

计算机组织与体系结构

第三讲

计算机科学与技术学院

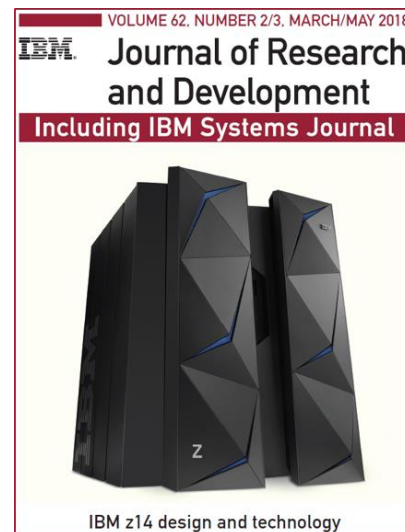
舒燕君

Recap

- 计算机的基本组成
 - ✓ 计算机硬件技术指标
- 计算机体系结构的概念
 - ✓ Amdal提出的经典体系结构概念
 - ✓ 现代计算机体系结构的概念
 - ✓ 计算机组成和物理实现
 - ✓ 系列机和兼容

2.1.1 计算机体系结构的概念

- 阿姆道尔 (C. M. Amdahl) 首次明确
- 计算机体系结构是程序员所看到的计算机的属性，即概念性结构与功能特性
 - 1964年4月，Architecture of the IBM System/360，发表在IBM Journal of Research and Development上
 - 计算机体系结构概念的经典定义



2.1.2 计算机体系结构、组成和实现

- 体系结构的概念用于描述计算机系统设计的技术、方法和理论，包括以下三个方面：
 - (1) 计算机指令系统
 - (2) 计算机组成
 - (3) 计算机硬件实现
- 涵盖处理器和多处理器、存储器、输入/输出系统、互联与通信等计算机系统设计的主要内容
- 还涉及到性能评价、编译和操作系统技术
- 通过定量分析的途径，学习掌握现代计算机体系结构研究的基本方法

计算机组成

- 指令集结构的逻辑实现
 - 数据通路宽度
 - 专用功能部件的设置
 - 功能部件的并行性
 - 缓冲和排队技术
 - 预测技术
 - 可靠性技术
 - 控制机构的组成，等等

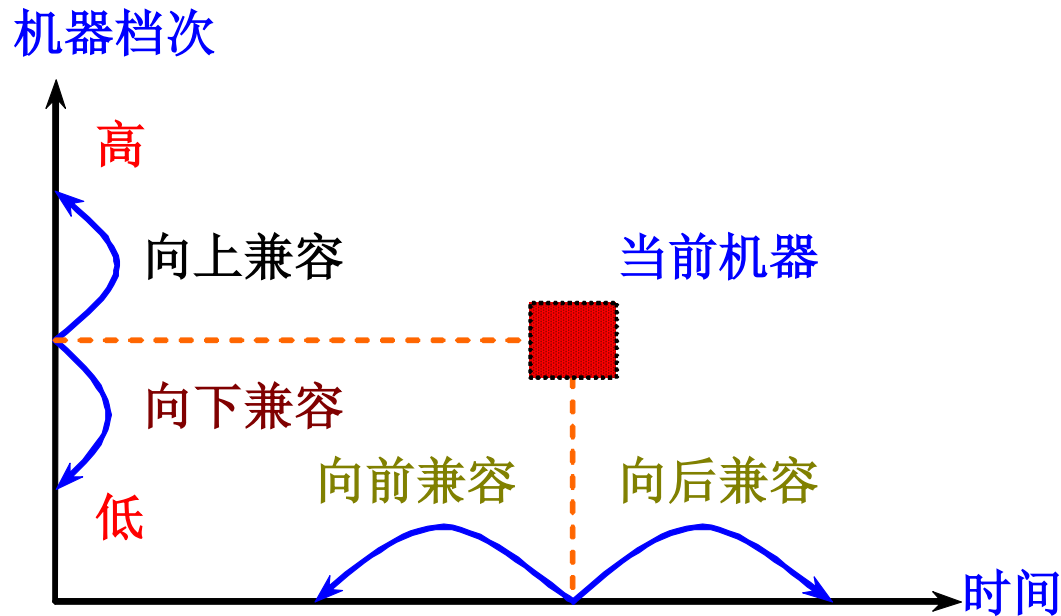
计算机的实现

- 处理器、主存的物理结构
- 器件的集成度和速度
- 信号传输
- 器件、模块、插件、底板的划分与连接
- 涉及的专用器件
- 电源、冷却
- 微组装技术
- 整机装配技术，等等



兼容性

- 向上(下)兼容指的是按某档机器编制的程序，不加修改的就能运行于比它高(低)档的机器
- 向前(后)兼容指的是按某个时期投入市场的某种型号机器编制的程序，不加修改地就能运行于在它之前(后)投入市场的机器



第 1 章 计算机系统概论

第 2 章 计算机系统量化分析基础

第 3 章 总线

第 4 章 指令系统

第 5 章 CPU设计

第 6 章 基本流水线技术

第 7 章 指令级并行

第 8 章 存储系统的结构与优化

第 9 章 IO系统

第 2 章 计算机系统量化分析基础

2.1 计算机体系结构的概念

2.2 计算机体系结构的发展

2.3 计算机系统设计和分析



2.2 计算机体系结构的发展

2.2.1 计算机分代和分型

2.2.2 软件的发展

2.2.3 应用的发展

2.2.4 相关核心技术的发展

2.2.5 体系结构的发展

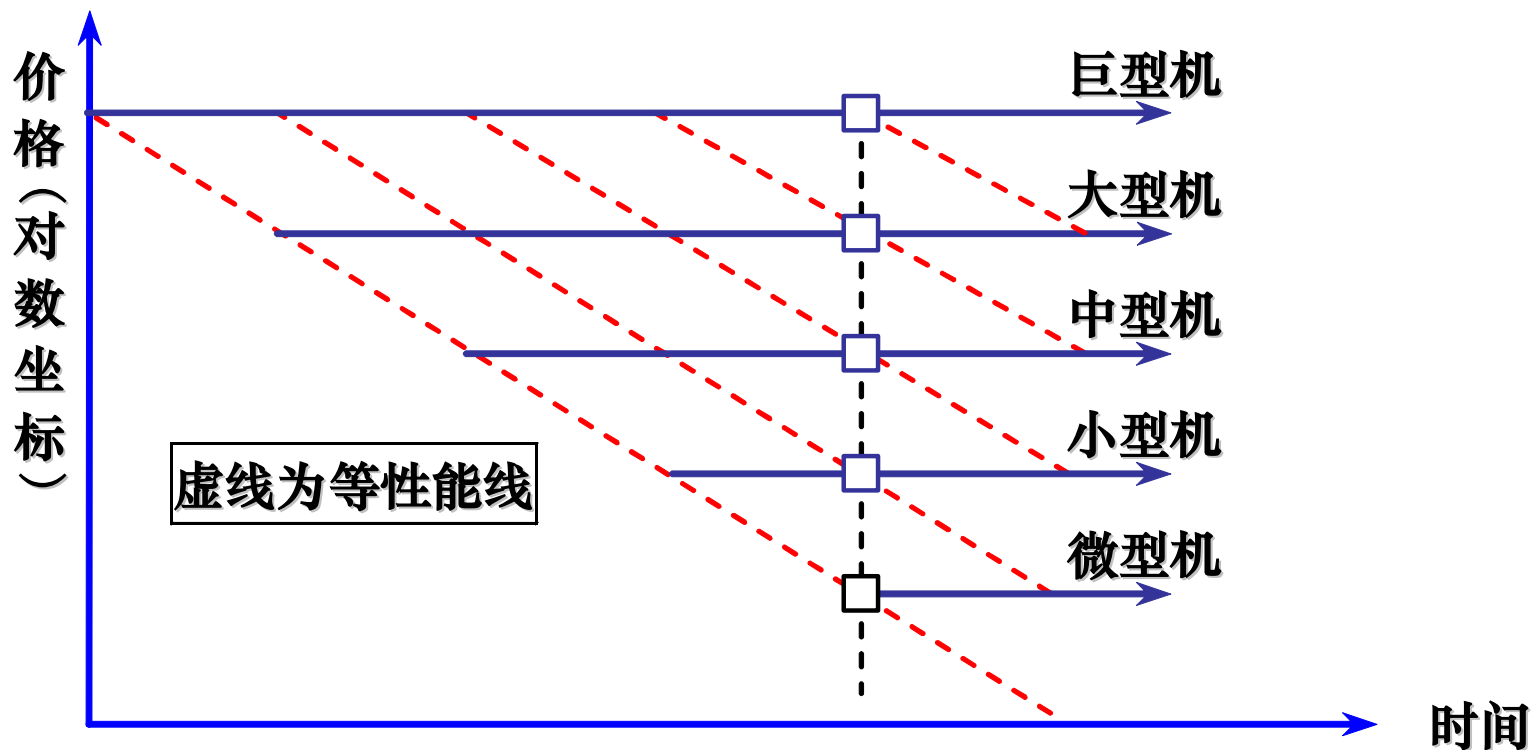
2.2.6 并行处理技术的发展

2.2.7 处理器芯片技术的发展

2.2.1 计算机的分代和分型

分代	器件	体系结构技术	软件技术	典型机器
第一代 (1945-1954)	电子管和继电器	存储程序计算机、 程序控制I/O	机器语言和汇编语言	普林斯顿ISA、 ENIAC、 IBM701
第二代 (1955-1964)	晶体管、磁芯、印刷电路	浮点数据表示、寻址技术、中断、 I/O处理机	高级语言和编译、批处理监控系统	Univac LARC、 CDC1604、 IBM7030
第三代 (1965-1974)	SSI和MSI、多层印刷电路、微程序	流水线、Cache、先行处理、系列计算机	多道程序和分时操作系统	IBM360/370、 CDC6600/7600、 DEC PDP-8
第四代 (1974-1990)	LSI和VLSI、半导体存储器	向量处理、分布式存储器	并行与分布处理	Cray-1、 IBM 3090、 DEC VAX9000、 银河系列
第五代 (1991-)	高性能微处理器、大规模高密度电路	指令级并行、SMP、MP、MPP、网络	可扩展并行与分布处理	SGI Jaguar、 IBM xServer 天河系列

2.2.1 计算机的分代和分型



技术和性能的“下移”。新型体系结构的设计一方面是合理地增加计算机系统中硬件的功能比例。另一方面则是通过多种途径提高计算机体系结构中的并行性。

2.2.2 软件的发展

程序和数据使用的存储器容量不断增大。

1. 计算机语言与编译技术

2. 操作系统

3. 软件工具与中间件

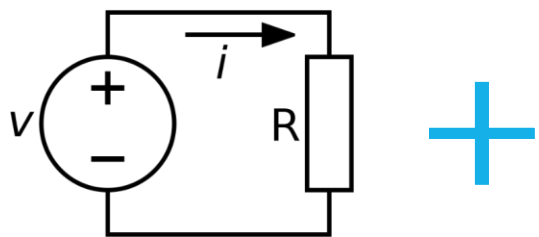
2.2.3 应用的发展

- 嵌入式计算机
包括移动设备和掌上计算机（关注功耗、成本等）
- 台式计算机
对性价比要求最高
- 服务器市场
可用性、高流量密度和可扩展性
- 数据中心
类似于服务器的要求，电能和散热系统很重要
- 巨型机
面向浮点运算的数据中心计算机，内部网络需要更高的带宽和更低的延迟

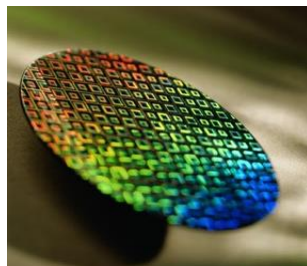
2.2.4 相关核心技术的发展

1. 集成电路逻辑技术 摩尔定律

芯片(Chip)是一种把电路（如电源、电阻、电容、电感、二极管、晶体管、集成电路和电键等，构成的网络、硬件）小型化的方式，并制造在半导体晶圆表面上。



电路



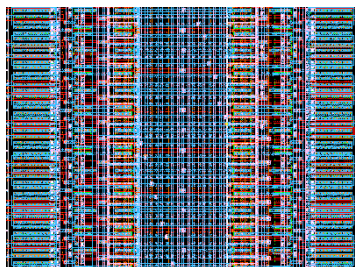
晶圆



芯片

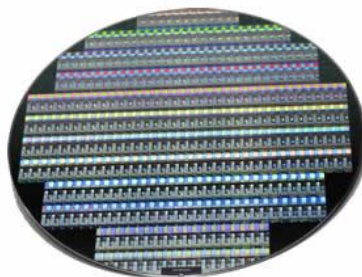
2.2.4 相关核心技术的发展

1. 集成电路逻辑技术 摩尔定律



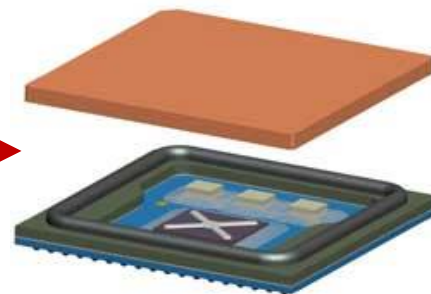
芯片设计

集成电路设计涉及对电子器件和器件间互连线模型的创建。



晶圆加工

晶圆加工是一系列化学处理步骤，使得电子电路逐渐形成在使用纯半导体材料制作的芯片上。

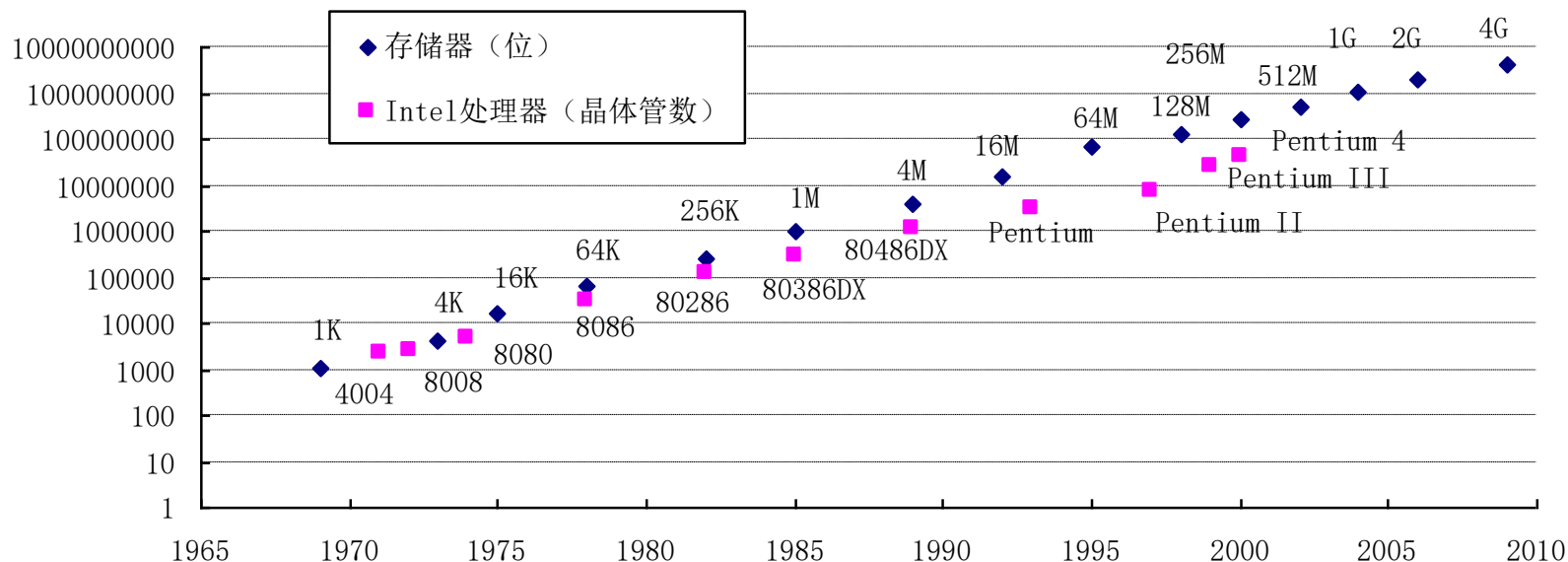


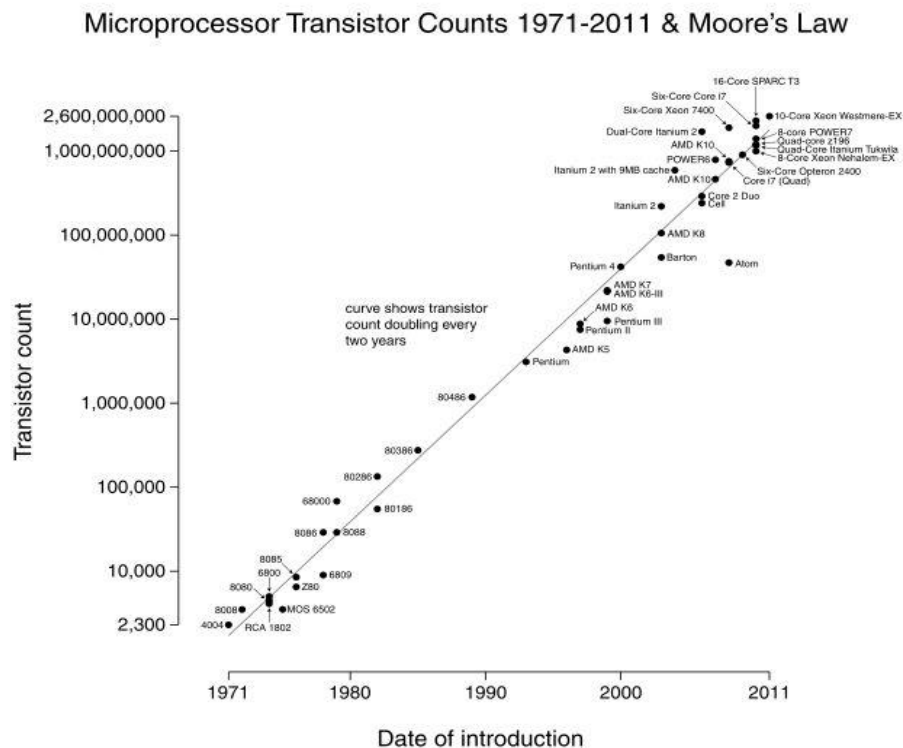
芯片封装测试

封装是将器件的核心晶粒封装在一个支撑物之内的过程，这个封装可以防止物理损坏以及化学腐蚀，并提供对外连接的引脚，之后将进行集成电路性能测试。

2.2.4 相关核心技术的发展

- 1965年，时任仙童公司研发实验室主任的摩尔（Gordon Mooer）在《Electronics》上撰文，认为集成电路密度大约每两年翻一番。
- 40年来，摩尔定律不但印证了集成电路技术的发展，也印证了计算机技术的发展。





2.2.4 相关核心技术产品的发展

2. DRAM (动态随机访问存储器)

单个DRAM模块的容量每年增加25-40%，增速逐年下降。

CA:AQA Edition	Year	DRAM growth rate	Characterization of impact on DRAM capacity
1	1990	60%/year	Quadrupling every 3 years
2	1996	60%/year	Quadrupling every 3 years
3	2003	40%–60%/year	Quadrupling every 3 to 4 years
4	2007	40%/year	Doubling every 2 years
5	2011	25%–40%/year	Doubling every 2 to 3 years

2.2.4 相关核心技术产品的发展

3. 闪存（快擦型存储器）

近几年，每年约50-60%的速度增长，大约每两年翻一番，大量用于便携式设备。

4. 磁盘

从04年开始，大约每年增长40%，约每3年翻一番。
SSD的成本的迅速下降。

5. 网络

2.2.5 体系结构的发展

1. 分布的I/O处理能力。
2. 保护的存储器空间。
3. 存储器组织结构的发展。
4. 并行处理技术
5. 指令集发展

2.2.6 并行处理技术的发展

1 并行性的概念

- **并行性**：计算机系统在同一时刻或者同一时间间隔内进行多种运算或操作。

只要在时间上相互重叠，就存在并行性。

- **同时性**：两个或两个以上的事件在同一时刻发生。
- **并发性**：两个或两个以上的事件在同一时间间隔内发生。

2.2.6 并行处理技术的发展

- 从**执行程序**的角度来看，并行性等级从低到高可分为：
 - **指令内部并行**：单条指令中各微操作之间的并行。
 - **指令级并行**：并行执行两条或两条以上的指令。
 - **线程级并行**：并行执行两个或两个以上的线程。
通常是以一个进程内派生的多个线程为调度单位。
 - **任务级或过程级并行**：并行执行两个或两个以上的过程或任务（程序段），以子程序或进程为调度单元。
 - **作业或程序级并行**：并行执行两个或两个以上的作业或程序。

2.2.6 并行处理技术的发展

- 从处理数据的角度来看，并行性等级从低到高可分为：
 - 字串位串：每次只对一个字的一位进行处理。
最基本的串行处理方式，不存在并行性。
 - 字串位并：同时对一个字的全部位进行处理，不同字之间是串行的。
开始出现并行性。
 - 字并位串：同时对许多字的同一位（称为位片）进行处理。
具有较高的并行性。
 - 全并行：同时对许多字的全部位或部分位进行处理。
最高一级的并行。

2.2.6 并行处理技术的发展

- 计算机系统结构的Flynn分类法
 - Flynn分类法 按照指令和数据的关系，把计算机系统的结构分为4类
 - 单指令流单数据流SISD
(Single Instruction stream Single Data stream)
 - 单指令流多数据流SIMD
(Single Instruction stream Multiple Data stream)
 - 多指令流单数据流MISD
(Multiple Instruction stream Single Data stream)
 - 多指令流多数据流MIMD
(Multiple Instruction stream Multiple Data stream)

2.2.6 并行处理技术的发展

2 提高并行性的技术途径

三种途径：

(1) 时间重叠

引入时间因素，让多个处理过程在时间上相互错开，轮流重叠地使用同一套硬件设备的各个部分，以加快硬件周转而赢得速度。

(2) 资源重复

引入空间因素，以数量取胜。通过重复设置硬件资源，大幅度地提高计算机系统的性能。

(3) 资源共享

这是一种软件方法，它使多个任务按一定时间顺序轮流使用同一套硬件设备。

2.2.6 并行处理技术的发展

单机系统中并行性的发展

(1) 在发展高性能单处理机过程中，起主导作用的是**时间重叠**原理。

实现时间重叠的基础：**部件功能专用化**

- ✓ 把一件工作按功能分割为若干相互联系的部分；
- ✓ 把每一部分指定给专门的部件完成；
- ✓ 然后按时间重叠原理把各部分的执行过程在时间上重叠起来，使所有部件依次分工完成一组同样的工作。

单机系统中并行性的发展

(2) 在单处理机中，**资源重复**原理的运用也已经十分普遍。

- 多体存储器
- 多操作部件
 - 通用部件被分解成若干个专用部件，如加法部件、乘法部件、除法部件、逻辑运算部件等，而且同一种部件也可以重复设置多个。
 - 只要指令所需的操作部件空闲，就可以开始执行这条指令（如果操作数已准备好的话）。

— 阵列处理机（并行处理机）

更进一步，设置许多相同的处理单元，让它们在同一个控制器的指挥下，按照同一条指令的要求，对向量或数组的各元素同时进行同一操作，就形成了阵列处理机。

(3) 在单处理机中，**资源共享**的概念实质上是用单处理机模拟多处理机的功能。

2.2.6 并行处理技术的发展

(2) 多机系统中并行性的发展

多机系统遵循时间重叠、资源重复、资源共享原理，发展为3种不同的多处理机：

同构型多处理机、异构型多处理机、分布式系统

耦合度 反映多机系统中各机器之间物理连接的紧密程度和交互作用能力的强弱。

- **最低耦合**：耦合度最低，除通过某种中间存储介质之外，各计算机之间没有物理连接，也无共享的联机硬件资源。
- **松散耦合系统（间接耦合系统）**：一般是通过通道或通信线路实现计算机之间的互连，可以共享外存设备（磁盘、磁带等）。机器之间的相互作用是在文件或数据集一级上进行。
- **紧密耦合系统（直接耦合系统）**：在这种系统中，计算机之间的物理连接的频带较高，一般是通过总线或高速开关互连，可以共享主存。

2.2.6 并行处理技术的发展

3. 并行计算的应用需求

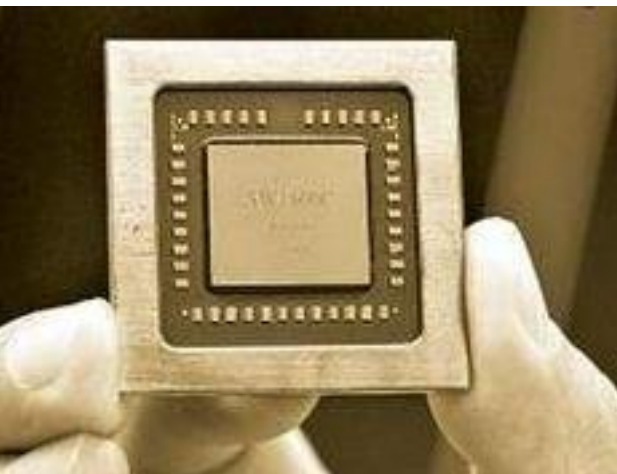
- 科学计算
- 娱乐业

对ZFLOPS计算能力和EB存储能力需求的领域在不断增多。

我国的高性能计算机

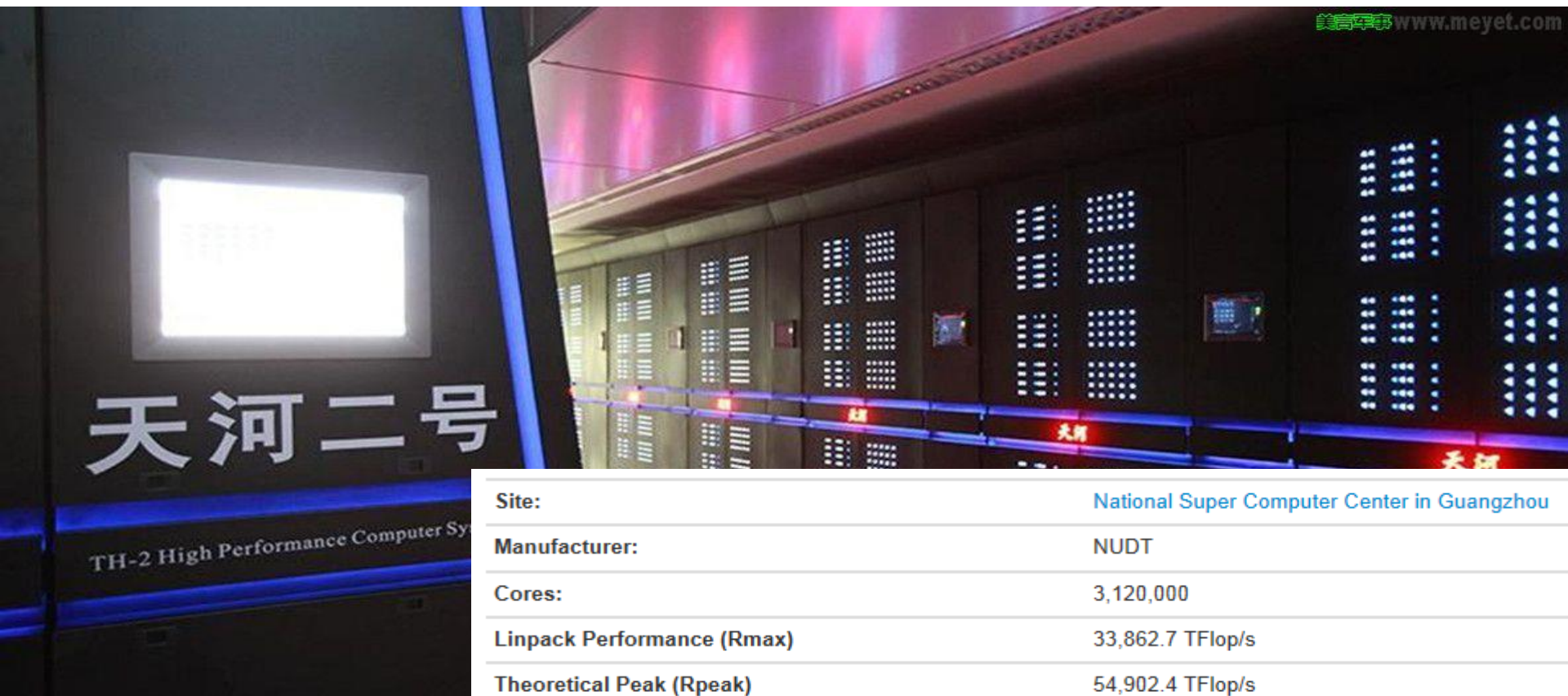
1. 银河-1(中国第一台大型机)
2. 银河-2
3. 银河-3
4. 天河一号
5. 天河二号
6. 神威
7. 曙光3000
8. 曙光4000
9. 曙光5000A

神威 太湖之光

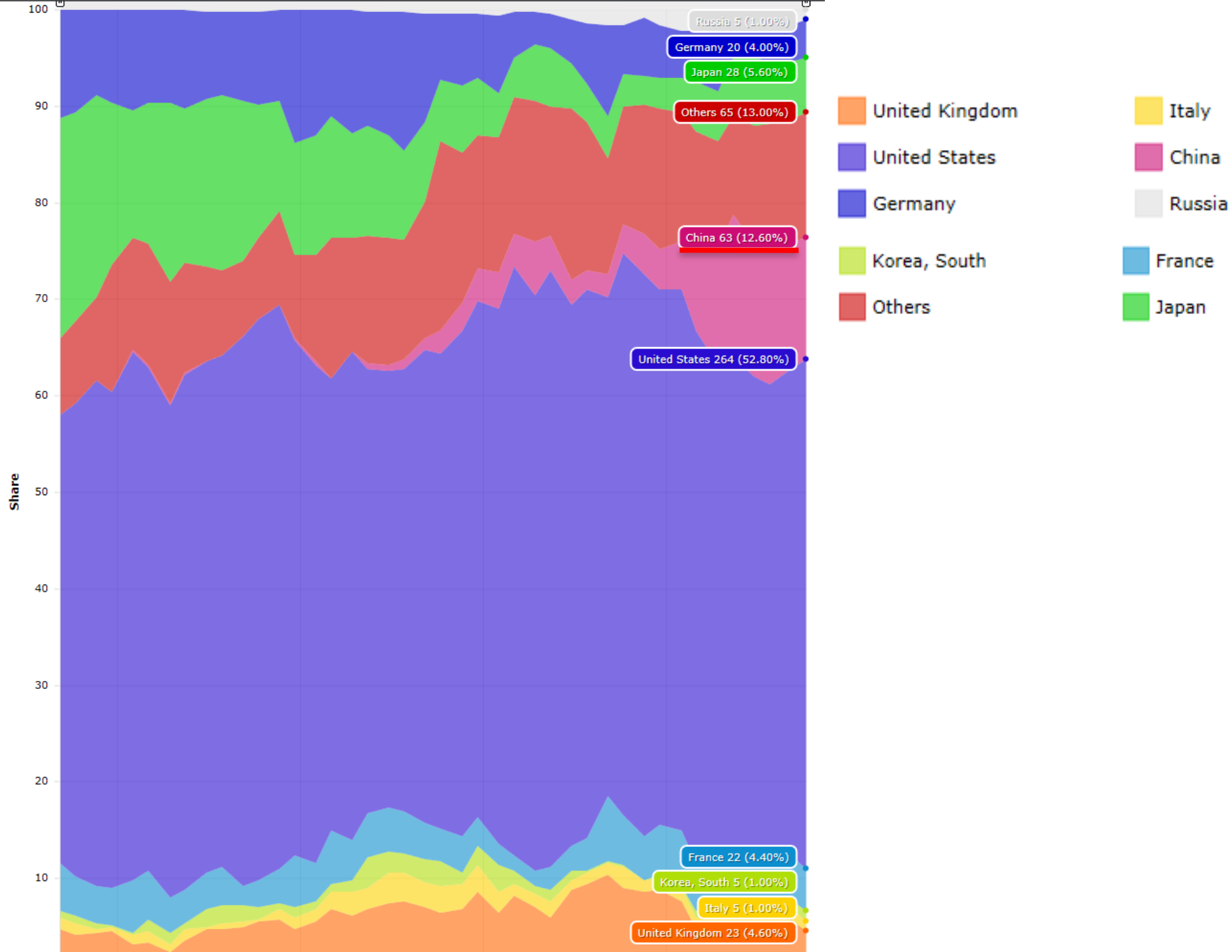


Site:	National Supercomputing Center in Wuxi
Manufacturer:	NRCP
Cores:	10,649,600
Linpack Performance (Rmax)	93,014.6 TFlop/s
Theoretical Peak (Rpeak)	125,436 TFlop/s
Nmax	12,288,000
Power:	15,371.00 kW (Submitted)
Memory:	1,310,720 GB
Processor:	Sunway SW26010 260C 1.45GHz
Interconnect:	Sunway
Operating System:	Sunway RaiseOS 2.0.5

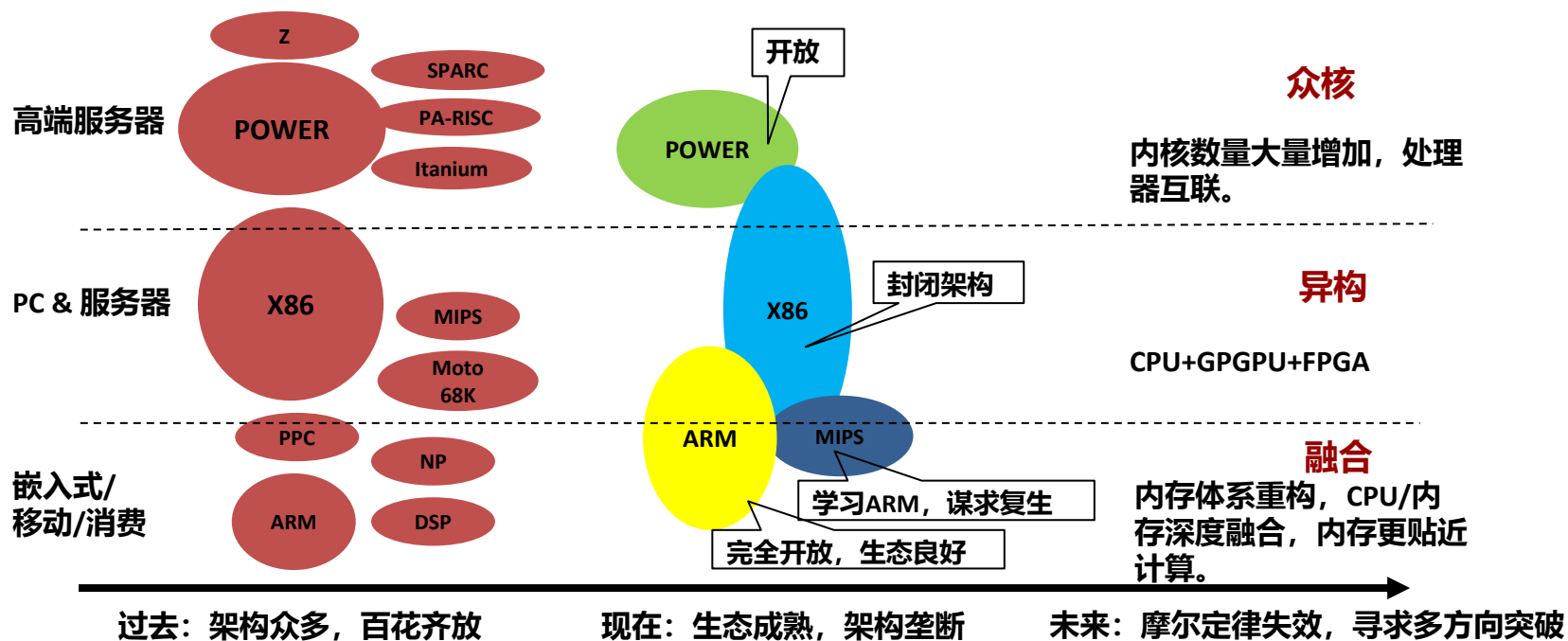
Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P



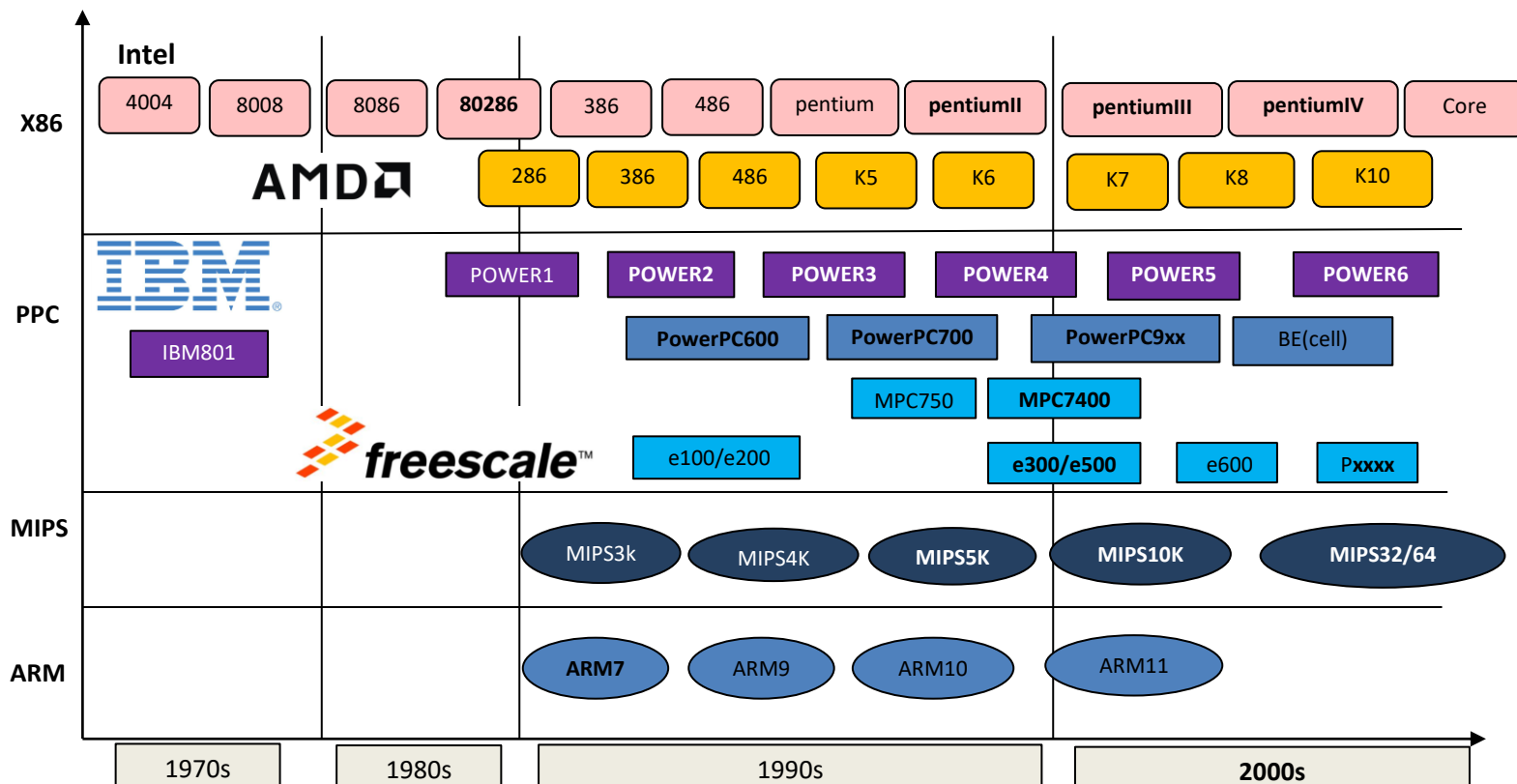
Site:	National Super Computer Center in Guangzhou
Manufacturer:	NUDT
Cores:	3,120,000
Linpack Performance (Rmax)	33,862.7 TFlop/s
Theoretical Peak (Rpeak)	54,902.4 TFlop/s
Power:	17,808.00 kW
Memory:	1,024,000 GB
Interconnect:	TH Express-2
Operating System:	Kylin Linux
Compiler:	icc
Math Library:	Intel MKL-11.0.0
MPI:	MPICH2 with a customized GLEX channel



2.2.7 处理器芯片技术的发展

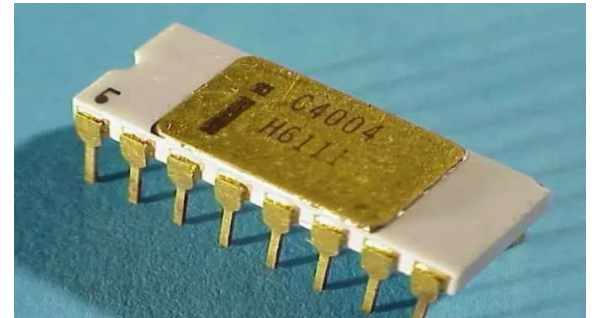


主流CPU发展路径



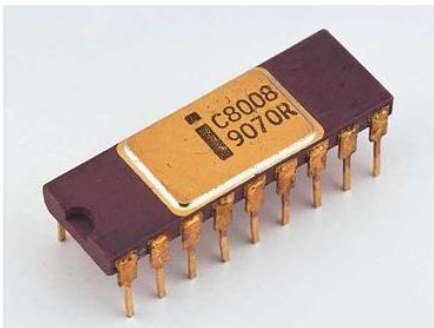
4位处理器： 4004

- 1971年11月15日，Intel公司的工程师霍夫发明了世界上第一个商用微处理器——4004，从此这一天被当作具有全球IT界里程碑意义的日子而被永远的载入了史册。
- 这款4位微处理器45条指令，每秒能执行5万条指令，运行速度为108kHz, 集成了2300只晶体管。



8/16/32位处理器

- 1972，1974年，8位处理器8008，8080推出。
- 1978年，首枚x86架构16位处理器8086推出，同年性能更出色的8088处理器推出。1982年，16位的80286处理器推出。
- 1985年，第一款32位处理器80386推出。1989年，最后一款以数字为编号的处理器80486推出。



Intel Pentium I/II/III/4处理器

- 1993年，英特尔发布了Pentium(奔腾)处理器。Pentium处理器集成了310万个晶体管，最初推出的初始频率是60MHz、66MHz，后来提升到200MHz以上。
- 1997年5月7日,英特尔发布Pentium II 处理器
- 1999年2月26日,英特尔发布Pentium III 处理器
- 2000年，英特尔发布Pentium 4处理器。



i3, i5, i7诞生

- 2011年3月,使用32nm工艺全新桌面级和移动端处理器采用了i3, i5和i7的产品分级架构。
- 其中i3主攻低端市场,采用双核处理器架构,约2MB二级缓存,售价500-800元人民币; i5处理器主攻主流市场,采用四核处理器架构, 4MB二级缓存,售价900-1500元人民币; i7主攻高端市场,四核八线程或六核+二线程架构,二级缓存不少于8MB,售价2100-7600元人民币。



MIPS处理器

- MIPS的意思是“无内部互锁流水级的微处理”(Microprocessor without interlocked piped stages), 其机制是尽量避免流水线中的数据相关问题。MIPS技术公司是美国著名的芯片设计公司, 它采用精简指令系统计算结构(RISC)来设计芯片。
- MIPS是出现最早的商业RISC架构芯片之一, 新的架构集成了所有原来MIPS指令集, 并增加了许多更强大的功能。MIPS自己只进行CPU的设计, 之后把设计方案授权给客户, 使得客户能够制造出高性能的CPU。

MIPS发展史

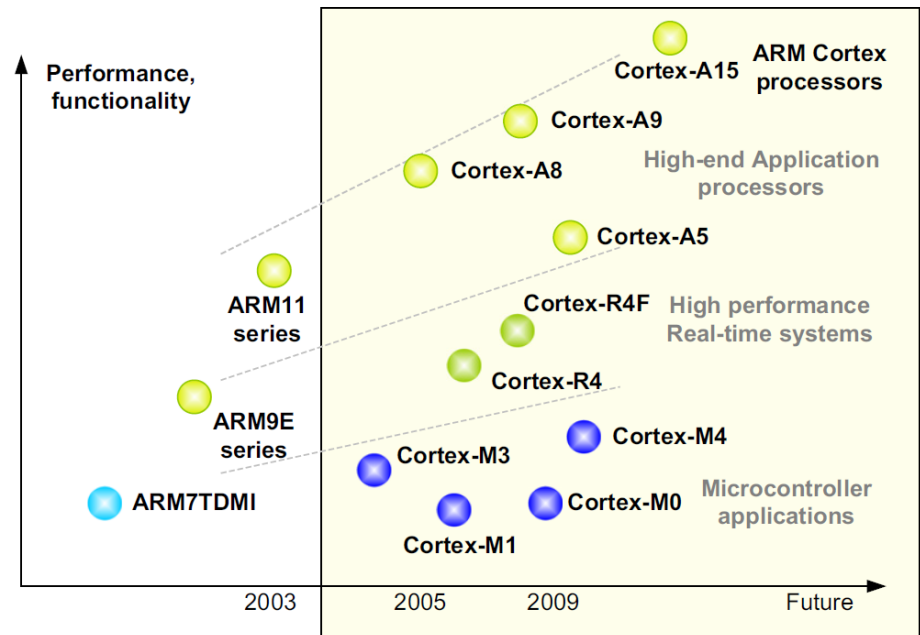
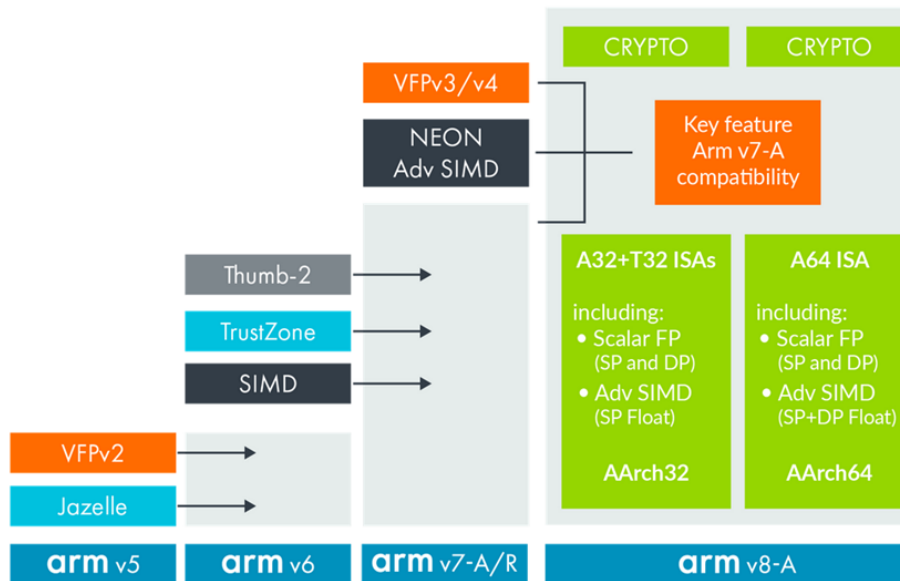
- 1984年，MIPS计算机公司成立，开始设计RISC处理器；
- 1986年推出R2000处理器。
- 1988年推出R3000处理器。
- 1991年推出第一款64位商用微处理器R4000；之后又陆续推出R8000（于1994年）、R10000（于1996年）和R12000（于1997年）等型号。
- 1999年，MIPS公司发布MIPS32和MIPS64架构标准，为未来MIPS处理器的开发奠定了基础。2000年，MIPS公司发布了针对MIPS32 4Kc的版本以及64位MIPS 64 20Kc处理器内核。
- 2007年8月16日—MIPS科技宣布，中科院计算机研究所的龙芯中央处理器获得其处理器IP的全部专利和总线、指令集授权。

POWERPC处理器

- POWER是1991年，Apple、IBM、Motorola组成的AIM联盟所发展出的微处理器架构。PowerPC是整个AIM平台的一部分，并且是到目前为止唯一的一部分。PowerPC系列是源自于POWER架构的设计,但进行了大量的改动。
- PowerPC处理器有广泛的实现范围，包括从诸如 Power4 高端服务器CPU 到嵌入式CPU市场（任天堂 GameCube 使用了 PowerPC）。



ARM架构发展史



国产ARM处理器： 鲲鹏

- 鲲鹏处理器是华为在2019年1月向业界发布的高性能处理器。
- 鲲鹏920是目前业界最高性能ARM-based处理器。该处理器采用7nm制造工艺，基于ARM架构授权，由华为公司自主设计完成。鲲鹏920以更低功耗为数据中心提供更强性能。



基于ARMv8架构的鲲鹏处理器



鲲鹏916

支持多路互联的ARM处理器

- 32核, 2.4 GHz主频
- SPECint性能匹配业界中端, 功耗低至75 W
- 支持4通道DDR4控制器
- 支持PCIe 3.0和SAS/SATA 3.0
- 集成板载GE/10 GE网络
- 支持2路互联



鲲鹏920

7nm制程, 数据中心ARM处理器

- 计算核数提升1倍, 最多64核
- SPECint性能提升超过2倍
- 内存通道数提升1倍, 支持8通道DDR4控制器
- 支持PCIe 4.0和CCIX
- 集成板载100 GE网络和加密、压缩等引擎
- 支持2路或4路互联

国产ARM处理器：飞腾

- 飞腾处理器，又称银河飞腾处理器，是由中国人民解放军国防科学技术大学研制的一系列嵌入式数字信号处理器（DSP）和中央处理器（CPU）芯片。其商业化推广是由天津飞腾信息技术有限公司负责。
- 早期的飞腾系列处理器使用的是SPARCv9架构，从飞腾-1500A开始，后续的处理​​器均使用ARMv8架构。



基于ARMv8架构的飞腾处理器

飞腾自主研发处理器

微内核

FTC660

FTC661

FTC662

FTC663

处理器



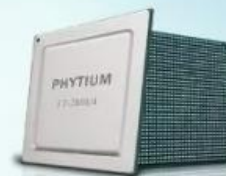
FT-1500A/4
FT-1500A/16



FT-2000A/2



FT-2000+/64



新款
FT-2000/4