操作系统

数学与信息学院: 张 猜

联系方式: zhangcai@scau. edu. cn

教材与参考资料

- 1、教材
- William Stallings著,陈向群 陈渝译,操作系统——精髓与设计原理(第八版),电子工业出版社,2012。
 2012年9月第1次印刷
- 2、参考资料
- (1)《操作系统教程 第4版》孙钟秀,费翔林,骆斌著, 高等教育出版社。2008。
- (2) 《操作系统概念 第6版》Abraham Silberschatz等著 , 郑扣根译, 高等教育出版社。2003。
- (3)《现代操作系统》(第二版)Andrew S. Tanenbaum等著,陈向群译,机械工业出版社。2005。
- (4)《操作系统教程》陈向群,杨芙清等著,北京大学出版社。2001。

网上资源

- 慕课
- http://www.zhihuishu.com

课程目标

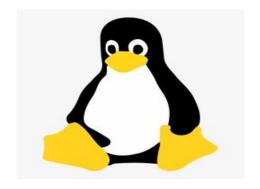
学习本课程后,学生应掌握计算机操作系统的有关基础 知识和基本原理,包括:

- ①操作系统特征与作用。
- ②进程的描述和控制,线程概念,进程调度算法,进程的同步与互斥,信号量,进程通信,死锁概念。
- ③分页和分段存储管理,虚拟存储器概念,请求分页和分段 存储管理。
- ④1/0控制,缓冲管理,设备分配,磁盘调度与分配。
- ⑤ 文件及目录的组织,文件保护和安全。
- ⑥ 分布式操作系统的进程管理、文件与资源管理、通信、协调等。

操作系统













操作系统











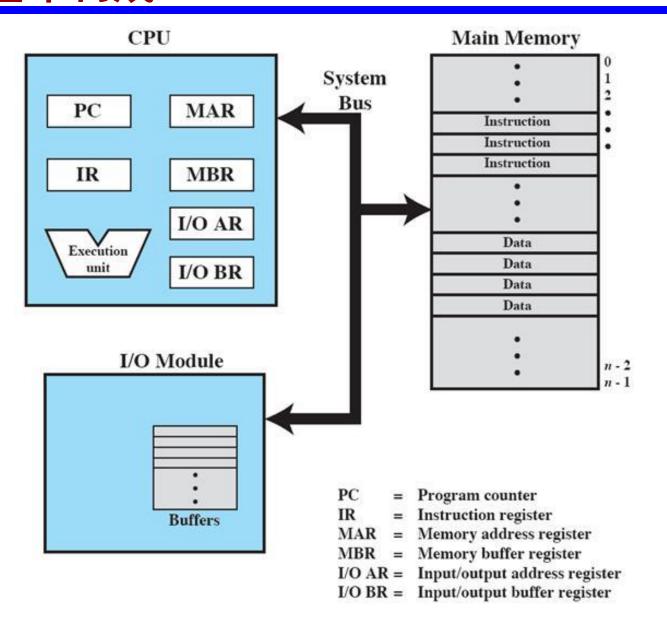




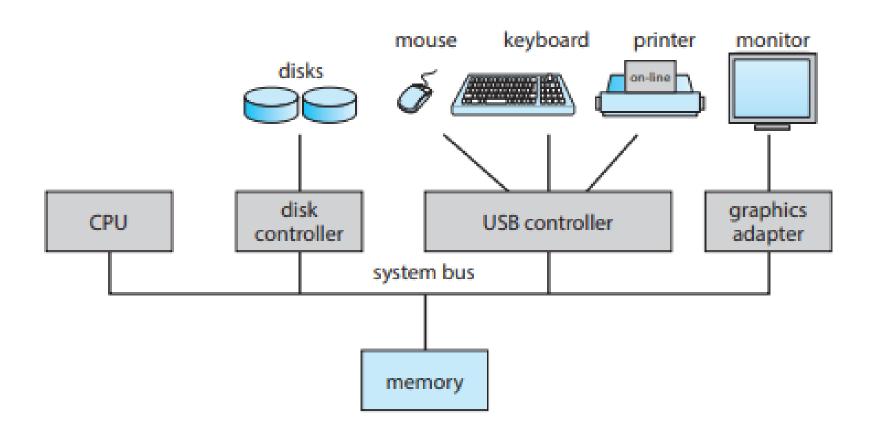
第1章 计算机系统概述

- 主要内容
 - 1.1 基本构成
 - 1.2 微处理器的发展演化
 - 1.3 指令的执行
 - 1.4 中断
 - 1.5 存储器的层次结构
 - 1.6 高速缓存
 - 1.7 直接内存存取
 - 1.8 多处理器和多核计算机组织结构

1.1 基本构成



1.1 基本构成



1.1 基本构成

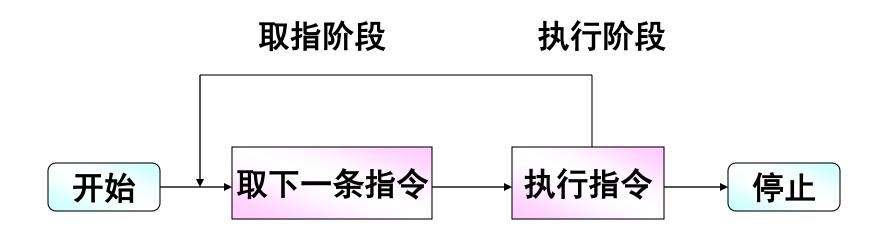
- 处理器:控制计算机操作,执行数据处理功能。 当只有一个处理器时,它通常指中央处理器(CPU)。
- 内存(实存储器或主存储器):存储数据和程序 。易失性。
- 输入/输出模块:在计算机和外部环境之间移动数据。
- 系统总线:为处理器、内存和输入/输出模块间 提供通信的设施。

1.2 微处理器的发展演化

- 微处理器
 - 一个芯片上容纳一个处理器内核
- 多核微处理器
 - 一个芯片上容纳多个处理器内核,共享执行单元☑双核、四核
 - 区每个核还可以配多个硬件线程——逻辑处理器
- 片上系统
 - CPU、高速缓存以及系统中的多数硬件都在同一芯片上
 - 满足便携式设备的需求



1.3 指令的执行



基本指令周期



0011011 1011011 0011001 0001011

1.3 指令的执行

指令操作大体上可以分为4类:

- 处理器-存储器
- 处理器-I/O
- 数据处理
- 控制

1.3 指令的执行



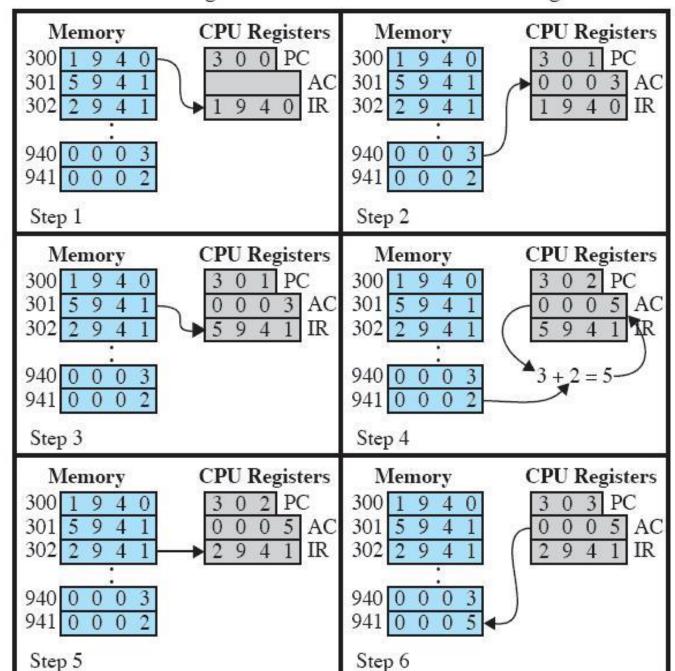
程序计数器(PC)=指令地址 指令寄存器(IR)=正在执行的指令 累加器(AC) =临时存储器 (c)内部CPU寄存器

0001=从存储器加载AC 0010=把AC的内容存储到存储器中 0101=从存储中加到AC中 (d)部分操作码列表

一台理想机器的特征

Fetch Stage

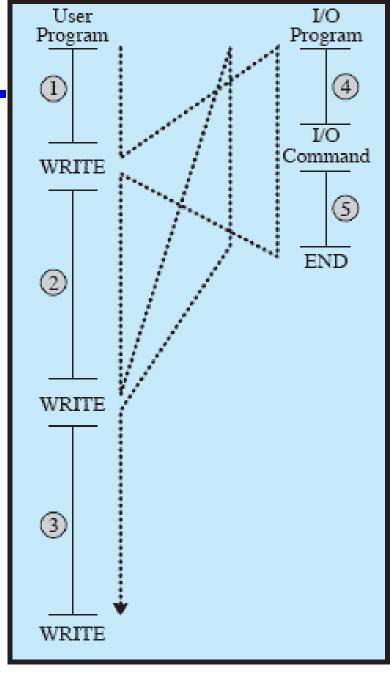
Execute Stage



1.4 中断

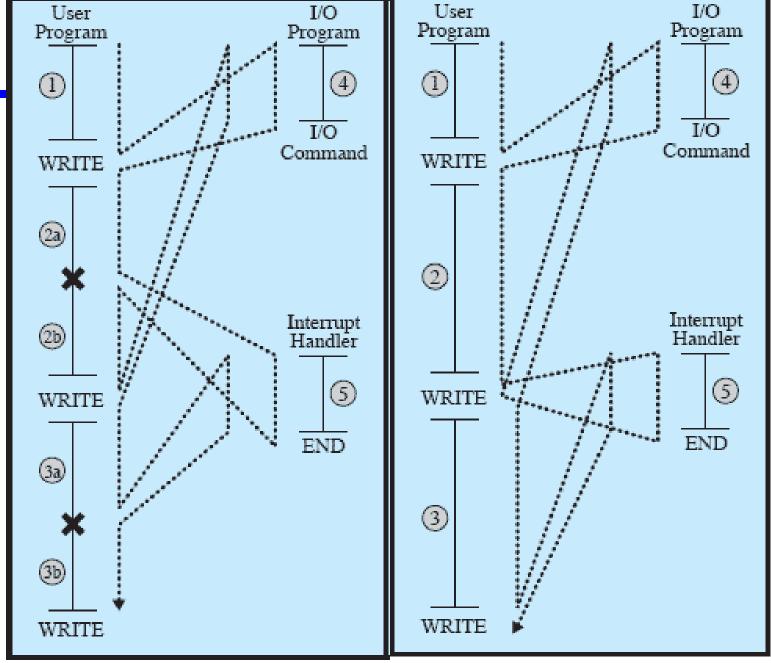
- 中断是一种机制,即允许其它模块(I/0、存储器)在处理器正常处理过程中打断其工作。
- 中断分类:
 - 程序中断
 - 时钟中断
 - I/0中断
 - 硬件失效中断
- 中断是提高处理器效率的一种手段。
 - 利用中断功能,处理器可以在 I / 0操作的执行过程中执行其他指令。
 - I/O操作和用户程序中指令的执行是并发的。

无中断



17

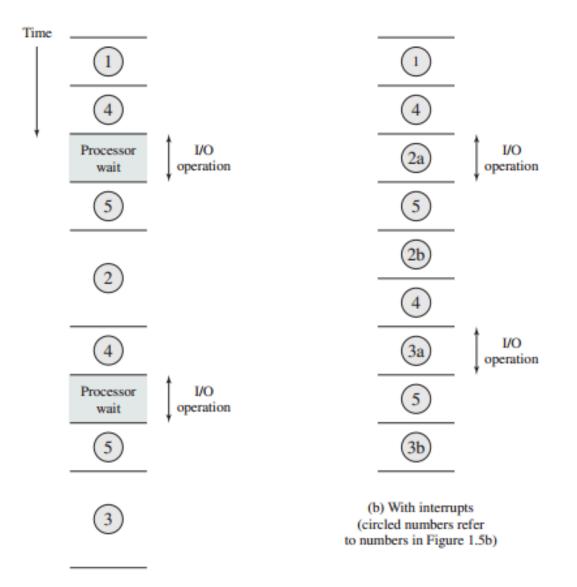
有中断



(b) Interrupts; short I/O wait

(e) Interrupts; long I/O wait

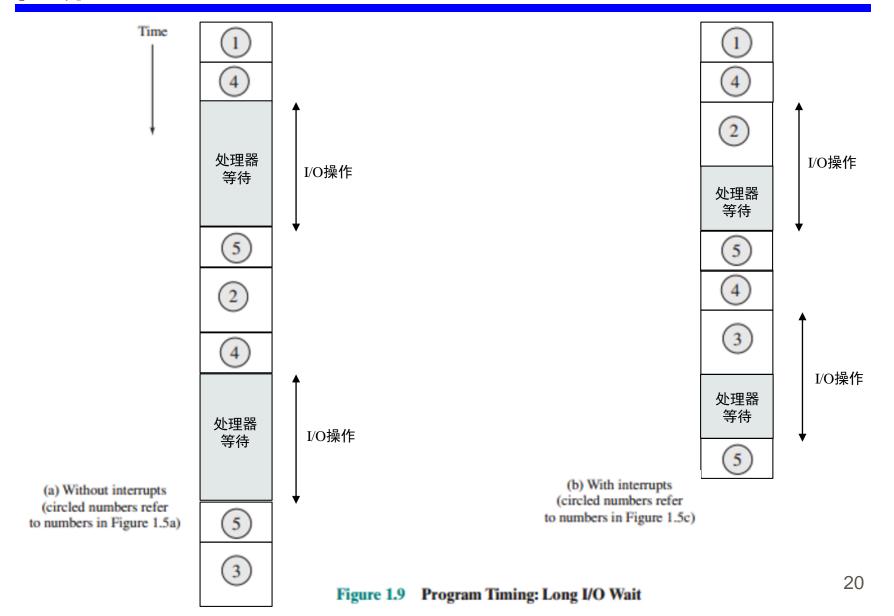
中断



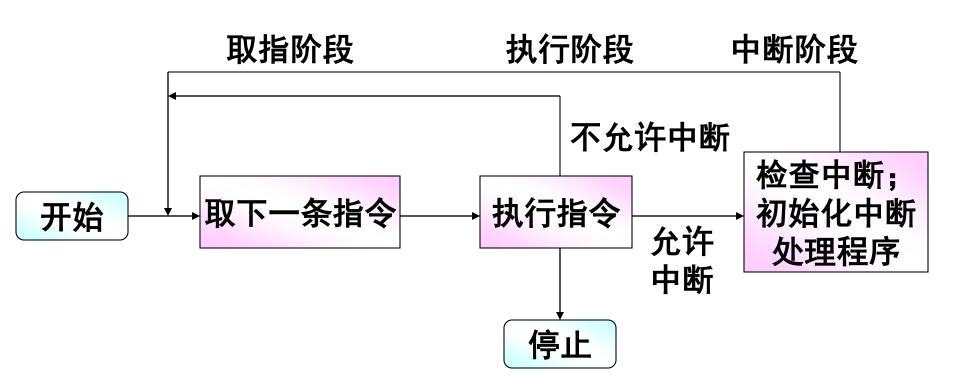
(a) Without interrupts (circled numbers refer to numbers in Figure 1.5a)

Figure 1.8 Program Timing: Short I/O Wait

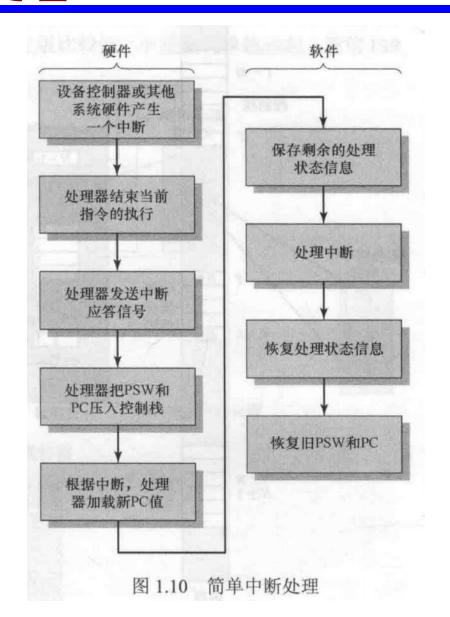
中断

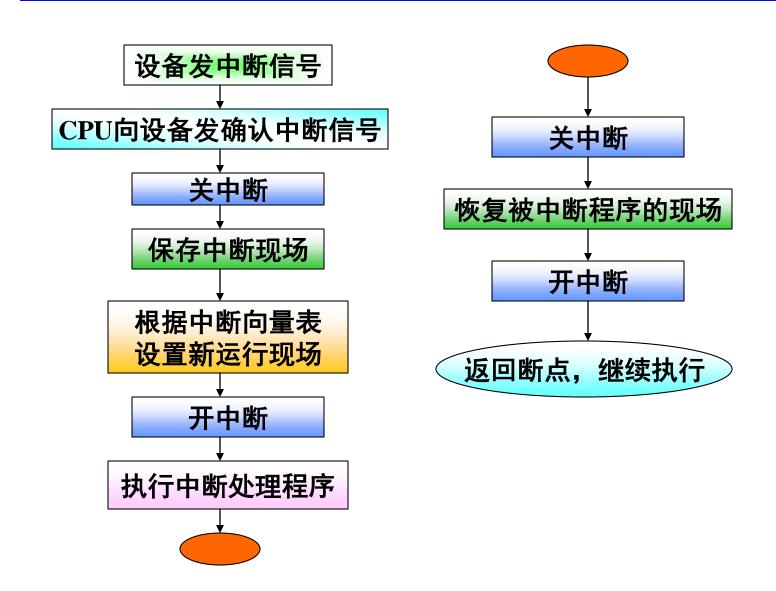


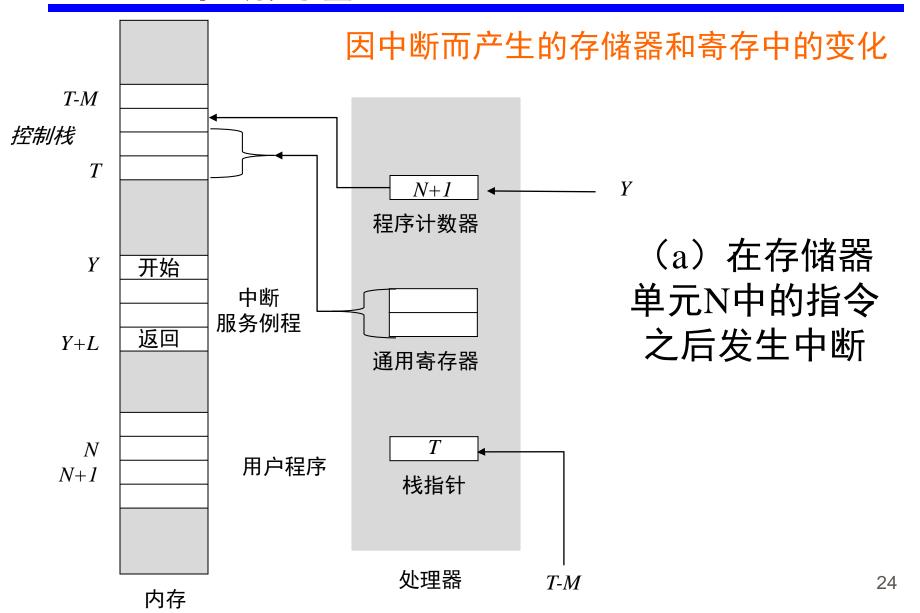
1.4.1 中断和指令周期

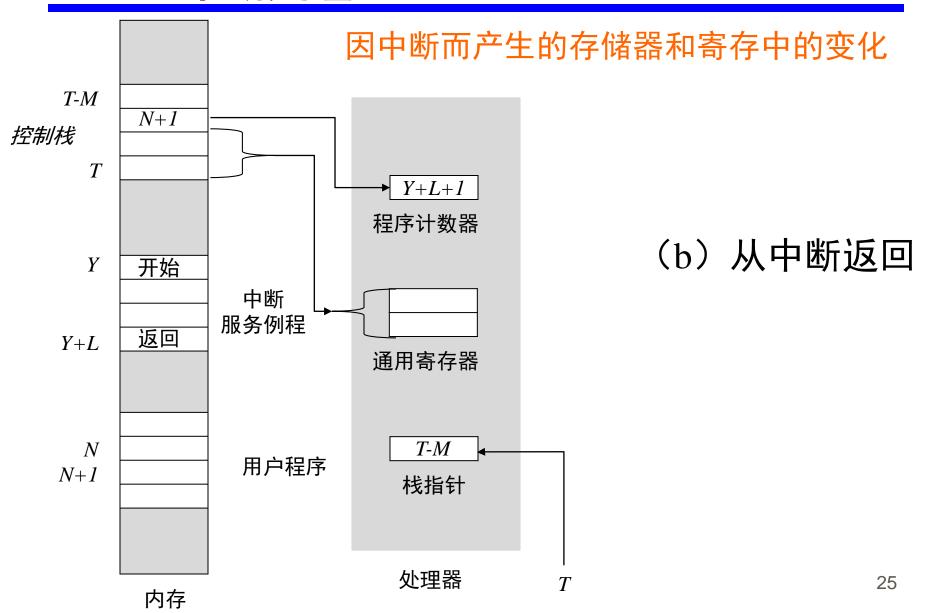


中断和指令周期









1.4.3 多个中断

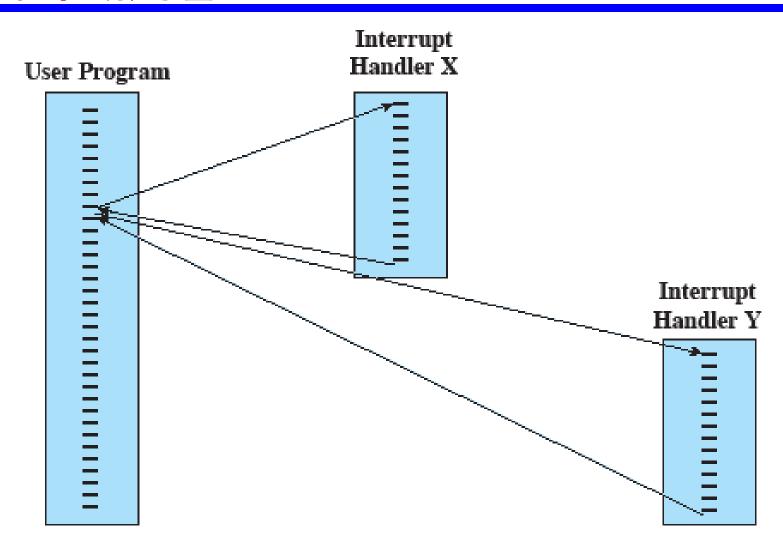
• 顺序中断处理

- 当正在处理一个中断时,禁止中断(对任何新的中断 请求信号不予理睬,处理完这个再处理下个)
- 缺点: 没有考虑相对优先级和时间限制的要求

• 嵌套中断处理

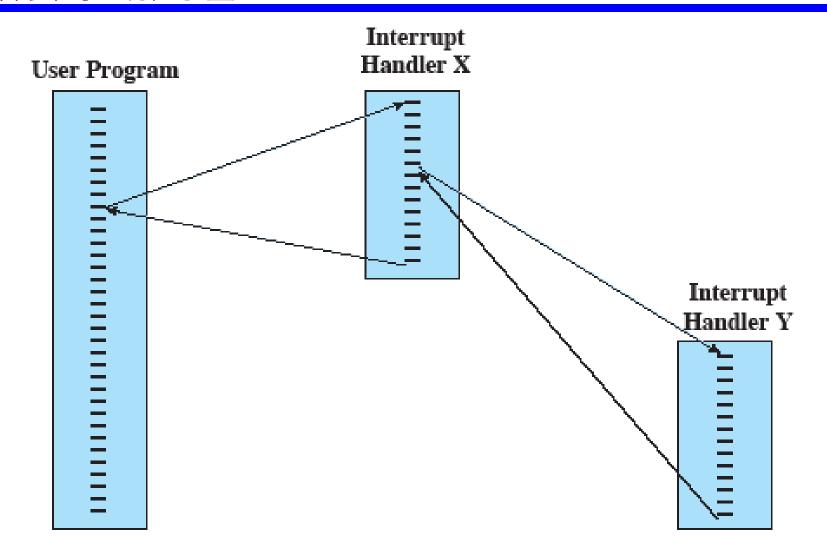
定义中断优先级,允许高优先级的中断打断低优先级的中断处理程序的运行。

顺序中断处理



(a) Sequential interrupt processing

嵌套中断处理



(b) Nested interrupt processing

中断优先级: 打印机(2) / 磁盘(4) / 通信线(5)。

多个中断处理例子

中断发起时间: 打印机(10)/磁盘(20)/通信线(15)。

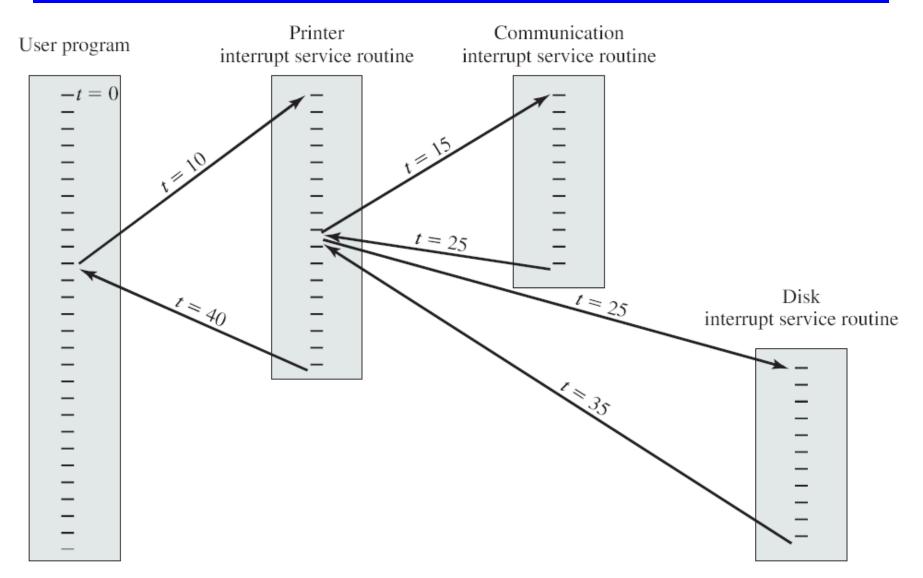
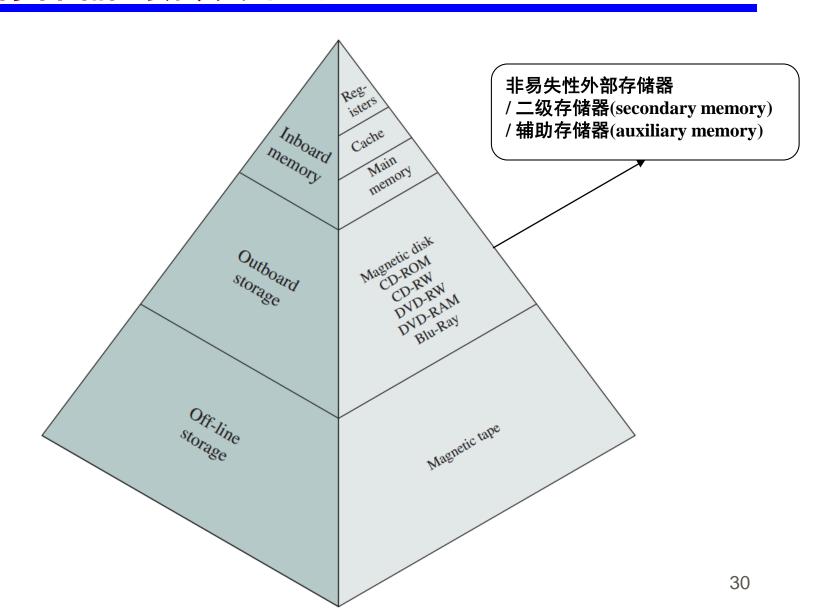


Figure 1.13 Example Time Sequence of Multiple Interrupts

1.5 存储器的层次结构



层次结构的特点

- 由上至下:
 - 每"位"的价格递减
 - 容量递增
 - 存取时间递增
 - 处理器访问存储器的频率递减
- 容量较大、价格较便宜的慢速存储器,是容量较小、价格较贵的快速存储器的后备。
- 存储器层次结构能够成功的关键: 低层访问频率 递减。

例

假定有一个二级存储器(内存+高速缓存),内存存取时间为1us,高速缓存存取时间为0.1us,且高速缓存的命中率为95%,则访问一个字节的平均存取时间为:

$$0.95\times0.1+0.05\times(0.1+1)=0.15$$
 (us)

访问的局部性原理

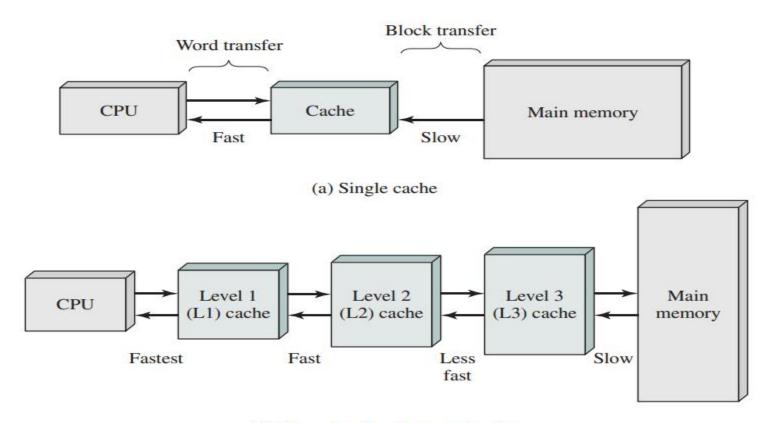
- 在执行程序期间,处理器的指令访存和数据访存 呈现"簇"状(一组数据集合),例如:
 - 循环: 重复访问一小范围的指令集合;
 - 数组操作: 存取一簇数据。
- 经过很长一段时间,程序访问的"簇"会改变, 但在较短时间内,处理器主要访问存储器中固定 的"簇"。
- 如二级存储结构,程序当前访问的"簇"暂时存 放在第一级存储器中,而第二级存储器包含所有 指令和数据。

1.6 高速缓存 1.6.1动机

- 指令执行期间,处理器需要多次访问内存;
- 处理器和内存的速度不匹配,处理器速度的提高 一直快于内存访问速度的提高——处理器执行指 令的速度受限;
- 利用局部性原理,在处理器和内存之间提供一个容量小而速度快的存储器——高速缓存。

1.6.2 高速缓存原理

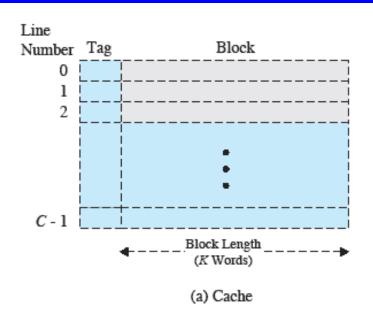
高速缓存试图使访问速度接近现有最快的存储器 ,同时保持价格便宜的大存储容量。



(b) Three-level cache organization

Figure 1.16 Cache and Main Memory

1.6.2 高速缓存原理



Memory address Block (K words) Block $2^{n} - 1$ Length (b) Main memory

As a simple example, suppose that we have a 6-bit address and a 2-bit tag. The tag 01 refers to the block of locations with the following addresses: 010000, 010001, 010010, 010011, 010100, 010111, 010110, 010111, 011000, 011001, 011010, 011111.

标签通常是地址 中较高的若干位

Figure 1.17 Cache/Main-Memory Structure

1. 6. 2 高速缓存原理

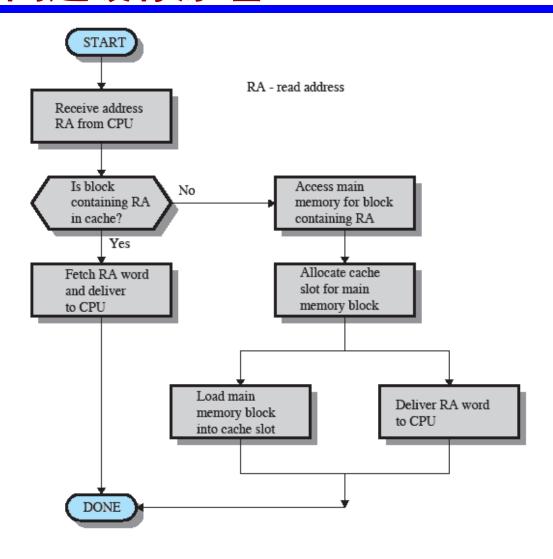


Figure 1.18 Cache Read Operation

1.6.3 高速缓存设计

- 高速缓存大小
- 块大小
- 映射函数
- 置换算法
- 写策略
- 高速缓存的级数

1.7 直接内存存取

- I/0操作三种可能的技术
 - 可编程I/O (程序控制I/O)
 - 中断驱动I/0
 - 直接内存存取(DMA)

可编程1/0

- I/0模块执行请求的动作并设置I/0状态寄存器中相应的位;
- 处理器执行I/0指令后,要定期检查I/0的状态, 以确定I/0操作是否已经完成。
- ▶问题:处理器在等待I/0操作完成期间需不断询问I/0模块的状态,严重降低了整个系统的性能

0

中断驱动1/0

- 处理器给I/0模块发送I/0命令,然后继续做其它 有用的工作;
- 当I/0模块准备好与处理器交换数据时,将打断 处理器的执行并请求服务;
- 处理器执行数据传送,然后恢复以前的执行过程。
- ▶问题:处理器仍需要主动干预在存储器和I/0模块之间的数据传送,任何数据传送都必须完全通过处理器。

直接内存存取(DMA)

- 当处理器要读或写一块数据时,给DMA模块产生 一条命令,发送以下信息:
 - 是否请求一次读或写;
 - 涉及的1/0设备的地址;
 - 开始读或写的存储器单元;
 - 需要读或写的字数。
- 处理器继续其它工作。DMA模块直接与存储器交 互,传送整个数据块,无须处理器参与;
- 传送完成,DMA模块发一个中断信号给处理器。
- ▶问题:在DMA传送过程中,当处理器需要访问总线时,因存在竞争,处理器将暂停一个总线周期,处理器的执行速度会变慢。

1.8 多处理器和多核计算机组织结构

- 并行处理——提高性能
 - 对称多处理器(SMP)
 - 多核计算机
 - 集群

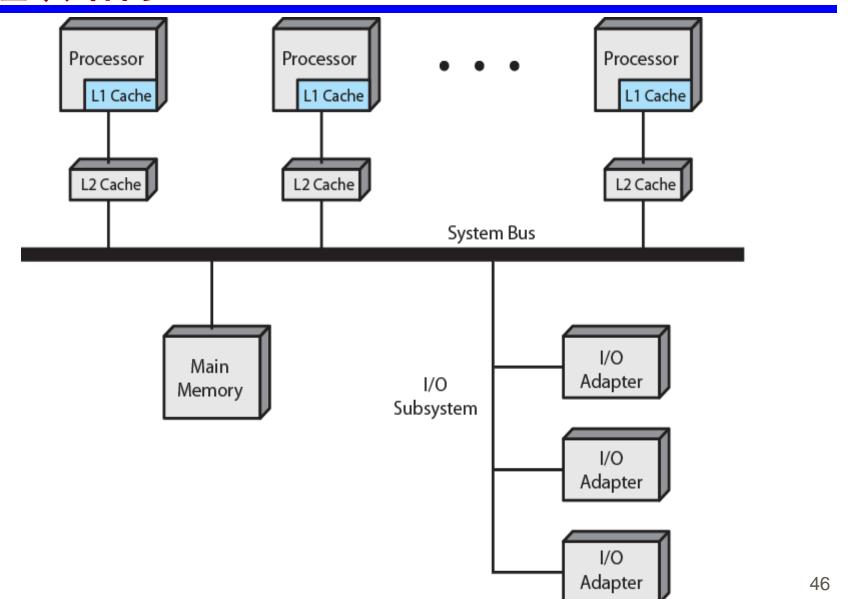
1.8.1 对称多处理器(SMP)

- SMP是具有以下特点的独立计算机系统:
 - 具有两个或两个以上可比性能的处理器;
 - 所有处理器共享内存和I/0设备,并通过总线或其他内部连接方式互连,从而每个处理器的访存时间大致上相同;
 - 所有处理器共享对I/0设备的访问,通过相同的通道或者可以连接到相同设备的不同通道;
 - 所有处理器可以执行相同的功能;
 - 整个系统由一个统一的操作系统控制,为多个处理器 及其程序提供各种级别的交互。
- 多处理器的存在对用户是透明的。

1.8.1 对称多处理器(SMP)

- 与单处理器相比, SMP优势:
 - 性能
 - 可用性
 - 渐近式增长
 - 可伸缩性
- 多处理器的存在对用户是透明的。

组织结构



1.8.2 多核计算机

- 多核(muliticore) 计算机是指将两个或多个处理器(核)组装在同一块芯片上。
- 每个核上通常会包含组成一个独立的处理器的所有零部件。

1.8.2 多核计算机

- 多核(muliticore) 计算机是指将两个或多个处理器(核)组装在同一块芯片上。
- 每个核上通常会包含组成一个独立的处理器的所有零部件。

作业

- 复习题 1.4, 1.5,
- 习题 1.13