

Norwegian University of Science and Technology (NTNU)
DEPT. OF COMPUTER AND INFORMATION SCIENCE (IDI)

Course responsible: Professor Lasse Natvig

Quality assurance of the exam: PhD-student Magnus Jahre

Contact person during exam: Lasse Natvig, Phone: 906 44 580

Deadline for examination results: 3rd of July 2010.

EXAM IN COURSE TDT4260 COMPUTER ARCHITECTURE

Saturday 12th of June 2010

Time: 0900 - 1300

Supporting materials: No written and handwritten examination support materials are permitted. A specified, simple calculator is permitted.

By answering in short sentences it is easier to cover all exercises within the duration of the exam. The numbers in parenthesis indicate the maximum score for each exercise. We recommend that you start by reading through all the sub questions before answering each exercise.

The exam counts for 80% of the total evaluation in the course. Maximum score is therefore 80 points.

Exercise 1) Amdahls law and caches (Max 25 points)

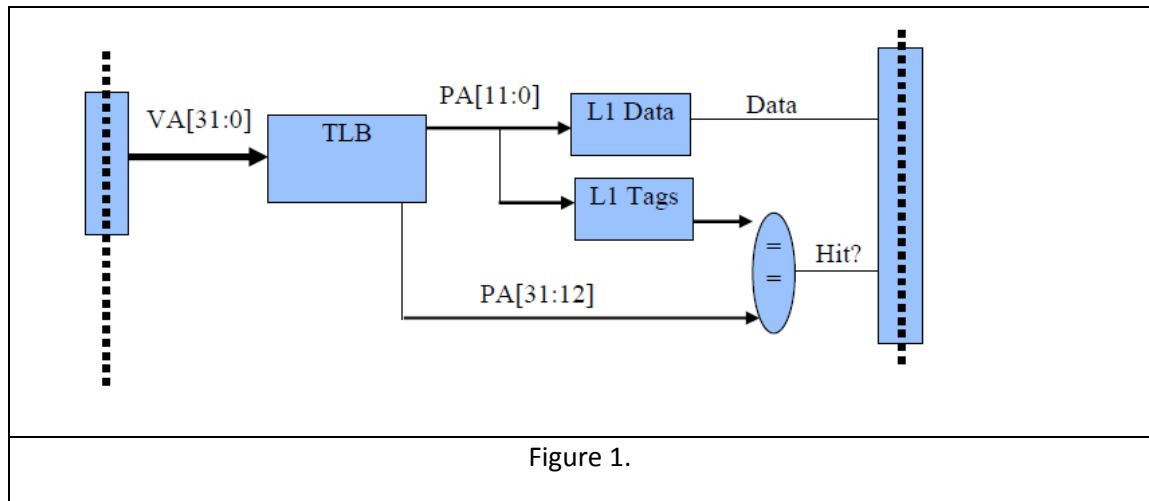
- a) (Max 6 points) Explain briefly Amdahl's law.

- b) (Max 4 points) What is the main difference between a homogeneous and heterogeneous multicore chip, and explain how Amdahl's law can be used as an argument in favor of heterogeneous multicore chips.

- c) (Max 5 points) Why does increasing the associativity of a cache often lower the miss rate? Give an explicit example.

- d) (Max 5 points) Most modern microprocessors have small level 1 data caches with a low associativity (2 for example). If increasing associativity lowers miss rate, then why does the level 1 cache often have a low associativity?

- e) (Max 5 points) John the architect is designing a pipeline. He is having trouble meeting timing, and the critical path is shown in Figure 1. In the path, a Virtual Address (VA) is taken from a register, then used to look up a Physical Address (PA) from the TLB. The Physical Address is then used to access a level 1 data cache. The TLB page size is 4 KB and the level 1 data cache is also 4 KB. What can John the architect do to improve timing?



Exercise 2) Instruction Level Parallelism (Max 10 points)

- a) (Max 5 points) Instructions per cycle (IPC), the number of instructions committed per cycle for a given benchmark, is often used to compare different architectures. Why is this a poor metric for architectural comparison? Give an example of a machine that achieves an IPC of 1.2 on a given benchmark that might have lower performance than a machine which achieves an IPC of 1.0 on the same benchmark.
- b) (Max 5 points) A computer architect is designing a new pipeline. She is considering a VLIW machine that can issue two instructions each cycle (ie. two-wide), and a two-wide in-order superscalar machine. Both machines have the same datapaths (ie, same number of ALUs, registers, etc). What is an advantage of choosing the VLIW machine? What is an advantage of choosing the superscalar design?

Exercise 3) Memory Systems (Max 10 points)

- a) (Max 6 points) Briefly explain three techniques that can be used to either reduce the hit time, increase bandwidth, reduce the miss penalty or reduce the miss rate in caches.
- b) (Max 4 points) Your task is to evaluate the performance effects of implementing non-blocking caches. Consider a processor with a single on-chip cache and a 100 clock cycle miss penalty to off-chip memory. In program A, the misses that stall the processor occur in bursts of 4 misses every 100 clock cycles. For program B, a single miss stalls the processor every 400 clock cycles. What percentage of time will the processor be stalled waiting for memory with Program A and B for (i) a blocking cache and (ii) a cache that can service 4 misses concurrently?

Exercise 4) Vector Processors and Interconnection Networks (Max 10 points)

- a) (Max 5 points) What makes vector processors fast at executing a vector operation?
- b) (Max 5 points) Interconnection networks fall into two main routing categories (source routing and distributed routing). In a source routed scheme, a header is generated which contains what to do at each switch point along the path of the packet. In a distributed routing scheme, the header simply contains the destination address, and the router calculates what to do at each switch point. What is an advantage of the source routed scheme? What is a disadvantage?

Exercise 5) Chip Multiprocessors (Max 25 points)

- a) (Max 6 points) In the paper Chip Multithreading: Opportunities and Challenges, by Spracklen & Abraham is the concept Chip Multithreaded processor (CMT) described. The authors describe three generations of CMT processors. Describe each of these briefly. Make simple drawings if you like.
- b) (Max 6 points) Explain briefly the research method called design space exploration (DSE). When doing DSE, explain how a cache sensitive application can be made processor bound, and how it can be made bandwidth bound.
- c) (Max 4 points) Discuss briefly the advantages and disadvantages of asymmetric multicore processors with homogenous instruction set architectures.
- d) (Max 5 points) Your task is to analyze the performance and power efficiency of a parallel program on two different machines. In Machine A, the chip area is used to provide 4 high-performance processing cores where each core can complete 2 units of work each second. In Machine B, the area is used to provide 16 cores that can complete 1 work unit each second. Both machines use the same amount of power. The program is mapped to one thread per core and consists of 32 units of work. Furthermore, it scales ideally and all communication overheads are negligible. What is the runtime of the program on Machine A and B? Which machine is most power efficient for this program?
- e) (Max 4 points) In Chip Multiprocessors, memory system units are often shared between processing cores. Commonly, the cores share the on-chip interconnect, the shared cache and the off-chip memory bus. Briefly describe how memory requests from different cores can interfere with each other in these units. How can this interference affect performance?

...---oooOOOooo---...