Big Data and Security

Jeff Borowitz, PhD

Lecturer

Sam Nunn School of International Affairs

Trends in Computer Hardware

Parts of a Computer

- CPU - executes instructions
  - E.g. 1 + 1 = 2
- Memory - store information for a little while during computation
  - Example: sorting a list of numbers
    - Go through the list, number by number
    - For each number, put it in it’s appropriate, sorted place in a new list
    - Repeat this process as long as you have items left in the old list
- Disk - store information for longer periods of time

Basic Computer Architecture

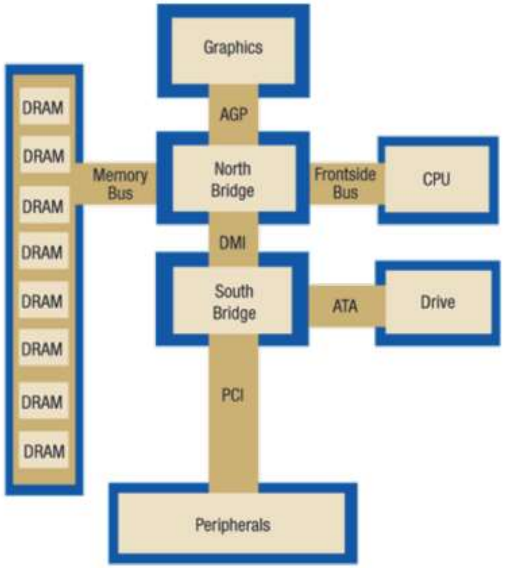
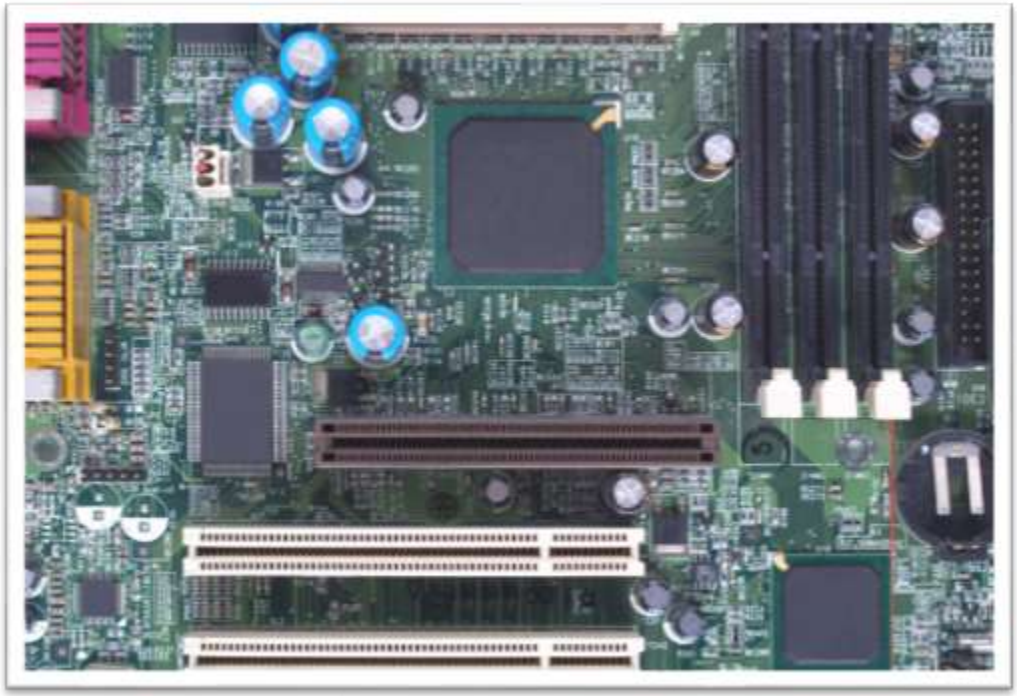


Figure 1 Major semiconductor content of standard motherboard.

A Motherboard



CPU

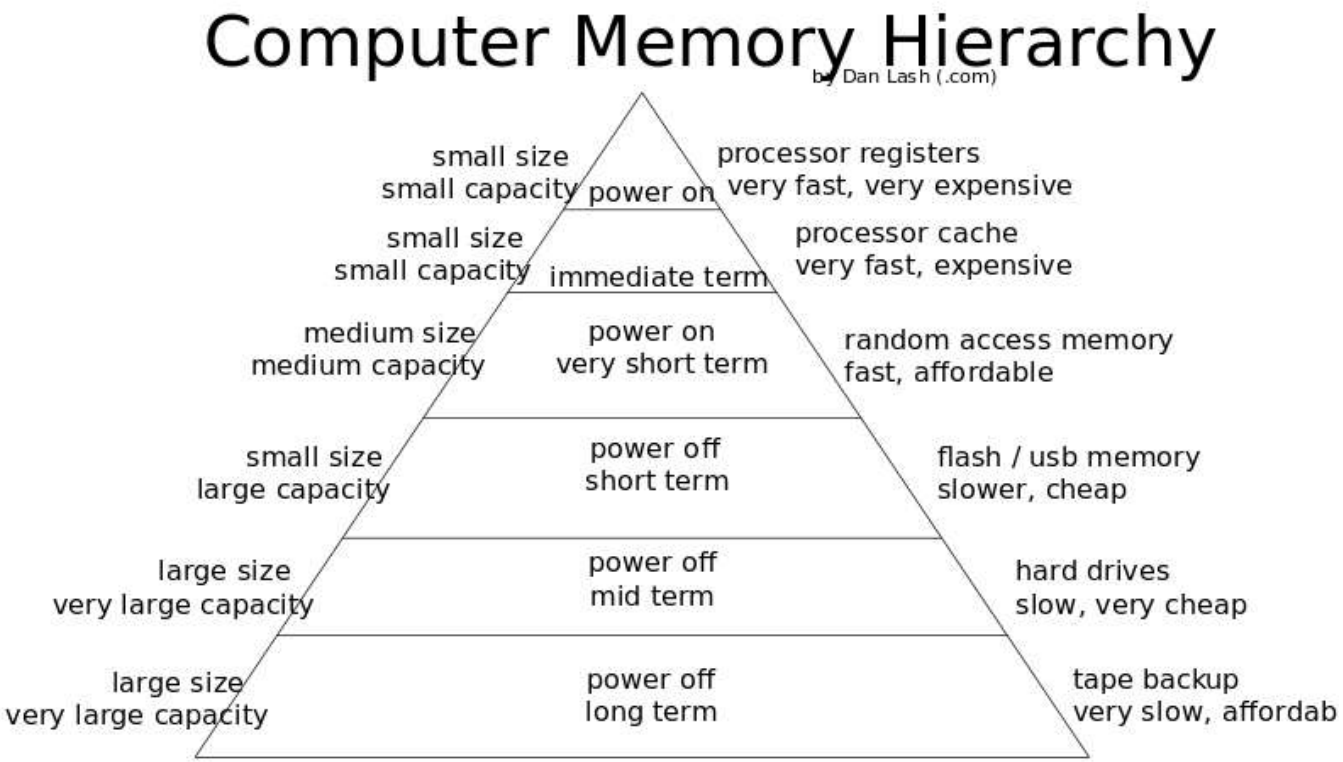
- CPU's do the following:
  - Read an instruction from memory and decode it
  - Find any associated data that is needed to process the instruction
  - Process the instruction
  - Write the results out
- Example: adding two digits
  - There are two digits, sent into the computer in binary form, across a set of pins. E.g. 12; 43
  - There is an in instruction sent to another set of pins. E.g. add
  - The result comes out in another set of pins E.g. 55, and is sent where it needs to go

CPU

- The way they implement these things are called microarchitectures
- Intel releases a new one of these every couple years (“Whiskey Lake”, “Xeon”)
- Key challenges: waiting for data to complete an instruction, conditional execution/branches
- Key solution: have multiple queues, guess at what branches will be followed

Storing Information

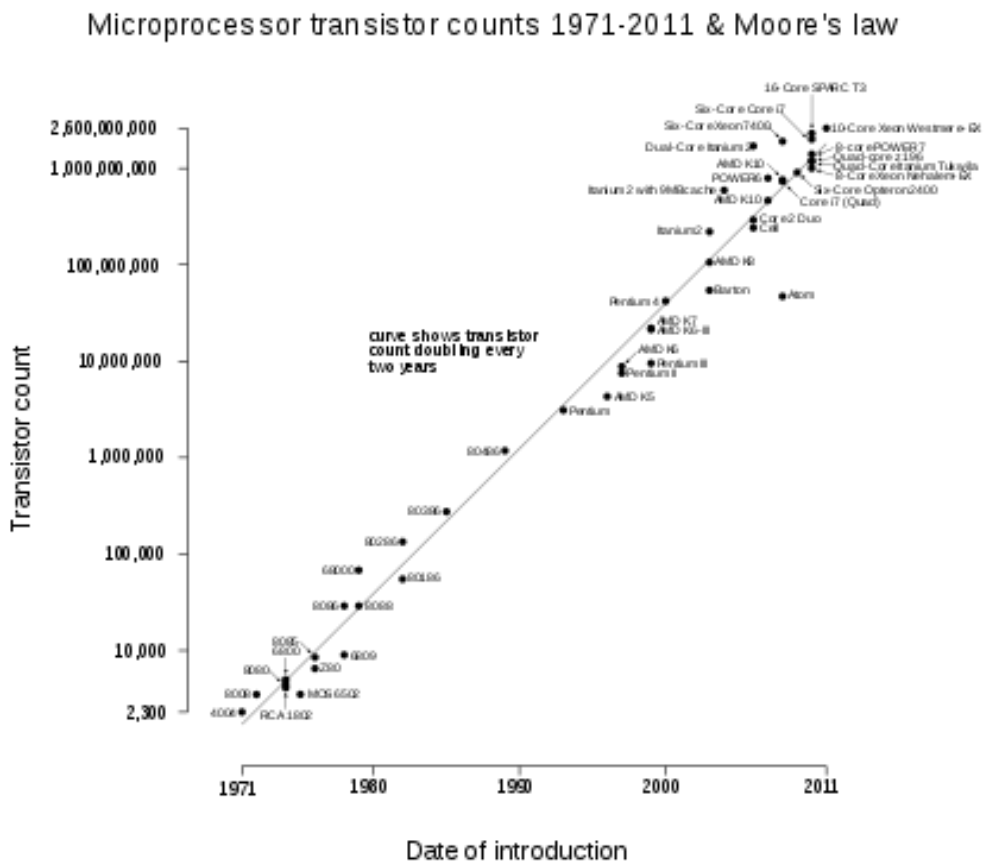
- To work quickly, computers need to access information quickly
- There are physical constraints on this: you can’t store all your data right next to your CPU!
- Levels of closeness, to look up a byte
  - L1 cache - 0.5ns
  - L2 cache - 7ns
  - Memory - 100ns
  - Disk - 10,000,000ns
  - *In memory elsewhere in a data center: 500,000 ns
  - *Solid state hard drives: 500,000 ns
- Note: A 1ghz CPU executes an instruction every nanosecond, including potentially looking up data



Source: ComputerMemoryHierarchy.png: User:Danlash at en.wikipedia.org

Moore’s Law and Other Technological Improvements

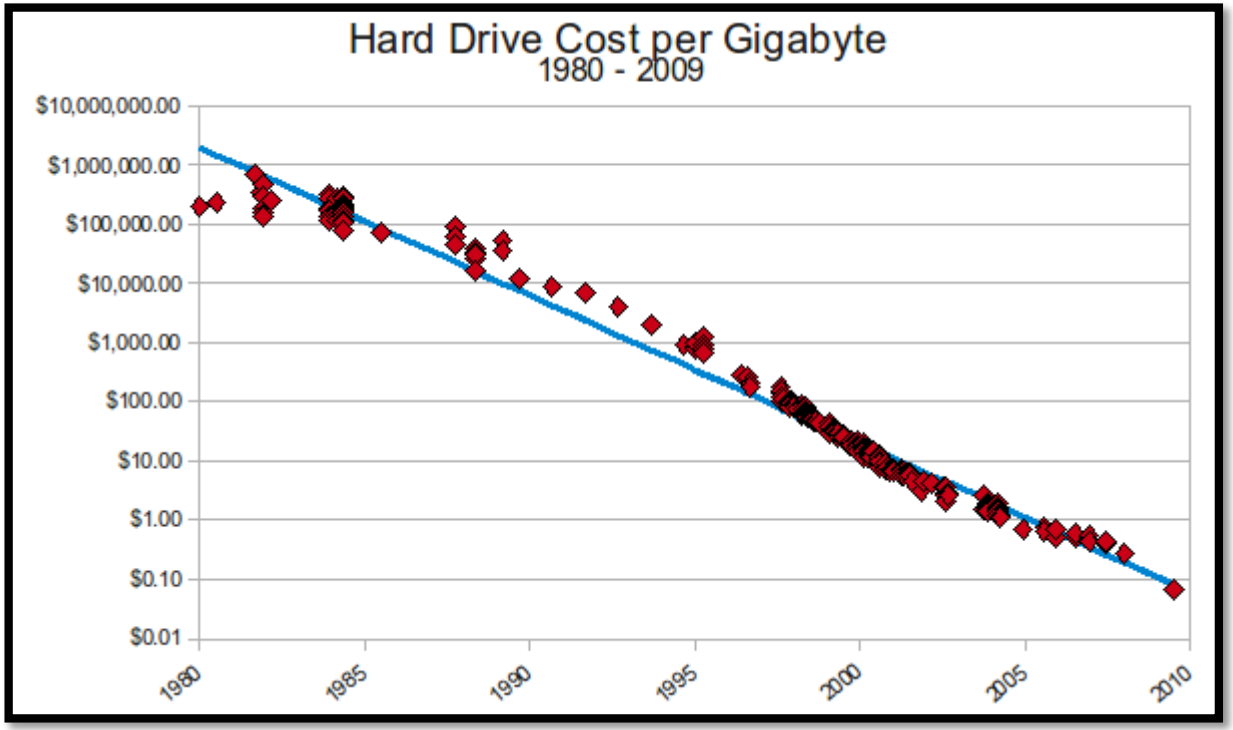
- The number of transistors per unit area can double every 18 months
- Kryder’s law: storage density on magnetic disks doubles every 18 months too



Source: User:Wgsimon at en.wikipedia.org

Georgia Tech

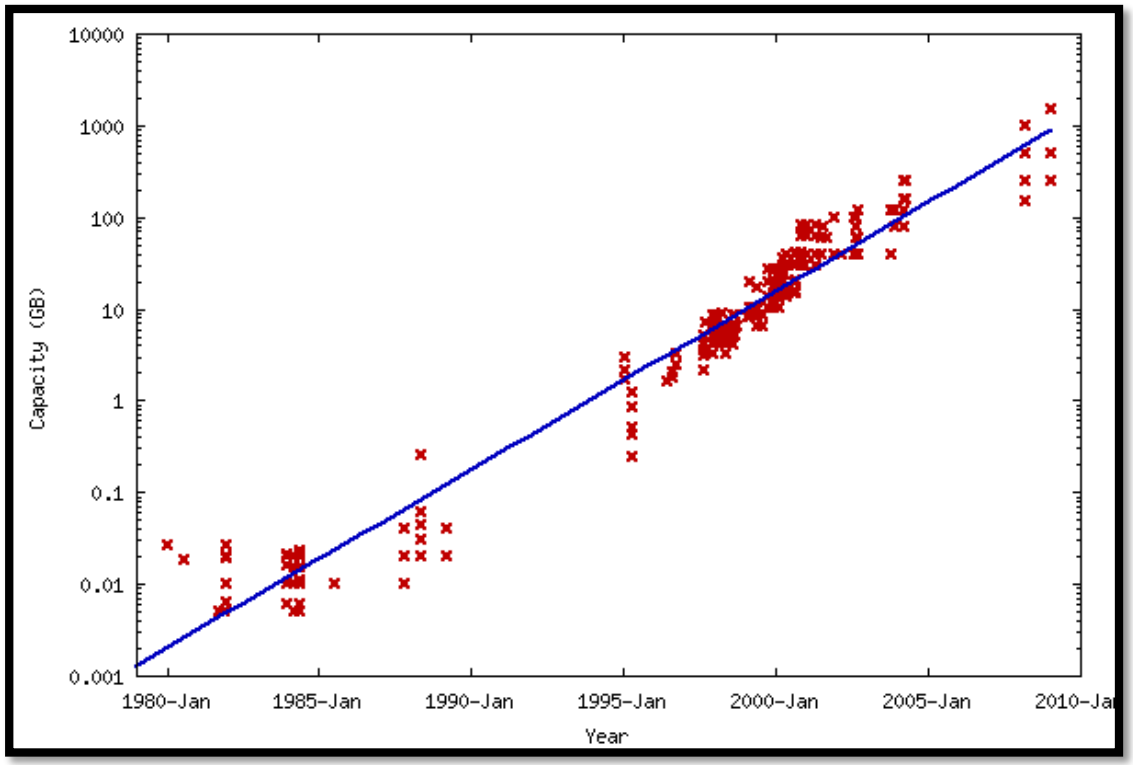
Hard Drive Prices Have Fallen Exponentially



Source: https://mkomo.com/cost-per-gigabyte

Georgia Tech

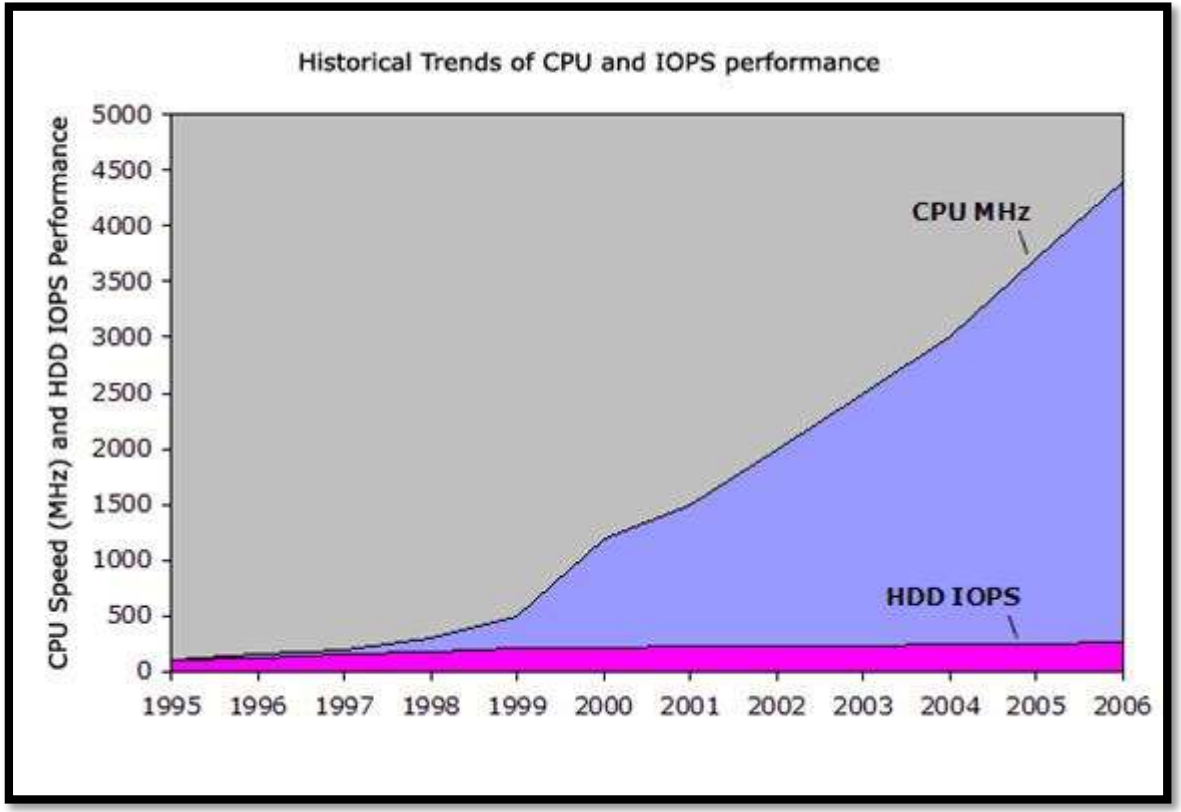
. . . While Capacity Has Grown Exponentially



Source: Created by User:Omegatron using the GIMP

Georgia Tech

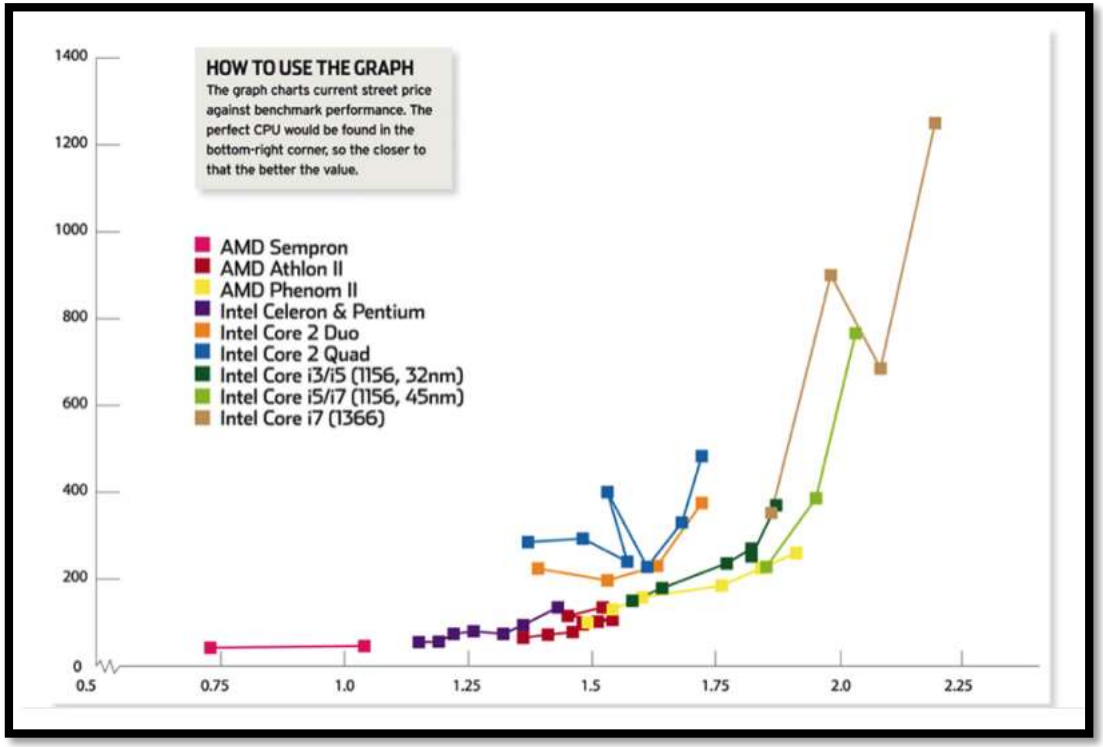
But Data Transfer Changes Slowly



Source: https://www.techbyter.com/2011/0828.html

Georgia Tech

Contemporary Performance is Exponentially More Costly



Georgia Tech

CPU Limits

- Moore's Law
- Power Dissipation
    - If you just shrunk transistors, power dissipation would decrease
    - But typically processors do more, and so power per area increases
- Single cores can't go faster without taking too much power
- Instead of more speed:
    - Multiple cores
    - More cache

Memory: Physical Limits

- Also depends on Moore's law-like progress
- Doesn't have acute overheating problems like CPUs: capacity just keeps growing

Hard Disk: Physical Limits

- Kryder's law suggests total storage will increase
- But platters can only spin so fast
- So accessing this large amount of data is likely to continue to be a big bottleneck
- One exception:
  - Sequential reads will get faster, since the head can read more data in a single rotation.
  - This can benefit performance of massive data operations which often need lots of sequential data

Summary

- Computers have 3 main components
  - CPU
  - Memory
  - Disk
- Major technological trends
  - CPU is growing cheaper but not faster
  - Memory/Disk space is growing
  - Disk access is not growing