

# 计算机组成原理与系统结构

## 第七章 控制器

<http://jpkc.hdu.edu.cn/computer/zcyl/dzkjdx/>





## 7.4 流水线原理



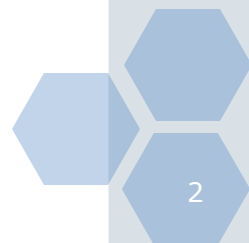
流水线基本概念



流水线分类



流水线性能分析





# 一、流水线基本概念

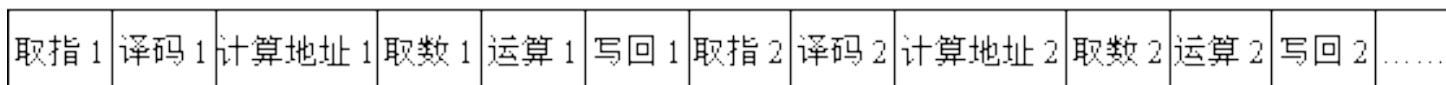
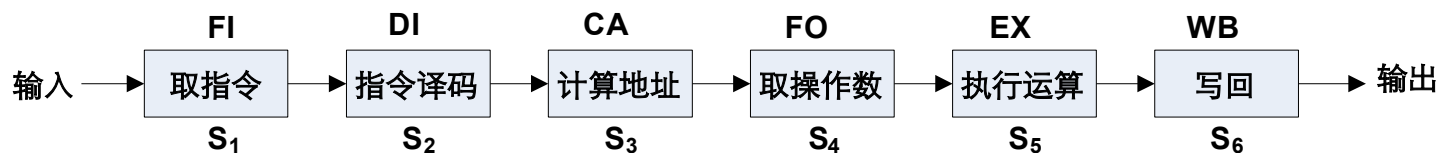
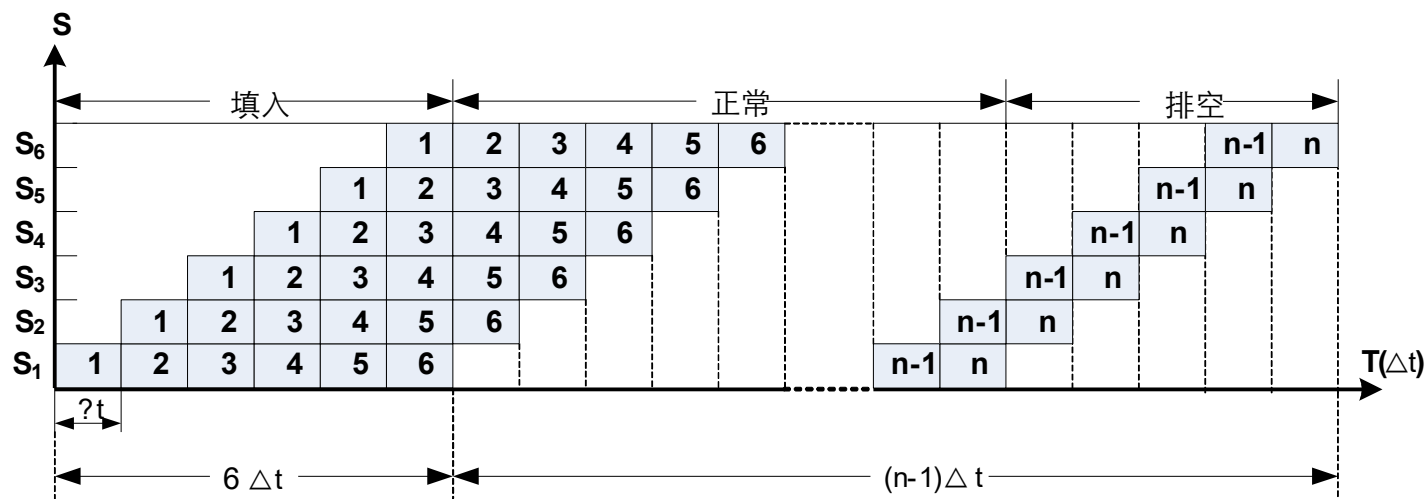


图10.1 指令的串行执行



## 指令流水线的连接图



## 指令流水线的时空图



# 一、流水线基本概念

## ❖ 流水线工作方式的特点：

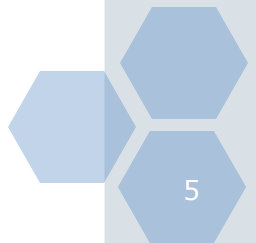
- 流水段划分越多，可同时运行的指令就越多，单位时间内可完成的指令也就越多，速度越快。
- 流水线上每个阶段的执行时间应尽量相等，否则会引起“**堵塞**”和“**断流**”。
- 流水线工作阶段分为建立（充入）、满载、排空三个阶段。流水线上必须等待一段时间，才能达到最大吞吐率，即**建立阶段**。建立时间等于一条指令的执行时间，又称为“**通过时间**”。只有流水线完全充满时，整个流水线的效率才得到充分发挥。



# 一、流水线基本概念

## ❖ 流水线工作方式的特点：

- 当编译形成的程序不能发挥流水线的作用，或存储器供应不上流动所需的指令和数据，或遇到程序转移指令等情况时，会造成流水线断流，使效率下降。

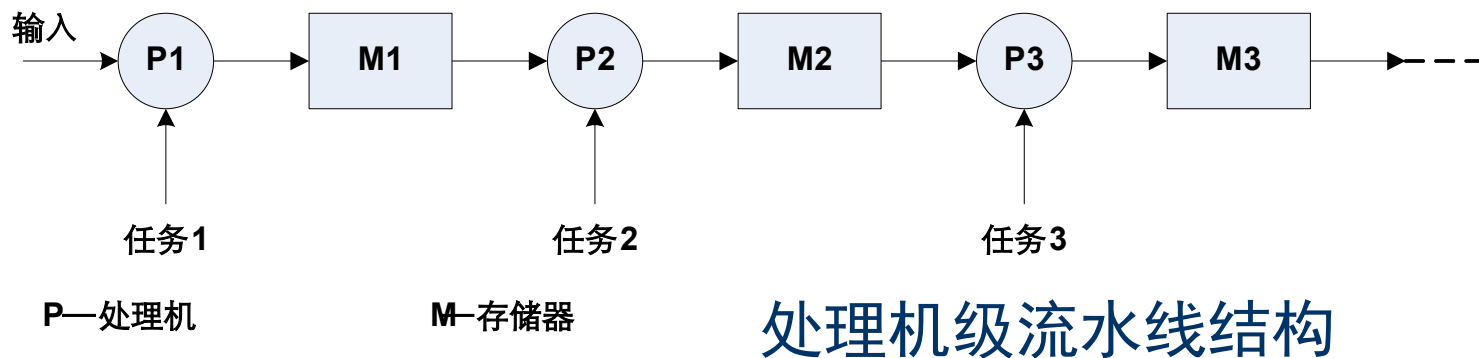




## 二、流水线分类

### 1. 按照任务级别分类

- 流水线按照任务级别，从低到高可以分为**操作部件级、指令级**和**处理机级**3种。

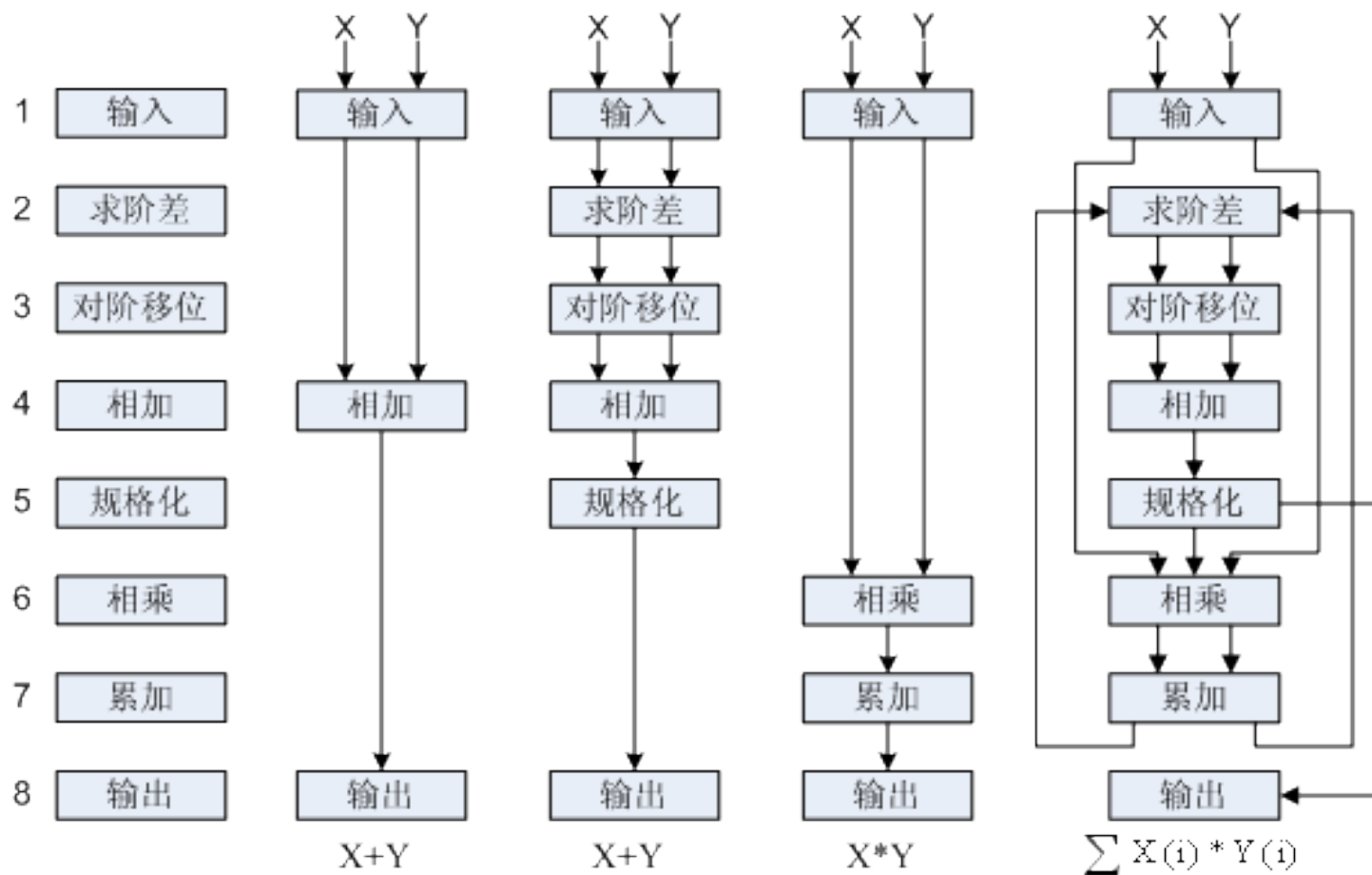


### 2. 按功能分类

- 流水线按照功能可分为**单功能流水线**和**多功能流水线**两种。



# TI-ASC 计算机的多功能流水线



(a) 流水段

(b) 定点加法

(c) 浮点加法

(d) 定点乘法

(e) 浮点点积



## 二、流水线分类

### ❖ 按工作方式分类

- 流水线按工作方式可以分为静态流水线和动态流水线。动态流水线一定是多功能流水线，而单功能流水线一定是静态的。
- 流水线按连接方式可以分为线性流水线和非线性流水线。
- 按处理机类型分类可以分为标量流水线和向量流水线。

标量流水线处理机也可以处理向量数据，但是它是采用流水处理方式对向量各元素（标量数据），使用标量指令进行处理。而向量流水线处理机内具有向量指令流水线，它能够对一个向量指令序列直接进行流水处理。







## 三、流水线性能分析

- ❖ 衡量流水线性能的主要技术指标有三个：  
**吞吐率、加速比和效率**

- 1 吞吐率 ( Throughput Rate )
- 2 加速比 ( Speedup Ratio )
- 3 效率 ( Efficiency )
- 4 举例





# 1、吞吐率（Throughput）

- ❖ 流水线的吞吐率是指：**单位时间内能处理的任务数或输出结果的数量。**
- ❖ 在指令级流水线中，吞吐率是指**单位时间内流水线所完成指令数或输出指令结果的数量。**



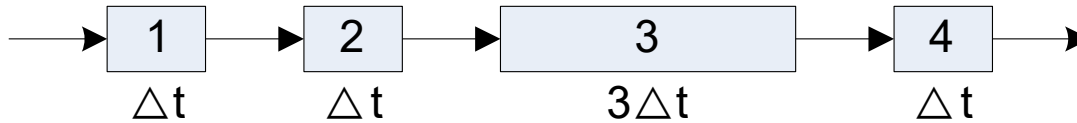


# 最大吞吐率

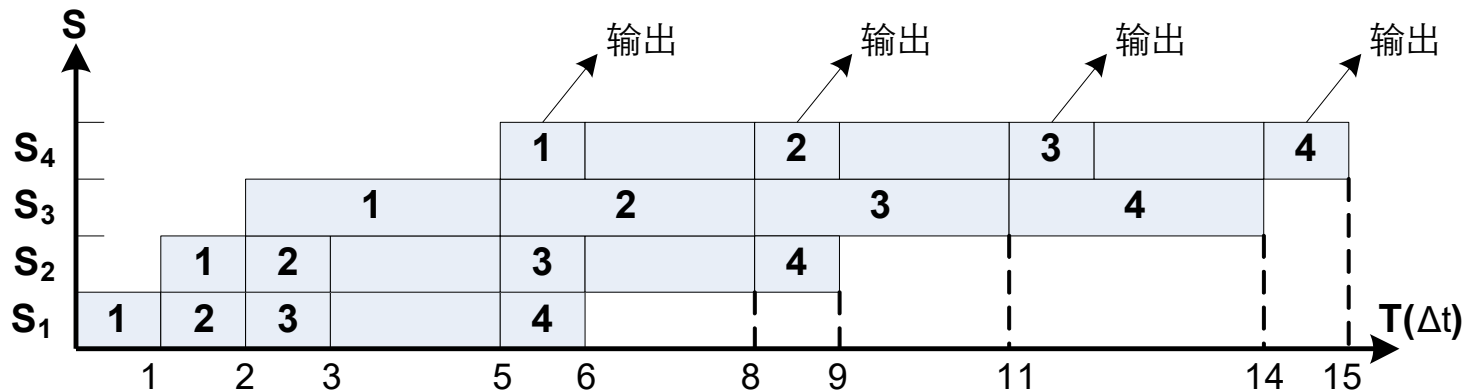
- ① 最大吞吐率：流水线在连续流动达到满载状态后所获得的吞吐率。
- ② 对于  $m$  段的指令流水线而言，若各段的时间均为  $\Delta t$ ，则最大吞吐率为： $TP_{MAX}=1/\Delta t$
- ③ 如果各功能段的操作时间不相等，那么吞吐率将取决于流水线中最慢的功能段的处理时间，其计算公式为： $TP_{MAX}=1/\max\{\Delta t_i\}$
- ④ 流水线中延迟时间最大的功能段将成为流水线中的瓶颈。
- ⑤ 消除流水线瓶颈的基本方法有两种：分离瓶颈段和重复设置瓶颈段



# 带有瓶颈的 4 段流水线



## 带有瓶颈的 4 段流水线

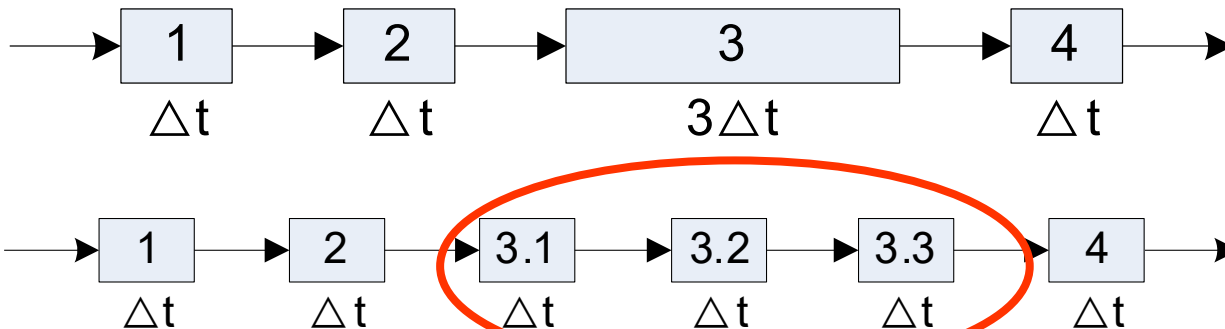


## 带有瓶颈的 4 段流水线时空图

# 分离瓶颈段

- I. 将瓶颈子过程进一步细分为若干个子过程，使每个子过程的延迟时间与其他子过程时间相等。
- II. 将功能段 3 划分为 3 个子过程  
3.1、3.2、3.3，使流水线段数增加为 6 段，每段的延迟时间均等于  $\Delta t$ ，这样，流水线的最大吞吐率增加为  $T_{PMAX} = 1 / \Delta t$

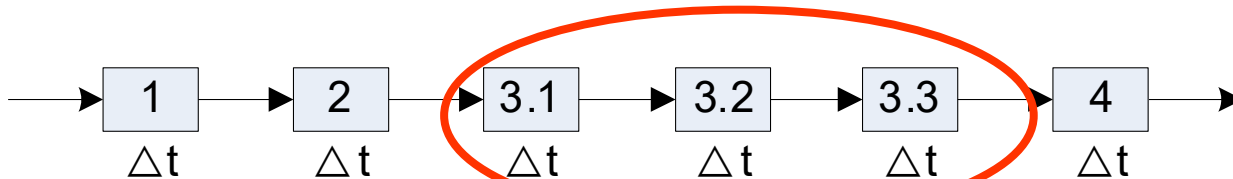
## 带有瓶颈的 4 段流水线



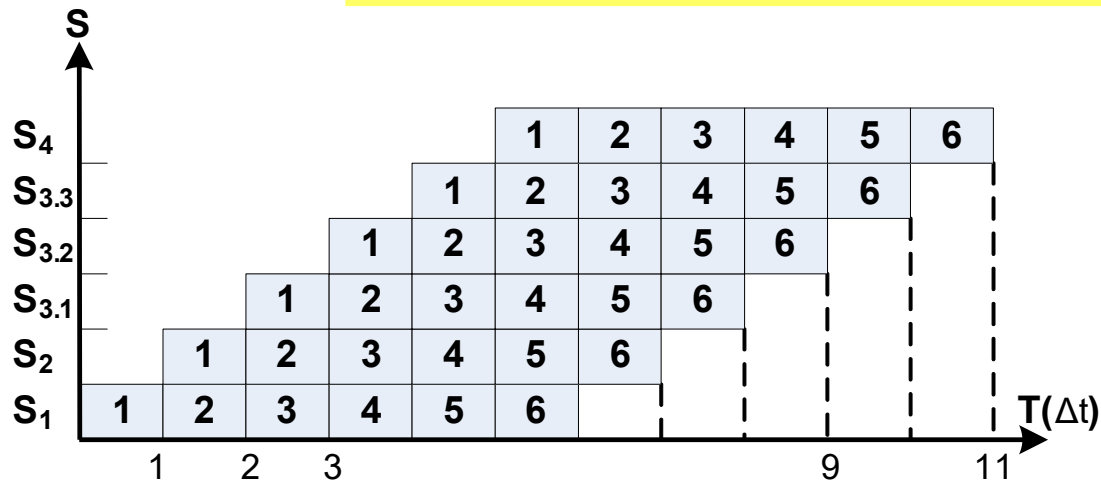
## 细分瓶颈段后的流水线



# 分离瓶颈段



## 细分瓶颈段后的流水线



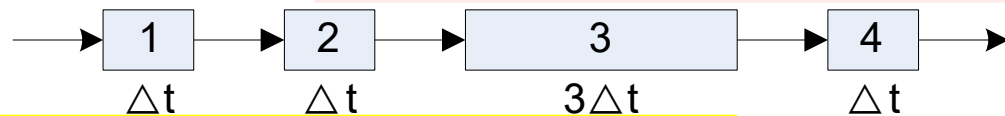
## 细分瓶颈段后的流水线时空图



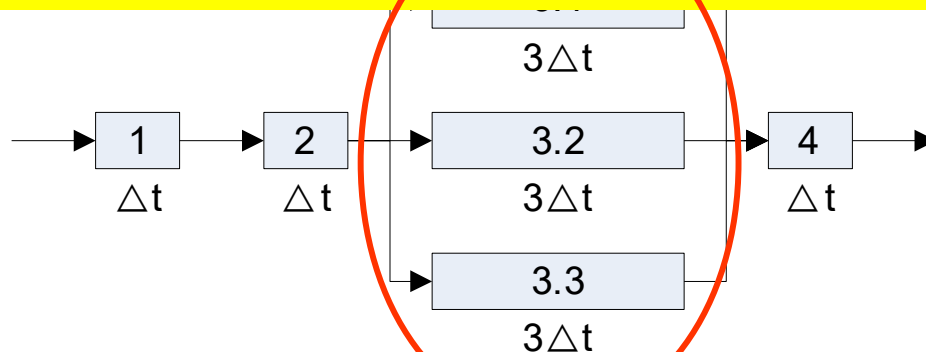
# 重复设置瓶颈段

- ❖ 在瓶颈段，并联设置多套功能部件，使它们轮流工作。当瓶颈子过程无法再细分时，采取这种方法来消除瓶颈。在功能段 3，设置 3 套同样的功能部件并行处理流入的任务，使流水线的最大吞吐率仍旧保持为  $TP_{MAX} = 1/\Delta t$

## 带有瓶颈的 4 段流水线



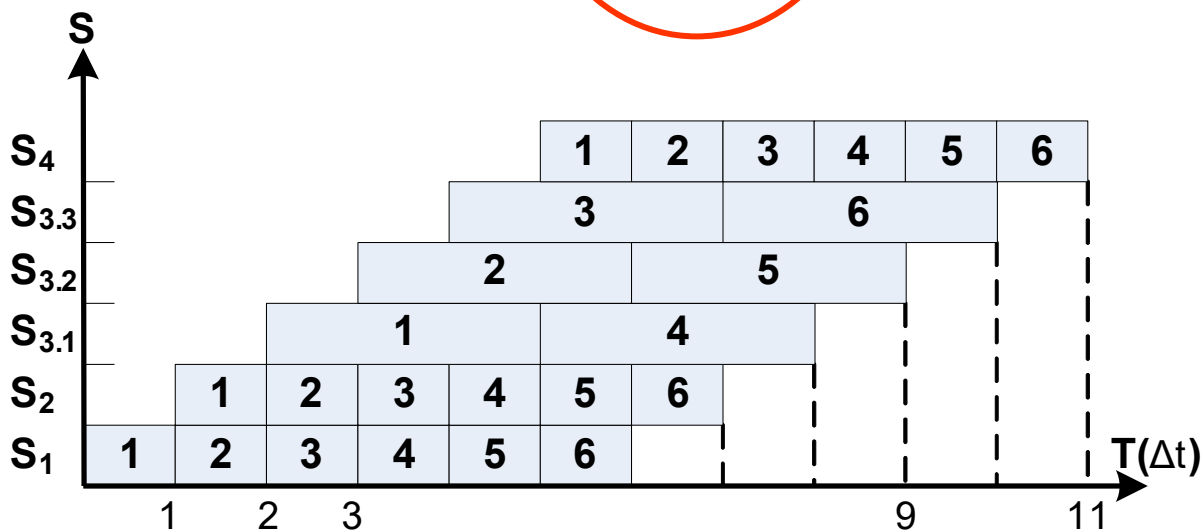
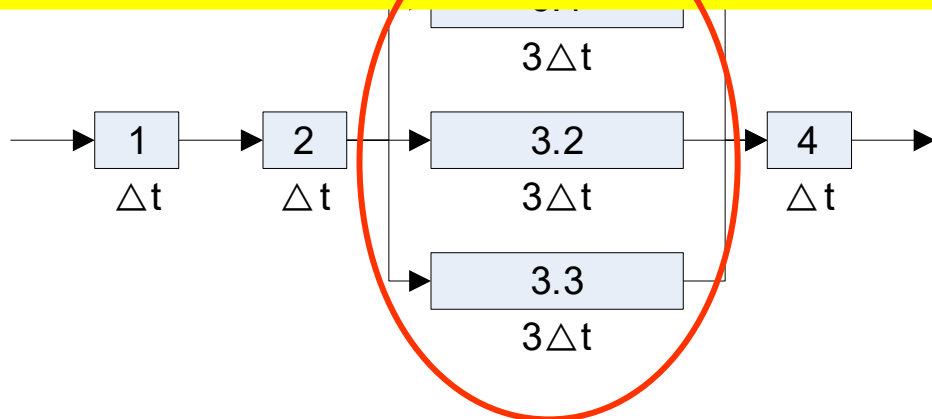
## 并联设置多套瓶颈段后的流水线





# 重复设置瓶颈段

## 并联设置多套瓶颈段后的流水线



并联设置多套瓶颈段后的流水线时空图







# 实际吞吐率

①实际吞吐率是指流水线完成  $n$  条指令的实际的吞吐率，其计算公式

$$TP = \frac{\text{完成的任务数}n}{\text{完成}n\text{个任务的时间}}$$

②实际吞吐率总是小于最大吞吐率

③对于  $m$  段的指令流水线，若各段的处理时间均为  $\Delta t$ ，连续处理  $n$  条指令，除第一条指令需  $m \cdot \Delta t$  外，其余  $(n-1)$  条指令，每隔  $\Delta t$  就有一个结果输出，即总共需时间  $m \cdot \Delta t + (n-1) \cdot \Delta t$ ，故实际吞吐率为：

$$TP = \frac{m \cdot \Delta t + (n-1) \cdot \Delta t}{m \cdot \Delta t + (n-1) \cdot \Delta t} = \frac{\Delta t \cdot (1 + \frac{m-1}{n})}{\Delta t \cdot (1 + \frac{m-1}{n})} = \frac{TP_{MAX}}{(1 + \frac{m-1}{n})}$$



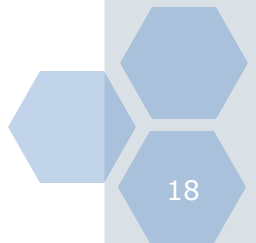
④当  $m \ll n$  时，实际吞吐率接近最大吞吐率



## 2、加速比（Speedup Ratio）

- ①流水线的加速比：指采用流水线方式后的工作速度与等效的串行方式的工作速度之比。
- ②对于  $n$  个任务而言，若采用串行方式工作所需的时间为  $T_1$ ，采用  $m$  段流水线方式工作所需的时间为  $T_2$ ，则加速比为：

$$S_p = \frac{T_1}{T_2}$$



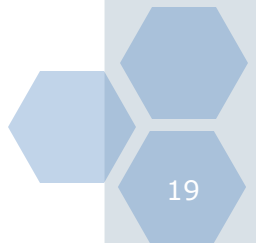


## 2、加速比（Speedup Ratio）

- ③如果  $m$  段流水线的各段时间均为  $\Delta t$ ，则完成  $n$  条指令在  $m$  段流水线上共需  $T = m \cdot \Delta t + (n - 1) \cdot \Delta t$  时间。在等效的顺序执行方式下所需时间为  $T = n \cdot m \cdot \Delta t$ 。加速比  $S_p$  为：

$$S_p = \frac{T_1}{T_2} = \frac{n \cdot m \cdot \Delta t}{m \cdot \Delta t + (n - 1) \cdot \Delta t} = \frac{n \cdot m}{m + n - 1} = \frac{m}{1 + \frac{m - 1}{n}}$$

- ④在  $n \gg m$  时， $S_p$  接近于  $m$ ，即当流水线各段时间相等时，其加速比的极限值等于流水线的段数。要获得高的加速比，应加大流水线的深度。





### 3、效率（Efficiency）

①流水线的效率指：流水线中各功能段（或设备）的利用率。是流水线各段的实际工作时间之和，与流水线各段被占用时间（从第一个任务流入到最后一个任务流出）之和的比值。

②衡量流水线效率的公式： $E = \frac{n \text{个任务占用的时空区}}{m \text{个功能段总的时空区}}$

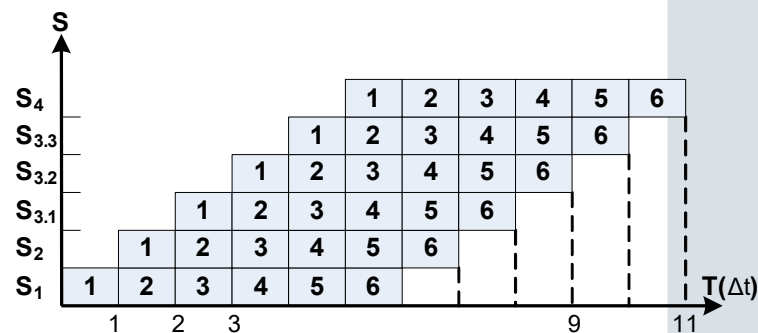
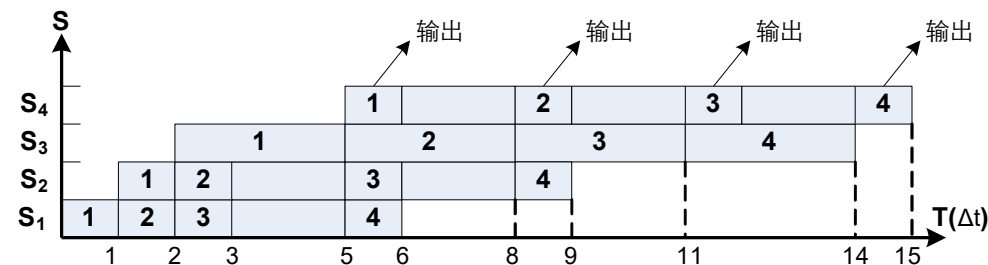
③在各段时间均等于 $\Delta t$ 的 $m$ 段流水线上连续地完成 $n$ 条指令，任务所占用的时空图面积为 $m \cdot n \cdot \Delta t$ ，流水线中各段总的时空区面积是 $m \cdot (m + n - 1) \cdot \Delta t$ ，那么流水线的效率为： $E = \frac{m \cdot n \cdot \Delta t}{m \cdot (m + n - 1) \cdot \Delta t} = \frac{n}{m + n - 1}$

当 $n \gg m$ 时， $S_p \approx m$ 。对比上两式，可知： $E = T$



# 举例：

1. 对照图 10.7 中的流水线时空图，假设  $\Delta t=80\text{ns}$ ，要处理的任务数（指令条数）为 20，要求计算图（a）和图（b）对应的流水线性能指标：**最大吞吐率、实际吞吐率、加速比、效率。**



（a）带有瓶颈的 4 段流水线时空图） 细分瓶颈段后的流水线时空图

解：

## ① 最大吞吐率

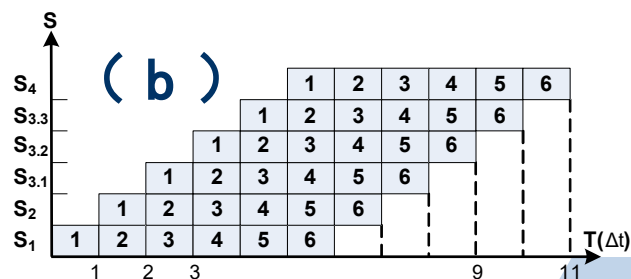
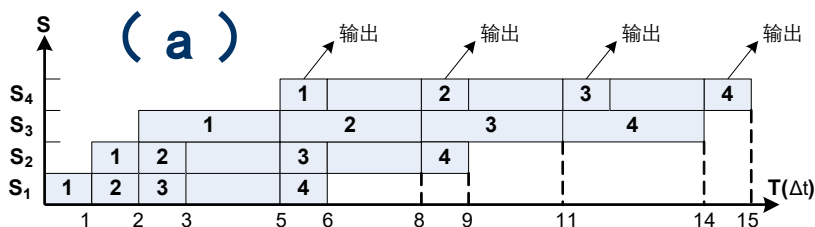
I. (a) 图流水线的最大吞吐率为： $TP_{MAX\ a} =$

$$1 / (3 \times 80ns) = 1 / (240ns)$$

$$\approx 4.17 \times 10^6 \text{ 条指令 / 秒}$$

II. (b) 图流水线的最大吞吐率为：

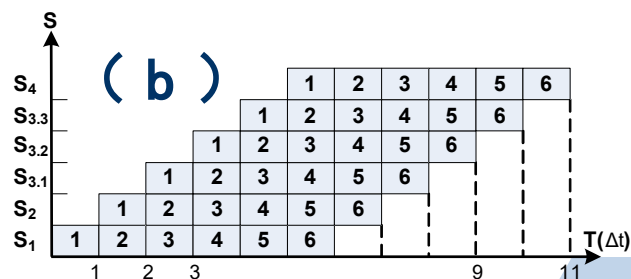
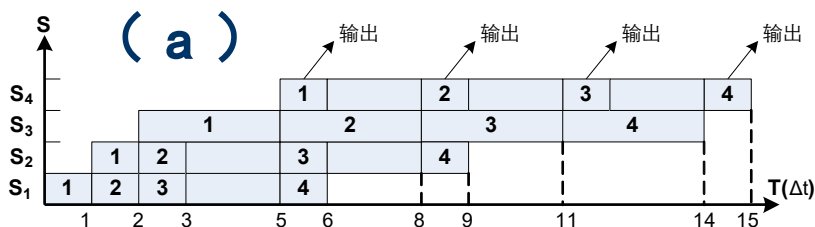
$$TP_{MAX\ b} = 1 / \Delta t = 1 / 80ns = 1.25 \times 10^7 \text{ 条指令 / 秒}$$



解：

## ② 实际吞吐率

- (a) 图流水线的实际吞吐率为： $TP_a = 20 / ((2 + 3 \times 20 + 1) \times 80ns) = 20 / (63 \times 80ns) \approx 3.97 \times 10^6$  条指令 / 秒
- (b) 图流水线的实际吞吐率为： $TP_b = 20 / ((6 + 20 - 1) \times 80ns) = 1/100ns = 10^7$  条指令 / 秒



解：

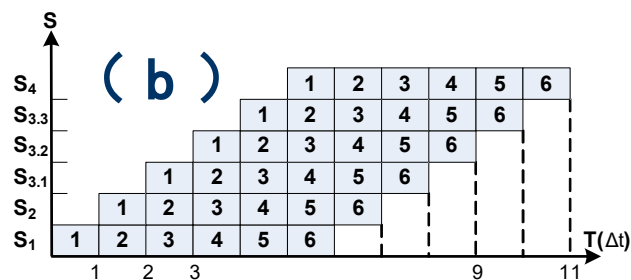
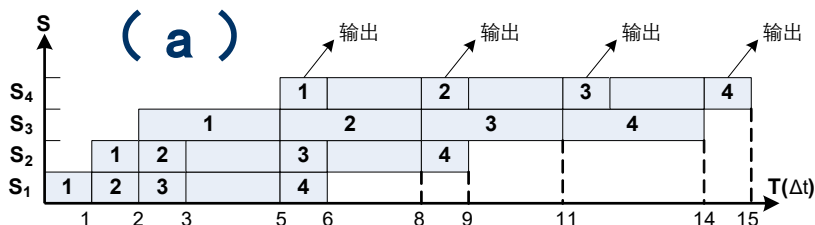
### ③ 加速比

- (a) 图流水线的加速比为： $SP_a$   

$$= 20 \times 6 \times 80\text{ns} / ( (2 + 3 \times 20 + 1) \times 80\text{ns} )$$

$$= 120 / 63 \approx 1.90$$
- (b) 图流水线的加速比为：  

$$SP_b = 20 \times 6 \times 80\text{ns} / ( (6 + 20 - 1) \times 80\text{ns} ) = 120 / 25 = 4.8$$





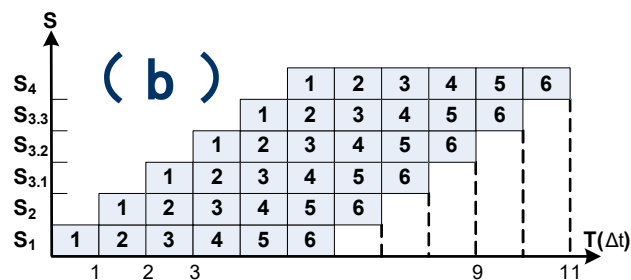
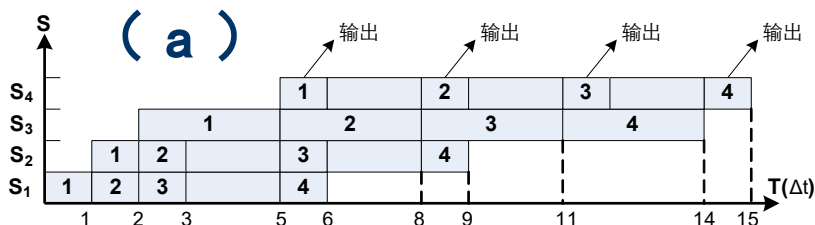
解：

#### ④ 效率

- (a) 图流水线的效率为：  $E_a$   

$$= 20 \times 6 \times 80\text{ns} / 4 \times ( (2 + 3 \times 20 + 1) \times 80\text{ns} ) = 120 / (4 \times 63) \approx 47.6\%$$
- (b) 图流水线的效率为：  

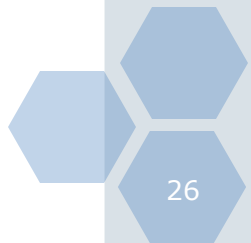
$$E_b = 20 \times 6 \times 80\text{ns} / 6 \times ( (6 + 20 - 1) \times 80\text{ns} ) = 120 / (6 \times 25) = 80\%$$





## 举例：

2. 假设指令流水线分指令预取（IF）、指令译码（ID）、地址生成（DA）、执行（EX）、回写（WB）5个过程段，每个过程段的延迟时间相同，共有10条指令连续输入此流水线。要求：
- ① 画出流水线的连接图。
  - ② 画出流水线时空图。
  - ③ 假设流水线的时钟周期为  $100\text{ns}$ ，求流水线的最大吞吐率和实际吞吐率。
  - ④ 求该流水处理器的加速比。
  - ⑤ 求该流水处理器的效率。



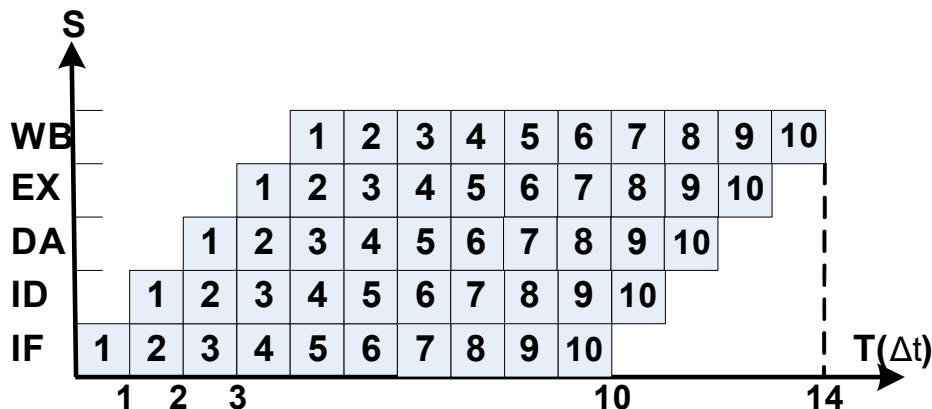


解：

### ① 指令流水线的连接图



### ② 10 条指令流入流水线的时空图

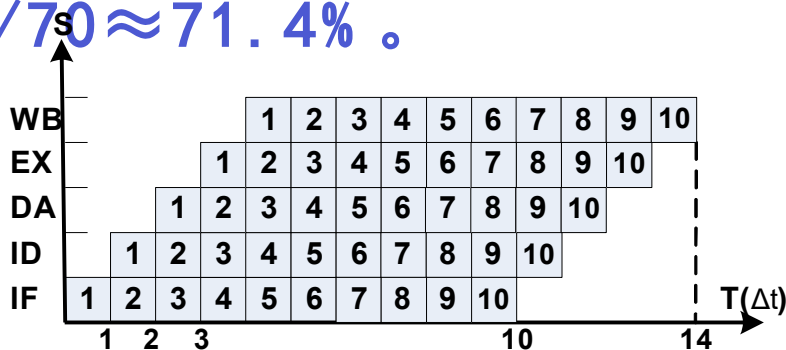


### ③ 最大吞吐率 $TP_{MAX}=1/100ns=10^7$ 条指令 / 秒

在 14 个时钟周期结束时，CPU 执行完 10 条指令，流水线的实际吞吐率为  $TP=10/(14 \times 100ns) = 10/1400ns \approx 7.14 \times 10^6$  条指令 / 秒

解：

- ④ 该流水处理器处理 10 条指令所需的时钟周期数为  $5 + (10 - 1) = 14$ ，串行顺序执行 10 条指令所需时钟周期数为  $10 \times 5 = 50$ ，所以流水处理器的加速比  $SP = 10 \times 5 \times 100\text{ns} / ((5 + 10 - 1) \times 100\text{ns}) = 50 / 14 \approx 3.57$
- ⑤ 流水处理器执行 10 条指令，各部件所需要的实际执行时间为  $10 \times 5 \times 100\text{ns}$ ，各部件被占用的时间为  $5 \times (5 + 10 - 1) \times 100\text{ns}$ ，所以流水处理器的效率  $E = 50 / 70 \approx 71.4\%$ 。





**The End !**

