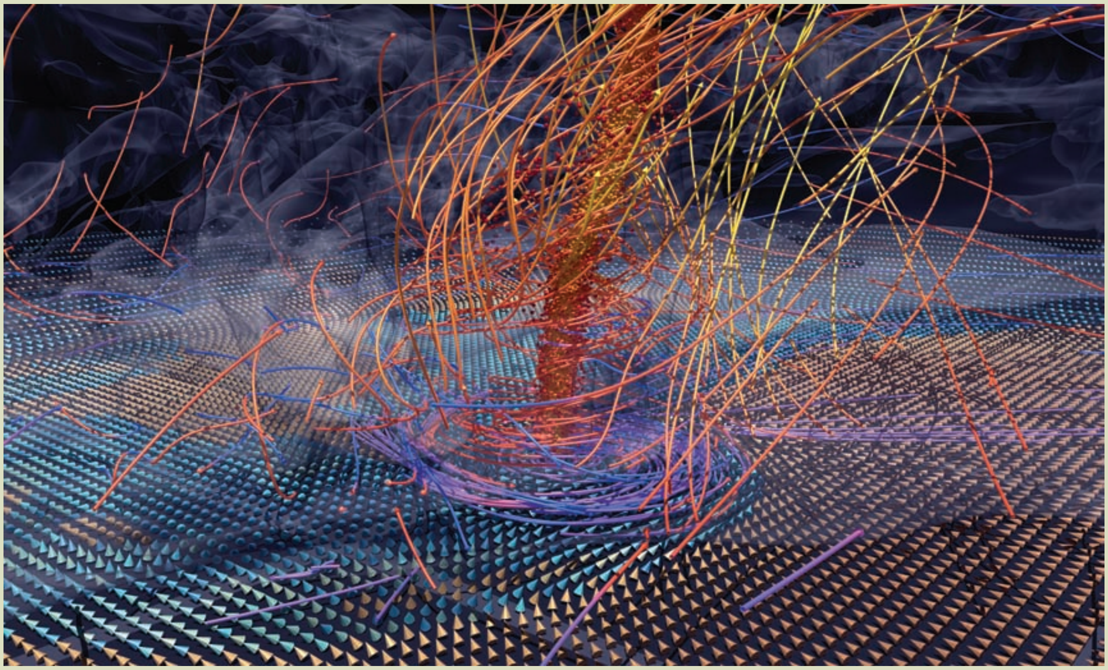
# CHAPTER 2 HIGH PERFORMANCE COMPUTING (2006-2010)

## I. What Does High Performance Computing Offer Science and Engineering?

What are the three-dimensional structures of all of the proteins encoded by the human genome, and how does structure influence their function in a human cell? What patterns of emergent behavior occur in models of very large societies? How do massive stars explode and produce the heaviest elements in the periodic table? What sort of abrupt transitions can occur in Earth's climate and ecosystem structure? How do these transitions occur, and under what circumstances? If we could design catalysts atom-by-atom, could we transform industrial synthesis? What strategies might be developed to optimize management of complex infrastructure systems? What kind of language processing can occur in large assemblages of neurons? Can we enable integrated planning and response to natural and man-made disasters that prevent or minimize the loss of life and property? These are just some of the important questions that researchers wish to answer using contemporary tools in a state-of-the-art High Performance Computing (HPC) environment.

Using HPC-based applications, researchers study the properties of minerals at the extreme temperatures and pressures that occur deep within the Earth. They simulate the development of structure in the early Universe. They probe the structure of novel phases of matter such as the quark-gluon plasma. HPC capabilities enable the modeling of life cycles that capture interdependen-



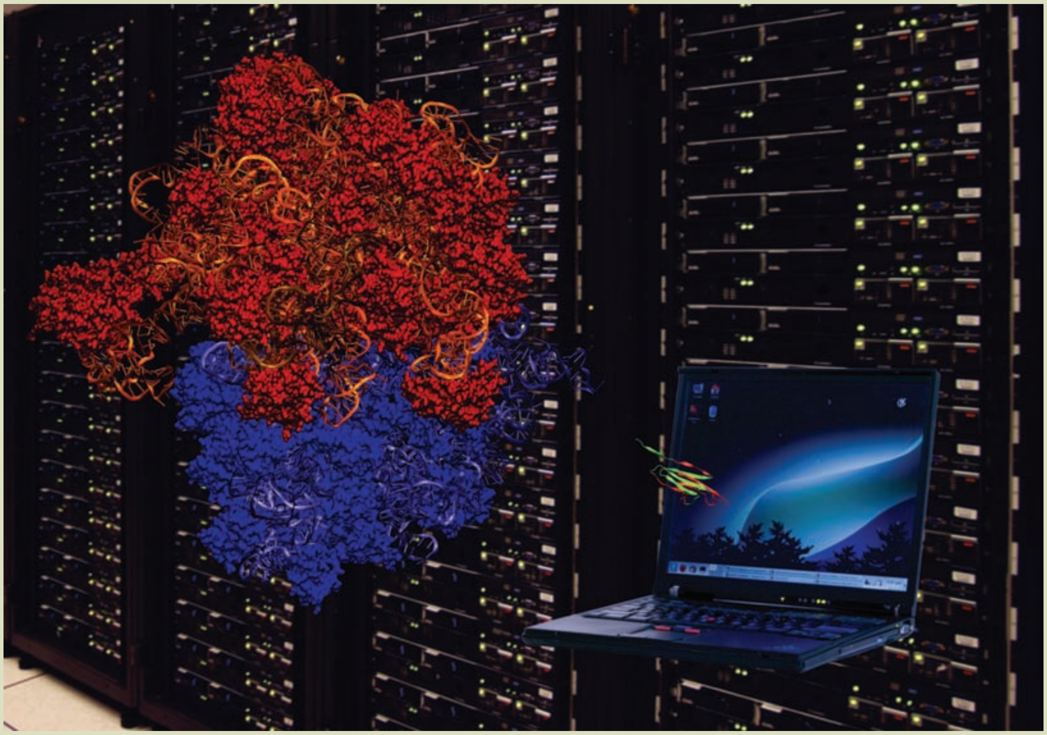
*The visualization above, created from data generated by a tornado simulation calculated on the NCSA computing cluster, shows the tornado by spheres colored according to pressure. Orange and blue tubes represent the rising and falling airflow around the tornado.*

*NCAR's blueice supercomputer, shown on the opposite page, enables scientists to enhance the resolution and complexity of Earth system models, improve climate and weather research, and provide more accurate data to decision makers.*

cies across diverse disciplines and multiple scales to create globally competitive manufacturing enterprise systems. And they examine the way proteins fold and vibrate after they are synthesized inside an organism. In fact, sophisticated numerical simulations permit scientists and engineers to perform a wide range of in silico experiments that would otherwise be too difficult, too expensive, or impossible to perform in the laboratory.

HPC systems and services are also essential to the success of research conducted with sophisticated experimental tools. Without the waveforms produced by the numerical simulation of black hole collisions and other astrophysical events, gravitational wave signals cannot be extracted from the data produced by the Laser Interferometer Gravitational Wave Observatory. High-resolution seismic inversions from the higher density of broad-band seismic observations furnished by the Earthscope project are necessary to determine shallow and deep Earth structure. Simultaneous integrated computational and experimental testing is conducted on the Network for Earthquake Engineering Simulation to improve seismic design of buildings and bridges. HPC is essential to extracting the signature of the Higgs boson and supersymmetric particles – two of the scientific drivers of the Large Hadron Collider – from the petabytes of data produced in the trillions of particle collisions.

Science and engineering research and education enabled by state-of-the-art HPC tools have a direct bearing on the nation's competitiveness. If investments in HPC are to have a long-term impact on problems of national need, such as bioengineering, critical infrastructure protection (for example, the electric power grid), health care, manufacturing, nanotechnology, energy, and transportation, then HPC tools must deliver high performance capability for a wide range of science and engineering applications.

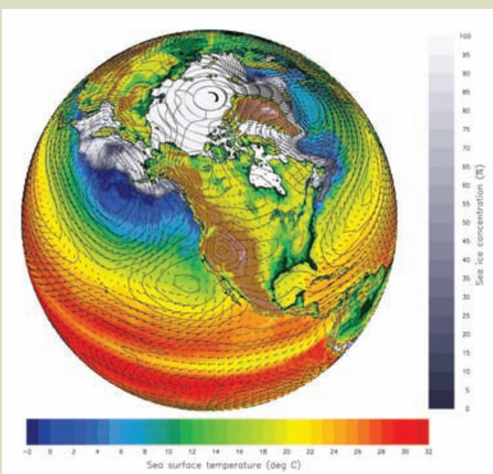


*A functioning ribosome, a complex of three large RNA molecules and fifty proteins with three million atoms, is simulated on the Texas Advanced Computing Center computer.*

## II. The Next Five Years: Creating a High Performance Computing Environment for Petascale Science and Engineering

NSF's five-year HPC goal is to enable petascale science and engineering through the deployment and support of a world-class HPC environment comprising the most capable combination of HPC assets available to the academic community. The petascale HPC environment will enable investigations of computationally challenging problems that require computers operating at sustained speeds on actual research codes of $10^{15}$ floating point operations per second (petaflops) or that work with extremely large data sets on the order of $10^{15}$ bytes (petabytes).

Petascale HPC capabilities will permit researchers to perform simulations that are intrinsically multi-scale or that involve multiple simultaneous reactions, such as modeling the interplay among genes, microbes, and microbial communities and simulating the interactions among the ocean, atmosphere, cryosphere and biosphere in Earth systems models. In addition to addressing the most computationally challenging demands of science and engineering, new and improved HPC software services will make supercomputing platforms supported by NSF and other partner organizations more efficient, more accessible, and easier to use.




*Results from the Parallel Climate Model, prepared from data in the Earth System Grid, depict wind vectors, surface pressure, seas surface temperature and sea ice concentration.*

NSF will support the deployment of a well-engineered, scalable, HPC infrastructure designed to evolve as science and engineering research needs change. It will include a sufficient level of diversity, both in architecture and scale of deployed HPC systems, to realize the research and education goals of the broad science and engineering community. NSF's HPC investments will be complemented by its simultaneous investments in data analysis and visualization facilities essential to the effective transformation of data products into information and knowledge.

The following principles will guide the agency's FY 2006 through FY 2010 investments:

- Science and engineering research and education priorities will drive HPC investments.
- Collaborative activities involving science and engineering researchers and private sector organizations are needed to ensure that HPC systems and services are optimally configured to support petascale scientific computing.
- Researchers and educators require access to reliable, robust, production-quality HPC resources and services.
- HPC-related research and development advances generated in the public and private sectors, both domestic and foreign, must be leveraged to enrich HPC capabilities.
- The development, implementation and annual update of an effective multi-year HPC strategy is crucial to the timely introduction of research and development outcomes and innovations in HPC systems, software and services.

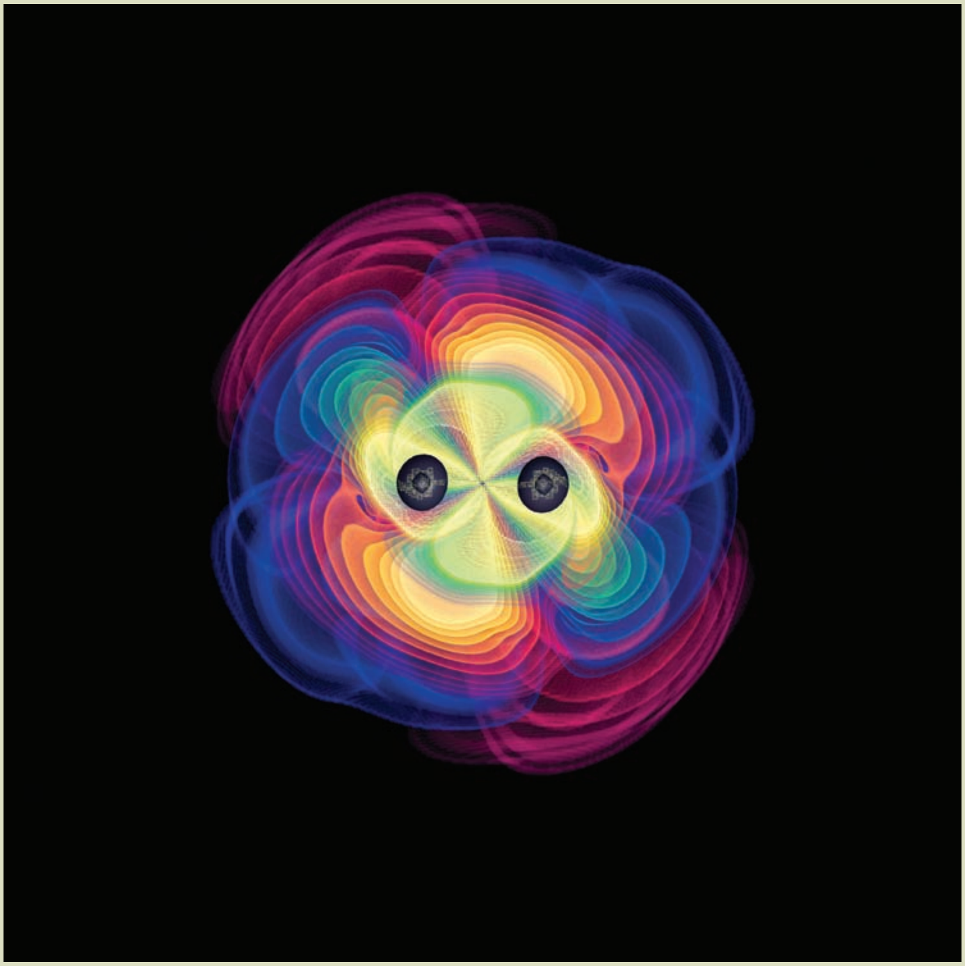
NSF's implementation plan to create a petascale environment includes the following three interrelated components:

### *1). Specification, Acquisition, Deployment and Operation of Science-Driven HPC Systems Architectures*

An effective computing environment designed to meet the computational needs of a range of science and engineering applications will include a variety of computing systems with complementary performance capabilities. By 2010, the petascale computing environment available to the academic science and engineering community is likely to consist of: (i) a significant number of systems with peak performance in the 50-500 teraflops range, deployed and supported at the local level by individual campuses and other research organizations; (ii) multiple systems with peak performance of 500+ teraflops that support the work of thousands of researchers nationally; and, (iii) at least one system capable of delivering sustained performance approaching $10^{15}$ floating point operations per second on real applications that consume large amounts of memory, and/or that work with very large data sets projects that demand the highest levels of computing performance. All NSF-deployed systems will be appropriately balanced and will include core computational hardware, local storage of sufficient capacity, and appropriate data analysis and visualization capabilities.

Over the FY 2006-2010 period, NSF will focus on HPC system acquisitions in the 100 teraflops to 10 petaflops range, where strategic investments on a national scale are necessary to ensure international leadership in science and engineering. Since different science and engineering codes may achieve optimal performance on different HPC architectures, it is likely that by 2010 the NSF-supported HPC environment will include both loosely coupled and tightly coupled systems, with several different memory models.

To address the challenge of providing the research community with access to a range of HPC architectures within a constrained budget, a key element of NSF's strategy is to participate in resource-sharing with other federal agencies. A



*This numerical simulation, created on the NCSA Itanium Linux Cluster by international researchers, shows the merger of two black holes and the ripples in space time that are born of the merger.*

strengthened interagency partnership will focus, to the extent practicable, on ensuring shared access to federal leadership-class resources with different architectures, and on the coordination of investments in HPC system acquisition and operation. The Department of Energy's Office of Science and National Nuclear Security Administration have very active programs in leadership computing. The Department of Defense's (DOD) High Performance Computing Modernization Office (HPCMOD) provides HPC resources and services for the DOD science and engineering community, while NASA is deploying significant computing systems that are also of interest to NSF PIs. NSF will explore enhanced coordination mechanisms with other appropriate federal agencies to capitalize on their common interests. It will seek opportunities to make coordinated and collaborative investments in science-driven hardware architectures in order to increase the diversity of architectures of leadership class systems available to researchers and educators around the country, to promote sharing of lessons learned, and to provide a richer HPC environment for the user communities supported by each agency.

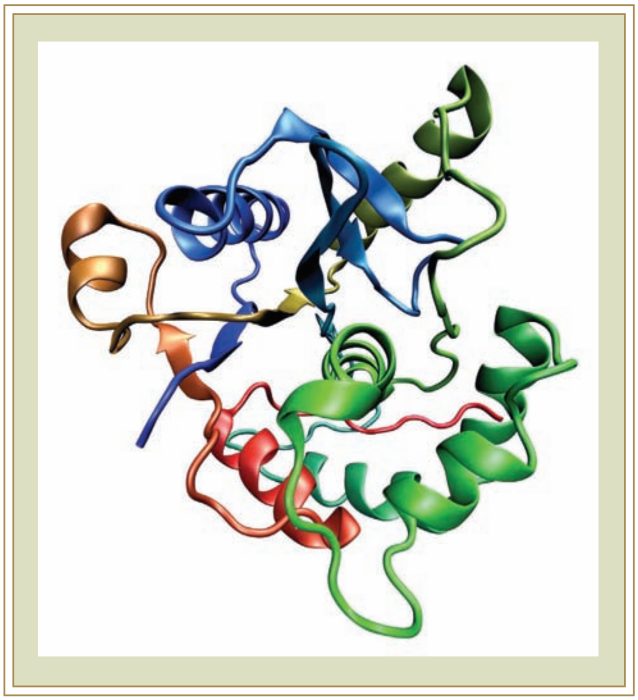
Strong partnerships involving universities, industry and government are also critical to success. NSF will also promote resource sharing between and among academic institutions to optimize the accessibility and use of HPC assets deployed and supported at the campus level.

In addition to leveraging the promise of Phase III of the Defense Advanced Research Projects Agency (DARPA)-sponsored High Productivity Computing Systems (HPCS) program, the agency will establish a discussion and collaboration forum for scientists and engineers—including computational and computer scientists and engineers—and HPC system vendors, in order to ensure that HPC systems are optimally configured to support state-of-the-art scientific computing. On the one hand, these discussions will keep NSF and the academic community informed about new products, product roadmap and technology challenges at various vendor organizations. On the other, they will provide HPC system vendors with insights into the major concerns and needs of the academic science and engineering community. These activities will lead to better alignment between applications and hardware both by influencing algorithm design and by influencing system integration.

***2). Development and Maintenance of Supporting Software: New Design Tools, Performance Modeling Tools, Systems Software, and Fundamental Algorithms.***

Many of the HPC software and service building blocks in scientific computing are common to a number of science and engineering applications. A supporting software and service infrastructure will accelerate the development of the scientific application codes needed to solve challenging scientific problems, and will help insulate these codes from the evolution of future generations of HPC hardware.

Supporting software services include the provision of intelligent development and problem-solving environments and tools. These tools are designed to provide improvements in ease of use, reusability of modules, and portable performance. Tools and services that take advantage of commonly-supported software tools can deliver similar work environments across different HPC platforms, greatly reducing the time-to-solution of computationally-intensive research problems by permitting local development of research codes that can then be rapidly transferred to, or incorporate services provided by, larger production



*Massachusetts Institute of Technology researchers are developing computational tools to analyze the structure of any protein, such as the human ubiquitin hydrolase (shown), for knots.*

environments. These tools, and workflows built from collections of such tools, can also be packaged for more general use. Applications scientists and engineers will also benefit from the development of new tools and approaches to debugging, performance analysis, and performance optimization.

Specific applications depend on a broad class of numerical and non-numerical algorithms that are widely used by many applications, including linear algebra, fast spectral transforms, optimization algorithms, multi-grid methods, adaptive mesh refinement, symplectic integrators, and sorting and indexing routines. To date, improved or new algorithms have been important contributors to performance improvements in science and engineering applications, the development of multi-grid solvers for elliptic partial differential equations being a prime example. Innovations in algorithms will have a significant impact on the performance of applications software. The development of algorithms for different architectural environments is an essential component of the effort to develop portable, scalable, applications software. Other important software services include libraries for communications services, such as MPI and OpenMP.

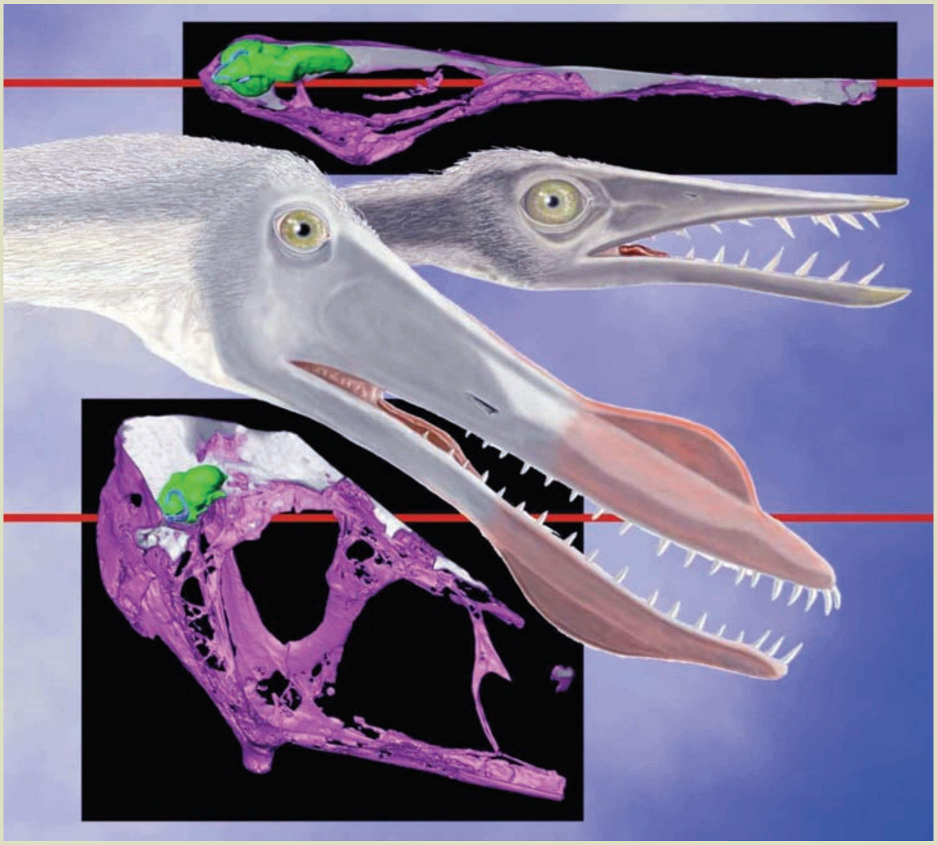
The development and deployment of operating systems and compilers that scale to hundreds of thousands of processors are also necessary. They must provide effective fault-tolerance and effectively insulate users from parallelization, as well as provide protection from latency management and thread management issues. To test new developments at large scales, operating systems and kernel researchers and developers must have access to the infrastructure necessary to test their developments at scale.

The software provider community will be a source for: applied research and development of supporting technologies; harvesting promising supporting software technologies from the research communities; performing scalability/reliability tests to explore software viability; developing, hardening and maintaining software where necessary; and facilitating the transition of commercially viable software into the private sector. It is anticipated that this community will also support general software engineering consulting services for science and engineering applications, and will provide software engineering consulting support to individual researchers and research and education teams as necessary.

The software provider community will be expected to promote software interoperability among the various components of the cyberinfrastructure software stack, such as those generated to provide modeling and simulation data, data analysis and visualization services, and networked resources and virtual organization capabilities. (See Chapters 3 and 4 in this document.) This will be accomplished through the creation and utilization of appropriate software test harnesses and will ensure that sufficient configuration controls are in place to support the range of HPC platforms used by the research and education community. The applications community will identify needed improvements in supporting software and will provide input and feedback on the quality of services provided.

NSF will seek guidance on the evolution of software support from representatives of academia, federal agencies and private sector organizations, including third party and system vendors. They will provide input on the strengths, weaknesses, opportunities and gaps in the software services currently available to the science and engineering research and education communities.

To minimize duplication of effort and optimize the value of HPC services provided to the science and engineering community, NSF's investments will be coordinated with those of other agencies. DOE currently invests in software infrastructure centers through the Scientific Discovery through Advanced Computing (SciDAC) program, while DARPA's investments in the HPCS program contribute significant systems software and hardware innovations. NSF will seek to leverage and add value to ongoing DOE and DARPA efforts in this area.



*Two skulls of separate species of pterosaurs were scanned at the High-Resolution X-ray Computed Tomography Facility at The University of Texas at Austin and the data was then fed to DigiMorph digital library to produce 2-D and 3-D structural visualizations.*

### *3). Development and Maintenance of Portable, Scalable Applications Software*

Today's microprocessor-based terascale computers place considerable demands on our ability to manage parallelism, and to deliver large fractions of peak performance. As the agency seeks to create a petascale computing environment, it will embrace the challenge of developing or converting key application codes to run effectively on new and evolving system architectures.

Over the FY 2006 through 2010 period, NSF will make significant new investments in the development, hardening, enhancement and maintenance of scalable applications software, including community models, to exploit the full potential of current terascale and future petascale systems architectures. The creation of well-engineered, easy-to-use software will reduce the complexity and time-to-solution of today's challenging scientific applications. NSF will promote the incorporation of sound software engineering approaches in existing widely-used research codes and in the development of new research codes. Multidisciplinary teams of researchers will work together to create, modify and optimize applications for current and future systems using performance modeling tools and simulators.

Since the nature and genesis of science and engineering codes varies across the research landscape, a successful programmatic effort in this area will weave together several strands. A new activity will be designed to take applications that have the potential to be widely used within a community or communities, to harden these applications based on modern software engineering practices, to develop versions for the range of architectures that scientists wish to use them on, to optimize them for modern HPC architectures, and to provide user support.