Problem 1 (25 points)

Please give the definitions for the following items.

(a) **performance**: In the field of computer engineering, an informal definition of performance is: it describes how fast a computer can run. For different applications, we need different metrics to measure the performance of computer systems. Generally, **response time** and **throughput** are the two most important aspects in evaluating the performance of a computer system. **Response time** refers to the time between the start and completion of a task running on a computer system. It is also called **execution time**. **Throughput** is the total amount of work done in a given time by a computer system. In many real computer systems, changing either execution time or throughput often affects the other.

(b) **latency**: Latency is a measure of the amount of time between the start of an action and its completion. For an instruction executed in the computer, its latency is the inherent execution time.

(c) **wall clock** time: Generally, it means the time of the real world, a kind of **absolute** time. It is used to measure the performance of a computer system.

(d) **weighted CPI**: Ideally, every instruction in an ISA has a fixed number of CPI. Suppose in MIPS, every instruction is executed in 5 pipeline stages, so their CPI is 5. In reality, each instruction of an ISA has a different CPI. For example, an "add" instruction may take 5 cycles, while a floating point divide instruction takes 18 cycles. Therefore, different instructions, have different CPIs. This is mainly caused by the inherent properties of different operations. "Floating point operations" are more complicated than "integer operations". So, "floating point instructions" have larger CPI than "integer instructions". Usually, this is called the weighted CPI.

(e) **system time**: The CPU time spent in the operating system performing tasks on behalf of the program.

Problem 2 (20 points)
(a) The performance data obtained by measuring CPI at run time is more accurate than measuring it by hand statically.

(b) Sometimes, you have to evaluate the performance of the specific computer architecture when the hardware is still not available, e.g. the pre-silicon stage. In this case, you can estimate the performance by calculating the CPI statically.

Problem 3 (20 points)

When using C1, I1 performs at 1.0667 times I2’s rate of performance.
When using C2, I2 performs at 1.0625 times I1’s rate of performance.
The combinations of (C1, I1), (C3, I1), and (C3, I2) tie for the highest level of performance among the possible choices.

Problem 4 (35 points)

f is in data section
e is in bss section
all others are in stack
The program prints: 11 - 11