

# ECSE 425 – Tutorial 2

Performance and Benchmarks

# Performance

- Which of those cars has the best performance?



Credit: {Brian Snelson, IFCAR}, Wikipedia

# Performance

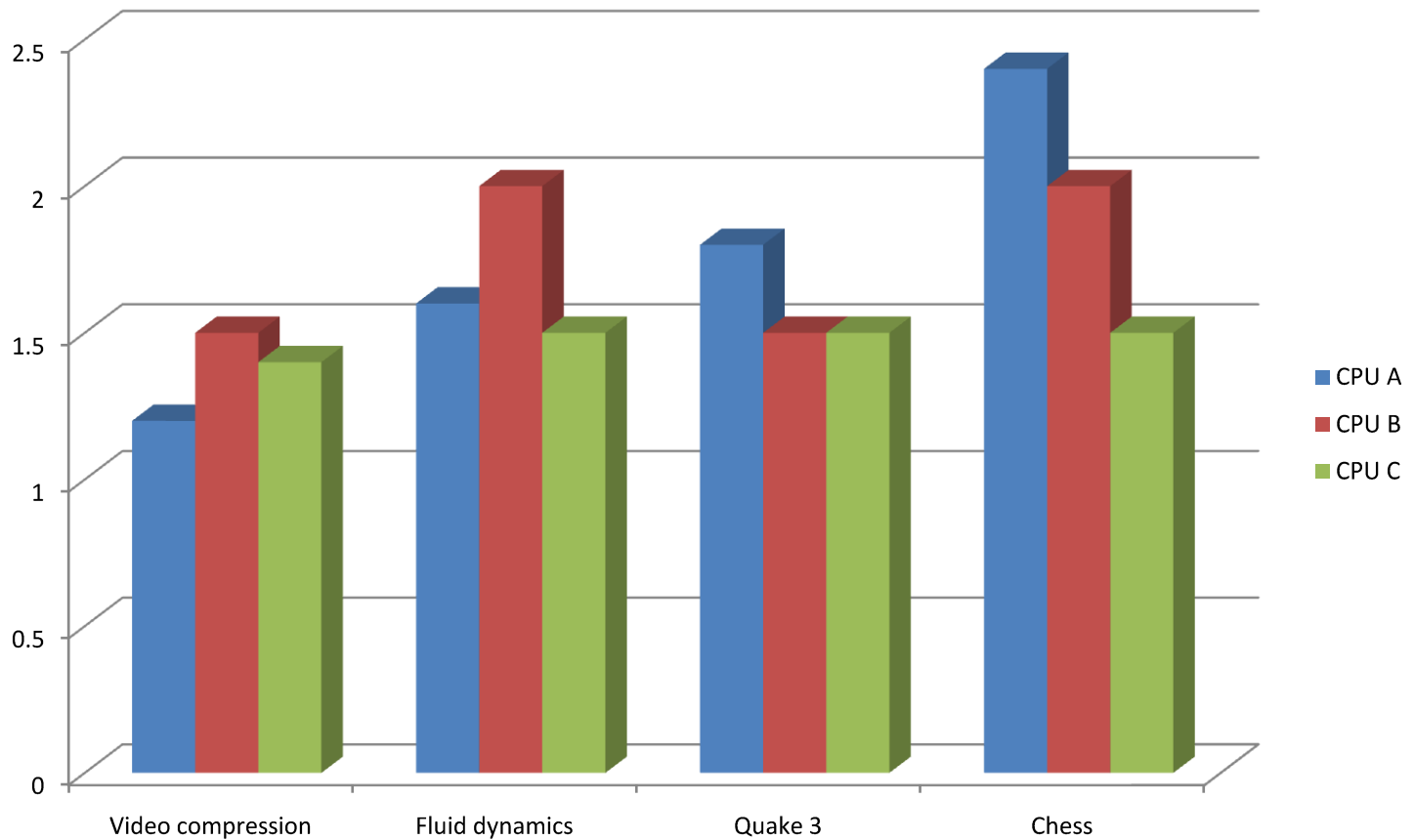
- How do you measure performance?
  - Throughput : number of operations/sec?
  - Latency : time before a result is obtained?
  - Execution time : wall-clock time taken to perform a task?
- Different types of users expect different things
  - When measuring performance, must take those different needs into account
    - Database? Web server? Rendering farm?

# Performance

- How do you measure performance?
  - **Benchmarks!**
- Good benchmark suites perform calculations using real-life programs, across a wide variety of domains
  - Video compression, games, ...
  - XML parsing, linear algebra, ...
  - Speech recognition, molecular dynamics, ...
  - ...

# Performance

- Sample benchmark results



# Performance

## Benchmark scores

- Often need to compute an average score
  - If the scores are expressed as the execution time, we use the arithmetic mean:

$$\mu = \frac{1}{n} \sum_{i=1}^n score_i$$

- If the scores are performance ratios, we use the geometric mean:

$$\mu = \sqrt[n]{\prod_{i=1}^n ratio_i}$$

# Performance

## Benchmark scores

- What is the mean score  $\mu$  of a benchmark suite where relative scores with respect to a reference machine are (ratio =  $t_{\text{new}}/t_{\text{ref}}$ ):
  - ratio<sub>1</sub>=4, ratio<sub>2</sub>=2, ratio<sub>3</sub>=8
- What is the mean execution time  $\mu$  of a benchmark suite where execution times are :
  - t<sub>1</sub>=4s, t<sub>2</sub>=2s, t<sub>3</sub>=8s

# Amdahl's Law

- How can you determine how much better a program performs given improvements in certain parts of this program?
  - This metric is known as the *speedup*
- Amdahl's law!
  - Speedup :  $n = \text{ExecutionTime}_{\text{old}} / \text{ExecutionTime}_{\text{new}}$



# Amdahl's Law

System with single enhancement:

$$\mathbf{time}_{new} = (1 - f_{enh})\mathbf{time}_{old} + \frac{f_{enh}\mathbf{time}_{old}}{s}$$

$$\mathbf{systemspeedup} = \frac{\mathbf{time}_{old}}{\mathbf{time}_{new}} = \frac{1}{(1 - f_{enh}) + \frac{f_{enh}}{s}}$$

System with multiple enhancements:

$$\mathbf{time}_{new} = \frac{f_1\mathbf{time}_{old}}{s_1} + \frac{f_2\mathbf{time}_{old}}{s_2}$$

$$\mathbf{systemspeedup} = \frac{\mathbf{time}_{old}}{\mathbf{time}_{new}} = \left( \frac{f_1}{s_1} + \frac{f_2}{s_2} \right)^{-1}$$

# Amdahl's Law

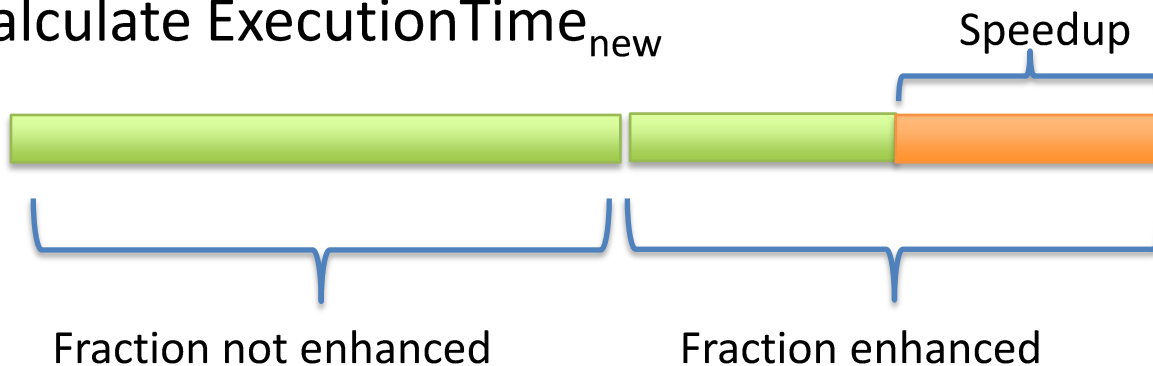
– Simpler method:

- $n = \text{ExecutionTime}_{\text{old}} / \text{ExecutionTime}_{\text{new}}$

- Assume  $\text{ExecutionTime}_{\text{old}} = 1$



- Calculate  $\text{ExecutionTime}_{\text{new}}$



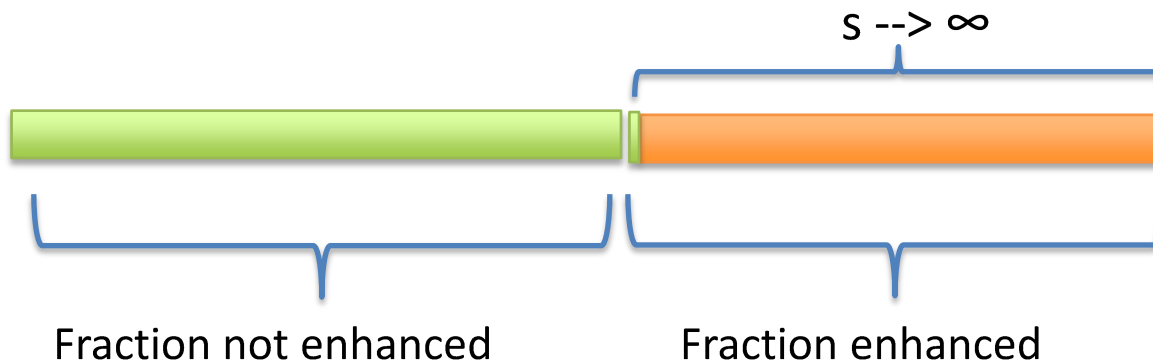
# Amdahl's Law

- Assume a processor spends 90% of its time performing integer arithmetic and 10% doing floating point calculations.
- If it were possible to speed-up integer calculations by 12% ( $s=1.136$ ) OR floating point calculations by 85% ( $s=6.666$ ), which one should you choose?

# Amdahl's Law

- What is the best speedup achievable for a given  $f_{enh}$ :

$$\lim_{s \rightarrow \infty} \left( \frac{1}{(1 - f_{enh}) + \frac{f_{enh}}{s}} \right) = \frac{1}{1 - f_{enh}}$$



# Amdahl's Law

- Assume a processor spends 90% of its time performing integer arithmetic and 10% doing floating point calculations.
- What is the best theoretical speed-up that you can obtain by improving the performance of the floating point unit infinitely?

# Processor Performance

- CPU time =  $IC * CPI * t_{cc}$
- IC (instruction count) = number of instructions in your program
- CPI (cycles per instruction) = number of clock cycles needed to perform an instruction
- CC (clock cycle time) = time taken by a single clock cycle
- Those metrics are tied to one another!

# Processor Performance

- Another way to view speedups:

- CPU time =  $IC * CPI * t_{cc}$

- $s = \text{Time}_{old} / \text{Time}_{new}$

- =  $IC_{old} * CPI_{old} * t_{cc\_old} / IC_{new} * CPI_{new} * t_{cc\_new}$

# Processor Performance

Assume a CPU with running a program:

–  $IC=20M$ ,  $CPI=2.6$ ,  $t_{cc}=1ns$

Suppose the following scenarios:

– Decrease  $t_{cc}$  to  $0.9ns$  and increase  $CPI$  to  $2.7$

– Increase  $t_{cc}$  to  $1.1ns$  and decrease  $CPI$  to  $2.1$

What is the best alternative?



# Processor Performance

Assume a CPU and workload with the following characteristics:

- 70% integer operations : 2 CC/instr. avg.
- 15% fp operations : 10 CC/instr. avg.
- 14.9% branches : 1 CC/instr. avg.
- 0.1% : SHA1 operation : 70 CC/instr

What is the average CPI for this CPU?

Is it better to improve  $CPI_{fp}$  to 9CC or  $CPI_{SHA1}$  to 20 CC?