Homework Assignment #1
SAMPLE SOLUTION

Q1. (5 points) Please briefly describe the steps that transform a program written in a high-level language such as C into a representation that is directly executed by a computer processor.

ANSWER: (more explanations and/or citing figures in the slides are acceptable too)
The program is compiled by a compiler into an assembly language program, which is then assembled into a machine-language program with binary representation of instructions that can be directly executed by a computer processor.

Q2. (5 points) What are the five classic components in all kinds of computers?

ANSWER:
Five classic components include input device, output device, memory, datapath (or arithmetic/logic component), and control component.

Q3. (10 points) Please briefly describe: 1) what the Moore's Law is; and 2) what the Pollack's rule is.

ANSWER:
Moore's Law is an observation and prediction of the growth rate of the number of transistors per chip, which states that the number of transistors per chip doubles roughly every 18 to 24 months.

Pollack's rule indicates that the performance improvement (i.e., the return of the investment) from microarchitecture advances is roughly proportional to the square root of the investment of resources (e.g., the number of transistors integrated on a chip).

Q4. (10 points) Please briefly explain the following acronyms/terminologies we discussed in the class: **SISD, SIMD, MISD, and MIMD.**

ANSWER:

- SISD stands for Single Instruction Single Data stream architecture, which represents a serial (Von Neumann) machine.
- SIMD stands for Single Instruction Multiple Data stream architecture, which is mostly suited for problems with high degree of regularity (e.g. GPU architecture).
- MISD stands for Multiple Instruction Single Data stream architecture, which is a hypothetical machine architecture with few actual examples exist.
- MIMD stands for Multiple Instruction Multiple Data stream architecture, which is the most common parallel computer architecture.

Q5. Consider three different processors P1, P2, and P3 executing the same instruction set. P1 has a 3 GHz clock rate and a CPI of 1.5. P2 has a 2.5 GHz clock rate and a CPI of 1.0. P3 has a 4.0 GHz clock rate and has a CPI of 2.2.

a. **(5 points)** Which processor has the highest performance expressed in instructions per second?
b. **(5 points)** If the processors each execute a program in 10 seconds, find the number of cycles and the number of instructions.
c. **(5 points)** We are trying to reduce the execution time by 30% but this leads to an increase of 20% in the CPI. What clock rate should we have to get this time reduction?

ANSWER:
a. in terms of instructions per second,
Performance of P1 (instructions/sec) = $3.0 * 10^9/1.5 = 2.0 * 10^9$
Performance of P2 (instructions/sec) = $2.5 * 10^9/1.0 = 2.5 * 10^9$
Performance of P3 (instructions/sec) = $4.0 * 10^9/2.2 = 1.8 * 10^9$
Thus, P2 has the highest performance in terms of the instructions per second.

b.
The number of cycles for P1 is: $10 * 3.0 * 10^9 = 3.0 * 10^{10}$ cycles
The number of cycles for P2 is: $10 * 2.5 * 10^9 = 2.5 * 10^{10}$ cycles
The number of cycles for P3 is: $10 * 4.0 * 10^9 = 4.0 * 10^{10}$ cycles

The number of instructions for P1 to execute in 10 seconds is: $10 * 3.0 * 10^9$ / CPI = $3.0 * 10^{10}$ / $1.5 = 2.0 * 10^{10}$ instructions
The number of instructions for P2 to execute in 10 seconds is: $10 * 2.5 * 10^9$ / CPI = $2.5 * 10^{10}$ / $1.0 = 2.5 * 10^{10}$ instructions
The number of instructions for P3 to execute in 10 seconds is: $10 * 4.0 * 10^9$ / CPI = $4.0 * 10^{10}$ / $2.2 = 1.82 * 10^{10}$ instructions

c.
Given that $CPI_{new} = CPI_{old} * 1.2$, then $CPI_{new}(P1) = 1.5 * 1.2 = 1.8$, $CPI_{new}(P2) = 1.0 * 1.2 = 1.2$, $CPI_{new}(P3) = 2.2 * 1.2 = 2.64$;

If the CPU execution time is reduced by 30%, then the CPU execution time is 7 seconds. Since the clock rate or the frequency = Instruction Count * CPI/CPU Time, then:
frequency(P1) = $2.0 * 10^{10} * 1.8/7 = 5.14$ GHz
frequency(P2) = $2.5 * 10^{10} * 1.2/7 = 4.29$ GHz
frequency(P3) = $1.82 * 10^{10} * 2.64/7 = 6.86$ GHz

Q6. (10 points) Consider two different implementations of the same instruction set architecture. The instructions can be divided into four classes according to their CPI (class A, B, C, and D). P1 with a clock rate of 2.5 GHz and CPIs of 1, 2, 3, and 3, and P2 with a clock rate of 3 GHz and CPIs of 2, 2, 2, and 2.

Given a program with a dynamic instruction count of 1.0E6 instructions divided into classes as follows: 10% class A, 20% class B, 50% class C, and 20% class D, which implementation is faster: P1 or P2? What is the global CPI (i.e., weighted average CPI) for each implementation?

ANSWER:

Since class A instructions are 10% of total 1.0E6 instructions, the instruction count for class A is thus: 10% * 1.0 * 10^6 = 1.0 * 10^5 instructions. Similarly, we can calculate that the instruction count for class B, C, and D, is 2.0 * 10^5 instructions, 5.0 * 10^5 instructions, and 2.0 * 10^5 instructions, respectively.

Since CPU Time = CPU Clock Cycles / Clock Rate, thus:
The CPU time for P1 is: CPU Clock Cycles / Clock Rate
= $\sum_{i=1}^{4}(CPI_i \times Instruction\ Count_i)$ / Clock Rate
= (1 * 1.0 * 10^5 + 2 * 2.0 * 10^5 + 3 * 5.0 * 10^5 + 3 * 2.0 * 10^5)/(2.5 * 10^9)
= 1.04 * 10^{-3} seconds

Similarly, the CPU time for P2 is: CPU Clock Cycles / Clock Rate
= $\sum_{i=1}^{4}(CPI_i \times Instruction\ Count_i)$ / Clock Rate
= (2 * 1.0 * 10^5 + 2 * 2.0 * 10^5 + 2 * 5.0 * 10^5 + 2 * 2.0 * 10^5)/(3.0 * 10^9)
= 6.67 * 10^{-4} seconds

Thus, P2 implementation is faster.

Since the global CPI (or the weighted average CPI) is Clock Cycles / Instruction Count, and Clock Cycles = CPU Time * Clock Rate, thus:
CPI(P1) = 1.04 * 10^{-3} * 2.5 * $10^9/10^6$ = 2.6
CPI(P2) = 6.67 * 10^{-4} * 3.0 * $10^9/10^6$ = 2.0

Q7. The results of the SPEC CPU2006 bzip2 benchmark running on an AMD Barcelona has an instruction count of 2.389E12, an execution time of 750 s, and a reference time of 9650 s.

a. **(5 points)** Find the CPI if the clock cycle time is 0.333 ns.
b. **(5 points)** Find the SPECratio.
c. **(5 points)** Suppose that we are developing a new version of the AMD Barcelona processor with a 4 GHz clock rate. We have added some additional instructions to the instruction set in such a way that the number of instructions has been reduced by 15%. The execution time is reduced to 700 s and the new SPECratio is 13.7. Find the new CPI.

ANSWER:
a.
Since CPI = CPU Time * Clock Rate / Instruction Count, and Clock Rate = 1/Cycle Time = 1 / 0.333ns = 3GHz, we have:
CPI = 750 * 3.0 * 10^9 / (2.389 * 10^{12}) = 0.94

b.
Since SPECratio = Execution Time$_{reference}$ / Execution Time$_{target}$, thus we have:
SPECratio(bzip2) = 9650 / 750 = 12.87

c.
Since CPI = Clock Cycles / Instruction Count = CPU Time * Clock Rate / Instruction Count, we have:
CPI = (700 * 4.0 * 10^9) / ((1-15%) * 2.389 * 10^{12}) = 1.38

Q8. (5 points) Suppose a new CPU reduces the voltage by 20% and reduces the frequency by 40%, while having the same capacitive load as the old CPU, how much power savings would be achieved compared with the old CPU?

ANSWER:
Since the CPU power consumption is proportional to: $Capacitive\ load \times Voltage^2 \times Frequency$, we have:

$$\frac{Power_{new}}{Power_{old}} = \frac{Capacitive\ load \times (Voltage \times 0.80)^2 \times (Frequency \times 0.60)}{Capacitive\ load \times Voltage^2 \times Frequency} = 0.384$$

In other words, the new CPU consumes 38.4% of the power that the old CPU consumes, which means the new CPU achieves a power savings of 61.6%.

Q9. (10 points) We discussed the paper "Amdahl's Law in the Multicore Era" by Prof. Mark Hill (University of Wisconsin-Madison) and Dr. Michael R. Marty (Google) in class and analyzed the speedup formula of one type of asymmetric architecture, as shown as below (Lecture 4, Slide 21). However, for this asymmetric multicore architecture (Lecture 4, Slide 18), we assume there's only one powerful *r*-BCE core and the rest are *n-r* base cores (i.e., 1 + *n* - *r* cores in total per chip). Assume we create another asymmetric multicore architecture, with **two *r*-BCE cores** (large cores) and **the rest are all *q*-BCE cores *(r > q)*** (small cores). What is the speedup formula for this asymmetric multicore architecture?

$$Speedup_{asymmetric}(f, n, r) = \frac{1}{\frac{1-f}{perf(r)} + \frac{f}{perf(r)+n-r}}$$

ANSWER:
Given an asymmetric multicore architecture, with two *r*-BCE cores (large cores) and the rest are all *q*-BCE cores *(r > q)* (small cores), since the (1- *f*) portion of the computation can only be solved by one core (i.e., cannot be benefited from the parallelism of multiple cores), thus that portion takes the below time to solve (*T* is the original execution time):

$$\frac{1-f}{perf(r)}T$$

For the portion that can be benefited from the parallelism of multiple cores, since now we have two *r*-BCE cores, each giving us *perf(r)* performance, and the rest are *(n – 2r)/q* number of *q*-BCE cores, each giving us *perf(q)* performance, thus that portion takes the below time to solve:

$$\frac{f}{2 * perf(r) + perf(q) * (n - 2 * r)/q} T$$

Thus, the speedup would be:

$$Speedup(f, n, r) = \frac{T}{\frac{1-f}{perf(r)} T + \frac{f}{2*perf(r)+perf(q)*(n-2*r)/q} T} = \frac{1}{\frac{1-f}{perf(r)} + \frac{f}{2*perf(r)+perf(q)*(n-2*r)/q}}$$

Q10. (15 points) Please pick one of the papers listed below, read the paper in detail, and write a short summary of the paper you studied. Please limit to maximum 400 words, and please focus on what problem is studied in the paper and what are the key conclusions.

[1] Gene M. Amdahl, "Validity of the Single-Processor Approach to Achieving Large Scale Computing Capabilities", 1967.

[2] John L. Gustafson, "Reevaluating Amdahl's Law", 1988.

[3] Hill & Marty, "Amdahl's Law in the Multicore Era", IEEE Computer 2008.

[4] X.-H. Sun and Y. Chen, "Reevaluating Amdahl's Law in the Multicore Era," Journal of Parallel and Distributed Computing, vol. 70, no. 2, pp. 183-188, Feb 2010.

N/A

THE END.