Computer Architecture (Fall 2022)

Introduction

Dr. Duo Liu (刘铎)
Office: Main Building 0626
Email: liuduo@cqu.edu.cn

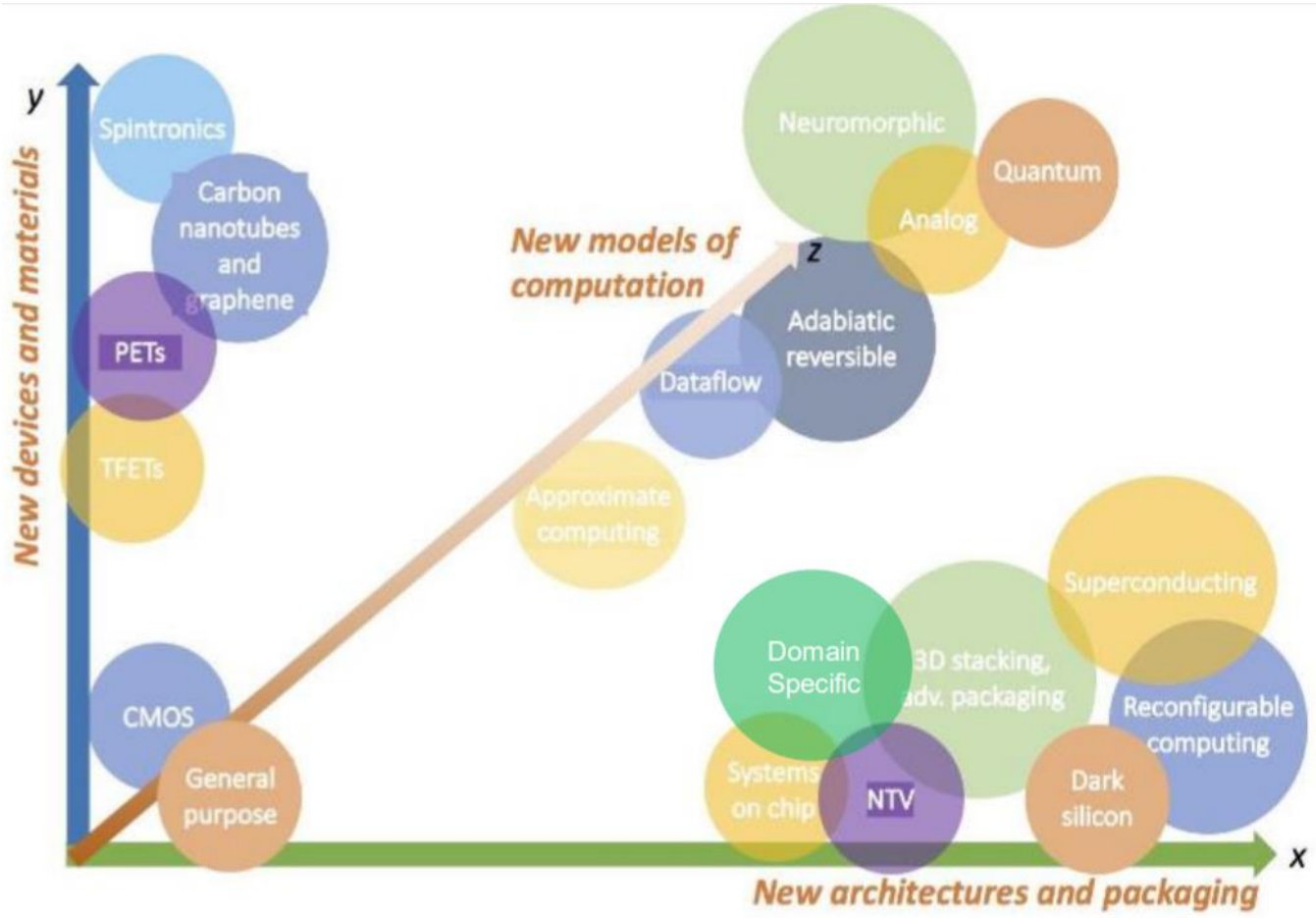
Outline of the Content

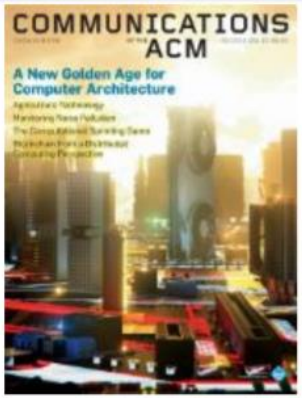
2017 Turing Award

John L. Hennessy　　David A. Patterson

计算机体系结构进入新的黄金时代！

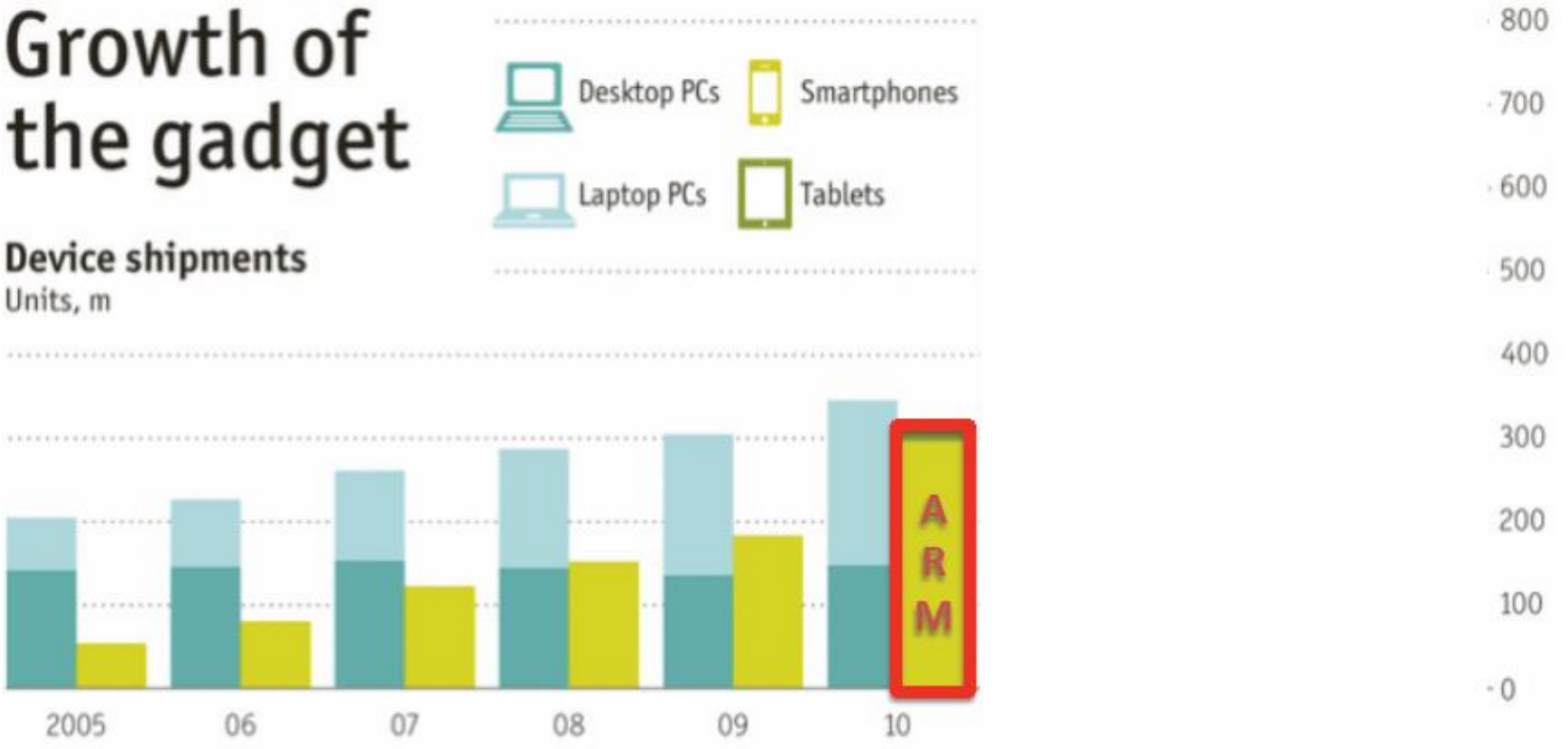
A New Golden Age for Computer Architecture





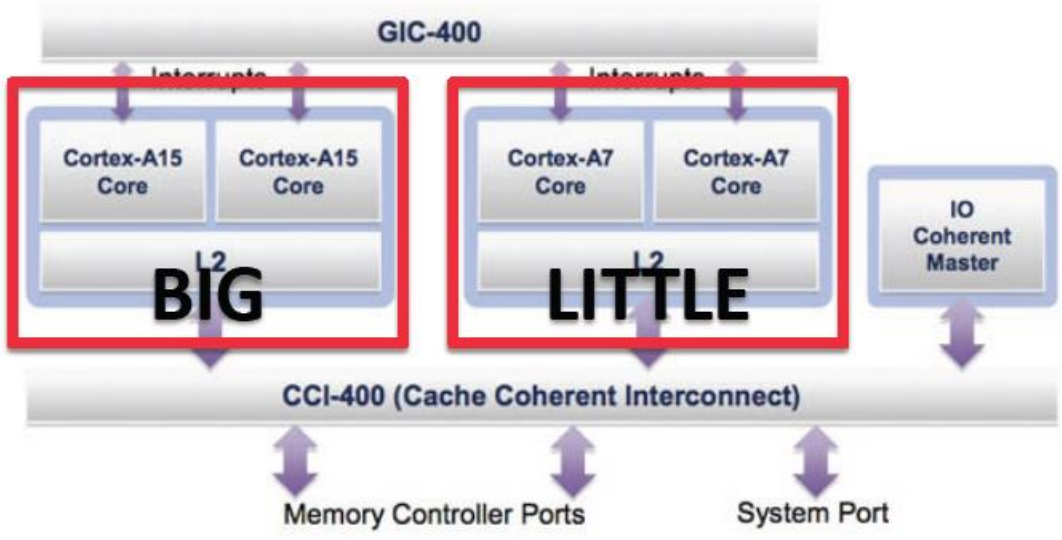
Who Knows What ARM is?

- Q: How many of you have an ARM computer?
  - A: **All of you**



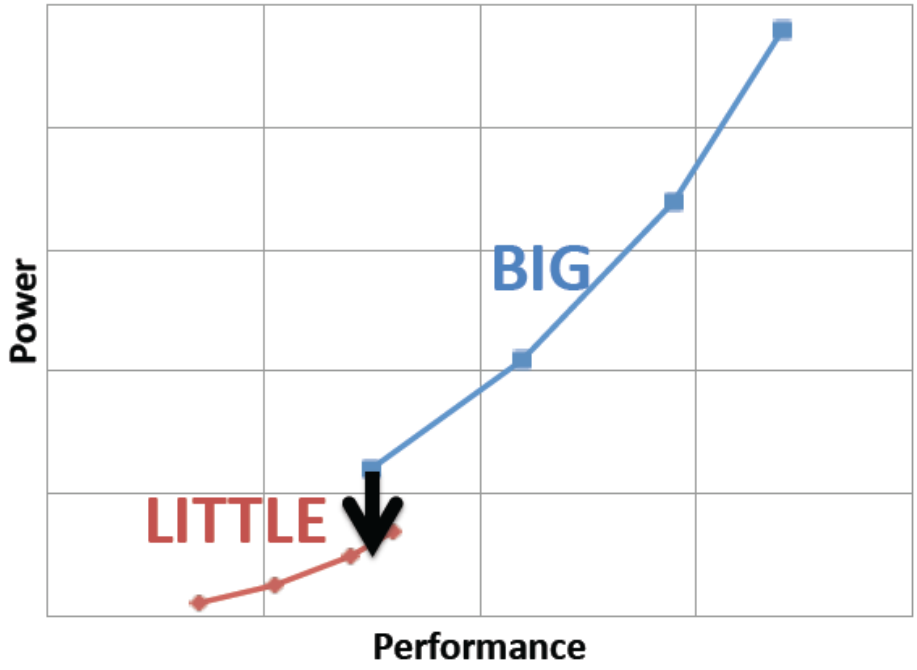
What Exactly Are They Doing?

- What is big.LITTLE?
- **Big cores** for high performance
- **Little cores** for low power



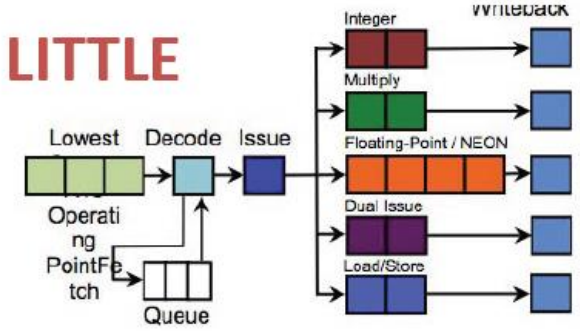
Why is ARM Doing This?

- **Power Efficiency** = calculations/energy



The Details

LITTLE

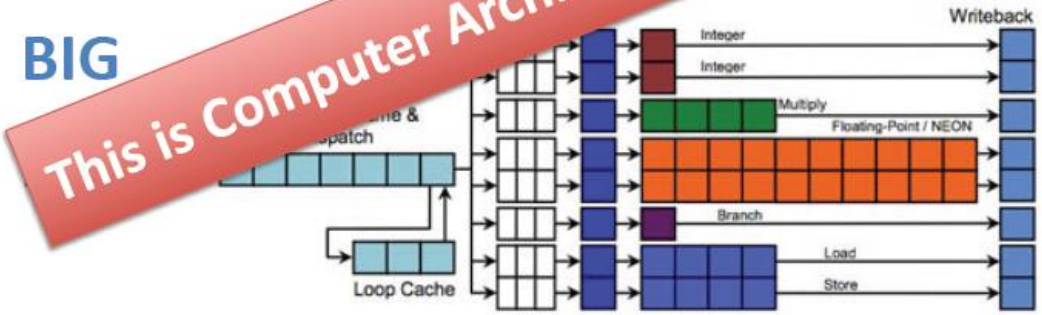


- Simple (→ fewer functional units)
- Short pipeline (→ slower clock)

BIG

- Complex
  - More functional units
  - Out of order execution
- Long pipeline
  - Faster clock
  - Bigger branch penalty

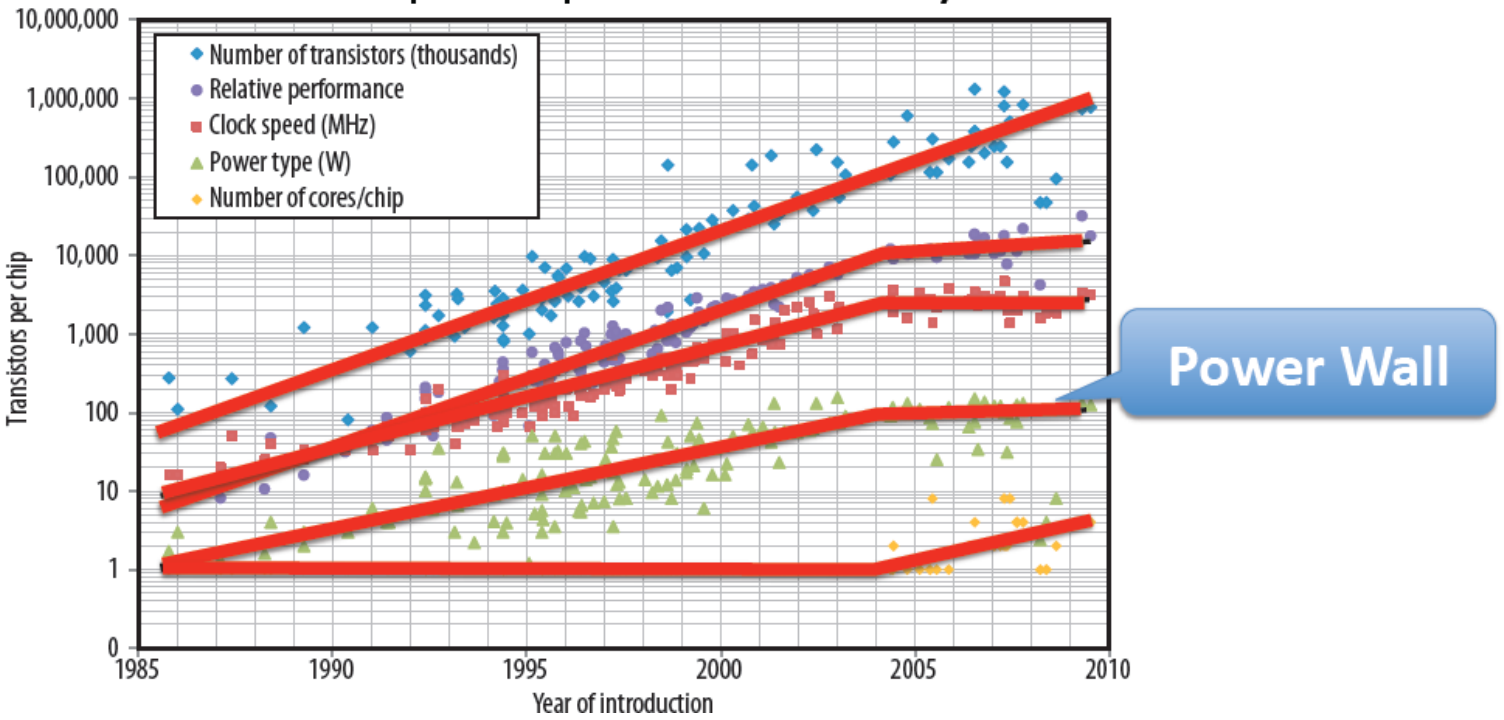
BIG



This is Computer Architecture!

Why Should They Do This?

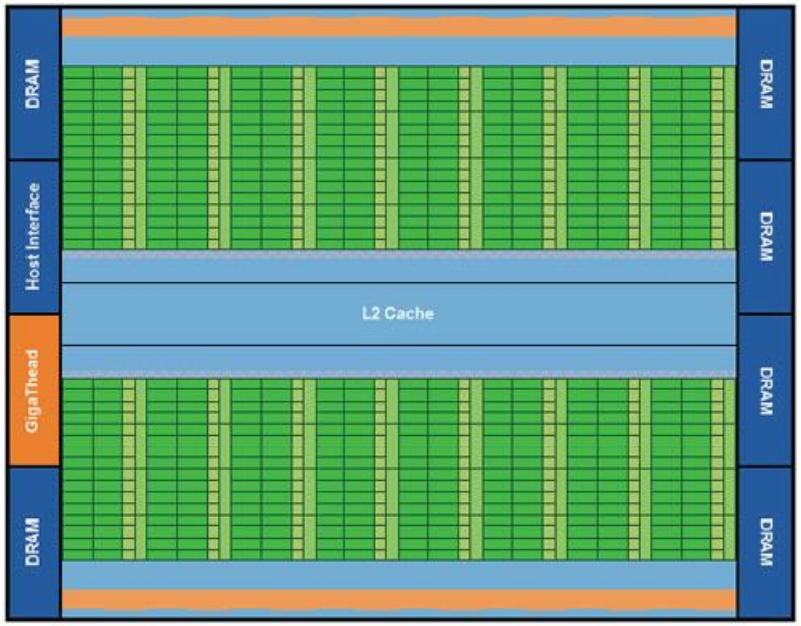
- Can't increase power:
  Need to improve power efficiency



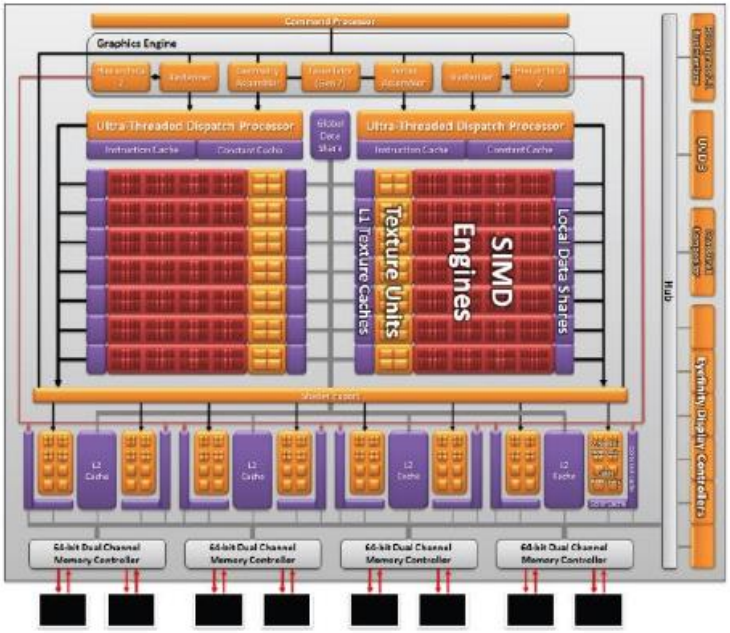
What are Others Doing?

- GPUs: Lots and lots of *very* small cores

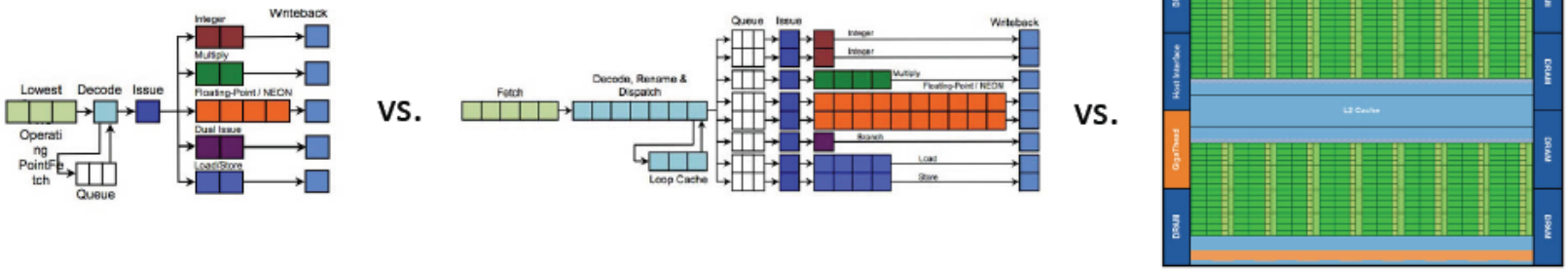
Nvidia Fermi



AMD Barts



This is Computer Architecture



- Understanding **performance and efficiency**
- **Design tradeoffs** in executing instructions
- Building the **hardware**
- Making it **programmable**

So, Why Should You Care?

- Computers are **evolving very fast**
- Need to understand how they work to understand why they are changing

- Computer Architecture is critical to performance and efficiency
- **Not just about designing hardware:**
  - How does big.LITTLE affect software?
  - How easy is it to program a GPU?

Example-1

How to select a proper processor?

What are your concerns?

- Price?
- Performance?
- Reliability?
- CPU, Memory, Storage, GPU?

…

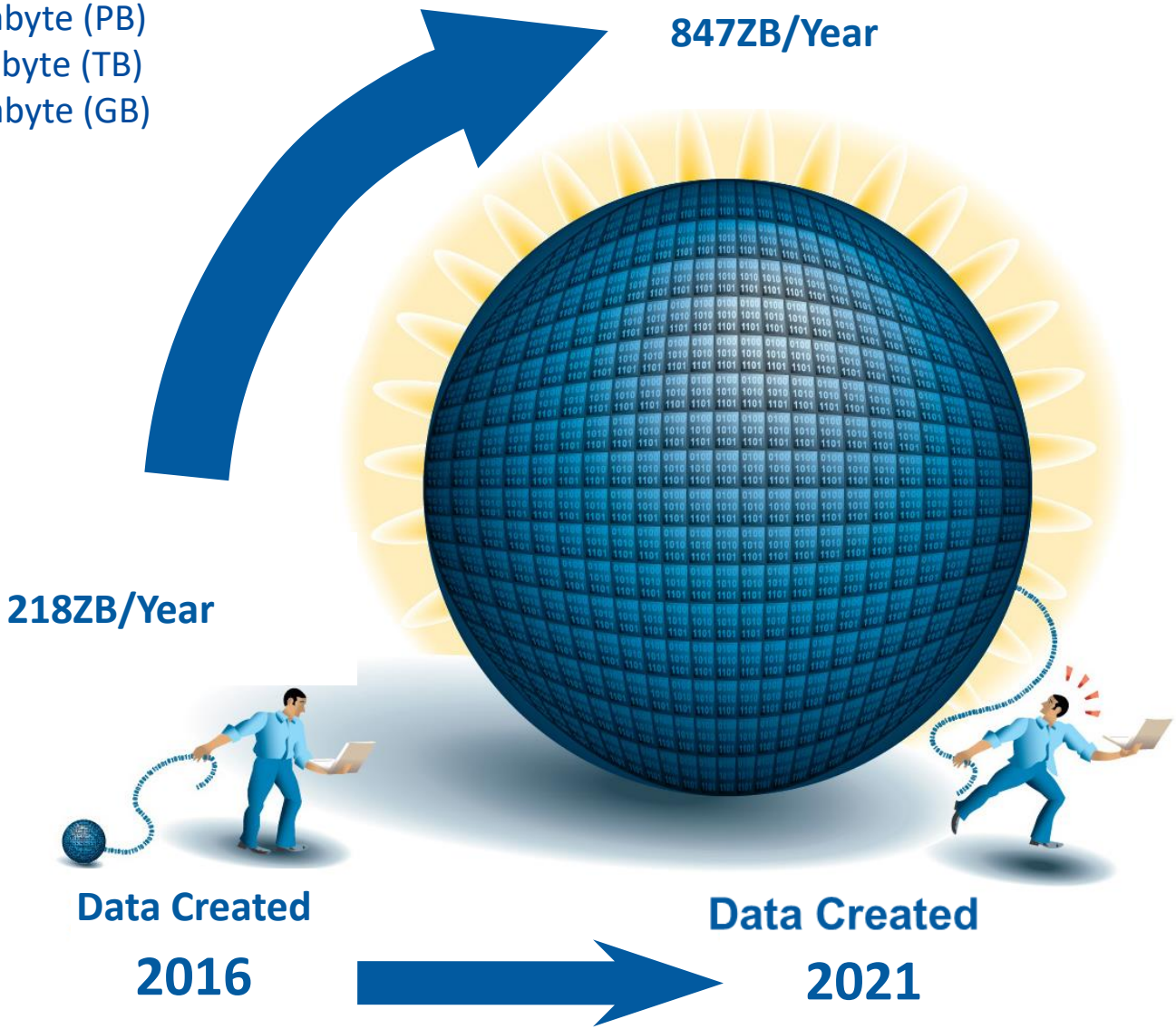
01 Introduction » Background

排名	研制厂商/单位	型号	安装地点	安装年份	应用领域	CPU核数	Linpack值(Tflops)	Linpack来源	Linpack峰值(Tflops)
1	国家并行计算机工程技术研究中心	神威太湖之光,40960*Sunway SW26010 260C 1.45GHz,自主网络	国家超级计算无锡中心	2016	超算中心	10649600	93015.0	Q	125436.0
2	国防科大	天河二号升级系统(Tianhe-2A),TH-IVB-MTX Cluster + 35584*Intel Xeon E5-2692v2 12C 2.2GHz+35584*Matrix-2000,TH Express-2	国家超级计算广州中心	2017	超算中心	427008	61445.0	Q	100679.0
3	联想	深腾8800系列,3800*Intel Xeon Gold 6133 20C 2.5GHz,25GbE	网络公司	2019	互联网/大数据	76000	3089.0	C	6080.0
4	联想	深腾8800系列,3680*Intel Xeon Gold 6133 20C 2.5GHz,25GbE	网络公司	2019	互联网/大数据	73600	2994.0	C	5888.0
5	联想	深腾8800系列,3600*Intel Xeon Gold 6133 20C 2.5GHz,25GbE	网络公司	2019	互联网/云计算	72000	2932.0	C	5760.0
6	联想	深腾8800系列,3400*Intel Xeon Gold 6133 20C 2.5GHz,25GbE	网络公司	2019	互联网/视频	68000	2780.0	C	5440.0
7	联想	深腾8800系列,612*Intel Xeon Gold 6240 18C 2.6GHz,EDR	信息中心	2019	科学计算	11016	2745.0	C	5201.0
8	联想	深腾8800系列,3200*Intel Xeon Gold 6133 20C 2.5GHz,25GbE	网络公司	2019	互联网/云计算	64000	2621.0	C	5120.0
9	国防科大	天河一号A,14336*Intel Hexa Core Xeon X5670 6C 2.93GHz,私有高速网络80Gb	国家超级计算天津中心	2010	超算中心	202752	2566.0	Q	4701.0
10	国家并行计算机工程技术研究中心	神威E级原型原型系统,1024*SW26010+ 260C 1.5GHz,SW-Net	国家超级计算济南中心	2018	超算中心	266240	2556.0	C	3130.0


13000+个
App下载
13.6万条
垃圾短信拦截
8万
净收入
152台
PC售出
iOS
25部
176部
13.9万张
照片上传
8.5万元
收入
2.8万
净收入
中国电信
7.3万笔
交易数
（2012双11）
SOLD OUT!!!
24.6万
净收入
中国移动
486笔
订单
IN 60 SECONDS..
1.4万
净收入
中国联通
1042笔
订单
（2012双11）
Add to Cart
12306
265万张
订单
10笔
订单
搜狐IT
it.sohu.com
1.25万
浏览人数
14.8万
独立访客
460万
网页打开
347万次
搜索
9.5万条
微博发送
1.1万GB
文件下载
搜狐IT
it.sohu.com
@iPhone一姐

Data Explosion

1 ZB = 1,024 Exabyte (EB)
1 EB = 1,024 Petabyte (PB)
1 PB = 1,024 Terabyte (TB)
1 TB = 1,024 Gigabyte (GB)



[Source: Cisco Global Cloud Index, 2016-2021]

Review: Some Basic Definitions

Metric prefixes

Prefix	Symbol	1000^m	10^n	Decimal	Short scale	Long scale	Since[n 1]
yotta	Y	1000^8	10^{24}	1 000 000 000 000 000 000 000 000	septillion	quadrillion	1991
zetta	Z	1000^7	10^{21}	1 000 000 000 000 000 000 000	sextillion	trilliard	1991
exa	E	1000^6	10^{18}	1 000 000 000 000 000 000	quintillion	trillion	1975
peta	P	1000^5	10^{15}	1 000 000 000 000 000	quadrillion	billiard	1975
tera	T	1000^4	10^{12}	1 000 000 000 000	trillion	billion	1960
giga	G	1000^3	10^9	1 000 000 000	billion	milliard	1960
mega	M	1000^2	10^6	1 000 000	million		1960
kilo	k	1000^1	10^3	1 000	thousand		1795
hecto	h	$1000^{2/3}$	10^2	100	hundred		1795
deca	da	$1000^{1/3}$	10^1	10	ten		1795
		1000^0	10^0	1	one		–

Review: Some Basic Definitions

Metric prefixes

Prefix	Symbol	1000^m	10^n	Decimal	Short scale	Long scale	Since[n 1]
		1000^0	10^0	1	one		–
deci	d	$1000^{-1/3}$	10^{-1}	0.1	tenth		1795
centi	c	$1000^{-2/3}$	10^{-2}	0.01	hundredth		1795
milli	m	1000^{-1}	10^{-3}	0.001	thousandth		1795
micro	μ	1000^{-2}	10^{-6}	0.000 001	millionth		1960
nano	n	1000^{-3}	10^{-9}	0.000 000 001	billionth	milliardth	1960
pico	p	1000^{-4}	10^{-12}	0.000 000 000 001	trillionth	billionth	1960
femto	f	1000^{-5}	10^{-15}	0.000 000 000 000 001	quadrillionth	billiardth	1964
atto	a	1000^{-6}	10^{-18}	0.000 000 000 000 000 001	quintillionth	trillionth	1964
zepto	z	1000^{-7}	10^{-21}	0.000 000 000 000 000 000 001	sextillionth	trilliardth	1991
yocto	y	1000^{-8}	10^{-24}	0.000 000 000 000 000 000 000 001	septillionth	quadrillionth	1991

1. ^ The metric system was introduced in 1795 with six prefixes. The other dates relate to recognition by a resolution of the CGPM.

Review: Some Basic Definitions

- Kilobyte – 2^{10} or 1,024 bytes
- Megabyte– 2^{20} or 1, 024 Kilobytes
  - sometimes “rounded” to 10^6 or 1,000,000 bytes
- Gigabyte – 2^{30} or 1, 024 Megabytes
  - sometimes rounded to 10^9 or 1,000,000,000 bytes
- Terabyte – 2^{40} or 1, 024 Gigabytes
  - sometimes rounded to 10^{12} or 1,000,000,000,000 bytes
- Petabyte – 2^{50} or 1024 Terabytes
  - sometimes rounded to 10^{15} or 1,000,000,000,000,000 bytes
- Exabyte – 2^{60} or 1024 Petabytes
  - Sometimes rounded to 10^{18} or 1,000,000,000,000,000,000 bytes
- Zettabyte – 2^{70} or 1024 Exabytes
  - Sometimes rounded to 10^{21} or 1,000,000,000,000,000,000, 000 bytes

01 Introduction » Course Content

How a computer is built

– Logic->circuits->**datapath**

- **How a computer is controlled**

– Basic operations->micro architecture->instructions (ISA)->**assembly**

- **Contents (in-order)**

– MIPS assembly
– Logic design(adders,ALU,control)
– Performance analysis
– Data path and pipelining
– Input/Output
– Caches
– Virtual memory

01 Introduction » Educational Goals

- Understand how modern computers are designed and implemented
  - computer architecture principles
  - design and implementation trade-offs
    - performance, cost, energy, design complexity
  - role of technology and market trends
    - not just microprocessor, but also servers and embedded systems
    - the arrival of multi-core architectures
  - case studies
    - not only classic processors, but also state-of-the art computers
- "SW-oriented students"
  - write better software applications
  - write better system's software
    - operating systems, compilers

01 Introduction » Prerequisites

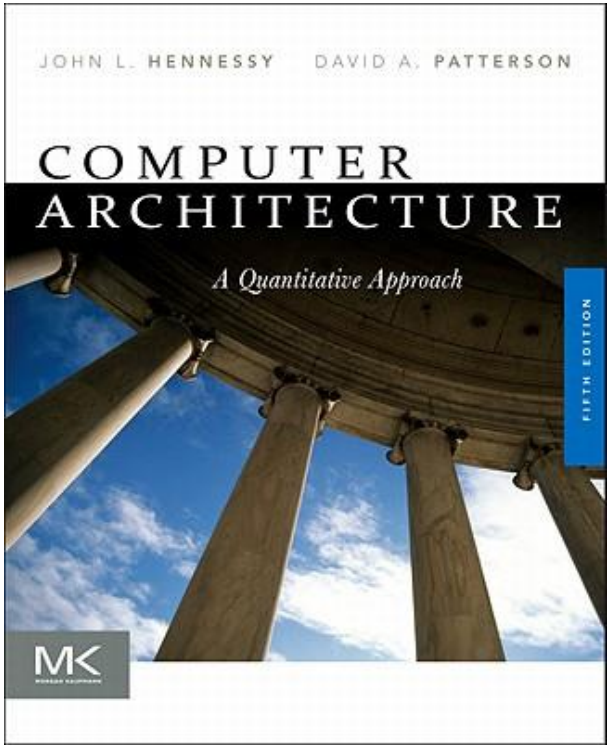
- Course in computer organization
  - processor organization, instruction set architecture
  - programming with an assembly language
  - ideally, basics of pipelining and caches
- Course on digital logic
  - logic gates, truth tables, Karnaugh maps
  - combinational vs. sequential logic, FSMs
- Background in programming and data structure
  - background in at least one high-level language
  - programming (C, C++, Java…)

01 Introduction » Major parts

- Introduction
  - Chapters 1.1-1.6
- Principles of Quantitative Analysis; MIPS
  - Chapters 1.8-1.12
- Pipelining, Hazards, Implementation Issues
  - Chapters A.1-A.7, A.9, A.12, C.1-C.4
- Instruction-Level Parallelism (ILP), Tomasulo
  - Chapters C.5-C.9, 3.1, 3.2, 3.7
- Memory-Hierarchy Design, Cache Optimization
  - Chapters B.1-B.3, 2.1, 2.3
- ILP: Branch Prediction, Superscalars
  - Chapters 3.3-3.6, 3.8, 3.9
- Memory Hierarchy Design, Virtual Memory
  - Chapters 2.4, B.4
- Mostly covered by the textbook Appendix A, B, C and Chapter 1, 2, 3

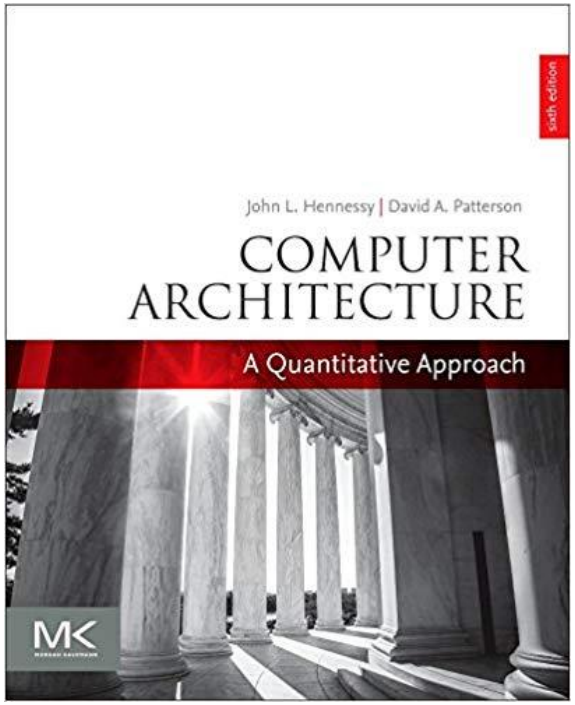
01 Introduction » Required Textbook

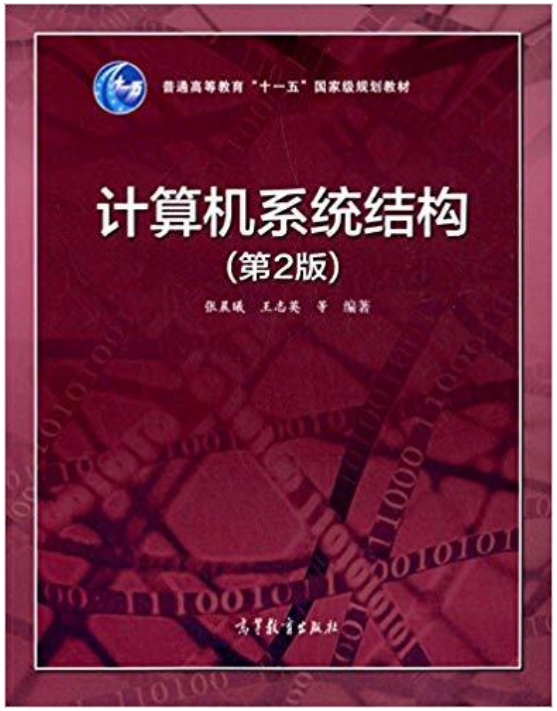
- Computer Architecture: A Quantitative Approach, 5th ed.," John L. Hennessey and David A. Patterson, Morgan Kaufman, 2011

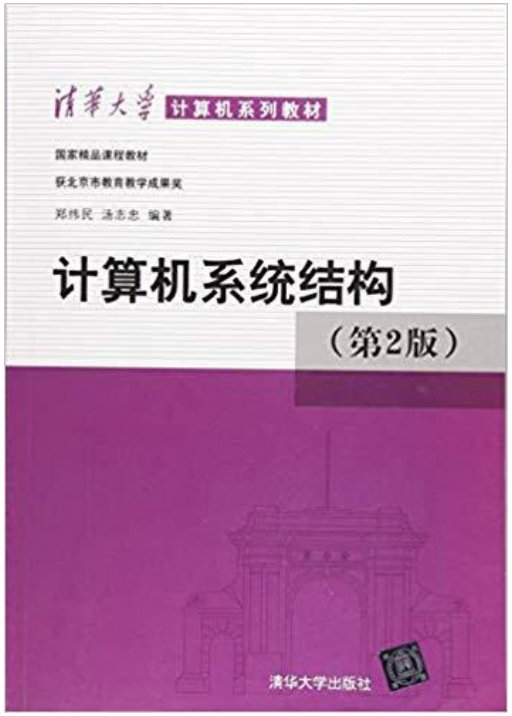


- Additional readings will be assigned in class:
  - journal and conference papers
  - excerpts from other books
  - technical manuals

01 Introduction » Required Textbook

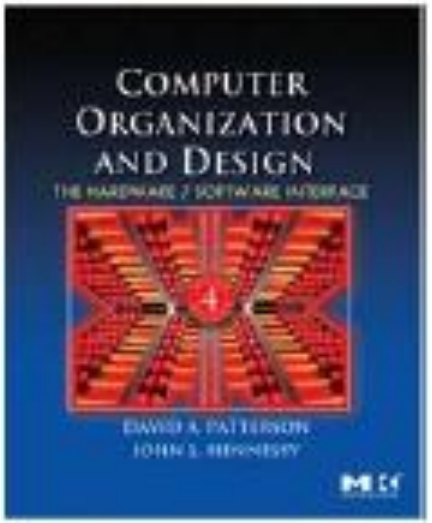






01 Introduction » Optional Textbooks

- D. Patterson and J. Hennessy, "Computer Organization & Design,"
  Morgan Kaufmann, 2008



- D. Harris and S. Harris, "Digital Design and Computer Architecture,"
  Morgan-Kaufmann, 2007

01 Introduction » Grading & Attendance

- **Grades** will be based on
  - Homework (about 15%)
  - In class quizzes (about 10%)
  - In class presentation (about 5%)
  - Midterm exam (about 10%)
  - Project / Experiments (about 20%)
  - Final exam (about 40%)
- Any material covered in class, assigned readings, handouts may appear in exam questions
- **Class Attendance:**
  - You are responsible for all material presented in class lectures
  - At times, lectures will diverge from the textbook
  - Regular class attendance is the best way to insure that you learn the material

01 Introduction » Grading & Policies

- Homework is due at the beginning of class on the due date

- Late policy
  - If you hand in something after the due date without the explicit written approval of an instructor or course TA, you might receive zero credit. In emergency only, contact the instructor in advance for an extension.

- General policy
  - Collaboration on solutions, or sharing or copying of solutions, is not allowed. No cooperation is allowed on exams. **This policy will be strictly enforced**

Office Hour

- Welcome to come to my office during the office hours:

Office: Main Building 0626
Email: liuduo@cqu.edu.cn

Office hours: Wednesday 3-5pm
or by appointment (send email)

Teaching Assistant of Prof. Duo Liu

- Haorui He (何浩睿)
  - Office: A区主教0414
  - Email: hehaorui@outlook.com
  - Phone: 15182706383
  - Office Hours
    - Saturday 3:00—5:00pm
- Mingfeng Jiang (蒋铭峰)
  - Office: A区主教0600
  - Email: 942513309@qq.com
  - Phone: 15320363541
  - Office Hours
    - Saturday 3:00—5:00pm