STANDARD PERFORMANCE EVALUATION CORPORATION

spec newsletter

Volume 5, Issue 4 *December 1993*

IN THIS ISSUE

RESULTS SUMMARIES

SPEC: A Five Year Retrospective

by Larry Gray
Hewlett Packard
Fort Collins, Colorado

Today it is nearly impossible to pick up a computer trade publication and not find mention of a SPEC benchmark or the performance metrics SPECint92, or SPECfp92. The performance rating methods established by SPEC are far from complete or perfect, but they surpass any others available during the last decade and therefore, have been adopted by virtually all major UNIX based computer vendors.

In October of 1993, SPEC marked its 5th year of existence and will celebrate its successes of acceptance, credibility and growth at the January '94 annual meeting in Silicon Valley, California.

A Little History

The founders of SPEC first met in September of 1988, at Stanley's Bar & Grill in Campbell, California. Their interest was piqued by Stan Baker and Ray Weiss of Electronic Engineering Times who, as computer industry watchers and reporters, were unable to understand how to compare the power of the emerging RISC-based computer systems. "What if the vendors agreed to a common method that was based on real computational work?" was the burning question of the day.

Each vendor representative came to that meeting with the unconventional idea of cooperation among competitors, albeit in the limited area of definition of new rules for system performance evaluation. Their common goal: To establish what has come to be known as a "level playing field"; system performance methods, measures, and metrics to be used by all members.

An interesting measure of change is reflected in the current member list compared to the list of founding members: Apollo, Hewlett-Packard, MIPS and Sun. Of those four, only two remain, albeit Apollo is now part of Hewlett-Packard and MIPS is part of Silicon Graphics.

Five years ago, the most commonly used performance measures were MIPS and MegaFLOPS derived from the Dhrystone and Linpack benchmarks. Unfortunately, it was also common knowledge that those measures were not directly comparable vendor to vendor, since each had its own "method" of computing MIPS. Although in decline, MIPS ratings are still in use today. To their credit, there are a few vendors who no longer publish or disclose MIPS ratings for their systems.

Technology has advanced so rapidly that, since the first newsletter, integer performance of RISC based workstation-class systems has increased

See **SPEC,** *page 3*

SPEC, *from page 1*

by a factor of up to 17 times, depending upon the measure selected. Using the Sun SPARCstation 1 as a base, (tested in September of 1989 and reported in the premiere issue of the *SPEC Newsletter*,) I compared it to the MIPS M/2000, the fastest system in the first newsletter. To form an even more interesting comparison, I also selected Digital's DECstation 3000 Model 500X and the IBM RS/6000 Powerserver Model 990 reported in this newsletter (see Table 1.)

The benchmarks from the SPEC Release 1 suite are compared with the same benchmarks which continue to be used in the current CINT92 suite.

Table 1: SPECratios

	Fall 1989		Winter 1993	
	Sun SS1	MIPS M/2000	DEC 500X	IBM 990
022.li	9.0	23.6	106.7	130.5
023.eqntott	9.7	17.8	139.2	164.2
008.espresso	8.9	8.1	113.5	93.8
Improvement factors				
	Fall 1989		Winter 1993	
	Sun SS1	MIPS M/2000	DEC 500X	IBM 990
22.li	1.0	2.6	11.9	14.5
023.eqntott	1.0	1.8	14.4	16.9
008.espresso	1.0	2.0	12.8	10.5

The SPEC benchmarks gave C and FORTRAN compiler developers application-like codes to study. New techniques for code optimization were developed that yielded (processor) architecture specific improvements in the execution efficiency of machine instruction streams. Performance improvements on the order of 2x have come from compiler, optimizer and source-code preprocessing software alone.

We believe the software advances to be a significant contribution to the industry because utilization of the software capabilities can be realized with applications that are functionally similar to the SPEC benchmarks.

The combination of new processors and compiler sophistication brought SPEC Release 1 benchmark execution times of some tests down into the range of four to eight seconds. One benchmark, 030.matrix300, was improved such that results were over 1,000 times faster than the SPEC Reference Time. SPEC members were then compelled to develop a new suite of benchmarks that would contain new codes, perform more work, run longer, and survive yet another projected doubling of system performance. The SPEC CINT92 and CFP92 suites were announced in January, 1992.

The incredible increases in computer system processing power are projected to continue through this decade and SPEC intends to release new suites of CPU benchmarks every three to four years.

SPEC Products

SPEC is best know for two types of products, its benchmarks and the *SPEC Newsletter.*

The newsletter is the SPEC vehicle for distributing results and information about performance evaluation, the benchmarks, and other performance topics. The newsletter, published quarterly, has grown to become a formidable publishing task. The first newsletter, published in the Fall of 1989, presented a total of 14 results pages. The largest newsletter to-date (June '93) was 155 pages, with 134 pages of benchmark results. Our subscriber list has grown to over 260 with addresses around the world.

Due to the extraordinary efforts over time of several volunteer newsletter editors, article authors, results submittors and the enthusiastic and capable staff of NCGA, we look forward to more frequent publication.

The following table provides a summary of the SPEC benchmark products and their introduction dates. Further definition of the benchmark suites can be found later in this newsletter.

To-date, over 800 benchmark tapes have been shipped to SPEC benchmark licensees who now number over 700.

Table 2: SPEC Benchmarks Summary

Suite Name	*Introduction Date*	*Metric*	*Type - Capability measured*
CPU Rel. 1.0	Oct.,1989	SPECmark89*	CPU-Floating Point & Integer
CPU Rel. 1.2b	Mar., 1990	SPECthruput*	CPU-Floating Point & Integer
SDM 1	May, 1991	SDET peak thruput	System-s/w development
SDM 1	May, 1991	Kenbus peak thruput	System-academic usage
Cint92	Jan., 1992	SPECint92	CPU-Integer-speed
Cfp92	Jan., 1992	SPECfp92	CPU-Floating Point-speed
Cint92	June, 1992	SPECrate_int92	CPU-System Capacity-integer
Cfp92	June, 1992	SPECrate_fp92	CPU-System Capacity-floating point
SFS 1	Mar., 1993	LADDIS ops/sec	System NFS Server Capacity

*NOW obsolete

SPEC benchmarks are recognized around the world for their value in providing performance metrics that are

See **SPEC,** *page 4*

SPEC, *from page 3*

timely, credible, and, in certain application environments, useful as predictors of application performance.

SPEC Membership

SPEC's membership has swelled from the original four founders to 30 member companies (as of 11/8/93) with much of the increase in the past two years and in spite of mergers. SPEC also has attracted five non-profit institutions as associate members and expects more associates to join in 1994. (The list of current members and associates can be found on page 13).

As an industry consortium, SPEC is staffed with volunteer leaders and a staff of no less than nine full-time engineers, with many more part-time enginee s that work on the various benchmark projects, documentation, run rules, reporting formats and many other details in the benchmark development and delivery processes.

A significant event in SPEC's history occurred mid-1993 with SPEC hiring its first full-time employee, Dianne Dean, as our Administrator of Operations. We are confident that this step will smooth the management of new products, member and subscriber growth and improve the productivity of SPEC's leaders. Other SPEC administrative affairs such as member and subscriber services and accounting are handled by NCGA, the SPEC administrator since 1991.

In conclusion, SPEC's contributions to the computer industry have been brought about by supplying application-based benchmark codes and the institution and compliance with run and reporting rules that insure one vendor's SPECint92 metric can be directly compared to that of another. Published results have credibility across the industry. SPEC has and will continue to fulfill its prime directive of providing a level performance evaluation playing field.

The Future

A wide variety of computer system performance evaluation challenges lie ahead. If you are not already a member, we invite you to join with us, cast your vote on issues and actively participate in what promises to be another extremely exciting five years in the world of open systems computing.

Larry Gray is a member of Hewlett-Packard's Cooperative Systems Evaluation team in Cupertino, Calif. He serves SPEC on its Board of Directors and is Chair of the SPEC Planning Committee.

Read *ALL* The Numbers

by Walter Bays
Sun Microsystems
Mountain View, CA

SPEC has always advised that you look at the individual benchmark SPEC ratios, not just at the composite SPECint92 and SPECfp92 metrics. The composite will most likely not give you as good of an indicator of performance as will examining the individual ratios with an understanding of your own workload. High-level performance characteristics (time to run an application) vary due to underlying variations in low level performance characteristics, e.g. differences in CPU pipelines, cache and TLB size and organization, memory system, compiler, and libraries. The advice that you read *ALL* the numbers is more important than ever now that multiprocessor systems are moving out of the machine room onto the desktop.

If a workload comprises multiple activities whether from multiple users, from multiple applications for a single user, or from multiple tasks in a single application, the SPECrate throughput metrics may be appropriate. But for a single task, the speed metrics are appropriate. (See "Clarification of Speed Versus Capacity" in the June 1993 *SPEC Newsletter*). This article deals only with the floating point speed metric SPECfp92, not SPECrate_fp92 or the integer metrics SPECint92 and SPECrate_int92.

It is possible to utilize multiple CPU's on a single-task application, and on the SPECfp92 benchmarks, by use of advanced compilers and preprocessors able to automatically compile applications to utilize multiple CPU's. All of SPEC's run rules still apply of course, so the benchmark cannot be rewritten to better use multiple CPU's, nor can directives giving hints to the compiler be inserted in the source code. The compiler must by itself discover opportunities for parallel execution, and generate code to exploit those opportunities. (We say the compiler automatically "parallelizes" the program.) On some programs the compiler may find no such opportunities, and only one CPU can be used.

These differences in effectiveness of parallelization lead to even greater variations in low-level performance characteristics among uniprocessors and multiprocessors. Added to the variations in relative CPU speeds are variations in how many CPU's can be effectively applied to a given algorithm. Hence there is greater variation in high-level performance characteristics from slowest application to fastest application, and from SPEC ratio to SPEC ratio.

These wide variations are no surprise to those in the supercomputing world where scientists have long restructured applications to obtain maximum benefit from vector and/or parallel systems, using both manual techniques and GUI based semi-automatic tools. But it may be new to some in the workstation and server world.

For example, consider "System M", a multiprocessor, with "System U", which has a single faster CPU. System M has a SPECfp92 of 118.5 while System U has a SPECfp92 of 112.5. I will simply call these "Systems M and U", because the point of this article is not competitive performance comparison but an examination of performance variations. It is important only that System M has four CPU's on one architecture and System U has a single CPU of a different architecture. Thus they have different low level performance characteristics. Table 1 below lists SPEC ratios for these two systems, plus a SPARCstation 10 Model 51 for reference. Benchmarks marked (*) were automatically parallelized on System M.

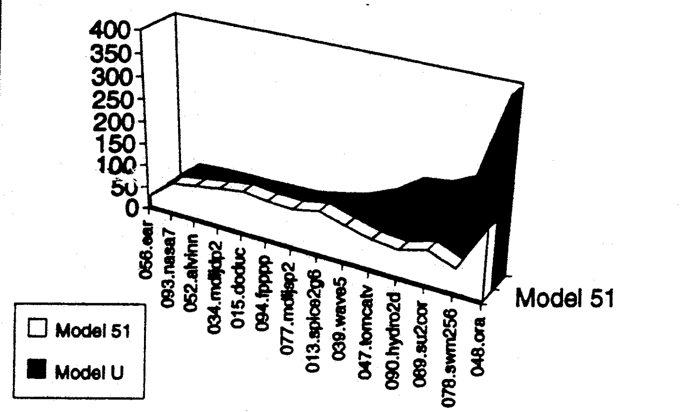
Table 1

	System M	System U	Model 51
013.spice2g6	54.8	58.3	56.2
015.doduc	68.4	86.5	81.2
034.mdljdp2	78.5	100.4	86.4
039.wave5*	76.9	72.0	57.9
047.tomcatv*	155.9	118.3	77.3
048.ora*	414.5	105.1	160.3
052.alvinn	176.0	228.2	175.6
056.ear	103.5	382.9	95.0
077.mdljsp2	40.5	49.6	42.2
078.swm256*	169.8	82.4	43.7
089.su2cor*	266.5	49.5	113.3
090.hydro2d*	228.7	133.9	79.1
093.nasa7*	102.1	138.4	93.5
094.fpppp	94.6	116.5	101.7
SPECfp92	118.5	112.5	83.0

If you concluded from SPECfp92 alone that the System M is as fast as System U, you would be missing much of the picture. In examining the SPEC ratios above, you see that on one benchmark, it is only 27% as fast as System U. On seven benchmarks, it is 81% as fast (geometric mean). And on six benchmarks, it is 181% as fast. This is depicted in Figure 1, in which the benchmarks are listed in order of increasing relative speed.

Figure 1

Speed relative to "System U"



Unsurprisingly, the six benchmarks where System M is as fast as System U were all parallelized. Among the seven benchmarks where it is 81% as fast, six were not parallelized, and one (093.nasa7) was only partially parallelized. Clearly a key consideration if you were considering these two systems, and were primarily concerned with single task performance, would be how well your applications could be parallelized.

Finally consider 056.ear, where System M is 27% as fast as System U. This benchmark was not parallelized, and the 4:1 speed difference comes from other differences in low level performance characteristics, hardware and software. This serves as a reminder that, even when comparing uniprocessor to uniprocessor, your best course is to understand your workload and **read *ALL* the numbers.**

Walter Bays is a manager in Sun Microsystems Computer Corporation's System Architecture and Performance Group in Mountain View, California. He is the Chairman of SPEC's Open Systems Steering Committee and serves on the SPEC Board of Directors.