

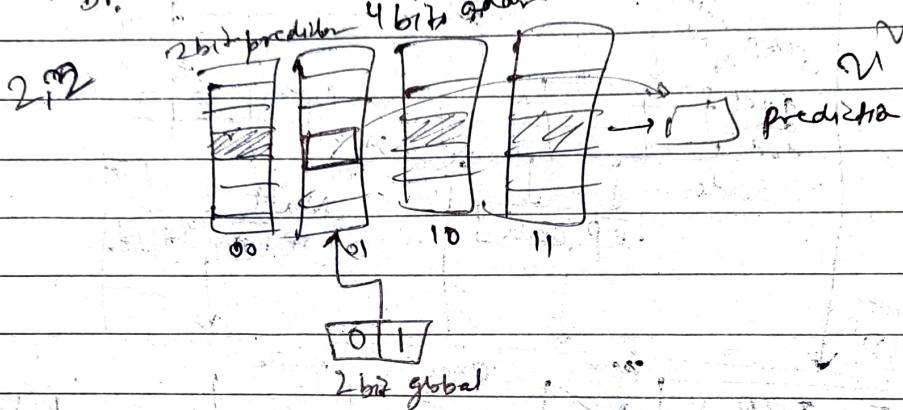
For eg.: if we maintain the info about the 3 earlier branches  $B_1, B_2$  and  $B_3$ , the behavior of the current branch now depends on how these earlier branches behaved.

Page No.:

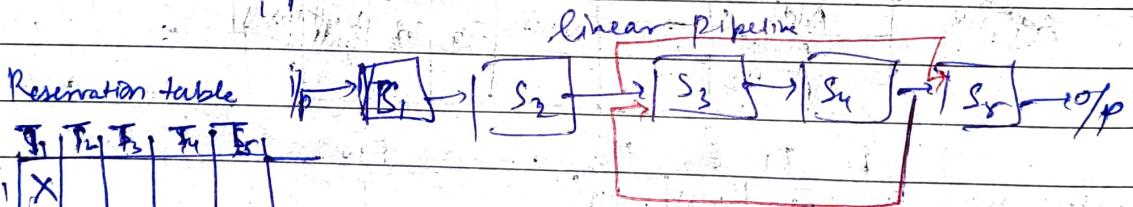
Date:

There are 8 different possibilities, from all 3 of them not taken (000) to all taken (111). There are 8 different predictors maintained and each one of them will be maintained as 1-bit or 2-bit predictor.

with global information with branches and each predictor is n-bit.

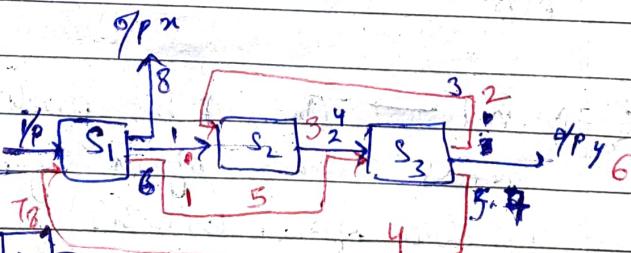


### Non-linear pipeline:-



	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>
S <sub>1</sub>	X				
S <sub>2</sub>		X			
S <sub>3</sub>			X		
S <sub>4</sub>				X	
S <sub>5</sub>					X

	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	T <sub>7</sub>	T <sub>8</sub>
S <sub>1</sub>	X			X	X			
S <sub>2</sub>		X	X					
S <sub>3</sub>		X	X	X				
S <sub>4</sub>	X			X				
S <sub>5</sub>		X						
S <sub>6</sub>			X					
S <sub>7</sub>	X	X	X	X				
S <sub>8</sub>	X	X	X	X				



(next cycle) X<sub>1</sub>, X<sub>2</sub> (for next cycle):

S <sub>1</sub>	X	⋮		X	⋮		X	⋮	
S <sub>2</sub>	X	⋮		X	⋮		X	⋮	
S <sub>3</sub>	X	⋮		X	⋮		X	⋮	

Y<sub>1</sub> Y<sub>2</sub>

**Simple cycle** - a latency cycle in which each state appears only once.  
 Some of the simple cycles are **greedy cycles**

**Greedy cycle** - in one whose edges are all made with minimum latencies from their ~~loop~~ state.

Date : / /

	1	2	3	4	5	6	7	8
$s_1$	X				X			
$s_2$		X	X					
$s_3$			X	X	X			

$$8-6=2, 8-1=7, 6-1=5$$

$$4-2=2$$

$$5-3=2, 7-5=2, 7-3=4$$

union = 2, 4, 5, 7 = forbidden (1) latency

MAL  $\geq$  sum of checkmarks  
in a row

MAL  $\geq 3$

(0) permissible =  $U - \text{forbidden}$

$$= \{1, 2, 3, 4, 5, 6, 7, 8\} - \{2, 4, 5, 7\}$$

$$= \{1, 3, 6, 8\}$$

4

8 G G G G G G G G  
8 7 6 5 4 3 2 1

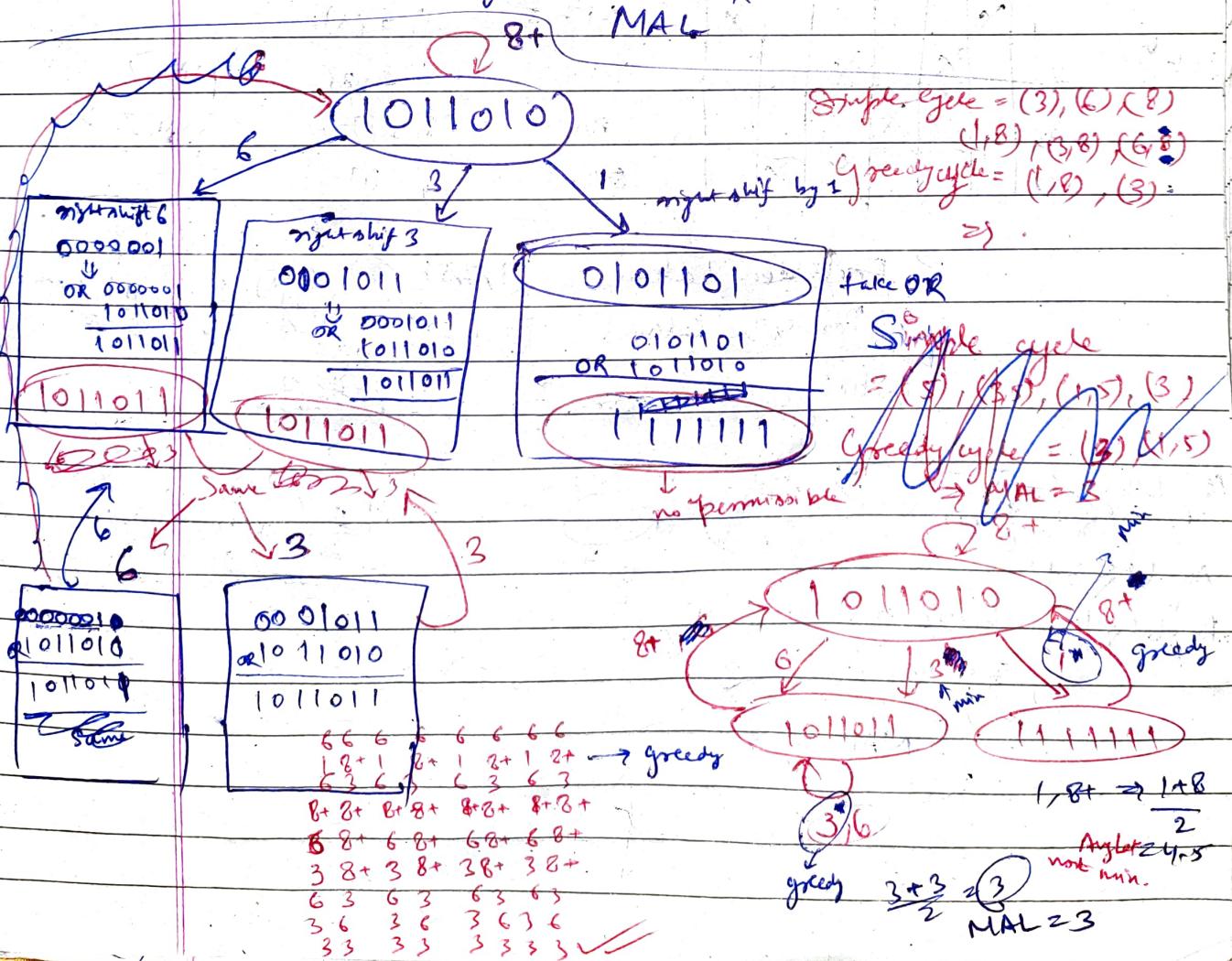
initial collision vector = 0 1 0 1 1 0 1 0

Avg latency  $\leq$   $\frac{\text{sum of ICV}}{\text{ICV} + 1}$   $\rightarrow$  upper bound

Avg Latency  $\leq 5$

minimum Avg latency (MAL)

throughput =  $\frac{1}{\text{MAL}} \times T$  ('Time period')



reservation for you	1	2	3	4	5	6	
$S_1$	X			X			
$S_2$		X					
$S_3$		X	X		X		

$$S_1 = \{4\}$$

$$S_2 = \{3\}$$

$$S_3 = \{2, 4\}$$

MAL = 3

$$\text{Forbidden latencies} = \{2, 4\}$$

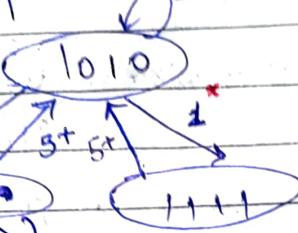
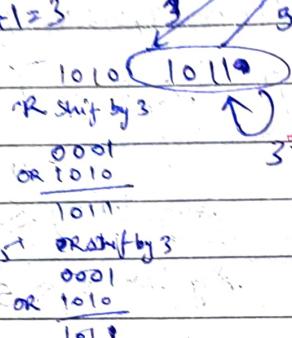
$$\text{Permissible latencies} = \{1, 3, 5, 6^+\}$$

(ICV) Initial collision vector = 001010

Avg latency  $\leftarrow \frac{\text{no of } 1's}{\text{ICV}}$

Avg lat  $\leq 2+1=3$

5+ = 1010



Simple cycle = (5)

(3, 5), (1, 5), (3)

Greedy cycle = (3), (1, 5)

Right by 1

0101  
OR 1010  
1111

Right by 5+

0000  
OR 1010  
1010

Kar View

	1	2	3	4	5	6
$S_1$	X			X		
$S_2$		X		X		
$S_3$			X			
$S_4$			X	X		

$$6-1 = 5 \quad S_1 = \{5\}$$

$$4-2 = 2 \quad S_2 = \{2\}$$

$$\text{null} \quad S_3 = \{\}$$

$$5+4 = 1 \quad S_4 = \{1\}$$

max no of checkmarks in a row

$$\text{Forbidden latencies} = \{1, 2, 5\}$$

$$\text{Permissible latencies} = \{3, 4, 6^+\}$$

Initial collision vector

010011

right shift by 3

R<sub>3</sub>  
6+

00001

OR 10011

00000

10011 (same)

Avg latency  $\leftarrow \frac{\text{no of } 1's}{\text{ICV}}$

Avg lat  $\leq 3+1$

Simple cycle = 3, 4, 6<sup>+</sup>

Greedy cycle = 3

MAL = 3

Q 6.9

	1	2	3	4	5	6
$S_1$	X				X	
$S_2$		X			X	
$S_3$			X			
$S_4$				X		
$S_5$		X				X

Final stage pipeline without clock edge

Page No.: 11

$$S_1 = 6-1 = 5 \quad T = 2 \text{ ns}$$

$$S_2 = 5-2 = 3 \quad \text{Time} = 1 \text{ ns}$$

$$S_3 = 3 \quad \text{Time} = 1 \text{ ns}$$

$$S_4 = 3 \quad \text{Time} = 1 \text{ ns}$$

$$S_5 = 6-2 = 4 \quad \text{Time} = 2 \text{ ns}$$

a) List the set of forbidden latencies and the collision vector

b) Draw a state transition diagram showing all possible initial sequence (cycles) without causing a collision in the pipeline.

c) List all the simple cycles from the state diagram.

d) Identify the greedy cycle among the simple cycles.

e) What is the min avg. latency (MAL) of this pipeline?

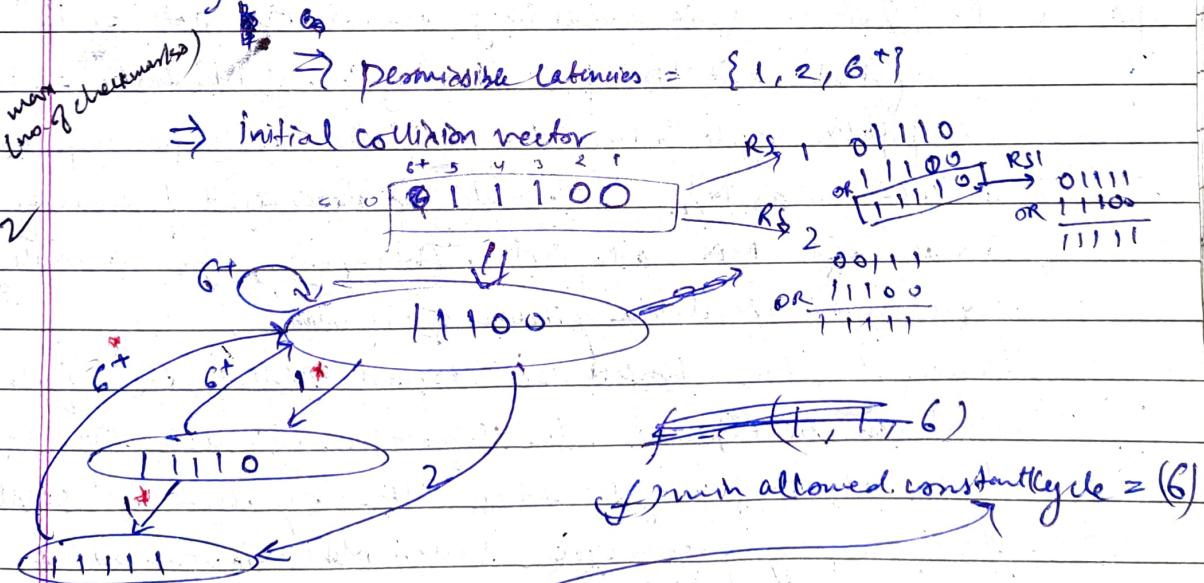
(f) min allowed constant cycle in this pipeline

(g) What will the max throughput of this pipeline?  $\frac{1}{T_{min}}$  = 6

(h) What will be the throughput if the minimum constant cycle is used?  $= \frac{1}{6} = 0.166$

Ans (a)

Forbidden latencies = {3, 4, 5}



(c) Simple cycle = (6), (1, 6), (1, 1, 6), (2, 6) A

(d) Greedy cycle = (1, 1, 6)

(Avg latency  $\leq 3(1) + 1$  in ICV)

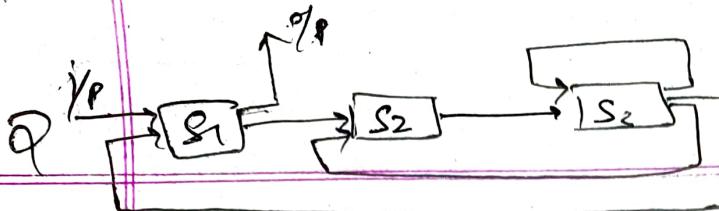
$$(e) MAL = \frac{1+1+6}{3} = \frac{8}{3} = 2.6$$

2) Avg Let  $\leq 4$

$$(f) \text{Max Throughput} = \frac{1}{\text{MAL}} = \frac{1}{2.6} = 0.385 \quad \text{Throughput} = \left[ \frac{n}{k + (n-1)T} \right] B = \frac{n}{k + (n-1)T} B = \frac{n}{k + nT - kT} B = \frac{n}{nT} B = \frac{1}{T}$$

$$(g) \left(\frac{1}{6}\right) = \text{Throughput} = 0.166$$

4 stages / or 2 forbidden latencies



	1	2	3	4	5	6	7	8
S1	X							
S2		X		X				
S3			X	X				
S4								
S5								

$$5-1 = 4$$

$$4-2 = 2$$

$$4-3 = 1$$

forbidden latency = {1, 2, 4}

permissible = {3, 5+}

initial collision vector =  $\begin{smallmatrix} 5 & 4 & 2 & 2 \\ 0 & 1 & 0 & 1 \end{smallmatrix}$

- 1011

$\Rightarrow 2 \leq MAL \leq 3+1$

$2 \leq MAL \leq 4$

simple cycle = (3), (5+)

greedy cycle = (3)

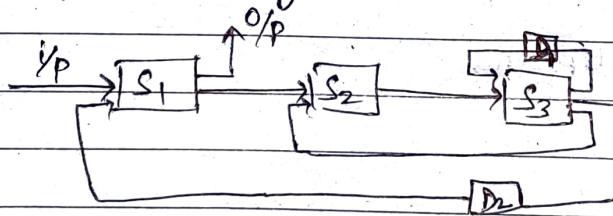
1011

3\*

$\Rightarrow MAL = 3$

make MAL reach its lower bound (2).

input delays :-



	1	2	3	4	5	6	7
S1	X			D2	D3	X	
S2		X		X			
S3			X	D1	X		
D1				D1			
D2					D2		

$$4-1 = 3$$

$$4-2 = 2$$

$$5-3 = 2$$

forb lat = {2, 3}

forb lat = {2, 3}

for permissible = {1, 3, 4, 5, 7+}

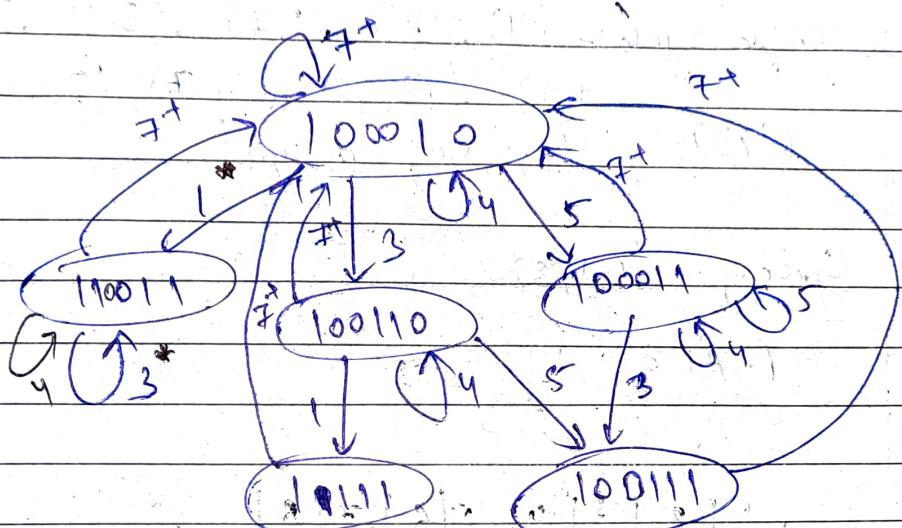
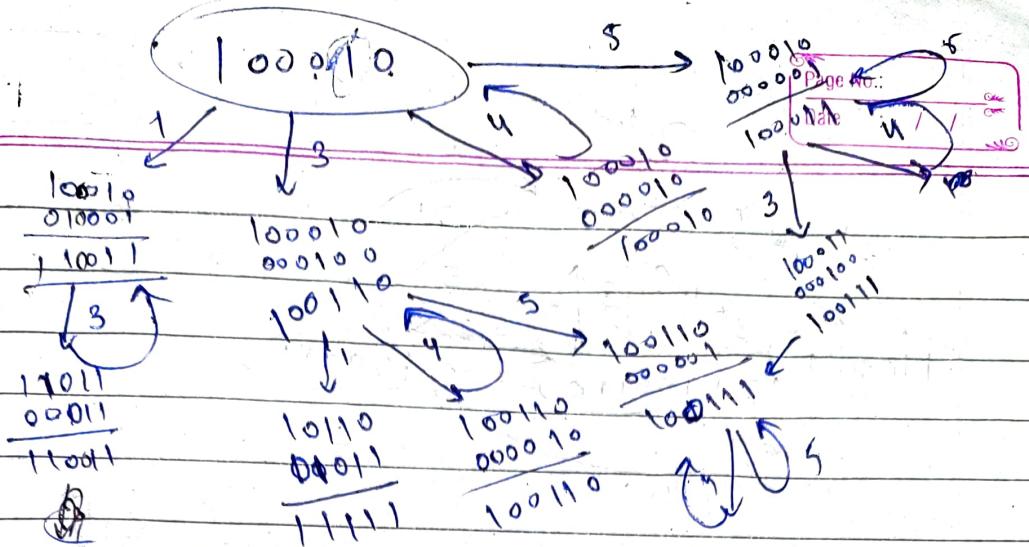
per = {1, 3, 4, 5, 7+}

ICV = 0100010

(100010)

$2 \leq MAL \leq 2$

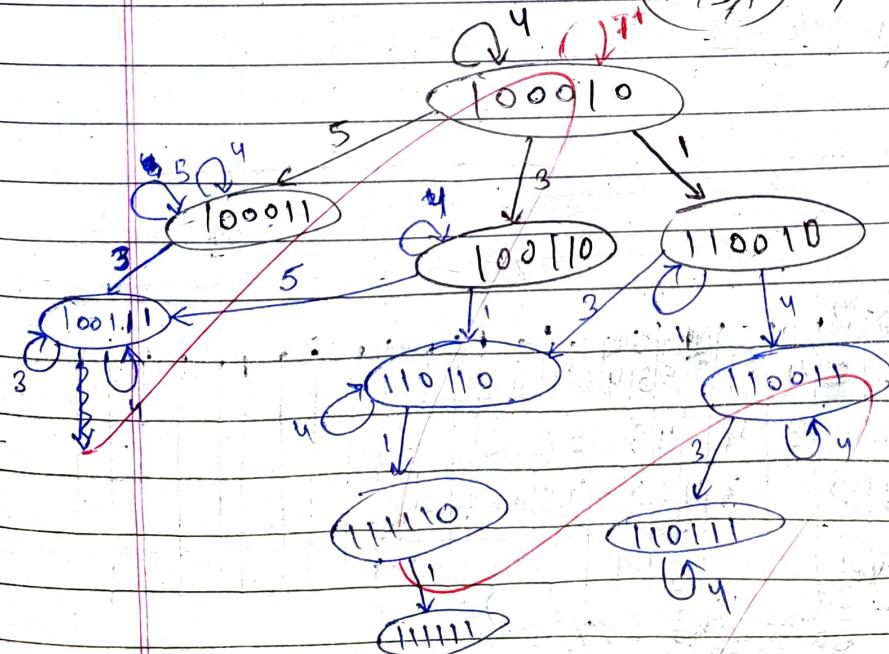
$\Rightarrow MAL = 2$

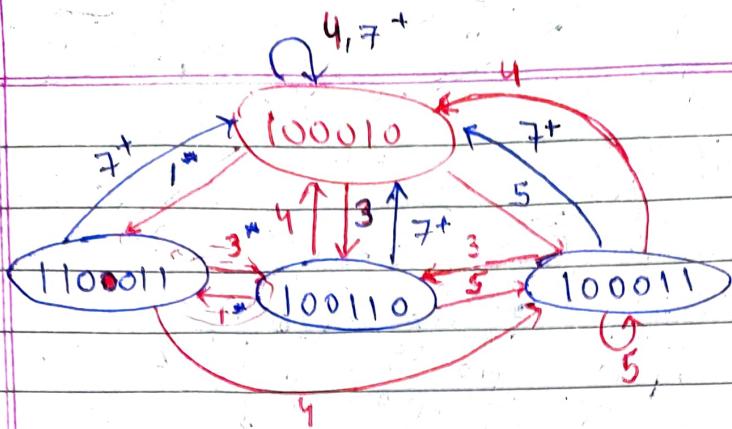


Simple cycle =  $(7)$ ;  ~~$(1, 3)$~~ ,  ~~$(1, 7)$~~ ,  ~~$(3, 7)$~~ ,  ~~$(5, 7)$~~   
 $(3, 1, 7)$ ,  $(3, 5, 7)$ ,  $(4)$ ,  $(5, 1, 7)$

$(5, 3, 7)$ , according to 1011

$2 \leq \text{MAL} \leq 4$





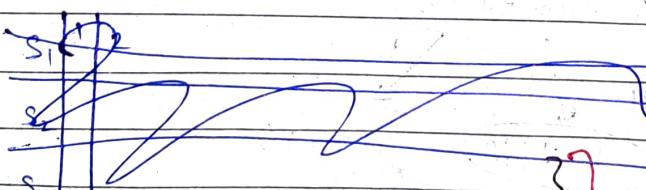
Simple cycles - (4), (7<sup>+</sup>), (1, 7), (3, 4), (5, 4), (5, 7), (3, 5, 4), (1, 3), (1, 3, 5, 4), (1, 4, 4)

Greedy = (1, 3)

$$MAL = \frac{1+3}{2} = 2$$

Q	1	2	3	4	5   6	7   8	
S <sub>1</sub>	x				x	x	
S <sub>2</sub>		x	x				
S <sub>3</sub>			x	x	x		

Latency cycle = 3



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
S <sub>1</sub>	1	2	1	3	1	2	4	3	5	3	4	6	4	5	7	3												
S <sub>2</sub>	1	1	2	2	3	3	4	4	4	5	5	6	6															
S <sub>3</sub>	1	1	2	1	2	3	3	4	3	4	5	4	5	6	6	5												

$$\text{efficiency} = \frac{8 \times 4 + 4 \times 1}{20 \times 3} = \frac{32}{60} = \frac{16}{30} = \frac{8}{15}$$

$$= \frac{32 + 12}{60} \times 100 = \frac{44}{60} \times 100 = \frac{22}{3} \% = 73.3\%$$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
$S_1$	$x_1$		$x_2$	$x_1$	$x_3$	$x_1$	$x_2$	$x_4$	$x_2$	$x_3$	$x_5$	$x_3$	$x_4$	$x_6$	$x_4$	$x_7$	$x_8$	$x_5$	$x_7$	$x_8$	
$S_2$		$x_1$	$x_1$	$x_2$		$x_2$	$x_3$		$x_3$	$x_4$		$x_4$		$x_5$	$x_6$		$x_6$	$x_7$	$x_6$	$x_7$	$x_8$
$S_3$		$x_1$	$x_1$	$x_2$	$x_1$	$x_2$	$x_3$	$x_2$	$x_3$	$x_4$	$x_3$	$x_4$	$x_5$	$x_4$	$x_5$	$x_6$	$x_5$	$x_6$	$x_7$	$x_8$	

8 filled out of 9

for latency cycle (3)

for latency cycle 0<sup>+</sup>

$$(18) \rightarrow \begin{cases} x_1 \rightarrow 1 \\ x_2 \rightarrow 9 \\ x_3 \rightarrow 10 \\ x_4 \rightarrow 18 \\ x_5 \rightarrow 19 \end{cases}$$

$$\begin{array}{l} x_6 \rightarrow 27 \\ x_7 \rightarrow 28 \\ x_8 \rightarrow 36 \end{array}$$

$$\text{Efficiency} = \frac{8}{9} \times 100 = 0.88$$

$$\Rightarrow 88.8\%$$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
$S_1$	$x_1$			$x_1$	$x_2$	$x_3$				$x_2$	$x_3$	$x_2$	$x_3$	$x_4$	$x_5$							
$S_2$		$x_1$	$x_1$			$x_2$	$x_3$	$x_2$	$x_3$					$x_4$	$x_5$	$x_4$	$x_5$	$x_6$				
$S_3$		$x_1$	$x_1$	$x_1$				$x_2$	$x_3$	$x_2$	$x_3$	$x_2$	$x_3$					$x_4$	$x_5$	$x_4$	$x_5$	$x_6$

1	23	24	25	28	27	28	29	30	31	32	33	34	35	36							
$S_1$	$x_4$	$x_5$	$x_4$	$x_5$	$x_6$	$x_7$				$x_6$	$x_7$	$x_6$	$x_7$		$x_8$	$x_7$	$x_8$	$x_7$	$x_8$		
$S_2$							$x_6$	$x_7$	$x_6$	$x_7$											
$S_3$		$x_5$	$x_4$	$x_5$				$x_6$	$x_7$	$x_6$	$x_7$	$x_6$	$x_7$								

$$\text{Efficiency} = \frac{16}{27} \times 100 = 59\%$$

$$\text{Efficiency} = \frac{16}{27} \times 100 = 59\%$$

or

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$S_1$	1	2			1	2	1	2	3	4			3	4	3	4	5	6							
$S_2$	1	2			1	2				3	4	3	4					5	6	5	6				
$S_3$	1	2	1	2	1	2				3	4	3	4	3	4				5	6	5	6	5	6	

$$\begin{cases} x_1 \rightarrow 1 \\ x_2 \rightarrow 2 \\ x_3 \rightarrow 10 \\ x_4 \rightarrow 11 \\ x_5 \rightarrow 19 \\ x_6 \rightarrow 20 \\ x_7 \rightarrow 28 \end{cases}$$

$$\text{Efficiency} = \frac{16}{27} \times 100 = 59\%$$

## Performance Loss

(Q1) A uniprocessor computer can operate in either scalar or vector mode. A certain benchmark program took time  $T$  to run on the computer. For that, it was found that 25% of  $T$  was attributed to vector mode. In the remaining time machine operated in scalar mode.

(Q2) Calculate the effective speedup under the given conditions as compared with when the vector mode is not used at all.

- Also calculate  $\alpha$  (the % of code that has been vectorized) in the above program.

In vector mode, computation can be done ~~at~~ faster than scalar mode.

$$\text{SpeedUp} = \frac{\text{Time without enhancement (i.e. scalar mode)}}{\text{Time with enhancement (i.e. system with vector computation)}}$$

$$= \frac{0.25T + \{0.25T \times g\}}{T} \quad \begin{matrix} \text{(corresponding scalar time} \\ \text{of vector)} \end{matrix}$$

$$= \frac{3T}{T} = 3$$

$$\alpha = \frac{0.25T \times g}{3T} = \frac{g}{9}$$

$$\begin{array}{c} \text{Total Time}(T) \\ \swarrow \alpha \quad \searrow \\ \text{sequential} \quad \text{parallel} \end{array}$$

$\nwarrow$ : Enhanced parallel processing  
 $\nearrow$ : .. of sequential

$$(1-\alpha)T \quad \frac{\alpha T}{n} \quad \begin{matrix} \text{with } n \text{ of processors from} \\ \text{parallel time decrease by } n \end{matrix}$$

$$\text{Time with enhancement} = (1-\alpha)T + \frac{\alpha T}{n}$$

$$\therefore \text{SpeedUp} = \frac{\text{Time without enhancement}}{\text{Time with enhancement}} = \frac{T}{(1-\alpha)T + \frac{\alpha T}{n}}$$

→ Jan 2011 No. 784

$$\boxed{\text{Speed Up} = \frac{1}{1 - \alpha + \frac{\alpha}{n}}}$$

→ This is Amdahl's law.

if  $f$  = static Port no, put  $1 - \alpha = f \Rightarrow \alpha = 1 - f$

$$\boxed{\text{Speed Up} = \frac{1}{f + \frac{1-f}{n}}}$$

$$S = \frac{1}{1 - \alpha + \frac{\alpha}{q}} \Rightarrow \alpha = \frac{3}{q}$$

# Generalized Amdahl's law  
 enhancement of  $\alpha_1$   $\xrightarrow{\text{hardware improvement}}$   
 " 2  $\alpha_2$   $\xrightarrow{\text{hardware improvement}}$   
 " "  
 enhancement in  $\alpha_n$   $\xrightarrow{\text{hardware improvement}}$

$$\text{overall speed up} = \frac{1}{1 - (\alpha_1 + \alpha_2 + \dots + \alpha_n) + \frac{\alpha_1}{f_1} + \frac{\alpha_2}{f_2} + \dots + \frac{\alpha_n}{f_n}}$$

fraction of reduced hardware enhancement

Q. ② Suppose we double the speed ratio b/w vector mode & scalar mode and by hardware improvement. Calculate the effective speedup factor can be achieved.

Q. ③ Suppose the same speedup obtained in part (b) by compiler must be obtained by compiler improvement.

reduced execution

2) Fraction of time no enhancement is used =

$$\frac{1 - (\alpha_1 + \alpha_2 + \dots + \alpha_m)}{1 - (\alpha_1 + \alpha_2 + \dots + \alpha_m) + \frac{\alpha_1}{S_1} + \frac{\alpha_2}{S_2} + \dots + \frac{\alpha_m}{S_m}}$$

ii) Suppose we double the speed ratios b/w the vector work and the scalar work by hardware improvement. Cal. the effective speedup that can be achieved.

iii) Suppose the same speedup obtained in part (ii) must be obtained by compiler improvement instead of hardware improvement. What could be new vectorization ratio  $\alpha$  that should be supported by the vectorizing compiler to achieve the same enhancement.

$$\text{Speedup} = \frac{1}{1 - \alpha + \alpha}$$

$$= \frac{1}{1 - 0.75 + \frac{0.75}{18}}$$

$$\text{Speedup} = 3.43$$

3.25 Let no of instructions to be executed  
 =  $K$  millions

then Parallel mode      sequential mode  
 $\alpha K$                          $(1-\alpha)K$

$$T = \frac{\alpha K}{mn} + \frac{(1-\alpha)K}{n}$$

$$\boxed{T = \frac{K(\alpha + (1-\alpha)m)}{mn}}$$

$$\Rightarrow \text{Effective MIPS} = \frac{K}{K[\alpha + (1-\alpha)m]} \cdot \frac{mn}{mn}$$

$$\boxed{\text{effective MIPS} = \frac{mn}{\alpha + (1-\alpha)m}}$$