CSCE 513 Computer Architecture **Test 2b** **November 23, 2009**

Name:_____ Email:_____

Instructions

. No Calculators!!

- Make sure your exam is complete. There should be 8 pages including this cover sheet and a collection of figures.
- No Calculators, cell phones, or other electronic devices.
- All questions are equally weighted.
- Answer in the space provided if at all possible.
- If a question is unclear please ask early in the test.
- Good Luck!

1. Describe in detail using either a flowchart or pseudocode the process of translation and then accessing the data specified by a physical address. Assume two level cache with both levels two-way associative and a fully associative TLB.

2. Cache performance etc.

   (a) What adaptation is necessary to support a virtually addressed cache?

   (b) Why is the typical cache structure two L1 caches and a unified L2 cache?

   (c) If an instruction "L.D R1, 1024($0)" is in an inner loop and addresses are such that both the instruction and the constant data reference 1024($0) mapped to the same set, if there is a unified L1 cache what problem does this present and how could it be handled (other than separate caches)?

   (d) Explain what aspect of cache performance write buffers improve and how do they it.

3. If wee assume a direct mapped cache with 32 sets (a small cache) and the program below such that

   - &a[0][0] = 0x7000
   - &b[0][0] = 0x8000
   - &c[0][0] = 0x9000

   and further that each of these addresses maps to set 0.
   Given the code below:

```
double a[64][64];
double b[64][64];
double c[64][64];
double sum = 0.0;
for(i=0; i < 64; ++i)
    for(j=0; j < 64; ++j)
        c[i][j] = a[i][j] + b[j][i];   // Note this is not the standard sum!!
```

   (a) What is the Hit ratio assuming blocks are 32B?

   (b) What is the Hit ratio assuming blocks are 64B?

   (c) Now assume the loops are switched with the "j-loop" becoming the outermost, what is the Hit ratio assuming blocks are 32B?

   (d) What is the Hit ratio assuming blocks are 64B?

4. Hand instrument the following code:

```
double a[64][64], b[64][64], c[64][64];

for(i=0; i < 64; ++i){

    for(j=0; j < 64; ++j){

        c[i][j] = a[i][j] + b[j][i];   // Note this is not the standard sum!!

    }

}
```

```
004001f0 <main>:
  4001f0:       43 00 00 00     addiu $29,$29,-24
  4001f8:       34 00 00 00     sw $31,16($29)
  400200:       02 00 00 00     jal 4004a8 <__main>
  400208:       42 00 00 00     addu $7,$0,$0
  400210:       a2 00 00 00     lui $11,4096
  400218:       43 00 00 00     addiu $11,$11,1008
  400220:       a2 00 00 00     lui $10,4097
  400228:       43 00 00 00     addiu $10,$10,1008
  400230:       a2 00 00 00     lui $9,4097
  400238:       43 00 00 00     addiu $9,$9,-31760
  400240:       42 00 00 00     addu $6,$0,$0
  400248:       55 00 00 00     sll $8,$7,0x3
  400250:       42 00 00 00     addu $5,$0,$9
  400258:       42 00 00 00     addu $4,$0,$10
  400260:       42 00 00 00     addu $3,$0,$11
  400268:       42 00 00 00     addu $2,$8,$3
  400270:       2b 00 00 00     l.d $f2,0($4)
  400278:       2b 00 00 00     l.d $f0,0($2)
  400280:       43 00 00 00     addiu $3,$3,512
  400288:       71 00 00 00     add.d $f2,$f2,$f0
  400290:       43 00 00 00     addiu $6,$6,1
  400298:       43 00 00 00     addiu $4,$4,8
  4002a0:       5c 00 00 00     slti $2,$6,64
  4002a8:       37 00 00 00     s.d $f2,0($5)
  4002b0:       43 00 00 00     addiu $5,$5,8
  4002b8:       06 00 00 00     bne $2,$0,400268 <main+0x78>
  4002c0:       43 00 00 00     addiu $10,$10,512
  4002c8:       43 00 00 00     addiu $9,$9,512
  4002d0:       43 00 00 00     addiu $7,$7,1
  4002d8:       5c 00 00 00     slti $2,$7,64
  4002e0:       06 00 00 00     bne $2,$0,400240 <main+0x50>
```

5. Cache Performance - In the system we are analyzing the memory has:

   - Separate L1 instruction and data caches, HitTime = Processor Cycle Time
   - 32KB L1 instruction cache with 1% miss rate, 64B blocks
   - 256KB L1 data cache with 5% miss rate, 16B blocks
   - 1M L2 unified cache with 64B blocks, local miss rate 10%, Hit Time for L2 = 10 cycles
   - the L2 $\Leftrightarrow$ Memory bus is 64 bits wide
   - Main Memory Access time is 100 cycles for the first portion of the block and subsequent parts are available on each subsequent bus cycle=10 CPU cycles.
   - Both L1 caches are four-way associative, L2 direct mapped.
   - Assume the only memory references are loads(5%) and stores(5%).

   I your answer do not do aritmetic, do not simplify, leave your answers as expressions so I can see your reasoning.

   (a) What percentage of memory references are for data?

   (b) Assuming that there are no misses in references to main memory what is the miss penalty for the L2 cache?

   (c) What is the average memory access time?

6. (a) Suppose that you have an interleaved memory and that L2 Cache blocks are 64B and the bus width is 64bits=8B. If the main memory access time to read 64 bits is 20 cycles and the subsequent reads necessary to transfer a block occur every 4 cycles then what is the time necessary to read and transfer a full L2 block.

   (b) Explain critical word first?

7. Virtual Memory/TLB - Given

   - 64 bit virtual addresses
   - the page size is 4KB,
   - 32 bit physical addresses
   - 1MB cache, only one level unified, 16 way associativity, 256B lines
   - 64 entry TLB, 8 way associative

   (a) What fields are in the TLB entries?

   (b) If the virtual address is 0xABCDEF9876548 and if the physical page number is 0x54 with leading zeroes not shown

      i. What is the page offset field?

      ii. What is the VPN?

      iii. How big is a PPN?

      iv. What is the physical address?

      v. What is the the cache "block offset" field?

      vi. What is "set-index" field?

      vii. What is the cache "tag" field?

   (c) How would a page fault be detected?

   (d) What happens on a TLB miss?