高级计算机结构

雷航

hlei@uestc.edu.cn

“什么是计算机系统结构?”

“计算机系统结构由程序设计者所看到的计算机系统的属性, 即概念性结构和功能特性, 这里所说的程序设计者是指机器语言或编译程序设计者”

(1964年, Amdahl(阿姆达尔))

“缔造了IBM360的辉煌, 奠定了IBM计算机帝国的江山, 被《计算机世界》列为改变世界的25人之一”

多级层次结构中机器语言级的结构,**是程序员所看到的计算机的属性**, 它是软/硬件之间的主要交界面

“过去, 计算机系统结构通常是指指令系统设计, 计算机设计的其他方面则称为实现, 这暗示计算机的实现技术不具有挑战性, 我们确信这种说法是不正确的。系统结构的设计者在其他方面所遇到的困难比指令系统设计遇到的困难更具有挑战性”

(John L hennessy, 约翰.亨利斯, 斯坦福大学校长(1999年-2016年), 美国工程院院士, 第一个商用RISC处理器(MISP处理器)的研发者), 2017年图灵奖获得者。

“计算机系统结构包含：
计算机指令系统、组成和硬件”

组成: 涵盖计算机设计的多方面, 如存储系统、CPU的设计等。

硬件: 计算机的具体逻辑设计等实现技术, 即使相同指令集的计算机, 也可能具有不同硬件实现。

(注: 不同的硬件实现将影响结构与性能)

– 系统结构的一种分类方法和设计准则

按“流”分类的方法: Flynn(1966年)根据系统的指令流和数据流对计算机系统进行分类。

(1) SISD 单指令流单数据流:

如传统的单处理机系统

(2) SIMD 单指令流多数据流:

如高端微处理器和并行处理机系统

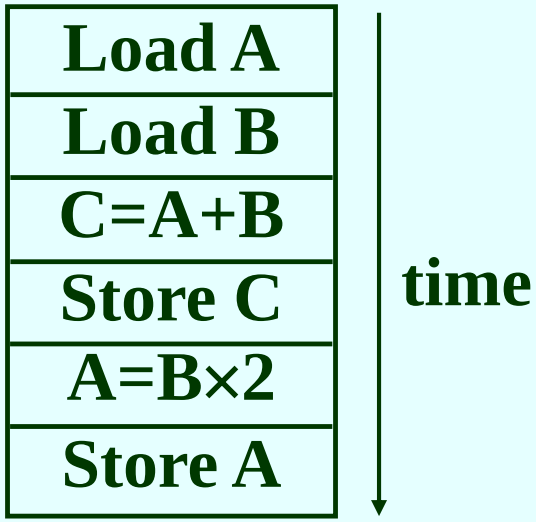
(3) MISD 多指令流单数据流:

实际上不存在, 但也有学者认为存在

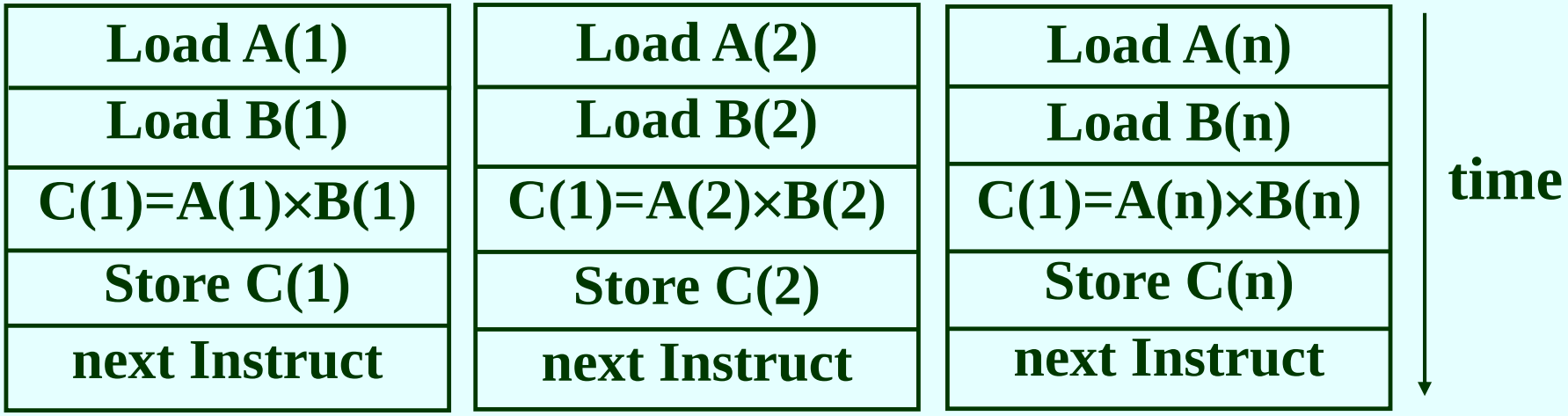
(4) MIMD 多指令流多数据流:

如大多数多处理机系统

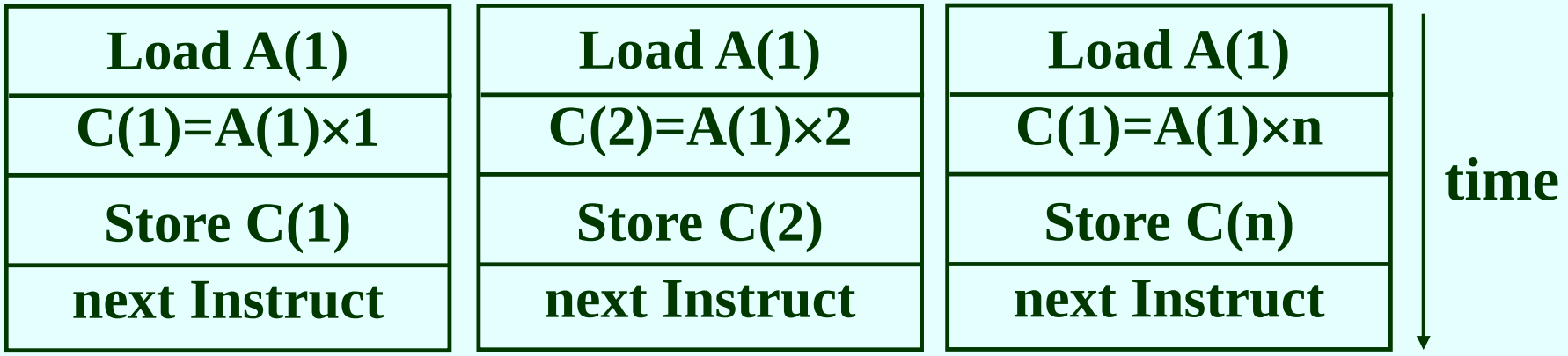
不同结构图例:



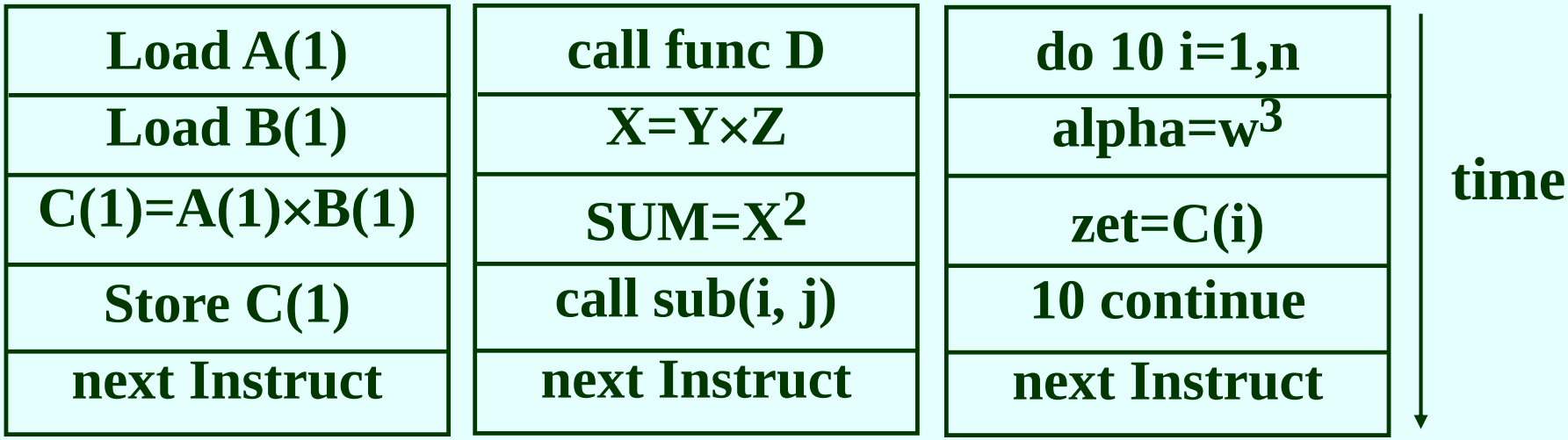
SISD结构



SIMD结构



MISD结构



MIMD结构

系统设计准则

1. 只加速使用频率高的部件

   最广泛采用的计算机设计准则。加快处理那些频繁出现的事件对系统的影响, 远比加速处理很少出现的事件的影响要大。

2. 阿姆达尔(Amdahl)定律

   系统中对某一部件采用更快执行方式, 所能获得的系统性能改进程度, 取决于这种执行方式被使用的频率, 或者所占总执行时间的比例。

3. 遵循程序访问的局部性规律

   包括时间局部性和空间局部性规律两个方面。

前 言

1、课程历史

现代微机系统

以Intel80286、80386为主

现代微机结构

Intel×86系列处理器、其它高端处理器、RISC技术、64位处理器、总线技术等

高级计算机结构

以《现代微机结构》内容为基础, 进一步结合计算机硬件技术发展状况进行讨论

2、本课程的主要内容和目的

主要内容

(1) 现代计算机硬件体系结构，着重介绍：

① 处理器体系结构(主流处理器：历史到现在)

② 采用的关键技术，如流水线、保护模式、超标量、指令重调度、超线程、多核技术、向量计算等，并以Intel处理器为典型代表。

(2) RISC处理器设计方法以及与CISC的比较

(3) 64位处理器

(4) 总线技术

(5) 计算机的一般性能评价方法等

学习目的

掌握现代计算机硬件系统发展状况、技术特征、新的技术和发展方向，了解新技术如何提高处理器以及计算机系统的性能，为进行计算机系统分析和研究、以及为其它专业课程的学习进行打下基础。同时，也从一个更高层面上学习计算机硬件技术。

课程特点

(1) 内容丰富，概念多，但难度不大；

(2) 系统结构与处理器相结合

前续课程

计算机组成原理、微机原理、操作系统

关于教材：《现代微处理器及总线技术》

第一章 概述

本章主要内容:

一、计算机的发展过程

二、计算机发展(硬件)发展的关键技术

三、处理器领域研究热点

四、计算机领域研究问题归纳

一、计算机的发展过程

(一) 计算机的发展阶段与类型

70年代前　70’年代　80’年代　90’年以后
大中型机　小型机　微型机　网络系统

计算机类型(面向应用):
嵌入式计算机　桌面计算机　服务器

网格/云计算等是以网络为基础的应用技术研究

(二) 计算机性能指标

(1) 字长

4位→8位→16位→32位→64位

第1台64位微处理器： DEC的Alpha21064

– 64位处理器: 64位内外部总线、64位寄存器
– 64位计算机系统
  - 内部总线和寄存器为64位
  - 外部总线64位
  - 配置64位操作系统
  - 64位数据一次性处理

字长对精度和速度的影响

(2) 速度

主频、处理器的结构、指令运行模式、Cache的容量、内存指标等诸多因素, 都会在不同程度上上影响计算机的速度。— 用什么来衡量速度?

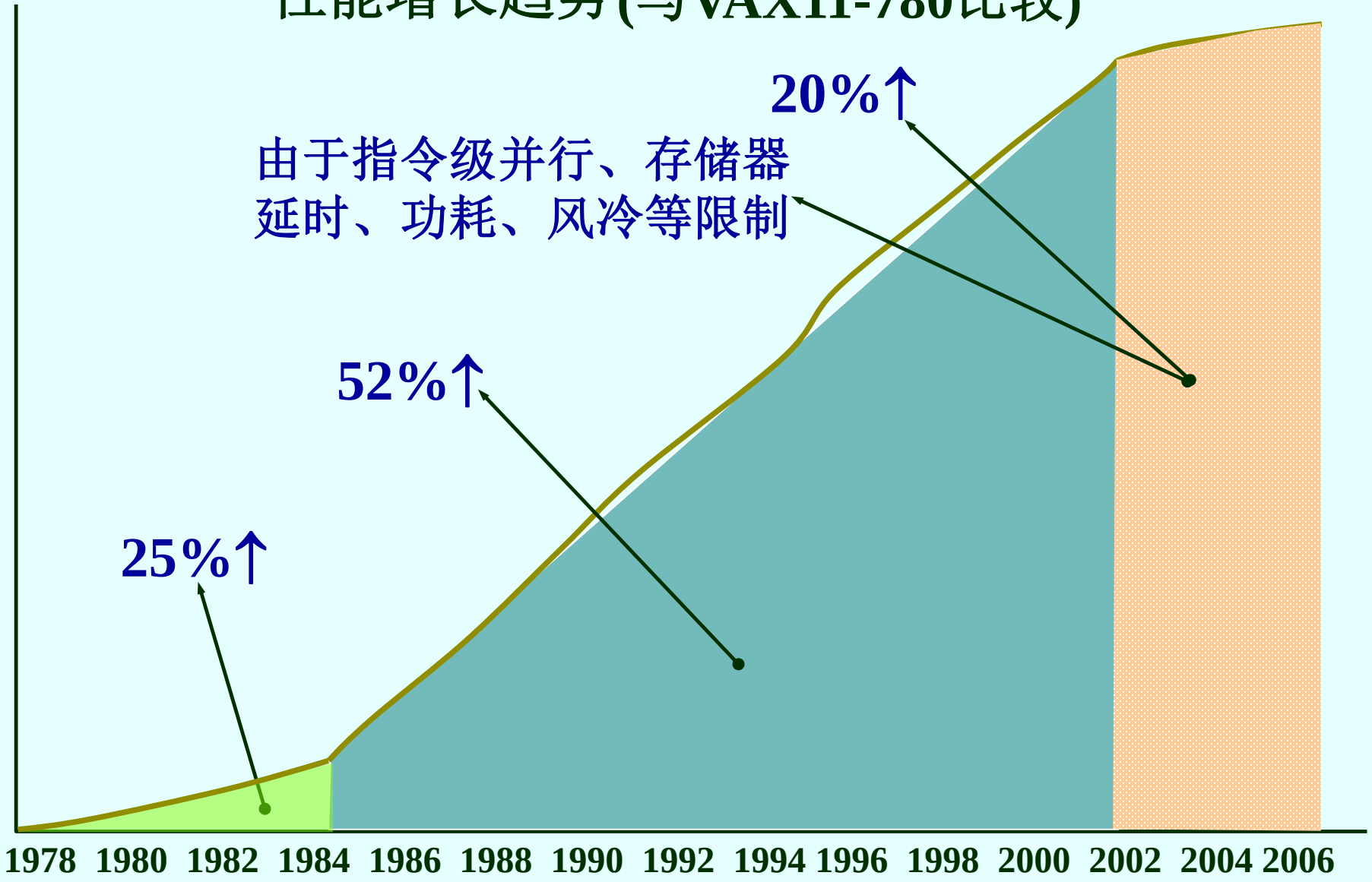
- 主频 → 决定主频的因素?
- MIPS: 百万条指令/每秒
- 基准测试程序, 比如SPEC

(3) 容量

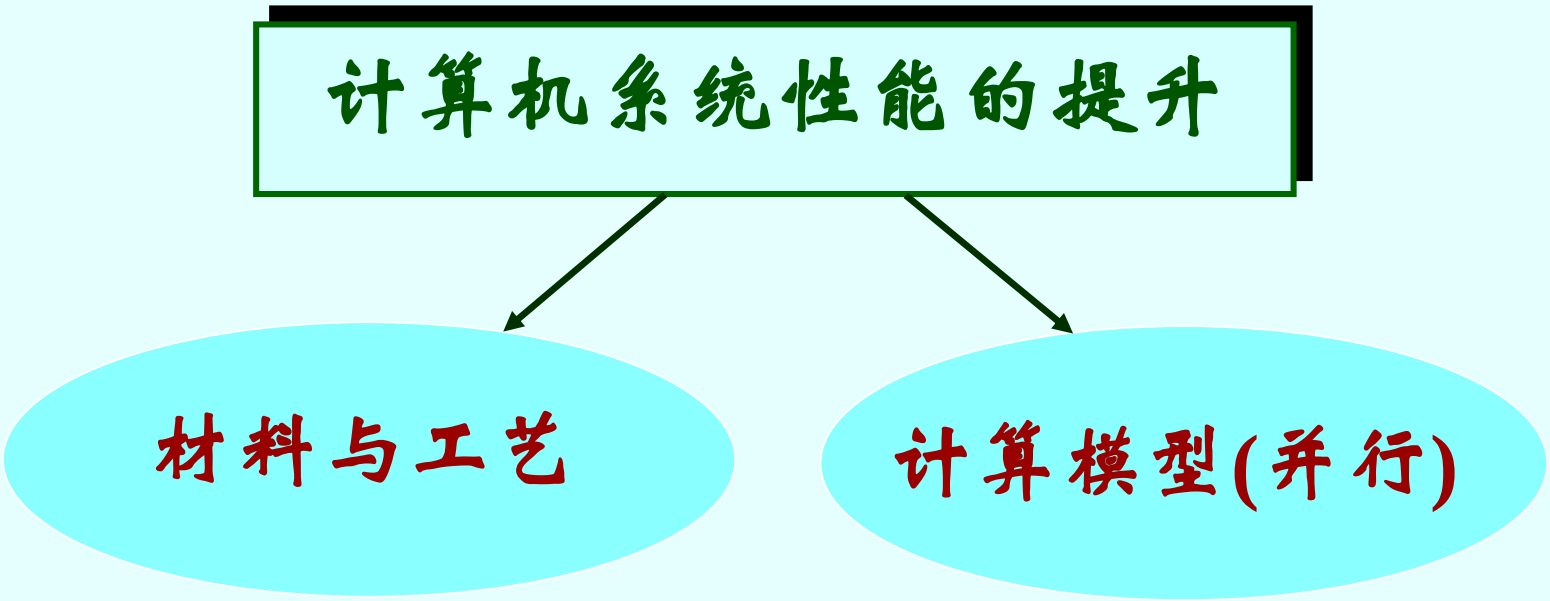
- 内存容量: 最大可达2^n (n为地址线的条数)
- 外存容量
- 高速缓存(Cache)容量: 不计入存储容量

性能增长趋势 ⟶

性能增长趋势(与VAX11-780比较)



二、计算机系统(硬件) 发展的关键技术



材料技术

集成技术与工艺

系统结构

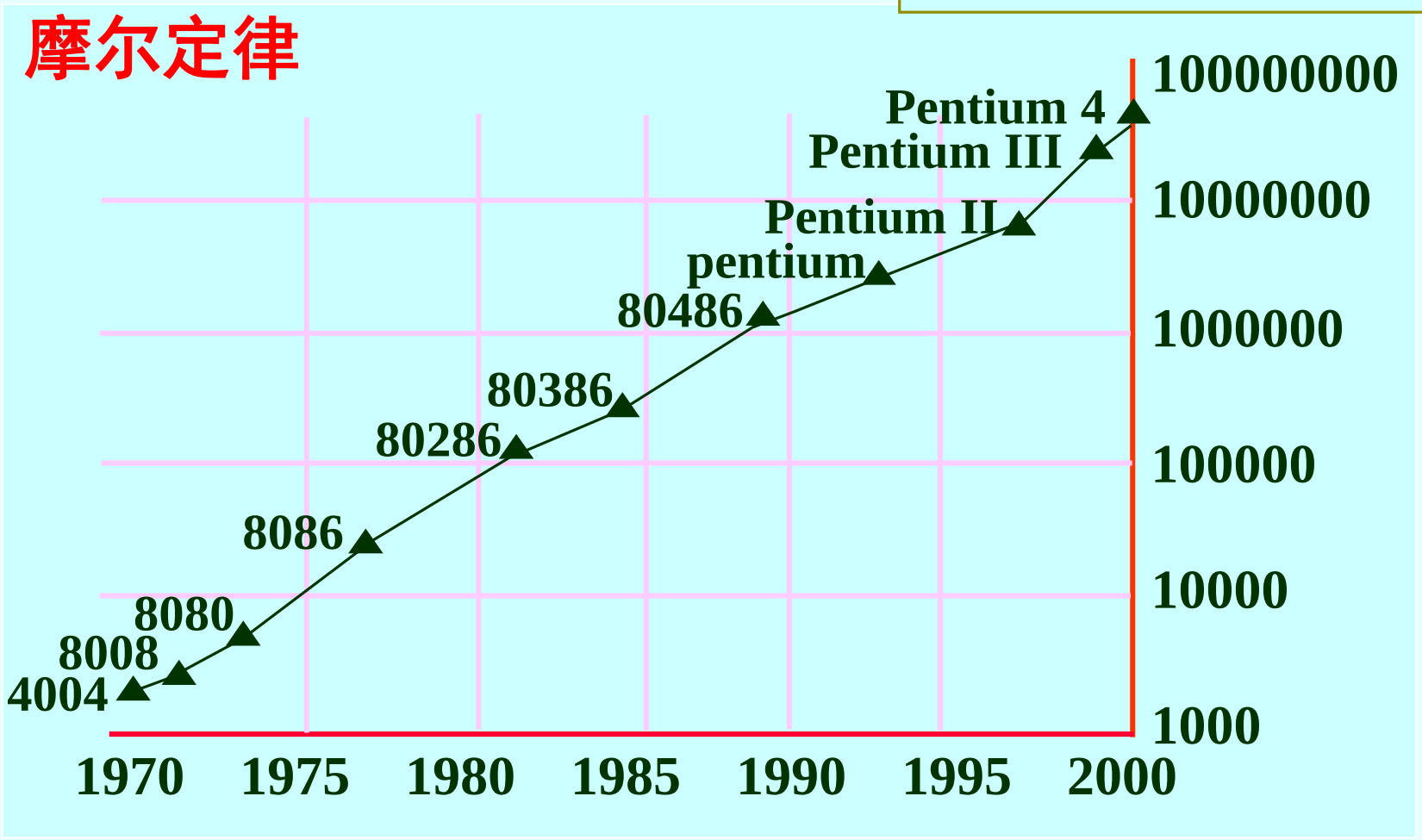
指令执行模式

数据流模式

(1) 从集成技术的角度

摩尔定律, 但在半导体基片上光刻电子元件的方式会遭遇极限(工艺和热量)。

第四代酷睿i7 约57亿



- CMOS(互补金属氧化物半导体电路)工艺

  直接在半导体基片上制作各种晶体管电路

- SOI制作工艺 (Silicon On Insulator)

  减少了晶体管的静电电容, 缩短了充放电时间, 提高了晶体管的切换速度; 降低了功耗。

  相比CMOS工艺, SOI可使速度提高35%~40%, 功耗降低50%~60%, 因此可以说SOI及相关技术可使摩尔定律继续起作用。

- 铜芯片

  铜导线来代替铝用于晶体管之间的互联, 相同条件下减少40%的功耗; 也可将铜导线与SOI相结合

- 应变硅技术

  加大硅原子之间的间距，减小电子通行阻力(减小了电阻)，从而降低了功耗，使速度得以提升。

- 多栅极晶体管技术

  将每个晶体管的栅数增加至两个或三个，使计算能力得以提高，并能降低功耗，减少电流间的相互干扰。

- 3D芯片技术

  在芯片设计中，将晶体管封装成两层或三层以上。形成立体封装，在基本同样大小的芯片里，晶体管数量可成倍增长，还可缩短晶体管之间金属连接导线的长度，有助于增强芯片性能。

- 记忆电阻器(忆阻—研发中, 第4种基本元件)

  一种有记忆功能的非线性电阻。通过控制电流的变化可改变其阻值,如果把高阻值定义为“1”,低阻值定义为“0”,从而可实现数据表示和存储。忆阻可使计算机可以反复开关,不必经过“导入”过程就能即刻回复到最近的结束状态。

但当前以及未来一段时间，处理器的发展方向不是增加单位面积的晶体管数量和提高主频，而是多核技术。首先，摩尔定律趋势已经变缓，由原来的1.5年一代变为2-3年一代。其次，过去每代微处理器主频是上代产品的两倍中，只有1.4倍来源于器件的按比例缩小，另外1.4倍来源于结构优化。

芯片设计越来越强调结构的层次化、功能部件的模块化和分布化，即每个功能部件相对简单，部件内部尽可能保持通信的局部性。

(2) 从体系结构的角度

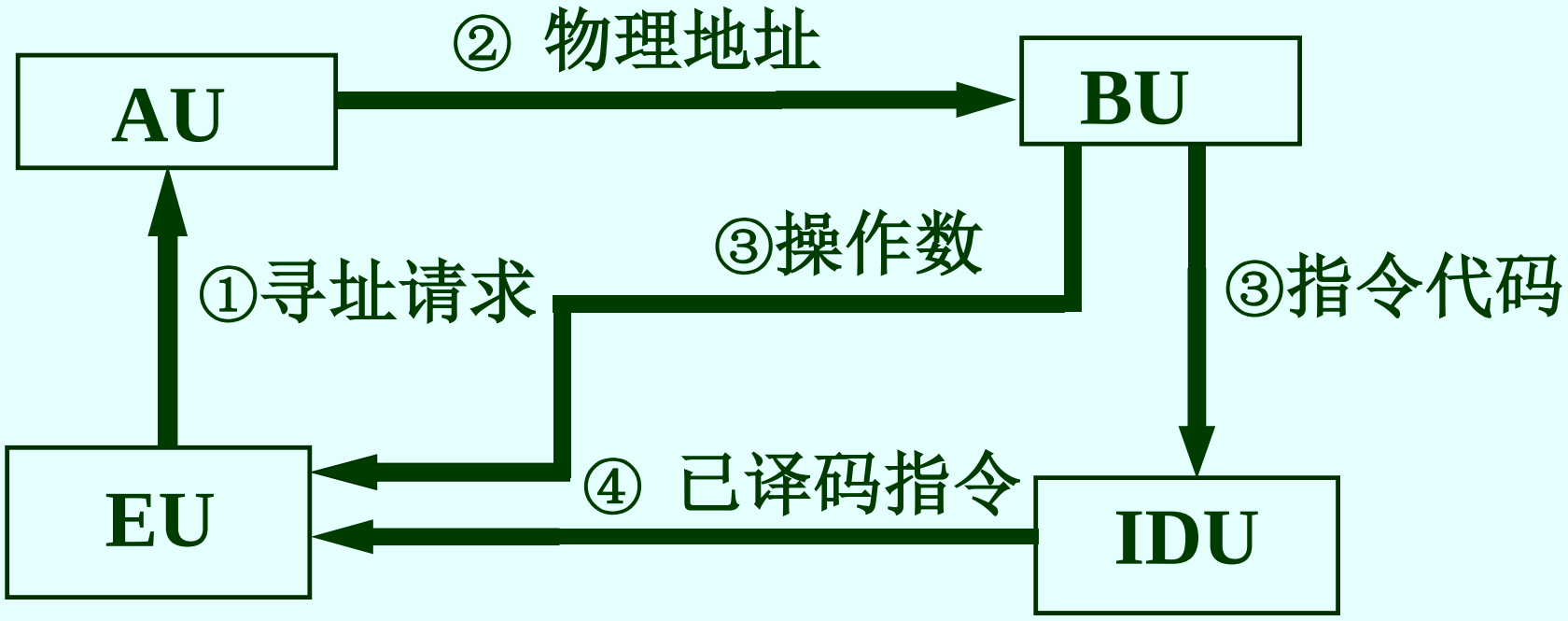
① 处理器体系结构

- 从标量结构演变到超级标量结构
- 从单数据流演变到多数据流
- 处理器内单一总线结构演变为多总线结构
- 单指令发射到多指令发射
- 超长指令字VLIW(Very Long Instruction Word)

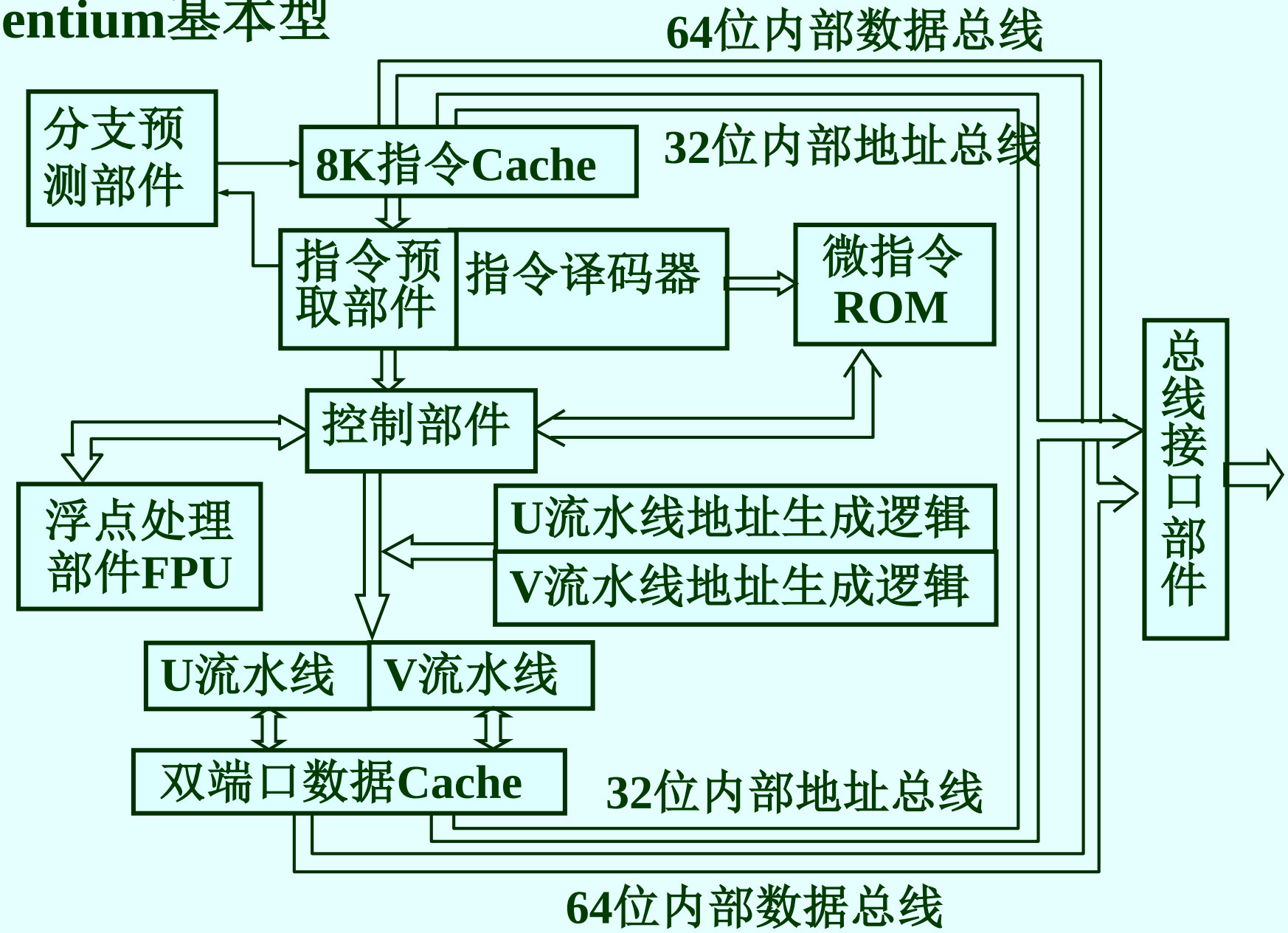
  即把多条指令组合在一起，以加快指令处理速度。比如：编译把“R1+R2→R3”和“R4+R5→R6”两条指令组合成一条指令(两条指令无寄存器相关)。
- 单核到多核技术

如：比较Intel 80286与Pentium的内部结构

— Intel 80286



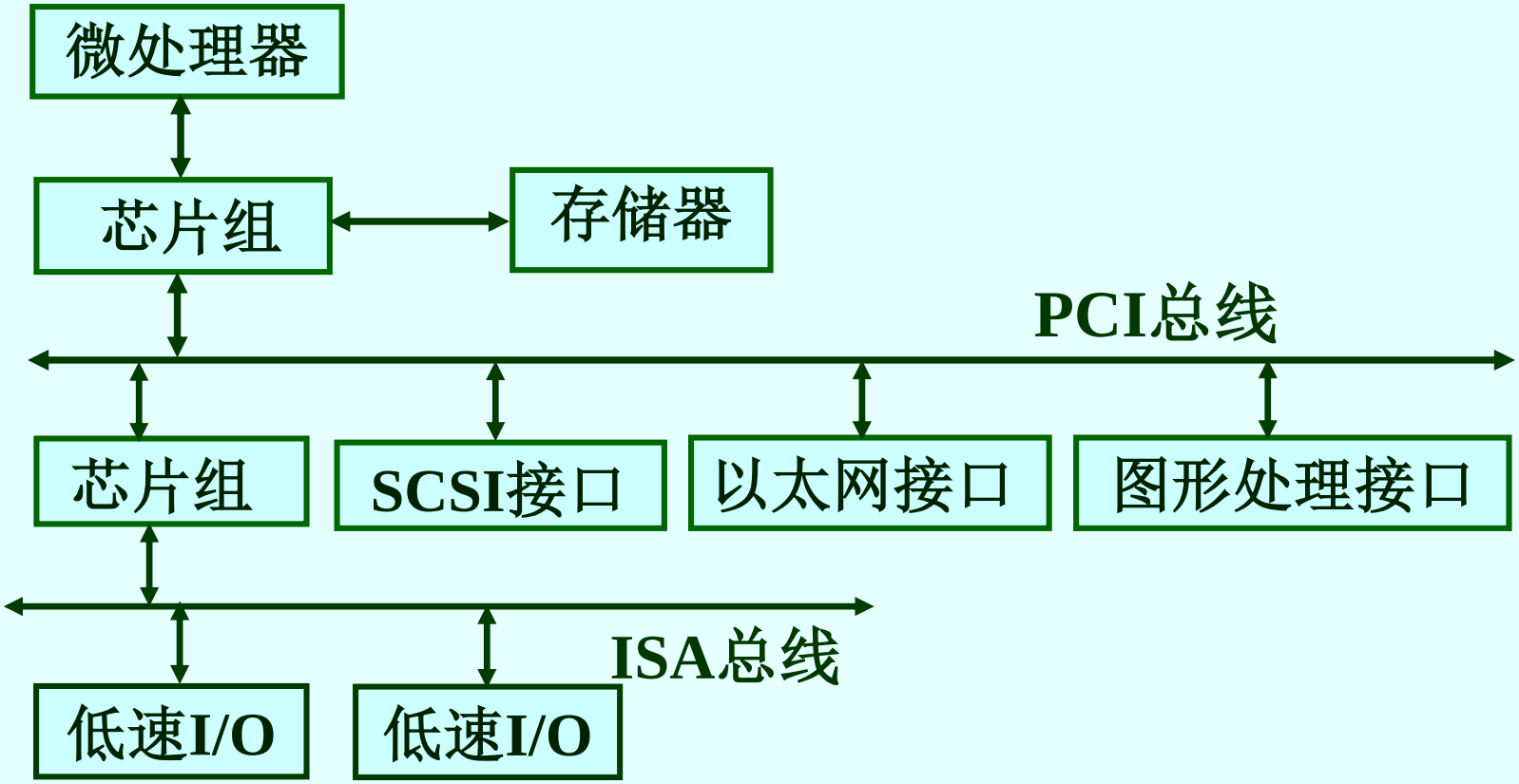
Pentium基本型


64位内部数据总线
分支预测部件
8K指令Cache
32位内部地址总线
指令预取部件
指令译码器
微指令ROM
总线接口部件
控制部件
浮点处理部件FPU
U流水线地址生成逻辑
V流水线地址生成逻辑
U流水线
V流水线
双端口数据Cache
32位内部地址总线
64位内部数据总线

② 系统体系结构

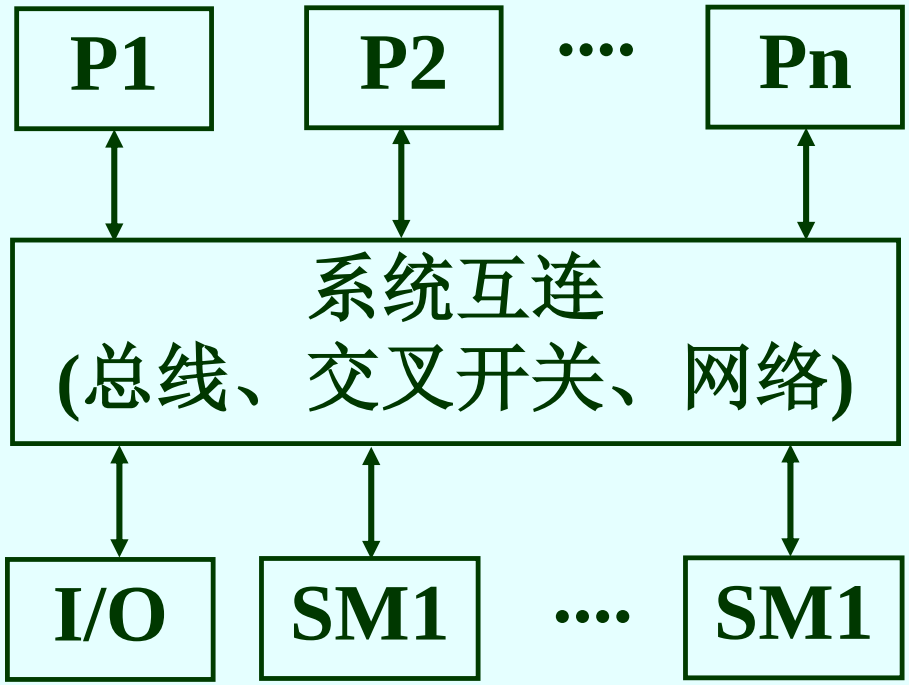
从单一总线结构 → 多总线结构 → 多处理器系统

多总线结构: 比如在一个系统中, PCI总线、ISA总线、EISA总线等并存。



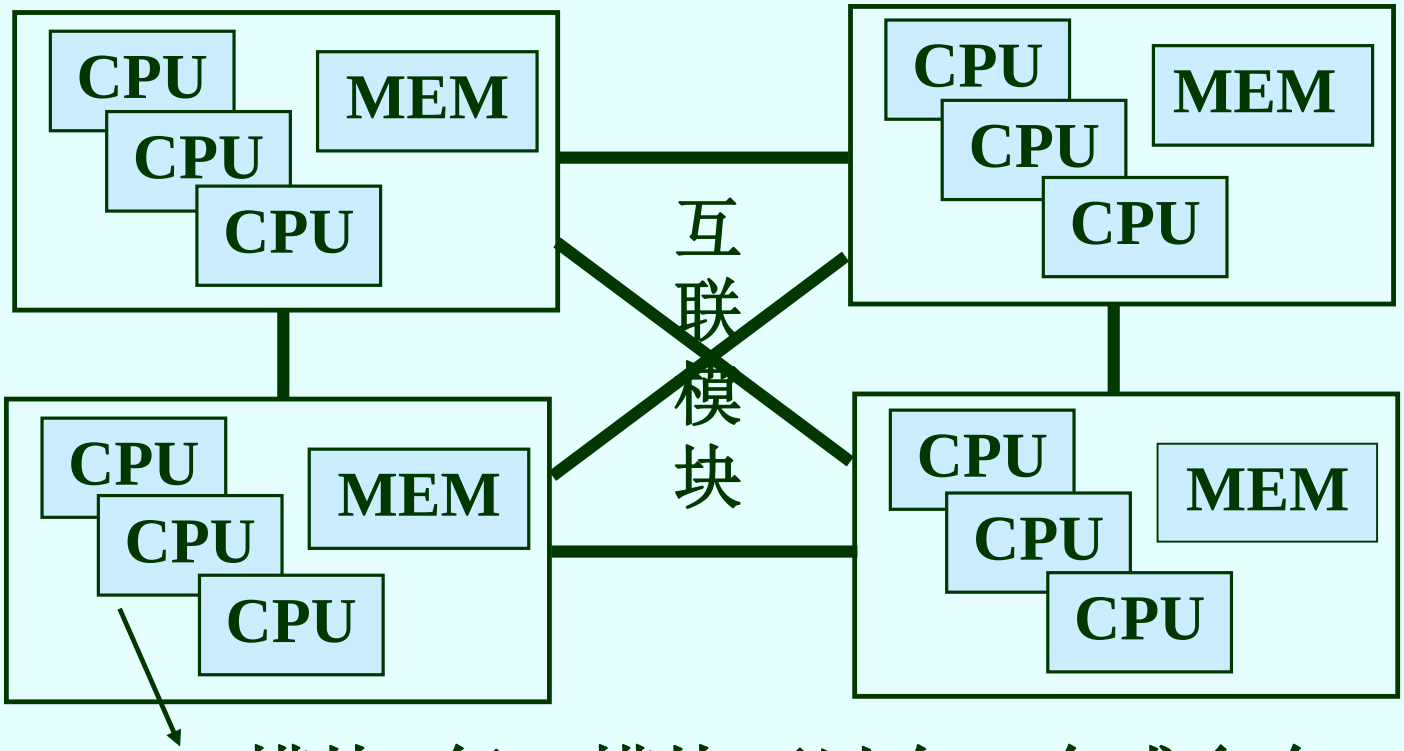
③ 多处理器系统

如 ➢ SMP技术: **Shared Memory multiProcessor**
Symmetry MultiProcessor



(Shared Memory)

- NUMA技术(非一致性访问分布共享存储):



CPU模块: 每一模块可以有一个或多个CPU, 具有独立的本地内存、I/O口等

每个CPU可以访问整个系统的内存(即共享内存, 这是NUMA系统与MPP系统的重要差别)

- **MPP(massively parallel processing)**

  不同CPU拥有自己独立的内存、硬盘等。通过专用操作系统和应用软件可以把一项庞大的任务拆分成多个子任务到不同的CPU。

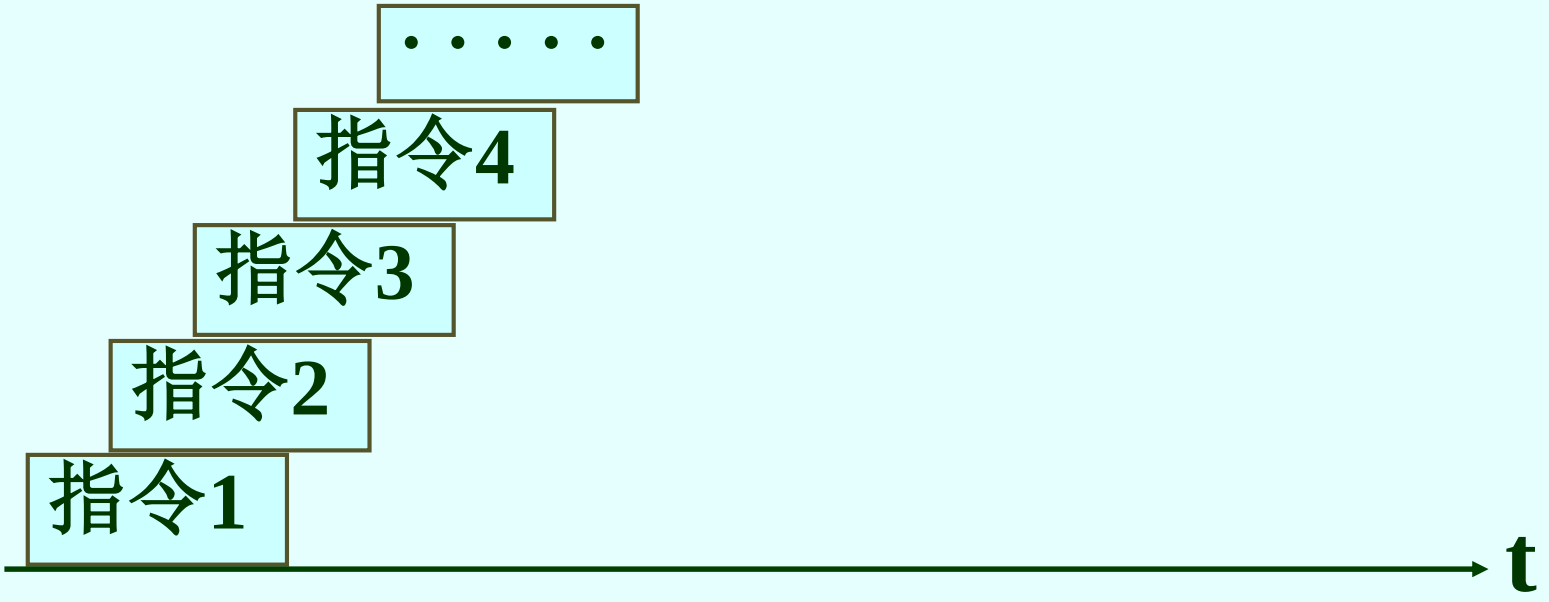
  MPP不需要共享内存、共享硬盘和其它的I/O设备(无需解决抢占内存和内存同步等问题。因此随CPU数量的增长, 系统的性能明显优于SMP。

(3) 指令执行模式

指令执行顺序的演变:

① 串行计算方式:

② 指令流水线:



③ EPIC(Explicity Parallel Instruction Computing)模式

EPIC体系结构的基本特点来自VLIW技术: 由编译器来决定指令执行方案。

在传统体系结构中, 条件分支往往是限制VLIW处理器性能发挥的瓶颈。

EPIC将分支指令拆分成三部分:

计算分支条件

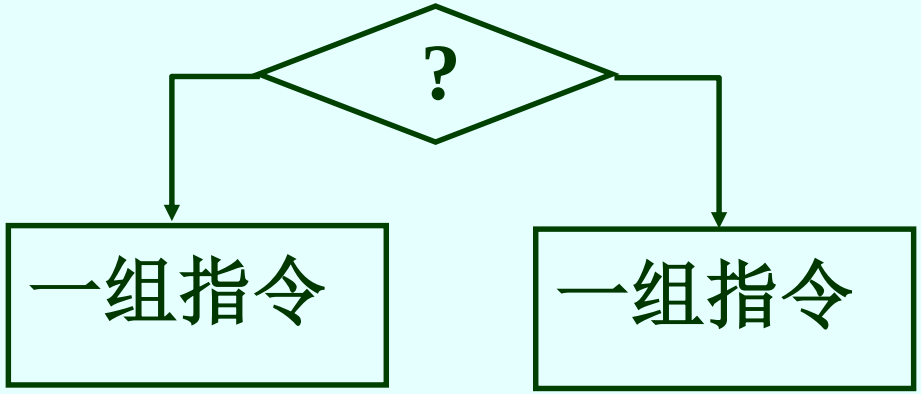
形成分支地址

从分支成功处和分支失败处取指令译码

各个部件可以重叠执行。

理论上讲, 让一组指令完全并行执行(EPIC的核心思想是“并行处理”), 比如:

对分支指令:
两个分支方向并行执行!



④ 进程级并行: 如Pentium4的超线程技术
多个线程并行执行

(4) 数据并行

典型情况如SIMD、MIMD

Load A(1)	Load A(2)	Load A(n)	
Load B(1)	Load B(2)	Load B(n)	
C(1)=A(1)×B(1)	C(1)=A(2)×B(2)	C(1)=A(n)×B(n)	SIMD 结构
Store C(1)	Store C(2)	Store C(n)	
next Instruct	next Instruct	next Instruct	

Load A(1)	call func D	do 10 i=1,n	
Load B(1)	X=Y×Z	alpha=w^3	
C(1)=A(1)×B(1)	SUM=X^2	zet=C(i)	MIMD 结构
Store C(1)	call sub(i, j)	10 continue	
next Instruct	next Instruct	next Instruct	

小结：近20年处理器技术发展主流及问题 —

- **90年代：增强指令并行性**
  - 指令级并行性存在很大限制，超标量走到尽头；
  - VLIW(超长指令字)兼容性问题，对编译要求高；
- **90年代末期：提高主频**
  - 流水线的细化和指令相关导致复杂性大大提高；
  - 存储器性能影响整体性能；
  - 功耗问题严重；
- **本世纪：多线程**
  - 增加了系统吞吐率(约15%)，但并未提高单个单线程的执行速度；
- **当前和可预见未来：多核**
  - 设置多个微处理器内核；
  - 线程级或进程级并行性来提高性能

三、处理器领域研究热点

1. 处理器

(1) 64位处理器

如酷睿系列处理器、AMD X86-64处理器、IBM的Power系列处理器等。

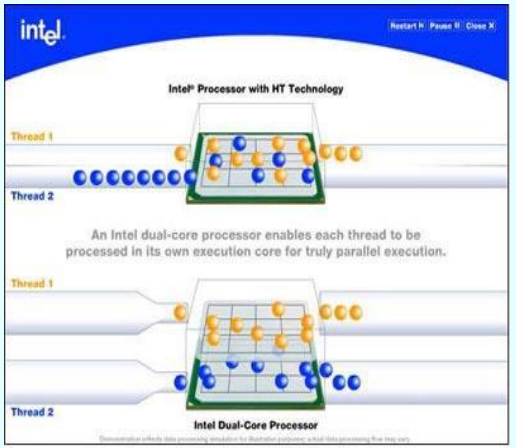
(2) 全64位计算平台

以64位计算模式为基础, 加上相应系统软件支持

应用层	• 64位编程模式和64位API • 64位的编译器
操作系统层	• 支持大于4GB的物理存储空间 • 支持大于4GB规模的文件 • 支持多个物理设备文件(如多个磁盘系统)
机器层	• 64位宽度以上的总线 • 64位Cache、高速图形板等
CPU芯片层	• 64位整数寄存器 • 64位浮点寄存器 • 64位宽度以上的数据总线

(3) 多核处理器

多核处理器(CMP)是传统多处理器系统的进一步发展, 也是集成电路技术发展的结果。



HT 技术与双核心技术比较

多核可以分为同构多核和异构多核两种:

同构多核 —

计算内核相同, 地位对等的称为“同构多核”。同构CMP大多数由通用的处理器组成, 多个处理器执行相同或者相似的任务。

异构多核—

计算内核不同, 地位不对等的称为“异构多核”, 异构多核大多采用“主处理核＋协处理核”的设计.

如：IBM公司的CELL处理器

AMD公司的Fusion方案 “CPU+GPU”

CPU和GPU功能融合, 形成APU

(Accelerated Processing Unit)

实现统一的CPU/GPU寻址空间、GPU使用可分页系统内存、GPU硬件可调度、GPU也支持高级编程语言(关键)。

但也存在以下一些观点：

多核处理器的最大问题是可编程性。尽管在并行计算上已探索了超过40年，但编写、调试、优化并行处理程序的能力还非常弱。

出于技术的挑战，多核强加给了产业，而产业界并没准备好。或许十年后，多核就到头了。.....并行计算的发展历史表明，并行性超过一定程度，程序就很难写。

即使能够不断增加同类型的CPU内核以加强并行处理能力，但整个系统的处理性能仍然会受到软件中必须串行执行的那部分的制约。

虽然Intel在很多年前已研发并展示了80核处理器原型，但目前还没有能够利用这一处理器的操作系统。

2. 操作系统

包括64位操作系统、支持多核处理器的操作系统、云计算操作系统、专用(领域)操作系统等。

3. 编程模式

包括支持64位处理器的编程模式、以及支持多核技术的编程模式等。

四、计算机领域研究问题归纳

1. 技术类

如:

- 集成电路技术, CPU和存储器设计制造技术等;
- 网络安全、网络管理技术等、中间件技术等;
- 多媒体技术信息处理和传输、虚拟现实技术;
- 操作系统技术、云计算模型与实现等。

2. 产品类

如:

- 硬件产品中的各种处理器、交换机、路由器
- 软件产品中的各类操作系统、WWW浏览器、开发工具、编程语言、通信软件等

3. 应用类

涉及计算机领域中的各种应用。

两大基础性研究领域

(1) 人工智能 (智能化计算机)

人工智能

引发的联想

人工智能是一门涉及数学、计算机技术、自动控制理论与工程、机电工程、生物学、心理学、语言学等多学科的交叉科学。

目前两种相反的观点：

1. 人工智能技术蓬勃发展

许多科学家断言，机器的智慧会迅速超过如爱因斯坦和霍金等的智慧之和....。最迟到本世纪中叶，计算机的智能就会超出人类的智能。

2. 人工智能为什么发展缓慢？

用数学的方式去瞬间(即非进化方式)搭建一个像人类那样高级的智能，这可能是太复杂了，也许根本就超出了人类的理解能力和能够把握的思维广度，所以我们迟迟没有大的进展也就不难理解了。

于是，越来越寄希望于全新的计算机技术：光子计算机、生物计算机和量子计算机。

发展状况:

计算机智能在如棋类、智力猜测、路径规划、信息检索等领域已经表现出超越人类的能力。由此可解释以下现象:按应用领域完成某种特定功能的机器人不断产生:

如:清扫机器人
助残机器人
手术机器人
战场机器人
娱乐机器人
……

原因 — 得益于信息存储、数值计算、统计推理、决策模型等的研究进展。

无突破领域——

模拟人类思维过程和智能行为(具有标志意义的智能属性)

由此可解释以下现象:在面对通用或复杂环境时,计算机所具有的智能属性与人类和科学目标相距甚远。因此在人类智能活动中,还无法改变机器人只能是作为辅助工具的角色定位。

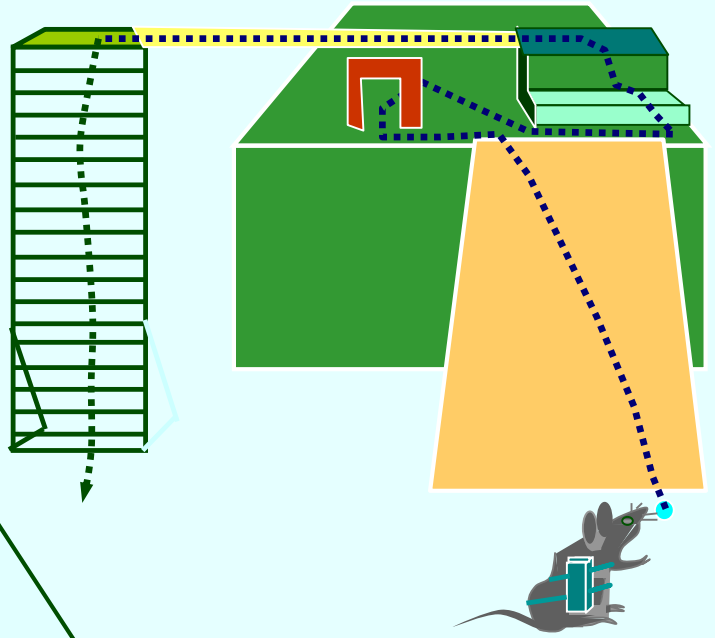
从“智能(智慧)”的角度,在以下领域,还有很多问题有待我们来研究,如:

- 知识表示 → 把信息表示为计算机可以理解的知识
- 知识获取 → 如何获得知识
- 自主学习 → 通过自身的学习不断提高其能力
- 自动推理 → 如树上有10只鸟,....
- 自然语言理解 → 你要是再不听话,会有你好看的...,或不会有你好看的

经过三年的训练,美国研制人工智能系统依然通不过八年级(相当于中国的初二年级)的科学课考试,最好成绩答对了60%的问题(得分59分)。(2016.2.19报道)

脑机混合智能：获取大脑皮层等活动信号，并实现与外部电子系统的交互，使生物脑与机器脑的互连形成混合智能

如已经研制成功的鼠机器人：外部机器向老鼠脑部输入控制指令，诱导其产生左、右、前进、停止等四种基本行为，完成爬梯、穿越障碍等导航任务，计算机可完全替代操作员



(2) 高性能 非“冯•诺依曼”体系结构 并行处理、神经网络计算机等试图突破。

“计算机技术的真正突破还有赖于物理学的突破”

量子计算机、生物计算机、基于超导技术的计算机、纳米计算机等。