Name: Student ID:	Computer Organization, Spring 2024 Final (140pt)

1. [10pt] Assume a processor's clock cycle time and the clock cycles per instruction both increase by 30%, and the total number of instructions to be executed decreases by 40%. To assess how will the above changes affect the overall execution time, please calculate the new execution time as a relative value to the original execution time? (Hint: Please compute the performance ratio of the new execution time to the original execution time.)

   執行時間變為原來的 1 x 0.6 x 1.3 x 1.3 = 1.014

2. [20pt] Assume there is a unified L1 cache, containing both instruction and data caches, attached to a pipelined processor, and the next lower level of the memory hierarchy is the main memory. A write-through scheme is implemented on this hierarchy to ensure data consistency between the cache and the main memory. The performance characteristics of the processor are as follows.
   - The base CPI with a perfect memory system = 1.5
   - The L1 cache miss rate = 5%
   - The memory access latency = 100 cycles
   - 30% of the instructions are data transfer instructions

   (1) [5pt] Please determine the effective CPI for the memory system.

   (2) [5pt] What can be added in the memory system design to reduce the memory write latency in the write-through scheme? (Hint: an extra hardware resource can be added.)

   (3) [5pt] What is the average memory access time when the processor has the clock cycle time of 1ns and the cache hit time of 1 cycle?

   (4) [5pt] Which of the following methods can be used to reduce the cache miss rate? (Hint: All choices must be correct, losing points for wrong answers)
   - (a.) Increase cache block size
   - (b.) Increase cache size
   - (c.) Increase associativity
   - (d.) Increase CPU clocking
   - (e.) Compiler optimization

   2.1

   write miss penalty = 100

   read miss penalty = 100

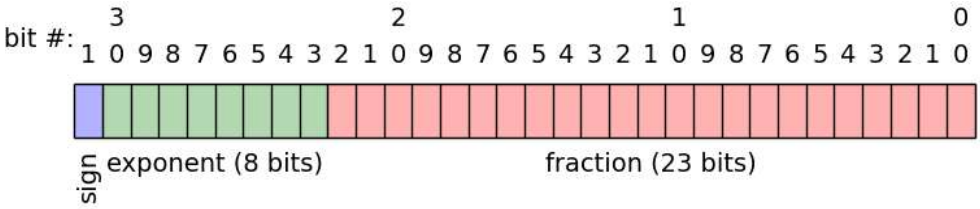
   effective = 1.5 + (1*0.05*100) + (0.3*0.05*100) = 8

2.2 write buffer

2.3 1+0.05*100

2.4 a,b,c,e

3. [10pt] The IEEE Standard for Floating-Point Arithmetic (IEEE 754) is a technical standard for floating-point arithmetic established in 1985 by the Institute of Electrical and Electronics Engineers (IEEE). This standard addresses many problems found in the diverse floating-point implementations that made them difficult to use reliably and portably. Many processors conform to the IEEE 754 standard for representing the floating-point numbers. For example, the bit sequence of -0.75 in a 32-bit register based on IEEE754 is 1011111101000...00. Please write down the hexadecimal strings of 65,536 and -3.25, respectively, using the IEEE 754 standard to represent floating-point data in a 32-bit register.



0x47800000

0xc0500000

4. [25pt] Considering an L1 cache design, there are a total of four cache blocks in this cache. Each block is one *word* long. The least recently used policy is adopted for selecting a victim cache block. Three cache variants are shown below.

Cache design 1

Mem. block addr.	Cache index	Hit/miss	Cache content after access			
			0	1	2	3
40	0	Miss	M[40]			
12	0	Miss	M[12]			
40	0	Miss	M[40]			
26	2	Miss	M[40]		M[26]	
12	0	miss	M[12]		M[26]	

Cache design 2

Mem. block addr.	Cache index	Hit/miss	Cache content after access			
			Set 0		Set 1	
40	0	miss	M[40]			
12	0	miss	M[40]	M[12]		
40	0	hit	M[40]	M[12]		
26	0	miss	M[40]	M[26]		
12	0	miss	M[12]	M[26]		

Cache design 3

Mem. block addr.		Hit/ miss	Cache content after access			
			Block 0	Block 1	Block 2	Block 3
40		miss	M[40]			
12		miss	M[40]	M[12]		
40		hit	M[40]	M[12]		
26		miss	M[40]	M[12]	M[26]	
12		hit	M[40]	M[12]	M[26]	

(1) [3pt] What are the names of the three cache designs?

Chapter 5 - PowerPoint p30

Direct mapped cache / Two-way set associative cache / Fully associative cache

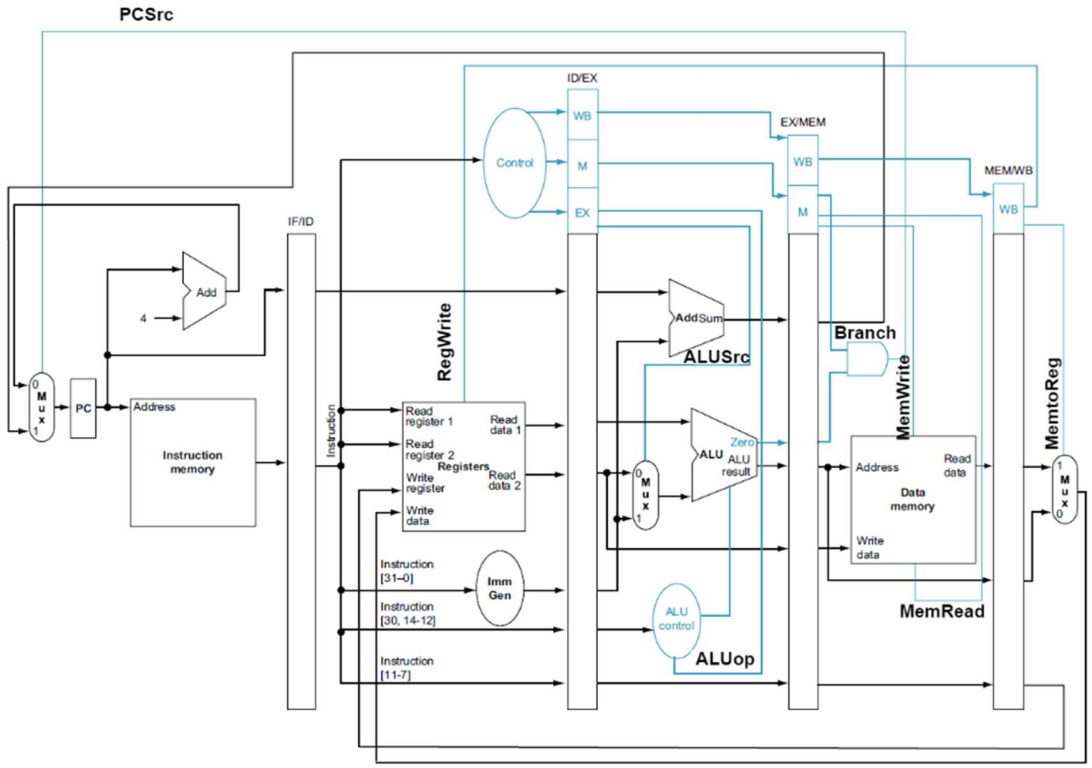
(2) [11pt] Given the access sequence of memory blocks, 40, 12, 40, 26, and 12, please fill out the blanks in the three cache designs. (Hint: You are required to give the answers for “cache index”, “hit/miss”, and “cache content after access”.)

(3) [6pt] Based on the above result, please report the numbers of cache hit and miss for each cache design. (Hint: your answer will look like this: Cache 1 with 2 hits/2 misses, Cache 2 with 2 hits/2 misses, and Cache 3 with 2 hits/2 misses.)

(4) [5pt] When the first-in-first-out policy is used for selecting a victim cache block, which cache design(s) is affected, in terms of the number of cache misses and hits? (Hint: your answer should be looked like this: The performance of Cache 1 is affected and it has 1 miss/2 hits.)

Two-way set associative cache - 3 misses / 2 hits

5. [60pt] The pipelined datapath with the control signals for a RISC-V processor is illustrated in the figure below. Please answer the following questions based on this RISC-V processor design.



(1) [20pt] Based on the information listed in the table (for ALUop), please help specify the values of the control signals (bits) generated by the Control unit in the above figure. The values of the control bits could be '0', '1', or 'X' (don't care condition).

Instruction opcode	ALUOp	Operation
lw	00	load word
sw	00	store word
beq	01	branch if equal
R-type	10	add
R-type	10	sub
R-type	10	and
R-type	10	or

[10pt] Please write down the control values for the lw instruction.

ALU Op (2 bits)	ALU Src	Branch	MemRead	MemWrite	RegWrite	MemtoReg
00	1	0	1	0	1	1

[10pt] Please write down the control values for the and instruction.

ALU Op (2 bits)	ALU Src	Branch	MemRead	MemWrite	RegWrite	MemtoReg
10	0	0	0	0	1	0

(2) [10pt] Given the instruction sequence in the following table, please identify the instruction(s) and the register(s) incurring data hazards. (Hint: Two data hazards exist.)

Seq. #	Instruction Seq.	CC1	CC2	CC3	CC4	CC5	CC6	CC7	CC8
1	addi x11, x11, 1	IF	ID	EX	MEM	WB			
2	addi x2, x2, 4		IF	ID	EX	MEM	WB		
3	lw x4, 0(x2)			IF	ID	EX	MEM	WB	
4	add x6, x4 ,x0				IF	ID	EX	MEM	WB

5.2

Chapter 4 - PowerPoint p66

data hazard 1 is between instructions 2 and 3, from x2

data hazard 2 is between instructions 3 and 4, from x4

(3) [10pt] Hardware and software techniques are available to alleviate the performance overhead (i.e., stalled cycles) incurred by the data hazards. These techniques can be added to the RISC-V processor design to accelerate the execution performance. Please answer the following questions.
[3pt] What is the hardware technique for the performance optimization?
[3pt] What is the software technique that can be used to alleviate the performance overhead?
[4pt] Assuming the hardware optimization technique is present, what is the instruction sequence optimized by the above software technique to achieve the best execution performance?

5.3

Forwarding hardware;

Code Scheduling;

addi, lw, addi, add (inst. seq. of 1, 3, 2, 4)

Instruction 2 接著 instruction 3 會發生的 load use data hazard，在使用 forwarding 之後仍然需要 stall 一個 clock cycle

(4) [20pt] Control hazards can affect the processor performance. Please answer the following questions.
[2pt] Assuming another instruction sequence lists in the following table, in which clock cycle (e.g., CC1) is the "branch decision" for the beq instruction determined? (Hint: The worst case scenario is assumed.)

5.4.1 CC4

[10pt] A branch predictor can be added to improve the pipelined execution of the RISC-V processor. Given a one-bit branch predictor, where the predictor is initialized as *taken* and the prediction is based

on the history, what is the prediction result of the following sequence for a branch instruction: T, N, N, T, N, N, T, N?

5.4.2 T-T-N-N-T-N-N-T

Correct or not: Y-N-Y-N-N-Y-N-N

[6pt] What is the accuracy of the branch predictor? (Hint: the number of correct predictions divided by the number of total predictions.)

5.4.3 accuracy: 3/8 (3 correct predictions out of eight shots)

[2pt] Given a perfect branch predictor with 100% accuracy, how many stalled cycles will a branch instruction cause on average?

Seq. #	Instruction Seq.	CC1	CC2	CC3	CC4	CC5	CC6	CC7	CC8	CC9
1	beq x1, x0, 20	IF	ID	EX	MEM	WB				
2	addi x11, x11, 1		IF	ID	EX	MEM	WB			
3	addi x2, x2, 4			IF	ID	EX	MEM	WB		
4	lw x4, 0(x2)				IF	ID	EX	MEM	WB	
5	add x6, x4 ,x0					IF	ID	EX	MEM	WB
6	and x14, x10, x0						IF	ID	EX	MEM

5.4.4 stalled cycles: 0 cycles

6. [15%] Assume two RISC-V processor models are available: one is a quad-core processor, and the other is a single-core processor with an 8-lane SIMD engine. These two processor models share the same instruction set architecture and both have the same pipeline design for in-order execution. In order to evaluate the hardware software interactions, we consider a matrix multiplication program in C, and a sequential version of the C-based program is illustrated below. The program can be divided into four code segments, S1, S2, S3, and S4. The execution time of the sequential version measured on the single-core RISC-V processor is: 2ms for S1, 8ms for S2, 16ms for S3, and 2ms for S4. Moreover, we have two processors, one for the quad-core processor and the other for the single-core, 8-lane SIMD processor. Besides, the matrix multiplication program has been converted into parallel versions for these two processors, respectively. Please answer the following questions.
   (1)[6pt] Which type of parallelism is exploited by each of the two processors?
   (2)[6pt] Please compute the speedups of the parallelized programs on these two processors, respectively.

(3)[3pt] Please determine which processor is faster, based on the answers in the previous question?

6.1 Task-level parallelism for 4-core. Data-level parallelism.

6.2 Task-level parallelism for S2 and S3. Speedup = (2+8+16+2)/(2+8+16/4+2) = 1.75

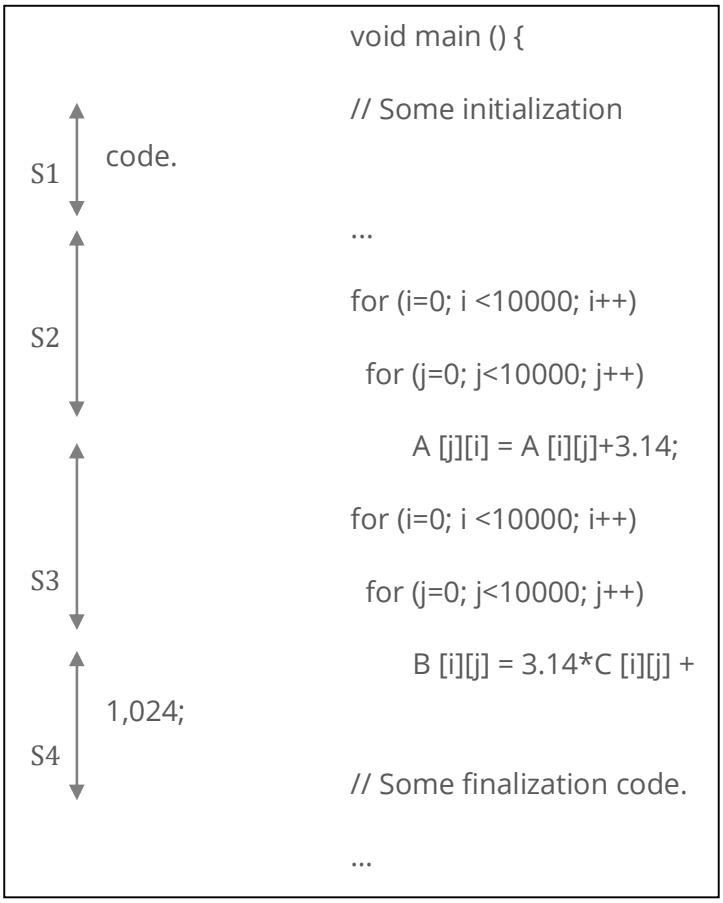
Data-level parallelism for S2 and S3. Speedup = (2+8+16+2)/(2+8+16/8+2)=2

S1 S4 無可加速空間,

S2 IJ 為相反的部分導致 8-lane SIMD 無法加速，因為若啟用加速可能會導致 data hazard，所以 S2 不啟用 SIMD 的加速

S3 可加速

6.3 SIMD engine is better.


void main () {
// Some initialization
S1
code.
...
S2
for (i=0; i <10000; i++)
for (j=0; j<10000; j++)
A [j][i] = A [i][j]+3.14;
S3
for (i=0; i <10000; i++)
for (j=0; j<10000; j++)
B [i][j] = 3.14*C [i][j] +
1,024;
S4
// Some finalization code.
...