

Exam in DAT 105 (DIT 051) Computer Architecture

Time: August 15, 2017 14 – 18 in the Mechanical Engineering Building

Person in charge of the exam: Per Stenström, Phone: 0730-346 340

Supporting material/tools: Chalmers approved calculator.

Exam Review: On August 24, 2017 (Per Stenstrom's office)

Grading intervals:

- **Fail:** Result < 24
- **Grade 3:** 24 <= Result < 36
- **Grade 4:** 36 <= Result < 48
- **Grade 5:** 48 <= Result

NOTE: Answers must be given in English

GOOD LUCK!

Per Stenström



[General disclaimer: All assumptions needed to do the assignments are provided. If you feel that sufficient facts are not provided to solve a problem, either 1) ask the teacher when she/he visits the exam, or 2) make your own additional assumptions. Additional assumptions will be accepted if they are reasonable and required to solve the problem. Always make sure to motivate your answers.]

ASSIGNMENT 1

1A) Consider a baseline microprocessor with a single processor run at 1 GHz. This baseline microprocessor is now upgraded and several alternatives are considered

- i) A single processor run at 2 GHz
- ii) Two processors run at 1 GHz each
- iii) Four processors run at 0.75 GHz

Consider two programs A and B. A is single threaded. B is multithreaded and enjoy a speedup at two processors a factor of two and speedup at four processors with a factor eight. What is the execution time and throughput improvement for the two programs on the three upgrades? **(6 points)**

1B) The geometric mean performance of a suite of five programs is S_5 . If we know the speedup of a sixth program, denoted S_6 , over the same reference machine as the suite of five programs, how do we calculate the geometric mean performance over the entire suite of six programs, S_6 ? **(3 points)**

1C) Assume that two architectural improvements A_1 and A_2 are applicable to 40% and 60%, respectively, of the execution of a program yielding a speedup on the portions they apply by a factor 60 and 40, respectively. What is the overall impact on performance of A_1 and A_2 ? **(3 points)**

ASSIGNMENT 2

Assume a processor with a simple five-stage pipeline (Instruction fetch, Instruction decode, Execute, Memory access, Write back). Furthermore, assume that all memory accesses complete in a single cycle (cache hit time is one cycle), that pipeline forwarding is used whenever possible to resolve RAW hazards, and that there are two branch delay slots (all branch computations are completed in the EX stage). Integer multiply operations are performed in two steps; the first in the Execute stage, and the second in the Memory Access stage, so multiply operations are fully pipelined like all other arithmetic instructions, but the result is not available until the end of the Memory access stage. Finally, for Branch instructions, branch conditions are established in the EX stage.

2A) Show a pipeline diagram of the pipeline, especially marking how operands can be forwarded. **(3 points)**

2B) Your assignment is to write a piece of MIPS assembly code to perform the following computation in as few clock cycles as possible.

```
for (int i=0; i<100; i=i+1) {
    x[i] = x[i]*y[i];
}
```

The x and y arrays are array of long integers (8 bytes). It is assumed that register R1 already is loaded with the start address of the x array, i.e. the address of x[0], and that you are allowed to change the value of R1.

The following is a first version of the code:

```

                ADDI    R2,R1,#800
LOOP:          LD      R4,0(R1)
                LD      R5,100(R5)
                MUL    R6, R4,R5
                SD     R6,0(R1)
                ADDI   R1, R1,8
                BNE    R1,R2,LOOP
```

Which RAW hazards are eliminated through operand forwarding and which are not and how many cycles are lost due to RAW hazards in each iteration? Finally, how much faster would the program run without any RAW hazards assuming only RAW hazards cause execution inefficiency?

(3 points)

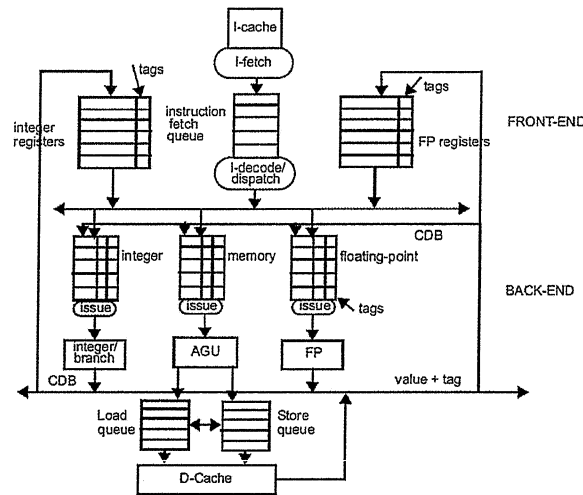
2C) How many cycles are lost in each iteration due to control hazards and how much faster would the program run without any control hazards assuming all RAW hazards are eliminated? **(3 points)**

2D) Branch prediction can also be used to eliminate control hazards. In the assumed pipeline the branch condition is computed in the EX stage. Explain how a 1-bit branch predictor can be used to remove control hazards and when it will correctly predict the branch outcome assuming it is initialized to assume the branch is not taken. Calculate the branch prediction accuracy for the example loop.

(3 points)

ASSIGNMENT 3

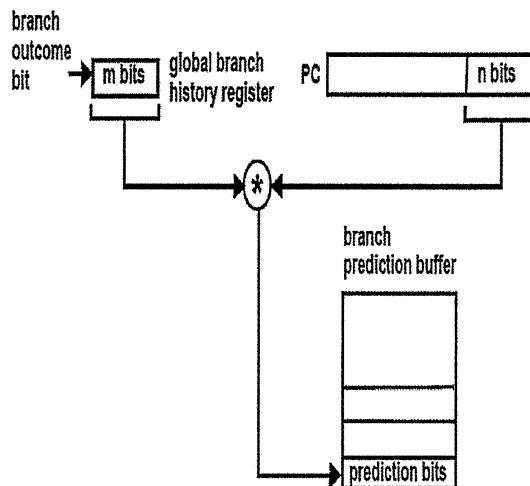
The diagram below shows a pipeline using the Tomasulo algorithm for dynamic instruction scheduling. The pipeline consists of 5 stages: I-Fetch, Dispatch, Issue, Execution/Cache Access, and CDB write. Assume that all integer and branch instructions execute in a single cycle whereas floating-point instructions execute in six cycles, thus the duration of the execution stage varies. A cache access (instruction and data) takes a single cycle whereas stores are decomposed into two instructions (one for address generation and one for cache access).



3A) How does the Tomasulo algorithm resolve RAW, WAR, and WAW hazards? Explain in detail how the use of tags assigned to each instruction and distributed over the CDB orchestrates hazard resolution. (3 points).

3B) How is the basic Tomasulo pipeline extended for speculative execution using branch prediction and a ReOrder Buffer (ROB)? (3 points).

3C) The figure below shows a two-level branch predictor. Let us use it to predict the outcome of one particular branch instruction using a single bit predictor and using $m=4$ bits of branch outcome. Initially, all prediction bits are cleared (not taken). If the 16 most recent executions of the branch are taken, what is the branch prediction accuracy? (3 points)



3D) Explain how explicit register renaming is used using a physical register file and then keeping track of the latest and the retired value of a particular register using a front-end and a back-end *Register Alias Table (RAT)*. **(3 points)**

ASSIGNMENT 4

4A)

A computer architect wants to establish how many misses in each of the three categories there are for a 4-block, two-way associative cache using LRU and for a 4-block fully associative cache. A program does the following block references: 0 1 2 3 4 0 5 1 8 4 9 5.

Establish the number of cold, capacity (using the OPT replacement policy) and conflict misses for the two caches. **(4 points)**

4B) How does the following techniques affect the cold and capacity miss rate? i) larger block size and ii) larger caches? **(2 points)**

4C) A computer architect wants to establish the relative performance between a system with a blocking and a non-blocking cache using software prefetching. In software prefetching, prefetch instructions are appropriately inserted in the code shown below. Assume that five instructions are executed per iteration where the CPI for each instruction is one assuming that it doesn't cause a cache miss. On the other hand, if the instruction causes a cache miss, CPI is 100. In addition, prefetch instructions result in CPI=2. Both caches have a block size of four words.

```
for (i=0; i<1000; i++)
    C+=A[i];
```

First show how the code is annotated with prefetch instructions. How much faster does the program run on the system with a non-blocking cache using software prefetching? A convincing explanation that is easy to follow is needed for full points. **(6 points)**

ASSIGNMENT 5

5A) Consider a multicore system comprising a number of processors (cores) on a chip that are connected to a single-level private cache. The private caches use the write-through write policy. $X_i=R_i$ and $X_i=W_i$ means a read and a write request to the *same* address from processor i , respectively, where $W_i=C$ means that the value C is written by processor i . Now consider the following access sequence:

W₁=0

R₁

R₂

W₁=1

R₂

What should be returned by the second read operation from processor 2 and what is the reason that the correct value is not returned given the cache write policy assumed?

(3 points)

5B) How can a simple change of the write-through cache make sure that the correct value is returned to processor 2? **(3 points)**

5C) Multithreading refers to a general technique to switch to another thread when a high-latency operation is encountered. Explain for the three example techniques a) block multithreading b) interleaved (or fine-grained) multithreading and c) simultaneous multithreading how a five-stage pipeline must be extended to support each of the techniques. **(6 points)**.

***** GOOD LUCK! *****