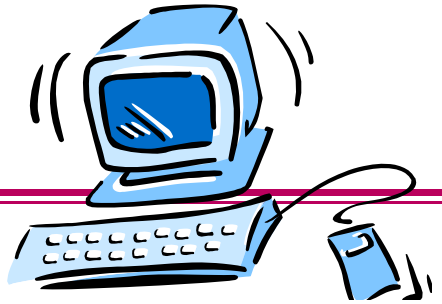


高性能计算技术

第一讲 引 言

何克晶
kjhe@scut.edu.cn

华南理工大学计算机学院

主要内容

- 什么是高性能计算？
- 课程简介
- 术语与定义
- 高性能计算发展现状及趋势

本课程是关于什么?

- 如何使得计算机可以更快地解决更大规模的问题（**bigger problems faster**）
- 直观的想法：可以用多台计算机
  - 如果一个程序在一台计算机上需要**100**个小时才能运行完，如果使用**100**台计算机，可能只要**10**个小时
  - 一台计算机只能处理**2GB**的数据集，如果使用**100**台计算机，可能可以处理**200GB**的数据集
- 如何能做到?

什么是高性能计算？

- 高性能计算（High Performance Computing）就是研究如何把一个需要非常巨大的计算能力才能解决的问题分成许多小的部分，分配给多个计算机进行处理，并把这些计算结果综合起来得到最终的结果的问题

- 高性能计算机是由多个可同时工作的处理器构成的计算机系统。
- 在一个高性能计算系统中，不同处理器同时运行同一程序的多个任务或进程，或者同时运行多个独立程序，以提高系统的运算速度、吞吐量或有效地利用系统的资源

智慧之语（Words of Wisdom）？

- “640KB [of memory] ought to be enough for anybody.”
  - Bill Gates， 微软总裁，1981.

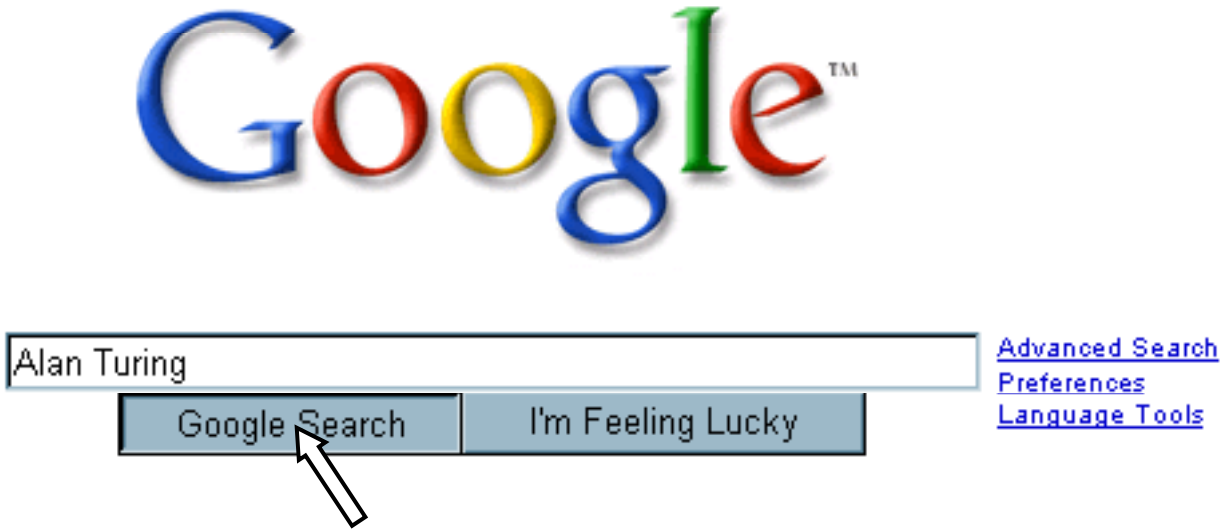
- 我们现在知道这是不正确的
  - 游戏（Games）
  - 数字视频、图像（Digital video/images）
  - 数据库（Databases）
  - 操作系统（Operating systems）
  - …….

信息孤岛

- 没有任何单个的服务器或搜索引擎能有效地覆盖不断增长的**web**内容
- **Internet**每年产生2×10^{18}（**2EB**）字节的信息
- 但每年只有3×10^{12} 字节信息可用（**0.00015%**）

来源: **Li Gong, IEEE Internet Computing, 2001**

一个简单的检索涉及到.....



- **200+ 处理器**
- **200+ TB数据库**
- **10^{10} 总的时钟周期**
- **0.1秒响应时间**
- **5¢ 广告收入**

Google的统计数据

- 基于2004年4月发布的Google IPO S-1表，Google拥有：
  - 719个机架（racks）
  - 63,272台机器
  - 126,544个CPUs
  - 253 THz的处理能力
  - 126,544 GB的RAM
  - 5,062 TB 的硬盘空间
  - 总计算能力：126–316万亿次（ teraflops）

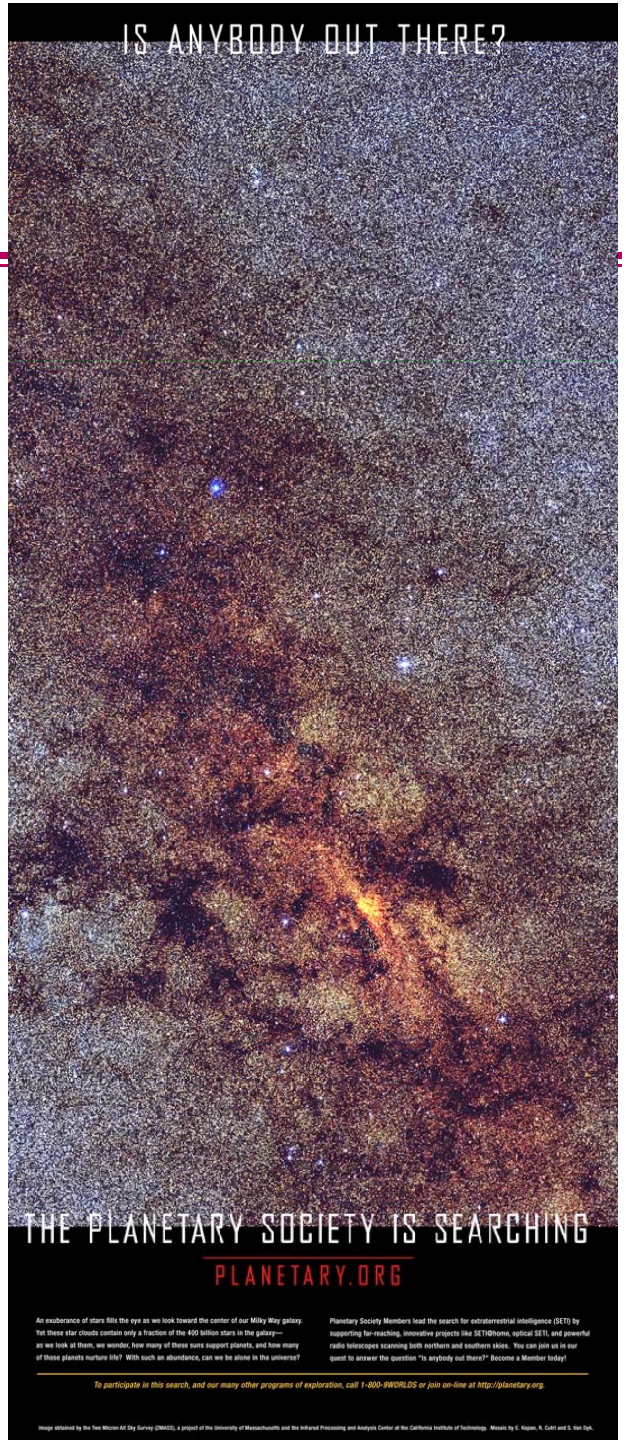
密码解密

- DES的56比特密钥有 $2^{56} = 7.2 \times 10^{16}$个可能的值，但强力搜索（brute force search）可能破译。
- 1997年1月28日，美国的RSA数据安全公司在RSA安全年会上公布了一项“秘密密钥挑战”竞赛，其中包括悬赏1万美元破译密钥长度为56比特的DES。美国克罗拉多洲的程序员Verser从1997年2月18日起，用了96天时间，在Internet上数万名志愿者的协同工作下，成功地找到了DES的密钥，赢得了悬赏的1万美元
- 1998年7月电子前沿基金会（EFF）使用一台25万美元的电脑在56小时内破译了56比特密钥的DES
- 1999年1月RSA数据安全会议期间，电子前沿基金会用22小时15分钟就宣告破解了一个DES的密钥
- 在现有的计算水平下，DES已经被宣布为不安全的密码

寻找星外文明：SETI@home

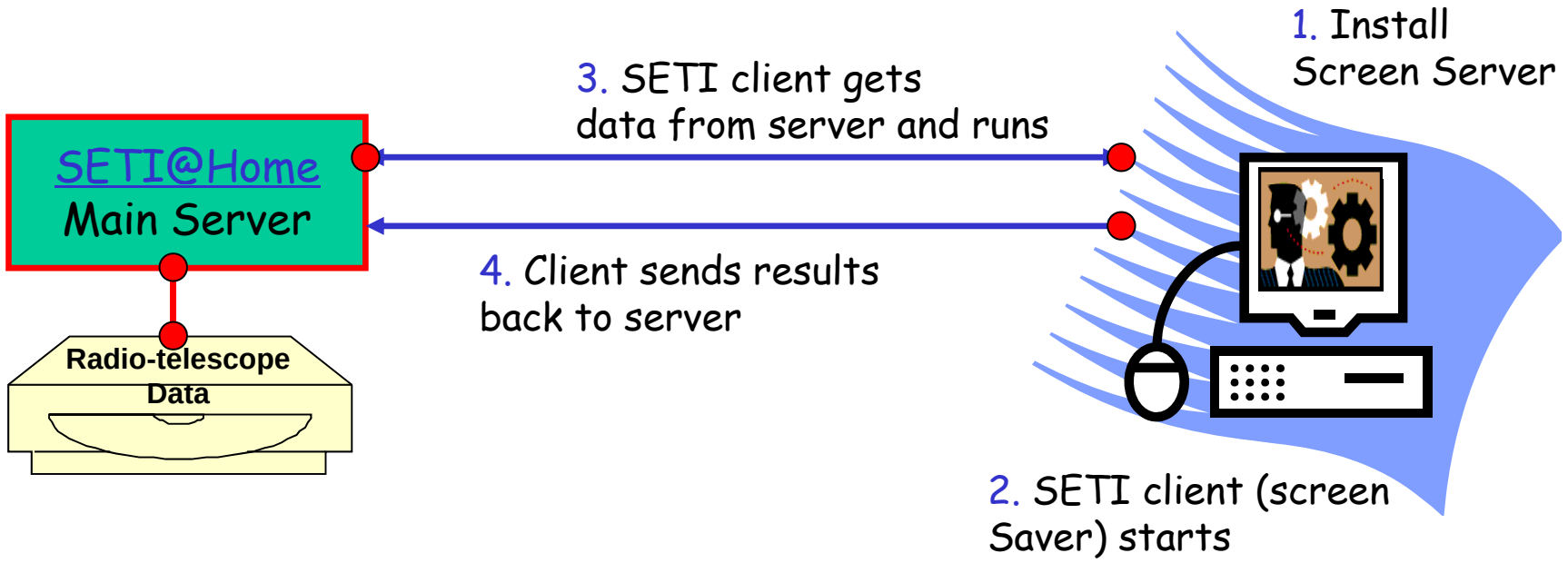
- **SETI（ Search for Extraterrestrial Intelligence ）：** 利用全球联网的计算机共同搜寻地外文明的科学实验计划
- 通过大规模并行计算完成来自其它宇宙文明社会电波信号的灵敏搜索
- **SETI@home**主要集中在检测窄频段信号，根据频段对数据进行分块，这些分块在本质上是相互独立的
- 对太空一个位置的观察得到的结果和另外一个位置得到的结果是相互独立的
- 因此可以把很大的数据集分成大量的小块，每一个计算机能够比较快的分析出其中的一块

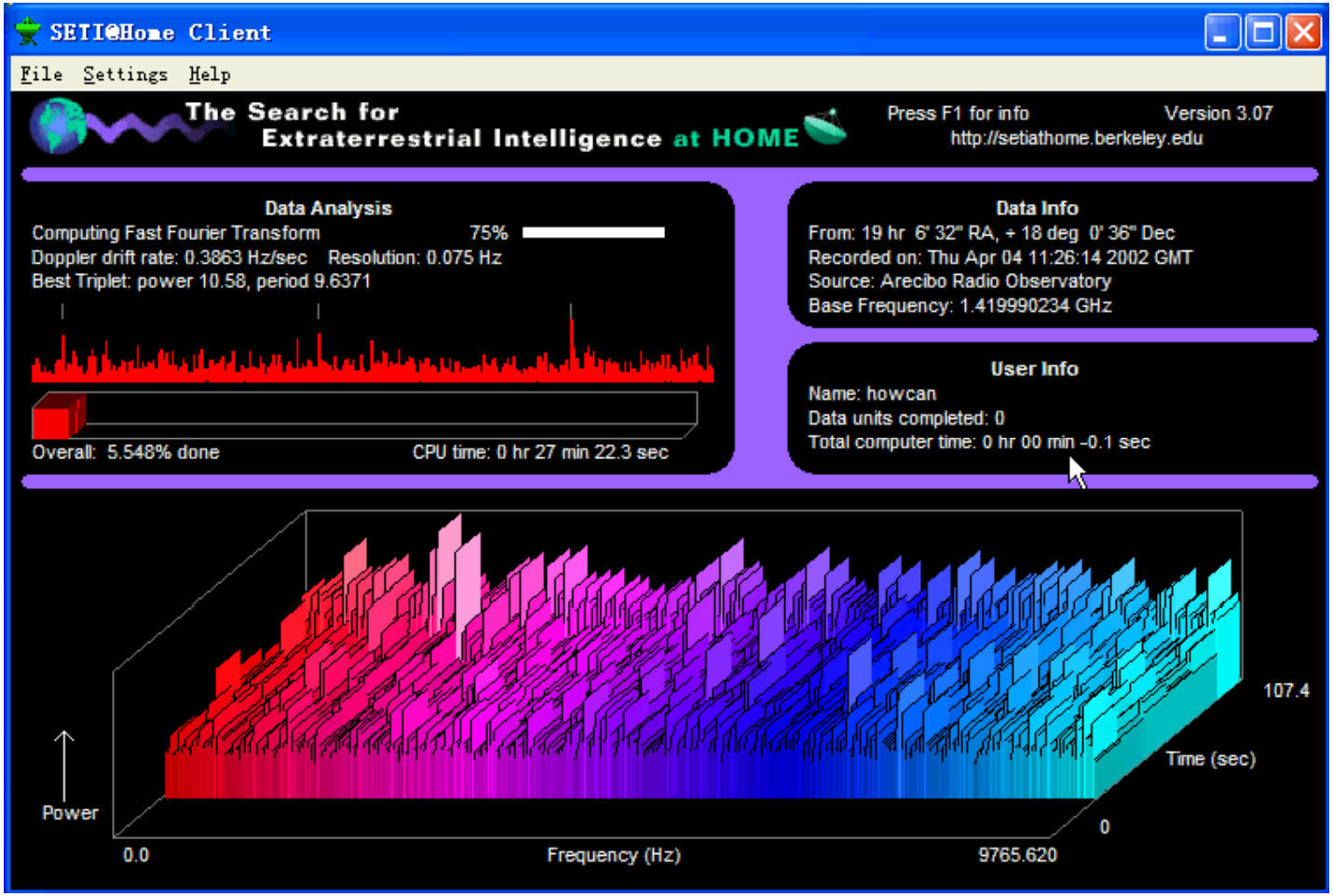
http://setiathome.berkeley.edu/



SETI@home

- 把工作分配到自愿贡献空闲cpu周期的机器处理
- 以屏保（**Screen Saver**）的方式运行





1999年5月17日开始正式运行

到Jan 13, 2006

用户数: 5436301

总的CPU时间：2433979.781 年

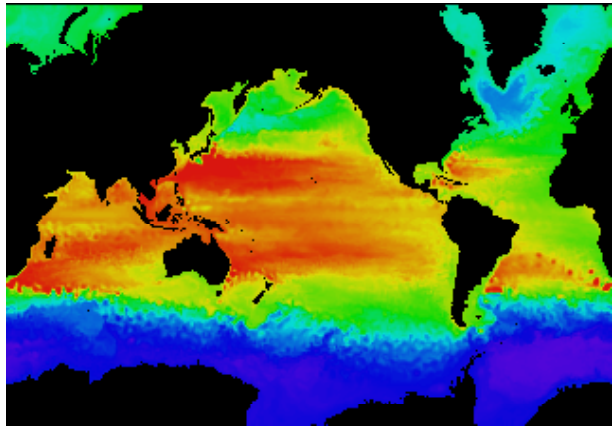
浮点计算次数：7.745086e+21

高性能计算的应用领域

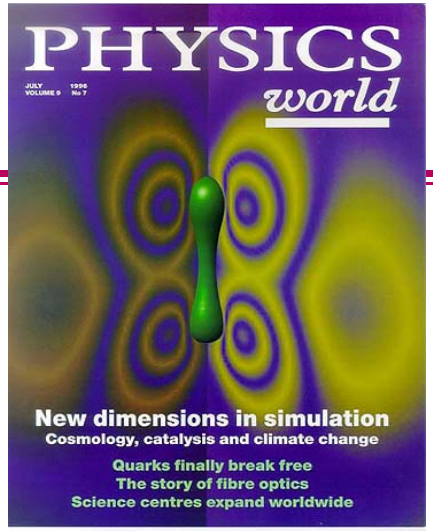
- 科学
  - 天气预报
  - 天体物理
  - 生物：生物形态学， 基因，蛋白质折叠、药物设计
  - 计算生物学
  - 计算材料科学与纳米科学
- 工程
  - （汽车）碰撞仿真
  - 半导体设计
  - 地震及结构建模
  - 计算流体力学（飞机设计）
  - 燃烧学（工程设计）
- 商业
  - 金融和经济构模
  - 事务处理
  - Web服务
  - 搜索引擎
  - 电子商务
  - 网络游戏
- 安全
  - 核武器－通过仿真（simulations）来做实验
  - 密码（Cryptography）

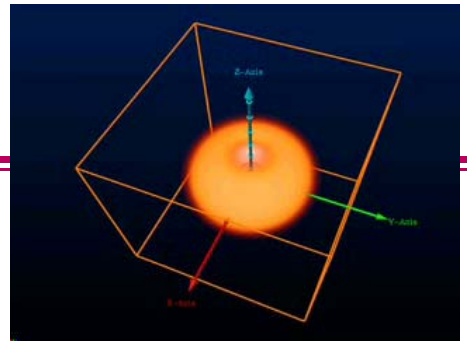
例子：天气预报

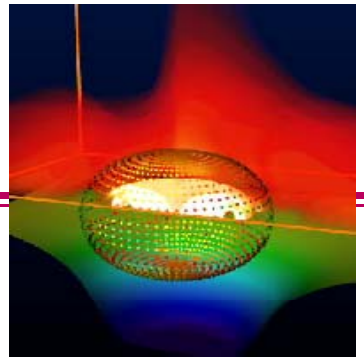
- 精确预报对计算能力有很高要求
  - 奥运场地按时间、地点的天气预报
- 计算问题：f (latitude, longitude, elevation, time) => temperature, pressure, humidity, wind velocity
- 方法：
  - 离散化，例如每10km一个测量点
  - 通过算法在给定时间t预测t+1时刻的气候

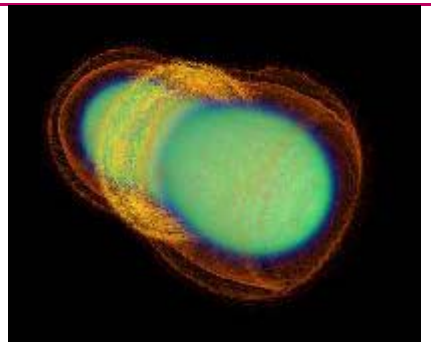


- 计算需求估计：
  - 实时预报：需要在60秒内进行 5×10^{11}个浮点运算，需要每秒80亿次浮点计算能力（ 8 Gflop/s）
  - 在24小时内预报一星期的天气 (7 days in 24 hours)，需要 每秒560亿次的浮点计算能力（56 Gflop/s）
  - 长期气候预测（50 years in 30 days），需要每秒4.8万亿次的浮点计算能力（4.8 Tflop/s）
  - 以12小时为单位的50年预测，需要288 Tflop/s
- 如果提高网格解析度则计算复杂性将呈8x,16x增加
- 更高的精确预测模型则需要综合考虑大气,海洋,冰川,陆地,加上地球化学等因素

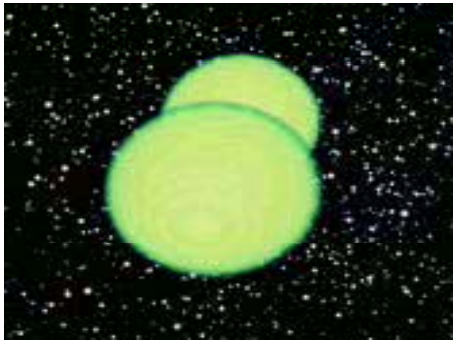

PHYSICS
world
JULY 1996
VOLUME 9 No 7
New dimensions in simulation
Cosmology, catalysis and climate change
Quarks finally break free
The story of fibre optics
Science centres expand worldwide

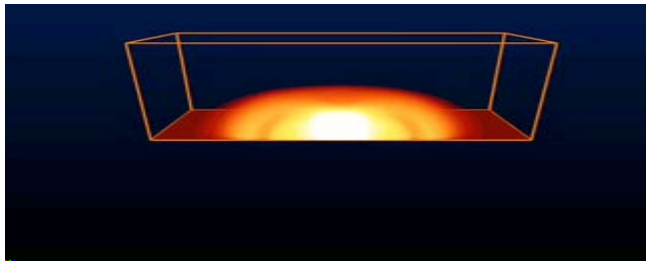


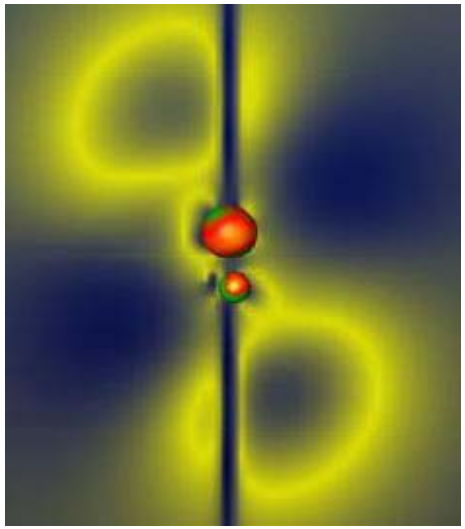


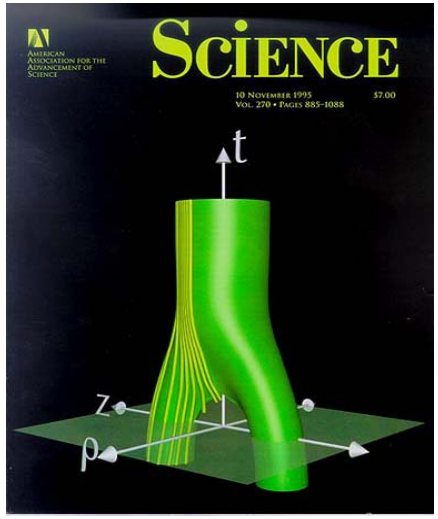


科学计算

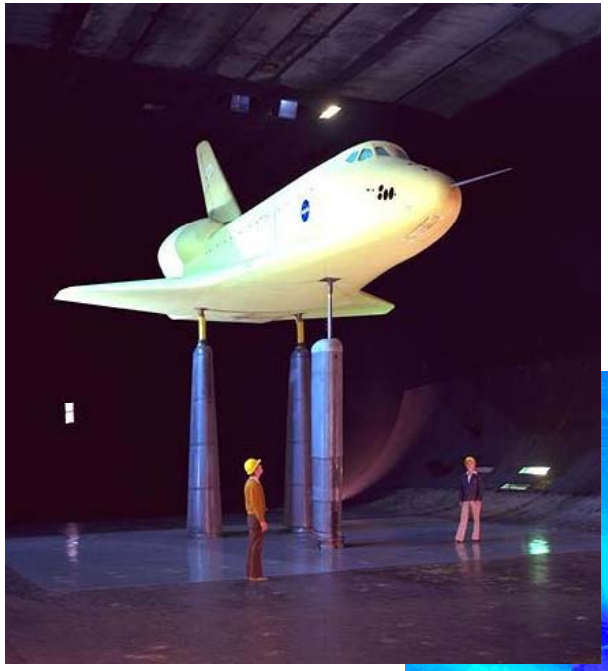




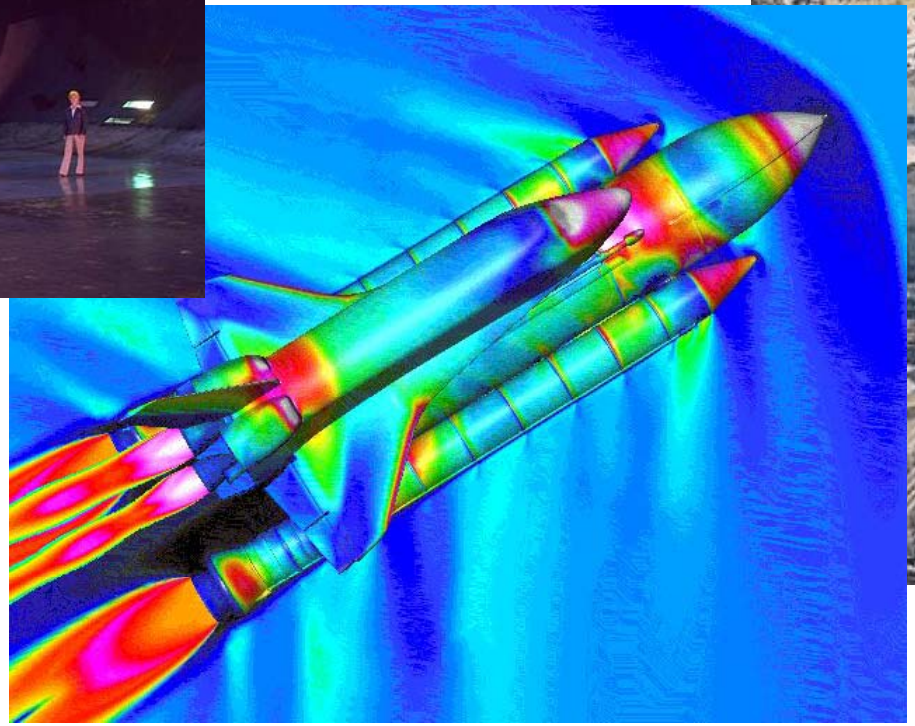



AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE
SCIENCE
10 NOVEMBER 1995
VOL. 270 • PAGES 885–1088
$7.00
t
z
ρ

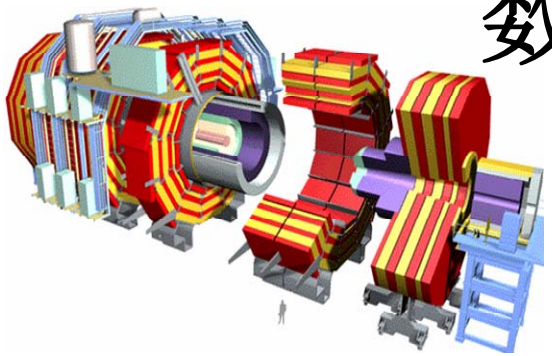
例子：计算流体力学
(Computational fluid dynamics ： CFD)



用计算机代替美国宇航局（**NASA**）的风洞



数据密集型计算

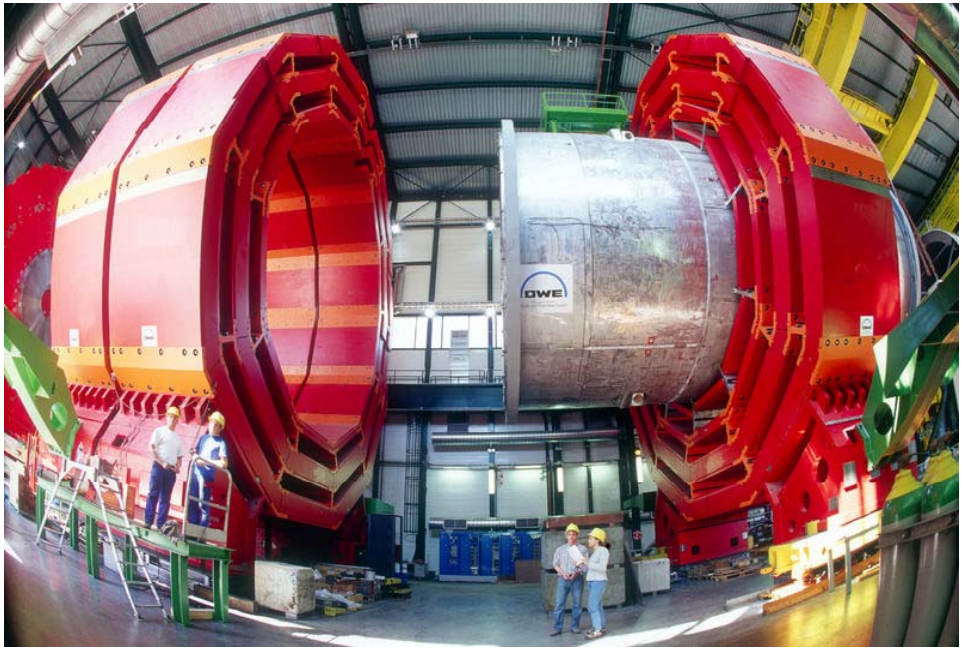


例子：大规模数据分析

欧洲核子研究中心（CERN）

- 干涉重力波天文台（Interferometer Gravitational Wave Observatory：LIGO）：由于大规模如星球的突然移动造成的空间和时间的微小扭曲

1TB/day (1024 GB/day), Year-long experiments

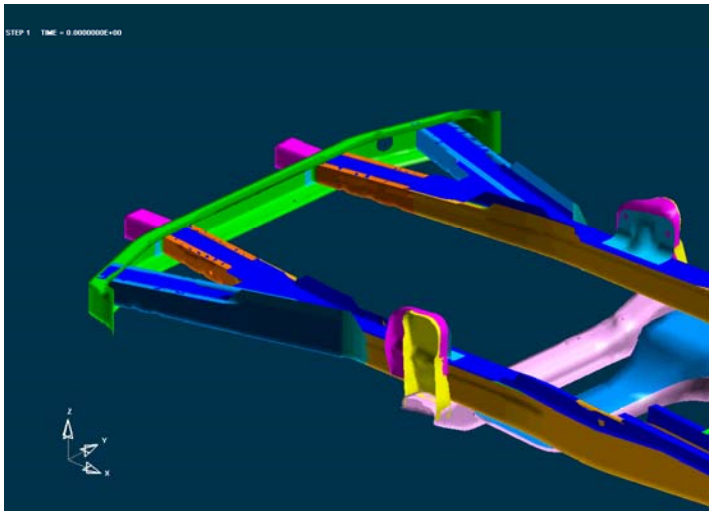


- 粒子探测器（The Compact Muon Solenoid）：用来寻找新粒子

10 GB/sec!!!
Many PB/year (1024 TB/year)

工程高性能计算

- 药物设计
- 新材料
- 石油勘探
- 流体动力学设计
- 碰撞仿真
- 强度设计
- 温度场/电磁场优化
- 建筑
- 工业设计
- 电力调度
- ……



例子：动漫与影视创作

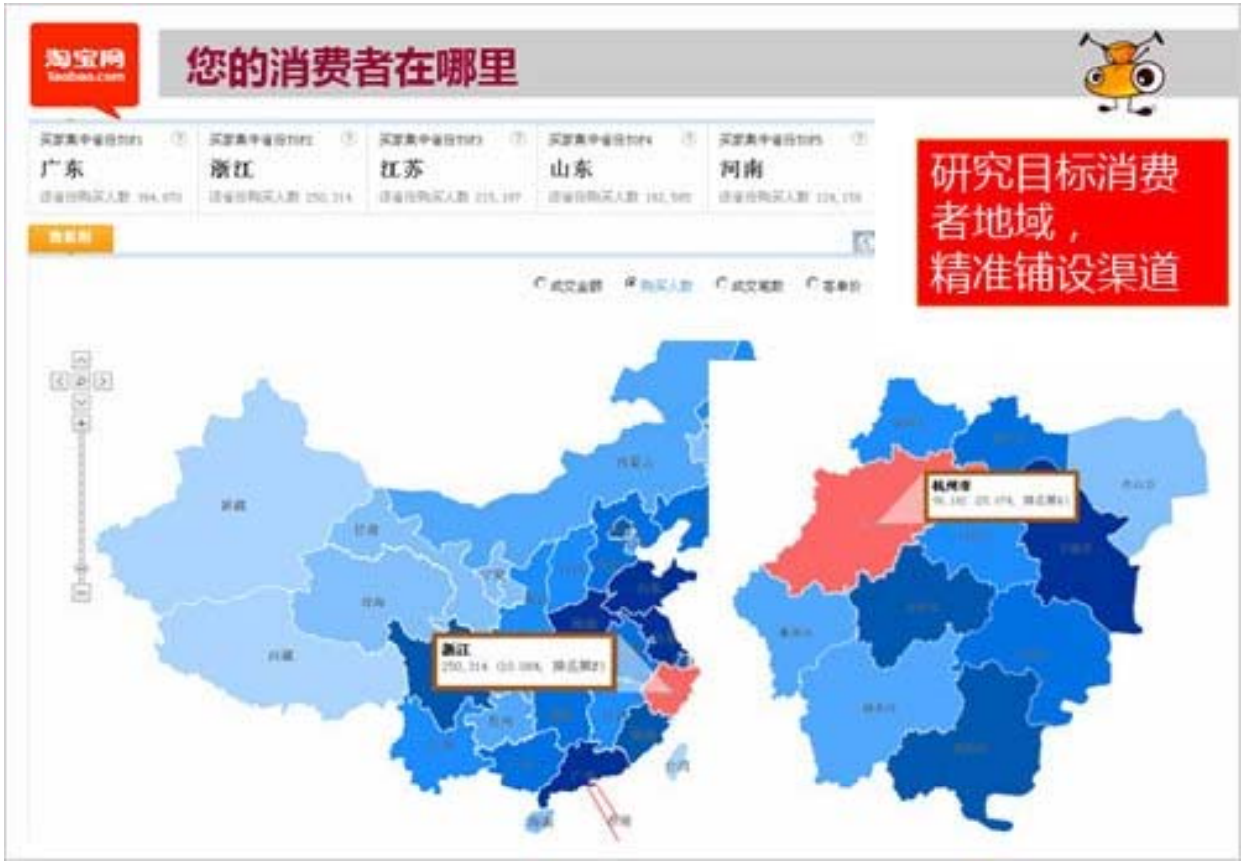
项目	长片动画影片
制作人员数量	80～100人
影片时长	90分钟
画面数量	90分钟×60秒×24帧=129600
每帧存储空间	12 MB
每部影片存储空间	2 TB
每帧计算时间	60分钟～300分钟
总计算量	5400天（129600小时）～27000天（648000小时）
200节点可完成数	27天～135天
废片率（重复计算）	每帧 3～25次

变形金刚：虚拟机器人造型，虚拟场景以及现场实拍特效完美的合成

- **渲染农场**（Render Farm）：分布式并行集群计算系统

商业高性能计算

- 决策支持，风险监测
- 数据挖掘
- 供应链优化



淘宝数据魔方

（淘宝网：每天产生数据量7T）

高性能计算应用分类

- 计算密集型（Compute-Intensive）应用
  - 大型科学工程计算与数值模拟
- 数据密集型（Data-Intensive）应用
  - 搜索引擎、数字图书馆、数据仓库、数据挖掘和计算可视化等
- 网络密集型（Network-Intensive）应用
  - 协同工作、遥控和远程医疗诊断等

驱动高性能计算应用的动力

- 科学计算
- 数据密集型计算
- （分布式）超级计算
- 合作工程
- 高吞吐率计算
  - 大规模模拟和参数研究
- 远程软件访问 / 租用软件
- 基于需求的计算

做得更快的方法

- 三种提高性能的方法
  - 努力工作（Work Harder）
    - Increase microprocessor performance
  - 工作得更有效率（Work smarter）
    - Better algorithm
  - 获取帮助（Getting help）
    - Parallel processing

单核CPU的瓶颈

单核CPU的极限突破

在CPU快速发展的20年里, CPU一次次地遭遇性能极限，但都又一次次地冲破了这个极限.从奔腾到奔腾2，CPU也突破了1GHZ，从奔腾2到奔腾3，CPU从1GHZ突破了2GHZ，从奔腾3到奔腾4，CPU也终于突破了3GHZ，现在最高的CPU主频已经高达3.8GHZ

单核CPU遭遇终极瓶颈

单核CPU好象停止了前进，到3.8G却怎么也超不过4G

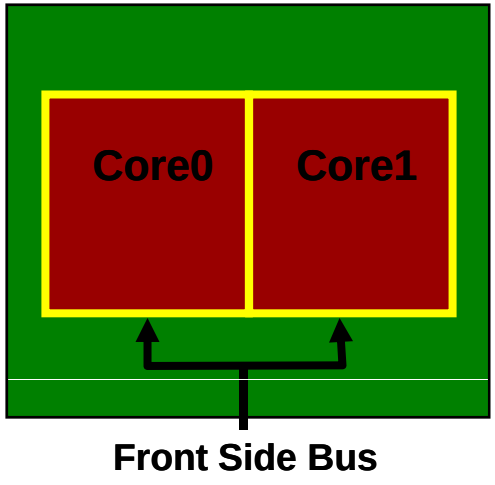
另寻出路

英特尔不得不承认奔腾系列已经遭遇主频极限，4G就象一场百年罕见地强降雪，将奔腾系列CPU的性能高速公路永远地封死了。因此，以英特尔为首的CPU军团不得不另外找一条更宽的性能高速公路来继续他们的CPU神话

多核时代来临

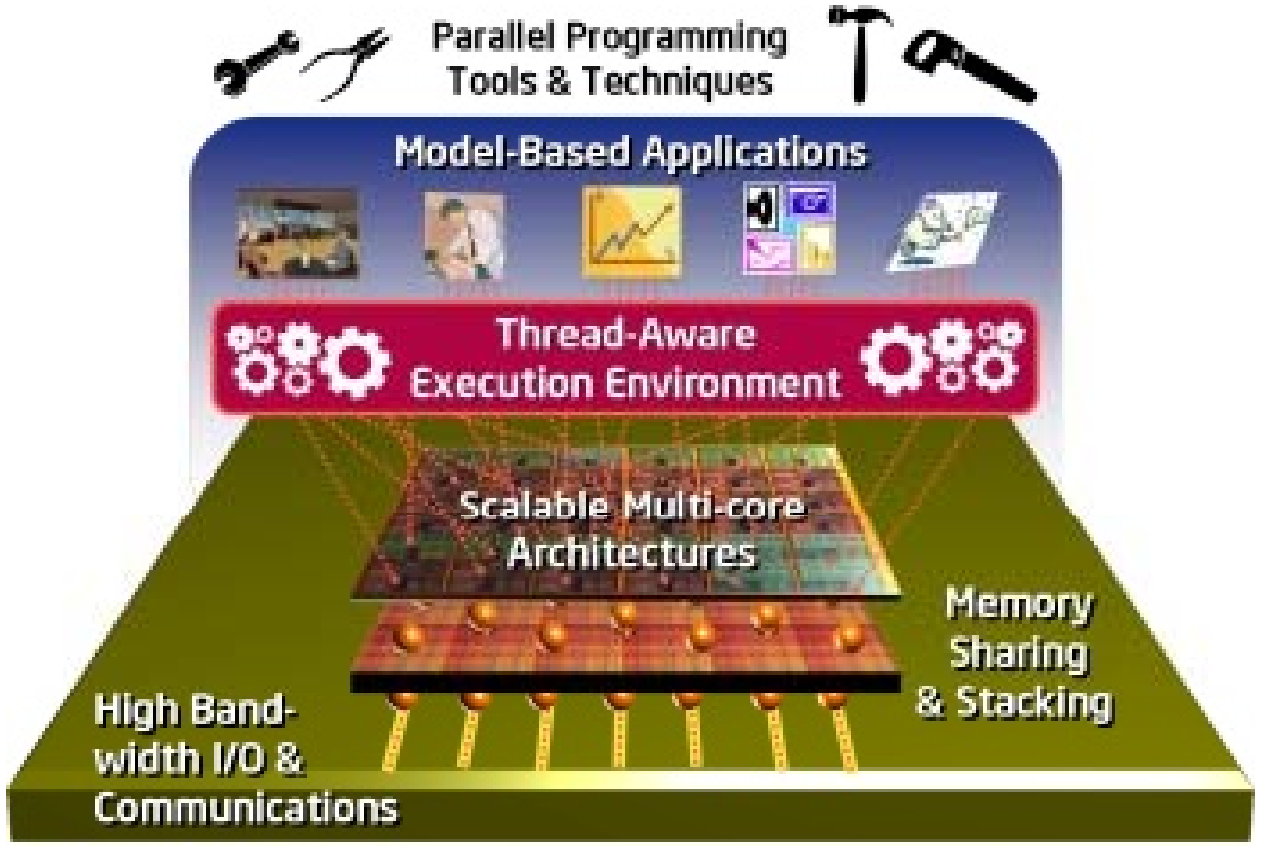
- AMD在2006年第一个推出了双核处理器，这种处理器的计算单元相互独立，但它们将共享CPU的一、二级缓存。这种CPU虽然没有两颗CPU的效率高，但它的性价比是非常高的

- 什么是多核处理器：两个或多个独立运行的内核集成于同一个处理器上



Intel Terascale*研究计划

- **TeraFLOP处理器**



http://techresearch.intel.com/articles/Tera-Scale/1421.htm

多核挑战软件开发

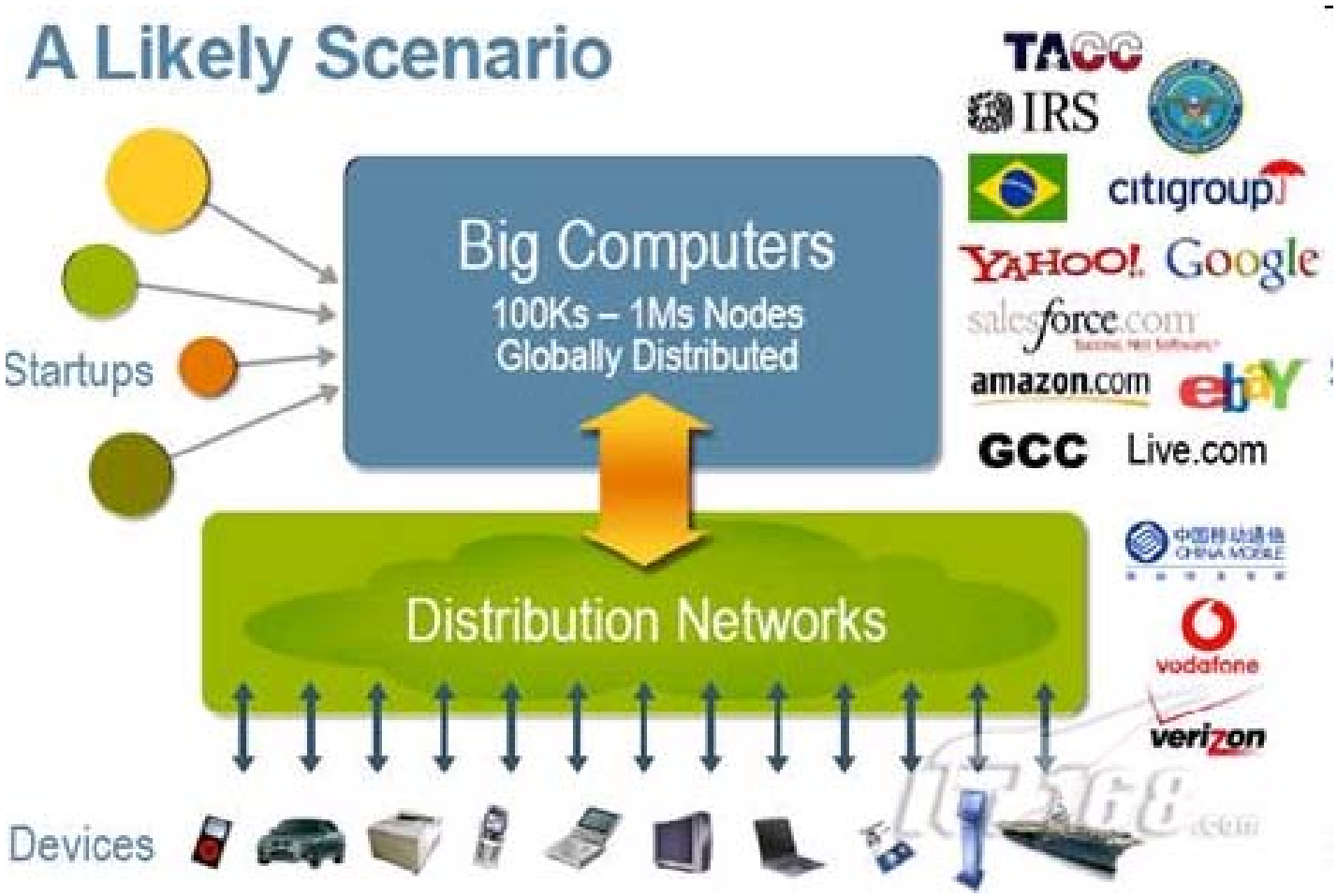
- 性能: 双核的运算速度一定会比单核的CPU快吗?
- 不一定。如果不针对多核进行软件开发，不仅多核提供的强大计算能力得不到利用，相反还有可能不如单核CPU好用。因为采用多核的CPU其每个内核的主频比主流的单核CPU通常要低一些，如果你的程 序只能发挥出一个内核效用的话，自然不如在单核CPU上快
- 要想发挥多核功能，设计的软件就首先要能做并行计算

讨论：以下看法是否正确？

- 高性能计算需要非常高性能的计算机
  - 算法/软件/时间决定了需要使用的机器
  - 即便是一台普通的台式机，也一样可以开展计算
- 高性能计算象魔法一样无所不能
  - 高性能计算是知识的程序化
  - 高性能计算只能对经济/社会效益起到助推器的作用

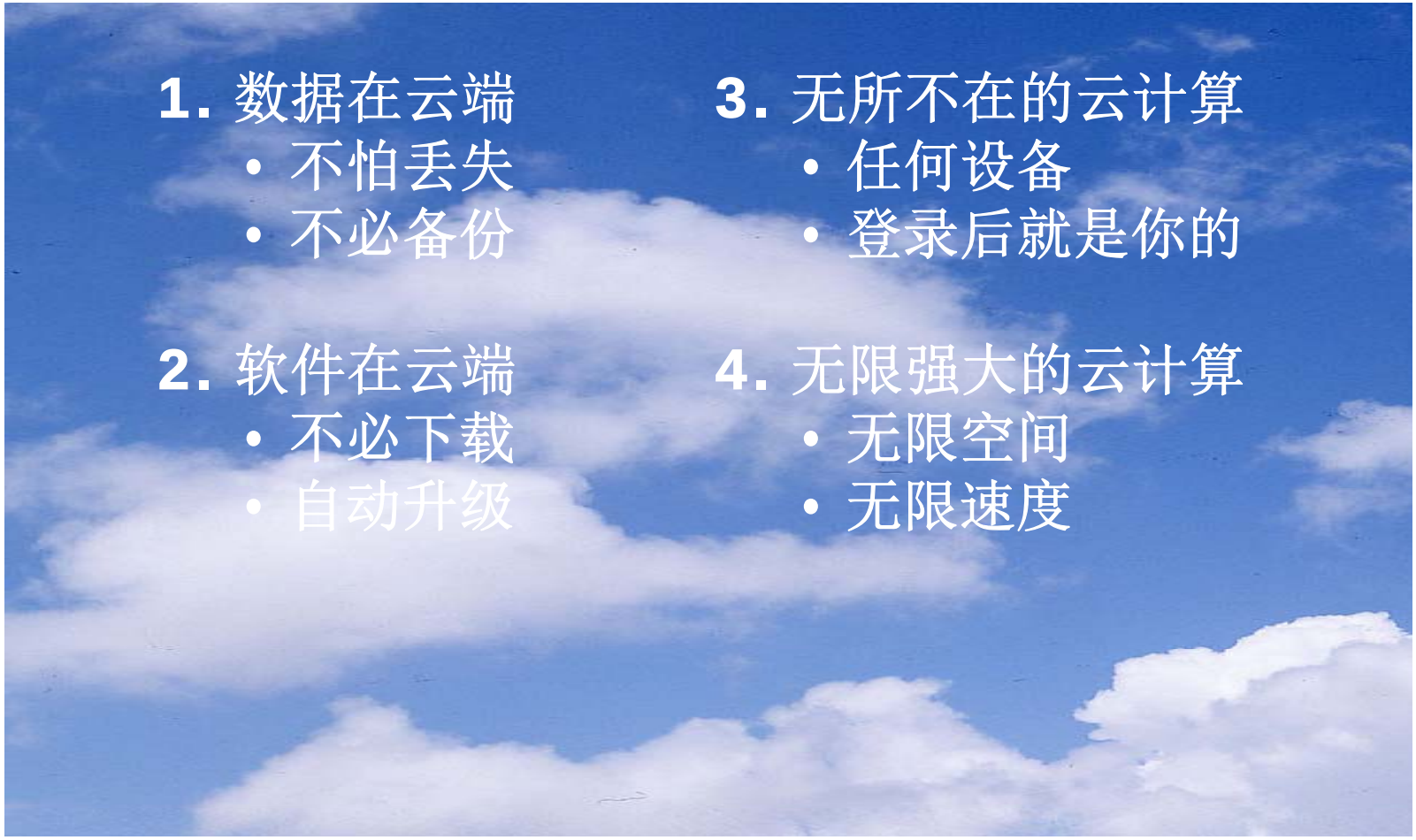
高性能计算并非高不可攀的普及时代已经来临！

云计算（Cloud Computing）



“云计算”——将所有的计算资源集中起来，并由软件实现自动管理（无需人为参与），并提供应用服务

云计算时代的到来


1. 数据在云端
• 不怕丢失
• 不必备份
2. 软件在云端
• 不必下载
• 自动升级
3. 无所不在的云计算
• 任何设备
• 登录后就是你的
4. 无限强大的云计算
• 无限空间
• 无限速度

主要内容

- 什么是高性能计算？
- 课程简介
- 术语与定义
- 高性能计算发展现状及趋势

课程内容

- 高性能计算系统及其结构模型
  - 高性能计算机的系统结构模型，对称多处理机（SMP）、大规模并行处理机（MPP）、集群系统（Cluster）和并行计算的性能评测；
- 并行算法设计
  - 并行算法的一般设计策略、基本设计技术和一般设计过程
- 并行程序的设计原理与方法
  - 并行程序设计基础、共享存储编程和分布存储编程以及并行程序设计环境与工具
- 高性能计算应用及发展趋势

课程要求

- 课堂讲授+上机实践
  - 共享存储编程
  - 消息传递编程
  - Hadoop平台和Map/Reduce编程
- 了解和掌握高性能计算的基本原理、技术及最新研究成果，具有高性能计算的理论基础和实践能力
- 应用高性能计算技术解决实际问题
- 评分：平时作业 (20%)，实验（20％），期末笔试(60%)

学习用书

- 课本：
  - 陈国良，并行计算——结构．算法．编程（修订版），高等教育出版社，2011
- 参考书：
  - Kai Hwang等著，陆鑫达等译，可扩展并行计算技术、结构与编程，机械工业出版社，2002
  - Maurice Herlihy等著，金海等译，多处理器编程技术，机械工业出版社，2009

华南理工大学高性能计算平台: SCUTGrid

http://grid.scut.edu.cn

主要内容

- 什么是高性能计算？
- 课程简介
- 术语与定义
- 高性能计算发展现状及趋势

高性能计算相关术语

- 高性能计算（**High-Performance Computing： HPC)**、超级计算（**Supercomputing**）
  - 很难定义，因为计算机的性能不断在提高
  - **Top 500**列表中的机器？
- 高性能计算和通信（**High-Performance Computing and Communications： HPCC**）
  - 分布式高性能计算、高速网络和**Internet**的使用
- 并行计算（**Parallel Computing**）
  - 使用多处理器系统的高性能计算
- 分布式计算（**Distributed Computing**）
  - 更着重于功能而不是性能的增加
- 网格计算（**Grid Computing**）
  - 分布式高性能计算（**Distributed, High Performance Computing: DHPC**），或称元计算（**Metacomputing**）
- 云计算（**Cloud Computing**）
  - 共享基础架构，将巨大的系统池连接在一起以提供各种服务

高性能计算的基础

- 计算（Compute）：仿真计算或合并数据源的数据处理能力
  - 浮点计算能力（FLOPS）
- 存储（Storage）：hierarchical from cache, to main memory, local disk, 磁盘阵列、磁带等...
  - 每秒读写的字节数（Mbytes/s)
- 通信（Communications）：内部网络、局域和广域网络
  - 带宽和时延
  - 相比于以太网的Mbit/s带宽和毫秒级时延，高性能计算系统具有Gb的网络带宽和微秒级时延

高性能计算的测度单位

- **Flops（floating point operations）：浮点计算操作**
- **Flop/s：每秒浮点计算操作**
- **Bytes：数据大小**

单位	描述
Mflop/s	**10^6 flop/sec （每秒百万次）**
Gflop/s	**10^9 flop/sec （每秒10亿次）**
Tflop/s	**10^{12} flop/sec （每秒万亿次）**
Pflop/s	**10^{15} flop/sec （每秒1千万亿次）**
Mbyte	**10^6 byte (also 2^{20} = 1048576)**
Gbyte	**10^9 byte (also 2^{30} = 1073741824)**
Tbyte	**10^{12} byte (also 2^{40} = 109952116277 76)**
Pbyte	**10^{15} byte (also 2^{50} = 1125899906842624)**

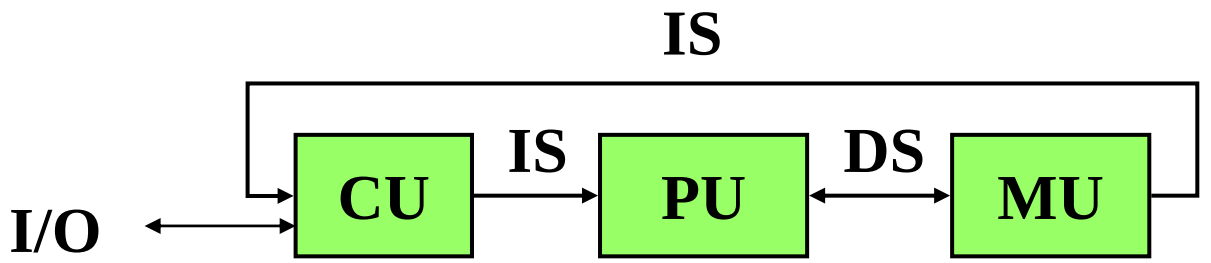
Flynn分类

- 基于指令（ instruction）和数据流（ data streams ）(1972)
  - 单指令单数据流： SISD (Single Instruction stream over a Single Data stream)
  - 单指令多数据流： SIMD (Single Instruction stream over Multiple Data streams)
  - 多指令单数据流： MISD (Multiple Instruction streams over a Single Data stream)
  - 多指令多数据流： MIMD (Multiple Instruction streams over Multiple Data stream)
- 普及程度：
  - MIMD > SIMD > MISD

SISD (Single Instruction Stream Over A Single Data Stream)

- SISD
  - 通用的串行机

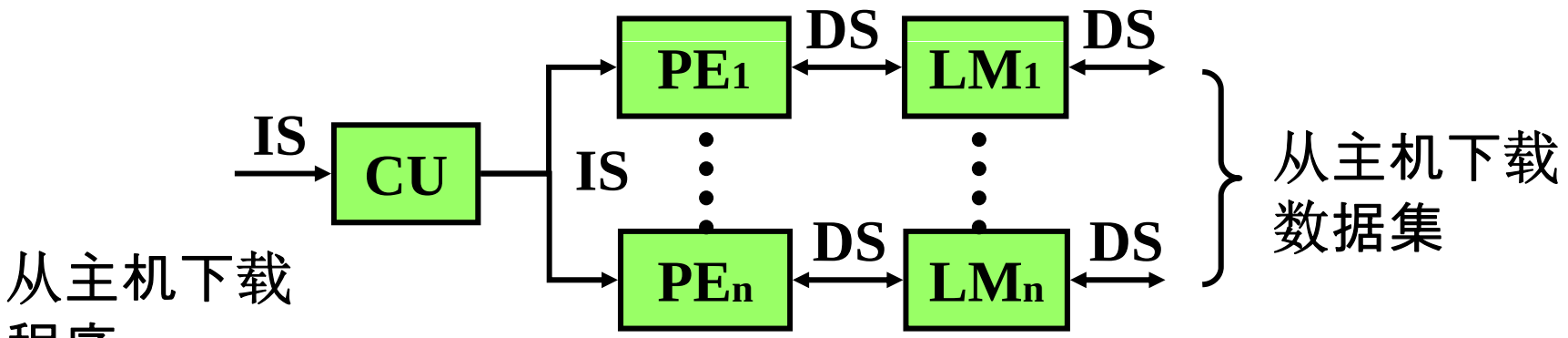
IS : 指令流（Instruction Stream）
DS : 数据流（Data Stream）
CU : 控制单元（Control Unit）
PU : 处理单元（Processing Unit）
MU : 存储单元（Memory Unit）



SIMD (Single Instruction Stream Over Multiple Data Streams)

- SIMD
  - 矢量机（Vector computers）
  - 专用计算机

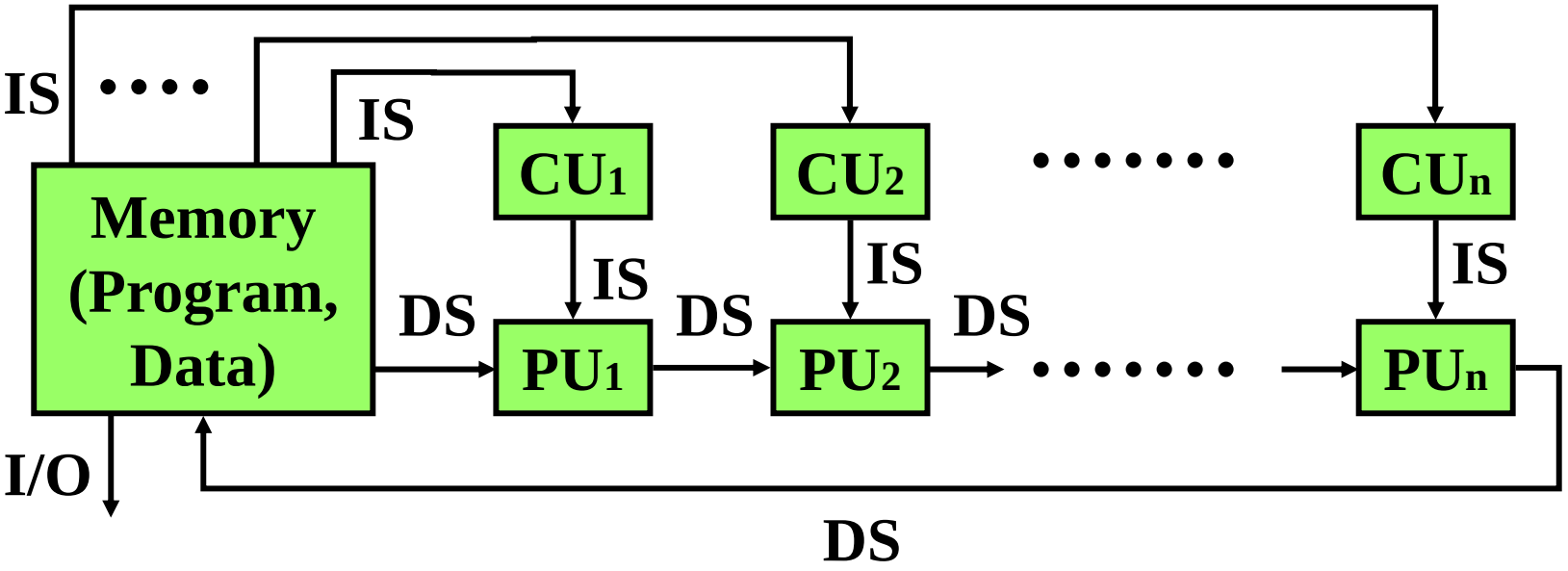
PE：处理模块（Processing Element）
LM：本地存储（Local Memory）



分布式存储的SIMD结构

MISD (Multiple Instruction Streams Over A Single Data Streams)

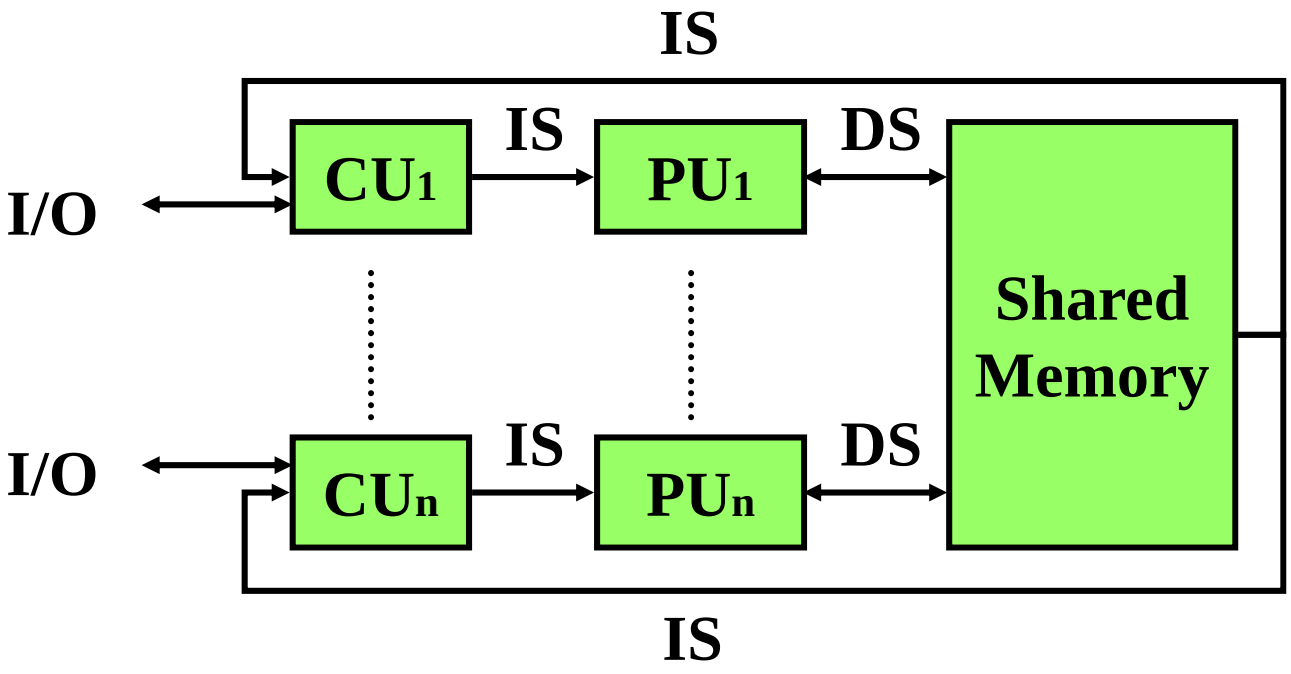
- MISD
  - 处理器阵列（Processor arrays）、脉动式阵列（systolic arrays）
  - 专用计算机



MISD结构（脉动式阵列）

MIMD (Multiple Instruction Streams Over Multiple Data Stream)

- MIMD
  - 通用的并行计算机



共享存储的MIMD结构

并行计算机体系结构

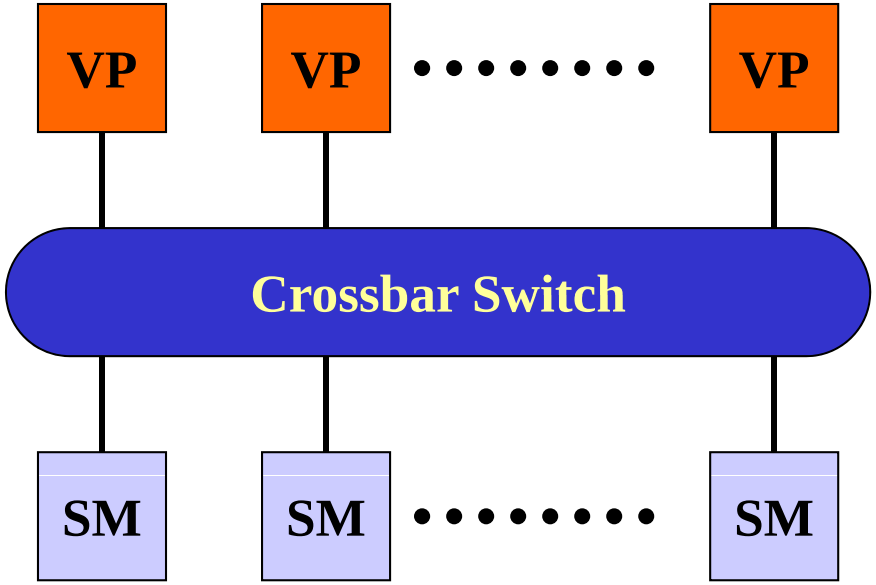
- 大部分并行计算机都是MIMD系统
  - PVP： Parallel Vector Processor
  - SMP： Symmetric Multiprocessors
  - MPP： Massively Parallel Processors
  - DSM ： Distributed Shared Memory (DSM)
  - 集群（Cluster）
  - 分布式系统（Distributed Systems）
- 可分为两种架构
  - 多处理器（ Multiprocessors ）架构
  - 多计算机（ Multicomputers ）架构

多处理器结构

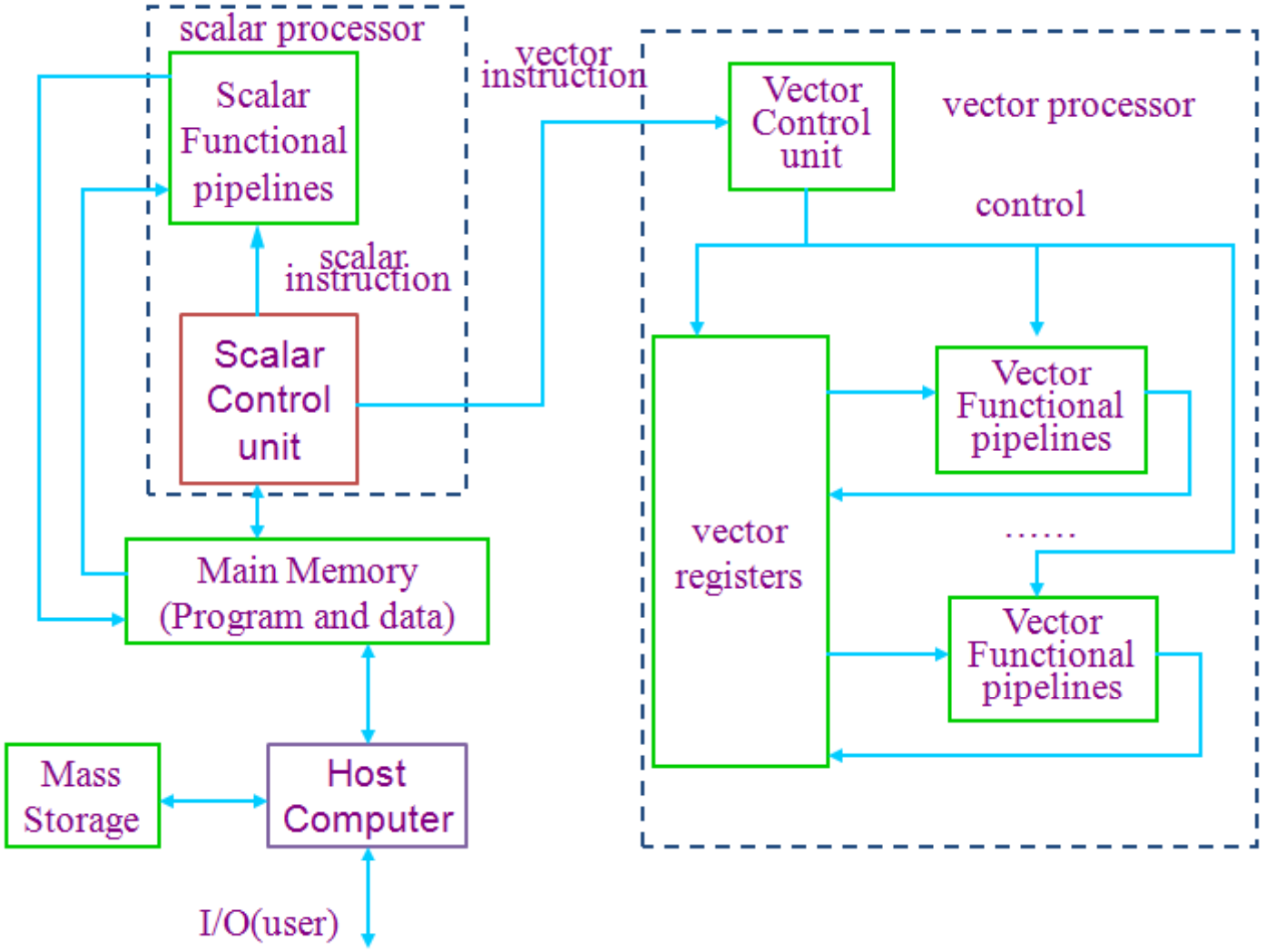
- 多处理器：共享存储空间(Shared Address Space Architecture)
  - PVP (Parallel Vector Processor)
    - 高性能的矢量（向量）处理器通过高带宽的交叉开关（crossbar switch）连接在一起
  - SMP (Symmetric Multiprocessor)
    - 商业微处理器（COTS：commercial off-the-shelf） 通过高速总线（bus）或交叉开关（crossbar switch）
  - DSM (Distributed Shared Memory)
    - 与SMP类似，但存储是物理分布在每个节点的

PVP (Parallel Vector Processor)

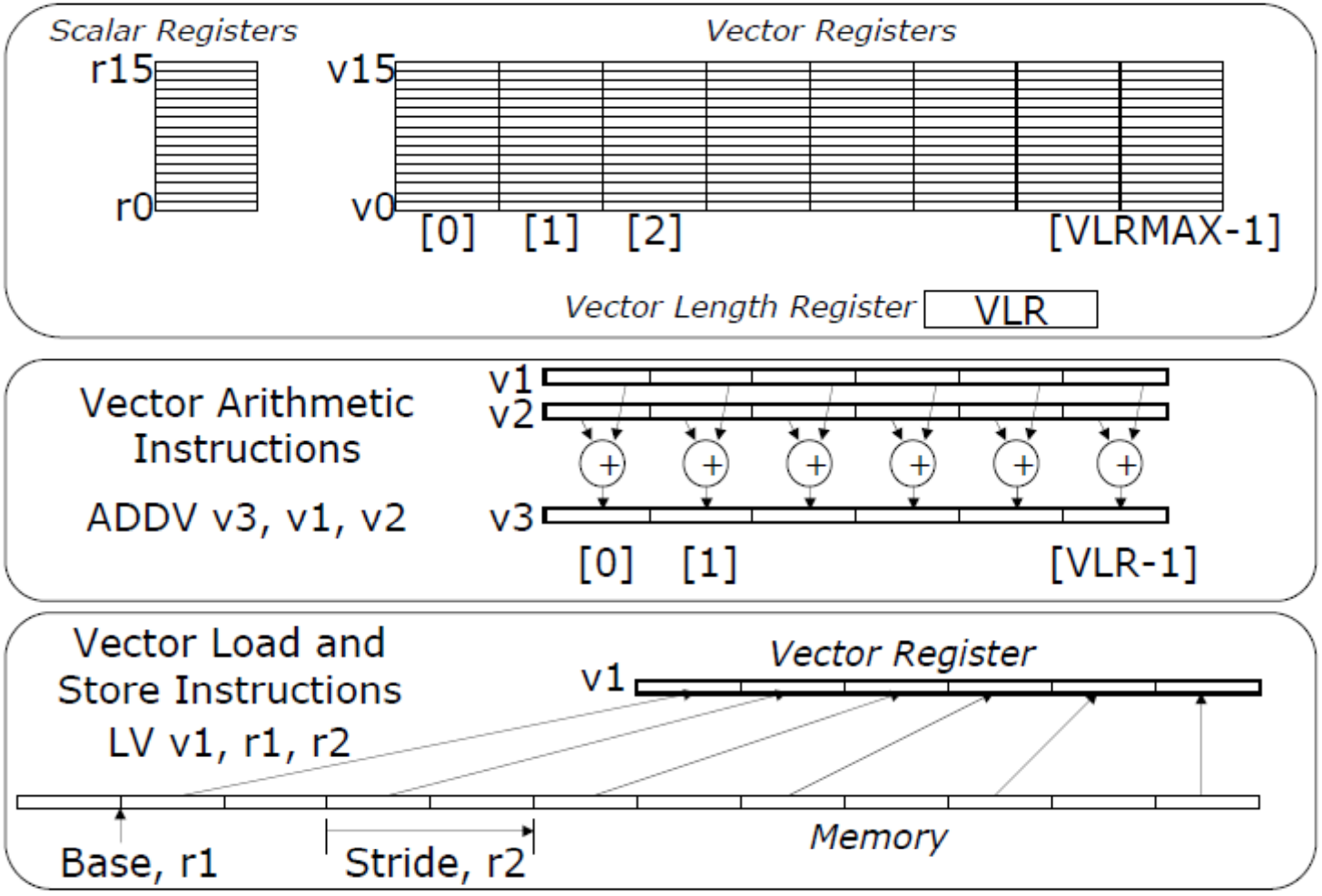
VP : 矢量处理器（Vector Processor）
SM : 共享存储（Shared Memory）



矢量机

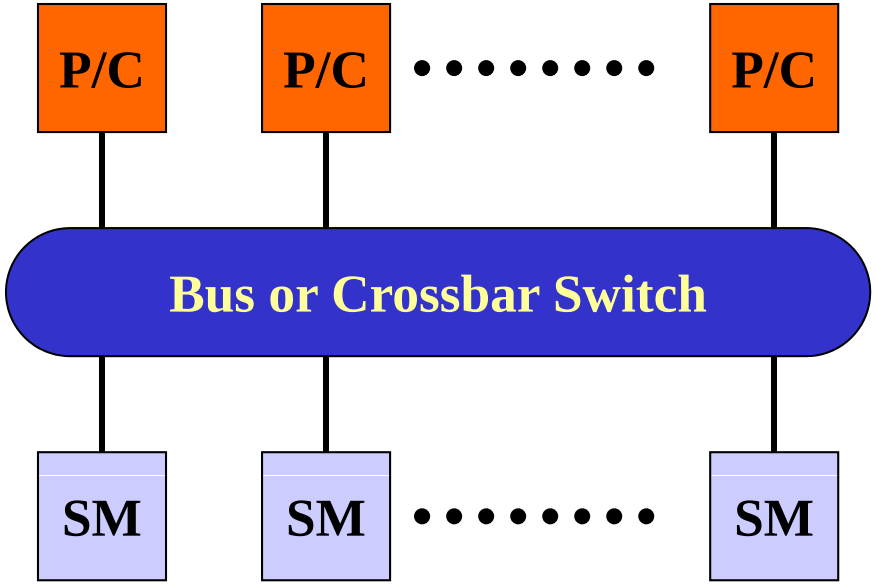

scalar processor
Scalar Functional pipelines
scalar instruction
Scalar Control unit
vector instruction
Vector Control unit
vector processor
control
vector registers
Vector Functional pipelines
.....
Vector Functional pipelines
Main Memory (Program and data)
Mass Storage
Host Computer
I/O(user)

矢量编程模型


Scalar Registers
r15
r0
Vector Registers
v15
v0
[0]
[1]
[2]
[VLRMAX-1]
Vector Length Register
VLR
Vector Arithmetic Instructions
ADDV v3, v1, v2
v1
v2
v3
[0]
[1]
[VLR-1]
Vector Load and Store Instructions
LV v1, r1, r2
Vector Register
v1
Base, r1
Stride, r2
Memory

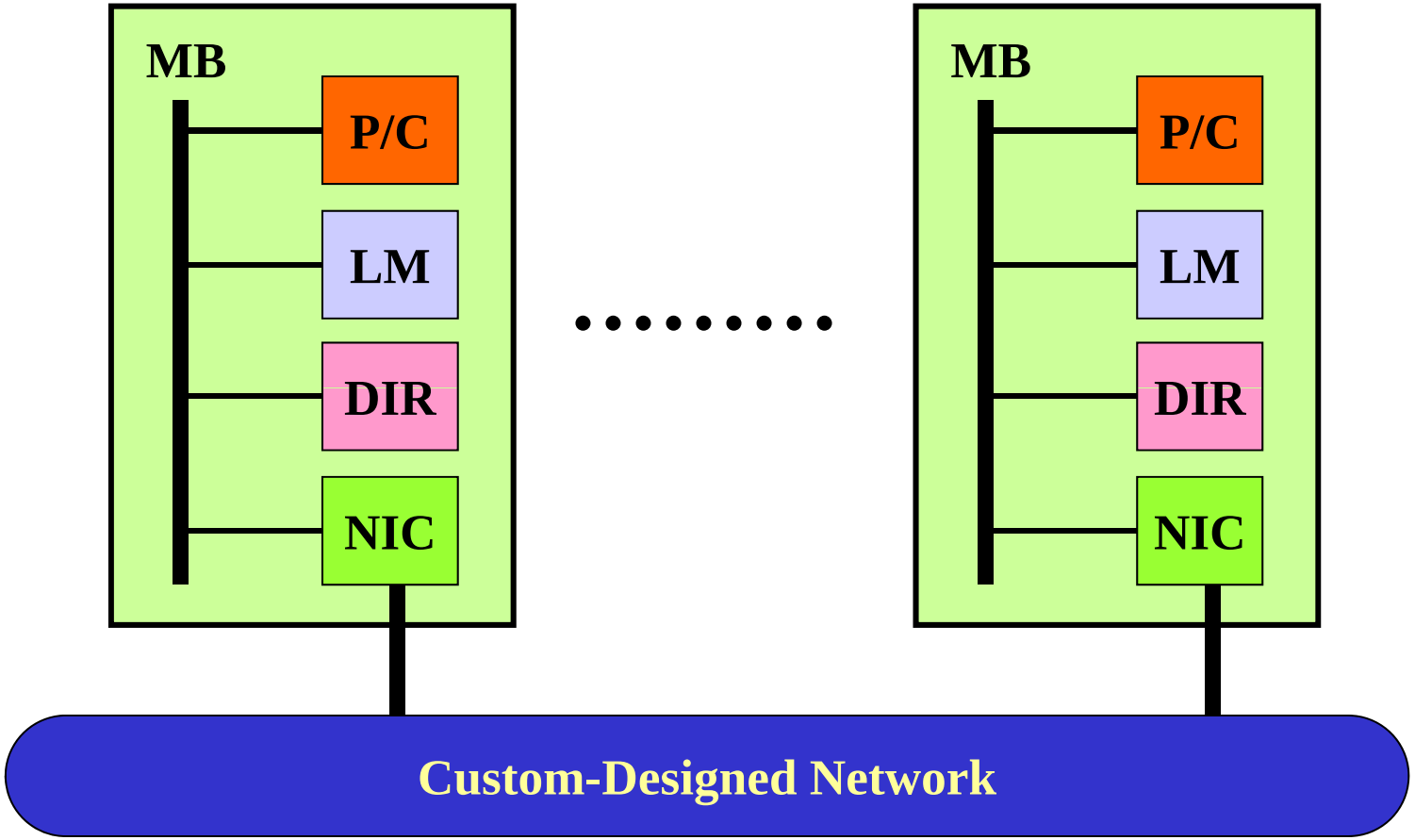
SMP (Symmetric Multi-Processor)

P/C：微处理器和缓存　（Microprocessor and Cache）



DSM (Distributed Shared Memory)

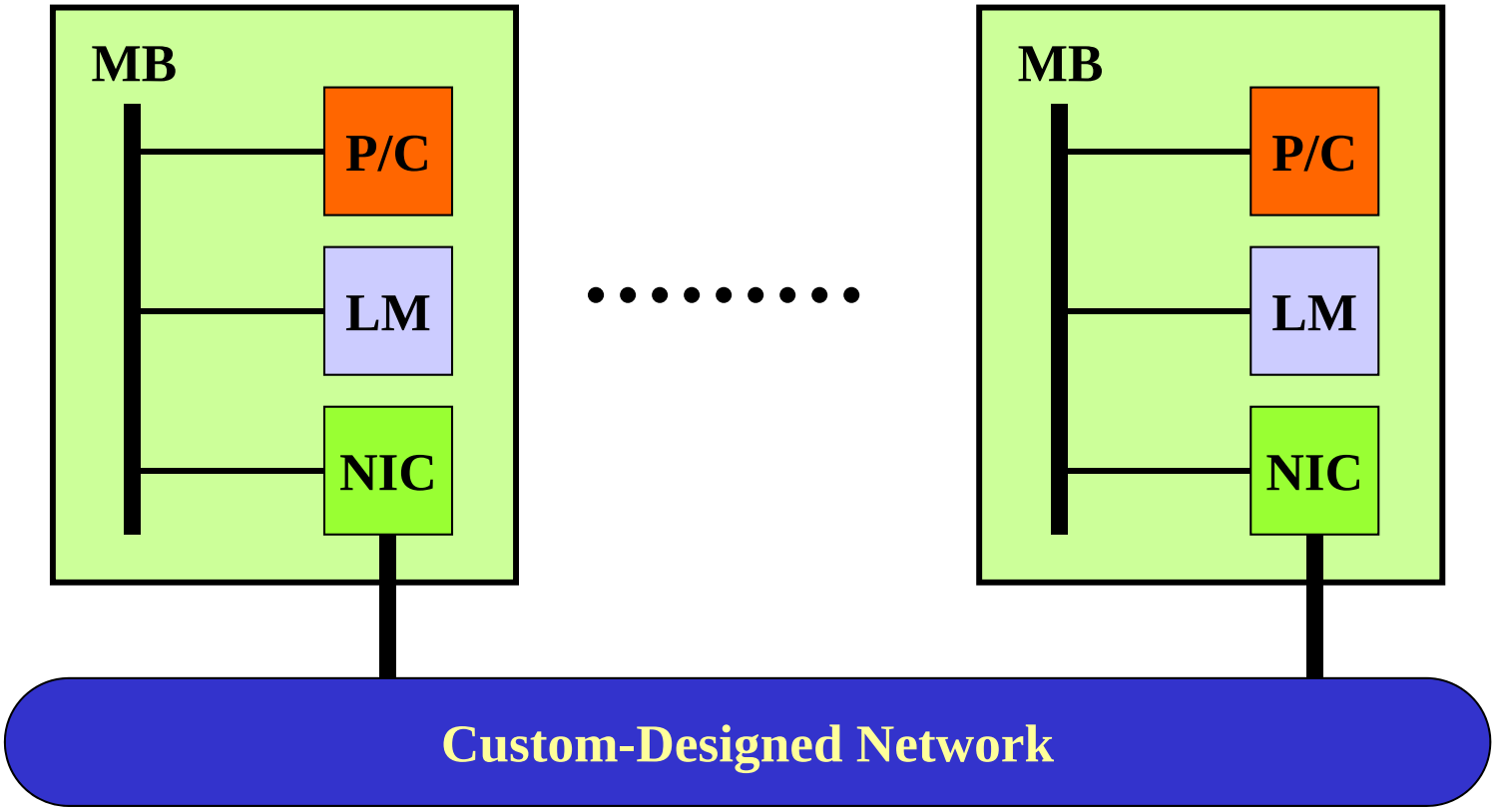
MB :存储总线（ Memory Bus），
DIR : 缓存目录（Cache Directory）
NIC : 网络接口电路（Network Interface Circuitry）



多计算机架构

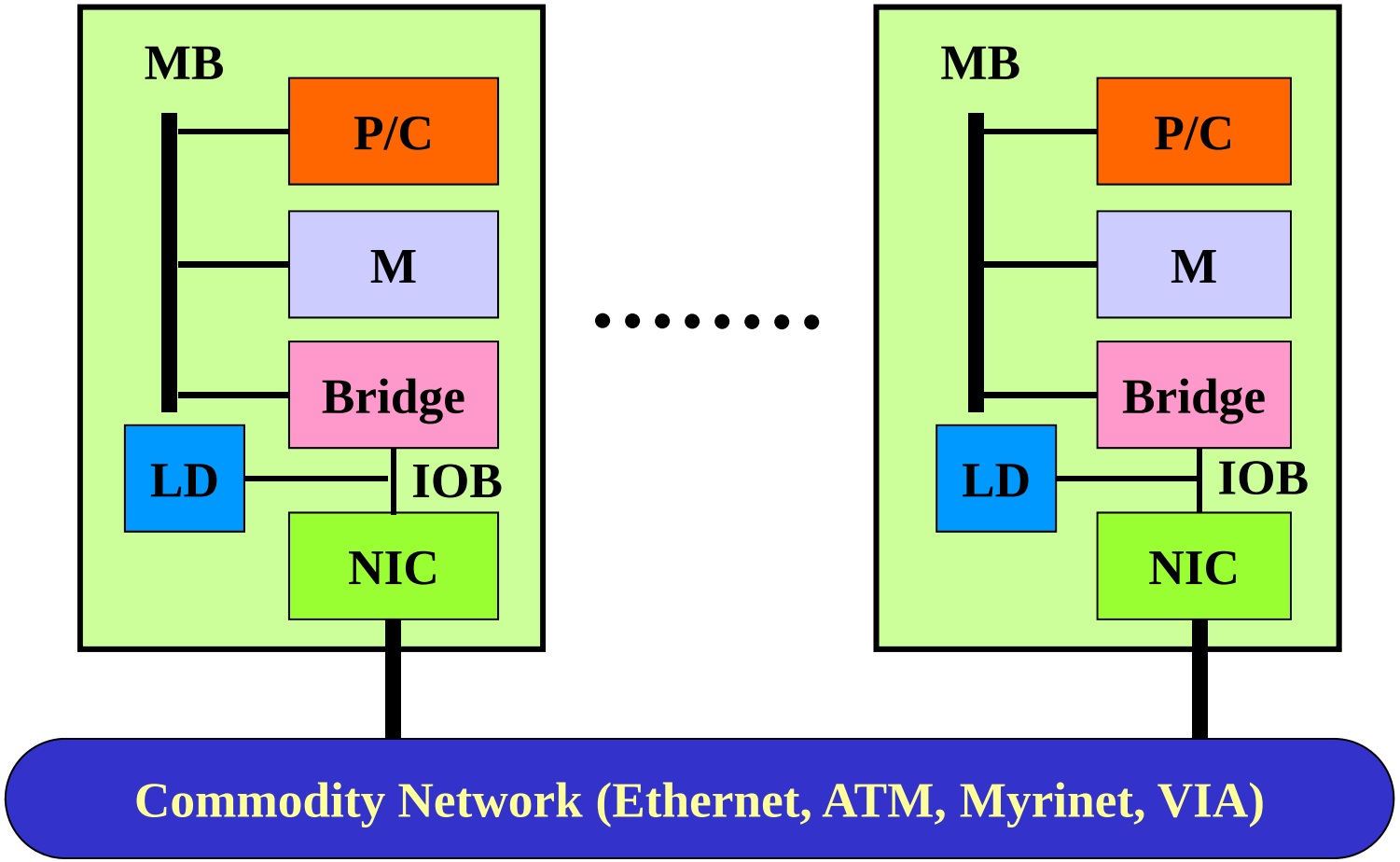
- 多计算机架构：消息传递结构（Message Passing Architecture）
- MPP (Massively Parallel Processing)
  - 处理器总数目> 1000
- Cluster
  - 系统中的每个节点的处理器少于 16个
- Constellation
  - 系统中的每个节点的处理器多于16个

MPP (Massively Parallel Processing)



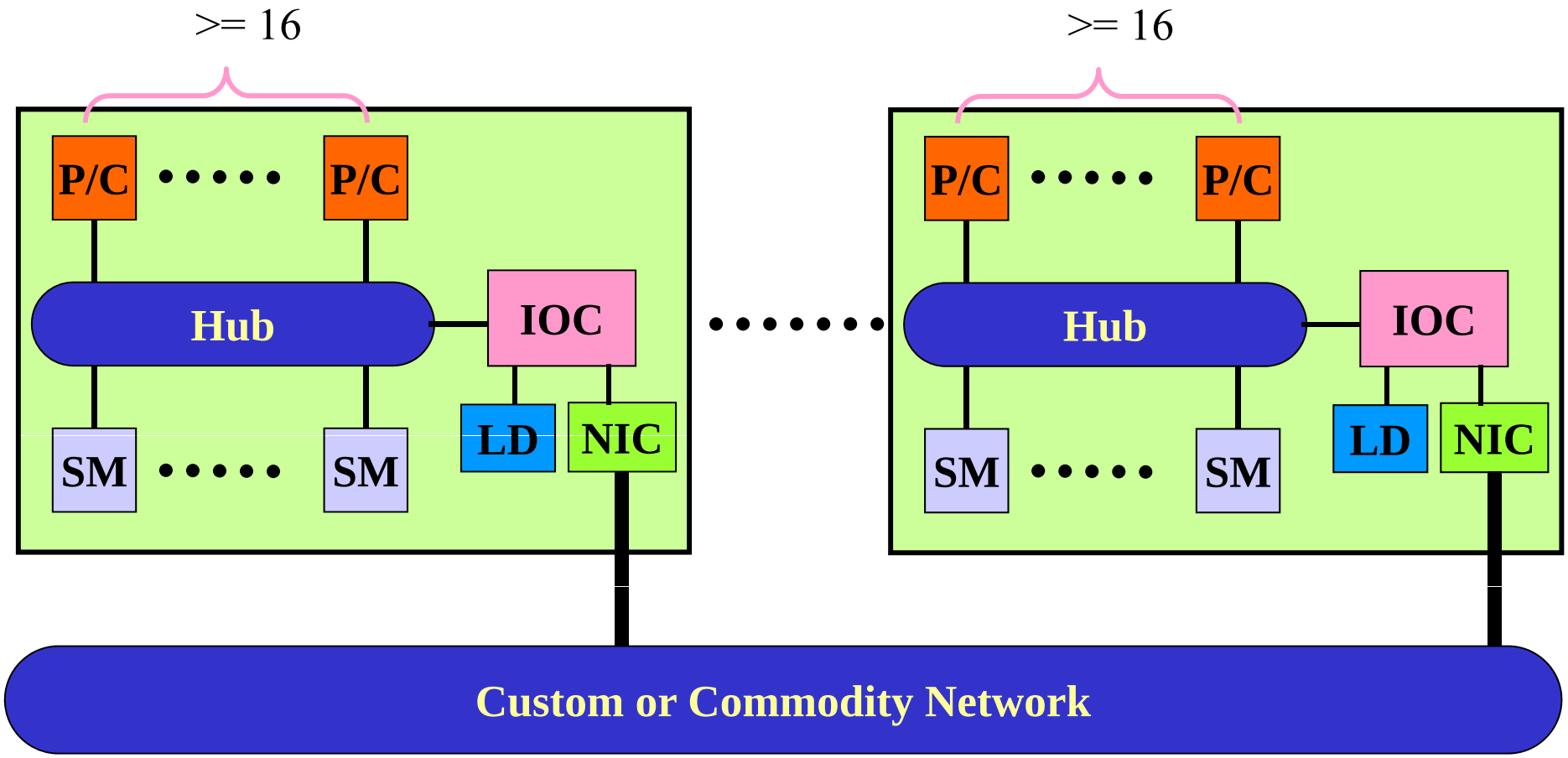
Cluster

LD：本地硬盘（Local Disk）
IOB：I/O总线（I/O Bus）



Constellation

IOC：I/O控制器（ I/O Controller）



主要内容

- 什么是高性能计算？
- 课程简介
- 术语与定义
- 高性能计算发展现状及趋势

世界上HPC计划

- 1993年美国HPCC（High Performance Computing & Communication，高性能计算和通信计划）：3T性能目标
- 1996年美国ASCI （ Accelerated Strategic Computing Initiative，加速战略计划创新）计划： 为主要设备供应商制造超大（Teraflop）计算机系统，主要用于仿真核武器实验

- 欧洲：ESPRIT, Alvey, Parallel Applications Programme, Europort, Fourth Framework etc
- 日本：日本的HPC计划，造就了许多大型的并行矢量机（NEC, Hitashi, Fujitsu）
- 澳大利亚APSC（Partnership for Advanced Computing ）计划：提供HPC的设备和培训

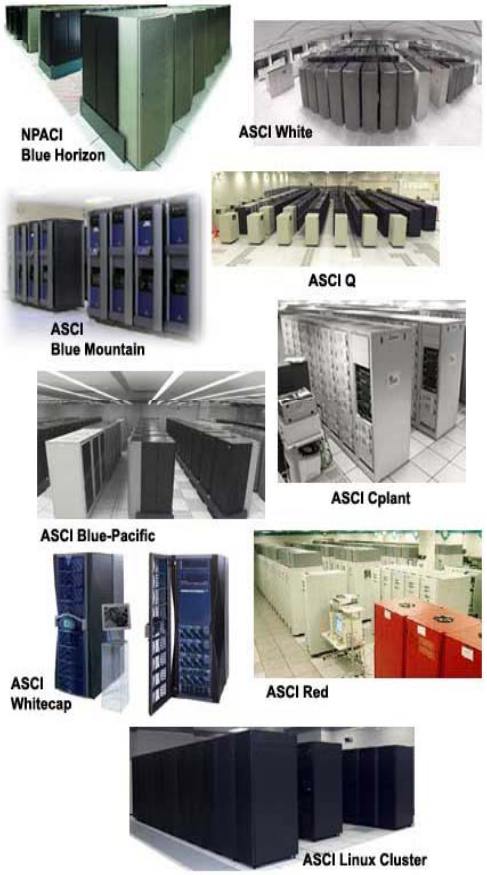
美国HPCC 计划

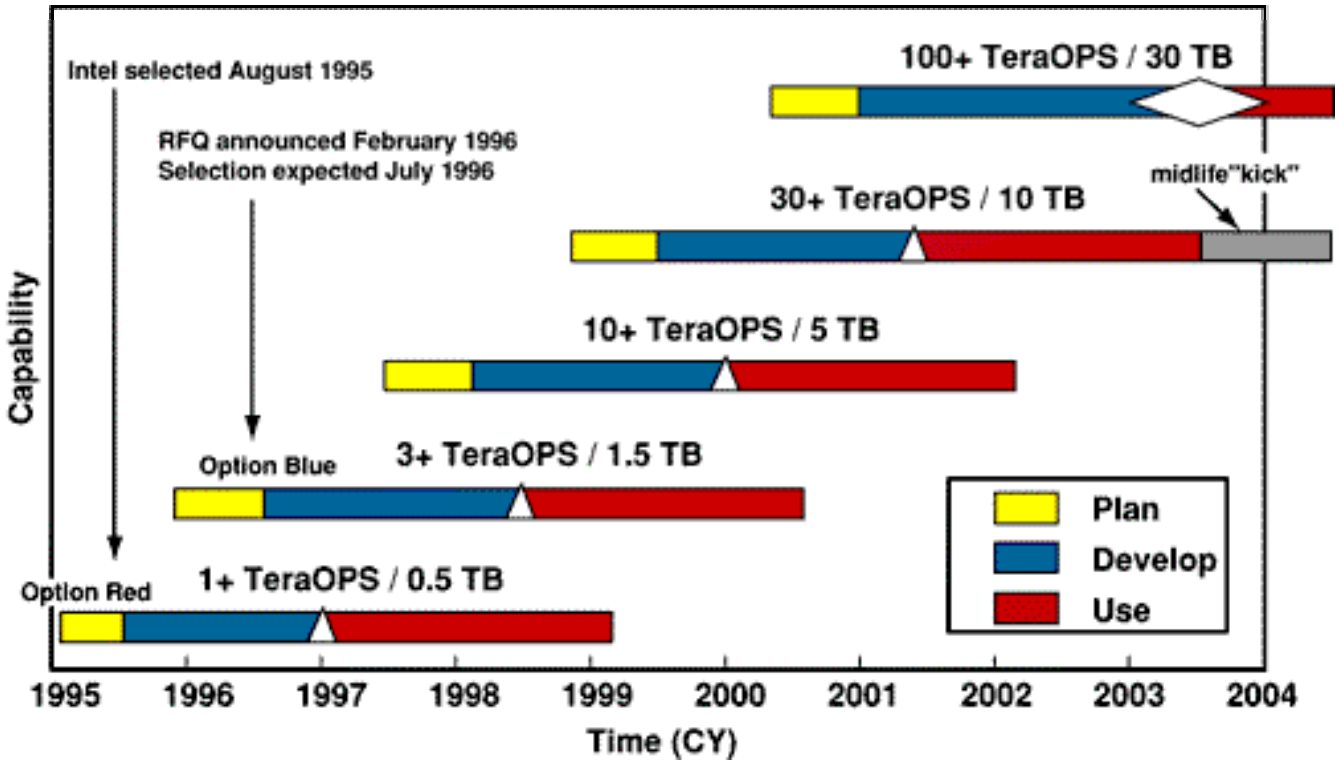
- HPCC二期 - June 2000
  - 显示并应用高性能计算环境来扩展我们的认识和能力，预测影响地球、太阳系和宇宙的物理、化学和生物过程
- 国家HPC Office (http://www.hpcc.gov）
- 重大挑战性课题（Grand Challenge Applications）（表1.1）
  - 大气、半导体、生物信息等
- 1993年：3T性能指标
- 2000年：1Pflops (1015)计算能力

ASCI计划

- ASCI：加速战略计划创新

（http://www.llnl.gov/asci/）





中国高性能计算机发展

- **1983**年，“银河Ⅰ号”巨型计算机研制成功，运算速度达每秒**1**亿次
- **1984**年，中国第一台**10**亿次巨型银河计算机Ⅱ型通过鉴定
- **1995**年，曙光**1000**大型机通过鉴定，其峰值可达每秒**25**亿次
- **2000**年，“神威**-I**”高性能计算机问世，峰值达每秒**384**亿次。我国成为继美国、日本之后，世界上第三个具备研制高性能计算机能力的国家。（1998年11月美国已研制出每秒3.1万亿次机）
- **2002**年**8**月，中科院第一台每秒万亿次的超级计算机联想深腾**1800**问世。在当年**TOP500**中排名**43**。（2001年美国已研制出每秒12.8万亿次机）
- **2003**年**11**月联想公司研制的“联想深腾**6800**高性能计算平台”，系统峰值**5.3**万亿次，当年**Top500**排名**14**（2002年4月，日本NEC公司研制出当时世界上运算速度最快的超级计算机“地球模拟器”，运算峰值为40万亿次）
- **2004**年**6**月曙光公司的研制的“曙光**4000A**”，系统峰值**10**万亿次，当年**Top500**排名**10**
- **2008**年**8**月，曙光百万亿次超级计算机“曙光**5000**”研制成功，成为第二个可研制百万亿次超级计算机的国家
- **2009**年**10**月，国防科大“天河一号”研制成功，在**Top500**排名第五
- **2010**年**5**月，曙光“星云”研制成功，在**Top500**排名第二，成为第二个可研制千万亿次超级计算机的国家，

TOP 500
（www.top500.org）

列出世界500强超级计算机，每年更新两次

TOP 10 (2012.6)

Rank	Site	Computer/Year Vendor	Cores	R_{max}	R_{peak}	Power
1	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom / 2011 IBM	1572864	16324.75	20132.66	7890.0
2	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect / 2011 Fujitsu	705024	10510.00	11280.38	12659.9
3	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	786432	8162.38	10066.33	3945.0
4	Leibniz Rechenzentrum Germany	**SuperMUC** - iDataPlex DX360M4, Xeon E5-2680 8C 2.70GHz, Infiniband FDR / 2012 IBM	147456	2897.00	3185.05	3422.7
5	National Supercomputing Center in Tianjin China	**Tianhe-1A** - NUDT YH MPP, Xeon X5670 6C 2.93 GHz, NVIDIA 2050 / 2010 NUDT	186368	2566.00	4701.00	4040.0
6	DOE/SC/Oak Ridge National Laboratory United States	**Jaguar** - Cray XK6, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA 2090 / 2009 Cray Inc.	298592	1941.00	2627.61	5142.0
7	CINECA Italy	**Fermi** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	163840	1725.49	2097.15	821.9
8	Forschungszentrum Juelich (FZJ) Germany	**JuQUEEN** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	131072	1380.39	1677.72	657.5
9	CEA/TGCC-GENCI France	**Curie thin nodes** - Bullx B510, Xeon E5-2680 8C 2.700GHz, Infiniband QDR / 2012 Bull	77184	1359.00	1667.17	2251.0
10	National Supercomputing Centre in Shenzhen (NSCS) China	**Nebulae** - Dawning TC3600 Blade System, Xeon X5650 6C 2.66GHz, Infiniband QDR, NVIDIA 2050 / 2010 Dawning	120640	1271.00	2984.30	2580.0

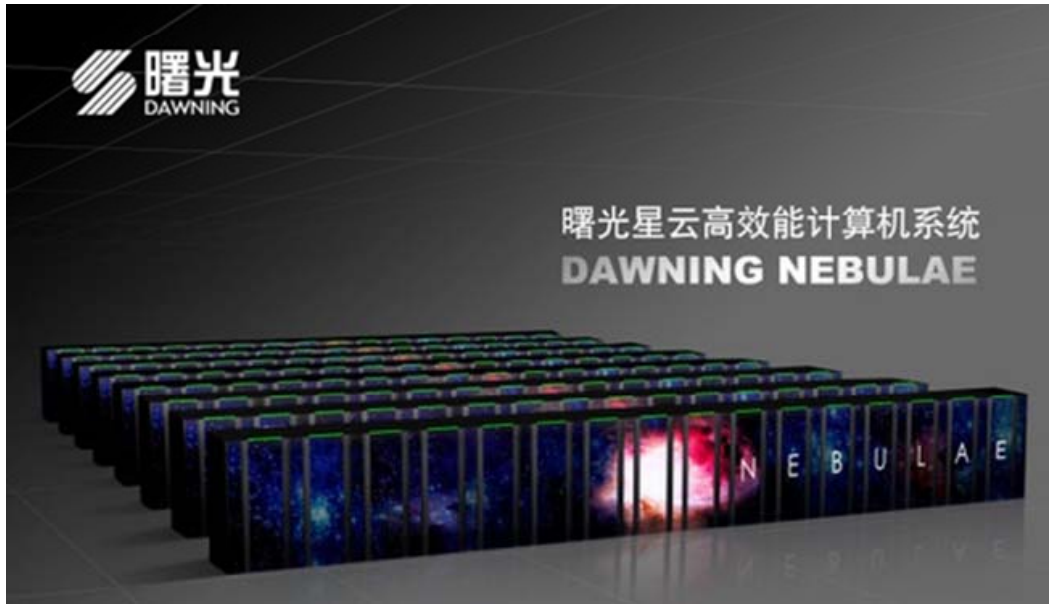
R_{max}：实测性能 (TFlops)，R_{peak}：理论峰值 (TFlops)，Power：电源消耗（KW）

Jaguar @ ORNL: 1.75 PF/s

- **Cray XT5-HE系统**
- **37,500 个4 核AMD Opteron 2.6 GHz处理器, 224,162 cores.**
- **存储：300 TB**
- **硬盘空间：10 PB**
- **硬盘带宽：240 GB/s**
- **互联网络：Cray's SeaStar2+**

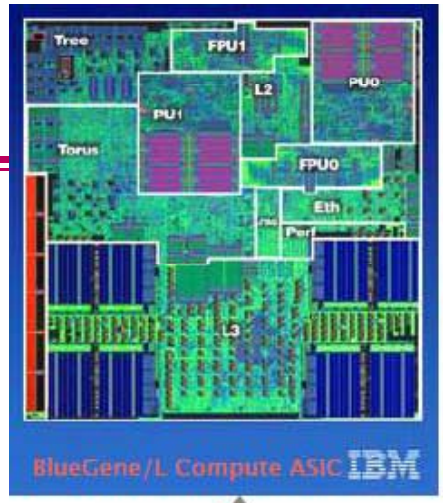
星云（Nebulae）

- 国内首台实测性能超千万亿次的超级计算机，每秒系统峰值达三千万亿次**(3PFlops)**，实测**Linpack**值达**1271**万亿次（**1.2PFlops**）
- **CPU**：**Intel Westmere**，**9280**个
- **GPU**： **Nvidia Fermi**，**4640**个
- 互联网络：**InfiniBand**
- 每瓦能耗：**498 MFLOPS/W**
- 占地：**600m²**
- 地点：国家超算深圳中心



IBM BlueGene/L

131,072 Processors



System
(64 racks, 64x32x32)
131,072 procs

Rack
(32 Node boards, 8x8x16)
2048 processors

Node Card
(32 chips, 4x4x2)
16 Compute Cards
64 processors

Compute Card
(2 chips, 2x1x1)
4 processors

Chip
(2 processors)

180/360 TF/s
32 TB DDR

2.9/5.7 TF/s
0.5 TB DDR

90/180 GF/s
16 GB DDR

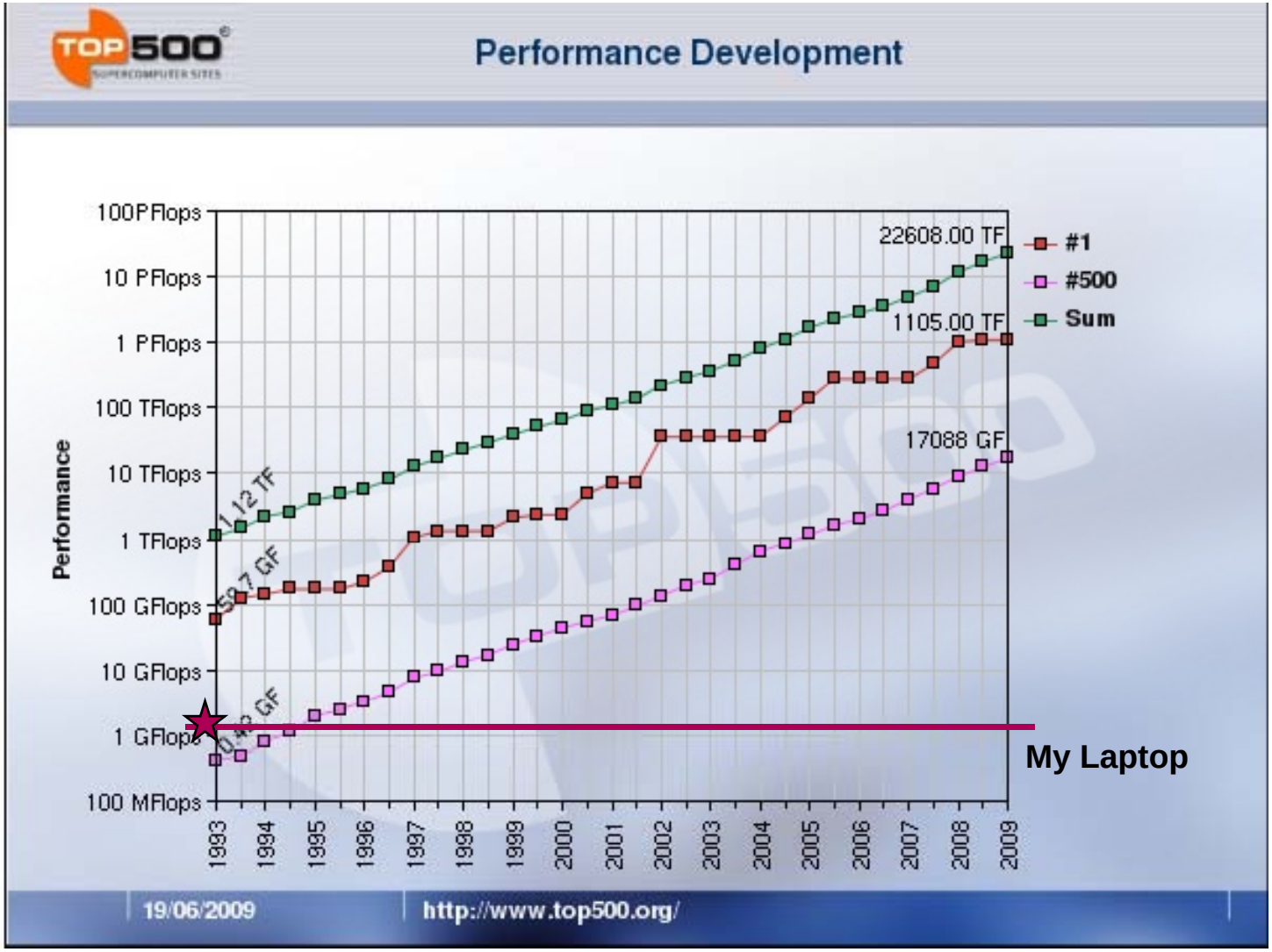
5.6/11.2 GF/s
1 GB DDR

2.8/5.6 GF/s
4 MB (cache)

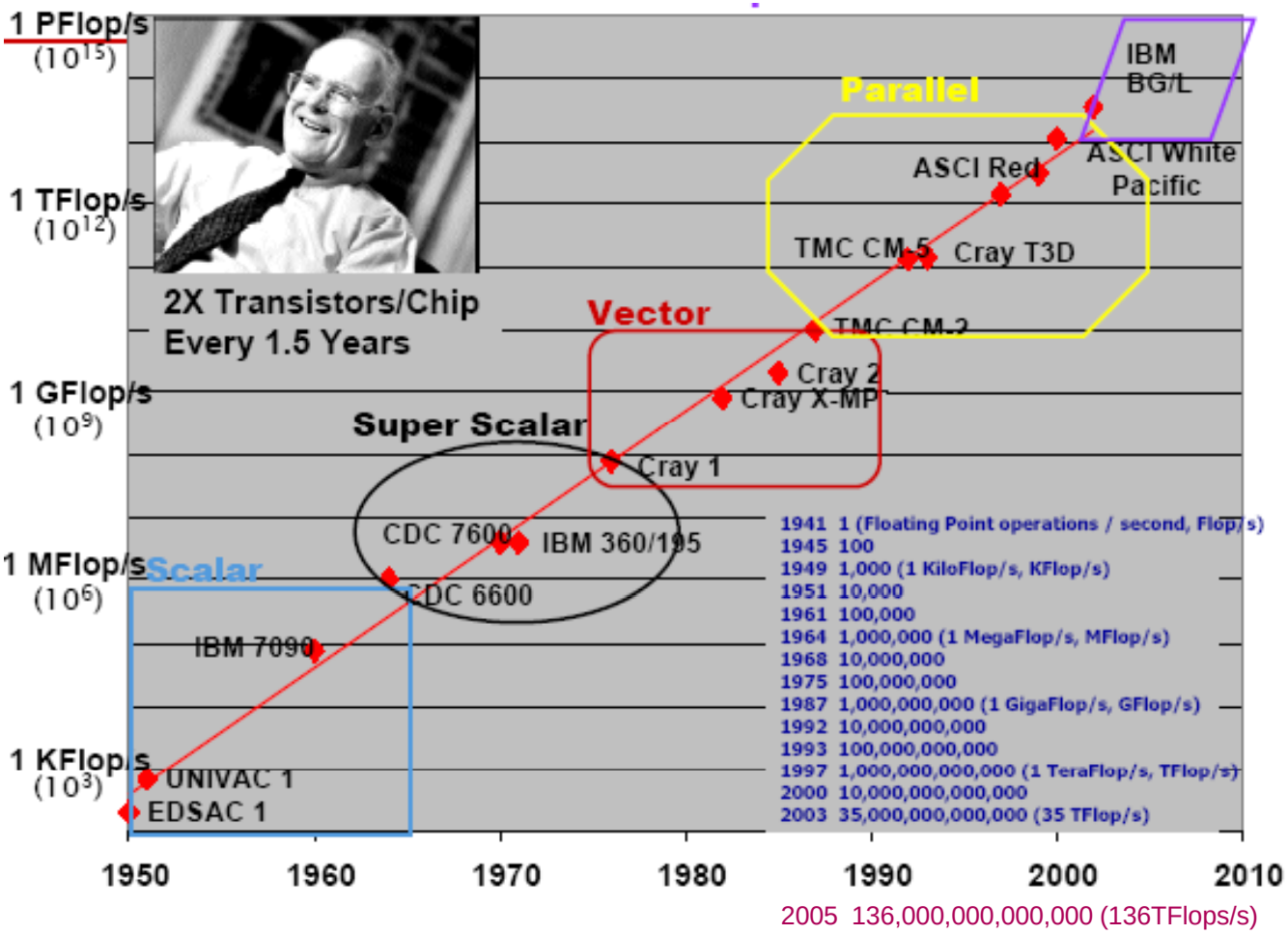
Full system total of
131,072 processors

"Fastest Computer"
BG/L 700 MHz 16K proc
8 racks
Peak: 45.9 Tflop/s
Linpack: 36.0 Tflop/s
78% of peak

透过Top500看HPC的发展趋势



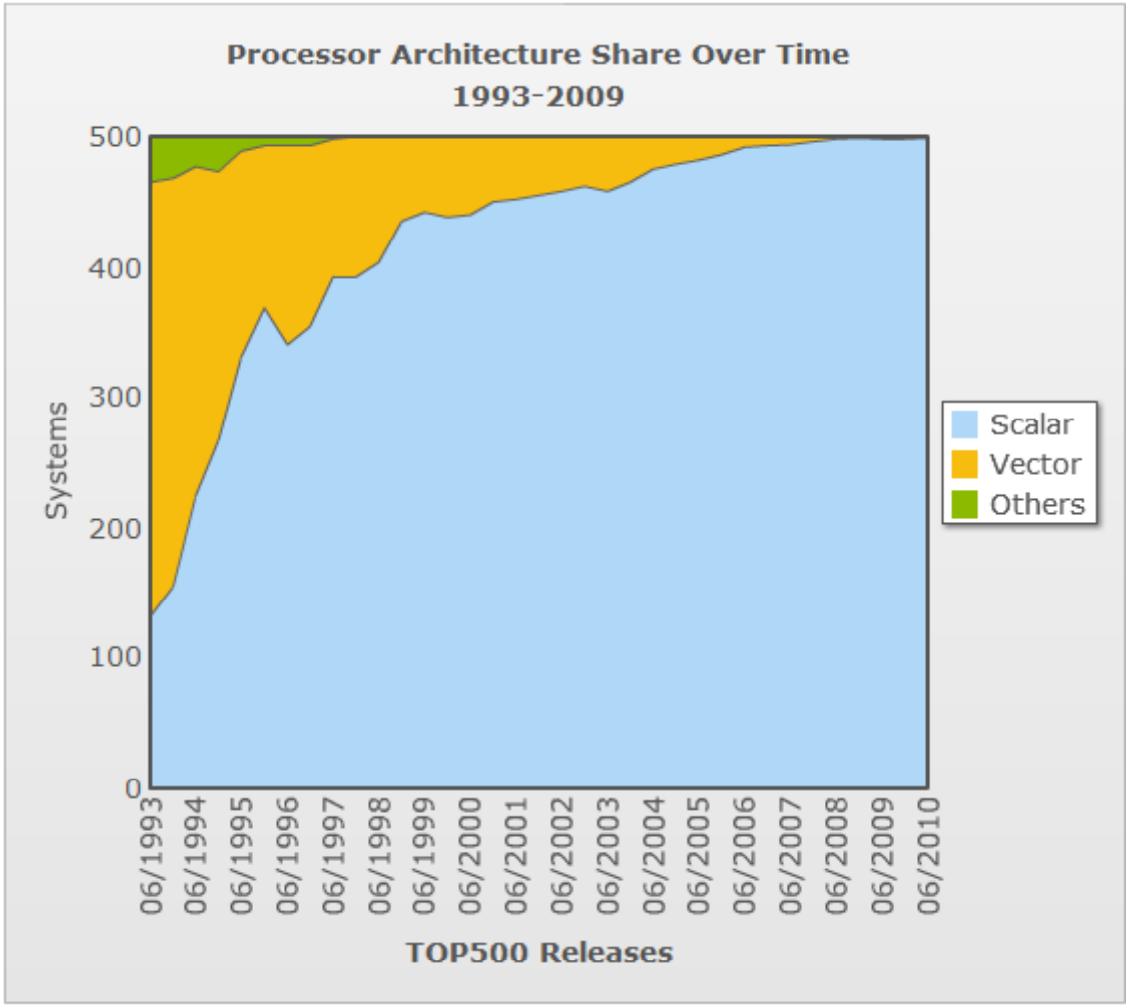
HPC历史发展与摩尔定律



1998(Tflops) →2008(Pflops) →2018(Eflops)

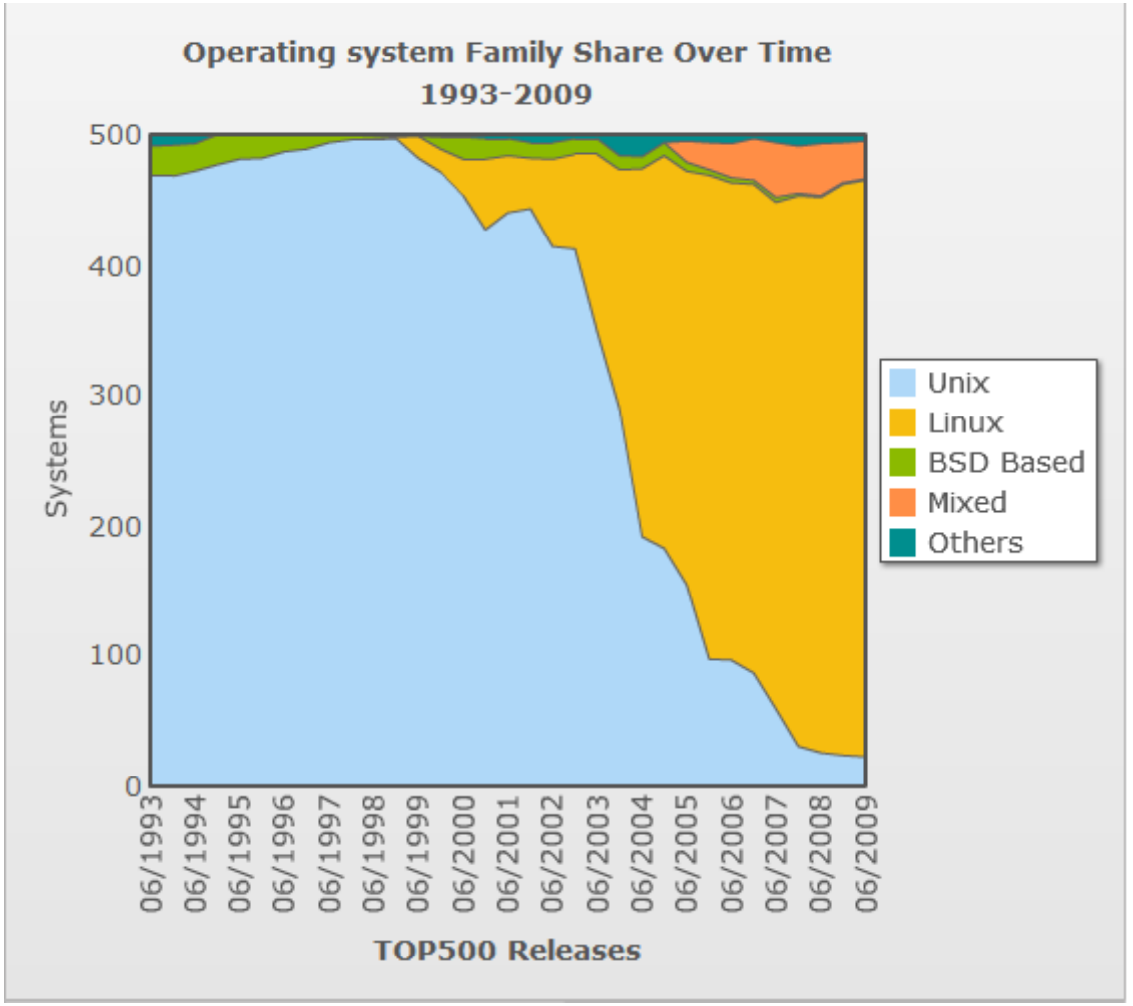
世界最快计算机每10年1000倍性能提升

处理器架构发展趋势



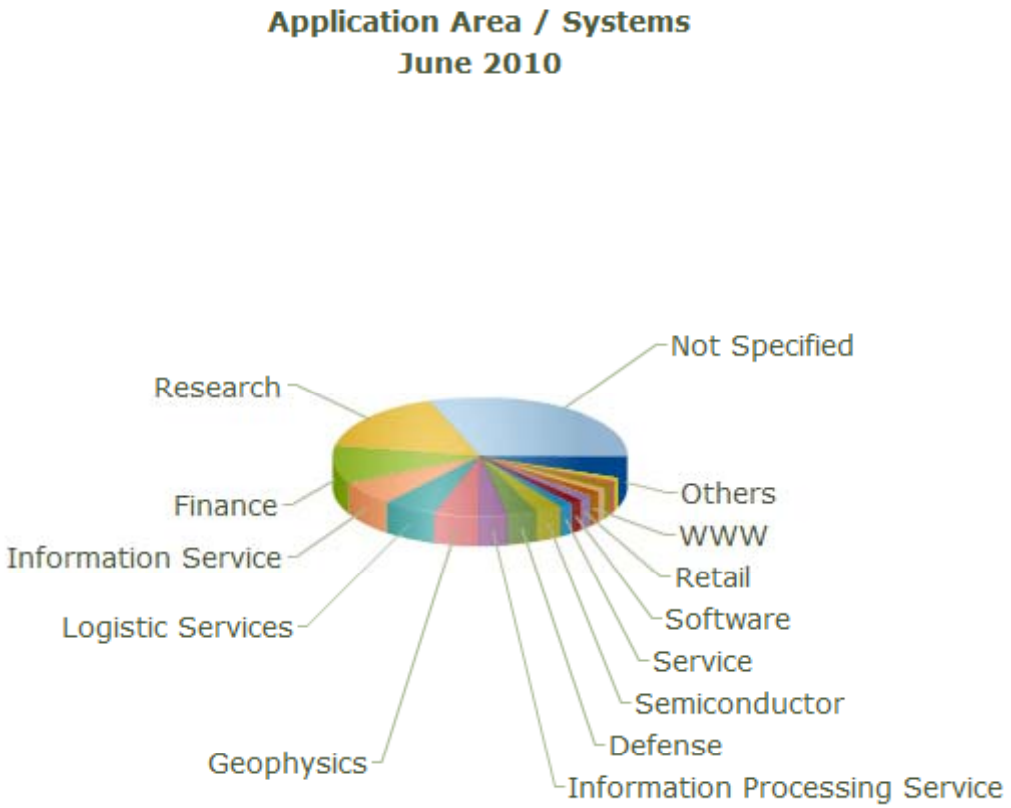
Vector: 0.2%, Scalar: 99.8% (2010.6)

操作系统发展趋势



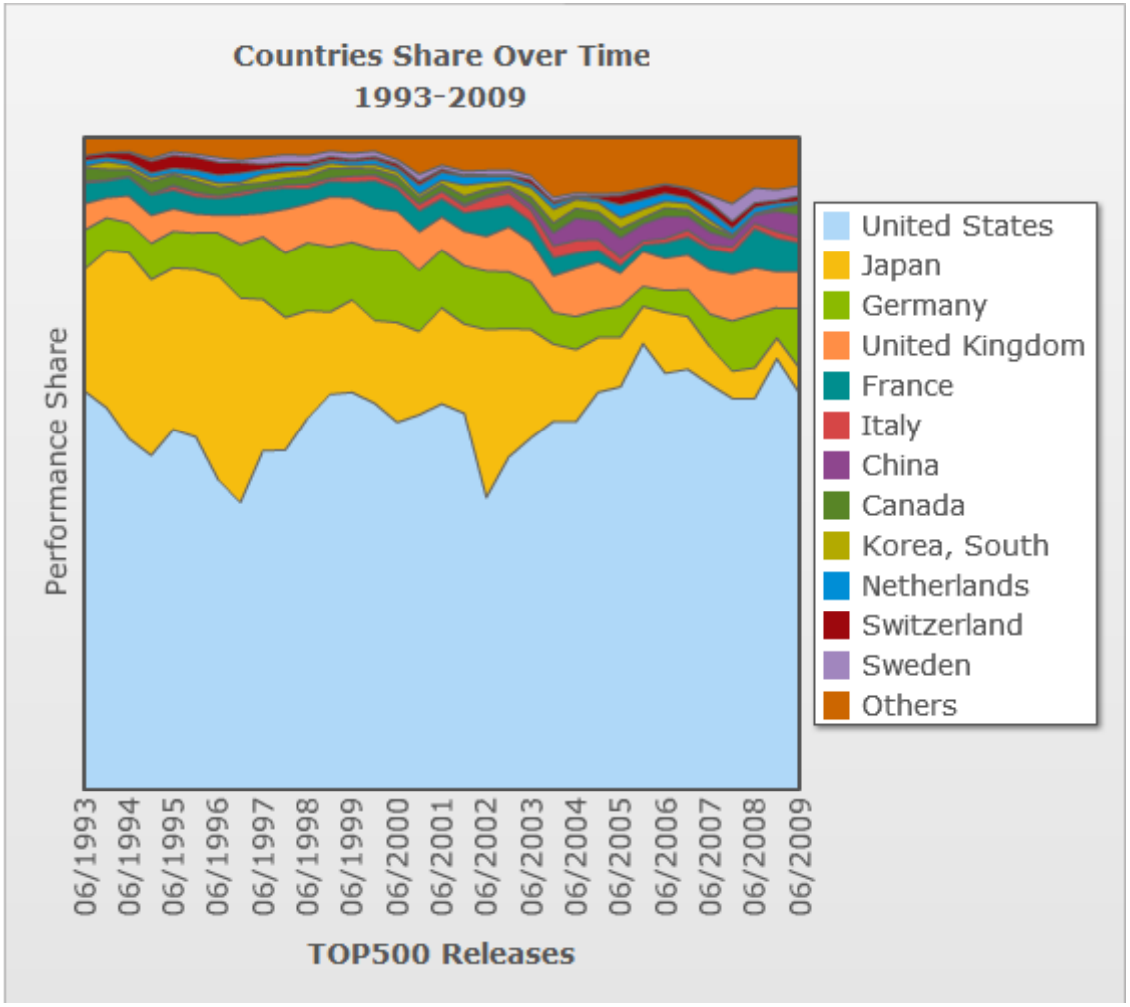
Linux: 91%, Mixed: 4.4%, Unix: 3.4%, Windows: 1% (2010.6)

应用领域的性能分布



非专用的：**30.8%**，研究： **16.4%**， 金融：**10.6%**，信息服务：**6.6%** （**2010.6**）

各个国家的性能分布



美国: 56.4%, 英国：7.6%，德国: 4.8%，法国4.6%，中国：4.8%，日本: 3.6%　(2010.6)

Top100 China 2012.2

Rank	Vendor	System	Installation Location	Application area	Num of Cores	Linpack (Gflops)	Peak(Gflops)	Efficiency
1	NUDT	TianHe OneA/7168x2 Intel Hexa Core Xeon X5670 2.93GHz + 7168 Nvidia Tesla M2050@1.15GHz+2048 Hex Core FT-1000@1GHz/ 80Gbps	National Supercomputer Center in Tianjin	Scientific Computing/Industry	202752	2566000	4701000	0.546
2	NPCETR	SunwayBlueLight /8575x16 Core SW1600@975MHz/QDR Infiniband	National Supercomputer Center in Jinan	Scientific Computing/Industry	137200	795900	1070160	0.744
3	NUDT	TianHe OneA-HN/2048x2 Intel Hexa Core Xeon X5670 2.93GHz + 2048 Nvidia Tesla M2050@1.15GHz/ 80Gbps	National Supercomputer Center in Changsha	Scientific Computing/Education	53248	771700	1343200	0.575
4	SUGON	SUGON NEBULA/ TC3600 Blade/2560x(2 Intel Hexa Core X5650 + Nvidia Tesla C2050 GPU)/QDR Infiniband	National Supercomputer Center in Shenzhen	Scientific Computing/Industry	52416	749200	1296320.26	0.578
5	IBM	xSeries x3650M3/Intel Xeon X56xx 2.53 GHz/Giga-E	Network Company	Internet Service	113040	636985	1143965	0.557

中国软件行业协会数学软件分会和国家863高性能计算机评测中心联合发布

中国Top 100 的应用领域(2009.11）

应用领域	数量（套）	份额	Linpack[GF/s]	峰值[GF/s]	平均效率	处理器数
能源	20	20%	254314.86	481548.42	54.39%	48176
游戏	15	15%	325516.00	592010.00	55.13%	62836
科学计算	14	14%	269046.67	359545.36	76.79%	32144
气象	10	10%	114200.37	152169.64	77.70%	16800
工业	9	9%	813762.46	1535635.16	70.76%	66792
政府部门	9	9%	112323.00	207515.96	55.88%	25674
教育	7	7%	95487.47	117343.52	79.56%	12108
电信	6	6%	100129.80	181019.78	57.77%	19849
生物信息	4	4%	34914.00	58464.40	62.68%	5596
地震	2	2%	37372.00	50066.08	76.15%	4608
电子商务	1	1%	16190.00	35200.00	46.00%	3520
动漫渲染	1	1%	12115.26	22131.20	54.70%	2080
金融保险	1	1%	9287.00	17920.00	51.80%	1600
视频计算	1	1%	9196.00	19200.00	47.90%	2400
总计	100	100%	220384.89	3829769.52	64.13%	304183

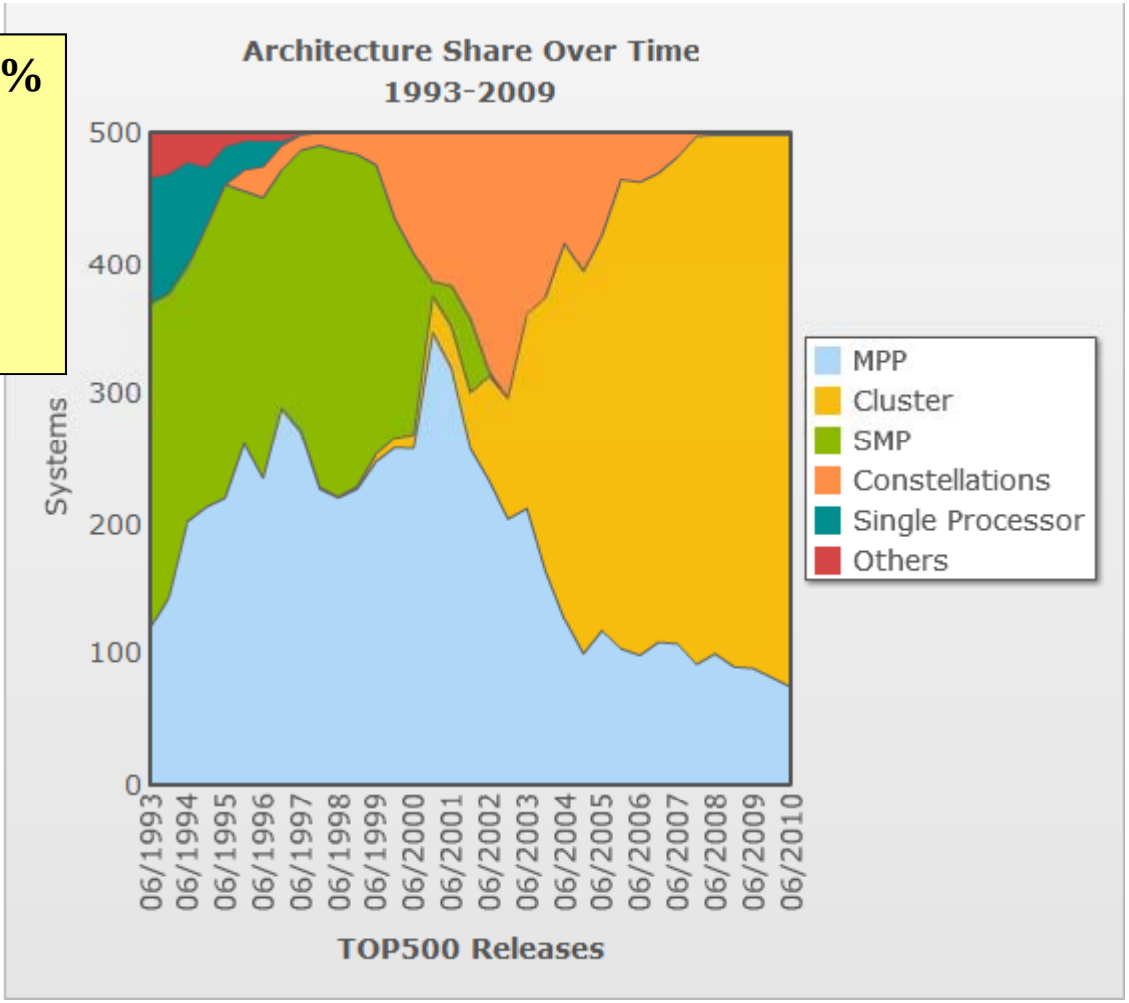
计算机系统体系结构发展趋势

Constellation:0.4%

MPP: 14.8%

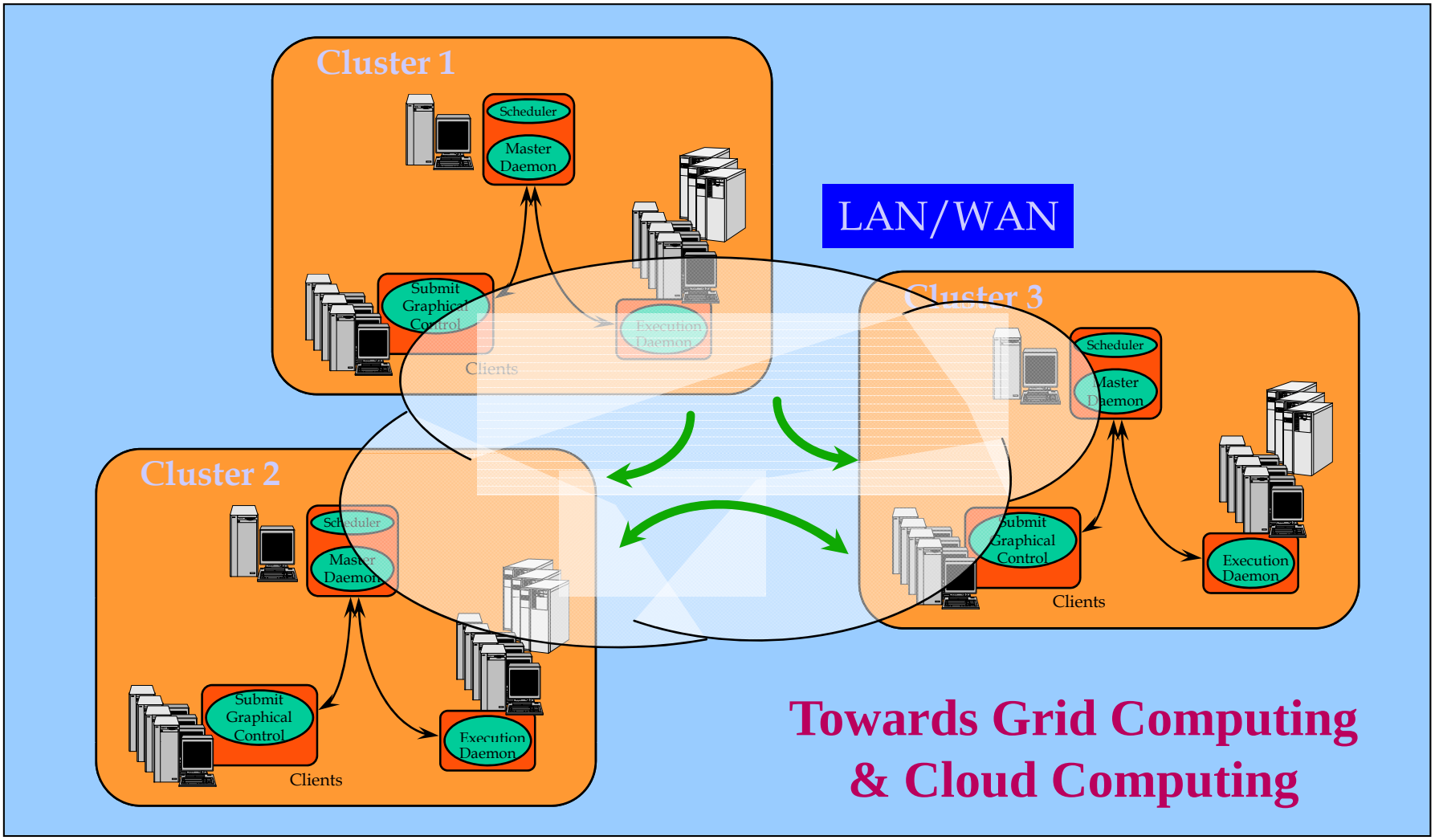
Cluster: 84.8%

(2009.6)



体系结构相对稳定，**Cluster**占有率不断提高

Cluster of Clusters



课程小结

- 高性能计算的需求
- 术语和定义
  - HPC, HPCC, Parallel Computing, Distributed Computing, Grid Computing,Metacomputing,Utility Computing
  - Cornerstone:Compute, Storage,Communication
  - Units of measurement
  - SISD,SIMD, MISD, MIMD
  - PVP, SMP, MPP,DSM,Cluster,Constellation
- 发展趋势
  - 体系结构、处理器、应用领域等

推荐读物和网站

- 阅读：
  - 《并行计算—结构、算法、编程》第一章

- 网站：
  - TOP500 Supercomputer Sites
    http://www.top500.org

下一讲

- 高性能计算机体系结构
  - 《并行计算—结构、算法、编程》第1，2章