CIS 655
HOMEWORK 1: Introduction and Performance Analysis

Name: Himani Patel

SU ID: 495936092

Email ID: htpatel@syr.edu

Research and Reading Assignment

A) Read and briefly summarize the two papers labeled with R1. Comment on which of the cited features listed for IBM 360 are still used in today's instruction sets? List some of the arguments made for and against RISC and explain if these arguments still hold true today.

Ans. In the paper titled 'Architecture of the IBM System/360', a detailed discussion is carried out about the designing objectives and the primary features of the IBM 360's architecture. Also, it significantly mentions the complications faced because of the compatibility exigency for different processors and the differing demands of real-time and logical information processing in the field of science, commerce, etc.

The cited features listed for IBM 360 that are still used in today's instruction sets are:

- Open-ended design: A reliable foundation is provided by the modern processor design for the future customer planning and customer programming, ongoing laboratory developments in the field of technology, system configuration and application & programming techniques.
- General-purpose function: In the processing of real-time, commerce and science related data, communications and language related data, logical data and data reduction, the system configurations for small, large, separate and mixed applications ought to be provided by the machine design individually. Specifically, its objective says that the following things must be provided:
    - Logical power of great generality
    - Code-independent operations
    - Manipulating individual bit separately

    - Referring to small units of bits by the general addressing system
- Efficient Performance: The configurations of all the individual systems or design would have to be able to meet the supremacy of the systems that are already specialized in functioning or performing or both.
- Intermodel Compatibility: Program compatibility have to be strictly followed at the program bit level, whether it is upward or downward, by all the models or system designs.

In 'The Case for the Reduced Instruction Set Computer' paper, the authors propose that the trend of the most common architectural change is not cost effective as always and it may be more harmful than being better. It also argues about the effectiveness of the VLSI computers implemented as RISC's which may be more than that of the CISC's.

The arguments made for RISC:

- Implementation Feasibility: A complex architecture like that of RISC can be less realized in a given technology than the less complex architecture of CISC.
- Design Time: RISC takes much less time to design and debug than CISC and thus it can likely use a superior technology.
- Speed: As RISC has a simpler design, it potentially gains in speed.
- Better use of chip area: The chip area used by RISC is less than that of CISC and the gained area can be used to make RISC more attractive.

The arguments made against RISC:

- It is believed that architecture simplification can be a backward step from supporting the high-level languages.
- Also, for all designing and debugging including system interactions, RISC does not support human readable high-level languages.

Due to the increase in improvements and research, many of these arguments do not hold true today. Also, the drawbacks of CISC are now well treated to beat the RISC.

B) Visit the Intel on-line microprocessor museum, and determine the rate of increase in transistor counts and clock frequencies in the 70's, 80's, 90's, 00's, and this decade. Also, create a plot of the number of transistors versus technology feature size using an MS Excel spreadsheet.

Ans. The transistor counts and the clock frequencies in the 70's, 80's, 90's, 00's and for this decade are listed below in the table:

Year	Name of the Processor	Transistor Counts	Clock Frequencies	Feature Size(micron)
1971	4004	2300	108000	10
1972	8008	3500	800000	10
1974	8080	4500	2000000	6
1978	8086	29000	5000000	3
1982	Intel 286 (80286)	134000	6000000	1.5
1985	Intel 386	275000	16000000	1.5
1989	Intel 486	12000000	25000000	1
1993	Pentium	31000000	66000000	0.8
1995	Pentium Pro	55000000	200000000	0.35
1997	Pentium 2	75000000	300000000	0.25
1998	Celeron	75000000	266000000	0.25
1999	Pentium 3	95000000	600000000	0.25
2000	Pentium 4	420000000	1500000000	0.18
2001	Itanium	410000000	2660000000	0.18
2001	Xeon	420000000	1700000000	0.18
2003	Pentium M	550000000	1700000000	0.09
2006	Core 2 Duo	291000000	2660000000	0.065
2007	Core 2 Quad	410000000	2400000000	0.045
2008	Atom	470000000	1860000000	0.045
2008	Core i7	731000000	4500000000	0.045
2010	2nd Generation	1160000000	3800000000	0.032
2012	3rd Generation	1400000000	2900000000	0.022
2012	Xeon Phi	7200000000	1700000000	0.022

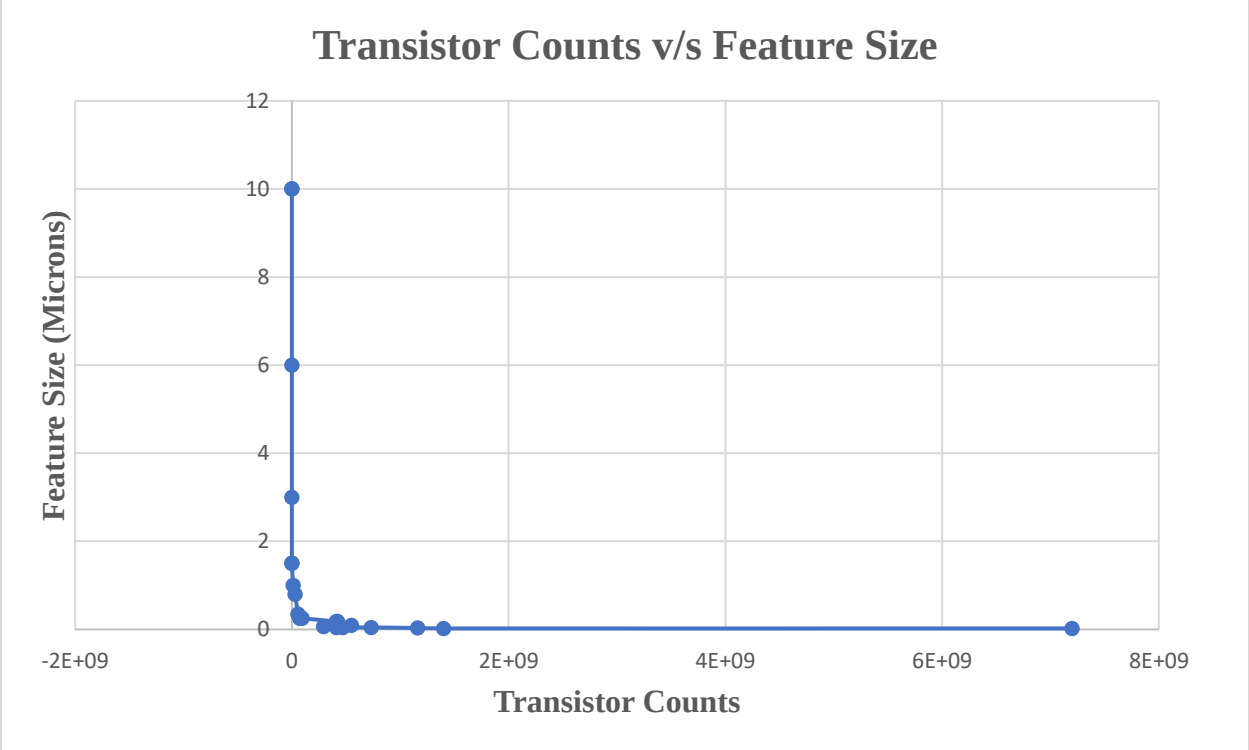
Rate of Increase in Transistor Counts:

- In 70's = 11.60
- In 80's = 88.55
- In 90's = 12.54
- In 00's = 1.76

Rate of Increase in Clock Frequencies:

- In 70's = 49.2
- In 80's = 3.166
- In 90's = 21.72
- In 00's = 1.23

Plot of the number of transistors versus the technology feature size:



Exercises

1. **Table below shows relevant chip statistics that influence the cost of several processors. Explore the effect of different possible design decisions for the Processor A and answer the below questions.**

Chip	Die Size	Estimated Defect Rate (per cm^2)	Manufacturing Size (nm)	Transistors (millions)
Processor A	400	0.30	130	276
Processor B	380	0.75	70	279
Processor C	199	0.75	70	233

   a) What is the die yield for Processor A? (Assume wafer yield is 100%, process-complexity factor is 5 for 130 nm technology)
   b) What might be the reasons that Processor A has a lower defect rate than the others?

   Ans.

   a) For Processor A, it is given that:

   Die Size = 400 mm^2 = 4.00 cm^2

   Defect Rate = 0.30 per cm^2

   Now, Die Yield = Wafer Yield * $\left(\frac{1}{1 + \text{Defects per unit area} * \text{Die Area}}\right)^N$, where N is the process complexity factor for Processor A.

   We are assuming that the Wafer Yield is 100%, and the process complexity factor is 5 for 130 nm manufacturing size.

   Therefore, N = 5 in the above formula.

   Thus, Die Yield for Processor A = $1 * \left(\frac{1}{1 + 0.3 * 4.00}\right)^5$

   Thus, Die Yield for Processor A = $1 * \left(\frac{1}{2.2}\right)^5$

   Thus, Die Yield for Processor A = 0.0194

   b) The reason for Processor A having a lower defect rate than the others is that the Manufacturing size of the Processor A is larger than that of the other processors. We know that the process complexity reduces as the size is increased which reduces the defect rate. Thus, this reason is true for Processor A having lower defect rate than the other processors.

2. **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. Sometimes, this is difficult to do. According to the trend in device scaling observed by Moore's law, the number of transistors on a chip in 2025 should be how many times the number in 2015?**

   Ans. The Moore's law states that the growth rate in the number of transistors on a chip is around 40% - 55% every year or it doubles every 18 months or 2 years.

   Let's assume that the growth rate in the number of transistors on a chip is 40%, then considering the transistors count in 2015 to be 1, the transistors count in 2016 will be 1 + 40% of 1, which is equal to 1 + 0.4 = 1.4.

   Thus, the number of transistors in 2025, which is 9 years from 2016, will be $(1.4)^{10} = 28.92$. It means that the number of transistors on a chip in 2025 year will be around 29 times the number of transistors on a chip in the year 2015.

   Now, assuming the growth rate in transistors count on a chip to be 55% and the number of transistors on a chip in 2015 to be 1, we get the number of transistors in 2016 as 1 + 55% of 1, that is 1 + 0.55 = 1.55.

   Therefore, in the year of 2025 i.e. 9 years from 2016, the number of transistors will be $(1.55)^{10} = 80.04$. Thus, the number of transistors on a chip in 2025 will be approximately 80 times the number of transistors on a chip in 2015.

3. **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. What is the speedup with N processors if 50% of the application is parallelizable, ignoring the cost of communication? What will be the speedup for a system with 1000 processors?**

   Ans. The Amdahl's law can be used to calculate the speedup of an application by improving some portions of it.

   The law states that:

$$\text{Speedup}_{\text{overall}} = \frac{1}{(1 - \text{Fraction}_{\text{enhanced}}) + \frac{\text{Fraction}_{\text{enhanced}}}{\text{Speedup}_{\text{enhanced}}}}$$

   Now, here $\text{Fraction}_{\text{enhanced}} = 50\% = 0.5$ and $\text{Speedup}_{\text{enhanced}} = N = 1000$

$$\text{Speedup}_{\text{overall}} = \frac{1}{(1 - 0.5) + \frac{0.5}{1000}}$$

$$\text{Speedup}_{\text{overall}} = \frac{1}{(0.5) + 0.0005} = \frac{1}{0.5005}$$

$\text{Speedup}_{\text{overall}} = 1.998$

Thus, the overall speedup for a system with 1000 processors is 1.998.

4. **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. Observe the effect of different design decisions on the necessary cooling, and thus the price, of a system. A cooling door for a rack costs $4,000 and dissipates 14 KW (into the room; additional cost is required to get it out of the room). How many servers with a Processor P2, 1 GB 240-pin DRAM, and a single 7,200 rpm hard drive can you cool with one cooling door? Use the table below for your power calculations.**

Component Type	Product	Performance	Power
Processor	P1	1.2 GHz	72-79 W peak
	P2	2 GHz	45-60 W
DRAM	MEM1	184-pin	3.7 W
	MEM2	240-pin	2.3 W
Hard Disk Drive	HDD1	5400 rpm	7.9 W read/seek, 2.9 W idle
	HDD2	7200 rpm	7.9 W read/seek, 4.0 W idle

Ans. From the given table, we know that:

A cooling door dissipates: 14KW of power = 14000W of power

Processor P2 consumes: maximum 60W of power

1 GB 240-pin DRAM consumes: 2.3W of power

7200 rpm Hard Disk Drive consumes: 7.9W of power

Thus, total power consumed by a single server is = Power consumed by Processor P2 + Power consumed by 1 GB 240-pin DRAM + Power consumed by 7200 rpm Hard Disk Drive

Therefore, total power consumed by a single server is = 60W + 2.3W + 7.9W = 70.2W

Number of servers that can be cooled = $\frac{\text{Power dissipated by a cooling door}}{\text{Total power consumed by a single server}}$

Number of servers that can be cooled with one cooling door = $\frac{14000}{70.2}$ = 199.43.

Thus, the number of servers with a Processor P2, 1 GB 240-pin DRAM and a single 7200 rpm Hard Disk Drive, that we can cool with one cooling door are approximately 199.

Case Studies

A) You have the following characteristics, as shown in the table below, on your company's processor for a certain benchmark, which runs at 400 MHz:

Instruction Type	Frequency (%)	Cycles
Arithmetic and Logical	30	1
Load and Store	20	2
Branches	40	3
Floating Point (FP)	10	5

You are asked to consider a cheaper, lower-performance version of this processor, by removing some of the FP hardware to reduce the die size. The wafer has a diameter of 10 cm, costs $1,000, and has a defect rate of 2/(cm^2). This wafer has a 75% yield. The current chip has a die size of 12 mm^2. The new chip becomes 10 mm^2, and FP instructions will now take 13 cycles to execute.

a) What are the old and new CPI (Cycles Per Instructions) and MIPS (Million Instructions Per Second) ratings running this benchmark?

b) What are the old and new die yields? What are the old and new costs per (working) processor? Please comment on the overall effect of the proposed hardware change on the cost and the performance of the processor. (Assume process-complexity factor is 4)

c) What would be the theoretical limit of the best possible overall speedup that we could ever get by only improving the FP unit, and what would be the CPI and MIPS ratings of this new processor?

Ans.

a) To calculate Cycles Per Instruction (CPI), we have:

$$\text{CPI} = \frac{\text{CPU clock cycles for a program}}{\text{Instruction Count}}$$

And, CPU clock cycles = $\sum_{i=1}^{n} IC_i * CPI_i$,
Where n = number of instruction cycles,
IC_i = the number of times i^{th} instruction is executed in the program,
CPI_i = clocks per instruction for the i^{th} instruction.

Thus, CPI = $i=1nICiInstruction\ Count * CPIi$

Now, old CPI = $\frac{30}{100} * 1 + \frac{20}{100} * 2 + \frac{40}{100} * 3 + \frac{10}{100} * 5 = 0.3 + 0.4 + 1.2 + 0.5 = 2.4$ cycles

It is given in the question that the company's processor runs at 400 MHz, thus clock rate = 400 MHz

Also, Instruction Execution Rate = $\frac{\text{Clock Rate}}{\text{Old CPI}} = \frac{400\ \text{MHz}}{2.4\ \text{cycles}} = 166.667$ MIPS

For new CPI, the FP instructions will now take 13 cycles to execute.
Thus,
New CPI = $\frac{30}{100} * 1 + \frac{20}{100} * 2 + \frac{40}{100} * 3 + \frac{10}{100} * 13 = 0.3 + 0.4 + 1.2 + 1.3 = 3.2$ cycles

And, Instruction Execution Rate = $\frac{\text{Clock Rate}}{\text{New CPI}} = \frac{400\ \text{MHz}}{3.2\ \text{cycles}} = 125$ MIPS

b) The formula for Die Yield is:

Die Yield = Wafer Yield * $\left(\frac{1}{1 + \text{Defects per unit area} * \text{Die Area}}\right)^N$, where N is the process complexity factor for Processor A.
Here, Wafer yield is given as 75% i.e. 0.75.
Process complexity factor = N = 4
Defects per unit area = 2 per cm^2
Die Area for old chip = 12 mm^2 = 0.12 cm^2
Die Area for new chip = 10 mm^2 = 0.1 cm^2

Old Die Yield = $0.75 * \left(\frac{1}{1 + 2 * 0.12}\right)^4 = 0.75 * \left(\frac{1}{1.24}\right)^4 = 0.75 * 0.4229 = 0.317$

New Die Yield = $0.75 * \left(\frac{1}{1 + 2 * 0.1}\right)^4 = 0.75 * \left(\frac{1}{1.2}\right)^4 = 0.75 * 0.4822 = 0.361$

Cost per processor = $\frac{\text{Cost of wafer}}{\text{Dies per wafer} * \text{Die Yield}}$

Dies per wafer = $\frac{\pi * \left(\frac{\text{Wafer Diameter}}{2}\right)^2}{\text{Die Area}} - \frac{\pi * \text{Wafer Diameter}}{\sqrt{2 * \text{Die Area}}}$

Dies per wafer for old chip = $\frac{\pi * \left(\frac{10}{2}\right)^2}{0.12} - \frac{\pi * 10}{\sqrt{2 * 0.12}} = 654.166 - 64.212 = 589.9$

Dies per wafer for new chip = $\frac{\pi * \left(\frac{10}{2}\right)^2}{0.1} - \frac{\pi * 10}{\sqrt{2 * 0.1}} = 785 - 70.246 = 714.75 = 715$

Old Cost per processor = $\frac{1000}{589.9 * 0.317} = \$5.347$

New Cost per processor = $\frac{1000}{714.75 * 0.361} = \$3.875$

Thus, it indicates that the proposed hardware change will reduce the cost to half of the old cost.

c) The theoretical limit for the best possible overall speedup that we can get is to take just one cycle for the execution.

$\text{Speedup}_{\text{enhanced}} = \frac{\text{Old clock cycle of FP instruction}}{\text{New clock cycle of FP}} = \frac{5}{1} = 5$

$\text{Fraction}_{\text{enhanced}} = \frac{\text{Clock cycle of FP instruction}}{\text{Total of all clock cycles of all operations}} = \frac{5}{1 + 2 + 3 + 5} = \frac{5}{11} = 0.4545$

According to Amdahl's Law:

$\text{Speedup}_{\text{overall}} = \frac{1}{(1 - \text{Fraction}_{\text{enhanced}}) + \frac{\text{Fraction}_{\text{enhanced}}}{\text{Speedup}_{\text{enhanced}}}}$

Now, the Speedup for 5 cycles of FP operation:

$$\text{Speedup}_{\text{overall}} = \frac{1}{(1 - 0.4545) + \frac{0.4545}{5}} = 1.57$$

The overall speedup for 13 cycles of FP operation:

$$\text{Speedup}_{\text{enhanced}} = 5$$

$$\text{Fraction}_{\text{enhanced}} = \frac{\text{Clock cycle of FP instruction}}{\text{Total of all clock cycles of all operations}} = \frac{13}{1 + 2 + 3 + 13} = \frac{13}{19} = 0.684$$

$$\text{Speedup}_{\text{overall}} = \frac{1}{(1 - 0.684) + \frac{0.684}{5}} = 2.208$$

$$\text{New CPI} = \frac{30}{100} * 1 + \frac{20}{100} * 2 + \frac{40}{100} * 3 + \frac{10}{100} * 1 = 0.3 + 0.4 + 1.2 + 0.1 = 2 \text{ cycles}$$

$$\text{New Instruction Execution Rate} = \frac{\text{Clock Rate}}{\text{New CPI}} = \frac{400 \text{ MHz}}{2 \text{ cycles}} = 200 \text{ MIPS}$$

B) Your company produces a mobile device. To extend the battery life in the newer version of the device, you are asked to elaborate on the idea to simply reduce the processor clock speed by 20%, and make no other changes. Stating your assumptions, describe whether this is a good idea or a bad idea, and why. Make sure to address both power and energy.

Ans. The entities power and energy are directly proportional to the Capacitive Load and Voltage. The equations for both are stated as:

$\text{Energy}_{\text{dynamic}} \alpha$ Capacitive Load * Voltage^2

$\text{Power}_{\text{dynamic}} \alpha$ Capacitive Load * Voltage^2 * Frequency Switch

If we reduce the processor clock speed by 20%, it reduces the program execution rate. This will consume high power as it increases the time to execute an instruction. Higher power consumption leads to increase in the running time of a device. Thus, it is not a good idea. Additionally, energy consumption becomes more due to the extended battery life which affects the efficiency and overall performance of a device. Therefore, reducing the processor clock speed by 20% is not a good idea. It can be useful for other aspects like reducing heat in a device, etc.