



西安电子科技大学
人工智能学院

计算机组成与体系结构

第7章 流水线技术与指令级并行

➤ 流水线性能度量

- ✓ 时-空图

- ✓ 吞吐率

➤ 流水线性能度量

- ✓ 加速比

- ✓ 效率

➤ 流水线性能分析



流水线性能测量

时-空图

时-空图

通过时间 $= (\text{流水线级数}-1) \times \tau$

$$\tau = (1 \sim n) \times T_{CLK}$$

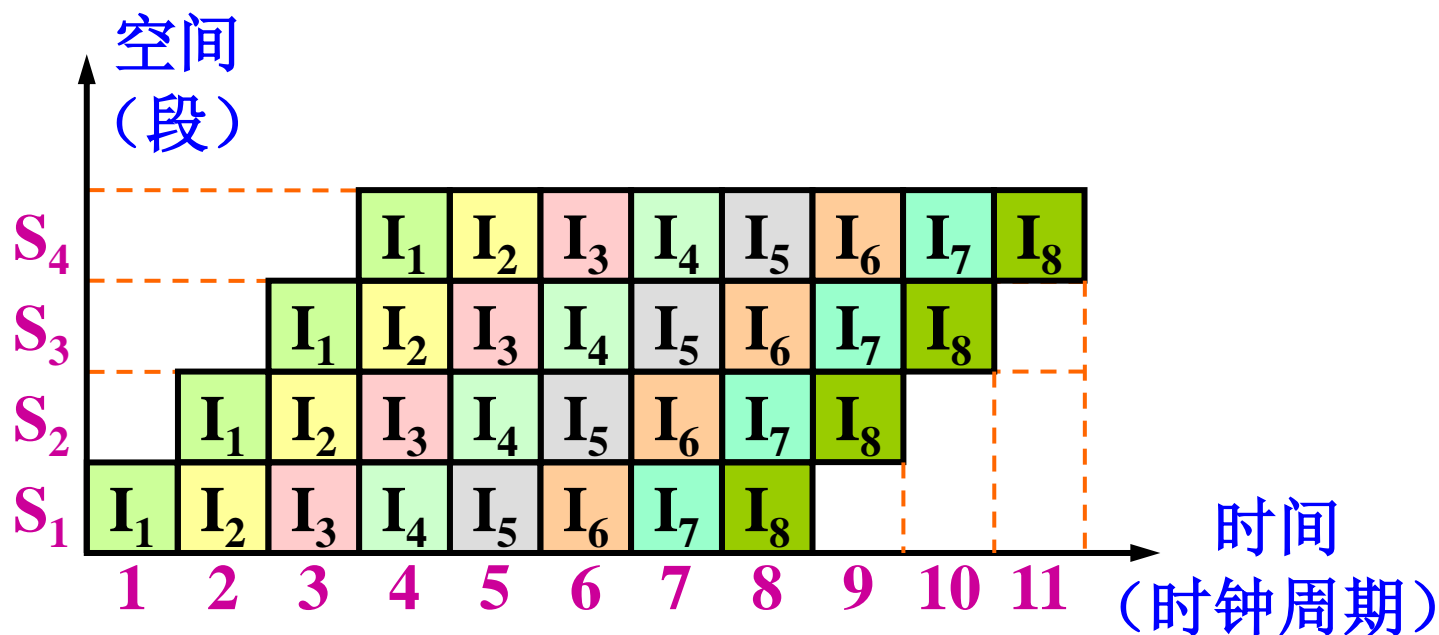


图 (a) 流水线时-空图

时-空图

通过时间 $= (\text{流水线级数}-1) \times \tau$

$$\tau = (1 \sim n) \times T_{CLK}$$

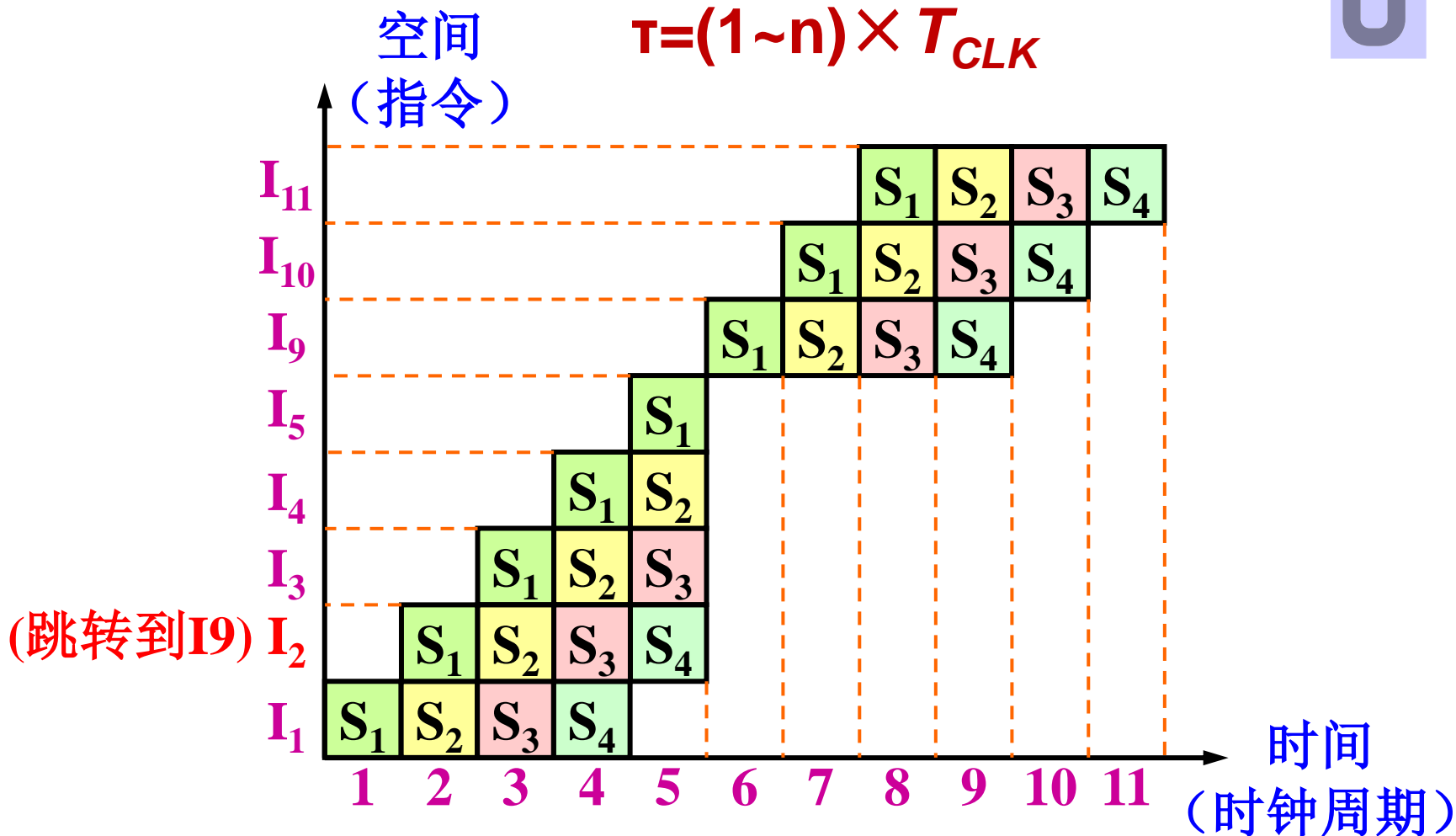


图 (b) 流水线时-空图



流水线性能测量

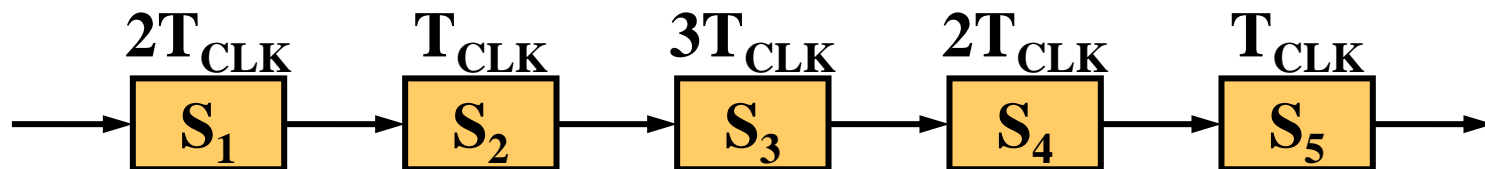
吞吐率

7.4.2 吞吐量

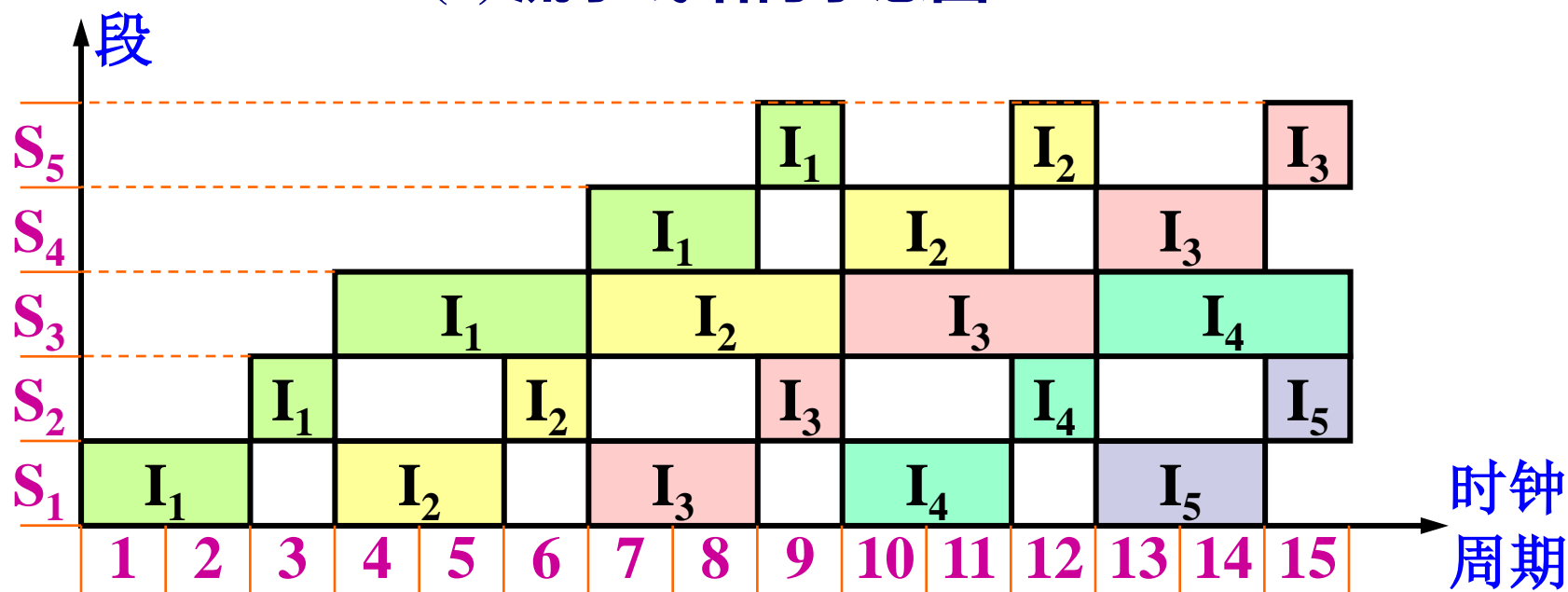
- ▶ **吞吐量**：单位时间内流水线所完成的**任务数**或**输出结果的数量**。
- ▶ **最大吞吐量** TP_{max} ：流水线在达到稳定状态后所得到的吞吐量。
 - ✓ 假设流水线各段运行时间相等，为1个时钟周期 T_{CLK} ，则： $TP_{max} = 1/T_{CLK}$
 - ✓ 假设流水线各段运行时间不等，第 i 段时间为 τ_i ，则

$$TP_{max} = 1/\max\{\tau_i\} = 1/\tau$$

7.4.2 吞吐率



(a) 流水线结构示意图



(b) 时-空图 (异步时钟控制)

图7.20 各段时间不等的流水线结构及时-空图

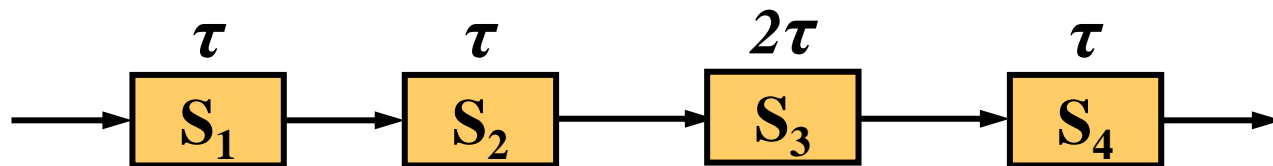
7.4.2 吞吐率

- 最大吞吐率取决于流水线中最慢一段所需的时间，所以该段成为流水线的瓶颈。
- 消除瓶颈的方法有：

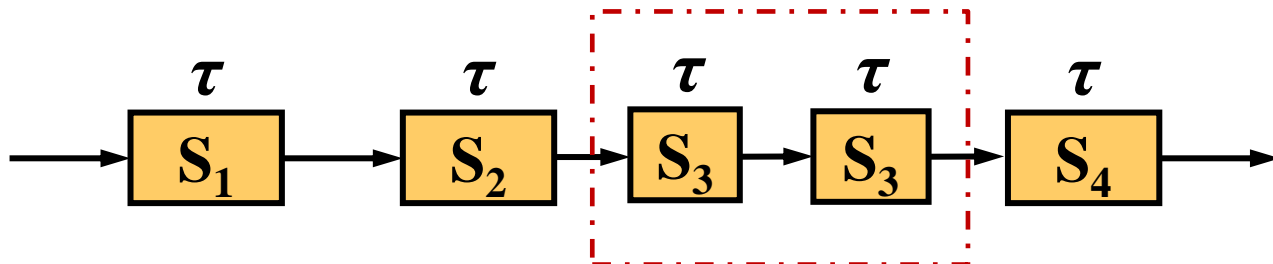
- ✓ 细分瓶颈段（首选）

- ✓ 重复设置瓶颈段

例：某流水线如下

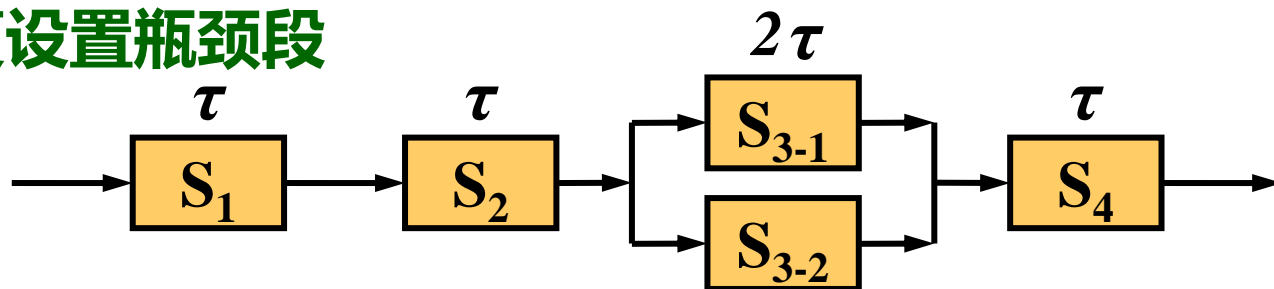


(1) 细分瓶颈段

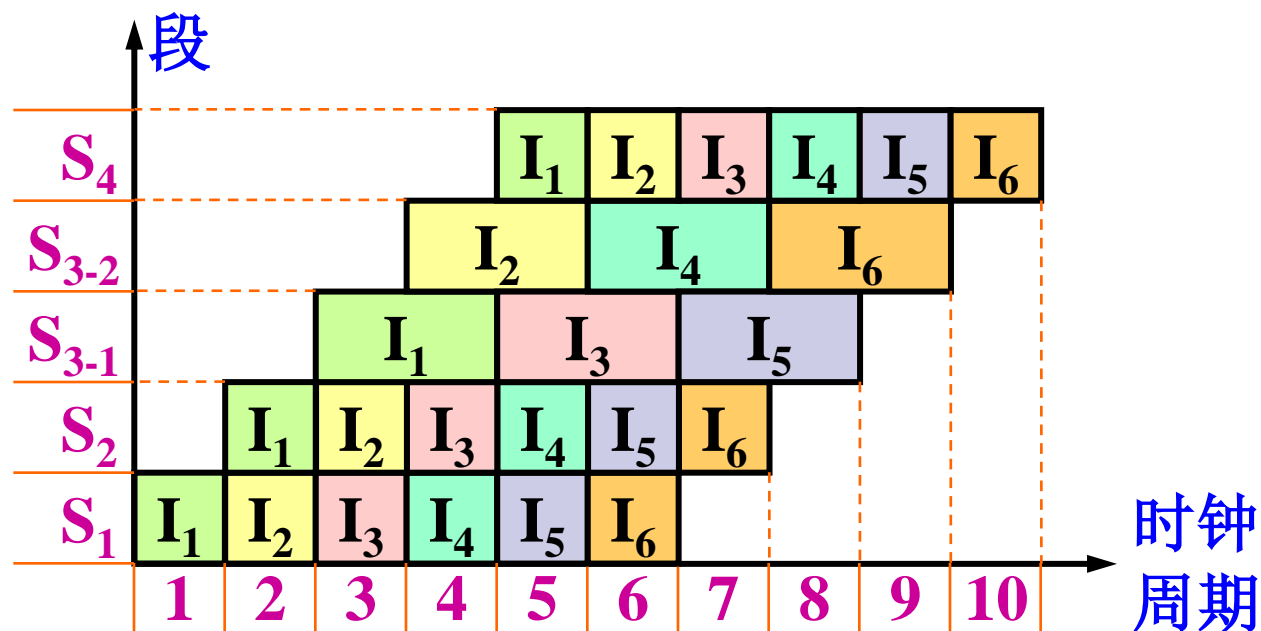


7.4.2 吞吐量

(2) 重复设置瓶颈段



(a) 流水线结构示意图



(b) 时-空图

图7.21 重复设置瓶颈段的流水线结构及时-空图

7.4.2 吞吐量

- **实际吞吐量**：若流水线由 m 段组成，完成 n 个任务的吞吐量称为实际吞吐量，记作 TP 。
- 使 TP 最大化，或使 TP 接近于 TP_{\max} ，是流水线实现中重点要解决的问题。
- 假设流水线 **各段运行时间相等**，为 1 个时钟周期 T_{CLK} ，在不出现流水线断流的情况下，完成 n 个任务所用时间为
$$T_n(m) = (m + (n - 1)) \times \tau = (m + (n - 1)) \times T_{CLK}$$
实际吞吐率为：

$$TP = \frac{n}{T_n(m)} = \frac{n}{(m + (n - 1)) \times T_{CLK}} = \frac{TP_{\max}}{1 + \frac{m - 1}{n}}$$

时空图

7.4.2 吞吐率

- 假设流水线各段运行时间不等，第*i*段时间为 τ_i ，则完成*n*个任务所用时间为

$$T_n(m) = \sum_{i=1}^m \tau_i + (n-1) \times \max\{\tau_i\}$$

实际吞吐率为

$$TP = \frac{n}{\sum_{i=1}^m \tau_i + (n-1) \times \max\{\tau_i\}}$$

时空图

当*n*很大时，*TP*趋近于 TP_{max} ，这说明流水技术适合于实现大量重复的时序过程。

7.4.2 吞吐量

- 对于**指令流水线**而言，吞吐量 **TP** 就是每秒执行的指令数，所以也可以用 **$MIPS$** 指标表示吞吐量，即

$$TP = MIPS = \frac{f_{CLK}}{CPI}$$

- 单流水线计算机系统

$$\therefore CPI_{\text{最佳}} = 1,$$

$$\therefore TP_{\text{max}} = f_{CLK}$$

随堂练习

某浮点运算流水线由4级构成，各级运行时间为100ns、150ns、200ns、100ns，则该流水线的最大吞吐率为 [填空1] (MF/s)。如果各级电路还可以进一步拆分，且流水线级数限定不超过12级，则该流水线合理的级数应设计为 [填空2]，此时流水线的最大吞吐率为 [填空3] (MF/s)。

某浮点运算流水线由4级构成，各级运行时间为100ns、150ns、200ns、100ns，则该流水线的最大吞吐率为

$$TP_{max} = 1/\max\{\tau_i\} = 1/200ns = 5(MF/s)$$

若以50ns拆分各级电路，则该流水线合理的级数应设计为

$$2+3+4+2=11 < 12$$

此时流水线的最大吞吐率为

$$TP_{max} = 1/\max\{\tau_i\} = 1/50ns = 20(MF/s)$$



流水线性能测量

加速比

7.4.3 加速比

若流水线为 m 段，加速比 S 定义为等功能的非流水线执行时间 $T(1)$ 与流水线执行时间 $T(m)$ 之比，即

$$S = S_n(m) = T_n(1) / T_n(m)$$

若每段运行时间均为 τ ，在不流水情况下，完成 n 个任务所需时间为： $T_n(1) = nm\tau$

在流水但不出现断流的情况下，完成 n 个任务所需时间为： $T_n(m) = m\tau + (n-1)\tau$ ，因此

$$S_n(m) = \frac{mn}{m+n-1} = \frac{m}{1 + \frac{m-1}{n}}$$

7.4.3 加速比

$$S_n(m) = \frac{mn}{m+n-1} = \frac{m}{1 + \frac{m-1}{n}}$$

$$S_{\max} = \lim_{n \rightarrow \infty} S_n(m) = \lim_{n \rightarrow \infty} \frac{m}{1 + \frac{m-1}{n}} = m$$

$$S_{\max} = \lim_{m \rightarrow \infty} S_n(m) = \lim_{m \rightarrow \infty} \frac{n}{1 + \frac{n-1}{m}} = n$$

➤ 增大指令流水线的**级数**和送入流水线的**指令数**均可以加速流水线的运行速度

7.4.3 加速比

- 指令流水线级数的增加会受到多方面因素的限制，如指令操作的可分解程度、实现成本（包括流水线控制电路的复杂度）、流水线段间延迟等，甚至当级数增加到一定程度时可能会对指令流水线的加速起反作用，实际计算机系统中常采用的指令流水线级数为几~十几级。
- 显然，如果流水线出现断流，加速比也会明显下降。



流水线性能测量

效率

7.4.4 效率

- 效率：流水线的设备利用率
- 由于流水线有通过（填充）时间和排空时间，还有各种引起断流的情况，所以流水线的各段并非一直满负荷工作，效率 $E < 1$ 。

假设流水线各段运行时间相等为 τ ，各段效率 e_i 也相等，即 $e_1 = e_2 = \dots = e_m = n\tau / T_n(m)$

则整个流水线效率 E 为：

$$E = \frac{\sum_{i=1}^m e_i}{m} = \frac{n\tau}{T_n(m)} = \frac{n}{m+n-1} = \frac{1}{1 + \frac{m-1}{n}}$$

当 $n \gg m$ 时， $E \approx 1$ 。

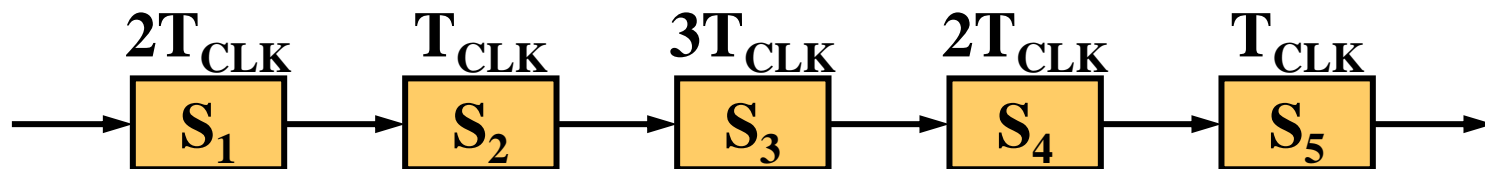
7.4.4 效率

- 效率：流水线的设备利用率
- 由于流水线有**通过（填充）时间**和**排空时间**，还有各种引起断流的情况，所以流水线的各段并非一直满负荷工作，效率 $E < 1$ 。

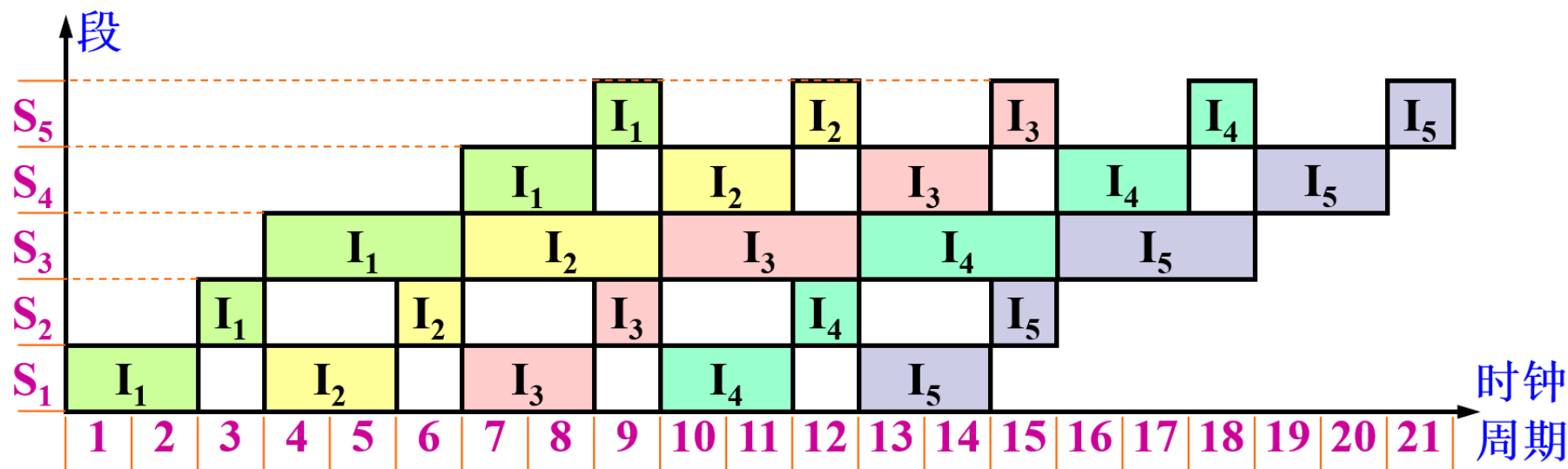
从时-空图上看，效率就是 n 个任务所占的时空区与 m 个段总的时空区之比。根据这个定义，可以计算流水线**各段时间不等时**的流水线效率为：

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

7.4.4 效率



(a) 流水线结构示意图



(b) 时-空图 (异步时钟控制)

$$\text{效率 } E = \frac{9 \times 5}{21 \times 5} = \frac{3}{7} = 42.86\%$$

7.4.5 吞吐率、加速比和效率的关系

- 效率是实际加速比 S 与最大加速比 m 之比。

$$E = n\tau / T_n(m) = mn\tau / (T_n(m) \cdot m) = S/m$$

- 当 τ 不变时，流水线的效率 E 与吞吐率 TP 呈正比

$$E = n\tau / T_n(m) = (n/T_n(m)) \cdot \tau = TP \cdot \tau$$

- 当 m 和 τ 不变时，流水线的加速比 S 与吞吐率 TP 呈正比。

$$TP = S/(m\tau)$$



流水线性能分析

7.4.6 流水线性能分析

【例7.2】在图7.18提供的5级RISC-V指令流水线上，7条指令分别在各段花费的处理时间如下（流水线寄存器延迟忽略）：
 $1\text{ps}=10^{-12}\text{s}$

指令类	取指令	寄存器读	ALU运算	数据访问	寄存器写	总时间
加载双字 (ld)	200ps	100ps	200ps	200ps	100ps	800ps
存储双字 (sd)	200ps	100ps	200ps	200ps		700ps
R型 (add, sub, and, or)	200ps	100ps	200ps		100ps	600ps
分支 (beq)	200ps	100ps	200ps			500ps

- (1) 试为该流水线确定时钟频率。
- (2) 某段程序有1000条指令，ld和sd指令各占15%，beq指令占10%，其余为R型指令。该段程序在图6.2的非流水结构中运行时间是多少？
- (3) 该段程序在图7.18的流水线上执行，若不考虑指令相关的影响，与非流水执行相比，加速比为多少？

【解】(1) 假设流水线采用同步时钟控制，根据式(7-2)，有

$$T_{LCLK} = \max\{t_i\} = \max\{200\text{ps}, 100\text{ps}, 200\text{ps}, 200\text{ps}, 100\text{ps}\} = 200\text{ps}$$

所以，流水线时钟频率 $f_{LCLK} = 1/T_{LCLK} = 1/200\text{ps} = 5\text{GHz}$ 。

(2) 在非流水结构上的运行时间为

$$T_n(1) = (800\text{ps} \times 15\% + 700\text{ps} \times 15\% + 600\text{ps} \times 60\% + 500\text{ps} \times 10\%) \times 1000 = 635000\text{ps}$$

(3) 在流水结构上的运行时间为

$$T_n(m) = (m + n - 1) \times T_{LCLK} = (5 + 1000 - 1) \times 200\text{ps} = 200800\text{ps}$$

所以，加速比 $S = T_n(1)/T_n(m) = 635000/200800 = 3.16$

结论

- ✓ 流水线控制时钟周期通常取 $T_{LCLK} = \max\{t_i\}$ ， $S_{理想} = m = 5$ ， $S_{实际} = 3.16$ 。
- ✓ 为了加大流水线的吞吐量，使流水线各段有较均衡的运行时间，减少不必要的等待时间，一种可行的方法就是增加流水线的级数。

7.4.6 流水线性能分析

【例7.5】某指令流水线由10级构成，每级处理时间为100ps。现要执行一段程序，该程序由50个结构相同、顺序执行的小程序段组成，每个小程序段由15条指令构成，其中第4条指令 I_4 为条件跳转指令，当条件为真时程序跳转到指令 I_{10} ，当条件为假时程序顺序执行。请计算：

(1) 当所有条件跳转指令均未发生条件跳转时，执行这段程序时流水线的实际吞吐率、加速比、效率各是多少？

(2) 当所有条件跳转指令均发生条件跳转时，执行这段程序时流水线的实际吞吐率、加速比、效率又是多少？

7.4.6 流水线性能分析

【解】 (1) 此时流水线的工作状况可参考图7.19(a)，因条件跳转未发生，所以50个小程序段中的所有指令顺序地在流水线上不断流地执行，所用时间为

$$T_1 = (m + n_1 - 1) \times \tau = (10 + 50 \times 15 - 1) \times 100\text{ps} = 75900\text{ps}$$

实际吞吐率为 $TP = \frac{n_1}{T_1} = \frac{50 \times 15}{(10 + 50 \times 15 - 1) \times 100\text{ps}} = 9881(\text{MIPS})$

加速比为 $S = \frac{n_1 m \tau}{T_1} = \frac{50 \times 15 \times 10}{10 + 50 \times 15 - 1} = 9.881$

效率为 $E = \frac{n_1}{m + n_1 - 1} = \frac{50 \times 15}{10 + 50 \times 15 - 1} = 0.988$

还可以用加速比/最大加速比计算

7.4.6 流水线性能分析

【解】（2）因发生了条件跳转，所以在执行50个小程序段时，每个小程序段只执行10条指令，且流水线会出现50次断流，每次断流时，流水线要进行重新填充操作，填充时间为 $(m-1) \times \tau$ ，此时流水线的工作状况可参考图7.19(b)，故完成规定程序所用时间为

$$\begin{aligned} T_2 &= (m+n_2-1) \times \tau + 50 \times (m-1) \times \tau \\ &= (10+50 \times 10-1) \times 100\text{ps} + 50 \times (10-1) \times 100\text{ps} = 95900\text{ps} \end{aligned}$$

实际吞吐率为 $TP = \frac{n_2}{T_2} = \frac{50 \times 10}{(10 + 50 \times 10 - 1 + 50 \times 9) \times 100\text{ps}} = 5213.8(\text{MIPS})$

加速比为 $S = \frac{n_2 m \tau}{T_2} = \frac{50 \times 10 \times 10}{10 + 50 \times 10 - 1 + 50 \times (10-1)} = 5.214$

效率为 $E = \frac{n_2}{m + n_2 - 1 + 50 \times (m-1)} = \frac{50 \times 10}{10 + 50 \times 10 - 1 + 50 \times (10-1)} = 0.5214$

➤ 结论：程序发生跳转会大大降低流水线性能。

7.4.6 流水线性能分析

【例7.6】某指令流水线结构如图7.20(a)所示，在流水线不断流的情况下，分析该流水线的吞吐率和加速比。

图7.20

【解】由于流水线各段运行时间不同，假设可以采用异步方式控制，即流水线各段的推进可以按照各段运行时间进行控制，则流水线工作的时空状态如图7.23所示。

由于最慢运行段S3的限制，使得S1、S2段不能及时向前推进，到执行第4条指令开始，对流水线各段推进的控制已出现规律性，该规律正是图7.20(b)时-空图所描述的流水线各段的时钟控制规律。

7.4.6 流水线性能分析

【例7.6】某指令流水线结构如图7.20(a)所示，在流水线不断流的情况下，分析该流水线的吞吐率和加速比？

图7.20

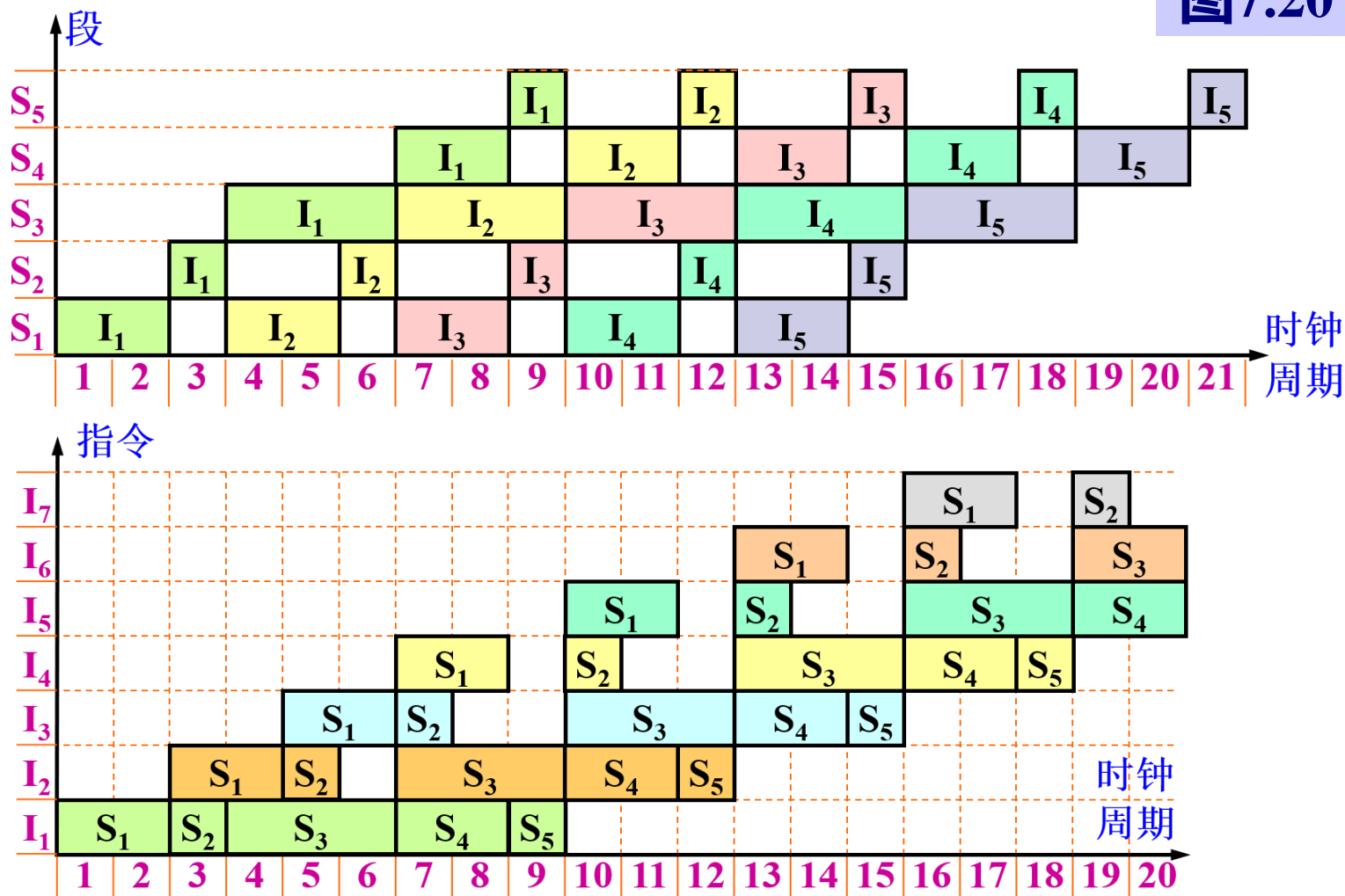


图7.23 图7.20(a)流水线时-空图的另一种表示

7.4.6 流水线性能分析

【例7.6】某指令流水线结构如图7.20(a)所示，在流水线不断流的情况下，分析该流水线的吞吐率和加速比？

图7.20

【解】

若流水线各段按照这种规律进行异步推进，则流水线达到稳定时，一条指令执行需要12个时钟周期，每3个时钟周期产生一条指令的处理结果，所以该流水线的性能为

$$TP_{max} = 1/(3T_{CLK})$$

$$S = (9T_{CLK})/(3T_{CLK}) = 3$$

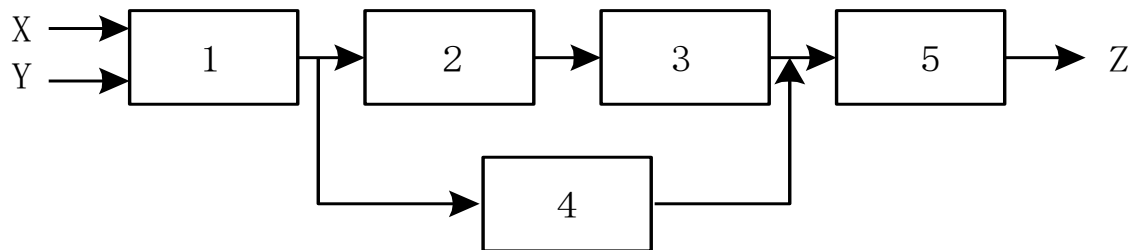
➤ 结论：流水线采用异步控制并没有给流水线性能带来改善，反而会增加控制电路的复杂性。所以，流水线采用的基本控制方式为同步方式。

	填充时间 (T_{CLK})	一条指令执行时间 (T_{CLK})	CPI	TP_{max}	S	控制电路
同步控制	12	15	3	$1/(3T_{CLK})$	3	简单
异步控制	8	12	3	$1/(3T_{CLK})$	3	复杂

随堂测试

下图为静态双功能浮点运算流水线，其中功能段1→4→5组成乘法流水线，功能段1→2→3→5组成加法流水线。假定流水线各段工作时间为 Δt ，流水线输出暂存于缓冲寄存器中，它们的延迟时间及功能切换时间均可忽略，现在要完成4次乘法任务后切换到加法流水线完成3次加法运算。

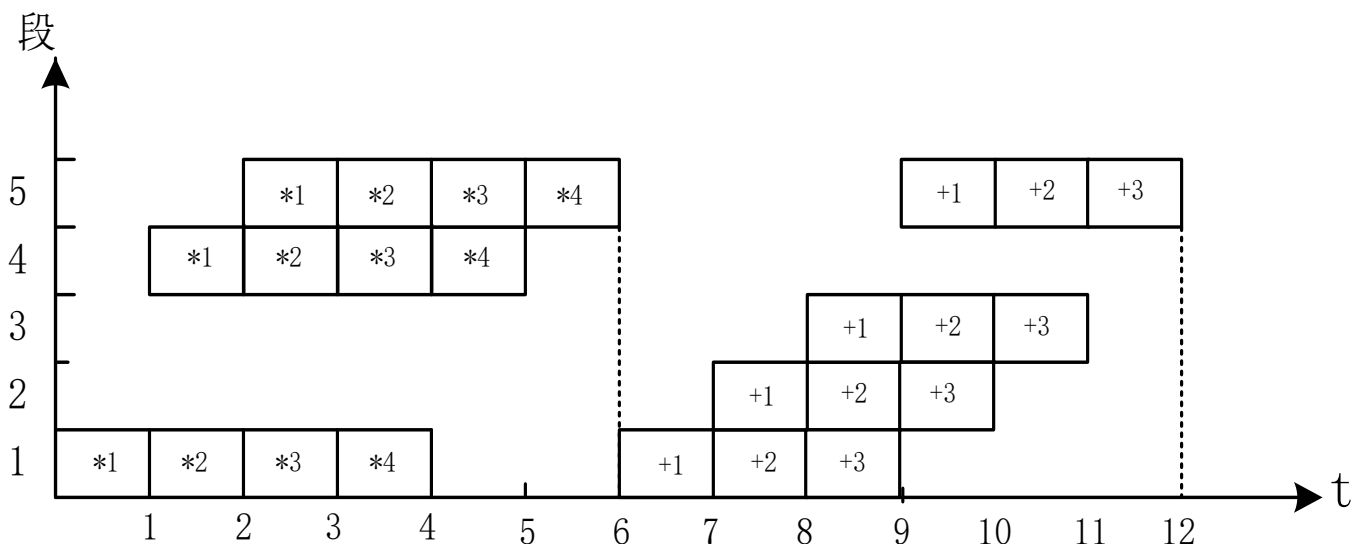
- (1) 画此流水线的时空图；
- (2) 计算加速比 S 及效率 E 。



答案

解：

(1)



(2) 非流水线处理时间：

$$4 \times 3 \times \Delta t + 3 \times 4 \times \Delta t = 24 \Delta t$$

流水线处理时间：

$$(3+4-1)\Delta t + (4+3-1)\Delta t = 12 \Delta t$$

加速比 $S = 24 \Delta t / 12 \Delta t = 2$

效率 $E = (4 \times 3 + 3 \times 4) / (12 \times 5) = 0.4$

如果加法运算能提前开始，结果会怎样？