Q1

The following is a simple function that divides the input by 73 and returns the result:

```
int div2(int num) {

    return num / 73;

}
```

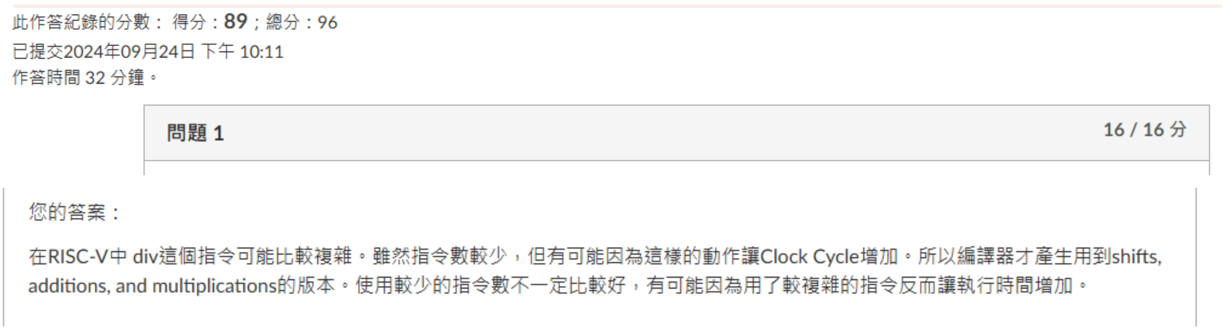
Here is the compiled assembly for RISC-V using GCC:
(https://godbolt.org/z/1Ynx94PjzLinks to an external site.)

```
div73(int):
        li      a5,-529514496
        addi    a5,a5,-2019
        mulh    a5,a0,a5
        srai    a4,a0,31
        add     a5,a5,a0
        srai    a5,a5,6
        sub     a0,a5,a4
        ret
```

This version contains 7 instructions.
Alternatively, the same function could be implemented in only 3 instructions:

```
div73(int):
        li      a1,73
        div     a0,a0,a1
        ret
```

In the 7-instruction version, there is no explicit division instruction. Instead, it uses shifts, additions, and multiplications to perform the division. Interestingly, the number "73" is not directly present in this version either.
Why does the compiler generate the 7-instruction version instead of the simpler 3-instruction version? Is using fewer instructions always preferable?
Discuss this with the concepts covered in Chapter 1.



此作答紀錄的分數： 得分：**89**；總分：96

已提交2024年09月24日 下午 10:11

作答時間 32 分鐘。

問題 1 **16 / 16 分**

您的答案：

在RISC-V中 div這個指令可能比較複雜。雖然指令數較少，但有可能因為這樣的動作讓Clock Cycle增加。所以編譯器才產生用到shifts, additions, and multiplications的版本。使用較少的指令數不一定比較好，有可能因為用了較複雜的指令反而讓執行時間增加。

Q2

A common pitfall is to utilize a subset of the performance equation as a performance metric. To illustrate this, consider the following two processors. P1 has a clock rate of 4.5 GHz, average CPI of 0.9, and requires the execution of 5.0E9 instructions. P2 has a clock rate of 3 GHz, an average CPI of 0.7, and requires the execution of 1.0E9 instructions.

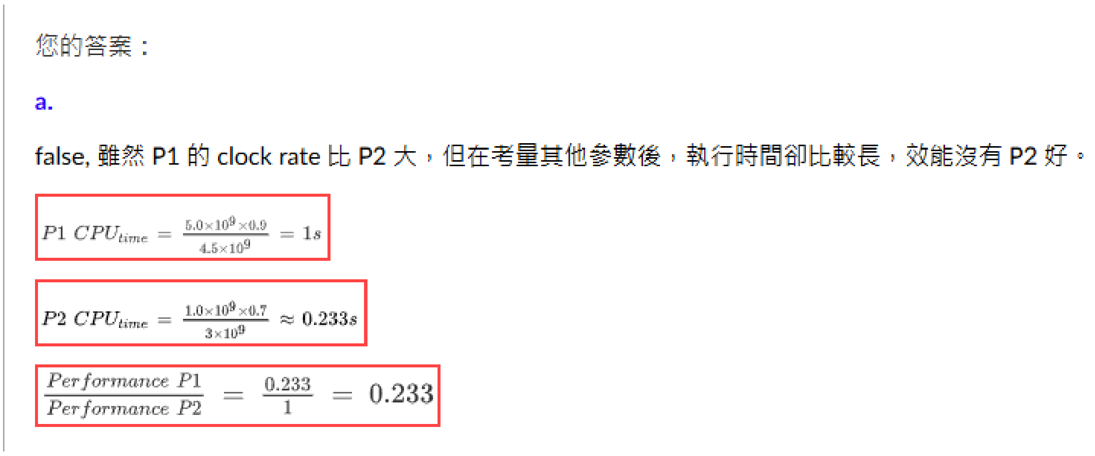
a. One usual fallacy is to consider the computer with the largest clock rate as having the highest performance. Check if this is true for P1 and P2.

b. Another fallacy is to consider that the processor executing the largest number of instructions will need a larger CPU time.

Considering that processor P1 is executing a sequence of 1.0E9 instructions and that the CPI of processors P1 and P2 do not change, determine the number of instructions that P2 can execute in the same time that P1 needs to execute 1.0E9 instructions.

c. A common fallacy is to use MIPS (millions of instructions per second) to compare the performance of two different processors, and consider that the processor with the largest MIPS has the largest performance. Check if this is true for P1 and P2.

問題 2 **16 / 16 分**



您的答案：

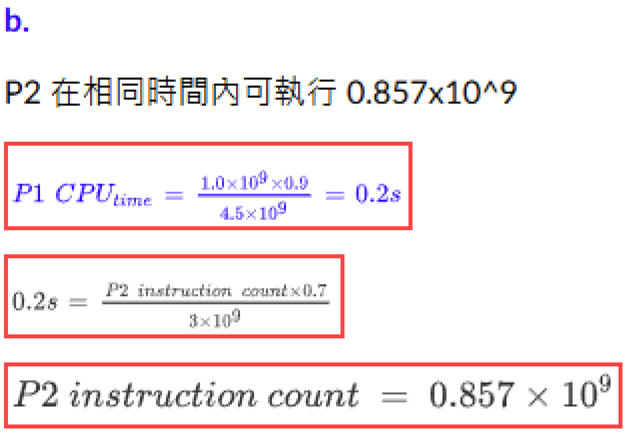
a.

false, 雖然 P1 的 clock rate 比 P2 大，但在考量其他參數後，執行時間卻比較長，效能沒有 P2 好。

$$\boxed{P1\ CPU_{time} = \frac{5.0\times10^9\times0.9}{4.5\times10^9} = 1s}$$

$$\boxed{P2\ CPU_{time} = \frac{1.0\times10^9\times0.7}{3\times10^9} \approx 0.233s}$$

$$\boxed{\frac{Performance\ P1}{Performance\ P2} = \frac{0.233}{1} = 0.233}$$



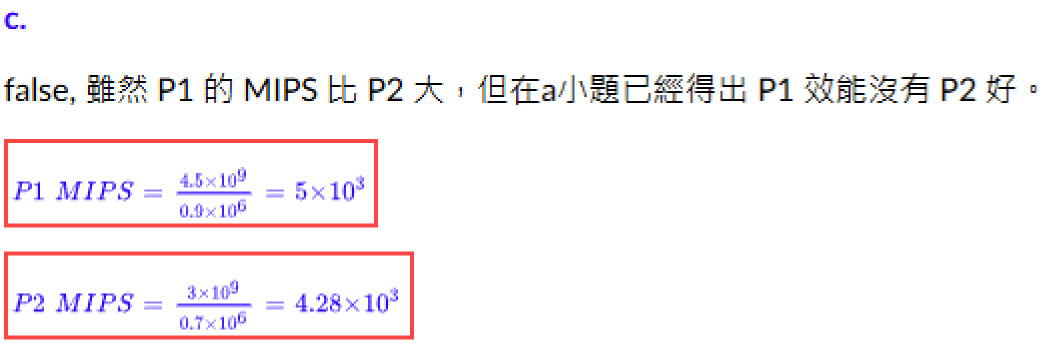
b.

P2 在相同時間內可執行 0.857x10^9

$$\boxed{P1\ CPU_{time} = \frac{1.0\times10^9\times0.9}{4.5\times10^9} = 0.2s}$$

$$\boxed{0.2s = \frac{P2\ instruction\ count\times0.7}{3\times10^9}}$$

$$\boxed{P2\ instruction\ count = 0.857\times10^9}$$



c.

false, 雖然 P1 的 MIPS 比 P2 大，但在a小題已經得出 P1 效能沒有 P2 好。

$$\boxed{P1\ MIPS = \frac{4.5\times10^9}{0.9\times10^6} = 5\times10^3}$$

$$\boxed{P2\ MIPS = \frac{3\times10^9}{0.7\times10^6} = 4.28\times10^3}$$

Q3

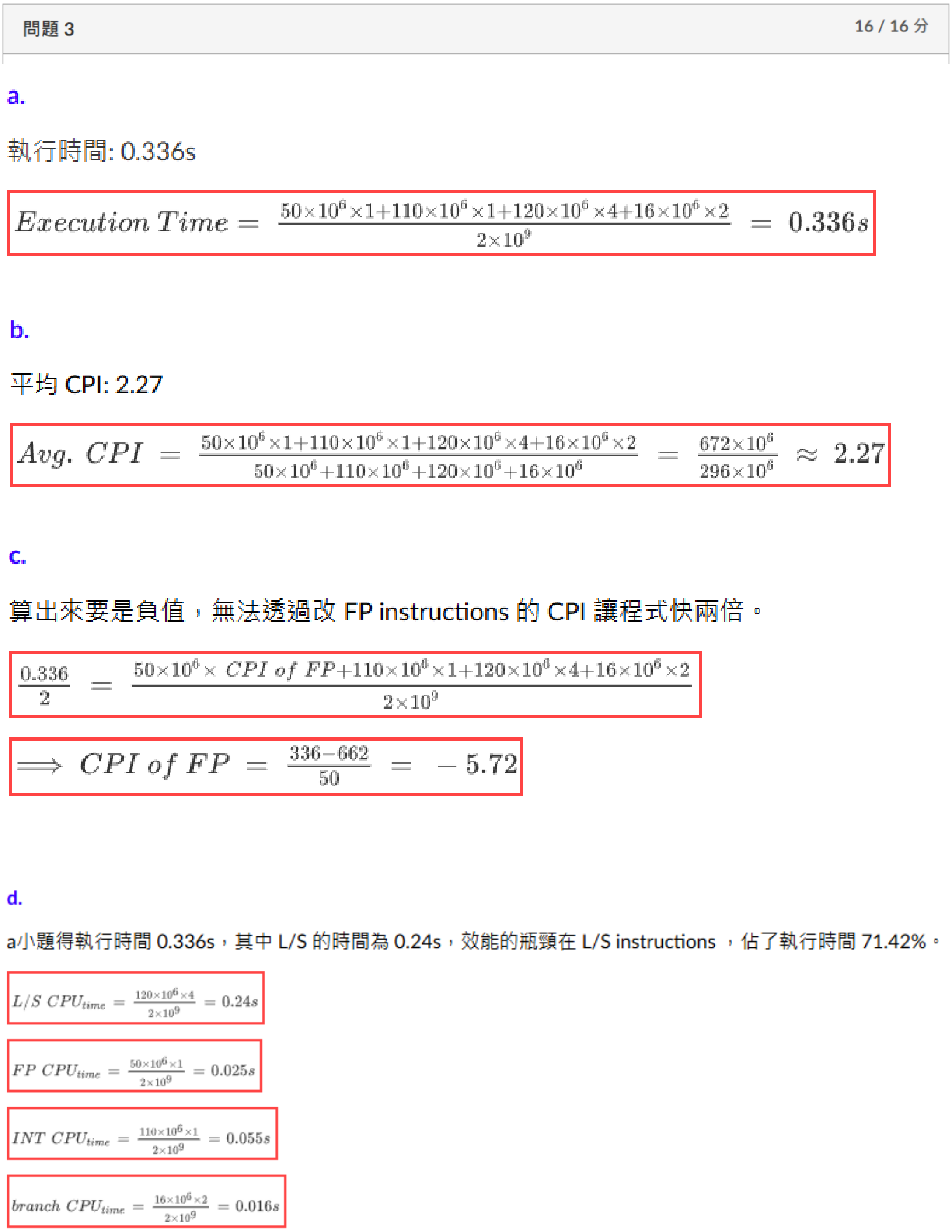
Assume a program requires the execution of 50×10^6 FP instructions, 110×10^6 INT instructions, 120×10^6 L/S instructions, and 16×10^6 branch instructions. The CPI for each type of instruction is 1, 1, 4, and 2, respectively. Assume that the processor has a 2 GHz clock rate.

a. What's the execution time of this program?

b. What is the average CPI when running this program?

c. By how much must we improve the CPI of FP instructions if we want the program to run two times faster?

d. What kinds of instruction takes up the execution time the most? by how much percentage? (What's the performance bottleneck?)



問題 3

16 / 16 分

a.

執行時間: 0.336s

$$\boxed{Execution\ Time = \frac{50\times10^6\times1+110\times10^6\times1+120\times10^6\times4+16\times10^6\times2}{2\times10^9} = 0.336s}$$

b.

平均 CPI: 2.27

$$\boxed{Avg.\ CPI = \frac{50\times10^6\times1+110\times10^6\times1+120\times10^6\times4+16\times10^6\times2}{50\times10^6+110\times10^6+120\times10^6+16\times10^6} = \frac{672\times10^6}{296\times10^6} \approx 2.27}$$

c.

算出來要是負值，無法透過改 FP instructions 的 CPI 讓程式快兩倍。

$$\boxed{\frac{0.336}{2} = \frac{50\times10^6\times\ CPI\ of\ FP+110\times10^6\times1+120\times10^6\times4+16\times10^6\times2}{2\times10^9}}$$

$$\boxed{\Longrightarrow CPI\ of\ FP = \frac{336-662}{50} = -5.72}$$

d.

a小題得執行時間 0.336s，其中 L/S 的時間為 0.24s，效能的瓶頸在 L/S instructions ，佔了執行時間 71.42%。

$$\boxed{L/S\ CPU_{time} = \frac{120\times10^6\times4}{2\times10^9} = 0.24s}$$

$$\boxed{FP\ CPU_{time} = \frac{50\times10^6\times1}{2\times10^9} = 0.025s}$$

$$\boxed{INT\ CPU_{time} = \frac{110\times10^6\times1}{2\times10^9} = 0.055s}$$

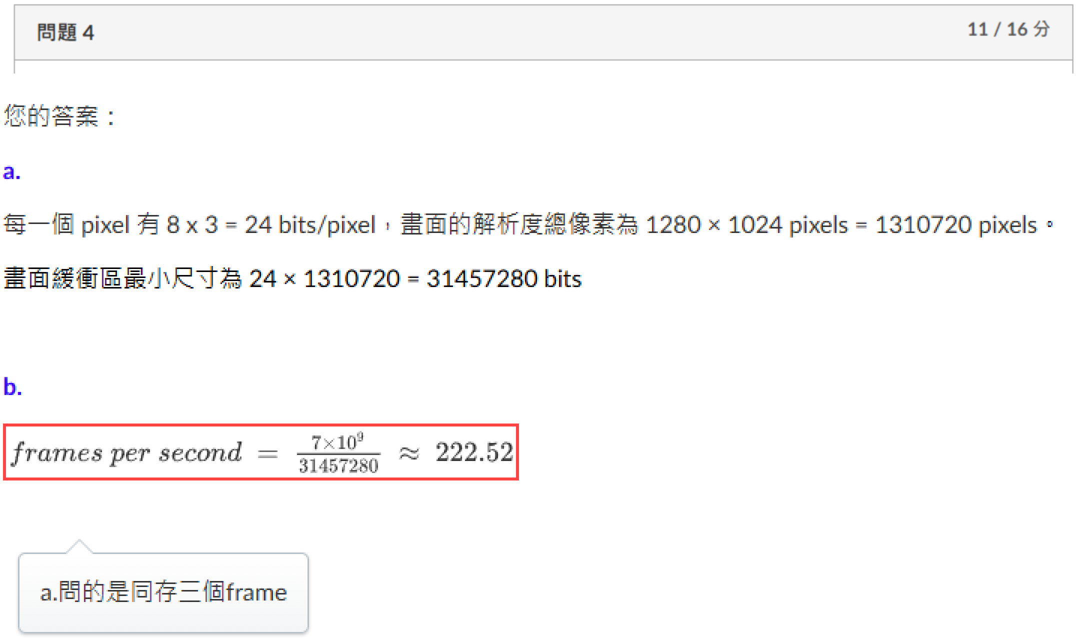
$$\boxed{branch\ CPU_{time} = \frac{16\times10^6\times2}{2\times10^9} = 0.016s}$$

Q4

A color display uses 8 bits per pixel for each of the primary colors: red, green, and blue. The frame resolution is 1280 × 1024 pixels.

a. If the frame buffer needs to store 3 frames simultaneously, what is the minimum size of the frame buffer?

b. Given a 7 Gbps network, what is the maximum number of frames per second that can be transmitted?



問題 4 11 / 16 分

您的答案：

a.

每一個 pixel 有 8 x 3 = 24 bits/pixel，畫面的解析度總像素為 1280 × 1024 pixels = 1310720 pixels。

畫面緩衝區最小尺寸為 24 × 1310720 = 31457280 bits

b.

$$\boxed{frames\ per\ second = \frac{7\times10^9}{31457280} \approx 222.52}$$

a.問的是同存三個frame

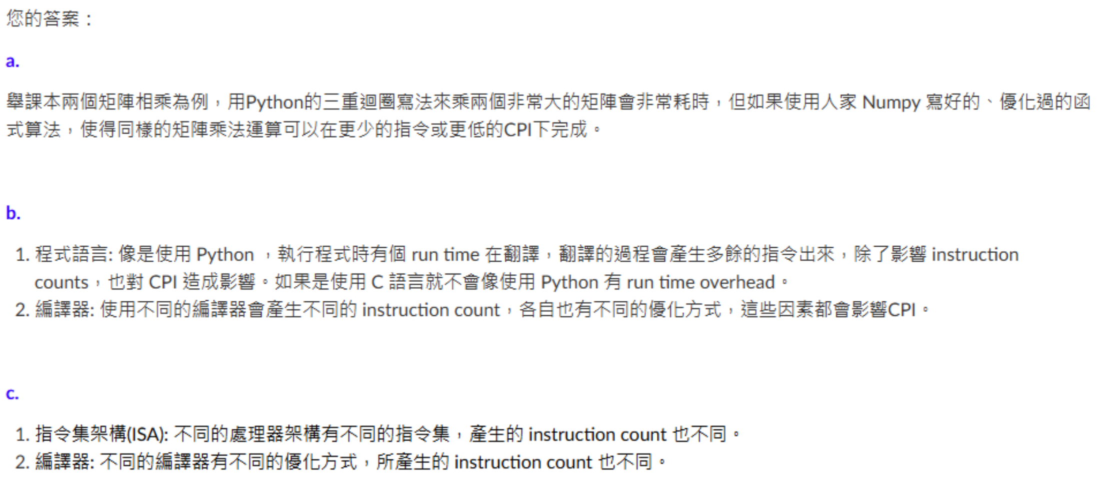
Q5

Clock rates, Cycles Per Instruction (CPI), and instruction counts are factors that influence performance.

a. How do algorithms influence Cycles Per Instruction (CPI)? Provide an example to support your explanation.

b. List and explain two additional factors that impact Cycles Per Instruction (CPI).

c. Identify two factors that affect instruction counts, even when the same algorithm is used. Explain why these factors have an impact.



您的答案：

a.

舉課本兩個矩陣相乘為例，用Python的三重迴圈寫法來乘兩個非常大的矩陣會非常耗時，但如果使用人家 Numpy 寫好的、優化過的函式算法，使得同樣的矩陣乘法運算可以在更少的指令或更低的CPI下完成。

b.

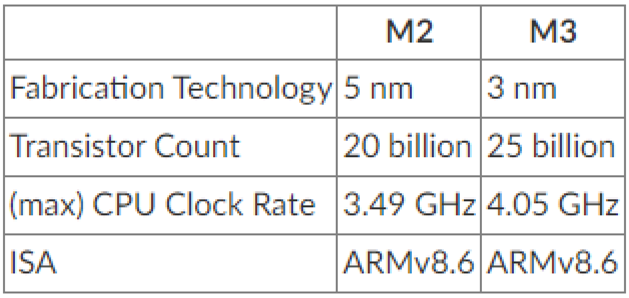
1. 程式語言: 像是使用 Python ，執行程式時有個 run time 在翻譯，翻譯的過程會產生多餘的指令出來，除了影響 instruction counts，也對 CPI 造成影響。如果是使用 C 語言就不會像使用 Python 有 run time overhead。
2. 編譯器: 使用不同的編譯器會產生不同的 instruction count，各自也有不同的優化方式，這些因素都會影響CPI。

c.

1. 指令集架構(ISA): 不同的處理器架構有不同的指令集，產生的 instruction count 也不同。
2. 編譯器: 不同的編譯器有不同的優化方式，所產生的 instruction count 也不同。

Q6

Let's compare Apple's M2 and M3 chips. Their specifications are as follows:



	M2	M3
Fabrication Technology	5 nm	3 nm
Transistor Count	20 billion	25 billion
(max) CPU Clock Rate	3.49 GHz	4.05 GHz
ISA	ARMv8.6	ARMv8.6

To evaluate their performance, we often run them through benchmarks. For instance, using the GeekbenchLinks to an external site. benchmark, we compare two MacBook Pro models—one with the M2 chip and the other with the M3 chip. Their single-core performance scores are 2614 and 3090, respectively.

a. Based on the hardware specifications above, please explain what contributes to this performance improvement.

b. Which chip is likely to be more energy-efficient? Provide reasoning based on the given data.

c. Given that both processors use the same ISA (ARMv8.6), does this mean their average CPI (Cycles Per Instruction) will be identical? Why or why not?

For detailed information, please refer to: M2 MacBook ProLinks to an external site., M3 MacBook ProLinks to an external site.

問題 6

14 / 16 分

您的答案：

a.

製程技術由 M2 的 5 nm 到 M3 的 3 nm，使得晶圓能容納晶體管數量上升到 250 億，以及 Clock rate 的提高都對效能的提升有所貢獻。

b.

M3 具有更好的能源效率。在 Geekbench 中 M3 的單核跟多核分數不管是在資料壓縮或是檔案處理等方面都比 M2 來的好，雖然 clock rate 的提升可能帶來更多的功耗，使其能源效率降低，但如果事情都處理得比 M2 快得多，那麼總體的能量消耗比起 M2 會更低。

c.

雖然是相同的 ISA ，但有可能在硬體層設計的不同，平均 CPI 不一定相同。

b. (-2) 可以多描述製程與電晶體數量可能帶來的影響