First SPEC95 Results Pump Up HP, IBM

Digital's Lead Narrowed, Sun Sags—Bandwidth Becomes Critical

by Linley Gwennap

SPEC has released the first measured results for its new SPEC95 benchmarks, including results from all five major RISC architectures as well as the x86. Although the new benchmark was designed to fix several problems in the SPEC92 suite, most of the new integer results are within 10% of the older results (accounting for a scaling factor). Digital and Sun had significant performance losses, however. On the new floating-point tests, vendors have gained or lost performance based mainly on cache size and memory bandwidth.

The biggest change between the benchmarks in SPEC92 and SPEC95 is an emphasis on larger data sets in the new tests *(see* ***091102.PDF****)*. As a result, the new tests generate far more cache misses. The differences between the older scores and the new ones tend to reflect the processor's ability (or inability) to access main memory quickly. The SPEC95 suite should be more indicative of the performance of real applications, which frequently access main memory.

In moving to SPEC95, we are listing only baseline scores, which limit compiler tuning, denoting these as SPECint95b and SPECfp95b. Several vendors apparently agree that the fully tuned "peak" SPEC scores have little relevance: Digital, IBM, and Sun did not even report peak results, sticking entirely with the baseline metric. For those vendors that reported peak results, the gain was less than 10% for all integer results and all but one FP result, as noted below. Our analysis compares baseline SPEC95 results to peak SPEC92 results using a scaling factor of 2.47%, which was calculated as the mean of the SPEC95/SPEC92 ratios for a variety of processors.

The data in Table 1 show a wide range. Digital's 300-MHz 21164 turns in the highest SPECint95b score at 7.33, while the PowerPC 601 rates just 2.37. There is even more variation on the floating-point side, with Digital again leading the pack at 11.6 while SuperSparc-2 trails at 2.14. For a given machine, the individual test scores vary widely, indicating that the averages may not represent the performance of particular applications.

Processor	Digital 21164	PowerPC 604	HP PA-7200	Digital 21064A	HP PA-7150	MIPS R4400SC	IBM Power2	Intel Pentium	Sun S'Sparc-2	PowerPC 601
System	AlphaSta. 600 5/300	IBM 43P RS/6000	HP9000 Mod J210	DEC 3000 Mod 900	HP9000 735/125	Siemens RM400	IBM 591 RS/6000	Xtended Xpress	S'Station Mod 20/71	IBM C10 RS/6000
Clock Rate	300 MHz	133 MHz	120 MHz	275 MHz	125 MHz	200 MHz	77 MHz	133 MHz	75 MHz	80 MHz
Cache (L1/L2)	8K/96K/4M	32K/512K	512K/0K	32K/2M	512K/0K	32K/4M	288K	16K/1M	36K/1M	32K/1M
099.go	10.3	5.56	7.40	5.99	6.87	4.73	4.91	4.70	3.57	3.15
124.m88Ksim	7.2	4.50	3.80	4.46	3.41	3.05	2.66	3.20	2.02	2.07
126.gcc	6.4	4.24	5.01	3.90	4.62	3.82	3.84	3.57	2.58	2.61
129.compress	6.7	4.00	4.34	4.20	3.94	4.50	4.90	3.64	2.50	2.69
130.li	6.6	3.82	4.28	3.50	4.12	3.57	3.03	4.16	2.05	2.18
132.ijpeg	6.9	5.63	3.33	4.39	2.78	3.75	4.27	2.78	2.32	2.30
134.perl	7.9	4.08	3.90	4.38	3.53	3.86	3.10	3.98	2.78	2.07
147.vortex	7.4	4.13	3.94	3.54	4.10	3.29	3.37	3.39	2.18	2.07
SPECint95b†	7.3	4.45	4.37	4.24	4.04	3.79	3.67	3.64	2.46	2.37
% change‡	–13%	+2%	+5%	–14%	+9%	+9%	+3%	–5%	–21%	–4%
101.tomcatv	14.7	3.63	14.0	8.39	5.05	—	23.1	4.01	3.57	4.41
102.swim	19.2	4.88	21.4	11.8	10.6	—	28.6	4.07	5.88	6.71
103.su2cor	6.9	1.73	3.25	3.98	2.19	—	8.20	1.70	1.65	2.25
104.hydro2d	7.8	1.59	4.26	3.83	2.19	—	5.40	1.59	1.32	1.54
107.mgrid	9.3	2.92	5.26	6.02	3.13	—	8.22	1.33	1.44	2.45
110.applu	6.5	2.35	4.47	3.71	2.88	—	8.28	0.67	1.10	2.25
125.turb3d	11.0	4.63	5.62	4.80	4.25	—	9.38	2.81	2.09	2.04
141.apsi	13.6	4.08	7.17	8.74	4.21	—	7.39	2.87	2.29	3.09
145.fpppp	21.4	8.24	12.1	8.85	11.1	—	13.3	5.07	2.05	4.53
146.wave5	14.7	3.01	12.5	7.73	8.23	—	17.0	3.49	2.85	3.29
SPECfp95b†	11.6	3.31	7.54	6.29	4.55	—	11.2	2.37	2.14	2.97
% change‡	–8%	–15%	+13%	–4%	–8%	—	+47%	–17%	–29%	+22%

Table 1. Relative to SPEC92, Sun and Digital lost ground on the new SPEC95 test, while the PA-7200, PowerPC 601, and Power2 improved. †baseline results ‡relative change from SPEC92 results scaled by a factor of 0.0247 (Source: SPEC except MDR scaling factor)

Sun Stunned, Digital Diminished

Sun's SuperSparc-2 took the biggest hit, losing 21% on the new integer rating and 29% on the new FP score compared with SPEC92. SuperSparc-2 accesses main memory across the sluggish (264-Mbyte/s) MBus but has a fast path to its secondary cache; thus, it looks better on the cache-bound SPEC92 than on SPEC95. Sun also noted that its initial SPEC95 scores rely on its own SunPro compiler, while its SPEC92 scores used the more advanced Apogee compilers. Sun's forthcoming UltraSparc processor accesses main memory at up to 1.3 Gbytes/s and so should not suffer the same performance loss as SuperSparc-2.

Also hurt on SPECint95 is Digital: its Alpha processors lose 13–14% compared with SPECint92. In this case, the problem is not lack of bandwidth but rather long latency; when measured in CPU cycles, the time needed to access main memory at 275–300 MHz is far greater than for a slower processor. While the 21164 continues to lead the pack in performance, the 21064A, which was ahead of all non-Alpha chips, now falls behind the PowerPC 604 and the PA-7200 on SPECint95.

HP's processors increased slightly, as their large primary caches do well on the new integer tests. The MIPS R4400 also gained a bit, probably because it was benchmarked with a huge 4M cache, and passed Pentium on the integer rankings. Among the major RISCs, only SPARC does not have a chip faster than Pentium.

IBM Powers Through SPECfp95

The new floating-point suite places even more of an emphasis on main-memory access than the integer tests. It also includes more double-precision math than the older SPECfp92 suite. IBM's Power2 got the biggest boost: a 47% increase over its scaled SPECfp92 score, trailing only the 21164 on SPECfp95b. On the new scale, the Power2 is more than three times better than the fastest PowerPC chip, showing why IBM continues to build Power2-based systems for FP-intensive applications. The multichip processor benefits from its phenomenal 2.3-Gbyte/s bandwidth to main memory.

The new FP suite gives the PowerPC 601 a bigger boost than the 604 due to the 601's unified L1 cache, which is more effective than split caches on programs with large data sets and small inner loops. In fact, the 604 beats the 601 by just 11%. The scores in Table 1 are not a fair comparison, however, as the 601 was tested with 1M of secondary cache, while the 604 uses only 512K.

As Table 2 shows, IBM measured a 120-MHz 604 with 1M of cache at 3.50 SPECfp95b, which is better than the 133-MHz 604's rating. With a 1M cache, the 133-MHz part should turn in at least 3.8 SPECfp95b, giving it a bigger lead over the 601. This example shows how sensitive SPECfp95 is to L2 cache size.

For More Information

For more information on the SPEC95 benchmarks, or for a subscription to the quarterly newsletter that details all SPEC95 results, contact the System Performance Evaluation Corporation (SPEC) at 703.698.9604; fax 703.560.2752 or e-mail *spec-ncga@cup.portal.com*.

The PA-7200, with a 768-Mbyte/s sustainable bandwidth to main memory, shines on the SPECfp95b tests. It gets a big boost relative to the PA-7150, its predecessor, which is constrained by its puny, 264-Mbyte/s path to main memory.

Pentium fares relatively poorly on the new FP suite; like the PowerPC chips, it must share a 528-Mbyte/s bus between L2 cache and main memory accesses. Unlike any other vendor, Intel gets a huge increase of 28% when hand-tuning for each test program is allowed. In particular, tweaking the compiler nearly doubles Pentium's performance on 102.swim and triples performance on 110.applu. Over time, other vendors may achieve higher numbers from compiler tweaking.

One issue that SPEC does not address is price. For example, the 21164 not only costs $2,936, but it requires at least 2M of expensive cache to buffer the 300-MHz CPU from the relatively slow main memory. Pentium, at $694, is tested with 1M of zero-wait-state cache. The 604-133 lists for $756 but with 512K of cache delivers better performance than the Pentium. The $253 R4600 delivers respectable performance with no external cache at all.

The new suite should give a better indication of processor performance on typical application software. The changes are not dramatic enough, however, to upset our overall view of processor performance. We applaud the vendors who have given up on excessive SPEC-specific compiler tuning and hope that others will follow their lead. The only problem then will be getting used to SPEC scores with a decimal point after the first digit. ♦

System	Processor	Cache (L1/L2)	SPEC95b† int	SPEC95b† fp
DEC 3000 Model 500	21064/150	16K/512K	2.15	3.65
AlphaStation 200 4/233	21064A/233	32K/512K	3.39	3.95
DEC 9000 Model 700	21064A/225	32K/2M	3.66	5.71
AlphaStation 250 4/266	21064A/266	32K/512K	4.18	5.78
AlphaStation 600 5/266	21164/266	112K/2M	6.30	10.0
AlphaStation 600 5/266	21164/266	112K/4M	6.43	10.6
HP 9000 Model 715/100	7100LC/100	256K/0K	2.89	3.47
HP 9000 Model 735/99	PA7100/99	512K/0K	3.27	3.98
HP 9000 Model J200	PA7200/100	512K/0K	3.64	6.28
IBM RS/6000 39H/3CT	Power2/67	160K/2M	3.28	9.44
IBM RS/6000 Model C20	PPC604/120	32K/1M	3.85	3.50
SNI RM200 Model 225	R4600/133	32K/0K	2.31	—

Table 2. Additional SPEC95 benchmark scores include systems based on the PA7100, 7100LC, 21064, and R4600 processors. †baseline results (Source: SPEC)

