COE 501: Computer Architecture

Problem Set 1: Fundamentals of Quantitative Design and Analysis

Solution

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?

Expected number of transistors per chip in 2020 = 1.17 Billion × $(1.35)^{10}$ = 23.5 Billion

b) (2 pts) The current increase in clock rates is only 4% 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.

Projected clock frequency in 2020 = 3.33 GHz × $(1.04)^{10}$ = 4.93 GHz

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.

$(1.25)^{10}$ = 9.3 (reduced to 8 since DRAM capacity is only increased by a power of 2)

Projected DRAM capacity in 2020 = 2 Gbits × 8 = 16 Gbits

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)?

Benchmarks	**Ultra time (sec)**	**Opteron time (sec)**	**SPEC ratio Opteron**	**Itanium2 time (sec)**	**SPEC ratio Itanium2**
swim	**3100**	**125.0**	**24.8**	**70.7**	**43.8**
mgrid	**1800**	**98.0**	**18.4**	**65.8**	**27.4**
apsi	**2600**	**150.0**	**17.3**	**231.0**	**11.3**
Geometric Mean			**19.9**		**23.8**

The Itanium2 has a higher geometric mean than the Opteron and should be chosen.

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?

Opteron: weighted average execution time = (125 + 98 + 3×150)/5 = 134.6 sec

Itanium2: weighted average execution time = (70.7 + 65.8 + 3×231)/5 = 165.9 sec

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

The Opteron processor is faster for the above mix (less average execution time).

The speedup factor of Opteron over Itanium2 = 165.9 / 134.6 = 1.23

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 $\times 10^6$)

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?

Time (P1) = $(4 \times 10^9) \times (0.2 \times 0.75 + 0.7 \times 0.8 + 0.1 \times 1.5) / (3 \times 10^9)$ = 1.1467 sec

Time (P2) = $(5 \times 10^9) \times (0.18 \times 1.3 + 0.7 \times 1.0 + 0.12 \times 1.3) / (4 \times 10^9)$ = 1.3625 sec

Processor P1 is faster than P2 for the given program (because less execution time)

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

MFLOPS (P1) = $0.7 \times (4 \times 10^9) / (1.1467 \times 10^6)$ = 2442 MFLOPS

MFLOPS (P2) = $0.7 \times (5 \times 10^9) / (1.3625 \times 10^6)$ = 2569 MFLOPS

Processor P2 has a higher MFLOPS rate for the given program

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

MIPS (P1) = $(4 \times 10^9) / (1.1467 \times 10^6)$ = 3488 MIPS

MIPS (P2) = $(5 \times 10^9) / (1.3625 \times 10^6)$ = 3670 MIPS

Processor P2 has a higher MIPS rate for the given program

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?

Speedup = $1/(0.2 + 0.8/N)$

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

Speedup = $1/(0.2 + 0.8/N + 0.005\times N)$

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?

Number of processors is doubled 3 times: $2^3 = 8$ processors

Communication overhead = (3×0.01) of original execution time

Speedup = $1/(0.2 + 0.8/8 + 3 \times 0.01) = 3.03$

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

Chip	Die Area (mm^2)	Defects per cm^2
Old IBM Power 5	389	0.03
New IBM Power 5	186	0.04

a) (2 pts) What is the yield for the Old and New IBM Power 5?

Use Bose-Einstein Formula: Yield = $(1 + \text{Defects per Area} \times \text{Die Area})^{-N}$ with $N = 6$.

Yield (Old IBM Power 5) = $(1 + 0.03 \times 3.89)^{-6} = 0.5157$

Yield (New IBM Power 5) = $(1 + 0.04 \times 1.86)^{-6} = 0.6501$

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 0.04 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 the same price of the old chips. Assume that it costs $2000 to fabricate a wafer in either the old and new fabrication facility. You were previously selling the old Power5 chips for 40% more than their cost.

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

Dies per wafer (old Power5) $\cong (\pi \times 15^2 / 3.89) - (\pi \times 30 / \text{sqrt}(2 \times 3.89)) \cong 148$

Estimated number of good dies (old Power 5) = 148 x 0.5157 = 76

Cost per die (old Power5) = $2000 / 76 = $26.32

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

Dies per wafer (new Power5) $\cong (\pi \times 15^2 / 1.86) - (\pi \times 30 / \text{sqrt}(2 \times 1.86)) \cong 331$

Estimated number of good dies (new Power 5) = 331 x 0.6501 = 215

Cost per die (new Power5) = $2000 / 215 = $9.30

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

Selling Price (old Power5) = $26.32 x 1.4 = $36.85

Profit (new Power5) = $36.85 – $9.30 = $27.55

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?

Rate of Sale (new Power5) = 3 x 500,000 = 1,500,000 per month

Profit = 1,500,000 x $27.55 = $41,325,000 per month

$1,000,000,000 / $41,325,000 per month = 24.2 months to regain the cost of the new fab.

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?

$Power_{new} = (0.4 + 0.6 \times 0.2)\ Power_{old} = 0.52\ Power_{old}$

This reduces power to 52% of the original power

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

$Power_{new} = (0.4 + 0.3 \times 0.2)\ Power_{old} = 0.46\ Power_{old}$

This reduces power to 46% of the original power

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

$Power \propto Capacitive\ Load \times Voltage^2 \times Frequency$

$Power_{new} = 0.8^2 \times 0.5\ Power_{old} = 0.32\ Power_{old}$

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?

MTTF (computer) = 10^9 / 20,000 = 50,000 hours

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

Availability = MTTF / (MTTF + MTTR) = 50,000 / (50,000 + 24) = 0.9995 = 99.95% of the time

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?

Failure Rate (cluster) = 100 x Failure Rate (computer) = 100 / 50,000

MTTF (cluster) = 50,000 / 100 = 500 hours = 20.83 days

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?

Failure Rate (each computer) = 1/2000 days

Failure Rate 1 (out of 1000 computers) = 1000/2000 days (any computer can fail)

Failure Rate 2 (out of 999 computers) = 999/2000 days (any computer out of 999 can fail)

Failure Rate 10 (out of 991 computers) = 991/2000 days (any computer out of 991 can fail)

Failure Rate (System) = (Failure Rate 1) x (Failure Rate 2) x … x (Failure Rate 10)

Failure Rate (System) = 1000 x 999 x … x 991 / 2000^{10}

MTTF (System) = 1/Failure Rate = 2000^{10} / (1000 x 999 x … x 991) = 1071 days