## **Computer Sciences Department**

## **University of Wisconsin-Madison**

## **CS/ECE 552 – Introduction to Computer Architecture**

## **In-Class Exercise (01/23)**

**Answers to all questions should be uploaded on Canvas**

1. [3 points] From the textbook

(a) [1 point] (Twist on Check Yourself 1.5) A key factor in determining the cost of an integrated circuit is volume. Which of the following are reasons why a chip made in high volume should cost less?

1. With high volumes, the manufacturing process can be tuned to a particular design, increasing yield.
2. It is less work to design a high-volume part than a low-volume part.
3. The masks used to make the chip are expensive, so the cost per chip is lower for higher volumes.
4. Engineering development costs are high and largely independent of volume, thus, the development cost per die is lower with high-volume parts.
5. High volume parts usually have smaller die sizes than low-volume parts and therefore have higher yield per wafer.

Given that (i), (iii), and (iv) are correct, why are (ii) and (v) incorrect?

**The higher volume indicates the higher market and general usage such as for person PCs. Low-volume part is usually used by government owned high-tech company and army such as NASA. Therefore, design a high-volume part requires less work.**

**Since the volume is large, given that the size of each wafer is the same, each die must be smaller. Therefore, since the research budget is distributed into more yields, and each wafer contains more die, each product will cost less. If the size of die is small, than more die can be put on each wafer and therefore the yield will be higher.**

(b) [1 point] (Section 1.6) A program *P* has 2500 instructions, each of which takes 2 cycles to complete. If we run *P* on a processor that has a clock speed of 100 MHz, how long does this program take to execute?

*Hint: see page 38 of the textbook.*

*2500\*2/100MHz = 5e-5 second*

(c) [1 point] (Section 1.7) Power is becoming an increasingly important issue to consider when designing processors. Suppose you are the chief architect at a small startup designing an IOT device. Given the tight deadline to ship your product, you only have enough time to make one optimization – reduce frequency by half or decrease the voltage by 25%. Which optimization will be more beneficial?

(3/4)^2 = 9/16 > ½ -> reduce frequency by half

2. [1 point] Consider a program running on your computer. You apply a technique that accelerates non-memory instructions in your program by 4x. This speeds up the entire program execution by 3x. Now would it be better for performance to (i) eliminate half of the memory instructions in your program or (ii) improve the technique such that it accelerates non-memory instructions by 6x?

(1) 1/(1/9\*1/2 + 8/9) = 18/17

(2) 1/(1-8/9 + 8/(9\*6)) = 27/7

3. [1 point] Consider a processor architecture with an average CPI of 1.8. You propose a change to the architecture that is able to eliminate half of all memory read instructions in the program but incurs a single-cycle overhead on all memory write instructions. Given the typical instruction breakdowns that you expect in your programs (table below), is this change a worthy trade-off for performance?

|  |  |
| --- | --- |
| **% of Program** | **Instruction Type** |
| 10% | Addition/subtraction |
| 5% | Multiplication |
| 40% | Memory read |
| 10% | Memory write |
| 15% | Logic |
| 20% | Branch |

(0.1 + 0.05 + 0.4\*0.5 + 0.15+0.2) \* 1.8 + 0.1\*2.8 = 1.54

1.8/1.54 = 1.169

4. [2 points] You build a computer that achieves a SPECratio of 9.5 for the benchmark astar when running at frequency f = 4GHz with a CPI of 1.79. Say that the number of instructions in astar were to increase by 10% and the CPI were to increase by 5%.

1. [1 point] By how much would astar’s execution time increase?

**1.1\*1.05 = 1.155**

1. [1 point] What is the new SPECratio?

9.5/1.155 = 8.225

5. [3 points]

(a) [1 point] Suppose we propose making a tradeoff in our architecture. Specifically, we propose making a change that reduces the performance of 20% of a program by a factor of 4 (call this “enhancement” *E3*) and improves the performance of another 12% of the program by a factor of 6 (call this enhancement *E4*). Assume the two enhancements are independent and affect different parts of the program (there is no overlap between the 20% and 12% of the program). What is the overall effect on the performance of the program? Is this tradeoff worthwhile?

**E3 = 1/(1-1/5+1/5\*4) = 0.625**

**E4 = 1.1111**

**1/(0.68+E3+E4) = 1.066 Yes**

(b) [2 points] Your colleague claims we can overcome *E3* by increasing the clock frequency by 10%. Is this sufficient? Why or why not? Assume the architecture stays the same and the change in frequency does not affect dynamic instruction count or CPI.

0.694 <1 No.