Performance

COMP375 Computer Architecture and Organization

What is Good Performance

- Which is the best performing jet?

Airplane	Passengers	Range (mi)	Speed (mph)
Boeing 737-100	101	630	598
Boeing 747	470	4150	610
BAC/Sud Concorde	132	4000	1350
Douglas DC-8-50	146	8720	544

The Concorde is the fastest.

The Boeing 747 is the largest.

The DC-8 has the longest range.

Benchmark at Home and Office

	Simple program	
Home	1.00	
Office	3.41	

Home CPU is an Intel 2.0 GHz Pentium 4
Office CPU is an Intel 3.16 GHz Core 2 Duo Pentium 4

Benchmark at Home and Office

	Simple program	Program with 2 threads
Home	1.00	0.99
Office	3.41	5.86

Home CPU is an Intel 2.0 GHz Pentium 4
Office CPU is an Intel 3.16 GHz Core 2 Duo Pentium 4

Defining Performance

- **Response Time** – The wall clock time it takes for a computer to complete a task. This is the performance measure of most interest to desktop users.
- **Throughput** – The amount of work that can be accomplished, such as web pages served per second. Servers are often measured in this manner.

Many Applications

- Some programs do a lot of I/O. They run best on a computer with lots of RAM and fast disks.
- Some programs are CPU bound. The run best on a computer with a fast CPU and lots of cache.

Why does having lots of RAM improve disk I/O access?

1. Better RAM to cache ratio
2. More space for disk caching
3. Faster bus access
4. It doesn't

Performance Aspects

- Integer arithmetic
- Floating Point arithmetic
- Network throughput
- Disk access time
- Disk throughput
- Audio and video processing
- Power usage

Clock Rate Isn't Everything

- Different architectures require a different number of Cycles per Instruction (**CPI**).
- CISC processors typically use more CPI, but may accomplish more per instruction.
- RISC processors usually have lower CPI, but it may take more instructions to accomplish a task.

Execution Time

- N = number of instructions executed to run a program
- Hz = Clock rate
- Execution time = N * CPI / Hz
- This gives the execution time, not counting any I/O, system time or time spent on other applications.

What might cause the number of cycles per second to vary?

1. Pipeline stalls
2. Caching
3. Virtual memory
4. Interrupts
5. All of the above
6. None of the above

MIPS

- Millions of Instructions Per Second
- ***M****eaningless* ***I****ndicator of* ***P****erformance*
- MIPS = clock rate / (CPI * 1,000,000)

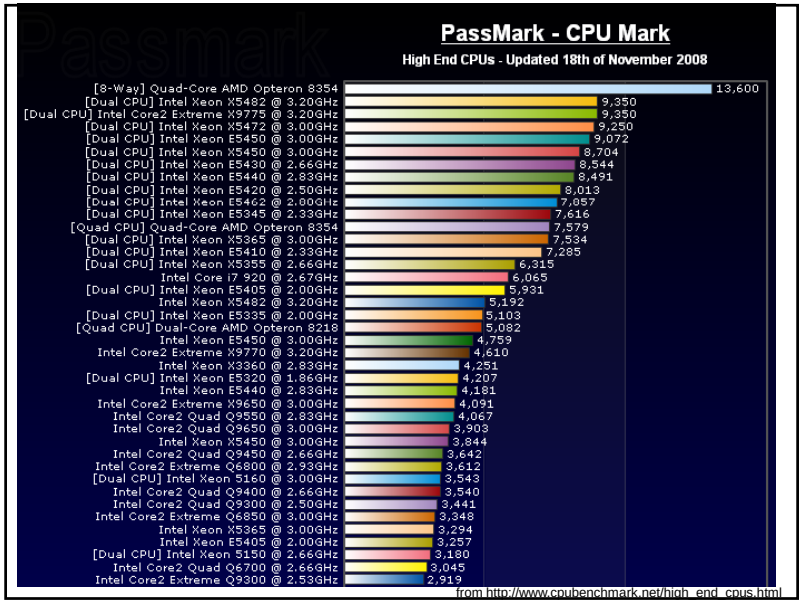
"The processor is guaranteed to not run faster than 25 MIPS." – honest salesman

FLOPS

- Floating Point Operations Per Second
- A common measure of high performance computer speed.
- Usually used with a metric such as TeraFLOPS or PetaFLOPS.

Fastest Computers

	Cores	Peak GFlops
Roadrunner	129,600	1,456,704
Jaguar - Cray XT5	150,152	1,381,400
Pleiades - SGI	51,200	608,829
BlueGene/L	212,992	596,378



Benchmarks

- A benchmark is a program run on different systems whose execution time is used as a measure of system performance.
- A benchmark tells you how long it will take to run THAT program with THAT data.
- Benchmark results may, or may not, correspond to the time it takes to run your application.

SPEC

- Standardized benchmarks from the System Performance Evaluation Corporation
- Several different versions
  - SPEC CINT2000
  - SPEC CFP2000
  - SPECweb2005
  - WinSPEC

Homogeneous Systems

Improvements for a given architecture:

- Increases in clock rate
- Improvements in CPU organization that lowers the CPI
- More cache and wider bus for faster memory access
- Compiler enhancements to reduce the number of instructions executed to perform an application.
- Adding custom instructions.

Amdahl's Law

- Improvements to part of a program can only improve that part of the program.
- Assume **P** is the fraction of the program that can run on **N** processors

$$speedup = \frac{1}{1 - P + \frac{P}{N}}$$

Speedup

- Speedup is the ratio of the speed of the parallel program compared to the speed of the sequential version.

$$Speedup = \frac{Original\ Execution\ Time}{Parallel\ Execution\ Time}$$

How to Improve Performance

- Buy more memory
- Buy faster disks
- If you are writing the application
  - use compiler optimization
  - read and write files in large blocks
  - use memory instead of files

Measure System Performance

- All operating systems provide tools to help measure system performance.
- Commonly measured values
  - CPU utilization
  - Number of I/Os to the disk / second
  - Number of page faults / second

SPARC Instruction Set

- The Sun SPARC processor has 32 registers.
- Register R0 is always zero and cannot be changed.
- Instructions are of the form

```
Add  R1, R2, R3
```

- Where R2 is added to R3 and the result stored in R1.

How can you implement **mov R_{dest}, R_{source}** ?

1. Add R_{dest}, R_{source}, R0
2. Add R0, R_{dest}, R_{source}
3. Sub R_{dest}, R0, R_{source}
4. Add R0, R0, R0
5. Cannot be done

3 Level Memory Speed

- Assume your have L1 and L2 cache and RAM with hit rates H_1, H_2 and H_3 *(Note: H1 + H2 + H3 = 1)*
- The average memory access time is:

$$avg = H_1*L1 + H_2*L2 + H_3*RAM$$