



# 《计组II》

## 第六章——流水线

---

李瑞 副教授

蒋志平 讲师

计算机科学与技术学院



# 自我介绍

蒋志平，讲师

工学博士，

毕业于西安交通大学电子与信息工程学院；

2018年初加入西电软件学院；

18年获得我校讲课(实验与系统)竞赛二等奖 (3/1xx)

18年获得《中国博士后创新人才计划》资助 (8/20)

18年获得《自然科学基金——青年基金》资助 (~20%)

研究方向：

基于认知无线电、可穿戴设备、视觉等技术的智能感知、隐私安全保护与智能交互等；



# 第6章 流水线技术

## 6.1 概述

### 6.1.1 流水线概述

#### 1、什么是流水线？

将一**重复**的**处理过程****分解**为若干**子过程**，每个子过程都可有效地在其专用功能段上**与其它子过程同时执行**，这种技术称为**流水线技术**。

## 洗衣房

- 张三、李四、王五、赵六每人有一包衣服需要洗涤、烘干、熨整



- 洗衣机需要 30分钟



- 烘干机需要30分钟



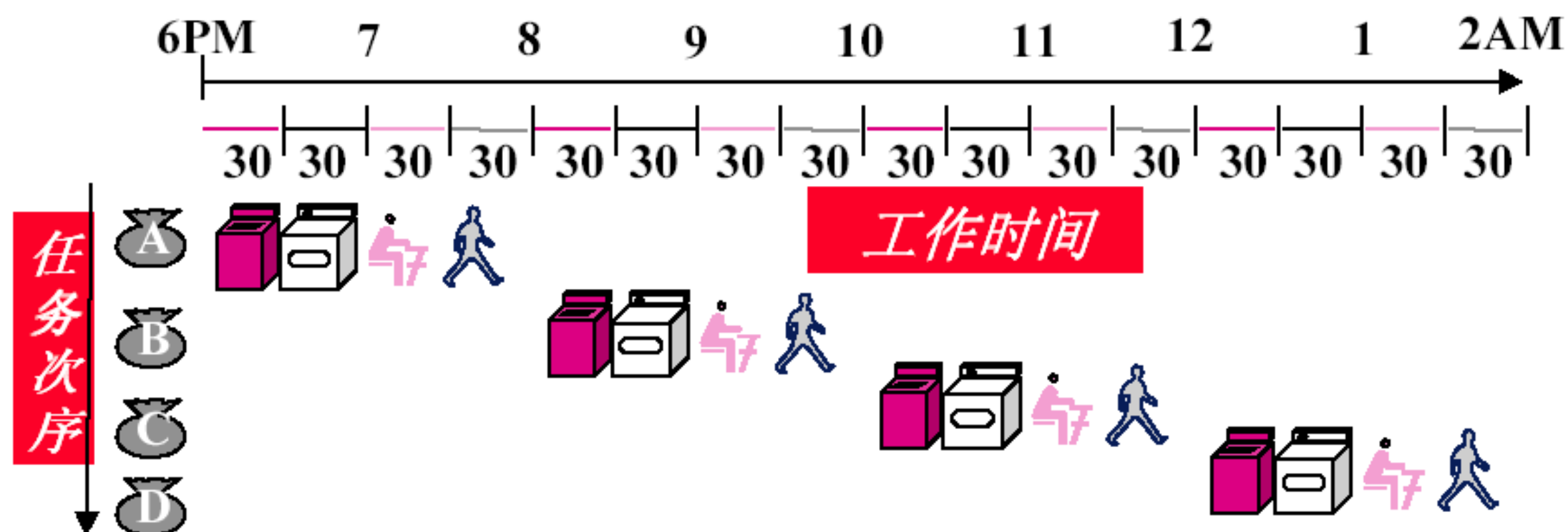
- 熨斗需要30分钟



- 洗衣工需要30分钟将衣物放到抽屉里



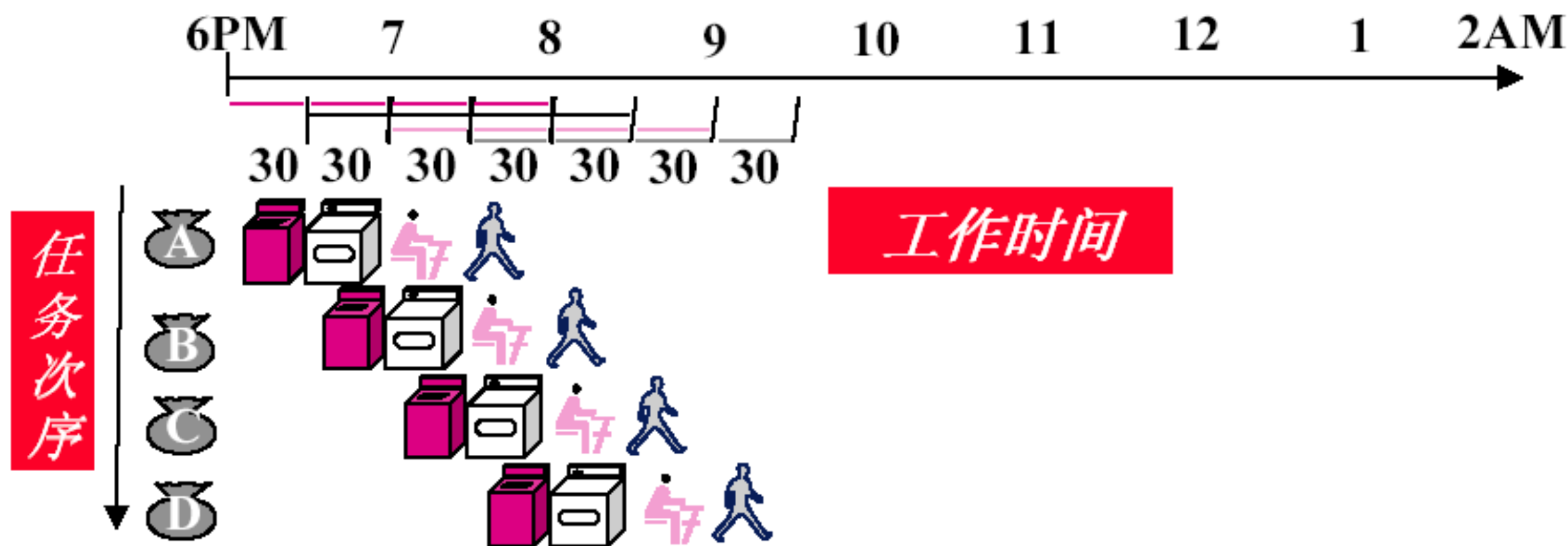
## 串行洗衣店



- 串行洗衣店需要8个小时完成4个工作量
- 如果他们了解流水技术，那么需要多长时间完成上述工作呢？

## 基本思想：流水举例

### 流水化的洗衣店：尽可能早地开始工作



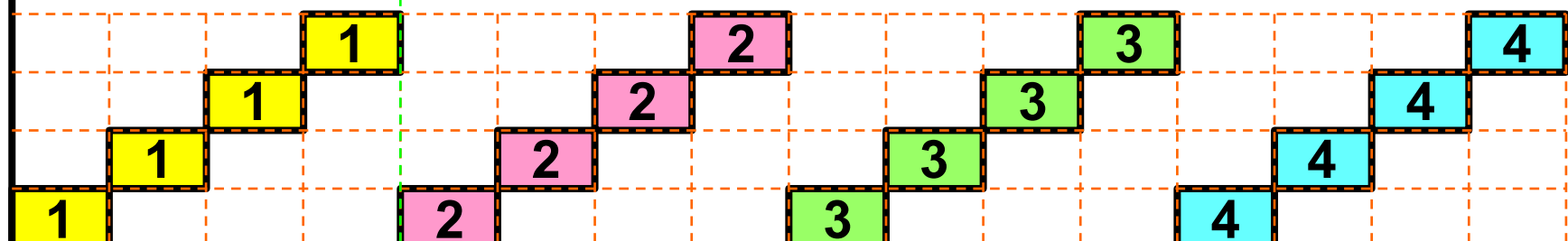
- 流水化洗衣店需要**3.5**个小时完成**4**个工作量

# 基本思想：流水举例

1: 张三  
2: 李四  
3: 王五  
4: 赵六

空间 (段号)

存放  
熨整  
烘干  
洗涤

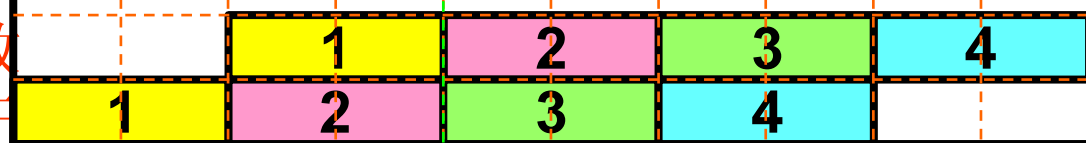


16 时间 (拍)

(a) 顺序方式

空间 (段号)

熨整、存放  
洗涤、烘干



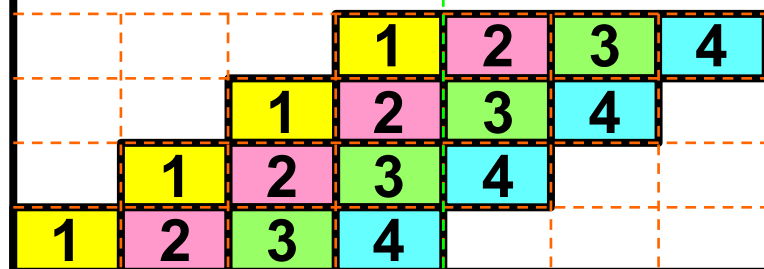
10

时间 (拍)

(b) 重叠方式

空间 (段号)

存放  
熨整  
烘干  
洗涤



$\Delta t$

7

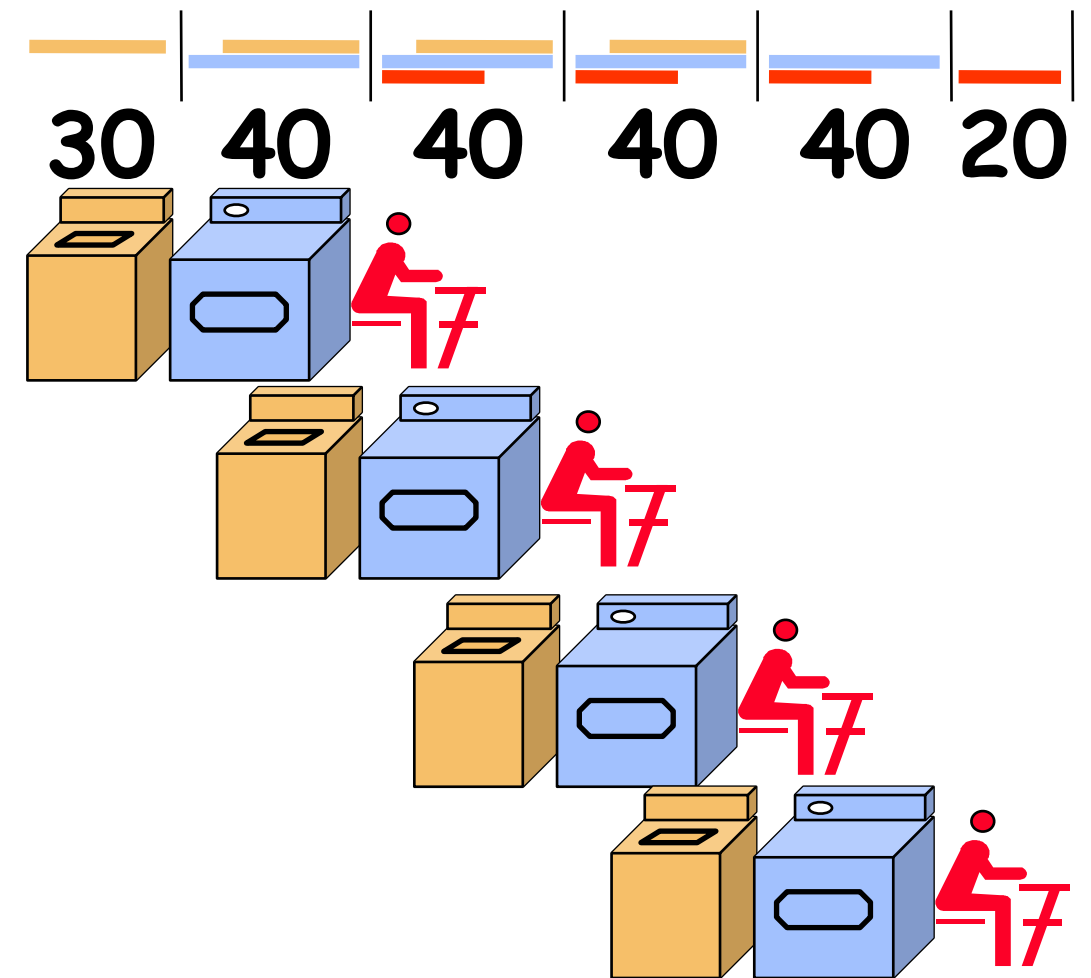
时间 (拍)

(c) 流水方式



### (3) 洗衣店的结论

- 流水线不能缩短单个任务的响应时间，但可以提高吞吐率；
- 流水线速度限制于最慢流水站的速度；
- 流水线中多个任务是并行处理的；
- 最大加速比 = 流水站数
  - 流水站速度不匹配
  - 流水线“填充”和“排空”时间





# 并行处理技术

## ■ 并行性的两种含义：

- **同时**性：同一时刻
- **并发**性：同一时间间隔

## ■ 并行处理技术的三种形式：

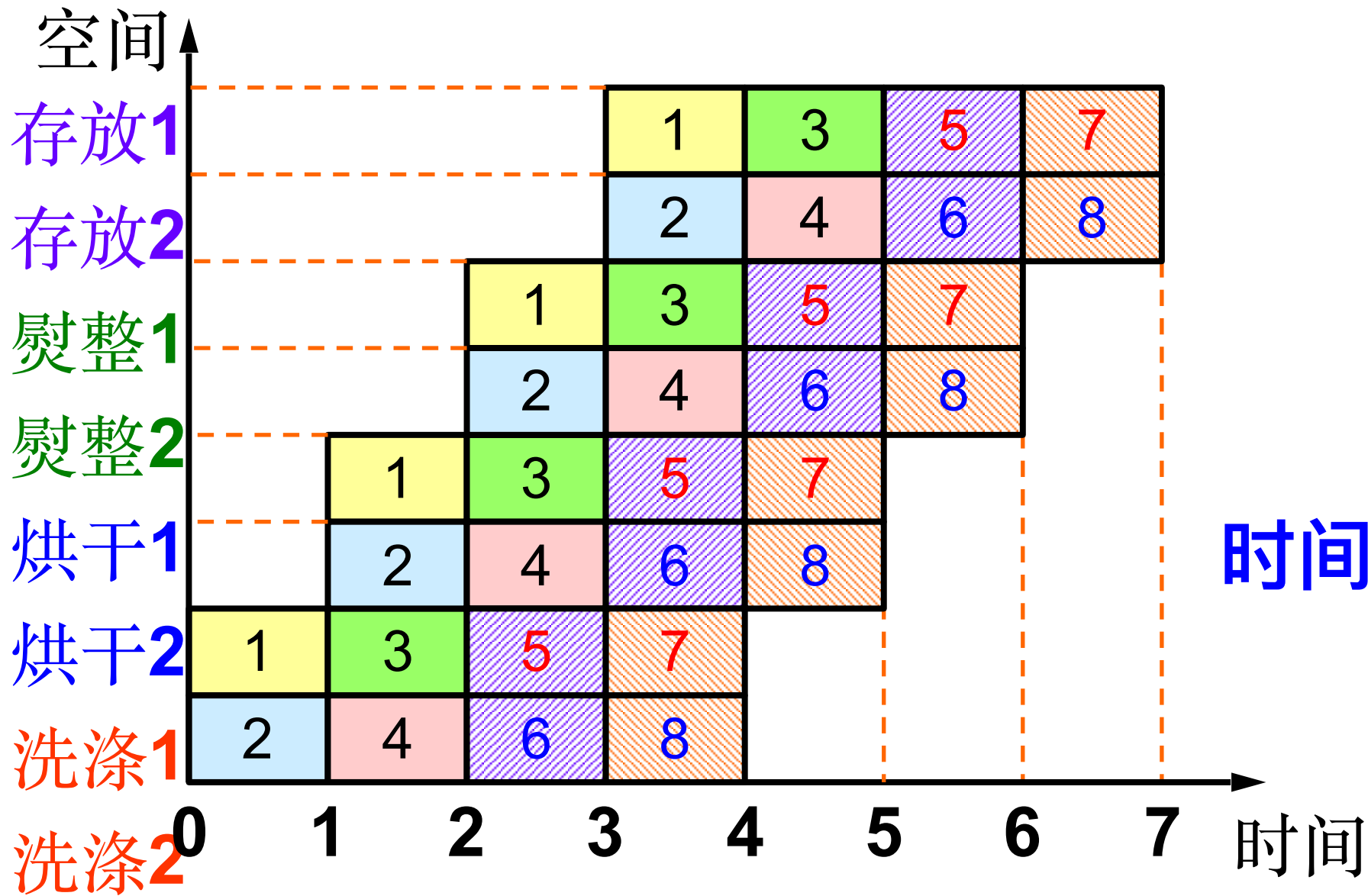
- 时间并行：**时间重叠**，流水
- 空间并行：**资源重复**
- 时间并行 + 空间并行：**超标量**流水

基本思想：流水举例



时间并行

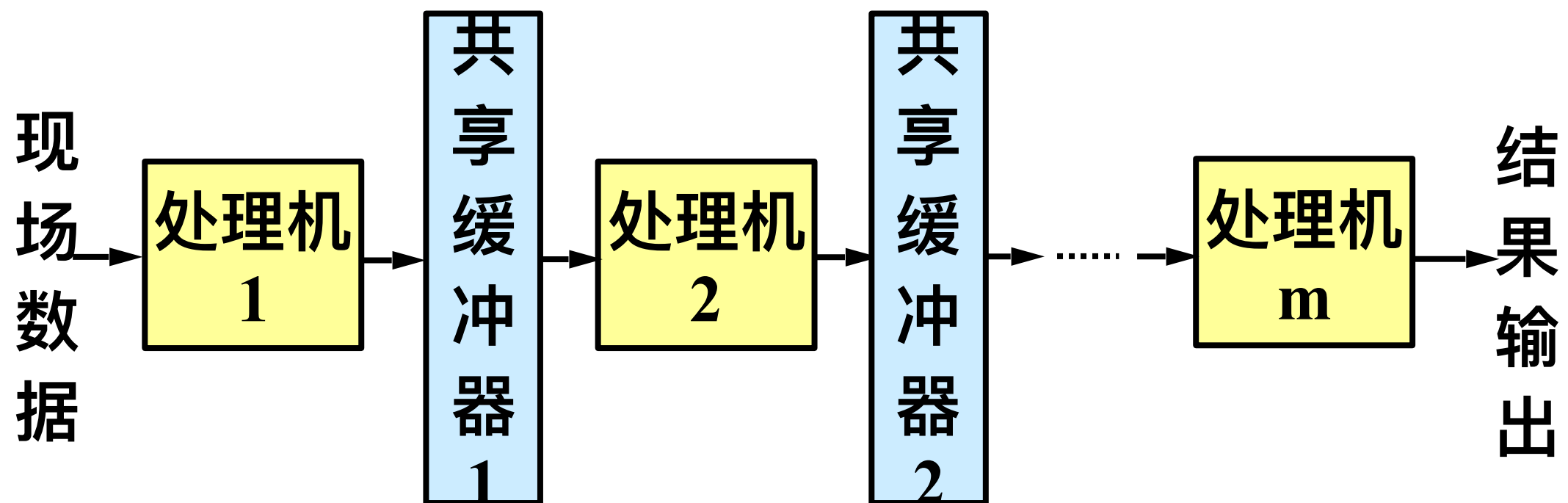
- 1: 张三
- 2: 李四
- 3: 王五
- 4: 赵六



时间并行 + 空间并行

## 2. 计算机中的流水线

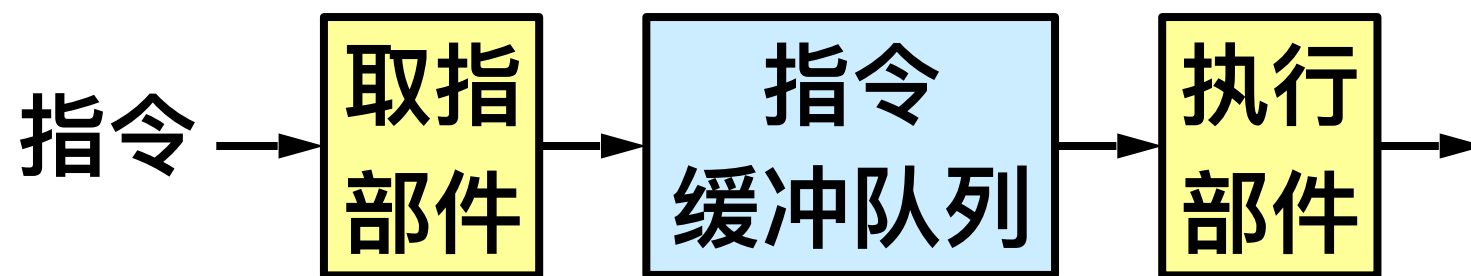
- 1. **系统级流水线/宏流水线**：在多（计算）机系统中由多个处理机串行构成的流水线。



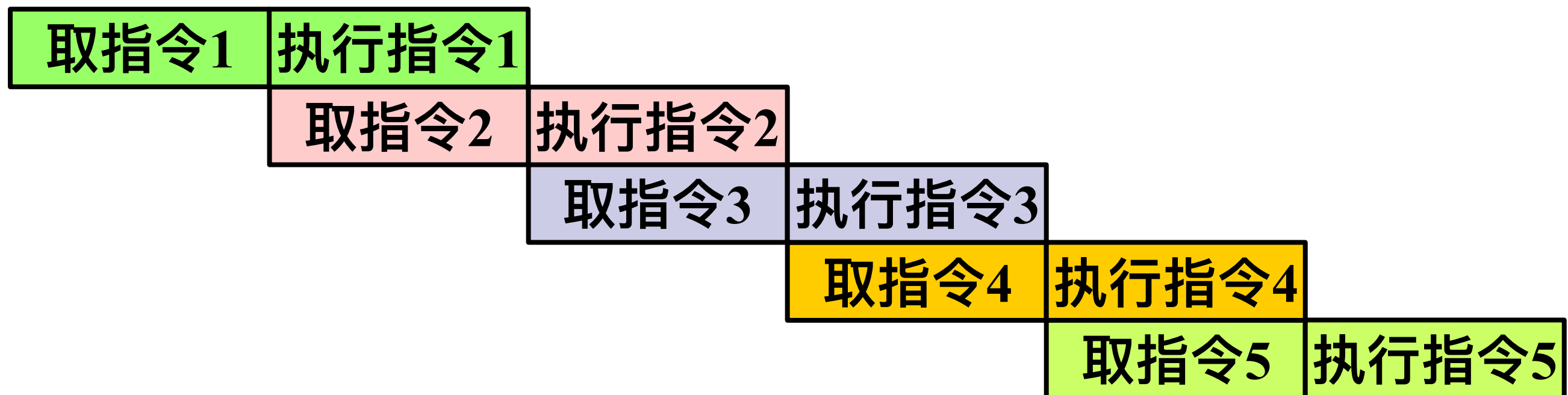
- 2. **处理器级流水线**

## 2. 计算机中的流水线

- 1. 系统级流水线/宏流水线：在多（计算）机系统中由多个处理机串行构成的流水线。
- 2. 处理器级流水线



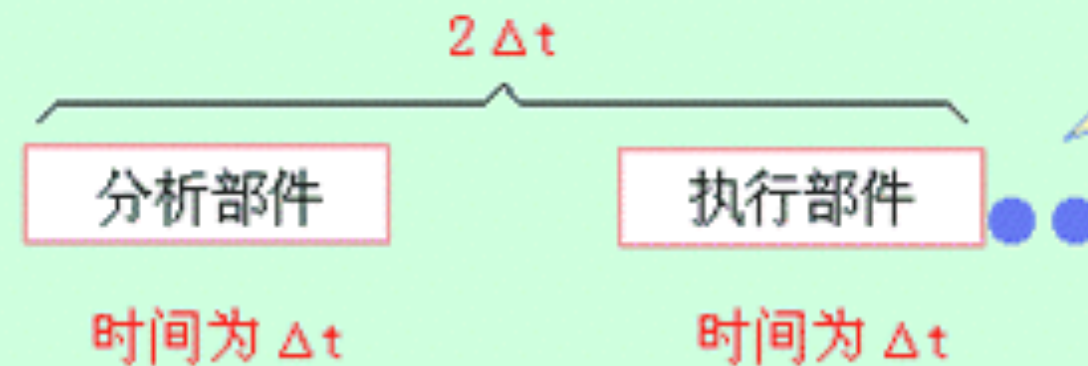
Intel 8086指令流水线



## 指令流水线



一条指令的解释过程



处理机的速度  
提高了一倍!

当分析部件完成上一条指令的“分析”后，就立即将之送入执行部件，同时分析部件可以开始处理下一条指令。

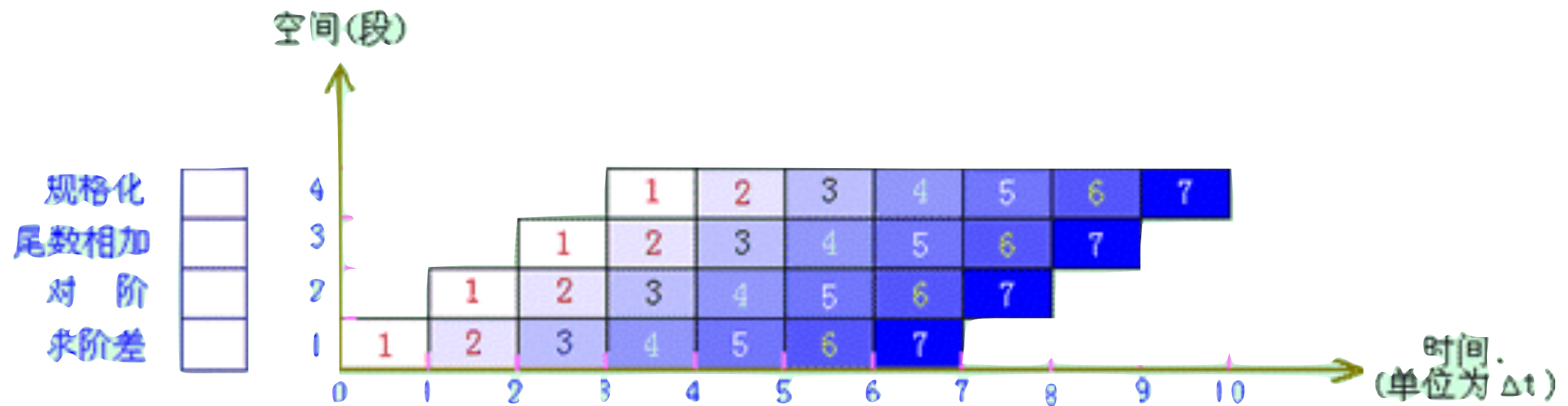
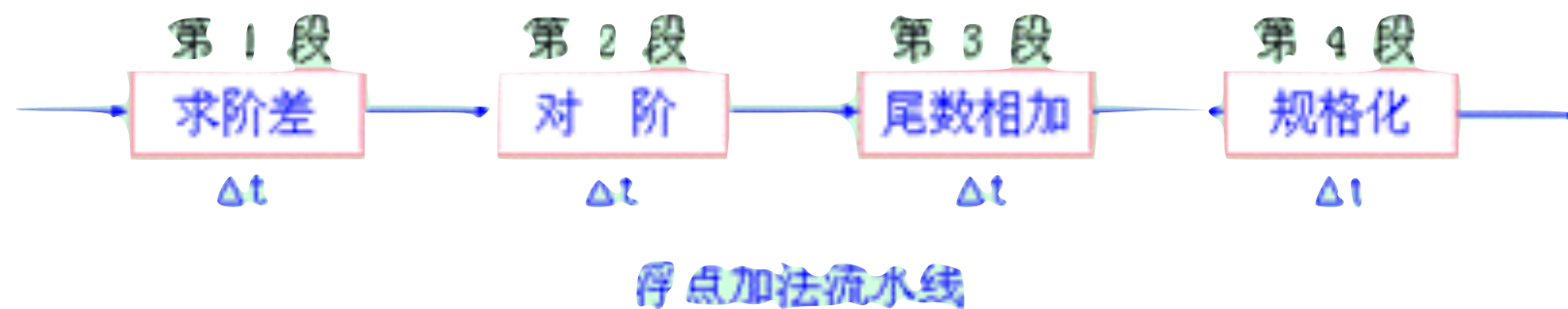
虽然从执行一条指令的全过程来看，仍需要  $2\Delta t$  的时间，但从机器的输出端来看，却是每隔一个  $\Delta t$  就能给出一条指令的执行结果。

## ➤ 流水线的特点

- ❑ 流水过程由多个相关的子过程组成，这些子过程称为流水线的“级”或“段”（Stage）。段的数目称为流水线的“深度”。
- ❑ 每个子过程由专用的功能段实现，各功能段的时间应基本相等，通常为1个时钟周期（1拍）。
- ❑ 流水线需要经过一定的通过时间才能稳定。
- ❑ 流水技术适合于大量重复的时序过程。

### 3. 时-空图

从时间和空间两个方面描述流水线的工作过程，横坐标表示时间，纵坐标表示各流水段。





## 6.1.1 流水线的分类

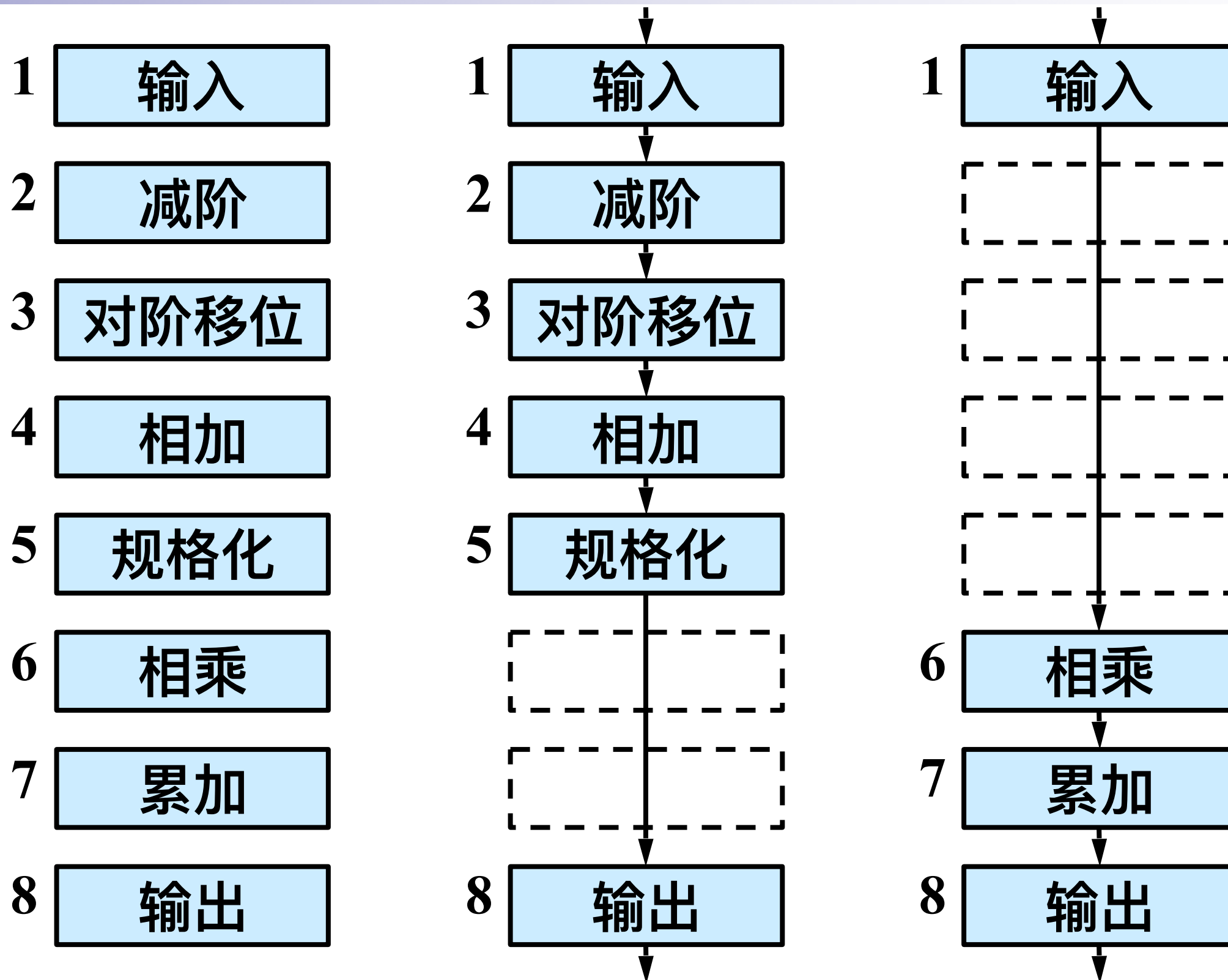
流水线可以按不同的观点进行分类。

### 1. 单功能 v.s 多功能

- **单功能流水线**，是指只能完成一种固定功能的流水线。
- **多功能流水线**，是指各段可以进行不同的连接，从而完成不同的功能。

# 7.1 流水线处理

## 二、流水线类型



(a) 流水线各功能段

(b) 浮点加、减运算

(c) 定点乘法运算

图7.5 TI ASC计算机的运算器流水线

## 2. 静态流水线和动态流水线

按同一时间内流水段的连接方式划分

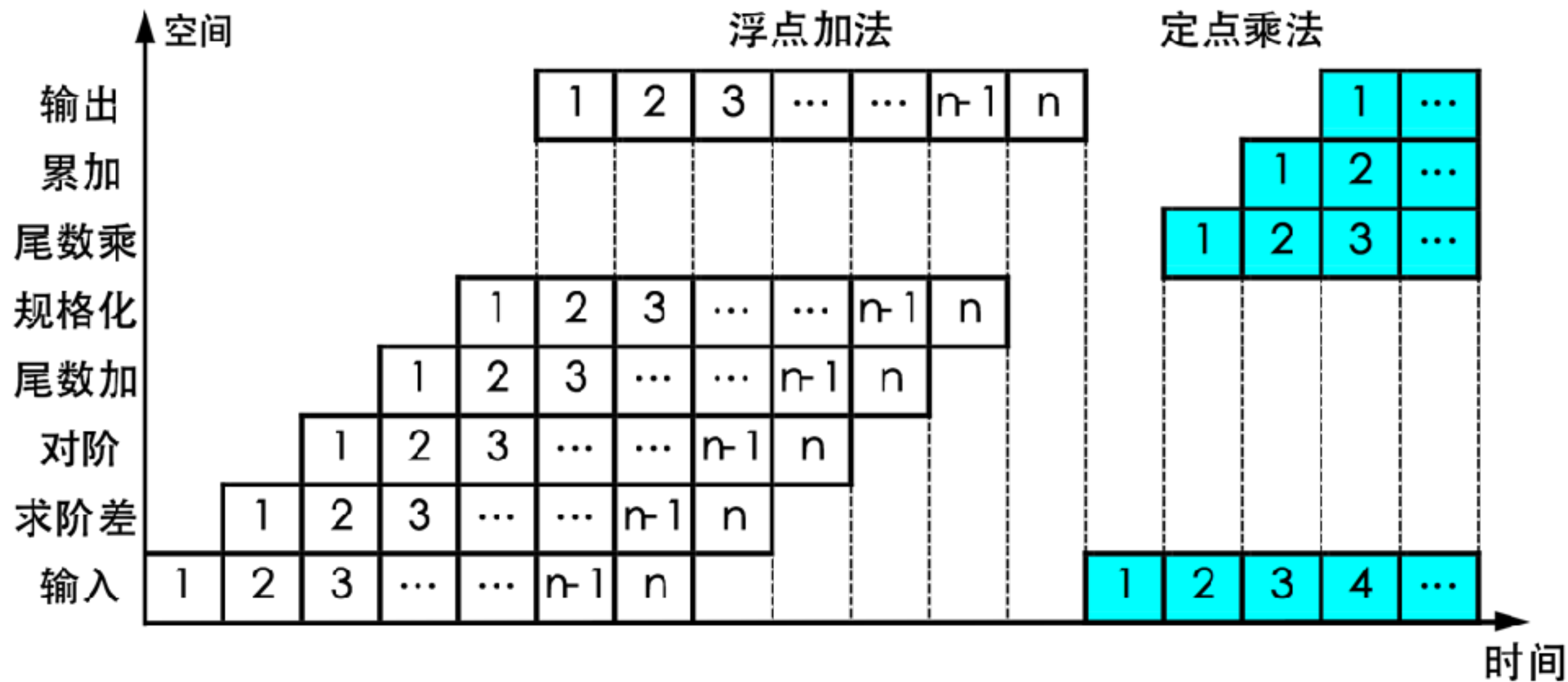
- **静态流水线**，是指在同一时间内，流水线的各段只能按同一种功能的连接方式工作。

例如：TI ASC的流水线

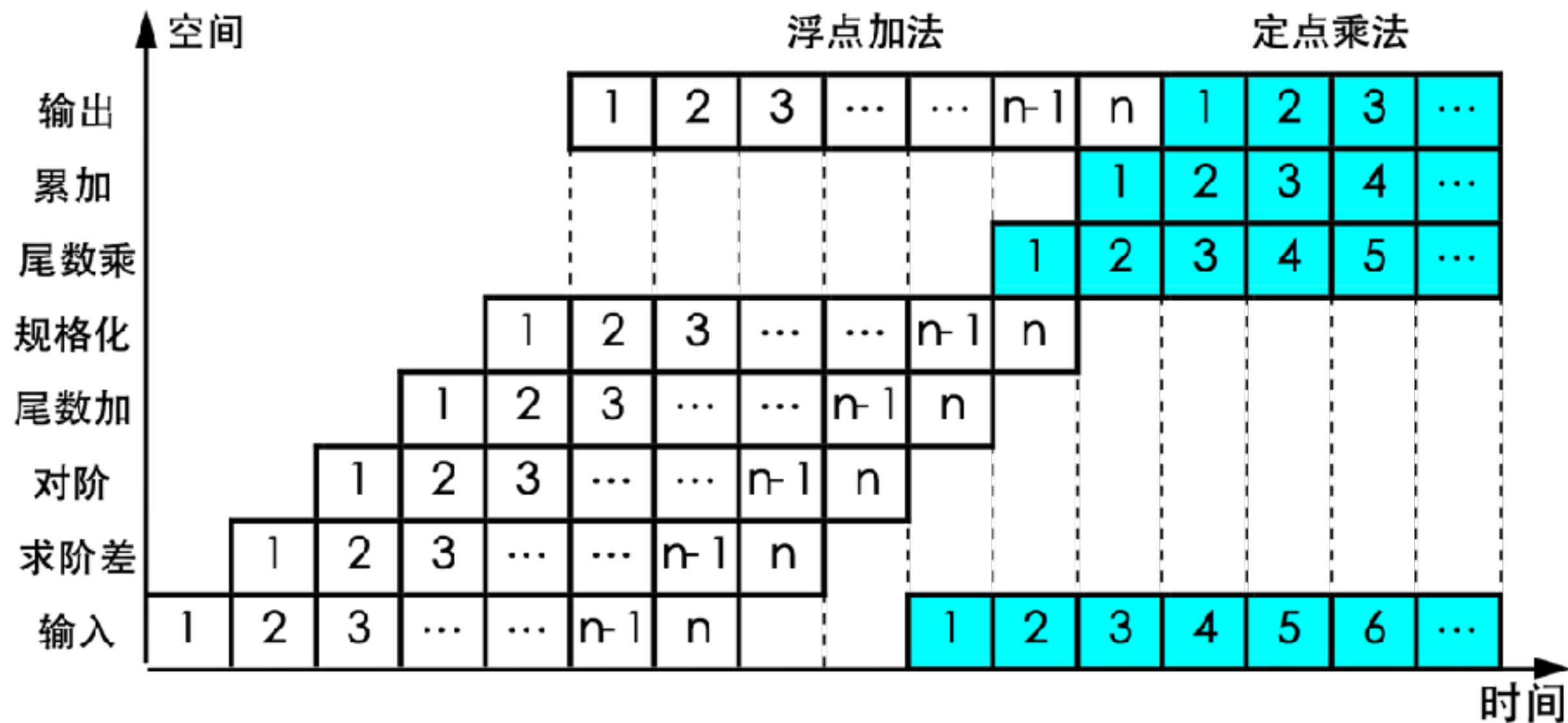
适合于处理一串相同的运算操作

- **动态流水线**，是指在同一时间内，当某些段正在实现某种运算时，另一些段却在实现另一种运算。

会使流水线的控制变得很复杂



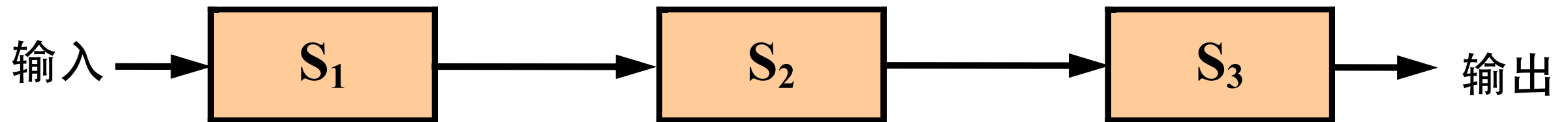
静态流水线时空图



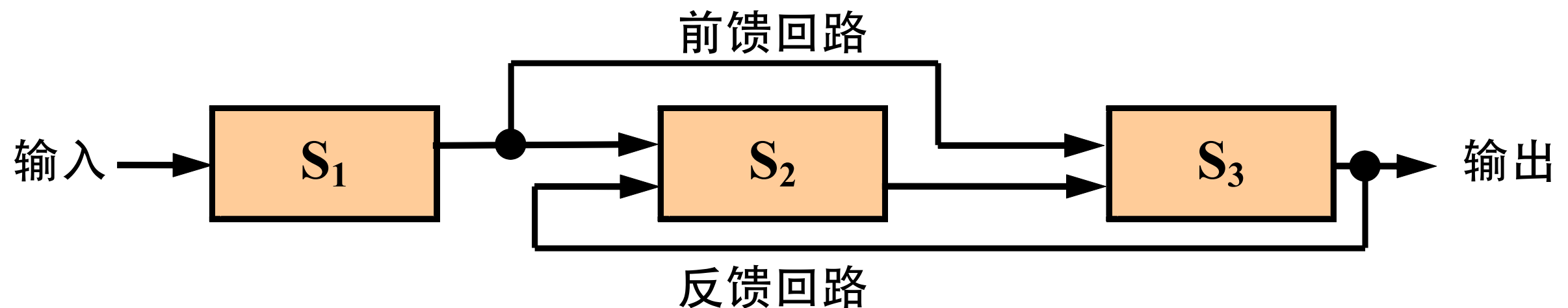
动态流水线时空图

■ 按流水线是否有反馈回路划分：

- 线性流水线
- 非线性流水线－流水线调度



一种简单的线性流水线

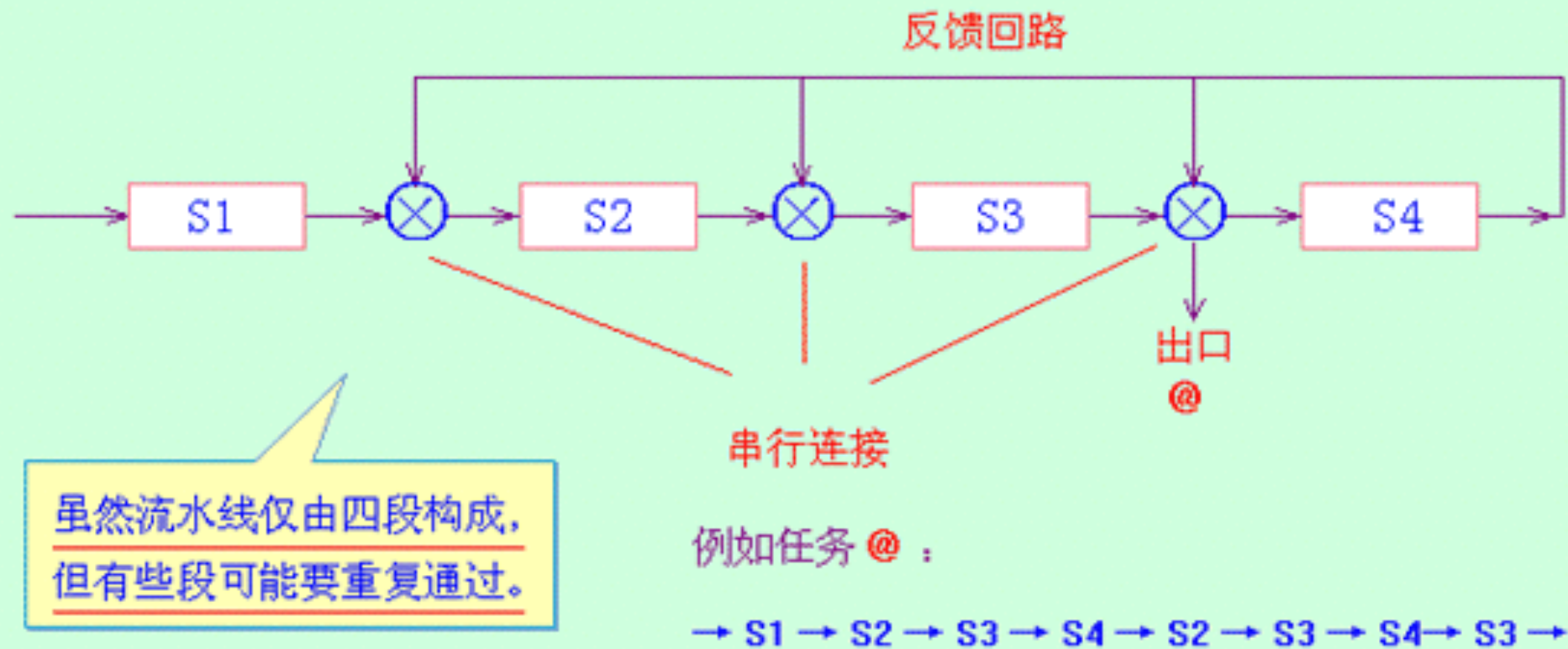


一种简单的非线性流水线

- ◆ 线性流水线能够用流水线连接图唯一表示。
- ◆ 非线性流水线必须用流水线连接图和流水线预约表共同表示。

## 非线性流水线

(举例)



### ➤ 存在流水线调度问题:

确定什么时候向流水线引进新的输入, 从而使新输入的数据和先前操作的反馈数据在流水线中不产生冲突, 此即所谓流水线调度问题。



## 5. 顺序流动流水线和乱序流动流水线

按照输出端任务流出顺序与输入端任务流入顺序是否相同划分

- **顺序流动流水线**是指流水线输出端任务流出的顺序与输入端任务流入的顺序**相同**。
- **乱序流动流水线**也可称为**无序流水线**、**错序流水线**，是指流水线输出端任务流出的顺序与输入端任务流入的顺序不一定相同。

## 6.2 流水线的性能分析

### 三项性能指标：吞吐率、加速比和效率

#### 6.2.1 吞吐率

**吞吐率**是指单位时间内流水线所完成的任务数或输出结果的数量。

- ◆ **最大吞吐率**  $TP_{max}$  是指流水线在达到稳定状态后所得到的吞吐率。
- ◆ 设流水线由  $m$  段组成，完成  $n$  个任务的吞吐率称为**实际吞吐率**，记作  $TP$ 。

## 1、最大吞吐率

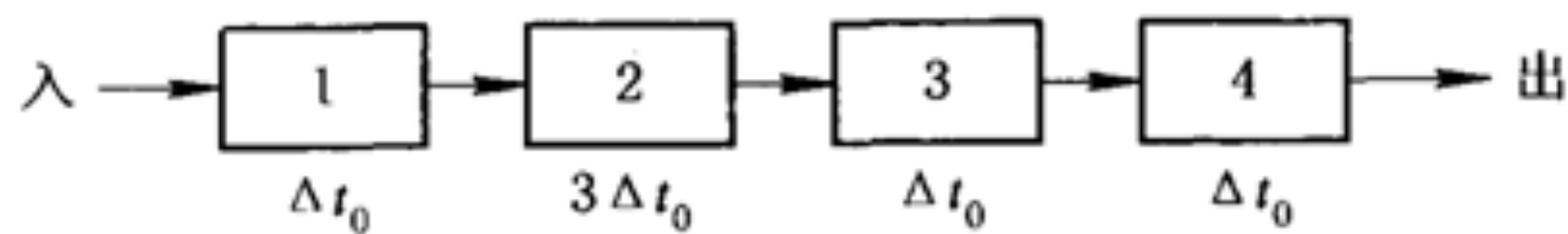
- ◆ 假设流水线各段的时间相等，均为  $\Delta t_0$ ，则：

$$TP_{max} = 1/\Delta t_0$$

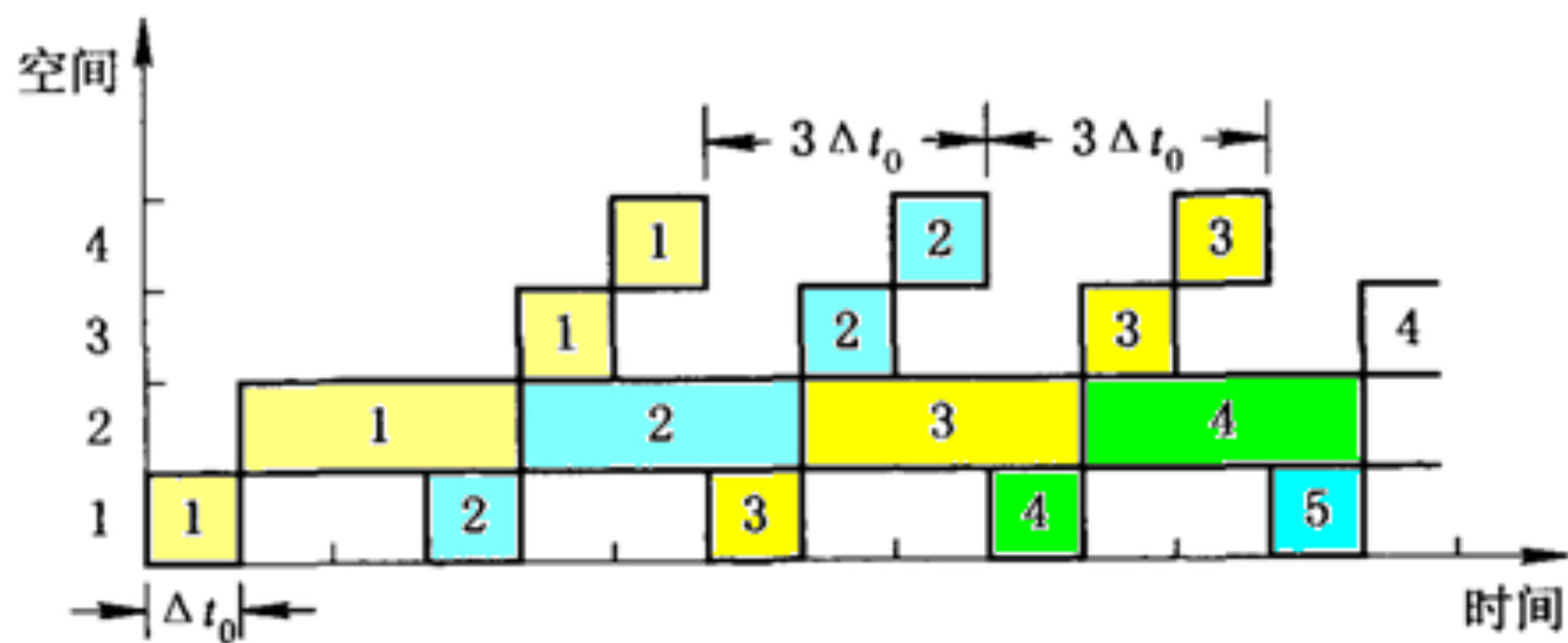
- ◆ 假设流水线各段时间不等，第i段时间为  $\Delta t_i$ ，则：

$$TP_{max} = 1/\max\{\Delta t_i\}$$

- 最大吞吐率取决于流水线中最慢一段所需的时间，该段成为流水线的瓶颈
- 消除瓶颈的方法
  - 细分瓶颈段
  - 重复设置瓶颈段

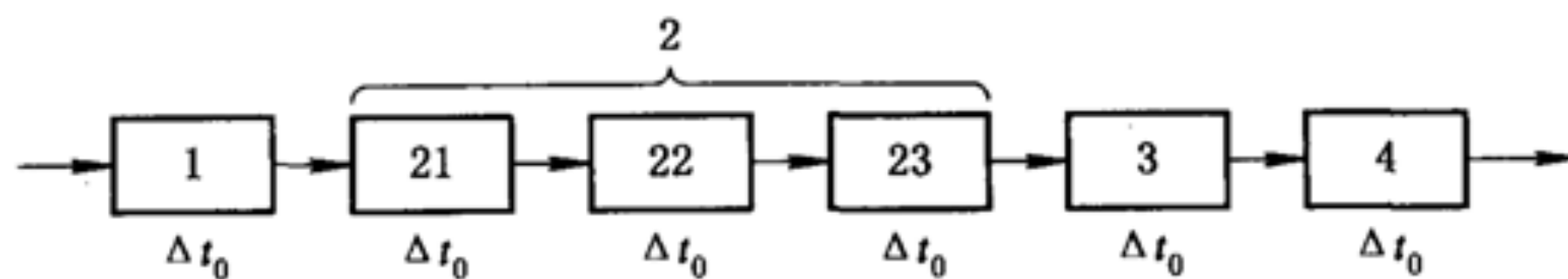


(a)

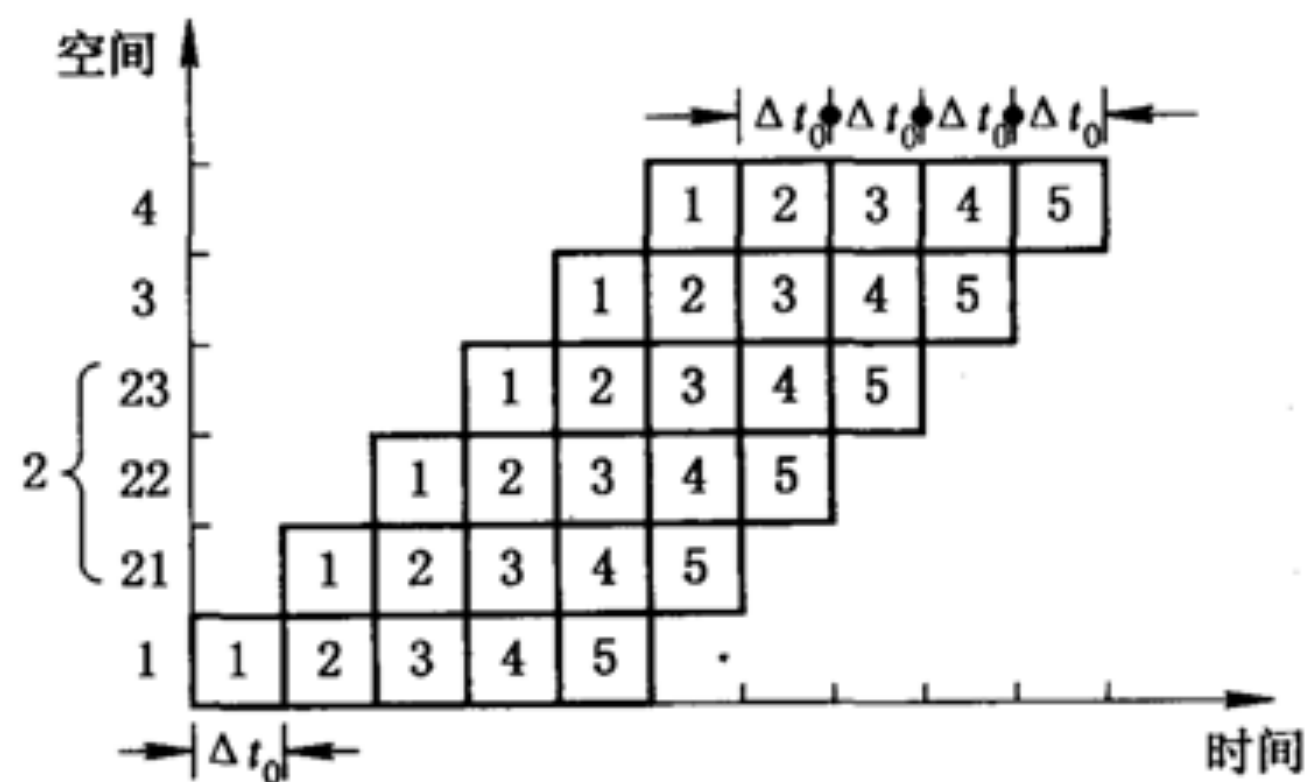


(b)

最大吞吐率取决于瓶颈段的时间

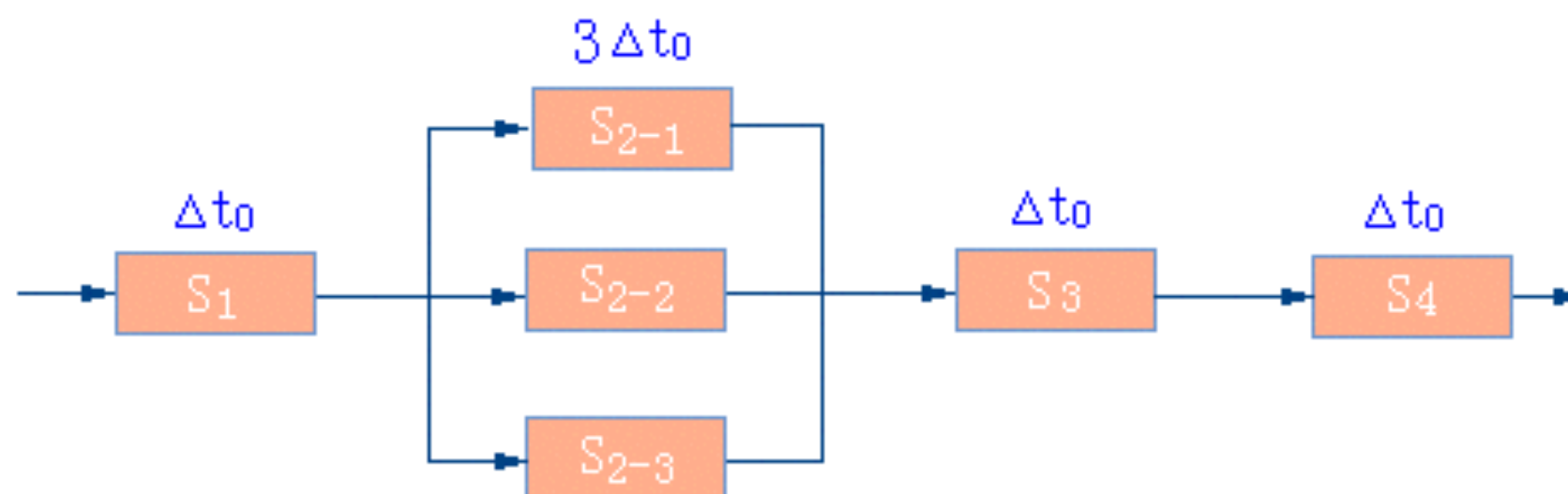


(a)

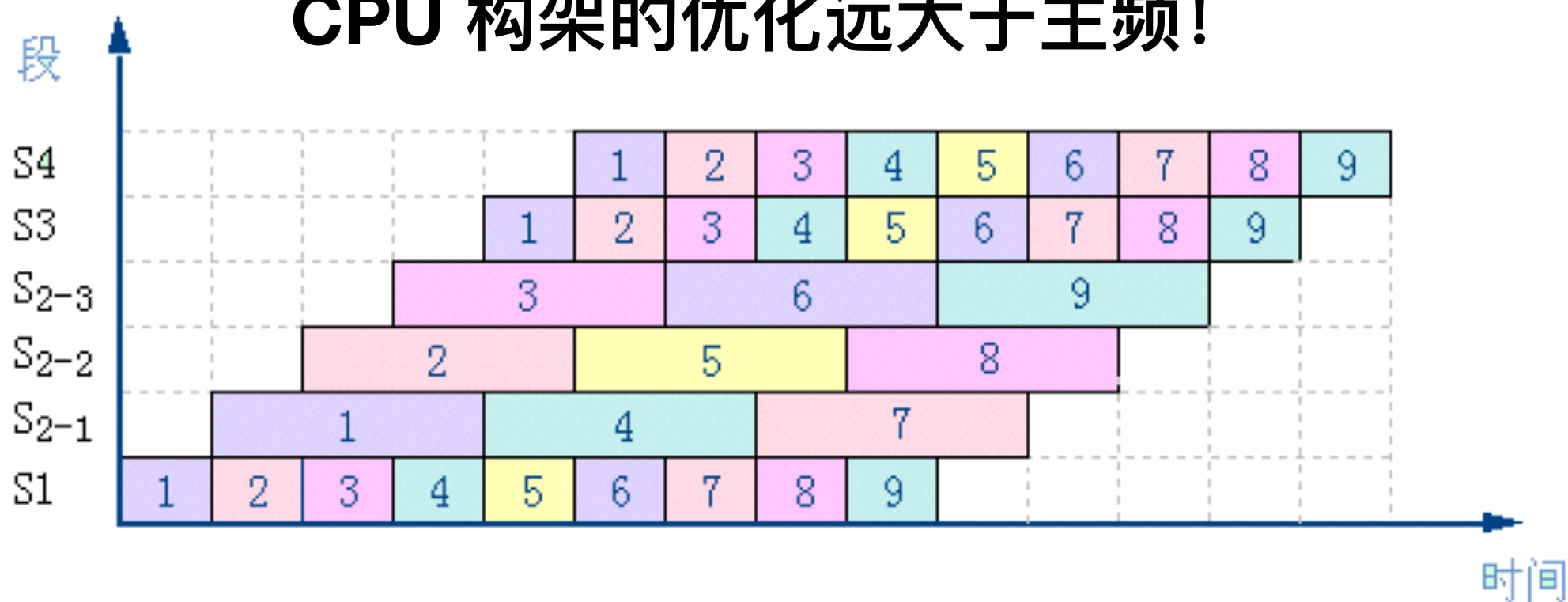


(b)

瓶颈子过程再细分



**CPU 构架的优化远大于主频!**



## 2、实际吞吐率

- ◆ 若各段时间相等（假设均为 $\Delta t_0$ ），则完成时间

$$T_{\text{流水}} = m \Delta t_0 + (n-1) \Delta t_0$$

实际吞吐率

$$\begin{aligned} TP &= \frac{n}{T_{\text{流水}}} = \frac{n}{m \cdot \Delta t_0 + (n-1) \Delta t_0} \\ &= \frac{1}{\Delta t_0 \left(1 + \frac{m-1}{n}\right)} = \frac{TP_{\max}}{1 + \frac{m-1}{n}} \end{aligned}$$

$$TP < TP_{\max}$$

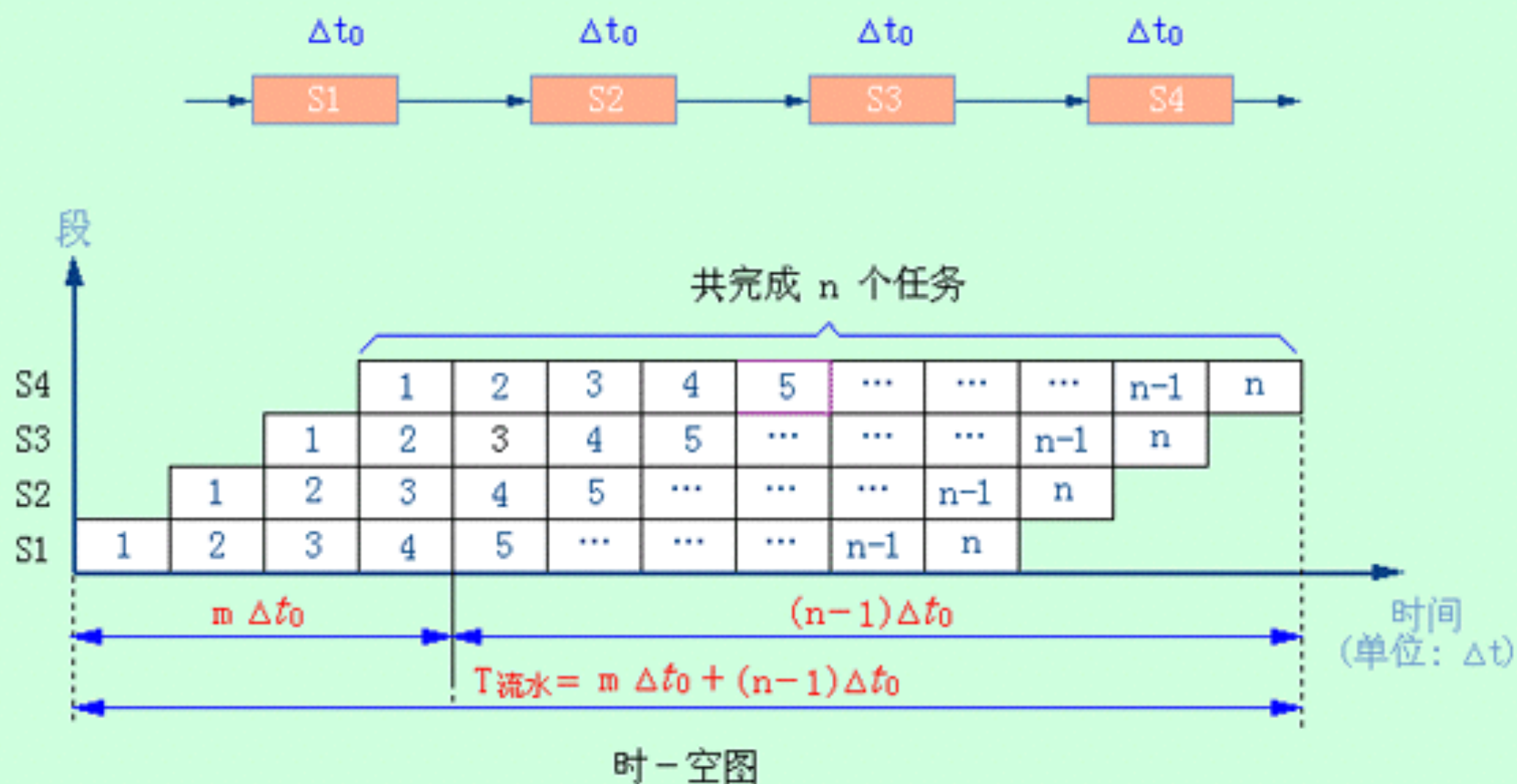
$$\text{当 } n \gg m \text{ 时, } TP \approx TP_{\max}$$



# 完成 $n$ 个任务所需的时间

## 流水线的时-空图

(各段时间相等)

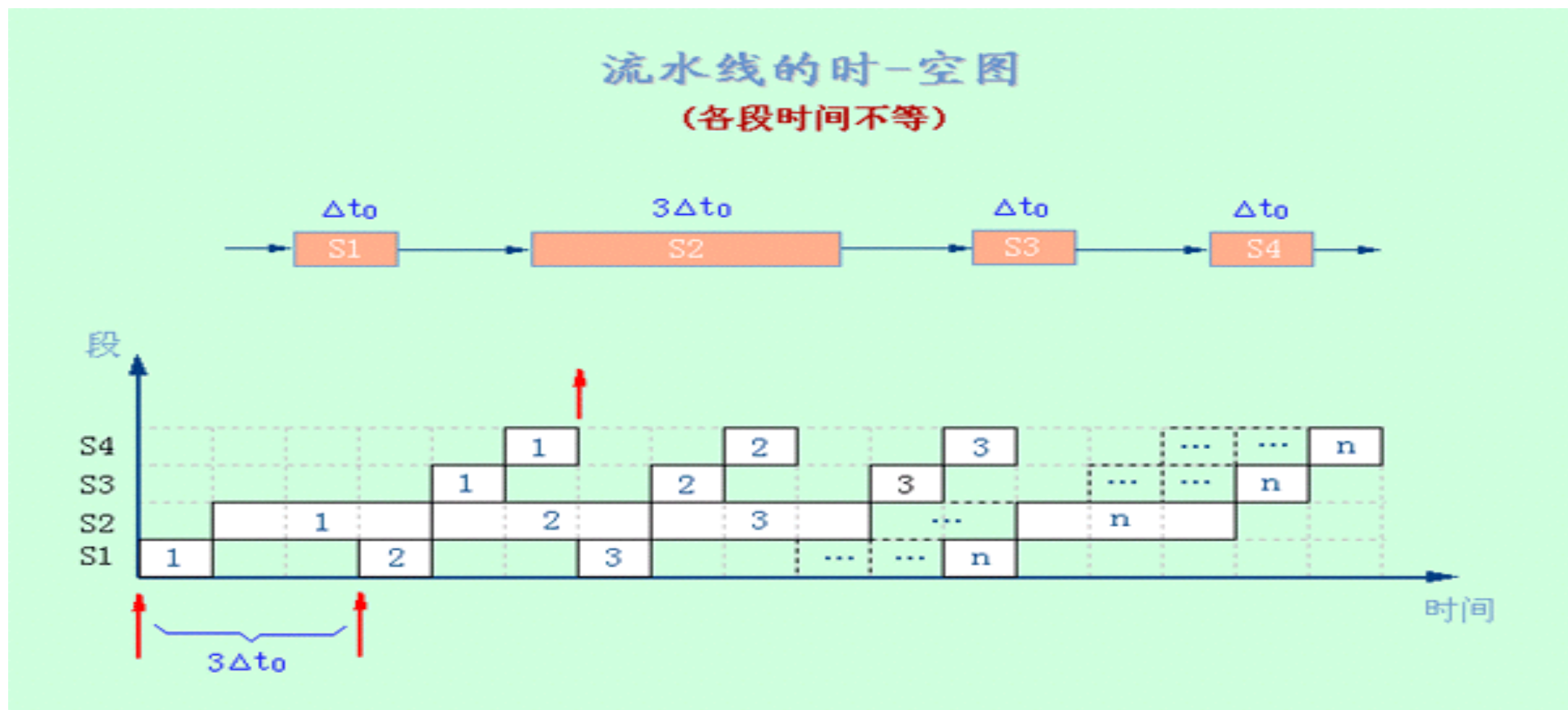


$$T_{\text{流水}} = m \Delta t_0 + (n-1) \Delta t_0$$

- 若各段时间不等（假设第*i*段为 $\Delta t_i$ ），则完成时间

$$T = \sum_{i=1}^m \Delta t_i + (n-1) \Delta t_j \text{ 这里, } \Delta t_j = \max\{\Delta t_i\}$$

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



## 6.2.2 加速比

**加速比**是指流水线速度与等功能的非流水线速度之比。

$$\text{加速比 } S = T_{\text{非流水}} / T_{\text{流水}}$$

- ◆ 若流水线为  $m$  段，每段时间均为  $\Delta t_0$

$$T_{\text{非流水}} = nm\Delta t_0, \quad T_{\text{流水}} = m\Delta t_0 + (n-1)\Delta t_0$$

$$S = \frac{mn}{m+n-1} = \frac{m}{1 + \frac{m-1}{n}}$$

可以看出：当  $n \gg m$  时， $S \approx m$

## · 各段时间不相等的流水线

当各段时间不相等的  $m$  段流水线, 顺序完成  $n$  个任务所用的时间  $T_{\text{顺}}$  和利用流水线完成  $n$  个任务所用时间  $T$  分别表示如下。

$$T_{\text{顺}} = n \sum_{i=1}^m \Delta t_i$$

$$T = \sum_{i=1}^m \Delta t_i + (n-1) \Delta t_{\max}$$

从而可以写出这种情况下的加速比如式所示

$$S = \frac{T_{\text{顺}}}{T} = \frac{n \sum_{i=1}^m \Delta t_i}{\sum_{i=1}^m \Delta t_i + (n-1) \Delta t_{\max}}$$

### 6.2.3 效率

效率指流水线的设备利用率。

由于流水线有 通过时间和排空时间，所以流水线的各段并非一直满负荷工作， $E < 1$

- ◆ 若各段时间相等，则各段效率也相等，

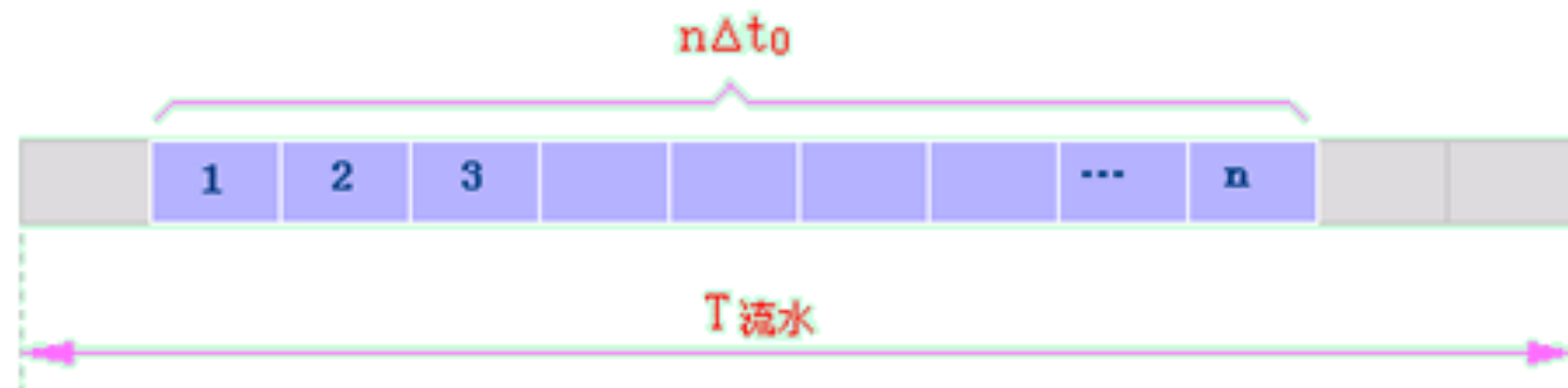
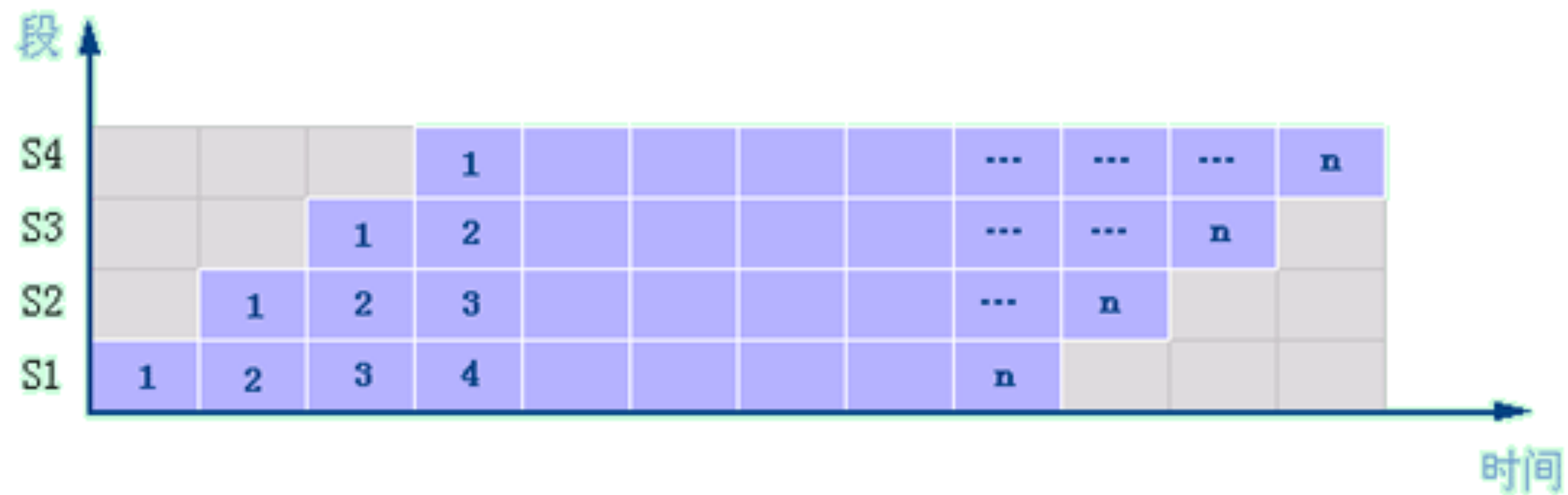
即  $e_1 = e_2 = e_3 = \dots = n \Delta t_0 / T_{\text{流水}}$

- ◆ 整个流水线效率

$$E = \frac{n \Delta t_0}{T_{\text{流水}}} = \frac{n}{m + n - 1} = \frac{1}{1 + \frac{m-1}{n}}$$

当  $n \gg m$  时， $E \approx 1$

## 流水段的效率



所以

$$e_2 = \frac{n\Delta t_0}{T_{流水}}$$

- ◆ 从时-空图上看，效率就是n个任务所占的时空区与m个段总的时空区之比
- ◆ 根据这个定义，可以计算流水线各段时间不等时的流水线效率

$$E = \frac{n \text{ 个任务占用的时空区}}{m \text{ 个段总的时空区}}$$



## 吞吐率、加速比和效率的关系

- ◆  $E = n \Delta t_0 / T_{\text{流水}} = mn \Delta t_0 / (T_{\text{流水}} m) = S/m$
- ◆ 效率是实际加速比S与最大加速比m之比。
- ◆  $E = n \Delta t_0 / T_{\text{流水}} = (n / T_{\text{流水}}) \cdot \Delta t_0 = TP \Delta t_0$

当 $\Delta t_0$ 不变时，流水线的效率与吞吐率呈正比。为提高效率而采取的措施，也有助于提高吞吐率。

- ◆  $TP = S / (m \Delta t_0)$

当m和 $\Delta t_0$ 不变时，流水线的加速比与吞吐率呈正比。为提高加速比而采取的措施，也有助于提高吞吐率。

**【例1】**某处理器中，浮点加法器采用4级流水线实现，流水线示意图如下图所示，每级处理时间 $\Delta t_0 = 250\text{ps}$ 。

请计算：

- (1) 该浮点加法器计算100组数据采用非流水和流水处理所用时间各是多少？
- (2) 采用流水处理的加速比是多少？
- (3) 采用流水处理的最大吞吐率是多少？

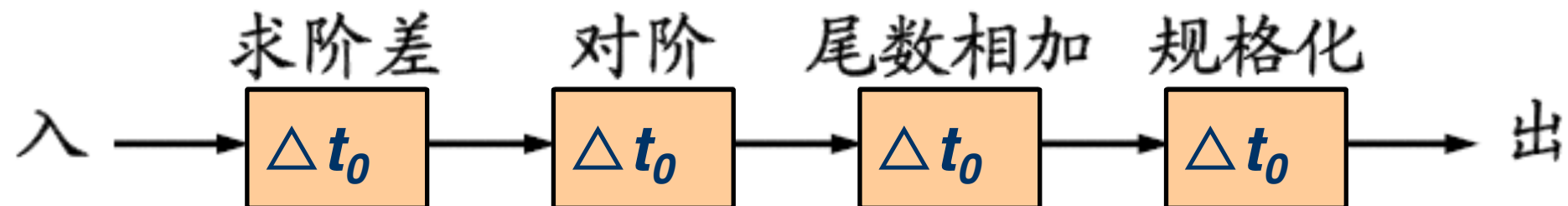


图 浮点加法流水线

**【解】**

(1) 采用非流水处理所用时间为

$$T_{\text{非流水}} = nm \Delta t_0 = 100 \times 4 \times 250 \text{ps} = 100 \text{ ns}$$

$$\begin{aligned} T_{\text{流水}} &= m \Delta t_0 + (n-1) \Delta t_0 \\ &= (4+100-1) \times 250 \text{ps} = 25.75 \text{ ns} \end{aligned}$$

(2) 加速比为

$$s = \frac{mn}{m+n-1} = \frac{4 \times 100}{4+100-1} = 3.88$$

(3) 最大吞吐率为

$$TP_{\text{max}} = 1/\Delta t_0 = 1/250 \text{ps} = 4 \text{ GFLOPS}$$

**【例2】** 假设在DLX的非流水实现和基本流水线中，5个功能单元的时间为：10，8，10，10，7（ns），流水额外开销为：1ns，求加速比S。

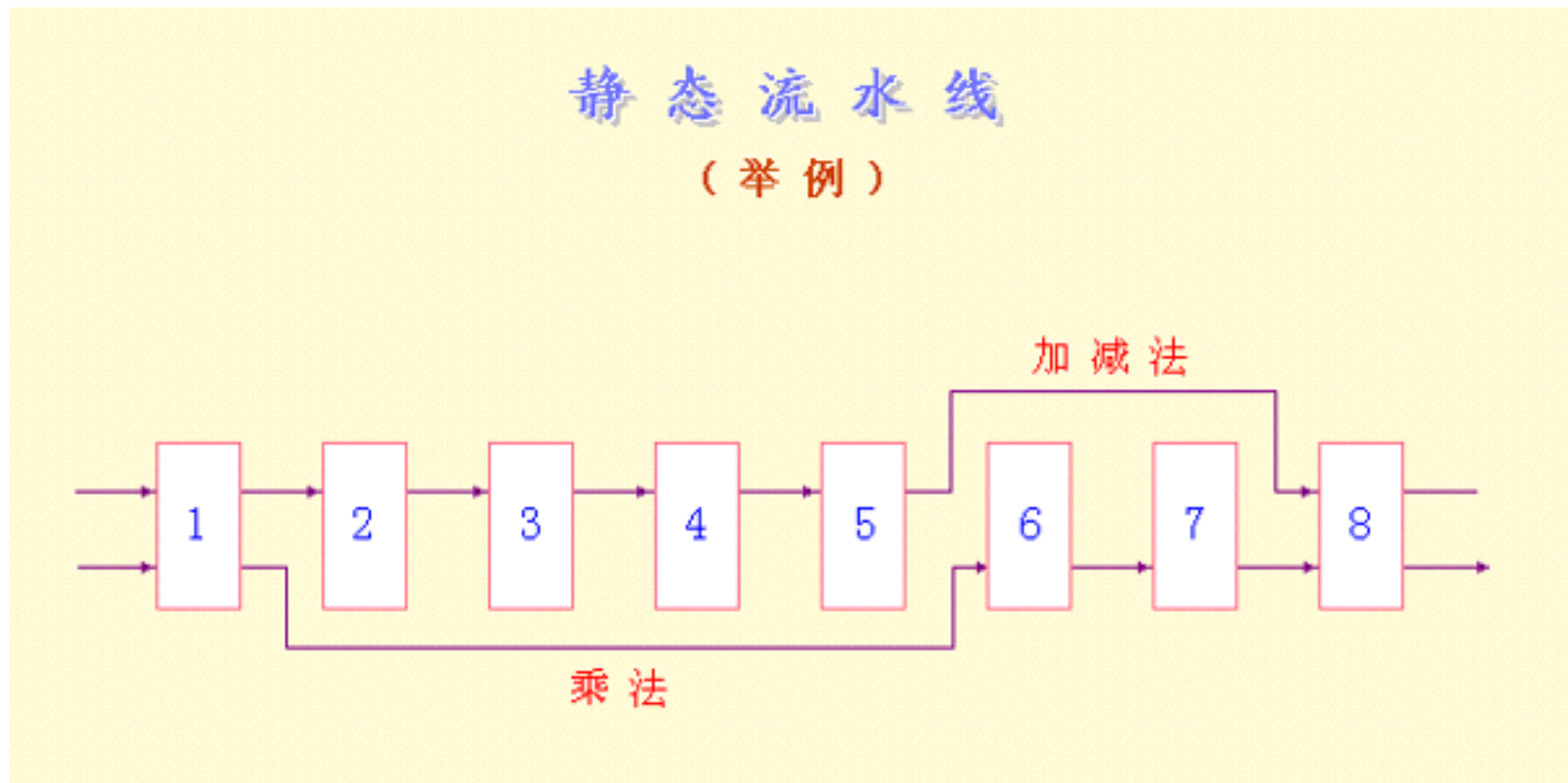
**【解】**

$$T_{\text{非流水}} = 10 + 8 + 10 + 10 + 7 = 45 \text{ (ns)}$$

$$T_{\text{流水}} = 10 + 1 = 11 \text{ (ns)}$$

$$S = 45 / 11 = 4.1$$

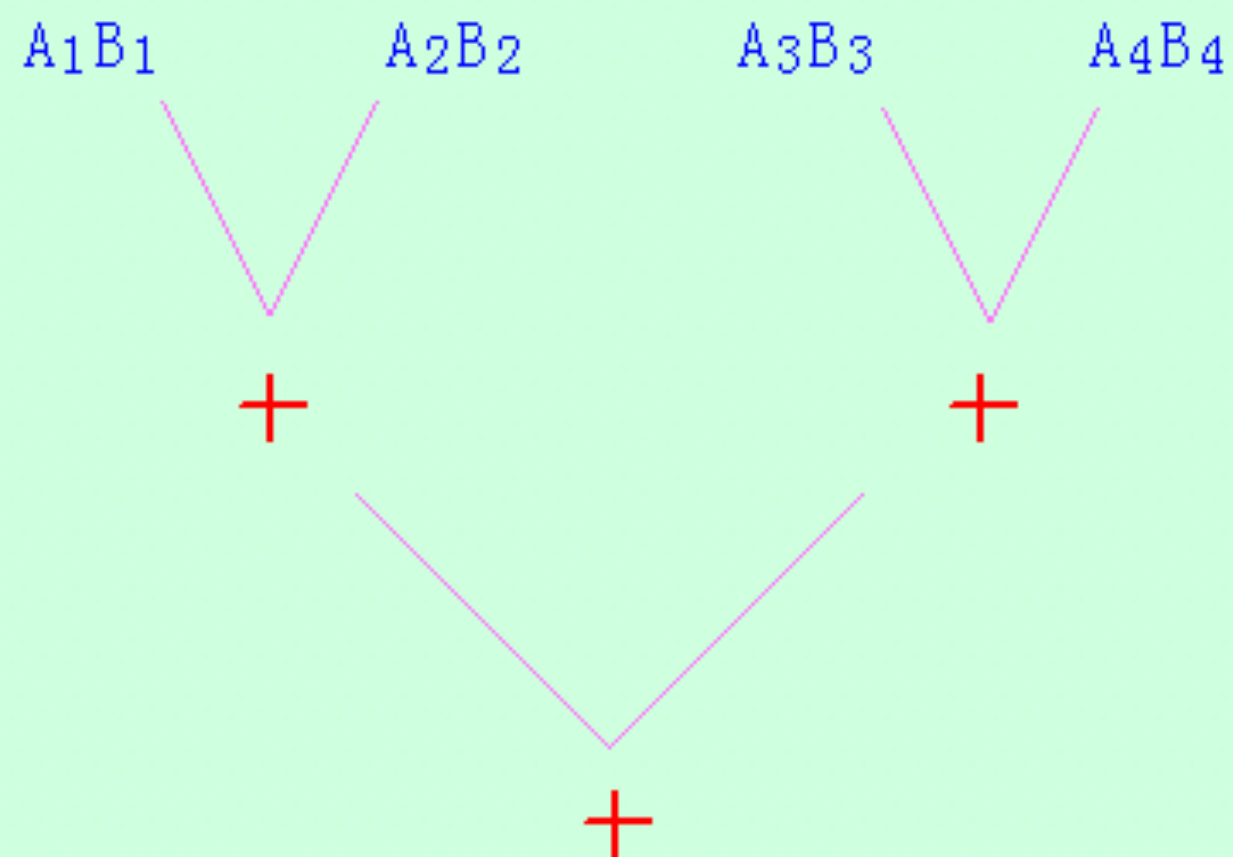
**【例3】** 在静态流水线上计算  $\sum A_i B_i$  ,  
求：吞吐率，加速比，效率。



# $\sum_{i=1}^4 A_i B_i$ 的计算过程

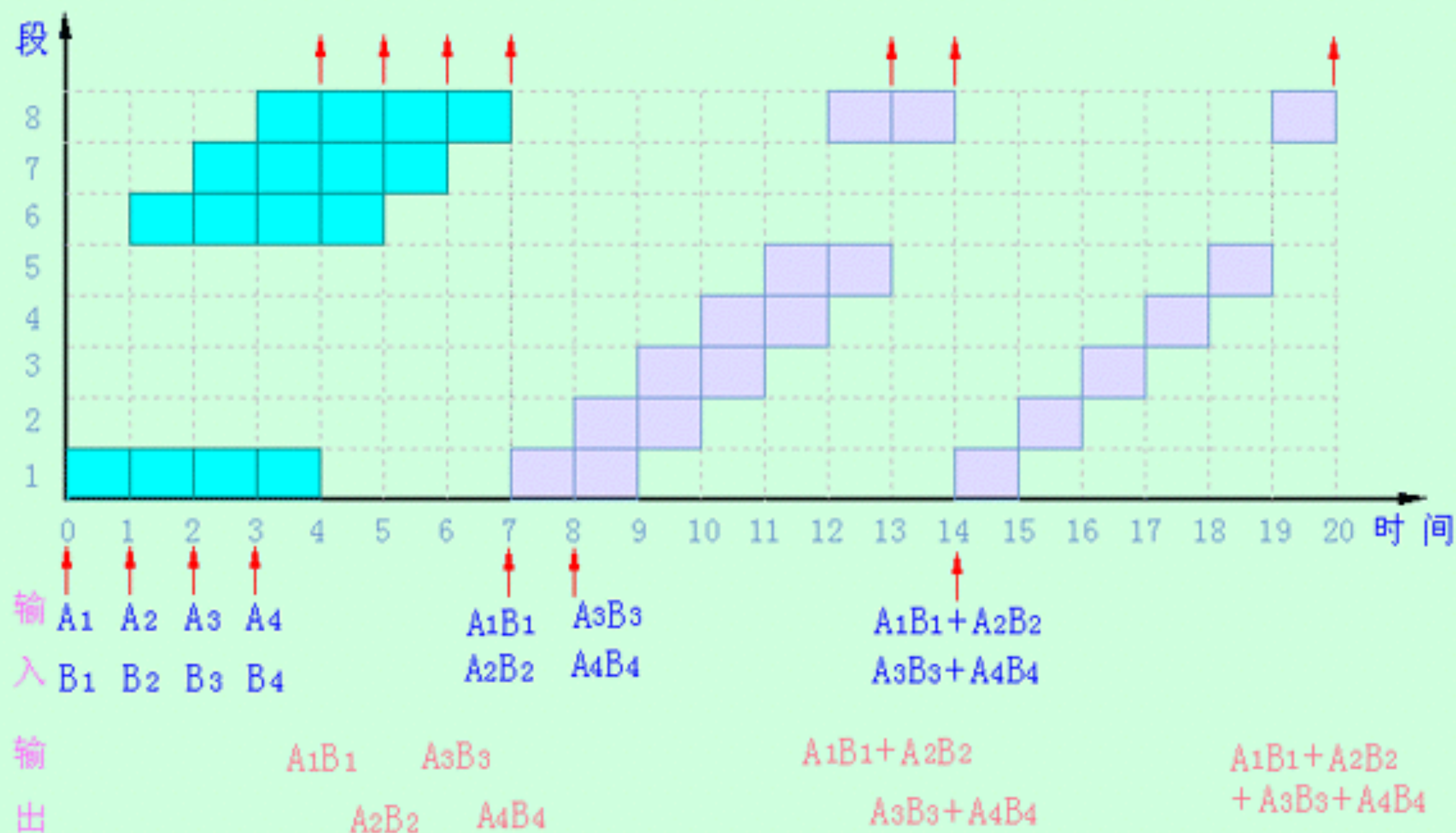
(例 3.1)

$$\sum_{i=1}^4 A_i B_i = A_1 B_1 + A_2 B_2 + A_3 B_3 + A_4 B_4$$



# 时空图

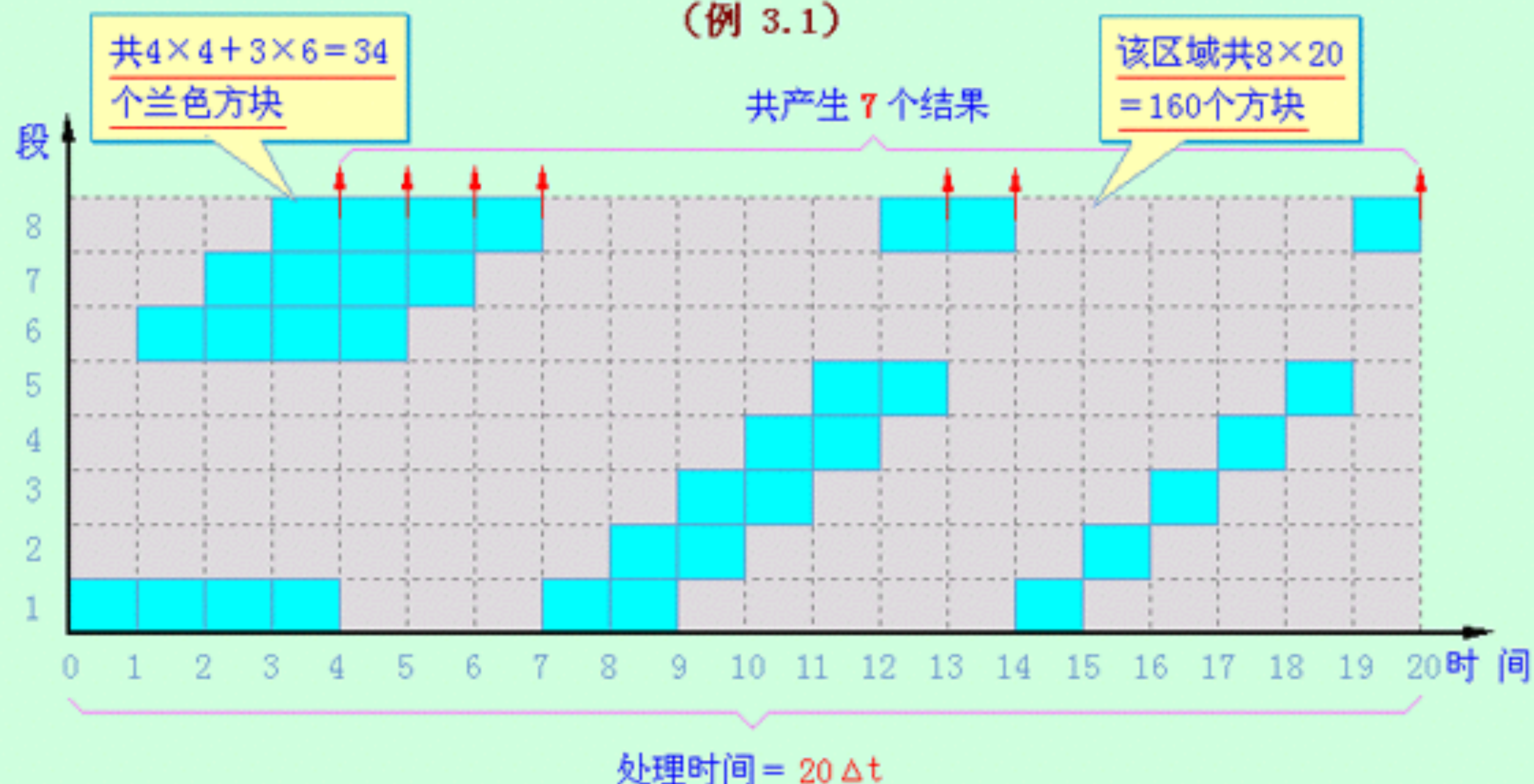
(例 3.1)





# 性能计算

(例 3.1)



吞吐率  $TP = \frac{7}{20 \Delta t}$

非流水处理所需时间 =  $4 \times 4 \Delta t + 3 \times 6 \Delta t = 34 \Delta t$

加速比  $S = \frac{34 \Delta t}{20 \Delta t} = 1.7$

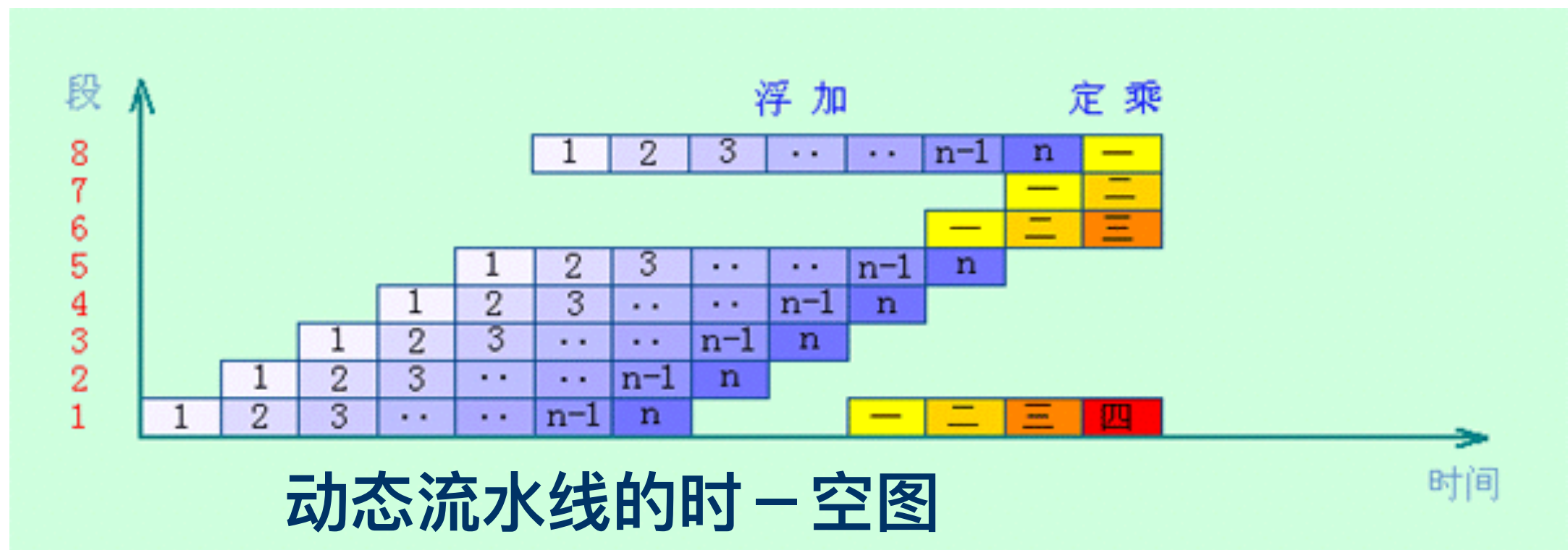
效率  $E = \frac{34}{160} = 0.21$



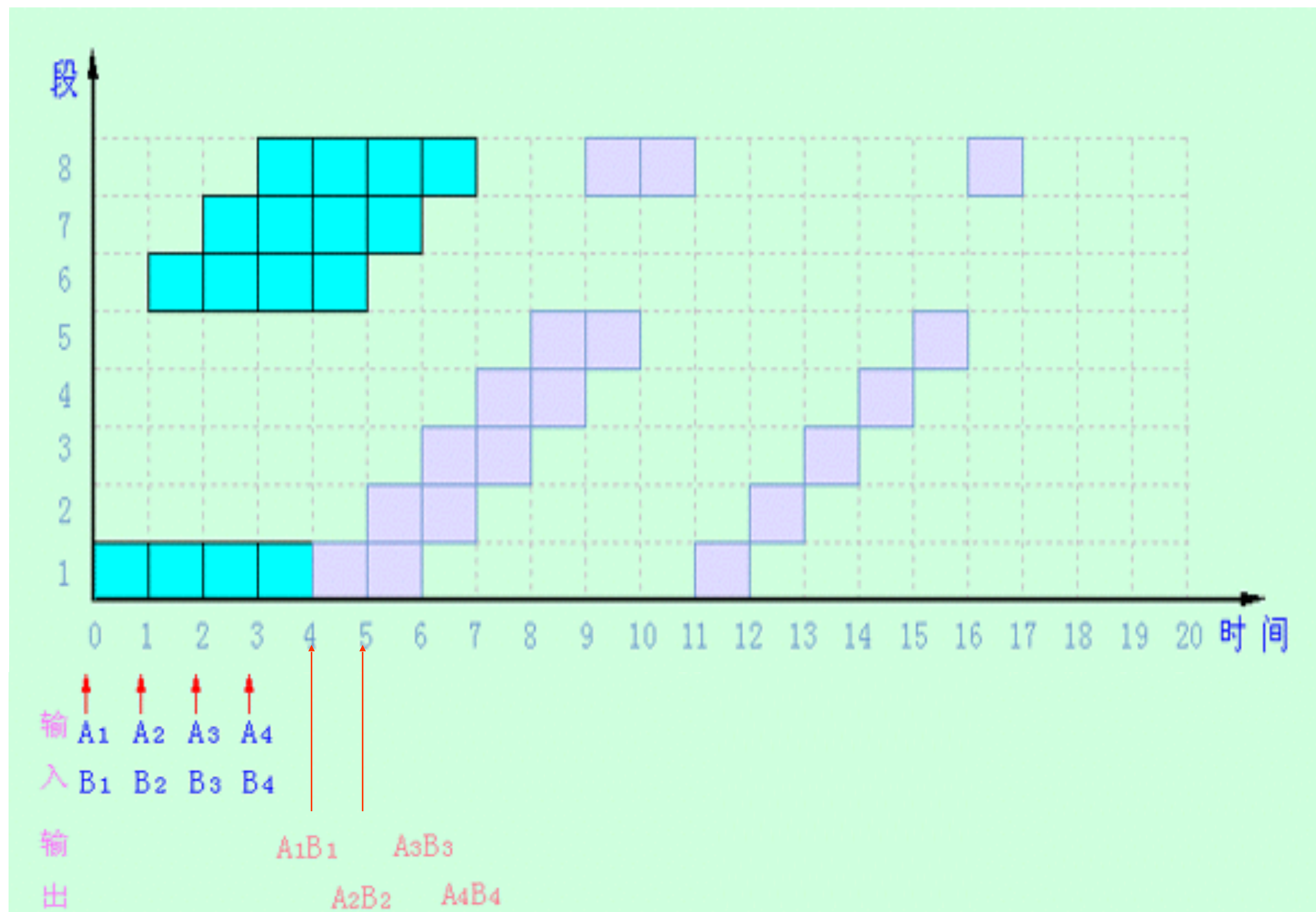
## 上述方案性能不高!

- 静态多功能流水线在对某种功能进行处理时，总有某些段处于空闲状态
- 功能切换增加了前一种功能的排空时间和后一种功能的通过时间
- 需要把输出回传到输入（相关）

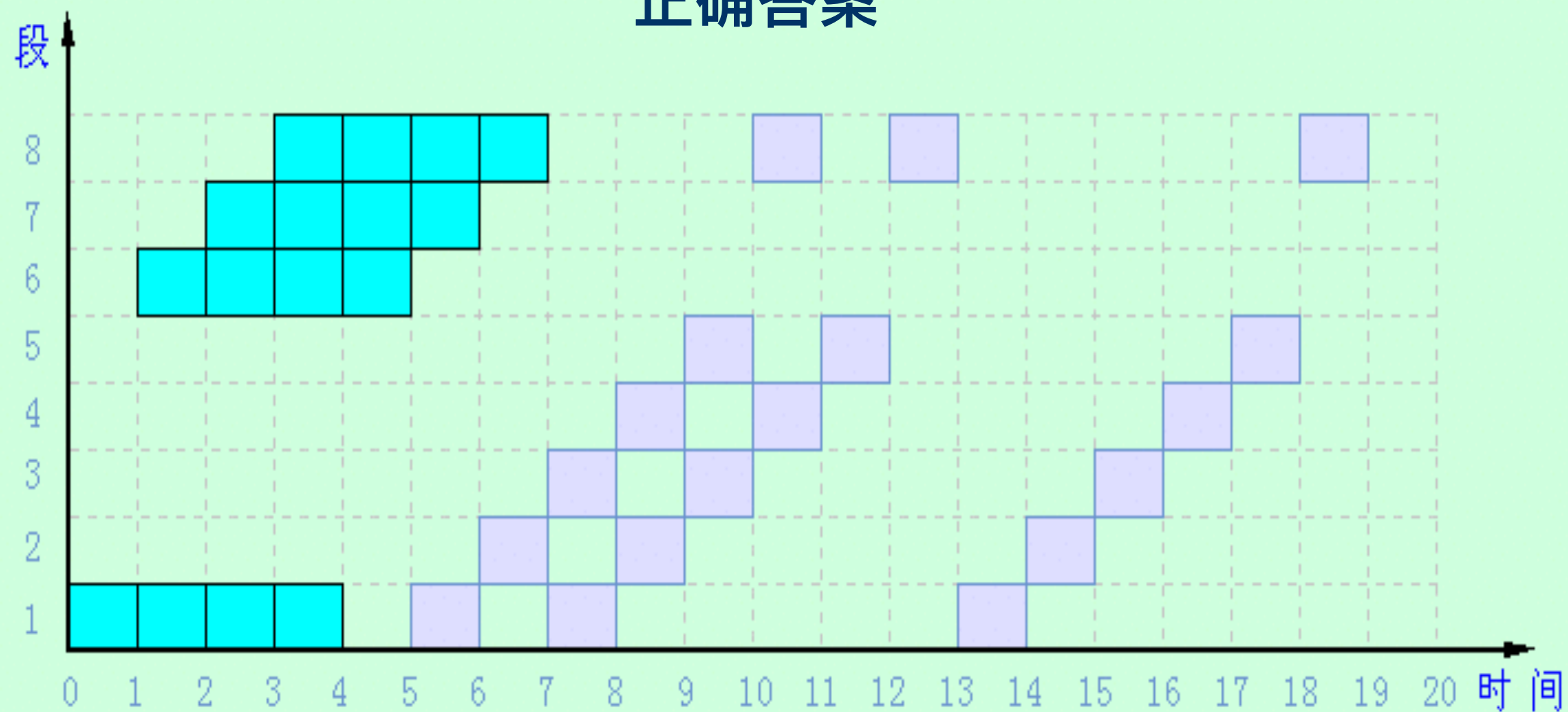
## 能否通过动态流水线改进其性能?



# 这样行不行？



# 正确答案



输入  $A_1$   $A_2$   $A_3$   $A_4$   
 $B_1$   $B_2$   $B_3$   $B_4$

输出  $A_1B_1$   $A_2B_2$   $A_3B_3$   $A_4B_4$   $A_1B_1 + A_2B_2$   $A_3B_3 + A_4B_4$   $A_1B_1 + A_2B_2 + A_3B_3 + A_4B_4$

## 小结：有关流水线性能的若干问题

- 流水线并不能减少（而且一般是增加）单条指令的执行时间，但能够提高吞吐率
- 增加流水线的深度可以提高流水线性能
- 流水线深度受限于流水线的延迟和额外开销
- 需要用高速锁存器作为流水线寄存器
- 指令之间存在的相关，限制了流水线的性能