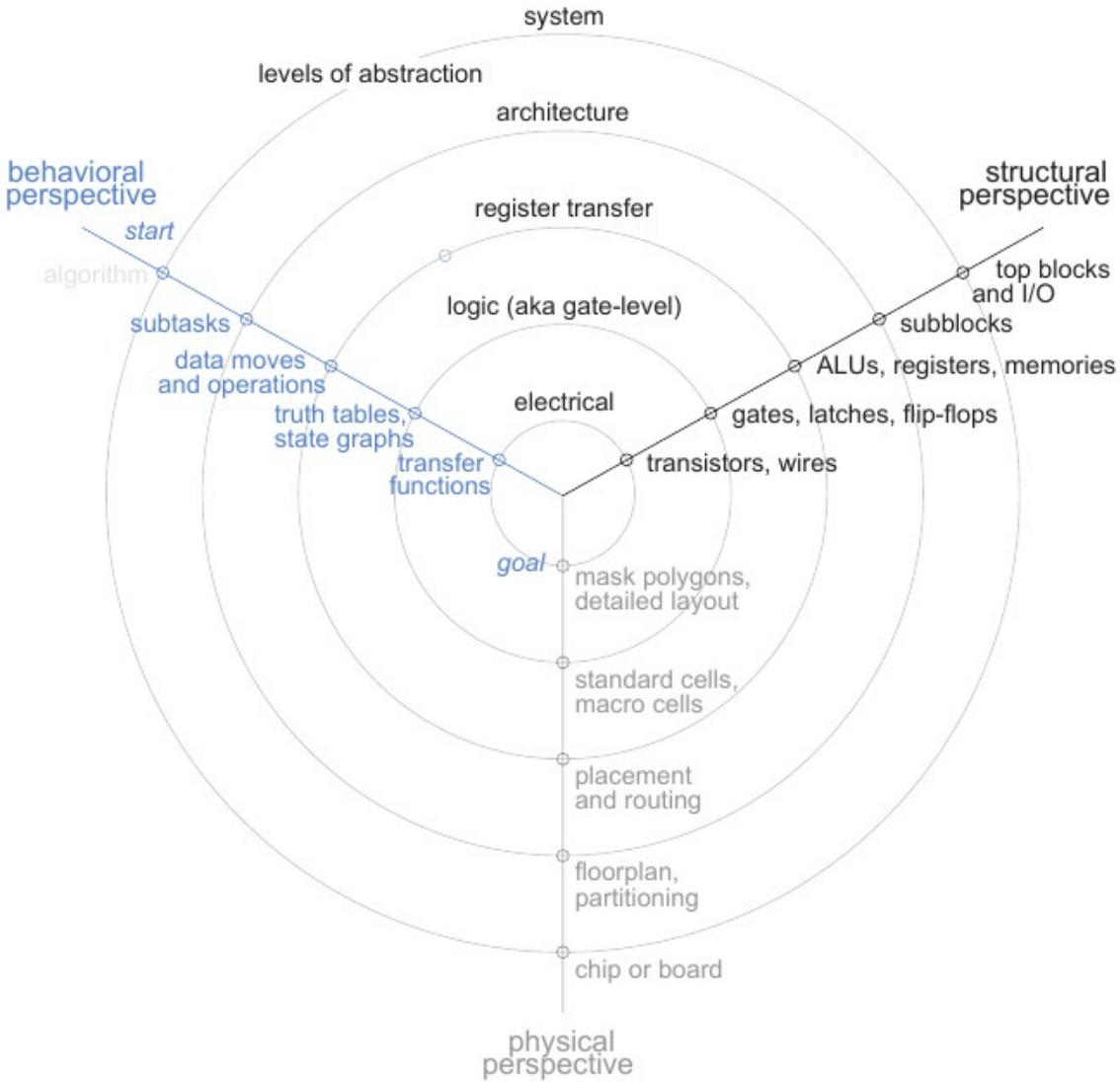
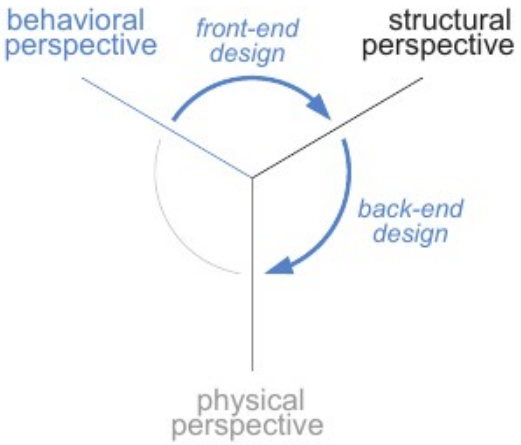
Introduction VLSI Design

Dr. Tassadaq Hussain
Computer Architect
Co-Founder PakAsic.com

Y-Chart of Digital Electronics Systems





Outline

- Silicon, pn-junctions and transistors
- A Brief History
- Operation of MOS Transistors
- CMOS circuits
- Introduction
- Fabrication steps for CMOS circuit

Introduction

Integrated circuits: many transistors on one chip.

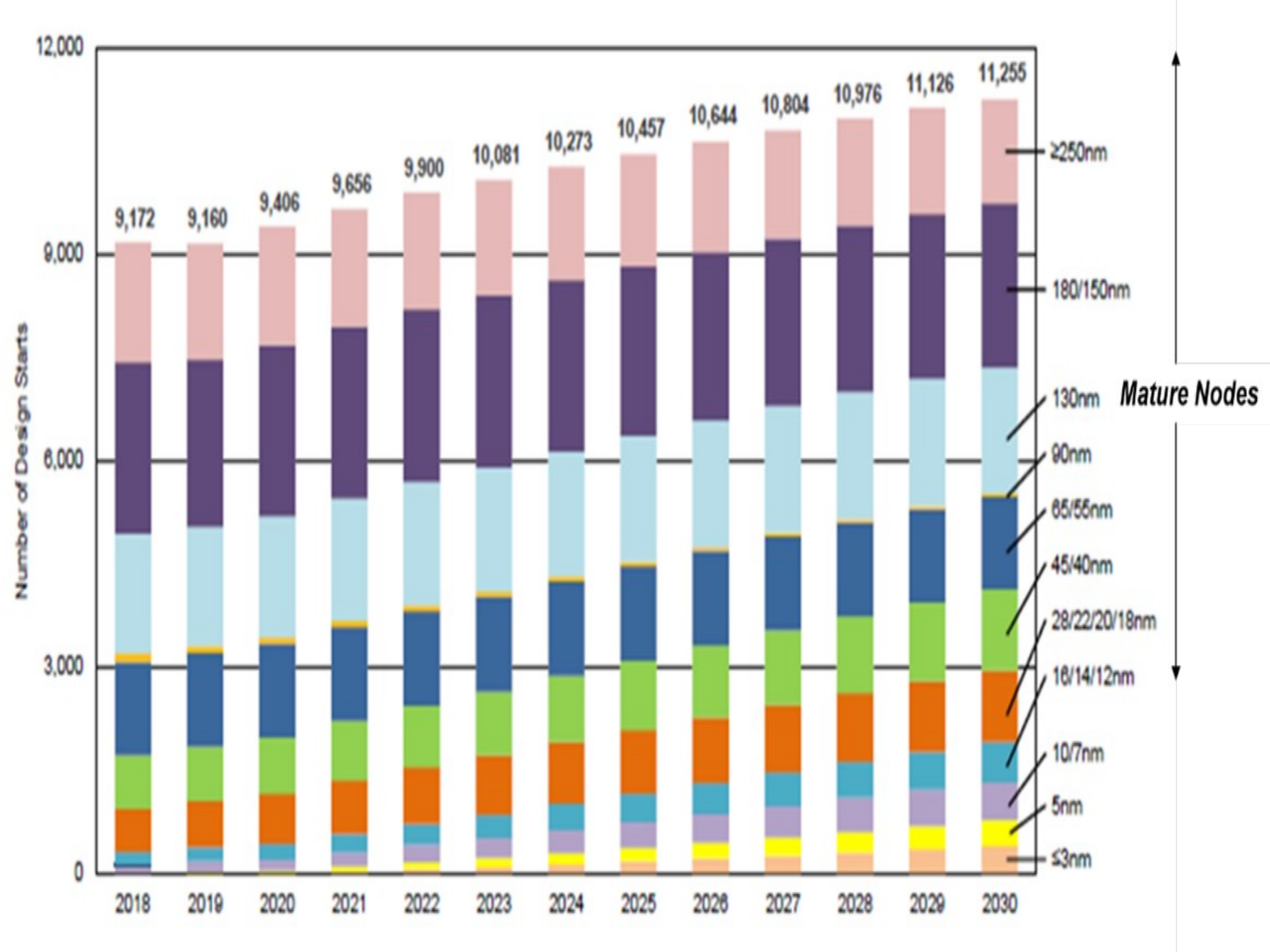
Very Large Scale Integration (VLSI)

Complementary Metal Oxide Semiconductor (CMOS)

Fast, cheap, “low-power” transistors circuits

WHY VLSI DESIGN?

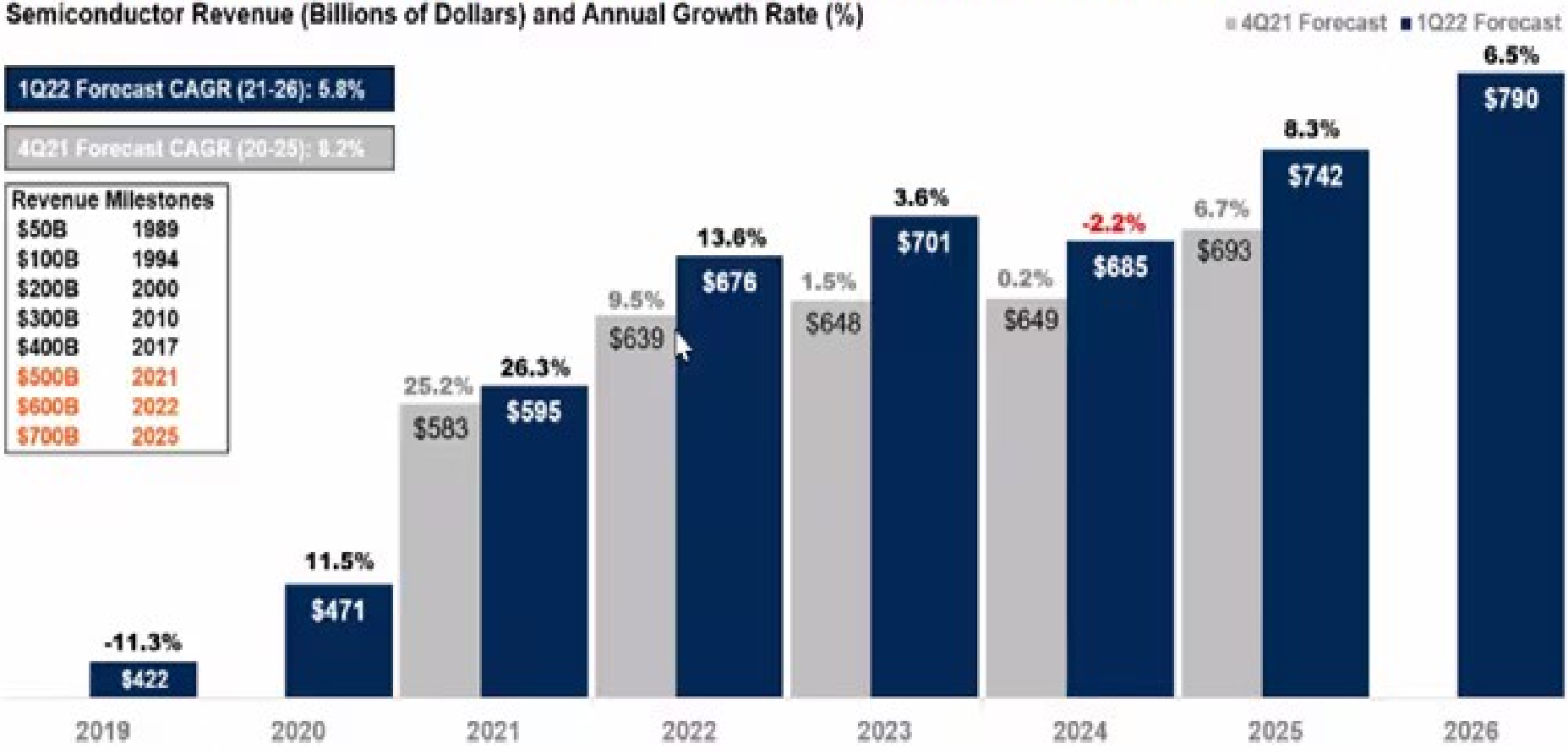
Money, technology, civilization


12,000
9,000
6,000
3,000
0
Number of Design Starts
9,172
9,160
9,406
9,656
9,900
10,081
10,273
10,457
10,644
10,804
10,976
11,126
11,255
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
≥250nm
180/150nm
130nm
90nm
65/55nm
45/40nm
28/22/20/18nm
16/14/12nm
10/7nm
5nm
≤3nm
Mature Nodes

Semiconductor Revenue Forecast, 1Q22 Update

Chip Shortage Will Ease Throughout 2022 Leading to Lower ASP Growth Compared to 2021

Semiconductor Revenue (Billions of Dollars) and Annual Growth Rate (%)



Source: Semiconductor Forecast Database, Worldwide, 1Q22 Update

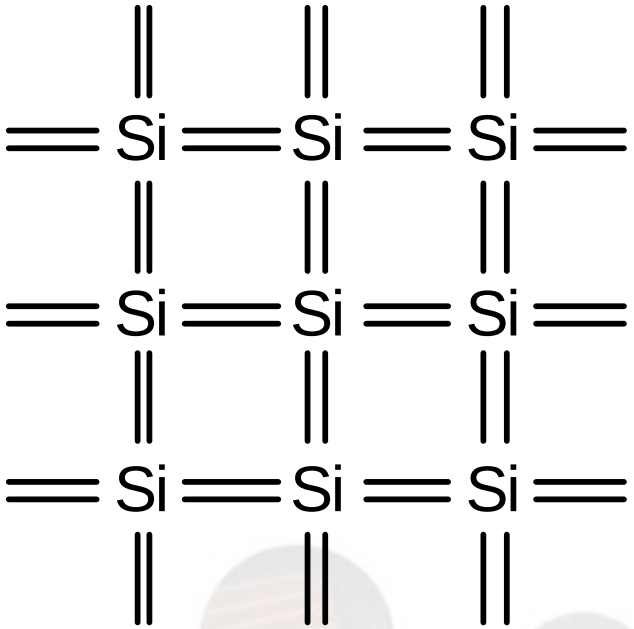
Gartner

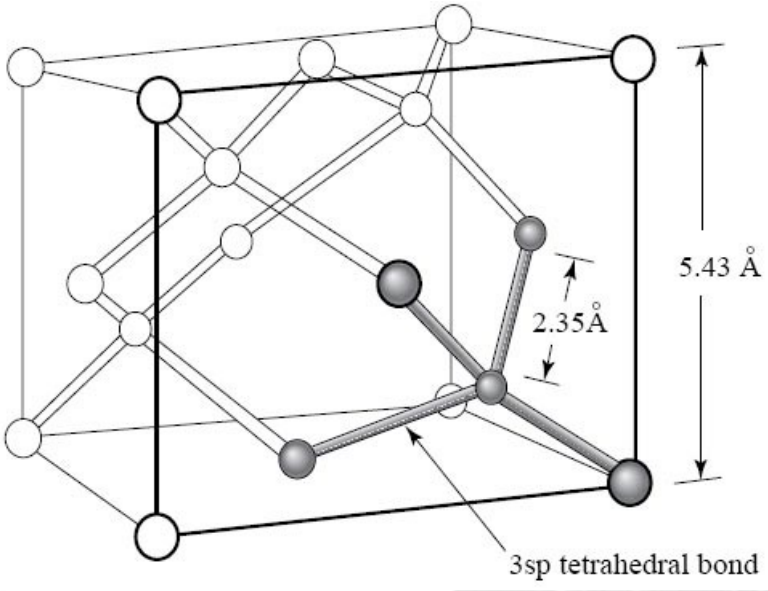
Digression: Silicon Semiconductors

Modern electronic chips are built mostly on silicon substrates

Silicon is a Group IV semi conducting material

crystal lattice: covalent bonds hold each atom to four neighbors



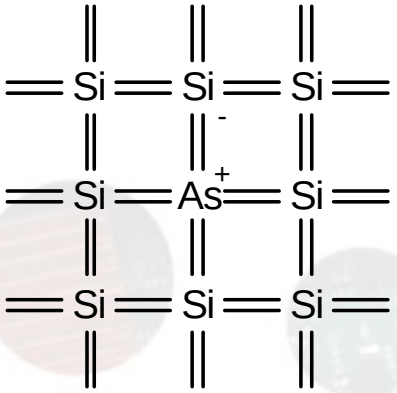


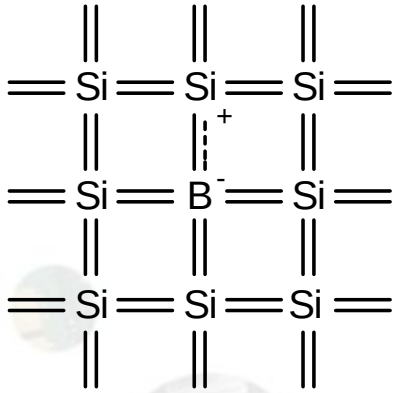
http://onlineheavytheory.net/silicon.html

Dopants

- Silicon is a **semiconductor at room temperature**
- Pure silicon has **few free carriers** and **conducts poorly**
- Adding **dopants increases** the conductivity drastically
- Dopant from **Group V (e.g. As, P)**: extra electron (n-type)
- Dopant from **Group III (e.g. B, Al)**: missing electron, called hole (p-type)

- Aluminium, Boron, Arsenic, Phosphorus



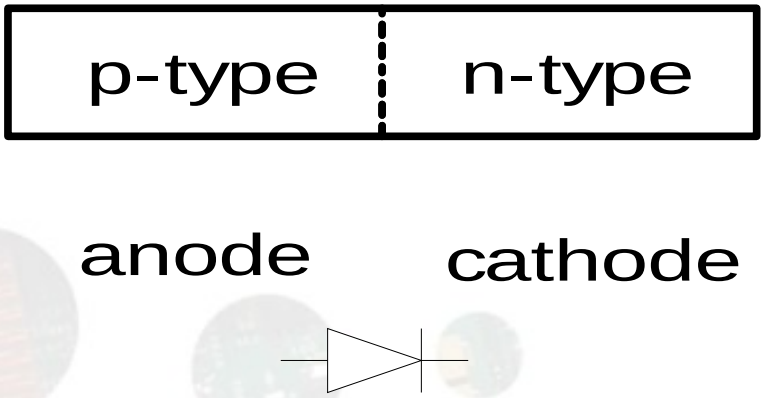


p-n Junctions

First semiconductor (two terminal) devices

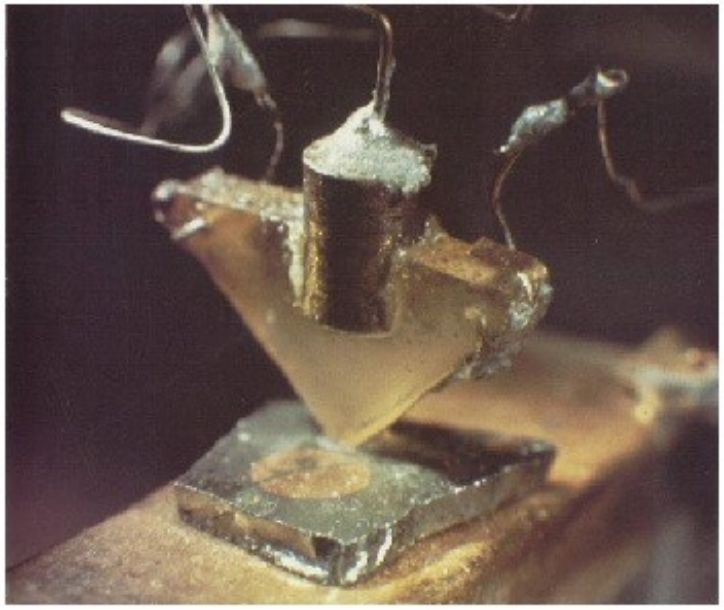
A junction between p-type and n-type semiconductor forms a diode.

Current flows only in one direction



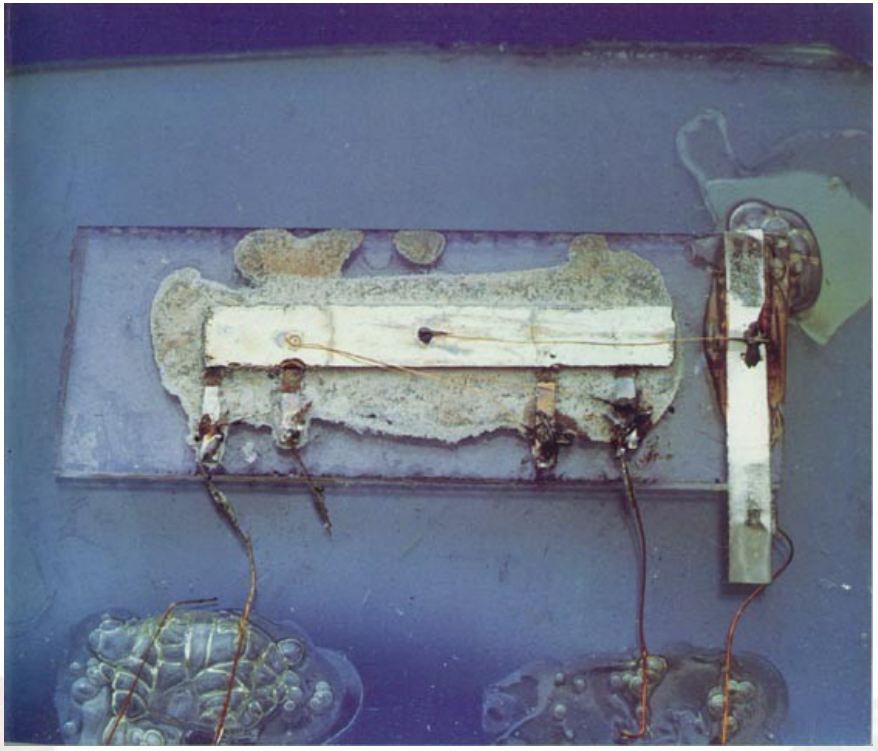
A Brief History
Invention of the Transistor

- Vacuum tubes ruled in first half of 20th century Large, expensive, power-hungry, unreliable
- 1947: first point contact transistor (3 terminal devices)
  - Shockley, Bardeen and Brattain at Bell Labs



A Brief History, contd..

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

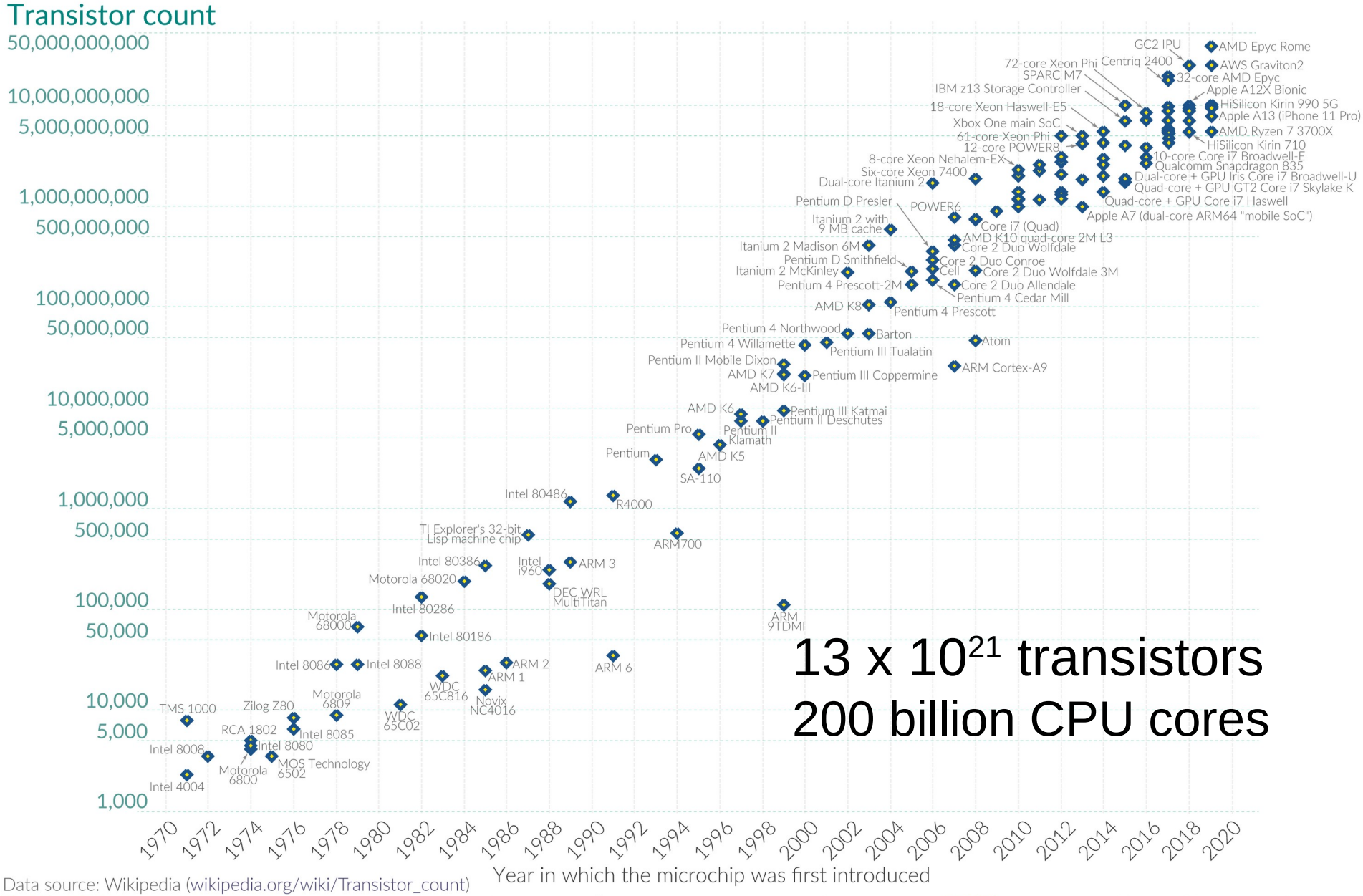


smithsonianchips.si.edu/ augarten/

Moore's Law: The number of transistors on microchips doubles every two years

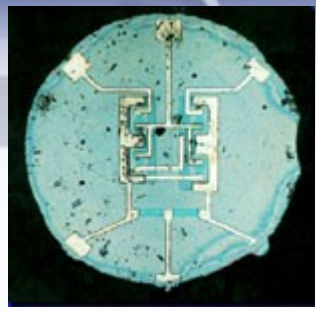
Our World in Data

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

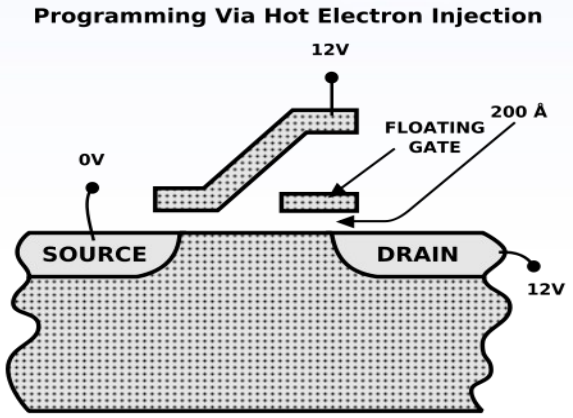


Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)

A Brief History, contd.



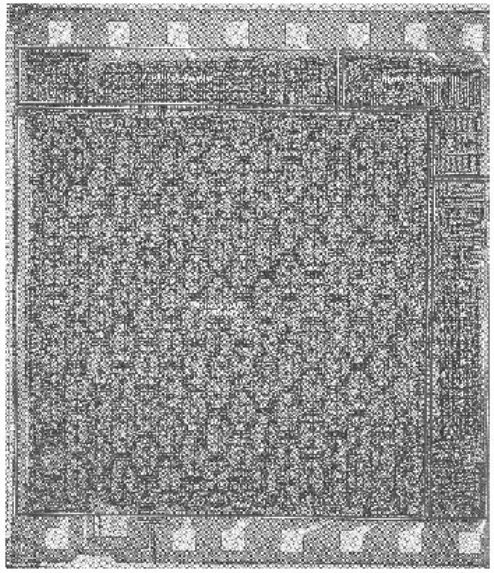
- First Planer IC built in 1961

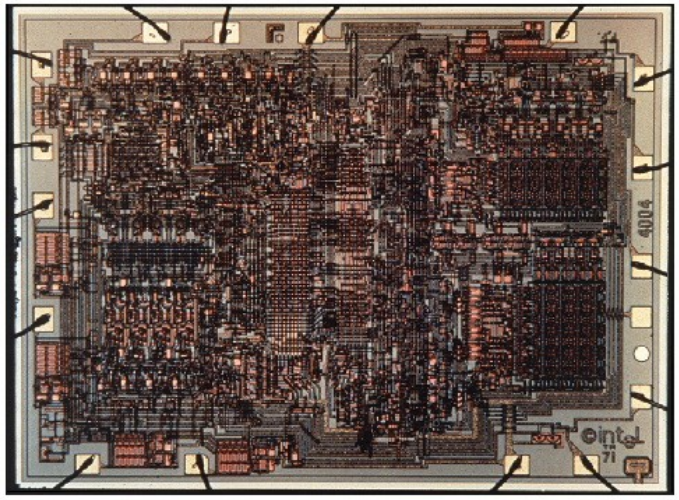


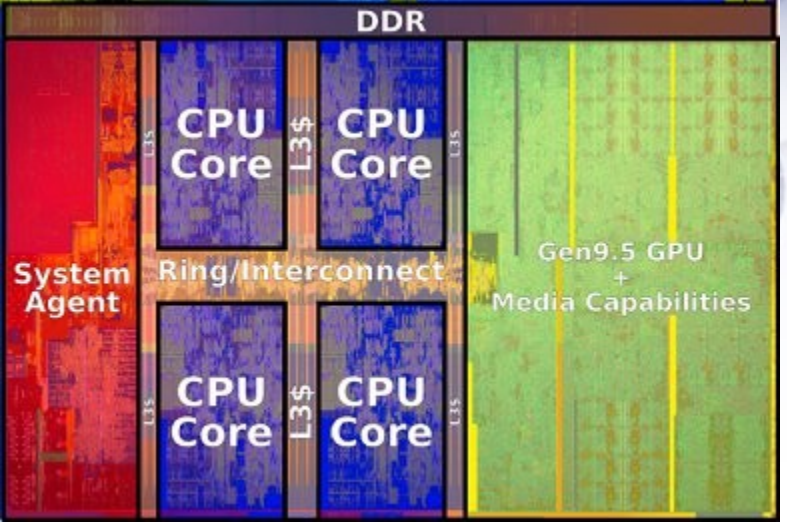
  - Intel Pentium 4 μprocessor (55 million transistors)
  - 512 Mbit DRAM (> 0.5 billion transistors)
  - 1 Tera Byte Flash holds (1 x 10^12 x 8) = 8 x trillion
- 53% compound annual growth rate over 45 years
  - No other technology has grown so fast so long
- Driven by miniaturization of transistors
  - Smaller is cheaper, faster, lower in power!
  - Revolutionary effects on society

MOS Integrated Circuits

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






DDR
CPU Core
CPU Core
CPU Core
CPU Core
L3$
L3$
L3$
L3$
System Agent
Ring/Interconnect
Gen9.5 GPU
+
Media Capabilities

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

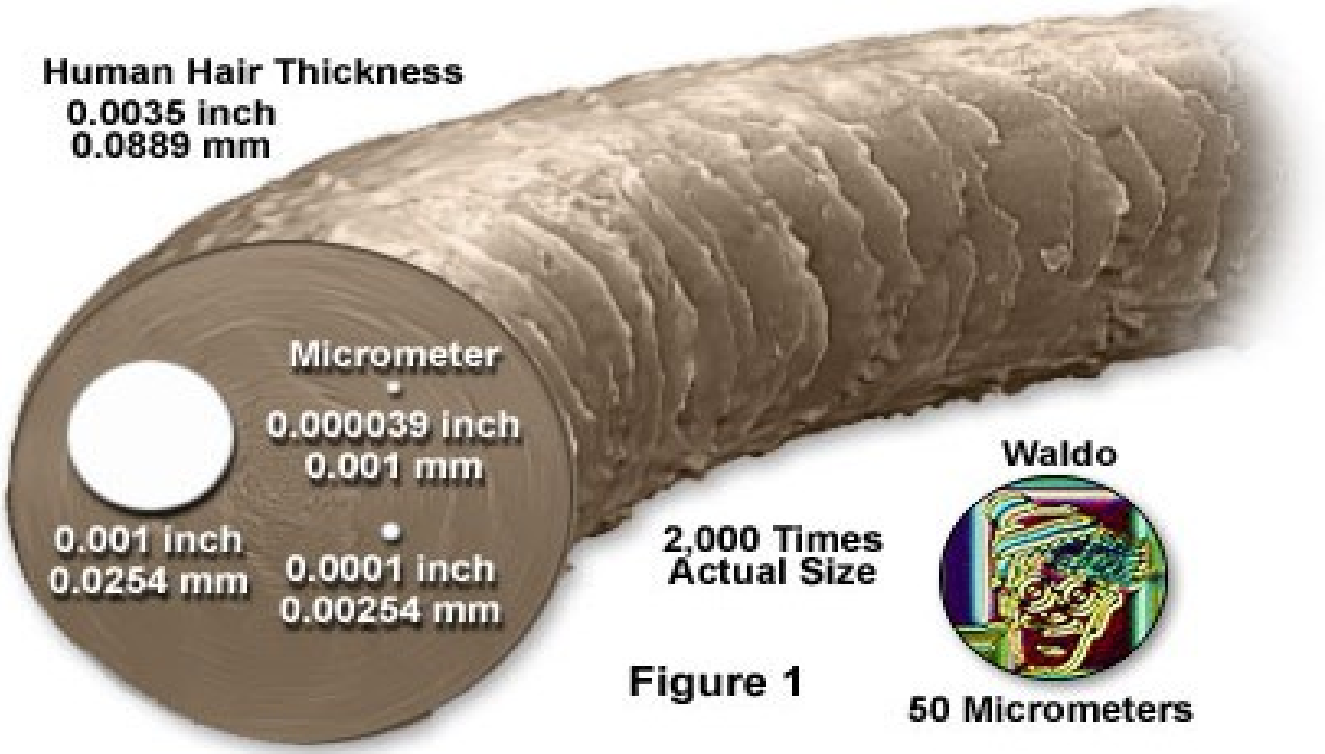


Figure 1

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

Transistor Types

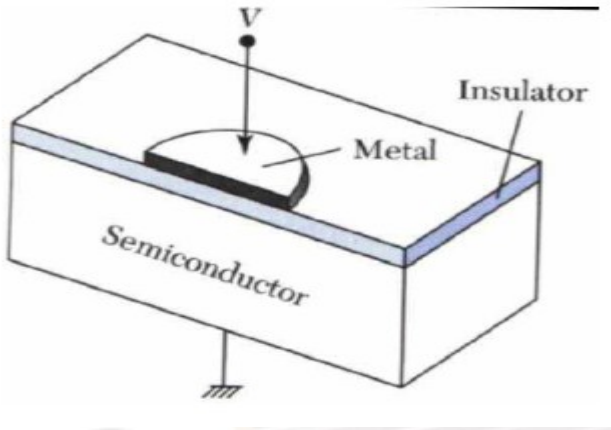
- Bipolar transistors
  - npn or pnp silicon structure
  - Small current into very thin base layer controls large currents between emitter and collector
  - Base currents limit integration density
- Metal Oxide Semiconductor Field Effect Transistors
  - nMOS and pMOS MOSFETS
  - Voltage applied to insulated gate controls current between source and drain
  - Low power allows very high integration
  - First patent in the ’20s in USA and Germany
  - Not widely used until the ’60s or ’70s

What is MOS

(metal-oxide-semiconductor)

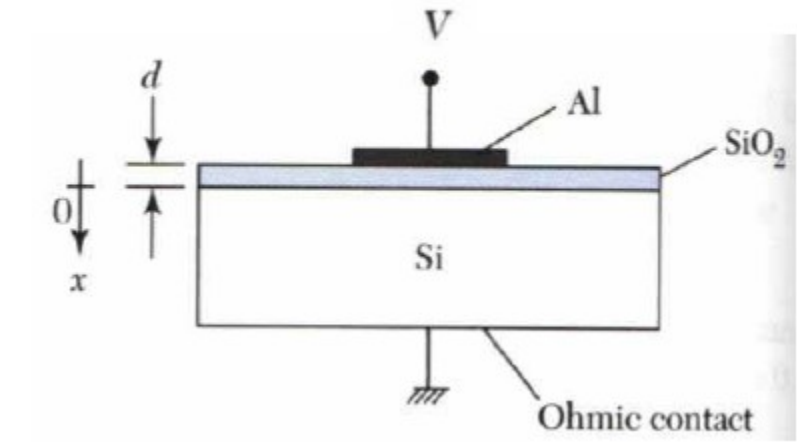
A basic MOS consisting of **three layers**.

- The top layer is a **conductive metal electrode.**
- The middle layer is an **insulator of glass or silicon dioxide.**
- The bottom layer is another **conductive electrode made out of crystal silicon**. This layer is a semiconductor whose conductivity changes with either doping or temperature.



Cross-section Structure of MOS

- **d** is the thickness of the oxide and V is the applied voltage on the metal field plate
- **V>(<)0** metal plate is positively (negatively) biased with respect to the ohmic contact



When an ideal MOS diode is biased with positive or negative voltages, three case may exist at the semiconductor surface

A. accumulation

B. depletion

C. inversion

MOS Transistors

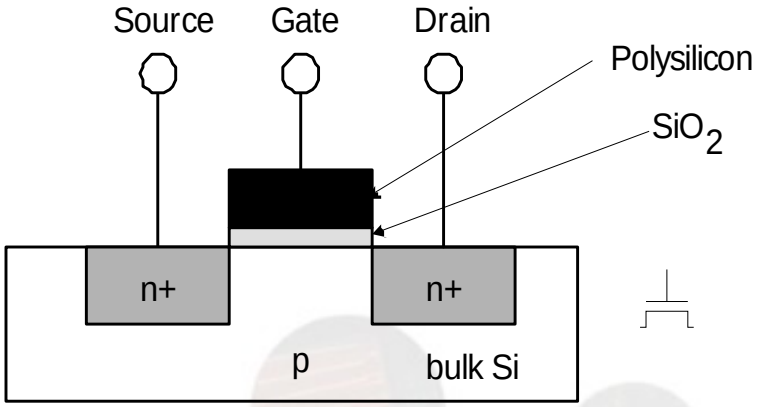
Four terminal device: gate, source, drain, body

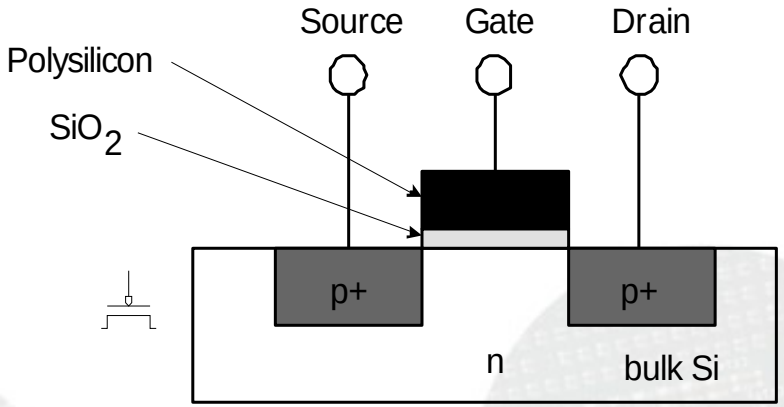
Gate – oxide – body stack looks like a capacitor

Gate and body are conductors (body is also called the substrate)

SiO_2 (oxide) is a “good” insulator (separates the gate from the body

Called metal–oxide–semiconductor (MOS) capacitor, even though gate is mostly made of poly-crystalline silicon (polysilicon)





NMOS Operation

Body is commonly tied to ground (0 V)

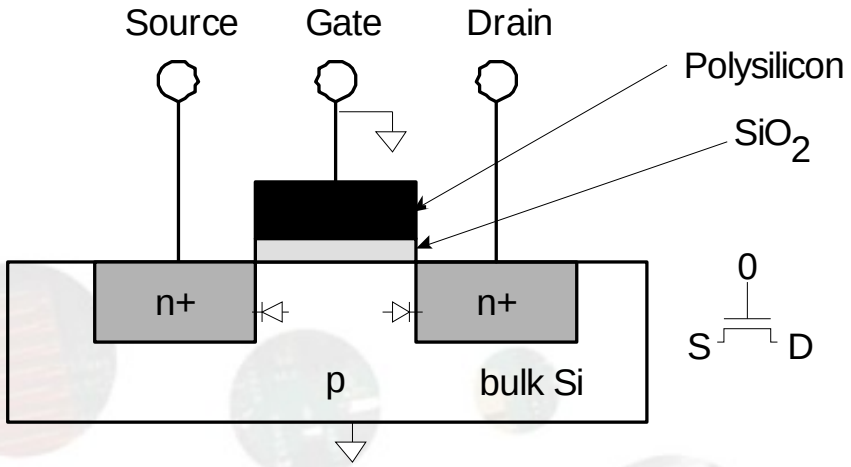
Drain is at a higher voltage than Source

When the gate is at a low voltage:

P-type body is at low voltage

Source-body and drain-body “diodes” are OFF

No current flows, transistor is OFF



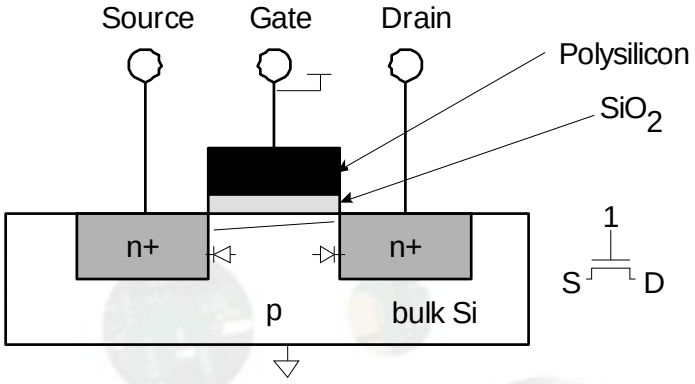
NMOS Operation Cont.

When the gate is at a high voltage: Positive charge on gate of MOS capacitor

Negative charge is attracted to body under the gate

Inverts a channel under gate to “n-type” (N-channel, hence called the NMOS) if the gate voltage is above a threshold voltage (VT)

Now current can flow through “n-type” silicon from source through channel to drain, transistor is ON



PMOS Transistor

Similar, but doping and voltages reversed

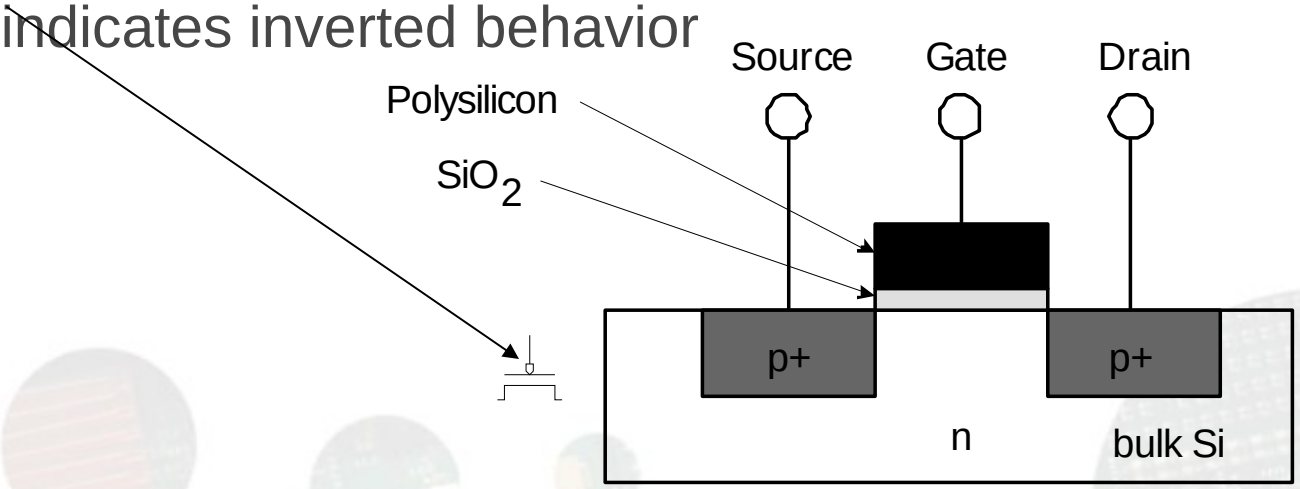
Body tied to high voltage (V_{DD})

Drain is at a lower voltage than the Source

Gate low: transistor ON

Gate high: transistor OFF

Bubble indicates inverted behavior



Power Supply Voltage

GND = 0 V

In 1980's, V_{DD} = 5V

V_{DD} has decreased in modern processes

High V_{DD} would damage modern tiny transistors

Lower V_{DD} saves power

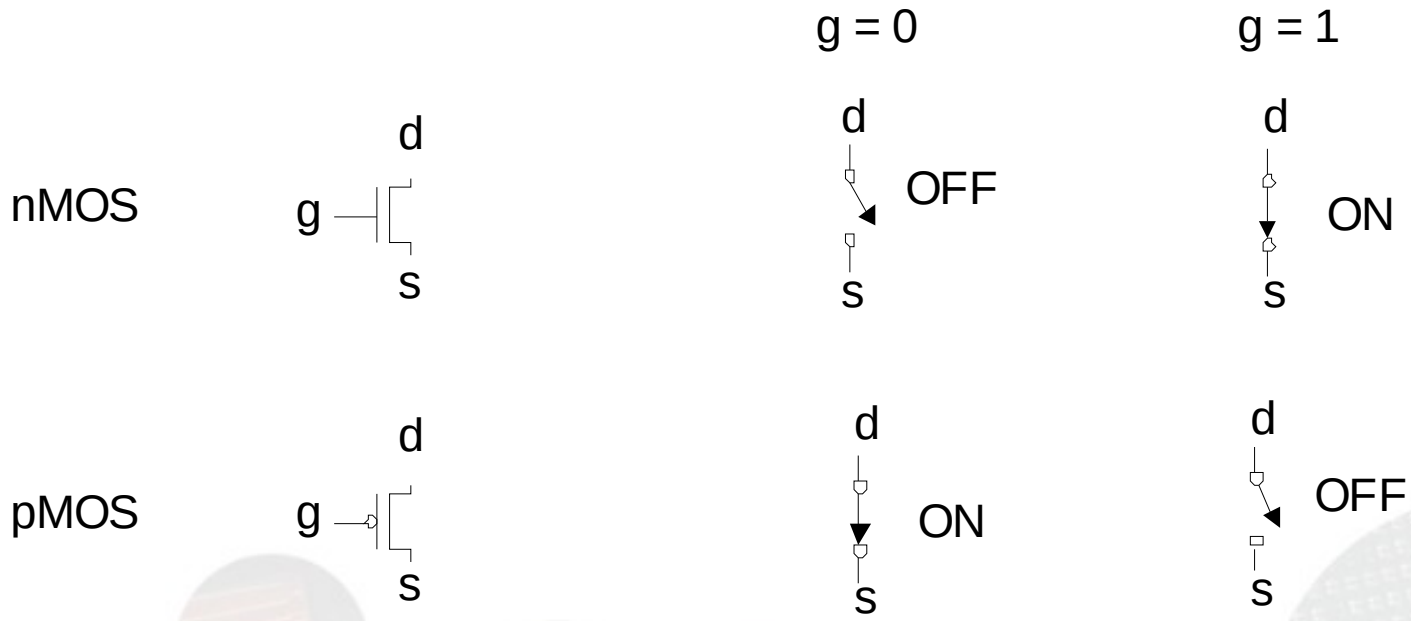
V_{DD} = 3.3, 2.5, 1.8, 1.5, 1.2, 1.0,

Effective power supply voltage can be lower due to IR drop across the power grid.

Transistors as Switches

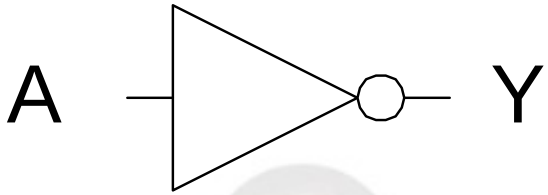
In Digital circuits, MOS transistors are electrically controlled switches

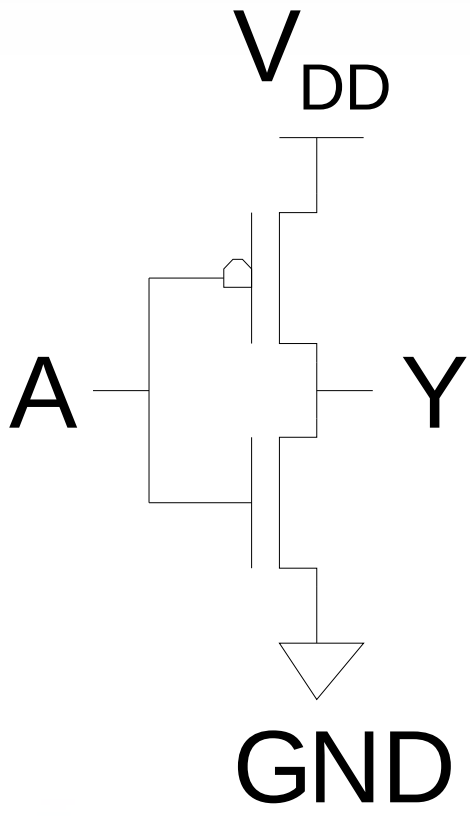
Voltage at gate controls path from source to drain



CMOS Inverter

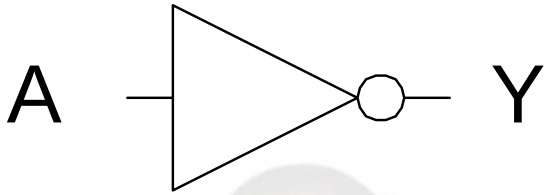
A	Y
0	
1	

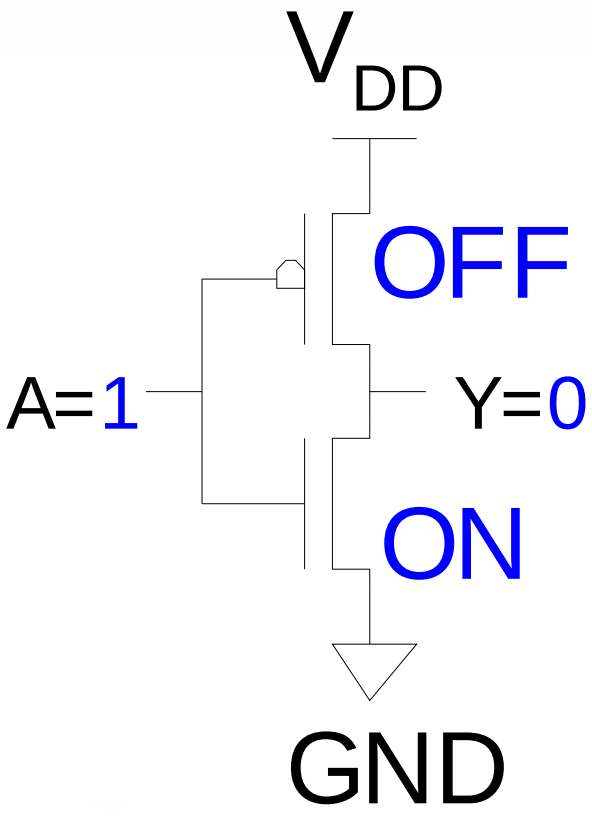




CMOS Inverter

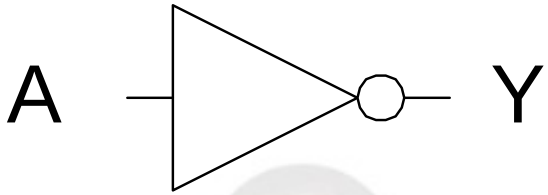
A	Y
0	
1	**0**

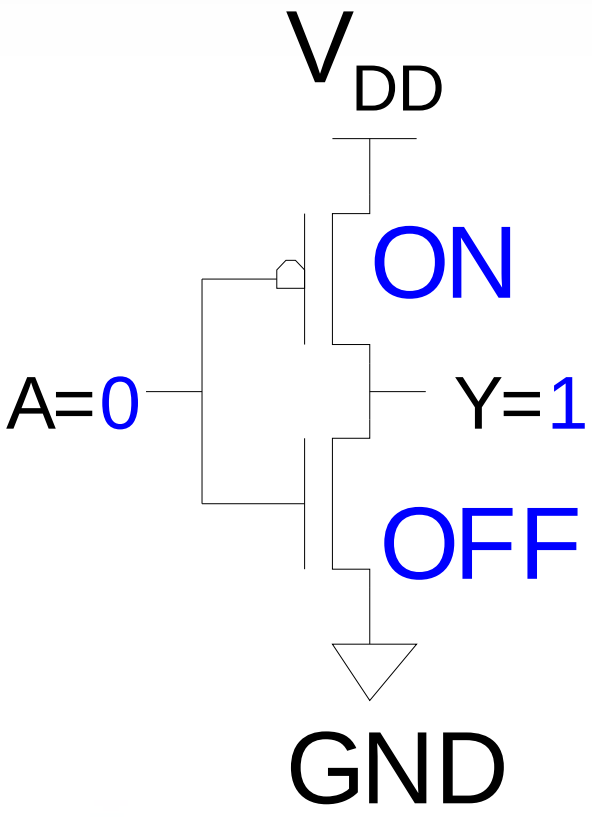




CMOS Inverter

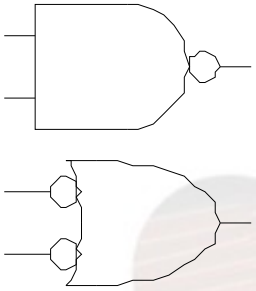
A	Y
0	**1**
1	0





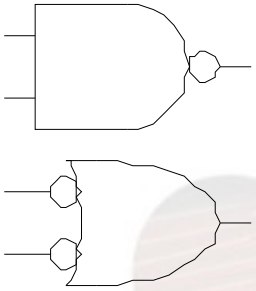
CMOS NAND Gate

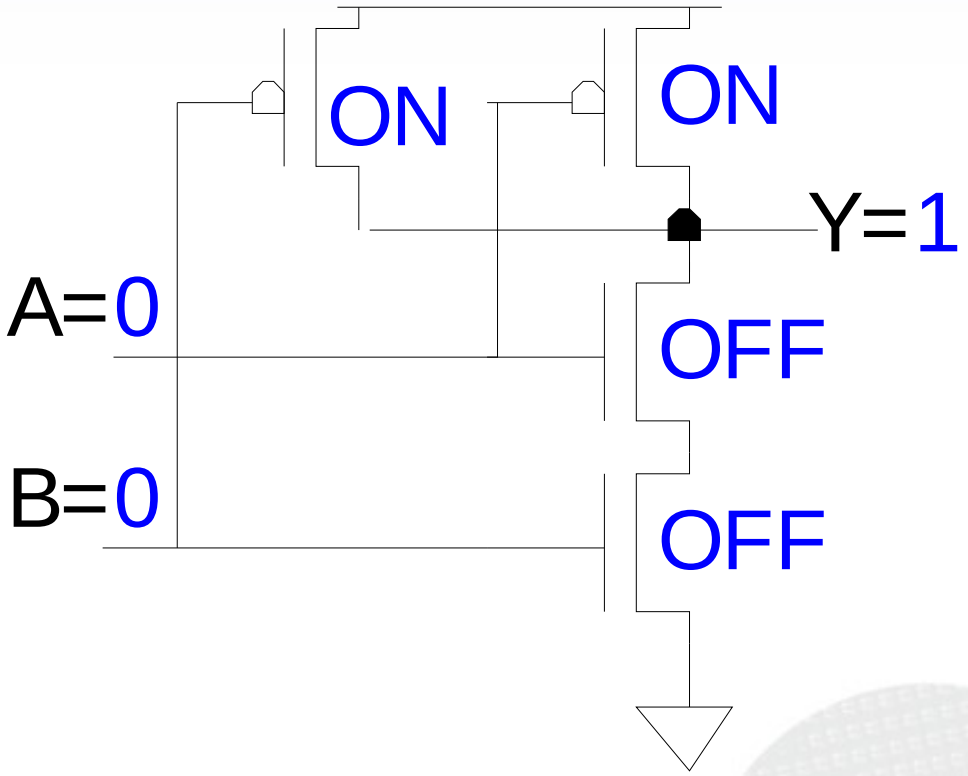
A	B	Y
0	0	
0	1	
1	0	
1	1	



CMOS NAND Gate

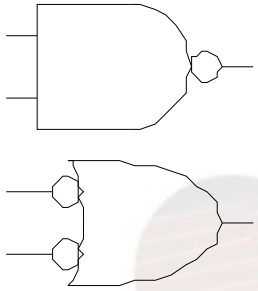
A	B	Y
0	**0**	**1**
0	1	
1	0	
1	1	

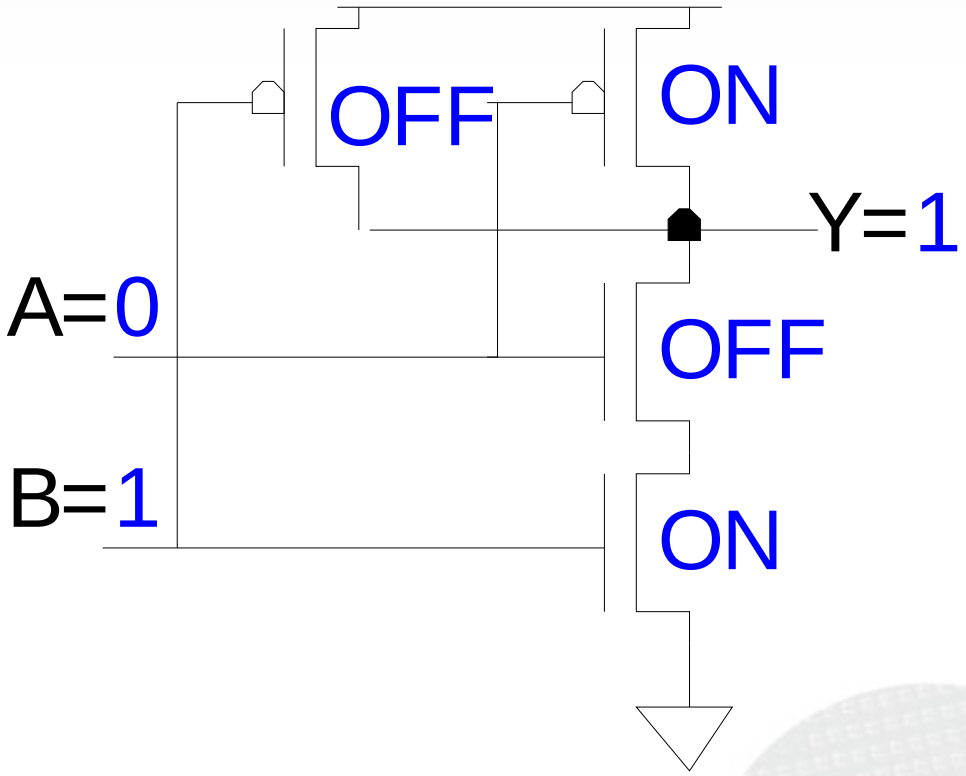




CMOS NAND Gate

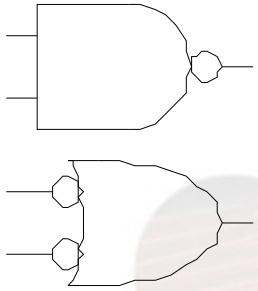
A	B	Y
0	0	1
0	**1**	**1**
1	0	
1	1	

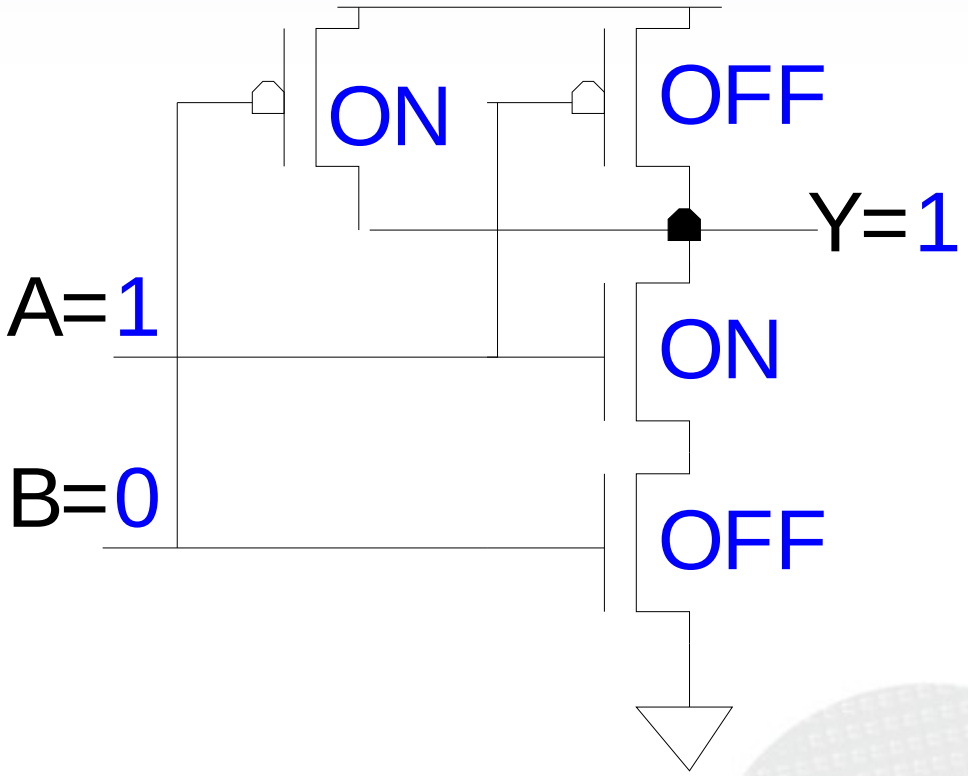




CMOS NAND Gate

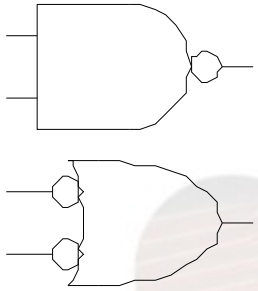
A	B	Y
0	0	1
0	1	1
1	**0**	**1**
1	1	

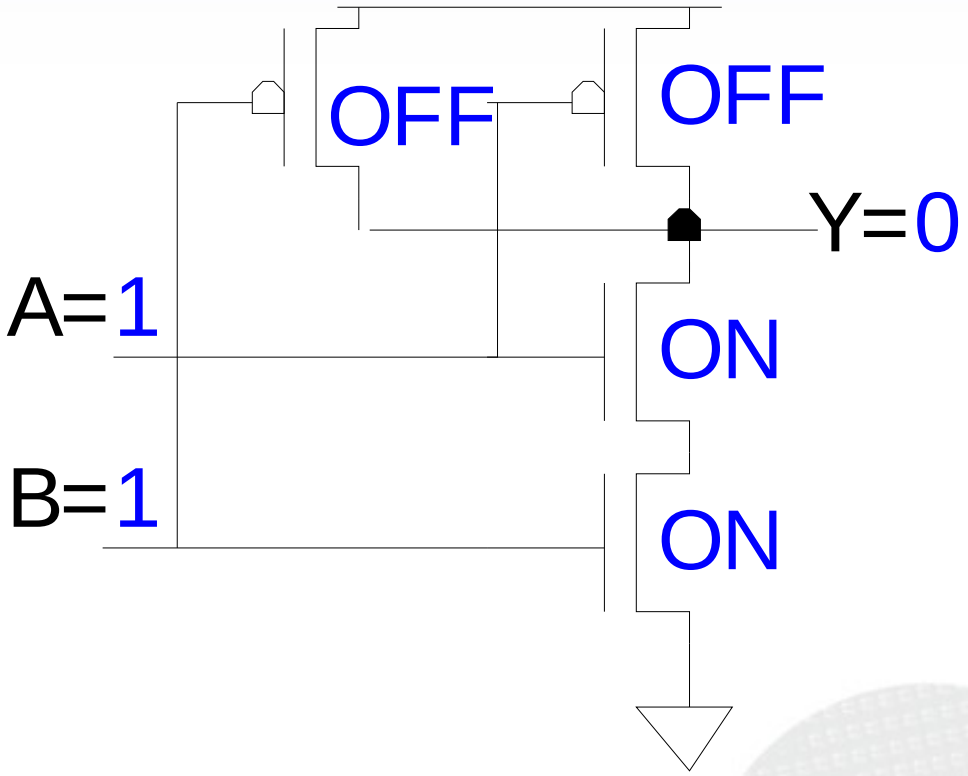




CMOS NAND Gate

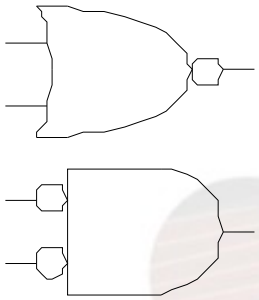
A	B	Y
0	0	1
0	1	1
1	0	1
1	**1**	**0**

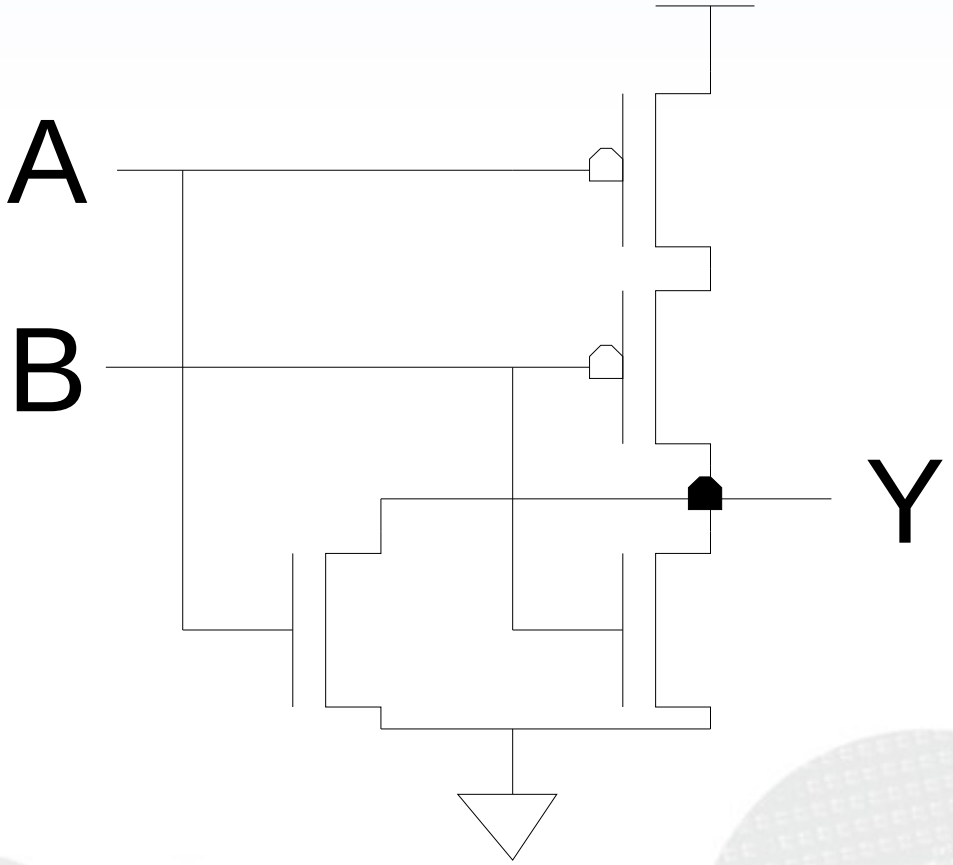




CMOS NOR Gate

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

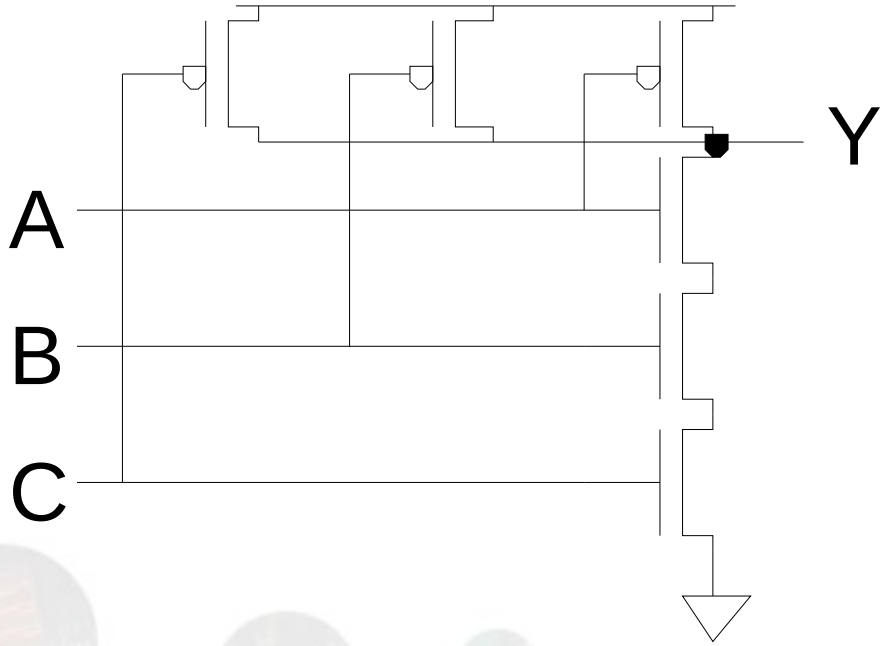




3-input NAND Gate

Y is pulled low if ALL inputs are 1

Y is pulled high if ANY input is 0



CMOS Fabrication

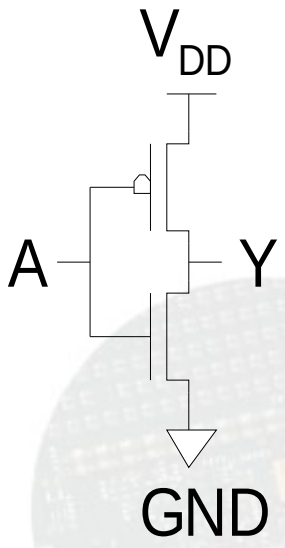
CMOS transistors are fabricated on silicon wafer

Wafers diameters (200-300 mm)

Lithography process similar to printing press

On each step, **different materials are deposited**, or patterned or etched.

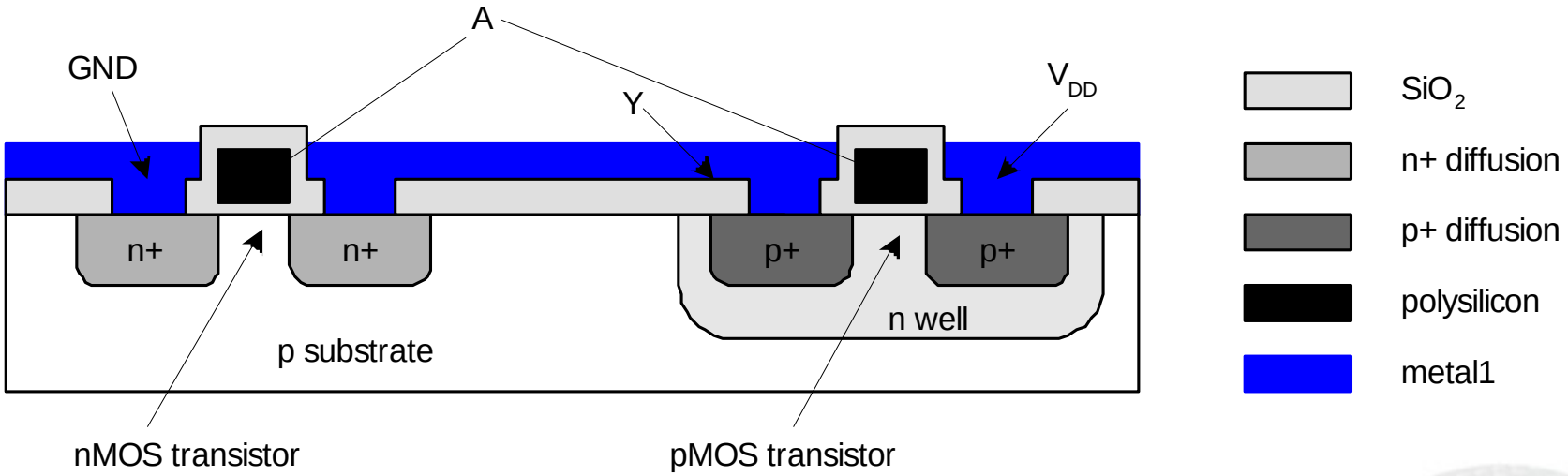
Easiest to understand by viewing both top and cross-section of wafer in a simplified manufacturing process



Inverter Cross-section

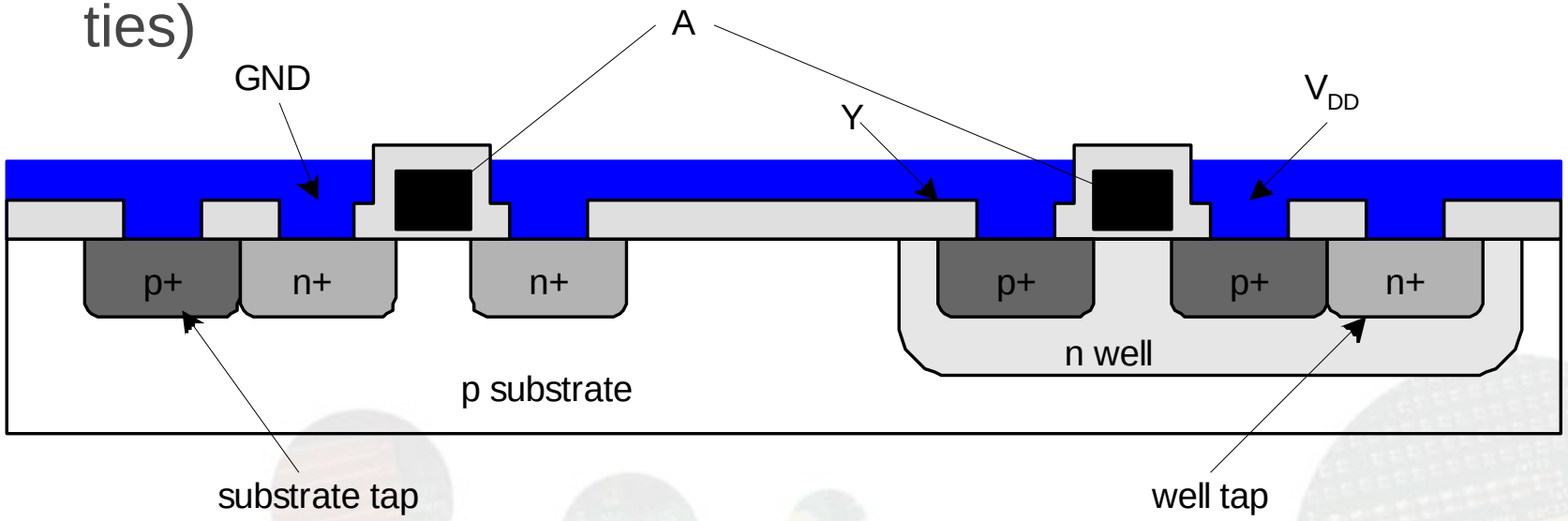
Typically use p-type substrate for nMOS transistors

Requires to make an n-well for body of pMOS transistors



Well and Substrate Taps

- Substrate must be tied to GND and n-well to V_{DD}
- Metal to lightly-doped semiconductor forms poor connection called Schottky Diode
- Use heavily doped well and substrate contacts/taps (or ties)

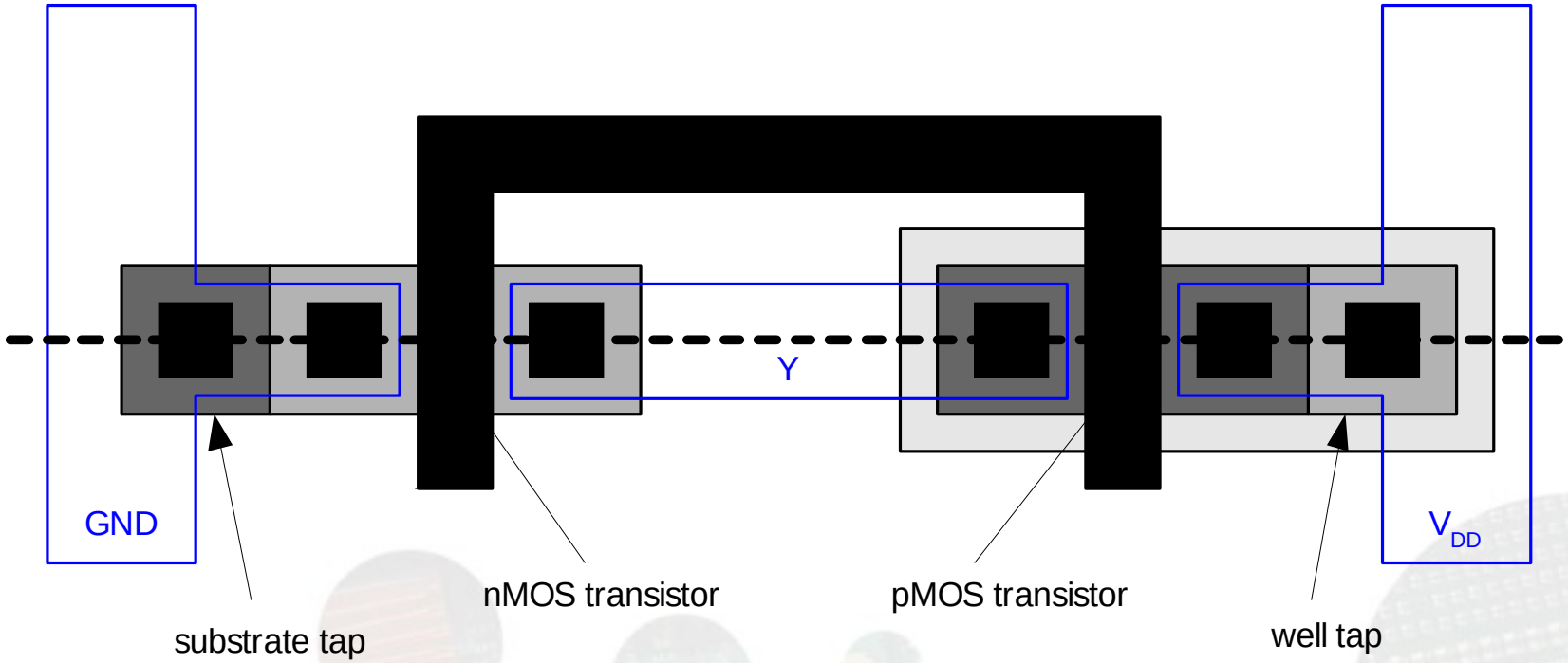


Inverter Mask Set

Top view

Transistors and wires are defined by *masks*

Cross-section taken along dashed line



Detailed Mask Views

Six masks

n-well

Polysilicon

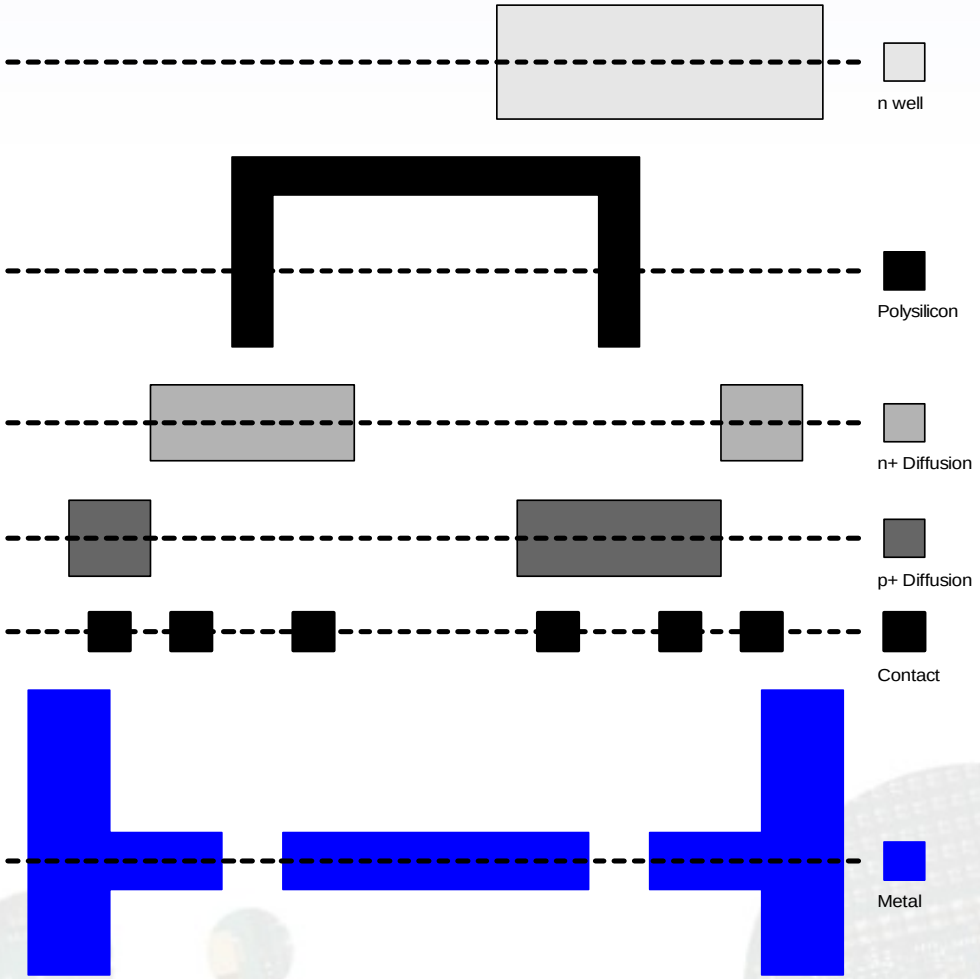
n+ diffusion

p+ diffusion

Contact

Metal

In reality >40 masks may be needed



Fabrication Steps

Start with blank wafer (typically p-type where NMOS is created)

Build inverter from the bottom up

First step will be to form the n-well (where PMOS would reside)

Cover wafer with protective layer of SiO_2 (oxide)

Remove oxide layer where n-well should be built

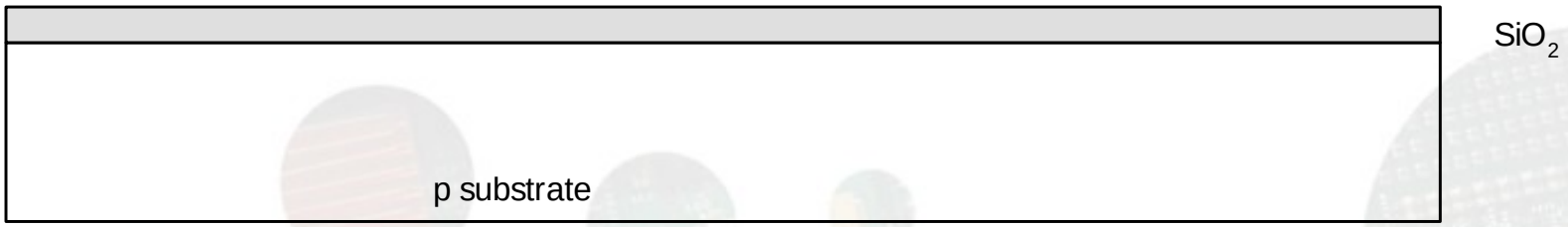
Implant or diffuse n dopants into exposed wafer to form n-well

Strip off SiO_2

Oxidation

Grow SiO_2 on top of Si wafer

900 – 1200 C with H_2O or O_2 in oxidation furnace



Photoresist

Spin on photoresist

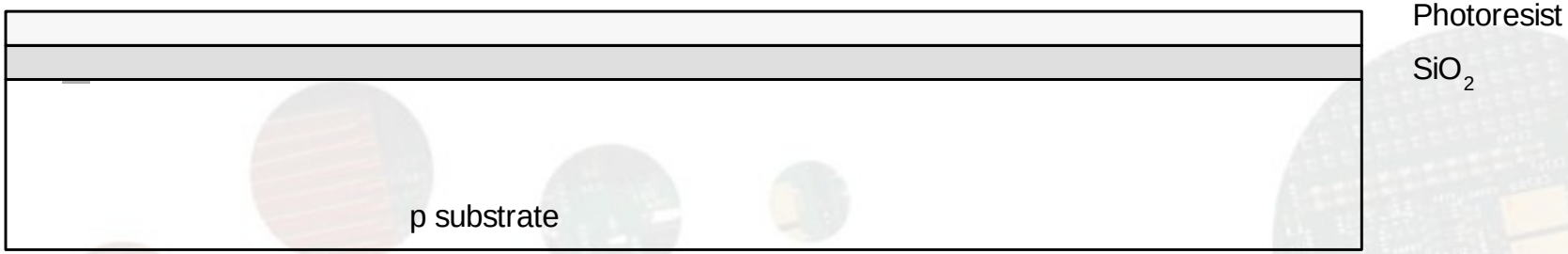
Photoresist is a light-sensitive organic polymer

Property changes where exposed to light

Two types of photoresists (positive or negative)

Positive resists can be removed if exposed to UV light

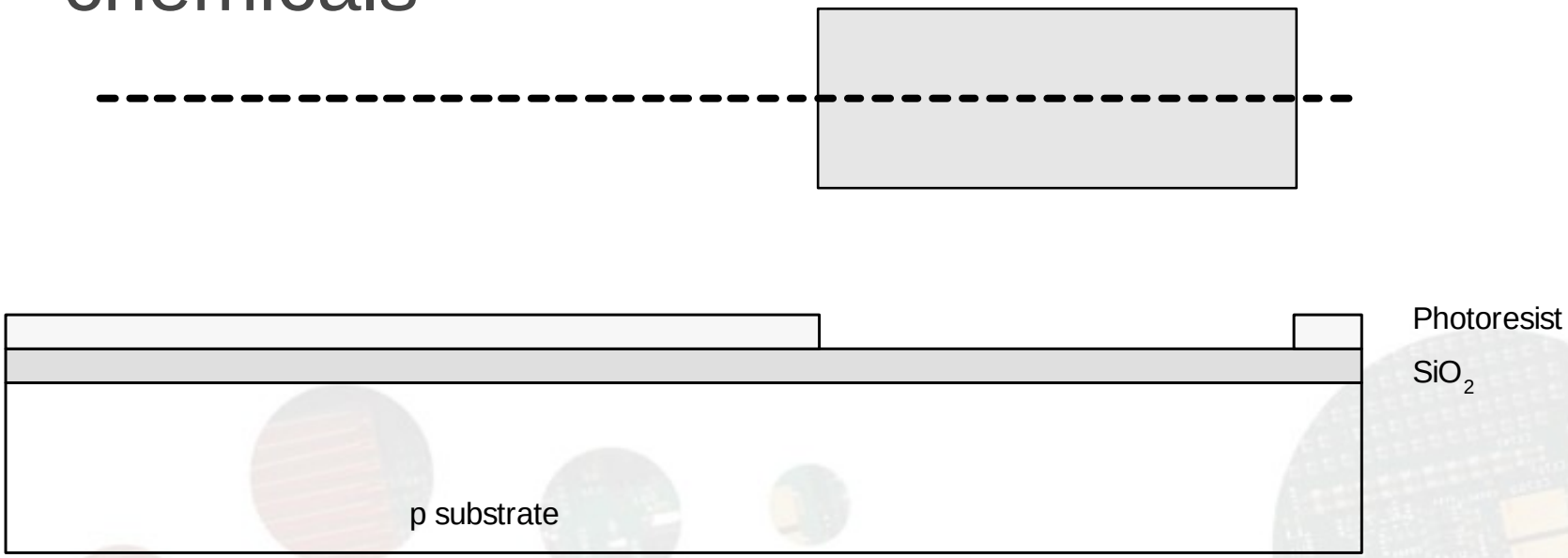
Negative resists cannot be removed if exposed to UV light



Lithography

Expose photoresist to Ultra-violate (UV) light through the n-well mask

Strip off exposed photoresist with chemicals



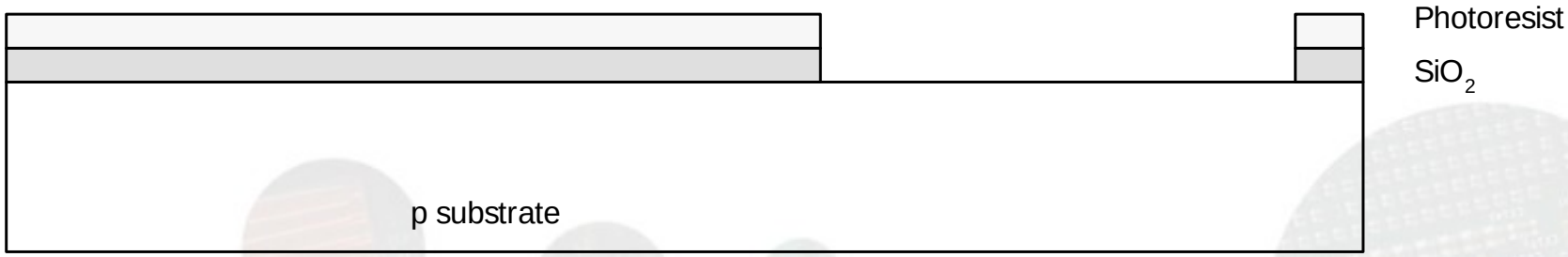
Etch

Etch oxide with hydrofluoric acid (HF)

Seeps through skin and eats bone; nasty stuff!!!

Only attacks oxide where resist has been exposed

N-well pattern is transferred from the mask to silicon-di-oxide surface; creates an opening to the silicon surface



Strip Photoresist

Strip off remaining photoresist

Use mixture of acids called piranah etch

Necessary so resist doesn't melt in next step

n-well

n-well is formed with diffusion or ion implantation

Diffusion

Place wafer in furnace with arsenic-rich gas

Heat until As atoms diffuse into exposed Si

Ion Implanatation

Blast wafer with beam of As ions

Ions blocked by SiO_2, only enter exposed Si

SiO_2 shields (or masks) areas which remain p-type

Strip Oxide

Strip off the remaining oxide using HF

Back to bare wafer with n-well

Subsequent steps involve similar series of steps

Polysilicon
(self-aligned gate technology)

Deposit very thin layer of gate oxide

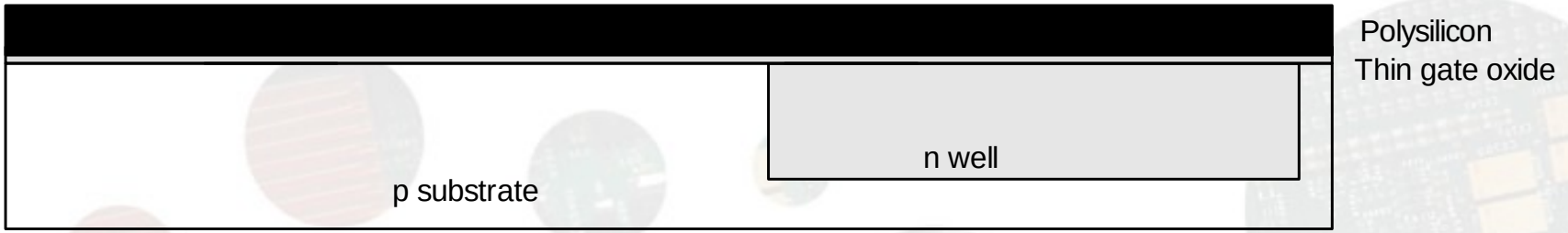
< 20 Å (6-7 atomic layers)

Chemical Vapor Deposition (CVD) of silicon layer

Place wafer in furnace with Silane gas (SiH_4)

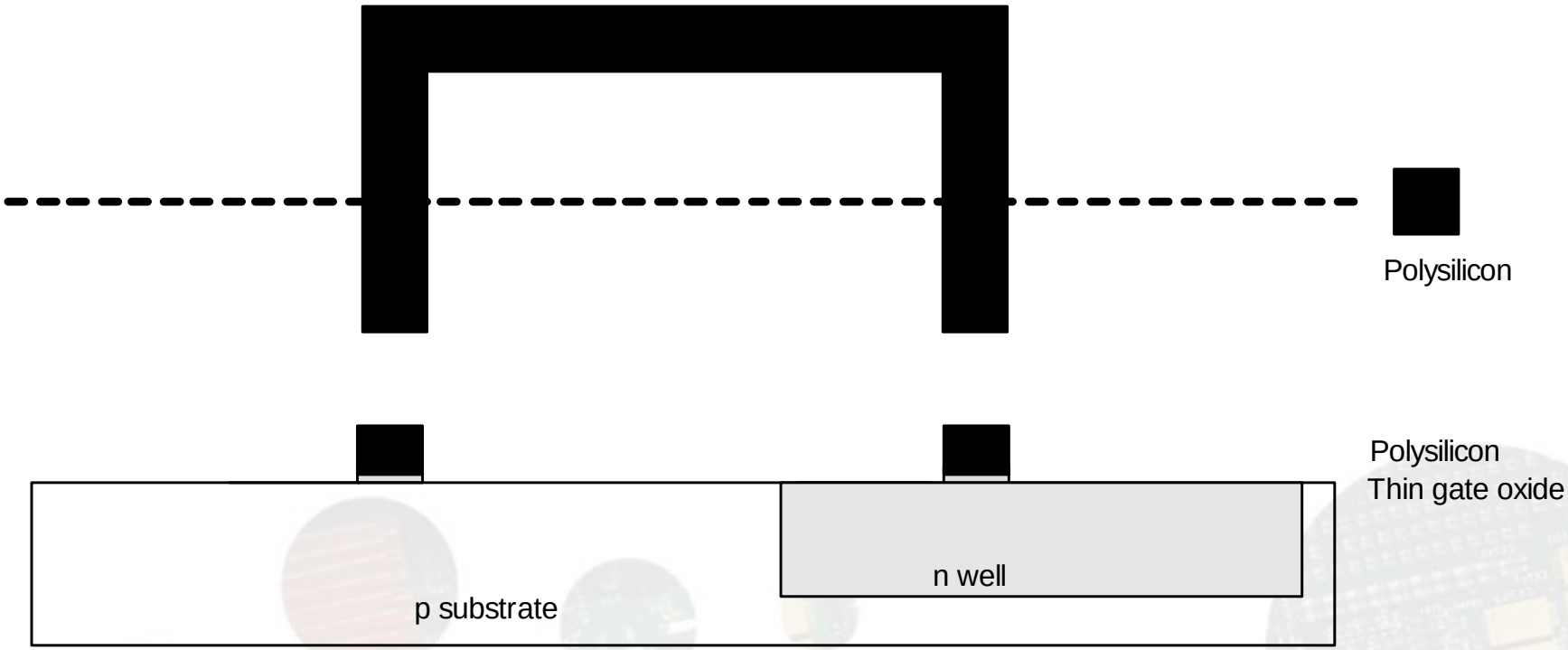
Forms many small crystals called polysilicon

Heavily doped to be good conductor



Polysilicon Patterning

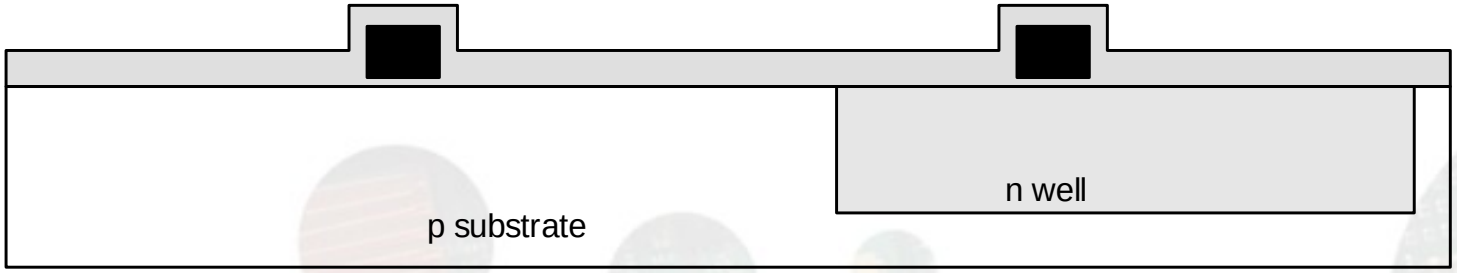
Use same lithography process discussed earlier to pattern polysilicon



Self-Aligned Process

Use gate-oxide/polysilicon and masking to expose where n+ dopants should be diffused or implanted

N-diffusion forms nMOS source, drain, and n-well contact

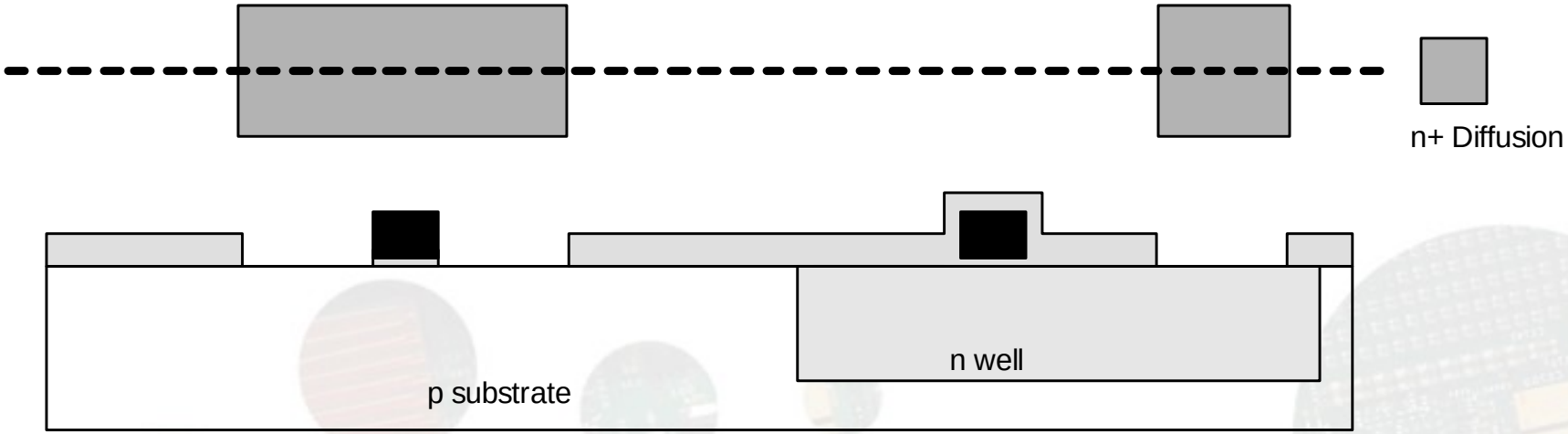


N-diffusion/implantation

Pattern oxide and form n+ regions

Self-aligned process where gate blocks n-dopants

Polysilicon is better than metal for self-aligned gates because it doesn't melt during later processing

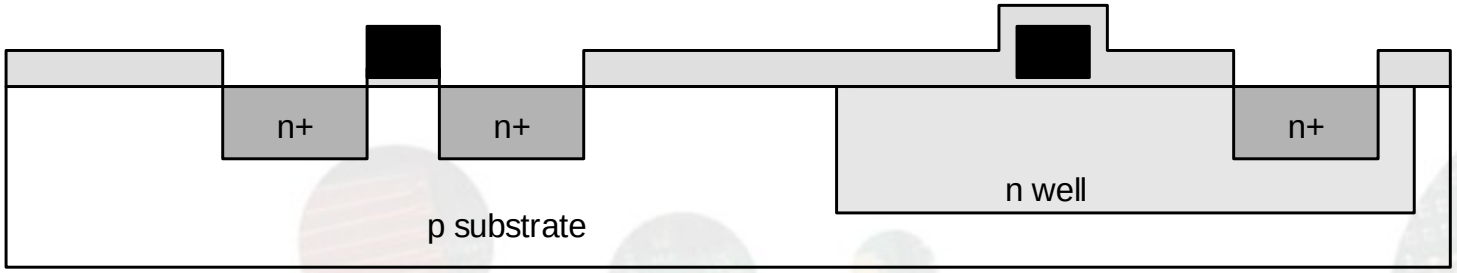


N-diffusion/implantation cont.

Historically dopants were diffused

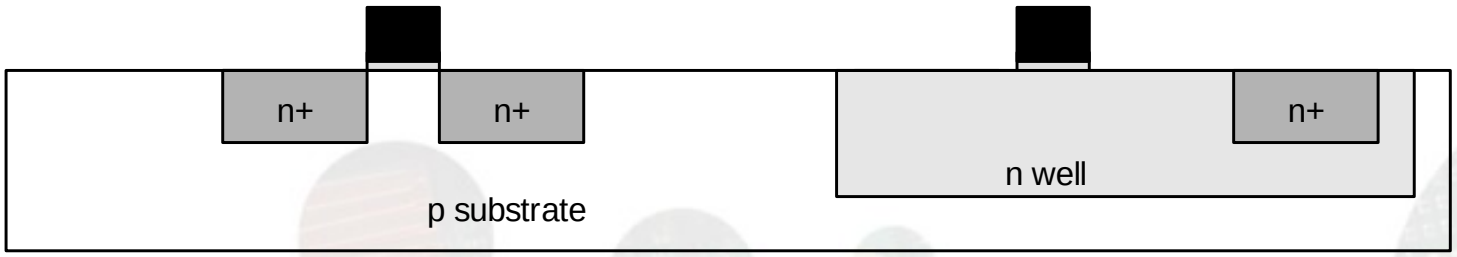
Usually high energy ion-implantation used today

But n+ regions are still called diffusion



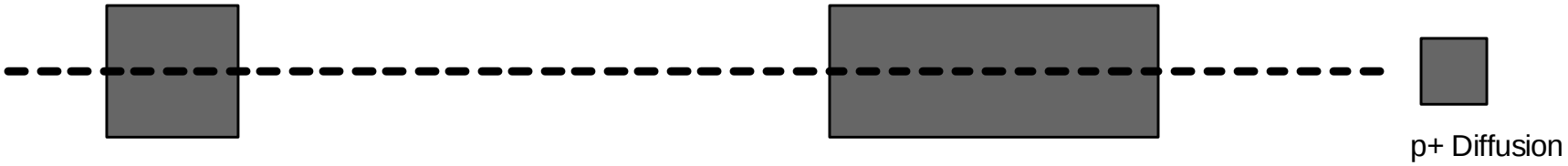
N-diffusion cont.

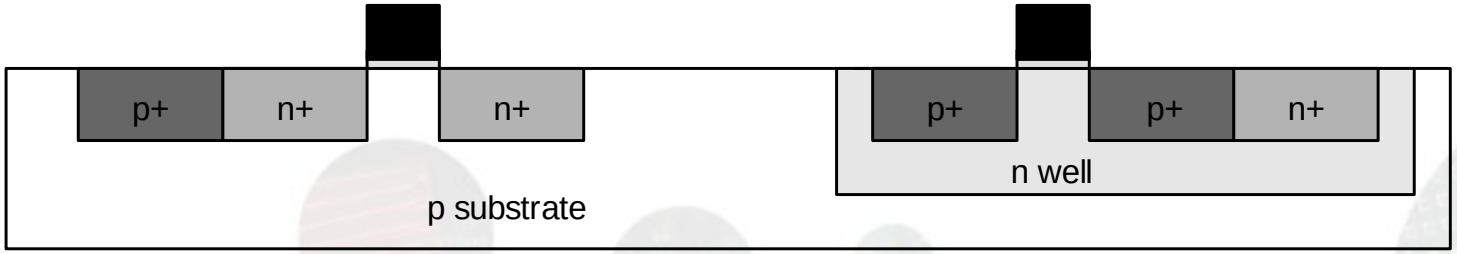
Strip off oxide to complete patterning step



P-Diffusion/implantation

Similar set of steps form p+ “diffusion” regions for PMOS source and drain and substrate contact



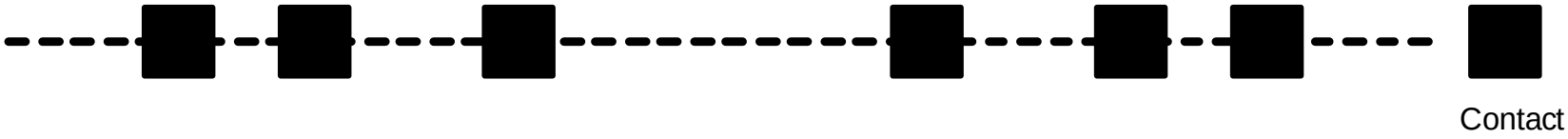


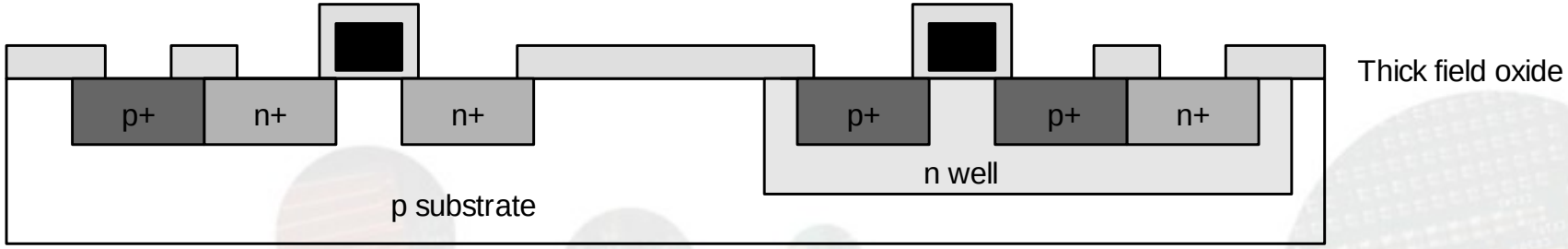
Contacts

Now we need to wire together the devices

Cover chip with thick field oxide (FO)

Etch oxide where contact cuts are needed



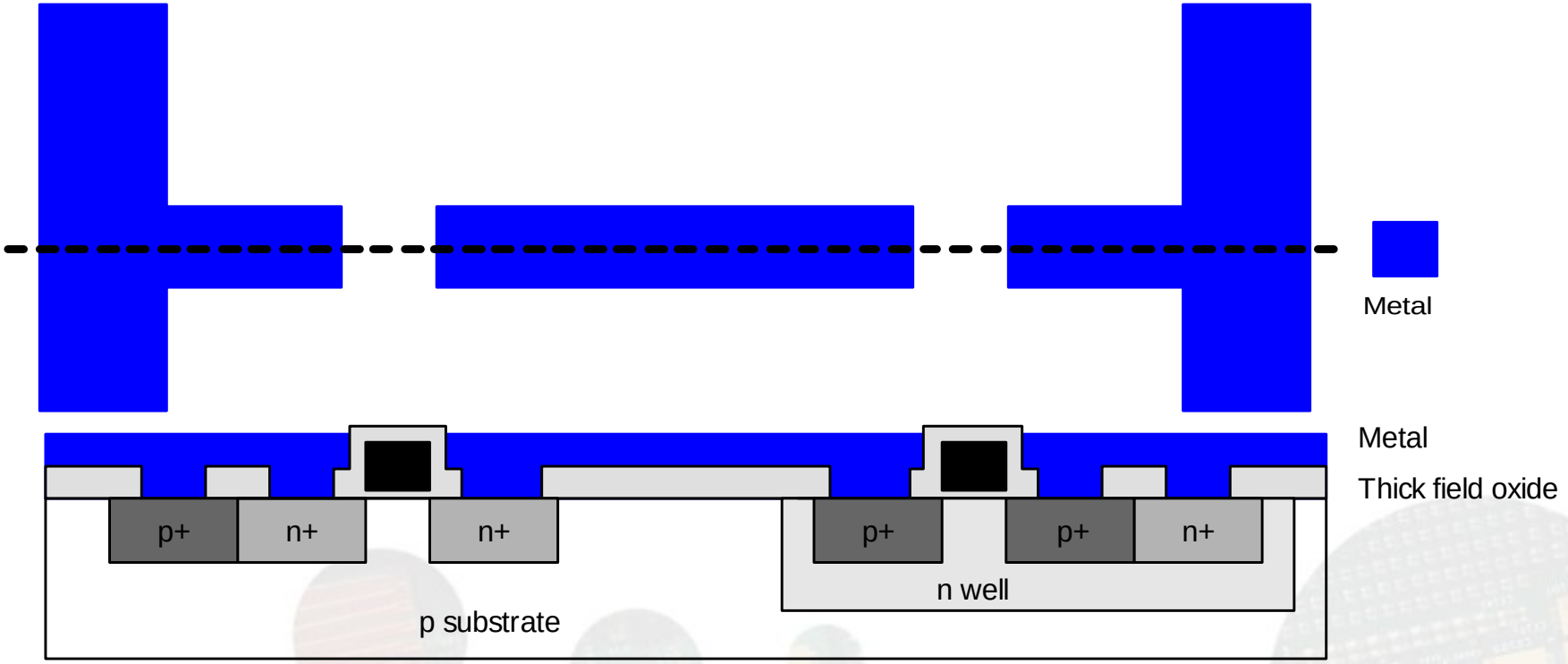


Metalization

Sputter on aluminum over whole wafer

Copper is used in newer technology

Pattern to remove excess metal, leaving wires



Physical Layout

Chips are specified with set of masks

Minimum dimensions of masks determine transistor size (and hence speed, cost, and power)

Feature size *f* = distance between source and drain

Set by minimum width of polysilicon

Feature size improves 30% every 3 years or so

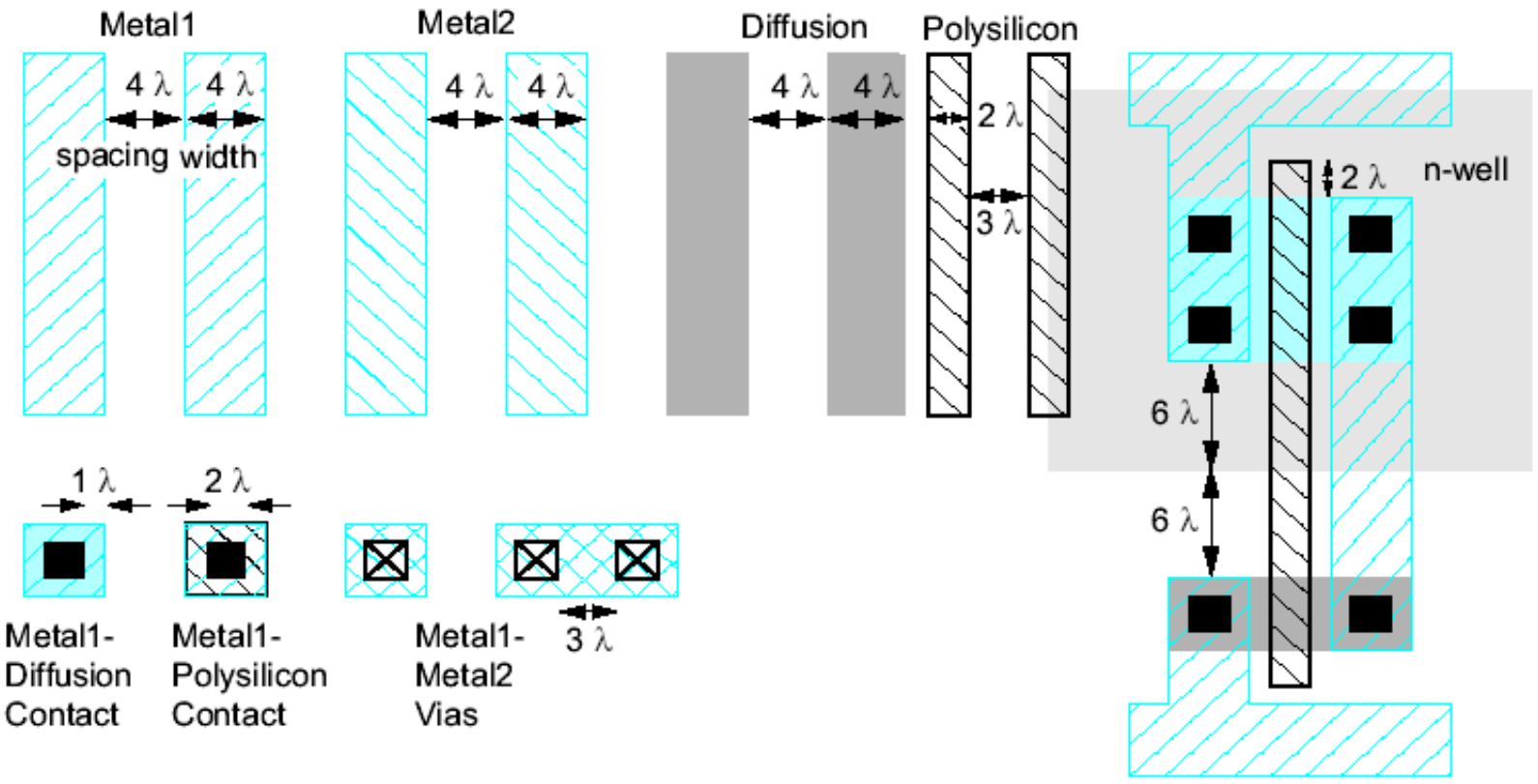
Normalize for feature size when describing design rules

Express rules in terms of $\lambda = f/2$

E.g. λ = 0.3 μm in 0.6 μm process

Simplified Design Rules

Conservative rules to get you started



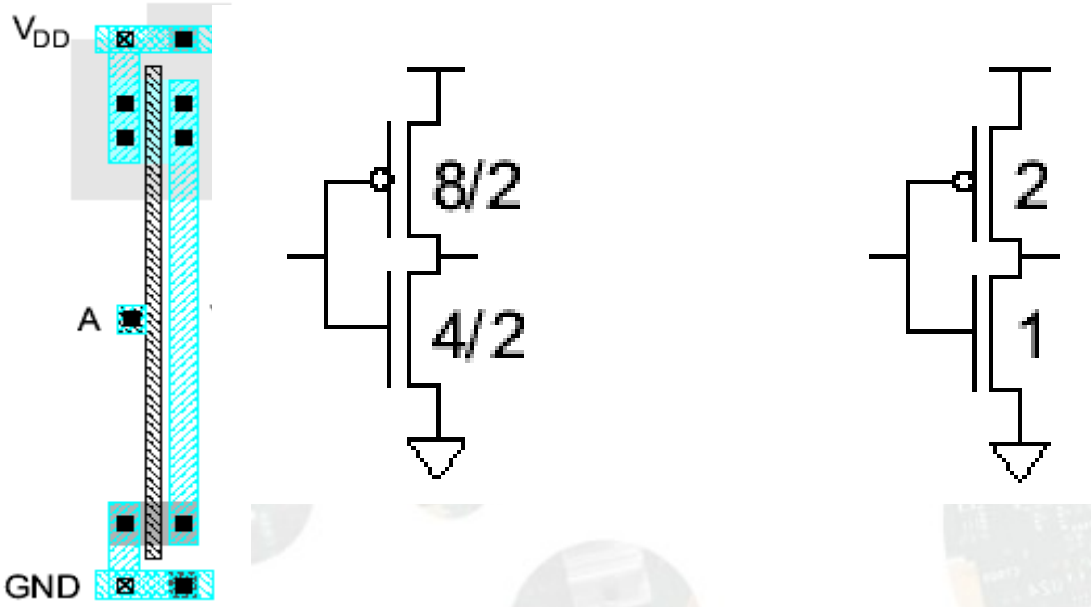
Inverter Layout

Transistor dimensions specified as Width / Length

Minimum size is 4-6λ/ 2λ, sometimes called 1 unit

In *f* = 0.25 μm process, this is 0.5-0.75 μm wide (W), 0.25 μm long (L)

Since $\lambda = f/2$, $\lambda = 0.125$ μm.



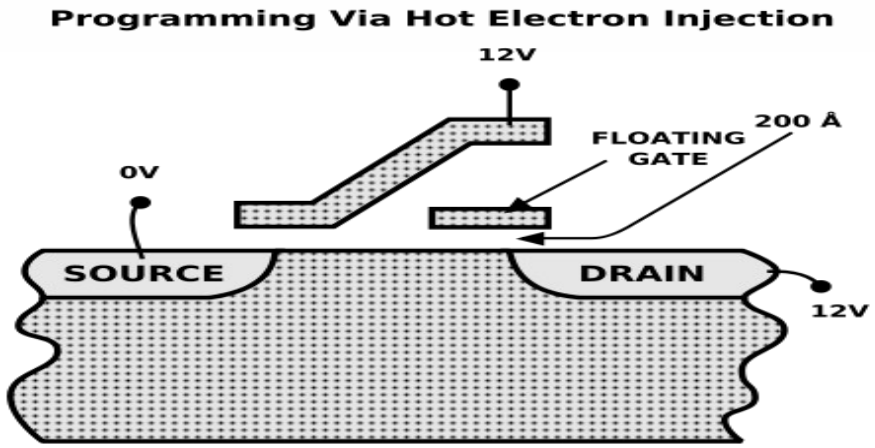
Summary

- MOS Transistors are stack of gate, oxide, silicon and p-n junctions
- Can be viewed as electrically controlled switches
- Build logic gates out of switches
- Draw masks to specify layout of transistors
- Now you know everything necessary to start designing schematics and layout for a simple chip!

Tasks

- Install OpenLane and OpenRoad
- Select a opensource digital system core
  - Processor
  - Convolution Filter
  - Computer Vision Based System on Chip
- Find out VLSI CAD tools for Graphene transistors
  - Understand Graphene based transistors
  - Develop a Graphene transistors Layout in CAD

A flash memory cell (single bit) resembles a standard MOSFET, except the transistor has two gates instead of one.



How many transistors would an N Tera Byte flash memory hold?

N is your Serial Number.