SPEC CPU Benchmarks: Looking Back and Looking Forward: Soliciting Feedback on SPEC's Next CPU Suite

International SPEC Benchmark Workshop 2010
October 8, 2010
Jeff Reilly, (Chair, SPEC CPU Subcommittee)
Anirudha Rahatekar, (Technical Rep, SPEC CPU Subcommittee)

With thanks to Alex Carlton, Daryl Gove, John Henning, Cloyce Spradling, the SPEC CPU Subcommittee and all those who contributed to SPEC's efforts

Agenda

- Level Setting on Definitions
- Where is SPEC CPU today?
- Benchmark design in ~30 minutes

Feel free to ask questions or provide comments!

Why Level Set?

What has the power to make or break a company or a career?

What has the power to generate heated controversy, hard feelings, and bold accusations?

A sex scandal?

Litigation?

Nope—try *benchmarks.*

http://arstechnica.com/cpu/2q99/benchmarking-1.html

There can be a lot of subtleties with benchmarking and it's important to be sure everyone is on the same page.

What Are Benchmarks?

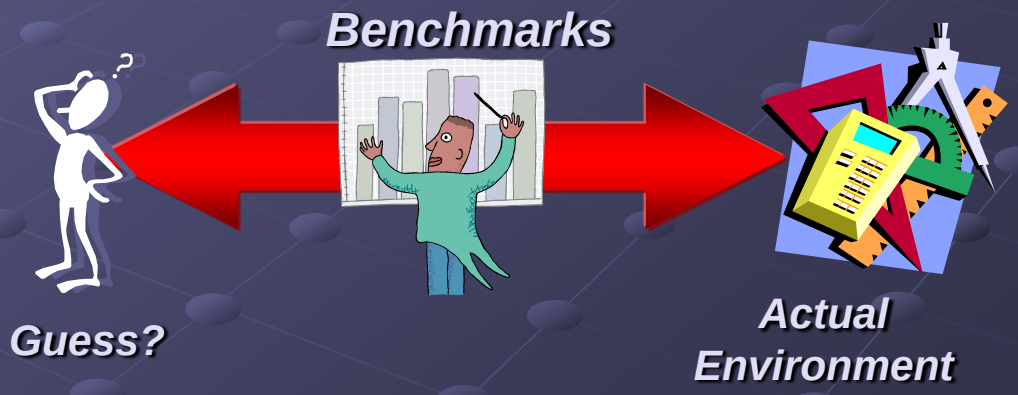
- A benchmark is a standard by which something can be measured or judged

- Examples Of Benchmarks:
  - Car Efficiency: miles per gallon
  - Sports Statistics: batting average
  - School: Grade point average

Benchmarks allow for evaluations or comparisons between two or more items

Why Use Benchmarks?

- Benchmarks lie between the extremes of “wild guess” and “actual environment”
- Benchmarks ideally, measure exactly want you want to evaluate but the following are issues...
  - Time
  - Money
  - Available data
  - Economy Of Scale



“Benchmarks provide successive approximations to reality” – J. Mashey
This requires understanding both the benchmark AND your needs!

A GOOD Benchmark Is...

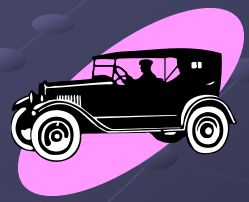
- Relevant
- Recognized
- Simple
- Portable
- Scaleable

Not all benchmarks (including some popular ones) have all of these characteristics! Always assess this...

Source: Jim Gray

What Makes A Good Benchmark Go Bad?

- Technology improvements
  - Hardware tends to evolve faster than software; scalability issues
- Introduction of unanticipated technology
  - Rule issues; test may no longer be meaningful.
- Misuse
  - Too many numbers, need for education
- Evolution of environment and usage models
  - Capture what is important to users today



An often overlooked fact is that benchmarks need to evolve with TIME… Or Yesterday's 'good' benchmark may NOT be today's 'good' benchmark.

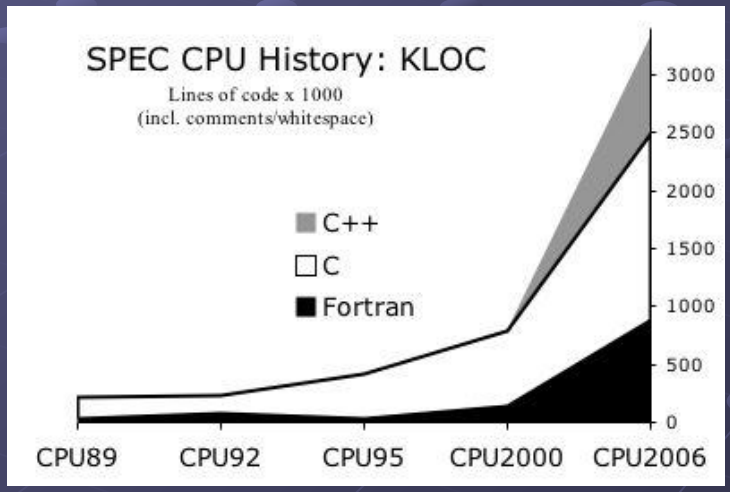
Background: SPEC CPU

- SPEC is an industry consortium (H/W, S/W, education, end-users) cooperating to develop benchmarks
  - CPU benchmarks are developed by the CPU Sub-committee of the Open Systems Group (OSG)
- Current members (as of October 2010) of the CPU Subcommittee:
  - AMD, Dell, Fujitsu, HP, IBM, Intel, Oracle, PGI
- Basic philosophy
  - SPEC: “establish, maintain and endorse a standardized set of relevant benchmarks that can be applied to the newest generation of high-performance computers.”
  - CPU Subcommittee: To develop CPU benchmarks that provide a comparative measure of CPU, compiler and memory performance with relevant, real-world applications across the widest range of platforms
- Decision making is meant to be by consensus; voting sets directions and establishes final release.

SPEC development is a team effort.

What Generally Drives SPEC CPU Evolution

- Run-time
  - Want meaningful workloads; want meaningful measurement interval; possible conflict with cost of benchmarking
- Application type
  - Want workloads that are meaningful in a performance context; want current versions
- Application size
  - Want workloads that are taxing for today's systems; enable demonstration of what is capable with coming systems
- Moving target
  - Provided as source code; provide new material for compilation



Source: J. Henning, SPEC

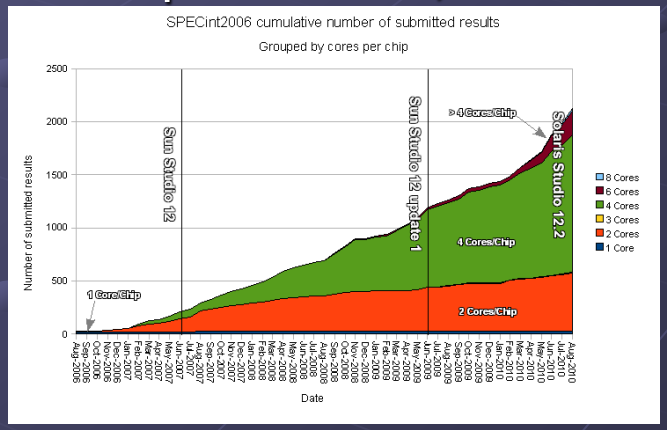
As computing platforms evolve, the criteria for judging them needs to evolve as well.

SPEC CPU Today: SPEC CPU2006

- Introduced in August 2006; Preceded by CPU89, CPU92, CPU95 and CPU2000; 8 metrics: integer/fp, speed/rate, baseline/peak

- Continued using two suites of benchmarks :
  - CINT2006: 12 benchmarks; 9 in C, 3 in C++
  - CFP2006: 17 benchmarks; 6 in FORTRAN, 4 in C, 3 in C++, 4 in a combination of C/FORTRAN

- 12411 results published at www.spec.org as of September 30, 2010:
  - 2236: CINT2006
  - 2191: CFP2006
  - 4256: CINT2006 Rate
  - 3728: CFP2006 Rate
  - 8 results per day for 4 years and 2 months



Source: http://www.darrylgove.com/

Meets the "good benchmark" definition of relevant, portable, recognized and scalable. Full details on http://www.spec.org

Technical Feedback on SPEC CPU2006?

- Run time is too long
- Too many workloads
- Too many metrics
- Workloads too small
- Workloads too big

Source: http://www.xkcd.com/303

In the remainder of the session, we will survey you on some of the questions the SPEC CPU subcommittee has been reviewing.

Audience Participation

Survey Question 1: Should SPEC change the technical development for the next SPEC CPU suite?

- Yes:

- No:

Audience Participation

Survey Question 2: Should SPEC CPU have :

- Separate integer and floating point metrics?

- No distinction between integer and floating point code?

- Other options?

Audience Participation

Survey Question 3: Should SPEC CPU :

- Separate baseline and peak metrics?

- Just baseline metrics?

- Just peak metrics?

- Other options?

Audience Participation

Survey Question 4: Should SPEC CPU have :

- Separate speed and rate metrics?

- Just speed metrics?

- Just rate metrics?

- Other options?

Audience Participation

Survey Question 5: If SPEC has speed and rate metrics, is it important for each invocation of the workload for rate to be the same as the invocation for speed?

- Yes:

- No:

Audience Participation

Survey Question 6: The memory footprint for a single copy of a SPEC CPU benchmark should be:

- 512MB or smaller

- 512MB-1GB

- 1GB-2GB

- 2GB to 4GB

- 4GB or more

Closing Comments

- When dealing across organizations, level set on definitions and expectations.
- Benchmarks do need to evolve with time.
- Effective development requires teamwork, defining expectations and compromise.

Thank you for your time!