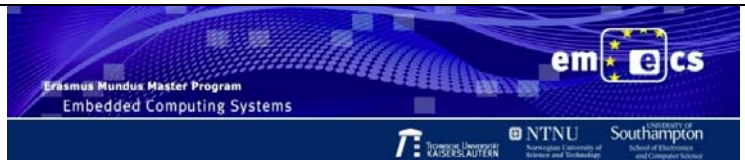
Norwegian University of Science and Technology (NTNU)
DEPT. OF COMPUTER AND INFORMATION SCIENCE (IDI)

Course responsible: Professor Lasse Natvig
Quality assurance of the exam: Associate Professor Magnus Jahre
Contact person during exam: Lasse Natvig, Phone: 906 44 580

Deadline for examination results: 27th of June 2011.

EXAM IN COURSE TDT4260/DT8803 COMPUTER ARCHITECTURE
Monday 6th of June 2011
Time: 1500 – 1900



This course is part of EMECS: Erasmus Mundus Master Program in Embedded Computing Systems. **The students are free to answer the exercises in English or Norwegian (Bokmål or Nynorsk).**

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 subquestions before answering each exercise.

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

Exercise 1) Measuring computer performance (Max 14 points)

a) (Max 6 points) The increase in processor performance on integer operations was quite stable in the period 1985 to 2002, but has changed after 2002. Explain briefly how the increase has been since 2002, and explain some of the reasons that have given the change.

b) (Max 4 points) Give a general (informal if you like) definition of each of the two concepts bandwidth (throughput) and latency (response time).

c) (Max 4 points) Explain how bandwidth has changed roughly during the last 20 years compared to latency for primary memory (main memory) and for hard disk technology.

Exercise 2) Loop unrolling and VLIW (Max 14 points)

a) (Max 4 points) Describe how you would do loop unrolling in the case where you do *not* know the number of iterations of the loop, called *n*.

b) (Max 6 points) Describe briefly the advantages and disadvantages of VLIW. Explain also briefly why it has been a popular architecture in embedded systems.

c) (Max 4 points) 1st generation VLIW systems operated in lock-step without hazard detection. Describe the main problem/disadvantage of this choice, and why caches make this problem larger.

Exercise 3) Caches and prefetching (Max 20 points)

a) (Max 4 points) Explain briefly why most computers today have a so-called memory hierarchy.

b) (Max 6 points) When measuring cache miss the misses can be categorized as “four C’s”. Two of these are Compulsory and Coherence. Explain briefly what kind of misses each of them represent. Give also the term for the two other C’s, and explain what kind of misses they represent.

c) (Max 6 points) Prefetching schemes can be described by answering three basic questions starting with “When”, “Where” and “What”. Discuss these three questions/dimensions and give an example showing that these three questions are not independent of each other.

d) (Max 4 points) Dahlgren et.al. proposed a prefetching policy called *sequential adaptive prefetching.* Explain the policy based on Figure 2, found below.

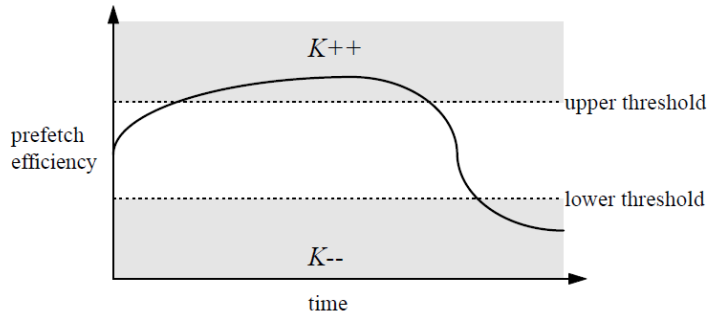


Figure 2: Sequential adaptive scheduling.

Exercise 4) Multi-cores architectures (Max 18 points)

a) (Max 4 points) The UltraSparc T1 (Niagara) has a pipeline where some of the subunits are replicated (copied) in 4 copies to make rapid context switches between 4 threads possible. Describe which units where replicated in this way.

b) (Max 4 points) Based on Figure 1, explain briefly the role of the unit called Home Engine and Remote Engine, as well as the Interconnect Links to the left in the figure.

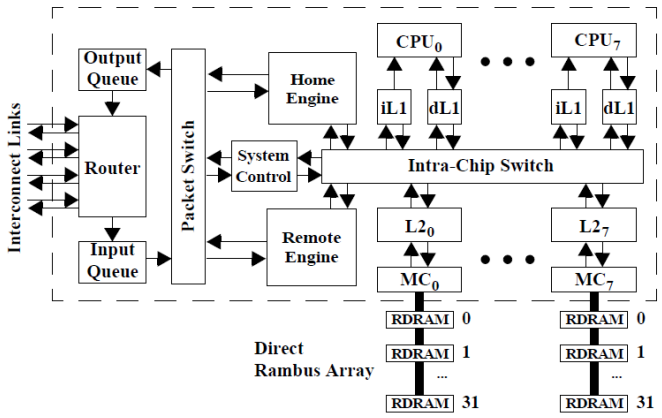


Figure 1: Piranha processing node

c) (Max 4 points) Describe briefly the role of the Piranha I/O node and how it is different from the processing node.

d) (Max 6 points) Describe briefly the protocol engines in Piranha and how they are implemented. Sketch a rough block diagram if you like. (Hint: They have three main stages).

Exercise 5) Interconnection networks, green computing and HPC (Max 14 points)

a) (Max 6 points) The Omega network is an example of a multistage interconnection network. Explain how such a network is a compromise between a bus and a crossbar, and sketch one such network (e.g. Omega) providing communication from any of 8 nodes to any of the same 8 nodes. (The exact details of the Omega "interconnection-pattern" is not needed)

b) (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?

c) (Max 3 points) GPFS is a filesystem used in supercomputers (HPC systems), also at NTNU. Describe GPFS briefly.

…---oooOOOooo---…