

Homework 1

Madiba Hudson-Quansah

Question 1

Energy Efficiency and Performance Trade-offs: As semiconductor companies like Intel, ASML, NVIDIA, and Qualcomm push the limits of Moore's Law, balancing energy efficiency with high-performance computing becomes a critical challenge. How can these companies leverage architectural innovations, advanced materials, and AI-driven optimization techniques to enhance computational power while minimizing energy consumption?

Solution: Companies can leverage architectural innovations by designing chips that are specifically optimized for energy efficiency, such as ARM-based architectures that are known for their low power consumption. Advanced materials like silicon carbide and gallium nitride can be used to create transistors that operate at higher voltages and temperatures, leading to improved performance without a significant increase in energy usage. AI-driven optimization techniques can be employed to dynamically adjust the power and performance levels of chips based on workload requirements, ensuring that energy is only consumed when necessary.

Question 2

Future of Computing Architectures: With the rise of heterogeneous computing, quantum computing, and neuro-morphic processors, the traditional von Neumann architecture faces increasing competition. How should companies like Intel, ASML, NVIDIA, and Qualcomm strategically invest in and develop next-generation computing architectures to maintain leadership in the evolving semiconductor landscape?

Solution:

Traditional silicon producers should invest in research and development to stay ahead of the curve. This includes exploring new materials, such as graphene or carbon nanotubes, which could potentially replace silicon in future chips. Additionally, companies should focus on developing hybrid architectures that combine traditional CPUs with specialized processors like GPUs and TPUs to optimize performance for specific workloads. Collaborations with research institutions and startups in the quantum computing and neuromorphic computing space can also provide valuable insights and innovations that can be integrated into their product lines.

Question 3

We wish to compare the performance of two different computers: M1 and M2. The following measurements have been made on these computers:

1. Which computer is faster for each program, and how many times as fast is it?
2. Find the instruction execution rate (instructions per second) for each computer when running program 1.
3. The clock rates for M1 and M2 are 3 GHz and 5 GHz, respectively. Find the CPI for program 1 on both machines.
4. Suppose that program 1 must be executed 1600 times each hour. Any remaining time should be used to run program 2. Which computer is faster for this workload and why? Performance is measured here by the throughput of program 2.

Solution:

1.

$$\begin{aligned}P1 &= \frac{2}{1.5} \\&= 1.33333 \\P2 &= \frac{10}{5} \\&= 2\end{aligned}$$

M2 is faster for program 1 by about 33%, and M1 is faster for program 2 by about 100%

2.

$$\begin{aligned}M1 &= \frac{5 \times 10^9}{2} \\&= 2.5 \times 10^9 \\M2 &= \frac{6 \times 10^9}{1.5} \\&= 4 \times 10^9\end{aligned}$$

3.

$$\begin{aligned}M1 &= \frac{6 \times 10^9}{5 \times 10^9} \\&= 1.2 \\M2 &= \frac{7.5 \times 10^9}{6 \times 10^9} \\&= 1.25\end{aligned}$$

4.

$$\begin{aligned}M1 &= 2 \times 1600 \\&= 3200 \text{ secs} \\&= 400 \text{ secs left} \\&= \frac{400}{5} \\&= 80 \text{ instructions an hour} \\M2 &= 1.5 \times 1600 &= 2400 \\&= 1200 \text{ secs left} \\&= \frac{1200}{10} \\&= 120 \text{ instructions an hour}\end{aligned}$$

\therefore M2 is faster for this work load as it has more throughput, i.e. more instructions executed each hour.

Question 4

Suppose you wish to run a program P with 7.5×10^9 instructions on a 5 GHz machine with a CPI of 1.2.

1. What is the CPU execution time?
2. When you run program P, it takes 3 seconds of wall time to complete. What is the percentage of the CPU time program P received?

Solution:

1.

$$\begin{aligned}\text{Execution Time} &= \frac{7.5 \times 1.2}{5} \\&= 1.8 \text{ secs}\end{aligned}$$

2.

$$\frac{1.8}{3} \times 100 = 60\%$$

Question 5

Consider two different implementations, M1 and M2, of the same instruction set. There are three classes of instructions (A, B, and C) in the instruction set. M1 has a clock rate of 6 GHz and M2 has a clock rate of 3 GHz. The CPI for each instruction class on M1 and M2 is given in the following table:

The above table also contains a summary of the usage of instruction classes generated by three different compilers: C1, C2, and C3. Assume that each compiler generates the same number of instructions for a given program.

1. Using C1 compiler on both M1 and M2, how much faster is M1 than M2?
2. Using C2 compiler on both M1 and M2, how much faster is M2 than M1?
3. If you purchase M1, which compiler would you use?
4. If you purchase M2, which compiler would you use?
5. Which computer and compiler combination gives the best performance and why?

Solution:

1.

$$\begin{aligned} M1 &= \frac{2 \times 0.4 + 3 \times 0.4 + 5 \times 0.2}{6} \\ &= 0.5 \\ M2 &= \frac{1 \times 0.4 + 2 \times 0.2 + 2 \times 0.4}{3} \\ &= 0.53333333333333 \\ \text{Speedup} &= \frac{0.53333333333333}{0.5} \\ j &= 1.06666666666667 \end{aligned}$$

\therefore M1 is faster than M2 by about 6.67%

2.

$$\begin{aligned} M1 &= \frac{2 \times 0.4 + 3 \times 0.2 + 5 \times 0.4}{6} \\ &= 0.56666667 \\ M2 &= \frac{1 \times 0.4 + 2 \times 0.2 + 2 \times 0.4}{3} \\ &= 0.533333333334 \\ \text{Speedup} &= \frac{0.56666667}{0.533333333334} \\ &= 1.0625 \end{aligned}$$

\therefore M2 is faster than M1 by about 6.25%

3. C1

4. C2

5. M1 and C1 as it has the least execution time for programs with the A, B, C classes of instructions.

Question 6

A benchmark program runs for 100 seconds. We want to improve the speedup of the benchmark by a factor of 3. We enhance the floating-point hardware to make floating-point instructions run 5 times faster. How much of the initial execution time would floating-point instructions have to account for to show an overall speedup of 3 on this benchmark?

Solution:

$$\begin{aligned}3 &= \frac{1}{(1-P) + \frac{P}{5}} \\3 \left((1-P) + \frac{P}{5} \right) &= 1 \\3 - 3P + \frac{3P}{5} &= 1 \\3 - 3P + 0.6P &= 1 \\3 - 2.4P &= 1 \\2 &= 2.4P \\\frac{2}{2.4} &= P \\\frac{5}{6} &= P\end{aligned}$$

∴ Floating point instructions would have to account for about 83.33% of the initial execution time to show an overall speedup of 3 on this benchmark.

Question 7

Consider two different implementations, M1 and M2, of the same instruction set. There are five classes of instructions (A, B, C, D and E) in the instruction set. M1 has a clock rate of 4 GHz and M2 has a clock rate of 6 GHz.

1. Assume that peak performance is defined as the fastest rate that a computer can execute any instruction sequence. What are the peak performances of M1 and M2 expressed in instructions per second?
2. If the number of instructions executed in a certain program is divided equally among the classes of instructions, except that for class A, which occurs twice as often as each of the others, how much faster is M2 than M1?

Solution:

1.

$$\begin{aligned}M1 &= \frac{4 \times 10^9}{1} = 4 \times 10^9 \\M2 &= \frac{6 \times 10^9}{2} = 3 \times 10^9\end{aligned}$$

2. The instruction distribution is therefore:

- A - $\frac{2}{6}$
- B - $\frac{1}{6}$
- C - $\frac{1}{6}$
- D - $\frac{1}{6}$
- E - $\frac{1}{6}$

$$\begin{aligned}
M1 &= \frac{1 \times \frac{2}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 3 \times \frac{1}{6}}{4} \\
&= \frac{7}{12} \\
M2 &= \frac{2 \times \frac{2}{6} + 2 \times \frac{1}{6} + 2 \times \frac{1}{6} + 4 \times \frac{1}{6} + 4 \times \frac{1}{6}}{6} \\
&= \frac{4}{9} \\
\text{Speedup} &= \frac{\frac{7}{12}}{\frac{4}{9}} \\
&= 1.3125
\end{aligned}$$

\therefore M2 is faster than M1 by about 31.25%