
AlphaServer 4100 Performance Characterization

The AlphaServer 4100 is the newest four-processor symmetric multiprocessing addition to DIGITAL's line of midrange Alpha servers. The DIGITAL AlphaServer 4100 family, which consists of models 5/300E, 5/300, and 5/400, has major platform performance advantages as compared to previous-generation Alpha platforms and leading industry midrange systems. The primary performance strengths are low memory latency, high bandwidth, low-latency I/O, and very large memory (VLM) technology. Evaluating the characteristics of both technical and commercial workloads against each family member yielded recommendations for the best application match for each model. The performance of the model with no module-level cache and the advantages of using 2- and 4-megabyte module-level caches are quantified. The profiles based on the built-in performance monitors are used to evaluate cycles per instruction, stall time, multiple-issuing benefits, instruction frequencies, and the effect of cache misses, branch mispredictions, and replay traps. The authors propose a time allocation-based model for evaluating the performance effects of various stall components and for predicting future performance trends.

The AlphaServer 4100 is DIGITAL's latest four-processor symmetric multiprocessing (SMP) midrange Alpha server. This paper characterizes the performance of the AlphaServer 4100 family, which consists of three models:¹⁻⁵

1. AlphaServer 4100 model 5/300E, which has up to four 300-megahertz (MHz) Alpha 21164 microprocessors, each without a module-level, third-level, write-back cache (B-cache) (a design referred to as *uncached* in this paper)
2. AlphaServer 4100 model 5/300, which has up to four 300-MHz Alpha 21164 microprocessors, each with a 2-megabyte (MB) B-cache
3. AlphaServer 4100 model 5/400, which has up to four 400-MHz Alpha 21164 microprocessors, each with a 4-MB B-cache

The performance analysis undertaken examined a number of workloads with different characteristics, including the SPEC95 benchmark suites (floating-point and integer), the LINPACK benchmark, AIM Suite VII (UNIX multiuser benchmark), the TPC-C transaction processing benchmark, image rendering, and memory latency and bandwidth tests.⁶⁻¹⁵ Note that both commercial (AIM and TPC-C) and technical/scientific (SPEC, LINPACK, and image rendering) classes of workloads were included in this analysis.

The results of the analysis indicate that the major AlphaServer 4100 performance advantages result from the following server features:

- Significantly higher bandwidth (up to 2.6 times) and lower latency compared to the previous-generation midrange AlphaServer platforms and leading industry midrange systems. These improvements benefit the large, multistream applications that do not fit in the B-cache. For example, the AlphaServer 4100 5/300 is 30 to 60 percent faster than the HP 9000 K420 server in the memory-intensive workloads from the SPECfp95 benchmark suite. (Note that all competitive performance data presented in this paper is valid as

of the submission of this paper in July 1996. The references cited refer the reader to the literature and the appropriate Web sites for the latest performance information.)

- An expanded very large memory (VLM). The maximum memory size increased from 2 gigabytes (GB) to 8 GB without sacrificing CPU slots. This increase in memory size benefits primarily the commercial, multistream applications. For example, the AlphaServer 4100 5/300 server achieves approximately twice the throughput of the Compaq ProLiant 4500 server and 1.4 times the throughput of the AlphaServer 2100 on the AIM Suite VII benchmark tests.
- A 4-MB B-cache and a clock speed of 400 MHz in the AlphaServer 4100 5/400 system. The larger B-cache size and 33 percent faster clock resulted in a 30 to 40 percent performance improvement over the AlphaServer 4100 5/300 system.

The performance improvement from the larger B-cache increases with the number of CPUs. For example, the AlphaServer 4100 5/300 system with its 2-MB B-cache design performs 5 to 20 percent faster with one CPU and 30 to 50 percent faster with four CPUs than the uncached 5/300E system. The majority of workloads included in this analysis benefit from the B-cache; however, the uncached system outperforms the cached implementation by 10 to 20 percent for large applications that do not fit in the 2-MB B-cache.

The performance counter profiles, based on the built-in hardware monitors, indicate that the majority of issuing time is spent on single and dual issuing and that a small number of floating-point workloads take advantage of triple and quad issuing. The load/store instructions make up 30 to 40 percent of all instructions issued. The stall time associated with waiting for data that missed in the various levels of cache hierarchy accounts for the most significant portion of the time the server spends processing commercial workloads.

Memory Latency

Memory latency and bandwidth have been recognized as important performance factors in the early Alpha-based implementations.^{16,17} Since CPU speed is increasing at a much higher rate than memory speed, the “memory wall” limitation is expected to become even more important in the future. Therefore, reducing memory latency and increasing bandwidth have been major design goals for the AlphaServer 4100 platform.¹ The AlphaServer 4100 achieved the lowest memory latency of all DIGITAL products based on

the Alpha 21164 microprocessor and all multiprocessor products by leading industry vendors. The major benefits come from the simpler interface, the use of synchronous dynamic random-access memory (DRAM) chips (i.e., synchronous memory), and the lower fill time.^{1,2} Figure 1 shows the measured memory load latency using the Imbench benchmark with a 512-byte stride.¹⁰ In this benchmark, each load depends on the result from the previous load, and therefore latency is not a good measure of performance for systems that can have multiple outstanding loads. (AlphaServer 4100 systems can have up to two outstanding requests per CPU on the bus.) The Imbench benchmark data indicates that the AlphaServer 4100 has the lowest memory latency of all industry-leading reduced-instruction set computing (RISC) vendors’ servers.

As shown in Figure 2, using a slightly different workload where there is no dependency between consecutive loads, the AlphaServer 4100 achieves even lower per-load latency, since the latency for the two consecutive loads can be overlapped. The plateaus in Figure 2 show the load latency at each of the following levels of cache/memory hierarchy: 8-kilobyte (KB) on-chip data cache (D-cache), 96-KB on-chip secondary instruction/data cache (S-cache), 2- and 4-MB off-chip B-caches (except for model 5/300E), and memory. The uncached AlphaServer 4100 5/300E achieves an 85 percent lower memory load latency than the previous-generation AlphaServer 2100. The AlphaServer 4100 5/300, with its 2-MB B-cache, increases memory latency 30 percent for load operations and 6 percent for store operations compared to the uncached 5/300E system because of the time spent checking for data in the B-cache. The synchronous memory shows one cycle lower latency than the asynchronous extended data out (EDO) DRAM (i.e., asynchronous memory), which results in 9 percent faster load operations and 5 percent faster store operations. Note that the cached AlphaServer 4100 and AlphaServer 8200 systems, which have the same clock speeds of 300 MHz, achieve comparable B-cache latency, while the memory latency for all AlphaServer 4100 systems is significantly lower than on both the AlphaServer 8200 and the AlphaServer 2100 systems. The latency to the B-cache in this test is lower on the AlphaServer 2100 than on the other AlphaServer systems due to 32-byte blocks (compared to 64-byte blocks in the 4100 and 8200 systems). Although not shown in this test, many applications can benefit from the larger cache block size. The 400-MHz AlphaServer 4100 system uses a 33 percent faster CPU and achieves 11 percent reduction in B-cache and memory latency compared to the 300-MHz AlphaServer 4100 system.

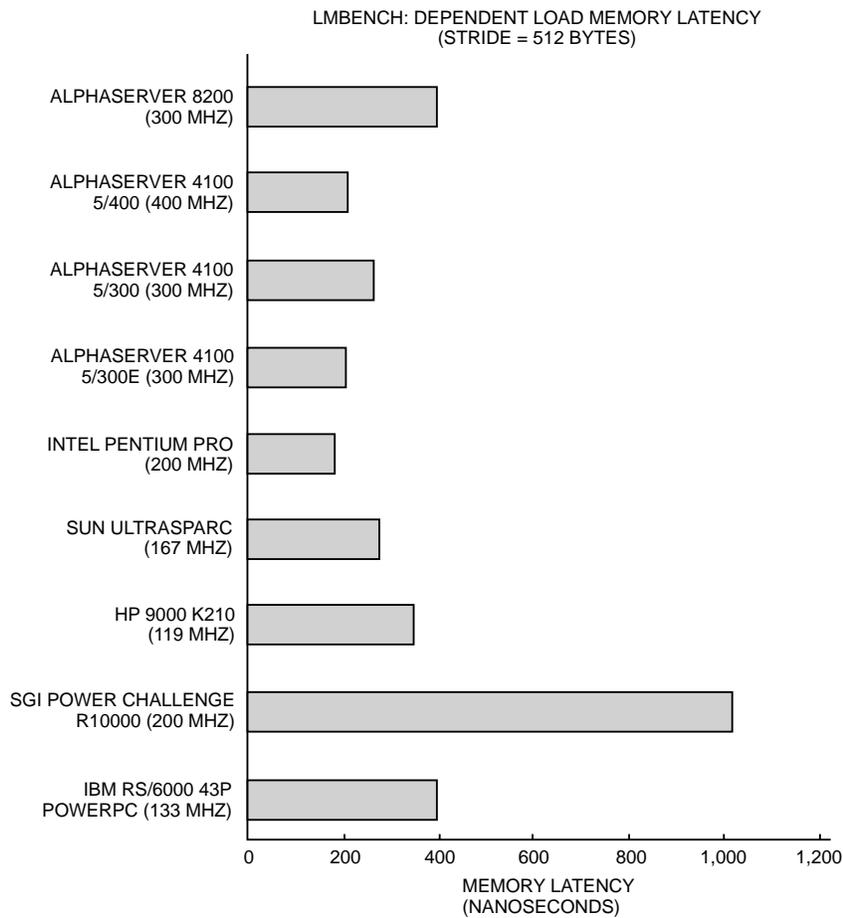


Figure 1
Imbench Benchmark Test Results Showing Memory Latency for Dependent Loads

Memory Bandwidth

The AlphaServer 4100 system bus achieves a peak bandwidth of 1.06 gigabytes per second (GB/s). The STREAM McCalpin benchmark measures sustainable memory bandwidth in megabytes per second (MB/s) across four vector kernels: Copy, Scale, Sum, and SAXPY.¹¹ Figure 3 shows measured memory bandwidth using the Copy kernel from the STREAM benchmark. Note that the STREAM bandwidth is 33 percent lower than the actual bandwidth observed on the AlphaServer 4100 bus because the bus data cycles are allocated for three transactions: read source, read destination, and write destination. The AlphaServer 4100 shows the best memory bandwidth of all multiprocessor platforms designed to support up to four CPUs. The platforms designed to support more than four CPUs (i.e., the AlphaServer 8400, the Silicon Graphics POWER CHALLENGE R10000, and the Sun Ultra Enterprise 6000 systems) show a higher bandwidth for four CPUs than the AlphaServer 4100. The STREAM bandwidth on the AlphaServer 4100 5/300 is 2.2 times higher than on the previous-generation AlphaServer 2100 5/250 (2.6 times higher

with the AlphaServer 4100 5/400). The uncached AlphaServer 4100 model shows 22 percent higher memory bandwidth than the cached model 5/300.

The AlphaServer 4100 memory bandwidth improvement from synchronous memory compared to EDO ranges from 8 to 12 percent. The synchronous memory benefit increases with the number of CPUs, as shown in Table 1.

Low memory latency and high bandwidth have a significant effect on the performance of workloads that do not fit in 2- to 4-MB B-caches. For example, the majority of the SPECfp95 benchmarks do not fit in the 2-MB cache. (Figure 20, which appears later in this paper, shows the cache misses.) The SPECfp95 performance comparison presented in Figure 4 shows that the uncached AlphaServer 4100 5/300E system outperforms the 2-MB B-cache model 5/300 in the benchmarks with the highest number of B-cache misses (tomcatv, swim, applu, and hydro2d). The performance of the uncached model 5/300E is comparable to that of the 4-MB B-cache model 5/400 for the swim benchmark. However, the benchmarks that fit better in the 4-MB cache (apsi and wave5) run significantly slower on the 5/300E than on the 5/400.

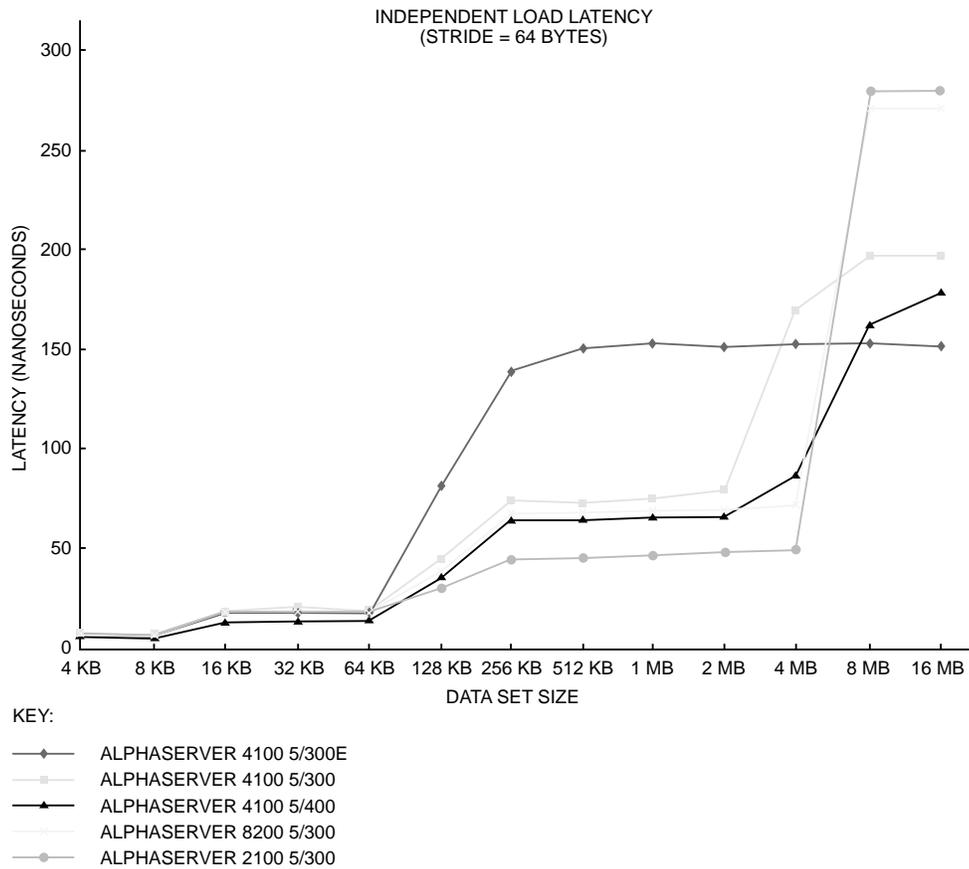


Figure 2
Cache/Memory Latency for Independent Loads

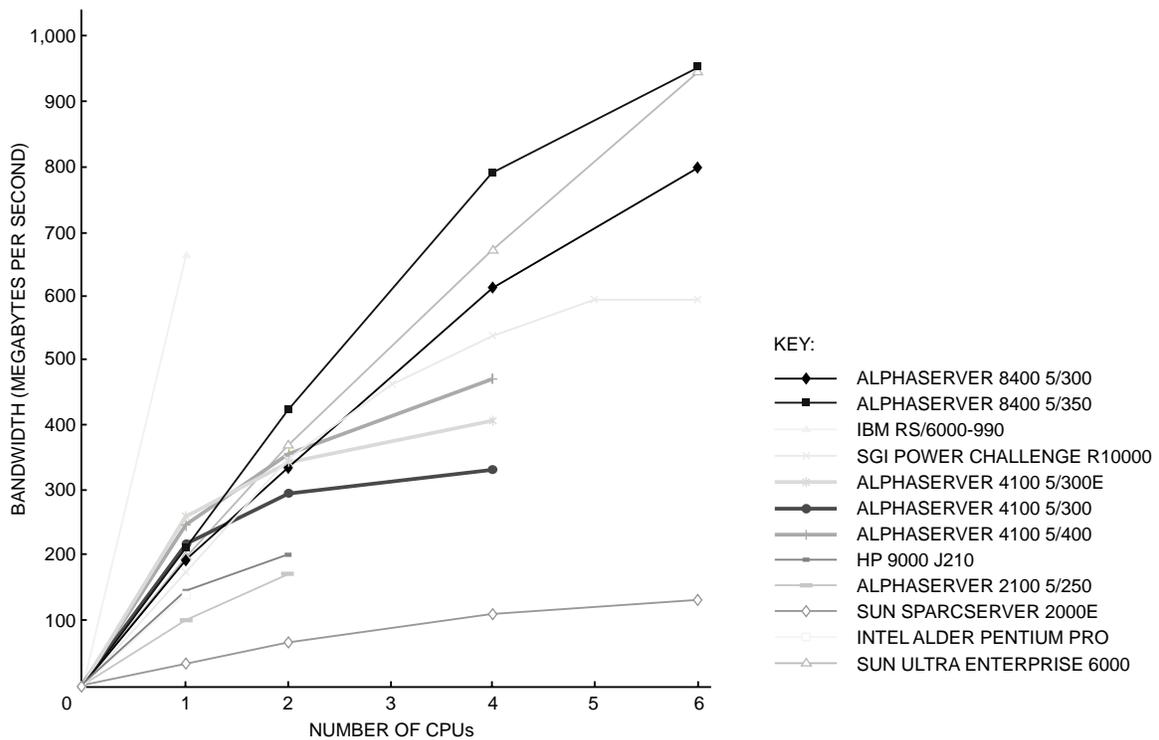


Figure 3
STREAM McCalpin Memory Copy Bandwidth Comparison

Table 1
Bandwidth Improvement from Synchronous Memory to Asynchronous Memory

	Number of CPUs			
	1	2	3	4
Bandwidth improvement	8%	8%	9%	12%

Figure 4 shows that the AlphaServer 4100 5/300 system has a significant (up to two times) performance advantage over the previous-generation AlphaServer 2100 system in the SPECfp95 benchmark tests with the highest number of B-cache misses. The SPECfp95 tests indicate that the 300-MHz AlphaServer 4100 is more than 50 percent faster than the HP 9000 K420 server, and the 400-MHz AlphaServer 4100 is twice as fast as the HP 9000 K420 in the SPECfp95 benchmarks that stress the memory subsystem.

SPEC95 Benchmarks

The SPEC95 benchmarks provide a measure of processor, memory hierarchy, and compiler performance. These benchmarks do not stress graphics, network, or I/O performance. The integer SPEC95 suite

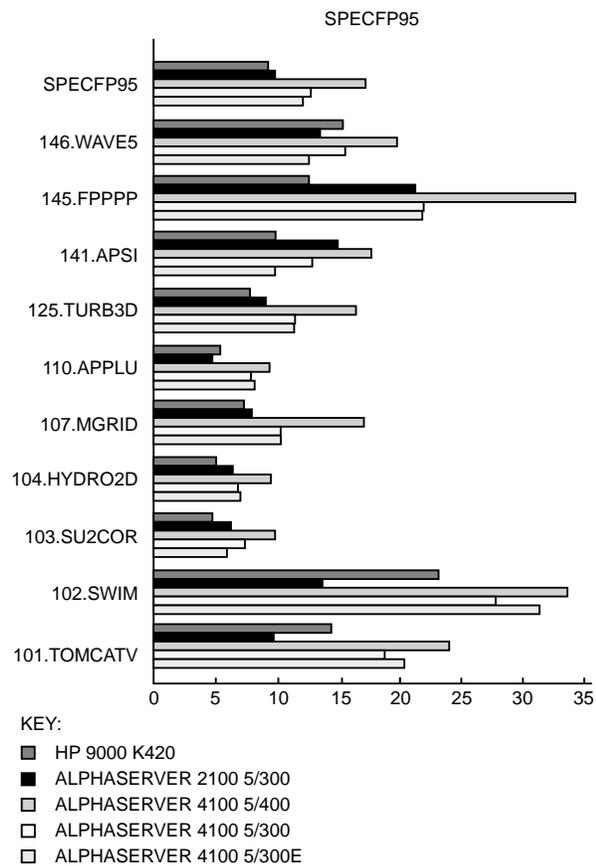


Figure 4
SPECfp95 Benchmarks Performance Comparison

(CINT95) contains eight compute-intensive integer benchmarks written in C and includes the benchmarks shown in Table 2.^{6,12}

The floating-point SPEC95 suite (CFP95) contains 10 compute-intensive floating-point benchmarks written in FORTRAN and includes the benchmarks shown in Table 3.^{6,12}

The SPEC Homogeneous Capacity Method (SPEC95 rate) measures how fast an SMP system can perform multiple CINT95 or CFP95 copies (tasks). The SPEC95 rate metric measures the throughput of the system running a number of tasks and is used for evaluating multiprocessor system performance.

Table 2
CINT95 Benchmarks (SPECint95)

Benchmark	Description
099.go	Artificial intelligence, plays the game of Go
124.m88ksim	A Motorola 88100 microprocessor simulator
126.gcc	A GNU C compiler that generates SPARC assembly code
129.compress	A program that compresses large text files (about 16 MB)
130.li	A LISP interpreter
132.jpeg	A program that compresses/decompresses an image
134.perl	A Perl interpreter that performs text and numeric manipulations
147.vortex	A database program that builds and manipulates three interrelational databases

Table 3
CFP95 Benchmarks (SPECfp95)

Benchmark	Description
101.tomcatv	A fluid dynamics mesh generation program
102.swim	A weather prediction shallow water model
103.su2cor	A quantum physics particle mass computation (Monte Carlo)
104.hydro2d	An astrophysics hydrodynamical Navier-Stokes equation
107.mgrid	A multigrid solver in a 3-D potential field (electromagnetism)
110.applu	Parabolic/elliptic partial differential equations (fluid dynamics)
125.turb3d	A program that simulates turbulence in a cube
141.apsi	A program that simulates temperature, wind, velocity, and pollutants (weather prediction)
145.fpppp	A quantum chemistry program that performs multielectron derivatives
146.wave5	A solver of Maxwell's equations on a Cartesian mesh (electromagnetics)

Figure 5 compares the SPEC95 performance of the AlphaServer 4100 systems to that of the other industry-leading vendors using published results as of July 1996. Figure 6 shows the same comparison in the multistream SPEC95 rates.¹² Note that all the SPEC95 comparisons in this paper are based on the peak results that include extensive compiler optimizations.¹² Figure 5 indicates that even the uncached AlphaServer 4100 5/300E performs better than the HP 9000 K420 system, and the AlphaServer 4100 5/400 shows approximately a two times performance advantage over the HP system. The AlphaServer 4100 5/300 SPECint95 performance exceeds that of the Intel Pentium Pro system, and the AlphaServer 4100 5/300 SPECfp95 performance is double that of the Pentium Pro. The AlphaServer 4100 5/400 is 1.5 times (SPECint95) and 2.5 times (SPECfp95) faster than the Pentium Pro system. The multiple-processor SPECfp95 on the AlphaServer 4100 is obtained by decomposing benchmarks using the KAP preprocessor from Kuck & Associates. Note that the cached four-CPU AlphaServer 4100 5/300 outperforms the Sun Ultra Enterprise 3000 with six CPUs in the SPECfp95 parallel test. The performance benefit of the cached versus the uncached AlphaServer 4100 is greater in multiprocessor configurations than in uni-processor configurations.

SPEC95 Multistream Performance Scaling

Figures 7 and 8 show SPEC95 multistream performance as the number of CPUs increases. The SMP scaling on the AlphaServer 4100 is comparable to that

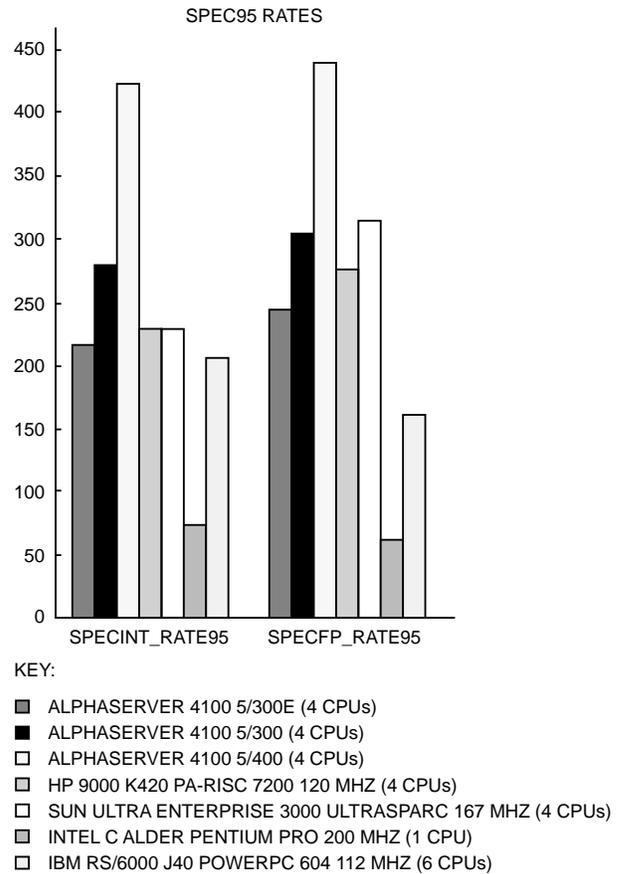


Figure 6
SPEC95 Throughput Results (SPEC95 Rates)

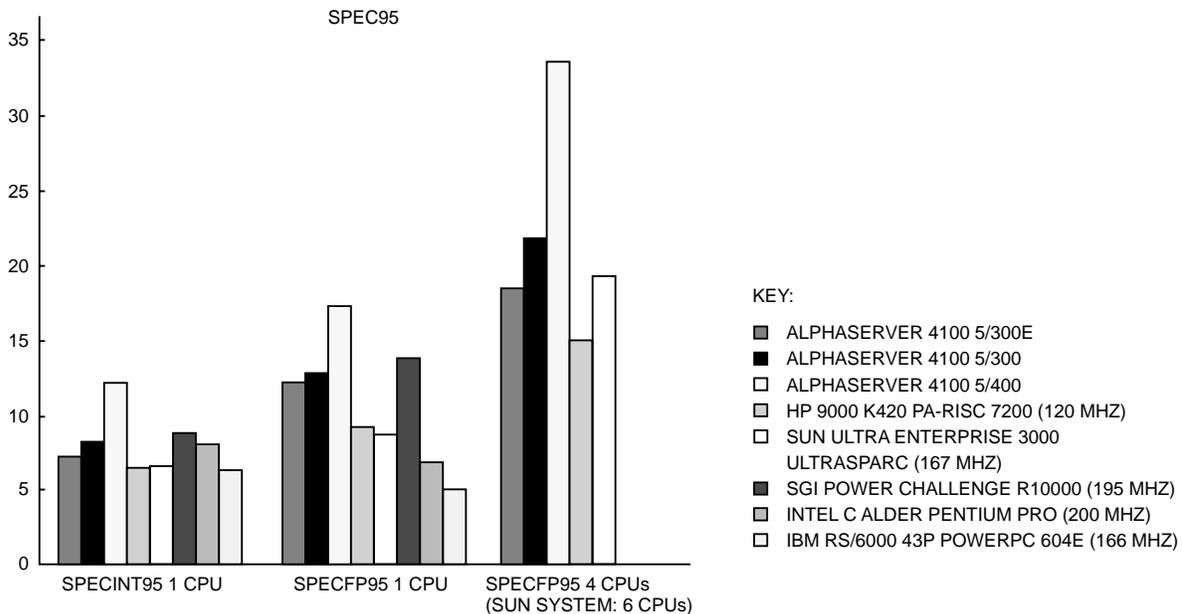


Figure 5
SPEC95 Speed Results

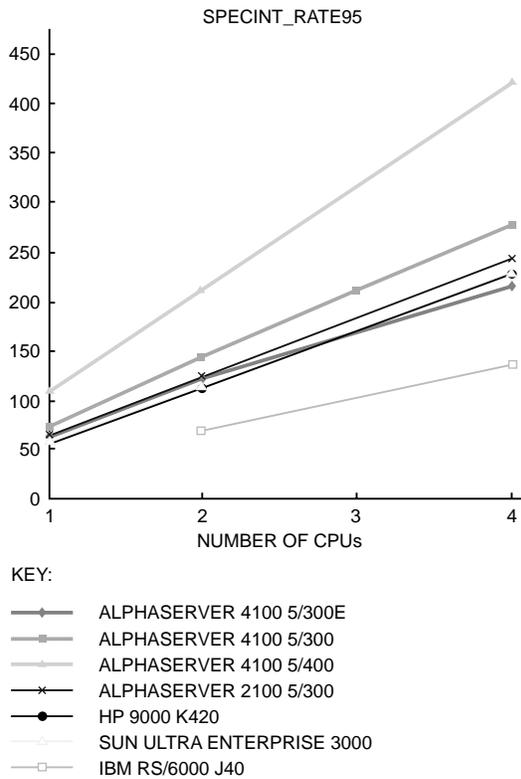


Figure 7
SPECint_rate95 Performance Scaling

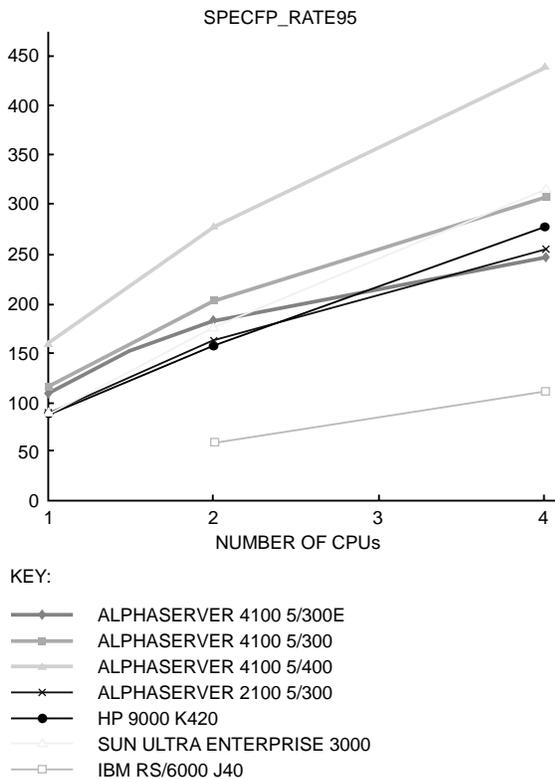


Figure 8
SPECfp_rate95 Performance Scaling

on the AlphaServer 2100 for integer workloads (that fit in the 5/300 2-MB B-cache). Note that SPECint_rate95 scales proportionally to the number of CPUs in the majority of systems, since these workloads do not stress the memory subsystem. The SMP scaling in SPECfp_rate95 is lower, since the majority of these workloads do not fit in 1- to 4-MB caches.

In the majority of applications, the AlphaServer 4100 5/300 and 5/400 systems improve SMP scaling compared to the uncached AlphaServer 4100 5/300E by reducing the bus traffic (from fewer B-cache misses) and by taking advantage of the duplicate tag store (DTAG) to reduce the number of S-cache probes. The cached 5/300 scaling, however, is lower than the uncached 5/300E scaling in memory bandwidth-intensive applications (e.g., tomcatv and swim). The analysis of traces collected by the logic analyzer that monitors system bus traffic indicates that the lower scaling is caused by (1) SetDirty overhead, where SetDirty is a cache coherency operation used to mark data as modified in the initiating CPU's cache; (2) stall cycles on the memory bus; and (3) memory bank conflicts.^{2,3}

Symmetric Multiprocessing Performance Scaling for Parallel Workloads

Parallel workloads have higher data sharing and lower memory bandwidth requirements than multistream workloads. As shown in Figures 9 and 10, the AlphaServer 4100 models with module-level caches improve the SMP scaling compared to the uncached AlphaServer 4100 model in the LINPACK 1000 × 1000 (million floating-point operations per second [MFLOPS]) and the parallel SPECfp95 benchmarks that benefit from 2- and 4-MB B-caches. Figure 9 indicates that the AlphaServer 4100 5/400 outperforms the SGI Origin 2000 system in the LINPACK 1000 × 1000 benchmark by 40 percent. Figure 10 indicates that the four-CPU AlphaServer 4100 5/400 shows better scaling than any other system in its class and outperforms the six-CPU Sun Ultra Enterprise 3000 system by more than 70 percent.

Very Large Memory Advantage: Commercial Performance

As shown in Figures 11 and 12, the AlphaServer 4100 performs well in the commercial benchmarks TPC-C and AIM Suite VII.^{13,14} In addition to the low memory and I/O latency, the AlphaServer 4100 takes advantage of the VLM design in these I/O-intensive workloads: with four CPUs, the platform can support up to 8 GB of memory compared to 1 GB of memory on the AlphaServer 2100 system with four CPUs and 2 GB with three CPUs.

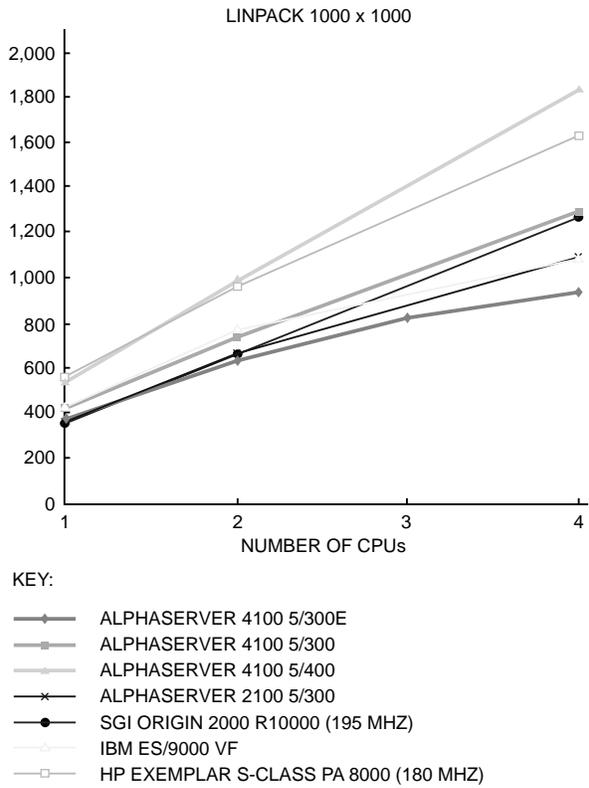


Figure 9
LINPACK 1000 × 1000 Parallel Performance Scaling

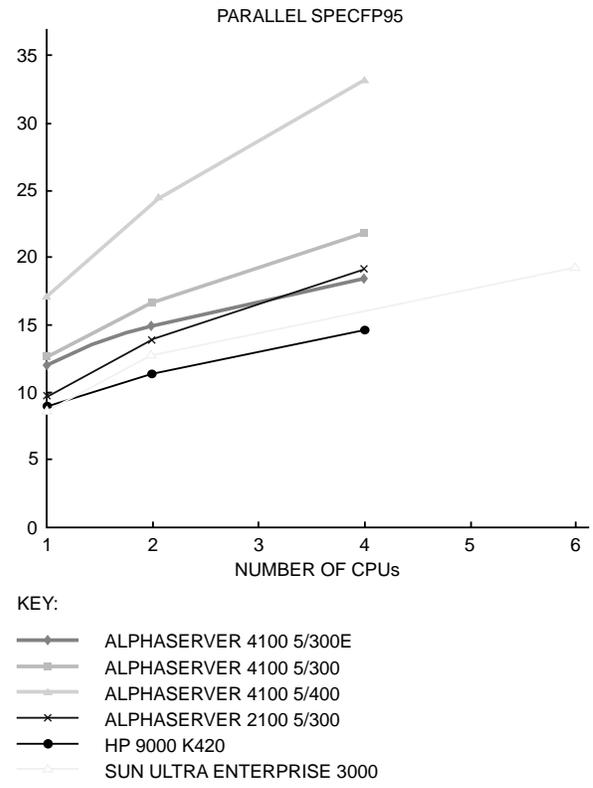


Figure 10
Parallel SPECfp95 Performance Scaling

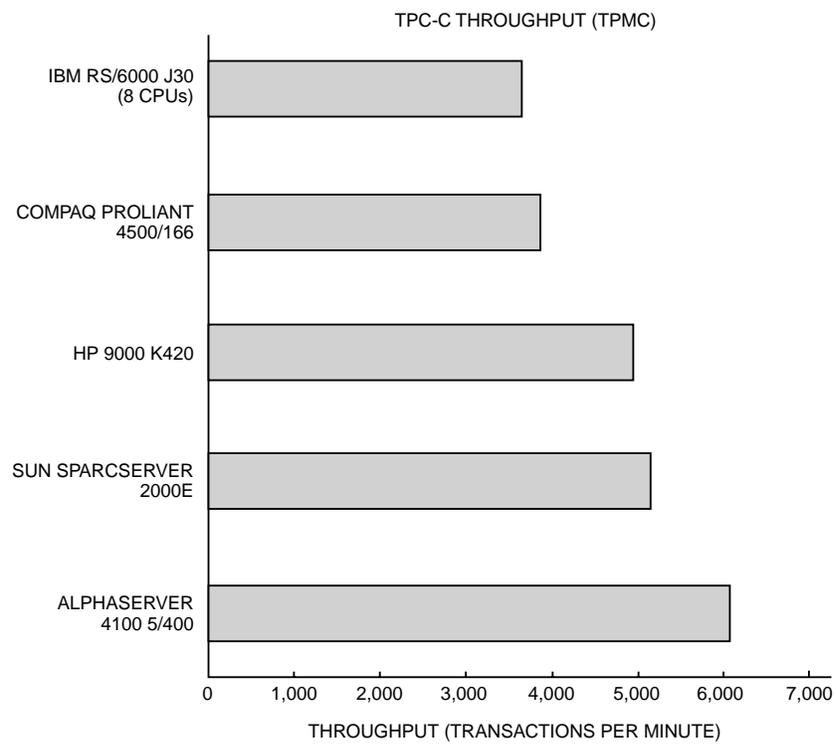
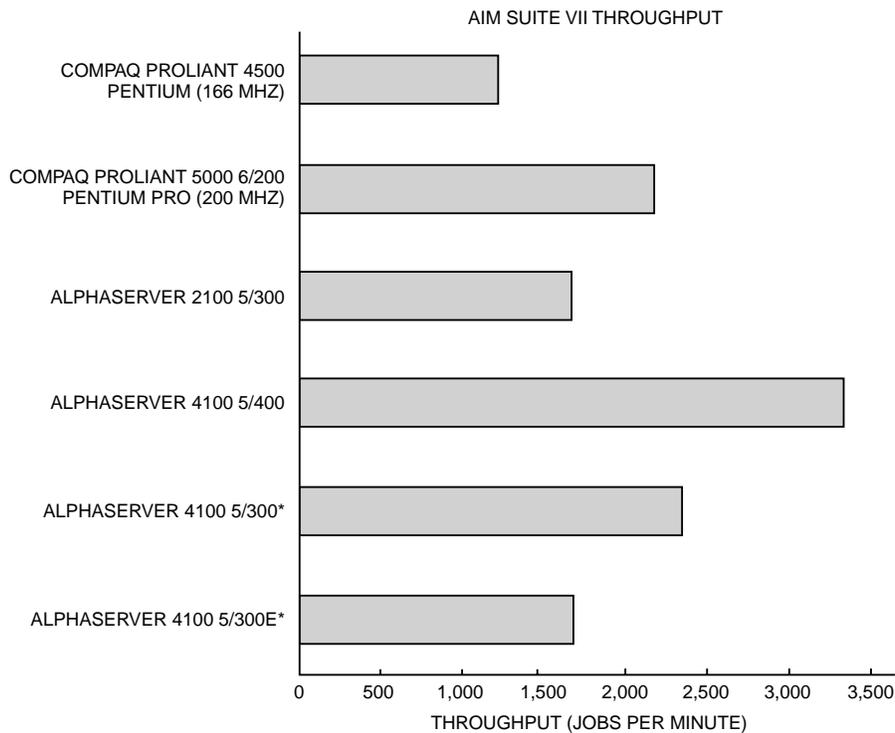


Figure 11
Transaction Processing Performance (TPC-C Using an Oracle Database)



*These internally generated results have not been AIM certified.

Figure 12
AIM Suite VII Multiuser/Shared UNIX Mix Performance

Figures 11 and 12 show the AlphaServer 4100 system's TPC-C performance (using an Oracle database) and AIM Suite VII throughput performance as compared to other industry-leading vendors. Note that the performance of the uncached AlphaServer 4100 5/300E is comparable to that of the 300-MHz AlphaServer 2100. (The AlphaServer 2100 system used in this test had three CPUs and 2 GB of memory, whereas the AlphaServer 4100 system had four CPUs and 2 GB of memory.)

With its 2-MB B-cache, the AlphaServer 4100 5/300 improves throughput by 40 percent in the AIM Suite VII benchmark tests as compared to the uncached AlphaServer 4100 5/300E. The AlphaServer 4100 5/400, with its 4-MB B-cache, benefits from its 33 percent faster clock and two times larger B-cache and provides 40 percent improvement over the AlphaServer 4100 5/300. Note that the AlphaServer 4100 5/300 and 5/300E results were obtained through internal testing and have not been AIM certified. The AlphaServer 5/400 results have AIM certification.

Compared to the best published industry AIM Suite VII performance, the AlphaServer 4100 5/300 throughput is almost twice that of the Compaq ProLiant 4500 server, and the AlphaServer 4100 5/400 throughput is more than 50 percent higher than that of the Compaq ProLiant 5000 server.¹⁴ At

the October 1996 UNIX Expo, the AlphaServer 4100 family won three AIM Hot Iron Awards: for the best performance on the Windows NT operating system (for systems priced at more than \$50,000) and for the best price/performance in two UNIX mixes—multiuser shared and file system (for systems priced at more than \$150,000).¹⁴

Cache Improvement on the AlphaServer 4100 System

Figures 13 and 14 show the percentage performance improvement provided by the 2-MB B-cache in the AlphaServer 4100 5/300 as compared to the uncached AlphaServer 4100 5/300E. Figure 13 shows the improvement across a variety of workloads; Figure 14 shows the improvement in individual SPEC95 benchmarks for one and four CPUs.

As shown in Figure 13, the 2-MB B-cache in the AlphaServer 4100 5/300 improves the performance by 5 to 20 percent for one CPU and 25 to 40 percent for four CPUs as compared to the uncached AlphaServer 4100 5/300E system. The benefits derived from having larger caches are significantly greater for four CPUs compared to one CPU, since large caches help alleviate bus traffic in multiprocessor systems.

The workloads that do not fit in the 2- to 4-MB B-cache (i.e., tomcatv, swim, applu) in Figure 14

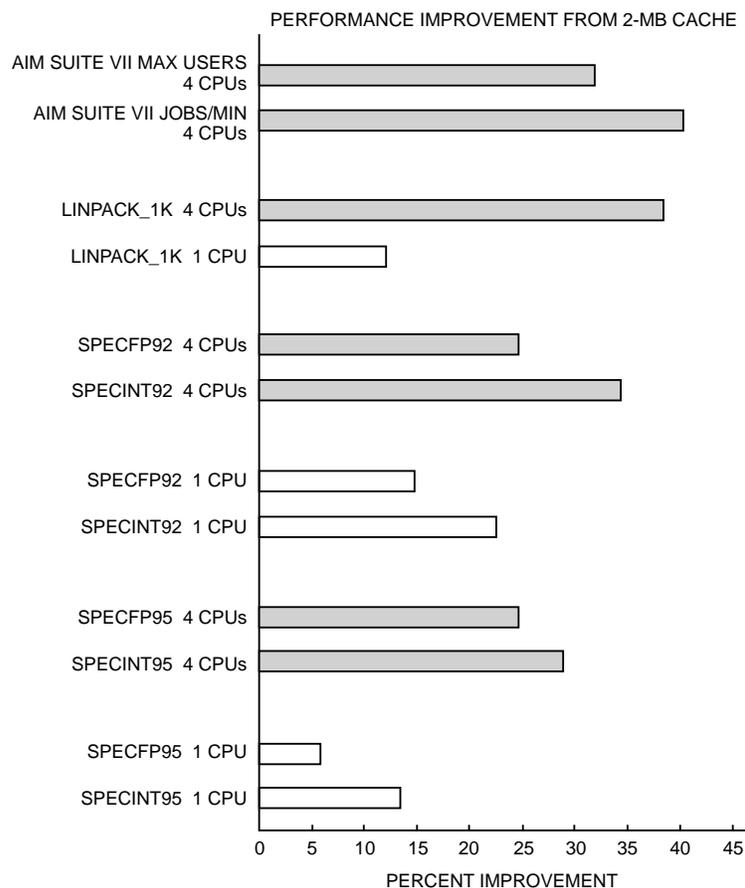


Figure 13
Performance Improvement across Various Workloads from a 2-MB B-Cache

run faster on the uncached AlphaServer 4100 than on the cached AlphaServer 4100 (up to 10 percent faster on one CPU and 20 percent faster on four CPUs) due to the overhead for probing the B-cache and the increase in SetDirty bandwidth. The majority of the other workloads benefit from larger caches.

The AlphaServer 4100 5/400 further improves the performance by increasing the size of the B-cache from 2 MB to 4 MB. In addition, the CPU clock improvement of 33 percent, B-cache improvement of 7 percent in latency and 11 percent in bandwidth, and the memory bus speed improvement of 11 percent together yield an overall 30 to 40 percent improvement in the AlphaServer 4100 model 5/400 performance as compared to that of the AlphaServer 4100 model 5/300.

Large Scientific Applications: Sparse LINPACK

The Sparse LINPACK benchmark solves a large, sparse symmetric system of linear equations using the conjugate gradient (CG) iterative method. The benchmark has three cases, each with a different type of preconditioner. Cases 1 and 2 use the incomplete

Cholesky (IC) factorization as the preconditioner, whereas Case 3 uses the diagonal preconditioner.

This workload is representative of large scientific applications that do not fit in megabyte-size caches. The workload is important in large applications, e.g., models of electrical networks, economic systems, diffusion, radiation, and elasticity. It was decomposed to run on multiprocessor systems using the KAP preprocessor.

Figure 15 shows that the uncached AlphaServer 4100 5/300E outperforms the AlphaServer 8400 by 41 percent for one CPU and by 9 percent for two CPUs because of higher delivered system bus bandwidth. However, the AlphaServer 4100 5/300E falls behind with three and four CPUs, as it does in the McCalpin memory bandwidth tests shown in Figure 3. Note that with one CPU, the 300-MHz uncached AlphaServer 4100 performs at the same level as the 400-MHz cached AlphaServer 4100 and performs 18 percent better than the 300-MHz cached AlphaServer 4100. This is an example of the type of application for which the cache diminishes the performance. The AlphaServer 4100 5/300E is a better match for this class of applications than the cached systems.

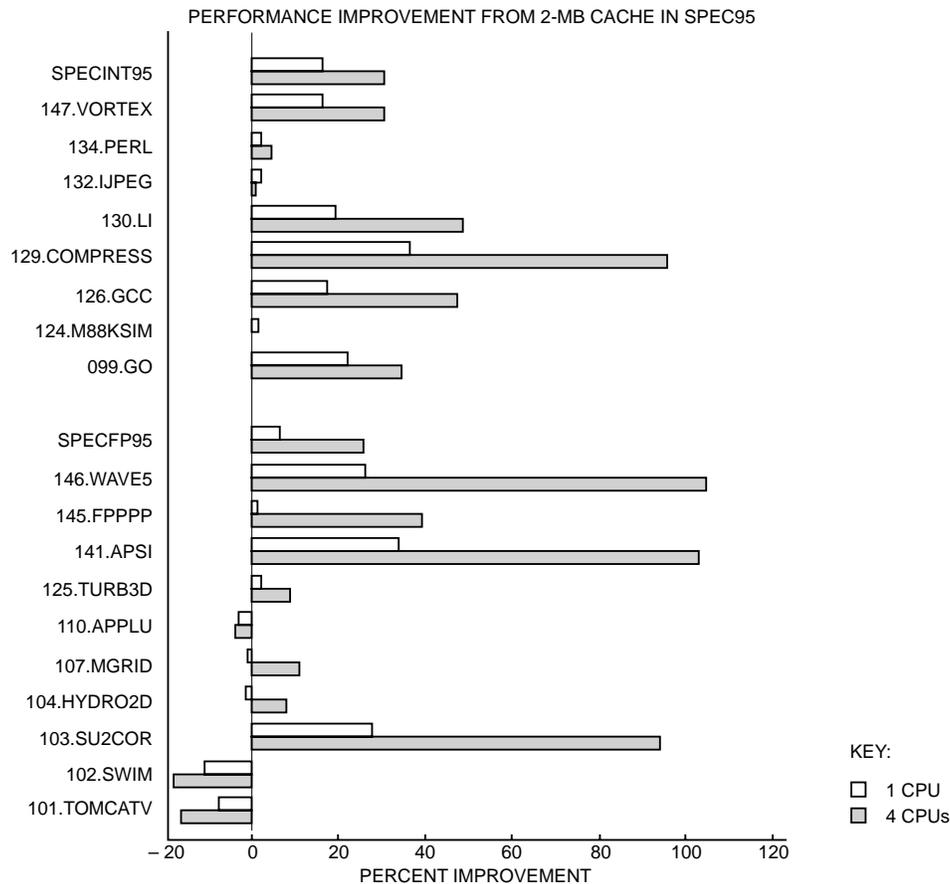


Figure 14
SPEC95 Performance Improvement from a 2-MB B-Cache

Image Rendering

The AlphaServer 4100 shows significant performance advantage in image rendering applications compared to the other industry-leading vendors. Figure 16 shows that the AlphaServer 4100 5/400 system is approximately 4 times faster than the Sun SPARC system that was used in the movie *Toy Story*, as measured in RenderMarks.¹⁵ The AlphaServer 4100 is 2.6 times faster than the Silicon Graphics POWER CHALLENGE system and 2.4 times faster than the HP/Convex Exemplar SPP-1200 system on the Mental Ray image rendering application from Mental Images. These image rendering applications take advantage of larger caches, and the performance improves as the cache size increases, particularly with four CPUs.

Performance Counter Profiles

The figures in this section, Figures 17 through 22, show the performance statistics collected using the built-in Alpha 21164 performance counters on the AlphaServer 4100 5/400 system. These hardware monitors collect various events, including the number and type of instructions issued, multiple issues, single

issues, branch mispredictions, stall components, and cache misses.^{3,16,17} These statistics are useful for analyzing the system behavior under various workloads. The results of this analysis can be used by computer architects to drive hardware design trade-offs in future system designs.

The SPEC95 cycles per instruction (CPI) data presented in Figure 17 shows lower CPI values for the integer benchmarks (CPI values of 0.9 to 1.5) than for the floating-point benchmarks (CPI values of 0.9 to 2.2). The CPI in commercial workloads (e.g., TPC-C) is higher than in the SPEC benchmarks, primarily because commercial workloads have a higher stall time, as shown in Figure 18. Note that the performance counter statistics were collected with four CPUs running TPC-C (with a Sybase database), while SPEC95 statistics were collected on a single CPU.

The Alpha 21164 has two integer and two floating-point pipelines and is capable of issuing up to four instructions simultaneously. The integer pipeline 0 executes arithmetic, logical, load/store, and shift operations. The integer pipeline 1 executes arithmetic, logical, load, and branch/jump operations. The floating-point pipeline 0 executes add, subtract,

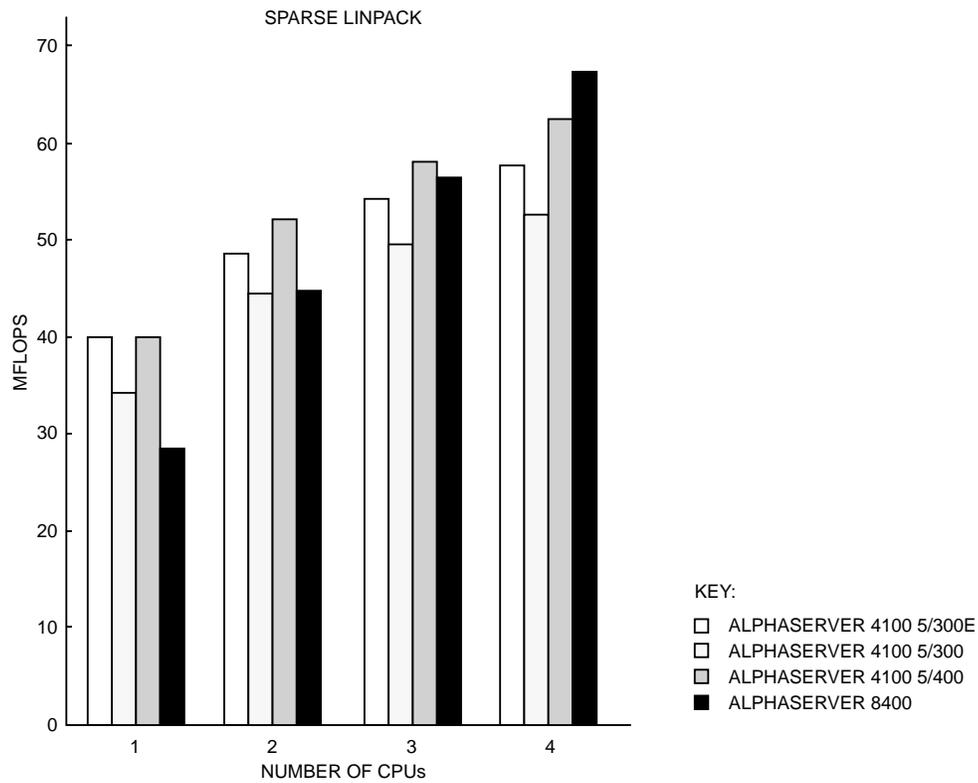


Figure 15
Sparse LINPACK Performance

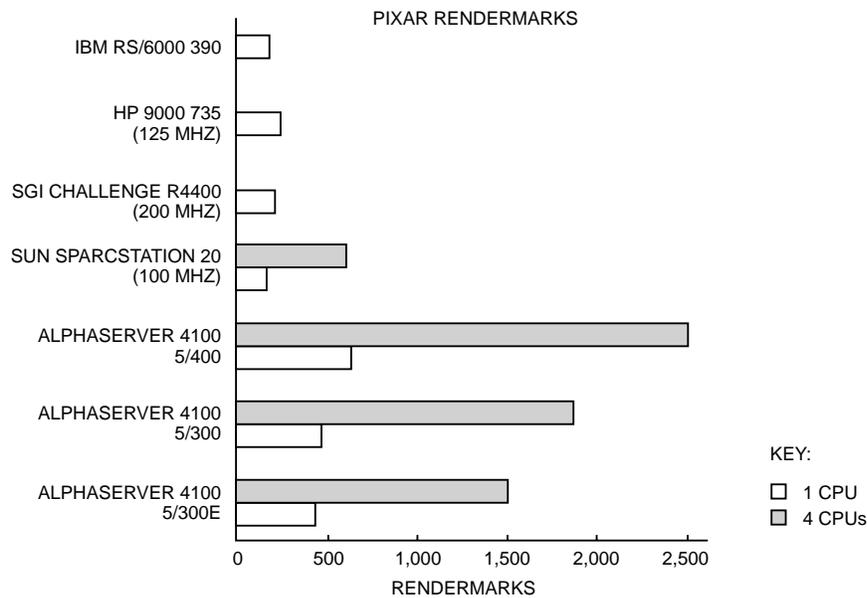


Figure 16
Image Rendering Performance

compare, and floating-point branch instructions. The floating-point pipeline 1 executes multiply instructions. The time distribution illustrated in Figure 18 indicates that most of the issuing time is spent in single

and dual issuing. Triple and quad issuing is noticeable in several floating-point benchmarks, but, on average, only 3 percent of the time is spent on triple and quad issuing in the SPECfp95 benchmarks.

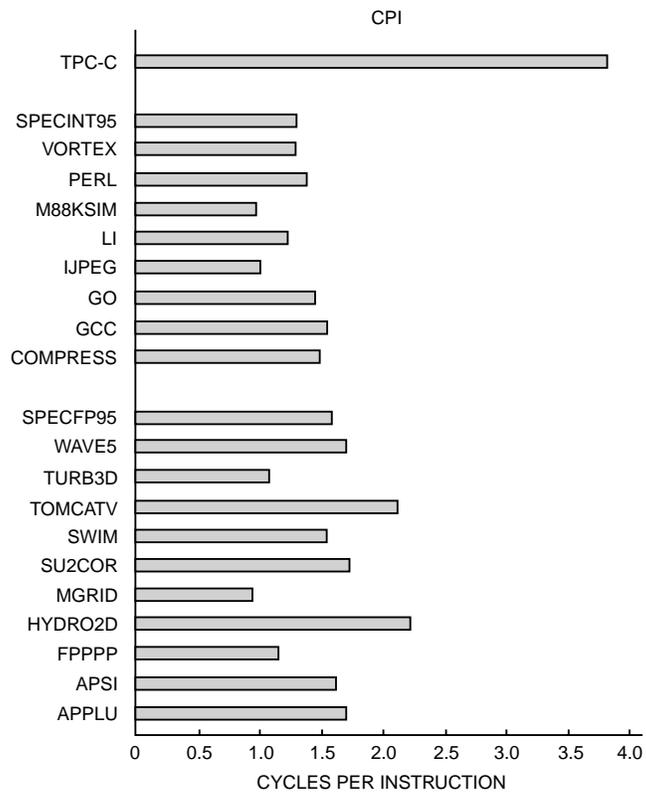


Figure 17
SPEC95 Cycles-per-instruction Comparison

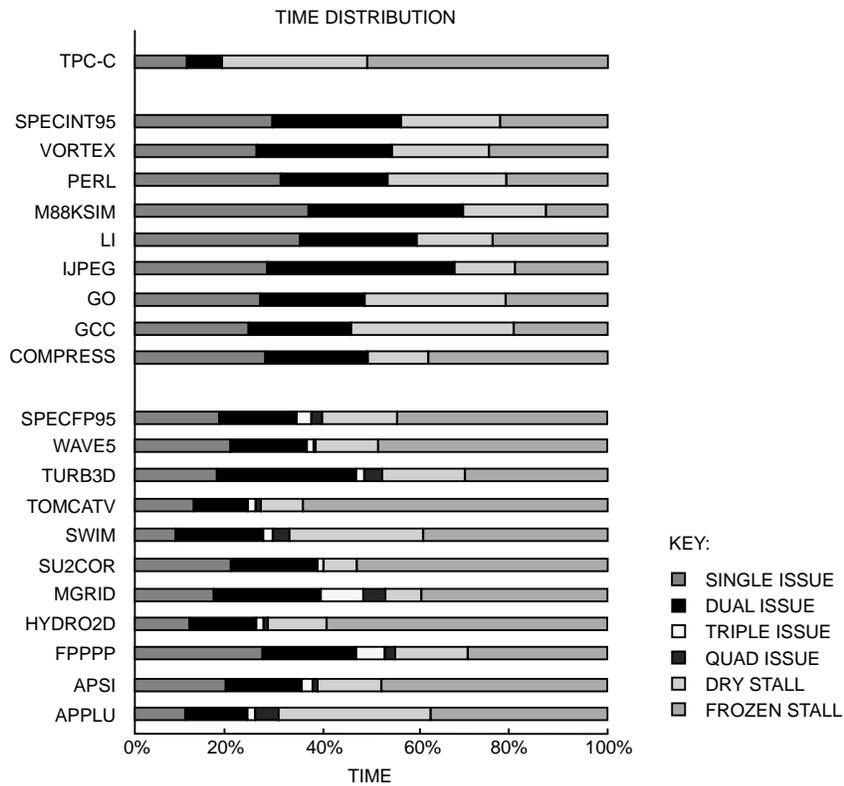


Figure 18
Issuing and Stall Time

The stall time (dry plus frozen stalls in Figure 18) is higher in the floating-point benchmarks than in the integer benchmarks and higher in the TPC-C benchmarks than in the SPEC95 benchmarks. Dry stalls include instruction stream (I-stream) stalls caused by the branch mispredictions, program counter (PC) mispredictions, replay traps, I-stream cache misses, and exception drain. Frozen stalls include data stream (D-stream) stalls caused by D-stream cache misses as well as register conflicts and unit busy. Dry stalls are higher in SPECint95 and TPC-C (mainly because of I-stream cache misses and replay traps), whereas frozen stalls are higher in SPECfp95 and TPC-C (mainly because of D-stream cache misses).

The Alpha 21164 microprocessor reduces the performance penalty due to cache misses by implementing a large, 96-KB on-chip S-cache.^{3,4} This cache is three-way set associative and contains both instructions and data. The four-entry prefetch buffer allows prefetching of the next four consecutive cache blocks on an instruction cache (I-cache) miss. This reduces the penalty for I-stream stalls. The six-entry miss address file (MAF) merges loads in the same 32-byte block and allows servicing multiple load misses with one data fill. A six-entry write buffer is used to reduce the store bus traffic and to aggregate stores into 32-byte blocks.^{3,4}

Figure 19 shows the instruction mix in SPEC95. The Alpha instructions are grouped into the following

categories: load (both floating-point and integer), store (both floating-point and integer), integer (all integer instructions, excluding ones with only R31 or literal as operands), branch (all branch instructions including unconditional), and floating-point (except floating-point load and store instructions). Figure 19 shows the percentage of instructions in each category relative to the total number of instructions executed. Note that load/store instructions account for 30 to 40 percent of all instructions issued. Integer instructions are present in both integer and floating-point benchmarks, but no floating-point instructions exist in the SPECint95 and commercial TPC-C workloads. The integer and commercial workloads execute more branches, while the branch instructions make up only a few percent of all instructions issued in the floating-point workloads.

The cache misses shown in Figure 20 are higher in the floating-point benchmarks than in the integer benchmarks. The I-cache misses are low in the floating-point benchmarks (except for fpppp) and higher in the SPECint95 benchmarks and the TPC-C benchmark. The D-cache misses are high in the majority of the benchmarks, which indicates that a larger D-cache would reduce D-stream misses. The TPC-C benchmark would benefit from a larger S-cache and faster B-cache, since the number of S-cache misses is high. The B-cache misses are negligible in the SPECint95 benchmarks and higher in the majority of

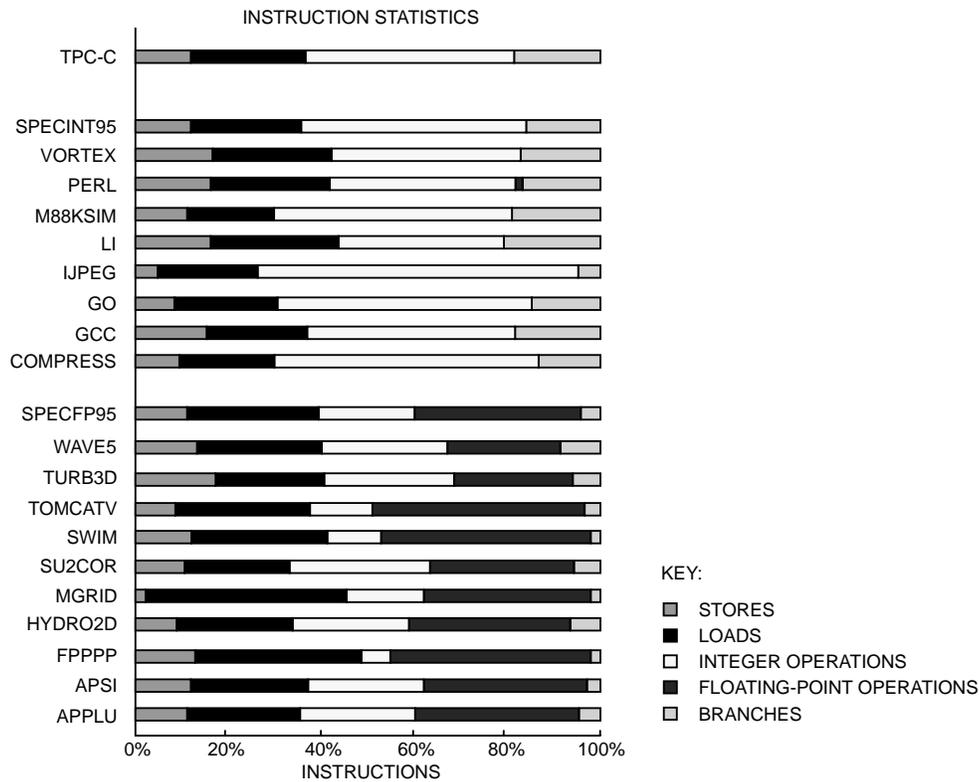


Figure 19
SPEC95 Instruction Profiles

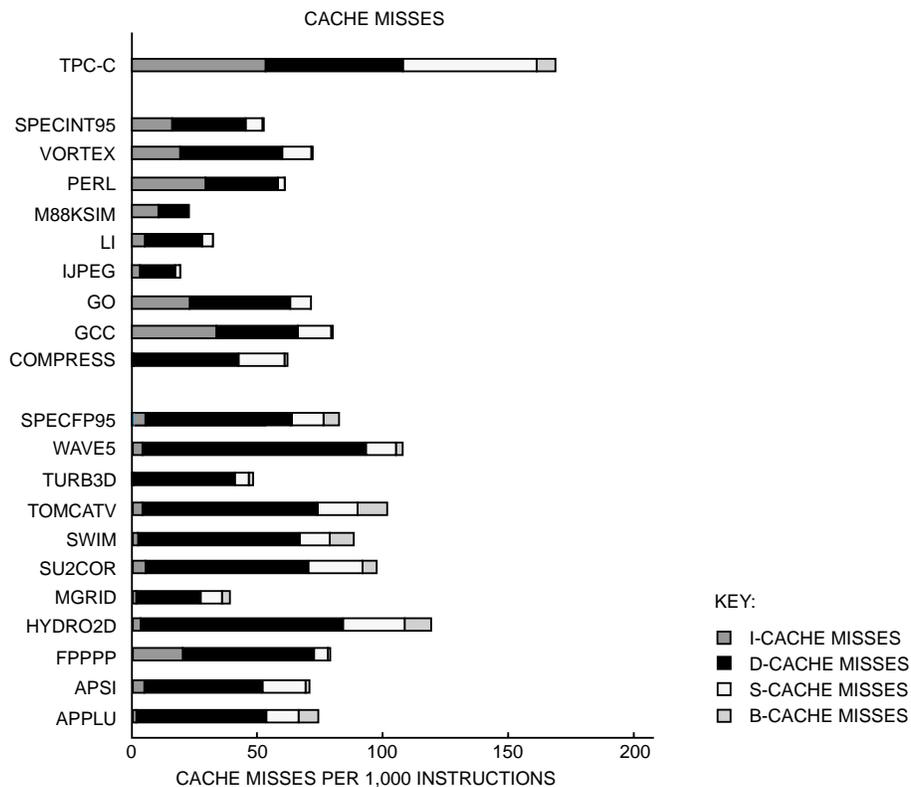


Figure 20
Cache Misses

the SPECfp95 TPC-C benchmarks. This data indicates that complex commercial workloads, such as TPC-C, are more profoundly affected by the cache design than simpler workloads, such as SPEC95.

The replay traps are generally caused by (1) full write-buffer (WB) traps (a full write buffer when a store instruction is executed) and full miss address file (MAF) traps (a full MAF when a load instruction is executed); and (2) load traps (speculative execution of an instruction that depends on a load instruction, and the load misses in the D-cache) and load-after-store traps (a load following a store that hits in the D-cache, and both access the same location).³ The replay traps and branch/PC mispredictions shown in Figure 21 are not the major reason for the high stall time in the commercial workloads (TPC-C), since traps and mispredictions are higher in some of the SPECint95 benchmarks than in TPC-C. Instead, a high number of cache misses (see Figure 20) correlates well with the high stall time and CPI (see Figure 17) in TPC-C.

Figure 22 shows the estimated stall components in SPEC95 and TPC-C. A time-allocation model is used to analyze the performance effect of different stall components. The total execution time is divided into two components: the compute component (where the CPU is issuing instructions) and the stall component (where

the CPU is not issuing instructions). The stall component is further divided into the dry and frozen stalls:

$$\begin{aligned} \text{time} &= \text{compute} + \text{stall} \\ \text{compute} &= \text{single} + \text{dual} + \text{triple} + \text{quad issuing} \\ \text{stall} &= \text{dry} + \text{frozen} \end{aligned}$$

$$\begin{aligned} \text{dry} &= \text{branch mispredictions} + \text{PC mispredictions} \\ &\quad + \text{replay traps} + \text{I-stream cache misses} \\ &\quad + \text{exception drain stalls} \end{aligned}$$

$$\begin{aligned} \text{frozen} &= \text{D-stream cache misses} \\ &\quad + \text{register conflicts and unit busy} \end{aligned}$$

The branch and PC mispredictions affect the performance of SPECint95 workloads (6 percent of the time is spent in branch and PC mispredictions in SPECint95) and have little effect on the performance of SPECfp95 workloads (less than 1 percent of the time) and the TPC-C benchmark (1.4 percent of the time). The SPECint95 workloads are affected primarily by the load traps, whereas the SPECfp95 benchmarks are affected by both load and WB/MAF traps. Note that the time spent on a load replay trap is overlapped with the load-miss time.

The S-cache and B-cache stalls are high in the SPECfp95 and TPC-C benchmarks, where the stall time is dominated by the B-cache and memory latencies. Note the high stall time resulting from waiting for

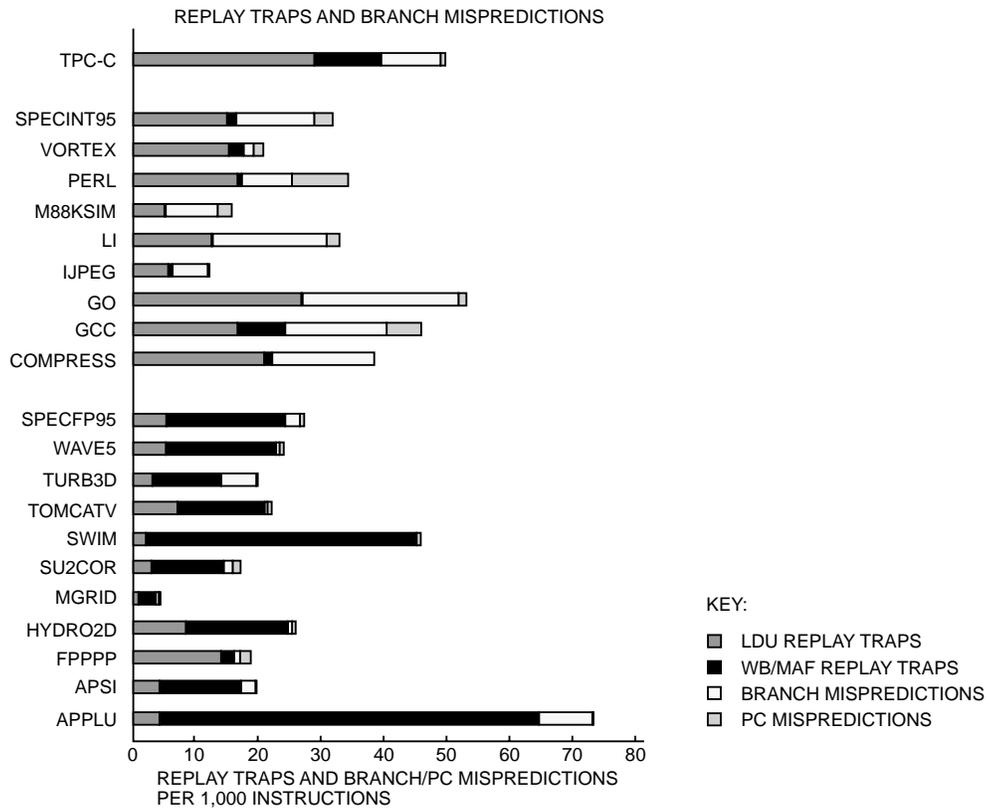


Figure 21
Replay Traps and Branch/PC Mispredictions

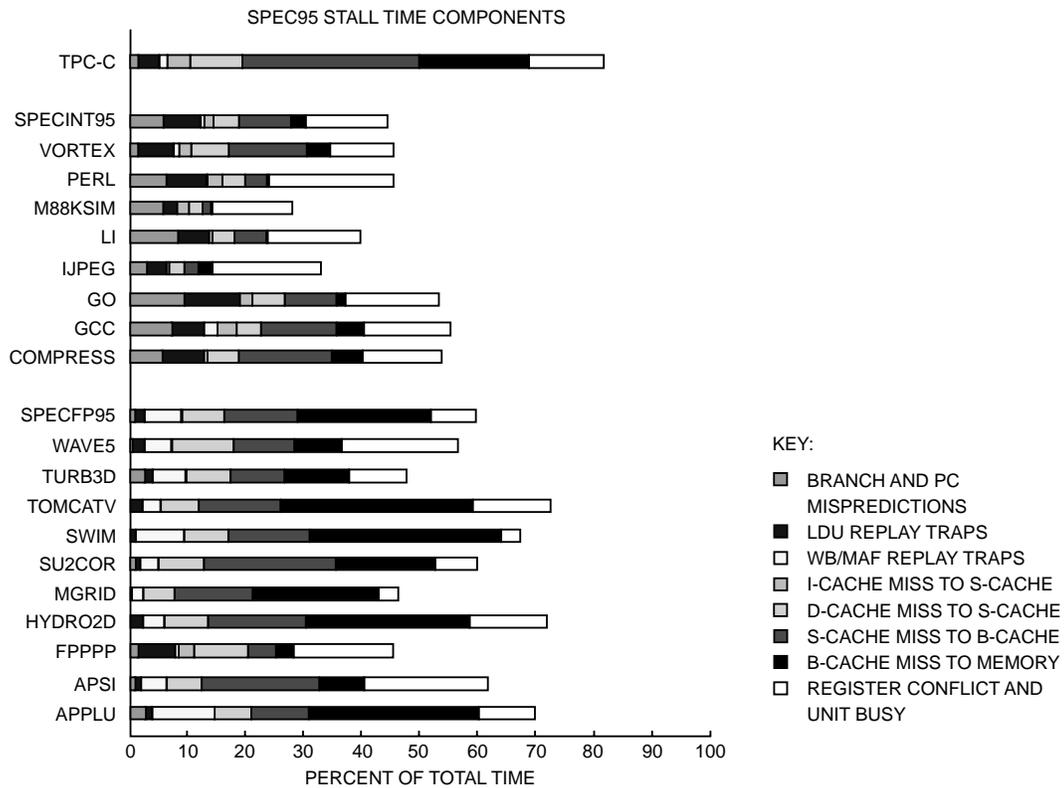


Figure 22
Estimated Stall Time Distribution

data from memory (close to 40 percent) in several of the SPECfp95 benchmarks that do not fit in a 4-MB cache. Although it contributes to the high SPECfp95 stall time, the memory component has a negligible effect on SPECint95 performance, since these benchmarks generate only a small number of B-cache misses (see Figure 20). Figure 22 indicates that stalls caused by cache misses are the largest component of the total stall time; therefore, reducing cache misses and improving cache and memory latencies would yield the largest performance benefit.

Once calibrated and validated with measurements, this model is an effective tool for evaluating the performance impact of various components on the overall system design. System architects can vary parameters, like the cache or memory access times or cache size, and adjust the appropriate stall component to predict performance of alternative designs without carrying out detailed and often time-consuming architectural simulations.

Conclusion

Using several performance metrics and a variety of workloads, we have demonstrated that the DIGITAL AlphaServer 4100 family of midrange servers provides significant performance improvements over the previous-generation AlphaServer platform and provides performance leadership compared to the leading industry vendors' platforms. The major AlphaServer 4100 performance strengths are the low memory and I/O latency and high memory bandwidth, the large-memory support (VLM), and the fast Alpha 21164 microprocessor. The work described in this paper has led to design changes that are expected to be implemented in future versions of the AlphaServer 4100 platform. The anticipated performance benefits will come from a faster CPU, faster and larger caches, faster memory, and improved memory bandwidth.

Acknowledgments

The authors would like to acknowledge the contributions of John Shakshober, Dave Stanley, Greg Tarsa, Dave Wilson, Paula Smith, John Henning, Michael Delaney, and Huy Phan for providing many of the benchmark measurements. In addition, special thanks go to Maurice Steinman, Glenn Herdeg, and Ted Gent for dedicating system resources and to Masood Heydari for supporting this work.

References

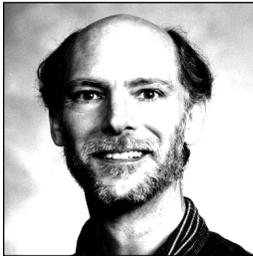
1. G. Herdeg, "Design and Implementation of the AlphaServer 4100 CPU and Memory Architecture," *Digital Technical Journal*, vol. 8, no. 4 (1996, this issue): 48–60.
2. M. Steinman, G. Harris, A. Kocev, V. Lamere, and R. Pannell, "The AlphaServer 4100 Cached Processor Module Architecture and Design," *Digital Technical Journal*, vol. 8, no. 4 (1996, this issue): 21–37.
3. *Alpha 21164 Microprocessor Hardware Reference Manual* (Maynard, Mass.: Digital Equipment Corporation, Order No. EC-QAEQA-TE, 1994).
4. J. Edmondson, P. Rubinfeld, and V. Rajagopalan, "Superscalar Instruction Execution in the 21164 Alpha Microprocessor," *IEEE Micro*, vol. 15, no. 2 (April 1995).
5. R. Sites, ed., *Alpha Architecture Reference Manual* (Burlington, Mass.: Digital Press, ISBN 1-55558-098-X, 1992).
6. SPEC95 Benchmarks (Manassas, Va.: Standard Performance Evaluation Corporation, 1995).
7. J. Dongarra, "Performance of Various Computers Using Standard Linear Equation Software" (Oak Ridge, Tenn.: Oak Ridge National Laboratory, 1996).
8. *UNIX System Price Performance Guide* (Menlo Park, Calif.: AIM Technology, Summer 1996).
9. J. Gray, ed., *The Handbook for Database and Transaction Processing Systems* (San Mateo, Calif.: Morgan Kaufman, 1991).
10. Information about the lmbench suite of benchmarks is available at http://reality.sgi.com/employees/lm_enr/lmbench/whatis_lmbench.html.
11. The STREAM benchmark program is described on-line by the University of Virginia, Department of Computer Science (Charlottesville, Va.) at <http://www.cs.virginia.edu/stream>.
12. The Standard Performance Evaluation Corporation (SPEC) makes available submitted results, benchmark descriptions, background information, and tools at <http://www.specbench.org>.
13. Information about the Transaction Processing Performance Council (TPC) is available at <http://www.tpc.org>.
14. Information about system performance benchmarking products from AIM Technology, Inc. (Menlo Park, Calif.) is available at <http://www.aim.com>.
15. Information about Pixar Animation Studio's RenderMark benchmark is available at <http://www.europe.digital.com/info/alphaserver/news/pixar.html>.
16. Z. Cvetanovic and D. Bhandarkar, "Characterization of Alpha AXP Performance Using TP and SPEC Workloads," *The 21st Annual International Symposium on Computer Architecture* (April 1994): 60–70.
17. Z. Cvetanovic and D. Bhandarkar, "Performance Characterization of the Alpha 21164 Microprocessor Using TP and SPEC Workloads," *The Second International Symposium on High-Performance Computer Architecture* (February 1996): 270–280.

Biographies



Zarka Cvetanovic

A consulting engineer in DIGITAL's Server Product Development Group, Zarka Cvetanovic was responsible for the performance characterization and analysis of the AlphaServer 4100, AlphaServer 8400/8200, AlphaServer 2100, DEC 7000, VAX 7000, and VAX 6000 systems, and for the performance modeling and definition of future AlphaServer platforms. Since joining DIGITAL in 1986, she has been involved in the development of fast database applications and efficient parallel applications for multiprocessor systems. Zarka received a Ph.D. in electrical and computer engineering from the University of Massachusetts, Amherst. She has published over a dozen technical papers at computer architecture conferences and in leading industry journals.



Darrel D. Donaldson

Darrel Donaldson is a senior consulting engineer and the technical leader and engineering manager for the AlphaServer 4100 project. He joined DIGITAL in 1983 and served as the lead technologist for the VAX 6000, VAX 7000, AlphaServer 7000, and AlphaServer 4100 projects. Darrel has a bachelor's degree in mathematics/physics from Miami University and a master's degree in electrical engineering from Cincinnati University, Cincinnati, Ohio. He holds 12 patents and has 10 patents pending, all related to protocols, signal integrity, and chip transceiver design for multiprocessor systems and non-volatile memory chip design. Darrel maintains membership in the IEEE Electron Devices Society and the Solid-State Circuits Society.