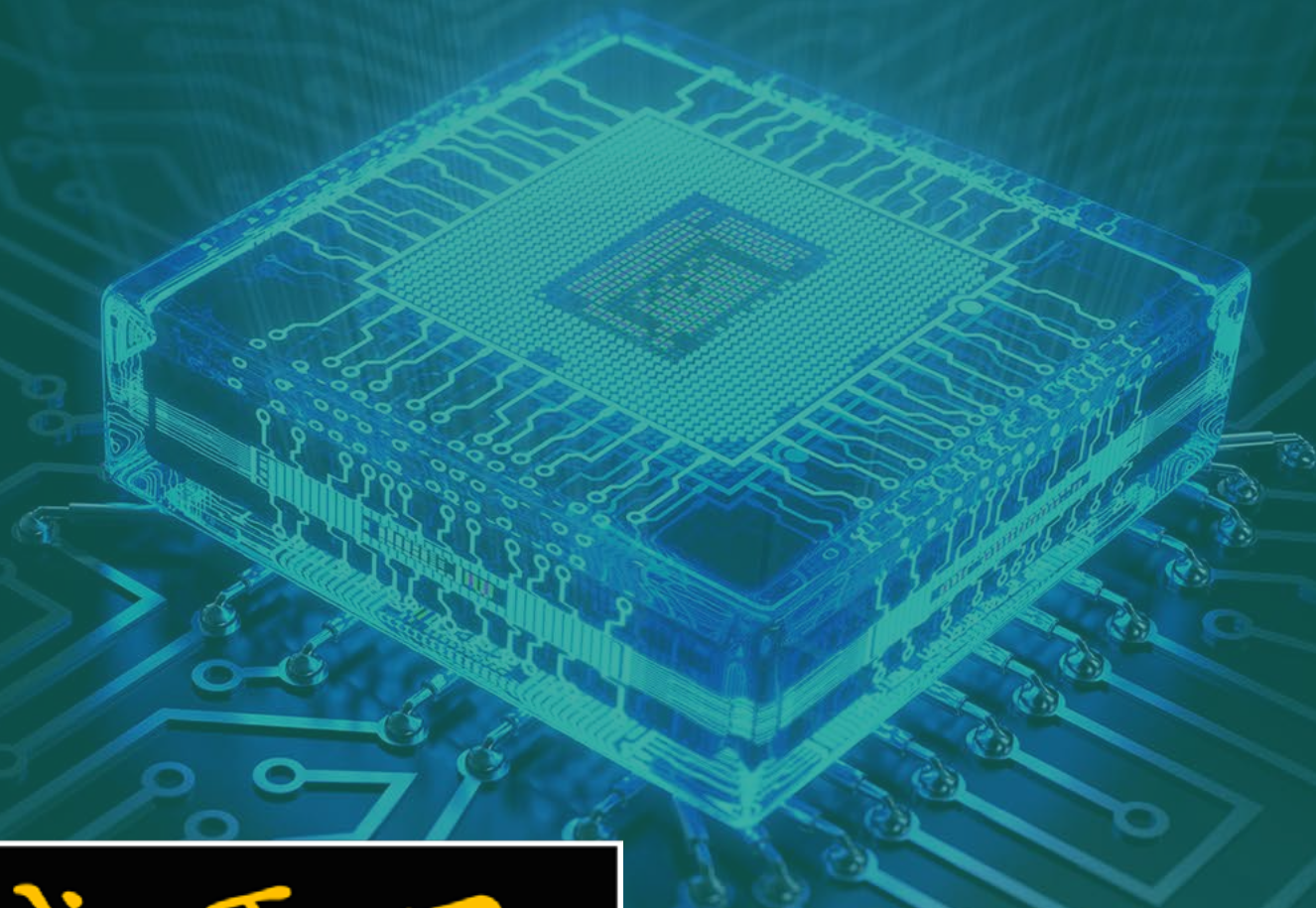




华中科技大学

计算机科学与技术学院

School of Computer Science & Technology, HUST



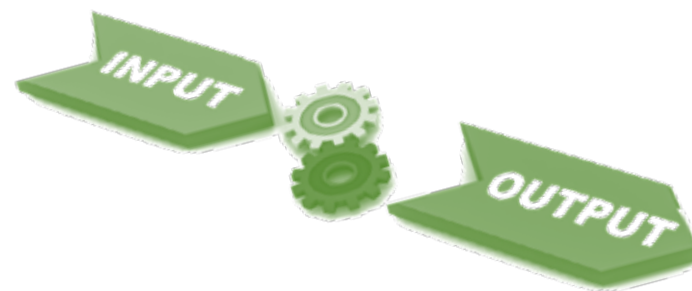
# 计算机组成原理



# 计算机组成原理



## 九、输入输出原理



# || 本章主要内容

n 9.1 输入输出设备与特性

n 9.2 I/O接口

n 9.3 数据传送控制方式

n 9.4 程序控制方式

n 9.5 程序中断方式

n 9.6 DMA方式

n 9.7 通道方式

n 9.8 常见I/O设备





## 9.1 输入输出设备与特性

- n 输入输出设备是计算机与人或者机器系统进行数据交互的装置，用于实现计算机内部二进制信息与外部不同形式信息的转换，简称外部设备或外设
  - p 输入设备：负责将数据、文字、图像、声音、电信号等转换成计算机可以识别的二进制信息，如键盘、鼠标、扫描仪、摄像头等；
  - p 输出设备：负责将计算机处理结果转换成数字、文字、图形、图像、声音或电信号，如显示器、打印机等；
  - p 输入输出设备：既能输入也能输出，如磁盘、网卡等。
- n 输入输出设备特性
  - p 异步性、实时性、独立性

# 输入/输出系统的组成与功能

n 外部设备、接口部件、总线以及相应的管理软件统称I/O系统

p I/O硬件

u 外设、控制器、I/O接口、总线

p I/O软件

u OS无关库，设备无关库，驱动

n 主要功能

p 完成计算机内部二进制信息与外部多种信息形式间的交流

p 保证CPU能够正确选择I/O设备并实现对其控制，与数据传输

p 利用数据缓冲、合适的数据传送方式，实现主机外设间速度匹配



# || 本章主要内容

n 9.1 输入输出设备与特性

■ 9.2 I/O接口

n 9.3 数据传送控制方式

n 9.4 程序控制方式

n 9.5 程序中断方式

n 9.6 DMA方式

n 9.7 通道方式

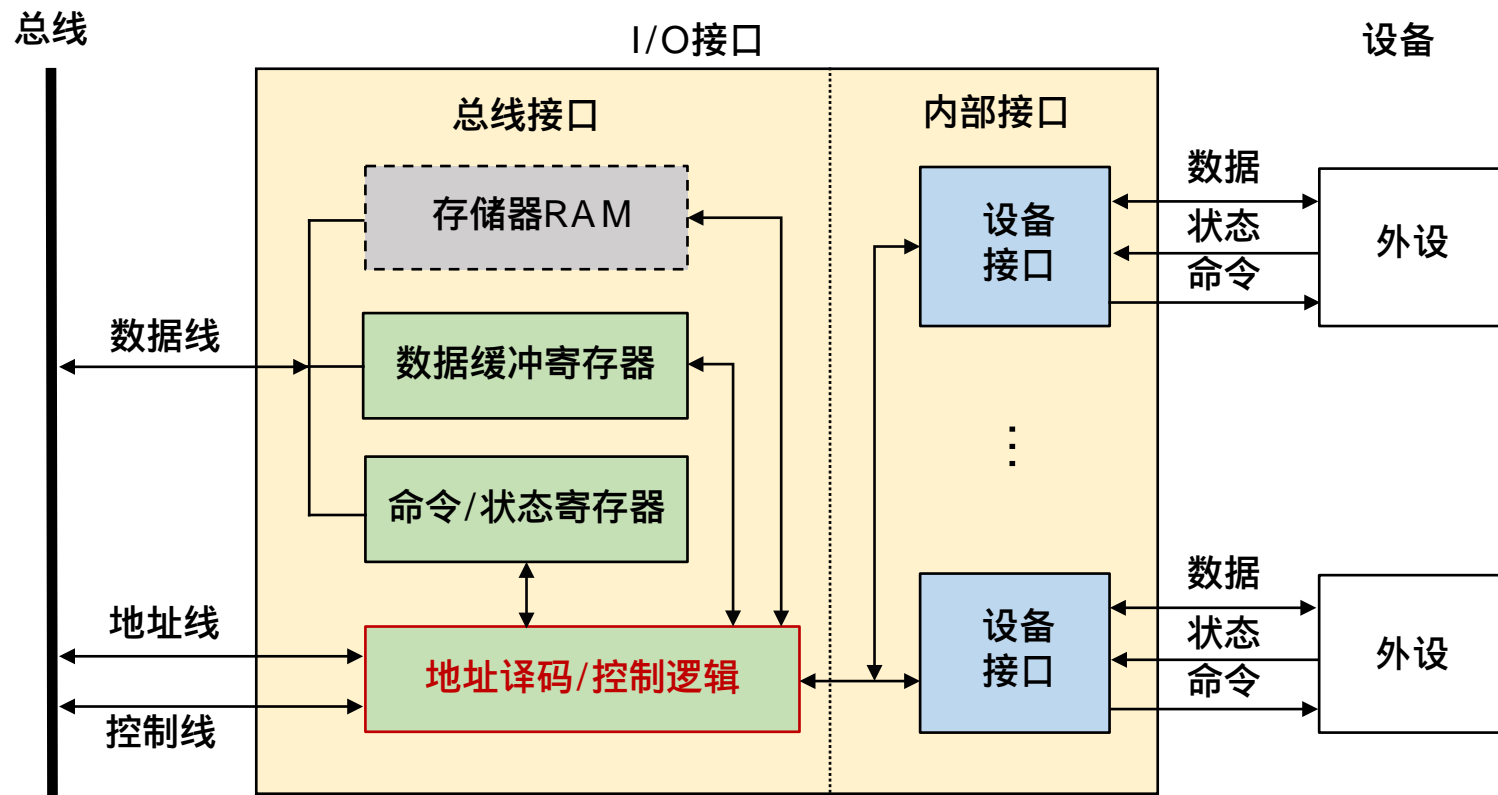
n 9.8 常见I/O设备



# I/O接口定义与功能

- n I/O接口：连接总线与I/O设备的物理和逻辑界面
  - p 既包括物理连接电路，也包括软件逻辑接口
  - p 所有设备均通过 I/O接口（总线接口）与总线相连
  - p CPU使用设备地址经总线与I/O接口通信访问I/O设备
  - p 标准接口有利于提升I/O系统的独立性，降低连接复杂度
- n I/O接口功能
  - p 设备寻址、数据交互、设备控制
  - p 状态检测、数据缓冲、格式转换

# I/O接口结构



## I/O接口的功能

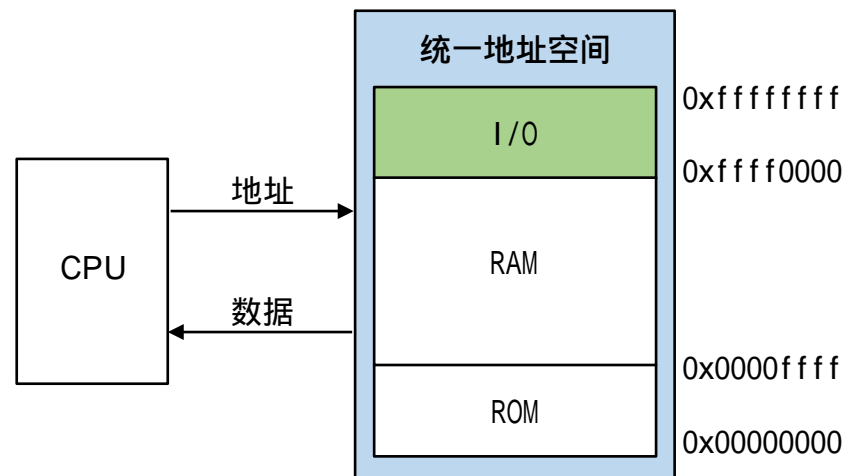
- 设备寻址
- 数据交互
- 设备控制
- 状态检测
- 数据缓冲
- 格式转换



# I/O接口编址

## n 统一编址

- p 内存映射编址 ( Memory-mapped )
- p 外设地址与内存地址统一编址，同一个地址空间
- p 不需要设置专用的I/O指令
- p 访存指令访问外设，访问什么设备取决于地址



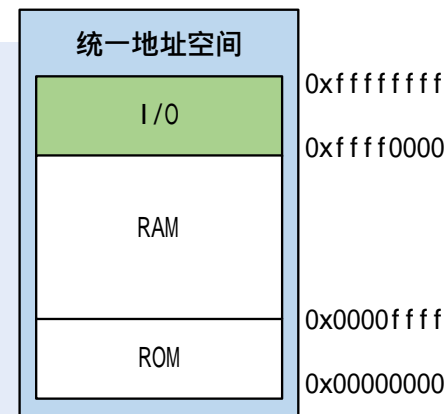
## n 独立编址

- p 端口映射编址 ( Port-mapped )
- p I/O地址空间与主存地址空间相互独立
- p I/O地址又称为I/O端口
- p 不同设备中的不同寄存器和存储器都有唯一的端口地址
- p 使用I/O指令访问外设

## 不同编址方式程序实例

n 统一编址：访存指令访问外设，具体访问什么设备取决于内存地址

|            |                  |                   |
|------------|------------------|-------------------|
| <b>lw</b>  | \$t0, 0x00000004 | # 从 ROM 读取一个字     |
| <b>sw</b>  | \$t0, 0x00000004 | # 写 ROM (会产生总线错误) |
| <b>lbu</b> | \$t0, 0x00010001 | # 从内存读取一个字节       |
| <b>sb</b>  | \$t0, 0xff000002 | # 写一个字节到内存        |
| <b>lbu</b> | \$t0, 0xffff0000 | # 从 I/O 设备读一个字节   |
| <b>sb</b>  | \$t0, 0xffff0004 | # 写一个字节到 I/O 设备   |



n 独立编址：I/O指令访问外设，具体访问什么设备取决于端口地址

|                     |   |
|---------------------|---|
| <b>OUT</b> DX, AL   | I/O 写：将 AL 寄存器中的字节写入 DX 寄存器对应的 I/O 端口地址 |
| <b>IN</b> AL, DX    | I/O 读：从 DX 寄存器对应的 I/O 端口地址读取一个字节        |
| <b>MOV</b> [BX], AL | 内存写：将 AL 寄存器中的值送到 BX 寄存器对应的内存地址         |

# I/O接口的软件

n 现代计算机中用户须通过OS间接访问设备，屏蔽设备细节，使用更方便

n 与OS无关的I/O库（用户态）

如C语言的标准I/O库 `stdio.h`、`printf`、`scanf`、`getchar`、`putchar`、  
`fopen`、`fseek`、`fread`、`fwrite`、`fclose`等

用户程序主要通过调用I/O库访问设备，方便程序在不同OS间移植

n 与设备无关的OS调用库（内核态）

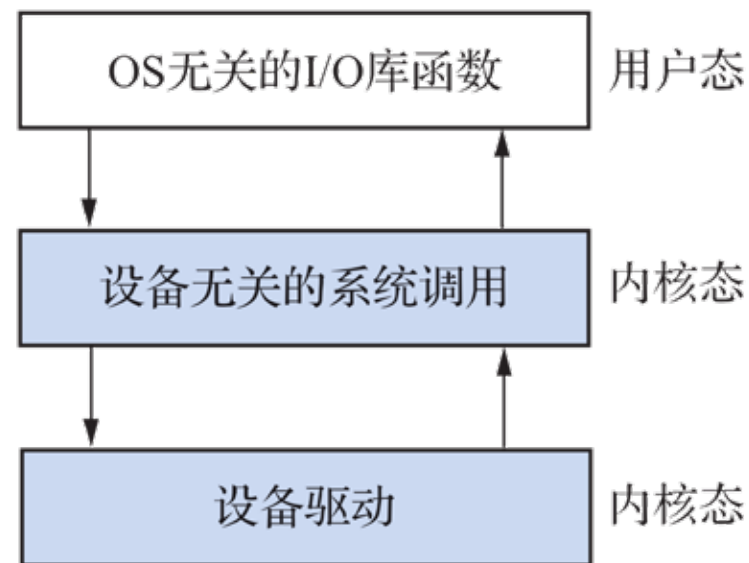
`open`、`read`、`write`、`seek`、`ioctl`、`close`

n 独立的设备驱动程序（内核态）

设备驱动程序是与设备相关的I/O软件部分

不同设备对应不同的驱动程序

遵循具体设备的I/O接口约定，包含设备接口细节



# I/O接口分类

## n 按数据传送方式

- p 并行、串行接口

## n 按接口的灵活性

- p 可编程接口、不可编程接口

## n 按通用性：通用、专用接口

## n 按总线传输的通信方式：同步、异步接口

## n 按访问外设的方式

- p 程序控制方式、程序中断方式、DMA及通道处理机接口

## || 本章主要内容

n 9.1 输入输出设备与特性

n 9.2 I/O接口

**n 9.3 数据传送控制方式**

n 9.4 程序控制方式

n 9.5 程序中断方式

n 9.6 DMA方式

n 9.7 通道方式

n 9.8 常见I/O设备







# 数据传输控制方式

## n 程序控制方式

p 程序查询方式、直接输入输出方式

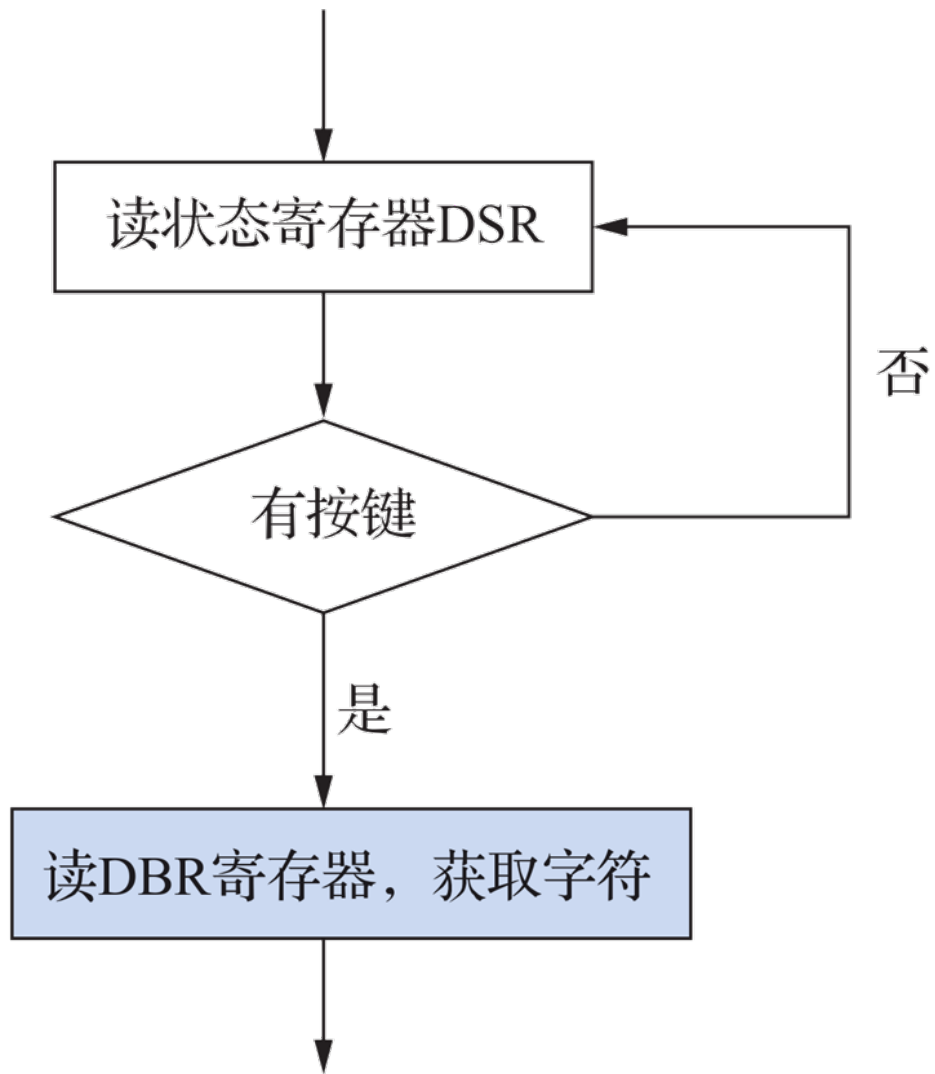
n 程序中断方式

n 直接内存访问方式

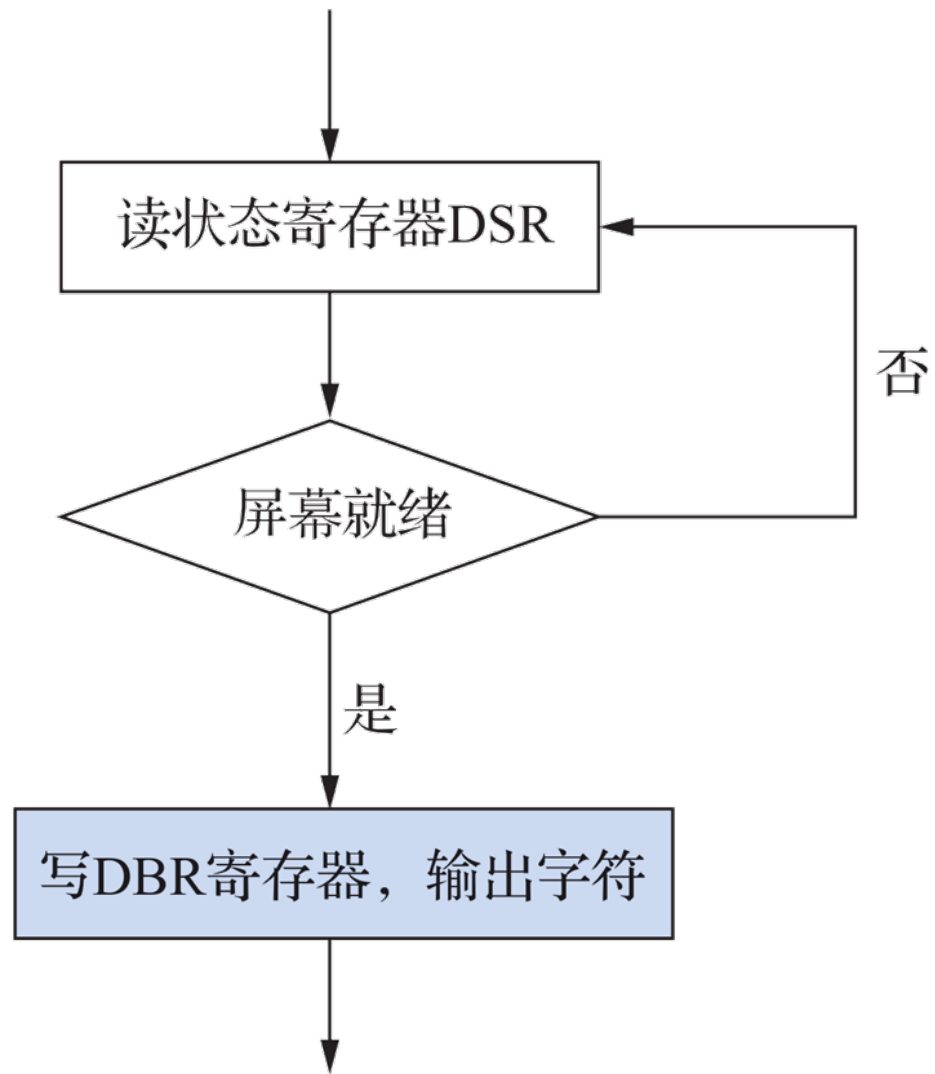
n 通道方式

n 外围处理机方式

## 简单设备查询流程

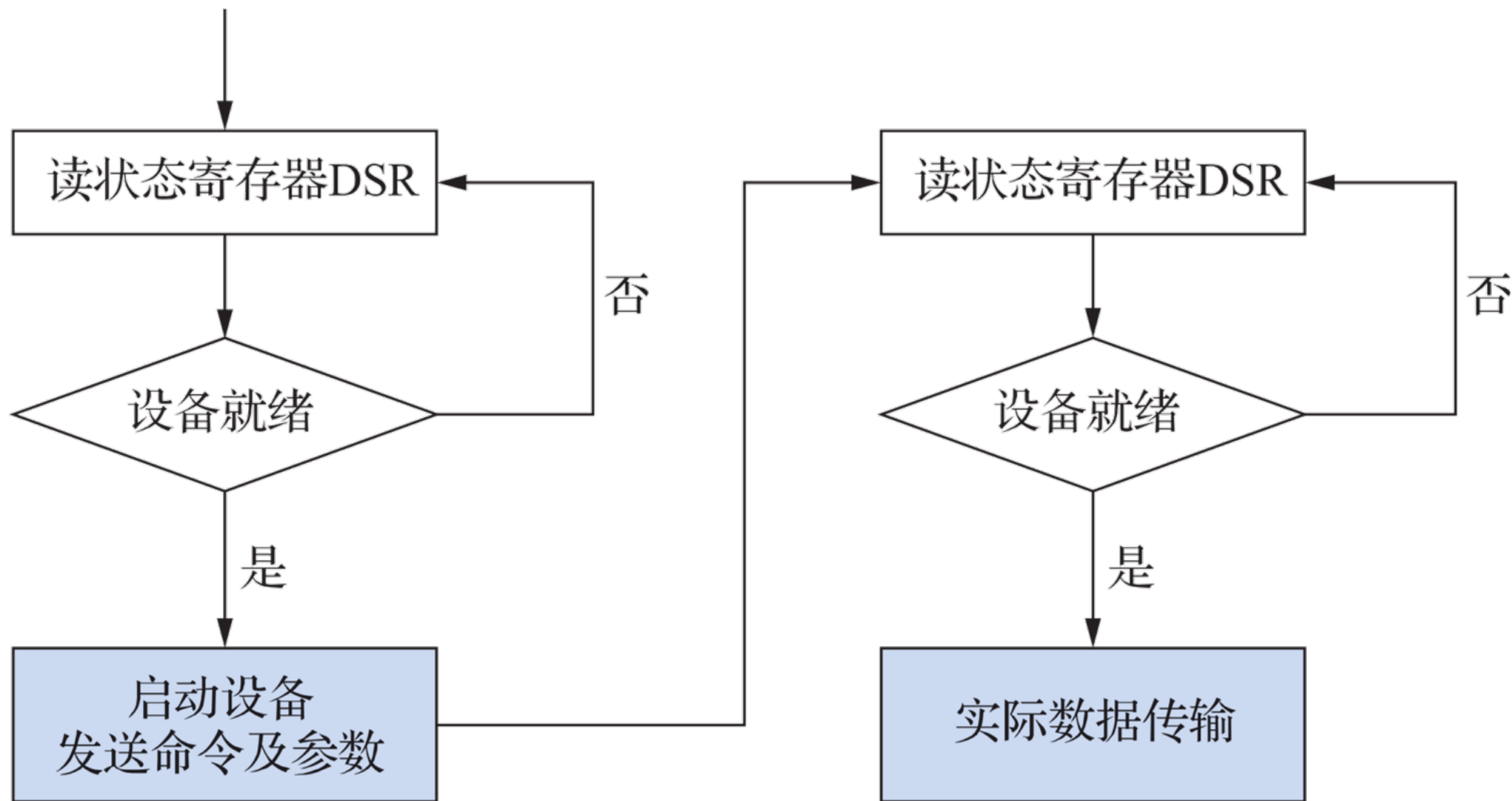


(a) 键盘程序查询流程



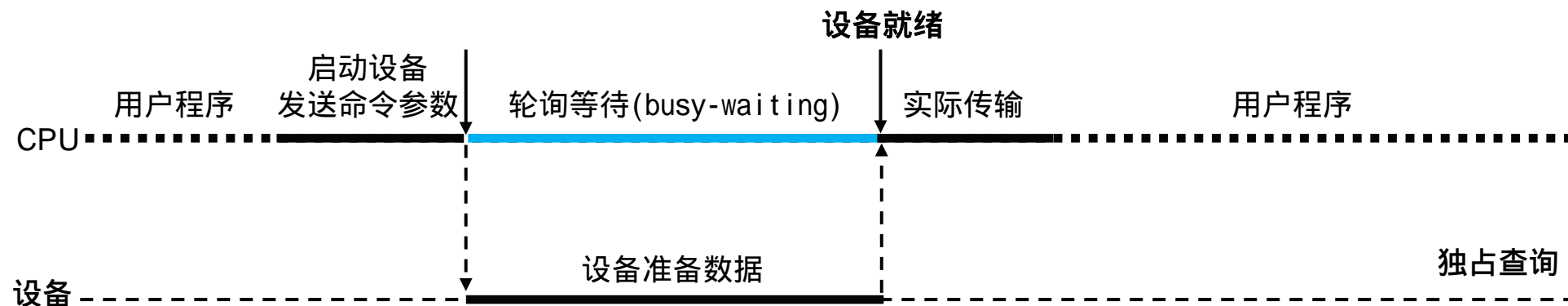
(b) 字符终端程序查询流程

## 复杂设备查询流程

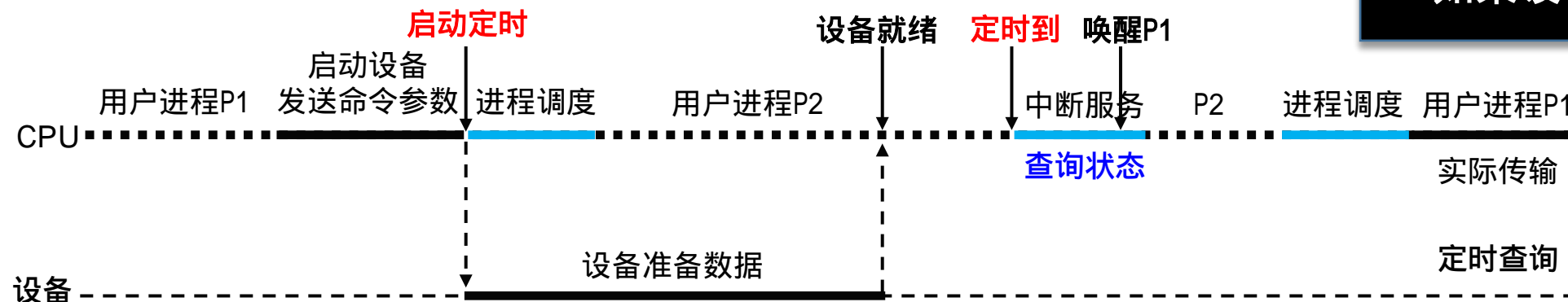




# 程序查询方式运行轨迹



如果设备过快？



## 例题

n 例 9.1 假设某程序查询方式的输入输出系统采用定时轮询方式，每次定时中断服务开销需要 400 个时钟周期，CPU 的时钟频率为 200MHz，包括鼠标和硬盘两个外部设备，求两种不同外部设备进行 I/O 操作时的 CPU 时间占用率。

(1) 鼠标以字节为单位进行数据传输，假设每秒必须进行 50 次轮询才能保证不会错过任何鼠标操作，每次轮询成功后的实际数据传输需要 13 个时钟周期。

n CPU 每秒用于鼠标 I/O 操作的时间为  $(400T + 13T) \times 50 = 20650T$

n 鼠标 I/O 的 CPU 时间占用率 =  $20650T / 1s = 0.01\%$

n 由此可见，对鼠标进行定时轮询操作基本不影响 CPU 的性能



## 例题

n 例 9.1 假设某程序查询方式的输入输出系统采用定时轮询方式，每次定时中断服务开销需要 400 个时钟周期，CPU 的时钟频率为 200MHz，包括鼠标和硬盘两个外部设备，求两种不同外部设备进行 I/O 操作时的 CPU 时间占用率。

(2) 硬盘以512字节的扇区为单位传输数据，启动阶段发送命令和参数需要90个时钟周期，实际传输阶段需要 1555 个时钟周期，CPU 访问硬盘的速率为 20MB/s。

n 每秒传输的次数为： $20\text{MB} / 512\text{B} = 39062.5$  次（数传率、频率均以 10 为基数）

n 理想情况下每次传输只定时轮询一次

n 每个扇区传输开销 = 启动开销90T + 中断开销400T + 数据传输开销 1555T = 2045T

n 总 CPU 占用率 =  $2045\text{T} \times 39062.5 / 1\text{s} = 39.94\%$ 。

## || 本章主要内容

n 9.1 输入输出设备与特性

n 9.2 I/O接口

n 9.3 数据传送控制方式

n 9.4 程序控制方式

■ 9.5 程序中断方式

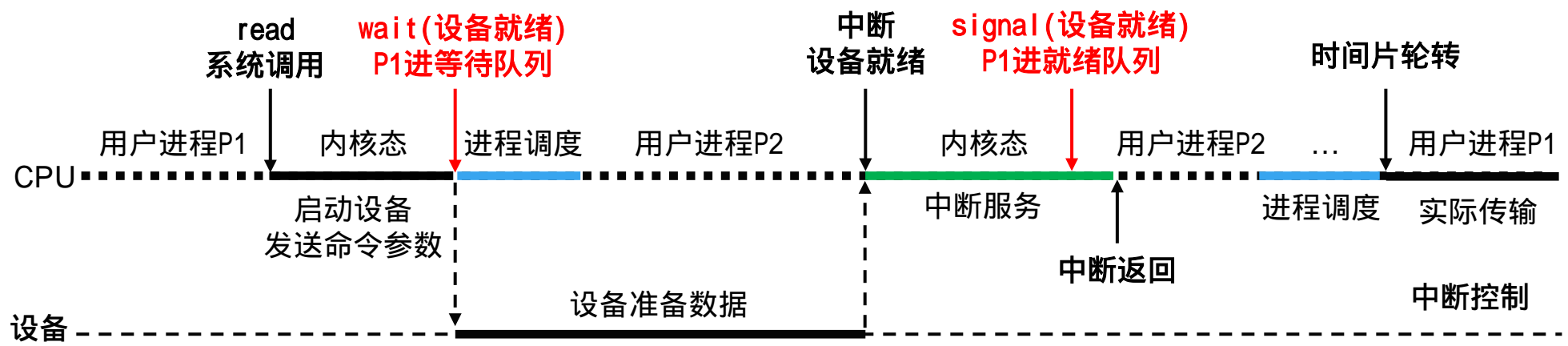
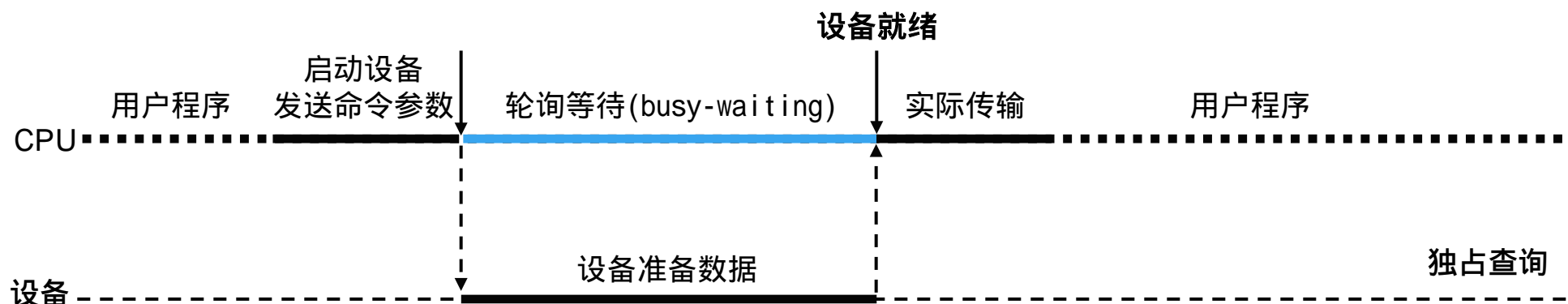
n 9.6 DMA方式

n 9.7 通道方式

n 9.8 常见I/O设备



# 中断控制方式



## 中断优势

- n 提高了CPU的使用效率

  - p 主动告知机制避免了反复查询设备状态

  - p 仍需CPU占用（中断服务子程序运行时间+中断开销）

- n 适合随机出现的服务

- n 需要专门的硬件

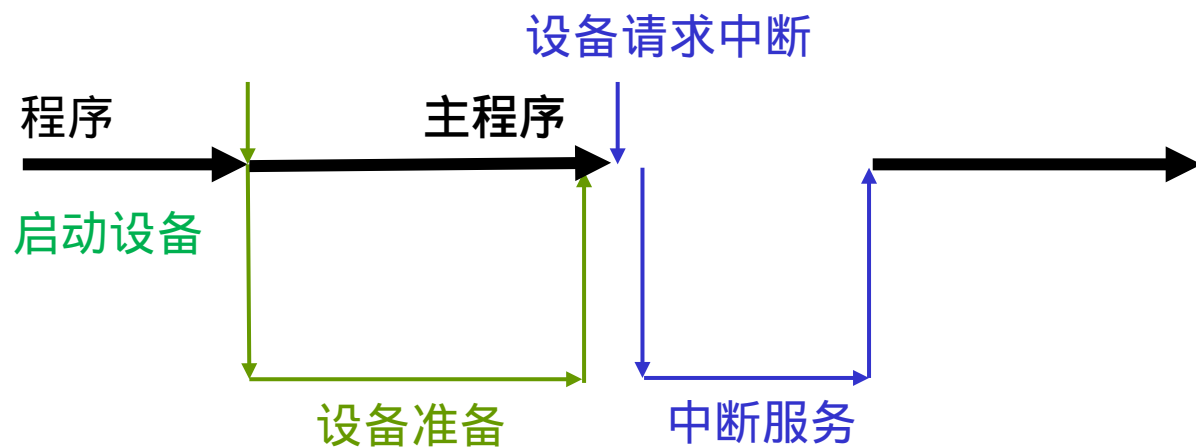
# || 程序中断方式

- n 中断基本概念
- n 程序中断基本接口
- n 中断仲裁方式
- n 中断控制器



## 中断基本概念

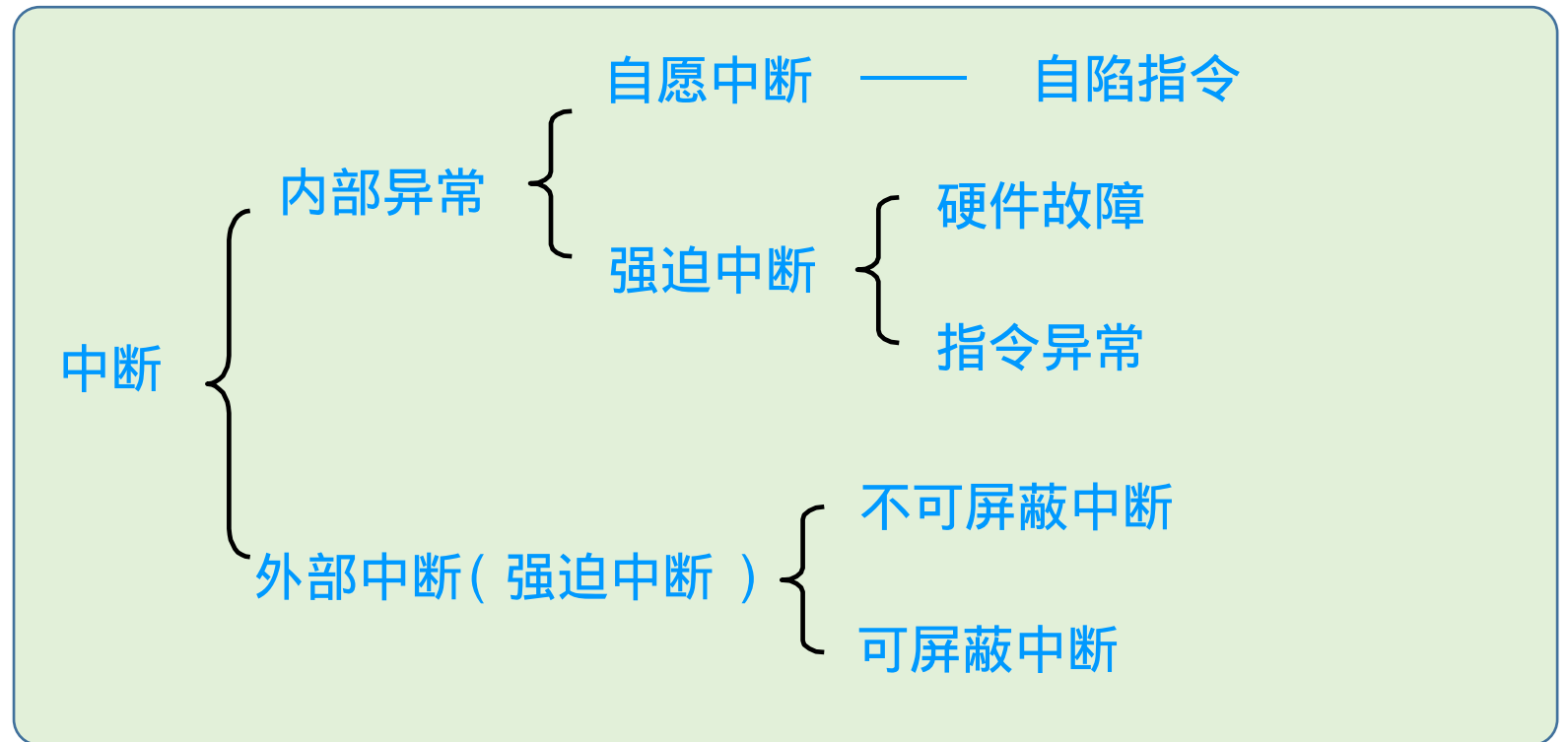
- n CPU暂时中止现程序的执行，转去执行为某个随机事件服务的中断处理子程序，处理完后自动恢复原程序的执行
  - p 原因：不处理会体验不好，会丢失数据，甚至造成灾难
- n 实现主机和外设准备阶段的并行工作
  - p 避免重复查询外设状态、提升工作效率



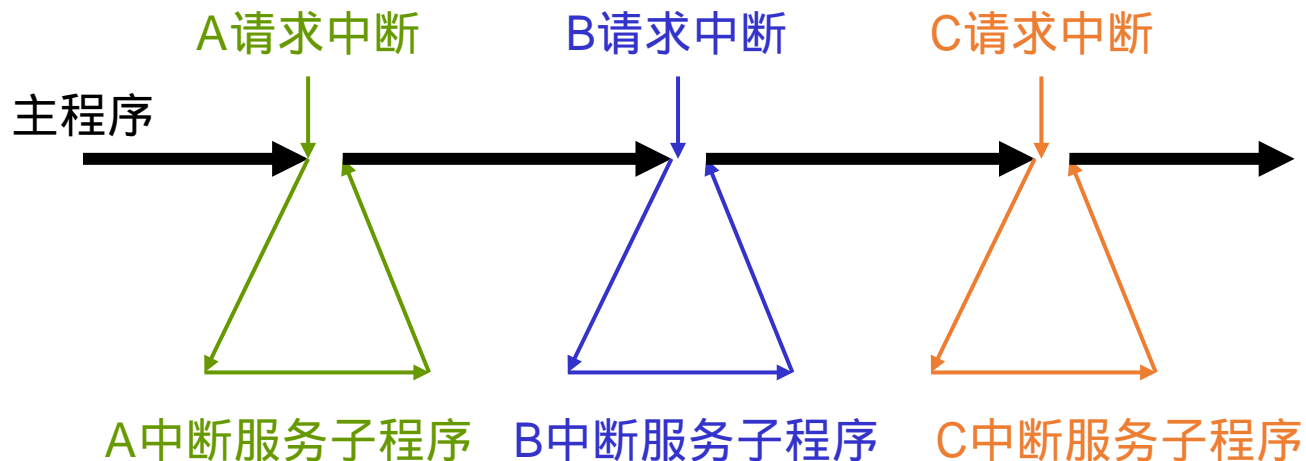
# 中断的分类与作用

n 中断技术赋予计算机应变能力，将有序的运行和无序的事件统一起来，大大增强了系统的处理能力

- p 主机外设并行工作
- p 程序调试
- p 故障处理
- p 实时处理
- p 人机交互



# 程序中中断处理示意图



n 子程序与中断服务子程序的异同？

p 共同点：调用后要返回

p 差异：调用方式（显式/随机），保存寄存器（中断更多），返回位置与方式（调用位置/中断位置）

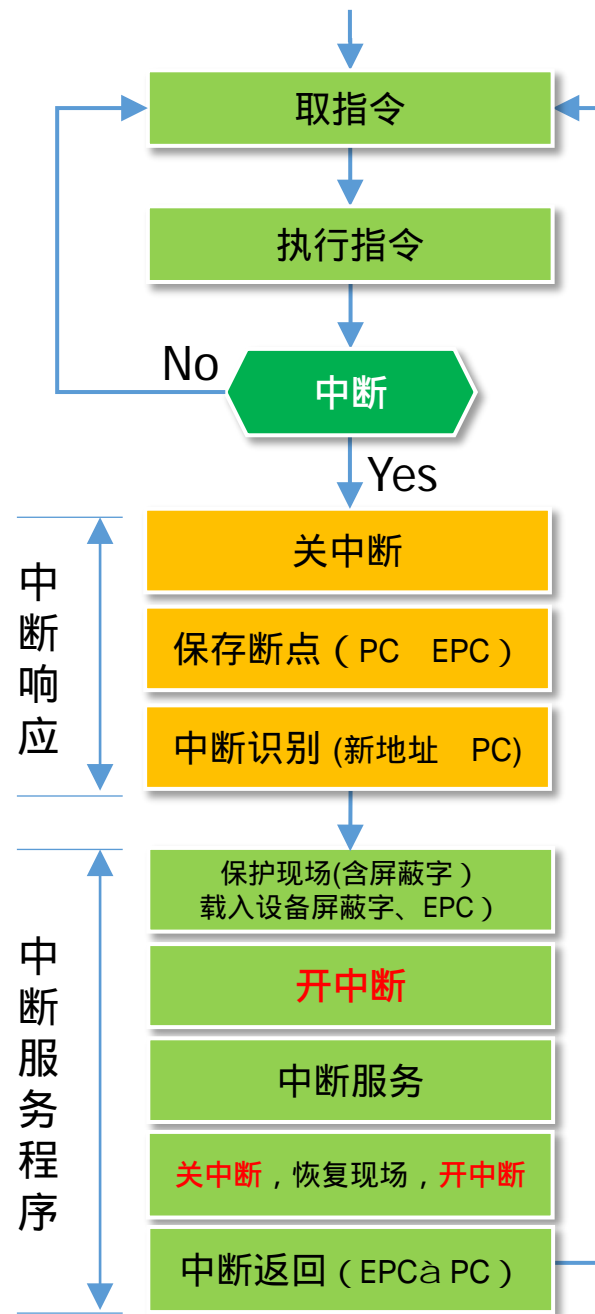
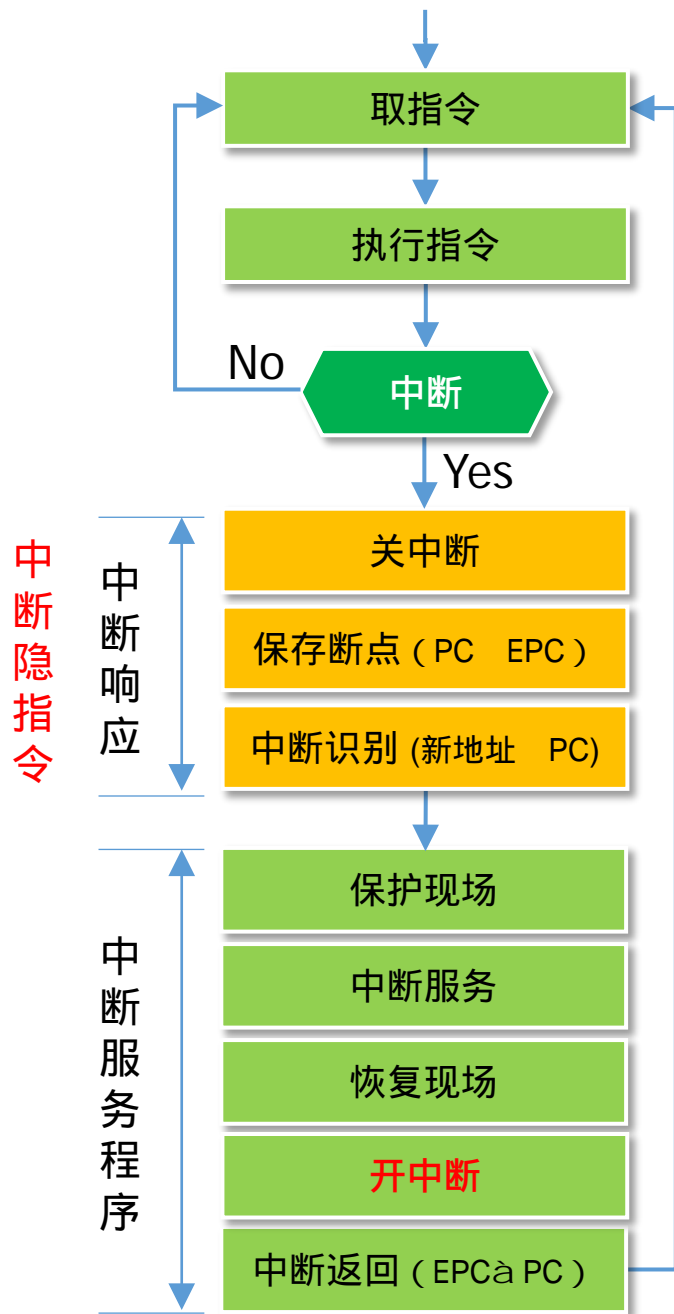
n 如果A，B，C同时产生中断？

p 中断优先级问题，中断仲裁

n 如果正在运行A中断服务子程序，又收到B中断？

p 中断嵌套

什么是主程序？



## 中断优先级

- n 多设备同时产生中断请求时，如何处理？
  - p 优先级高的先响应，优先级低的后响应
  - p CPU优先级随不同中断服务程序而改变
    - u 执行某设备中断服务子程序
    - u CPU优先级就与该设备优先级一样
  - p 优先级高的中断请求可否中断优先级低的程序？



# 单级中断与多级中断

n 高优先级中断请求能否中断运行中的程序呢？

n 系统硬件、软件开销的权衡

## p 单级中断

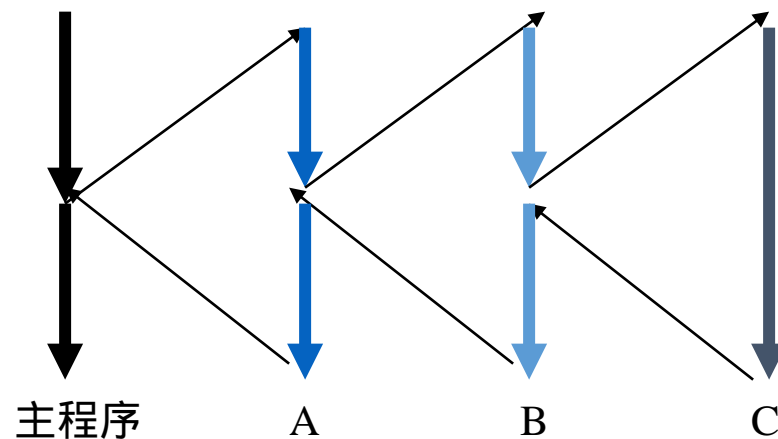
u 所有中断源均属同一级，离CPU近的优先级高

u CPU处理某个中断时，不响应其他中断

## p 多重中断

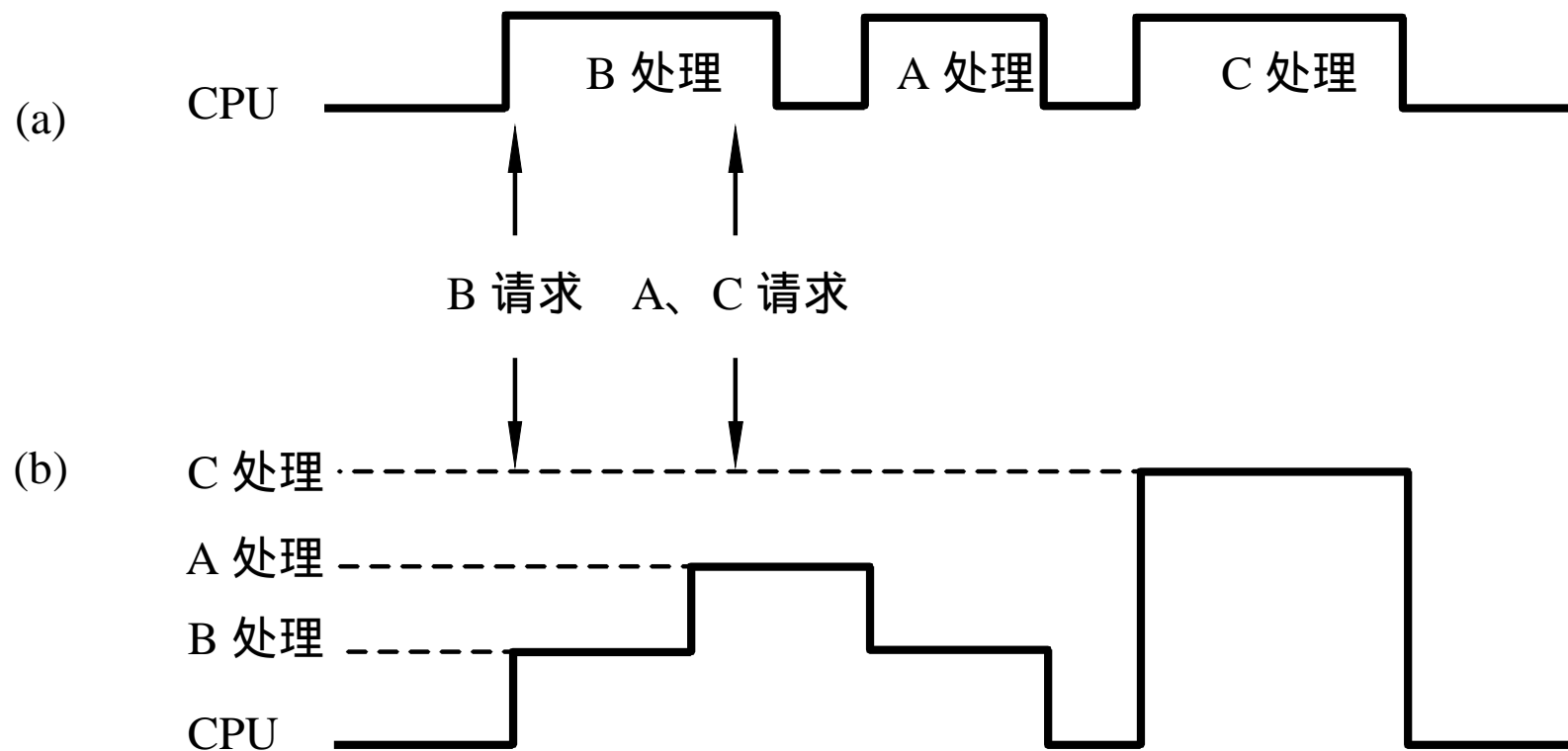
u 优先级高的中断可以打断优先级低的中断服务程序

u 中断嵌套



## 同时中断请求的处理方法

$A > B > C$



## 划分优先级的一般规律

- n 硬件故障中断属于最高级，其次是程序错误中断
- n 非屏蔽中断优于可屏蔽中断
- n DMA请求优先于I/O设备传送的中断请求
- n 高速设备优于低速设备
- n 输入设备的中断优于输出设备
- n 实时设备优先于普通设备

# 中断屏蔽

## n 响应优先级

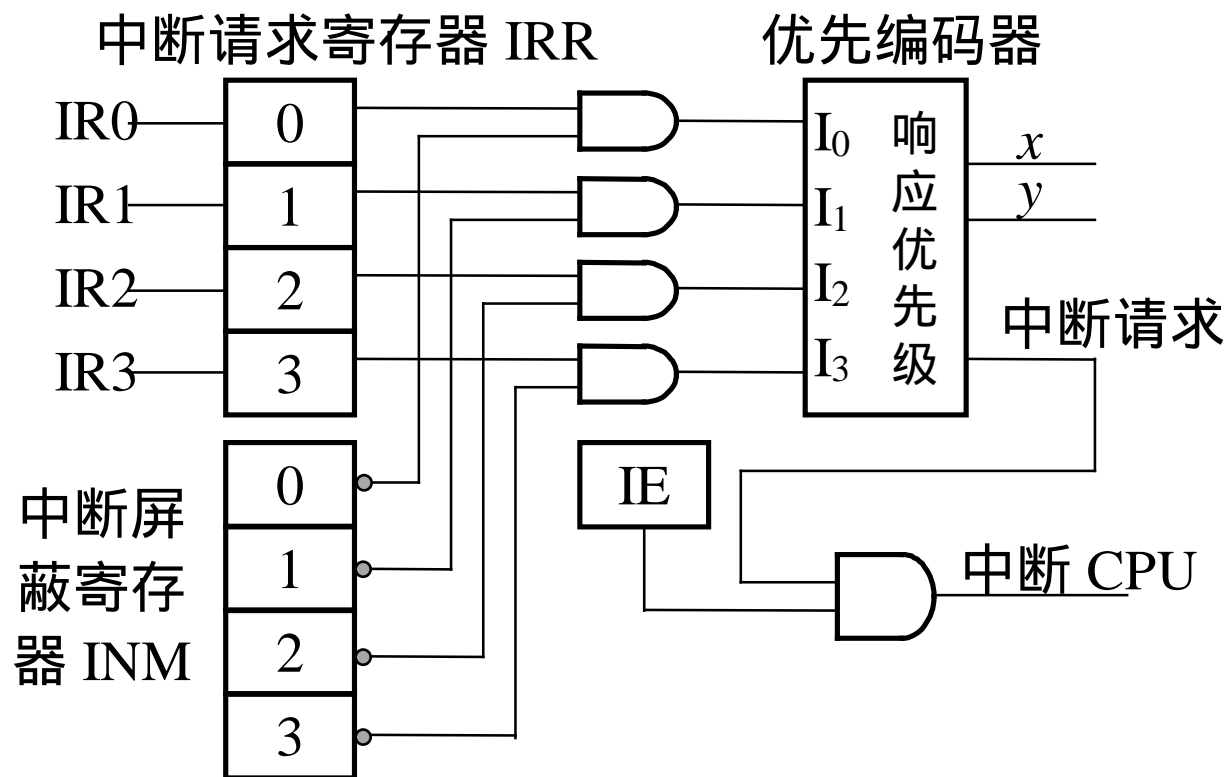
- p CPU对各设备中断请求进行响应，并准备好处理的先后次序
- p 这种次序往往在硬件线路上已固定，不便于变动

## n 处理优先级

- p CPU实际对各中断请求处理的先后次序
- p 如果不使用屏蔽技术，响应的优先级就是处理优先级

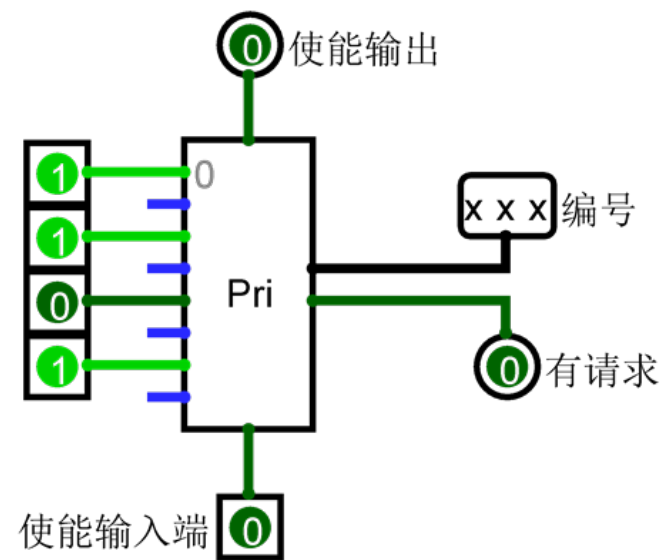
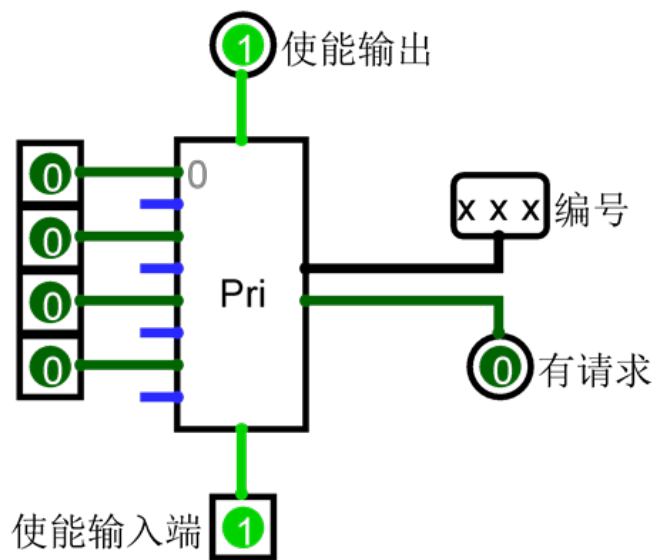
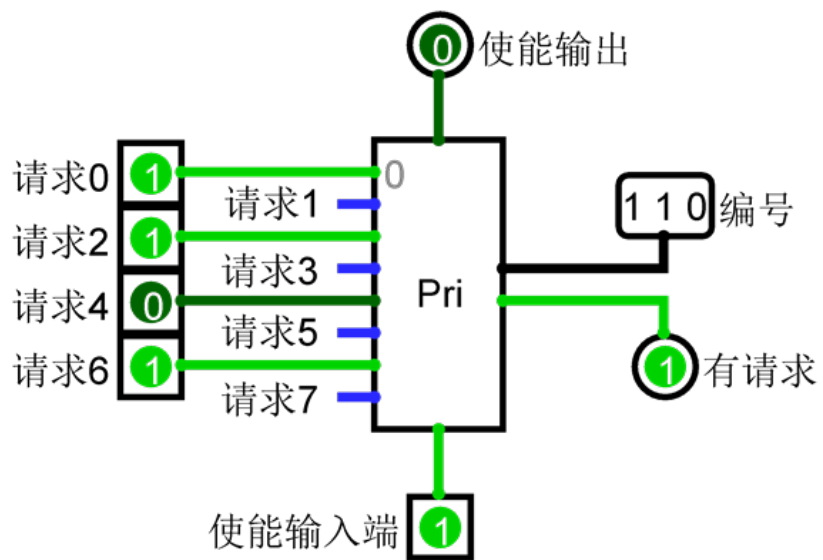
## n 中断屏蔽技术可动态改变各设备的处理优先级

## 中断屏蔽方式



- n 优先编码器决定响应优先级，中断屏蔽改变处理优先级
- n 当CPU执行某个设备的中断服务程序时，如何设置中断屏蔽字？

# || 优先编码器



## 中断屏蔽位

### n 中断请求寄存器IRR

- p 对应位为1表示相应外设发出了中断请求
- p 中断字，中断码

### n 中断屏蔽寄存器INM

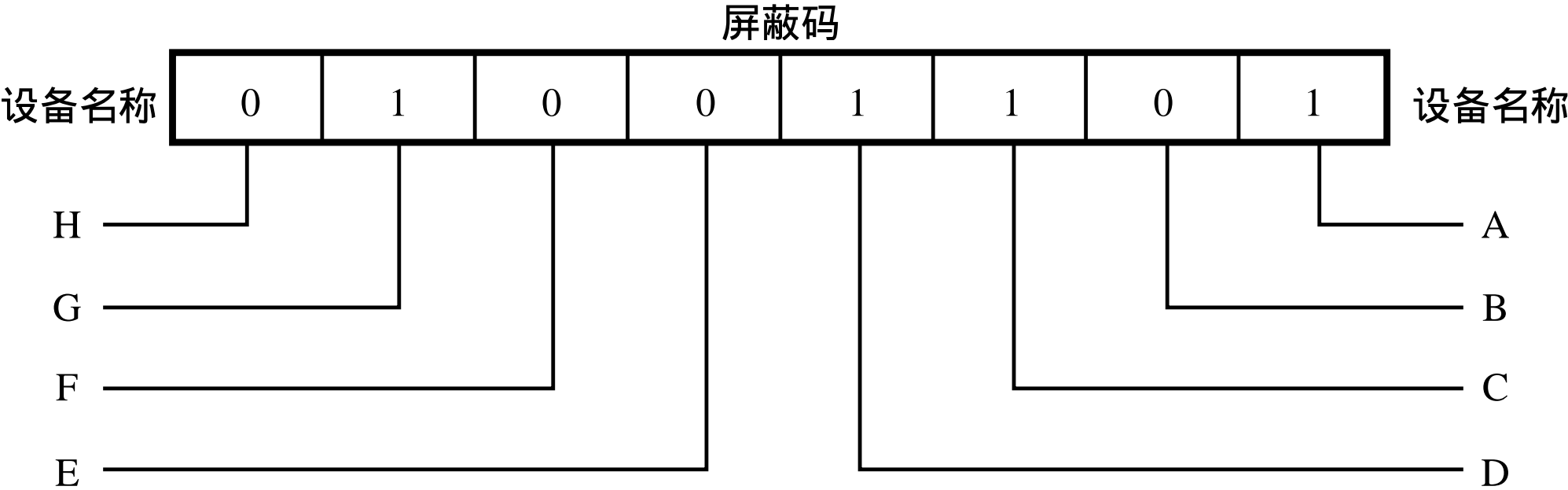
- p 对应位为1设置屏蔽，否则取消屏蔽
- p 每个设备都有自己独立的中断屏蔽字
- p CPU执行某设备的中断服务子程序时将其中断屏蔽字载入INM
- p 不可屏蔽中断不受中断屏蔽寄存器的控制

### n 中断允许触发器IE（中断使能寄存器）



# 屏蔽码

- n 控制各设备接口的屏蔽触发器，可改变处理次序。
- n 运行某个设备的中断服务程序时载入对应的屏蔽码



## 例子

- n 假定硬件原来的响应顺序为 $0 > 1 > 2$ ，试设置各设备中断屏蔽字，将中断优先级改为 $1 > 2 > 0$ 。

| 设备/屏蔽字 | L0 | L1 | L2 |
|--------|----|----|----|
| L0     | 1  | 0  | 0  |
| L1     | 1  | 1  | 1  |
| L2     | 1  | 0  | 1  |

## || 优先级实现---中断仲裁

n 同一时刻可能有多个设备同时发出中断请求，响应谁？

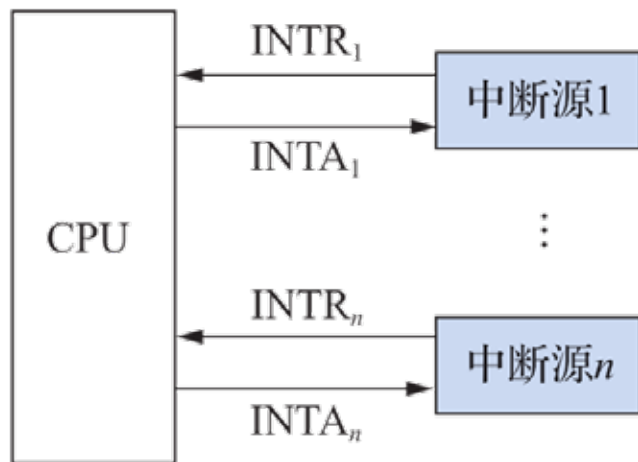
p 独立请求

p 链式查询

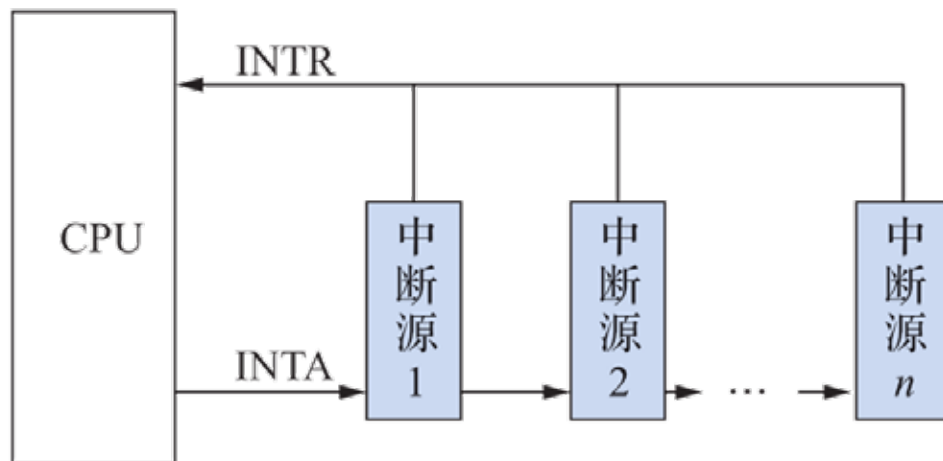
p 中断控制器方式

p 分组链式结构

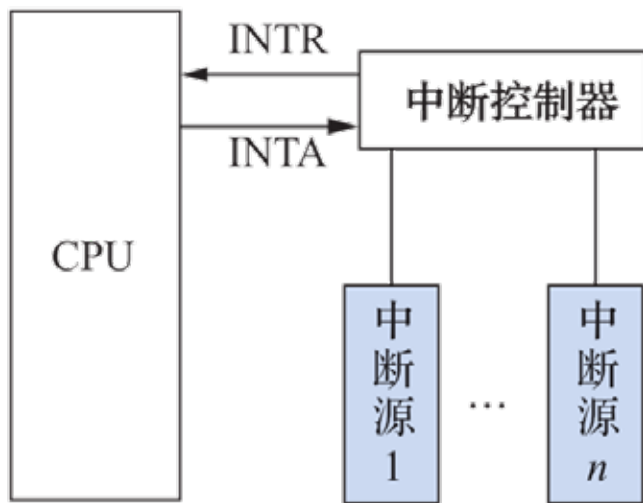
# 中断请求信号的传输方式



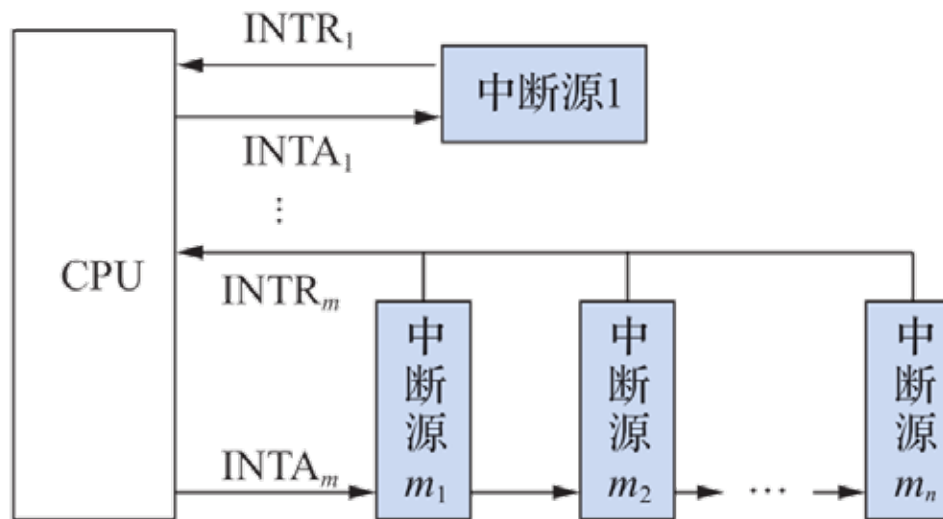
(a) 独立请求方式



(b) 链式请求方式



(c) 中断控制器方式



(d) 分组链式

## 中断识别（寻找入口地址）

### n 独立请求：向量中断

p 将服务程序入口(中断向量)组织在中断向量表中；响应时由硬件直接产生中断号（向量地址），查中断向量表取得服务程序入口，转入相应服务程序。

u 硬件查询法

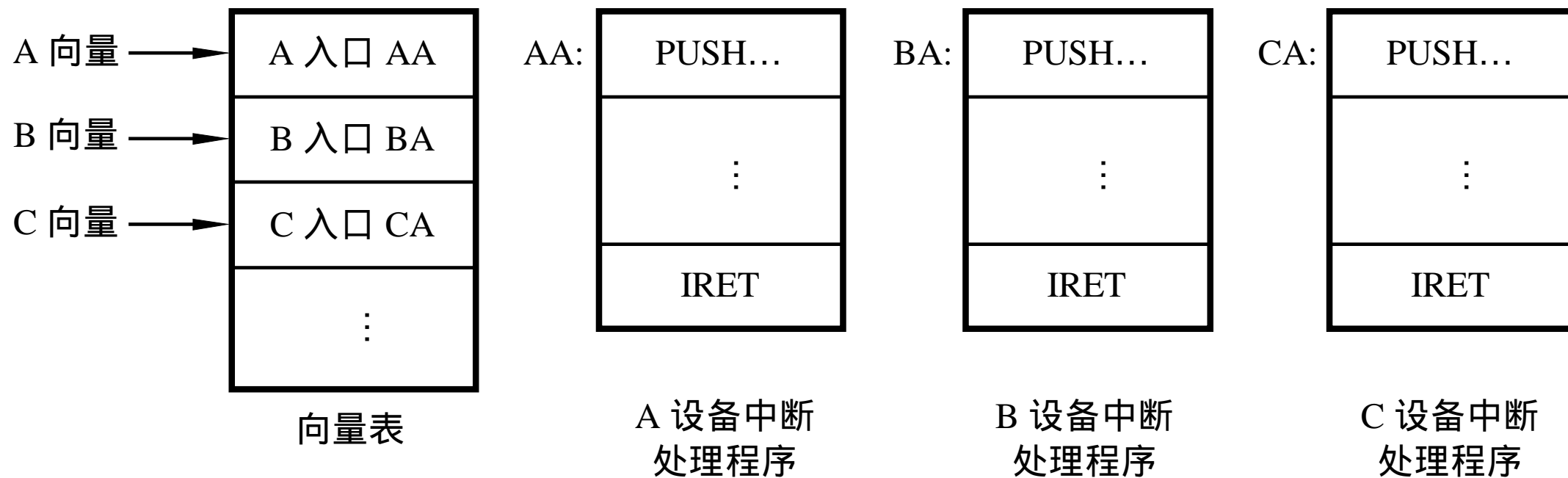
### n 中断共享：非向量中断

p 将服务程序入口组织在查询程序中

p 响应时执行查询程序查询中断源，查询特定端口(GPIO)识别中断源

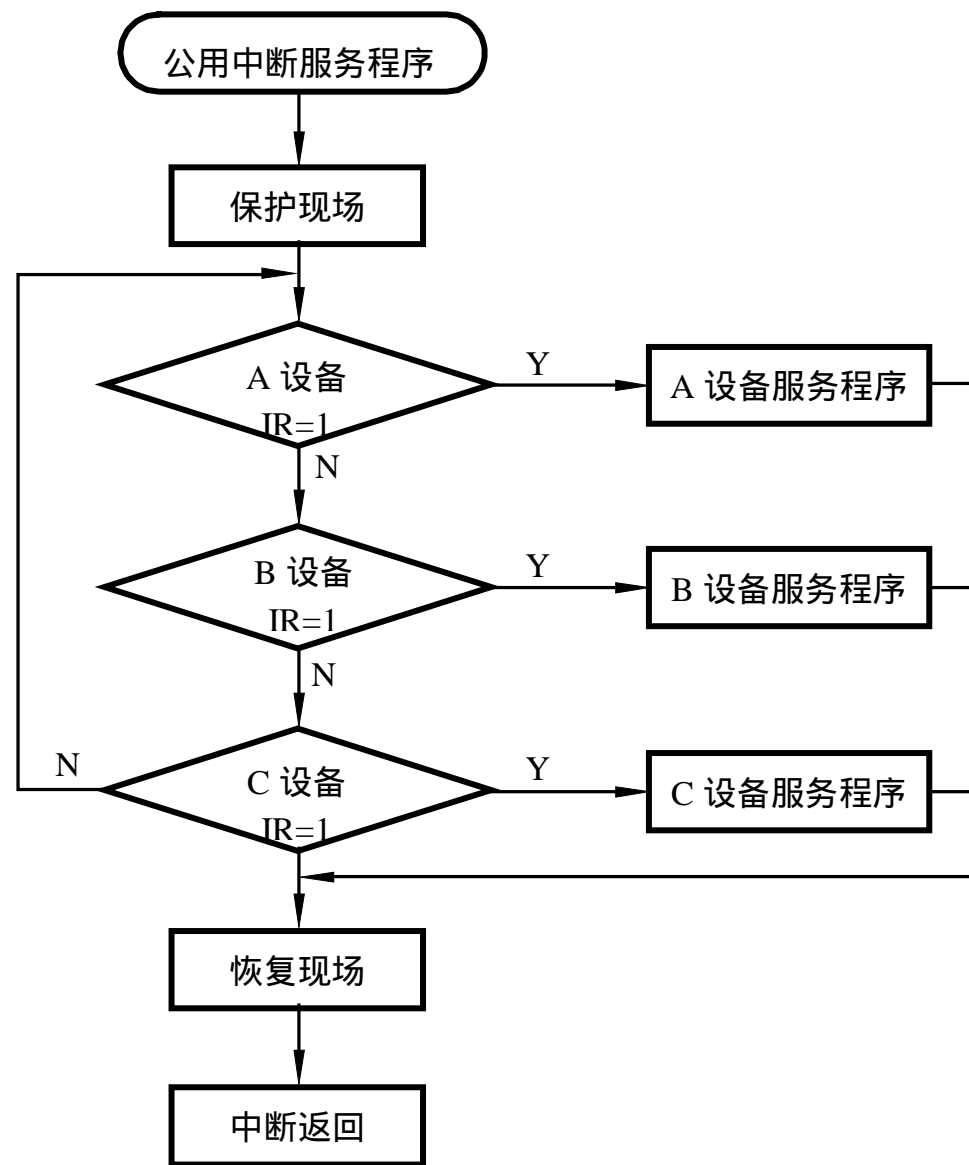
u 程序识别（软件方法）

# 中断向量法





# 程序识别



# 中断处理中的问题

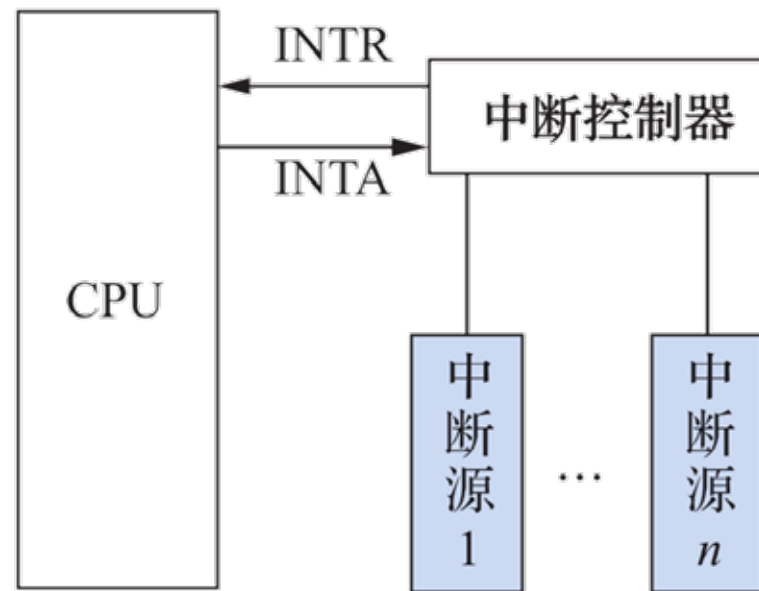
## n 中断响应条件

- p 中断允许触发器IE=1
- p 对应的中断未被屏蔽
- p 无更高优先级的DMA请求
- p 中断嵌套必须优先级更高
- p 指令已经执行完最后一个机器周期
  - u 保证指令执行的完整性

## n 保存现场，恢复现场

- p 中断程序用到的通用寄存器，EPC，屏蔽字
- p 缺页异常的断点和外部中断断点不一致

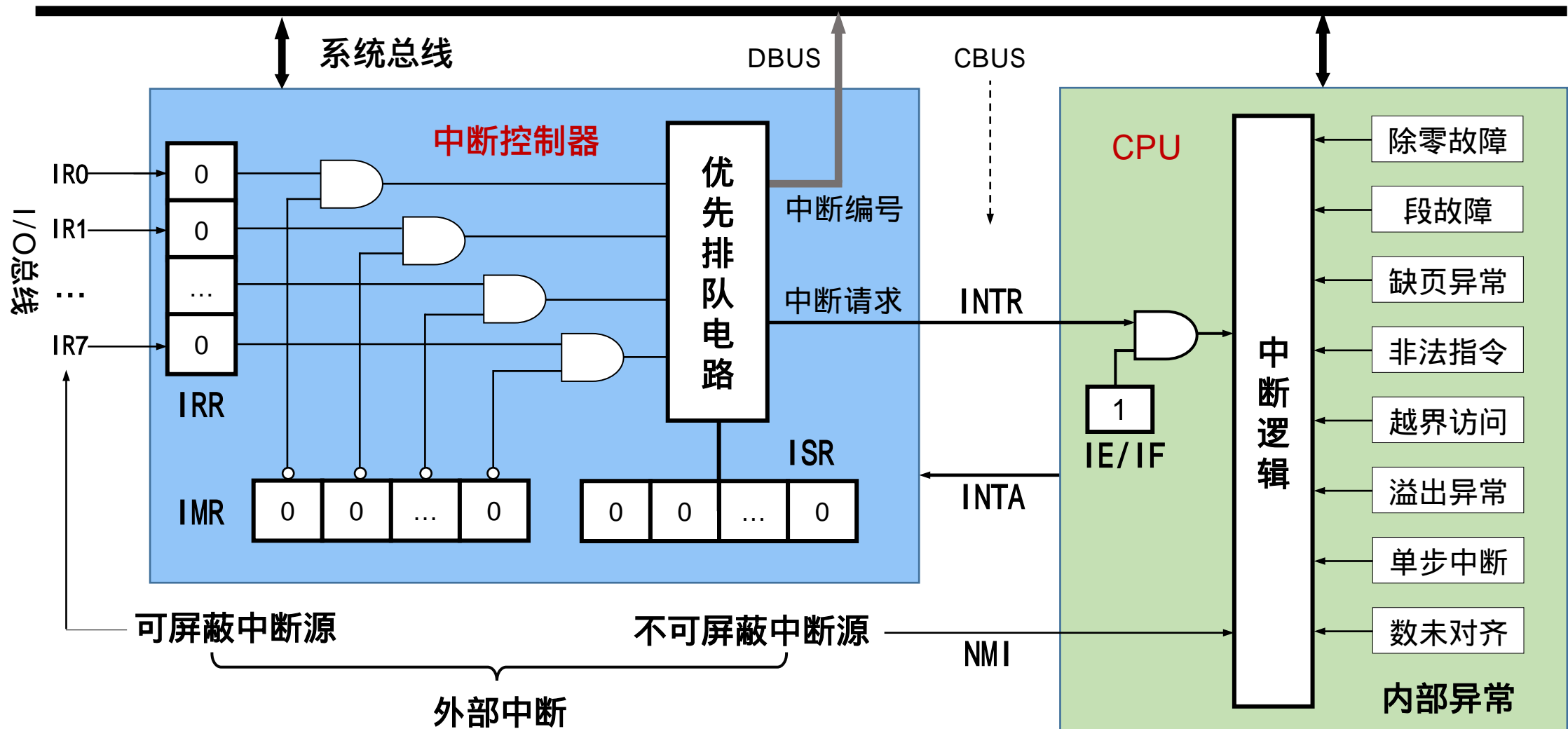
## n 中断过程由软硬件结合完成



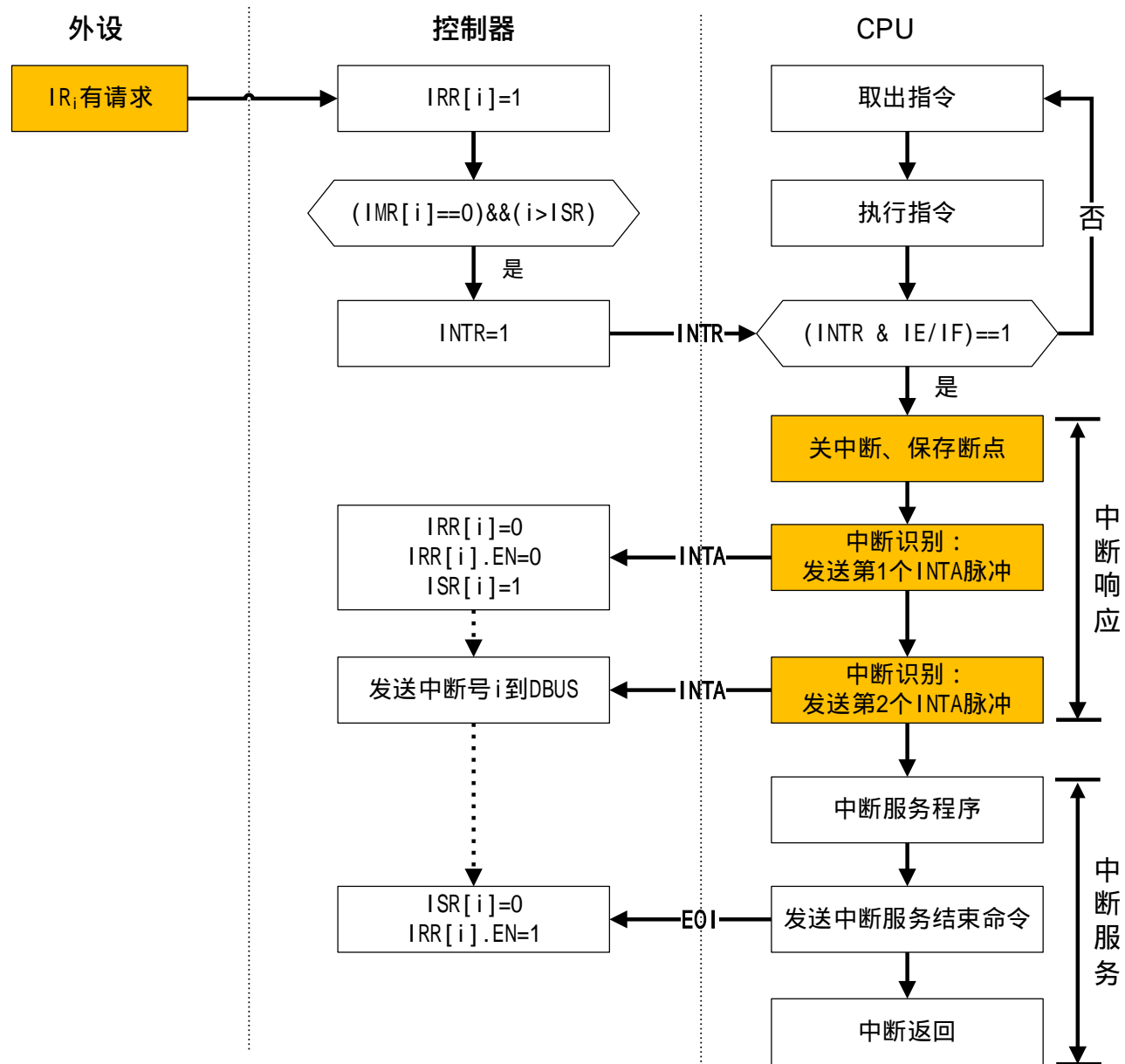
(c) 中断控制器方式



# 中断硬件接口



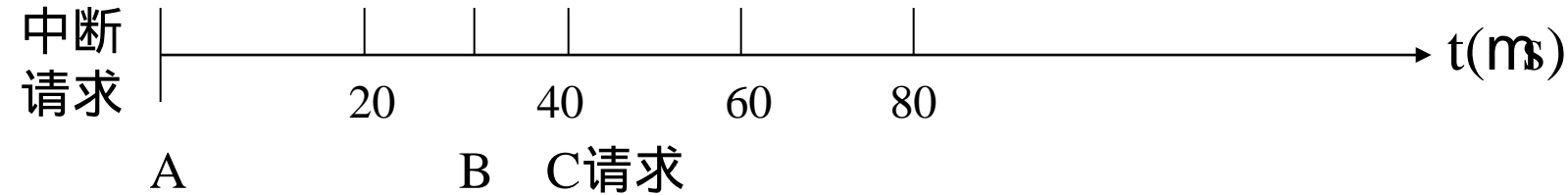
# 中断控制器工作流程



13.A、B、C是与主机连接的3台设备，在硬件排队线路中，它们的响应优先级是 $A > B > C > \text{CPU}$ ，为改变中断处理的次序，将它们的中断屏蔽字设为：

| 设备  | 屏蔽码 |   |   |
|-----|-----|---|---|
|     | A   | B | C |
| A   | 1   | 1 | 1 |
| B   | 0   | 1 | 0 |
| C   | 0   | 1 | 1 |
| CPU | 0   | 0 | 0 |

请按下图所示时间轴给出的设备中断请求时刻，画出CPU执行程序的轨迹。A、B、C中断服务程序的时间宽度均为20 ms。

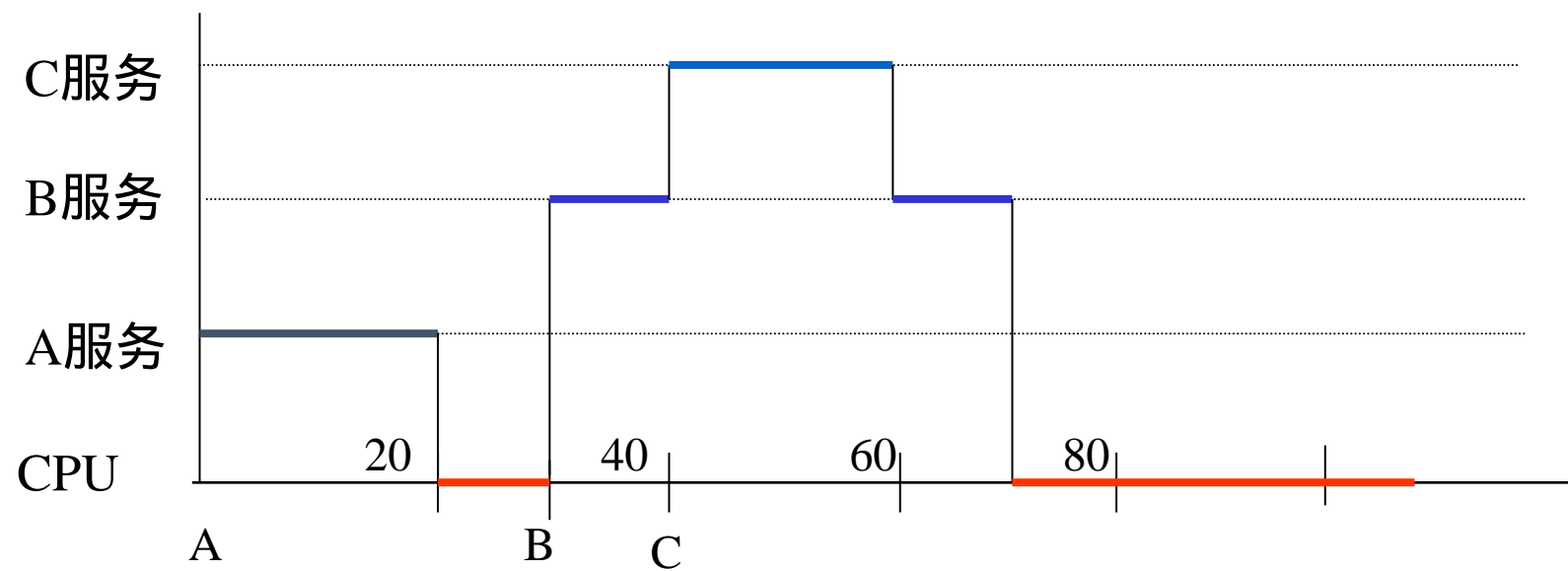




解：从中断屏蔽字看出，其处理优先级为：

$A > C > B > \text{CPU}$

故CPU运行轨迹如下：



## || 本章主要内容

- n 9.1 输入输出设备与特性
- n 9.2 I/O接口
- n 9.3 数据传送控制方式
- n 9.4 程序控制方式
- n 9.5 程序中断方式
- n 9.6 DMA方式**
- n 9.7 通道方式
- n 9.8 常见I/O设备



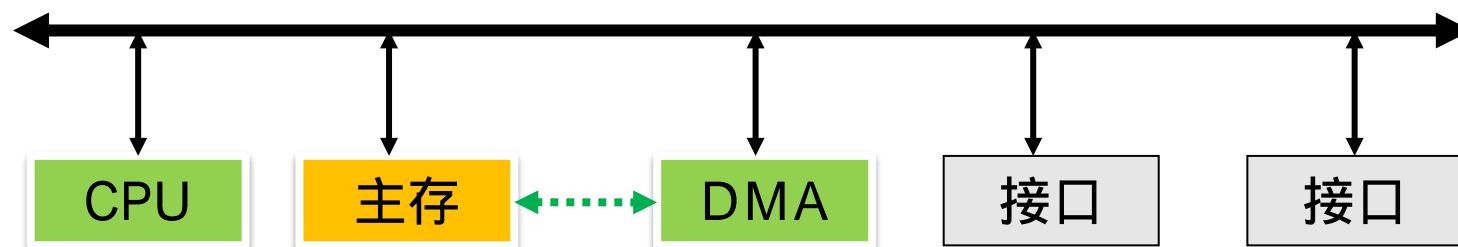
# DMA基本概念

## n 中断方式

- p 传送一个数据执行一次中断服务子程序（几十条指令）
- p 效率低下，不适合于高速传输的系统

## n DMA方式

- p 外设与主存间建立一个由DMA硬件管理的数据通路（虚拟通路，还是通过系统总线）
- p CPU不介入外设与主存的数据传送操作
- p 减少CPU开销，提升效率



# || DMA方式

- n DMA基本概念
- n DMA传输方式
- n 基本DMA控制器

# 内存争用

## n 访存冲突

- p DMA控制器直接访问内存
- p CPU执行主程序（需要访内）

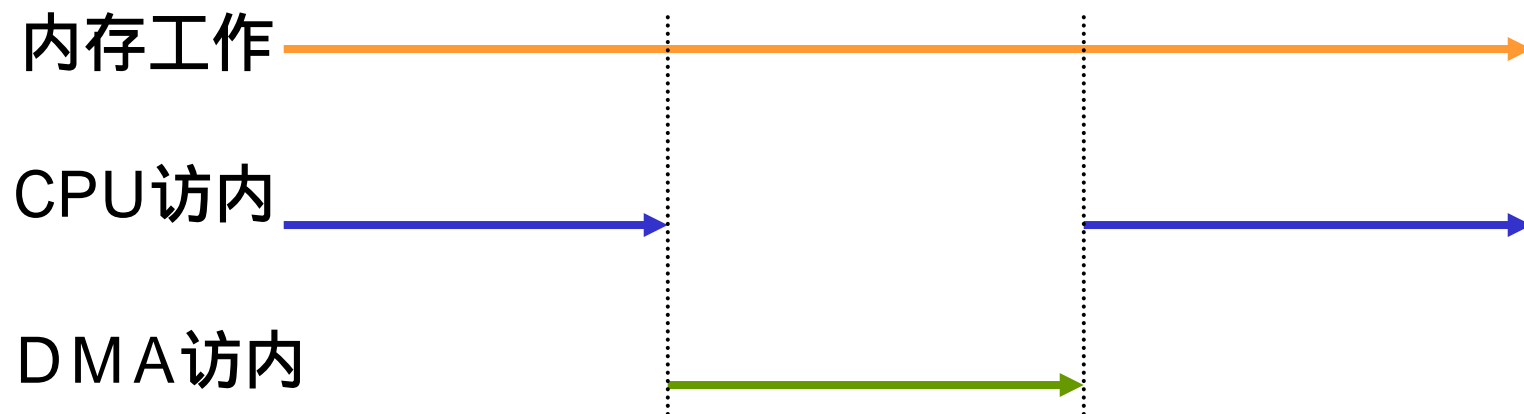
## n 如何处理这种冲突？

- p 停止CPU使用主存
- p DMA与CPU交替使用主存
- p 周期挪用法

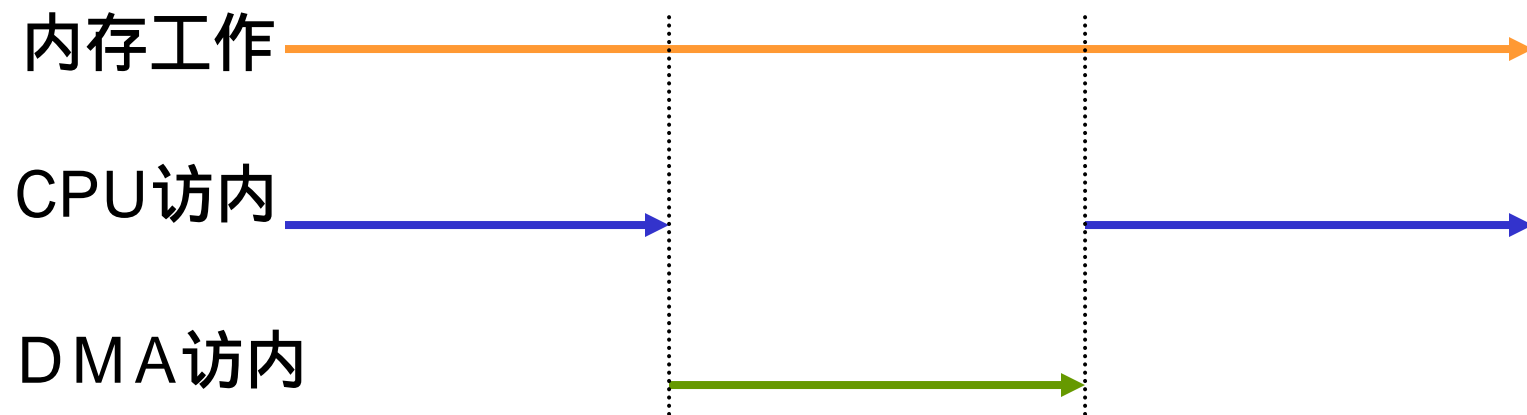


## || 停止CPU使用主存

- n DMA传送数据时，CPU停止使用主存
- n 一批数据传送结束后，DMA再交还主存使用权
- n DMA传送过程中，CPU处于等待状态



## 停止CPU访内

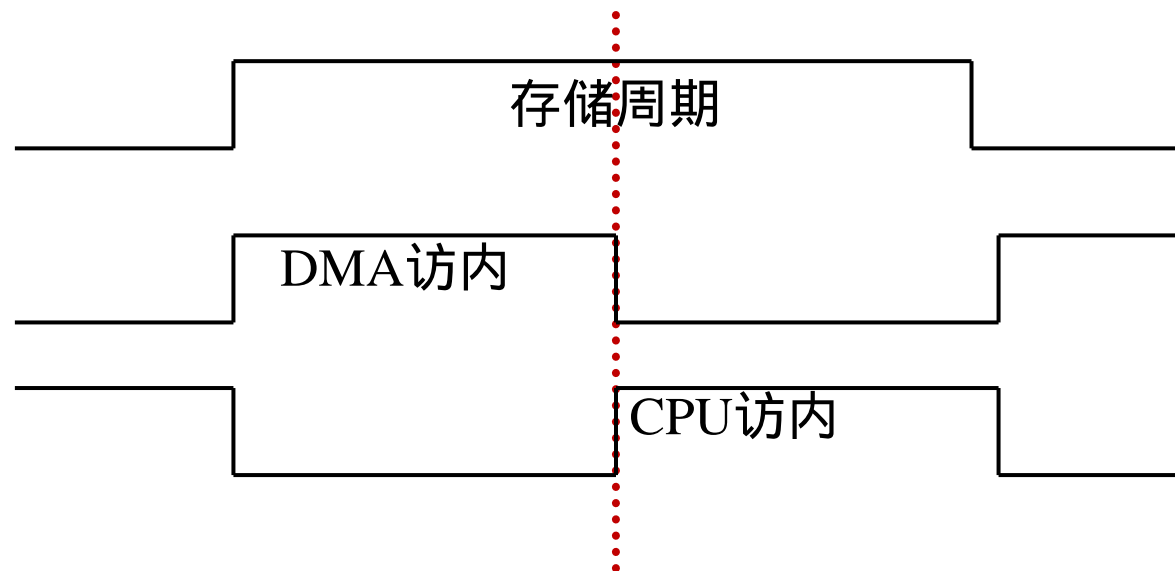


1. DMA批量数据传输周期过长，CPU长期无法访内
2. 外设传送两个数据的时间间隔大于存储周期，内存未充分利用



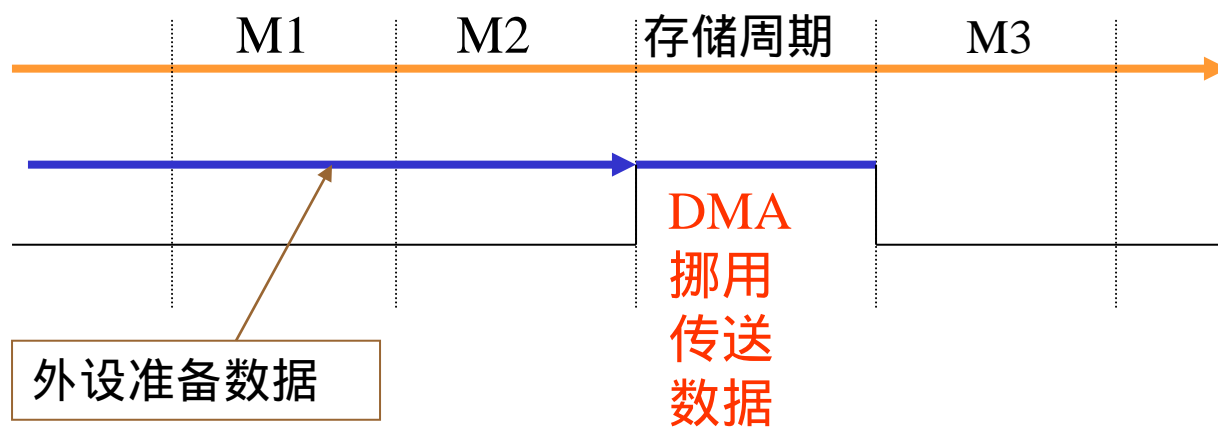
# DMA与CPU交替使用主存

- n 每个存储周期分成两段
  - p 一段用于DMA访问主存
  - p 一段用于CPU访问主存
- n 无主存使用权移交过程

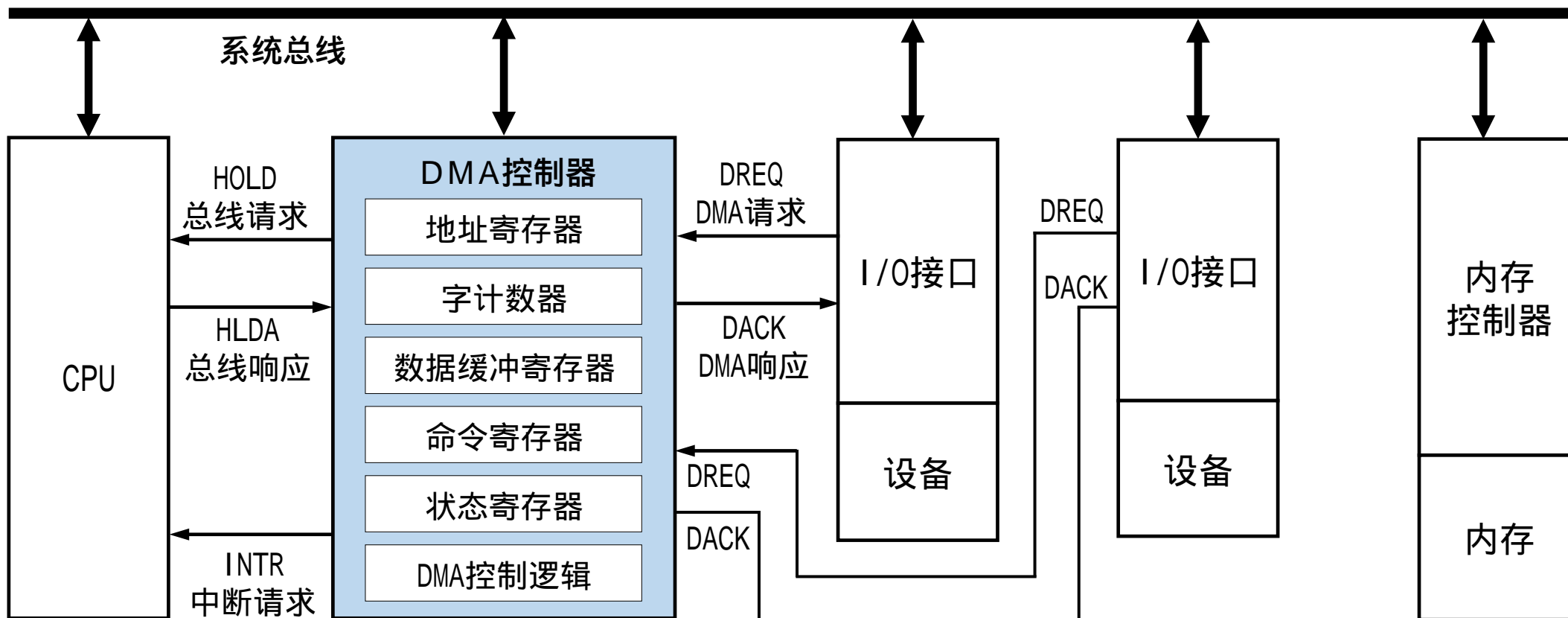


## 周期挪用法

- n DMA要求访问主存时，CPU暂停一个或多个存储周期。一个数据传送结束后，CPU继续运行。
- n CPU现场没有变动，仅延缓了指令的执行
  - p 周期挪用，或称周期窃取。
- n 如发生访存冲突，则DMA优先访问。



# DMA控制器



# DMA传输流程

n 准备阶段

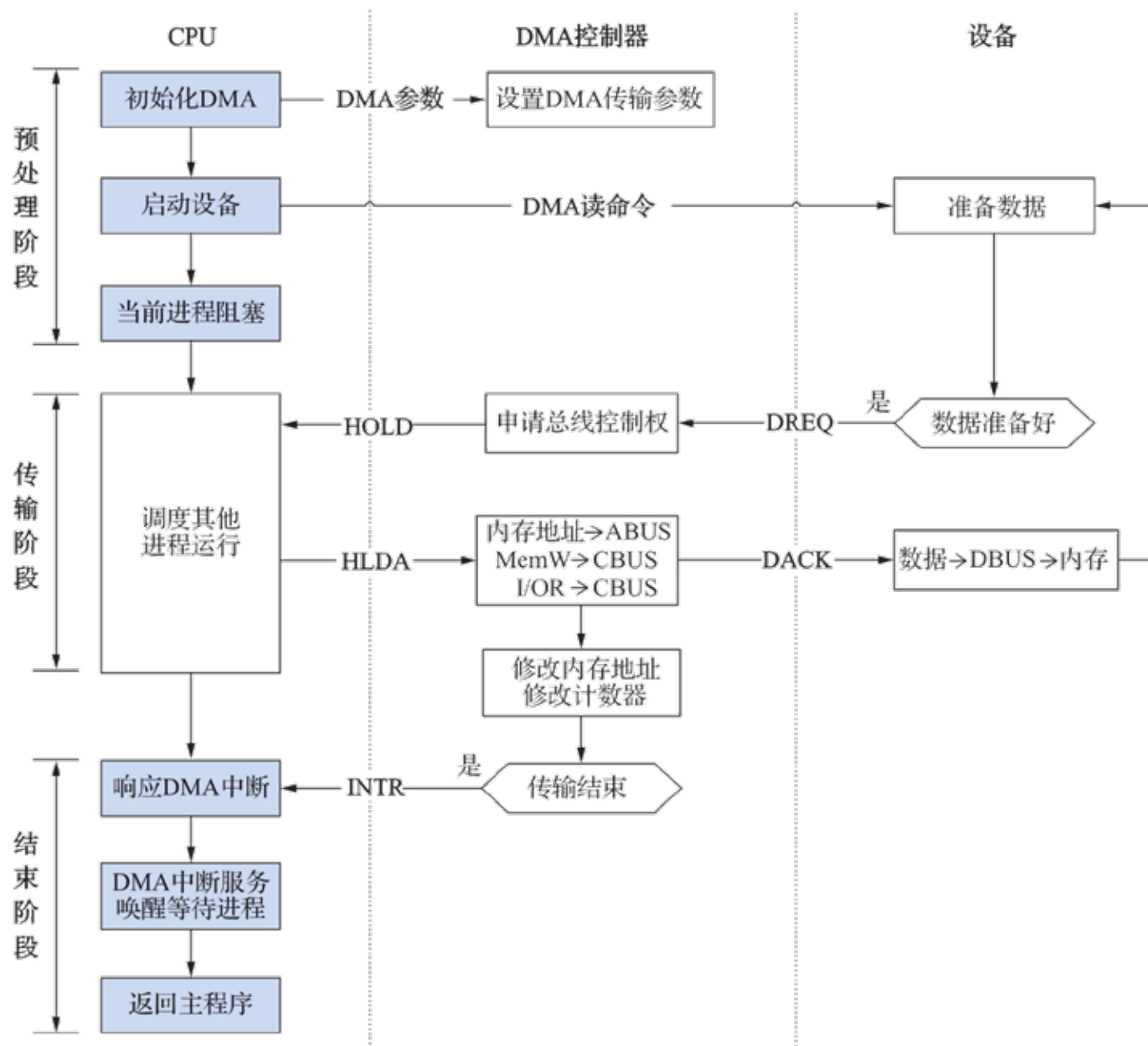
p CPU干预

n 传送阶段

p CPU不参与

n 结束阶段

p CPU干预



## || DMA主要操作过程（准备阶段）

- n 主机通过CPU指令向DMA接口发送必要的传送参数，并启动DMA工作。
  1. 数据传送的方向
  2. 数据块在主存的首地址
  3. 数据在外设存储介质上的地址
  4. 数据的传送量

## || DMA主要操作过程（传送阶段）

- n 宏观 DMA是连续传送一批数据    微观 每传送一个数据，发一次DMA请求
- n 传输过程（周期挪用）
  1. 设备准备数据：当设备接收到 CPU 的 DMA 命令后就可以开始准备数据
  2. 设备发送 DMA 请求：数据准备好后就通过 DREQ 控制线向 DMAC 发出 DMA 请求。
  3. DMAC 申请总线：DMAC 收到 DMA 请求后立即将 HOLD 信号置“1”，向 CPU 申请总线控制权。
  4. 总线授权：CPU 在机器周期结束后响应总线使用申请，让出总线控制权，并发出总线授权信号 HLDA 通知 DMAC。



## || DMA主要操作过程（传送阶段）

5. DMA 数据传输：收到 HLDA 信号将内存地址放置在地址总线上；设置控制总线读写命令控制信号，并向设备DMA 应答信号 DACK。设备收到 DACK 信号后会和内存完成一个机器字的数据交换
6. 传输控制：设备传输完一次数据后会继续重复第 1 步到第 5 步的工作，DMAC 在每次传输时还需要负责维护内存地址和传输计数器，并撤除 HOLD信号释放总线
7. 数据传输结束时，DMAC 会通过 INTR 信号线发送一个 EOP (End Of Process) 的 DMA 中断请求信号，告知 CPU 传输完成

## || DMA主要操作过程（结束阶段）

- n DMA在两种情况下都进入结束阶段。
  - p 正常结束，一批数据传送完毕
  - p 非正常结束，DMA故障
- n 结束阶段DMA向主机发出中断请求
- n CPU执行中断服务程序
  - p 查询DMA接口状态，根据状态进行不同处理

## || DMA与程序中中断的区别

- n 中断通过**程序传送数据**，DMA靠**硬件来实现**。
- n 中断时机为**两指令之间**，DMA响应时机为**两存储周期之间**。
- n 中断不仅具有数据传送能力，还能处理异常事件。DMA只能进行数据传送。
- n DMA仅挪用了**一个存储周期**，不改变CPU现场。
- n DMA请求的**优先权比中断请求高**。CPU优先响应DMA请求，是为了避免DMA所连接的高速外设丢失数据。
- n DMA利用了中断技术

## 例题

- n 例 9.5 某磁盘采用 DMA 方式与 CPU 交换信息，其传输速率为 20MB/s。若 DMA 的预处理阶段需要 200 个时钟周期，DMA 完成传输后的中断处理阶段需要 400 个时钟周期。如果 DMA 平均传输的数据块长度为 512B，问：磁盘工作时，200MHz 的处理器进行 DMA 传输时的 CPU 占用率是多少？如果采用 4KB 的传输单位呢？（不考虑 DMAC 与 CPU 争用内存对 CPU 性能的影响）

每秒需要执行的 DMA 次数为  $20\text{MB}/512\text{B} = 39062.5$  次。

每次 DMA 传输所需要的 CPU 时间为  $(200+400)T = 600T$ 。

则 DMA 传输所占用 CPU 时间的比例为  $600T \times 39062.5 / 1\text{s} = 11.72\%$ 。

如果采用 4KB 的大块传输，则传输次数降低为原来的 1/8。

DMA 操作占用 CPU 时间的比例为  $600T \times (20\text{MB}/4\text{KB}) / 1\text{s} = 1.46\%$ 。

## 例题

n 某计算机CPU主频500MHz，CPI为5。假定某外设的数据传输率为0.5MB/s,采用中断方式与主机进行数据传送，以32位为传输单位，对应的中断服务程序包含18条指令，中断服务的其他开销相当于2条指令的执行时间。

(1) 在中断方式下，CPU用于外设I/O的时间占整个CPU时间的百分比是多少？

传输32bit，需一次中断，

所需CPU开销 $T_{I/O} = (18+2) \times CPI \times T = 20 \times 5 / 500\text{MHz}$

传输32bit，需要的总时间 $T_{\text{total}} = 32 / 8 / 0.5\text{MB/s}$

CPU用于外设I/O时间占整个CPU时间比例 =  $T_{I/O} / T_{\text{total}} = 2.5\%$

## 例题

- n 某计算机CPU主频500MHz，CPI为5。假定某外设的数据传输率为0.5MB/s,采用中断方式与主机进行数据传送，以32位为传输单位，对应的中断服务程序包含18条指令，中断服务的其他开销相当于2条指令的执行时间。
- n 当外设的数传率为5MB/s时，改用DMA方式。假定DMA传送块大小为5000B，且DMA预处理和后处理的总开销为500个时钟周期，则CPU用于该外设I/O的时间占整个CPU时间的百分比是多少（假定DMA与CPU之间没有访存冲突）

DMA传输阶段不需要占用CPU时间。

传输5000B，需一次DMA，所需CPU开销 $T_{IO}=500 \times T=500/500\text{MHz}$

传输5000B 需要的总时间 $T_{total}=5000/5\text{MB/s}$

CPU用于外设I/O时间占整个CPU时间比例=  $T_{IO}/T_{total}=0.1\%$





华中科技大学  
计算机科学与技术学院  
School of Computer Science & Technology, HUST

THANKS

计算机组成原理

