COMPARING CPU PERFORMANCE BETWEEN AND WITHIN PROCESSOR FAMILIES

Lee A. Butler
US Army Research Laboratory

Travis Atkison
Department of Defense

Ethan Miller
University of Maryland,
Baltimore County

Abstract

Our study compares CPU performance on RISC and CISC uni and multiprocessors of varying speeds, and shows that the Instruction Set Architecture (ISA) style no longer matters. Our study ran the BRL-CAD™ benchmark on over a dozen computers with several different ISAs. We discuss the performance similarities and differences both between ISAs and within each ISA family. We found that the two dominant factors in CPU performance for our program were clock speed and the number of functional units; the complexity of the instruction set made little difference.

Introduction

The goal of this research project was to assess currently available architectures for their suitability to perform BRL-CAD™ analysis. The primary computational activity for this type of analysis involves raytracing complex geometry. This process can be performed in either a serial or parallel manner. The BRL-CAD™ software package includes a benchmark suite [MUSS88] to help users assess their system performance requirements.

This study was conducted in preparation for upgrading our computing environment. This environment typically consists of three classes of hardware: Desktop, Workstation, and Server. The Desktop class consists of machines that will be used only occasionally or for small-scale analyses. Such machines are generally low-cost single-processor systems with a medium amount of memory (e.g., 64 MB by current standards). The Workstation class consists of machines that will be used on a regular basis for modeling and moderate-scale analyses. Selection of these systems is based primarily on performance. They are typically high-end uniprocessor or multiprocessor systems with two to four CPUs and typically more memory than the Desktop category (e.g., 512 MB-1 GB in current systems). The Server is our last class of hardware. This type of system is designed to handle large-scale analyses. Such systems typically have eight or more CPUs with memory capacities ranging from 2 GB to 128 GB in current systems. Performance, not cost, is usually the primary selection criteria.

Architectures tested include systems available from SGI, Sun Microsystems, and a variety of Intel-derived platforms. For perspective, several older systems were also tested. Historically, architectural differences among these CPU families have accounted for variations in performance.

Overview of BRL-CAD™ Suite

The benchmark consists of computing six reference images from Computer Aided Design (CAD) models (see Figure 1). These optical renderings are compared with reference images and verified for correctness. The images consist of 24 bit RGB pixel values (8 bits for each color channel). Images are considered correct if pixel channel values differ by no more than 1 from the reference. This accommodates the variability due to differences in floating point representation of the various CPU architectures.

The six reference models used represent differing levels of complexity. The first three (moss, world, & star) represent simple validation of correctness tests. The latter three models (Bldg391, M35, and Sphflake) strongly resemble those used in production analysis. The inclusion of these models assures that there is a strong correlation between performance on the benchmark and performance in actual use.

The benchmark reports the number of ray/geometry intersections performed per second during the raytracing phase of the rendering. Time spent reading the model into memory and performing setup operations is not included in the performance evaluation. This gives a more realistic prediction of actual performance, since typical applications spend
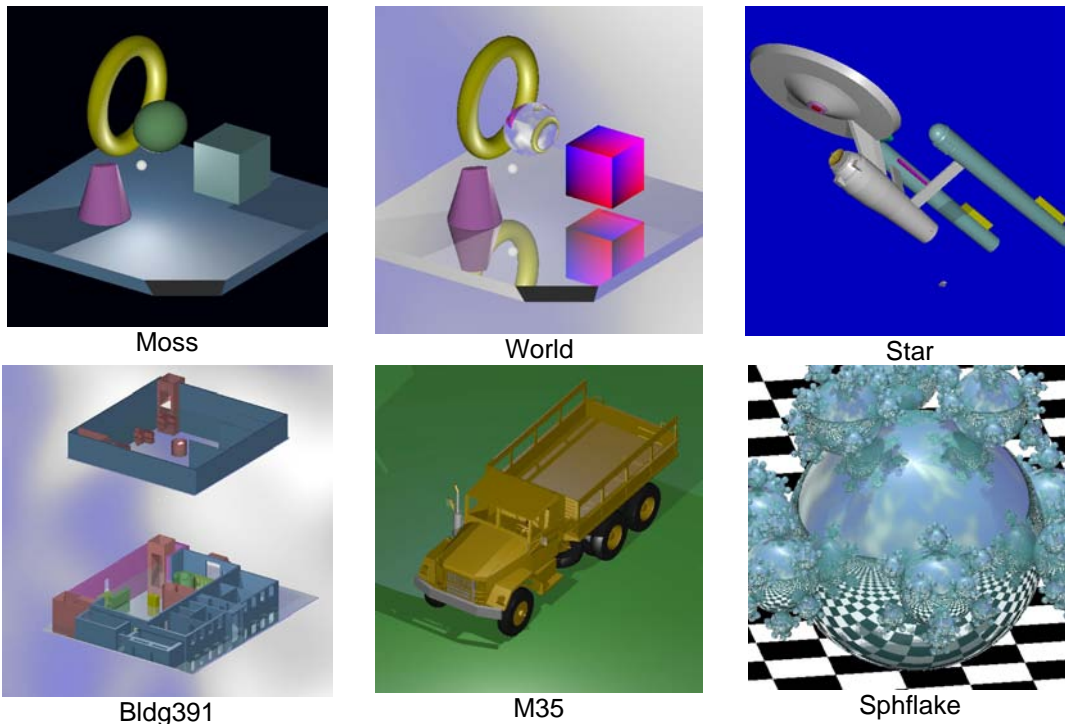
Figure 1. Benchmark Reference Images.

relatively little time in I/O and setup compared to raytracing. The amount of time the benchmarks spend doing I/O and setup is about equal to that spent raytracing.

Experiment

The benchmark suite was run on a variety of systems made available for this experiment. The suite was run on systems under minimal load conditions. In most cases, it was the only task being run. For each system, the suite was run a minimum of three times, with average values and standard deviations computed. The suite was run in single-processor mode on all systems. On multiprocessor systems, additional tests were performed using varying numbers of CPUs. For comparison purposes, all results were normalized to the results from a Pentium running at 166 MHz.

In our experiment, the architecture families tested included the Intel x86, MIPS, and Sparc. The x86 architecture machines ran under the Linux OS environment, with the EGCS version of GNU's gcc used to compile the suite. In contrast, the suite run on the MIPS architecture was compiled on an SGI R4400 system running Irix 5.3 with version 5.3, of the compiler tools. Finally, on the Sparc architecture, version 4.2 of the compiler was used under the Sun Solaris 2.6 OS. The BRL-CAD™ benchmark does not benefit from new instructions introduced by vendors to support application domains such as multimedia

applications. As a result, we did not recompile the suite for each machine tested within an architecture family.

Table 1 shows the versions of operating systems used to perform the testing.

Architecture	Operating Systems
Intel	Linux 6.1 and 6.2
MIPS	Irix 6.5 and 5.3
Sparc	Solaris 2.6

Table 1. Operating Systems for each Architecture

Architectural information on each of the CPUs tested was obtained. Based upon past experience, this information was likely to provide key insights into the results of the tests. As we shall see, this type of information proved less useful than expected.

Single-CPU Results

The results from many of our single-CPU tests (normalized with a Pentium 166) can be seen in Figure 2. What immediately stands out from this figure is that the AMD Athlon 700 MHz shows a clear advantage, being roughly two times faster than the nearest competitor. Next in line is the Pentium-III 500 MHz, which is slightly faster than the top-of-the-line MIPS or SPARC offerings. It is interesting to note that neither the Pentium nor Athlon CPUs represents the vendors' current top offerings.
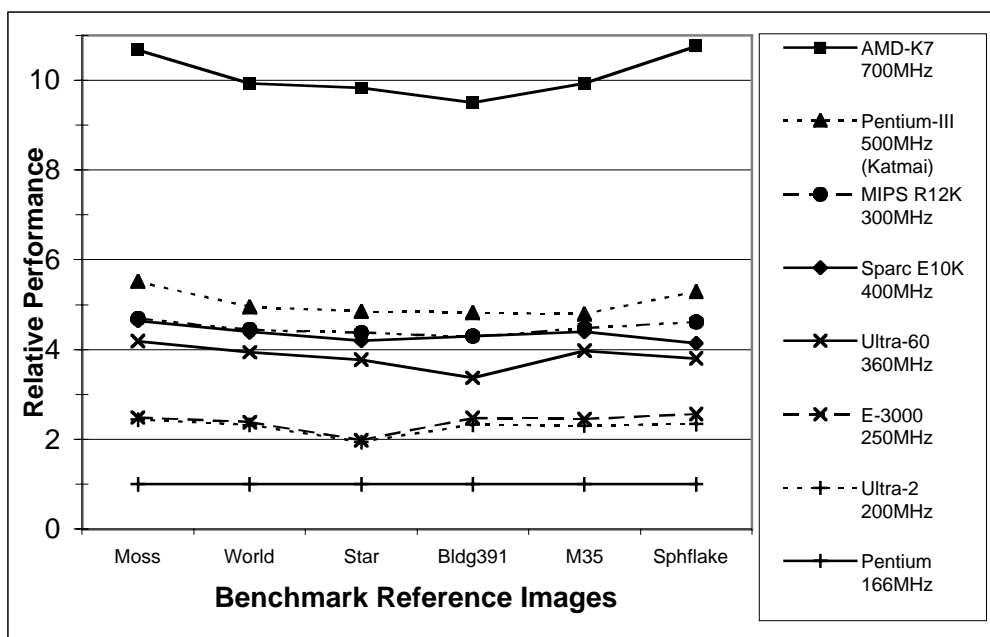
Figure 2. Performance Normalized with Pentium 166MHz.

Another interesting observation is the slight dips in performance on the E3000 and Ultra-2 on the "Star" benchmark and the Ultra-60 on the "Bldg391" benchmark. Examining the architecture of these machines reveals very minor differences. The E3000 and Ultra-2 employ a five-state cache system, which might have accounted for their performance loss. However, there was nothing unusual about the Ultra-60 architecture that might explain its drop in performance on the "Bldg391" benchmark.

A more concise view of the benchmark results can be seen in Figure 3. The values for the benchmarks have been averaged for each system. The systems are presented in historical order and by architecture family. The results for all systems tested are presented in this figure. For Clarity some of these were not shown in Figure 2.

There are two items of particular interest in Figure 3. The first is the discrepancy between the two Pentium-II 266MHz systems. One of these systems was a desktop with a Pentium-II CPU while the other was a laptop using a Mobile Pentium-II CPU. An examination of these two CPUs reveals nothing in their design that could account for this performance variation. Since laptops are frequently designed to sacrifice performance in favor of lower power consumption, it is likely that the difference is the result of some aspect of the laptop external to the CPU.

The second item of note in Figure 3 is the relative performance of the Ultra-2 vs. the Ultra-5. The former is an UltraSparc-II 200 MHz CPU, while the latter is an UltraSparc-IIi 270 MHz CPU. The Ultra-2 is the culmination of its generation of processor architecture, with the Ultra-5 being the low end of the succeeding generation of processor architecture. It seems that the focus has now shifted from leveraging architecture to maximizing clock speed to attain increased performance. Thus, we infer that the Ultra-5 CPU is designed to run at much higher clock rates than that of the Ultra-2. As a result, it is less aggressive about performing as many operations per clock cycle as possible. Figure 4, where clock speed has been factored out of the performance values supports this hypothesis.

Looking further at Figure 4, it becomes clear that AMD has achieved its performance in the K6-2 family primarily through increasing the clock rate at which its CPU can run. The clock-independent performance of the K6-2 is only slightly better than that of the Pentium-166. The slight difference can be accounted for by an increase in cache size.

In addition to delivering the highest performance level, Athlon systems were priced at or below other desktop systems. This makes it both the performance and price/performance leader in the desktop class.

Workstation Results

The results for systems with one to four CPUs are shown in Figure 5. For comparison, the values for a single processor Pentium-III 500 MHz and Athlon 700 MHz are also displayed. From this graph, we can see that the single Athlon outperforms all dual processor systems. It takes three E10000 processors to exceed the performance of one Athlon and four of the MIPS
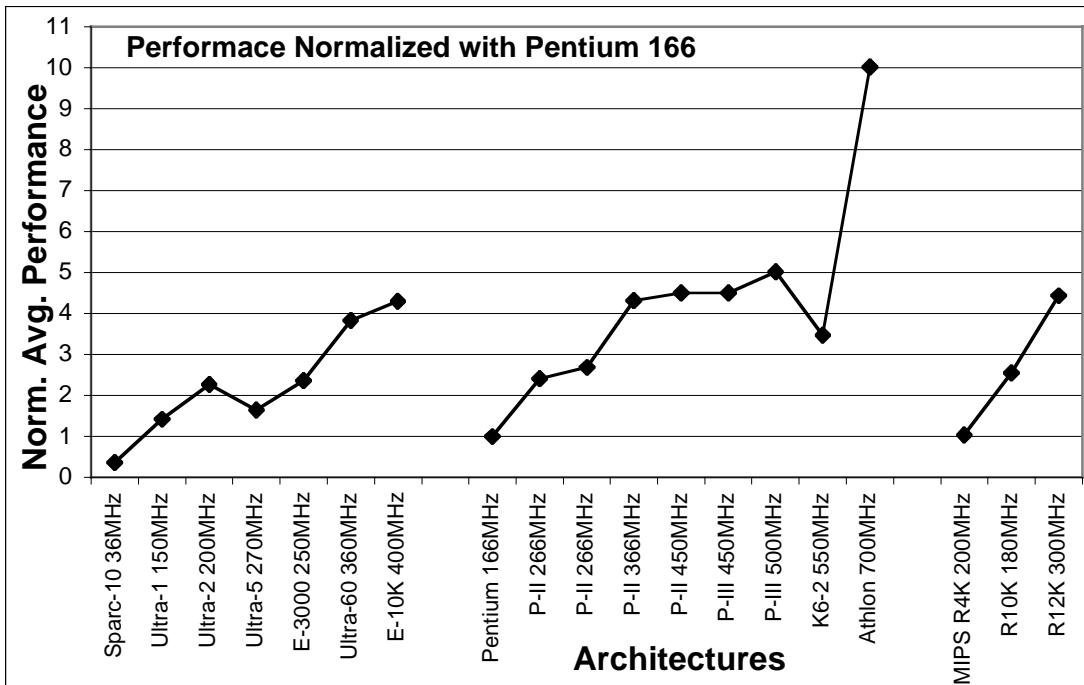
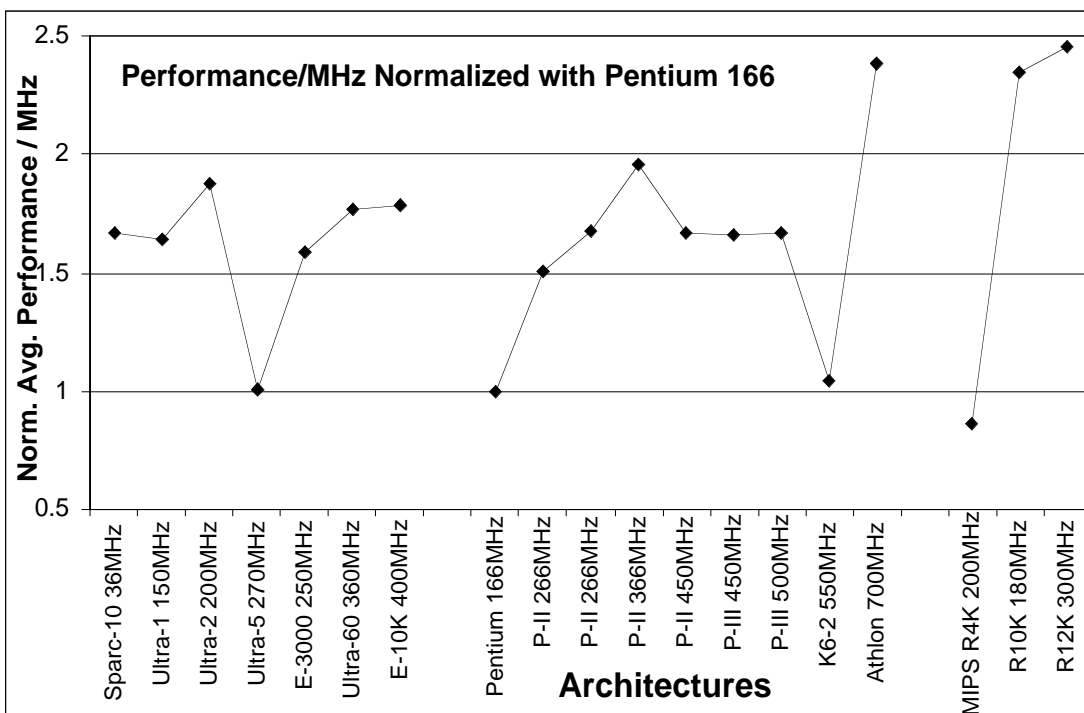Figure 3. Average Relative Performance.

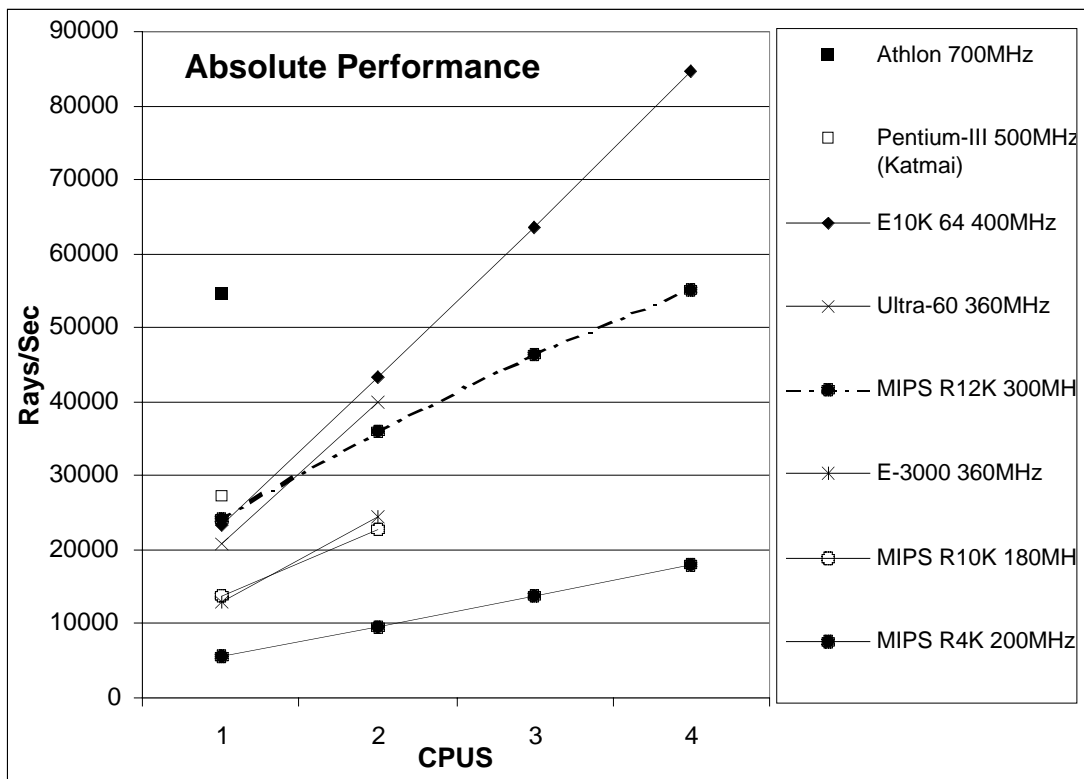Figure 4.   MHz Normalized Relative Performance.

Figure 5. Absolute Performance of one to four CPUs.

R12K system processors to equal the Athlon performance.

The multi-CPU systems are all more expensive than the Athlon and provide, at most, a small increase in performance when running our application even in parallel mode. The price/performance ratio for current systems is not as attractive as it has been historically.

Server Results

Finally, we examined performance on systems with large numbers of processors. Historically, this category has been filled with systems ranging from multi-CPU servers from Sun and SGI to Cray-class supercomputers. Today, Sun and SGI remain the two key providers of this class of machine. We were fortunate to obtain access to large-scale implementations of both architectures for testing. Testing was performed on a 64-CPU Sun Enterprise 10000, a 64-CPU SGI Origin 2000, and for comparison purposes, a 20-CPU SGI Challenge.

Sun Enterprise 10000

The Sun/Sparc machine tested was a 64-processor Enterprise 10000 with 400 MHz UltraSparc-IIi CPUs and 64 GB of main memory. The E10K system has a uniform memory architecture (UMA), which behaves as a single main memory that has a uniform access time from each processor [HENN96]. The results in

Figure 6 show a constant performance improvement with the addition of CPUs to the benchmark. This demonstrates the highly parallel nature of our raytracing application.

SGI Origin 2000

The SGI/MIPS machine tested was a 64-processor Origin 2000 with 300MHz R12K CPUs. The O2K is a non-uniform memory architecture (NUMA) machine. In this architecture, memory is distributed with each CPU. These physically separate memories can be addressed as one logically shared address space.

This means that any processor can make a memory reference to any memory location. Access time depends on the location of a data word in memory [HENN96]. The results in Figure 7 show that performance does not increase uniformly with the number of CPUs.

Three of the benchmark models show a definite ceiling in performance. The effect of this ceiling is apparent between 10 and 12 CPUs. All of the models show a great deal of variability in runtime between experimental runs. In addition, there is performance variability between similar numbers of CPUs. The shape of this graph can be explained by a detailed understanding of the Origin 2000 system architecture.
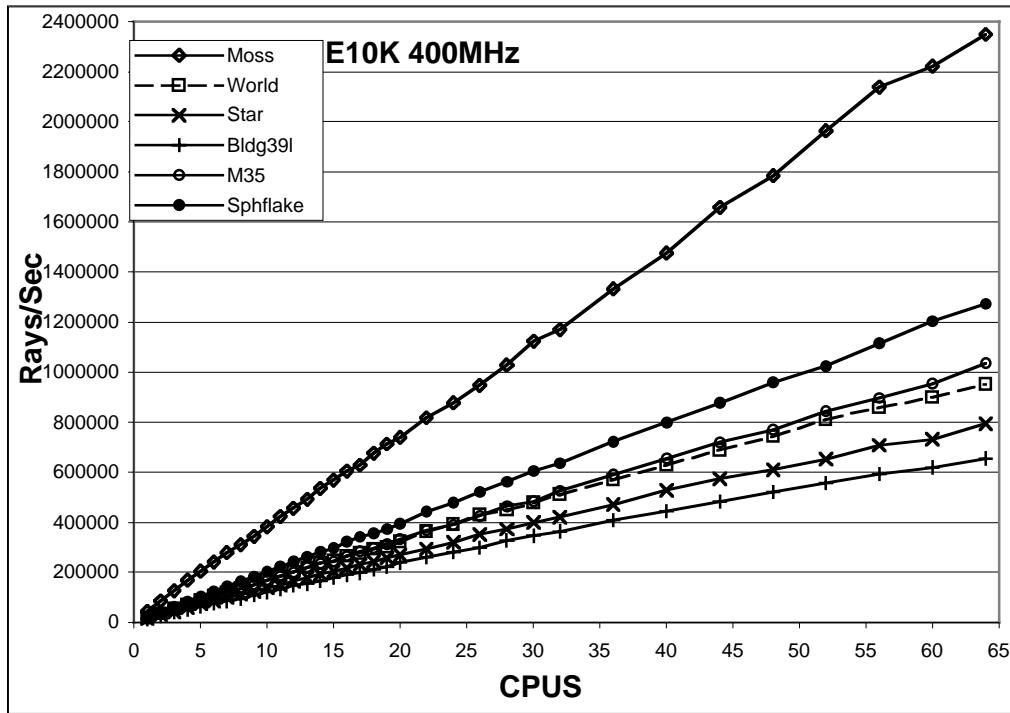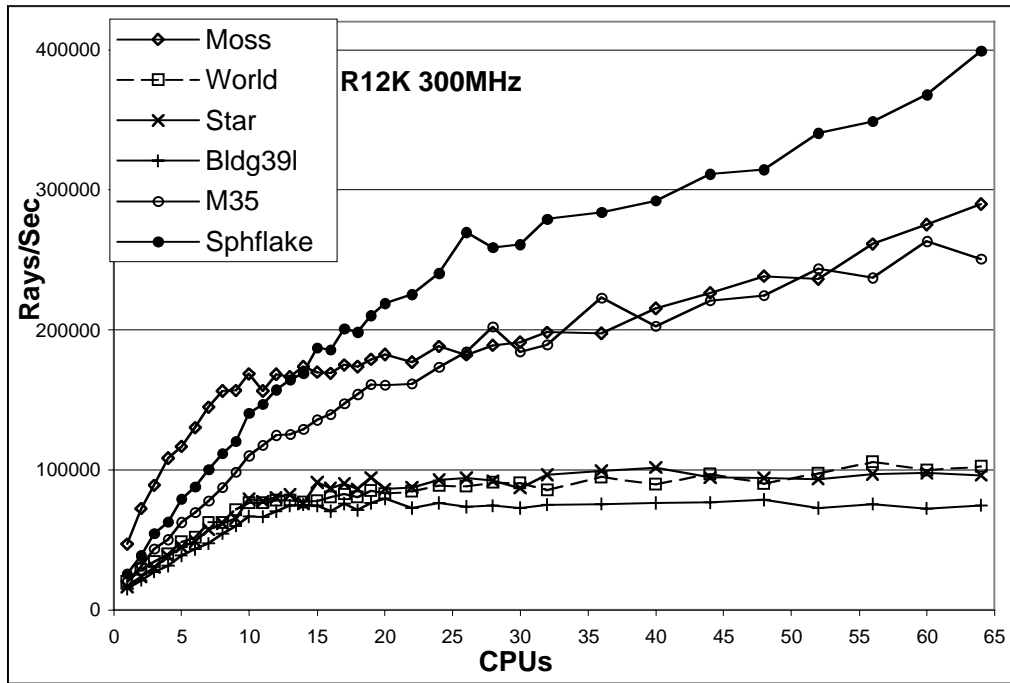
Figure 6. Sun Enterprise 10000.
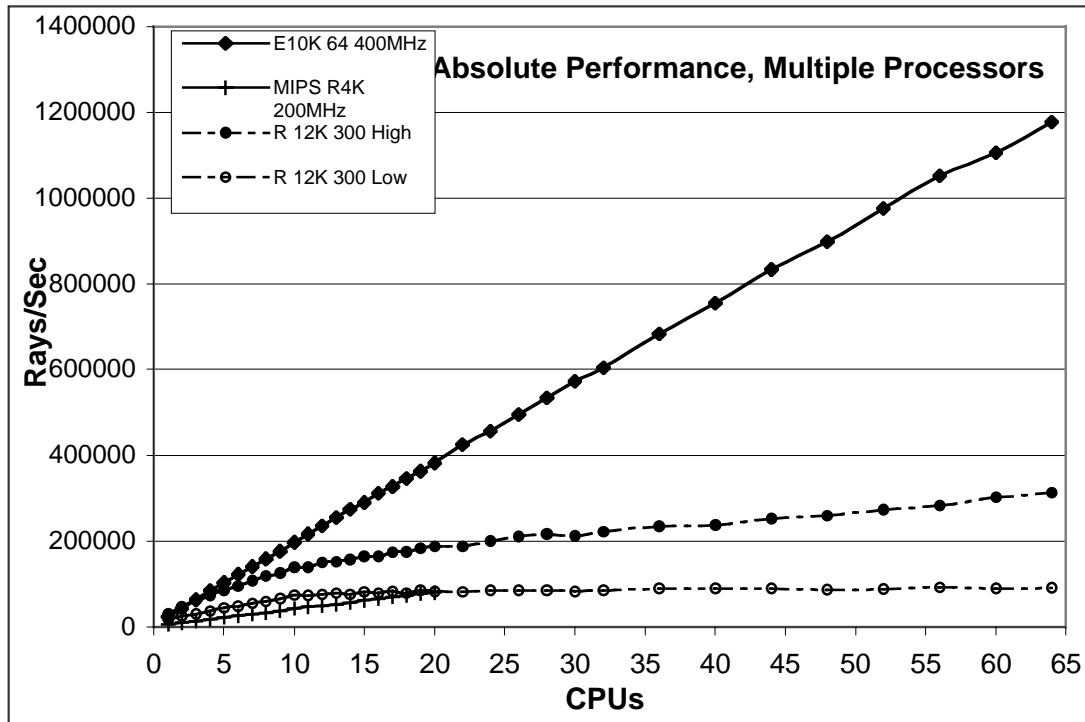

Figure 7.   SGI Origin 2000.

Figure 8. Server-Class Machines.

In this system, each processor module contains two CPUs, some local memory, and two remote memory access ports. Up to six CPUs can be interconnected without a router module for the DSM crossbar switch. Up to eight CPUs can be placed in a single backplane. To accommodate more than eight CPUs, multiple backplanes are interconnected through their memory bus. Accesses from one CPU module to another are satisfied in the order of the bus distance to the CPU module containing the desired data. Thus, when large numbers of CPUs contend for common data, CPUs in a cluster neighborhood can starve remote CPUs of access. The great variability of performance in higher numbers of CPUs can be explained by the distribution of the computation across the available nodes of the system. Jobs appear to be randomly distributed across the available CPUs. Furthermore, during execution, the operating system scheduling algorithm may move a job to a new CPU without any apparent reason.

SGI Challenge

Unlike its successor, the SGI Challenge machine has a UMA. The 200 MHz R4400 CPUs of this machine lack many of the features of its later brethren. For example, it is not super-scalar, and does not offer speculative execution. Figure 8 shows the results of all server-class machines plotted together.

For the Origin 2000, two separate lines have been plotted. This reflects the bimodal distribution of performance. Note that the R4400 CPU exhibits the same properties as the E10K when running the same code as the Origin 2000.

Conclusions

Examining current architectures in detail reveals relatively minor differences between the various CPU offerings. Current offerings in all architecture families tested employ many of the same techniques for achieving improved performance. These include pipelined super-scalar execution, branch prediction, and speculative execution. Memory architecture and clock speeds appear to determine differences in performance level. Performance increases in Sparc and MIPS architectures have not kept pace with those using the Intel instruction set.

Clearly, the boundary between Desktop- and Workstation-class machines has been completely blurred. In the near term, these roles are best filled by systems using the Athlon or similar processors.

For large-scale servers, Sun's Enterprise 10000 systems are the authors' clear choice. While the performance of individual processors of both E10K and Origin 2000 systems are similar, better efficiency

is observed when adding CPUs to the Sun system than the SGI. This is due, in large part, to the different memory architectures.

Examining Figure 4, it appears that the Athlon and MIPS CPUs perform a similar amount of work per clock cycle. We can speculate that if MIPS could increase the clock rate at which its CPUs currently perform, performance would rival that of AMD's Athlon. Likewise, the Sparc architects appear to have redesigned their CPU to do less work per clock cycle. This is a measure that might have been taken in preparation for dramatically increasing clock speed.

The BRL-CAD™ benchmark is represented as a "black-box" type of benchmark. The models represent a variety of increasingly complex geometry. It is expected that performance would vary with complexity; however, our tests show that there is additional variability that the designers of the suite did not report. This variability becomes apparent on the Origin 2000.

Future Work

At the time these tests were performed, it was not possible to obtain access to systems based either on Intel's Coppermine Pentium-III CPU or the PowerPC processor used by Apple in its products. Both of these have received favorable performance reviews; therefore, it would be desirable to include data from testing such platforms in our analysis.

A moderate amount of effort was applied to identify any minor changes to the software that might improve the performance on the Origin 2000. Such efforts, however, proved futile. When presented with our preliminary findings, representatives of SGI expressed interest in examining the applications code and were optimistic that they could substantially improve the performance of the application on their systems. In addition, we are about to receive one of SGI's new SN1 architecture systems. It would be desirable to perform this series of tests on this new system. Based upon architectural changes in this generation of system, performance may be somewhat improved.

The benchmark was originally developed using a VAX 11/780 as the reference machine. The typical runtime for computing a single reference image is reported at approximately 3 hours. Tests performed on a 64 CPU Origin 2000 took less than 4 seconds. This duration is decreasing to the point at which viable statistics can no longer be gathered. In the near future, the benchmark should be updated to meet the rigorous demands of today's faster processor systems. This will extend runtimes to the point where valid statistical analysis can continue to be performed.

References

[HENN96]   J. L. Hennesy and D. Patterson, "Computer Architecture: A Quantitative Approach" Morgan Kaufman Publishers, Inc. San Francisco, CA, 1996

[MUUSS88]   M. Muuss, P. Dykstra, K. Applin, G. Moss, P. Stay, C. Kennedy "Solid Modeling System and Ray-Tracing Benchmark" BRL-CAD™ Release 4.0 Volume V, U. S. Army Research Laboratory, APG, MD 21005-5066, 1991

Bibliography

Intel Corp., "Intel Architecture Software Developers Manual, Vol. 1 Basic Architecture," URL: http://developer.intel.com/design/pentiumii/manuals/243190.htm, 1999.

Intel Corp., "Intel Architecture Software Developers Manual, Vol. 2 Instruction Set Reference," URL: http://developer.intel.com/design/pentiumii/manuals/243191.htm, 1999.

Intel Corp., "Intel Architecture Software Developers Manual, Vol. 3 System Programming Guide," URL: http://developer.intel.com/design/pentiumii/manuals/243192.htm, 1999

Standard Performance Evaluation Corp., "SPEC Benchmark Results Database," URL: http://www.spec.org/results.html

Gee, et al., "Cache Performance of the SPEC92 Benchmark Suite," IEEE Micro, Vol. 13, no. 4, Aug. 1993, pp. 17-27

S. Dasgupta and E. Servan-Schreiber, "Cache Behavior of the SPEC95 Benchmark Suite," URL: http://http.cs.berkeley.edu/~dasgupta/paper/rep/rep.html, 1996

A. Lebeck and D. Wood, "Cache Profiling and the SPEC Benchmarks: A Case Study," IEEE Computer, vol. 27, no. 10, Oct. 1994, pp. 15-26

Sun Microsystems Inc., "The Ultra-SPARC Users Manual," 1997, URL: http://www.sun.com/microelectronics/manuals/ultrasparc/802-7220-02.pdf

Sun Microsystems Inc., "The Ultra-SPARC IIi Users Manual," 1997, URL: http://www.sun.com/microelectronics/manuals/805-0087.pdf

Sun Microsystems Inc., "The Ultra-SPARC IIi Users Manual Addendum," 1997, URL: http://www.sun.com/microelectronics/UltraSPRAC-IIi/dos/ultrasparc-IIi_add.pdf

Sun Microsystems Inc., "The Super-SPARC II Users Manual," 1994, URL: http://www.sun.com/microelectronics/manuals/STP1021UG.pdf

Sun Microsystems Inc., "The Ultra-SPARC IIi Data Sheet," 1999

Sun Microsystems Inc., "The Ultra-SPARC II Data Sheet," 1999

MIPS Technologies, Inc., "MIPS R4400 Microprocessor Product Information," 1996

MIPS Technologies, Inc., "MIPS R10000 Microprocessor Product Overview," 1994

MIPS Technologies, Inc., "MIPS R10000 Microprocessor User's Manual Version 2.0," 1996

MIPS Technologies, Inc., "MIPS R5000 Microprocessor Technical Backgrounder," 1996

MIPS Technologies, Inc., "Differences between R10000 and R4400," 1995

MIPS Technologies, Inc., "MIPS R12000 Microprocessor Product Overview"

MIPS Technologies, Inc., "MIPS R10000 Technical Brief"

MIPS Technologies, Inc., "MIPS R12000 Microprocessor User Manual"