

尚德机构

# 计算机系统结构

讲师：孙小涵

学习是一种信仰！ IN LEARNING WE TRUST

SUNLAND





# 讲师介绍

- 主讲老师：孙小涵（尚德机构-小涵老师）
- 主讲课程：计算机类、数学类
- 邮箱：sunxiaohan@sunlands.com



# 课程章节

## 计算机系统结构

第1章 计算机系统结构概论

第2章 数据表示、寻址方式与指令系统

第3章 存储、中断、总线与I/O系统

第4章 存储体系

第5章 标量处理机

第6章 向量处理机

第7章 多处理机

第8章 数据流计算机和归约机

## 第2章 数据表示、寻址方式与指令系统

## 第2章 数据表示、寻址方式与指令系统

●	数据表示	★ ★
●	寻址方式	★
●	指令系统的设计和优化	★
●	指令系统的发展和改进	★ ★

## 2.4指令系统的发展和改进

### 2.4.3按RISC方向发展和改进指令系统

#### 1.CISC的问题（简答）

- 1) 指令系统**庞大**，一般指令在200条以上。
- 2) 许多指令的操作**繁杂**，执行速度很低，甚至不如用几条简单、基本的指令组合实现。
- 3) 由于指令系统庞大，使高级语言编译程序选择**目标指令的范围太大**，因此，难以优化生成高效机器语言程序，编译程序也**太长、太复杂**。
- 4) 由于指令系统庞大，各种指令的使用频度都不会太高，且**差别很大**，其中相当一部分指令的**利用率很低**。

## 2.4指令系统的发展和改进

### 2.4.3按RISC方向发展和改进指令系统

#### 2.设计RISC的基本原则（简答）

- 1) 确定指令系统时，只选择使用频度很高的那些指令，再增加少量能有效支持操作系统、高级语言实现及其他功能的指令，大大减少指令条数，一般使之不超过100条。
- 2) 减少指令系统所用寻址方式种类，一般不超过两种。简化指令的格式限制在两种之内，并让全部指令都是相同长度。
- 3) 让所有指令都在一个机器周期内完成。
- 4) 扩大通用寄存器数，一般不少于32个，尽量减少访存，所有指令只有存（STORE）、取（LOAD）指令访存，其他指令一律只对寄存器操作。
- 5) 为提高指令执行速度，大多数指令都用硬联控制实现，少数指令才用微程序实现。
- 6) 通过精简指令和优化设计编译程序，简单、有效地支持高级语言的实现。

## 2.4指令系统的发展和改进

### 2.4.3按RISC方向发展和改进指令系统

3.设计RISC结构采用的基本技术（简答）

- 1) 按设计RISC的一般原则来设计。
- 2) 逻辑实现采用硬联和微程序相结合。
- 3) 在CPU中设置大量工作寄存器并采用重叠寄存器窗口。
- 4) 指令用流水和延迟转移。
- 5) 采用高速缓冲存储器Cache,设置指令Cache和数据Cache分别存放指令和数据。
- 6) 优化设计编译系统。



补充：（考察频率较低，答案不唯一）

指令	频度pi	扩展操作码			
I1	0.4	0	0		
I2	0.3	0	1		
I3	0.15	1	0		
I4	0.05	1	1	0	0
I5	0.04	1	1	0	1
I6	0.03	1	1	1	0
I7	0.03	1	1	1	1

## 第3章 存储、中断、总线与I/O系统

## 第2章 数据表示、寻址方式与指令系统

●	存储系统的基本要求和并行主存系统	★
●	中断系统	★★
●	总线系统	★
●	I/O系统	★★

## 3.1 存储系统的基本要求和并行主存系统

**本节主要内容：**

**并行主存系统的组成形式，极限频宽和实际频宽的计算  
使用并行主存的组成技术提高主存实际频宽的可能性、和  
发展存储体系的必要性**

## 3.1 存储系统的基本要求和并行主存系统

### 3.1.1 存储系统的基本要求（填空）

对存储系统的基本要求是**大容量**、**高速度**和**低价格**。

在存储器所用期间一定的条件下，**容量越大**，因其延迟增大会使**速度降低**；容量越大，存储器总**价格会越大**；存储器**速度越快**，**价格也越高**。

## 3.1 存储系统的基本要求和并行主存系统

### 3.1.2 并行主存系统（综合）

能并行读出多个CPU字的**单体多字和多体单字、多体多字**的**交叉访问**主存系统被称为**并行主存系统**。

主存最大频宽  $B_m = m \times W / T_M$

## 3.1 存储系统的基本要求和并行主存系统

### 3.1.2 并行主存系统

【1610真题】设主存采用模 $m$ 多分体交叉存取，每个分体的存取周期为 $T_M = 2\mu s$ ，要求主存实际频宽为 $8MB/s$ ，但实际频宽只能达到最大频宽的 $0.6$ 倍。

1. 若分体宽度 $W = 4$ 字节，则主存模数应取多少才能满足要求? ( $m$ 取2的幂)
2. 若主存模数为 $8$ ，则分体宽度应为多少才能满足要求?

## 3.1 存储系统的基本要求和并行主存系统

### 3.1.2 并行主存系统

【1610真题】设主存采用模 $m$ 多分体交叉存取，每个分体的存取周期为 $T_M = 2\mu s$ ，要求主存实际频宽为 $8MB/s$ ，但实际频宽只能达到最大频宽的 $0.6$ 倍。

1. 若分体宽度 $W = 4$ 字节，则主存模数应取多少才能满足要求？( $m$ 取 $2$ 的幂)

主存最大频宽 $B_m = m \times W / T_M$

$0.6 \times m \times 4 / 2 \geq 8$ ；解得 $m \geq 6.667$

所以：主存模数应取 $8$ 才能满足要求。



## 3.1 存储系统的基本要求和并行主存系统

### 3.1.2 并行主存系统

【1610真题】设主存采用模 $m$ 多分体交叉存取，每个分体的存取周期为 $T_M=2\mu s$ ，要求主存实际频宽为 $8MB/s$ ，但实际频宽只能达到最大频宽的 $0.6$ 倍。

2. 若主存模数为 $8$ ，则分体宽度应为多少才能满足要求？

$$0.6 \times 8 \times W/2 \geq 8; \text{ 解得 } W \geq 3.333$$

所以：分体宽度应取 $4$ 字节。

## 3.1 存储系统的基本要求和并行主存系统

### 3.1.2 并行主存系统（综合）

能并行读出多个CPU字的**单体多字和多体单字、多体多字**的**交叉访问**主存系统被称为**并行主存系统**。

主存最大频宽

$$B_m = m \times W / T_M$$

每个存储周期能访问到的平均字数

$$B = \frac{1 - (1 - \lambda)^m}{\lambda}$$

## 3.1 存储系统的基本要求和并行主存系统

### 3.1.2 并行主存系统

$$B = \frac{1 - (1 - \lambda)^m}{\lambda}$$

【1710真题】程序存放在模32单字交叉存储器中，设访存申请队的转移概率 $\lambda = 25\%$ ，

1. 求每个存储周期能访问到的平均字数。
2. 当模为16呢？由此可得到什么结论？

## 3.1 存储系统的基本要求和并行主存系统

### 3.1.2 并行主存系统

$$B = \frac{1 - (1 - \lambda)^m}{\lambda}$$

【1710真题】程序存放在模32单字交叉存储器中，设访存申请队的转移概率 $\lambda = 25\%$ ，

1. 求每个存储周期能访问到的平均字数。
2. 当模为16呢？由此可得到什么结论？

当  $m = 32$  时，将  $m = 32, \lambda = 25\%$  代入上式，则

$$B = \frac{1 - (1 - \lambda)^m}{\lambda} = \frac{1 - (1 - 25\%)^{32}}{25\%} \approx 4$$

即平均每个周期能访问到 4 个字。

## 3.1 存储系统的基本要求和并行主存系统

### 3.1.2 并行主存系统

$$B = \frac{1 - (1 - \lambda)^m}{\lambda}$$

【1710真题】程序存放在模32单字交叉存储器中，设访存申请队的转移概率 $\lambda = 25\%$ ，

1. 求每个存储周期能访问到的平均字数。4
2. 当模为16呢？由此可得到什么结论？

当  $m = 16$  时，将  $m = 16, \lambda = 25\%$  代入上式，则

$$B = \frac{1 - (1 - \lambda)^m}{\lambda} = \frac{1 - (1 - 25\%)^{16}}{25\%} \approx 3.96$$

即每个存储周期平均能访问到 3.96 个字。

## 3.2 中断系统

**本节主要内容：**

**中断分类和分机的目的（简答）**

**设置中断级屏蔽位的作用及中断嵌套的原则**

**按中断处理要求的次序设置中断级屏蔽位状态**

**发生中断请求时，CPU程序执行状态的转切过程**

**中断系统软硬件功能分配状况**

## 3.2 中断系统

### 什么是中断系统（单选、填空）

响应和处理各种中断的软、硬件总体称为中断系统。

在计算机中，中断可分为内部中断、外部中断和软件中断三类。

中断系统是整个计算机系统不可缺少的重要组成部分。

## 3.2.1 中断的分类和分级

### 1. 中断的分类（单选）

IBM 370系统就将中断分成机器校验、管理程序调用、程序性、外部、输入/输出和重新启动6类。





## 3.2.1 中断的分类和分级

### 1. 中断的分类（单选）

机器校验中断是告诉程序发生了设备故障。

可用64位机器校验中断码指明故障原因和严重性

包含有电源故障、运算电路的误动作、主存出错、通道动作故障、处理器的各种硬件故障等。

## 3.2.1 中断的分类和分级

### 1. 中断的分类（单选）

访管中断是在用户程序需要操作系统介入时，通过执行“**访管**”指令时发生的，访管原因由“访管”指令中的**8**位码指明。



## 3.2.1 中断的分类和分级

### 1. 中断的分类（单选）

程序性中断是包括指令和数据的格式错、程序执行中出现异常（非法指令、目态下使用管态指令、主存访问方式保护、寻址超过主存容量、各种溢出、除数为0、有效位为0等）以及程序的事件、监督程序对事件的检测引起的中断等。

## 3.2.1 中断的分类和分级

### 1. 中断的分类（单选）

外部中断来自计算机外部，它包括各种**定时器中断**、**外部信号中断**及**中断键中断**。

各种定时器中断用以计时、计费、控制等；外部信号中断主要用于与其他计算机和系统的联系；中断键则用于操作员对计算机的干预。这些外部中断又可分为两类：一类是若未被响应，则继续保留；另一类是如不响应，则不再保留。

## 3.2.1 中断的分类和分级

### 1. 中断的分类

输入/输出中断是CPU与I/O设备及通道联系的工具，在输入/输出操作完成或者I/O通道或者设备产生故障时发出。



## 3.2.1 中断的分类和分级

### 1. 中断的分类（简单了解）

重新启动中断是为操作员或另一台CPU要启动一个程序所用。  
CPU不能禁止这种中断。



### 3.2.1 中断的分类和分级

## 2. 中断的分级（单选、简答）

中断系统按中断源的级别高低来响应。

通常优先级最高的中断定为第1级，其次是第2级，再次是第3级.....

通常机器校验为第1级，程序性和管理程序调用为第2级，外部为第3级，输入/输出为第4级，重新启动为最低级。（顺序）

## 3.2.2 中断的响应次序与处理次序

### 1. 中断的响应次序（单选）

中断的响应次序是在同时发生多个不同中断类的中断请求时，中断响应硬件中的**排队器**所决定的响应次序。





### 3.2.2 中断的响应次序与处理次序

#### 2. 中断的处理次序（综合）

中断的处理要由中断处理程序来完成，而中断处理程序在执行前或执行中是可以被中断的



## 3.2.2 中断的响应次序与处理次序

### 2. 中断的处理次序（综合）

【1404真题】机器有5级中断，中断响应次序为 $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$ ，现要求实际中断处理次序为 $2 \rightarrow 3 \rightarrow 1 \rightarrow 5 \rightarrow 4$ 。

1. 设计各级中断处理程序的中断级屏蔽位的状态，令“0”为开放，“1”为屏蔽；
2. 若运行用户程序时，同时发生1、3级中断请求，而在1级中断服务未完成时，又发生2、3、4、5级中断，请画出处理机执行程序全过程的示意图（标出交换PSW的时间）。

## 3.2.2 中断的响应次序与处理次序

### 2. 中断的处理次序（综合）

【1404真题】机器有5级中断，中断响应次序为 $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$ ，现要求实际中断处理次序为 $2 \rightarrow 3 \rightarrow 1 \rightarrow 5 \rightarrow 4$ 。

1. 设计各级中断处理程序的中断级屏蔽位的状态，令“0”为开放，“1”为屏蔽；

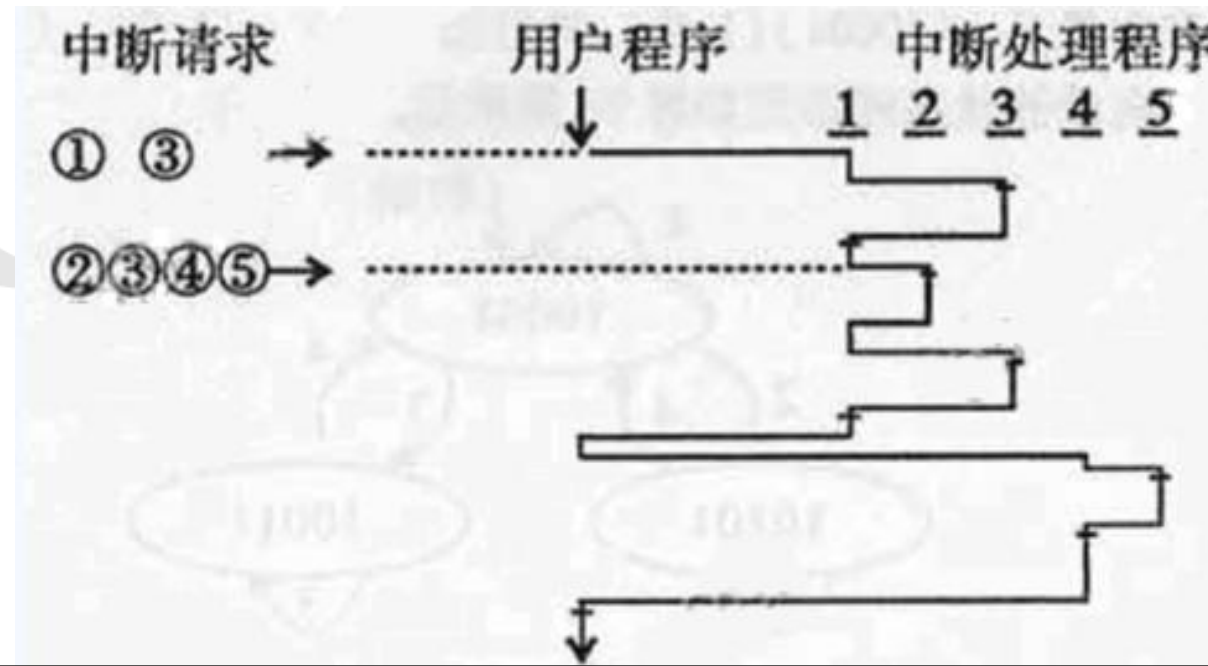
中断处理 程序级别	中断级屏蔽位				
	1 级	2 级	3 级	4 级	5 级
第 1 级	1	0	0	1	1
第 2 级	1	1	1	1	1
第 3 级	1	0	1	1	1
第 4 级	0	0	0	1	0
第 5 级	0	0	0	1	1

## 3.2.2 中断的响应次序与处理次序

### 2. 中断的处理次序（综合）

【1404真题】机器有5级中断，中断响应次序为 $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$ ，现要求实际中断处理次序为 $2 \rightarrow 3 \rightarrow 1 \rightarrow 5 \rightarrow 4$ 。

2. 若运行用户程序时，同时发生1、3级中断请求，而在1级中断服务未完成时，又发生2、3、4、5级中断，请画出处理机执行程序全过程的示意图（标出交换PSW的时间）。



### 3.2.2中断的响应次序与处理次序

## 2.中断的处理次序（综合）

【1510真题】设中断级屏蔽位“1”对应于开放，“0”对应于屏蔽，各级中断处理程序的中断级屏蔽位设置如题表所示。（10分）

1. 当中断响应优先次序为1→2→3→4时，其中断处理次序是什么？

中断处理程序级别	中断级屏蔽位			
	第 1 级	第 2 级	第 3 级	第 4 级
第 1 级	0	0	0	0
第 2 级	1	0	1	1
第 3 级	1	0	0	1
第 4 级	1	0	0	0

### 3.2.2中断的响应次序与处理次序

## 2.中断的处理次序（综合）

【1510真题】设中断级屏蔽位“1”对应于开放，“0”对应于屏蔽，各级中断处理程序的中断级屏蔽位设置如题表所示。（10分）

1. 当中断响应优先次序为1→2→3→4时，其中断处理次序是什么？

中断处理顺序为1→4→3→2

中断处理程序级别	中断级屏蔽位			
	第 1 级	第 2 级	第 3 级	第 4 级
第 1 级	0	0	0	0
第 2 级	1	0	1	1
第 3 级	1	0	0	1
第 4 级	1	0	0	0

### 3.2.2 中断的响应次序与处理次序

## 2. 中断的处理次序（综合）

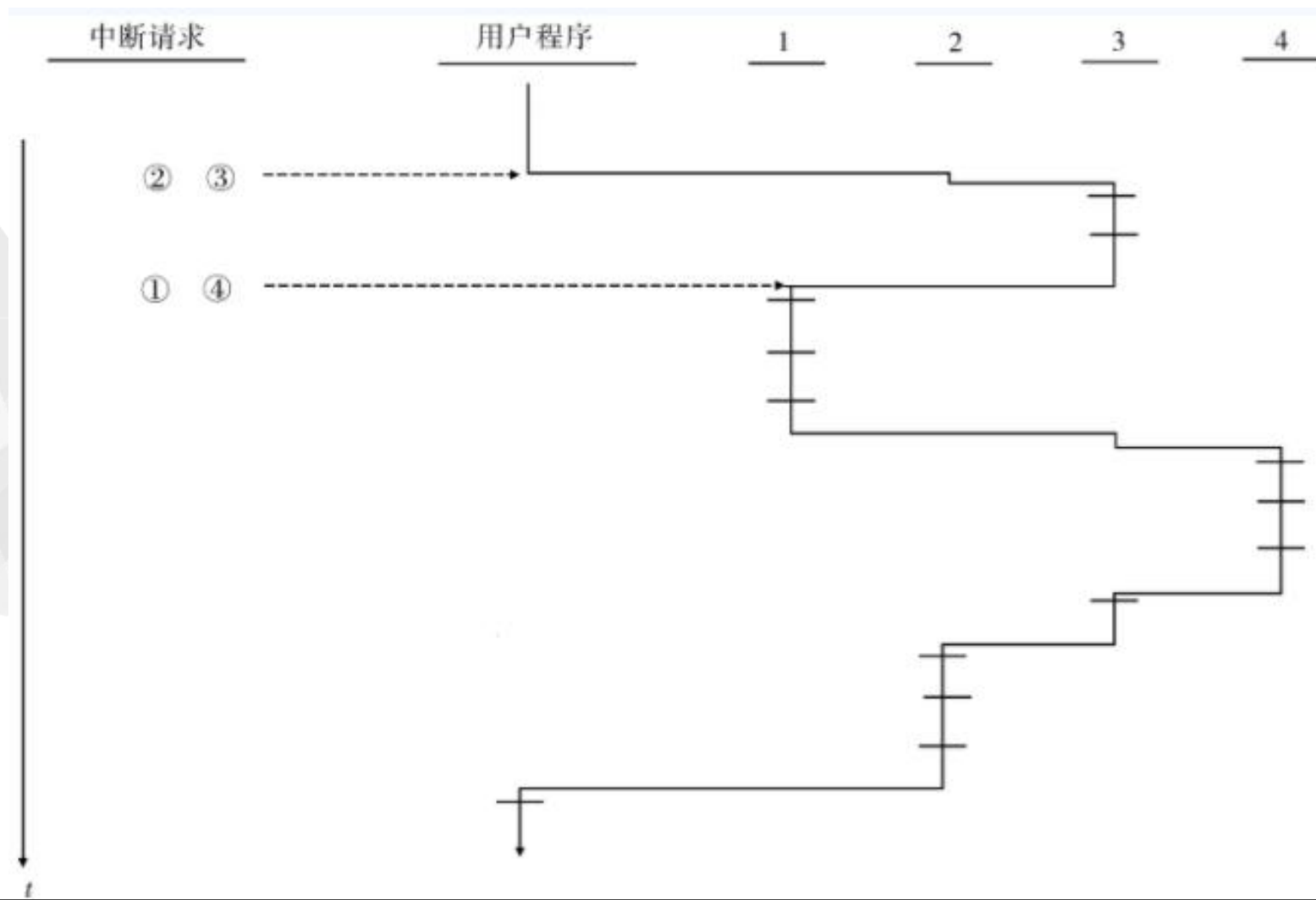
【1510真题】设中断级屏蔽位“1”对应于开放，“0”对应于屏蔽，各级中断处理程序的中断级屏蔽位设置如题表所示。（10分）

中断处理顺序为1→4→3→2

如果所有的中断处理各需3个单位时间，中断响应和中断返回时间相对中断处理时间少得多。当计算机正在运行用户程序时，同时发生第2、3级中断请求，过两个单位时间后，又同时发生1、4级中断请求，试画出程序运行过程示意图。

## 3.2.2 中断的响应次序与处理次序

### 2. 中断的处理次序（综合）







那么意气风发地  
走在成功的道路上

## 真题练练手

1、实现中断响应次序的器件称为 ( ) 1704

A:计数器

B:排队器

C:定时器

D:触发器



那么意气风发地  
走在成功的道路上

## 真题练练手

1、实现中断响应次序的器件称为 ( ) 1704

A:计数器

**B:排队器**

C:定时器

D:触发器

答案： B

### 3.2.3 中断系统的软、硬件功能分配

#### 1. 中断系统的功能分配（填空）

中断系统的软、硬件功能的实质是中断**处理程序软件**和中断**响应硬件**的功能分配。



### 3.2.3 中断系统的软、硬件功能分配

## 2. 中断系统的功能（简答）

中断请求的保存和清除、优先级的确定、中断断点及现场的保存、对中断请求的分析和处理以及中断返回等。  
中断现场包括软件状态和硬件状态。

## 3.3 总线系统

**本节主要内容：**

**专用和非专业总线的定义、优缺点及适用场合**

**非专业总线中3种总线控制方式**

**同步和异步通信控制方式**

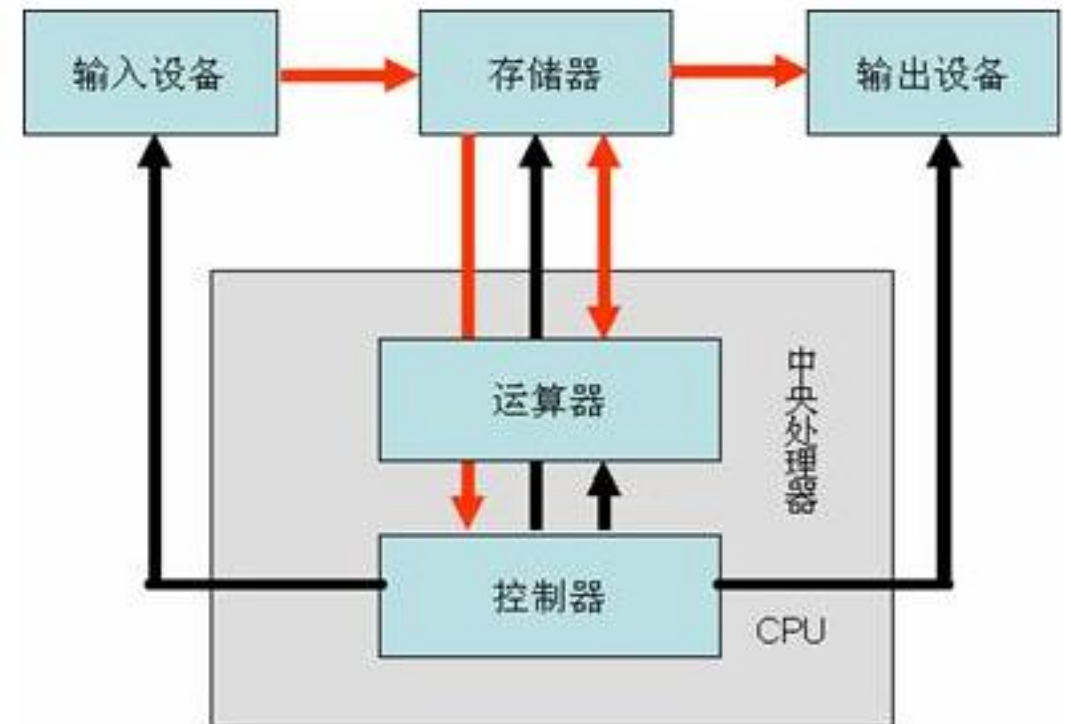
**数据宽度的定义、分类及优缺点**

**数据宽度与数据通路宽度的不同**

## 3.3总线系统

### 1.什么是总线（简单了解）

总线是用于互连计算机、CPU、存储器、I/O接口及外围设备、远程通信设备间信息传送通路的集合。



## 3.3总线系统

### 2.什么是总线系统（简单了解）

总线与其相配合的附属控制电路统称为总线系统。

按信息传送功能、性能的不同，有**数据线**，**地址线**，命令、时序和中断信号等**控制/状态线**，电源线，地线及备用线等。

## 3.3总线系统

### 3.什么是总线的分类（单选）

总线按在系统中的位置分 芯片级（CPU芯片内的总线）、板级（连接插件板内的各个组件，也称局部总线或内部总线）和系统级（系统间或主机与I/O接口或设备之间的总线）等3级。



## 3.3总线系统

### 3.什么是总线的分类（填空）

就总线允许信息传送的方向来说，可以有**单向传输**和**双向传输**两种。

双向传输又有**半双向**和**全双向**的不同。

## 3.3总线系统

### 3.什么是总线的分类（填空、简答）

总线按用法可分为**专用**和**非专用**两类。

只连接一对**物理部件**的总线称专用总线。

其优点是

- 1.多个部件可以同时收/发信息，不争用总线，**系统流量高**；
- 2.通信时不用指明源和目的，**控制简单**；
- 3.任何总线的失效只会使连于该总线的两个部件不能直接通信，但它们仍可通过其他部件间接通信，因而**系统可靠**。

专用总线的缺点是**总线数多**。

## 3.3总线系统

### 4.什么是总线的控制

集中式总线的控制（单选、填空）

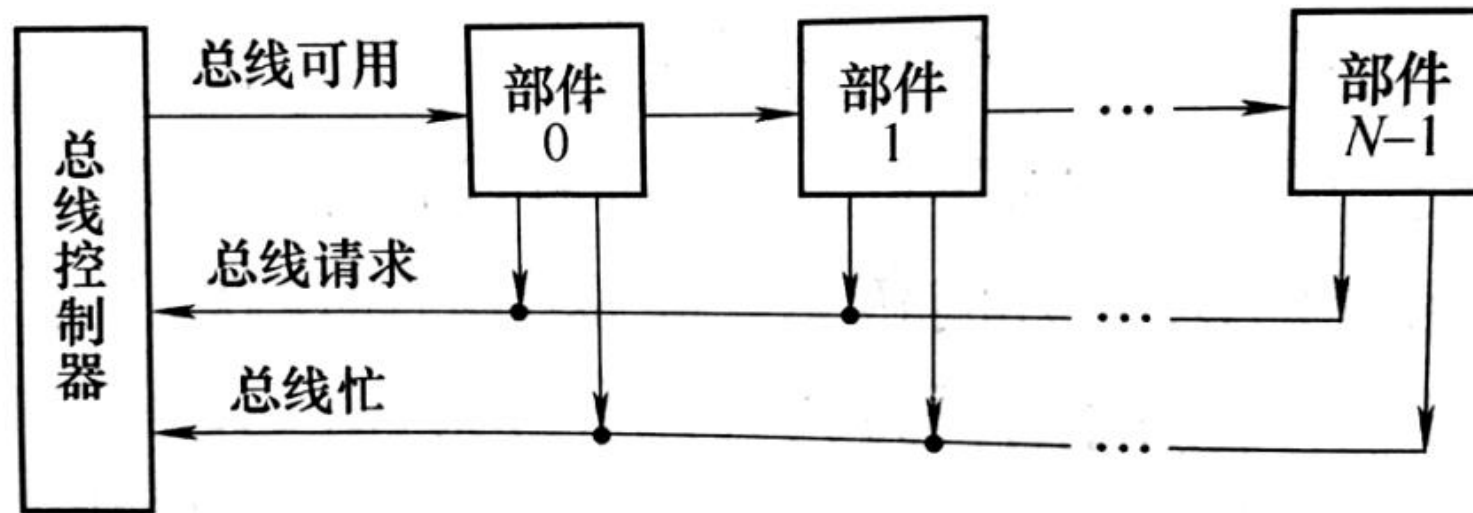
优先次序的确定可以有串行链接、定时查询和独立请求3种不同的方式，也可以是它们的结合。

采用何种方式取决于控制线数目、总线分配速度、灵活性、可靠性等因素的综合权衡。

## 3.3 总线系统

### 4. 什么是总线的控制

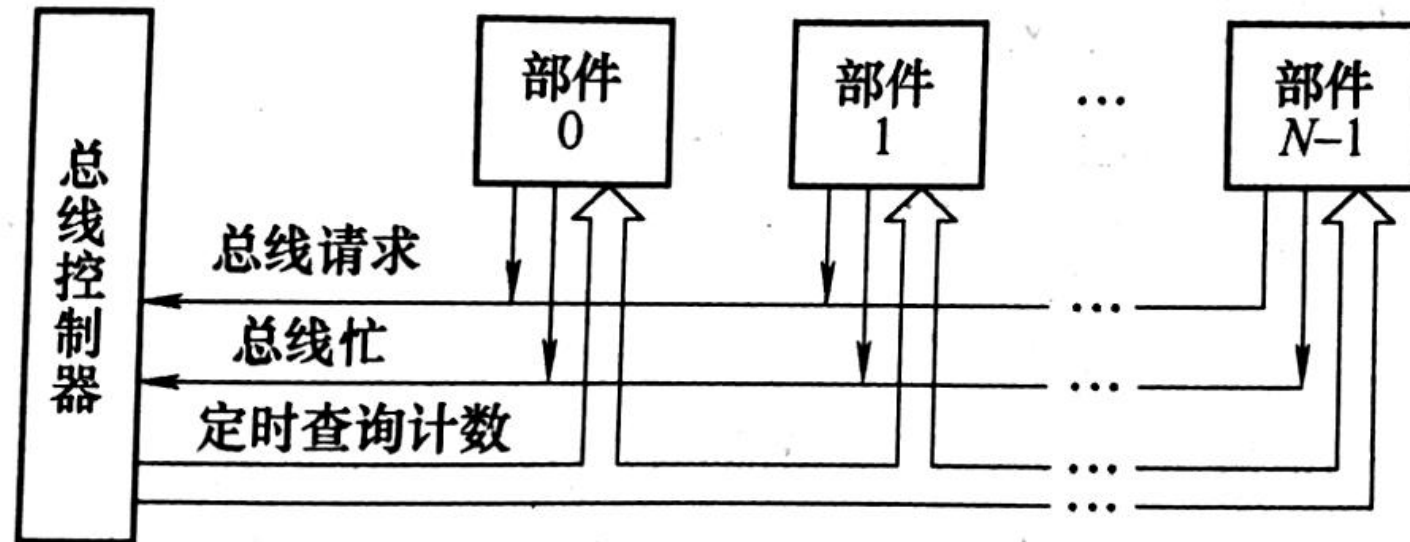
串行链接方式获得使用总线权的优先次序是由“总线可用”线所接部件的物理位置来决定的，离总线控制器越近的部件其优先级越高。



## 3.3 总线系统

### 4. 什么是总线的控制

集中式定时查询方式：



## 3.3总线系统

### 4.什么是总线的控制

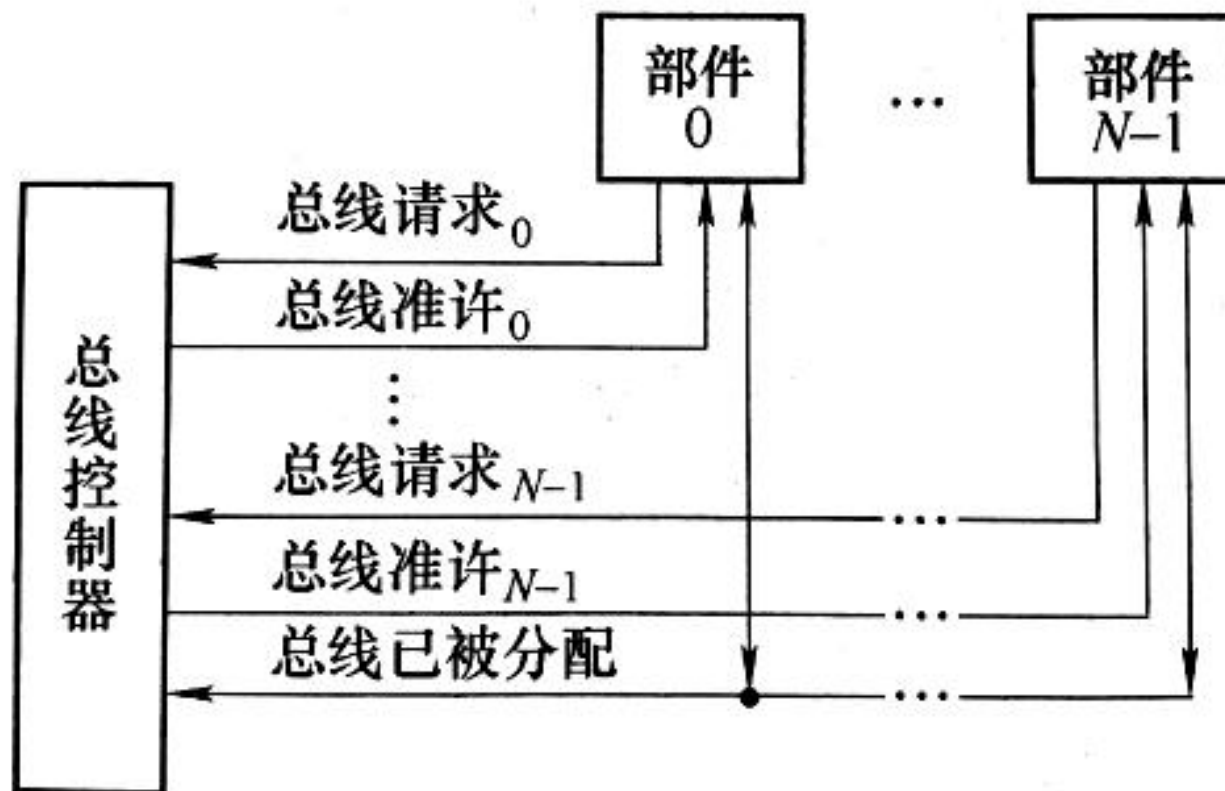
定时查询的缺点（单选）

是控制线的线数较多，需 $2 + \lceil \log_2 N \rceil$ 根；可以共享总线的部件数受限于定时查询线的线数（编址能力），扩展性稍差；控制较为复杂；总线分配的速度取决于计数信号的频率和部件数，不能很高。

## 3.3 总线系统

### 4. 什么是总线的控制

集中式独立请求方式。



## 3.3总线系统

### 4.什么是总线的控制（优缺点：简答题，课下自己背背）

集中式串行链接

集中式定时查询

集中式独立请求方式。





那么意气风发地  
走在成功的道路上

## 真题练练手

1、总线按在系统中的位置分的级别数为 ( ) 1904

A:3级

B:4级

C:5级

D:6级



那么意气风发地  
走在成功的道路上

## 真题练练手

1、总线按在系统中的位置分的级别数为 ( ) 1904

A:3级

B:4级

C:5级

D:6级

答案： A



那么意气风发地  
走在成功的道路上

## 真题练练手

2、关于非专用总线三种控制方式中，下列叙述错误的是（ ） 0704

- A:集中式定时查询，所有部件共用同一条“总线忙”线
- B:集中式定时查询，所有部件都用同一条“总线请求”线
- C:集中式独立请求，所有部件都用同一条“总线请求”线
- D:集中式串行链接，所有部件都用同一条“总线请求”线



那么意气风发地  
走在成功的道路上

## 真题练练手

2、关于非专用总线三种控制方式中，下列叙述错误的是（ ） 0704

A:集中式定时查询，所有部件共用同一条“总线忙”线

B:集中式定时查询，所有部件都用同一条“总线请求”线

C:集中式独立请求，所有部件都用同一条“总线请求”线

D:集中式串行链接，所有部件都用同一条“总线请求”线

答案：C



那么意气风发地  
走在成功的道路上

## 真题练练手

3、总线控制定时查询方式的控制线的线数为 ( ) 1604 0604

A:  $\lceil \log_2 N \rceil$

B:  $1 + \lceil \log_2 N \rceil$

C:  $2 + \lceil \log_2 N \rceil$

D:  $3 + \lceil \log_2 N \rceil$



那么意气风发地  
走在成功的道路上

## 真题练练手

3、总线控制定时查询方式的控制线的线数为 ( ) 1604 0604

A:  $\lceil \log_2 N \rceil$

B:  $1 + \lceil \log_2 N \rceil$

C:  $2 + \lceil \log_2 N \rceil$

D:  $3 + \lceil \log_2 N \rceil$

答案: C



那么意气风发地  
走在成功的道路上

## 真题练练手

4、列选项中，不属于集中式总线控制方式的是（） 1610

A:环形链接

B:串行链接

C:定时查询

D:独立请求



那么意气风发地  
走在成功的道路上

## 真题练练手

4、列选项中，不属于集中式总线控制方式的是（） 1610

A:环形链接

B:串行链接

C:定时查询

D:独立请求

答案: A



### 3.3.3总线的通信技术（填空）

1.同步通信

2.异步通信



### 3.3.4数据宽度与总线线数

#### 1.数据宽度（单选、填空）

数据宽度是I/O设备取得I/O总线后所传送数据的总量。

数据宽度有单字（单字节）、定长块、可变长块、单字加定长块和单字加可变长块等之分。

单字（单字节）宽度适合于输入机、打印机等低速设备。

定长块宽度适合于磁盘等高速设备，可以充分利用总线带宽。

可变长块宽度适合于高优先级的中高速磁带、磁盘等设备，

对于挂有速度较低而优先级较高的设备的总线，可以采用单字加定长块传送。

采用单字加可变长块的传送，是一种灵活有效却复杂、开销大的方法。

### 3.3.4数据宽度与总线线数

## 2.总线线数（单选、填空）

总线要有发送/接收电路、传输导线或电缆、转接插头和电源等，在满足性能前提下应尽量减少线数。总线线数可通过采用线的功能组合、编码和并/串一串/并转换来减少。

总线标准一般包括机械、功能、电气及过程（同步）等4个方面的标准

## 3.4 I/O系统

本节主要内容：

输入/输出系统的基本概念

I/O系统的3种方式

I/O处理机的两种方式

通道处理机的输入/输出过程

通道的3种类型

通道的极限流量计算，外设对通道要求的流量计算

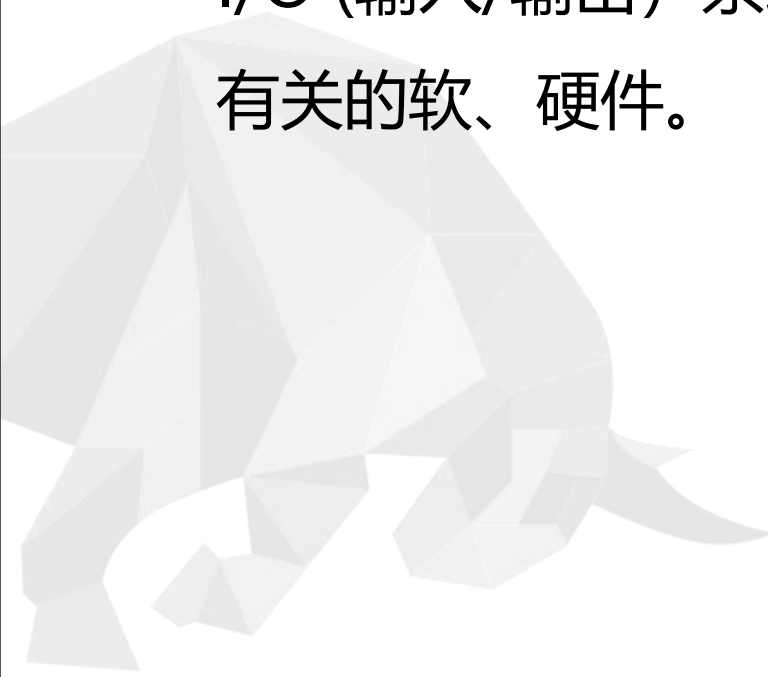
I/O系统给的流量计算

字节多路通道的流量计算、通道工作周期设计、通道响应和处理各台外设请求时刻的时空图

## 3.4. 1 I/O系统概述

### 1.什么是I/O系统（简单了解）

I/O（输入/输出）系统包括输入/输出设备、设备控制器及与输入/输出操作有关的软、硬件。



## 3.4. 1 I/O系统概述

### 2.I/O系统的发展（单选、填空）

输入/输出系统的发展经历了3个阶段，相对应于3种方式，即程序控制I/O（包括全软件的、程序查询的、中断驱动的）、直接存储器访问（DMA）及I/O处理机方式。它们可分别用于不同的计算机系统，也可用于同一系统。

## 3.4. 1 I/O系统概述

### 3.输入/输出设备的分类（简单了解）

输入/输出设备分**外存**和**传输设备**两大类。外存有磁盘、磁带、光盘等。传输设备有键盘、鼠标、光笔、显示器、各种打印/印字机、声音输入/输出设备、图形扫描器、网络驱动器等。

### 3. 4. 2通道处理机的工作原理和流量设计

#### 1.通道处理机的工作原理

启动I/O指令是主要的输入/输出指令，属于管态指令





### 3. 4. 2通道处理机的工作原理和流量设计

#### 1.通道处理机的工作原理（单选、填空）

根据通道数据传送期中信息传送方式的不同，可分为**字节多路**、**数组多路**和**选择**3类通道。

字节多路通道适用于连接大量的像**光电机**等字符类低速设备。

数组多路通道适合于连接多台**磁盘**等高速设备。

选择通道适合于连接优先级高的**磁盘**等高速设备，

### 3. 4. 2通道处理机的工作原理和流量设计

#### 2.通道流量的设计

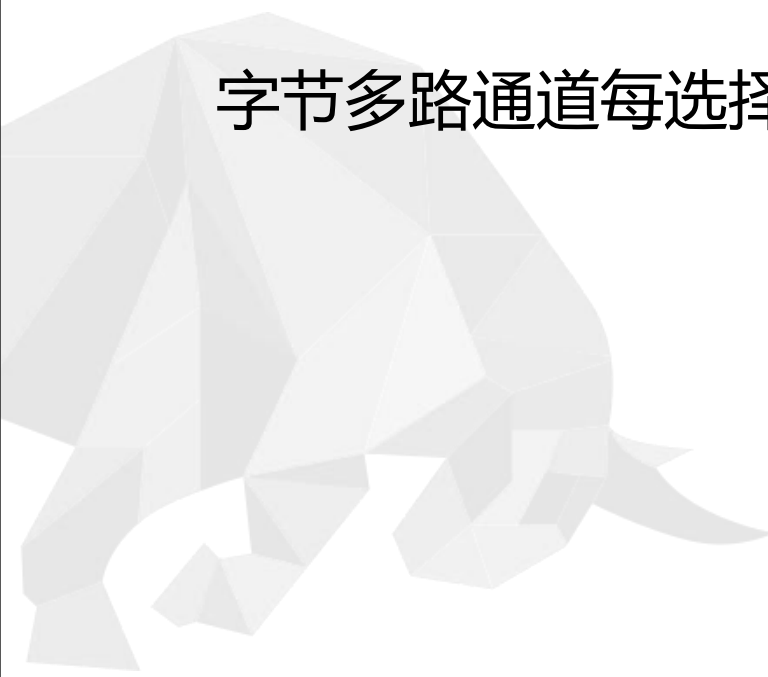
通道流量是通道在数据传送期内，单位时间内传送的字节数。它能达到的最大流量称通道极限流量。



### 3. 4. 2通道处理机的工作原理和流量设计

#### 2.通道流量的设计（综合）

字节多路通道每选择一台设备只传送一个字节，其通道极限流量


$$f_{\max \bullet byte} = \frac{1}{T_S + T_D}$$

### 3. 4. 2通道处理机的工作原理和流量设计

#### 2.通道流量的设计（综合）

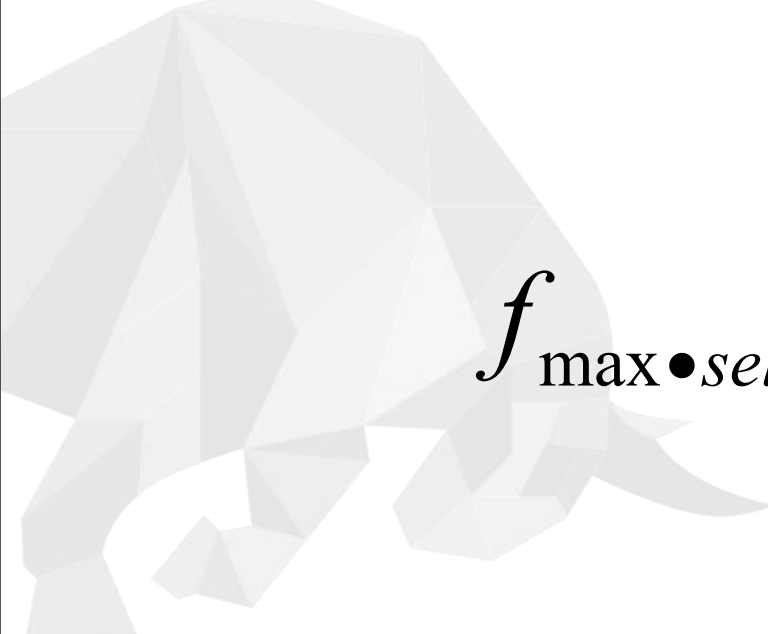
数组多路通道每选择一台设备可传送完K个字节。如果要传送N个字节，就得分「N/K」次传送才行，每次传送都要选一次设备，通道极限流量

$$f_{\max \bullet block} = \frac{K}{T_S + KT_D} = \frac{1}{\frac{T_S}{K} + T_D}$$

### 3. 4. 2通道处理机的工作原理和流量设计

#### 2.通道流量的设计

选择通道每选择一台设备就把N个字节全部传送完，通道极限流量


$$f_{\max \bullet select} = \frac{N}{T_S + NT_D} = \frac{1}{\frac{T_S}{N} + T_D}$$

### 3. 4. 2通道处理机的工作原理和流量设计

#### 2.通道流量的设计

【1604真题】如果通道在数据传送期中,选择设备的时间 $T_s$ 为 $10\mu s$ ,传送一个字节数据需要的时间 $T_d$ 为 $0.5\mu s$ 。

1. 某低速设备每隔 $500\mu s$ 发出一个字节数据请求,至多可接几台这种设备

### 3. 4. 2通道处理机的工作原理和流量设计

#### 2.通道流量的设计

低速设备每隔  $500\mu s$  发出一个字节数据传送请求,挂低速设备的通道应按字节多路通道方式工作,极限流量:

$$f_{\max \cdot \text{byte}} = 1/(T_s + T_D)$$

如果所挂的台数为  $m$ ,设备的速率  $f_i$  实际就是设备发出字节传送请求的间隔时间的倒数,相同设备,速率之和为  $mf_i$ 。为使不丢失信息,应满足

$$\frac{1}{T_s + T_D} \geq mf_i$$

于是  $m$  应满足:

$$m \leq \frac{1}{(T_s + T_D) \cdot f_i} = \frac{500}{(10 + 0.5)} = 48 \text{ (台)}$$

所以至多可挂 48 台设备。

### 3. 4. 2通道处理机的工作原理和流量设计

#### 2.通道流量的设计

【1604真题】如果通道在数据传送期中,选择设备的时间 $T_s$ 为 $10\mu s$ ,传送一个字节数据需要的时间 $T_d$ 为 $0.5\mu s$ 。

2. 对于如表所示的低速设备,一次通信传送的字节数不少于1024个字节,则哪些设备可挂?哪些设备不可挂?

A ~ F 设备的速率

设备	A	B	C	D	E	F
设备速率 $f_i / B \cdot \mu s^{-1}$	1/0.1	1/0.6	1/0.2	1/0.25	1/0.8	1/0.55



### 3. 4. 2通道处理机的工作原理和流量设计

#### 2.通道流量的设计

如 表所示为低速设备,此通道是选择通道,如果通道上挂有  $m$  台设备,则选择通道的极限流量为:

$$f_{\max-\text{select}} = \frac{n}{(T_s + nT_D)} = \frac{1}{\frac{T_s}{n} + T_D} = \frac{1}{\frac{10}{n} + 0.5}$$

限制通道上所挂的设备速率

$$f \leq \frac{1}{\frac{10}{n} + 0.5} B \cdot \mu s^{-1}, n \geq 1024$$

$$\text{即: } f_i \leq 1.96B \cdot \mu s^{-1}$$

根据 表可知,只能挂 B、E、F 设备, A、C、D 设备因为速率超过  $f_{\max-\text{select}}$ , 所以不能挂。



那么意气风发地  
走在成功的道路上

## 真题练练手

1、字节多路通道适用于连接 ( ) 1704

A:大量低速设备

B:少量低速设备

C:大量高速设备

D:少量高速设备



那么意气风发地  
走在成功的道路上

## 真题练练手

1、字节多路通道适用于连接 ( ) 1704

A:大量低速设备

B:少量低速设备

C:大量高速设备

D:少量高速设备

答案：A



那么意气风发地  
走在成功的道路上

## 真题练练手

2、根据通道数据传送期中信息传送方式的不同,通道可分为 ( ) 、 ( ) 和数组多路三类通道。1410



那么意气风发地  
走在成功的道路上

## 真题练练手

2、根据通道数据传送期中信息传送方式的不同,通道可分为 ( ) 、 ( ) 和数组多路三类通道。1410

答案：字节多路 选择



那么意气风发地  
走在成功的道路上

## 真题练练手

3、输入 / 输出系统的3种控制方式是（    ）、直接存储器访问及（    ）。

1510



那么意气风发地  
走在成功的道路上

## 真题练练手

3、输入 / 输出系统的3种控制方式是（    ）、直接存储器访问及（    ）。

1510

答案：程序控制I/O    I/O处理机方式

## 第4章 存储体系



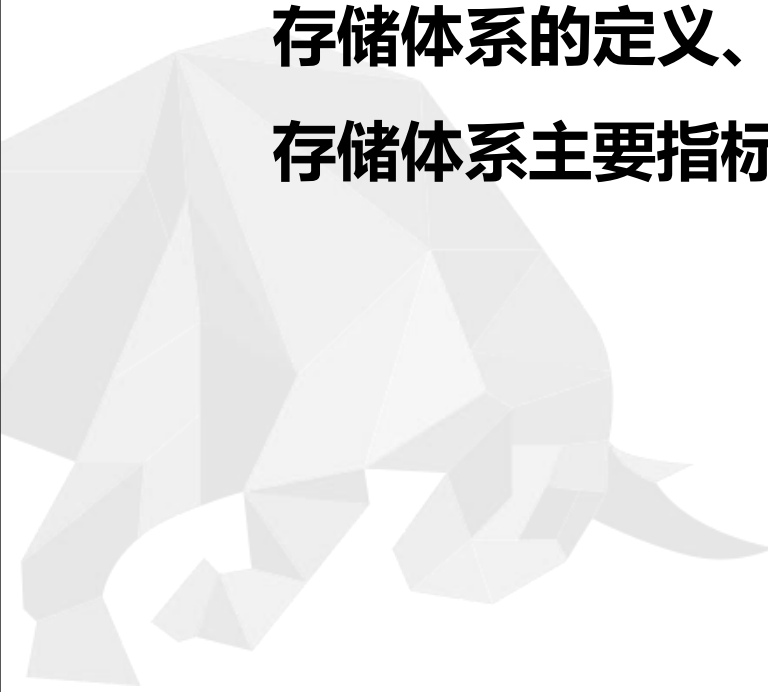
# 第4章 存储体系

●	基本概念	★	☆
●	虚拟存储器	★	☆
●	高速缓冲存储器	★	★
●	三级存储体系	★	

## 4.1 基本概念

**本节主要内容：**

**存储体系的定义、分支及依据  
存储体系主要指标参数的计算**



## 4.1.1 存储体系及其分支

### 1. 存储体系（简单了解）

存储体系（即存储层次）是让构成存储系统的几种不同的存储器（ $M_1 \sim M_n$ ）之间，配上辅助软、硬件或辅助硬件，使之从应用程序员角度来看，它们在逻辑上是一个整体。

## 4.1.1 存储体系及其分支

### 2. 虚拟存储器（单选、填空）

虚拟存储器是因主存容量满足不了要求而提出来的。在主存和辅存之间，增设辅助的软、硬件设备，让它们构成一个整体，所以也称为主存—辅存存储层次，

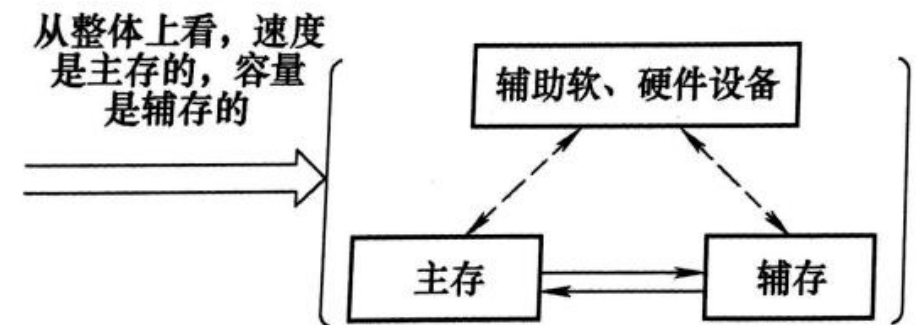


图 4-1 主存—辅存存储层次

## 4.1.1 存储体系及其分支

### 3.cache存储器（单选、填空）

因主存**速度**满足不了要求而引出了Cache存储器。在CPU和主存之间增设高速、小容量、每位价格较高的Cache，用辅助硬件将Cache和主存构成整体，

Cache存储器不仅对**应用程序员**是透明的，  
而且对**系统程序员**也是透明的。

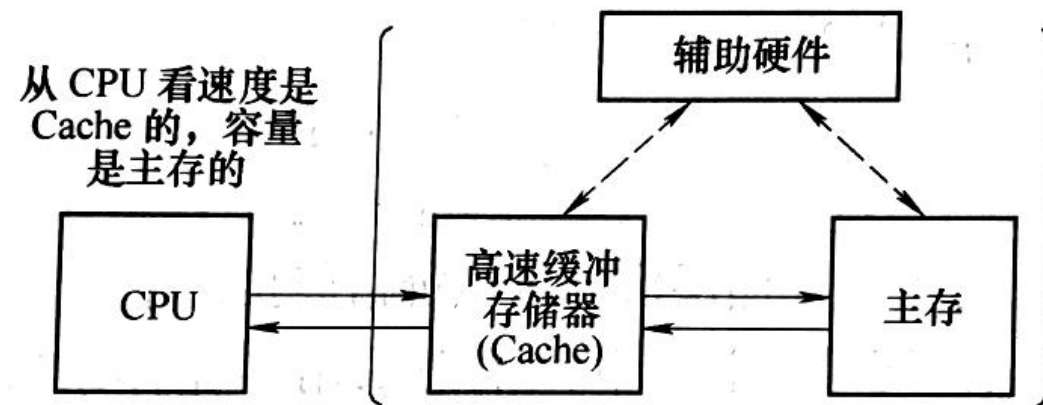


图 4-2 Cache—主存存储层次

## 4.1.1 存储体系及其分支

### 4. 多级存储层次（简单了解原理）

由二级存储层次可组合成如图4-3所示的多级存储层次。希望从CPU角度看是一个整体，有接近于最高层M1的速度、最低层Mn的容量，并有接近于最低层Mn的每位价格。

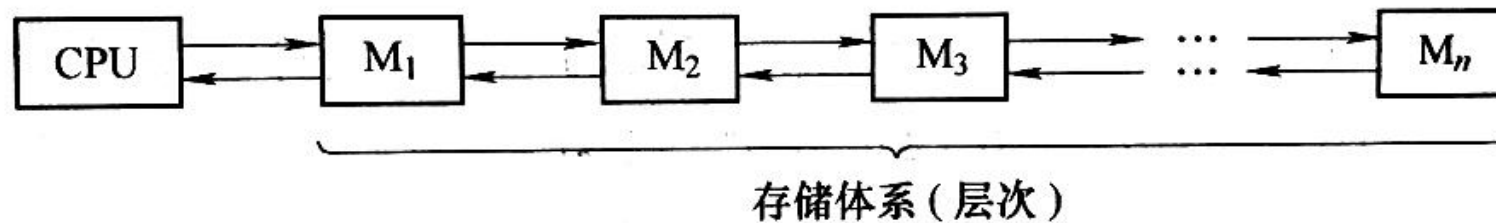


图 4-3 多级存储层次

## 4.1.2存储体系的构成依据（单选、填空、简答）

为了使存储体系能有效地工作，当CPU要用到某个地址的内容时，总希望它已在速度最快的M1中，这就要求能预知未来被访问信息的地址，这对存储体系的构成是非常关键的。

这种预知的可能性是基于计算机程序具有局部性，包括时间上的局部性和空间上的局部性。



那么意气风发地  
走在成功的道路上

## 真题练练手

1、Cache存储器是为了解决主存满足不了要求的性能指标是（） 1904

A:容量

B:价格

C:字长

D:速度





那么意气风发地  
走在成功的道路上

## 真题练练手

1、Cache存储器是为了解决主存满足不了要求的性能指标是（） 1904

A:容量

B:价格

C:字长

D:速度

答案： D



那么意气风发地  
走在成功的道路上

## 真题练练手

2、在主存-辅存存储层次结构中,从CPU上看, ( ) 接近于主存, ( ) 是辅存的。1810



那么意气风发地  
走在成功的道路上

## 真题练练手

2、在主存-辅存存储层次结构中,从CPU上看, ( ) 接近于主存, ( ) 是辅存的。1810

答案：速度 容量



那么意气风发地  
走在成功的道路上

## 真题练练手

3、CPU要能预知未来被访问信息的地址是基于计算机程序具有的特性是（）

1710

A:快速性

B:稳定性

C:局部性

D:可靠性



那么意气风发地  
走在成功的道路上

## 真题练练手

3、CPU要能预知未来被访问信息的地址是基于计算机程序具有的特性是（）

1710

A:快速性

B:稳定性

C:局部性

D:可靠性

答案：C

## 4.1.3 存储体系的性能参数

### 4. 多级存储层次（考纲要求，从来没考）

以图4-4所示的二级存储体系（M1，M2）为例来分析。设 $c_i$ 为 $M_i$ 的每位价格， $S_{mi}$ 为以位计算的存储容量，为CPU访问到 $M_i$ 中的信息所需的时间。为评价存储层次性能，引入存储层次的每位价格 $c$ 命中率 $H$ 和等效访问时间 $T_A$ 。

存储层次的每位价格

$$c = \frac{c_1 S_{M_1} + c_2 S_{M_2}}{S_{M_1} + S_{M_2}}$$

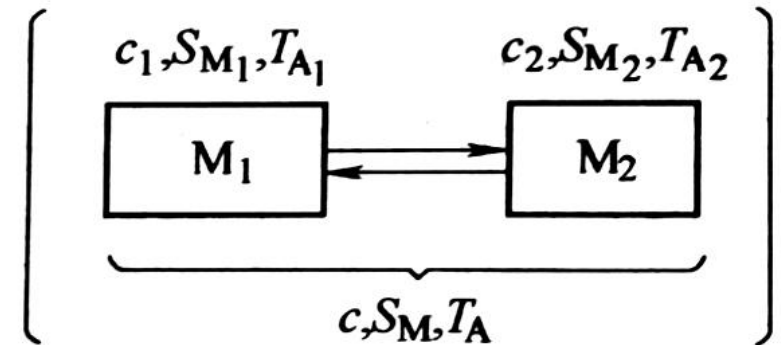


图 4-4 二级存储体系的评价

## 4.1.3 存储体系的性能参数

### 4.多级存储层次（单选）

命中率 $H$ 定义为CPU产生的逻辑地址能在 $M1$ 中访问到（命中到）的概率。

命中率 $H$ 与程序的地址流、所采用的地址预判算法及的容量都有很大关系

即 $H$ 越接近于1越好

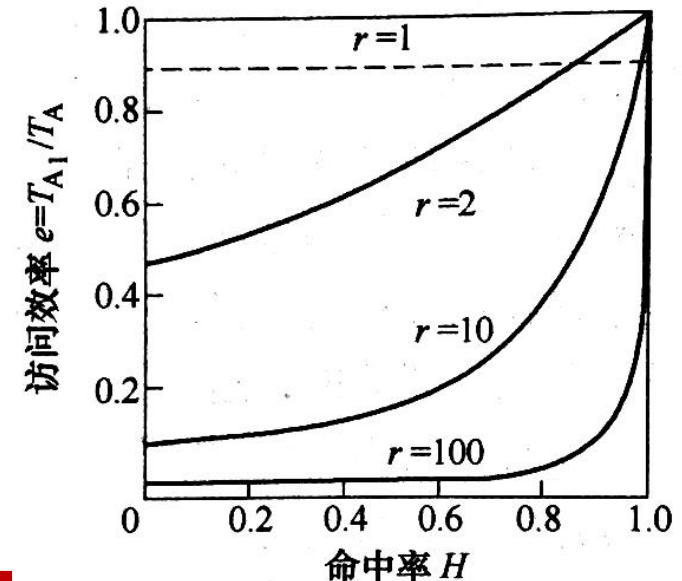


图4-5 对于不同的 $r$ ，命中率 $H$ 与访问效率 $e$ 的关系

## 4.2 虚拟存储器

**本节主要内容：**

**3种虚拟管理方式的原理、地址映像规则、映像表机构、地址变换过程、各自的优点和问题。**

**段页式虚拟存储器**

**页式虚拟存储器**

**用FIFO、LRU、OPT算法页面替换的过程模拟**

**堆栈型替换算法定义、种类**

**PFF替换算法、分析虚拟存储器，综合评估和改进页式虚拟存储器性能的办法。**



## 4.2.1 虚拟存储器的管理方式（填空）

根据存储映像算法的不同，可有多种不同的存储管理方式的虚拟存储器，其中主要有段式、页式和段页式3种。

## 4.2.1 虚拟存储器的管理方式

### 1. 段式管理（填空）

程序都有模块性，一个复杂的大程序总可以分解成多个在逻辑上相对独立的模块。这些模块可以是主程序、子程序或过程，也可以是数据块。

为了进行段式管理，每道程序在系统中都有一个段（映像）表来存放该道程序各段装入主存的状况信息。

## 4.2.1 虚拟存储器的管理方式

### 1. 段式管理

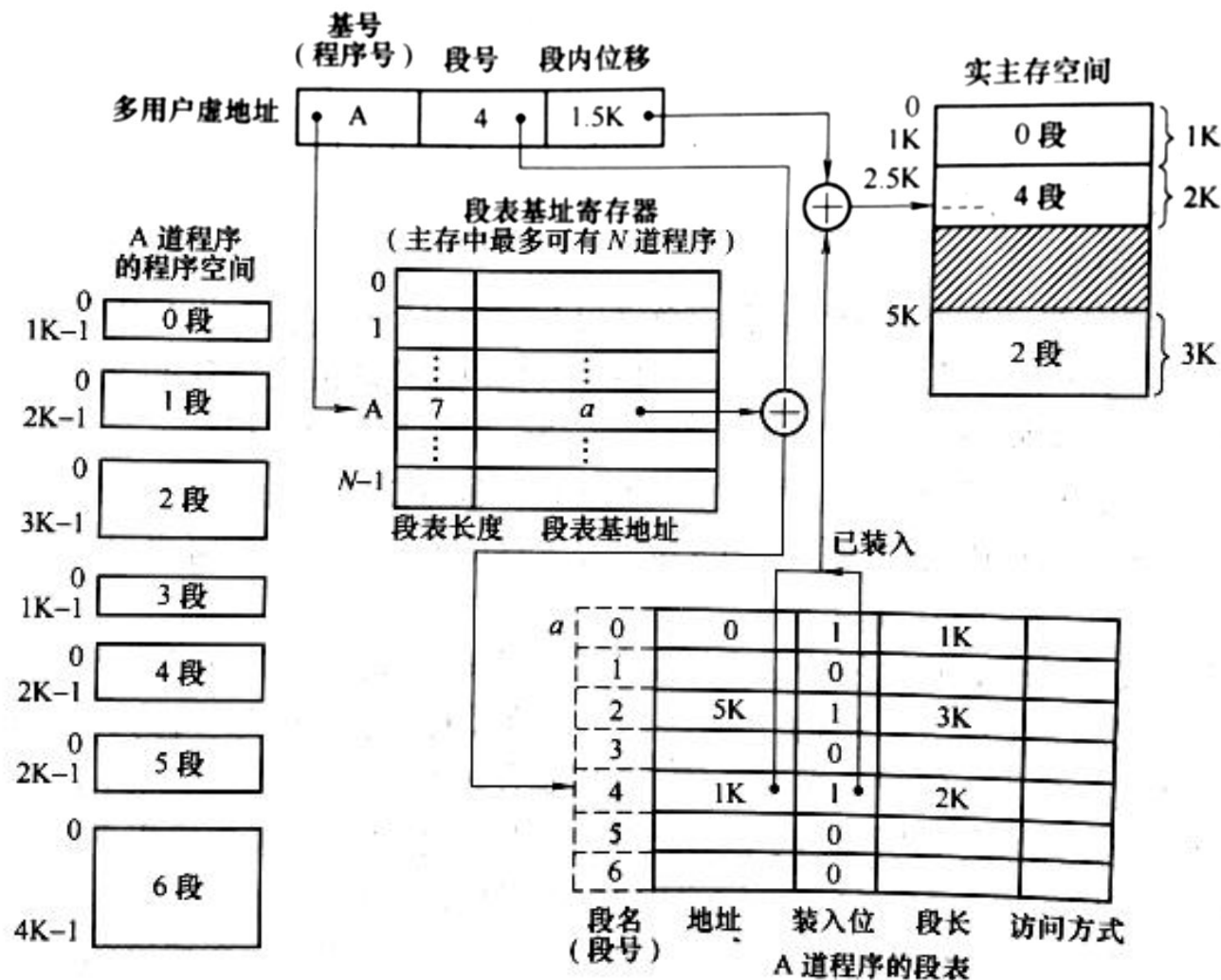


图 4-6 段式管理的定位映像机构及地址的变换过程

## 4.2.1 虚拟存储器的管理方式

### 1. 段式管理

段式中每个段**独立**，有利于程序员灵活实现段的链接，段的扩大、缩小和修改，而**不影响**到其他的段；

每段只包含**一种类型**的对象，如过程或是数组、堆栈、标量等集合，易于针对其特定类型实现保护；

把**共享**的程序或数据单独构成一个段，从而易于实现多个用户、进程对共用段的管理，等等。

口诀：独一共



那么意气风发地  
走在成功的道路上

## 真题练练手

- 1、根据所用的存储映像算法，虚拟存储器管理方式主要有段式、（     ）和（     ）三种。1204   1504



那么意气风发地  
走在成功的道路上

## 真题练练手

1、根据所用的存储映像算法，虚拟存储器管理方式主要有段式、（     ）和（     ）三种。1204 1504

答案：页式   段页式



那么意气风发地  
走在成功的道路上

## 真题练练手

2、一个复杂的大程序可以分解成多个在逻辑上相对独立的模块，这些模块可以是主程序、（ ）或（ ）也可以是数据块。1710



那么意气风发地  
走在成功的道路上

## 真题练练手

2、一个复杂的大程序可以分解成多个在逻辑上相对独立的模块，这些模块可以是主程序、（ ）或（ ）也可以是数据块。1710

答案：子程序 过程



## 4.2.1 虚拟存储器的管理方式

### 2. 页式管理（填空）

页式存储是把主存空间和程序空间都机械地等分成固定大小的页（页面大小随计算机而异，一般在512B到几KB之间），按页顺序编号。  
与段式一样，计算机是采用多道程序方式工作的。

## 4.2.1 虚拟存储器的管理方式

### 2. 页式管理

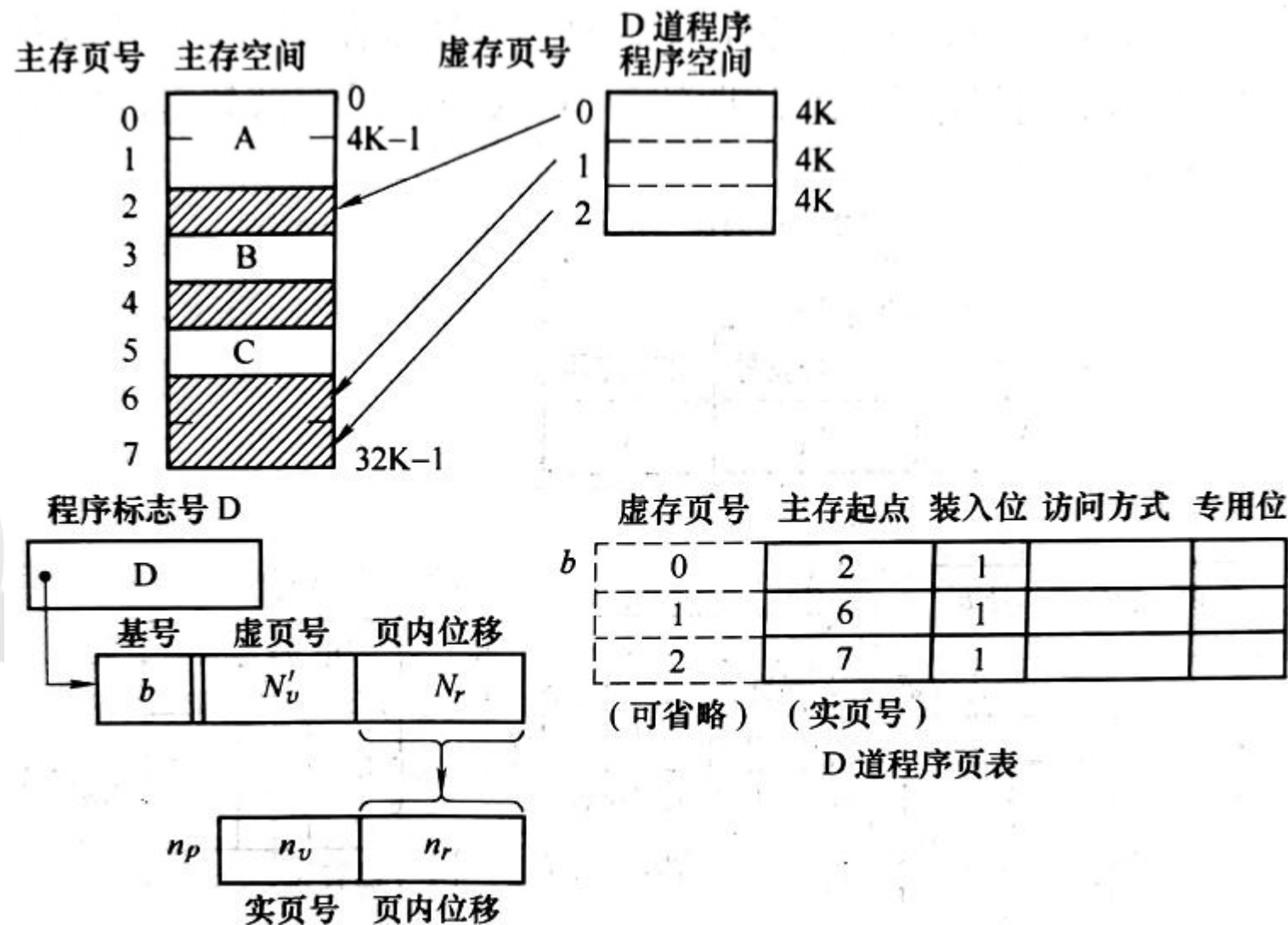


图 4-1 采用页式存储后 D 道程序仍可装入

## 4.2.1 虚拟存储器的管理方式

### 3. 段页式管理（填空）

段页式 = 段表 + 页表



## 4.2.1 虚拟存储器的管理方式

### 3. 段页式管理

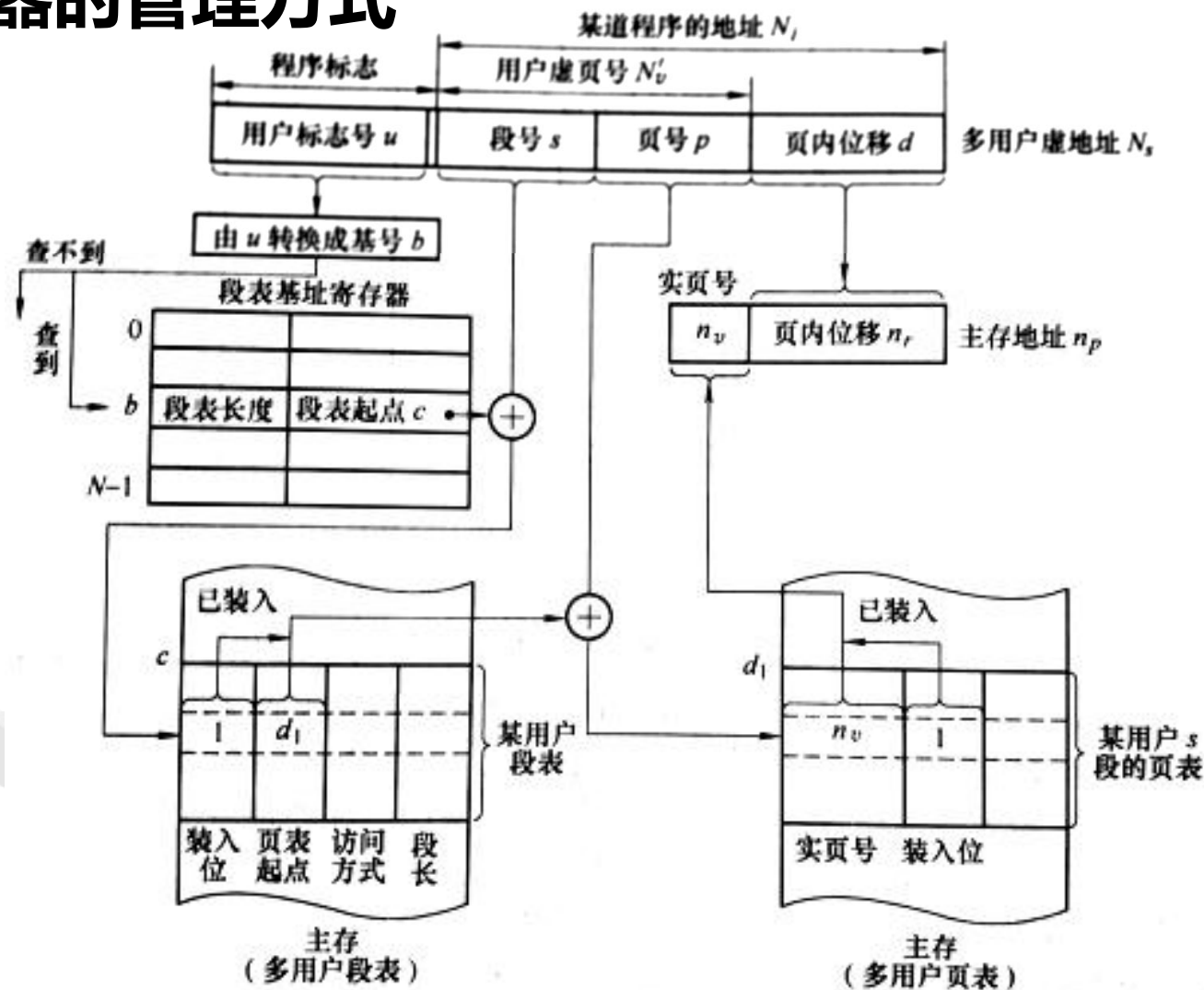


图 4-9 段页式管理的定位映像机构及其地址的变换过程



那么意气风发地  
走在成功的道路上

## 真题练练手

1、虚拟存储器的页式管理是把（ ）空间和（ ）空间机械等分成固定大小的页，按页顺序编号。 1804



那么意气风发地  
走在成功的道路上

## 真题练练手

1、虚拟存储器的页式管理是把（ ）空间和（ ）空间机械等分成固定大小的页，按页顺序编号。 1804

答案：主存 程序



那么意气风发地  
走在成功的道路上

## 真题练练手

2、段页式虚拟存储器的内部地址映像表机构有（ ）和（ ）两部分。

0807



那么意气风发地  
走在成功的道路上

## 真题练练手

2、段页式虚拟存储器的内部地址映像表机构有（ ）和（ ）两部分。

0807

答案：断表 页表



## 4.2.2页式虚拟存储器的构成

### 1.地址的映像和变换（单选）

页式虚拟存储器是采用页式存储和管理的主存—辅存存储层次。

地址的映像是将每个虚存单元按某种规则（算法）装入（定位于）实主存，建立起多用户虚地址NS与实（主）存地址np之间的对应关系。

由于是把大的虚存空间压缩到小的主存空间，主存中的每一个页面位置应可对应多个虚页，能对应多少个虚页与采用的映像方式有关。

## 4.2.2 页式虚拟存储器的构成

### 2. 全相联映像

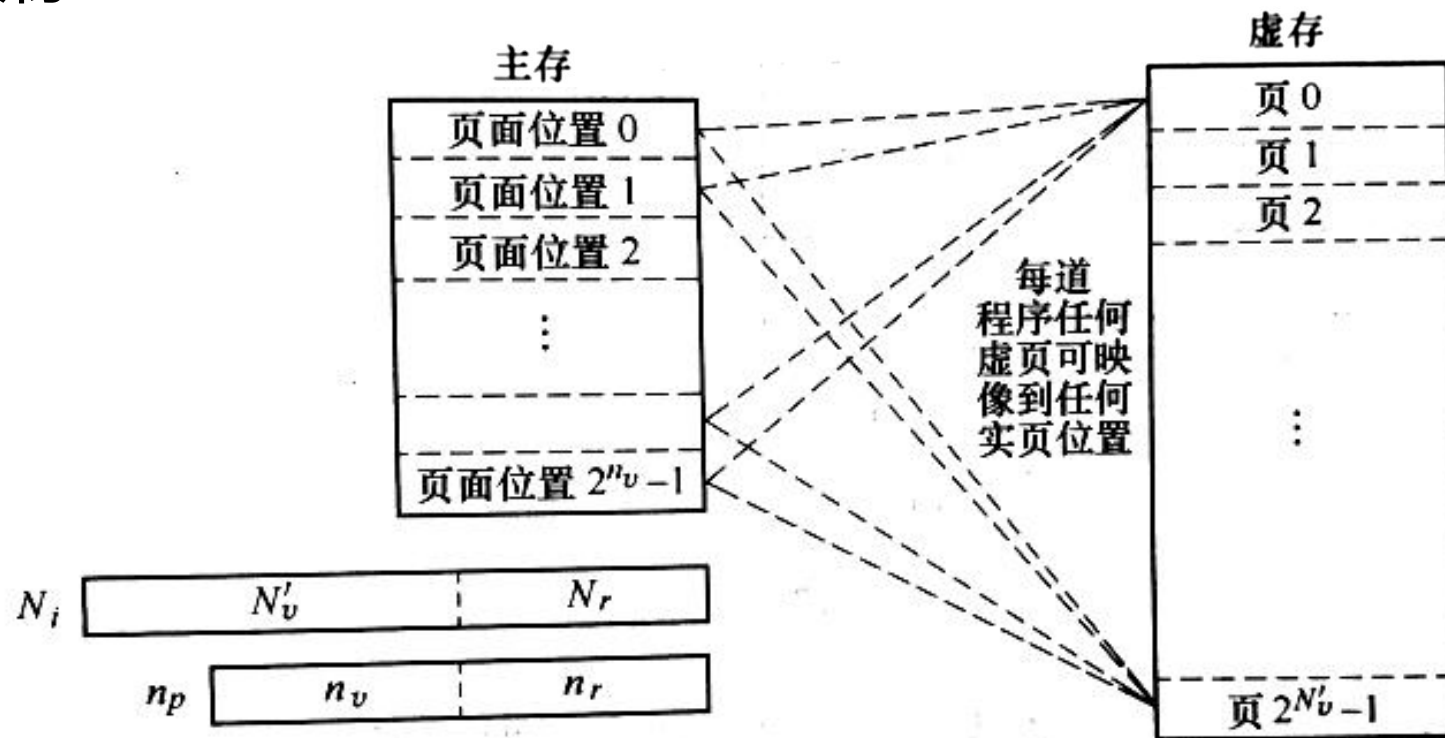


图 4-11 全相联映像

## 4.2.2页式虚拟存储器的构成

### 3.相联目录表法（单选）

把页表压缩成只存放已装入主存的那些虚页（用基号 $b$ 和标识）与实页位置（ $nv$ ）的对应关系，如图4-12所示，该表最多为 $2^{nv}$ 行。我们称它为相联目录表法，简称目录表法。该表采用按内容访问的相联存储器构成。

按内容访问相联的不同于按地址访问的

随机存储器

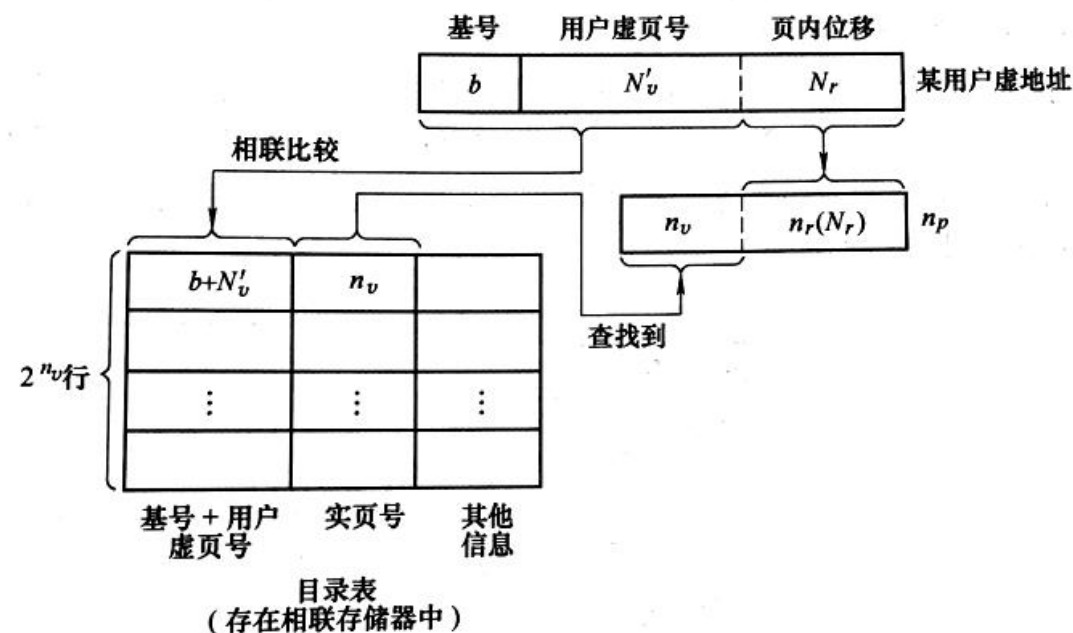


图 4-12 目录表法

## 4.2.2页式虚拟存储器的构成

### 4.页面替换算法（填空）

替换算法的确定主要看主存是否有高的命中率，也要看算法是否便于实现，辅助软、硬件成本是否低。

目前已研究过多种替换算法，如随机算法、先进先出算法、近期最少使用（近期最久未用过）算法等。

## 4.2.2页式虚拟存储器的构成

### 4.页面替换算法（了解原理）

(1) 随机算法 (Random, RAND)是用软的或硬的随机数产生器产生主存中要被替换页的页号。

这种算法简单，易于实现

## 4.2.2页式虚拟存储器的构成

### 4.页面替换算法（了解原理）

（2）先进先出算法（First-In First-Out, FIFO)是选择最早装入主存的页作为被替换的页。

这种算法实现方便，只要在操作系统为主存管理所设的主存页面表中给每个实页配一个计数器字段

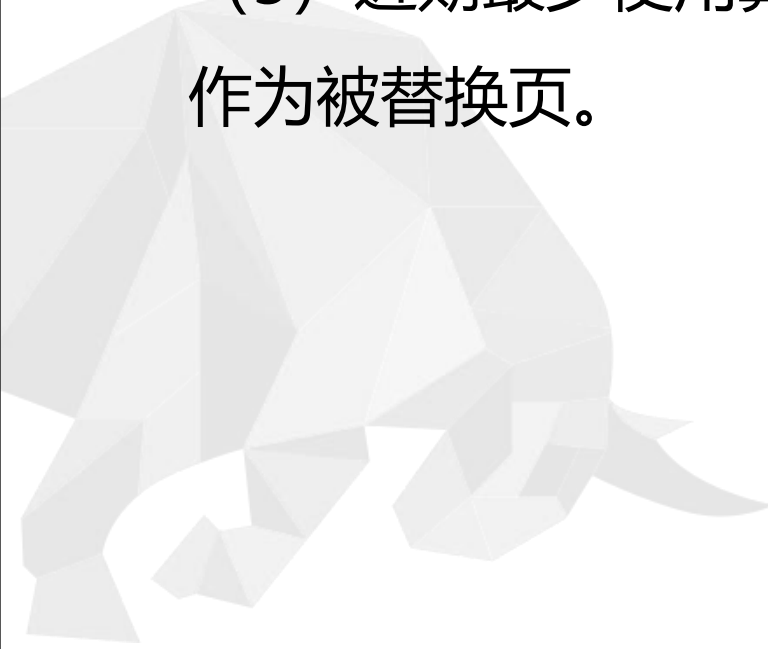
实页号	(计数器)						其他信息
	占用位	程序号	段页号	使用位	程序优先位	$H_s$	
0							
1							
⋮							
$2^n-1$							

图 4-14 主存页面表

## 4.2.2页式虚拟存储器的构成

### 4.页面替换算法（了解原理）

(3) 近期最少使用算法 (Least Recently Used, LRU) 是选择近期最少访问的页作为被替换页。



## 4.2.2页式虚拟存储器的构成

### 4.页面替换算法

【例】设有一道程序，有1~5页，执行时的页地址流（即依次用到的程序页页号）为

2, 3, 2, 1, 5, 2, 4, 5, 3, 2, 5, 2

若分配给该道程序的主存有3页，图4-15所示为分别用FIFO、LRU、OPT这3种替换算法对这3页的使用和替换过程。其中用\*号标记出按所用算法选作下次应被替换的页号。



## 4.2.2页式虚拟存储器的构成

### 4.页面替换算法(填空)

结论1命中率与所选用替换算法有关。LRU算法要优于FIFO算法。命中率也与页地址流有关。

时间 / 页地址流	1	2	3	4	5	6	7	8	9	10	11	12
先进先出 (FIFO)	2 2 2	2 3 2	2 3 2	2* 3 1	5 3* 1	5 2 1*	5* 2 4	5* 2 4	3 2* 4	3 2* 4	3 5 4*	3* 5 2
	调进	调进	命中	调进	替换	替换	替换	命中	替换	命中	替换	替换
近期最少使用 (LRU)	2 2 2	2 3 2	2 3 2	2 3* 1	2* 5 1	2 5 1*	2 5* 4	2* 5 4	3 5 4*	3 5* 2	3* 5 2	3* 5 2
	调进	调进	命中	调进	替换	命中	替换	命中	替换	替换	命中	命中
优化 (OPT)	2 2 2	2 3 2	2 3 2	2 3 1*	2 3* 5	2* 3 5	4* 3 5	4* 3 5	4* 3 5	2 3* 5	2 3 5	2 3 5
	调进	调进	命中	调进	替换	命中	替换	命中	命中	替换	命中	命中

图 4-15 3 种替换算法对同一页地址流的替换过程

## 4.2.2页式虚拟存储器的构成

### 4.页面替换算法（填空）

结论2命中率与分配给程序的  
的主存页数有关。

时间 / 页地址流	1	2	3	4	5	6	7	8	9	10	11	12
先进先出 (FIFO)	2	2	2	2*	5	5	5*	5*	3	3	3	3*
命中 3 次												
	调进	调进	命中	调进	替换	替换	替换	命中	替换	命中	替换	替换
近期最少使用 (LRU)	2	2	2	2	2*	2	2	2*	3	3	3*	3*
命中 5 次												
	调进	调进	命中	调进	替换	命中	替换	命中	替换	替换	命中	命中
优化 (OPT)	2	2	2	2	2	2*	4*	4*	4*	2	2	2
命中 6 次												
	调进	调进	命中	调进	替换	命中	替换	命中	命中	替换	命中	命中

图 4-15 3 种替换算法对同一页地址流的替换过程

## 4.2.2页式虚拟存储器的构成

### 5.堆栈型替换算法

设A是长度为L的任意一个页地址流，t为已处理过t-1个页面的时间点，n为分配给该地址流的主存页数， $B_t(n)$ 表示在t时间点、在n页的主存中的页面集合，表示到t时间点已遇到过的地址流中相异页的页数。如果替换算法满足下式，则是堆栈型的替换算法。

$$n < L_t, \text{ 时, } B_t(n) \subset B_t(n+1)$$

$$n \geq L_t, \text{ 时, } B_t(n) = B_t(n+1)$$

## 4.2.2页式虚拟存储器的构成

### 5.堆栈型替换算法

- 1、LRU算法在庄村中保留的是 $n$ 个最近使用的页，他们又总是包含在 $n+1$ 个最近使用的页中，所以LRU算法是堆栈型算法。
- 2、OPT算法也是堆栈型算法

## 4.2.3页式虚拟存储器实现中的问题

### 1.页面失效的处理（单选）

要访问的虚页不在实际主存中时,就会发生页面失效。



## 4.2.3页式虚拟存储器实现中的问题

### 2.提高虚拟存储器等效访问速度的措施

(1)存储层次的等效访问时间是 $TA=HT1+(1-H)T2$ ,式中, $H$ 为主存命中率, $T1$ 、 $T2$ 分别是主、辅存访问时间。

(2)要提高存储层次等效访问速度,可采取的措施有:

- ①当等效访问时间远大于主存访问周期时,可采取提高主存命中率的方法;
- ②当主存命中率 $H$ 已经很高时,可提高主存的访问速度,以降低 $T1$ ;
- ③加快内部地址映像和变换,如采用快-慢表层次,增大快表命中率等。

## 4.2.3页式虚拟存储器实现中的问题

### 3.影响主存命中率和CPU效率的某些因素

程序地址流、替换算法以及分配给程序的实页数（主存容量）不同都会影响命中率。



那么意气风发地  
走在成功的道路上

## 真题练练手

1、虚拟存储器地址变换是指（ ） 1910

A:将实地址变换成虚地址

B:静态再定位时将程序的逻辑地址变换成主存的实地址

C:程序执行时将虚地址变换成对应的实地址

D:将指令的符号地址变换成二进制地址





那么意气风发地  
走在成功的道路上

## 真题练练手

1、虚拟存储器地址变换是指（ ） 1910

A:将实地址变换成虚地址

B:静态再定位时将程序的逻辑地址变换成主存的实地址

C:程序执行时将虚地址变换成对应的实地址

D:将指令的符号地址变换成二进制地址

答案：C



那么意气风发地  
走在成功的道路上

## 真题练练手

2、不属于堆栈型替换算法的是（ ） 1910

A:近期最久未使用算法

B:OPT算法

C:先进先出算法

D:近期最少使用算法



那么意气风发地  
走在成功的道路上

## 真题练练手

2、不属于堆栈型替换算法的是（ ） 1910

A:近期最久未使用算法

B:OPT算法

C:先进先出算法

D:近期最少使用算法

答案：C



那么意气风发地  
走在成功的道路上

## 真题练练手

3、页式虚拟存储器中的CPU要用到的指令或数据不在（ ）时会发生（ ）。1910



那么意气风发地  
走在成功的道路上

## 真题练练手

3、页式虚拟存储器中的CPU要用到的指令或数据不在（ ）时会发生（ ）。1910

答案：主存 页面失效



那么意气风发地  
走在成功的道路上

## 真题练练手

4、虚拟存储器的页式管理是把（ ）空间和（ ）空间机械等分成固定大小的页，按页顺序编号。 1804



那么意气风发地  
走在成功的道路上

## 真题练练手

4、虚拟存储器的页式管理是把（ ）空间和（ ）空间机械等分成固定大小的页，按页顺序编号。 1804

答案：主存 程序

## 4.3高速缓冲存储器

**本节主要内容：**

**cache存储器的组成与工作原理，与虚拟存储器对比**

**全相联、直接、组相联地址映像规则**

**cache块替换的原理**

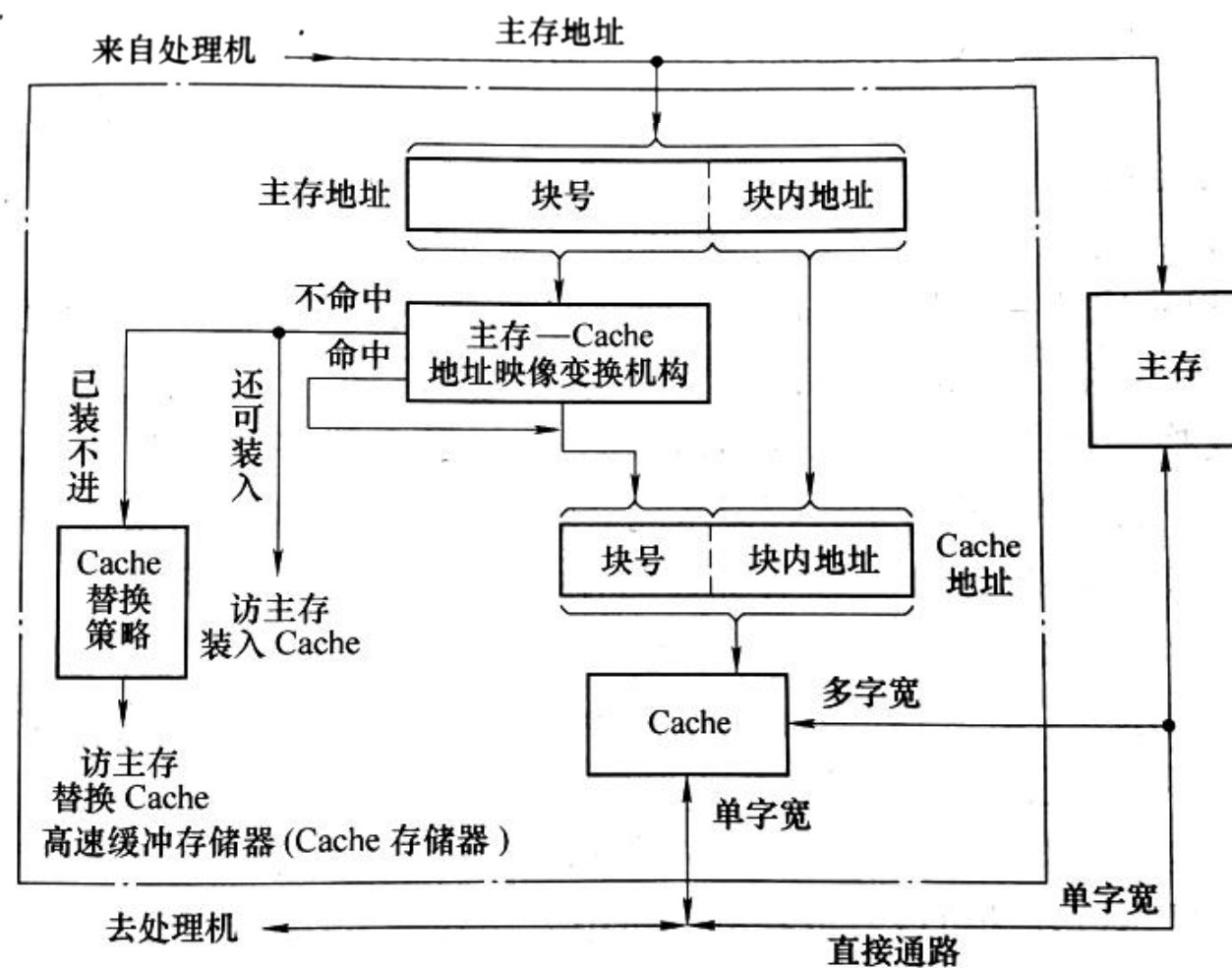
**各种替换算法装入或替换的过程图，计算cache的块命中率**

**cache存储器的等效访问速度与cache容量及速度的关系**



## 4.3.1工作原理和基本结构

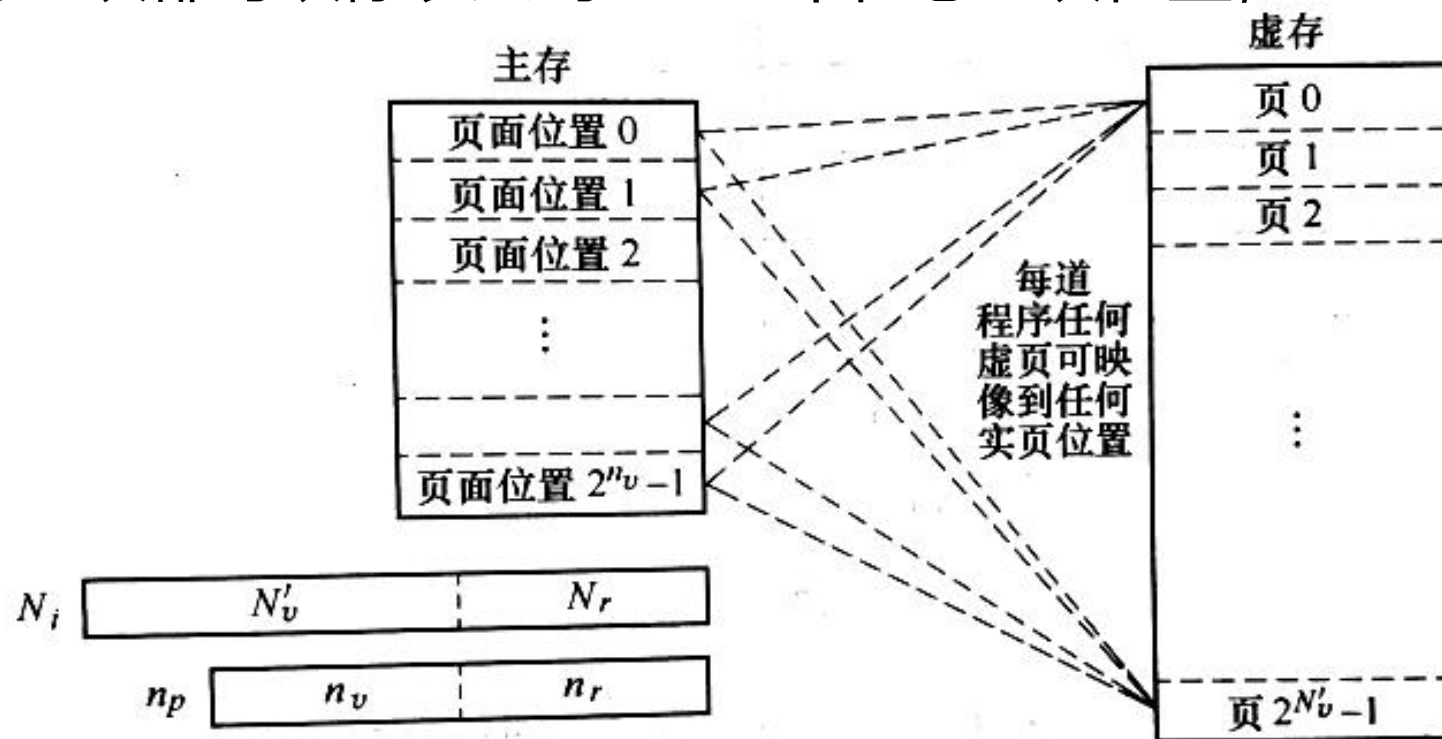
高速缓冲（Cache）存储器是为弥补主存速度的不足，在处理机和主存之间设置一个高速、小容量的Cache，构成Cache—主存存储层次，使之从CPU角度来看，速度接近于Cache，容量却是主存的。



## 4.3.2地址的映像与变换

### 1.全相联映像和变换

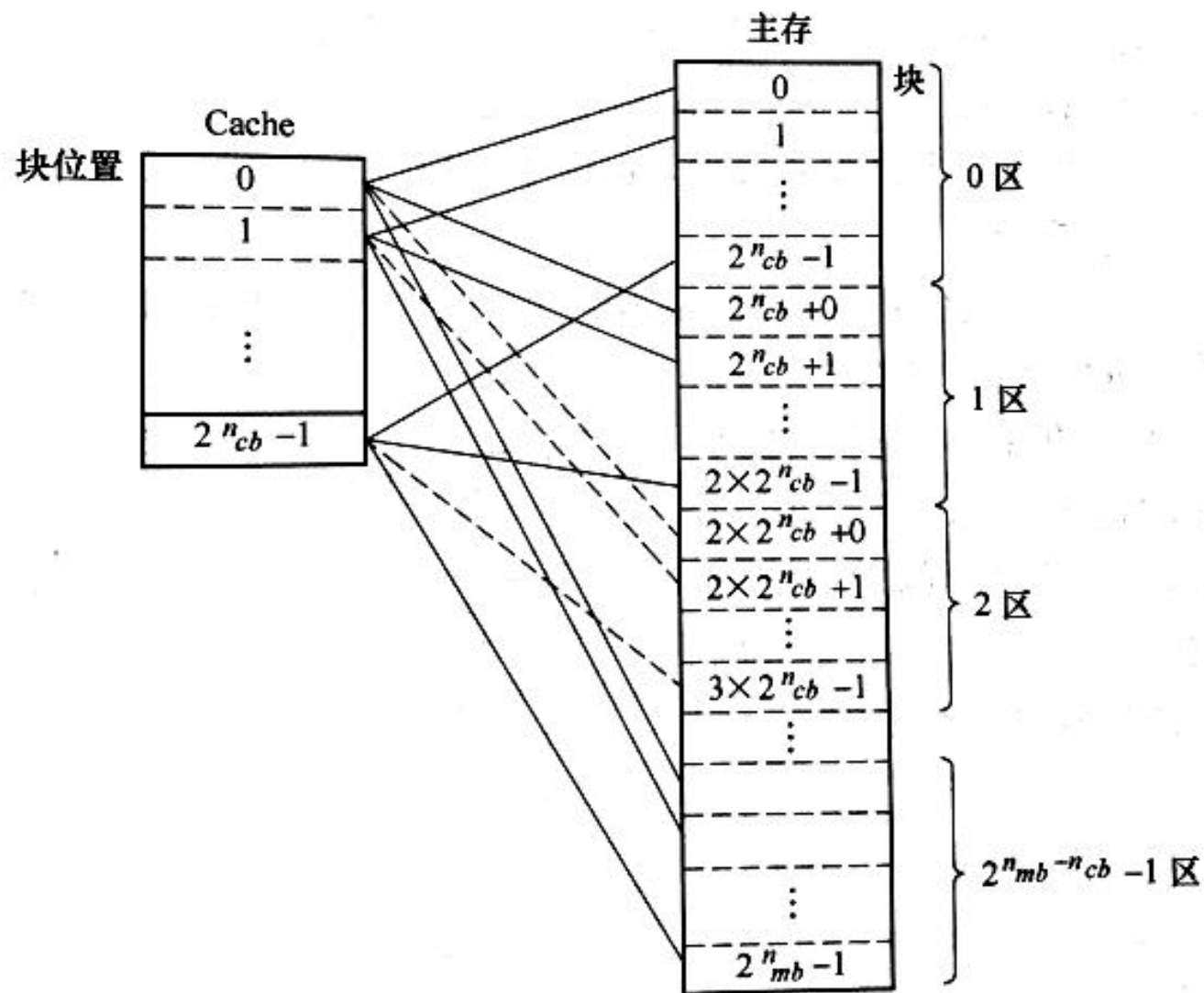
主存中任意一块都可映像装入到Cache中任意一块位置，



## 4.3.2地址的映像与变换

### 2.直接映像及其变换

把主存空间按Cache大小等分成区，每区内的各块只能按位置——对应到Cache的相应块位置上



## 4.3.2地址的映像与变换

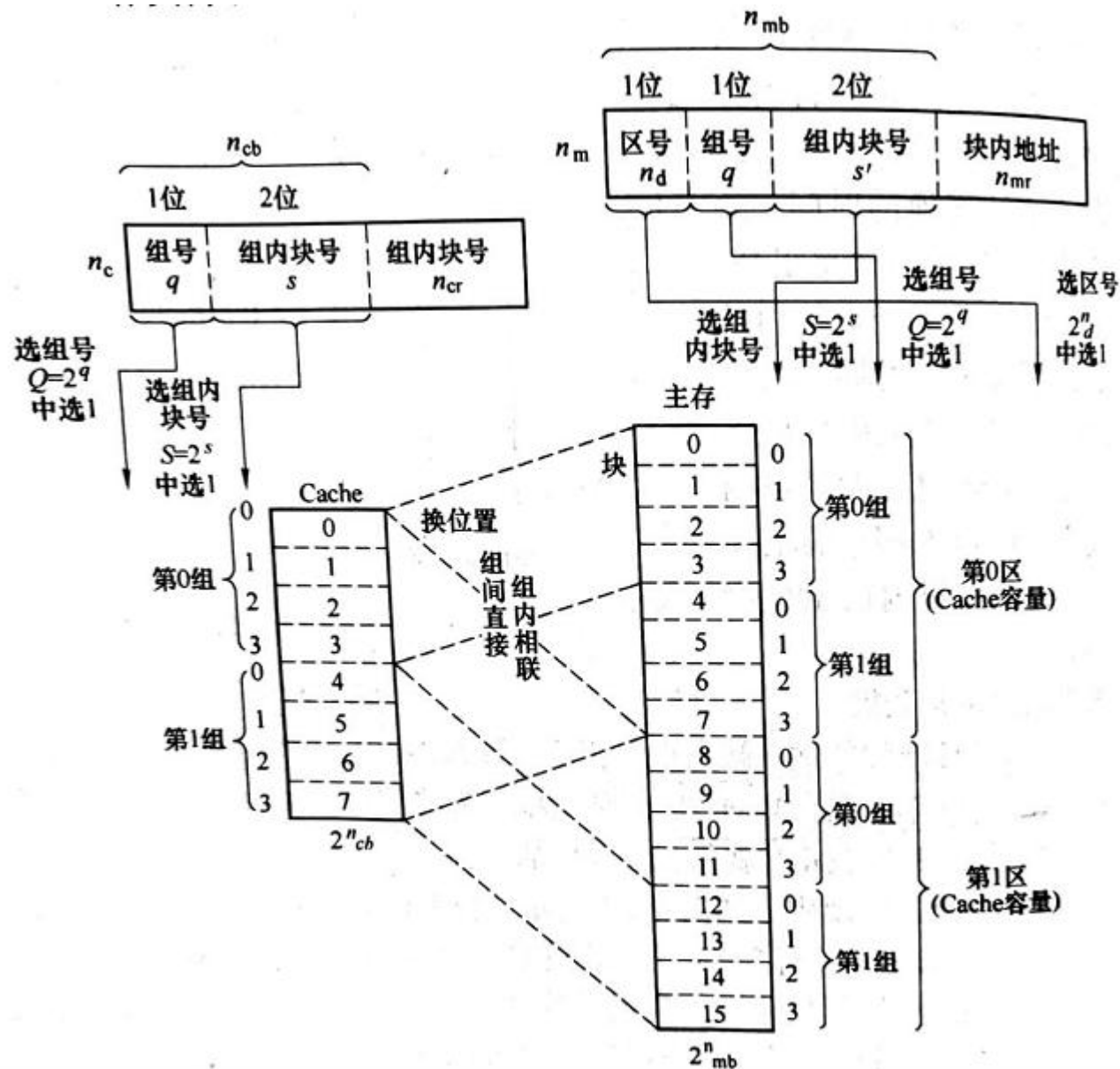
### 3.组相联映像及其变换

全相联映像和直接映像的优、缺点正好相反，那么能否将两者结合，采用一种映像规则，既能减少块冲突概率，提高Cache空间利用率，又能使地址映像机构及地址变换速度比全相联的简单和快速

组相联映像指的是各组之间是直接映像，而组内各块之间是全相联映像。

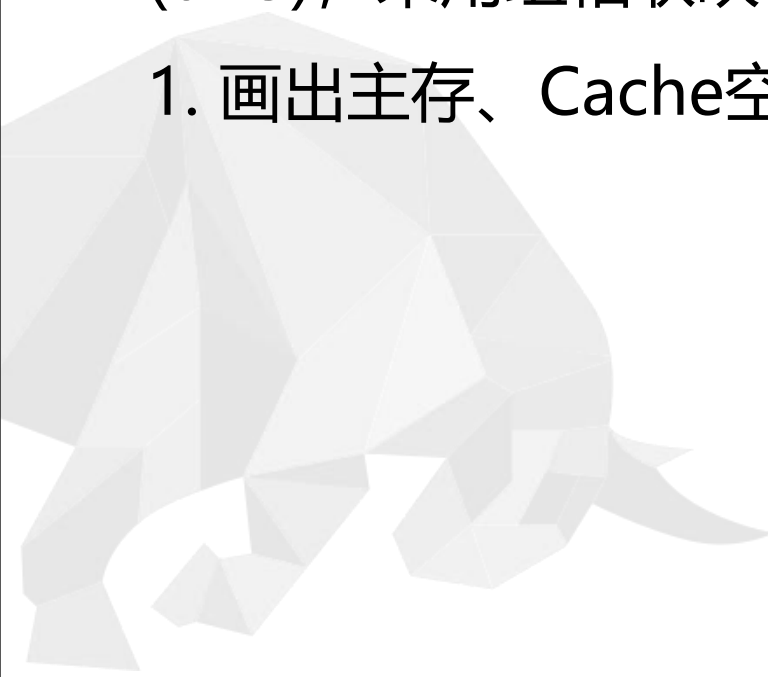
## 4.3.2地址的映像与变换

### 3.组相联映像及其变换



【1204真题】有一个Cache-主存存储层次，主存共8个块(0~7)，Cache有4个块(0~3)，采用组相联映像，组内块数为2块，替换算法为LRU算法。

1. 画出主存、Cache空间块的映像对应关系示意图；

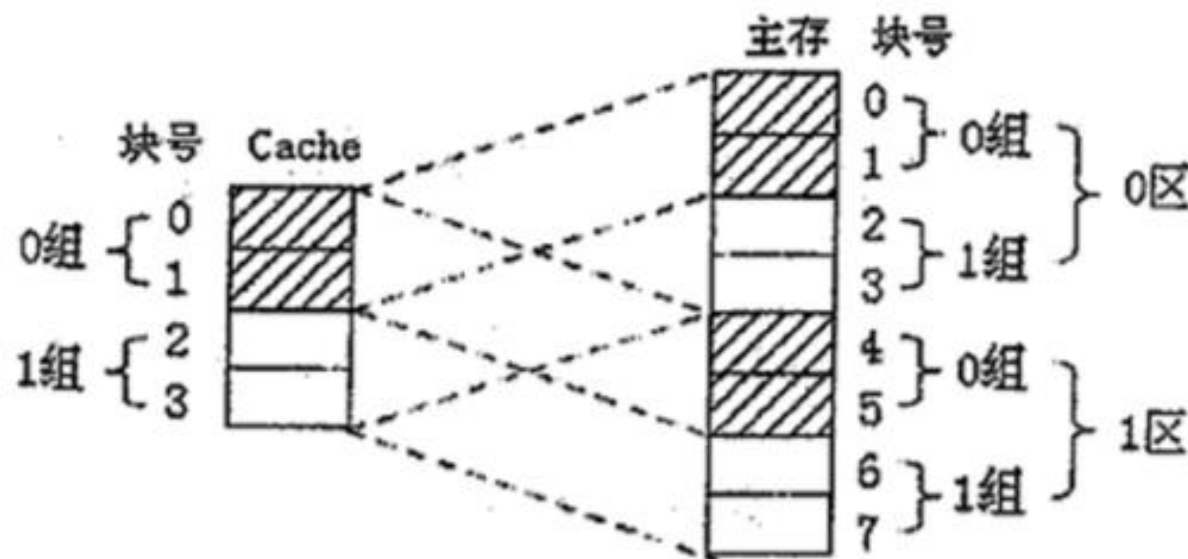




【1204真题】有一个Cache-主存存储层次，主存共8个块(0~7)，Cache有4个块(0~3)，采用组相联映像，组内块数为2块，替换算法为LRU算法。

1. 画出主存、Cache空间块的映像对应关系示意图；

主存 - Cache 空间块的映像关系：



主存的第 0、1、4、5 块只可映像装入或替换 Cache 中的第 0、1 块内容。主存的第 2、3、6、7 块只可映像装入或替换 Cache 中的第 2、3 块内容。

## 4.3.4 Cache存储器的透明性及性能分析

### 1. Cache存储器的透明性分析及解决办法（填空、简答）

一般可有写回法和写直达法两种。

写回法也称为抵触修改法。它是在CPU执行写操作时，信息只写入Cache,仅当需要替换时，才将改写过的Cache块先写回主存，然后再调入新块。



## 4.3.4 Cache存储器的透明性及性能分析

### 1. Cache存储器的透明性分析及解决办法（填空、简答）

一般可有写回法和写直达法两种。

写直达法也称存直达法。它是利用Cache存储器在处理机和主存之间的直接通路，每当处理机写入Cache的同时，也通过此通路直接写入主存。

## 4.3.4 Cache存储器的透明性及性能分析

### 2.Cache的取算法（单选）

Cache所用的取算法基本上是按需取进法，即在Cache块失效时才将要访问的字所在的块取进。

适当选择好Cache的容量、块的大小、组相联的组数和组内块数，是可以保证有较高的命中率的。

## 4.3.4 Cache存储器的透明性及性能分析

### 2.Cache的取算法

为了便于硬件实现，通常在访问主存第  $i$  块（无论是否已取进Cache）时，只预取顺序的第  $i + 1$  块。

何时取进该块，可有恒预取和不命中时预取两种方法。恒预取是只要访问到主存第  $i$  块，无论Cache是否命中，恒预取第  $i + 1$  块。不命中时预取则是只当访问主存第  $i$  块在Cache不命中时，才预取主存中的第  $i + 1$  块。Amdahl 470 V/8就采用不命中时预取方法。

## 4.3.4 Cache存储器的透明性及性能分析

### 2.Cache的取算法

采用预取法并非一定能提高命中率，它还和块的大小及预取开销的大小有关。



## 4.3.4 Cache存储器的透明性及性能分析

### 3.Cache存储器的性能分析

评价Cache存储器的性能主要是看命中率的高低，而命中率与块的大小、块的总数（即Cache的总容量）、采用组相联时组的大小（组内块数）、替换算法和地址流的簇聚性等有关。

Cache本身的速度与容量都会影响 Cache存储器的等效访问速度。

【1904真题】考虑一个1000个字的程序,其访问虚存的地址流为16、219、136、156、332、480、503、868、916、999。若页面大小为200字,主存容量为400字,采用FIFO替换算法,请按访存的各个时刻,写出其虚页地址流,计算主存命中率。



【1904真题】考虑一个1000个字的程序,其访问虚存的地址流为16、219、136、156、332、480、503、868、916、999。若页面大小为200字,主存容量为400字,采用FIFO替换算法,请按访存的各个时刻,写出其虚页地址流,计算主存命中率。

页面大小为 200 字,主存容量为 400 字,可知实存页数为 2。根据虚存的地址流,可得其虚页地址流为:

0,1,0,0,1,2,2,4,4,4

采用 FIFO 替换算法替换时的实际装入和替换过程如答 27 图所示。

虚地址	16	219	136	156	332	480	503	868	916	999
虚页地址	0	1	0	0	1	2	2	4	4	4
n=2	0	0*	0*	0*	0*	2	2	2*	2*	2*
		1	1	1	1	1*	1*	4	4	4
			H	H	H		H		H	H

其中: “\*” 标记的是候选替换的虚页号, H 表示命中。

命中率  $H = 6/10 = 60\%$

## 4.4 三级存储体系

**本节主要内容：**

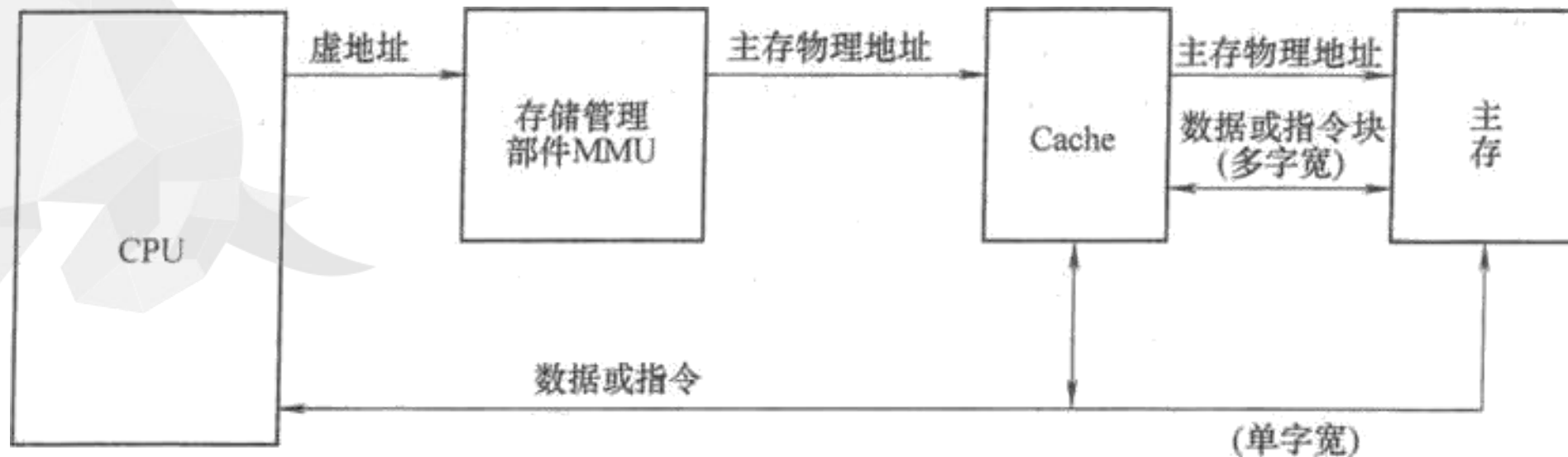
**三级存储体系的组织，要求达到“领会”层次**





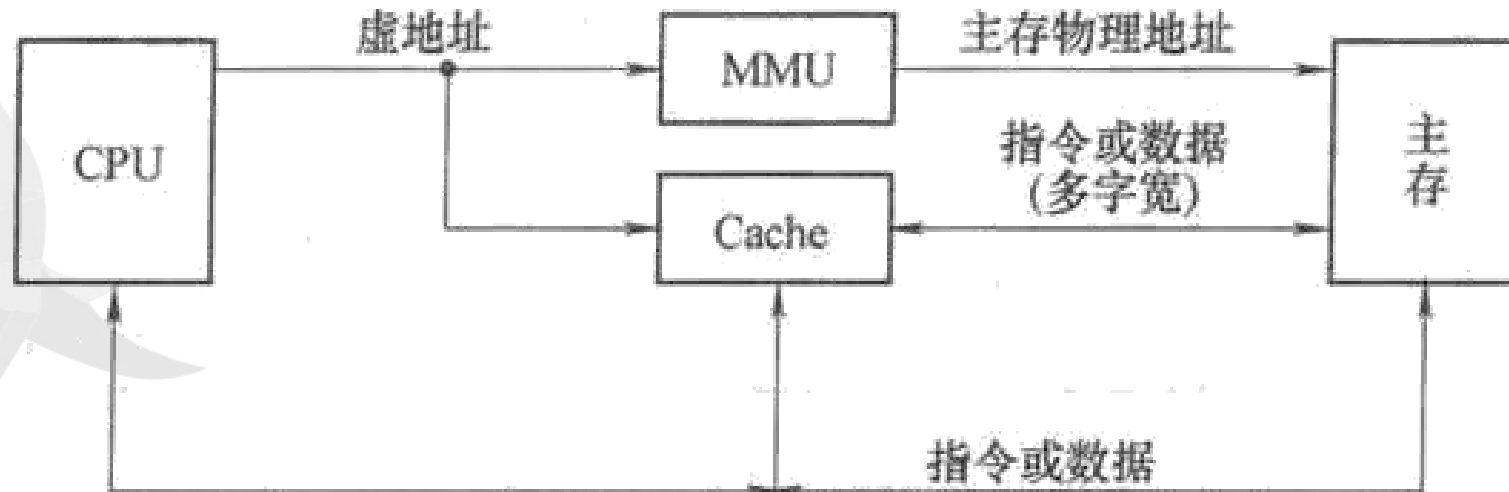
## 4.4.1 物理地址Cache

物理地址Cache是由 “Cache—主存” 和 “主存—辅存” 两个独立的存储层次组成,



## 4.4.2 虚地址Cache

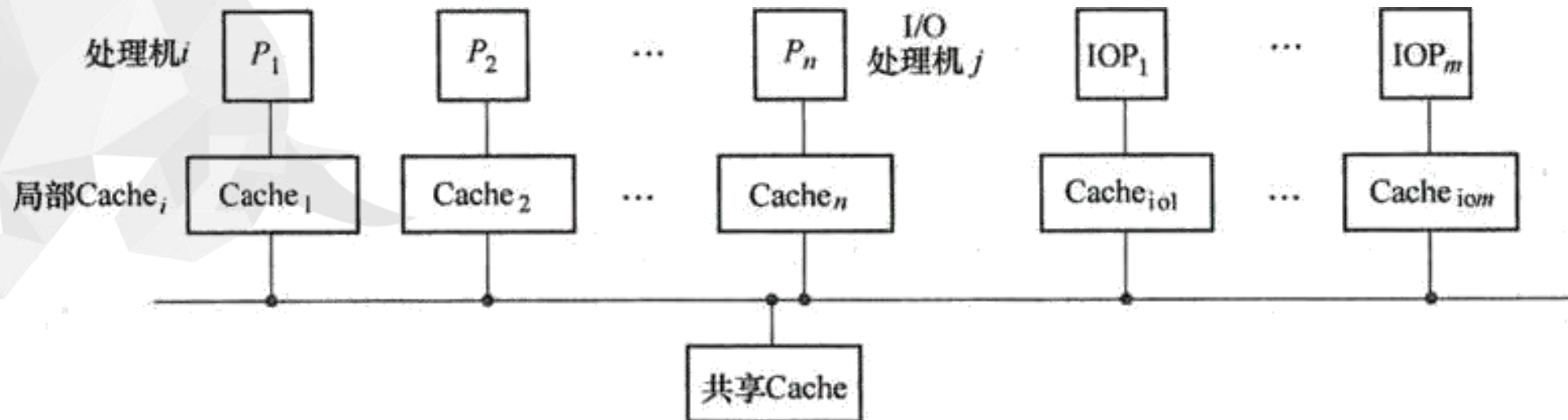
虚地址Cache是将Cache—主存—辅存直接构成三级存储层次形式，其组成形式如图



### 4.4.3 全Cache

全Cache是最近出现的组织形式，尚不成熟，尚未商品化。它没有主存，只用Cache 与辅存中的一部分构成“Cache-辅存”存储体系。

Cache存储系统的等效访问时间要接近于Cache的，容量是虚地址空间的容量。图4-34是在多处理机实现的一种方案。



## 第5章 标量处理机

## 第2章 数据表示、寻址方式与指令系统



## 5.1 重叠方式

**本节主要内容：**

**顺序方式与重叠方式的定义和特点。**

**重叠方式解决访存冲突的办法**

**“一次重叠”的定义和好处**

**条件转移指令与后续指令之间的相关及处理办法**

**指令、主存数、通用寄存器组的数和变（基）址相关的定义及处理办法**

**指令间微操作重叠的时间关系**

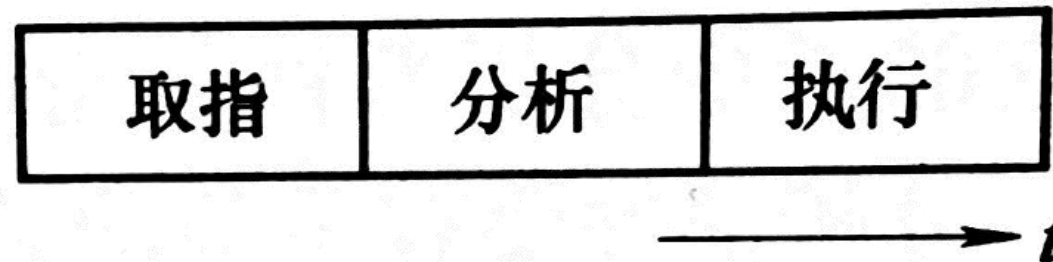
**计算执行完若干条指令所需的时间**

## 5.1.1 重叠原理与一次重叠

### 1. 重叠的原理（填空）

指令的重叠解释使机器语言程序的执行速度会比采用顺序解释的有较大的提高。

解释一条机器指令的微操作可归并成取指令、分析和执行三部分



## 5.1.1 重叠原理与一次重叠

### 2. 顺序解释（单选）

顺序解释指的是各条指令之间**顺序串行**（执行完一条指令后才取下条指令）地进行，每条指令内部的各个微操作也顺序串行地进行。

顺序解释的优点是**控制简单**，转入下条指令的**时间**易于控制。但缺点是上一步操作未完成，下一步操作便不能开始，速度上不去，计算机各部件的利用率低。

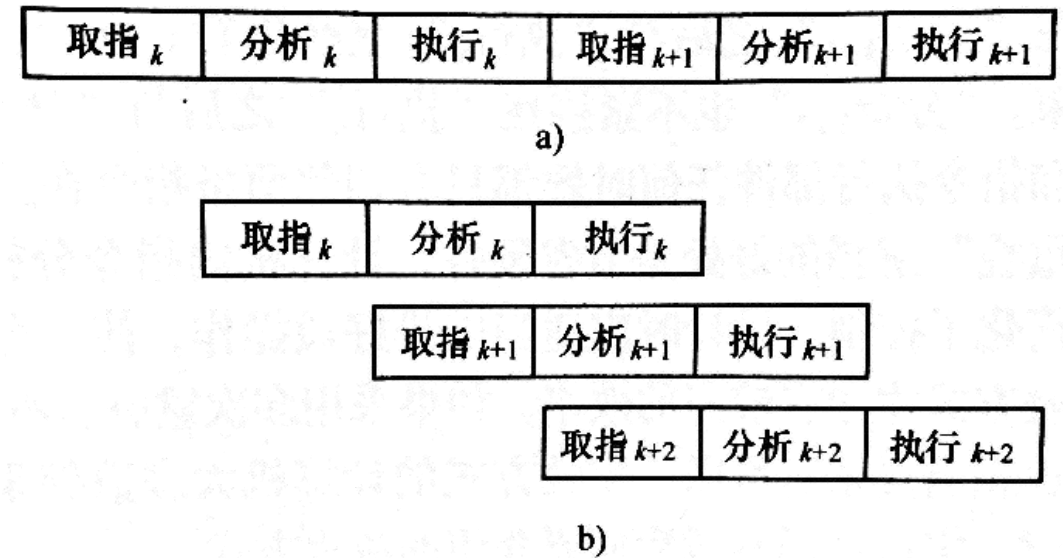


图 5-2 指令的顺序与重叠解释

a) 顺序解释 b) 重叠解释的一种方式



## 5.1.1 重叠原理与一次重叠

### 3. 一次重叠（单选）

- 1) 实际上“分析”和“执行”所需的时间常不相同，还需在硬件中解决控制上的同步，保证任何时候都只是“**执行k**”与“**分析k+1**”重叠；
- (2) 这种指令分析部件和指令执行部件任何时候都只有**相邻两条指令**在重叠解释的方式为“一次重叠”；

## 5.1.2 相关处理

### 1. 转移指令的处理（简单了解）

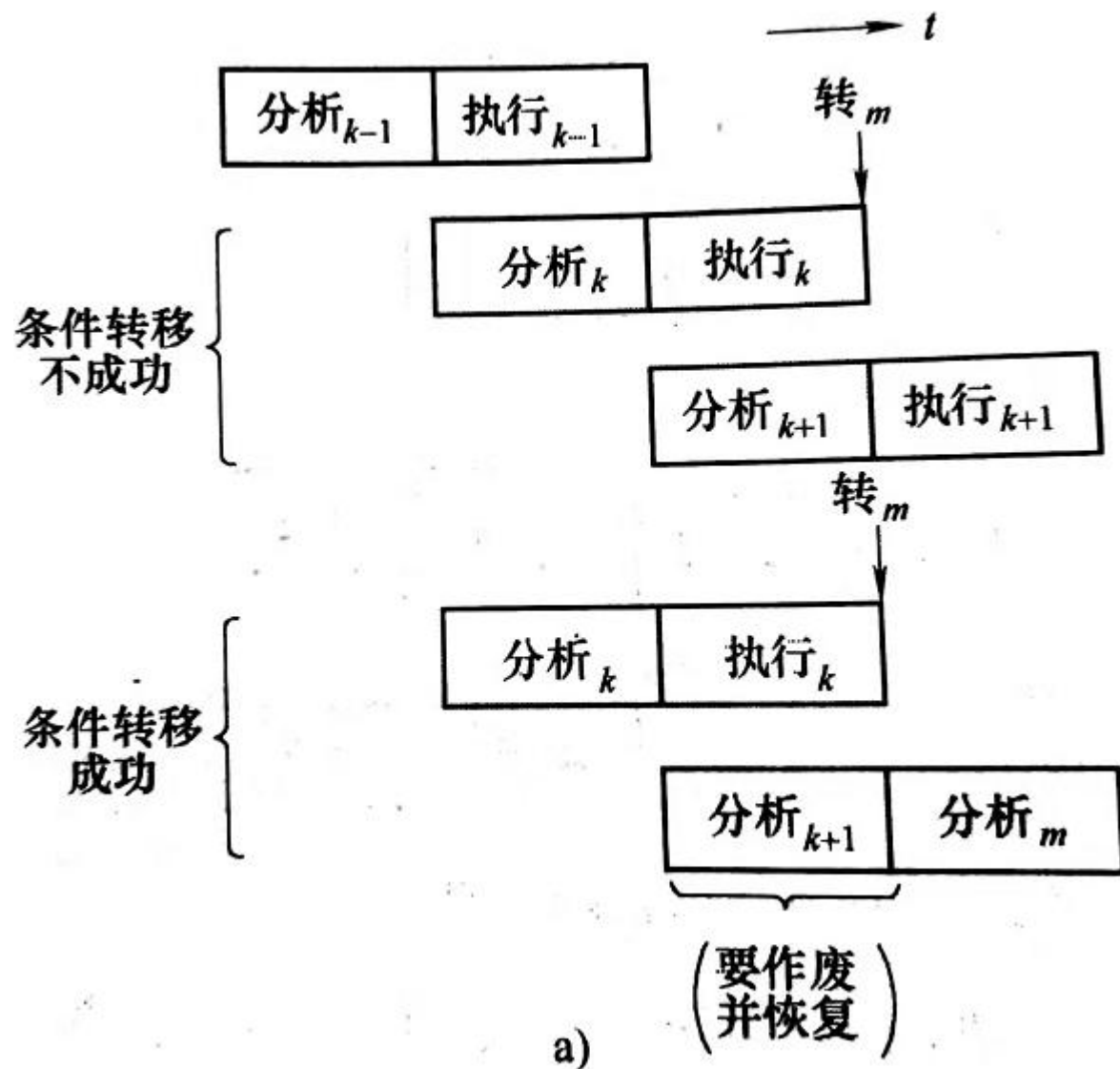
当程序中遇到条件转移时，一旦条件转移成功，重叠解释实际变成了顺序解释。



## 5.1.2 相关处理

### 2.指令相关的处理（简单了解）

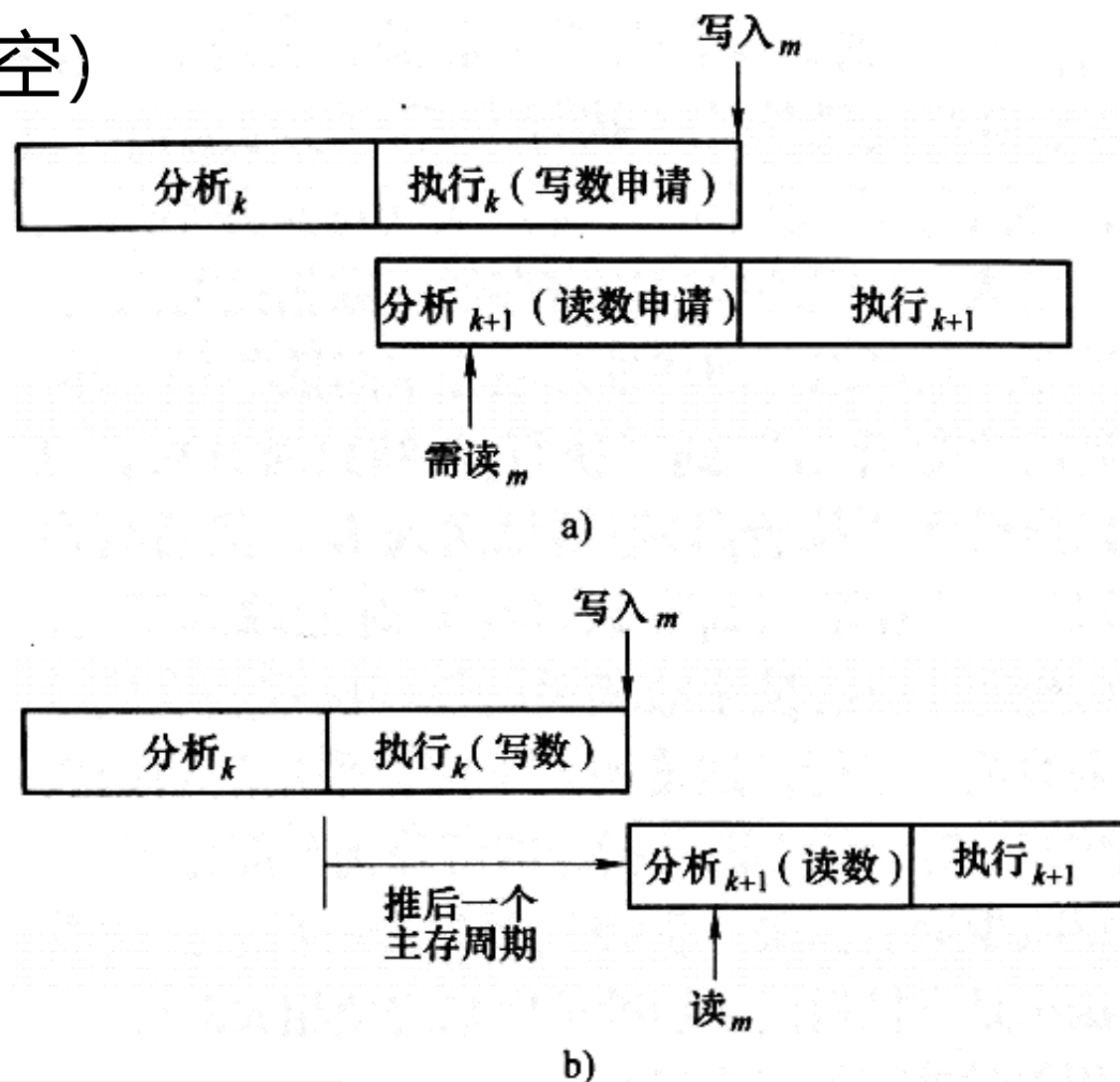
对于有指缓的计算机，由于指令是提前从主存取进指缓的，为了判定是否发生了指令相关，需要对多条指令地址与多条指令的运算结果地址比较，看是否有相同的，这是很复杂的。



## 5.1.2 相关处理

### 3.主存空间数相关的处理（单选、填空）

主存空间数相关是相邻两条指令之间出现对主存**同一单元**要求**先写而后读**的关联。如图5-4a所示。如果让“执行 $k$ ”与“分析 $k+1$ ”在时间上重叠，就会使“分析 $k+1$ ”读出的数不是第 $k$ 条指令执行完应写入的结果而出错。要想不出错，只有推后“分析 $k+1$ ”的读。



## 5.1.2 相关处理

### 4.通用寄存器组相关的处理（简单了解）

一般的计算机中，通用寄存器除了存放源操作数、运算结果外，也可能存放形成访存操作数物理地址的变址值或基址值。因此，通用寄存器组的相关又有操作数的相关和变址值或基址值的相关两种。

或

操作码	$L_1$	$L_3$	$B_2$	$d_2$
-----	-------	-------	-------	-------

操作码	$L_1$	$L_3$		$L_2$
-----	-------	-------	--	-------

## 5.1.2 相关处理

### 4.通用寄存器组相关的处理（单选、填空）

推后“分析 $k+1$ ”和设置“相关专用通路”是解决重叠方式相关处理的两种基本方法。前者是以降低速度为代价，使设备基本上不增加。

或

操作码	$L_1$	$L_3$	$B_2$	$d_2$
-----	-------	-------	-------	-------

操作码	$L_1$	$L_3$		$L_2$
-----	-------	-------	--	-------



那么意气风发地  
走在成功的道路上

## 真题练练手

1、关于指令执行的“一次重叠”，下列说法正确的是（ ） 1910

- A: “一次重叠”是同时解释很多条指令
- B: “一次重叠”是同时解释相邻两条指令
- C: “一次重叠”是同时解释一条指令
- D: “一次重叠”是解释两条指令



那么意气风发地  
走在成功的道路上

## 真题练练手

1、关于指令执行的“一次重叠”，下列说法正确的是（ ） 1910

- A: “一次重叠” 是同时解释很多条指令
- B: “一次重叠” 是同时解释相邻两条指令
- C: “一次重叠” 是同时解释一条指令
- D: “一次重叠” 是解释两条指令

答案： B





那么意气风发地  
走在成功的道路上

## 真题练练手

2、指令间“一次重叠”是指 ( ) 0707

- A: “取指 $k+1$ ”与“分析 $k$ ”重叠
- B: “分析 $k+1$ ”与“执行 $k$ ”重叠
- C: “分析 $k$ ”与“执行 $k+1$ ”重叠
- D: “执行 $k$ ”与“取指 $k+1$ ”重叠



那么意气风发地  
走在成功的道路上

## 真题练练手

2、指令间“一次重叠”是指 ( ) 0707

A: “取指 $k+1$ ”与“分析 $k$ ”重叠

B: “分析 $k+1$ ”与“执行 $k$ ”重叠

C: “分析 $k$ ”与“执行 $k+1$ ”重叠

D: “执行 $k$ ”与“取指 $k+1$ ”重叠

答案: B



那么意气风发地  
走在成功的道路上

## 真题练练手

3、标量处理机顺序解释的优点是 ( ) 1810

A:控制简单

B:速度快

C:效率高

D:利用率离



那么意气风发地  
走在成功的道路上

## 真题练练手

3、标量处理机顺序解释的优点是 ( ) 1810

A:控制简单

B:速度快

C:效率高

D:利用率离

答案：A



那么意气风发地  
走在成功的道路上

## 真题练练手

4、对指令间“一次重叠”描述不正确的是（） 1804

A:仅“执行k”与“分析k+1”重叠

B:“分析k+1”完成后立即开始“执行k+1”

C:应尽量使“分析k+1”与“执行k”时间相等

D:只需要一套指令分析部件和执行部件



那么意气风发地  
走在成功的道路上

## 真题练练手

4、对指令间“一次重叠”描述不正确的是（） 1804

A:仅“执行k”与“分析k+1”重叠

B:“分析k+1”完成后立即开始“执行k+1”

C:应尽量使“分析k+1”与“执行k”时间相等

D:只需要一套指令分析部件和执行部件

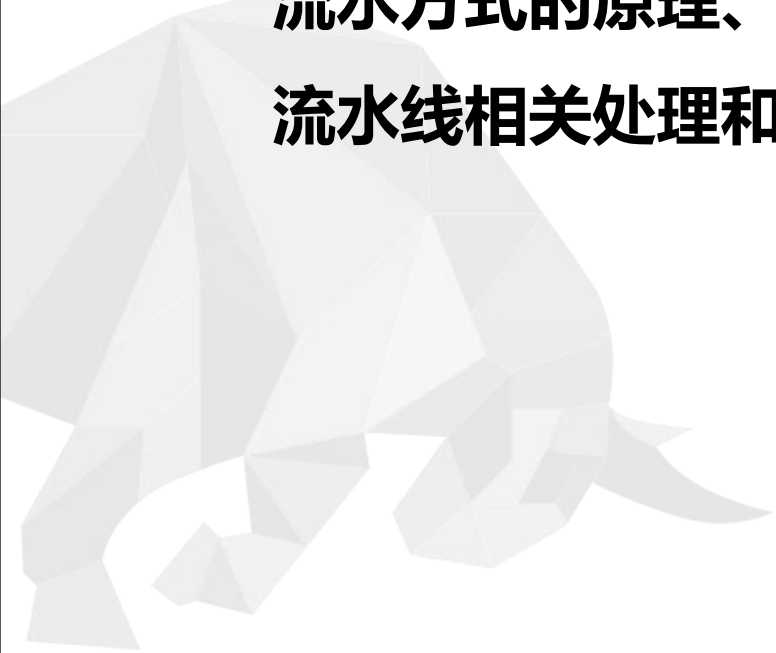
答案：B

## 5.2 流水方式

**本节主要内容：**

**流水方式的原理、分类（识记=单选、填空）**

**流水线相关处理和性能瓶颈消除（领会=简答）**



## 5.2.1 基本概念

### 1. 工作原理

“分析 $k+1$ 与“执行 $k$ ”的一次重叠是把指令的解释过程分解成“分析”与“执行”两个子过程，在独立的分析部件和执行部件上时间重叠地进行。

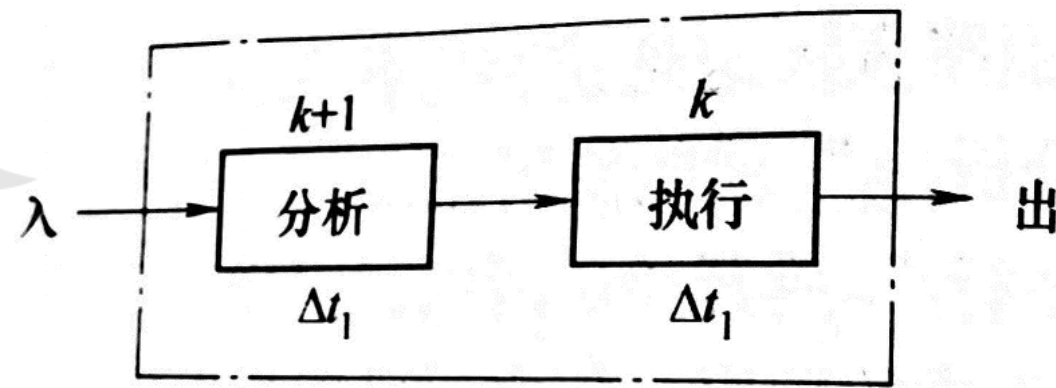
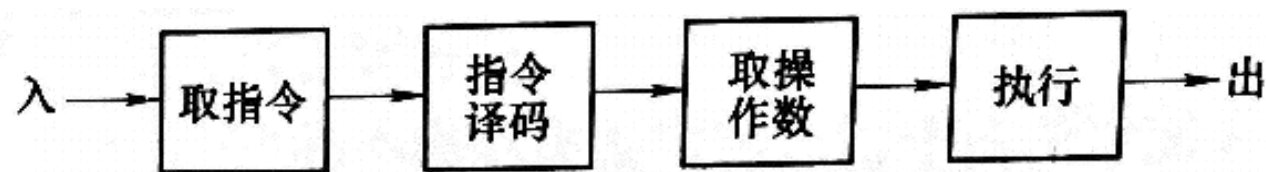


图5-11 指令分解为“分析”与“执行”子过程

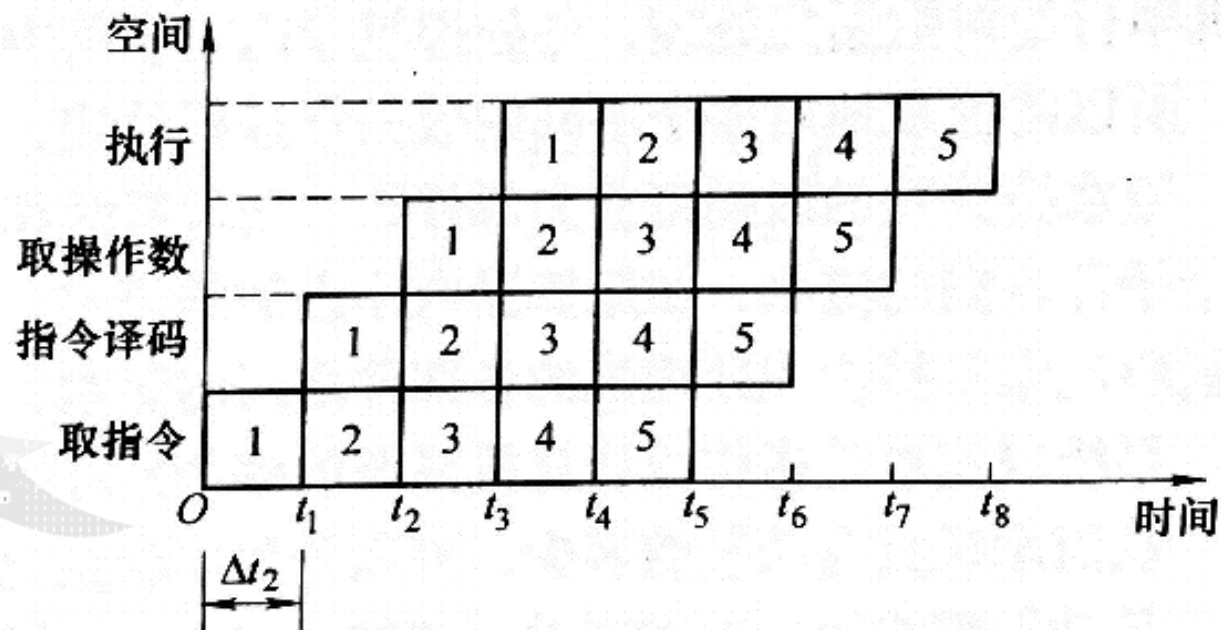


## 5.2.1 基本概念

### 1. 工作原理



a)



b)

图5-12 流水处理

a) 指令解释的流水处理    b) 流水处理的时 (间) 空 (间) 图

## 5.2.1 基本概念

### 2.流水的分类（填空）

流水按处理的级别可分为**部件级**、**处理机级**和**系统级**。

从流水线具有功能的多少，可以分为**单功能流水线**和**多功能流水线**。

## 5.2.1 基本概念

### 2.流水的分类（单选、填空）

按多功能流水线的各段能否允许同时用于多种不同功能连接流水，可把流水线分为静态流水线和动态流水线。

静态流水线在某一时间内各段只能按一种功能连接流水，只有等流水线全部流空后，才能切换成按另一种功能连接流水。

动态流水线的各功能段在同一时间内可按不同运算或功能连接。

## 5.2.1 基本概念

### 2.流水的分类（单选）

从计算机所具有的数据表示角度，可以把流水线处理机分为**标量流水机**和**向量流水机**。

标量流水机没有向量数据表示，只能用标量循环方式来处理向量和数组，如**Amdahl 470 V/6**及后面要介绍的IBM 360/91。向量流水机指的是计算机有向量数据表示，设置有向量指令和向量运算硬件，能流水地处理向量和数组中的各个元素。向量流水机是向量数据表示和流水技术的结合

## 5.2.1 基本概念

### 2.流水的分类（单选）

从流水线中各功能段之间是否有反馈回路的角度，可以把流水线分为**线性**流水线和**非线性**流水线。



## 5.2.2标量流水线的主要性能（填空、选择）


标量流水处理机的性能主要是

- 1.吞吐率TP和加速比SP
- 2.效率

## 5.2.2 标量流水线的主要性能

### 1. 吞吐率TP和加速比SP

吞吐率是流水线单位时间里能流出的任务数或结果数。


$$T_p = \frac{n}{m\Delta t + (n-1)\Delta t}$$

## 5.2.2标量流水线的主要性能

### 1.吞吐率TP和加速比SP

【1410真题】流水线由4个功能部件组成，每个功能部件的延迟时间为 $\Delta t$ ，当输入5个数据后，间歇 $5\Delta t$ 又输入5个数据，如此周期性地工作，画出时空图，并求此时流水线的吞吐率。

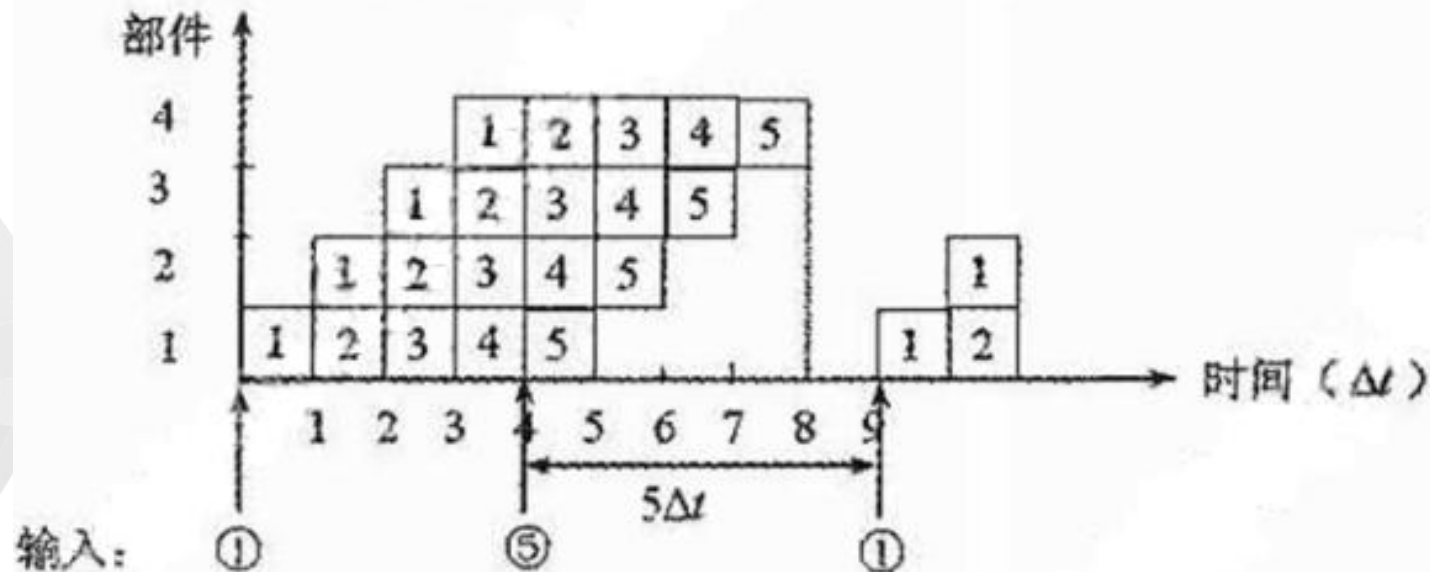


## 5.2.2标量流水线的主要性能

### 1.吞吐率TP和加速比SP

$$T_p = \frac{n}{m\Delta t + (n-1)\Delta t}$$

时空图如 图所示。



按周期性,流水线吞吐率为:  $T_p = \frac{5}{8\Delta t}$

## 5.2.2标量流水线的主要性能

### 2.效率（综合）

【1304真题】向量  $A$  和  $B$  各有 6 个元素，计算向量点积  $A \cdot B = \sum_{i=1}^6 a_i * b_i$ 。

1. 若在顺序方式下，一次“加”需  $4\Delta t$ ，一次“乘”需  $3\Delta t$ ，求执行完  $A \cdot B$  所需的时间；

## 5.2.2标量流水线的主要性能

### 2.效率（综合）

【1304真题】向量  $A$  和  $B$  各有 6 个元素，计算向量点积  $A \cdot B = \sum_{i=1}^6 a_i * b_i$ 。

1. 若在顺序方式下，一次“加”需  $4\Delta t$ ，一次“乘”需  $3\Delta t$ ，  
求执行完  $A \cdot B$  所需的时间；  
需  $3\Delta t \times 6 + 4\Delta t \times 5 = 38\Delta t$



那么意气风发地  
走在成功的道路上

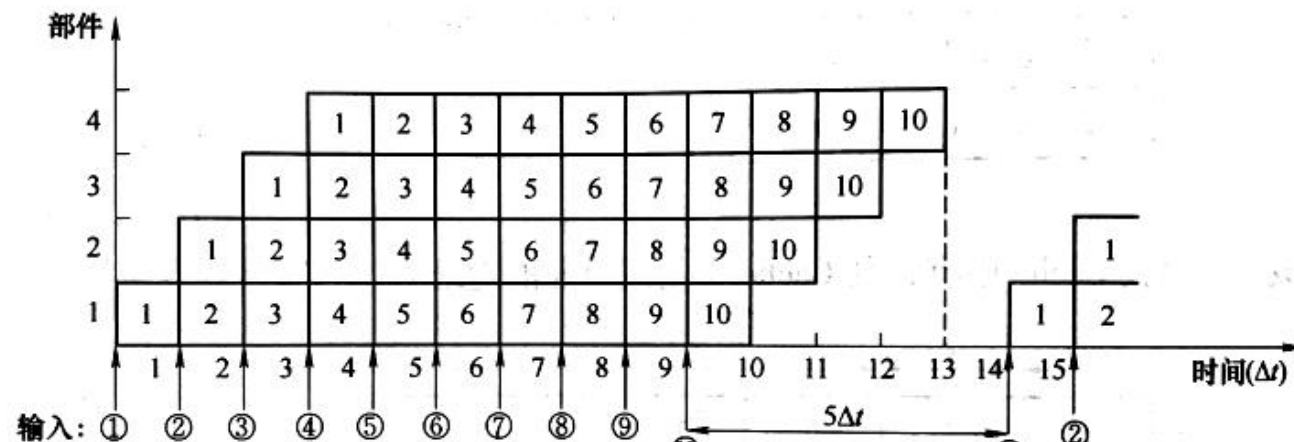
## 真题练练手

2、流水线由4个功能部件组成，每个功能部件的延迟时间为，当输入10个数据后，间接5又输入10个数据，如此周期性地工作，求此时流水线的吞吐率，并画出其时空图。



那么意气风发地  
走在成功的道路上

## 真题练练手



2、流水线由4个功能部件组成，每个功能部件的延迟时间为，当输入10个数据后，间接5又输入10个数据，如此周期性地工作，求此时流水线的吞吐率，并画出其时空图。

答案：按题意可得4个功能部件流水时的时空关系如附图15所示  
所以，按周期性工作时的流水线平均吞吐率为：

$$T_p = \frac{10}{13\Delta t}$$



尚德机构

# ▶ 答疑时间 ◀





尚德机构

▶ THANK YOU ◀

