THE EDITORS' VIEW

Wanted: Benchmarks for Windows NT
When Worlds Collide, Old Benchmarks Lose Their Value

As Intel processors and RISC processors inexorably collide in the Windows NT market, the old ways of comparing performance no longer suffice. While the PC and workstation worlds each have their own benchmarks, these measurements become inadequate or even impossible to run on the next generation of systems. NT, the first PC operating system to run on multiple architectures, forces us to re-examine our arsenal of tests to determine which, if any, are appropriate for this new market.

The vanishing distinction between PCs and workstations means that PC benchmarks cannot be used for broad comparisons, since they run only on x86-compatible processors. "MIPS" has too many meanings to be useful, and Dhrystone tests too little of the processor and is susceptible to compiler "tricks." Benchmarks for transaction processing and scientific calculations tend not to be representative of desktop system performance.

The most widely available benchmark is the SPEC suite. These tests have been ported to most versions of UNIX, as well as a few other operating systems, on nearly every major CPU. Thus, SPEC is often used to compare processors in a variety of markets.

Unfortunately, the SPEC suite was originally developed solely for the workstation market. As such, the first tests (SPECmark89) were weighted heavily (60%) toward floating-point programs. While this proportion may reflect the needs of a typical CAD engineer or theoretical physicist, it is not appropriate for mainstream desktop users. Floating-point power accelerates 3-D imaging but it does little to assist the windows-based GUIs and simple graphics that most people use.

Recognizing this situation, SPEC revamped the test suite last year, separating the integer and floating-point portions into two new numbers called SPECint92 and SPECfp92, respectively. Examining these two scores clearly differentiates between companies—such as DEC, HP, and IBM—that offer very high FP performance, and other vendors—SPARC, MIPS, and Intel—that are merely adequate in that area. In fact, the former group, once thought to hold a tremendous performance advantage over the latter, has little or no advantage when only SPECint92 is compared.

This leads to some creative marketing. In a recent press release, DEC used the term "aggregate SPECmark" to refer to the obsolete SPECmark89 rating, and went on to discuss its leadership in aggregate SPECmarks per dollar. This leadership is based solely on Alpha's fast floating-point performance, which most of its target customers will not need to use. These customers might as well try to commute to work in an F-16. DEC would be better off touting its leading SPECint92 ratings *(see **0705MSB.PDF**)*.

Floating point could become more important in the future if animation becomes popular. For motion video that consists of a series of stored frames, integer performance (or more accurately, memory-to-screen bandwidth) is all that's required, but some techniques create new images by mathematically altering and combining stored data. The latter method requires numerous floating-point calculations, which may be handled by the CPU or by a special processor. Until the future of multimedia becomes clear, SPECfp92 means little unless you have a heavy scientific or technical workload.

SPECint92 provides a good first estimate when comparing the performance of different desktop platforms for general use. It would be even more useful if the tests were ported to Windows NT, since vendors such as Acer *(see **070501.PDF**)* are building systems that don't run UNIX. At some point, though, the best measure of comparison will be native NT benchmarks.

The ideal test would stress various portions of the system: processor, memory, graphics, and disk would be a good start. It must, like NT itself, be processor-independent. It should use actual Windows NT system calls and simulate the workload of a "typical" user. Since many of us aren't typical, it should provide a breakdown of results so that individual users can customize the results for their own applications. This test would combine the best features of BAPCo and the SPEC suite.

One area that requires special focus is DOS emulation under Windows NT. RISC advocates say they have superior performance on native NT code, but they must emulate an x86 system when running other applications. Vendors make claims that their RISC chips can run x86 code as fast as a 386, or even a 486, but there are no tests to verify these claims. Running BAPCo under emulation would be interesting, but the current version may require some modification to work with NT.

DEC and MIPS must demonstrate superior price/performance over the x86 to gain a significant share of the NT market. Until better benchmarks come along, SPECint92 is the best way to compare the performance of these processors for general-purpose systems. Like many other tests, SPECfp92 is helpful in certain niche markets but not for general use.♦

Linley Gwennap