Evaluating High-Performance Computer Technology through Industrially Significant Applications

Rudolf Eigenmann
Purdue University
School of Electrical and Computer Engineering
West Lafayette, Indiana 47907
Phone: 317-494-1741
Email: eigenman@ecn.purdue.edu

and

Siamak Hassanzadeh
Sun Microsystems
2550 Garcia Avenue
Mountain View, California 94043
Phone: 415-336-0118
Email: siamak.hassanzadeh@corp.sun.com

Abstract

The paper gives an overview and a brief status of the activities of the High-Performance Group of the Standard Performance Evaluation Corporation. SPEC/HPG has recently released a first suite of benchmarks, called SPEChpc96, that will be used to evaluate high-performance computer systems across the wide spectrum of available systems. The benchmark suite includes industrially significant applications in different areas. The effort is broadly supported by industrial and academic institutions. The paper also describes roles that participants of this benchmarking activity can play.

Keywords: Benchmarking, Performance Evaluation, High-Performance Computers, Computational Applications, SPEC, SPEChpc96.

1 The SPEC High-Performance Group: A New Benchmarking Initiative

In January 1994 the High-Performance Group of the Standard Performance Evaluation Corporation (SPEC/HPG) was founded with the mission to establish, maintain and endorse a suite of benchmarks representative of real-world, high-performance computing applications. Current SPEC/HPG members include Convex Computer Corp. (now Hewlett-Packard Convex Technology Division), Cray Research, Digital Equipment Corp., Electronic Data Systems (EDS), Fujitsu America, Hewlett-Packard, International Supercomputing Technology Institute (ISTI, France), Kuck & Associates, NEC/HNSX Supercomputers, Silicon Graphics, Sun Microsystems, the Parkbench organization (represented by the University of Tennessee), the University of Illinois, the University of Michigan, the University of Minnesota, and Purdue University.

Several efforts joined forces to form SPEC/HPG and to initiate a new benchmarking venture that is supported broadly. Founding partners include the member organizations of SPEC, former members of the Perfect Benchmarks effort, and representatives of area-specific benchmarking activities. Other benchmarking organizations have joined the SPEC/HPG committee since its formation.

Figure 1 illustrates the various contributors to SPEC/HPG. Since its foundation in the late 1980's, the Standard Performance Evaluation Corporation has become the accepted leader in evaluating the performance of workstations. SPEC's mission is to provide tools that help both manufacturers and users gain more insight into the performance that can be expected from their computer systems. SPEC is broadly industry-based. Its membership has grown to more than 35 organizations, including computer vendors, systems integrators, research firms and academic institutions. Benchmark test results, such as performance numbers of the most recent SPEC95 suite, are reported each quarter in the SPEC

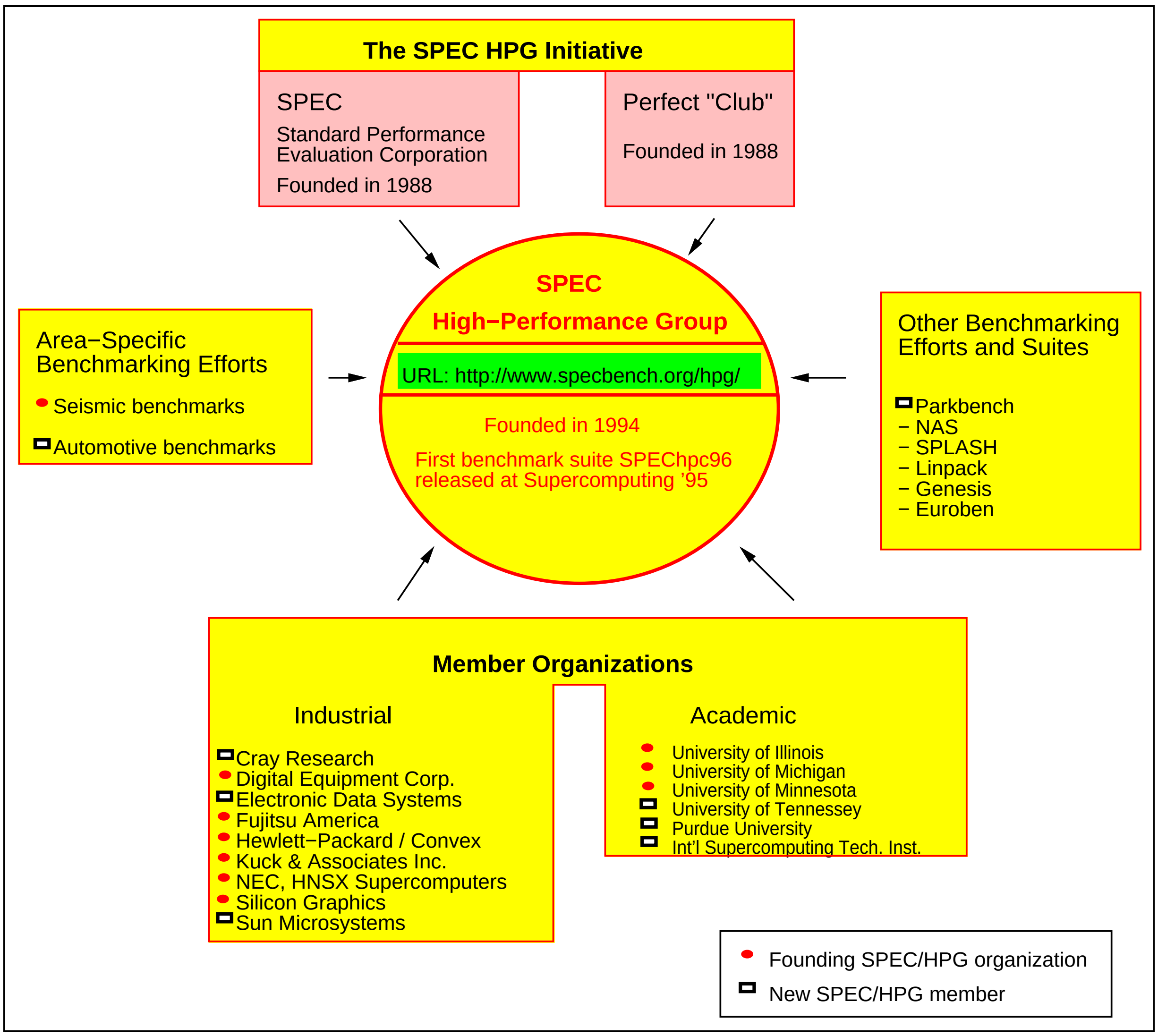

Figure 1: The various contributors and the membership of the SPEC High-Performance Group.

Newsletter[1]. With the formation of the High-Performance Group, SPEC extends its activities to address performance evaluation needs across the large spectrum of today's high-performance systems.

The Perfect Benchmarks® effort was initiated in 1988 [3, 4] to provide a balanced set of realistic scientific and engineering application programs and to use them for evaluating high-performance computers. In coordination with the vendors of such systems, performance results were obtained and documented in several reports [3, 11]. This effort led to the first comprehensive evaluation of the sustainable application performance of supercomputers at that time. While the Perfect Benchmarks are still in-use by many researchers, the need for an updated set of programs with larger data sets and the objective to maintain the benchmarking effort over a long time period led to the search for new benchmarking partners, which resulted in the foundation of the SPEC/HPG committee.

The Seismic Benchmarking initiative, informally known as the "ARCO suite" [10] is one area-specific benchmarking effort with the objective to provide a suite of application codes representative of the seismic processing industry. Parts of this suite have already been included in several other benchmarking activities. Thanks to the participation of this effort, SPEC/HPG could start with an important example of an industrially significant application and the corresponding methodology of benchmarking high-performance computers.

Another effort to create an area-specific suite is described in [7]. The objective of this project is to create benchmarks for evaluating high-performance computers used in the automotive industry. A representative of this effort has joined SPEC/HPG, and forms of collaboration are being discussed. For example, this effort may provide the automotive area benchmarks of the SPEChpc96 suite.

The *Parkbench* (PARallel Kernels and BENCHmarks) committee was founded in 1992 by a group of interested parties from universities, laboratories and industries. Members from both Europe and the USA [8] are participating in this effort with the goal of establishing a comprehensive set of parallel benchmarks and to set standards for benchmarking methodology. Maintaining a repository for the benchmarks and results is part of the objective. The Parkbench activity is represented in SPEC/HPG and the two efforts attempt to complement each other to the extent possible.

In summary, SPEC/HPG is a new benchmarking effort that is supported broadly by industrial and academic institutions. It includes representatives from manufacturers of most computer architectures currently offered on the market. It brings together benchmarking experts with those who understand market situations and customer needs of high-performance systems. It involves specialists of the computational applications used in the benchmarks as well as researchers who push forward new technology. Hence, it combines broad knowledge of the performance evaluation discipline with the necessary resources to establish and maintain a comprehensive benchmarking effort.

2 The Need for High-Performance Benchmarks

Over the past decade, huge investments have been made world-wide into the design of faster computer systems. As a result, we can now find individual computers that perform more than a hundred-million elementary operations per second. When coupling a number of such processors to build a parallel computer, vendors usually multiply the two numbers, resulting in "peak performance" figures that amount to tens or even hundreds of gigaflops[2]. Thanks to recent advancements in the field, such speeds are not only achievable by complex and very expensive systems that have fast, special-purpose technology, but new generations of commodity, off-the-shelf computer chips are rapidly approaching these peak performance levels as well.

As a result, "supercomputing" is widely expected to become what has been predicted for some time: a mainstream technology. Indeed, impressive performance demonstrations of various systems with a variety of application programs are possible. Thus far, however, these successes have not yet led to

[1] subscription can be mailed to SPEC, Suite 200, 2722 Merrilee Drive, Fairfax, VA 22031-4499, USA, or email to spec-ncga@cup.portal.com

[2] 1 gigaflop=one billion floatingpoint operations per second

the widespread application of high-performance computing technology by mainstream users. In fact – contrary to the expectation – the results of a previous effort to evaluate computer systems based on a balanced set of real applications (the Perfect Benchmarks [4]) had uncovered the fact that the sustained application performance has improved only little over a decade.

Evaluation of new computing technology and new computing paradigm is a challenging task, beyond the realm of possibility for average users. Most users of high-performance computing (HPC) systems are finding that they are spending a great amount of resources to evaluate such systems as part of upgrading their HPC facilities. On the other hand most available public domain benchmarks have not addressed the specific needs of most high-end computing users. Thus, there is a need to develop benchmarks representative of real world HPC applications for standardized, cross-platform performance comparison on a level playing field. We believe, such performance information enhances usability and adoptibility of new HPC technology within the marketplace.

We conclude, that there is also a need for a new benchmarking approach that integrates and drives forward the many useful benchmarking contributions of previous efforts. We need a combination of the evaluation effort of distributed-memory systems (led by the Parkbench organization) with shared-memory benchmarking (supported, among others, by the Perfect Club). We need overall system performance numbers combined with detailed studies of the benchmark codes and system characteristics (such results came out of both the Parkbench and Perfect efforts, as well as from many other benchmarking projects [9, 16, 5, 15, 6, 1, 14]). We need the application expertise that helps us understand the codes and ensure that they represent the current state of computational problem solving technology (such as the automotive and seismic industry benchmarks). And we need broad vendor support and the experience of an established organization (such as SPEC) for continuously obtaining new benchmark results, enforcing fair metrics, and disseminating the results to the user community.

3 Status and Plans of SPEC/HPG

3.1 First version of SPEChpc96 released

The SPEChpc96 benchmark suite			
Code	Area	Programming language	#lines
SPECchem96	Molecular modeling	Fortran 77 and C / PVM	110'000
SPECseis96	Seismic processing	Fortran 77 and C / PVM	20'000

Candidate codes		
Program name	Area	Programming language
AESSWM	Weather modeling	Fortran 77 / PVM
LANS3D	Air flow simulation	Fortran 77 / PVM
COMPACT	Mechanical engineering	Fortran 77 / PVM

Table 1: SPEChpc96 benchmarks and candidate codes

At the Supercomputing '95 conference in San Diego, California, the first release of SPEChpc96 was announced publicly. Together with the benchmark codes, a first set of performance numbers supplied by several participating vendors was presented. Benchmarks within the SPEChpc96 suite represent specific application areas. The first two benchmarks are SPECseis96, a seismic processing application, and SPECchem96, a computational chemistry application. Table 1 lists the codes of the SPEChpc96 suite together with candidate applications that are currently being considered for inclusion in the near future.

SPECseis96 is based on a benchmark suite originally developed at Atlantic Richfield Corp. (ARCO)[10]. It is an industrial application representative of modern seismic processing programs used in the search for oil and gas. SPECchem96 is based on GAMESS (General Atomic and Molecular Electronic Structure System), an improved version of programs that came from the Department of Energy's National Resource for Computations in Chemistry[13]. Many of the functions found in GAMESS are duplicated in commercial packages used in the pharmaceutical and chemical industries for drug design and bonding analysis. Both SPECseis96 and SPECchem96 incorporate a wide range of problem sizes.

SPEChpc96 can be used for performance comparisons over a broad range of high-performance computing systems, including multiprocessor systems, workstation clusters, distributed memory parallel systems, and traditional vector and vector parallel supercomputers. Performance measurements for these benchmarks are guided by a set of run rules. They allow for flexible adaptation of the codes to the various platforms and for a certain degree of program optimization. But they also restrict the participants from optimizing the codes excessively. Examples of allowed optimizations are the insertion of parallelization directives and the replacement of subroutine calls by equivalent library calls. On the other hand, manual tuning of the codes at the machine (assembly) level is disallowed.

The SPEChpc96 benchmarks are released in two versions: a serial version and a message-passing version. The current codes are written in Fortran 77 with several C routines. The PVM [2] libraries are used in the message passing code variants. Benchmarkers are free to use either version as a starting point for their measurements. Typically, for distributed-memory systems that do not provide a global address space, the message-passing variant will be used. For shared-memory parallel systems one might start with the serial version and add parallel directives as needed for taking advantage of the system resources. The provision of a program variant that expresses parallelism for global-address-space systems, is planned at a later date.

In addition to the two adopted SPEChpc96 codes, there are several codes that are being considered for inclusion in the suite in the near future. Table 1 lists three of these codes and the corresponding application areas. These codes are currently being studied for their suitability as high-performance computer benchmarks. Although the specific codes may change as a result of this evaluation process, the areas covered are considered important and will be addressed by future SPEChpc benchmarks.

3.2 Ongoing projects and future plans

The announcement of SPEChpc96 at the Supercomputing 95 conference included preliminary performance numbers. Final numbers are being produced by most participating vendors in an ongoing effort. The reports will be made available together with the SPEC newsletter in Spring 1996.

The run rules will need refinement as this benchmarking project proceeds. The initial set of rules was deemed appropriate by the HPG members as of November 1995. The process of experimenting with the codes and producing final performance figures is expected to uncover potential improvements to the rules, and such updates will be among the agenda items of the SPEC/HPG committee work.

Much of the work of running the benchmarks and generating the performance numbers is mechanical in nature. Tools that facilitate this process are desirable. A preliminary set of such tools are being released together with the codes. They facilitate compiling, linking, and executing the codes. As the codes are being used by a wide range of benchmarkers, improvements to the tools will become necessary and will be incorporated. Furthermore, the current plan is to extend the tools so that they can facilitate the production of publishable performance reports.

One of the most important ongoing activities of SPEC/HPG is the incorporation of new benchmarks into the growing suite. Currently, there exist only two application areas: chemistry and seismic processing. We envision that a variety of additional areas such as structural analysis, climate modeling, computational fluid dynamics, computational physics and electronics, will be added to the suite in the future. A continuous review of the codes will also be necessary in order to assure that the codes stay current with the practices in their respective disciplines.

Development of benchmarks is an iterative process. Thus suggestions on improving the quality and usability of SPEChpc96 suite is highly encouraged. Incorporating feedback from benchmarkers within and from outside the SPEC/HPG community will hence be another important task.

Another ongoing SPEC/HPG effort is the development of benchmark variants for shared-address-space systems. As discussed above, the current SPEChpc96 benchmark codes are only available in the serial and message-passing variant. Porting a benchmark to parallel systems that provide a global address space involves either inserting parallelization directives in the serial code or adapting the parallel message-passing variant to the environment of the given system. This porting effort could be reduced through the provision of a third variant that includes such program modifications in a generic form. To this end, we have started to study codes [12] and evaluate possible languages for expressing the new benchmark variants. We will continue this effort in coordination with the increasing number of manufacturers offering system architectures that provide a shared address space.

4 Roles that Participants of HPG can Play

Code sponsorship: A new benchmark code will only be adopted by SPEC/HPG if there is an expert sponsor available. This expert will assure that the code is in fact significant to the specific industrial area, that necessary modifications of the code will not impair its relevance, and that the automatic validation of the code, which is part of the benchmarking run, is done adequately. The sponsor will also assist the members in technical questions about porting code.

Reviewing benchmarking rules and methodology: The rules guiding the benchmarking process are always at issue. In fact, the participants have the responsibility to review these rules periodically and discuss suggested improvements.

System evaluation: Evaluation of computing platforms is primarily done by the participating system vendors. However, other participants can assist in this process. For example, performance evaluation of systems from manufacturers that are not represented in SPEC/HPG can be carried out by academic centers having access to such systems.

Code characterization: SPEC/HPG's objective of using industrially significant benchmark applications comes at the costs of having to deal with large codes. While other benchmark suites typically include several hundred to several thousand lines of code, large industrial applications can be over 100,000 lines long. By characterizing these codes and making the obtained information available to the benchmarkers, their task can become more manageable. This, in turn, furthers the overall HPG effort. Such characterizations can include basic statistics (lines of codes, subroutines, execution profiles), algorithmic characterizations, optimization diaries (applied optimizations and their performance impact), etc. Means for making this information available to the benchmarkers have yet to be defined, and tools may be created.

Evaluation of new technology: The SPEChpc96 codes are intended to measure commercial computer systems and to serve as an evaluation tool for new technologies. The latter goal is primarily driven by the HPG academic members. SPEC/HPG will actively support research efforts that are based on SPEChpc96. We will facilitate access to all information resulting from such efforts by making available code characterization, and by encouraging researchers and sponsor organizations to put increasing emphasis on new technology development that is driven by the needs of significant computational applications.

5 Conclusions

The goals of SPEC/HPG are set high. We are working toward a comprehensive evaluation of all available high-performance computer systems, using applications that are industrially significant and representative of computationally-intensive disciplines. Thanks to a broad support by major computer manufacturers and thanks to the active involvement of research organizations this goal is within reach. Achieving our objective will have a significant impact on the computer industry. For the first time it will be possible to compare systems from very different manufacturers in a quantitative way. Customers will find performance numbers that reflect their use of the system with some accuracy. New software and hardware components can be evaluated with a set of realistic applications and new concepts can be derived from the observed need in such representative codes. As a result, the computer industry will not only become more competitive but also significantly more user-oriented.

References

[1] D. H. Bailey, E. Barszcz, L. Dagum, and H. Simon. NAS parallel benchmark results. In *Proc. Supercomputing '92*, pages 386–393. IEEE Computer Society Press, 1992.

[2] A. Beguelin, J. Dongarra, A. Geist, R. Manchek, S. Otto, and J. Walpole. PVM: Experiences, current status and future direction. In *Proceedings of Supercomputing '93*, page 765, 1993.

[3] M. Berry, D. Chen, P. Koss, D. Kuck, L. Pointer, S. Lo, Y. Pang, R. Roloff, A. Sameh, E. Clementi, S. Chin, D. Schneider, G. Fox, P. Messina, D. Walker, C. Hsiung, J. Schwarzmeier, K. Lue, S. Orszag, F. Seidl, O. Johnson, G. Swanson, R. Goodrum, and J. Martin. The Perfect Club Benchmarks: Effective Performance Evaluation of Supercomputers. *Int'l. Journal of Supercomputer Applications*, 3(3):5–40, Fall 1989.

[4] George Cybenko, Lyle Kipp, Lynn Pointer, and David Kuck. Supercomputer Performance Evaluation and the Perfect Benchmarks. *Proceedings of ICS, Amsterdam, Netherlands*, March 1990.

[5] J. J. Dongarra. The Linpack benchmark: An explanation. In A. J. van der Steen, editor, *Evaluating Supercomputers*, pages 1–21. Chapman and Hall, London, 1990.

[6] C. A. Addison et. al. The GENESIS distributed-memory benchmarks. In J. J. Dongarra and W. Gentzsch, editors, *Computer Benchmarks*, pages 257–271. Elsevier Science Publishers, Amsterdam, 1991.

[7] Myron Ginsberg. Creating an automotive industry benchmark suite for assessing the effectiveness of high-performance computers. In *Proc. Ninth Int'l. Conf. on Vehicle Structural Mechanics and CAE*, page 290, 1995.

[8] R. W. Hockney and M. Berry (Editors). PARKBENCH report: Public international benchmarking for parallel computers. *Scientific Programming*, 3(2):101–146, 1994.

[9] F. McMahon. The Livermore kernels: A computer test of the numerical performance range. In J.L. Martin, editor, *Performance Evaluation of Supercomputers*, pages 143–186. Elsevier Science Publishers, Amsterdam, 1988.

[10] C. C. Mosher and S. Hassanzadeh. ARCO seismic processing performance evaluation suite, User's Guide. ARCO Technical report, Plano, TX., 1993.

[11] Lynn Pointer. Perfect: Performance Evaluation for Cost-Effective Transformations Report 2. Technical Report 964, Univ. of Illinois at Urbana-Champaign, Cntr for Supercomputing Res & Dev, March 1990.

[12] Bill Pottenger and Rudolf Eigenmann. Targeting a Shared-Address-Space version of the seismic benchmark Seis1.1. Technical Report 1456, Univ. of Illinois at Urbana-Champaign, Cntr. for Supercomputing Res. & Dev., September 1995.

[13] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elber, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgonery. The general atomic and molecular electronics structure systems. *Journal of Computational Chemistry*, 14(11):1347–1363, 1993.

[14] Jaswinder Pal Singh, Wolf-Dietrich Weber, and Anoop Gupta. Splash: Stanford parallel applications for shared-memory. *Computer Architecture News*, 20(1):5–44, 1992.

[15] A. J. van der Steen. The benchmark of the EuroBen group. *Parallel Computing*, 17:1211–1221, 1991.

[16] R. P. Weicker. An overview of common benchmarks. *IEEE Computer*, 23(12):65–75, December 1990.