Written test on 11^{th} of December 2019

Course: Computer Architecture and Engineering Course no. 02155

Exam time: 2 hours

No aids – Allowed: pocket calculator

Weighting: Each of the 5 problems counts 20%. All questions in a problem are weighted equally.

Answers are evaluated based on correctness, completeness, and conciseness (i.e., answers must be correct, show how the result is obtained, and contain only the necessary text). If some specification is missing, just state your own specification and continue the problem according to your stated specification.

Problem 1

Answer the following questions.

Question 1.1

Why is it important that the source register addresses are always at the same position in the instruction encoding of the RISC-V instruction set?

Question 1.2

Name the two types of locality and give one example for each when it happens.

Question 1.3

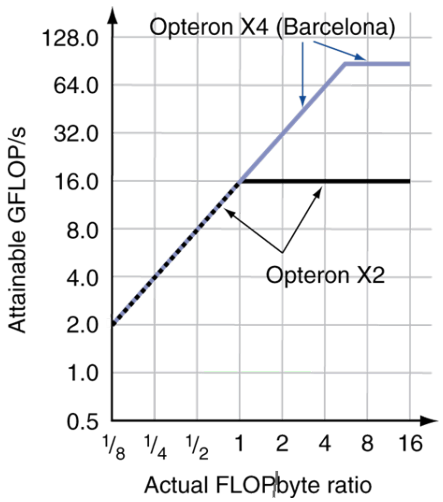
When does a page fault occur?

Question 1.4

What is the use of the Translation Look-aside Buffer (TLB)?

Question 1.5

In this "roofline" diagram for two Opteron processors X2 and X4, explain what happens for X2 and X4 when running a kernel with arithmetic intensity 2.0.
What modifications can be done to improve the performance of X4?



Problem 2

Question 2.1

Are there structural hazards in the 5-stage RISC-V pipeline? If so, what can you do to mitigate the issue without introducing stalling?

Question 2.2

Write RISC-V assembly code for the following fragment of C code. State if you write assembly code for the 32-bit or the 64-bit version of RISC-V (Venus and your simulator is a 32-bit RISC-V). Assume the variables are according to the RISC version either 32-bit or 64-bit long (same as the register width).

```
A[12] = h + A[8];
```

`h` is in `x21`, base address of `A` is in `x22`. The base address is where the array starts in memory

Question 2.3

Consider following loop:

```
LOOP:
lw x10, 0(x13)
lw x11, 8(x13)
add x12, x10, x11
subi x13, x13, 16
bnez x12, LOOP
```

Assume branches are always correctly predicted (i.e., no stalls due to branches), that there are no delay slots, and that the pipeline has full forwarding support.

Show a pipeline execution diagram for the first two iterations. Mark with a circle (e.g., Ⓔ) the pipeline stages that do not perform useful work. Branches are resolved in EX stage.

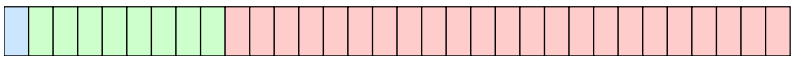
How often while the pipeline is full do we have a cycle in which all five pipeline stages are doing useful work? (Begin with the cycle when the `subi` is in the IF stage. End with the cycle when the `bnez` is in the IF stage.)

... continued on the next page

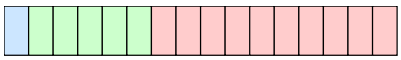
Problem 3

The IEEE standard 754 for floating-point includes the specifications for

binary32, or single precision:



binary16, or half precision:



Specifications		
Format	**binary16**	**binary32**
Storage (bits)	16	32
Precision (bits)	11	24
Trailing significand f (bits)	10	23
Total exponent length (bits)	5	8
E_{max}	15	127
bias	15	127

Given the two decimal numbers

$$X = 1000 \quad \text{and} \quad Y = -993$$

represented in floating-point, answer the following questions:

Question 3.1

Write the **binary32** and **binary16** representation of both X and Y (four values). You may represent the values in hexadecimal, if you wish.

Question 3.2

Compute the addition of X and Y for **binary32** by using the floating-point addition algorithm. Write the **binary32** representation of the result.

Question 3.3

Compute the product of X and Y for **binary16** by using the floating-point multiplication algorithm. Write the **binary16** representation of the result.

... continued on the next page

Problem 4

Question 4.1

Media applications that play audio or video files are part of a class of workloads called "streaming" workloads; i.e., they bring large amounts of data but do not reuse much of it. Consider a video streaming workload that accesses 1024 KB working set sequentially with the following address stream (byte addresses):

0, 2, 4, 6, 8, 10, 12, 14, 16, ...

Assume a 32 KB direct-mapped cache with a 32-byte block. What is the miss rate for the address stream above? How is this miss rate sensitive to the size of the cache or the working set? How would you categorize the misses (compulsory, conflict, or capacity) this workload is experiencing?

Question 4.2

Re-compute the miss rate when the cache block size is 16 bytes, 64 bytes, and 128 bytes. What kind of locality is this workload exploiting?

Question 4.3

Assume as 32-bit processor and following cache configuration: 2-way associative, 512 cache lines, 16 byte per cache line.

What are sizes of the address fields for tag, index, and offset? How large is the cache? How many bits are used in total for such a cache implementation and how many bits are used for the data storage.

... continued on the next page

Problem 5

A multicore processor executes an application A1 on a single core C0 in 100 μs.

When the application A1 is executed in two cores C0 and C1, the execution time is reduced to 75 μs.

Question 5.1
Determine what fraction of the application A1 is parallelizable.

Question 5.2

a) What would be the speed-up if the application A1 is run in 4 cores?

b) What is the maximum achievable speed-up for application A1?

After redesign, the serial portion of the code is halved in the new application A2. However, the execution on a single core of the new application A2 still takes 100 μs.

Question 5.3
Determine the execution time when A2 is executed in the two cores C0 and C1.

The multicore processor supports frequency and voltage scaling. Consider two power modes (for the single core):

mode	P_{ave} $[W]$	f_{clock} [GHz]
performance	10	2.0
low-power	4	1.0

Where P_{ave} is the average power dissipation in the single core when clocked at frequency f_{clock}. The execution time of A2 in the single core at $f_{clock} = 2$ GHz is 100 μs.

Question 5.4
Determine the execution time for the four combinations of the power modes when A2 is executed in the two cores C0 and C1.

Question 5.5
Determine the combination of power modes in C0 and C1 giving the minimum energy consumption for the execution of application A2.

End of the Exam