CS698Y: Modern Memory Systems
Lecture-1 (Introduction)

Biswabandan Panda

biswap@cse.iitk.ac.in

https://www.cse.iitk.ac.in/users/biswap/CS698Y.html

BACK TO
SCHOOL

WELCOME!

Course Staff

Instructor: Biswa (~~Biswabandan, Sir, Prof., Dr., Er., *-Biswa~~)

Website: www.cse.iitk.ac.in/biswap

Contact: KD 203, biswap@cse.iitk.ac.in
Office Hours: Mon/Thurs: 12 noon, by appointment

Teaching and Research Interests: Computer Architecture and Systems

What, When, and Where of CS698Y

When: Mon/Thurs 10.30-12 Hrs, Where: KD 103, What: You know it

Course website: www.cse.iitk.ac.in/~biswap/CS698Y.html

Piazza: For online discussions

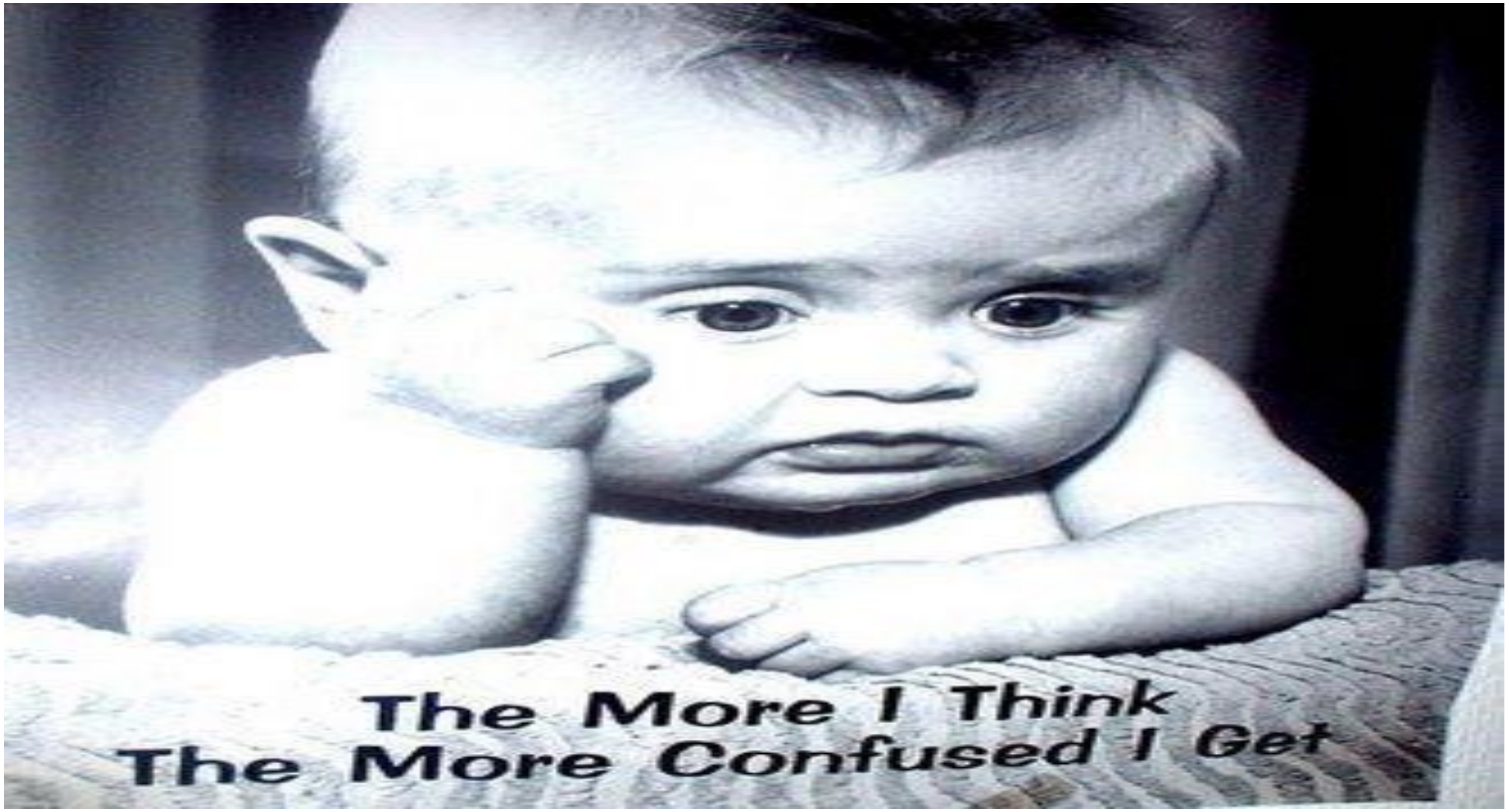
Submission of Assignments: Canvas

Register ASAP (*Wait* till you see the next few slides)

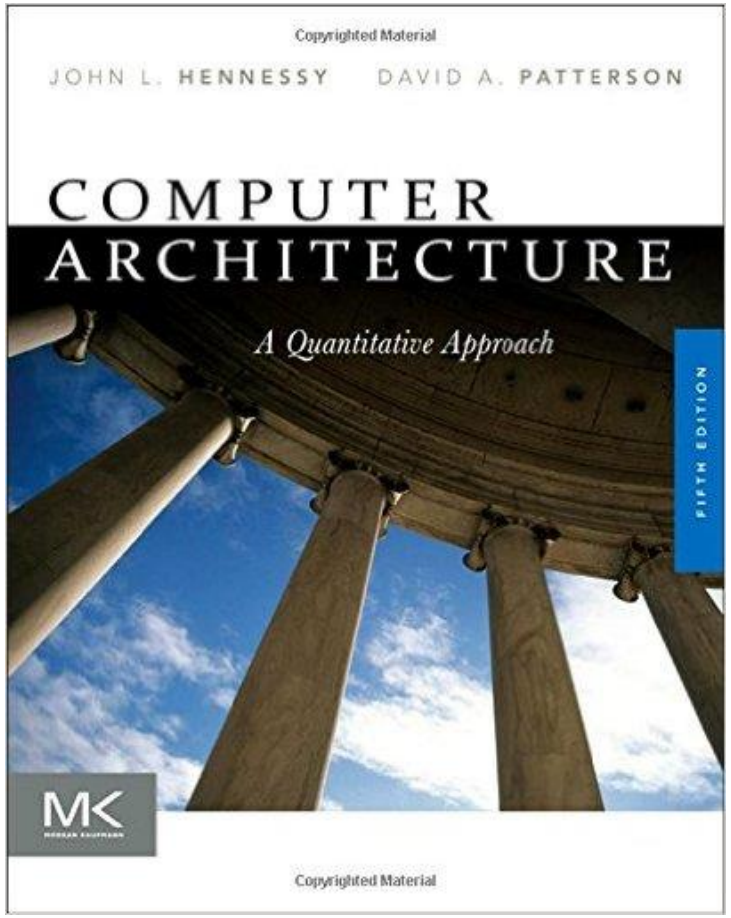
Let's Get Started

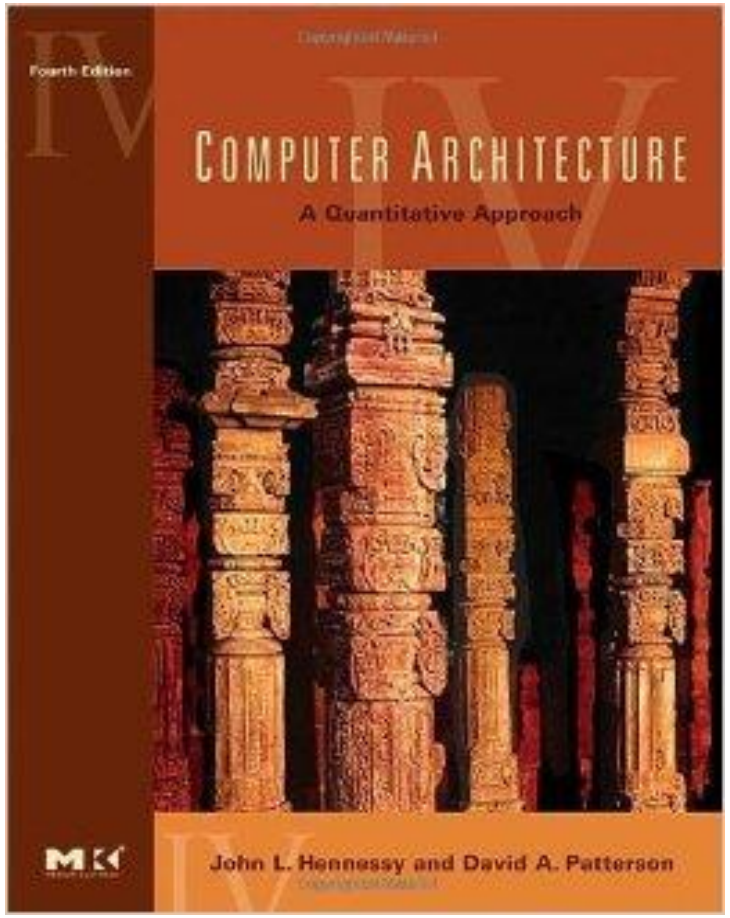
SNCF

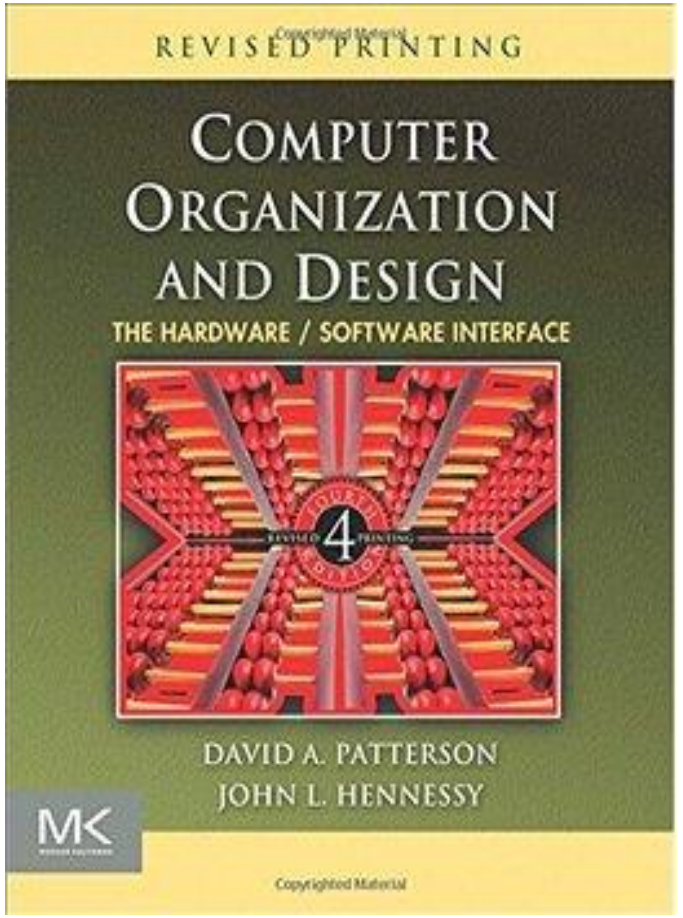
What is the takeaway?



Read this?


Copyrighted Material
JOHN L. HENNESSY DAVID A. PATTERSON
COMPUTER ARCHITECTURE
A Quantitative Approach
FIFTH EDITION
MK
Copyrighted Material


Fourth Edition
COMPUTER ARCHITECTURE
A Quantitative Approach
MK
John L. Hennessy and David A. Patterson


REVISED PRINTING
COMPUTER ORGANIZATION AND DESIGN
THE HARDWARE / SOFTWARE INTERFACE
4
DAVID A. PATTERSON
JOHN L. HENNESSY
MK
Copyrighted Material

Prerequisite

- Instruction pipelining
- LOAD/STORE, PC
- Cache, L1/L2
- Tag/Index/Offset
- Direct/Associative mapping
- SRAM/DRAM
- Latency/Throughput
- Virtual/Physical address
- Process/Thread
- Programming in C/C++

Score yourself

10 – Expert
5 – Knowledgeable
0 – No Knowledge

Your score

> 60	– *Welcome*
> 30 & ≤ 60	– Let's Talk
< 30	– Next Time

What is Expected From You?

No open-screens (no nomophobics): No open smart-phones (phones) & laptops/tablets. Keep your phones in silent mode.

Open-screens will affect (distract) you, your friends, and me.

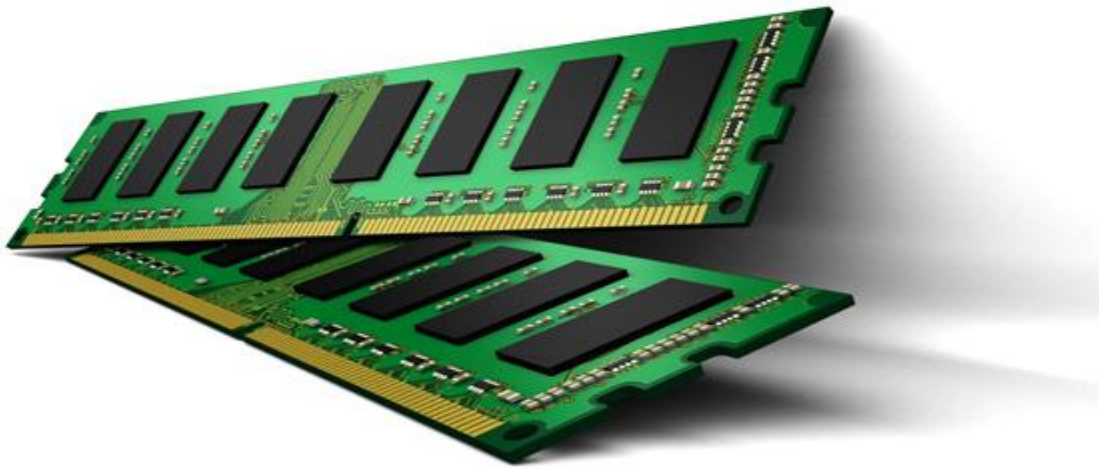
Ask questions & participate in in-class discussions (worth 10 points ☺)

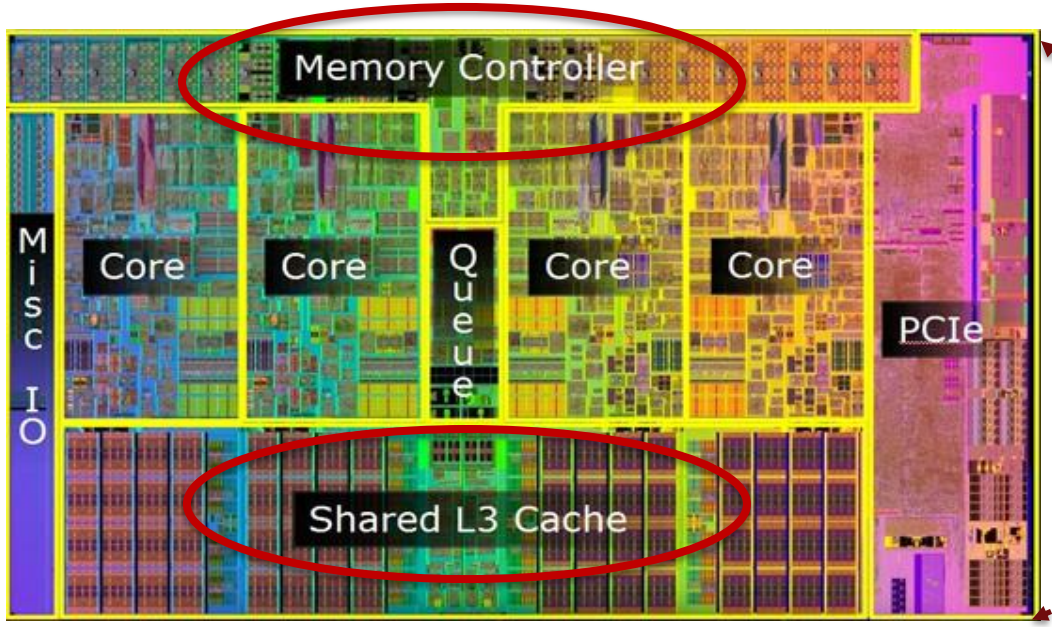
Paper reading and writing reviews/reports

Understand, implement, and analyze ideas

Slides will not contain everything. So attend lectures.

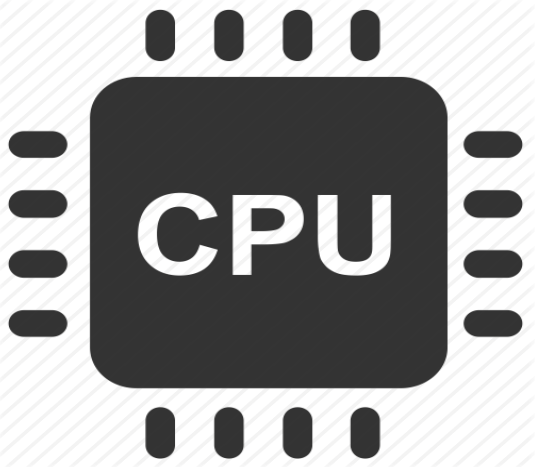
Memory System



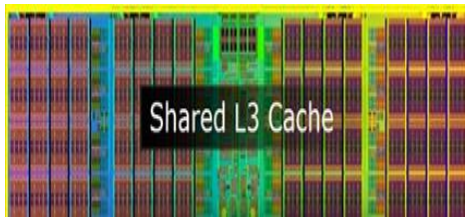


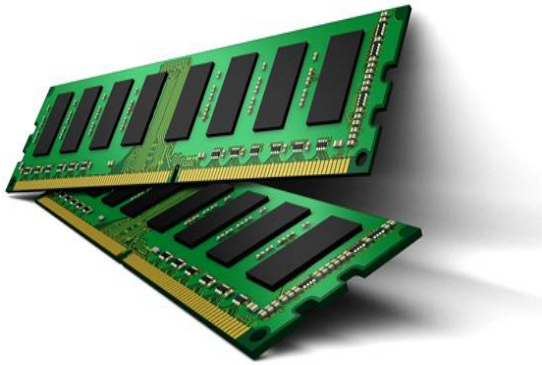
Source: Intel

Why Memory Systems (CS698Y) ?

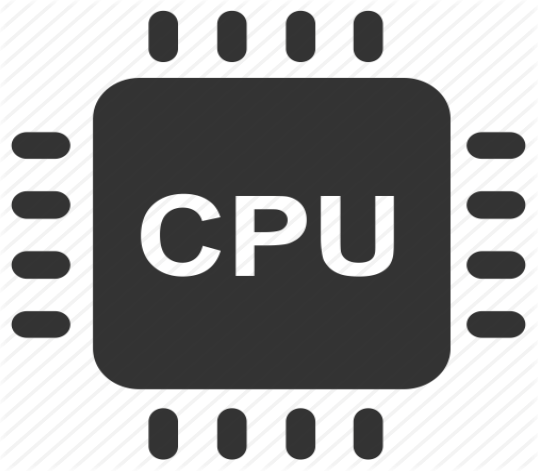


Huge gap?



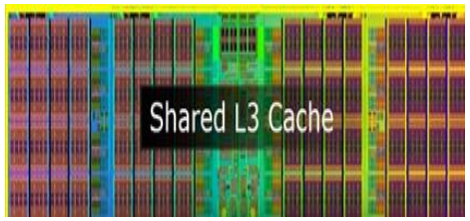


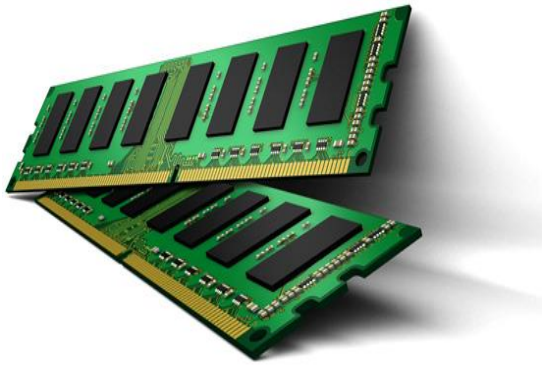
Why Memory Systems (CS698Y) ?



Huge gap?

Demand-supply



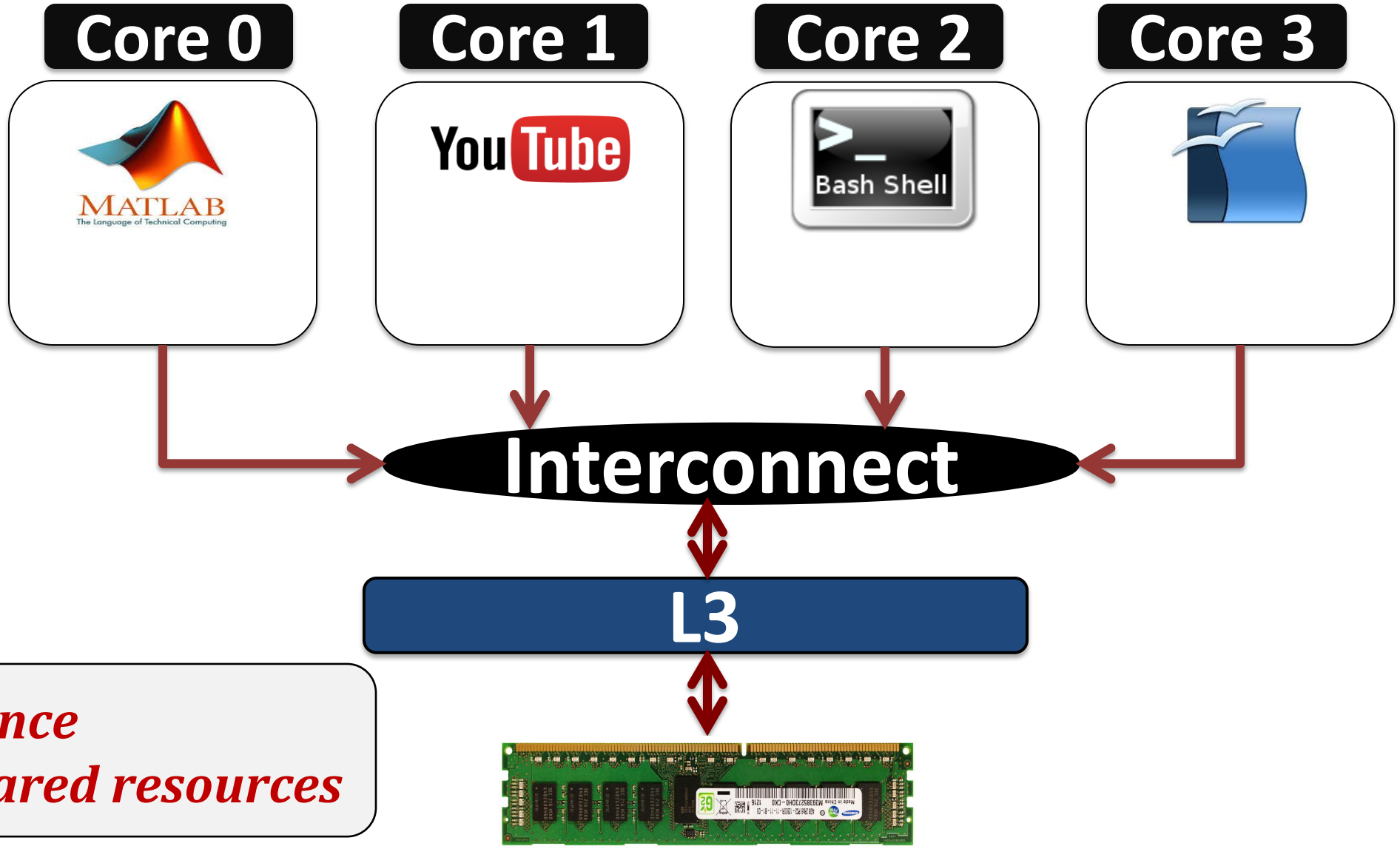


- Latency - 100s of cycles
- Energy and Power
- Fixed bandwidth - GB/sec
- Reliability and Security
- Capacity - GBs
- Fairness and Quality of Service

LATENCY

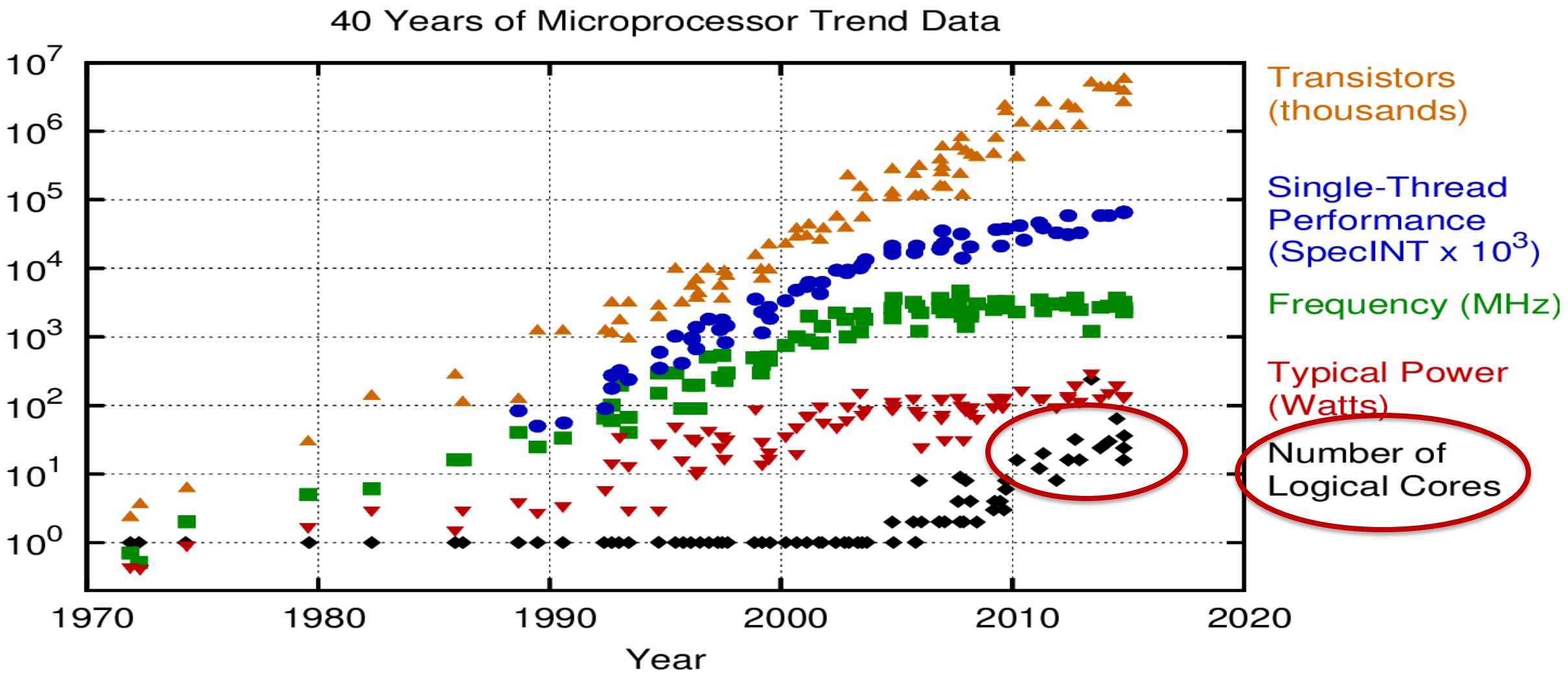
Performance

Interference @Shared Resources



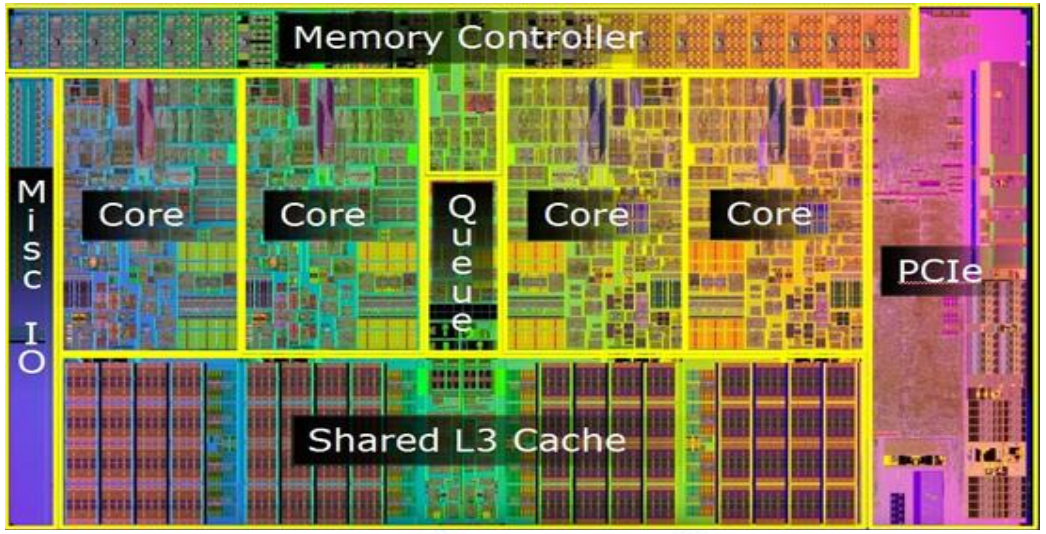
Interference at the shared resources

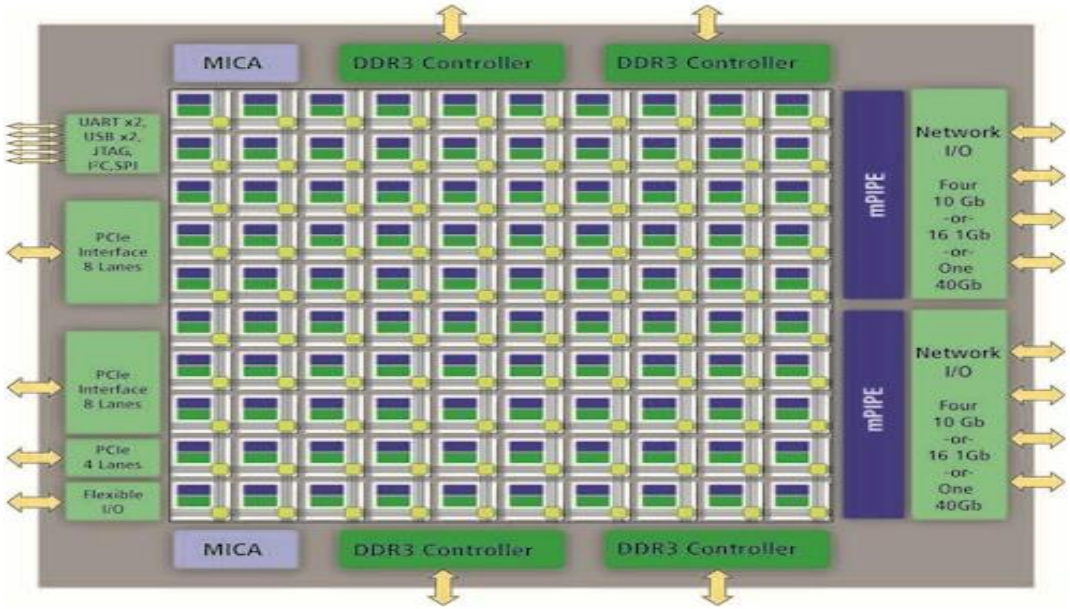
Trend on Core Count

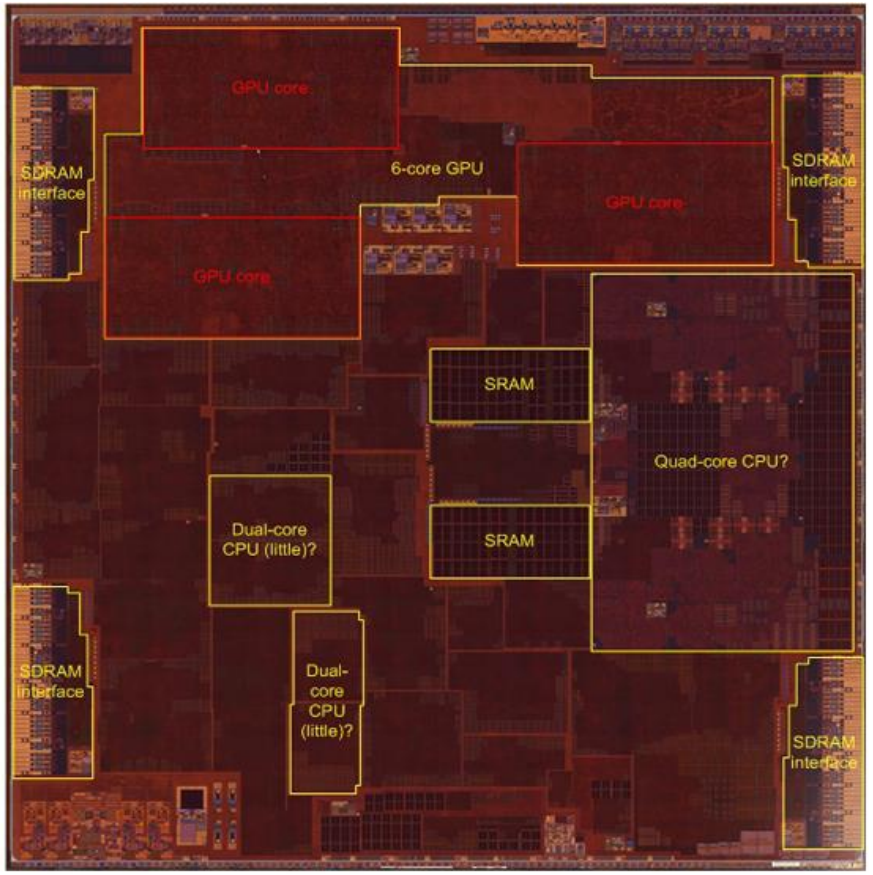


Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2015 by K. Rupp

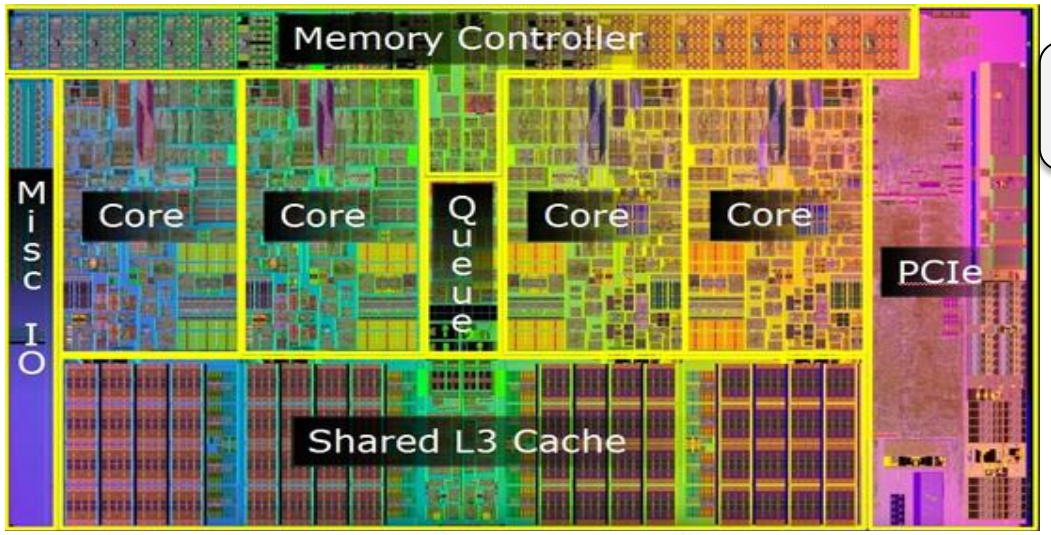
Multicores


Memory Controller
Misc IO
Core
Core
Queue
Core
Core
PCIe
Shared L3 Cache

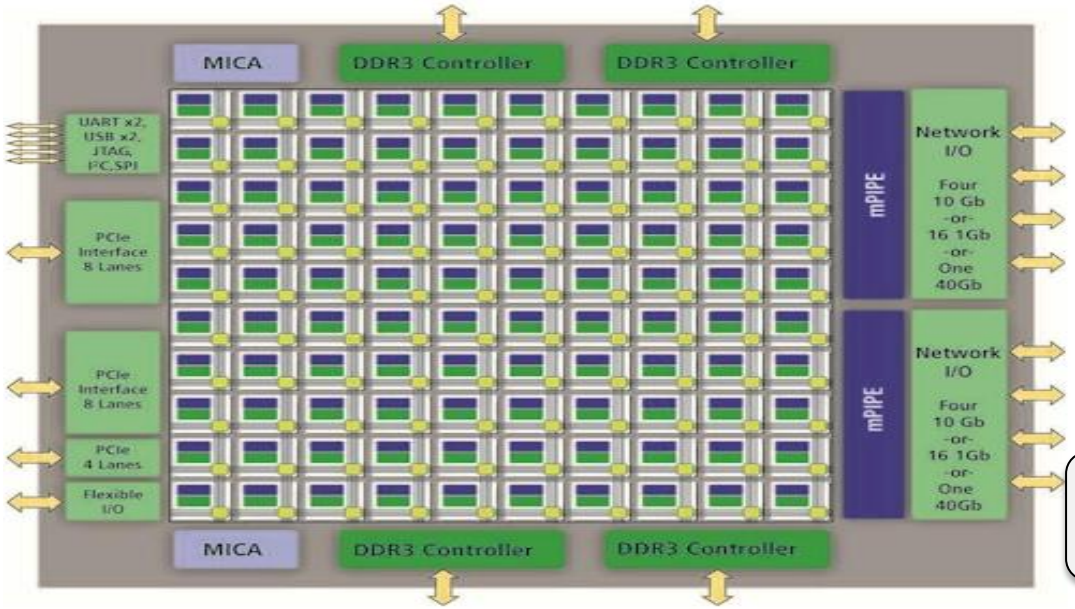

MICA
DDR3 Controller
DDR3 Controller
UART x2, USB x2, JTAG, I²C, SPI
PCIe Interface 8 Lanes
PCIe Interface 8 Lanes
PCIe 4 Lanes
Flexible I/O
mPIPE
mPIPE
Network I/O
Four 10 Gb -or- 16 1Gb -or- One 40Gb
Network I/O
Four 10 Gb -or- 16 1Gb -or- One 40Gb
MICA
DDR3 Controller
DDR3 Controller


GPU core
SDRAM interface
6-core GPU
GPU core
SDRAM interface
GPU core
SRAM
Quad-core CPU?
Dual-core CPU (little)?
SRAM
SDRAM interface
Dual-core CPU (little)?
SDRAM interface

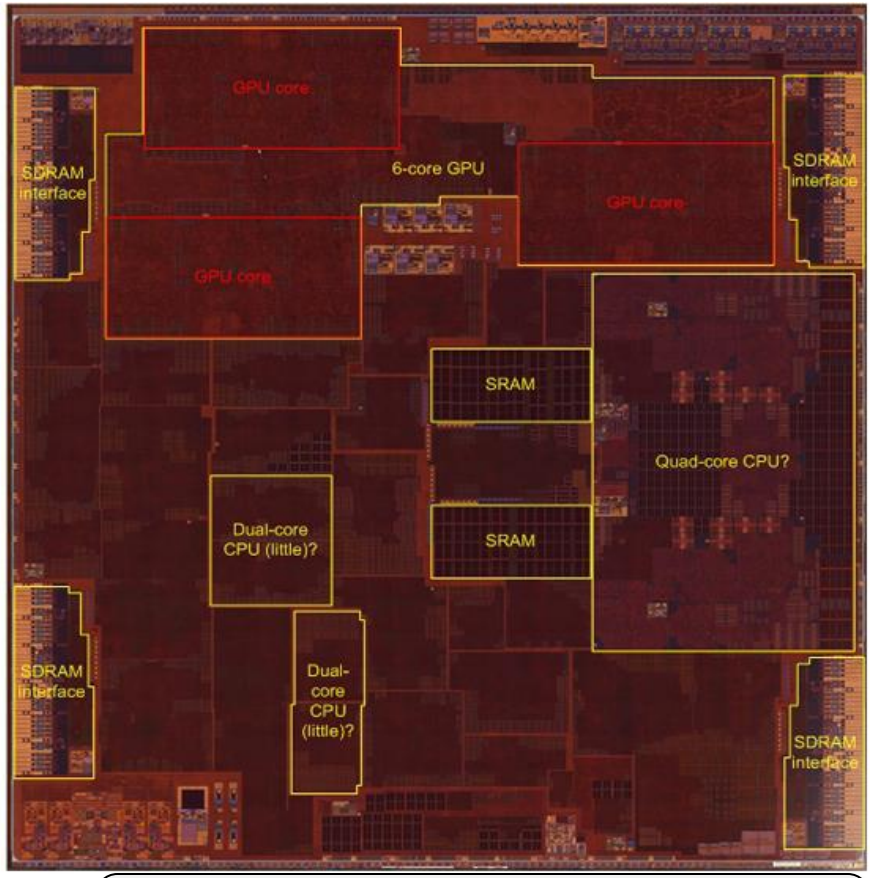
Multicores



4-core system (intel core i series)



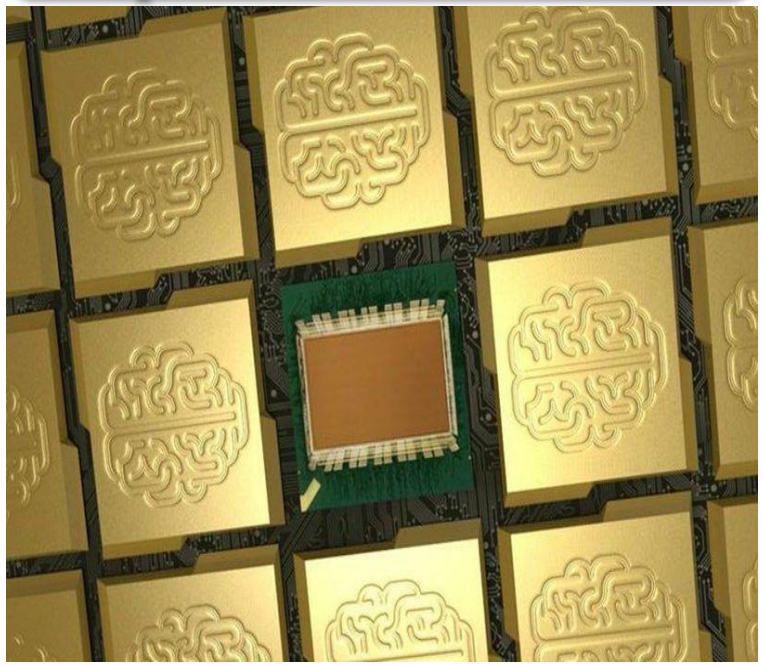
100-core system (Tilera)



Heterogeneous cores (iphone 7)

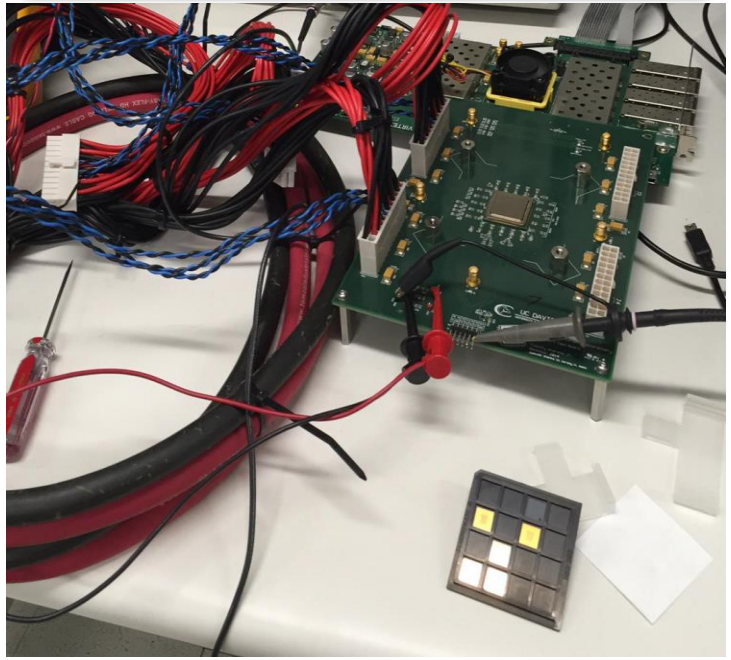
Multi to Many

IBM's brain like chip: 4096 cores



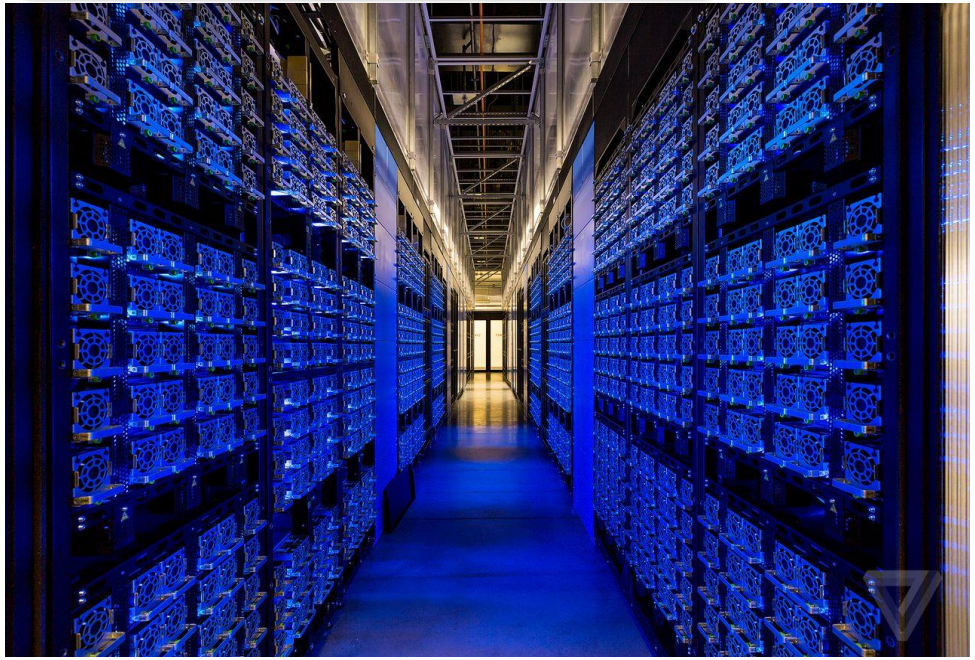
Source: IBM

1000 core chip from UC Davis



Source: UC Davis

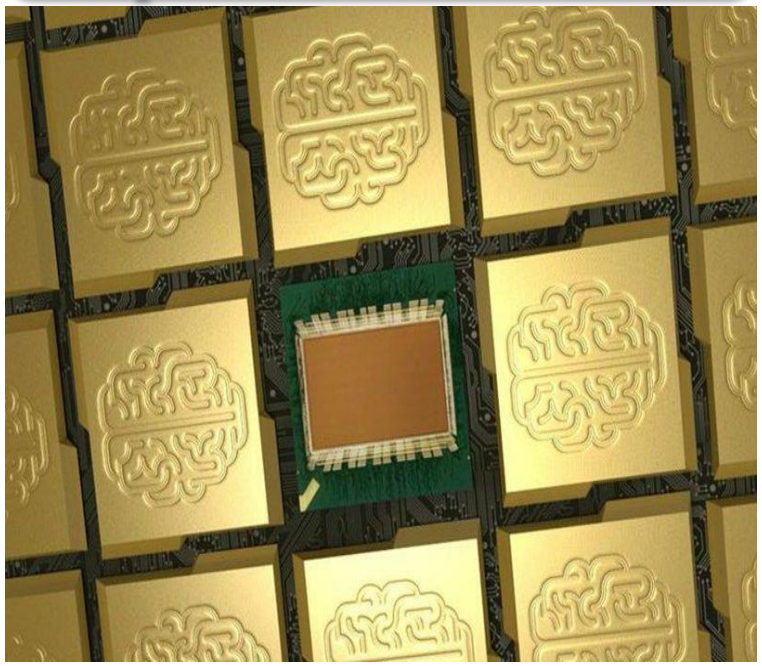
??



Source: The Verge

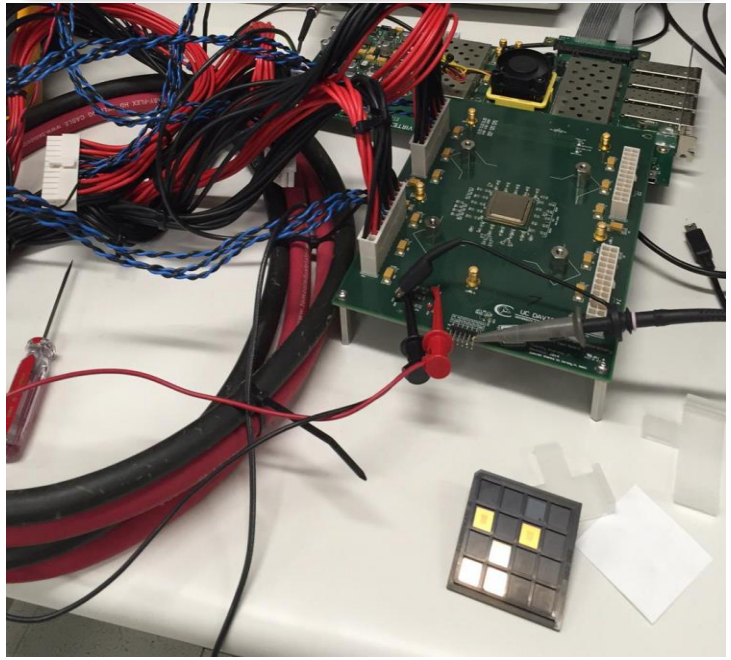
Multi to Many

IBM's brain like chip: 4096 cores



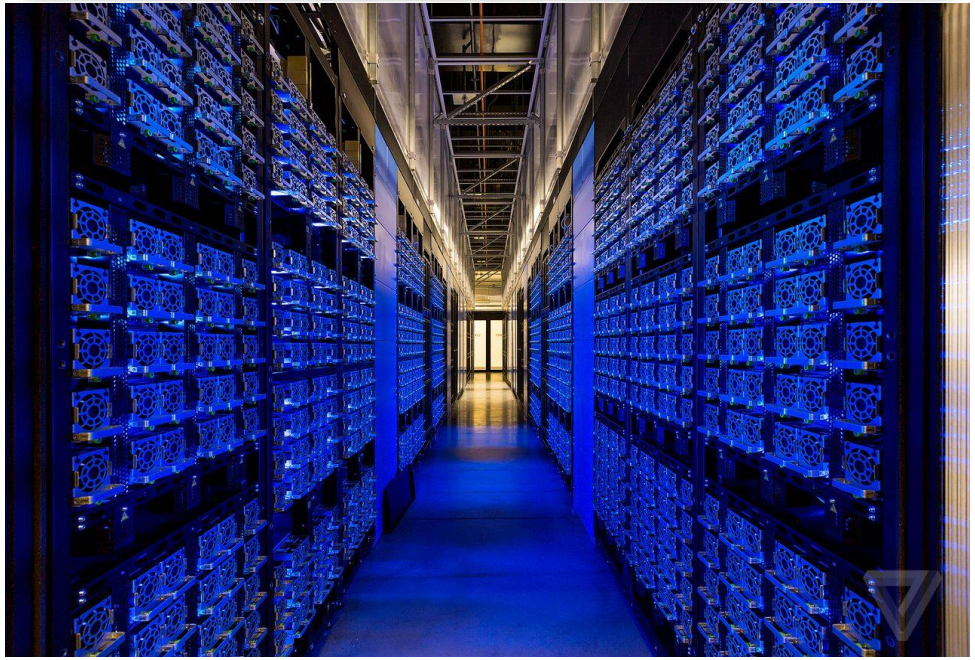
Source: IBM

1000 core chip from UC Davis



Source: UC Davis

Datacenter @Facebook



Source: The Verge

Ideally ??

Source:
Erlang@SinaWeibo

Interference ??

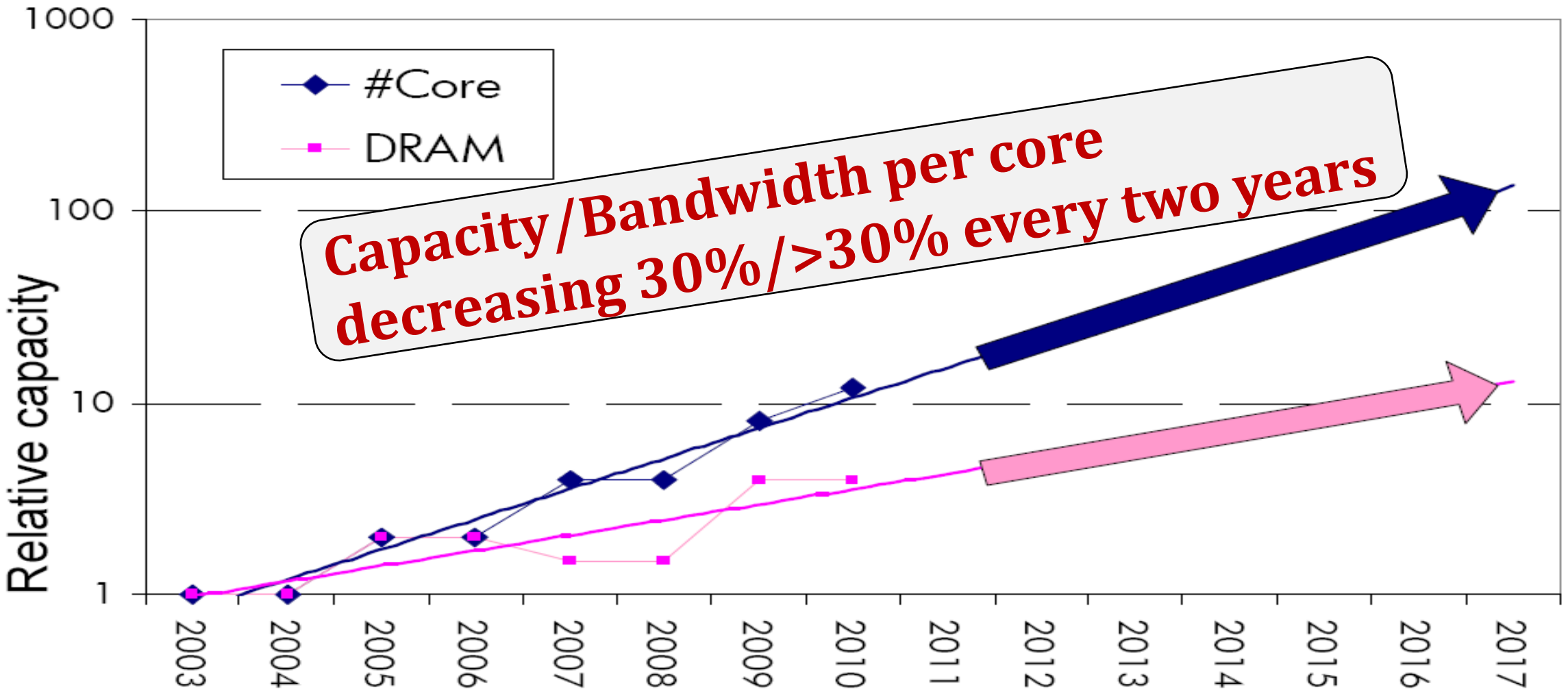
Source:
Erlang@SinaWeibo

#cores doubling every two years. *More interference ??*

CAPACITY & BANDWIDTH

Data-intensive Applications

Trend on Memory Capacity & Bandwidth



Source: Lim et al., ISCA 2009

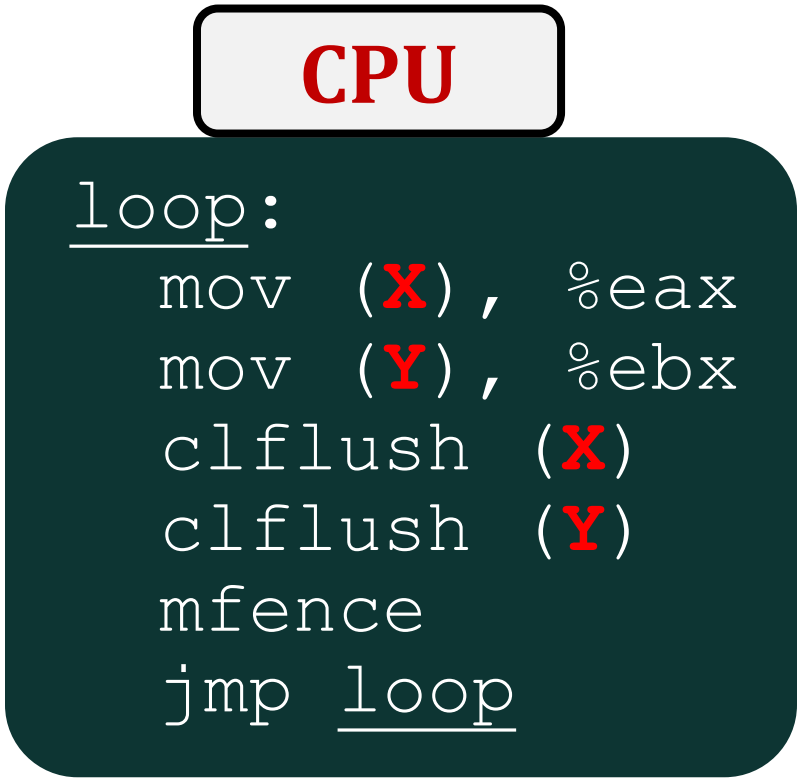
ENERGY & POWER

Energy bills & heating

RELIABILITY

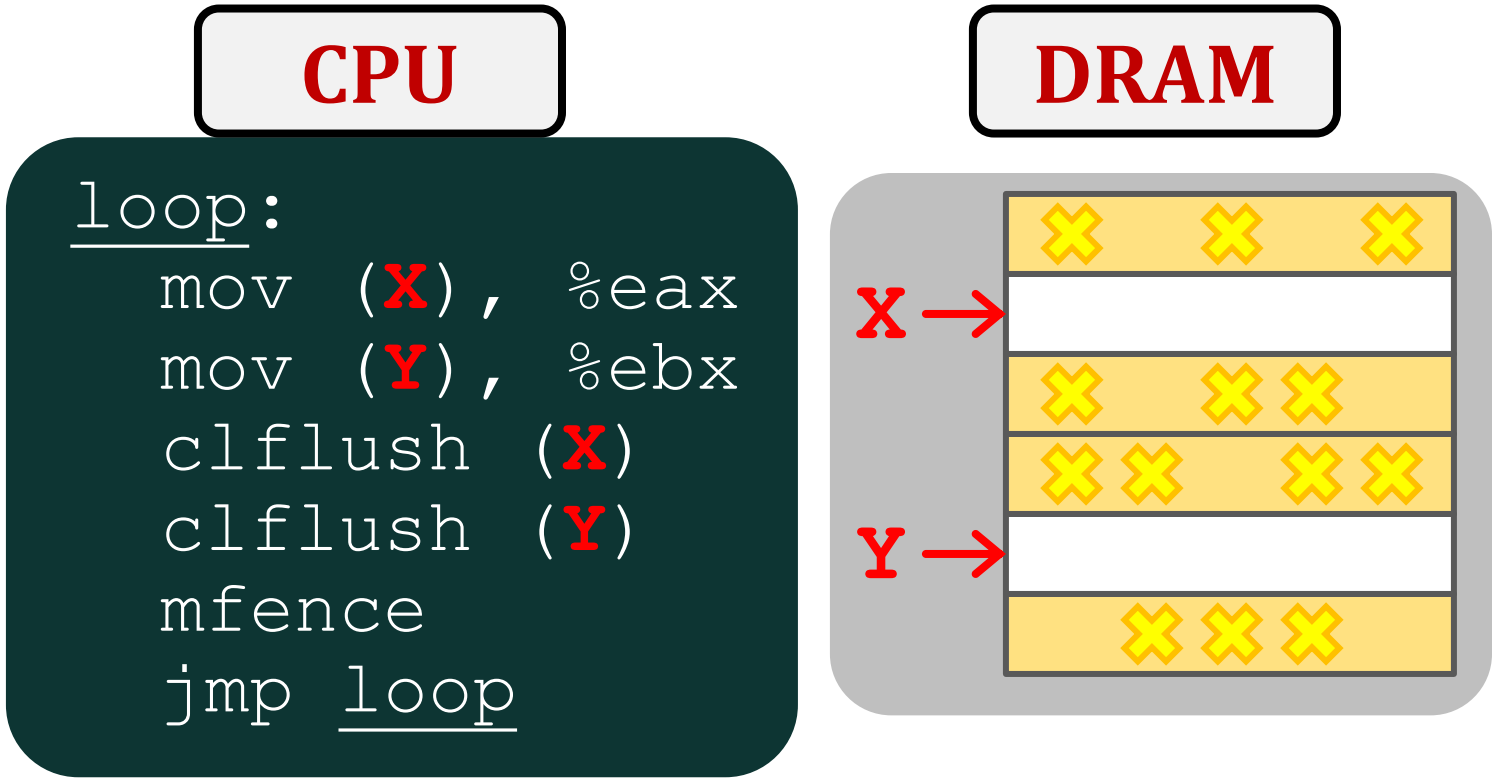
Loss of data

Row-Hammer Problem



Source: Kim et al., ISCA 2014

Row-Hammer Problem



Source: Kim et al., ISCA 2014

Row-Hammer Problem

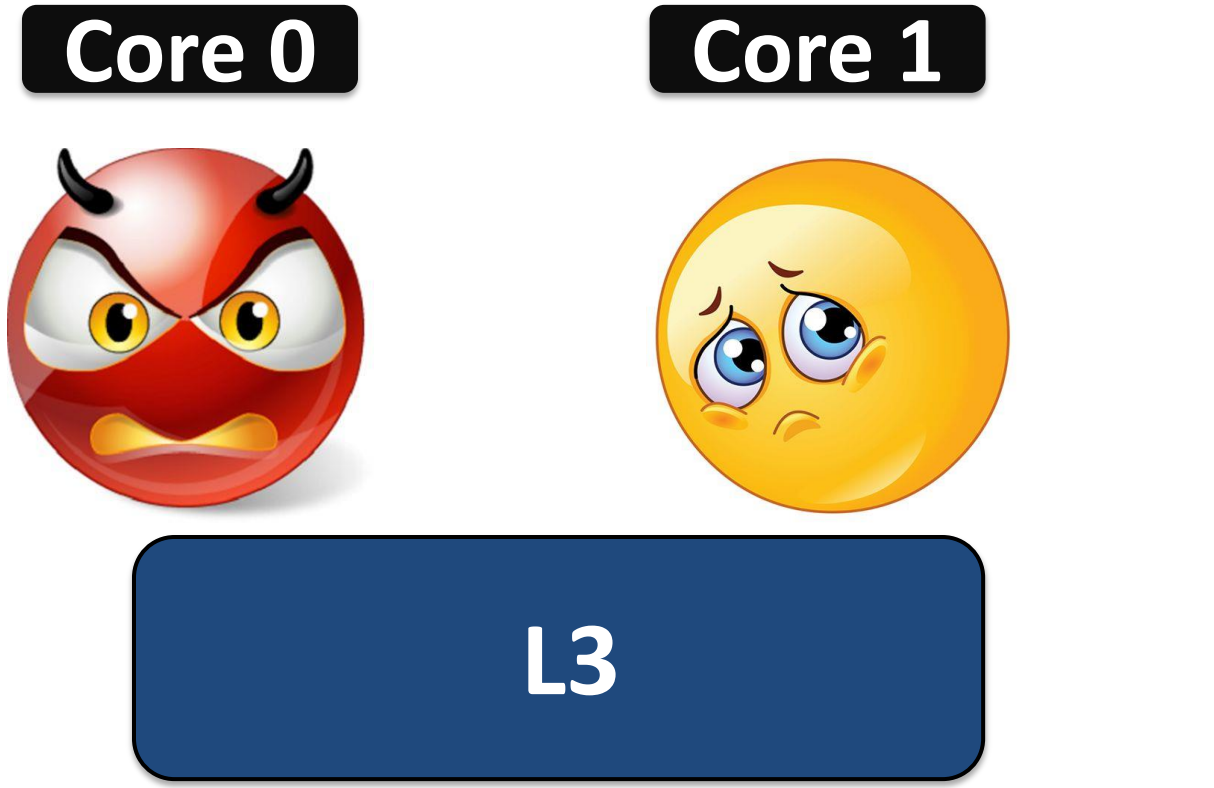
MODEL	#DEVICES	#VULNERABLE
ARMv7 (32-bit) devices		
LG Nexus 4	1	1
LG Nexus 5	15	12
Motorola Moto G (2013)	1	1
Motorola Moto G (2014)	1	1
OnePlus One	2	2
Samsung Galaxy S4	1	1
Samsung Galaxy S5	2	1
ARMv8 (64-bit) devices		
HTC Desire 510	1	0
Lenovo K3 Note	1	0
LG G4	1	1
LG Nexus 5X	1	0
Samsung Galaxy S6	1	0
Xiaomi Mi 4i	1	0

Source: fossbytes.com

SECURITY

Information leakage

Side/Covert-channel Attack



Source: Pinterest

FAIRNESS & QOS

Fair Slowdown

Minimum Performance Guarantee

Ideal Memory Systems

Latency – L

Bandwidth – H

Capacity – H

Energy and Power – L

Reliable and Secure – H (not at the cost of previous four)

Fairness and Quality of Service – H

Cost ??

CS 698Y is not a

Theory course

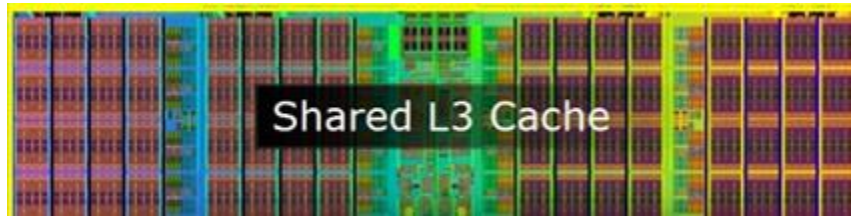
Digital/analog (☺) circuit level course

Course on HDL/VHDL modeling of memory systems

Course on microprocessors

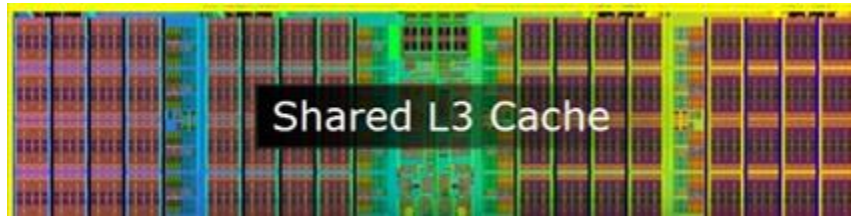
10,000 Feet View of CS698Y

Caches



10,000 Feet View of CS698Y

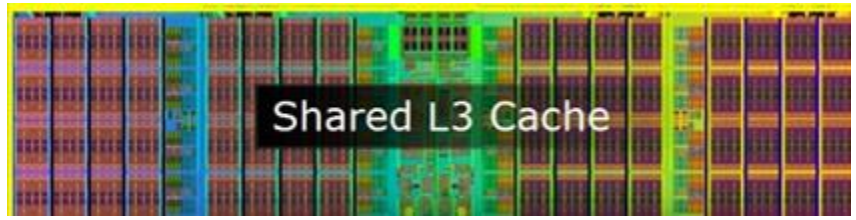
Caches



DRAM

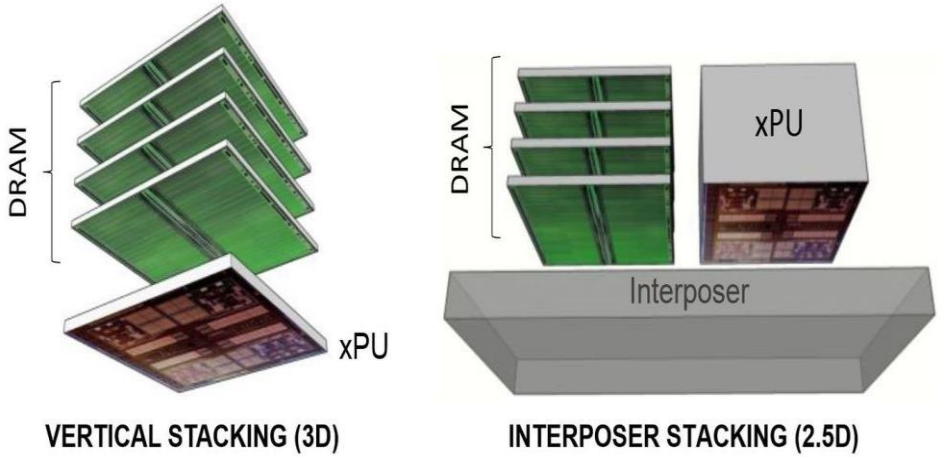
10,000 Feet View of CS698Y

Caches



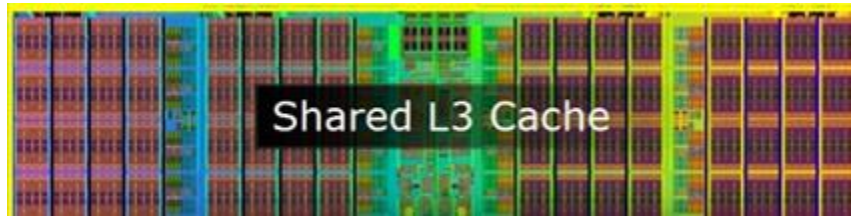
DRAM

Memory stacking



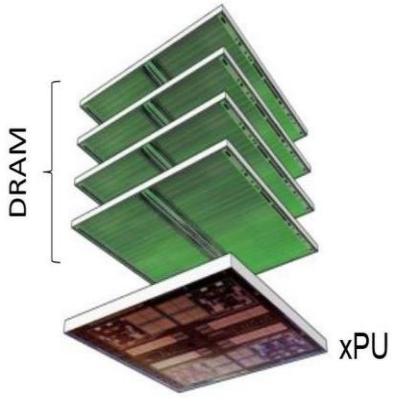
10,000 Feet View of CS698Y

Caches

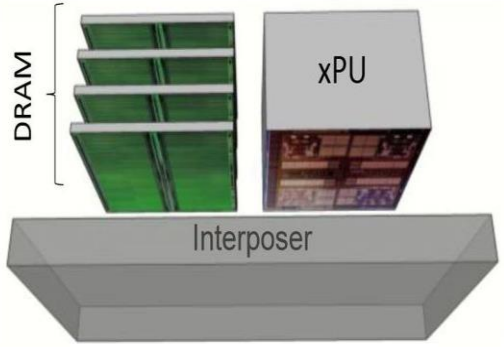


DRAM

Memory stacking

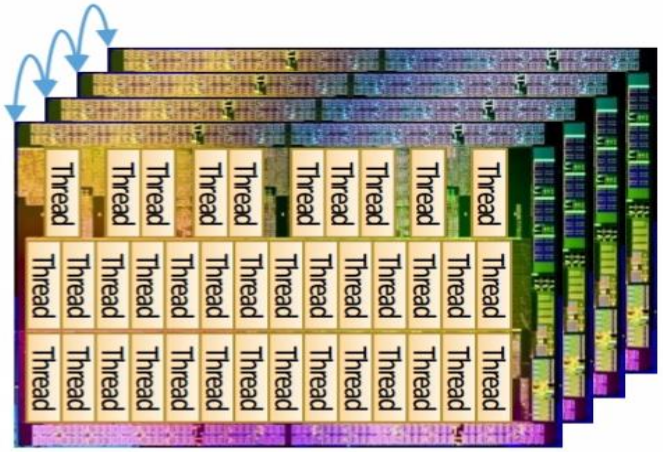


VERTICAL STACKING (3D)



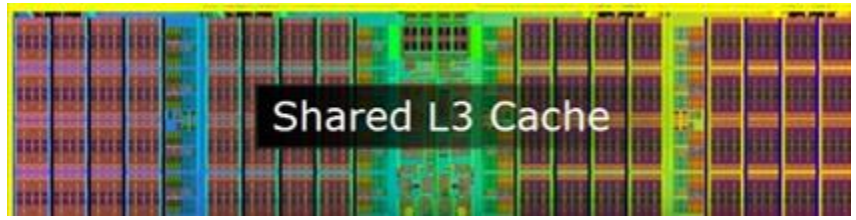
INTERPOSER STACKING (2.5D)

Processing in Memory



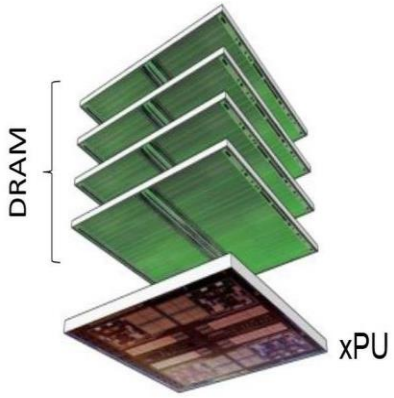
10,000 Feet View of CS698Y

Caches

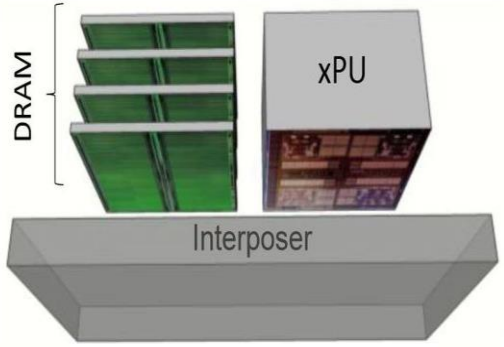


DRAM

Memory stacking

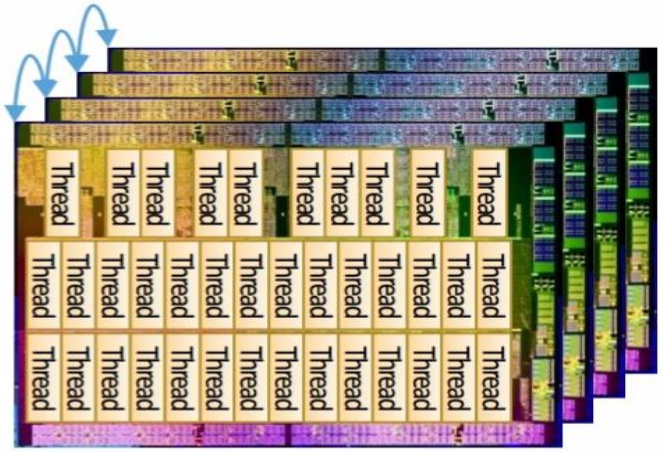


VERTICAL STACKING (3D)

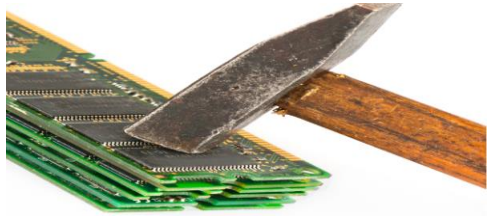


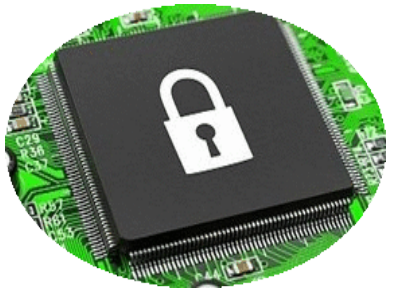
INTERPOSER STACKING (2.5D)

Processing in Memory



Secure and reliable Cache and DRAM

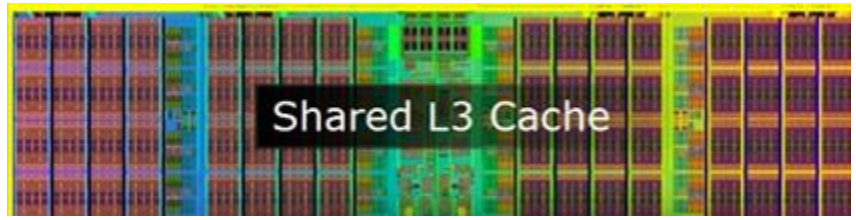




Source:Hackaday

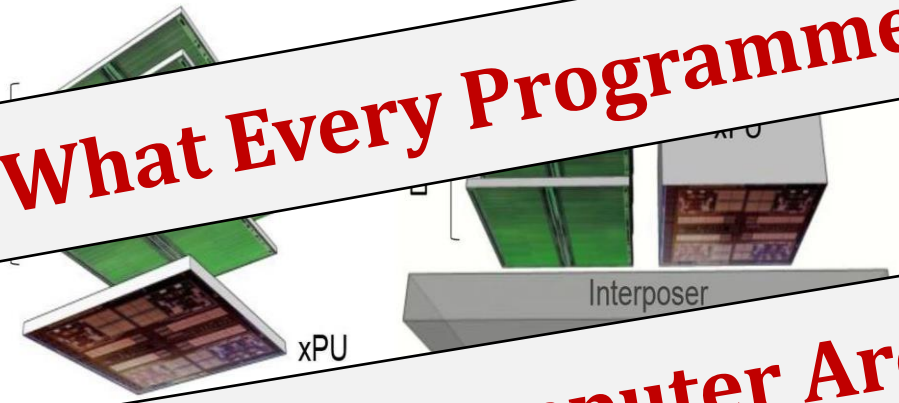
10,000 Feet View of CS698Y

Caches



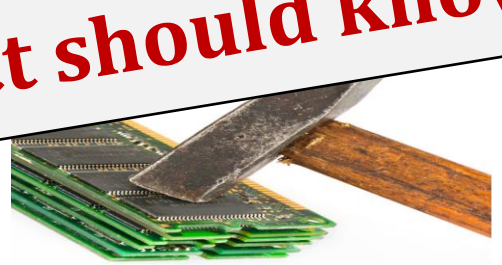
DRAM

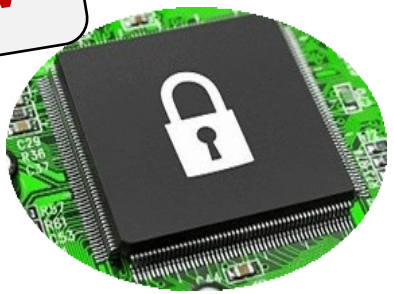
Memory stacking



Processing in Memory

Secure and reliable





Source:Hackaday

Application-driven Approach

What Every Programmer should know

What Every Computer Architect should know

Takeaways (learning objectives) from CS698Y

Understand and appreciate **the memory systems**

Analyze and evaluate
the performance (memory system) bottlenecks

Research **on memory systems**

Learning Points

Option-I:

30 = (3 × 10) = 3 programming assignments

40 = (2 × 20) = Quiz 1.0 and Quiz 2.0 (Optional Quiz 1.1 and Quiz 2.1) = max (Quiz 1.x) + max (Quiz 2.x)

20 = (2 × 10) = 2 paper reviews

10 = Classroom and Piazza participation

Option-II:

30 = (3 × 10) = 3 programming assignments

20 = (1 × 20) = max (Quiz 1.0, Quiz 1.1)

30 = (1 × 30) = 1 research project (weekly meetings)

10 = (2 × 5) = 2 paper reviews

10 = Classroom and Piazza participation

Bonus points for finding typos in slides

Iterative Assessment

Choose your option wisely ☺

NEXT TWO LECTURES

BASICS OF PROCESSOR & CACHE

ASSIGNMENT-0

Submit it by tonight (11:59 PM)

"It takes two to speak the truth - one to speak and another to hear" - **Henry David Thoreau**

Thank You & Have a Good day