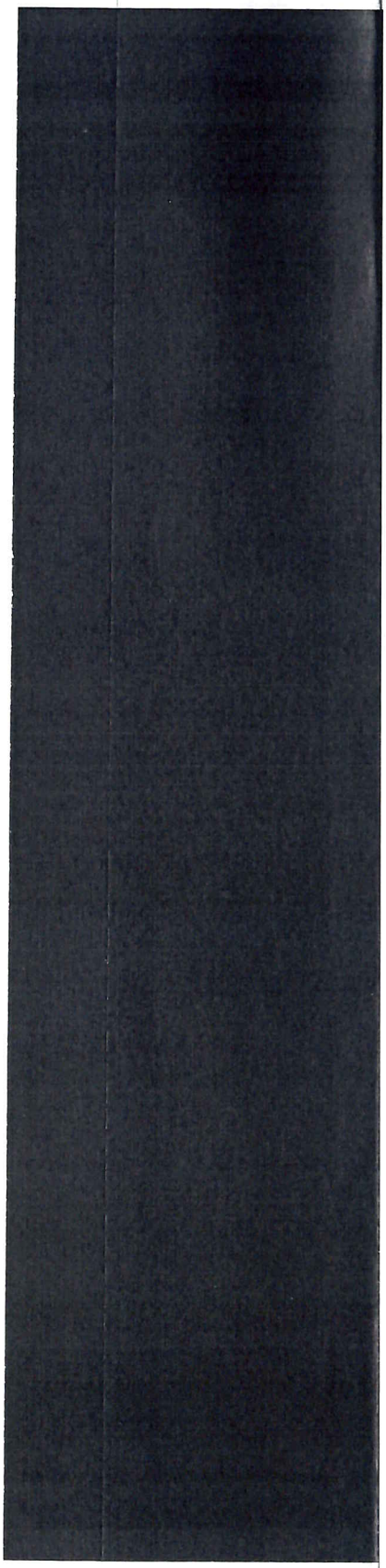
CHALMERS
EXAMINATION / TENTAMEN

Course code/kurskod	Course name/kursnamn			
DAT105	Computer Architecture			
Anonymous code Anonym kod		Examination date Tentamensdatum	Number of pages Antal blad	Grade Betyg
DITOSI-0004-GDT		23/10/23	13	

* I confirm that I've no mobile or other similar electronic equipment available during the examination.
Jag intygar att jag inte har mobiltelefon eller annan liknande elektronisk utrustning tillgänglig under eximinationen.

Solved task Behandlade uppgifter No/nr		Points per task Poäng på uppgiften	Observe: Areas with bold contour are to completed by the teacher. Anmärkning: Rutor inom bred kontur ifylles av lärare.
1	X	11	
2	X	10	
3	X	9	
4	X	9	
5	X	12.	
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
Bonus poäng			
Total examination points Summa poäng på tentamen			



3

(1A) Arithmetic mean of execution times for P1-P3 of A, B? Fastest?

$$T_{\bar{A}} = \frac{2+10+2}{3} = \frac{14}{3} = 4,67\,s$$ VII

For computer model B, we don't have the execution times but we can calculate them taking into account the speedup over machine A.

$$S = \frac{T_A}{T_B} \quad ; \quad T_B = \frac{T_A}{S}$$

$$T_{B1} = \frac{2}{0.5} = 4$$

$$T_{B2} = \frac{10}{5} = 2$$

$$T_{B3} = \frac{2}{0.5} = 4$$

$$T_{\bar{B}} = \frac{4+2+4}{3} = \frac{10}{3} = 3,33\,s$$ //

(1A) $$S = \frac{T_{\bar{B}}}{T_{\bar{A}}} = \frac{3.33}{4.67} = 0,71$$

Machine B is faster than the Machine by around a 29%.

Machine B is farter because even though P1 and P3 are slower, there is a significant speedup on P2 over machine A.

3

(1B) Geometric mean speedup? Fastest? Consistent with 1A?

First, we will need to calculate the speedup of B over reference machine R:

$$S_B = \frac{T_R}{T_B} ;$$

We can get T_R through the speedup of A over R (SA): $S_A = \frac{T_R}{T_A} \Rightarrow T_R = S_A \cdot T_A$

$$S_{B1} = \frac{2 \cdot 10}{4} = 5s$$

$$S_{B2} = \frac{10 \cdot 10}{2} = 50s$$

$$S_{B3} = \frac{2 \cdot 10}{4} = 5s$$

Then, we calculate the geometric mean speedup:

(4)

$$S_A = \sqrt[3]{10 \cdot 10 \cdot 10} = 10$$

$$S_B = \sqrt[3]{5 \cdot 50 \cdot 5} = 10,77$$

$$S = \frac{10}{10.77} = 0.93$$

Machine B is still faster than A but only by a 7%. This is not consistent with the 29% from 1A, and this is due because the geometric mean tends to soften the impact of the outliers.

(1C) MPKI for P1 and P2 on A and B.

Both machines have 1 GHz operating frequency.

The number of instruction cache acceses correspond to the number of instructions.

First, we calculate the number of cycles it would take for each of the programs to execute on each of the machines. Then, we will calculate the total time spent servicing misses and from there we will be able to determine the number of misses.

1 GHz = 10^9 cycles/s
Miss penalty = 100ns
} For both machines

1 GHz ⇒ 1 ns/cycle

For A: $T_1 = 2s$, $T_2 = 10s$ 8 clocks / instruction

Process 1

$N_{cycles} = 10^8 \cdot 8$

$T_{exe} = 8 \cdot 10^8 \cdot 1 = 8 \cdot 10^8$ ns

$2 \cdot 10^9 - 8 \cdot 10^8 = 1.2 \cdot 10^9$ ns servicing misses

$\frac{1.2 \cdot 10^9}{100} = 1.2 \cdot 10^7$ total misses

$\frac{1.2 \cdot 10^7}{10^8} = 0,12$ misses/instr

⇒ 120 MPKI

Process 2

$N_{cycles} = 10^9 \cdot 8$

$T_{exe} = 10^9 \cdot 8 \cdot 1 = 8 \cdot 10^9$ ns

$10 \cdot 10^9 - 8 \cdot 10^9 = 2 \cdot 10^9$ ns

$\frac{2 \cdot 10^9}{100} = 2 \cdot 10^7$ total misses

$\frac{2 \cdot 10^7}{10^{10^9}} =$ 0,02 0,2 miss/instr

⇒ 200 MPKI 20 MPKI

3

1C

For B: 1 clock/instr

Process 1

$T_1 = 4$

$N_{cycles} = 10^8 \cdot 1$

$T_{exe} = 10^8$ ns

$4 \cdot 10^9 - 10^8 = 3.1 \cdot 10^9$ ns servicing misses

$\frac{3.1 \cdot 10^9}{100} = 3.1 \cdot 10^7$ misses

$\frac{3.1 \cdot 10^7}{10^8} = 0.31$ misses/instr

=> 310 MPKI

Process 2

$T_2 = 2$

$N_{cycles} = 10^9 \cdot 1$

$T_{exe} = 10^9$ ns

$2 \cdot 10^9 - 10^9 = 10^9$ ns

$\frac{10^9}{100} = 10^7$ misses

$\frac{10^7}{10^9} = 0,01$ miss/instr

=> 10 MPKI

3)

3

2A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
I1	IF	ID	EX	ME	WB															
I2		IF	ID	EX	ME	WB														
I3			IF	ID	FP1	FP2	FP3	FP4	FP5	ME	WB									
I4				IF	ID	-	-	-	FP1	FP2	FP3	FP4	FP5	ME	WB					
I5					IF	-	-	-	ID	EX	ME	WB								
I6						IF	-	-	-	ID	-	-	-	EX	ME	WB				

from the ID stage of I1 to the ME-stage of I6, it will take 15 cycles.

(2B)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
I1	IF1	IF2	ID	EX1	EX2	ME1	ME2	WB															
I2		IF1	IF2	ID	EX1	EX2	ME1	ME2	WB														
I3			IF1	IF2	ID	ID	ID	FP1	FP2	FP3	FP4	FP5	ME1	ME2	WB								
I4				IF1	IF2	-	-	ID	-	-	-	-	FP1	FP2	FP3	FP4	FP5	ME1	ME2	WB			
I5					IF1	IF2	-	-	-	-	-	-	ID	EX1	EX2	ME1	ME2	WB					
I6						IF1	IF2	-	-	-	-	-	-	ID	-	-	-	EX1	EX2	ME1	ME2	WB	

Structural Hazard!

It takes 18 cycles from the ID of I1 to ME1 of I6, that is, 3 cycles more than the 2A pipeline, but as we have double the frequency, this pipeline will be faster, because we can execute double the cycles in the same time.

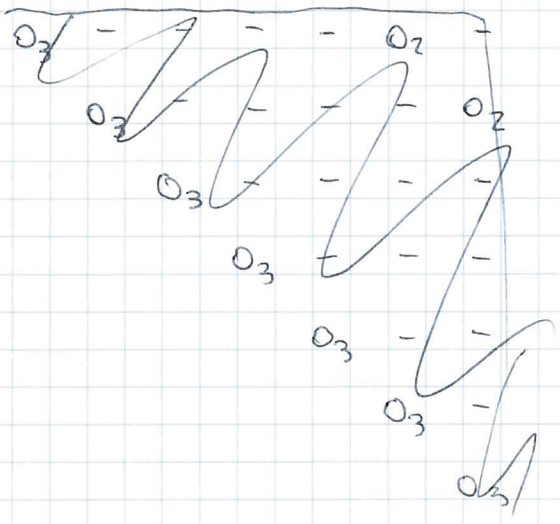
3

3

(2C)

```
LOOP: LD   F0, 0(R1)      O1
      ADD  F4, F0, F1     O2
      SD   F4, 0(R1)      O3
      ADD  R1, R1, #8
      SUBI R3, R3, #1
      BNEZ R3, LOOP
```

	C1	C2	C3	C4	C5	C6	C7	C8	
I_1	O_1								prolog
I_2	-	O_2							
I_3	O_2	-	O_1						
I_4	-	O_2	-	O_1					
I_5	-	-	O_2	-	O_1				
I_6	-	-	-	O_2	-	O_1			
I_7	-	-	-	-	O_2	-	O_1		
I_8	O_3	-	-	-	-	O_2	-	O_1	Kernel



(3)

Instructions wood for the Kernel?

3

9

(3A)

i) The ROB tracks the instructions so they ~~are issued~~ respect the data dependencies. The RAT assigns an entry to each of the ~~registers~~ instructions in the ROB, say i1 and i2, so as soon as the first register is available ~~the~~ we will be able to use it for the second instruction. fetching it from the ROB. (2)

ii) We will have an entry for R1 on I1, I2, I3.

(3C) Three instructions will have been executed. As in the first iteration we will have misspredicted the branch, we will have to flush the ROB ~~of~~ after the branch along with the pipeline. (2)

(3D) We will misspredict twice (1st and 2nd iterations) and once on the last one, so out of 100 times, we will miss $\frac{3}{100}$ = 3% of the time, and a 97% of the time we will have done a correct branch predichor. (2)

3

(3B)

	IF	ID	IS	Ex_1	Ex_2	COB	CT
I1		0	1	2	2	3	4
I2	0	1	~~3~~	~~4~~	6	~~7~~	8
I3	1	2	7	8	9	–	
I4	2	3	4	5	5	6	9
I5	3	4	5	6	6	7	10
I6	4	5	6	6	7	8	11

3.

3

(4A) 9 Exclusive memory-hierarchy (2)

i) False. We cannot find any block existing on both levels at the same time.

ii) True. For exclusive memory-hierarchy, blocks on the upper level won't be found on a lower level, as only one copy of a block is possible at the same time.

iii) False. Same as i)

iv) True. Same as ii), we can only have 1 copy of a block at the same time.

In conclusion:
For exclusive memory hierarchy, we can only have one copy of a block in the cache. This copy can be either on the upper level or the lower level, but not both at the same time.

(4B) First, we will need to calculate the execution time in cycles assuming a blocking cache. (4)

Every four iterations of the for, we will experience a cache miss, so $5 \cdot (3 + 24) = 390$ (3 iterations → 1st it) will be the cycles it will take to execute them. In total, $390 \cdot 250 = 9750$ cycles.

For the non-blocking cache, we will only miss on the first iteration, so we will have again $24 + 3 \cdot 5 = 39$ for the 4 first iterations. After the initial miss we won't be missing anymore, as with the non-blocking cache we will be able to prefetch the instructions before they are needed.

3

(4B) In this way, we will be able to drastically reduce the number of cycles spent servicing misses. For the rest of the iterations (that is, 996), all of the instructions will be executed in $996 \cdot 5 = 4980$ cycles.

In total, $4980 + 39 = 5019$ cycles

The speedup of the non-blocking cache over the blocking one is $\frac{9750}{5019} = 1.94$

So, we can affirm that the non-blocking cache is almost 2 times faster than the blocking one.

(4C)

Replace					1	2	3	4	1	2
Sequence	1	2	3	4	5	6	7	8	9	10
Outcome	M	M	M	M	M	M	M	M	M	M

Replace	9	10	7	8	1	2	17	18
Sequence	1	2	3	4	17	18	1	2
Outcome	M	M	M	M	M	M	M	M

Cold misses: Number of total uniques blocks accessed.

In this case, (12).

Capacity misses: Misses that would not happen in an infinite cache.

In a fully associative cache with OPT, total − cold = 12 − 12 = (0)

Conflict misses: Total − cold − capacity = 18 − 12 − 0 = (6)

3

(5A) Write-back policy

	x (P1)	x (P2)	X
R1	[0]	–	0
W1=1	1	–	0
R2	1	[0]	0
W2=2	1	2	0
R2	1	[2]	0
R1	[1]	2	0

R_2 will return 2, while R_1 will return 1. This is not correct because after a write to the same variable, we would expect that both processors would return the same value, but both processors are reading their own private copies of X.

✓ (3)

(5B) We can make sure that the correct value is returned to processor 2 by invalidating the old values of X on a write request, In this way, we force processor 2 to get a ~~read of 2 X from~~ valid copy of X whenever P2 wants to issue a write or request.

✓ (3)

3

(5C) A thread switch stage (TS in short) must be implemented between the fetch and decode stages.

We will also need a thread switch which will decide the thread to be executed on each instance.

Also, we will need N program counters (in our case, 4) and the same number of register files, one for each thread.

Each thread will be accompanied by a thread ID to manage data hazards.

6