

计算机体系结构

习题解答

第1章 基础知识

■ 翻译

- 先用转换程序将高级机器级上的程序整个地变换成低级机器级上可运行的等效程序，然后再在低级机器级上去实现的技术

第1章 基础知识

■ 解释

- 在低一级机器级上用它的一串语句或指令来仿真高一级机器上的一条语句或指令的功能，通过对高一级机器语言程序中的每条语句或指令逐条解释来实现的技术

第1章 基础知识

■ 计算机系统结构

- 程序员所看到的计算机的属性，即程序员为编写出能在机器上正确运行的程序所必须了解到的机器的**概念性结构和功能特性**。从层次结构出发，计算机系统结构指层次结构中传统机器级的系统结构，研究的是软、硬件之间的功能分配，以及对传统机器级界面的定义

第1章 基础知识

■ 透明性

- 本来存在的事物或属性，从某个角度看却好象不存在

■ 软件兼容

- 程序可以不加修改地运行在各档机器上，区别仅在于运行时间不同
- 它分为向上兼容和向下兼容，以及向前兼容和向后兼容

第1章 基础知识

■ 模拟

- 用一种机器(A)的机器语言解释实现另一种机器(B)的指令系统,使A具有B的指令系统,从而实现软件移植的方法。被模拟的机器称为虚拟机,模拟用的机器称为宿主机,解释程序称为模拟程序

第1章 基础知识

■ 仿真

- 用一种机器(A)的微程序直接解释实现另一种机器(B)的指令系统，从而实现软件移植的方法
- 被仿真的机器称为目标机，进行仿真的机器称为宿主机，解释微程序机器称为仿真微程序

第1章 基础知识

■ 并行性

- 指可以同时进行运算或操作的特性，它有两重含义：

- 同时性：同一时刻发生

- 并发性：同一时间间隔内发生

第1章 基础知识

■ 系列机

- 同一厂家生产的具有相同系统结构、但有不同组成和实现的一系列不同型号的机器

兼容机

- 不同厂家生产的具有相同系统结构、但有不同组成和实现的一系列不同型号的机器

第1章 基础知识

■ 紧耦合系统

- 多台计算机之间通过总线或高速开关互连，共享主存，并有较高的信息传输速率的多机系统

■ 时间重叠

- 引入时间概念，让多个处理过程轮流使用同一套硬件设备的各部分，基本上不需要重复设置硬件设备

第1章 基础知识

■ 1-1

第4级	$N^3/M^3 * K秒$
第3级	$N^2/M^2 * K秒$ （同理）
第2级	$N/M * K秒$ ($X/M * N * 1/X * K = N/M * K秒$)
第1级	$K秒$ （设为X条， 每条需要 $1/X * K秒$ ）

第1章 基础知识

■ 1-6

■ 透明的是：

指令缓冲器

时标发生器

乘法器

主存地址寄存器

先行进位链

移位器

第1章 基础知识

■ 1-8

(1) 可以。因为它虽然是属于计算机系统结构的内容。新增加的数据类型和指令，不会影响到已有指令所写的程序正确运行，只是现在用新增加的指令来写程序，会使计算机的性能和效率变得更好。

第1章 基础知识

(2) 不可以。中断的分级和中断的响应次序等中断机构都属于计算机系统结构的内容。中断分级由原来的4级增加到5级应当还是允许的，关键是重新调整了中断响应的优先次序，这就使原有程序的中断响应次序发生了改变，会影响原有程序工作的正确性。

第1章 基础知识

(3)可以。Cache存储器属于计算机组成，它不会改变原有的系统程序和应用程序，不会影响到它们的正确性。只是有了它之后，系统的性能会显著的提高。

(4)可以。浮点数尾数的下溢处理不属于计算机系统结构，而是计算机组成设计所需要考虑的内容。

第1章 基础知识

(5)不可以。指令的操作码、字段格式、寻址方式等都是计算机系统结构的内容。如果将它们改变，就会直接影响以前编写的程序不能正确运行。

(6)可以。数据通路宽度是计算机组成的内容。

第1章 基础知识

(7)可以。单总线改为多总线是计算机组成的内容。

(8)不可以。通用寄存器的使用是属于计算机系统结构的内容。0号通用寄存器改为堆栈指示器，将使原先程序中0号寄存器中的内容改变直接影响到堆栈指针的位置发生变化，造成程序无法正常工作。

1-9 如果某一计算任务用向量方式求解比用标量方式求解要快20倍，称可用向量方式求解部分所花费时间占总的百分比为可向量化百分比。请写出加速比与可向量化比例两者的关系曲线。

解：

$$S = \frac{1}{(1-F) + F/20} = \frac{20}{20 - 19 \times F}$$

在习题1-9中，为达到加速比2，可向量化的百分比应为多少？

解：

$$2 = \frac{1}{(1 - F) + (F / 20)}$$

$$F = \frac{10}{19} = 0.53$$

1-10 用一台40MHz处理机执行标准测试程序，它含的混合指令数和相应所需的时钟周期数如下：

指令类型	指令数	时钟周期数
整数运算	45000	1
数据传送	32000	2
浮点	15000	2
控制传送	8000	2

求有效CPI、MIPS速率和程序的执行时间。

[解答]

$$\begin{aligned} \text{CPI} &= \sum_{i=1}^n (\text{CPI}_i * \frac{I_i}{IC}) \\ &= (45000 * 1 + 32000 * 2 + 15000 * 2 + 8000 * 2) / \\ &\quad (45000 + 32000 + 15000 + 8000) = 1.55 \text{ 周期/指令} \end{aligned}$$

$$\text{速率}_{\text{MIPS}} = f / (\text{CPI} * 10^6) = 40 * 10^6 / (1.55 * 10^6) = 25.81$$

$$\begin{aligned} t &= (45000 * 1 + 32000 * 2 + 15000 * 2 + 8000 * 2) * (\frac{1}{40 * 10^6}) \\ &= 3.875 * 10^{-3} \text{ 秒} \end{aligned}$$

1-11 假设在一台40MHz处理机上运行200,000条指令的目标代码，程序主要由四种指令组成。根据程序跟踪实验结果，已知指令混合比和每种指令所需的指令数如下：

指令类型	CPI	指令混合比
算术和逻辑	1	60%
高速缓存命中的加载/存储	2	18%
转移	4	12%
高速存储缺失的存储器访问	8	10%

(a)计算在单处理机上用上述跟踪数据运行程序的平均CPI。

(b)根据(a)所得的CPI,计算相应的MIPS速率。

第1章

[解答]

$$(a) \text{ CPI} = 1 * 60\% + 2 * 18\% + 4 * 12\% + 8 * 10\% \\ = 2.24$$

$$(b) \text{ MIPS} = \frac{f}{\text{CPI} * 10^6} = \frac{40\text{MHz}}{2.24 * 10^6} = 17.86$$

第2章 指令系统

■ 数据表示

- 由硬件直接识别和引用的数据类型

第2章 指令系统

■ 动态再定位

- 动态再定位采用基址寻址方法。当把程序装入主存时，对程序不做任何修改（变换），直接装入主存，同时将程序在主存中的起始地址 a 存入对应该道程序的基址寄存器。在执行指令时，通过地址加法器将逻辑地址加上基址寄存器的内容(程序基点地址)，形成物理地址，然后访存

第2章 指令系统

■ 功能型指令

- 真正执行数据变换的指令，如加、减、乘、除等

■ 哈夫曼压缩概念

- 当各种事件发生的概率不均等时，使用概率高的事件用短代码表示，使用概率低的事件用长代码表示，就会使平均位数缩短。哈夫曼编码是最优化的编码方法

第2章 指令系统

■ CISC

- 增强原有指令的功能，设置更为复杂、但功能更强的新指令以取代原先由软件子程序完成的功能，实现软件功能的硬化。按这种途径和方向发展，会使机器的指令系统越来越庞大和复杂，因此称采用这种途径设计而成的CPU的计算机为复杂指令集计算机CISC

第2章 指令系统

■ RISC

- 减少指令总数，简化指令功能，以降低硬件设计的复杂度，提高指令的执行速度。按这种途径和方向发展，会使机器的指令系统精炼简单，因此称采用这种途径设计而成的CPU的计算机为精简指令集计算机
RISC

第2章 指令系统

2-1

- **数据表示**是指由硬件直接识别和引用的数据类型
- **数据结构**反映的是各种数据元素或信息单元之间的结构关系

第2章 指令系统

- 数据结构是通过软件映像成机器所具有的各种数据表示实现的
- 数据表示是数据结构的组成元素
- 数据表示为数据结构提供不同程度的支持，反映在效率和方便程度的不同
- 因此，数据结构与数据表示是软、硬交界面

第2章 指令系统

习题2-4

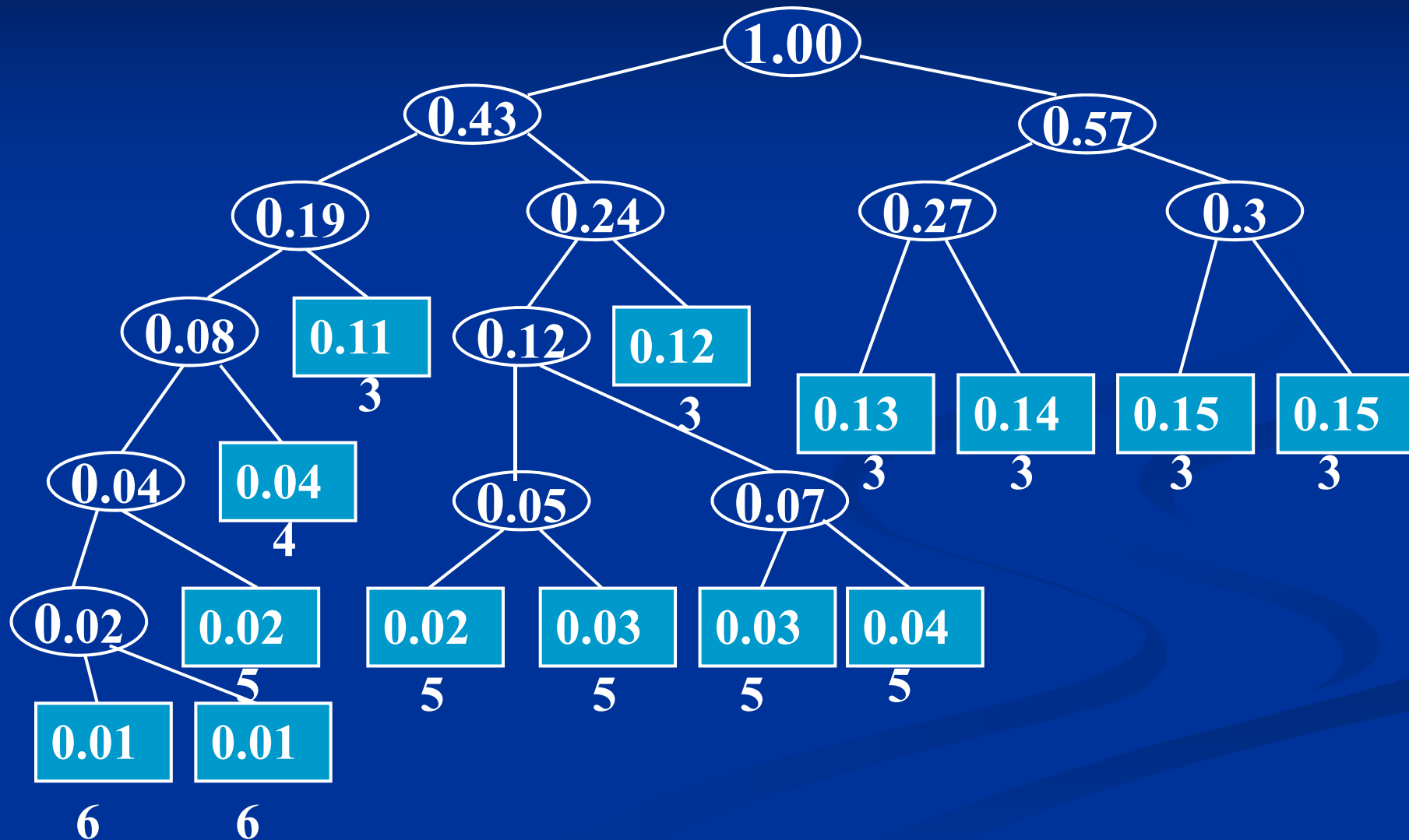
1、等长二进制编码

平均码长为 $\lceil \log_2 14 \rceil + 1 = 4$

2、Huffman编码

首先构造Huffman树

第2章 指令系统



第2章 指令系统

平均码长为

$$\begin{aligned}\Sigma P_i * L_i &= (0.01 + 0.01) \times 6 \\ &+ (0.02 + 0.02 + 0.03 + 0.03 + 0.04) \times 5 \\ &+ 0.04 \times 4 + (0.11 + 0.12 + 0.13 \\ &+ 0.14 + 0.15 + 0.15) \times 3 \\ &= 3.38\end{aligned}$$

第2章 指令系统

■ 3、只有两种码长的扩展操作码编码

使用频度高的用短代码表示。根据 **Huffman** 编码和指令使用频度，可以设计出具有3位和6位码长的操作码。操作码编码如下：

第2章 指令系统

编码	指令使用频度	编码	指令使用频度
000	0.15	111000	0.04
001	0.15	111001	0.03
010	0.14	111010	0.03
011	0.13	111011	0.02
100	0.12	111100	0.02
101	0.11	111101	0.01
110	0.04	111110	0.01

第2章 指令系统

平均码长为

$$\begin{aligned}\Sigma P_i * L_i &= (0.15 + 0.15 + 0.14 + 0.13 \\ &+ 0.12 + 0.11 + 0.04) \times 3 \\ &+ (0.04 + 0.03 + 0.03 + 0.02 \\ &+ 0.02 + 0.01 + 0.01) \times 6 \\ &= 3.48\end{aligned}$$

3-5位操作码编码如下:

编码	指令使用频度	编码	指令使用频度
000	0.15	11000	0.04
001	0.15	11001	0.04
010	0.14	11010	0.03
011	0.13	11011	0.03
100	0.12	11100	0.02
101	0.11	11101	0.02
		11110	0.01
		11111	0.01

第2章 指令系统

平均码长为

$$\begin{aligned}\Sigma P_i * L_i &= (0.15 + 0.15 + 0.14 + 0.13 \\ &+ 0.12 + 0.11) \times 3 \\ &+ (0.04 + 0.04 + 0.03 + 0.03 + 0.02 \\ &+ 0.02 + 0.01 + 0.01) \times 5 \\ &= 3.40\end{aligned}$$

2-5 一个处理机共有10条指令，各指令在程序中出现的概率如下：

指令	概率	Huffman	2/8扩展	3/7扩展
I1	0.25			
I2	0.20			
I3	0.15			
I4	0.10			
I5	0.08			
I6	0.08			
I7	0.05			
I8	0.04			
I9	0.03			
I10	0.02			

- (1) 计算这10条的操作码最短平均长度。
- (2) 采用Huffman编码法编写这10条指令的操作码，并计算操作码的平均长度和信息冗余量。
- (3) 采用2/8扩展编码法编写这10条指令的操作码，并计算操作码的平均长度和信息冗余量。
- (4) 采用3/7扩展编码法编写这10条指令的操作码，并计算操作码的平均长度和信息冗余量。

第2章

(1) 最优Huffman

$$H = - \sum P_i \times \log_2 P_i$$

$$= 0.25 \times 2 + 0.20 \times 2.322 + 0.15 \times 2.737 +$$

.....

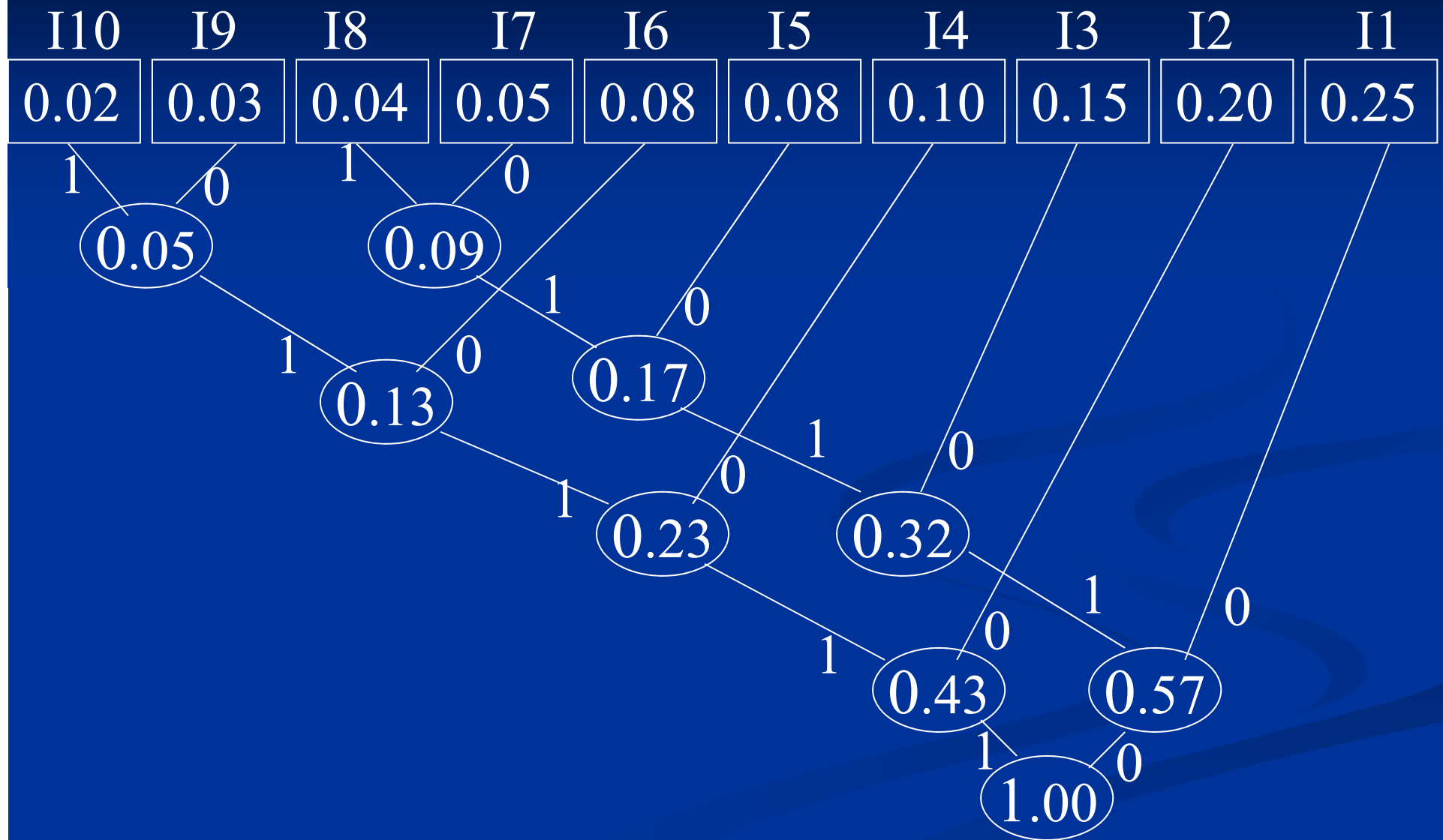
$$= 2.96$$

(2) 采用最小概率合并法

$$H' = \sum P_i \times L_i$$

$$\begin{aligned} &= 0.25 \times 2 + 0.20 \times 2 + 0.15 \times 3 + 0.10 \times 3 + \\ &\quad 0.08 \times 4 + 0.08 \times 4 + 0.05 \times 5 + 0.04 \times 5 + \\ &\quad 0.03 \times 5 + 0.02 \times 5 \\ &= 2.99 \end{aligned}$$

$$\text{信息冗余} = 1 - 2.96/2.99 = 1\%$$



指令	概率	Huffman
----	----	---------

I1	0.25	00
----	------	----

I2	0.20	10
----	------	----

I3	0.15	010
----	------	-----

I4	0.10	110
----	------	-----

I5	0.08	0110
----	------	------

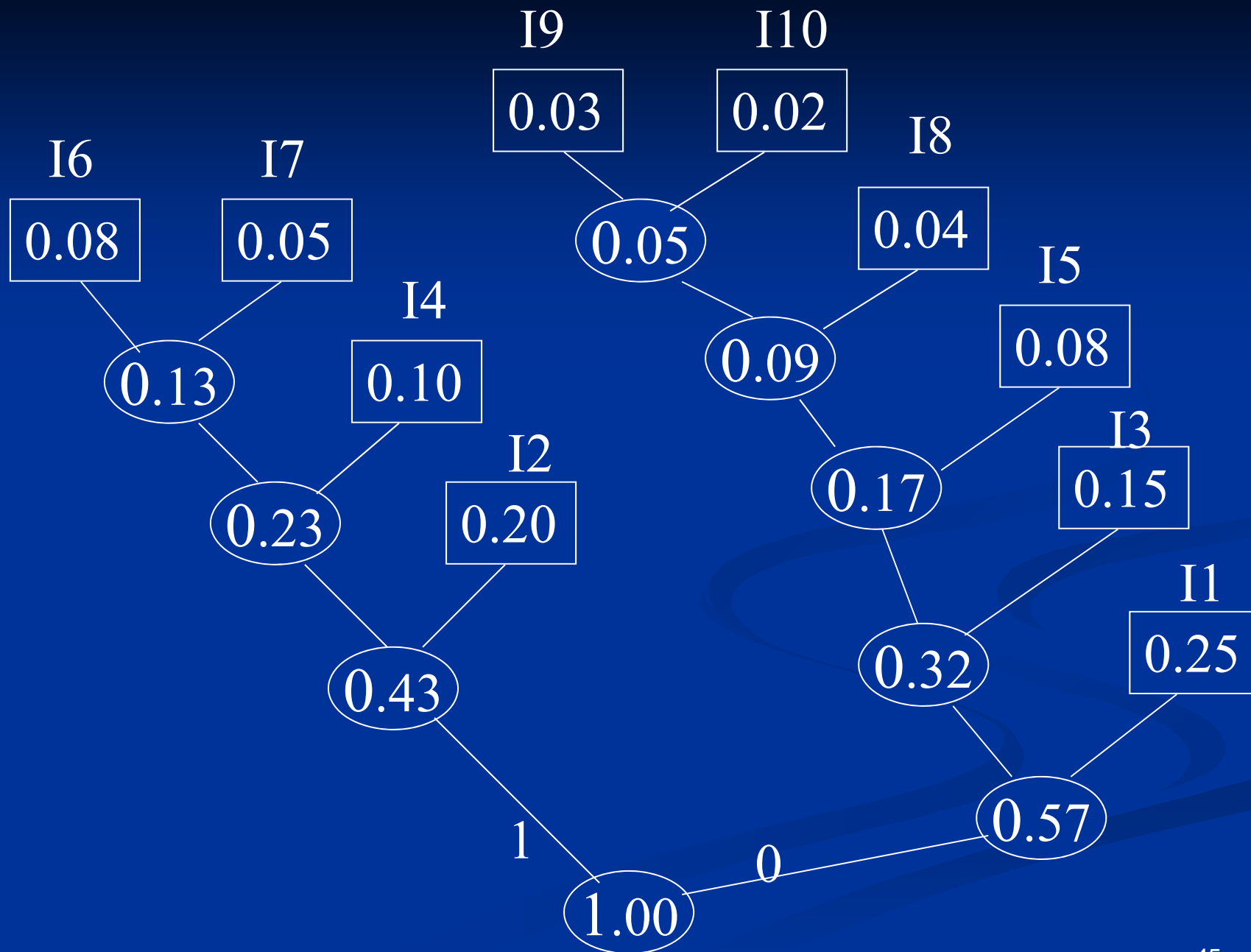
I6	0.08	1110
----	------	------

I7	0.05	01110
----	------	-------

I8	0.04	01111
----	------	-------

I9	0.03	11110
----	------	-------

I10	0.02	11111
-----	------	-------



(2) 采用最小概率合并法

$$H' = \sum P_i \times L_i$$

$$\begin{aligned} &= 0.25 \times 2 + 0.20 \times 2 + 0.15 \times 3 + 0.10 \times 3 + \\ &\quad 0.08 \times 4 + 0.08 \times 4 + 0.05 \times 4 + 0.04 \times 5 + \\ &\quad 0.03 \times 6 + 0.02 \times 6 \\ &= 2.99 \end{aligned}$$

$$\text{信息冗余} = 1 - 2.96/2.99 = 1\%$$

指令	概率	Huffman
----	----	---------

I1	0.25	00
I2	0.20	10
I3	0.15	010
I4	0.10	110
I5	0.08	0110
I6	0.08	1111
I7	0.05	1110
I8	0.04	01110
I9	0.03	011111
I10	0.02	011110

(3) 采用2/8扩展法

$$H = \sum P_i \times L_i$$

$$= (0.25 + 0.20) \times 2 +$$

$$(0.15 + 0.10 + 0.08 + 0.08 + 0.05 + 0.04 +$$

$$0.03 + 0.02) \times 4$$

$$= 3.1$$

$$\text{信息冗余} = 1 - 2.96 / 3.1 = 4.5\%$$

指令 概率 2/8扩展

I1	0.25	00
I2	0.20	01
I3	0.15	1000
I4	0.10	1001
I5	0.08	1010
I6	0.08	1011
I7	0.05	1100
I8	0.04	1101
I9	0.03	1110
I10	0.02	1111

(4) 采用3/7扩展法

$$H = \sum P_i \times L_i$$

$$\begin{aligned} &= (0.25 + 0.20 + 0.15) \times 2 + \\ &\quad (0.10 + 0.08 + 0.08 + 0.05 + 0.04 + \\ &\quad 0.03 + 0.02) \times 5 \\ &= 3.2 \end{aligned}$$

$$\text{信息冗余} = 1 - 2.96 / 3.2 = 7.5\%$$

指令	概率	3/7扩展
----	----	-------

I1	0.25	00
I2	0.20	01
I3	0.15	10
I4	0.10	11000
I5	0.08	11001
I6	0.08	11010
I7	0.05	11011
I8	0.04	11100
I9	0.03	11101
I10	0.02	11110

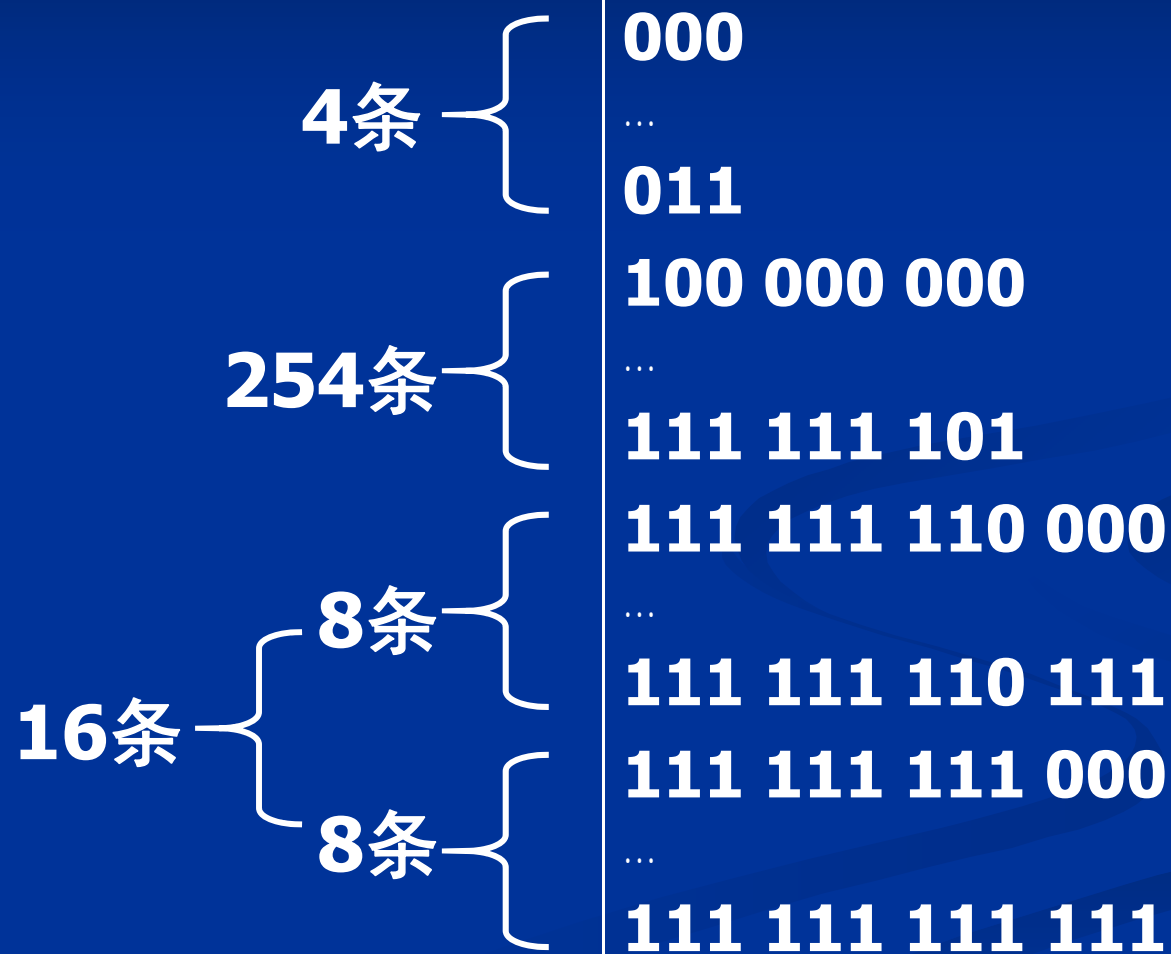
指令	概率	Huffman	2/8扩展	3/7扩展
I1	0.25	00	00	00
I2	0.20	10	01	01
I3	0.15	010	1000	10
I4	0.10	110	1001	11000
I5	0.08	0110	1010	11001
I6	0.08	1110	1011	11010
I7	0.05	01110	1100	11011
I8	0.04	01111	1101	11100
I9	0.03	11110	1110	11101
I10	0.02	11111	1111	11110

第2章 指令系统

习题2-6

- 已知指令字长为**12位**，地址码长为**3位**
- 则三地址指令的操作码码长为**3位**，单地址指令的操作码码长为**9位**，零地址指令的操作码码长为**12位**
- 因此，只要按**3-9-12**扩展操作码进行编码，看是否能容纳所有指令即可。
一种可能的编码方案如下：

第2章 指令系统



第2章 指令系统

- 可以看出，三地址指令和零地址指令不存在问题，但无法容纳**255**条单地址指令。但能容纳**254**条单地址指令

第3章 总线、中断、输入输出系统

3-3

根据题意，

$$T_s = 9.8\mu s, T_D = 0.2\mu s$$

因此，该通道的极限流量为

$$1 / (T_s + T_D) = 1 / (9.8 + 0.2) = 100 \text{KBps}$$

低速设备每隔500 μs 发送一个字节数据传送请求，因此，每个设备的实际流量为

$$1 / 500 = 2 \text{KBps}$$

第3章 总线、中断、输入输出系统

根据 $f_{\max} \geq f_{\text{实际}}$ ，该通道最多可以连接

$$100/2=50 \text{ 台}$$

这样的低速设备

第3章 总线、中断、输入输出系统

根据题意，该通道的极限流量为

$$f_{\max} =$$

$$1024/9.8 + 1024 * 0.2 = 1/0.2096$$

对比给出的数据传送请求的时间间隔，可以确定：

B、C、E、F 可以挂在此通道上，

A、D 不可以（比**0.2096**快）

3-8 一个字节多路通道连接有5台设备，它们的数据传输率如下表：

设备名称	D1	D2	D3	D4	D5
数据传输速率 (KB/S)	100	33.3	33.3	20	10
服务优先级	1	2	3	4	5

- (1) 计算这个字节多路通道的实际工作量。
- (2) 为了使通道能够正常工作，请设计通道的最大流量和工作周期。
- (3) 当这个字节多路通道工作在最大流量时，5台设备都在0时刻同时向通道发出第一次传送数据的请求，并在以后的时间里按照各自的数据传输速率连续工作。画出通道分时为各台设备服务的时间关系图，并计算这个字节多路通道处理完各台设备的第一次数据服务请求的时刻。

第3章

解:

(1) 实际工作流量

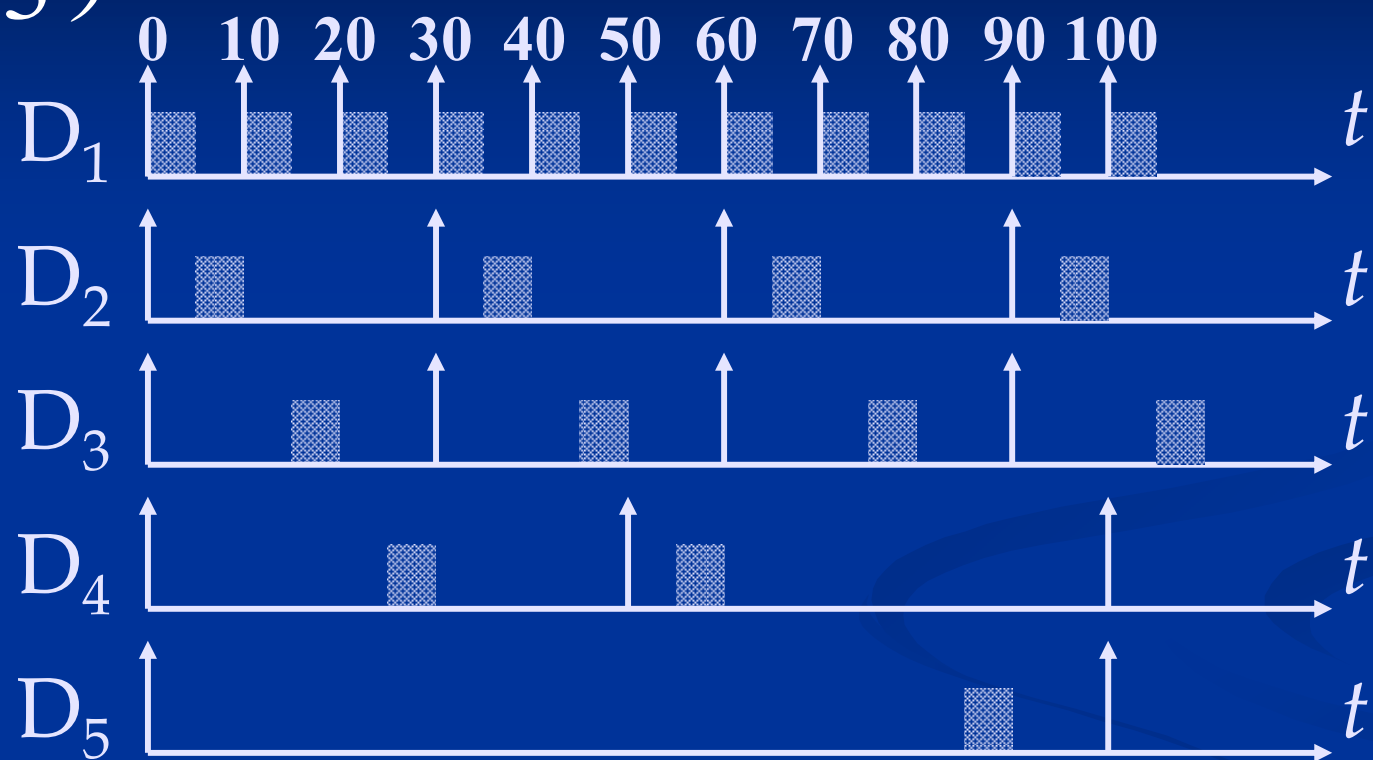
$$f_{\text{BYTE}} = 100 + 33.3 + 33.3 + 20 + 10 = 196.6 \text{ KB/s}$$

(2) 最大流量 = 200 KB/s

$$\text{工作周期 } t = 1 / f_{\text{BYTE}} = 1 / 200 \text{ K} = 5 \mu\text{s}$$

第3章

(3)



通道处理完各台设备的第一次数据服务请求的时刻分别为： $5\mu s$ ， $10\mu s$ ， $20\mu s$ ， $30\mu s$ ， $90\mu s$ 。

3-9 一个字节多路通道连接有4台设备，每台设备发出输入输出服务请求的时间间隔、它们的服务优先级和发出第一次服务请求的时刻如下表：

设备名称	D1	D2	D3	D4
发服务请求间隔	10 μ s	75 μ s	15 μ s	50 μ s
服务优先级	1	4	2	3
发出第一次请求时间	0 μ s	70 μ s	10 μ s	20 μ s

- (1) 计算这个字节多路通道的实际流量和工作周期。
- (2) 在数据传送期间，如果通道选择一次设备的时间为 $3\mu\text{s}$ ，传送一个字节的时间为 $2\mu\text{s}$ ，画出这个字节多路通道响应各设备请求和为设备服务的时间关系图。
- (3) 从(2)时间关系图中，计算通道处理完成各设备第一次服务请求的时刻。
- (4) 从(2)时间关系图中看，这个字节多路通道能否正常工作？
- (5) 在设计一个字节多路通道的工作流量时，可以采用哪些措施来保证通道能够正常工作？

解:

(1) 数据传输率 $D1=1/10=100\text{KB/s}$

$$D2=1/75=13.3\text{KB/s}$$

$$D3=1/15=66.7\text{KB/s}$$

$$D4=1/50=20\text{KB/s}$$

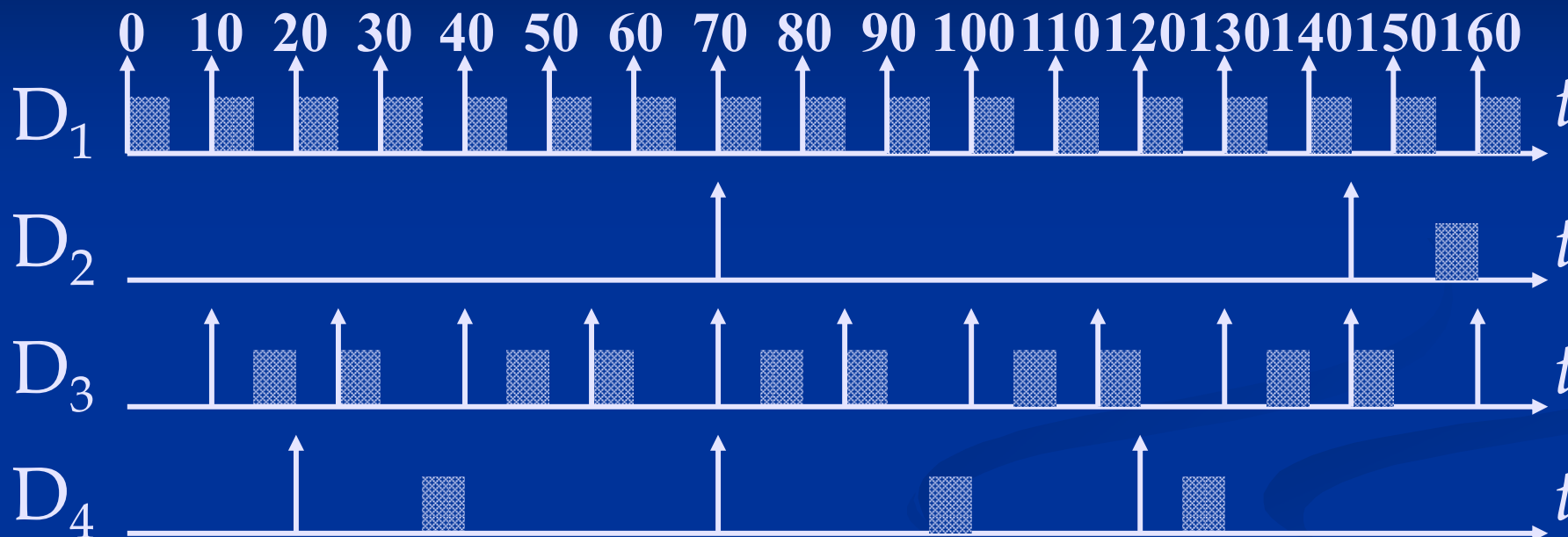
实际工作流量

$$f_{\text{BYTE}}=100+13.3+66.7+20=200\text{KB/s}$$

工作周期 $t=1/f_{\text{BYTE}}=1/200\text{K}=5\mu\text{s}$

第3章

(2)



(3) 通道处理完各台设备的第一次数据服务请求的时刻分别为： $5\mu s$ ， $160\mu s$ ， $20\mu s$ ， $40\mu s$ 。

(4) 此字节多路通道不能正常工作，因为D2的第一次服务请求被响应之前，设备已经发出第二次服务请求，这就可能导致丢失数据。

(5) 可以采用3种措施来保证通道能够正常工作：

增加通道的最大流量。

动态改变设备的优先级。

增加一定的数据缓冲器。

第4章 存储体系

■ 地址的变换

- 程序按某种地址映像关系装入主存后，运行时如何把多用户虚地址变换成实地址

第4章 存储体系

■ 存储层次（体系）

- 利用多种不同的存储器，形成M1至Mn的不同层次的存储器系统。并利用I/O处理机和操作系统、硬件等技术，由增设的辅助软、硬件自动完成地址变换、程序定位等工作，使这些存储器形成一个整体。

第4章 存储体系

■ 存储层次（体系）

- 从整体上看，速度接近M1，容量是Mn的，位价格接近廉价的Mn。将这样的存储器系统称为存储体系。

第4章 存储体系

■ 程序局部性

- 程序在执行时所用到的指令和数据的分布不是随机的，而是相对地簇聚成块或页。它包括时间局部性和空间局部性。

第4章 存储体系

- **时间局部性**是指最近的未来要用到的信息可能就是当前正在使用的信息——这是由程序的循环造成的
- **空间局部性**是指最近的未来要用到的信息可能就是当前信息的相邻信息——这是由程序的顺序执行造成的

第4章 存储体系

4-1

- 在页式虚拟存储器中，若所要访问的虚页不在主存，我们称此现象为**页面失效**。把两个或两个以上的虚页想进入主存中的同一页面位置的现象称为**页面争用或实页冲突**。

第4章 存储体系

当所要访问的虚页不在主存，而所要进入主存中的页面又未被占用时，两者不会同时发生

若所要进入主存中的页面已经被占用，则两者会同时发生。对全相联映象，只有主存全被占满时才会同时发生

4-2 对于一个由两个存储器**M1**和**M2**构成的存储系统，设**M1**的命中率为**h**，两个存储器的存储容量分别为**s1**和**s2**，访问速度分别为**t1**和**t2**，每千字节的价格分别为**c1**和**c2**？

(1) 在什么条件下，整个存储系统的每千字节平均价格会接近于**c2**？

(2) 写出这个存储系统的等效访问时间**ta**的表达式。

(3) 假设存储系统的访问效率**e=t1/ta**，两个存储器的速度比**r=t2/t1**。试以速度比**r**和命中率**h**来表示访问效率**e**。

(4) 如果 $r=100$ ，为了使访问效率 $e>0.95$ ，要求命中率 h 是多少？

(5) 对于(4)所要求的命中率实际上很难达到。假设实际的命中率只能达到 0.96 。现采用一种缓冲技术来解决这个问题。当访问 $M1$ 不命中时，把包括被访问数据在内的一个数据块都从 $M2$ 取到 $M1$ 中，并假设被取到 $M1$ 中的每个数据平均可以被重复访问 5 次。请设计缓冲深度（即每次从 $M2$ 取到 $M1$ 中的数据块的大小）。

4-2 解:

(1)

$$C = \frac{C_1 \cdot S_1 + C_2 \cdot S_2}{S_1 + S_2}$$

$S_2 \gg S_1$ 时, $C \approx C_2$

(2)

$$t_a = h \times t_1 + (1-h) \times t_2$$

第4章

(3)

$$e = \frac{t1}{ta} = \frac{t1}{ht1 + (1-h)t2} = \frac{1}{h + (1-h)\frac{t2}{t1}} = \frac{1}{h + (1-h)r}$$

(4)

r=100时,

$$e = \frac{1}{100 - 99h}$$

$$0.95 = 1/(100 - 99h)$$

$$h = 0.99947$$

(5)

$$H' = 1 - \frac{1-H}{n} = \frac{H+n-1}{n}$$

$$0.99947 = 1 - \frac{1-0.96}{n}$$

$$n=75.47$$

$$5m=n$$

$$m=15.1$$

4-3 在页式虚拟存储器中，一个程序由 **P1~P5** 共**5**个页面组成。在程序执行过程中依次访问到的页面如下：

P2,P3,P2,P1,P5,P2,P4,P5,P3,P2,P5,P2

假设系统分配给这个程序的主存有**3**个页面，分别采用**FIFO**、**LRU**和**OPT**三种页面替换算法对这**3**页主存进行调度。

(1) 画出主存页面调入、替换和命中的情况表。

(2) 统计三种页面替换算法的页命中率。

第4章

4-3 解:

(1) FIFO

1 2 3 4 5 6 7 8 9 10 11 12
2 3 2 1 5 2 4 5 3 2 5 2

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

中 * * * 中 * 中 * *

第4章

(2) LRU

1 2 3 4 5 6 7 8 9 10 11 12

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

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

中 * 中 * 中 * * 中 中

(2) LRU堆栈法

1 2 3 4 5 6 7 8 9 10 11 12

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

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

中 * 中 * 中 * * 中 中

第4章

(3) OPT

1 2 3 4 5 6 7 8 9 10 11 12
2 3 2 1 5 2 4 5 3 2 5 2

2	2	2	2	2	2*	4*	4*	4*	2	2	2
	3	3	3	3*	3	3	3	3	3	3	3
			1*	5	5	5	5	5	5	5	5

中 * 中 * 中 中 * 中 中

第4章

(2)

FIFO算法页 $H=3/12=25\%$

LRU算法页 $H=5/12=41.7\%$

OPT算法页 $H=6/12=50\%$

4-4 假设在一个采用组相联映象方式的**Cache**中，主存由**B0~B7**共**8**块组成，**Cache**有两组，每组两块，每块的大小为**16**个字节，采用**LRU**块替换算法。在一个程序执行过程中依次访问这个**Cache**的块地址流如下：

B6,B2,B4,B1,B4,B6,B3,B0,B4,B5,B7,B3

(1) 写出主存地址的格式，并标出各字段的长度。

(2) 写出**Cache**地址的格式，并标出各字段的长度。

(3) 画出主存与**Cache**之间各个块的映像对应关系。

(4) 采用**LRU**替换算法，计算**Cache**的块命中率。

(5) 如果改为全相联映像方式，再做(4)，可以得出什么结论？

(6) 如果在程序执行过程中，每从主存装入一块到**Cache**，则平均要对这个块访问**16**次。请计算在这种情况下**Cache**命中率。

第4章

4-4 解:

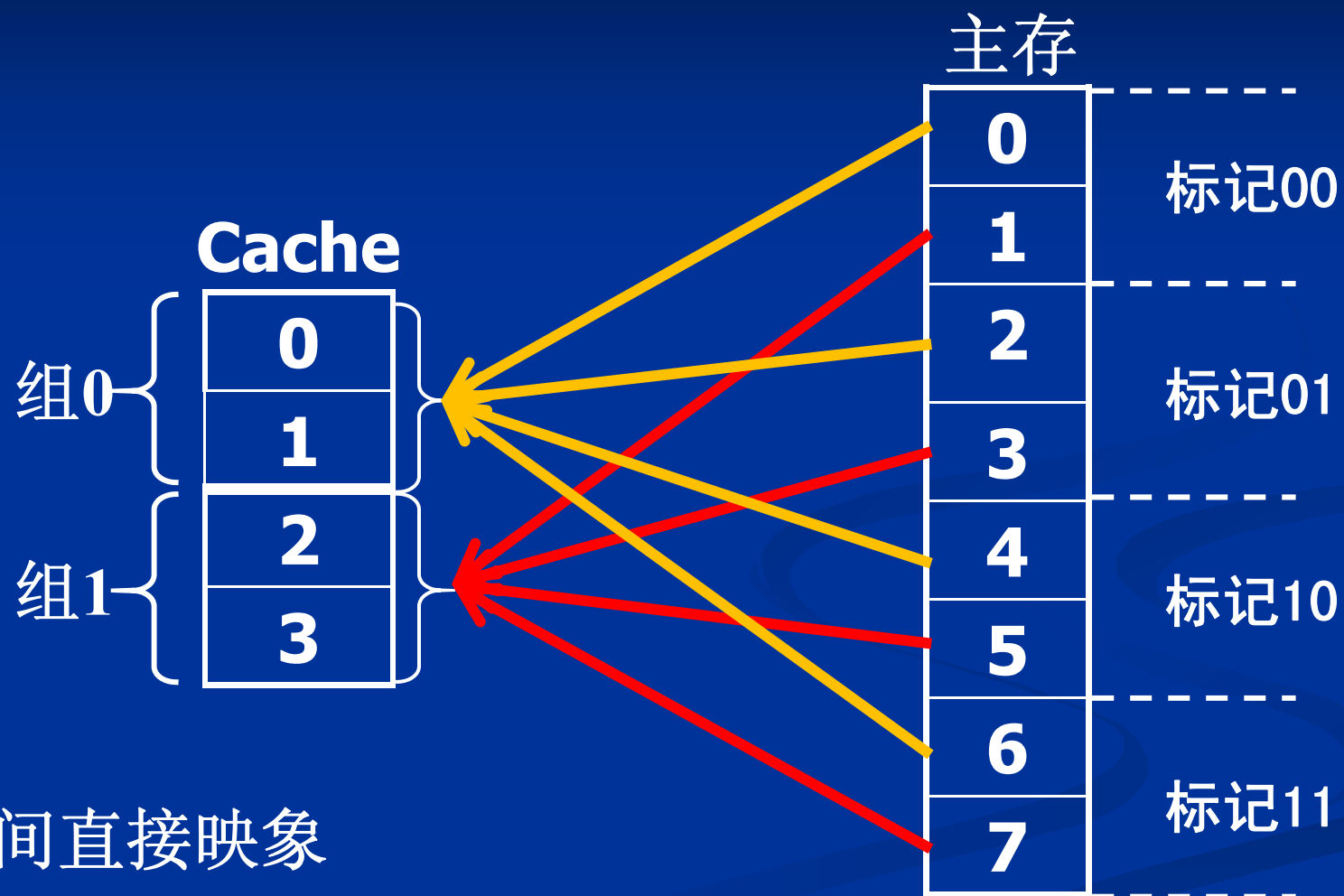
(1) 主存地址格式

标记	组号	块内地址
2	1	4

(2) Cache地址格式

组号	组内块号	块内地址
1	1	4

(3)



组间直接映象

组内全相联映象

(4) LRU利用堆栈法

1 2 3 4 5 6 7 8 9 10 11 12

6 2 4 1 4 6 3 0 4 5 7 3

6	2	4	4	4	6	6	0	4	4	4	4
	6	2	2	2	4	4	6	0	0	0	0
			1	1	1	3	3	3	5	7	3
						1	1	1	3	5	7

* 中 *

* * * * *

块命中率 $H=1/12=8.3\%$

(5) LRU

1 2 3 4 5 6 7 8 9 10 11 12

6 2 4 1 4 6 3 0 4 5 7 3

6	6	6	6	6	6	6	6	6	5	5	5
	2	2	2	2	2	3	3	3	3	7	7
		4	4	4	4	4	4	4	4	4	4
			1	1	1	1	0	0	0	0	3

中 中 * * 中 * * *

块命中率 $H=3/12=25\%$

(5)LRU利用堆栈法

1 2 3 4 5 6 7 8 9 10 11 12

6 2 4 1 4 6 3 0 4 5 7 3

6	2	4	1	4	6	3	0	4	5	7	3
	6	2	4	1	4	6	3	0	4	5	7
		6	2	2	1	4	6	3	0	4	5
			6	6	2	1	4	6	3	0	4

中 中 * * 中 * * *

块命中率 $H=3/12=25\%$

(6)

$$H' = 1 - \frac{1-H}{n} = \frac{H+n-1}{n}$$

$$H' = 1 - (1 - 0.083) / (16 * 16) = 0.996$$

$$H' = 1 - (1 - 0.25) / (16 * 16) = 0.997$$

第4章 存储体系

4-5

- 页式虚拟存储器采用全相联地址映射与变换
- 会发生页面失效的虚页号有：2，3，5，7

第4章 存储体系

虚地址	虚页号	装入位	页内地址	实地址
0	0	1	0	3072
3728	3	0	失效	无
1023	0	1	1023	4095
1024	1	1	0	1024
2055	2	0	失效	无
7800	7	0	失效	无
4096	4	1	0	2048
6800	6	1	656	656

第4章 存储体系

■ 4-6

页地址流		4	5	3	2	5	1	3	2	2	5	1	3
堆 栈 内 容	S(1)	4	5	3	2	5	1	3	2	2	5	1	3
	S(2)		4	5	3	2	5	1	3	3	2	5	1
	S(3)			4	5	3	2	5	1	1	3	2	5
	S(4)				4	4	3	2	5	5	1	3	2
	S(5)						4	4	4	4	4	4	4
	S(6)												
实 页 数	$n=1$									H			
	$n=2$									H			
	$n=3$					H				H			
	$n=4$					H		H	H	H	H	H	H
	$n \geq 5$					H		H	H	H	H	H	H

模拟结果表明,使用 LRU 替换算法替换,对该程序至少应分配 4 个实页。如果只分配 3 个实页,其页命中率只有 $2/12$,太低;而分配实页数多于 4 页后,其页命中率不会再有提高。所以,分配给该程序 4 个实页即可,其可能的最高命中率为 $H=7/12$ 。

第4章 存储体系

4-7

■ 地址字段及其对应关系

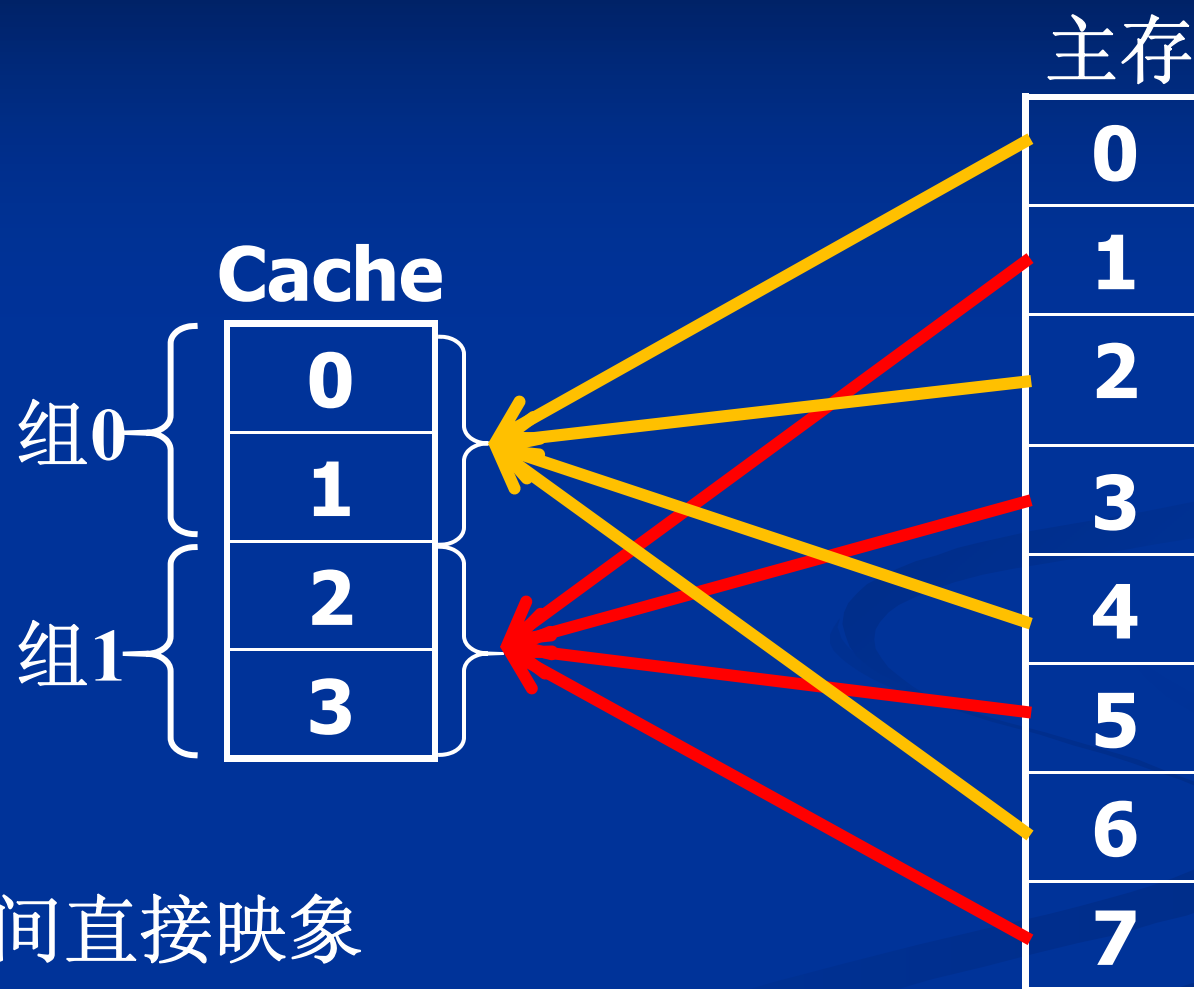
Cache地址



主存地址



第4章 存储体系



组间直接映象

组内全相联映象

第4章 存储体系

时刻	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
地址流	1	2	4	1	3	7	0	1	2	5	4	6	4	7	2
0		2	4	4	4	4	0	0	2	2	4	6	4	4	2
1			2	2	2	2	4	4	0	0	2	4	6	6	4
2	1	1	1	1	3	7	7	1	1	5	5	5	5	7	7
3					1	3	3	7	7	1	1	1	1	5	5
				命中		替	替	替	替	替	替	替	命中	替	替
						换	换	换	换	换	换	换		换	换

第4章 存储体系

块失效同时又发生块争用的时刻是：

**6, 7, 8, 9, 10, 11, 12, 14,
15**

Cache的命中率为

$2/15=13.3\%$

第5章 重叠、流水、向量处理机

■ 流水线吞吐率

- 指流水线单位时间内能处理的指令条数或能输出的结果数

■ 流水线效率

- 指流水线中的设备实际使用时间占整个运行时间之比，也称为流水线的设备时间利用率

第5章 重叠、流水、向量处理机

■ 全局性相关

- 由转移指令引起的相关，会影响后续指令和指缓中的指令。它对流水线的影响是全局的，指缓中的指令可能要全部作废，流水线会断流，使吞吐率和效率下降等。我们把由转移造成的指令之间关联称为全局性相关

第5章 重叠、流水、向量处理机

■ 局部性相关

- 指令相关、主存操作数相关、通用寄存器组操作数相关和通用寄存器组基址或变址值相关等只影响相关的两条或几条指令，并不会改变已经预取到指令缓冲器中指令，其影响是局部的，因此称之为局部性相关

第5章 重叠、流水、向量处理机

■ 先写后读相关

- 机器同时解释执行多条指令时，这些指令对同一存储单元要求先写入而后读出的关联

第5章 重叠、流水、向量处理机

■ 向量处理机

- 向量适合于流水处理。一般将向量数据表示与流水处理方式结合在一起，构成向量流水处理机，也称其为向量处理机，以提高主要面向向量数组计算类应用的计算机的速度性能

5-1 假设一条指令的执行过程分为“取指令”、“分析”和“执行”三段，每一段的时间分别为 Δt 、 $2\Delta t$ 和 $3\Delta t$ 。在以下各种情况下，分别写出连续执行 n 条指令所需要的时间表达式。

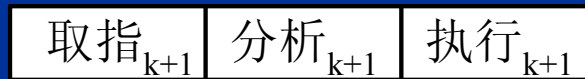
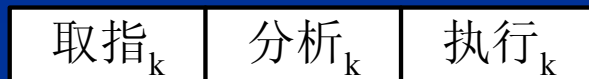
- (1) 顺序执行方式。
- (2) 仅“取指令”和“执行”重叠。
- (3) “取指令”、“分析”和“执行”重叠。

5-1 解:

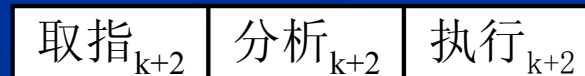
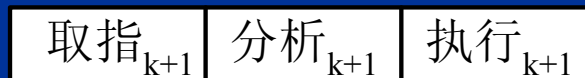
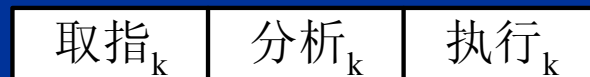
第5章



(a)



(b)



(c)

第5章

$$(1) (t_{\text{取指}} + t_{\text{分析}} + t_{\text{执行}}) \times n$$

$$T = (\Delta t + 2 \Delta t + 3 \Delta t) \times n = 6n\Delta t$$

$$(2) t_{\text{取指}} + n \times t_{\text{分析}} + (n-1) \times \max\{t_{\text{取指}}, t_{\text{执行}}\} + t_{\text{执行}}$$

$$T = \Delta t + (2 \Delta t + 3 \Delta t) \times n = (1 + 5n)\Delta t$$

$$(3) t_{\text{取指}} + \max\{t_{\text{取指}}, t_{\text{分析}}\} + (n-2) \times \max\{t_{\text{取指}}, t_{\text{分析}}, t_{\text{执行}}\} + \max\{t_{\text{分析}}, t_{\text{执行}}\} + t_{\text{执行}}$$

$$T = \Delta t + 2 \Delta t + 3 \Delta t \times n = (3 + 3n)\Delta t$$

第5章 重叠、流水、向量处理机

5-2

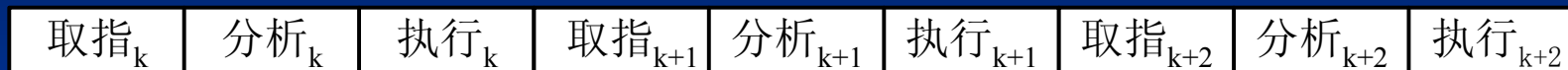
假设指令的解释分取指、分析和执行3步，每步的时间相应为 $t_{\text{取指}}$ 、 $t_{\text{分析}}$ 、 $t_{\text{执行}}$ ，

(1) 分别计算下列几种情况下，执行完100条指令所需时间的一般关系式：

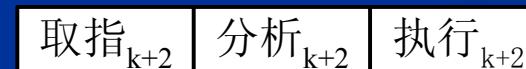
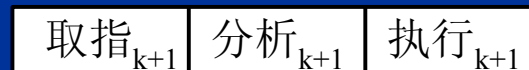
- ① 顺序方式；
- ② 仅“执行 k ”与“取指 $k+1$ ”重叠；
- ③ 仅“执行 k ”、“分析 $k+1$ ”、“取指 $k+2$ ”重叠。

(2) 分别在 $t_{\text{取指}}=t_{\text{分析}}=2$ 、 $t_{\text{执行}}=1$ 及 $t_{\text{取指}}=t_{\text{执行}}=5$ 、 $t_{\text{分析}}=2$ 两种情况下，计算出上述各结果。

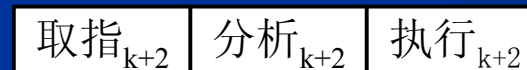
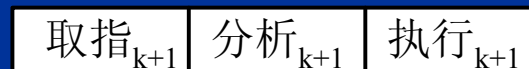
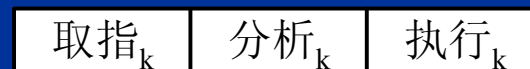
第5章 重叠、流水、向量处理机



(a)



(b)



(c)

第5章 重叠、流水、向量处理机

(1) 执行100条指令所需的时间:

- ① $100 \times (t_{\text{取指}} + t_{\text{分析}} + t_{\text{执行}})$
- ② $t_{\text{取指}} + 100 \times t_{\text{分析}} + 99 \times \max\{t_{\text{取指}}, t_{\text{执行}}\} + t_{\text{执行}}$
- ③ $t_{\text{取指}} + \max\{t_{\text{取指}}, t_{\text{分析}}\} + 98 \times \max\{t_{\text{取指}}, t_{\text{分析}}, t_{\text{执行}}\} + \max\{t_{\text{分析}}, t_{\text{执行}}\} + t_{\text{执行}}$

(2) 当 $t_{\text{取指}} = t_{\text{分析}} = 2$ 、 $t_{\text{执行}} = 1$ 时, 可求得执行100条指令所需时间:

- ① 500
- ② 401
- ③ 203

$t_{\text{取指}} = t_{\text{执行}} = 5$ 、 $t_{\text{分析}} = 2$ 时, 可求得执行100条指令所需时间:

- ① 1200
- ② 705
- ③ 510

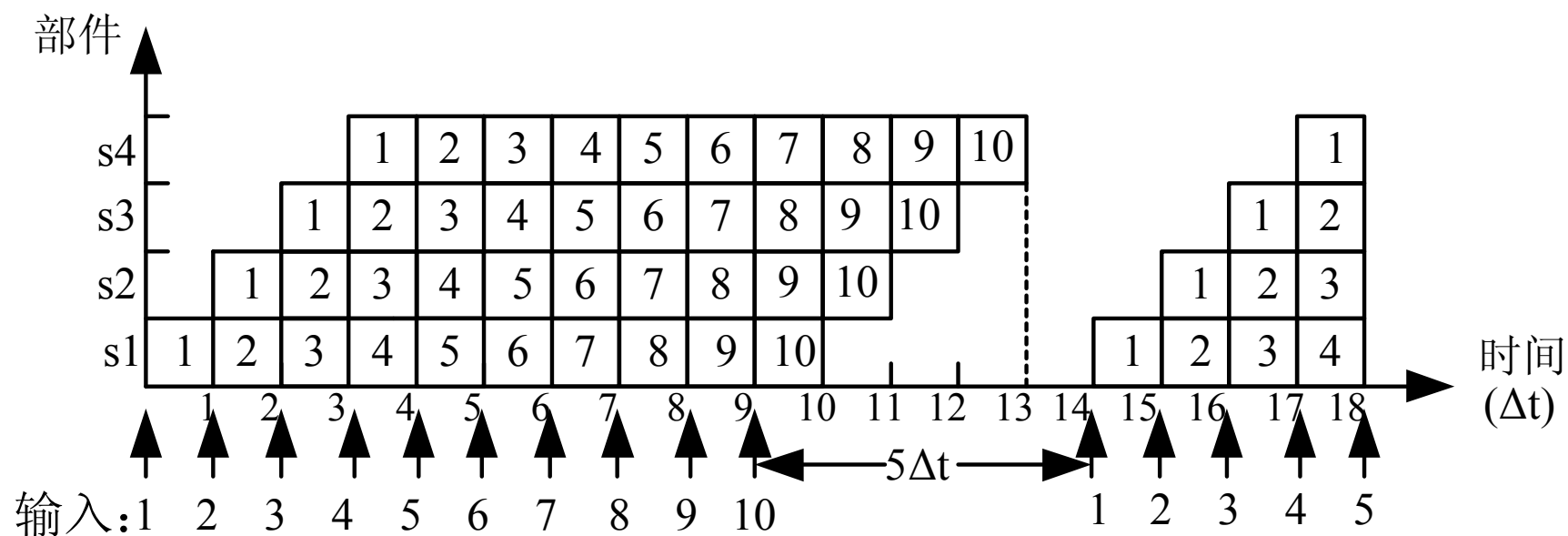
第5章 重叠、流水、向量处理机

5-3

流水线由4个功能部件组成，每个功能部件的延迟时间为 Δt 。当输入10个数据后，间歇 $5\Delta t$ ，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，并画出时空图。

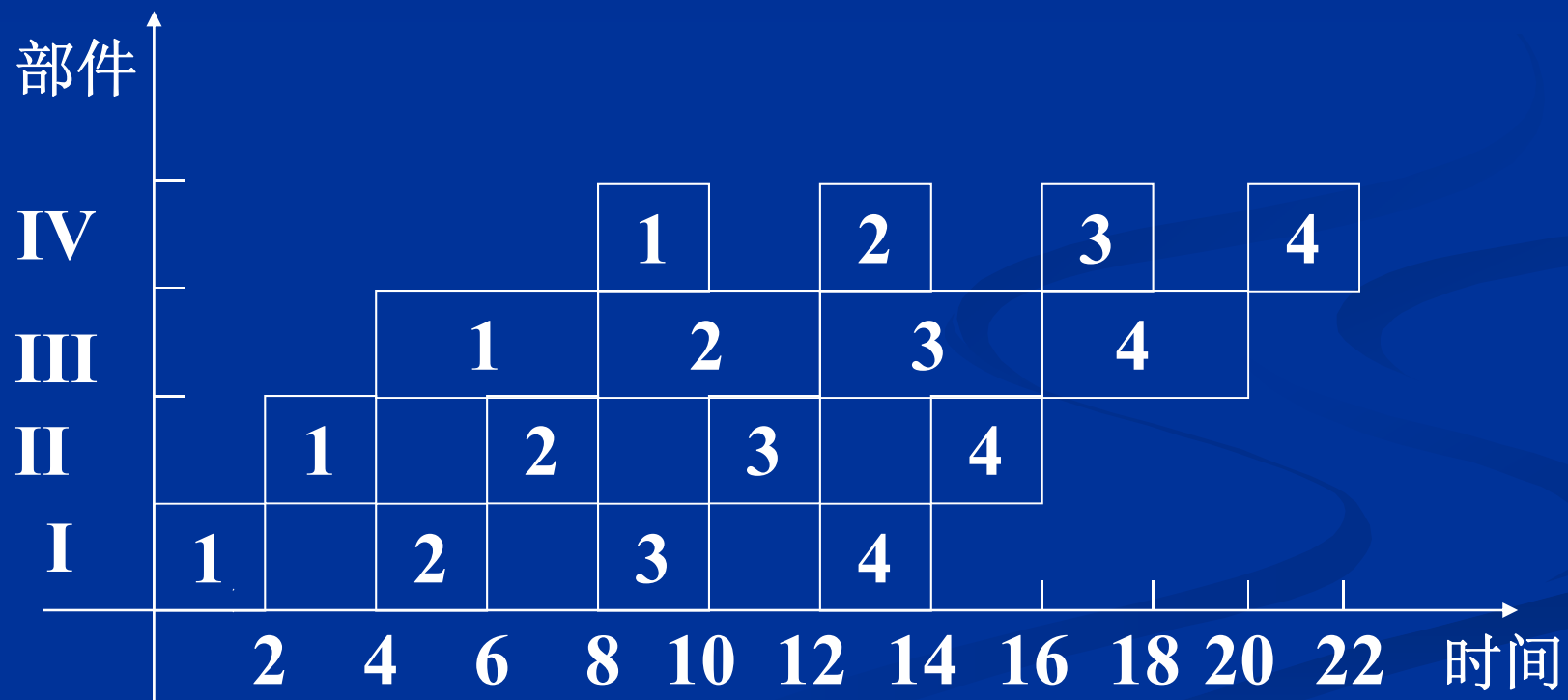
第5章 重叠、流水、向量处理机

$$TP = 10/14 \Delta t = 5/7 \Delta t$$



第5章 重叠、流水、向量处理机

■ 5-4



第5章 重叠、流水、向量处理机

■ 实际吞吐率

$$TP = 4 / 22\Delta t = 2 / 11\Delta t$$

■ 效率

$$\begin{aligned}\eta &= (3 \times 4 \times 2\Delta t + 4 \times 4\Delta t) / 4 \times 22\Delta t \\ &= 40\Delta t / 88\Delta t = 5/11 = 45.5\%\end{aligned}$$

第5章 重叠、流水、向量处理机

- 为加快流水，使流水线每隔 $2\Delta t$ 流出一个结果，应减少 III 段的经过时间至 Δt 。此时，流水线的实际吞吐率和效率为：

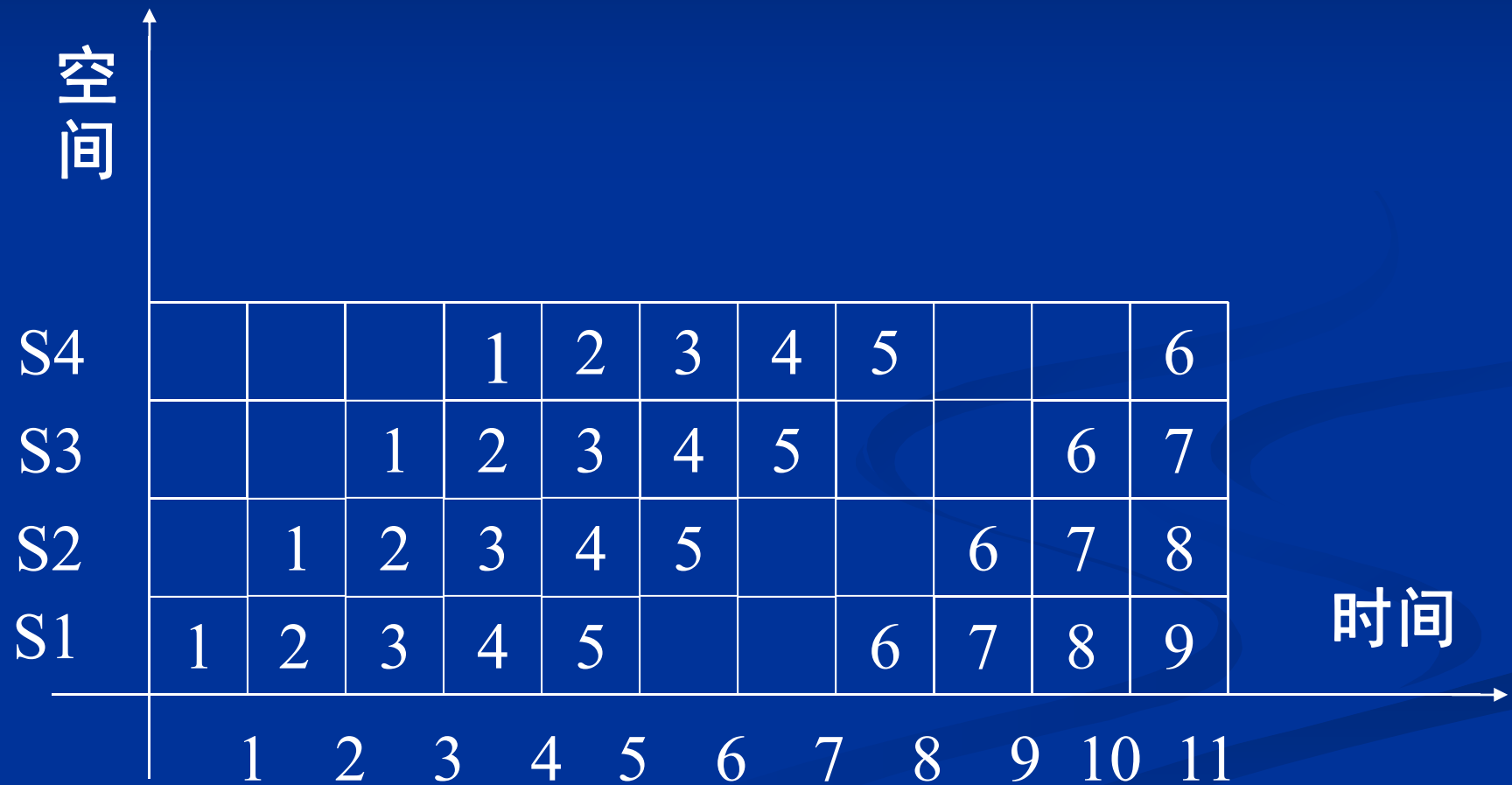
$$TP = 4 / 14\Delta t = 2 / 7\Delta t$$

$$\begin{aligned}\eta &= (3 \times 4 \times 2\Delta t + 8 \times 1\Delta t) / 4 \times 14\Delta t \\ &= 32\Delta t / 56\Delta t = 4 / 7 = 57\%\end{aligned}$$

5-5一条线性流水线由**4**个流水段组成，每个流水段的延迟时间都相等，都为 Δt 。开始5个 Δt ，每隔一个 Δt 向流水线输入一个任务，然后停顿2个 Δt ，如此重复。求流水线的实际吞吐率、加速比和效率。

第5章

5-5 解:



第5章

$$TP = 5 / (7\Delta t) = 0.714 / \Delta t$$

$$S = (5 \times 4\Delta t + 2\Delta t) / 7\Delta t = 22\Delta t / (7\Delta t) = 3.14$$

$$E = 4 \times 5\Delta t / (4 \times 7\Delta t) = 0.714 = 71.4\%$$

第5章 重叠、流水、向量处理机

5-6

- 为提高流水线效率可采用哪两种主要途径来克服速度瓶颈？现有3段流水线，各段经过时间依次为 Δt 、 $3\Delta t$ 、 Δt 。
- (1) 分别计算在连续输入3条指令时和30条指令时的吞吐率和效率。
- (2) 按两种途径之一改进，画出流水线结构示意图，同时计算连续输入3条指令和30条指令时的吞吐率和效率。

(1) 连续输入3条指令时的吞吐率

$$\begin{aligned} TP_3 &= 3 / (5\Delta t + (3-1) * 3 \Delta t) \\ &= 3 / 11 \Delta t \end{aligned}$$

$$\begin{aligned} \text{效率}\eta_3 &= (3 * 5\Delta t) / (3 * 11\Delta t) = 5/11 \\ &= 45.5\% \end{aligned}$$

连续输入30条指令时的吞吐

$$\begin{aligned} TP_{30} &= 30 / (5\Delta t + (30-1) * 3\Delta t) \\ &= 15/46 \Delta t \end{aligned}$$

$$\begin{aligned} \text{效率}\eta_{30} &= (30 * 5\Delta t) / (3 * 92\Delta t) = 25/46 \\ &= 54.3\% \end{aligned}$$

为提高流水线效率，可采取瓶颈段再细分和瓶颈段并联两种途径。连续输入3条指令和30条指令时的吞吐率和效率分别为

$$TP_3 = 3 / (5\Delta t + (3-1)\Delta t) = 3/7 \Delta t$$

$$\eta_3 = (3 * 5\Delta t) / (5 * 7\Delta t) = 3/7 \\ = 42.9\%$$

$$TP_{30} = 30 / (5\Delta t + (30-1) * \Delta t) \\ = 15/17 \Delta t$$

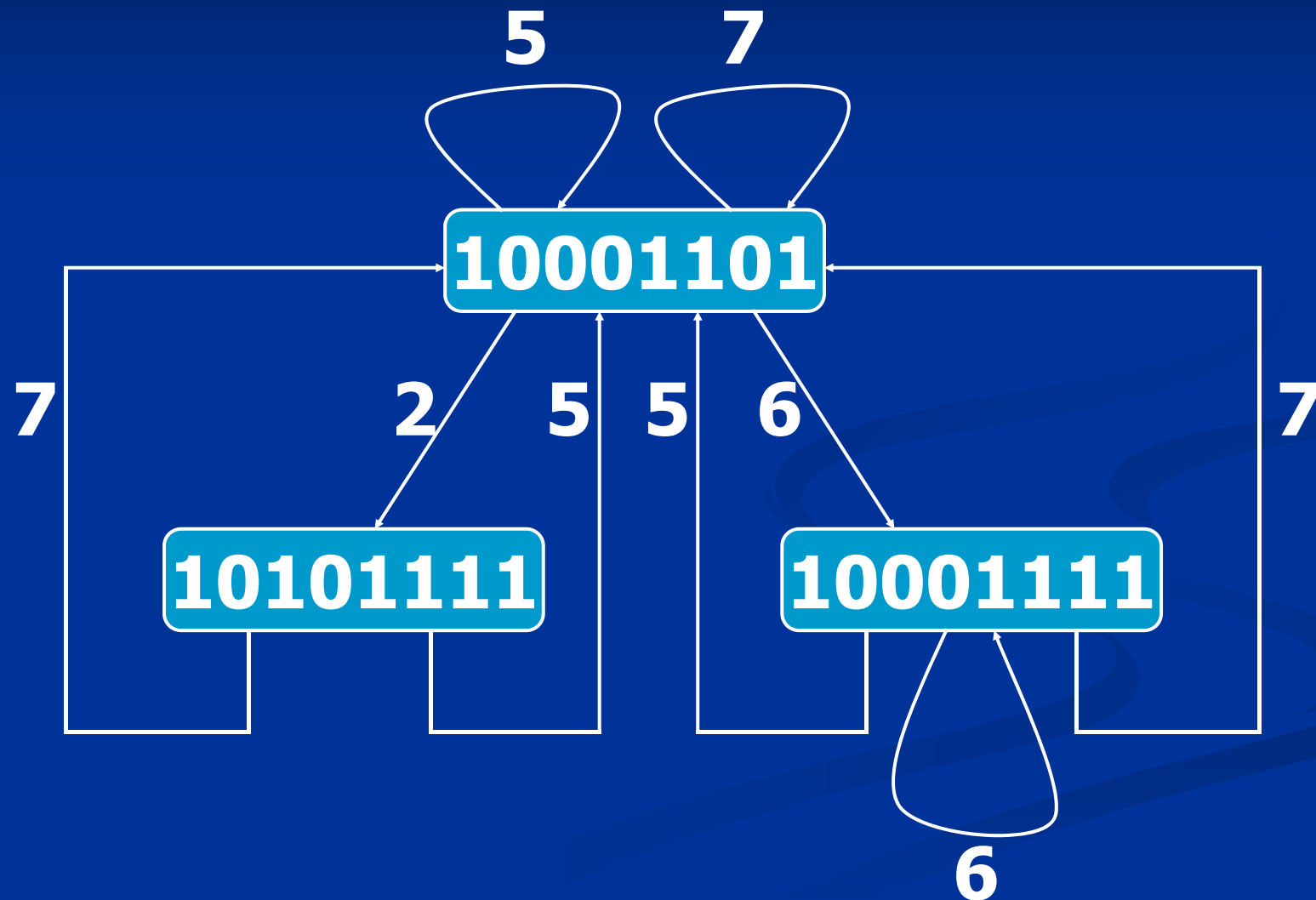
$$\eta_{30} = (30 * 5\Delta t) / (5 * 34\Delta t) = 15/17 \\ = 88.2\%$$

第5章 重叠、流水、向量处理机

5-12

- 禁止表 $F = \{1, 3, 4, 8\}$
- 冲突向量 $C = (10001101)$
- 流水线状态图如下:

第5章 重叠、流水、向量处理机



第5章 重叠、流水、向量处理机

调度方案	平均间隔拍数	调度方案	平均间隔拍数
(5)	5	(6, 5)	5.50
(7)	7	(6, 7)	6.50
(2, 5)	3.50	(6)	6
(2, 7)	4.50		

第5章 重叠、流水、向量处理机

- 最小平均延迟=3.5 拍
- 最大吞吐率 $TP_{MAX}=1/3.5\Delta t$
- 调度方案为 (2, 5)

第5章 重叠、流水、向量处理机

- 按（2，5）调度方案，当输入6个任务时，任务调度及完成情况如下：

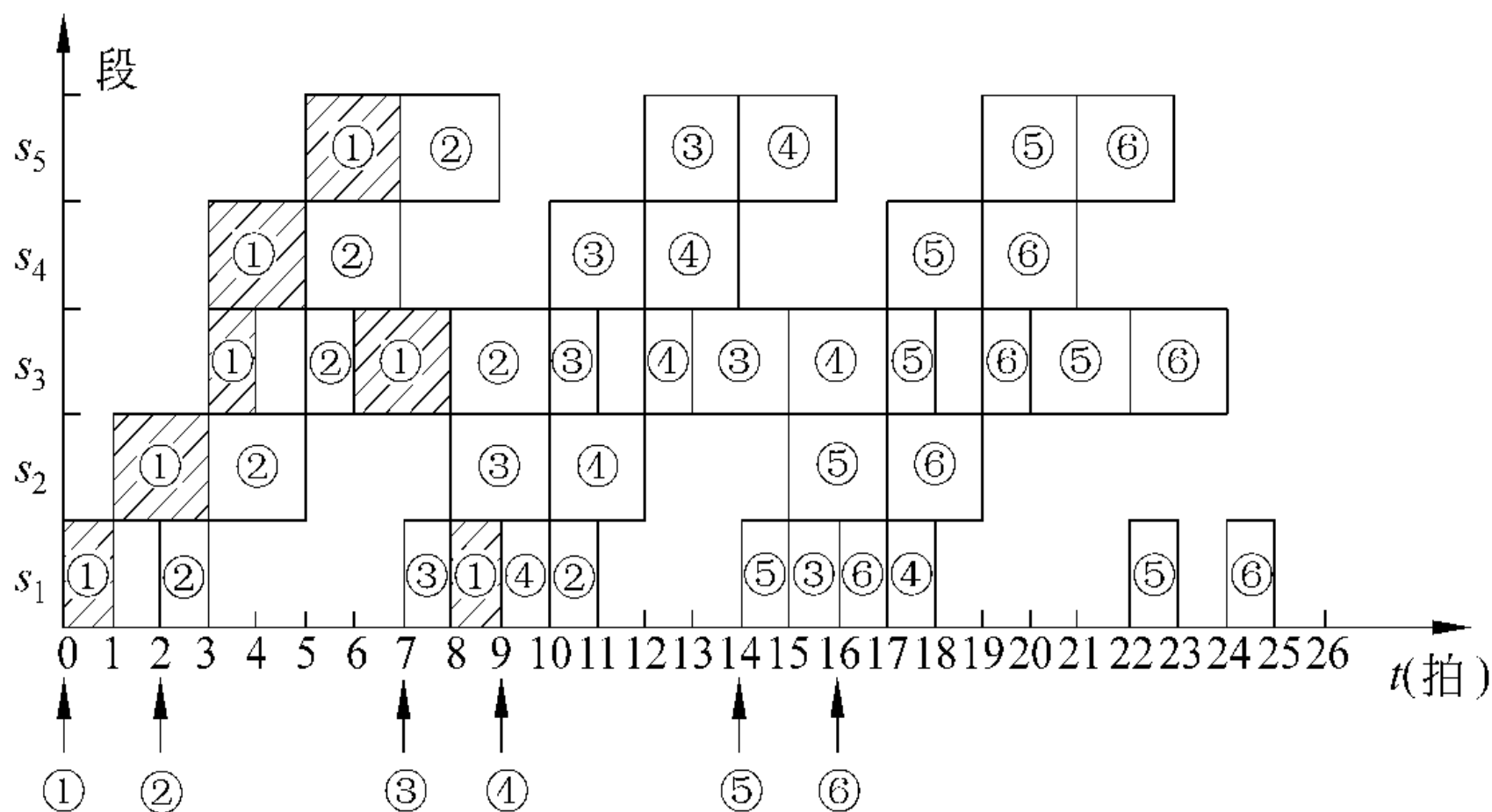
任务： 1 2 3 4 5 6

所需拍数： 9 2 5 2 5 2

- 实际吞吐率

$$TP = 6 / (9 + 2 + 5 + 2 + 5 + 2) \Delta t = 6 / 25 \Delta t$$

第5章 重叠、流水、向量处理机

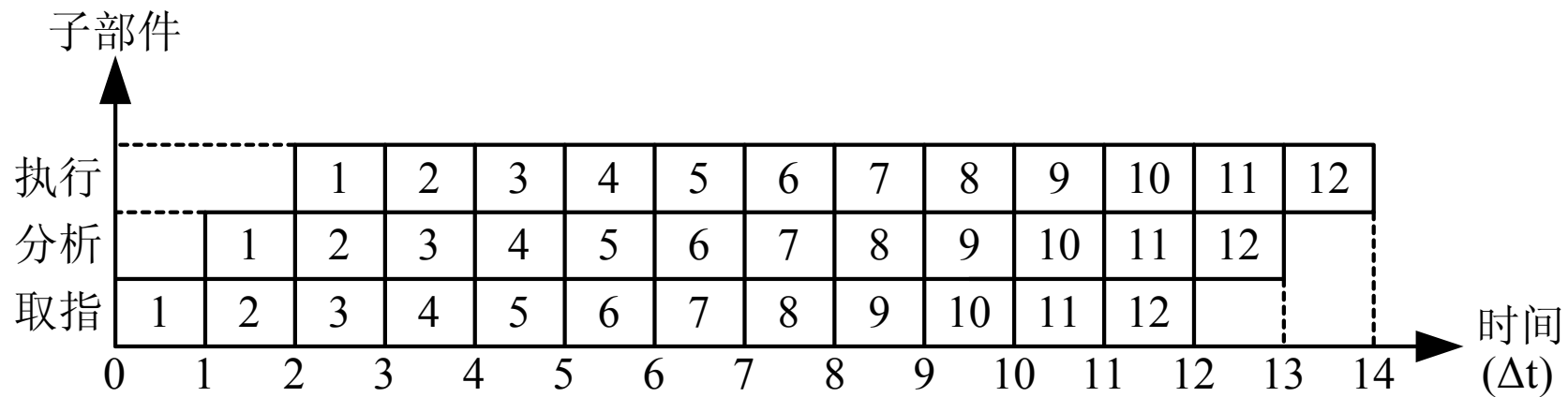


第5章

5-17 设指令由取指、分析、执行三个子部件组成。每个子部件经过时间为 Δt ，连续执行**12**条指令。请分别画出在常规标量流水处理机及度 m 均为**4**的超标量处理机、超长指令字处理机、超流水线处理机上工作的时空图，分别计算它们相对常规标量处理机的加速比**SP**。

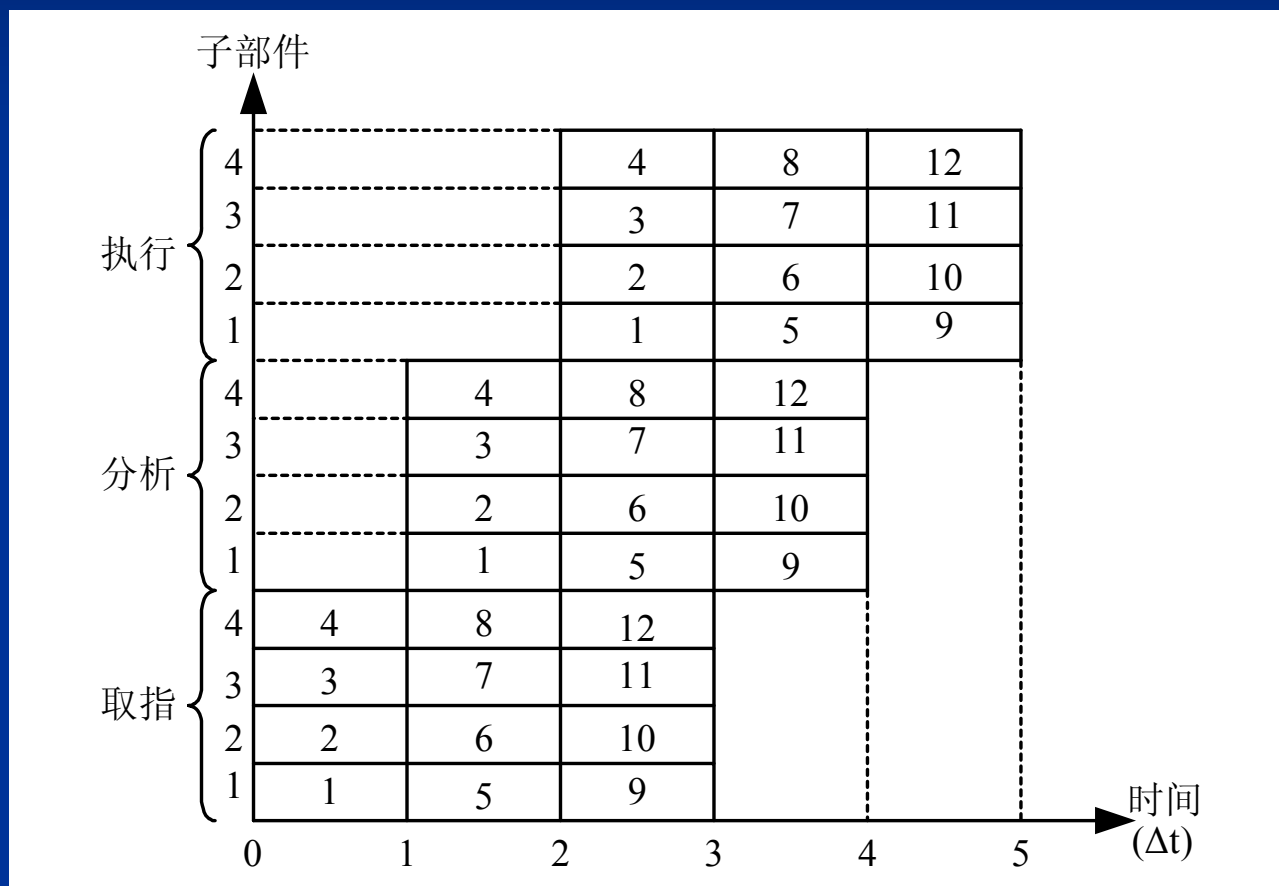
第5章

常规标量流水处理机



第5章

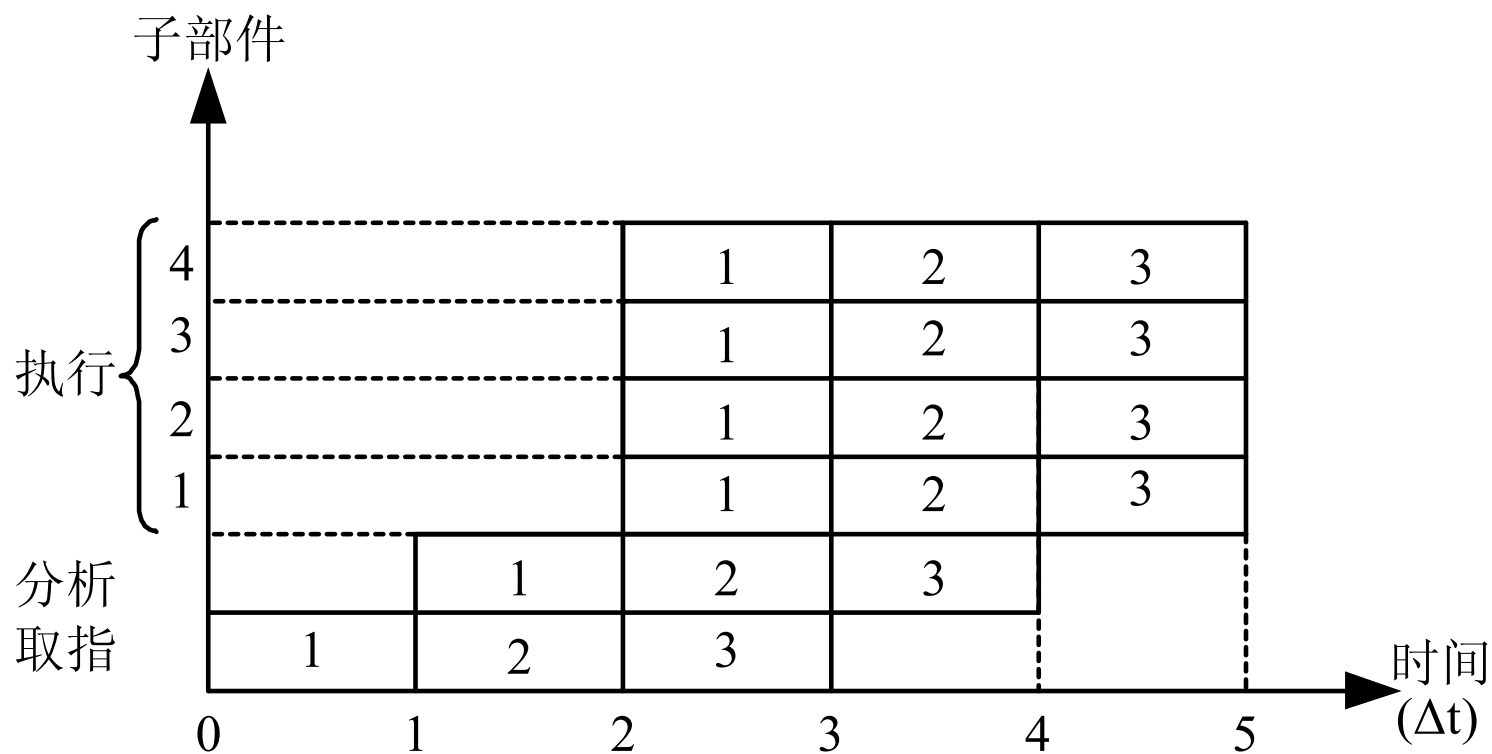
度 $m=4$ 的超标量处理机



加速比 $SP = 14 \Delta t / 5 \Delta t = 2.8$

第5章

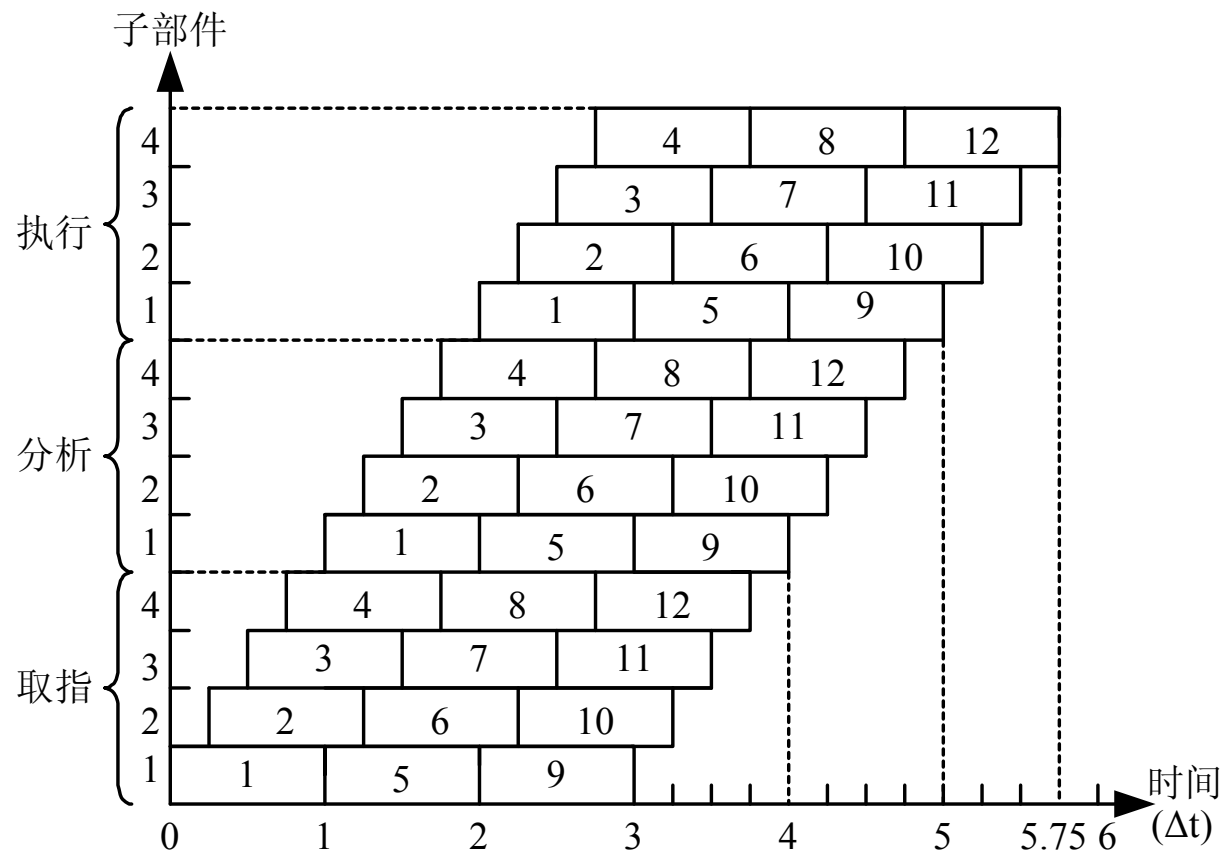
度 $m=4$ 的超长指令字处理机



加速比 $SP = 14 \Delta t / 5 \Delta t = 2.8$

第5章

度 $m=4$ 的超流水线处理器

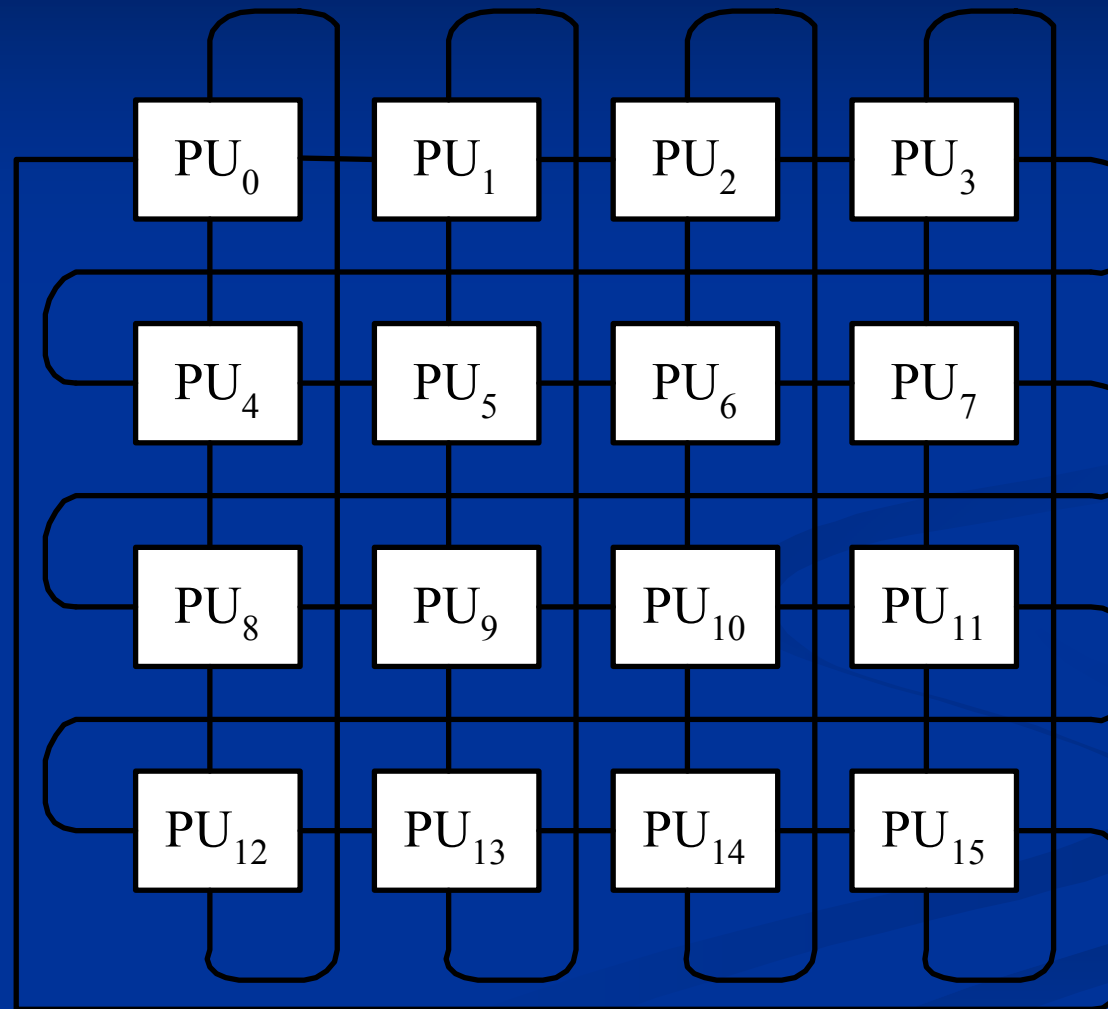


$$\text{加速比SP} = 14 \Delta t / 5.75 \Delta t = 2.43$$

第6章

6-1 画出16台处理器仿ILLIAC IV的连接模式进行互连的互连结构图，列出PE0分别经一步、二步和三步传送，能将信息传送到各处理器号。

第6章



第6章

PE0 (PU0) 经一步可将信息传送至PU1、PU4、PU12、PU15。

PE0 (PU0) 至少需经二步才能将信息传送至PU2、PU3、PU5、PU8、PU11、PU13、PU14。

PE0 (PU0) 至少需经三步才能将信息传送至PU6、PU7、PU9、PU10。

第6章

6-2 编号为**0**、**1**、...、**15**的**16**个处理器用单级互连网络互连。当互连函数分别为

(1) Cube3

(2) PM2+3

(3) PM2-0

(4) Shuffle

(5) Shuffle(Shuffle)

第6章

16个处理器的编号可用4位二进制P3P2P1P0表示，其中第13号处理器的二进制编号为1101。

(1) $\text{Cube3} = P3'P2P1P0 = 0101_2 = 5$ 。

(2) $PM2+3 = j+2^3(\text{mod } 16) = 13+8(\text{mod } 16) = 5$

(3) $PM2-0 = j-2^0(\text{mod } 16) = 13-1(\text{mod } 16) = 12$

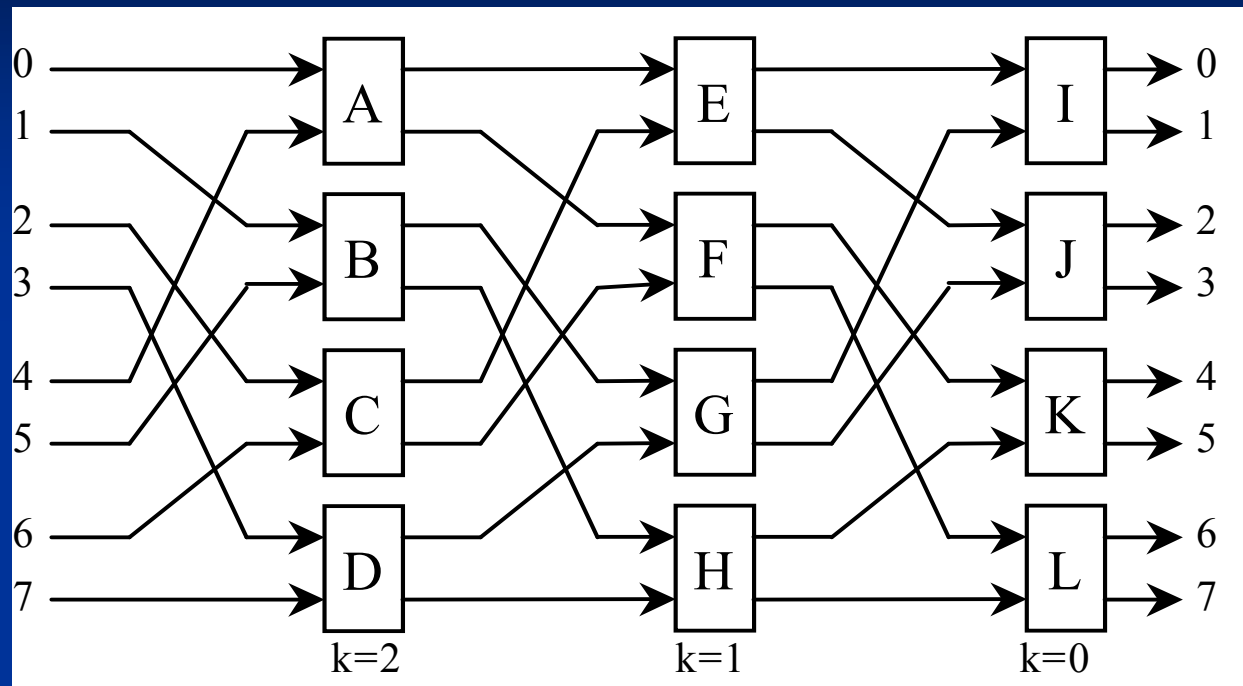
(4) $\text{Shuffle} = P2P1P0P3 = 1011_2 = 11$

(5) $\text{Shuffle}(\text{Shuffle}) = P1P0P3P2 = 0111_2 = 7$

第6章

6-6 画出**0~7**号共**8**个处理器的**3**级混洗交换网络，在图中标出实现将**6**号处理器数据播送到**0~4**号，同时将**3**号处理器数据播送到其余**3**个处理器时 各有关交换开关的状态。

第6章



每级4个开关。

C、E、H、L:

I、J:

D、F:

K:

下播
上播
交换
直通

第7章

练习：7-2, 7-6, 7-7