Computer Architecture

Fundamentals of Quantitative design and analysis

Name: Eddy Kimani
Registration Number: SCT212-0596/2021

Introduction

- Since the late 1970s till the early 200s, computers have had a rapid advance in cost and performance, achieving an annual growth rate of 52%
- This corresponds with Moore’s law, which predicted the doubling of transistors in chips after every 2 years
- A new design approach known as RISC(Reduced Instruction Set Computer) architecture was also created, which used simpler instruction sets compared to its predecessors

Classes of Computers

1. Personal Mobile Devices -> These devices try to maximize battery life and performance over a small form factor. Examples include smart phones and tablets
2. Desktop Computing -> These devices are built for tasks that require much more processing power that what a typical personal mobile device can handle. They focus on balancing between price and performance in tasks like web browsing, writing documents and playing graphics intensive games
3. Servers -> These devices focus on handling high throughput, since they are meant to handle a large volume of data and requests 24/7
4. Clusters/Warehouse computers -> These are groups of interconnected servers that work together as a single system. These are the massive data centers running cloud services. They are designed for massive scalability, power efficiency, fault tolerance and storage capacity
5. Embedded computers -> These are the systems integrated in everyday machines from cars to appliances to industrial equipment. They are specific purpose computers needed to be highly reliable and energy efficient

Parallelism

- Due to limitations in power consumption, the industry started shifting to parallelism, where instead of focusing on the single processor performance of a single chip, multiple processors (cores) were introduced on one chip
- The classes of parallelism include:
  - Data level parallelism, where multiple data points are operating at once
  - Task level parallelism, where independent tasks are executed simultaneously
  - Request level parallelism, where many independent requests are handled
- Categories of parallel architecture include:
  - Single Instruction Single Data (SISD) -> A single processor executes a single instruction at a time on a single data point
  - Single Instruction Multiple Data (SIMD) -> A single instruction is applied to multiple data points simultaneously
  - Multiple Instruction Single Data (MISD) -> Multiple instructions are carried out a single data point
  - Multiple Instruction Multiple Data (MIMD) -> Multiple instructions are connected to multiple data points concurrently

Defining Computer Architecture

- Different computer architectures have evolved to meet specific user needs and requirements
- The instruction set architecture is the actual programmer visible instruction set in the book that serves as the boundary between software and hardware
- The most common types of instruction set architectures include:
  - X86 -> This is the most influential architecture which has been implemented in most modern devices we use today. The architecture has evolved over the years from 16 bits to 64 bits
  - ARM -> This is a newer architecture designed with a focus on simplicity and power efficiency in mind

Trends in Technology

- If an instruction set architecture is to be successful, it must be designed to survive rapid changes in computer technology
- There have been 5 implementation technologies, which have changed at a dramatic pace:
  - Integrated circuit logic technology -> Transistor density has doubled every 18-24 months, according to Moore's law
  - Semiconductor DRAM -> Memory capacity has grown 25 – 40% per year
  - Semiconductor Flash -> This is the fast non-volatile memory used in SSDs and mobile devices that grows at 50 – 60% per year
  - Magnetic Disk Technology -> Traditional hard drives have improved in storage capacity but lag in speed
  - Network Technology -> Network performance depend on the performance of switches and the transmission system
- Data transfer speed (bandwidth) improves much faster than response time (latency)

Trends in Power and Energy in Integrated Circuits

- Power consumption is still the main constraint in CPU design
- However, there are techniques that are being used to improve energy efficiency. They include:
  - Dynamic voltage and Frequence Scaling (DVFS) -> This adjusts power based on a given workload
  - Turning off idle components which saves energy when not in user
  - Efficient cooling techniques to prevent overheating in high-performance systems
  - Using multicore processors instead of increasing the clock speeds of a single processor

Trends in Cost

- Over the years computing power has become cheaper than ever, but development costs are rising
- Moore’s law is slowing down, making cost efficiency more important
- This has been caused by:
  - The complexity of manufacturing smaller transistors -> More transistors equals higher cost
  - Memory and storage prices -> DRAM and Flash semiconductors are becoming cheaper
  - Energy consumption -> Power efficiency reduces operational costs
  - Demand for specialized hardware