CS2410: Computer Architecture (graduate)
University of Pittsburgh
Fall Semester, 2011

Assignment #1 (due September 19)

NOTE: (1) This is an individual assignment. (2) No late submission will be accepted.

PART I (70 points)

1. (30 points) Case Study 1: Chip Fabrication Cost, *Chapter 1 of Hennessy and Patterson.*

2. (40 points) Case Study 4: Performance, *Chapter 1 of Hennessy and Patterson.*

PART II (30 points) Read the following two papers (posted on the course web page): (1) Ronet et al. and (2) Freitas and Wilcke. Briefly summarize the papers first.

• Discuss new (interesting) things that you learned.

• Do you think you'll have a 64-core x86 processor in your notebook by 2021? Justify your answer. Of course, there is no single answer to this question.

• Do you think the traditional "rotating" hard drives will disappear (replaced by storage class memory) by 2021 (from your desktop or notebook or departmental servers)? Justify your answer.

Submit your work at the class or directly to the mailbox of the instructor (box #276), located in the mail room on 5th floor, Sennott Square.

Case Studies with Exercises by Diana Franklin

Case Study 1: Chip Fabrication Cost

Concepts illustrated by this case study

- Fabrication Cost
- Fabrication Yield
- Defect Tolerance through Redundancy

There are many factors involved in the price of a computer chip. New, smaller technology gives a boost in performance and a drop in required chip area. In the smaller technology, one can either keep the small area or place more hardware on the chip in order to get more functionality. In this case study, we explore how different design decisions involving fabrication technology, area, and redundancy affect the cost of chips.

1.1 [10/10/Discussion] <1.5, 1.5> Figure 1.22 gives the relevant chip statistics that influence the cost of several current chips. In the next few exercises, you will be exploring the trade-offs involved between the AMD Opteron, a single-chip processor, and the Sun Niagara, an 8-core chip.

a. [10] <1.5> What is the yield for the AMD Opteron?

b. [10] <1.5> What is the yield for an 8-core Sun Niagara processor?

c. [Discussion] <1.4, 1.6> Why does the Sun Niagara have a worse yield than the AMD Opteron, even though they have the same defect rate?

1.2 [20/20/20/20/20] <1.7> You are trying to figure out whether to build a new fabrication facility for your IBM Power5 chips. It costs $1 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. Assume it costs $500 to fabricate a wafer in either technology. You were previously selling the chips for 40% more than their cost.

Chip	Die size (mm^2)	Estimated defect rate (per cm^2)	Manufacturing size (nm)	Transistors (millions)
IBM Power5	389	.30	130	276
Sun Niagara	380	.75	90	279
AMD Opteron	199	.75	90	233

Figure 1.22 Manufacturing cost factors for several modern processors. $\alpha = 4$.

a. [20] <1.5> What is the cost of the old Power5 chip?

b. [20] <1.5> What is the cost of the new Power5 chip?

c. [20] <1.5> What was the profit on each old Power5 chip?

d. [20] <1.5> What is the profit on each new Power5 chip?

e. [20] <1.5> If you sold 500,000 old Power5 chips per month, how long will it take to recoup the costs of the new fabrication facility?

1.3 [20/20/10/10/20] <1.7> Your colleague at Sun suggests that, since the yield is so poor, it might make sense to sell two sets of chips, one with 8 working processors and one with 6 working processors. We will solve this exercise by viewing the yield as a probability of no defects occurring in a certain area given the defect rate. For the Niagara, calculate probabilities based on each Niagara core separately (this may not be entirely accurate, since the yield equation is based on empirical evidence rather than a mathematical calculation relating the probabilities of finding errors in different portions of the chip).

a. [20] <1.7> Using the yield equation for the defect rate above, what is the probability that a defect will occur on a single Niagara core (assuming the chip is divided evenly between the cores) in an 8-core chip?

b. [20] <1.7> What is the probability that a defect will occur on one or two cores (but not more than that)?

c. [10] <1.7> What is the probability that a defect will occur on none of the cores?

d. [10] <1.7> Given your answers to parts (b) and (c), what is the number of 6-core chips you will sell for every 8-core chip?

e. [20] <1.7> If you sell your 8-core chips for $150 each, the 6-core chips for $100 each, the cost per die sold is $80, your research and development budget was $200 million, and testing itself costs $1.50 per chip, how many processors would you need to sell in order to recoup costs?

Case Study 2: Power Consumption in Computer Systems

Concepts illustrated by this case study

- Amdahl's Law
- Redundancy
- MTTF
- Power Consumption

Power consumption in modern systems is dependent on a variety of factors, including the chip clock frequency, efficiency, the disk drive speed, disk drive utilization, and DRAM. The following exercises explore the impact on power that different design decisions and/or use scenarios have.

c. [20] <1.2> When the site returned, the number of users allowed to visit the site at one time was limited. Imagine that it was limited to 50% of the customers who wanted to access the site. Assume that each server costs $7500 to purchase and set up. How many servers, per day, could they purchase and install with the money they are losing in sales?

d. [20] <1.2, 1.9> Gap.com had 2.6 million visitors in July 2004 [AP 2005]. On average, a user views 8.4 pages per day on Gap.com. Assume that the high-end servers at Gap.com are running SQLServer software, with a TPCC benchmark estimated cost of $5.38 per transaction. How much would it cost for them to support their online traffic at Gap.com.?

1.9 [10/10] <1.8> The main reliability measure is MTTF. We will now look at different systems and how design decisions affect their reliability. Refer to Figure 1.25 for company statistics.

a. [10] <1.8> We have a single processor with an FIT of 100. What is the MTTF for this system?

b. [10] <1.8> If it takes 1 day to get the system running again, what is the availability of the system?

1.10 [20] <1.8> Imagine that the government, to cut costs, is going to build a supercomputer out of the cheap processor system in Exercise 1.9 rather than a special-purpose reliable system. What is the MTTF for a system with 1000 processors? Assume that if one fails, they all fail.

1.11 [20/20] <1.2, 1.8> In a server farm such as that used by Amazon or the Gap, a single failure does not cause the whole system to crash. Instead, it will reduce the number of requests that can be satisfied at any one time.

a. [20] <1.8> If a company has 10,000 computers, and it experiences catastrophic failure only if 1/3 of the computers fail, what is the MTTF for the system?

b. [20] <1.2, 1.8> If it costs an extra $1000, per computer, to double the MTTF, would this be a good business decision? Show your work.

Case Study 4: Performance

Concepts illustrated by this case study

- Arithmetic Mean
- Geometric Mean
- Parallelism
- Amdahl's Law
- Weighted Averages

In this set of exercises, you are to make sense of Figure 1.26, which presents the performance of selected processors and a fictional one (Processor X), as reported by *www.tomshardware.com*. For each system, two benchmarks were run. One benchmark exercised the memory hierarchy, giving an indication of the speed of the memory for that system. The other benchmark, Dhrystone, is a CPU-intensive benchmark that does not exercise the memory system. Both benchmarks are displayed in order to distill the effects that different design decisions have on memory and CPU performance.

1.12 [10/10/Discussion/10/20/Discussion] <1.7> Make the following calculations on the raw data in order to explore how different measures color the conclusions one can make. (Doing these exercises will be much easier using a spreadsheet.)

a. [10] <1.8> Create a table similar to that shown in Figure 1.26, except express the results as normalized to the Pentium D for each benchmark.

b. [10] <1.9> Calculate the arithmetic mean of the performance of each processor. Use both the original performance and your normalized performance calculated in part (a).

c. [Discussion] <1.9> Given your answer from part (b), can you draw any conflicting conclusions about the relative performance of the different processors?

d. [10] <1.9> Calculate the geometric mean of the normalized performance of the dual processors and the geometric mean of the normalized performance of the single processors for the Dhrystone benchmark.

e. [20] <1.9> Plot a 2D scatter plot with the x-axis being Dhrystone and the y-axis being the memory benchmark.

f. [Discussion] <1.9> Given your plot in part (e), in what area does a dual-processor gain in performance? Explain, given your knowledge of parallel processing and architecture, why these results are as they are.

Chip	# of cores	Clock frequency (MHz)	Memory performance	Dhrystone performance
Athlon 64 X2 4800+	2	2,400	3,423	20,718
Pentium EE 840	2	2,200	3,228	18,893
Pentium D 820	2	3,000	3,000	15,220
Athlon 64 X2 3800+	2	3,200	2,941	17,129
Pentium 4	1	2,800	2,731	7,621
Athlon 64 3000+	1	1,800	2,953	7,628
Pentium 4 570	1	2,800	3,501	11,210
Processor X	1	3,000	7,000	5,000

Figure 1.26 Performance of several processors on two benchmarks.

1.13 [10/10/20] <1.9> Imagine that your company is trying to decide between a single-processor system and a dual-processor system. Figure 1.26 gives the performance on two sets of benchmarks—a memory benchmark and a processor benchmark. You know that your application will spend 40% of its time on memory-centric computations, and 60% of its time on processor-centric computations.

a. [10] <1.9> Calculate the weighted execution time of the benchmarks.

b. [10] <1.9> How much speedup do you anticipate getting if you move from using a Pentium 4 570 to an Athlon 64 X2 4800+ on a CPU-intensive application suite?

c. [20] <1.9> At what ratio of memory to processor computation would the performance of the Pentium 4 570 be equal to the Pentium D 820?

1.14 [10/10/20/20] <1.10> Your company has just bought a new dual Pentium processor, and you have been tasked with optimizing your software for this processor. You will run two applications on this dual Pentium, but the resource requirements are not equal. The first application needs 80% of the resources, and the other only 20% of the resources.

a. [10] <1.10> Given that 40% of the first application is parallelizable, how much speedup would you achieve with that application if run in isolation?

b. [10] <1.10> Given that 99% of the second application is parallelizable, how much speedup would this application observe if run in isolation?

c. [20] <1.10> Given that 40% of the first application is parallelizable, how much *overall system speedup* would you observe if you parallelized it?

d. [20] <1.10> Given that 99% of the second application is parallelizable, how much overall system speedup would you get?