







CS305: Computer Architecture

Empirical Evaluation-II

<https://www.cse.iitb.ac.in/~biswa/courses/CS305/main.html>

<https://www.cse.iitb.ac.in/~biswa/>

Evaluation

- To compare Processor A with Processor B by running programs
- How many programs? 
- The programs that you care.  
- What if I want to build a new one (processor, caches, DRAM) ? 

World of Benchmarks

- SPEC CPU 2017 (<https://www.spec.org/cpu2017/>)

The **SPEC CPU® 2017** benchmark package contains SPEC's **next-generation, industry-standardized, CPU intensive** suites for measuring and comparing **compute intensive performance, stressing a system's processor, memory subsystem and compiler.**

SPECspeed: used for comparing time for a computer to complete single tasks

SPECrate: measure the throughput or work per unit of time.

World of Benchmarks

CloudSuite (<https://www.cloudsuite.ch/>)

CloudSuite is a benchmark suite for **cloud services**. The benchmarks are based on real-world software stacks and represent real-world setups.

PARSEC (<https://parsec.cs.princeton.edu/>)

Benchmark suite composed of **multithreaded** programs. The suite focuses on emerging workloads and was designed to be representative of next-generation shared-memory programs for chip-multiprocessors.

World of Benchmarks

MobileBench

(<https://mobilebench.engineering.asu.edu/>)

comprising a selection of representative **smart phone applications**.

Many more application domain specific: Graph processing, ML perf,

Pitfalls of Benchmarks

Benchmark not representative of all

Your workload is I/O bound → SPEC CPU is useless

Benchmark is too old

Need to be periodically refreshed

Non-benchmarks

- Application kernels: A small code fragment or part of the program
- Synthetic benchmark : Not part of any real program!!
- Micro-benchmark
- *OK! So, I will create a chip and then evaluate these benchmarks*

World of Simulators

- Functional Simulator: Used to **verify the correct** execution of the program. Can not be used for performance evaluation.
- Performance simulators:
 - (i) Trace-driven: ChampSim
(<https://github.com/ChampSim/ChampSim>)
 - (ii) Execution-driven: gem5, Multi2sim

Functional simulator is part of the performance simulators.

Evaluation Continued

Pick a *relevant* benchmark suite

Measure IPC of each program

Summarize the performance using:

Arithmetic Mean (AM)

Geometric Mean (GM)

Harmonic Mean (HM)

Which one to choose?

Example

| | IMTEL | ABM | AND |
|------------|-------|-----|-----|
| App. one | 10 | 20 | 30 |
| App. two | 20 | 30 | 40 |
| App. three | 30 | 40 | 10 |

Which machine performs better over IMTEL and why?

Contd.

| | ABM | AND |
|------------|------|------|
| App. one | 2 | 3 |
| App. two | 1.5 | 2 |
| App. three | 1.3 | 0.3 |
| A.M. | 1.60 | 1.76 |
| G.M. | 1.57 | 1.21 |
| H.M. | 1.54 | 0.72 |

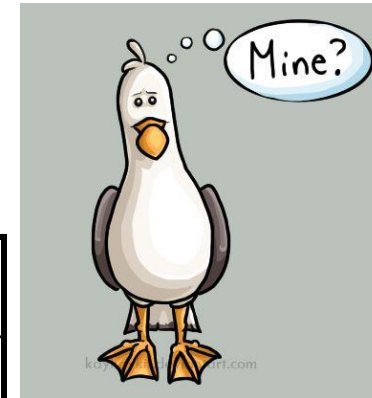


AM on ratios

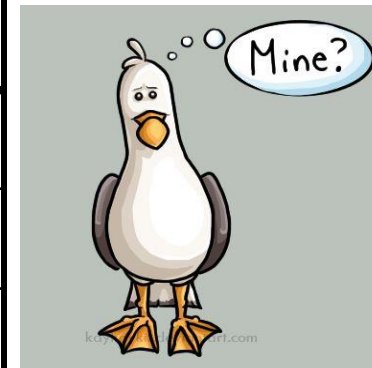
| | X | Y |
|--------|------|-----|
| App. 1 | 1 | 100 |
| App. 2 | 1000 | 10 |

| Normalized to X | X | Y |
|-----------------|---|--------|
| App. 1 | 1 | 100 |
| App. 2 | 1 | 0.01 |
| AM | 1 | 50.005 |

| Normalized to Y | X | Y |
|-----------------|--------|---|
| App. 1 | 0.01 | 1 |
| App. 2 | 100 | 1 |
| AM | 50.005 | 1 |



Y is 50 times faster than X



X is 50 times faster than Y

Nandi