Case Studies with Alternate Exercises
Case Studies with Alternate 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/Discussion] <1.6> 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 effect of different possible design decisions for the IBM Power5.

a. [10] <1.6> What is the yield for the IBM Power5?

b. [Discussion] <1.6> Why does the IBM Power5 have a lower defect rate than the Niagara and Opteron?

1.2 [20/20/10/20] <1.6> It costs $1 billion to build a new fabrication facility. You will be selling a range of chips from that factory, and you need to decide how much capacity to dedicate to each chip. Your Woods chip will be 150 mm² and will make a profit of $20 per defect-free chip. Your Markon chip will be 250 mm² and will make a profit of $25 per defect-free chip. Your fabrication facility will be identical to that for the Power5. Each wafer has a 300 mm diameter.

a. [20] <1.6> How much profit do you make on each wafer of Woods chip?

b. [20] <1.6> How much profit do you make on each wafer of Markon chip?

c. [10] <1.6> Which chip should you produce in this facility?

d. [20] <1.6> If your demand is 50,000 Woods chips per month and 25,000 Markon chips per month, and your facility can fabricate 150 wafers a month, how many wafers should you make of each chip?

<table>
<thead>
<tr>
<th>Chip</th>
<th>Die size (mm²)</th>
<th>Estimated defect rate (per cm²)</th>
<th>Manufacturing size (nm)</th>
<th>Transistors (millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM Power5</td>
<td>389</td>
<td>.30</td>
<td>130</td>
<td>276</td>
</tr>
<tr>
<td>Sun Niagara</td>
<td>380</td>
<td>.75</td>
<td>90</td>
<td>279</td>
</tr>
<tr>
<td>AMD Opteron</td>
<td>199</td>
<td>.75</td>
<td>90</td>
<td>233</td>
</tr>
</tbody>
</table>

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

1.3 [20/20] <1.6> Your colleague at AMD suggests that, since the yield is so poor, you might make chips cheaper if you placed an extra core on the die and only threw out chips on which both processors had failed. We will solve this exercise by viewing the yield as a probability of no defects occurring in a certain area given the defect rate. Calculate probabilities based on each Opteron 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] <1.6> What is the probability that a defect will occur on no more than one of the two processor cores?

b. [20] <1.6> If the old chip cost $20 dollars per chip, what will the cost be of the new chip, taking into account the new area and yield?

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.

1.4 [20/10/20] <1.5> Figure 1.23 presents the power consumption of several computer system components. In this exercise, we will explore how the hard drive affects power consumption for the system.

<table>
<thead>
<tr>
<th>Component type</th>
<th>Product</th>
<th>Performance</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Sun Niagara 8-core</td>
<td>1.2 GHz</td>
<td>72-79W peak</td>
</tr>
<tr>
<td></td>
<td>Intel Pentium 4</td>
<td>2 GHz</td>
<td>48.9-66W</td>
</tr>
<tr>
<td>DRAM</td>
<td>Kingston X64C3AD2 1 GB</td>
<td>184-pin</td>
<td>3.7W</td>
</tr>
<tr>
<td></td>
<td>Kingston D2N3 1 GB</td>
<td>240-pin</td>
<td>2.3W</td>
</tr>
<tr>
<td>Hard drive</td>
<td>DiamondMax 16</td>
<td>5400 rpm</td>
<td>7.0W read/seek, 2.9W idle</td>
</tr>
<tr>
<td></td>
<td>DiamondMax Plus 9</td>
<td>7200 rpm</td>
<td>7.9W read/seek, 4.0W idle</td>
</tr>
</tbody>
</table>

Figure 1.23 Power consumption of several computer components.
a. [20] <1.5> Assuming the maximum load for each component, and a power supply efficiency of 80%, what wattage must the server’s power supply deliver to a system with an Intel Pentium 4 chip, 2 GB 240-pin Kingston DRAM, and one 7200 rpm hard drive?

b. [10] <1.5> How much power will the 7200 rpm disk drive consume if it is idle roughly 60% of the time?

c. [20] <1.5> Given that the time to read data off a 7200 rpm disk drive will be roughly 75% of a 5400 rpm disk, at what idle time of the 7200 rpm disk will the power consumption be equal, on average, for the two disks?

1.5 [10/10/20] <1.5> One critical factor in powering a server farm is cooling. If heat is not removed from the computer efficiently, the fans will blow hot air back onto the computer, not cold air. We will look at how different design decisions affect the necessary cooling, and thus the price, of a system. Use Figure 1.23 for your power calculations.

a. [10] <1.5> A cooling door for a rack costs $4000 and dissipates 14 KW (into the room; additional cost is required to get it out of the room). How many servers with an Intel Pentium 4 processor, 1 GB 240-pin DRAM, and a single 7200 rpm hard drive can you cool with one cooling door?

b. [10] <1.5> You are considering providing fault tolerance for your hard drive. RAID 1 doubles the number of disks (see Chapter 6). Now how many systems can you place on a single rack with a single cooler?

c. [20] <1.5> Typical server farms can dissipate a maximum of 200 W per square foot. Given that a server rack requires 11 square feet (including front and back clearance), how many servers from part (a) can be placed on a single rack, and how many cooling doors are required?

1.6 [Discussion] <1.8> Figure 1.24 gives a comparison of power and performance for several benchmarks comparing two servers: Sun Fire T2000 (which uses Niagara) and IBM x346 (using Intel Xeon processors). This information was reported on a Sun Web site. There are two pieces of information reported: power and speed on two benchmarks. For the results shown, Sun’s Fire T2000 is clearly superior. What other factors might be important, and thus cause someone to choose the IBM x346 if it were superior in those areas?

<table>
<thead>
<tr>
<th></th>
<th>Sun Fire T2000</th>
<th>IBM x346</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power (watts)</td>
<td>298</td>
<td>438</td>
</tr>
<tr>
<td>SPECjbb (op/s)</td>
<td>63,378</td>
<td>39,985</td>
</tr>
<tr>
<td>Power (watts)</td>
<td>330</td>
<td>438</td>
</tr>
<tr>
<td>SPECWeb (composite)</td>
<td>14,001</td>
<td>4,348</td>
</tr>
</tbody>
</table>

*Figure 1.24 Sun power / performance comparison as selectively reported by Sun.*
1.7  [20/20/20/20] <1.6, 1.9> Your company’s internal studies show that a single-core system is sufficient for the demand on your processing power. You are exploring, however, whether you could save power by using two cores.

   a.  [20] <1.9> Assume your application is 80% parallelizable. By how much could you decrease the frequency and get the same performance?

   b.  [20] <1.6> Assume that the voltage may be decreased linearly with the frequency. Using the equation in Section 1.5, how much dynamic power would the dual-core system require as compared to the single-core system?

   c.  [20] <1.6, 1.9> Now assume the voltage may not decrease below 25% of the original voltage. This voltage is referred to as the “voltage floor,” and any voltage lower than that will lose the state. What percent of parallelization gives you a voltage at the voltage floor?

   d.  [20] <1.6, 1.9> Using the equation in Section 1.5, how much dynamic power would the dual-core system require as compared to the single-core system when taking into account the voltage floor?

---

1.8  [10/20/20] <1.2> On August 24, 2005, three Web sites managed by the Gap—Gap.com, OldNavy.com, and BananaRepublic.com—went down for improvements [AP 2005]. The other is for Amazon.com, which was not down, but has better statistics on high-load sales days. The exercises combine the two data sets and require estimating the economic cost to the shutdown.

   a.  [10] <1.2> If the sites were down for just December 12th, what would the cost be?

   b.  [20] <1.2> Imagine that 20% of their business is due to return customers. During the downtime, imagine they lost 1% of their return customers. How much money would they have in their budget to have bought and tested their new upgrade software on separate servers rather than taking down their site for two-week maintenance?
c. [20] <1.2> OldNavy.com had 4.2 million visitors in July 2005 [AP 2005]. On average, a user views 8.9 pages per day on OldNavy.com. Given that a high-end server running SQLServer software, running the database benchmark TPCC, costs $5.38 per transaction, how much does it cost for them to support their online traffic at the two Web sites?

1.9 [10/10] <1.7> 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] <1.7> We have a single processor with an FIT of 150. What is the MTTF for this system?

b. [10] <1.7> If it takes two days to get the system running again, what is the availability of the system?

1.10 [20] <1.7> 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 100 processors? Assume that if one fails, they all fail.

1.11 [20/20/20] <1.2, 1.7> 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] <1.7> If a company has 20,000 computers, and it experiences catastrophic failure only if 2/5 of the computers fail, what is the MTTF for the system?

b. [20] <1.2, 1.7> Given Amazon’s highest quarterly income, and assuming that all computers are equally and fully loaded (which is not true), how much money is lost, per day, to computers failing?

<table>
<thead>
<tr>
<th>Company</th>
<th>Time period</th>
<th>Amount</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gap</td>
<td>3rd qtr 2004</td>
<td>$4 billion</td>
<td>Sales</td>
</tr>
<tr>
<td></td>
<td>4th qtr 2004</td>
<td>$4.9 billion</td>
<td>Sales</td>
</tr>
<tr>
<td></td>
<td>3rd qtr 2005</td>
<td>$3.9 billion</td>
<td>Sales</td>
</tr>
<tr>
<td></td>
<td>4th qtr 2005</td>
<td>$4.8 billion</td>
<td>Sales</td>
</tr>
<tr>
<td></td>
<td>3rd qtr 2004</td>
<td>$107 million</td>
<td>Online sales</td>
</tr>
<tr>
<td></td>
<td>3rd qtr 2005</td>
<td>$106 million</td>
<td>Online sales</td>
</tr>
<tr>
<td>Amazon</td>
<td>3rd qtr 2005</td>
<td>$1.86 billion</td>
<td>Sales</td>
</tr>
<tr>
<td></td>
<td>4th qtr 2005</td>
<td>$2.98 billion</td>
<td>Sales</td>
</tr>
<tr>
<td></td>
<td>4th qtr 2005</td>
<td>108 million</td>
<td>Items sold</td>
</tr>
<tr>
<td></td>
<td>Dec 12, 2005</td>
<td>3.6 million</td>
<td>Items sold</td>
</tr>
</tbody>
</table>

Figure 1.25 Statistics on sales for Gap and Amazon. Data compiled from AP [2005], Internet Retailer [2005], Gamasutra [2005], Seattle PI [2005], MSN Money [2005], Gap [2005], and Gap [2006].
Chapter 1  Case Studies with Alternate Exercises by Diana Franklin

7

c. <1.7> If it costs an extra $500, 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 [10/20/20] <1.8> 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 fastest application for each benchmark.

b. [20] <1.8> 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 memory benchmark.

c. [20] <1.8> Calculate the geometric standard deviation of the dual-processor performance on the memory benchmark. What does this suggest about how much the choice of the particular processor makes to performance?

1.13 [10/10/Discussion] <1.8> 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 30% of its time on memory-centric computations, and 70% of its time on processor-centric computations.

a. [10] <1.8> Calculate the weighted performance of the benchmarks for the Pentium 4 and Athlon 64 X2 3800+.
b. [10] <1.8> How much speedup do you anticipate getting if you move from using a Pentium 4 to an Athlon 64 X2 3800+ on a memory-intensive application suite?

c. [Discussion] <1.8> You are using a dual-core Athlon processor, and you are choosing between two ways to implement the same algorithm. The first is to create a large lookup table to store 4K words of data. When you need the result, you look up the answer. The second method would be to calculate the result in a very tight loop. What are the advantages and disadvantages of each implementation?

1.14 [10/10/20/20] <1.9> 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 75% of the resources, and the other only 25% of the resources.

a. [10] <1.9> Given that 60% of the first application is parallelizable, how much speedup would you achieve with that application if run in isolation?

b. [10] <1.9> Given that 95% of the second application is parallelizable, how much speedup would this application observe if run in isolation?

c. [20] <1.9> Given that 60% of the first application is parallelizable, how much overall system speedup would you observe if you parallelized it, but not the second application?

d. [20] <1.9> How much overall system speedup would you achieve if you parallelized both applications, given the information in parts (a) and (b)?