CSUN | CALIFORNIA STATE UNIVERSITY NORTHRIDGE

COMP 122

DR JEFF SOFTWARE
INDIE APP DEVELOPER

COMP122

Spring 2023 Rev 2-10-23

Lectures on Vol 1

Computer Architecture

& ASSEMBLY Programming

By

Dr Jeff Drobman

website → drjeffsoftware.com/classroom.html

email → jeffrey.drobman@csun.edu

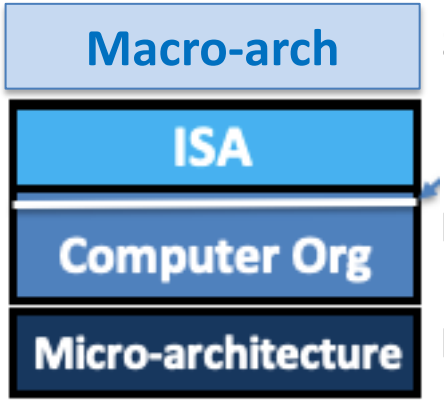
Index

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

COMP122

Course Description (122)

- **Computer Architecture/Organization (COMP122/ 222)**
  - **CPU, FPU, GPU org** (ALU, registers, addressing)
  - **ISA's: MIPS, ARM, x86**
  - **Memory models**
    - MLM- caches
    - Virtual memory
  - CPU status (PSW) & clock sync
  - Interrupts, Exceptions, Syscall
  - Cores & Threads
  - Pipelines (ICU)
  - *Microprogramming (Am2900)*
  - Logic & State Machines (FSM)
  - CPU performance/benchmarks
- **Computer Arithmetic (COMP222)**
  - **ALU**: Full adder
  - Mult/Div (Booth's algorithm)
  - Error codes (ECC, CRC, parity)
- **Parallel & *Micro* Architecture (COMP222)**
  - Multi-core, Multi-threading, *superscalar*
  - SIMD/MIMD/SPMD



System arch – *Cores*

Instructions (Primitives)

***Software** Interface*

Execution Units
- ALU, ICU, Reg

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

(COMP122)

- Software Tools
  - IDE's/Assemblers
  - OS, RTOS, ***Monitors***
  - ***Simulators***
- Debug Tools
  - ICE/Logic Analyzers
  - *Disassemblers*

Section

Intro
Models

Ordinals

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

10^(29)/2^(100) **geo**

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

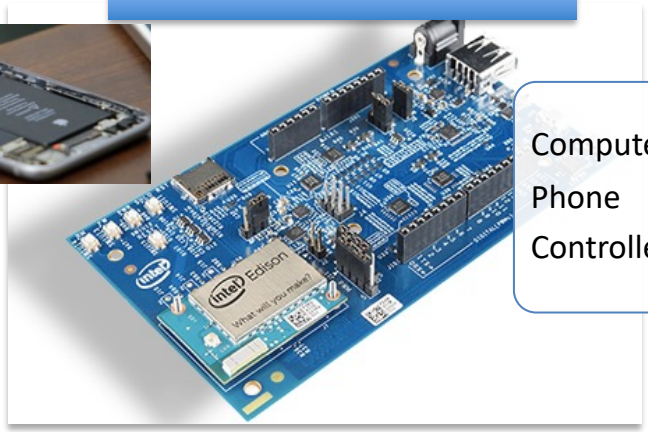
Ordinal	Power of 2	Power of 10	Actual
1K	2^{10}	10^{3}	1024
1M	2^{20}	10^{6}	1,048,576
1G	2^{30}	10^{9}	1.074×10^{9}
1T	2^{40}	10^{12}	1.0995×10^{12}

Name	2^n	M/G	Actual
byte	2^{8}	--	256
short	2^{16}	64K	65,536
word	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}

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

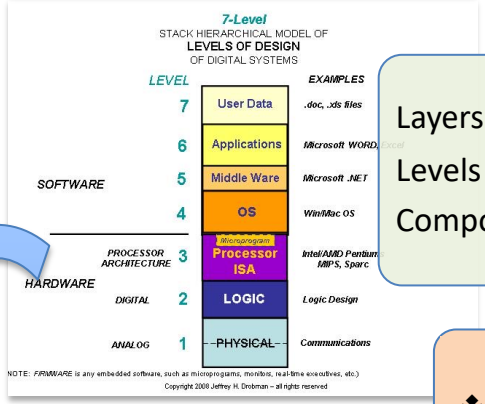
Digital Systems High Level

Hardware



Computer
Phone
Controller

Software



Layers
Levels
Composition

- CODE
- DATA

External Storage

Hard disk
Flash disk
Flash
EEPROM

Computer **Engineering**

- **Architecture**
- Physical design

Computer **Science**

- **Algorithms**
- Theory

Realms of Software

~70% of all software

❖ Applications
- ❑ Desktop
- ❑ Mobile (Apps)
- ❑ Web

❖ APIs (Frameworks)

❖ Web
- ❑ Markup
- ❑ Applications
- ❑ SQL databases

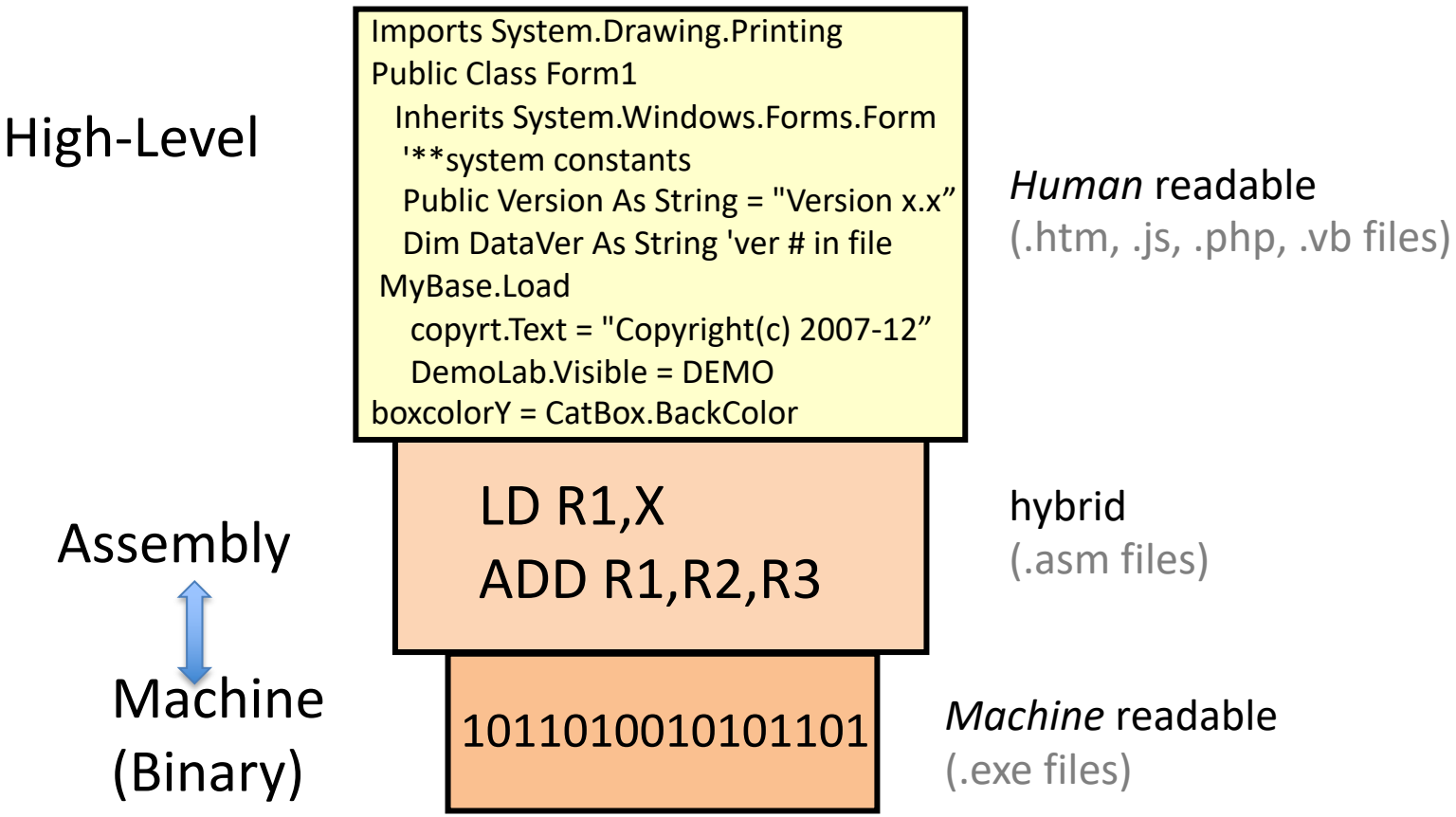
❖ Client-Server model
❖ Language "stacks" (e.g., LAMP)

❖ Embedded Control
- ❑ Small (8-bit)
- ❑ Medium (16-bit)
- ❑ Large (32/64-bit)

❖ From TV remotes to
❖ Autonomous cars and
❖ Robots

➢ Common required properties
- Performance
- Reliability (bug free)
- *Security*

Software *Levels*

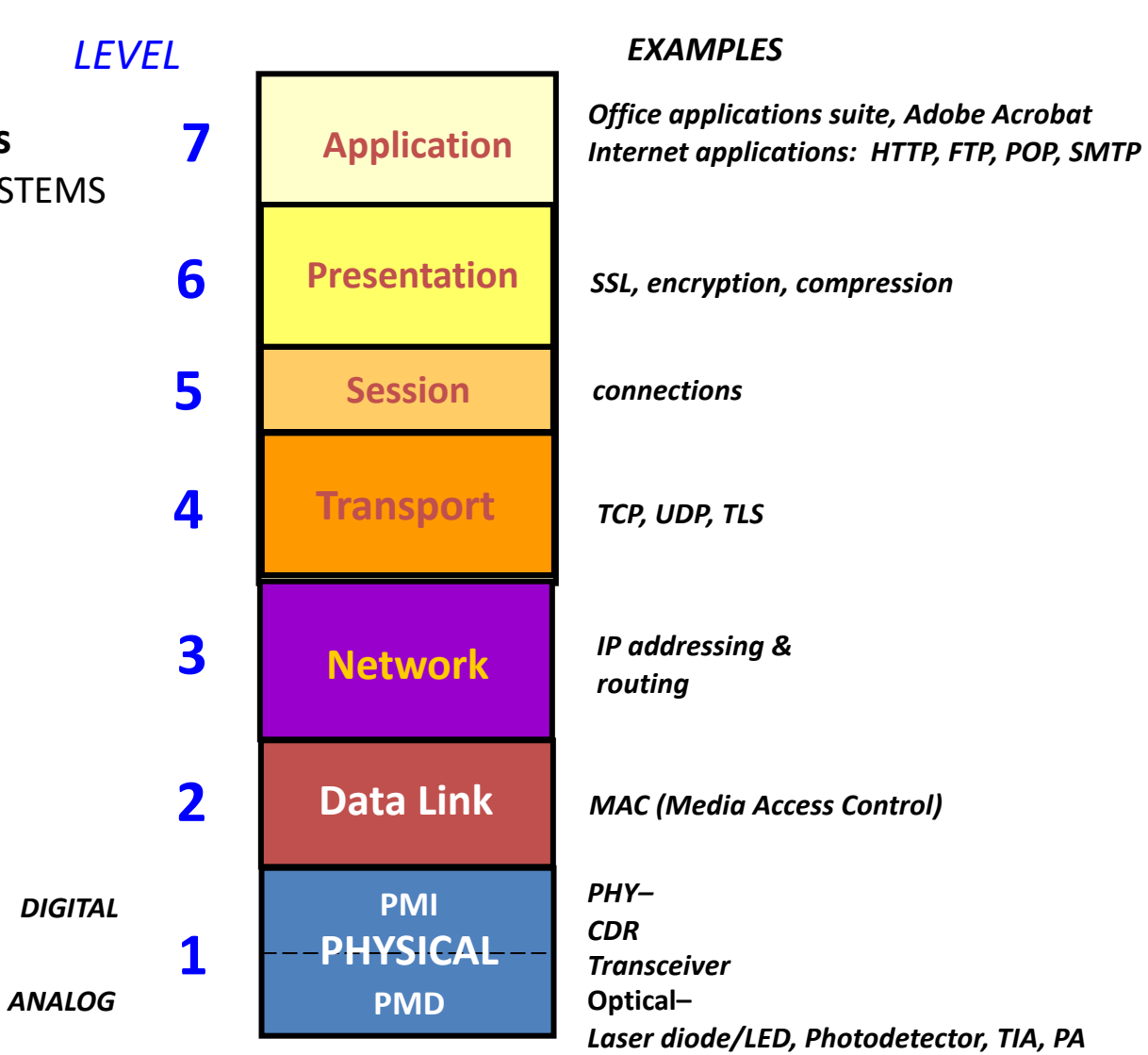


CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Com Protocol Layers

7-Level
OSI MODEL
of **Protocol Layers**
IN COMMUNICATIONS SYSTEMS

Protocol STACK Model

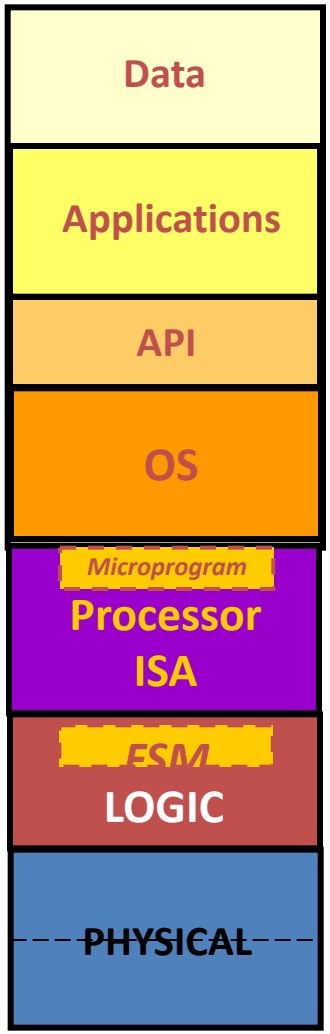


CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Hardware-Software *Layers*

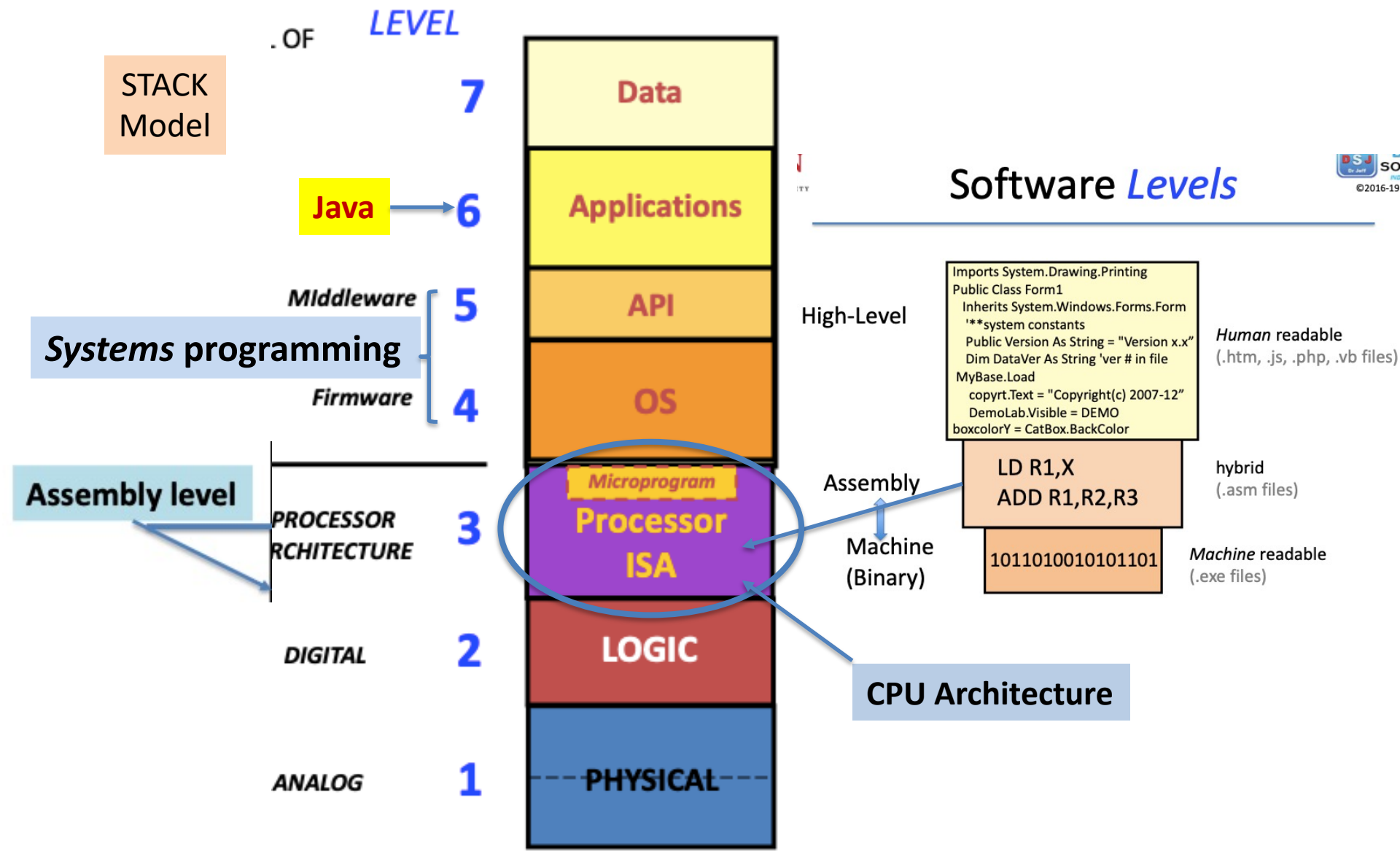
7-Level
STACK HIERARCHICAL MODEL OF
LEVELS OF DESIGN
OF DIGITAL SYSTEMS

STACK Model



(NOTE: *FIRMWARE* is any embedded software, such as microprograms, monitors, real-time executives, etc.)

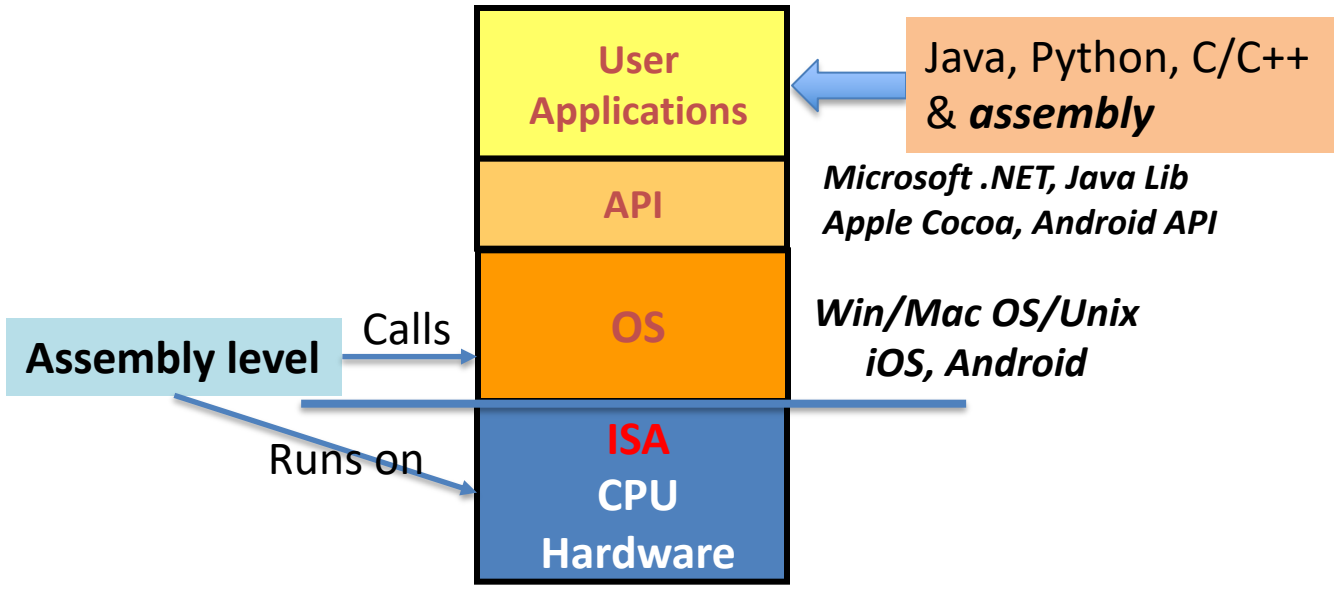
Levels of System Architecture



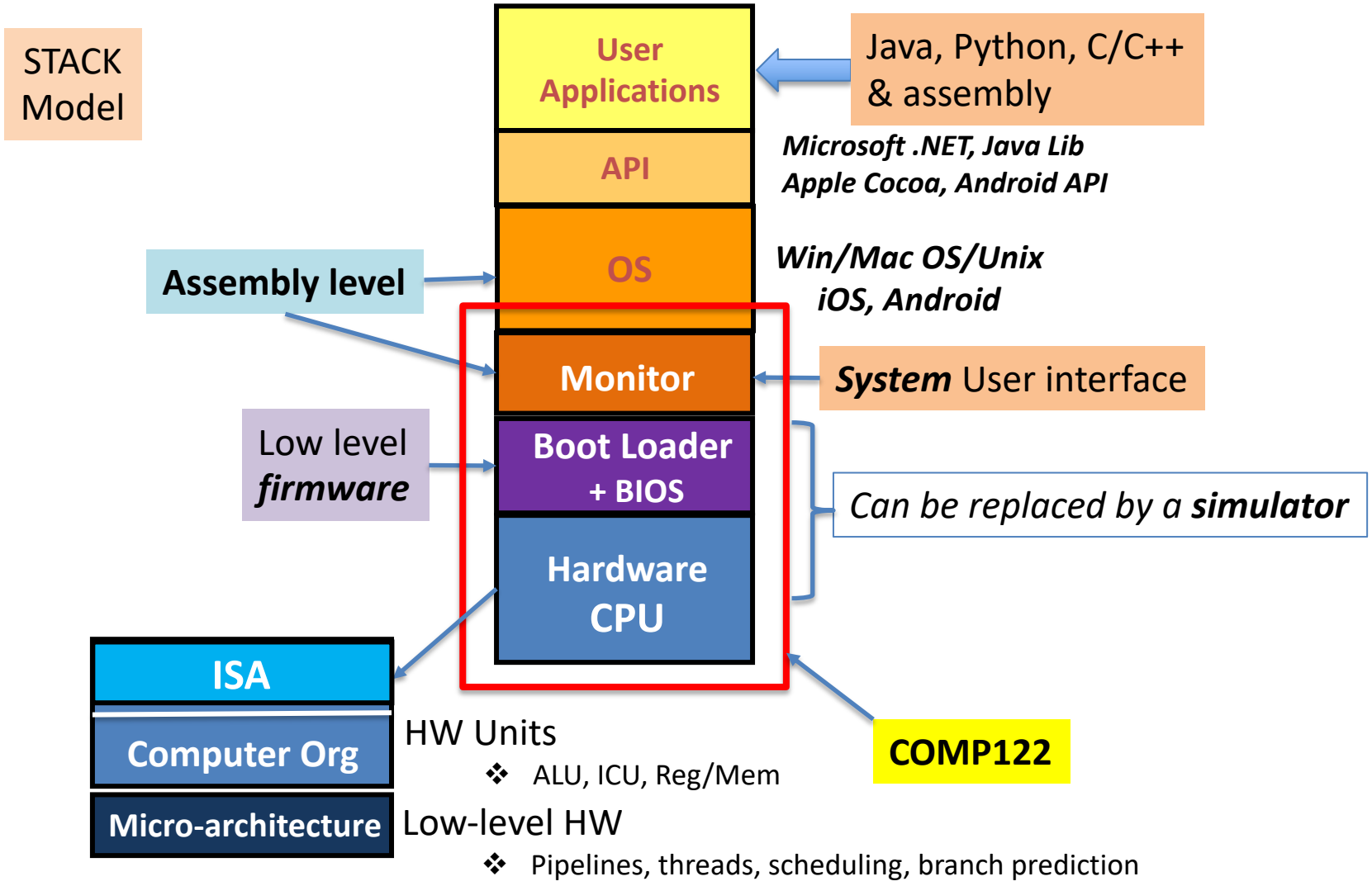
Software *Layers*

STACK Model

Simple View

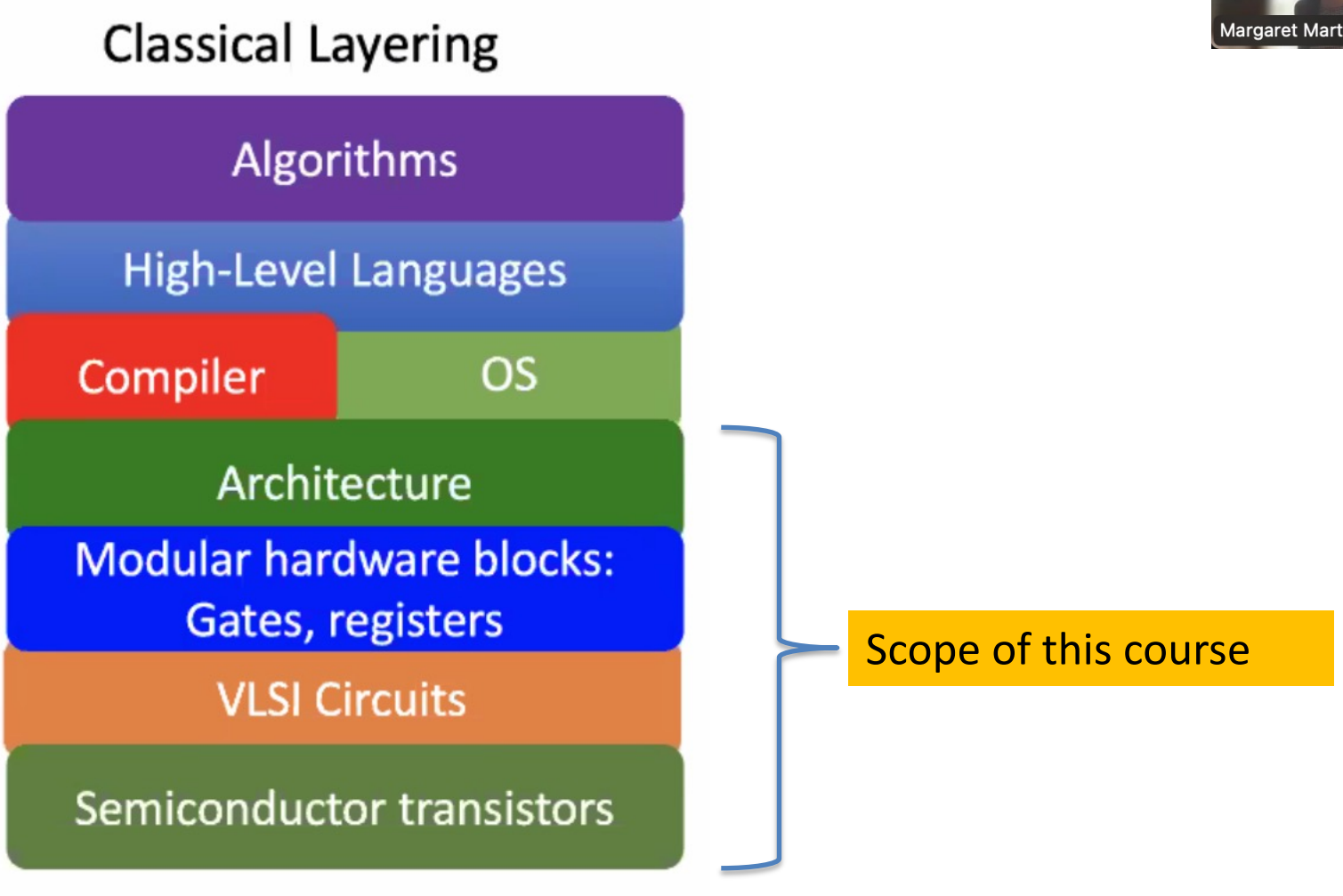


Hardware/Software *Low Level*



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

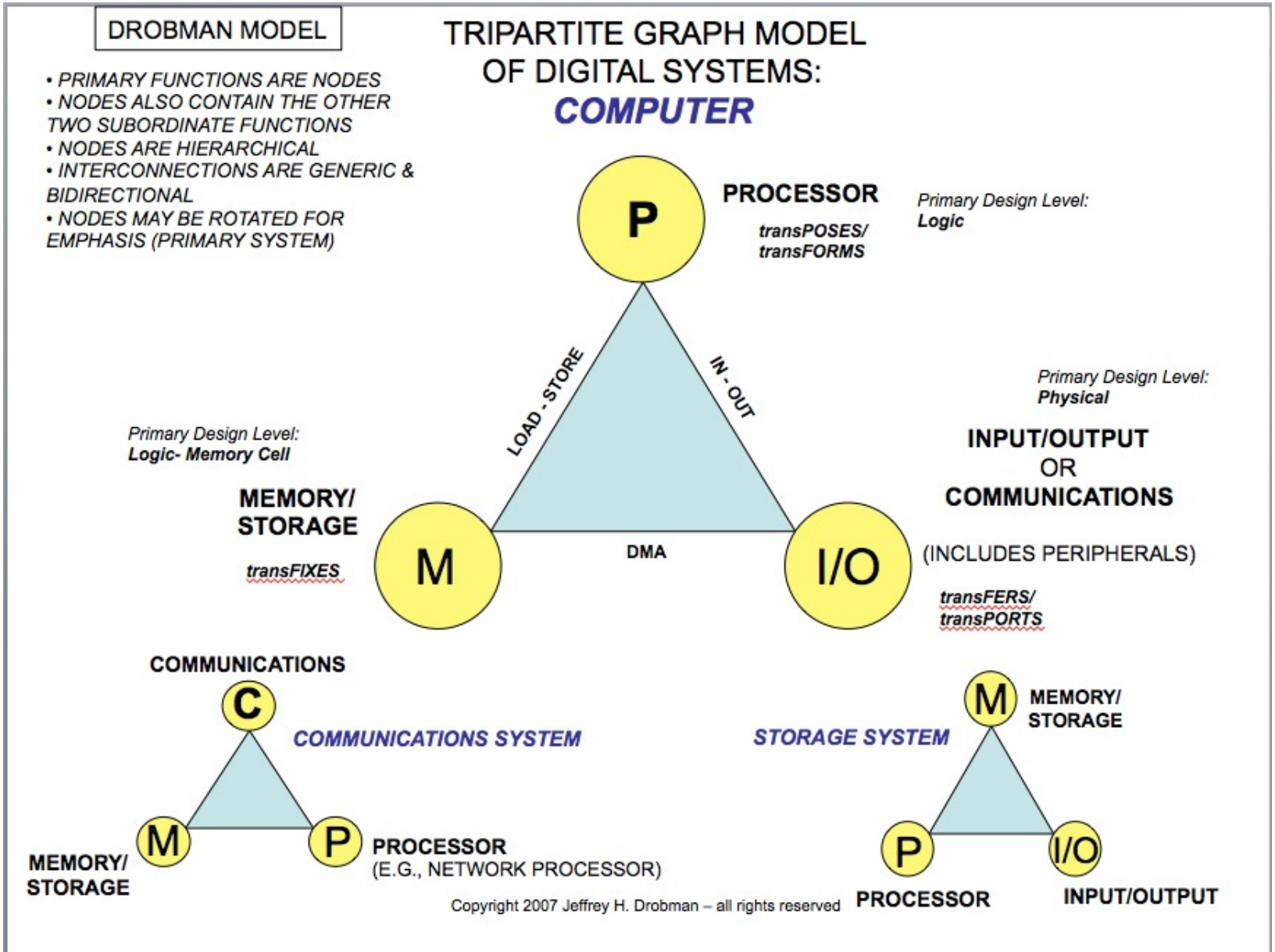
Computation Stack Mode.



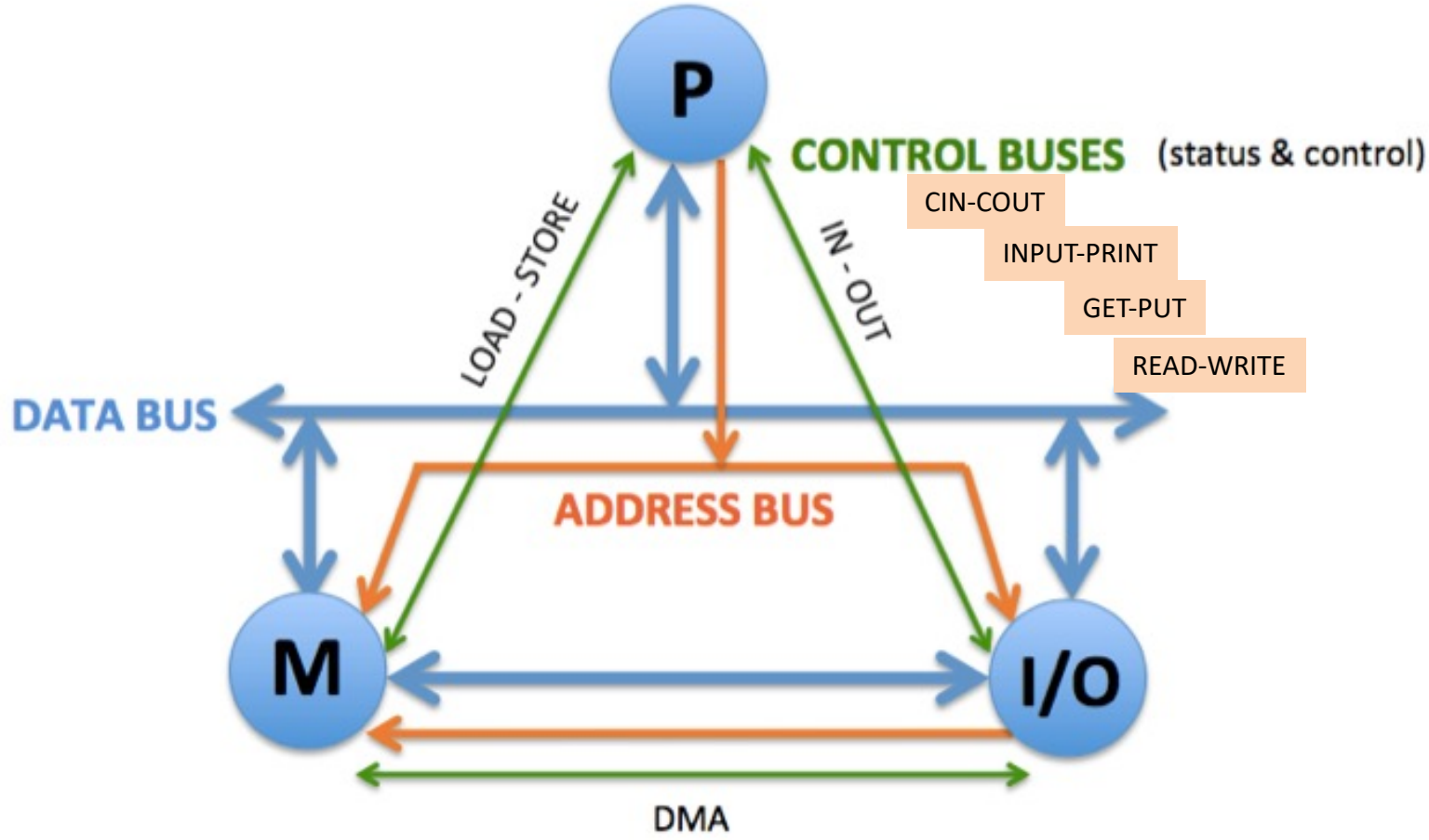
Chip Specs

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

Hardware-System Model



Hardware-Bus Model



NON-MULTIPLEXED BUSES

Computer Org

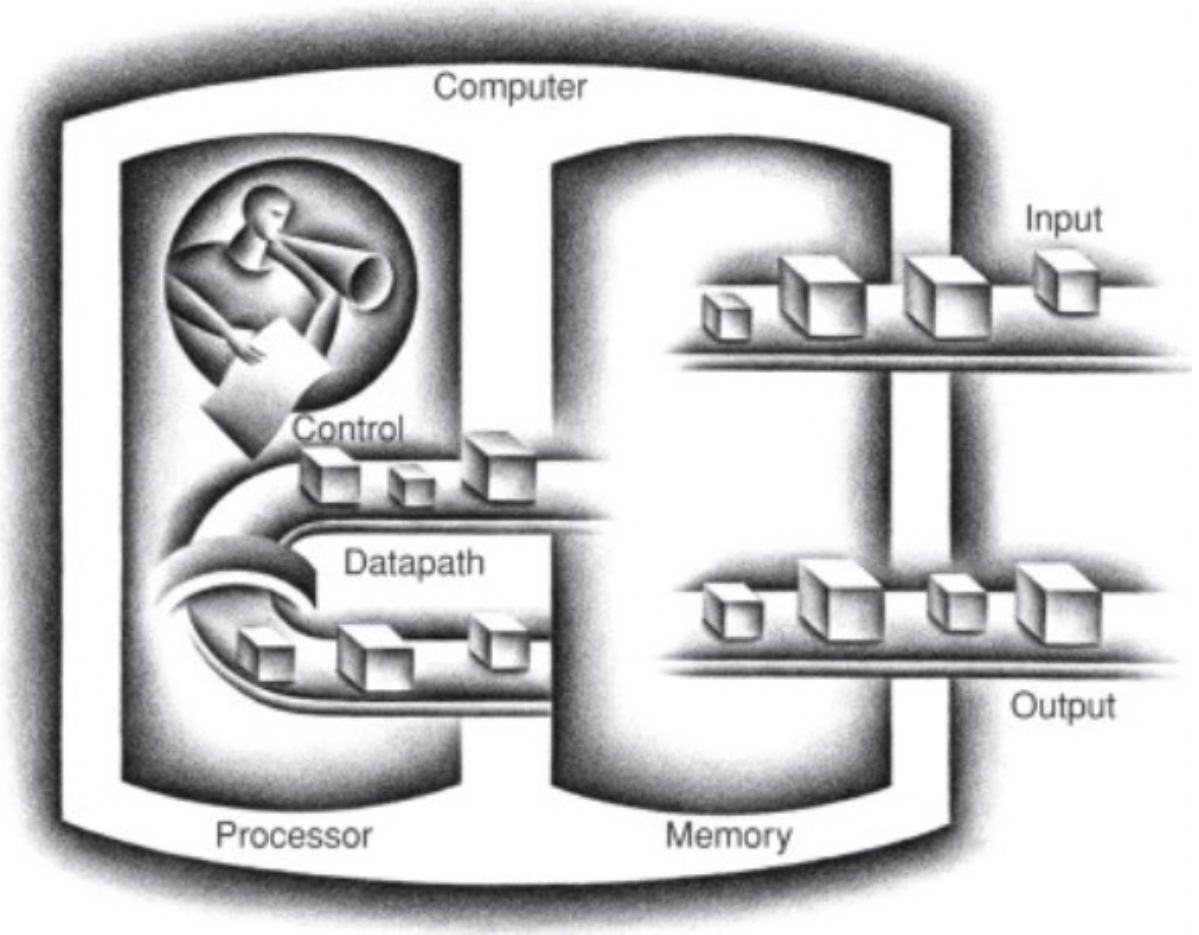
P&H Ch 1

Figure 1.4.1: The organization of a computer, showing the five classic components (COD Figure 1.5).

The processor gets instructions and data from memory. Input writes data to memory, and output reads data from memory. Control sends the signals that determine the operations of the datapath, memory, input, and output.

- CPU
  - Processor
  - Memory
  - I/O

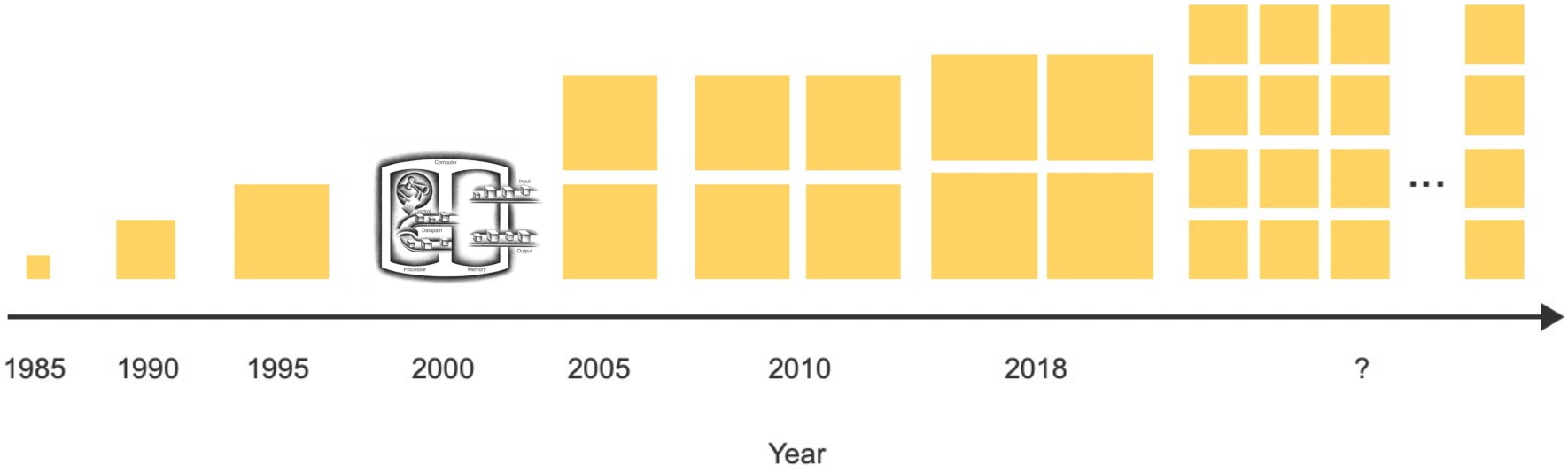
- Datapath
- Control



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Computer Org: Multi-Core

P&H Ch 1

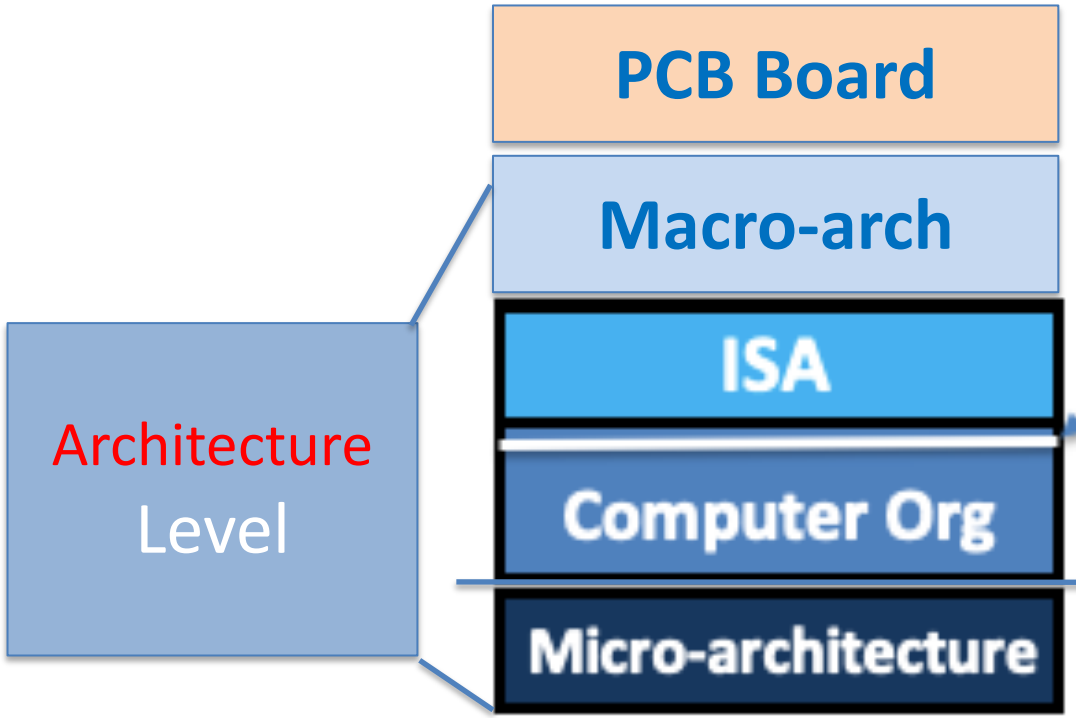


CSUN CALIFORNIA STATE UNIVERSITY NORTHRIDGE

COMP122

Computer Architecture

4-Layer Stack Model



Chipsets, Chiplets and Sockets

System arch – *Cores, caches*

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

122

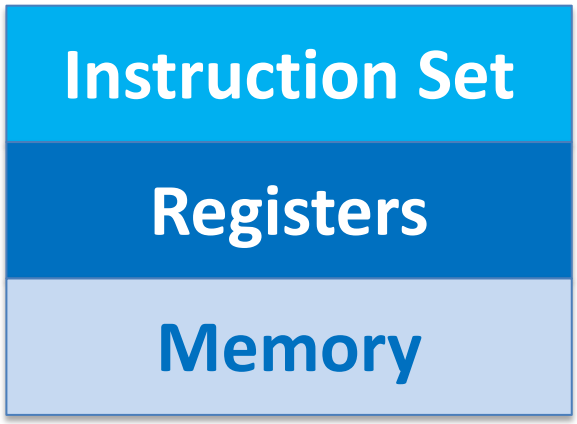
Execution Units
- ALU, ICU, Reg

Below "see" level

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

222

ISA

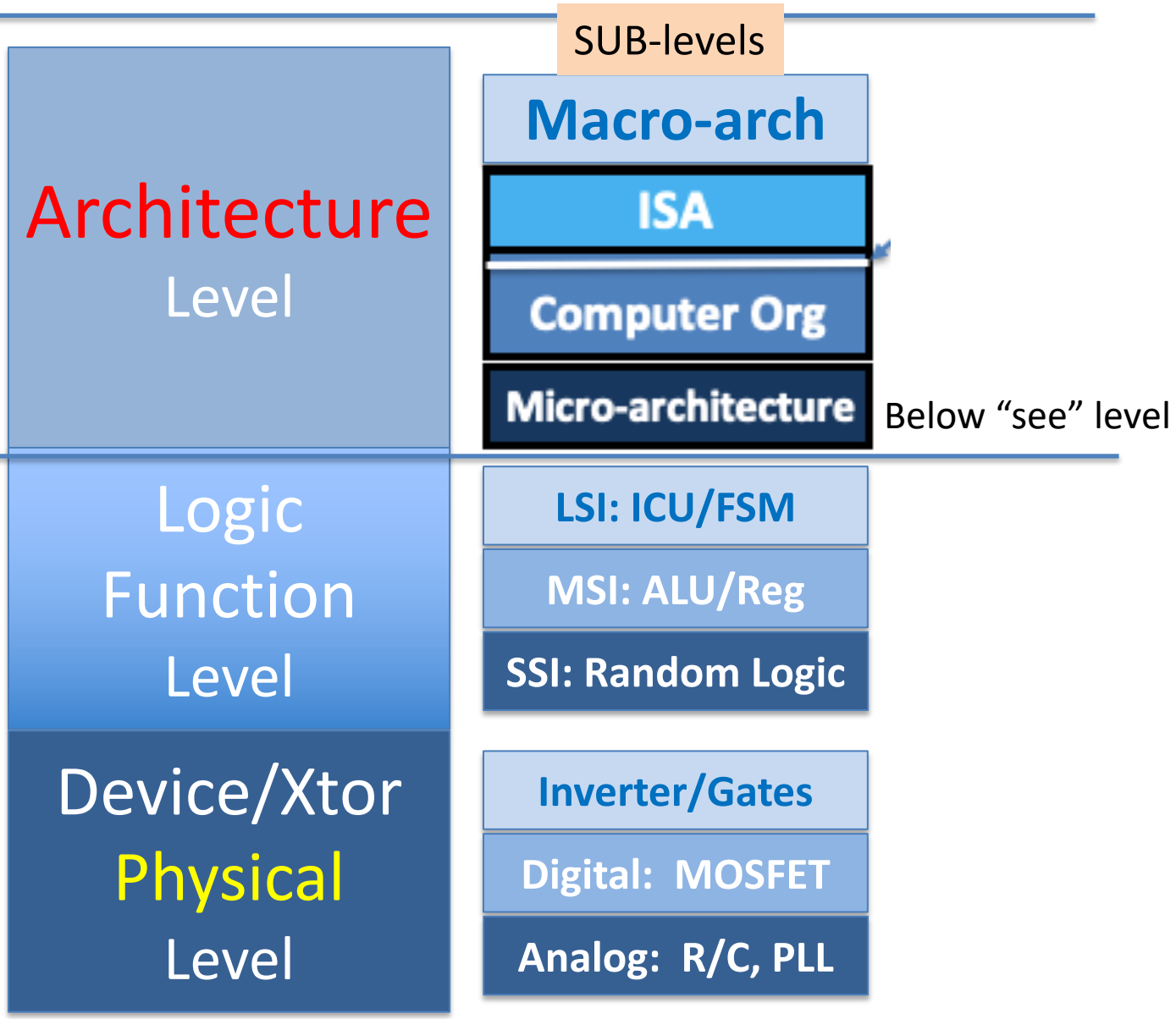


Instructions (Primitives, pseudos)

GR/Dedicated

Segmentation
Virtual ← →Physical

Transistors to Chips: Levels

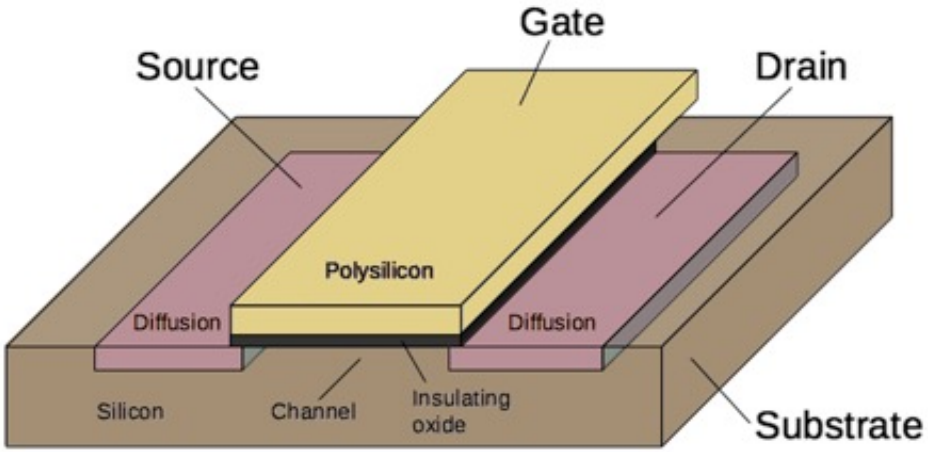


Physical Level: MOSFET

Device/Xtor
Physical
Level

Inverter/Gates

Digital: MOSFET



Structure of a MOSFET in the integrated circuit.

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

P→N→C MOS

What is an NMOS transistor?

Digital: MOSFET

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

Answered just now

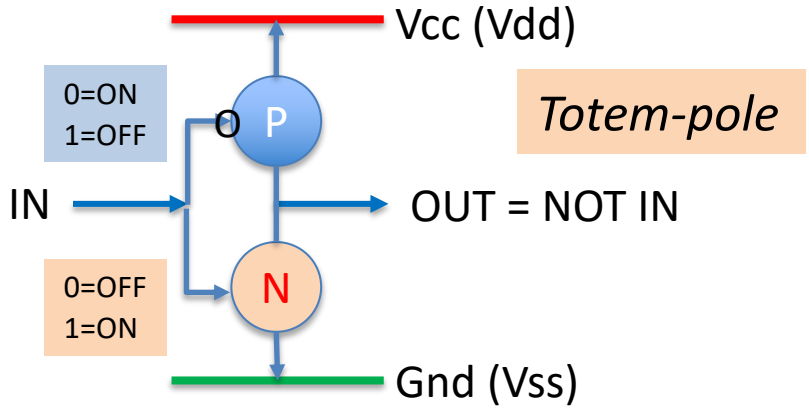
"MOS" is a planar (2D) FET structure that uses a "Gate" voltage to switch a FET on/off by opening or closing a conductive "channel" between a current Source and Drain. The Source, Drain and Channel are of the same semiconductor type (P or N) in order to have a straight closed connection. the industry, via pioneer Intel, first used "P channel" MOS using a negative supply and gate voltage. but since "N channel" is faster, and uses a positive supply and gate voltage, Intel switched to it. for about the first 10 years, all MOS was NMOS. then along came CMOS, which uses both P and N MOSFET's in a push-pull totem pole structure (one ON, one OFF) — to save power, while just as fast as NMOS.

Complementary

CMOS

INVERTER

Inverter/Gates



Totem-pole

P→N→C MOS

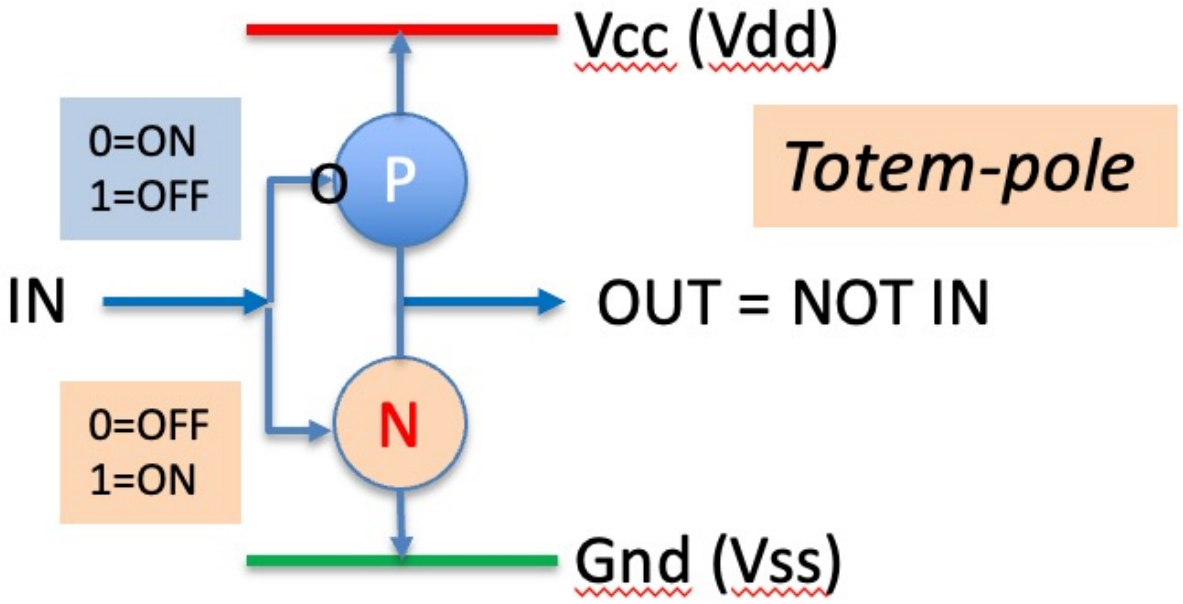
Device/Xtor
Physical
Level

Inverter/Gates

Complementary

CMOS

INVERTER



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Transistors to Computers

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Quora

If computers are really just many (billions) of on/off switches, how do they perform operations?

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

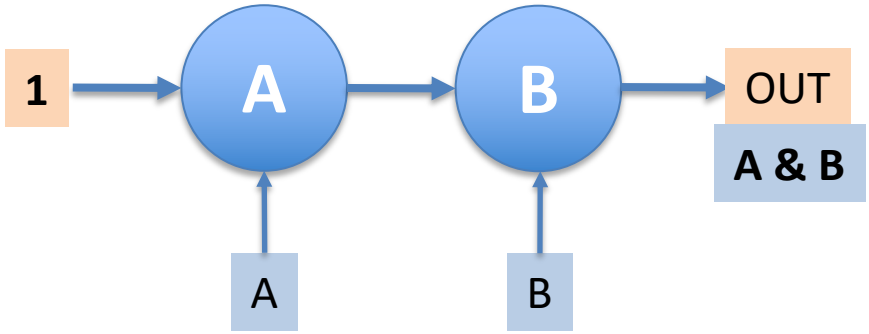
Answered just now

via a multi-level hierarchy of digital logic. transistors are combined to form logic “gates” of simple logic functions (AND, OR, NOT). the gates are combined to form more complex functions such as decoders, ALUs, and multiplexers. these functional blocks are then combined further into ever more complex logic blocks such as EU’s and then CPU cores. also, random logic implements the ICU as an FSM which includes pipelining. besides logic, computers have “storage” in the form of registers and memory (at up to 4 levels) via DRAM and SRAM cells formed from transistors (and a capacitor).

Logic Gates: AND, OR

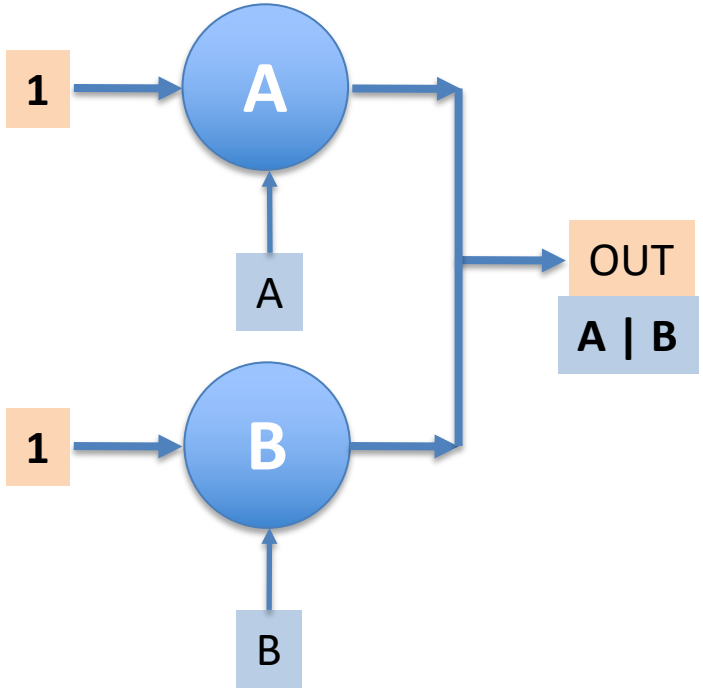
AND

SERIES



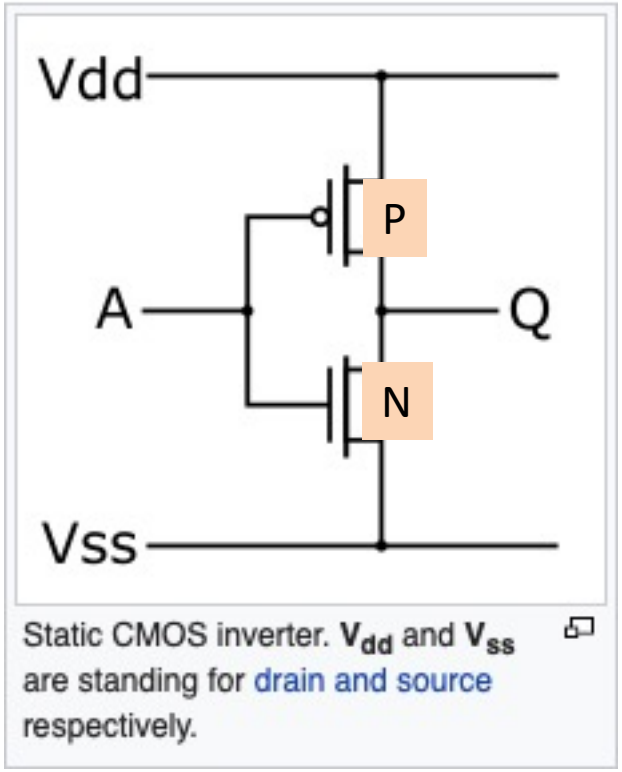
OR

PARALLEL

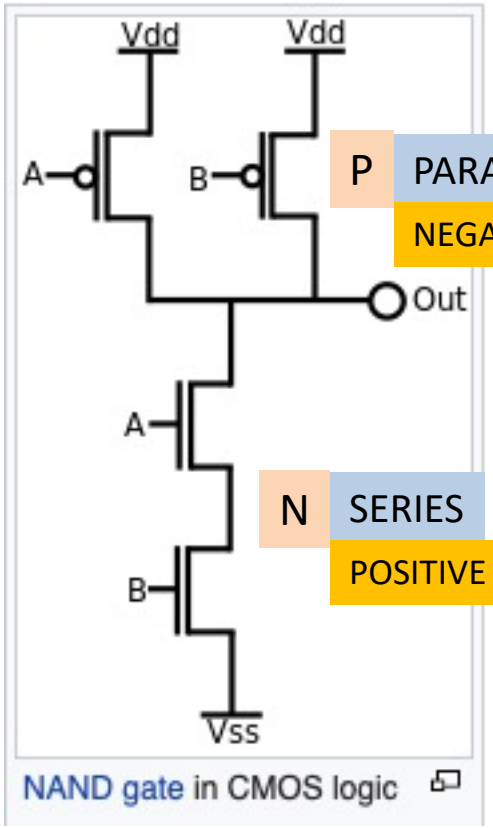


CMOS Gates

MOSFET



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



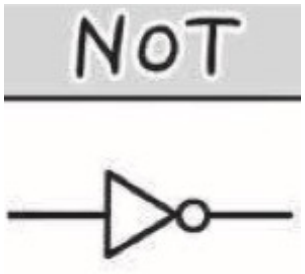
NAND gate in CMOS logic

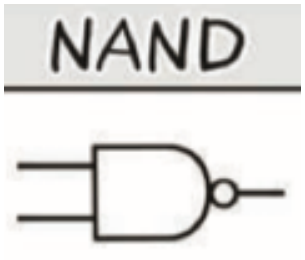
P PARALLEL

NEGATIVE OR

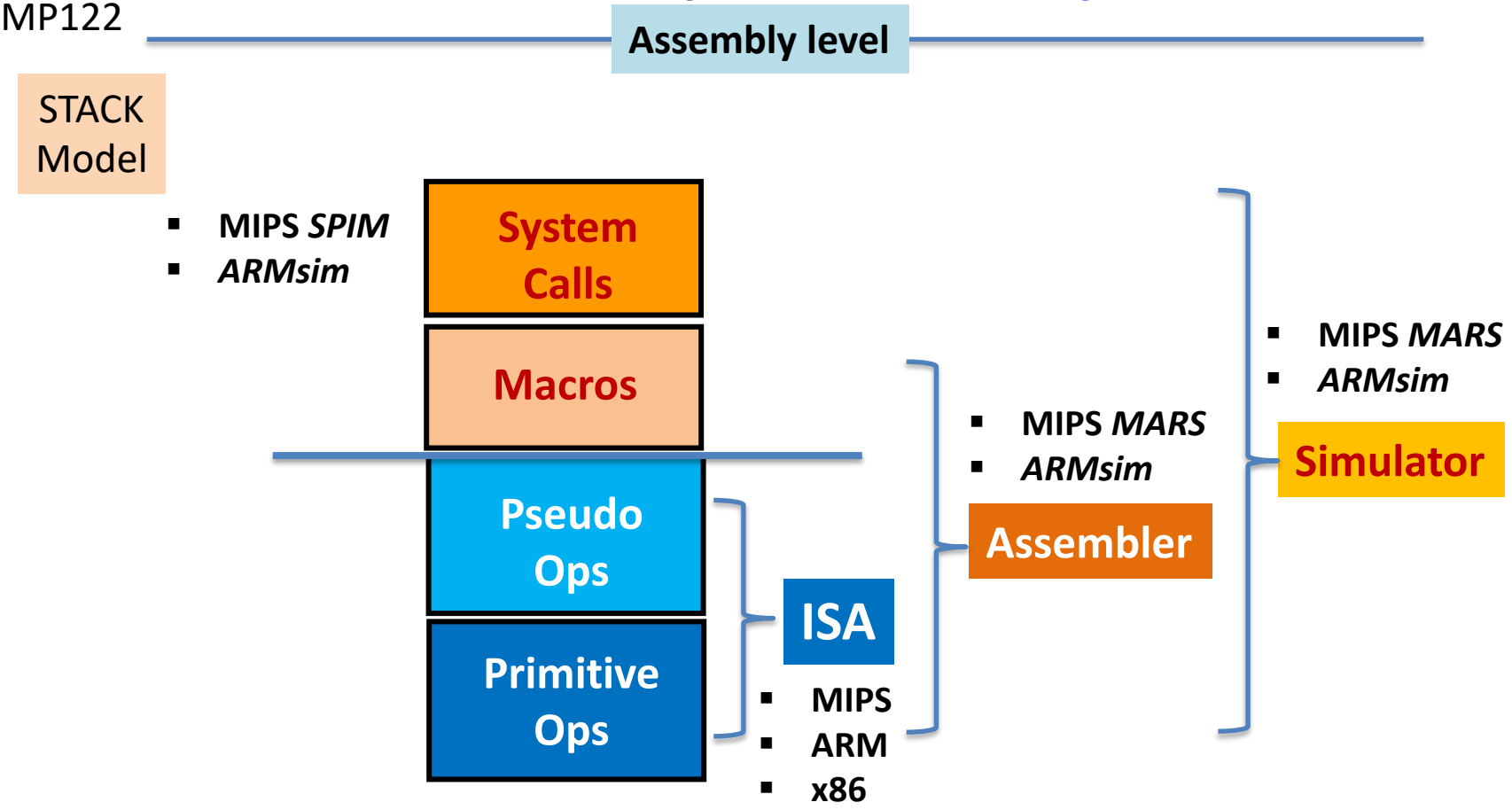
N SERIES

POSITIVE AND





Assembly Level *Layers*



Levels of Instructions

Instruction Set

- Subroutines
  - Block of code that can be "called"
- ***Macros***
  - Block of code that will be substituted *in situ*
- ***Pseudo*** instructions
  - Group of 1 or more ***primitives*** abstracted to higher level
- ***Primitives***
  - Native ***machine instructions*** (in the ISA set)
- *Micro* instructions
  - Complete set of all *control bits* per clock cycle
  - Now = *primitive* (both execute in 1 clock cycle, per *RISC*)
  - Old CISC: Each *primitive* assigned a micro-coded subroutine
  - Can be "horizontal" = Long Instruction Word (VLIW) for *parallel* ISA's

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

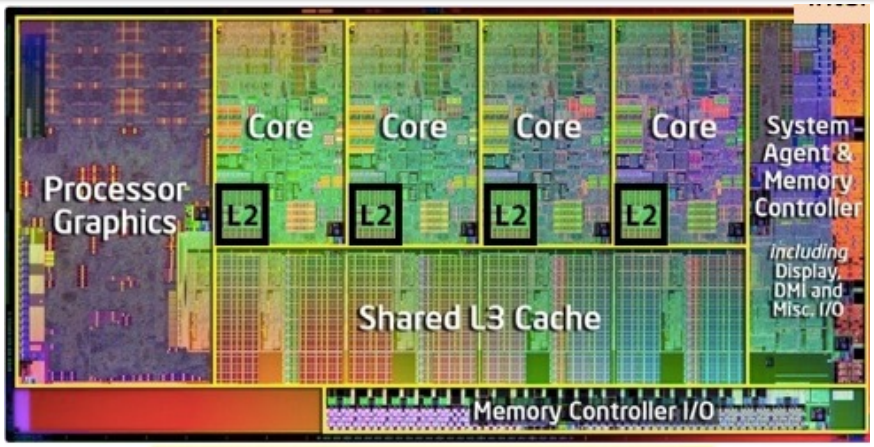
COMP122

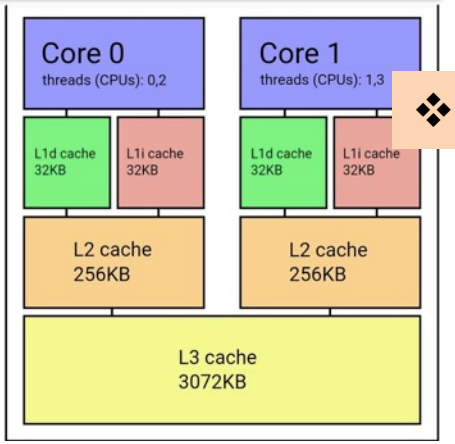
4 Levels of CPU *Architecture*

COMP122/222

❖ Macro

System= *Multi-core SoC*



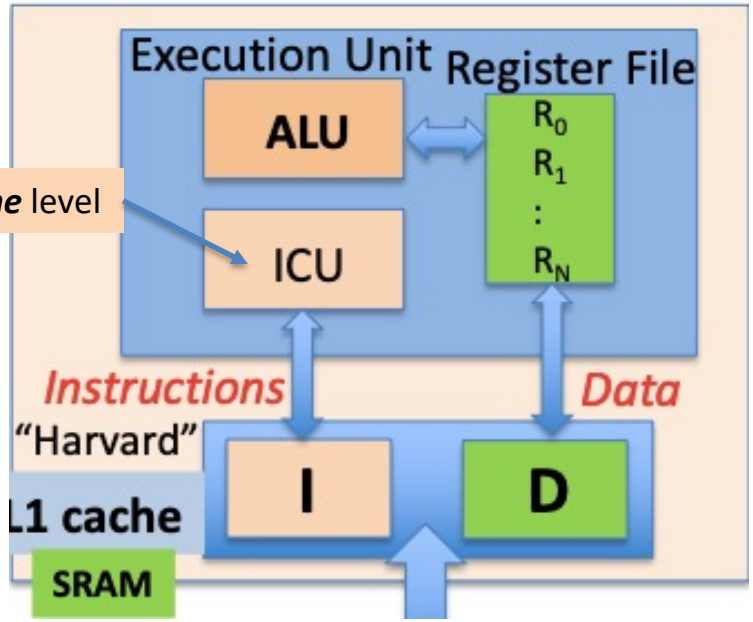


❖ Floorplan

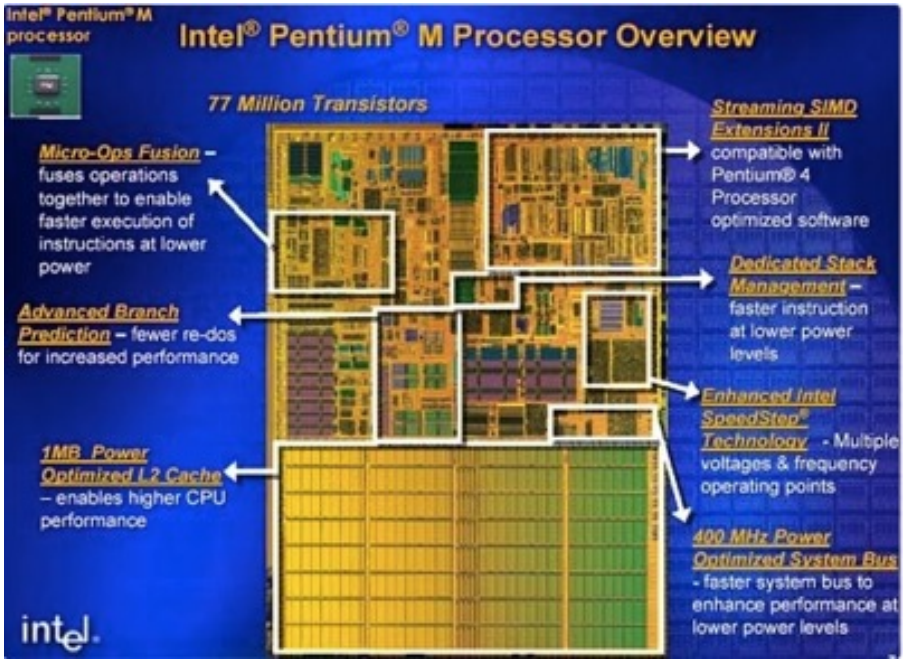
COMP122

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

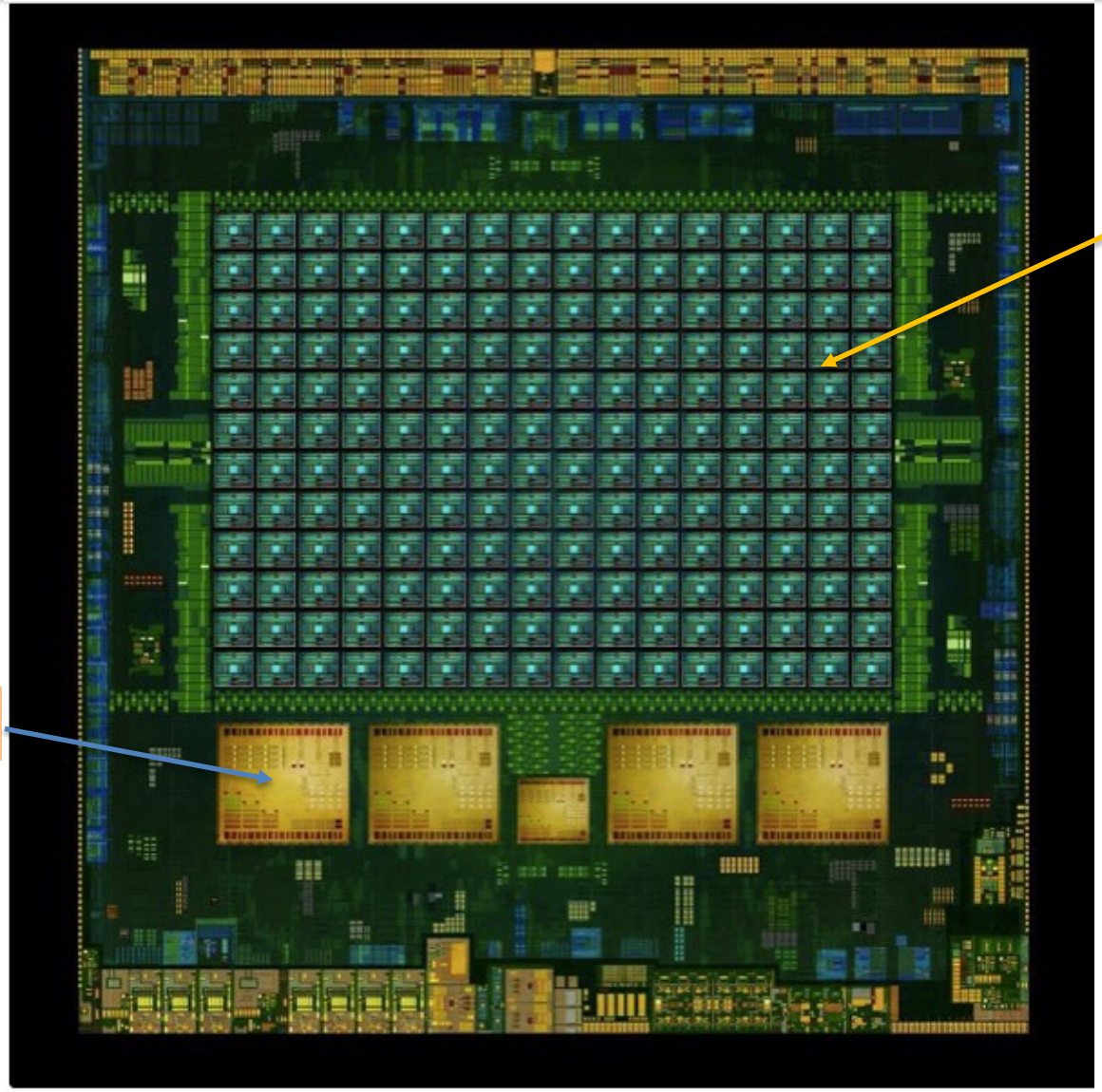
Pipeline level



COMP222 ❖ Micro ***Pipeline*** level



SoC = CPU + GPU



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

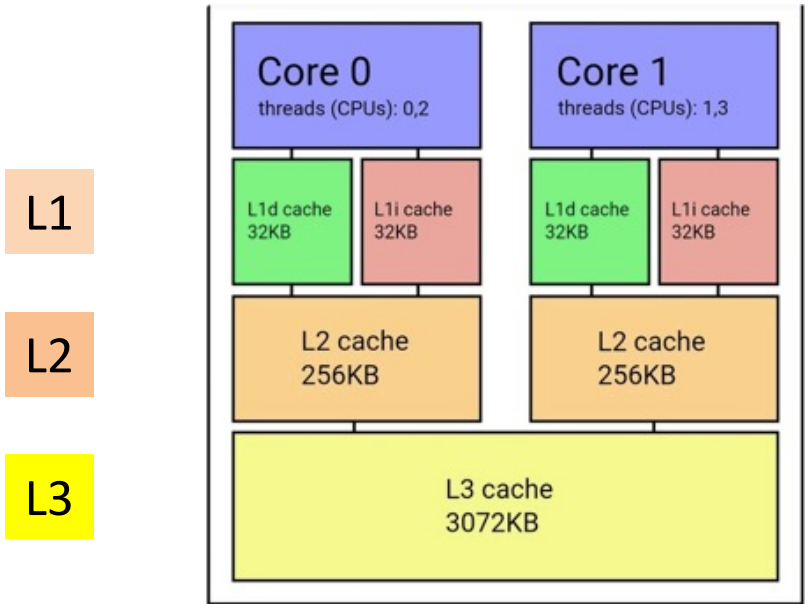
Cache Levels

Why is cache memory divided into levels?

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

Answered just now

there are generally 3 levels of cache used today. the original RISC CPU's used 2 levels, with L1 on-chip and L2 off-chip. L1 must be Harvard style, with separate caches for I and D, both of which must be fast enough to operate at the CPU clock frequency. there is a limit to L1 cache sizes, based on the speed requirement. but DRAM main memory is so slow relative to the CPU, that it makes sense to provide an L2 cache that is larger but slower than L1, but still way faster than main DRAM. multi-core has added a shared L3 cache.

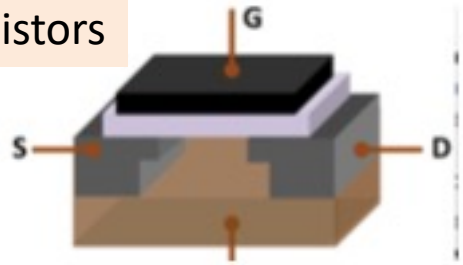


CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

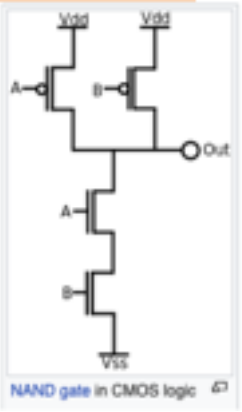
3 Levels of Integration

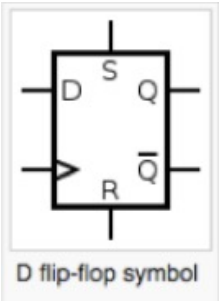
Hardware Building Blocks

0- Transistors

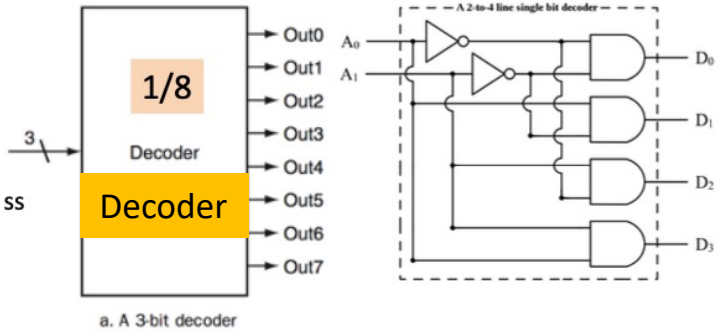


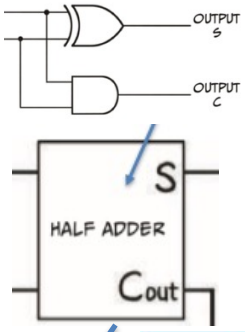
1- Gates & FF's



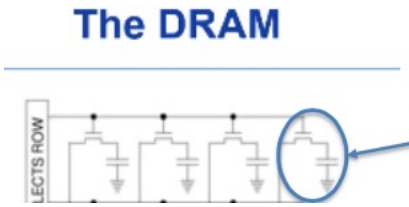


2-Single Functions



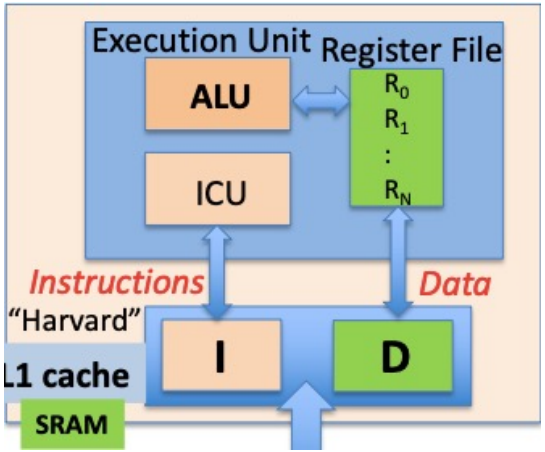


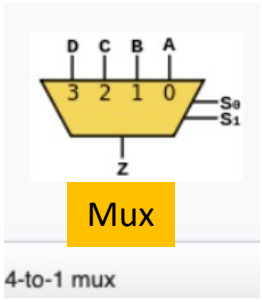
Memory Cells

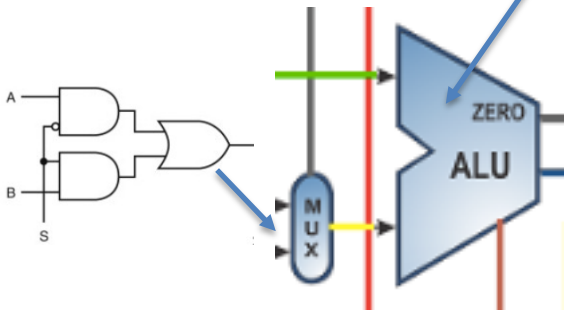


- 1 transistor
- 1 cap (parasitic)

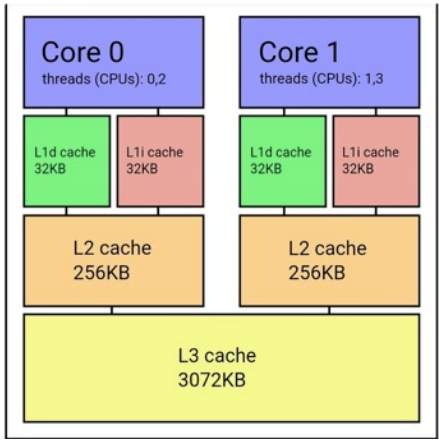
3-CPU







3-Multi-core CPU



CPU's Are Small?

Quora

CPUs are very small and they don't contain much material. So why does it take so many years to develop new technology when all it really is some sort of change in materials or it's position? Genuine question.

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

Answered just now

CPU's may be small in physical size, but very large in complexity and performance. each CPU core utilizes up to a half billion transistors. this has taken computer scientists decades to perfect the architecture, and process chemists decades to shrink these transistors so much.

Computer Architecture

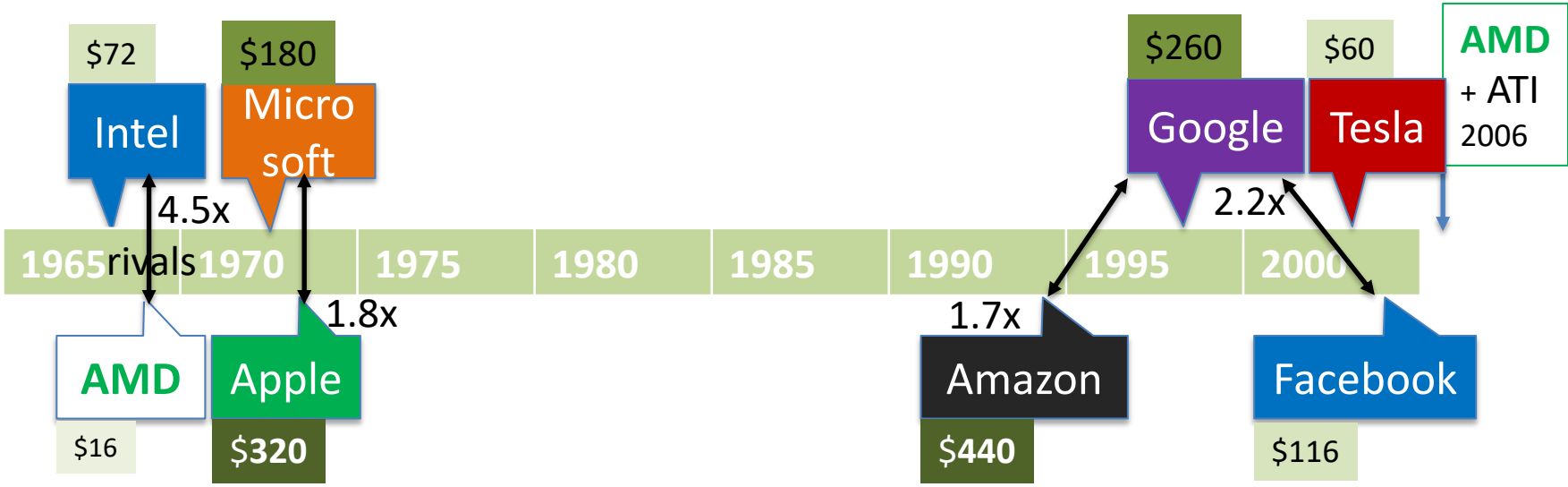
Tech Landscape

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Tech Titan Timeline

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Annual Revenue in $B | Historical Perspective | As of 4Q2021



Eras: chips | PCs | Sales | Search | Social

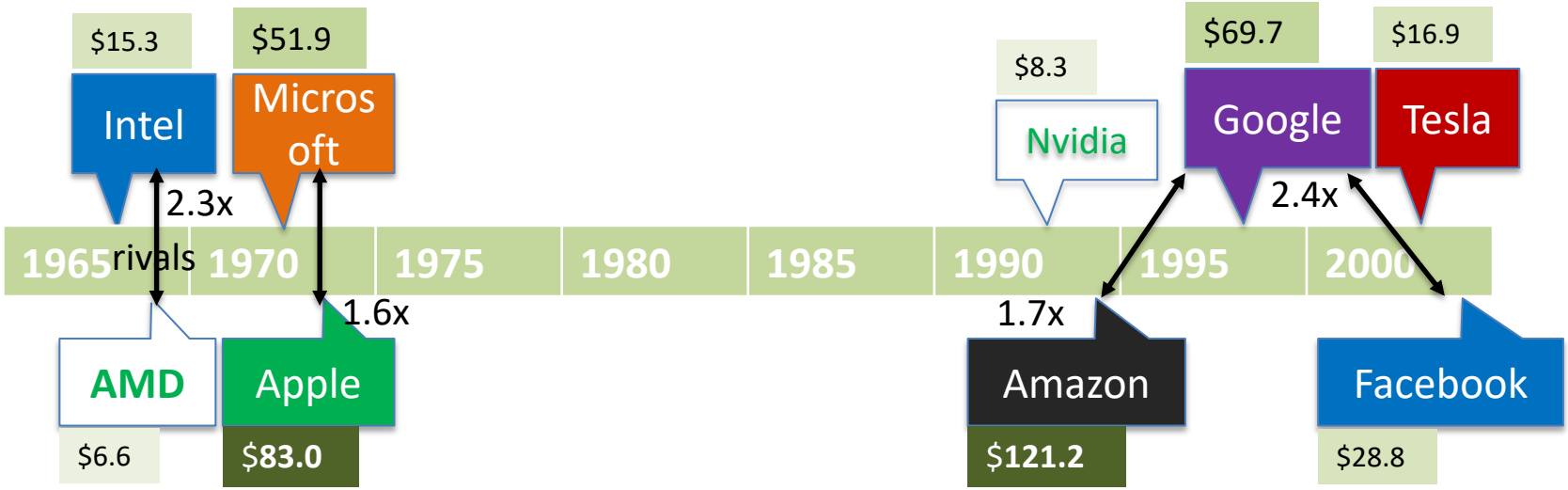
delivered

Internet
1969

WWW
1989

Tech Titan Timeline

Qtr Revenue in $B | Current Perspective | **Q2** 2022

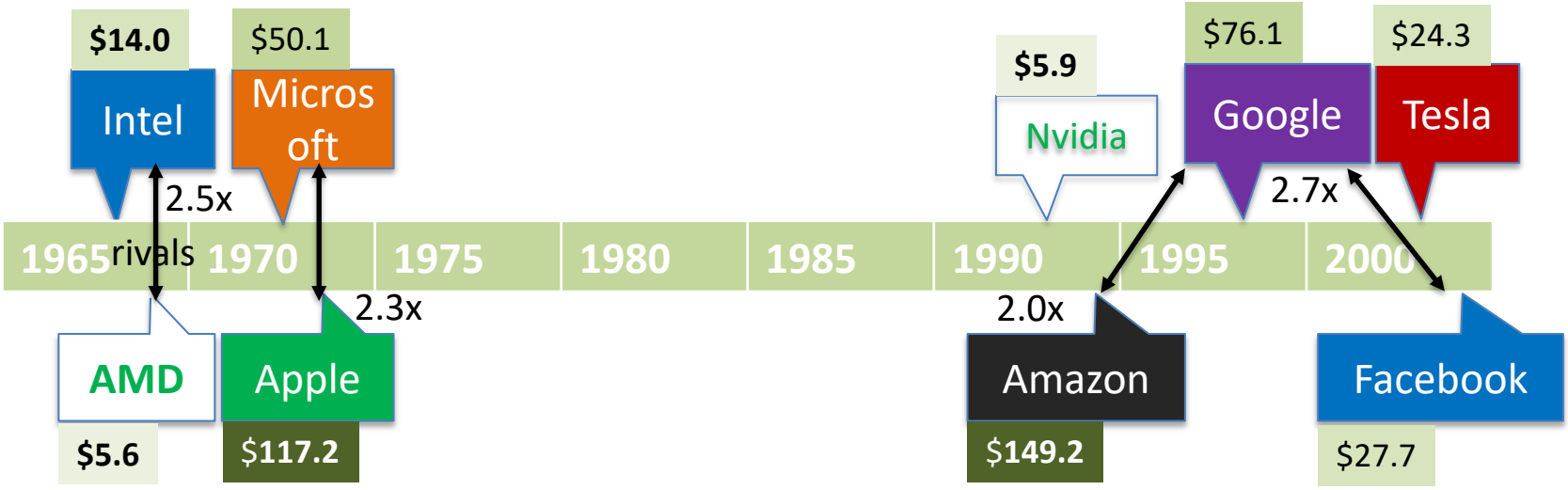


- Other *Industrials*
  - **GM** $36 → 2x Tesla
  - **IBM** $14.2 → ~Intel
  - QCOM $10.7
  - NXPI $3.3

- Other *Services*
  - **Netflix** $8.0
  - **PayPal** $6.8

Tech Titan Timeline

Qtr Revenue in $B | Current Perspective | **Q4** 2022



- Other *Industrials*
  - **GM** $43.1 → 2x Tesla
  - **Ford** $37.2
  - **IBM** $16.7 → ~Intel
  - QCOM $9.5
  - TI $4.2
  - NXPI $3.3

- Other *Services*
  - **Netflix** $8.0
  - Visa $7.9
  - **PayPal** $7.4

ISA/SoC Landscape

CPU & GPU Cores

- MIPS — Focus
- ARM
  - **Apple** — Mobile
  - Qualcomm
  - Samsung*
  - *Google*

- x86 — Laptop, Desktop, Server
  - Intel*
  - AMD
- ARM
  - Apple
  - *Nvidia*

*has own fab

- iOT
  - *Amazon*
  - *Google*
- Cars
  - *Tesla*

- Foundry (mfr)
  - TSMC
  - Samsung
  - GlobalFoundries (AMD)
  - Intel (new)

AMD vs Intel: CPU Families

Market Segment	AMD	Intel
Desktop	Ryzen 4K/ 5K	Core i3/5/7/9 (12th gen)
Laptop	Ryzen 4000	Ice Lake
Gaming	Ryzen Threadripper +Radeon	Core Extreme
Server/Workstn	Epyc	Xeon

According to the company, the AMD Ryzen 4700 G series desktop processor offers up to 2.5x multi-threaded performance compared to the previous generation, up to 5% greater single-thread performance than the Intel Core i7-9700, up to 31% greater multithreaded performance than the Intel Core i7-9700, and up to 202% better graphics performance than the Intel Core i7-9700.

AMD vs Intel

Quora

Drazen Zoric · Follow
Lives in Cork, Ireland · 5h

Overall x86 CPU Share (ALL CPUs)

Overall x86 CPU Share Includes IoT and SoC	2022 Q1 Current Quarter	2021 Q4 Prior Quarter	2021 Q1 Year Ago Quarter
	Share	Share	Share
Intel	72.3%	74.4%	79.3%
AMD	27.7%	25.6%	20.7%
VIA	0.0%	0.0%	0.0%
Total	100.0%	100.0%	100%

AMD vs Intel

Quora

Drazen Zoric · Follow
Lives in Cork, Ireland · 5h

Server CPU Share excluding IoT

Server CPU Share	2022 Q1 Current Quarter	2021 Q4 Prior Quarter	2021 Q1 Year Ago Quarter	Share Change (points)	Share Change (points)
	Share	Share	Share	Quarter	Year
Intel	88.4%	89.3%	91.1%	- 0.9	- 2.7
AMD	11.6%	10.7%	8.9%	+ 0.9	+ 2.7
Total	100.0%	100.0%	100.0%		

Desktop CPU Share excluding IoT

Desktop PC CPU Share	2022 Q1 Current Quarter	2021 Q4 Prior Quarter	2021 Q1 Year Ago Quarter	Share Change (points)	Share Change (points)
	Share	Share	Share	Quarter	Year
Intel	81.7%	83.8%	80.6%	- 2.1	+ 1.1
AMD	18.3%	16.2%	19.3%	+ 2.1	- 1.0
VIA	0.0%	0.0%	0.1%	+ 0.0	- 0.0
Total	100.0%	100.0%	100.0%		

AMD vs Intel

Quora

Drazen Zoric · Follow
Lives in Cork, Ireland · 5h

Mobile CPU Share excluding IoT

Mobile CPU Share	2022 Q1	2021 Q4	2021 Q1	Share	Share
	Current Quarter	Prior Quarter	Year Ago Quarter	Change (points)	Change (points)
	Share	Share	Share	Quarter	Year
Intel	77.5%	78.4%	82.0%	- 0.9	- 4.4
AMD	22.5%	21.6%	18.0%	+ 0.9	+ 4.4
Total	100.0%	100.0%	100.0%		

Yeah, Intel lost 2 - 7% but still sells 4 - 8 times more CPUs. AMD will never be able to close this gap. Things in Intel changed with 12th gen when they have again fastest CPUs. AMD Zen4 might take a lead but in few months 13th gen is out which will be better.
I saw ridiculous AMD Zen4 pricing, insane $800 for 7950X. If Intel lowers 13900 it will regain share.

Next year when Intel switches to HA EUV, 14th gen, it will have also better laptop CPUs which will be lower power.

Question is what is going on with Sapphire Rapids. It has 12 respins and still some 500 bugs. AMD already released Epyc Genoa with 96 cores what will threaten Intel in server and already did in supercomputer segments.

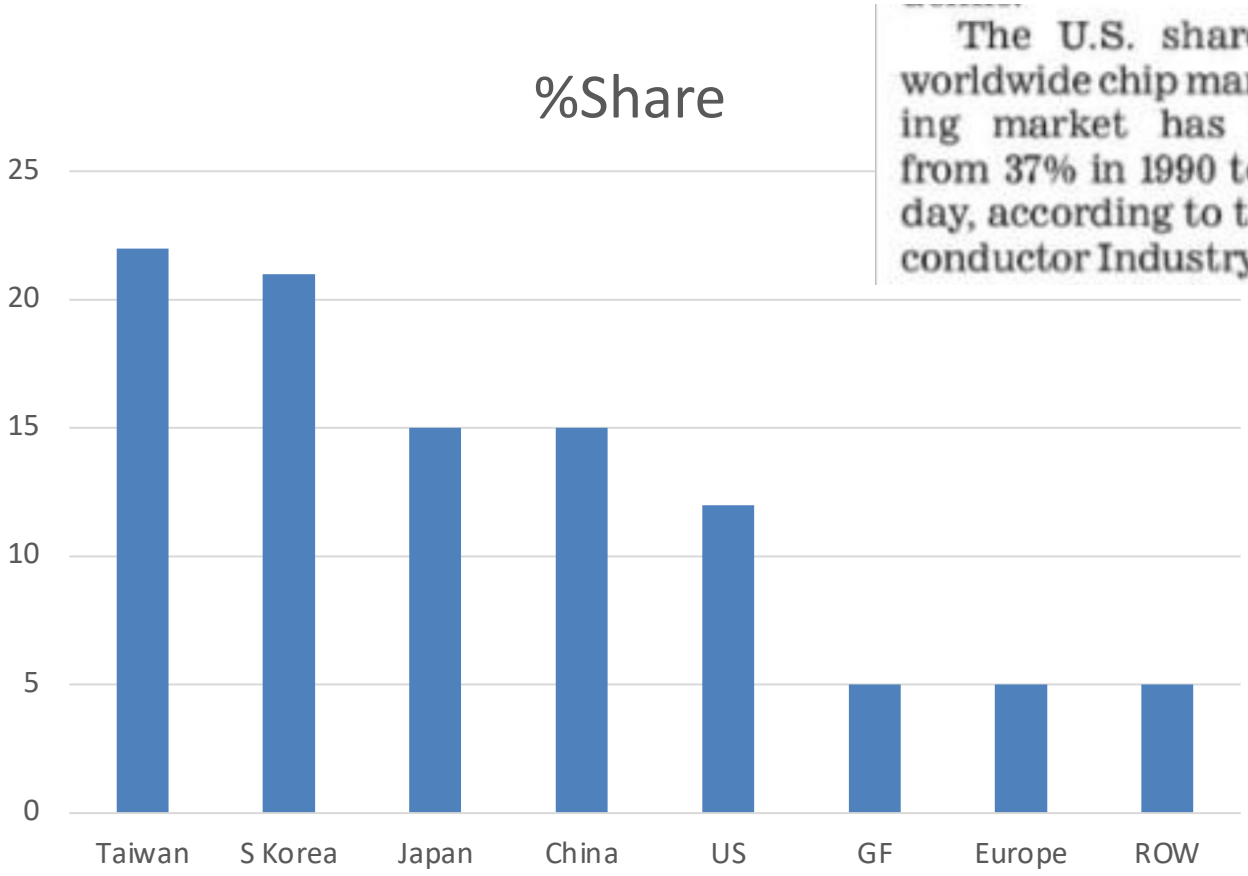
CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

CPU Leaders: AMD vs Intel

Apple Segments

	Q1 2021	Q4 2021	Q1 2022	y/y % change	q/q % change
iPhone Revenue $M	65597	38868	71628	9.19%	84.29%
Mac Revenue $M	8675	9178	10852	25.10%	18.24%
iPad Revenue $M	8435	8252	7248	-14.07%	-12.17%
Wearables, H&A Revenue $M	12971	8785	14701	13.34%	67.34%
Services Revenue $M	15761	18277	19516	23.82%	6.78%
Total Revenue $M	111439	83360	123945	11.22%	48.69%

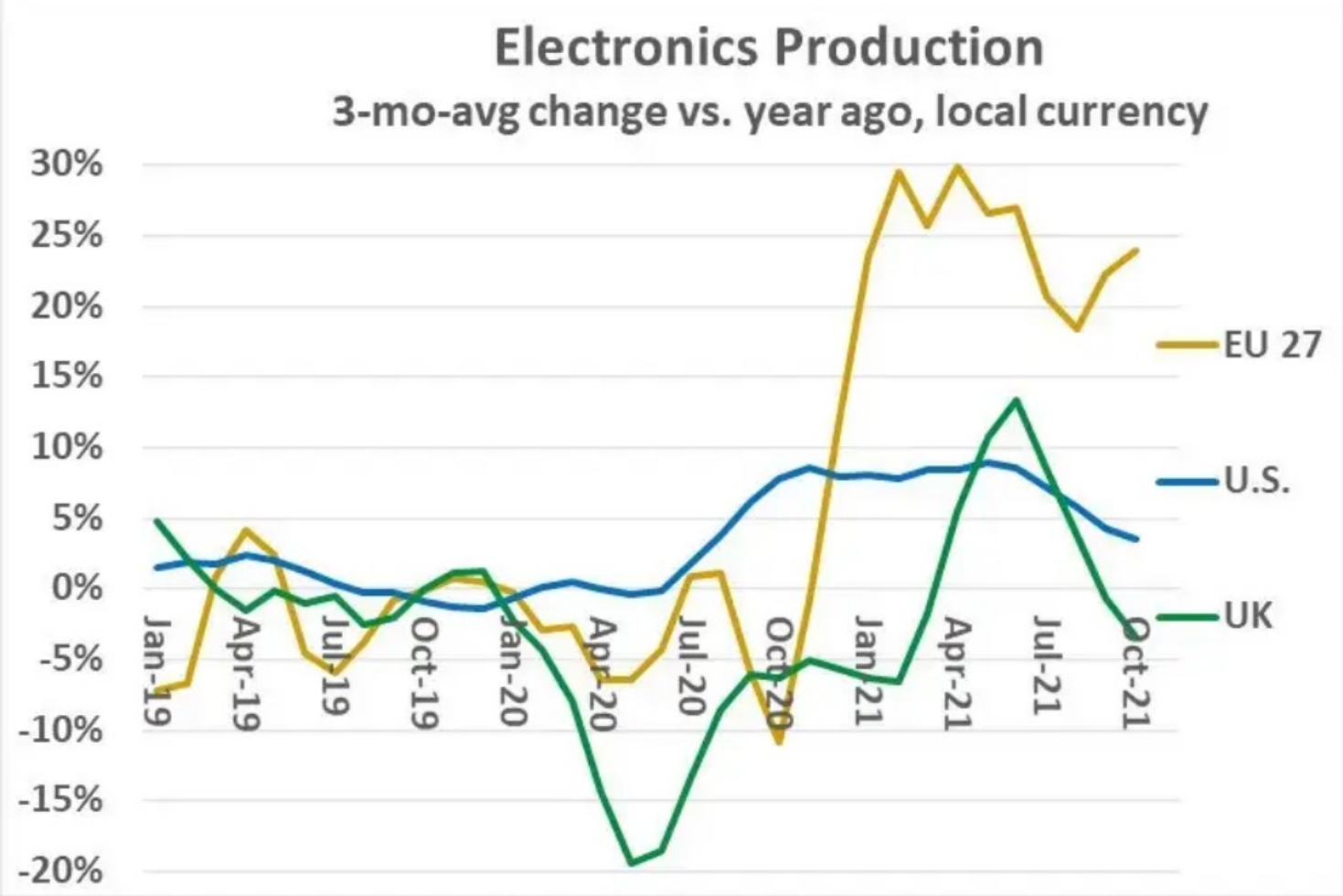
WW Fab Share by Region



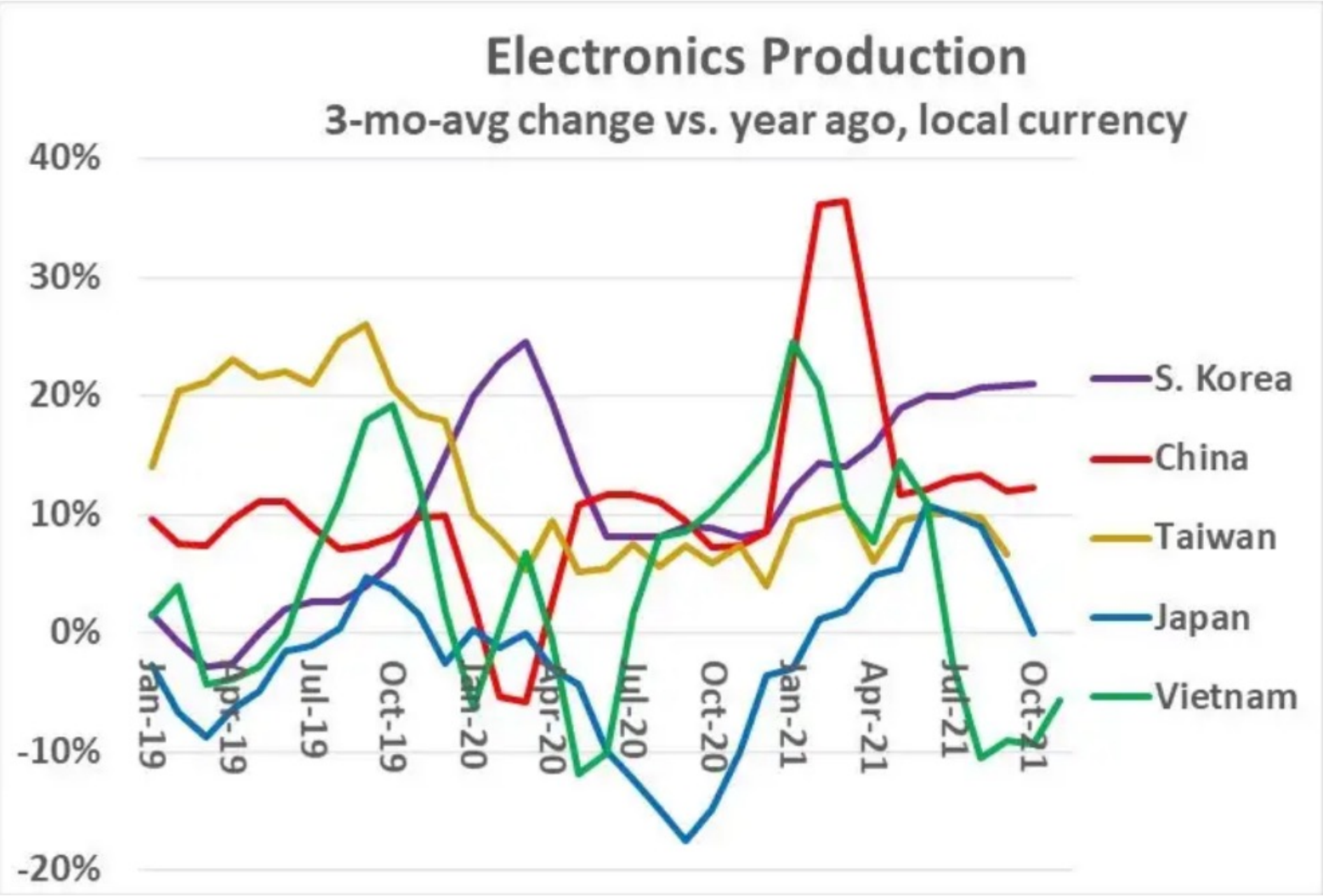
The U.S. share of the worldwide chip manufacturing market has declined from 37% in 1990 to 12% today, according to the Semiconductor Industry Assn.

Source: LA Times/SIA 1/22/22

Industry Regions

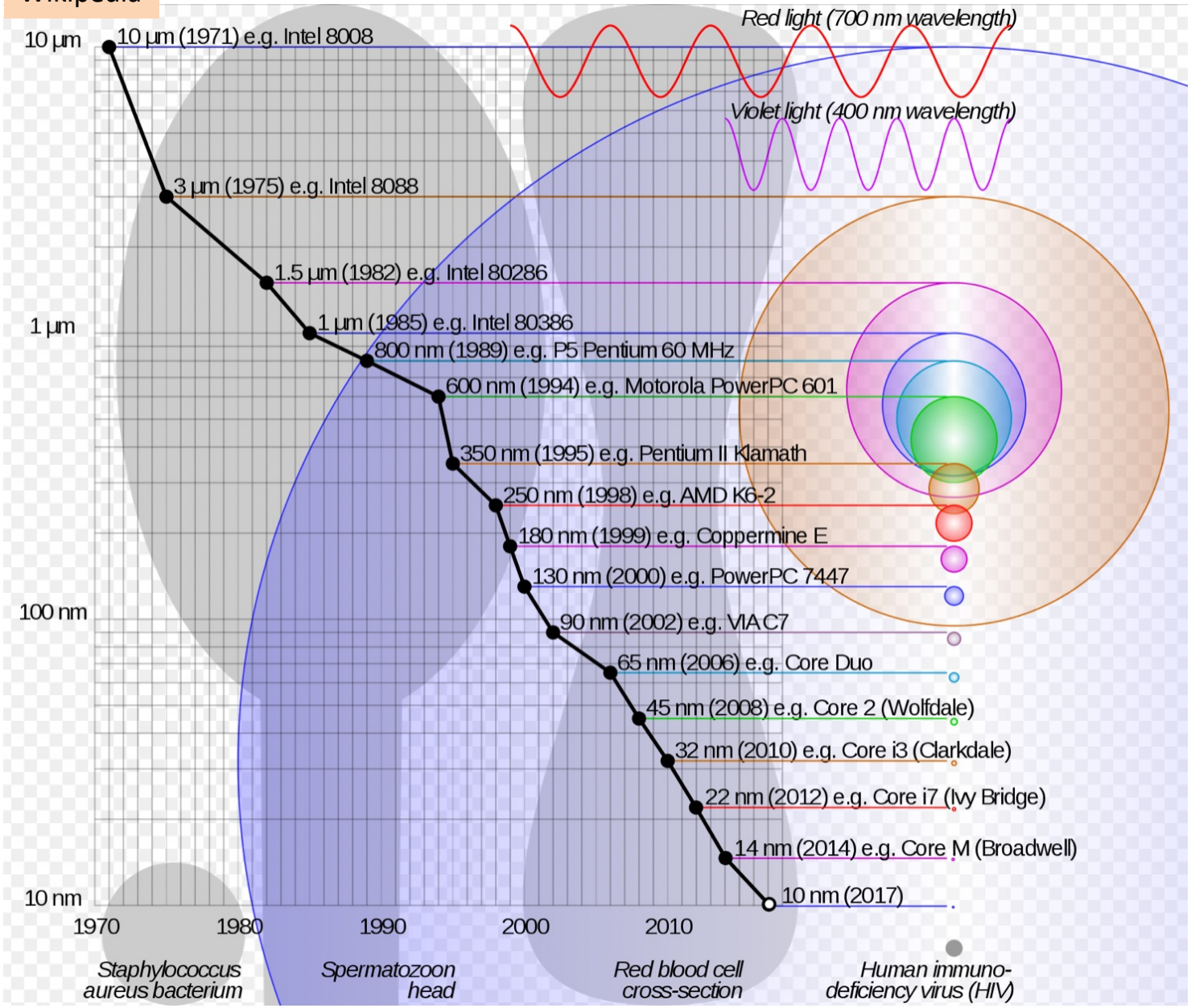


Asia Growth



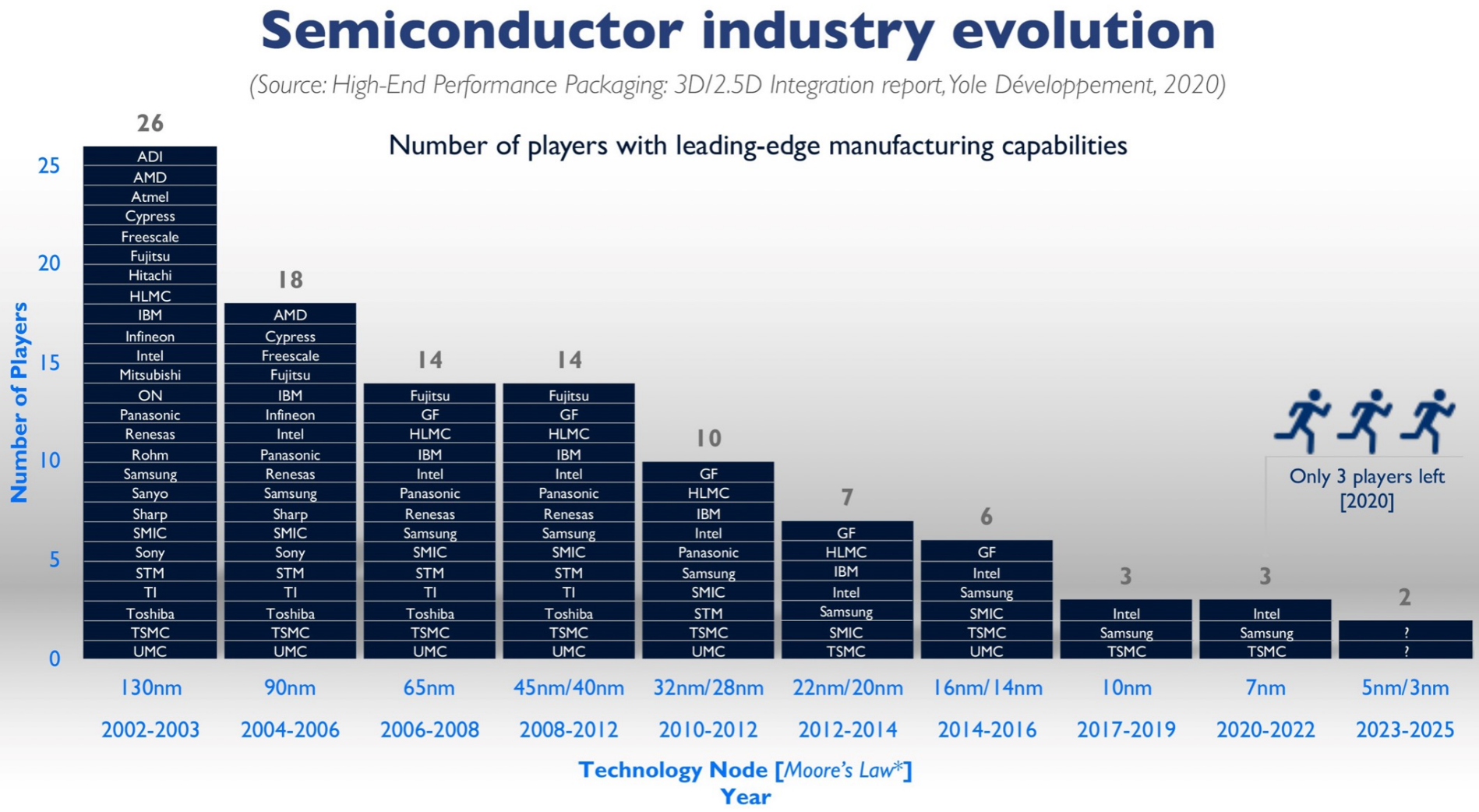
Process Timeline: 10um→10nm

Wikipedia



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Fab Timeline



Chip *Design* Leaders

Table 1: Global Top Ten IC Design Company Revenue Ranking, 1Q22 (Unit: US$1 Million)

1Q22 Rank	1Q21 Rank	Company	1Q22 Revenue	1Q21 Revenue	YoY
1	1	Qualcomm	9,548	6,281	52%
2	2	NVIDIA	7,904	5,173	53%
3	3	Broadcom	6,110	4,849	26%
4	5	AMD	5,887	3,445	71%
5	4	MediaTek	5,007	3,805	32%
6	9	Marvell	1,412	821	72%
7	6	Novatek	1,281	929	38%
8	8	Realtek	1,044	822	27%
9	-	Will Semiconductor	744	815	-9%
10	-	Cirrus Logic	490	294	67%
	7	Xilinx	-	851	-
-	10	Dialog	-	366	-
Total Revenue			39,427	27,342	44%

Notes

1. This top ten ranking only accounts for companies ahead of public financial reporting.
2. Qualcomm revenue only includes QCT; NVDIA excludes OEM/IP revenue; Broadcom revenue only includes semiconductors; Will Semiconductor revenue only includes semiconductor design and sales.
3. NT$:US$ exchange rate: 1Q22 - 28.50:1; 1Q21 - 28.39:1
4. RMB:US$ exchange rate: 1Q22 - 6.336:1; 1Q21 - 6.483:1

Source: TrendForce, Jun. 2022

Chip *Sales* Leaders

2Q21 Top 10 Semiconductor Sales Leaders ($M, Including Foundries)

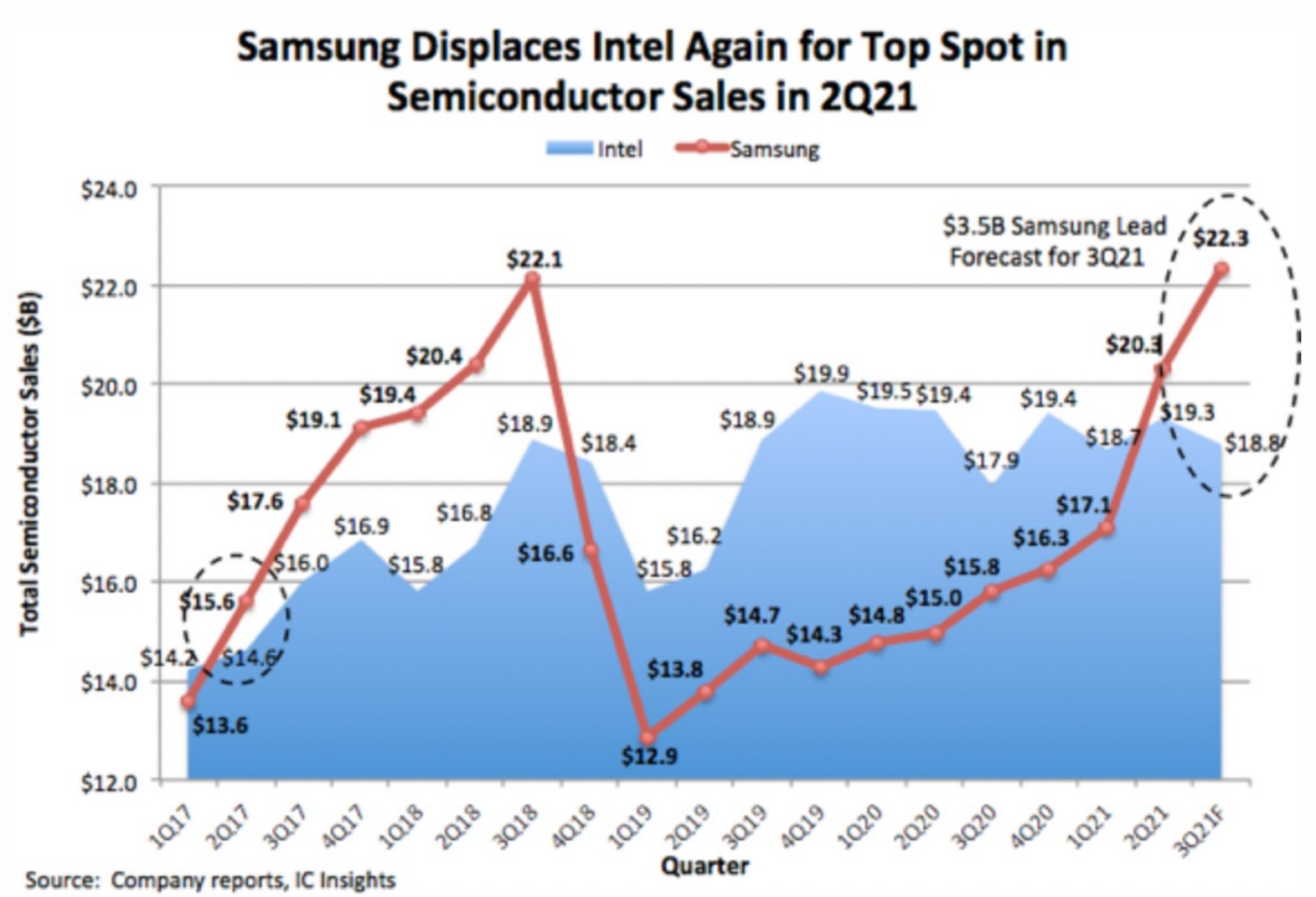
2Q21 Rank	1Q21 Rank	Company	Headquarters	1Q21 Total IC	1Q21 Total O-S-D	1Q21 Total Semi	2Q21 Total IC	2Q21 Total O-S-D	2Q21 Total Semi	2Q21/1Q21 % Change
1	2	Samsung	South Korea	16,152	920	17,072	19,262	1,035	20,297	19%
2	1	Intel	U.S.	18,676	0	18,676	19,304	0	19,304	3%
3	3	TSMC (1)	Taiwan	12,911	0	12,911	13,315	0	13,315	3%
4	4	SK Hynix	South Korea	7,270	358	7,628	8,762	451	9,213	21%
5	5	Micron	U.S.	6,629	0	6,629	7,681	0	7,681	16%
6	6	Qualcomm (2)	U.S.	6,281	0	6,281	6,472	0	6,472	3%
7	8	Nvidia (2)	U.S.	4,842	0	4,842	5,540	0	5,540	14%
8	7	Broadcom Inc. (2)	U.S.	4,364	485	4,849	4,400	490	4,890	1%
9	10	MediaTek (2)	Taiwan	3,849	0	3,849	4,496	0	4,496	17%
10	9	TI	U.S.	3,793	235	4,028	4,030	269	4,299	7%
—	—	Top-10 Total		84,767	1,998	86,765	93,262	2,245	95,507	10%

(1) Foundry (2) Fabless

Source: Company reports, IC Insights' *Strategic Reviews* database

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Chip Sales: Samsung vs Intel



Microprocessor Leaders

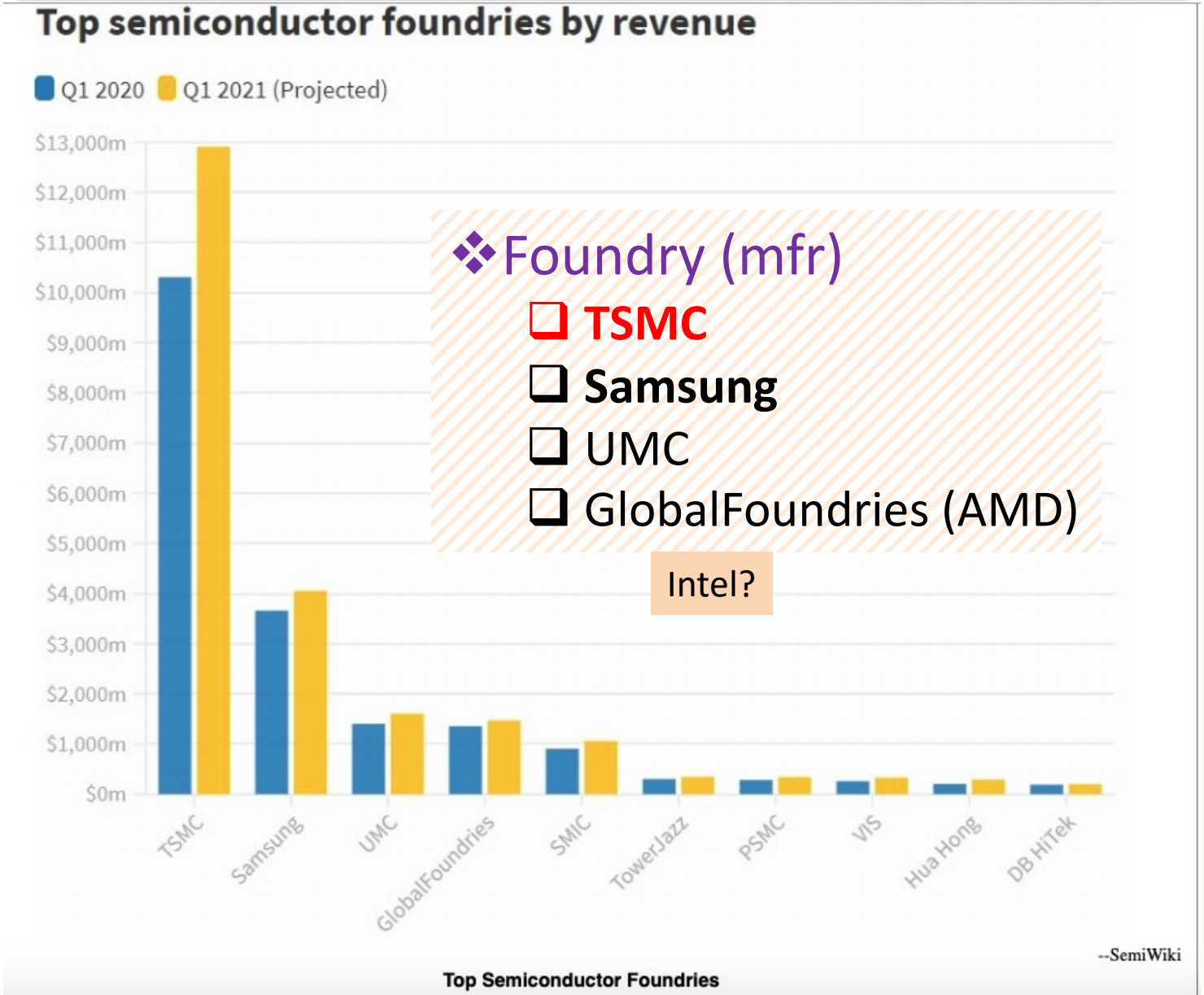
Leading MPU Suppliers ($B)

2021 Rank	Company	Headquarters	2020	2021	21/20 % Chg	2021 % Marketshare
1	Intel	U.S.	50.6	52.3	3%	50.9%
2	Apple*	U.S.	10.5	13.4	27%	13.0%
3	Qualcomm	U.S.	7.4	9.4	26%	9.1%
4	AMD	U.S.	5.9	9.2	56%	8.9%
5	MediaTek	Taiwan	2.7	4.1	51%	4.0%

*Custom designs for Apple's products that are made by IC foundries.

Source: Company reports, IC Insights

TSMC



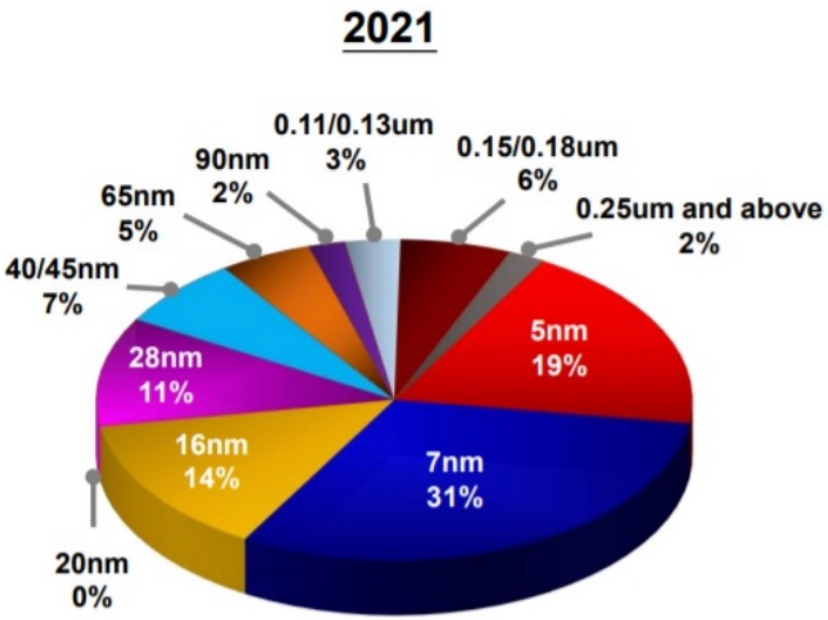
- Foundry (mfr)
  - **TSMC**
  - **Samsung**
  - UMC
  - GlobalFoundries (AMD)

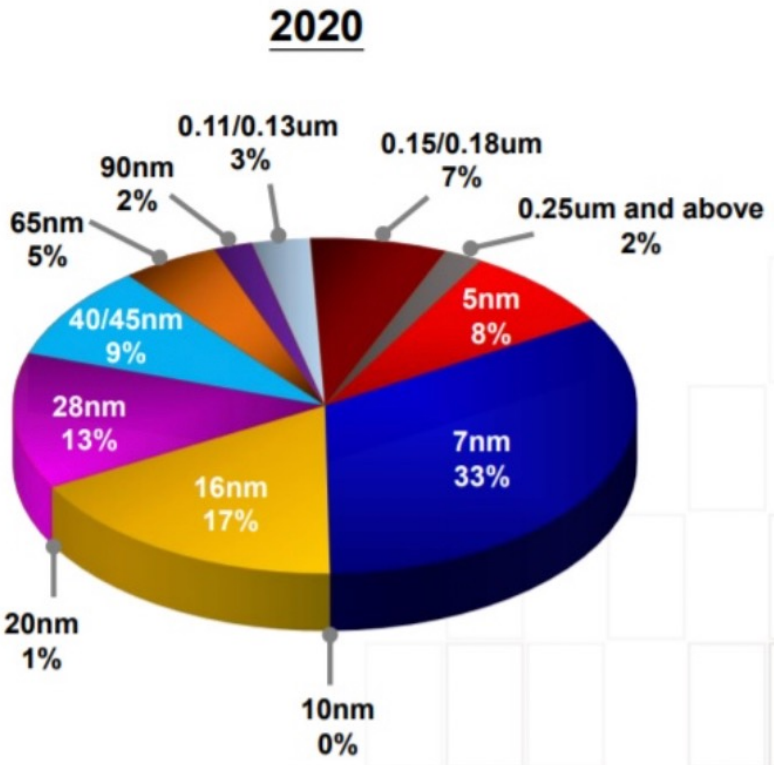
Intel?

TSMC Node Segments

Semi Wiki

Revenue by Technology

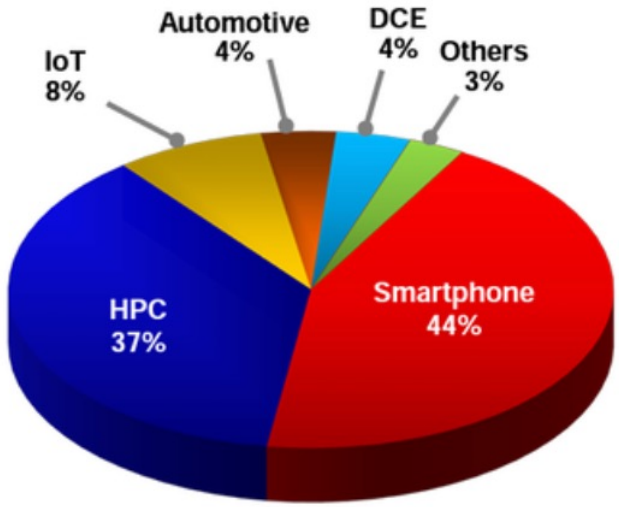


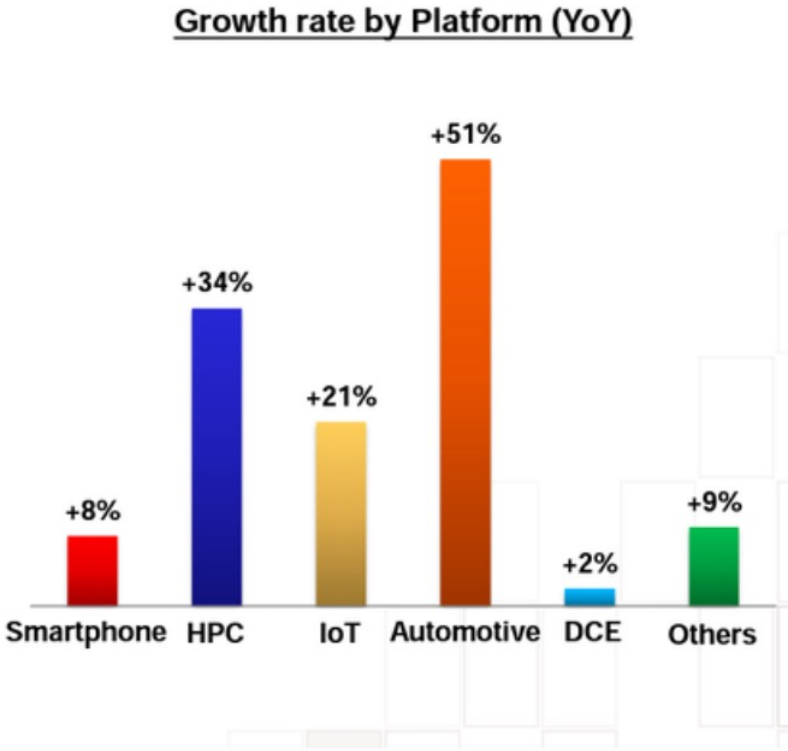


TSMC Revenue Segments

Semi Wiki

2021 Revenue by Platform





TSMC's current business mix

Samsung Processes

Logic

IC Knowledge

Samsung Keynote at IEDM
by Scotten Jones on 01-27-2022 at 6:00 am

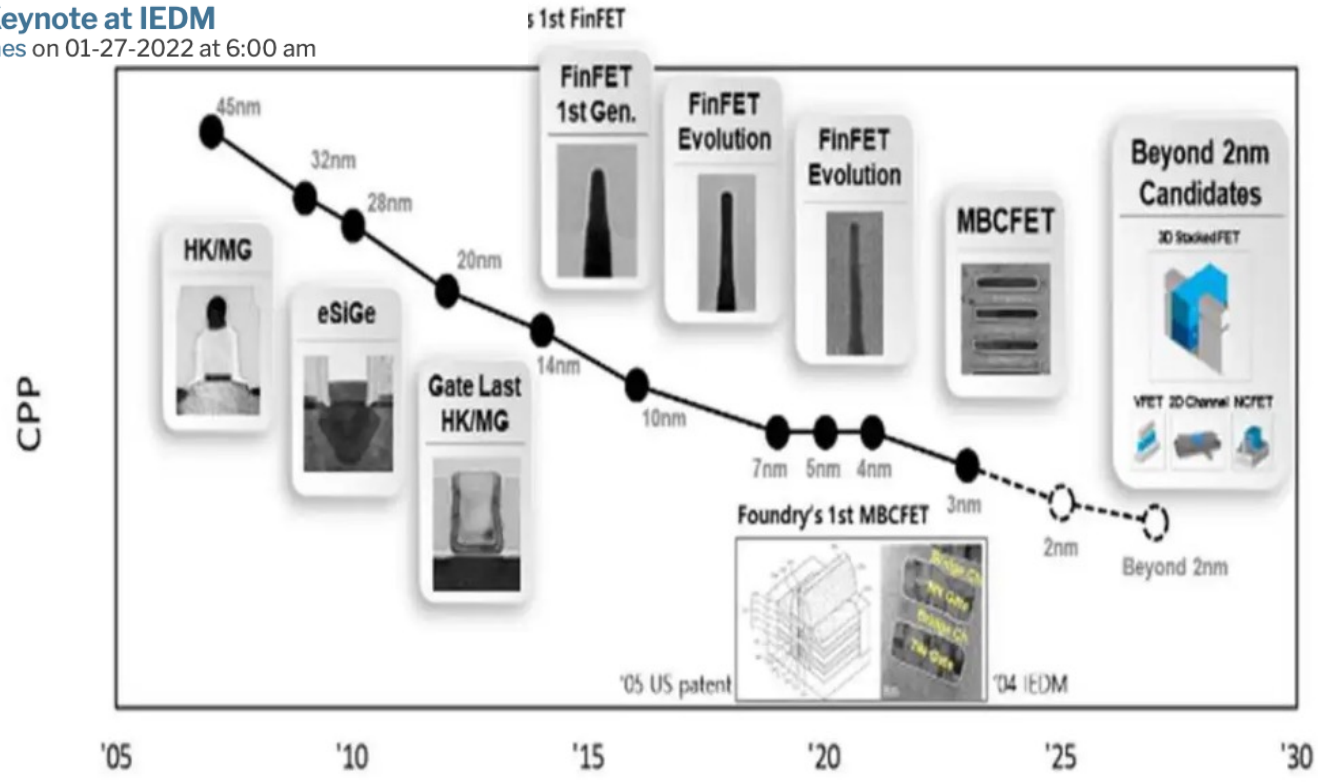


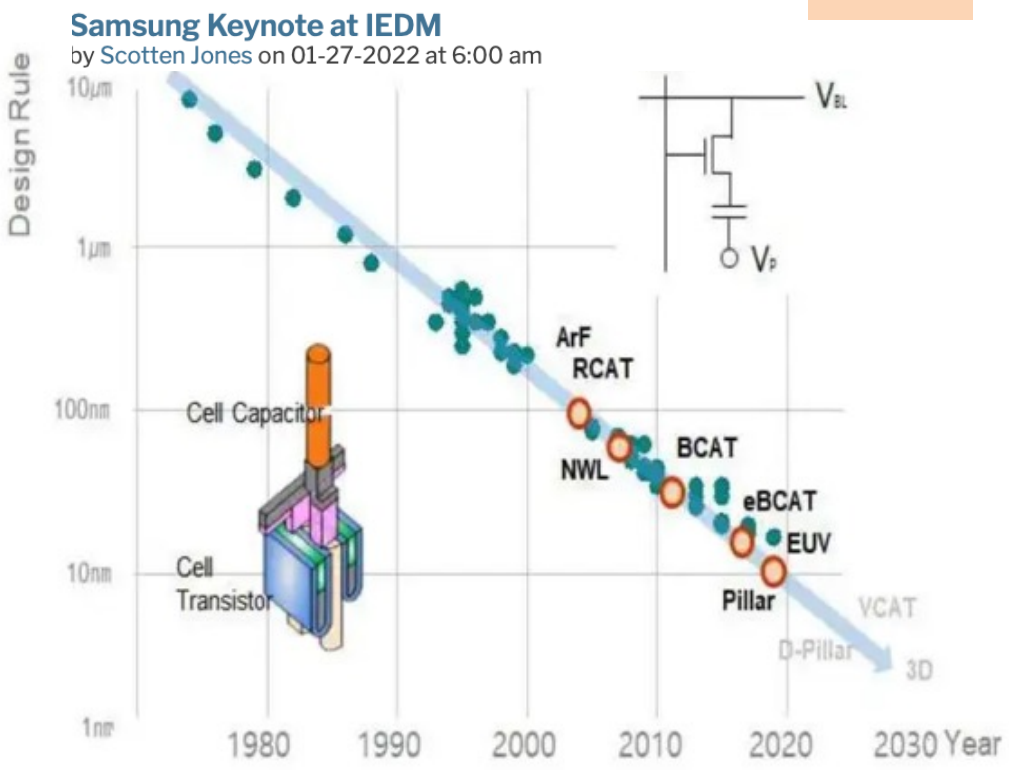
Figure 1. Logic Roadmap.

n figure 1 we can see how the contacted poly pitch (CPP) of logic processes has caled over time. In the planar era we saw high-k metal gate (HKMG) introduced by ntel at 45nm and by the foundries at 28nm as well as innovations like embedded

Samsung Processes

IC Knowledge

DRAM



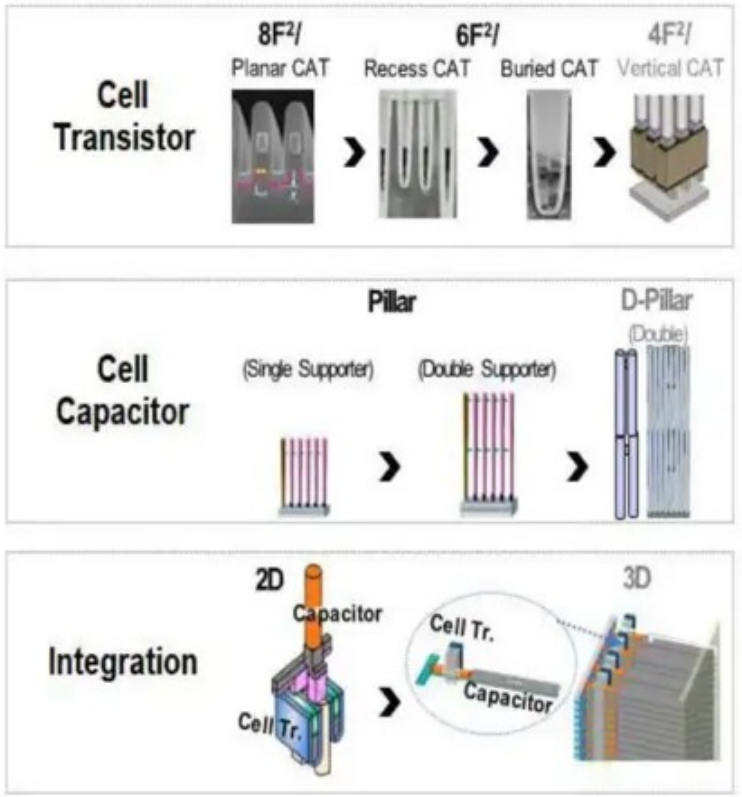


Figure 2 DRAM Roadmap

With EUV already ramping up in DRAM, the next challenges are shrinking the memory cell. Samsung is anticipating staking two layers of capacitors soon. A switch

Samsung Processes

IC Knowledge

NAND (Flash)

Samsung Keynote at IEDM
by Scotten Jones on 01-27-2022 at 6:00 am

· NAND.

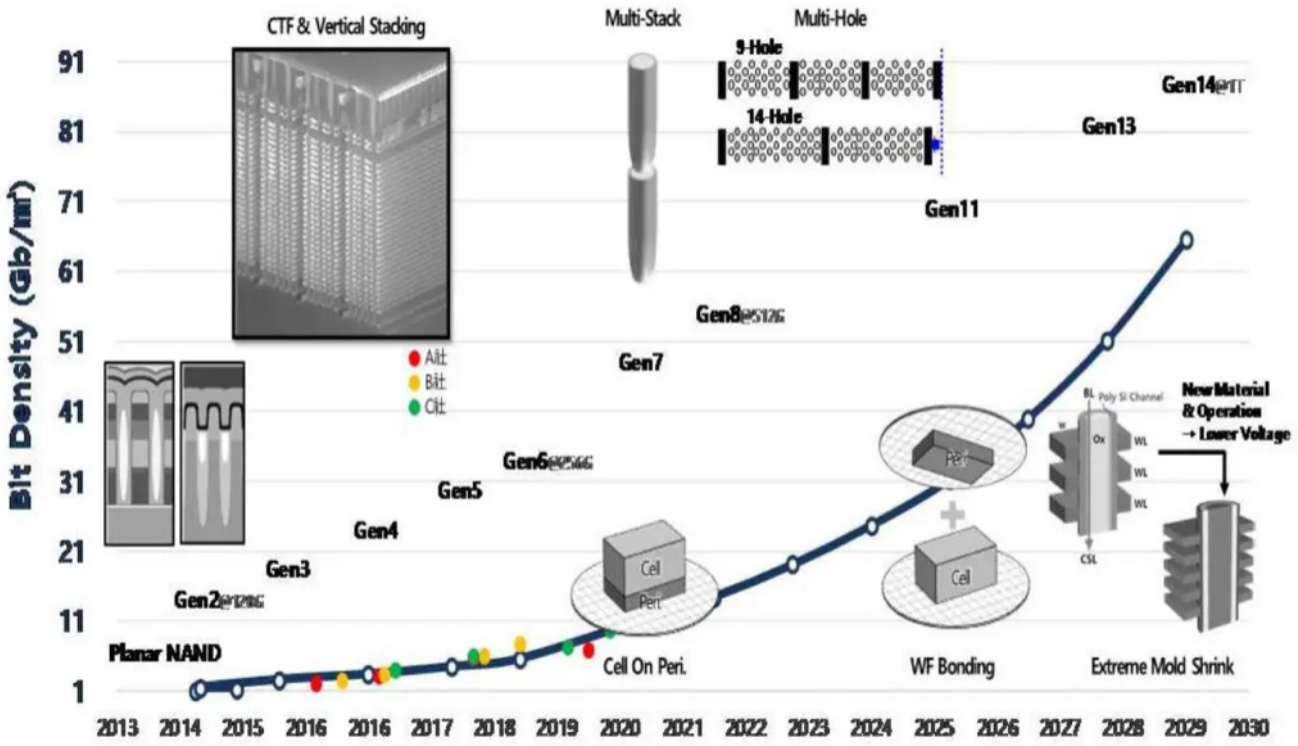
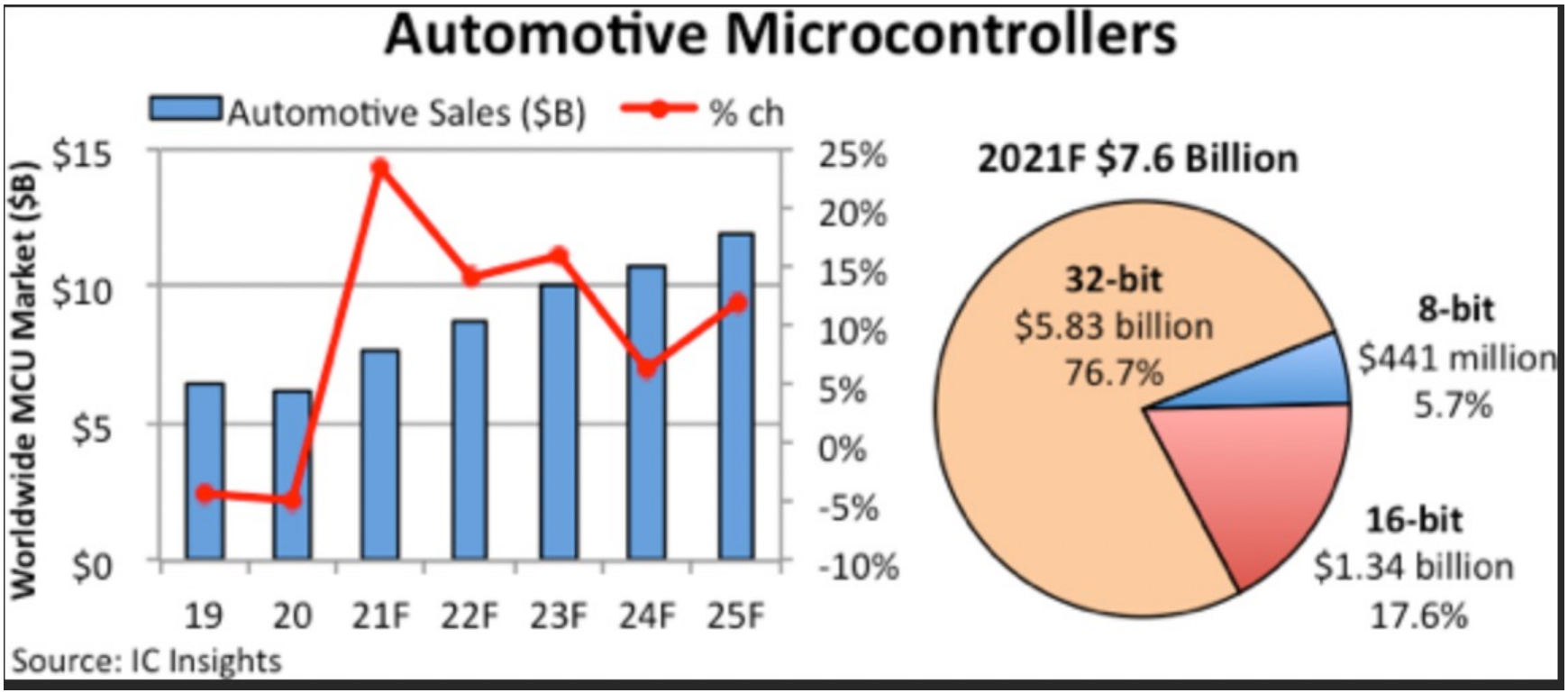


Figure 3 NAND Roadmap

Samsung's latest 3D NAND is a 176-layer process that uses string stacking for the first time (first time string stacking for them, others have been string stacking for multiple generations) and peripheral under the array for the first time (once again the

MCU in Automotive



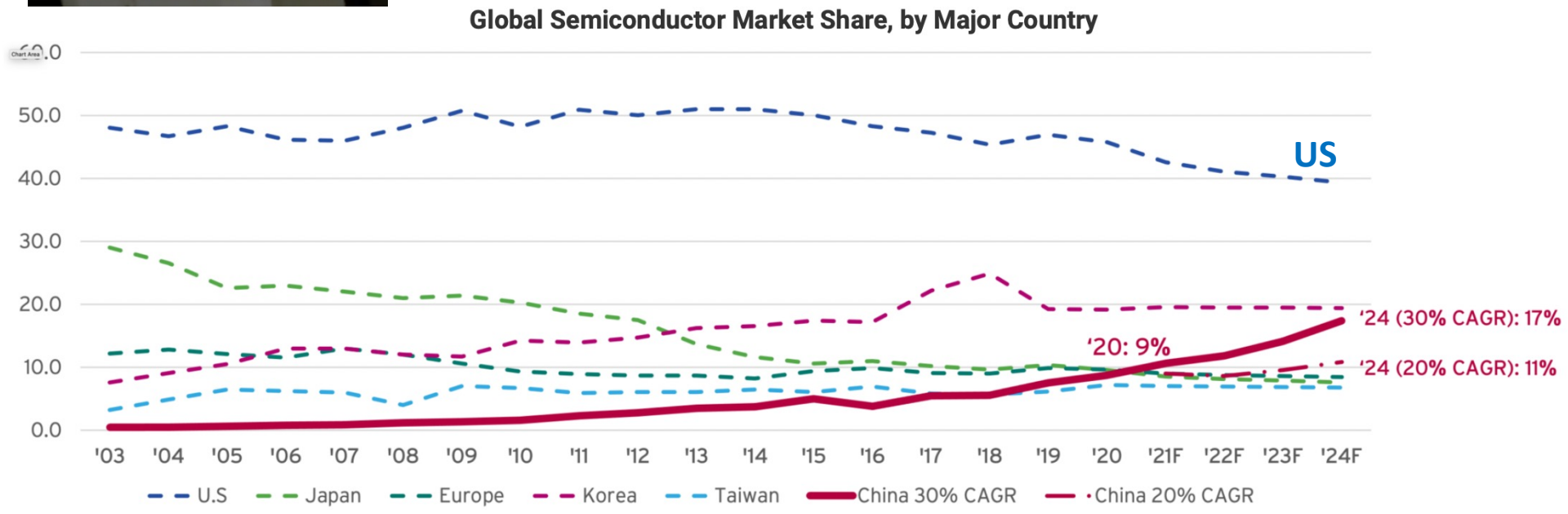
US Semi Market Share

SIA SEMICONDUCTOR INDUSTRY ASSOCIATION

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Global Market Shares

SIA SEMICONDUCTOR INDUSTRY ASSOCIATION



Source: Company financials, SIA analysis, WSTS, Omida

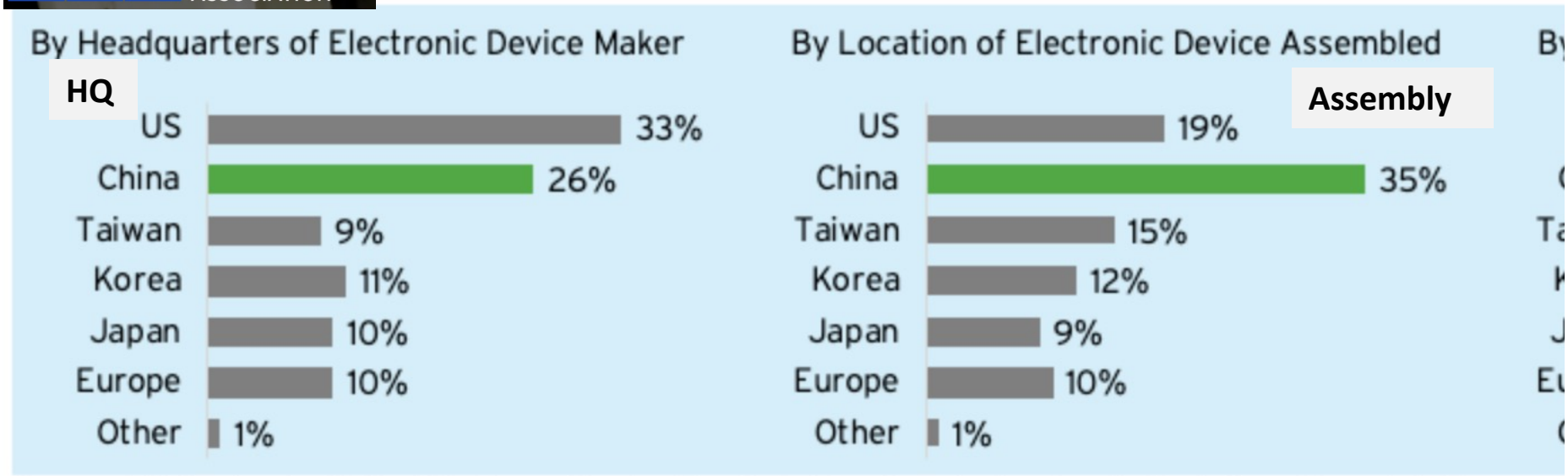
CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

COMP122

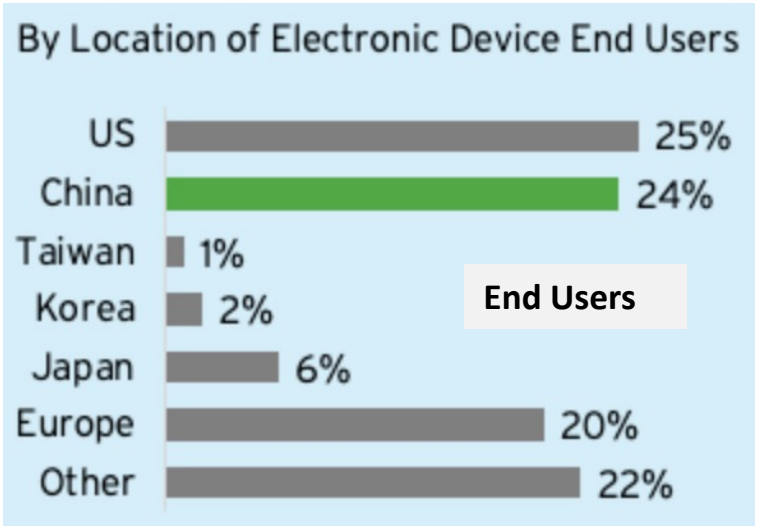
Global Chip Sales Segments

SIA SEMICONDUCTOR INDUSTRY ASSOCIATION

Global Semiconductor Sales by Geographic Area, 2019 (%)



Source: UN Comtrade; BCG x SIA: Strengthening the Global Semiconductor Supply Chain in an Uncertain Era



Computer Architecture

Computer History

- See separate slide set on ***History of Tech*** Vol 1

Father of Computers

Quora

Who was the real father of the modern computer, Alan Turing or John von Neumann? Why?

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

Answered just now

I choose von Neumann, in that we have used the "von Neumann" architecture as the basic architecture since the 1st stored program digital computer in 1948 the EDVAC. Turing defined an automata theory, but not an architecture — although he did contribute to the design of the UK's Colossus computer ca 1944.

1st Computer

Quora

How and when did digital computers come into existence?

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

Charles Babbage designed his own *computer* -- more of a *calculator* -- in the 19th century as a steam powered "Analytical Engine" but did not have the electronics available to complete it as an "electronic computer" (or *calculator*).

the ***first computer*** is generally regarded as the first all-electronic, digital and programmable computer, ENIAC in 1944. UK's Colossus also a close 2nd. ENIAC evolved into the the 1st commercial computers, the "UNIVAC" line.
they used vacuum tubes and relays for logic, mag drums and mag core memory (mag disk in 1954), and cables for programming (later punched cards).

ENIAC was designed at the U of Penn by Mauchly and Eckert. its successor was EDVAC in 1948 based on John von Neumann's stored program, unified memory architecture. we still refer to that basic architecture as a "von Neumann" architecture. most of the basic design philosophy of a program stored in a unified memory has continued to be used to this day.

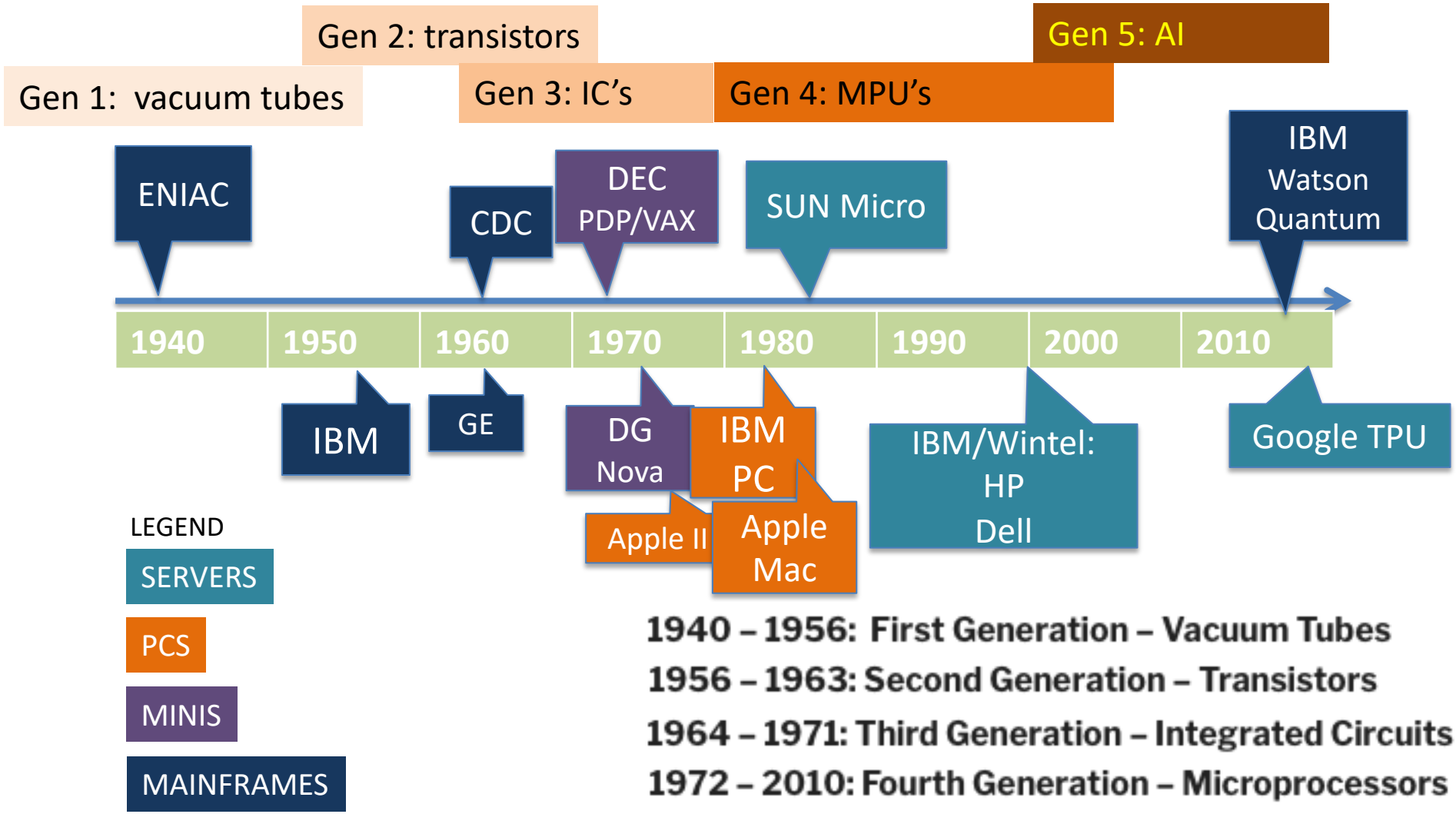
the "first" computer, ENIAC, in 1944 was programmed with patch cables. in the next decade, the 1950's, IBM style (Hollerith) punch cards were used. a programmer used a keypunch machine to punch out assembly or high-level language code onto a deck of cards. the cards used a "card reader" to input the program into the computer. there had to also be some type of operating system to handle the card reader input and produce printed output.

at roughly the same time in the UK, in 1944, their government built the Colossus (Whirlwind) to decrypt the German Enigma code, with help from Alan Turing. but that

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Computer Generations

TIMELINE



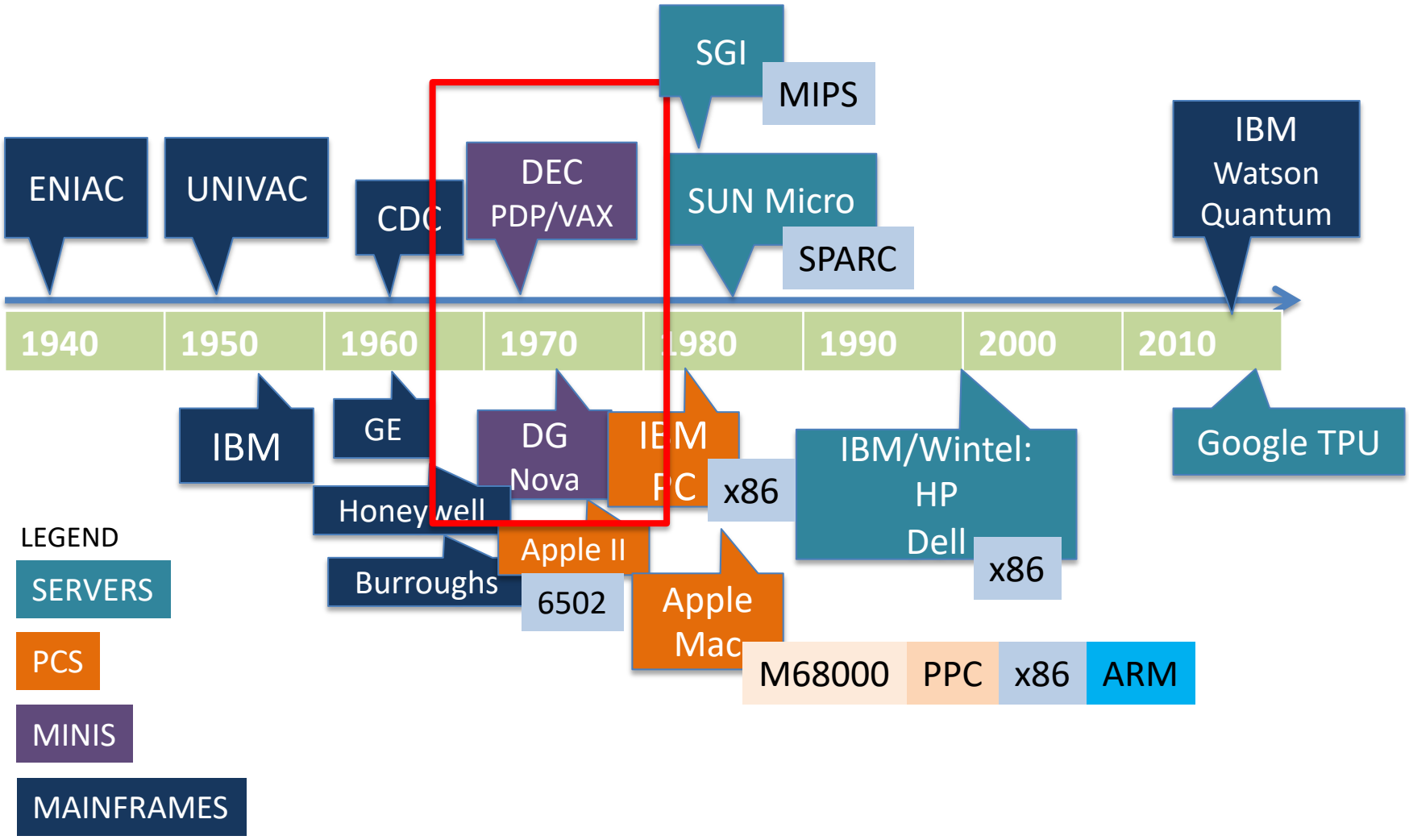
1940 – 1956: First Generation – Vacuum Tubes
1956 – 1963: Second Generation – Transistors
1964 – 1971: Third Generation – Integrated Circuits
1972 – 2010: Fourth Generation – Microprocessors
2010- : Fifth Generation – Artificial Intelligence

Computers

Minicomputers of the 1970's
Used Am2900 *bit-slices*

TIMELINE

Computers of the 1990's
Used RISC CPU's



Old Computer ISA's

WIKIPEDIA The Free Encyclopedia

Execute instruction

From Wikipedia, the free encyclopedia

In computer instruction set architecture (ISA), an **execute instruction** is a machine language instruction which treats data as a machine instruction and executes it.

It can be considered a fourth mode of instruction sequencing after ordinary sequential execution, branching, and interrupting.[1]

Computer models [edit]

V·T·E

Many computers designed in the 1960s included *execute* instructions: the IBM 7030 Stretch (mnemonic: `EX`, `EXIC`),[2][1] the IBM 709[1] and IBM 7090 (`XEC`),[3], the PDP-1 (`XCT`),[4] the CDC 924 (`XEC`),[5] the PDP-6/PDP-10 (`XCT`), the IBM System/360 (`EX`),[6] the GE-600/Honeywell 6000 (`XEC`, `XED`),[7] the SDS 9 Series (`EXU`).[8][9]

Fewer 1970s designs included execute instructions. An execute instruction was proposed for the PDP-11 in 1970,[10] but never implemented for it[11] or its successor, the VAX.[12] The Nuclear Data 812 minicopmuter (1971) includes an execute instruction (`XCT`).[13]

The TMS9900 microprocessor (1976) has an *execute* instruction (`X`).[14]

Modern processors do not include *execute* instructions because they interfere with pipelining and other optimizations.

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Old Computer ISA's

WIKIPEDIA
The Free Encyclopedia

Execute instruction

From Wikipedia, the free encyclopedia

Applications [edit]

The execute instruction has several applications:[1]

- Late binding
  - Implementation of call by name and other thunks.[1]
  - A table of execute targets may be used for dynamic dispatch of the methods or virtual functions of an object or class, especially when the method or function may often be implementable as a single instruction.[11]
  - An execute target may contain a hook for adding functionality or for debugging; it is normally initialized as a NOP which may be overridden dynamically.
  - An execute target may change between a fast version of an operation and a fully traced version.[17][18][19]
- Tracing, monitoring, and emulation
  - This may maintain a pseudo-program counter, leaving the normal program counter unchanged.[1]
- Executing dynamically generated code, especially when memory protection prevents executable code from being writable.
- Emulating self-modifying code, especially when it must be reentrant or read-only.[10]

DEC PDP/VAX

DEC PDP-11

1970 Wiki

PDP-11

From Wikipedia, the free encyclopedia
(Redirected from DEC PDP-11)

This article is about the PDP-11 series of minicomputers. For the PDP-11 processor architecture, see PDP-11 archit

The **PDP-11** is a series of 16-bit minicomputers sold by Digital Equipment Corporation (DEC) from 1970 into the 1990s, one of a succession of products in the PDP series. In total, around 600,000 PDP-11s of all models were sold, making it one of DEC's most successful product lines. The PDP-11 is considered by some experts[1][2][3] to be the most popular minicomputer ever.

The PDP-11 included a number of innovative features in its instruction set and additional general-purpose registers that made it much easier to program than earlier models in the PDP series. Additionally, the innovative Unibus system allowed external devices to be easily interfaced to the system using direct memory access, opening the system to a wide variety of peripherals. The PDP-11 replaced the PDP-8 in many real-time applications, although both product lines lived in parallel for more than 10 years. The ease of programming of the PDP-11 made it very popular for general-purpose computing uses as well.

The design of the PDP-11 inspired the design of late-1970s microprocessors including the Intel x86[] and the Motorola 68000. Design features of PDP-11 operating systems, as well as other operating systems from Digital Equipment, influenced the design of other operating systems such as CP/M and hence also MS-DOS. The first officially named version of Unix ran on the PDP-11/20 in 1970. It is commonly stated that the C programming language took advantage of several low-level PDP-11–dependent programming features,[4] albeit not originally by design.[5]

An effort to expand the PDP-11 from 16 to 32-bit addressing led to the VAX-11 design, which took part of its name from the PDP-11.

DEC PDP-11

No dedicated I/O instructions [edit]

Early models of the PDP-11 had no dedicated bus for input/output, but only a system bus called the Unibus, as input and output devices were mapped to memory addresses.

An input/output device determined the memory addresses to which it would respond, and specified its own interrupt vector and interrupt priority. This flexible

Interrupts [edit]

The PDP-11 supports hardware interrupts at four priority levels. Interrupts are serviced by software service routines, which could specify whether they themselves could be interrupted (achieving interrupt nesting). The event that causes the interrupt is indicated by the device itself, as it informs the processor of the address of its own interrupt vector.

Interrupt vectors are blocks of two 16-bit words in low kernel address space (which normally corresponded to low physical memory) between 0 and 776. The first word of the interrupt vector contains the address of the interrupt service routine and the second word the value to be loaded into the PSW (priority level) on entry to the service routine.

Instruction set orthogonality [edit]

See also: PDP-11 architecture

The PDP-11 processor architecture has a mostly orthogonal instruction set. For example, instead of instructions such as *load* and *store*, the PDP-11 has a *move* instruction for which either operand (source and destination) can be memory or register. There are no specific *input* or *output* instructions; the PDP-11 uses memory-mapped I/O and so the same *move* instruction is used; orthogonality even enables moving data directly from an input device to an output device. More complex instructions such as *add* likewise can have memory, register, input, or output as source or destination.

Most operands can apply any of eight addressing modes to eight registers. The addressing modes provide register, immediate, absolute, relative, deferred (indirect), and indexed addressing, and can specify autoincrementation and autodecrementation of a register by one (byte instructions) or two (word instructions). Use of relative addressing lets a machine-language program be position-independent.

DEC PDP/VAX

DEC PDP-11

1st **LSI**-chip Computer | 1970 | Wiki

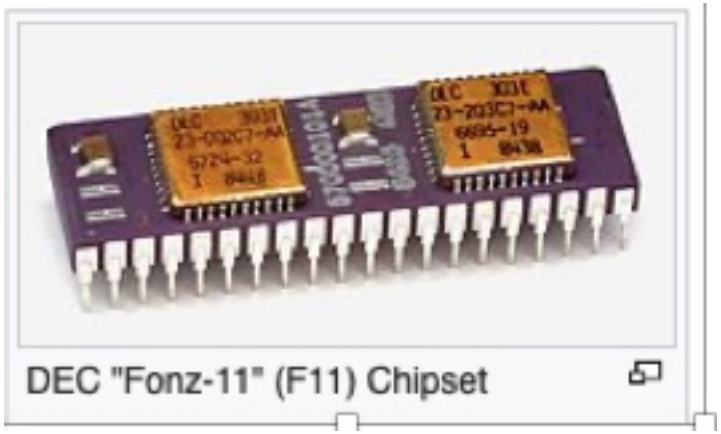
PDP-11/40. The processor is at the bottom. A TU56 dual DECtape drive is installed above it.

DEC PDP-11

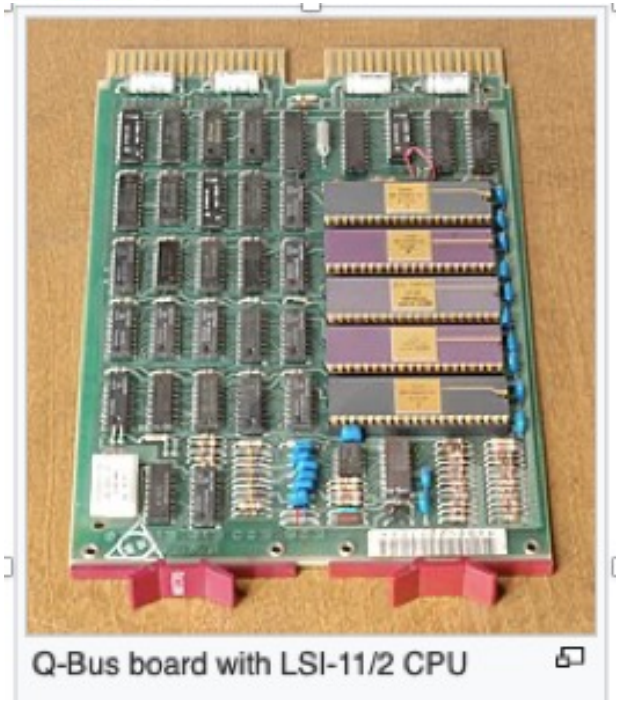
LSI-11 [edit]

The LSI-11 (PDP-11/03), introduced in February 1975[10] is the first PDP-11 model produced using large-scale integration; the entire CPU is contained on four LSI chips made by Western Digital (the MCP-1600 chip set; a fifth chip can be added to

The CPU microcode includes a debugger: firmware with a direct serial interface (RS-232 or current loop) to a terminal. This lets the operator do debugging by typing commands and reading octal numbers, rather than operating switches and reading lights, the typical debugging method at the time. The operator can thus examine and modify the computer's registers, memory, and input/output devices, diagnosing and perhaps correcting failures in software and peripherals (unless a failure disables the microcode itself). The operator can also specify which disk to boot from.



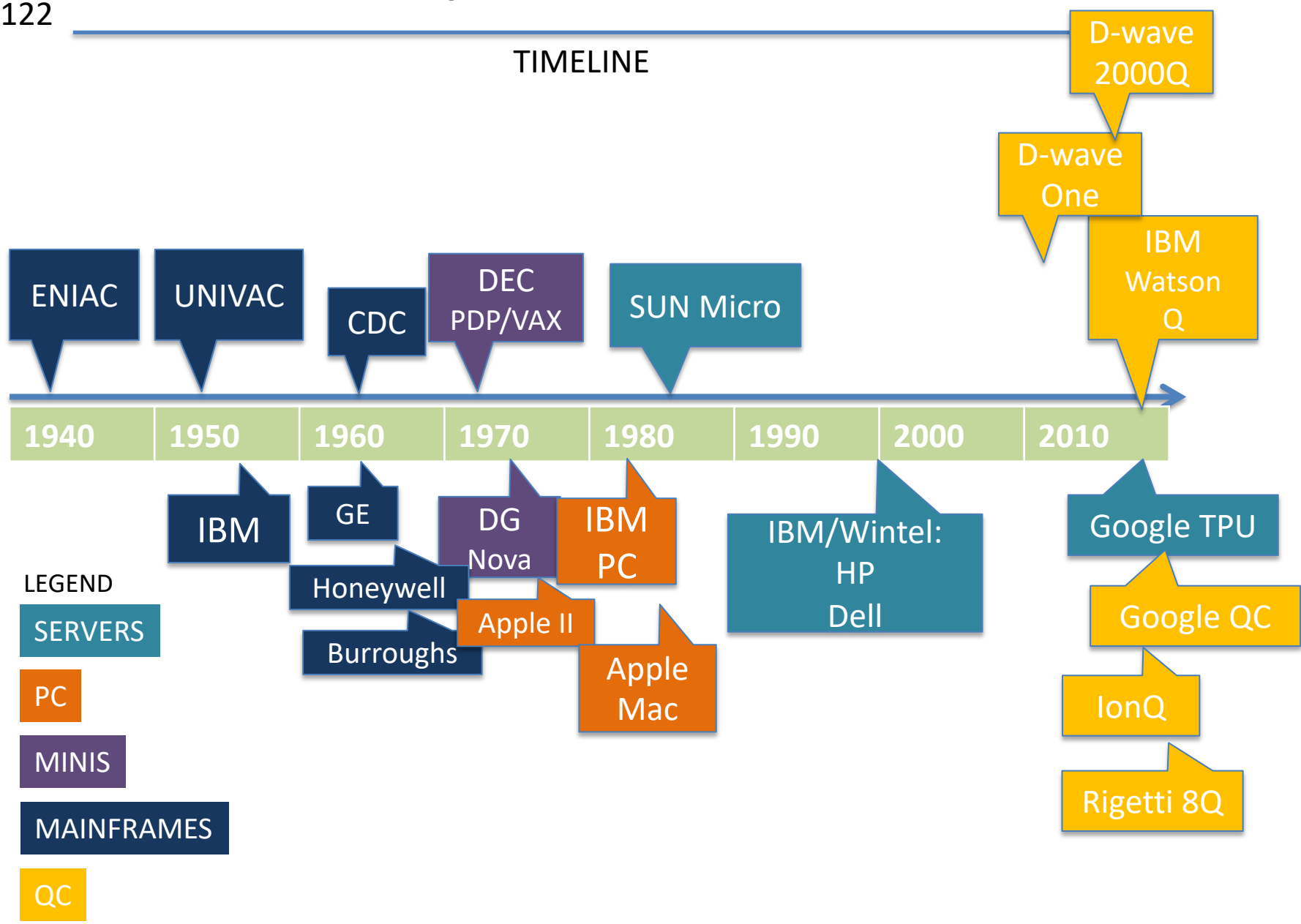
DEC "Fonz-11" (F11) Chipset



Q-Bus board with LSI-11/2 CPU

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Computers & QC's



Quantum Computers (QC)

- Google
- IBM
- Intel
- Microsoft

Outlook

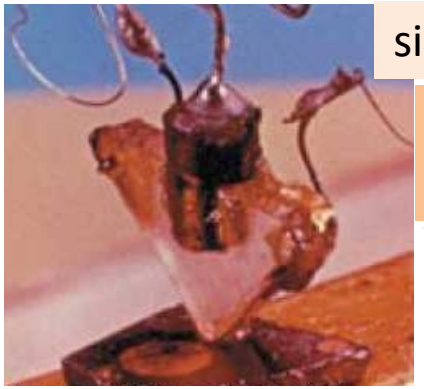
Businesses are hoping the advancement of quantum computers—by tech giants such as Google, IBM, and Intel, as well as startups such as Rigetti Computing—will lead to unprecedented scientific and technical breakthroughs in the coming years. They're eyeing applications from new chemical reactions for the development of drugs, fertilizers, and batteries, to the improvement of optimization algorithms and mathematical modeling.

Computer Architecture

IC History

- See separate slide set on ***Transistors***

The Transistor



size = ~1 inch

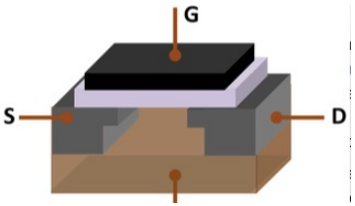
1947 ushered in the era of ***Microelectronics***

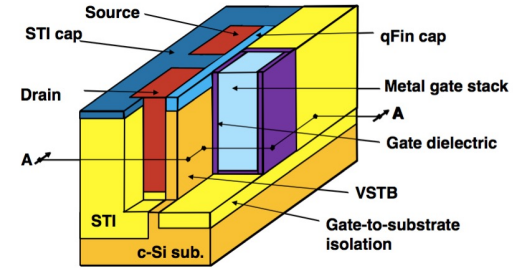
A **transistor** is a semiconductor device used to amplify or switch electronic signals and electrical power. It is composed of semiconductor material usually with at least three terminals for connection to an external circuit. A voltage or current applied to one pair of the transistor's termina

- 1947- Bipolar point/junction
- 1959- Planar bipolar [10]*
- 1964- MOS (P-channel) [100]
- 1972- MOS (N-channel) [1,000]
- 1978- CMOS [4,000]
- 1990- sub-micron [10,000]
- 2000- 100 nm [100,000]
- 2011- FinFET [1,000,000]
- 2019- 7nm [10,000,000,000]

*no. of transistors

Transistors have been shrunk every 2 years according to ***Moore's Law***

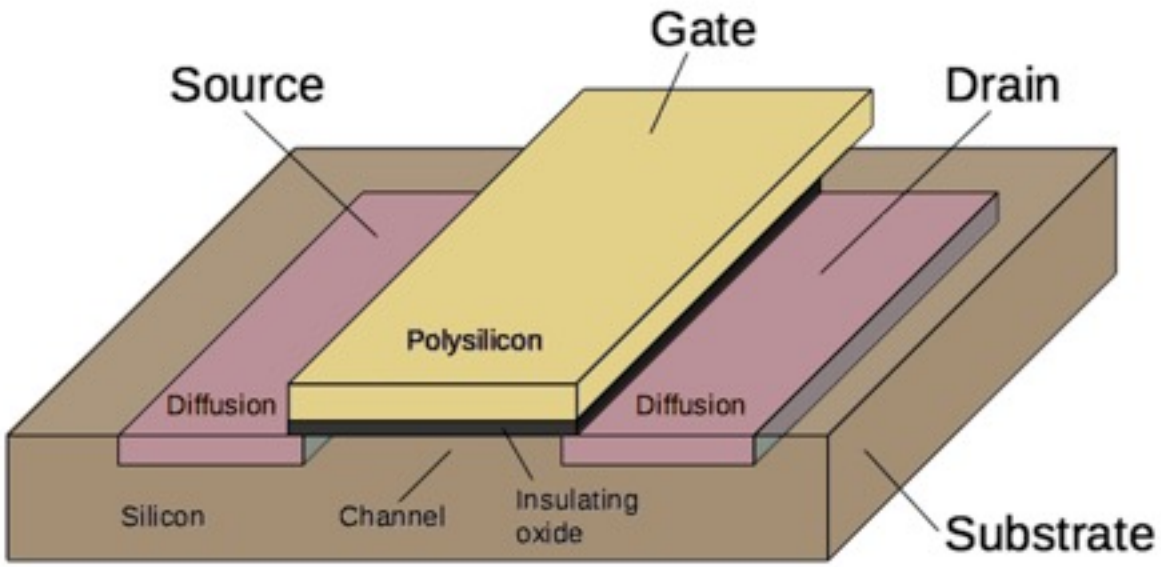




- size = 10 nm = 4×10^{-7} inches
- yields → ~1M devices per cm^2

MOS Transistor

WikiSemi



Structure of a MOSFET in the integrated circuit.

MOS Transistor

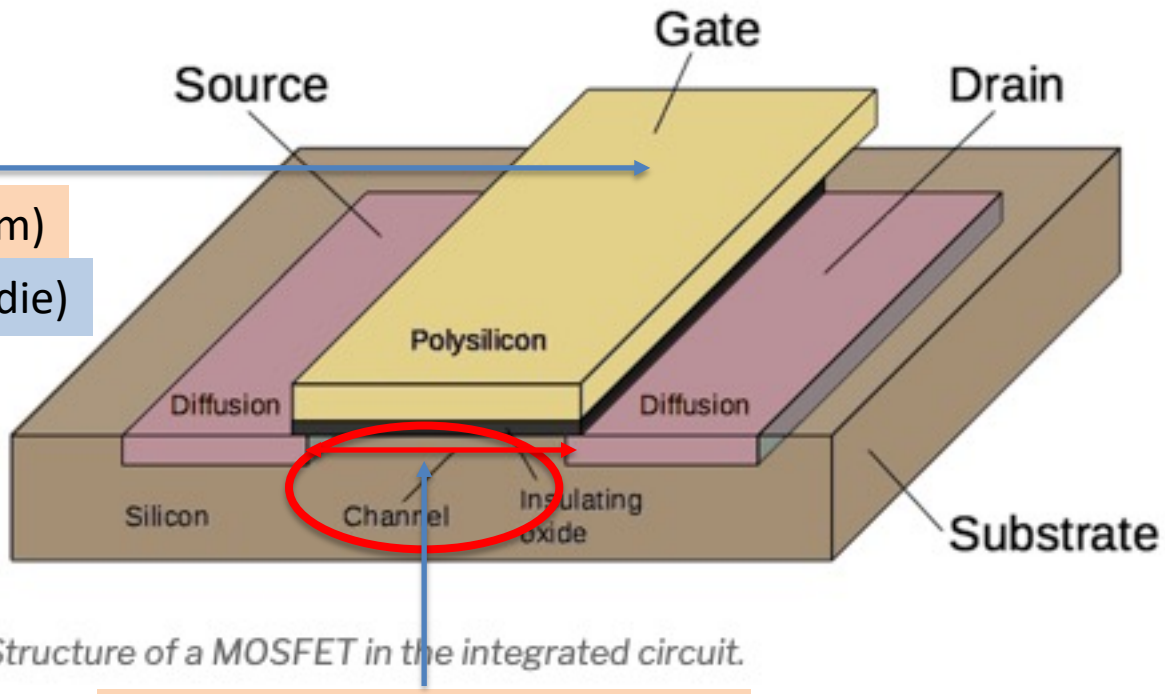
WikiSemi

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

Gate

Pitch (e.g.14nm)

Density (xtor/die)



Structure of a MOSFET in the integrated circuit.

Smallest feature size (e.g. 5nm)

Performance (F_t)

Atoms in 7nm

Quora

Al Kordesch, Semiconductor Device Modeling

Answered Feb 1, 2019

How many atoms are in a typical transistor in a chip?

Short Answer: 49,000 atoms!

Apple's iPhone XS uses 7 nanometer transistors. So let's estimate how many atoms are in one of them. Excluding the connecting wires and other parts, I'm just going to calculate the size of the active part, the "channel" under the gate. The volume of the channel is about (7 nm long) x (7 nm deep) x (20 nm wide). The atomic density of silicon is 5E+28 atoms per cubic meter. So let's go!

Number of atoms n = volume x density

n = (980E-27) x (5E+28) = 49,000 atoms.

Atomic radius = .111nm → 4.5 atoms/nm → 5/nm

Cubic: 5x5x5 = **125** atoms/cu nm

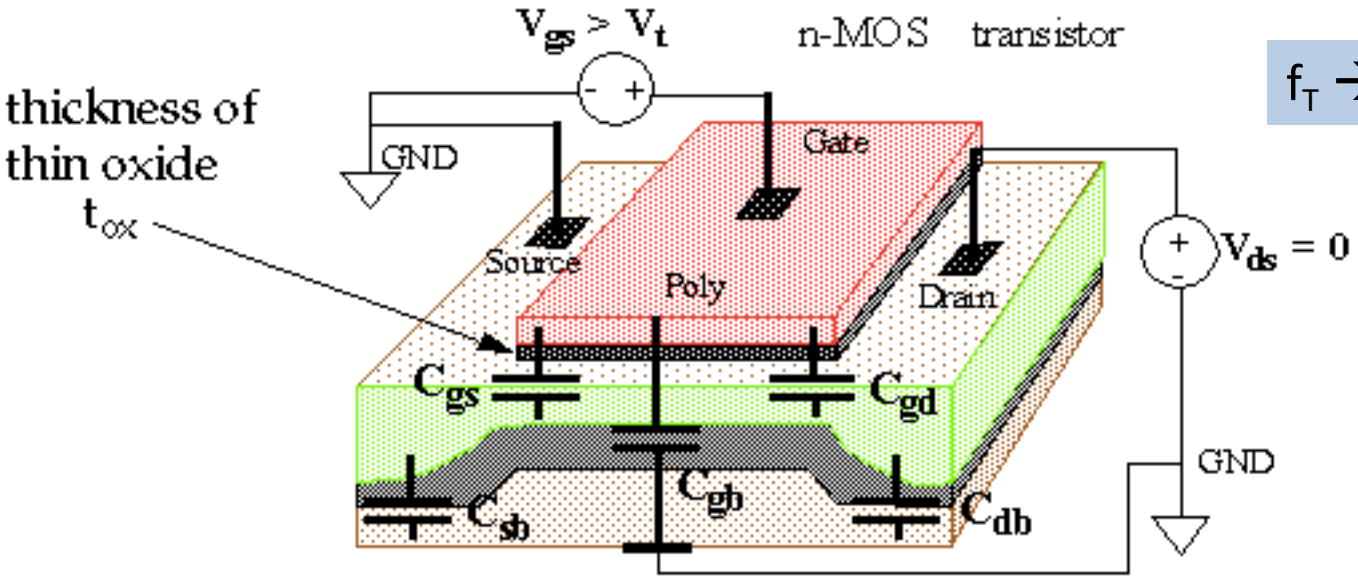
@7n: ~1000 cu nm

MOS Transistors

The channel will have a length (distance from one electrode to the other) and a width (imagine this diagram coming out of the screen). The electrodes have geometries. The gate doesn't always span the full width of the channel and is its own critical dimension.

The sizes of each of these things are critical dimensions. They all have an effect on the performance of the device because they will contribute parasitic capacitance and resistance.

$e^{-RC/T}$ Parasitic capacitances

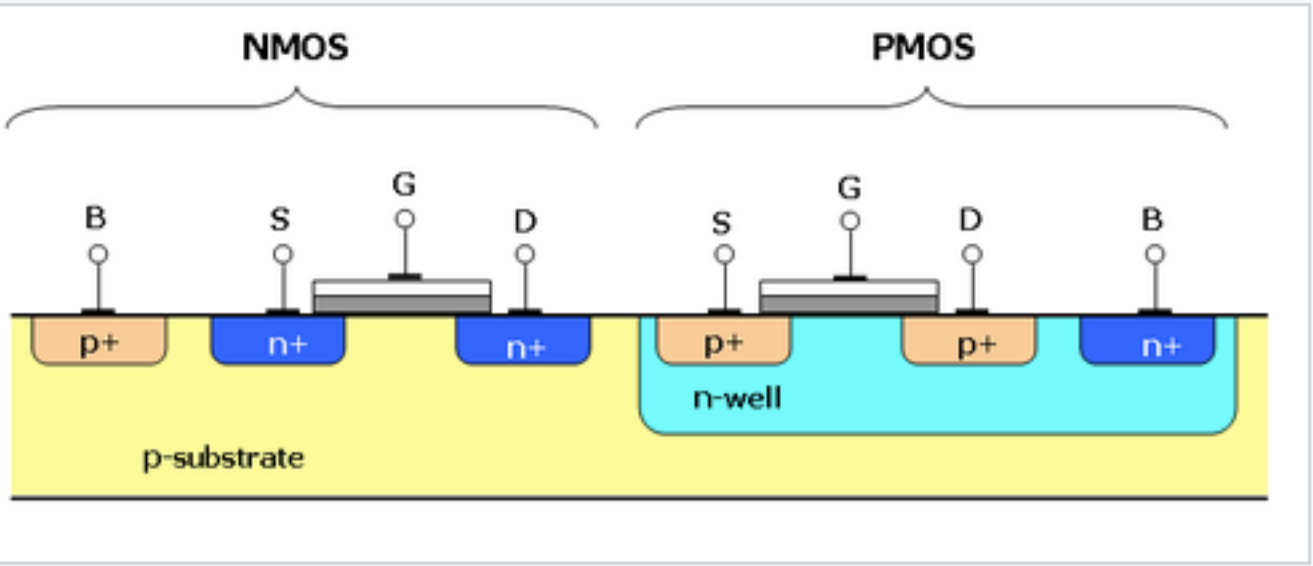


$f_T \rightarrow$ lengths

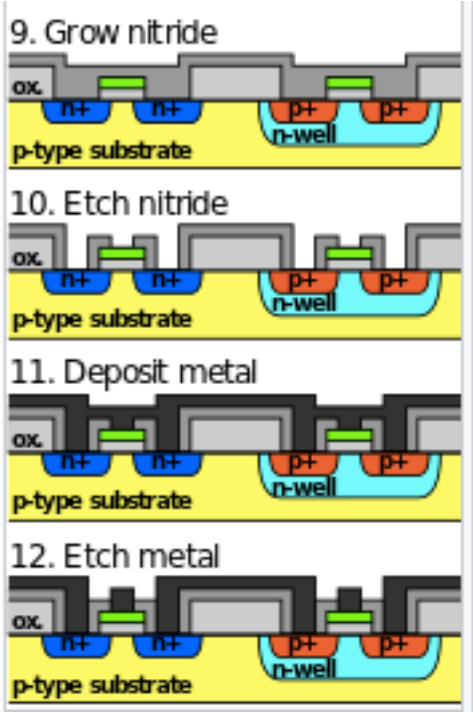
Parasitics depicted by University of New Mexico

CMOS Transistors

MOSFET



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



Last 4 steps

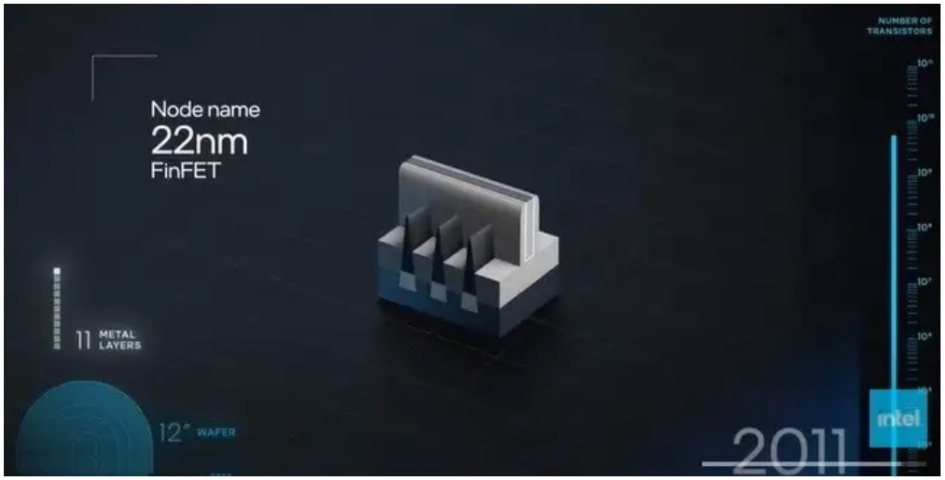
CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Intel MOSFET

Intel video

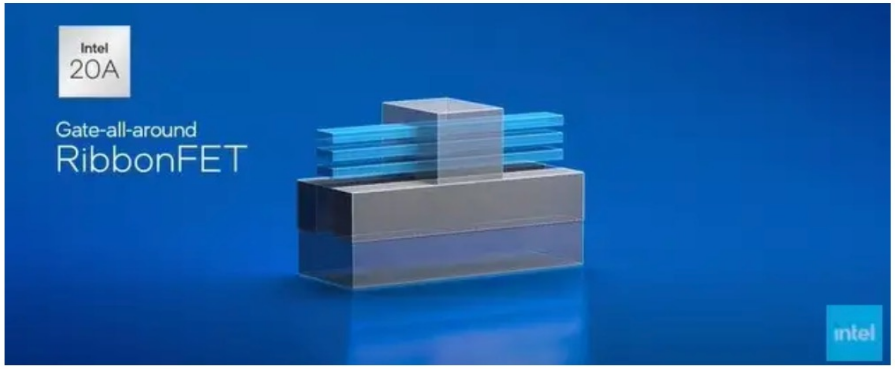
https://www.youtube.com/watch?v=Z7M8etXUEUU&t=47s

FinFET: 2011



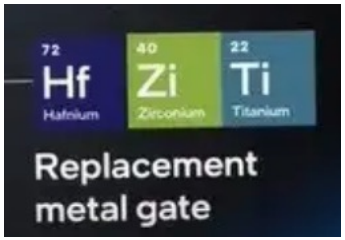
Intel amazed the industry with its aggressive adoption of a new transistor topology at the 22nm process node – the FinFET (also known as the "tri-gate FET").

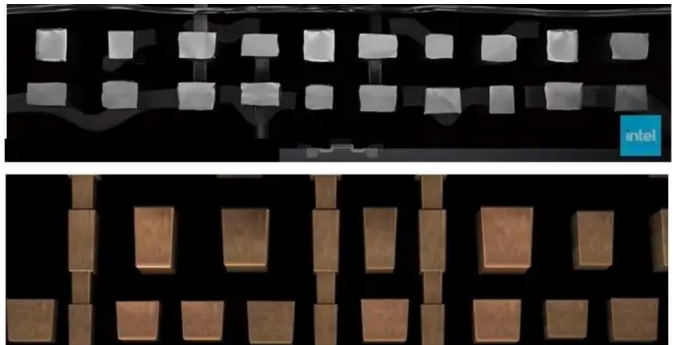
Gate-All Around (GAA) Ribbon FET: Intel 20A in 2024



To further improve the electrostatic gate control over the channel, another major evolution in the transistor topology is emerging to replace the FinFET. A gate-all-around configuration involves a vertical stack of electrically isolated silicon channels. The gate dielectric and gate input utilize an atomic layer deposition (ALD) process flow to surround all channel surfaces in the stack.

Intel will be releasing their GAA *Ribbon FET* 20A process in 1H 2024.





Si Valley History

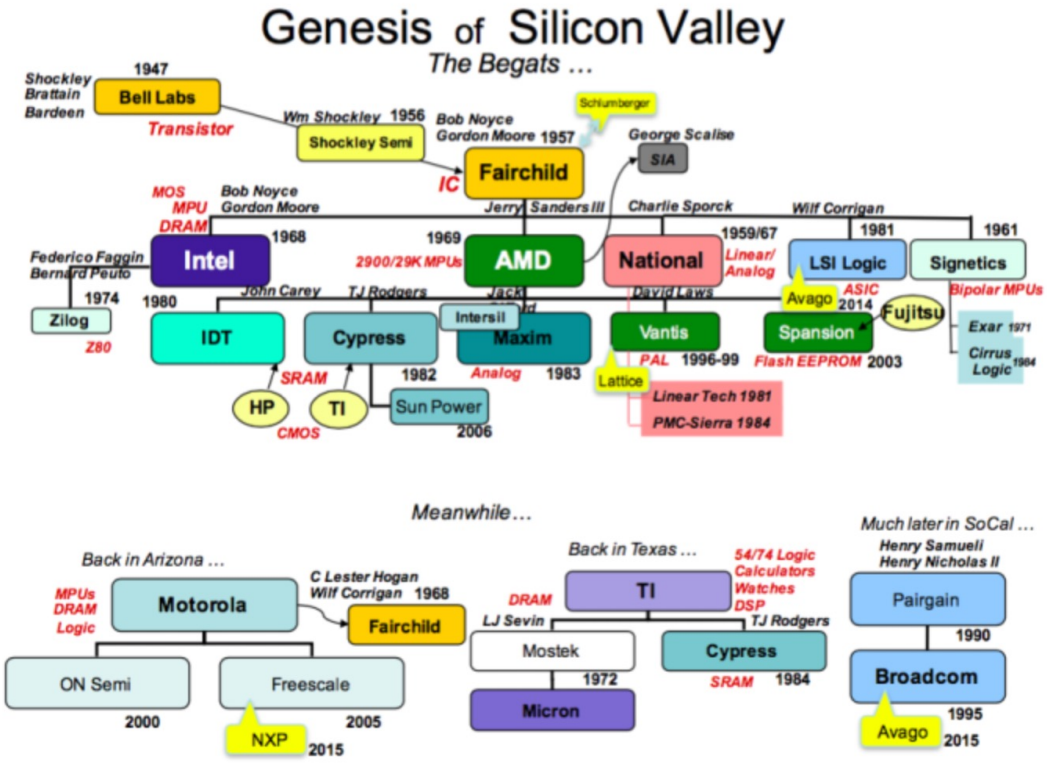
Genesis: A Silicon Valley Tale

TECH HISTORY ARTICLE

BY DR JEFF DROBMAN

Highlights

- Fairchild founding
- Intel founding
- AMD history
- AMD – Intel rivalry
- Search for CMOS
- RISC CPU Architecture
- Legendary Parties & Conferences
- Anecdotes
- Valley Significant Others
- Genesis org-chart
- Process Technology Evolution
- Anniversaries of Technologies



Shockley Labs

Beginning of ***Silicon Valley***

Mountain View, CA 1956

+

Stanford U

Genesis of Silicon Valley

The Begats & Bygones

CSUN CALIFORNIA STATE UNIVERSITY NORTHRIDGE

DR JEFF SOFTWARE INDIE APP DEVELOPER

COMP130

1947 Bell Labs — Shockley, Brattain, Bardeen — Transistor

Wm Shockley* 1956 Shockley Semi

Bob Noyce*, Gordon Moore 1957 Fairchild — IC

Schlumberger

George Scalise — SIA

Sperry Rand

Philips NXP

Jerry Sanders III — Charlie Sporck* — Wilf Corrigan

MOS, MPU, DRAM

Bob Noyce*, Gordon Moore

1968 Intel

1969 AMD — 2900/29K MPUs

1959/67 National — Linear/Analog

1981 LSI Logic

1961 Signetics — Bipolar MPUs

Federico Faggin, Bernard Peuto

1974 Zilog — Z80

1980

John Carey* — TJ Rodgers — Jack Gifford* — David Laws

IDT — Renesas

Cypress — SRAM — 1982

Intersil — Maxim — Analog — 1983

TI

Avago — ASIC — 2014

Fujitsu

Vantis — PAL — 1996-99

Spansion — RIP 2009 — Flash EEPROM — 2003

Exar 1971

Cirrus Logic 1984

Lattice

Linear Tech 1981

PMC-Sierra 1984

Analog Devices

Microsemi

HP — CMOS

TI

Sun Power — 2006

Legend—
Parents
Acquirers
*deceased

Meanwhile…

Back in Arizona …

MPUs, DRAM, Logic

Motorola

C Lester Hogan*, Wilf Corrigan 1968 Fairchild

ON Semi — 2000

Freescale — 2005

NXP — 2015

Back in Texas …

TI — DRAM, LJ Sevin, TJ Rodgers

54/74 Logic, Calculators, Watches, DSP

Mostek — 1972

Cypress — SRAM — 1984

Micron — Vin Prothro

Dallas SC

Maxim

Much later in SoCal …

Henry Samueli, Henry Nicholas II

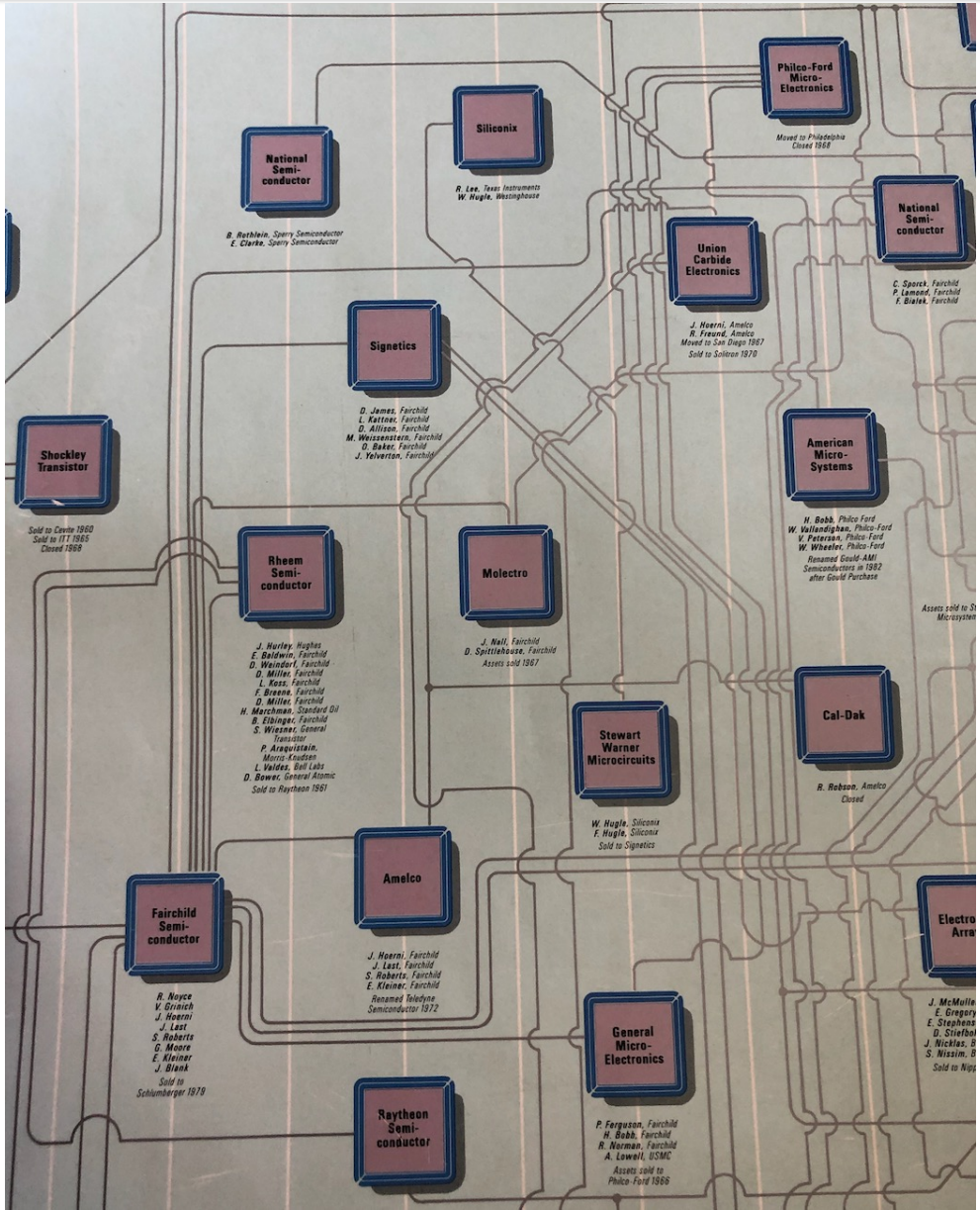
Pairgain — 1990

Broadcom — 1995

Avago — 2015

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

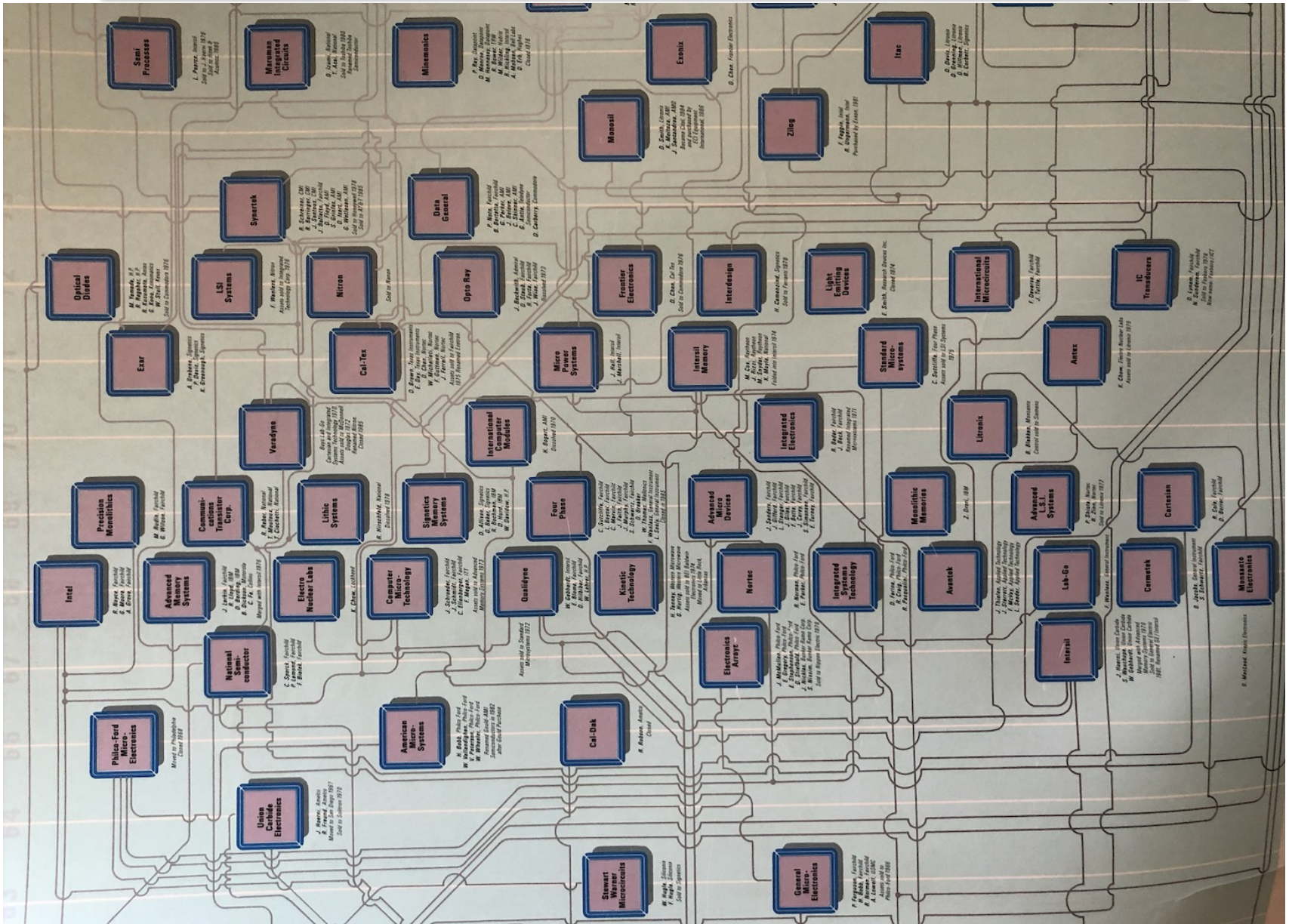
Si Valley History



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Si Valley History

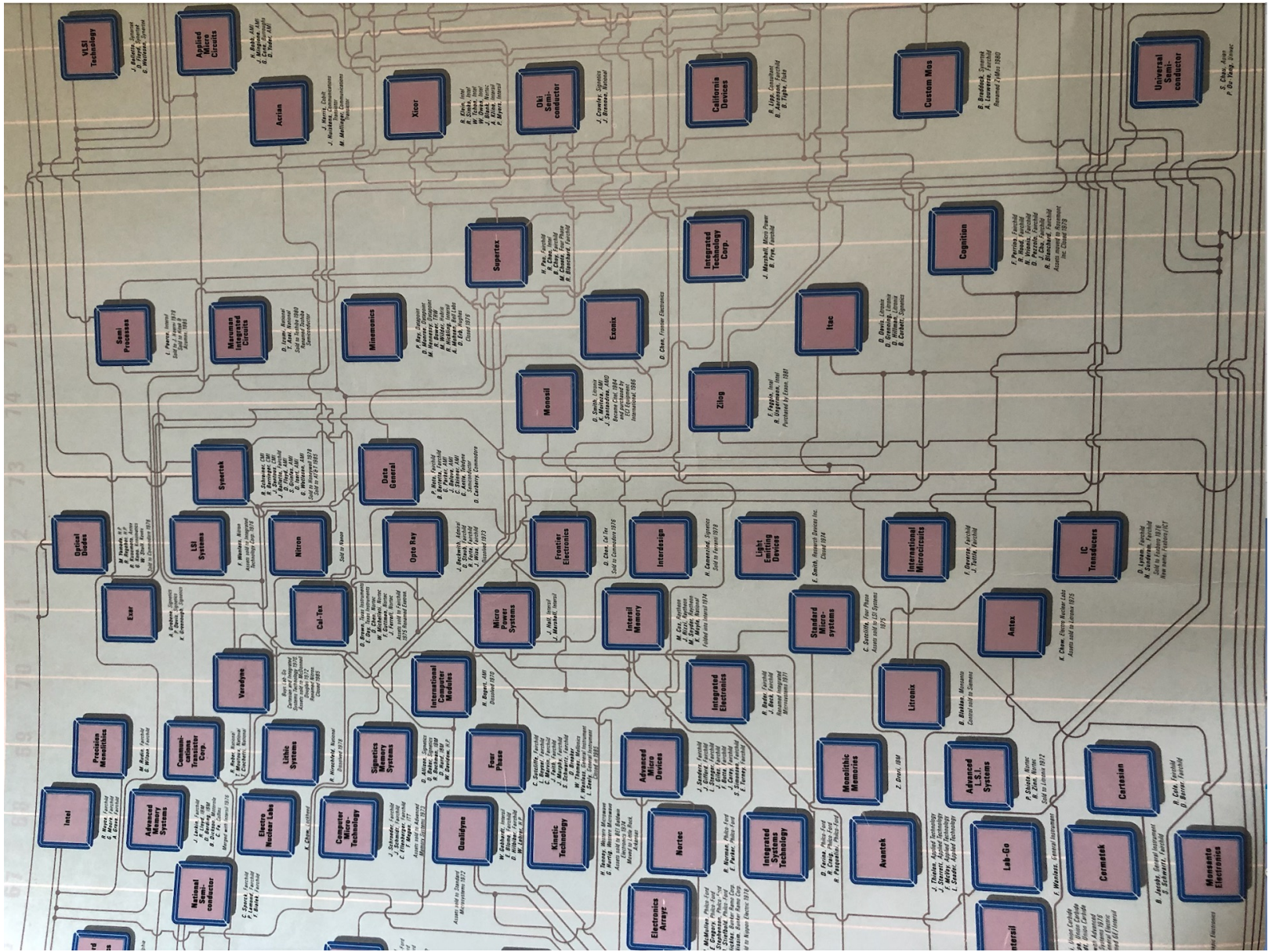
DR JEFF SOFTWARE
INDIE APP DEVELOPER



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Si Valley History

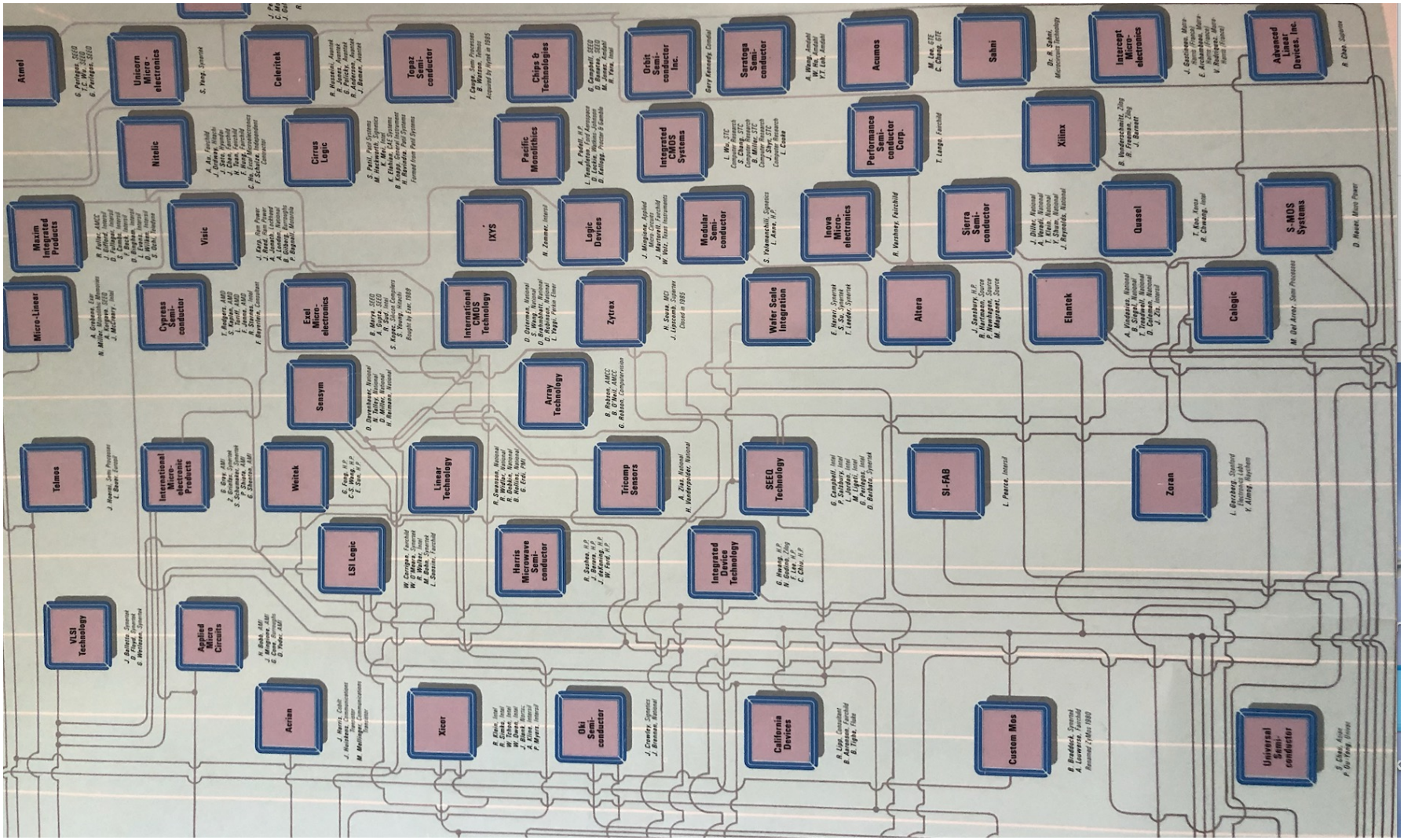
DR JEFF SOFTWARE
INDIE APP DEVELOPER



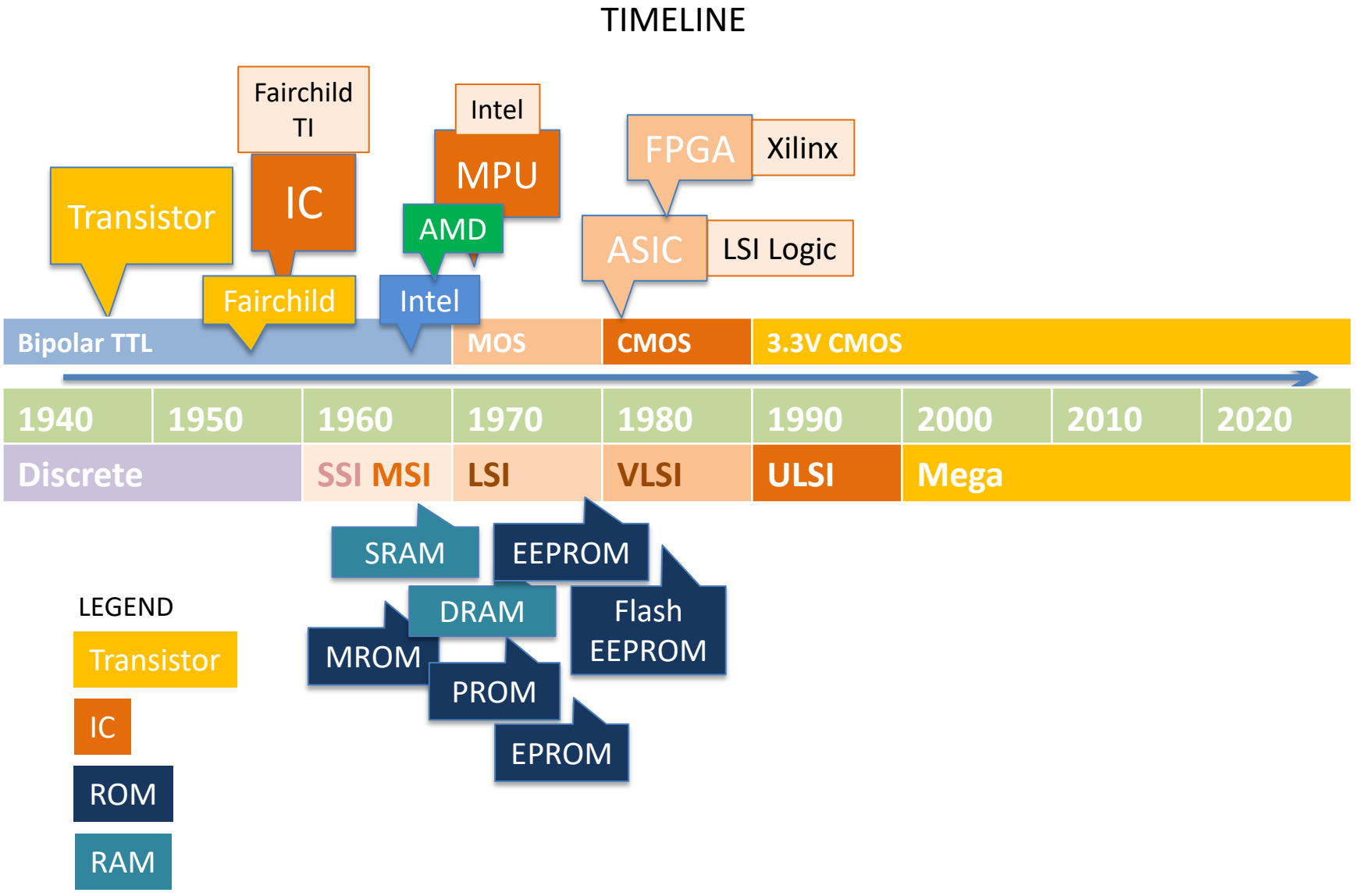
CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Si Valley History

DR JEFF SOFTWARE
INDIE APP DEVELOPER



IC Technology



Computer Architecture

Microprocessor History

MPU Part Numbers

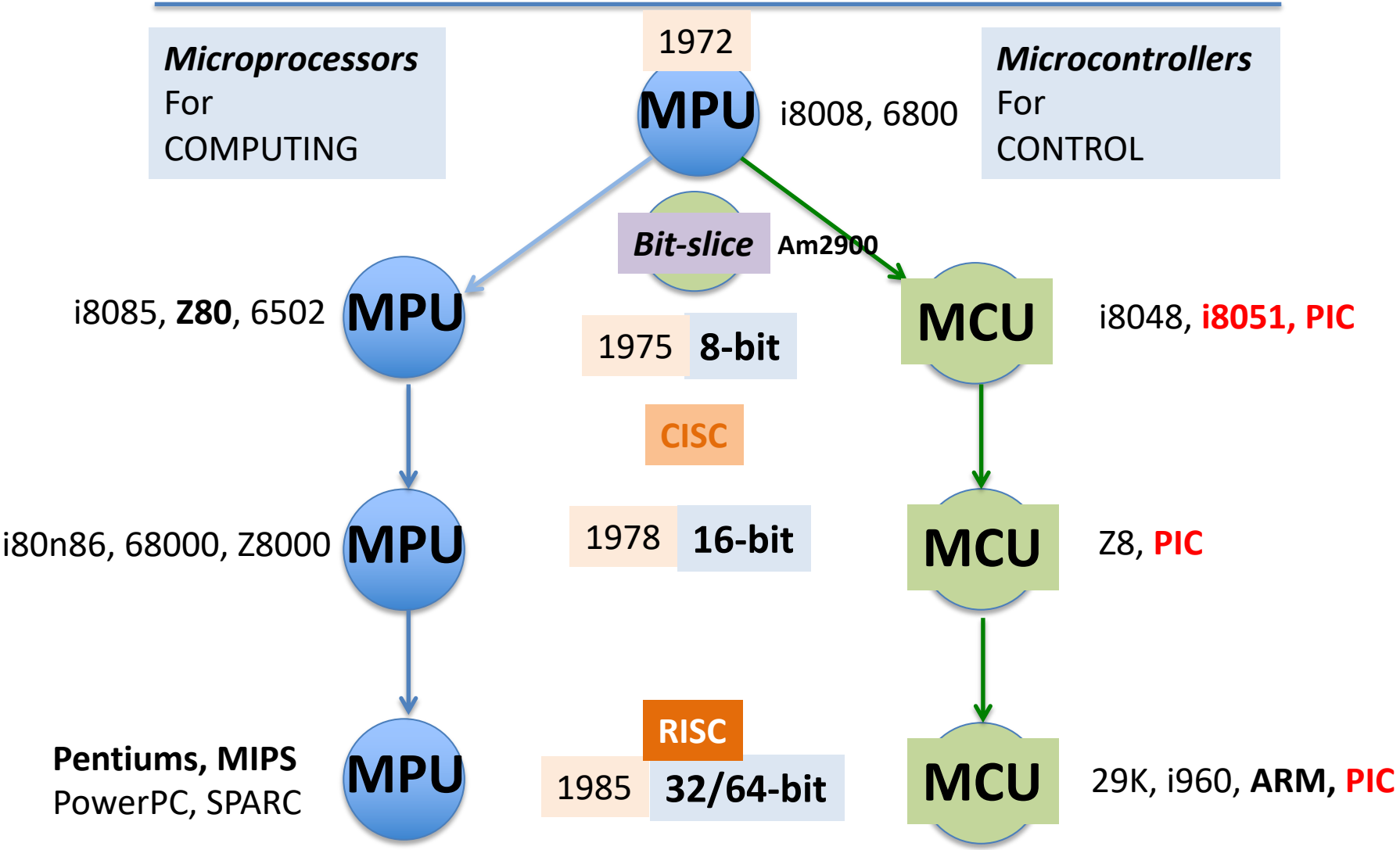
Quora

Was there a historical, technological or marketing significance behind the model numbers used in early microprocessors (i.e. 4004, 8008, 8080, 6800, 6502, 2900, et. al.)?

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

yes, there was a method for issuing part numbers — for ALL chips. leading (MS) digit for product line, 2nd digit for product family, so 9400 (Fairchild logic), 6500, 6800, 2900 (AMD bit-slice), were families. Intel started using 8008 for its 1st 8-bit microprocessor, then went 8080 for 2nd generation, 8085 for 3rd generation. Intel set its "x86" precedent when it strangely numbered its 1st 16-bit microprocessor "i8086". then each subsequent generation used the 2nd MS digit: 80186, 80286, 80386 (1st 32-bit), 80486, 80586. When Intel sued AMD over copying part numbers, a judge ruled part numbers cannot be trademarked. (Because the industry long had a practice of copying part numbers to indicate compatibility.) So Intel decided to forever stop using part numbers for CPU's, and started using "Pentium" and later names like "Celeron" etc. AMD had to likewise adopt unique names like Athlon, Ryzen, etc.
Note that Mot went from 8-bit 6800 to 16-bit 68000 (add a 0), but that would not be extended. Zilog did likewise going from its 8-bit Z80 to 16-bit Z8000 (add 00!) to sound stronger.

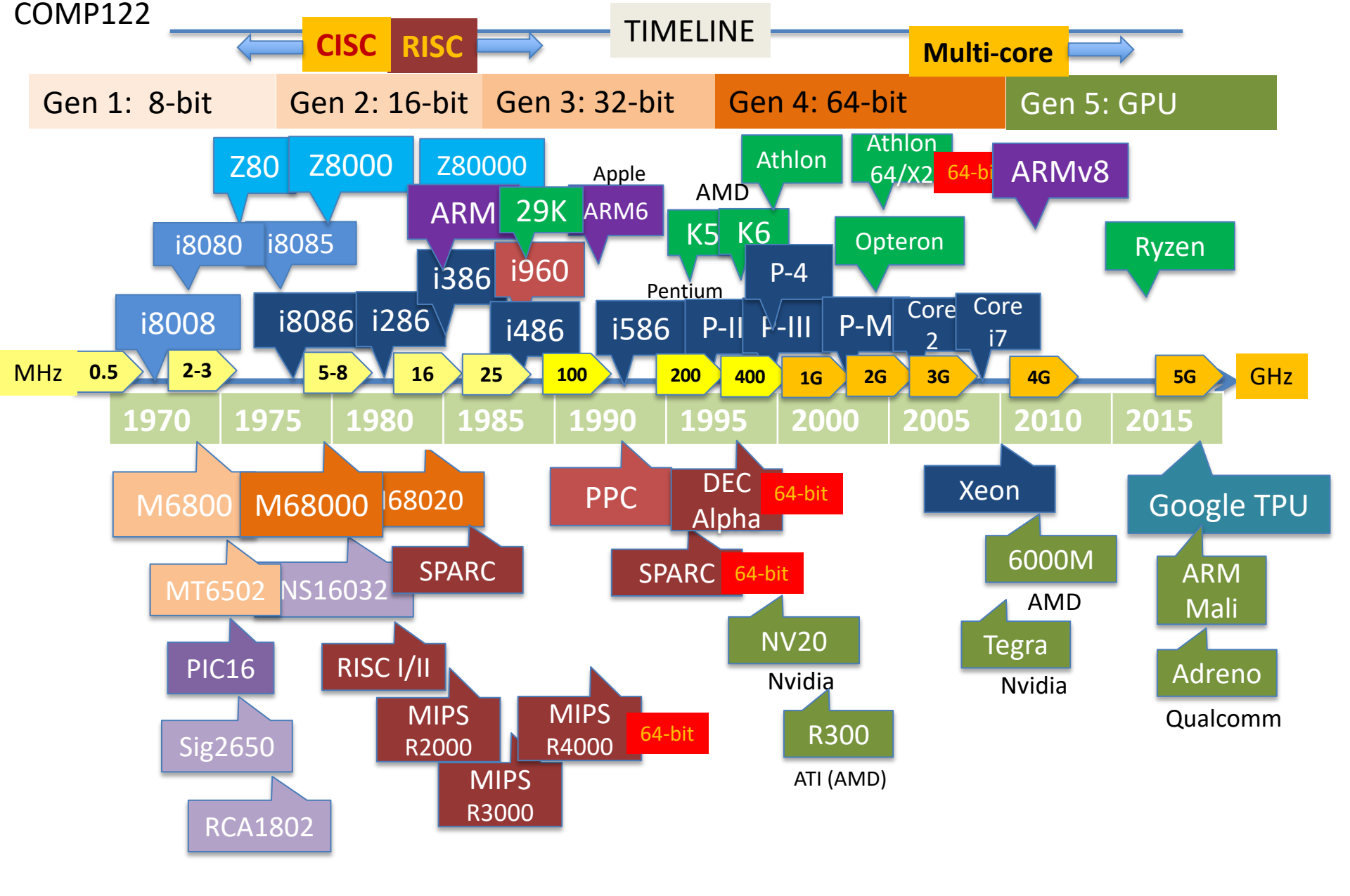
MPU/MCU Generations



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

COMP122

MPU Generations



Embedded Control

Microprocessors
For
COMPUTING

- All 32/64-bit CPUs
- Large *data processing* applications
  - Employee records
  - Accounting
  - Payroll
- Operating systems (OS)
- "Apps" (applications)
  - PC/Mac
  - Mobile (phones, tablets)
  - Web apps
  - Cloud apps (SaaS)

Focus is ***Memory*** for large Data Files

Large DRAM, Disk, Flash

Microcontrollers
For
CONTROL

- *Real-time*
- *All-in-one*

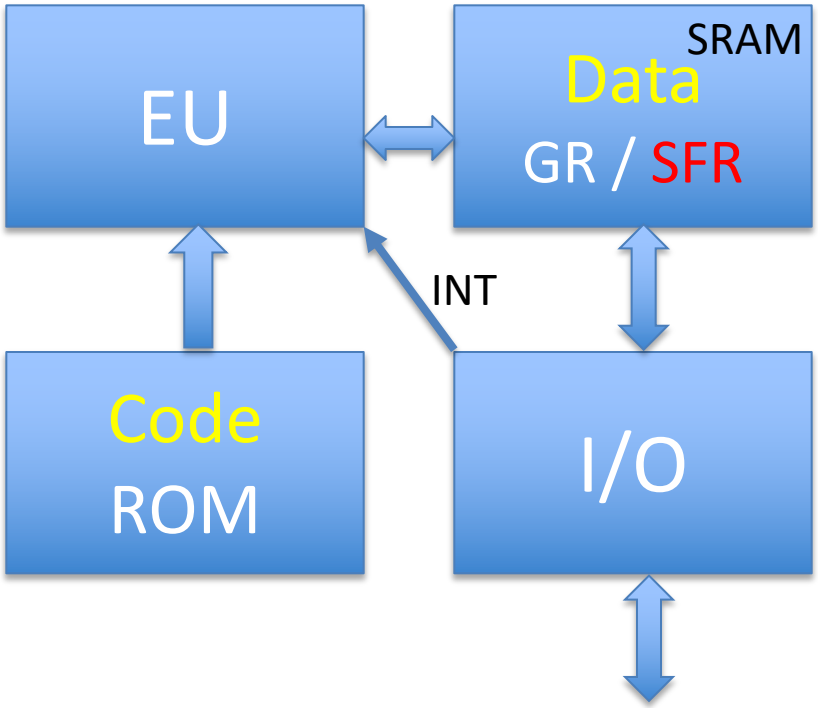
- Small *embedded control* applications (8-bit MCU)
  - Appliances
  - Disk controllers
  - Remote controllers
  - Garage/gate openers
- Tiny
- Low power
- Low cost
- Medium *embedded control* (16-bit MCU)
  - User devices (iPods, phones, etc.)
  - Car/Airplane engine control
  - Car/Airplane braking & safety
  - Car transmission control
  - Home Automation (HAN)
- Large *embedded control* (32/64-bit MCU)
  - Car/Airplane entertainment
  - Car/Airplane navigation, systems management
  - Printers (MF)
  - Communications gear (WiFi, cable TV boxes)

Focus is ***I/O*** – *Interrupts*

MCU Block Diagram

8/16/32-bit

BASIC MODEL



All on one cheap chip

- No Cache
- No External RAM

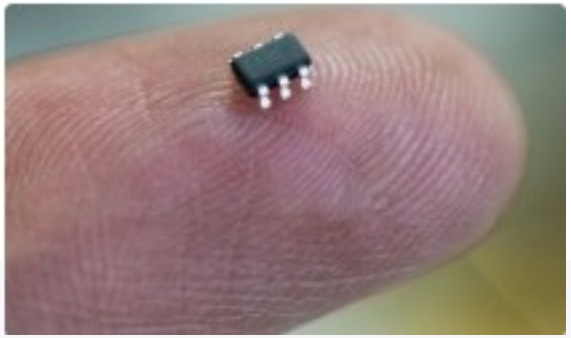
MCU

Quora

Meanwhile the number of microcontrollers estimated to be shipped in 2019 was estimated at around 27 billion, twelve times as many as the total number of microprocessors. As of 2017, the split was 40% for 32-bit, 33% 8-bit, and 24% 16-bit.

MCU = 12x MPU

So it can be estimated there were somewhere around nine billion 8-bit microcontrollers shipped in 2019. They are predominantly used in embedded systems that have a specific task, such as a small (air fryer, microwave oven) or large (washing machine) appliance; automobile cruise control; intelligent thermostat; etc.



Jeff Drobman

Just now

as of 5 years ago (when I last checked), the i8051 was still popular along with the PIC16 and 18 (16-bit). many models sold at <$1. Atmel's AVR is a popular microcontroller family that is customizable.

Small/Cheap MCU's

Quora

Another interpretation is for small/cheap microcontroller, such as the list in

https://theorycircuit.com/top-5-smallest-microcontrollers/

- **ATtiny20**
- **PSoC 4000**
- **KL03 (Arm)**
- **PIC12LF1552** ➢ MicroTech PIC family
- **C8051T606**

These are all in the vicinity of 3mm x 3mm x 0.5mm in size and sub-50 Mhz clocks, 1.5–16k flash, and 128 bytes to 2k of RAM.

Power seems to be in the 25–200 uA range depending on how careful you are.

CISC vs RISC:

Complex/Reduced Instruction Set Architecture

- Microprocessor History
  - 1971-85: **CISC** (8/16-bit)
    - Intel i4004 (4-bit)
    - Intel i8008 (8-bit) → i8080 → i8085, Z80 → i8086 (16-bit) → "x86"
    - Motorola 6800 (8-bit) → 6502 → 68000 (16-bit)
    - IBM PC used i8088 (8/16-bit) in 1981 → i80n86 ("x86") → *Pentiums* (now RISC)
  - 1985-2000: **RISC** – (32/64-bit)
    - SPARC* (UC Berkeley→ Sun/Oracle)
    - MIPS* (Stanford)
    - PowerPC (Motorola/IBM)
    - AMD 29K
    - Intel i960
    - ARM*

*still exist

Register Arch/Org

Hennessy & Patterson

Figure 2.21.1: The number of general-purpose registers in popular architectures over the years (COD Figure e2.21.1).

CISC

RISC

Machine	Number of general-purpose registers	Architectural style	Year
EDSAC	1	Accumulator	1949
IBM 701	1	Accumulator	1953
CDC 6600	8	Load-store	1963
IBM 360	16	Register-memory	1964
DEC PDP-8	1	Accumulator	1965
DEC PDP-11	8	Register-memory	1970
Intel 8008	1	Accumulator	1972
Motorola 6800	2	Accumulator	1974
DEC VAX	16	Register-memory, memory-memory	1977
Intel 8086	1	Extended accumulator	1978
Motorola 68000	16	Register-memory	1980
Intel 80386	8	Register-memory	1985
ARM	16	Load-store	1985
MIPS	32	Load-store	1985
HP PA-RISC	32	Load-store	1986
SPARC	32	Load-store	1987
PowerPC	32	Load-store	1992
DEC Alpha	32	Load-store	1992
HP/Intel IA-64	128	Load-store	2001
AMD64 (EMT64)	16	Register-memory	2003

CISC vs RISC Performance

- CISC → *CPI* = ~5-9 (typ)
- RISC → *CPI* = ~1.4 (typ) → 5X faster

Single core, single pipeline
(no instruction level parallelism)

Single-cycle execution → +Delays for Load, Branch

- Pipeline architecture
- Memory access limited (Load-Store)

i8086 History

WikiSemi

History of the 8086

The path to the 8086 was not as direct and planned as you might expect. Its earliest ancestor was the Datapoint 2200, a desktop computer/terminal from 1970. The Datapoint 2200 was before the creation of the microprocessor, so it used an 8-bit processor built from a board full of individual TTL integrated circuits. Datapoint asked Intel and Texas Instruments if it would be possible to replace that board of chips with a single chip. Copying the Datapoint 2200's architecture, Texas Instruments created the TMX 1795 processor (1971) and Intel created the 8008 processor (1972). However, Datapoint rejected these processors, a fateful decision. Although Texas Instruments couldn't find a customer for the TMX 1795 processor and abandoned it, Intel decided to sell the 8008 as a product, creating the microprocessor market. Intel followed the 8008 with the improved 8080 (1974) and 8085 (1976) processors. (I've written more about early microprocessors here.)

Datapoint 2200 computer. Photo courtesy of Austin Roche.

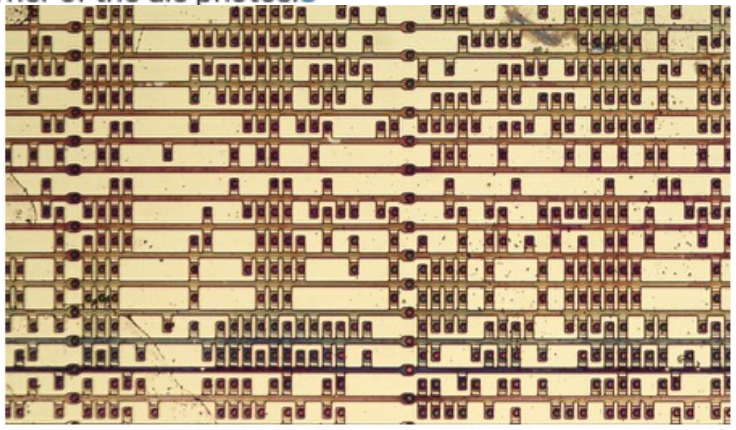
i8086 History

WikiSemi

Microcode

One of the hardest parts of computer design is creating the control logic that tells each part of the processor what to do to carry out each instruction. In 1951, Maurice Wilkes came up with the idea of microcode: instead of building the control logic from complex logic gate circuitry, the control logic could be replaced with special code called microcode. To execute an instruction, the computer internally executes several simpler micro-instructions, which are specified by the microcode. With microcode, building the processor's control logic becomes a programming task instead of a logic design task.

Microcode was common in mainframe computers of the 1960s, but early microprocessors such as the 6502 and Z-80 didn't use microcode because early chips didn't have room to store microcode. However, later chips such as the 8086 and 68000, used microcode, taking advantage of increasing chip densities. This allowed the 8086 to implement complex instructions (such as multiplication and string copying) without making the circuitry more complex. The downside was the microcode took a large fraction of the 8086's die; the microcode is visible in the lower-right corner of the die photos.3



A section of the microcode ROM.

i8086 History

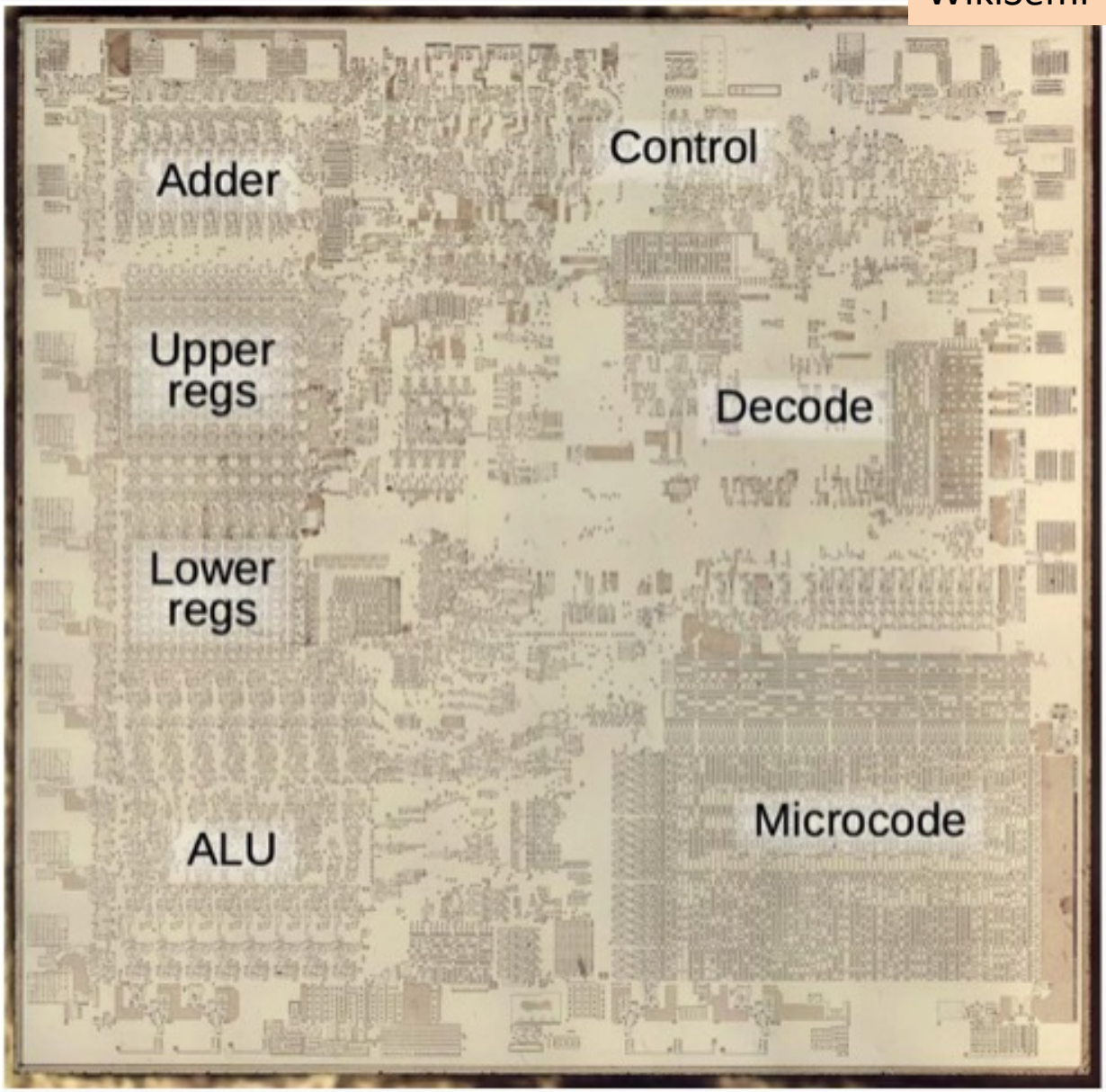
WikiSemi

Why did the IBM PC pick the Intel 8088 processor?7 According to Dr. David Bradley, one of the original IBM PC engineers, a key factor was the team's familiarity with Intel's development systems and processors. (They had used the Intel 8085 in the earlier IBM Datamaster desktop computer.) Another engineer, Lewis Eggebrecht, said the Motorola 68000 was a worthy competitor6 but its 16-bit data bus would significantly increase cost (as with the 8086). He also credited Intel's better support chips and development tools.5

In any case, the decision to use the 8088 processor cemented the success of the x86 family. The IBM PC AT (1984) upgraded to the compatible but more powerful 80286 processor. In 1985, the x86 line moved to 32 bits with the 80386, and then 64 bits in 2003 with AMD's Opteron architecture. The x86 architecture is still being extended with features such as AVX-512 vector operations (2016). But even though all these changes, the x86 architecture retains compatibility with the original 8086.

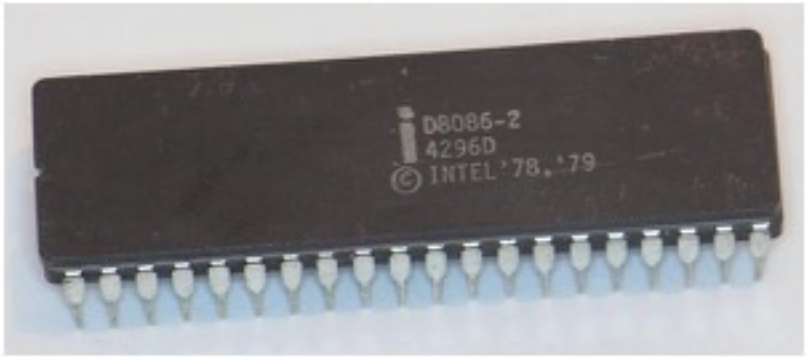
i8086 Die (etched)

WikiSemi

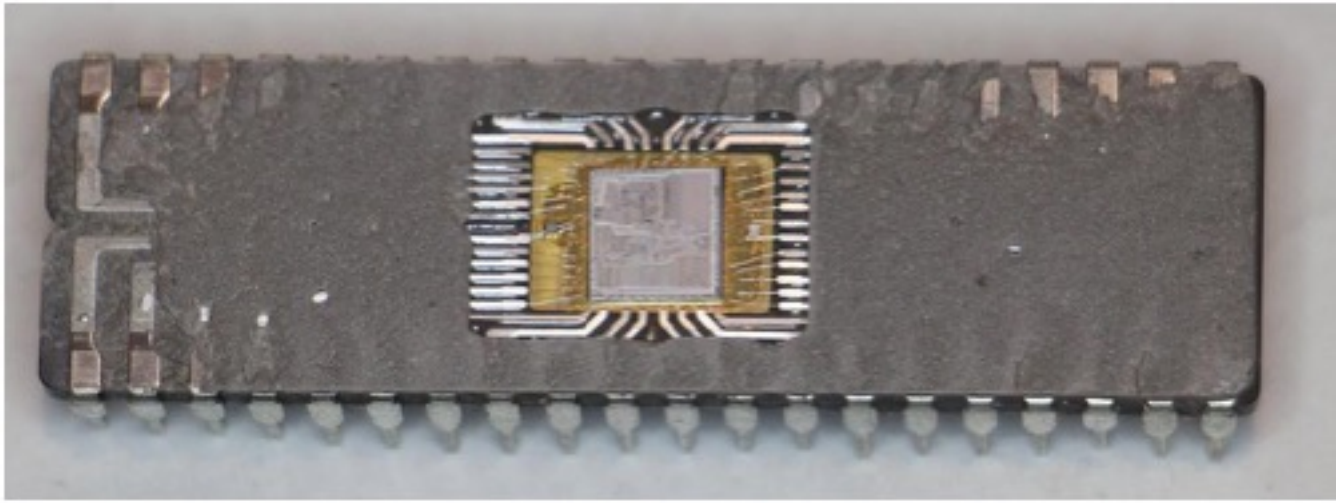


i8086 Packaged

WikiSemi



The 8086 chip, in 40-pin ceramic DIP package.



The 8086 die is visible in the middle of the integrated circuit package.

i8086 16-bit MPU

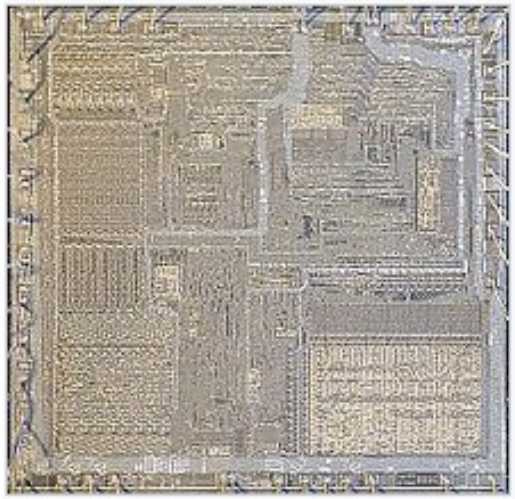
1st 16-bit MPU

1978

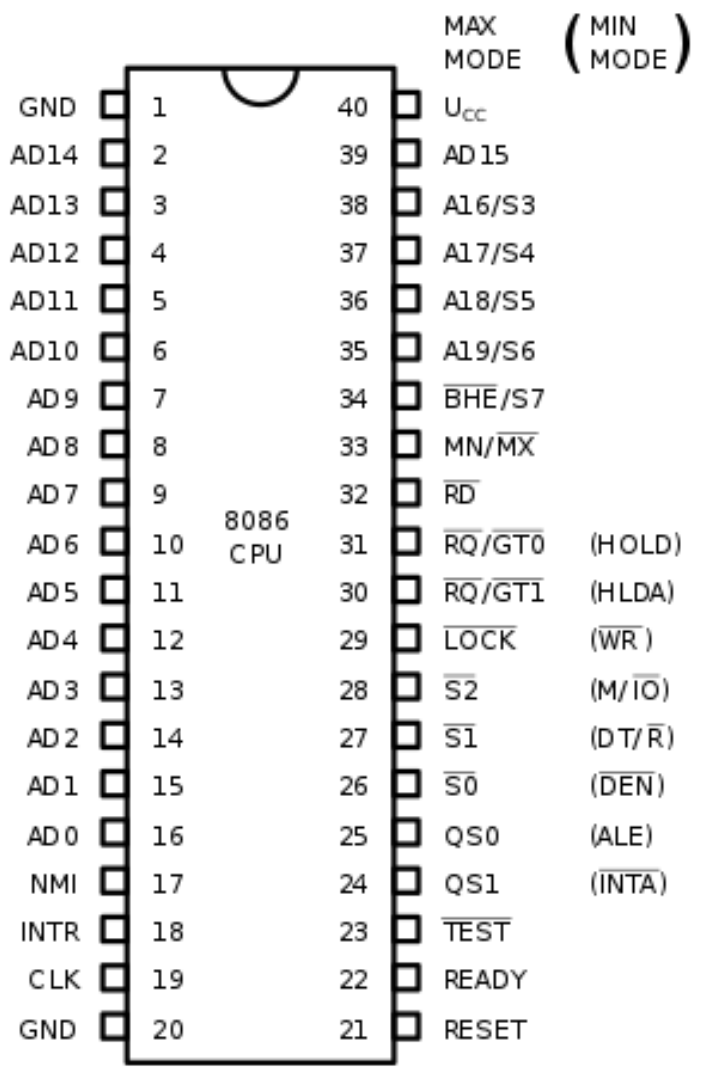
Intel 8086

A rare Intel C8086 processor in purple ceramic DIP package with side-brazed pins

General Info	
Launched	1978
Discontinued	1998[1]
Common manufacturer(s)	Intel, AMD, NEC, Fujitsu, Harris (Intersil), OKI, Siemens AG, Texas Instruments, Mitsubishi, Panasonic (Matsushita)
Performance	
Max. CPU clock rate	5 MHz to 10 MHz
Data width	16 bits
Address width	20 bits
Architecture and classification	
Min. feature size	3 μm
Instruction set	x86-16
Physical specifications	
Transistors	29,000
Co-processor	Intel 8087
Package(s)	40 pin DIP



Intel 8086 CPU die image

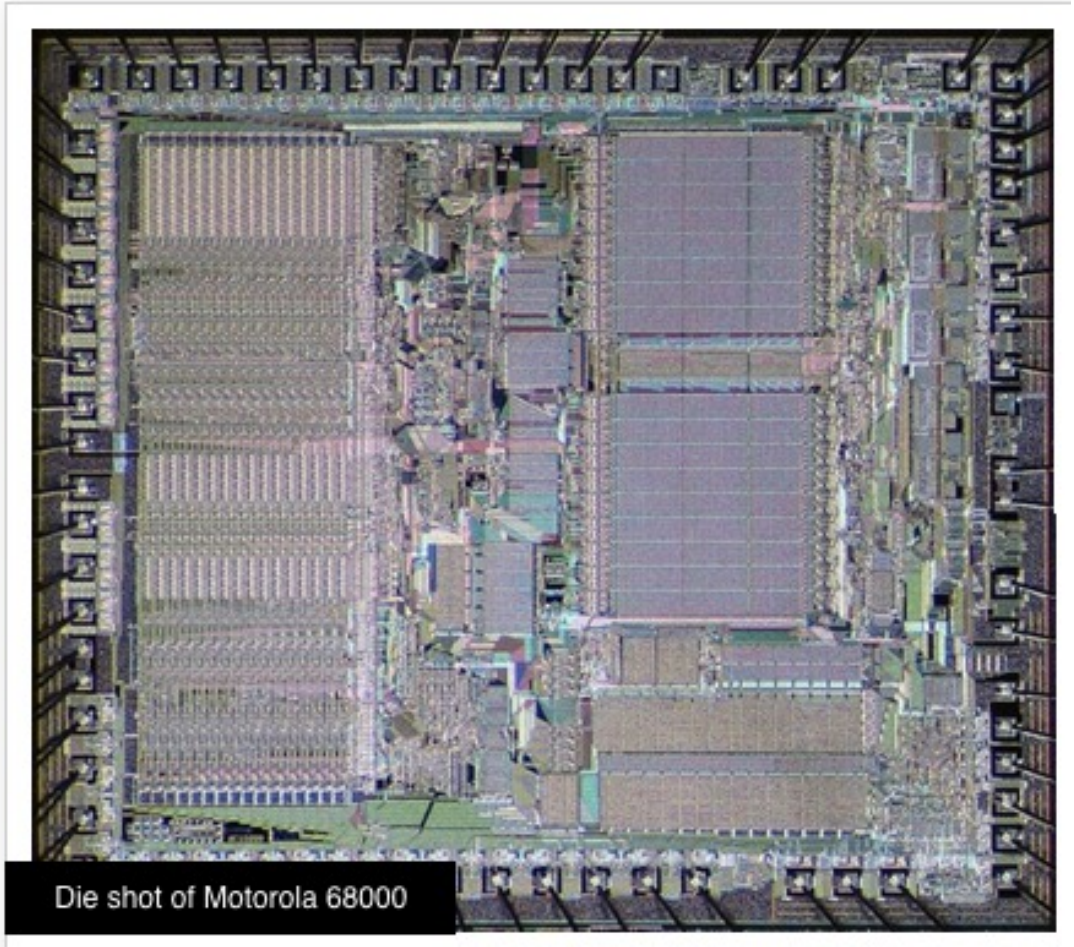


The 8086 pin assignments in min and max mode

M68000 16-bit MPU

1980

Motorola introduces the 68000 microprocessor



Die shot of Motorola 68000

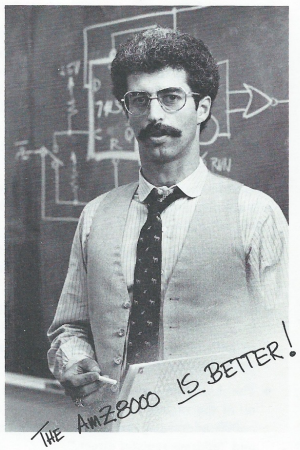
CPU ISA's

- x86
  - i8088
  - Pentium
    - Intel P, M
    - AMD K5-8
- MIPS
  - R3000/4000
  - MIPS32/64
- ARM
  - Cortex (A, M)
  - ARMv7/8
- RISC-V

The AmZ8000* Family

AmZ8000 VS 68000 REGISTER ARCHITECTURE

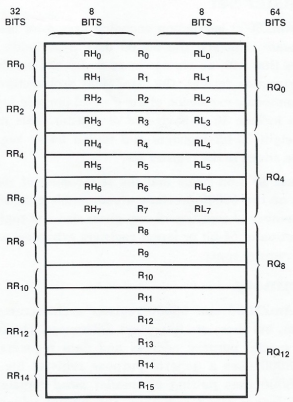
MAY 1981



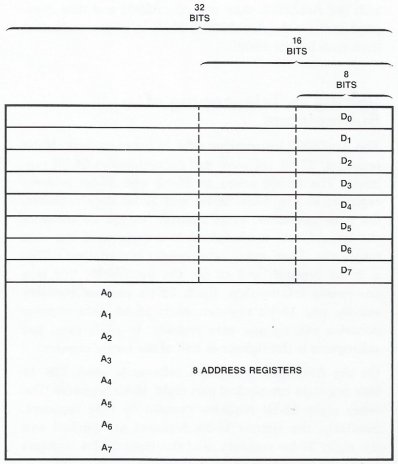
The International Standard of Quality guarantees these electrical AQLs on all parameters over the operating temperature range: 0.1% on MOS RAMs & ROMs; 0.2% on Bipolar Logic & Interface; 0.3% on Linear ICs; 0.3% on Logic & other memories.

The AmZ8000 and the 68000 take quite different approaches to register architecture. The principal points of difference are:

- General purpose vs. special purpose registers
- Pairing vs. telescoping of subregisters
- Extensibility of the registers sets



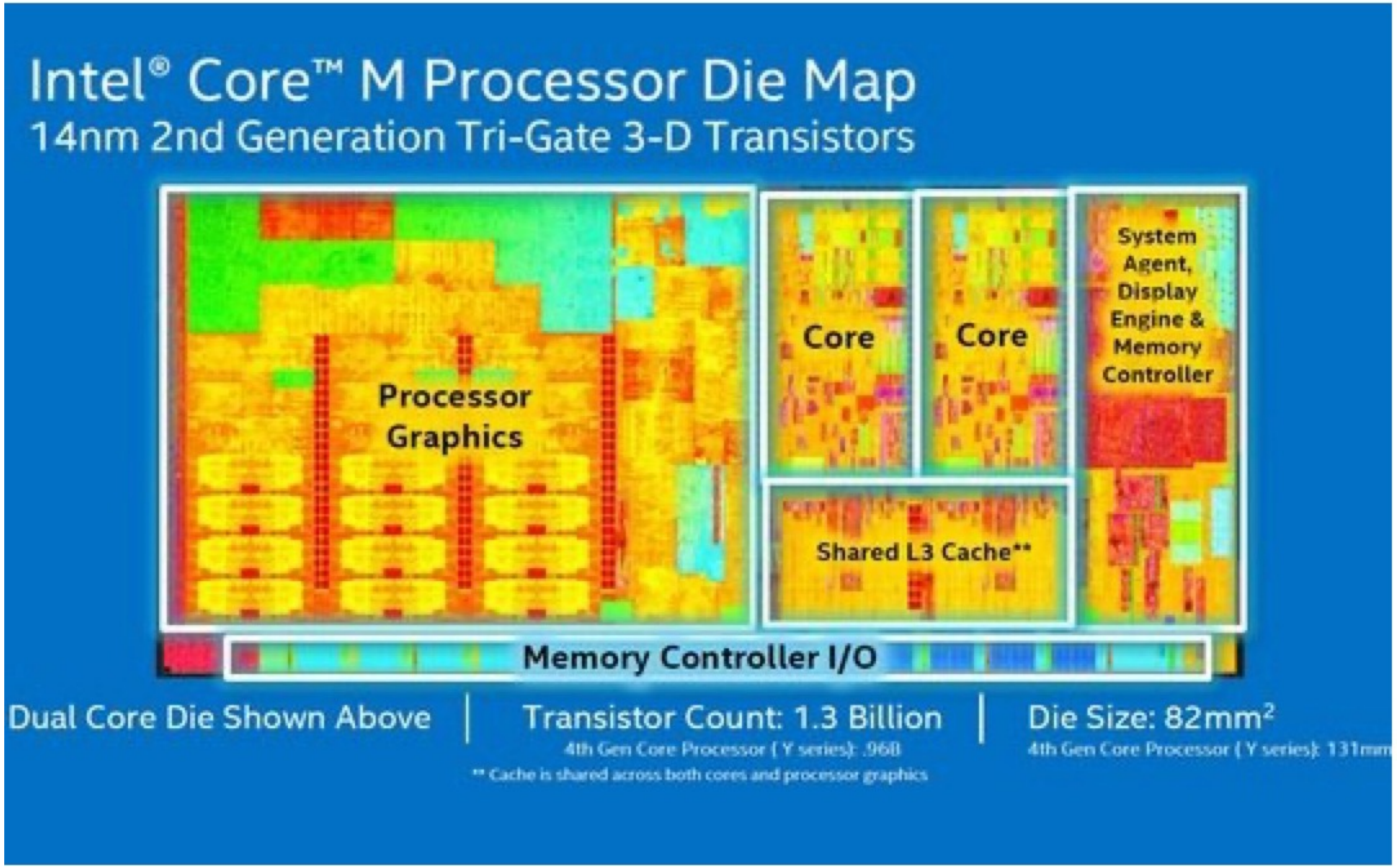
AmZ8000
16 General Purpose Registers can be used as
8 byte plus 8 word registers
or 16 word registers
or 8 long word registers
or 4 64-bit registers



MC68000
8 Data Registers can be used as
8 byte registers
or 8 word registers
or 8 long word registers

Advanced Micro Devices

Intel Core M Die



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

COMP122

Intel 12th Gen

12th Gen Intel® Core™ Desktop Processors

12th Generation Intel® Core™ i7 Processors

Product brief: 12th Gen Intel® Core™ desktop processors →

Product brief: Intel® Z690 Chipset →

NOW!

2 Products COMPARE ALL

Compare	Product Name	Status	Launch Date	# of Cores	Max Turbo Frequency	Cache	Processor Graphics ‡
☐	Intel® Core™ i7-12700KF Processor (25M Cache, up to 5.00 GHz)	Launched	Q4'21	12	5.00 GHz	25 MB Intel® Smart Cache	
☐	Intel® Core™ i7-12700K Processor (25M Cache, up to 5.00 GHz)	Launched	Q4'21	12	5.00 GHz	25 MB Intel® Smart Cache	Intel® UHD Graphics 770

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

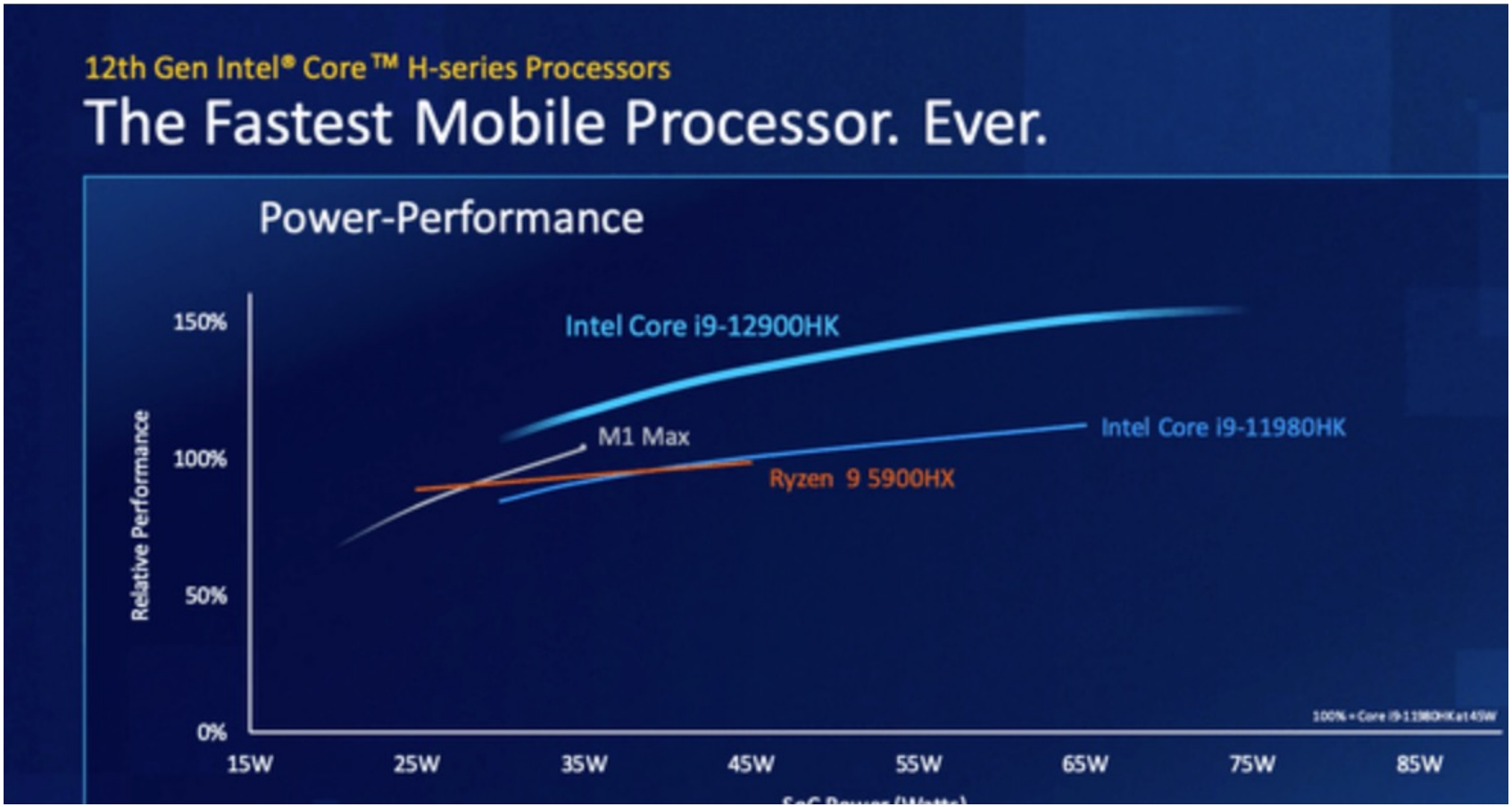
Intel 12th Gen

12th Gen Intel® Core™ Desktop Processors

12TH GEN INTEL® CORE™ DESKTOP PROCESSORS COMPARISONS

	Intel® Core™ i9-12900K & i9-12900KF[4]	Intel® Core™ i7-12700K & i7-12700KF[4]	Intel® Core™ i5-12600K & i5-12600KF[4]
Max Turbo Frequency [GHz]	Up to 5.2	Up to 5.0	Up to 4.9
Intel® Turbo Boost Max Technology 3.0 Frequency [GHz]	Up to 5.2	Up to 5.0	n/a
Single P-core Turbo Frequency [GHz]	Up to 5.1	Up to 4.9	Up to 4.9
Single E-core Turbo Frequency [GHz]	Up to 3.9	Up to 3.8	Up to 3.6
P-core Base Frequency [GHz]	3.2	3.6	3.7
E-core Base Frequency [GHz]	2.4	2.7	2.8
Processor Cores (P-cores + E-cores)	16 (8P + 8E)	12 (8P + 4E)	10 (6P + 4E)
Intel® Hyper-Threading Technology[5]	Yes	Yes	Yes
Total Processor Threads	24	20	16
Intel® Thread Director[1]	Yes	Yes	Yes
Intel® Smart Cache (L3) Size [MB]	30	25	20
Total L2 Cache Size [MB]	14	12	9.5

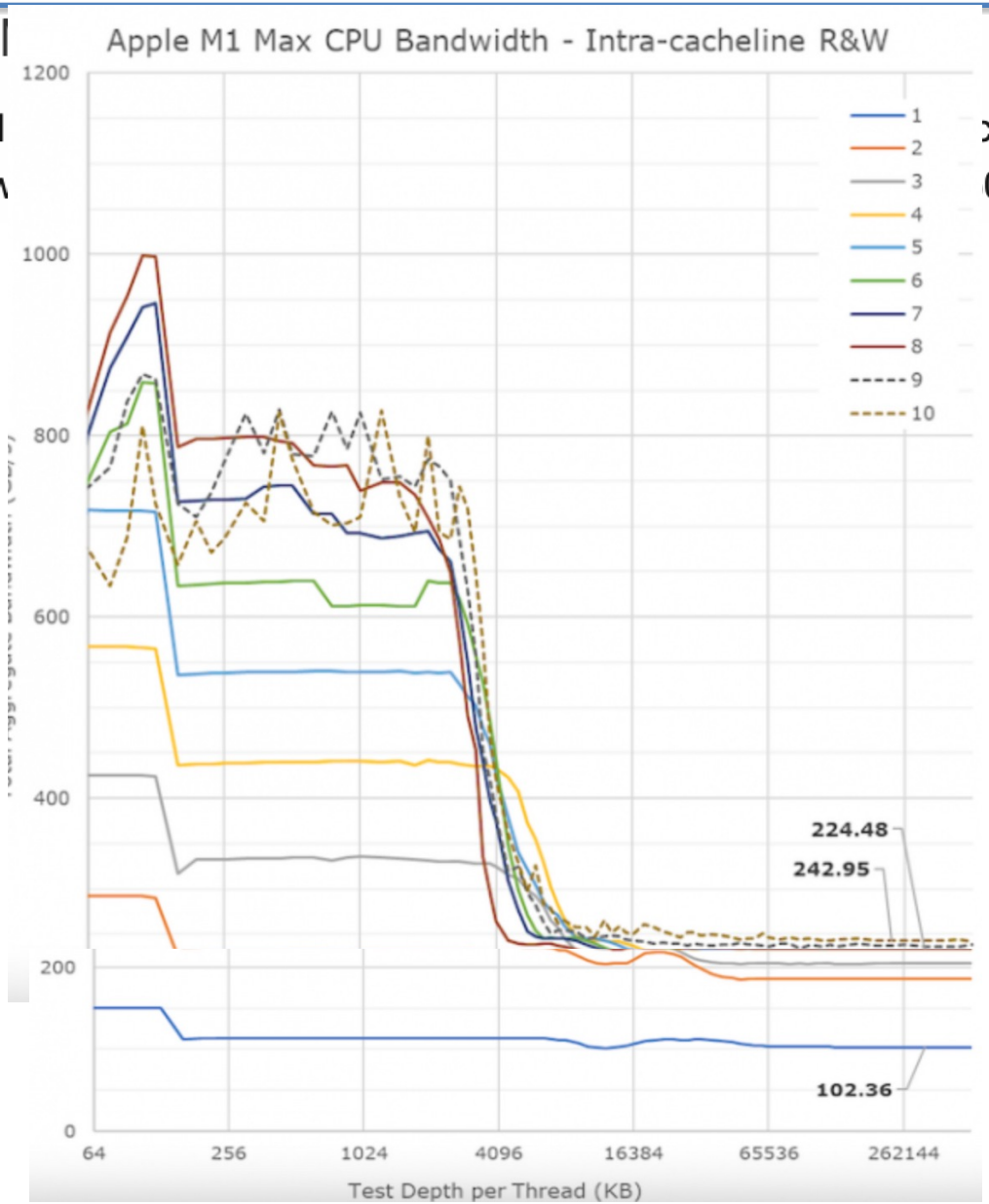
Intel 12G vs M1 Benchmarks



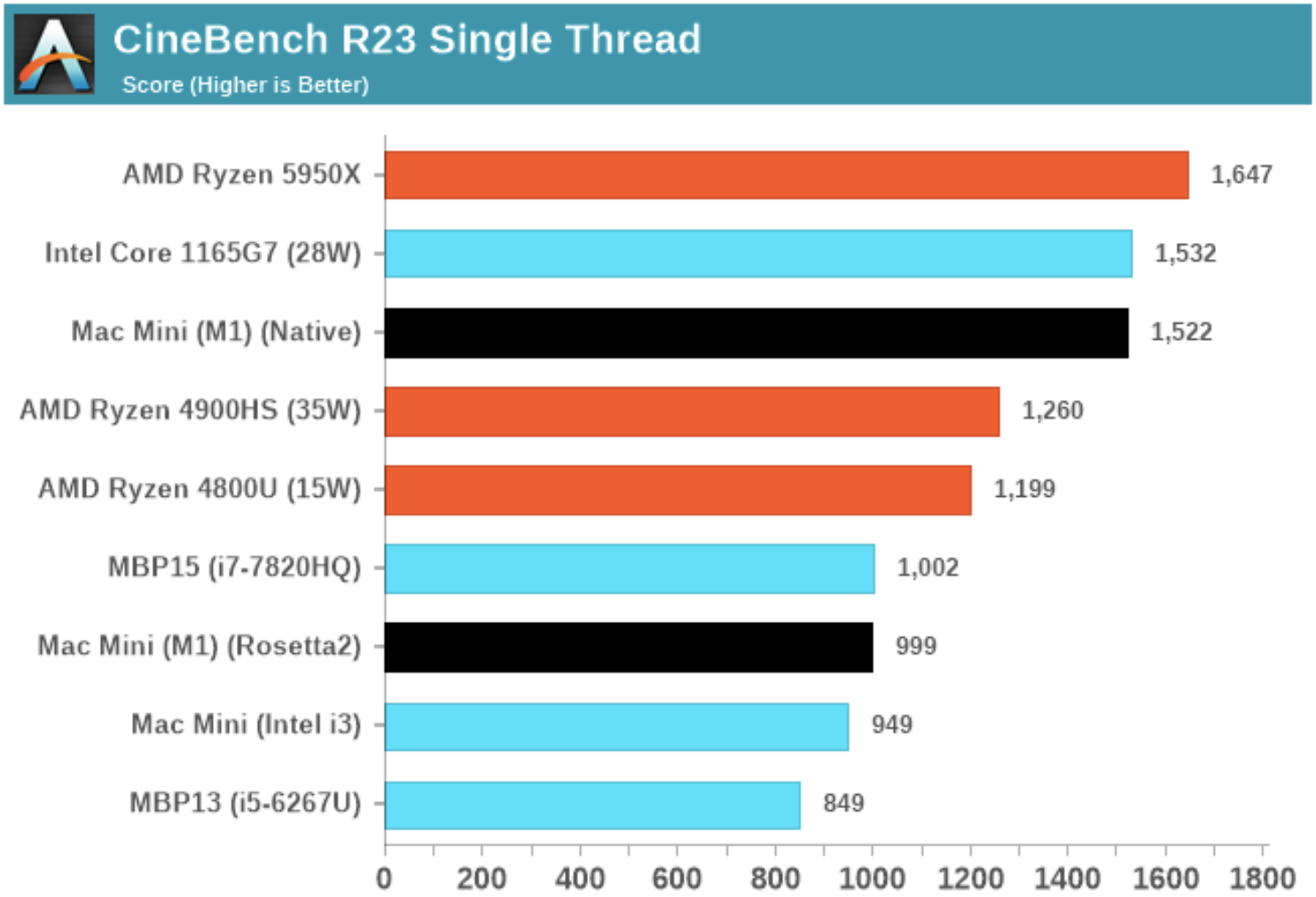
Intel 12G vs M1 Benchmarks

How Apple's I

When Apple's M1 ccess the system's theoretical bandw 0GB/s of it.



Benchmarks



The AMD Ryzen 5950X is clocked much higher. The M1 runs at 3.2 GHz. The Ryzen runs at 3.4 to 4.9 GHz. Let us use this to compute performance per GHz. For the M1 we get:

Computer Architecture

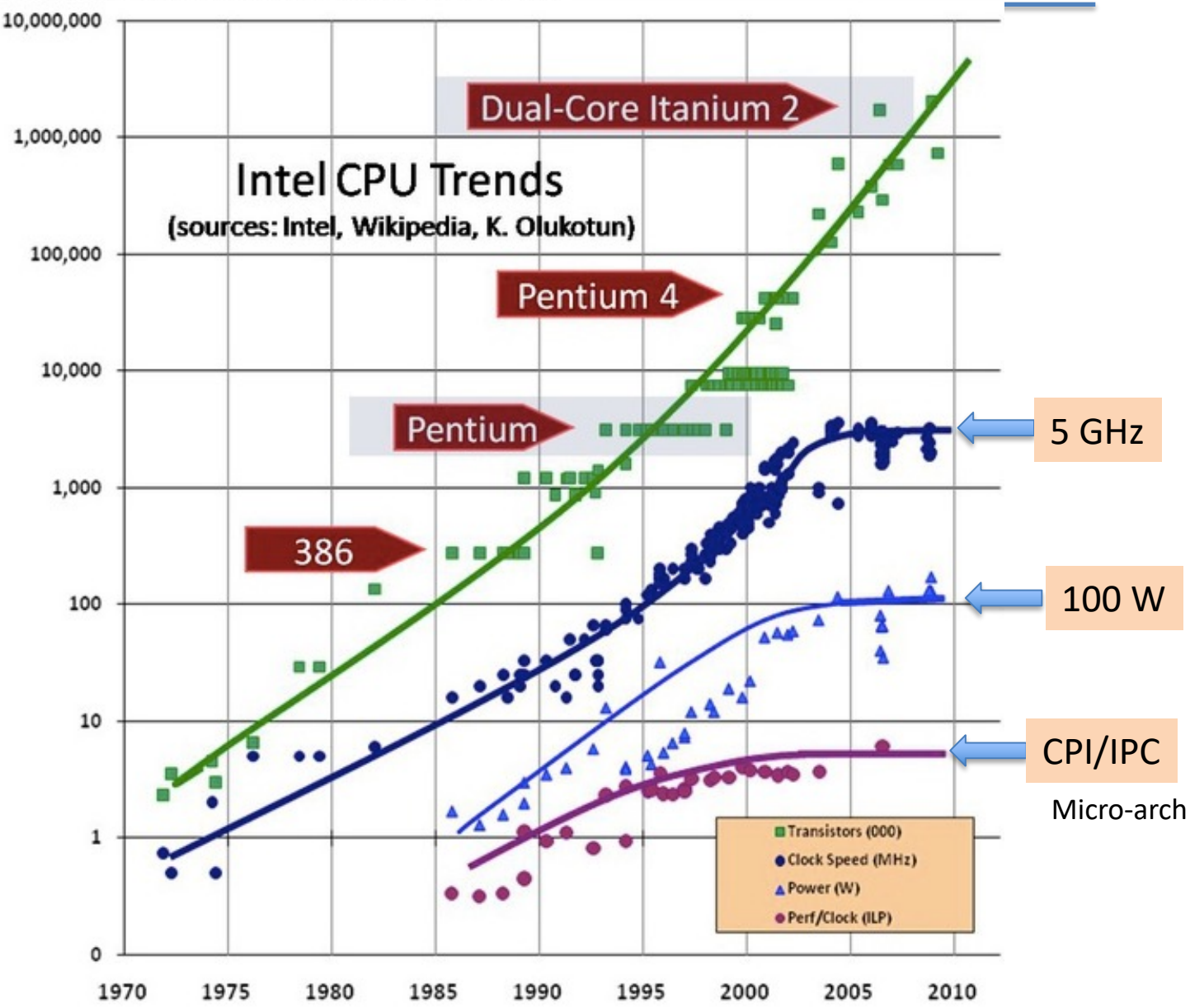
IC & Microprocessor Trends

CPU Trends

Quora

Steve Baker, Blogger at LetsRunWithIt.com (2013-present)

Answ



CPU Trends

Quora

Steve Baker, Blogger at LetsRunWithIt.com (2013-present)
Answered March 2, 2021

So Moore's Law for clock speeds has been pretty much over for 20 years.

The only way we can still make significantly faster chips is to use more transistors - which we have PLENTY of...but that doesn't translate into raw speed.

To use more transistors, all you can really do is to add more cores - add more cache memory - or try to build a more sophisticated way to run machine code.

- Adding more cores doesn't buy you much if your software can't use them all (and most software cannot).
- Increasing the amount of cache produces an incidental speed up for some algorithms and for badly written code - but for well-written programs, it can often have little to no effect on performance.
- So we're left with using more transistors to get smarter at running instructions.

Efforts to do that have included things like branch prediction, parallel execution at the micro-code level, speculative evaluation, all sorts of devious tricks.

But the trouble with these things is that they keep on resulting in things like fundamental CPU bugs and security issues.

Speculative evaluation (for example) was the cause of the Spectre, Meltdown, SPOILER and Foreshadow malware attacks - which were essentially impossible to defend against because the bug was hard-wired into the CPU core.

It's extremely hard to implement fancier CPU features without inadvertently opening a new security hole or some other horrific bug.

We truly are seeing the end of CPU speed improvements.

CPU Trends

Quora

Jeff Drobman · Just now

very good commentary, and spot on about transistor frequency plateau due to thermal limitations, as well as selective use of CPU turbo boost with dark silicon. seems the focus has changed from compute performance to power management. that said, the performance side has shifted toward SIMD vector and tensor arithmetic along with more use of GPU's with super high thread counts (1000s).

How will CPUs continue to get better after we can't shrink transistors anymore?

Jeff Drobman

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

the trend has been to add more cores, both CPU and GPU. and different types such as P (power) and E (efficient). add ***parallelism*** into CPU via SIMD/AVX vectors, a little more multi-issue slots with high EU count; into GPU's by higher thread count; and new TPU's with matrix multiply units.

CPU Trends

Quora

Steve Baker, Blogger at LetsRunWithIt.com (2013-present)
Answered March 2, 2021

SO WHAT IS THE FUTURE?

The future seems to be in more specialized processors:

GPU

- The GPU architecture - originally intended for graphics processing - has proven to be immensely powerful. Instead of having a handful of entirely unrelated and highly sophisticated CPU cores - you build hundreds of much simpler GPU "cores" which operate more or less together in lockstep. By sharply limiting the functionality - but radically increasing their numbers - we can write specialized "shader" software that runs at speeds that CPU software can only dream of. Hence GPU cores are now used for things that have nothing to do with graphics. Everything from artificial intelligence to bitcoin mining. They don't help with every algorithm, but in areas where they DO help - you can get two orders of magnitude speedup with a relatively cheap chip.

AI

- Specialized AI processors - the Tesla AI chip for example - take that even further. Performing *JUST* the neural networking algorithm at the heart of all AI - but doing so with VAST numbers of even simpler processors (not much more than multiply-accumulators). This means that they can run AI processing hundreds of times faster than even a GPU chip. But that's ALL it can do. In order to run conventional programs, the Telsa chip has to have several conventional CPU cores on the same chip to feed and generally manage the AI system.

QC

- Quantum computers - which are truly insanely fast - but are only capable of running VERY specialized algorithms that require extreme parallelism.

CPU Trends

Quora

How many orders of magnitude have computers advanced in 50 years?

➢ 100,000x perf Single core

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

Answered 1m ago

the 1st microprocessors were introduced by Intel in 1972, so yes, 50 years ago. they ran at <1 MHz and took about 10 clocks per instruction. now, we have up to 5 GHz (5000x), multi-core with MT can at least double the throughput (2x), and we achieve about 1.4 cycles per instruction per thread. and this is just CPU cores, not including GPU cores. add it up, we get close to 100,000x more performance (MIPS) = 5 orders for CPU chips.

<1 MHz → 5 GHz

CPI = 10 → 1.4

CISC → RISC

RISC-V

Heikki Kultala, Technical leader, SoC architecture at Nokia (2020-present)

Answered February 6

No, RISC-V is 1980s done correctly, 30 years later.

It still concentrates on fixing those problems that we had in 1980s (making instruction set that is easy to pipeline with a simple pipeline), but we mostly don't have anymore, because we have managed to find other, more practical solutions to those problems.

And it's "done correctly" because it abandons the most stupid RISC features such as delay slots. But it ignores most of the things we have learned after that.

ARMv8 is much more advanced and better instruction set which makes much more sense from a technical point of view. Many common things require much more RISC-V instruction than ARMv8 instructions. The only good reason to use RISC-V instead of ARM is to avoid paying licence fees to ARM.

Example Arch: My Mac

Hardware Overview:

Model Name:	MacBook Air
Model Identifier:	MacBookAir5,2
Processor Name:	Dual-Core Intel Core i5
Processor Speed:	1.8 GHz
Number of Processors:	1
Total Number of Cores:	2
L2 Cache (per Core):	256 KB
L3 Cache:	3 MB
Hyper-Threading Technology:	Enabled
Memory:	4 GB
Boot ROM Version:	421.0.0.0.0
SMC Version (system):	2.5f9
Serial Number (system):	C02JHC5TDRVC
Hardware UUID:	385C5076-CFB8-5720-8

Other CPU Chips

Apple
A/M Series
ARM
SoC

See separate slide set: ***SoC***

New Apple A14/iPads

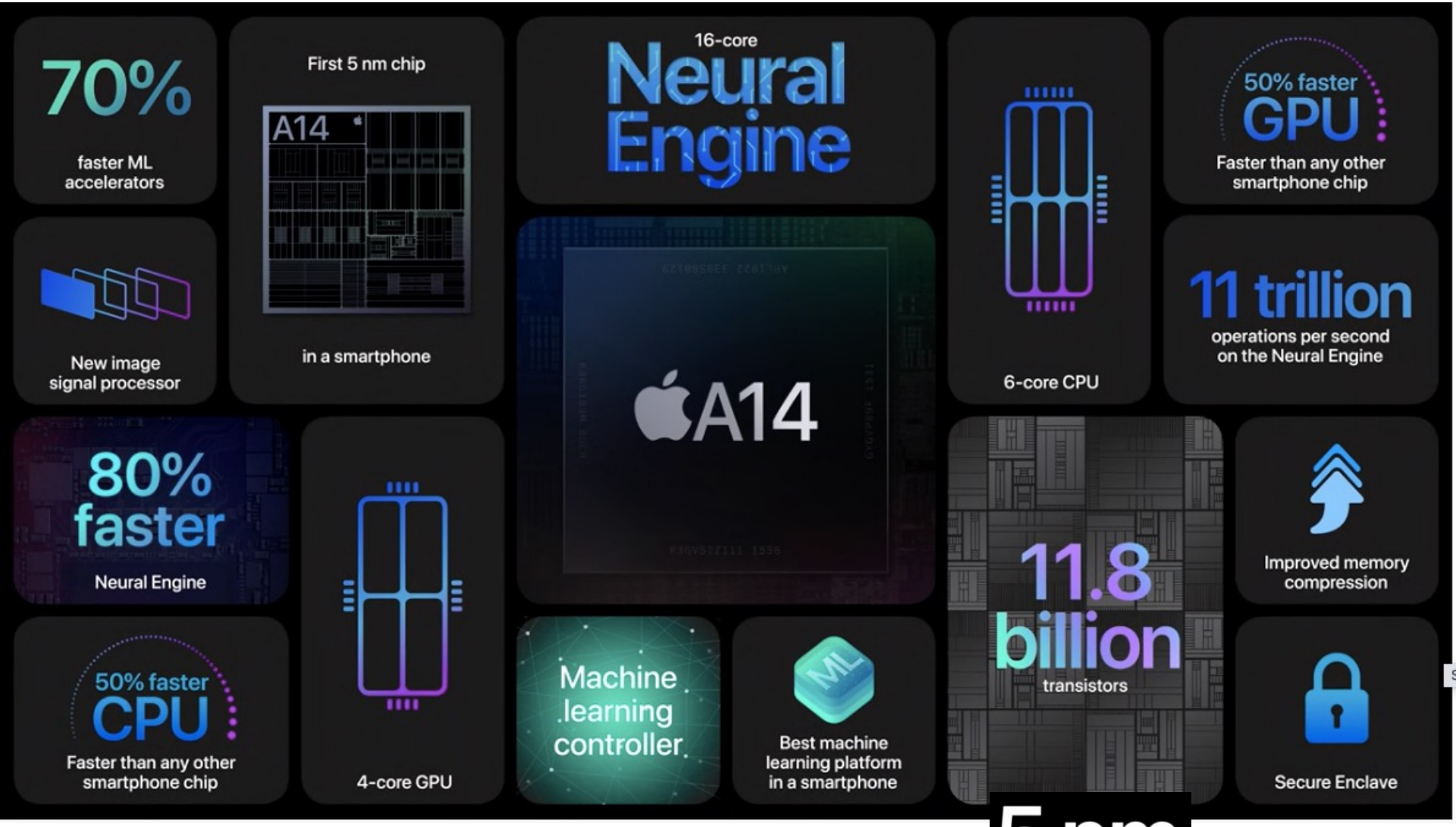
CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

COMP122

Apple A14

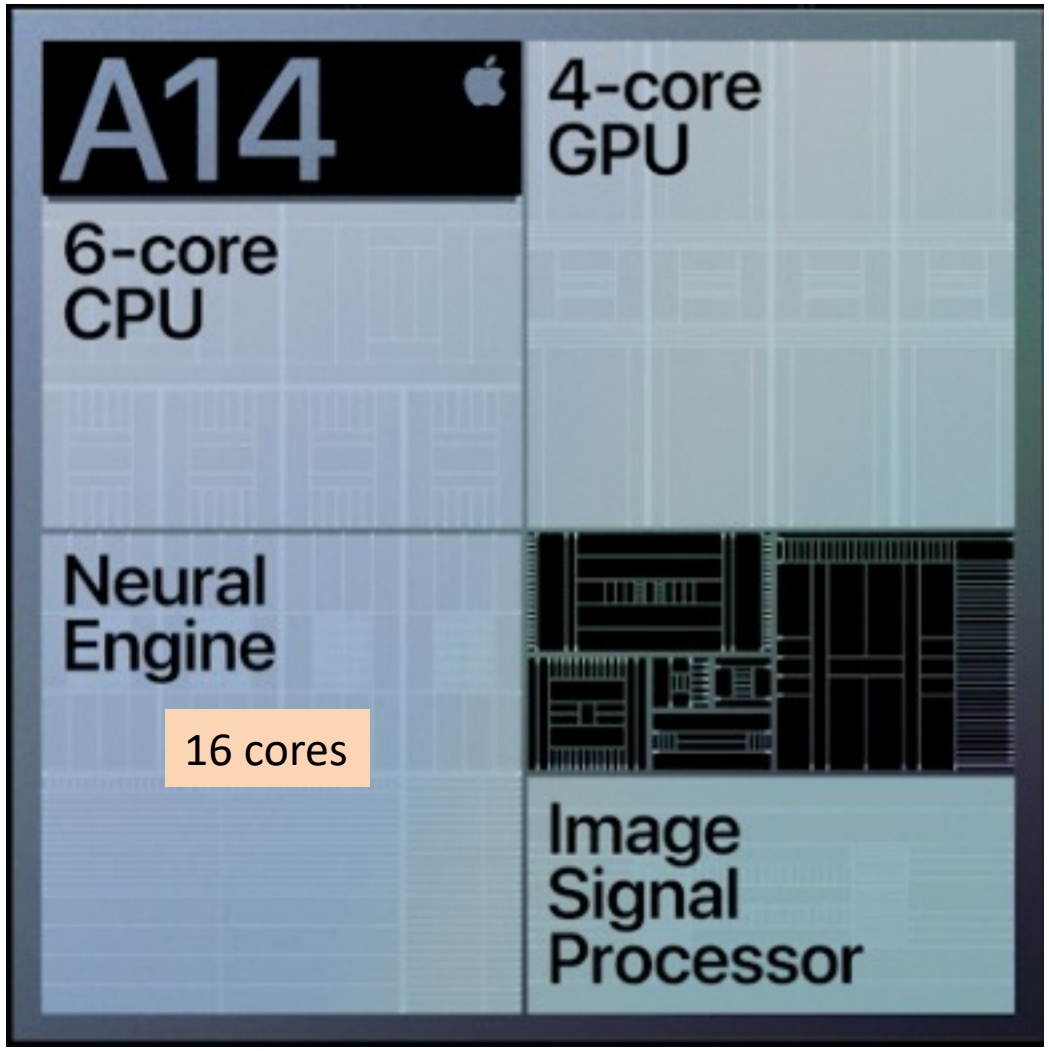
October 13, 2020

A14 Bionic



Apple A14

October 13, 2020

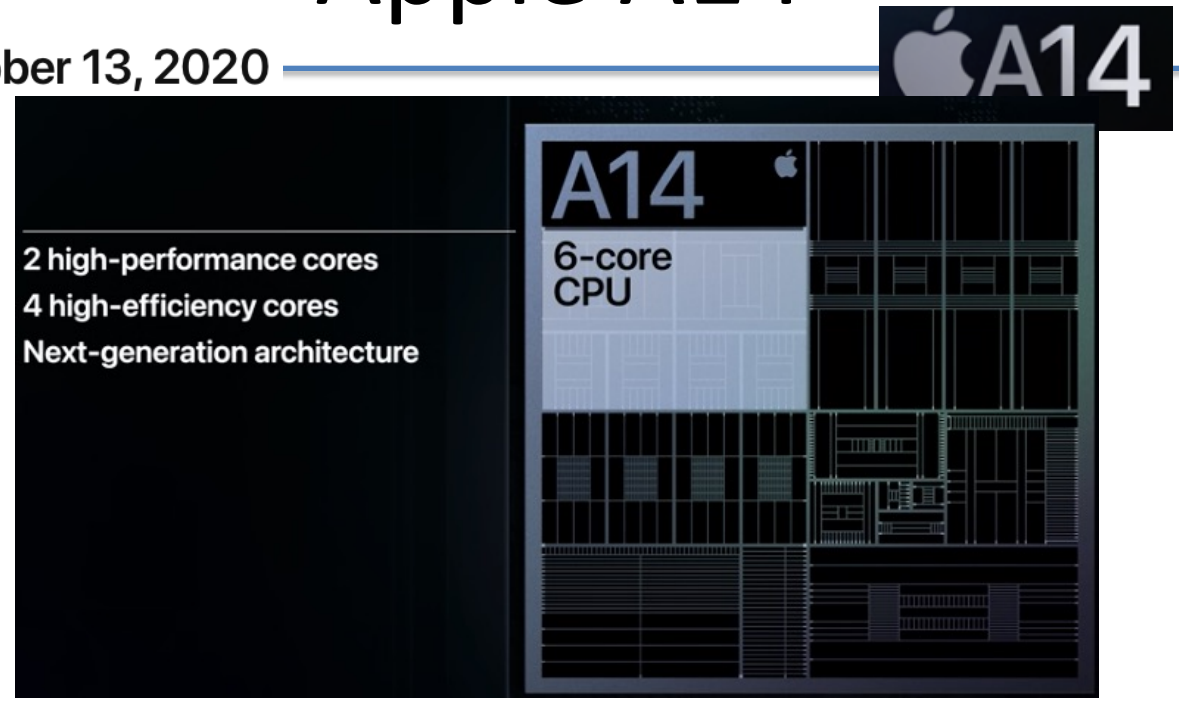


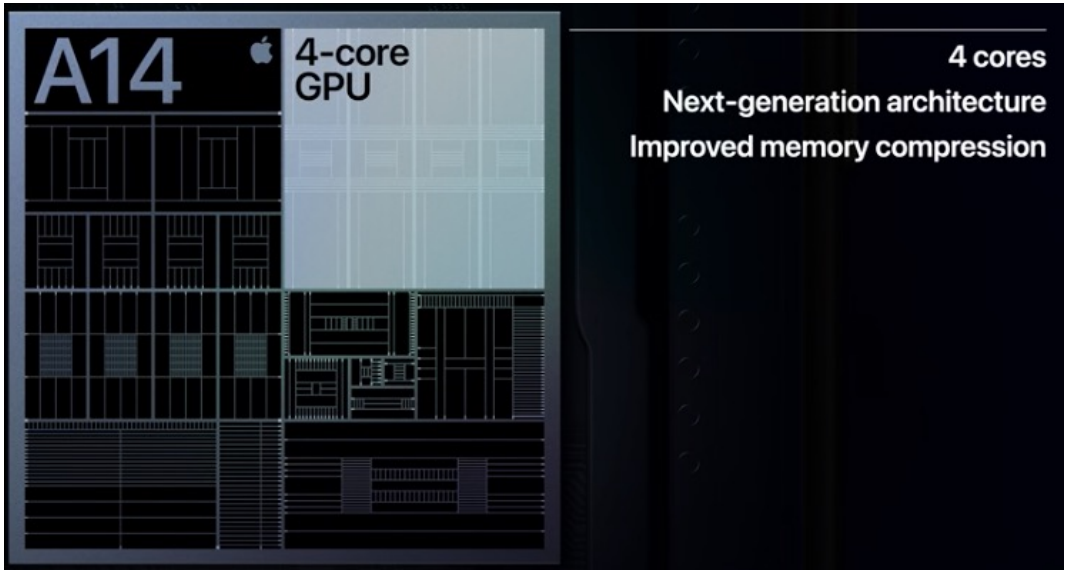
CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

COMP122

Apple A14

October 13, 2020



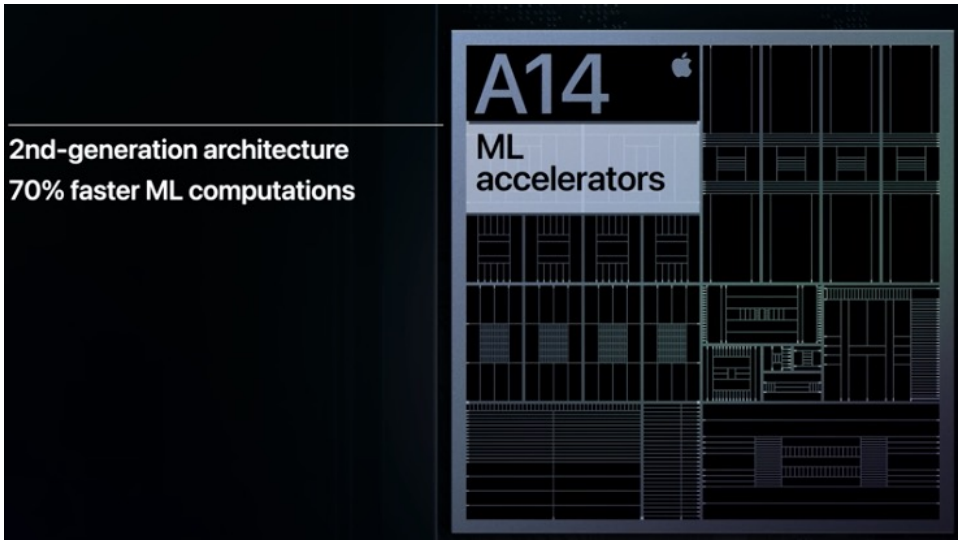


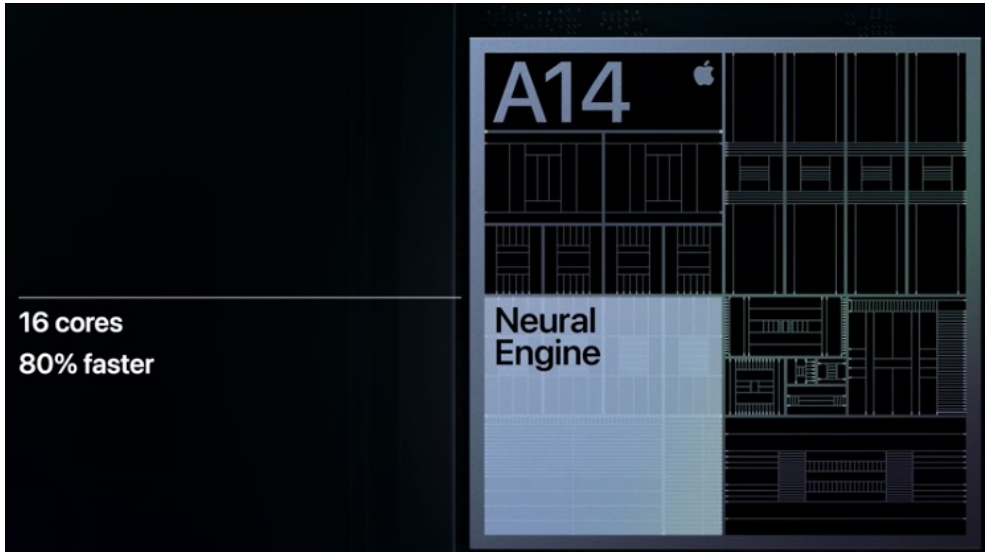
CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

COMP122

Apple A14

October 13, 2020





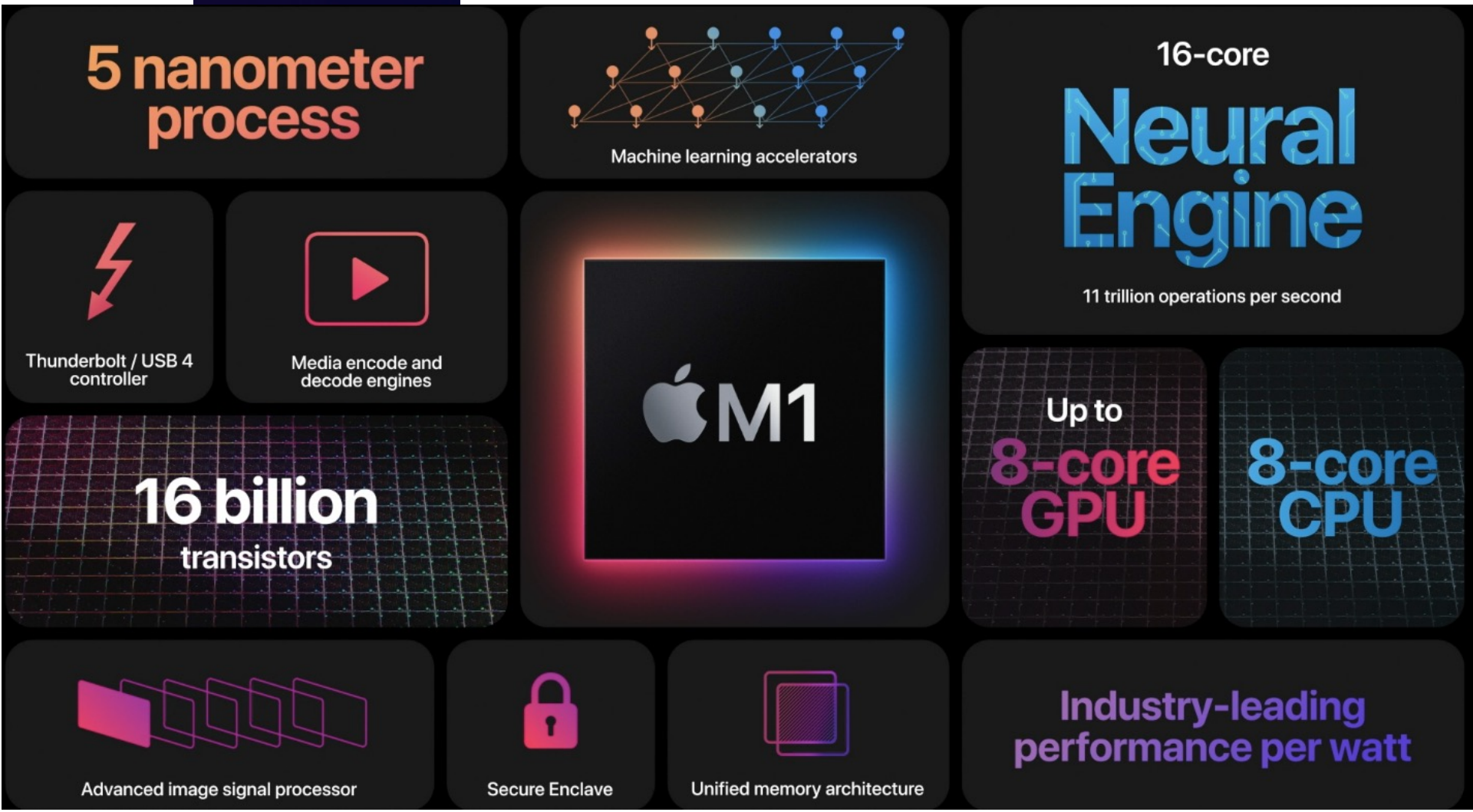
CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

COMP122

Apple Event

DR JEFF SOFTWARE
INDIE APP DEVELOPER

November 10, 2020



Apple M1

5-nanometer process

The first personal computer chip built with this cutting-edge technology.

16 billion transistors

The most we've ever put into a single chip.

Apple M1 Module

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

COMP122

Apple M1

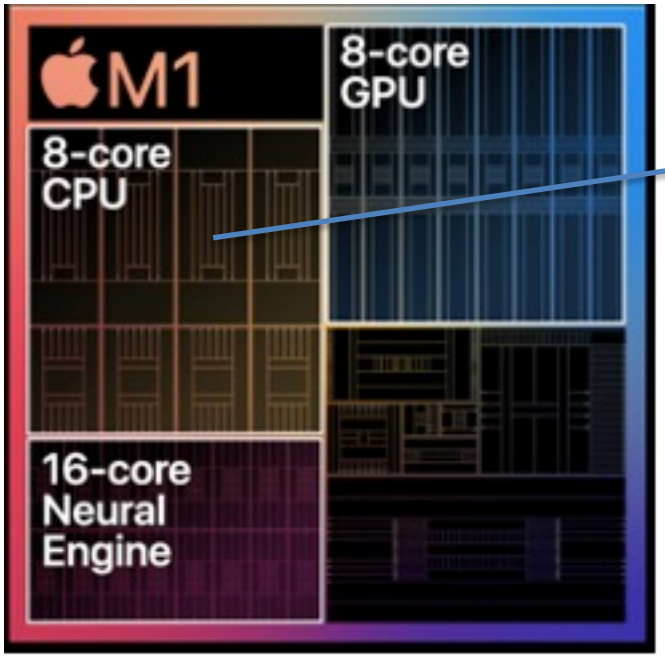
November 10, 2020

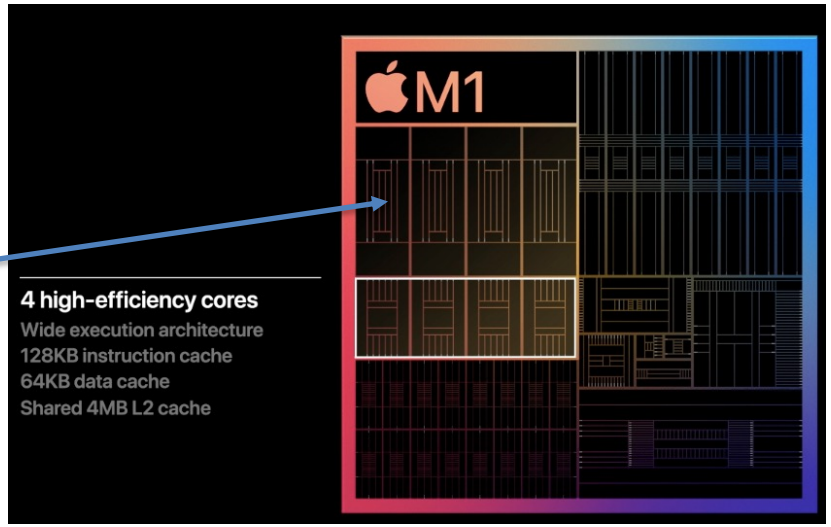
11 trillion
Operations per second

11 Tera FLOPS

- Cores
  - 8 CPU
  - 8 GPU
  - 16 NPU

- CPU cores
  - 4 Hi Perf (20W)
  - 4 Hi Efficiency (1.3W low power)





Other CPU Chips

Nvidia

See separate slide set: ***SoC***

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
COMP122

Nvidia + ARM = HPC

Apr 9, 2021

Nvidia to Make Server Processor, Targets Intel Profit Center

(Bloomberg) -- Nvidia Corp. unveiled its first server microprocessors, extending a push into Intel Corp.'s most lucrative market with a chip aimed at handling the most complicated computing work. Intel shares fell about 4% and Nvidia jumped on the news.

Nvidia's stock rallied further, to a gain of about 6%, after the company said first-quarter revenue "is tracking" above its previous forecast.The graphics chipmaker has designed a central processing unit, or CPU, based on technology from Arm Ltd., a company it's trying to acquire from Japan's SoftBank Group Corp. The Swiss National Supercomputing Centre and U.S. Department of Energy's Los Alamos National Laboratory will be the first to use the chips in their computers, Nvidia said Monday at an online event.

Nvidia has focused mainly on graphics processing units, or GPUs, which are used to power video games and intensive computing tasks in data centers. CPUs, by contrast, are a type of chip that's more of a generalist and can do basic tasks like running operating systems. Expanding into this product category opens up more revenue opportunities for Nvidia.

ARM CPU

Super-Computers

Other CPU Chips

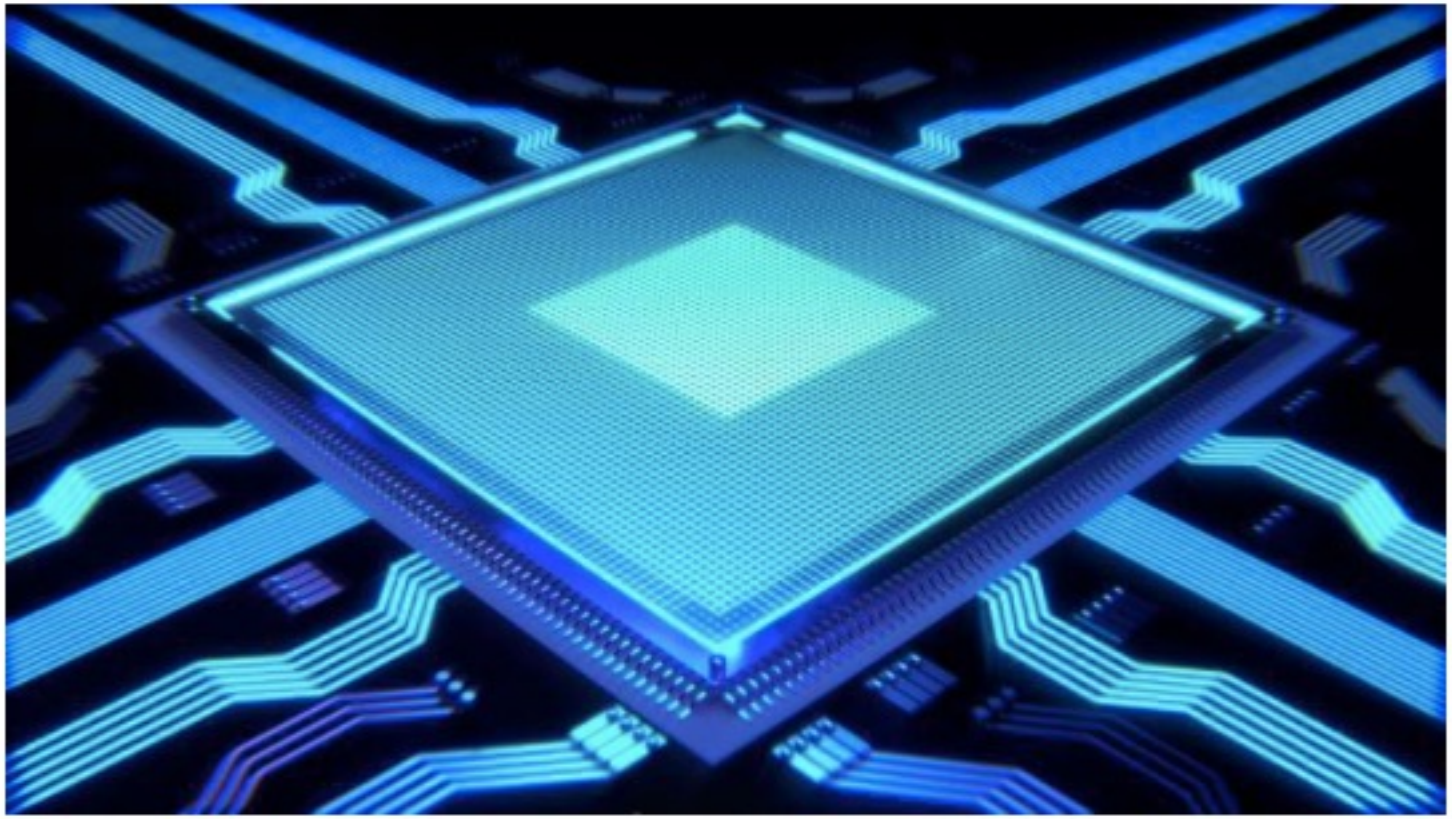
Google

See separate slide set: ***SoC***

Google Chips

Google trains chips to design themselves

by Peter Grad , Tech Xplore



Other CPU Chips

Tesla

See separate slide set: ***SoC***

Tesla Computers

Summary [edit]

Name	Autopilot hardware 1	Enhanced Autopilot hardware 2.0[a]	Enhanced Autopilot hardware 2.5 (HW2.5)[b]	Full self-driving computer (FSD) hardware 3[c]
Hardware	**Hardware 1**	**Hardware 2**[71]		**Hardware 3**
Initial availability date	2014	October 2016	August 2017	April 2019
Computers				
Platform	MobilEye EyeQ3[119]	NVIDIA DRIVE PX 2 AI computing platform[120]	NVIDIA DRIVE PX 2 with secondary node enabled[39]	Two identical Tesla-designed processors
Sensors				
Forward Radar	160 m (525 ft)[69]		170 m (558 ft)[69]	
Front / Side Camera color filter array	N/A	RCCC[69]	RCCB[69]	
Forward Cameras	1 monochrome with unknown range	3: Narrow (35°): 250 m (820 ft) Main (50°): 150 m (490 ft) Wide (120°): 60 m (195 ft)		
Forward Looking Side Cameras	N/A	Left (90°): 80 m (260 ft) Right (90°): 80 m (260 ft)		
Rearward Looking Side Cameras	N/A	Left: 100 m (330 ft) Right: 100 m (330 ft)		
Sonars	12 surrounding with 5 m (16 ft) range	12 surrounding with 8 m (26 ft) range		

Tesla designs