Date: 19/03/19

Q1 → CPI (cycles per insts.)

→ 1.5 GHz means 0.67ns per cycle. 0.67 × 5 = 3.33 ns/insk.

→ Effective CPI = Base CPI + Stall CPI

↓ if everything works fine always '1'.

↓ depends on hazards etc.
↳ memory stalls + branch stalls + load-ALU stalls

⇒ 1 + (0.3 × 0.05 × 50) + (0.2 × 0.3 × 2) + (0.1 × 1) ⇒ 1.97

→ 1 GHz after pipelining. ie. 1 × 1.97 = 1.97 ns/insts.

→ Speedup = $\frac{3.33}{1.97}$ = 1.69

↳ $\frac{\text{Ex\_up}}{\text{Ex-p}}$

Q2 without operand forwarding

ID of n^{th} inst. can happen only after WB of $n-1^{th}$ insts.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
L.D	IF	ID	EX	ME	WB									
ADD.D		IF	-	-	-	ID	EX	ME	WB					
L						IF	-	-	-	ID	EX	ME	WB	
A										IF	-	-	-	ID

total 8001 cycles.

CPI = $\frac{8001}{2000}$ = 4.0005

Date:

with op forwarding

in LD.D , op is ready at end of MEM and passed to EX of ADD.D

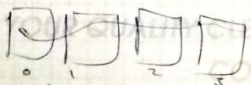
in ADD.D, op is ready at end of EX stage and passed to EX of LD.D .

	1	2	3	4	5	6	7	8	9	10	11	12	13
LD	IF	ID	EX	MEM	WB								
ADD		IF	–	ID	EX	ME	WB						
LD				IF	ID	EX	ME	WB					
				IF	–	ID	EX	ME	WB				

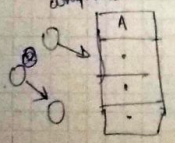
CPI = $\frac{3004}{2000}$ = 1.502

Q3

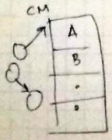
① Pseudo LRU



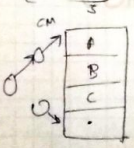
comp. miss



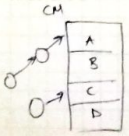
CM



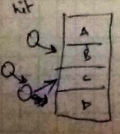
CM



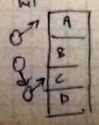
CM



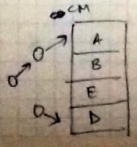
hit



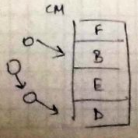
hit



CM

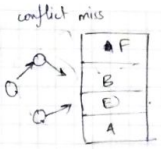


CM



Date:

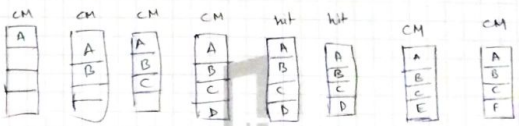
conflict miss



miss rate : $\frac{\text{comp. miss + conflict miss}}{\text{total memory access}}$ = $\frac{6 \text{ + } 3}{14}$ = $\frac{9}{14}$

② LIFO

CM CM CM CM hit hit CM CM



~~6~~ 6 CMP- miss

1 comp. - miss

7 hits

Q4

32 bit processor i.e. word length is 32-bit (4 byte)

8 x 1024 bit I-cache instruction : 22, 23, 24, 25
data : 260, 261, 273

L1 I-cache

L1 D-cache

L2 of chip

main memory to L2 happens then instr. go to I-cache
data go to D-cache

Date:

L-2

(block size is 16 words)

22/16, 23/16, 24/16, 25/16 = B1 block

260/16, 261/16 = B16 block, 275/16 = B17 block

these blocks of MM are based on L-2 cache.

# of sets in L-2 : $\frac{\text{cache size}}{\text{blocksize} \times \text{associativity}}$ = 512 sets

B1 % 512 = Set 1 = S1

B16 % 512 = Set 16 = S16

B17 % 512 → Set 17 = S17

B1 is fetched and kept in any 1 of the 4 blocks in S1. usually in way-0 as all 4 are empty initially

L-1 I-cache

(block size is 8 words)

22/8, 23/8 = B2

24/8, 25/8 = B3

sets in I-cache = $\frac{\text{cache size}}{\text{block size} \times \text{associativity}}$ = 256 sets

(associativity → 1 as it is direct mapping)

B2 % 256 — S2

B3 % 256 — S3

Date:

L-1 D-cache

260/8 = B32

261/8 = B32

275/8 = B34

sets = 256 sets (associativity = 2)

B32 % 256 → S32

B34 % 256 → S34

only 1 block of S32 and S34 is occupied.

Q5

I-cache miss rate = 2%

D-cache " " = 4%

CPI = 2 without memory stalls

miss penalty = 100 cycles for all misses

D-cache is accessed only in MEM stage.

36% are load/store → D-cache

IMP

D-cache : 36 out of 100 inst. $\frac{36}{136}$

I-cache : 100 out of 100 inst. $\frac{100}{136}$

CPI ideal = Base CPI = 2

Actual CPI = Base CPI + stall CPI

Stall CPI = (% use of I-cache × stall of IC) + % use of DC × stall of DC

% use of I-cache = $\frac{100}{136}$

miss rate × miss penalty = 0.02 × 100

% use of DC = $\frac{36}{136}$

miss rate × miss penalty = 0.04 × 100

Date:

$$\text{Speedup} = \frac{2 + \text{Stall CPI}}{2} = \frac{\text{real CPI}}{\text{ideal CPI}}$$

Q6

Avg. memory access time

$$\text{AMAT} = \text{hit time} + (\text{miss rate} \times \text{miss penalty})$$

hit time: ↓ in ns; also called access time

miss penalty: ↓ in ns

Q8

block size = 128 B = 32 words

miss penalty = $50 + (31 \times 10) = 360$