

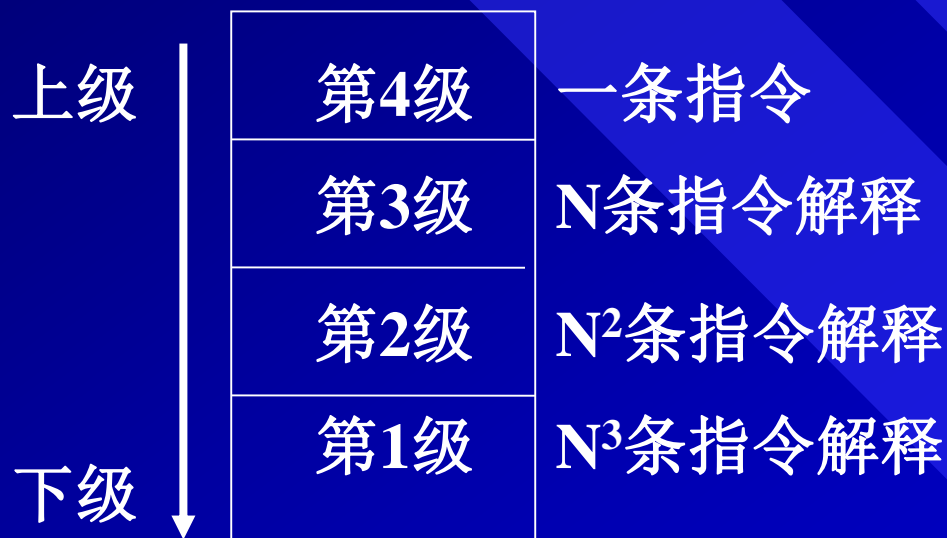
计算机系统结构作业题解

第1章

作1.2 如有一个经解释实现的计算机，可以按功能划分成4级。每一级为了执行一条指令需要下一级的 N 条指令解释。若执行第一级的一条指令需 $K(ns)$ 时间，那么执行第2、3、4级的一条指令各需要用多少时间(ns)？

解： ∵ 第二级的一条指令需第1级的 N 条指令解释
∴ 第二级的一条指令执行时间为 $NKns$ ；
第三级的一条指令执行时间为 N^2Kns ；
第四级的一条指令执行时间为 N^3Kns 。

本题有两个问题应特别注意：第一个问题是“上一级”与“下一级”的关系，即哪是上一级，哪是下一级？在图1.1中第3级是第2级的“上一级”，第1级又是第2级的“下一级”。第二个问题是该计算机是一个“经解释实现的计算机”，上一级的程序在下一级上实现不是经翻译完成，只能是解释。



作1.3 有一个计算机系统可按功能划分成4级，各级的指令都不相同，每一级的指令都比其下一级的指令在效能上强M倍，即第i级的一条指令能完成第i-1级的M条指令的计算量。现若需第i级的N条指令解释第i+1级的一条指令，而有一段第1级的程序需要运行Ks，问在第2、3和4级上的一段等效程序各需要运行多长时间(s)？

解：

第2级上的一段等效程序运行时间为： $\frac{N}{M} \times K(s)$

第3级上的一段等效程序运行时间为： $(\frac{N}{M})^2 \times K(s)$

第4级上的一段等效程序运行时间为： $(\frac{N}{M})^3 \times K(s)$

作1.7 从机器（汇编）语言程序员看，哪些对应用程序员透明？

指令地址寄存器，指令缓冲器，时标发生器，条件码寄存器，乘法器，主存地址寄存器，磁盘外设，先行进位链，移位器，通用寄存器，中断字寄存器。

答：对机器语言程序员透明，指的是这些器件是机器语言程序员不可修改、不可控制。因此指令缓冲器，时标发生器，乘法器，先行进位链，移位器。

作1.6 什么是透明性概念?对计算机系统结构, 下列哪些是透明的?哪些是不透明的?

存贮器的模 m 交叉存取; 浮点数据表示; I / O系统是采用通道方式还是外围处理机方式; 数据总线宽度; 字符行运算指令; 阵列运算部件; 通道是采用结合型的还是独立型的; PDP-11系列中的单总线结构; 访问方式保护; 程序性中断; 串行、重叠还是流水控制方式; 堆栈指令; 存贮器最小编址单位; Cache存贮器。

分析: 有关系统结构属性所包括的内容, 对系统结构都不透明。

- 对于计算机系统结构透明的是：存储器的模 m 交叉存取、数据总线宽度、阵列运算部件、通道是采用结合型还是独立型、**PDP-11**系列的单总线结构、串行、重叠还是流水控制方式、**Cache**存储器。
- 对于计算机系统结构不透明的是：浮点数据表示、**I/O**系统是采用通道方式还是外围处理机方式、字符型运算指令、访问方式保护、程序性中断、堆栈指令、存储器最小编址单位。

例1.1 假设将某系统的某一部件的处理速度加快到**10**倍，但该部件的原处理时间仅为整个运行时间的**40%**，则采用加快措施后能使整个系统的性能提高多少？

解：由题意可知 $f_e=0.4$, $r_e=10$, 根据Amdahl定律

$$S_P = \frac{T_e}{T_o} = \frac{1}{(1 - 0.6) + 0.4 / 10} = \frac{1}{0.64} = 1.56$$

作1.13 假设高速缓存Cache工作速度为主存的5倍，且Cache被访问命中的概率为90%，则采用Cache后，能使整个存储系统获得多高的加速比？

解： $f_e=0.9$, $r_e=5$

$$S_P = \frac{T_e}{T_o} = \frac{1}{(1 - 0.9) + 0.9 / 5} = \frac{1}{0.28} = 3.57$$

作1.11 某工作站采用时钟频率为15MHz、处理速率为10MIPS的处理机来执行一个已知混合程序。假定每次存储器存取为1周期延迟、试问：

(1) 此计算机的有效CPI是多少？

(2) 假定将处理机的时钟提高到30MHz，但存储器子系统速率不变。这样，每次存储器存取需要两个时钟周期。如果30%指令每条只需要一次存储存取，而另外5%每条需要两次存储存取，还假定已知混合程序的指令数不变，并与原工作站兼容，试求改进后的处理机性能。

解 (1)

$$CPI_{old} = \frac{f_{old}}{MIPS \times 10^6} = \frac{15 \times 10^6}{10 \times 10^6} = 1.5$$

(2) 依题意可知：30%的指令需要一次存储存取，则这些指令在处理器提高时钟频率之后需要增加1个时钟周期；另外5%的指令需要增加2个时钟周期。

$$CPI_{new} \rightarrow CPI_{old} \cdot 30\% \cdot 1 + 5\% \cdot 2 \rightarrow 1.9$$

$$MIPS_{new} \rightarrow \frac{f_{new}}{CPI_{new} \cdot 10^6} \rightarrow \frac{30 \cdot 10^6}{1.9 \cdot 10^6} \rightarrow 15.79$$

改进后性能提高情况可用CPU时间之比表示：

$$T_{CPU(old)} / T_{CPU(new)} \rightarrow \frac{CPI_{old} \cdot I_N / f_{old}}{CPI_{new} \cdot I_N / f_{new}} \approx 1.58$$

作1.15 假定利用增加向量模块来提高计算机的运算速度。计算机处理向量的速度比其通常的运算要快20倍，将可用向量处理部分所花费的时间占总时间的百分比称为可向量化百分比。

- (1) 求出加速比 S 和向量化百分比之间的关系式。
- (2) 当要得到加速比为2时的可向量化百分比 F 为多少？
- (3) 为了获得在向量模式所得到的最大加速比的一半，可向量化百分比 F 为多少？

(2) 由 (1) 式有

解 (1) :
$$S \leq \frac{1}{(1 - F) + (F / 20)} \leq \frac{20}{19 + F} \quad (1)$$

由Amdahl定律

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

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

(3) 由题意可知

$$10 = \frac{1}{(1 - F) + (F / 20)}$$

$$F = \frac{18}{19} = 0.95$$

题1.1 某计算机系统同时采用两种措施改进性能，使得两个功能部件的性能分别提高到原来的 r_{e1} 倍和 r_{e2} ，这两个部件在运行时使用的时间比例分别为 f_{e1} 和 f_{e2} 。试分析系统性能提高的总体加速比。

解：

$$S_p = \frac{1}{(1 - f_{e1} - f_{e2}) + f_{e1} / r_{e1} + f_{e2} / r_{e2}}$$

- **例1.2** 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	指令条数	时钟周期数
● 整数运算	45000	1
● 数据传送	32000	2
● 浮点运算	15000	2
● 控制传送	8000	2
● 求有效CPI、MIPS速率和程序的执行时间。		

- 解：依题意可知 $I_N=10^5$ 条， $n=4$

$$CPI = \sum_{i=1}^n (CPI_i \times \frac{I_i}{I_N})$$

$$= \sum_{i=1}^4 (1 \times 0.45 + 2 \times 0.32 + 2 \times 0.15 + 2 \times 0.08)$$

$$= 1.55$$

$$MIPS = \frac{f_C}{CPI \times 10^6} = \frac{40 \times 10^6}{1.55 \times 10^6} \approx 25.8$$

$$T_{CPU} = I_N \times CPI \times T_C$$

$$= 10^5 \times 1.55 \times 1 / 40 \times 10^{-6} = 3.875(ms)$$

题1.2 设有两台机器A和B，对条件转移采用不同的方法。CPU_A采用比较指令和条件转移指令处理方法，若条件转移指令占总执行指令数的20%，比较指令也占20%。CPU_B采用比较和条件转移指令合一的方法，占执行指令数的20%。若规定两台机器执行条件转移指令需2T，其它指令需1T。CPU_B的条件转移指令比CPU_A慢25%，现比较CPU_A合和CPU_B哪个工作速度更快？

解: $CPI_A = 0.2 \times 2 + 0.8 \times 1 = 1.2$

$$CPU_A \text{ 时间} = IC_A \times CPI_A \times T_A = 1.2 T_A \times IC_A$$

IC_A 是 CPU_A 的指令条数, 由于 CPU_B 无比较指令, 因此 $IC_B = 0.8 IC_A$, 若 $IC_A = 100$, 则 $IC_B = 80$, 而 CPU_B 的条件转移指令仍是 20 条, 所以占比例为 $20/80 = 0.25 = 25\%$

$$CPI_B = 0.25 \times 2 + 0.75 \times 1 = 1.25$$

又因为 CPU_B 的 T_B 比 CPU_A 的 T_A 慢 25%, 所以 $T_B = 1.25 T_A$

$$\begin{aligned} CPU_B &= IC_B \times CPI_B \times T_B = 0.8 IC_A \times 1.25 \times 1.25 T_A \\ &= 1.25 T_A \times IC_A \end{aligned}$$

可见, $CPU_A \text{ 时间} < CPU_B \text{ 时间}$, CPU_A 比 CPU_B 工作速度快。

上例中， T_B 只比 T_A 慢10%，则哪个CPU更快些？

解：

$$T_B = 1.1T_A$$

$$\begin{aligned}\text{CPU}_B \text{时间} &= 0.8IC_A \times 1.25 \times 1.1T_A \\ &= 1.1T_A \times IC_A\end{aligned}$$

因此 CPU_B 时间 $<$ CPU_A 时间，则 CPU_B 更快些。

题1.3 某向量计算机系统中，标量指令的平均CPI是1，向量运算指令的平均CPI是64，系统加快向量部件的速度后使向量运算速度提高到原来的2倍，某一测试程序执行时的向量运算指令数量占全部指令数的10%，问计算机系统运行这个测试程序的整体性能比原来提高多少？

解：

$$S_p = \frac{1}{(1-f_e) + (f_e / r_e)} = \frac{1}{0.9 + \frac{0.1}{2 \times 64}} = 1.11$$

作1.12 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	指令混合百分比
算术和逻辑运算	1	60%
Cache命中的加载/存储	2	18%
转移	4	12%
Cache失效时访问主存	8	10%

- (1) 计算在单处理机上用上述跟踪数据运行程序的平均CPI。
- (2) 根据(1)所得CPI，计算相应的MIPS 速率。

- 解：依题意可知 $I_N = 2 \times 10^5$ 条， $n=4$,

$$CPI = \sum_{i=1}^n (CPI_i \times \frac{I_i}{I_N})$$

$$= \sum_{i=1}^4 (1 \times 0.6 + 2 \times 0.18 + 4 \times 0.12 + 8 \times 0.1)$$

$$= 2.24$$

$$MIPS = \frac{f_c}{CPI \times 10^6} = \frac{40 \times 10^6}{2.2 \times 10^6} \approx 17.86$$

$$T_{CPU} = I_N \times CPI \times T_C$$

$$= 2 \times 10^5 \times 2.24 \times \frac{1}{40} \times 10^{-6} = 11.2(ms)$$

第2章

题2.1 一种浮点数有1位符号位，阶码为7位移码，尾数8位与符号位一起采用原码的规格化表示，基数为2，该浮点数可表示的最大数为____，最大数与最接近它的数据（次最大数）的差值为____，可表示的最小数为____，最小数与最接近它的正数（次最小数）的差值为_____。

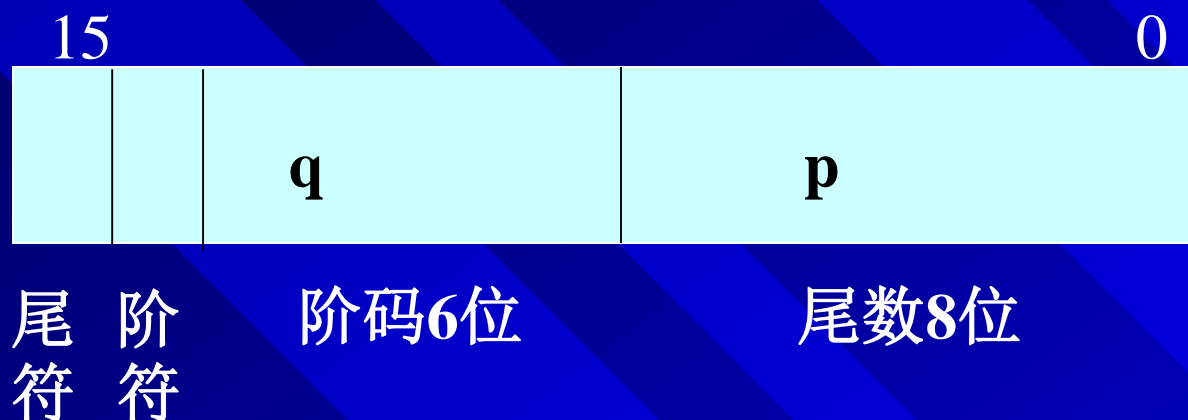
解：最大数 $(1 - 2^{-8})2^{63}$

最小正数 $2^{-1} \times 2^{-64} = 2^{-65}$

最大数与次大数的差值 $2^{-8} \times 2^{63} = 2^{55}$

最小正数与次小正数的差值 $(2^{-1} + 2^{-8})2^{-64}$

解：阶码为7位移码表示，1位符号位，尾数8位，原码规格化表示，基数为2，其格式为：



浮点数N $N = m \times r_m^e \quad e = r_e^q$

尾数基值 $r_m = 2$ (二进制) 阶码的基值 $r_e = 2$,
尾数长度 $p = 8$ (不包括符号位), 阶码长度 $q = 6$ (不包括符号位),
规格化表示的正数的范围:

规格化表示的正数的范围:

最小尾数为

$$r_m^{-1} = 2^{-1} = 1/2$$

最大尾数为

$$1 - r_m^{-p} = 1 - 2^{-8} = (0.11111111)_2$$

最大阶码为

$$2^q - 1 = 2^6 - 1 = 63$$

最小阶码为

$$-2^{-q} = -2^{-6} = -64$$

可表示的最大正浮点数为

$$(1 - r_m^{-p}) \times r_m^{(2^q - 1)} = (1 - 2^{-8}) 2^{63}$$

可表示的最小正浮点数为

$$r_m^{-1} \times r_m^{-2^q} = 2^{-1} \times 2^{-64}$$

可表示的正阶、正尾规格化数的个数为

$$2^p \times 2^q \times 1/2 = 2^{13}$$

规格化表示的负数的范围:

最小尾数（原码）为

$$-(1 - r_m^{-1}) = -(1 - 2^{-8})$$

最大尾数为

$$-r_m^{-1} = -1/2 = -(0.1)_2$$

最大阶码为

$$2^q - 1 = 2^6 - 1 = 63$$

最小阶码为

$$-2^{-q} = -2^{-6} = -64$$

可表示的最大负浮点数为

$$-r_m^{-1} \times r_m^{-2^q} = -2^{-1} \times 2^{-64} = -2^{-65}$$

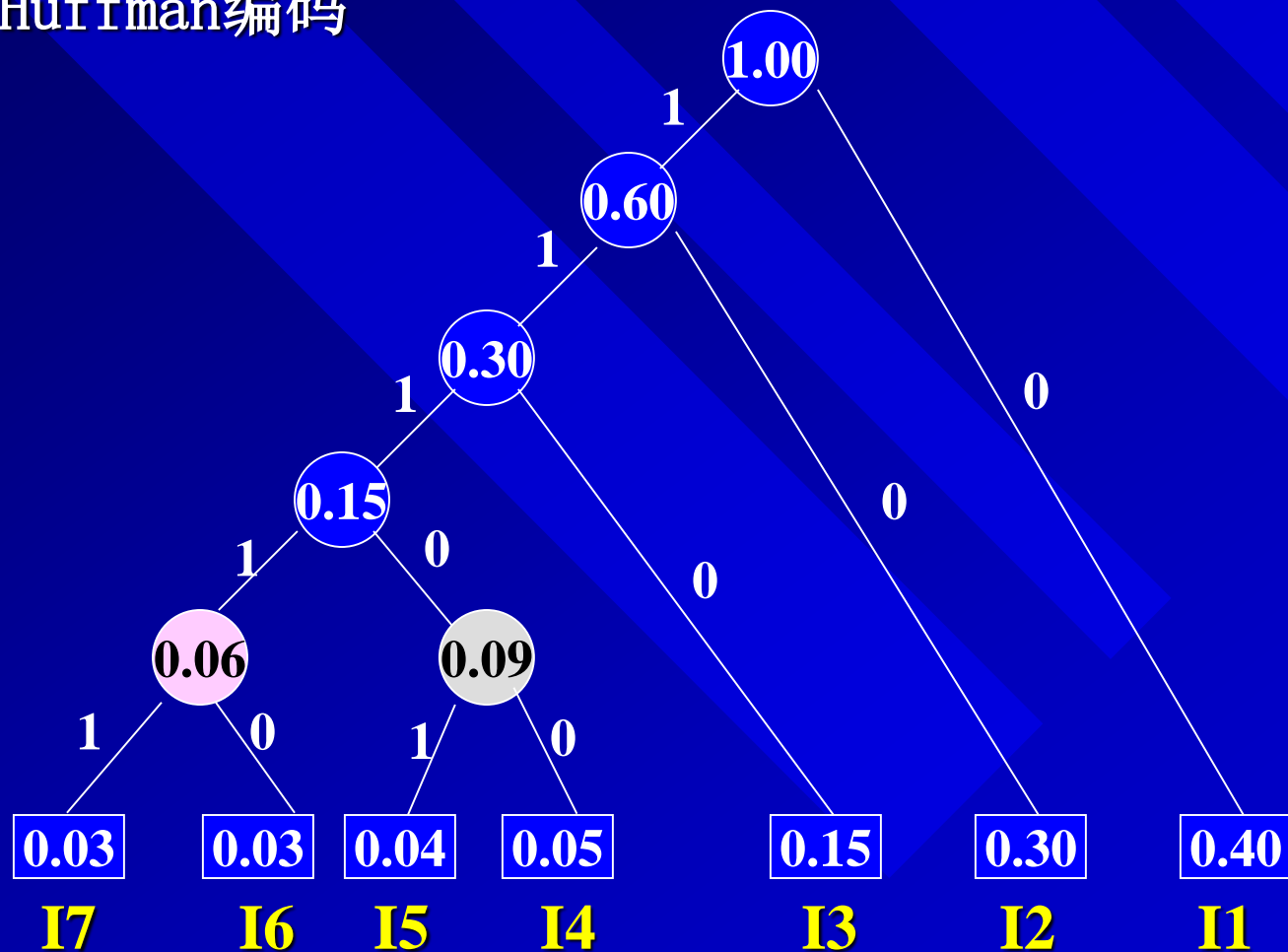
可表示的最小负浮点数为

$$-(1 - r_m^{-p}) \times r_m^{(2^q - 1)} = -(1 - 2^{-8})2^{63}$$

例2.1 某指令系统各指令使用频度如下：

I1	I2	I3	I4	I5	I6	I7
0.4	0.3	0.15	0.05	0.04	0.03	0.03

1. Huffman编码



由此可得到哈夫曼编码如下：

I1: 0 I2: 10 I3: 110 I4: 11100

I5: 11101 I6: 11110 I7: 11111

平均码长

$$L=0.4*1+0.3*2+0.15*3+0.05*5+0.04*5 \\ +0.03*5+0.03*5 = 2.20\text{位}$$

$$\text{信息冗余量 } R = (2.20 - 2.17) / 2.20 = 1.36\%$$

指令长度个数=4

2. 扩展哈夫曼编码

- I1, I2, I3 用两位: 00, 01, 10
- I4, I5, I6, I7 用四位: 1100, 1101, 1110, 1111

$$L = (0.4 + 0.3 + 0.15) * 2 + (0.05 + 0.04 + 0.03 + 0.03) * 4$$
$$= 2.30 \text{ 位}$$

$$\text{信息冗余量} = (2.30 - 2.20) / 2.30 = 0.0435 = 4.35\%$$

操作码的扩展（等长扩展）

指 令	频 度 (P_i)	操作码OP 使用哈夫曼 编码	OP长 度 l_i	huffma n扩展编 码	OP长度 l_i
I1	0.40	0	1	0 0	2
I2	0.30	1 0	2	0 1	2
I3	0.15	1 1 0	3	1 0	2
I4	0.05	1 1 1 0 0	5	1 1 0 0	4
I5	0.04	1 1 1 0 1	5	1 1 0 1	4
I6	0.03	1 1 1 1 0	5	1 1 1 0	4
I7	0.03	1 1 1 1 1	5	1 1 1 1	4

平均码长:

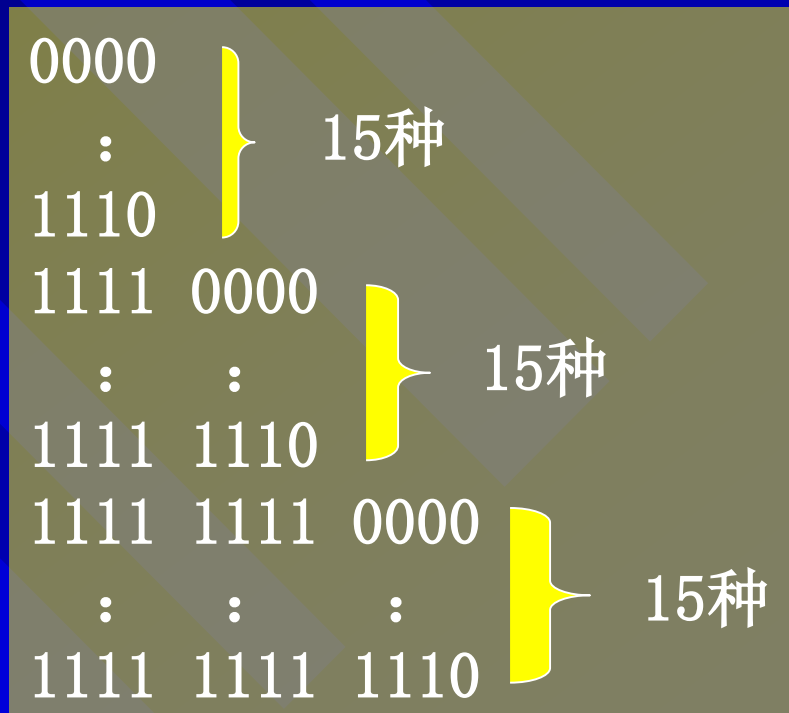
2.2

2.3

例2.2 指令系统共有42种指令，前15种使用频率平均为0.05，中间13种使用频率平均为0.015，最后14种使用频率平均为0.004。如何编码？

解：因频率分布有三种，故码长可有三种；

因每段指令数基本相同，故可采用**等长扩展**(4-8-12位)，保留特征码的每段指令数相同(15-15-15)方法。结果如图所示；



结果：采用15-15-15扩展方法，最后一种编码用于扩展，每段0000~1110用于编码，1111用于扩展。

例2.3 指令系统共有74种指令，前4种使用频率平均为0.12，中间15种使用频率平均为0.02，最后55种使用频率平均为0.004。如何编码？

解：同上例方法，码长可有三种；

因每段指令数成比例
(1: 4)，故可采用**等长
扩展方法(3-6-9位)扩展**，
保留标志位方法，结果
如图所示；

0xx 4种

1xx 0xx 16种

1xx 1xx 0xx 64种

结果：采用4-16-64扩展方法，编码第一位用于扩展，
每段0XX用于编码，1XX用于扩展。

4-16-64平均码长

$$=0.12*4*3+0.02*15*6+0.004*55*9=5.22;$$

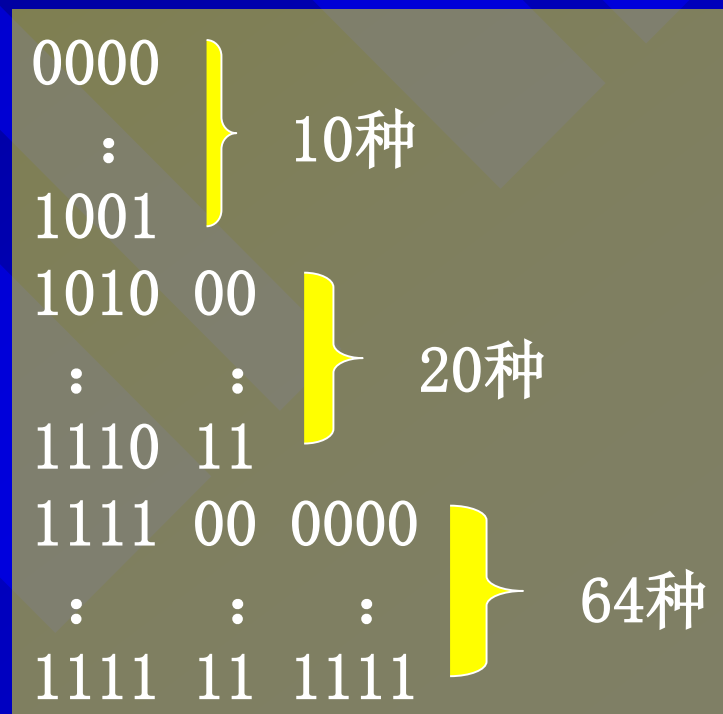
例2.4 指令系统共有78种指令，前10种使用频率平均为0.049，中间18种使用频率平均为0.02，最后50种使用频率平均为0.003。如何编码？

解：同上例方法，码长可有三种；因每段指令数比例为1：2：5，故不可采用等长扩展，采用不等长编码（4-6-10位）较能减少平均码长。

第一种采用4位编码中前10种（0000~1001）；

第二种采用第一种频率编码中的后5种编码（1010~1110）与扩展的2位共20种编码；

第三种采用第一种频率编码中的最后一种（1111）与扩展的6位共64种编码。



作2.3 设某机阶码6位、尾数48位，阶符和数符不在其内，当尾数分别以2、8、16为基时，在非负阶、正尾数、规格化数情况下，求出其最小阶、最大阶、阶的个数、最小尾数值、最大尾数值、可表示的最小值和最大值及可表示的规格化数的总个数

解： $p=6$ 、 $m=48$ 时，在非负阶、规格化、正尾数情况下，尾基 $r_m=2$ 、8、16时的各个参数的计算结果如下表所示。

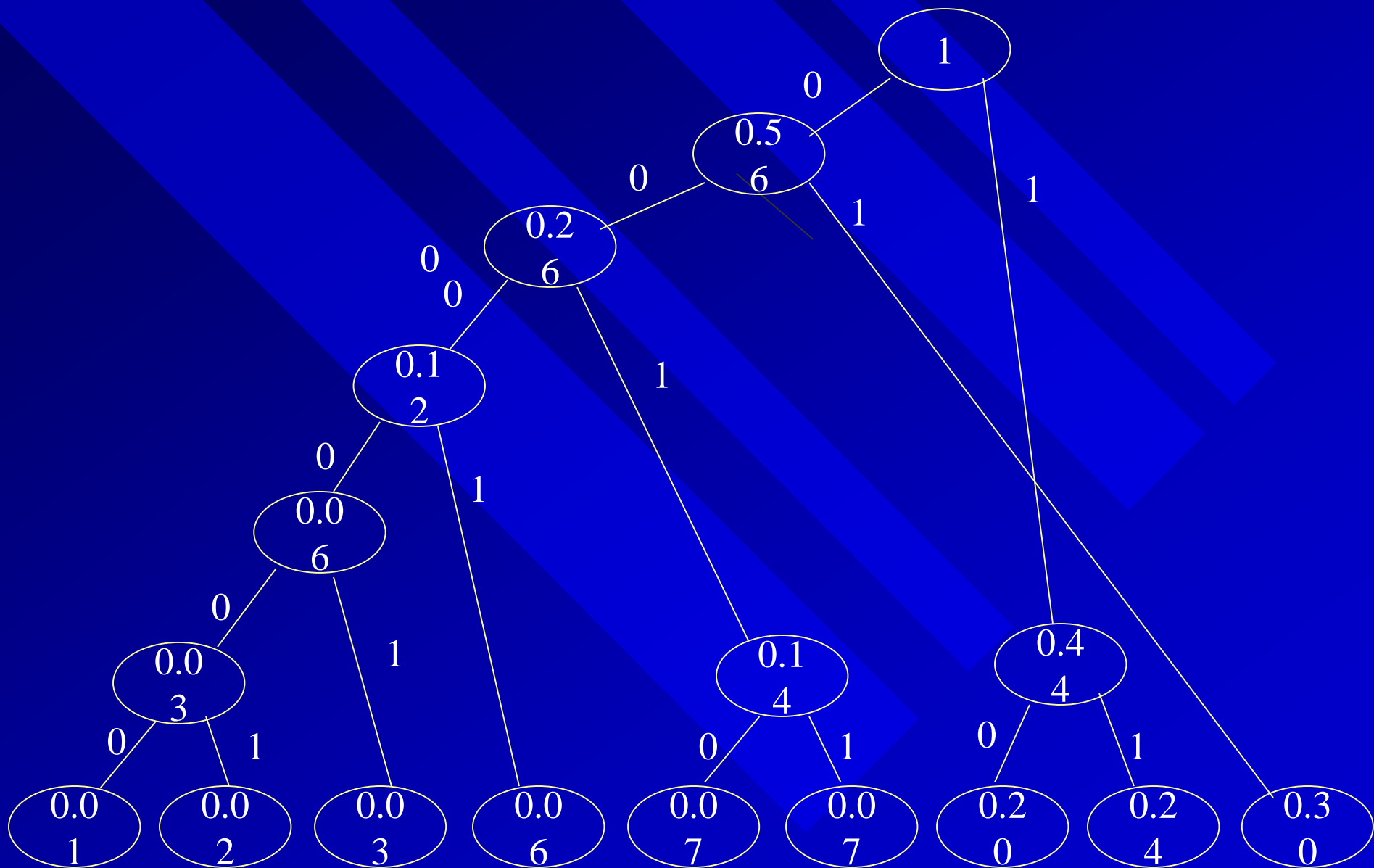
非负阶、正尾数、规格化		尾基 r_m (p=6位, m=48位)		
		2 (48位)	8 (16位)	16 (12位)
最小阶值	0	0	0	0
最大阶值	2^p-1	63	63	63
阶的个数	2^p	64	64	64
尾数最小值	r_m^{-1}	1/2	1/8	1/16
尾数最大值	$1 - r_m^{-m}$	$1-2^{-48}$	$1-8^{-16}$	$1-16^{-12}$
最小值	r_m^{-1}	1/2	1/8	1/16
最大值	$r_m^{2^p-1} \bullet (1-r_m^{-m})$	$2^{63} \bullet (1-2^{-48})$	$8^{63} \bullet (1-8^{-16})$	$16^{63} \bullet (1-16^{-12})$
数的个数	$2^p \bullet r_m^m \bullet \frac{r_m-1}{r_m}$	2^{53}	$7 \bullet 2^{51}$	$15 \bullet 2^{50}$

作2.15 某模型机有9条指令，其使用频率为：

ADD（加）	30%	SUB（减）	24%
JOM（按负转移）	6%	STO（存）	7%
JMP（转移）	7%	SHR（右移）	2%
CIL（循环左移）	3%	CLA（清加）	20%
STP（停机）	1%		

要求有两种指令字长，都按双操作数指令格式编，采用扩展操作码，并限制只能有两种操作码码长。设该机有若干个通用寄存器，主存为16位宽，按字节编址，采用整数边界存贮，任何指令都在一个主存周期中取得，短指令为寄存器—寄存器型，长指令为寄存器—主存型，主存地址应能变址寻址。

解：(1) Huffman树的形式如图所示。



● 由上图可得到的Huffman编码为:

ADD (加)	30%	01
SUB (减)	24%	11
CLA (清加)	20%	10
JOM (按负转移)	6%	0001
STO (存)	7%	0011
JMP (转移)	7%	0010
CIL (循环左移)	3%	00001
SHR (右移)	2%	000001
STP (停机)	1%	000000

因此, 操作码的平均码长为:

$$l = \sum_{i=1}^9 p_i \times l_i = 2.61 \text{ 位}$$

● (2) 采用2-5扩展的操作码编码为:

ADD(加)	30%	00
SUB(减)	24%	01
CLA(清加)	20%	10
JOM(按负转移)	6%	11000
STO(存)	7%	11001
JMP(转移)	7%	11010
SHR(右移)	2%	11011
CIL(循环左移)	3%	11100
STP(停机)	1%	11101

因此, 操作码的平均码长为:

$$\sum_{i=1}^9 p_i \times l_i = (0.30 + 0.24 + 0.20) \times 2 + 0.26 \times 5 = 2.78 \text{位}$$

(3) 该机允许使用的可编址的通用寄存器个数为 $2^3=8$ 个

(4) 短指令为寄存器-寄存器型，格式如下：



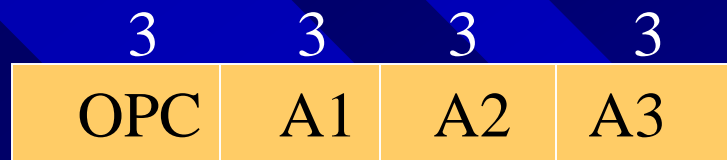
长指令为寄存器-主存型，格式如下：



(5) 访主存操作数地址寻址的最大相对位移量为64个字节(-32~+31个字节)

例2.5 若某机要求有：三地址指令4条，单地址指令192条，零地址指令16条。设指令字长为12位，每个地址码长3位。问能否以扩展操作码为其编码？

解： 1. 三种指令格式字如下：



三地址指令4条

000 xxx xxx xxx

⋮

011 xxx xxx xxx

100 000 000 xxx

⋮

⋮

111 111 101 xxx

111 111 110 000

⋮

111 111 111 111

三地址4条

一地址192条

零地址16条



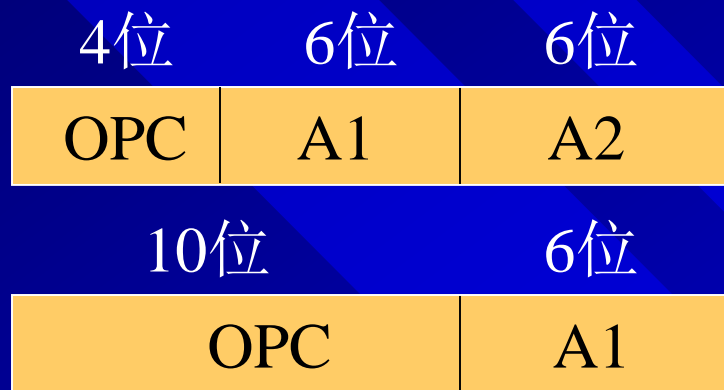
单地址指令192条



零地址指令16条

2.某机指令字长16位。设有单地址指令和双地址指令两类。若每个地址字段为6位，且双地址指令有X条。问单地址指令最多可有多少条？

解：二种指令格式字如下：



由于双地址指令有X条，单地址指令最多可有：

$$(2^4 - X) \times 2^6 \text{ 条 } (X \neq 16)$$

作2.16 某处理机的指令字长为16位，有双地址指令、单地址指令和零地址指令三类，并假设每个地址字段的长度为16位。

(1) 如果双地址指令有15条，单地址和零地址指令的条数基本相同，问单地址指令和零地址指令各有多少条？并且为这三类指令分配操作码。

(2) 如果三类指令的比例为1: 9: 9，问双地址指令、单地址指令和零地址指令各有多少条？并且为这三类指令分配操作码。

解:

(1) 双地址指令15条, 地址码: 0000~1110

单地址指令 $2^6 - 1 = 63$ 条, 地址码: 1111 000000

1111 111110

零地址指令64条, 地址码: 1111 111111 000000

1111 111111 111111

(2) 双地址指令14条, 地址码: 0000~1101

单地址指令 $2^6 \times 2 - 2 = 126$ 条, 1110 000000

1110 111110

零地址指令128条, 地址码: 1111 111110 000000

1111 111111 111111

第3章

作3.2 设一个任务在计算机中的处理时间为50秒，CPU处理时间为30秒，I/O处理时间为20秒。如果将CPU的处理速度提高4倍，则总的处理时间将是多少？

解：如使CPU的速度增加4倍，则CPU的处理时间为：
 $T_{cpu}=30/4=7.5s$

则总的处理时间为： $T=T_{cpu}+T_{i/o}=7.5+20=27.5s$

题3.1 假设一台计算机的I/O处理占10%，当其CPU性能改变，而I/O性能保持不变，系统总体性能会出现什么变化？

(1) 如果CPU的性能提高10倍

(2) 如果CPU的性能提高100倍

解：假设原来的程序执行时间为1个单位时间。如果CPU的性能提高10倍，则程序的计算（包含I/O处理）时间为

$$(1-10\%) / 10 + 10\% = 0.19$$

即整机性能只能提高约5倍，差不多有50%的CPU性能浪费在I/O上。

如果CPU的性能提高100倍，则程序的计算时间为

$$(1-10\%) / 100 + 10\% = 0.109$$

而整机性能只能提高约10倍，表示有90%的性能浪费在I/O上了。

例3.1 某字节交叉多路通道连接6台设备，其数据
传送速率如下表所示。

设备号	1	2	3	4	5	6
传送速率 (B/ms)	50	50	40	25	25	10
二次请求的 间隔时间 (μ S)						

- (1) 在上表中填出设备相应二次请求传送字节的间隔时间。
- (2) 当所有设备同时要传送数据时，求其对通道要求的总流量 f_{byte}

(3) 让通道以极限流量 $f_{\text{max. byte}} = f_{\text{byte}}$ 的工作周期工作，通道的工作周期(即TS+TD的时间间隔)是多少？

(4) 让通道中所挂设备速率越高的数据传送请求被响应的优先级越高。画出6台设备同时发送请求到下次同时发送请求期间里，通道响应和处理完各设备请求时刻的示意图。哪个设备丢失了信息？提出一种不丢失信息的解决办法。

解: (1)

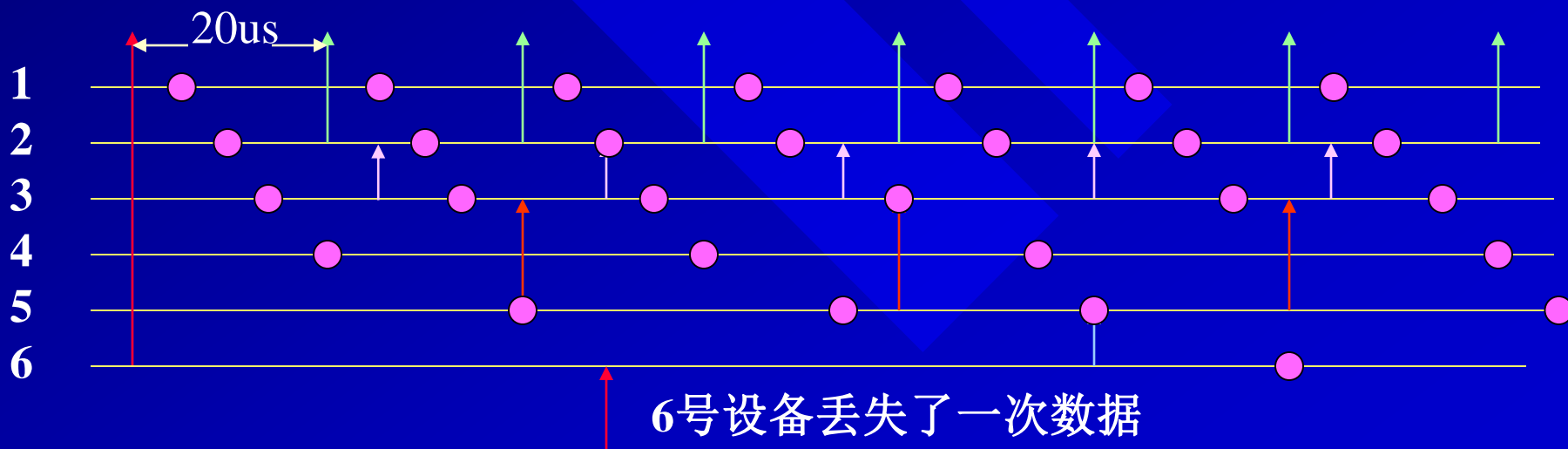
设备号	1	2	3	4	5	6
传送速率(B/ms)	50	50	40	25	25	10
二次请求的间隔时间(μ S)	20	20	25	40	40	100

(2) 总容量

$$f_{byte} = \sum_{i=1}^6 f_i = 200B/ms$$

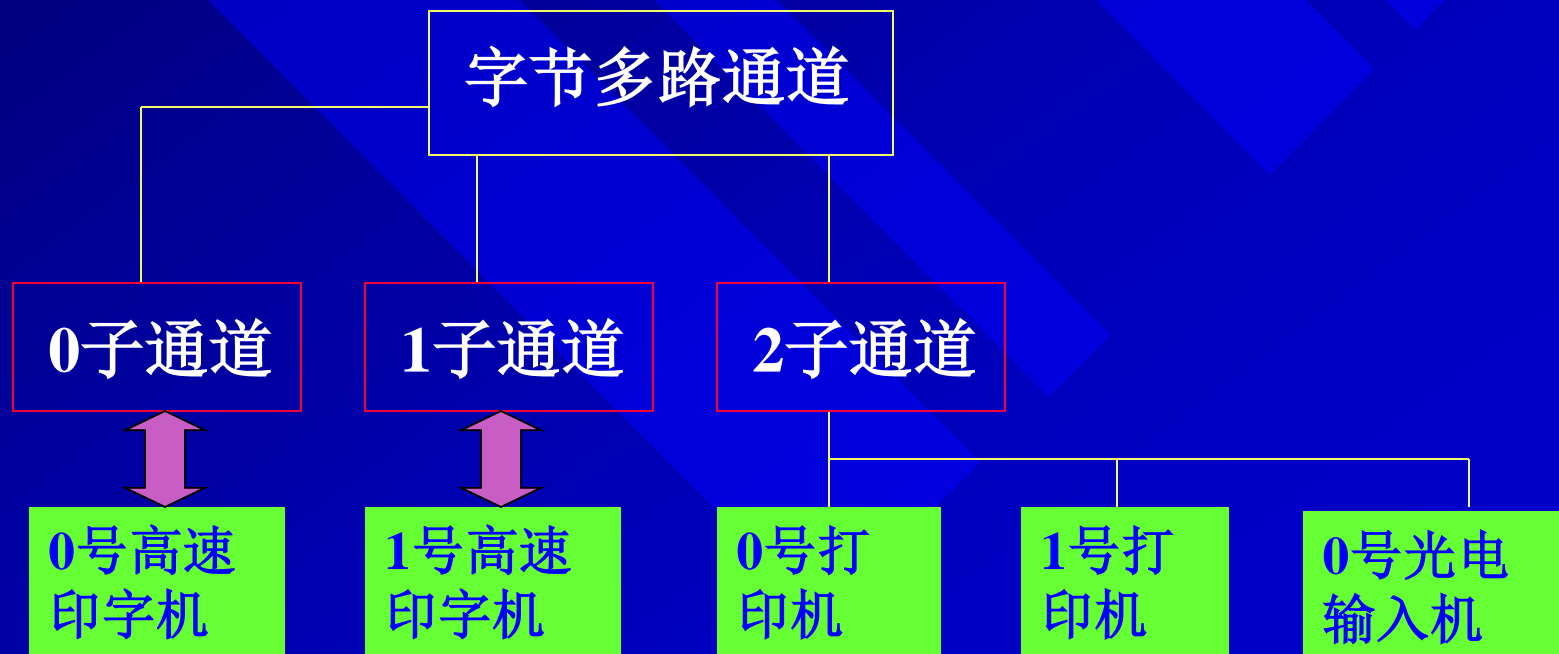
(3) 传送周期

$$T_S + T_D = 1ms/200B = 5\mu S$$



- 方法1：增加通道的最大流量，保证连接在通道上的所有设备的数据传送请求能够及时得到通道的响应
- 方法2：动态改变设备的优先级
- 方法3：增加一定数量的数据缓冲器，特别是对优先级比较低的设备

例3.2 印字机各占一个子通道，0号打印机、1号打印机和0号光电输入机合用一个子通道。假定数据传送期内高速印字机每隔25 μ s发一个请求，低速打印机每隔150 μ s发一个字节请求，光电输入机每隔800 μ s发一个字节请求，则这5台设备要求通道的实际流量为多少？



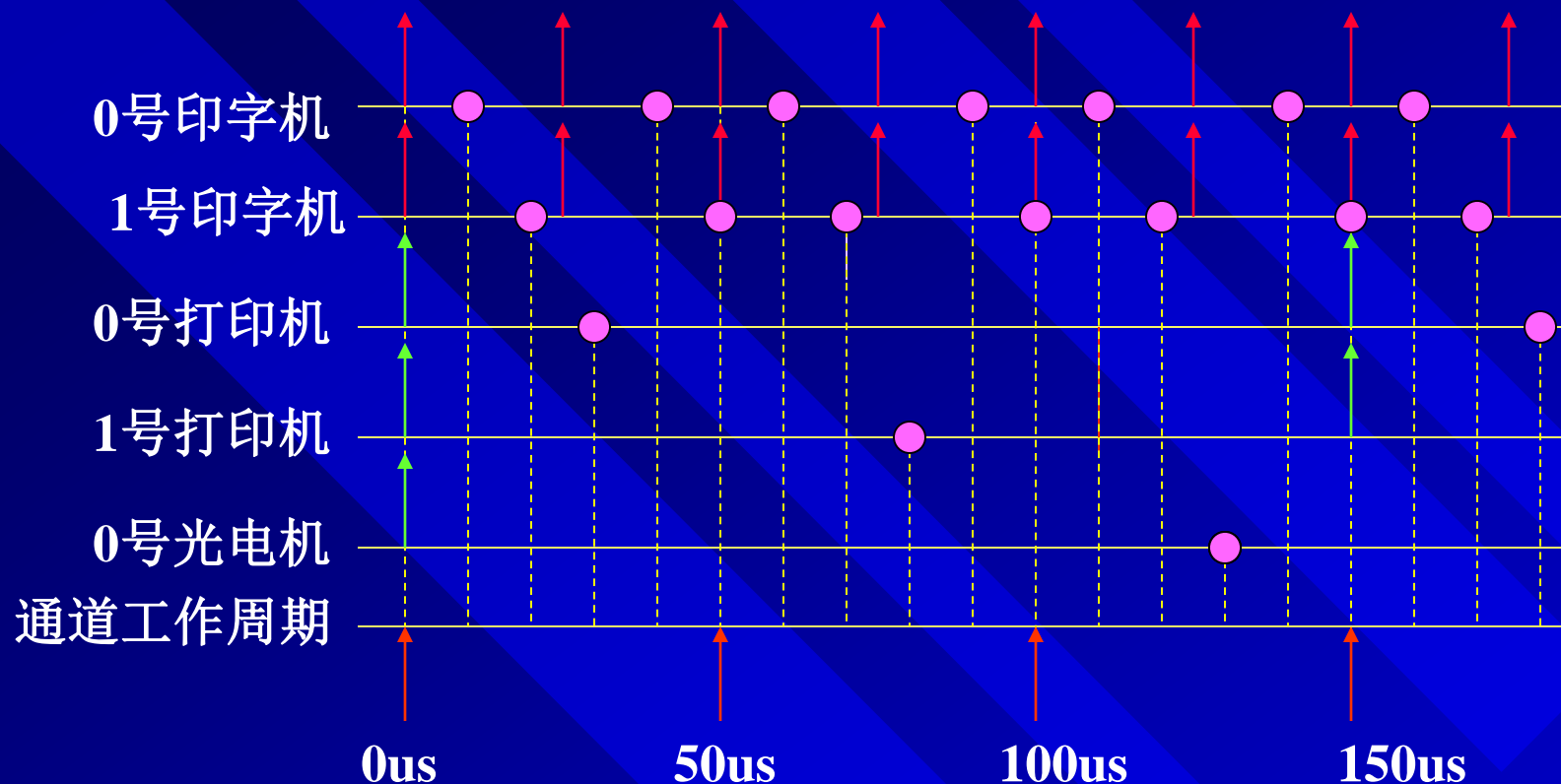
解：5台设备要求通道的数据流量为：

$$f_{byte} = \sum f = \frac{1}{25} + \frac{1}{25} + \left(\frac{1}{150} + \frac{1}{150} + \frac{1}{800} \right) = 0.095 MB/s$$

可将该通道设计成0.1MB/s，即所设计的工作周期为：

$$t = \frac{1}{f_{byte}} = \frac{1}{T_S + T_D} = 10\mu s$$

这样各设备的请求就能及时得到响应和处理，不会丢失信息。



↑ 表示设备提出申请的时刻

● 表示通道处理完设备申请的时刻

优先级次序：0号印字机、1号印字机、0号打印机、1号打印机、0号光电机

例3.3 设通道在数据传送期中，选择设备需 $4.9\ \mu\text{S}$ ，
传送一个字节数据需 $0.1\ \mu\text{S}$ 。

(1) 其低速设备每隔 $250\ \mu\text{S}$ 发出一个字节数据
传送请求，问最多可接多少台这种设备？

(2) 若有A~E共5种高速设备，要求字节传送
的间隔时间如下表所示，其时间单位为 μS 。若
一次通信传送的字节数不少于1024个字节，问哪
些设备可挂在此通道上？哪些则不能？

设备	A	B	C	D	E
时间间隔 (μS)	0.13	0.1	0.11	0.2	0.3

解： (1) 低速设备应接字节多路通道

$$f_{\max \bullet \text{byte}} = \frac{1}{T_S + T_D} = \frac{1}{4.9 + 0.1} = \sum_{i=1}^n f_{\text{byte} \bullet i} = \frac{n}{250}$$

$n = 250/5 = 50$ 台，所以， $n \leq 50$ 台，即最多可接 50 台这种设备

(2) 根据题意，此通道为选择通道

$$f_{\max \bullet \text{select}} = \frac{1}{T_D + \frac{T_S}{n}} = \frac{1}{0.1 + \frac{4.9}{n}}$$

其中， $n \geq 1024$ ，应使 $f_{\text{select} \bullet i} \leq f_{\max \bullet \text{select}}$

由此可得出通道工作周期为： $T \approx 0.1048(\mu\text{s})$

所以，只有 A、C、D、E 可挂在此通道上，B 则不行。

- **作3.2** 设一个任务的处理时间为64秒，CPU在这期间始终忙于处理，I/O处理时间为36秒。为提高系统性能，有两种方案：使CPU的速度增加1倍，或者使CPU和I/O的处理速度同时增加1倍。计算这两种情况下的处理时间。

解：如使CPU的速度增加1倍，则CPU的处理时间为：

$$T_{cpu}=64/2=32$$

则总的处理时间为： $T=T_{cpu}+T_{i/o}-T_{overlap}$

$$\because T_{overlap} \leq \min\{T_{cpu}, T_{i/o}\}$$

$$\therefore T \geq 32+36-32=36$$

当两者速度同时增加1 倍时：

$$T_{cpu}=64/2=32 \quad T_{i/o}=18 \quad \text{则：}$$

$$T \geq 32+18-18=32$$

- 作3.10 (1) 字节多路通道、数组多路通道、选择通道，它们一般用什么数据宽度来进行通信？
- (2) 如果通道在数据传送期中选择设备需 $9.8\mu\text{s}$ ，传送一个字节数据需 $0.2\mu\text{s}$ ，某低速设备每隔 $500\mu\text{s}$ 发出一个字节数据传送请求，问至多可接几台这种低速设备？对于如下A~F6种高速设备，一次通信传送的字节数不少于1024个字节，问哪些设备可以挂在此通道上？哪些则不能？其中A~F设备每发一个字节数据传送请求的时间间隔分别为

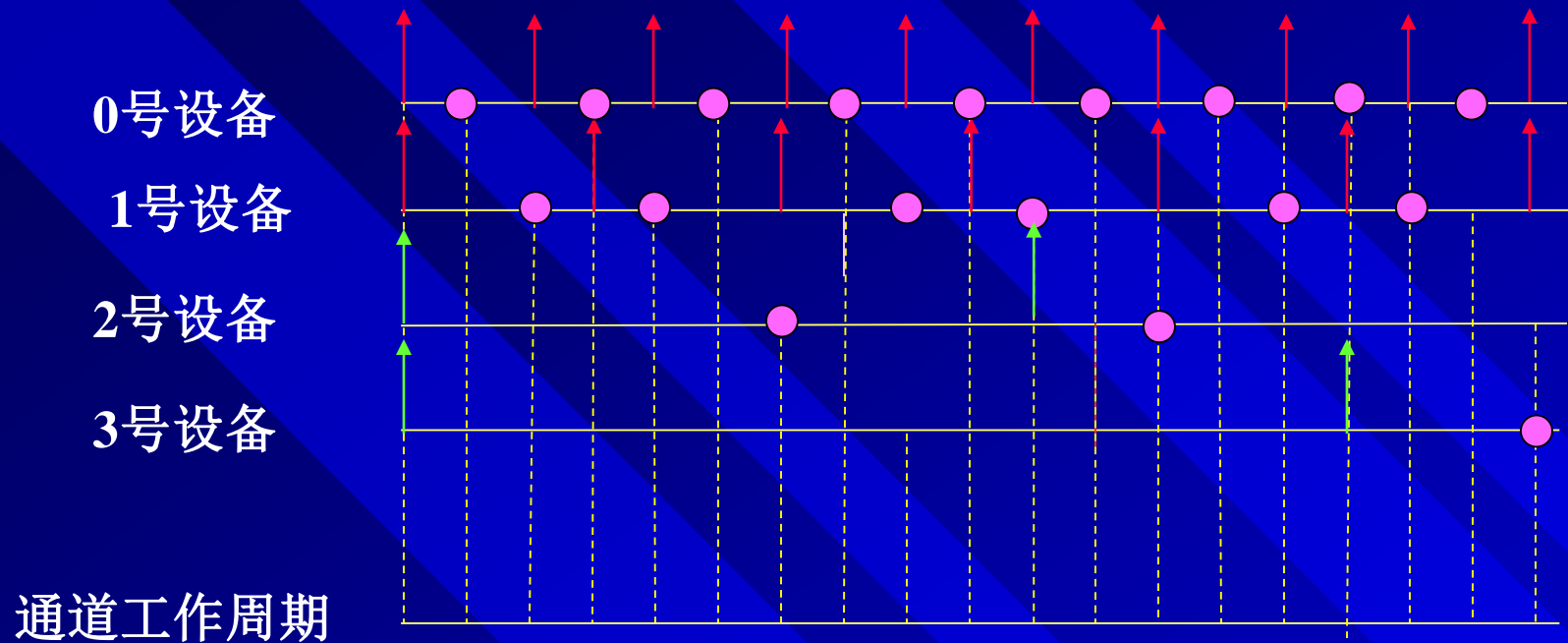
设备	A	B	C	D	E	F
发申请间隔 (μs)	0.2	0.25	0.5	0.19	0.4	0.21

解：（1） 字节通道的数据流量为：

$$f_{byte} = \sum f = \frac{1}{10} + \frac{1}{75} + \frac{1}{15} + \frac{1}{50} = 0.25 MB / s$$

（2） 传送周期

$$T_s + T_d = 3\mu s + 2\mu s = 5 \mu s$$



(3) 通道处理完成各设备第一次服务请求的时刻:

0号设备: 5us, 1号设备: 10us, 2号设备: 30 us, 3号设备: 丢失

↑ 表示设备提出申请的时刻

● 表示通道处理完设备申请的时刻

优先级次序: 0号设备、1号设备、2号设备、3号设备

(4) 从时间关系图中看，这个字节通道不能正常工作，因为以现在的工作频率，不可能连接过多的设备。

(5) 改进的措施：

- 1) 增加此通道的工作频率；
- 2) 改变设备的优先级；
- 3) 增加一定数量的缓冲器，保存优先级低设备可能丢失的数据。

第4章

- **例4.1** 假设高速缓存Cache工作速度为主存的5倍，且Cache被访问命中的概率为90%，则采用Cache后，能使整个存储系统获得多高的加速比？

解：

$$S_p = \frac{T_{A2}}{T_A} = \frac{5T_{A1}}{HT_{A1} + (1-H)T_{A2}} = \frac{5}{0.9 + 0.5} = 3.57$$

- **例4.2** 假设高速缓存Cache的访问周期为50ns，主存的访问周期为400ns，且Cache被访问命中的概率为95%，则采用Cache后，能使整个存储系统等效的访问周期为多少？获得多高的加速比？

解：

$$T_A = HT_{A1} + (1-H)T_{A2} = 0.95 \times 50 + 0.05 \times 400 = 67.5ns$$

$$S_p = \frac{T_{A2}}{T_A} = \frac{400}{67.5} = 5.9$$

- **例4.3** 设某用户虚存共有8页, 主存有4页, 每页大小为1KB. 试根据页表计算出虚地址1023和6800的主存实地址。

页表

虚页号	实页号	装入位
0	3	1
1	1	1
2	2	0
3	3	0
4	2	1
5	1	0
6	0	1
7	0	0

提示：注意页表中虚、实页对应关系



解：页号与地址对应关系

第0页	0—1023
第1页	1024—2047
第2页	2048—3071
第3页	3072—4095
第4页	4096—5119
第5页	5120—6143
第6页	6144—7167
第7页	7168--8191

虚页号 = 虚地址 % 1024

虚地址**1023**，虚页号为0，页内位移为1023；根据虚页号查页表得知实页

号为3，且装入位为1。

主存实地址 $PA = 3072 + 1023 = 4095$

虚地址**6800**，虚页号为6，页内位移为656；根据虚页号查页表得知实页号为0，且装入位为1。

主存实地址 $PA = 0 + 656 = 656$

每页首地址 = 页号 × 每页大小

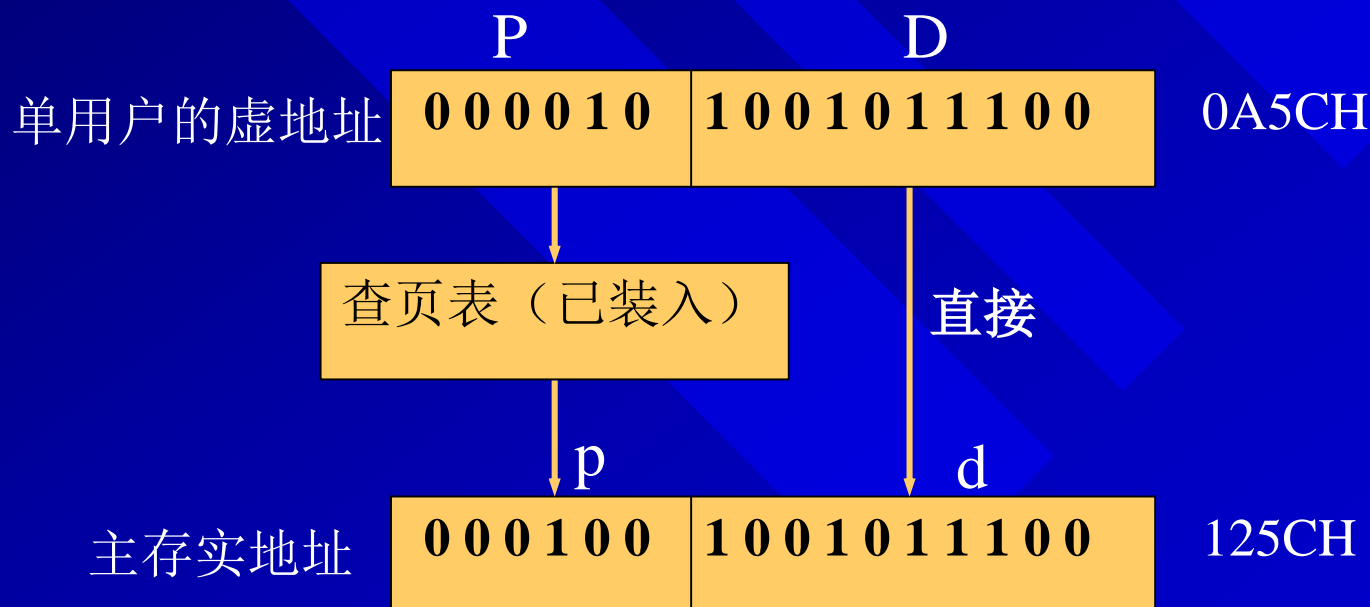
- **例4.4** 虚存32页, 主存16页, 每页为1KB。某时刻已调入主存的实页与虚页对应如下:

虚页号 0 1 2 8

实页号 5 10 4 7

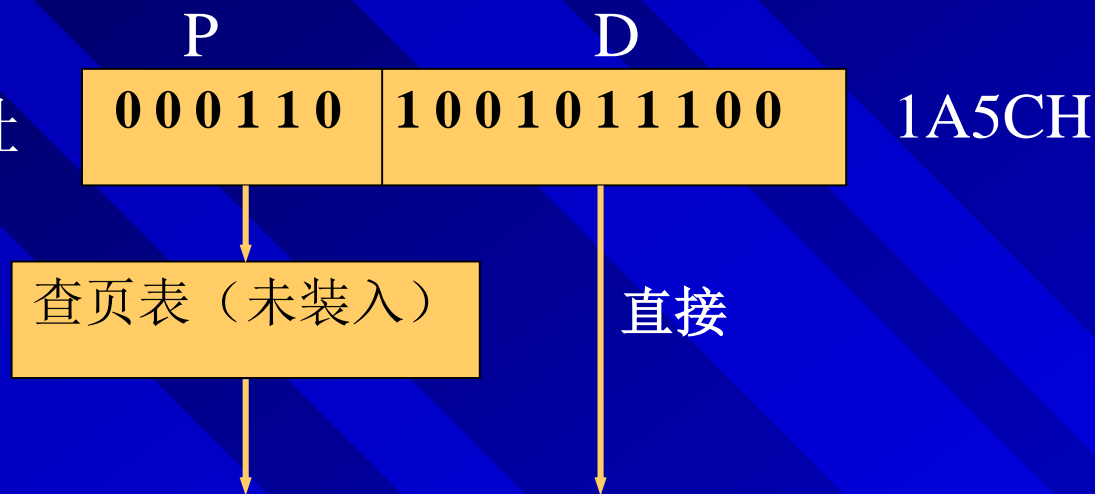
求虚地址0A5CH和1A5CH对应的实地址。

解:



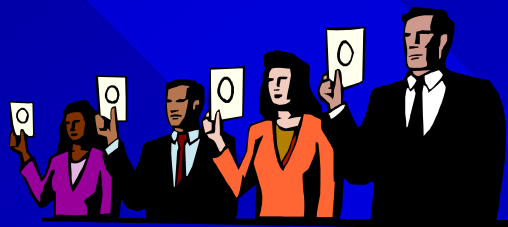
解:

单用户的虚地址

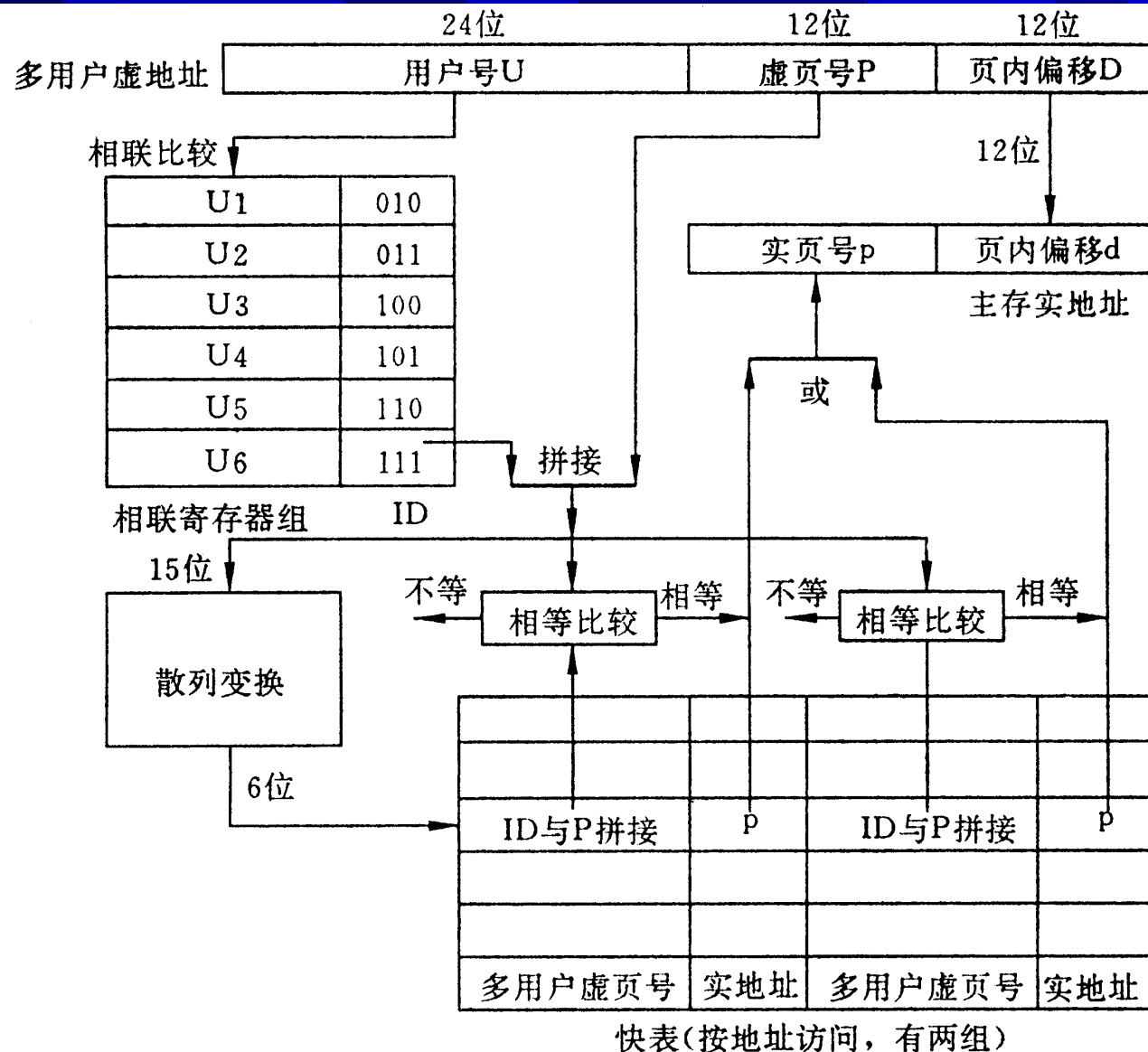


页面失效，无对应的主存实地址。

用户的虚页并未装入主存中，当执行该虚页程序时，找不到对应的实页。



例4.5 虚拟存储器举例—IBM370/168



$$2^{12}=4K$$

$$2^{12}=4KB$$

15位 页数?

20位

64行

24+24位

相等比较
快表行数?
快表每行的位数?

如何减少散列冲突?
快表为何采用两组?
相等比较的作用?

例4.6

假若一计算机的主存储器按64块组织，块大小为8个字，高速缓存有8块。试按下述要求画图回答问题。

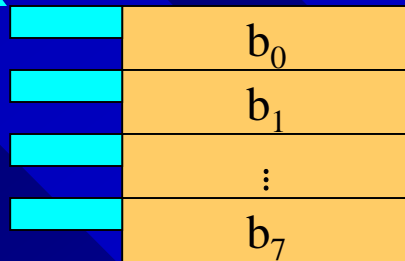
(1) 画出全相联映像示意图、指定标记字段和主存地址格式。

(2) 画出直接映像示意图、指定标记字段和主存地址格式。

(3) 画出2路组相联映像示意图、指定标记字段和主存地址格式。

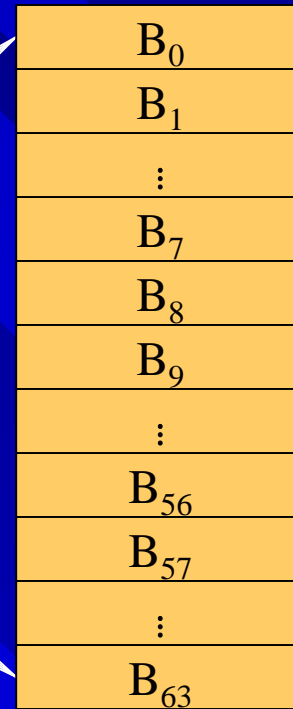
解：(1) 全相联映像

标记



标记6位的高速缓存

\vdots



主存储器

主存地址 (块号) 标记 (6位) 字 (3位)

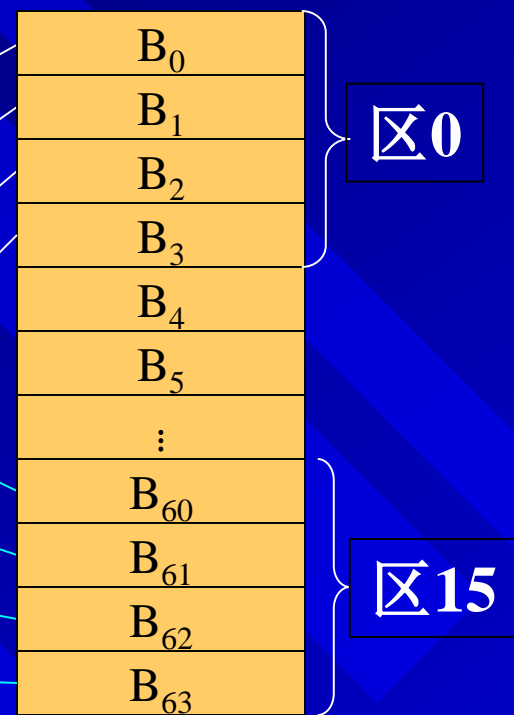
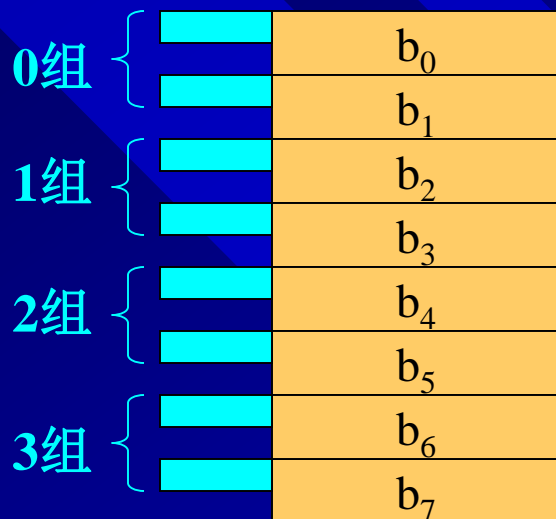
缓存地址 行号 (3位) 字 (3位)

(2) 直接映像



(3) 2路组相联映像

标记4位的高速缓存



主存储器



● 作4.6 解:

(1) 由于虚地址中有2位表示段号、2位表示页号，所以此地址空间共有 $2^2 \times 2^2 = 16$ 个虚页。

段号	段0	段1	段2	段3
访问方式	只读	可读/执行	可读/写/执	可读/写
虚页0	实页9	在辅存内	页表不在主存	实页14
虚页1	实页3	实页0	页表不在主存	实页1
虚页2	在辅存内	实页15	页表不在主存	实页6
虚页3	实页12	实页8	页表不在主存	在辅存内

(2) 根据上页表，题中所示操作的虚地址访问情况如下

其中，实地址=实页号 \times 页大小+页内位移。如下表第1行中，
实地址=3 \times 2048+1=6145

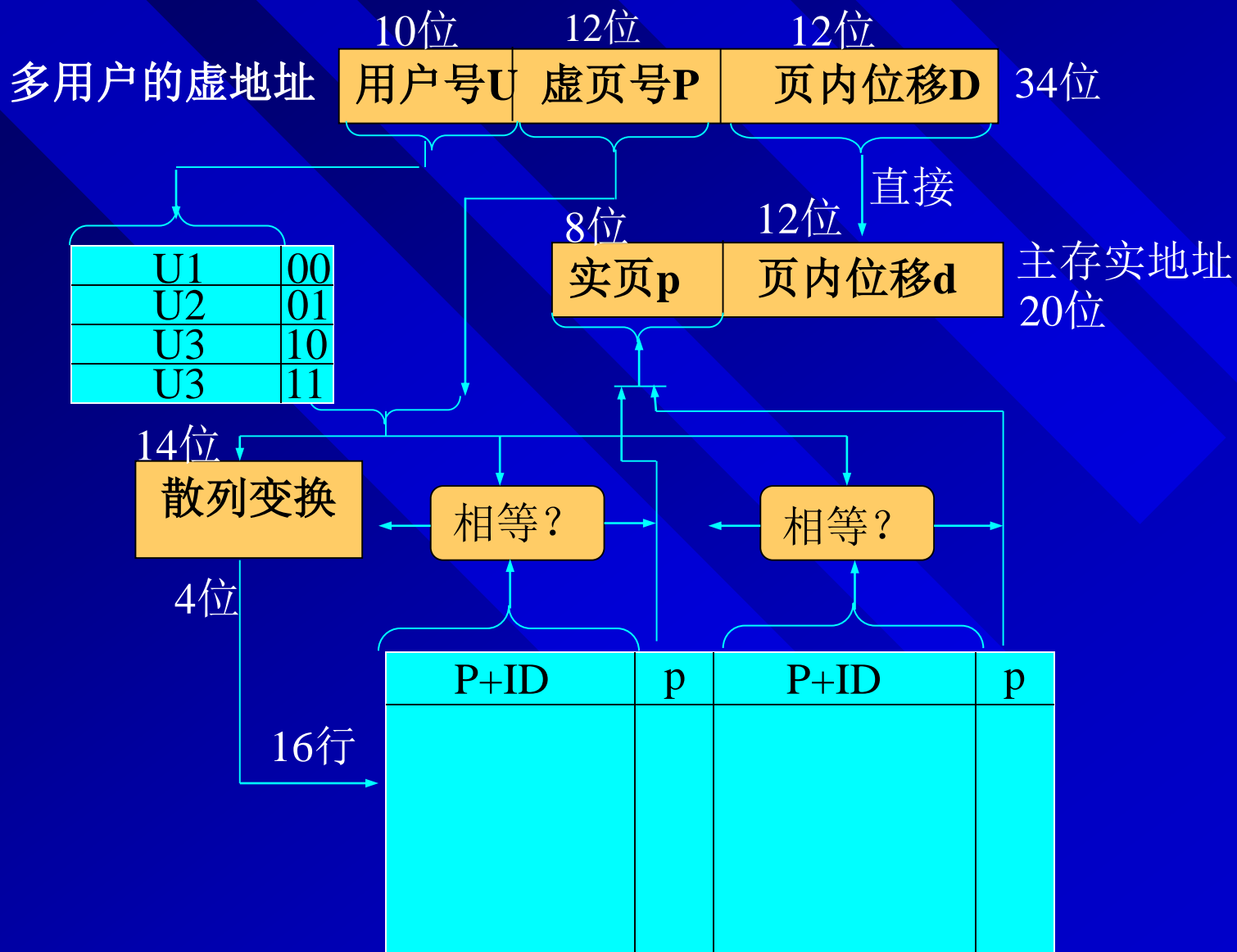
方式	段	页	页内位移	段失效	页失效	实页号	实地址	保护失效
取数	0	1	1	无	无	3	6145	无
取数	1	1	10	无	无	0	10	无
取数	3	3	2047	无	有	无	无	/
存数	0	1	4	无	无	3	6148	有
存数	2	1	2	有	/	无	无	/
存数	1	0	14	无	有	无	无	/
转移至此	1	3	100	无	无	8	16484	无
取数	0	2	50	无	有	无	无	/
取数	2	0	5	有	/	无	无	/
转移至此	3	0	60	无	无	14	28732	有

●作4.8 解:

(1) \because 可有1K个任务, 短期内只有4个用户,
 \therefore 指示用户号的地址字段 $U=10$ 位; $ID=2$ 位;
又 \because 每个用户程序空间可达4096页, 每页512B,
 $\therefore P=12, D=(9+3)=12$ 位;
又 \because 主存地址长度为20位, \therefore 实页号 $p=20-12=8$ 位。

- (2) 每个相联寄存器的相联比较位数为 $U=10$ 位;
- (3) 每个相联寄存器的总位数为 $U+ID=12$ 位;
- (4) 散列变换硬件的输入和输出位数为14位和4位;
- (5) 每个相等比较器的位数为 $P+ID=12+2=14$ 位;
- (6) 快表的总容量为 $16 \times (12+2+8) \times 2=704$ (位)。

虚实地址经快表变换的逻辑示意图



● 作4.12 解:

依题意可知，每块大小为 $4 \times 4 = 16\text{B}$ ，用4位二进制表示。
采用4个外相等比较电路，指示块号用2位二进制表示。
指示组号地址为 $10 - 2 - 4 = 4$ 位。

又因为主存容量为256KB，主存总位数18位。所以，指示区号的位数为 $18-10=8$ 位。

地址对应关系如下:



相联目录表如下：

16行 {	E+B	b	E+B	b	E+B	b	E+B	b

相联目录表行数： $2^q=2^4=16$ 行；

每行总位数： $E+B+b=8+2+2=12$ 位；

每个相等比较电路的位数： $E+B=8+2=10$ 位。

- **题4.1** 在一个Cache存储系统中， Cache的访问周期为10ns， 主存储器的访问周期为60ns， 每个数据在Cache中平均重复使用4次。当块大小为1个字节时，存储系统的访问效率只有0.5，现在要通过增加块大小，使存储系统的访问效率达到0.94。
- (1) 当存储系统的访问效率为0.5时，计算命中率和等效访问周期。
- (2) 为了使存储系统的访问效率达到0.94，命中率和等效访问周期应该提高到多少？
- (3) 为了使存储系统的访问效率从0.5提高到0.94，块的大小至少增加到几个字？

解：（1）当存储系统的访问效率为0.5时，由表达式

$$e = \frac{T_{A1}}{T_A} = \frac{T_{A1}}{HT_{A1} + (1-H)T_{A2}} = \frac{1}{H + (1-H)T_{A2}/T_{A1}}$$

可求出命中率为

$$0.5 = \frac{1}{H_1 + (1-H_1)60/10} \quad H_1 = 0.8$$

等效访问周期为

$$\begin{aligned} T_1 &= T_c \times H_c + (1-H_c) \times T_m \\ &= 10 \times 0.8 + 60 \times 0.2 = 20(ns) \end{aligned}$$

$$\text{或由 } e = \frac{T_{A1}}{T_A} = \frac{T_c}{T_1} = \frac{10}{T_1} = 0.5 \quad \text{得} \quad T_1 = 20(ns)$$

(2) 当存储系统的访问效率提高到0.94时, 命中率应该提高到 H_2

$$0.94 = \frac{1}{H_2 + (1 - H_2)60/10} \quad H_2 = 0.987$$

等效访问周期应提高为

$$T_2 = 10 \times 0.987 + 60 \times (1 - 0.987) = 10.64(ns)$$

$$\text{或由 } e = \frac{T_{A1}}{T_A} = \frac{T_c}{T_2} = \frac{10}{T_2} = 0.94 \quad \text{得 } T_1 = 10.64(ns)$$

(3) 为了使存储系统的访问效率由0.5提高到0.94，块大小应为B个字。则有

$$H_2 = \frac{H_1 + n - 1}{n}$$

其中n为Cache的块大小与数据重复使用次数的乘积， H_1 是原来的命中率， H 是块大小增加后的命中率。

$$H_2 = \frac{H_1 + 4B - 1}{4B}$$

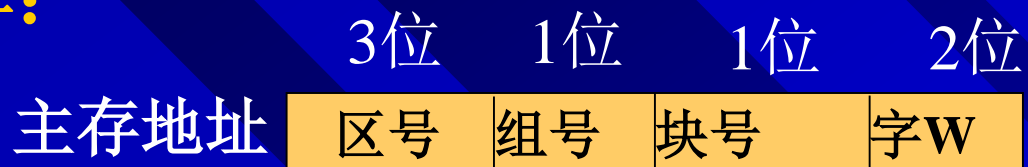
在上式中代入相关参数，可求出

$$B = 4$$

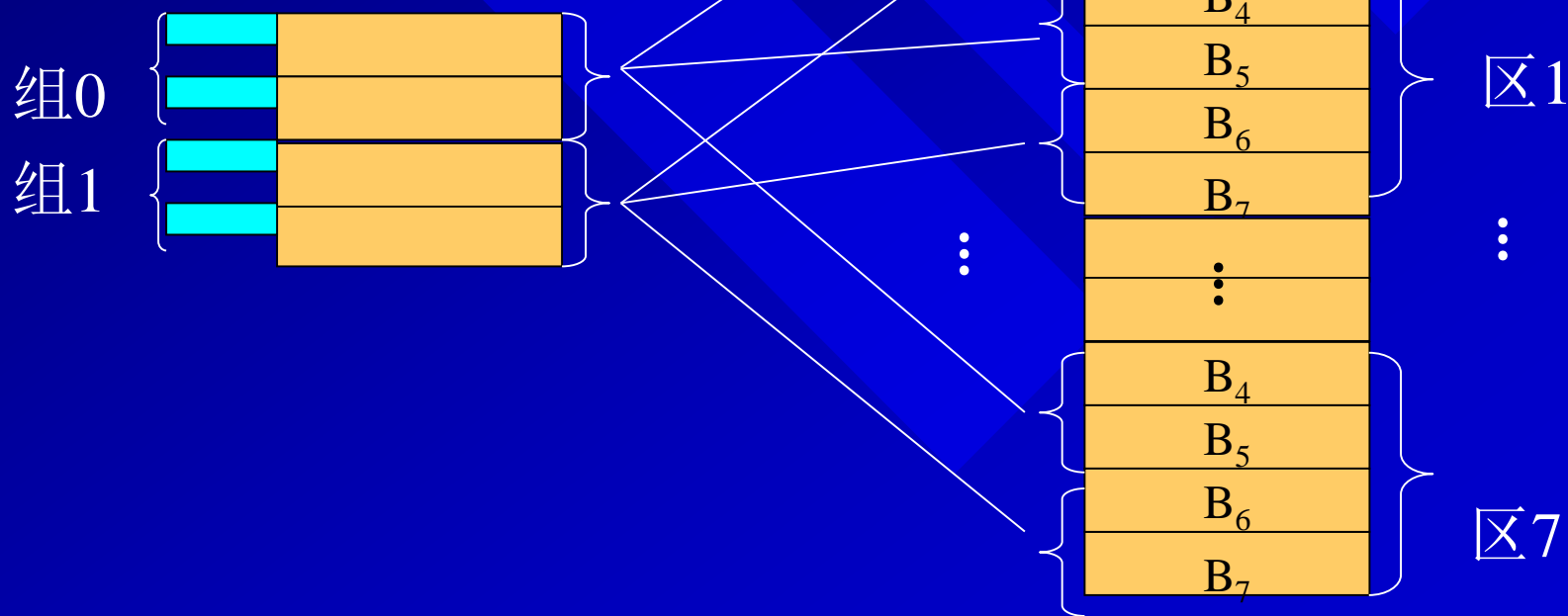
● **题4.2** 对于下述访存字节地址序列：

1,14,50,89,20,17,19,56,19,11,14,43,15,16,9,17
标出每次访存后的cache存储空间的分配情况
和命中情况。假定cache是2路组相联的，
采用FIFO替换策略，每块是4个32位的字。
Cache的容量是16字，初始cache为空。

解:



主存储器



(2路组相联FIFO)

时间 t 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

字地址流

	1	14	50	89	20	17	19	56	19	11	14	43	15	16	9	17
组0 {	1	1	1	1*	20	20*	19	19	19	19	19	19	19	19	19*	17
			50	50	50*	17	17	17	17	17	17	17	17*	16	16	16
组1 {		14	14	14	14	14	14*	56	56	56*	14	14*	15	15	15	15
				89	89	89	89	89	89*	11	11*	43	43	43*	9	9

命中1次 调进 调进 调进 调进 替换 替换 替换 替换 命中 替换 替换 替换 替换 替换 替换 替换

- 假设对指令 Cache 的访问占全部访问的 75%；而对数据 Cache 的访问占全部访问的 25%。Cache 的命中时间为 1 个时钟周期，失效开销为 50 个时钟周期，在混合 Cache 中一次 load 或 store 操作访问 Cache 的命中时间都要增加一个时钟周期，32KB 的指令 Cache 的失效率为 0.39%，32KB 的数据 Cache 的失效率为 4.82%，64KB 的混合 Cache 的失效率为 1.35%。又假设采用写直达策略，且有一个写缓冲器，并且忽略写缓冲器引起的等待。试问指令 Cache 和数据 Cache 容量均为 32KB 的分离 Cache 和容量为 64KB 的混合 Cache 相比，哪种 Cache 的失效率更低？两种情况下平均访存时间各是多少？

解：（1）根据题意，约 75% 的访存为取指令。

因此，分离 Cache 的总体失效率为： $(75\% \times 0.15\%) + (25\% \times 3.77\%) = 1.055\%$ ；

容量为 128KB 的混合 Cache 的失效率略低一些，只有 0.95%。

（2）平均访存时间公式可以分为指令访问和数据访问两部分：

平均访存时间 = 指令所占的百分比 \times （读命中时间 + 读失效率 \times 失效开销）
+ 数据所占的百分比 \times （数据命中时间 + 数据失效率 \times 失效开销）

所以，两种结构的平均访存时间分别为：

分离 Cache 的平均访存时间 = $75\% \times (1 + 0.15\% \times 50) + 25\% \times (1 + 3.77\% \times 50)$

= $(75\% \times 1.075) + (25\% \times 2.885) = 1.5275$

混合 Cache 的平均访存时间 = $75\% \times (1 + 0.95\% \times 50) + 25\% \times (1 + 1 + 0.95\% \times 50)$

= $(75\% \times 1.475) + (25\% \times 2.475) = 1.725$

因此，尽管分离 Cache 的实际失效率比混合 Cache 的高，但其平均访存时间反而较低。

分离 Cache 提供了两个端口，消除了结构相关。

某 RISC 处理机的工作主频为 500MHz，有一个 Cache 和一个主存储器，Cache 的存取周期为 2ns，主存储器的存取周期为 20ns，Cache 的命中率为 99%，有 20% 的 LOAD/STORE 指令，并假设处理机速度的瓶颈完全在存储系统。

- (1) 求理想情况(Cache 的命中率为 100%) 下的 CPI。
- (2) 计算该 RISC 处理机的实际 MIPS 速率。
- (3) 如果处理机的工作主频提高到 1GHz，计算实际 CPI 和 MIPS 速率。

- 解答:
- (1) $CPI = (2ns + 2ns \times 20\%) \times 500MHz = 1.2$
- (2)
 $MIPS = 1 / (2ns \times 99\% + 20ns \times 1\% + 2ns \times 20\% \times 99\% + 20ns \times 20\% \times 1\%)$
 $= 1 / 2.616ns = 382.3$
- (3)
 $CPI = (2ns \times 99\% + 20ns \times 1\% + 2ns \times 20\% \times 99\% + 20ns \times 20\% \times 1\%) \times 1GHz$
 $= 2.616ns \times 1GHz = 2.616$
- $MIPS = 1 / 2.616ns = 382.3$

给定以下的假设，

- (1) 对指令Cache的访问占全部访问的75%，对数据Cache的访问占25%；
- (2) Cache的命中时间为1个时钟周期，失效开销为50个时钟周期；
- (3) 在混合Cache中一次load或store操作访问Cache的命中时间都要增加1个时钟周期；
- (4) 32KB的指令Cache的失效率为0.39%，32KB的数据Cache的失效率为4.82%，64KB的混合Cache的失效率为1.35%；
- (5) 采用写直达策略，且有一个写缓冲器，并且忽略写缓冲器引起的等待。

要求：

- (1) 对于指令Cache和数据Cache容量均为32 KB的分离Cache，计算其失效率，且计算其平均访存时间；
- (2) 计算容量为64 KB的混合Cache的平均访存时间；
- (3) 分别从失效率和平均访存时间两个方面，比较分离Cache与混合Cache的性能

解:

(1) 对于分离Cache的总体失效率为:

$$(75\% \times 0.39\%) + (25\% \times 4.82\%) = 1.4975\%$$

平均访存时间公式可以分为指令访问和数据访问两部分:

平均访存时间 = 指令所占的百分比 \times (指令命中时间 + 指令失效率 \times 失效开销) + 数据所占的百分比 \times (数据命中时间 + 数据失效率 \times 失效开销)

$$\text{分离Cache的平均访存时间} = 75\% \times (1 + 0.39\% \times 50) + 25\% \times (1 + 4.82\% \times 50)$$

$$= 1.74875$$

$$(2) \text{混合Cache的平均访存时间} = 75\% \times (1 + 1.35\% \times 50) + 25\% \times (1 + 1 + 1.35\% \times 50)$$

$$= 1.925$$

(3) 从失效率方面来看, 分离Cache为1.4975%, 其大于混合Cache的失效率1.35%, 混合Cache的性能优于分离Cache。

但从平均访存时间来看, 分离Cache为1.74875个时钟周期, 其小于混合Cache的1.925个时钟周期, 分离Cache的性能优于混合Cache。

因此, 尽管分离Cache的实际失效率比混合Cache的高, 但分离Cache提供了两个端口, 消除了结构冲突, 其平均访存时间反而较低。

- 假设某台计算机的特性及性能为：

- (1) 存储器总线宽度为1个字（32位），送地址需要4个时钟周期，每个字的访问时间为24个时钟周期，传送1个字的数据需4个时钟周期；

- (2) Cache块大小为1个字时，Cache失效率为3%；

- (3) 平均每条指令访存1.2次；

- (4) 在不考虑Cache失效时，平均CPI为2。

- 要求：

- (1) 计算Cache失效开销，存储器的带宽以及在考虑Cache失效时的CPI；

- (2) 若Cache块大小为2个字时，Cache失效率为2%，计算相应的CPI；

- (3) 若Cache块大小为2个字时，讨论在采用2路多体交叉存取以及将存储器和总线宽度增加一倍时，对提高性能（用CPI说明）各有何作用？

解:

(1) 当Cache块大小为一个字时,

Cache失效开销为 $4+24+4=32$ 个时钟周期。

存储器的带宽为每个时钟周期 $4/32=1/8$ 字节。

在考虑Cache失效时的CPI为 $2 + (1.2 \times 3\% \times 32) = 3.15$ 个时钟周期。

(2) 若Cache块大小为2个字时, Cache失效率为2%,

相应的CPI为 $2 + (1.2 \times 2\% \times 2 \times 32) = 3.54$ 个时钟周期。

(3) 若Cache块大小为2个字且采用2路多体交叉存取时,

相应的CPI为 $2 + 1.2 \times 2\% \times (4 + 24 + 8) = 2.86$ 个时钟周期。

性能相对于不采用2路多体交叉存取, 提高了 $(3.54 - 2.86) / 3.54 = 19.21\%$ 。

若Cache块大小为2个字且采用64位总线和存储器, 不采用多体交叉时,

相应的CPI为 $2 + 1.2 \times 2\% \times 1 \times 32 = 2.77$ 个时钟周期。

性能相对于不采用存储器和总线宽度增加一倍, 提高了 $(3.54 - 2.77) / 3.54 = 21.75\%$ 。

第5章

作5.2 设一条指令的执行过程分为“取指令”、“分析”和“执行”三段，每一段的执行时间分别为 Δt 、 $2\Delta t$ 和 $3\Delta t$ 。在下列各种情况下，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
- (4) 先行控制方式

- 解:

- (1) 顺序执行需要的时间如下:

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

- (2) 取指令和执行重叠, 即一次重叠执行方式, 我们假设第 $n+1$ 条指令的取指令和第 n 条指令的执行同时结束, 那么所需要的时间为:

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

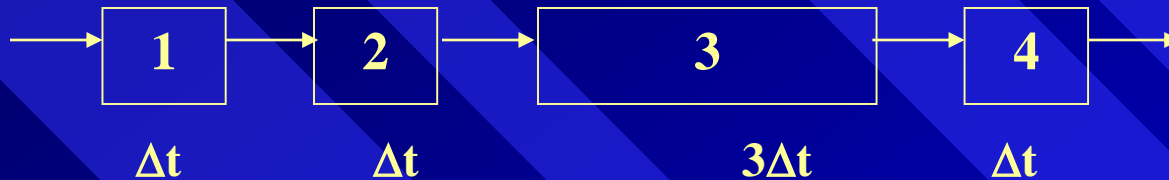
- (3) 取指令、分析和执行重叠

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

- (4) 先行控制方式

$$T = t_{\text{分析1}} + \sum_{i=1}^n t_{\text{执行}} = \Delta t + 3n\Delta t$$

例5.1 带有瓶颈部件的4功能段流水线, $\Delta t_1 = \Delta t_2 = \Delta t_4 = \Delta t$, $\Delta t_3 = 3\Delta t$, 4个任务、10个任务时TP, E、SP。

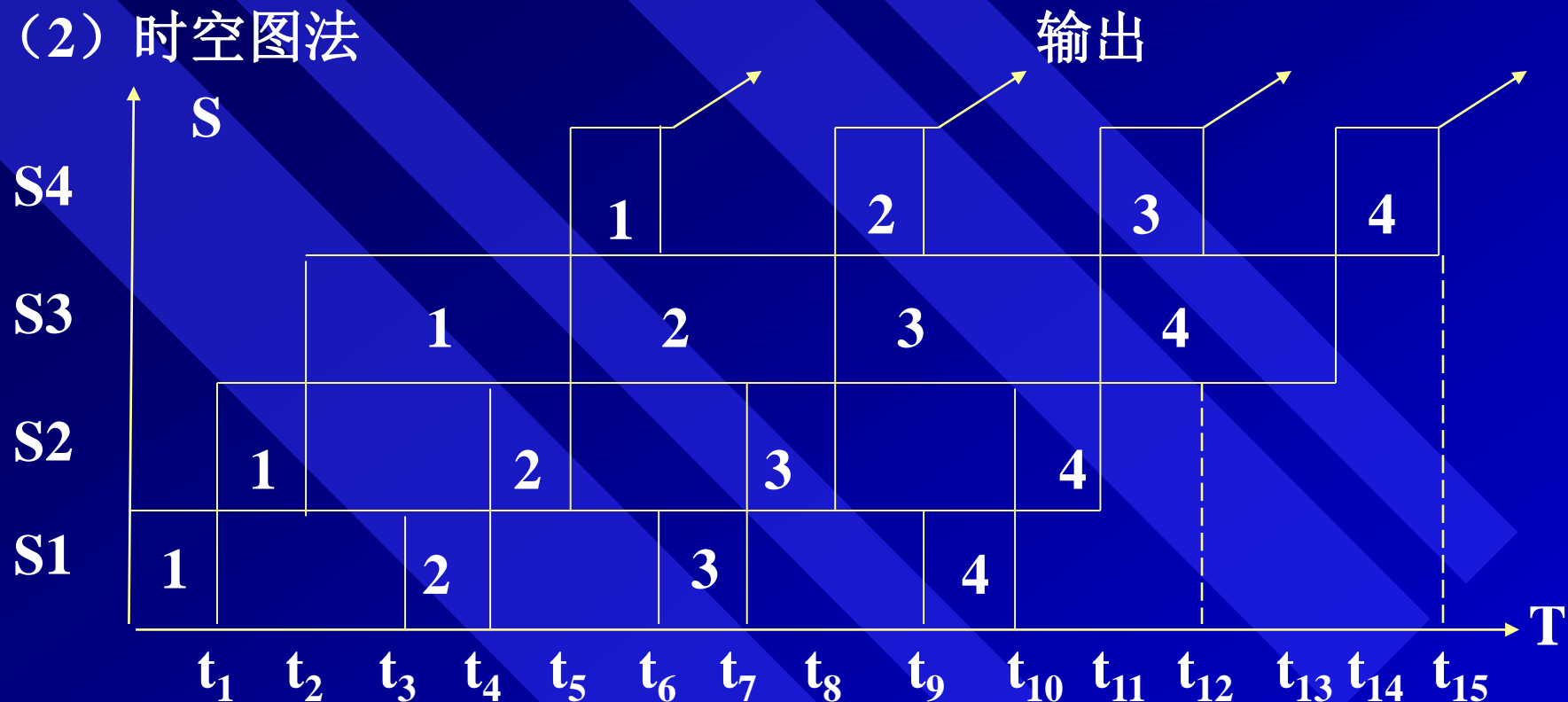


(1) 分析法: 各段时间不等

$$TP = \frac{n}{\sum_{i=1}^m \Delta t_i + (n-1) \Delta t_j}$$

$$n = 4 \text{ 时}, TP = \frac{4}{(6+9)\Delta t} = \frac{4}{15\Delta t}$$

(2) 时空图法



$$E = \frac{\text{n个任务实际占用的时-空区}}{\text{M各段总的时-空区}}$$

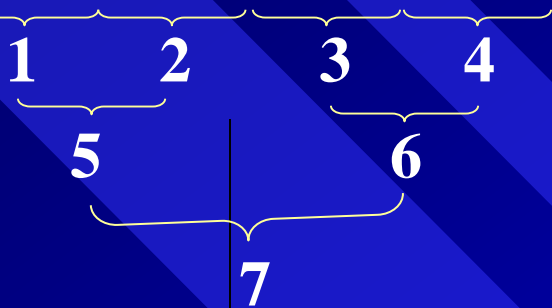
$$= \frac{4 \times 6 \Delta t}{4 \times 15 \Delta t} = \frac{6}{15} = 0.4 = 40\%$$

$$Sp = \frac{n * \sum_{i=1}^m \Delta t_i}{\sum_{i=1}^m \Delta t_i + (n-1) * \Delta t_j} = \frac{4 * 6 \Delta t}{15 \Delta t} = \frac{24}{15} = 1.6$$

例5.2 以浮点加法运算为例（四段流水线）各段时间相等，求吞吐率、效率。

求 $Z=A+B+C+D+E+F+G+H$ ，TP、E、Sp（注意有相关）

$$Z=A+B+C+D+E+F+G+H$$

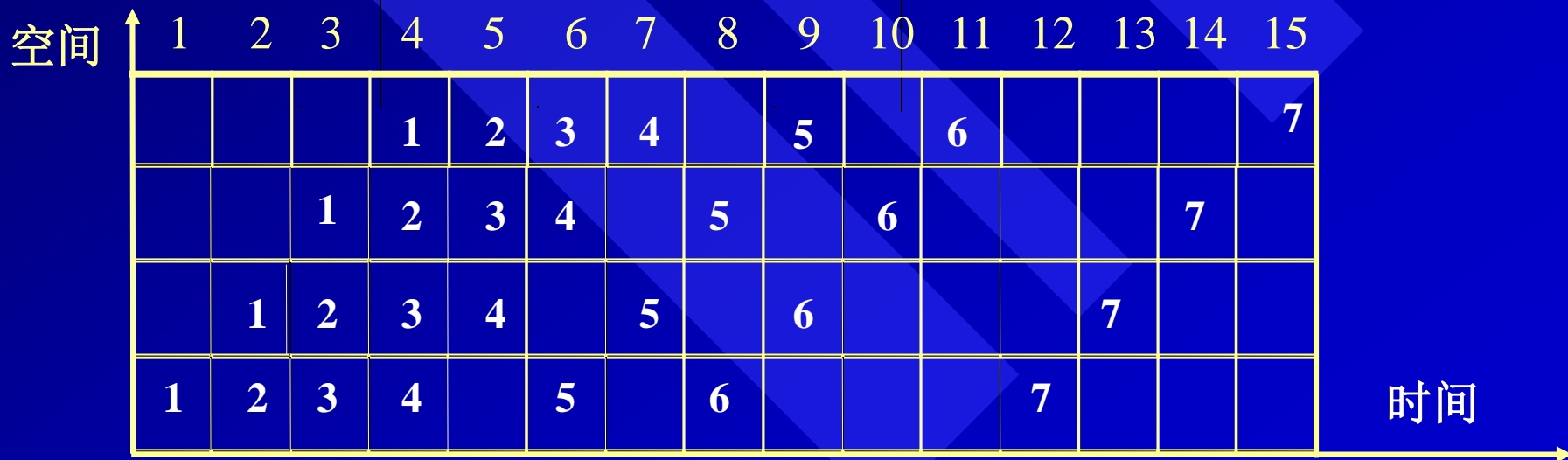


解：

$$TP=7/15\Delta t$$

$$E=7*4/(15*4)=7/15=46\%$$

$$Sp=4*7/15=28/15=1.87$$



流水线的效率不高，原因在于存在着数据相关，有空闲功能段。

例5.3 ASC计算机多功能算术运算流水线各段时间相等，6次浮点加、5次定点乘的吞吐率，效率，加速比 **m=8, n=11**



分析: $T_{\text{加}} = 6 + (6-1) \times 1 = 11(\Delta t)$ $T_{\text{乘}} = 4 + (5-1) \times 1 = 8(\Delta t)$

时间

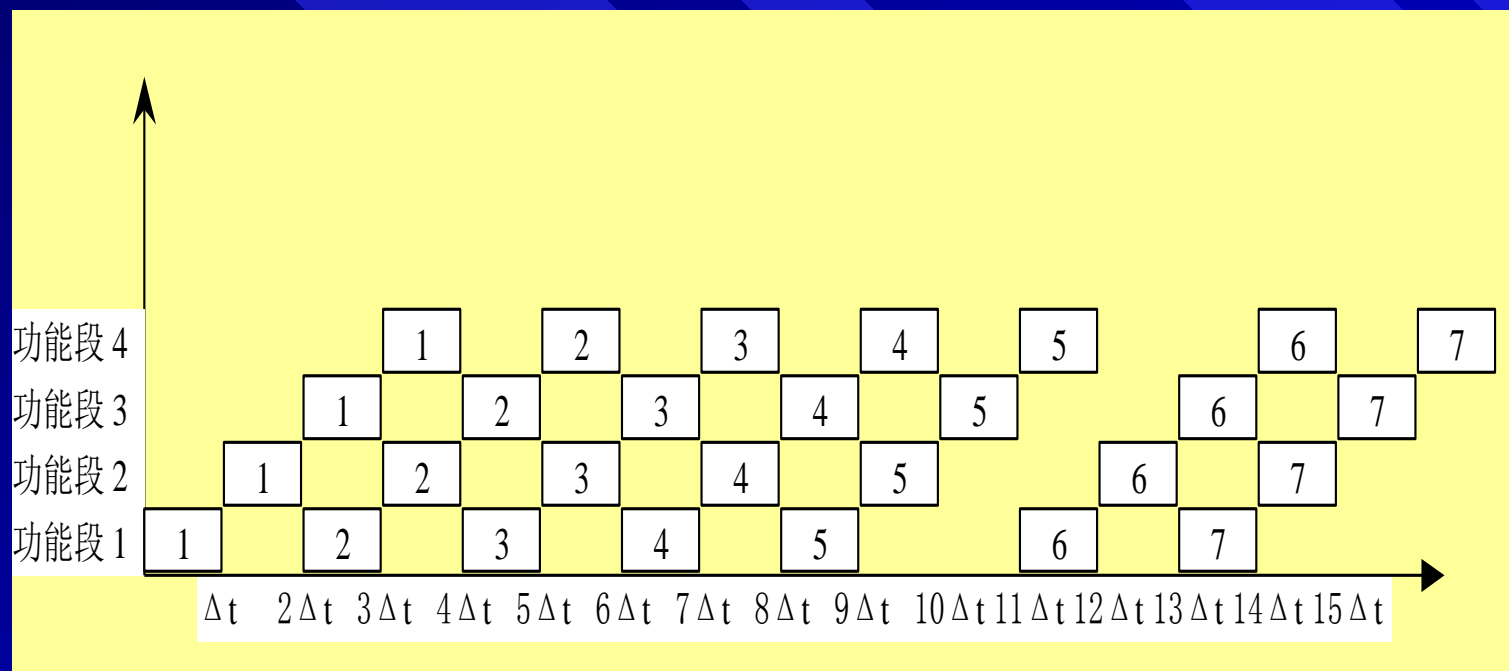
则 $TP = 11 / (11 + 8) \Delta t = 11 / 19 \Delta t$

$E = (6 \times 6 + 5 \times 4) \Delta t / (19 \times 8 \Delta t) = 36.8\%$

$Sp = (6 \times 6 + 5 \times 4) \Delta t / 19 \Delta t = 56 / 19 = 2.94$

题5.1 一条线性流水线有4个功能段组成，每个功能段的延迟时间都相等，都为 Δt 。开始5个 Δt ，每间隔一个 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。

[解答]流水线的时空图如下：



我们可以看出，在 $(11n+1) \Delta t$ 的时间内，可以输出 $5n$ 个结果，如果指令的序列足够长 ($n \rightarrow \infty$)，并且指令间不存在相关，那么，吞吐率可以认为满足：

$$Tp = \frac{5n}{(11n+1)\Delta t} = \frac{5}{(11+1/n)\Delta t} = \frac{5}{11\Delta t} (n \rightarrow \infty)$$

加速比为：

$$S = \frac{5n \times 4\Delta t}{(11n+1)\Delta t} = \frac{20n}{11n+1} = \frac{20}{11+1/n} = \frac{20}{11} (n \rightarrow \infty)$$

从上面的时空图很容易看出，效率为：

$$E = \frac{20n\Delta t}{4 \times (11n+1)\Delta t} = \frac{5}{11+1/n} = \frac{5}{11} (n \rightarrow \infty)$$

例5.4 用一条5个功能段的浮点加法器流水线计算

$$F = \sum_{i=1}^{10} A_i$$

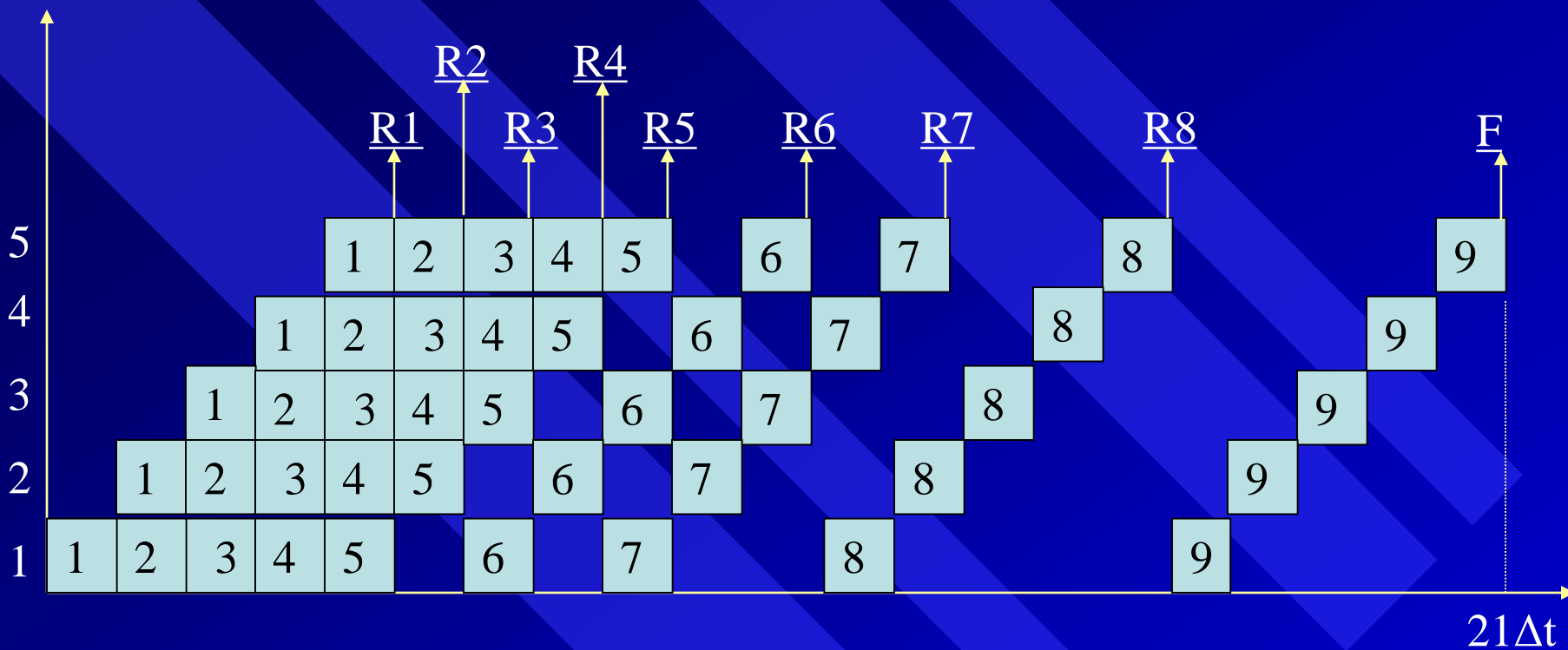
每个功能段的延迟时间均相等，流水线的输出端和输入端之间有直接数据通路，而且设置有足够的缓冲寄存器。要求用尽可能短的时间完成计算，画出流水线时空图，并计算流水线的实际吞吐率、加速比和效率。

[解答]首先需要考虑的是，10个数的和最少需要做几次加法。我们可以发现，加法的次数是不能减少的：9次；于是我们要尽可能快的完成任务，就只有考虑如何让流水线尽可能充满，这需要消除前后指令之间的相关。由于加法满足交换率和结合率，我们可以调整运算次序如以下的指令序列，我们把中间结果寄存器称为R，源操作数寄存器称为A，最后结果寄存器称为F，并假设源操作数已经在寄存器中，则指令如下：

- I1: $R1 \leftarrow A1 + A2$
- I2: $R2 \leftarrow A3 + A4$
- I3: $R3 \leftarrow A5 + A6$
- I4: $R4 \leftarrow A7 + A8$
- I5: $R5 \leftarrow A9 + A10$
- I6: $R6 \leftarrow R1 + R2$
- I7: $R7 \leftarrow R3 + R4$
- I8: $R8 \leftarrow R5 + R6$
- I9: $F \leftarrow R7 + R8$

这并不是唯一可能的计算方法。假设功能段的延迟为 Δt 。时空图如下，图中的数字是指令号。

部件m



$$R1=A1+A2$$

$$R5=A9+A10$$

$$R2=A3+A4$$

$$R6=R1+R2$$

$$R8=R5+R6$$

$$R3=A5+A6$$

$$R7=R3+R4$$

$$R4=A7+A8$$

$$R7=R3+R4$$

$$F=R7+R8$$

整个计算过程需要 $21 \Delta t$ ，所以吞吐率为：

$$Tp = \frac{9}{21\Delta t}$$

加速比为：

$$S = \frac{9 \times 5\Delta t}{21\Delta t} = \frac{45}{21} = 2.14$$

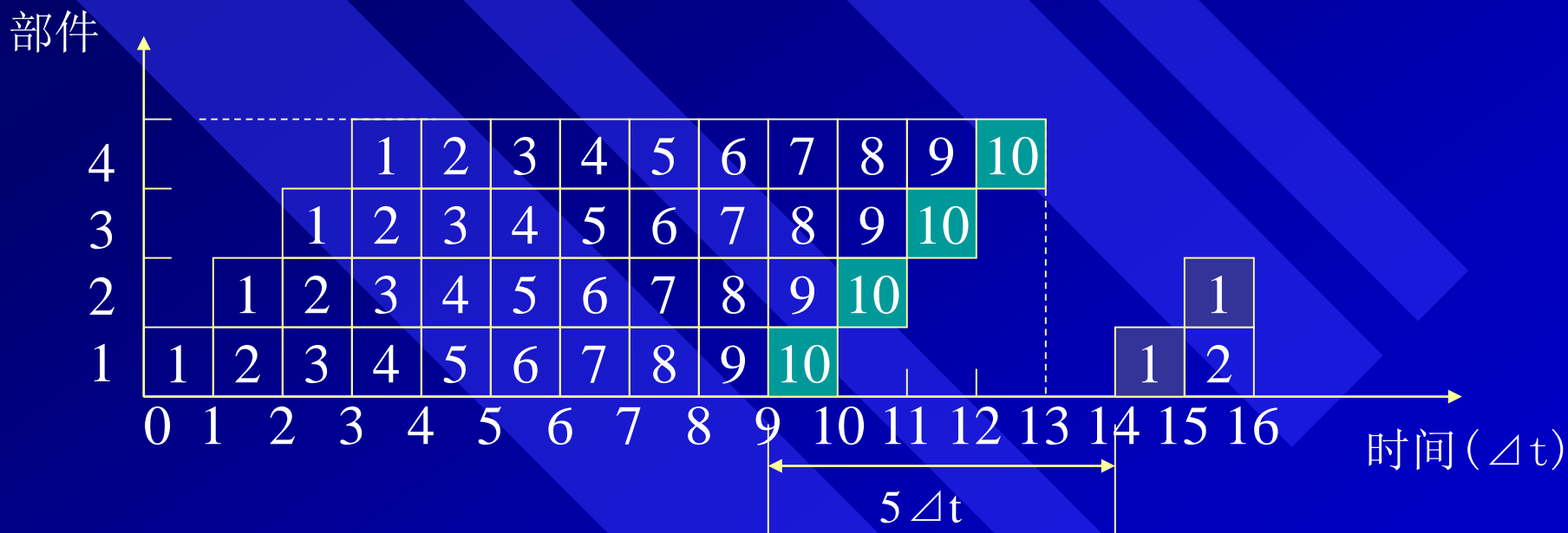
效率为：

$$E = \frac{9 \times 5\Delta t}{5 \times 21\Delta t} \approx 42.86\%$$

作5.5 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，并画出其时空图。

[分析] 所谓输入10个数据后，间歇 $5\Delta t$ ，又输入10个数据的含义应当是以输入时间为基准，即从第10个数据输入时算起，隔 $5\Delta t$ 后又开始输入新一轮数据。

[解答]按题意可得4个功能部件流水时的时空关系如下图所示

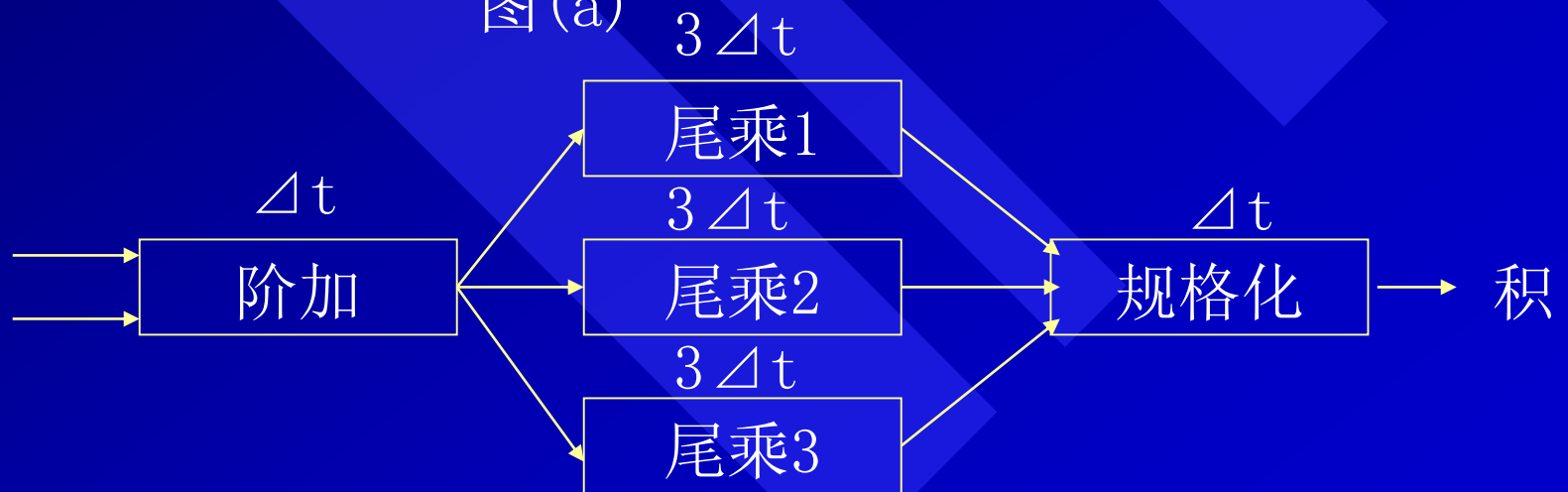


所以，按周期性工作时的流水线平均吞吐率为
 $T_p = 10 / (14 \Delta t) = 5 / (7 \Delta t)$

作5.6 有一个浮点乘流水线如下图(a)所示, 其乘积可直接返回输入端或暂存于相应缓冲寄存器中, 画出实现 $A*B*C*D$ 的时空图以及输入端的变化, 并求出该流水线的吞吐率和效率; 当流水线改为下图(b)形式实现同一计算时, 求该流水线的效率及吞吐率。



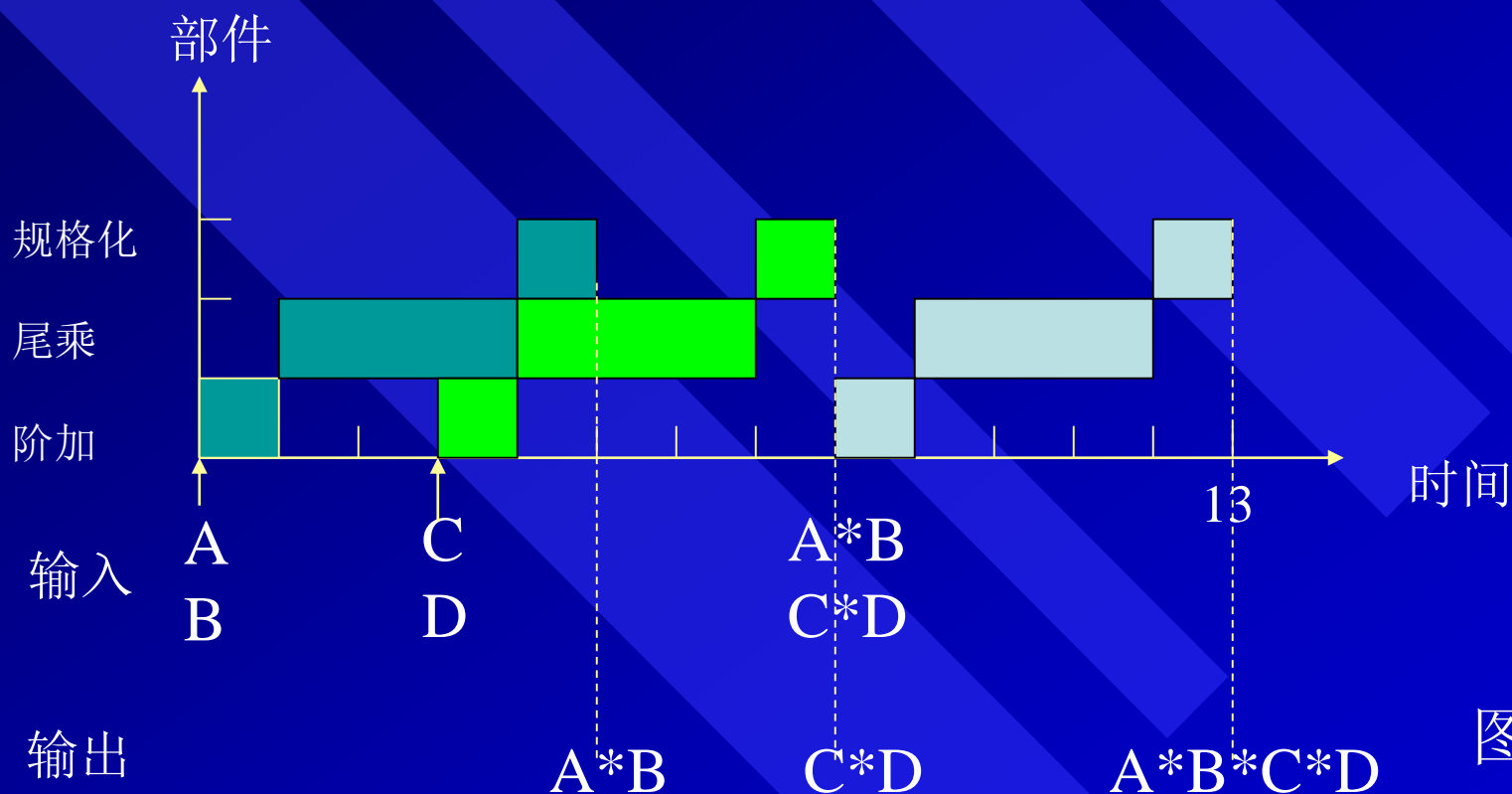
图(a)



图(b)

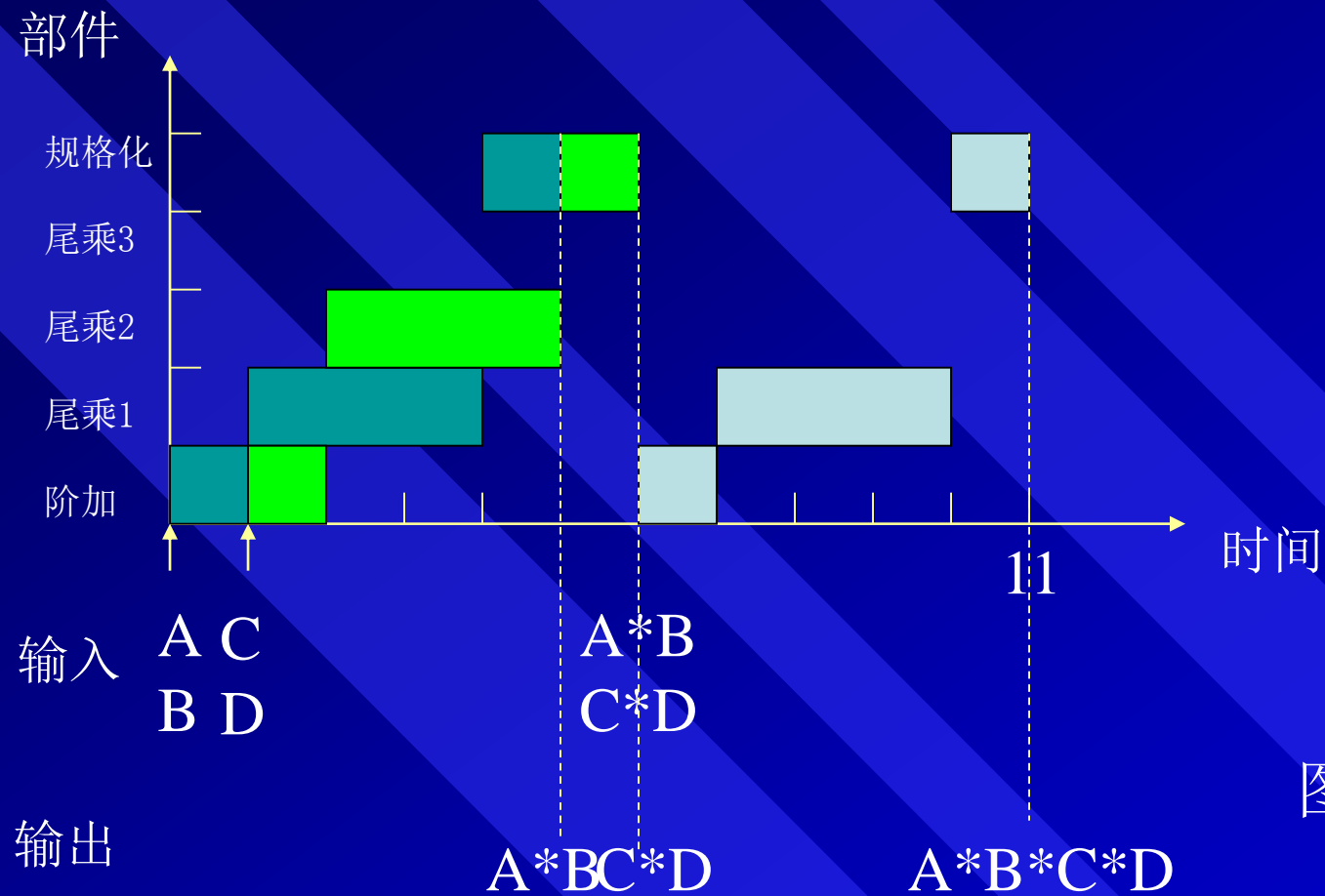
[分析]为了减少运算过程中的操作数相关, $A*B*C*D$ 应改为采用 $((A*B)*(C*D))$ 的算法步骤进行运算。

[解]按图(a)组织, 实现 $A*B*C*D$ 的时空关系如下图(A)所示。



$$\text{吞吐率} TP = 3 / (13 \Delta t)$$

$$\text{效率} E = (3 \times 5 \Delta t) / (3 \times 13 \Delta t) = 5 / 13 = 38.5\%$$



图(B)

流水线按图(b)组织时，实现 $A*B*C*D$ 的时空关系如图(B)

吞吐率 $TP=3/(11 \Delta t)$

效率 $E = (3 \times 5 \Delta t) / (5 \times 11 \Delta t) = 3/11 = 27.3\%$

例5.5（类似题5.8）一条线性静态多功能流水线由6个功能段组成，加法操作使用其中的1、2、3、6功能段，乘法操作使用其中的1、4、5、6功能段，每个功能段的延迟时间均相等。流水线的输入端与输出端之间有直接数据通路，而且设置有足够的缓冲寄存器。现在用这条流水线计算：

$$F = \sum_{i=1}^6 (A_i \times B_i)$$

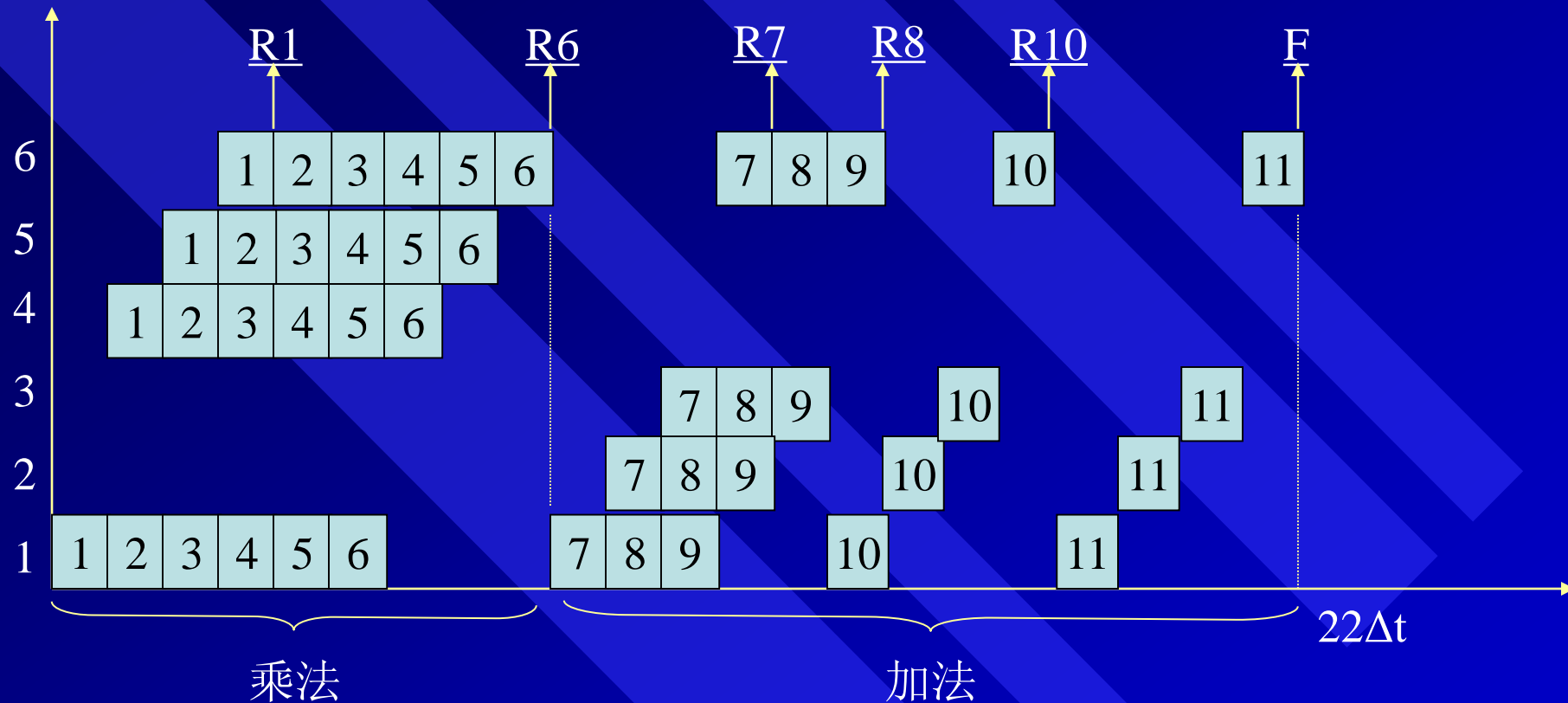
画出流水线时空图，并计算流水线的实际吞吐率、加速比和效率。

解：为了取得较高的速度，我们需要一次将乘法作完，设源操作数存放在寄存器A、B中，中间结果存放在寄存器R中，最后结果存放在寄存器F中，则执行的指令序列如下所示：

```
I1:      R1←A1*B1
I2:      R2←A2*B2
I3:      R3←A3*B3
I4:      R4←A4*B4
I5:      R5←A5*B5
I6:      R6←A6*B6
I7:      R7←R1+R2
I8:      R8←R3+R4
I9:      R9←R5+R6
I10:     R10←R7+R8
I11:     F←R9+R10
```

这并不是唯一可能的计算方法。假设功能段的延迟为 Δt 。时空图（不完全）如下，图中的数字是指令号。

部件m



$$R1=A1*B1$$

$$R2=A2*B2$$

$$R3=A3*B3$$

$$R4=A4*B4$$

$$R5=A5*B5$$

$$R6=A6*B6$$

$$R7=R1+R2$$

$$R8=R3+R4$$

$$R9=R5+R6$$

$$R10=R7+R8$$

$$F=R9+R10$$

作5.4 在一台单流水线多操作部件的处理机上执行下面的程序，取指令、指令译码各需要1个时钟周期，MOVE, ADDT和MUL操作各需要2个、3个和4个时钟周期。每个操作都在第一个时钟周期从寄存器中读取操作数，在最后1个时钟周期把运算结果写到通用寄存器中。

K: MOVE R1,R0 ; $R1 \leftarrow R0$

K+1: MUL R0 , R2 ,R1 ; $R0 \leftarrow (R1) \times (R2)$

K+2: ADD R0,R2 , R3 ; $R0 \leftarrow (R2) + (R3)$

- (1) 就程序本身而言，可能有哪几种相关？
- (2) 在程序实际执行过程中，有哪几种相关会引起流水线的停顿？
- (3) 画出指令执行过程的流水线时空图，并计算机执行完这3条指令共使用了多少个时钟周期？

解 (1)

K: **MOVE R1,R0 ; R1←R0**

K+1: **MUL R0 , R2 ,R1 ; R0← (R1)×(R2)**

K+2: **ADD R0,R2 , R3 ; R0← (R2) +(R3)**

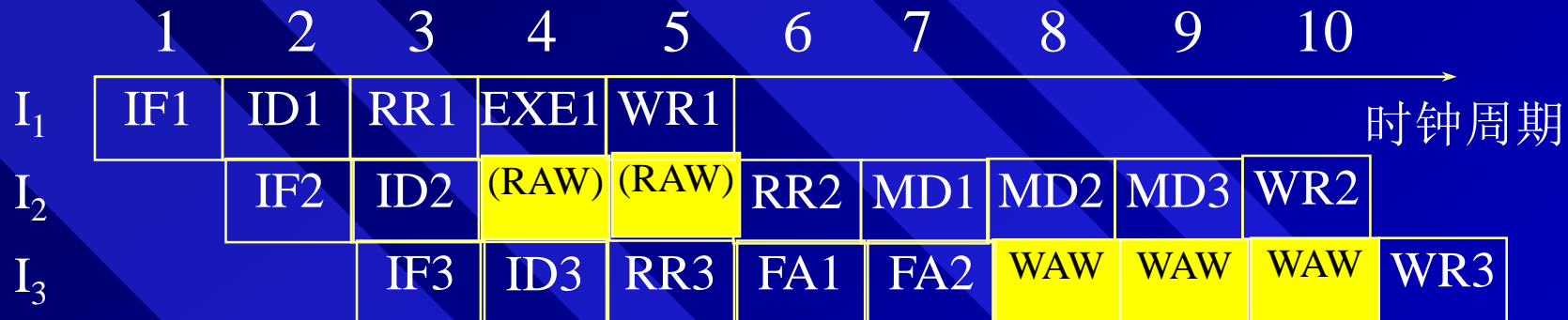
K、K+1 存在写读相关，读写相关；

K+1、K+2存在写写相关。

(2) K、K+1 的写读相关，会引起流水线的停顿；

K+1、K+2的写写相关会引起流水线的停顿。

(3)



共使用11个
时钟周期

K: MOVE R1,R0 ; R1←R0

K+1: MUL R0 , R2 ,R1 ; R0← (R1)×(R2)

K+2: ADD R0,R2 , R3 ; R0← (R2) +(R3)

作5.13 一台非流水处理机A的工作时钟频率为25MHz，它的平均CPI为4。处理器B是A的改进型，它有1条5段的线性指令流水线。由于锁定电路延迟及时钟扭斜效应，它的工作时钟频率仅为20MHz，问：

(1) 若在A和B两个处理器上执行100条指令的程序，则处理器B对A的加速比是多少？

(2) 在执行上述程时，计算A，B处理器各自的MIPS速率是多少？

解：已知： $MIPS_B=20$

$$MIPS_A = \frac{R_C}{CPI \times 10^6} = \frac{25M}{4M} = 6.25$$

$$S_P = \frac{MIPS_B}{MIPS_A} = \frac{20}{6.25} = 3.2$$

题5.2 某超标量流水线计算机每个时钟发射两条指令，其4级指令流水线中包含IF、ID、EXE、WB流水级，各个流水段都是花费一个时钟周期。ALU指令在EXE流水段完成算术运算，乘除指令比ALU指令多花费2个时钟周期，访存指令在EXE功能段完成访存操作。流水中无相关专用通路。指出下列指令序列中的数据相关性，分别画出指令流水线在有序执行有序写回、无序执行无序写回情况下的时空图。

1: LOAD R0 , A ; $R0 \leftarrow \text{主存}(A)\text{单元}$

2: ADD R1 , R0 ; $R1 \leftarrow (R1) + (R0)$

3: LOAD R2 , B ; $R2 \leftarrow \text{主存}(B)\text{单元}$

4: MUL R3 , R4 ; $R3 \leftarrow (R3) \times (R4)$

5: AND R4 , R5 ; $R4 \leftarrow (R4) \wedge (R5)$

6: ADD R2 , R5 ; $R2 \leftarrow (R2) + (R5)$

解（1）相关性分析:1、2 存在写读相关；4、5存在读写相关；3、6存在写读相关；3、6存在写写相关。

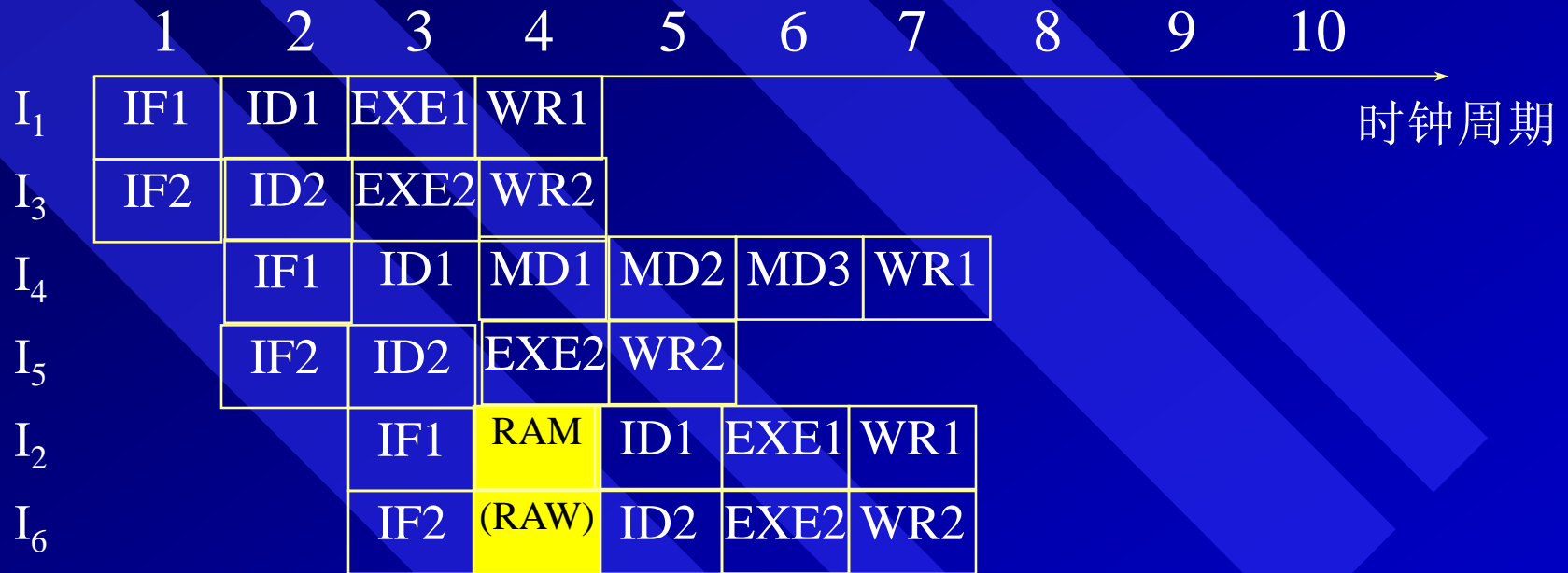
(2) 相关性分析: 按序发射按序完成之调度

	1	2	3	4	5	6	7	8	9	10	
I_1	IF1	ID1	EXE1	WR1							时钟周期
I_2	IF2	(RAW)	(RAW)	(RAW)	ID2	EXE2	WR2				
I_3		IF1	ID1	EXE1	顺序	顺序	顺序	WR1			
I_4		IF2	ID2	MD1	MD2	MD3	顺序	顺序	WR2		
I_5			IF1	ID1	EXE1	顺序	顺序	顺序	顺序	WR1	
I_6			IF2	(RAW)	(RAW)	(RAW)	(RAW)	(RAW)	ID2	EXE2	WR2

主共使用11
个时钟周期

- 1: **LOAD R0 , A ; $R0 \leftarrow \text{主存(A)单元}$ (1,2RAW)**
- 2: **ADD R1 , R0 ; $R1 \leftarrow (R1) + (R0)$**
- 3: **LOAD R2 , B ; $R2 \leftarrow \text{主存(B)单元}$ (3,6RAW,WW)**
- 4: **MUL R3 , R4 ; $R3 \leftarrow (R3) \times (R4)$ (4,5WAR)**
- 5: **AND R4 , R5 ; $R4 \leftarrow (R4) \wedge (R5)$**
- 6: **ADD R2 , R5 ; $R2 \leftarrow (R2) + (R5)$**

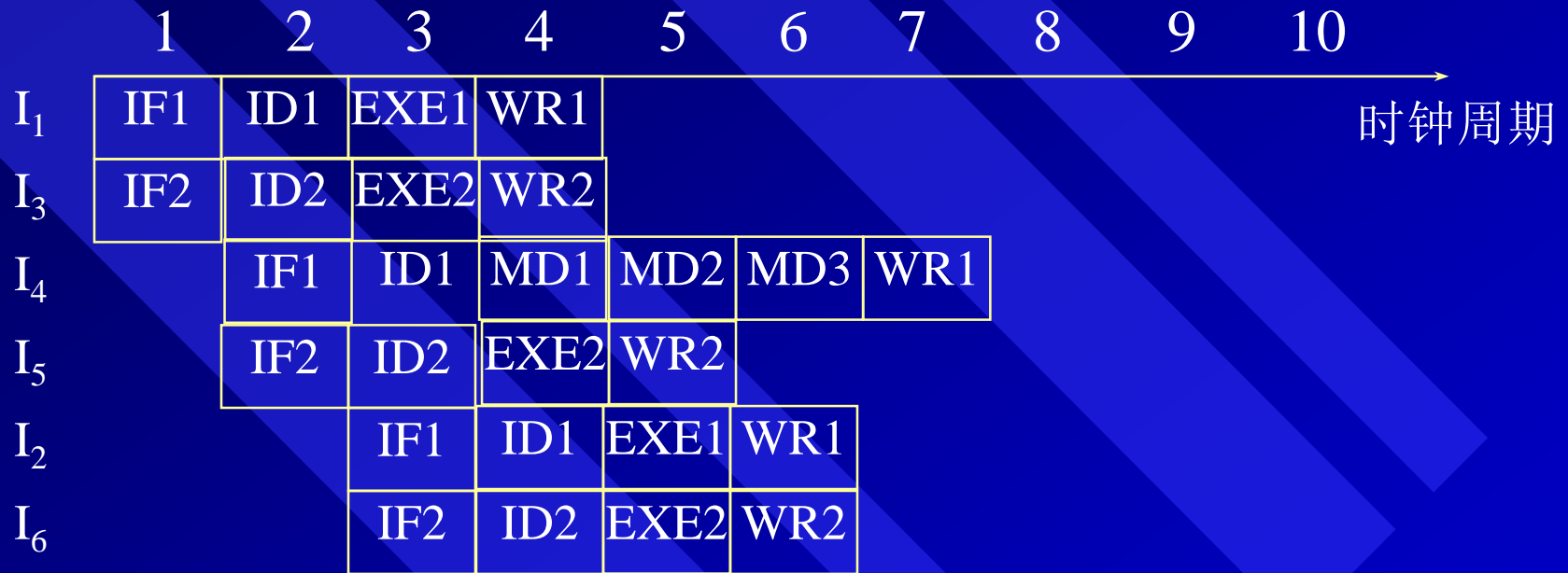
(3) 相关性分析: 无序发射无序完成之调度



主共使用7
个时钟周期

- 1: **LOAD R0 , A ; $R0 \leftarrow \text{主存(A)单元}$** (1,2RAW)
- 2: **ADD R1 , R0 ; $R1 \leftarrow (R1) + (R0)$**
- 3: **LOAD R2 , B ; $R2 \leftarrow \text{主存(B)单元}$** (3,6RAW),WW
- 4: **MUL R3 , R4 ; $R3 \leftarrow (R3) \times (R4)$** (4,5WAR)
- 5: **AND R4 , R5 ; $R4 \leftarrow (R4) \wedge (R5)$**
- 6: **ADD R2 , R5 ; $R2 \leftarrow (R2) + (R5)$**

(4) 相关性分析: 无序发射无序完成之调度 (设置专用数据通路)



主共使用7
个时钟周期

- 1: **LOAD R0 , A ; R0←主存(A)单元** (1,2RAW)
- 2: **ADD R1 , R0 ; R1← (R1) +(R0)**
- 3: **LOAD R2 , B ; R2←主存(B)单元** (3,6RAW),WW
- 4: **MUL R3 , R4 ; R3← (R3) ×(R4)** (4,5WAR)
- 5: **AND R4 , R5 ; R4← (R4) ∧ (R5)**
- 6: **ADD R2 , R5 ; R2← (R2) +(R5)**

作5.17 在**CRAY-1**机上，设向量长度均为**64**，所有浮点功能部件的执行时间分别为：相加需**6**拍，相乘需**7**拍，求倒数近似值需**14**拍，从存储器读数据需**6**拍，打入寄存器机启动功能部件各需**1**拍，问下列个指令组，组内的哪些指令可以链接？，哪些指令不可以链接？不能链接的原因是什么？并分别计算各指令组全部完成所需的拍数。

(1) $V_0 \leftarrow \text{存储器}$
 $V_1 \leftarrow V_2 + V_3$
 $V_4 \leftarrow V_5 \times V_6$

(2) $V_2 \leftarrow V_0 \times V_1$
 $V_3 \leftarrow \text{存储器}$
 $V_4 \leftarrow V_2 + V_3$

(3) $V_0 \leftarrow \text{存储器}$
 $V_2 \leftarrow V_0 \times V_1$
 $V_3 \leftarrow V_2 + V_0$
 $V_5 \leftarrow V_3 + V_4$

(4) $V_2 \leftarrow \text{存储器}$
 $V_1 \leftarrow 1/V_0$
 $V_3 \leftarrow V_1 \times V_2$
 $V_5 \leftarrow V_3 + V_4$

解:

组(1) 三条指令可并行执行。

$$T=1+7+1+64-1=72(\text{拍})。$$

组(2) 前二条指令可并行执行, 前两条与第三条指令可链接执行。

$$T=(1+7+1+1+6+1)+63=80(\text{拍})。$$

组(3) 前3条指令可链接执行, 后一条指令只能串行(加法部件冲突)

$$T=(8+9+8+63)+8+63=159(\text{拍})。$$

组(4) 前二条指令可并行执行, 后两条可链接, 这样前两条与后两条指令可链接执行。

$$T=(14)+(9)+(8)+63=94(\text{拍})。$$

题5.3 对于以下计算表达式:

- $I = A + B + C + D \times E \times F + G + H$
- 其中A,B...,I 都是寄存器名。
- (1) 指出其中的并行性, 即哪些操作可以并行执行;
- (2) 对于采用5级指令流水线的计算机, 试安排完成表达式计算的操作顺序, 并写出指令序列, 使得流水线的停顿时间最少;
- (3) 对于一台VLIW计算机, 假定其有2个加法部件和一个乘法部件, 都能在一个流水周期中完成计算, 试将上述表达式用最少的超长指令序列表示。

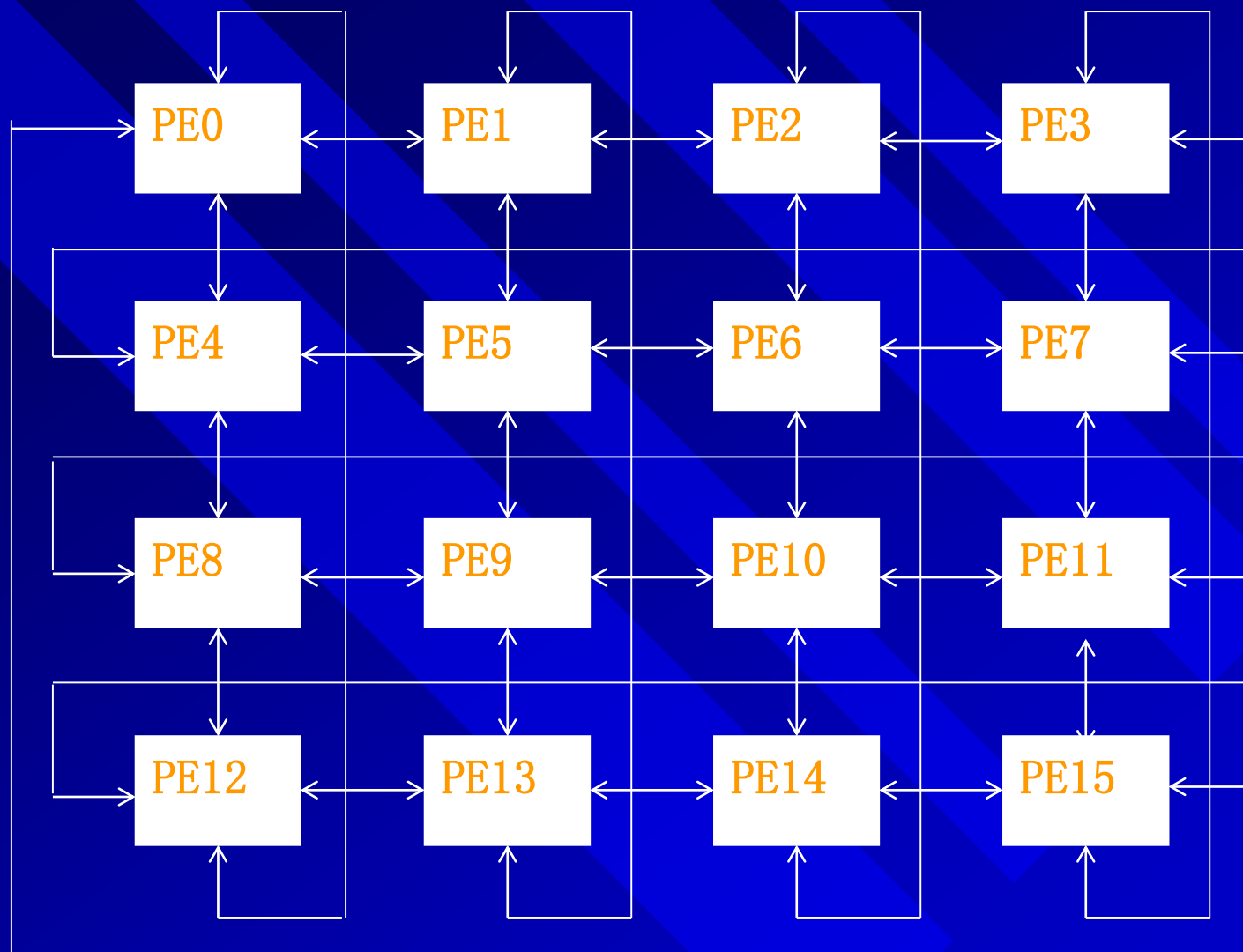
解 (1)

$$I = A + B + C + D \times E \times F + G + H$$

- 可并行操作的运算为:
- $A + B + C + G + H$
- $D \times E \times F$
- (3)
- ① $A + B, G + H, D \times E$
- ② $A + B + C$, $D \times E \times F$
- ③ $A + B + C + G + H$
- ④ $A + B + C + G + H$ + $D \times E \times F$

第6章

- 作6.1 画出16台处理器仿ILLIAC IV的模式进行互连的互连结构图，列出PE0分别只经一步、二步和三步传送能将信息传送到的各个处理器号。
- 解：16台处理机仿ILLIAC IV的模式互连结构图如下页图所示。由图可知：
- PE0经一步可将信息传送到的处理器号有PE1、PE4、PE12、PE15。
- PE0经二步可将信息传送到的处理器号有PE2、PE3、PE5、PE8、PE11、PE13、PE14。
- PE0经三步可将信息传送到的处理器号有PE6、PE7、PE9、PE10。
- 可见，最多经3步就可以将PE₀的信息传送到任何其它的处理器上，所以其最大的距离为3步。



16台处理器仿ILLIAC IV机的互连机构

- **例6.1** 编号为0、1、 \dots 14、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)时，第13号处理器各连至哪一个处理器？
- [分析]：由于是16个处理器组成的单级互连网络，每个处理器的二进制地址编号为4位。根据互连函数关系式，即可求出与第13号处理器相连的处理器号。

- 解:

- (1) $Cube_3(x_3x_2x_1x_0) = \overline{x_3}x_2x_1x_0$ 即 $Cube(1101) = 0101$

第13号处理器连至5号处理器;

- (2) $PM2_{+3}(13) = (13 + 2^3) \bmod 16 = 5$

即第13号处理器连至第5号处理器;

- (3) $PM2_{-0}(13) = (13 - 2^0) \bmod 16 = 12$

即第13号处理器连至第12号处理器;

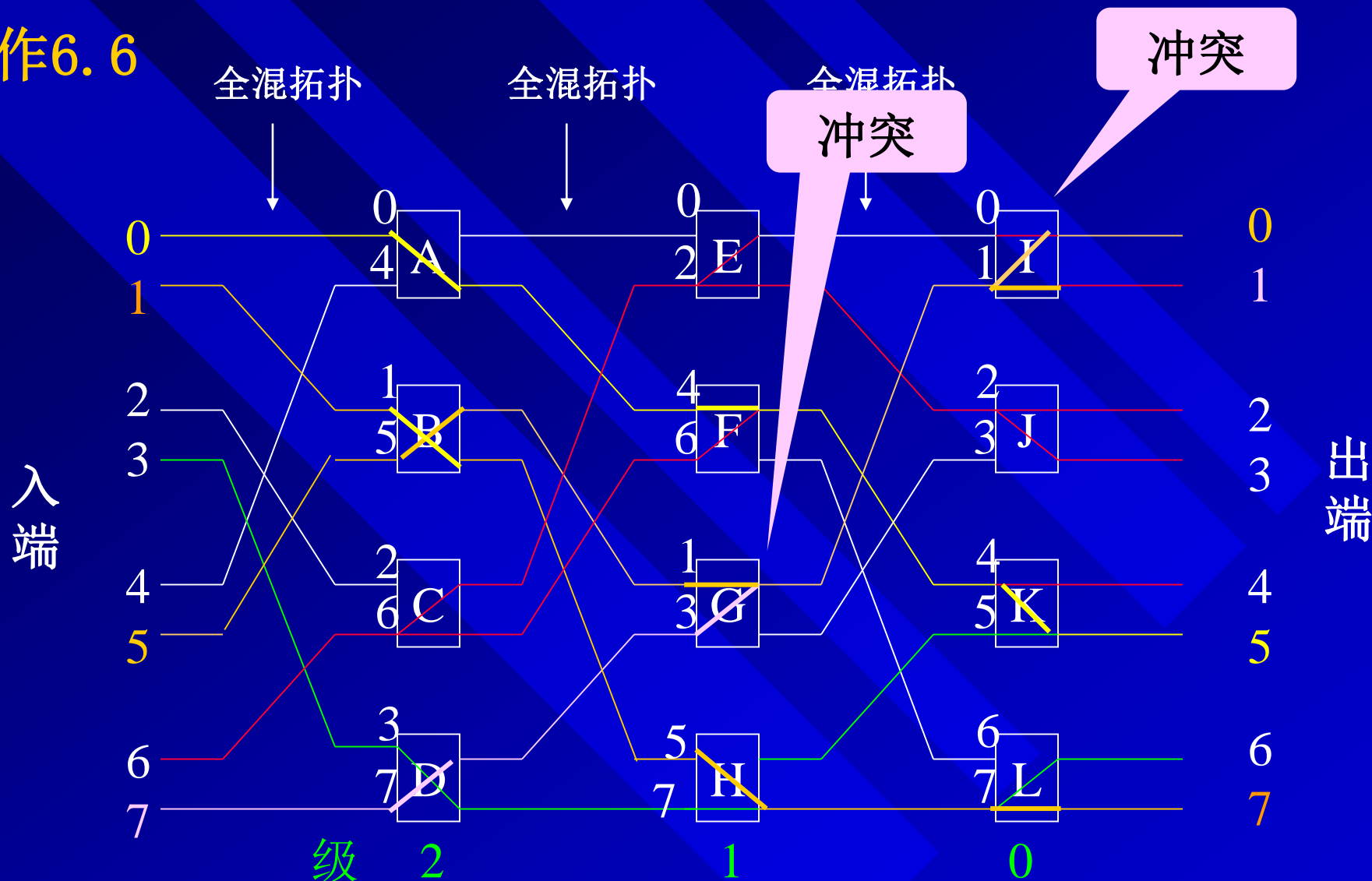
- (4) $Shuffle(x_3x_2x_1x_0) = x_2x_1x_0x_3$ 即 $Shuffle(1101) = 1011$

- 第13号处理器与第11号处理器相连;

- (5) $Shuffle(Shuffle(x_3x_2x_1x_0)) = Shuffle(x_2x_1x_0x_3) = x_1x_0x_3x_2$

- 第13(1101)号处理器连至第7(0111)号处理器。

作6.6



5→0与7→1冲突

0→5与1→7不冲突

N=8多级混洗交换网络

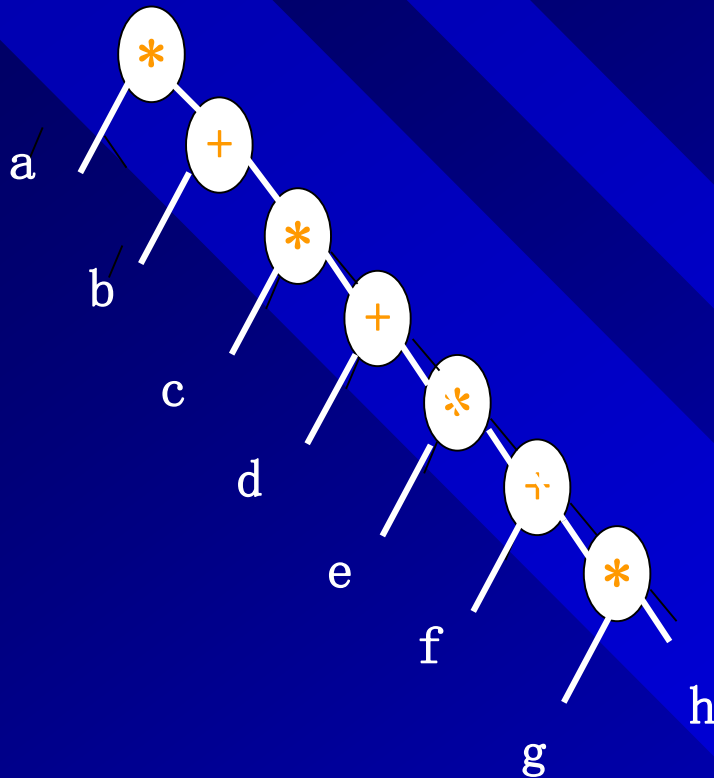
- **作6.18** 由霍纳（Horner）法则给定的表达式如下： $E=a(b+c(d+e(f+gh)))$ ，利用减少树高的办法来加速运算，要求：
 - （1）画出树形流程图；
 - （2）确定 T_p 、 P 、 S_p 、 E_p 等诸值。
- 解：（1）利用交换律、结合律、分配律，将原始表达式转换成如下表达式：

$$E=a(b+cd)+ace(f+gh)$$

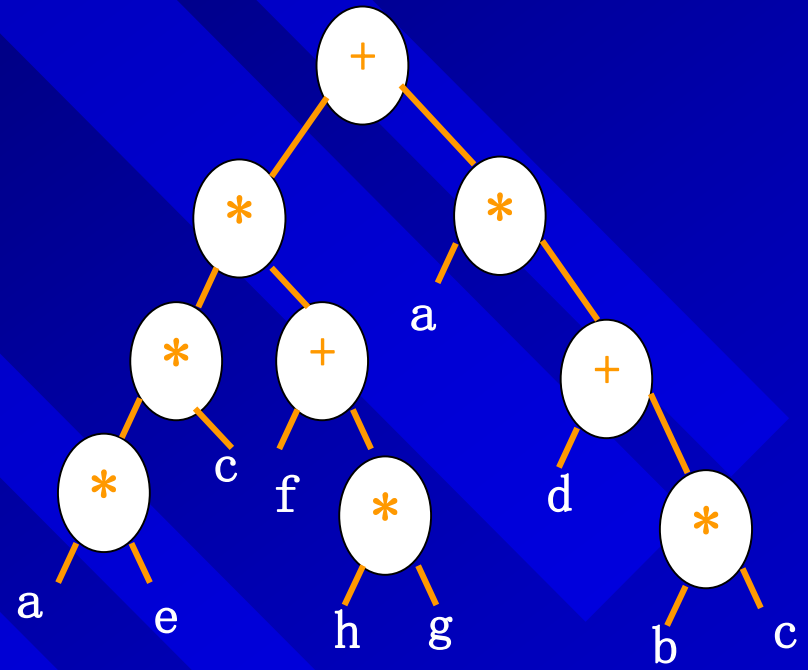
- 在单处理机上的计算的树形图如图7.4(a)所示，在多台处理机上并行运算的树形图如图7.4(b)所示。

$$E = a(b + c(d + e(f + gh)))$$

$$E = a(b + cd) + ace(f + gh)$$



(a)

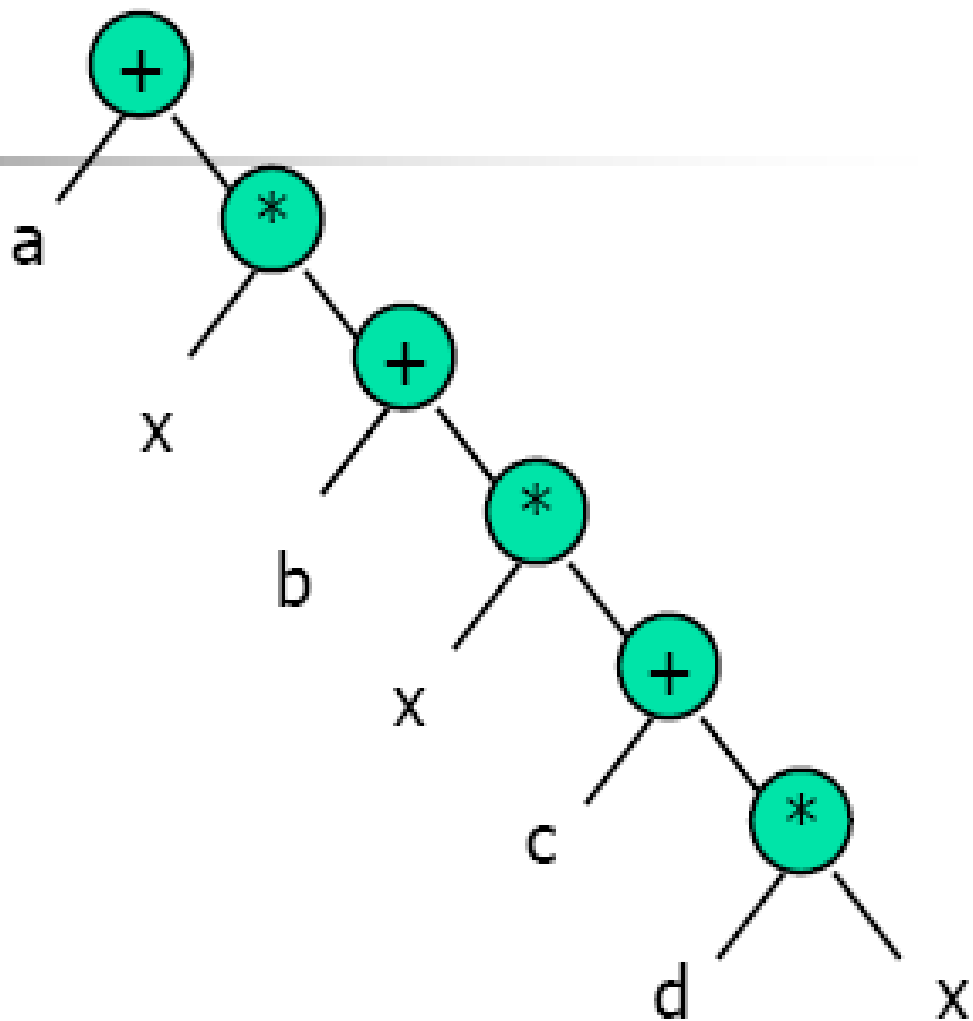


(b)

$$p = 3, T_p = 4, Sp = \frac{T_1}{T_p} = \frac{7}{4}, Ep = \frac{Sp}{p} = \frac{7}{12}$$

举例

- $E1 = a + bx + cx^2 + dx^3$
- 利用 Horner 法:
 $E1 = a + x(b + x(c + x(d)))$
- 需 3 个乘加循环, 6 级运算
- 适合于单处理机
- 用树形流程图



- 举例1(续)
- $E1=a+bx+cx^2+dx^3$
- 用3台处理机，需4级运算
- 级数(高度) $T_p=4$
- 处理机数 $P=3$
- 加速比 $S_p = \text{顺序运算级数 } T_1 / P$
 P 台处理机运算的级数 T_p
 $= 6/4 = 3/2$
- 效率 $E_p = S_p / P = 1/2$
- 即运算的加速总是伴随着效率的下降

