COMP 4611 Homework #1 Solutions

Q1. [22 Points] Cost and Price

You are trying to figure out whether to construct a new fabrication facility for your IBM Power5 chips. It costs $1.5 billion to build a new fabrication facility. The benefit of the new fabrication is that you predict that you will be able to sell 3 times as many chips at 2 times the price of the old chips. The new chip will have an area of 186 mm^2, with a defect rate of .7 defects per cm^2. Assume the wafer has a diameter of 300 mm. In both the old and new fabrications, it costs $1000 to fabricate a wafer, and the packaging and testing cost is $20 per chip (after final testing). You were selling the old chips for 40% more than their cost. Assume $\alpha = 4$, Wafer Yield = 1.

The old fabrication process' parameters are shown in the following table.

Chip	Die size (mm^2)	Estimated defect rate (per cm^2)	Manufacturing feature size (nm)	Transistors (millions)
IBM Power5	389	.30	130	276

Note: In this question, you will use equations based on empirical study and approximations. For simplicity, when you need to round to integer numbers, use the mathematical round function (e.g., round 3.24 to 3.2 or round 3.51 to 4) without considering whether the floor or ceiling function would be more appropriate.

Remember not all calculation results need to be rounded to integer numbers.

a) [2 Points] With the old fabrication, how many dies can we get from a wafer before we test individual dies? (Round the result to an integer number)

$$Dies\_per\_wafer = \frac{pi \times (30/2)^2}{3.89} - \frac{pi \times 30}{\sqrt{2 \times 3.89}} = 182 - 34 = 148$$

b) [3 Points] With the old fabrication, what is the die yield?

$$Yield = \left(1 + \frac{0.30 \times 3.89}{4}\right)^{-4} = 0.36$$

c) [3 Points] What is the cost of the old Power5 chip?

$$Cost\_per\_die = \frac{\$1000}{148 \times 0.36} = \$18.77$$

$$Chip\_\cos t = \$18.77 + 20 = \$38.77$$

d) [2 Points] With the new fabrication, how many dies can we get from a wafer before we test individual dies? (Round the result to an integer number)

$$Die\_per\_Wafer = \frac{\pi(300/2)^2}{186} - \frac{\pi \cdot 300}{\sqrt{2 \cdot 186}} = 380 - 49 = 331$$

e) [2 Points] With the new fabrication, what is the die yield?

$$Die\_Yield = 1 \cdot \left(1 + \frac{0.007 \cdot 186}{4}\right)^{-4} = 0.32$$

f) [2 Points] What is the cost of the new Power5 chip?

$$Die\_Cost = \frac{\$1000}{331 \cdot 0.32} = \$9.44$$

$$\boxed{Chip\_\cos t = \$9.44 + 20 = \$29.44}$$

g) [2 Points] What was the selling price of each old Power5 chip?

$$\$38.77 \cdot (1 + 40\%) = \$54.28$$

h) [2 Points] What is the selling price of each new Power5 chip? What is the difference between the selling price and the cost of the new chip?

$$selling\_price = \$54.28 \cdot 2 = \$108.56$$

$$difference\_above\_\cos t = \$108.56 - \$29.44 = \$79.12$$

i) [4 Points] Suppose 50% of the difference between the selling price and the chip cost is your profit. If you sold 500,000 old Power5 chips per month, how long would it take for the accumulated profit to recoup the costs of the new fabrication facility?

$$\frac{1.5 \cdot 10^9}{\frac{79.12}{2} \cdot (500000 \cdot 3)} = 25.28 months$$

Q2. [18 Points] Performance Metrics

An application running on a 1GHz pipelined processor has 55% load-store, 30% arithmetic, and 15% branch instructions. The individual CPIs of these instructions are 5, 4 and 4, respectively.

a) [6 Points] Determine the overall CPI of this program execution on the given processor.

$$CPI = 5 \cdot 55\% + 4 \cdot 30\% + 4 \cdot 15\% = 4.55$$

b) A new embedded version of the processor is being modified to operate at 600 MHz. In this new version, the individual CPIs of load-store and arithmetic instructions are remaining unchanged. However, the individual CPI of branch instructions is getting stretched to 6 clock cycles. A new compiler is also developed for the new processor which eliminates 25% of load-store and 5% of arithmetic instructions for the given application (e.g., if there were 1 million load-store instructions before, now the program compiled by the new compiler would have only 750000).

b1) [6 Points] Determine the overall CPI of this program execution on the new processor together with the new compiler technology.

$$\frac{5 \cdot \left(55\% \cdot (1 - 25\%)\right) + 4 \cdot \left(30\% \cdot (1 - 5\%)\right) + 6 \cdot 15\%}{55\% \cdot (1 - 25\%) + 30\% \cdot (1 - 5\%) + 15\%} = 4.84$$

b2) [6 Points] Determine the factor by which the application will run faster or slower on the new processor with the new compiler technology.

$$
\begin{aligned}
Speedup &= \frac{Old\_Execution\_Time}{New\_Execution\_Time} \\
&= \frac{I_{old} \cdot CPI_{old} \cdot Clock\_Cycle_{old}}{I_{new} \cdot CPI_{new} \cdot Clock\_Cycle_{new}} \\
&= \frac{I_{old} \cdot 4.55 \cdot \left(1/10^9\right)}{\left(I_{old} \cdot \left(55\% \cdot (1 - 25\%) + 30\% \cdot (1 - 5\%) + 15\%\right)\right) \cdot 4.84 \cdot \left(1/600 \cdot 10^6\right)} \\
&= 0.67
\end{aligned}
$$

Q3. [20 Points] Amdahl's Law

Suppose there is a program which takes 200 seconds to execute. Of this time, 30% is used for multiplication, 60% for memory access instructions and 10% for other tasks. Suppose your goal is to enhance the performance of a processor by 2 times and there are two ways of doing so: either make multiply instructions run faster than before, or memory access instructions run faster than before, but not both.

a) [6 Points] How much shall we improve the memory access instructions in order to achieve the performance enhancement? Is it possible?

$$Speedup = \frac{1}{(1-60\%)+\dfrac{60\%}{x}} = 2$$

$x = 6$

Thus, it is possible if we improve the memory access instructions by 6 times.

b) [6 Points] How much shall we improve the multiplication in order to achieve the performance enhancement? Is it possible?

$$Speedup = \frac{1}{(1-30\%)+\dfrac{30\%}{x}} = 2$$

$x = -1.5$

Thus, it is impossible.

c) [8 Points] Suppose we now improve the memory access by 5 times (Time(old execution) / Time(new execution) = 5). Unfortunately, the design makes multiplication slow down by 20% (i.e., Time(old execution) / Time(new execution) = 5/6). What will the overall speedup be?

$$Speedup = \frac{T_{old}}{T_{new}} = \frac{T_{old}}{\left(\dfrac{60\%}{5}+30\%\cdot(1+20\%)+10\%\right)\cdot T_{old}} = 1.72$$

Q4. [20 Points] Geometric versus Arithmetic Mean

The following is a set of individual benchmark scores for each of the programs in the integer portion of the SPEC2000 benchmark.

Benchmark	Score before improvement	Score after improvement
164.gzip	10	12
175.vpr	14	16
176.gcc	23	28
181.mcf	36	40
186.crafty	9	12
197.parser	12	120
252.eon	25	28
253.perlbmk	18	21
254.gap	30	28
255.vortex	17	21
256.bzip2	7	10
300.twolf	38	42

a) [6 Points] Calculate the speedup after the improvement using the arithmetic mean.

$$Arithmetic\_mean_{old} = \frac{10+14+23+36+9+12+25+18+30+17+7+38}{12} = 19.92$$

$$Arithmetic\_mean_{new} = \frac{12+16+28+40+12+120+28+21+28+21+10+42}{12} = 31.5$$

Speedup = 31.5/19.92 = 1.58

b) [6 Points] Calculate the speedup after the improvement using the geometric mean.

$$Geometric\_mean_{old} = \sqrt[12]{10\times14\times23\times36\times9\times12\times25\times18\times30\times17\times7\times38} = 17.39$$

$$Geometric\_mean_{new} = \sqrt[12]{12\times16\times28\times40\times12\times120\times28\times21\times28\times21\times10\times42} = 24.42$$

Speedup = 24.42/17.39 = 1.40.

c) [8 Points] Note that there is a difference between the above two improvement ratios, what is the main reason?

The arithmetic mean is much more sensitive to large changes in one of the values in the set than the geometric mean. Most of the individual benchmarks see relatively small changes as we add the improvement to the architecture, but 197.parser improves from 12 to 120 while all other scores are no higher than 42. This causes the arithmetic mean to increase by almost 60 percent, while the geometric mean increases by only 40 percent. This reduced sensitivity to individual values is why many benchmarking systems prefer the geometric mean for averaging the results of multiple benchmarks, since one particularly large score in the set of benchmark scores has less of an impact on the overall score.

Q5. [20 Points] Endian-ness

Is the Internet Big Endian or Little Endian? Give your answer and provide justification and evidence. The evidence can be a protocol, a piece of software or other real-world technical used in the Internet.

The Internet is Big Endian. The IP protocol, which is a fundamental part of the TCP/IP protocols used in the Internet, aligns bytes in a big-endian manner and transmits the most significant bit first in data communication. The following specification from RFC 791 (Internet Protocol) specifies these facts:

```
"Whenever an octet represents a numeric quantity the left most bit in the
diagram is the high order or most significant bit.  That is, the bit labeled
0 is the most significant bit.  For example, the following diagram represents
the value 170 (decimal).
```

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|1 0 1 0 1 0 1 0|
+-+-+-+-+-+-+-+-+
```

Significance of Bits

Figure 11.

Similarly, whenever a multi-octet field represents a numeric quantity the
left most bit of the whole field is the most significant bit. When a multi-
octet quantity is transmitted the most significant octet is transmitted
first."