COE 501: Computer Architecture

Problem Set 1: Fundamentals of Quantitative Design and Analysis

1) (6 pts) One challenge for architects is that the design created today will require several years of implementation, verification, and testing before appearing on the market. This means that the architect must project what the technology will be like several years in advance.

a) (2 pts) Assuming a 35% increase per year in the number of transistors (according to Moore's Law), if the core i7 chip had 1.17 billion transistors in 2010, how many transistors are expected to be integrated on a processor chip in 2020?

b) (2 pts) The current increase in clock rates is only 1% per year, what is the projected clock frequency in 2020, given that the core i7 chip in 2010 had a 3.33 GHz clock frequency.

c) (2 pts) The rate of growth for DRAM capacity has also slowed down. The rate is about 25% increase in capacity per year recently. If the DRAM capacity was 2 Gbits in 2010, what is the projected capacity of a DRAM chip in 2020? Round the result to the nearest power of 2.

2) (6 pts) The following table shows the execution times of three SPECfp2000 benchmarks on the SUN Ultra, the AMD Opteron, and the Intel Itanium2.

Benchmarks	Ultra time (sec)	Opteron time (sec)	Itanium2 time (sec)
swim	3100	125.0	70.7
mgrid	1800	98.0	65.8
apsi	2600	150.0	231.0

a) (3 pts) For these three benchmarks, find the geometric mean of the SPEC ratios for the Opteron and the Itanium2, using the Ultra time as a reference. Based on the geometric mean, which processor would you choose (Opteron or Itanium2)?

Your company is trying to decide between purchasing the Opteron or Itanium 2. You have analyzed your company's applications. It will be running applications similar to apsi 3 times more frequently than applications like swim and mgrid.

b) (2 pts) What is the weighted average execution time for this mix of applications for the Opteron and the Itanium 2?

c) (1 pt) Which processor is faster for this mix of applications and by what speedup factor?

3) (6 pts) A common performance figure is MFLOPS (Millions of Floating-point Operations Per Second), defined as:

MFLOPS = Number of FP operations / (Execution Time × 10^6)

Consider a program running on two different processors P1 and P2. Statistics are shown in the following table. L/S are the Load/Store and FP are the Floating-Point instructions.

Processor	Instruction Count	% of Instructions			CPI			Clock Rate
		L/S	FP	Branch	L/S	FP	Branch	
P1	4×10^9	20%	70%	10%	0.75	0.8	1.5	3 GHz
P2	5×10^9	18%	70%	12%	1.3	1.0	1.3	4 GHz

a) (2 pts) Find the execution time of the program on P1 and P2. Which processor is faster?

b) (2 pts) Find the MFLOPS rate for the program on P1 and P2. Which processor has a higher MFLOPS rate?

c) (2 pts) Find the MIPS rate for the program on P1 and P2. Which processor has a higher MIPS rate?

4) (6 pts) When parallelizing an application, the ideal speedup is speeding up by the number of processors. This is limited by two things: percentage of the application that can be parallelized and the cost of communication. Amdahl's law takes into account the former but not the latter.

a) (2 pts) What is the speedup achieved with *N* processors if 80% of the application is parallelizable, ignoring the cost of communication?

b) (2 pts) What is the speedup with *N* processors if the communication overhead is 0.005 × *N* of the original execution time?

c) (2 pts) What is the speedup with 8 processors if, for every time the number of processors is doubled, the communication overhead is increased by 1% of the original execution time?

5) (10 pts) The following table shows the manufacturing factors for two IBM Power 5 chips.

Chip	Die Size (mm^2)	Defect Rate per cm^2
Old IBM Power 5	389	.30
New IBM Power 5	186	.70

a) (2 pts) What is the yield for the Old and New IBM Power 5? (use $\alpha = 4$)

It costs $1 billion to build a new fabrication facility for the new IBM Power 5 chips. The new IBM Power 5 chip will have an area of 186 mm^2 with a defect rate of .70 defects per cm^2. The wafer has a diameter of 30 cm. You predict that you will be able to sell 3 times as many chips in the new fabrication facility at 2 times the price of the old chips. Assume that it costs $500 to fabricate a wafer in either the old and new fabrication facility. You were previously selling the chips for 40% more than their cost.

b) (2 pts) What is the cost of the old Power5 chip?

c) (2 pts) What is the cost of the new Power5 chip?

d) (2 pt) What is the profit on each new Power5 chip?

e) (2 pts) If you sold 500,000 old Power5 chips per month, how long will it take to regain the costs of the new fabrication facility?

6) (6 pts) A server farm such as Google provides enough compute capacity for the highest request rate of the day. Imagine that most of the time these servers operate at only 40% capacity. These servers could be turned off, but they would take too long to restart in response to more load. A new power-saving system has been proposed that allows for a quick restart of the servers, but requires 20% of the original power while in this low-power "sleep" state.

a) (2 pt) How much power savings would be achieved by placing 60% of the servers in the "sleep" state?

b) (2 pt) How much power savings would be achieved by placing 30% of the servers in the "sleep" state and 30% off?

c) (2 pts) How much power savings would be achieved by reducing the voltage by 20% and frequency by 50%?

7) (10 pts) Availability is the most important consideration for designing servers, followed closely by scalability and throughput.

a) (2 pts) We have a single computer system with a failures in time (FIT) of 20,000. What is the mean time to failure (MTTF) for this computer?

b) (2 pts) It takes one day to repair and get the system running again, what is the availability of the computer system?

c) (2 pts) A cheap cluster is built out of 100 identical computers as described in part (a). If one computer fails, the cluster fails. What is the MTTF for the cluster?

d) (4 pts) In a server farm, a single computer failure does not cause the entire system to crash. Instead, it will reduce the number of requests that can be satisfied at any one time.

A company has 1000 computers, each with MTTF of 2000 days. If a computer fails, it is not repaired. The system experiences a catastrophic failure when 10 computers fail. Then, all 10 computers are repaired. What is the MTTF of the system?