CSUN | CALIFORNIA STATE UNIVERSITY NORTHRIDGE

DR JEFF SOFTWARE
INDIE APP DEVELOPER

COMP222

COMP 222

Spring 2023 Rev 2-21-23

Computer Organization

(Architecture)

Lecture 1A: Intro

Dr Jeff Drobman

website ➡ *drjeffsoftware.com/classroom.html*

email ➡ *jeffrey.drobman@csun.edu*

Index (vol. 1A)

Section

Numbers

Binary/Hex vs Decimal

Is there any point where using binary, hexadecimal or octal instead of decimal is better or worse from a computational standpoint?

Jeff Drobman
Lecturer at California State University, Northridge (2016–present) · Just now ·

Binary is the radix of choice due to digital logic being in 2 states. Decimal is usually binary encoded in BCD and has no benefit.

Binary is a base 2 number system. *Octal* is merely *representing* binary as 3-bit digits, and hex as 4-bit digits. Sort of like zooming in or out of same data.

Ordinals

Powers of 2 <> 10: 10:3

Technical ordinals

10^(-24)	yacto
10^(-21)	zepto
10^(-18)	atto
10^(-15)	femto
10^(-12)	pico
10^(-9)	nano
10^(-6)	micro
10^(-3)	milli
10^(-2)	centi
10^(-1)	deci
10^(+1)	deka
10^(+2)	hecto
10^(+3)/2^(10)	kilo
10^(+6)/2^(20)	mega
10^(+9)/2^(30)	giga
10^(+12)/2^(40)	tera
10^(+15)/2^(50)	peta
10^(+18)/2^(60)	exa
10^(+21)/2^(70)	zetta
10^(+24)/2^(80)	yotta

Gazillions

10^(+6)	million
10^(+9)	billion
10^(+12)	trillion
10^(+15)	quadrillion
10^(+18)	quintillion
10^(+21)	sexillion
10^(+24)	septillion
10^(+27)	octillion
10^(+30)	nonillion
10^(+33)	decillion
10^(+36)	undecillion
10^(+39)	duodecillion
10^(+42)	tredecillion
10^(+45)	quattuordecillion
10^(+48)	quindecillion
10^(+51)	sexdecillion
10^(+54)	septendecillion
10^(+57)	octodecillion
10^(+60)	novemdecillion
10^(+63)	vigintillion
10^(+100)	googol
10^(+303)	centillion
10^(10^(+100))	googolplex

Ordinal	Power of 2	Power of 10	Actual
1K	2^{10}	10^{3}	1024
1M	2^{20}	10^{6}	1,048,576
1G	2^{30}	10^{9}	1.074×10^{9}
1T	2^{40}	10^{12}	1.0995×10^{12}

Name	2^n	M/G	Actual
byte	2^{8}	--	256
short	2^{16}	64K	65,536
integer	2^{32}	4B	4.3×10^{9}
long	2^{64}	16 Q	1.84×10^{19}
IPv6	2^{128}	340 uD	3.4×10^{38}

GiB/TiB ($2^{30}/2^{40}$)

Decimal	Abbreviation	Value	Binary term	Abbreviation	Value	% Larger
kilobyte	KB	10^3	kibibyte	KiB	2^{10}	2%
megabyte	MB	10^6	mebibyte	MiB	2^{20}	5%
gigabyte	GB	10^9	gibibyte	GiB	2^{30}	7%
terabyte	TB	10^{12}	tebibyte	TiB	2^{40}	10%
petabyte	PB	10^{15}	pebibyte	PiB	2^{50}	13%
exabyte	EB	10^{18}	exbibyte	EiB	2^{60}	15%
zettabyte	ZB	10^{21}	zebibyte	ZiB	2^{70}	18%
yottabyte	YB	10^{24}	yobibyte	YiB	2^{80}	

Actual
1024
1,048,576
$1.074x10^9$
$1.0995x10^{12}$

Ordinal	Power of 2	Power of 10	Actual
1K	2^{10}	10^3	1024
1M	2^{20}	10^6	1,048,576
1G	2^{30}	10^9	$1.074x10^9$
1T	2^{40}	10^{12}	$1.0995x10^{12}$

Signed Numbers

Binary	Unsigned	Sign Magnitude	Excess-127	Two's Complement
00000000	0	0	-127	0
00000001	1	1	-126	1
⋮	⋮	⋮	⋮	⋮
01111110	126	126	-1	126
01111111	127	127	0	127
10000000	128	-0	1	-128
10000001	129	-1	2	-127
⋮	⋮	⋮	⋮	⋮
11111110	254	-126	127	-2
11111111	255	-127	128	-1

ASCII Codes

Binary, hexadecimal, and decimal equivalents for each character in "Hello World"

Character	Binary	Hexadecimal	Decimal
H	01001000	48	72
e	01100101	65	101
l	01101100	6C	108
l	01101100	6C	108
o	01101111	6F	111
	00100000	20	32
W	01010111	57	87
o	01101111	6F	111
r	01110010	62	98
l	01101100	6C	108
d	01100100	64	100
NUL	00000000	00	0

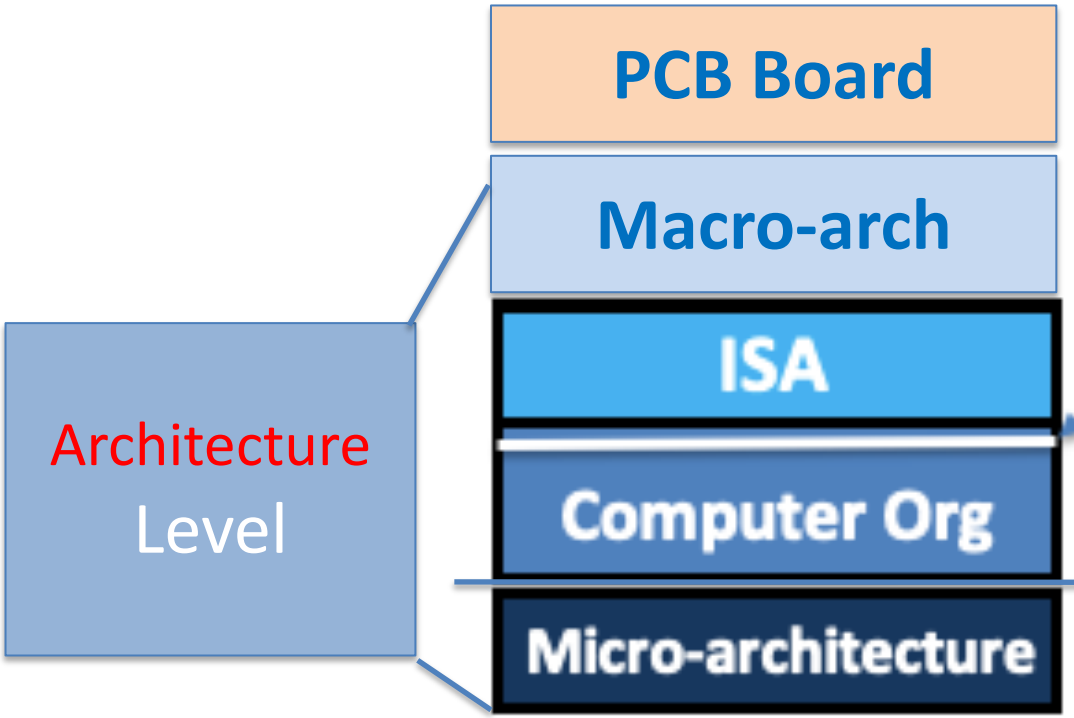
Section

Logic

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP222

Computer Architecture

4-Layer Stack Model



Chipsets, Chiplets and Sockets

System arch – *Cores, caches*

Instructions (Primitives)
***Software** Interface*

122

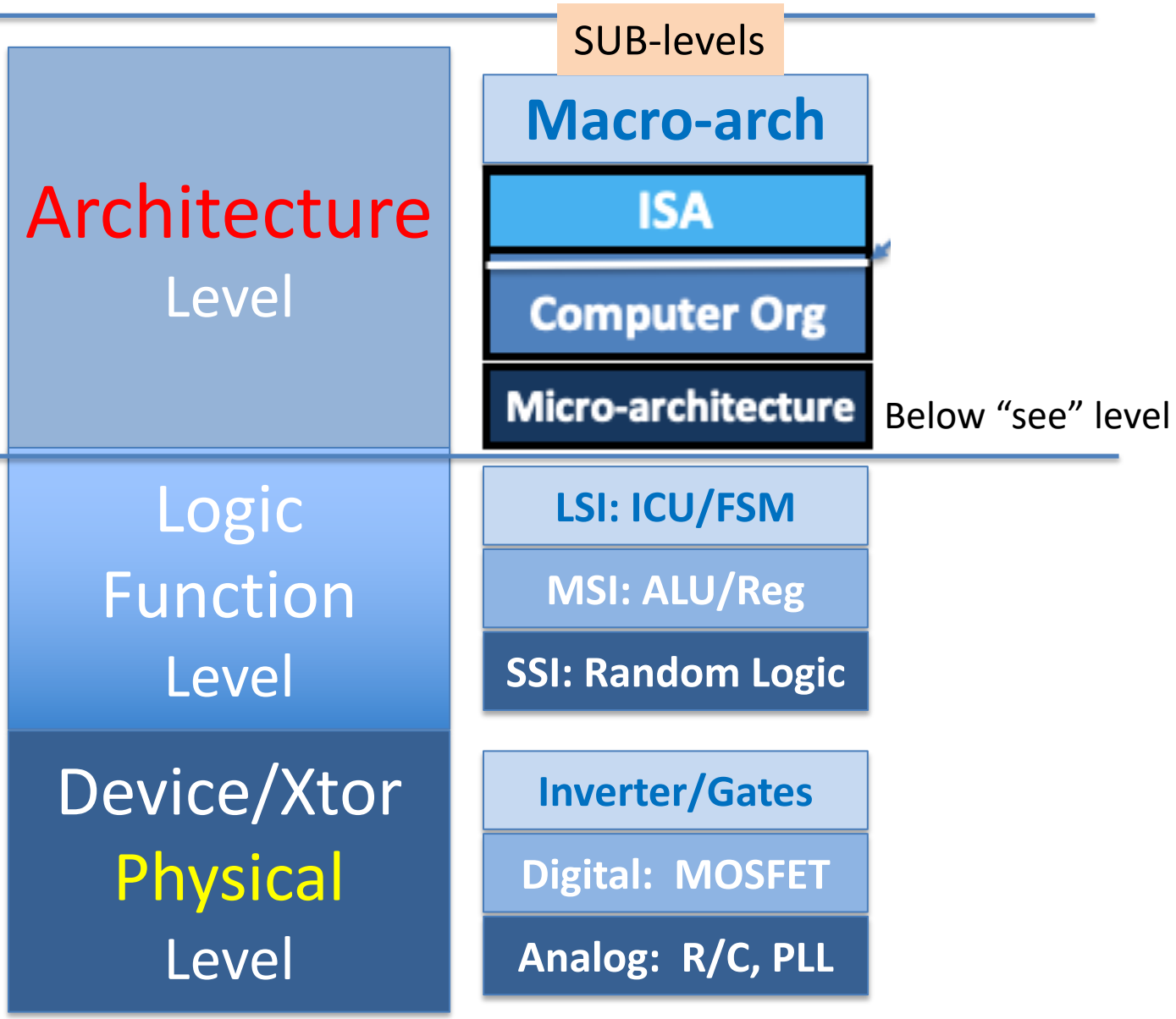
Execution Units
- ALU, ICU, Reg

Below "see" level

Low-level execution
- Pipelines, threads
- scheduling
- branch prediction

222

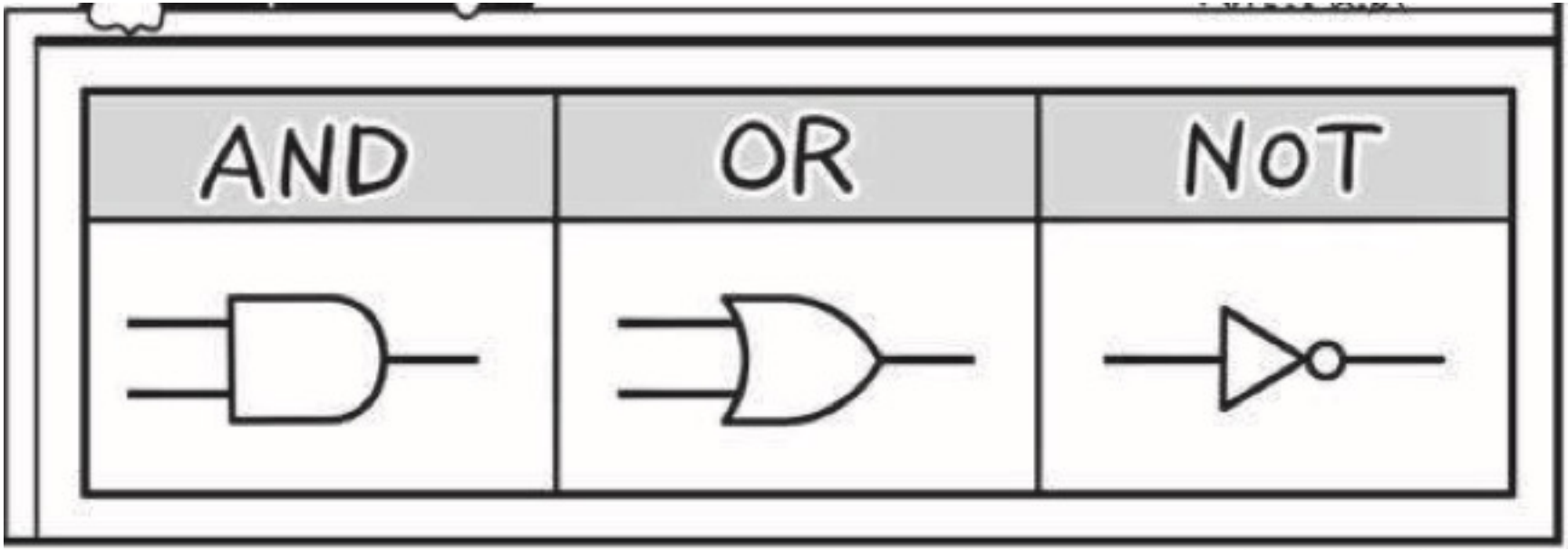
Transistors to Chips: Levels

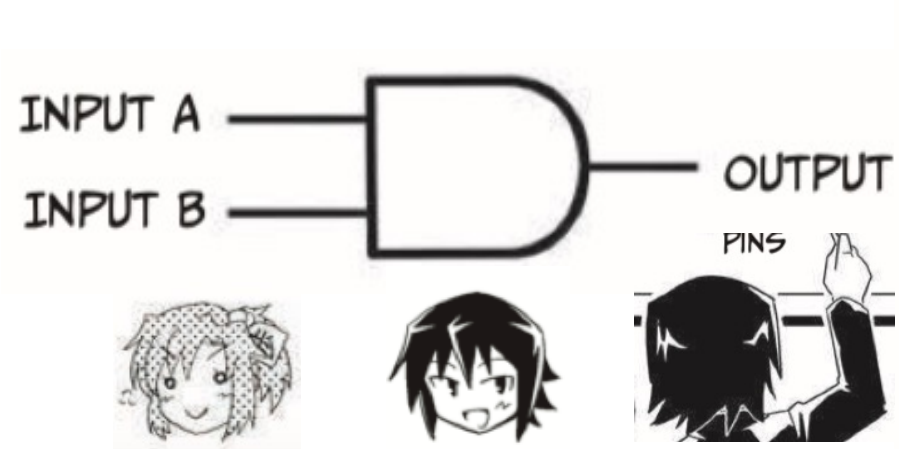


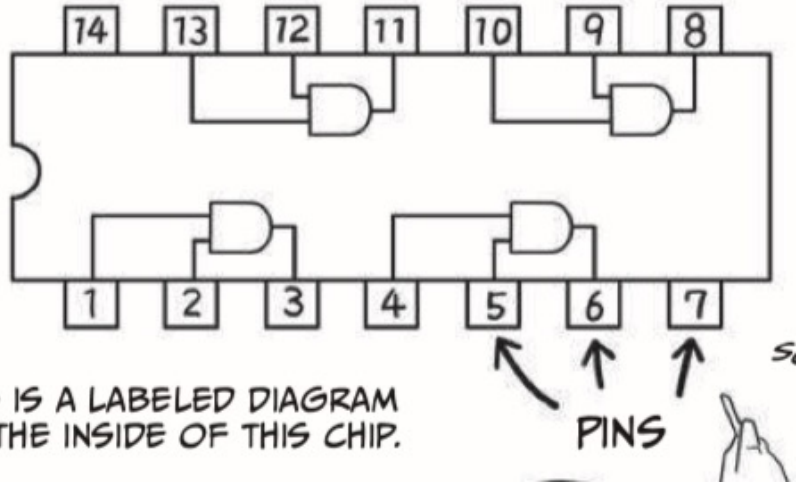
Logic Universal Set

Manga Guide

https://nostarch.com/download/MangaGuidetoMicroprocessors_sample_Chapter2.pdf







Logic Gates: Polarity

DeMorgan's Law

Manga Guide

That's it! It also means that we can use De Morgan's laws to show our circuits in different ways. Using this technique, it's easy to simplify schematics when necessary.

BOTH OF THESE ARE NAND GATES!

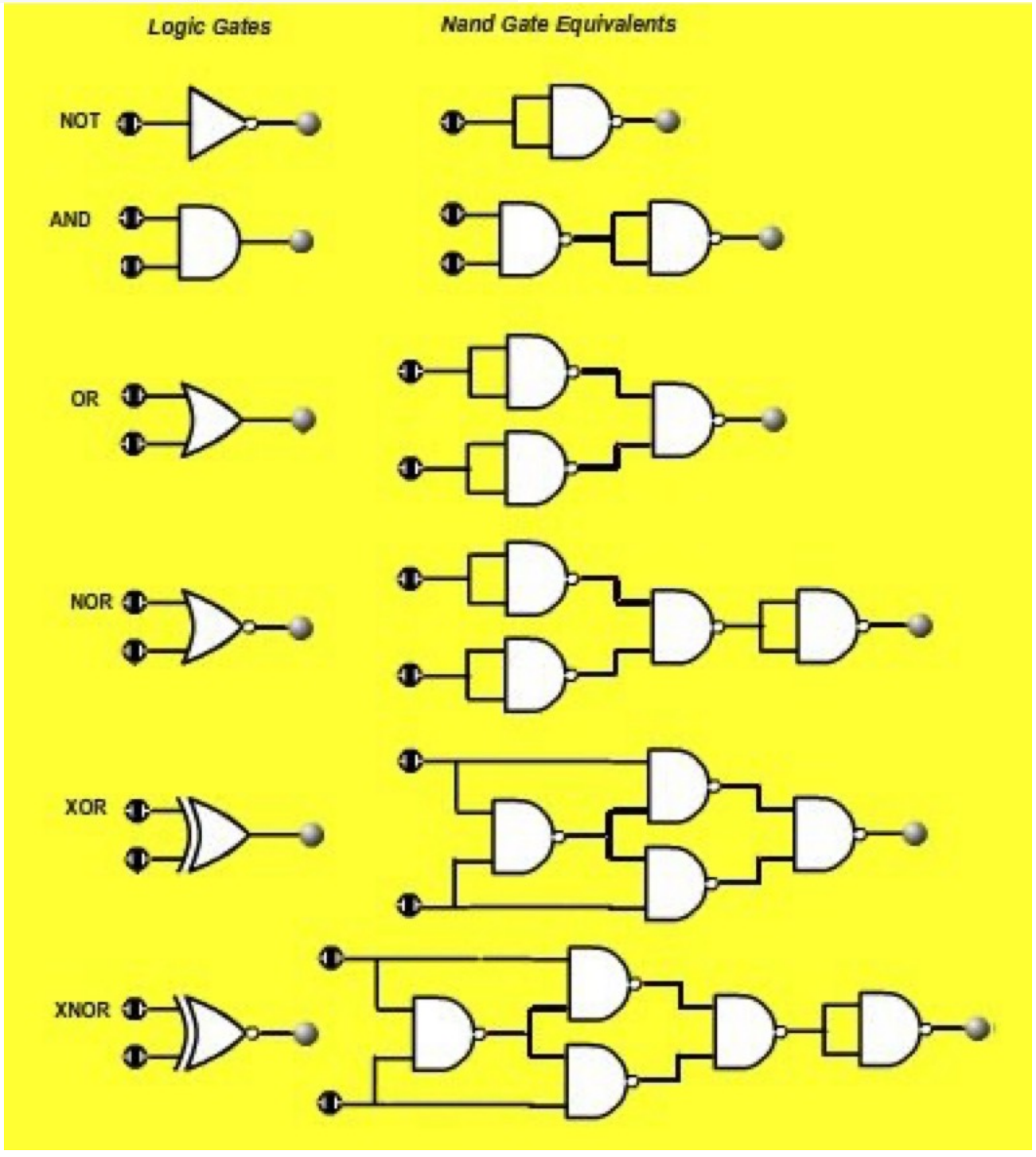
BOTH OF THESE ARE NOR GATES!

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP222

NAND Gates

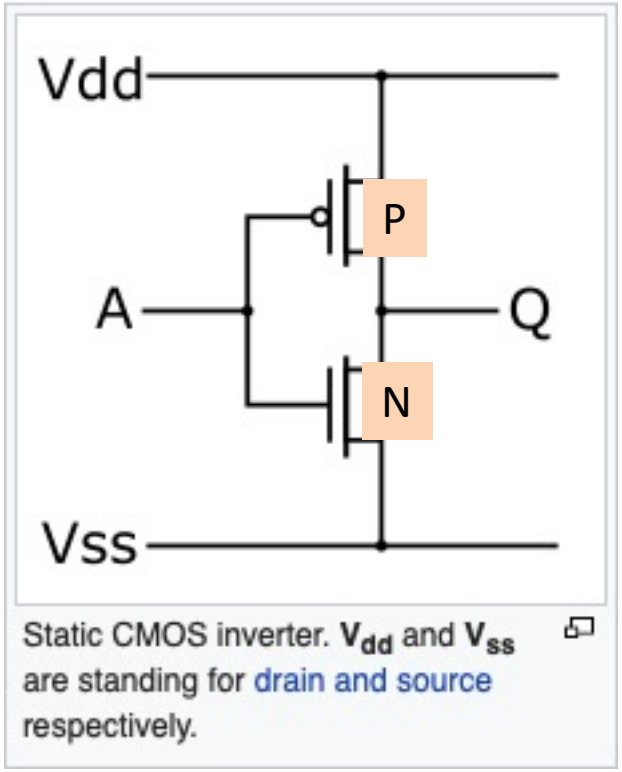
Quora

Tom Crosley · Follow
Embedded systems programmer for 45 years

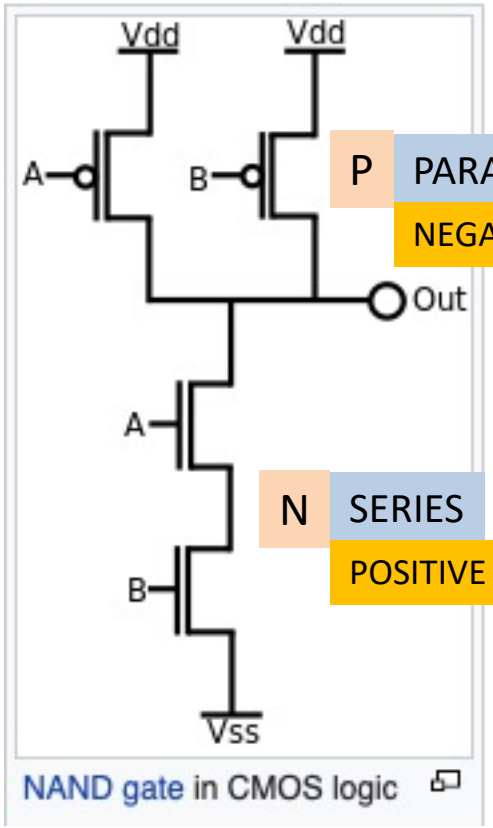


CMOS Gates

MOSFET



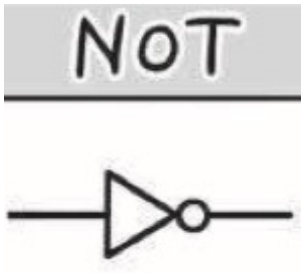
Static CMOS inverter. V_{dd} and V_{ss} are standing for drain and source respectively.

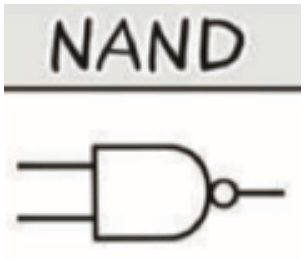


NAND gate in CMOS logic

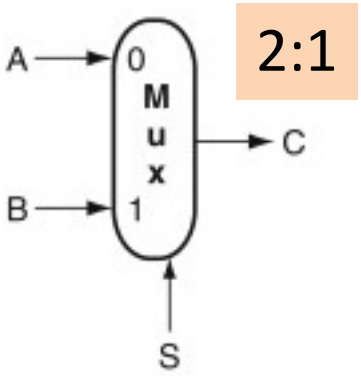
P PARALLEL
NEGATIVE OR

N SERIES
POSITIVE AND

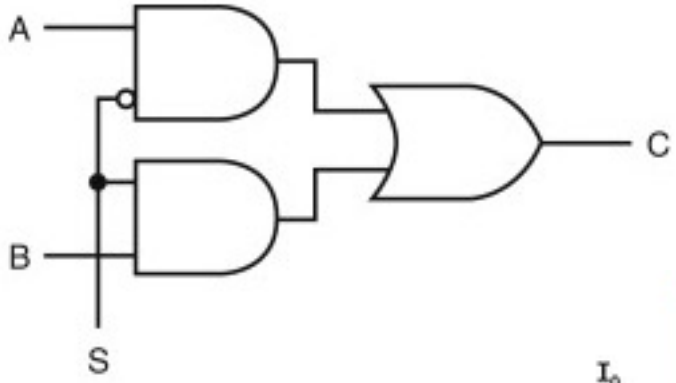




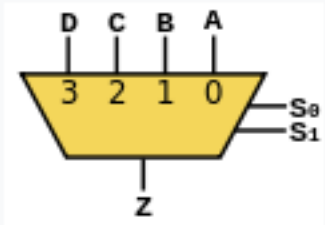
MUX



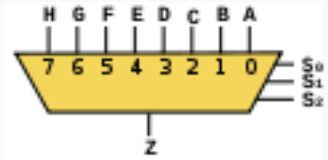
2:1



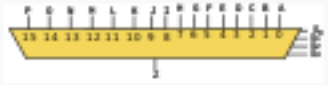
The following 4-to-1 multiplexer is constructed from 3-state buffers and AND gates



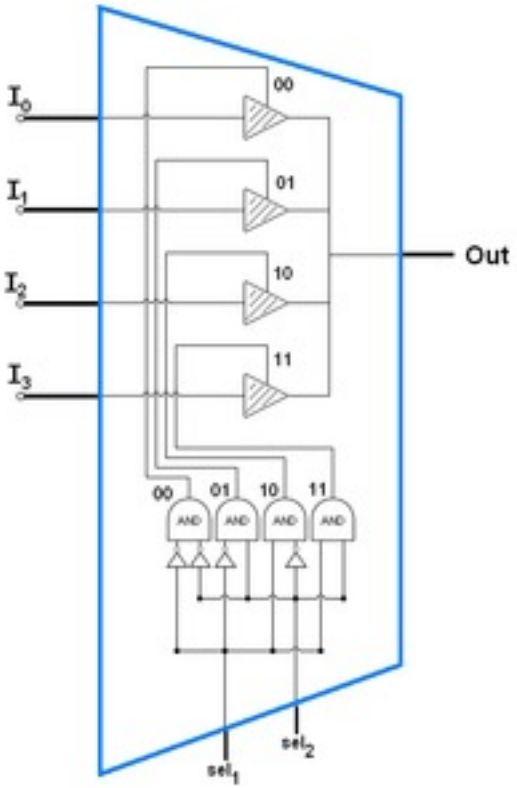
4-to-1 mux



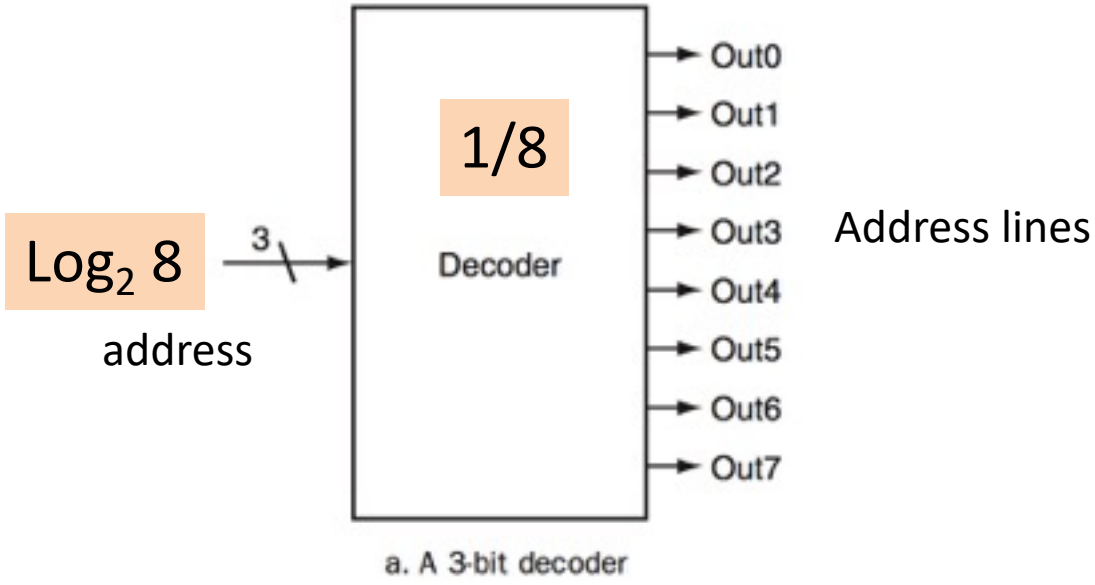
8-to-1 mux



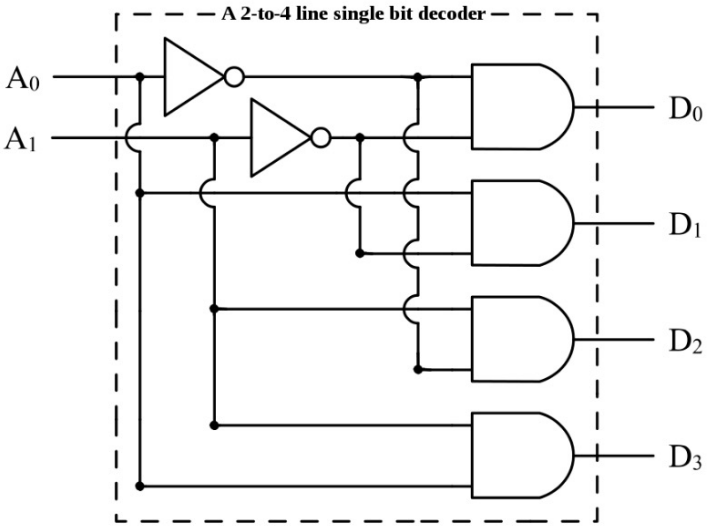
16-to-1 mux



Decoder



a. A 3-bit decoder



Truth Table

A_1	A_0	D_3	D_2	D_1	D_0
0	0	0	0	0	1
0	1	0	0	1	0
1	0	0	1	0	0
1	1	1	0	0	0

Minterm Equations

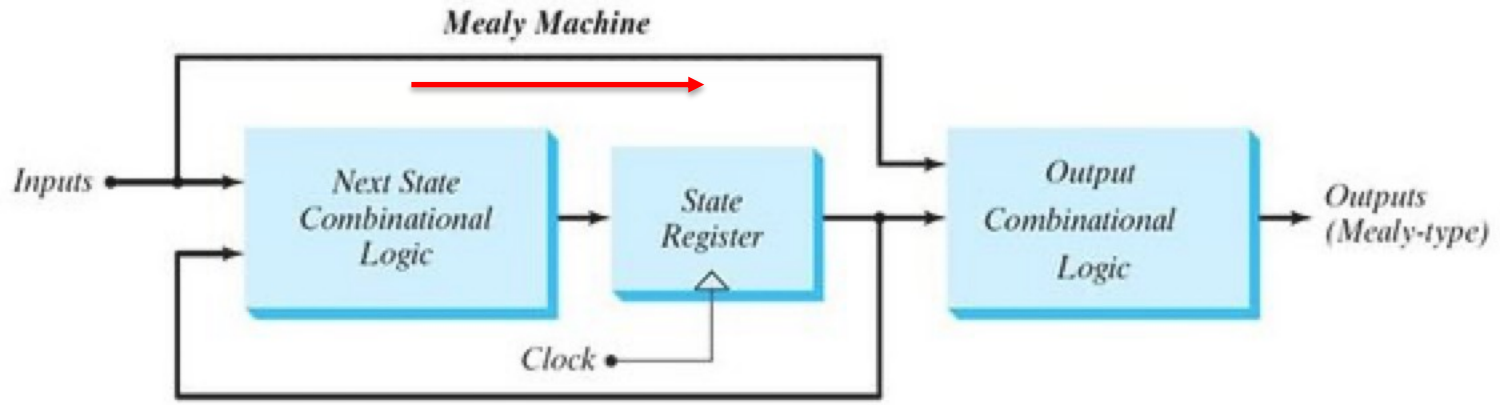
$$D_0 = \overline{A_1} \cdot \overline{A_0}$$

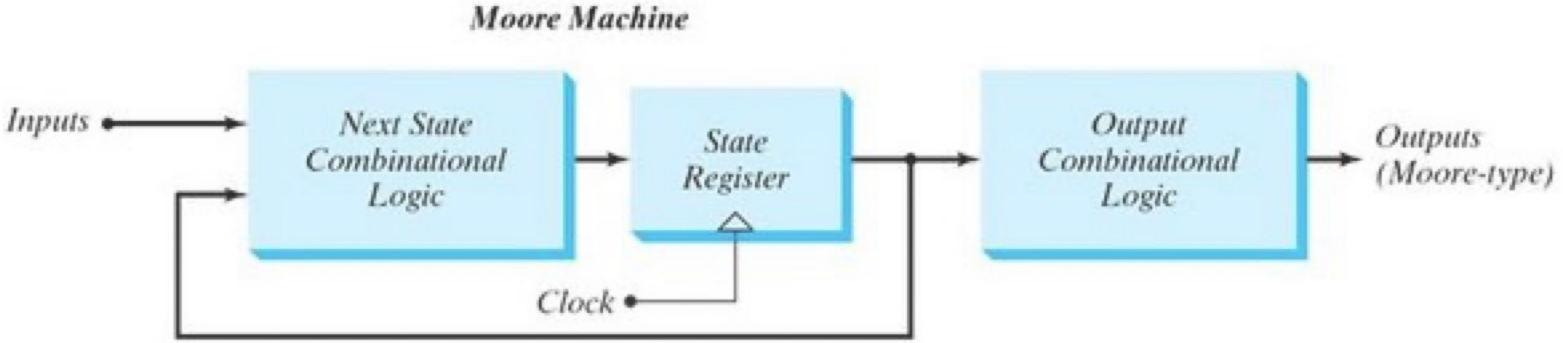
$$D_1 = \overline{A_1} \cdot A_0$$

$$D_2 = A_1 \cdot \overline{A_0}$$

$$D_3 = A_1 \cdot A_0$$

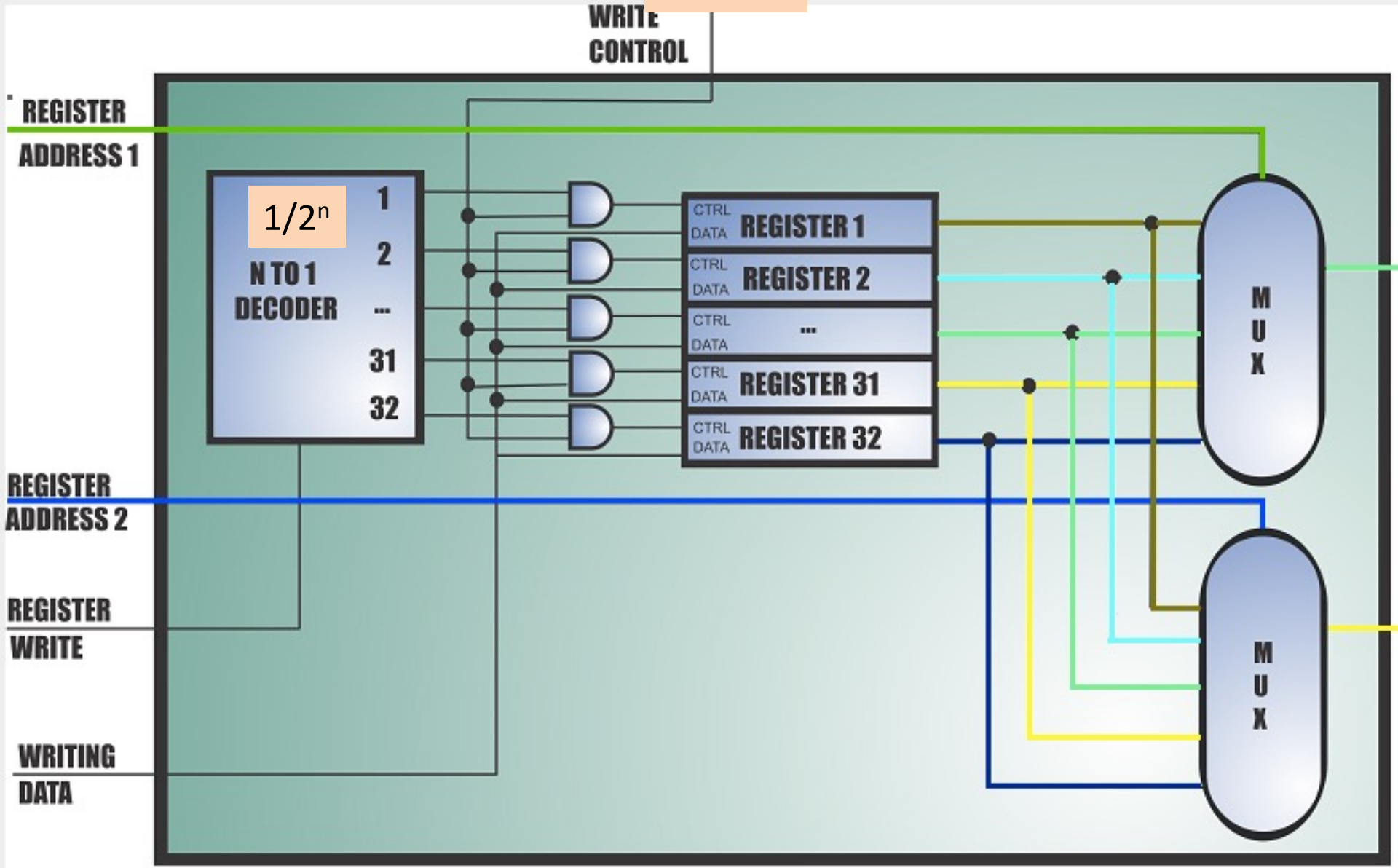
Mealy-Moore FSM



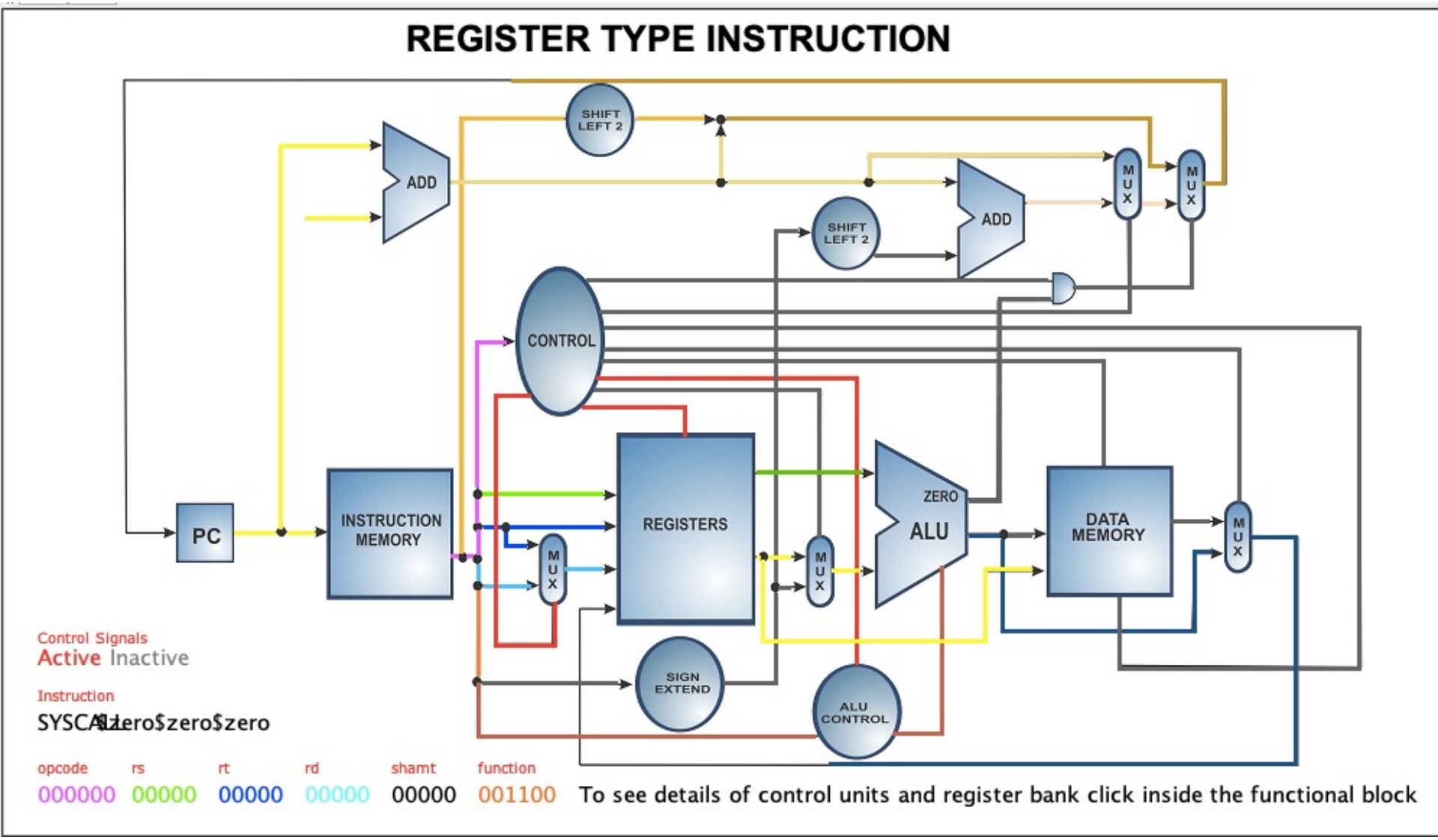


Register File on MARS

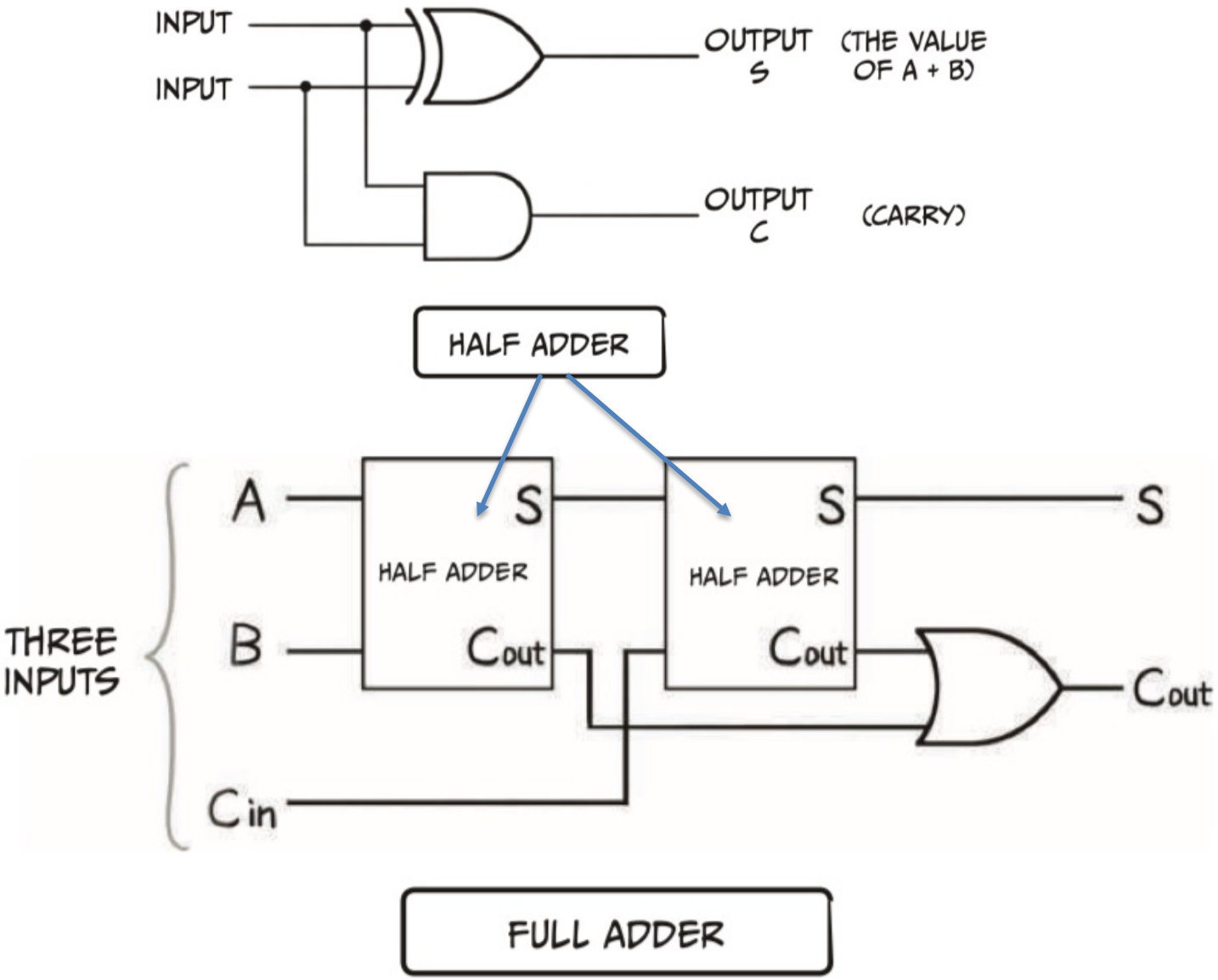
Decoders



MIPS on MARS



Adders



Multiplication

Quora

Jeff Drobman · Just now

multiplication is usually done completely in hardware, via a 2D array of "XY(i) + C" multiplier modules, whereby each row generates a partial product of the next signed digit of the multiplier times the multiplicand. shifting occurs in the hardware placement of each row. this array can also be pipelined, so multiple operations can be performed in sequential concurrency.

(See the 1971 **Am2505** 2x4-bit multiplier slice, and my personal MS thesis.)

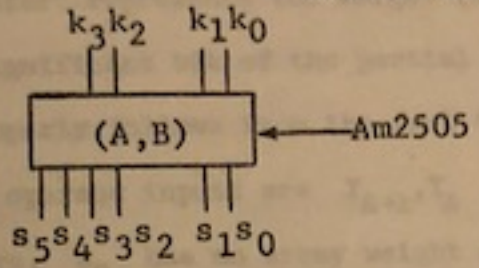
Am2505 Multiplier

Notation: (from [1], [15], [16])

multiplier pair = Y_{A+1}, Y_A

multiplicand group = $X_{B+3}, X_{B+2}, X_{B+1}, X_B$

"module designator" = (A,B)



2x4-bit slices

8-bit x 8-bit multiply

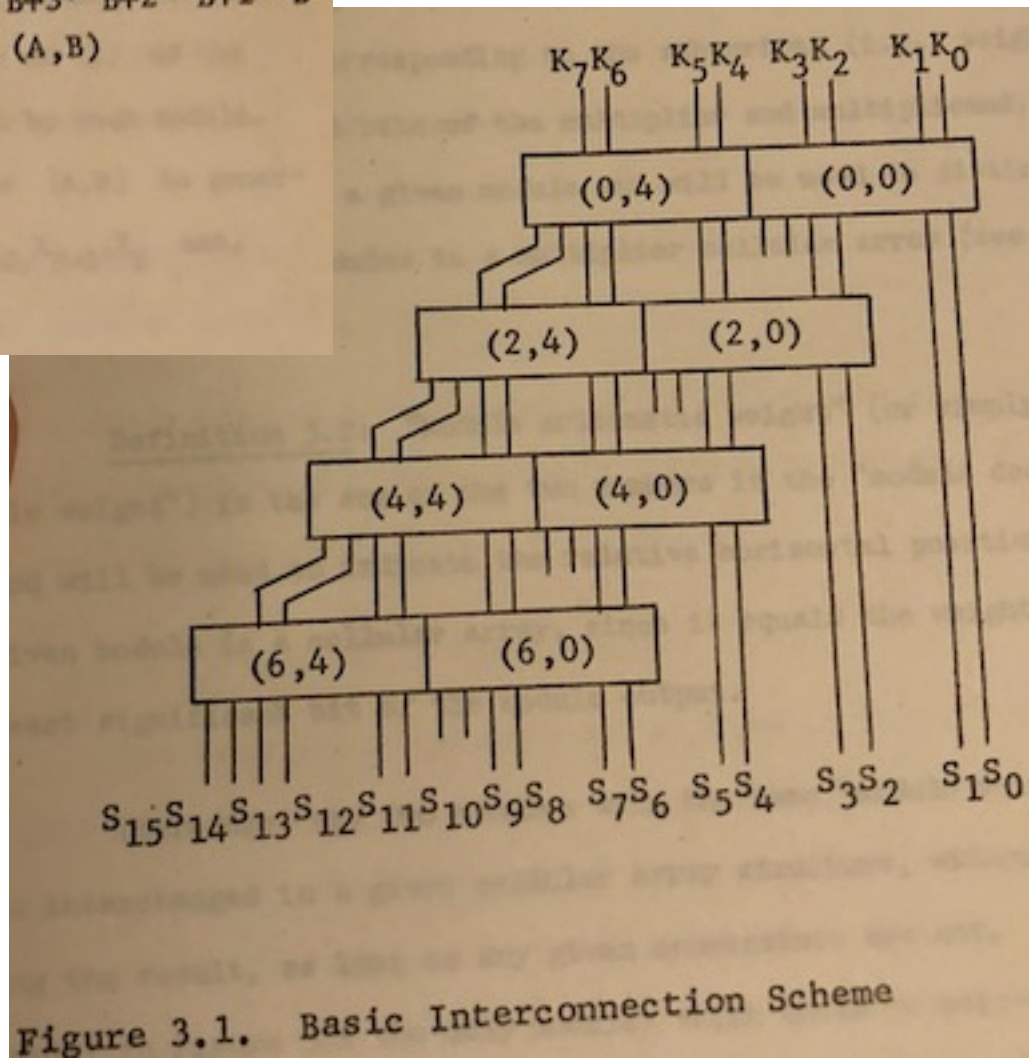


Figure 3.1. Basic Interconnection Scheme

Division

Non-Restoring Div

How do calculators calculate binary division?

Jeff Drobman, Lecturer at California State University, Northridge (2016-present)

Answered just now

the most common division algorithm used in the past was "non-restoring". but there are others, as listed in Wikipedia:

"Division algorithms fall into two main categories: slow division and fast division. Slow division algorithms produce one digit of the final quotient per iteration. Examples of slow division include restoring, non-performing restoring, non-restoring, and SRT division. Fast division methods start with a close approximation to the final quotient and produce twice as many digits of the final quotient on each iteration. Newton–Raphson and Goldschmidt algorithms fall into this category."

Section

VLSI vs ASIC vs FPGA

VLSI vs ASIC vs FPGA

- 3 options for chips

❖VLSI

- ***Fully*** Custom
  - *Building blocks*: Designer IP (all levels)
  - *Tools*: Licensed from EDA vendors

❖ASIC

- ***Semi*** Custom
  - *Building blocks*: Manufacturer IP
  - *Tools*: Provided by Mfr (ASIC vendor)

❖FPGA

- ***Programmable*** Custom (SRAM)
  - *Building blocks*: logic gates (NAND, NOR)
  - *Tools*: Lab Programmers, software

FPGA vs PLD

- **FPGA**
  - ***Field Programmable*** (SRAM based)
    - *Building blocks*: logic gates (NAND, NOR)
    - *Tools*: Lab Programmers, software
- **PLD**
  - **PLA**
    - AMD *Mach* family (merged with PAL)
  - **PAL**
    - MMI invented, bought by AMD
    - AMD spun off as Vantis (bought by Lattice)
  - **CPLD**
    - *Complex* PLD

Programmable Logic – FPGA

From Wikipedia, the free encyclopedia

FPGAs

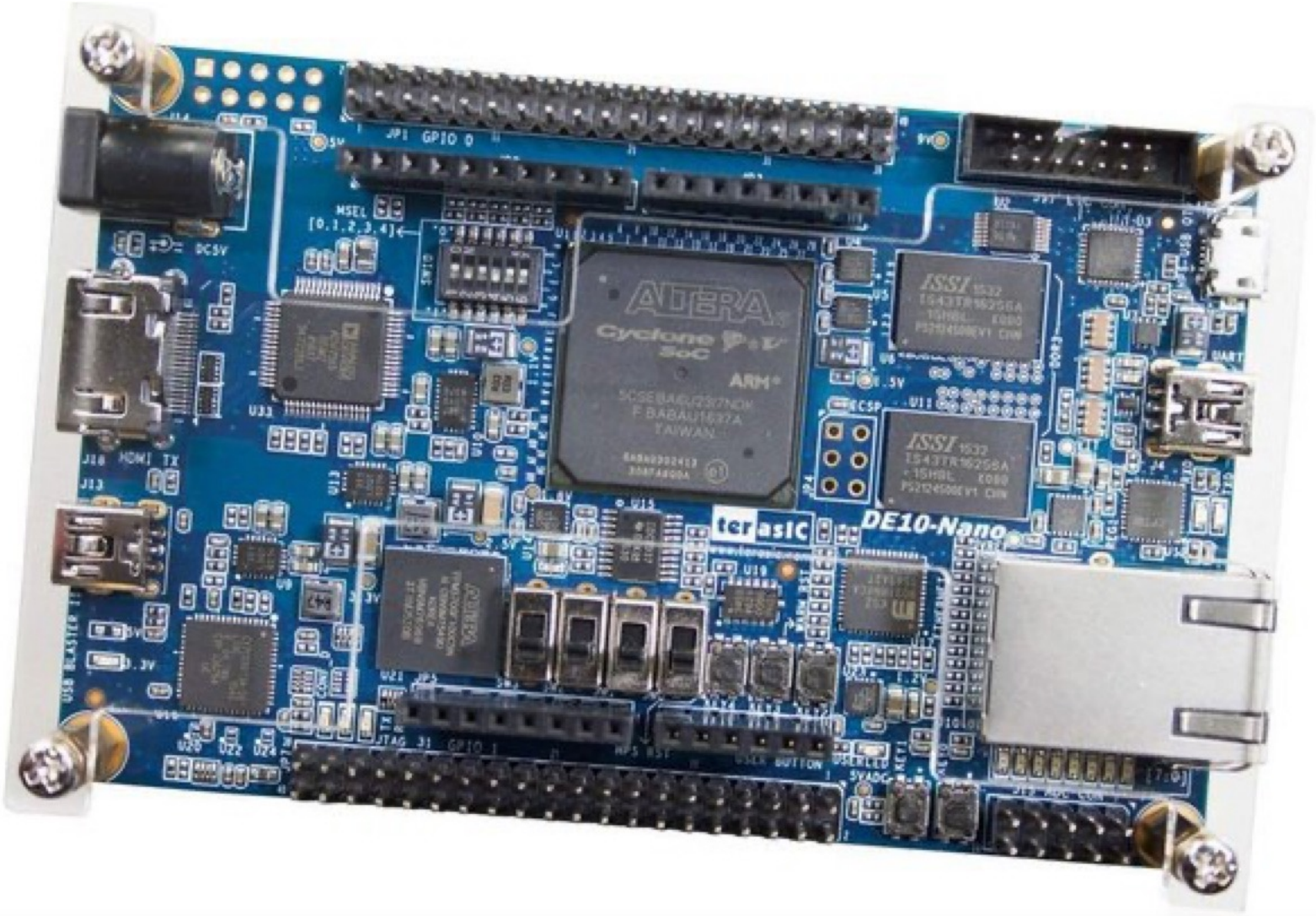
FPGA

A **field-programmable gate array** (**FPGA**) is an integrated circuit designed to be configured by a customer or a designer after manufacturing – hence the term *field-programmable*. The FPGA configuration is generally specified using a hardware description language (HDL), similar to that

HDL

Quora FPGA

For example this is generic gaming console built using single FPGA:



CPU/GPU vs ASIC

Quora

What is the difference between ASICs and GPUs/CPUs?

Jeff Drobman
Lecturer at California State University, Northridge (2016–present) · Just now

in general, ASICs use smaller functional building blocks than “cores”, which are much larger, complete units. ASIC’s are “application specific” — to a single application, and are not *programmable*.

Whereas CPU and GPU cores are software *programmable* — so *general* purpose (not application specific).

Section

Chip Design

Transistors

Making Transistors

Doping **P** & **N**

\+ -

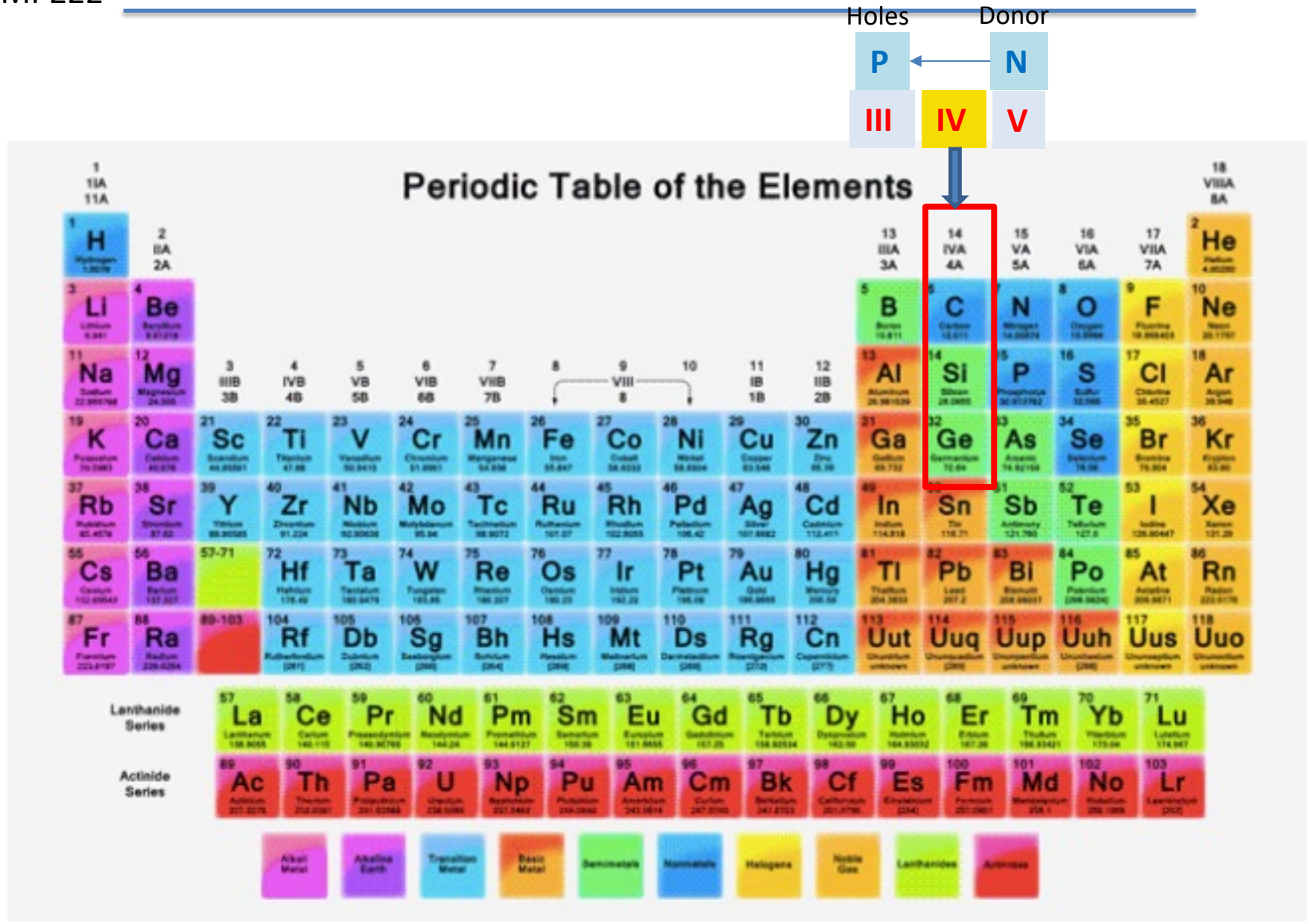
How does impurity mix with a material to form positive and negative semiconductor material?

Jeff Drobman
Lecturer at California State University, Northridge (2016–present) · Just now ·

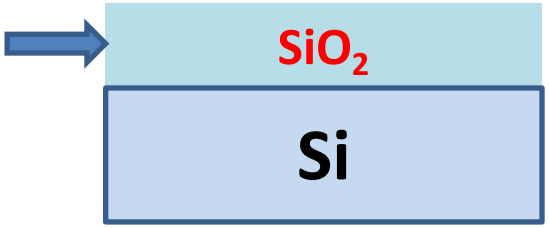
P and N dopants (III and V valence) are chemically *diffused* as gases into the silicon substrate in ovens. sometimes *ion implantation* is used instead.

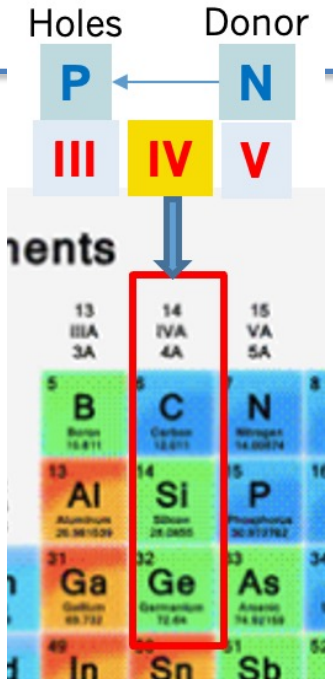
CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP222

Silicon Semi



Silicon Semi





Quora

Chris Bevis · Follow
Lives in Silicon Valley (1997–present) · Updated Sep 9

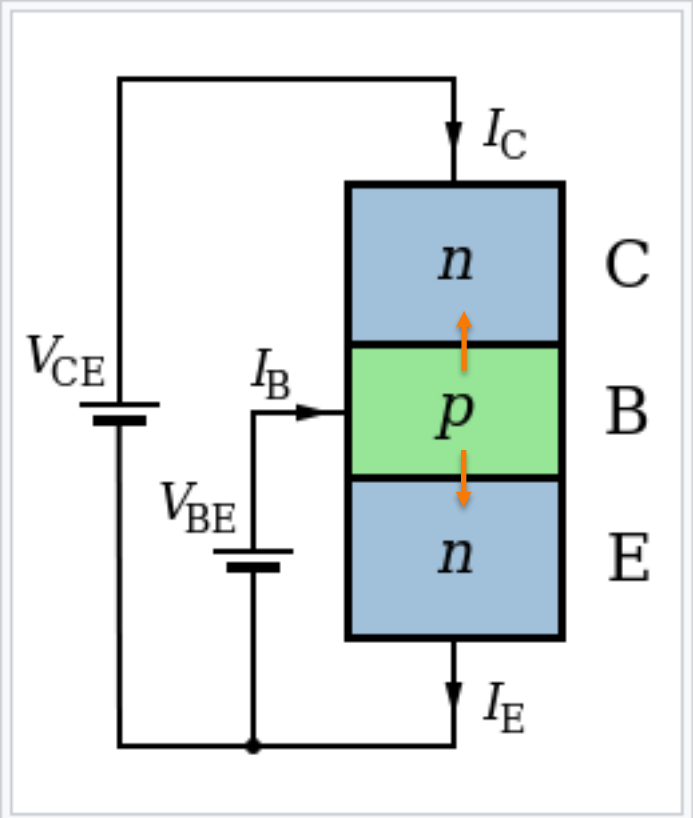
Related **Why is silicon mostly used in tech companies than germanium?**

The earliest semiconductor devices were made of germanium. This started to change when the first planar IC's were developed. It is easy to grow a stable, insulating oxide film on silicon but not on germanium, so silicon was the material of choice. One of the key breakthroughs in silicon processing was made by Andy Grove and a colleague: the Grove-Deal model of SiO2 growth kinetics.

These days, most advanced devices have channels made of an alloy of Si and Ge doped with boron (SiGeB). The percentage of Si and Ge is carefully controlled and varied over the height of the channel to produce strain which enhances the mobility of electrons or holes.

Bipolar Transistors

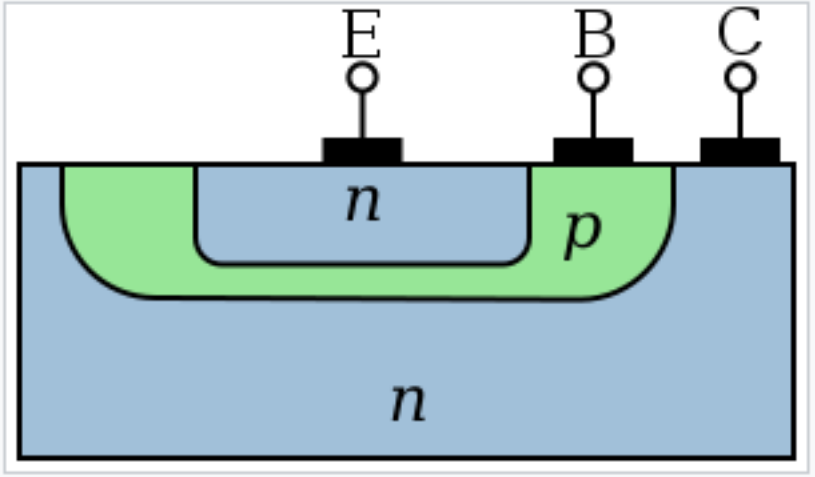
BJT



Structure and use of NPN transistor. Arrow according to schematic.

PLANAR

Structure [edit]



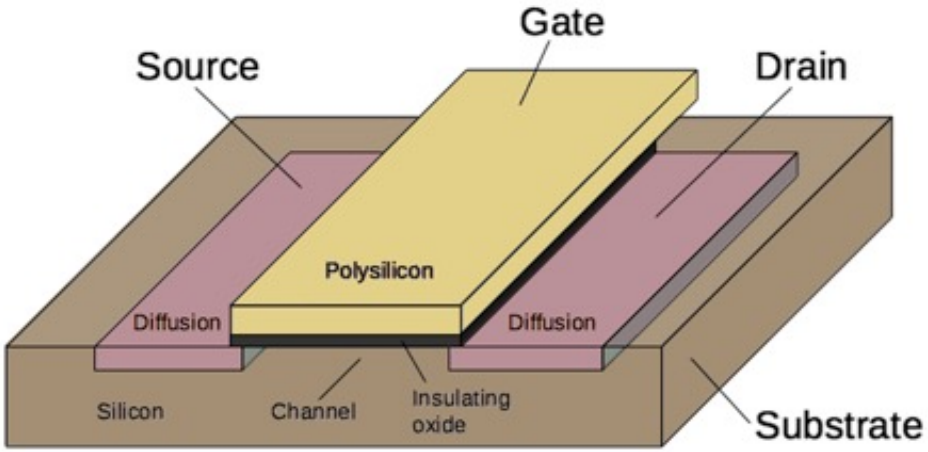
Simplified cross section of a planar *NPN* bipolar junction transistor

Physical Level: MOSFET

Device/Xtor
Physical
Level

Inverter/Gates

Digital: MOSFET



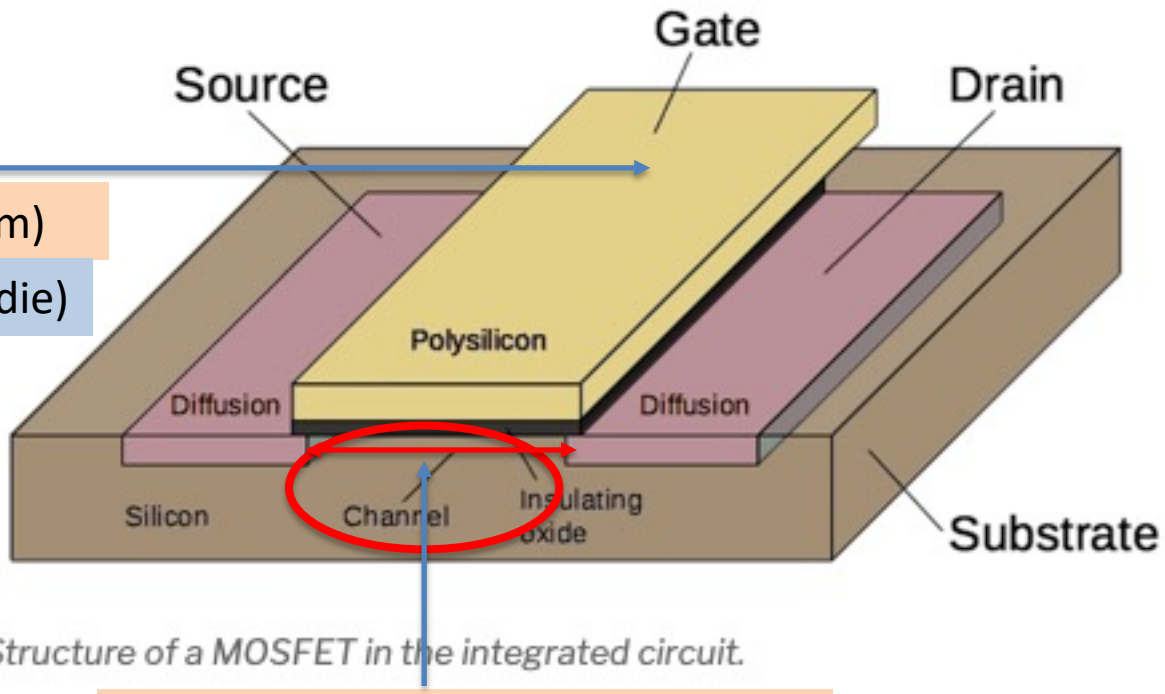
Structure of a MOSFET in the integrated circuit.

(see separate slide set ***Transistors***)

MOS Transistor

$V_g = 5 \rightarrow 3.3 \rightarrow 1.8$ V

Gate



Structure of a MOSFET in the integrated circuit.

Pitch (e.g.14nm)

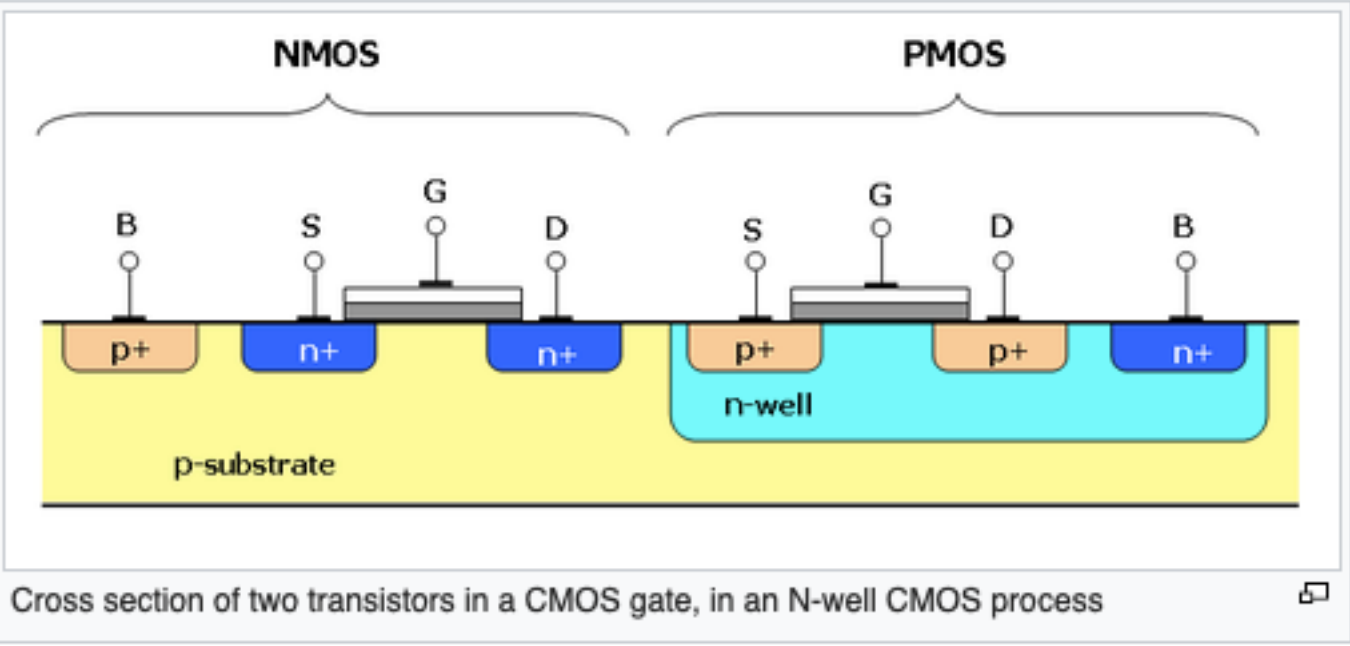
Density (xtor/die)

Smallest feature size (e.g. 5nm)

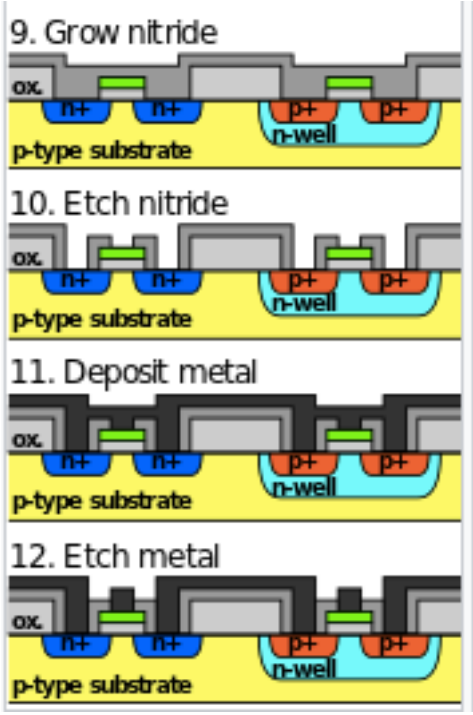
Performance (F_t)

CMOS Transistors

MOSFET



Cross section of two transistors in a CMOS gate, in an N-well CMOS process



Last 4 steps

IC Process & Interface

TTL compatible — **5V → 3V**

❖ *Bipolar*

5V

➢ RTL→DTL→TTL → *Schottky* TTL→ *LS* TTL

1960	1965	1970	1975	1980	1985	1990

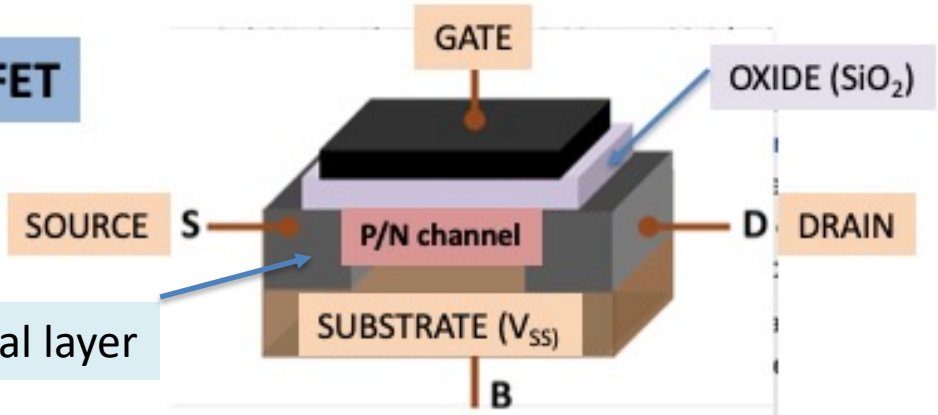
❖ *MOS*

➢ PMOS→NMOS → **CMOS** → CMOS (TTL I/O)→ 3.3V

5V → 3V

BiCMOS

MOSFET



Epitaxial layer

The **metal–oxide–semiconductor field-effect transistor**, also known as the **metal–oxide–silicon transistor**, is a type of field-effect transistor that is fabricated by the controlled oxidation of a semiconductor, typically silicon. It has an insulated gate, whose voltage determines the conductiv

Section

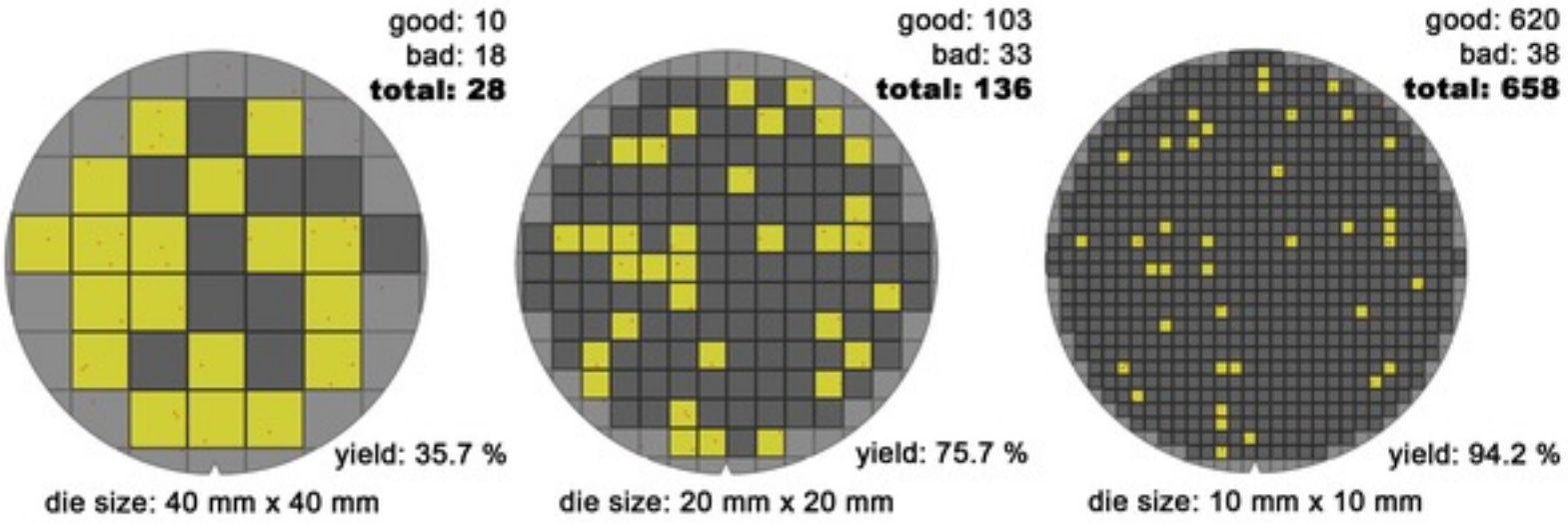
Chip Design

Fab

Chip Specs

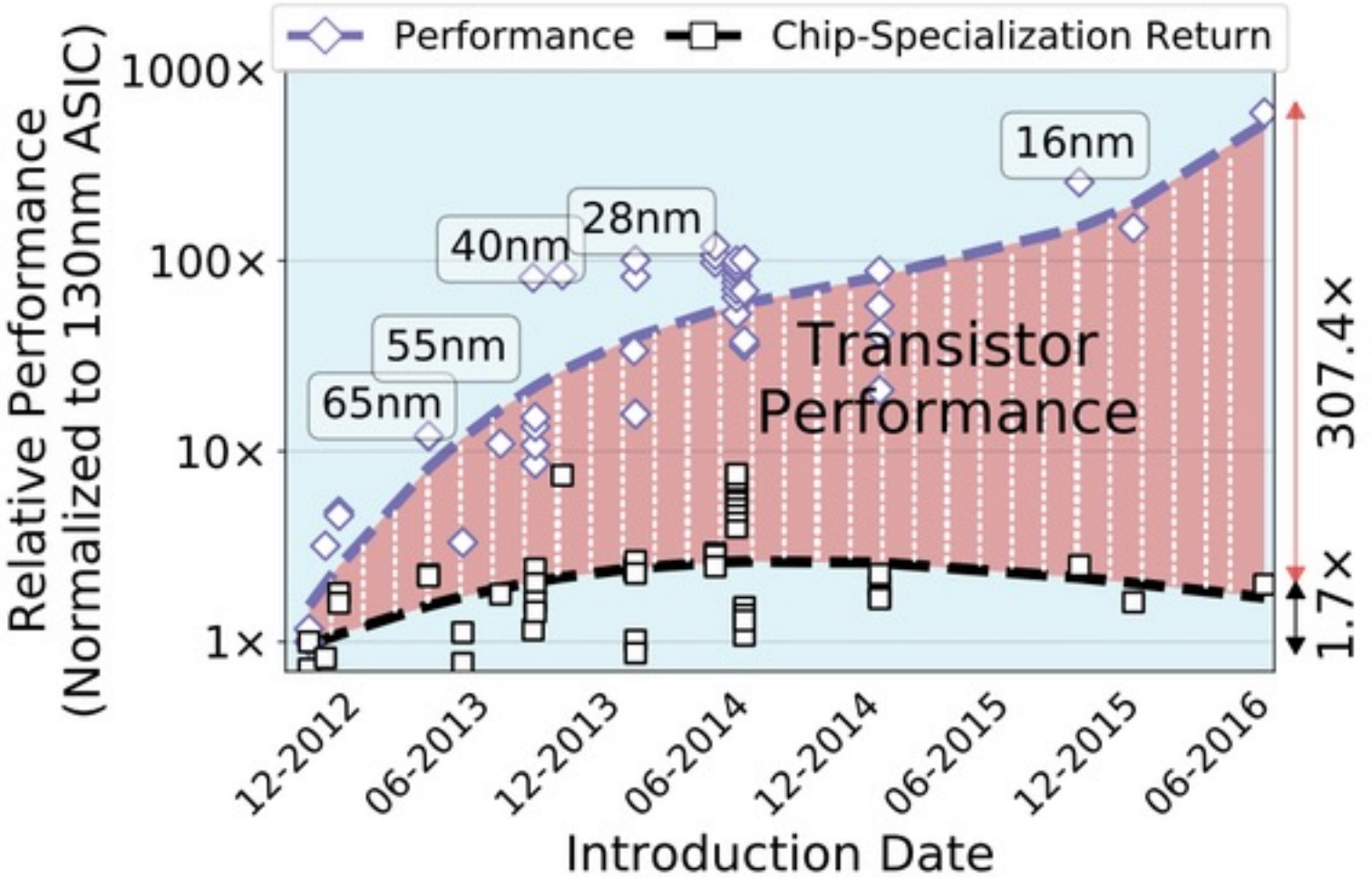
- Architectural
- Functional
- Mechanical
- Electrical (DC)
- Timing (AC)
- Thermal (theta JA, JC, CA)

Wafers: Yield



yellow shows bad dice (a function of defect density)

Chip Specializations



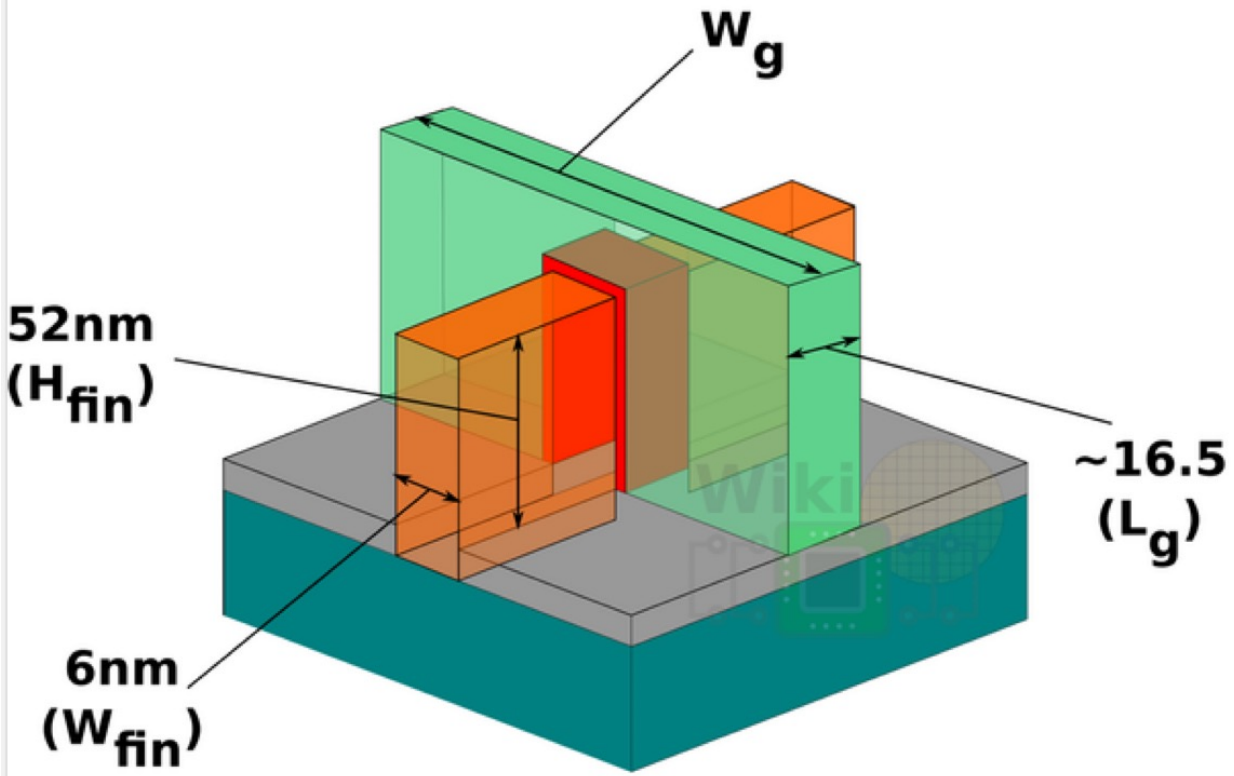
On-chip specialization or microarchitctural changes merely account for a fraction of the performance gains.

Chip Fab

7 nm TSMC

What's a 5nm FinFET transistor look like?

Well... I didn't find a good pic for 5nm, but I found a pic for 7nm here:

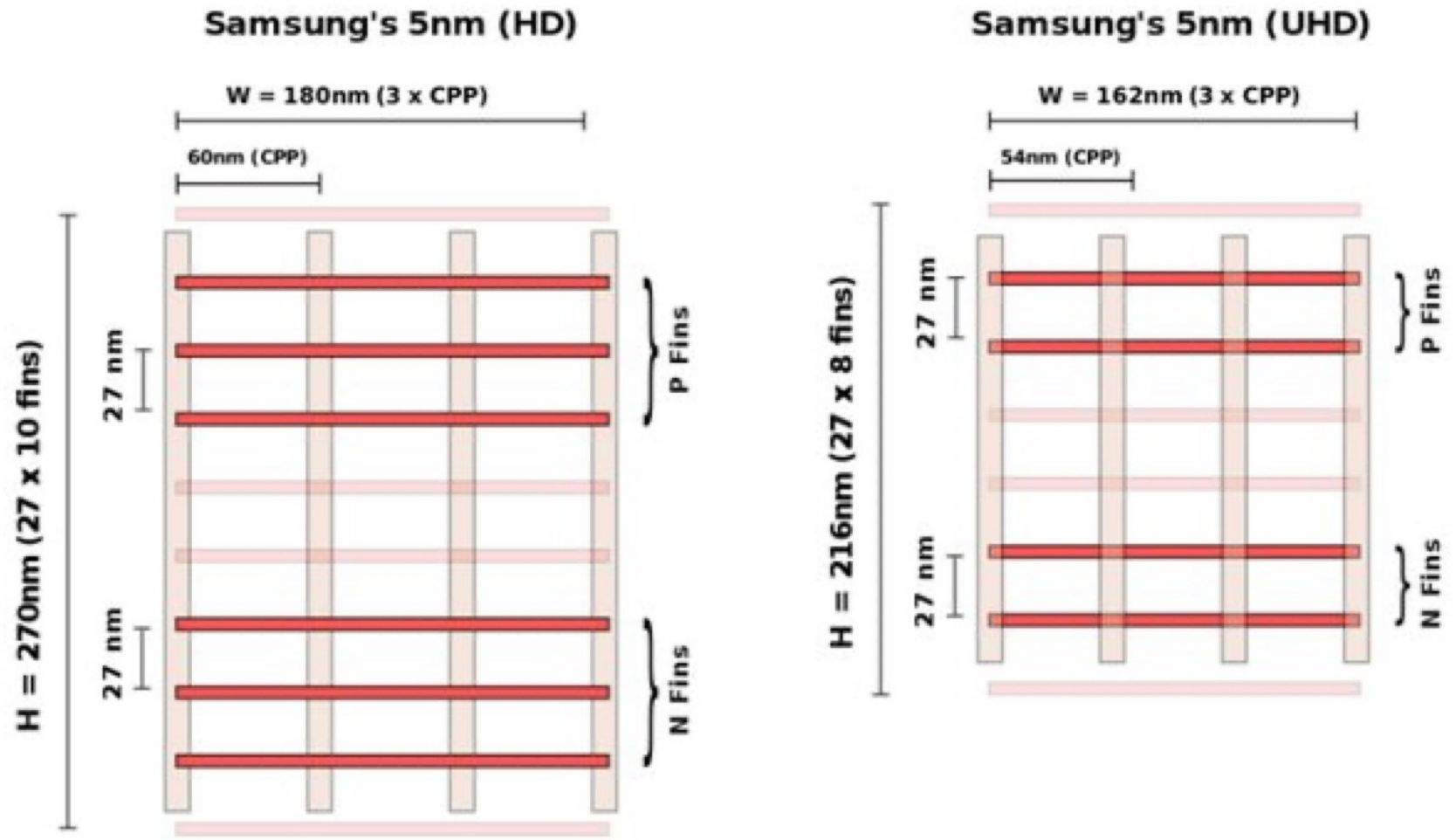


We can assume 5nm is slightly smaller on all axes.

The first thing to note is that the gate length (L_g) is 16.5nm. The width of the gate will vary, but I am going to go out on a limb and guess it is no smaller than $3W_{\text{fin}}$, or 18nm. And the overall structure is tall: 52nm plus another few nm

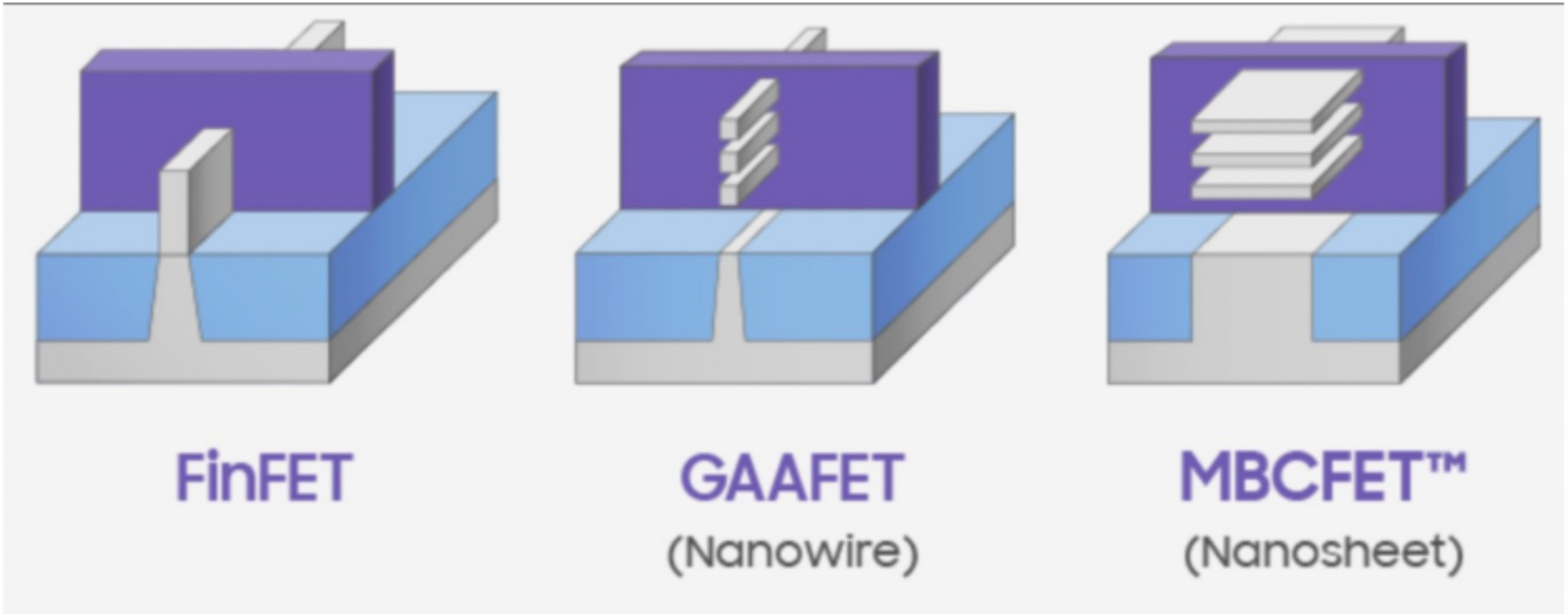
Making FinFET Transistors

CMOS = P & N



Above are CMOS gates having N and P transistors, roughly 5nm node has transistor size of 130 x 90 nm. Far, far, far away from 5nm name.

New Processes



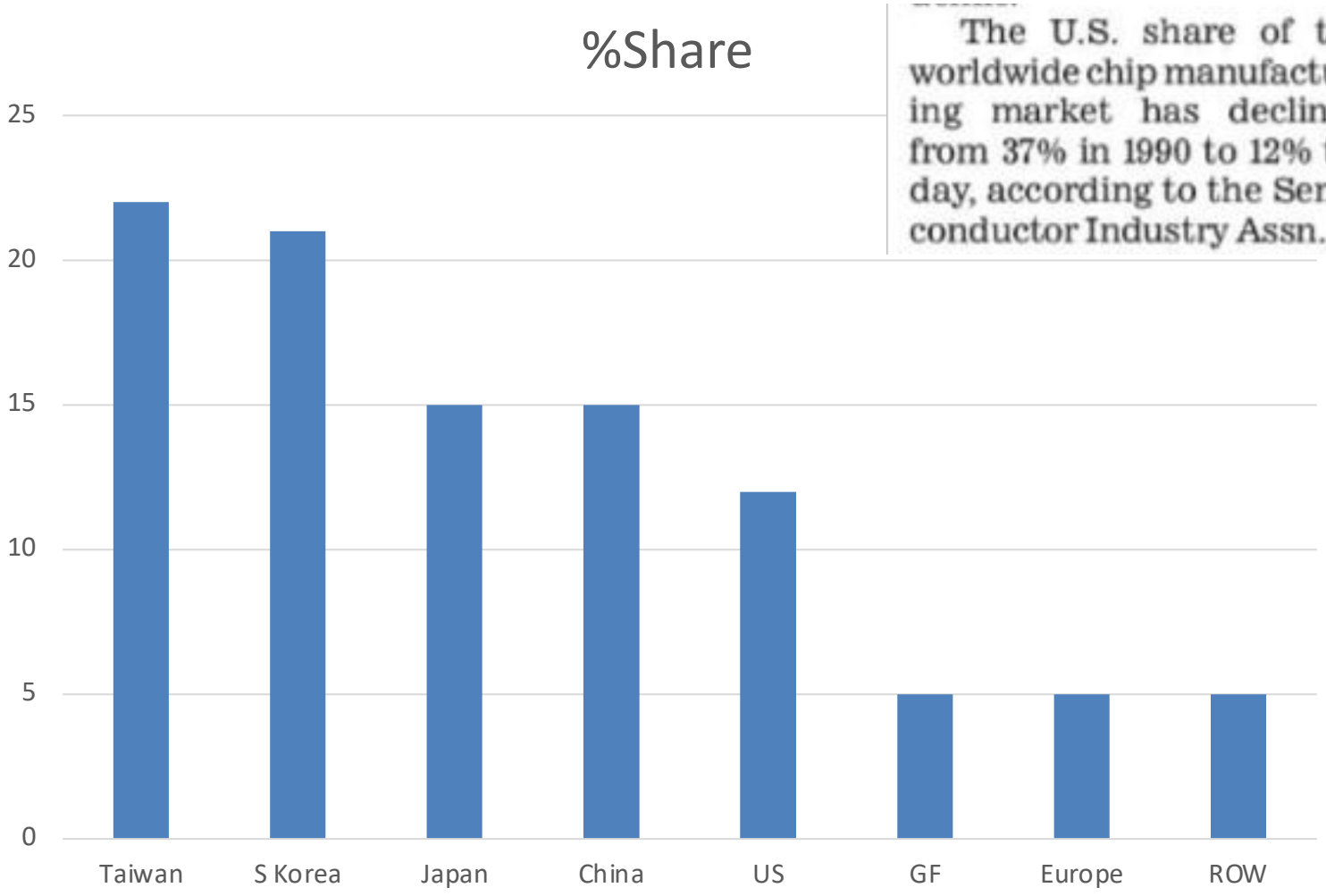
- It is all about the **Gate**

Wafer Fabs Today

1968	❖ Intel
1978	❖ Micron**
1980	❖ Samsung
1987	❖ TSMC* (1st foundry)
2009	❖ AMD, IBM → Global Foundries*
2010	❖ Chartered → Global Foundries*
	❖ SMIC (China)

*Pure Foundry
**Internal use only

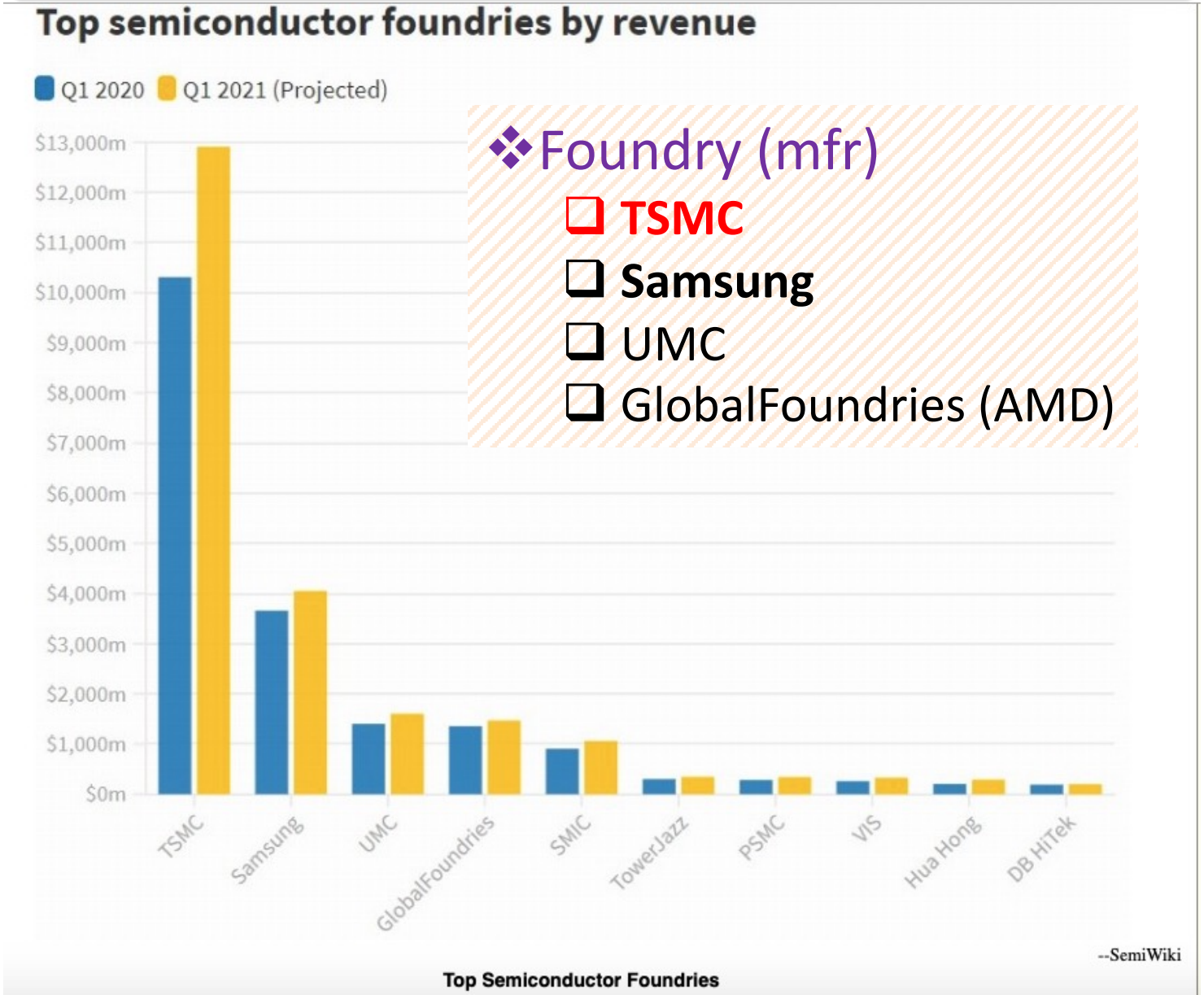
WW Fab Shares



The U.S. share of the worldwide chip manufacturing market has declined from 37% in 1990 to 12% today, according to the Semiconductor Industry Assn.

Source: LA Times/SIA 1/22/22

Foundry Sizes 2020-1

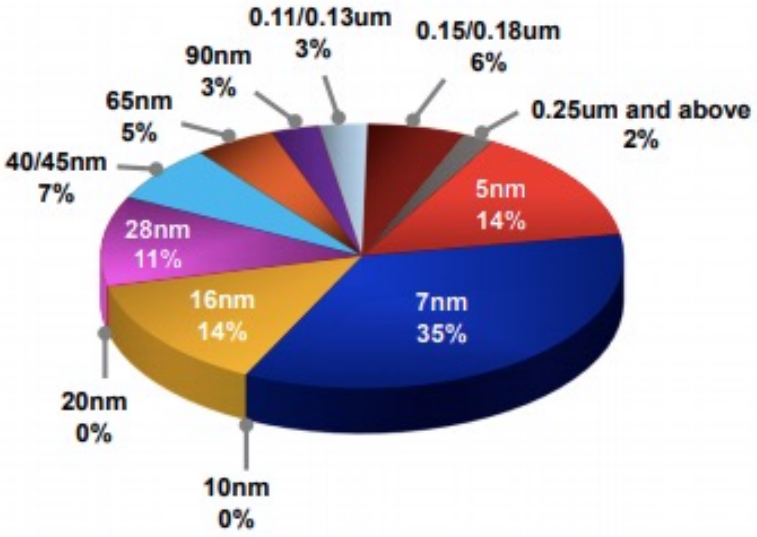


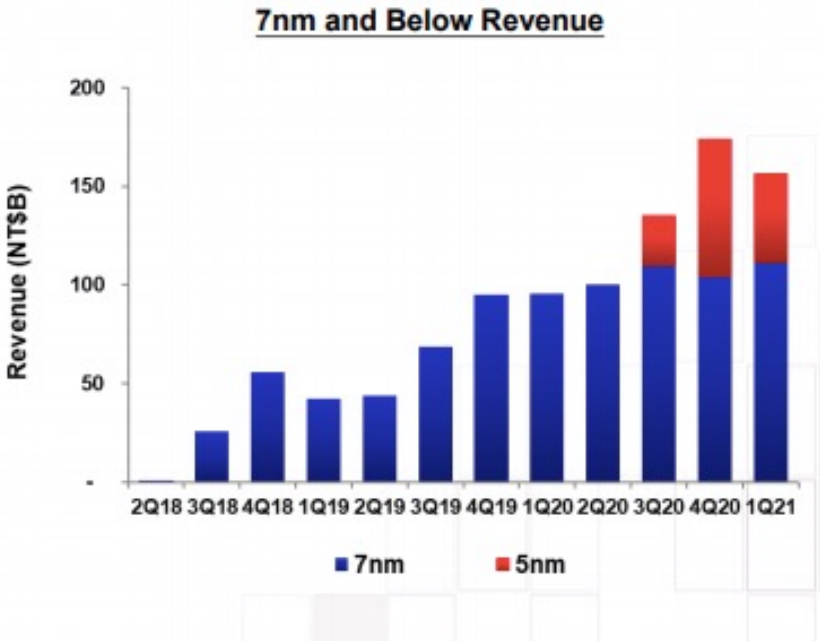
- Foundry (mfr)
  - **TSMC**
  - **Samsung**
  - UMC
  - GlobalFoundries (AMD)

TSMC

Unleash Innovation

1Q21 Revenue by Technology





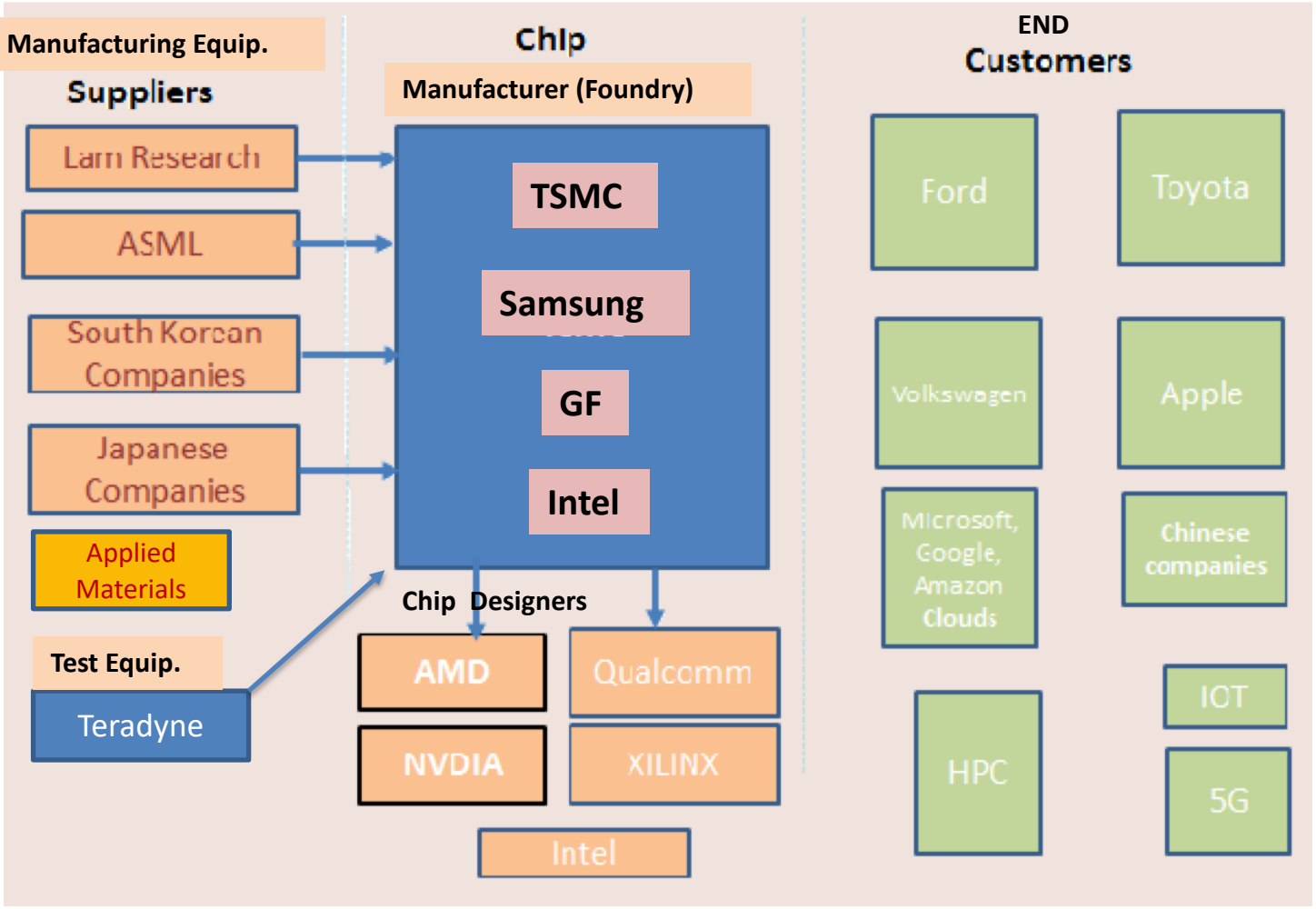
TSMC Customers

Table 4 – TSMC Customer Share of Revenues 2019-2021			
	2019	2020	2021
Apple	24.0%	24.2%	25.4%
Hi-Silicon	15.0%	12.8%	0.0%
Qualcomm	6.1%	9.8%	7.6%
NVIDIA	7.6%	7.7%	5.8%
Broadcom	7.7%	7.6%	8.1%
AMD	4.0%	7.3%	9.2%
Intel	5.2%	6.0%	7.2%
Mediatek	4.3%	5.9%	8.2%
Source: The Information Network (www.theinformationnet.com)			

Foundry Supply Chain

Quora

Harvey King
born in Taiwan, grew up in US, family rooted in mainland China for centuries.

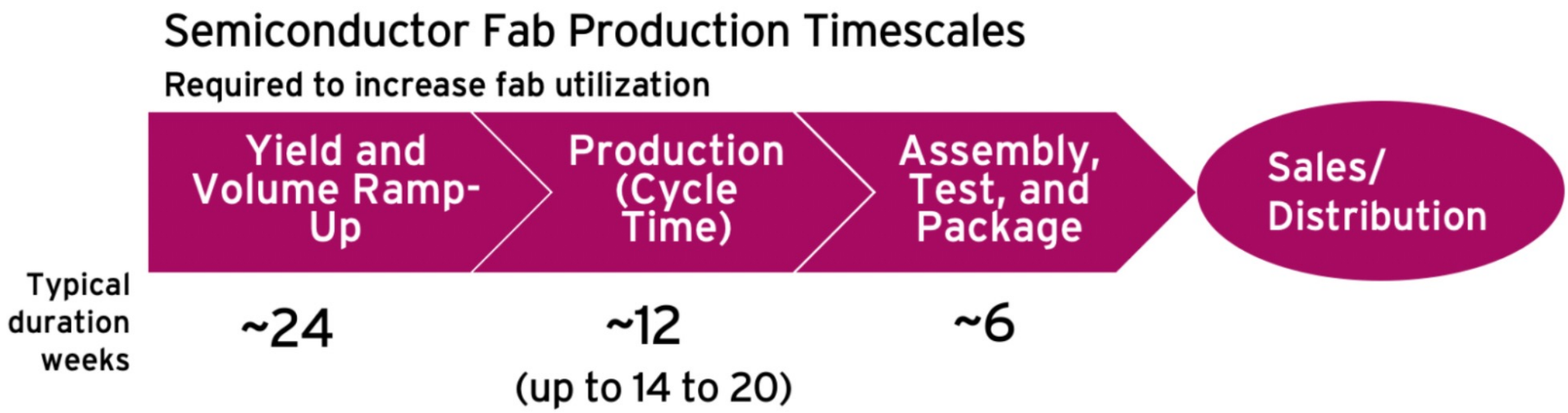


CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

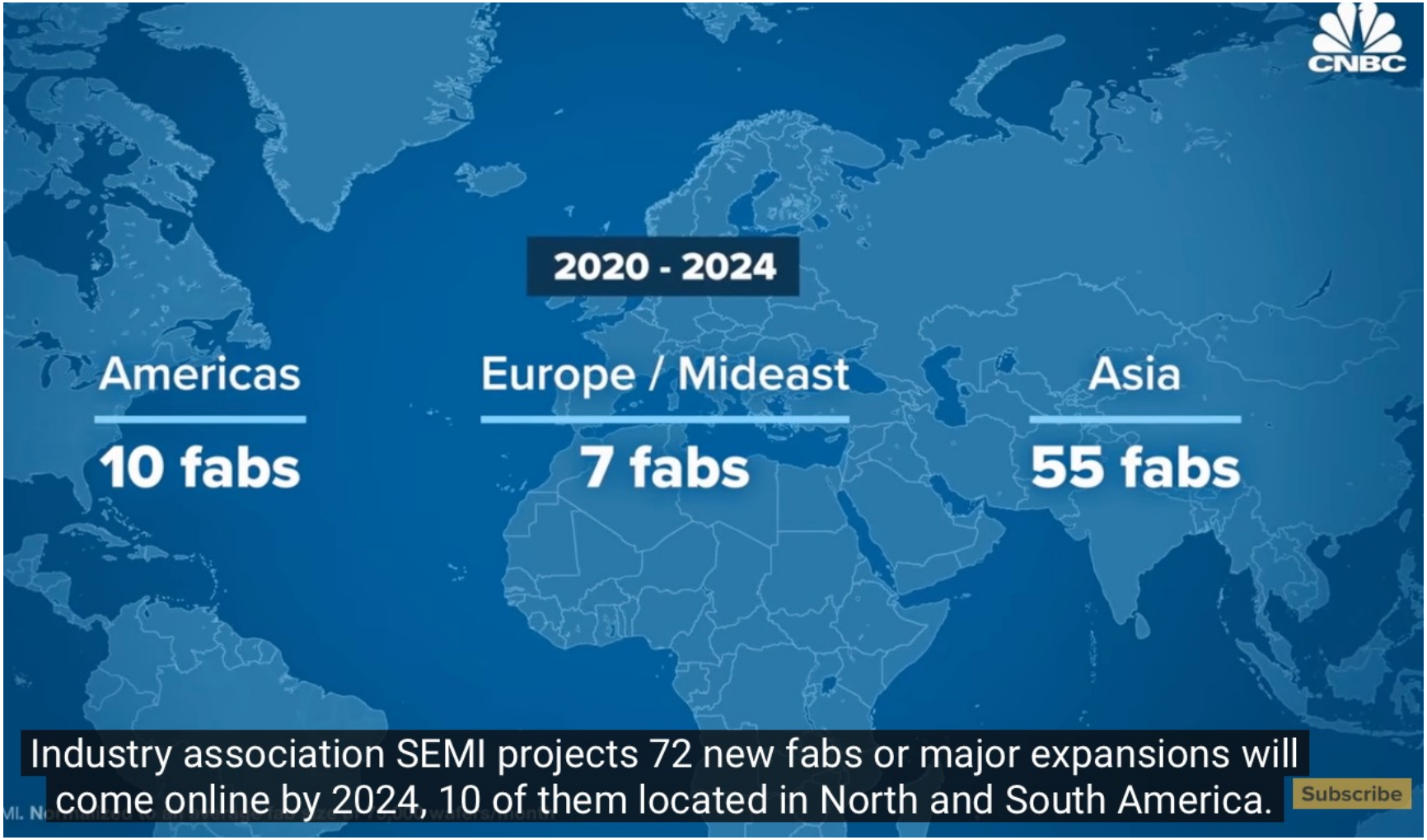
COMP222

Fab Timescales

SIA SEMICONDUCTOR INDUSTRY ASSOCIATION



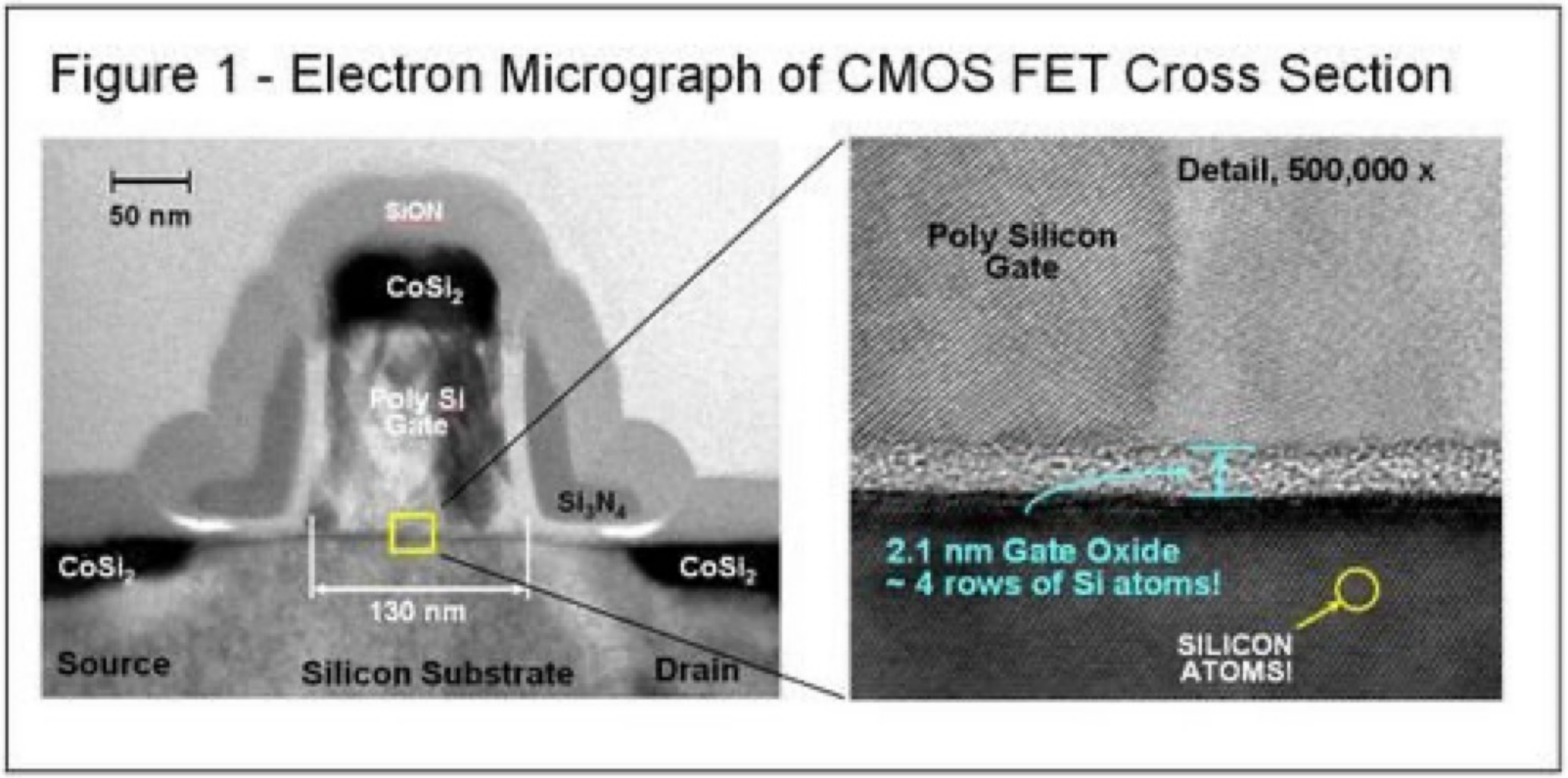
Global 72 New Fabs



Metal Interconnect

Quora

Aluminum → Copper → Cobalt

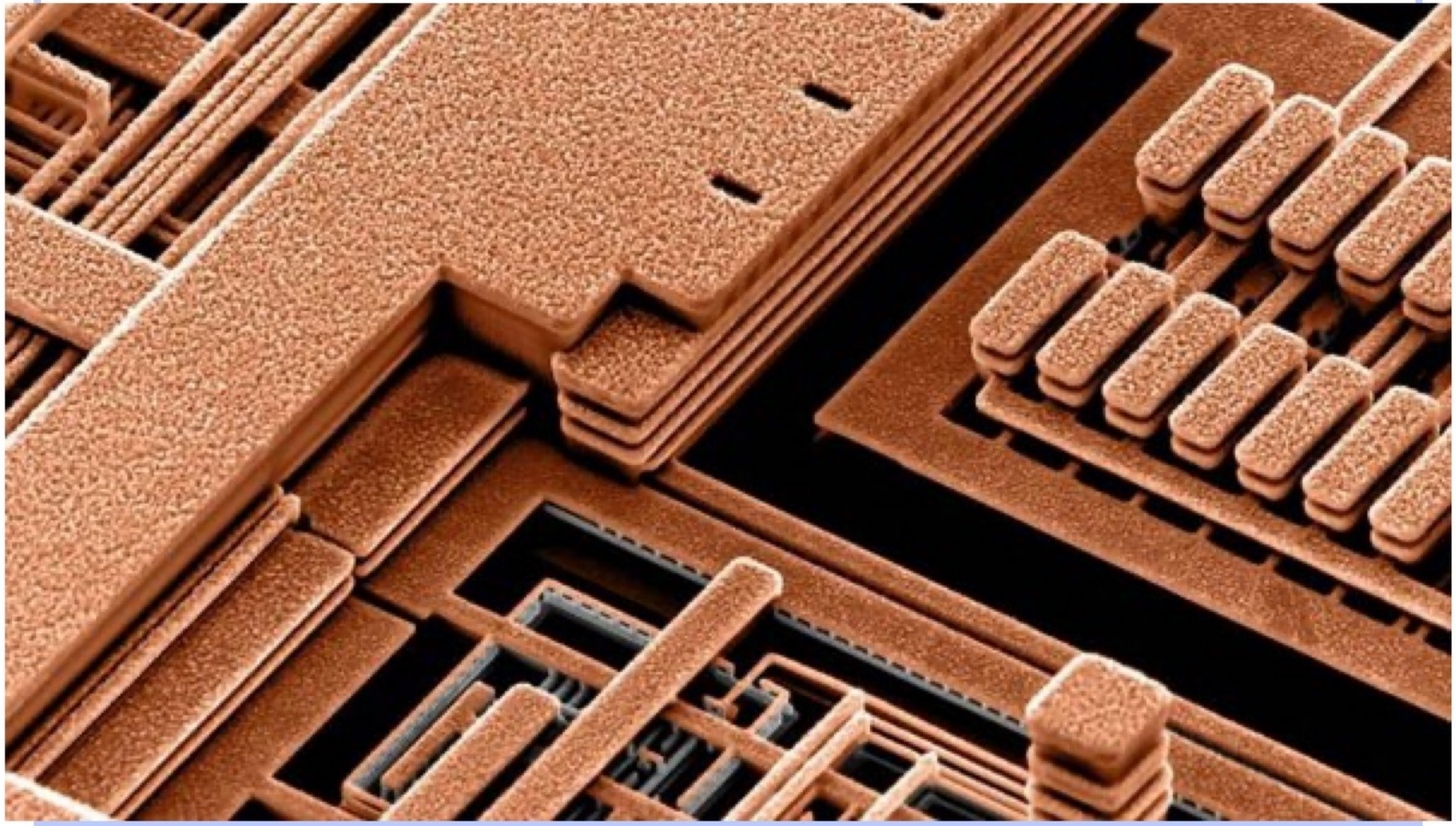


Metal Interconnect

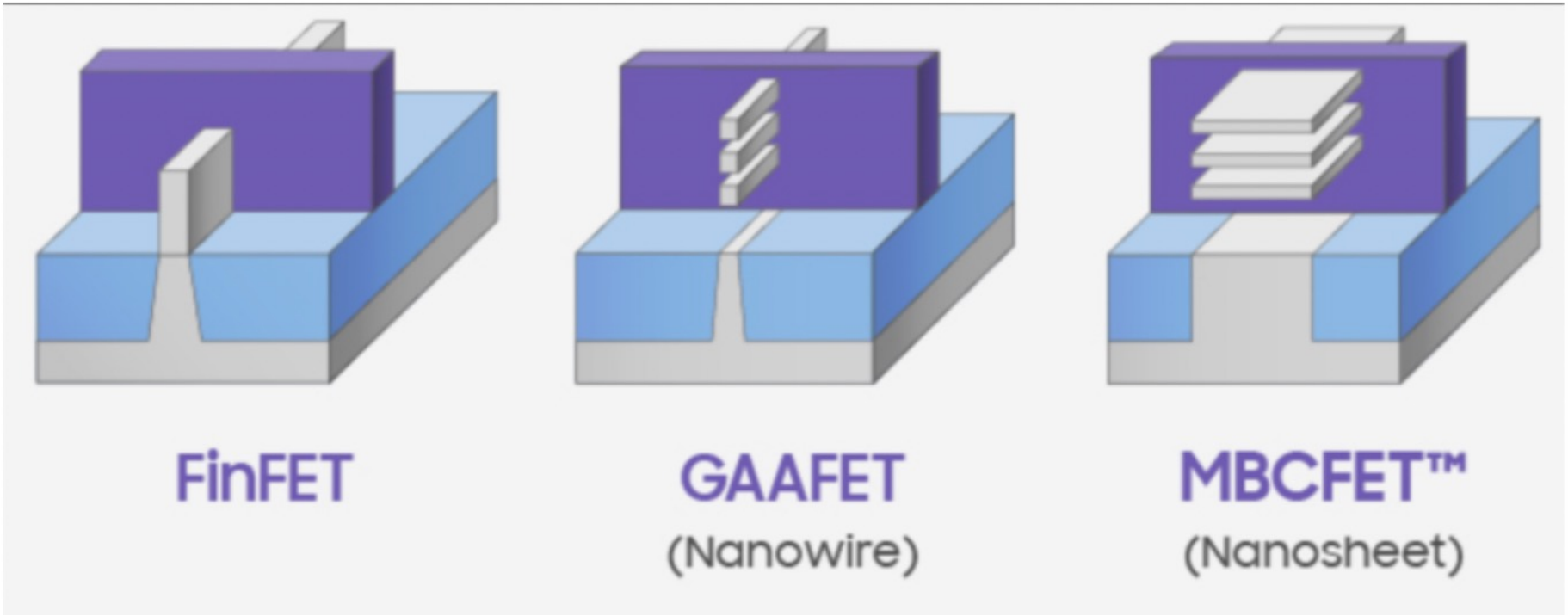
Quora

Aluminum → Copper → Cobalt

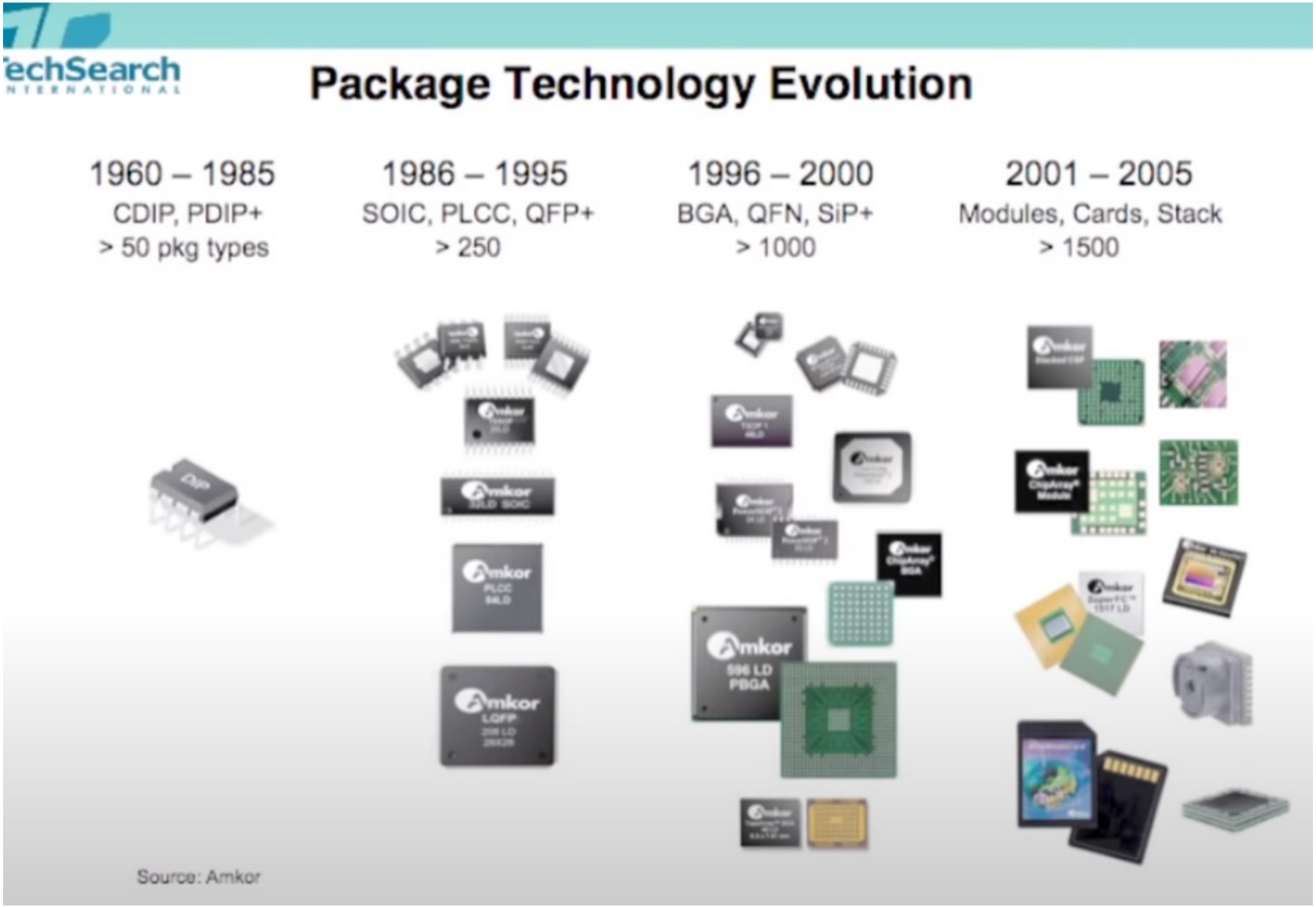
The transistors connect to each other using metal wires called interconnects. There are several layers of them (more then 8). Usually the space between them is filled with an insulator to provide structural rigidity, and expel air which can expand when heated and damage the chip, but IBM very kindly took the insulator out of one of their chips to show what the interconnect looks like:



New Processes



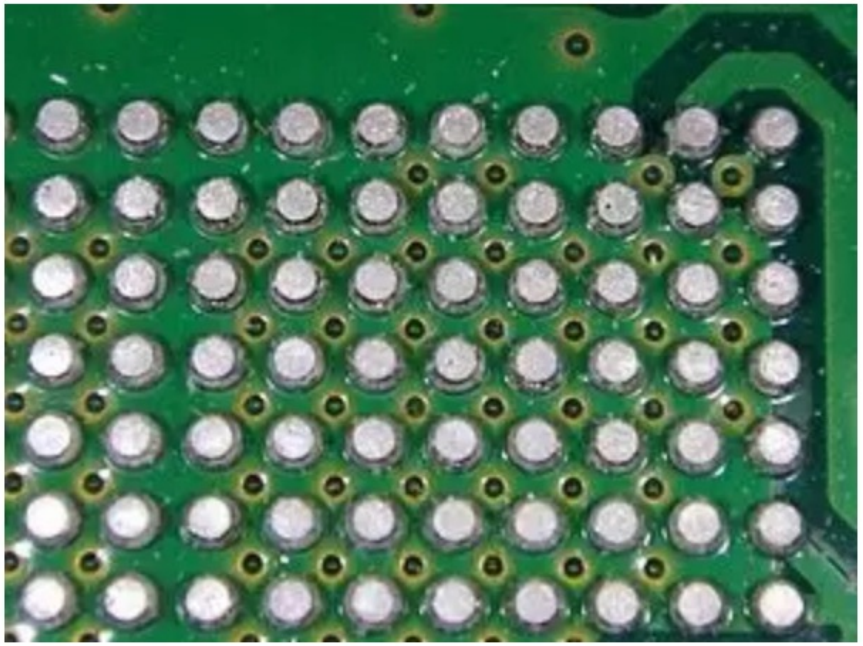
Chip Packaging



Chip Packaging

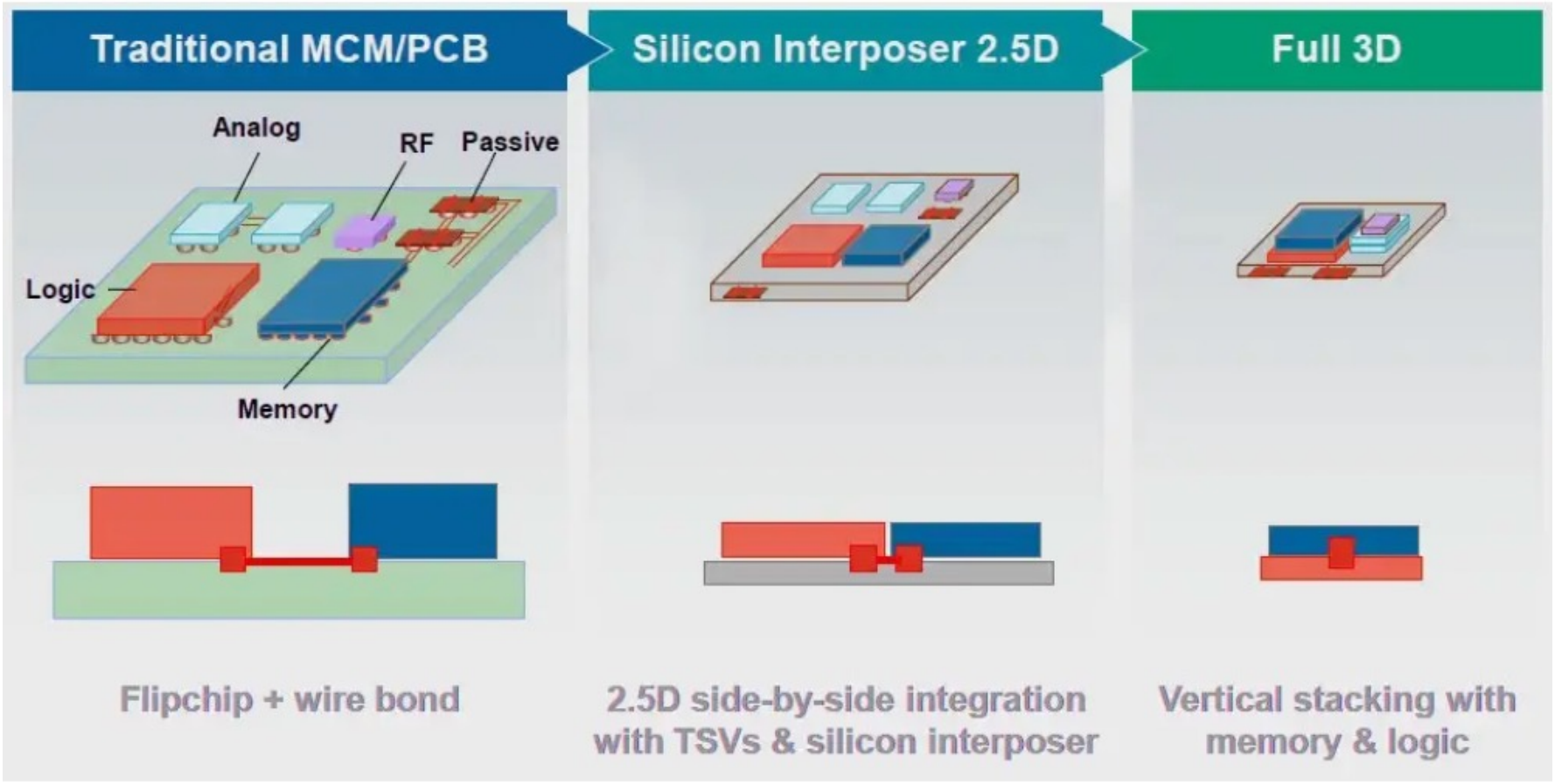
Ball Grid Packaging and Chip Scale Packaging (1990s – 2000s)

As the demands of semiconductor speed continue to pick up, so does the need for better packaging. While QFN (quad-flat no-leads) and other Surface Mounted technologies clearly continue to proliferate, I want to introduce you to the beginning of a package design that we will have to know about in the future. This is the beginning of the solder balls – or broadly Ball Grid Array (BGA) packaging.



Those balls or bumps are called solder bumps/balls

Chip Packaging (MCM)



Section

CPU Cores

CPU Function

What is the most purely mathematical description you could give of the workings of a CPU?

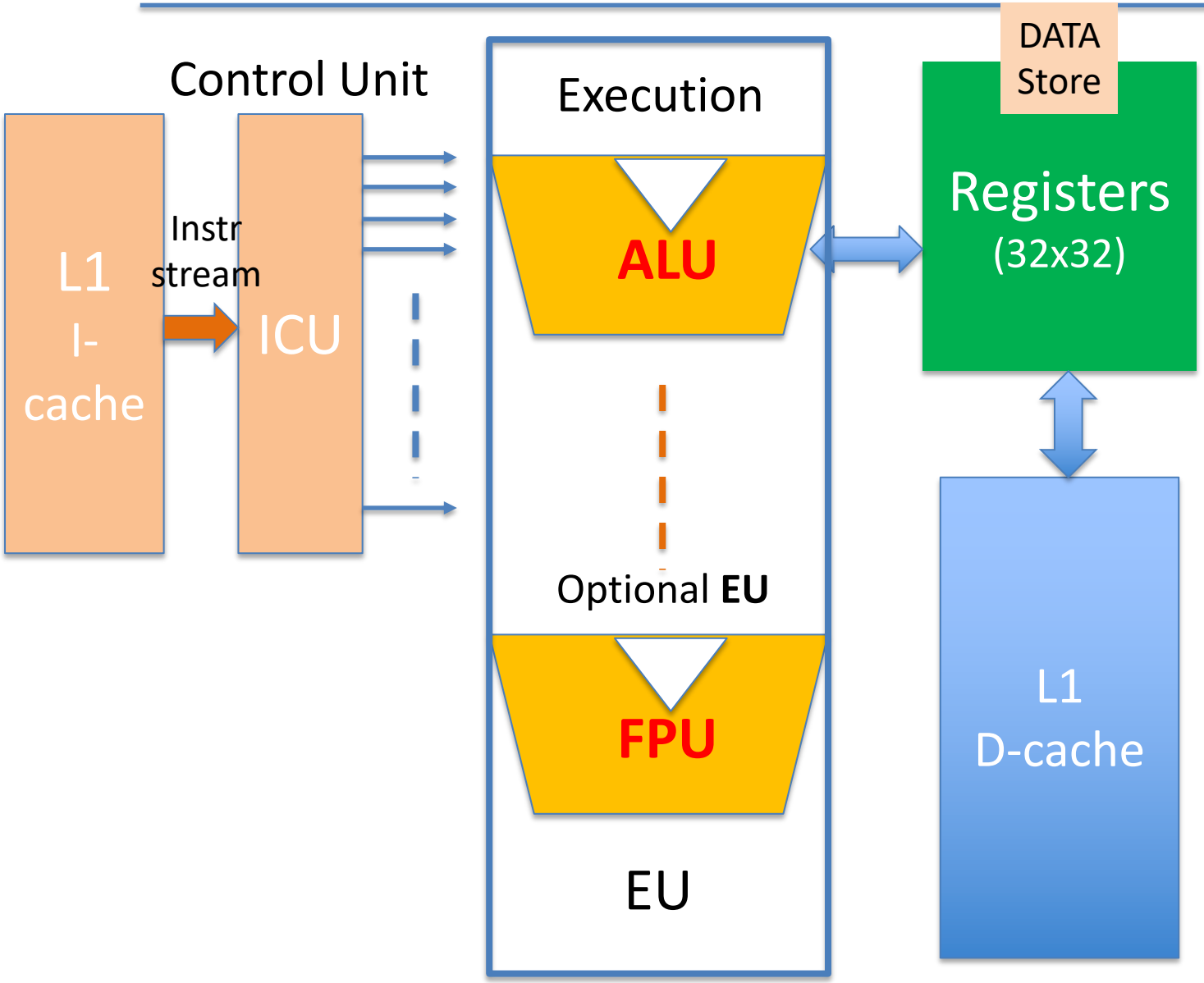
Jeff Drobman
Lecturer at California State University, Northridge (2016–present) · Just now

CPU_state = function(instruction, last_state)

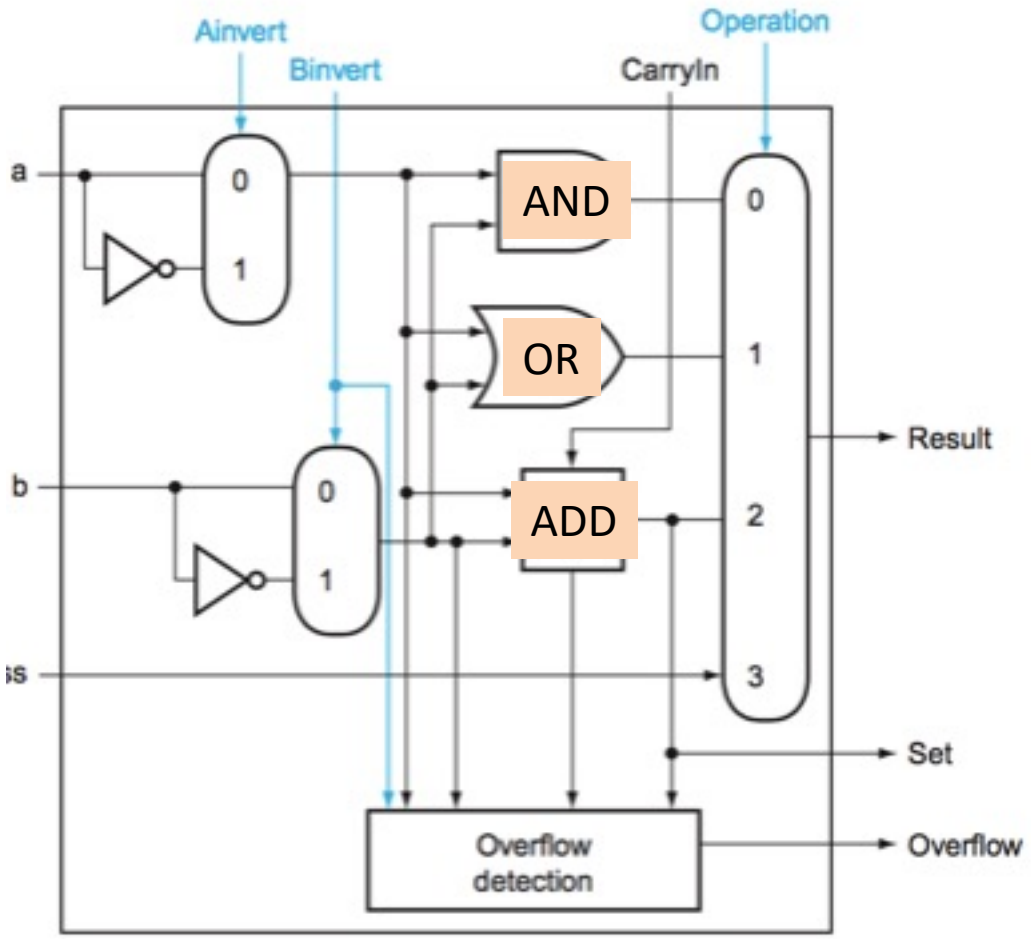
FSM definition

CPU_state = Function(instruction, last_state)

CPU = Data + Control



MIPS/MARS ALU



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

4 Levels of CPU *Architecture*

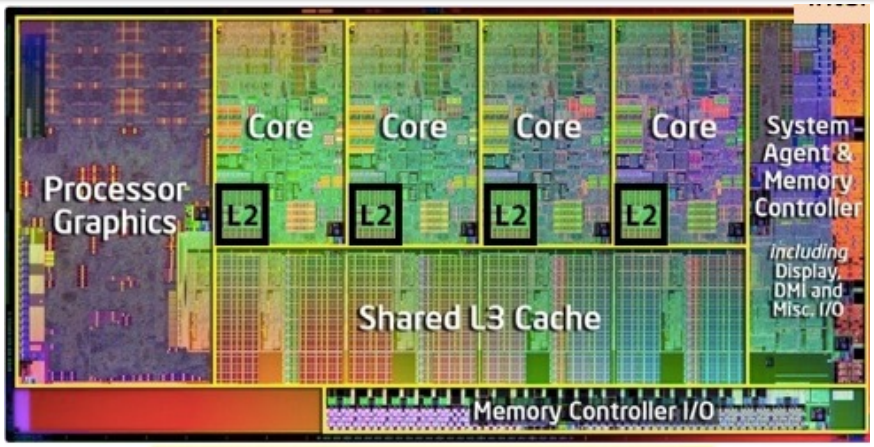
DR JEFF SOFTWARE
INDIE APP DEVELOPER

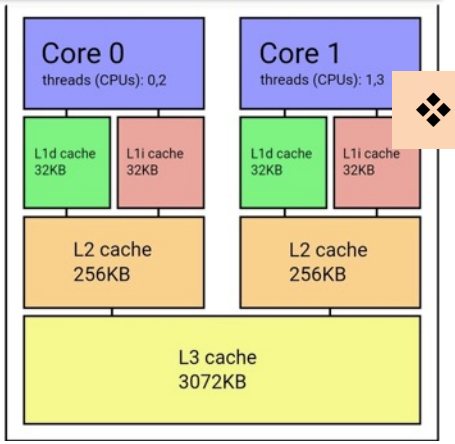
COMP222

COMP122/222

❖ Macro

System= *Multi-core SoC*



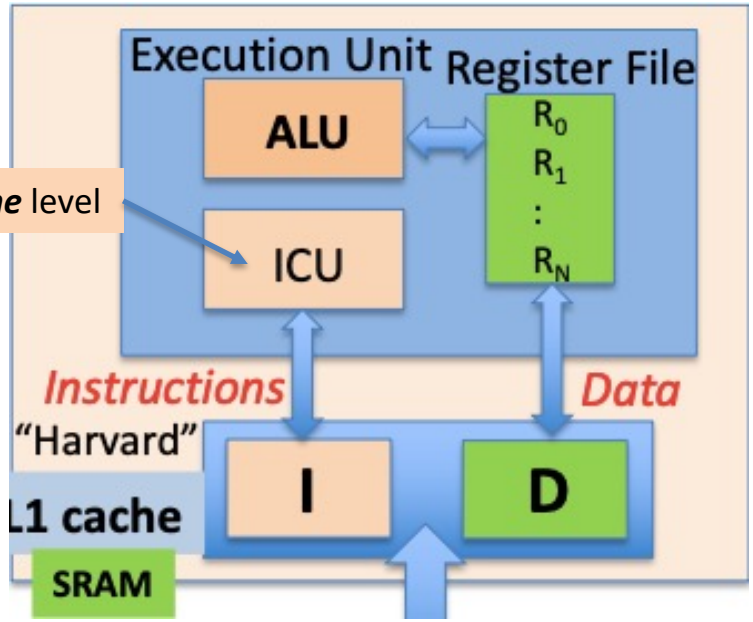


❖ Floorplan

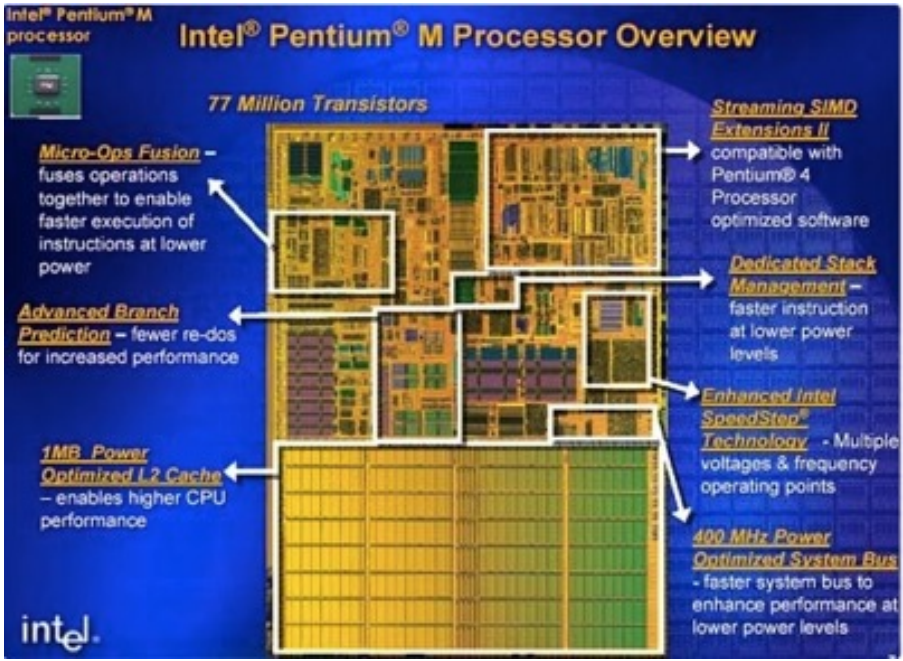
COMP122

❖ **Org + ISA** ***CPU Core*** internals

Pipeline level



COMP222 ❖ Micro ***Pipeline*** level



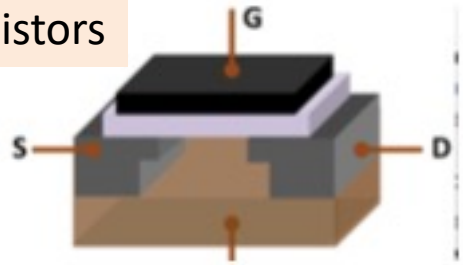
CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP222

3 Levels of Integration

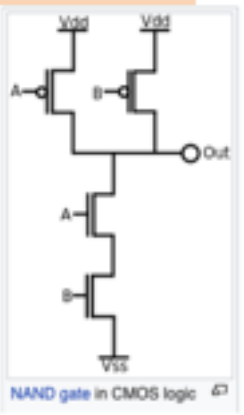
DR JEFF SOFTWARE
INDIE APP DEVELOPER

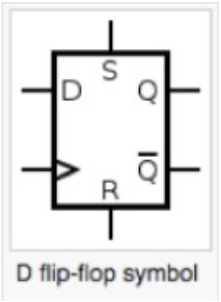
Hardware Building Blocks

0- Transistors

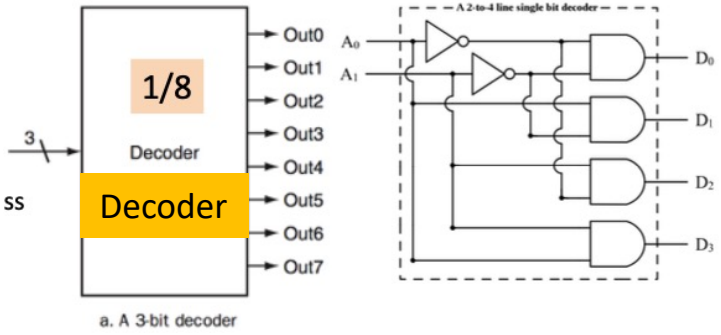


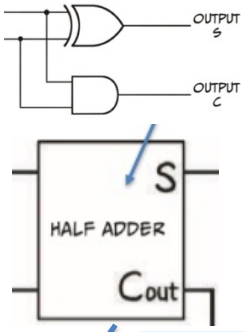
1- Gates & FF's



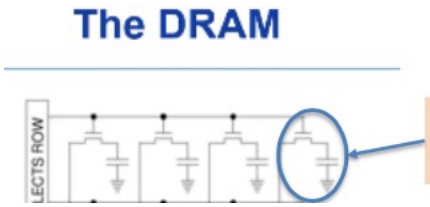


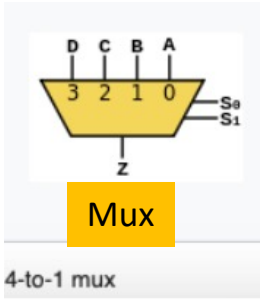
2-Single Functions

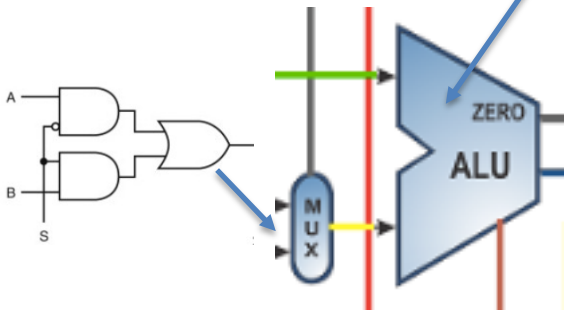




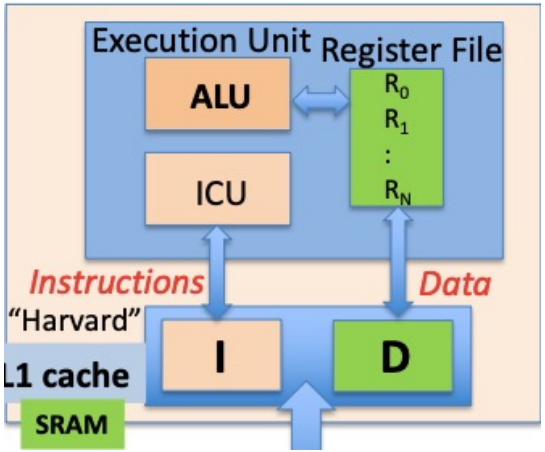
Memory Cells



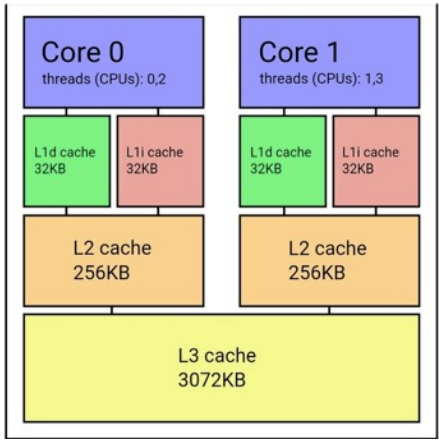




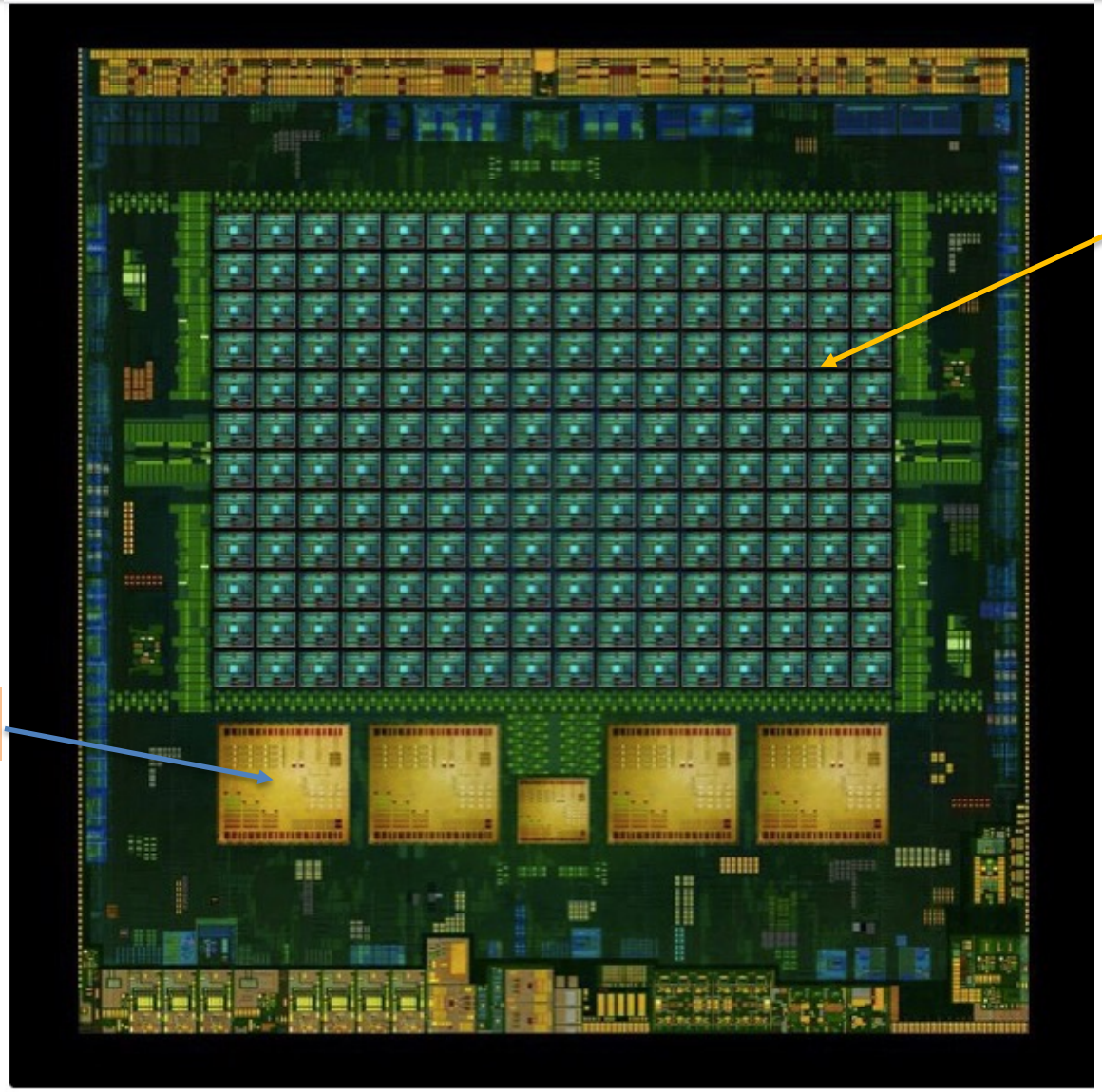
3-CPU



3-Multi-core CPU



SoC = CPU + GPU



This SOC has four CPU cores (ARM Cortex) and 192 GPU cores (Kepler).

CPU Cores: P & E

Mobile & Laptop (battery power)

BIG (large cache) — P — High Freq (4-5 GHz)

LITTLE (small cache) — E — Low Freq (2 GHz)

Clock domains

Efficiency (Power)

Desktop (Server)

P

Performance

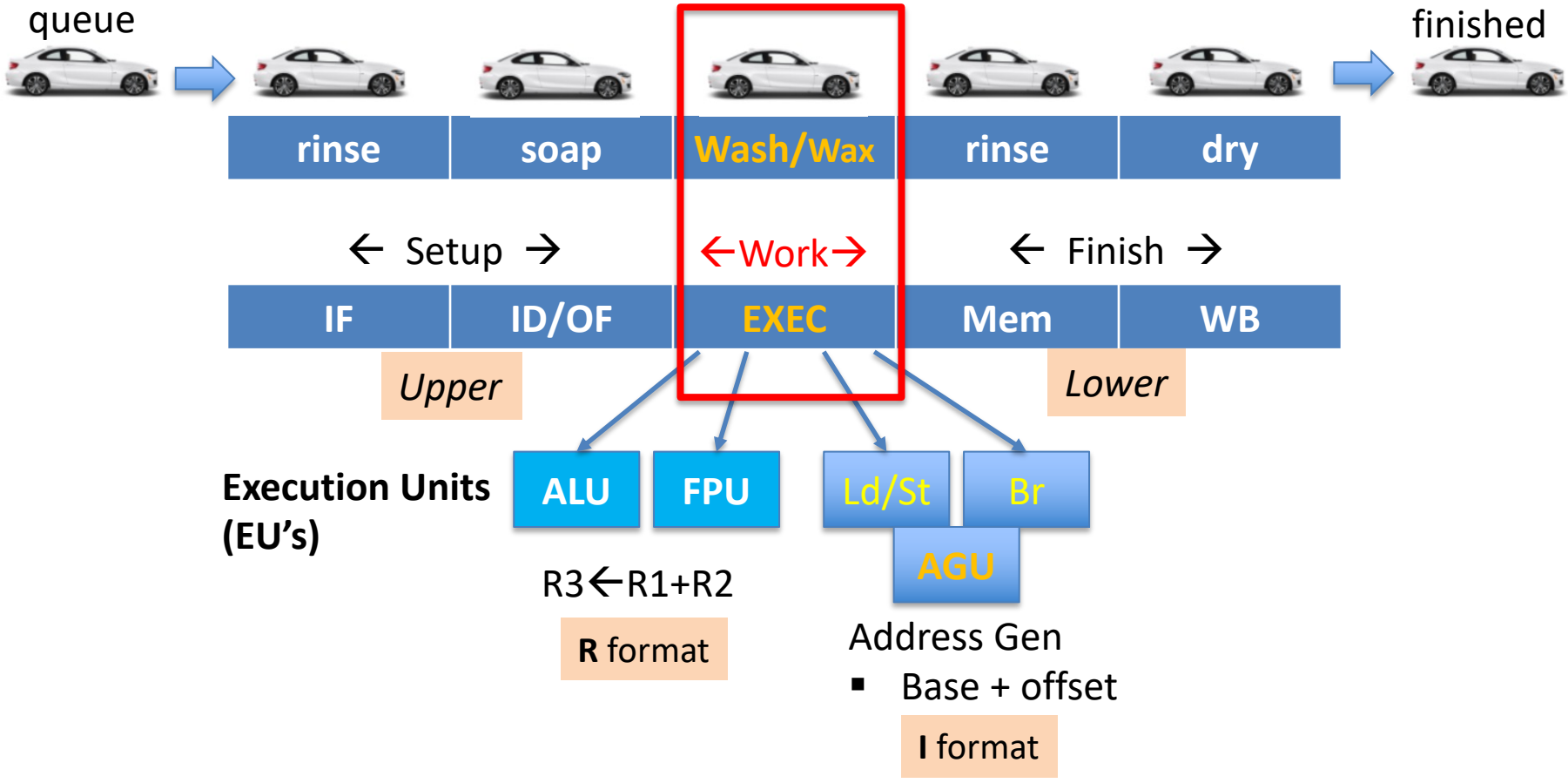
Section

Pipelines

MIPS RISC Pipeline

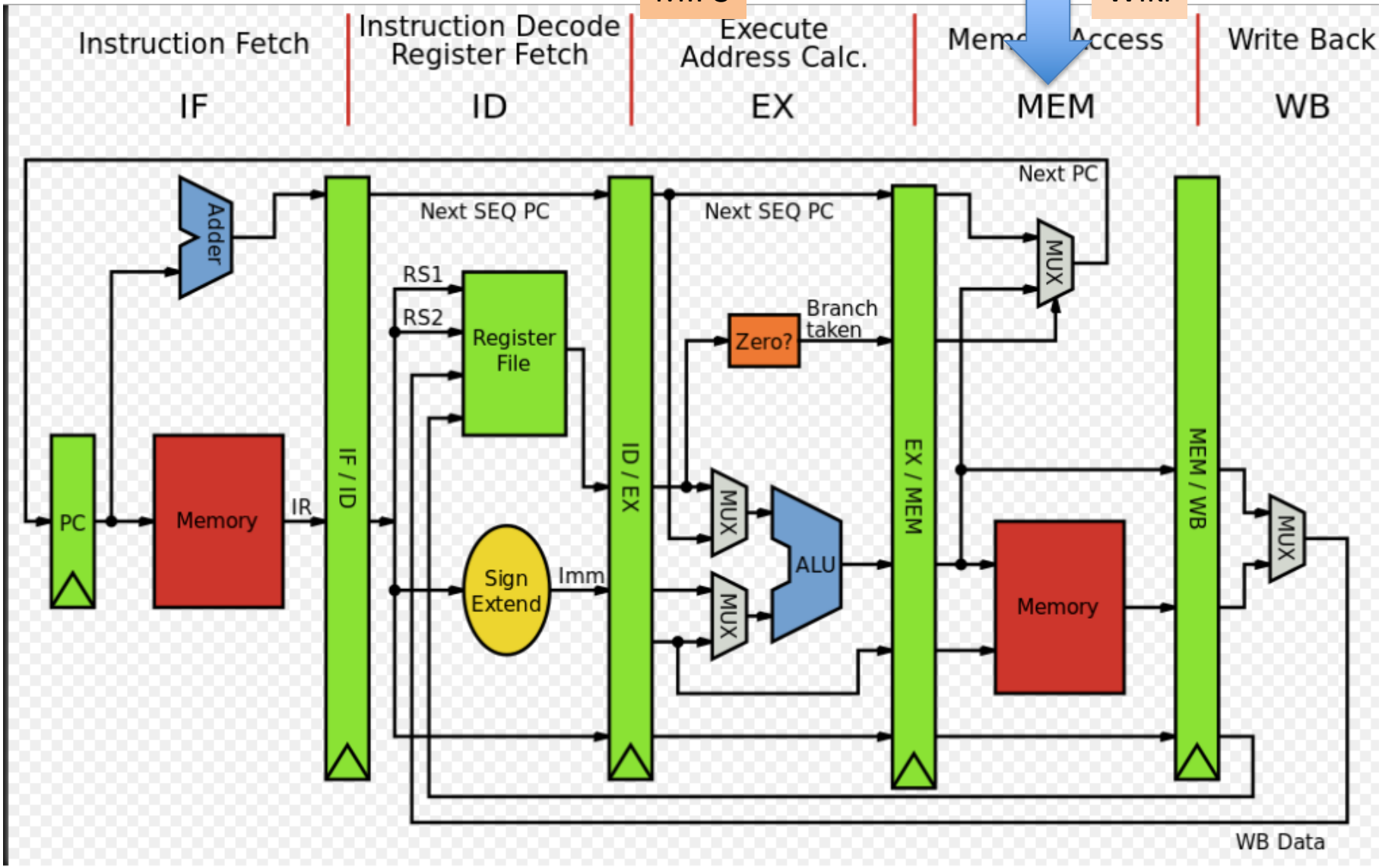
5 Stages

Each stage takes only 1/5 of instruction cycle: **clock F => 5x**



MIPS Pipelined Org

MIPS Wiki



MIPS, showing the five stages (instruction fetch, instruction decode, execute, memory access and write back).