III. OPPORTUNITIES IN THE SCIENTIFIC DISCIPLINES

A. Condensed Matter Physics

Opportunities for Computational Physics

Much of the U.S. industrial economy is based on materials properties (metals, plastics, semiconductors, chemicals, etc.), so the potential economic impact of developing a capability to predict the properties of new materials is huge. Potential applications include the search for materials with special properties such as magnetism, superconductivity, hardness, etc. An understanding of microscopic systems is central to many areas of science and engineering such as chemistry, materials science, nanoscience, molecular biology, device physics, geoscience, astrophysics, and others of special relevance to NSF. The richness of condensed matter physics (CMP) arises from the diversity of materials and properties that are currently being investigated—more so than in many of the other physical sciences—but this results in a fragmentation of the computational techniques and codes. A large fraction of the HPCC resources are devoted to simulations of many-body quantum or classical systems because simulation is able to deal with the complexity of "real materials" and predict a variety of properties.

CMP simulation is divided between quantum-level and classical descriptions. In both arenas there are challenges and opportunities. In quantum-level calculation, which deals directly with electronic properties, a sea change has occurred in the last decade: the replacement of semi-empirical potentials with density functional methods. This sea change has occurred because of fundamental scientific discoveries (such as more accurate density functionals, both static and time-dependent), new algorithms (for example, *ab initio* methods like the Car-Parrinello approach), methods to treat large numbers of electrons efficiently, and, of course, the increase in available computational resources. Methods such as path integral Monte Carlo, which go beyond density functional theory to treat correlated quantum systems (i.e., where we need to treat electron-electron interactions accurately), allow us to adjust and calibrate the density functional methods. These methods are computationally more expensive, but indispensable in many cases. More accurate and efficient algorithms would have an enormous scientific and technological impact.

An example of the current capabilities of these methods is provided by recent achievements in understanding "hot dense hydrogen." The properties of hydrogen, the most abundant element, are crucial for understanding the formation of the Jovian planets, brown dwarfs, and other stars, and also for inertially confined fusion. However, the relevant part of the phase diagram is at energies and densities difficult to access experimentally (e.g., between temperatures of 1000 K and 20,000 K and at pressures from 50 to 1000 GPA). This is precisely when the theoretical description is most difficult: the molecular solid has melted, molecules are beginning to dissociate because of the pressure and temperature, and the result is a "soup" of many possible chemical species interacting quantum-mechanically. It is now possible to address this problem with first-principles simulations (starting with nothing more than the masses and charges of the electron and protons) with both the fundamental path integral techniques and the faster, but approximate, density functional methods. Recently it was shown that both of these methods give reliable predictions of what happens after a shock wave passes through hydrogen (the Hugoniot equation), reliable enough to assert that analysis of existing laser shock experiments was incorrect. In fact, new experimental results, published in November 2001, verified the simulation predictions. This is a milestone for computational condensed matter physics—when the simulations on a complex many-body system can be trusted as much as the experiments. The time is now ripe to compute thermodynamic properties of the common materials that stars and planets are made of: hydrogen, helium, oxygen, carbon, silicon, iron, etc. in regions of high temperature and density. However, this type of interdisciplinary work is not usually supported since it does not lie entirely in DMR or in PHYS. What is needed is a long-term "virtual computational facility" with funding for the scientists developing the methods and codes.

Turning now to the classical simulations, Moore's law implies that we can double length and time scale every six years. (Simulations are really in 4D space-time; hence doubling the length scale requires eight times the computational resources.) In the foreseeable future, we will not be able to simulate a macroscopic object particle by particle. To treat problems with both microscopic and macroscopic time or length scales, we need to develop multi-scale approaches, where the focus of the computation is on interesting regions of space-time, and the other regions are treated with a continuum description. It is difficult to develop such an approach and also to control errors, but it is the only known way of handling such problems. Aside from the mathematical difficulty, the code complexity is an order of magnitude more, since both particle level and continuum level needed to be stitched together.

Even at the macroscopic level, we cannot entirely work only with classical mechanics, since the quantum world can interact with the macroscopic world. Examples are systems undergoing a change of electronic state, such as during a chemical reaction or in the propagation of a crack tip through a solid. Currently quantum simulations are performed only for hundreds to thousands of electrons, not nearly large enough to describe chemical reactions in solution. For those systems, we need to be able to spawn a quantum calculation upon demand from within a classical simulation at a longer length scale. The key intellectual problem is to match the two descriptions seamlessly to avoid creating artificial interfaces.

Problems and Recommendations

Progress in computational physics has been constrained by available resources and research support. For example, the multi-scale approach is "cutting edge" but requires special support since it cuts across discipline boundaries. There is a need to nurture individuals who are developing methods, codes, and standards for the general research community and to change the culture to support university-based code development, for example, by supporting a larger team than is typical in university-based CMP research. Rather than supporting only highly innovative research, NSF needs to fund continuity and build on past successes. We need to be able to identify the best algorithms (through "bake-offs") and use these to make a sturdy, understandable code. The current system does not do this. Progress will also come from more collaboration between the physics, applied mathematics, and computer science communities. There are several examples of CMP problems that can benefit from expertise in such areas as improved mesh schemes for real-space Schrödinger/Poisson solvers, improved Monte Carlo sampling techniques, algorithms for high performance computers, construction of mathematical libraries, advanced software and languages. Although historically there has been such interaction from time to time, if a fruitful collaboration were to develop, it is hard to say where the long-term funding would come from and whether it would pass a “double jeopardy” review from both the physics and mathematics communities. Whether or not a computational physics program exists, progress will be made in computational physics, since there are a huge number of potential applications. However, without an appropriate funding program, progress will occur more slowly or elsewhere (in other countries, at national labs, etc.).

B. High Energy Physics

Quantum Chromodynamics

The Standard Model of High Energy Physics consists of theories of the strong, electromagnetic, and weak interactions. It has been enormously successful, having passed every experimental test. Nevertheless, high energy physicists believe that the Standard Model is the “low energy” limit of a more general theory which unifies all fundamental interactions, including gravity. Major experimental efforts are in progress in the United States and abroad to understand the physical phenomena predicted by the Standard Model, to make precision tests of it, and to look for new physics that goes beyond it. A knowledge of the predictions of quantum chromodynamics (QCD), the sector of the Standard Model describing the strong interactions, is crucial for these efforts. At present the only method of performing non-perturbative calculations of QCD from first principles and with controlled systematic errors is through large-scale numerical simulations within the framework of lattice gauge theory.

Lattice gauge theory calculations have demonstrated important qualitative features of QCD, such as quark confinement and chiral symmetry breaking. They have also yielded quantitative results of steadily increasing accuracy. Recent refinements of numerical algorithms coupled with major improvements in the capabilities of massively parallel computers have brought lattice QCD simulations to a new level. It is now possible to calculate a few crucial quantities to an accuracy comparable to their experimental determination. The strong coupling constant and the masses of the *c* and *b* quarks are notable examples. Furthermore, the experience gained to date allows lattice gauge theorists to predict the computing resources needed for accurate determinations of a broad range of fundamental quantities.

The study of the weak decays of strongly interacting particles (hadrons) is likely to be particularly fruitful for lattice QCD over the next several years. A significant part of the experimental programs at the major high energy physics laboratories is devoted to the study of such decays. However, some of the fundamental parameters of the Standard Model can only be extracted from these experiments with the aid of lattice calculations of the effects of strong interactions on processes induced by weak interactions. Example of such processes include the leptonic and semileptonic decays of *B* mesons, and CP-violating decays of *K* mesons. In most cases, the uncertainties in the relevant Standard Model parameters are now, or soon will be, dominated by those in the lattice calculations (Figure 1). Terascale computers will enable enormous improvements in these calculations, reducing their uncertainties to levels comparable to experimental ones, thereby improving our knowledge of some of the least well determined parameters of the Standard Model. By over-determining these parameters through the study of a variety of processes, we expect to be able to make precise tests of our current theories.

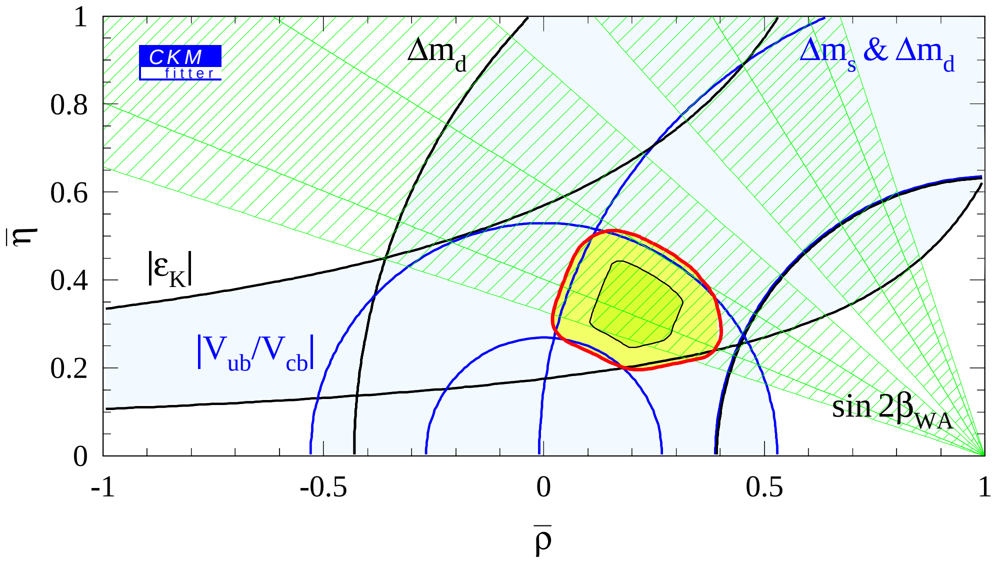


FIG. 1. The allowed region in the $\bar{\rho}-\bar{\eta}$ plane. Also shown are the individual constraints, and the world average sin 2β. (K. Anikeev et al., Fermilab-Pub-01/197)

One of the longstanding aims of lattice QCD calculations is to determine the masses, decay properties, and internal structure of strongly interacting particles. The masses of the lightest hadrons are very well known, so calculations of them serve as tests of lattice methods. However, a number of particles uniquely predicted by QCD, such as glueballs and particles with exotic quantum numbers, have not yet been observed. Accurate lattice calculations of their masses and decay properties would greatly aid experimental searches for them presently in progress. There is a wealth of experimental data that probes the internal structure of the nucleon, and further experiments are in progress. Lattice calculations will provide predictions or postdictions for much of this data. Work is already in progress to evaluate the electromagnetic form factors of the nucleon, as well as the moments of its quark density, spin, and momentum distributions.

At low temperatures and densities, quarks and gluons, the fundamental entities of QCD, are confined in elementary particles, such as protons and neutrons. At very high temperatures or densities, one expects a phase transition or crossover from this ordinary strongly interacting matter to a plasma of quarks and gluons. The quark-gluon plasma is believed to have been a dominant state of matter in the early development of the universe, and may exist today in the cores of neutron stars. Its observation and characterization is the primary physics goal of the Relativistic Heavy Ion Collider at Brookhaven National Laboratory. In order to make such an observation, it is important to determine the nature of the transition, the properties of the plasma, and its equation of state. Lattice QCD calculations are the only means of making a priori predictions about the quark-gluon plasma in the vicinity of the transition. They have already yielded a determination of the temperature at which the transition occurs and considerable information regarding the properties of the plasma. However, definitive predictions regarding the nature of the transition and the equation of state of the plasma require the terascale computations that will be made over the next several years.

The very large increases in computing resources available to academic scientists through the NSF's PACI Program and the DOE supercomputer centers will enable major progress in lattice QCD, as well as in many other areas of computational science. However, in order to fully capitalize on the investments being made in new hardware, it is necessary to support the development and maintenance of the software that will run on it, and the training of young scientists who will make use of it. An NSF program in computational physics that focuses on those aspects of research and human resource development that are not easily funded through current programs in the Physics Division would have a major impact.

An NSF program in computational physics could have a particularly important impact on the development and maintenance of software. The development of portable, efficient software for a variety of ever-changing supercomputer architectures is a challenging task, which requires command of physics, algorithms, hardware architecture, and programming techniques. Investments in this area can yield large scientific dividends. Clearly, a 10% improvement in the efficiency of the major codes run at the PACI centers would yield a scientific return equivalent to what would come from a 10% increase in the hardware budget, but at a much lower cost. Large, complex codes are increasingly being shared by entire scientific communities. Investments in such community codes would have particularly broad impact. It should be noted that the maintenance of complex codes used by broad communities is a time-consuming task that requires ongoing support. In lattice QCD studies, the overwhelming fraction of the computer resources are spent in generating large lattices (data sets), which can be used for a variety of physics applications. These lattices are beginning to be shared to maximize the science coming from them. Support for archives that would allow such data sets to be more easily and widely distributed would also have an important impact.

Lattice QCD projects are large undertakings. As is the case in many areas of computational science, work on them is increasingly being performed by large, geographically distributed groups. Although senior members of such groups can obtain funding through current NSF programs in their disciplines, support for multi-institutional group activities would be a natural role for a program in computational physics. Examples include support of extended summer meetings for collaborative research; support for postdoctoral research associates and graduate students working within the group under the direction of senior members at different institutions; and support for intermediate sized computers to be used for code development, data analysis, and simulations that do not require the powerful supercomputers located at the national centers but are too large for desktop machines.

Experimental High-Energy and Nuclear Physics

The role of computation in experimental high-energy and nuclear physics (HENP) has grown steadily over the last 30 years, and the rate of growth is increasing. An ever-expanding area of intellectually challenging computational science lies in the physics analysis of massive data sets produced by advanced detectors and accelerators.

To set the scale, in 1971 a typical large experiment involved 10 physicists and required as much as 100,000 lines of FORTRAN code. In 2001, a typical large experiment involves 500 physicists who create 7 million lines of C++, Perl, and Java code. The large experiments of 2011 will each involve 2000 physicists.

The foundation for many of the challenges in HENP computation is the non-deterministic nature of the quantum world. High-statistics data are needed to make quantitative measurements. High-statistics simulation is also vital, since physics itself draws an impenetrable veil hiding the detailed processes of individual interactions. Raising the collision-energy frontier has produced the majority of HENP's major discoveries. This inevitably demands increasingly large, complex, and costly detectors and accelerators. International collaboration has proved extremely effective in amassing the intellectual and financial resources needed for these major experimental programs.

Three challenges in the computational science of experimental HENP will be examined below:

1. The challenge of large-scale data management, driven by the need for precise measurements of probabilities in a quantum world.
2. The challenge of distributed data management and analysis, driven also by the imperative towards international collaboration, resulting in a "Grid" approach to data analysis that existed even before Grids became fashionable.
3. The challenge posed by the need for high-quality, long-lived scientific software, most notably software embodying all existing knowledge about particle interactions with matter.

Large-Scale Data Management

Way back in the mists of time (the 1960s and early 1970s), most groundbreaking experimental high-energy physics was done by measuring pictures. Photographs of particle interactions and decays in a bubble chamber were a wonderful way for a graduate student to study the frontier of physics, and were capable of pushing back the frontier, if high data rates and real-time rejection of 'boring' pictures were not essential.

Modern detectors can still produce pictures, but these are now a representation of information acquired by millions of high-speed sensitive devices. The pictures are used to understand the performance of the devices and of the feature-extraction or pattern-recognition software. They are not used to arrive at physics results, since human perusal of a billion pictures is impossible and, in any case, not to be trusted. Extracting deep understanding from a database containing billions of non-pictures has become a science in itself.

The Data Challenge

The challenge begins in the real-time systems that, taking an LHC[1] experiment as an example (Figure 2), must detect particle collisions happening at 40 MHz and, based on a limited readout of "trigger" signals, select no more than 0.25% of these to be fully digitized. The resultant digitized flow of over 2000 petabytes per year will be reduced to affordably manageable proportions by a series of increasingly compute-intensive filters. A few petabytes per year of fully digitized collisions will be written to persistent storage.

[1] LHC: Large Hadron Collider under construction at CERN, Geneva, with substantial U.S. participation in the accelerator construction and the experimental program. Operation is expected to start in 2007.

HENP Data Models and Access Patterns

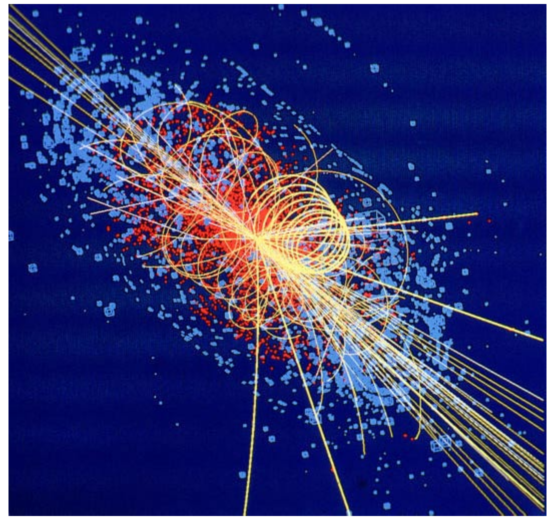


FIG. 2. Simulation of an LHC event with a Higgs boson.

Even bubble chamber pictures were rarely digitized and stored as images—the data volume would have been crippling. Already 35 years ago, the stored data consisted of concise descriptions of features such as tracks and vertices with structural or reference links describing their relationships. The data models describing collisions in today's detectors have hundreds of classes, with many relations between them. For much of the last 35 years, such models were mapped uneasily onto FORTRAN data types, and it was with some relief that HENP began to turn to object-oriented analysis, design, and languages in the early 1990s.

Understanding data-access patterns is vital when architecting data-analysis systems. The raw data, the first representations of collisions written to persistent storage, are rarely accessed. Extensive, computationally intensive reconstruction (that is, pattern recognition or feature extraction) must be performed before the "events" (the data arising from a collision) can be compared with a fundamental physics hypothesis. Being resource-intensive, reconstruction of the raw data is performed rarely and produces persistent reconstructed data that are then accessed much more frequently. Even when the physicists arrive at better detector calibrations and pattern-recognition algorithms, reconstruction is repeated at most once per year. Further tiers of more and more intensely accessed persistent data are created by processing the reconstructed data, creating more compact representations of high-level features. The overall picture is a data hierarchy ranging from the vast and immutable raw data, accessed infrequently, to frequently recalculated analysis data that hundreds of physicists query at the maximum rate technology can support. The corresponding access rates range from 0.01 to 1 gigabytes/s for raw data to 0.4 to 40 gigabytes/s for the most actively used analysis data; the lower numbers are typical of today, and the larger are expectations for the LHC. While these rates would not be particularly hard to sustain as aggregates of streaming access to disk-resident data sets, the reality is that the physicists' queries typically retrieve a sparely distributed collection of few-kilobyte data objects and often appear indistinguishable from totally random access to the hardware supporting the queries.

An Example of Data Management: The BaBar Experiment

A principal goal of the BaBar experiment is the study of CP violation, the small matter-antimatter asymmetry that gives rise to our existence and is far from understood. The PEP-II accelerator at SLAC started delivering collisions to the BaBar detector in 1999 and published the first measurement of CP violation in the neutral B–antiB meson system in the scientific and popular media in July 2001. The drive towards understanding CP violation requires many billions of B–antiB events. PEP-II accelerator physicists are racing Moore's Law by increasing the collision rate as fast as technology delivers more computing and data-handling capability.

Already in 1999, the BaBar data handling was straining at the limits of object-database technology. The computational demands of reconstruction required that hundreds of processors work in parallel and write data into the object database at an aggregate rate of tens of megabytes a second. Achieving this rate without relaxing the requirements that the database organization be optimized for physics analysis proved to be a major challenge. Intense work by a team of computer scientists and physicists succeeded in improving performance by a factor 20 before the end of the 1999, only to be met with a higher and higher data rate each year.

The current size of the BaBar database is about 450 terabytes, arguably the largest database in existence. The database system is still growing rapidly and now (late 2001) comprises close to 2000 database client machines, 100 database servers and tens of terabytes of disk cache integrated with petabyte-capable robotic tape storage. Tens of auxiliary servers—lock servers, journal servers, catalog servers, clustering hint servers, etc.—perform essential database functions and improve performance. The fragility caused by added complexity must be compensated by continued architectural and engineering improvements to increase reliability.

This exciting exploration of data-intensive science results in an exploration of uncharted territory in computer science. An indication of just how uncharted it is (and perhaps also that some estimates could have been better) is obtained by comparing what BaBar is doing now with the estimates in the 1995 technical design report. The plans for detector construction were proved correct to a high precision, producing never-before-constructed devices on time, on budget, and performing as planned. However, BaBar is now using three times the predicted number of database servers, complemented with 30 times as many client processors and 60 times as much disk cache as predicted. The principal cause of these embarrassing factors was a failure to anticipate all the benefits of the huge increases in hardware capability delivered by Moore's-Law-like evolution. Relatively cheap processing power and storage opened up new opportunities. The opportunities rapidly became necessities, since they were seen as cost-effective ways to maximize the return on the investments in PEP-II and the BaBar detector. The opportunities led into uncharted territory because Moore's-Law-like evolution does not produce smooth scaling of complex systems. In particular, while processing power and disk capacity continue to advance exponentially, the ability of hardware to support random-access I/O has advanced slowly and has been a major challenge for scientists in the BaBar database-development team.

Grids: Distributed Data Management and Analysis

High-energy physicists were quick to realize that, although they did not invent the term "Grid" in distributed computing, the Data Grid concept neatly encompasses everything that they have been striving to do for decades in support of distributed collaboration. The rapid widening of interest in Grids has presented a marvelous opportunity to ally computer scientists and physicists in revolutionizing the science of distributed data-intensive science. Typical of the HEP plans to which Grid clearly applies is the hierarchy of computer centers that is planned for LHC data analysis. An experiment has a single "Tier-0" center, in this case CERN, the site where the collisions occur, are filtered, are recorded on persistent storage, and undergo a first pass of reconstruction. The Tier-0 center will be linked by ~10 Gbps circuits to Tier-1 centers, typically national computer centers with state-of-the art technology for data-intensive computing. Tier-2 centers are run by university consortia or large universities and will be encouraged to experiment with innovative technologies in collaboration with computer scientists. Tier-3 centers are the compute servers and data-cache servers run by individual university groups.

The science of exploiting this hierarchy of centers has many challenges and tantalizing opportunities. The Particle Physics Data Grid is a DOE-funded collaboration of physicists and computer scientists aiming to advance both the science of Grids and its impact on physics by putting existing or near-term Grid technology into the unforgiving world of the mainstream data handling of current experiments like BaBar, and the demanding world of the preparations for LHC. The NSF-funded Grid Physics Network (GriPhyN) project is working at a frontier of computer science by committing to architect and implement "Virtual Data." As noted above, almost all HENP data analysis queries data that are the result of a previous computation. These data may be regarded as "virtual" in that revolutionary improvements in the effectiveness of the Data Grid for physics are likely if the Grid itself makes the decisions to instantiate, move, store, or replicate the data needed to respond to user queries.

Data Grids for HENP are not just a computer science challenge. Coordination of all the U.S., European, and worldwide Grid projects impacting HENP, together with coordination across HENP experiments involving a total of 10,000 physicists, is a requirement for success and rivals the computer science challenges in difficulty. A significant fraction of Grid funding is already being used for coordination activities, and this fraction must almost certainly increase.

NSF and DOE are putting major resources, approaching $10M per year, into PPDG, GriPhyN, and the U.S. component of the International Virtual Data Grid Laboratory (iVDGL) in the belief that the strong coupling to HENP will drive this area of science to make advances with impact well beyond HENP, perhaps extending to society as a whole.

HENP Software: Simulation

Precise simulation is required to pull physics out of the impenetrable quantum-mechanical clouds shrouding individual collisions, the severely obscuring effects of detector imperfections, and the stochastic processes through which particles generate signals in detectors. Creating software that embodies current understanding of all processes occurring when particles interact with matter, and is able to simulate interactions in detectors with millions of components, is a daunting undertaking. That is why otherwise staunchly independent collaborations have long cooperated in software development. Perhaps the leading current development effort is the GEANT4 Collaboration, in which an international team of over 140 physicists and computer scientists has been working for almost eight years on the development of an object oriented simulation toolkit (Figure 3). Almost from the beginning, the European Space Agency has joined high-energy and nuclear physics in developing GEANT4. In the last two years, there has been a rapid growth in collaboration with the medical community. As with Data Grids, coordinating this development and science is becoming daunting, but the potential benefits to science and medicine make the retention of limited focus inexcusable.

Summary of HENP Data Analysis Opportunities

Advancing the science of computation is integral to the mission of high-energy and nuclear physics. Opportunities exist to go beyond the bounds of HENP alone by funding research requiring a collaborative effort spanning HENP, computer science, and other data-intensive sciences. In the field of Grids, such funding is already in place, and the benefits are beginning to appear.

Large-scale scientific data management will increasingly underpin much of tomorrow's leading science. HENP is likely to strain at the limits of the science and technology of data management for at least a decade. There are exciting opportunities to fund collaboration between HENP, computer science, and the other increasingly data-intensive disciplines and achieve revolutionary advances for physics and other sciences.

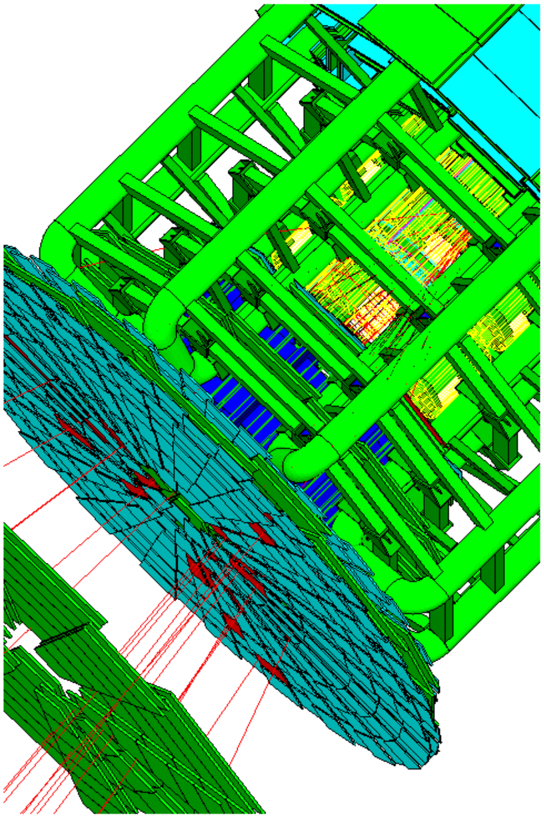


FIG. 3. GEANT4 display of simulated particle tracks in part of the ATLAS detector at the LHC.

Finally, the availability of well-architected software embodying current scientific knowledge is a key to pushing back frontiers and widening the applicability of existing discoveries. Funding should encourage the creation of high-quality software, architected and implemented for a wide range of scientific applications.

Accelerator Science

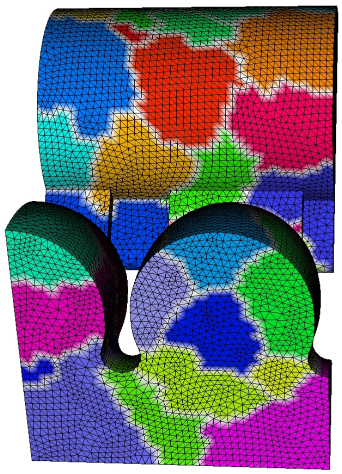
Particle accelerators are among the most important and most complex scientific instruments in the world. The nation's accelerators—including its high-energy/nuclear facilities, synchrotron light sources, and spallation neutron sources—are critical to research in fields such as high energy physics, nuclear physics, materials science, chemistry, and the biosciences. The scientific discoveries and technological advances made possible by accelerators impact both the basic and applied sciences. Accelerators have also been proposed, or are already playing a role, in addressing national needs related to the environment, energy, and national security. Examples include the accelerator transmutation of waste, accelerator-driven fission and fusion energy production, accelerator production of tritium, and proton radiography for stockpile stewardship. Beyond these large-scale applications, particle accelerators and the technology associated with them have many uses that are highly beneficial to society. Examples include irradiation and sterilization of biological hazards, medical isotope production, particle beams for medical irradiation therapy, superconducting magnets for medical magnetic resonance imaging, scintillator technology for medical diagnostics, ion implantation, and beam lithography. All told, particle accelerators have had, and will continue to have, a profound impact on U.S. leadership in science and technology, and on improving the quality of people's lives.

The NSF now operates or contributes to several small, medium, and large-scale accelerator facilities such as the Cornell Electron Synchrotron Ring (CESR) and the Cornell High-Energy Synchrotron Source (CHESS), the National Superconducting Cyclotron Laboratory (NSCL) at Michigan State University, and the Indiana University Cyclotron Facility (IUCF). The successful development of accelerator facilities involves investments in the three principal elements of scientific research: theory, experiment, and computation. The availability of high performance, large memory parallel supercomputers has made large-scale computing an indispensable tool for designing next-generation accelerators and for performing research on new accelerator technologies. Large-scale simulation enables numerical experiments on systems for which physical experimentation would be prohibitively expensive or technologically unfeasible. In situations that involve beams in extreme environments (like the ultra-high field environment of laser/plasma accelerators) or that push the boundaries of existing technologies (like the ultra-high-brightness beams of proposed fourth-generation light sources), computation provides a window to explore and ultimately gain insight into the fundamental behavior of beams.

The development of next-generation accelerators and new accelerator technologies will require advances in computational accelerator science. Computational accelerator science includes the development of grid generation tools, mathematical algorithms, computational methods, and visualization tools, all of which must be implemented on and optimized for parallel computing environments. Applications of computational accelerator science fall mainly into three areas: electromagnetic modeling of geometrically complex 3D accelerator structures and components, simulation of beam dynamics in accelerators, and simulation of "advanced concepts," often involving a combination of particle beams, lasers, and plasmas. The following describes challenges and opportunities in each of these areas.

Electromagnetic Modeling

Accelerator physicists and engineers are faced with increasingly stringent requirements on electromagnetic components as machines continually strive towards higher energy and current, and greater efficiency. In one next-generation linear collider scheme, for example, the frequency of the accelerating field must be accurate to within 1 part in 10,000, which is comparable to fabrication tolerance. This requirement is to be met in a complex cavity geometry that optimizes the field gradient while suppressing wakefield effects. The computational design of such a structure involves a huge number of degrees of freedom, and can only be performed on very large memory, high performance parallel supercomputers. Figure 4a shows a one million degree-of-freedom geometric model of an optimized accelerating cavity design based on an unstructured grid and partitioned for load balancing. The use of advanced computing in designing electromagnetic components enables simulation to become a cheaper and faster alternative to the expensive, time-consuming process of repeated fabrication and testing. Terascale computing provides the opportunity to address even more challenging design issues that arise not only in other future accelerators, such as fourth generation light sources, but also in potential upgrades to existing facilities such as cyclotrons and storage rings. For example, more complex rf structures are being proposed for cyclotron upgrades as well as cyclotrons for medical applications. Figure 4b shows the current density obtained by parallel computations in a preliminary design for a cyclotron project at the Paul Scherrer Institute. The demand for computing resources for this and similar projects is expected to increase dramatically as higher levels of complexity are included for both electrical and mechanical purposes. Without new simulation tools using supercomputers, such computer-aided design endeavors would not be considered possible.



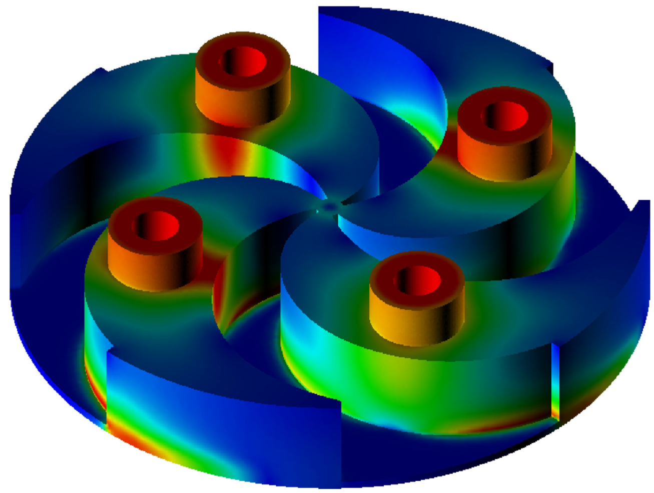


FIG. 4a (left). Geometric model of an optimized linear collider accelerating cavity design based on an unstructured grid and partitioned for load balancing.
FIG. 4b (right). Results of a parallel simulation showing the current density, indicated by color, in a preliminary design for a PSI cyclotron project.

Beam Dynamics Modeling

Large-scale computing provides a powerful tool to explore the complex behavior of charged particle beams in accelerators. Critical issues that can be studied using advanced computing include (1) beam stability and phenomena affecting beam quality in storage rings, (2) physics of intense charged particle beams including halo formation, and (3) techniques for manipulating beams in phase space, including emittance control. All of these are important issues for NSF facilities. For example, the beam-beam effect is one of the key factors limiting the luminosity of electron storage rings like CESR; space-charge effects must be well understood in order to increase the beam intensity in cyclotrons like those at NSCL and IUCF; and cooling methods such as electron cooling are already in use at facilities such as IUCF and may play an important role in other accelerators such as the Relativistic Heavy Ion Collider and the Tevatron.

For all three of these beam dynamics areas, modeling on parallel supercomputers is essential to gain insight and understanding of these systems, and for design decisions aimed at evaluating and reducing risk, reducing cost, and optimizing accelerator performance in future upgrades. The availability of terascale computers has opened the door to performing accelerator simulations that were, as little as 10 years ago, thought to be well beyond the realm of possibility. For example, the Fokker-Planck equation provides a model of the multiple small angle scattering associated with electron cooling. But as recently as 1997, self-consistent 3D computer simulation based on this model was said to be “completely impractical in terms of number of particles, computation time, and statistical fluctuations.” Now, however, thanks to the availability of terascale resources and the development of new algorithms targeted to terascale platforms, such calculations are a reality.

Simulation of Laser- and Plasma-Based Accelerators

Conventional accelerators operate with gradients on the order of 1 to 20 million volts per meter (MeV/m). Efforts are under way to push the limits of conventional rf technology to close to 100 MeV/m, but even that is extremely challenging because, at such high gradients, conventional structures are prone to rf breakdown. Given the fact that high energy accelerators cannot grow in size indefinitely, it will be necessary to develop new, high-gradient accelerator technologies in order to continue to advance the energy frontier.

One possible approach, which is being pursued by several groups funded by the NSF, is to use the extremely high fields that can be generated in lasers and plasmas as a means to reach very high gradients. Such laser- and plasma-based concepts have already achieved ultra-high gradients in laboratory experiments—up to several hundred GeV/m—but these gradients have been sustained only over very short distances (of order millimeters). The challenge is to control and stage high-gradient sections so that one can produce high-quality, high-energy beams in a less costly, more compact configuration that would be impossible using conventional technology. Not only would such gradients make it possible to reach ultra-high energies; such a drastic increase in gradients would also allow a reduction in accelerator size so that, though operating at low power, tabletop accelerators might someday achieve beam energies approaching those now found only at a few national facilities.

The ability to place such compact accelerators in university departments, government research organizations, high-technology businesses, and hospitals, would have staggering consequences for science, industry, and medicine. The development of these systems into useable accelerators is a very high-risk, high-return undertaking. Thanks to the confluence of three things—successful small-scale experiments, the availability of terascale computing resources, and the availability of parallel 3D codes for modeling laser/plasma accelerators—it is now possible for full-scale simulations to play a pivotal role in guiding experiments (Figure 5) and in making laser- and beam-plasma accelerators a reality. In addition, the fundamental physics inherent in ultra-intense laser- and beam-plasma interactions is rich in nonlinear, ultra-fast, and relativistic physics. The insight gained from large-scale particle-in-cell codes is essential for unraveling this new physics.

Summary of Opportunities in Accelerator Science

Given the great value of particle accelerators to the nation, it is imperative that the most advanced high performance computing tools and resources be brought to bear on the challenging and important problems facing the field. Continuing the exciting progress of accelerator science and technology into the twenty-first century will require a coherent program of research and development involving accelerator physics, mathematics, computer science, and engineering. A new NSF program in computational science that includes computational accelerator science, performed in concert with theoretical and experimental programs, will lead to new discoveries in accelerator science and beam-based science. Beyond the intrinsic value of basic research in these fields, these discoveries will open the door to the development of innovative accelerator concepts that will provide the nation's researchers with the twenty-first century tools—twenty-first century accelerators—needed for continued progress and leadership in the sciences and engineering.

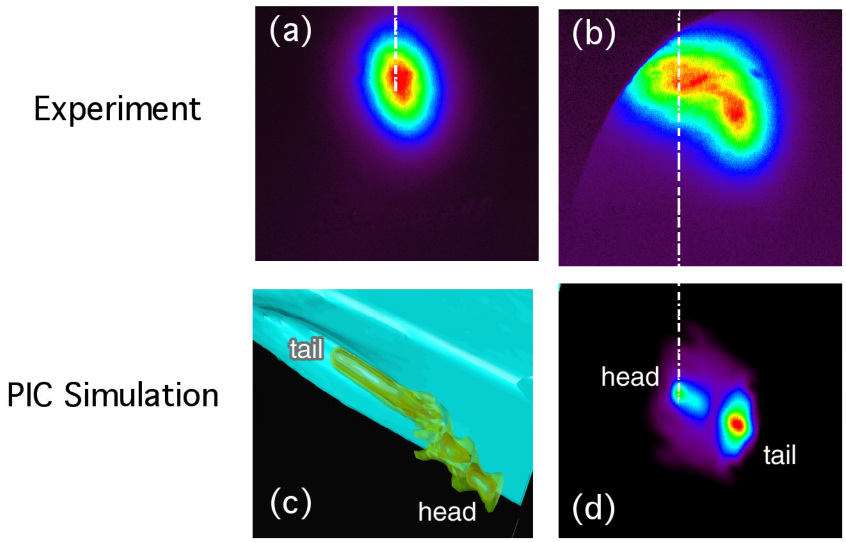


FIG. 5. Experimental and computational results of an electron beam refracted by a laser-ionized plasma. (a) Head-on view of the electron beam with the laser off; (b) head-on view with the laser on; (c) visualization of simulation results showing a perspective view of the refracted beam with the laser on; (d) head-on view of the simulation results with the laser on—the amount of deflection of the tail is in good agreement with measurement. From P. Muggli et al., "Boundary effects: Refraction of a particle beam," *Nature* **411**, 43 (2001).

C. Astrophysics and Relativity

While observation has always been fundamental to astronomy, controlled experiments in astronomy are extremely rare. Therefore, in astronomy, computer simulations have taken over the traditional scientific role of controlled experiments by making it possible to test scenarios, so long as the underlying physical laws are known. Observations still provide a check, but they show the results of processes that we cannot control in a laboratory. Furthermore, the evolutionary time scales for most astronomical systems are so long that we see these systems as if frozen in time. Constructing evolutionary models purely from observation is therefore difficult. By observing many different systems of the same type, such as stars or galaxies, we can see many different stages of development and attempt to put them into a logical order, but we cannot watch a single system evolve. To provide the evolutionary model that ties the different observed stages together using known physical laws and properties of matter, a computer simulation is usually required. For example, we could conjecture that the collision of two spiral galaxies might be so disruptive that it would result in the formation of an elliptical galaxy. Observations of many different sets of colliding galaxies might make this argument quite convincing, but only a computer simulation could show the entire time progression for the same two colliding galaxies and might thus establish conclusively that Newton's law of gravitation, applied to this complex set of circumstances, implies this result.

Perhaps the greatest success of early computational astrophysics has been the theory of stellar evolution. This theory, for example, explains how a red giant star can evolve from a solar-type precursor as the result of exhausting hydrogen fuel in its core. Stellar evolution theory gives us an excellent example of why astrophysicists have been forced to rely on computer simulation. Although one can perform laboratory experiments to determine the properties of the gaseous constituents in a star like the sun, one cannot build an experimental star in the laboratory and watch it evolve. To perform that experiment, one must resort to computer simulation. Although one can make some simple arguments and estimates without using a computer, the physics involved in stellar evolution theory is complex and nonlinear. Therefore one does not get very far in developing the theory without a computer. The gas dynamical equations that must be solved are nonlinear, particularly during stages of relatively rapid evolution such as the birth or death of the star, and they are coupled to complex systems of nuclear reactions as well as to complex small-scale phenomena such as turbulent convection. Near the surface of the star, radiative transfer needs to be treated in order to compute the appearance of the star for comparison with observations. Magnetic fields undoubtedly play an essential role in the process of star formation, so that they must be included as well. If the star ultimately explodes as a supernova, exotic nuclear reactions and the transport of neutrinos need to be incorporated in the simulation. Computer simulations enable us to put all these very complex physical ingredients together and, with sufficient computer power and grid resolution, to compute the observable behavior that they imply under a range of circumstances. This ability to carry out numerical experiments over simulated times ranging from millions to billions of years makes the computer indispensable to the astrophysicist.

Computational astrophysics has benefited enormously from the NSF supercomputer center program. The new Terascale systems—the one recently put in place at Pittsburgh and the still more ambitious TeraGrid—offer the possibility of particularly exciting and rapid advances. These new computing systems represent so great a leap in computational power that they make possible simulations that not only refine previous studies but also add new dimensions. The new computing power can be used to literally add a spatial dimension, taking, for example, a 2D simulation of a supernova explosion into 3D. The new systems can be used to add treatments of new and important phenomena to a simulation; for example, magnetic fields could be added to global simulations of solar convection to address the operation of the dynamo that drives the sunspot cycle. For some problems, such as the development of large-scale structure in the expanding universe, simply getting more of the system under study into the computational problem domain by dramatically increasing the size of the computational grid should have a significant impact on scientific discovery. Alternatively, one might choose to simulate the same-sized system, using the new computational power to treat structures on a much wider range of length and time scales. Here the cosmological problem is an excellent example, since it contains within it scales of interest ranging from a single star to a large cluster of galaxies.

To take advantage of the new opportunities for scientific discovery that are created by the NSF investments in terascale computing hardware, astrophysicists will need to adapt and enhance their simulation codes considerably. The need to handle multiscale problems is a recurrent theme among the various disciplines represented in this report, and astrophysics is no exception. However, for astrophysics, handling multiple physical processes within a single simulation code is an equally important theme. The primary components of astrophysical computation are: (1) unsteady, compressible gas dynamics, with or without magnetic fields; (2) dynamics of gravitation for continuum systems or for systems of point particles, and possibly under relativistic conditions; (3) radiative transfer, either for continuum or line radiation; and (4) nuclear reactions. The present state of the art in computational astrophysics can involve computations that, while extremely difficult, leave out one or more of the above components in order to make a particular problem tractable. As computing machinery becomes steadily more powerful, astrophysicists are able to combine more of these computational components into single simulation codes, so that they can perform increasingly realistic and complex computer experiments.

While incorporating these additional components into the simulation codes, astrophysicists must still attempt to treat increasing ranges of length and time scales. At the same time, these new, more powerful codes must achieve scalable performance on increasingly large and complex networks of machines. One can therefore easily appreciate that the process of code development and maintenance can become so complex and time consuming that it begins to replace access to computer power as the main factor limiting scientific progress in computational astrophysics. Here is a prime opportunity for timely NSF investment. An NSF computational physics program can invest not only in the research teams that will run these new simulation codes on the new terascale equipment, but also in the teams that will write and maintain these codes. Such investments will surely bear fruit in scientific discovery. At the workshop in September, a number of avenues for such discovery were discussed, and these are reviewed below.

Stellar Astrophysics

Until recently, the study of stellar evolution has been dominated by 1D models. Over the last several years, the rapidly increasing power of the machines at the NSF supercomputer centers has enabled astrophysicists to begin taking into account the true 3D nature of stars. The inadequacy of 1D models is most evident at the earliest stage of stellar evolution, the process of star formation. The formation of protoplanetary disks around protostars, the development of strong jet outflows, and, of course, the accumulation of planets make it clear that star formation is a 3D process. Proper simulations of this process must, at a minimum, include compressible gas dynamics, gravity, magnetic fields, and radiative transfer. Treating all these phenomena accurately in a single code in 3D is daunting enough. However, the vast difference in scale between the radius of a protostar and the radius of its protoplanetary disk presents a still greater challenge. Not only will a successful simulation require a strongly adaptive spatial grid, but it will also have to deal with the huge difference in dynamical time scales for the central object and its surrounding disk. The new terascale computing systems offer the hope that 3D simulations of star formation that overcome all these challenges might emerge in the next few years. However, the great difficulty of the problem suggests that only sustained team efforts will be successful here. The question of how stars form, and particularly stars with planets, is of great interest because of its obvious relevance to the likelihood of other intelligent life existing in the neighborhood of our galaxy. Efforts to understand star formation in greater detail are particularly timely now, because modern telescopes are beginning to detect evidence of planets orbiting nearby stars.

The 3D nature of stars also plays an essential role near the end of stellar evolution, in a supernova explosion. The explosion of SN1987A in the Large Magellanic Cloud gave us the extremely unusual opportunity to observe the death of a star in great detail. The early emergence of heavy elements in the remnant made it clear that 1D models of this phenomenon cannot apply.

Arnett and his collaborators have since applied computer simulations in 2D that demonstrate the potentially decisive role of Rayleigh-Taylor instabilities in getting materials formed in the exploding star's deep interior far out into the ejecta and ultimately mixed into the interstellar medium, where they can become incorporated in new stars and planets. Recent laser experiments at the Lawrence Livermore National Laboratory provide some more direct means to test the computational methods used in this problem, but it is clear that the mixing process set up by the Rayleigh-Taylor instability can properly be described only in 3D.

Another fluid-dynamical instability that plays an essential role in stellar evolution is convection. When the conduction of heat outward from the interior by light is too inefficient, the heat is instead transported by material motion in convection. The extent to which convective eddies overshoot the confines of the region of radii in which the gas is formally convectively unstable can have important consequences for the mixing of the various chemical constituents in the star. Also, near the surface of a star like the sun, where the heat transported by convection from the interior escapes as light, the structure of this visible region is profoundly affected by convection. Our proximity to the sun allows us to observe these surface phenomena in great detail, providing an enormous data base for testing computational models. The computational models are our principal tool for explaining these surface phenomena, such as the 22-year sunspot cycle, since it is clear that driving mechanisms, such as a solar dynamo, are buried from our view beneath the surface layers. Recently helioseismology has given us valuable information on the subsurface structure, which is helping to constrain the models.

Both the NSF and NASA designated understanding turbulent convection in stars as a computational Grand Challenge during the last decade. The state of this art was discussed by Juri Toomre, a leader of both the NSF and NASA Grand Challenge teams, at the September workshop. Present computing equipment and present hydrodynamics and magnetohydro-dynamics (MHD) methods have so far been able to bring us global simulations of convection in spherical shells at resolutions where the turbulence of the gas is well developed (Figure 6). These calculations are beginning to illuminate the process in which convection redistributes the angular momentum in the sun's convection zone, producing the differential rotation that we see. This differential rotation, with the equatorial regions rotating faster than the polar regions, must drive the solar cycle of magnetic activity. To illuminate that process, further and far more difficult calculations will need to be performed, but this goal now seems within reach in the next few years.

Convection in the sun is restricted to about the outer third of the star in radius. In giant stars, particularly the highly luminous ones on the so-called asymptotic giant branch, convection reaches from the surface all the way down to the very small and very hot central core. Simplified models of these stars have allowed recent computations to simulate the entire stellar envelope, revealing a global convection circulation coupled to pulsation of the envelope (Figure 6). To address mass loss from the star and the ejection of the envelope material to form a planetary nebula will require not only more realistic stellar models in such calculations but also the addition of detailed interactions of the escaping light with the gas near and above the visible stellar surface. Such calculations will require more complex simulation codes and more computer power, but these goals are attainable within the next few years.

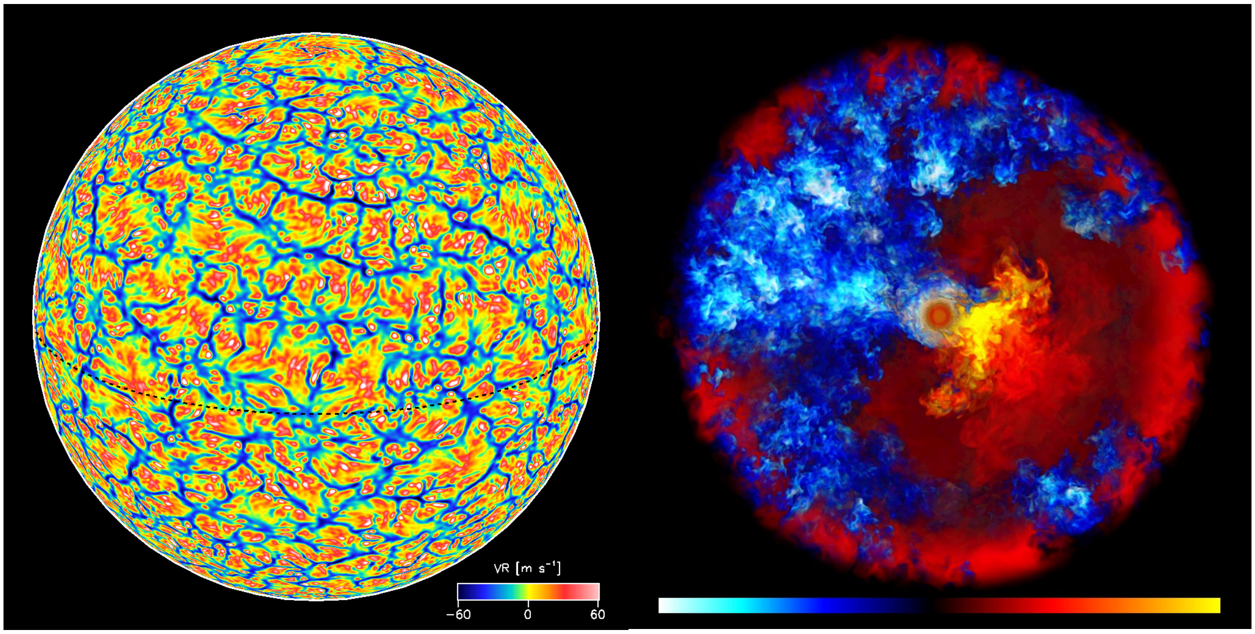


FIG. 6. On the left is shown the radial component of the velocity at the upper surface of a thin, convectively unstable gaseous layer representing the upper convection zone of a star like the sun. This simulation by Brun and Toomre was performed using an anelastic spherical harmonic code running on the NSF supercomputers. Only a sector of longitudes was simulated here, and periodic boundary conditions were applied. The cellular structures visible at the surface are to be compared to the solar supergranulation. At the right, temperature deviations from the average at each radius are shown in a simplified model of the convective envelope of a 3 solar mass star in its asymptotic giant branch (AGB) phase. This simulation, by Jacobs, Porter, and Woodward using the NCSA computers, includes the entire stellar envelope with a grid of a billion cells. The star has been sliced in half along a plane of approximate symmetry to reveal a global dipole circulation that carries relatively cool gas (blue and aqua) downward, so that it flows around the hot central region of the star, is heated, and rises as relatively warm, buoyant gas (red and yellow) on the other side of the star. From inspecting the temperature or velocity distribution near the stellar surface, this highly organized interior flow is barely evident.

In recent years researchers have begun to investigate a variety of transient events in the life of a star using multidimensional computer simulations. Many of the most impressive simulations have been carried out by the University of Chicago's Center for Astrophysical Thermonuclear Flashes. This group has used some of the world's largest supercomputers and has aggressively employed adaptive mesh refinement techniques to handle the large ranges of length and time scales in problems such as the propagation of helium detonations on the surfaces of neutron stars. The bulk of these calculations published to date have been carried out in two dimensions. As the power of supercomputers continues to increase, three-dimensional simulations of this type should become feasible.

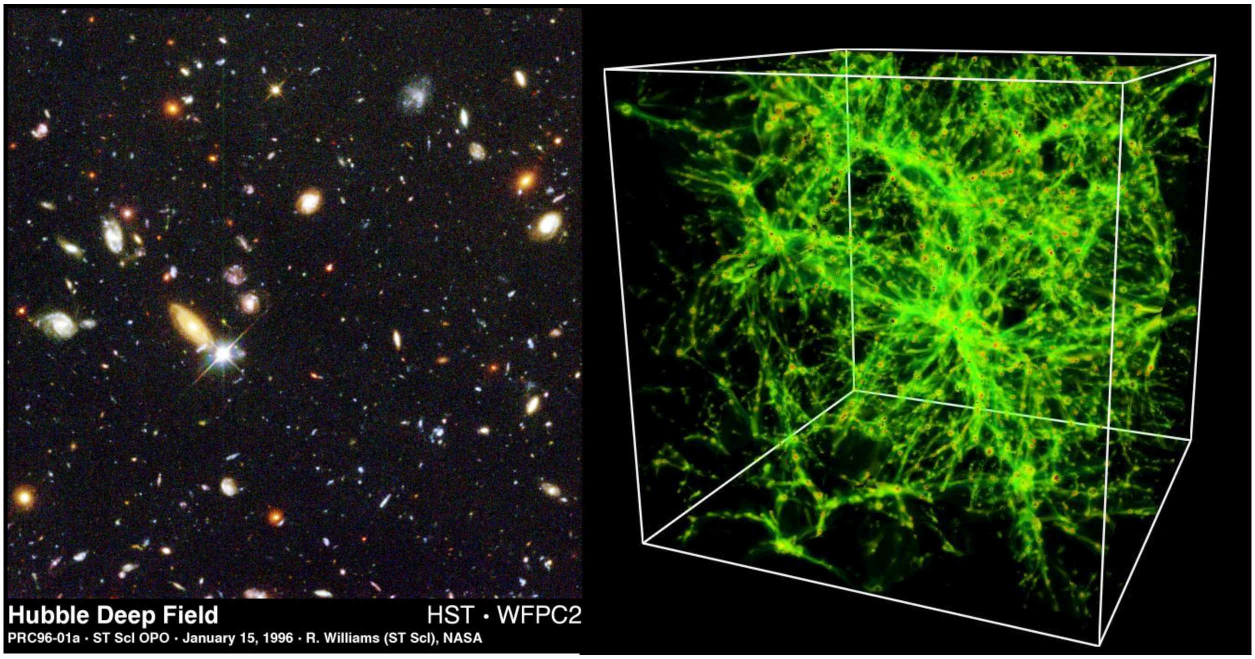


FIG. 7. At the left, the Hubble Space Telescope image contains extremely distant galaxies that are only one or two billion years old (although the larger objects are nearer and older). At the right, a simulation by Cen and Ostriker on a 512^3 grid of the development of large-scale structure through the action of gravity is one of many computer experiments in recent years that have allowed researchers to find ranges in the parameters of cosmological models that result in structures in agreement with the exciting new observations of the early universe. In the simulation, green shows overdensities of a factor of about 10, while the red regions are overdense by a factor of about ten thousand. Thus the red spots in the image at the right show the locations where galaxies are forming.

Cosmology

Over the last decade, a wealth of new observations have made cosmology one of the most active fields of modern astrophysics. NASA's Cosmic Background Explorer (COBE) satellite and the Hubble Space Telescope gave us images of structure in the very early universe and at times when galaxies were very young (Figure 7). Also, our new ability to detect type-1 supernovae in extremely distant galaxies has allowed us to measure distances in a way that can be compared with some confidence to redshifts, so that we have been able to estimate the rate at which the expansion of the universe appears to be changing. These new observations, and the expectation of far more information from the next generation of observing equipment, have generated tremendous excitement. They make possible for the first time meaningful comparisons between observations and the results of numerical simulations of the development of structure driven by gravitational forces in the very early universe. The observations place sufficient constraints upon the models for the expanding universe that one is justified in concentrating on a relatively small range of parameters and studying the action of gravitation in detail. The results of such studies can then meaningfully be compared to observations not only of the ultimate results, the structures that we now see at our own late era, but also to observations corresponding to the era of galaxy formation.

The development under the action of gravity of large-scale structure in the early universe is a highly nonlinear and complex process. It is profoundly affected by strong shock waves that decelerate and heat gas falling into growing density condensations as well as by the radiation released from early generations of stars, which ionize and heat the surrounding gas (Figure 8). The interplay between the formation of collapsed objects like stars and galaxies, their back reaction upon the remaining gas through both gravitation and radiation, and the distinctly different dynamics of gravitating point particles and gas make these simulations extremely difficult.

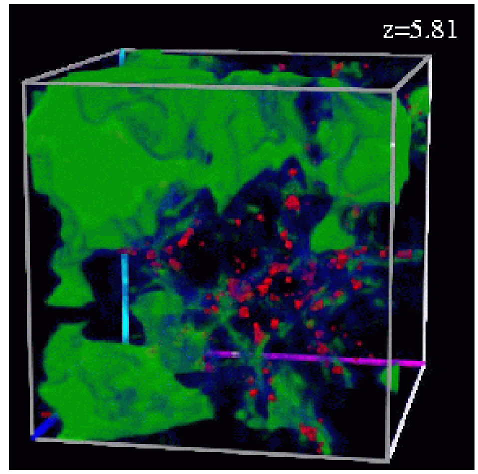


FIG. 8. A simulation of the reionization of the intergalactic medium, created by Razoumov et al. using the NCSA supercomputers. The green regions indicate the neutral gas that is being ionized by the radiation from newly formed galaxies.

Adaptive grid techniques are used to handle the tremendous difference in scale between small galaxies and large galaxy clusters. However, if such a calculation is to follow the formation of large-scale structures up to the present era, it soon becomes impractical to attempt to follow the detailed behavior of large star-forming clouds or of small galaxies on their own natural time scales. Such small scale behavior must either be simulated by recourse to a simplified model or, inaccurately, by refusing to adapt the grid to resolve it. Building careful simulations of structure development in the early universe is thus a large enterprise, which must ultimately draw upon many related studies of the detailed processes involved. From these studies of small-scale phenomena, one may design and test models that are simple enough to permit the simulations of the global dynamics to proceed. This undertaking will require a great deal of work and a great many simulations investigating in detail such processes as galaxy collisions and feedback on the developing dynamics from radiation emitted by early generations of stars.

There have been a number of very encouraging successes in this area in recent years. It has been known for many years that we cannot account for all of the gravitating mass in the universe by adding up all the luminous matter that we see, even if we look in the electromagnetic spectrum well beyond the small range of visible wavelengths. Simulations such as the one in Figure 7, based on the “cold dark matter” model, in which the missing mass is represented by unseen particles with a small initial velocity dispersion, yield distributions of visible matter in the early universe that correspond well with the observations of absorption lines in the spectra of very distant quasars (the “Lyman alpha forest”). These same quasar spectra now seem to indicate that at very early times, around a redshift of 6, the intergalactic medium was largely neutral, and it became reionized at later epochs, presumably as a result of the intense ionizing radiation from newly formed stars and galaxies. This reionization of the intergalactic medium is a natural result of the models, and simulations of this effect are beginning to appear (Figure 8). These calculations are extremely difficult and time consuming because of the added dimensionality that radiative transfer adds to the problem. Once again we can expect significant improvements to result form the much more powerful computing hardware that is being installed to build the NSF's TeraGrid.

Relativity

It is the relentless pull of gravity that drives the generation of large-scale structure in the universe. The formation of stars and galaxies increases the matter density above the average value by many orders of magnitude. As stars evolve, burn their nuclear fuel, and eventually collapse, even greater concentrations of matter can be formed which we call black holes. These objects, with gravitational fields so powerful that even light cannot escape, are perhaps the most exotic structures in the universe. The general belief that they exist in nature comes from theoretical arguments and from observations of their indirect effects. The general theory of relativity that predicts their existence also predicts that they can be the sources of strong oscillations in the gravitational field, observable on earth as gravitational waves. To observe these waves, extremely sensitive detectors are being built, including NSF's LIGO (Laser Interferometer Gravitational-Wave Observatory, Figure 9).

FIG. 9. The Hanford Observatory, one of LIGO's two laser interferometers for detecting gravitational waves.

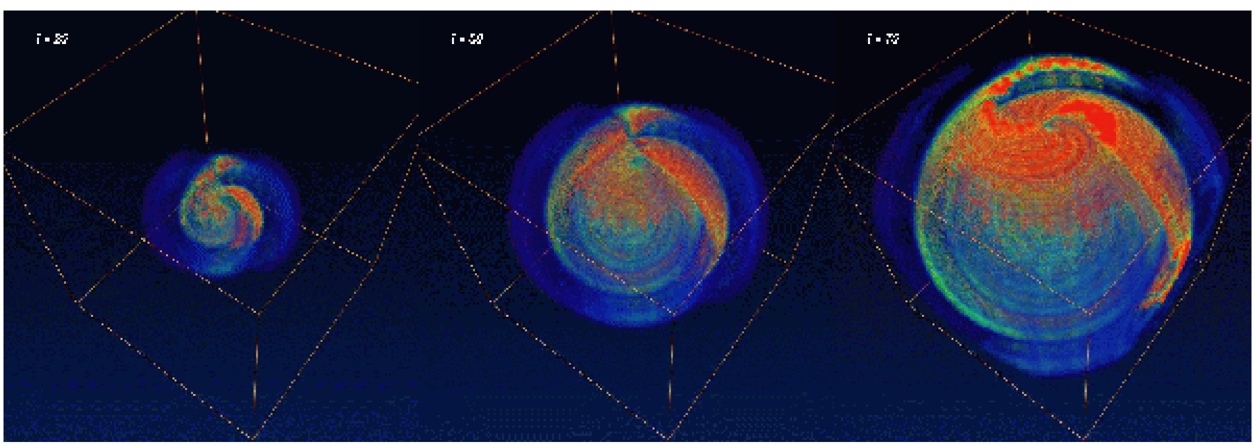


FIG. 10. Three successive snapshots of the developing gravitational wave from the coalescence of two black holes in a binary system. Computation by Baker et al. of the Albert Einstein Institute's Lazarus project.

LIGO is being constructed to observe gravitational waves generated in the supernova collapse of stellar cores to form either neutron stars or black holes, and in the collisions and coalescences of neutron stars or black holes. To help pick these signals out of the noise, LIGO will use theoretically generated waveforms as a guide. One such waveform (Figure 10) was recently computed by Baker et al. of the Albert Einstein Institute's Lazarus project using the Cactus code to simulate the black hole merger and using a perturbation technique to compute the gravitational radiation from the resulting distorted black hole. Such calculations push the limits of present-day computer power. As new detectors like LIGO begin producing data, these simulated waveforms will be our only means of identifying the sources of gravitational waves, such as black holes, and their distance from us.

D. Computational Plasma Physics

Plasmas comprise over 99% of the visible universe and are rich in complex, collective phenomena. A major component of research in this area is the quest for harnessing fusion energy, the power source of the sun and other stars, which occurs when forms of the lightest atom, hydrogen, combine to make helium in a very hot (100 million degrees centigrade) ionized gas or *plasma.* The development of a secure and reliable energy system that is environmentally and economically sustainable is a truly formidable scientific and technological challenge facing the world in the twenty-first century. This demands basic scientific understanding that can enable the innovations to make fusion energy practical.

Future research will require the accelerated development of computational tools and techniques that vitally aid the acquisition of the scientific understanding needed to develop predictive models which can prove superior to empirical scaling. This is made possible by the impressive advances in high performance computing which will allow simulations of increasingly complex phenomena with greater fidelity. Viewed in this way, advanced scientific computing, in tandem with theory and experiment, is a powerful new tool for discovery.

Recent progress and future directions for advanced simulations in magnetically confined plasmas are highlighted here to demonstrate that plasma science is both effectively utilizing and contributing to the exciting progress in information technology and scientific computing. This is clearly a time of excellent opportunity for investments in computational plasma science to drive research with the greatest promise for accelerating scientific understanding, innovation, and discovery.

Nature of Challenges and Recent Advances

Although the fundamental laws that determine the behavior of plasmas, such as Maxwell's equations and those of classical statistical mechanics, are well known, obtaining their solution under realistic conditions is a nonlinear scientific problem of extraordinary complexity. In a magnetically confined plasma, the interplay between the complex trajectories of individual charged particles and the collective effects arising from the long-range nature of electromagnetic forces lead to a wide range of waves, oscillations, and instabilities characterizing the medium. As a result, there is an enormous range of temporal and spatial scales involved in plasmas of interest. As illustrated in Figure 11, the relevant physical processes can span over ten decades in time and space. Effective prediction of the properties of plasma systems (such as energy-producing fusion experiments) depends on the successful integration of many complex phenomena spanning these vast ranges. This is a formidable challenge that can only be met with advanced scientific computing in tandem with theory and experiment.

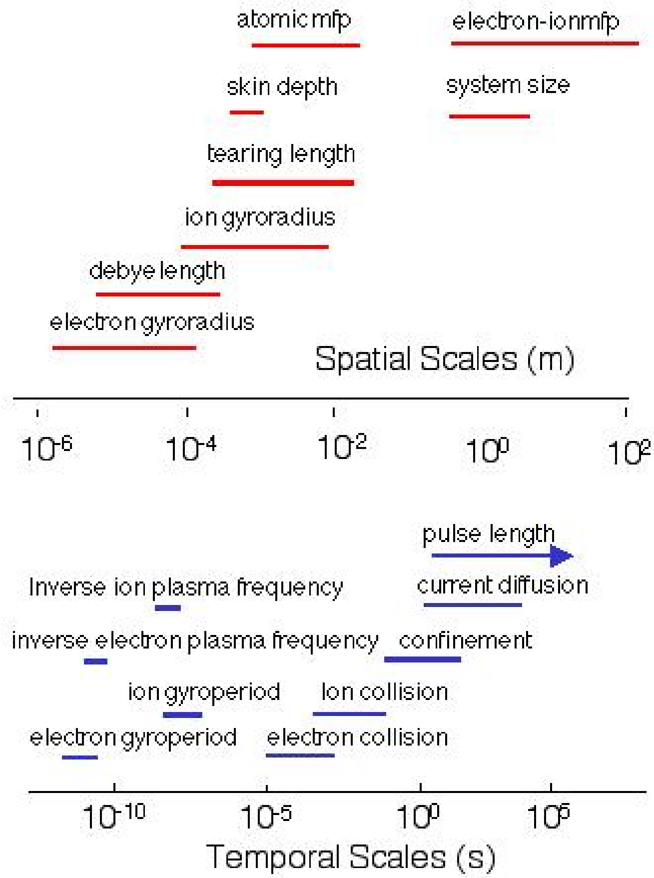


FIG. 11. Widely disparate temporal and spatial scales in fusion energy sciences research.

Magnetically confined plasmas are naturally subject to both large- and small-scale disturbances (*instabilities)* which thermodynamically relax the system to a lower energy state. In order to generate more power output than it takes to keep the plasma hot in fusion experiments, for example, it is necessary to first understand these complex, collective phenomena, and then devise the means to control them. The larger-scale (*macro*) instabilities can produce rapid topological changes in the confining magnetic field, resulting in a catastrophic loss of power density. In addition, smaller-scale (*micro*) instabilities can also prevent efficient hot plasma confinement by causing the turbulent transport of energy and particles. Because of the complexity of the kinetic, electromagnetic, and atomic physics equations describing the behavior of plasmas, researchers in this field have a long history of productive use of advanced computation and modeling. Beginning with the establishment of the predecessor to the Department of Energy (DOE) National Energy Research Scientific Computing Center (NERSC) 27 years ago, the U.S. plasma physics community has been actively and productively engaged in the simulation and modeling of plasma confinement and the interactions of plasma with its surroundings.

As described in the recent National Research Council report[2] assessing plasma physics, the scientific challenges related to magnetically confined plasmas can be categorized into four areas:

[2] "An Assessment of the Department of Energy's Office of Fusion Energy Sciences Program," National Research Council, Fusion Science Assessment Committee, 2001, *Final Report*, Washington, D.C.: National Academy Press.

(1) macroscopic stability, (2) wave-particle interactions, (3) microturbulence and transport, and (4) plasma-material interactions.

1. Because charged particles, momentum, and heat move more rapidly along the magnetic field than across it, magnetic fusion research has focused on magnetic traps in which the magnetic field lines wrap back on themselves to cover a set of nested toroidal surfaces called magnetic flux surfaces (because each surface encloses a constant magnetic flux). Macroscopic stability is concerned with large-scale spontaneous deformations of magnetic flux surfaces. These major displacements or macroinstabilities are driven by the large electrical currents flowing in the plasma and by the plasma pressure.
2. Wave-particle interactions deal with how particles and plasma waves interact. Detailed calculation of particle motions in background electromagnetic fields are needed to assess the application of waves to heat the plasma as well as address the dynamics of energetic particles resulting from intense auxiliary heating and/or alpha particles from possible nuclear fusion reactions.
3. Microturbulence and the associated transport come from fine-scale turbulence, driven by inhomogeneities in the plasma density and temperature, which can cause particles, momentum, and heat to leak across the flux surfaces from the hot interior to be lost at the plasma edge.
4. Finally, plasma-material interactions determine how high-temperature plasmas and material surfaces can co-exist.

Progress in scientific understanding in all of these areas contributes to integrated design considerations for fusion devices and demands significant advances in physics-based modeling capabilities. Indeed, advanced scientific codes are a realistic measure of the state of understanding of all natural and engineered systems.

The developmental path for computational codes as validated tools for scientific discovery in plasma physics can be visualized as shown in Figure 12. This multidisciplinary collaborative process begins with basic theoretical research laying the foundations for the mathematical formulation of the physical phenomena of interest observed in experiments. Computational scientists produce the codes which solve these equations, ideally using the best possible algorithms which efficiently utilize modern high-performance computers. Opportunities clearly exist to provide valuable support for partnerships with applied mathematicians who could provide the basic mathematical algorithms and with the computer scientists who could provide the requisite computer systems software.

The computational scientists must then engage the research applications scientists in the critical scientific code validation phase where the newly computed results are compared against experimental/observational data. This is a major challenge involving a hierarchy of benchmarking criteria which begin with cross-checks against analytic theory, empirical trends, and suggestive “pictorial” levels of agreement. It then graduates to sensitivity studies, where agreement is sought when key physical parameters are simultaneously varied in the simulation and experiment/observation. At the next level, richer physics validation is dependent on the availability of advanced experimental diagnostics which can produce integrated measurements of key physical quantities such as spectra, correlation functions, heating rates, and other variables of interest.

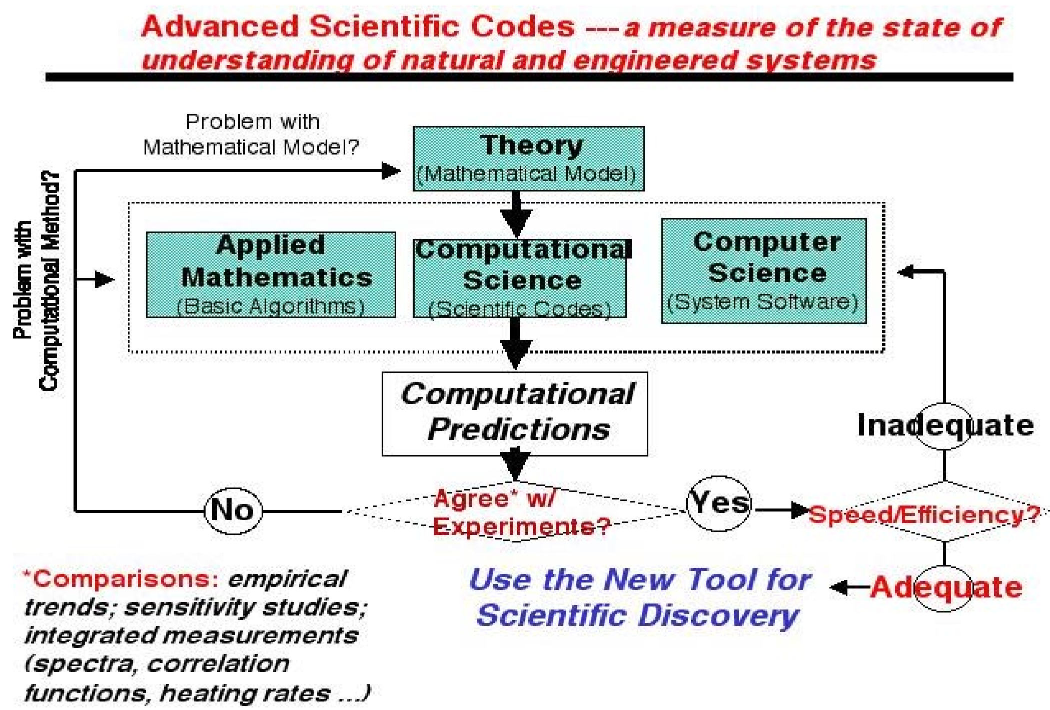


FIG. 12. Development of high performance codes as validated tools for scientific discovery.

If the simulation/experimental data comparisons are unsatisfactory at any of these validation levels, the work flow moves back to: (1) the theorists (in consultation with experimentalists) if the problem seems to be with the mathematical model, and (2) computational scientists (in consultation with applied mathematicians and computer scientists) if the problem appears to be with the computational method. Even when the theory/experiment comparisons prove satisfactory, code performance criteria for speed and efficiency could dictate another round in the computational science box.

If all criteria are met, then the new "tool for scientific discovery" can be effectively utilized for interpreting experimental data, designing new experiments, and even predicting new phenomena of interest. This cycle of development will, of course, be repeated as new discoveries are encountered. To support timely and effective code development, significant new investments must be made in these interdisciplinary components of computational plasma physics.

Simulation domains have both minimum and maximum limits on spatial and temporal resolution so that any given plasma simulation can only address a finite range of space and time scales. In the past, this issue has been dealt with by deriving simplified sets of equations, or "reduced equations," that are valid for restricted but important ranges of time and space scales. Examples of these are the gyrokinetic equations[3] and the MHD equations.[4] While the reduced equations have enabled progress in the past, they have fundamental restrictions on their regions of validity. In actual laboratory or natural plasmas, phenomena occurring on different time and space scales interact and influence one another. It thus becomes essential to utilize more general equations

[3] P. Rutherford and E. Frieman, Phys. Fluids **11**, 569 (1968).
[4] J. P. Freidberg, *Ideal Magnetohydrodynamics,* Plenum Press, New York and London (1987).

which are valid on a wider range of space and time scales and which will accordingly increase the simulation domains.

At the most fundamental level, a plasma can be described by kinetic equations for the distribution function within a six-dimensional (plus time) phase-space of each particle species. These kinetic equations are coupled to each other through self-consistent electric and magnetic fields. The simulation techniques used in plasma physics fall into two broad categories: particle-in-cell models and fluid models.

The particle-in-cell methods proceed by integrating a (possibly reduced) kinetic equation in time by advancing *marker* particles along a representative set of characteristics within the (possibly reduced) phase space. Particle-in-cell simulation techniques developed over the last 20 years include *finite-sized particles*,[5] which reduce the "noise" caused by discrete marker particles; *gyrokinetics,*[6] a reduction of the full kinetic equation to a five-dimensional phase space which removes high-frequency motion not important to turbulent transport; and *delta-f,*[7] a prescription for integrating the gyrokinetic equation along characteristics which further reduce the discrete particle noise. These advances have reduced the number of particles required to faithfully represent the physics, and have dramatically increased the accuracy and realism of the particle-in-cell simulation technique.

The fluid models proceed by advancing velocity moments of the kinetic equation in time. The best known of these are the extended-MHD models,[8] which represent the plasma as one or more interacting conducting fluids. This higher-level description frees the model of many fine-scale resolution requirements and makes feasible the simulation of large-scale motions and instabilities. Extensive theoretical analysis over the years has led to refinements of the fluid models and improved the closure relations so that many nonfluid effects, such as particle motion and wave-particle resonances, can be represented at some level.

Many key macroscopic simulation problems in the plasma sciences share with fluid simulation challenges in other fields the common features of extreme temporal and spatial stiffness, severe spatial anisotropy, and complex boundary conditions. These characteristics make them among the most challenging problems in computational physics. Aided by the successful implementation of unstructured mesh algorithms and the application of advanced visualization resources to deal with complex three-dimensional toroidal structures, recent nonlinear simulations of an internal magnetic reconnection event in the National Spherical Torus Experiment (NSTX) are shown on Figure 13, where the sequence of images depicts the evolution of an internal MHD instability, showing how the hot inner region of the plasma (red area) interchanges with the cold outer region (green area) via magnetic reconnection. When compared with soft X-ray measurements of the thermal response, the simulation results show encouraging agreement, with good prospects for improvement when these calculations are carried out at higher resolution on more powerful computing platforms.

[5] J. Dawson, Reviews of Modern Physics, **55**, 403 (1983).
[6] W. Lee, J. Comput. Phys. **72**, 243 (1987).
[7] A. Dimits and W. Lee, J. Comput. Phys. **107**, 309 (1993).
[8] W. Park et al., Phys. Plasmas **6**, 1796 (1999).

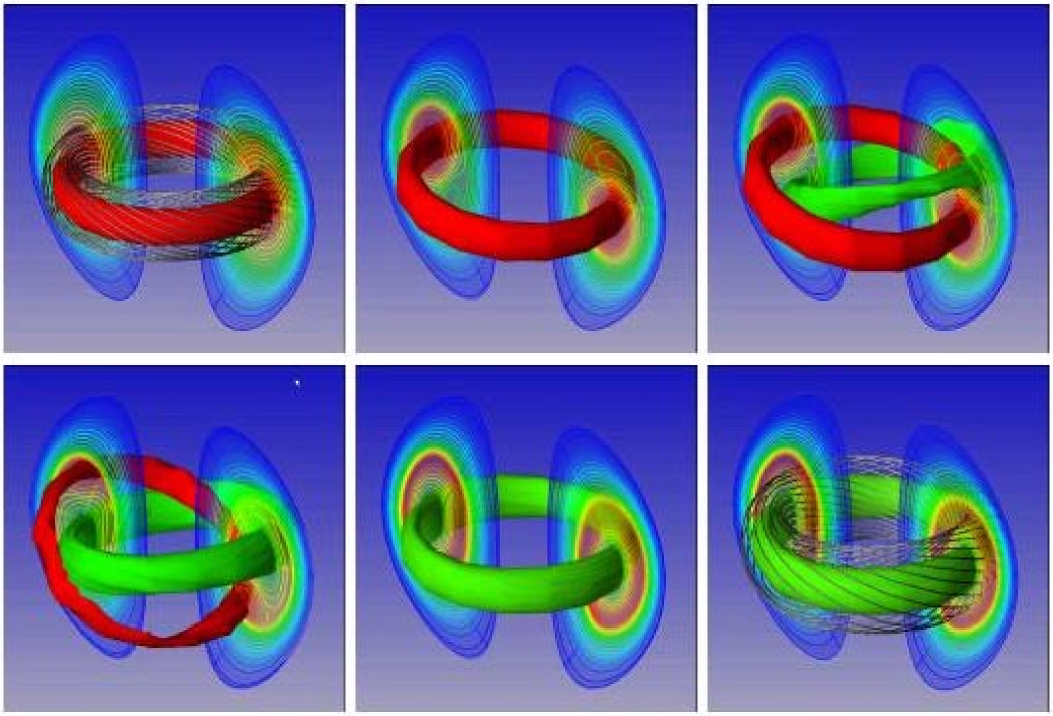


FIG. 13. Nonlinear simulation of internal magnetic reconnection event in the National Spherical Torus Experiment.

Understanding turbulent plasma transport is not only an important practical problem but is generally recognized as a true scientific grand challenge which is particularly well suited to be addressed by terascale MPP computational resources. Improved models with efficient grids aligned with the magnetic field have now been developed to address realistic 3D (toroidal) geometry. This involves the multi-scale challenge of capturing the physics both on the small scale of the fluctuations (microinstabilities) and the large scale of the equilibrium profile variations. Such simulations are being counted on to advance the scientific understanding of the turbulence responsible for the increased ("anomalously large") transport of particles, momentum, and heat, which are experimentally observed to be significantly greater than levels expected from the collisional relaxation of toroidally-confined plasmas. This is particularly important for magnetic fusion research because the effective size (and therefore cost) of an ignition experiment will be determined largely by the balance between fusion self-heating and turbulent transport losses.

With the advent of teraflops-scale MPP computers, high-resolution simulations of the fundamental equations governing turbulent transport become feasible. Although progress in capturing the ion dynamics has been impressive, the description of the electrons is still being upgraded to include important kinetic effects such as trapping in equilibrium magnetic wells, drift motions, and wave-particle resonances. Significant challenges also remain in extending the present capabilities for dealing with electrostatic perturbations to include magnetic perturbations, especially in cases where they are sufficiently large to alter the actual geometric properties of the self-consistent magnetic field. In such circumstances, microinstabilities can drive currents parallel to the equilibrium magnetic field, which in turn produce magnetic perturbations in the perpendicular direction. These kinetic electromagnetic waves can modify the stability properties of the electrostatic modes or act as separate instabilities which can alter the magnetic topology. In this sense, the kinetic simulations would encounter the major multi-scale task of also dealing with the larger-scale phenomena associated with the aforementioned MHD studies.

A good example of progress in kinetic simulations involves studies of electrostatic turbulence suppression produced by self-generated zonal flows within the plasma. Results from particle-in-cell global gyrokinetic simulations show that the suppression of turbulence caused by prominent instabilities driven by ion temperature gradients (ITG) is produced by a shearing action which destroys the finger-like density contours which promote increased thermal transport in a 3D toroidal system.[9] This dynamical process is depicted by the sequences shown in Figure 14. The lower panels, which show the nonlinear evolution of the turbulence in the absence of flow, can be compared with the upper panels, which illustrate the turbulence decorrelation caused by the self-generated E × B flow. This is also a good example of the effective use of one of the fastest non-classified supercomputers in the world (the 5 teraflops IBM-SP at NERSC).

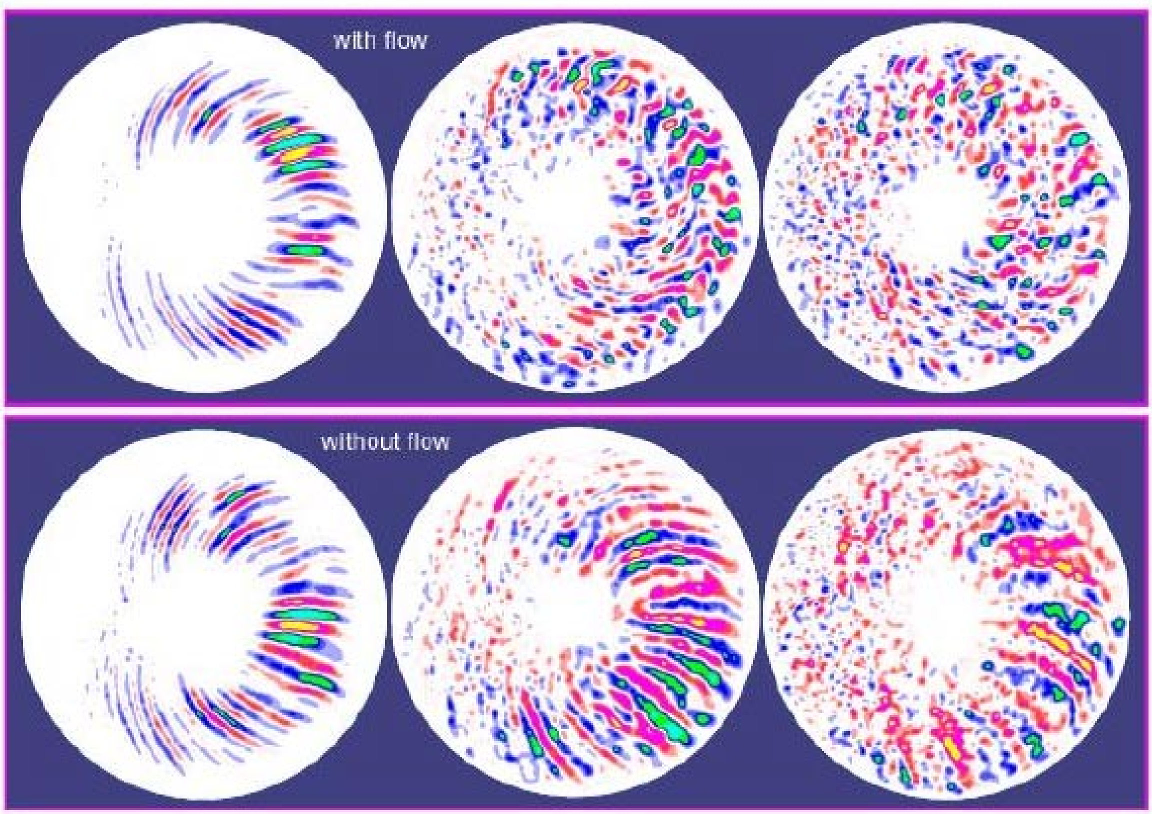


FIG. 14. Turbulence reduction via sheared plasma flow compared to case with flow suppressed.

The most recent simulations of this type used one billion particles with 125 million grid-points over 7000 time-steps to produce significant new physics results. For the larger sized reactor-scale plasmas of the future, the present simulations indicate that the relative level of turbulent heat loss from electrostatic turbulence does not increase with size. In addition to addressing experimental validation challenges, the interplay between analytic theory and advanced simulations will be

[9] Z. Lin et al., Science **281**, 1835 (1998).

increasingly important. For example, impressive progress in physics understanding of the nonlinear processes associated with zonal flow dynamics has resulted both from simulations directed by analytic theory as well as from simulation results which have inspired new analytic models.[10]

Terascale Computing Challenges and Opportunities

Addressing the most challenging scientific issues in plasma physics requires advanced code development efforts which are important for most areas of research. The primary task involves enhancing the physics models and developing more efficient algorithms to efficiently deal with the associated problems. Challenges include: (1) multi-scale physics such as kinetic electromagnetic dynamics which have been discussed in the previous section; (2) improved algorithms; and (3) scalability of codes. With regard to item (2), in addition to making sure presently working algorithms are scalable to new computing platforms, innovative numerical algorithms will have to be invented to make progress. In addition, powerful approaches such as adaptive mesh refinement for higher dimensionality phase-space need to be actively pursued. Item (3) deals with the ability to efficiently implement existing codes on the most powerful MPP supercomputers to enable simulations of larger problems. The plasma science community has had success in developing codes for which computer run-time and problem size scale well with the number of processors on massively parallel machines. For example, as depicted in Figure 15, the microturbulence code GTC has recently demonstrated full scalability for 2000 processors.

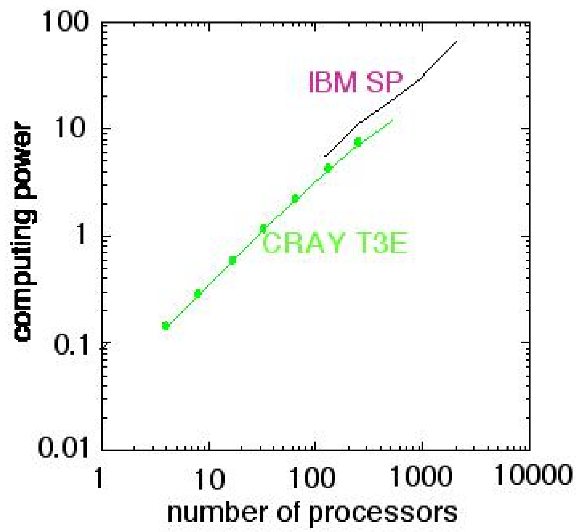


FIG. 15. Scalability of the GTC micro-turbulence code. Y-axis: number of particles (in millions) which move one step in one second.

It should be emphasized that a consequence of the effective utilization of supercomputers is the tremendous amount of data generated. The microturbulence example described earlier in this paper alone produced over 2 terabytes of data. Hence, coordinated efforts will be essential not only to help accelerate progress on the impressive state-of-the-art physics advances but also to provide the necessary tools for data visualization, mining, and manipulation. Means of dealing with the data glut in the interactive exploratory visualization of terascale simulations, including image rendering, must be developed. Another key priority involves developing a set of diagnostic and visualization tools that will allow real-time interaction with the simulated data. This is important for assisting applications scientists in testing theories/hypotheses and in answering specific questions about the key physics within the computational models. In order to realize the benefits from advancements in understanding, it will be necessary to periodically update existing

[10] M. Rosenbluth and F. Hinton, Phys. Rev. Lett **80**, 724 (1998); L. Chen et al., Phys. Plasmas **7**, 3129 (2000); P. Diamond et al., Nucl. Fusion **41**, 1067 (2001).

integrated models to ensure that they reflect the fresh insights gained from these new "tools for discovery."

Improvements in plasma diagnostic techniques have made it increasingly feasible to demonstrate more in-depth correlations between experimental results and theoretical models. For example, the development of diagnostic instruments capable of making high-resolution measurements of electric and magnetic fields and cross-sectional measurements of turbulent fluctuations have greatly improved the basic understanding of the mechanisms controlling plasma confinement. As in dealing with the output from terascale simulations, maximizing the effectiveness of simulation/experiment comparisons will also necessitate addressing critical computer science and enabling technology (CSET) issues in the area of data management and visualization. Indeed, the power of advanced computing to solve challenging problems can be fully exploited only if a capable infrastructure is established and effective software tools are made available. Terascale computing requires complementary software that scales as well as the hardware and which provides an efficient code development environment. In general, networking, database, visualization, and other infrastructure tools are critically needed to strengthen the coupling between terascale simulations with theory and experiment.

The applications development challenge for terascale computing will invariably involve multiple research institutions. Hence, efficiency will require system integration and the availability of middleware software that allows maximal use of available computing platforms to support joint application development projects. Accordingly, modern object-oriented code development methods are needed to facilitate sharing code development efforts among collaborators from numerous research groups.

The increased emphasis on advanced computations in fusion energy sciences can also stimulate mutually beneficial scientific alliances with other applications areas which could serve to raise the overall standards for scientific programming as well as enable sharing with other fields the insights gained in the process of obtaining successful solutions to scientific problems of extraordinary complexity. This can complement efforts to actively develop and nurture opportunities for attracting, educating, and retaining bright young talent essential for the future technological health of the field. Computational plasma science carries an exciting vision which can serve this role. While it is encouraging that many key recent advances have involved major contributions from young scientists, greater efforts are required to educate the next generation of researchers with capabilities that cross-cut traditional boundaries.

Concluding Observations on Opportunities in Plasma Science

Advanced computations are demonstrably aiding progress toward gaining the physics knowledge needed in all areas of plasma science.[11] The plasma physics community is both effectively utilizing and contributing to exciting advances in information technology and scientific computing. This has produced a more stimulating environment for transforming research and accelerating scientific understanding, innovation, and discovery. However, the emerging research opportunities in this field are not being adequately addressed by existing programs. Timely investments will enable productive partnerships involving laboratories, universities, and industries, and also a sharing with other fields of the insights gained in the process of obtaining

[11] W. M. Tang, "Advanced Computations in Plasma Physics," Physics of Plasmas **9**, 1856 (2002).

successful solutions to scientific problems of extraordinary complexity. Indeed, advanced computation is a natural bridge for fruitful collaborations between scientific disciplines which will be of mutual benefit to all areas.

Advanced scientific computations have the potential to revolutionize plasma physics research by significantly improving scientific understanding of experimental data, by stimulating new theoretical ideas, and by helping produce innovations leading to new discoveries and to the most attractive and viable designs for future facilities. This impact will in turn help raise the visibility of computational physics across all fields. Computational plasma science is also helping to attract, educate, and retain bright young talent essential for the future of the field. In general, the present is an optimal time to initiate strong new programs in computational plasma science.

E. Atomic and Molecular Physics

In atomic and molecular physics, the basic equations are known, at least at nonrelativistic energies, where the majority of practically important applications occur. The Schrödinger equation and its semiclassical and classical limits are known to describe the structure and dynamics of atomic and molecular systems. However, these systems, even relatively small ones, show a rich diversity of phenomena that arise from the intrinsically complex many-body physics of systems of charged particles. Moreover, they pose a fundamental problem that remains a challenge to computational research: dynamics involving electronic excitation and energy transfer.

Much of the progress in this area in the last two decades, at least the portion dealing with electronic structure and dynamics, has benefited from the progress made in quantum chemistry towards accurate descriptions of electronic structure. But there is a key and critical difference between much of the research in atomic and molecular physics and that in quantum chemistry. The difference is that atomic and molecular physics frequently focuses on processes and dynamics that involve electronically excited states and the responses of molecular or atomic systems to perturbations and collisions—while great progress can be made in chemistry by considering only the ground electronic states of complex molecules and ensembles of molecules.

Other aspects of atomic and molecular physics involve the dynamics of nuclear motion in such processes as reactive collisions of atoms and molecules, cold collisions and the phenomena associated with Bose condensation, photodissociation, electron-impact dissociation, and so forth.

Finally, the underpinnings of many ideas for quantum computing, and the understanding of the systems which display the quantum coherence and control on which it depends, depend on accurate and complete theoretical and computational treatments of interacting molecules and atoms.

Cold Collisions and Bose-Einstein Condensation

Because experiments dealing with laser cooling and trapping of atoms and the production of Bose-Einstein condensates (BEC) have received worldwide attention and were recognized with Nobel Prizes in Physics in 1997 and 2001, these topics have come to dominate current research efforts in atomic, molecular, and optical physics at many institutions. The presentations at meetings of the American Physical Society's Division of Atomic, Molecular and Optical Physics in recent years have become increasingly dominated by those working in the area of cold collisions. The experimental activity has spawned an associated set of theoretical efforts (Figure 16) that seek to produce a complete and predictive treatment of the phenomena associated with ultracold atoms and molecules and the properties of pure and mixed condensates as well as their interaction with surfaces and each other.

There are a host of basic issues and computational challenges associated with the dynamical interaction of cold atoms, since the collision physics is dominated by subtle effects of weak forces and long-range interactions. Colliding neutral atoms, confined in a trap, can be photoassociated to produce exotic molecular states of purely long range that have never been seen by conventional techniques. The structure and dynamics of such exotic states of cold molecules is fertile ground for computational theory. There are also computational challenges that revolve around developing an atomic-level simulation from first principles of the correlated motion of a large number of atoms in a condensate. The theoretical and computational workhorse for studying many of the properties of a weakly interacting BEC, including the formation of quantized vortices and the generation and propagation of soliton waves, is the nonlinear Gross-Pitaevskii wave equation, which is used to investigate the internal dynamics of the condensates. Experimental thrusts toward the production of larger condensates and the possibility of producing a molecular BEC are already prompting theoretical investigations of condensates that go beyond the Gross-Pitaevskii equation. This rapidly evolving field will continue to offer a broad array of theoretical and computational challenges.

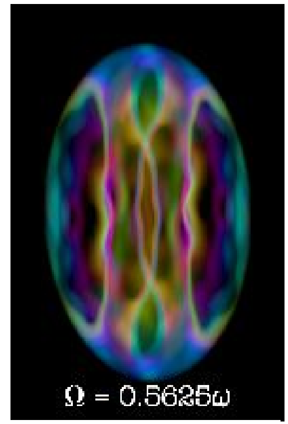


FIG. 16. Bose-Einstein condensate.

The Few- and Many-Body Physics of Electrons

Collisions of electrons with molecules and atoms create the energetic species that drive chemical and physical changes of matter in environments that range from plasmas to living tissue. The molecules used to etch semiconductor materials do not react with silicon surfaces unless they are subjected to electronic collisions in the low-temperature, high-density plasmas used in plasma etching and in plasma-enhanced chemical vapor deposition. Breaks in the DNA of living systems caused by ionizing radiation have been shown to be due primarily to collisions with secondary electrons that attach to the components of DNA molecules or to the water around them and drive bond dissociation. Secondary electron cascades in mixed radioactive/chemical waste drive much of the chemistry that determines how those materials age, change, and interact with the natural environment. Electron collisions create the reactive molecular fragments in the plasma devices being developed and used to destroy undesirable compounds or remediate NO_x in combustion exhausts.

To understand any of these chemical and physical changes, we require an understanding of the fundamental processes that underlie them. A key point is that at the collision energies of interest, ranging up to about 100 eV, there is *an infinite number of electronic states* that are coupled by the collisions. That simple fact is the fundamental reason that theory and modeling of these collisions remains a challenging but essential problem in atomic and molecular physics. At these energies. the electron dynamics are highly correlated, and none of the high-energy or perturbative approximations that simplify the problem at higher collision energies apply.

At these intermediate and low energies, electronic collisions are uniquely effective in transferring energy to and from the electronic degrees of freedom of the target atom or molecule. That is the fundamental reason that new developments in modern fluorescent lighting and plasma displays are distinguished by their energy efficiency. The fact that the incident electron not only has the same mass as those that bind the target molecule together, but is physically indistinguishable from them, means that at low energies the coupling to the molecule's internal degrees of freedom is especially strong.

The computational study of electron-molecule collisions in important technological contexts is not merely a bright prospect, it is a current reality. There are many useful calculations that can be performed with currently available variational methods on a variety of technologically significant polyatomic species. These techniques have already been shown to be capable of providing highly accurate cross sections in a few important cases involving small polyatomic targets where experimental data was available for comparison. In many other cases, especially those involving reactive species and complex polyatomic target gases, theory has proved to be the "only game in town" and thus the sole source of critically needed collision data. While improvements to existing methodologies are under way, studies using existing codes will continue to be of immediate practical benefit.

There are practical limitations on what can be expected from current methods without substantial future investment. Nearly all calculations to date are fixed nuclei approximations. For very large target molecules, the calculations are currently limited to the use of simple target wave functions. Electronic excitation can be studied using only a small number of coupled states, and the extent to which polarization effects (which are very important at energies below a few electron volts) can be accurately treated depends very strongly on the complexity of the target. An investment now will allow investigators to leverage existing methods and computational platforms to make an immediate impact on both plasma processing and environmental chemistry. Benchmarking against experimental measurements will be essential to bring the credibility that will be needed for a sustained future effort.

Advances in electronic structure theory over the last fifteen years, combined with the emergence of terascale computing platforms, on which many of those structure codes run, have provided an entirely new context in which to attack the electron-polyatomic molecule scattering problem. The last ten years have also seen separate advances in electron scattering theory that have made it possible to carry out multichannel electronic excitation calculations on polyatomics. In addition, access to terascale computing platforms will become more broadly available to researchers in the next five years, completing the arsenal for addressing these problems successfully. This combination of advances arising in different research communities has set the stage for the first comprehensive theoretical attack on problems in electronic collisions.

Collisions with Photons and Intense Field Effects

Essentially all of the problems that are involved in collisions with electrons at intermediate energies arise as well in photoionization and photodissociation, even with incident intensities that do not enter the regime of intense field effects. These problems are becoming more tractable experimentally, and the detailed dynamics of multiple ionization and photofragmentation are being studied with imaging techniques such as COLTRIMS (cold target recoil ion momentum spectroscopy), providing both a need and an opportunity for computationally intensive investigations.

Large-scale simulations are necessary to reveal the physics of intense field effects, which involve many photons and which are dominated by nonlinear effects, such as above threshold ionization, high harmonic generation, and multiple ionization. Even for a system with two or three electrons, these calculations are challenging and subtle. New effects, like "jets" of outgoing electrons, have been predicted in such calculations. The solution of the time-dependent Schrödinger equation for a small *molecule* including the radiation field is still a large-scale computational problem which must involve nuclear dynamics as well. It is in these simulations that the dynamics of intense field effects will be elucidated. Even the visualization of the time-dependent dynamics of these highly correlated multielectronic systems is a significant computational and intellectual challenge.

The Motion of Atoms and Molecules Coupled to Electron Dynamics

Although it is generally considered to be a part of chemical physics instead of molecular physics, the reactive scattering of molecules is a closely related field. One of the presentations at the workshop focused on reactive scattering and new quantum approaches that allow the treatment of complex reacting systems. Those systems can involve multiple electronic states, and the nuclear dynamics thereby proceeds on multiple electronic surfaces on which the dynamics is coupled. This entire field is closely related to the study of important processes such as electron impact dissociation and dissociative attachment, which are the focus of molecular physics. The common computational and theoretical challenges and approaches that these investigations share are another example of how modern computational physics knits together, in an intellectual as well as practical sense, research in the various subdisciplines of physics.

In many elementary processes of practical interest in low-temperature plasmas and in radiation damage to materials and living systems, it is the dissociation of polyatomic molecules that is the key step. Only the rudiments of those dissociative collisions are understood, and it is not an exaggeration to state that it is currently well beyond the state of the art to predict the branching ratios of the fragments and their excited states. The central question of what reactive molecular fragments are present can only be answered by an understanding of how they are generated. The transient nature of the products, and the detailed nature of the required information about the states of the neutral fragments, makes the relevant experimental determinations on polyatomic dissociation extremely challenging. There is a paucity of even benchmark experimental data.

All of these processes happen both in gas and condensed phases. They are not yet understood even in the gas phase, and the alteration of the gas phase processes in condensed media remains an almost completely unanswered question. In the key application areas of environmental remediation and radiation damage to living tissue, they happen almost exclusively in the liquid or solid phases. Thus a key challenge for this field is the connection between investigations in the gas phase and the still more difficult work that remains to be done in clusters, at interfaces, and in condensed media.

F. Nuclear Physics

Many forefront questions in contemporary theoretical nuclear physics can only be addressed using computational methods. For example, understanding the confinement of quarks and the structure of hadrons requires lattice QCD calculations; solving the quantum many-body problem for nuclei requires quantum Monte Carlo calculations; and understanding the origin of the elements in supernova explosions requires multi-dimensional simulations. Theoretical work on these questions is crucial to obtain the full physics potential of the investments that have been made at Jefferson Lab and the Relativistic Heavy-Ion Collider (RHIC) as well as new investments that are recommended for the Rare Isotope Accelerator (RIA) and the underground neutrino physics lab. Recent advances in computational physics and computer technology represent great opportunities for breakthroughs in nuclear science.

Lattice QCD is crucial for answering fundamental questions in strong interaction physics: What are the physical mechanisms of confinement and chiral symmetry breaking? What is the origin of the nucleon mass and spin? What is the quark and gluon structure of the nucleon? And what is the phase structure of strong-interaction matter as functions of temperature and baryonic density? QCD is an integral part of the Standard Model and has been accepted universally as the fundamental theory of strong interactions. The QCD vacuum and its hadronic excitations are intricate quantum mechanical systems composed of strongly coupled, ultra-relativistic quarks and gluons. Understanding nonperturbative QCD is one of the most difficult challenges in modern theoretical physics. Lattice field theory is the only approach at present to solve, rather than model, the strong interaction systems. In the short term, lattice QCD will have decisive impact on interpreting experimental data coming from Jefferson Lab, RHIC, and other experimental facilities. In the long run, it is an indispensable tool in the grand scheme of understanding matter from the level of quarks to the level of the cosmos.

Multi-dimensional supernova simulations are essential to understand the origin of heavy elements and the mechanism of supernova explosions. Supernova explosions are not just spectacular events in the history of the Universe, these explosions are thought to have produced about half of the heavy nuclei found in nature. Understanding the physical mechanism in supernovae is an outstanding computational and theoretical “grand challenge” problem in nuclear physics and astrophysics. This problem encompasses nuclear theory, astrophysics, and computer science, requiring modeling of the nuclear equation of state up to at least four times nuclear density. Simulations involve hydrodynamics, convection, and shock wave propagation, and the neutrino-nucleus microphysics that is crucial to both the explosion mechanism and associated nucleosynthesis.

Significant progress has been made in the large-scale numerical simulations of supernova explosions. The first semi-realistic two-dimensional simulations of supernova explosions have been performed. This could be an important step in understanding the mixing apparent in the ejecta of observed supernovae. Full Boltzmann neutrino transport has been implemented in one-dimensional models. Progress has been made in descriptions of the progenitor, e.g., multi-dimensional models that account for convection and rotation. Improved electron capture and beta decay rates and improved neutrino opacities have made the input microphysics much more realistic. Progress has been made in modeling the r-process, including improved weak interaction rates, a better understanding of the effects of mass formula uncertainties and phenomena such as the vanishing of shell closures, and inclusion of neutrino postprocessing effects.

Yet there are many open questions in supernova modeling that need to be addressed. The key theoretical problem is to develop a supernova standard model that incorporates realistic neutrino transport and microphysics. Current one-dimensional models generally fail to explode. This could reflect some flaw in our understanding of the physics, or the importance of doing multi-dimensional simulations. In current multi-dimensional simulations, a variety of physics—neutrino heating, convection, rotation, magnetic fields, general relativity—are also inadequately modeled. It is not known which of these effects may be essential to successful explosions. Nor is it clear how dependent (or independent) explosions may be on the class of progenitor star. The development of multi-dimensional models with realistic neutrino transport and microphysics is possible once terascale computational machines are available.

Understanding nuclear structure and nuclear reactions continue to be major goals in nuclear science. Because of extreme complication of the nuclear many-body problems, nuclear theorists for many years have resorted to various models for which the error control is poor and improvements are hard. Quantum Monte Carlo methods now open the door to the precision calculation of nuclear structure in terms of the interaction between the basic constituents, protons and neutrons. Outstanding progress has been made in *ab initio* calculations in quantum many-body systems. We can now tackle low-energy nuclear physics with a general assumption that nuclei are bound states of nucleons interacting through one-, two-, and perhaps three-body potentials. In the low-energy domain, other contaminants, such as deltas, excess pions, etc. may be absorbed through interaction potentials. Thus, the starting point of a fundamental nuclear description is the many-body Schrödinger equation with phenomenological potentials.

Since the early 1980s, quantum Monte Carlo (variational and Green's function Monte Carlo) calculations with spin-isospin dependent interactions have become possible for nuclei with a few nucleons. Successful applications have been made initially in $A = 3$ and 4, then to $A \leq 6$, $A \leq 8$, and more recently $A = 10$ systems. Because the configuration space grows rapidly as a function of the atomic number, the quantum Monte Carlo calculations depend critically on large-scale computational resources if they are to be extended to a large system. These calculations revolutionize our understanding of many aspects of the nuclear structure. They give us a clear picture of the origin of the nuclear binding. They elucidate the role and form of three-body forces and the importance of relativistic corrections. They yield important physical observables such as the nucleon distribution functions, electron scattering form factors, response functions, and the nuclear shapes. Many of the these results have important applications in astrophysics. The current calculations have errors in binding energy at a level of less than 1 percent. Ultimately, with development of new algorithms based on a cluster expansion and effective field theory ideas, accurate *ab initio* calculations for many nuclei in the chart of the nuclear isotopes may become possible. This not only will have a tremendous impact on the physics of RIA, but also will bring nuclear physics to a new era.

The major increases in computing power being provided by the NSF PACI Program and the DOE supercomputer centers will have a major impact in all of the research areas outlined above. However, to fully capitalize on the investments being made in computer hardware, it is necessary to also make investments in software and, most importantly, in the young scientists who will carry forward the research. An NSF program in computational physics could play a major role in this regard.

G. Applied Mathematics and Computer Science

Applied mathematics and computer science can and must play a significant role in any computational physics initiative. We illustrate this point by reviewing several areas where mathematics and computer science have made significant contributions in the past and should play a role in the future.

Applied Mathematics

Mathematicians have had great success in creating solid foundations for scientific discovery with precise mathematical assertions that provide fundamental understanding and which enable the systematic design of algorithms.

One such area is numerical linear algebra, where Householder, Wilkinson, and others developed the backward error analysis that provides a complete understanding of the effect of roundoff error, permitting the design and implementation of robust algorithms for the solution of linear equations. Another such area is partial differential equations, where von Neumann, Lax, Godunov, and others developed the basic convergence theory that stability plus consistency implies convergence; even for discontinuities such as shocks, discrete conservation, and convergence, this theory implies that the correct weak solution will be computed. These are fundamental statements that apply to many of the scientific areas computational physicists are working in today.

Another area of contribution is numerical software. Over the last twenty years, portable, widely available, and robust software that is optimized for performance has been developed in several sufficiently mature areas. Examples include dense linear algebra (Linpack, Eispack, and the recent versions LAPACK, ScaLAPACK, and Atlas/PHiPAC, which automatically tune for cache performance); nonlinear systems and optimization (MINPACK); and ordinary differential equations (numerous packages such as Gear, VODE/PVODE, and DASSL). The quality of these tools is sufficiently high that they form the core of higher-level environments, such as MATLAB for linear algebra, NEOS for optimization, and CHEMKIN for detailed chemistry simulations. The finite element method, which originated in the engineering community, was put on sound footing by mathematicians. Adaptive finite element packages, both commercial and free (NASTRAN and PLTMG, for example), are available to solve elliptic equations.

More recently, algorithm developments such fast Fourier transforms (FFTs), multigrid, the fast multipole method, and high-resolution Riemann solver-based methods have revolutionized computation in the relevant fields. The combination of mathematical analysis, geometric insight, and careful algorithm design is at the heart of these methods.

A variety of newer algorithm developments have the potential to continue this level of impact. Block-structured adaptive mesh refinement (Figure 17), initially applied to hyperbolic equations such as those in fluid and solid mechanics, electromagnetics, and plasma physics, has recently been extended to density functional theory and the Vlasov equations. Similar advances with interface tracking result from the use of level-set and volume-of-fluid methods and embedded boundary techniques. These advances enable more precise and accurate solutions to problems with phase boundaries or interfaces between different media, and greatly improve the representation of surfaces in photolithography, etching, and deposition processes.

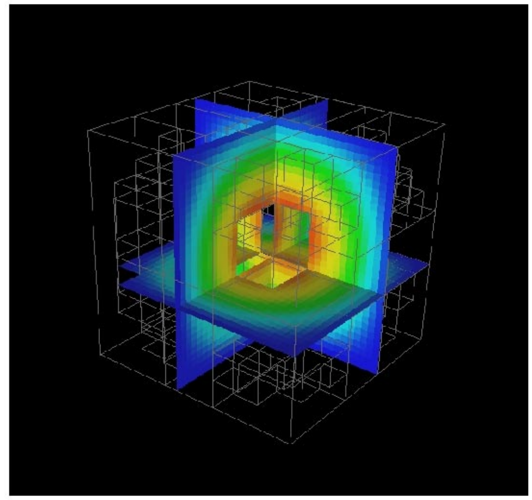


FIG. 17. Adaptive mesh refinement.

This section has so far concentrated on contributions that mathematicians are already making. There are difficult problems in computational physics that haven't (yet) had much computational mathematical input. One such area is the coupling of multiple physics models. Examples include climate modeling (coupling fluid mechanics, thermodynamics, moisture, radiation, and of course the coupling between the atmosphere and oceans), plasma physics (particularly in regimes in which combinations of fluid and kinetic behavior occur), and problems arising in computational astrophysics and cosmology, such as general relativity and radiation hydrodynamics. In order to develop methods that are computationally tractable, it is necessary to decompose such coupled computations into components using fractional step methods and related approaches. To obtain methods that are demonstrably stable and preserve the accuracy of the underlying component discretizations requires a fundamental understanding of the mathematical structure of the coupled problem.

Computational quantum mechanics is another area in which there are opportunities for new mathematical ideas, particularly for time-dependent problems and other problems with large numbers of degrees of freedom. Just as in the case of classical partial differential equations, the goal is to develop a fundamental understanding of the structure of a solution, which can then be exploited in designing numerical methods. Monte Carlo methods for these and other problems have experienced such developments, including Swendsen-Wang and related algorithms, umbrella sampling and simulated annealing, multigrid Monte Carlo, Green's function Monte Carlo, and high-order accurate quadrature rules for integrals over function spaces. The success of equilibrium statistical mechanics makes us believe that it is possible to remove degrees of freedom in time-dependent problems to obtain accurate representations of coarse-grained dynamics, with applications to fluid turbulence and to many-body problems. Possible approaches include ideas from nonequilibrium statistical mechanics, renormalization group approaches, and stochastic modeling, with techniques from mathematical statistics.

Interdisciplinary projects attacking any of these problems involving combinations of mathematicians, physicists, chemists, computer scientists, and statisticians, could well have unexpected, exciting results.

Computer Science

The above discussion has focused on mathematics, since scientific algorithm and software development often resides there. However, there are equally compelling challenges and needs for collaborations with and contributions from computer science. Computer scientists are typically interested in language design, abstraction management, software development practices, large-scale data analysis, and scientific visualization, all of which are relevant to large-scale scientific software.

Other technical research issues for computer scientists include the performance issues of writing general purpose codes in high-level languages versus domain specific software, and how to design for multiple applications using layered abstractions. Compiler issues, parallel and distributed computing, usability and portability remain important ongoing issues of great importance to computational scientists. Recent examples of this kind of input can be found in PETSC, MPI, and the development of Beowulf clusters. As hardware becomes more complex, it is harder for working scientists to take advantage of it. Fruitful and successful collaborations to improve this situation would be mutually beneficial to both computational physics and computer science.

Very large databases containing terabytes (and soon petabytes) of scientific data pose challenging problems, both for computational physics and computer science. Concomitantly, extracting insight from this data, either via automated techniques or scientific visualization, poses a plethora of unsolved problems and research opportunities at large scale.

There are a number of institutional challenges that NSF may be able to address. For computational science to be successful, applied mathematicians and computer scientists must join with physical scientists as peers in the research enterprise. Software development is difficult and time consuming. However, the long-term health of computational science depends critically on greatly expanding the number of scientists using computing successfully and well. Finally, graduate education must encompass all three parts of the business. How will space be made in the disciplinary curriculum? Who will develop and teach these courses? These institutional issues of integrating mathematicians and computer scientists must be resolved as well.