SPEC CPU2017 – Next-Generation Compute Benchmark

James Bucek
Hewlett Packard Enterprise
james.bucek@hpe.com

Klaus-Dieter Lange
Hewlett Packard Enterprise
klaus.lange@hpe.com

Jóakim v. Kistowski
University of Würzburg
joakim.kistowski@uni-wuerzburg.de

ABSTRACT

Description of the new features of the SPEC CPU2017 industry standard benchmark and its metric calculations.

CCS CONCEPTS

• **General and reference** → **Performance** • General and reference → Metrics • Hardware → Power and energy • Software and its engineering → Compilers

KEYWORDS

Performance, SPEC, CPU, Compiler, Power, Energy, Rate, Speed

ACM Reference format:

James Bucek, Klaus-Dieter Lange, Jóakim v. Kistowski. 2018. SPEC CPU2017 – Next-Generation Compute Benchmark. In *ICPE '18: ACM/SPEC International Conference on Performance Engineering Companion, April 9 – 13, 2018, Berlin, Germany.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3185768.3185771

1 SPEC CPU2017 OVERVIEW

SPEC CPU2017 [1] is the next-generation compute intensive benchmark from the Standard Performance Evaluation Corporation (SPEC) [1]. SPEC CPU2017 is designed in a similar manner to its predecessor, SPEC CPU2006 [2], as both benchmarks are designed to stress a system's processor(s), memory subsystem, and compiler. Both SPEC CPU2006 and SPEC CPU2017 conducted a search program to find real world, compute bound programs, which: could be ported to multiple architectures, solve problems of various sizes, have reasonably predictable code paths, and allow platform/architecture dependent code to be disabled (unless explicitly needed for portability).

The programs and applications that were submitted to the search program were put through multiple porting and testing steps before their inclusion. Similar to SPEC CPU2006, SPEC CPU2017 also classifies benchmarks as integer or floating point benchmarks so a user can run or compare the results from a subset of all the benchmarks within the full suite that are most relevant to the user's environment. SPEC CPU2017 provides an accurate measure of compute-intensive performance across the ever-changing tech industry. Individual benchmarks within SPEC CPU2017 are ported to multiple architectures (such as 32- and 64-bit versions of AMD64, ARMv8, AArch64, Intel IA32, Power ISA, SPARC, and more), can run on various operating systems such as UNIX based operating systems (AIX, Linux, and Solaris) as well as Windows, and are supplied within the SPEC CPU2017 suite as source-code [1].

2 CHANGES IN SPEC CPU2017

Although SPEC CPU2017 does classify individual benchmarks as either integer or floating point benchmarks, SPEC CPU2017 does not use this to primarily designate the suites, which was previously the case with SPEC CPU2006 (ex. SPECint_rate metrics). Instead, SPEC CPU2017 is divided into four suites, including 43 benchmarks. The four suites - SPECspeed 2017 Integer, SPECspeed 2017 Floating Point, SPECrate 2017 Integer, and SPECrate 2017 Floating Point (ex. SPECrate2017_int metrics) - may or may not have speed and rate pairs of benchmarks that are similar to each other. Benchmark pairs that are similar to each other between a respective SPECspeed and SPECrate suite will still be different due to required differences in compile flags, run rules, and the amount of system memory required due to workload sizes. For SPECspeed suites, parallelization can be used in the form of OpenMP and/or Autopar [1], while compiler parallelization is explicitly disallowed for SPECrate suites due to run rules such as 2.2.6 [1]. SPECspeed suites require a minimum of 16 GB of system memory, while SPECrate suites require a minimum of 1 GB/copy (compiled in 32-bit space) or 2 GB/copy (compiled in 64-bit space) [1].

For a compliant run, SPEC CPU2017 still allows a user to run each benchmark within each suite for three iterations, with the median runtime from each of iteration of every benchmark selected by the suite tools to calculate the overall geometric mean score. In order to lower the cost of benchmarking, SPEC CPU2017 allows the user the choice to run each benchmark within each suite for two iterations, instead of three. In the case of running only two iterations, the slower iteration of the two will automatically be selected to calculate the overall geometric mean score.

SPEC CPU2017 has also allowed the peak run-time environment to differ by benchmarks in some cases. In addition to changing the run time environment from compile-time flags, a user may now change the run-time environment through features in the configuration file such as env_vars and submit. Similar to the SPEC CPU2006 rule that allowed a user to run different number of copies for SPECrate peak metrics [2], and the SPEC CPU2017 run rule for the number of SPECrate copies, a user can choose to run different number of threads between benchmarks within a suite for the SPECspeed peak metric [1].

One of the biggest new features of the SPEC CPU2017 benchmark suite is the option to collect non-comparable power statistics: Energy (kJ), Average Power, and Maximum Power.

[1]Press release: https://spec.org/cpu2017/press/release.html.

ICPE'18 Companion, April 9-13, 2018, Berlin, Germany

ACM ISBN 978-1-4503-5629-9/18/04…$15.00
https://doi.org/10.1145/3185768.3185771

These power statistics are used to calculate experimental metrics of energy usage for each of the possible eight metrics. The power statistics are collected with the help of PTDaemon [3].

SPEC CPU2017 lowered the cost of benchmarking for some larger systems as compared to SPEC CPU2006 since the setup for SPECrate metrics is now faster and smaller - due to the use of hard links. For SPEC CPU2017, input files are hard linked to the first copy of a SPECrate2017 metric. By comparison, in SPEC CPU2006 a copy of all input files were made for each SPECrate copy [1].

SPEC CPU2017 now requires a mandatory collection of data from the sysinfo program, which will allow for a more accurate full disclosure report since the information collected is now encrypted in the result file and not in user editable fields. This helps users of SPEC CPU2017 to know whether results accurately disclose the hardware and software configuration by comparing against details which sysinfo automatically collects. Sysinfo collects details about the processor (name, core count, threads per core, number of NUMA nodes, byte order, cache sizes, flags, etc.), NUMA node information when the run was executed, OS information (including the release version and kernel version), some memory configuration information, and the BIOS version (where applicable). The programming language standards were updated (C99, Fortran-2003, C++2003) as compared to SPEC CPU2006 (C99, Fortran-95, C++ 98) to ensure that benchmarks accepted into SPEC CPU2017 were more modern representations of real software.

3 METRICS

In SPEC CPU2017, a SPECspeed metric is a time-based metric and a SPECrate metric is a throughput metric. For SPECspeed metrics, one copy of a single or multi-threaded benchmark is executed, while recording the length of time needed to complete with a correct answer. The time the SPECspeed benchmark took to complete on the system under test (SUT), is then divided into the time the same SPECspeed benchmark took to complete on the reference machine producing a ratio. This ratio is then used as the score for that iteration of the particular SPECspeed benchmark. For SPECrate metrics, single or multiple copies of single threaded benchmarks are run. As with SPECspeed benchmarks, SPECrate benchmarks are timed with respect to how long it takes the benchmark to complete with a correct answer. Similar to SPECspeed benchmarks, SPECrate benchmarks have a ratio computed by dividing their completion time into the completion time required on the reference machine. This ratio computed for SPECrate metrics is then multiplied by the number of copies of the benchmark run. This makes SPECrate metrics a throughput metric, instead of a time only metric, see Table 1.

4 SPEC CPU2017 DISCLOSURE

The initial SPEC CPU2017 results using the reference system [4] all have "Base" metrics (under the "Results" column) of "1.00". For users familiar with SPEC CPU2006, most of the disclosure will look very similar. The full disclosure still has a graph in the middle of the first page to show the ratios for all benchmarks and the overall geometric mean ratio. The full disclosure contains familiar areas with information such as hardware and software configuration fields, a results table with the time and ratio from each iteration of each benchmark, compiler invocation details, portability flags used, optimization flags used, other compiler flags used, and areas for notes (operating system notes, general notes, and platform notes).

Each of the four reference system results contain the new, experimental, and non-comparable energy statistics; as well as choosing to utilize the new run rule (1.2.1) which allows running two iterations for a compliant run. Figure 1 shows the details of the energy statistics that SPEC CPU2017 will calculate, as well as only using two of three possible iterations.

Base Results Table

Benchmark	Copies	Seconds	Ratio	Energy (kJ)	Energy Ratio	Average Power	Maximum Power	Seconds	Ratio	Energy (kJ)	Energy Ratio	Average Power	Maximum Power	Seconds	Ratio
503.bwaves_r	1	**10027**	**1.00**	**10900**	**1.00**	**1090**	**1120**	10026	1.00	10900	1.00	1090	1120		
507.cactuBSSN_r	1	1264	1.00	1390	1.00	1100	1120	**1266**	**1.00**	**1390**	**1.00**	**1100**	**1120**		
508.namd_r	1	**949**	**1.00**	**1030**	**1.00**	**1090**	**1100**	949	1.00	1030	1.00	1090	1100		
510.parest_r	1	**2616**	**1.00**	**2840**	**1.00**	**1090**	**1100**	2615	1.00	2850	1.00	1090	1100		
511.povray_r	1	**2335**	**1.00**	**2530**	**1.00**	**1080**	**1090**	2334	1.00	2530	1.00	1080	1090		
519.lbm_r	1	1026	1.03	1170	1.03	1140	1140	**1053**	**1.00**	**1200**	**1.00**	**1140**	**1140**		
521.wrf_r	1	**2239**	**1.00**	**2440**	**1.00**	**1090**	**1100**	2239	1.00	2440	1.00	1090	1100		
526.blender_r	1	1521	1.00	1650	1.00	1080	1090	**1523**	**1.00**	**1650**	**1.00**	**1080**	**1100**		
527.cam4_r	1	1748	1.00	1900	1.00	1090	1120	**1748**	**1.00**	**1900**	**1.00**	**1090**	**1110**		
538.imagick_r	1	2486	1.00	2690	1.00	1080	1090	**2486**	**1.00**	**2690**	**1.00**	**1080**	**1090**		
544.nab_r	1	**1683**	**1.00**	**1820**	**1.00**	**1080**	**1090**	1682	1.00	1820	1.00	1080	1090		

Figure 1: Two iteration result with energy statistics

ACKNOWLEDGMENTS

The authors acknowledge the members of the SPEC CPU Subcommittee who have contributed to the design, development, testing, and overall success of SPEC CPU2017. SPEC and its benchmark names are registered trademarks of the Standard Performance Evaluation Corporation (SPEC).

REFERENCES

[1] SPEC CPU2017 home page: www.spec.org/cpu2017

[2] SPEC CPU2006 home page: www.spec.org/cpu2006

[3] SPEC PTDaemon: https://spec.org/power/docs/SPEC-PTDaemon_Design.pdf

[4] SPEC CPU2017 initial results: https://spec.org/cpu2017/results/res2017q2/

Table 1: SPECspeed and SPECrate metrics differences

Calculating SPECspeed Metrics	Calculating SPECrate Metrics
1 copy of each benchmark in a suite is run	The tester chooses how many concurrent copies to run
The tester may choose how many OpenMP threads to use	OpenMP is disabled
For each benchmark, a performance ratio is calculated as: time on a reference machine / time on the SUT	For each benchmark, a performance ratio is calculated as: number of copies * (time on a reference machine / time on the SUT)
Higher scores mean that less time is needed	Higher scores mean that more work is done per unit of time
Example: • The reference machine ran 600.perlbench_s in 1775 seconds • A particular SUT took about 1/5 the time, scoring about 5 • More precisely: 1775/354.329738 = 5.009458	Example: • The reference machine ran 1 copy of 500.perlbench_r in 1592 seconds • A particular SUT ran 8 copies in about 1/3 the time, scoring about 24 • More precisely: 8*(1592/541.52471) = 23.518776
For both SPECspeed and SPECrate, in order to provide some assurance that results are repeatable, the entire process is repeated. The tester may choose: a. To run the suite of benchmarks three times, in which case the tools select the medians b. Or to run twice, in which case the tools select the lower ratios (i.e. slower)	
For both SPECspeed and SPECrate, the selected ratios are averaged using the geometric mean, which is reported as the overall metric.	