Heterogeneous Architectures
Heterogeneous Computing

Professor: Dr. Joel Fuentes - jfuentes@ubiobio.cl

Assistants:
- Daniel López - daniel.lopez1701@alumnos.ubiobio.cl
- Sebastián González - sebastian.gonzalez1801@alumnos.ubiobio.cl

Course Webpage: http://www.face.ubiobio.cl/~jfuentes/classes/ch

Contents

- History of processors

- The rise of Heterogenous Systems

- Accelerators

History

- What are accelerators?

- Why is it important to learn to program different architectures?

- Why is it important to optimize our programs?

History

- Software and performance optimization was common since the computing resources were limited.
- A lot of programs simply would not run without being fully optimized.

IBM System/360
Courtesy of alihodza on Flickr.
Used under CC-BY-NC.

Launched:	1964
Clock rate:	33 KHz
Data path:	32 bits
Memory:	524 Kbytes
Cost:	$5,000/month

DEC PDP-11
Courtesy of jonrb on Flickr.
Used under CC-BY-NC.

Launched:	1970
Clock rate:	1.25 MHz
Data path:	16 bits
Memory:	56 Kbytes
Cost:	$20,000

Apple II
Courtesy of mwichary on Flickr.
Used under CC-BY.

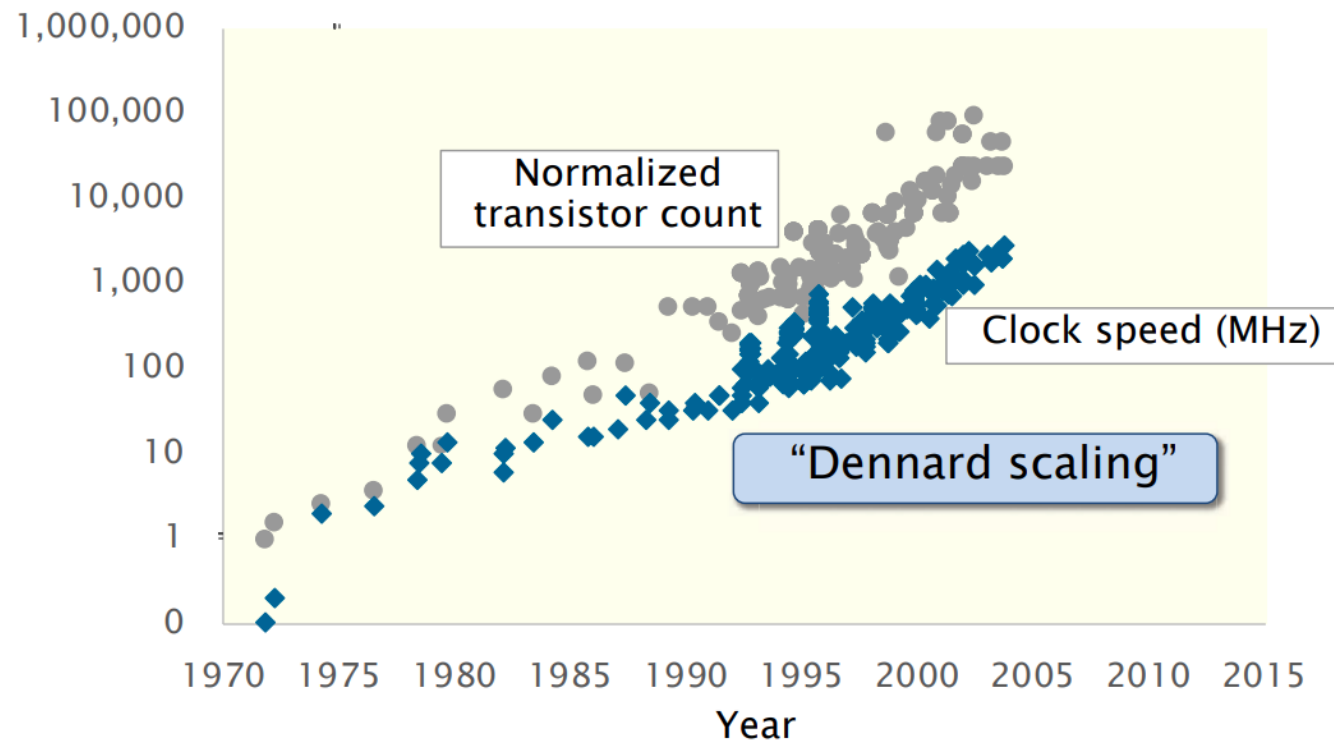|	
Launched:	1977
Clock rate:	1 MHz
Data path:	8 bits
Memory:	48 Kbytes
Cost:	$1,395

History

- Moore´s law: the number of transistors inside a microchip doubles every two years.
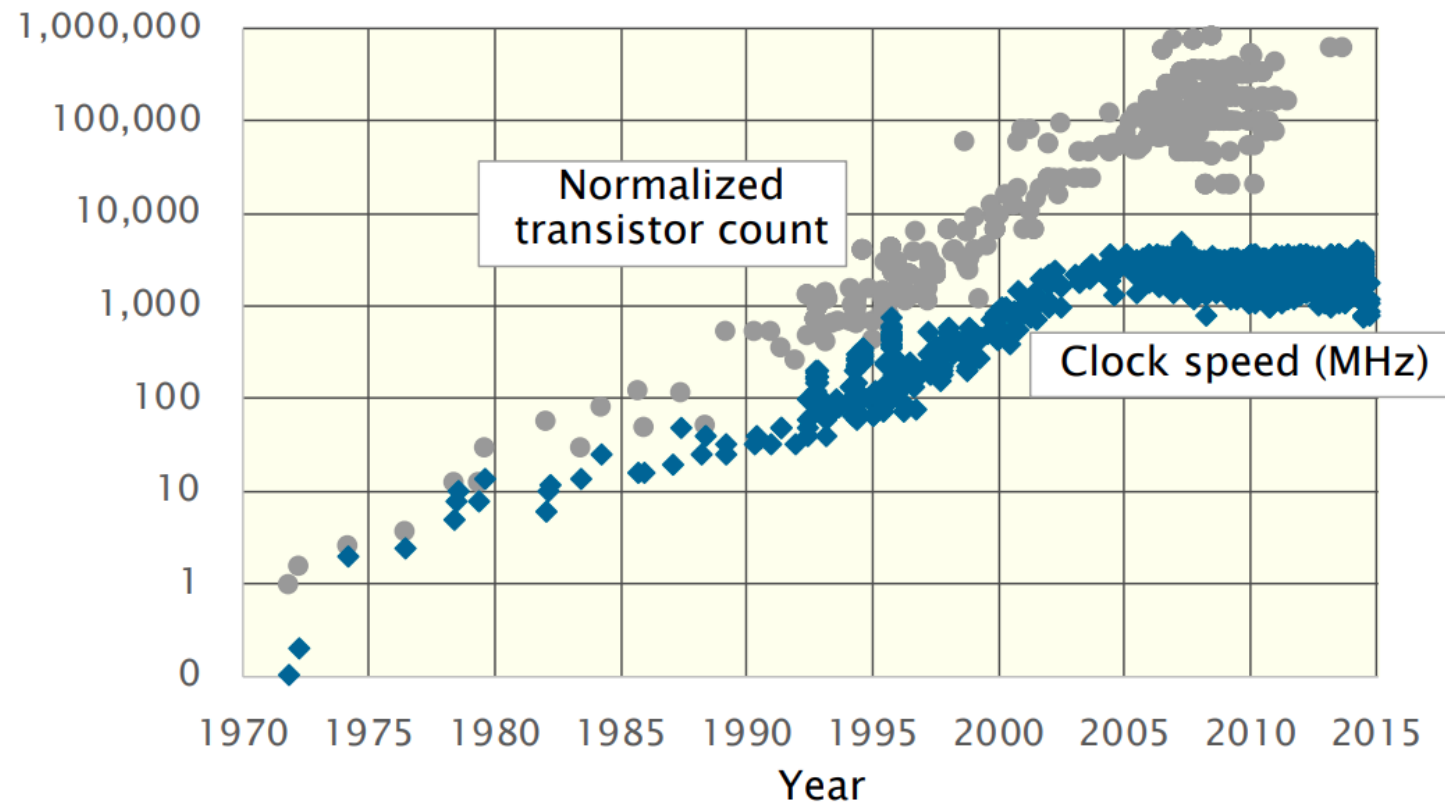- Data up to 2004:
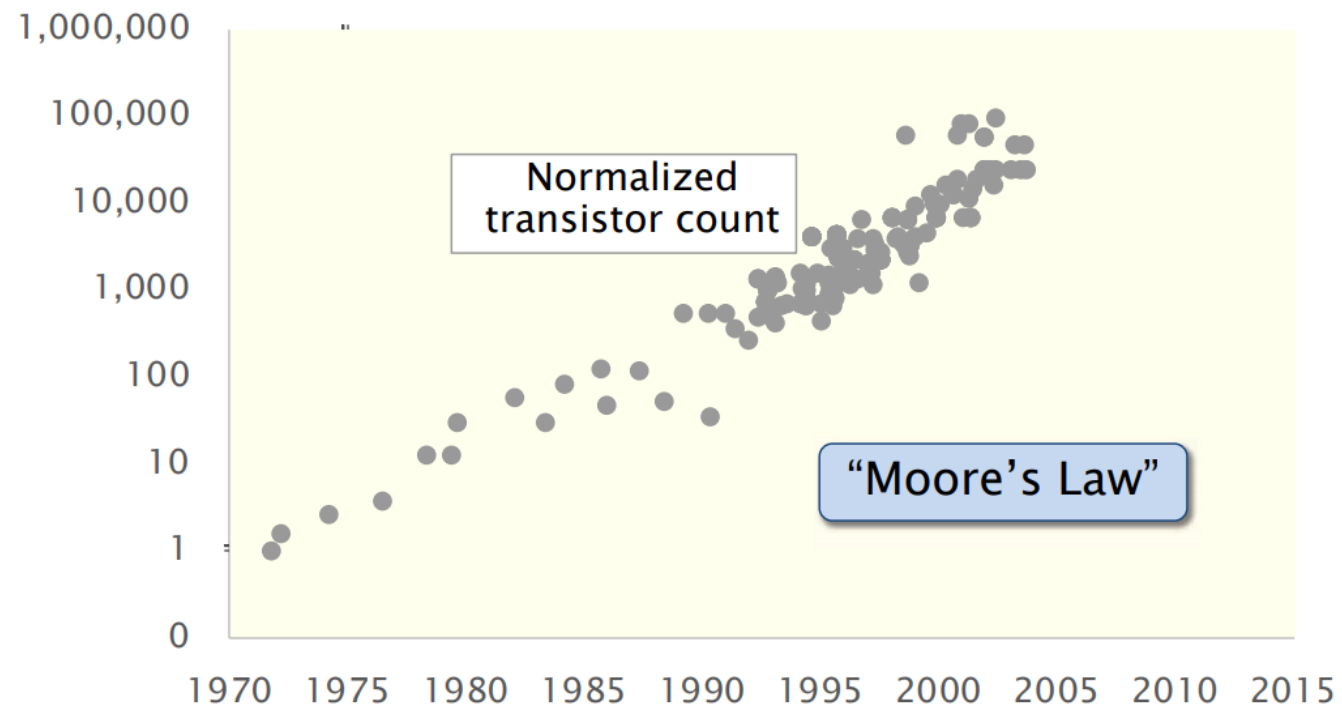
History

- Dennard scaling
- Data up to 2004

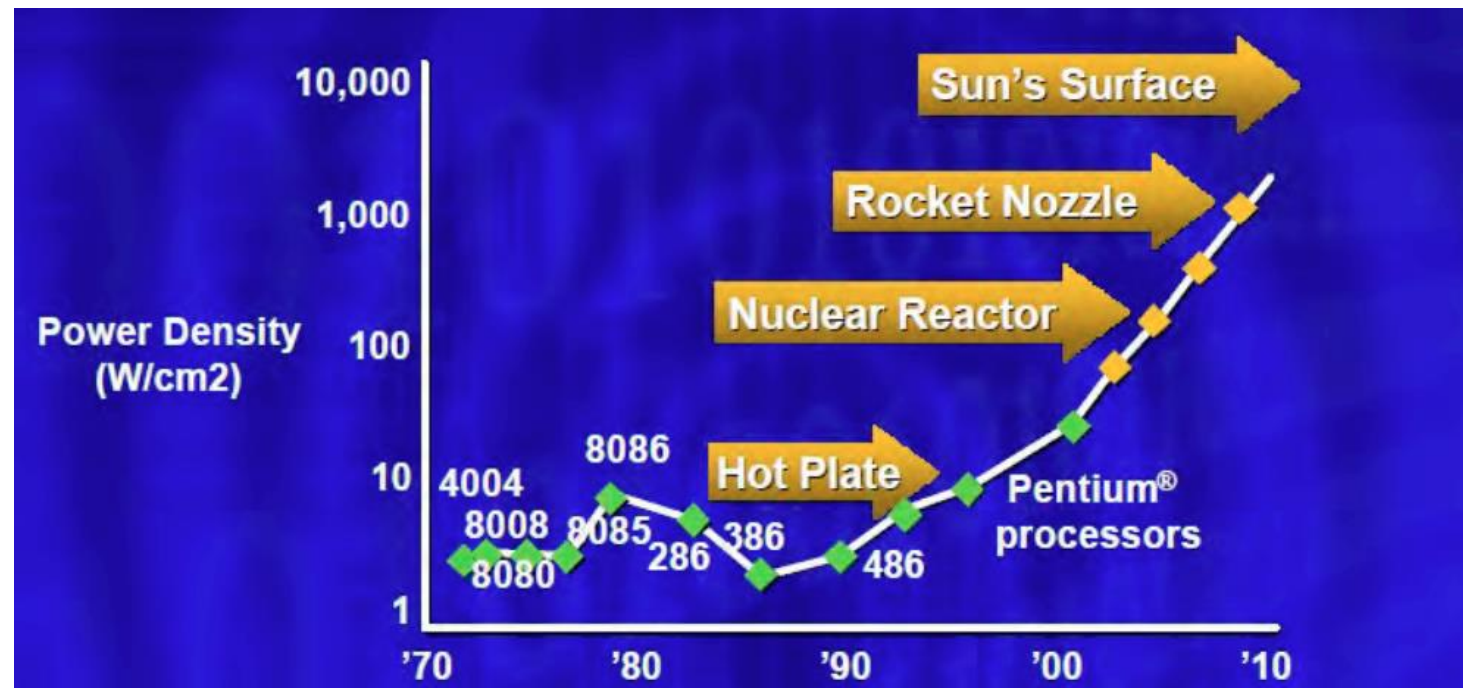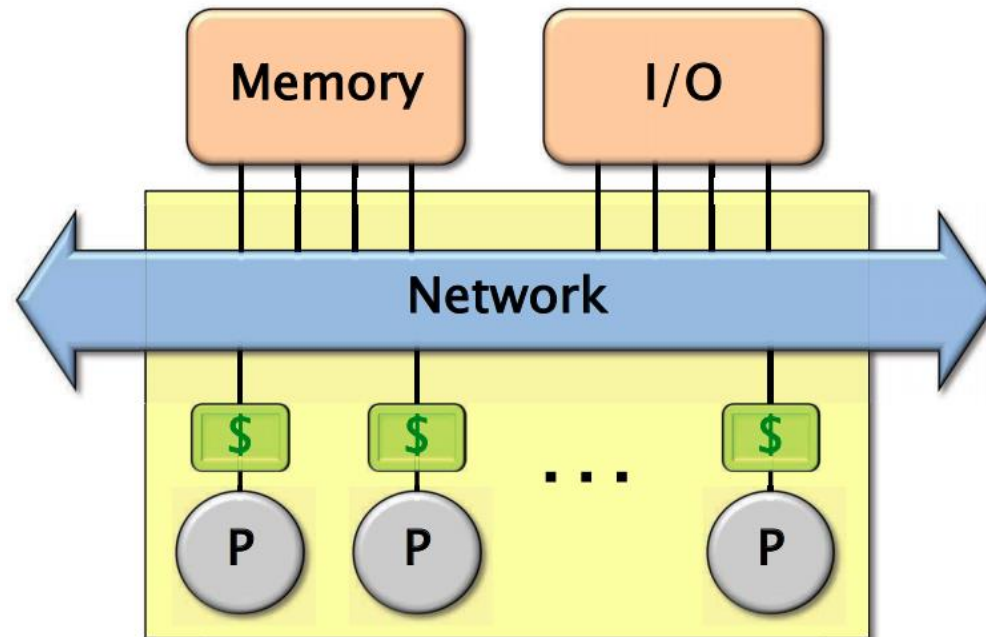# History

- Since 2004:

History

- Power density

Chart courtesy : Pat Gelsinger, Intel Developer Forum, 2004

- Power density increase if the growth in frequency would have kept its incremental trend of 25-30% each year.
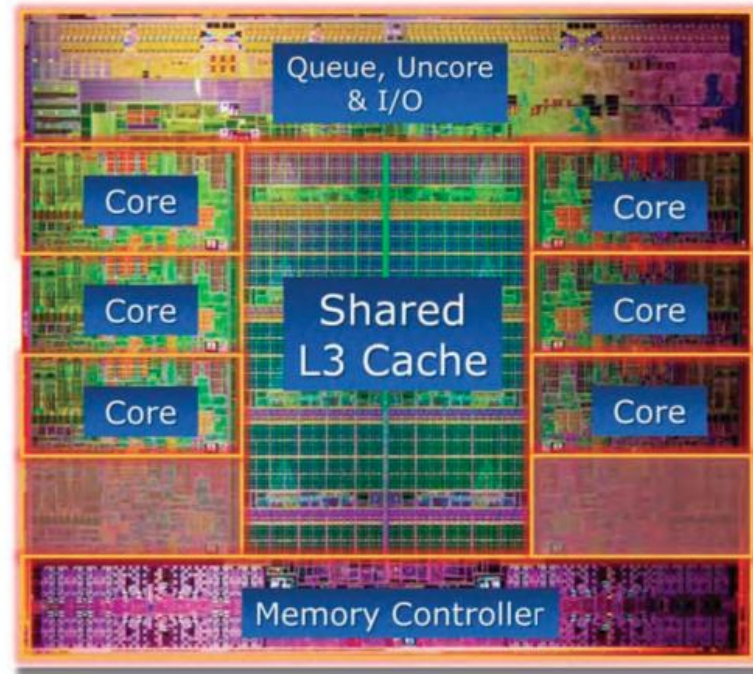
History

- Industry solution: Multi-core

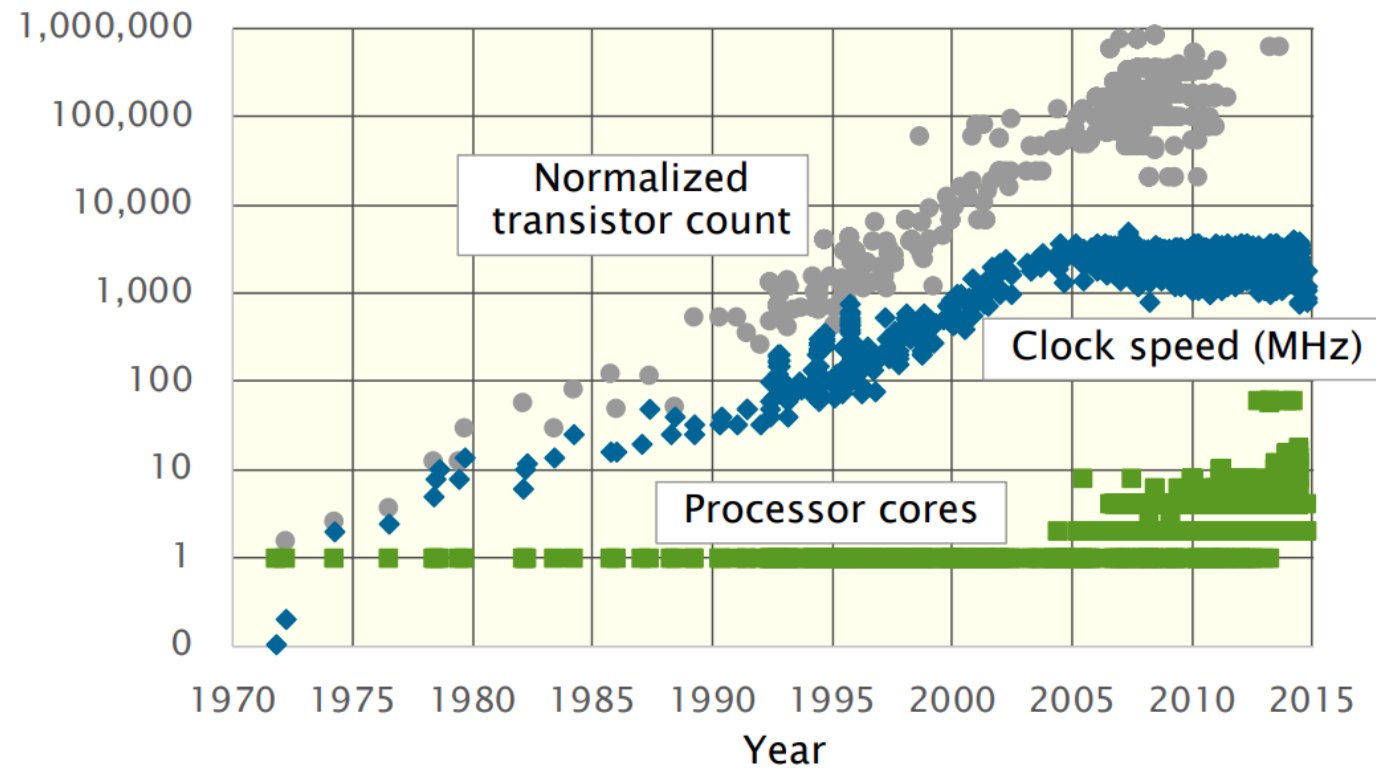# History

- Industry solution: Multi-core

Intel Core i7 3960X (Sandy Bridge), 2011
- 6 cores
- 3.3 GHz
- 15 MB cache L3

- In order to scale up performance, manufacturers added more processing cores inside the processor

History

- Industry solution: Multi-core
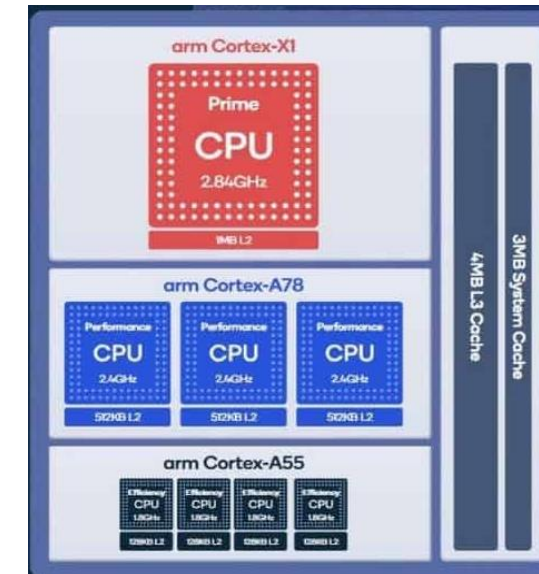
Heterogeneous Computing

- Nowadays Moore's law keeps increasing the computers' performance.
- However, now architectures have small and large multi-core processors, complex memory hierarchy, vector processor units, GPUs, FPGAs, IA units, etc.
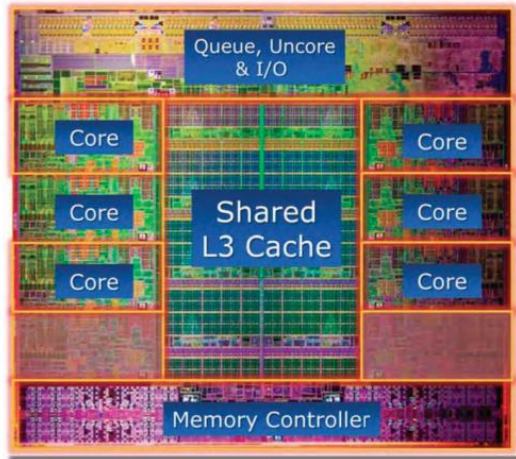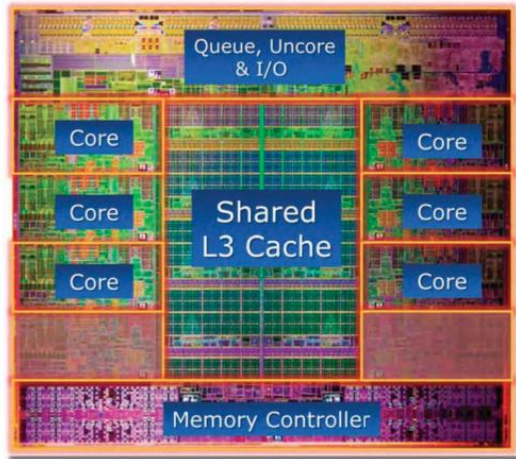
Intel Core i7

GPU Nvidia RTX 3080

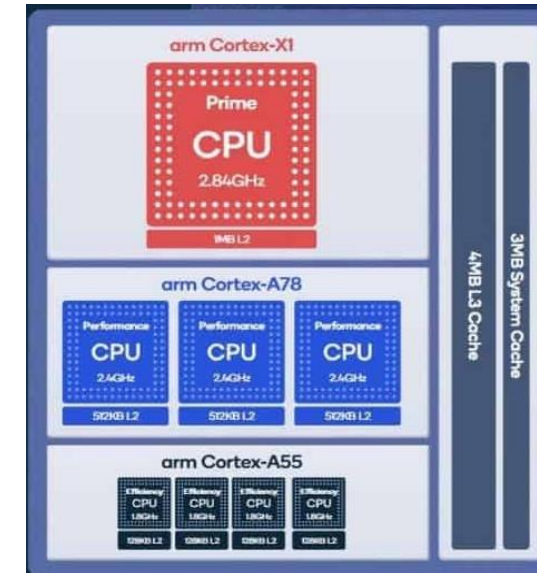Qualcomm Snapdragon 888 (ARM)

Heterogeneous Computing

- Software must be adapted to fully use this hardware efficiently

Intel Core i7

GPU Nvidia GT200

Qualcomm Snapdragon 888 (ARM)

Heterogeneous Computing

- Heterogenous architecture in a smartphone

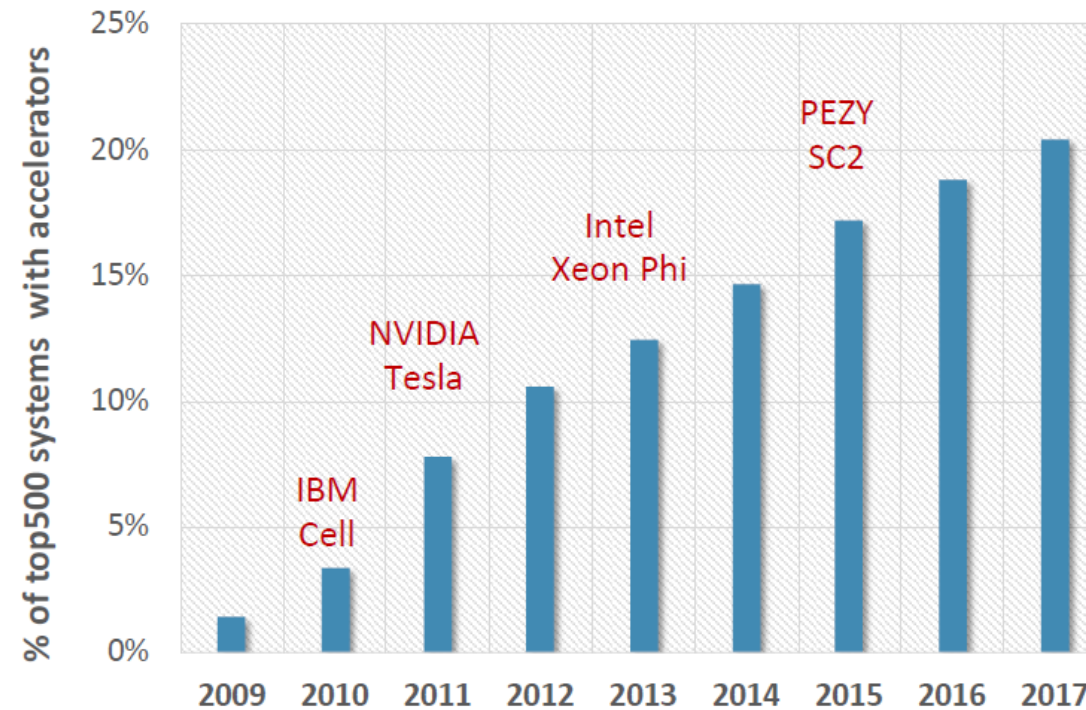8 cores, 2 levels of performance

Heterogeneous Computing

- Heterogenous architecture in a smartphone
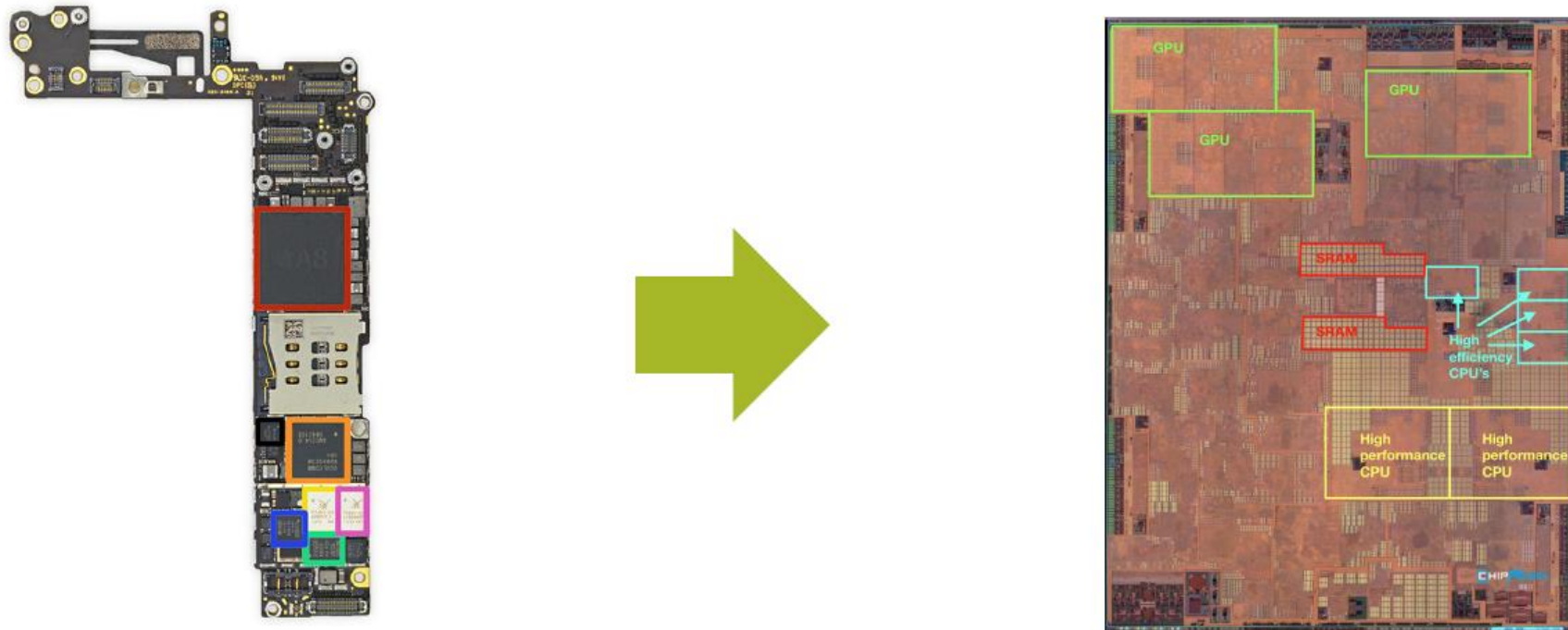
Heterogenous Computing

- Heterogenous accelerators in the world's most powerful systems

In this first unit we'll review:

- Memory hierarchy

- Heterogenous architectures

  - Multi-core CPUs

  - GPUs

  - FPGAs

  - TPUs (tensor processing unit)