计算机系统结构

Computer Systems and Architecture

李峰
fli@sdu.edu.cn
https://funglee.github.io

Short Bio

- Feng Li (李峰)
- Education
  - 2010 – 2015, Ph.D., Nanyang Technological University, Singapore.
  - 2007 – 2010, M.S., Shandong University, China.
  - 2003 – 2007, B.S., Shandong Normal University, China.
- Employment
  - 2015.11 – Now, Assistant Professor, Shandong University, China.
  - 2014.12 – 2015.11, Research Fellow, National University of Singapore, Singapore
- Research Interests
  - Applied Optimization (应用优化)
  - Wireless Networking (无线网络)
  - Mobile Sensing and Computing (移动感知及计算)
  - Distributed Algorithms and Systems (分布式算法及系统)

课程介绍

1. 课程名称

- Computer Architecture
- 计算机系统结构，计算机体系结构
- 建筑物的设计或式样，通常指一个系统的外貌

2. 研究内容

- 从外部来研究计算机系统
- 使用者所看到的物理计算机的抽象
- 编写出能够在机器上正确运行的程序所必须了解到的计算机属性

3. 学习目的

- 建立计算机系统的完整概念
- 学习计算机系统的分析方法和设计方法
- 了解计算机系统的最新研究成果

4. 课程安排

- 课内：48学时；实验：16学时
- 成绩 = 考试成绩*80% + 平时成绩（包括实验及作业）*20%

5. 参考书：

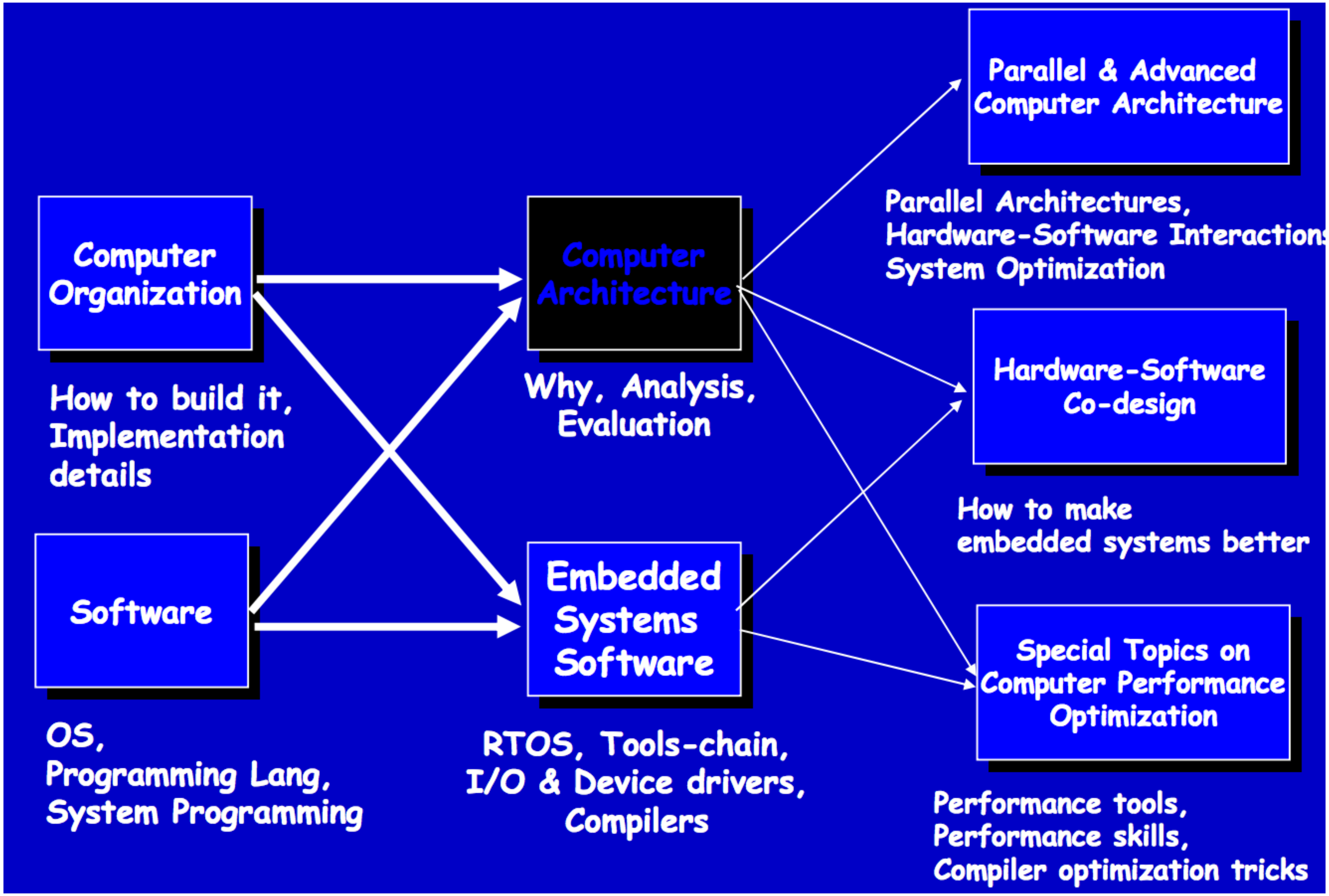
1) David A. Patterson and John L. Hennessy, Computer Architecture: A Quantitative Approach
2) David A. Patterson and John L. Hennessy, Computer Organization and Design: The Hardware/Software Interface
3) 郑纬民、汤志忠，计算机系统结构

6. 课程网站：https://funglee.github.io/csa/csa.html

7.课程大纲

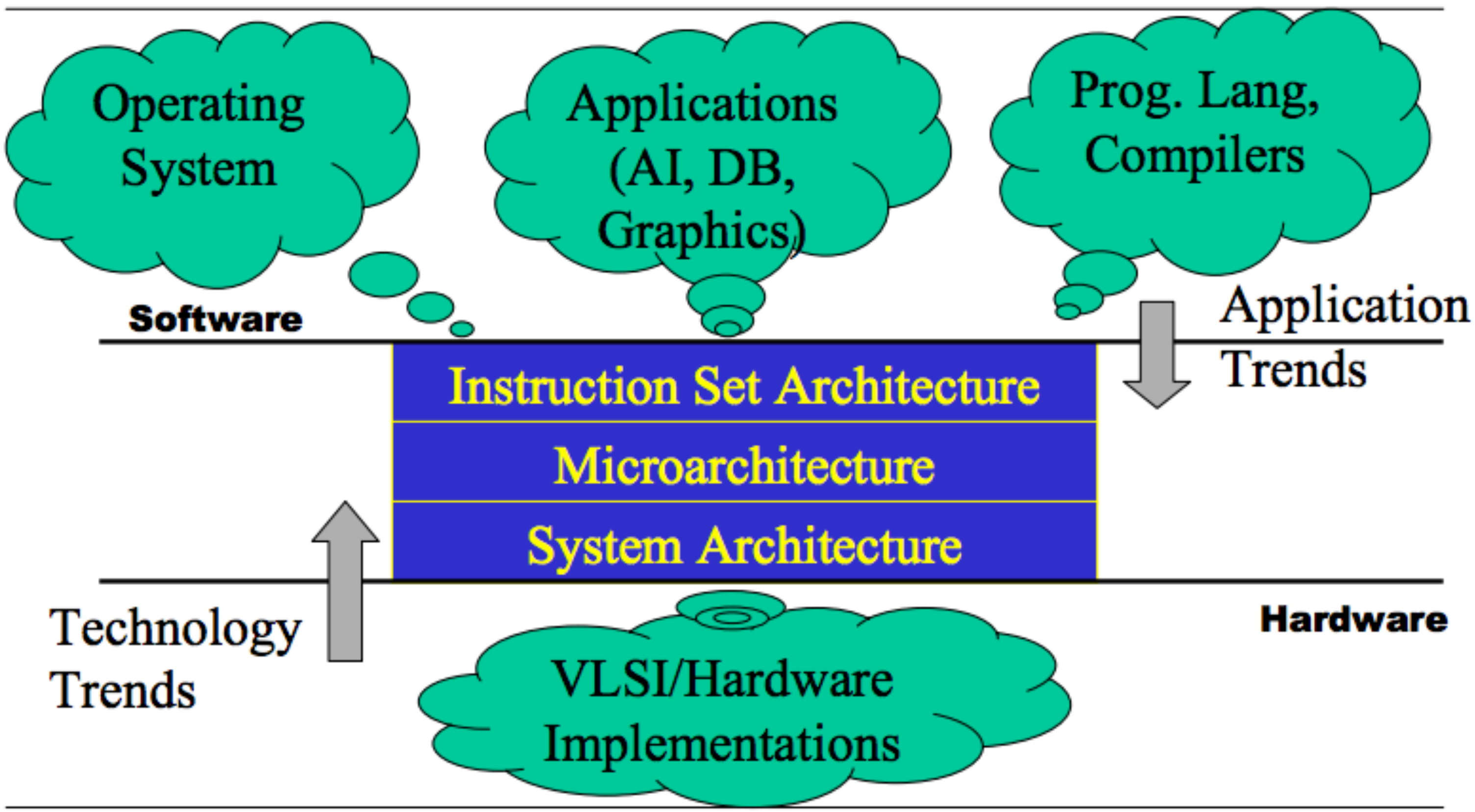
- 基本概念（Basic Concepts）
- 指令系统（Instruction Systems）
- 存储体系（Memory Architecture）
- I/O系统（I/0 Systems）
- 标量处理机（Scalar Processors）
- 向量处理机（Vector Supercomputer）
- 多处理机系统（Multiprocessor Systems）

8. 相关课程



Computer Organization
How to build it, Implementation details
Software
OS, Programming Lang, System Programming
Computer Architecture
Why, Analysis, Evaluation
Embedded Systems Software
RTOS, Tools-chain, I/O & Device drivers, Compilers
Parallel & Advanced Computer Architecture
Parallel Architectures, Hardware-Software Interactions System Optimization
Hardware-Software Co-design
How to make embedded systems better
Special Topics on Computer Performance Optimization
Performance tools, Performance skills, Compiler optimization tricks

计算机体系结构是什么？

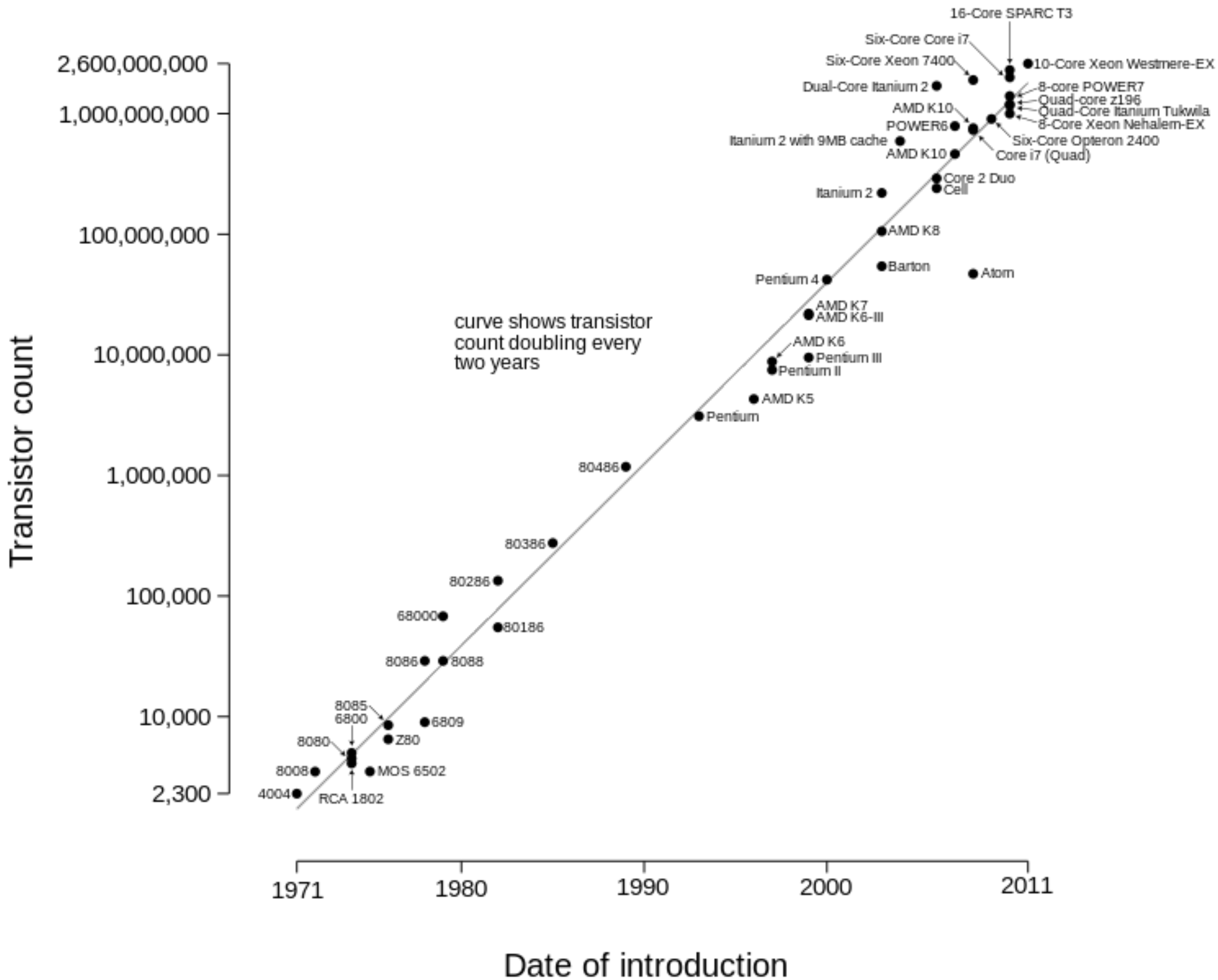


Operating System
Applications (AI, DB, Graphics)
Prog. Lang, Compilers
Software
Application Trends
Instruction Set Architecture
Microarchitecture
System Architecture
Technology Trends
Hardware
VLSI/Hardware Implementations

摩尔定律（Moore’s law）

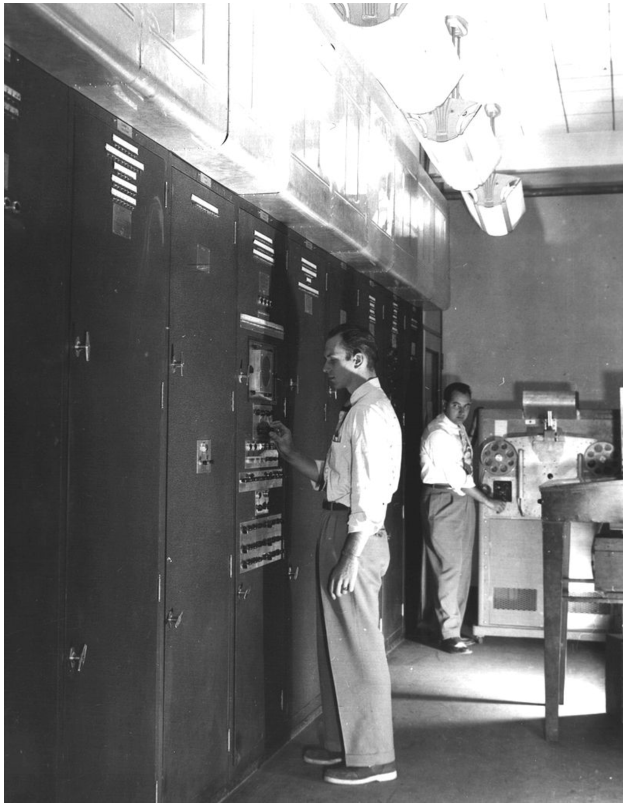
- The number of transistors in a dense integrated circuit doubles approximately every two years (by Gordon E. Moore, 1975)
- The capabilities of many digital electronic devices are strongly related to Moore’s law: processing speed, memory capacity, sensors and even the number and size of pixels in digital cameras

Microprocessor Transistor Counts 1971-2011 & Moore's Law



1950s: Early Designs

- CPUs were customized and used as part of a larger computer
  - Uniqueness
  - Poor compatibility



1960s: The Computer Evolution and CISC

- Price and Performance is the main concern
- A family of computers which can run the same software but with different performance
- CISC (Complex Instruction Set Computing) is a processor design where single instructions can execute several low-level operations or are capable of multi-step operations or addressing modes within single instructions
- Examples: PDP-11, Motorola 68000

1970s: Large-Scale Integration

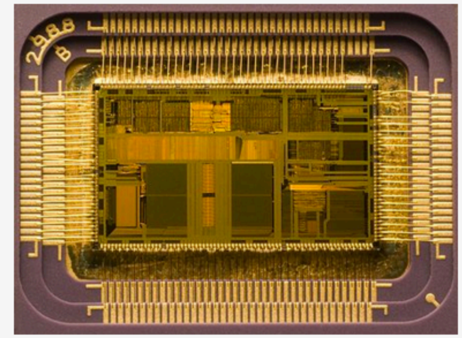
- Intel 4004 in 1970
  - 740 kHz, 4-bit BCD-oriented, 10um, 2300 transistors
- Intel 8008 in 1972
  - 0.2-0.8 MHz, 8-bit, 10um, 3500 transistors
- MOS Technology 6502 in 1975
  - 1-2 MHz, 8-bit,
- VAX-11/780 in 1977
  - 5 MHz, 32-bit,
- Intel 8086 and Intel 8088 in 1978 (the first x86 chips)
  - 5 -10MHz, 16-bit, 3um, >20000 transistors

Early 1980s: the lesson of RISC

- However, only a small set of these complex instructions were used frequently by most computer language compilers and interpreters.
- There, RISC (Reduced Instruction Set Computing) was emerged in the early 1980s, where a simplified instruction set is adopted.
    - Examples: DEC Alpha, AMD Am29000, ARC, ARM, Atmel AVR, Intel i860 and i960, MIPS, Motorola 88000, Power, …

Mid-to-late 1980s: Exploiting Instruction Level Parallelism

- Instruction pipelining is a technique that implements a form of parallelism called instruction-level parallelism within a single processor. It results in much faster CPU throughput
- Unfortunately, this architecture increases hardware complexity, resulting in higher cost, larger circuits, higher power consumption.
- Examples: Intel 80386, 80486



The exposed die of an Intel 80486DX2 microprocessor

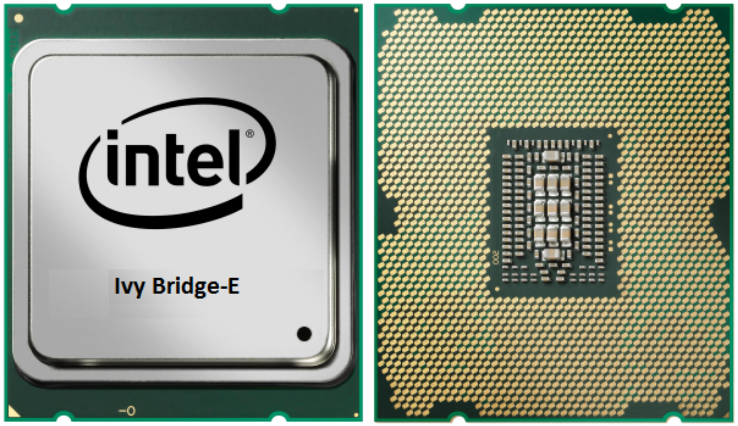
Produced	From 1989 to 2007
Common manufacturer(s)	Intel, IBM, AMD, Texas Instruments, Harris Semiconductor, UMC, SGS Thomson
Max. CPU clock rate	16 MHz to 150 MHz
FSB speeds	16 MHz to 50 MHz
Min. feature size	1μm to 0.6μm

1990 to today

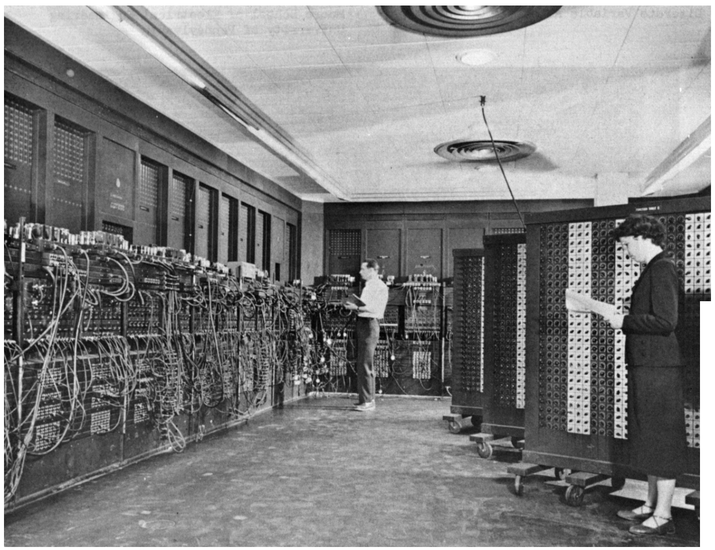
- VLIM (Very Long Instruction Word)
  - Scheduling instructions is performed by compilers and programs rather than hardware.
  - Complicated compiler but simpler hardware
- EPIC (Explicitly Parallel Instruction Computing)
  - Highly improve VLIM by increasing instruction throughput
  - Indicate the dependency between the “bundles” of instructions
- Multi-threading
- Multi-core
- Open source processors
- Asynchronous CPUs (clockless CPU)
- Optical processors

Ivy Bridge

- It is the codename for a line of processors based on the 22nm manufacturing process developed by Intel in 2011
- Examples: 15-core Xeon Ivey Bridge-EX, 4-core Core i7 Ivy
- Around 4,310,000,000 transistors
- 22 nm Tri-gate transistor



History of Computing and Computers



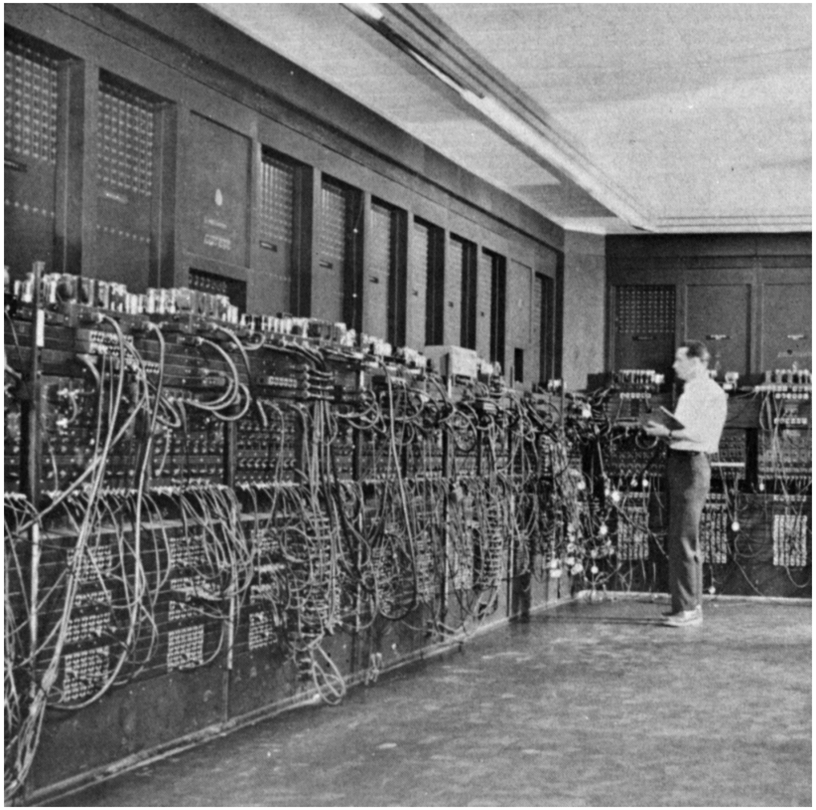
1946

2015

The First Electronic General-Purpose Computer

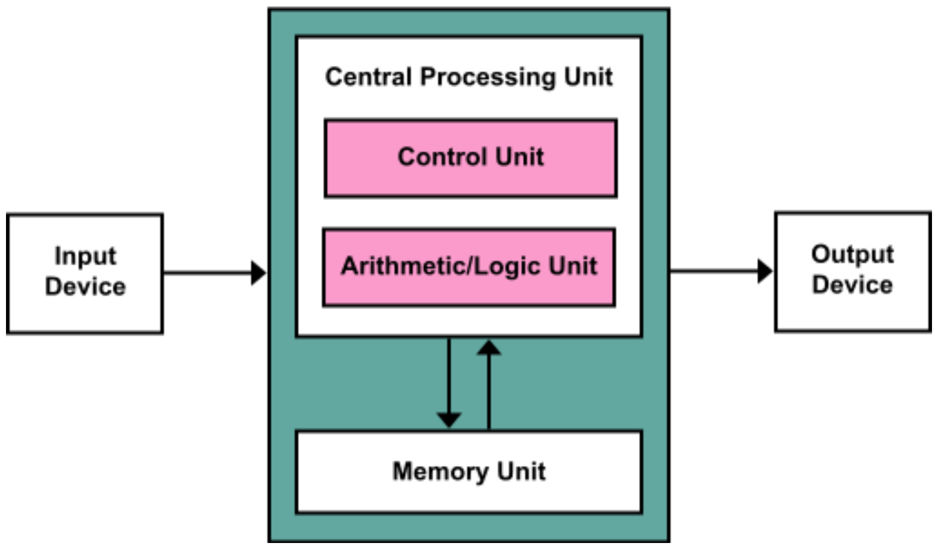
ENIAC (Electronic Numerical Integrator and Computer)

- Designed and built by Eckert and Mauchly at the University of Pennsylvania during 1943-45.
- It was Turing-complete, digital, and could solve “a large class of numerical problems” through reprogramming.
    - 30 tons, 72 square meters, 200KW
    - Read in 120 cards per minute
    - Addition took 20us, division 6ms
    - 1000 times faster than electro-mechanical machines
- Applications: ballistic calculations



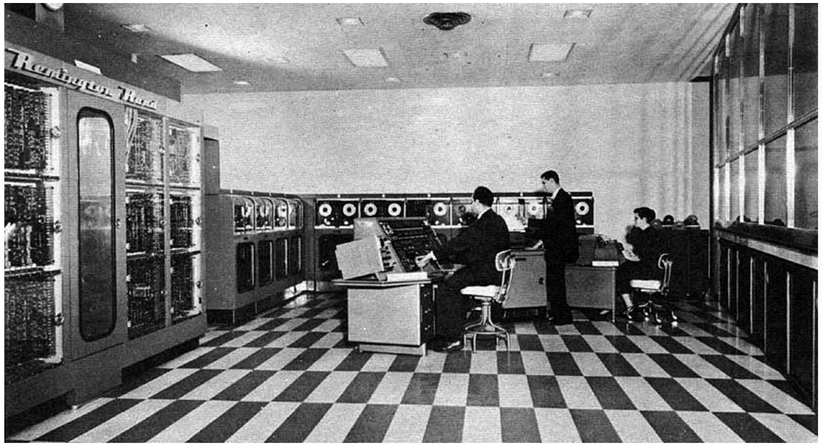
Von Neumann Architecture

- Proposed by John von Neumann in 1945



The First Commercial Computer

- UNIVAC I (UNIVersal Automatic Computer I) designed by J. Presper Eckert and John Mauchly in 1951
  - 5200 vacuum tubes
  - 29000 pounds (13 tons)
  - 125W
  - 1905 operations per second
  - 2.25 MHz clock
  - $4.3m \times 2.4m \times 2.6m$



More …

- More computer firsts:
  - 1947 : The invention of the transistor;
  - 1958 : The invention of the integrated circuit;
  - 1971 : A young company called Intel produced the very first microprocessor:
    - Intel 4004 and contained around 2300 transistors on a single chip;
    - Chip technology is now so advanced that we are close to having one billion transistors on a single chip.
  - 1981 : The first Personal Computer.
- Now:
  - Desktop, laptop, supercomputer, smart phones, smart watches, smart glass,
  - … …

The TOP500 project ranks and details the 5000 most powerful (non-distributed) computer systems in the world.

Top 10 positions of the 46th TOP500 in November 2015

Rank	Rmax Rpeak (PFLOPS)	Name	Computer design Processor type, interconnect	Vendor	Site Country, year	Operating system
1	33.863 54.902	*Tianhe-2*	**NUDT** Xeon E5–2692 + Xeon Phi 31S1P, TH Express-2	NUDT	National Supercomputing Center in Guangzhou China, 2013	Linux (Kylin)
2	17.590 27.113	*Titan*	**Cray XK7** Opteron 6274 + Tesla K20X, Cray Gemini Interconnect	Cray Inc.	Oak Ridge National Laboratory United States, 2012	Linux (CLE, SLES based)
3	17.173 20.133	*Sequoia*	**Blue Gene/Q** PowerPC A2, Custom	IBM	Lawrence Livermore National Laboratory United States, 2013	Linux (RHEL and CNK)
4	10.510 11.280	*K computer*	**RIKEN** SPARC64 VIIIfx, Tofu	Fujitsu	RIKEN Japan, 2011	Linux
5	8.586 10.066	*Mira*	**Blue Gene/Q** PowerPC A2, Custom	IBM	Argonne National Laboratory United States, 2013	Linux (RHEL and CNK)
6	8.101 11.079	Trinity	**Cray XC40** Xeon E5-2698v3, Cray Aries Interconnect	Cray Inc.	DOE/NNSA/LANL/SNL United States, 2015	Linux (CLE)
7	6.271 7.779	*Piz Daint*	**Cray XC30** Xeon E5–2670 + Tesla K20X, Aries	Cray Inc.	Swiss National Supercomputing Centre Switzerland, 2013	Linux (CLE)
8	5.640 7.404	Hazel Hen	**Cray XC40** Xeon E5-2680v3, Cray Aries Interconnect	Cray Inc.	HLRS - Höchstleistungsrechenzentrum, Stuttgart Germany, 2015	Linux (CLE)
9	5.537 7.235	*Shaheen II*	**Cray XC40** Xeon E5–2698v3, Aries	Cray Inc.	King Abdullah University of Science and Technology Saudi Arabia, 2015	Linux (CLE)
10	5.168 8.520	*Stampede*	**PowerEdge C8220** Xeon E5–2680 + Xeon Phi, Infiniband	Dell	Texas Advanced Computing Center United States, 2013	Linux (CentOS)[13]

FLOPS (Floating-point operations per second), PFLOPS = 10^{15} FLOPS

Tianhe-2

- 16000 computer nodes, each comprising two Intel Ivy Bridge Xeon processors and three Xeon Phi coprocessor chips.
- Each node is equipped with memory of 88 GiB
- HD array 12.4 PiB
- Price: 3.9 million US$
- Applications: scientific computing

Cloud Computing

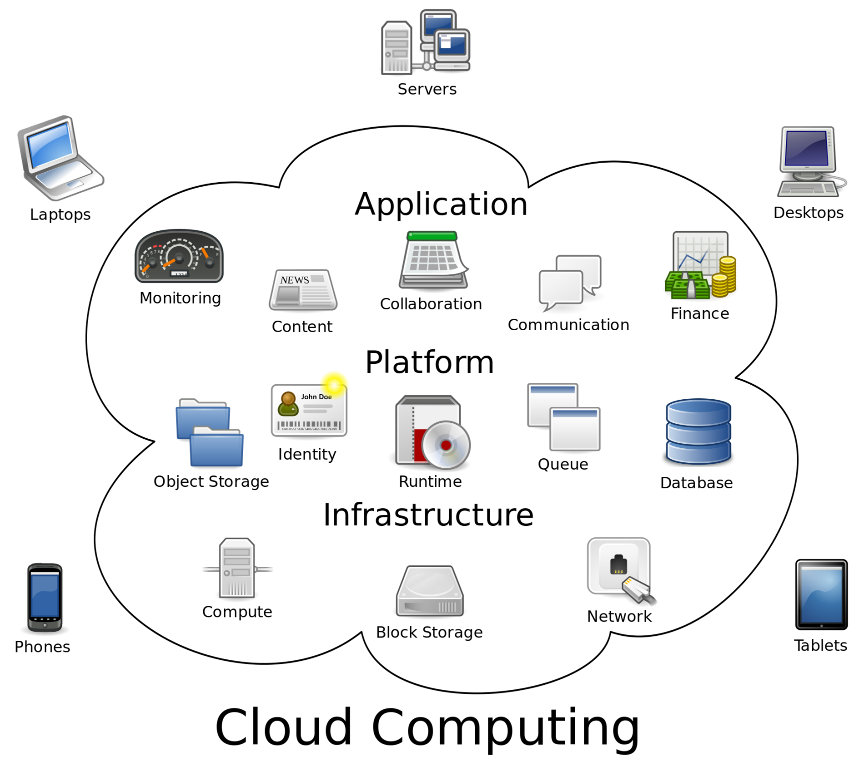
- Everyone is talking about Cloud Computing, but what is it?
  - Computing service is managed, scheduled, and delivered to users over Internet.
  - For example
    - Google Drive
    - One Drive
    - Hotmail
    - Gmail

Characteristics

- On demand self-service
- Access to networks anywhere, anytime, on any devices
- Location independent resource pooling
- Deployment flexibility
- Pay as you go

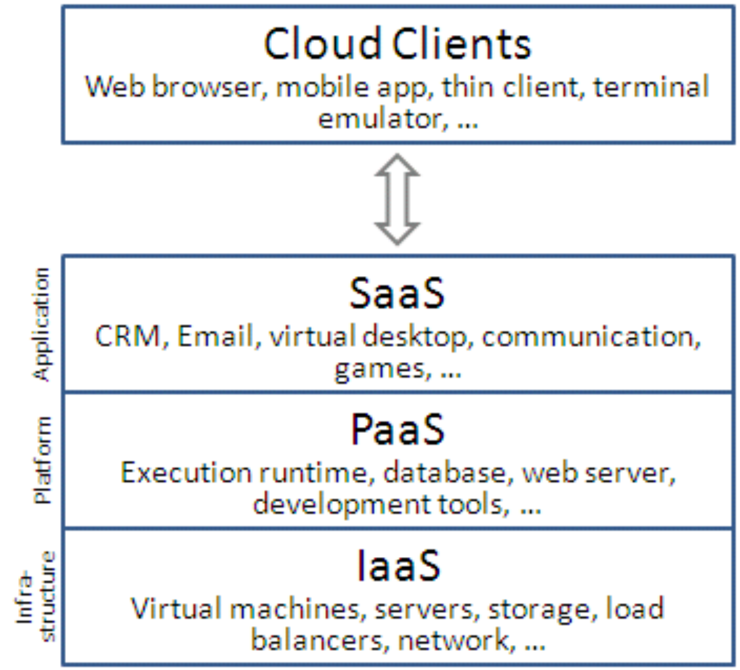
Infrastructures for Cloud Computing

- Development of computing capability
- Virtualization technology
- Distributed Storage
- Fast internet access



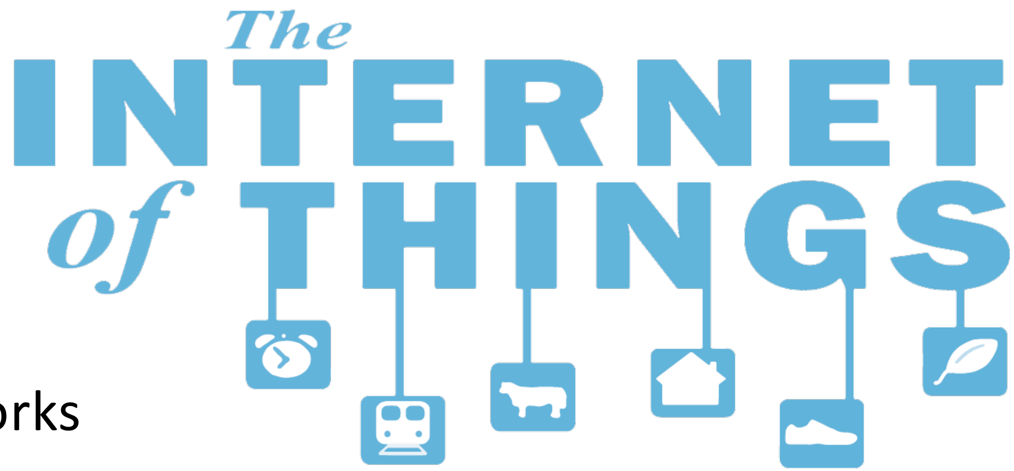
Services of Cloud Computing

- SaaS: Software as a Service
  - Gmail, Hotmail, Flickr, OfficeLive
- PaaS: Platform as a Service
  - Amazon EC2, Microsoft Azure
- IaaS: Infrastructure as a Service
  - AT&T Hosting and Storage
  - Amazon EC2



Internet of Things (IoT)

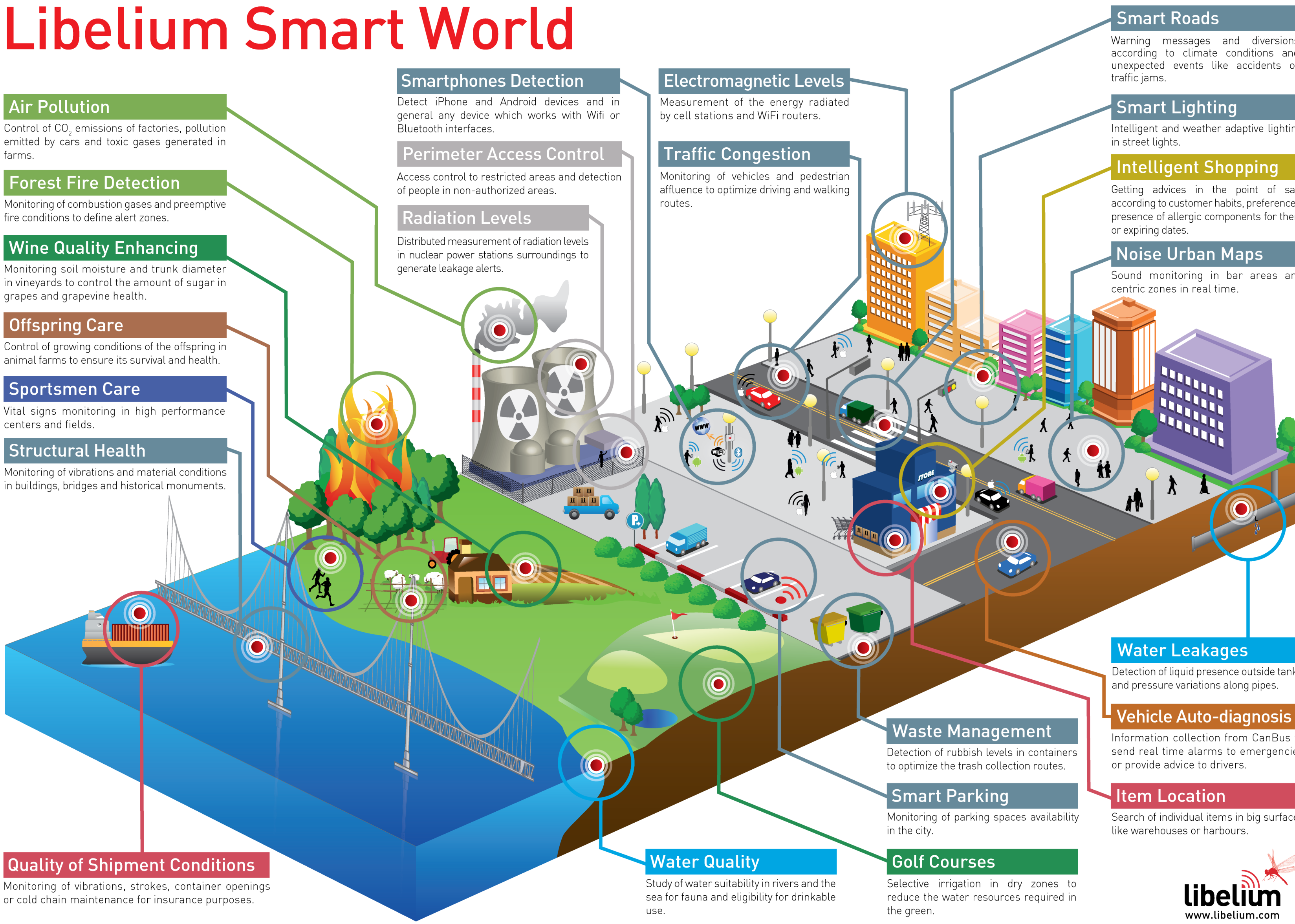
- Smart + X
  - Smart City
  - Smart Traffic
  - Smart Building
  - Smart …
- Wireless Sensor Networks
  - Sensor Motes
  - Mobile Phones
- RFID Systems



IoT Systems include…

- Sensors
  - We look at the world through sensors, e.g., light sensors, cameras, microphones, motion sensors, accelerators, gyroscopes, magnetic sensors, barometers, GPS.
- Networks and communications
  - The sensed data are transmitted, stored and processed in a networked fashion, e.g., WAN, MAN, LAN, PAN.
  - Various communication techniques are combined in the systems, e.g., 3G, 4G, Bluetooth, WiFi, ZigBee, RFID.
- Applications, people and processes
  - All the data are fed back to applications, people and processes for further process and analysis, and finally are sued to make better decision, e.g., remote monitoring, mobile apps, security, supply chain management, locating and tracking, control and automation.

Libelium Smart World



Thanks!