NYU Tandon School of Engineering

Fall 2022, ECE 6913

Homework Assignment 1

Instructor Contact : ajb20@nyu.edu

Course Assistants & Graders: Varadraj Kakodkar (vns2008), Kartikay Kaushik (kk4332), Siddhanth Iyer (si2152), Swarnashri Chandrashekar (sc8781), Karan Sheth (kk4332), Haotian Zheng (hz2687), Haoren Zhang (kk4332), Varun Kumar (vs2411)

Homework Assignment 1 [released: Tuesday Sept 6th] [due Friday Sept 16th by 11:59 PM]

You *are allowed* to discuss HW assignments with anyone. You are *not allowed* to share your solutions with other colleagues in the class. Please feel free to reach out to the Course Assistants or the Instructor during office hours or by appointment if you need any help with the HW.
Please enter your responses in this Word document after you download it from NYU Brightspace. *Please use the Brightspace portal to upload your completed HW.*

---

1. Consider three different processors P1, P2, and P3 executing the same instruction set. P1 has a 3 GHz clock rate and a CPI of 1.5. P2 has a 2.5 GHz clock rate and a CPI of 1.0. P3 has a 4.0 GHz clock rate and has a CPI of 2.2.

a. Which processor has the highest performance expressed in instructions per second?

b. If the processors each execute a program in 10 seconds, find the number of cycles and the number of instructions.

c. We are trying to reduce the execution time by 30%, but this leads to an increase of 20% in the CPI. What clock rate should we have to get this time reduction?

2. Assume for arithmetic, load/store, and branch instructions, a processor has CPIs of 1, 12, and 5, respectively. Also assume that on a single processor a program requires the execution of 2.56E9 **arithmetic** instructions, 1.28E9 **load/store** instructions, and 256 million **branch** instructions. Assume that each processor has a 2 GHz clock frequency.

Assume that, as the program is parallelized to run over multiple cores, the number of arithmetic and load/store instructions per processor is divided by **0.7** × **p** (where p is the number of processors) but the number of branch instructions per processor remains the same

a. Find the total execution time (ET) for this program on 1, 2, 4, and 8 processors, and show the relative speedup of the 2, 4, and 8 processors result relative to the single processor result.

b. If the CPI of the arithmetic instructions were doubled, what would the impact be on the execution time of the program on 1, 2, 4, or 8 processors?

c. To what should the CPI of load/store instructions be reduced in order for a single processor to match the performance of four processors using the original CPI values?

3. Consider the three different processors P1, P2, and P3 executing the same instruction set. P1 has a clock cycle time of 0.33 ns and CPI of 1.5; P2 has a clock cycle time of 0.40 ns and CPI of 1.0; P3 has a clock cycle time of 0.3 ns and CPI of 2.8.

a. Which has the highest clock rate? What is it?

b. Which is the fastest computer? If the answer is different than above, explain why. Which is slowest?

c. How do the answers for a and b reflect the importance of benchmarks?

4. You are designing a system for a real-time application in which specific deadlines must be met. Finishing the computation faster gains nothing. You find that your system can execute the necessary code, in the worst case, twice as fast as necessary.

a. How much energy do you save if you execute at the current speed and turn off the system when the computation is complete?
b. How much energy do you save if you set the voltage and frequency to be half as much?

5. Consider two different implementations of the same instruction set architecture. The instructions can be divided into four classes according to their CPI (classes A, B, C, and D). P1 with a clock rate of 2.0 GHz and CPIs of 1, 2, 2, and 1, and P2 with a clock rate of 4 GHz and CPIs of 2, 3, 4, and 4.

a. Given a program with a dynamic instruction count of 1.0E6 instructions divided into classes as follows: 10% class A, 20% class B, 50% class C, and 20% class D. Which is faster: P1 or P2 (in total execution time)?

b. What is the global CPI for each implementation?

c. Find the clock cycles required in both cases.

d. Which processor has the highest throughput performance (instructions per second) ?

e. Which processor do you think is more energy efficient? Why?

6. **a.** What is the difference between CISC and RISC architectures? Give some examples wherein you think CISC architectures are better suited for than RISC architectures and vice versa.

b. Describe in your own words why you think RISC V would be a better alternative compared to ARM or x86 architectures?

c. What do you think are the challenges faced by RISC V architecture going forward?

7. In this exercise, assume that we are considering enhancing a machine by adding vector hardware to it. When a computation is run in vector mode on the vector hardware, it is 15 times faster than the normal mode of execution. We call the percentage of time that could be spent using vector mode the *percentage of vectorization.* Vectors are discussed in Chapter 4, but you don't need to know anything about how they work to answer this question!

a. Draw a graph that plots the speedup as a percentage of the computation performed in vector mode. Label the y-axis "Net speedup" and label the x-axis "Percent vectorization."
b. What percentage of vectorization is needed to achieve a speedup of 2?
c. What percentage of the computation run time is spent in vector mode if a speedup of 2 is achieved?
d. What percentage of vectorization is needed to achieve one-half the maximum speedup attainable from using vector mode?
e. Suppose you have measured the percentage of vectorization of the program to be 70%. The hardware design group estimates it can speed up the vector hardware even more with significant additional investment. You wonder whether the compiler crew could increase the percentage of vectorization, instead. What percentage of vectorization would the compiler team need to achieve in order to equal an addition 2× speedup in the vector unit (beyond the initial 15×)?

8. In a server farm such as that used by Amazon or eBay, a single failure does not cause the entire system to crash. Instead, it will reduce the number of requests that can be satisfied at any one time.

a. If a company has 10,000 computers, each with a MTTF of 35 days, and it experiences catastrophic failure only if 1/3 of the computers fail, what is the MTTF for the system?

b. If it costs an extra $1000, per computer, to double the MTTF, would this be a good business decision? Show your work.

9. a. A program (or a program task) takes 150 million instructions to execute on a processor running at 2.7 GHz. Suppose that 70% of the instructions execute in 3 clock cycles, 20% execute in 4 clock cycles, and 10% execute in 5 clock cycles. What is the execution time for the program or task?

b. Suppose the processor in the previous question part is redesigned so that all instructions that initially executed in 5 cycles and all instructions executed in 4 cycles now execute in 2 cycles. Due to changes in the circuitry, the clock rate also must be decreased from 2.7 GHz to 1.5 GHz. What is the overall percentage improvement?

10. Availability is the most important consideration for designing servers, followed closely by scalability and throughput.

a. We have a single processor with a failure in time (FIT) of 100. What is the mean time to failure (MTTF) for this system?

b. If it takes one day to get the system running again, what is the availability of the system?

c. Imagine that the government, to cut costs, is going to build a supercomputer out of inexpensive computers rather than expensive, reliable computers. What is the MTTF for a system with 1000 processors? Assume that if one fails, they all fail.

11. Server farms such as Google and Yahoo! provide enough compute capacity for the highest request rate of the day. Imagine that most of the time these servers operate at only 60% capacity. Assume further that the power does not scale linearly with the load; that is, when the servers are operating at 60% capacity, they consume 90% of maximum power. The servers could be turned off, but they would take too long to restart in response to more load. A new system has been proposed that allows for a quick restart but requires 20% of the maximum power while in this "barely alive" state.

a. How much power savings would be achieved by turning off 60% of the servers?
b. How much power savings would be achieved by placing 60% of the servers in the "barely alive" state?
c. How much power savings would be achieved by reducing the voltage by 20% and frequency by 40%?
d. How much power savings would be achieved by placing 30% of the servers in the "barely alive" state and 30% off?

12. Assume for a given processor the CPI of arithmetic instructions is 1, the CPI of load/store instructions is 10, and the CPI of branch instructions is 3. Assume a program has the following instruction breakdowns: 100 million arithmetic instructions, 20 million load/store instructions, 20 million branch instructions.

a. Suppose that new, more powerful arithmetic instructions are added to the instruction set. On average, through the use of these more powerful arithmetic instructions, we can reduce the number of arithmetic instructions needed to execute a program by 25%, while increasing the clock cycle time by only 10%. Is this a good design choice? Why?
b. Suppose that we find a way to double the performance of arithmetic instructions. What is the overall speedup of our machine? What if we find a way to improve the performance of arithmetic instructions by 10 times?