Spiral 1 / Unit 8

Transistor Implementations

CMOS Logic Gates

Spiral Content Mapping

Spiral	Theory	Combinational Design	Sequential Design	System Level Design	Implementation and Tools	Project
1	• Performance metrics (latency vs. throughput) • Boolean Algebra (Pt. 1) • Canonical Representations	• Decoders and muxes • Synthesis with min/maxterms • Synthesis with Karnaugh Maps	• Edge-triggered flip-flops • Registers (with enables)	• Encoded State machine design	• Structural Verilog HDL • CMOS gate implementation • Fabrication process	
2	• Boolean algebra for analysis and optimization (DeMorgan's theorem)	• Synthesis with memory • Adder and comparator design	• Bistables, latches, and Flip-flops • Counters • Memories	• One-hot state machine design • Control and datapath decomposition • Single-cycle CPU	• MOS Theory • Capacitance, delay and sizing • Memory constructs	
3	• Shannon's theorem	• Synthesis with muxes (Shannon's theorem)		• HW/SW partitioning • Bus interfacing	• Power and other logic families • EDA design process	

Outcomes

- I know the difference between combinational and sequential logic and can name examples of each.
- I understand latency, throughput, and at least 1 technique to improve throughput
- I can identify when I need state vs. a purely combinational function
  - I can convert a simple word problem to a logic function (TT or canonical form) or state diagram
- I can use Karnaugh maps to synthesize combinational functions with several outputs
- I can design a working state machine given a state diagram
- I can implement small logic functions with complex CMOS gates

DEMORGAN'S THEOREM

DeMorgan's Theorem

$$F = (\overline{X}+Y) + \overline{Z} \bullet (Y+W)$$

To find F', invert both sides of the equation and then use DeMorgan's theorem to simplify…

$$\overline{F} = \overline{(\overline{X}+Y) + \overline{Z} \bullet (Y+W)}$$

$$\overline{F} = \overline{(\overline{X}+Y)} \bullet \overline{(\overline{Z} \bullet (Y+W))}$$

$$\overline{F} = (\overline{\overline{X}} \bullet \overline{Y}) \bullet (\overline{\overline{Z}} + \overline{(Y+W)})$$

$$\overline{F} = (X \bullet \overline{Y}) \bullet (Z + (\overline{Y} \bullet \overline{W}))$$

Generalized DeMorgan's Theorem

$$F'(X_1,\ldots,X_n,+,\bullet) = F(X_1',\ldots,X_n',\bullet,+)$$

To find F', swap AND's and OR's and complement each literal. However, you must maintain the original order of operations.

Note: This parentheses doesn't matter (we are just OR'ing X', Y, and the following subexpression)

$$F = (\overline{X}+Y) + \overline{Z} \bullet (Y+W)$$

$$F = \overline{X}+Y + (\overline{Z} \bullet (Y+W))$$

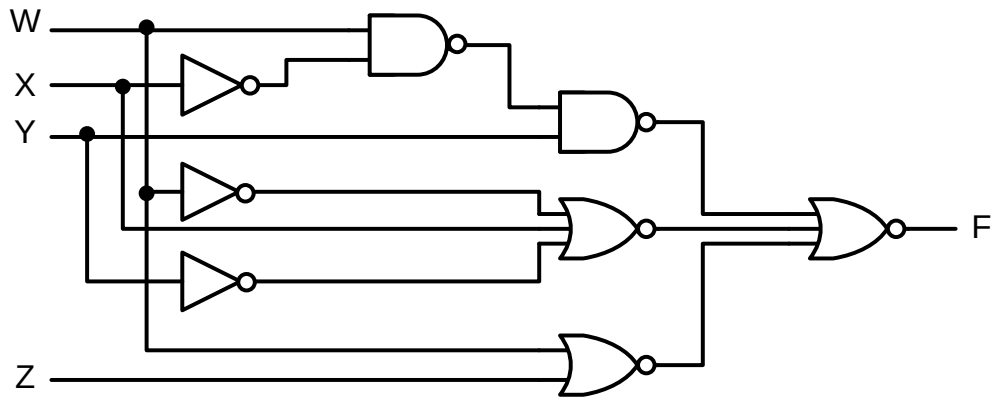
Fully parenthesized to show original order of ops.

$$\overline{F} = X \bullet \overline{Y} \bullet (Z + (\overline{Y} \bullet \overline{W}))$$

AND's & OR's swapped
Each literal is inverted

DeMorgan's Theorem Example

- Cancel as many bubbles as you can using DeMorgan's theorem.

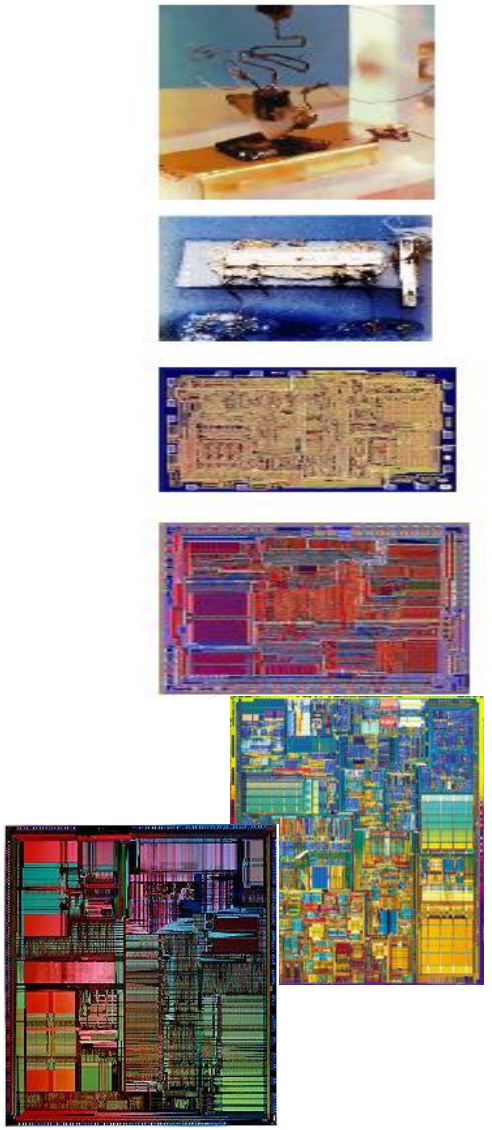


With focus on MOS Transistors

SEMICONDUCTOR TECHNOLOGY

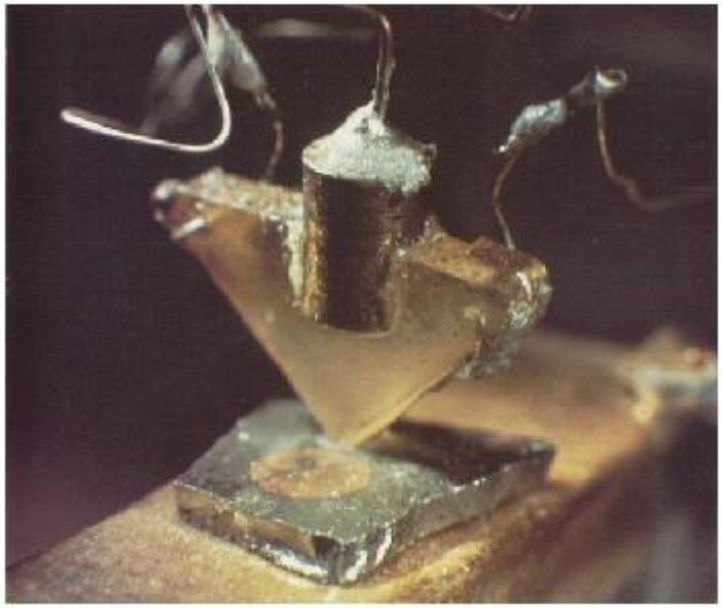
Evolution of transistor in ICs

- BJT invention, Bell Labs, 1947
- Single transistor, TI, 1958
- CMOS gate, Fairchild, 1963
  - First processor, Intel, 1970
- Very Large Scale Integration, 1978
  - Up to 20k transistor
- Ultra Large Scale Integration, 1989
  - More than 1 million per chip
- System-on-Chip, 2002-2015
  - Millions to several billion transistors



Invention of the Transistor

- Vacuum tubes ruled in first half of 20th century
  Large, expensive, power-hungry, unreliable
- 1947: first point contact transistor
  - John Bardeen and Walter Brattain at Bell Labs
  - See *Crystal Fire*
    by Riordan, Hoddeson

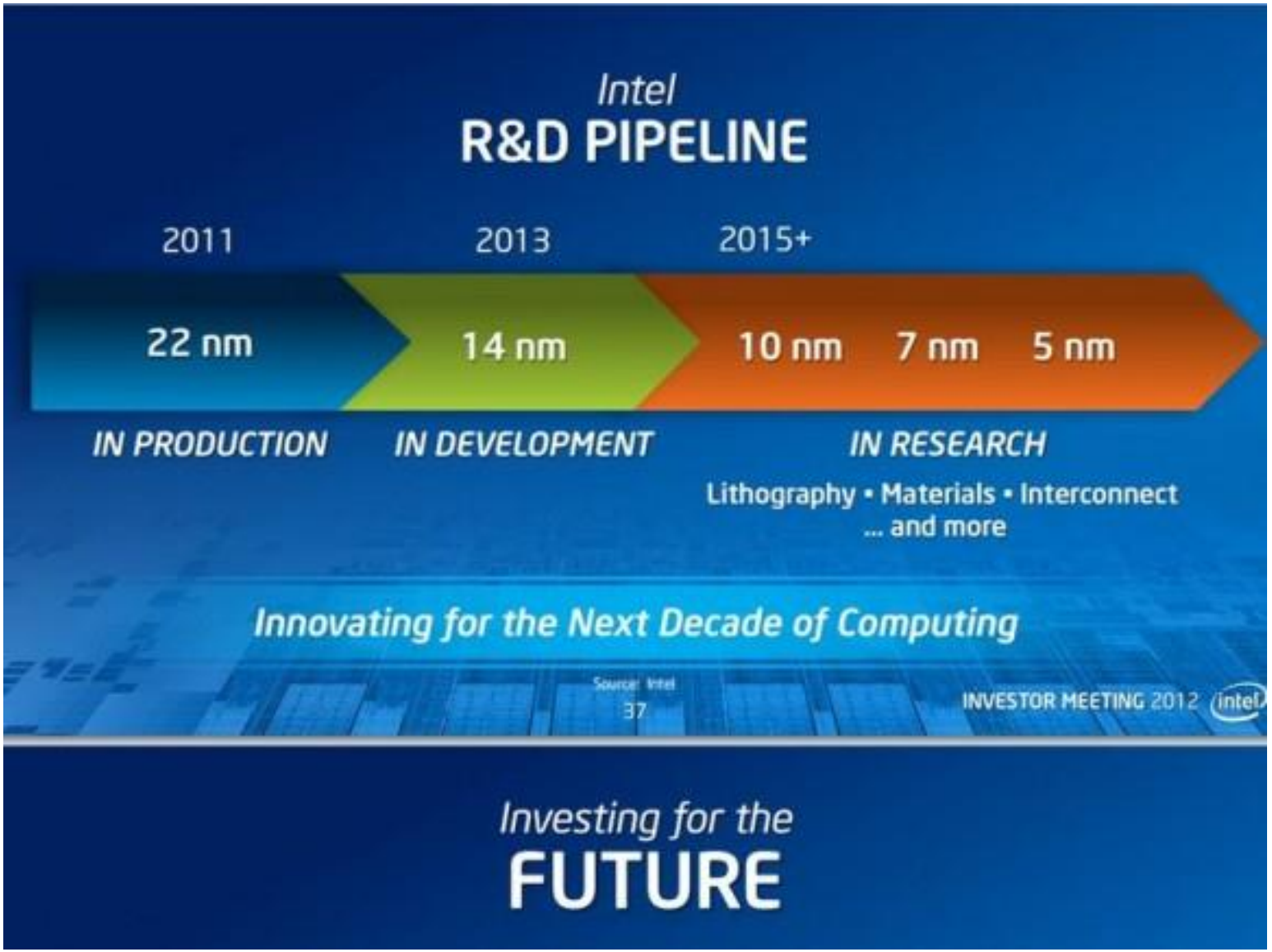


Growth Rate

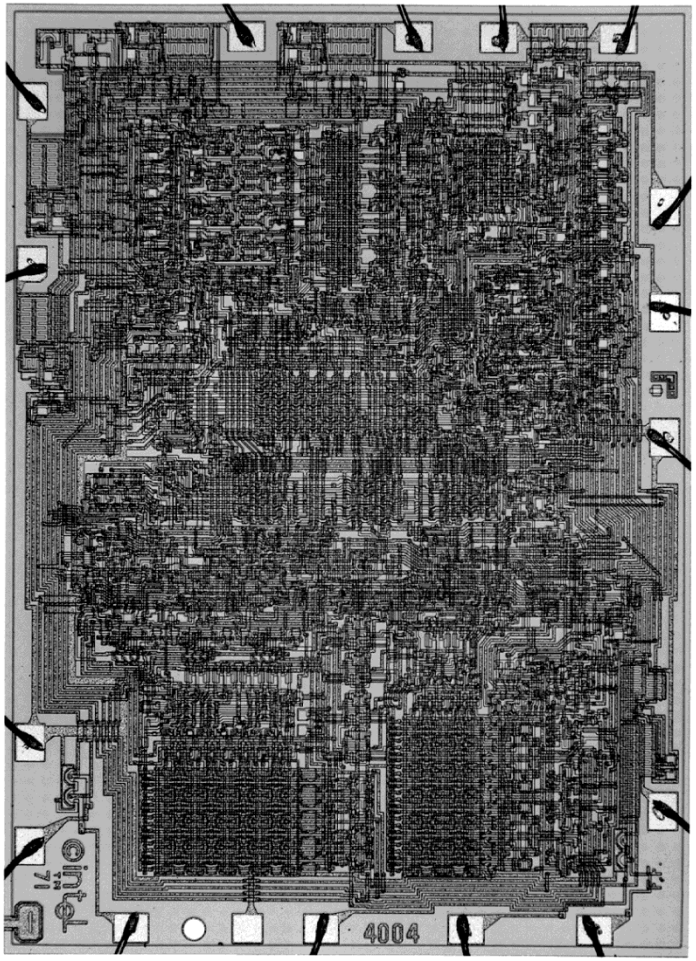
- 53% compound annual growth rate over 50 years
  - No other technology has grown so fast so long
- Driven by miniaturization of transistors
  - Smaller is cheaper, faster, lower in power!
  - Revolutionary effects on society

[Moore65]
Electronics Magazine

Minimum Feature Size



Intel 4004 Micro-Processor



1971
1000 transistors
1 MHz operation

Intel Core I7

2nd Gen. Intel Core i7 Extreme Processor for desktops launched in Q4 of 2012

- #cores/#threads: 6/12
- Technology node: 32nm
- Clock speed: 3.5 GHz
- Transistor count: Over one billion
- Cache: 15MB
- Addressable memory: 64GB
- Size: 52.5mm by 45.0mm mm^2

ARM Cortex A15

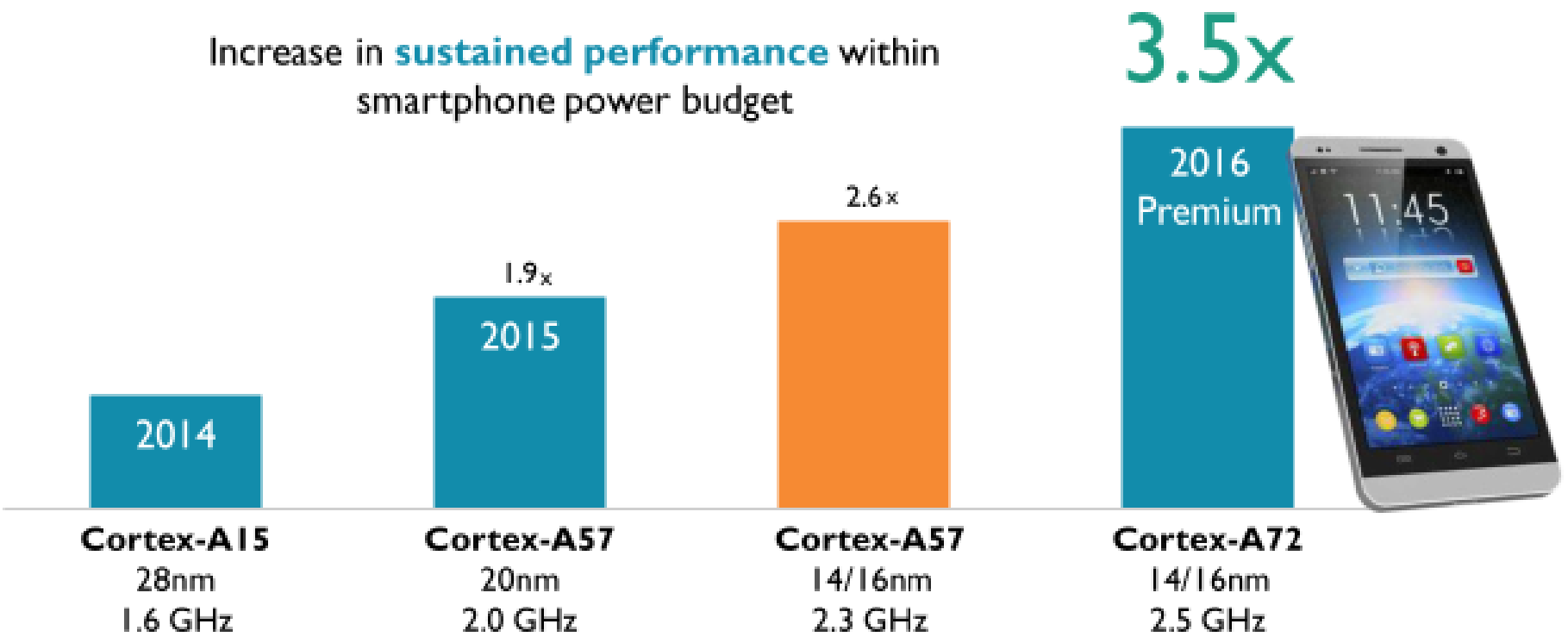
ARM Cortex A15 in 2011 to 2013

- 4 cores per cluster, two clusters per chip
- Technology node: 22nm
- Clock speed: 2.5 GHz
- Transistor count: Over one billion
- Cache: Up to 4MB per cluster
- Addressable memory: up to 1TB
- Size: 52.5mm by 45.0mm

Cortex-A72

Cortex-A72: Accelerating Usable Performance

Increase in **sustained performance** within smartphone power budget

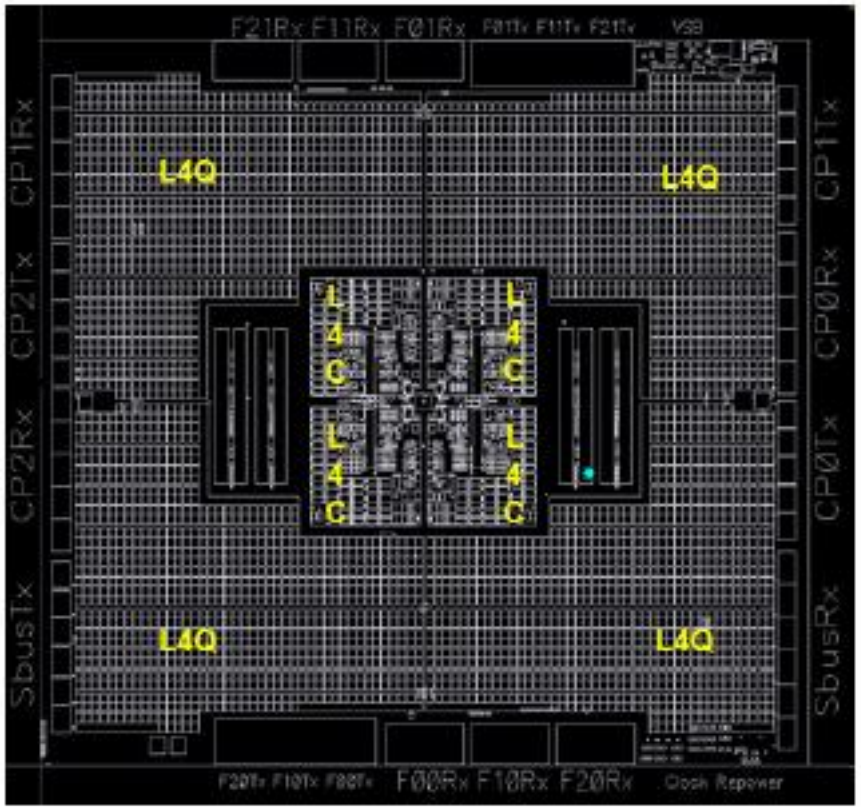


ARM

IBM z13 Storage Controller

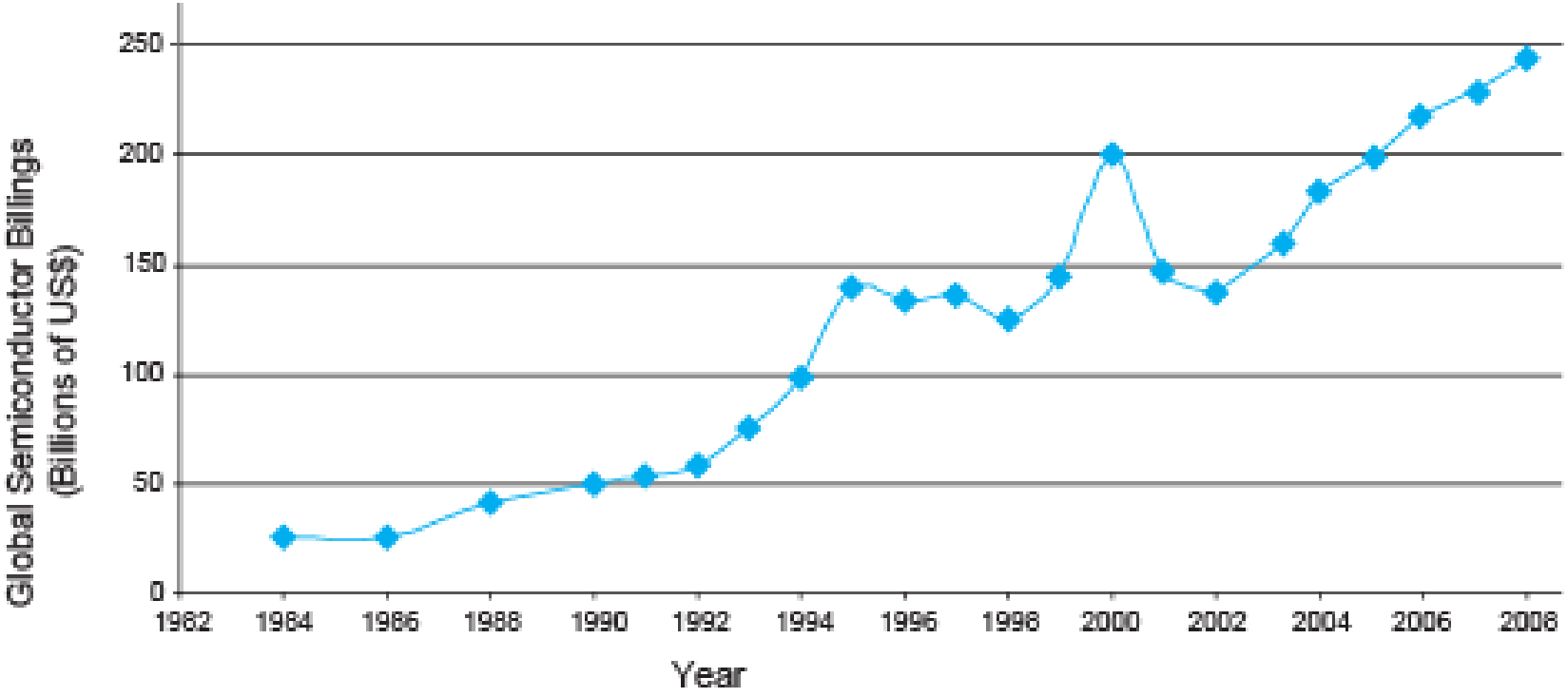
z13 Storage Control (SC) Chip Detail

- **CMOS 14S0 22nm SOI Technology**
  - 15 Layers of metal
  - 7.1 Billion transistors
  - 12.4 Miles of copper wire
- **Chip Area –**
  - 28.4 x 23.9 mm
  - 678 mm²
  - 11,950 power pins
  - 1,707 Signal Connectors
- **eDRAM Shared L4 Cache**
  - 480 MB per SC chip (Non-inclusive)
  - 224 MB L3 NIC Directory
  - 2 SCs = 960 MB L4 per z13 drawer
- **Interconnects (L4 – L4)**
  - 3 to CPs in node
  - 1 to SC (node – node) in drawer
  - 3 to SC nodes in remote drawers
- **6 Clock domains**

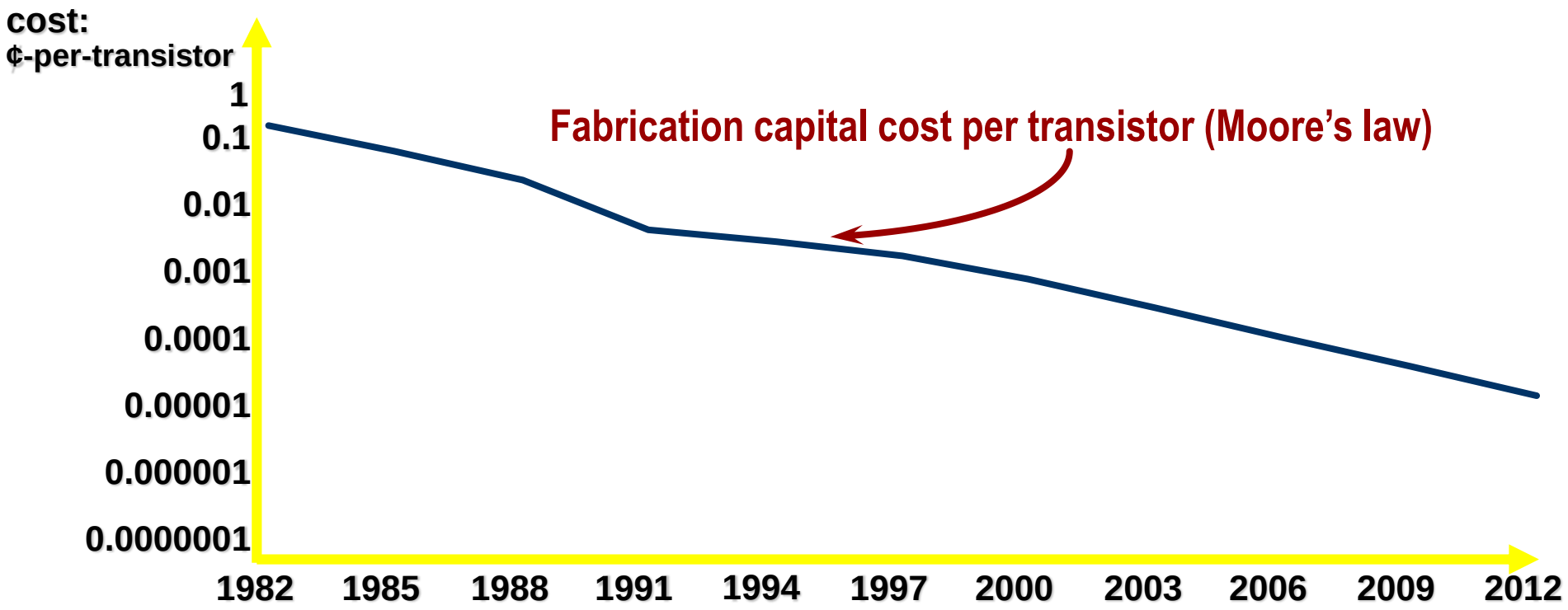


Annual Sales

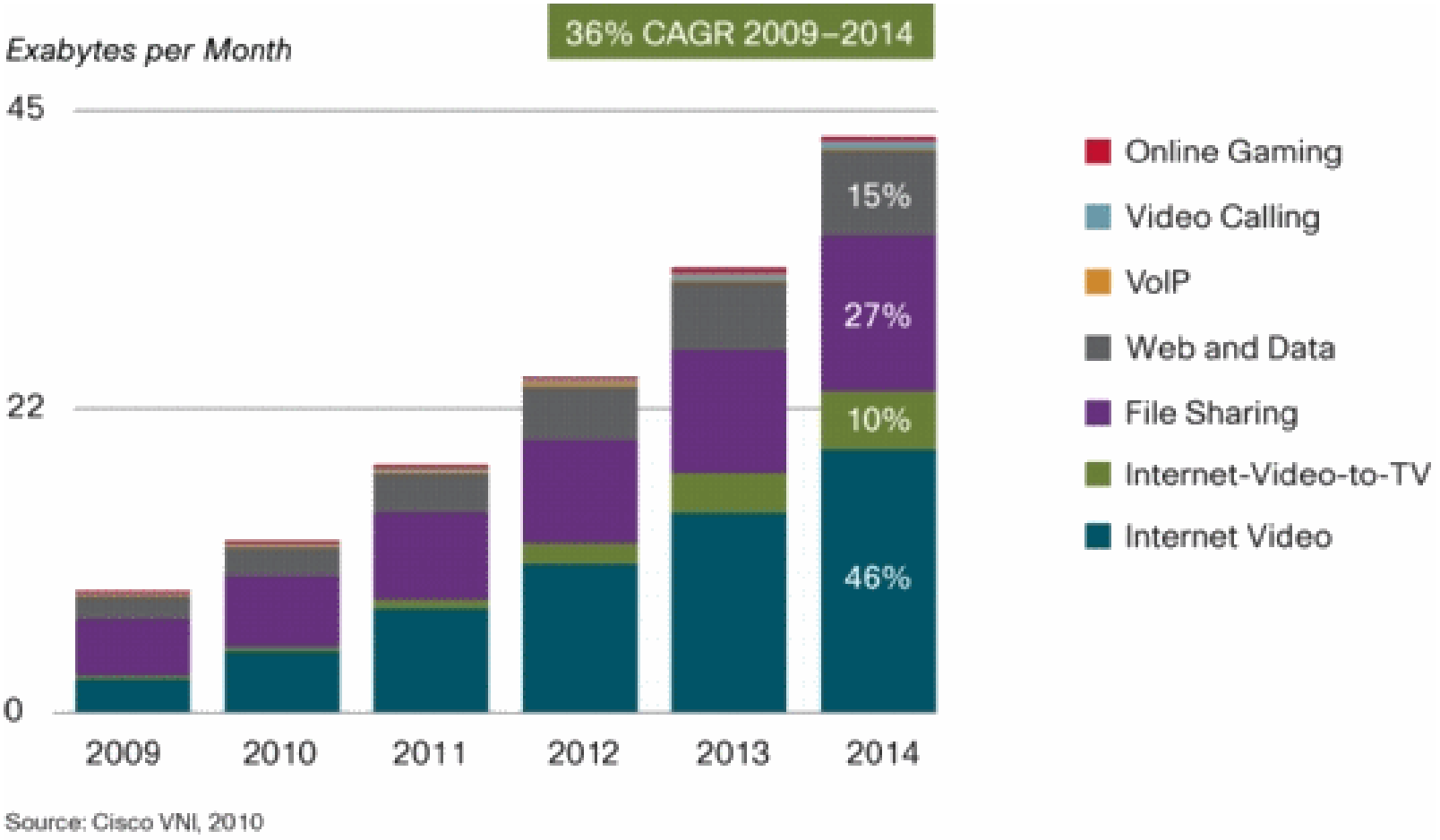
- >10^{19} transistors manufactured in 2008
  - 1 billion for every human on the planet



Cost per Transistor



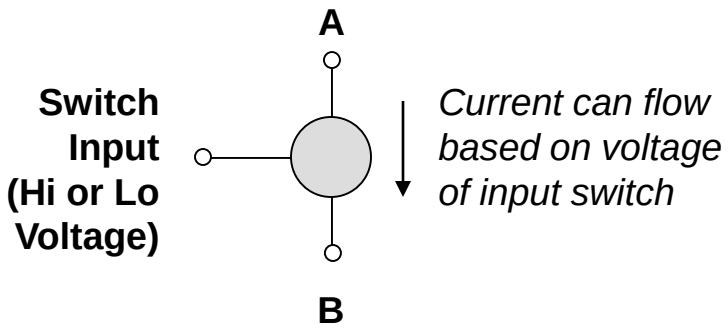
Internet Traffic Growth



TRANSISTOR BASICS

Transistors As Switches

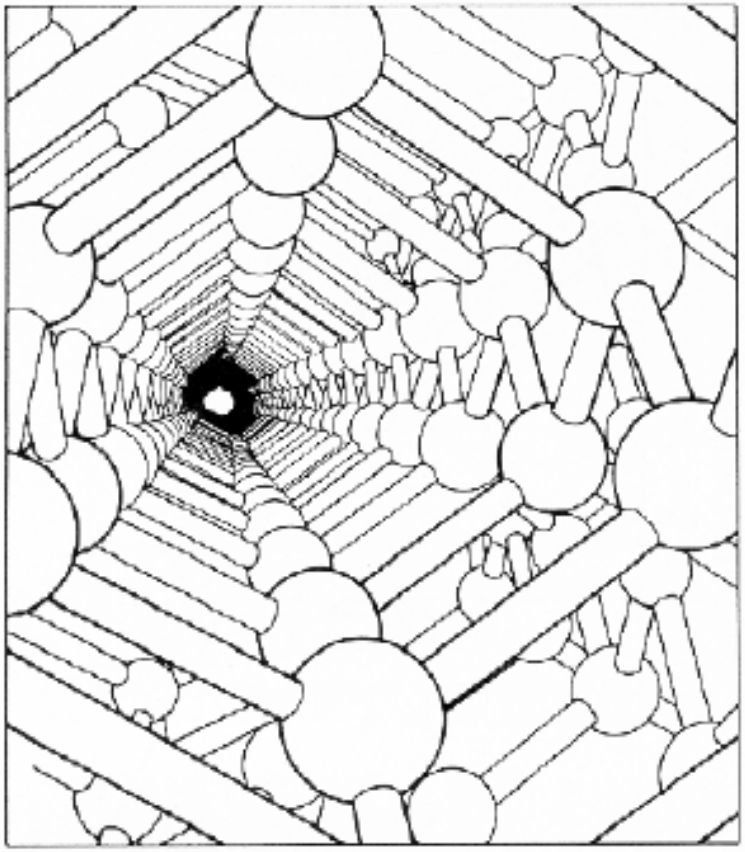
- Transistor act as a form of switch (on / off)
- Different physical structures lead to different kinds of transistors
  - Bipolar Junction Transistor (BJT)
    - Initial technology back in the late 40's – 60's
  - Metal-Oxide-Semiconductor Field Effect Transistor (MOSFET)
    - Dominates the digital IC market today
- All transistors essentially function similarly with 3 nodes/terminals:
  - 1 node serves as the switch value allowing current to flow between the other 2 nodes (on) or preventing current flow between the other 2 nodes (off)
  - Example: if the switch input voltage is 5V, then current is allowed to flow between the other nodes



Semiconductors

While there are numerous semiconductor materials available, by far the most popular material is **Silicon**.

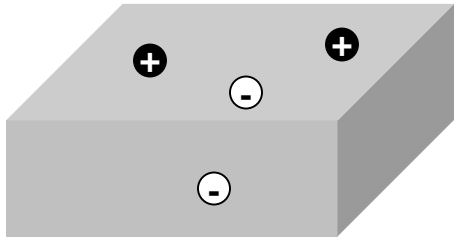
GaAs, InP and SiGe are compound semiconductors that are used in specialized devices.



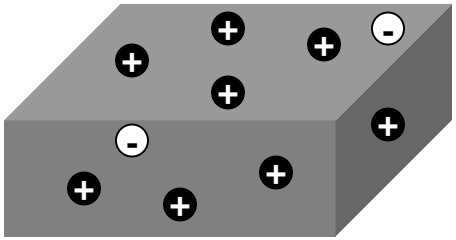
The success of a semiconductor material depends on how easy it is to process and how well it allows reliable high-volume fabrication.

Semiconductor Material

- Semiconductor material is not a great conductor material in its pure form
  - Small amount of free charge
- Can be implanted ("doped") with other elements (e.g. boron or arsenic) to be more conductive
  - Increases the amount of free charge

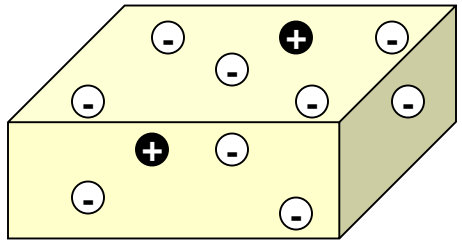


Pure Silicon



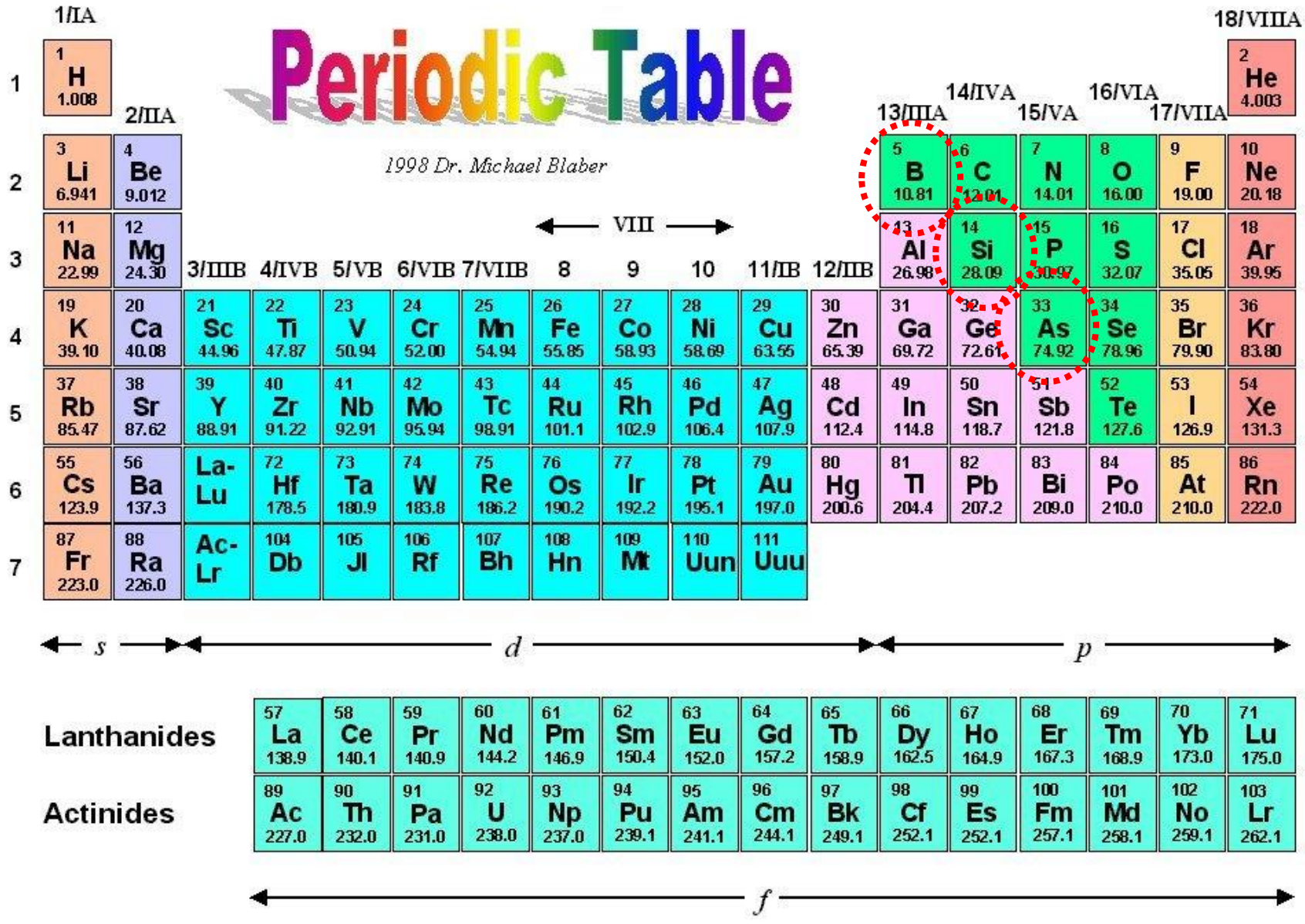
P-Type Silicon
(Doped with boron)

Electron acceptors

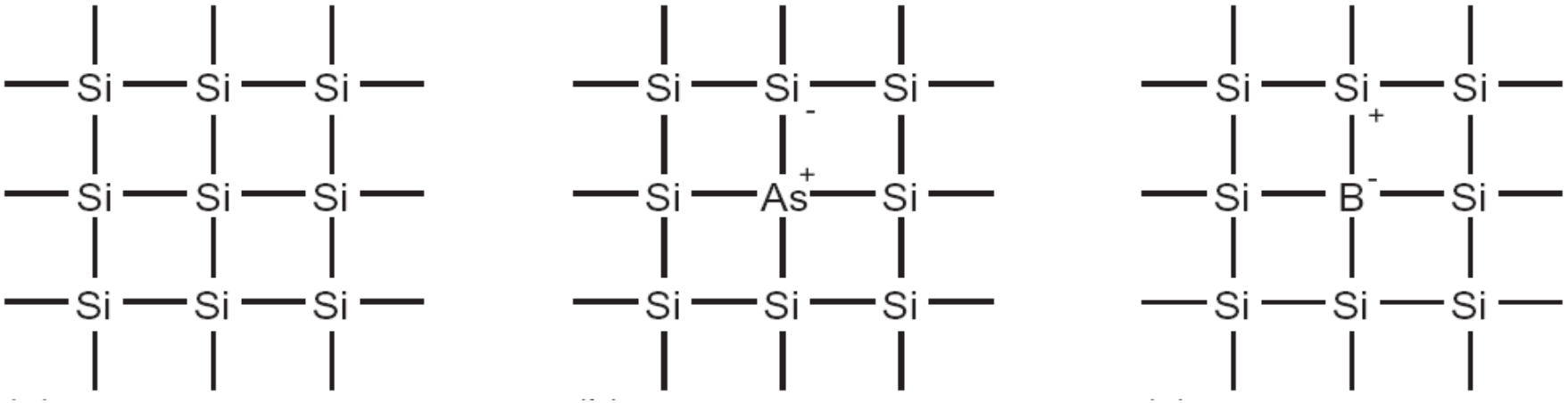


N-Type Silicon
(Doped with arsenic)

Electron donors


Periodic Table
1998 Dr. Michael Blaber
1/IA
2/IIA
3/IIIB
4/IVB
5/VB
6/VIB
7/VIIB
VIII
8
9
10
11/IB
12/IIB
13/IIIA
14/IVA
15/VA
16/VIA
17/VIIA
18/VIIIA
1 H 1.008
2 He 4.003
3 Li 6.941
4 Be 9.012
5 B 10.81
6 C 12.01
7 N 14.01
8 O 16.00
9 F 19.00
10 Ne 20.18
11 Na 22.99
12 Mg 24.30
13 Al 26.98
14 Si 28.09
15 P 30.97
16 S 32.07
17 Cl 35.05
18 Ar 39.95
19 K 39.10
20 Ca 40.08
21 Sc 44.96
22 Ti 47.87
23 V 50.94
24 Cr 52.00
25 Mn 54.94
26 Fe 55.85
27 Co 58.93
28 Ni 58.69
29 Cu 63.55
30 Zn 65.39
31 Ga 69.72
32 Ge 72.61
33 As 74.92
34 Se 78.96
35 Br 79.90
36 Kr 83.80
37 Rb 85.47
38 Sr 87.62
39 Y 88.91
40 Zr 91.22
41 Nb 92.91
42 Mo 95.94
43 Tc 98.91
44 Ru 101.1
45 Rh 102.9
46 Pd 106.4
47 Ag 107.9
48 Cd 112.4
49 In 114.8
50 Sn 118.7
51 Sb 121.8
52 Te 127.6
53 I 126.9
54 Xe 131.3
55 Cs 123.9
56 Ba 137.3
La-Lu
72 Hf 178.5
73 Ta 180.9
74 W 183.8
75 Re 186.2
76 Os 190.2
77 Ir 192.2
78 Pt 195.1
79 Au 197.0
80 Hg 200.6
81 Tl 204.4
82 Pb 207.2
83 Bi 209.0
84 Po 210.0
85 At 210.0
86 Rn 222.0
87 Fr 223.0
88 Ra 226.0
Ac-Lr
104 Db
105 Jl
106 Rf
107 Bh
108 Hn
109 Mt
110 Uun
111 Uuu
s
d
p
Lanthanides
57 La 138.9
58 Ce 140.1
59 Pr 140.9
60 Nd 144.2
61 Pm 146.9
62 Sm 150.4
63 Eu 152.0
64 Gd 157.2
65 Tb 158.9
66 Dy 162.5
67 Ho 164.9
68 Er 167.3
69 Tm 168.9
70 Yb 173.0
71 Lu 175.0
Actinides
89 Ac 227.0
90 Th 232.0
91 Pa 231.0
92 U 238.0
93 Np 237.0
94 Pu 239.1
95 Am 241.1
96 Cm 244.1
97 Bk 249.1
98 Cf 252.1
99 Es 252.1
100 Fm 257.1
101 Md 258.1
102 No 259.1
103 Lr 262.1
f

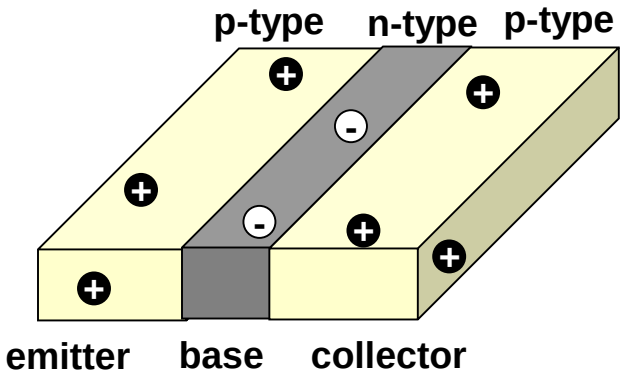
Silicon Lattice and Dopant Atoms



- Pure silicon: 3-D lattice of atoms (a cubic crystal) and a poor conductor
- Conductivity can be raised by adding either donors or acceptor
  - Donors: Group V dopant impurities, which have more free electrons than silicon
    - The resulting material is called n-type
  - Group III dopants impurities which have lack of electrons
    - The resulting material is called p-type

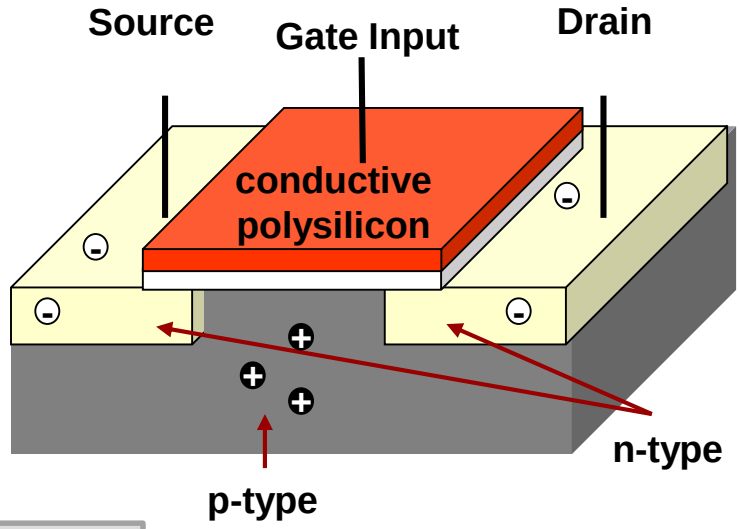
Transistor Types

- Bipolar Junction Transistors (BJT)
  - npn or pnp silicon structure
  - Small current into very thin base layer controls large currents between emitter and collector
  - However the fact that it requires a current into the base means it burns power (P = I*V) and thus limits how many we can integrate on a chip (i.e. density)



npn BJT

- Metal Oxide Semiconductor Field Effect Transistors
  - nMOS and pMOS MOSFETS
  - Voltage applied to insulated gate controls current between source and drain
    - Gate input requires no constant current…thus low power!

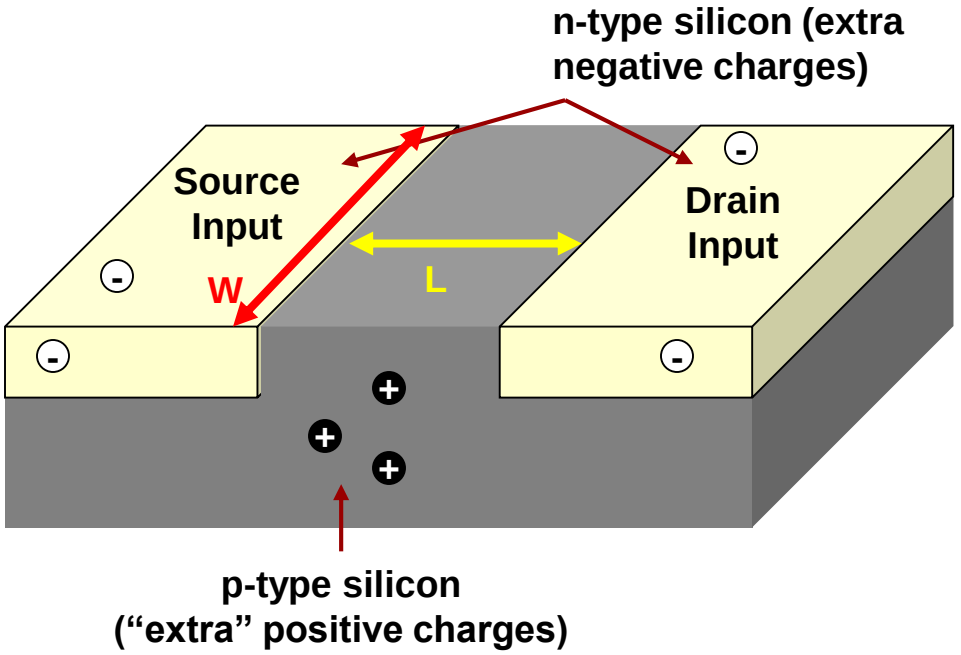


N-type MOSFET

We will focus on MOSFET in this class

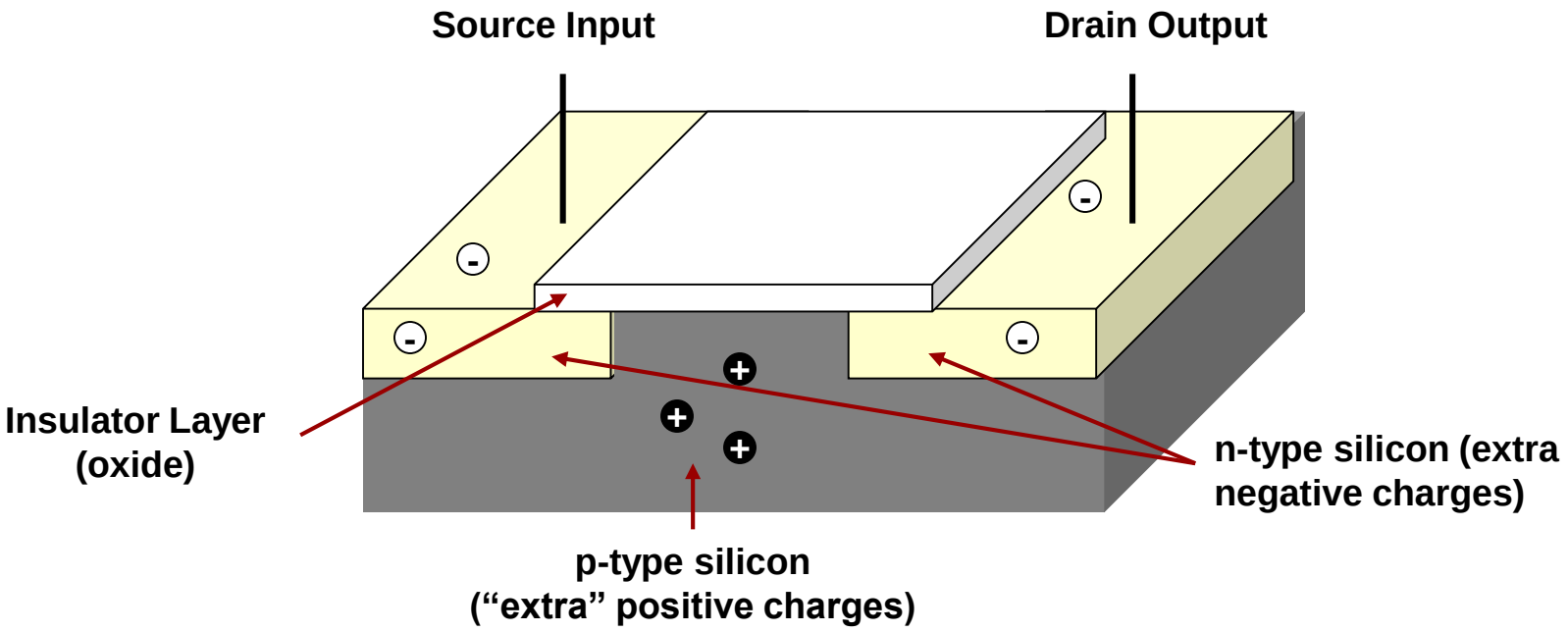
NMOS Transistor Physics

- Transistor is started by implanting two n-type silicon areas, separated by p-type



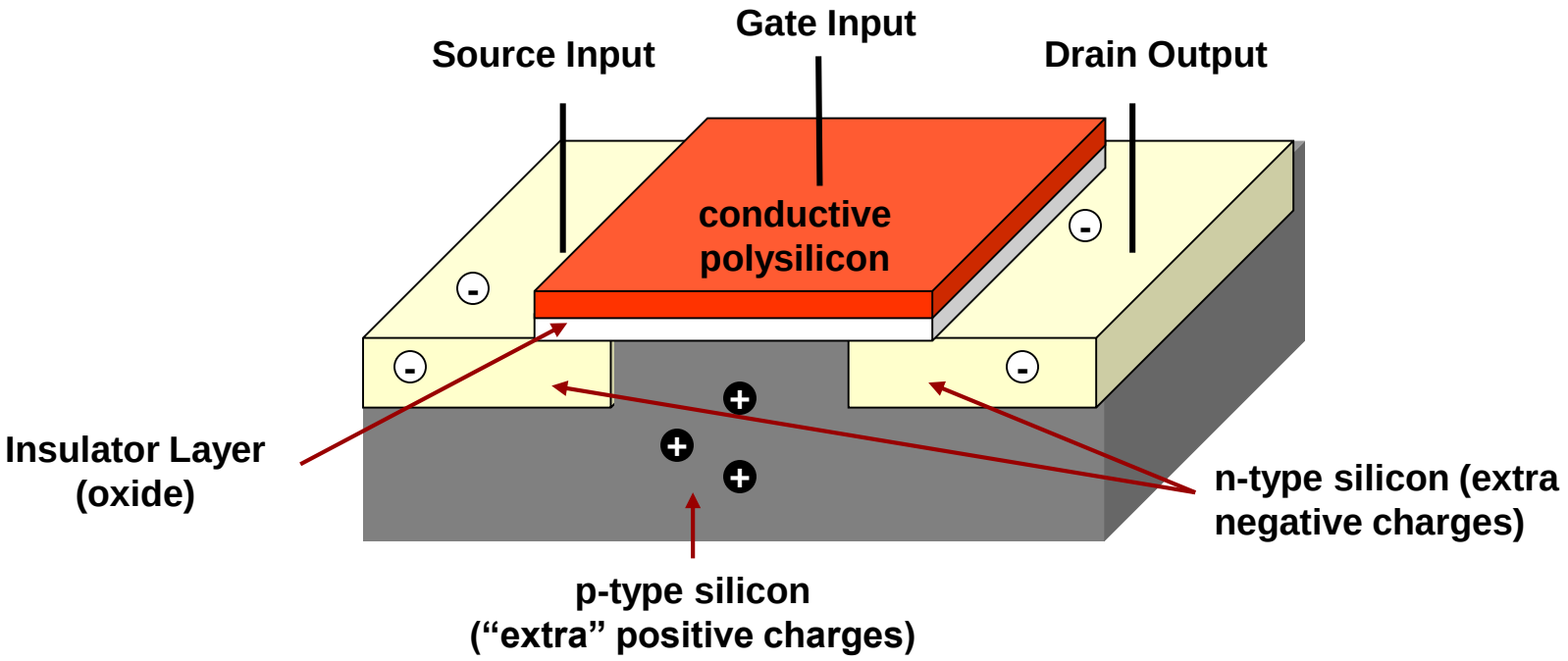
NMOS Transistor Physics

- A thin, insulator layer (silicon dioxide or just “oxide”) is placed over the silicon between source and drain



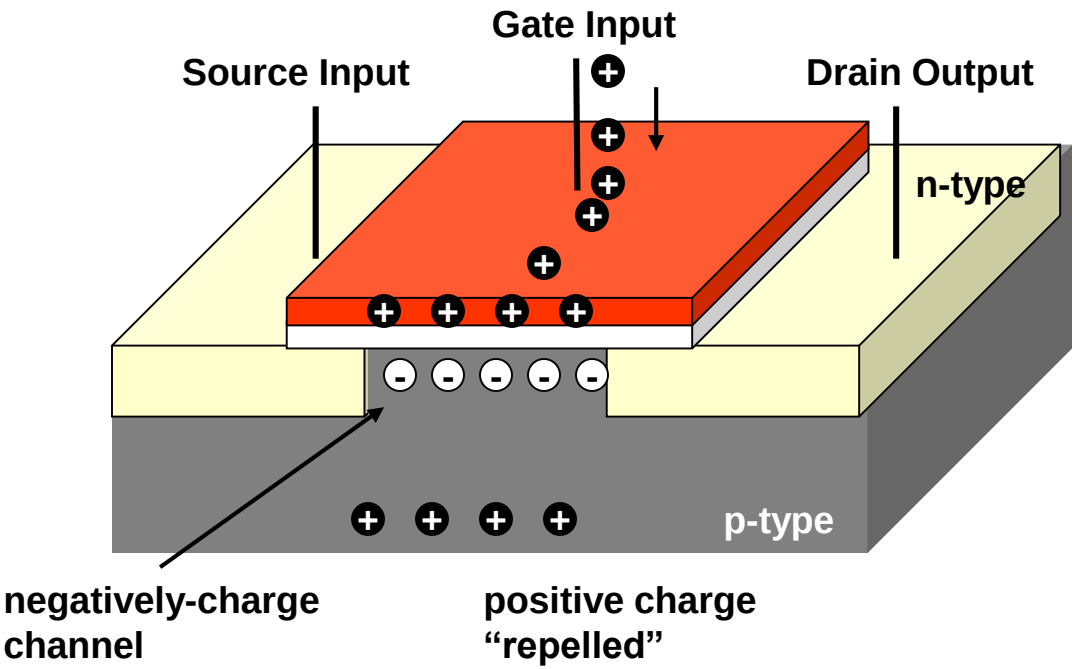
NMOS Transistor Physics

- A thin, insulator layer (silicon dioxide or just “oxide”) is placed over the silicon between source and drain
- Conductive polysilicon material is layered over the oxide to form the gate input



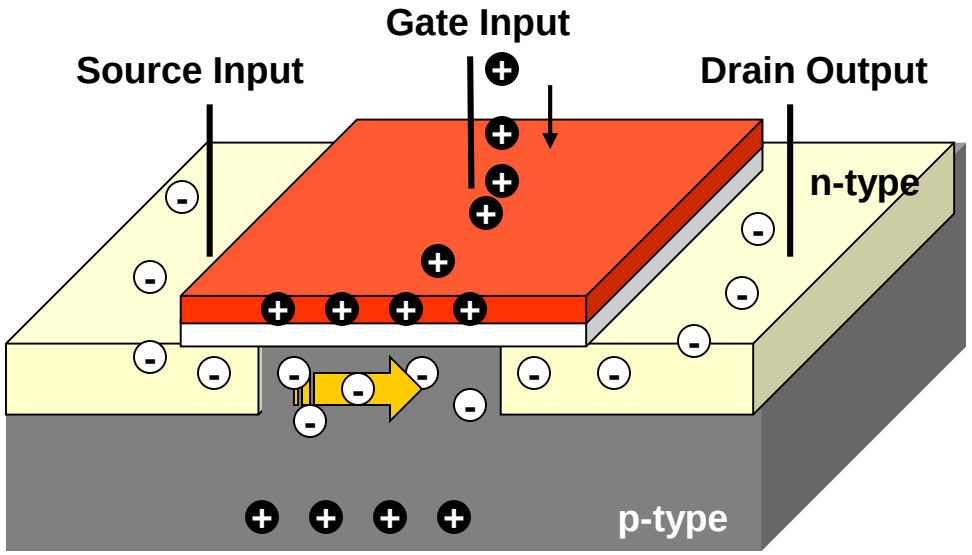
NMOS Transistor Physics

- Positive voltage (charge) at the gate input *repels* the extra positive charges in the p-type silicon
- Result is a negative-charge channel between the source input and drain



NMOS Transistor Physics

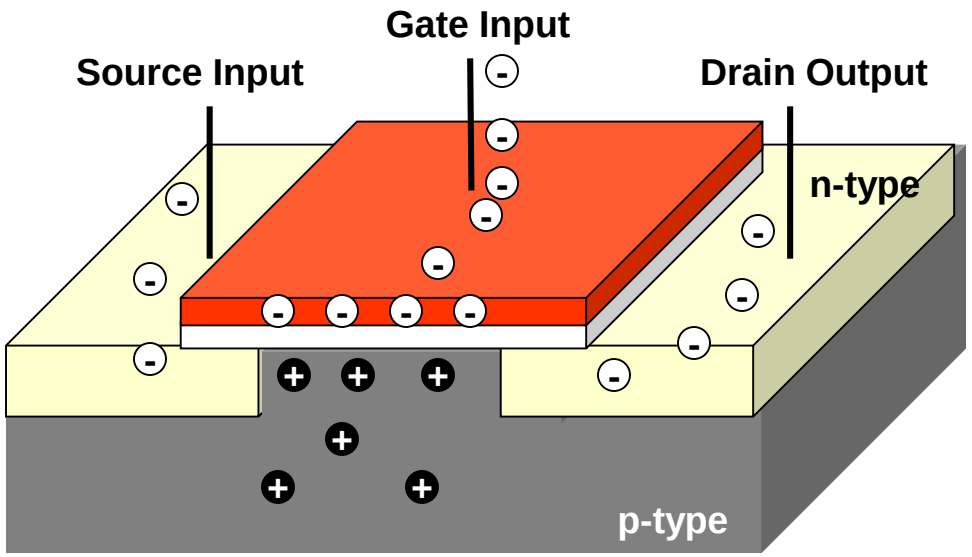
- Electrons can flow through the negative channel from the source input to the drain output
- The transistor is "on"



Negative channel between source and drain = Current flow

NMOS Transistor Physics

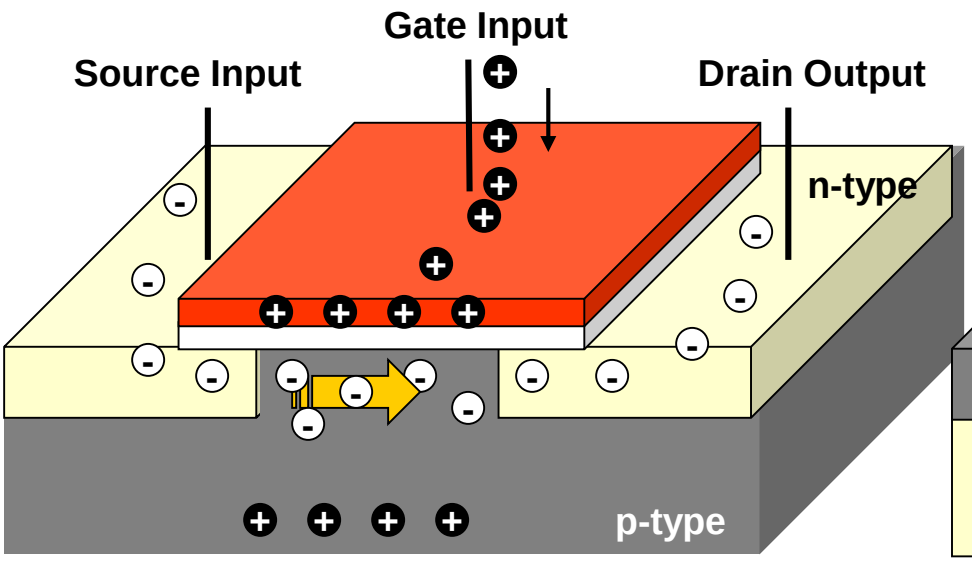
- If a low voltage (negative charge) is placed on the gate, no channel will develop and no current will flow
- The transistor is "off"



No negative channel between source and drain = No current flow

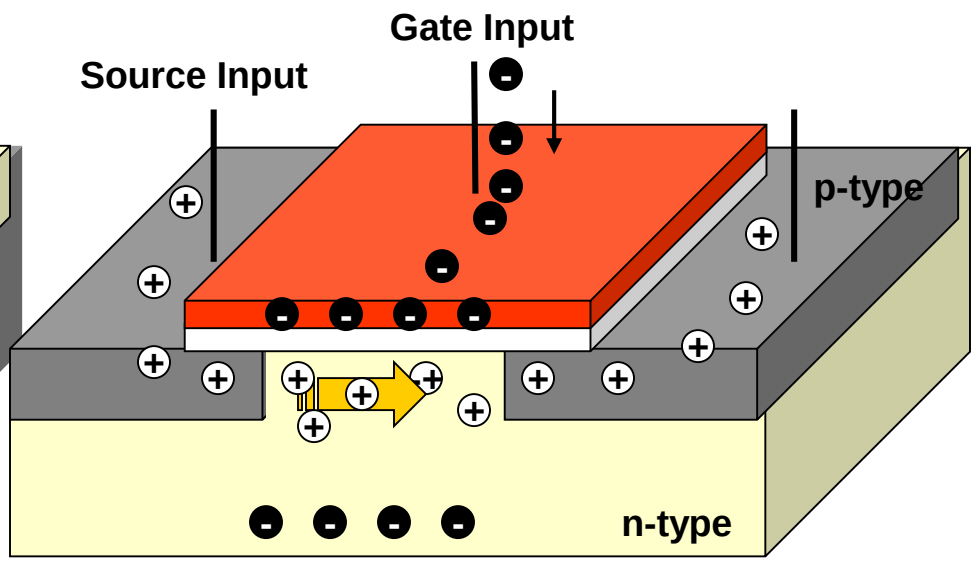
PMOS vs. NMOS

- PMOS transistors can also be made that are on when the gate voltage is low and off when it is high



Negative channel between source and drain = Current flow

NMOS



"Positive" channel between source and drain = Current flow

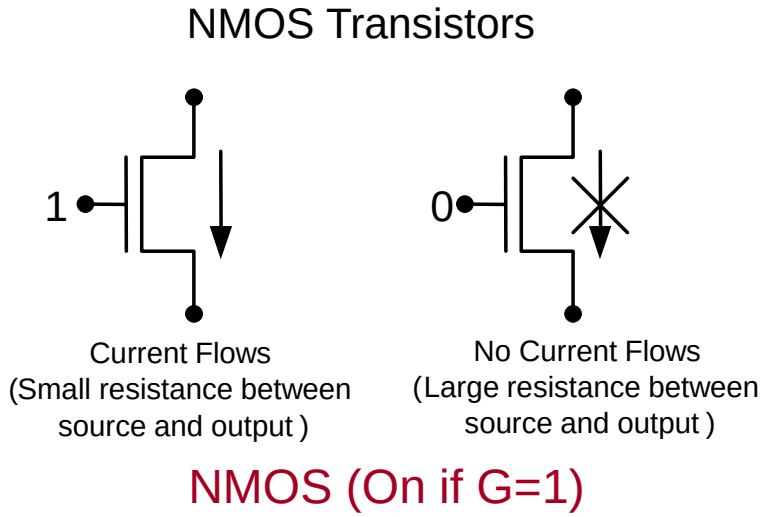
PMOS

Understanding physical constraints

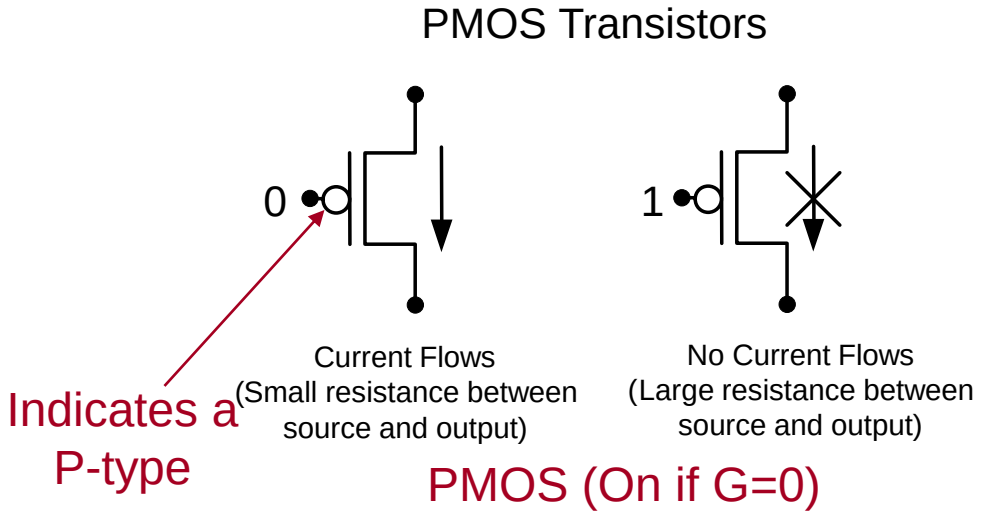
CMOS TRANSISTOR LEVEL IMPLEMENTATION

NMOS and PMOS Transistors

- NMOS conducts when gate input is at a high voltage (logic '1')

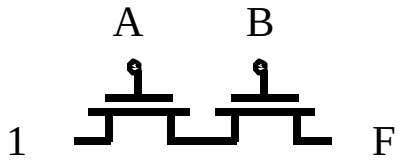


- PMOS conducts when gate input is at a low voltage (logic '0')

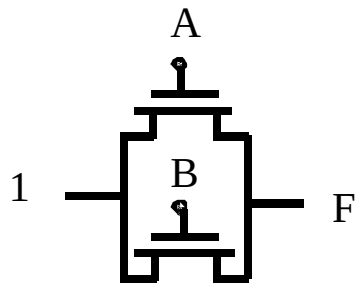


NMOS Transistors in Series/Parallel Connection

- Transistors can be thought as a switch controlled by its gate signal
- NMOS switch closes when switch control input is high



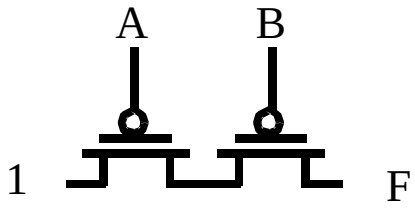
F = 1 if A and B



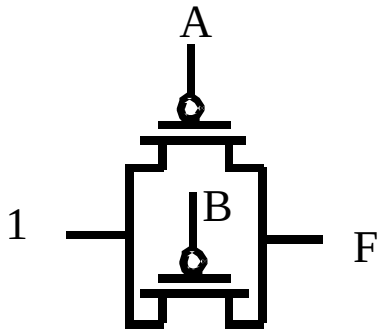
F = 1 if A OR B

PMOS Transistors in Series/Parallel Connection

- PMOS switch closes when switch control input is low



$F = 1$ if $\overline{A}$ AND $\overline{B} = \overline{A + B}$

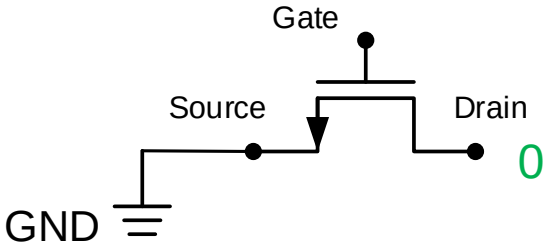


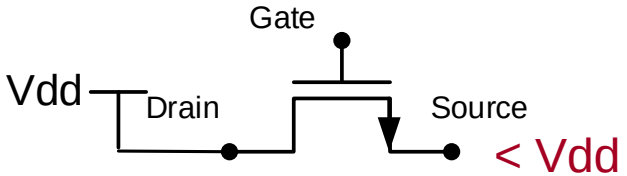
$F = 1$ if $\overline{A}$ OR $\overline{B} = \overline{AB}$

We All Have Our Strengths

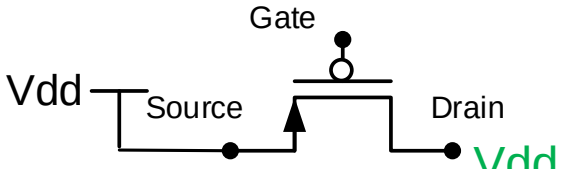
- NMOS are:
  - Good at pulling the output voltage DOWN to 0
  - Bad at pulling the output voltage up to 1
- PMOS are:
  - Good at pulling the output voltage up to 1
  - Bad at pulling the output voltage down to 0

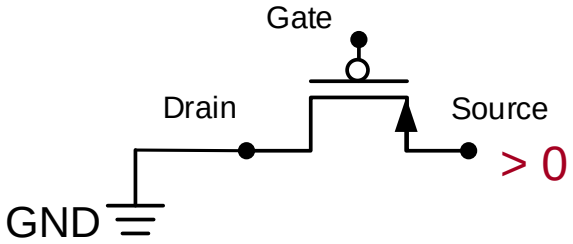
NMOS





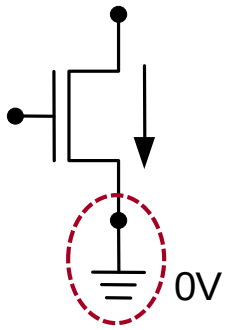
PMOS



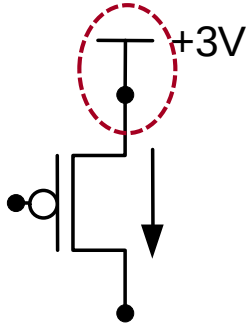


NMOS and PMOS Transistors

- NMOS transistors work best when one terminal is connected to a low voltage source, pulling the other terminal down to that voltage
  - Normally, source terminal is connected to GND=0V
- PMOS transistors work best when one terminal is connected to a high voltage source, pulling the other terminal down to that voltage
  - Normally, source terminal is connected to power supply voltage (+5V, +3V, etc.)



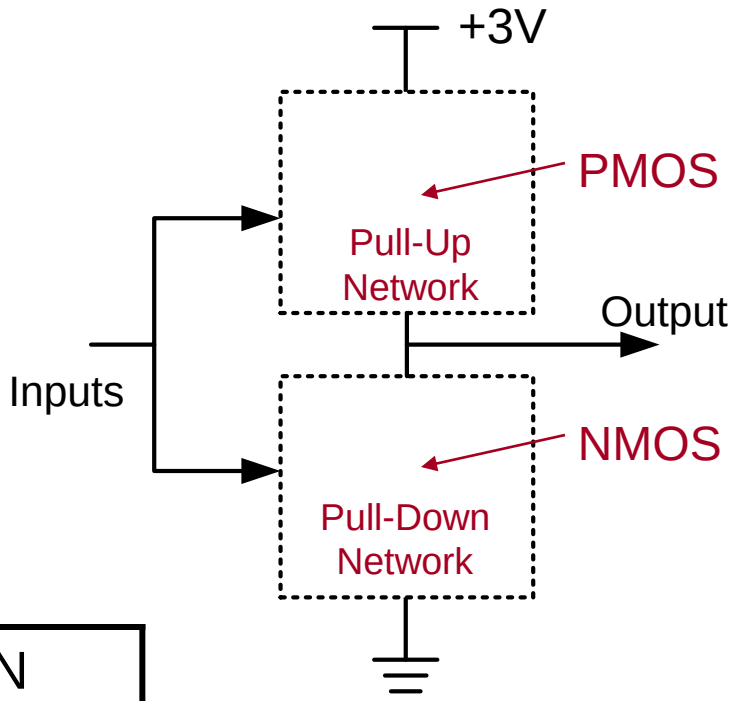
NMOS



PMOS

CMOS

- Complimentary MOS (CMOS)
  - Use PMOS to connect output to high voltage source
    - We call this the Pull-Up Network
  - Use NMOS to connect output to low voltage source (usually = GND)
    - We call this the Pull-Down Network
  - Either PMOS or NMOS should create a conductive path to output, but not both



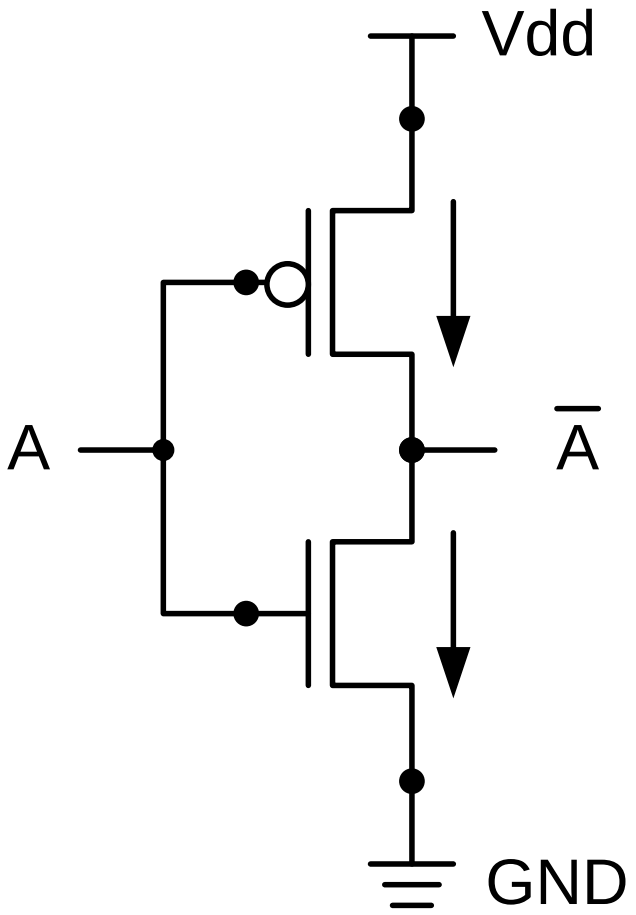
	Pull-up OFF	Pull-up ON
Pull-down OFF	Z (float)	1
Pull-down ON	0	X (crowbar)

Signal Strength

- *Strength* of signal
  - How close it approximates ideal voltage source
- V_{DD} and GND rails are strongest 1 and 0
- nMOS passes strong 0
  - But degraded or weak 1
- pMOS passes strong 1
  - But degraded or weak 0
- Thus nMOSes are best for the pull-down network, pMOSes are best for the pull-up network

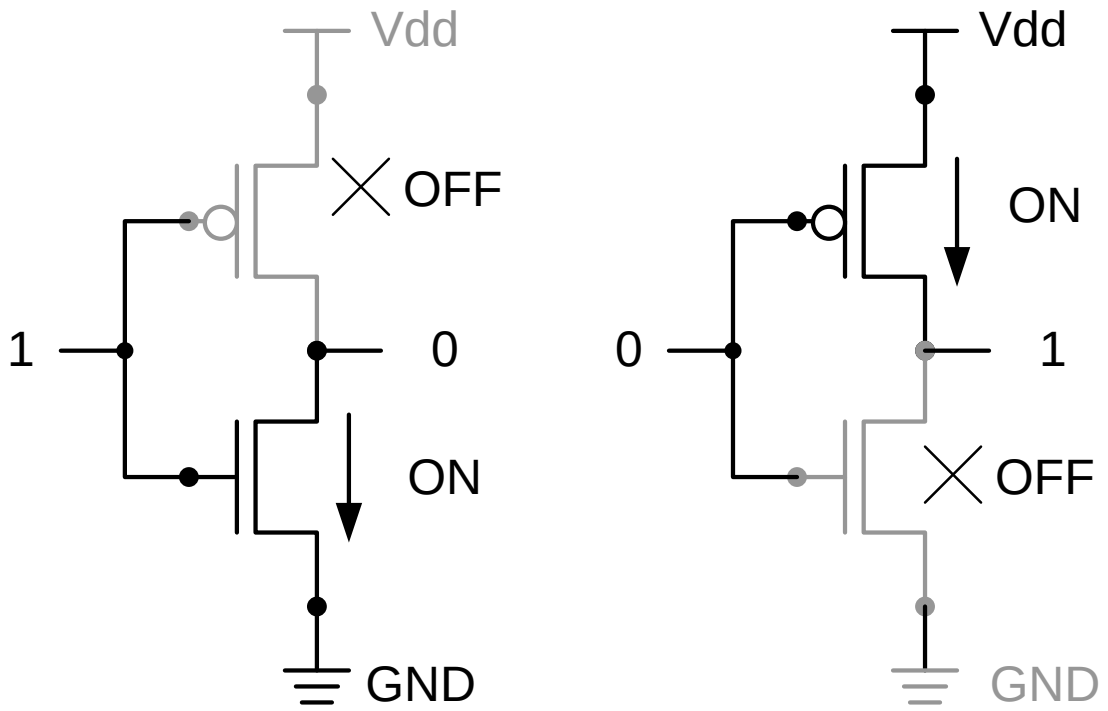
CMOS Inverter

- Inverter can be formed using one PMOS and NMOS transistor
- The input value connects to both gate inputs
- The output is formed at the junction of the drains



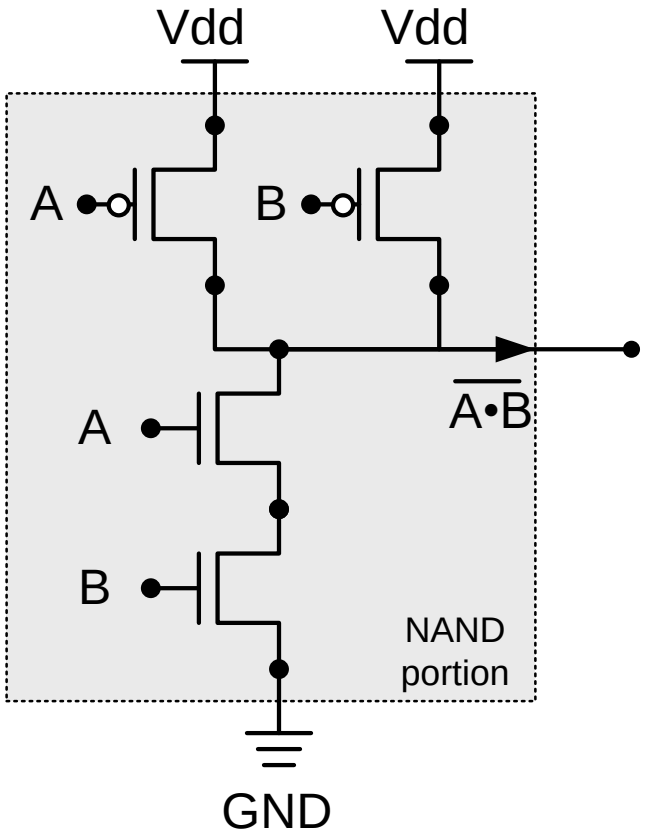
CMOS Inverter

- When input is 1, NMOS conducts and output is pulled down to 0V (GND)
- When input is 0, PMOS conducts and output is pulled up to 3V (V_{DD})



CMOS 'NAND' Gate

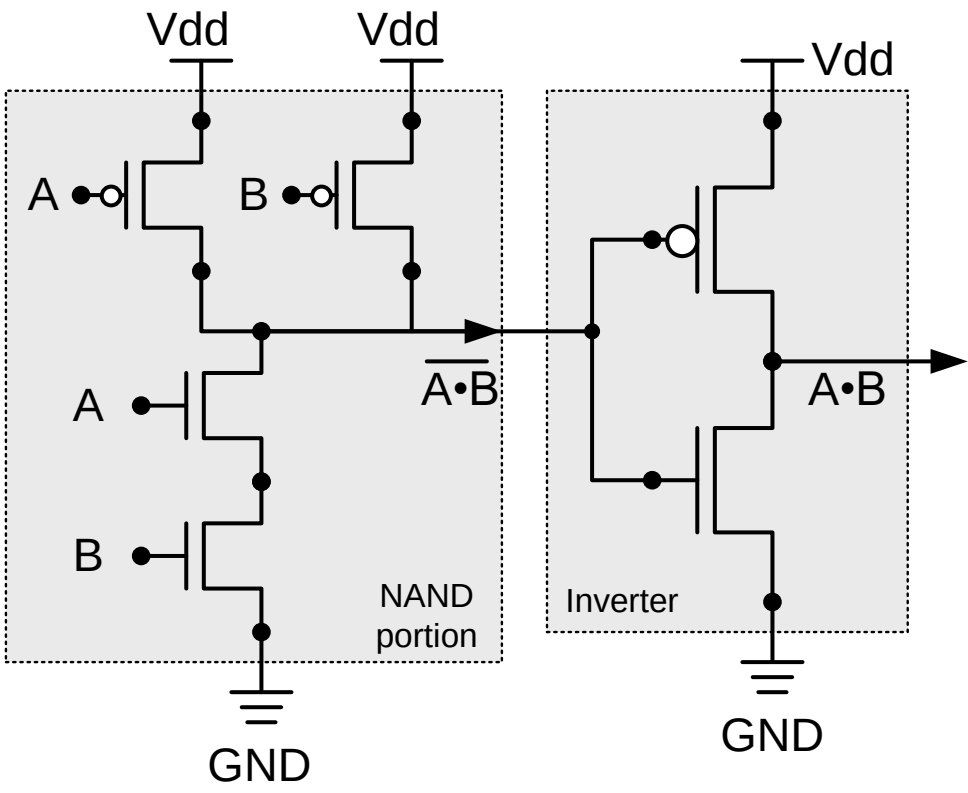
- If A and B = 1, the output of the first circuit is pulled to 0 (opposite of AND function)
- If A or B = 0, the output of the first circuit is pulled to 1 (opposite of AND function)
- Rule of *Conduction Complements*
  - Pull-up network is the dual (complement) of pull-down
  - Parallel -> series, series -> parallel



NAND

CMOS 'AND' Gate

- If A and B = 1, the output of the first circuit is pulled to 0 (opposite of AND function)
- If A or B = 0, the output of the first circuit is pulled to 1 (opposite of AND function)
- Inverter is then used to produce true AND output

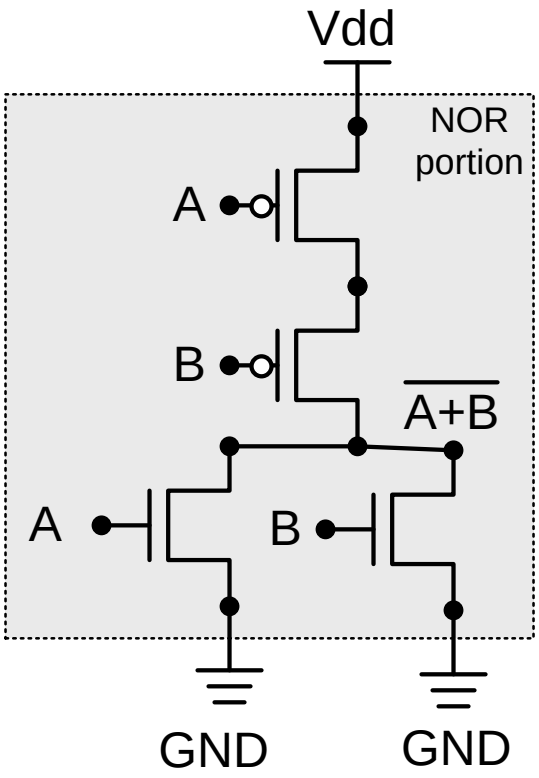


NAND

Inverter to produce AND

CMOS 'NOR' Gate

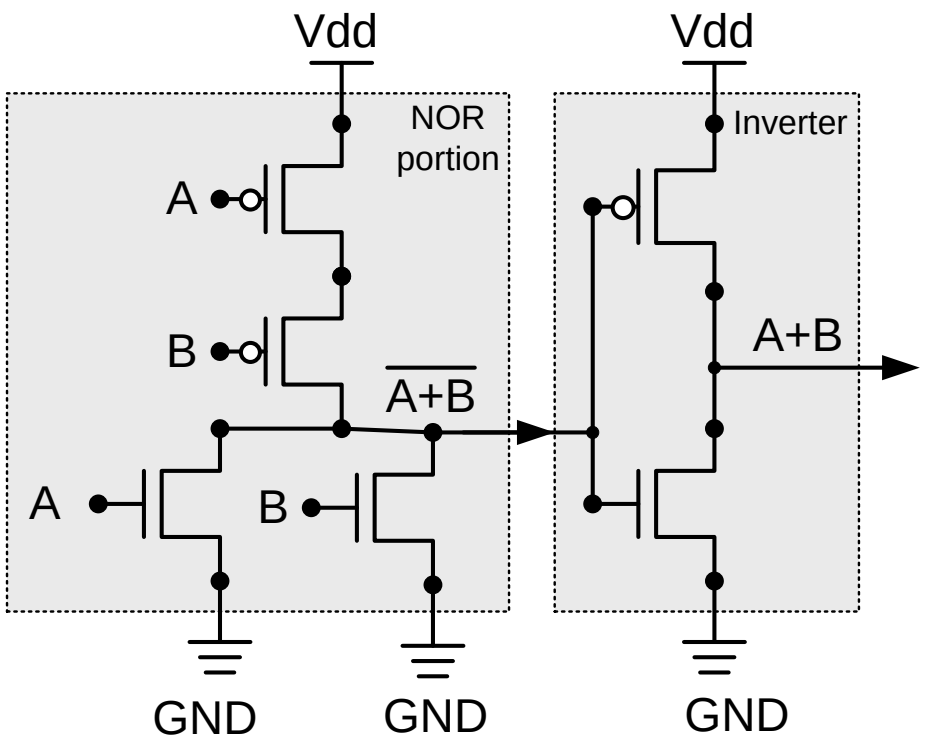
- If A or B = 1, the output of the first circuit is pulled to 0 (opposite of OR function)
- If A and B = 0, the output of the circuit is pulled to 1 (opposite of OR function)
- Rule of *Conduction Complements*
  - Pull-up network is the dual (complement) of pull-down
  - Parallel -> series, series -> parallel



NOR

CMOS 'NOR' Gate

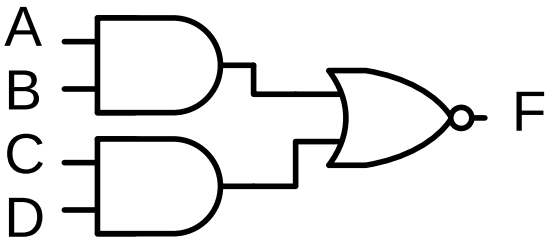
- If A or B = 1, the output of the first circuit is pulled to 0 (opposite of OR function)
- If A and B = 0, the output of the circuit is pulled to 1 (opposite of OR function)
- Inverter is then used to produce true OR output

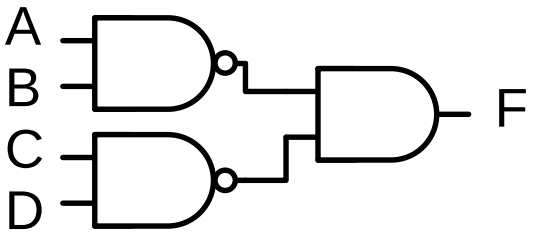


OR

Compound Gates

- How could you build this gate?
- You could try building each gate separately
  - Two AND gates = 2*6 transistors
  - One NOR gate = 4 transistors
- With DeMorgan's
  - Two NAND gates = 2*4 transistors
  - One AND gate = 6 transistors
- Or you could take build it as a single compound gate.

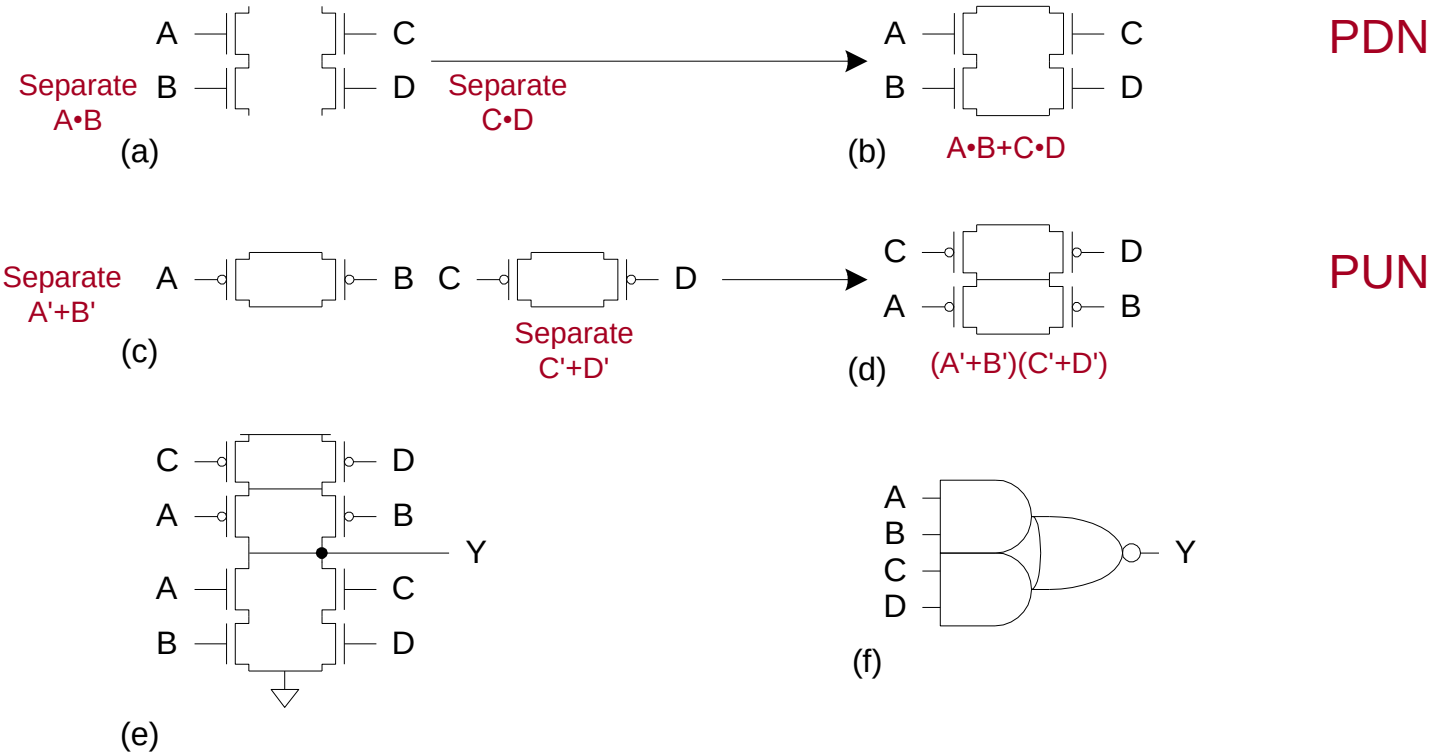




Compound Gates

- *Compound gates* can do any inverting function
- Ex: AND-OR-INVERT (AOI)

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

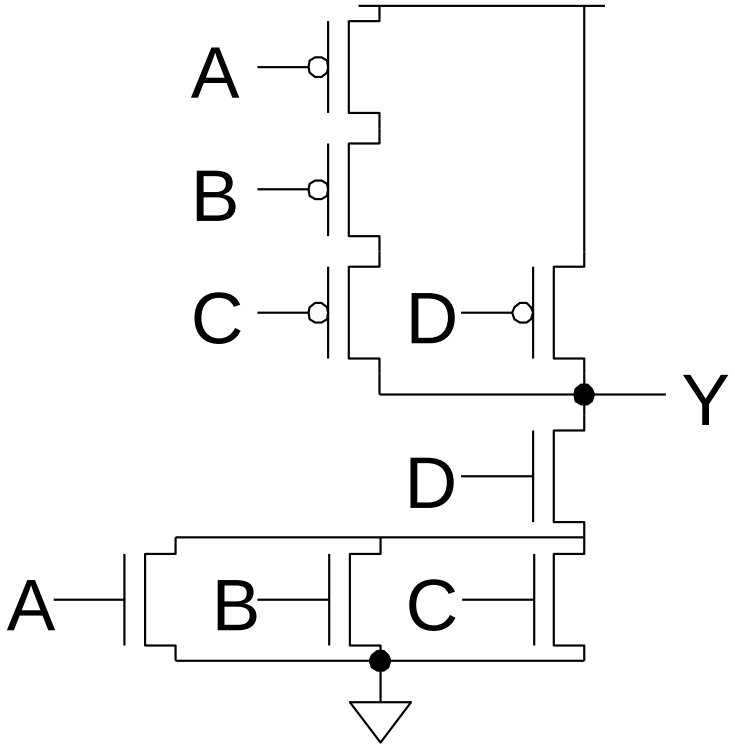


Compound Gate Approach

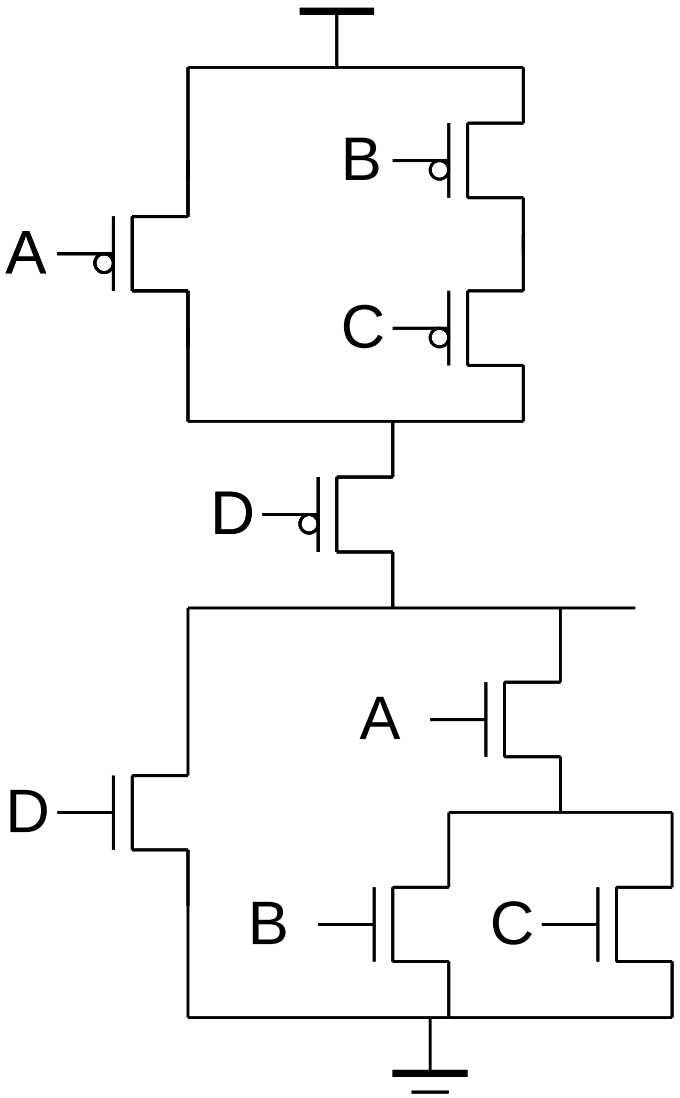
- For an inverting function just look at the expression (w/o the inversion) and…
  - Implement the PDN using:
    - Series connections for AND
    - Parallel connections for OR
  - Implement PUN as dual of PDN
    - Swap series and parallel
- If function is non inverting just add an inverter at the output

Compound Gate Example

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

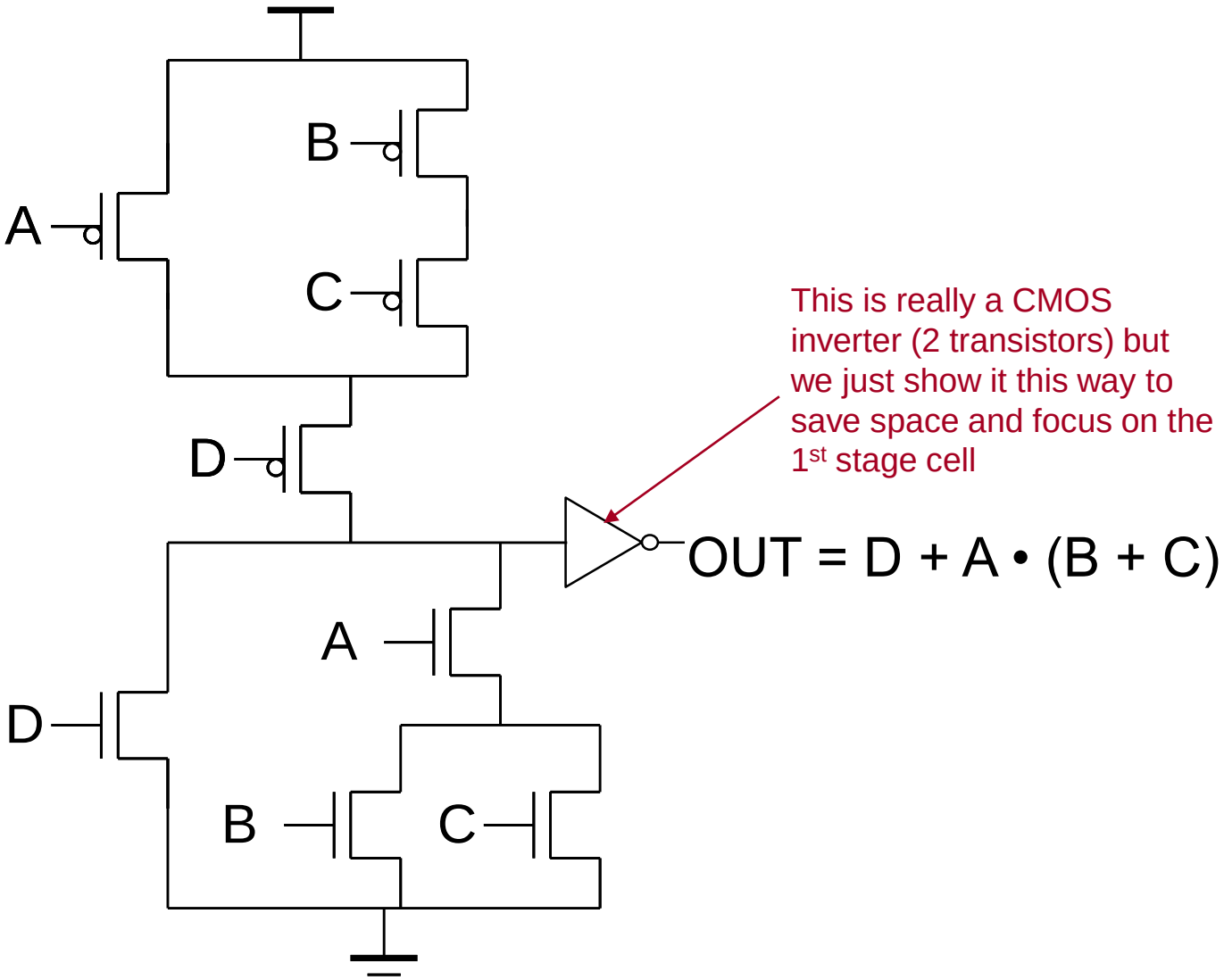


Compound Gate Example



$$OUT = \overline{D + A \bullet (B + C)}$$

Compound Gate Example (cont.)



Another Compound Gate Example

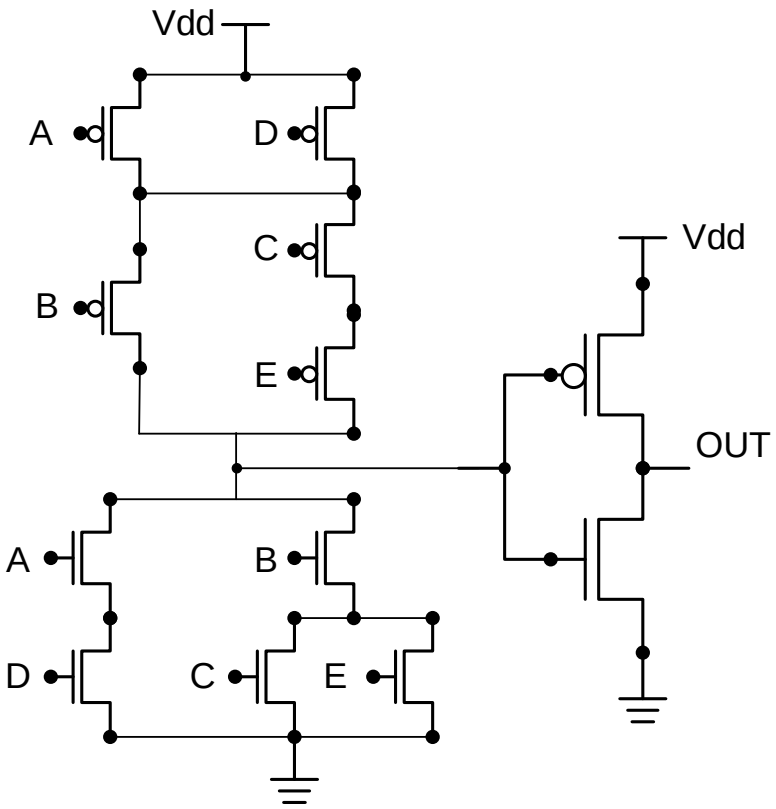
$$OUT = A \cdot D + B(C + E)$$

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

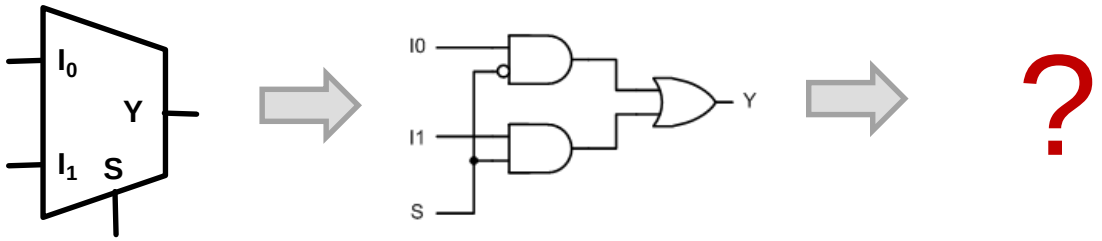
Add an inverter at the output

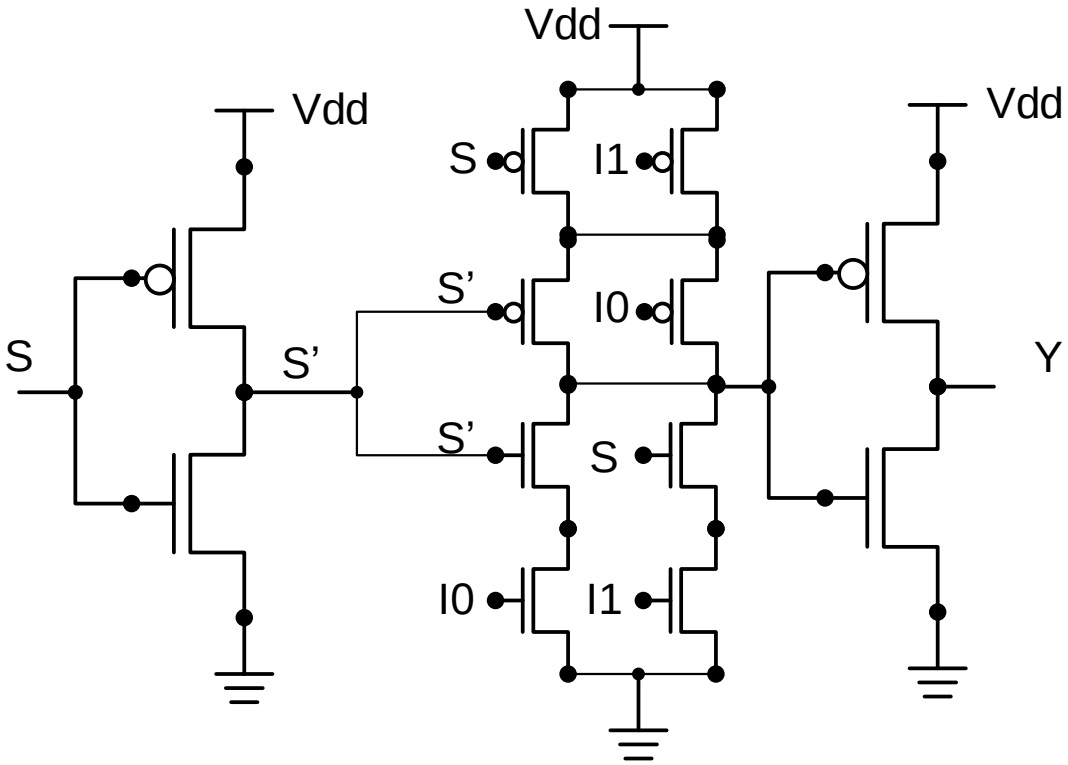
Implement inverting function using compound CMOS gate

OR apply DeMorgan's theorem with the inner inversion and just build the resulting circuit



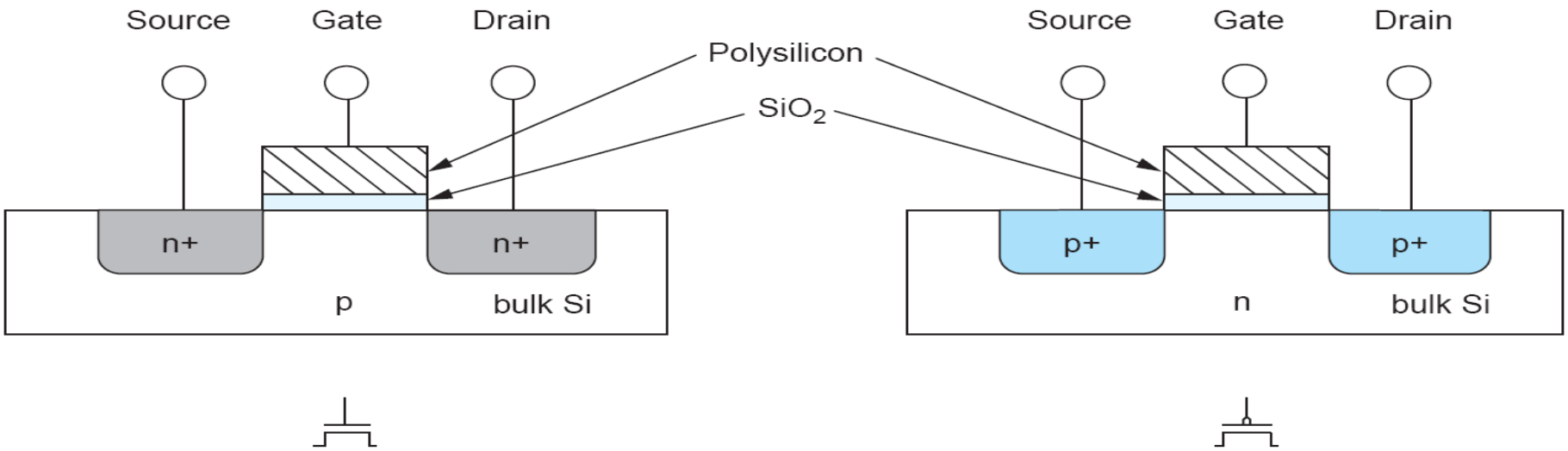
Build a 2-to-1 mux at the Transistor Level

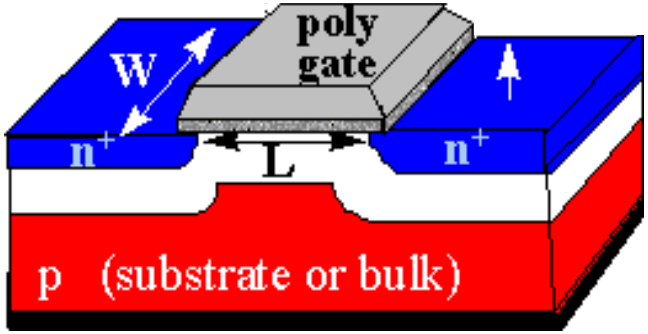




FABRICATION

MOS Layout Structure



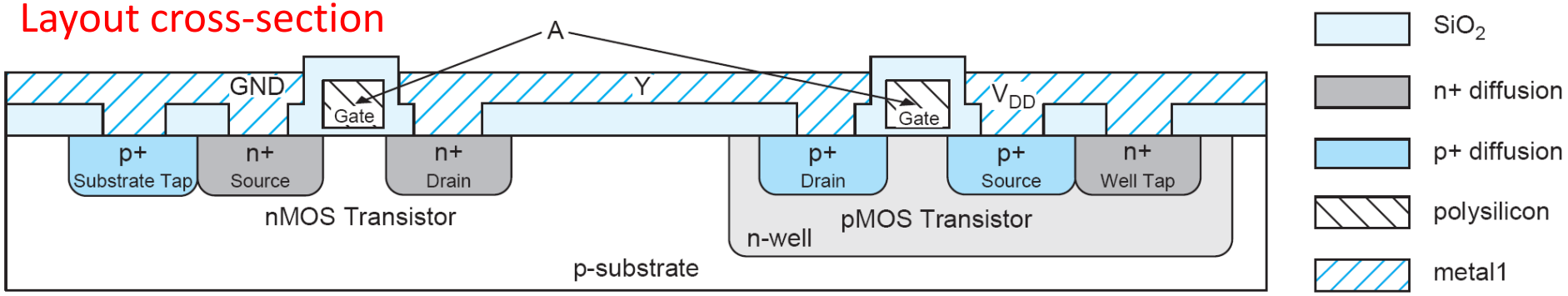


L: Channel Length

W: Channel Width

CMOS Layout Structure

Layout cross-section

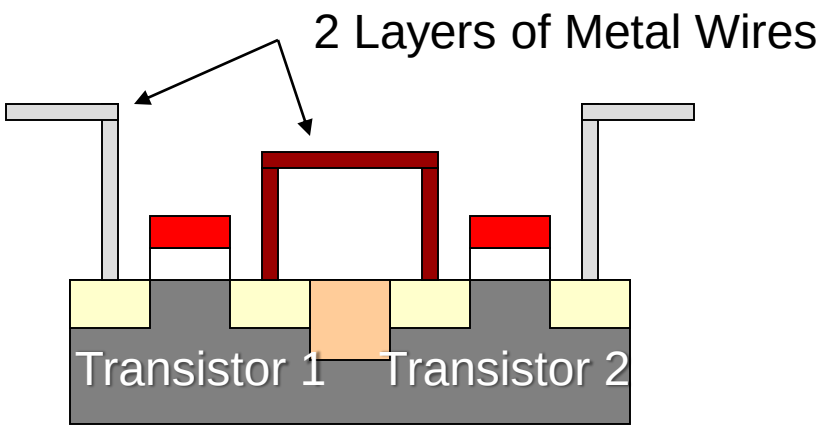


Schematic

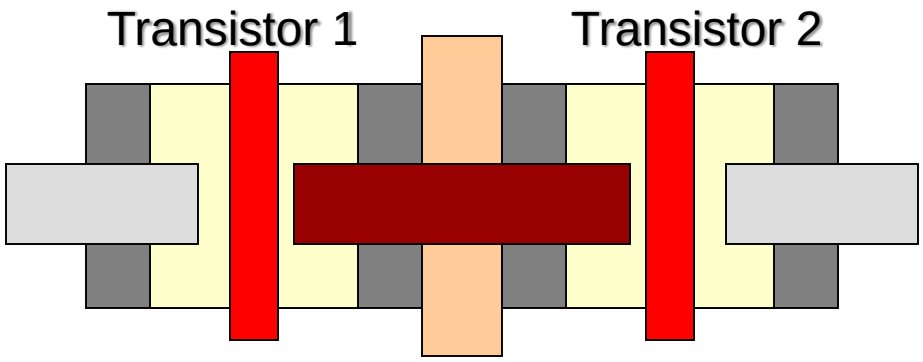
- Both n-channel (NMOS) and p-channel (PMOS) transistors are built on the same chip substrate
  - Well: A special region created in which the semiconductor type is opposite the substrate's type
    - Example: n-well
      - CMOS fabrication technology to create a n-type substrate inside the already p-type substrate
      - The n-well is used to create the PMOS transistors

Layers

- Start from the bottom up
  - Build the n- and p-type material areas on the silicon
  - Lay the insulator layer (oxide) over the silicon
  - Place the polysilicon (gate) on top of the oxide
  - Connect wires to the source, gate, and drain use layers of metal above the gate



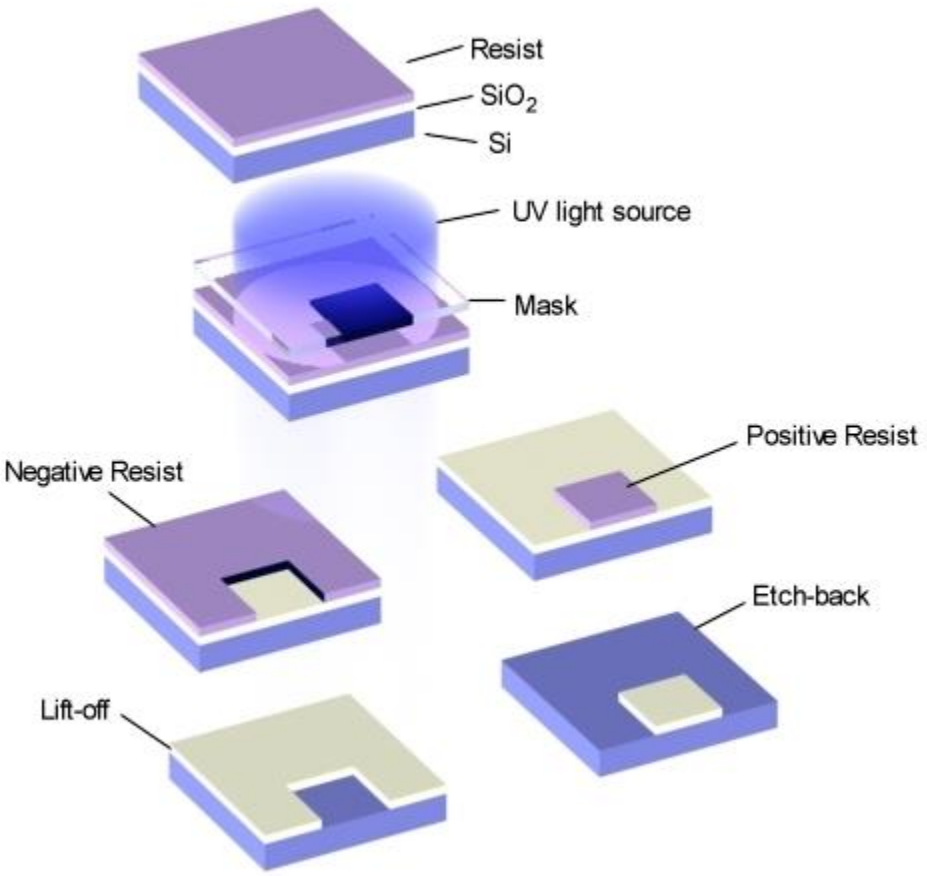
Side View



Top View

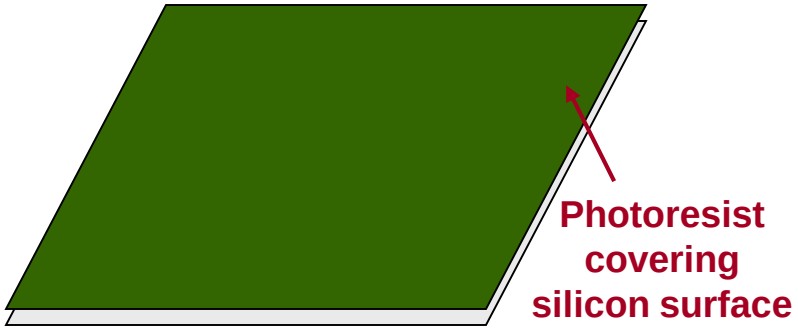
Photolithography

- An IC consists of several layers of material that are manufactured in successive steps
- Lithography is used to selectively process the layers where the 2-D mask geometry is copied on the surface
- Once the desired shape is patterned with photoresist the unprotected areas are etched away
- Lift-off and etching are different techniques to remove and shape



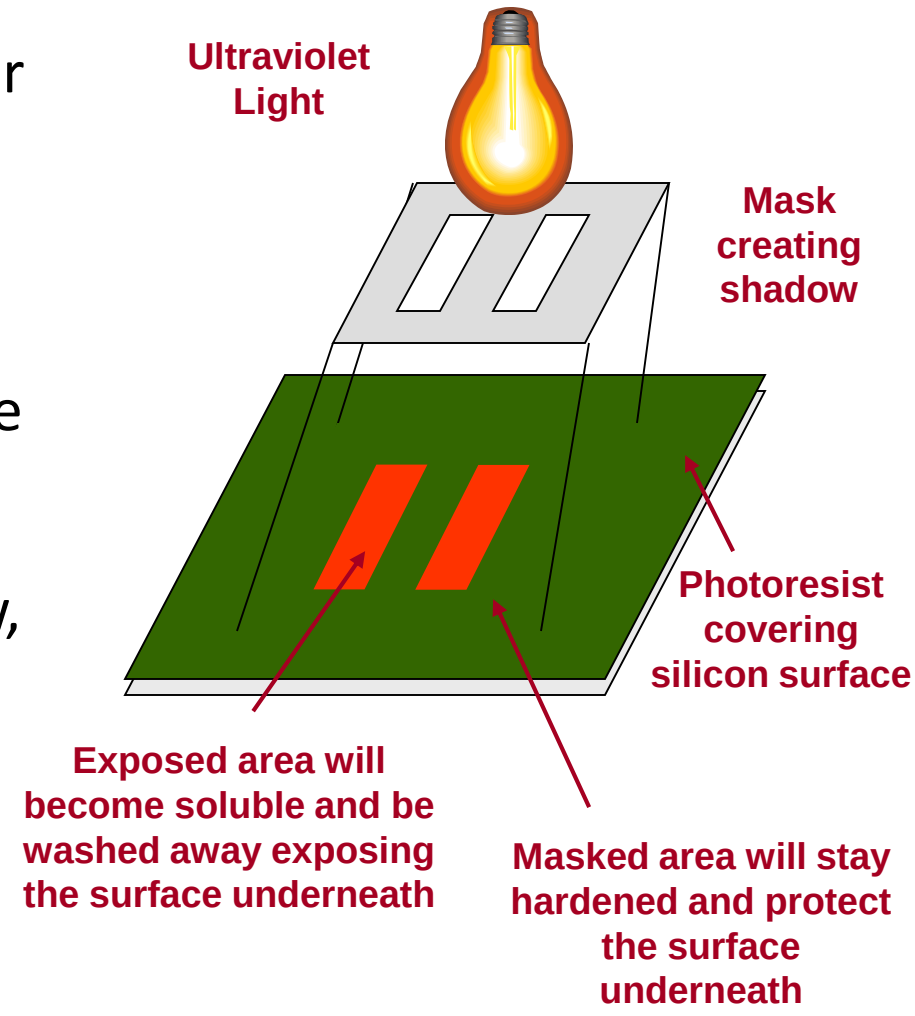
Photolithography

- Expose only specific areas of the chip for layer deposition or etching
- A layer of “photoresist” material is deposited on the chip
- “Photoresist” becomes soluble when exposed to ultraviolet light



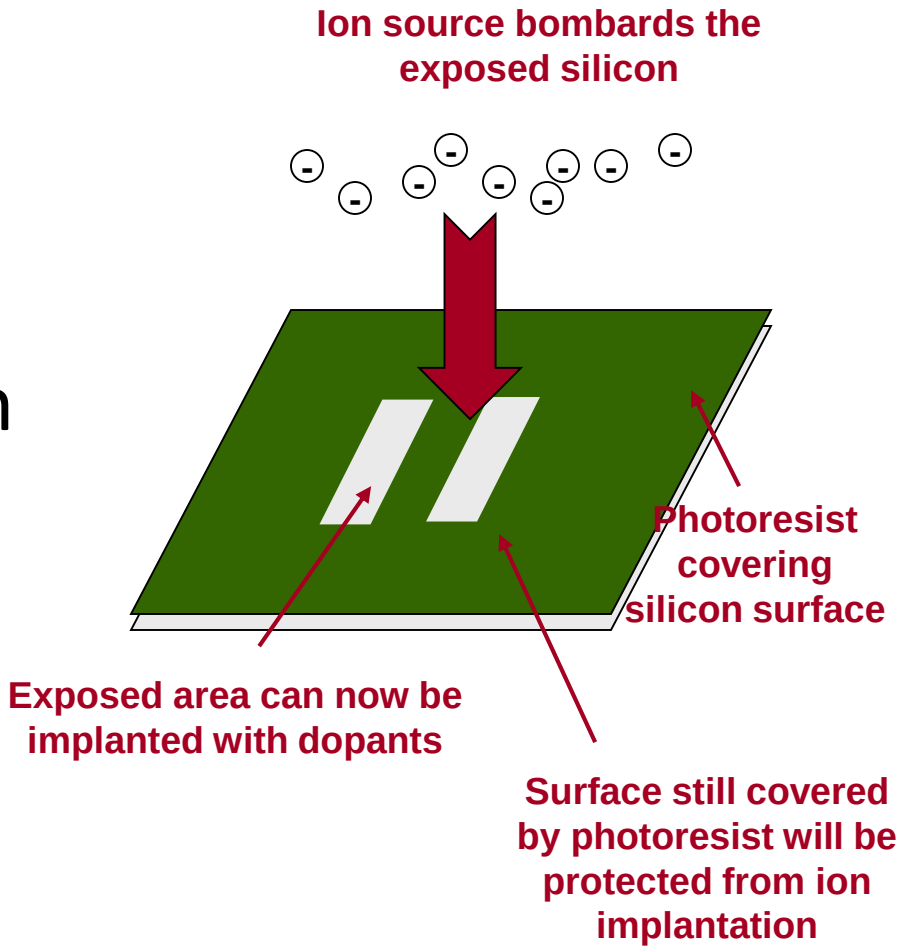
Photolithography

- Expose only specific areas of the chip for layer deposition or etching
- A layer of "photoresist" material is deposited on the chip
- "Photoresist" becomes soluble when exposed to ultraviolet light
- Using a mask to cast a shadow, some portions of photoresist can be kept while the remainder is washed away



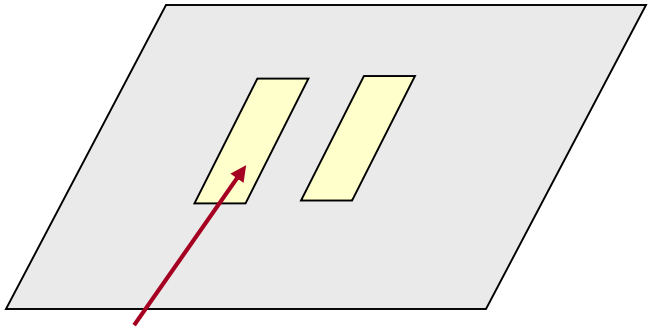
Ion Implantation

- After washing away soluble photoresist, silicon in the shape of the mask is exposed
- Can be implanted with ions to make n- or p-type material



Resulting Material

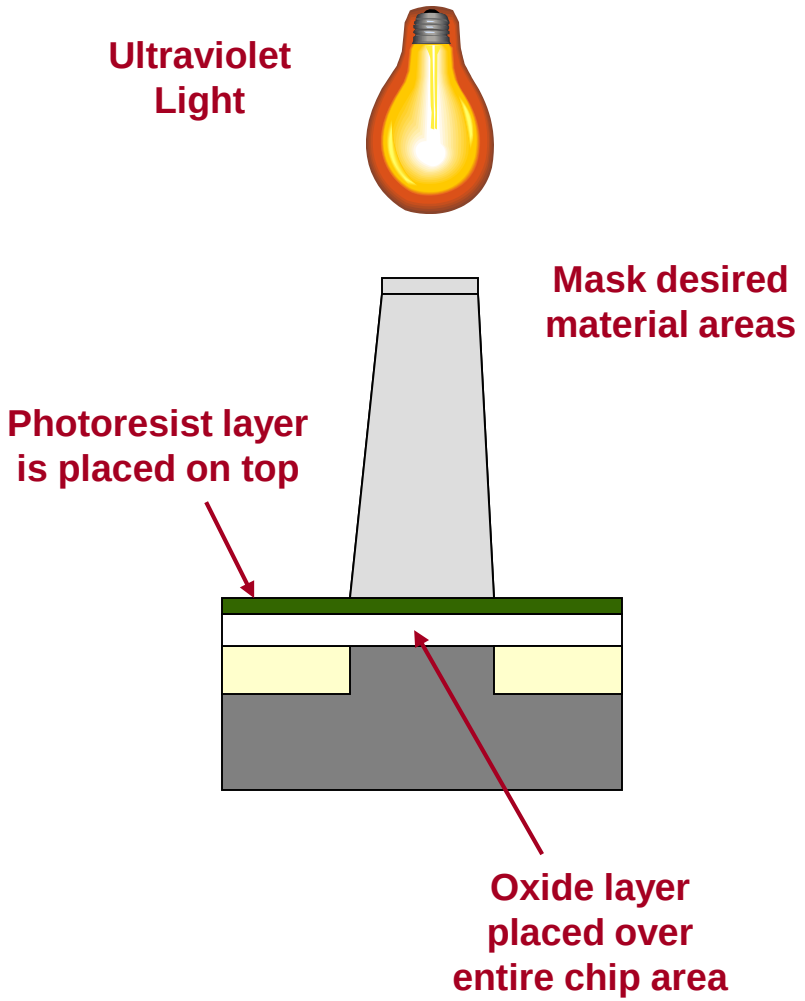
- After implantation, remaining photoresist can be exposed and washed away leaving n-type silicon in the appropriate areas



n-type “doped” silicon

Layer Deposition

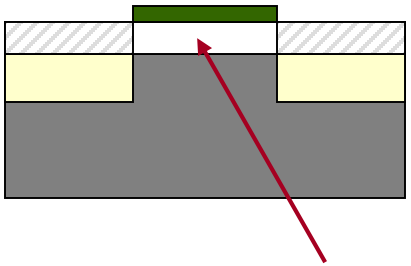
- For layers above the surface (oxide, gate polysilicon, and metal wires), a similar but slightly different process is used
  1. Entire layer of material is deposited over entire area
  2. Covered with photoresist
  3. Mask is used to indicate where material is desired



Layer Deposition

- For layers above the surface (oxide, gate polysilicon, and metal wires), a similar but slightly different process is used
  1. Entire layer of material is deposited over entire chip
  2. Covered with photoresist
  3. Mask is used to indicate where material is desired
  4. Wash away exposed photoresist
  5. Use chemical/mechanical etching process to remove exposed oxide

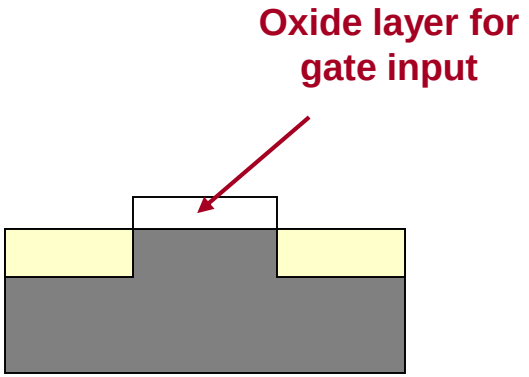
Etching process removes exposed oxide material but cannot penetrate photoresist material



Oxide layer placed over entire chip area

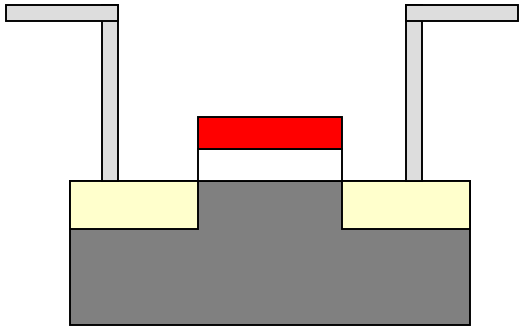
Layer Deposition

- For layers above the surface (oxide, gate polysilicon, and metal wires), a similar but slightly different process is used
  1. Entire layer of material is deposited over entire chip
  2. Covered with photoresist
  3. Mask is used to indicate where material is desired
  4. Wash away exposed photoresist
  5. Use chemical/mechanical etching process to remove exposed oxide
  6. Remaining photoresist can be removed exposing oxide in the desired location

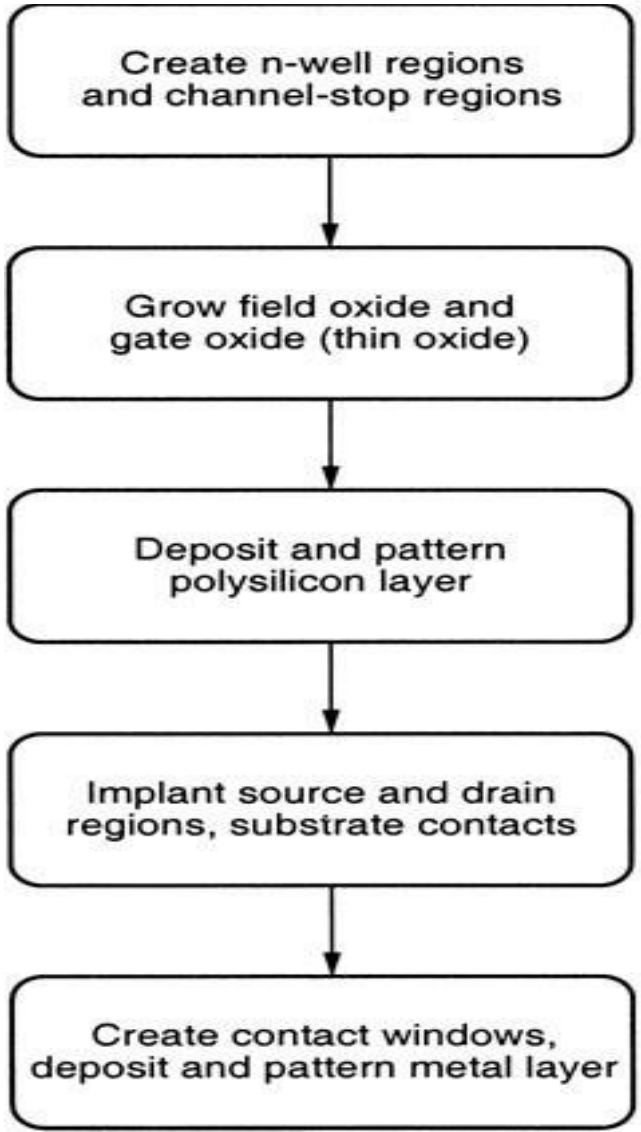


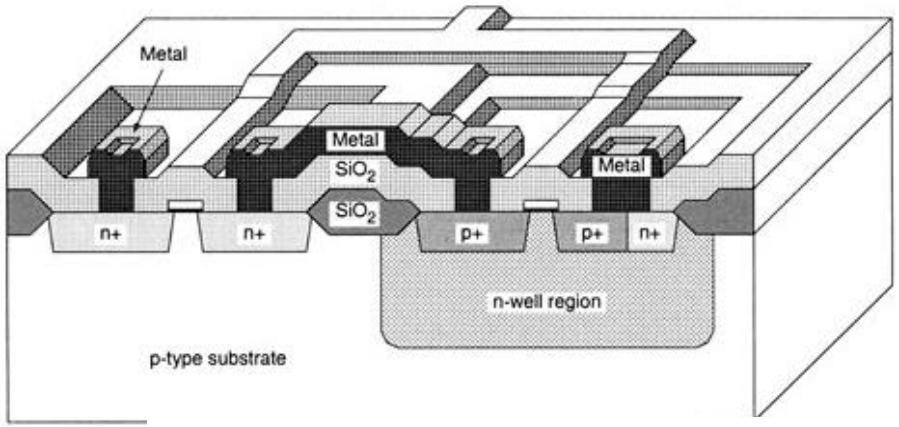
Layer Deposition

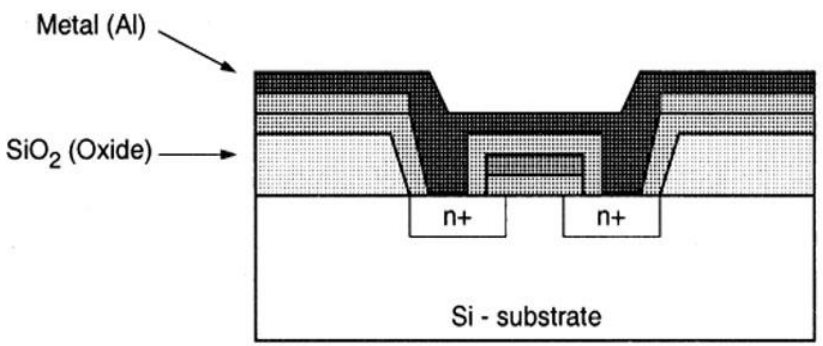
- Process is repeated for gate (polysilicon) and metal wire layers
- A separate mask is required for each layer to indicate where the substance should be kept and where it should be etched away

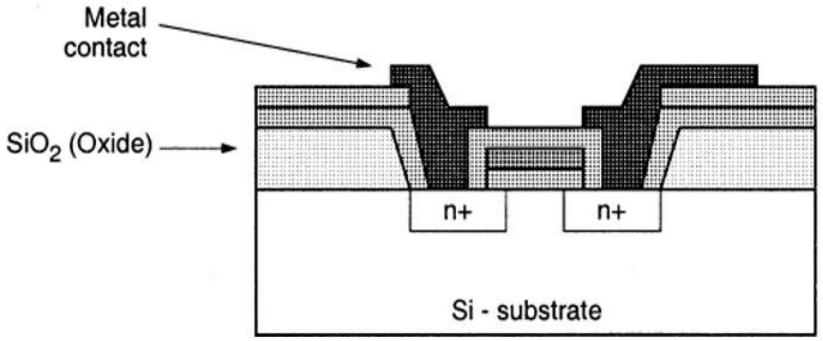


Simplified CMOS Fabrication Process

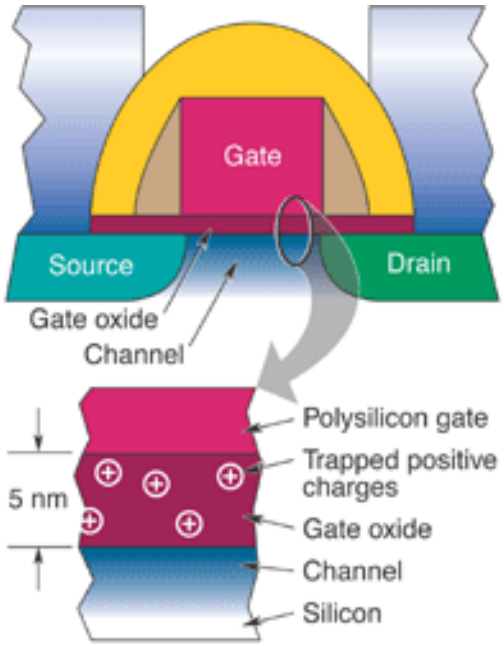


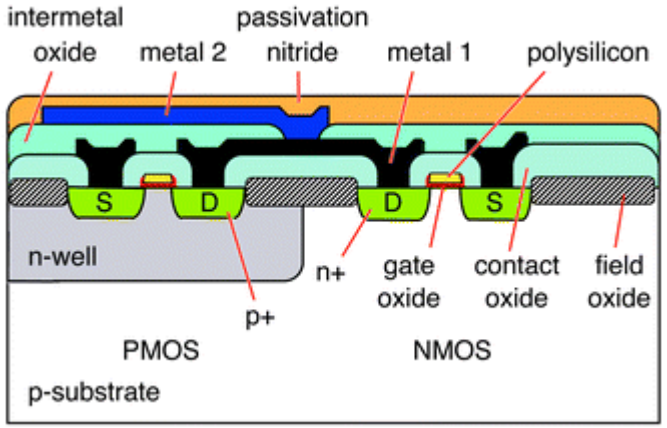


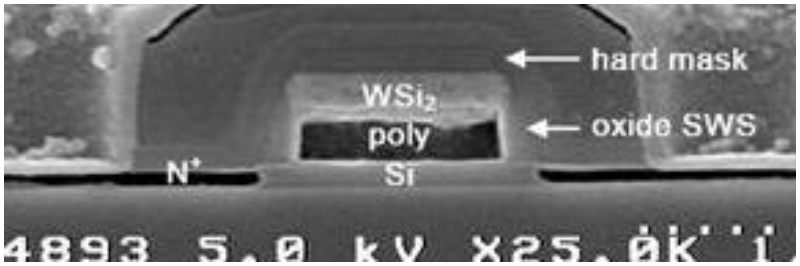


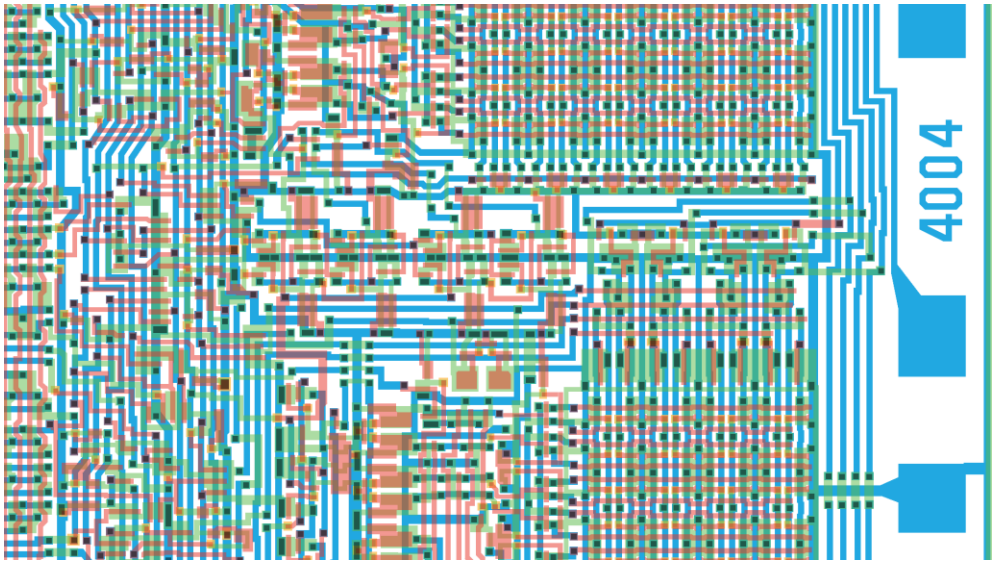


Fabrication Images









http://pubs.rsc.org/services/images/RSCpubs.ePlatform.Service.FreeContent.ImageService.svc/ImageService/Articleimage/2003/AN/b208563c/b208563c-f1.gif

http://www.4004.com/assets/4004-east-mask-detail-hdcrop.gif