COMP228-DD Computer Organization & Design Winter 2020
Assignment 1
Due: January 29, 2020

Submit on Moodle before 18:00h and typed hardcopy in class.

1. [16 marks] Amdahl's Law

a) On a uniprocessor, serial portion A of program P consumes x% of the time, while parallel portion B consumes the remaining (100 - x)%. On a parallel computer, portion A slows down by a factor of 2, while portion B speeds up by the number of processors. The theoretical maximum speedup is 250 times. How many actual processors are needed to achieve at least 75% of the maximum speedup? How many actual processors are needed to achieve at least 90% of the maximum speedup?

- at least 75% of the maximum speedup: P = _____ processors
- at least 90% of the maximum speedup: P = _____ processors

b) Portion A of program P takes 'x' seconds; portion B takes 'y' seconds. 'x + y = 100'. A is totally sequential, while B is totally parallel. With an infinite number of processors, we get a speedup of 's'. It takes 5 2/3 imaginary processors (6 real processors) to achieve a speedup of at least 's/2'. What are the values of 'x' and 'y'? (They are integers). (x,y) = _____

2. [16 marks] Performance and Power Efficiency

A heterogeneous multicore chip adapts to the particular requirements by turning on a certain number of each of its two classes of core. Alpha cores generate 12 GFs/s and dissipate 2 W (watts) of power. Beta cores generate 1 GF/s and dissipate 0.1 W. We seek to maximize aggregate chip performance while still achieving a power efficiency of at least 6 2/3 GFs/s/W. Electrical considerations dictates that precisely 500 cores to be turned on.

a) Indicate the number of cores turned on. _____ Alpha cores, _____ beta cores
b) What is chip peak performance and chip power dissipation? _____ TFs/s, _____ W

3. [20 marks] Architecture and Implementation.

A data scientist needs a MIPS assembly-language program for data analysis. He has three arrays of floating-point numbers. Array 'a' is a source array, and 'b' and 'c' are result arrays. Array 'a' contains '2n' floating-point numbers. 'r1' is the address of 'a[0]'. 'r4' is the address of the byte immediately following 'a', i.e., the address of the imaginary element 'a[2n]'. Each of 'b' and 'c' can store 'n' floating-point numbers. 'r2' is the address of 'b[0]'. 'r3' is the address of 'c[n-1]'. The algorithm is: walk through array 'a' copying even-index elements to 'b' in 'b' order, after multiplying by 'f2', and odd-index elements to 'c' in reverse order (i.e., from high 'c' index to low 'c' index), after multiplying by 'f0'. Write the MIPS code. Use only the following instructions.

```
l.d   f6,n(r2)    // load to 'f6' eight bytes starting at Mem[r2+n]
s.d   f6,n(r2)    // store f6's value to memory starting at Mem[r2+n]
```

```
mul.d f4,f0,f2    // write 'f0*f2' into 'f4'
addi  r1,r1,n     // add integer 'n' to 'r1'
subi  r1,r1,n     // subtract integer 'n' from 'r1'
bne   r1,r2,loop  // if 'r1 /= r2' then goto 'loop'
```

4. [12 marks] Architecture and Implementation

Say that a processor has a "fetch unit" and a "branch unit". Both units require access to the same pipeline data store ("processor variable"). What is this data store? And why do both units require access to it?

5. [16 marks] Address-register Sizes.

An n-bit register can hold 2^n distinct bit patterns. As such, it can only be used to address a memory whose number of addressable units (typically, bytes) is less than or equal to 2^n. In what follows, use the power-of-two interpretation of Greek prefixes. I.e., K = 2^10, M = 2^20, etc. In this question, register bit sizes need not be a power of two.

a) What is the minimum size of an address register for a computer with 8 GBs of memory? _____ bits

b) What is the minimum size of an address register for a computer with 35 TBs of memory? _____ bits

c) What is the minimum size of an address register for a computer with 1.05 EBs of memory?____ bits

d) What is the minimum size of an address register for a computer with 0.5 (1/2) EB of memory? ___ bits

6. [20 marks] Utility-Cost Tradeoffs

A single NEC SX-9 vector processor unit (VPU) has a peak arithmetic performance of 102.4 GFs/s. The memory system supplies a peak memory bandwidth of 162 GBs/s. The VPU dissipates 40 W, while the memory system dissipates 55 W. Together, both components dissipate 95 W. Every program has an arithmetic intensity (AI), measured in flops/byte, that indicates how much performance is achievable for a given bandwidth. The equation is: achievable GFs/s = min{AI*BW, peak GFs/s}.

a) What is the minimum AI that achieves peak arithmetic performance on an SX-9 VPU?
AI = _____ flops/byte

b) What is the power efficiency of this computer system at peak performance?
PE = _____ GFs/s/W

c) A special customer buys two memory systems for the one VPU, thus doubling both the bandwidth and the power of the memory. What is the new minimum AI that achieves peak arithmetic performance?
AI = _____ flops/byte

d) What is the power efficiency of the new system at peak performance?
PE = _____ GFs/s/W

e) Since power efficiency is a good thing, why would a customer spend more money to get less of it?