

# Computer Architecture

## 01. Introduction

Jianhua Li (李建华)

College of Computer and Information  
Hefei University of Technology

slides are adapted from CA course of wisc, princeton, mit, berkeley, etc.

**The uses of the slides of this course are for educational purposes only and should be used only in conjunction with the textbook. Derivatives of the slides must acknowledge the copyright notices of this and the originals.**



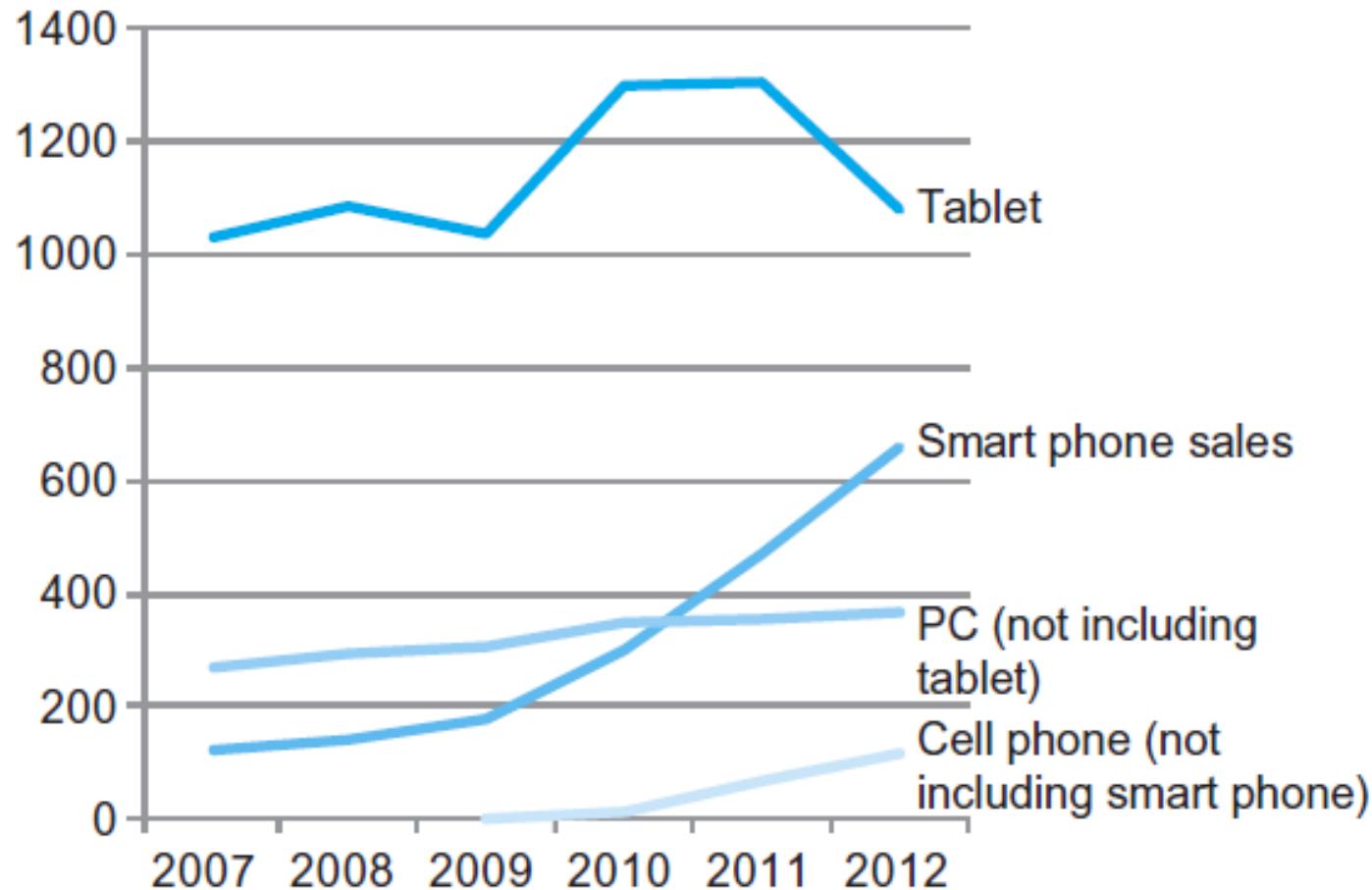
# The Computer Revolution

- Progress in computer technology
  - Underpinned by Moore's Law
- Makes novel applications feasible
  - Computers in automobiles
  - Cell phones
  - Human genome project
  - World Wide Web
  - Search Engines
- Computers are pervasive

# Classes of Computers

- Personal Mobile Device (PMD)
  - e.g. smartphones, tablet computers
  - Emphasis on energy efficiency and real-time
- Desktop Computer
  - Emphasis on price-performance
- Servers
  - Emphasis on availability, scalability, throughput
- Clusters / Warehouse Scale Computers
  - Used for “Software as a Service (SaaS)”
  - Emphasis on availability and price-performance
- Embedded Computers
  - Emphasis: price

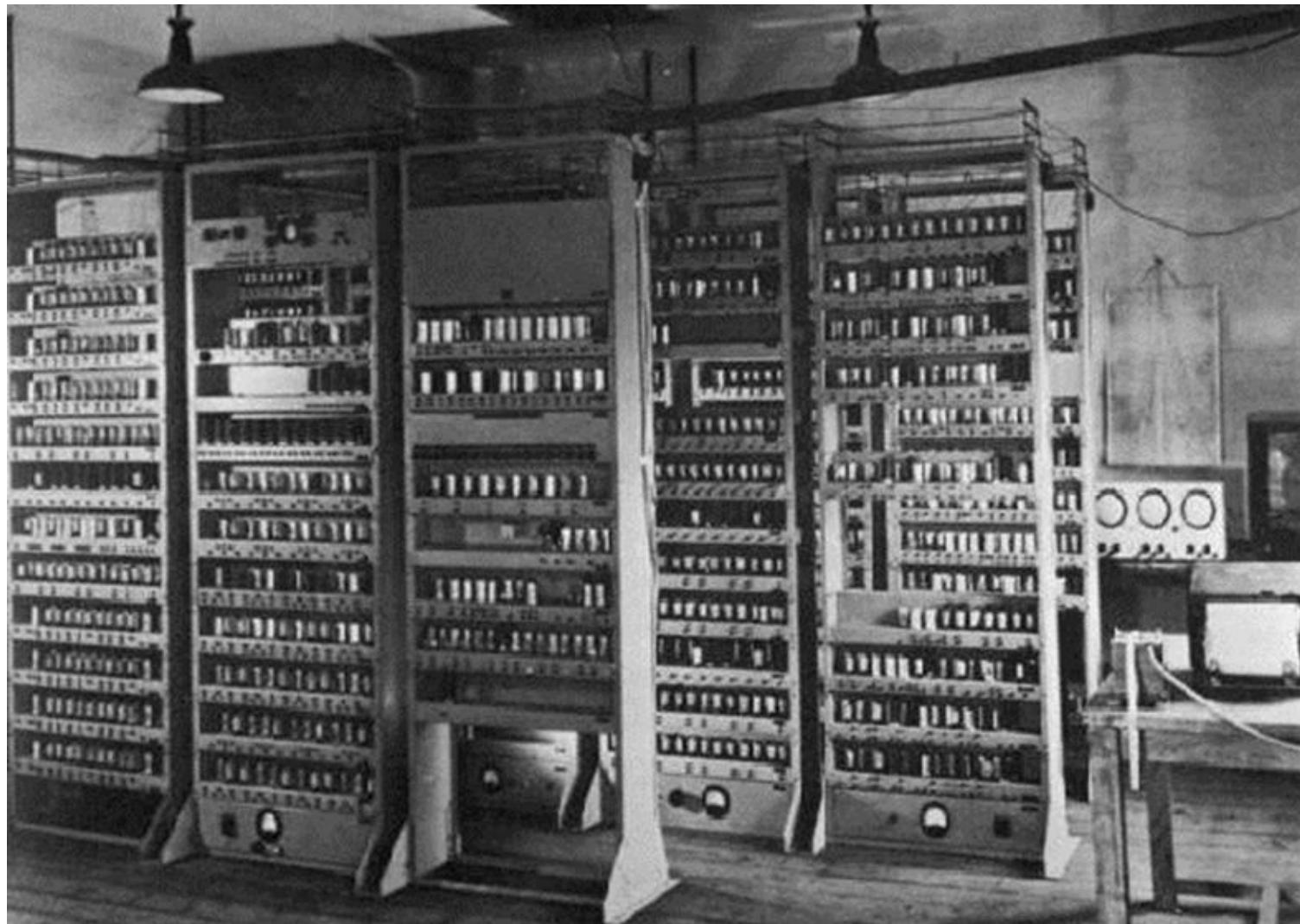
# The PostPC Era



# The PostPC Era

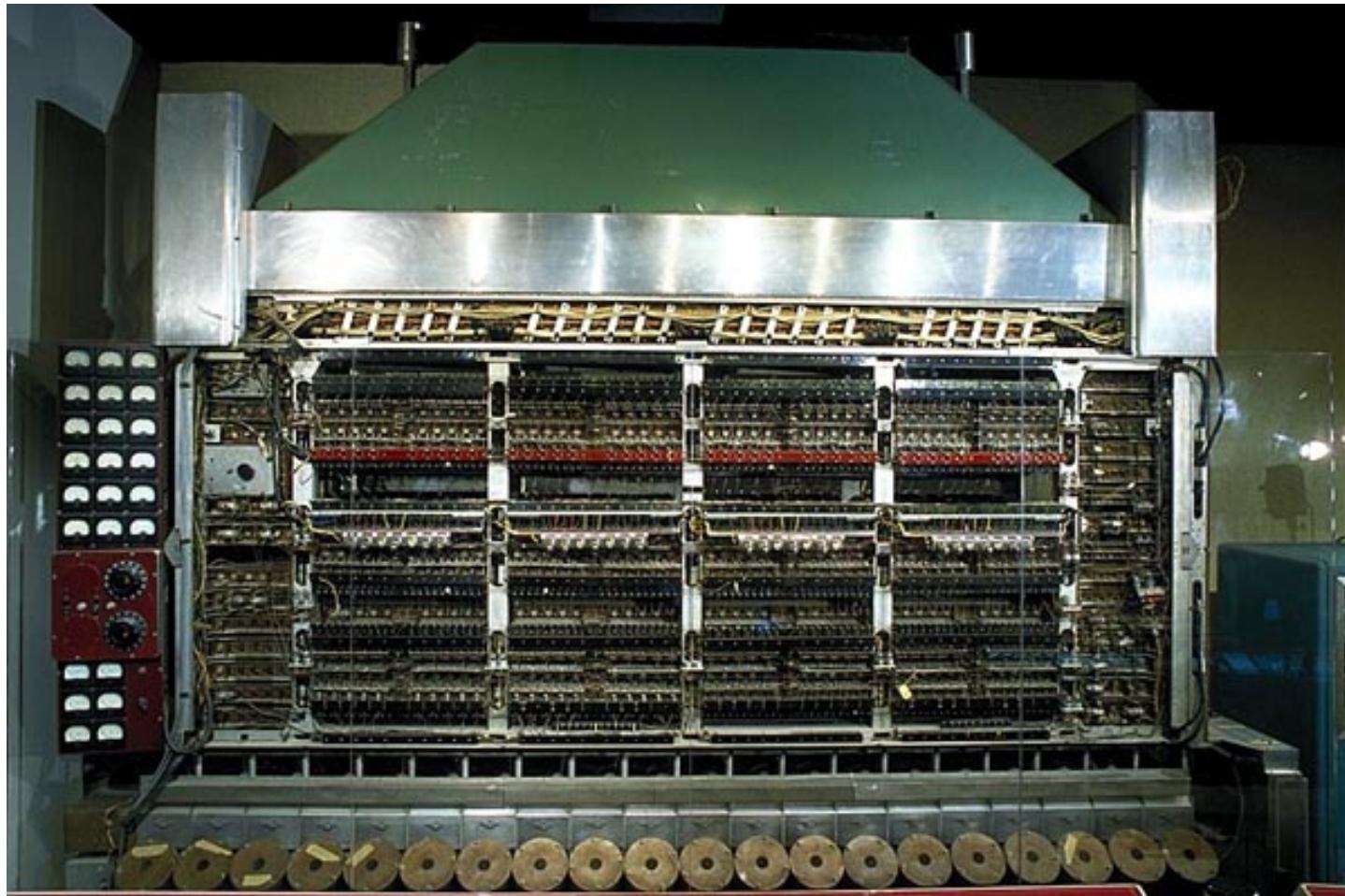
- Personal Mobile Device (PMD)
  - Battery operated
  - Connects to the Internet
  - Hundreds of dollars
  - Smart phones, tablets, electronic glasses
- Cloud computing
  - Warehouse Scale Computers (WSC)
  - Software as a Service (SaaS)
  - Portion of software run on a PMD and a portion run in the Cloud
  - Amazon and Google

# 历史上的计算机…



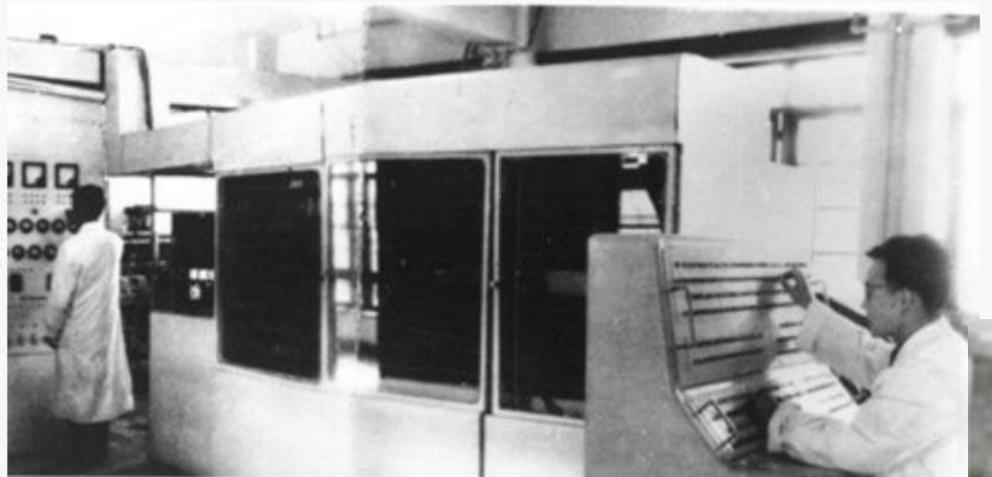
EDSAC, University of Cambridge, UK, 1949

# 历史上的计算机…

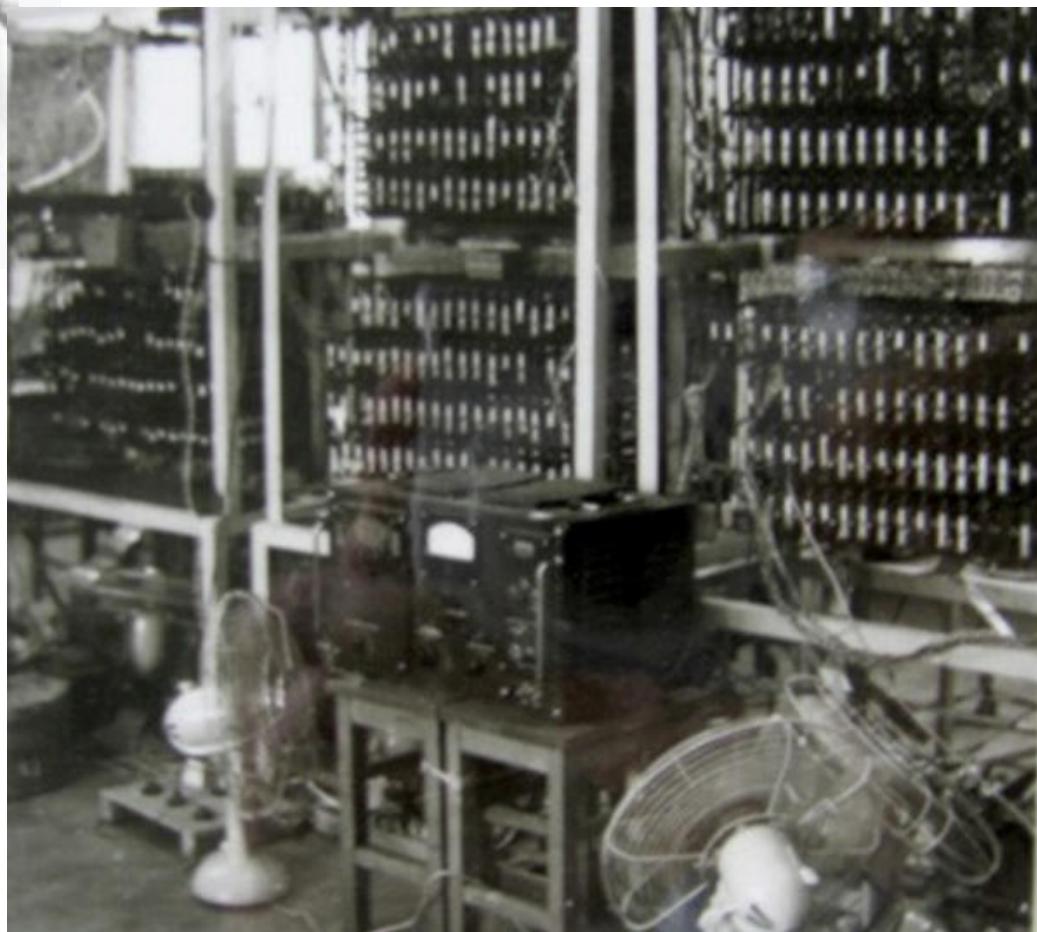


IAS Machine. Design directed by John von Neumann.  
First booted in Princeton NJ in 1952  
Smithsonian Institution Archives (Smithsonian Image 95-06151)

# 历史上的计算机…



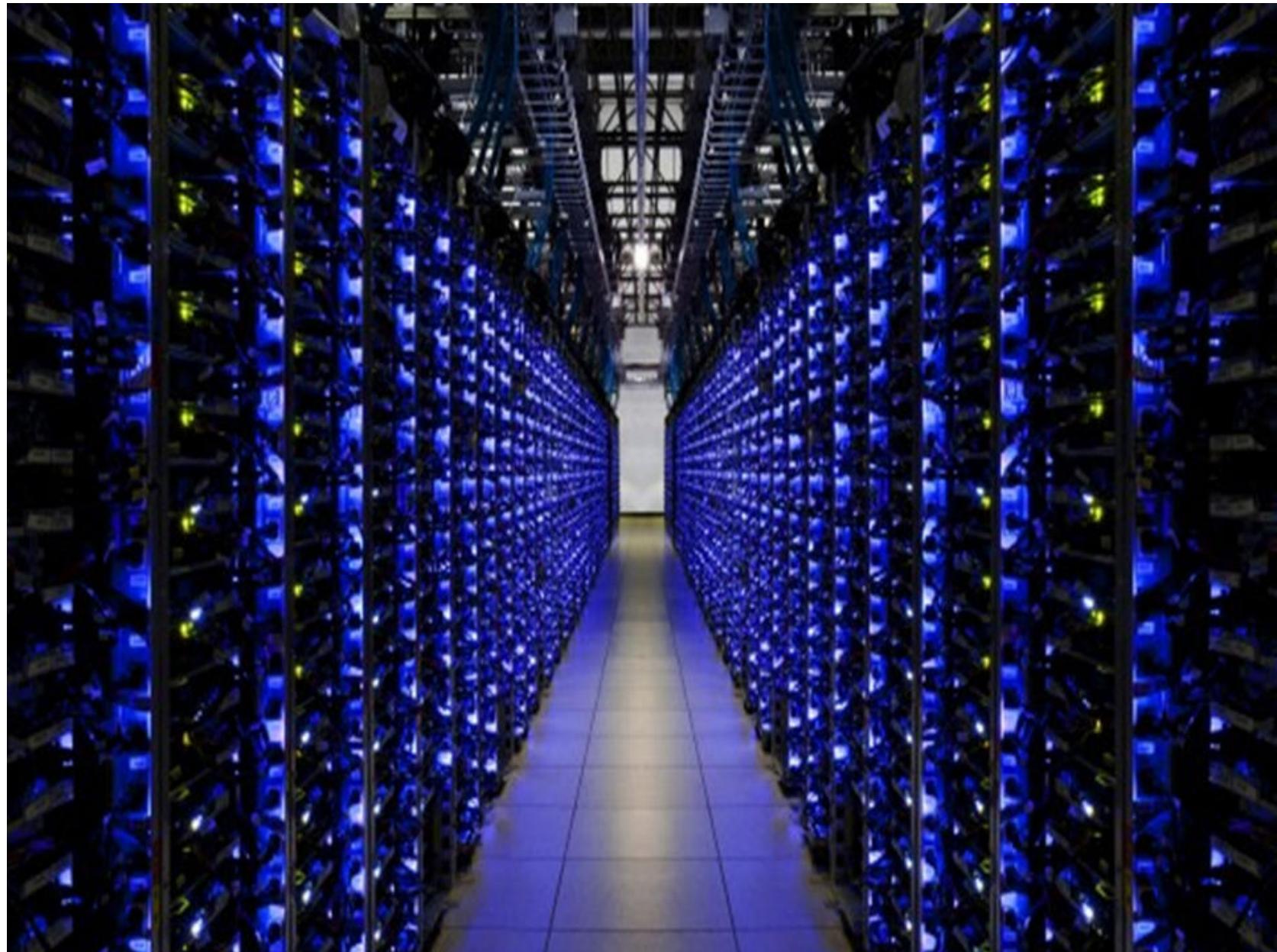
107 (KD-1) 计算机在科大安装调试



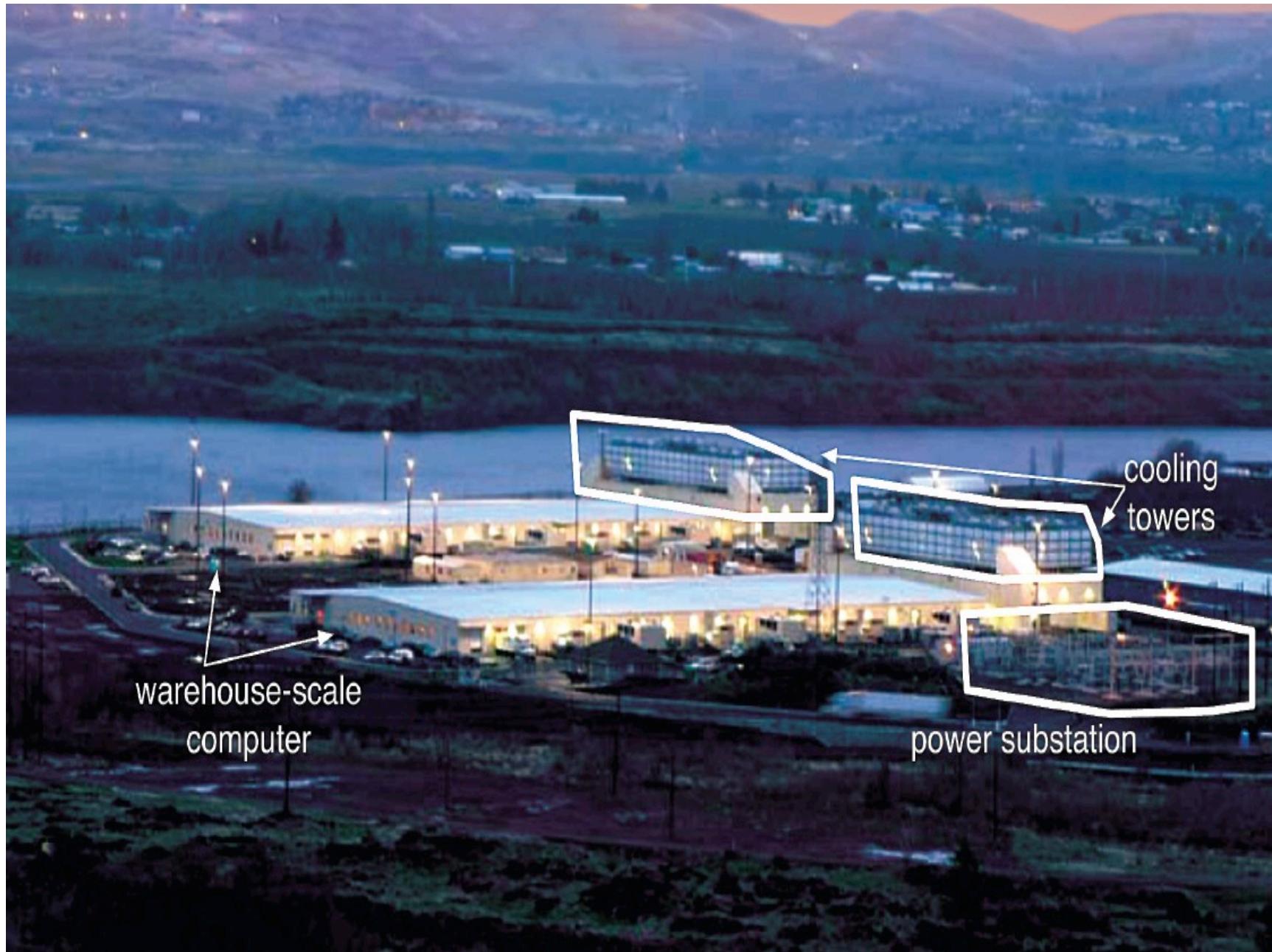
# 现代的计算机...



现代的计算机...

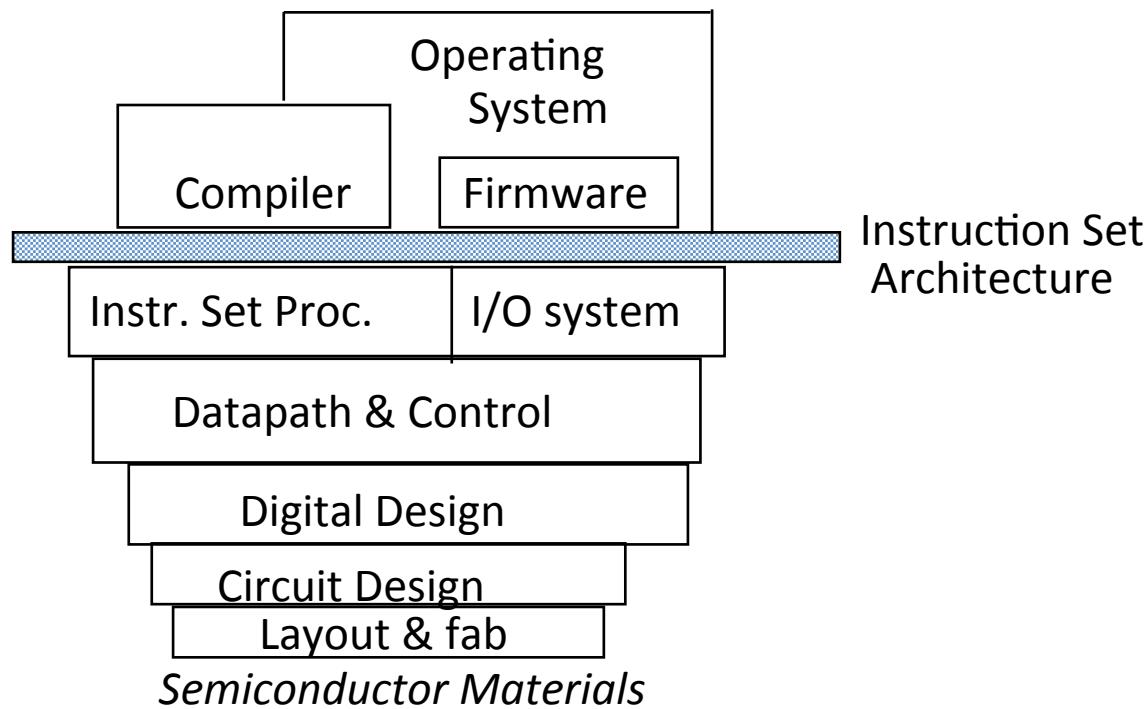


# 现代的计算机...



# What is Computer Architecture?

*Applications*

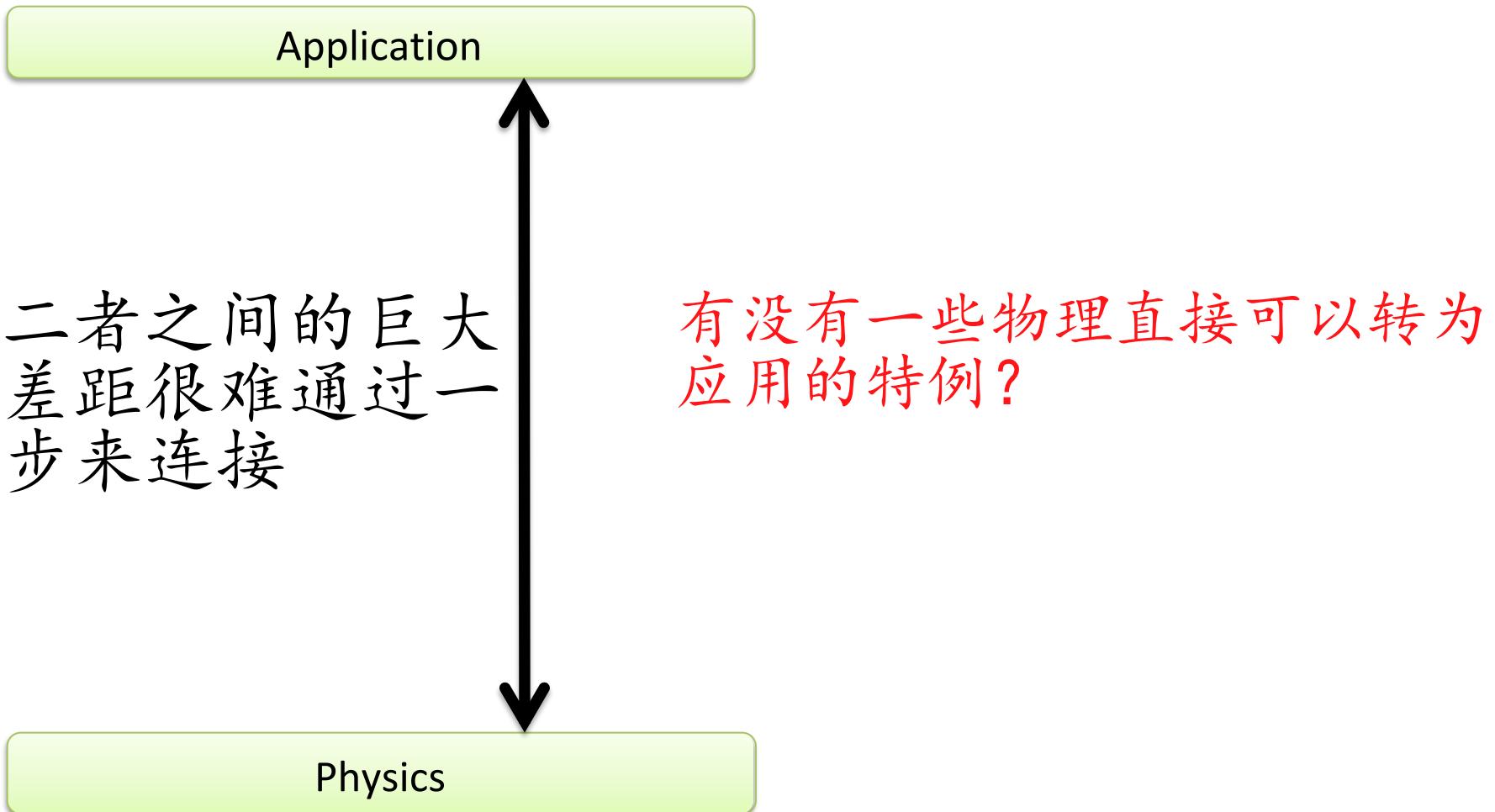


# What is Computer Architecture?

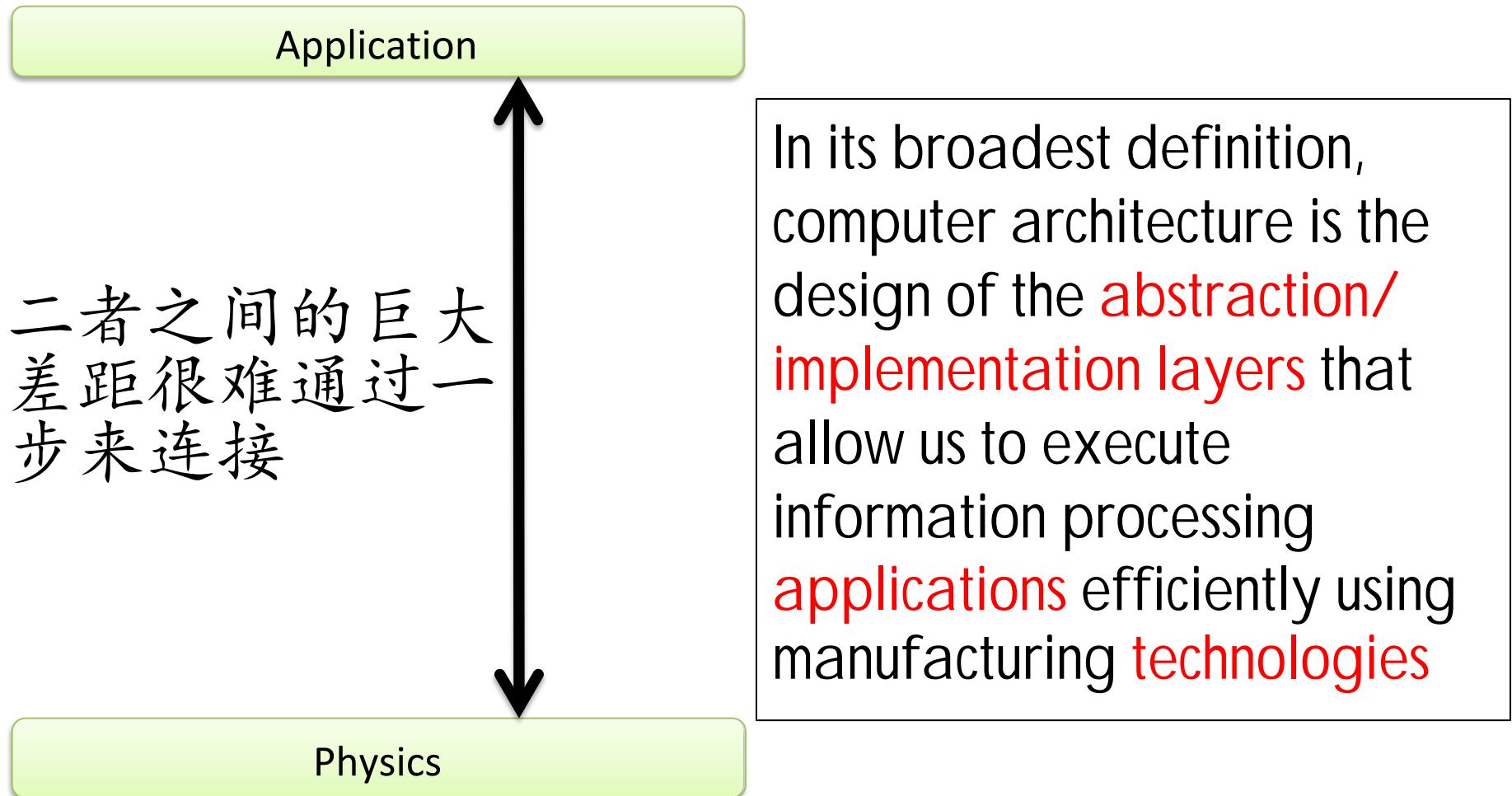
Application

Physics

# What is Computer Architecture?



# What is Computer Architecture?



# Abstractions in Modern Computing Systems

Application

Algorithm

Programming Language

Operating System/Virtual Machines

Instruction Set Architecture

Microarchitecture

Register-Transfer Level

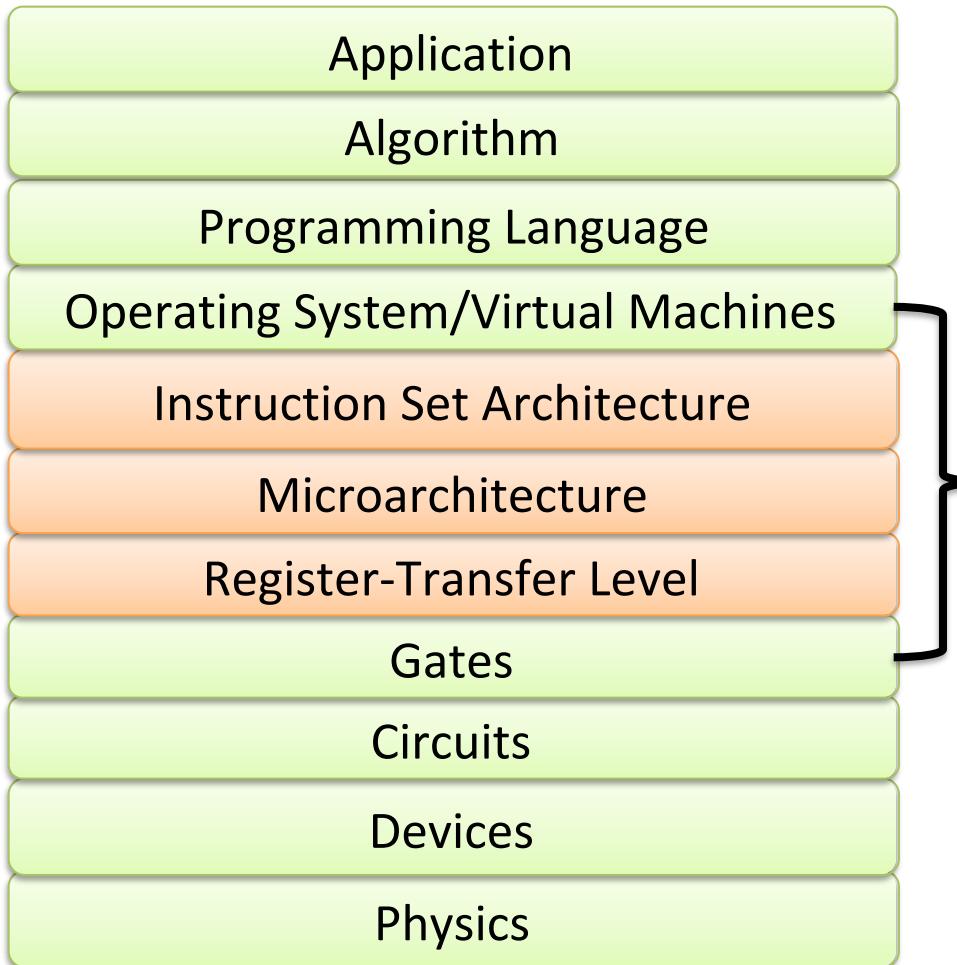
Gates

Circuits

Devices

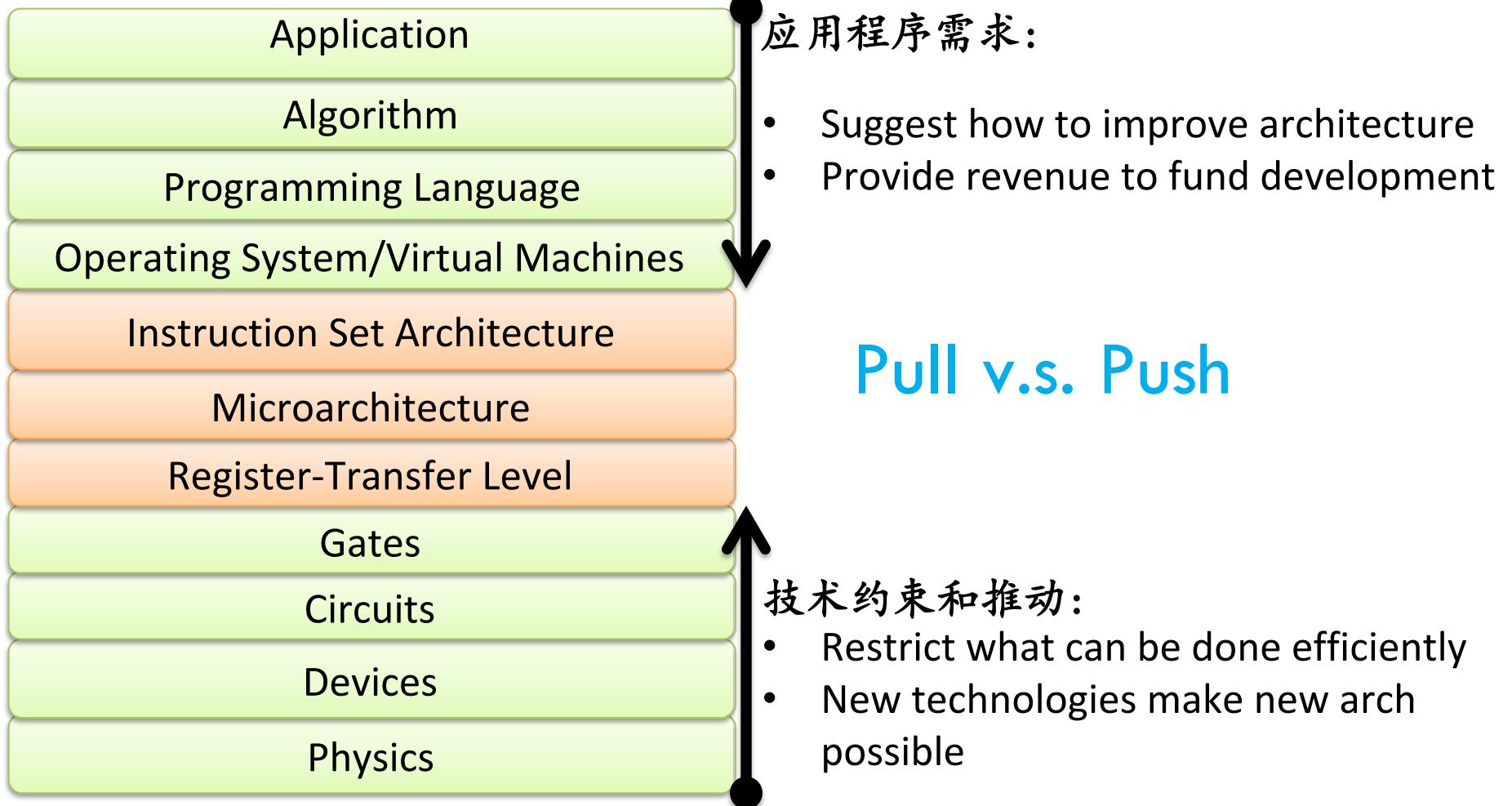
Physics

# Abstractions in Modern Computing Systems

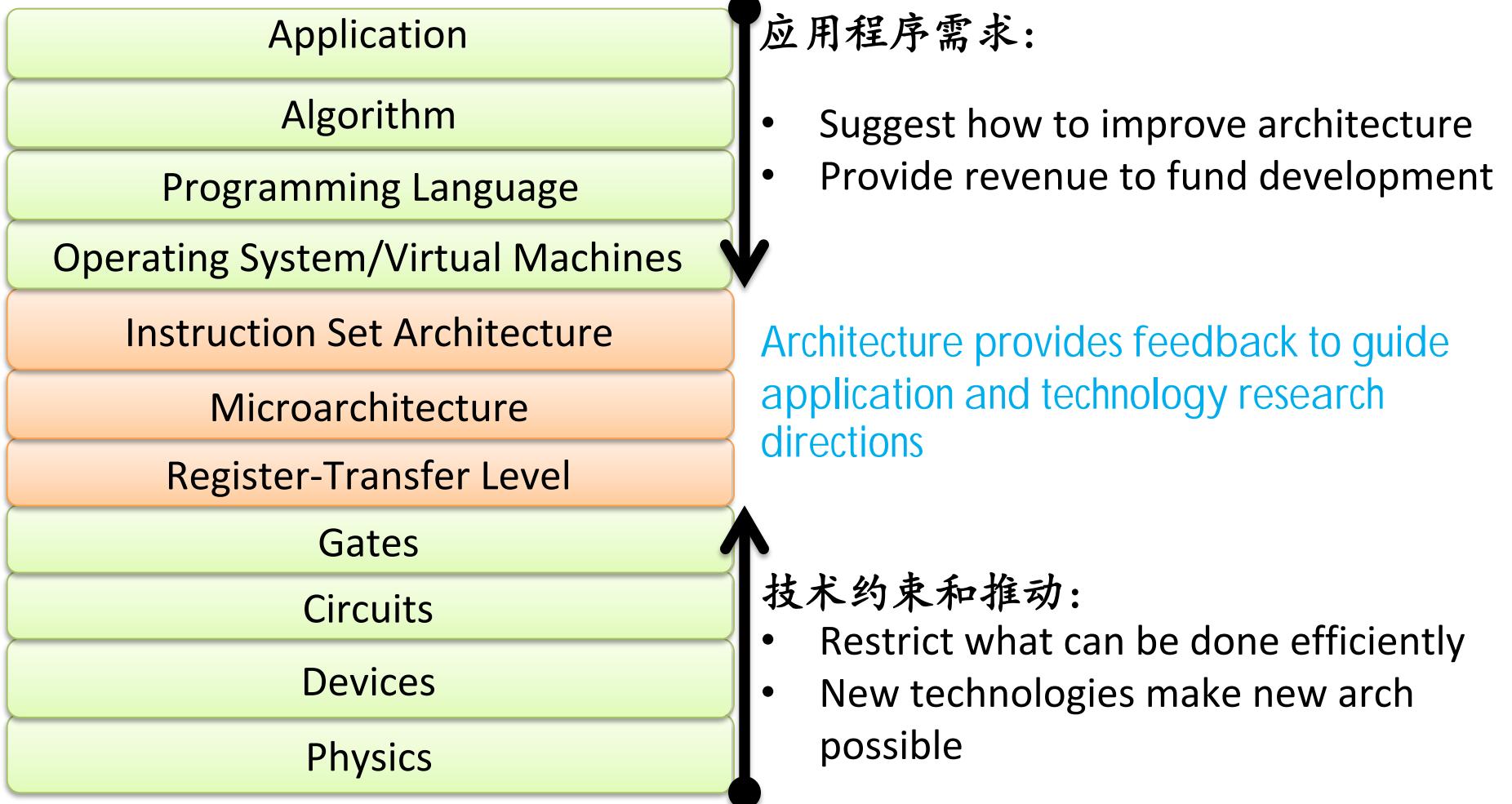


这些是计算机体系结构所关注的内容。

# Computer Architecture is Constantly Changing



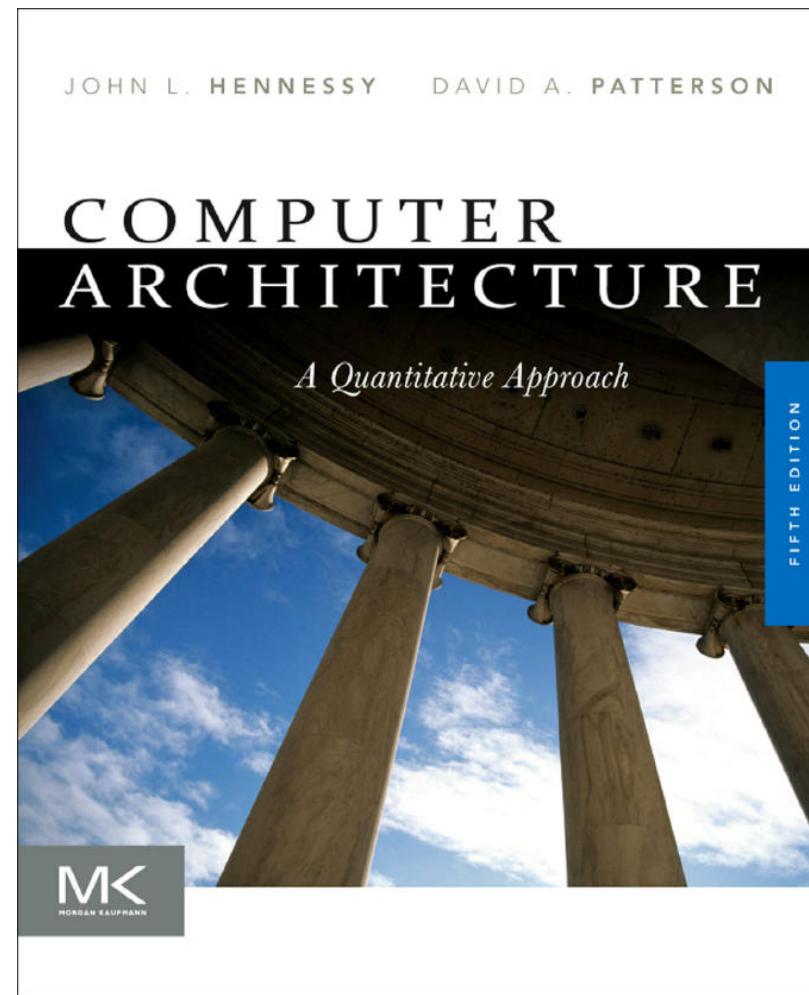
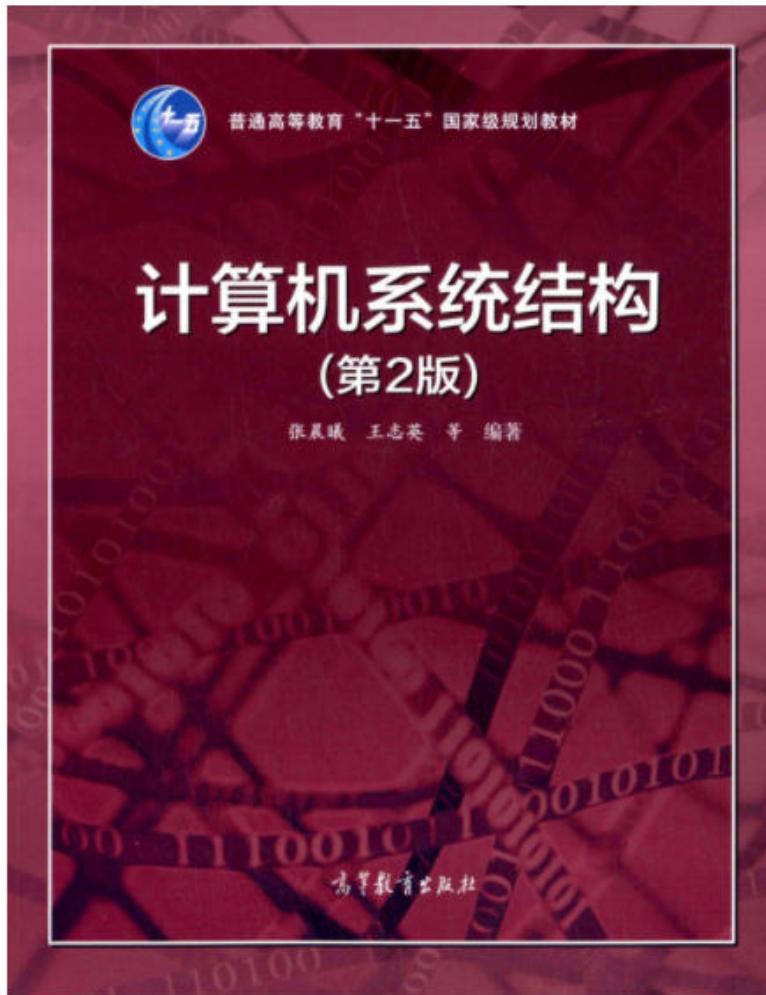
# Computer Architecture is Constantly Changing



# 课程信息

## 课程教材：

参考教材，强烈推荐。



# 课程信息

授课教师：



个人简介：

李建华，男，博士，副研究员。计算机与信息学院，情感计算与系统结构所教师。

邮箱：jhli AT hfut.edu.cn

通讯地址：合肥市蜀山区丹霞路485号合肥工业大学翡翠湖校区翠教楼A806 邮编：230601

# 课程信息

考核方式：

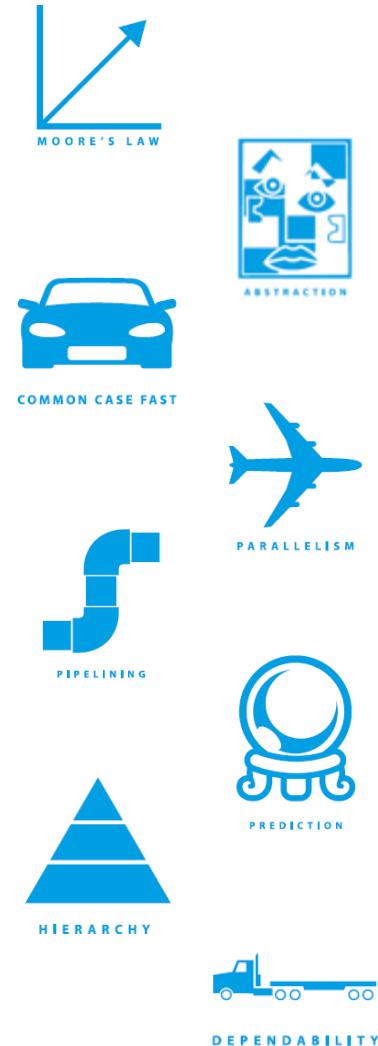
代码	名称	占比
EM6	作业撰写	10
EM11	出勤率	10
EM16	实验报告	20
EM3	期末考试	40
EM4	课堂测试	20

- ✓ 4次课程作业
- ✓ 3个课程实验
- ✓ 1次课堂测试

1个课程报告，10%。

# Great Ideas in Computer Architectures

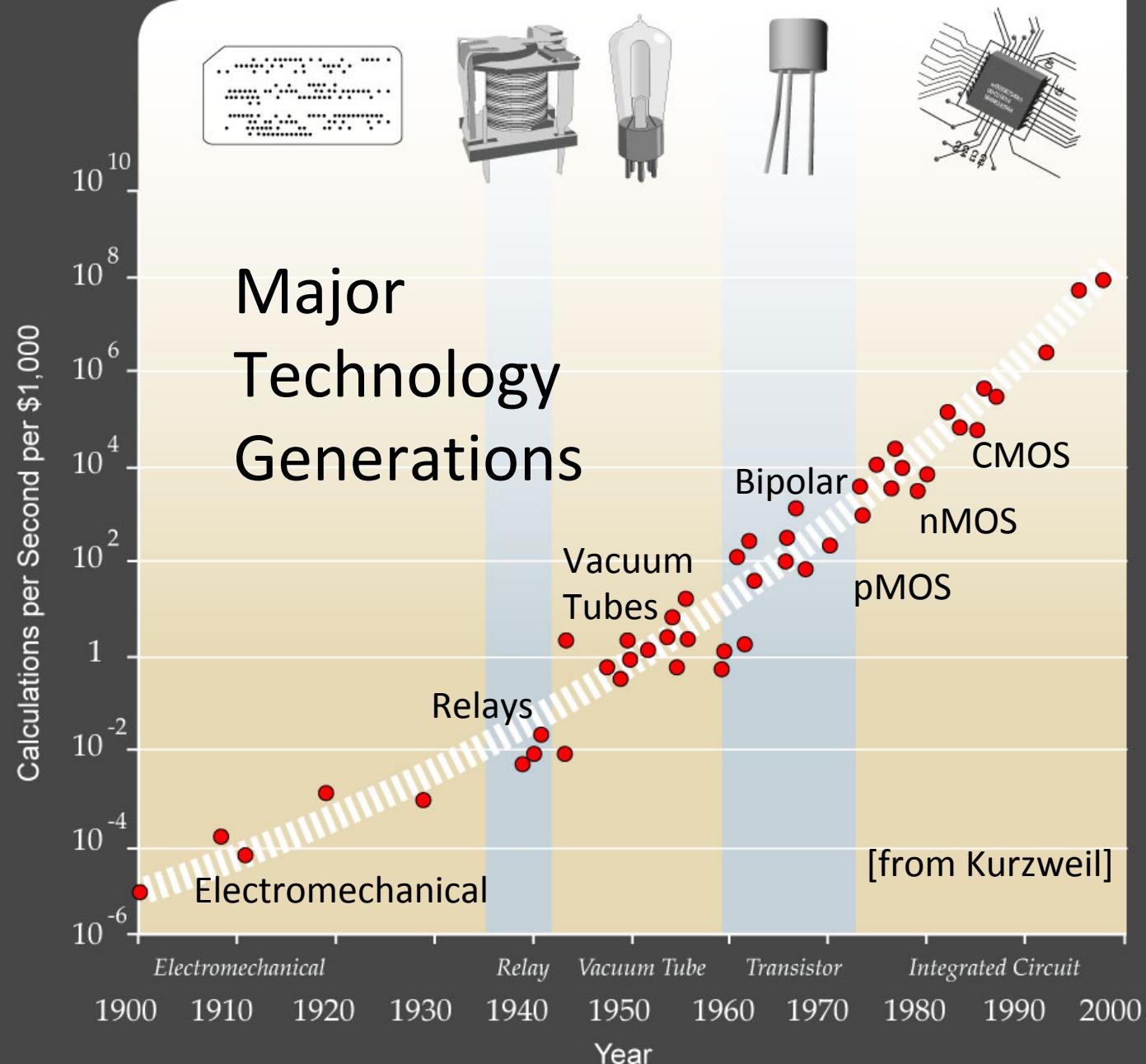
1. Design for ***Moore's Law***
2. Use ***abstraction*** to simplify design
3. Make the ***common case fast***
4. Performance via ***parallelism***
5. Performance via ***pipelining***
6. Performance via ***prediction***
7. ***Hierarchy*** of memories
8. ***Dependability*** via redundancy



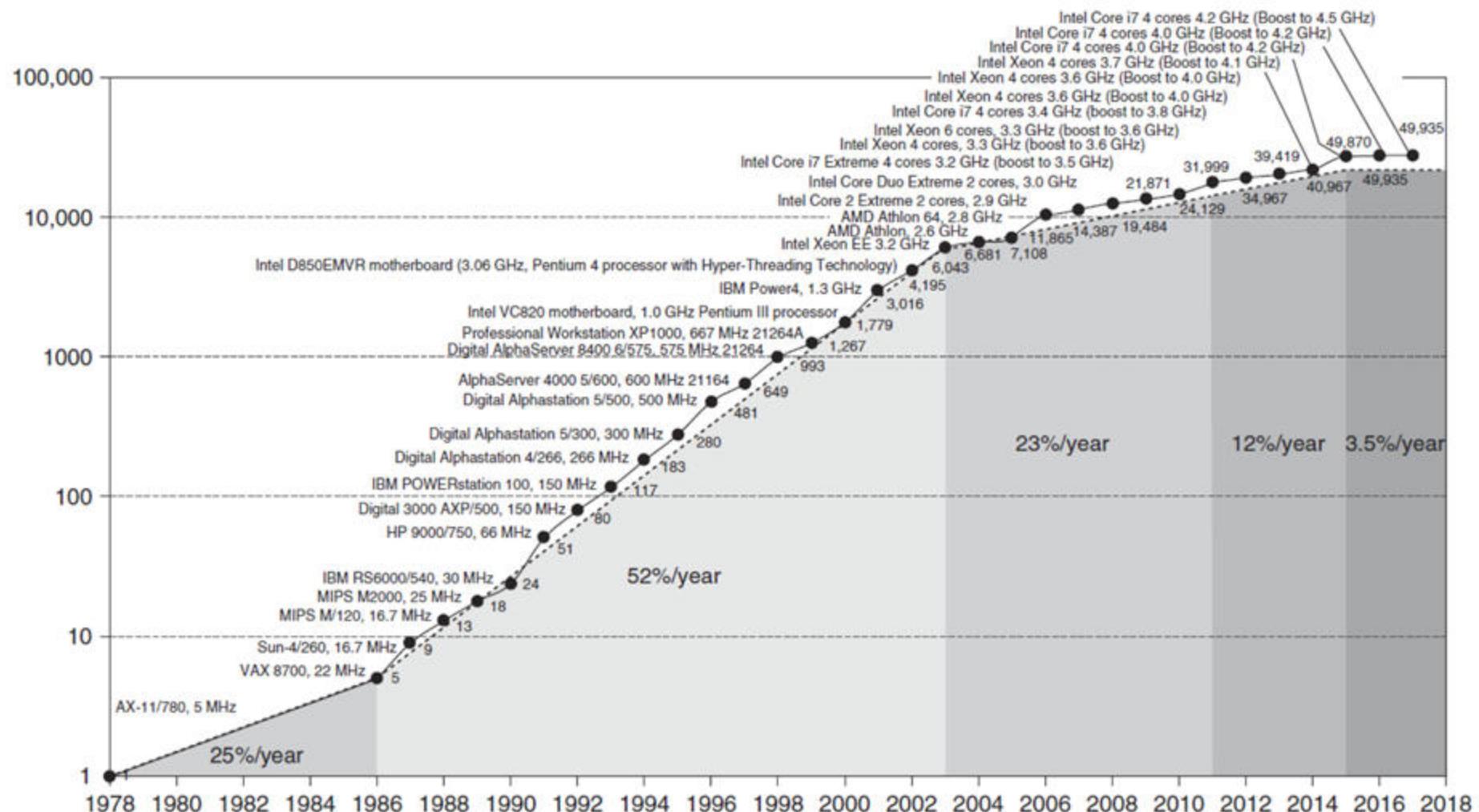
## Moore's Law

### The Fifth Paradigm

Logarithmic Plot



# Sequential Processor Performance

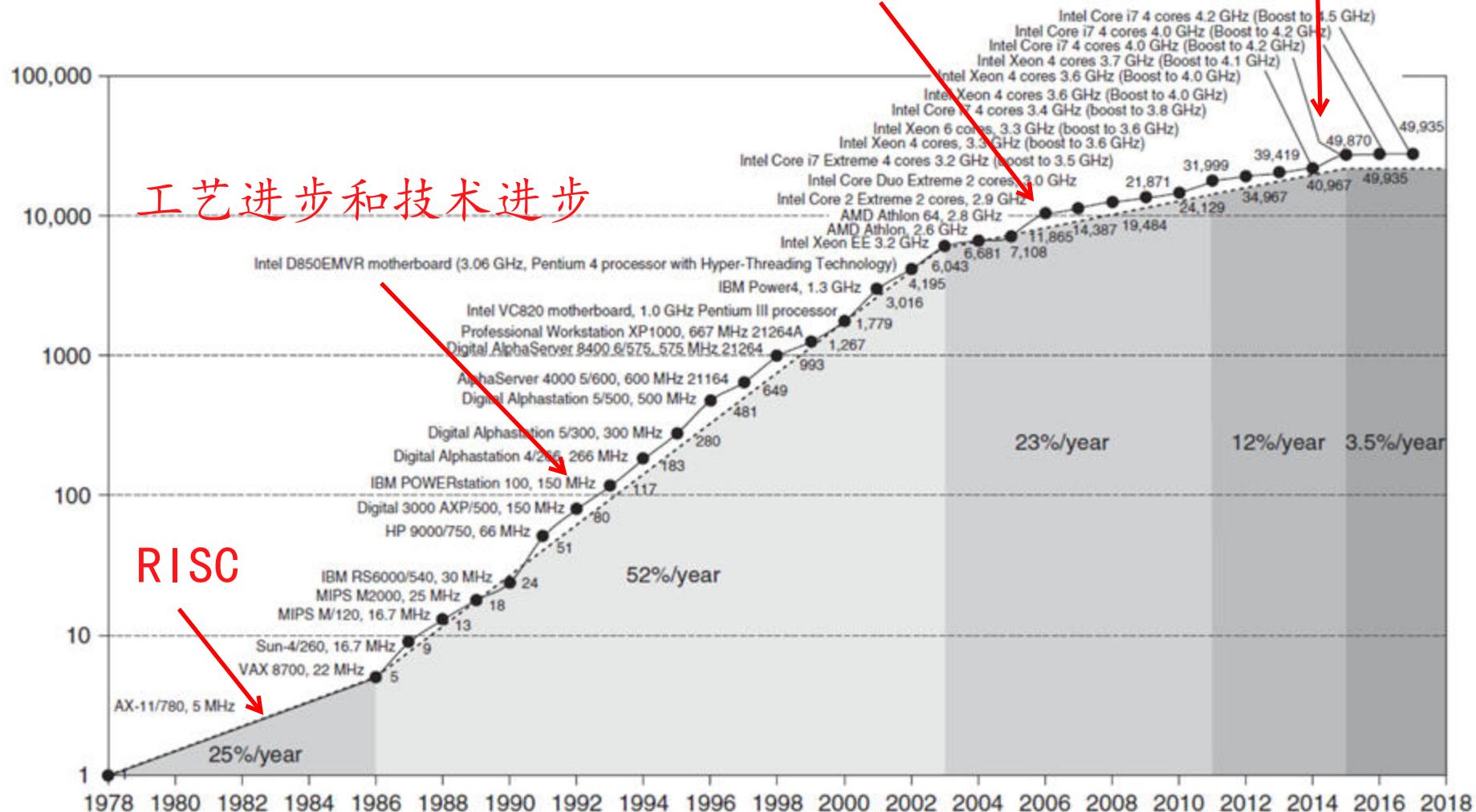


From Hennessy and Patterson 6e Image Copyright © 2019, Elsevier Inc. All rights Reserved.

摩尔定律和丹纳德缩放定律的终结

# Sequential Processor Performance

转向多核架构



# 课程内容

## 计算机组成

### Computer Organization

- Basic Pipelined Processor

~50,000 Transistors

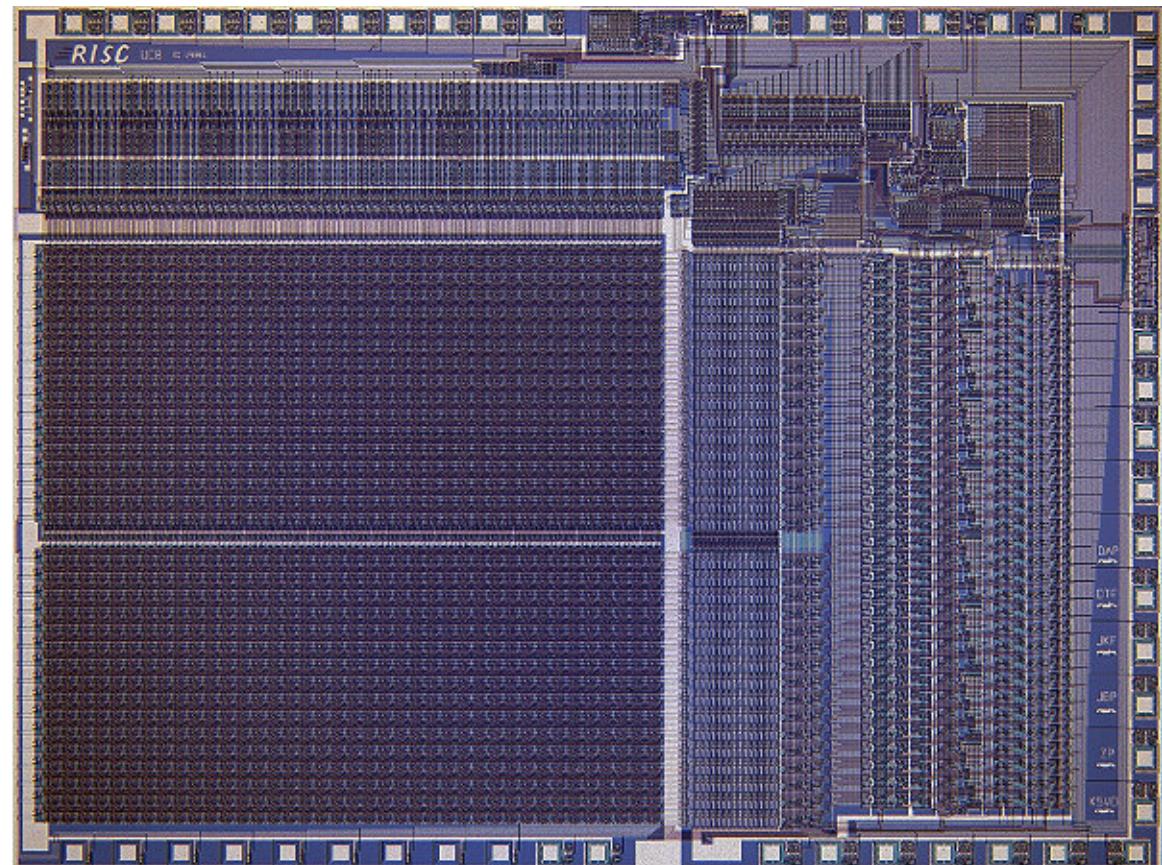
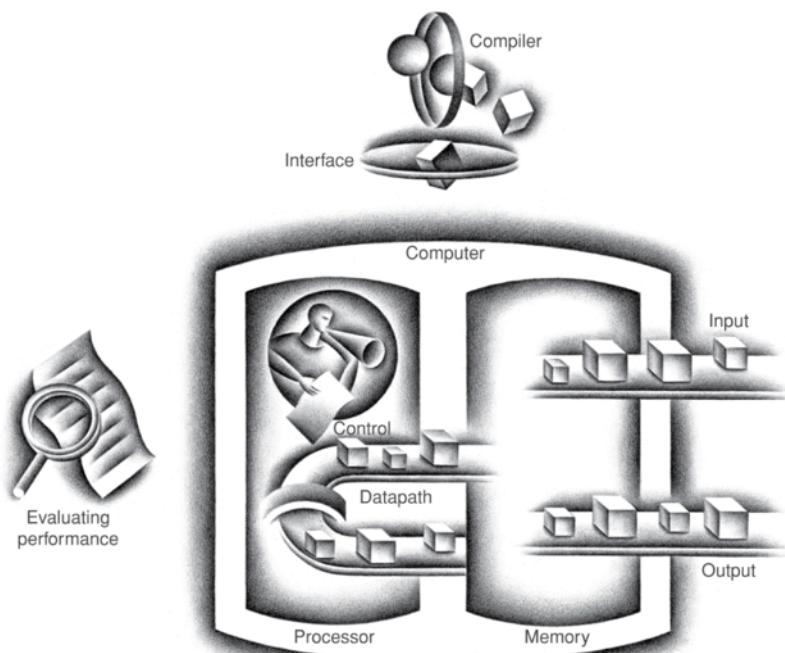


Photo of Berkeley RISC I, © University of California (Berkeley)

# Components of a Computer

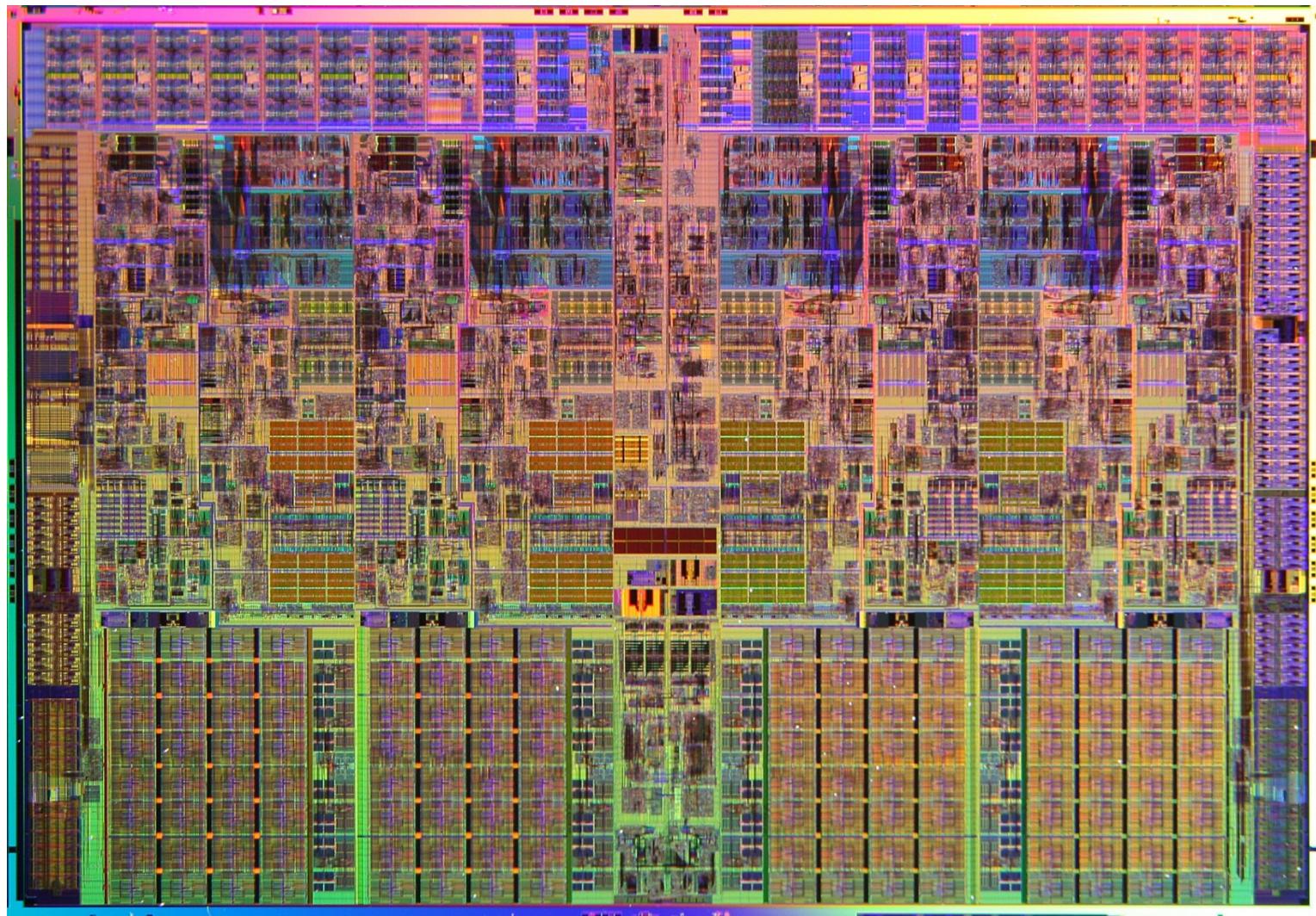
## The BIG Picture



- Same components for all kinds of computer
  - Desktop, server, embedded
- Input/output includes
  - User-interface devices
    - Display, keyboard, mouse
  - Storage devices
    - Hard disk, CD/DVD, flash
  - Network adapters
    - For communicating with other computers

# 课程内容

## 计算机体系结构

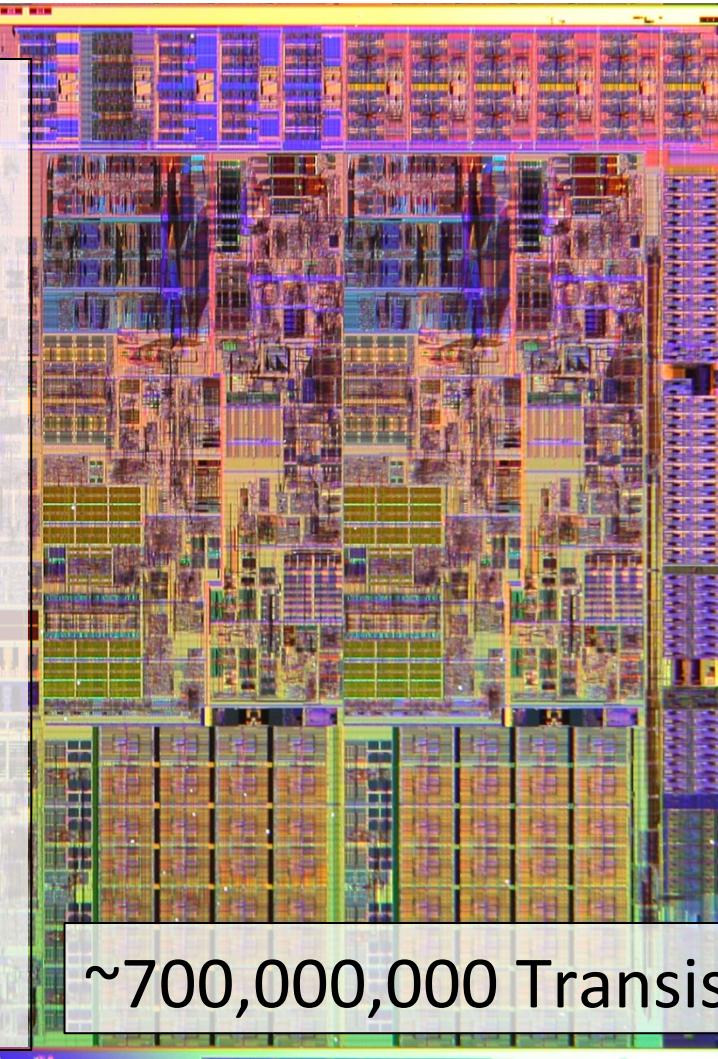


Intel Nehalem Processor, Original Core i7, Image Credit Intel:  
[http://download.intel.com/pressroom/kits/corei7/images/Nehalem\\_Die\\_Shot\\_3.jpg](http://download.intel.com/pressroom/kits/corei7/images/Nehalem_Die_Shot_3.jpg)

# 课程内容

## 计算机体系结构

- Instruction Level Parallelism
  - Superscalar
  - Very Long Instruction Word (VLIW)
- Long Pipelines (Pipeline Parallelism)
- Advanced Memory and Caches
- Data Level Parallelism
  - Vector
  - GPU
- Thread Level Parallelism
  - Multithreading
  - Multiprocessor
  - Multicore
  - Manycore



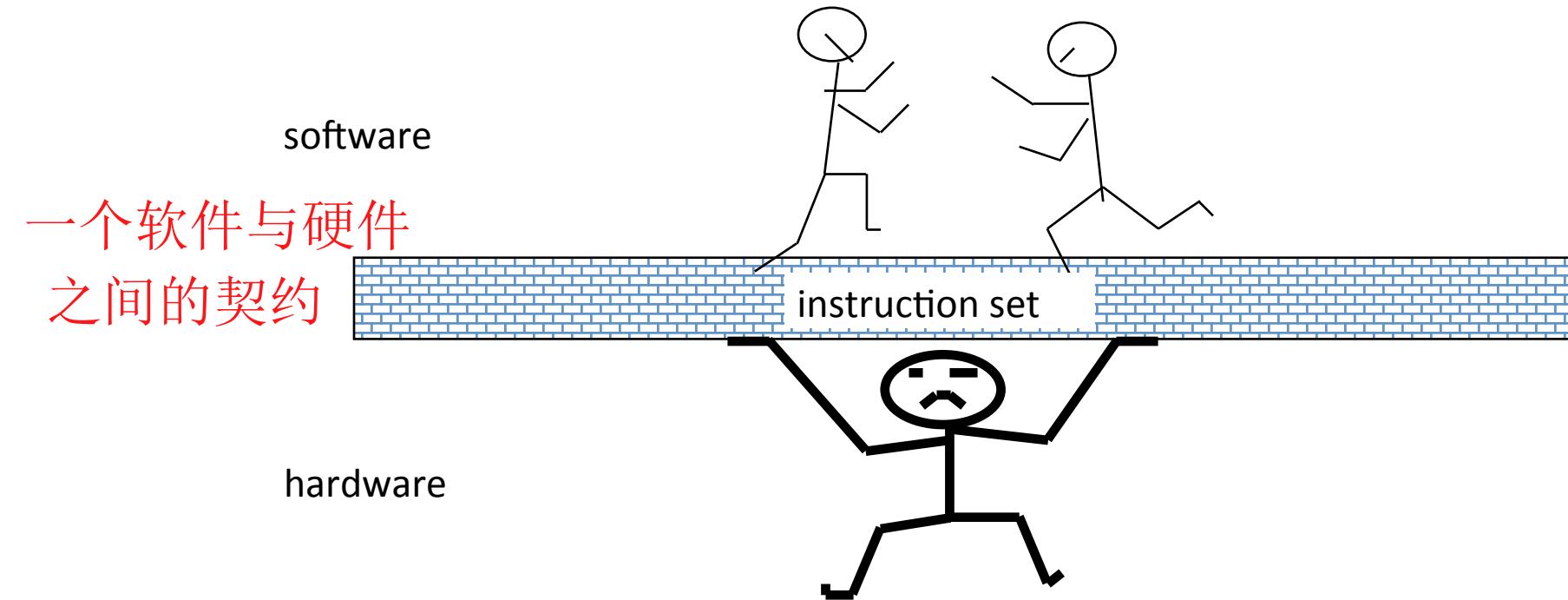
Intel Nehalem Processor, Original Core i7, Image Credit Intel:  
[http://download.intel.com/pressroom/kits/corei7/images/Nehalem\\_Die\\_Shot\\_3.jpg](http://download.intel.com/pressroom/kits/corei7/images/Nehalem_Die_Shot_3.jpg)

# Architecture vs. Microarchitecture

“Architecture”/Instruction Set Architecture:

- Programmer visible state (Memory & Register)
- Operations (Instructions and how they work)
- Execution Semantics (interrupts)
- Input/Output
- Data Types/Sizes

# Instruction Set Architecture



- Properties of a good abstraction
  - Lasts through many generations (portability)
  - Used in many different ways (generality)
  - Provides convenient functionality to higher levels
  - Permits an efficient implementation at lower levels

# Architecture vs. Microarchitecture

“Architecture”/Instruction Set Architecture:

- Programmer visible state (Memory & Register)
- Operations (Instructions and how they work)
- Execution Semantics (interrupts)
- Input/Output
- Data Types/Sizes

Microarchitecture/Organization:

- Tradeoffs on how to implement ISA for some metric (Speed, Energy, Cost)
- Examples: Pipeline depth, number of pipelines, cache size, silicon area, peak power, execution ordering, bus widths, ALU widths

# 软件的发展

up to 1955

Libraries of numerical routines

- Floating point operations
- Transcendental functions
- Matrix manipulation, equation solvers, . . .

1955-60

*High level Languages - Fortran 1956  
Operating Systems -*

- Assemblers, Loaders, Linkers, Compilers
- Accounting programs to keep track of usage and charges

# 软件的发展

up to 1955

Libraries of numerical routines

- Floating point operations
- Transcendental functions
- Matrix manipulation, equation solvers, . . .

1955-60

*High level Languages - Fortran 1956*

*Operating Systems -*

- Assemblers, Loaders, Linkers, Compilers
- Accounting programs to keep track of usage and charges

Machines required *experienced operators*

- Most users could not be expected to understand these programs, much less write them
- Machines had to be sold with a lot of resident software

# IBM的兼容性问题

By early 1960's, IBM had 4 incompatible lines of computers!

701      ⇒      7094

650      ⇒      7074

702      ⇒      7080

1401     ⇒      7010

# IBM的兼容性问题

By early 1960's, IBM had 4 incompatible lines of computers!

701	⇒	7094
650	⇒	7074
702	⇒	7080
1401	⇒	7010

Each system had its own

- Instruction set
- I/O system and Secondary Storage:  
magnetic tapes, drums and disks
- assemblers, compilers, libraries,...
- market niche business, scientific, real time, ...

这会导致什么问题？

# IBM的兼容性问题

By early 1960's, IBM had 4 incompatible lines of computers!

701	⇒	7094
650	⇒	7074
702	⇒	7080
1401	⇒	7010

Each system had its own

- Instruction set
- I/O system and Secondary Storage:  
magnetic tapes, drums and disks
- assemblers, compilers, libraries,...
- market niche business, scientific, real time, ...

⇒ IBM 360

# IBM 360: A General-Purpose Register (GPR) Machine

- Processor State
  - 16 General-Purpose 32-bit Registers
    - *may be used as index and base register*
    - *Register 0 has some special properties*
  - 4 Floating Point 64-bit Registers
  - A Program Status Word (PSW)
    - *PC, Condition codes, Control flags*
- A 32-bit machine with 24-bit addresses
  - But no instruction contains a 24-bit address!
- Data Formats
  - 8-bit bytes, 16-bit half-words, 32-bit words, 64-bit double-words

# IBM 360: A General-Purpose Register (GPR) Machine

- Processor State
    - 16 General-Purpose 32-bit Registers
      - *may be used as index and base register*
      - *Register 0 has some special properties*
    - 4 Floating Point 64-bit Registers
    - A Program Status Word (PSW)
      - *PC, Condition codes, Control flags*
  - A 32-bit machine with 24-bit addresses
    - But no instruction contains a 24-bit address!
  - Data Formats
    - 8-bit bytes, 16-bit half-words, 32-bit words, 64-bit double-words
- The IBM 360 is why bytes are 8-bits long today!*

# IBM 360: Initial Implementations

	<i>Model 30</i>	...	<i>Model 70</i>
<i>Storage</i>	8K - 64 KB		256K - 512 KB
<i>Datapath</i>	8-bit		64-bit
<i>Circuit Delay</i>	30 nsec/level		5 nsec/level
<i>Local Store</i>	Main Store		Transistor Registers
<i>Control Store</i>	Read only 1 $\mu$ sec		Conventional circuits

IBM 360 instruction set architecture (ISA) completely hid the underlying technological differences between various models.

# IBM 360: Initial Implementations

	<i>Model 30</i>	...	<i>Model 70</i>
<i>Storage</i>	8K - 64 KB		256K - 512 KB
<i>Datapath</i>	8-bit		64-bit
<i>Circuit Delay</i>	30 nsec/level		5 nsec/level
<i>Local Store</i>	Main Store		Transistor Registers
<i>Control Store</i>	Read only 1 $\mu$ sec		Conventional circuits

IBM 360 instruction set architecture (ISA) completely hid the underlying technological differences between various models.

Milestone: The first true ISA designed as portable hardware-software interface!

# IBM 360: Initial Implementations

	<i>Model 30</i>	...	<i>Model 70</i>
<i>Storage</i>	8K - 64 KB		256K - 512 KB
<i>Datapath</i>	8-bit		64-bit
<i>Circuit Delay</i>	30 nsec/level		5 nsec/level
<i>Local Store</i>	Main Store		Transistor Registers
<i>Control Store</i>	Read only 1 $\mu$ sec		Conventional circuits

IBM 360 instruction set architecture (ISA) completely hid the underlying technological differences between various models.

Milestone: The first true ISA designed as portable hardware-software interface!

*With minor modifications it still survives today!*

# IBM 360: Over 50 years later...

## The zSeries z14 Microprocessor



Image Credit: IBM  
Courtesy of International Business  
Machines Corporation, © International  
Business Machines Corporation.

- 5.2 GHz in IBM 14nm SOI CMOS technology
- 6.1 billion transistors in 696 mm<sup>2</sup>
- 64-bit virtual addressing
  - original S/360 was 24-bit, and S/370 was 31-bit extension
- 10-core design
- 6-fetch/cycle
- 10-issue/cycle out-of-order superscalar pipeline
- Out-of-order memory accesses
- Redundant datapaths
  - every instruction performed in two parallel datapaths and results compared
- 128KB L1 I-cache, 128KB L1 D-cache on-chip
- 2MB private I-cache L2 per core
- 4MB private D-cache L2 per core
- On-Chip 128MB eDRAM L3 cache
- Up to 672MB eDRAM L4

# Same Architecture Different Microarchitecture

## AMD Phenom X4

- X86 Instruction Set
- Quad Core
- 125W
- Decode 3 Instructions/Cycle/Core
- 64KB L1 I Cache, 64KB L1 D Cache
- 512KB L2 Cache
- Out-of-order
- 2.6GHz

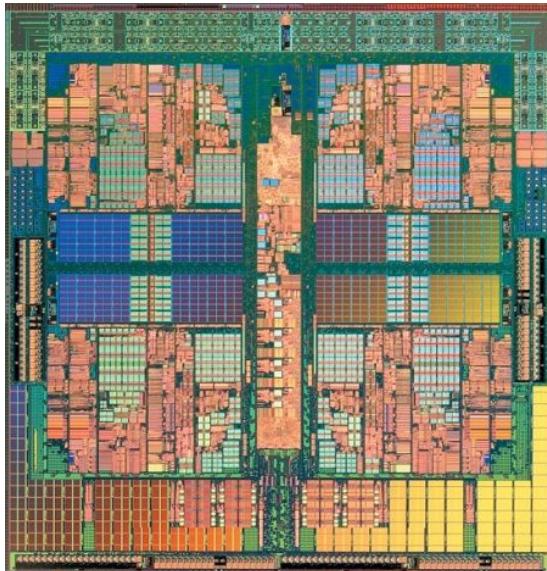


Image Credit: AMD

## Intel Atom

- X86 Instruction Set
- Single Core
- 2W
- Decode 2 Instructions/Cycle/Core
- 32KB L1 I Cache, 24KB L1 D Cache
- 512KB L2 Cache
- In-order
- 1.6GHz

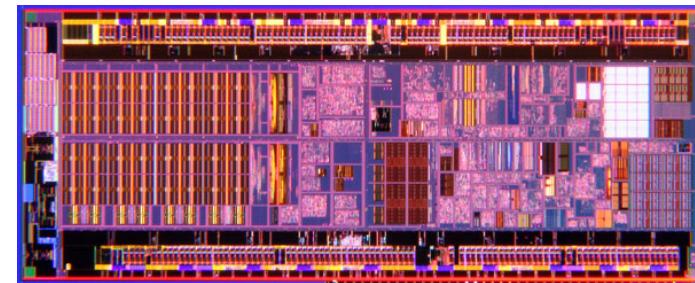


Image Credit: Intel

# Different Architecture Different Microarchitecture

## AMD Phenom X4

- X86 Instruction Set
- Quad Core
- 125W
- Decode 3 Instructions/Cycle/Core
- 64KB L1 I Cache, 64KB L1 D Cache
- 512KB L2 Cache
- Out-of-order
- 2.6GHz

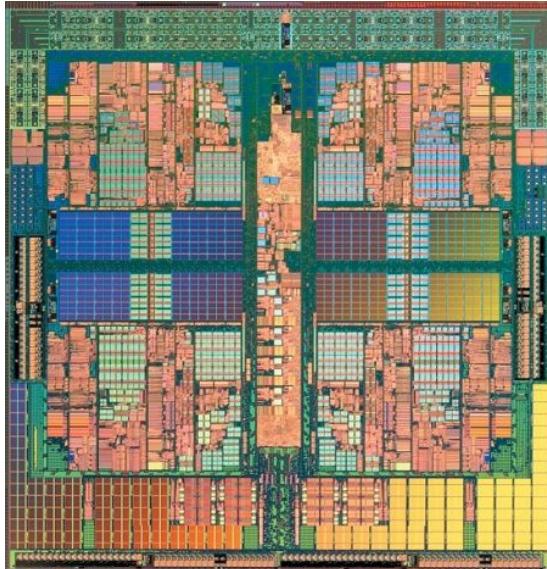


Image Credit: AMD

## IBM POWER7

- Power Instruction Set
- Eight Core
- 200W
- Decode 6 Instructions/Cycle/Core
- 32KB L1 I Cache, 32KB L1 D Cache
- 256KB L2 Cache
- Out-of-order
- 4.25GHz

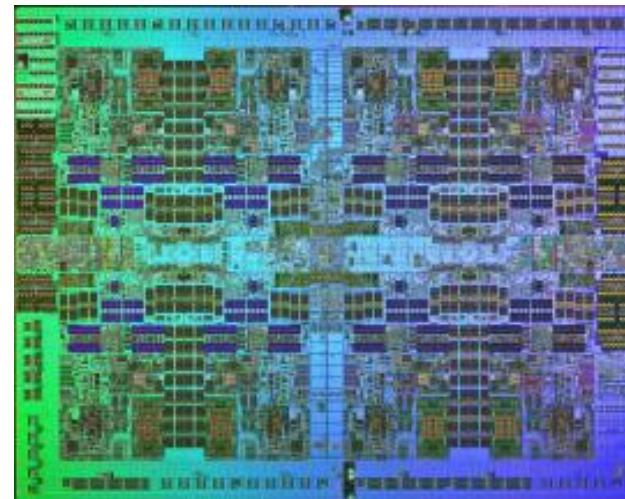
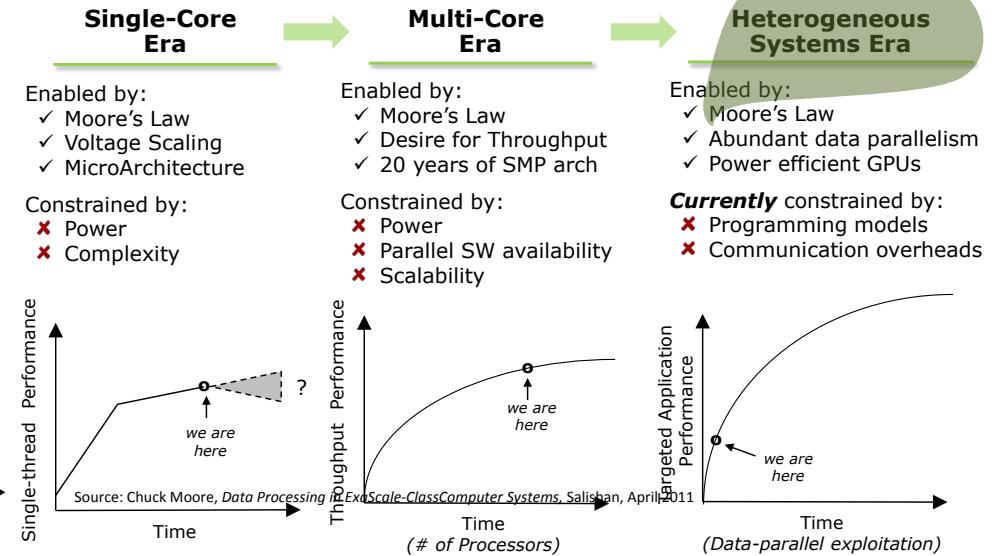
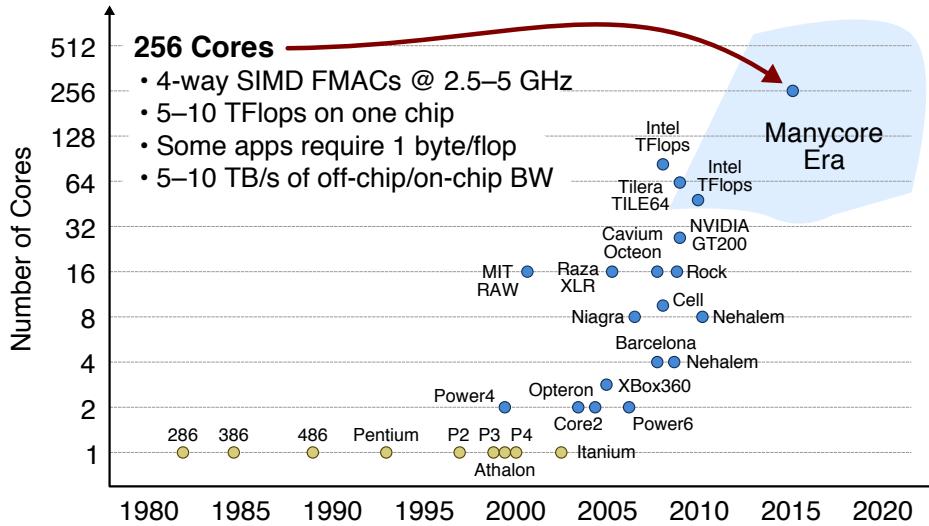


Image Credit: IBM  
Courtesy of International Business Machines Corporation, © International Business Machines Corporation.

# Architectural Challenges



- Massive (ca. 4X) increase in concurrency
  - Multicore (4 - <100) → Manycores (100s – 1ks)
- Heterogeneity
  - System-level (accelerators) vs chip level (embedded)
- Compute power and memory speed challenges (two walls)
  - *500x compute power and 30x memory of 2PF HW*
  - *Memory access time lags further behind*

