

# 计算机系统结构

## 第三讲 流水线的性能指标 (2)

冯丹

武汉光电国家研究中心

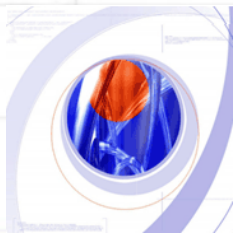


Computer Architecture

# 1. 效率

## — 效率 (E)

- 流水线中的设备实际使用时间与整个运行时间的比值
- 由于流水线有通过时间和排空时间，所以在连续完成 $n$ 个任务的时间内，各段并不是满负荷地工作



## 2-1.效率分析：各段时间均等

- 各段的效率 $e_i$ 相同

$$e_1 = e_2 = \cdots = e_k = \frac{n\Delta t}{T_k} = \frac{n}{k+n-1}$$

- 整条流水线的效率

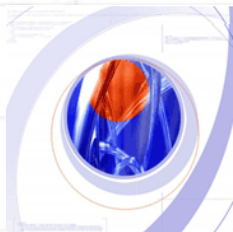
$$E = \frac{e_1 + e_2 + \cdots + e_k}{k} = \frac{ke_1}{k} = \frac{kn\Delta t}{kT_k} \quad \text{即}$$

$$E = \frac{n}{k+n-1}$$

- 最高效率

$$E_{\max} = \lim_{n \rightarrow \infty} \frac{n}{k+n-1} = 1$$

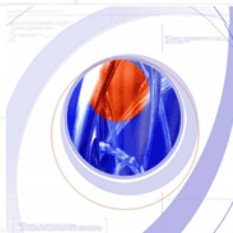
- 当 $n \gg k$ 时,  $E \approx 1$



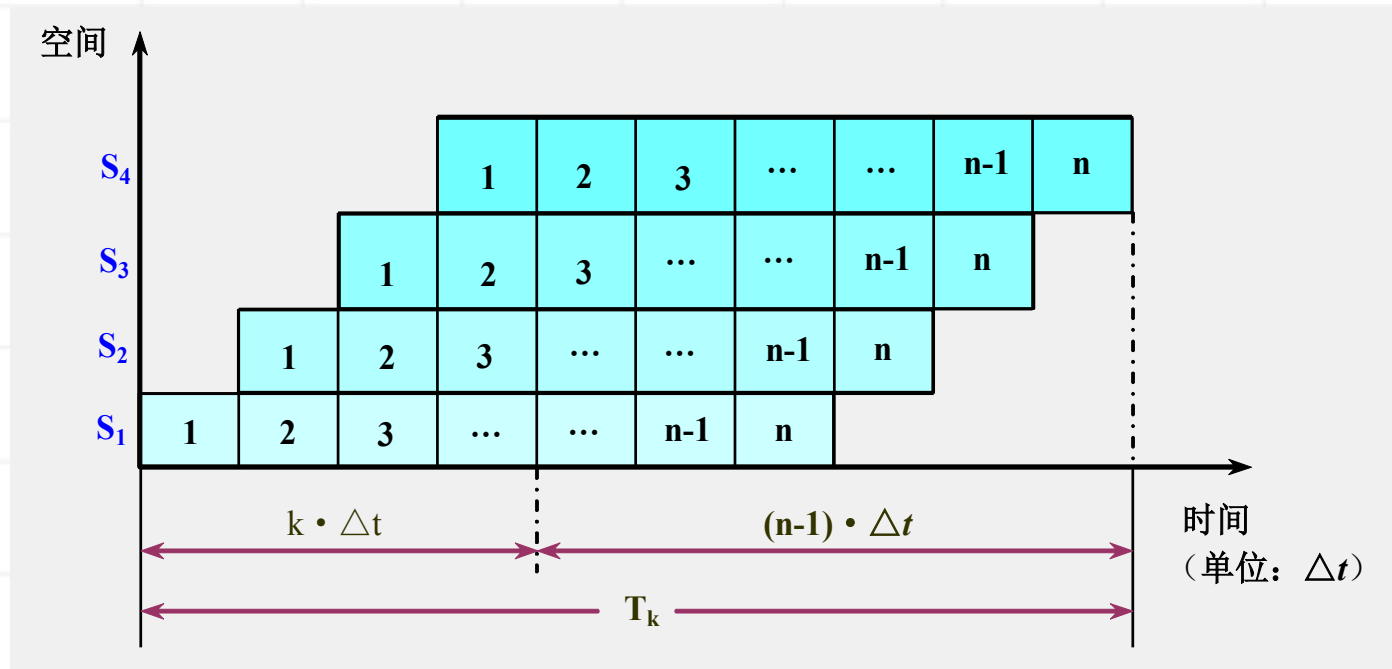
## 2-2.效率与其它指标的关系

- 当流水线各段时间相等时，流水线的效率与吞吐率成正比
  - $E = TP \cdot \Delta t$
- 流水线的效率是流水线的实际加速比 $S$ 与它的最大加速比 $k$ 的比值

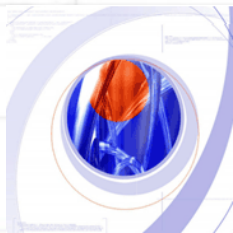
$$E = \frac{S}{k}$$



### 3. 更直观的描述效率

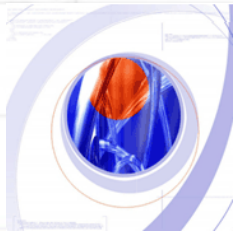


$$E = \frac{n \text{ 个任务实际占用的时空区}}{k \text{ 个段总的时空区}}$$



## 4.效率分析：各段时间不全相等

$$E = \frac{n \cdot \sum_{i=1}^k \Delta t_i}{k \left[ \sum_{i=1}^k \Delta t_i + (n-1) \cdot \max(\Delta t_1, \Delta t_2, \dots, \Delta t_k) \right]}$$

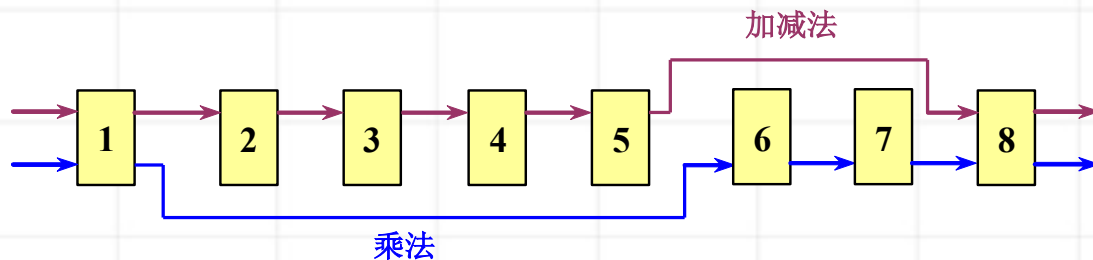


## 5-1. 多功能流水线性能指标计算

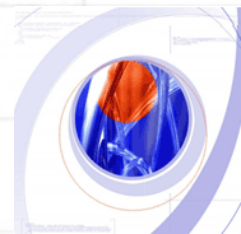
例 设在下图所示的静态流水线上计算：

$$\prod_{i=1}^4 (A_i + B_i)$$

流水线的输出可以直接返回输入端或暂存于相应的流水寄存器中，试计算其吞吐率、加速比和效率。



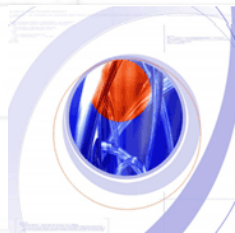
(每段的时间都为 $\Delta t$ )



## 5-2. 多功能流水线性能指标计算

解：（1）选择适合于流水线工作的算法

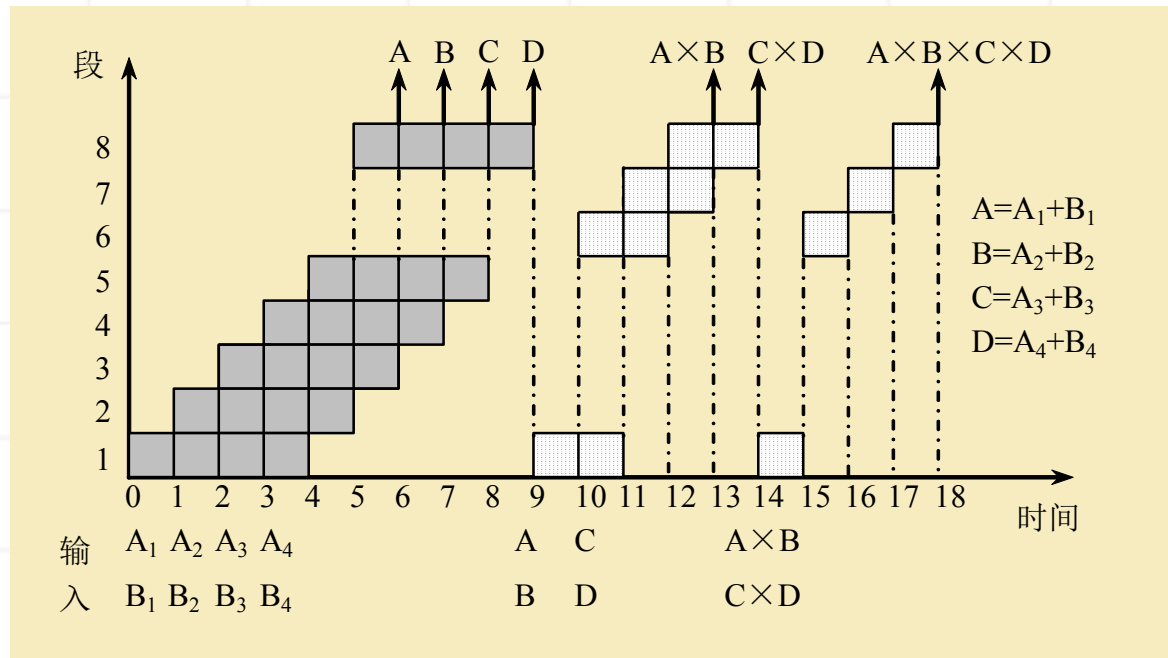
- 先计算 $A_1+B_1$ 、 $A_2+B_2$ 、 $A_3+B_3$ 和 $A_4+B_4$ ；
- 再计算 $(A_1+B_1) \times (A_2+B_2)$ 和 $(A_3+B_3) \times (A_4+B_4)$ ；
- 然后求总的乘积结果。





## 5-3. 多功能流水线性能指标计算

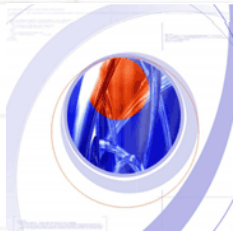
(2) 画出时空图



$$TP = \frac{7}{18\Delta t}$$

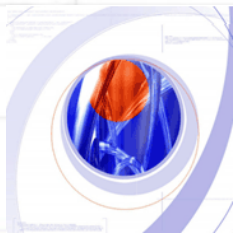
$$S = \frac{36\Delta t}{18\Delta t} = 2$$

$$E = \frac{4 \times 6 + 3 \times 4}{8 \times 18} = 0.25$$



## 5-4. 多功能流水线性能指标计算

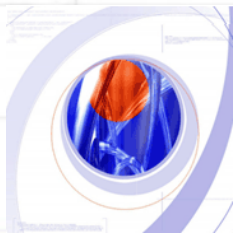
- 总结例题：影响（多功能）流水线性能的原因
  - 多功能流水线在做某一种运算时，总有一些段是空闲的；
  - 静态流水线在进行功能切换时，要等前一种运算全部流出流水线后才能进行后面的运算；
  - 运算之间存在关联，后面有些运算要用到前面运算的结果；
  - 流水线的工作过程有建立与排空部分。



## 6-1. 流水线设计中的重要问题

### — 瓶颈问题

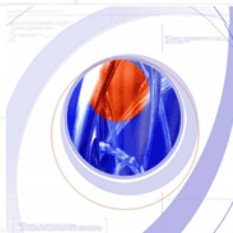
- 理想情况下，流水线在工作时，其中的任务是同步地每一个时钟周期往前流动一段。
- 当流水线各段不均匀时，**机器的时钟周期取决于瓶颈段的延迟时间。**
- 在设计流水线时，要尽可能使各段时间相等。



## 6-2. 流水线设计中的重要问题

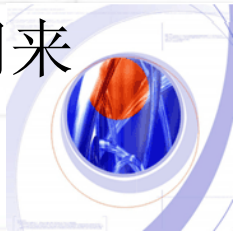
### — 流水线的额外开销

- 流水寄存器需要建立时间和传输延迟
  - 建立时间：在触发写操作的时钟信号到达之前，寄存器输入必须保持稳定的时间。
  - 传输延迟：时钟信号到达后到寄存器输出可用的时间。
- 时钟偏移开销
  - 流水线中，时钟到达各流水寄存器的最大差值时间。  
(时钟到达各流水寄存器的时间不是完全相同)



## 6-3. 流水线设计中的重要问题

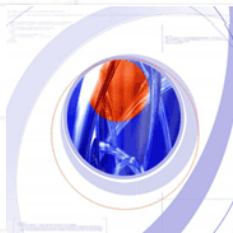
- 由流水线性能与开销得到的重要结论
  - 流水线并不能减少（而且一般是增加）单条指令的执行时间，但却能提高吞吐率。
  - 增加流水线的深度（段数）可以提高流水线的性能。
  - 流水线的深度受限于流水线的额外开销。
    - 当时钟周期小到与额外开销相同时，流水已没意义。因为这时在每一个时钟周期中已没有时间来作有用的工作。



## 6-4. 流水线设计中的重要问题

### — 冲突问题

- 前面产生了冲突的例子
- 流水线设计中要解决的**重要问题之一**



谢谢大家

