

5.2.3流水机器的相关处理和控制机构 (1)

相关的分类

- 局部相关：指令相关、主存操作数相关和通用寄存器组的操作数相关及基址值变址值相关被称为局部相关。
- 全局相关：转移指令和其后的指令之间存在关联，使之不能解释，其造成的流水线吞吐率和效率的下降要比指令相关、主存操作数相关和通用寄存器组的操作数相关及基址值变址值相关严重的多，所以被称为全局相关

1. 局部性相关的处理

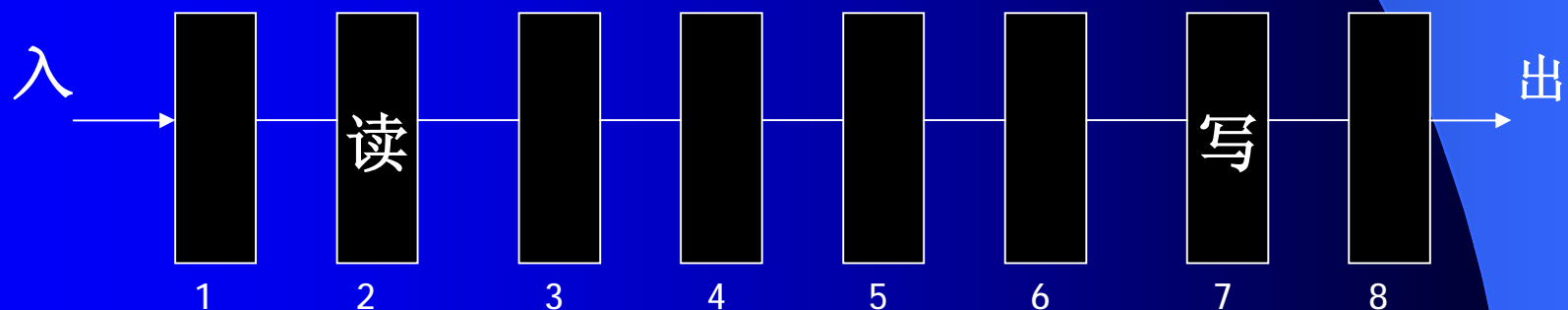
- 局部相关的基本处理方法
 - (1) 推后后续指令对相关单元的读；
 - (2) 设置相关专用通路。

5.2.3流水机的相关处理和控制机构 (2)

流水线中流动顺序的安排和控制

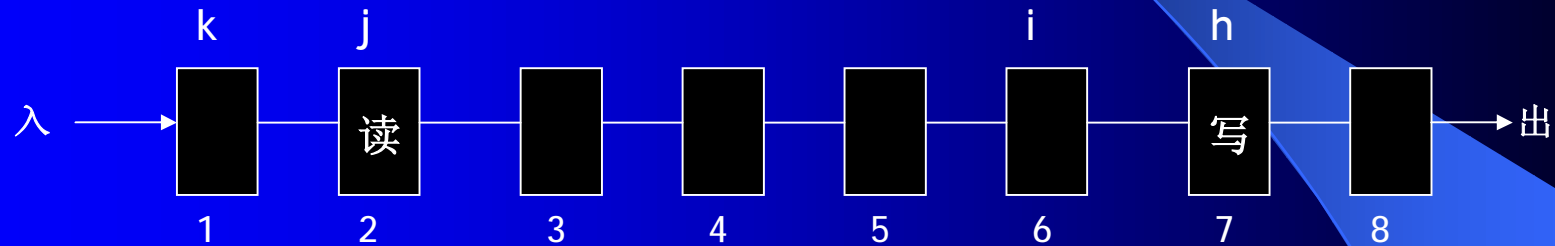
- 同步流动方式：任务或指令在流水线中的流动顺序与其流入流水线的顺序相同。
- 异步流动方式：任务或指令在流水线中的流动顺序与其流入流水线的顺序不同。

例如有一个8段的流水线，其中第2段为读段，第7段为写段，如下图所示：

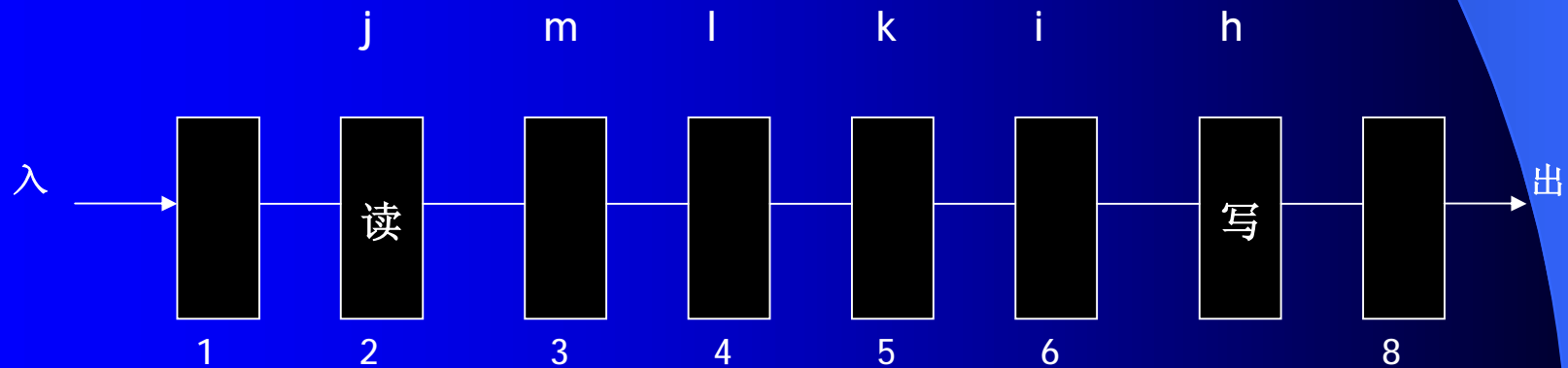


5.2.3流水机器的相关处理和控制机构 (3)

有一串指令h、i、j、k、l、m、n...依次流入，当指令j的源操作数地址与指令h的目的操作数地址相同时，就发生了先写后读的操作数相关。
按顺序流动方式处理时如下：

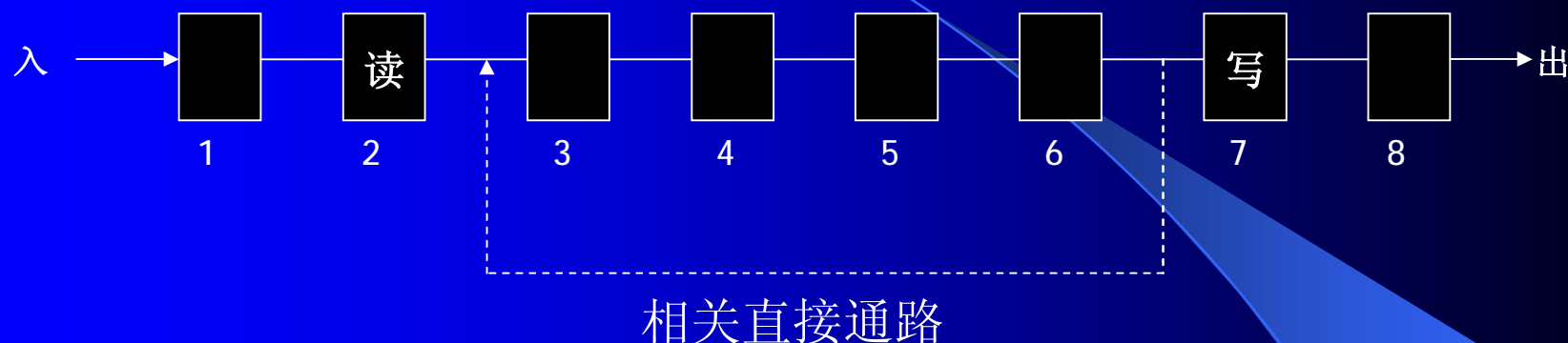


采用异步流动方式



5.2.3流水机器的相关处理和控制机构 (4)

设置相关专用通路



利用异步流动并设置相关专用通路，进一步提高流水效率。

5.2.3流水机器的相关处理和控制机构 (5)

异步流水线要处理的三种相关

- 先写后读相关：顺序流动和异步流动都可能出现的相关；
- 先读后写相关：要求指令序列中在先的指令先读，在后的指令后写
- 写—写相关：要求在先的指令先写，后面的指令后写。

5.2.3流水机的相关处理和控制机构 (6)

全局性相关的处理

全局性相关的概念

- 已进入流水线的转移指令（尤其是条件转移指令）和其后续指令之间的相关。

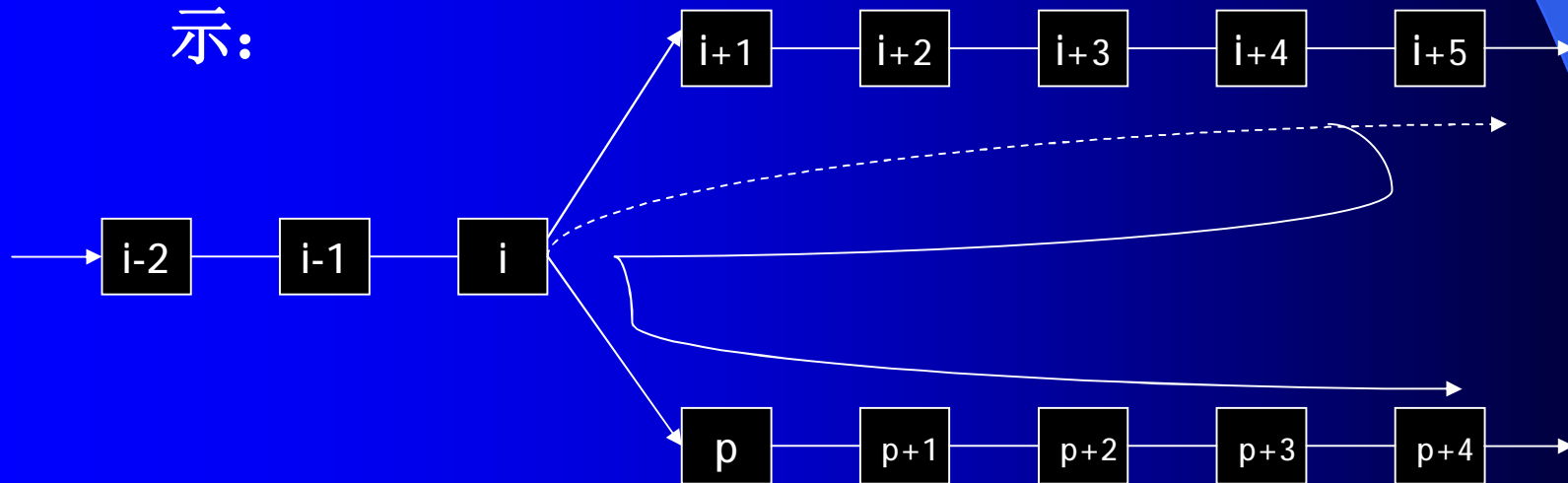
全局性相关的处理方法

- 猜测法：猜测指令的跳转方向，并按猜测结果使流水线继续向前流动。若指令*i*是条件转移指令，它有两个分支，一个分支是*i*+1、*i*+2、*i*+3...，按原来的顺序继续执行，称转移不成功分支。另一个分支是转向

，然后继续执行

+1、

+2...，成为转移成功分支。如下图所示：



全局性相关的处理 (2)

全局性相关的处理

– 猜测法—如何猜?

- ┆ 当转向两个分支概率相等时, 选择不成功转移分支;
- ┆ 两个分支概率不均等时, 猜选概率高的分支。

– 猜错了怎么办

- ┆ 已进入流水线的一些指令作废, 重新取指, 并执行
- ┆ 需要解决的关键问题—如何恢复分支处的现场
 - 进入流水线的分支只进行指令译码和取操作数, 不进行运算。
 - 进入流水线的分支进行运算, 但其运算结果不送回
 - 采用后援寄存器保存可能被破坏的原始数据, 如果猜测方向不对时, 再从后援寄存器中恢复原始状态。

┆ 提高恢复原分支点速度的方法

- 设置多套指令缓冲器

全局性相关的处理 (4)

全局性相关的处理

– 加快和提前形成条件码

- ┆ 加快单条指令内部条件码的形成
- ┆ 在一段程序内提前形成条件码，这特别适合于循环型程序在判断循环是否继续时的转移情况。

– 采用延迟转移技术

- ┆ 用软件方法进行静态指令调度的技术。

– 加快短循环程序的处理

- ┆ 将长度小于指令缓冲器容量的短循环程序整个一次性放入指令缓冲器，并暂停预取指令，从而减少访主存重复取指的次数。
- ┆ 而是由于循环分支概率高，因此让循环出口端的条件转移指令恒猜循环分支，以减少因条件分支造成的流水线断流。

流水机的中断处理

主要需要处理问题

– 断点现场的保护和恢复

┆ 不精确断点法

不论指令*i*在流水线的哪一段，都不允许尚未进入流水线的后续指令再进入，但已在流水线中的指令继续流动到执行完毕，然后才转入中断处理程序。

┆ 精确断点法

不论指令*i*是在流水线的哪一段响应中断，给中断处理程序的现场都是对应*i*的，*i*之后流入流水线的指令的原有现场都能恢复。

流水线的调度 (1)

主要讨论的问题:

- 在单功能非线性流水线中, 究竟要间隔几拍在送入下一个任务才既不发生功能段使用冲突又能是流水线有较高的吞吐率和效率?

二维预约表:

- 设一个单功能非线性流水线由 k 段组成, 每个任务通过流水线需要 N 拍。若任务在第 n 拍需要用到第 k 段, 就在第 n 列和第 k 行的交点用 \checkmark 表示。

拍号

	1	2	3	4	5	6	7	8	9
1	\checkmark								\checkmark
2		\checkmark	\checkmark					\checkmark	
3				\checkmark					
4					\checkmark	\checkmark			
5							\checkmark	\checkmark	

段号

流水线的调度 (2)

- 延迟禁止表

流水线中各段对同一任务流过时会争用同一段的拍间隔数汇集在一起，所形成的集合称为延迟禁止表F。本例中： $F1=\{8\}$ ， $F2=\{1,5,6\}$ ， $F3=\{\}$ ， $F4=\{1\}$ ， $F5=\{1\}$ ，因此有： $F=\{1, 5, 6, 8\}$ 。

- 冲突向量

后继新任务间隔各种不同的拍数送入流水线时，是否会发生功能段的使用冲突。

$(C_{N-1}, \dots, C_i, \dots, C_2, C_1)$

$C_i=0$ 后继任务间隔*i*拍进入流水线，不会发生功能段的使用冲突；

$C_i=1$ 后继任务间隔*i*拍进入流水线，会发生功能段的使用冲突；

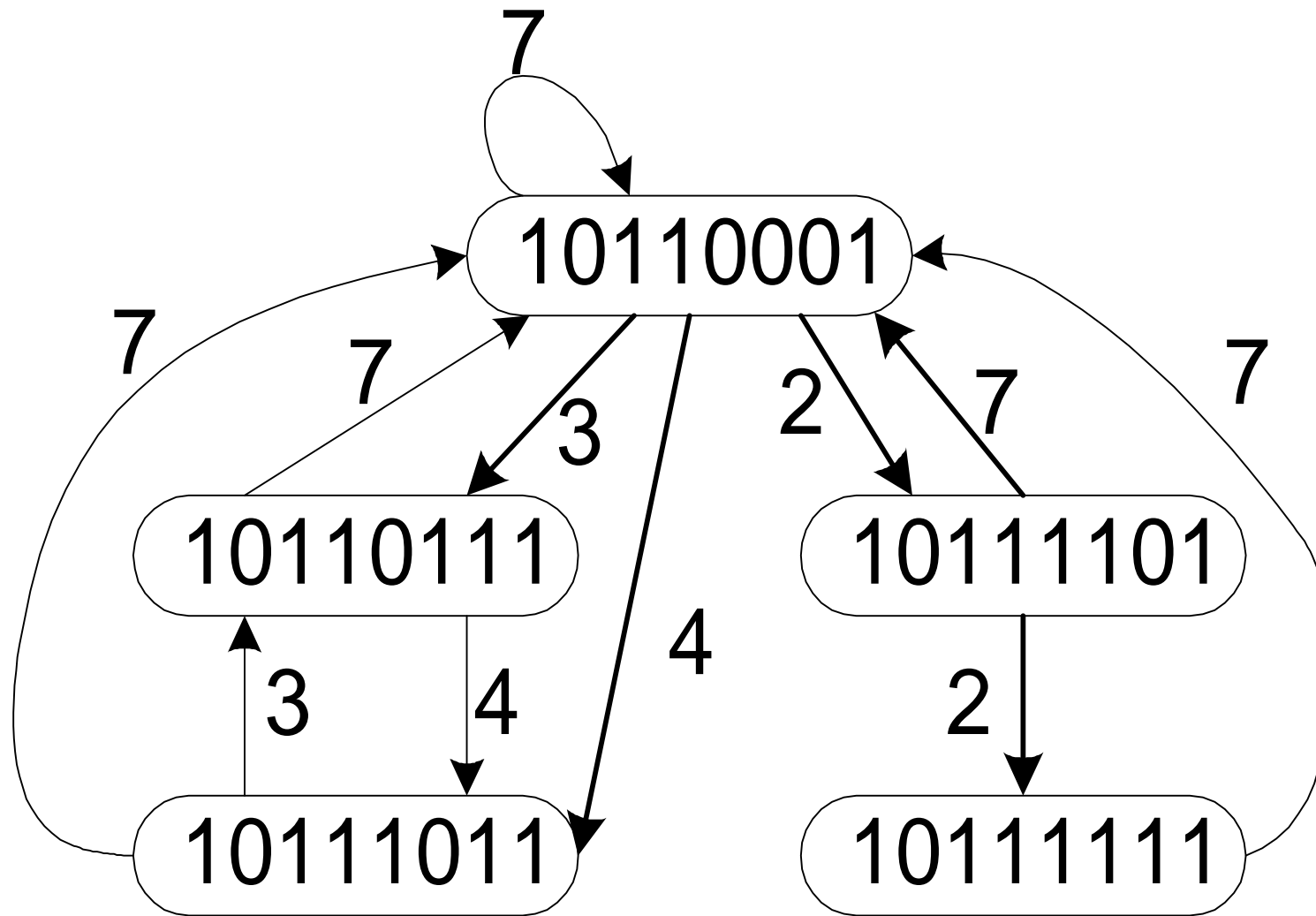
- 初始冲突向量：

第一个任务进入流水线时形成的冲突向量。本例为
(10110001)

流水线的调度 (3)

- 当没有新任务进入流水线时冲突向量的变化
冲突向量放在一个移位寄存器中，每拍逻辑右移1位，左面补0，随着任务在流水线中的推进，不断动态的形成当前的冲突向量。
- 当有新的任务进入流水线时，冲突向量的变化
第三个任务进入流水线后，既不能与早已进入流水线的任务争用功能段，也不能与刚刚进入流水线的任务发生功能段争用。
- 单功能流水线的状态图：
表示了流水线冲突向量之间的转换关系

流水线的调度 (4)



流水线的调度 (5)

各种调度方案的平均间隔拍数

调度方案	平均间隔拍数	调度方案	平均间隔拍数
(2,2,7)	3.67	(3,7)	5.00
(2,7)	4.50	(4,3,7)	4.67
(3,4)	3.50	(4,7)	5.50
(4,3)	3.50	(7)	7.00
(3,4,7)	4.67		

5.3 向量的流水处理与向量流水处理机

- 向量处理机 (Vector Processor): 由于向量内部各个元素之间 (分量) 之间很少相关, 且对这些元素 (分量) 一般是执行同一种操作, 比较容易发挥出流水线的效能, 所以一般是将向量数据表示和流水线技术结合起来, 构成向量流水处理机, 简称为向量处理机 (Vector Processor) 以提高主要面向向量数组计算类的计算机的速度性能。

5.3.1 向量的流水处理

向量的处理方式

– 需解决的关键问题

找到最能充分发挥流水线效能的处理方式

– 向量的不同处理方式

例：要计算 $D=A*(B+C)$,其中A、B、C、D都是具有N个元素的向量，应该采用什么样的处理方式才能最充分发挥流水线的效能呢？

• 横向（水平）处理方式

逐个求D向量元素，即访存取 a_i, b_i, c_i 计算出 d_i 再取

$a_{i+1}, b_{i+1}, c_{i+1}$

– 横向处理方式存在的主要问题：很难发挥流水线的效率。

• 纵向（垂直）处理

对整个向量只有将相同的操作都执行完之后，在转去执行别的操作。首先计算

$a_i + b_i \rightarrow k_i (i \text{ 从 } 1 \text{ 到 } N)$ ，然后再 $k_i * c_i \rightarrow d_i (i \text{ 从 } 1 \text{ 到 } N)$

5.3.1 向量的流水处理(2)

■ 向量流水机的分类

- 面向存储器—存储器型结构的流水线处理机
源向量和目标向量都存放在主存中

关键技术：提高主存的频宽。

- 面向寄存器—寄存器型结构的流水线处理机

流水线的输入端和输出端连接在一个寄存器组上。源向量和目标向量存储在寄存器组中。对于大向量可以进行分段处理。

5.3.2 向量流水计算机

■ 向量处理机的指令系统

– 指令的分类

■ 向量指令和标量指令

– 向量指令的分类

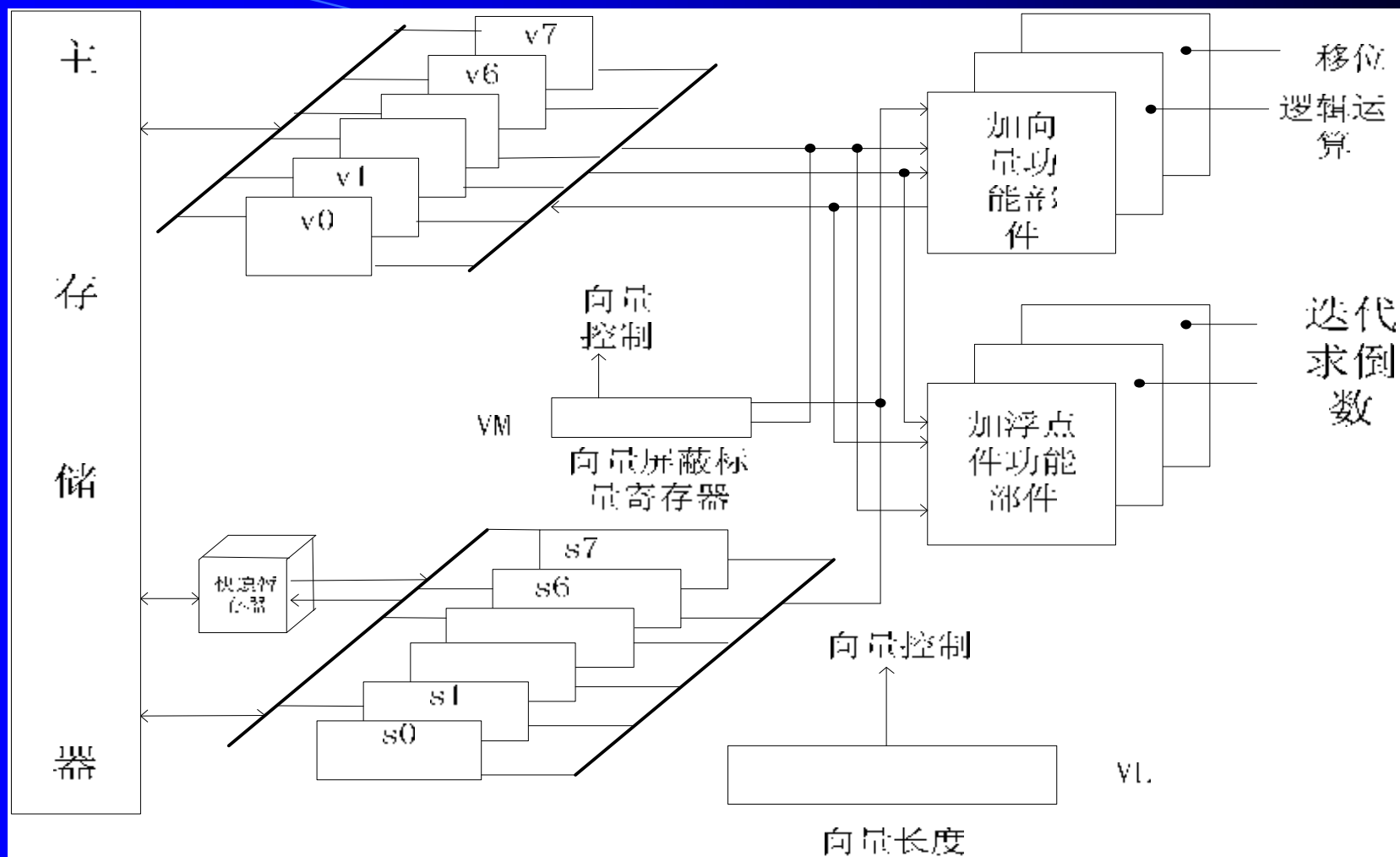
■ 向量V1运算得向量V2, 如 $V2 = \text{SIN}(V1)$;

■ 向量V1运算得标量S, 如 $S = \sum_{i=1}^n V_i$;

■ 向量V1与向量V2运算得向量V3, 如 $V3 = V1 \wedge V2$;

■ 向量V1与标量S运算得向量V2, 如 $V_2 = S * V_1$;

■ 向量比较指令, 向量压缩指令, 归并指令, 向量传送指令等。



注意：向量寄存器组与功能部件之间不是一般的总线联系

图 5.32 CRAY-1的向量流水处理部分简图

2. 向量流水处理机的结构(1)

CRAY-1的向量流水部分结构上的特点

- 向量运算由6个流水线单功能部件组成;
- 向量寄存器共分为8组, 每组64个64位的寄存器;
- 每组寄存器都有连结到6个功能部件的单独总线
- 每个寄存器组在一个时钟周期内可接受一个结果分量, 并可以为下一次操作再提供源分量。

Vi冲突

- 并行工作的的各向量指令的源向量或结果向量使用了相同的Vi。(可能出现源向量冲突)

$$V4 \leftarrow V1 + V2$$

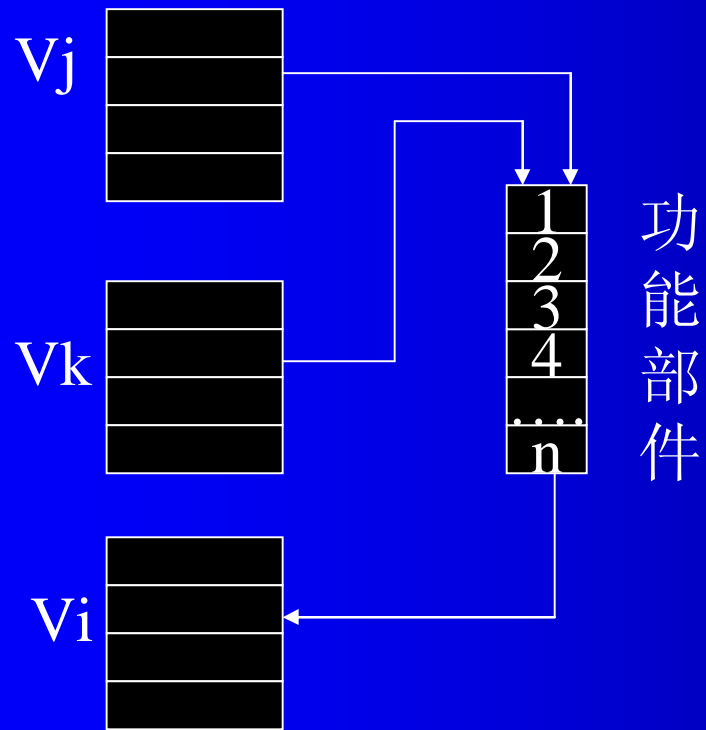
$$V5 \leftarrow V1 \wedge V3$$

• 功能部件冲突

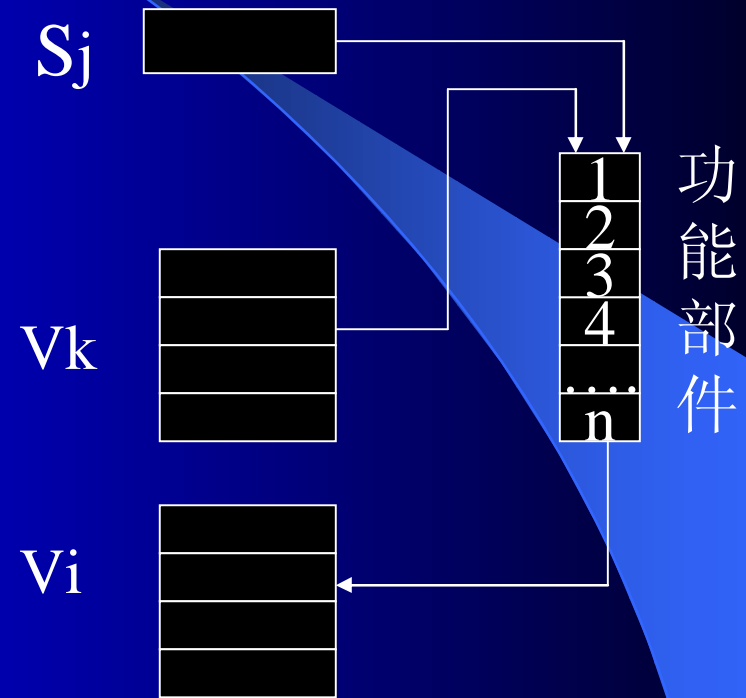
- 同一功能部件被一条以上的要求并行工作的向量指令使用。 $V4 \leftarrow V2 * V3$, $V5 \leftarrow V1 * V6$

2. 向量流水处理机的结构(2)

CRAY-1中的4种向量指令



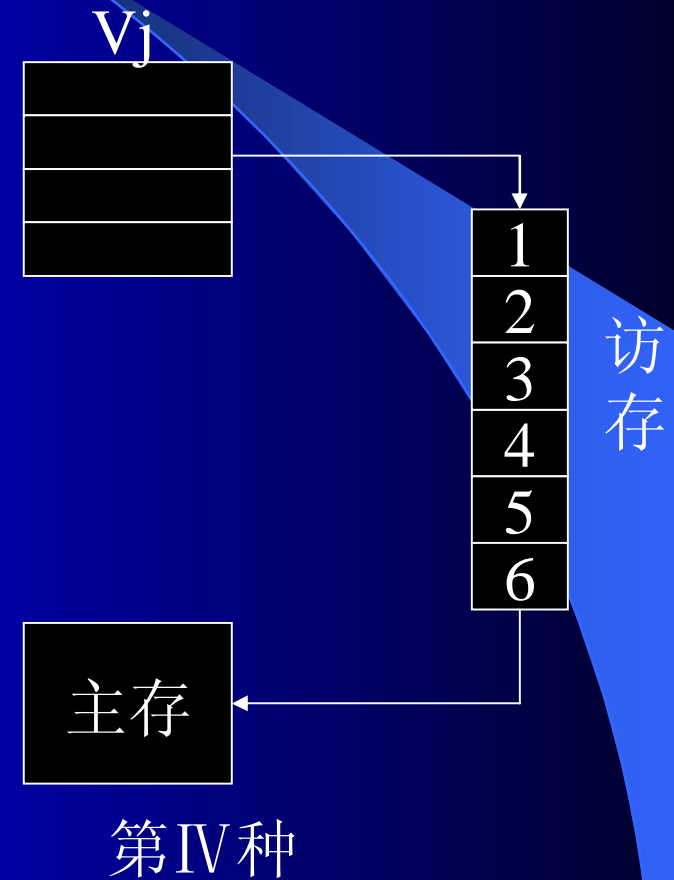
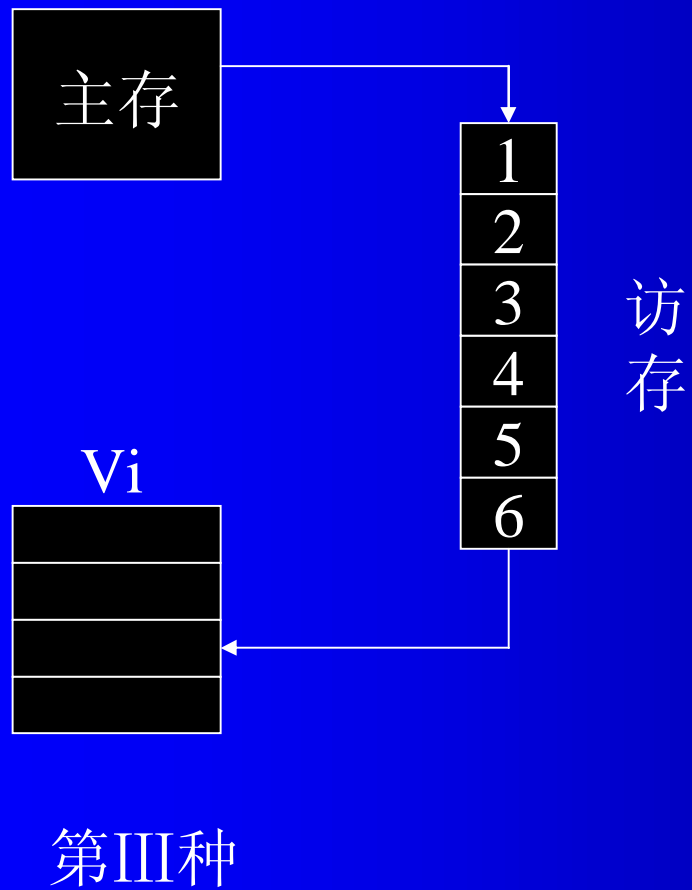
第 I 种



第 II 种

2. 向量流水处理机的结构(3)

CRAY-1中的4种向量指令



2. 向量流水处理机的结构(4)

CRAY-1向量处理的一个显著特点

- 只要不出现功能部件冲突和源向量冲突，通过链接机构可使有数据相关的向量指令仍能重叠并行处理

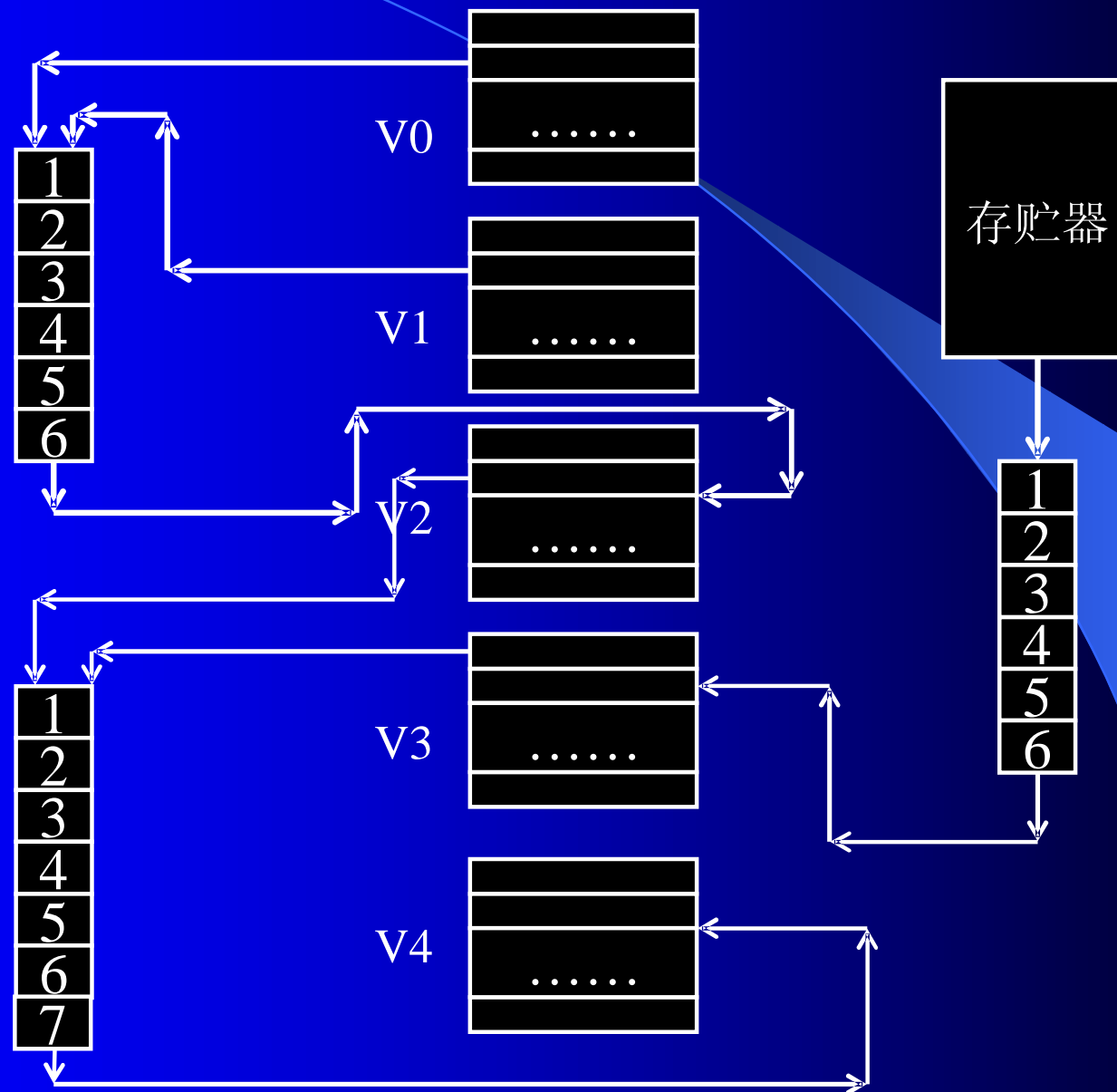
例如：向量运算 $D=A*(B+C)$ 若向量长度 $N \leq 64$ ，向量为浮点数，则在B、C取到V0、V1后，就可以用以下3条向量指令求解：

$V3 \leftarrow \text{存贮器 (访存取A向量)}$

$V2 \leftarrow V0+V1$ (B向量和C向量浮点加)

$V4 \leftarrow V2*V3$ (浮点乘，存D向量)

2. 向量流水处理机的结构(5)



2. 向量流水处理机的结构(6)

从访存开始直至把第一个结果分量存入V4所需的拍数

$$1 \left\{ \begin{array}{l} \text{启动访存} \\ \text{送浮点加部件} \end{array} \right\} + 6 \left\{ \begin{array}{l} \text{访存} \\ \text{浮加} \end{array} \right\} + 1 \left\{ \begin{array}{l} \text{存V3} \\ \text{存V2} \end{array} \right\} + 1 \left\{ \begin{array}{l} \text{送浮乘部件} \\ \text{送浮乘部件} \end{array} \right\} + 7 \left\{ \text{浮乘} \right\} \\ + 1 \left\{ \text{存V4} \right\} = 17 \text{拍}$$

此后，每拍就可取得一个结果分量存入V4，一共只需17+ (N-1)拍就可以执行完这3条向量指令，获得全部结果分量。

显然这要比第一、二条指令执行完，所需分量全部送入V2、V3后，才开始执行第三条指令要快得多，因为后者需1+6+1+N-1+1+7+1+N-1=15+2N拍

5.4 指令级高度并行的超级处理机 (1)

■ 超标量处理机

— 常规流水处理机

假设一条指令包含取指令、译码、执行、存结果4个子过程，每个子过程经过时间为 t 。

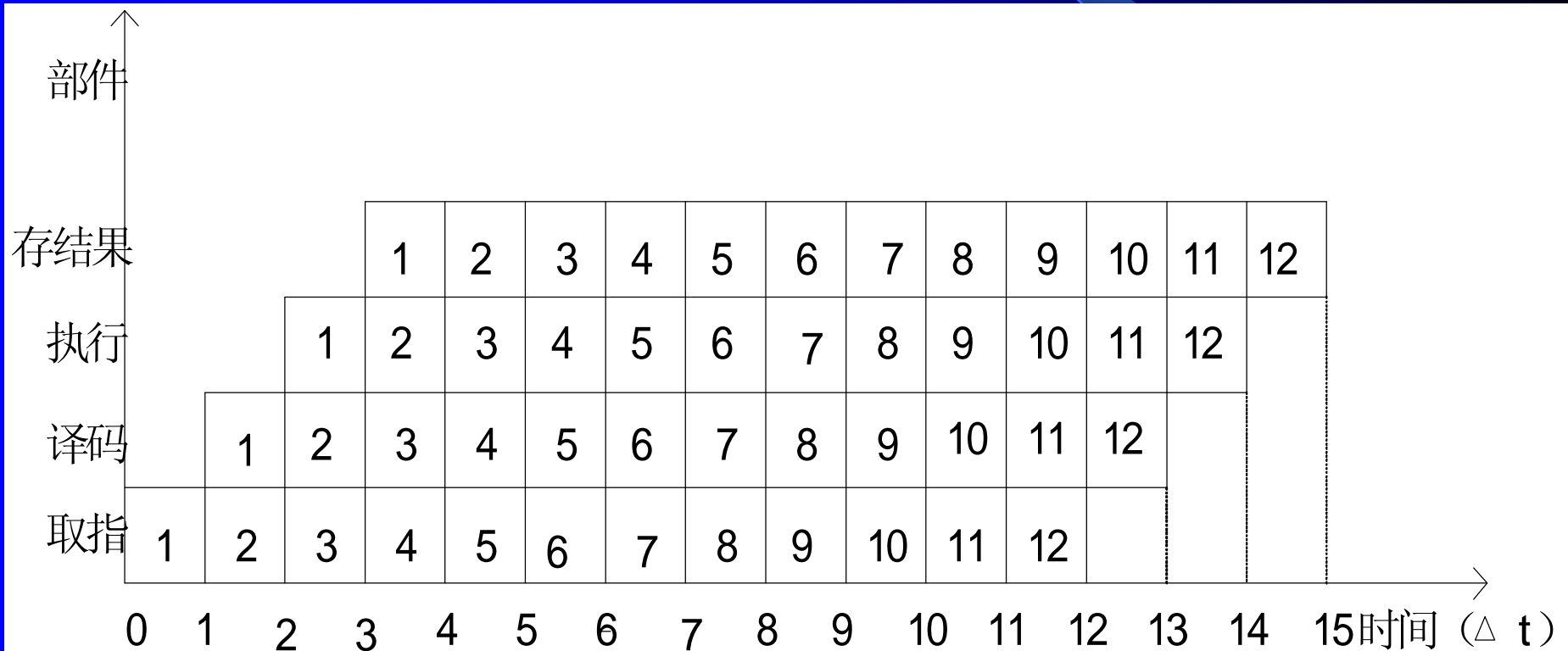


图5.33常规 (度=1) 的表量流水机时 -空图

5.4 指令级高度并行的超级处理机 (2)

超标量处理机

- 采用多指令流水线，每个 t 同时流出 m 条指令（称度为 m ）。

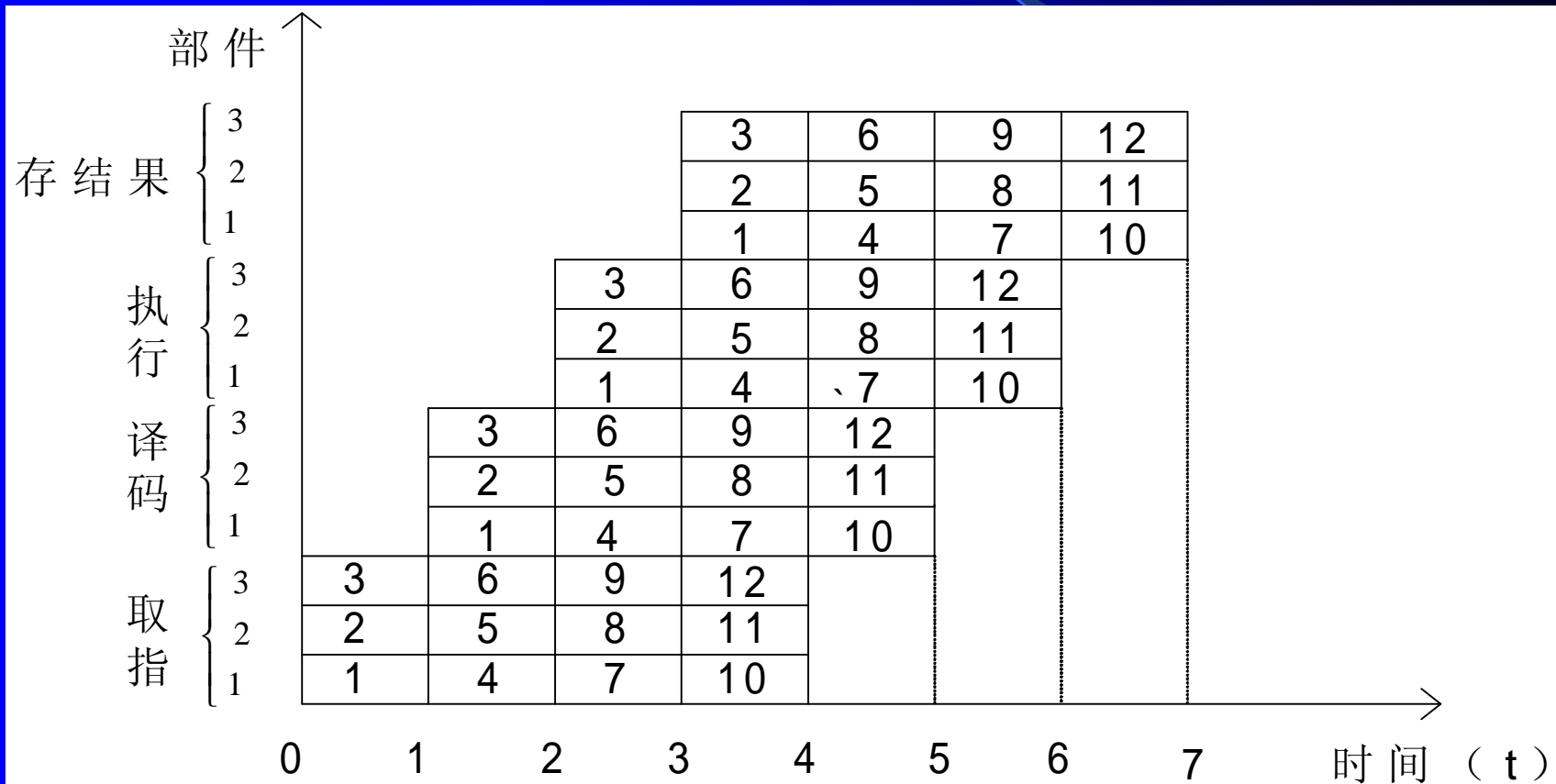


图 5.34 度 $m=3$ 的超标量处理机时 - 空图

5.4 指令级高度并行的超级处理机 (3)

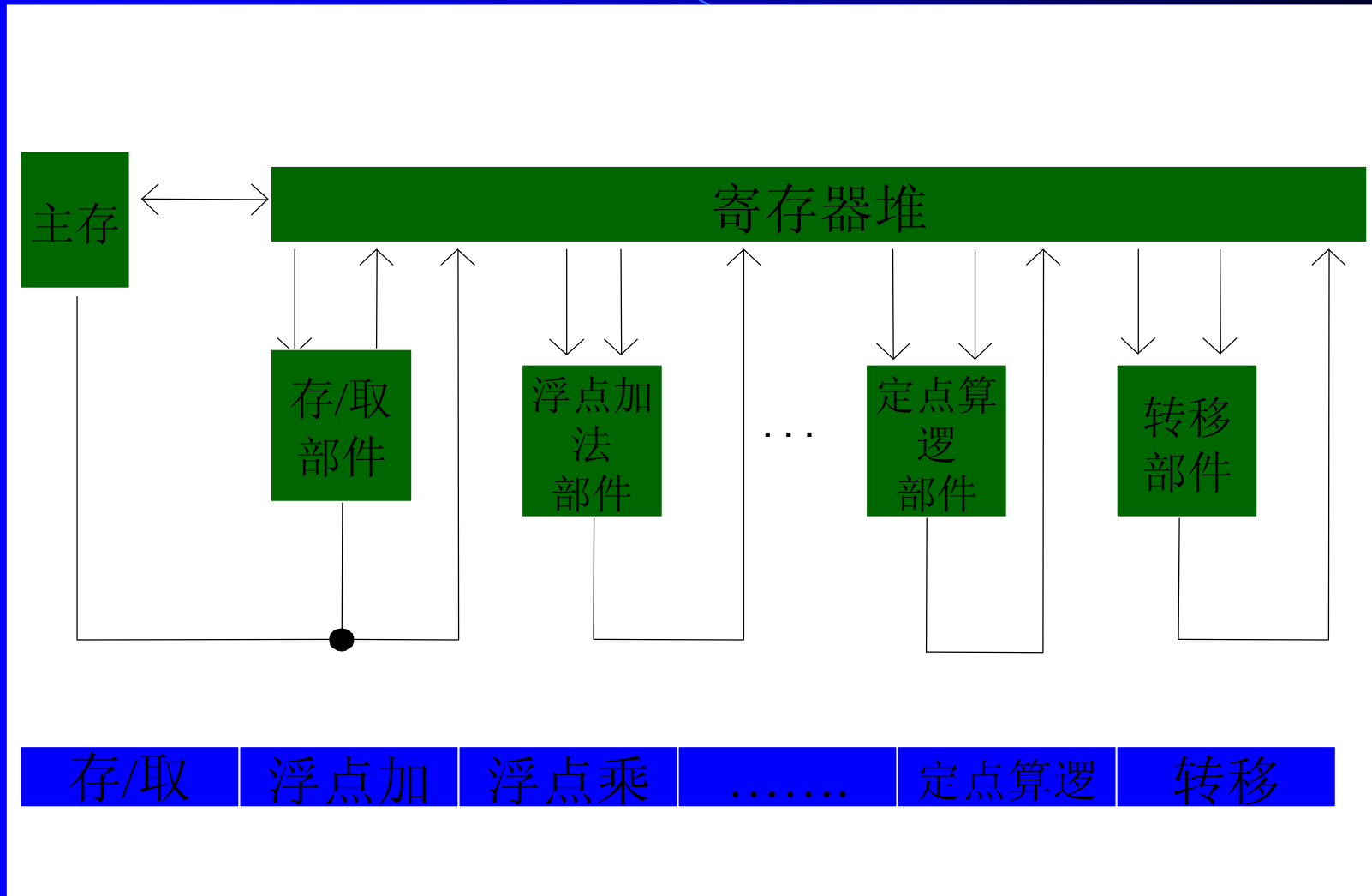
■ 超常指令字处理机

– 特点

- (1) 指令字长可达数百位;
- (2) 多个功能部件并发工作;
- (3) 共享寄存器堆;
- (4) 编译时编译程序找出指令间潜在的并行性, 将多个能并行执行的不相关或无关的操作先行压缩组合在一起, 形成一条有多个操作段的超长指令;
- (5) 这条超长字指令控制机器中多个相互独立的功能部件并行操作。每个操作码段控制其中的一个功能部件, 相当于同时执行多条指令;
- (6) 是一种单指令多操作码多数据的系统结构(SIMOMD)。

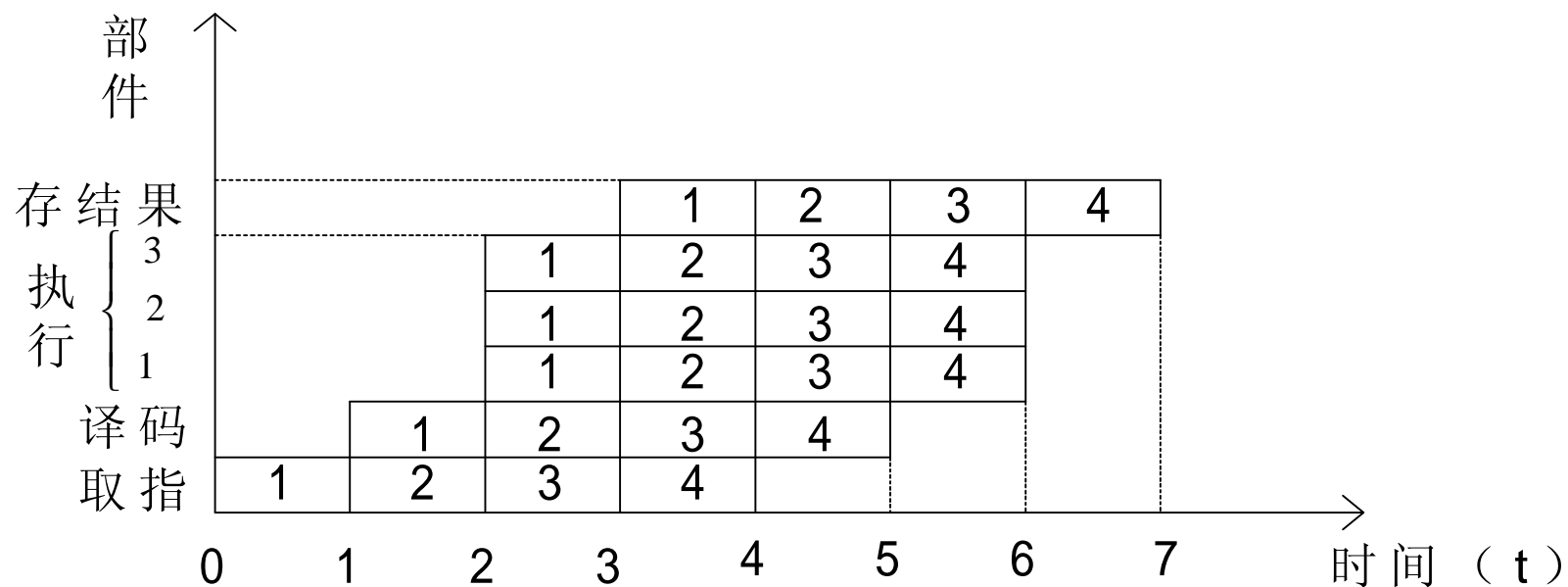
5.4 指令级高度并行的超级处理机 (4)

(a) 典型的VLIW处理机组成和指令格式



(a) 典型的VLIW处理机组成和指令格式

5.4 指令级高度并行的超级处理机 (5)



(b) 度 $m=3$ 时的执行时-空图

图 5.35 超长指令字 (VLIM) 处理机

5.4 指令级高度并行的超级处理机 (6)

超流水处理机

– 关键技术：多相时钟技术，高速时钟机制

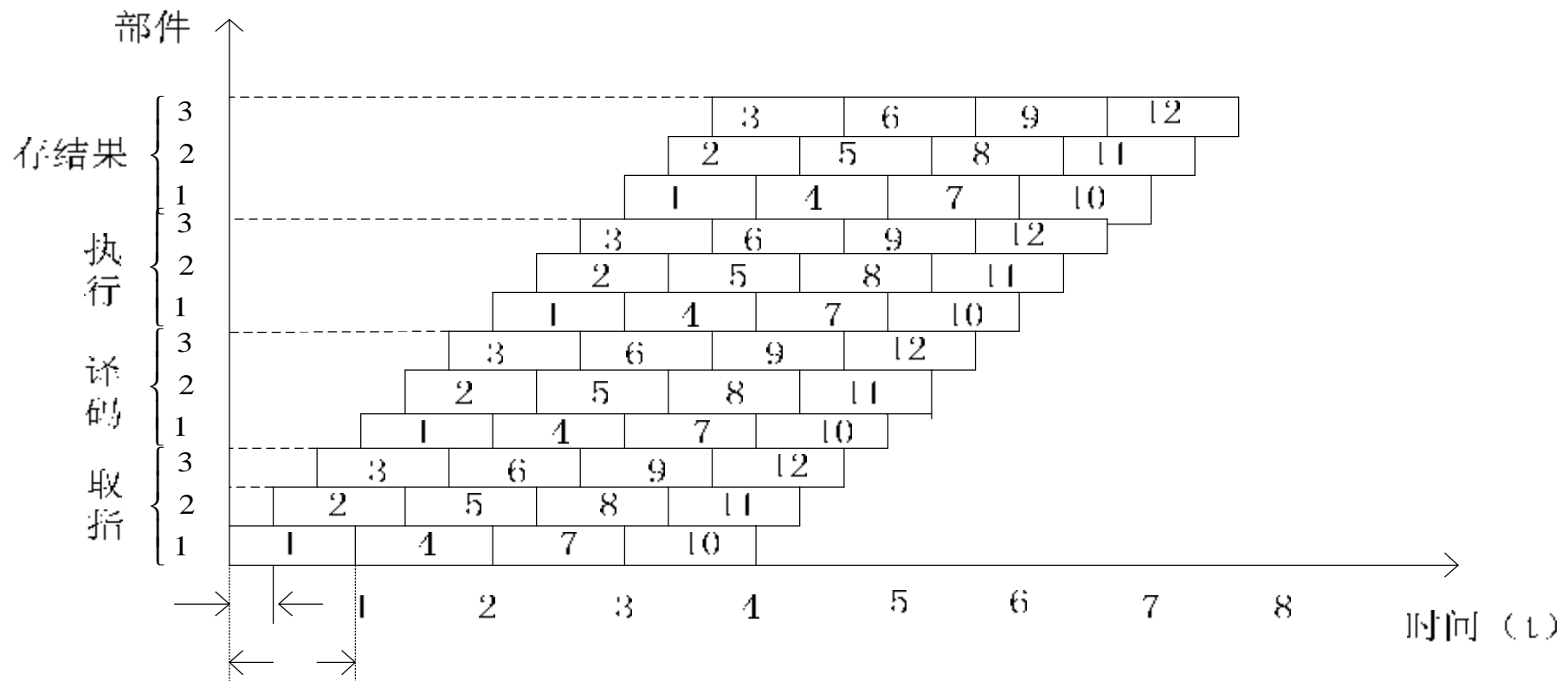


图5.36 每流出一条指令，且度 $m=3$ 的超流水线处理机时-空图