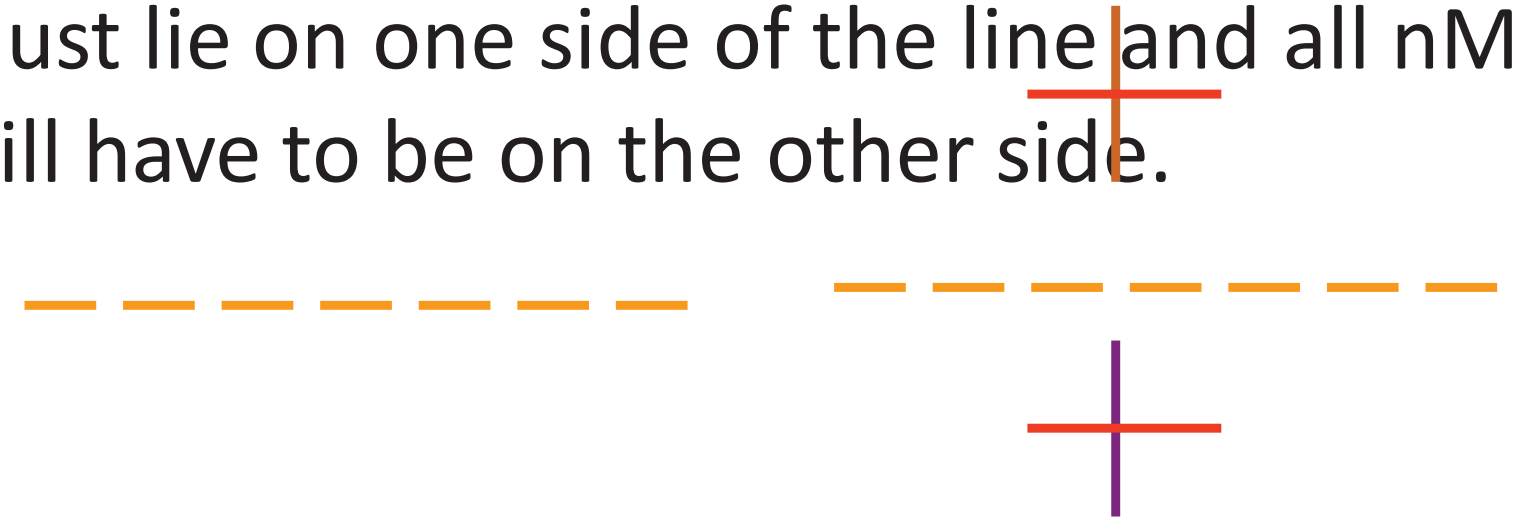
Stick Diagrams – Some rules

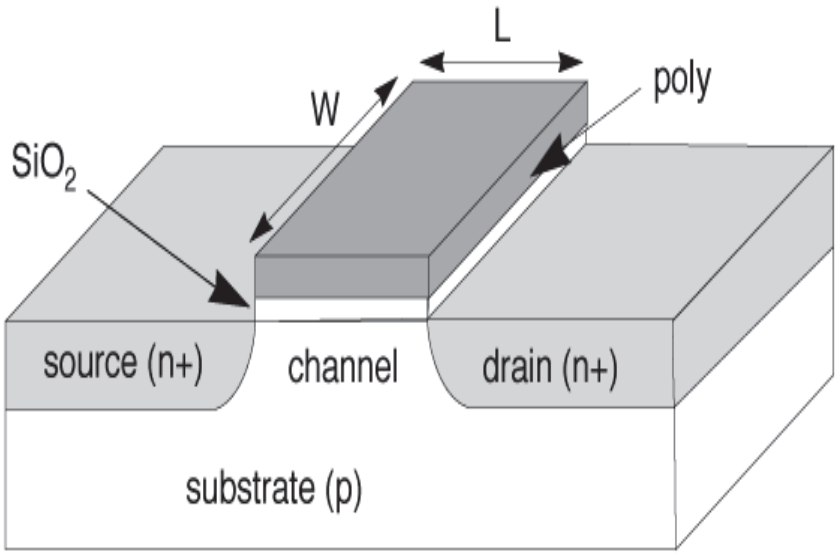
- Rule 3.
- When a poly crosses diffusion it represents a transistor.

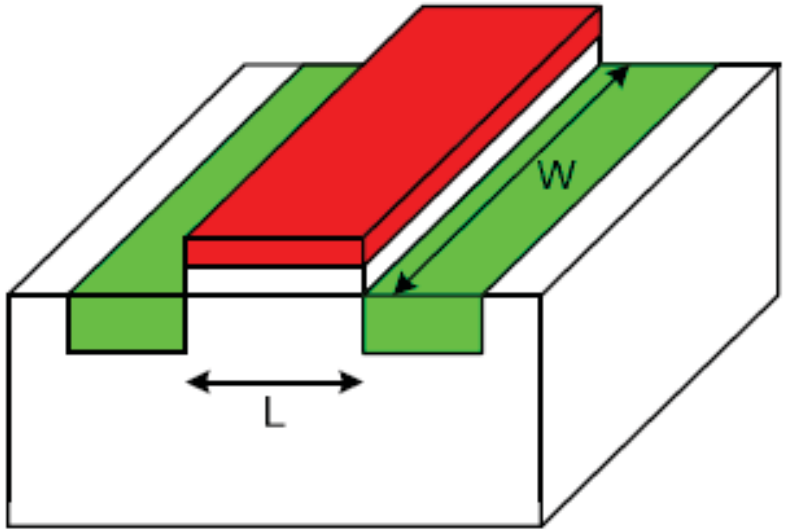
Note: If a contact is shown then it is ***not*** a transistor.

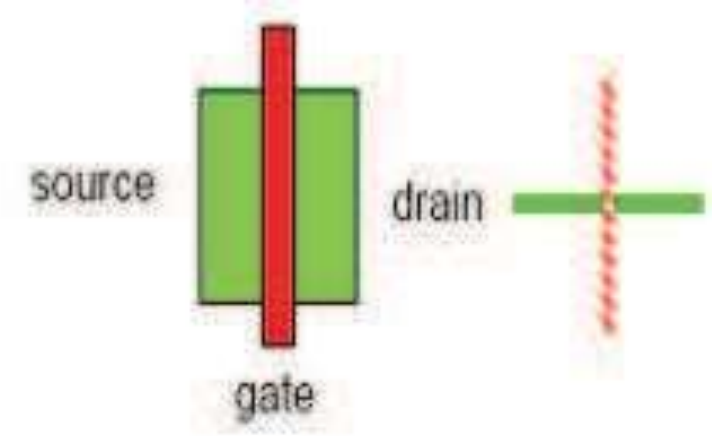
Stick Diagrams – Some rules

- Rule 4.
- In CMOS a demarcation line is drawn to avoid touching of p-diff with n-diff. All pMOS must lie on one side of the line and all nMOS will have to be on the other side.



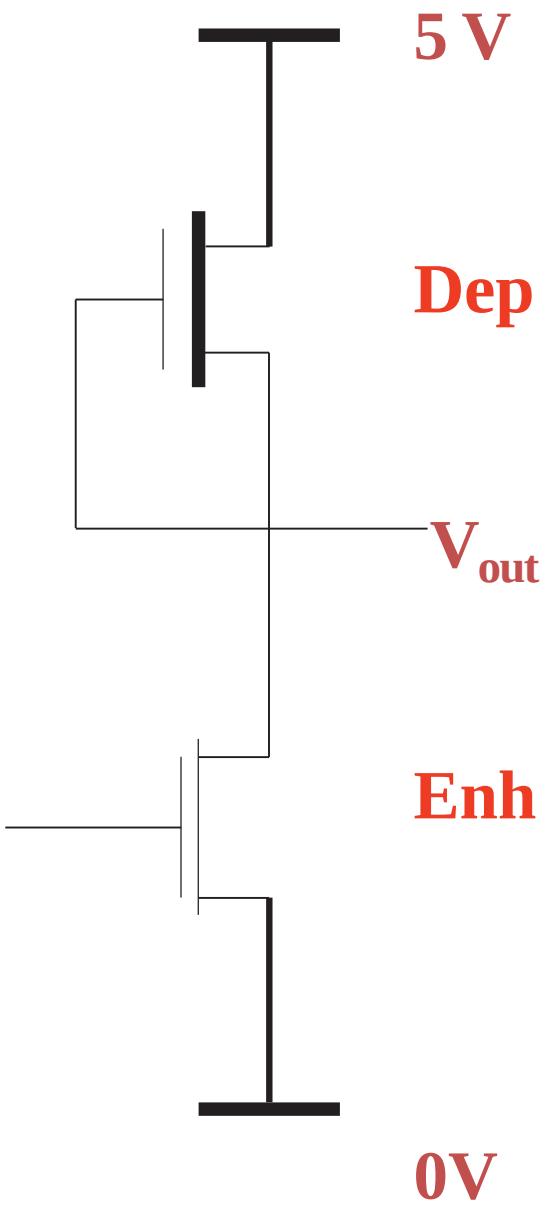


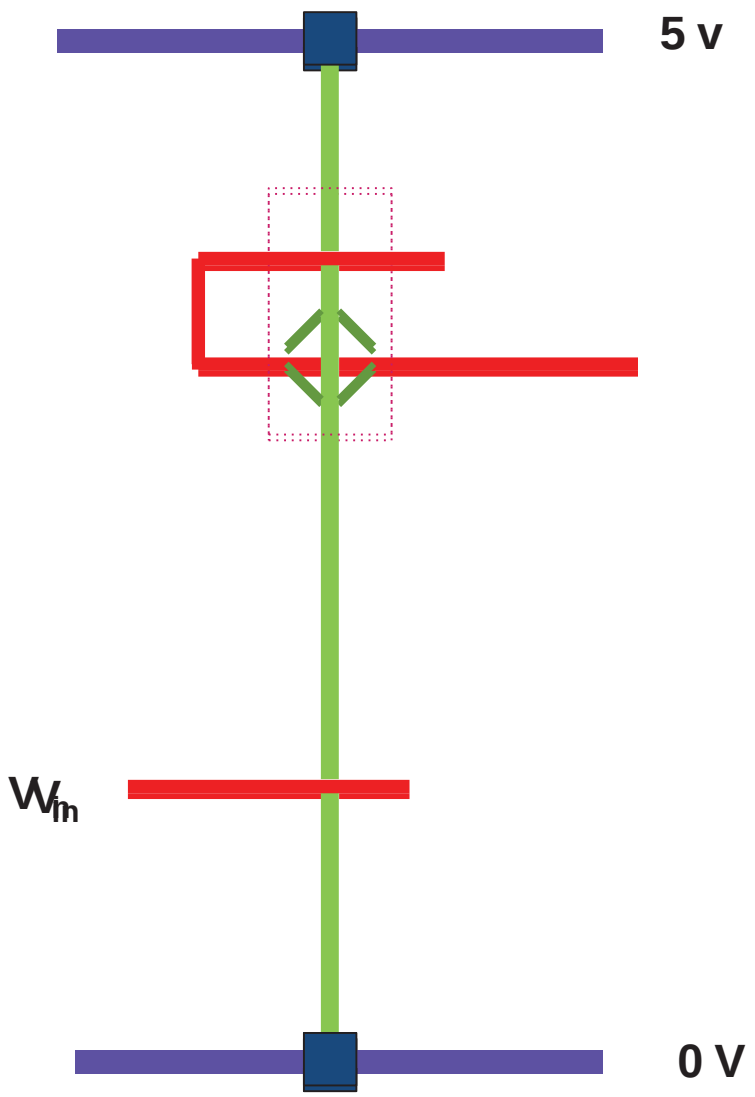




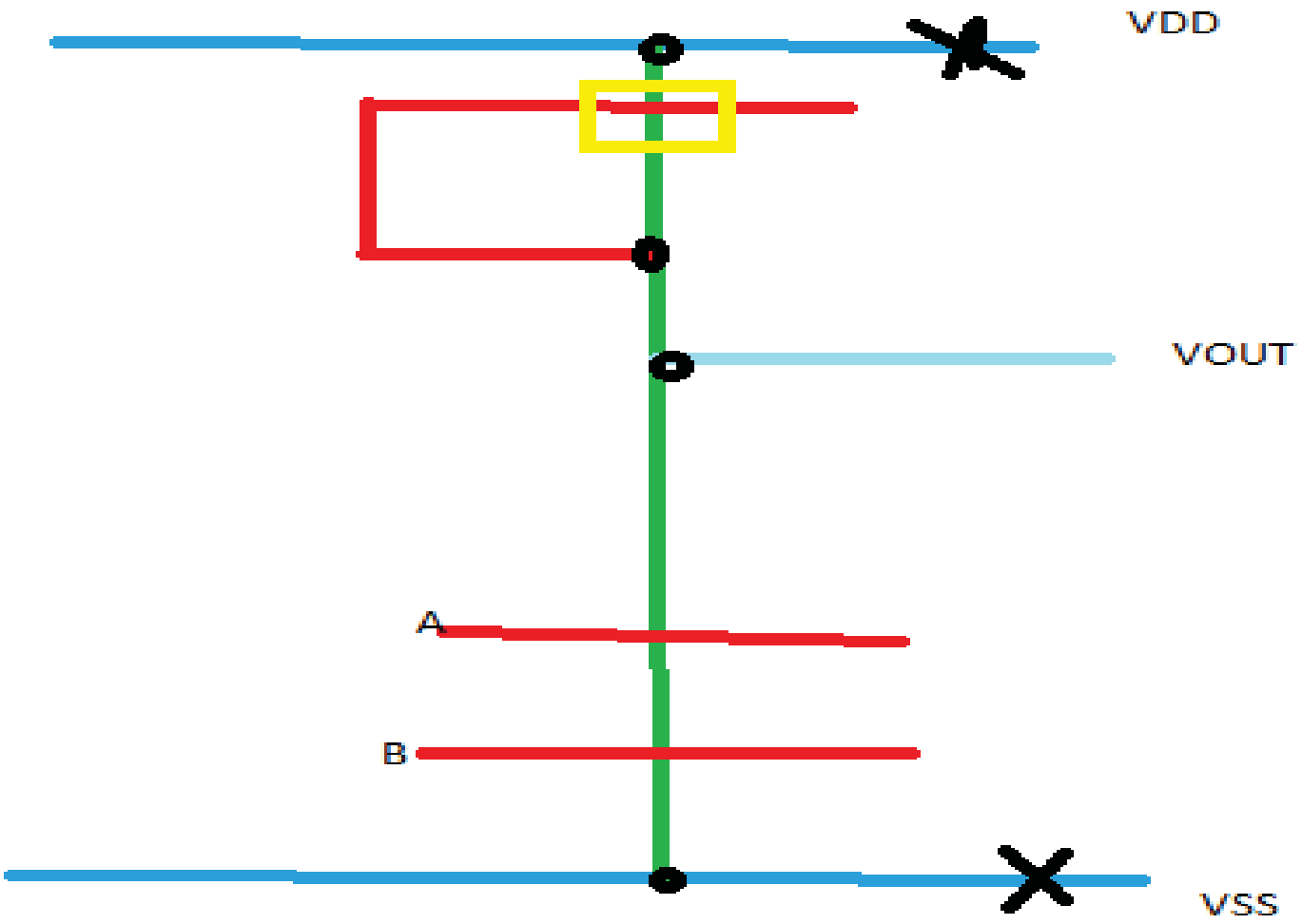
Red (Poly) crosses over Green (Active)

NMOS INVERTER

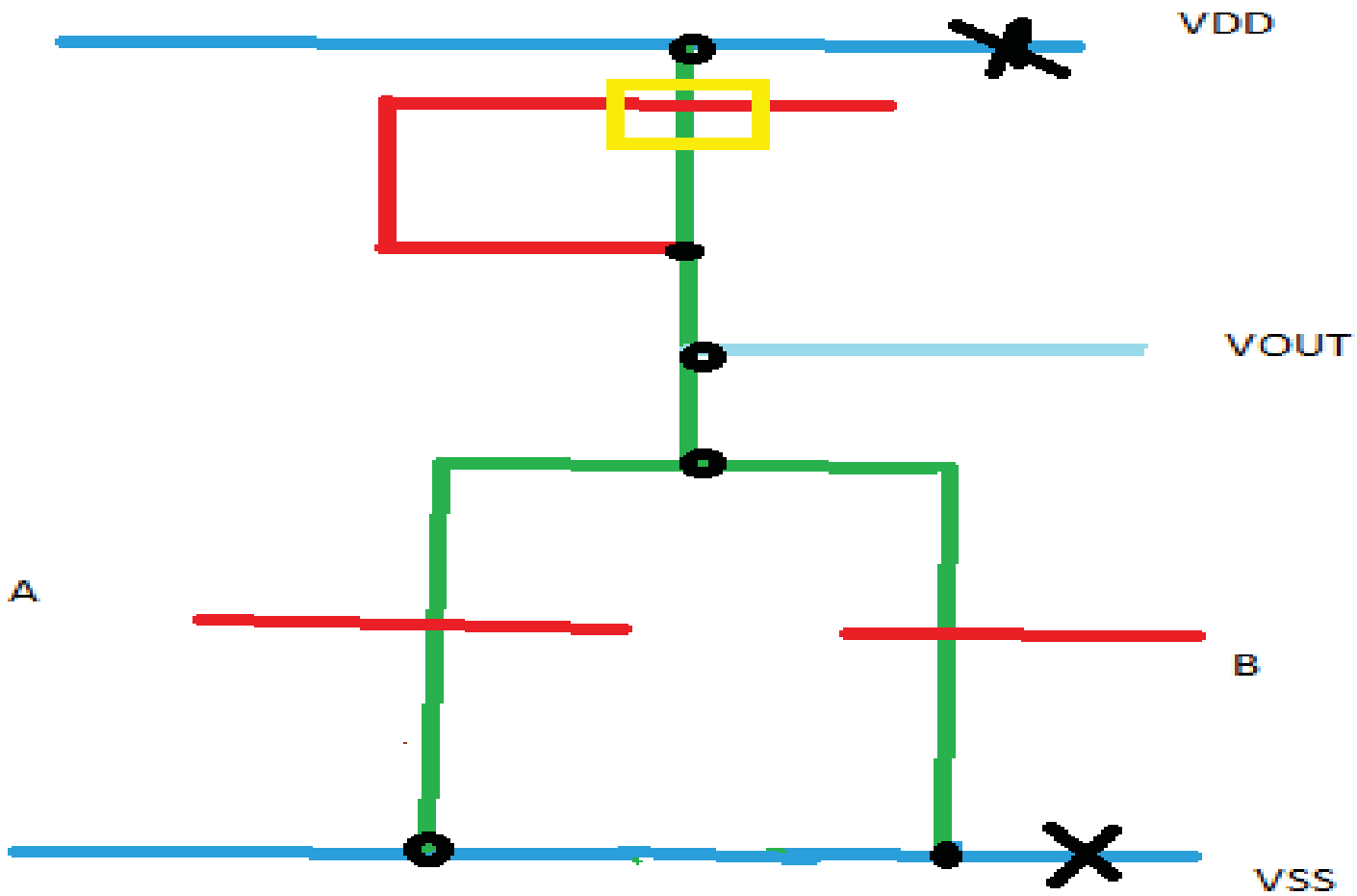




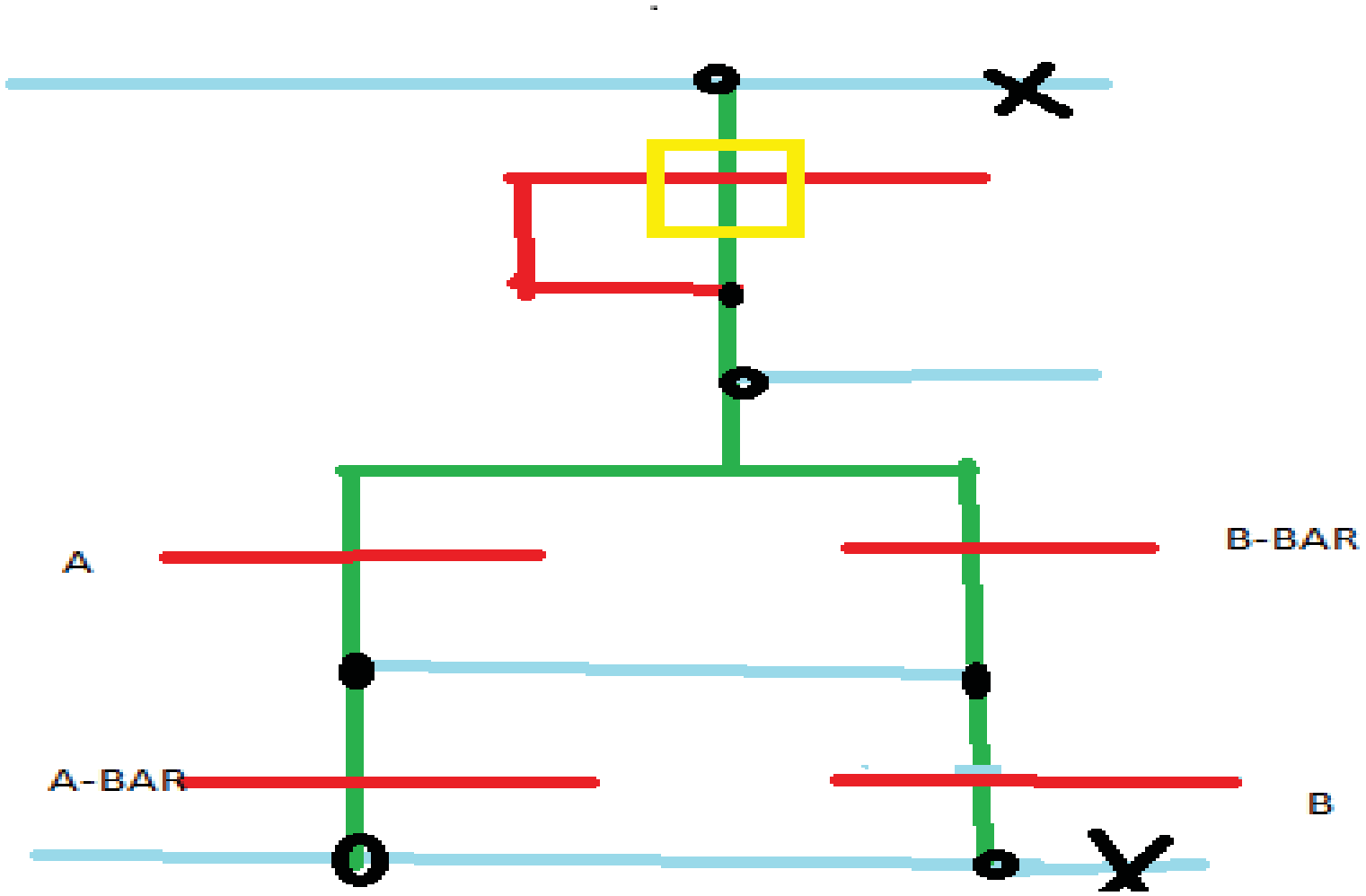
NMOS-NAND



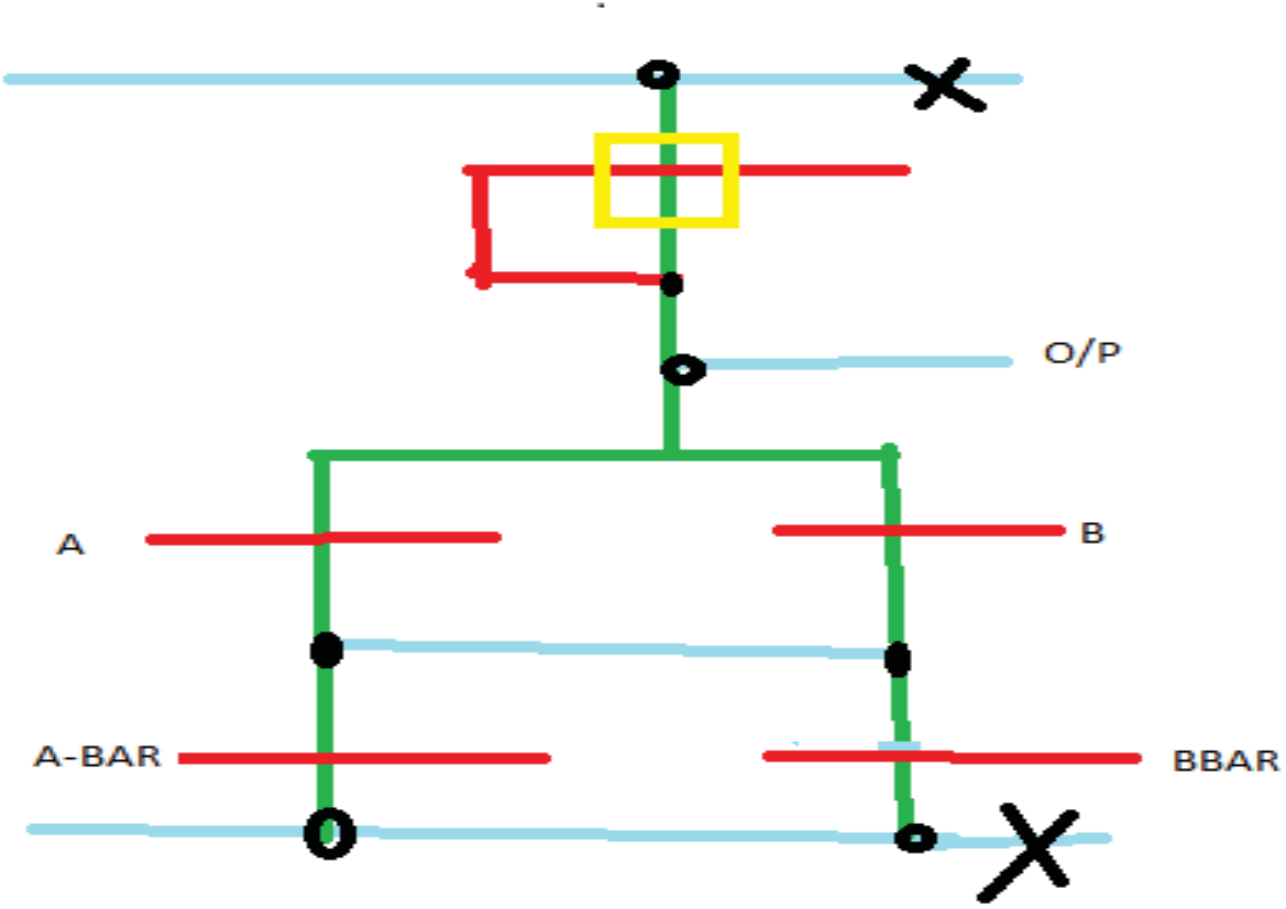
NMOS-NOR



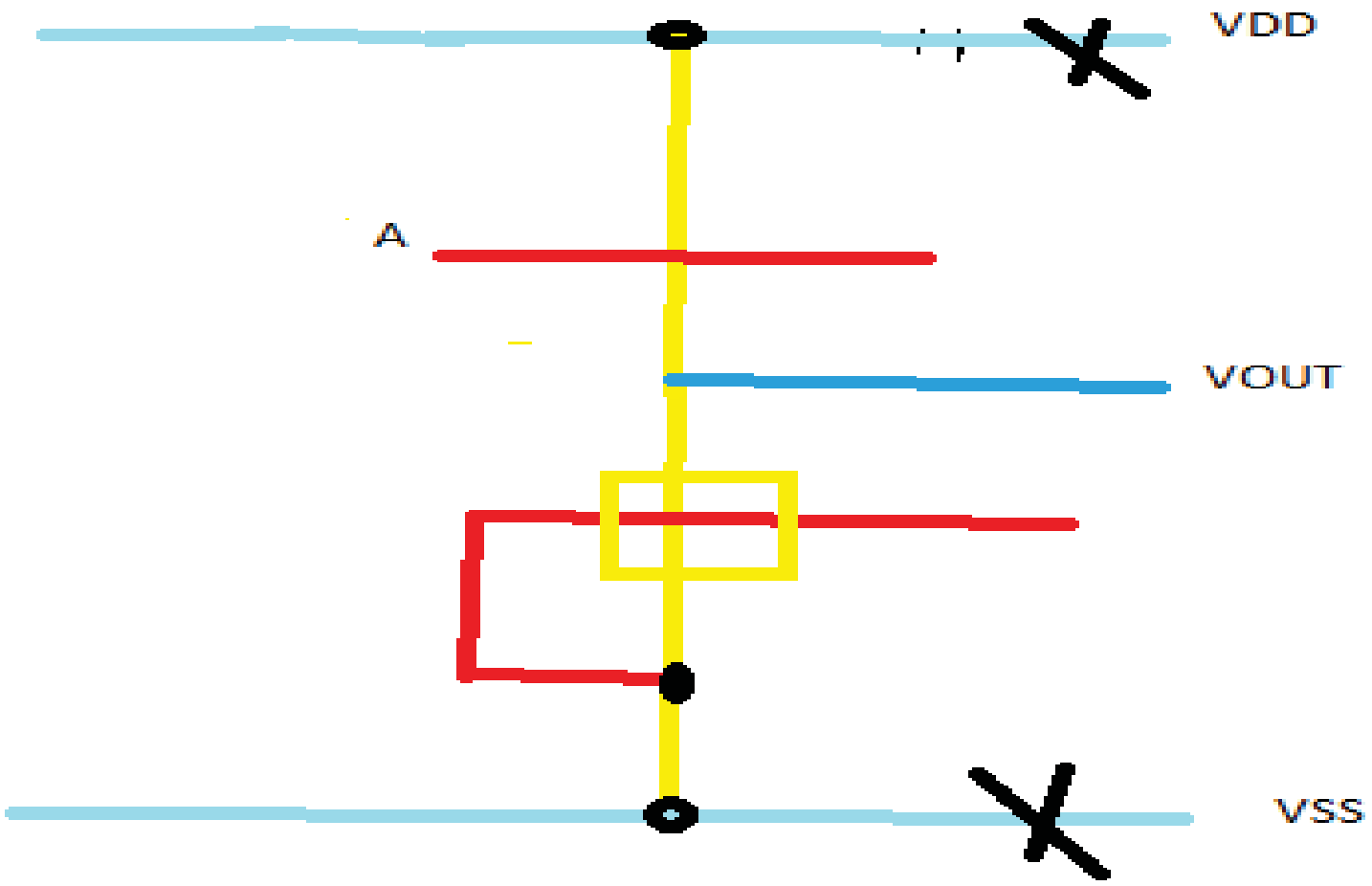
NMOS EX-OR



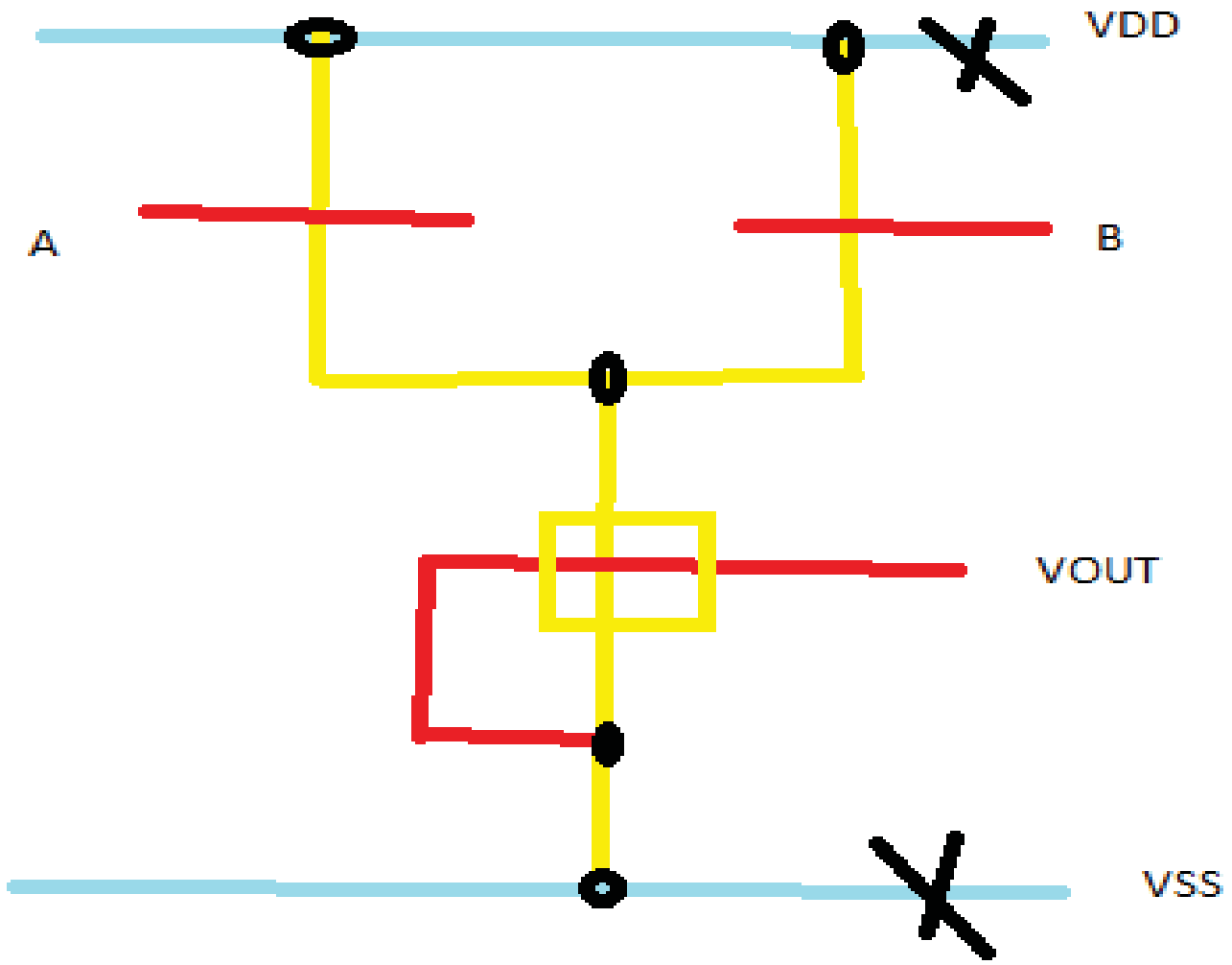
NMOS EX-NOR



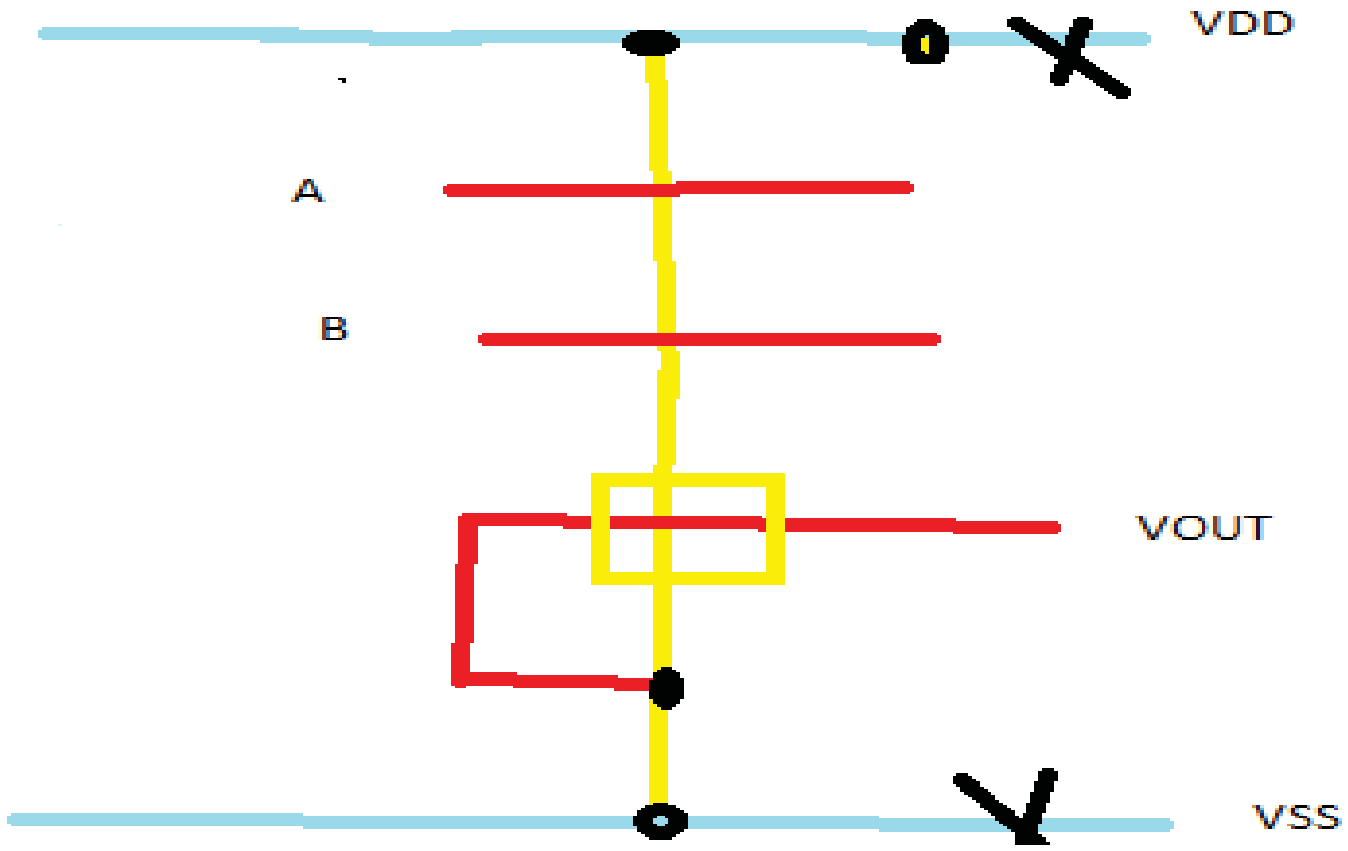
PMOS-INVERTER



PMOS NAND

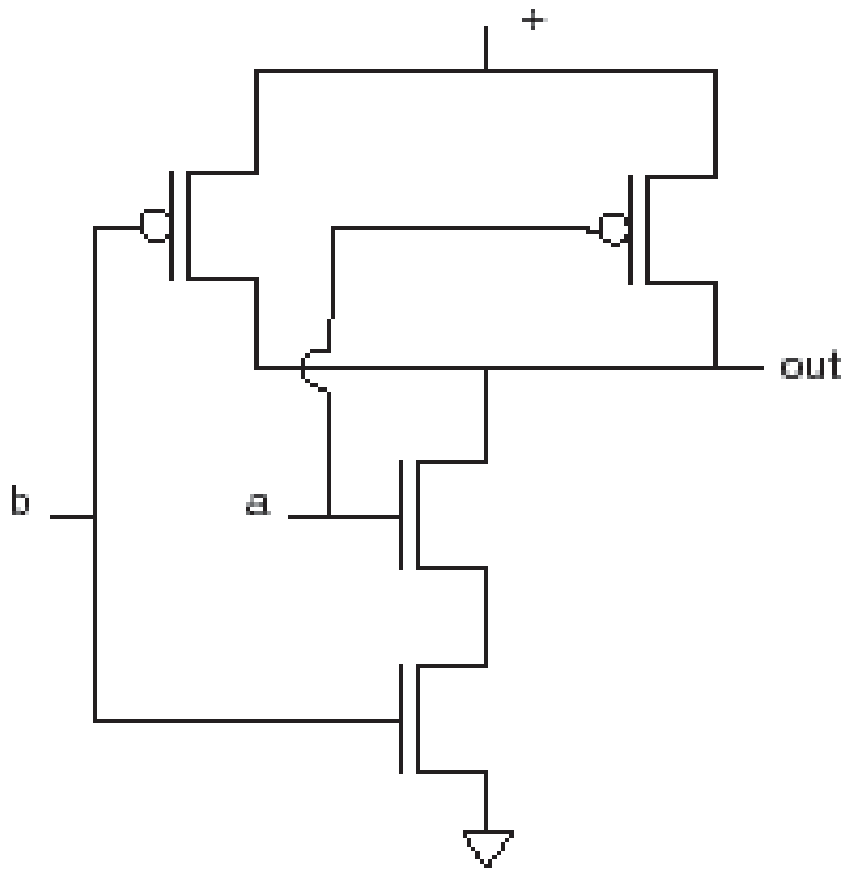


PMOS-NOR

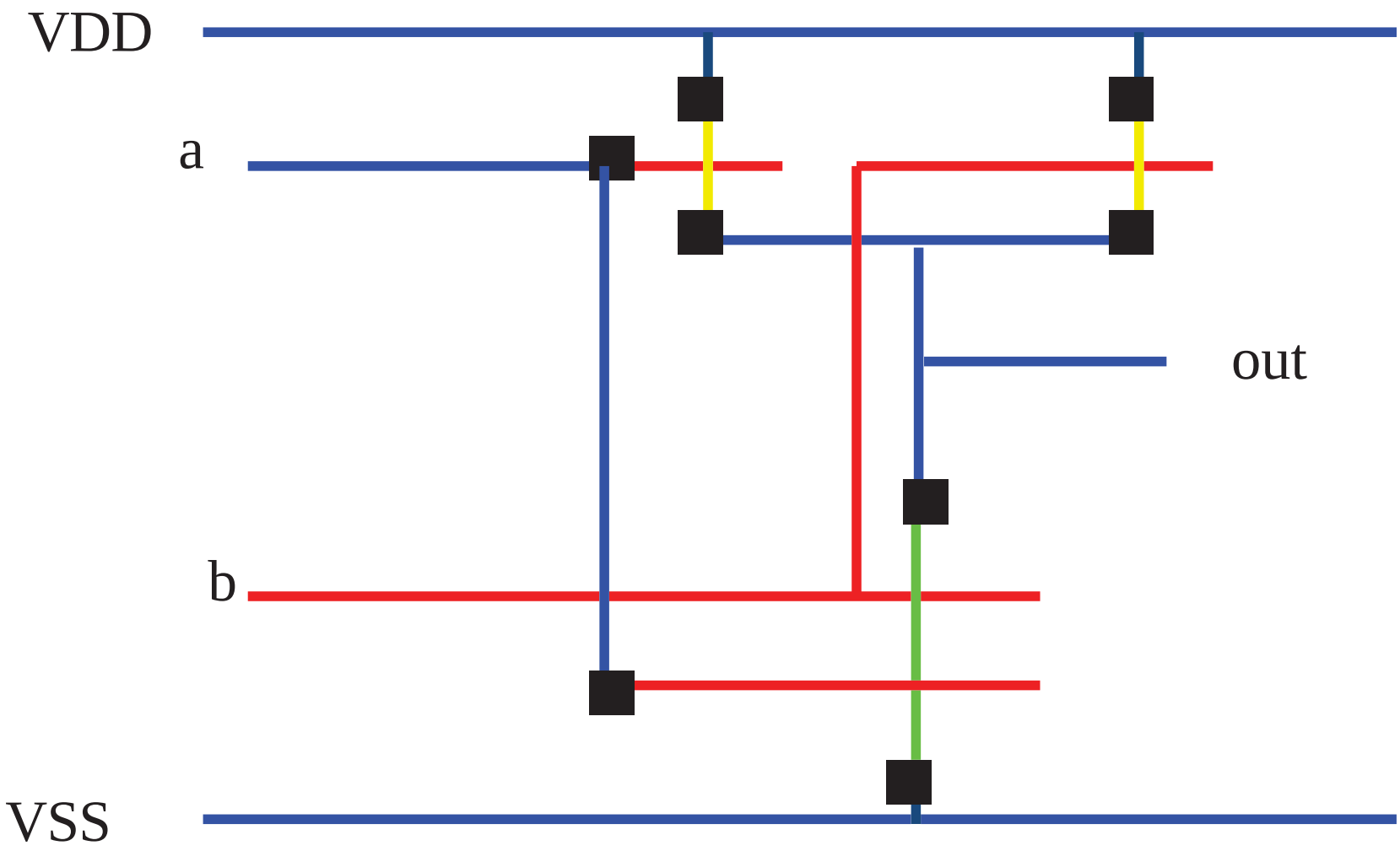


Sticks design CMOS NAND:

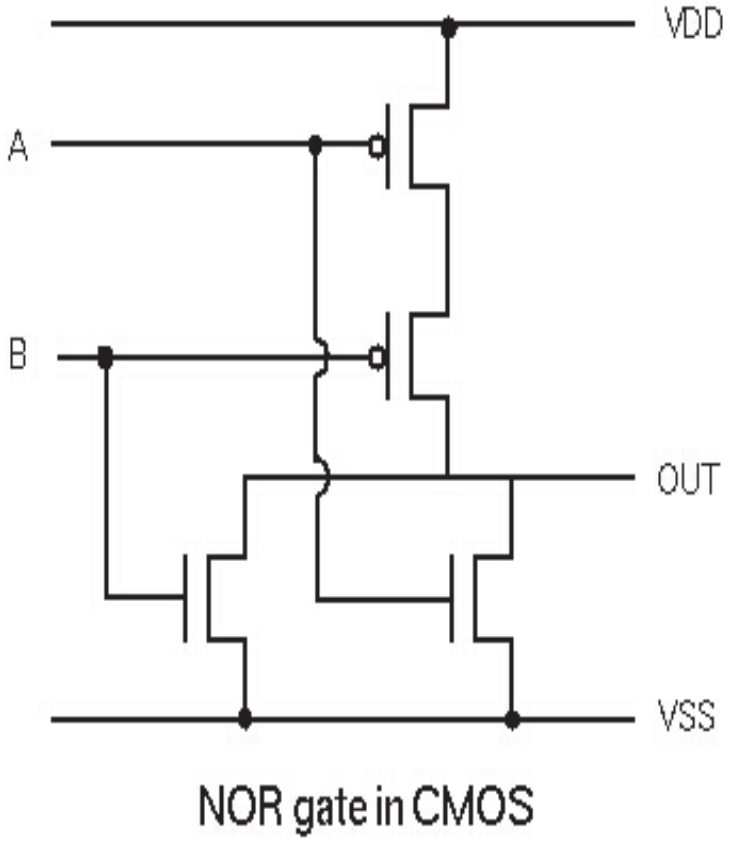
- Start with NAND gate:



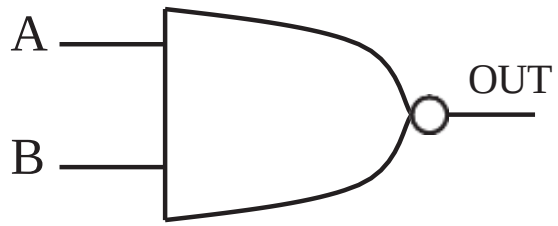
NAND sticks



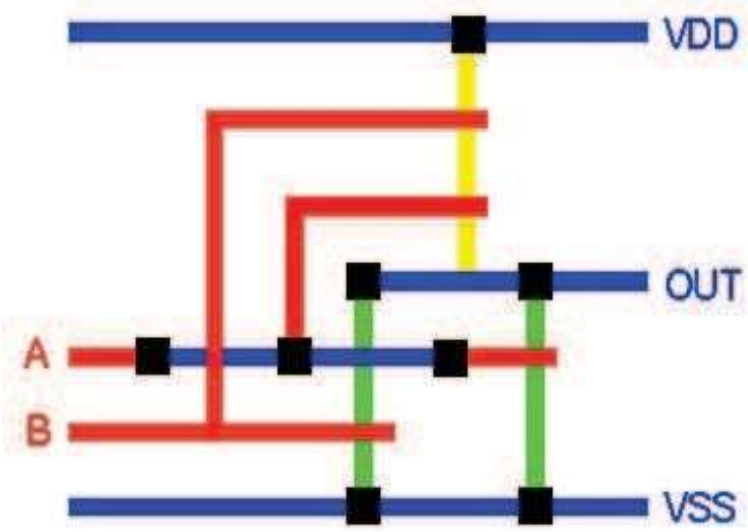
Stick Diagram - Example



NOR gate in CMOS

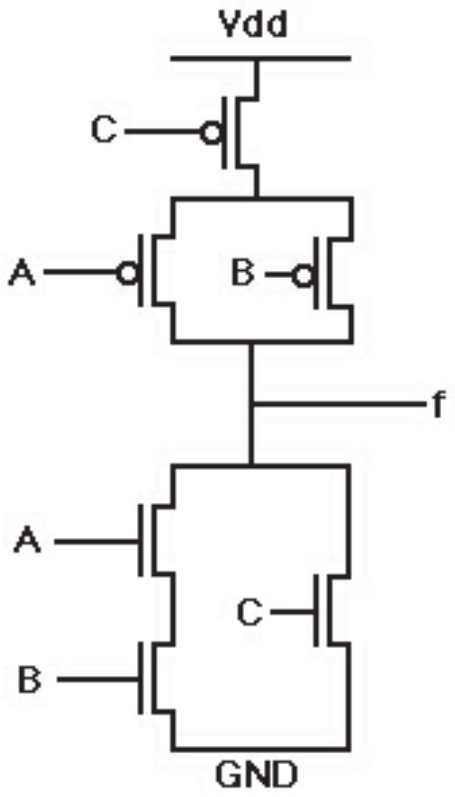


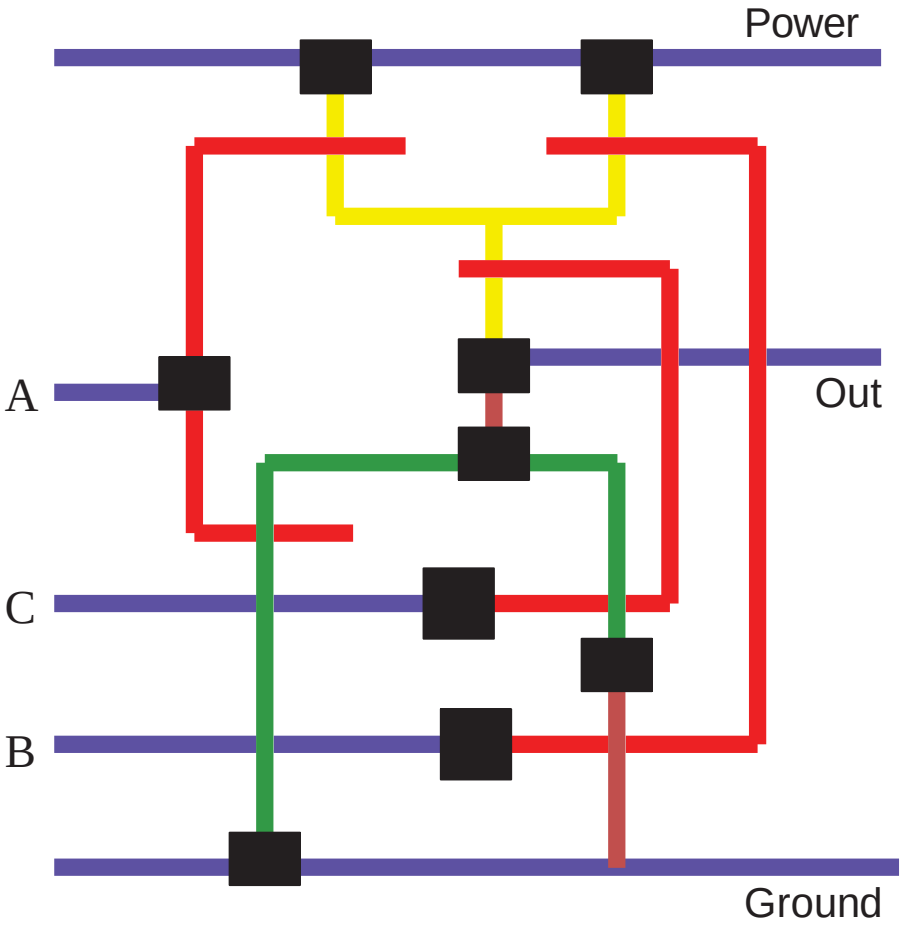
NOR Gate



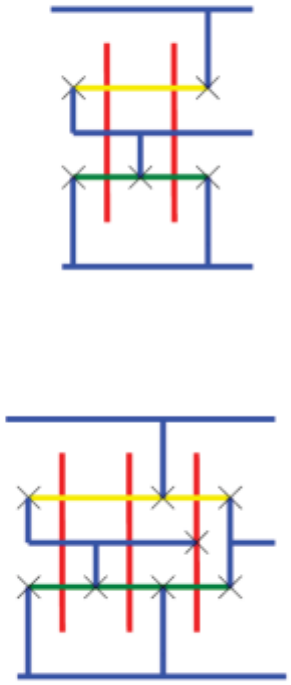
Stick Diagram - Example

Example: $f = \overline{(A \bullet B) + C}$

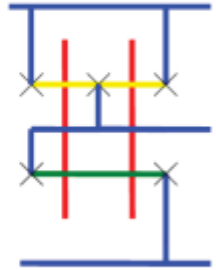


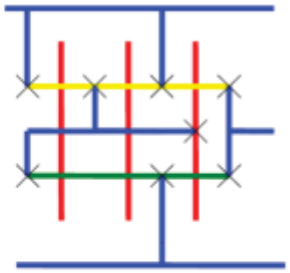


2 I/P OR GATE

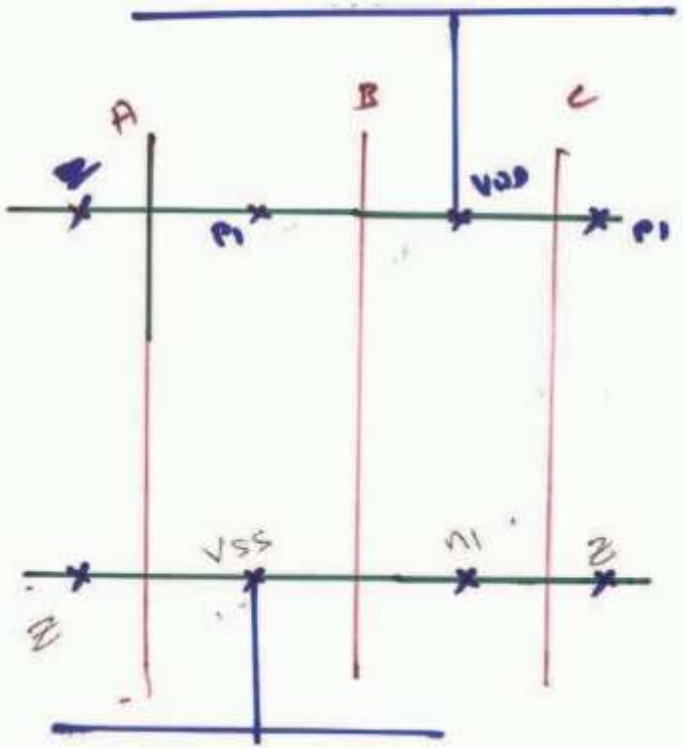


2 I/P AND

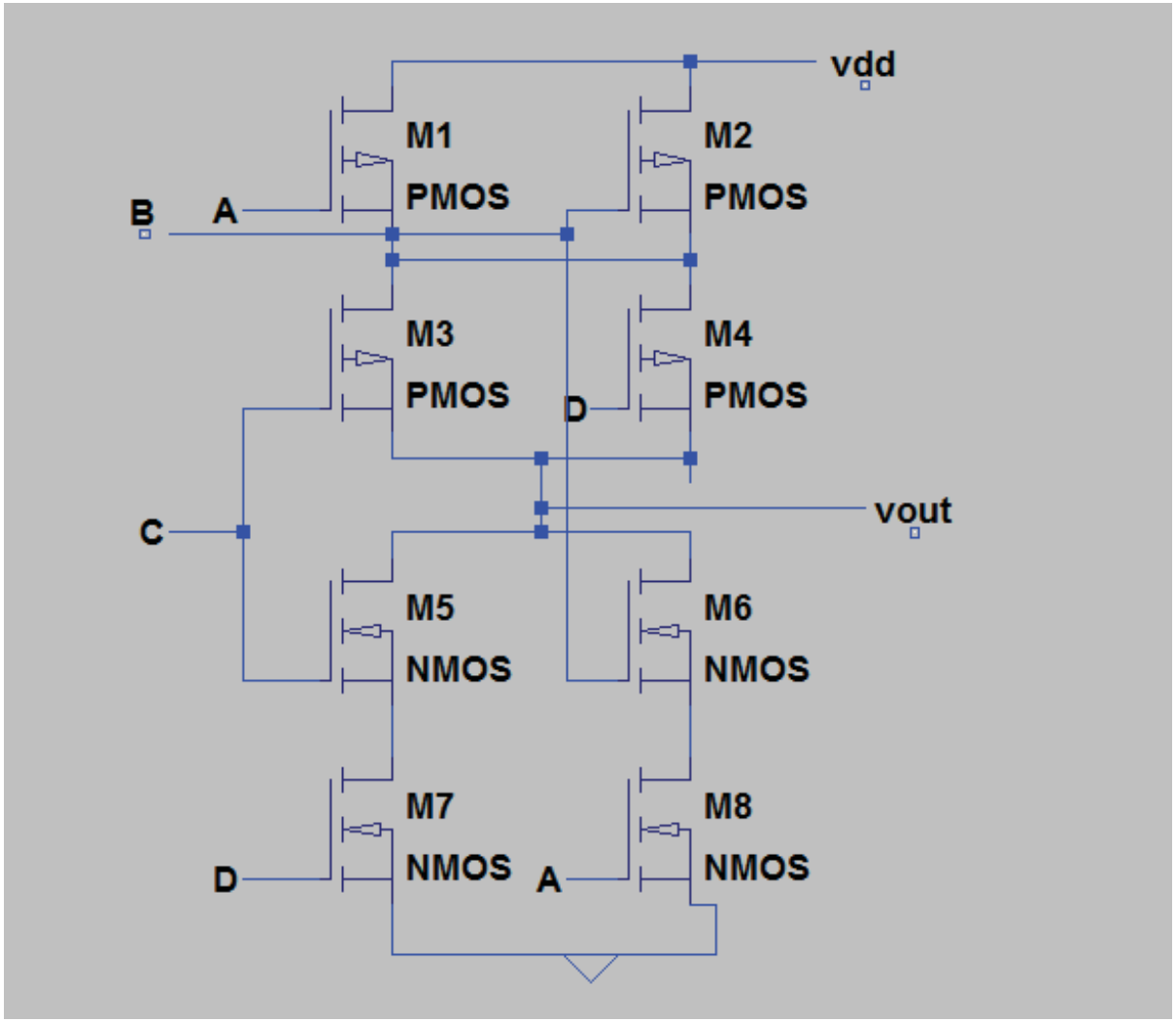




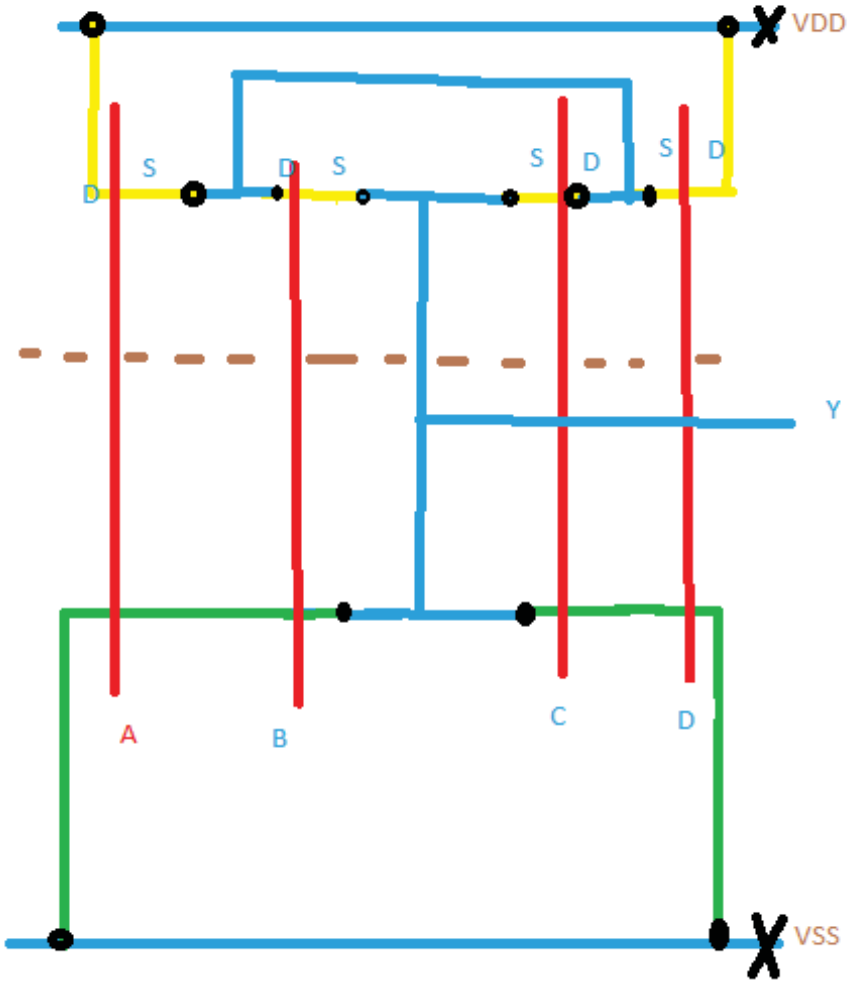
$$Z = \overline{A + BC}$$

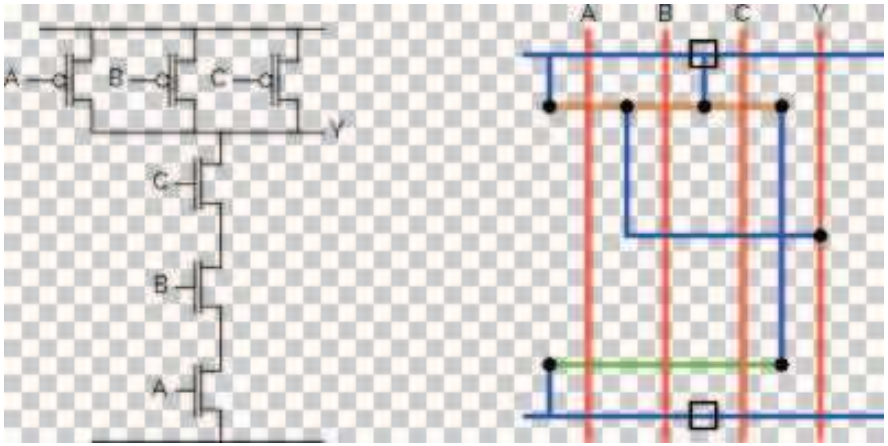


Y=(AB+CD)’


vdd
M1
PMOS
M2
PMOS
B
A
M3
PMOS
M4
PMOS
D
vout
C
M5
NMOS
M6
NMOS
M7
NMOS
M8
NMOS
D
A

Y=(AB+CD)’ STICK




A
B
C
Y
C
B
A
A
B
C
Y

Design Rules

- Design rules are a set of geometrical specifications that dictate the design of the layout masks
- A design rule set provides numerical values
  - For minimum dimensions
  - For minimum line spacings
- Design rules must be followed to insure functional structures on the fabricated chip
- Design rules change with technological advances (www.mosis.org)

Silicon Foundry

- A standard
- A foundry allows designers to submit designs using a state-of-the-art process
- Each foundry state simpler set of design rules called lambda design rules
- All widths, spacings, and distances are written in the form
  - Value = mZ
  - TSMC (Thailand Semiconductor Manufacturing Corporation)

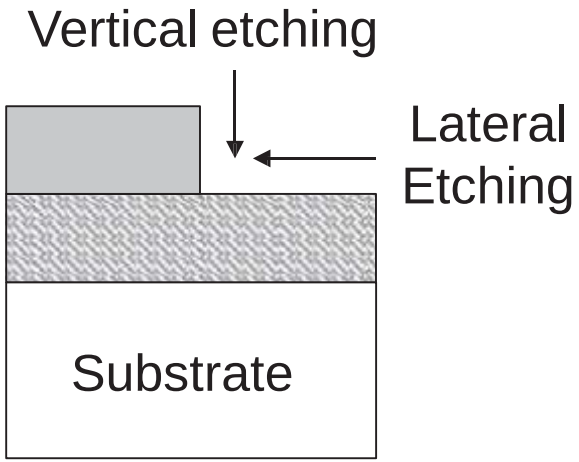
Design Rules Classification

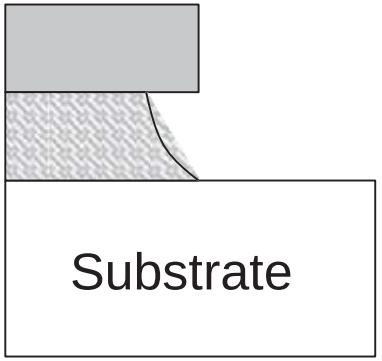
- Minimum width
- Minimum spacing
- Surround
- Extension

Physical Limitations

- Line width limitation of an imaging system
  - The reticle shadow projected on the photoresist does not have sharp edges due to optical diffraction
- Etching process problem
  - Undercutting of the resist due to lateral etching decreases the resolution

Etching Process Problem

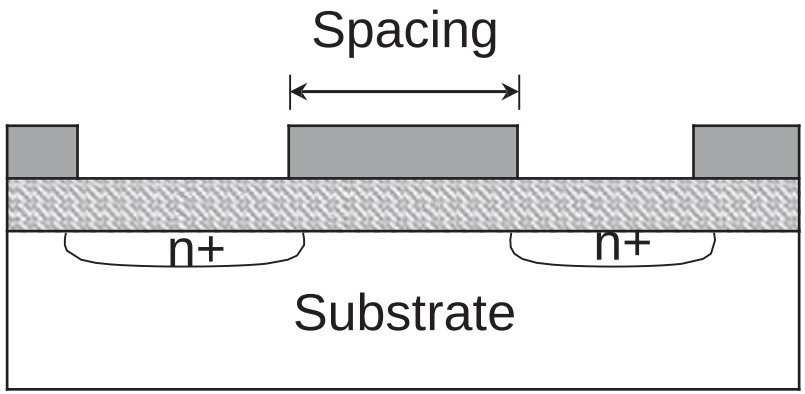




Isotropic etch

Depletion Region

- If depletion regions of adjacent pn junctions touch, then
  - The current blocking characteristics are altered
  - Current can flow between the two



Limits on n+ spacing

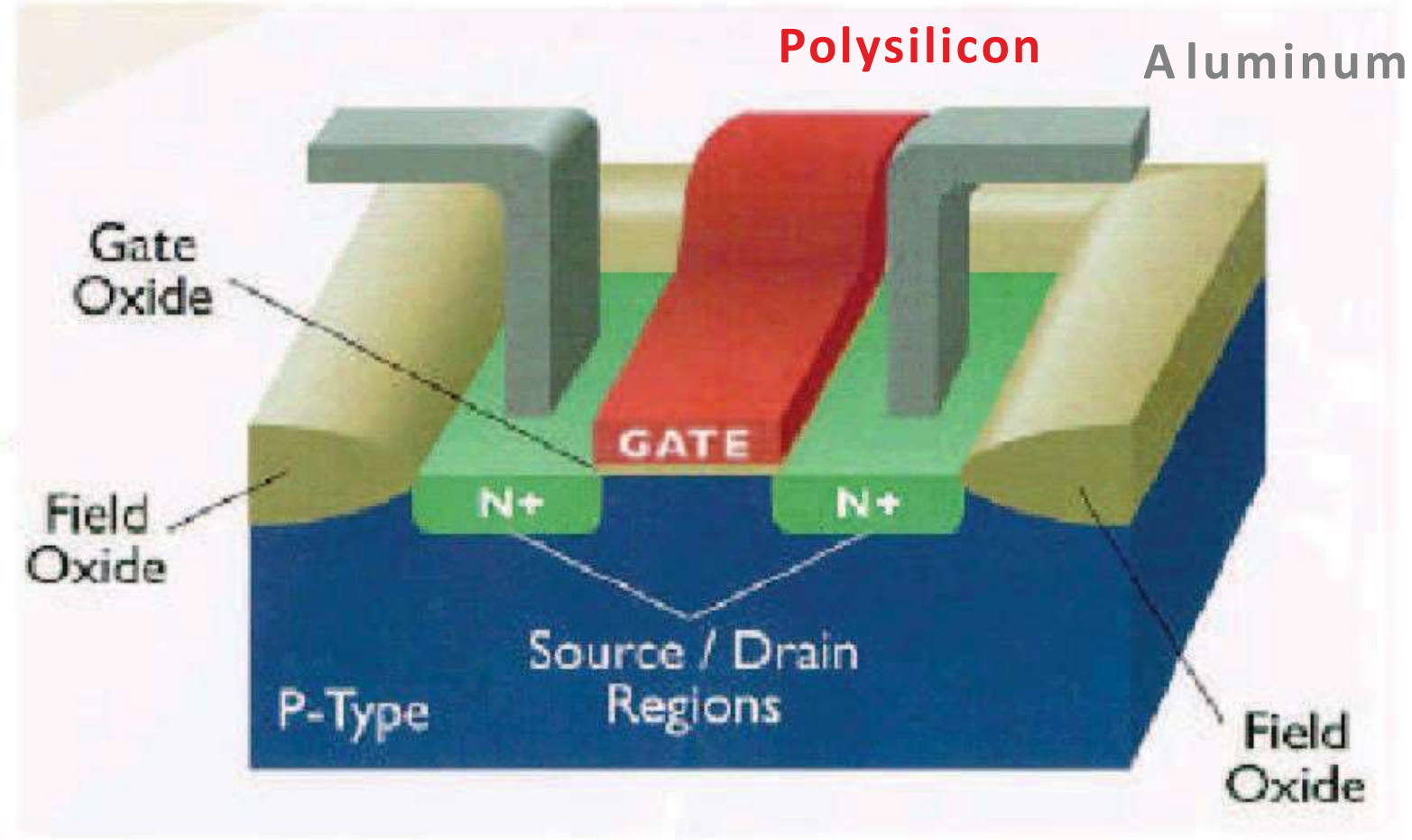
Electrical Capacitive Coupling

- This occurs between closely spaced conducting lines
- This leads to a problem called crosstalk
  - A portion of the electrical energy is coupled to another causing noise
  - This is a major problem in high-density design

Electrical Rules

- An example of an electrical rule is the allowed width of a metal interconnect line
  - To avoid electromigration effects
  - The design rule set will stipulate the maximum current flow level permitted

3D Perspective



Design rules and Layout

- Why we use design rules?
  - Interface between designer and process engineer
  - Guidelines for constructing process masks

Design Rules

Minimum length or width of a feature on a layer is **2Z**

Why?

To allow for shape contraction

Minimum separation of features on a layer is 2Z

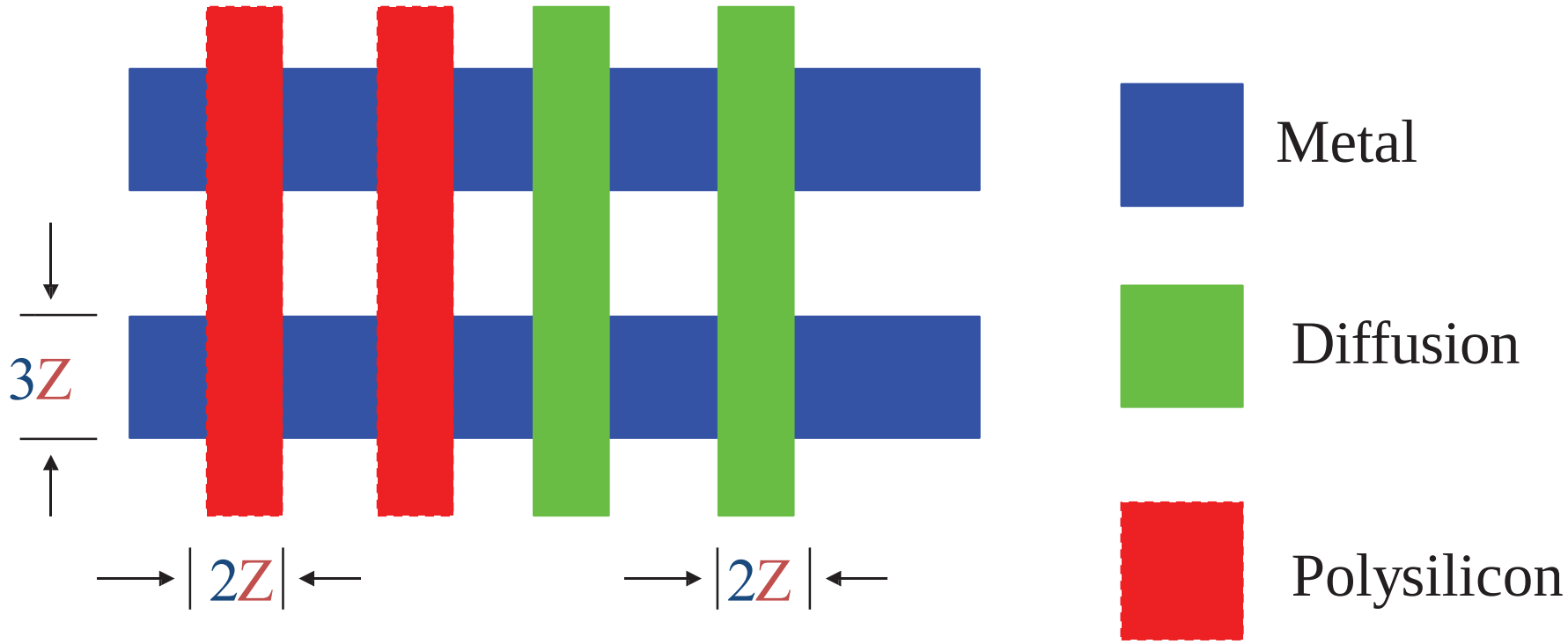
Why?

To ensure adequate continuity of the intervening materials.

Design Rules

Minimum width of PolySi and diffusion line **2**Z

Minimum width of Metal line **3**Z as metal lines run over a more uneven surface than other conducting layers to ensure their continuity

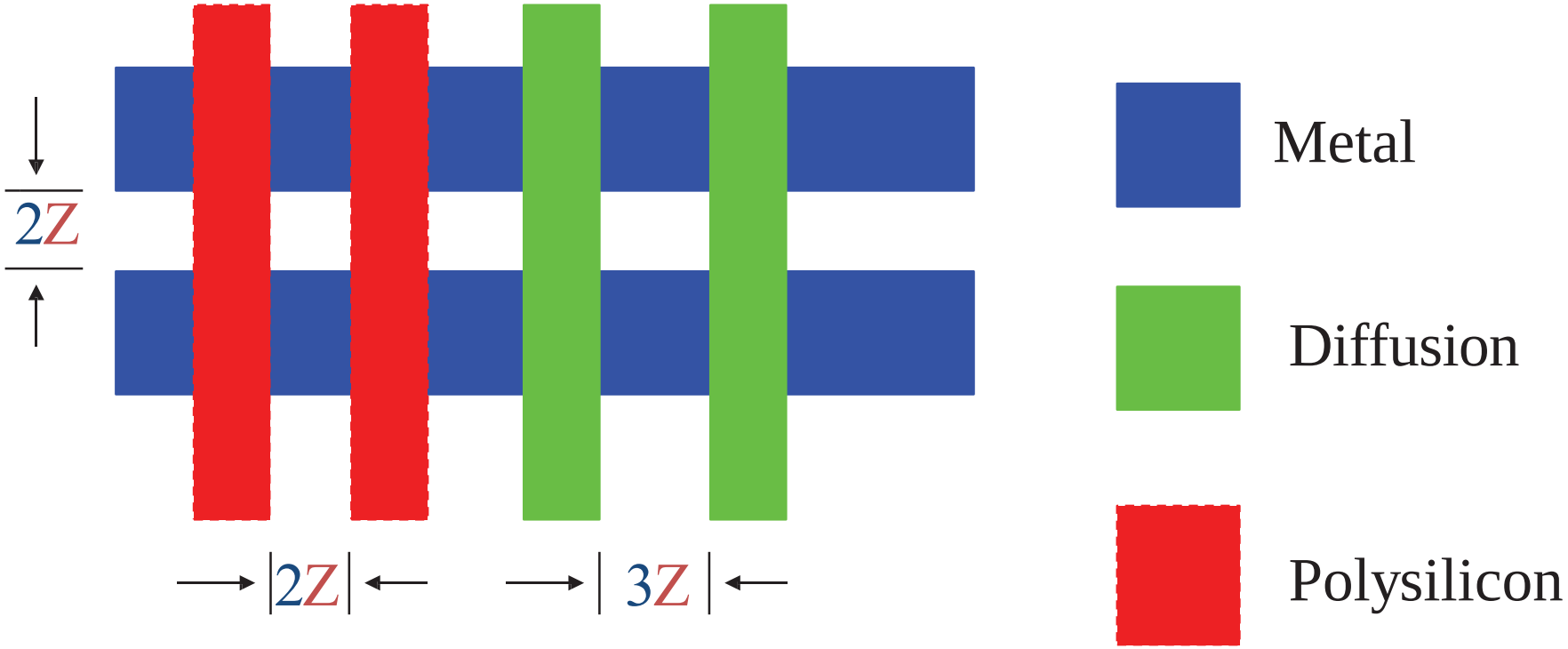


Design Rules

PolySi – PolySi space $2Z$

Metal - Metal space $2Z$

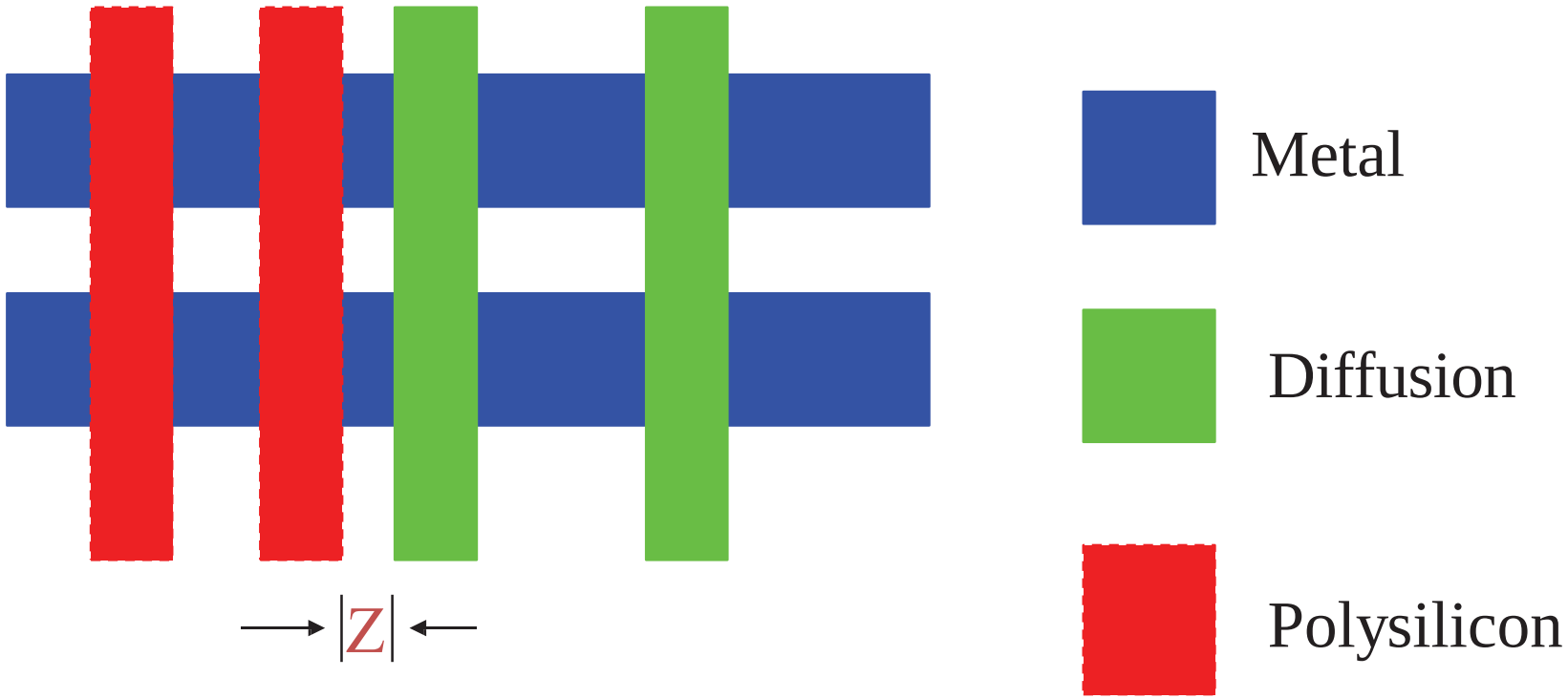
Diffusion – Diffusion $3Z$ **To avoid the possibility of their associated regions overlapping and conducting current**



Design Rules

Diffusion – PolySi Z To prevent the lines overlapping to form unwanted capacitor

Metal lines can pass over both diffusion and polySi without electrical effect. Where no separation is specified, metal lines can overlap or cross



Metal Vs PolySi/Diffusion

- Metal lines can pass over both diffusion and polySi without electrical effect
- It is recommended practice to leave Z between a metal edge and a polySi or diffusion line to which it is not electrically connected

Review:

- **poly-poly spacing** $2Z$
- **diff-diff spacing** $3Z$

  (depletion regions tend to spread outward)
- **metal-metal spacing** $2Z$
- **diff-poly spacing** Z

Note

- Two Features on different mask layers can be misaligned by a maximum of 2l on the wafer.
- If the overlap of these two different mask layers can be catastrophic to the design, they must be separated by at least 2l
- If the overlap is just undesirable, they must be separated by at least l

When a transistor is formed?

Gate is formed where polySi crosses diffusion with thin oxide between these layers.

Design rules

min. line width of polySi and diffusion 2Z

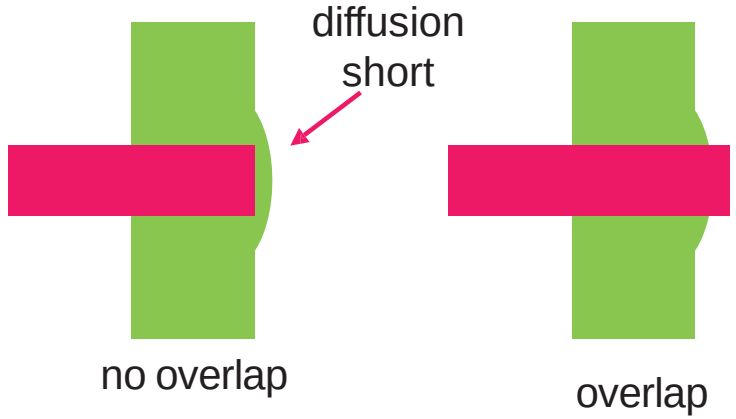
drain and source have min. length and width of 2Z

And

PolySi extends in the gate region…

The polySi of the gate extends 2Z beyond the gate area on to the field oxide to prevent the drain and source from shorting.

- **Diffusion Problems**



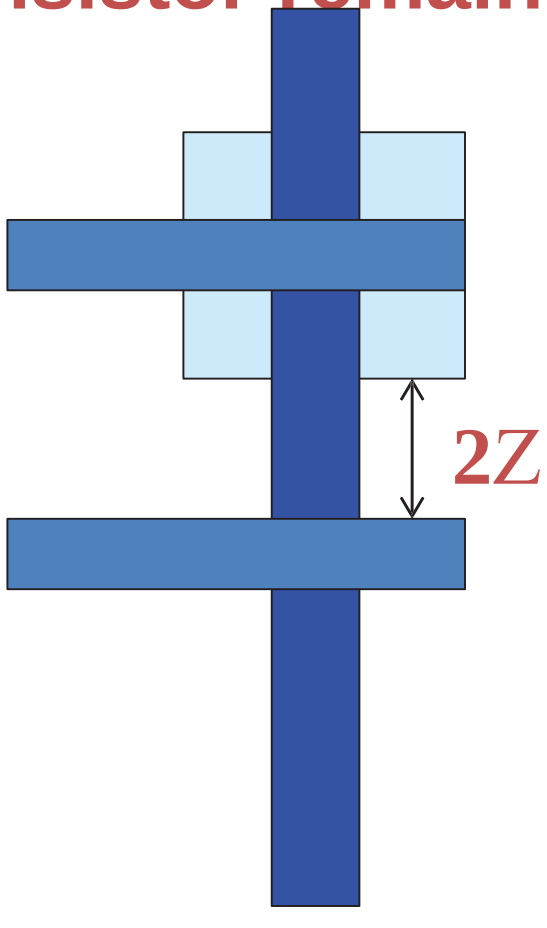
Depletion Transistor

We need depletion implant

An implant surrounding the Transistor by 2Z

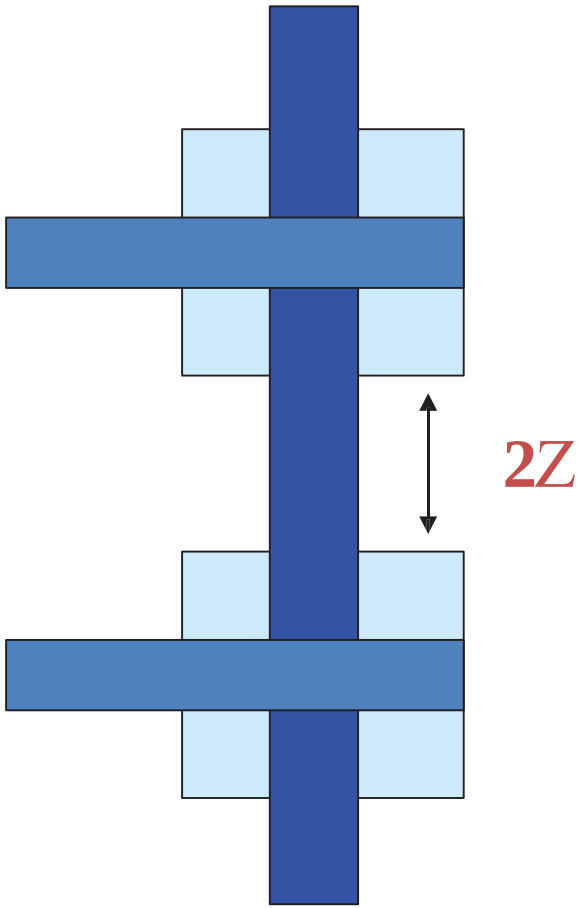
Ensures that no part of the transistor remains in the enhancement mode

A separation of 2Z from the gate of an enhancement transistor avoids affecting the device.



Depletion Transistor

Implants are separated by 2Z to prevent them from merging



Butting Contact

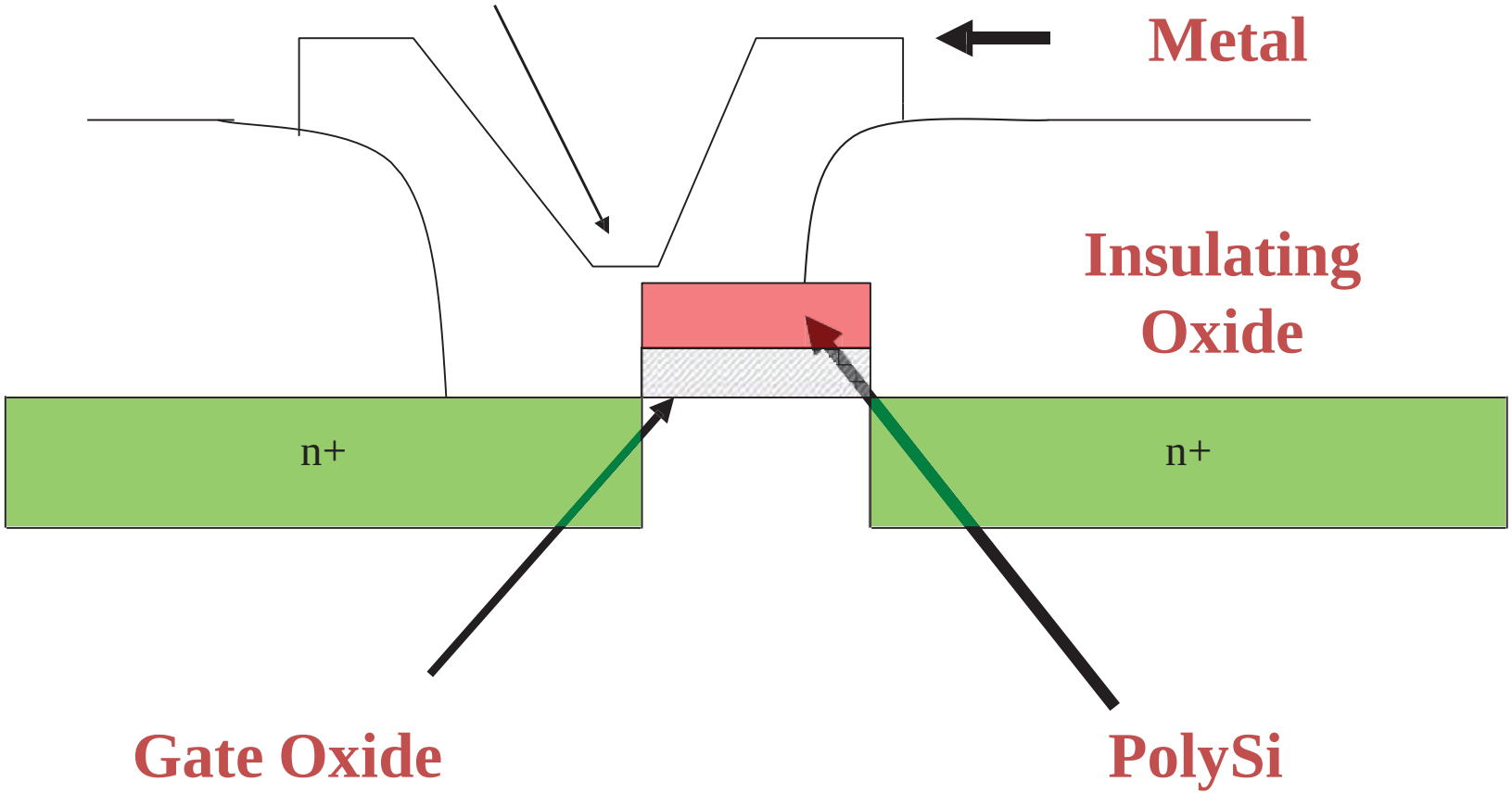
The gate and source of a depletion device can be connected by a method known as **butting contact.** Here metal makes contact to both the diffusion forming the source of the depletion transistor and to the polySi forming this device's gate.

Advantage:

No buried contact mask required and avoids associated processing.

Butting Contact

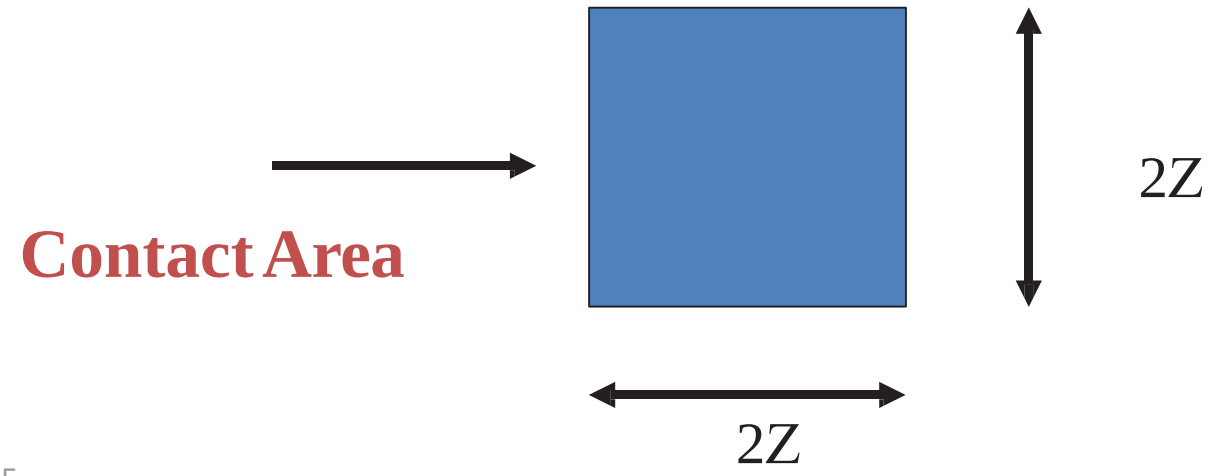
Problem: **Metal descending the hole has a tendency to fracture at the polySi corner, causing an open circuit.**



Buried Contact

It is a preferred method. The buried contact window defines the area where oxide is to be removed so that polySi connects directly to diffusion.

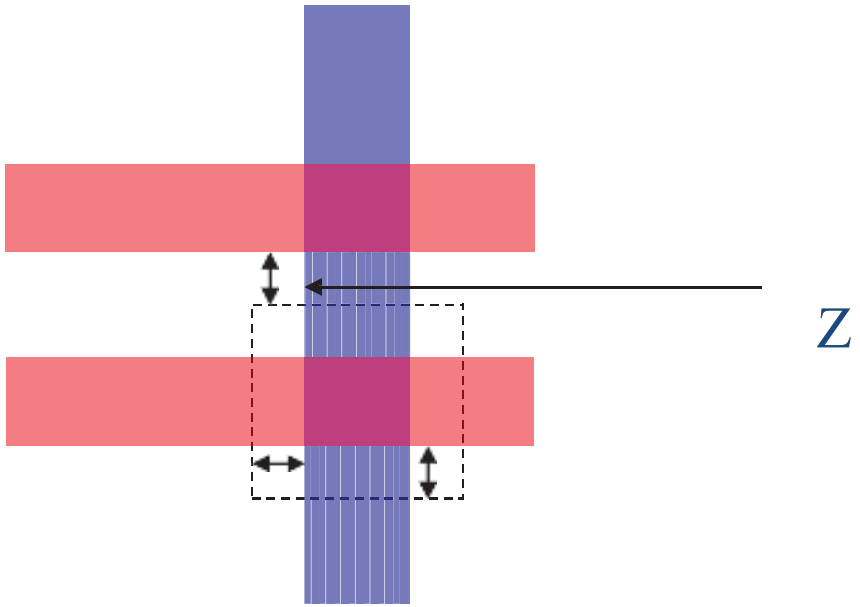
Contact Area must be a min. of 2Z*2Z to ensure adequate contact area.



Buried Contact

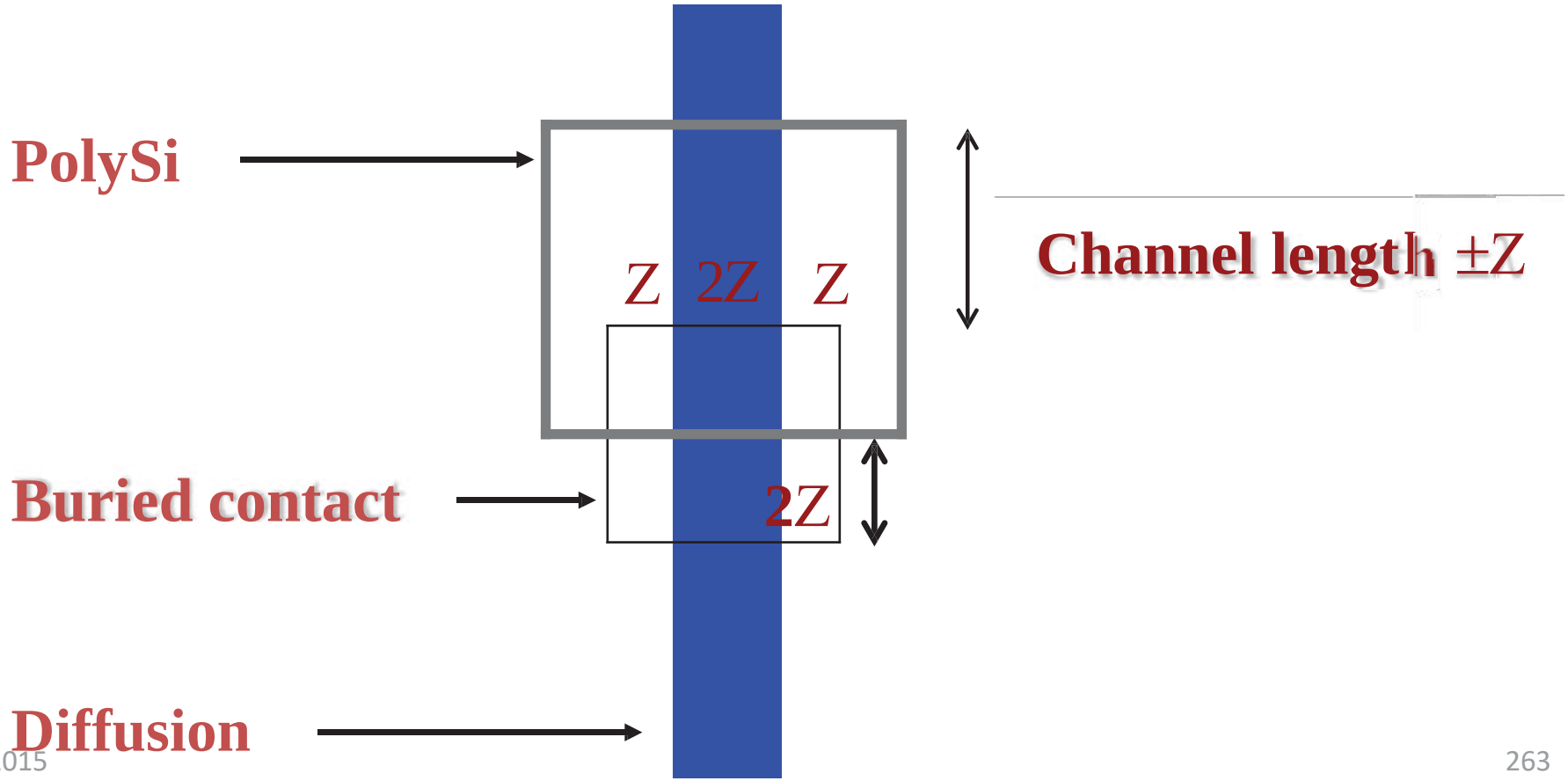
The buried contact window surrounds this contact by Z in all directions to avoid any part of this area forming a transistor.

Separated from its related transistor gate by Z to prevent gate area from being reduced.



Buried Contact

Here gate length is depend upon the alignment of the buried contact mask relative to the polySi and therefore vary by ±Z.

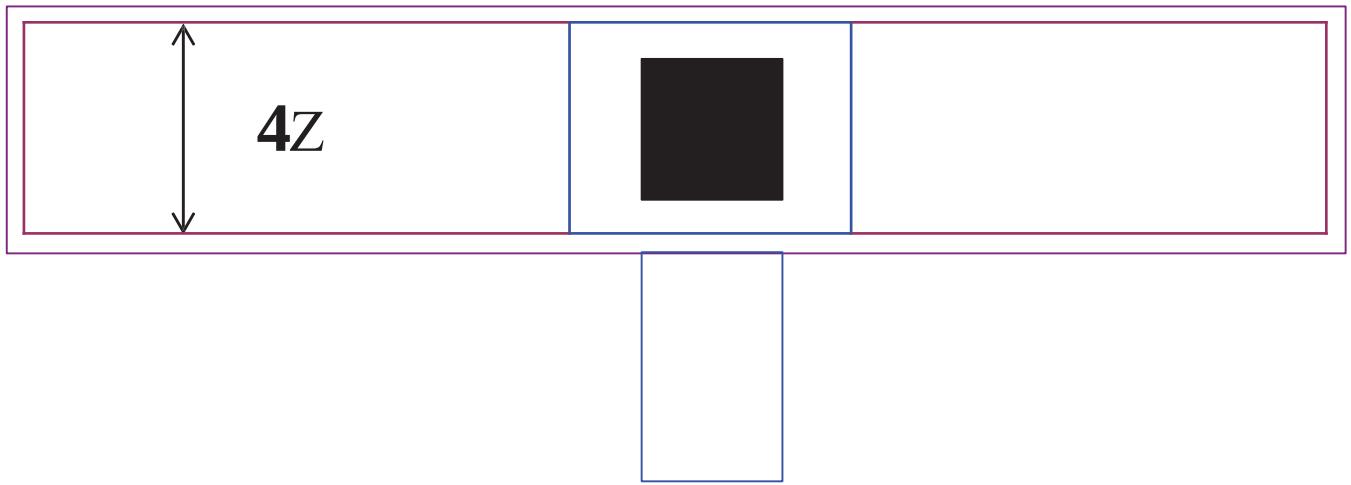


Contact Cut

Metal connects to polySi/diffusion by contact cut.

Contact area: 2Z×2Z

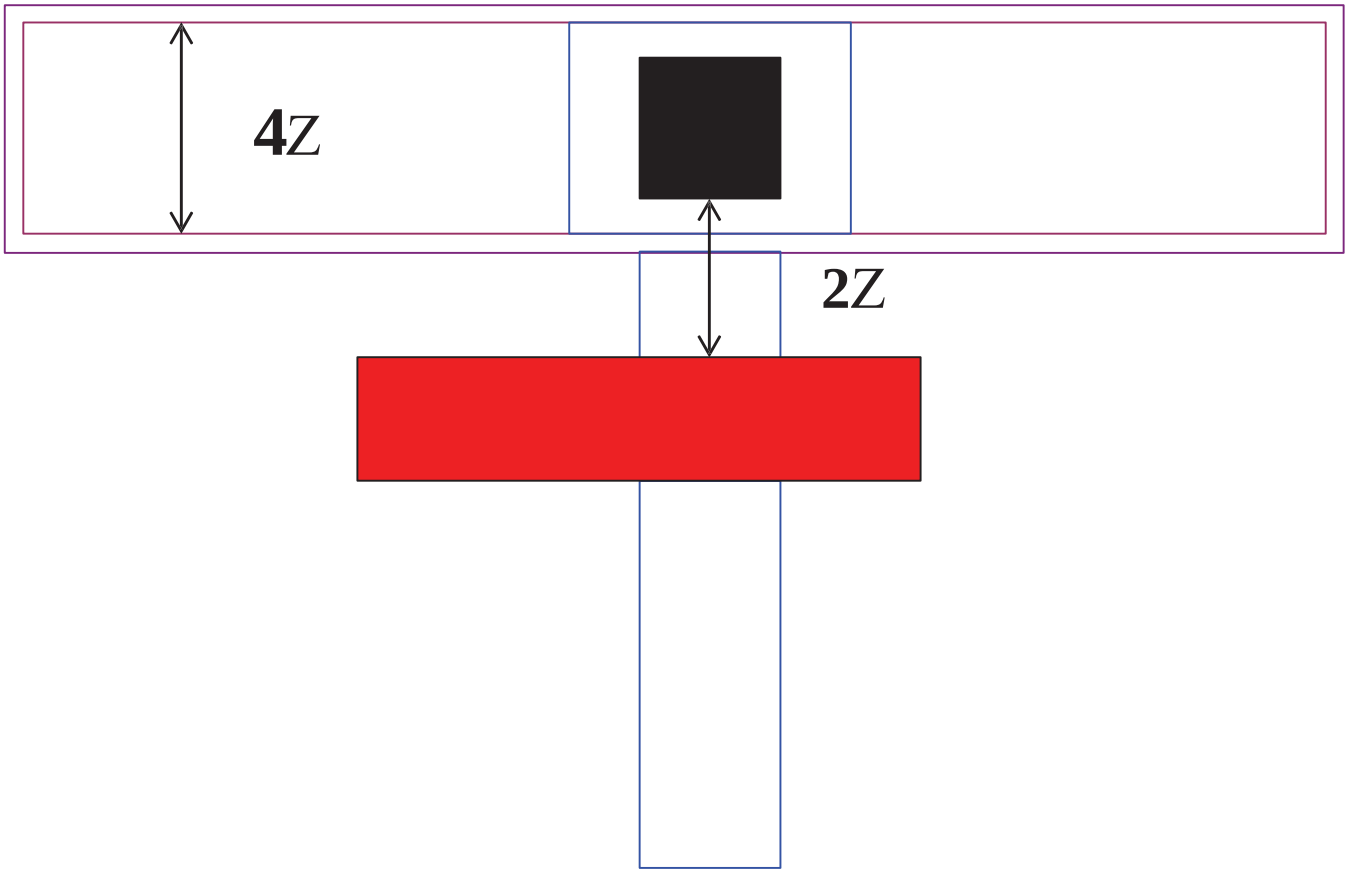
Metal and polySi or diffusion must overlap this contact area by Z so that the two desired conductors encompass the contact area despite any mis-alignment between conducting layers and the contact hole



Contact Cut

Contact cut – any gate: 2Z apart

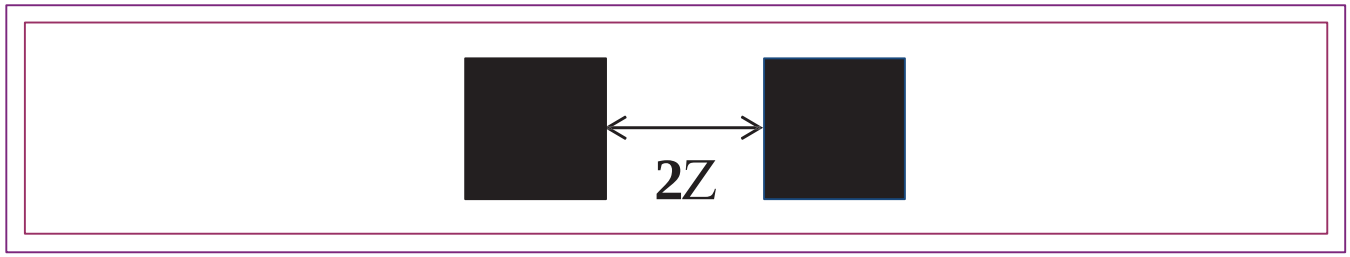
Why? No contact to any part of the gate.



Contact Cut

Contact cut – contact cut: 2Z apart

Why? To prevent holes from merging.



Rules for CMOS layout

Similar to those for NMOS except No

1. **Depletion implant**
2. **Buried contact**

Additional rules

1. **Definition of n-well area**
2. **Threshold implant of two types of transistor**
3. **Definition of source and drains regions for the NMOS and PMOS.**

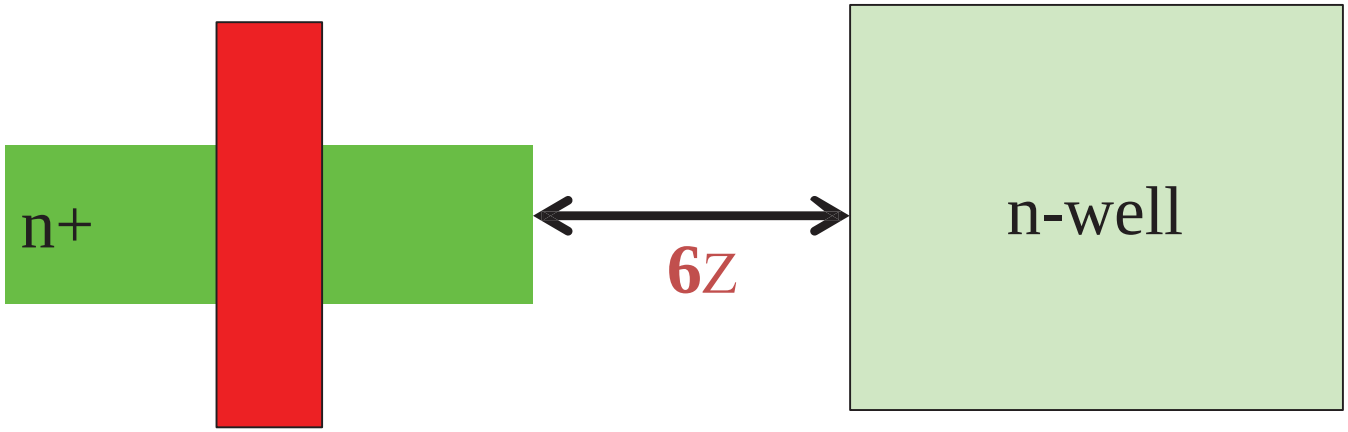
Rules for CMOS layout

To ensure the separation of the PMOS and NMOS devices,

n-well supporting PMOS is 6Z away from the active area of NMOS transistor.

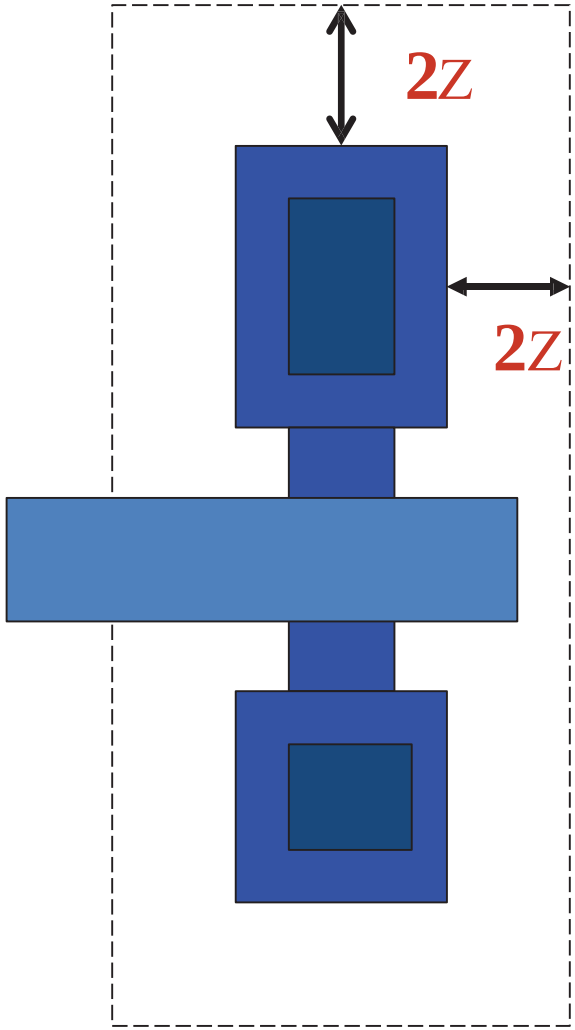
Why?

Avoids overlap of the associated regions



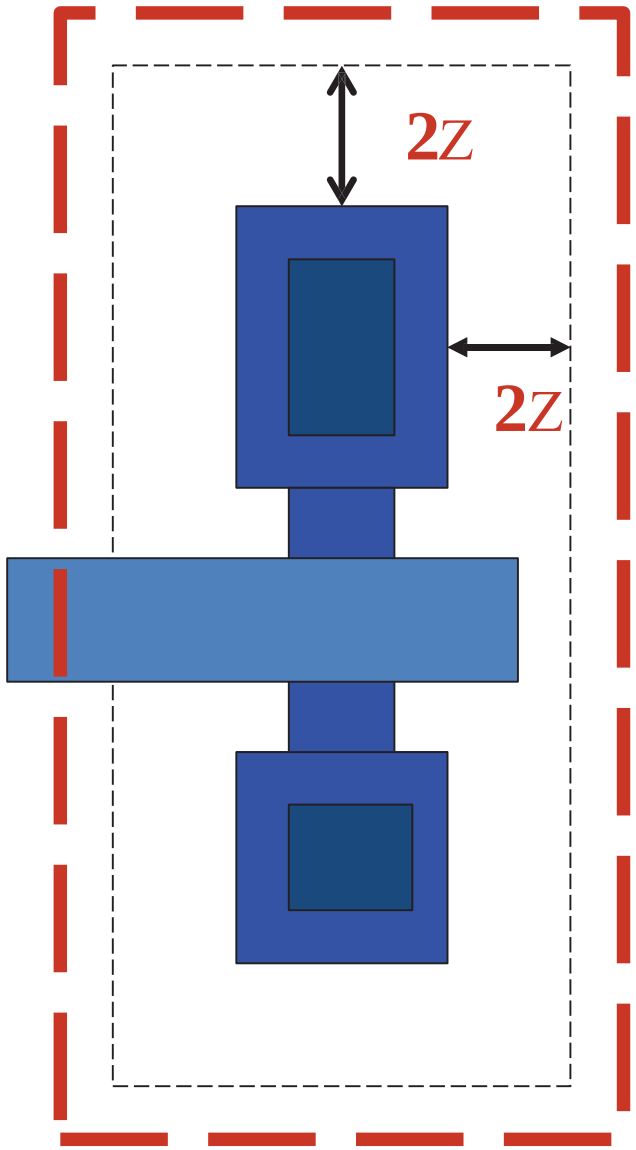
Rules for CMOS layout

N-well must completely surround the PMOS device's active area by 2Z



Rules for CMOS layout

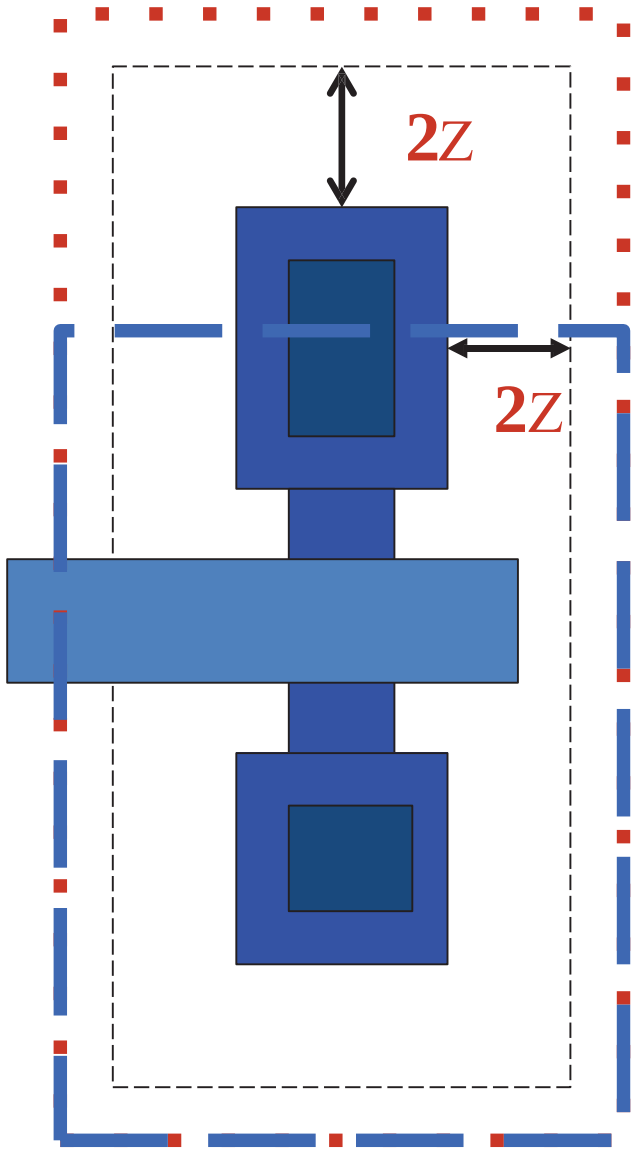
The threshold implant mask covers all n-well and surrounds the n-well by Z



Rules for CMOS layout

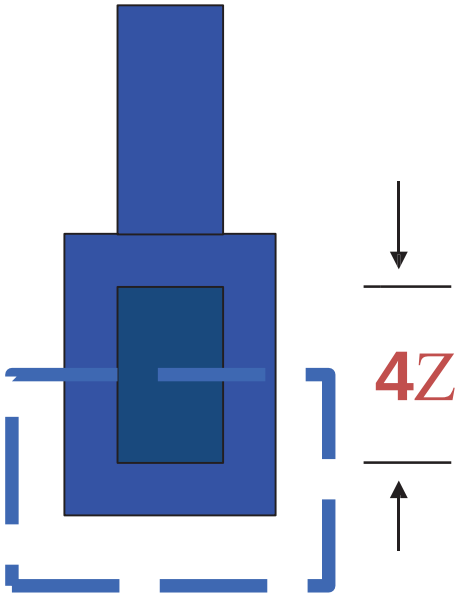
The p^+ diffusion mask defines the areas to receive a p^+ diffusion.

It is coincident with the threshold mask surrounding the PMOS transistor but excludes the n-well region to be connected to the supply.



Rules for CMOS layout

A p^+ diffusion is required to effect the ground connection to the substrate. Thus mask also defines this substrate region. It surrounds the conducting material of this contact by Z.



Rules for CMOS layout

Total contact area = $2Z \times 4Z$

Neither NMOS nor CMOS usually allow contact cuts to the gate of a transistor, because of the danger of etching away part of the gate

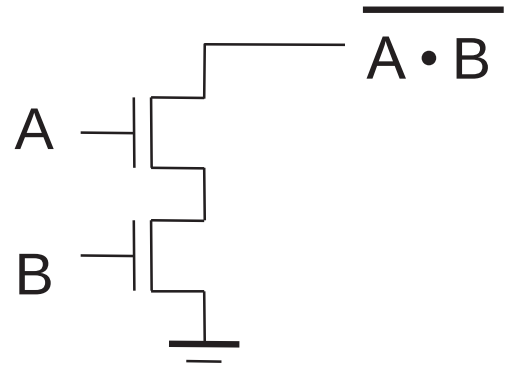
UNIT-IV
GATE LEVEL DESIGN

Topics

- Logic gates and other complex gates
- Switch logic
- Alternate gate circuits
- Time delays
- Driving large capacitive loads
- Wiring capacitances
- Fan-in and fan-out, Choice of layers

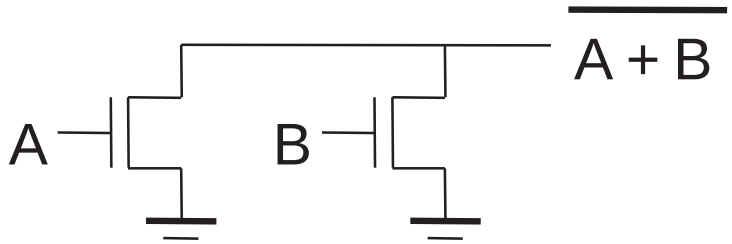
NMOS Gate construction

•NMOS devices in series implement a NAND function



A	B	F
0	0	1
0	1	1
1	0	1
1	1	0

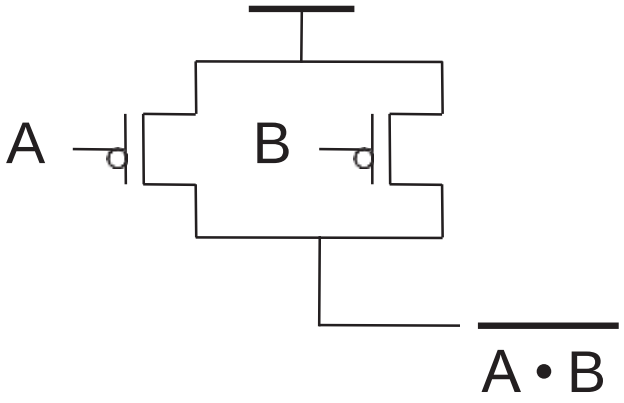
•NMOS devices in parallel implement a NOR function



A	B	F
0	0	1
0	1	0
1	0	0
1	1	0

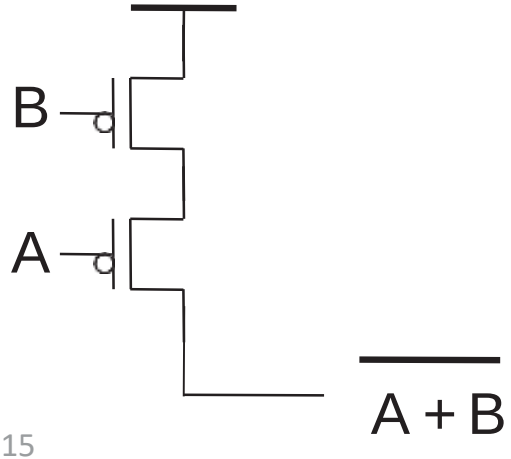
PMOS Gate construction

•PMOS devices in parallel implement a NAND function



A	B	F
0	0	1
0	1	1
1	0	1
1	1	0

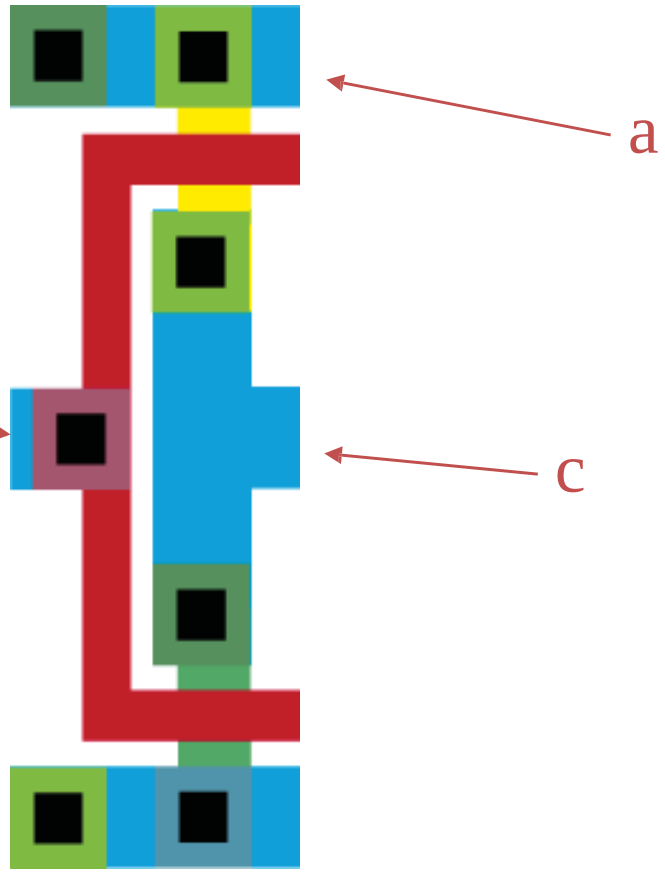
•PMOS devices in series implement a NOR function



A	B	F
0	0	1
0	1	0
1	0	0
1	1	0

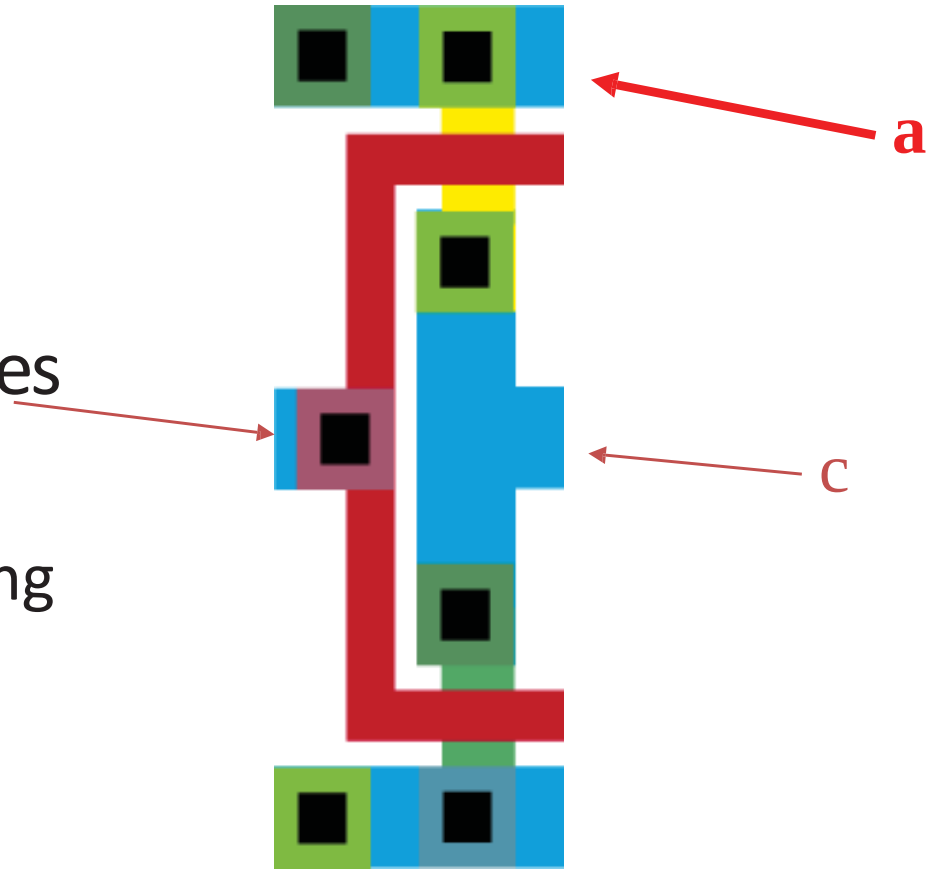
Parasitics and Performance

- Consider the following layout:
- What is the impact on performance of parasitics
  - At point a (VDD rail)?
  - At point b (input)?
  - At Point c (output)?



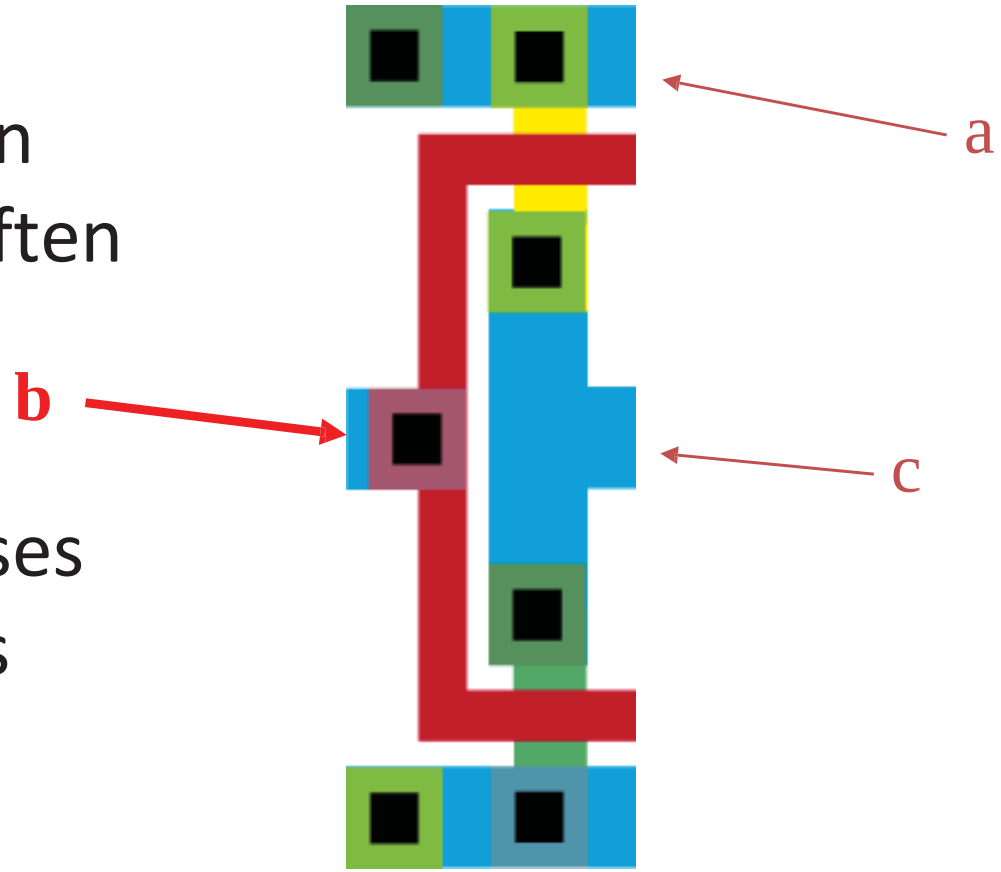
Parasitics and Performance

- a - power supply connections
  - capacitance - no effect on delay
  - resistance - increases delay (see p. 135)
    - minimize by reducing difffusion length
    - minimize using parallel vias



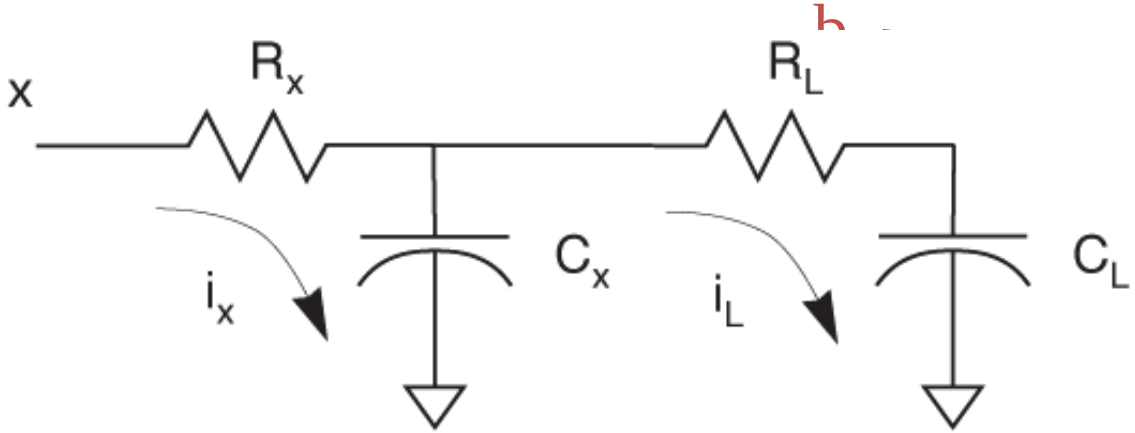
Parasitics and Performance

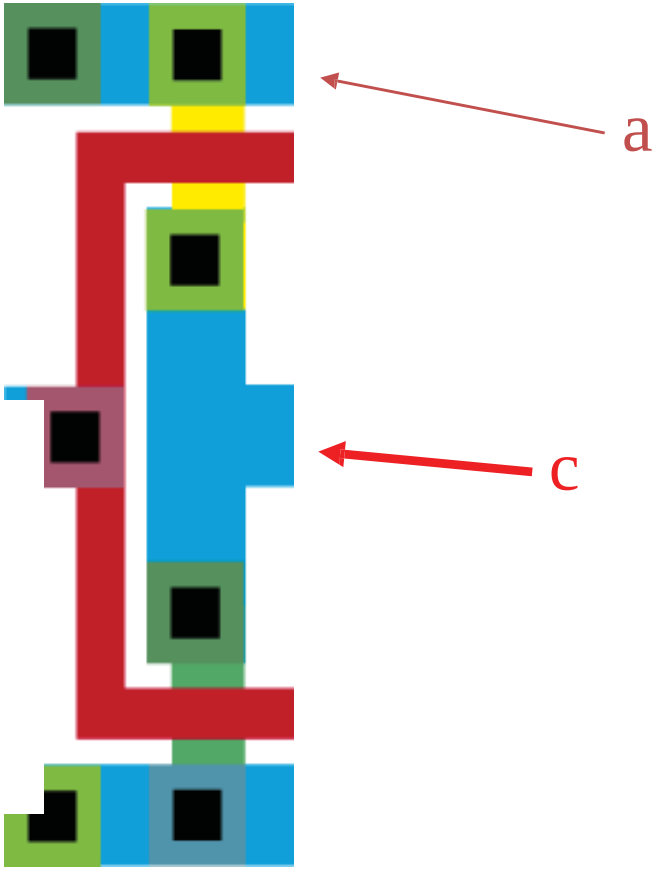
- b - gate input
  - capacitance increases delay on previous stage (often transistor gates dominate)
  - resistance increases delay on previous stage



Parasitics and Performance

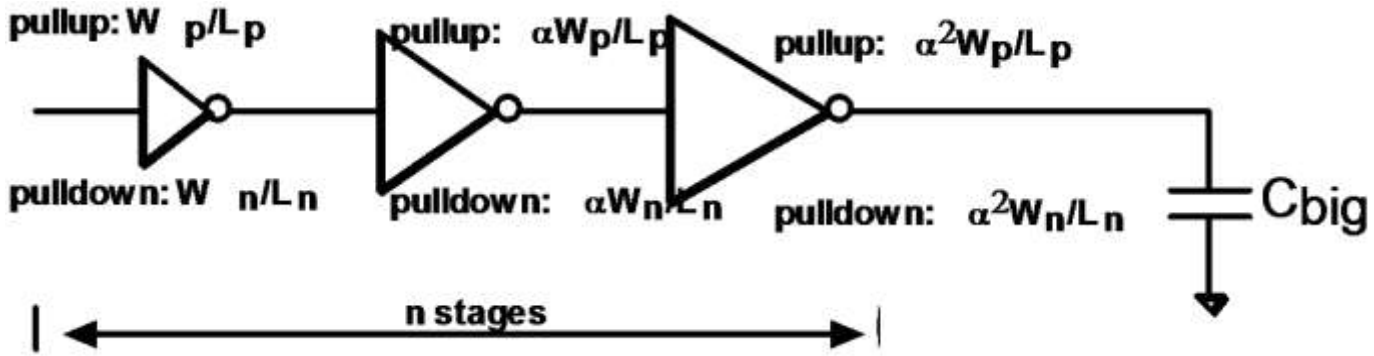
- c - gate output
  - resistance, capacitance increase delay
  - Resistance & capacitance "near" to output causes additional delay





Driving Large Loads

- Off-chip loads, long wires, etc. have high capacitance
- Increasing transistor size increases driving ability (and speed), but in turn increases gate capacitance
- Solution: stages of progressively larger transistors
  - Use nopt = ln(Cbig/Cg).
  - Scale by a factor of a=e



Summary: Static CMOS

- Advantages
  - High Noise Margins (VOH=VDD, VOL=Gnd)
  - No static power consumption (except for leakage)
  - Comparable rise and fall times (with proper sizing)
  - Robust and easy to use
- Disadvantages
  - Large transistor counts (2N transistors for N inputs)
    - Larger area
    - More parasitic loading (2 transistor gates on each input)
  - Pullup issues
    - Lower driving capability of P transistors
    - Series connections especially problematic
    - Sizing helps, but increases loading on gate inputs

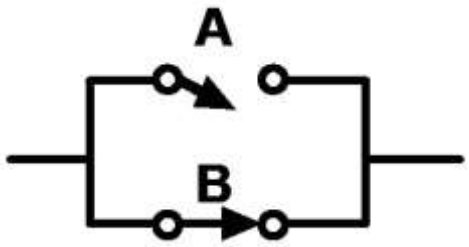
Alternatives to Static CMOS

- Switch Logic
- nmos
- Pseudo-nmos
- Dynamic Logic
- Low-Power Gates

Switch Logic

- Key idea: use transistors as switches
- Concern: switches are bidirectional

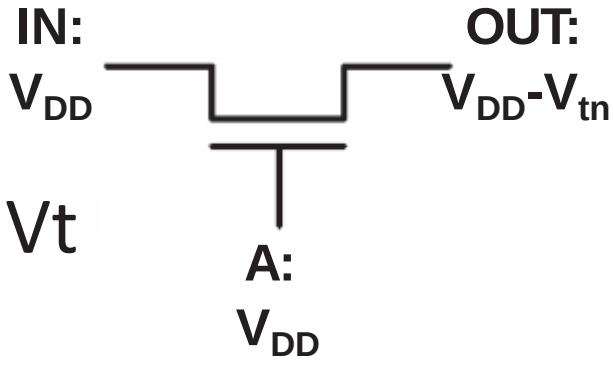
AND



OR

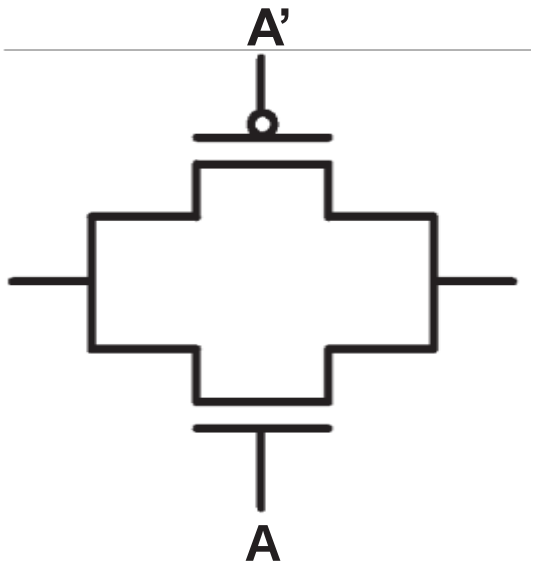
Switch Logic - Pass Transistors

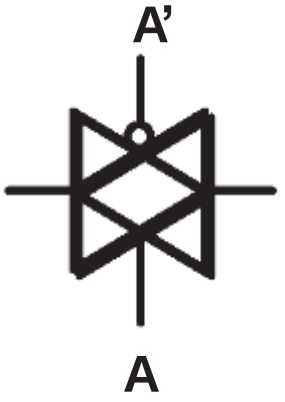
- Use n-transistor as “switches”
- “Threshold problem”
  - Transistor switches off when Vgs < Vt
  - VDD input -> VDD-Vt output
- Special gate needed to “restore” values



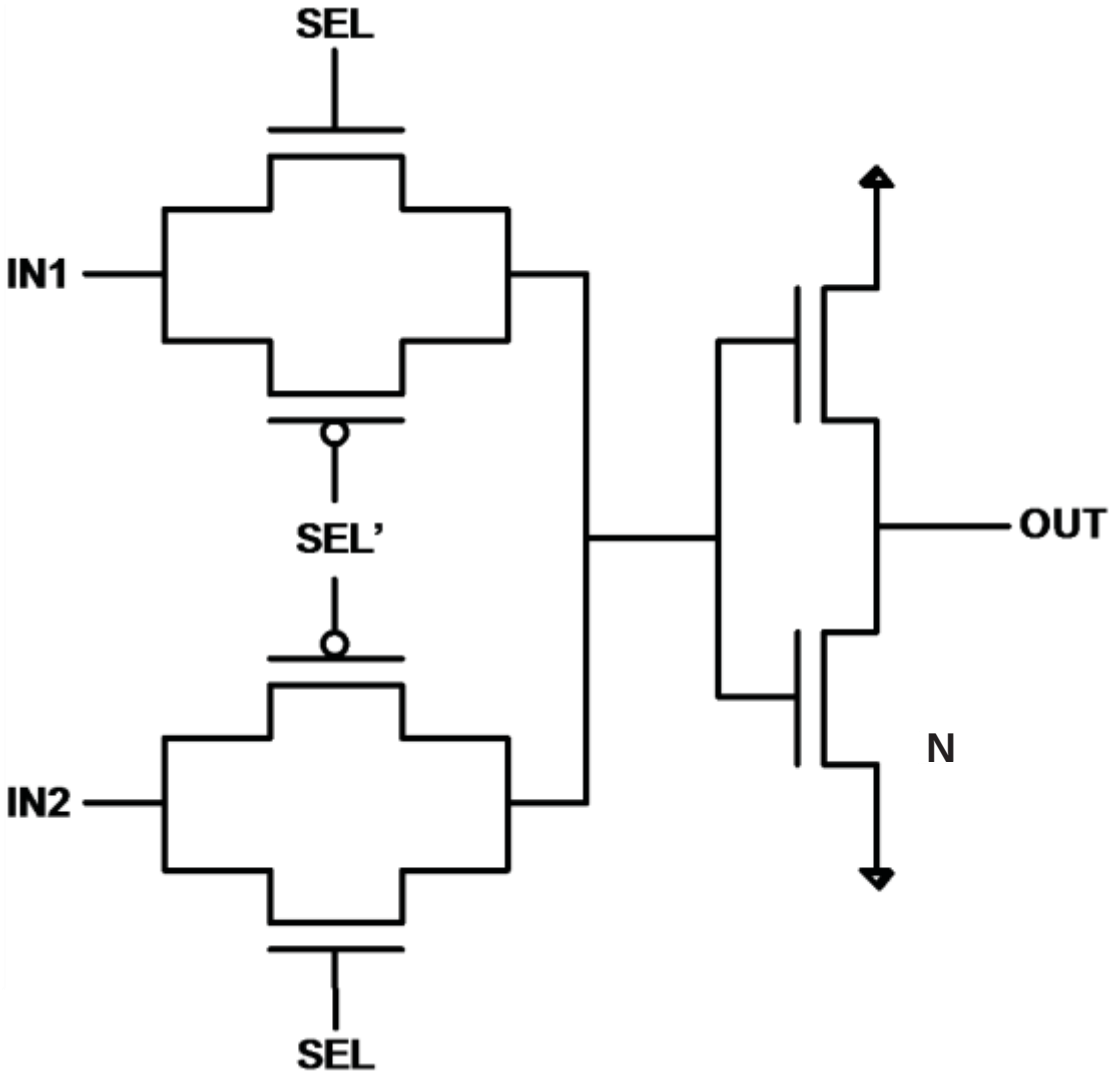
Switch Logic - Transmission Gates

- Complementary transistors - n and p
- No threshold problem
- Cost: extra transistor, extra control input
- Not a perfect conductor!



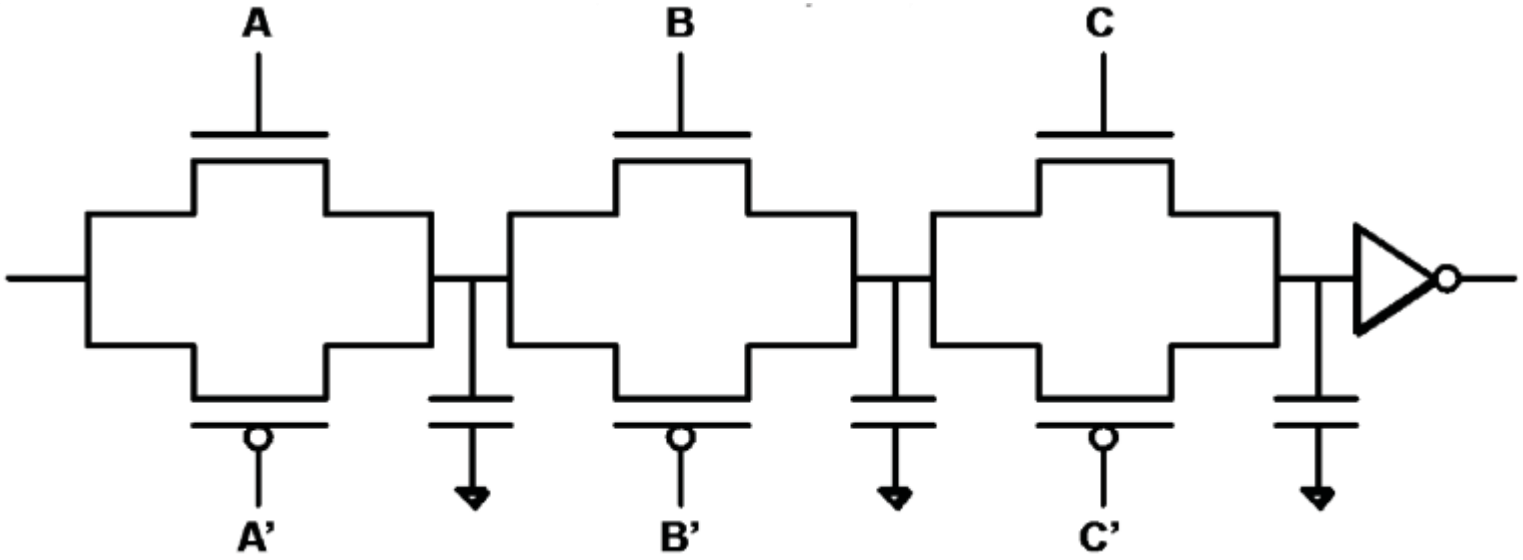


Switch Logic Example - 2-1 MUX



Charge Sharing

- Consider transmission gates in series
  - Each node has parasitic capacitances
  - Problems occur when inputs change to redistribute charge
  - Solution: design network so there is always a path from VDD or Gnd to output

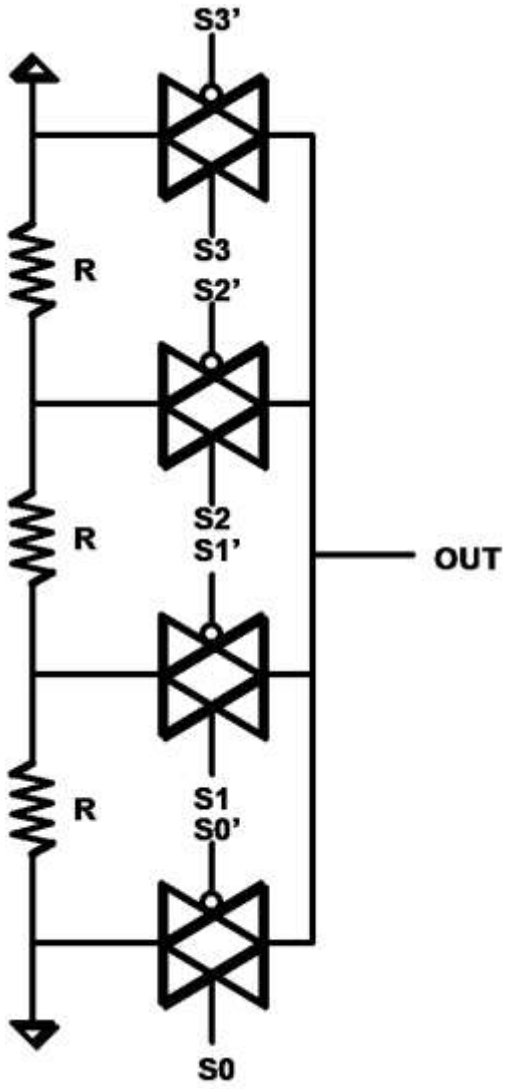


Aside: Transmission Gates in Analog

- Transmission Gates

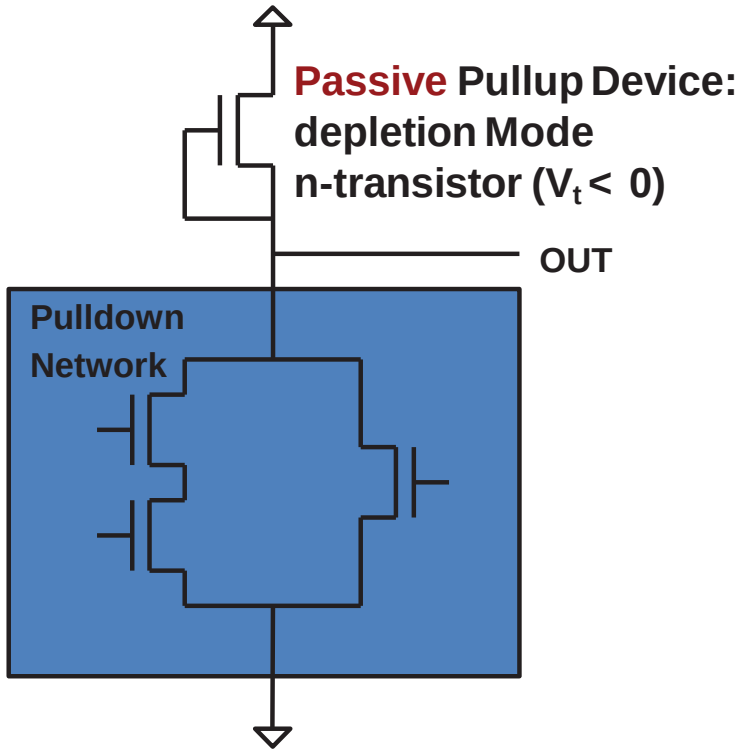
work with analog values, too!

- Example:
  Voltage-Scaling D/A Converter



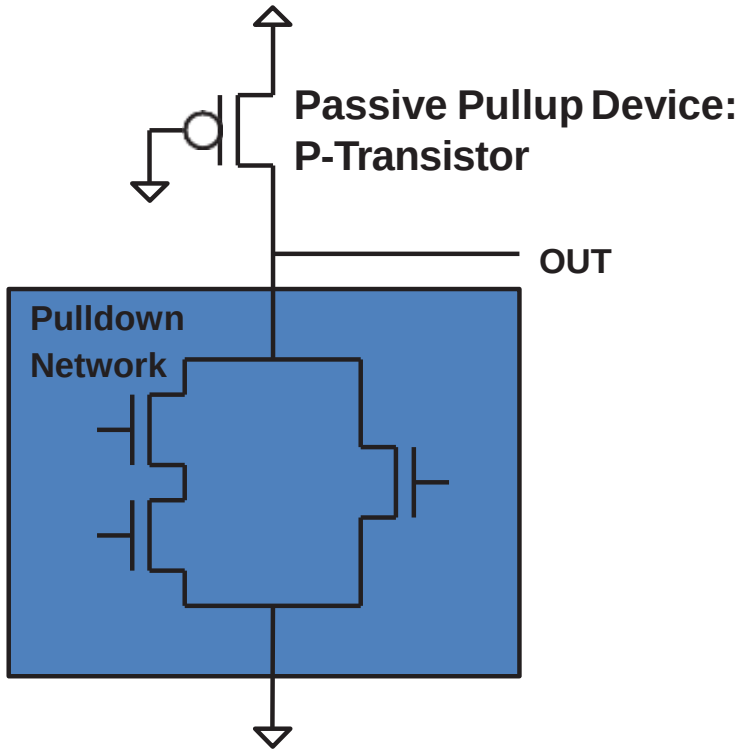
NMOS Logic

- Used before CMOS was widely available
- Uses only n transistors
  - Normal n transistors in pull-down network
  - depletion-mode n transistor (Vt < 0) used for pull-up
  - "ratioed logic" required
- Tradeoffs:
  - Simpler processing
  - Smaller gates
  - higher power!
  - Additional design considerations for ratioed logic



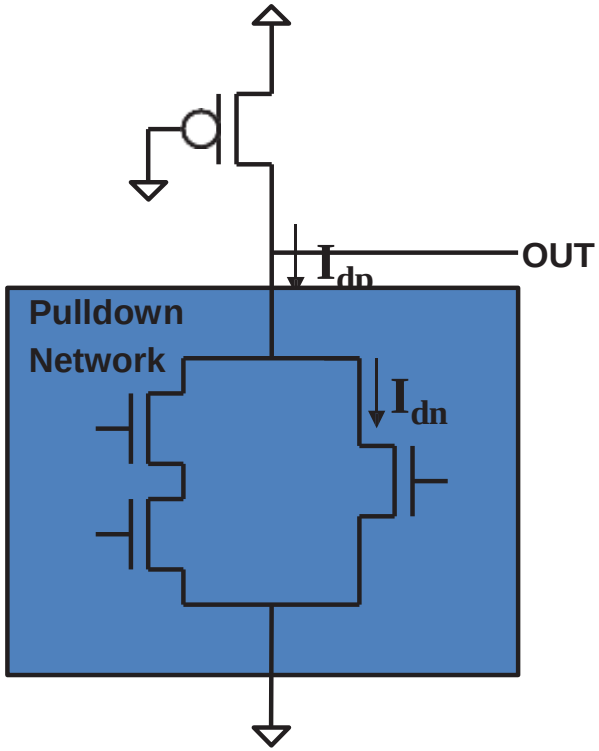
Pseudo-nmos Logic

- Same idea, as nmos, but use p-transistor for pullup
- "ratioed logic" required for proper design (more about this next)
- Tradeoffs:
  - Fewer transistors -> smaller gates, esp. for large number of inputs
  - less capacitative load on gates that drive inputs
  - larger power consumption
  - less noise margin (VOL > 0)
  - additional design considerations due to ratioed logic



Rationed Logic for Pseudo-nmos

- Approach:
  - Assume VOUT=VOL =0.25*VDD
  - Assume 1 pulldown transistor is on
  - Equate currents in p, n transistors
  - Solve for ratio between sizes of p, n transistors to get these conditions
  - Further calculations necessary for series connections



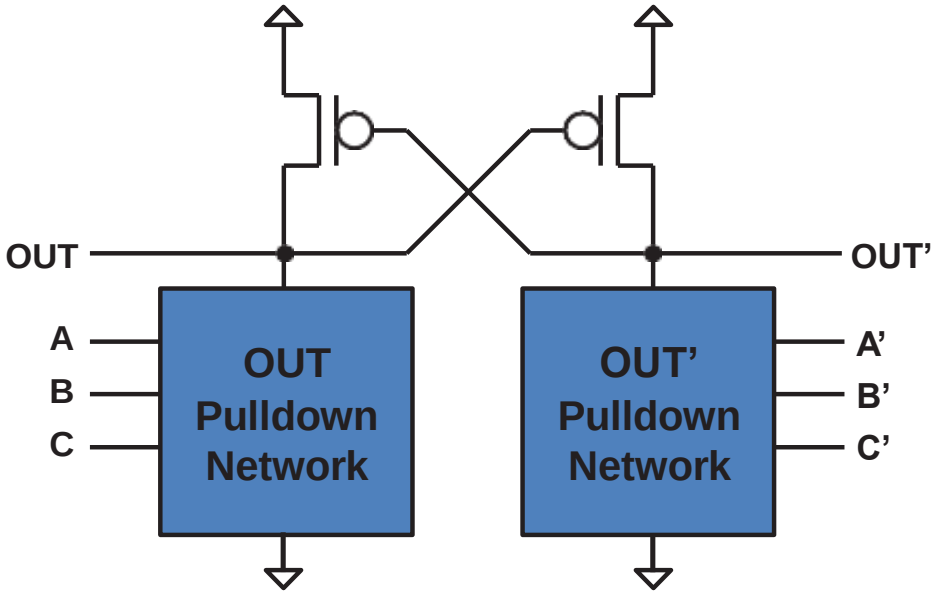
$$I_{dn} = I_{pn}$$

$$\frac{1}{2}k'_n \frac{W_n}{L_n}\left(V_{gs,n} - V_{tn}\right)^2 = \frac{1}{2}k'_p \frac{W_p}{L_p}\left[2\left(V_{gs,p} - V_{tp}\right)V_{ds,p} - V_{ds,p}^2\right] \quad (EQ\ 3-21)$$

$$\frac{W_p/L_p}{W_n/L_n} = 3.9 \quad (EQ\ 3-22) - Assuming\ V_{DD} = 3.3V$$

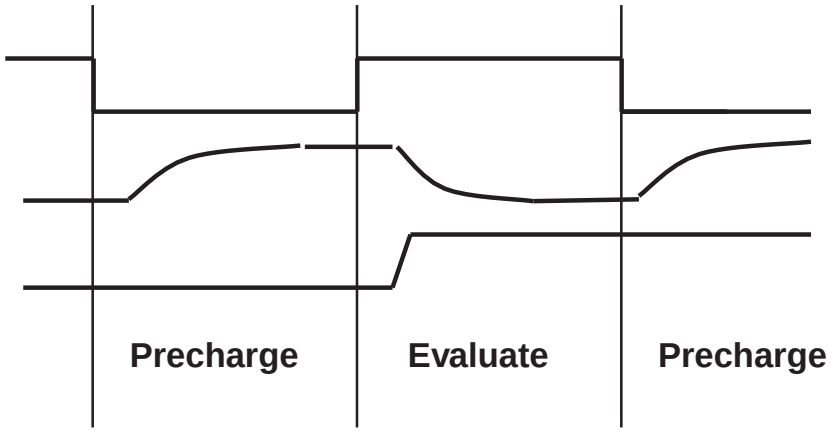
DCVS Logic

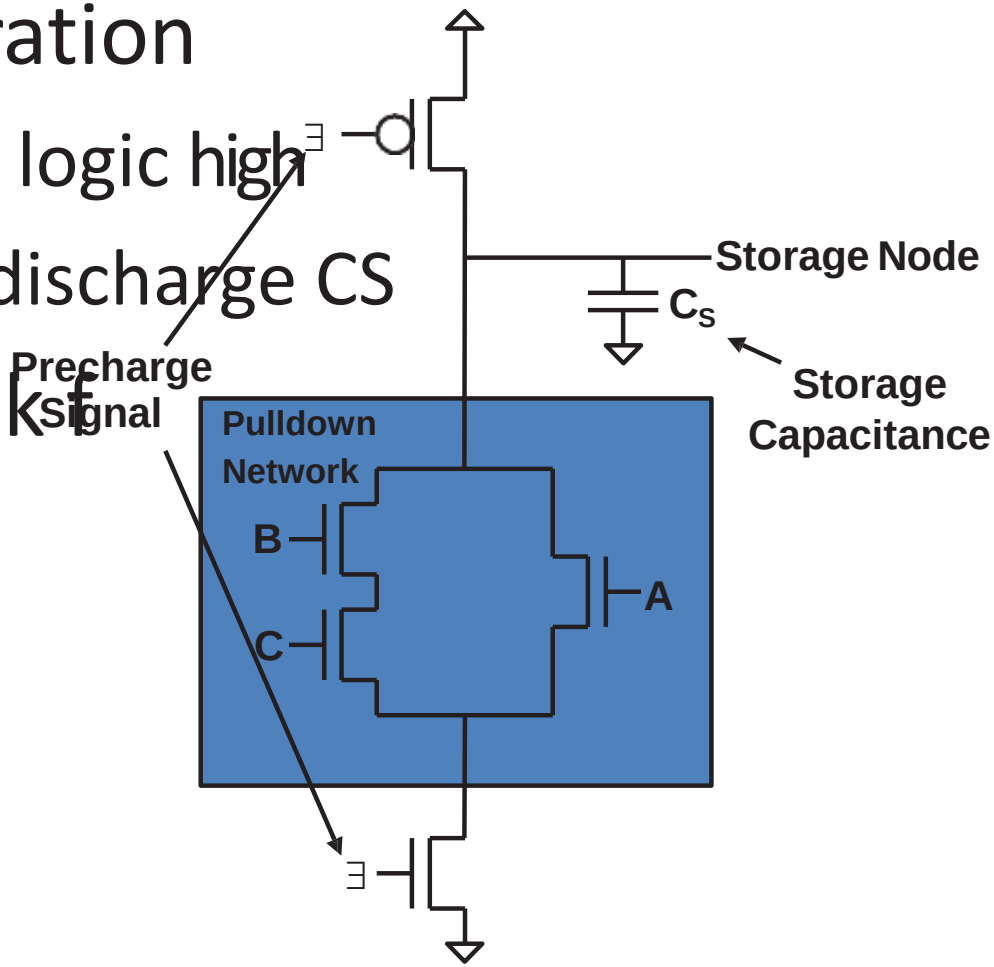
- DCVS - Differential Cascode Voltage Switch
- Differential inputs, outputs
- Two pulldown networks
- Tradeoffs
  - Lower capacitative loading than static CMOS
  - No ratioed logic needed
  - Low static power consumption
  - More transistors
  - More signals to route between gates



Dynamic Logic

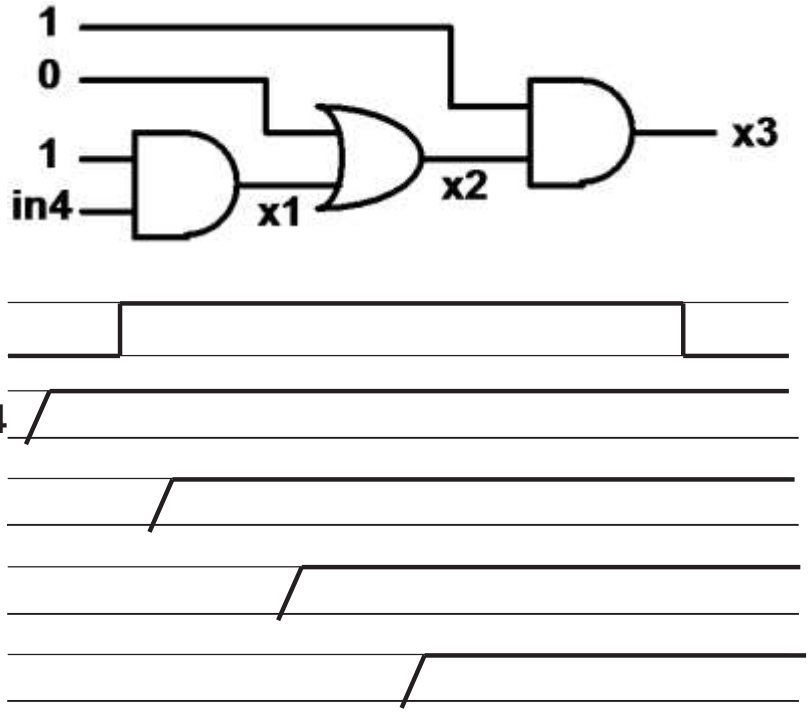
- Key idea: Two-step operation
  - precharge - charge CS to logic high
  - evaluate - conditionally discharge CS
- Control - precharge clock

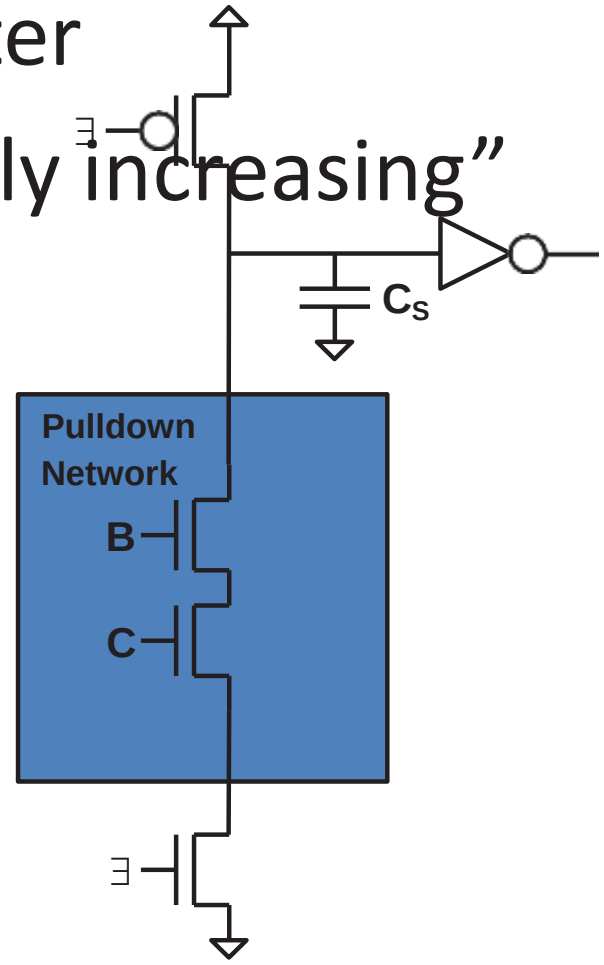




Domino Logic

- Key idea: dynamic gate + inverter
- Cascaded gates - “monotonically increasing”





Domino Logic Tradeoffs

- Fewer transistors -> smaller gates
- Lower power consumption than pseudo-nmos
- Clocking required
- Logic not complete (AND, OR, but no NOT)

More Techniques for Saving Power

- Reduce VDD (tradeoff: delay)
- Multiple Power Supplies
  - High VDD for “fast” logic
  - Low VDD for “slow” logic
- Dealing with leakage currents
  - Multiple-Threshold CMOS (MTCMOS)
  - Variable-Threshold CMOS (VTCMOS)

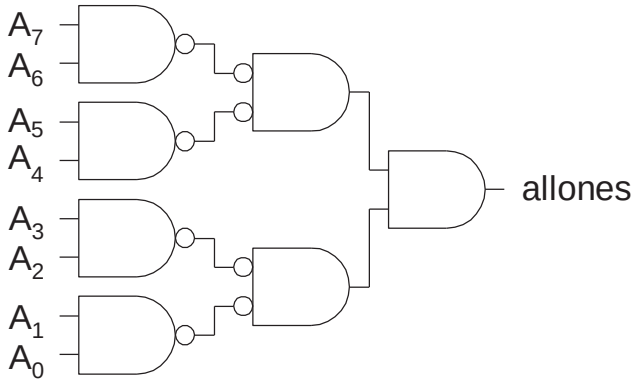
UNIT V
DATAPATH SUBSYSTEMS

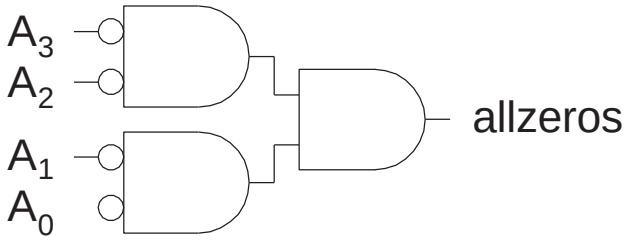
Topics

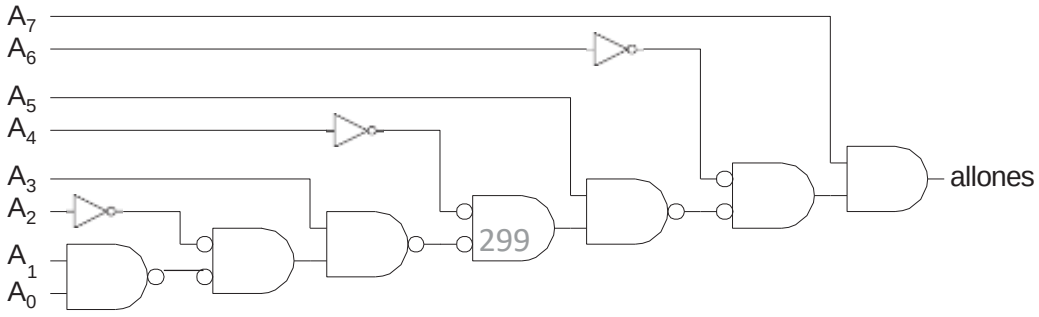
- Sub system design,
- Shifters,
- Adders,
- ALUs,
- Multipliers,
- Parity generators,
- Comparators,
- Zero/One detectors,
- Counters.

1's & 0's Detectors

- 1's detector: N-input AND gate
- 0's detector: NOTs + 1's detector (N-input NOR)





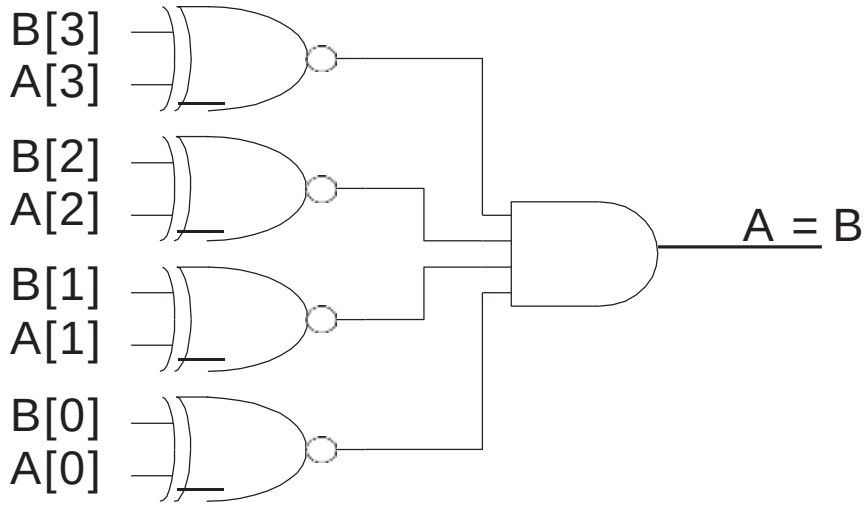


Comparators

- 0's detector: A = 00…000
- 1's detector: A = 11…111
- Equality comparator: A = B
- Magnitude comparator: A < B

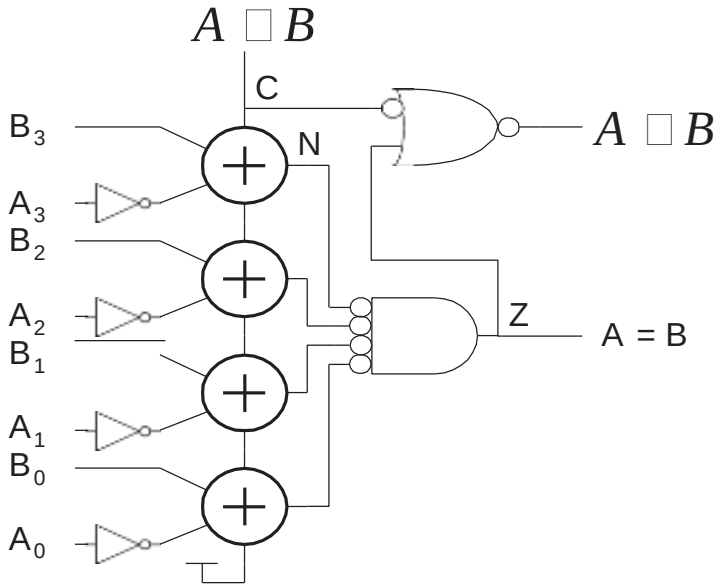
Equality Comparator

- Check if each bit is equal (XNOR, aka equality gate)
- 1's detect on bitwise equality



Magnitude Comparator

- Compute B – A and look at sign
- B – A = B + ~A + 1
- For unsigned numbers, carry out is sign bit



Signed vs. Unsigned

- For signed numbers, comparison is harder
  - C: carry out
  - Z: zero (all bits of B – A are 0)
  - N: negative (MSB of result)
  - V: overflow (inputs had different signs, output sign σ B)
  - S: N xor V (sign of result)
  -

Relation	Unsigned Comparison	Signed Comparison
$A = B$	Z	Z
$A \neq B$	$\overline{Z}$	$\overline{Z}$
$A < B$	$C \cdot \overline{Z}$	$\overline{S} \cdot \overline{Z}$
$A > B$	C	S
$A \leq B$	C	$\overline{S}$
$A \geq B$	$\overline{C} + Z$	$S + Z$

Shifters

- Logical Shift:
  - Shifts number left or right and fills with 0's
    - 1011 LSR 1 = 0101 1011 LSL1 = 0110
- Arithmetic Shift:
  - Shifts number left or right. Rt shift sign extends
    - 1011 ASR1 = 1101 1011 ASL1 = 0110
- Rotate:
  - Shifts number left or right and fills with lo
    - 1011 ROR1 = 1101 1011 ROL1 = 0111

Adders

- Single-bit Addition
- Carry-Ripple Adder
- Carry-Skip Adder
- Carry-Look ahead Adder
- Carry-Select Adder
- Carry-Increment Adder
- Tree Adder

Barrel Shifter

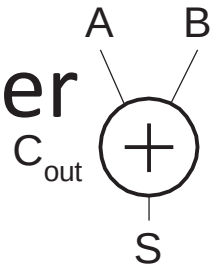
- Barrel shifters perform right rotations using wrap-around wires.
- Left rotations are right rotations by N – k = k + 1 bits.
- Shifts are rotations with the end bits masked off.

Single-Bit Addition

- Half Adder

$S = A \heartsuit B$

$C_{out} = A \cdot B$

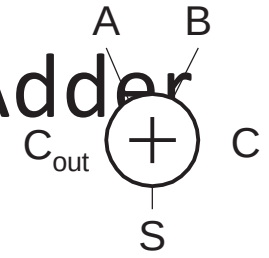


A	B	C_{out}	S
0	0		
0	1		
1	0		
1	1		

Full Adder

$S = A \heartsuit B \heartsuit C$

$C_{out} = MAJ(A, B, C)$



A	B	C	C_{out}	S
0	0	0		
0	0	1		
0	1	0		
0	1	1		
1	0	0		
1	0	1		
1	1	0		
1	1	1		

PGK

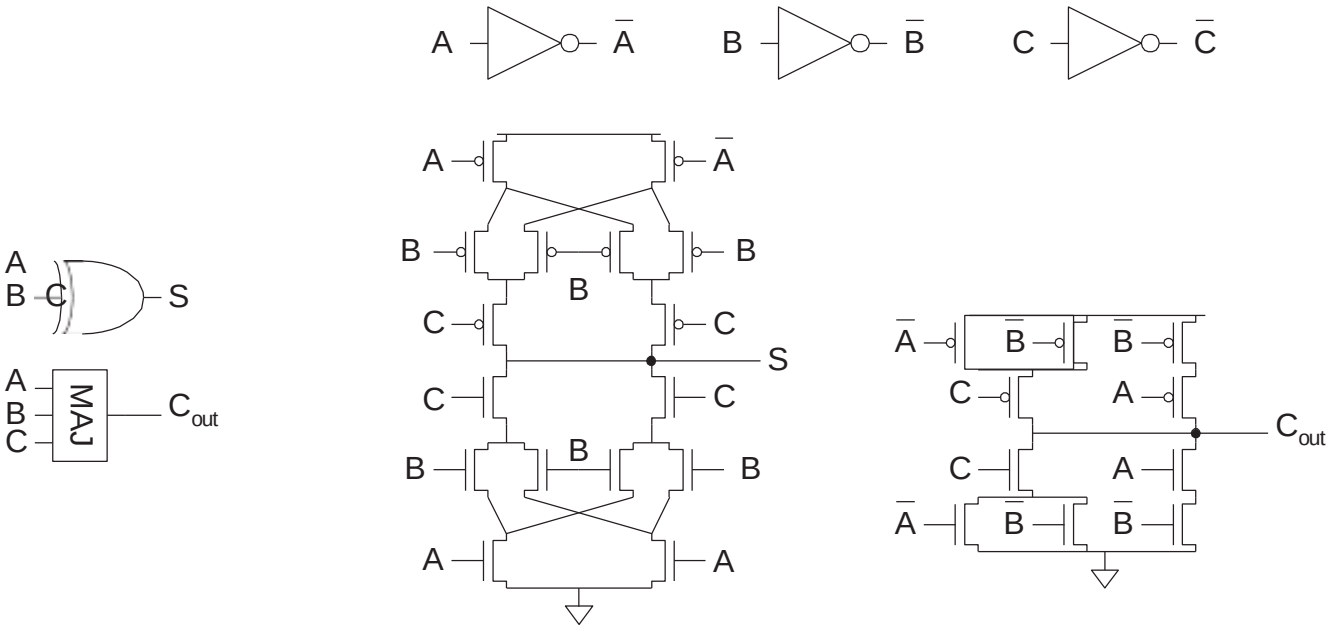
- For a full adder, define what happens to carries
  - (in terms of A and B)
  - Generate: Cout = 1 independent of C
    - G = A • B
  - Propagate: Cout = C
    - P = A ⊕ B
  - Kill: Cout = 0 independent of C
    - K = ~A • ~B

Full Adder Design I

- Brute force implementation from eqns

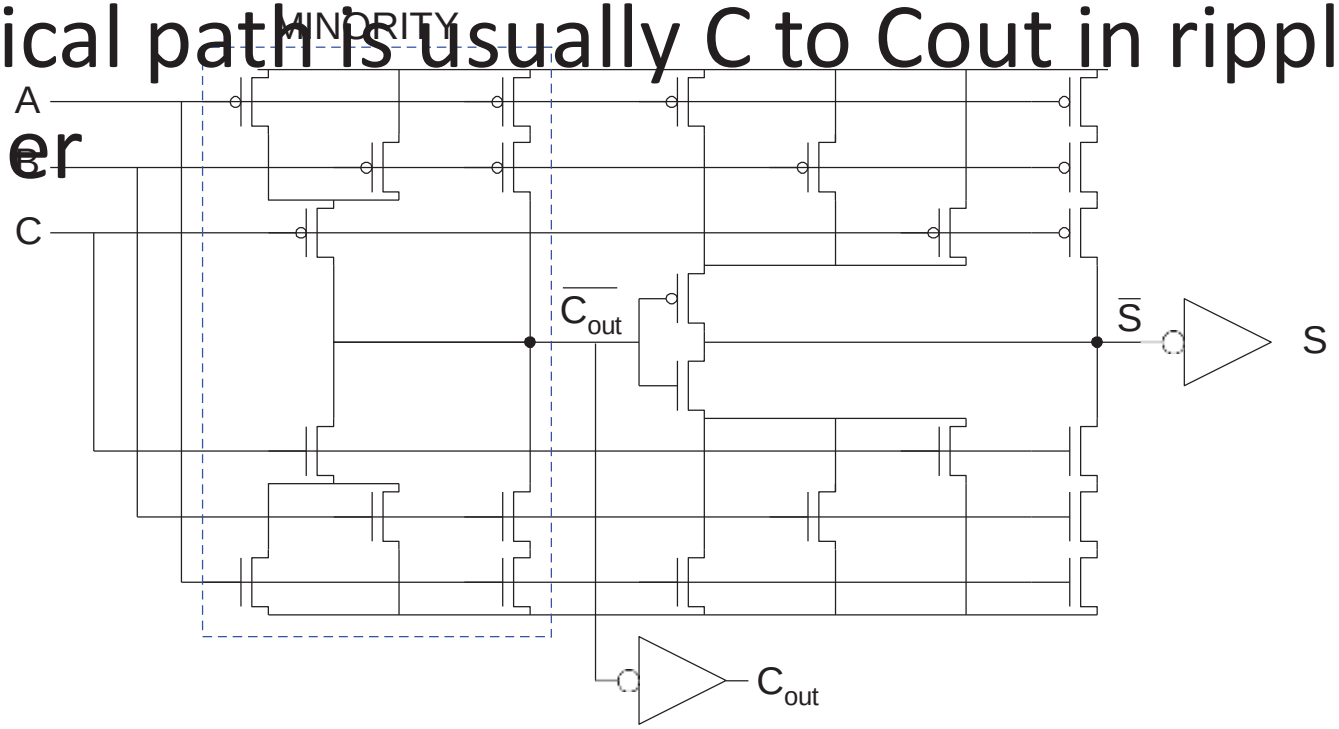
$$S = A \blacktriangledown B \blacktriangledown C$$

$$C_{\text{out}} = MAJ(A,B,C)$$



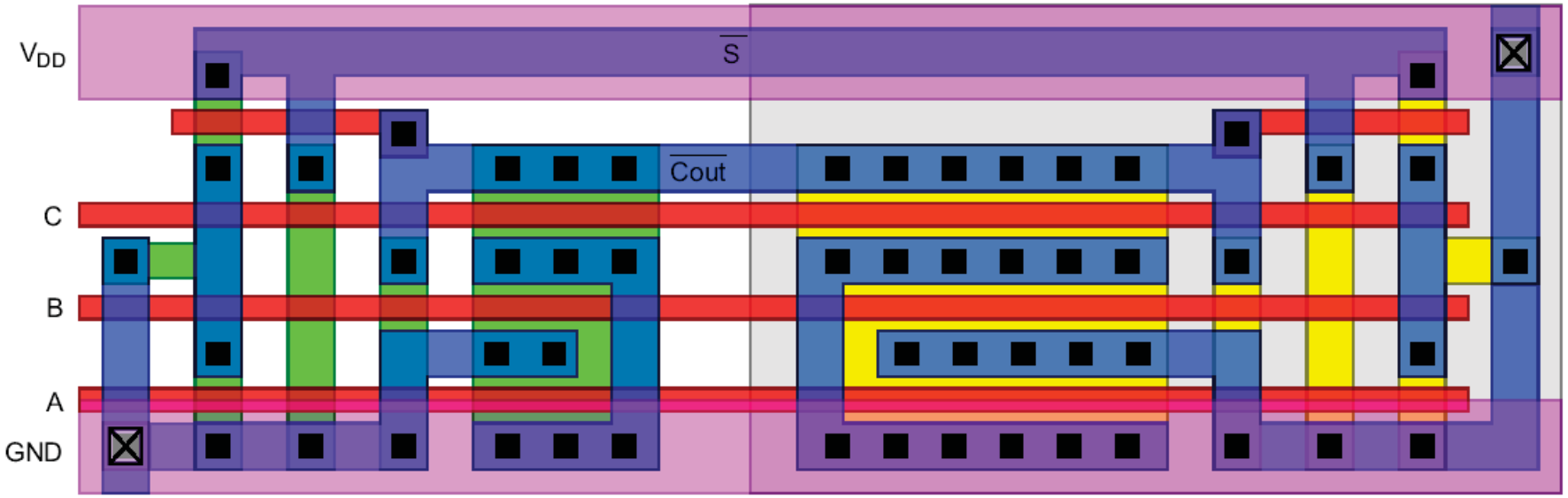
Full Adder Design II

- Factor S in terms of Cout
- S = ABC + (A + B + C)(~Cout)
- Critical path is usually C to Cout in ripple adder
-



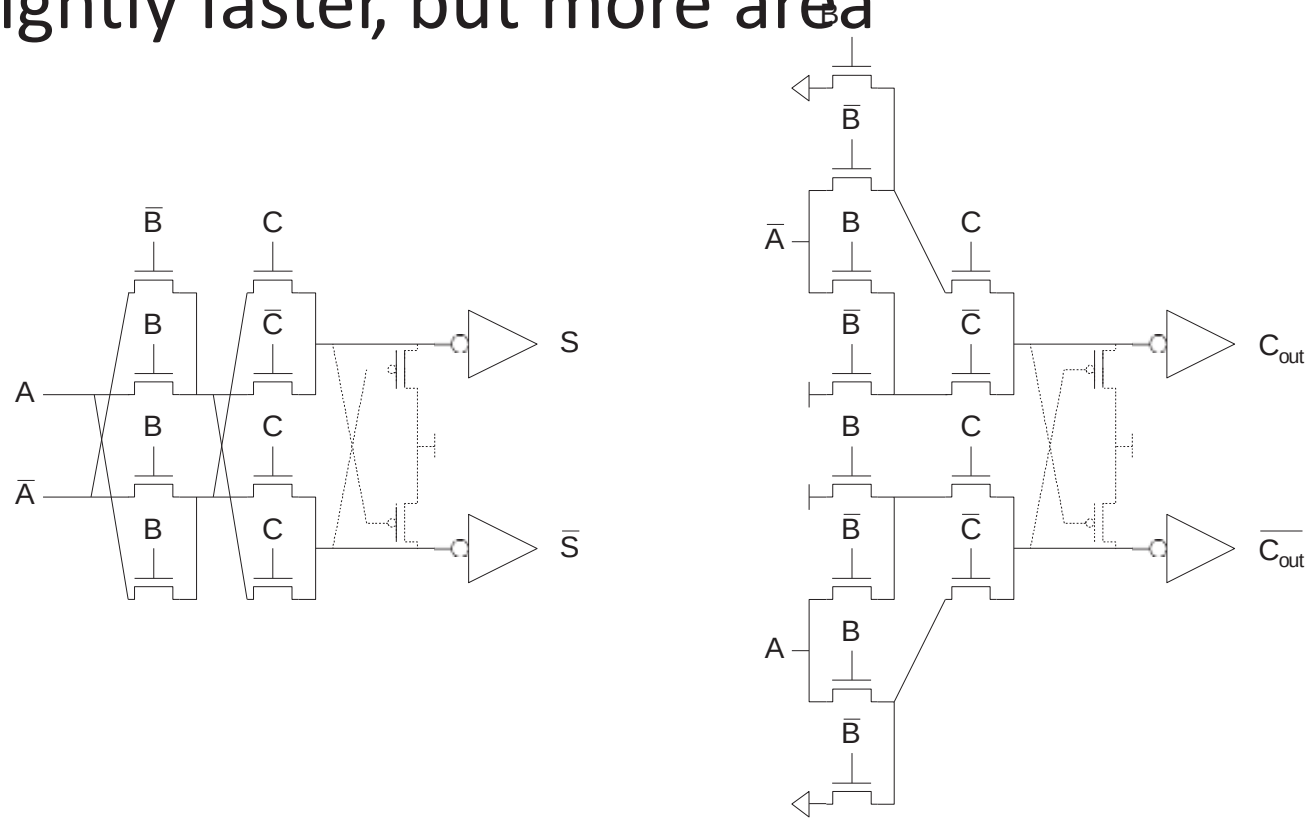
Layout

- Clever layout circumvents usual line of diffusion
  - Use wide transistors on critical path
  - Eliminate output inverters



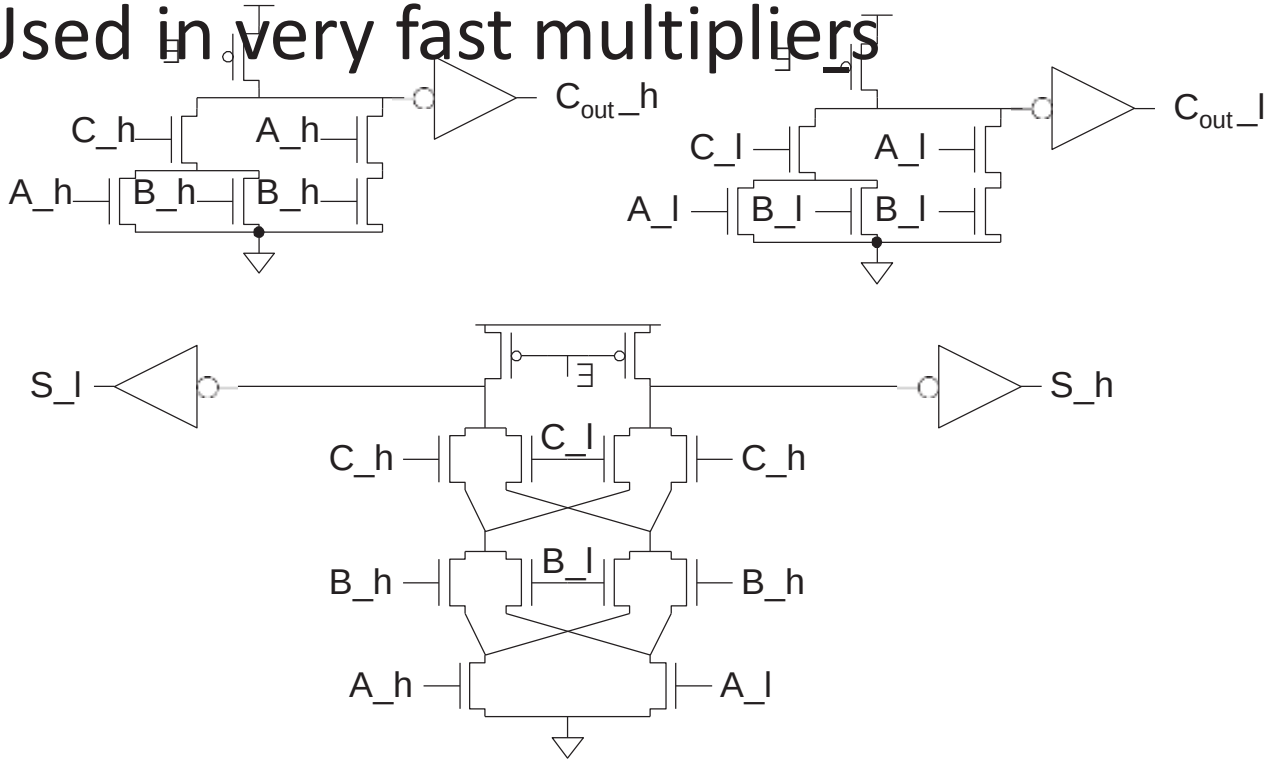
Full Adder Design III

- Complementary Pass Transistor Logic (CPL)
  - Slightly faster, but more area



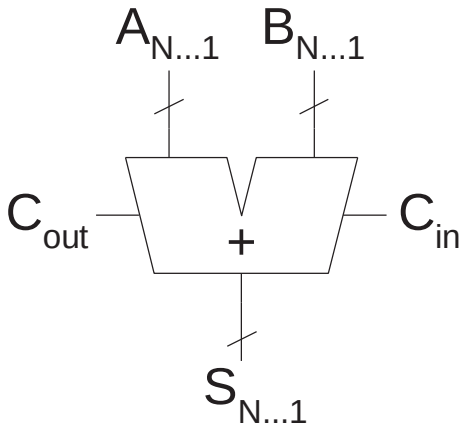
Full Adder Design IV

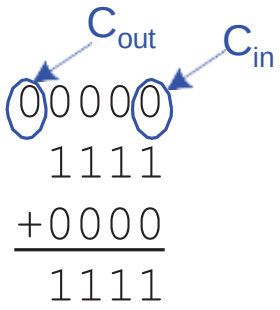
- Dual-rail domino
  - Very fast, but large and power hungry
  - Used in very fast multipliers

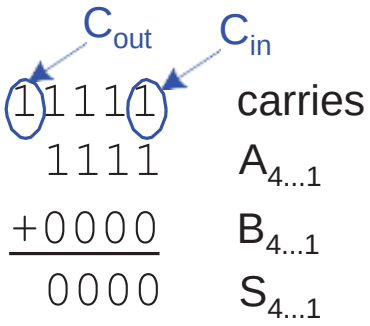


Carry Propagate Adders

- N-bit adder called CPA
  - Each sum bit depends on all previous carries
  - How do we compute all these carries quickly?

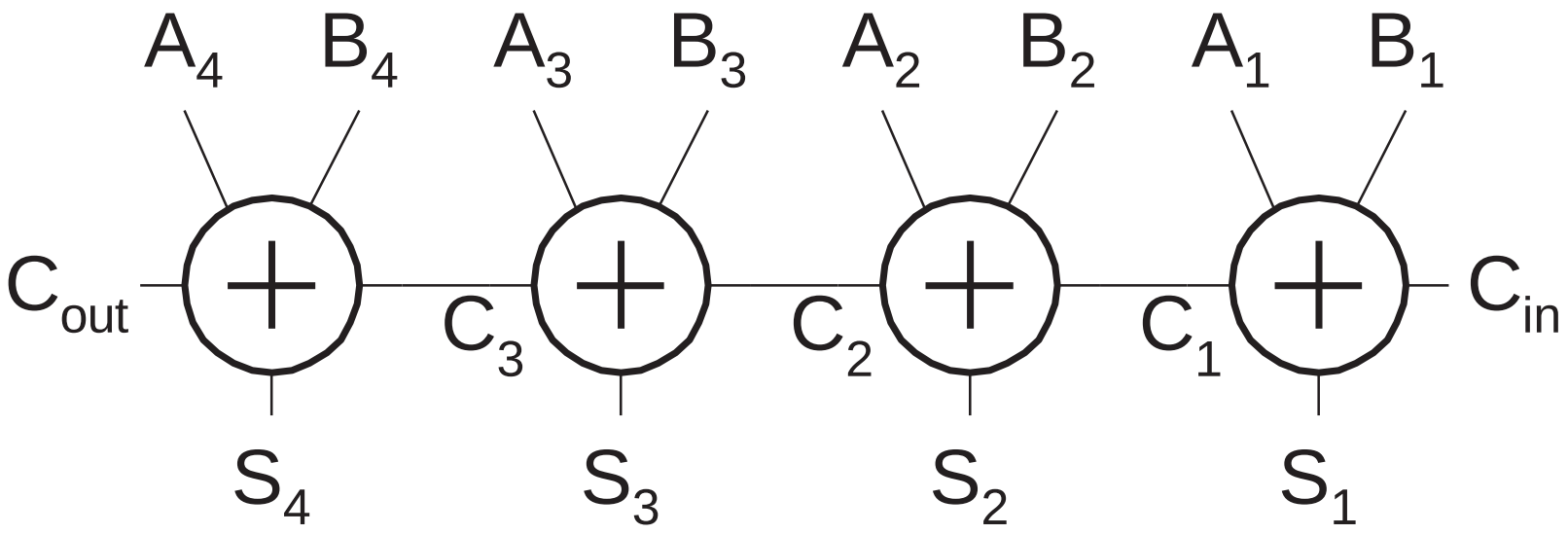






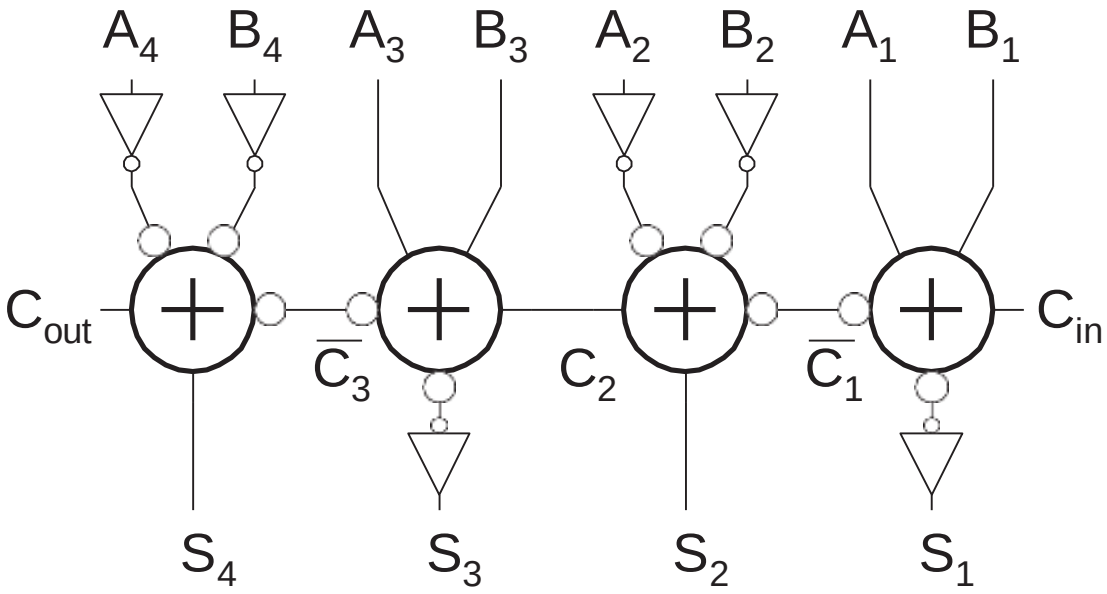
Carry-Ripple Adder

- Simplest design: cascade full adders
  - Critical path goes from Cin to Cout
  - Design full adder to have fast carry delay



Inversions

- Critical path passes through majority gate
  - Built from minority + inverter
  - Eliminate inverter and use inverting full adder



Generate / Propagate

- Equations often factored into G and P
- Generate and propagate for groups spanning i:j

$$G_{i:j} = G_{i:k} + P_{i:k} \cdot G_{k-1:j}$$

$$P_{i:j} = P_{i:k} \cdot P_{k-1:j}$$

- Base case

$$G_{i:i} = \div$$

$$G_{0:0} = \div$$

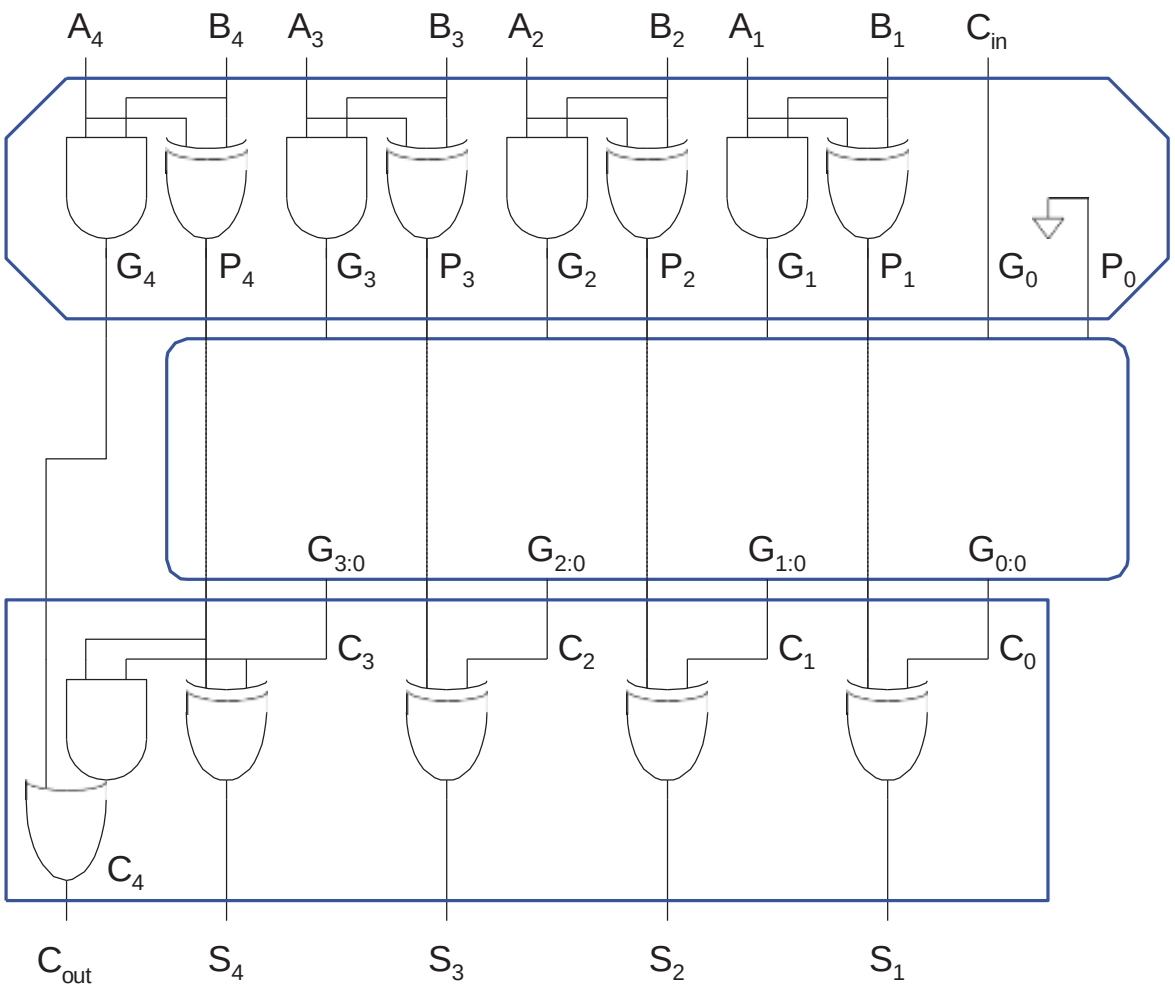
$$P_{i:i} = \div$$

$$P_{0:0} = \div$$

- Sum:

$$S_i =$$

PG Logic



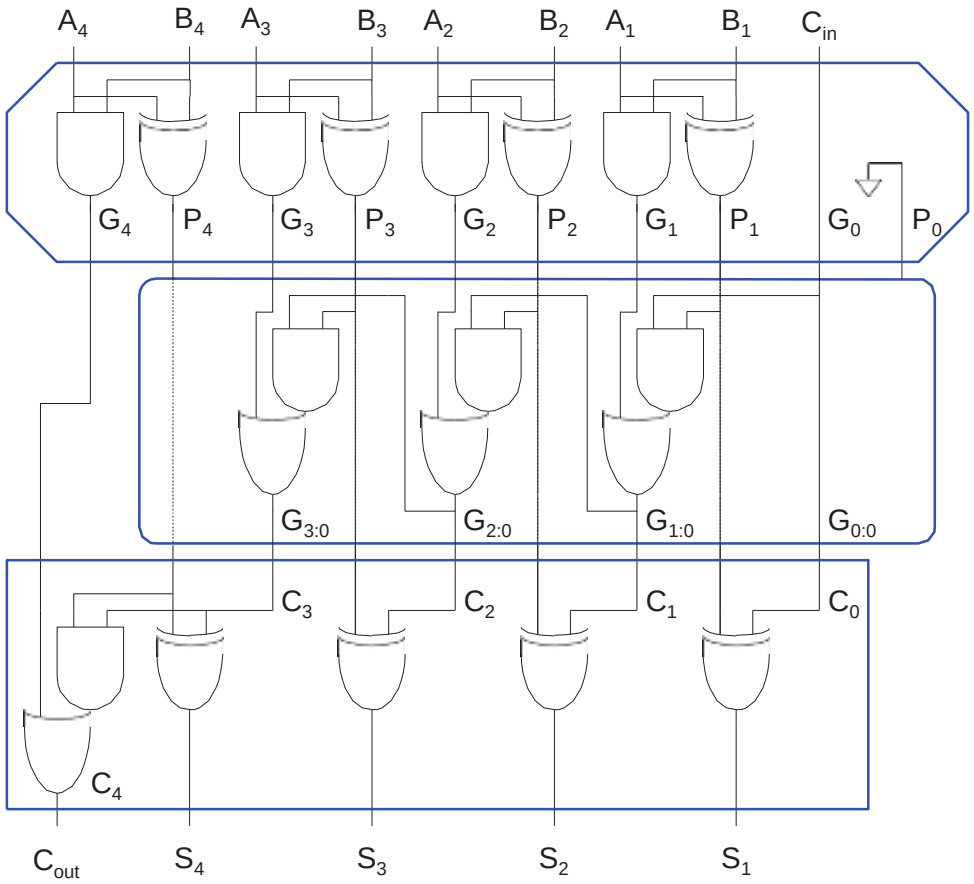
1: Bitwise PG logic

2: Group PG logic

3: Sum logic

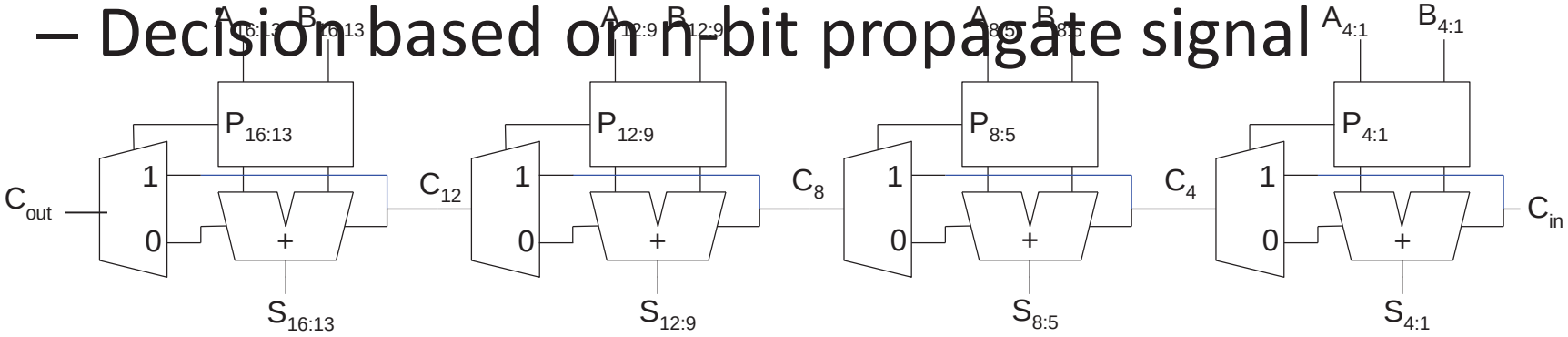
Carry-Ripple Revisited

$$G_{i:0} = G_i + P_i \cdot G_{i-1:0}$$



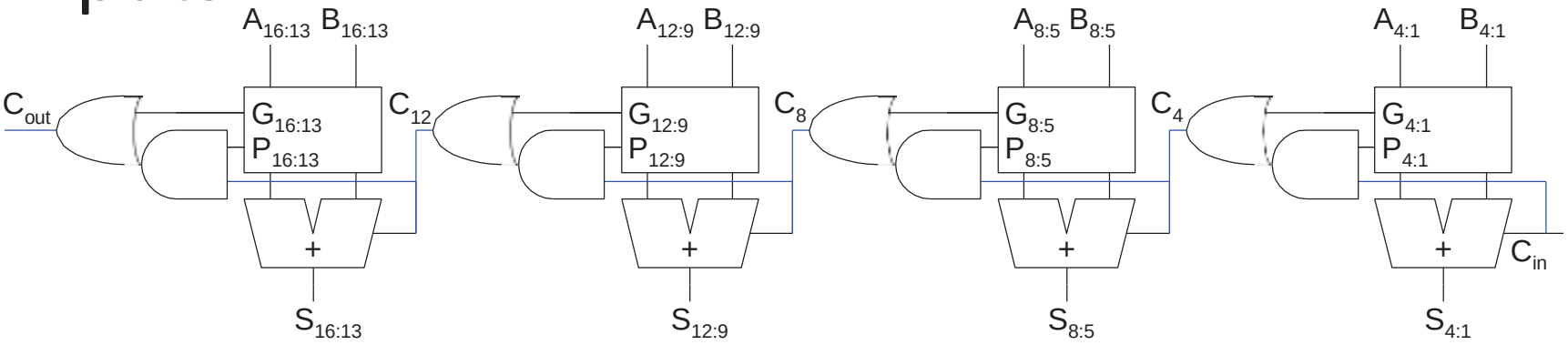
Carry-Skip Adder

- Carry-ripple is slow through all N stages
- Carry-skip allows carry to skip over groups of n bits
  - Decision based on n-bit propagate signal



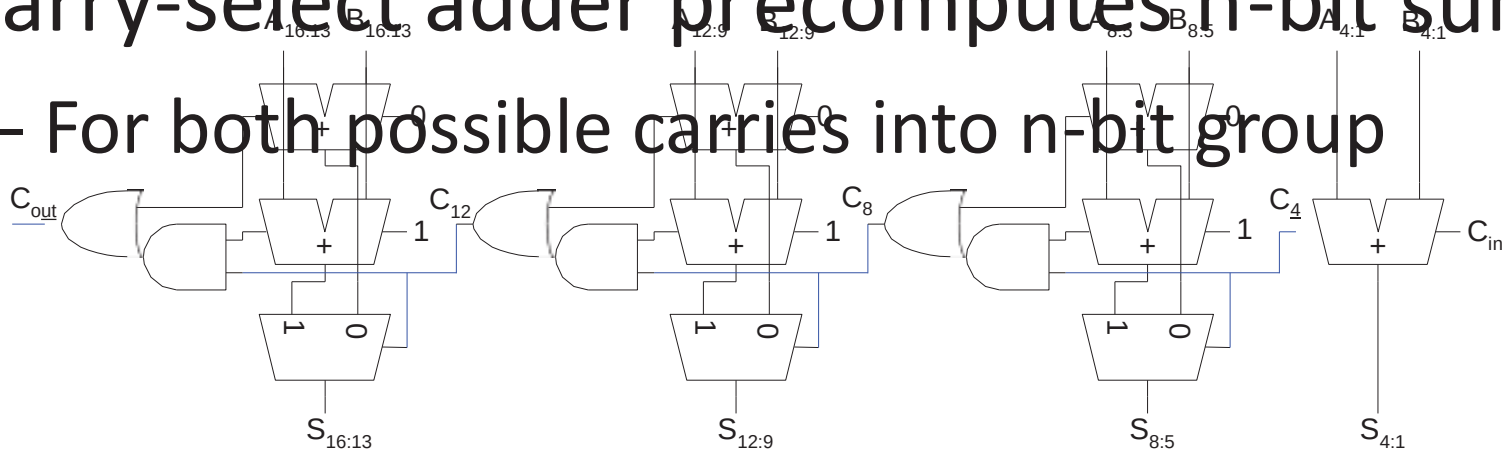
Carry-Lookahead Adder

- Carry-lookahead adder computes Gi:0 for many bits in parallel.
- Uses higher-valency cells with more than two inputs.



Carry-Select Adder

- Trick for critical paths dependent on late input X
  - Precompute two possible outputs for X = 0, 1
  - Select proper output when X arrives
- Carry-select adder precomputes n-bit sums
  - For both possible carries into n-bit group



Tree Adder

- If lookahead is good, lookahead across lookahead!
  - Recursive lookahead gives O(log N) delay
- Many variations on tree adders

Tree Adder Taxonomy

- Ideal N-bit tree adder would have
  - L = log N logic levels
  - Fanout never exceeding 2
  - No more than one wiring track between levels
- Describe adder with 3-D taxonomy (l, f, t)
  - Logic levels: L + l
  - Fanout: 2f + 1
  - Wiring tracks: 2t
- Known tree adders sit on plane defined by
- $l + f + t = L-1$

Summary

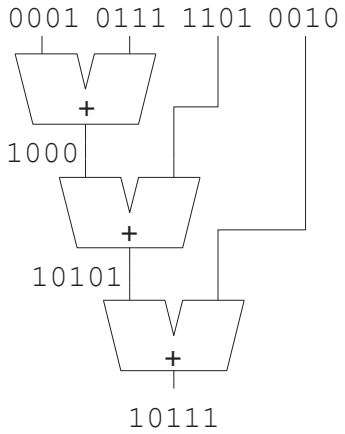
Adder architectures offer area / power / delay tradeoffs.

Choose the best one for your application.

Architecture	**Classification**	**Logic Levels**	**Max Fanout**	**Tracks**	**Cells**
Carry-Ripple		N-1	1	1	N
Carry-Skip n=4		N/4 + 5	2	1	1.25N
Carry-Inc. n=4		N/4 + 2	4	1	2N
Brent-Kung	(L-1, 0, 0)	$2\log_2 N - 1$	2	1	2N
Sklansky	(0, L-1, 0)	$\log_2 N$	N/2 + 1	1	$0.5\ N\log_2 N$
Kogge-Stone	(0, 0, L-1)	$\log_2 N$	2	N/2	$N\log_2 N$

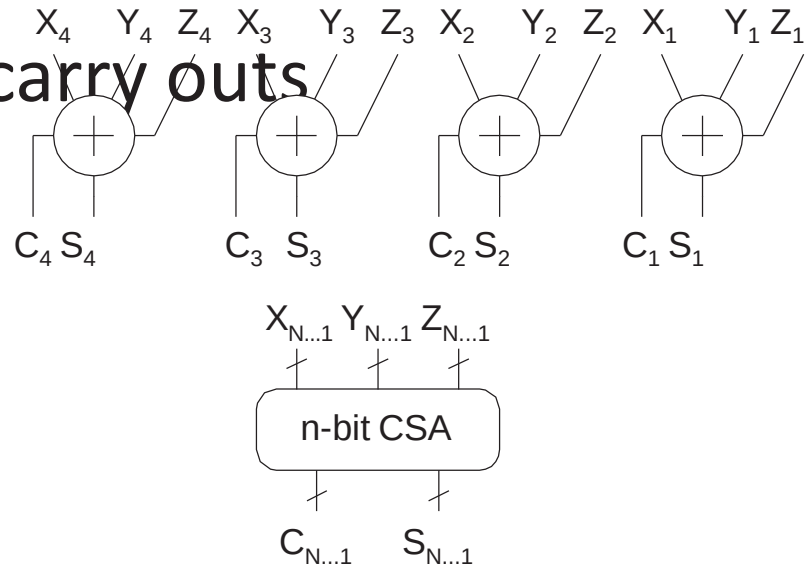
Multi-input Adders

- Suppose we want to add k N-bit words
  - Ex: 0001 + 0111 + 1101 + 0010 = 10111
- Straightforward solution: k-1 N-input CPAs
  - Large and slow



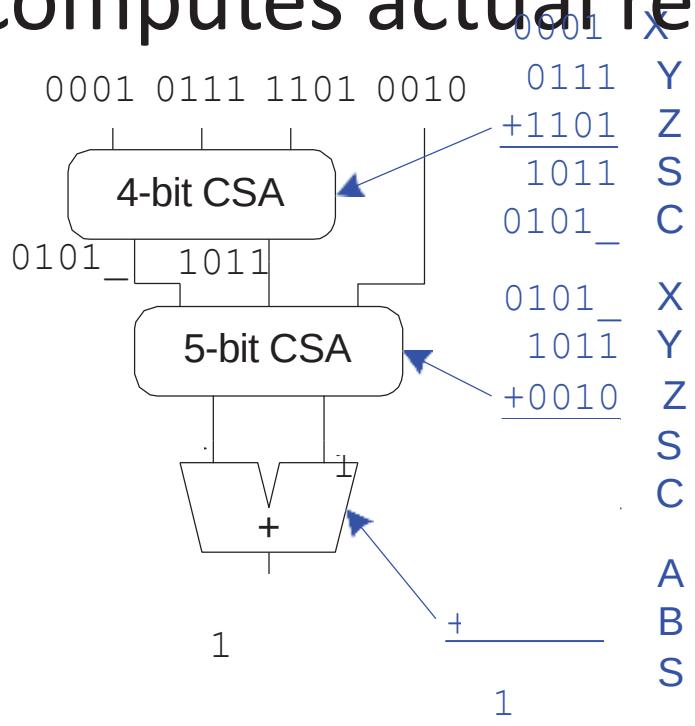
Carry Save Addition

- A full adder sums 3 inputs and produces 2 outputs
  - Carry output has twice weight of sum output
- N full adders in parallel are called carry save adder
  - Produce N sums and N carry outs



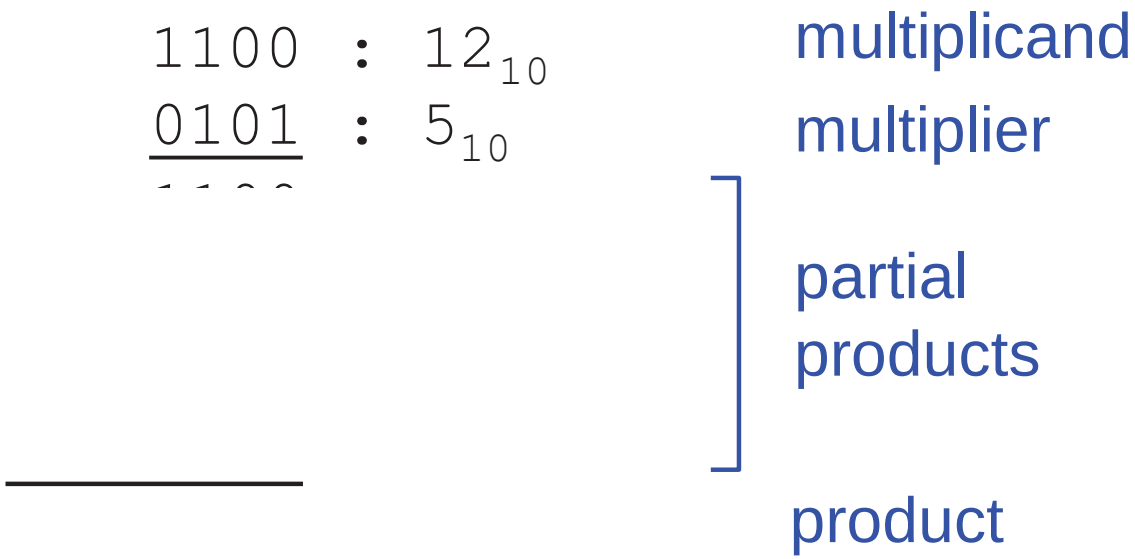
CSA Application

- Use k-2 stages of CSAs
  - Keep result in carry-save redundant form
- Final CPA computes actual result



Multiplication

- Example:



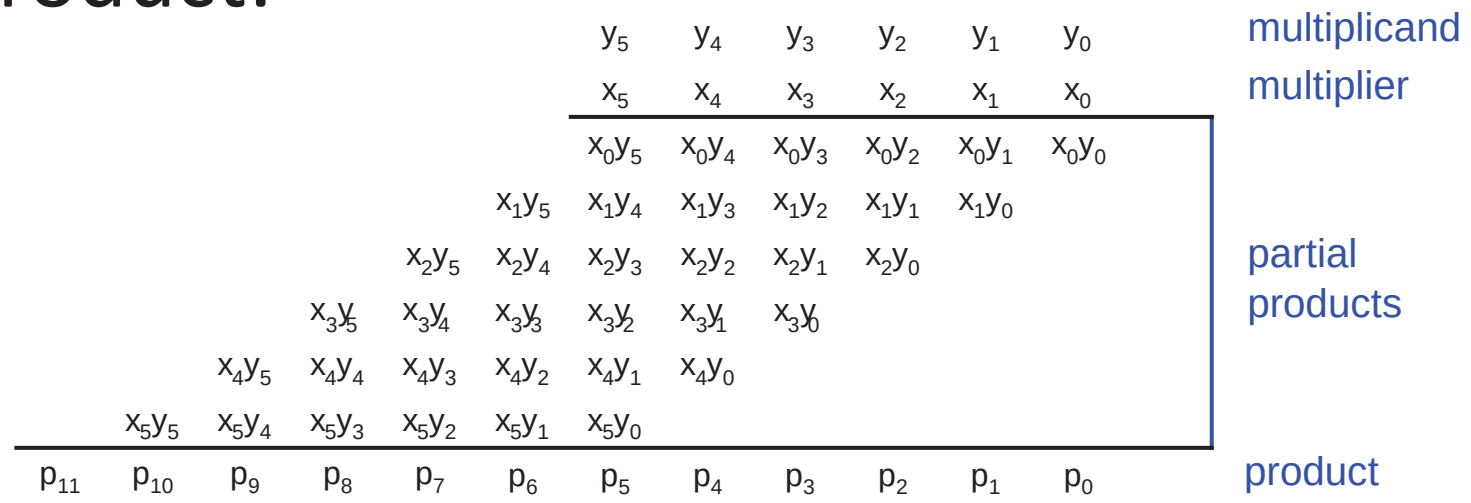
- M x N-bit multiplication
  - Produce N M-bit partial products
  - Sum these to produce M+N-bit product

General Form

- Multiplicand: Y = (yM-1, yM-2, …, y1, y0)
- Multiplier: X = (xN-1, xN-2, …, x1, x0)

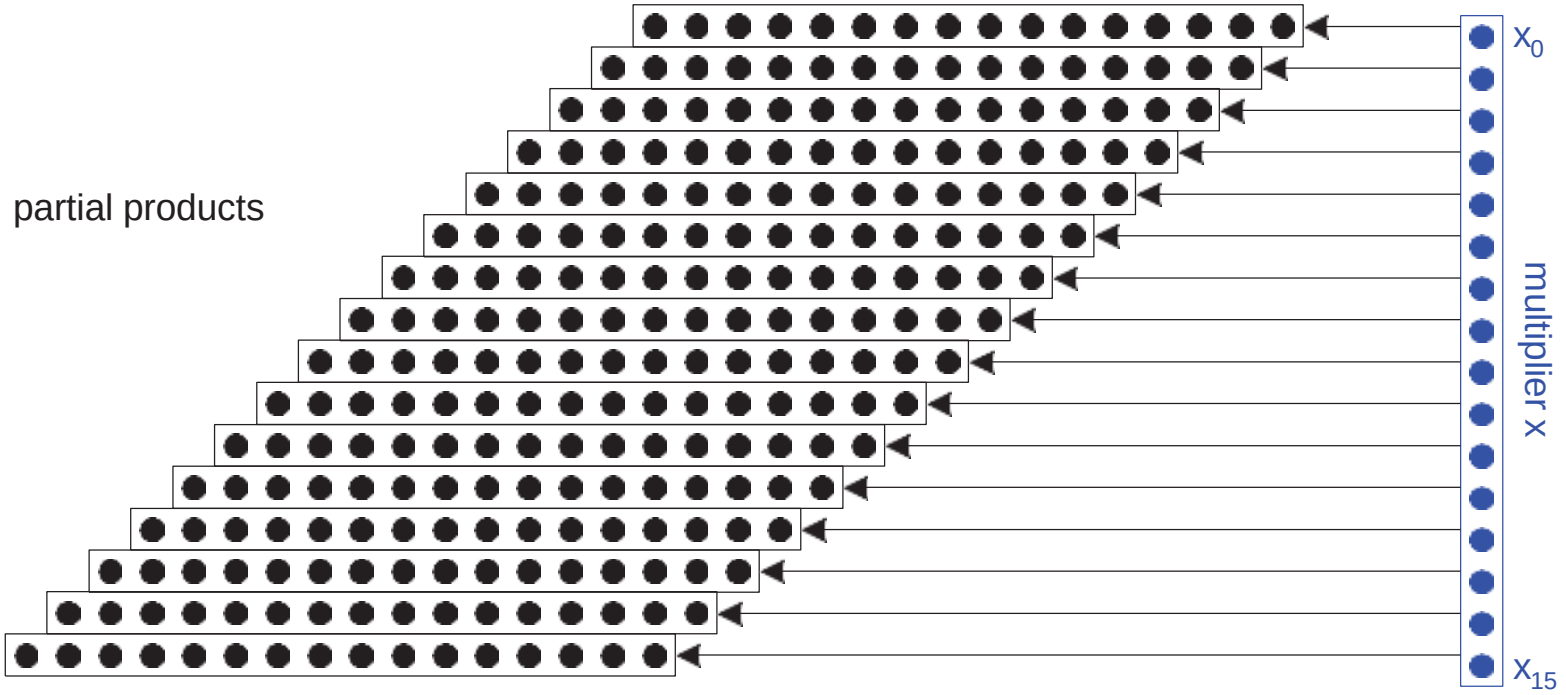
$$P = \left\{ \sum_{j=0}^{M-1} y_j 2^j \right\} \left\{ \sum_{i=0}^{N-1} x_i 2^i \right\} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} x_i y_j 2^{i+j}$$

- Product:

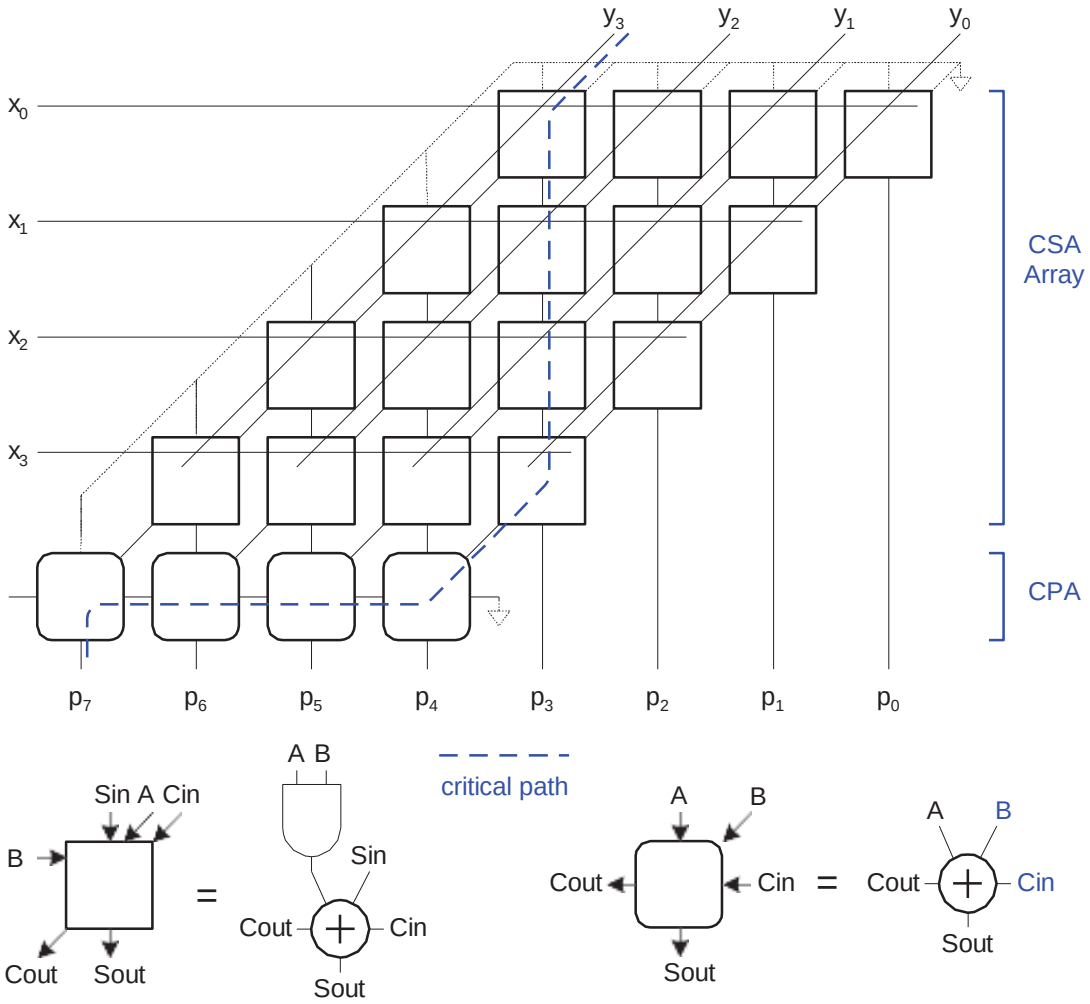


Dot Diagram

- Each dot represents a bit

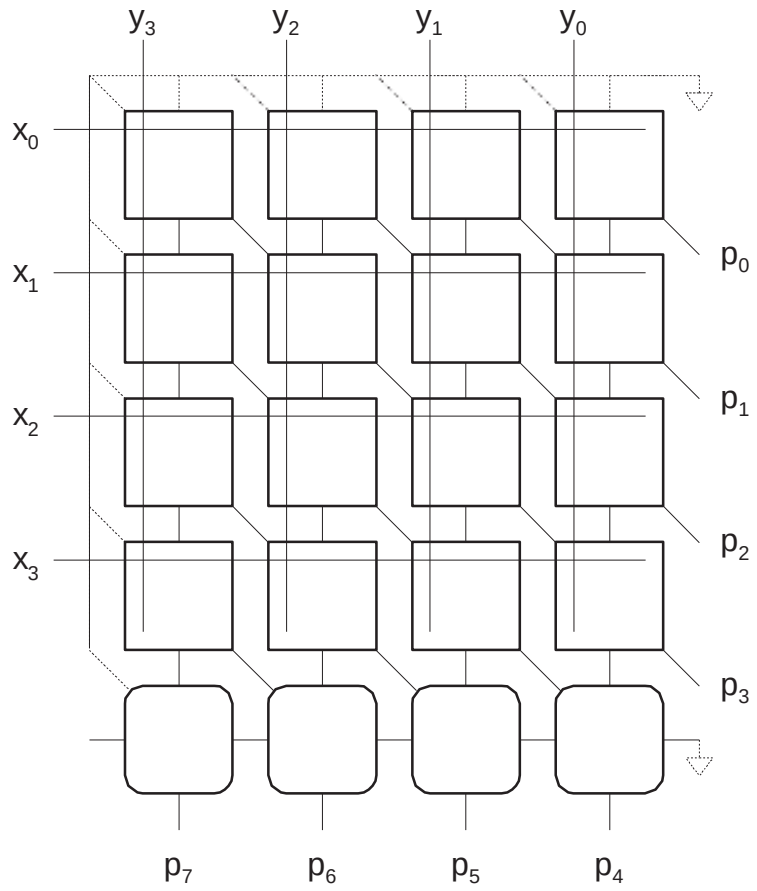


Array Multiplier



Rectangular Array

- Squash array to fit rectangular floorplan



Fewer Partial Products

- Array multiplier requires N partial products
- If we looked at groups of r bits, we could form N/r partial products.
  - Faster and smaller?
  - Called radix-2r encoding
- Ex: r = 2: look at pairs of bits
  - Form partial products of 0, Y, 2Y, 3Y
  - First three are easy, but 3Y requires adder ☹

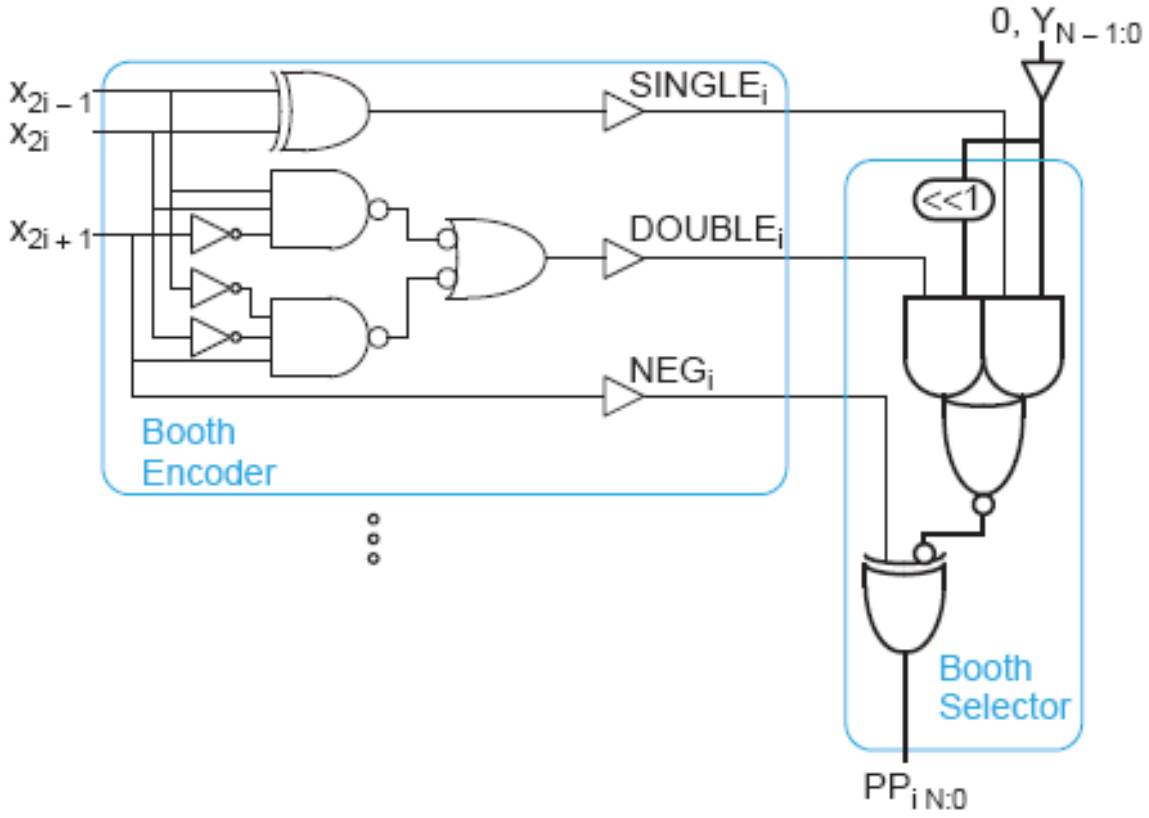
Booth Encoding

- Instead of 3Y, try –Y, then increment next partial product to add 4Y
- Sir

Inputs			Partial Product	Booth Selects		
x_{2i+1}	x_{2i}	x_{2i-1}	PP_i	SINGLE_i	DOUBLE_i	NEG_i
0	0	0	0	0	0	0
0	0	1	Y	1	0	0
0	1	0				0
0	1	1		0	1	0
1	0	0	$-2Y$	0	1	
1	0	1				
1	1	0				
1	1	1	$-0\ (=0)$	0		

Booth Hardware

- Booth encoder generates control lines for each PP
  - Boo



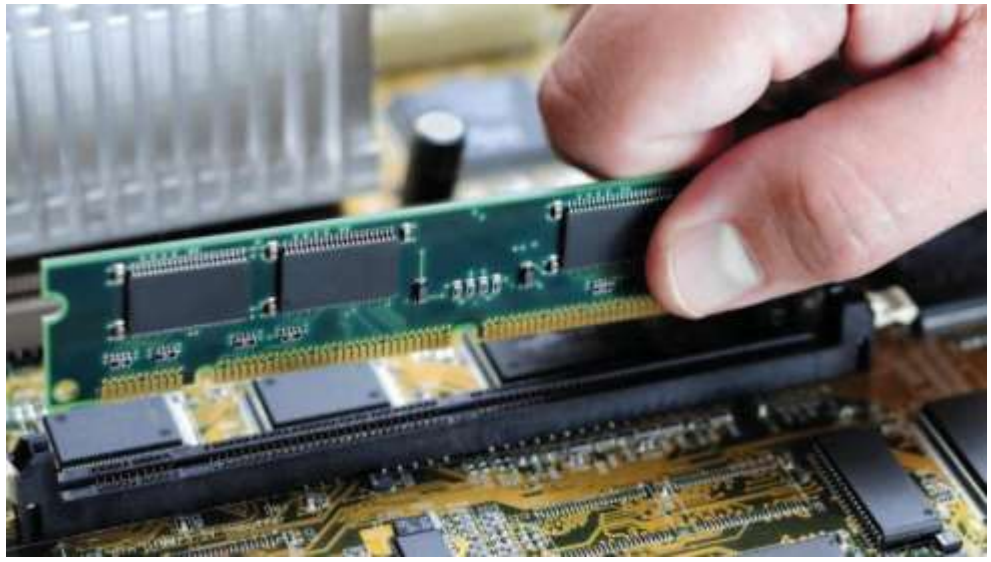
Advanced Multiplication

- Signed vs. unsigned inputs
- Higher radix Booth encoding
- Array vs. tree CSA networks

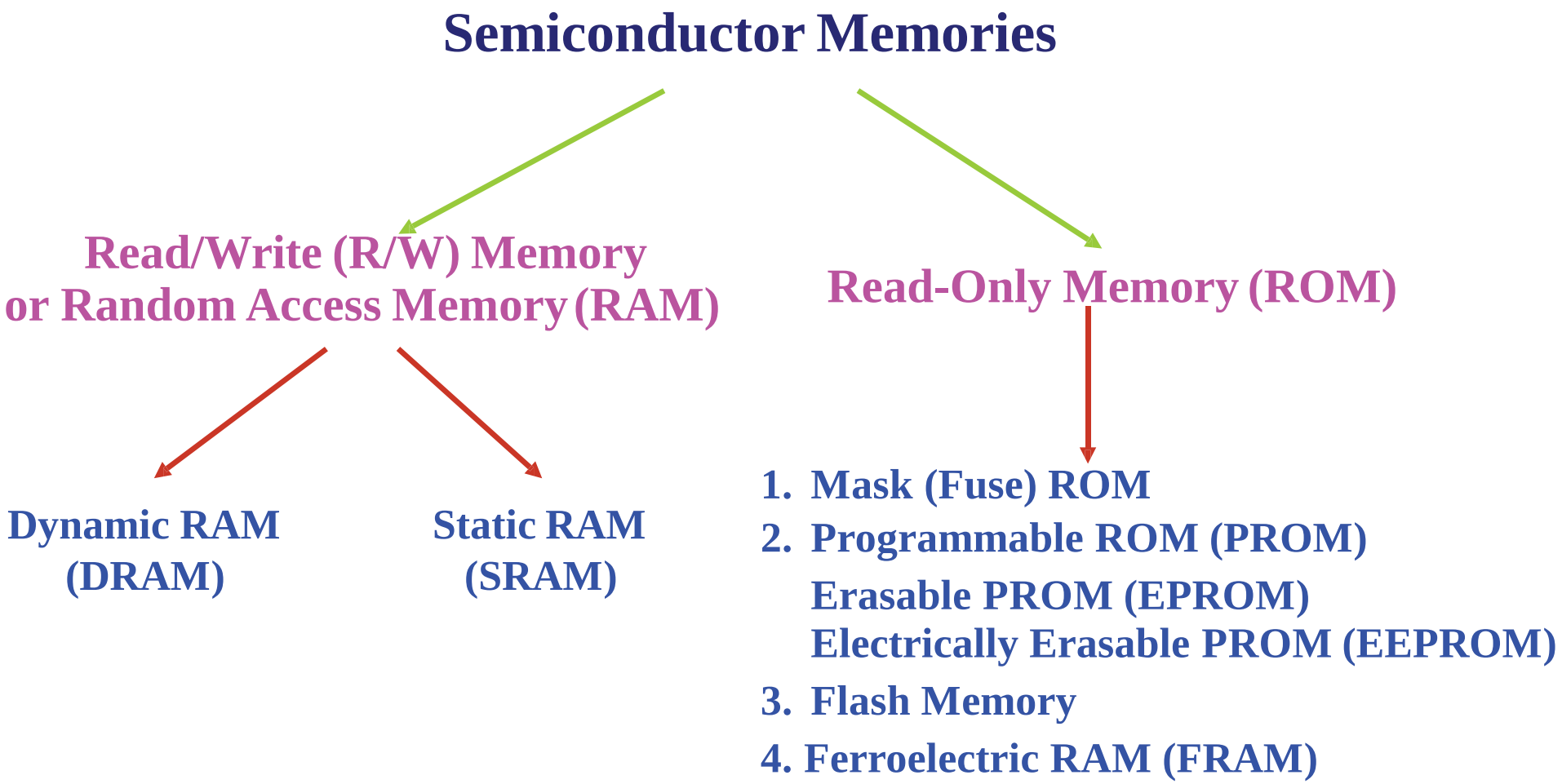
Unit VI
Array Sub Systems

Topics:

- SRAM
- DRAM
- ROM
- Serial Access Memories
- Content Addressable Memory



Semiconductor Memory Types



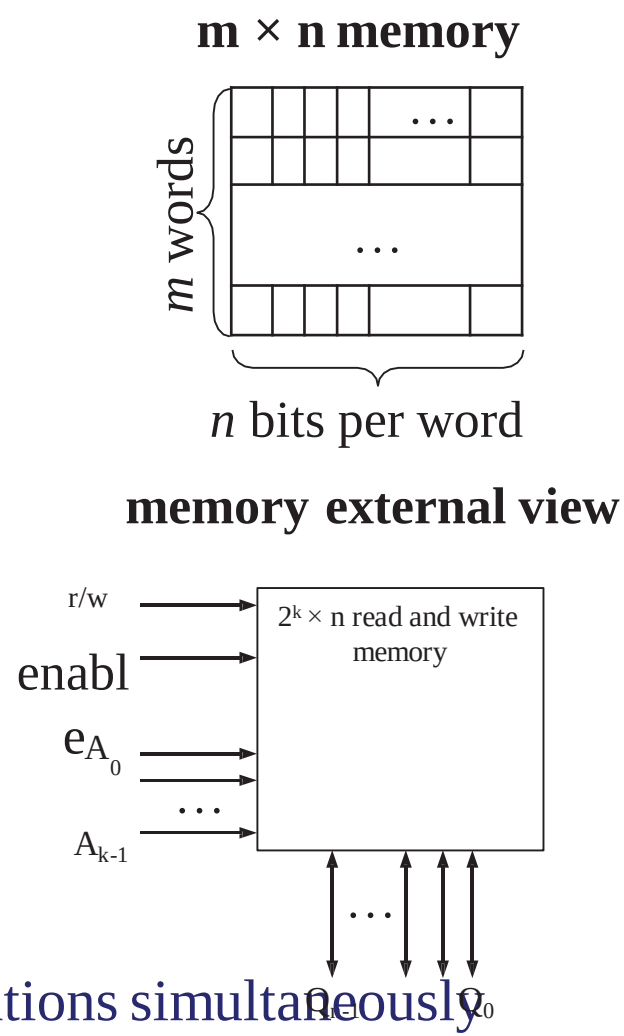
Semiconductor Memory Types (Cont.)

- Design Issues
  - » Area Efficiency of Memory Array: # of stored data bits per unit area
  - » Memory Access Time: the time required to store and/or retrieve a particular data bit.
  - » Static and Dynamic Power Consumption

- Requirements
  - » Easy reading
  - » Easy Writing
  - » High density
  - » Speed, more speed and still more speed

Memory Architecture

- Stores large number of bits
  - m x n: m words of n bits each
  - k = Log2(m) address input signals
  - or m = 2k words
  - e.g., 4,096 x 8 memory:
    - » 32,768 bits
    - » 12 address input signals
    - » 8 input/output data signals
- Memory access
  - r/w: selects read or write
  - enable: read or write only when asserted
  - multiport: multiple accesses to different locations simultaneously



Semiconductor Memory Types (Cont.)

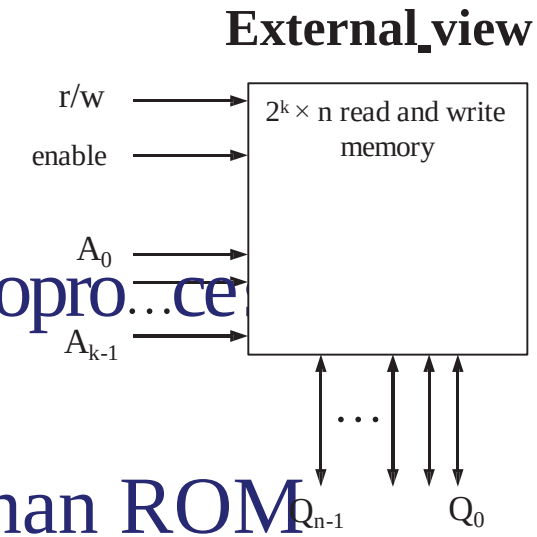
- RAM: the stored data is volatile

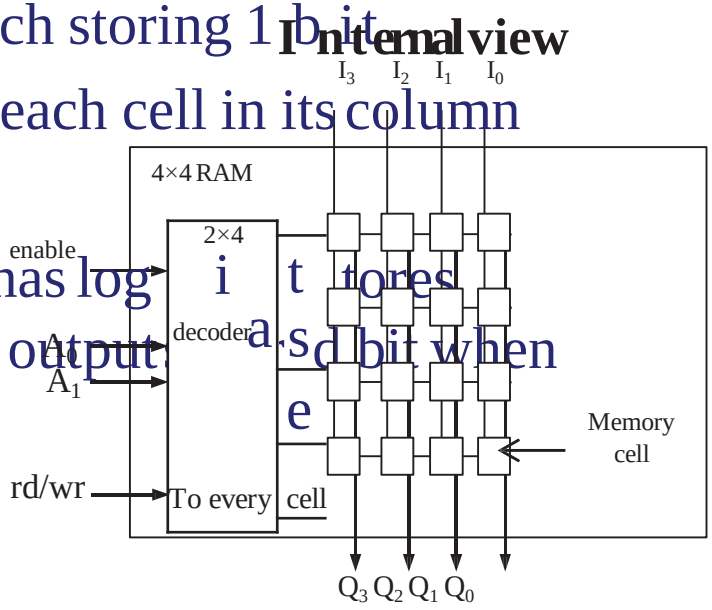
  - DRAM
    - » A capacitor to store data, and a transistor to access the capacitor
    - » Need refresh operation
    - » Low cost, and high density → it is used for main memory

  - SRAM
    - » Consists of a latch
    - » Don't need the refresh operation
    - » High speed and low power consumption →it is mainly used for cache memory and memory in hand-held devices

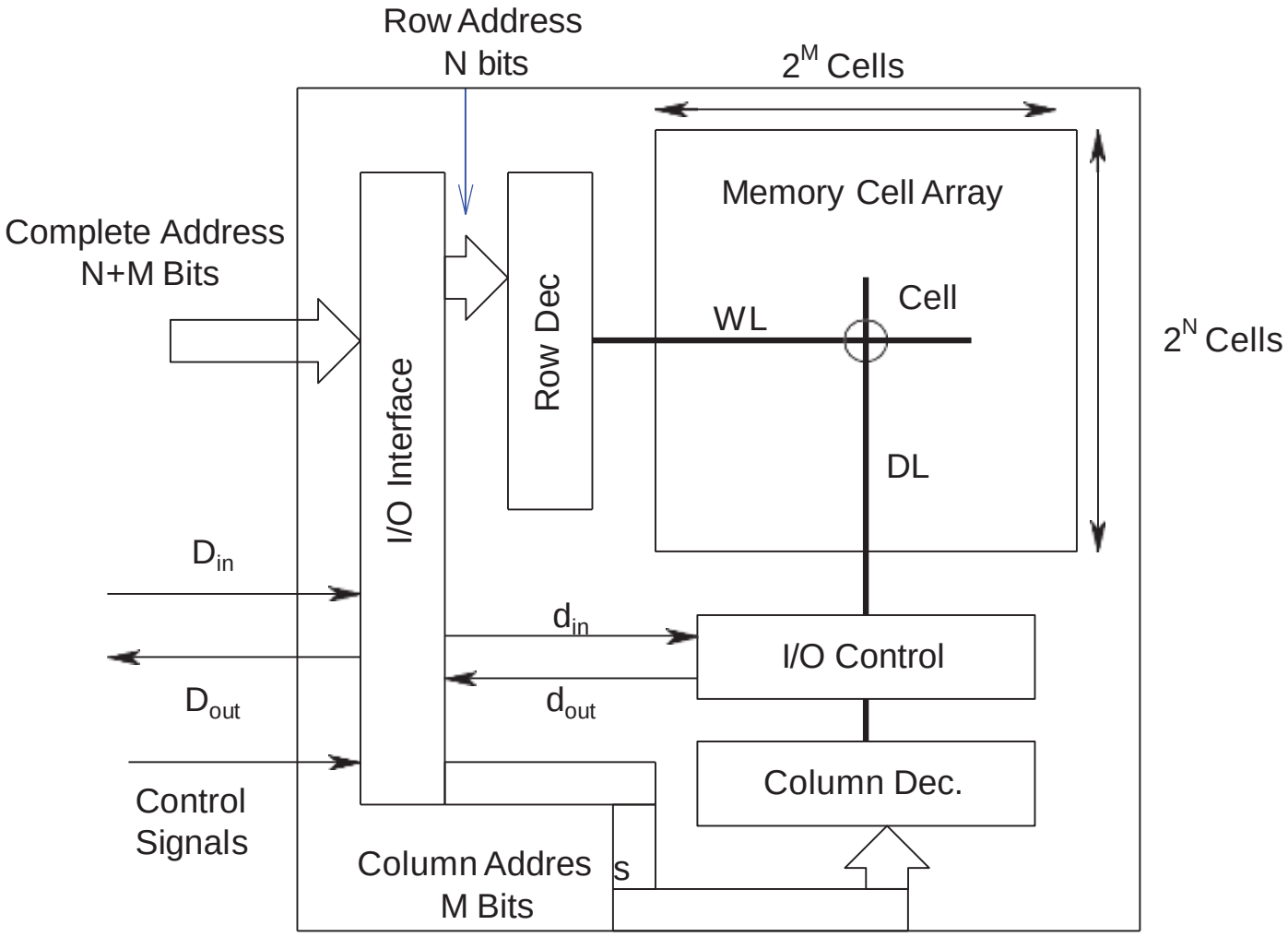
RAM: "Random-access" memory

- Typically volatile memory
  - bits are not held without power supply
- Read and written to easily by microprocessor during execution
- Internal structure more complex than ROM
  - a word consists of several memory cells, each storing 1 bit
  - each input and output data line connects to each cell in its column
  - rd/wr connected to every cell
  - when row is enabled by decoder, each cell has logic to store input data bit when rd/wr indicates write or output data bit when rd/wr indicates read



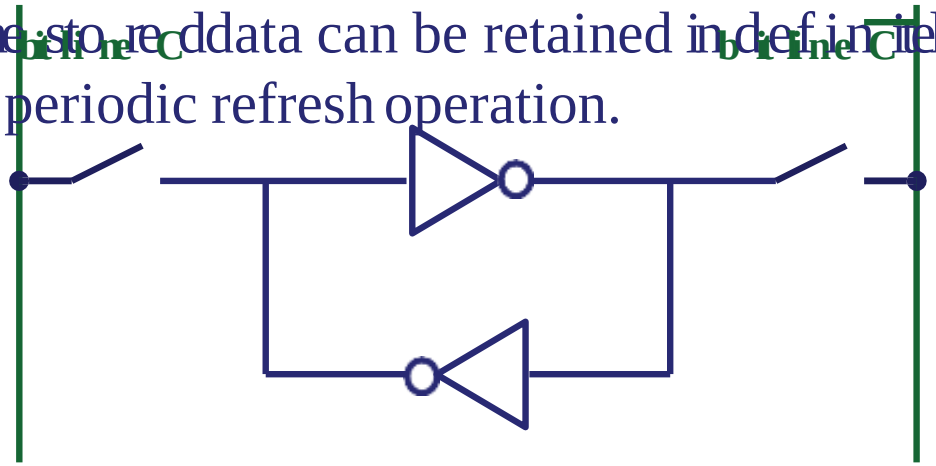


Memory Chip Configuration

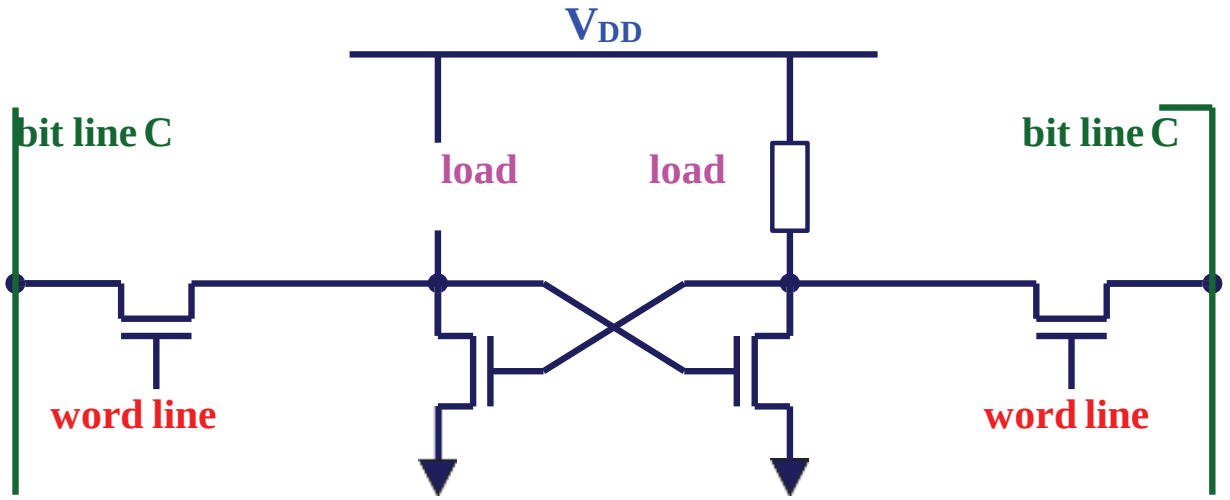


Static Random Access Memory (SRAM)

- SRAM: The stored data can be retained indefinitely, without any need for a periodic refresh operation.

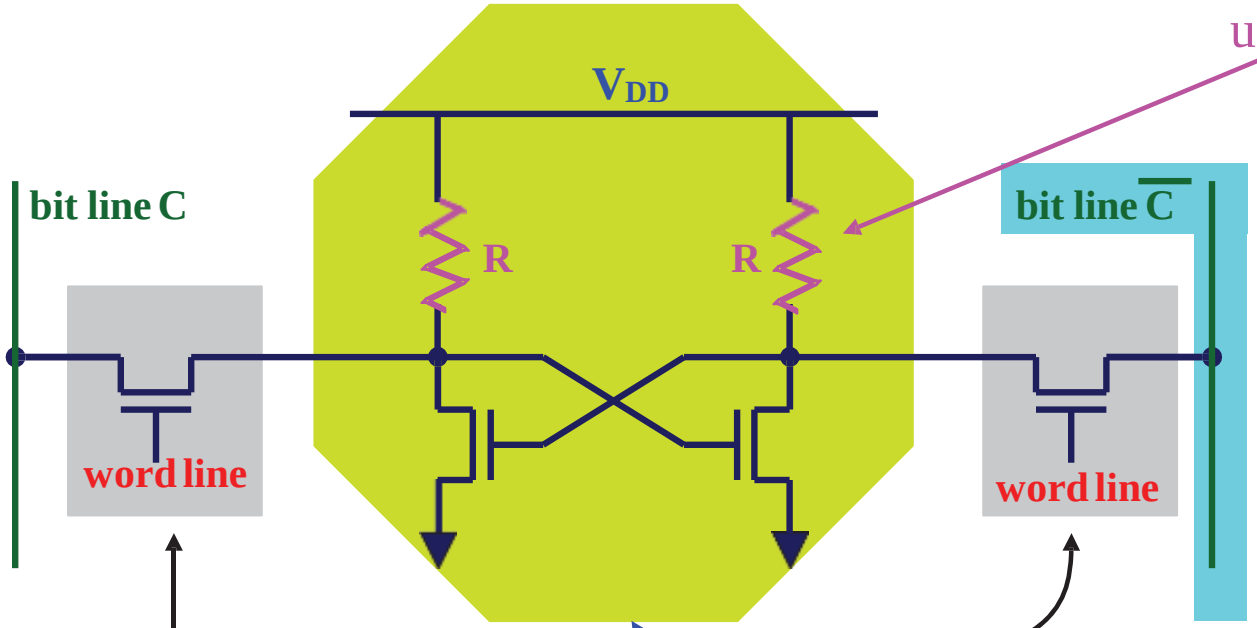


1-bit SRAM cell

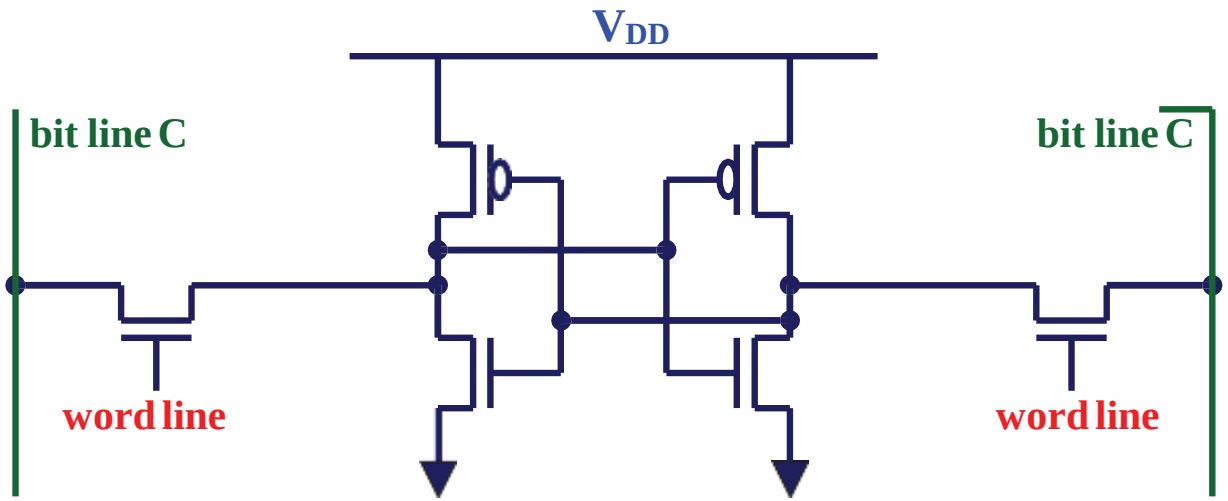


- **Complementary Column** arrangement is to achieve a more reliable SRAM operation

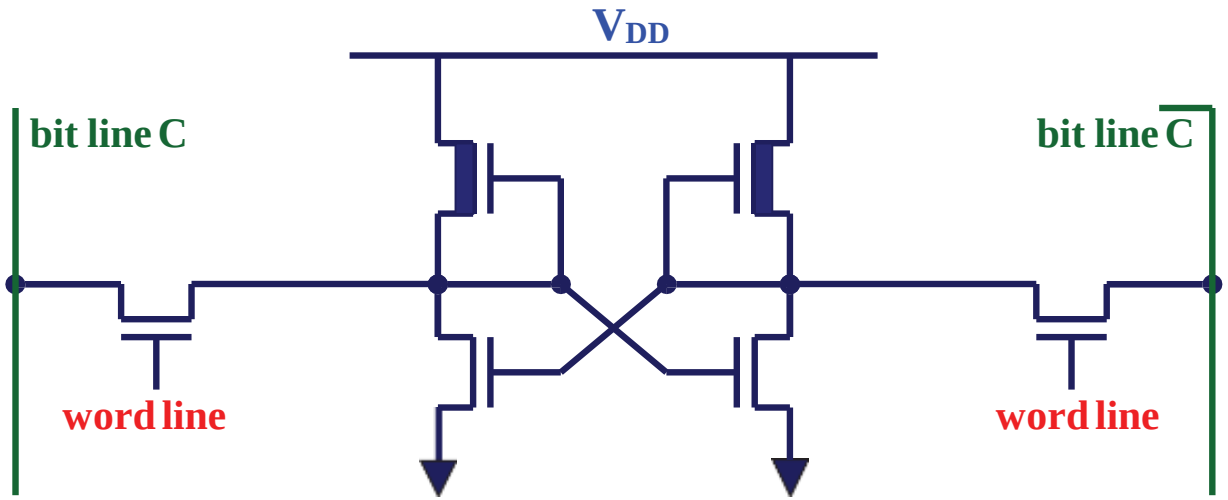
4T-SRAM(Resistive-Load SRAM Cell)



6T-SRAM



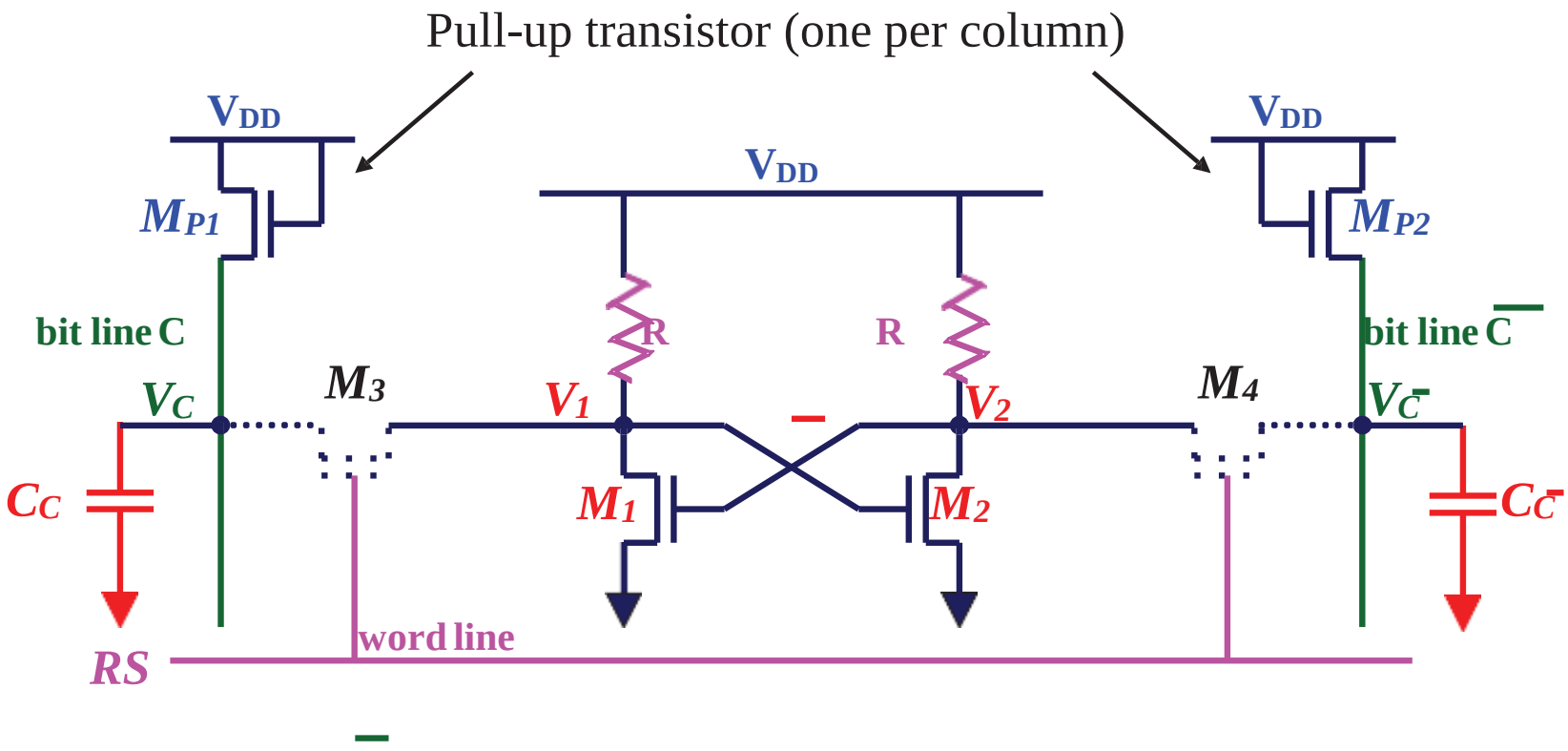
Full CMOS SRAM Cell



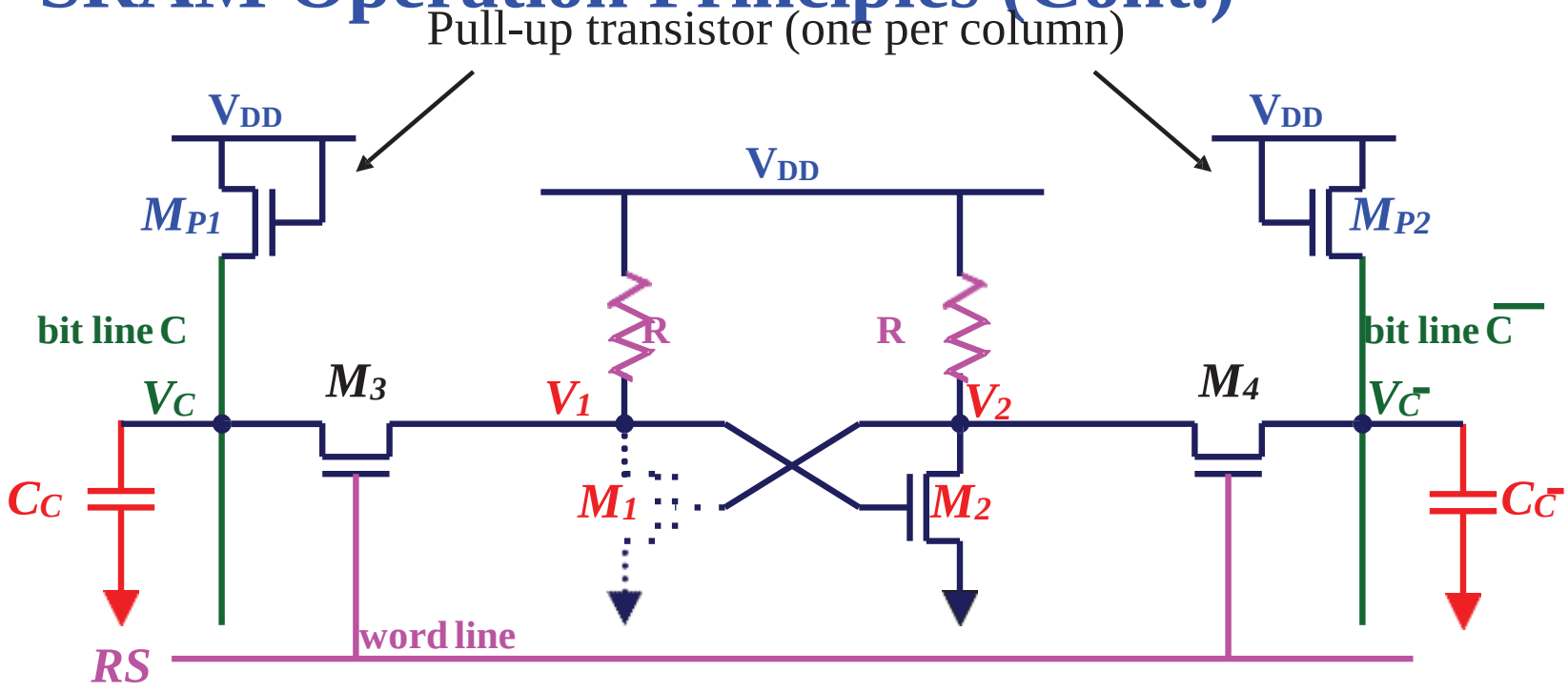
Depletion-Load SRAM Cell

SRAM Operation Principles

- RS=0: The word line is not selected. M3 and M4 are OFF
- One data-bit is held: The latch preserves one of its two stable states.
- If RS=0 for all rows: CC and CC are charged up to near VDD by pulling up of MP1 and MP2 (both in saturation)

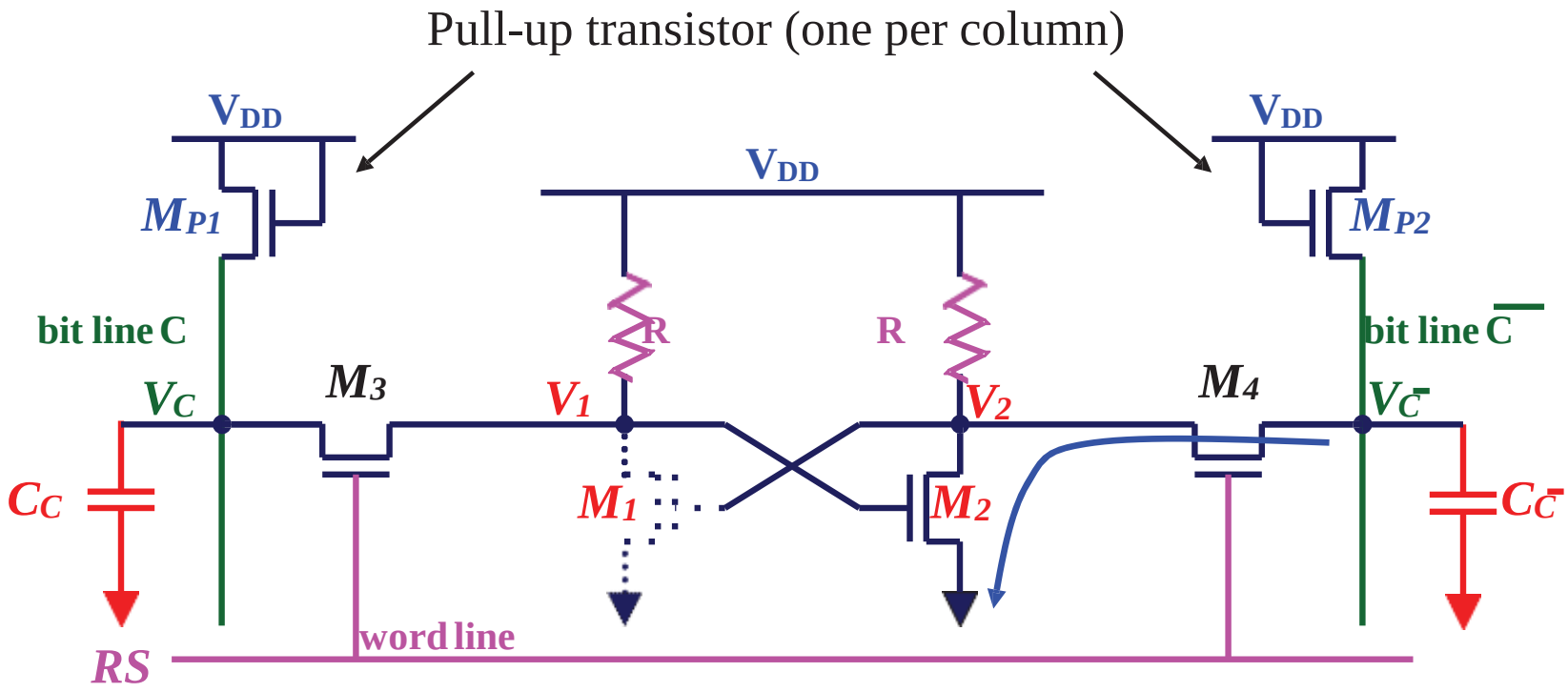


SRAM Operation Principles (Cont.)



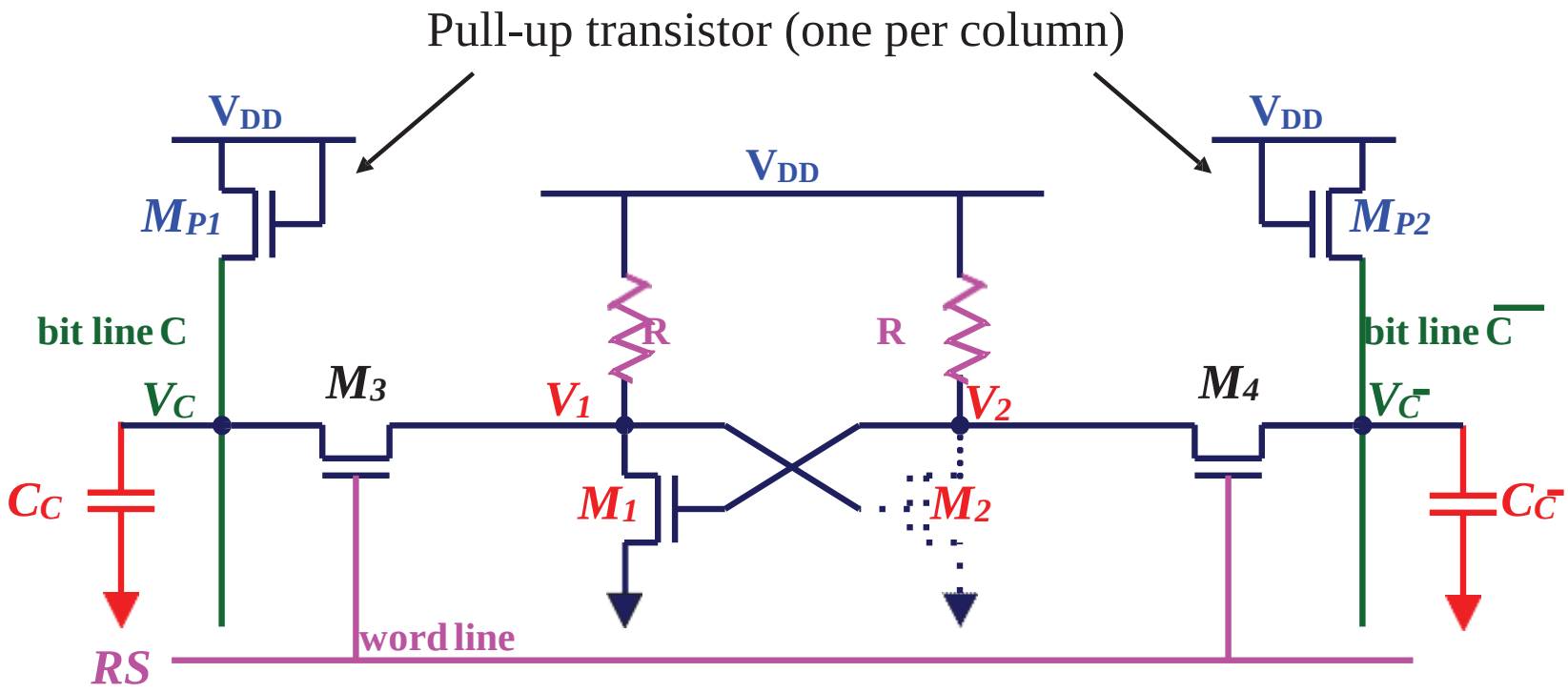
- RS=1: The word line is now selected. M3 and M4 are ON
- Four Operations
- Write "1" Operation (V1=VOL, V2=VOH at t=0-):
- VC → VOL by the data-write circuitry. Therefore, V2 → VOL, then M1 tur off $\bar{V}1$→VOH and M2 turns on pulling down V2 → VOL.

SRAM Operation Principles (Cont.)



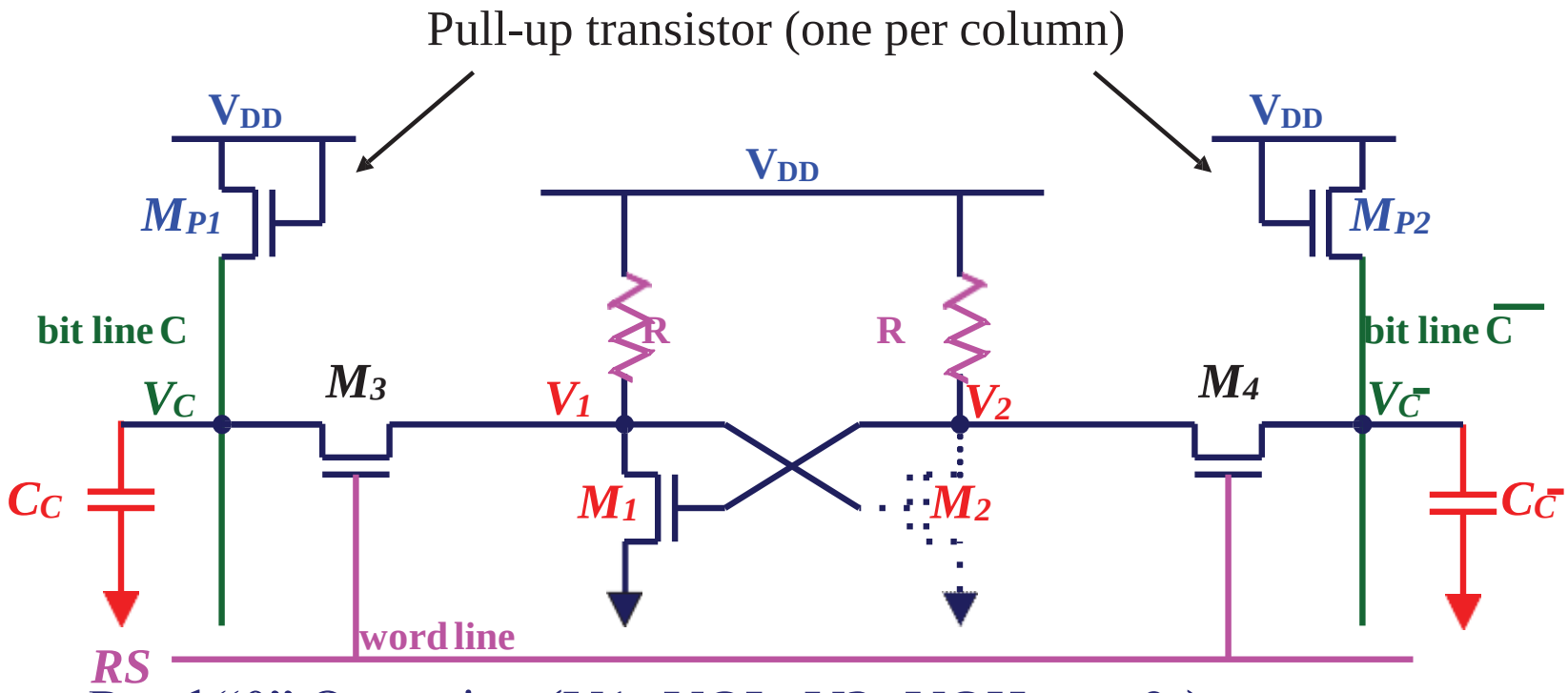
- Read "1" Operation (V1=VOH, V2=VOL at t=0-):
- VC retains pre-charge level, while VC → VOL by M2 ON. Data-read circuitry detects small voltage difference VC – VC > 0, and amplifies it as a "1" data output.

SRAM Operation Principles (Cont.)



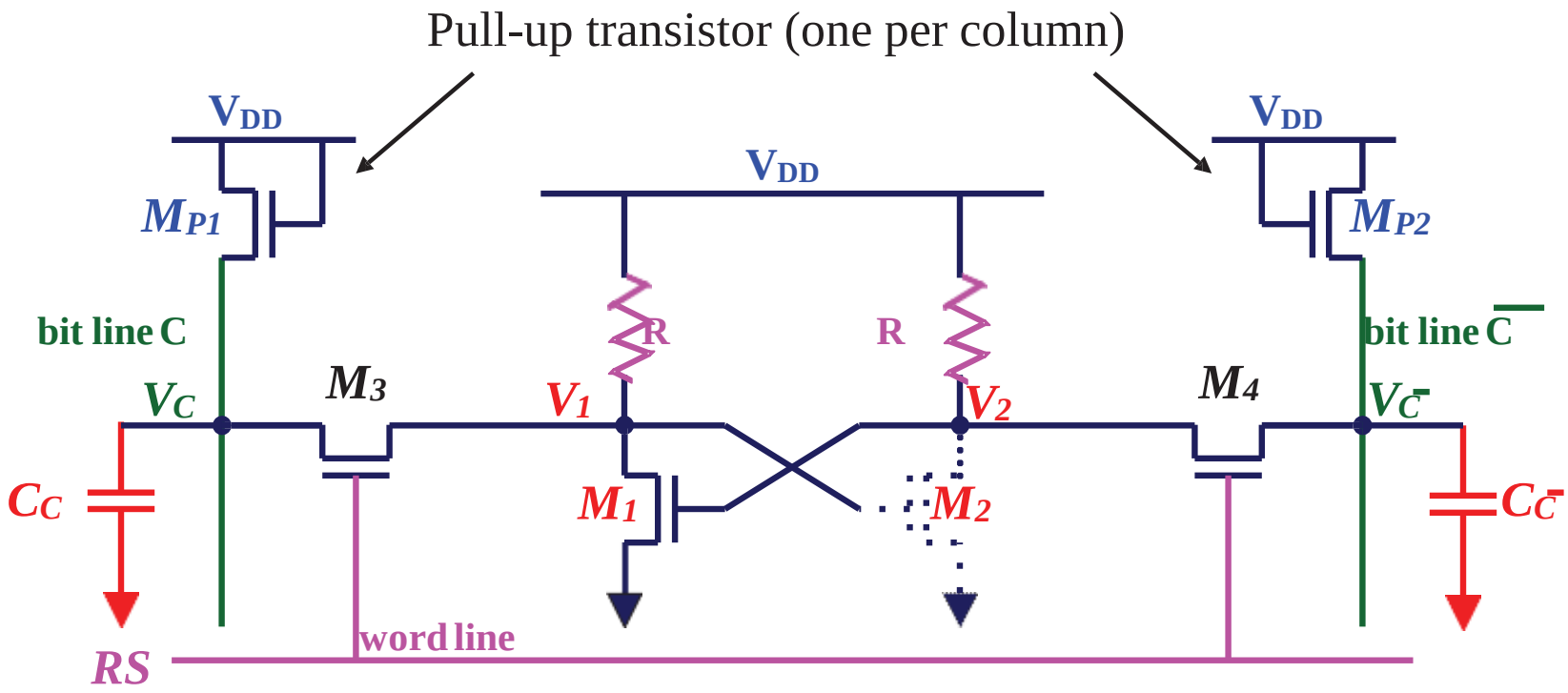
- Write "0" Operation (V1=VOH, V2=VOL at t=0-):
- VC → VOL by the data-write circuitry.
- Since V1 → VOL, M2 turns off, therefore V2 → VOH.

SRAM Operation Principles (Cont.)



- Read "0" Operation (V1=VOL, V2=VOH at t=0-):
- VC retains pre-charge level, while VC → VOL by M1 ON.
- Data-read circuitry detects small voltage difference VC – VC < 0, and amplifies it as a "0" data output.

SRAM Operation Principles (Cont.)

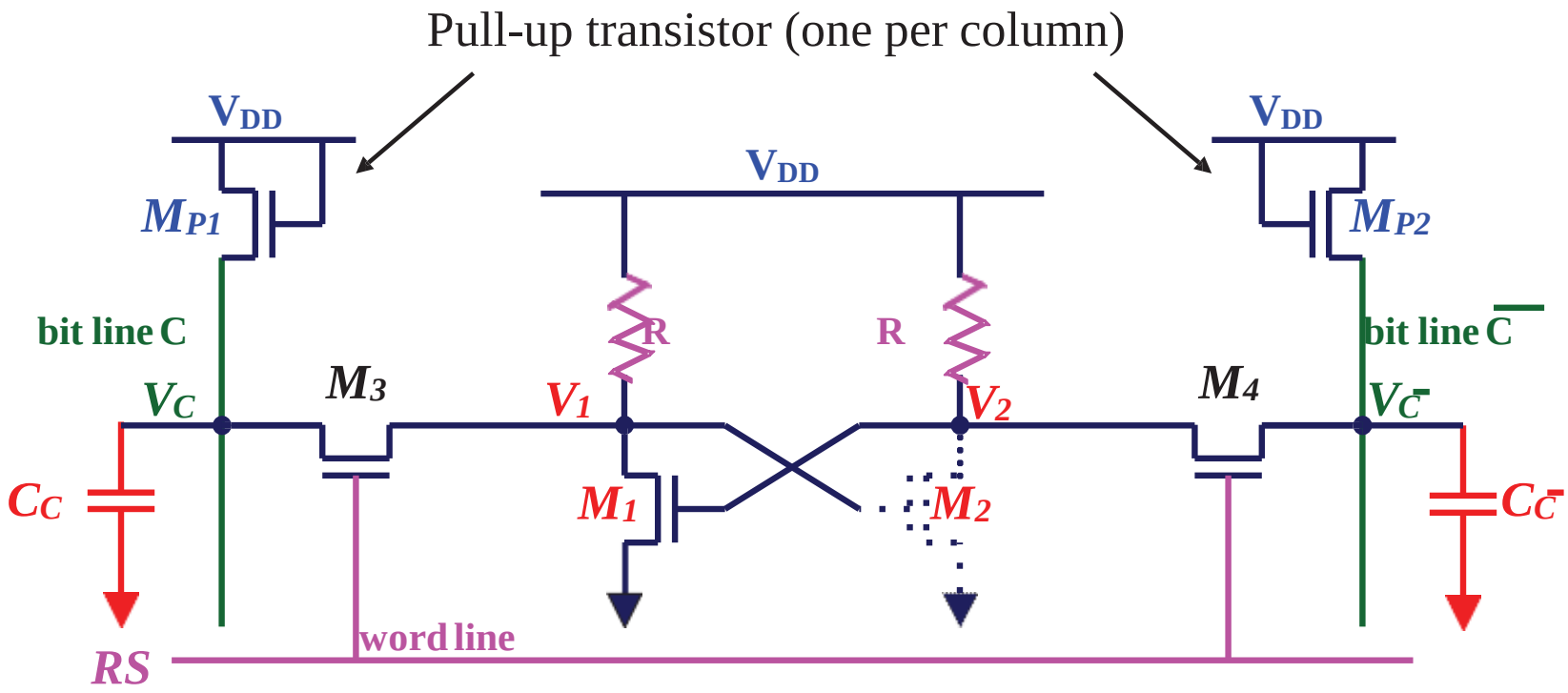


3. Write "0" Operation ($V_1=V_{OH}$, $V_2=V_{OL}$ at $t=0^-$):

$V_C \rightarrow V_{OL}$ by the *data-write circuitry*.

Since $V_1 \rightarrow V_{OL}$, M_2 turns off, therefore $V_2 \rightarrow V_{OH}$.

SRAM Operation Principles (Cont.)

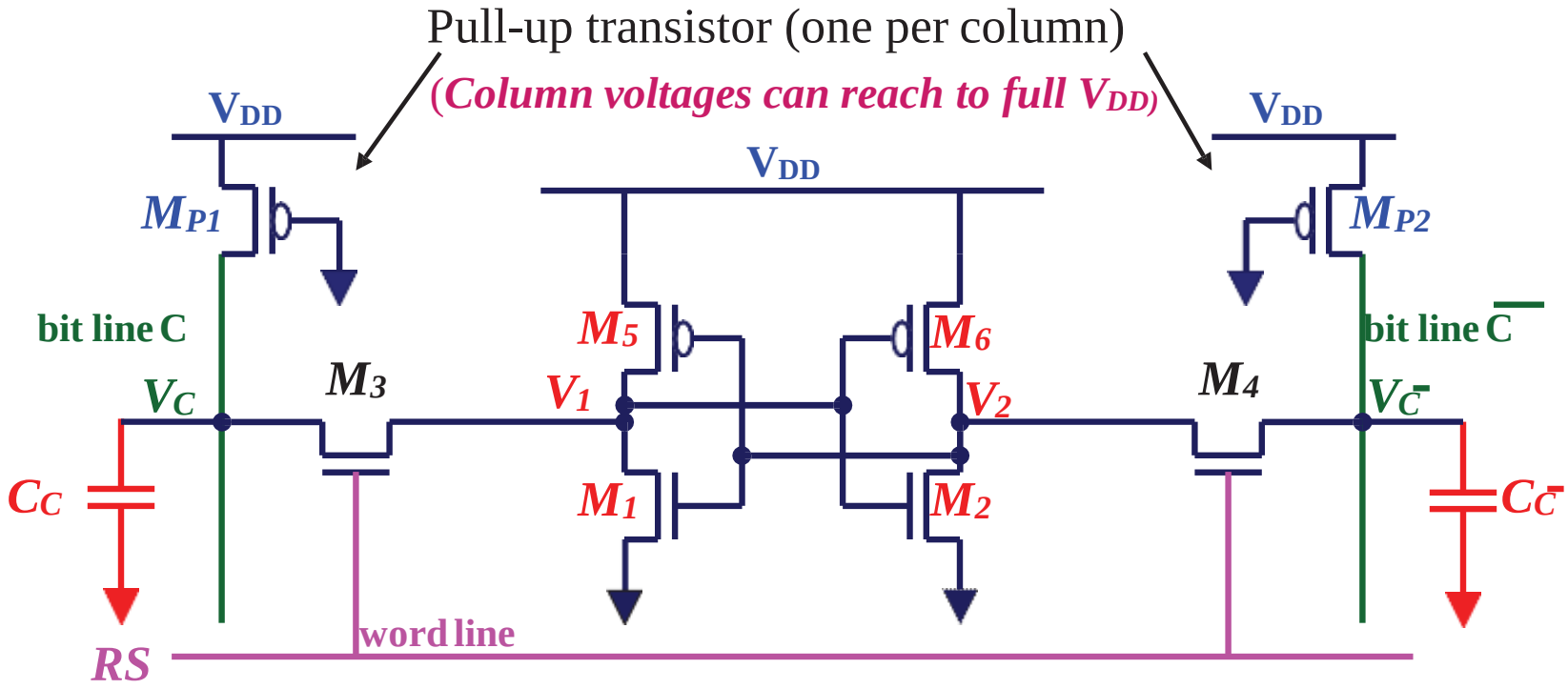


4. Read "0" Operation ($V_1=V_{OL}$, $V_2=V_{OH}$ at $t=0^-$):

$V_{\bar{C}}$ retains pre-charge level, while $V_C \rightarrow V_{OL}$ by M_1 *ON*.

Data-read circuitry detects small voltage difference $V_C - V_{\bar{C}} < 0$, and amplifies it as a "*0*" data output.

Circuit of CMOS SRAM Cell



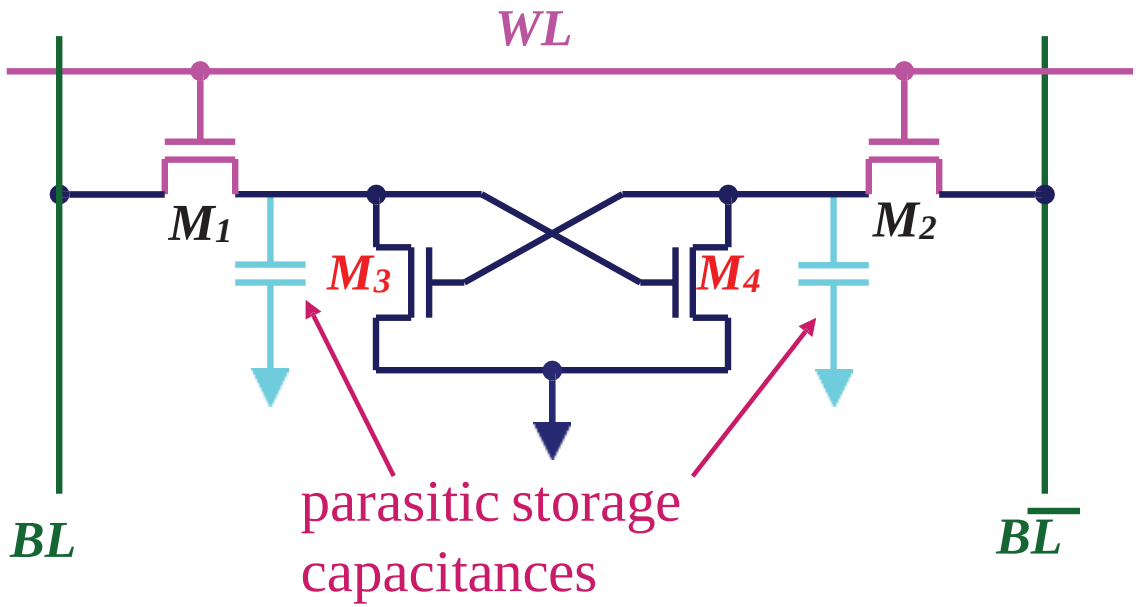
- **Advantages**
  - Very **low standby power** consumption
  - **Large noise margins** than ***R***-load **SRAMS**
  - **Operate at lower supply voltages** than ***R***-load **SRAMS**
- **Disadvantages**
  - **Larger die area**: To accommodate the n-well for pMOS transistors and polysilicon contacts. The area has been reduced by using multi-layer polysilicon and multi-layer metal processes
  - **CMOS more complex process**

Dynamic Read-Write Memory (DRAM) Circuits

- **SRAM:** 4~6 transistors per bit

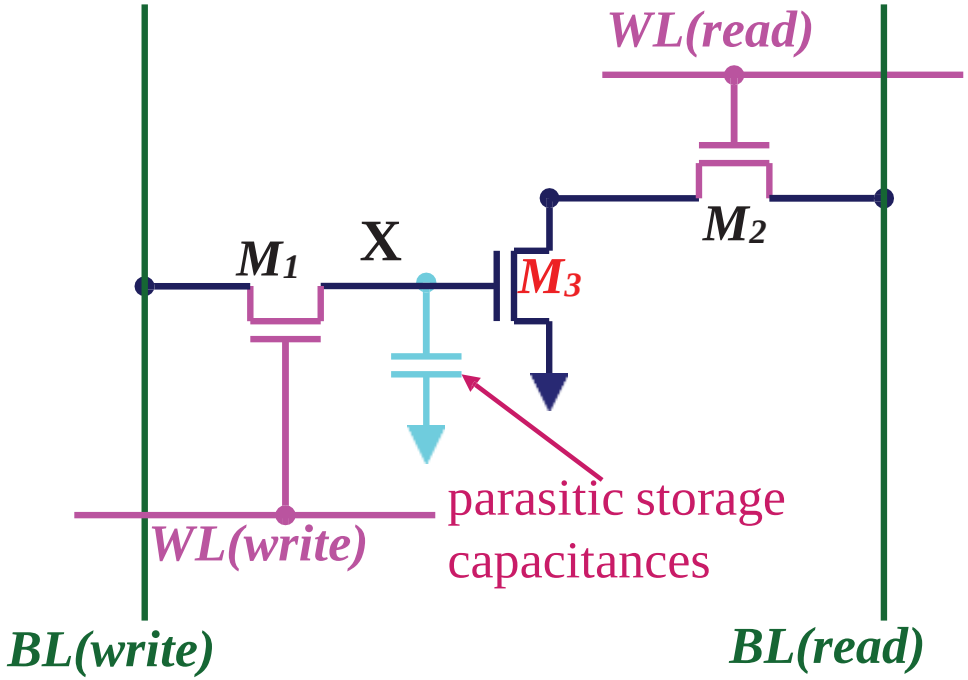
  4~5 lines connecting as charge on capacitor
- **DRAM:** Data bit is stored as charge on capacitor Reduced die area

  Require periodic refresh



Four-Transistor DRAM Cell

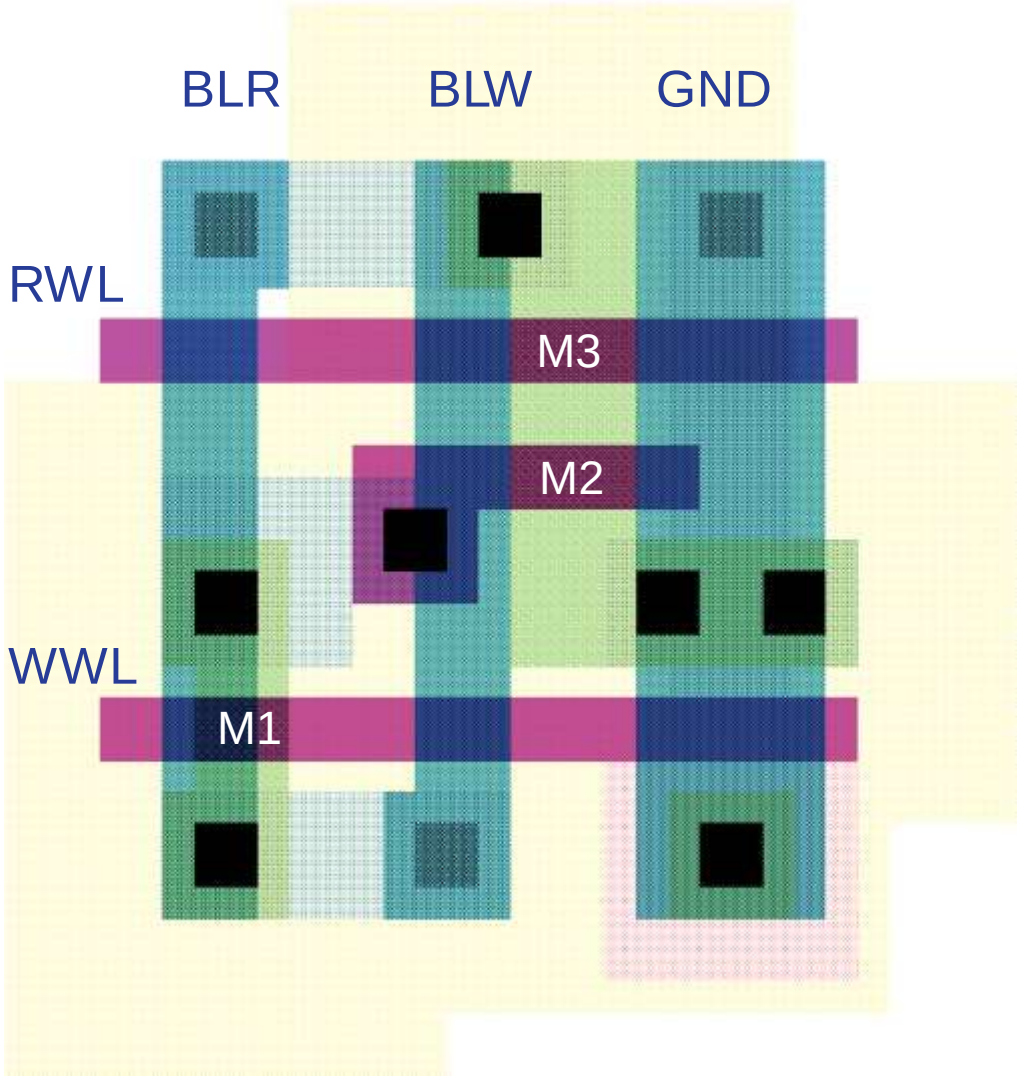
DRAM Circuits (Cont.)



Three-Transistor DRAM Cell

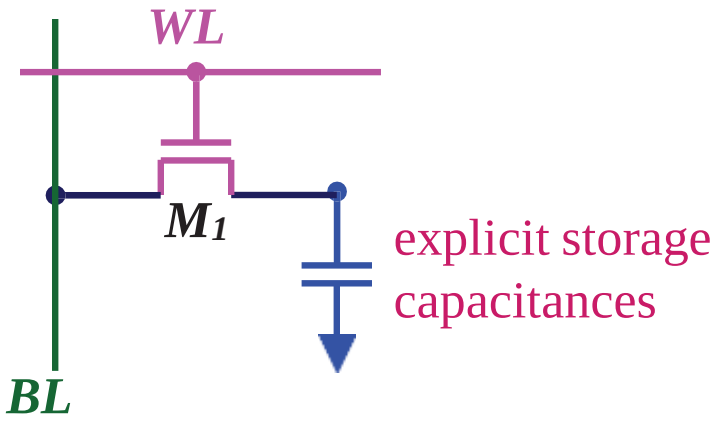
- No constraints on device ratios
- Reads are non-destructive
- Value stored at node X when writing a "1" = V_{WWL}-V_{Tn}

3T-DRAM — Layout



Source: Digital Integrated Circuits 2nd

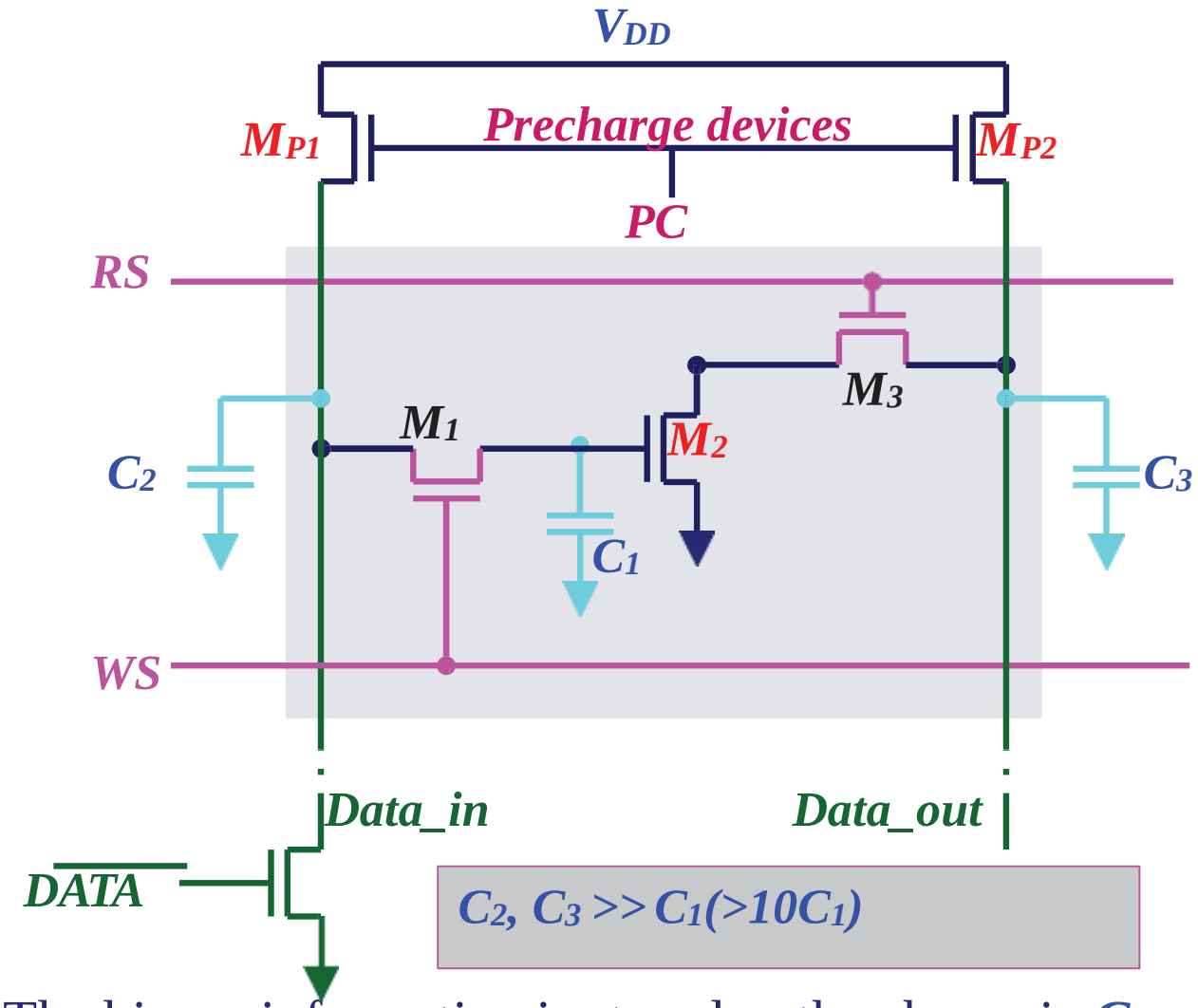
One-Transistor DRAM Cell



One-Transistor DRAM Cell

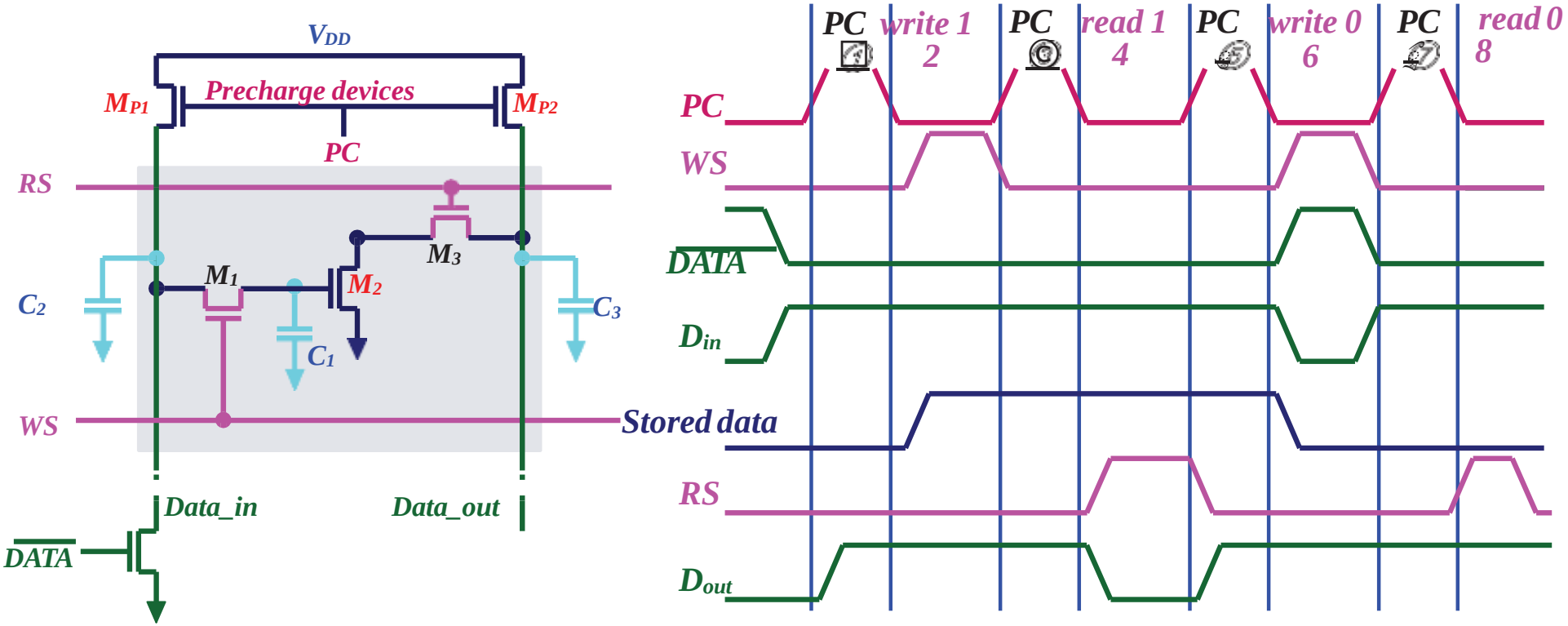
- **Industry standard** for high density dram arrays
- **Smallest** component count and silicon area per bit
- Separate or "**explicit**" capacitor (dual poly) per cell

Operation of Three-Transistor DRAM Cell



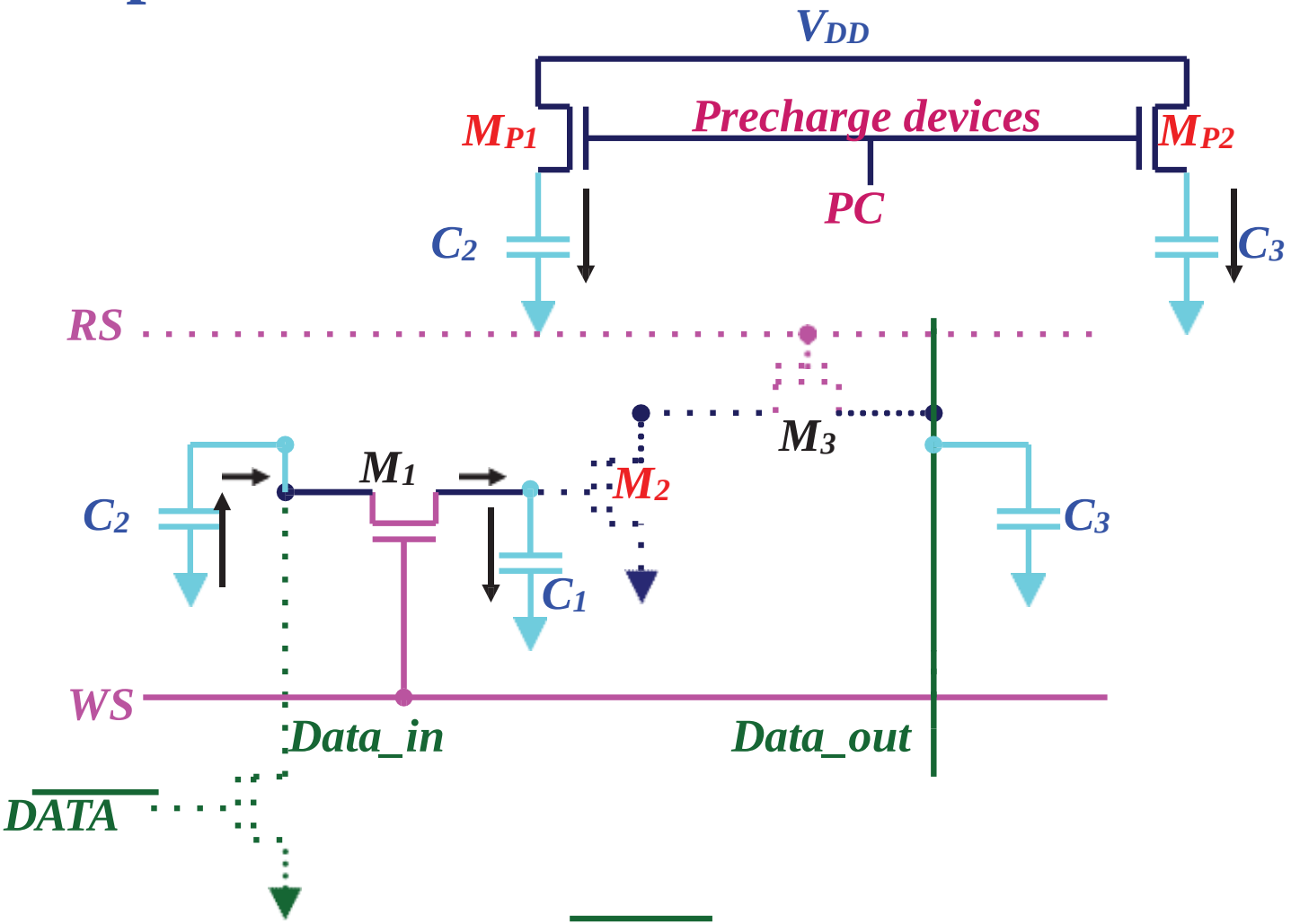
- The binary information is stored as the charge in C_1
- ***Storage transistor M_2*** is on or off depending on the charge in C_1
- **Pass transistors M_1 and M_3:** access switches
- Two separate bit lines for "data read" and "data write"

Operation of Three-Transistor DRAM Cell (Cont.)



- The operation is based on a **two-phase non-overlapping clock scheme**
  - » The precharge events are driven by $\exists_1$, and the "read" and "write" operations are driven by $\exists_2$.
  - » Every "read" and "write" operation is preceded by a precharge cycle, which is initiated with ***PC*** going **high**.

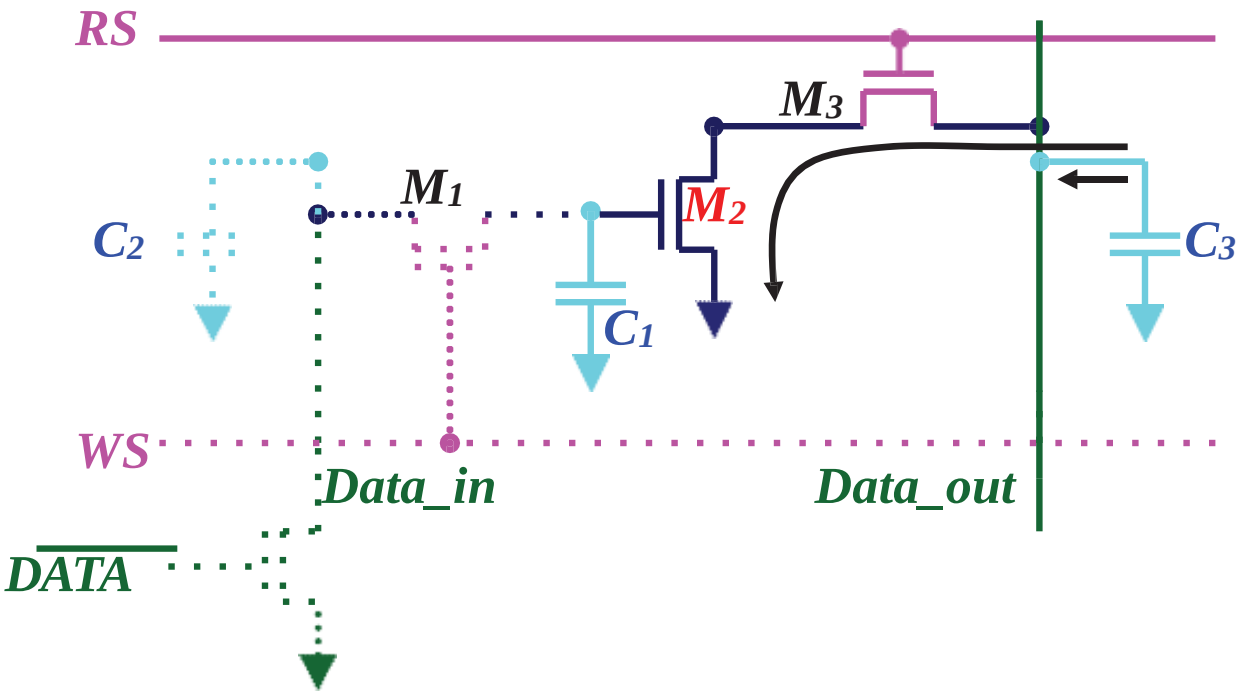
Operation of Three-Transistor DRAM Cell (Cont.)



Pre-charge Cycle

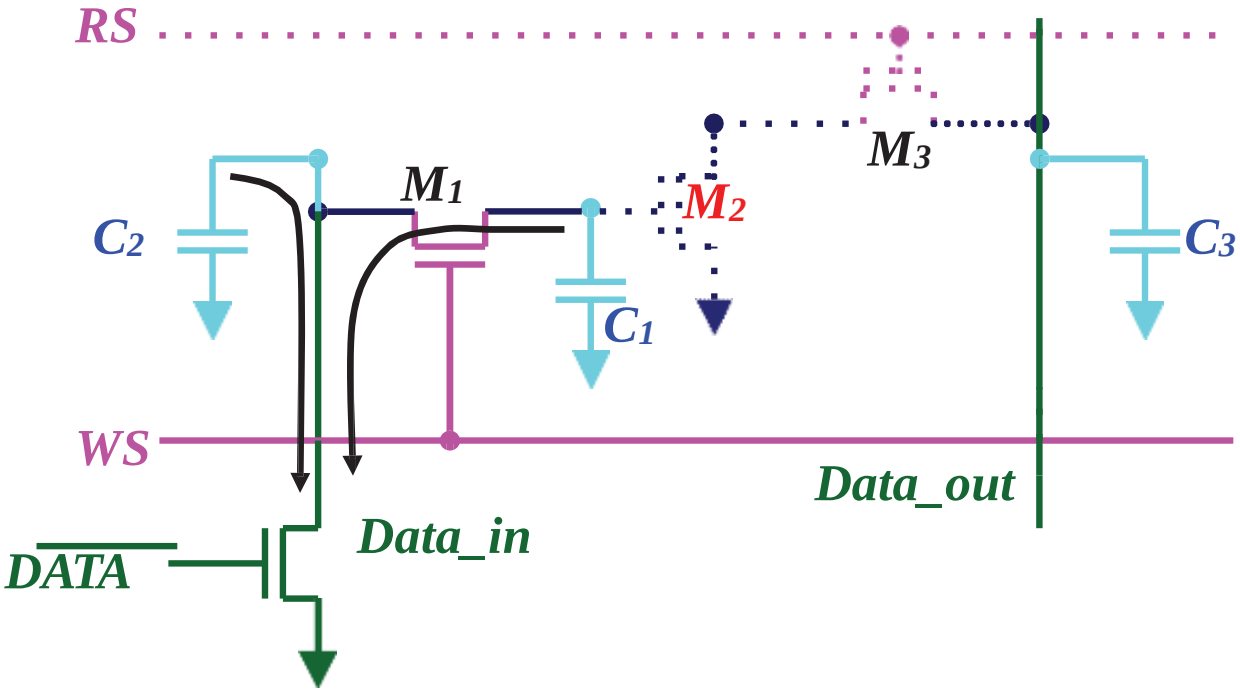
- **Write "1" OP**: $\overline{DATA} = 0$, $WS = 1$; $RS = 0$
  - » C_2, C_1 Share charge due to M_1 **ON**
  - » Since $C_2 >> C_1$, the storage node C_1 attains approximately the same logic level.

Operation of Three-Transistor DRAM Cell (Cont.)



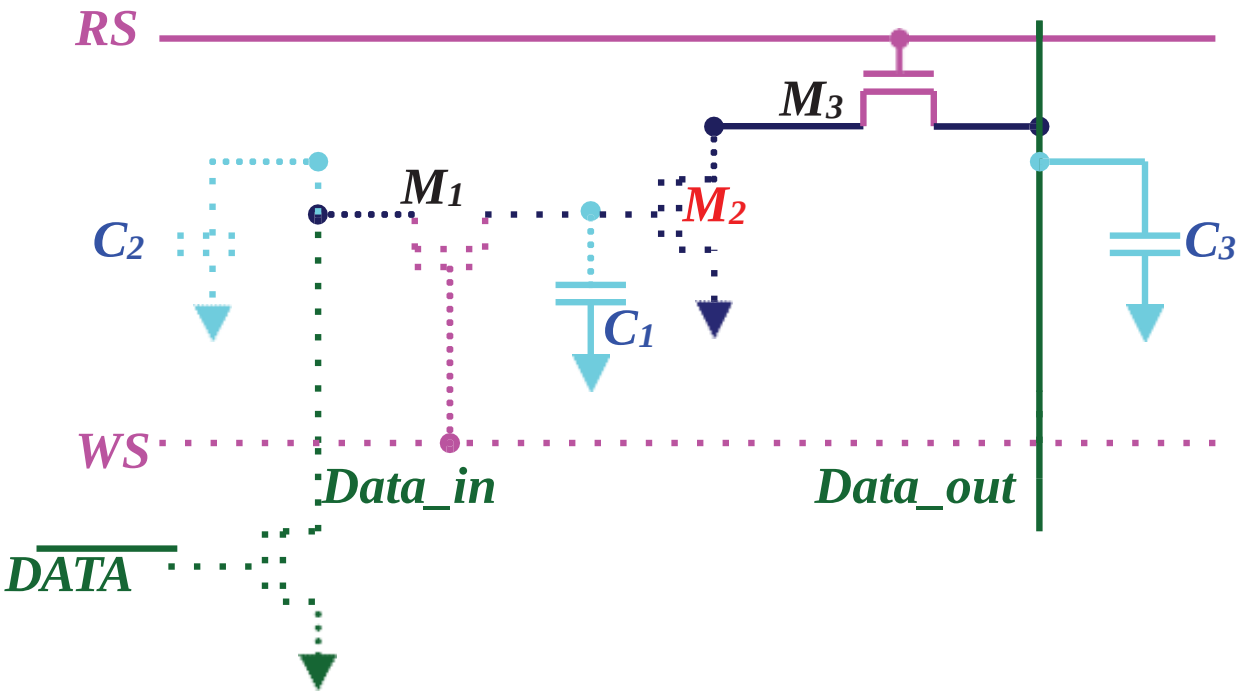
- **Read "1" OP**: $\overline{DATA} = 0$, $WS = 0$; $RS = 1$
  - » M_2, M_3 $ON \rightarrow C_3$, C_1 discharges through M_2 and M_3, and the falling column voltage is interpreted bt the "data read" circuitry as a stored logic "1".

Operation of Three-Transistor DRAM Cell (Cont.)



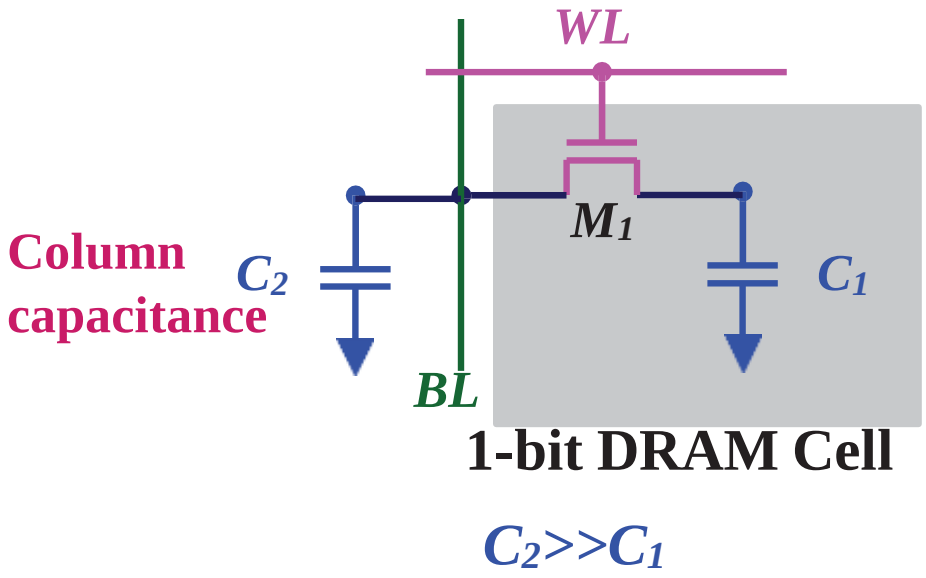
- **Write "0" OP**: $\overline{DATA} = 1$, $WS = 1$; $RS = 0$
  - » M_2, M_3 $ON \rightarrow C_2$ and C_1 discharge to 0 through M_1 and *data_in nMOS*.

Operation of Three-Transistor DRAM Cell (Cont.)



- **Read "0" OP**: $\overline{DATA} = 1$, $WS = 0$; $RS = 1$
  - » C_3 does not discharge due to M_2 **OFF**, and the logic-high level on the ***Data_out*** column is interpreted by the data read circuitry as a stored "0" bit.

Operation of One-Transistor DRAM Cell

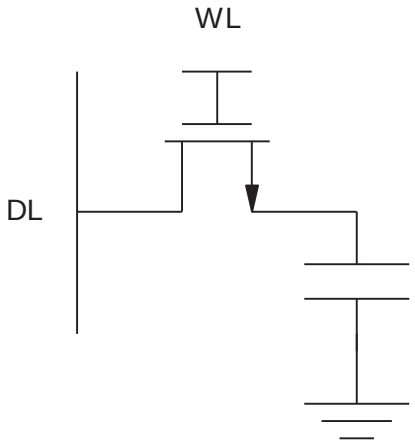


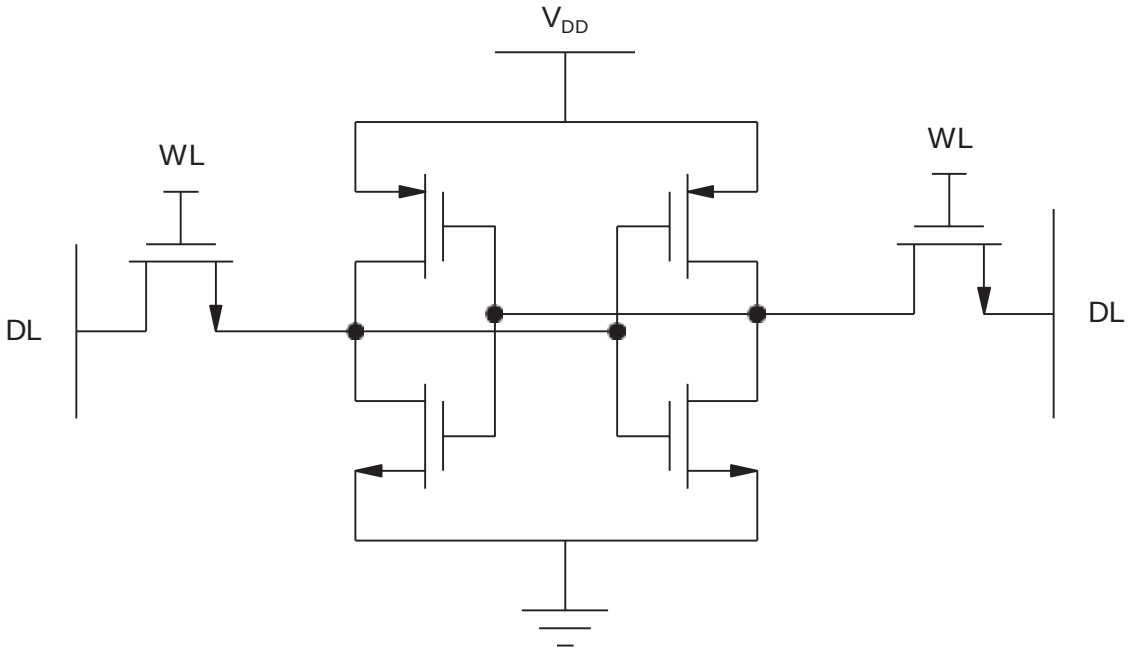
- **Write "1" OP:** ***BL = 1***, ***WL = 1*** (***M1*** **ON**)→***C1*** charges to "1"
- **Write "0" OP:** ***BL = 0***, ***WL = 1*** (***M1*** **ON**)→***C1*** discharges to "0"
- **Read OP:** destroys stored charge on ***C1*** → destructive refresh is needed after every data read operation

RAM

- DRAM

SRAM





Semiconductor Memory Types (Cont.)

- **ROM: 1, nonvolatile memories**

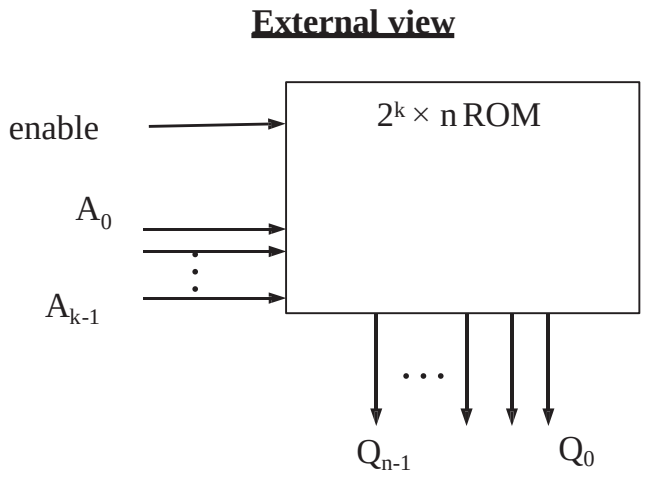
  2, only can access data, cannot to modify data

  3, lower cost: used for permanent memory in printers, fax, and game machines, and ID cards

  - ***Mask ROM*:** data are written **during** chip fabrication by a **photo mask**
  - ***PROM*:** data are written electrically **after** the chip is fabricated.
    - » ***Fuse ROM*:** data **cannot** be erased and modified.
    - » ***EPROM and EEPROM*:** data **can be rewritten**, but the number of subsequent re-write operations is limited to 10^4-10^5.
      - ***EPROM* uses ultraviolet rays** which can penetrate through the crystal glass on package to erase whole data simultaneously.
      - ***EEPROM* uses high electrical voltage** to erase data in 8 bit units.
  - ***Flash Memory*:** similar to EEPROM

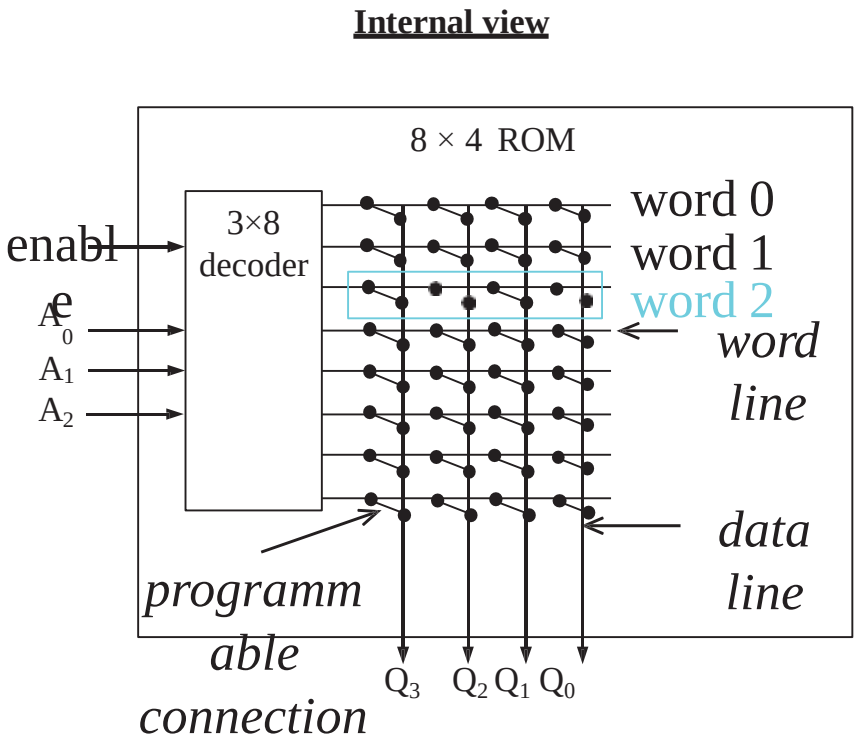
ROM: “Read-Only” Memory

- Nonvolatile
- Can be read from but not written to, by a processor in an microcomputer system
- Traditionally written to, “programmed”, before inserting to microcomputer system
- Uses
  - Store software program for general-purpose processor
  - Store constant data (parameters) needed by system
  - Implement combinational circuits (e.g., decoders)



Example: 8 x 4 ROM

- Horizontal lines = words
- Vertical lines = data
- Lines connected only at circles
- Decoder sets word 2's line to 1 if address input is 010
- Data lines Q_3 and Q_1 are set to 1 because there is a "programmed" connection with word 2's line
- Word 2 is not connected with data lines Q_2 and Q_0
- Output is 1010



Memory – ROM

- Nonvolatile Memory

-SRAM and DRAM and attractive due to their speed

-however, they are volatile which means when the power is removed, the data is lost

-for a microcomputer, we need a nonvolatile storage device so that upon power-up, the computer knows what to do.

-currently, the most popular semiconductor ROM is Flash (or EEprom)

- before looking at the details of a Flash transistor, let's first look at the different types of ROM arrays and addressing modes

ROM Arrays

- There are two basic types of ROM arrays

  1) NOR-based ROM
  2) NAND-based ROM

NOR-based ROM

- All Column Lines are pulled-up using a PMOS transistor (or resistor)

- The Row Lines are connected to the gates of NMOS transistors at the intersection of Row and Column Lines

- The presence or absence of the NMOS transistors dictates whether a 1 or a 0 is stored

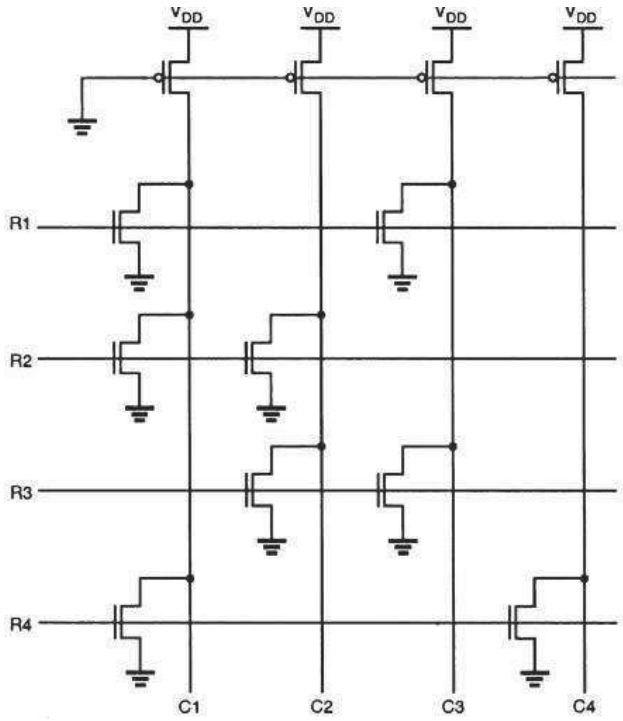
- If the NMOS transistor is present, it will pull down the Column Line when its gate is driven high by the Row Line

- If the NMOS transistor is absent, the Column Line will not be pulled down, so it will remain pulled up by the PMOS's

NOR-based ROM

- In order to Read from the array, the Row line is asserted and the desired Column line is observed

- a NOR-based ROM is similar to a Hex Keypad



R1	R2	R3	R4	C1	C2	C3	C4
1	0	0	0	0	1	0	1
0	1	0	0	0	0	1	1
0	0	1	0	1	0	0	1
0	0	0	1	0	1	1	0

Memory – ROM

NAND-based ROM

R1	R2	R3	R4	C1	C2	C3	C4
0	1	1	1	0	1	0	1
1	0	1	1	0	0	1	1
1	1	0	1	1	0	0	1
1	1	1	0	0	1	1	0

- NAND-based ROM is a different array architecture

- it uses a depletion-load NMOS as the pull-up transistor

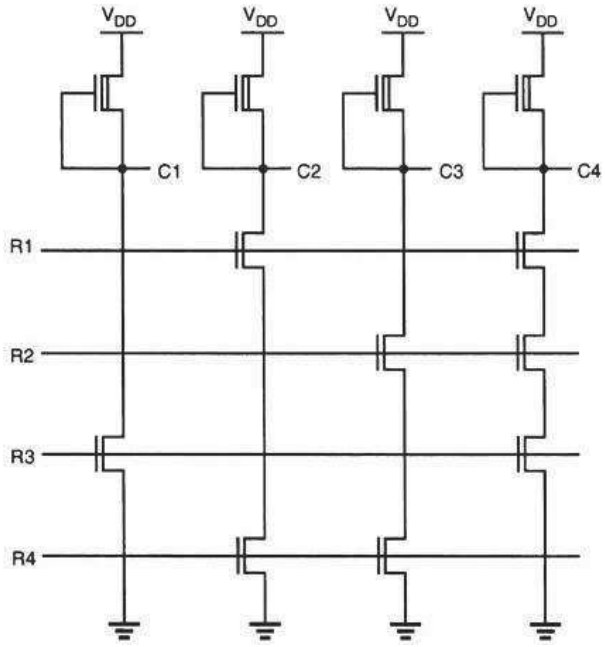
- the Column NMOS's are connected in series with the column lines (i.e. a NAND configuration)

- If an NMOS exists in the Column line and the Row line is asserted, the NMOS will pull the Column Line down and represent a stored '0'

- If an NMOS is absent on the Column line and the Row line is asserted, the Column Line will remain pulled high by the depletion NMOS and represent a stored '1'

- since all of the NMOS's are in series, in order to Read from a Row, all other Rows much be turned ON

- this means in order to distinguish the Row we are asserting, we write a '0' to it



NAND-based ROM

R1	R2	R3	R4	C1	C2	C3	C4
0	1	1	1	0	1	0	1
1	0	1	1	0	0	1	1
1	1	0	1	1	0	0	1
1	1	1	0	0	1	1	0

- In this configuration, if an NMOS is present, it will represent a "stored 1" since in order to address its location, the Row line is driven to a '0' and the NMOS not turned on. This leaves the Column line pulled HIGH

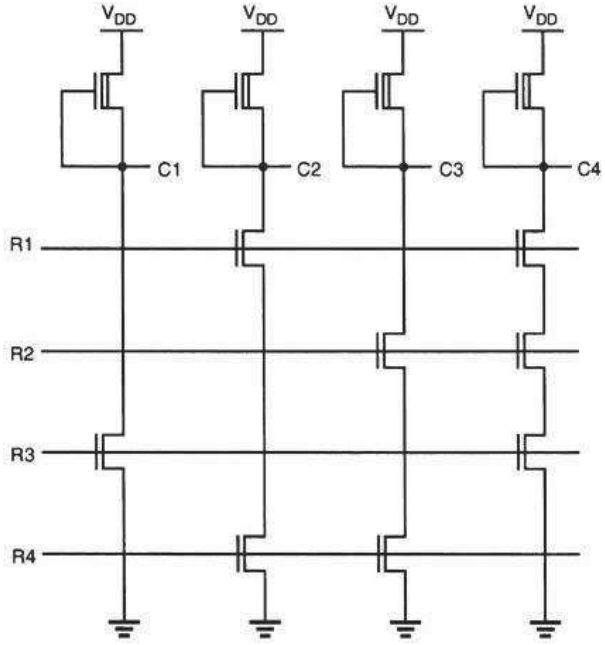
- if an NMOS is absent, it will represent a "stored 0" since all of the other Row NMOS's are turned on and will pull the Column Line LOW

- this gives the opposite behavior as in a NOR-based ROM

	NOR	NAND
NMOS present	0	1
NMOS absent	1	0

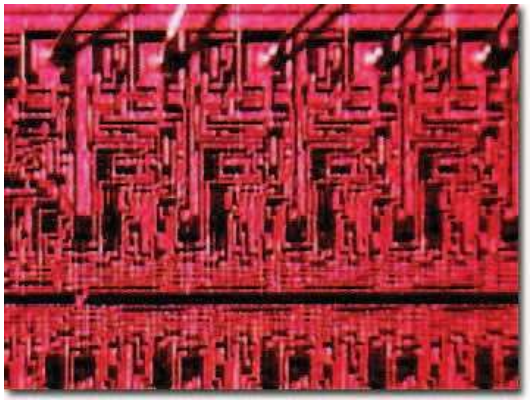
- it also gives a complementary addressing scheme

	NOR	NAND
Address Row Line by driving:	1	0
All other Row Lines driven to:	0	1



Mask-programmed ROM

- Connections “programmed” at fabrication
  - set of masks
- Lowest write ability
  - only once
- Highest storage permanence
  - bits never change unless damaged
- Typically used for final design of high-volume systems
  - spread out NRE (non-recurrent engineering) cost for a low unit cost

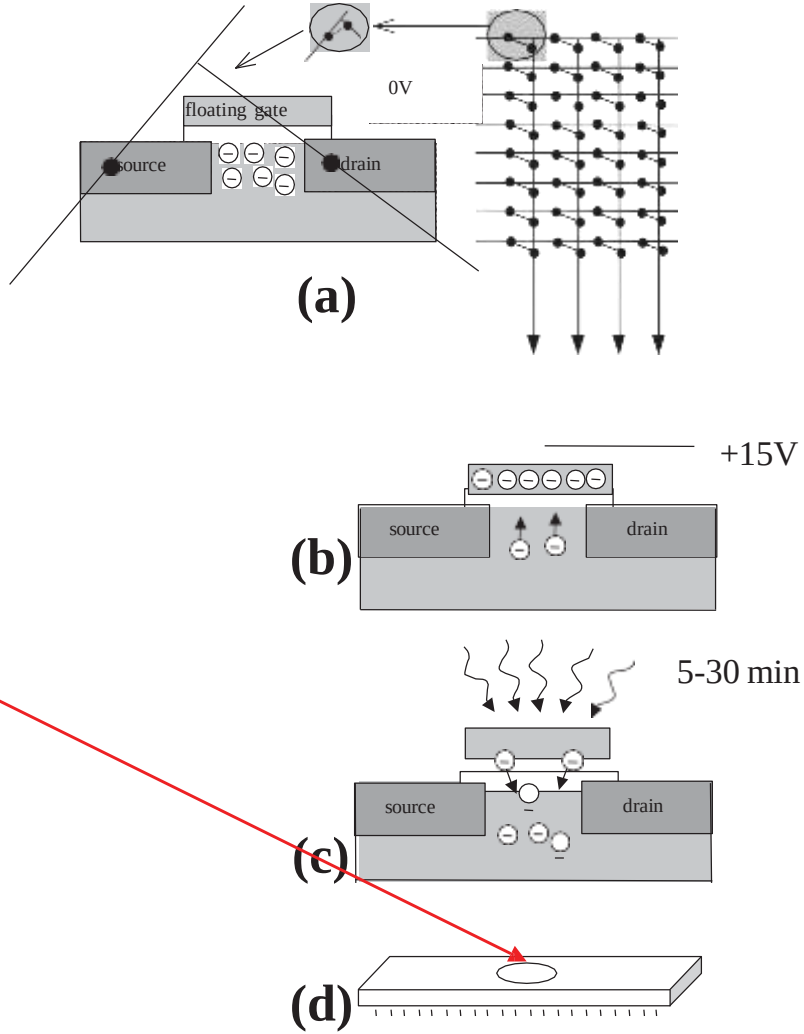


OTP ROM: One-time programmable ROM

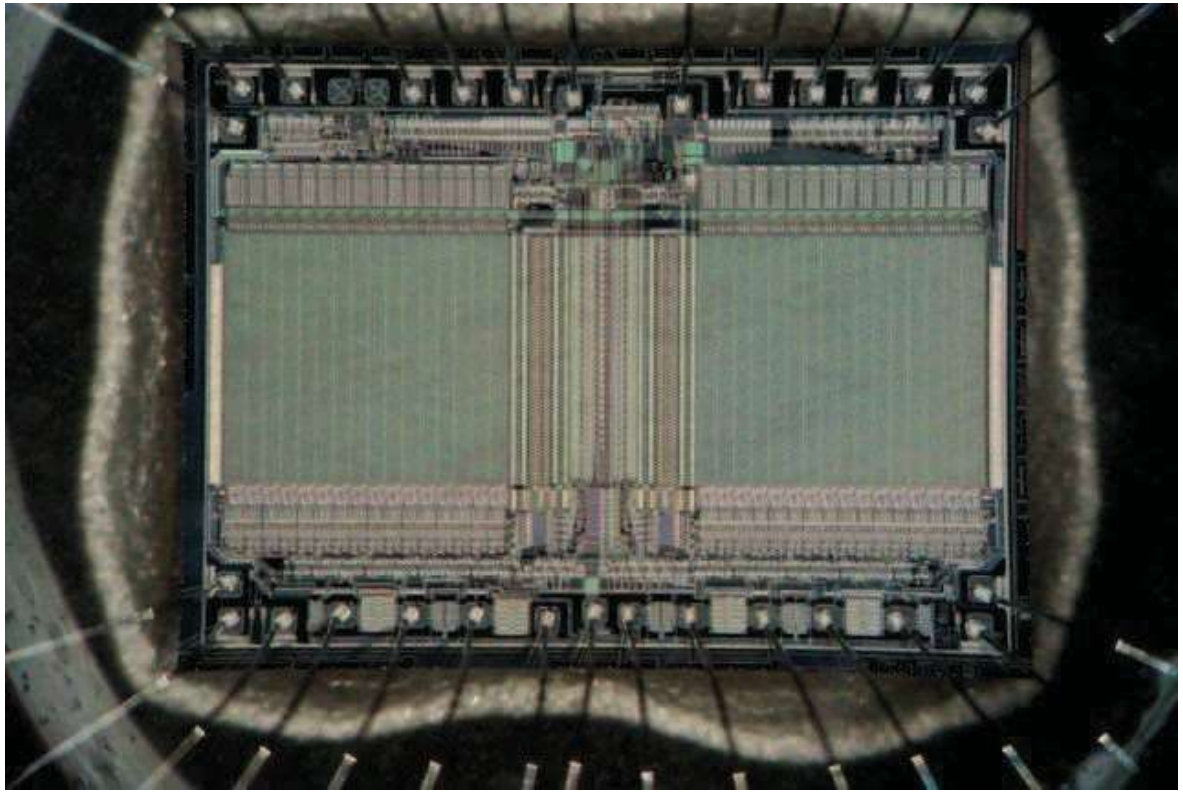
- Connections “programmed” after manufacture by user
  - user provides file of desired contents of ROM
  - file input to machine called ROM programmer
  - each programmable connection is a fuse
  - ROM programmer blows fuses where connections should not exist
- Very low write ability
  - typically written only once and requires ROM programmer device
- Very high storage permanence
  - bits don’t change unless reconnected to programmer and more fuses blown
- Commonly used in final products
  - cheaper, harder to inadvertently modify

EPROM: Erasable programmable ROM

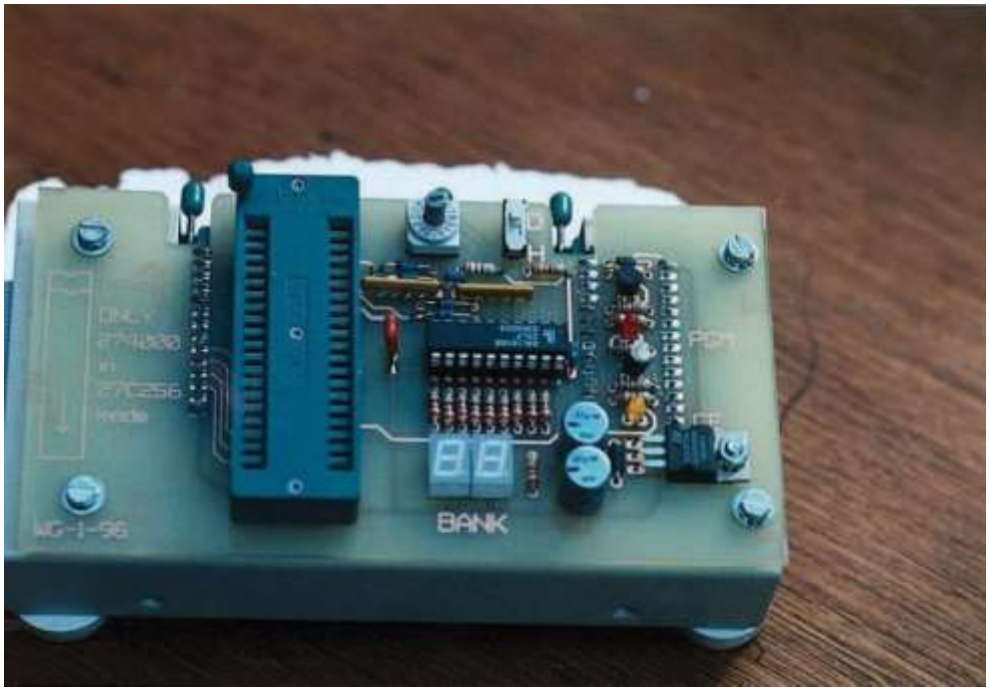
- **Programmable component is a MOS transistor**
  - Transistor has "floating" gate surrounded by an insulator
  - **(a)** Negative charges form a channel between source and drain storing a logic 1
  - **(b)** Large positive voltage at gate causes negative charges to move out of channel and get trapped in floating gate storing a logic 0
  - **(c)** (Erase) Shining UV rays on surface of floating-gate causes negative charges to return to channel from floating gate restoring the logic 1
  - **(d)** An EPROM package showing quartz window through which UV light can pass
- **Better write ability**
  - can be erased and reprogrammed thousands of times
- **Reduced storage permanence**
  - program lasts about 10 years but is susceptible to radiation and electric noise
- **Typically used during design development**



Sample EPROM components



Sample EPROM programmers



EEPROM: Electrically erasable programmable ROM

- Programmed and erased electronically
  - typically by using higher than normal voltage
  - can program and erase individual words
- Better write ability
  - can be in-system programmable with built-in circuit to provide higher than normal voltage
    - » built-in memory controller commonly used to hide details from memory user
  - writes very slow due to erasing and programming
    - » “busy” pin indicates to processor EEPROM still writing
  - can be erased and programmed tens of thousands of times
- Similar storage permanence to EPROM (about 10 years)
- Far more convenient than EPROMs, but more expensive

FLASH

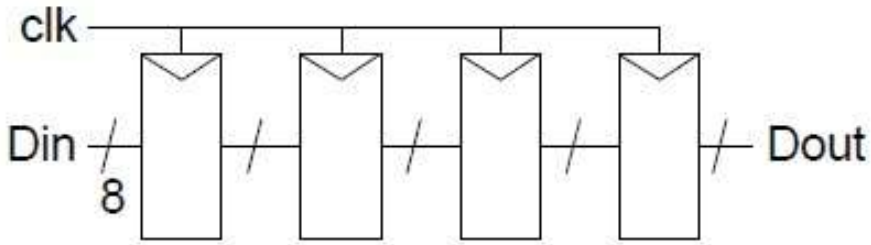
- Extension of EEPROM
  - Same floating gate principle
  - Same write ability and storage permanence
- Fast erase
  - Large blocks of memory erased at once, rather than one word at a time
  - Blocks typically several thousand bytes large
- Writes to single words may be slower
  - Entire block must be read, word updated, then entire block written back
- Used with embedded microcomputer systems storing large data items in nonvolatile memory
  - e.g., digital cameras, MP3, cell phones

Serial Access Memories

- Serial access memories do not use an address
- – Shift Registers
- – Serial In Parallel Out (SIPO)
- – Parallel In Serial Out (PISO)
- – Queues (FIFO, LIFO)

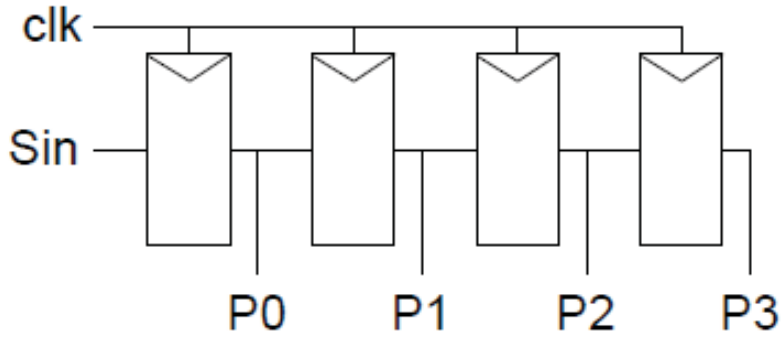
Shift Register

- *Shift registers* store and delay data
- Simple design: cascade of registers
  - Watch your hold times!



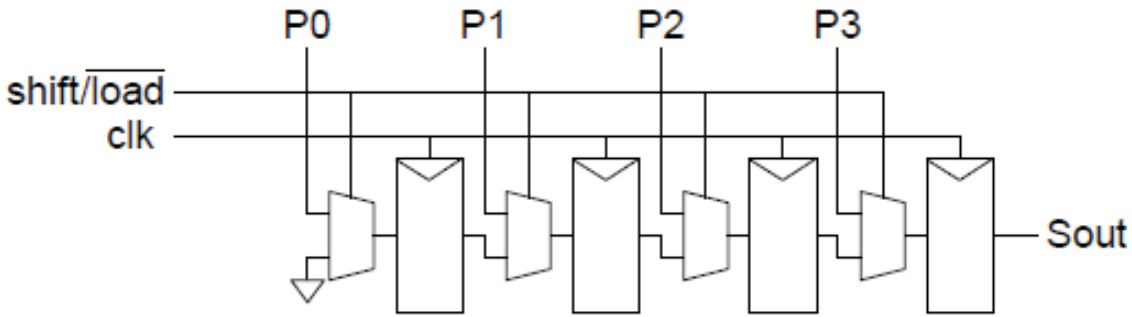
Serial In Parallel Out

- 1-bit shift register reads in serial data
  - After N steps, presents N-bit parallel output



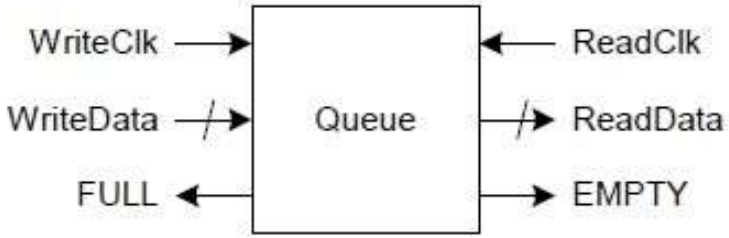
Parallel In Serial Out

- Load all N bits in parallel when shift = 0
  - Then shift one bit out per cycle



Queues

- *Queues* allow data to be read and written at different rates.
- Read and write each use their own clock, data
- Queue indicates whether it is full or empty
- Build with SRAM and read/write counters (pointers)



- FIFOs are commonly used in electronic circuits for buffering and flow control which is from hardware to software.
- In its hardware form, a FIFO primarily consists of a set of read and write pointers, storage and control logic.
- Storage may be SRAM, flip-flops, latches or any other suitable form of storage. For FIFOs of non-trivial size, a dual-port SRAM is usually used, where one port is dedicated to writing and the other to reading.
- A synchronous FIFO is a FIFO where the same clock is used for both reading and writing. An asynchronous FIFO uses different clocks for reading and writing.

- **FIFO full/empty**
- A hardware FIFO is used for synchronization purposes. It is often implemented as a circular queue, and thus has two pointers:
- Read Pointer/Read Address Register
- Write Pointer/Write Address Register

- FIFO Empty
- When the read address register reaches the write address register, the FIFO triggers the Empty signal.
- FIFO FULL
- When the write address register reaches the read address register, the FIFO triggers the FULL signal.

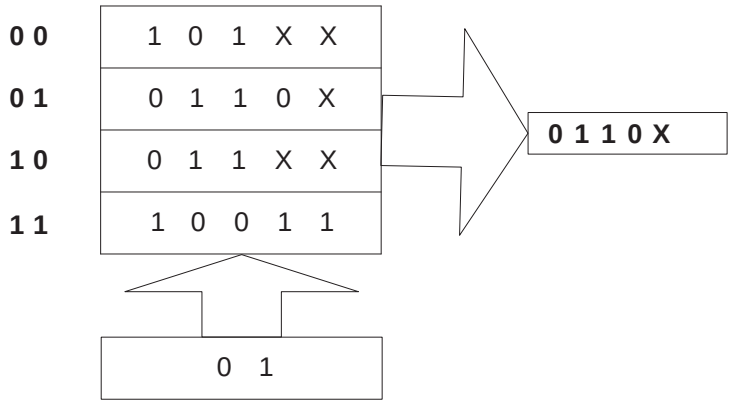
FIFO, LIFO Queues

- *First In First Out* (FIFO)
  - Initialize read and write pointers to first element
  - Queue is EMPTY
  - On write, increment write pointer
  - If write almost catches read, Queue is FULL
  - On read, increment read pointer
- *Last In First Out* (LIFO)
  - Also called a *stack*
  - Use a single *stack pointer* for read and write

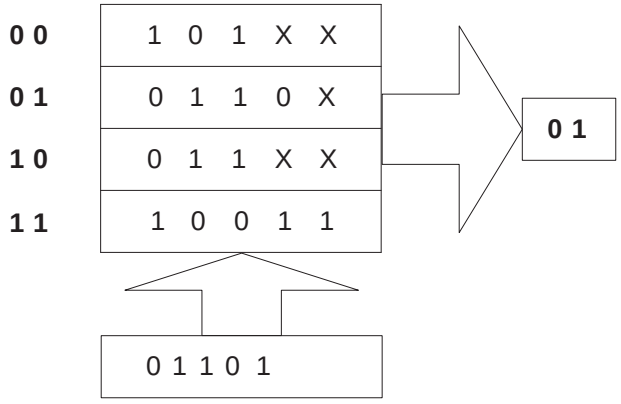
Content Addressable Memories

What is CAM?

- Content Addressable Memory is a special kind of memory!
- Read operation in traditional memory:
  - Input is address location of the content that we are interested in it.
  - Output is the content of that address.
- In CAM it is the reverse:
  - Input is associated with something stored in the memory.
  - Output is location where the associated content is stored.



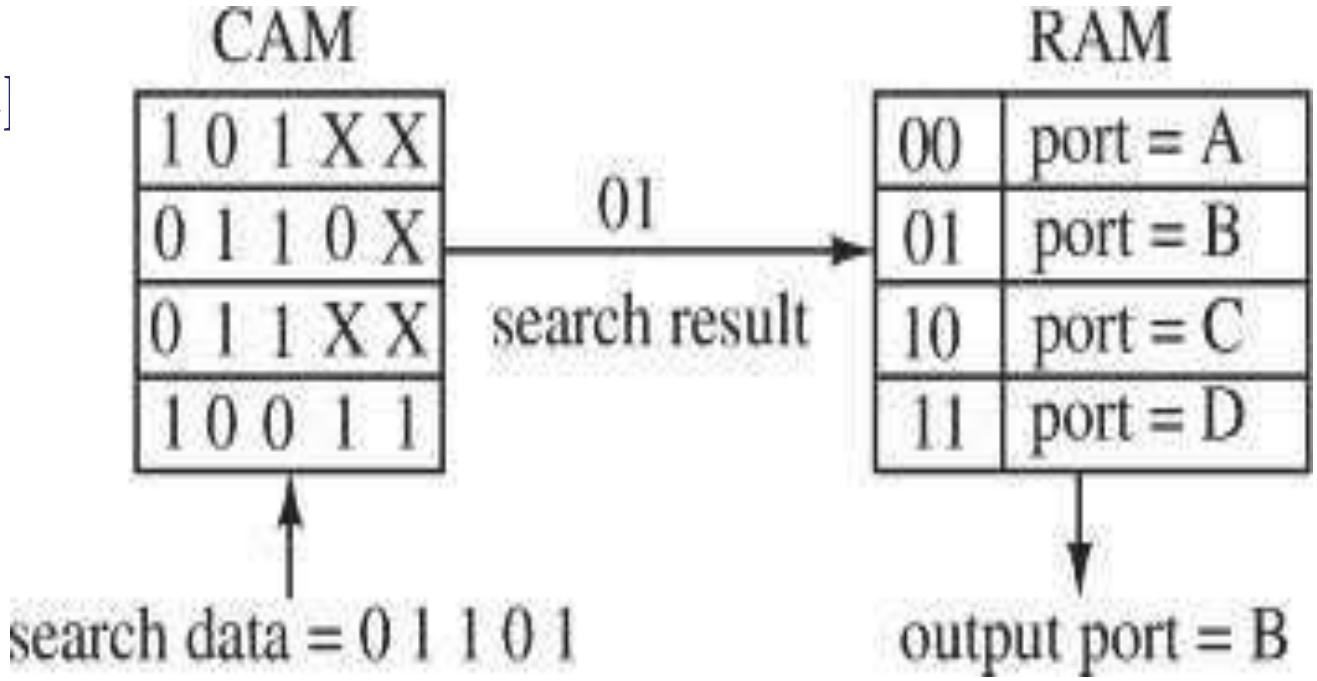
Traditional Memory



Content Addressable Memory

CAM for Routing Table Implementation

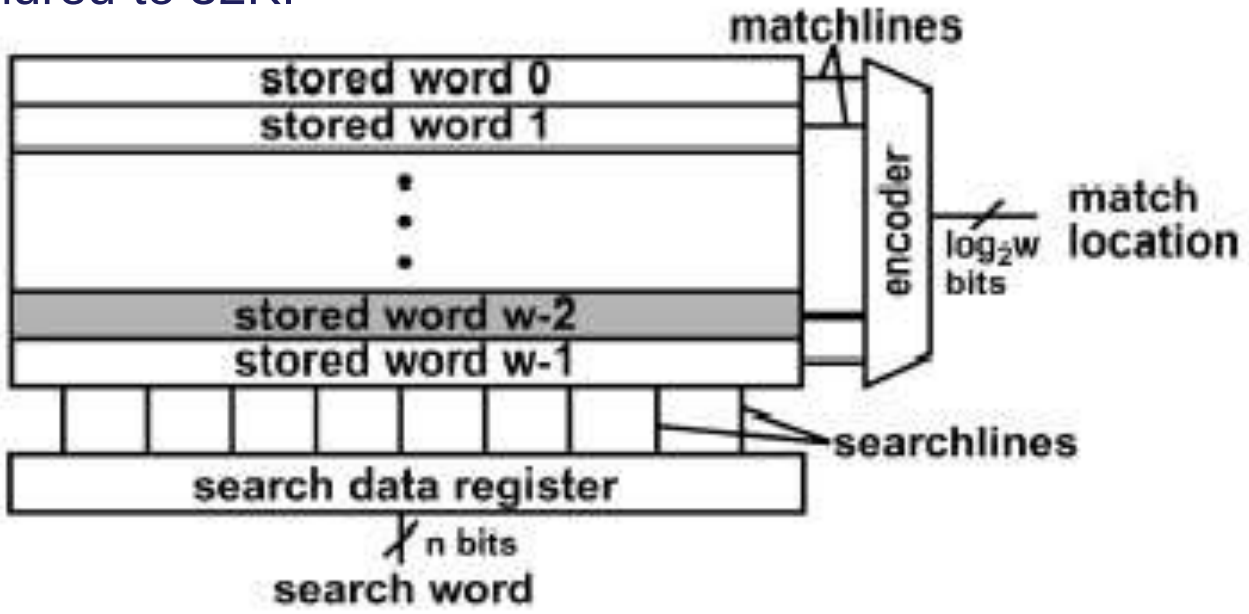
- CAM can be used as a search engine.
- We want to find matching contents in a database or Table.
- Exam



Source: http://pagiamtzis.com/cam/camintro.html

Simplified CAM Block Diagram

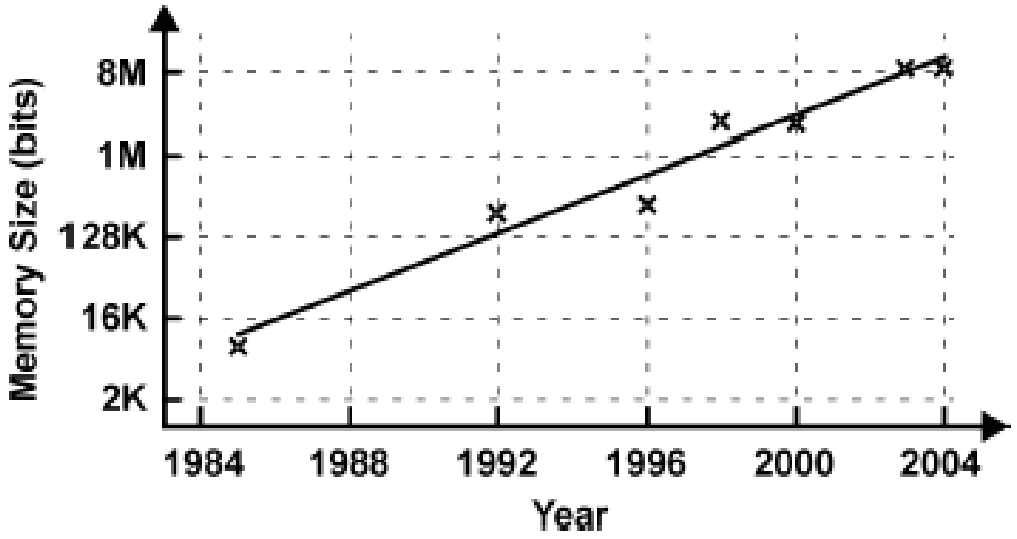
- The input to the system is the *search word*.
- The search word is broadcast on the *search lines*.
- *Match line* indicates if there were a match btw. the search and stored word.
- Encoder specifies the match location.
- If multiple matches, a priority encoder selects the first match.
- *Hit signal* specifies if there is no match.
- The length of the search word is long ranging from 36 to 144 bits.
- Table size ranges: a few hundred to 32K.
- Address space : 7 to 15 bits.



Source: K. Pagiamtzis, A. Sheikholeslami, "Content-Addressable Memory (CAM) Circuits and Architectures: A Tutorial and Survey," *IEEE J. of Solid-state circuits. March 2006*

CAM Memory Size

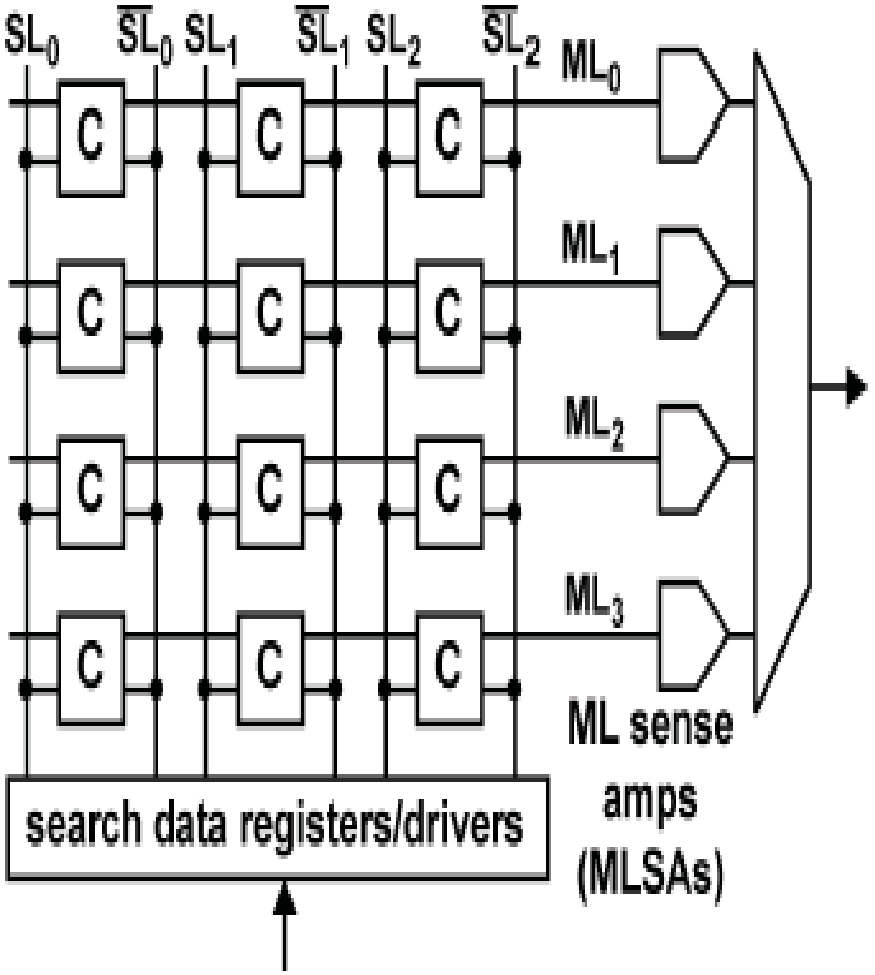
- Largest available around 18 Mbit (single chip).
- Rule of thumb: Largest CAM chip is about half the largest available SRAM chip.
  - A typical CAM cell consists of two SRAM cells.
- Exponential growth rate.



Source: K. Pagiamtzis, A. Sheikholeslami, "Content-Addressable Memory (CAM) Circuits and Architectures: A Tutorial and Survey," *IEEE J. of Solid-state circuits. March 2006*

CAM Basics

- The search-data word is loaded into the search-data register.
- All match-lines are pre-charged to high (temporary match state).
- Search line drivers broadcast the search word onto the differential search lines.
- Each CAM core compares its stored bit against the bit on the corresponding search-lines.
- Match words that have at least one missing bit, discharge to ground.



Source: K. Pagiamtzis, A. Sheikholeslami, "Content-Addressable Memory (CAM) Circuits and Architectures: A Tutorial and Survey," *IEEE J. of Solid-state circuits. March 2006*

Type of CAMs

- Binary CAM (BCAM) only stores 0s and 1s
  - Applications: MAC table consultation. Layer 2 security related VPN segregation.
- Ternary CAM (TCAM) stores 0s, 1s and don't cares.
  - Application: when we need wilds cards such as, layer 3 and 4 classification for QoS and CoS purposes. IP routing (longest prefix matching).
- Available sizes: 1Mb, 2Mb, 4.7Mb, 9.4Mb, and 18.8Mb.
- CAM entries are structured as multiples of 36 bits rather than 32 bits.

CAM Advantages

- They associate the input (comparand) with their memory contents in one clock cycle.
- They are configurable in multiple formats of width and depth of search data that allows searches to be conducted in parallel.
- CAM can be cascaded to increase the size of lookup tables that they can store.
- We can add new entries into their table to learn what they don’t know before.
- They are one of the appropriate solutions for higher speeds.

CAM Disadvantages

- They cost several hundred of dollars per CAM even in large quantities.
- They occupy a relatively large footprint on a card.
- They consume excessive power.
- Generic system engineering problems:
  - Interface with network processor.
  - Simultaneous table update and looking up requests.

Unit VII

SEMICONDUCTOR INTEGRATED CIRCUIT DESIGN

- Programmable Logic Array (PLA)
- Programmable Array Logic(PAL)
- FPGAs
- CPLDs
- Standard cells

Programmable logic devices (PLD)

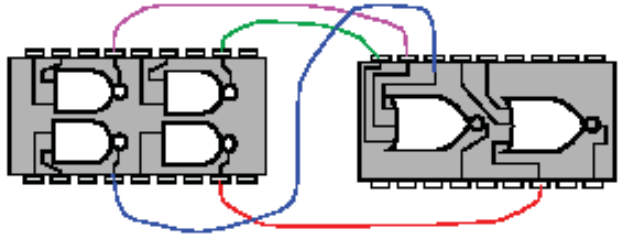
- Design Approach
- Parameters influencing low power design

Constructing Digital Circuits

Hand Wired Circuits

Cirri 1970-85

- Make 2 to 4 silicon gates in a package.
- Connect with wires.

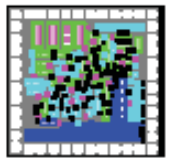


VLSI circuits

Start with a silicon wafer and make:

- the gates
- the interconnections on top

both made together.

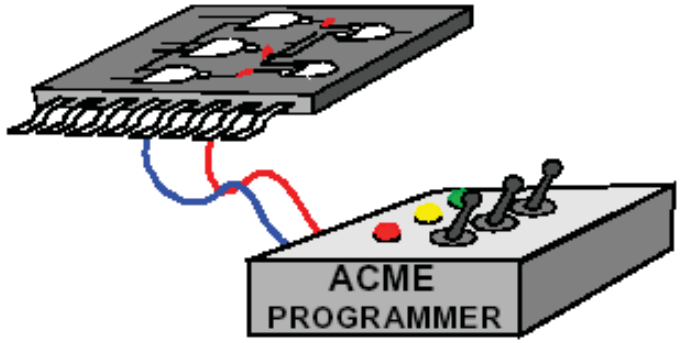


Field Programmable circuits

Start with a silicon wafer and make:

- gates with no connections.
- Make connections later using:

1) electrical means
   - blow fuses, grow anti fuses
   - use memory to hold connections
2) deposit metal lines on top of silicon.

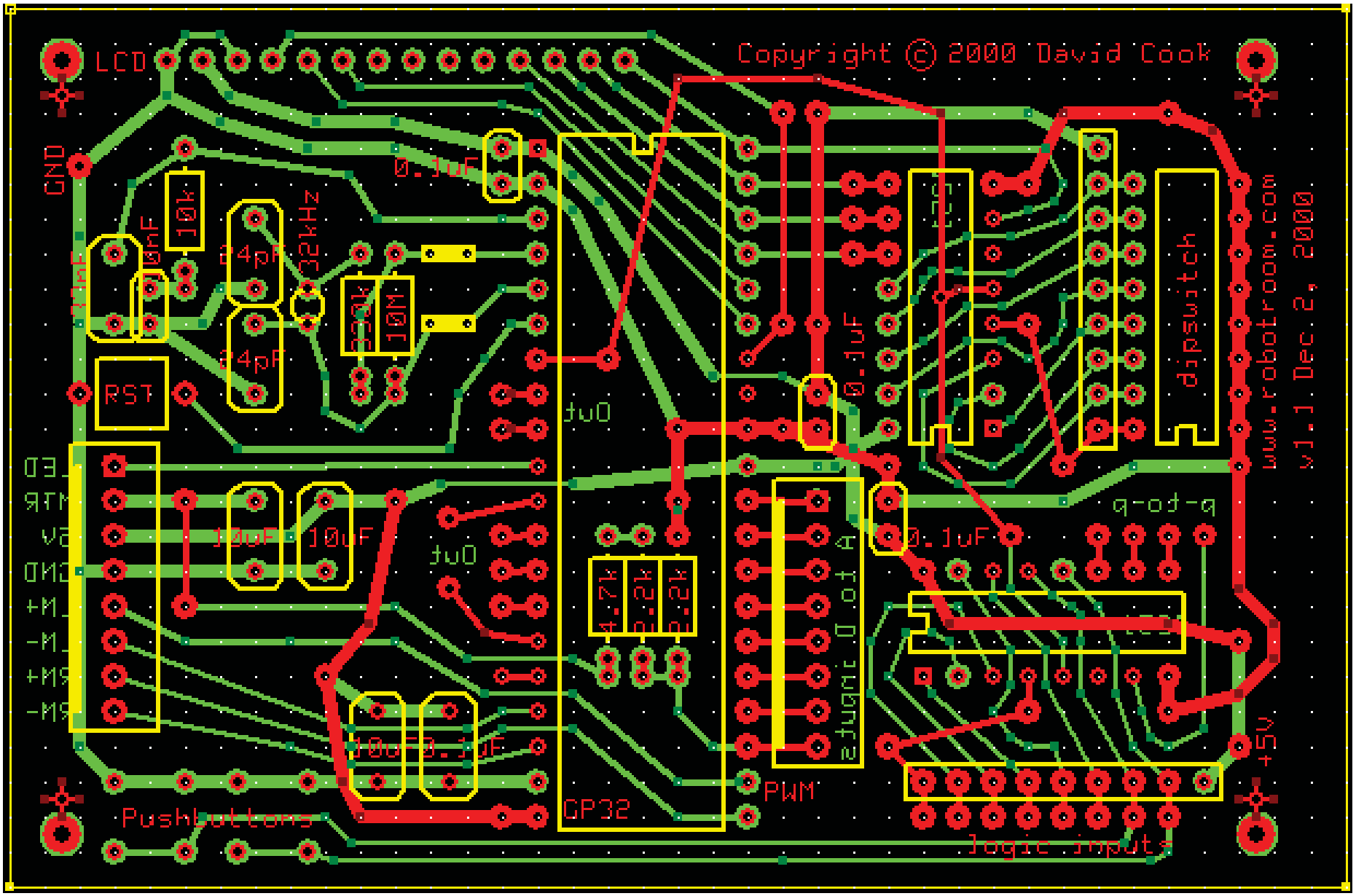


PLD

Programmable logic is defined as a device with configurable logic and flip-flops linked together with programmable interconnect.

Why we are going for PLDs

- Problems by Using Basic Gates
- Many components on PCB:
  - As no. of components rise, nodes interconnection complexity grow exponentially
  - Growth in interconnection will cause increase in interference, PCB size, PCB design cost, and manufacturing time


LCD
Copyright © 2000 David Cook
GND
0.1uF
10k
33nF
24pF
32kHz
24pF
470k
10M
RST
JUO
10uF
10uF
JUO
4.7k
4.2k
4.2k
0.1uF
dipswitch
www.robotroom.com
v1.1 Dec 2, 2000
Pushbuttons
GP32
PWM
+5V
logic inputs