CMSC611: Advanced Computer Architecture
Homework 1 Solutions

Question 1: (25 points)

Based on given facts, we first need to find out the number of dies per wafer by

$$\begin{aligned}\text{DiePerWafer} &= \frac{\pi\times(\text{WaferDiameter}/2)^2}{\text{DieArea}} - \frac{\pi\times\text{WaferDiameter}}{\sqrt{2\times\text{DieArea}}}\\ &= \frac{3.14\times(30/2)^2}{1.5\times1.5} - \frac{3.14\times30}{\sqrt{2\times1.5\times1.5}}\\ &\approx 270\end{aligned}$$

Since we could ignore bad wafers, the wafer yield is 100%. So, the die yield is

$$\begin{aligned}\text{DieYield} &= \text{WaferYield}\times\left(1+\frac{\text{DefectsPerUnitArea}\times\text{DieArea}}{\alpha}\right)^{-\alpha}\\ &= 100\%\times\left(1+\frac{0.04\times1.5\times1.5}{4}\right)^{-4}\\ &\approx 0.9148\end{aligned}$$

Now, we could estimate the die cost as

$$\begin{aligned}\text{DieCost} &= \frac{\text{WaferCost}}{\text{DiesPerWafer}\times\text{DieYield}}\\ &= \frac{6000}{270\times0.9148}\\ &\approx 24.29\end{aligned}$$

If we want to make a profit of 40% of costs, the selling price should cover both the cost and the profit, which is \$24.29 × (1 + 40%) ≈ \$34.

Question 2: (25 points)

a) In order to compute the average CPI, all given instruction types need to be considered as

$$\begin{aligned}\text{AverageCPI}_A &= \frac{\text{CPUClockCycles}_A}{\text{InstructionCount}_A}\\ &= \frac{(1\times6+4\times12+2\times14+5\times9+2\times15)\times10^9}{(6+12+14+9+15)\times10^9}\\ &\approx 2.80\end{aligned}$$

$$\begin{aligned}\text{AverageCPI}_B &= \frac{\text{CPUClockCycles}_B}{\text{InstructionCount}_B}\\ &= \frac{(2\times8+1\times10+4\times13+3\times14+4\times18)\times10^9}{(8+10+13+14+18)\times10^9}\\ &\approx 3.05\end{aligned}$$

b) The total execution time for computer A is

$$\begin{aligned}\text{ExecutionTime}_A &= \frac{\text{ClockCycles}_A}{\text{ClockRate}_A} \\ &= \frac{(1\times6+4\times12+2\times14+5\times9+2\times15)\times10^9}{3\times10^9} \\ &\approx 52.33s\end{aligned}$$

For computer B, it is

$$\begin{aligned}\text{ExecutionTime}_B &= \frac{\text{ClockCycles}_B}{\text{ClockRate}_B} \\ &= \frac{(2\times8+1\times10+4\times13+3\times14+4\times18)\times10^9}{2.7\times10^9} \\ &\approx 71.11s\end{aligned}$$

Computer A has a smaller execution time, thus it is faster than computer B.

c) In this question, we only consider the mixed instruction types with given percentages to calculate the execution time for both computer A and B.

$$\begin{aligned}\text{ExecutionTime}_A &= \frac{\text{InstructionCounts}\times\text{AverageCPI}_A}{\text{ClockRate}_A} \\ &= \frac{50\%\times4+30\%\times5+20\%\times2}{3\times10^9}\times\text{InstructionCounts} \\ &= 1.3\times10^{-9}\times\text{InstructionCounts}\end{aligned}$$

$$\begin{aligned}\text{ExecutionTime}_B &= \frac{\text{InstructionCounts}\times\text{AverageCPI}_B}{\text{ClockRate}_B} \\ &= \frac{50\%\times1+30\%\times3+20\%\times4}{2.7\times10^9}\times\text{InstructionCounts} \\ &\approx 0.81\times10^{-9}\times\text{InstructionCounts}\end{aligned}$$

Computer B has a smaller execution time than computer A. The speedup is

$$\text{Speedup} = \frac{\text{ExecutionTime}_A}{\text{ExecutionTime}_B} \approx 1.6$$

Therefore, computer B is faster than computer A by 1.6x.

Question 3: (25 points)

a) Suppose the percentage of vectorization is *X*. Based on Amdahl's Law, we have

$$\text{Speedup}_{\text{overall}} = \frac{1}{\left(1-\text{Fraction}_{\text{enhanced}}\right)+\frac{\text{Fraction}_{\text{enhanced}}}{\text{Speedup}_{\text{enhanced}}}}$$

$$2 = \frac{1}{\left(1-X\right)+\frac{X}{20}}$$

$$X \approx 52.63\%$$

b) The maximum speedup can be achieved if the percentage of vectorization is 100%. So the maximum speedup is

$$\begin{aligned}\text{MaxSpeedup}_{\text{overall}} &= \frac{1}{\left(1-\text{Fraction}_{\text{enhanced}}\right)+\frac{\text{Fraction}_{\text{enhanced}}}{\text{Speedup}_{\text{enhanced}}}}\\ &= \frac{1}{\left(1-100\%\right)+\frac{100\%}{20}}\\ &= 20\end{aligned}$$

If only one-half of the maximum speedup is needed with the percentage of vectorization as *X*, using Amdahl's Law

$$\text{Speedup}_{\text{overall}} = \frac{1}{\left(1-\text{Fraction}_{\text{enhanced}}\right)+\frac{\text{Fraction}_{\text{enhanced}}}{\text{Speedup}_{\text{enhanced}}}}$$

$$\frac{20}{2} = \frac{1}{\left(1-X\right)+\frac{X}{20}}$$

$$X \approx 94.74\%$$

c) Based on Amdahl's Law, the speedup that the hardware group could achieve is

$$\begin{aligned}\text{Speedup}_{\text{overall}} &= \frac{1}{\left(1-\text{Fraction}_{\text{enhanced}}\right)+\frac{\text{Fraction}_{\text{enhanced}}}{\text{Speedup}_{\text{enhanced}}}}\\ &= \frac{1}{\left(1-70\%\right)+\frac{70\%}{2\times 20}}\\ &\approx 3.1496\end{aligned}$$

For the compiler crew, they have to increase the original percentage of vecorization to *X* for achieving the same speedup as what the hardware group does. *X* can be found by using Amdahl's Law again as

$$\text{Speedup}_{\text{overall}} = \frac{1}{\left(1 - \text{Fraction}_{\text{enhanced}}\right) + \frac{\text{Fraction}_{\text{enhanced}}}{\text{Speedup}_{\text{enhanced}}}}$$

$$3.1496 = \frac{1}{\left(1 - X\right) + \frac{X}{20}}$$

$X \approx 71.84\%$

So, the increase of the percentage of vectorization is 1.84%. Comparing to a significant additional engineering investment required by the hardware group, adding a small amount of the percentage of vectorization costs less. It is defiantly worth investing the compiler crew!

Question 4: (25 points)

a) Based on listed benchmark results for two servers,

$$\text{ExecutionTime}_{\text{Dell}} = 498 + 593 + 528 + 711 = 2330s$$

$$\text{ExecutionTime}_{\text{HP}} = 475 + 564 + 612 + 687 = 2338s$$

Dell server is slightly faster than the HP server.

b) Based on the benchmark result of "403.gcc", Dell server costs less execution time than HP server. So, Dell server should be better than HP server on compiling C programs.

c) Based on the benchmark result of "464.h264ref", HP server costs less execution time than Dell server. So, HP server should be better than Dell server on video compression.

d) We can break down the workload of this problem into different proportions. The program runs 20 edit/test iterations for debugging. For each iteration, the application uses bzip to decompress video sources for 6 times and applies h264 to compress videos for 6 times, while the code needs to be compiled by gcc only once. After that, the code is checked for once by a source control system, which performances can be estimated by the Perl benchmark. With those facts, we can find out the percentage of workload for different benchmarks as

400.perlbench: 1 / [20 × (6 + 6 + 1) + 1] = 0.38%
401.bzip2: (20 × 6) / [20 × (6 + 6 + 1) + 1] = 45.98%
403.gcc: (20 × 1) / [20 × (6 + 6 + 1) + 1] = 7.66%
464.h264ref: (20 × 6) / [20 × (6 + 6 + 1) + 1] = 45.98%

So, the performances of two servers are

$$\begin{aligned}\text{Performance}_{\text{Dell}} &= \frac{1}{\text{ExecutionTime}_{\text{Dell}}} \\ &= \frac{1}{0.38\% \times 498 + 45.98\% \times 593 + 7.66\% \times 528 + 45.98\% \times 711} \\ &\approx 1.56 \times 10^{-3}\end{aligned}$$

$$\begin{aligned}\text{Performance}_{\text{HP}} &= \frac{1}{\text{ExecutionTime}_{\text{HP}}} \\ &= \frac{1}{0.38\% \times 475 + 45.98\% \times 564 + 7.66\% \times 612 + 45.98\% \times 687} \\ &\approx 1.6 \times 10^{-3}\end{aligned}$$

The HP server is recommended for this work.