Computer Architecture. Week 1

Muhammad Fahim, Alexander Tormasov

Innopolis University

m.fahim@innopolis.ru
a.tormasov@innopolis.ru

September 3, 2020

Topic of the lecture

- Introduction to computer architecture

Topic of the tutorial

- Vertical overview from high-level language to assembler

Topic of the lab

- Journey from high-level language to assembler

Content of the class

- Computer Technology
- Classes of Computers
- What is Computer Architecture?
- Overall Structure
- Arithmetic Logic Unit (ALU)
- Control Unit (CU)
- The Registers
- The Bus
- Main Memory
- Disk Storage
- Anatomy: 5 Components of any computer
- I/O Devices
- Evolution in Computer Architecture Design
- Moore's Law

I think it's fair to say that personal computers have become the most empowering tool we have ever created. They are tools of communication, they are tools of creativity, and they can be shaped by their user.

– Bill Gates, February 24, 2004

Classes of Computers (1/2

- Personal Mobile Device (PMD - smart phones, tablet etc.)
  - Emphasis on energy efficiency and real-time
- Desktop Computing
  - Emphasis on price-performance
- Servers
  - Emphasis on availability, scalability, throughput

Classes of Computers (2/2

- Clusters/Warehouse Scale Computers
  - Used for “Software as a Service (SaaS)”, cloud
  - Emphasis on availability and price-performance
  - Sub-class: Supercomputers, emphasis: floating-point performance and fast internal networks
- Embedded Computers
  - Emphasis: price, robustness (critical mission/space/nuclear/etc.)

What is Computer Architecture?

The science and art of **designing**, selecting, and interconnecting **hardware components** and designing the hardware/software **interface** to create a computing system that meets functional, performance, energy consumption, cost, and other specific goals.

Problem Solution Stack in Modern World

Problem
Algorithm
Data Structure
User Programs
System Programs
Architecture/ISA
Microarchitecture
Circuits
Electrons

ISA: Instruction Set Architecture

Architecture vs Microarchitecture

- Architecture
  - Programmer's view of computer
    - Defined by instructions and operand locations
- Microarchitecture
  - How to implement an architecture in hardware

The Von Neumann Architecture/Model

- Also called stored program computer (instructions in memory). It has two key properties:
- Stored program
  - Instructions stored in a linear memory array (directly adressable!)
  - Memory is unified between instructions and data
    - The interpretation of a stored value depends on the control signals
- Sequential instruction processing
  - One instruction processed (fetched, executed, and completed) at a time
  - Program counter (instruction pointer) identifies the current instruction.
  - Program counter is advanced sequentially except for control transfer instructions

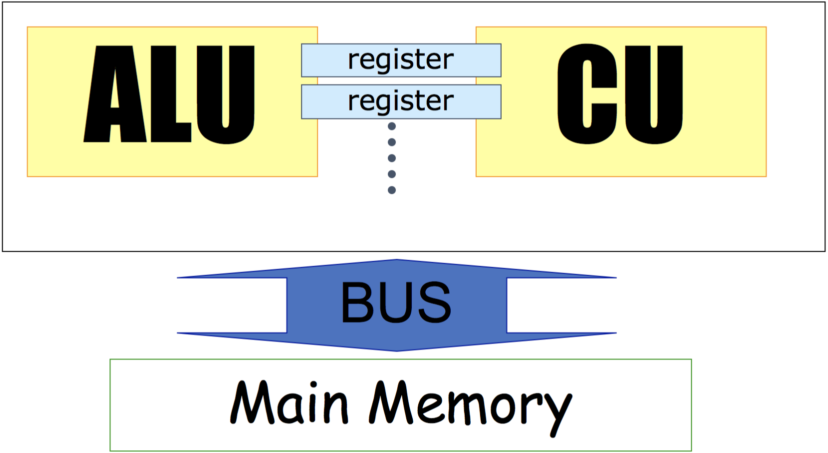
Microarchitecture

- How the underlying implementation actually executes instructions
- Microarchitecture can execute instructions in any order as long as it obeys the semantics specified by the ISA when making the instruction results visible to software
- Programmer should see the order specified by the ISA

Examples

- All major instruction set architectures today use Von-Neumann model (with minor variations, such as separate instruction and data caches, and multi-core clusters).
  - For example: x86, ARM, MIPS, SPARC, Alpha, POWER, ...
- Underneath (at the microarchitecture level), the execution model of almost all implementations (or, microarchitectures) is very different
  - Pipelined instruction execution: Intel 80486
  - Multiple instructions at a time: Intel Pentium
  - Out-of-order execution: Intel Pentium Pro
  - Separate instruction and data caches
- But, what happens underneath that is not consistent with the von Neumann model is not exposed to software

Overall Structure



The Control Unit

- It is the core of the sequencing of operations
- Picks the new operation to be executed
- Decodes it
- Coordinates its execution

The Arithmetic and Logic Unit (ALU)

- Is the core of the “computation”
- Performs arithmetic, logic and shift operations
- All other operations are combinations of these basic operations
- Works on numbers in base 2, usually (more detail in next lecture)

The Registers

- Are the places where we put the data we need for the actual execution
- Limited in size and very fast to access

The Bus

- The interconnection between the different pieces
- There are different kinds, supporting different speed and sizes

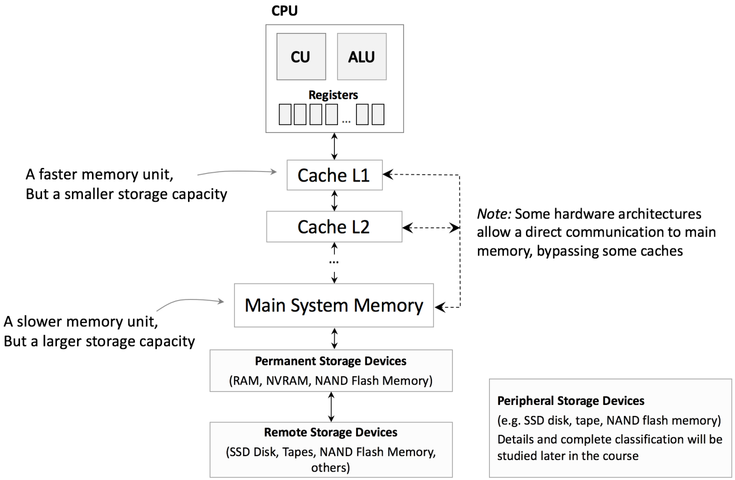
Main Memory

- Addressed directly – sometimes said “randomly”, hence RAM
- Fast access, no as fast as register, but still fast
- Volatile structure

Disk Storage

- Slower to access
- Sequential in accessing nature (while randomly accessible)
- Larger capacity
- Permanent storage

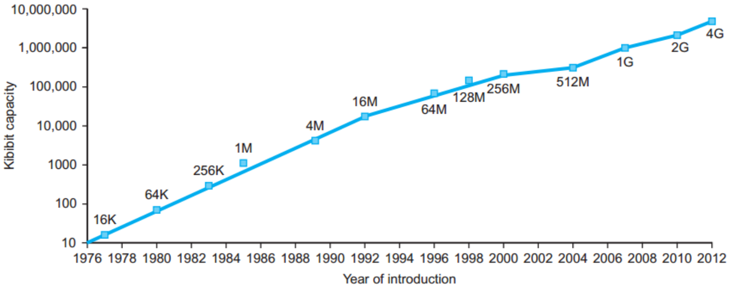
Anatomy of Computer



Input Devices

Output Devices

Memory Capacity Single-Chip DRAM



- The y-axis is measured in kibibits (2^{10} bits).
- The DRAM industry quadrupled capacity almost every three years
- In recent years, the rate has slowed down and is somewhat closer to doubling every two years to three years

Computer Technology: Dramatic Change!

State-of-the-art PC:

- Processor clock speed: 10,000 MHz (10.0 GHz)
- 100X performance in last decade.
- Memory capacity: 10,000 MB (10.0 GB)
- Disk capacity: 20,000 GB (20.0 TB)
- New units! Mega =>Giga, Giga =>Tera =>Peta =>Exa =>Zeta =>Yotta

  NOTE: 10GHz clock speed is not practical due to heat dissipation problems. On modern computers the clock rate achieved over 3.5Ghz and later was scaled back due to aforementioned problems. The current direction of improving performance is different from increasing clock rate: adding more parallelism to microarchitecture, adding hardware support for multithreading, using multi-core clusters, and specialized hardware, optimized for certain tasks (like GPU for 3D graphics)

Computer Technology: Dramatic Change!

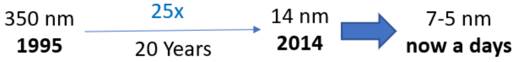
- Smartphone (iPhone 11 Pro):
  - Processor: Hexa-core and GPU (4-core graphics)
  - Internal Memory capacity: 3 GB RAM
  - Storage capacity: Up to 512 GB
- Notebook (Dell XPS'17):
  - Processor: 1.8GHz Intel Core i7 (10th Generation)
  - Memory capacity: 32GB
  - Disk capacity: 1TB SSD
- Server (Intel Xeon Server)
  - Processor: 2.2 GHz Xeon Processor E7
    (60 MB Cache, 24 Cores, 48 Threads)
  - Max. Memory capacity: 3TB
  - Disk capacity: 12 TB

NOTE: The more cores a processor has, the more sets of instructions the processor can receive and process at the same time, which makes the computer faster.

Computer Technology: Dramatic Change!

- Processor
  - 2X in speed every 1.5 years (since 85)
  - 100X performance in last decade.
- Memory
  - DRAM capacity: 2x / 2 years (since 96)
  - 64x size in last decade.
- Disk
  - Capacity: 2X / 1 year (since 97)
  - 250X size in last decade.

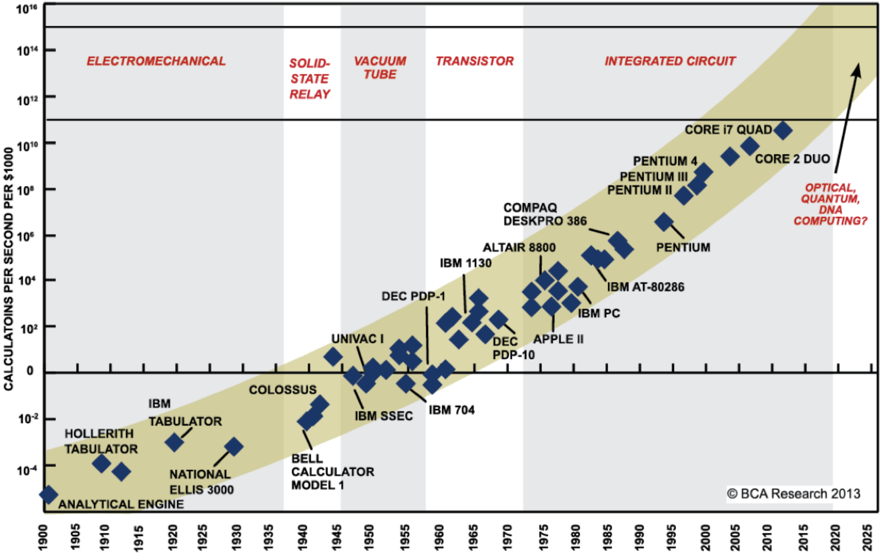
Processor



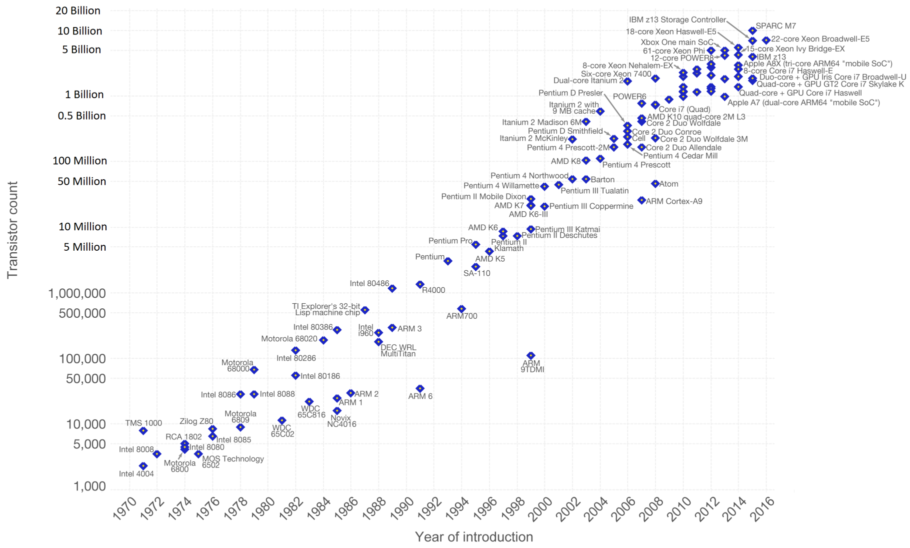
- Real performance rising $<$ 25x
  - Multicore system use 25-50% of peak performance
  - Why 8-cores smartphone is better than 4-cores one?
- **Window of opportunity:** customization and new generation of EDA
  - SoC customization for the application allows to achieve a best performance with minimal cost and power consumption
  - New system level design technologies and EDA tools allow to achieve the best time to market in the industry

NOTE: EDA is Electronic design automation
SoC: System on a chip

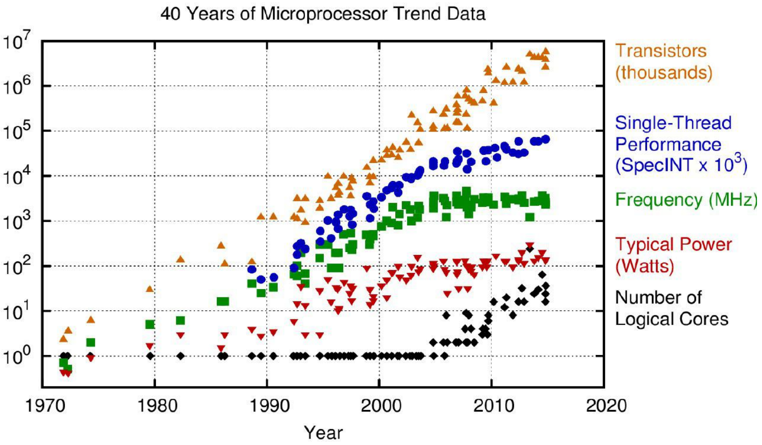
Microprocessor Complexity (Moore's Law)



No. of transistors on IC chips (Moore's Law)



Processor Performance



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

Processor Performance

In the next lecture we will learn:

- Performance evaluations is a complex issue
- We cannot judge performance by clock frequency or average number of instructions per cycle alone
- Processors with the same frequency can have different performances.
- For fair comparison a multitude of parameters should be taken into account (including clock rate, throughput, memory performance, etc.)
- Ideally, one would compare performance running his own application on different computers. However it is not practical, so the industry uses instead synthetic benchmarks that mimic the behaviour of typical user programs

Summary

- Computer architecture: Programmer's view of computer
- Microarchitecture: How to implement an architecture in hardware
- The Von Neuman model: Describes a design architecture for an electronic digital computer
- Computer structure
- Evolution in technology

Acknowledgements

- This lecture was created and maintained by Muhammad Fahim, Giancarlo Succi and Alexander Tormasov