



并行与分布式程序设计

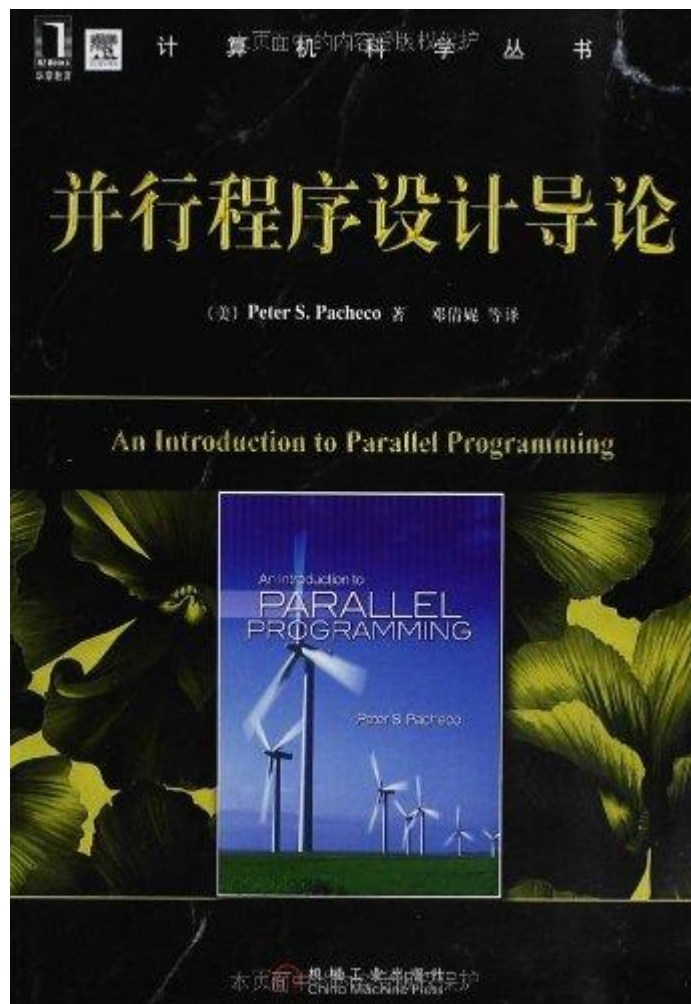
授课教师: 孙永谦

sunyongqian@nankai.edu.cn

参考王刚、任明明《并行程序设计》

参考教材

- 《并行程序设计导论》
机械工业出版社，2012
- 课件参考了犹他大学
cs4230课程
《并行程序设计》

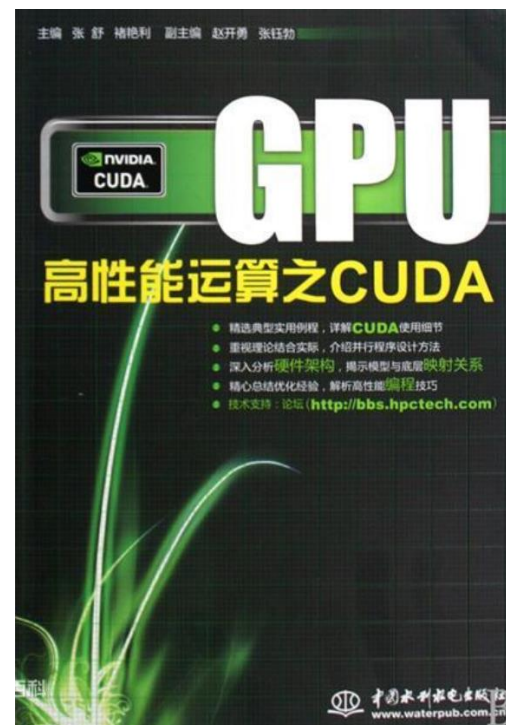


其他参考书

《并行计算导论（原书第2版）》，Ananth Grama等著，张武等译，机械工业出版社，2004

《MPI并行程序设计实例教程》，张武生等著，清华大学出版社，2009

《GPU高性能运算之CUDA》，张舒著，中国水利水电出版社，2009



沟通方式

○ 联系方式

□ 微信群:

□ qq群:



□ 孙永谦: sunyongqian@nankai.edu.cn

□ 实验室主页: <http://nkcs.iops.ai/>

○ 课件:

□ 主要在qq群里分享



课程要求

○ 先导课

- C/C++程序设计
- 计算机体系结构知识

○ 课程安排

- 理论课+实验课（8~16周）

○ 成绩

- 40%： 编程作业（不能抄袭，包括从网络）
书面作业
- 60%： 期末闭卷考试



课程概述

- 重要问题要求强大的计算机
 - 强大的计算机必然是并行机
 - 培养**并行编程人才**越来越重要
 - 一些并行程序员也应是**性能优化专家**
- 开发**高性能**并行应用
- 本课程涉及的并行架构
 - SSE/AVX、多核、GPU、集群



课程目标

- 学习并行系统上的编程
 - 用并行思维思考问题，编写**正确的**并行程序
 - 理解软件到并行架构的映射→实现**高性能和高伸缩性**
- 亲自动手获得编程经验
 - 在实际硬件上编写实际应用
 - 设计并行算法
- 讨论当前的并行计算环境
 - 新架构和编程模型，发展趋势



课程的重要性

- 多核、众核时代已经来临并将持续
 - 为什么？技术发展趋势、应用需求推动
- 很多程序员需要开发并行软件
 - 但仍有很多人并未接受并行编程训练
 - 学习如何充分利用并行计算资源
- 用处
 - 求职
 - 研究生学习
- 教学重点
 - 核心概念、常用编程模型、更广泛的内容



第1讲 绪论



提纲

- 推动并行计算的因素
- 并行计算的应用
- 超级计算机硬件的发展
- 软件技术面临的挑战
- 众核技术/GPU的发展

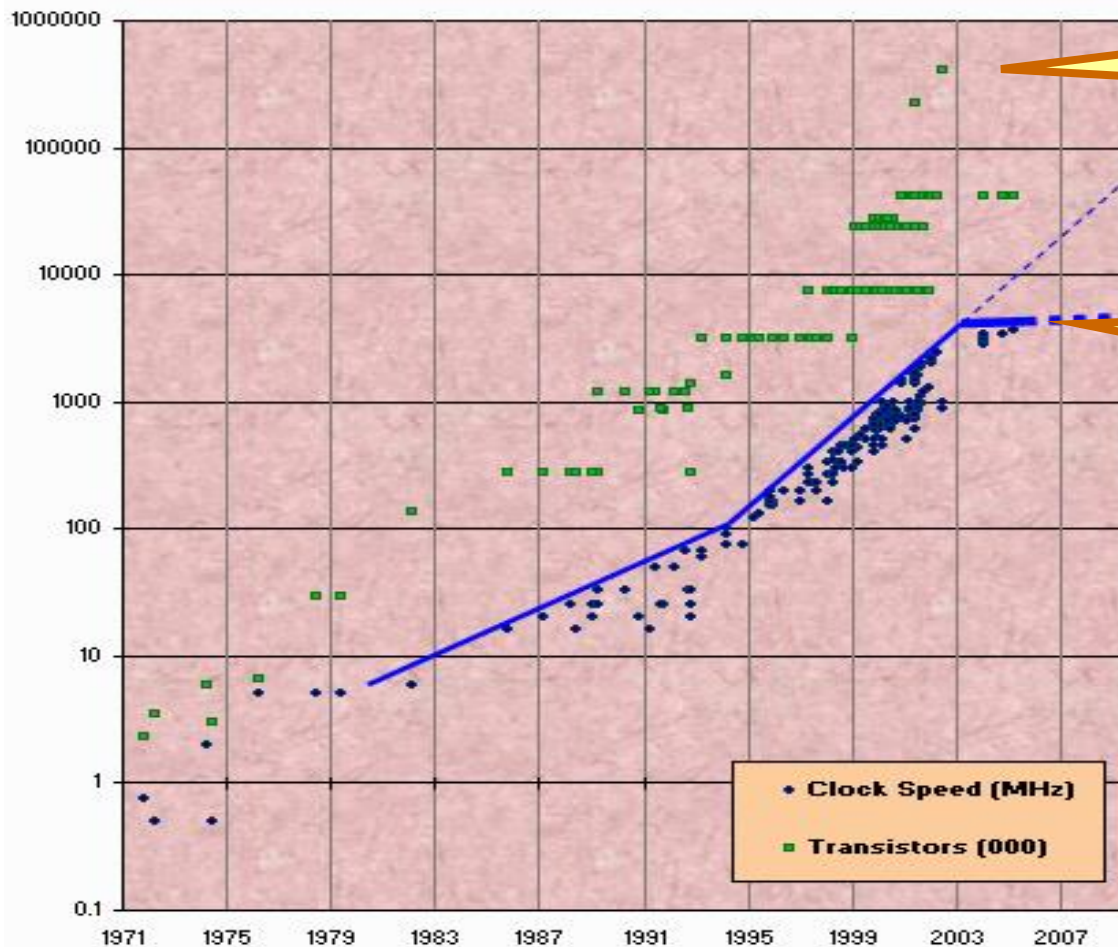


绪论提纲

- 推动并行计算的因素
- 并行计算的应用
- 超级计算机硬件的发展
- 软件技术面临的挑战
- 众核技术/GPU的发展

推动并行计算的因素

○ 处理器能力

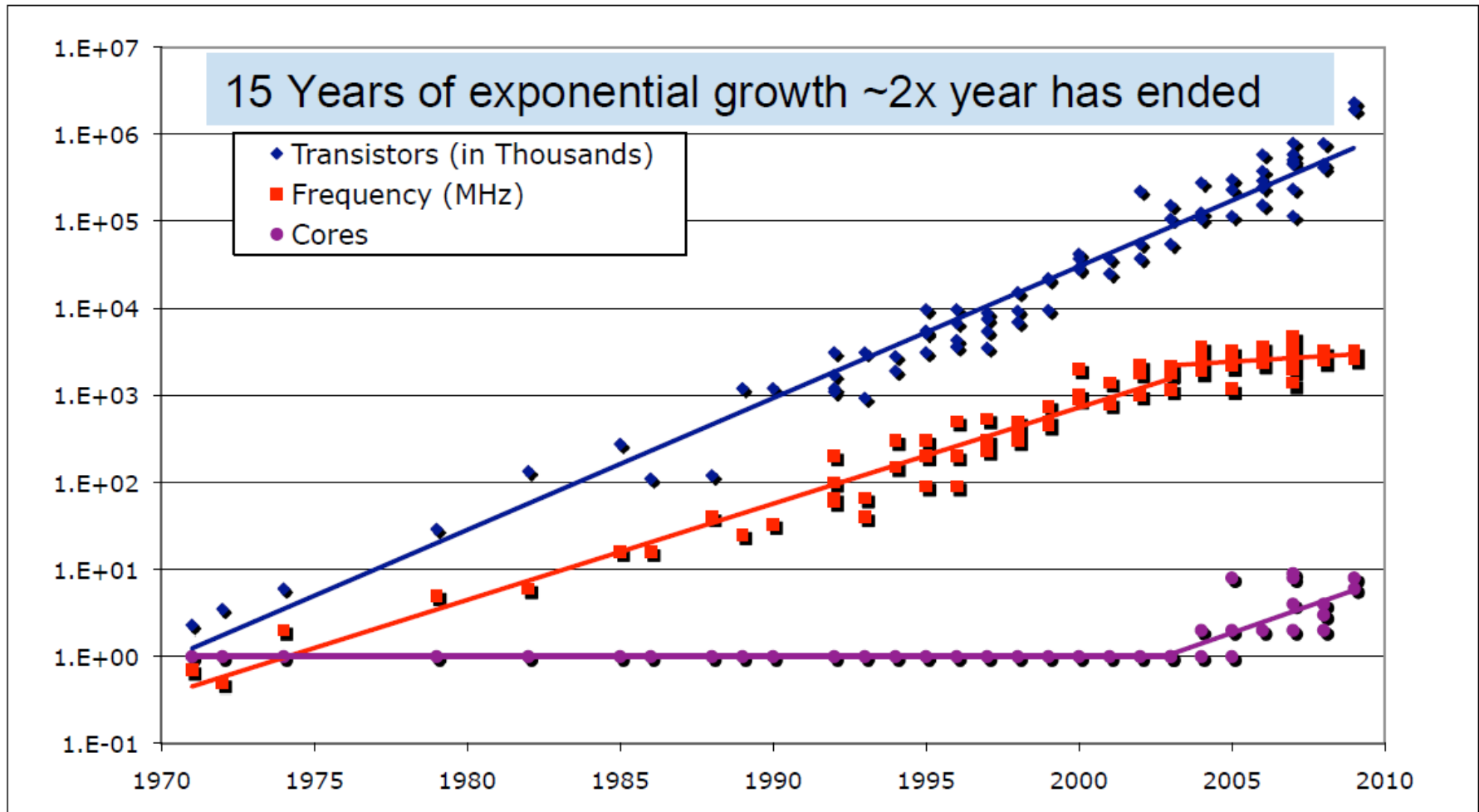


晶体管
集成密度
仍在提高

时钟频率
提高速度
急剧放缓

幻灯片来源: Maurice Herlihy

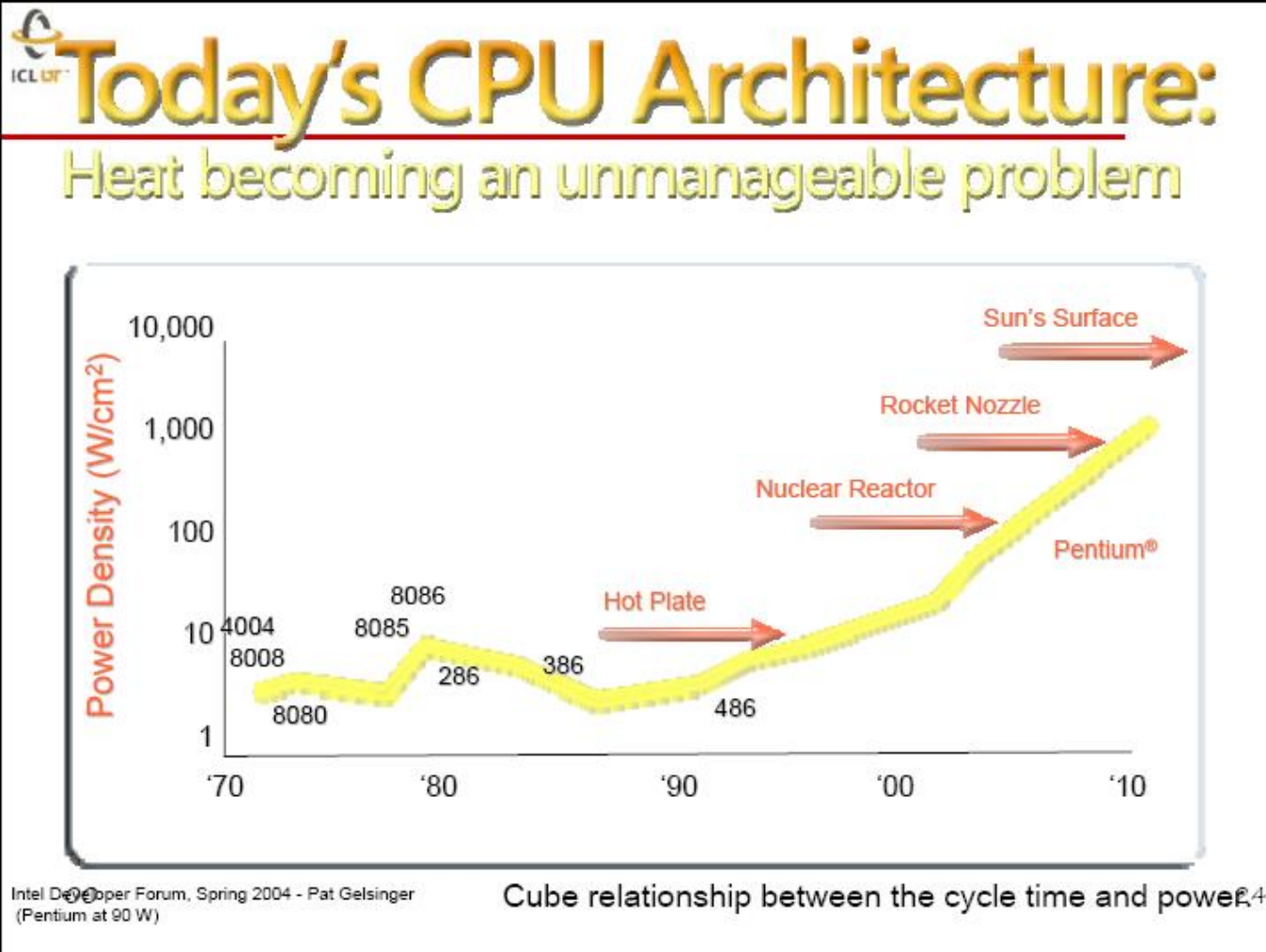
频率已不是处理器发展的主角



Data from Kunle Olukotun, Lance Hammond, Herb Sutter,
Burton Smith, Chris Batten, and Krste Asanović
Slide from Kathy Yelick

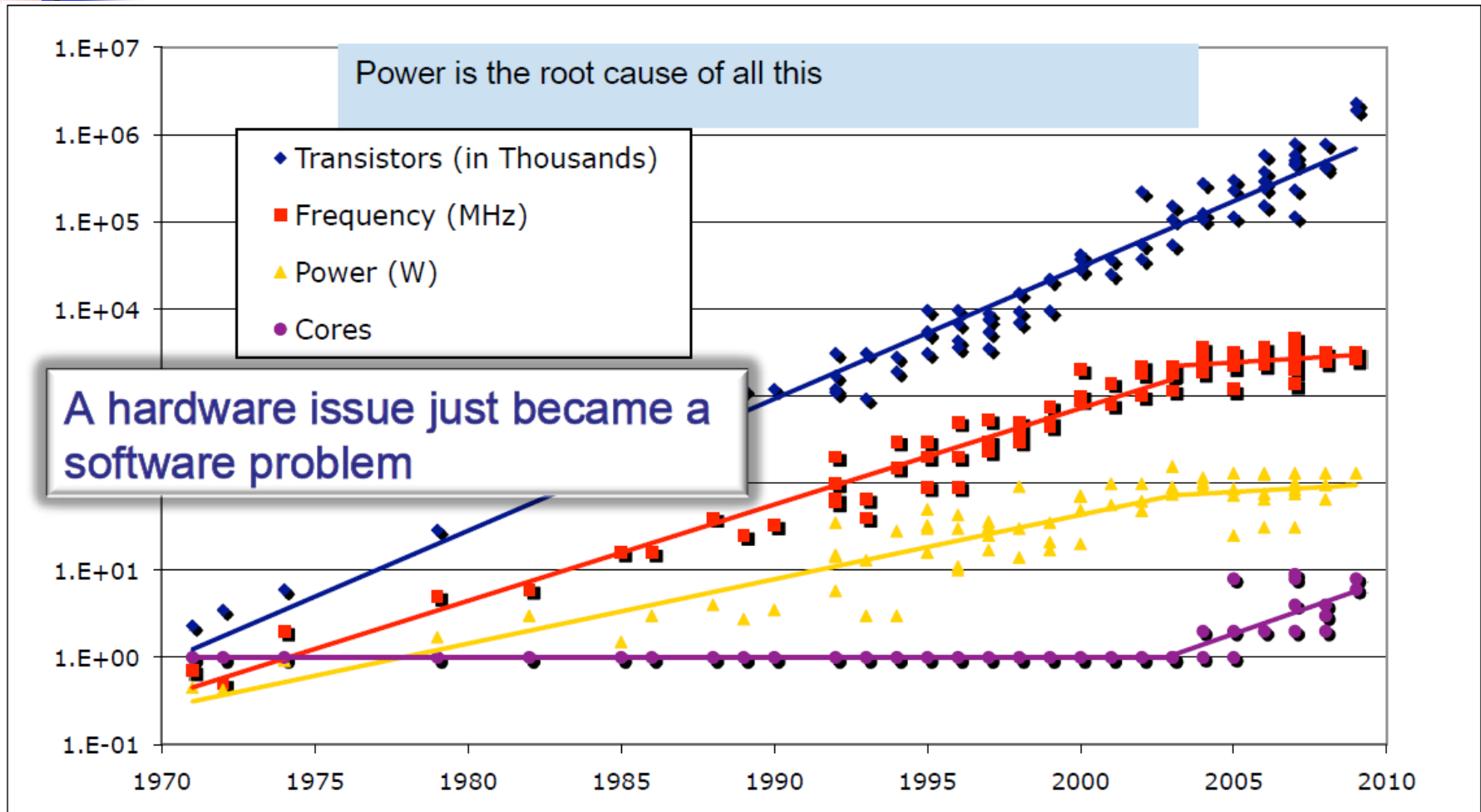
为什么?

功耗/散热的限制



Jack Dongarra, An Overview of High Performance Computing, HPC Asia 2005, Beijing, China, November 29, 2005.

性能上升放缓



Data from Kunle Olukotun, Lance Hammond, Herb Sutter,
Burton Smith, Chris Batten, and Krste Asanović
Slide from Kathy Yelick



多核、众核发展趋势

- 转变“更复杂的处理器设计、更快的时钟频率”的发展思路
- 并行架构更容易设计
- 充分利用资源
- 巨大的功耗优势

All Computers are Parallel Computers.

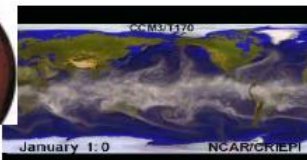
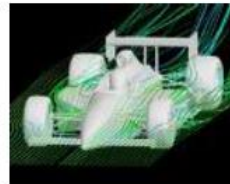
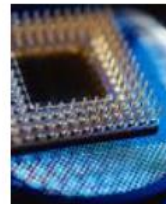
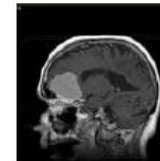


提纲

- 推动并行计算的因素
- 并行计算的应用
- 超级计算机硬件的发展
- 软件技术面临的挑战
- 众核技术/GPU的发展

- Of the 500 Fastest Supercomputer
 - Worldwide, Industrial Use is > 56%

- Aerospace
- Automotive
- Biology
- CFD
- Database
- Defense
- Digital Content Creation
- Digital Media
- Electronics
- Energy
- Environment
- Finance
- Gaming
- Geophysics
- Image Proc./Rendering
- Information Processing Service
- Information Service
- Life Science
- Media
- Medicine
- Pharmaceuticals
- Research
- Retail
- Semiconductor
- Telecomm
- Weather and Climate Research
- Weather Forecasting





科学仿真

- 传统科学/工程研究模式

- 进行理论或纸面设计
- 进行实验或构建系统

- 局限

- 太难——建造大型风洞
- 太贵——建造“用完就扔的”大型客机
- 太慢——等待气候变化、银河系演化
- 太危险——武器、药物设计，气候实验

- 科学计算研究模式

- 使用高性能计算机系统仿真现象
 - 依赖于已知的物理定律和高效的数值计算方法



对计算能力的不断追求

- 科学仿真持续推动对计算系统的需求
 - 提高结果精度
 - 提高计算速度（例如，气候建模、灾难建模）
- 中国(和美国等)持续追求更大规模系统
 - 上述原因
 - 保持竞争力
- 商用计算机领域也是如此
 - 更强、更快、更便宜

幻灯片来源: Jim Demmel

实例：全局气候建模

- 问题描述：构造函数

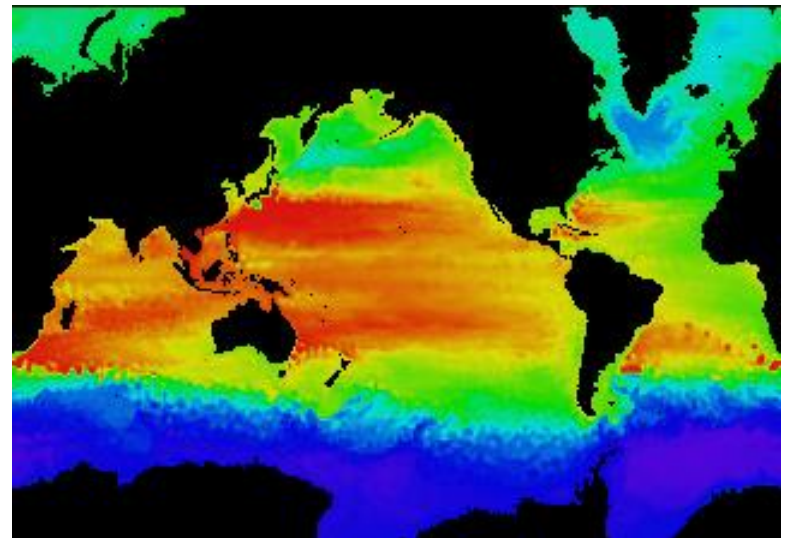
$f(\text{纬度}, \text{精度}, \text{海拔}, \text{时间}) \rightarrow \text{温度}, \text{压力}, \text{湿度}, \text{风力}$

- 方法：

- 离散化大气层，例如每1英里³一个测量点
- 设计算法，对给定 t ，预测 $t+\delta t$ 的天气

- 应用：

- 预测重要气候趋势，
如厄尔尼诺现象
- 设定空气排放标准



幻灯片来源：Jim Demmel

Source: <http://www.epm.ornl.gov/chammp/chammp.html>



为什么需要并行计算

- 大气层→1英里³的网格，考虑地表向上10英里的范围→共 5×10^8 个单元！
- 计算一定时间间隔，每个单元的物理量的变化——模拟大气运动
- 每次每个单元计算需200FLO，共 10^{11} FLO
- 时间间隔10分钟，计算10天内大气运动
- 100MFlops的计算机需100天
- 若想10分钟内完成——1.7TFlops的计算能力



其他例子：海洋建模

- POP, Parallel Ocean Program,
- 基于标准BCS模型
- MPI消息传递平台, F90语言
- 极点配置网格, 大气—海洋—海冰模型
- 大西洋高精度模拟, CM-5, 512个处理器, 4个月
- SGI Origin 2000 万亿次并行机, 全球高精度模拟需6个月
- J. K. Dukowicz and R. D. Smith, “implicit free-surface method for the Bryan-Cox-Semtner ocean model”, Journal of Geophysical Research, 99(C4):7991-8014, Apr 1994.

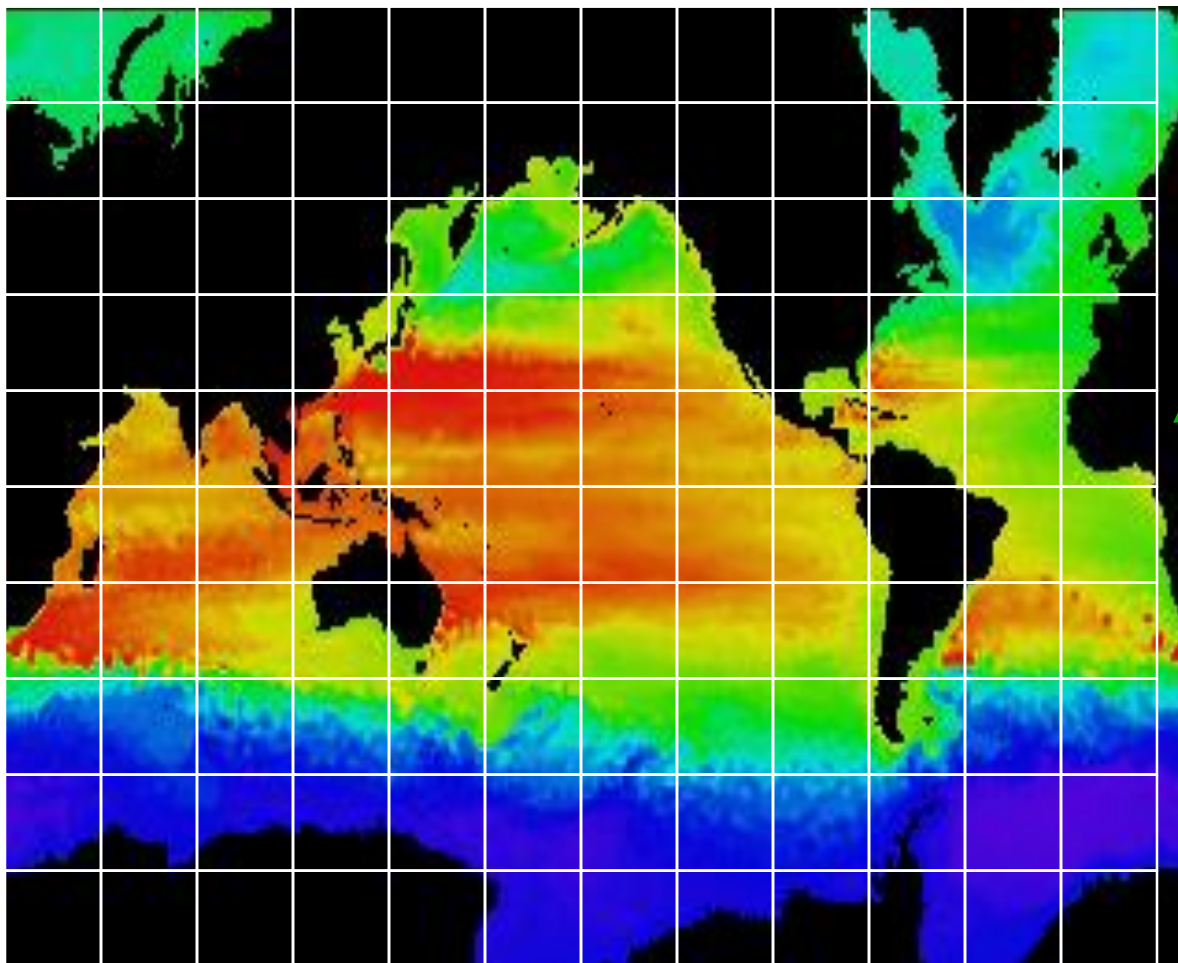


科学仿真的一般方法

- 将物理或概念空间离散化为网格
 - ▣ 规则划分更简单，可能需要自适应方法
- 在网格上进行局部计算
 - ▣ 给定昨天天气温和天气模型，今天温度期望？
- 网格局部结果互相交互
 - ▣ 综合局部天气结果来理解全球天气模式
- 重复若干时间步
- 可能对结果进行其他计算
 - ▣ 根据天气模型，哪些地区需要进行灾害疏散

离散化示例

另一个
处理器
并行计算
这个局部



某处理器
计算这个
局部

网格中相邻区域的处理器交换它们的计算结果.



更多例子：星系演化

○ 模拟天体运动

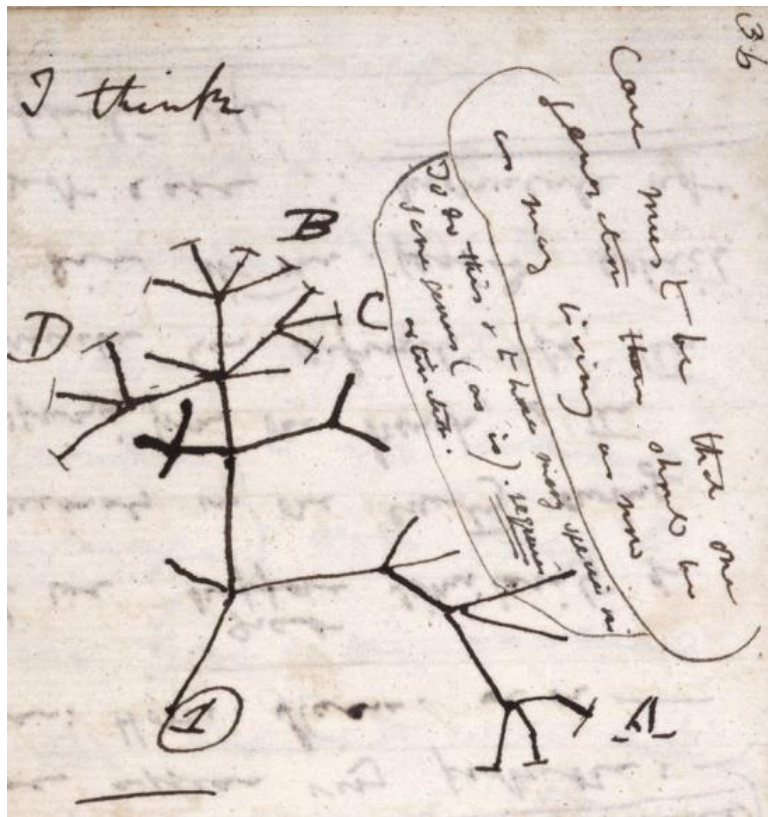
- 每个时间步，计算N个星体之间的引力
 - 每个星体的运动趋势
 - 若干时间步内星系的运动

N-body问题

- 计算N个点的相互引力： $O(N^2)$
- 10^{11} 个星体， $1\mu s$ 计算一个， 10^9 年！
- 近似算法： $O(N\log N)$ ， 也需1年

更多例子：生物信息学

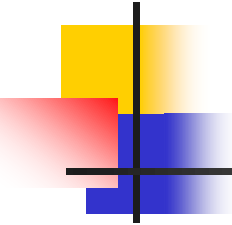
○ 种系发生算法MrBayes的GPU算法





种系发生算法MrBayes

- 基于现有物种的基因推断演化树结构
- 85个物种，基因序列长度13087
运行时间70小时！
- 当前基因测序技术发展迅速
基因序列长度达到数十万、上百万



更多例子：强子对撞机实验

- 利用欧洲高能物理实验室的强子对撞机进行发现希格斯玻色子标记的实验
 - 实验规模史无前例：全球几百个组织，5000多名物理学家
 - 每次实验产生PB级的试验数据
 - 美国：NSF物理网、DOE例子物理数据网格、NSF国际虚拟数据网格实验室
 - 欧洲：EU数据网格工程、英国GridPP项目、意大利INFN网格、Nordugrid网格

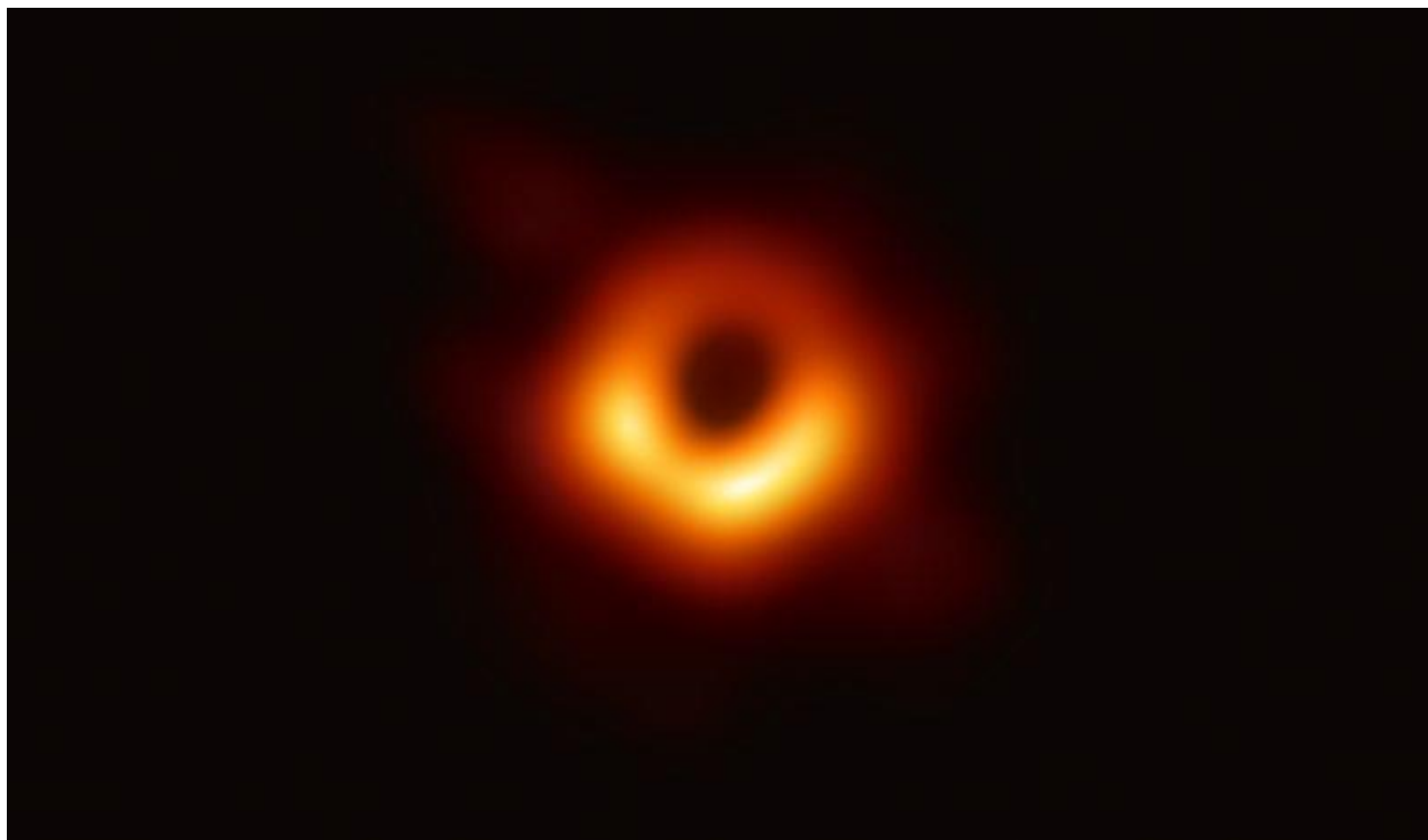


更多例子：天文学

- 天体望远镜采集的极大量数据的分析
 - 数字天空测量：10年内，10TB→1000TB
 - 美国国际虚拟天文台：每年500TB
 - 激光干涉仪重力天文台：每年250TB
 - VISTA望远镜：每晚250GB，每年10TB，10年几千TB
 - 关键是分析这些数据，建立大型数据仓库，提供全球用户访问的一致接口

更多例子：天文学

- 第一张黑洞照片（2019.4.11, M87黑洞）





更多例子：天文学

- 第一张黑洞照片（2019.4.11, M87黑洞）
 - 八个探测望远镜构成视界望远镜（EHT），跨越地球直径
 - 每台望远镜都记录了超1 PB（100万GB）的数据，数据量加起来足有36PB
 - 借助超计算机协助，但核算进程仍然耗费了科学家们近两年的时间。



更多例子：医学

- X光片、CT等的在线存储，图像分析
- UK e-Diamond项目
- 生物医学情报学研究网
- 美国国家数字乳房X射线照片档案
- 欧洲的MammoGrid
- ...



更多例子：商业应用

- Web服务为代表：大量静态、动态内容
 - 需要高性价比、伸缩性强的服务器
 - 多处理器PC、Linux集群
 - 华尔街，同时处理几十万用户的交易、上百万订单
IBM SP、SUN Ultra PC，全世界最大的超级计算机网络大多在此
 - Google，服务器数量数百万
百度，一个机房数万台服务器



更多例子：人工智能

○ AlphaGo的硬件与性能

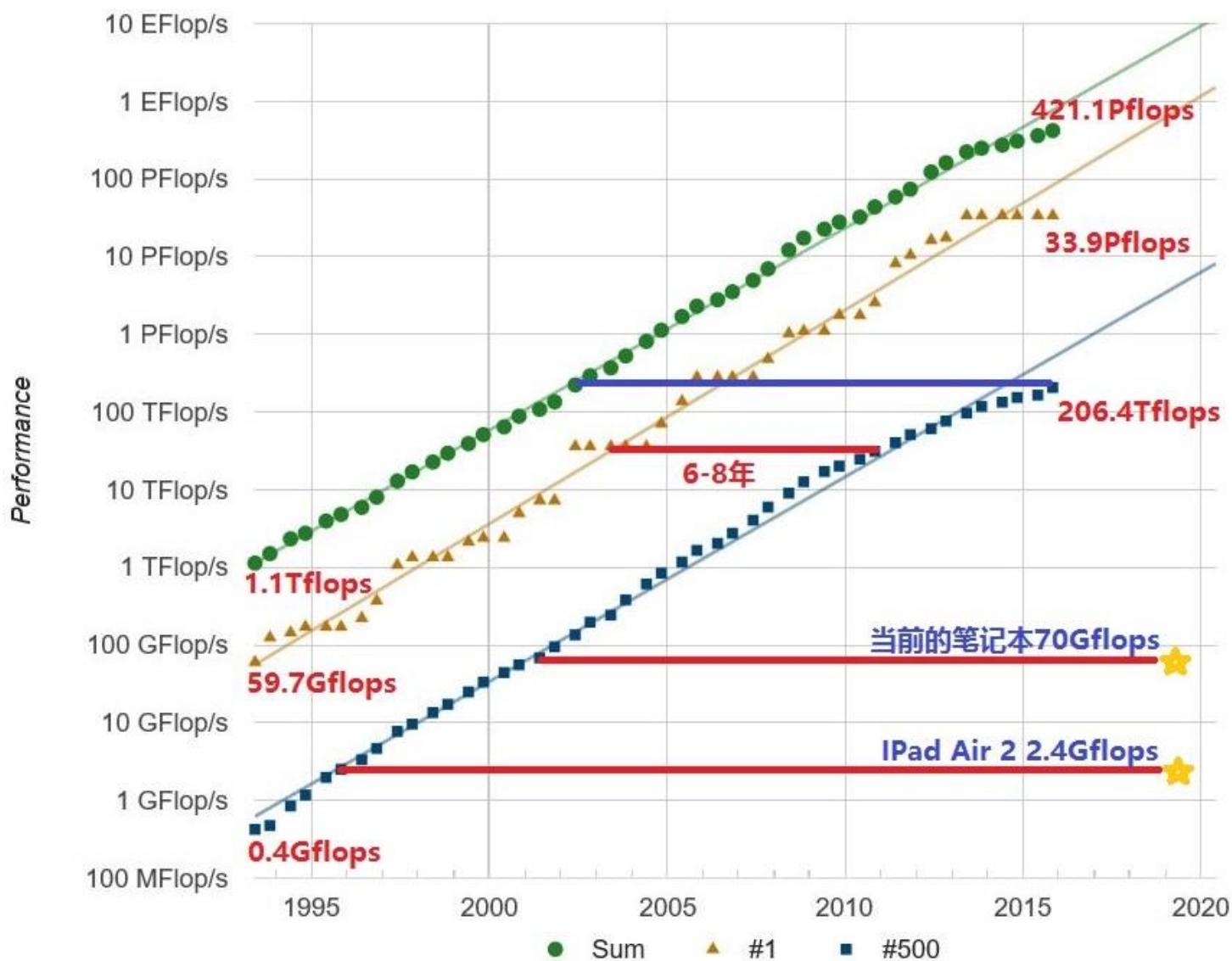
Configuration and performance				
Configuration	Search threads	No. of CPU	No. of GPU	Elo rating
Asynchronous	40	48	1	2,151
Asynchronous	40	48	2	2,738
Asynchronous	40	48	4	2,850
Asynchronous	40	48	8	2,890
Distributed	12	428	64	2,937
Distributed	24	764	112	3,079
Distributed	40	1,202	176	3,140
Distributed	64	1,920	280	3,168



提纲

- 推动并行计算的因素
- 并行计算的应用
- 超级计算机硬件的发展
- 软件技术面临的挑战
- 众核技术/GPU的发展

并行计算机的发展——top500





超级计算机排名Top10

本届排名	上届排名	名称	国家
1	1	Supercomputer Fugaku	日本
2	2	Summit	美国
3	3	Sierra	美国
4	4	Sunway TaihuLight	中国
5		Perlmutter	美国
6	5	Selene	美国
7	6	Tianhe-2A	中国
8	7	JUWELS Booster Module	德国
9	8	HPC5	意大利
10	9	Frontera	美国

截止至2021/06

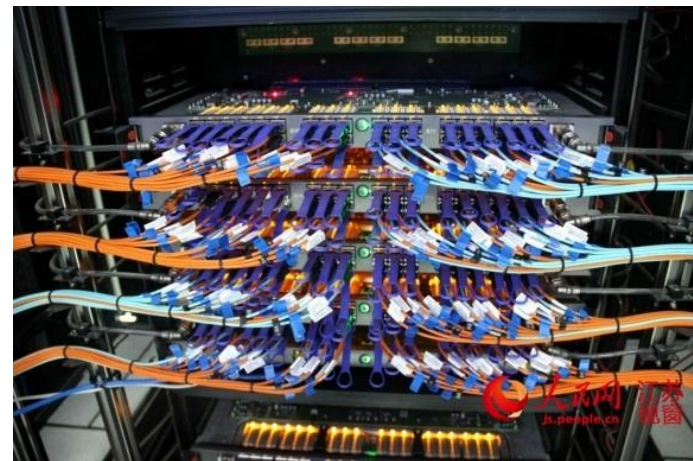
富岳

- Fugaku
- 日本理化学研究所(RIKEN) 与富士通公司共同开发
- 建成时间：2019年底，2020年6月进Top500 排名榜首
- 处理器：ARM架构的富士通A64FX，核心数达7630848个
- 峰值性能：513,855 TFlop/s (51.3855亿亿次/秒)
- 持续性能：442,010 TFlop/s (44.2亿亿次/秒)
- 操作系统：Red Hat Enterprise Linux



神威·太湖之光

- Sunway TaihuLight
- 国家并行计算机工程技术研究中心研制
- 安装在国家超级计算无锡中心
- 建成时间：2016年
- 40960个自研的“申威26010”众核处理器，10649600核心
- 峰值性能：12.5436亿亿次/秒
- 持续性能：9.3014亿亿次/秒
- 操作系统：Sunway RaiseOS 2.0.5



E级超算

○ 天河三号

- 2018年8月，原型机在国家超级计算济南中心完成部署并正式启用
- 运算峰值：百亿亿次/秒
- 预计2020年完成研制部署

○ 中科曙光

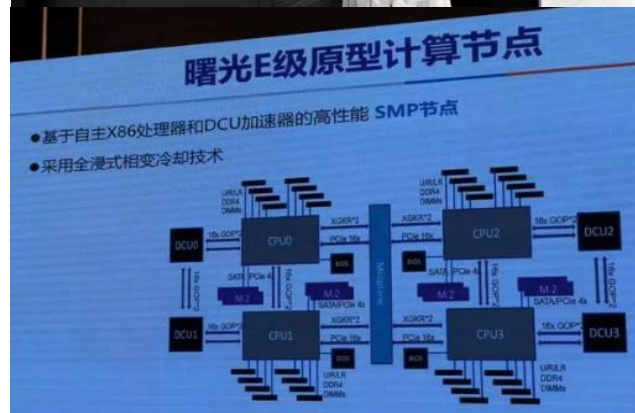
- 预计也在2020年完成研制部署

○ 美国“极光（Aurora）”

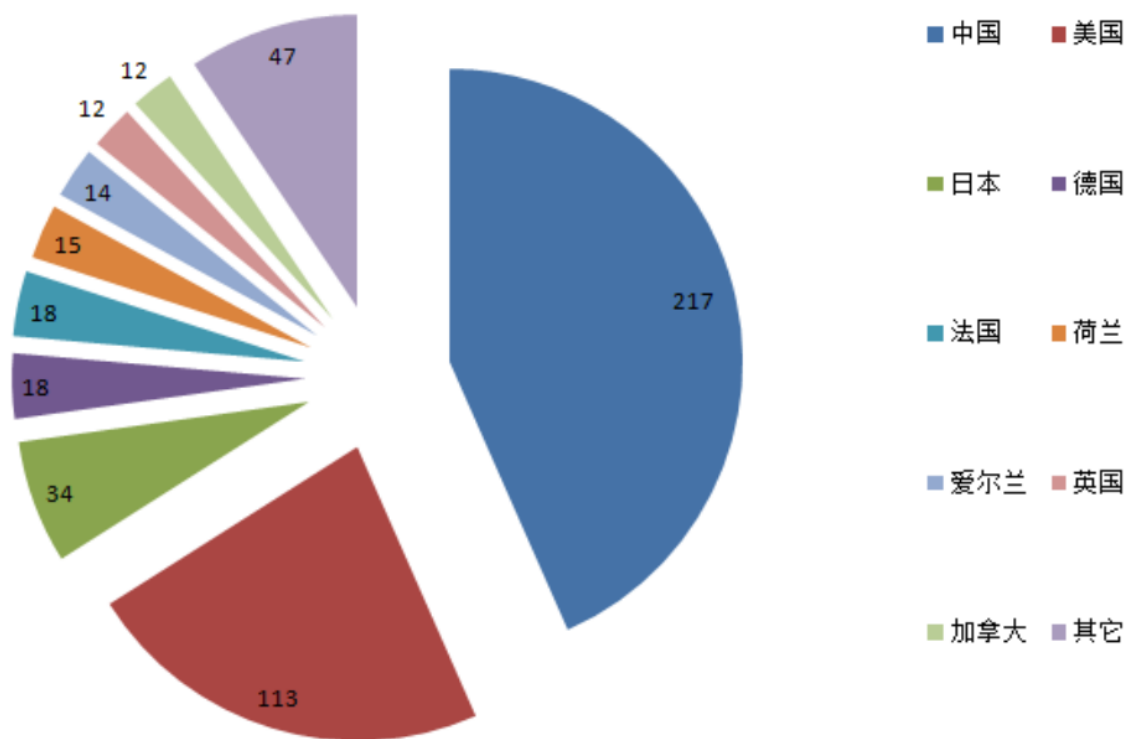
- 预计2021年上线

○ 日本“后京”（Post-K）

- 预计2020~2021年部署



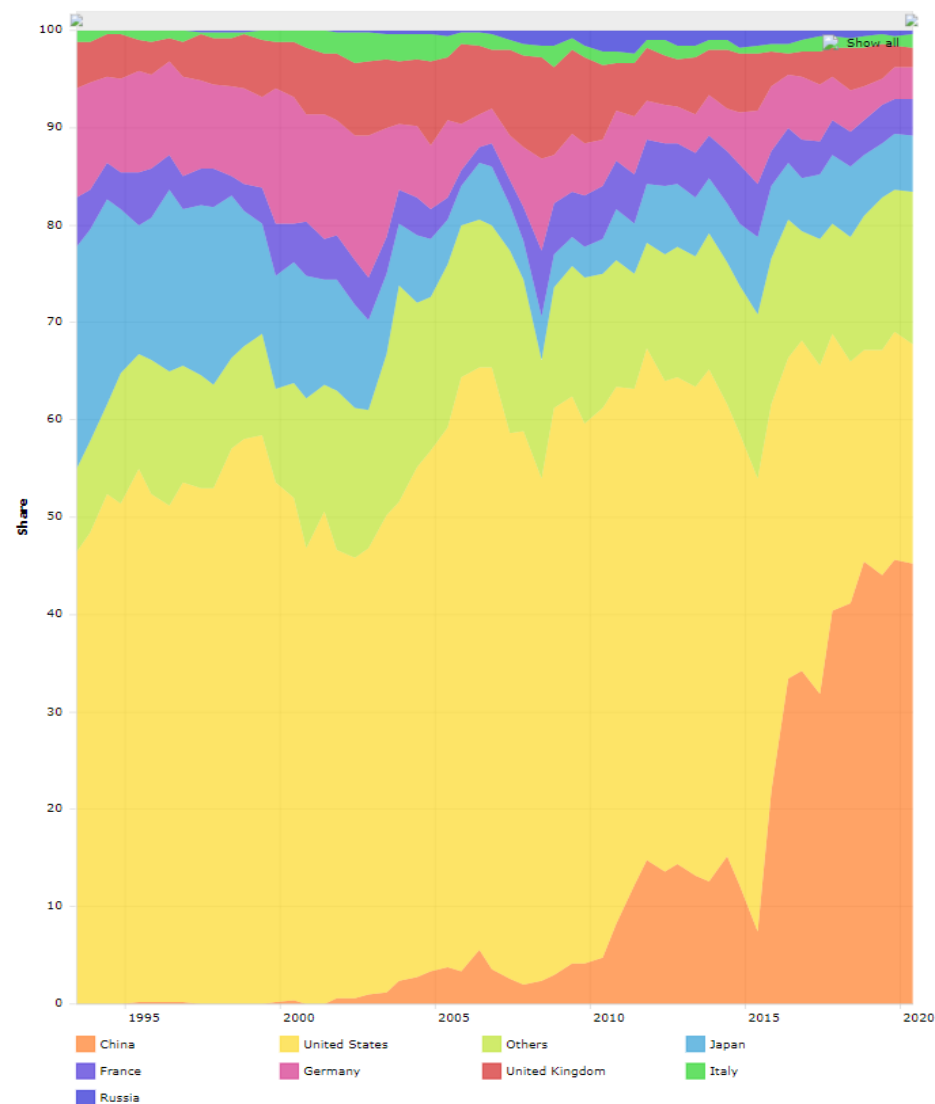
超级计算机分布



截止至2020/11

超级计算机的发展

Countries - Systems Share



Countries - Performance Share

