

# Memorandum of Understanding between NERSC Lawrence Berkeley National Laboratory and ITWM Fraunhofer Gesellschaft

Document Version: 0.3 9/26/2000

The National Energy Research Scientific Computing Division (NERSC) at the Lawrence Berkeley National Laboratory is a leading-edge scientific computing center, providing high-end computing services and performing leading-edge computational science research, operated under the auspices of the Office of Science in the U.S. Department of Energy. The Institut für Techno- und Wirtschaftsmathematik (ITWM) in Kaiserslautern, Germany provides a critical link between the academic fields of applied mathematics and computational science on one hand, and the world of practical industrial engineering on the other.

In recent discussions, these institutions have identified several areas of mutual interest:

## Computer Benchmarking and Performance Analysis

The field of technical computing is changing rapidly. Until a few years ago, mainframe vector supercomputers, exemplified by the Cray Y-MP and T-90, dominated the field, particularly in practical, production computing. Beginning about five years ago, these systems began to be displaced by highly parallel arrays of workstation technology nodes, of which the IBM SP is the most popular. At the present time, there is increasing interest in personal computer clusters, epitomized by the "Beowulf" design -- a system constructed exclusively from off-the-shelf, commodity parts and software, typically PC nodes, with an Ethernet or Myrinet network, running the Linux operating system.

With such a wide variety of computer systems available to choose from, it has become increasingly difficult to devise "fair" benchmark tests that can objectively compare such disparate designs. Today the most commonly cited scientific computing benchmark is the scalable Linpack benchmark, whose results are compiled in the Top500 list. However, even its proponents recognize that this is a very limited measure of performance -- it is a single computational kernel that is not very demanding in irregular main memory access or network bandwidth, and, more to the point, is not very typical of real-world scientific computing (even real-world high-end scientific computing).

The NAS Parallel Benchmarks (NPB) suite, which was first introduced about ten years ago, is a much more diverse and realistic measure of performance. Sadly, however, the suite has fallen into disrepair -- the team that originally worked on this suite have dispersed and NASA has not continued to evolve it. As a result,

the NPB has become rather outdated (the problem sizes are not appropriate for today's high-end systems), and documented results are hard to find.

Thus there is a growing consensus in the field that one, or possibly several new benchmarks are needed. It is essential that these benchmarks be:

1. Scalable. The problems are easily adjustable to run on a wide range of systems, from modest-sized workstations and PC clusters of today to the petaflops-scale systems that will likely prevail ten years from now.
2. Portable. These benchmarks must be publicly available from a central web site, and easily downloaded and installed. Running the benchmarks should not require extensive effort or specialized mathematical or scientific expertise.
3. Realistic. They must provide an accurate picture of real-world technical computation, including parts that involve, among other things, (a) irregular data access, (b) main memory bandwidth, (c) substantial long-distance communication, and (d) system-level facilities, such as input/output, process scheduling and system management.
4. Widely recognized and supported. Even the most expertly crafted benchmark suite will not have a significant impact unless it is widely recognized and well supported, with a large number of up-to-date results readily available on a public web site. Thus a significant amount of continuing support, as well as some "salesmanship", will be required.

To this end, it is proposed that both parties will participate in serious activities in this arena, cooperating and complementing each other's activities as much as possible. The cooperation should be based on a common project plan, which specifies the task of each institution. As a starting point a white paper describing the requirements and project steps will be developed.

Some specific suggested activities include:

- Further development of the ESP benchmark, incorporating highly scalable and publicly available jobs into the suite, and making the entire package easily available from a web site.
- Development of new NPB-class benchmarks, promotion of them, and support of them until they become a community standard.
- Development of application-derived benchmarks which reflect the needs of industrial users.
- Execution of the benchmarks developed by both parties on classical high-end supercomputer systems, as well as any cluster-type systems that may be available.

The main activities of the project partners are in the following fields:

**NERSC:** ESP Benchmark, NPB-class benchmarks, classical supercomputers, highly parallel supercomputers

**ITWM :** Application benchmarks, PC clusters

To facilitate cooperation, a common software source tree will be established (located at NERSC) and a private discussion space using the BSCW system at the ITWM will be set up.

NERSC and ITWM agree that this arrangement is mutually beneficial and that neither party need provide financial support to the other. Exchanges of visits by personnel from each institution are expected, and these will be scheduled as needed to facilitate technical collaboration.

References:

1. Adrian Wong, Leonid Oliker, William Kramer, Teresa Kaltz and David Bailey, "ESP:A System Utilization Benchmark", manuscript available from [www.nersc.gov/~dhbailey](http://www.nersc.gov/~dhbailey).
2. "MAGMASOFT: How Parallel Computing Improves Foundry Business", SuParCup 1999.

## PC Clusters

Another area of considerable common interest is Linux-based PC clusters. As mentioned above, there is a growing interest in PC cluster systems. Such systems may provide a very cost-effective alternative, particularly appropriate for small- and medium-sized laboratory systems, but potentially even for high-end supercomputer systems of the future. There are however many hurdles that must be overcome:

1. System management. Typical research cluster systems lack many of the system management tools that are available on large commercial systems.
2. Baseline software distribution. Aside from the Linux operating system itself, there is no central repository of field-tested, battle-hardened software appropriate for cluster systems.
3. Performance and optimization tools. Much is needed in this area to make these systems truly usable for production-quality computation.

It is proposed that both parties work on PC cluster software components, share their software with the other party, and seek to form a community-wide system for testing, validating and distributing usable PC cluster software.

Planned activities:

- Establishment of a common source tree
- Elaboration of a work plan based on a white paper describing requirements
- First software release in Q1 2001 (small systems)

Dr. Horst Simon  
Director of NERSC

Dr. Franz-Josef Pfreundt  
Member of Board, ITWM

Berkeley, CA

9/27/2000