Creating Science-Driven Computer Architecture: A New Path to Scientific Leadership

Horst D. Simon, C. William McCurdy William T.C. Kramer - Lawrence Berkeley National Laboratory
Rick Stevens, Argonne National Laboratory
Mike McCoy, Mark Seager, Lawrence Livermore National Laboratory
Thomas Zacharia, Jeff Nichols, Oak Ridge National Laboratory
Ray Bair, Scott Studham, Pacific Northwest National Laboratory
William Camp, Robert Leland, Sandia National Laboratories
John Morrison, Bill Feiereisen, Los Alamos National Laboratory

1. The Challenge: Maintaining American Leadership in High End Computing

We believe that it is critical for the future of high end computing in the United States to bring into existence a new class of computational capability that is optimal for science. In recent years scientific computing has increasingly become dependent on hardware that is designed and optimized for commercial applications. Science in this country has greatly benefited from the improvements in computers that derive from advances in microprocessors following Moore's Law, and a strategy of relying on machines optimized primarily for business applications. However within the last several years, in part because of the challenge presented by the appearance of the Japanese Earth Simulator, the sense has been growing in the scientific community that a new strategy is needed. A more aggressive strategy than reliance only on market forces driven by business applications is necessary in order to achieve a better alignment between the needs of scientific computing and the platforms available.

The United States should undertake a program that will result in scientific computing capability that durably returns the advantage to American science, because doing so is crucial to the country's future. Such a strategy must also be sustainable. New classes of computer designs will not only revolutionize the power of supercomputing for science, but will also affect scientific computing at all scales. What is called for is the opening of a new frontier of scientific capability that will ensure that American science is greatly enabled in its pursuit of research in critical areas such as nanoscience, climate prediction, combustion, modeling in the life sciences, and fusion energy, as well as in meeting essential needs for national security.

The creation of the HECRTF set the right signal. As a nation we must create a new class of computing capability for the country by undertaking the research and development necessary to build supercomputers, the associated software and systems applications environment optimized for science in partnership with the American computer industry, as well as new algorithms and applications that will run on them. The combined strengths of the national laboratories and the university computer science community must be brought to bear on this problem to ensure its success within five years or less. These new computers must be made available to the *entire* U.S. scientific community, especially in the development stages, so as to expose these machines to the rigorous testing that American scientists must be able to give them if there is to be broad-based confidence in their full-scale deployment. Furthermore renewed emphasis must be placed on computer science research in this class of computer design for science, because the computer and applications architectures research communities are critical to the success of its larger mission in science and national security.

We recognize that the computer industry cannot participate in the long term unless sufficient markets exist for these machines. Therefore, while our central goal is faster scientific application time-to-solution, our joint goal is to develop architectures that industry can adapt to other markets to assure the sustainability of this national effort for science.

In this white paper we propose a strategy for accomplishing this mission, pursuing different directions of hardware development and deployment, and establishing a highly capable networking and grid infrastructure connecting these platforms to the broad research community.

2. A Strategy for Creating a New Class of Computer Architectures for Scientific Computing

In the 1980s and early 1990s there were more than twenty U.S. companies producing supercomputers that were designed for scientific and technical applications. Among them were Alliant, BBN, Convex, Cray Research, Cray Computer, IBM, Intel Supercomputer Division, Kendall Square, and Thinking Machines. The primary role of national laboratories and universities at that time was to establish performance requirements, evaluate offerings and select the best for use in scientific research. The market for high-performance scientific and technical computing was a significant focus of the computer industry.

Today the situation is radically different because the market for commercial Web and data servers has grown to completely overshadow the market for high performance scientific computing. Supercomputers in use now largely consist of clusters of commodity servers, connected by networks that have not increased in capability at the same rate as the processors they connect.

Today's situation calls for a strategy that creates a new class of supercomputing machinery by leveraging Moore's Law and the technology that underlies commercially viable computers and the microprocessors they are based on, instead of simply utilizing its existing implementations. Most crucially, this strategy must provide a new way to couple scientific applications requirements to the development of computer architectures, thereby opening a sustainable path to petaflops/s-level performance and beyond.

2.1 Sustained Cooperative Development of New Computer Architectures

We propose a new type of development partnership with computer vendors that goes beyond the evaluation of the offerings that those vendors are currently planning for the next decade. We therefore propose in this paper a comprehensive strategy that includes development partnerships with multiple vendors. Those partnerships will bring to bear:

1. teams of scientists and computational mathematicians who will modify and optimize their applications for future systems through the use of performance modeling, simulators and prototypes of new hardware

2. teams of computer architects from major U.S. computer vendors who will interact directly with the scientific applications teams, and

3. teams of computer scientists who will work with both applications scientists and computer architects to analyze and abstract the requirements of scientific applications so that they can be addressed in hardware and to develop the software environments that will allow scientists to extract the maximum performance and capability from that hardware.

This strategy is directed at challenging and partnering with vendors to create architectures that perform to a target level on a specific suite of scientific applications. Unlike the current approach

that has become standard for both scientific and commercial computing, this new approach does not merely abstract requirements of example applications as the primary mechanism for making performance-related design decisions. That strategy fails for scientific computing because the core algorithms for scientific applications are vastly more diverse than those in commercial applications and because they continue to be combined in new ways as scientific understanding and problem solving approaches progress. The strategy is to balance general-purpose processor architectures that achieve the best average performance on a broad set of applications, and special-purpose hardware that achieves the best performance on a specific set of applications.

This new strategy requires sustained partnerships because the development time for new computer architectures is from three to six years and one may need to iterate over several generations of new hardware to achieve the desired outcome. Because of the long duration of the typical microprocessor development cycle, from design to commercial manufacture, the general purpose microprocessors for 2005 are already designed and in the stages of implementing their manufacture. U.S. vendors cannot commit to significant changes in their development plans without a sustained commitment from their partners such as those in the national laboratories. They cannot succeed with those changes if they are not able to constantly reevaluate and modify their plans in partnership with the scientific community. The national laboratories are a natural home for such partnerships, which will also include extensive participation of the university computer science and applications communities. The laboratories can sustain and manage the partnerships and deliver both software and in-depth analysis of applications performance to meet development milestones.

This strategy accepts that design changes and deviations made by vendors in their architectures from what would otherwise be their mainline business offerings will involve added cost that may or may not be amortized over the business brought by the scientific community. The success of this strategy therefore depends crucially on the balance between the magnitude of government investment in development of the optimizing technologies and the market over which the added costs are to be amortized. Achieving this balance is at the heart of a sustainable strategy.

2. 2 A Focus on Sustained Performance of Scientific Applications

Within a decade we believe it will be possible to provide scientific computing in the U.S. with orders of magnitude increase in sustained performance. The development partnerships we will implement are motivated by the simple goal of achieving maximum sustained performance of scientific applications. The most successful architectures will enable a range of scientific applications to reach those sustained performance levels, but optimization for specific scientific applications is also a benefit of the strategy.

This strategy is not the same as simply "standing up" very large computers. It is intended to change the offerings of computer vendors at all levels, from the laboratory scale, single-principal-investigator scale, to the scale of supercomputers. We propose to change the building blocks of high performance scientific computing and the communications fabrics that enable their integration into large-scale systems. If we can improve applications performance of the building blocks, the nodes that consist of a few processors or tens of processors, to routinely reach sustained performances of 30 to 50 percent of their peak capability, the core productivity of American computational science will increase dramatically at all scales. The ultimate metric for success is time-to-solution.

The current measure of the upper levels of scientific computing is a peak speed of tens of teraflop/s (trillions of floating point operations per second), of which typical applications extract

5 or 10 percent (or less). The strategy we propose here is directed toward the goal of providing peak capabilities at a petaflop/s (1000 teraflop/s) by 2007–2009, of which 30 to 50 percent will be accessible to scientific users.

Doing so could require between 50,000 and 100,000 processors depending on the level of integration achieved. No computer hardware and software architecture known today will scale to that size and provide high levels of sustained performance. The situation can be dramatically improved through greater levels of hardware integration (smaller node counts) and fault tolerant systems and applications software. As the hardware environment is refined, it is equally necessary to emphasize the development of software environments and tools that scale to high processor counts and scientific applications must take into account and accommodate extreme levels of parallelism. The Nation must step up to this challenge, and the strategy we offer can meet it.

2.3 A Strategy to Pursue Multiple Science-Driven Architectures

We propose a comprehensive strategy that makes use of the entire complex of national laboratories and investment in university scientific and computer science communities. The proposed initial expression of the program will be in several sites. Even in this initial implementation, we propose a national collaboration among these initial platform sites, and between the platform sites and the computational science and computer science community.

Each of the platform sites would engage in a significant development partnership with the vendor that will bring these new architectures into existence in a form, at a scale, and on a schedule that the scientific community can exploit successfully. Several options for such partnerships already exist, and could be leveraged on in the future. These options are being pursued at some level today at the institutions represented by the authors of this white paper and constitute viable new architecture directions, but they are not the only viable options. These options are:

- Using custom components at all levels in an architecture known to be successful in scientific applications, *parallel vector processing*. The initial stages of this effort have started in 2003 with the evaluation of the Cray X1 at Oak Ridge National Laboratory. The follow-on X2 will be available in 2006 and a hybrid architecture that combines vector processors and superscalar processors will be capable of peta-scale computing around 2008.

- Using modified commercial microprocessors in a new architecture that will provide much better memory and communications performance and will be programmable in the same way as the first option, *ViVA* or *Virtual Vector Architecture*. IBM has partnered with Lawrence Berkeley National Laboratory and Lawrence Livermore National Laboratory to implement early versions of this architecture and deliver "Blue Planet", a 160 teraflop/s mature implementation in the second half of 2005.

- Exploiting the "system-on-a-chip" architecture that is being explored most visibly in the IBM Blue Gene project. This architecture is arguably the most promising for reaching the petaflop/s goal of this proposal; however, its suitability for general scientific use has not yet been demonstrated. This 4path is extremely cost-effective to pursue and provides the best long-term bet currently known to the scientific community. LLNL will take delivery of the first Blue Gene/L platform in 2004 and make it available to the research community and its development partners, in particular Argonne National Laboratory.

- Expanding on the Hewlett-Packard / Intel relationship to modify industry standard processors to meet the needs of the scientific community. By making minor modifications to commodity processors it is possible to provide an economic solution to meet the high performance

computing needs of the scientific community. Pacific Northwest National Laboratory, HP and Intel have established a roadmap as part of the DARPA High Productivity Computing System Program that will be capable of sustaining petascale computing by the end of the decade.

- Realizing the scalability inherent in the Red Storm architecture developed at Sandia National Laboratories in partnership with Cray and AMD. The 10 thousand processor, 40 teraflops system planned for mid 2004 is designed to be upgradeable to a 30 thousand processor, 180 teraflops system that retains favorable architectural balance and strong price performance using design principles proven over several machine generations. These include a very high bandwidth, low latency network based on minimal custom design, and non-intrusive system software tailored to support very high performance.

- Exploring at Los Alamos with other government agencies and departments the promise of leveraging the extremely high volume consumer and embedded processor markets. This effort is in conjunction with projects to develop high bandwidth low latency electrical or optical interconnects in partnership with American industry and vendors.

Further opportunities that represent distinct architecture directions, all of which could be elaborated into more than one generation of science-driven computing machinery include

- Exploration of alternative technologies that are in development at universities today

- Leveraging the new DARPA High Productivity Computing Systems program and the architectures developed by the participants Cray, HP, IBM, SGI, and SUN.

- Specific vendor discussions currently taking place under non-disclosure agreements (NDA).

2.4 Embedding High End Capabilities in a National Grid Infrastructure

Regardless of which architectures are deployed in the next three years to enable the U.S. scientific community to maintain international leadership in computational science, it will be critical to integrate these new computing platforms with the emerging computational infrastructure that provides integrated access to large-scale data, computing platforms, instruments and users. It is critical that these systems are deployed in a manner that creates a national footprint and fully integrates them with the emerging research Grid.

A number of the scientific disciplines require a new scale of networking connectivity to exploit this new frontier of computing. The strength of American science is due in no small part to the infrastructure that supports it, and networking technology is rapidly reaching a level of performance that will change fundamentally our expectations of accessibility of our computational resources and our ability to move tasks and data between them transparently. We therefore recommend a dramatic step forward in the deployment of a national high-performance networking infrastructure that complements the advanced architectures that will be developed.

2.5 A New Investment in the Computer Science Research and Scientific Research Communities

Two other classes of investment are critical to the success of this proposed new program. They must not be neglected in the initial description of the initiative to ensure leadership in computing for science.

- A long-term program of computer science research in computer architecture for science that reestablishes a viable national university research community in this discipline as well as a core capability in this area in the national laboratories.

- An investment across the entire range of the computational science research portfolio to enable researchers funded by participating agencies to make use of new architectures as they become available as prototypes and production installations.

In the 1980s and early 1990s there were roughly 50 groups of computer scientists in universities across the United States investigating new computer architectures specifically directed at high performance scientific computing. They were primarily funded by DARPA as part of the High Performance Computing Initiative. Today there are only a handful of such groups, and their interest, like that of the larger computer science community, is turning primarily to new technologies that exploit Web and Grid computing as well as a pervasive digital infrastructure for the country.

Today we must accept the responsibility of reestablishing a national research community in this discipline if it is to provide a sustainable path to leadership in scientific computing. University researchers must be able to depend on stable interest from a funding agency in this arena if they are to devote their careers to it. Several funding agencies have related programs, but this will be a new effort requiring new investments and program structures.

Within DOE the first steps in this direction were taken by the SciDAC initiative. Those were only small steps by comparison with what is required to bind the national scientific research communities, together with the applied mathematics community, to the project of creating new computing architectures for science. Those communities are ready for this challenge, but the effort required to explore multiple architectures and programming models is well beyond their present resources. This initiative must be driven by scientific applications and managed to use them to drive the creation of new computing resources.

Corresponding author:

Horst D. Simon
Mail Stop 50B-4230
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
(510) 486-7377
hdsimon@lbl.gov