Computer Architecture
(Spring 2020)

<u>Introduction</u>

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

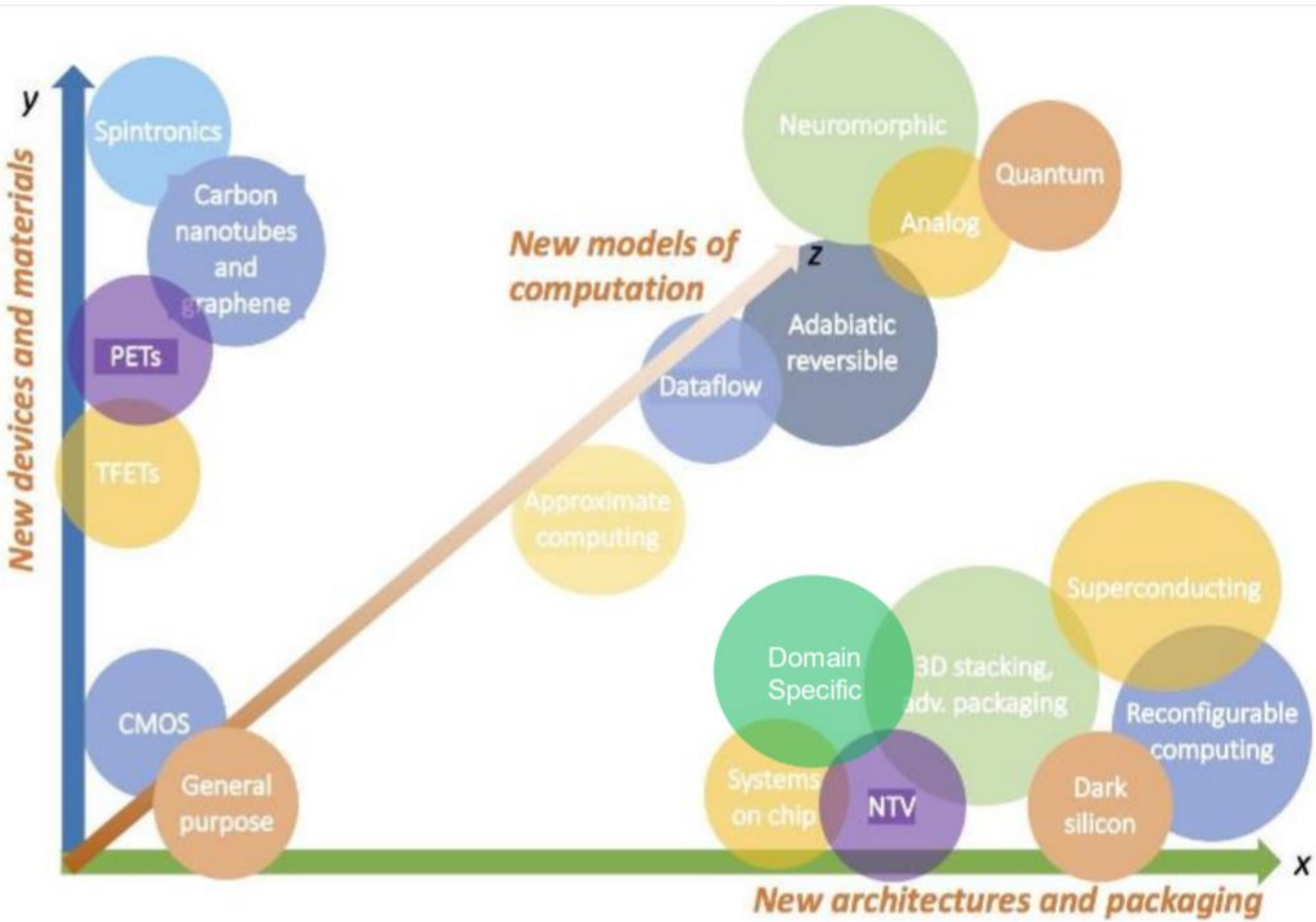
2017 Turing Award

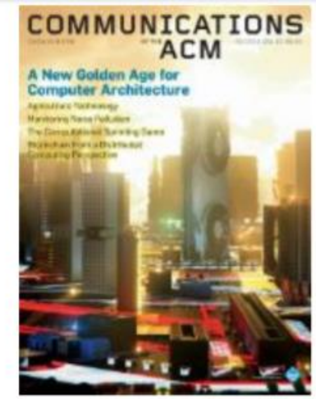
John L. Hennessy

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





About the Course

How a computer is built
– Logic->circuits->**datapath**
- **How a computer is controlled**
– Basic operations->micro architecture->instructions (ISA)->**assembly**

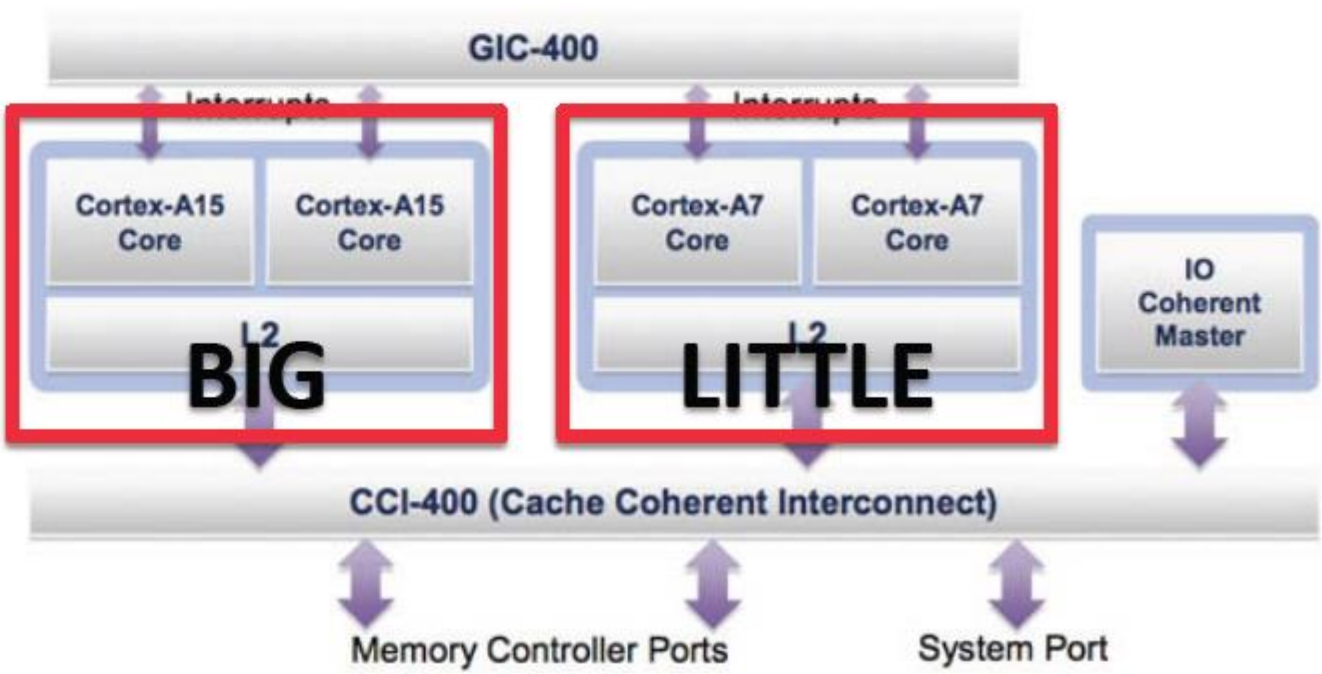
- **Contents (in-order)**
– MIPS assembly
– Logic design(adders,ALU,control)
– Performance analysis
– Data path and pipelining
– Input/Output
– Caches
– Virtual memory

Who Knows What ARM is?

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

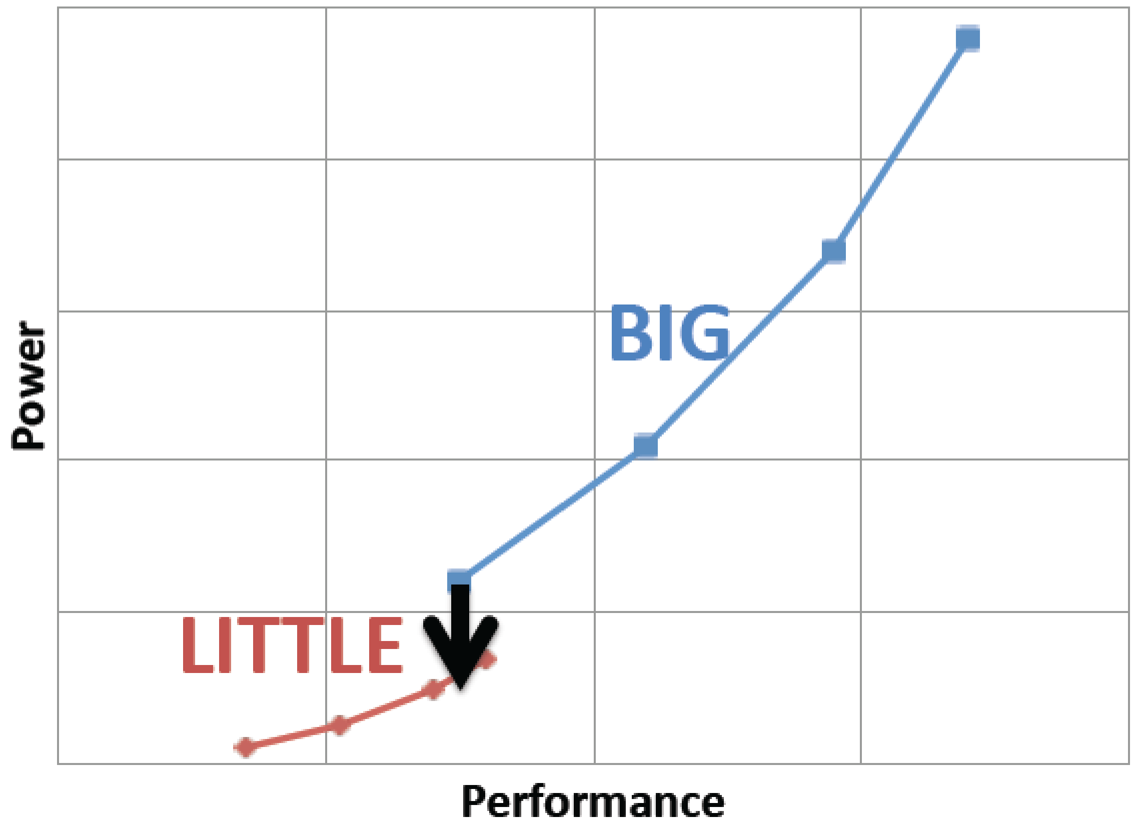
What Exactly Are They Doing?

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



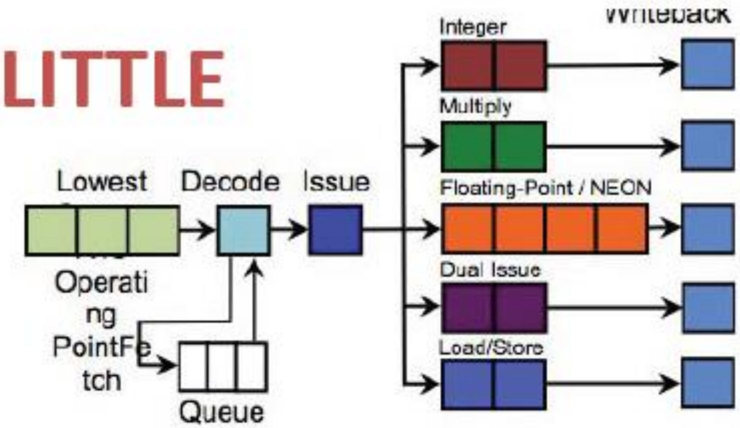
Why is ARM Doing This?

- **Power Efficiency** = calculations/energy

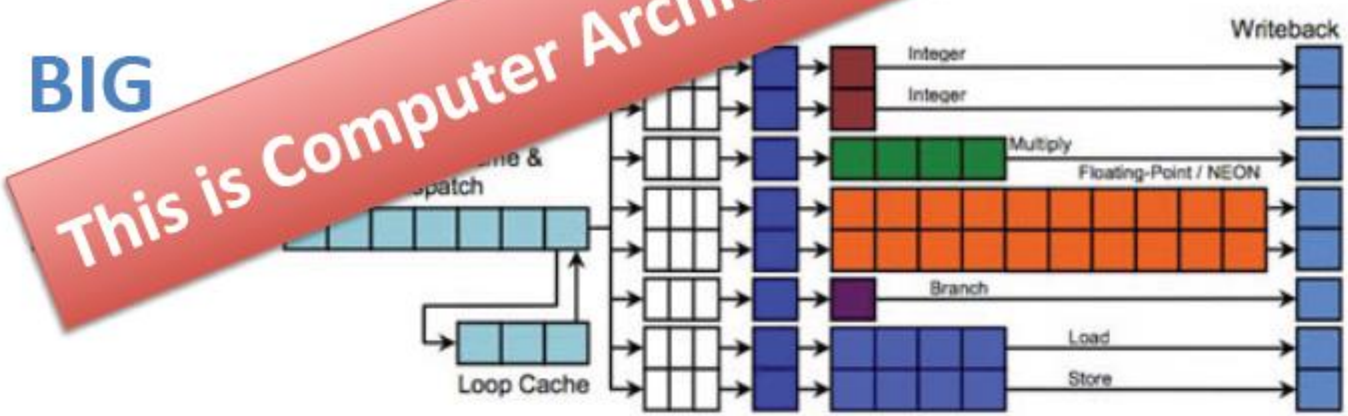


The Details

LITTLE



BIG



This is Computer Architecture!

LITTLE

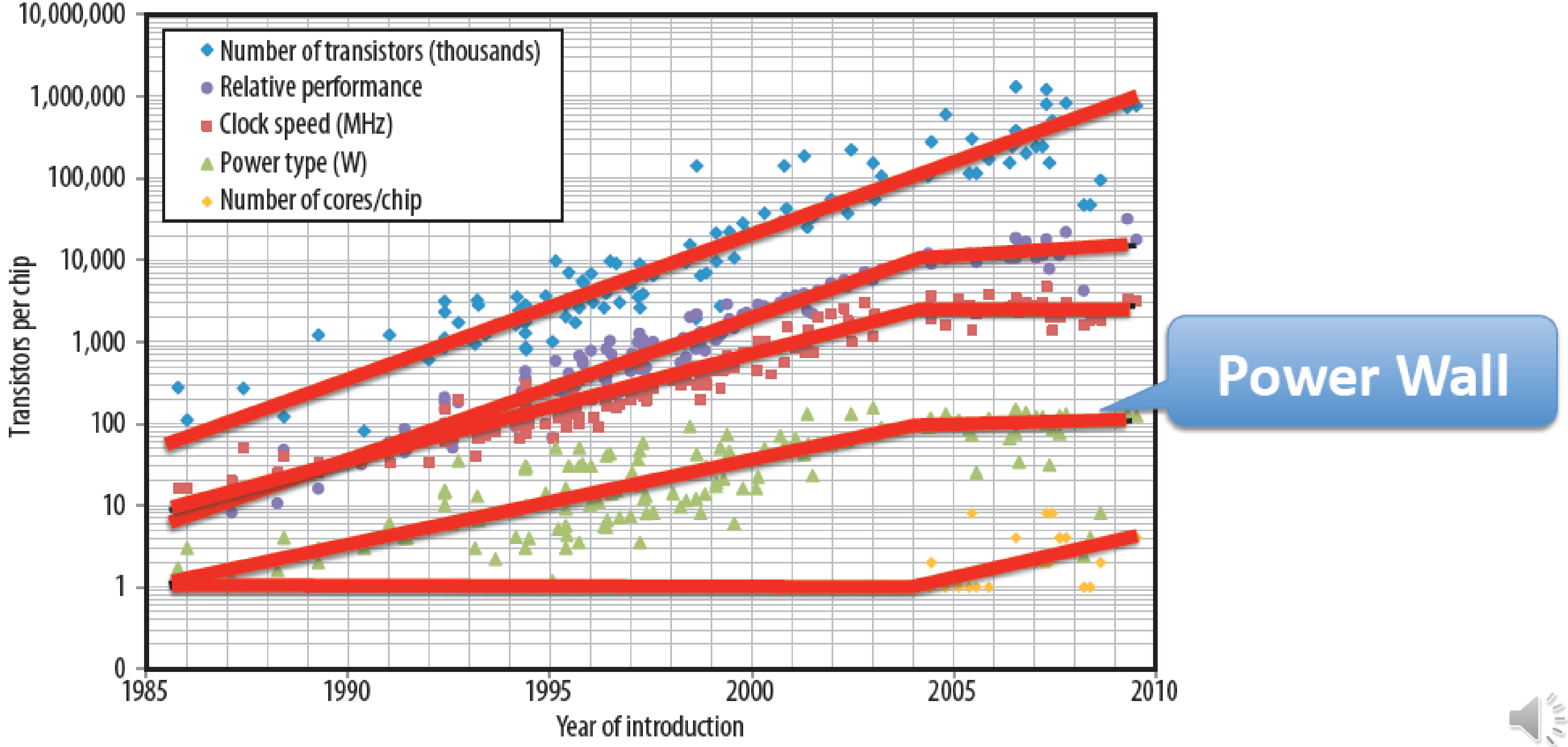
- Simple (→ fewer functional units)
- Short pipeline (→ slower clock)

BIG

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

Why **Should** They Do This?

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

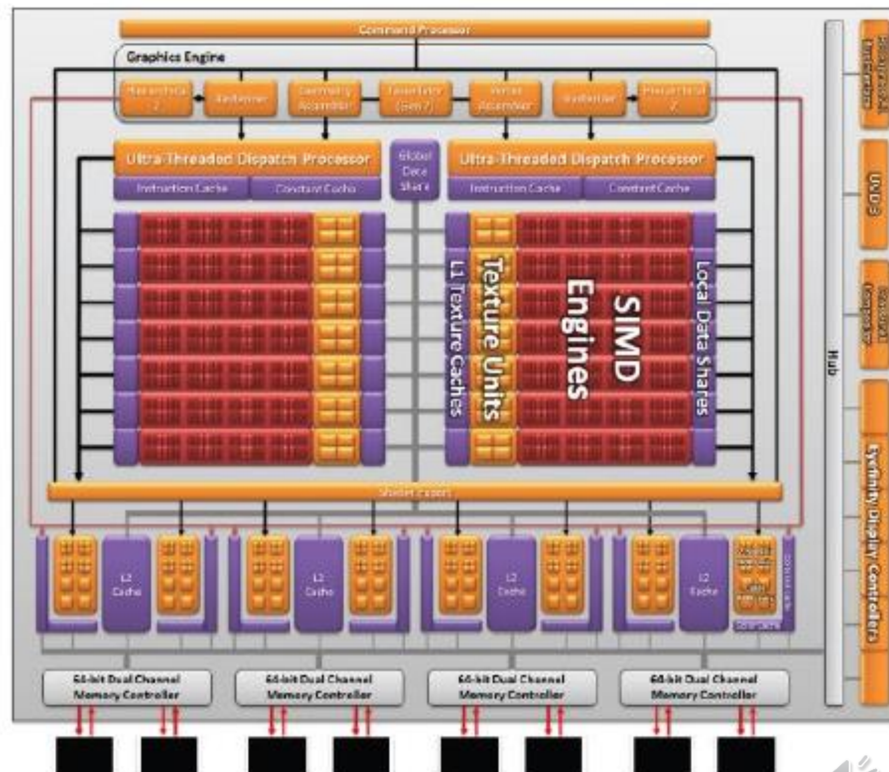


What are Others Doing?

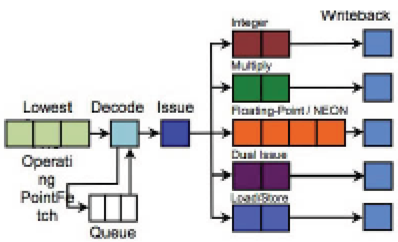
- GPUs: Lots and lots of *very* small cores

Nvidia Fermi

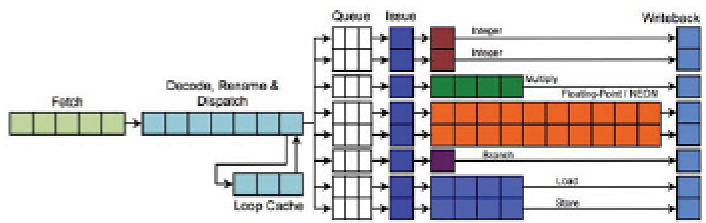
AMD Barts



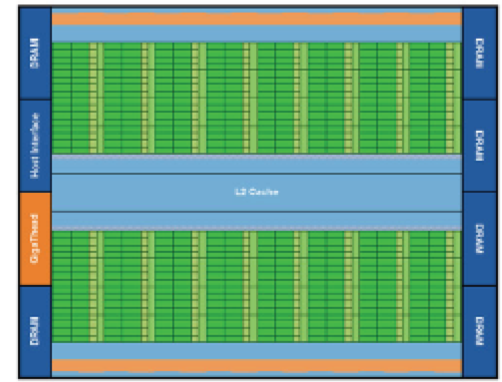
This is Computer Architecture



vs.



vs.



- 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?

The more you learn about the computer architecture, the higher efficiency you can achieve!

Example-1

How to select a proper processor?

What are your concerns?

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

…

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

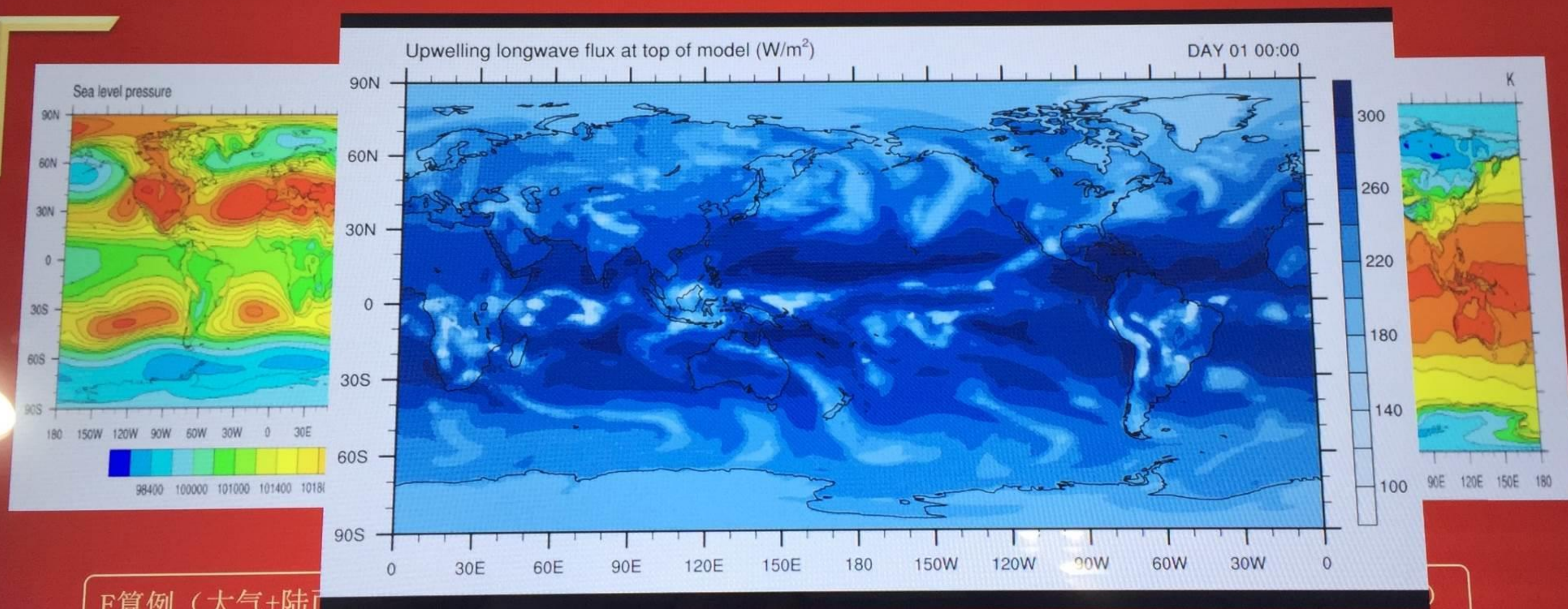
“神威·太湖之光”应用课题列表

序号	领域/课题名称		承担单位	规模
1	气象	全球大气非静力云分辨模拟	中科院软件所、清华大学等	整机
2	海洋	高分辨率海浪数值模式	国家海洋局海洋一所	整机
3	材料	钛合金微结构演化相场模拟	中科院计算机网络信息中心	整机
4	生物	骆驼基因组功能注释及同源基因家族构建	上海科学院上海生物信息技术研究中心	整机
5	航天	航天飞行器全流域数值模拟	国家计算流体力学实验室	整机
6	核聚变	托克马克磁约束核聚变装置中的逃逸电子轨迹模拟	中国科技大学核学院	整机
7	航天	可压缩边界层湍流及转捩模拟	中科院力学所	半机
8	生物	蛋白质分子动力学模拟	北京大学	半机
9	电磁	FDTD时域有限差分法电磁模拟软件	西安电子科技大学	半机
10	电磁	PO物理光学法电磁模拟软件	西安电子科技大学	半机
11	船舶	潜艇声学特性模拟计算	中船重工集团702所	半机
12	船舶	大型海上浮动平台波浪载荷计算	中船重工集团702所	半机
13	动漫	真实感动漫渲染	山东大学	半机
14	核模拟	低能段GMNT软件	中科院近代物理研究所	半机
15	生物	测序数据精确对比	山东大学	半机
16	物理	Crystal-MD分子动力学模拟	北京科技大学	半机
17	能源	三维波动方程正演模拟	中国地质大学（武汉）	半机
18	航空	湍流的直接数值模拟	北京大学	半机
19	航天	高超声速内外流数值模拟	中国空气动力研究发展中心	32768核组
20	气象	GRAPES全球数值预报模式	中国气象局	16384核组
21	气候	地球系统模式	清华大学、北京师范大学	16384核
22	航空	航空飞行器模拟程序	清华大学航空航天学院	16384核
23	航空	飞机、高速列车流场分析	中科院深圳先进技术研究院	10000核
24	气象	城市污染扩散	中科院深圳先进技术研究院	10000核
25	生命科学	脑血流模拟	中科院深圳先进技术研究院	10000
26	生物	NUMD分子动力学计算	上海药物所	8192核
27	航空	[illegible]飞机外型全流场优化设计	631所	8192
28	航空	大型客机减阻与激波控制	南京航空航天大学	8192
29	航空	航空发动机高精度数值模拟	浙江大学	8192
30	能源	风电场设计评估云服务平台	远景能源公司	409

地球系统模式

清华大学地学中心

国内首次实现**百万核**规模的全球**10公里**高分辨率地球系统模式数值模拟，全面提高我国应对极端气候事件和自然灾害的减灾防灾能力，为我国参与国际气候环境外交谈判，提升我国在气候变化领域国际影响力提供重要基础

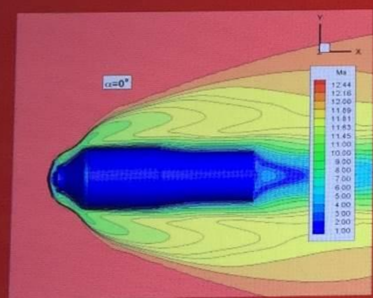


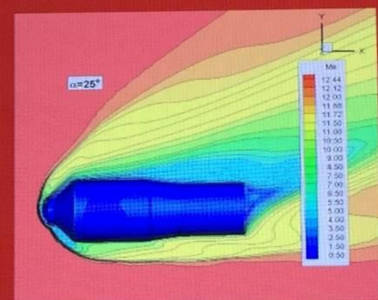
F算例（大气+陆

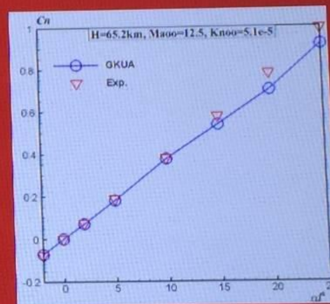
航天飞行器统一算法数值模拟

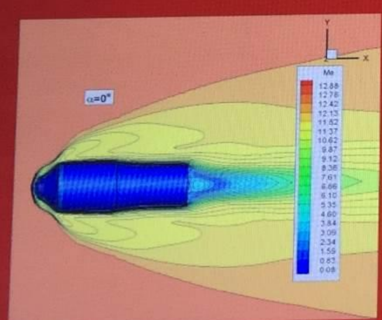
国家计算流体力学实验室

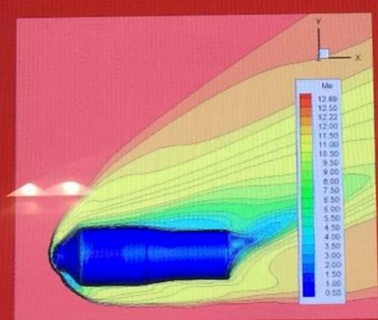
完成**天宫-1**飞行器两舱简化外形（长度10余米、横截面直径近3.5米）陨落飞行H=65km、62km、 Ma=13绕流状态大规模并行计算，使用16384个处理器在**20天**内完成常规需要**12个月**的计算任务，计算结果与风洞实验结果符合较好。

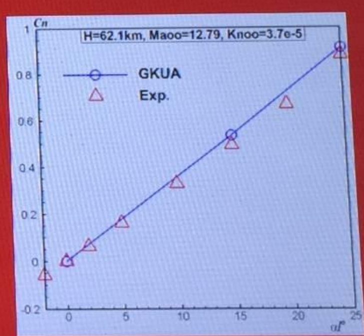













天河
天河二号
TH-2 High Performance Computer System

Course Information – 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

Course Information – Prerequisites

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

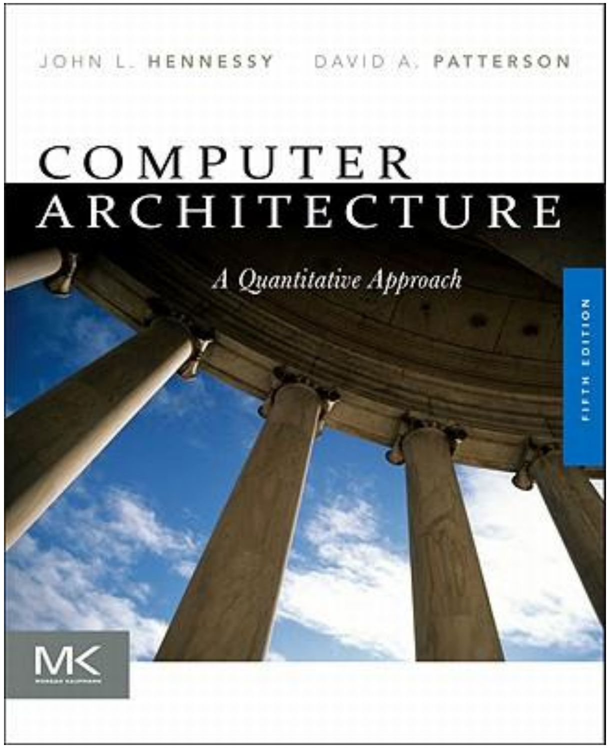
Course Information – Major parts

- 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

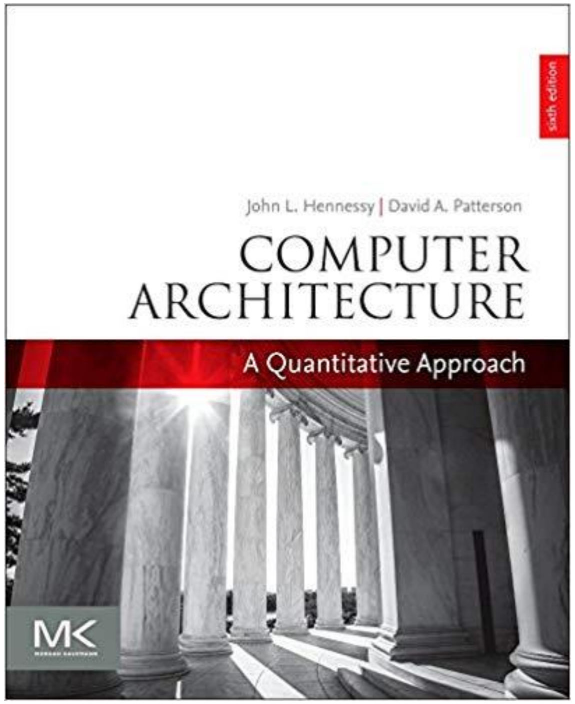
Course Information – Required Textbook

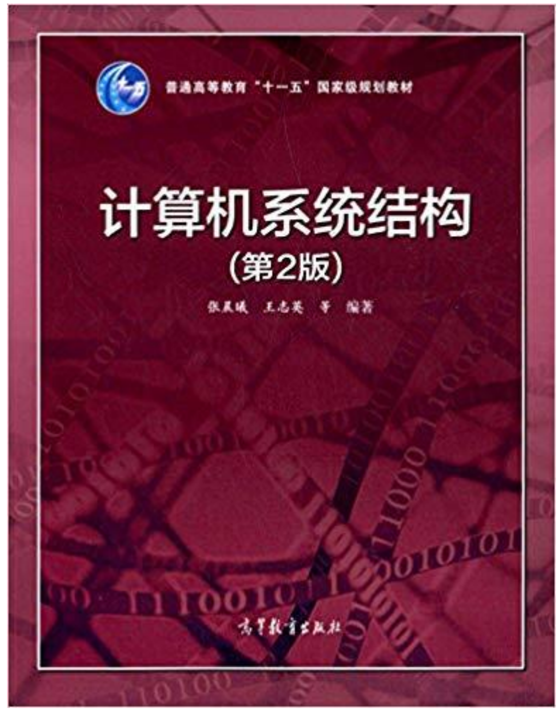
- Computer Architecture: A Quantitative Approach, 5th ed.," John L. Hennessey and David A. Patterson, Morgan Kaufman, 2011

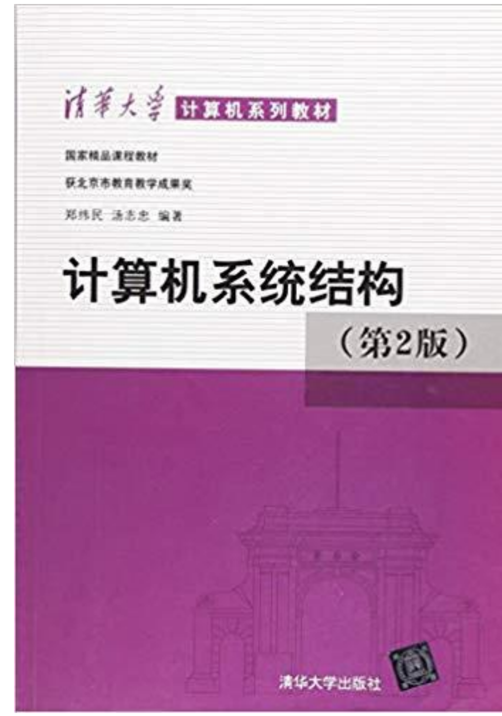
- Additional readings will be assigned in class:
  - journal and conference papers
  - excerpts from other books
  - technical manuals



Course Information – Required Textbook

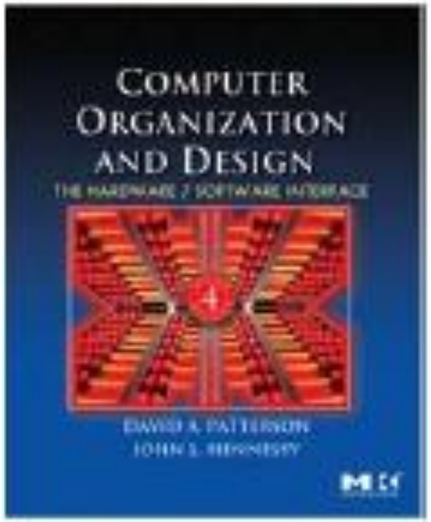






Course Information - Optional Textbooks

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



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

Course Information - Grading & Attendance

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

Course Information - 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

- Longpan Luo(罗龙攀)
  - Office: A区主教0600
  - Email: 793345571@qq.com
  - Phone: 15213408742
  - Office Hours
    - Saturday 3:00—5:00pm