STANDARD PERFORMANCE EVALUATION CORPORATION

spec newsletter

Volume 4, Issue 1 *March 1992*

INSIDE THIS ISSUE

RESULTS SUMMARIES

SPEC Announces New Benchmark Suites CINT92 and CFP92

On Wednesday, January 15, 1992, the Standard Performance Evaluation Corporation (SPEC) announced the availability of its newest CPU-intensive integer and floating-point benchmark suites.

These two benchmark suites build upon the strengths of the now popular SPEC Release 1 CPU suite and continue to provide users with a thorough and equitable method for evaluating the CPU performance of high-end computer systems.

"These new suites increase both the breadth and the depth of understanding of the CPU performance of computer systems used in the commercial and scientific arena," said Kaivalya Dixit, President of SPEC and an engineering program manager at Sun Microsystems.

The new integer suite (**CINT92**) is comprised of 6 real-world application benchmarks. **CINT92** represents application areas in: circuit theory, LISP interpreter, logic design, text compression algorithms, spread-sheet, and software development.

The new floating-point suite (**CFP92**) is comprised of 14 real- world application benchmarks, five of which are single precision. **CFP92** represents application areas in: circuit design, Monte-Carlo simulation, quantum chemistry, optics, robotics, quantum physics, astrophysics, weather prediction, and other scientific and engineering problems.

"The 20 benchmarks contained in the new CPU suites doubles the number from Release 1, allowing more diverse user environments and application characteristics to be represented. The benchmarks have been enhanced to keep pace with the higher-performance systems that have entered the marketplace", said John Laskowski, SPEC Board member, and manager of the Performance Evaluation Center at IBM.

"Computer system performance is increasing by at least 50% per year. The continuing advances in hardware and software technology require concurrent advances in system evaluation tools and techniques. New benchmarks to measure overall system throughput, networking and disk input/output are being readied for release in 1992 and 1993," said Larry Gray, SPEC Board Member and performance manager at Hewlett-Packard.

"The year 1991 saw major progress in understanding performance

(Continued on page 3)

SPEC Announces New Benchmark Suites

(Continued from page 1)

measurements with the SPEC Release 1 benchmark suite. Dramatic improvements in architecture, both hardware and software have significantly reduced runtimes of several Release 1 benchmarks. The new benchmarking suites have remedied potential problems of very short run times and unrealistic optimizations from constant propagation," said Jeff Reilly, the release manager of the new suites, and performance engineer at Intel.

The new suites provide a wider "SPECtrum" of performance characterization by adding new application benchmarks to both the integer and floating-point suites. The new suites are decidedly better because:

- Integer and floating-point benchmarks are in two separate suites.
- **CINT92** includes additional application areas like a compression algorithm, and a spreadsheet. These benchmarks make system calls and spend some time in the operating system just like many real applications.
- **CFP92** includes single-precision applications in C and FORTRAN.
- **CFP92** includes new application areas like quantum chemistry, optics, robotics (neural network), shallow water models (weather prediction), quantum physics and astrophysics.

SPEC determined that the original SPECmark rating methodology (composite of **SPECint** and **SPECfp**) is not suitable for the new suites because:

- The two separate suites have fundamentally different workload characteristics,
- There are significantly more floating-point benchmarks (14) than integer benchmarks (6). This is appropriate, as floating-point benchmarks show substantially more variability in performance. Hence more benchmarks are needed to characterize machines,
- Separate measures for the two suites give customers in different market segments a better idea of their own performance. Many vendors found it useful to separate integer and floating-point subsets, which came to be officially called **SPECint89** and **SPECfp89**,
- They solve current problems of either inflating or deflating measured performance.

The new metrics SPEC has chosen are **"SPECint92"** and **"SPECfp92"**. They are measures of integer and floating-point processor performance. Each metric is standalone - a composite of results from the **CINT92** and **CFP92** benchmark suites, respectively.

"SPECint92 is a broadened version of **SPECint89**, that completely overshadows the infamous 'mips-rating'. Integer benchmarks are considered by SPEC to be the least common denominator among divergent processor and system architectures, as most processors spend at least some time running pure integer code (operating systems, compilers, DBMS, most commercial code, etc). **SPECfp92** indicates floating point performance over a wide range of scalar and vector floating point, and is relevant to many scientific tasks, large parts of engineering, and some parts of commercial work", said Bud Funk, SPEC Steering Committee Chairman and Benchmark test manager at Unisys.

"Despite the popularity of the SPECmark, SPEC has always advocated use of all individual benchmarks to characterize performance. We continue to recommend this practice for the six SPEC ratios that make up **SPECint92**, and the 14 that make up **SPECfp92**, as the variations in these numbers provide valuable information. (For example, as compilers have attacked 030.matrix300 – a SPEC Release 1 benchmark, it was important that people report the individual benchmarks, lest the source of improvements be misunderstood.)", said John Mashey, VP of technology at MIPS.

SPEC asks for your support for these new benchmark suites and in adopting these new metrics. SPEC very much appreciates your continued support.

CINT92 and CFP92 Questions and Answers

By Subra Balan (IBM), Kalvalya Dixit (SUN), Walter Bays (SUN)

WHAT ARE THE NEW SPEC CINT92 and CFP92 BENCHMARK SUITES?

CINT92 and **CFP92** are CPU-intensive benchmark suites. **CINT92** is the integer suite comprised of six benchmarks written in the C language. **CFP92** is the floating-point suite comprised of 14 benchmarks, two of which are written in C and the rest in FORTRAN. Five of the 14 are single precision and the rest are double precision. All of these benchmarks are based on real applications. Detailed descriptions of the individual benchmarks may be found in the December 1991 issue of the *SPEC Newsletter* (Volume 3, Issue 4).

WHAT NEW APPLICATIONS ARE AVAILABLE IN THE TWO BENCHMARK SUITES?

Both a data compression application and a spreadsheet application have been added to **CINT92**. Several applications have been added to **CFP92**. These include robotics, astrophysics, weather, and quantum physics.

ARE THE SPEC RELEASE 1 BENCHMARKS PART OF THESE NEW SUITES?

Three of the four integer benchmarks (008.espresso, 022.li and 023.eqntott) are part of the new **CINT92** suite. The longer running 085.gcc replaces the fourth (001.gcc). Three of the six floating-point benchmarks(013.spice2g6, 015.doduc, and 047.tomcatv) are part of the new **CFP92**. 093.nasa7 is a more robust version of 020.nasa7 (Release 1 benchmark) and 094.fpppp is a longer running version of 042.fpppp (Release 1 benchmark).

HOW ARE THE NEW BENCHMARK SUITES DIFFERENT FROM THE SPEC RELEASE 1 SUITE?

Although nine of the 20 benchmarks are from SPEC Release 1, these nine have been made more robust and less susceptible to compiler optimizations. Also, the run time for some of the benchmarks has been increased to keep pace with the faster CPUs available today. Additionally, the floating-point suite includes two C language benchmarks and five single-precision benchmarks that did not exist in SPEC Release 1. The larger number of benchmarks and the diversity of applications they represent make these a significantly improved set of benchmarks compared to the SPEC Release 1 suite.

IS THE ORIGINAL SPEC RELEASE 1 SUITE OBSOLETE?

The original SPEC Release 1 suite will continue to be useful to those customers whose applications are represented by one or more of the 10 individual benchmarks. Additionally, there is a large database of results available for the SPEC Release1 suite that will continue to be useful in comparing systems. Over a period of time, more results will be available for the new suites, and as chips and architectures get faster, the **CINT92** and **CFP92** suites will take the place of the SPEC Release 1 suite.

WHAT ARE THE METRICS FOR THE NEW BENCHMARK SUITES?

SPECint92 is the composite metric for **CINT92**. It is the geometric mean of the SPECratios of the six integer benchmarks. The SPECratio for a benchmark on a given system is the quotient derived by dividing the SPEC Reference Time for that benchmark (run time on a DEC VAX 11/780) by the run time for the same benchmark on that particular system.

SPECfp92 is the composite metric for **CFP92** and is the geometric mean of the SPECratios of the fourteen floating-point benchmarks.

WHY IS THERE NO SPECmark FOR THE NEW SUITES?

SPECmark is the composite metric of performance based on all 10 of the benchmarks in SPEC Release 1. The relative performance of different systems may be different for integer and floating-point depending on the architecture. That is, two machines with the same SPECmark could flip flop positions on integer and floating-point performance. Therefore, in order to better position systems for the user community, SPEC adopted the SPECint and SPECfp sub-composites in December 1991. SPECint is the geometric mean of the four integer benchmarks in SPEC Release 1 and SPECfp that for the six floating-point benchmarks. Positioning of systems could be performed more accurately with these than with the overall SPECmark.

In keeping with the same intent of complete disclosure of the performance characteristics of the various systems, SPEC has chosen not to adopt an overall performance metric similar to the SPECmark for the new suites.

Further, there are 14 floating-point benchmarks

(more needed due to the higher variability in results across different architectures) and only six integer benchmarks. Combining these into a meaningful single number that would provide the same information as these two composites may not be practical.

HOW WOULD I USE THESE TWO BENCHMARK SUITES IN COMPARING SYSTEMS?

SPEC has always maintained that the best way of comparing two or more systems is using the actual user application. Where this is not feasible, standard benchmarks such as those from SPEC are a second alternative. While systems may be compared using the overall **SPECint92** and **SPECfp92** composites, there are some risks in doing so. SPEC recommends that the users study the profile of each individual benchmark and select one or more that are most representative of their environment and compare systems based on their performance on these benchmark(s). The profiles for the **CINT92** and **CFP92** benchmarks are available in the December 1991 issue of the *SPEC Newsletter* (Volume 3, Issue 4). Also see the article 'Interpreting SPEC Release 1 Results in the December 1991 issue of the *SPEC Newsletter.*

WHEN SHOULD THE INTEGER SUITE BE USED AND WHEN SHOULD THE FLOATING-POINT SUITE BE USED?

The floating-point suite may be used to compare the floating-point-intensive (typically engineering and scientific applications) environment. The integer suite may be used for those environments that are not floating-point intensive. It is a good indicator of the base CPU performance in a commercial environment (CPU performance is one of the many components such as disk and terminal subsystems, memory, and OS services that contribute to performance in a commercial environment).

WHY DID SPEC SEPARATE THE BENCHMARKS INTO TWO DIFFERENT PRODUCTS?

The characteristic of integer and floating-point applications are quite different from one another. User applications also typically tend to be floating-point intensive or integer intensive. Therefore, structuring the new suites this way allows SPEC to address each of these two important user segments.

WHY DID SPEC NOT INCLUDE 030.MATRIX300 IN THE NEW CFP92 SUITE?

There are several reasons why SPEC did not include 030.matrix300 in the new CFP92 suite. They are:

- New technologies (architecture, hardware and software) have significantly reduced the run time on this benchmark. On some platforms, the benchmark completes the run in six seconds. Short run times are difficult to reproduce and prone to measurement errors,

- 030.matrix300 is a kernel benchmark and the new suite contains applications that use code similar to 030.matrix300 code. In general, SPEC wants to use real applications as benchmarks rather than kernels. 093.nasa contains a kernel (matrix multiply) which performs computations similar to 030.matrix300,

- 030.matrix300 has no protection from aggressive compilers against constant propagation and dead-code elimination techniques.

ARE THESE NEW SUITES JUST A REACTION TO THE PERFORMANCE INFLATION BY 030.MATRIX300?

These new releases have been in development for about two years. Recently, several companies have shown impressive performance improvements on matrix manipulation codes by using vector preprocessors. SPEC allows such techniques to speed up the benchmarks, just as they benefit the end user.

The 030.matrix300 benchmark was not included in the **CFP92** suite primarily because it is a synthetic benchmark, and thus is more susceptible to various optimizations than are real end-user applications. However, a number of the **CFP92** benchmarks are real applications that include matrix manipulation code which can benefit from preprocessing in a realistic manner.

WHAT IS THE DIFFERENCE BETWEEN 020.NASA7 and 093.NASA7?

NASA7 is a separate benchmark which is in both SPEC Release 1 and in **CFP92**. It does include matrix multiplication like 030.matrix300, but does many other computations as well. For **CFP92**, 020.nasa7 was modified to split the code into separate files, to add input data, and to add output data. These changes guard against unrealistic compiler optimizations. The reason the SPEC ratios are lower is simply that the machine which generates the reference times, a DEC VAX-11/780, now

(Continued on next page)

runs the benchmark faster. (Performance tuning applies to older machines as well as new ones.) Because of the source code modifications and the new reference time, nasa7 is given a new identification number in **CFP92** (093.nasa7) to guard against confusion with the old SPEC ratios from 020.nasa7.

DO THE NEW SUITES REPLACE THE SPEC RELEASE 1 SUITE?

SPEC strongly recommends use of the new suites to characterize CPU performance because it addresses some of the shortcomings of SPEC Release 1. The SPEC Release 1 will be out of the price list some time in July 1992. SPEC will respect and meet customers' needs and gracefully retire the SPEC Release1 suite.

DO THE NEW SUITES REPLACE THE SPEC SDM SUITE?

No, the new suites measure component-level CPU performance. SPEC SDM measures system-level performance.

WHAT IS THE DIFFERENCE BETWEEN THESE SUITES, SPEC SDM, AND COMMERCIAL BENCHMARKS?

These suites address measurement of two aspects of a system, integer and floating-point CPU speed. The SPEC SDM benchmarks address measurement of overall system throughput with given workloads. Some commercial benchmarks also address system throughput measurement. People evaluating computer systems need to look at both kinds of measurements.

THE OLDER SPECint AND NEWER SPECint92 POSITION SYSTEMS ALMOST IDENTICALLY. WHY SHOULD ANYONE BUY AND USE THE NEW CINT92 SUITE?

The primary reason for using the new **CINT92** is that it contains two new applications (e.g. spreadsheet and compression algorithms). Additionally, the new version of the GNU Compiler benchmark has a significantly larger workload. Every new application fills a void in the performance SPECtrum. Six different SPEC ratios give a more complete picture of integer performance.

WHAT DO THESE BENCHMARK SUITES COST? IF I HAVE A SPEC RELEASE 1 LICENSE, IS THERE A LOWER UPGRADE PRICE?

The product pricing in U.S. dollars is as follows :

Product	Existing Customers*	New Customers
CINT92	$300	$425
CFP92	$400	$575
Combination	$600	$900

** Customers with an existing SPEC Release 1 license may upgrade to the new suites at these upgrade prices for a period of six months after the* **CINT92** *and* **CFP92** *announced on 01/15/92.*

WHY HAS THE PRICE INCREASED FROM $300 FOR RELEASE 1 TO $900 FOR THE NEW INTEGER AND FLOATING-POINT SUITES?

The new SPEC benchmark suites represent a significant improvement in both the breadth of workloads and the quality of the products. The price reflects the cost of development and projected cost of supporting the suite for our customers during the life of the benchmarks. The price of these benchmark suites is very reasonable considering these factors. Additionally, the increase in the number of customers, and the number of benchmark suites has increased SPEC's administrative costs.

Subra Balan works at the IBM Performance Evaluation Center in Roanoke, TX.

Kaivalya Dixit is the SPEC president and an engineering program manager for SUN Microsystems, Inc., of Mountain View, California.

Walter Bays is a performance engineer for Sun Microsystems, Inc., of Mountain View, California.