

第1-1章

定量设计与分析
基础

- 1.1 引言
- 1.2 计算机的类别
- 1.3 定义计算机体系结构以及计算机设计的任务是什么？
- 1.4 技术趋势
- 1.5 集成电路的功耗趋势
- 1.6 成本趋势
- 1.7 可靠性
- 1.8 性能的测量、报告与总结
- 1.9 计算机设计的定量原则
- 1.10 综合应用

什么是计算机体系结构

- 什么是计算机体系结构 (CA) ?
- 计算机体系结构 (CA) 设计师的任务是什么？

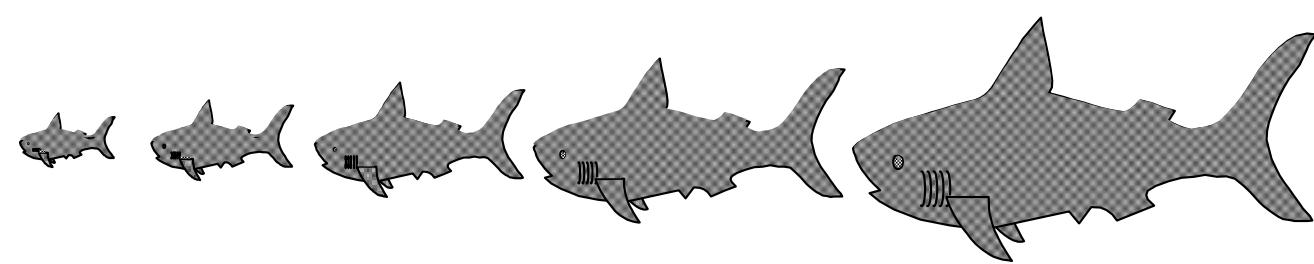
- 帕特森的演讲：“计算机体系结构的50年”
>><https://www.bilibili.com/video/av46710093/>
- 计算的未来——与约翰·亨尼西的对话
>><https://www.bilibili.com/video/av23283946/>

- 技术
> 从长远来看，提升CPU性能的关键因素。

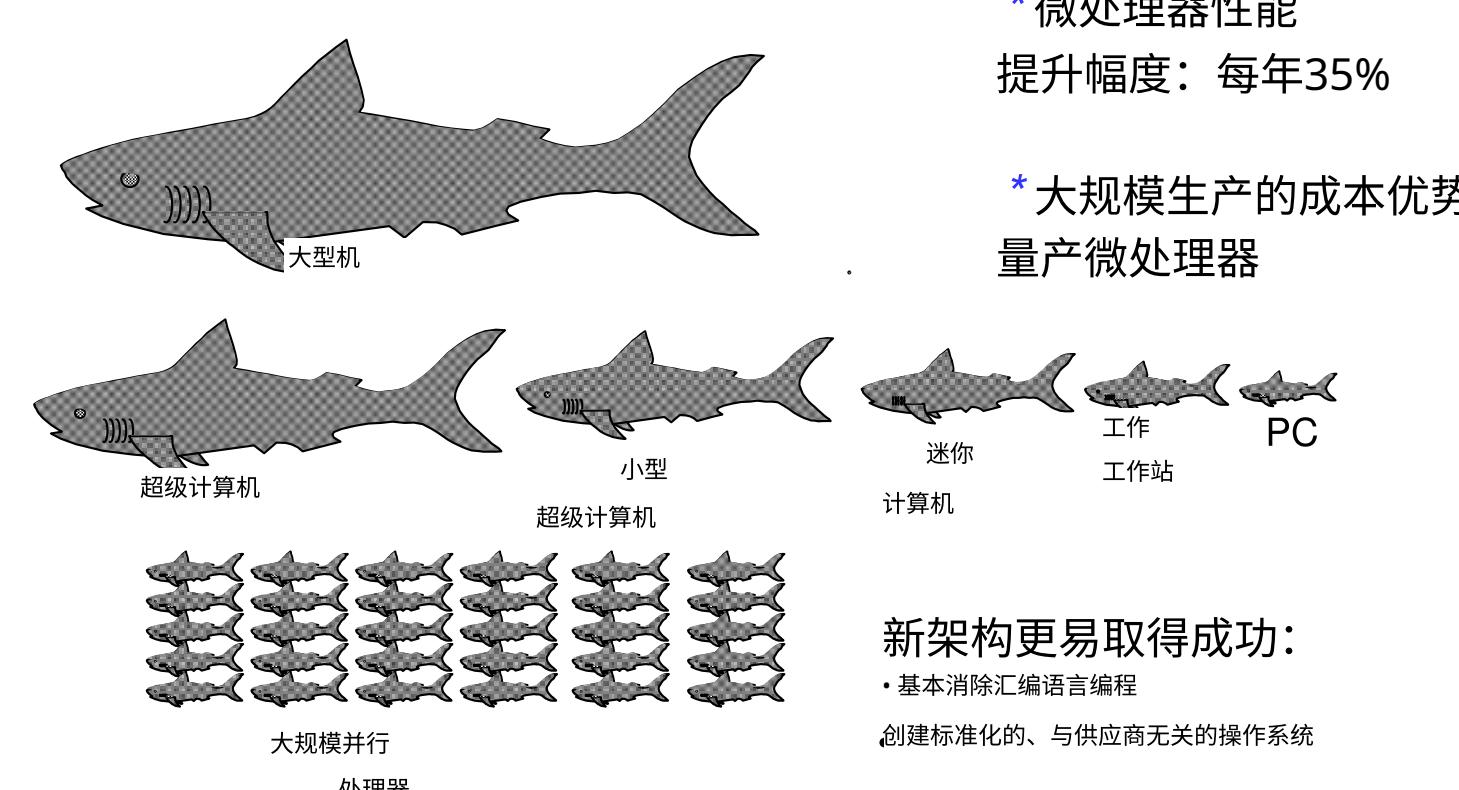
计算机市场的演变

- 最初情况：

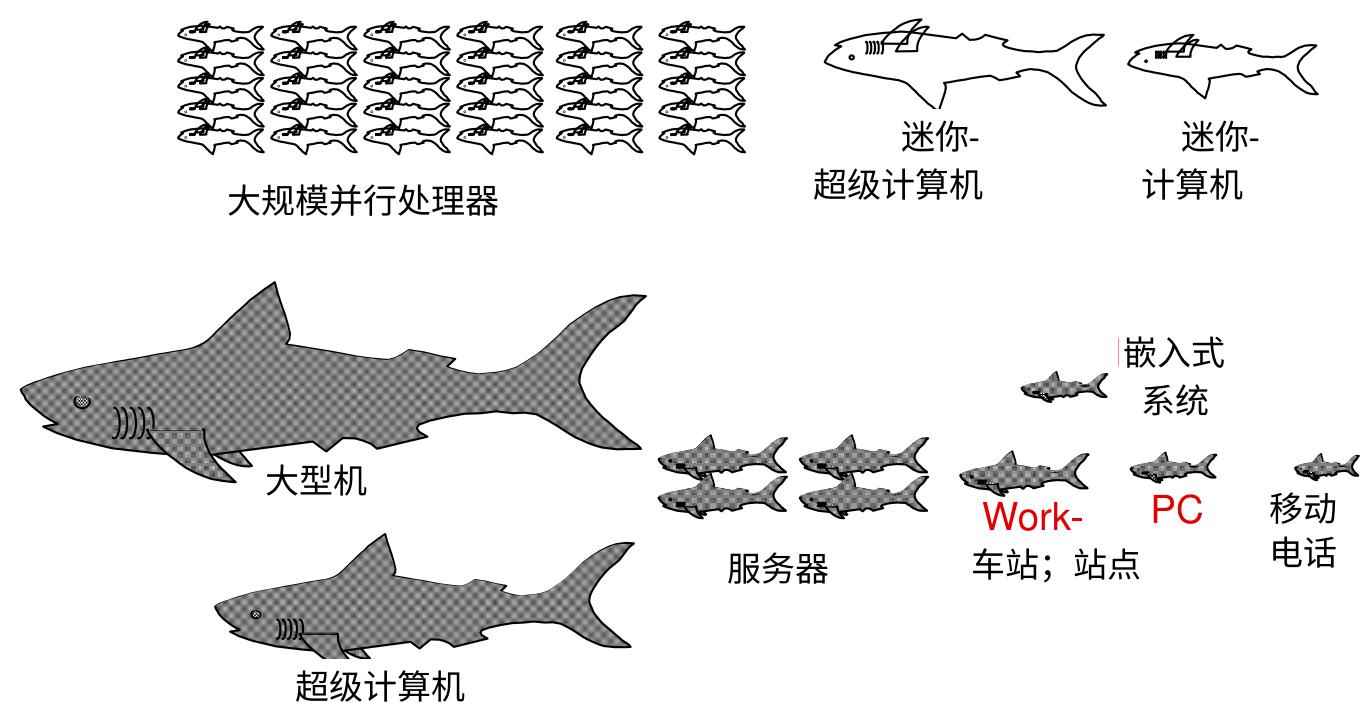
大鱼吃小鱼



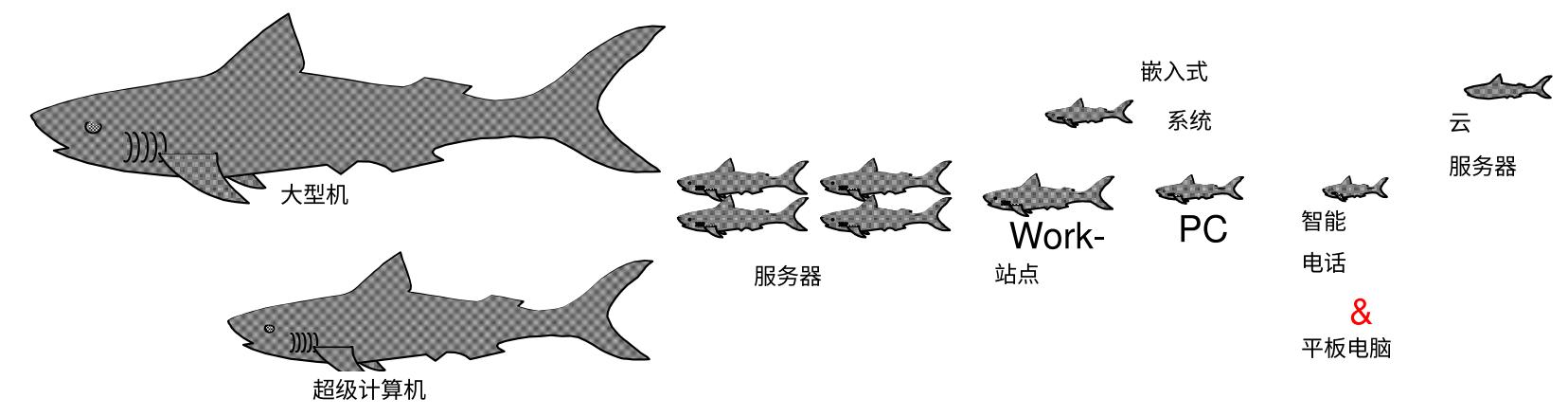
计算机市场的演变



计算机市场的演变

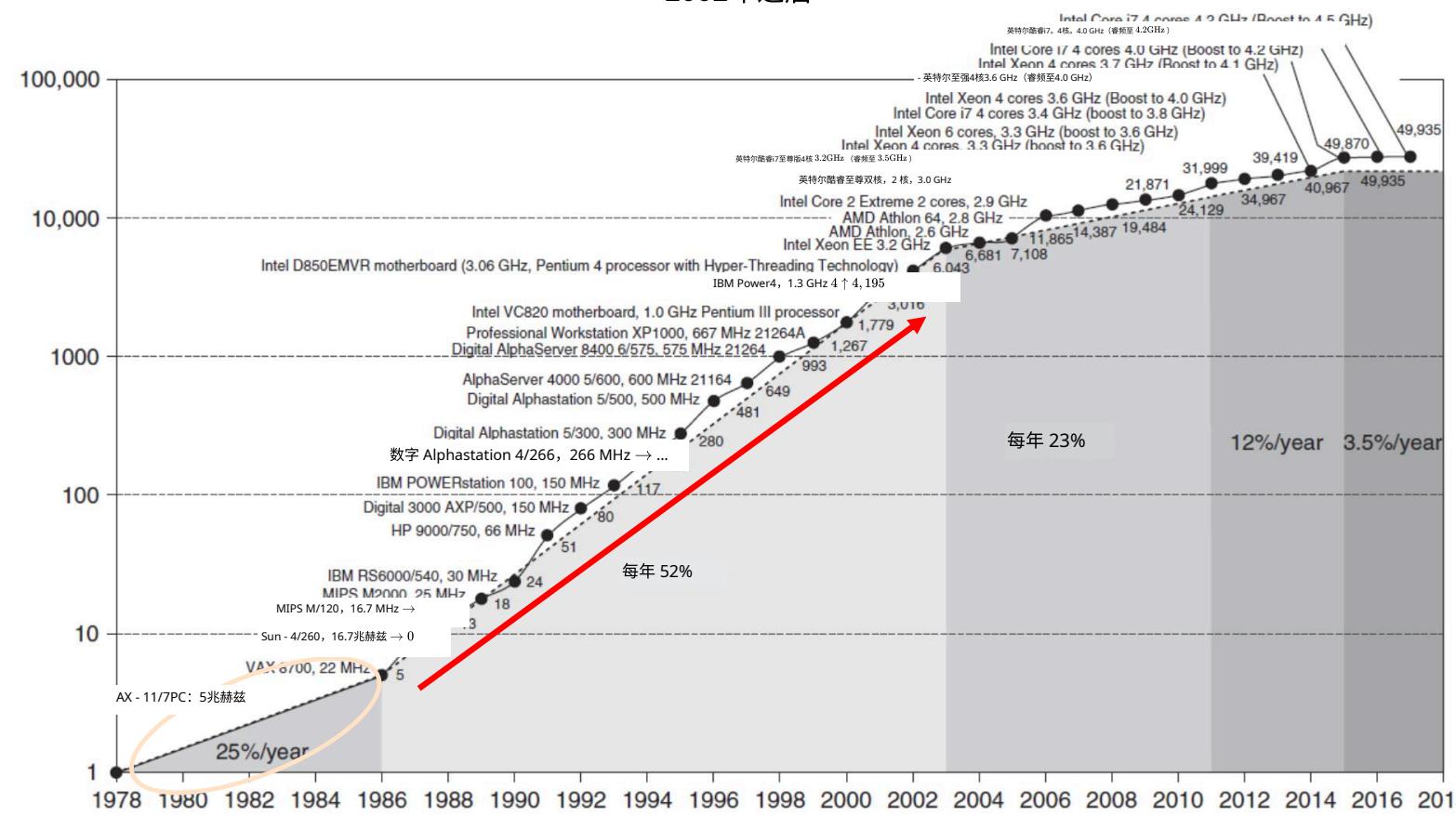


最新的计算机市场



惊人的CPU性能提升

VAX : 1978年至1986年每年提升25%; RISC+x86: 52%/year 1986 to 2002; RISC+x86: 22%/year 2002年之后



在 20th 世纪期间急剧增长率的影响

> 技术进步

互补金属氧化物半导体超大规模集成电路在成本和性能方面优于旧技术（晶体管 - 晶体管逻辑电路、发射极耦合逻辑电路）

> 计算机体系结构的进步

精简指令集计算机 (RISC)、超标量、乱序执行、推测执行、超长指令字 (VLIW)、独立磁盘冗余阵列 (RAID)

> 增强了计算机用户的能力：

如今微处理器的性能 > 20 年前的超级计算机

> 新型计算机

个人电脑、工作站 → 手机、笔记本电脑可访问包含数百万台服务器的仓库

> 基于微处理器的计算机在计算机设计领域占据主导地位

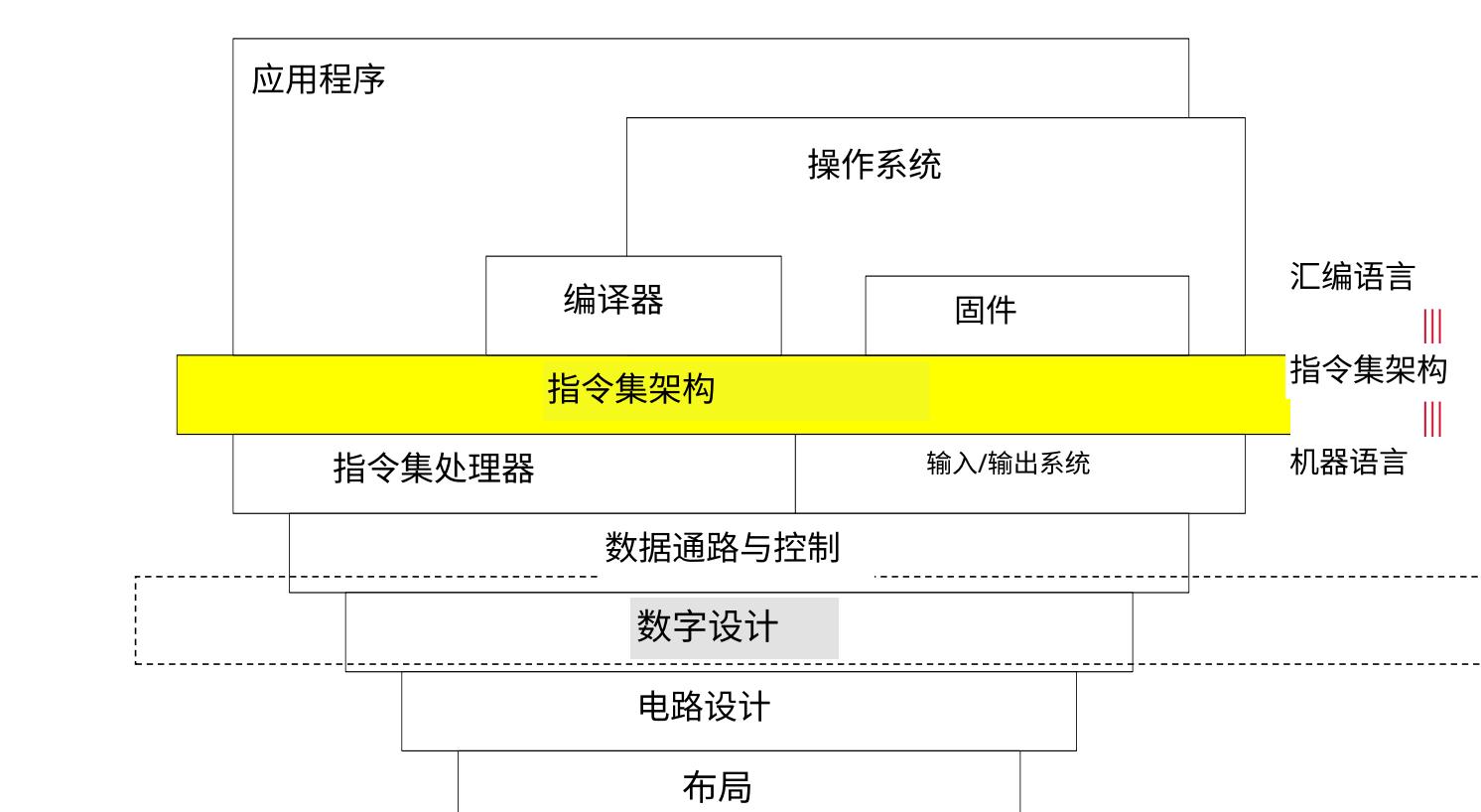
小型计算机 → 服务器、大型机、超级计算机 → 微处理器集合

> 更改软件

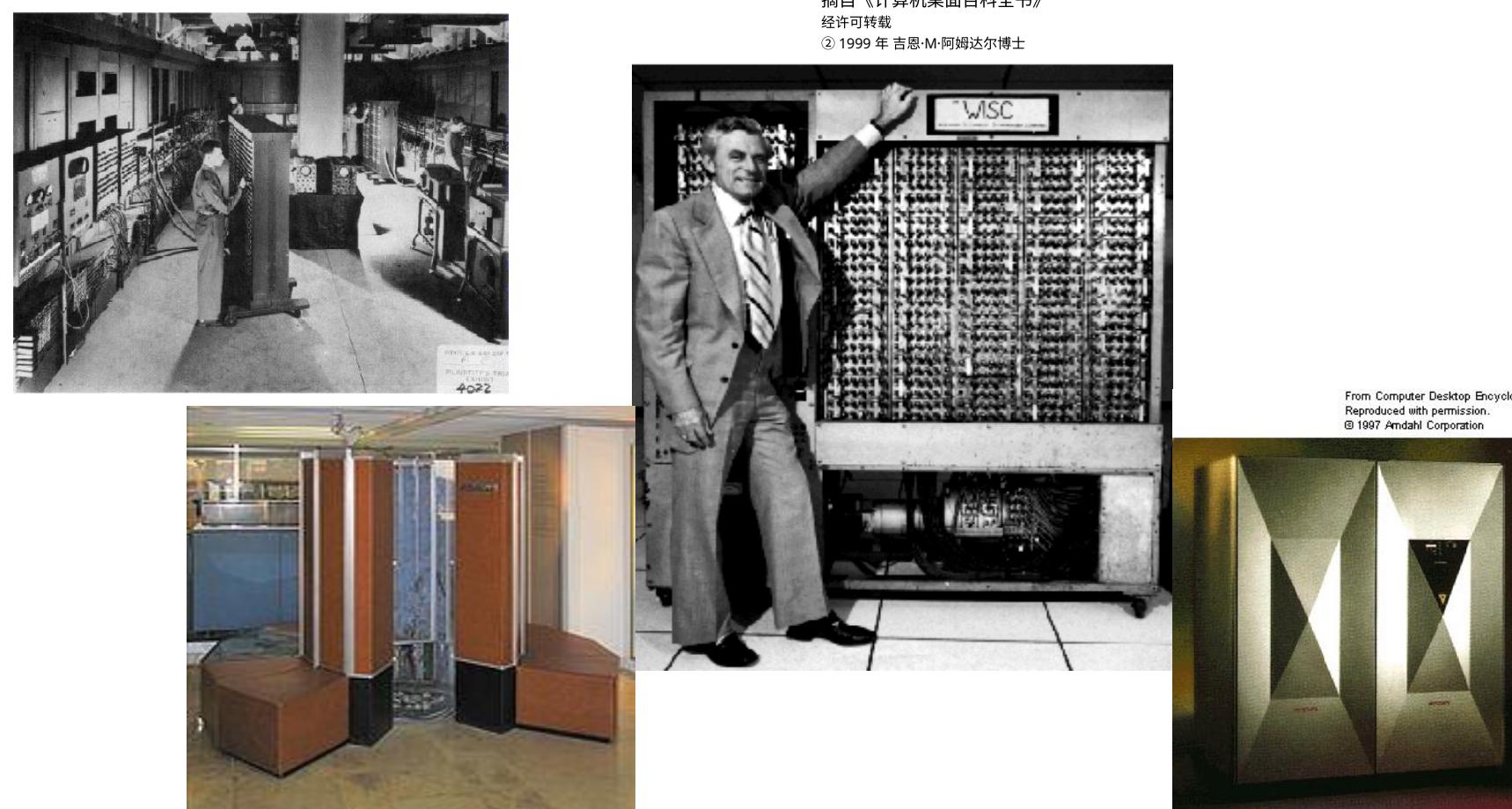
为提高生产力而进行性能权衡 C, C++ → Java, Scala; JavaScript, Python

改变应用程序的性质：科学计算 → 语音、图像、视频

计算机体系结构 = 指令集架构 (ISA) ?



外观差异很大



该图告诉了我们什么？

□ 性能改进：

25%：技术改进比计算机体系结构的进步更稳定。

» 特征尺寸、时钟速度

152%：精简指令集计算机 (RISC) 出现后，计算机设计既强调体系结构创新，也注重技术改进的有效利用。改进。

• 计算机体系结构在性能改进中起着重要作用

> 流水线、动态调度、乱序执行、分支预测、推测执行、超标量、超长指令字、预测指令

什么是计算机体系结构

□ 概念演变

程序员所看到的 [计算] 系统的属性，即

概念结构和功能行为，区别于数据流和控制的组织、逻辑设计以及物理实现。

阿姆达尔、布拉乌和布鲁克斯，1964年

指令集架构的七个维度

□ 指令集架构的类别

□ 内存寻址

□ 寻址模式

□ 操作数的类型和大小

□ 操作

□ 控制流指令

□ 对指令集架构 (ISA) 进行编码

截然不同的指令集架构

OPDP-11

-LC3

IBM 360

-80x86

VAX

-PowerPC

ICRAY

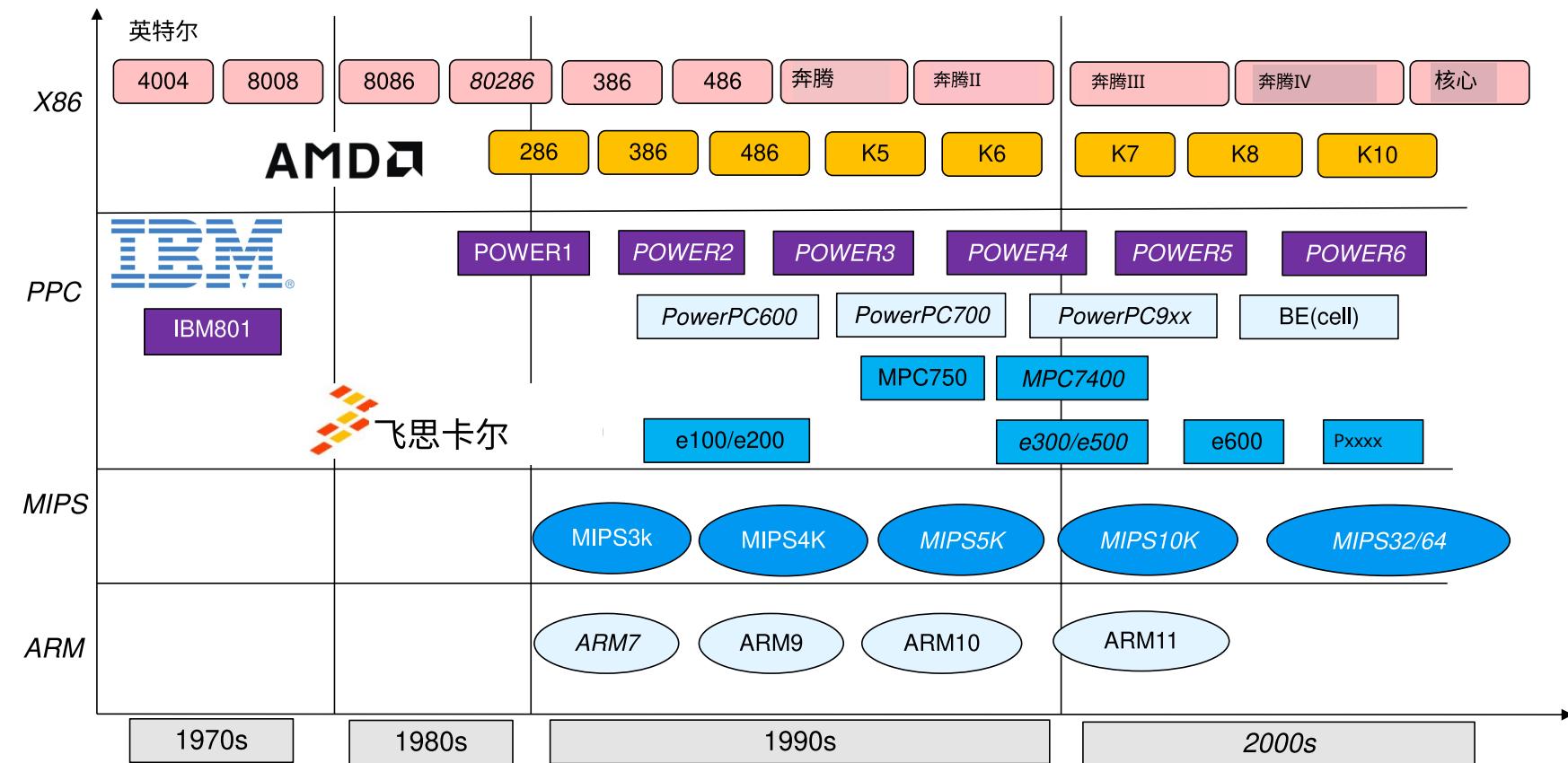
-MIPS

...

-ARM

-RISC V

主流CPU发展路径



新的概念？



□ 计算机体系结构是一门选择硬件组件并将其互连以创建满足功能、性能、成本和功耗目标的计算机的科学与艺术。

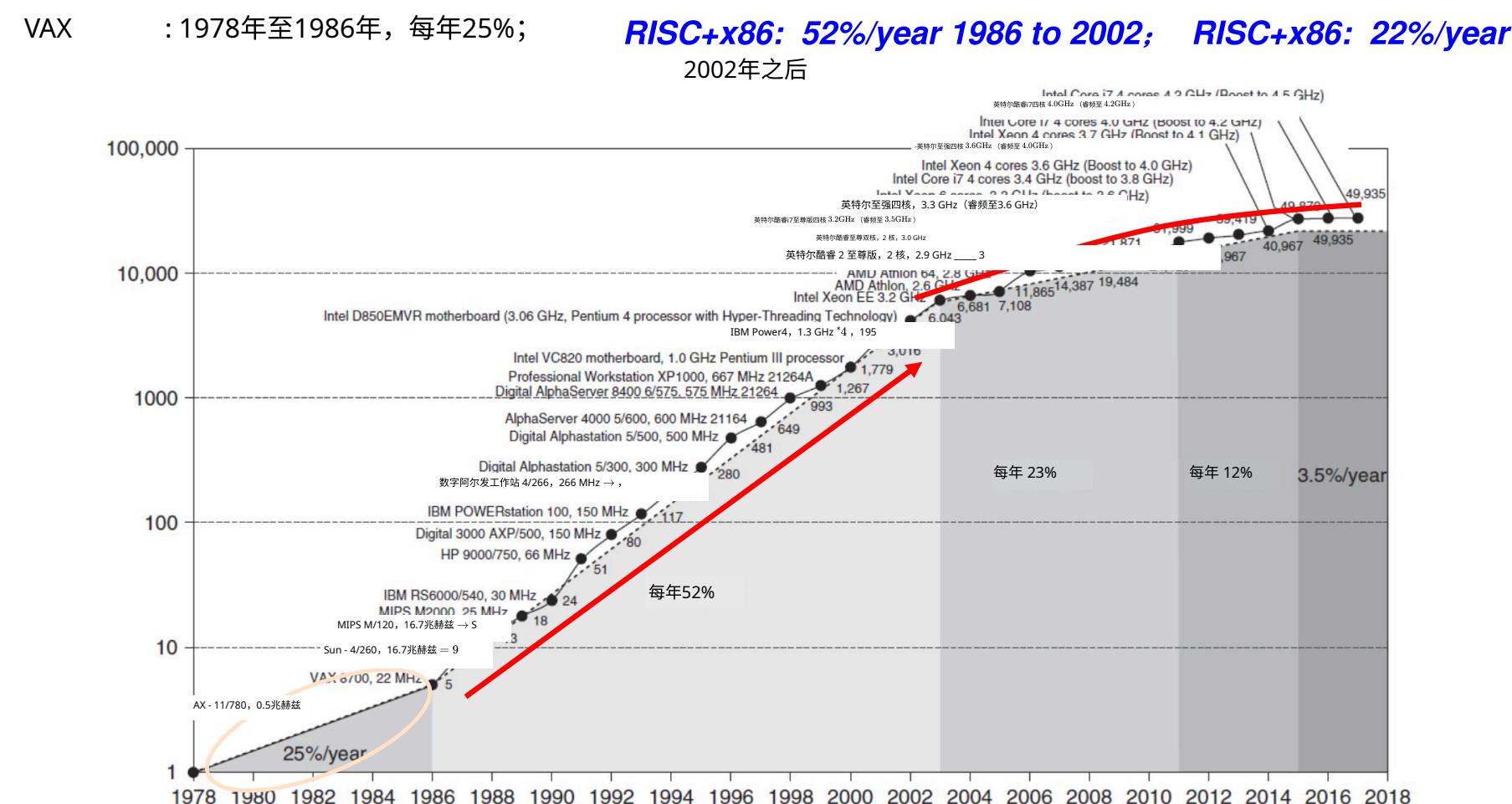
□ 它是计算机各部分需求和设计实现的蓝图与功能描述，主要关注中央处理器（CPU）的内部运行方式以及对内存地址的访问方式。

计算机体系结构

2 计算机体系结构至少包含三个主要子类别：^[1]

- 指令集架构，
- 微架构，也称为计算机组织，是对系统更底层、更具体和详细的描述，涉及系统的组成部分如何相互连接，以及它们如何进行互操作以实现指令集架构（ISA）。¹
- 系统设计，涵盖计算系统内的所有其他硬件组件，例如：
 - 逻辑实现
 - 电路实现
 - 物理实现

惊人的CPU性能提升



“三堵墙”的挑战

- 指令级并行性墙
 - > 寻找更多指令级并行（ILP）硬件的收益递减（必须利用显式线程和数据并行性）
- 内存墙
 - > CPU与CPU芯片外内存之间的速度差距日益增大。内存延迟将成为计算机性能的主要瓶颈。
- 功耗墙
 - > 运行频率每增加一倍，功耗就增加一倍的趋势

计算机设计任务

- 确定新机器的重要属性，在满足成本、功耗、可用性等约束条件的同时，最大限度地提高性能。
 - > 指令集架构设计
 - > 职能型组织
 - 计算机设计的高级方面，即内存系统、总线架构和CPU内部设计。
 - > 逻辑设计（硬件）
 - > 实现（硬件）

为什么2004年后有所下降？三个变化

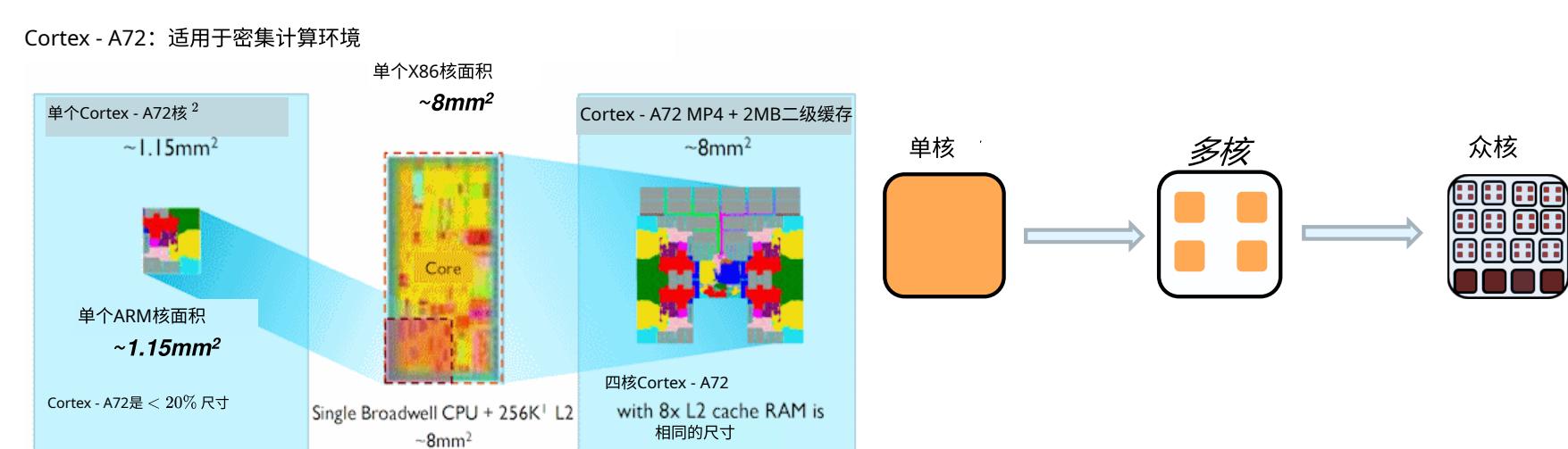
- 技术
 - > 登纳德缩放定律的终结：功耗成为关键限制因素——即使增加晶体管数量，给定硅片面积的功率密度仍保持恒定
 - » 摩尔定律放缓：晶体管成本（即使未使用）
- 架构方面
 - > 利用指令级并行性和单处理器时代的局限性与低效性。
 - > 阿姆达尔定律及其影响终结了“轻松”的多核时代

应用重点转移

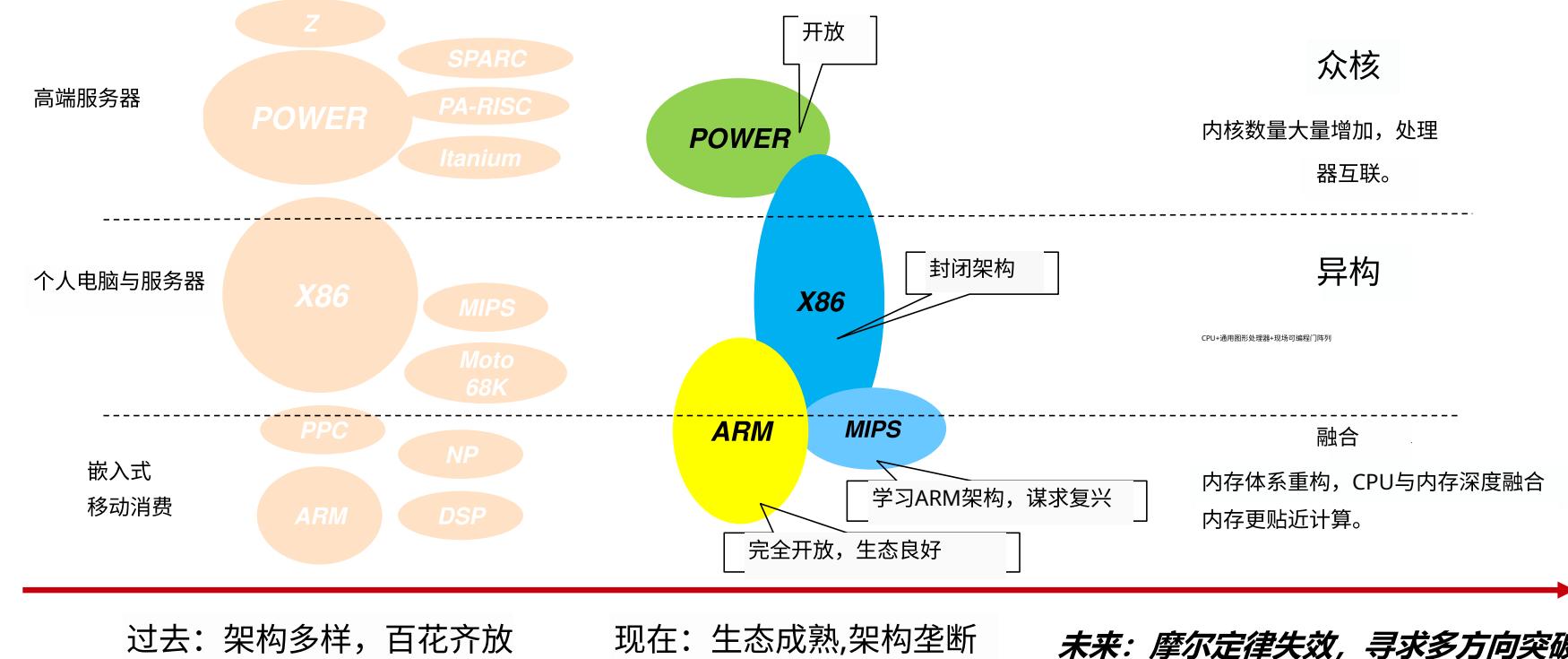
- > 从桌面设备到个人设备、移动设备和超大规模云计算：新的限制。

处理器芯片的发展趋势

- 工艺、主频遇到瓶颈后，开始通过增加核数的方式来提升性能；
- 芯片的物理尺寸有限制，不能无限制地增加；
- ARM的众核横向扩展空间优势明显。



现代CPU发展的一些趋势特点



ARM 服务器级别处理器一览

HISILICON	Hi1612	32 核, 2.1GHz 16nm	Hi1616	32 核, 2.4GHz 16nm	Hi1620	48 核, 3.0GHz 7nm
CAVIUM	雷霆 X	48核, 2.5GHz 28nm	雷神 X2	32 - 54摄氏度, 3.0吉赫兹 14nm		
	2400	48核, 2.2 - 2.6吉赫兹 14nm				
	AMPERE	X - 基因3	32核, 3.3吉赫兹 16nm			
飞腾	FT1500	16核, 1.6GHz 28nm	FT2000+	64核, 2.3吉赫兹 16nm		横轴代表性能

如今, 一个片上系统 (SoC) 上集成多种指令集架构 (ISA)

- 应用处理器 (通常为 ARM 架构)
- 图形处理器
- 图像处理器
- 无线电数字信号处理器 (DSP)
- 音频数字信号处理器
- 安全处理器
- 电源管理处理器
- 某些片上系统 (SoC) 上有超过十几种指令集架构 (ISA) ——每种都有独特的软件栈

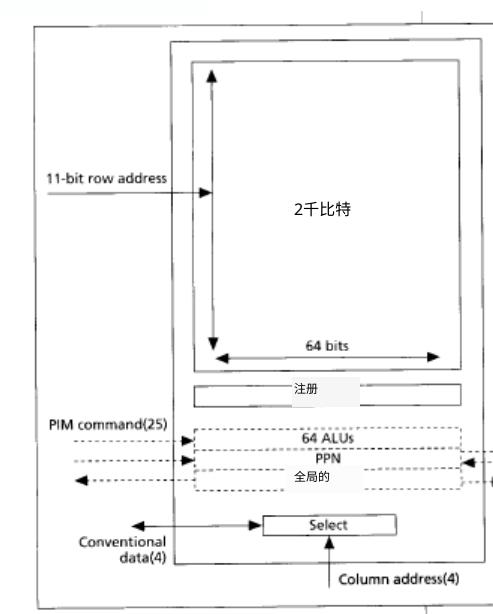
为什么?

应用处理器的指令集架构太大, 对加速器而言缺乏灵活性
> 从不同地方购买的IP, 每个都有专属的指令集架构 (ISA)
> 工程师构建自主研发的指令集架构 (ISA) 内核

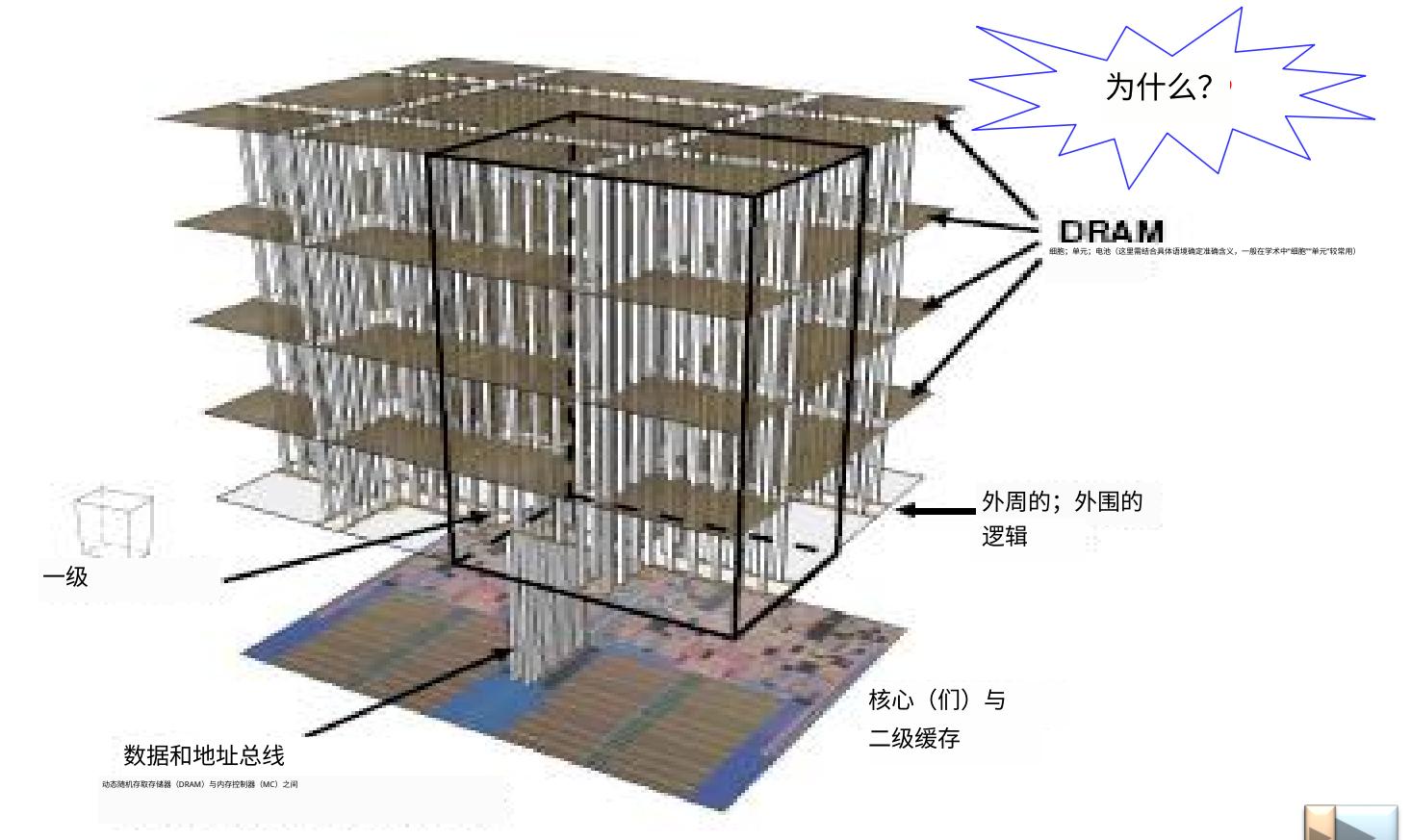
计算型随机存取存储器 (RAM) /存内计算 (PIM)

- 存内处理器 (PIM)
> 存内处理 (PIM, 有时也称为存内处理器) 是将处理器与随机存取存储器 (RAM) 集成在单个芯片上。其结果有时即所谓的PIM芯片。

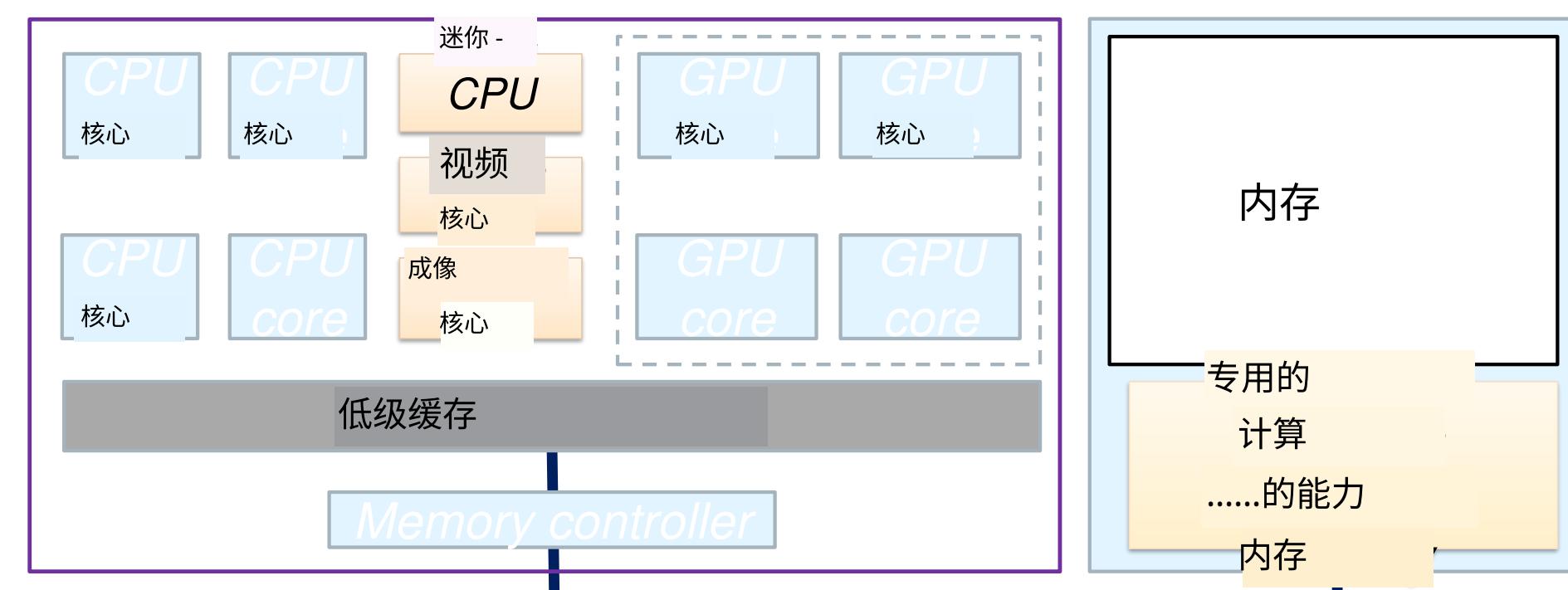
3 1995年。4 《IEEE计算机》:
内存内处理: Terasys大规模并行PIM阵列。



计算机体系统结构发展趋势



将内存用作加速器



按位运算: DRAM内复制、清零、与、或、非
2025/3/31

在哪里探索并行性?

■ 隐式, 编译器和硬件

→ 显式, 程序员

所以, 你们程序员必须了解硬件中的并行性。

并且在设计算法和编程时探索并行性!

编程!

application Algorithm Language Compiler Hardware

硬件 → 硬件和编译器 → 编译器和程序员

指令级并行 (ILP)、循环级并行 → 线程级并行 (TLP)、数据级并行 (DLP)

硬件与软件协同设计

关于计算机体系结构的更多内容

中国最快的计算机

□ 2011年天河一号A - 大规模并行处理系统

□ 2008年曙光5000A

> 30720个节点 * AMD皓龙1.9GHz四核处理器

> 内存: 122.88TB, 无限带宽网络, 180.6万亿次浮点运算每秒

> 操作系统: Windows HPC 2008

> 2008年11月全球超级计算机500强排名第10

□ 2004年曙光4000A

> 11万亿次浮点运算每秒

> 2004年6月全球超级计算机500强排名第10

□ 2003年深腾6800

> 5.324 万亿次浮点运算每秒

□ 2002年深腾1800

> 2.04 万亿次浮点运算每秒

□ 2000年银河四号

> 1024个CPU

> 1万亿次浮点运算每秒

最快的计算机有哪些?

>><http://top500.org/> 2011年6月

1. “京”计算机, SPARC64 VIIIfx 2.0GHz, Tofu互连 2. 天河一号A - 大规模并行处理系统, X5670 2.93Ghz 6核, 英伟达GPU, 飞腾1000 8核 3. 美洲虎 - 克雷XT5 - HE皓龙6核2.6 GHz 4. 星云 - 曙光TC3600刀片服务器, 英特尔X5650, 英伟达特斯拉C2050 GPU 5. 筑波2.0 - 惠普ProLiant SL390s G7至强6核X5670, 英伟达GPU, Linux/Windows 6. 天空 - 克雷XE6 8核2.4 GHz 7. 昂星团 - 思杰Altix ICE 8200EX/8400EX, 至强超线程四核3.0/至强5570/5670 2.93Ghz, 无限带宽

8. 霍普 - 克雷XE6 12核2.1 GHz

9. Tera - 100 - 布尔bullx超级节点S6010/S6030

10台路runner - BladeCenter QS22/LS21集群, PowerXCell 8i

3.2 GHz / Opteron DC 1.8 GHz, Voltaire Infiniband

来自<http://top500.org/> 2015年6月的世界
最快超级计算机

RANK	SITE	SYSTEM	CORES	[TFLOP/S]	[TFLOP/S]	[kW]
1	中国广州国家超级计算中心	天河二号(银河二号) - TH-IVB - FEP集群, 英特尔至强E5-2692 12核2.200GHz, TH Express - 2, 英特尔至强融核3151P	3,120,000	33,862.7	54,902.4	17,808
	NUDT					
2	美国能源部/科学办公室/橡树岭国家实验室	泰坦 - 克雷XK7, 风龙6274 16核2.200GHz, 克雷双子座互连, 英伟达K20X 克雷公司	560,640	17,590.0	27,112.5	8,209
3	美国能源部/国家核安全管理/劳伦斯利弗莫尔国家实验室	红杉 - 蓝色基因/Q, Power BQC 16C 1.60 GHz, 定制	1,572,864	17,173.2	20,132.7	7,890
4	日本理化研究所高级计算科学研究所 日本	K计算机, SPARC64 VIIIfx 2.0GHz, 豆腐互连 富士通	705,024	10,510.0	11,280.4	12,660
5	美国能源部/科学办公室/阿贡国家实验室 美国	米拉 - 蓝色基因/Q, Power BQC 16C 1.60GHz, 定制	786,432	8,586.6	10,066.3	3,945
6	瑞士国家超级计算中心 (CSCS)	皮兹·戴恩特 - 克雷XC30, 至强E5-2670 8核2.600GHz, 白羊座互连, 英伟达K20X 克雷公司	115,984	6,271.0	7,788.9	2,325
7	沙特阿拉伯阿卜杜拉国王科技大学	沙欣二号 - 克雷XC40, 至强E5-2698v3 16核2.3GHz, 白羊座互连 克雷公司	196,608	5,537.0	7,235.2	2,834
8	美国得克萨斯大学高级计算中心	Stampede - PowerEdge C8220, 至强E5-2680 8核2.700GHz, 无限带宽FDR, 英特尔至强融核SE10PDett	462,462	5,168.1	8,520.1	4,510
9	于利希研究中心 德国 (FZJ)	JUQUEEN - 蓝色基因/Q, Power BQC 16核1.60GHz, 定制互连	458,752	5,008.9	5,872.0	2,301
10	美国能源部/国家核安全管理/劳伦斯利弗莫尔国家实验室 美国	火神 - 蓝色基因/Q, Power BQC 16核1.60GHz, 定制互连	393,216	4,293.3	5,033.2	1,972

世界上最快的超级计算机

数据来源: <http://top500.org/> 2013年6月

排名	站点	系统	核心	最大实测性能 (万亿次浮点运算每秒)	峰值性能 (万亿次浮点运算每秒)	功率 (kW)
1	国防科技大学	天河二号(银河二号) - TH-IVB - FEP集群, 英特尔至强E5-2692 12核2.200GHz, TH Express - 2, 英特尔至强融核3151P 国防科技大学	3,120,000	33,862.7	54,902.4	17,808
2	美国能源部科学办公室/橡树岭国家实验室	泰坦 - 克雷XK7, 风龙6274 16核2.200GHz, 克雷双子座互连, 英伟达K20X 克雷公司	560,640	17,590.0	27,112.5	8,209
3	美国能源部/国家核安全管理/劳伦斯利弗莫尔国家实验室	红杉 - 蓝色基因/Q, Power BQC 16C 1.60 GHz, 定制	1,572,864	17,173.2	20,132.7	7,890
4	日本理化研究所高级计算科学研究所	K计算机, SPARC64 VIIIfx 2.0GHz, 豆腐互连 富士通	705,024	10,510.0	11,280.4	12,660
5	美国能源部/科学办公室/阿贡国家实验室 美国	米拉 - 蓝色基因/Q, Power BQC 16C 1.60GHz, 定制	786,432	8,586.6	10,066.3	3,945
6	美国得克萨斯大学高级计算中心	Stampede - PowerEdge C8220, 至强E5-2680 8核2.700GHz, 无限带宽FDR, 英特尔至强融核SE10PDett	462,462	5,168.1	8,520.1	4,510
7	于利希研究中心 德国 (FZJ)	JUQUEEN - 蓝色基因/Q, Power BQC 16核1.60GHz, 定制互连	458,752	5,008.9	5,872.0	2,301
8	美国能源部/国家核安全管理/劳伦斯利弗莫尔国家实验室	火神 - 蓝色基因/Q, Power BQC 16核1.60GHz, 定制互连	393,216	4,293.3	5,033.2	1,972
9	莱布尼茨计算中心 德国	SuperMUC - iDataPlex DX360M4, 至强E5-2680 8核2.700GHz, 无限带宽FDR	147,456	2,897.0	3,185.1	3,423
10	国家超级计算天津中心 中国	天河一号A - 国防科大银河MPP, 至强X5670 6核2.93GHz, 英伟达2050	186,368	2,566.0	4,701.0	4,040

世界上最快的超级计算机 来自<http://top500.org/>
2016年6月

RANK	SITE	SYSTEM	核心	最大实测性能 (万亿次浮点运算每秒)	峰值性能 (万亿次浮点运算每秒)	功率 (kW)
1	中国无锡国家超级计算中心	神威·太湖之光 - 神威大规模并行处理系统, 神威SW26010 260核1.45GHz, 神威NRPc	10,649,600	93,014.6	125,435.9	15,371
2	中国广州国家超级计算中心	天河二号(银河二号) - TH-IVB-FEP集群, 英特尔至强ES-2692 12核2.200GHz, 天河二号高速互联通信系统, 英特尔至强融核3151P NUDT	3,120,000	33,862.7	54,902.4	17,808
3	美国能源部科学办公室/橡树岭国家实验室 美国	泰坦 - 克雷XK7, 风龙6274 16核2.200GHz, 克雷双子座互连网	560,640	17,590.0	27,112.5	8,209
4	美国能源部/国家核安全管理/劳伦斯利弗莫尔国家实验室 美国	红杉 - 蓝色基因/Q, Power BQC 16核1.60GHz, 定制	1,572,864	17,173.2	20,132.7	7,890
5	日本理化研究所高级计算科学研究所 日本	K计算机, SPARC64 VIIIfx 2.0GHz, Tofu互连, 富士通	705,024	10,510.0	11,280.4	12,660
6	美国能源部科学办公室/阿贡国家实验室	米拉 - 蓝色基因/Q, Power BQC 16C 1.60GHz, 定制	786,432	8,586.6	10,066.3	3,945
7	美国能源部/国家核安全管理/劳伦斯利弗莫尔国家实验室 美国	三位一休 - 克雷XC40, 至强E5-2698v3 16C 2.3GHz, 白羊座互连, 克雷公司	301,056	8,100.9	11,078.9	
8	瑞士国家超级计算中心 (CSCS)	皮兹·戴恩特 (Piz Daint) ——克雷XC30, 至强E5-2670单核心2.600GHz, 白羊座互连, 英伟达 克雷公司	115,984	6,271.0	7,788.9	2,325
9	HLRS - 斯图加特高性能计算中心 德国	海泽尔·亨 (Hazel Hen) ——克雷XC40, 至强E5-2680v3 12核2.5GHz, 白羊座互连 克雷公司	185,088	5,640.2	7,403.5	
10	沙特阿拉伯阿卜杜拉国王科技大学	沙欣二号 (Shaheen II) ——克雷XC40, 至强E5-2698v3 16核2.3GHz, 白羊座互连 克雷公司	196,608	5,537.0	7,235.2	2,834

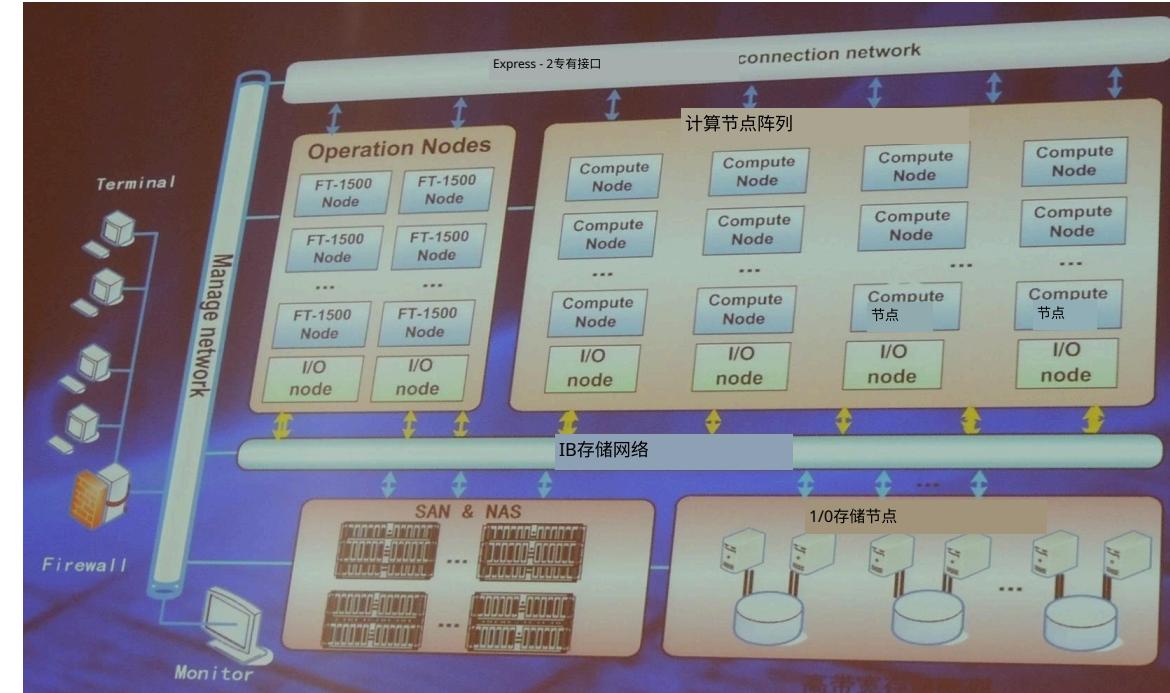
世界上最快的超级计算机 来自
<http://top500.org/> 2017年6月

TH - 2概述

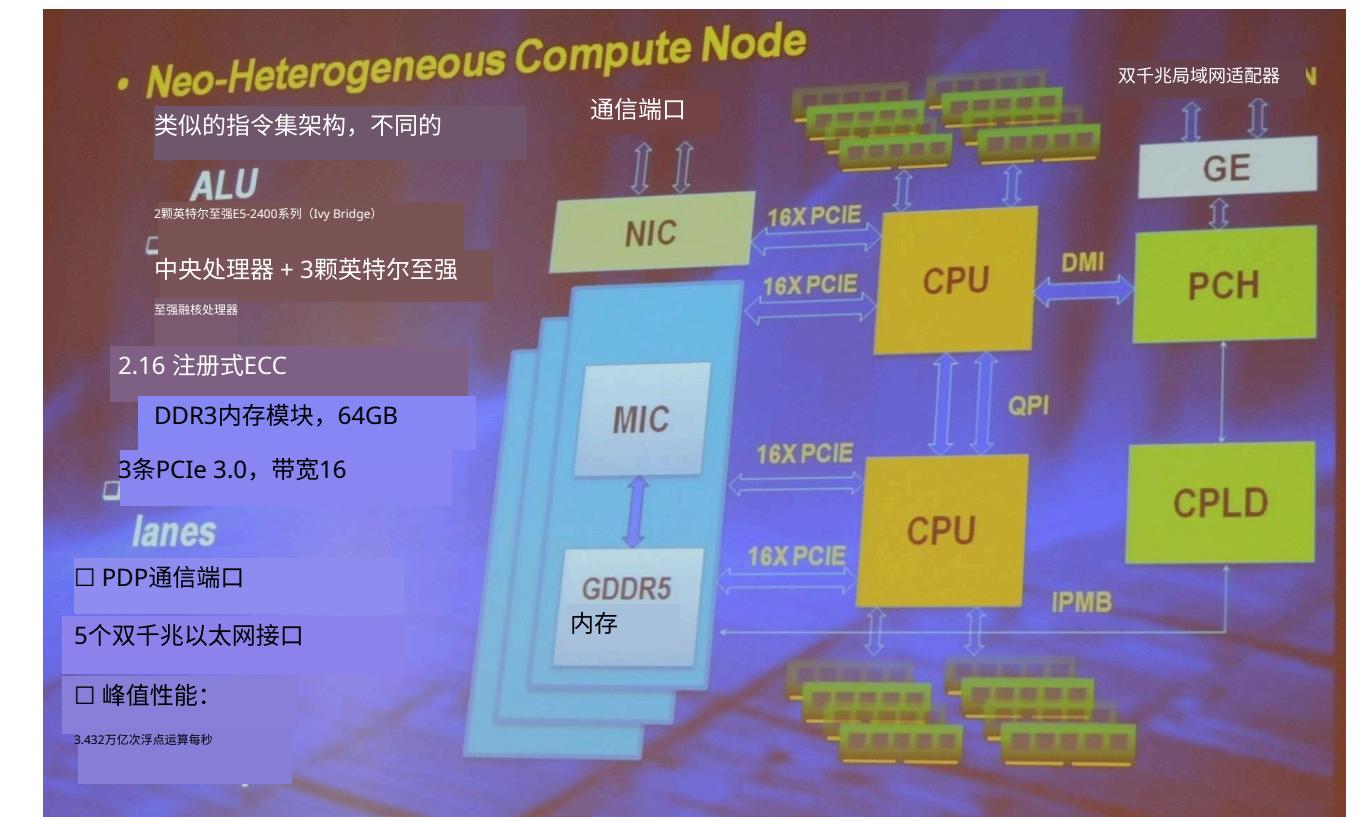
项目 (此条目可能拼写有误, 推测为items)	配置
处理器	32000颗英特尔至强CPU + 48000颗至强Phi + 4096颗FT CPU峰值性能为54.9千万亿次浮点运算每秒, 持续性能为33.9千万亿次浮点运算每秒
互连	专有高速互连网络TH Express - 2
内存	总计1.4PB
存储	全局共享并行存储系统, 12.4PB
机柜	125+13+24+8 = 170 计算/通信/存储/服务机柜
电源	17.8MW
冷却	封闭空气冷却系统

排名	系统	核心数	最大实测性能 (万亿次浮点运算每秒)	峰值性能 (万亿次浮点运算每秒)	功率 (kW)
1	神威·太湖之光 - 神威大规模并行处理系统, 神威SW26010 260核1.45GHz, 神威, 国家发展和改革委员会 中国无锡国家超级计算中心	10,649,600	93,014.6	125,435.9	15,371
2	天河二号(银河二号) - 天河IVB - FEP集群, 英特尔至强E5-2692 12核2.200GHz, 天河二号互联网, 英特尔至强融核3151P, 中国广州国家超级计算中心 NUDT	3,120,000	33,862.7	54,902.4	17,808
3	皮兹·丹特 - 克雷XC50, 至强E5-2690v3 12核2.6GHz, 白羊座互联网, 英伟达特斯拉P100, 克雷公司 瑞士国家超级计算中心 (CSCS) 瑞士	361,7			

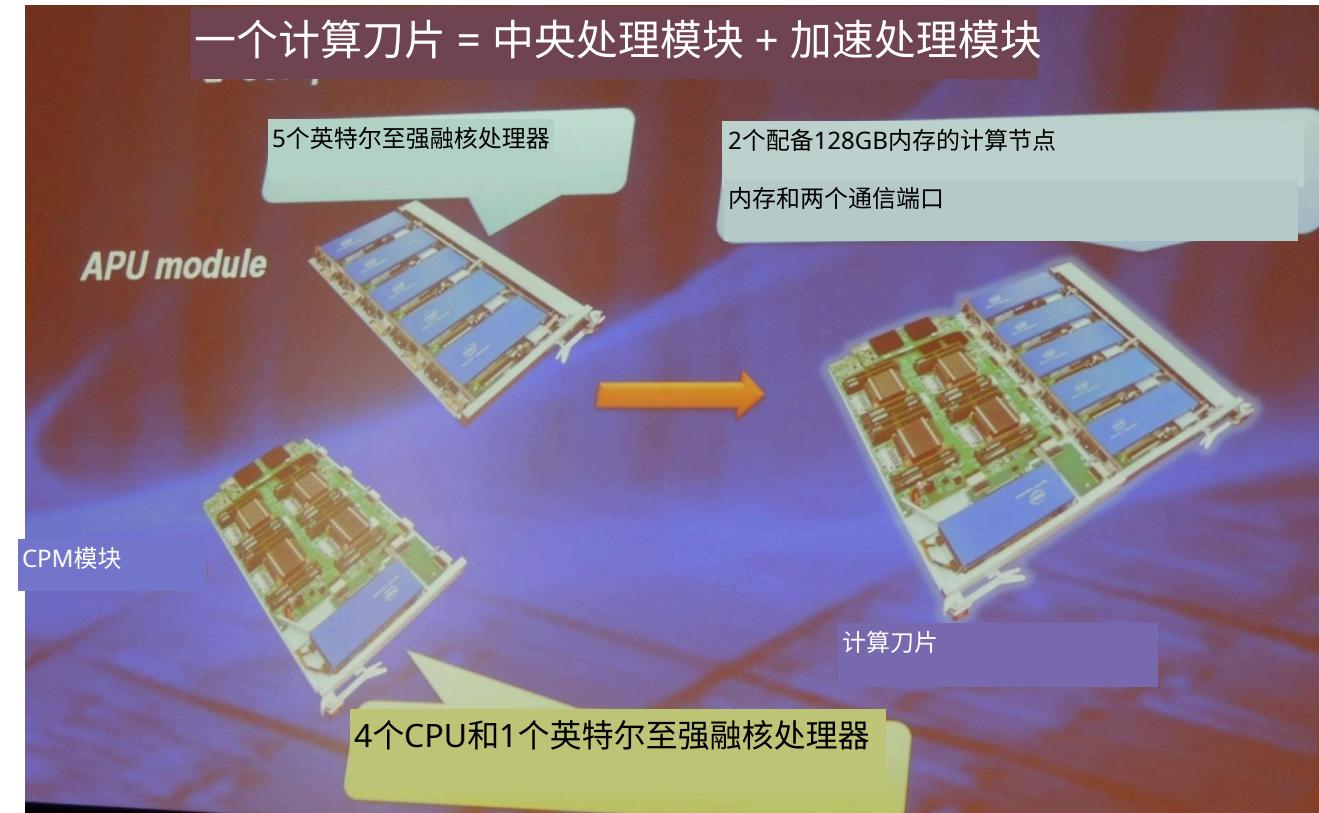
TH - 2硬件子系统



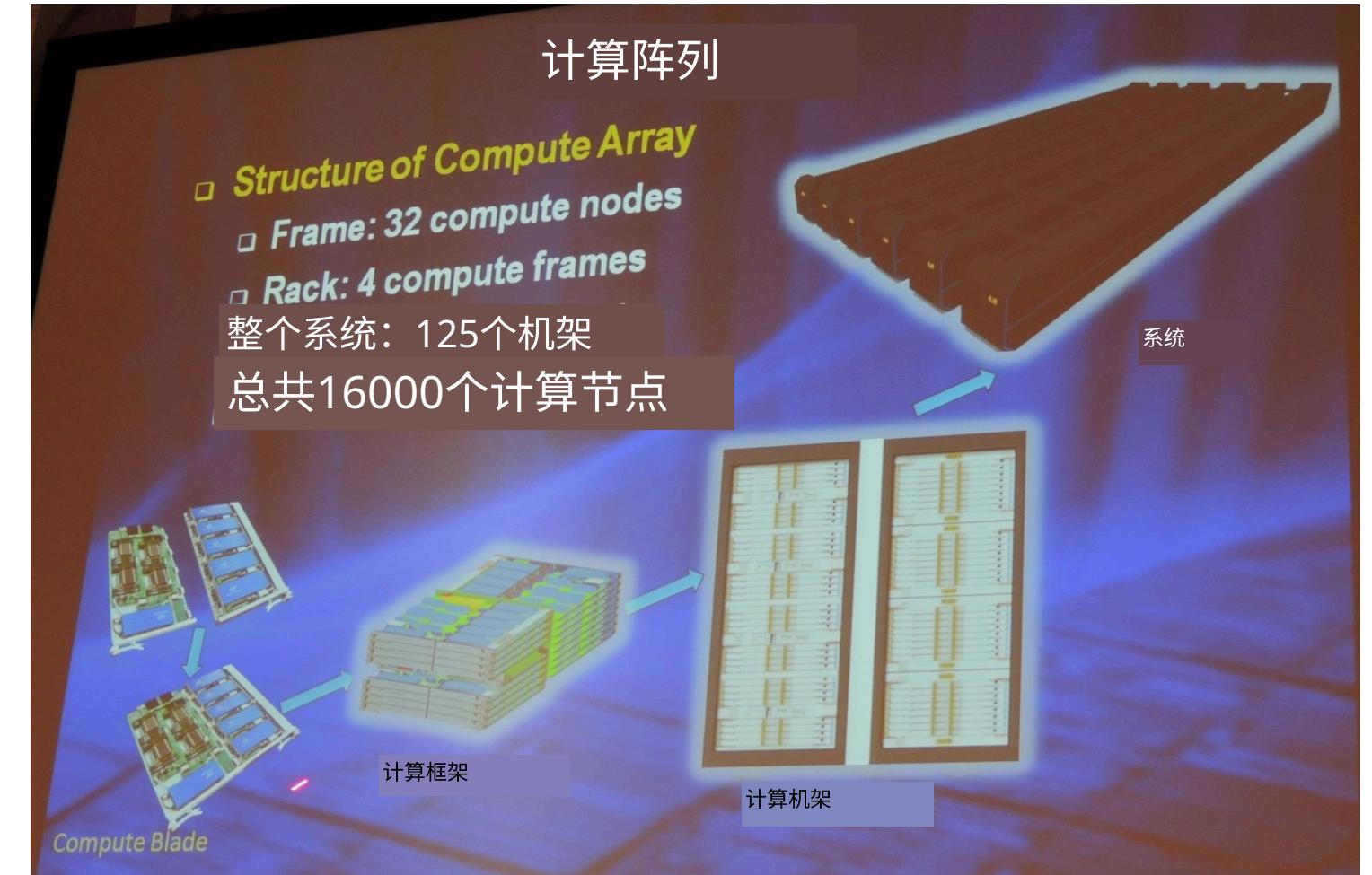
天河二号计算机节点



天河二号计算节点



天河二号计算阵列



关于太湖之光超级计算机

• 40960个中国自主设计的SW26010多核64位RISC处理器

每个SW26010处理器有X (256个处理核心 + 4个辅助核心) = 10,649,600CPU核心

• 处理核心具备 64 KB 的数据暂存内存 (指令内存为 16 KB)，并通过片上网络进行通信。

• 操作系统：基于Linux的神威睿思操作系统2.0.5，拥有其自定义实现的OpenACC2.0。



2020年6月最新排名前十

第一名 富岳 (富岳)	第四名 神威·太湖之光 (Sunway TaihuLight) 中国
日本 制造商：富士通 处理器核心：7299072个；峰值(Rmax)：415530 TFlop/s 简介： Fugaku超算原来被称为“Post K”，是继往的世界第一计算机产品。它的复兴时代，采用ARM架构的是ASIFX处理器。性能为第二名Summit的2.5倍。	中国 制造商：国家并行计算机工程技术研究中心 处理器核心：10649600个；峰值 (Rmax)：930157万亿次浮点运算每秒 简介： 我介绍的神威·太湖之光超级计算机连续四次位列全球超级计算机500强榜首，该系统全部使用中国自主知识产权的处理器芯片
第二名 Summit (美国)	
多国制造商：IBM 处理器核心：2414592个；峰值(Rmax)：148600 TFlop/s 简介： 作为超级计算机，Summit比神威·太湖之光快60%，比此前的Titan——美国冠军保持者快8倍，而在其内部，近28000块英伟达Volta GPU提供了95%的算力。	
第三名 塞拉 (美国)	第五名 天河二号 (中国)
处理器核心：1572480个；峰值(Rmax)：94640 TFlop/s 简介： 高性能计算机类国际服务规模例检实验室已经明确了接下来要做的事情，助力科学家在资源协调、材料发现、医疗保健等领域里的研究探索，其在癌症研究方面用于名为“癌症分布式学习环境 (CANDLE)”的项目	中国 承建商：国防科大 处理器核心：4981760个；峰值 (Rmax)：61445万亿次浮点运算每秒 简介： 天河二号历经6次蝉联冠军，采用麒麟操作系统；目前使用英特尔处理器 将来计划用国产处理器替换，不仅应用于电力勘探工程、载人航天军科科探项目，还在石油勘探、汽车飞机的设计制造、基因测序等民用方面大展身手 手
第六名 超级计算机5 (意大利)	
处理器核心：66910个；峰值 (Rmax)：304507万亿次浮点运算每秒 简介： 由DELL EMC公司为ENI能源公司打造的世界上功能最强大的工业级超计算机、它的潜在体系结构使分子模拟算法特别有效。	

世界上最快的超级计算机，数据来源：
<http://top500.org/> 2018年6月

排名系统	核心	最大实测浮点运算速度 [万亿次浮点运算每秒]	峰值浮点运算速度 [万亿次浮点运算每秒]	功耗 [kW]
1 顶点 - IBM Power System AC922、IBM POWER9 22C 3.07GHz、英伟达Volta GV100、双轨Mellanox EDR无限带宽网络、美国能源部/科学办公室/橡树岭国家实验室 美国	2,282,544	122,300.0	187,659.3	8,806
2 神威·太湖之光 - 神威大规模并行处理系统，神威SW26010 260核 1.45GHz，神威，国家发展和改革委员会，中国无锡国家超级计算中心	10,649,600	93,014.6	125,435.9	15,371
3 塞拉 - IBM Power System S922LC, IBM POWER9 22核 3.1GHz, 英伟达Volta GV100, 双轨Mellanox EDR无限带宽网络, IBM, 美国能源部/国家核安全管理局/劳伦斯利弗莫尔国家实验室, 美国	1,572,480	71,610.0	119,193.6	
4 天河二号A - 天河IVB - FEP集群，英特尔至强E5 - 2692v2 12核 2.2GHz，天河二号高速互连通信系统，Matrix - 2000，国防科技大学，中国广州国家超级计算中心	4,981,760	61,444.5	100,678.7	18,482
5 人工智能桥梁 黄金614 国家 Japan	6,616,649			
6 代恩峰 NVIDIA A1 瑞士 僧侣团U	326.3	2.272		
7 Titan - C 英伟达K20x 克雷公司 美国能源部科学办公室/橡树岭国家实验室	112.5	8,209		
“顶点”的计算能力比神威·太湖之光要快60%，比前美国超级计算机“泰坦 (泰坦)”要快8倍。				
8 红杉 - 蓝色基因/Q, Power BQC 16C 1.60 GHz, 定制版, IBM 美国能源部/国家核安全管理局/劳伦斯利弗莫尔国家实验室	1,572,864	17,173.2	20,132.7	7,890
9 三位一体 - 克雷XC40，英特尔至强融核7250 68C 1.4GHz，白羊座互连，克雷公司 美国 能源部/国家核安全管理局/洛斯阿拉莫斯国家实验室/桑迪亚国家实验室	979,968	14,137.3	43,902.6	3,844
10 科里 - 克雷XC40，英特尔至强融核7250 68核 1.4GHz，白羊座互连	622,336	14,014.7	27,880.7	3,939

世界上最快的超级计算机
数据来源：<http://top500.org/> 2022年6月

排名	系统	内核	最大实测浮点运算速度 [万亿次浮点运算每秒]	峰值浮点运算速度 [万亿次浮点运算每秒]	功耗 [kW]
1.0 美国 - 前沿* - 慧与-克雷-EX235a, - AMD优化第三代：霄龙64核2GHz, - AMD Instinct MI250X, “弹弓-11”, “慧与”美国	8,730,112	1,102.00	1,685.65	21,100	
2.0 日本 - 超级计算机——富岳——超级计算机富岳, - A64FX - 48核 - 2.26赫兹, - 豆腐互连D, *	7,630,848	442.01	537,219,899.4		
3.0 芬兰 - LUMI ° - 慧与 - 克雷 EX235a, - AMD 优化 - 第三代：霄龙 64 核 - 2GHz, - AMD 本能 MI250X, 弹弓 11, “HPE”欧洲高性能计算/芬兰气象与地球物理研究所/	1,110,144	151.90	214,352,942		
4.0 美国 - “顶点” - IBM 动力系统 AC922, - IBM POWER9 22 核 - 3.07GHz, - NVIDIA 伏特 GV100, - 双轨迈络思 EDR 无限带宽, * IBM 美国能源部/科学办公室/橡树岭国家实验室	2,414,592	148.60	200,7910,096		
5.0 美国 - “塞拉”* - IBM 动力系统 AC922, - IBM POWER9 22 核 - 3.1GHz, - NVIDIA 伏特 GV100, - 双轨迈络思：EDR 无限带宽, * IBM / - NVIDIA / - 迈络思 - 美国能源部/国家核安全管理局/劳伦斯利弗莫尔国家实验室	1,572,480	94.64	125,717,438		
6.0 中国 - 神威·太湖之光* - 神威大规模并行处理系统， - 神威SW26010 - 260C - 1.45GHz, - 神威, 国家发展和改革委员会/4 无锡国家超级计算中心	10,649,600	93.01	125,4415,371		
7.0 美国 - “伯尔马特”- 慧与公司：克雷EX235n, - AMD霄龙：7763 - 64C - 2.45吉赫兹, - 英伟达A100 - SXM4 - 40GB, - 弹弓10, * 慧与公司 - 美国能源部/科学办公室/劳伦斯伯克利国家实验室/国家能源研究科学计算中心	761,856	70.87	93,752,589		
8.0 美国 - “Selenne” - NVIDIA - DGX - A100, “AMD - EPYC - 7742 - 64C - 2.25GHz, - NVIDIA - A100, - Mellanox - HDR - 最低配置：“Nvidia” - 英伟达公司:	555,520	63.46	79,222,646#		
9.0 China - 天河二号A - TH - IVB - FEP集群， - 英特尔 - 至强 - E5 - 2692v2 - 12C - 2.2GHz, - TH - Express - 2, - Matrix - SDOS, “国防科大”广州国家超级计算中心	4,981,760	61.44	100,6818,482		
10.0 France - 弹弓 - 11, “HPE” - 国家高性能计算设备 - 国家计算机中心：高等教育 (GENCI - CINES) ->	46.10	61,61921			

排名	系统	核心数	最大浮点运算速率 (千万亿次浮点运算/秒)	R峰 (千万亿次浮点运算每秒)	功率 (kW)
1	前沿 - 慧与 (HPE) Cray EX235a, AMD优化的第三代EPYC 64核2GHz处理器, AMD Instinct MI250X, 弹弓-11网络, 美国能源部科学办公室/橡树岭国家实验室, 美国	8,699,904	1,206.00	1,714.81	22,786
2	极光 - 慧与 (HPE) Cray EX - 英特尔百亿亿次级计算刀片, 至强CPU Max 9470 52核2.4GHz处理器, 英特尔数据中心GPU Max, 弹弓-11网络, 英特尔, 美国能源部科学办公室/阿贡国家实验室, 美国	9,264,128	1,012.00	1,980.01	38,698
3	鹰 - 微软Nv5, 至强铂金8480C 48核2GHz处理器, 英伟达H100, 英伟达无限带宽NDR网络, 微软Azure, 微软Azure, 美国	2,073,600	561.20	846.84	
4	富岳超级计算机 - 富岳超级计算机, A64FX 48核 2.2GHz, Tofu互连D。富士通理化研究所计算科学中心 日本	7,630,848	442.01	537.21	29,899
5	LUMI - HPE Cray EX235a, AMD优化的第三代EPYC 64核 2GHz, AMD Instinct MI250X, Slingshot - 11, HPE 欧洲高性能计算/芬兰科学计算中心 芬兰	2,752,704	379.70	531.51	7,107

芯片技术

台湾的天下，排名世界第一的日月光，还跟着一堆实力不俗的小弟：矽品、力成、南茂、欣邦、京元电子。

中国大陆的三大封测巨头：长电科技、华天科技、通富微电

刻蚀机，中国的情况要好很多，16纳米已经量产，7-10纳米也在上

离子注入机70%的市

场份额是美国应用材料公司光刻机，荷兰阿斯麦公司(ASML) 唯一高端光刻机生产商(12,24,40 台/年)

高纯硅要求99.9999999%，几乎全靠进口，传统霸主依然是德国瓦克和美国赫姆洛克 (美日合资)

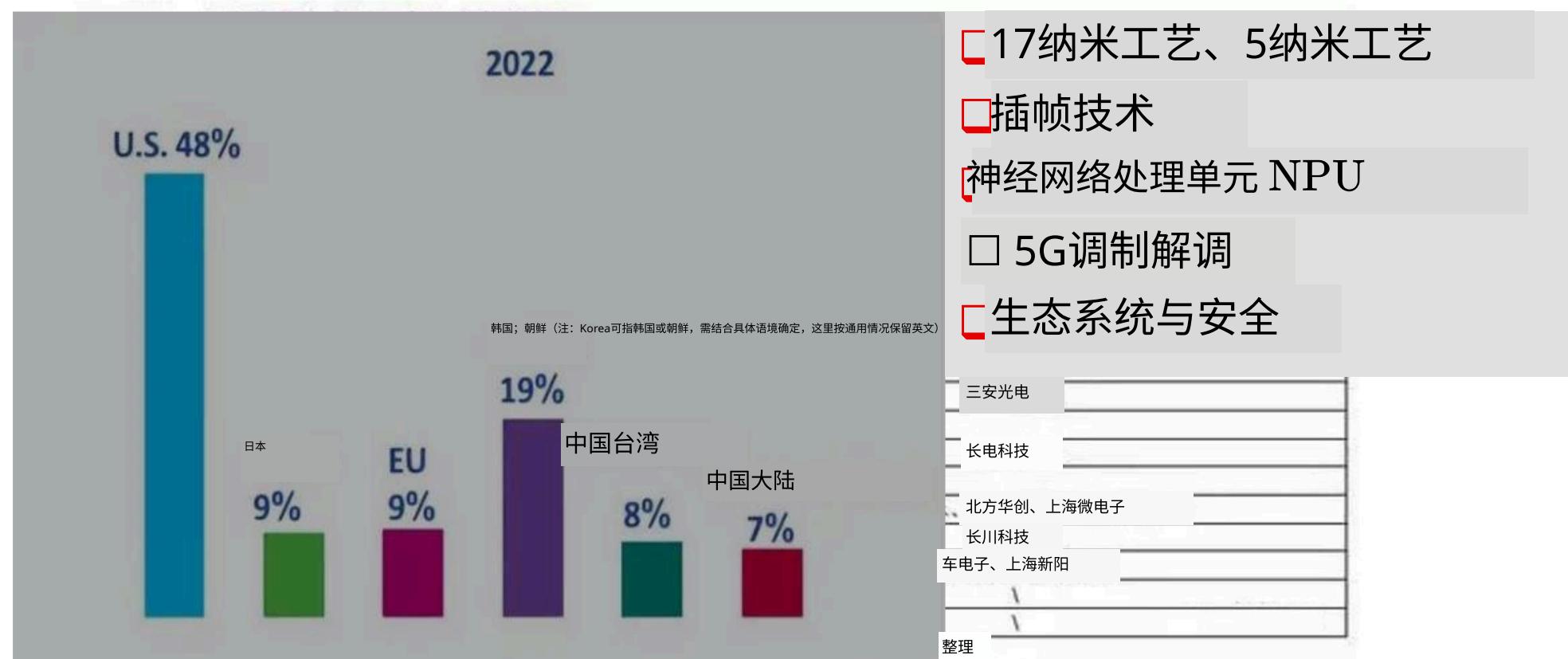
一块300毫米直径的晶圆16纳米工艺可以做出100块芯片，10纳米工艺可以做出210块芯片，价格可便宜了一半

美国英特尔、韩国三星、日本东芝、意法半导体；中国台湾地区的：旺宏电子；中国华润微电子等。但中国高端芯片几乎空白

设计与制造 晶圆代工厂：
台积电、中芯国际
高通、博通、AMD，
中国台湾的联发科，
中国大陆的华为海思

芯片技术

中国芯片的总体水平差不多处在刚刚实现零突破的阶段，虽然涉足的领域都规模较小，但前景依然值得期待。



埃克特 - 莫奇利奖获得者

2024 胡文美
 2023年 昆勒·奥卢科通
 2022年 马克·霍洛维茨
 2021年 玛格丽特·马托诺西
 2020年 路易斯·A·巴尔索
 2019年 马克·D·希尔

胡因在多代处理器架构的设计和应用方面做出的开创性和基础性贡献而受到认可。他的基础性和开创性贡献对三代处理器架构产生了广泛影响：超标量、超长指令字和面向吞吐量的多核处理器 (GPU)。

1988 渡边哲 (注: T. Watanabe一般音译, 这里按常见日本人名姓在前名在后处理)
 1997 罗伯特·托马苏洛
 1996 耶鲁·帕特
 1995年 克劳福德·约翰
 1994年 桑顿·詹姆斯·E.
 1993年 大卫·J·库克
 1992年 迈克尔·J·弗林
 1991年 伯顿·J·史密斯
 1990年 肯尼斯·E·巴彻

建筑界的大人物(2)



□2001 亨尼西，约翰
 因其是MIPS计算机系统的创始人兼首席架构师，并为具有里程碑意义的MIPS R2000微处理器的研发做出贡献。

什么是“大香蕉”？

莫奇利奖

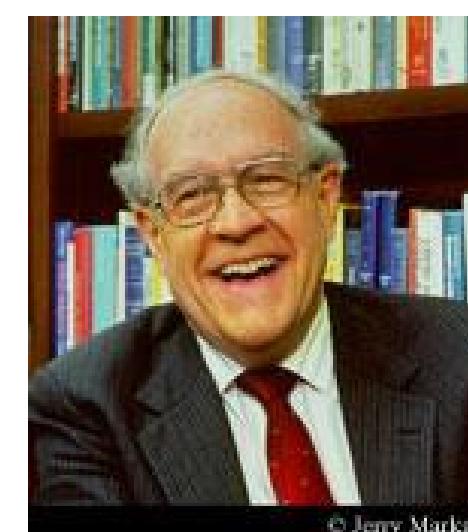
>> <http://www.computer.org/portal/web/awards/eckert>
 由美国计算机协会 (ACM) 和电气与电子工程师协会计算机学会 (IEEE Computer Society) 联合管理。该奖项奖金为5000美元，旨在表彰在计算机和数字系统架构领域做出的贡献，目前计算机架构领域被认为涵盖计算和数字系统的软硬件联合设计与分析。

架构领域的大人物(1)



□2007 马特奥·巴莱罗
<http://personals.ac.upc.edu/mateo/>
 因在指令级并行性和超标量处理器设计方面的重要贡献。

建筑界的大人物(3)



© Jerry Markatos

弗雷德里克·P·布鲁克斯

<http://www.cs.unc.edu/~brooks/>

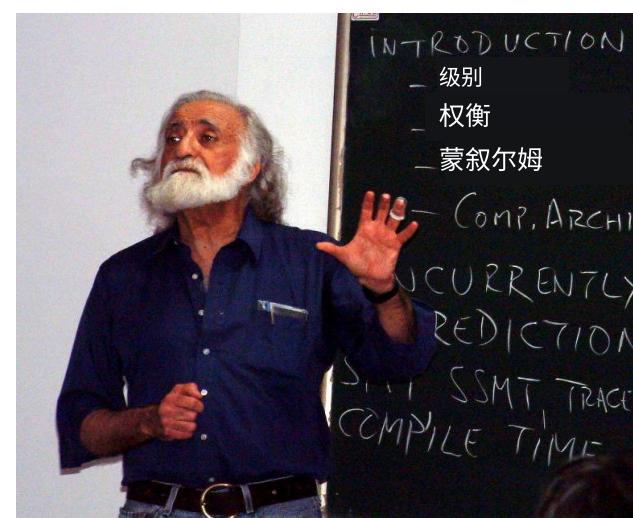
2004年埃克特 - 莫齐利奖“因对计算机体系结构的定义，以及对计算机系列概念和指令集设计原则的贡献；因在指令排序方面的开创性贡献，包括中断系统和执行指令；以及因对IBM 360指令集的贡献

体系结构。”
 □1999年ACM图灵奖
 对计算机体系结构、操作系统和软件工程做出了具有里程碑意义的贡献。

建筑界的大人物(5)

- 1996 耶鲁·帕特尔
因在指令级并行性和超标量处理器设计方面的重要贡献。

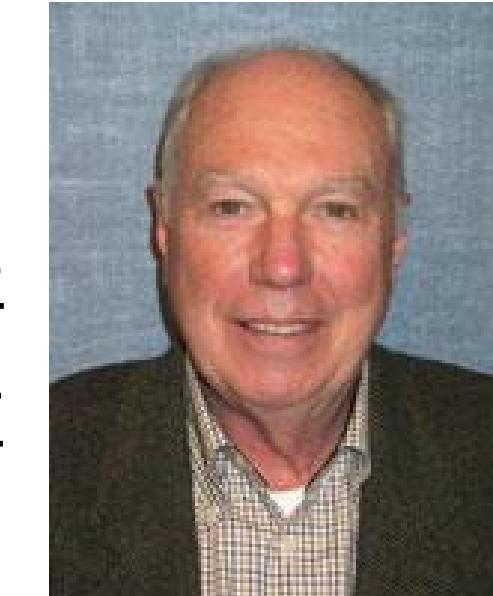
- 计算机系统导论 (2013 - 2017年, 2019年)



建筑界的大人物(6)

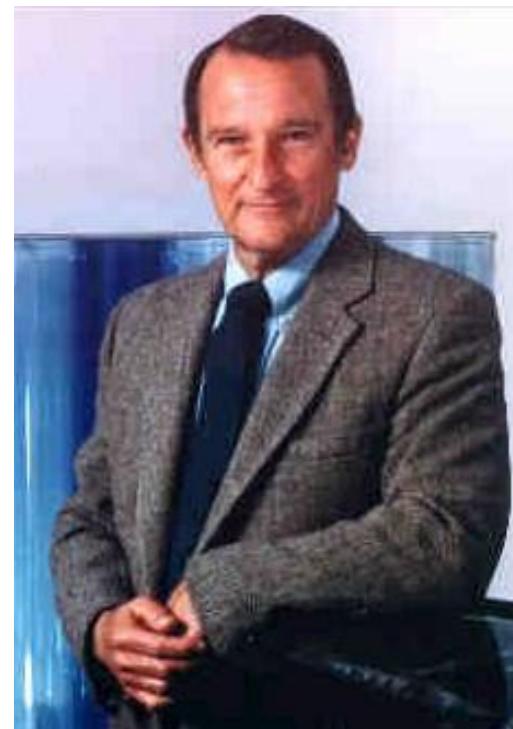
- 1992 迈克尔·J·弗林
<http://www.cpe.calpoly.edu/IAB/flynn.html>

□ 因其在处理器组织与分类、计算机算术和性能评估方面做出的重要开创性贡献。



建筑界的大人物 (7)

- 1989年 西摩·克雷
- 因其在推动超级计算机设计方面取得的一系列成就。

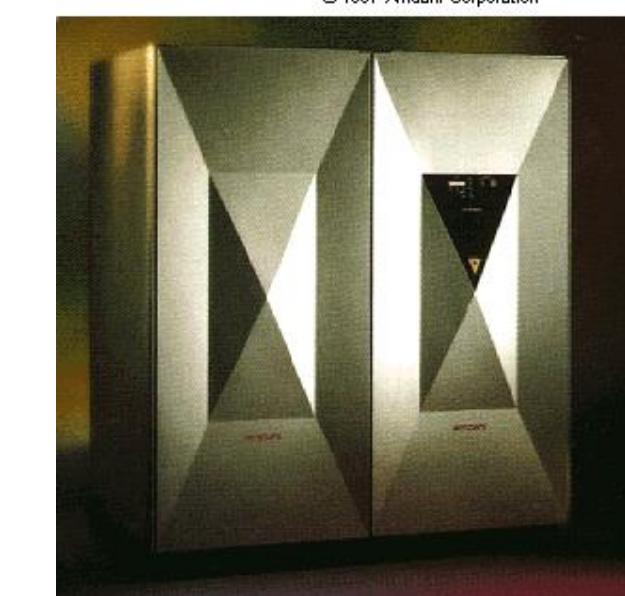


建筑界的大人物(8)

摘自《计算机桌面百科全书》
经授权转载。
② 1999 年 吉恩·M·阿姆达尔博士



1975 年，阿姆达尔博士站在威斯康星整体同步他设计的计算机 (WISC) 于1950年设计。它于1952年制造完成。(图片由吉恩·M·阿姆达尔博士提供。)



在计算机体系结构领域关于前瞻，以及

顶级会议和期刊

□ 顶级会议：

- > ISCA

> 微架构会议 (MICRO)

> 体系结构支持编程语言和操作系统大会 (ASPLOS)、超算大会 (SC)、高性能计算机体系结构研讨会 (HPCA)、设计自动化大会 (DAC)

□ 顶级期刊：

> 《计算机系统汇刊》(TOCS)：美国计算机协会计算机系统汇刊

> 《存储汇刊》(TOCS)：美国计算机协会存储汇刊

> 《计算机辅助设计汇刊》(TCAD)：电气与电子工程师协会集成电路与系统计算机辅助设计汇刊

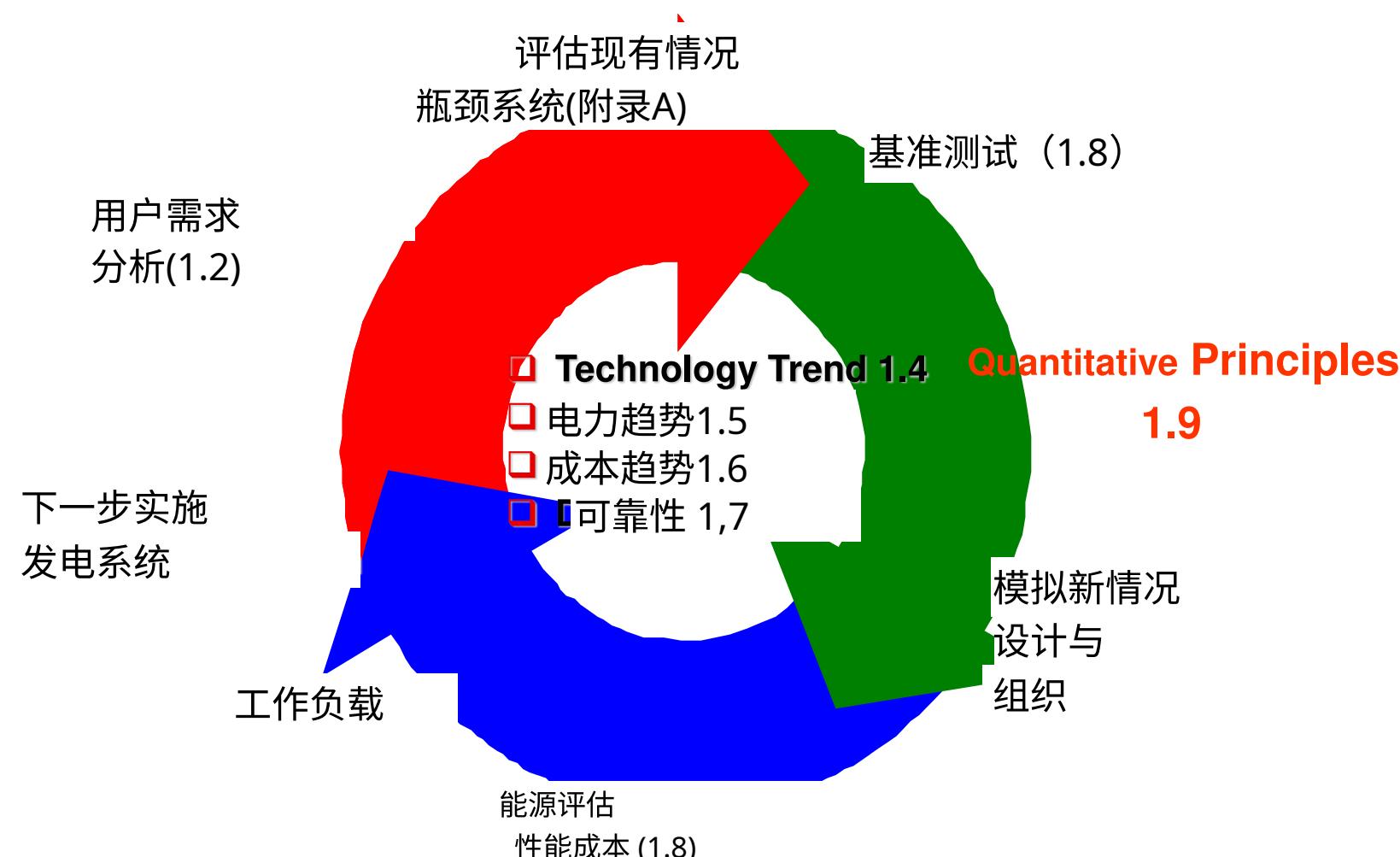
> 技术委员会：《电气与电子工程师协会计算机汇刊》

并行与分布式系统性能指标：《电气与电子工程师协会并行与分布式系统汇刊》

- 《计算机体系结构与代码优化汇刊》

□ 细胞自动机设计师的任务是什么？

计算机设计工程生命周期



第1章主题

- 1.1 引言
- 1.2 计算机类别
- 1.3 定义计算机体系结构以及计算机设计的任务是什么?
- 1.4 技术趋势
- 1.5 集成电路的功耗趋势
- 1.6 成本趋势
- 1.7 可靠性
- 1.8 测量、报告和总结性能
- 1.9 计算机设计的定量原则
- 1.10 总结归纳

总结：计算机设计任务

□考虑因素：

> 功能和非功能需求

> 实现复杂度

•复杂的设计完成时间更长

•复杂的设计必须提供更高的性能才能具有竞争力

> 技术趋势

•不仅要考虑当前可用的技术，还要考虑系统

准备发布时将可用的技术。（稍后会详细介绍）

集成电路功耗趋势 > 成本趋势 □ 论据 > 评估现有系统

的瓶颈 □ 定量原则

阅读作业

□第1章，附录A。

□第1章的作业1将发布在网站上

□ 结束！

□ 谢谢！