Analysis

A strategic view of SciDAC-2: growth ahead

Dr Horst Simon and Dr Rick Stevens

During the past five years, the SciDAC-1 program from the the US Department of Energy (DOE) has shown that important scientific accomplishments are possible through simulation and modeling with the focused collaboration and active partnership of domain scientists, applied mathematicians, and computer scientists.

Successes have been seen in fields such as accelerator design, chemistry, combustion, climate modeling, and fusion. SciDAC activities have demonstrated that large-scale simulation offers a cost-effective way of answering a number of scientific questions on the fundamental structure of matter, the production of heavy elements in supernovae, and the functions of enzymes, for instance.

The approach to scientific enterprise in the US is in the midst of two fundamental transformations. One is the increasing application of computational simulation to long-standing challenge problems. The other is the rapid adoption of advanced infrastructure for the capture, storage, transmission, sharing, and analysis of large-scale experimental data.

Together, these two technologies are dramatically improving our nation's ability to solve important engineering problems and make critical discoveries in many scientific domains. Such advances were

PHOTOLIBRARY

Climate modeling is just one of the many areas that SciDAC-2 will focus on.

considered intractable in the past because of their extreme complexity, multidisciplinary nature, or lack of data-analysis capability.

In spite of tremendous progress, however, much remains to be done, and even more compelling opportunities for scientific discovery lie ahead. In FY 2004, the DOE's Office of Science (SC) launched an aggressive program to develop and deploy leadership-class computing facilities and announced a 20-year scientific facilities roadmap that will provide a rich scientific infrastructure for the next two decades.

Each of the SC programs — Basic Energy Sciences, Biological and Environment Research, Fusion Energy Sciences, High-Energy Physics, and Nuclear Physics — has identified a crucial need for sustained advances in scientific computing. The creation of additional SciDAC-like scientific application partnerships is required for fulfillment of these programs' missions.

To address these needs, it was recommended that the SC's Office of Advanced Scientific Computing Research (ASCR) make significant investments for FY 2007 to establish a SciDAC-2 program. SciDAC-2 can build on the success of SciDAC-1 and extend the program in three important ways:

- Strengthen the scientific application partnerships that formed the core of the successful SciDAC-1 program. The major source of acceleration in simulation-based science has been the strength and depth of partnerships among application domains, computer science, and applied mathematics.
- Involve experimental science. The use of advanced computing technologies to accelerate scientific discovery is not limited to modeling and simulation; it can also be used to improve experimental science. New data-management and analysis tools are required for both small and large-scale experiments such as those that will be carried out in the facilities outlined in the SC's 20-year roadmap. SciDAC-2 should improve the productivity and capability of experimental science through the development and application of advanced data and analysis capabilities, and computation and technologies that support and automate experiments.
- Expand to new scientific communities. Some areas of science do not yet fully benefit from large-scale computing; some communities are not fully aware of successes and methods from other scientific areas; and in other cases methods, tools, or conceptual approaches are relatively immature or inadequately developed. SciDAC-2 should improve outreach activities, for instance by creating SciDAC Institutes aimed at the broad inclusion of new communities, and improving workforce development to ensure a robust supply of next-generation researchers.

It was also recommended that these investments should be supported by computational and experimental science application investments from DOE programs and by additional ASCR infrastructure investments in leadership-class computing facilities, the National Energy Research Scientific Computing Center, and the Energy Sciences Network. The resulting research portfolio will accelerate progress in advanced energy systems, biotechnology, nanotechnology, and environmental modeling.

Dr Horst Simon is Associate Laboratory Director for Computing Sciences, LBNL.
Dr Rick Stevens is Acting Associate Laboratory Director for Physical, Biological and Computing Sciences, Argonne.

Perspectives and funding

SciDAC-2: on the path to petascale computing

Dr Walter M. Polansky

The driving force behind the initiation of the SciDAC program was a growing perception that modeling and simulation can advance the understanding of scientific challenges. Through SciDAC, perception became reality. Researchers developed state-of-the-art simulation codes to run on DOE supercomputers and gain substantial insight into the areas of accelerator design, chemistry, combustion, climate, and fusion.

The key to the overall success of SciDAC has been the unleashing of the power of multidisciplinary teams; experts in the relevant scientific discipline, computer scientists, and applied mathematicians have come together. The teams were able to chart courses for progress that could not be traversed in any other way, and forge a new paradigm for simulation science. In addition, SciDAC built a unique foundation in applied mathematics, computer science and high-performance computing technology that could lead to further scientific advances.

The research landscape is currently being shaped in two areas, with profound implications. One area is enabled by the increasing application of modeling and simulation to long-standing scientific problems. The second area is driven by the rapid adoption and adaptation of advanced technologies for data storage and for data transfer associated with large-scale experimental facilities or massive data sets.

The SciDAC program is in a position to help shape that research landscape. Scientific application partnerships could be enriched in existing areas and introduced to other disciplines. The need for scientific data storage and data transmission has outstripped current methodologies; innovative approaches could emerge from research partnerships that marry hardware with the highly specialized needs of scientific users. The partnership concept could reach out to engage a broader community of scientists in the activities of scientific discovery through advanced computation. Finally, mechanisms could be established to bring together a "critical mass" of experts from multiple disciplines to tackle problems as they emerge within a discipline.

SciDAC-2 has been formulated to capitalize on these opportunities. As envisioned, SciDAC-2 will be an ensemble of research partnerships whose activities and results will be coordinated to leverage advances and optimize overall progress along two major lines:

- The creation of a comprehensive, scientific computing software infrastructure that fully integrates applied mathematics, computer science, and computational science in the physical, biological, and environmental sciences for scientific discovery at the petascale.
- A new generation of data management and knowledge discovery tools for data-intensive science, including data sets obtained from experimental facilities and from high-end simulations.

SciDAC-2 seeks to advance scientific discovery through computing in 13 areas of science: accelerator science and simulation, astrophysics, climate modeling and simulation, computational biology, fusion science, groundwater reactive transport modeling and simulation, high-energy physics, high-energy and nuclear physics with petabytes, materials science and chemistry, nuclear physics, quantum chromodynamics, radiation transport, and turbulence.

Advances through simulations in these 13 topical science areas will be provided via enabling computational technologies, which have been structured into the following three categories:

- The Centers for Enabling Technologies, which will address mathematical and computing systems software issues.
- The Scientific Applications Partnerships, which seek to fund activities that form a partnership between computational mathematics and computer science with a science application domain.
- SciDAC Institutes, which are university-led centers of excellence intended to complement the Centers for Enabling Technology and the Scientific Applications Partnerships, as well as centers that may be formed under specific science application domains. A key feature of a SciDAC Institute will be a dimension of training and outreach in high-performance computing topics, including topics for graduate students and post-doctoral staff.

The Notice inviting grant applications (www.sc.doe.gov/grants/FAPN06-04.html) and the Program Announcement to DOE National Laboratories (www.sc.doe.gov/grants/LAB06_04.html) were issued on December 23, 2005. The Notice and the Program Announcement describe the scientific challenges. The research community will provide the ideas, through proposals and grant applications, for meeting those challenges. SciDAC-2 awards will be made in the summer of 2006 following a thorough peer review.

Under the SciDAC-2 calls for proposals, letters of intent were due on January 23, 2006. Proposals and grant applications are due on March 6, 2006. The overall funding available for SciDAC-2 is approximately $70,000,000 this fiscal year: $66,000,000 from the Office of Science (SC) and $4,000,000 from the Department's National Nuclear Security Administration. All SC programs are participating in SciDAC-2: these are Advanced Scientific Computing Research; Basic Energy Sciences, Biological and Environmental Research; Fusion Energy Sciences, High Energy Physics; and Nuclear Physics. In addition, the National Science Foundation is a potential funding source for grant applications in the areas of climate modeling, high-energy physics, and nuclear physics.

Collaborative research proposals are not only encouraged, they are expected. The majority of the applications and proposals submitted are expected to be for scientific simulation teams with several institutions. The number, size and duration of each will be determined by the quality and content of the proposals selected for funding. However, to meet the goals of SciDAC-2, a relatively small number of awards is expected.

SciDAC-2 represents a commitment to propel simulation science to the petascale and to broaden the prospects for scientific advancement. The scientific challenge problems are there to be solved and petascale computers are on the horizon. The final terms in the SciDAC-2 "equation for success" represent the dedicated research scientists, with outstanding credentials, ready to partner with colleagues in other disciplines to bring about scientific discovery through simulation and modeling.

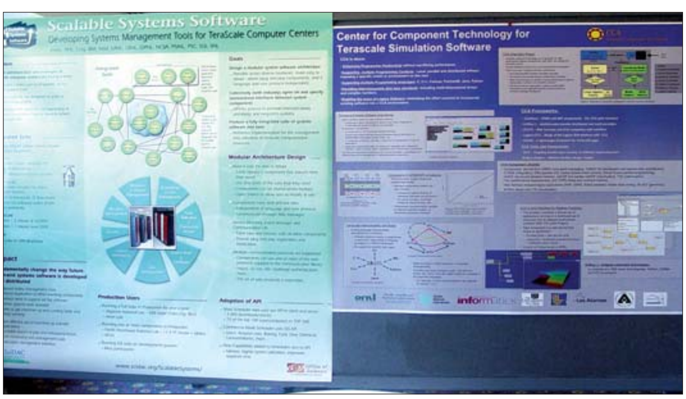
Dr Walter M. Polansky is Senior Technical Advisor for Project Management at the Office of Advanced Scientific Computing Research, Office of Science, DOE.

Conference: **SciDAC 2005**

Meeting grows in scope as project progresses

Dr Anthony Mezzacappa

On June 26–30, 2005, at the Grand Hyatt on Union Square in San Francisco, several hundred computational scientists from around the world came together to celebrate computational science. Scientists from the SciDAC program as well as other agencies and nations were joined by applied mathematicians and computer scientists to highlight the many occasions in the past year on which computation has successfully led to scientific discovery in a variety of fields: lattice quantum chromodynamics, accelerator modeling, chemistry, biology, materials science, Earth and climate science, astrophysics, and combustion and fusion energy science.



Sharing science at SciDAC 2005: the poster session at the conference.

The advances made in numerical methods and computer science, and the multidisciplinary collaboration cutting across science, mathematics, and computer science that enabled these discoveries, were also highlighted.

The scope and import of the meeting was well summarized in the opening remarks of Dr Michael Strayer, Associate Director, Office of Science (ASCR) and Director, SciDAC: "SciDAC is also undergoing a transformation. This meeting is a prime example. Last year it was a small programmatic meeting tracking progress in SciDAC. This year, we have a major computational science meeting with a variety of disciplines and enabling technologies represented. SciDAC 2005 should position itself as a new corner stone for Computational Science and its impact on science."

The scientific program began with one of the DOE's great traditions and core missions: energy research. Computation has been seminal to the critical advances that have been made in this arena. Of course, deciphering our world, whether for its own sake or for practical purposes, will require explorations on all of its scales. Computational science has been an important tool in this arena. From explorations of quantum chromodynamics — the fundamental theory that describes how quarks make up the protons and neutrons of which we are composed — to explorations of the complex biomolecules that are the building blocks of life and explorations of some of the most violent phenomena in the universe, computation has provided not only significant insight, but often the only means by which we have been able to explore and begin to understand these complex, multicomponent systems.

While our ultimate target remains scientific discovery, at a fundamental level the world is mathematical. Equations ultimately govern the evolution of the systems of interest to us, be they physical, chemical, or biological systems. The development and choice of discretizations of these underlying equations is often a critical deciding factor in whether or not one is able to model such systems stably, faithfully, and practically, while the algorithms that solve the resultant discrete equations are the complementary, critical ingredient in the recipe to model the natural world. The use of parallel computing platforms, especially at the terascale, and the trend toward even larger numbers of processors, continue to present significant challenges in the development and implementation of these algorithms.

Computational scientists often speak of their "workflows." A workflow, as the name suggests, is the sum total of all complex and interlocking tasks, from simulation set up, execution, and I/O, to visualization and scientific discovery. For the computational scientist, enabling such workflows presents myriad, significant challenges, and it is computer scientists who are called upon at such times to address these challenges. Simulations are currently generating data at the staggering rate of tens of terabytes per simulation, over the course of days. In the next few years, these data-generation rates are expected to climb exponentially to hundreds of terabytes per simulation, performed over the course of months. The output, management, movement, analysis, and visualization of these data will be our key to unlocking their meaning. And there is no hope of generating such data to begin with, or of scientific discovery, without stable computing platforms and a sufficiently high and sustained performance of scientific applications codes on them.

Thus, scientific discovery in the realm of computational science at the terascale and beyond will occur at the intersection of science, applied mathematics, and computer science. The SciDAC program was constructed to mirror this reality, and the 2005 conference was just the beginning. ● *This year's conference, SciDAC 2006, will be held on June 25–30 in Denver, Colorado. Chairman: Bill Tang, PPPL.*

Dr Anthony Mezzacappa is Group Leader for Theoretical Astrophysics, Physics Division, ORNL, and Chair, SciDAC 2005.

For more information

SciDAC 2005 Conference Proceedings
www.iop.org/EJ/toc/1742-6596/16/1.

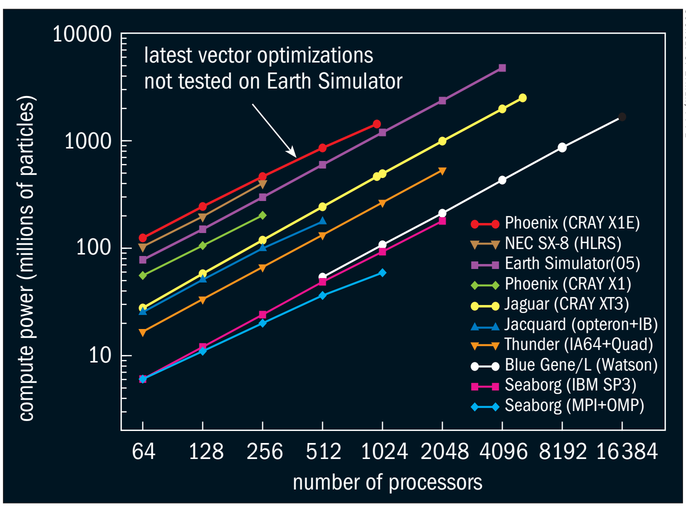
SciDAC 2006 home page
www.scidac.org/Conference2006.

Computing highlights

The GTC code scales new heights

Dr Stephane Ethier, PPPL

Owing to the introduction of a new multi-level Message Passing Interface (MPI)-based parallel algorithm in which both the simulation volume and the particles are split between processors, the gyrokinetic toroidal code (GTC; see feature "Simulating star power on Earth," p40) scales easily to several thousand processors on the most powerful parallel supercomputers available, including the IBM Blue Gene/L and modern parallel-vector systems such as the CRAY X1E and the Earth Simulator in Japan.



SOURCE: S. ETHIER, PPPL

Compute power for simulations versus number of processors; data Nov. 2005.

Unlike most of the other gyrokinetic codes used to study plasma turbulence in toroidal fusion devices, the GTC solves all of its equations in real space, rather than in Fourier (spectral) space. While this introduces some algorithmic challenges, it has the major advantage of avoiding the heavy global communication between processors that is necessary for full nonlinear turbulence calculations when carried out in Fourier space.

This has enabled the GTC to demonstrate impressive scaling capabilities — running on more than 16,000 processors on the IBM Blue Gene/L computer — and perform very high-resolution simulations, pushing 13 billion particles at 7.2 Tflop on 4,096 processors on the Earth Simulator.

Dr Stephane Ethier is a Computational Physicist in CPPG at the Princeton Plasma Physics Laboratory.

Conferences

SC|05 is a window on DOE research

William Kramer

The 2005 Supercomputing Conference (SC|05) was held in November in Seattle and provided a unique view of the thriving high-performance computing community. More than 9,700 scientists, engineers, researchers, educators, senior managers, programmers, and system managers from the world's leading computing installations and companies came to showcase innovative developments that are sparking new ideas and industries, as well as reinvigorating older ones. For the first time, SC|05 brings a focus on analytics — the software that, along with the computing storage and networking resources, solves large-scale real-world problems.

SC|05 had the most competitive technical program on record, featuring the latest developments in fields such as computational systems and information architectures. With 62 papers, a series of major plenary sessions and over 20 tutorials, the 2005 conference was one of the strongest ever.

In addition to high-performance computing, networking, and storage, SC|05 introduced a focus on HPC analytics, which brings physical computing, storage, and networking resources together with software and algorithms to produce new insights and tools. Its uses include analyzing tremendous amounts of data for everything from genomics to astrophysics and cybersecurity.

The technical sessions started with a keynote by Bill Gates, who talked about how computers are changing the way in which science is being done. Gates talked of how data acquisition and data analysis are intricately connected through a series of remote sensing networks to massively parallel systems for simulation.

Dozens of remote locations throughout the world were connected in real time over the Internet to showcase achievements in the arts and sciences. And SC Desktop enabled the entire technical program to be available to anyone at their desktop. More than 140 teachers participated in the educational program.

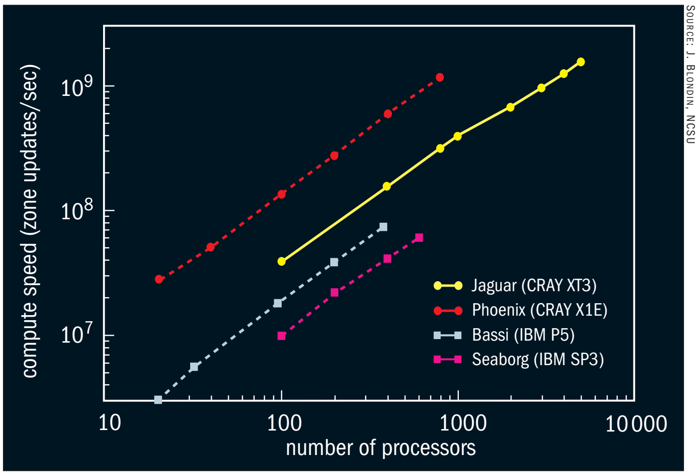
SC|05 was an exciting event for the DOE community. Researchers from the DOE, including many involved with SciDAC, participated in 10 of the 20 tutorials, 23 papers, 12 posters and three of six panels. DOE staff were involved with two SCinet bandwidth challenges, three Storcloud challenges, and three HPC analytics challenges.

Dr John Bell of LBNL, who is involved with SciDAC, received the prestigious Sid Fernbach award. A team of researchers from LLNL and IBM won the Gordon Bell prize for their paper "100+ TFlop Solidification Simulations on BlueGene/L." DOE researchers were on teams that won the Storcloud, the SCinet Bandwidth and the HPC Analytics challenges.

William Kramer is NERSC Center General Manager and General Chair, SC|05.

Computing highlights

Studying supernova shock waves



The compute speed of the VH-1 hydrodynamics code used to study the dynamics of the stalled supernova shock wave in three dimensions in the TSI project.

Dr John Blondin

This plot shown in the graph left represents a "weak" scaling study, in which the computational workload (i.e. the size of the problem) is increased along with the number of processors. The test problem that was used involves a strong shock wave propagating across the volume of a cartesian box.

A typical TSI run uses of the order of one billion zones, and a full time update can be performed in roughly 1 s on either of the new Cray platforms.

Dr John Blondin is Professor of Physics at North Carolina State University.

Funding

The INCITE program builds on success in 2005

Barbara Helland

As one of the world's leading sponsors of scientific research, the US Department of Energy (DOE) launched a major program in 2003 to allocate millions of hours of supercomputing time to address some of the most challenging research problems in physics, chemistry, genetics, and energy. Called Innovative and Novel Computational Impact on Theory and Experiment (INCITE), the program seeks computationally intensive large-scale research projects each year that can make high-impact scientific advances via a substantial allocation of computer time and data storage. Such large allocations are critical to advancing understanding in areas such as astrophysics, global climate change, fusion energy and combustion.

The INCITE program specifically encourages proposals from non-DOE researchers, including researchers in industry.

In the first two years of the competitive program, the DOE Office of Science received and reviewed 75 proposals, requesting a total of nearly 160 million hours of processing time. About two-thirds of the proposals came from university researchers.

In 2003, three projects received a total of 4.9 million processor hours at the National Energy Research Scientific Computing Center (NERSC) at Lawrence Berkeley National Laboratory to undertake studies to advance our understanding of the composition of the universe; the chemical process by which plants convert sunlight to energy while removing carbon dioxide from the atmosphere; and the turbulent forces that affect everything from weather to industrial processes.

In INCITE's second year, the selected projects received a total of 6.5 million processor hours at NERSC and aimed to gain greater insight into how stars and solar systems form; look into ways to reduce pollution; and advance our knowledge about how proteins express genetic information.

In 2005, building on past successes, INCITE expanded a single supercomputing facility at LBNL to include five supercomputers at four DOE labs in Berkeley, CA; Chicago, IL; Oak Ridge, TN; and Richland, WA. Specifically, 10% of the Cray Leadership-class computers at Oak Ridge National Laboratory (ORNL); 10% of the IBM Power 3 at LBNL; and 10% of the resources on the IBM BlueGene machine at Argonne National Laboratory, together with 5% of the HP MPP system at Pacific Northwest National Laboratory, were dedicated to the expanded INCITE program. And in a new collaboration between IBM and ANL, IBM agreed to augment Argonne's INCITE computer capacity with compute cycles on IBM's Blue Gene system at the IBM T. J. Watson Research Center in Yorktown Heights, NY.

As a result of the call for proposals, 43 computationally intensive, large-scale research projects were submitted, requesting over 95 million processor hours. In January, 2006, the DOE awarded 18.2 million hours of computing time across the four sites to 15 INCITE projects. For the first time ever, four of the proposals receiving awards were from industry: Boeing; Dreamworks Animation; General Atomics; and Pratt Whitney.

Academic, research institutions and others to receive computing time are: Auburn University; California Institute of Technology; Fisk University, Harvard University; Howard Hughes Medical Institute; Rollins College; Tech-X Corp.; University of Alaska, Fairbanks; University of California, Berkeley; University of California, Davis; University of California, San Diego; University of Colorado; University of Strathclyde; and University of Washington. Scientists at the Lawrence Berkeley, Lawrence Livermore, Oak Ridge, and Los Alamos National Laboratories will also receive computing time.

Those receiving INCITE awards in the coming year will use their allocations to work on diverse projects, including:

- improving aircraft efficiency
- learning about disease
- advancing fusion power
- studying climate change
- developing stronger materials
- simulating molecular collisions
- developing computing tools
- studying water and how light affects it in biological systems
- modeling protein structure
- demystifying dark energy
- simulating particle accelerators.

The call for 2007 INCITE proposals is expected to be released in mid-spring, 2006. For more information about the INCITE program and the 2006 awardees, see http://hpc.science.doe.gov.

Barbara Helland is Program Manager, Office of Advanced Scientific Computing Research.