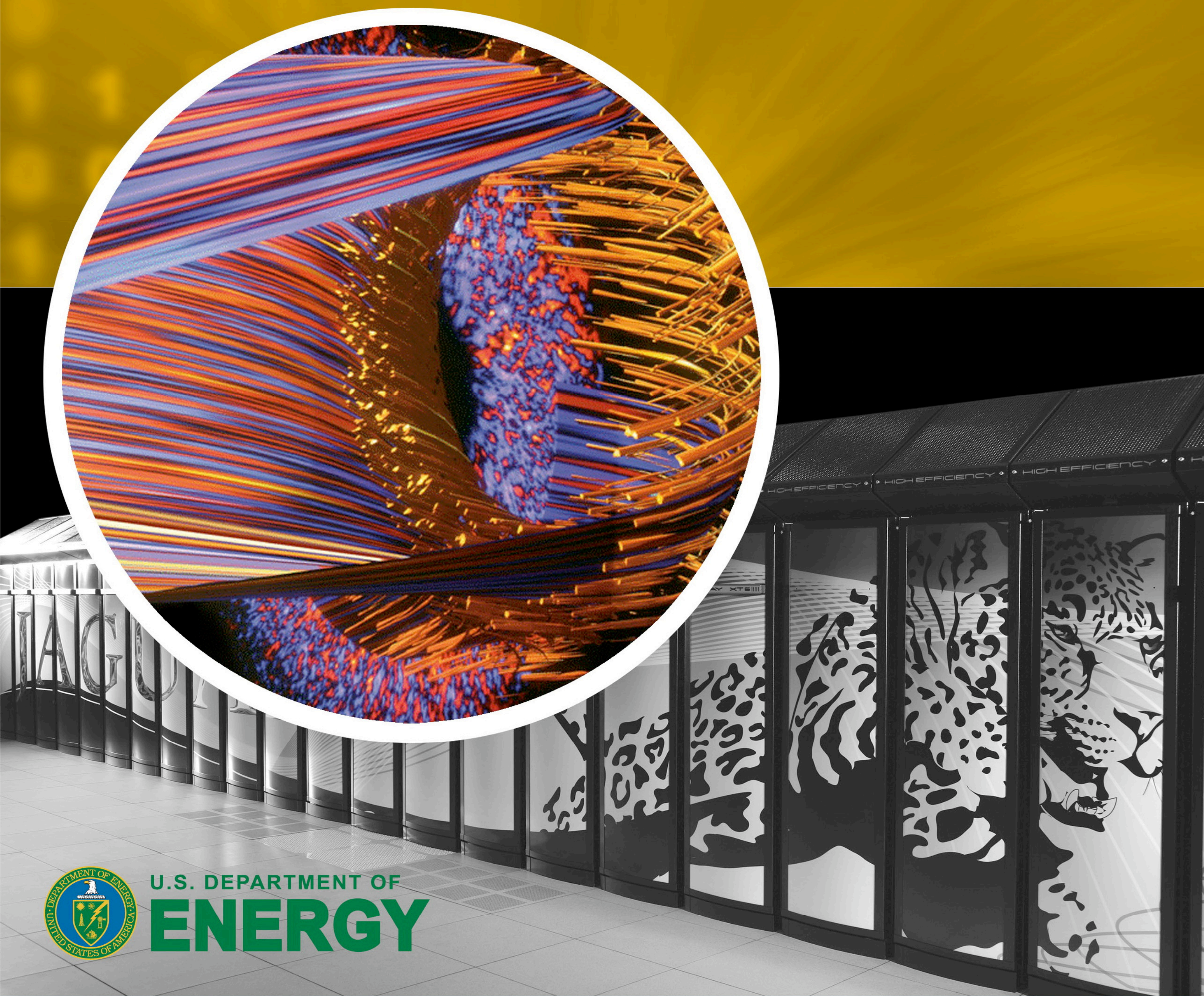
Scientific Grand Challenges

FUSION ENERGY SCIENCES AND THE ROLE OF COMPUTING AT THE EXTREME SCALE

March 18-20, 2009 • Washington, D.C.



U.S. DEPARTMENT OF ENERGY

Sponsored by the Office of Fusion Energy Sciences and the Office of Advanced Scientific Computing Research

DISCLAIMER

On the cover: Oak Ridge National Laboratory's Cray XT5TM supercomputer. The computer, dubbed Jaguar, is the largest in the U.S. Department of Energy's Office of Science. The Jaguar's computing power makes it among the most powerful open scientific systems in the world. Future reports in the Scientific Grand Challenges workshop series will feature different Office of Science computers on their covers.

SCIENTIFIC GRAND CHALLENGES: FUSION ENERGY SCIENCE AND THE ROLE OF COMPUTING AT THE EXTREME SCALE

Report from the Workshop Held March 18-20, 2009

Sponsored by the U.S. Department of Energy, Office of Fusion Energy Sciences and the Office of Advanced Scientific Computing Research

Chair, **William Tang**
Princeton Plasma Physics Laboratory

Co-Chair, **David Keyes**
Columbia University

Lead, Burning Plasma/ITER Science Challenges, **Ned Sauthoff**
Oak Ridge National Laboratory

Co-Lead, Burning Plasma/ITER Science Challenges, **Nikolai Gorelenkov**
Princeton Plasma Physics Laboratory

Co-Lead, Advanced Physics Integration Challenges, **John R. Cary**
University of Colorado and Tech-X Corporation

Co-Lead, Advanced Physics Integration Challenges, **Arnold H. Kritz**
Lehigh University

Co-Lead, Plasma-Material Interaction Science Challenges, **Steven Zinkle**
Oak Ridge National Laboratory

Co-Lead, Plasma-Material Interaction Science Challenges, **Jeffrey N. Brooks**
Purdue University

Co-Lead, Laser-Plasma Interactions and High-Energy Density Laboratory Physics,
Riccardo Betti, University of Rochester

Co-Lead, Laser-Plasma Interactions and High-Energy Density Laboratory Physics,
Warren Mori, University of California, Los Angeles

Co-Lead, Basic Plasma Science/Magnetic Reconnection Physics, **Amitava Bhattacharjee**,
University of New Hampshire

Co-Lead, Basic Plasma Science/Magnetic Reconnection Physics, **William Daughton,**
Los Alamos National Laboratory

Lead, Algorithms for Fusion Energy Sciences at Extreme Scale, **David Keyes***,*
Columbia University

Lead, Data Analysis, Management, and Visualization in Fusion Energy Science,
Arie Shoshani, Lawrence Berkeley National Laboratory

Co-Lead, Mathematical Formulations, **Phil Colella**,
Lawrence Berkeley National Laboratory

Co-Lead, Mathematical Formulations, **Ravi Samtaney**,
Princeton Plasma Physics Laboratory

Lead: Programming Models, Frameworks, and Tools, **Ewing Lusk**,
Argonne National Laboratory

Representative, Office of Fusion Energy Sciences, **John Mandrekas**

Representative, Office of Advanced Scientific Computing Research, **Lali Chatterjee**

EXECUTIVE SUMMARY

Energy and matter, which are the basic components of our physical world, are manifested in various ways under different physical conditions and are affected by various processes. The fusion of light nuclides forms the basis of energy release in the universe, which can potentially be harnessed and used as a clean and sustainable supply of energy. The mission of the U.S. Department of Energy's (DOE) Office of Fusion Energy Sciences (FES) program is to expand the fundamental understanding of matter at very high temperatures and densities, and to build the scientific foundations needed to develop a fusion energy source. A key need is the timely development of a predictive integrated simulation capability for magnetically confined fusion plasmas that are properly validated against experiments in regimes relevant for practical fusion energy production. Additional objectives are to pursue scientific opportunities and grand challenges in plasma science—including high-energy density plasma science—to better understand the universe and to enhance national security and economic competitiveness (Eckstrand 2009).[1]

A technical workshop to discuss forefront questions in fusion energy science and the potential role of high-performance computing at the extreme scale to help resolve associated scientific issues was held March 18-20, 2009, at the Hilton Washington D.C. North/Gaithersburg hotel in Maryland. DOE's Office of FES and the Office of Advanced Scientific Computing Research (ASCR) co-sponsored this collaborative workshop where over 120 participants contributed to the identification of leading scientific problems in fusion energy science with a further focus on those problems that require scientific computing capabilities at the extreme scale to accelerate progress.

Over the next decade, significant advances will be made in computing technology that will allow increases in computing performance by orders of magnitude. These advances will have a major impact on the ability to solve critical scientific and technological problems. The purpose of this technical workshop was to accomplish the following:

1. Identify forefront scientific problems in fusion energy science that could be aided by computing at the extreme scale over the next decade
2. Establish specifics of how and why new high-performance computing capabilities will address issues at the frontiers of fusion energy science
3. Provide fusion energy scientists with the opportunity to influence the development of high-performance computing
4. Provide the FES community with plans for development of future high-performance computing capability in collaboration with ASCR.

The workshop provided a forum for the exchange of ideas from multiple scientific disciplines, including fusion energy scientists, computer scientists, and applied mathematicians. These participants provided the interdisciplinary expertise required to identify and address challenges in fusion energy science and high-performance computing with an emphasis on the use of extreme-scale computing to accelerate progress in fusion energy science research toward significant advances and discoveries.

[1]Eckstrand S. 2009. "Fusion Energy Sciences Program." Presented at the Scientific Grand Challenges in Fusion Energy Sciences and the Role of Computing at the Extreme Scale workshop, March 18-20, Gaithersburg, Maryland. U.S. Department of Energy's Office of Fusion Energy Sciences, Washington, D.C. Available at http://www.ofes.fusion.doe.gov/more_html/FESAC/ProgPrioritiesReport.pdf. Accessed January 4, 2010.

Technical panel discussions focused on five major fusion energy science scientific grand challenges areas where extreme-scale computing can significantly accelerate progress. These included the following:

1. Burning Plasma/ITER Science Challenges
2. Advanced Physics Integration Challenges
3. Plasma-Material Interaction Science Challenges
4. Laser-Plasma Interactions and High-Energy Density Laboratory Physics
5. Basic Plasma Science/Magnetic Reconnection Physics.

This report gives a description of key science issues and opportunities for these five major science areas. The current status is illustrated with examples drawn from active research that have effectively used large-scale computational resources at the national computing centers. The report also provides descriptions of priority research directions (PRDs) and crosscutting challenges with opportunities featured for collaboration with scientists from multidisciplinary fields.

FUSION ENERGY SCIENCE PANEL FINDINGS

Panel 1 - Burning Plasma/ITER Science Challenges

As fusion research enters a new era of burning plasma experiments on the reactor scale, it becomes increasingly urgent to develop experimentally validated predictive capabilities that can produce accurate and robust simulations. This is particularly important for mitigating the risk associated with achieving—in a timely manner—the desired plasma performance in major investments such as the international ITER project. At the highest level, two main concerns in producing the required capabilities involve addressing the larger spatial and longer energy-confinement time scales. Assessments based on fundamental, first-principles physics considerations indicate that scales spanning the small gyro-radius of the ions to the radial dimension of the plasmas will need to be addressed when properly simulating the dynamics in burning plasma. Compared to present-day experiments, an order of magnitude greater spatial resolution is needed to account for the larger plasmas of interest, and the major increase expected in the plasma energy confinement time (~1 second in the ITER device), together with the longer pulse of the discharges in these superconducting systems, will demand simulations of unprecedented aggregate floating point operations.

Panel 2 - Advanced Physics Integration Challenges

Computational modeling is expected to have a major impact on the fusion plasma science program. Because of the high cost of each discharge in burning plasma experiments, planning experimental campaigns and analysis of data demands simulations with unprecedented physics fidelity. Traditionally, computational fusion energy science has addressed separate areas such as macroscopic stability; energetic particles (from auxiliary heating sources including radio-frequency waves and neutral-beam injection and also as products from the fusion reactions); microturbulence and associated “anomalous” transport; and edge plasma physics (where atomic processes are important). Each of these areas has currently demonstrated at varying levels of efficiency the capability of productively using existing leadership class facilities. With extreme-scale computational power, it will be possible to couple improved versions of these large-scale simulations to produce an experimentally validated *integrated* simulation capability for scenario modeling of the whole device.

Panel 3 - Plasma-Material Interaction Science Challenges

Plasma and material interactions are one of the most critical scientific issues for fusion power, affecting the following: 1) lifetime of plasma-facing components due to sputter and transient erosion; 2) plasma contamination by eroded material; 3) tritium co-deposition in eroded/re-deposited material; and 4) operating limits on core plasma (beta, confinement, edge temperature/density, duty factor, etc.) as a result of the above factors. A related critical topic is bulk material performance and optimization. Gaining understanding and predictive capabilities in this vitally important area requires addressing simultaneously complex and diverse physics occurring over a wide range of lengths (angstroms to meters) and times (femtoseconds to days). This will require further development of not only detailed physics models and computational strategies at each of these scales but also algorithms and methods to strongly couple them in a way that can be robustly validated. While present research confined to each of these scales, or pioneering approaches to couple two or more of them, already push the state of the art in technique and available computational power, simulations spanning multiple scales needed for major future fusion energy projects (e.g., ITER and DEMO [Demonstration Reactor]) will require extreme-scale computing platforms and integrated physics, and computer science advances.

Panel 4 – Laser-Plasma Interactions and High-Energy Density Laboratory Physics

Recent technological advances in lasers, particle beams, and Z-pinches have made it possible to generate plasmas with unprecedented energy densities in the laboratory. Understanding the properties and behavior of such plasmas constitutes the science area that is called high-energy density laboratory plasmas (HEDLP). This rapidly emerging science area is extremely rich in basic science phenomena as well as potential applications such as inertial fusion energy science (IFES). IFES is one possible approach towards producing a clean and sustainable supply of energy. A recent DOE Office of Science and National Nuclear Security Administration panel (*Advancing the Science of High-Energy Density Laboratory Plasmas*[1]) produced prioritized lists of compelling science opportunities in basic HEDLP, issues for IFES, and related opportunities for advanced computing to make a major impact. Many of these opportunities include processes that can demand fully kinetic models involving multiscale science issues spanning micro- to meso-time and space scales. For example, in some IFES experiments, a millimeter-scale pellet of deuterium and tritium is compressed to 1000 times solid density over nanosecond-time scales and lasers with wavelengths of microns or smaller propagate through centimeter-scale plasmas.

Panel 5 - Basic Plasma Science/Magnetic Reconnection Physics

The liberation of magnetic field energy through the process of magnetic reconnection is at the core of a diverse range of plasma phenomena including solar flares, geomagnetic substorms, sawteeth oscillations and disruptions in tokamaks, extragalactic jets, and a wide variety of astrophysical settings. In the past decade, most of the theoretical and simulation efforts have been directed at relatively small two-dimensional systems using both fluid and kinetic descriptions. Presently, it remains unclear how these

[1] *Advancing the Science of High Energy Density Laboratory Plasmas*. 2009. Prepared by the Fusion Energy Science Advisory Committee for the U.S. Department of Energy and National Nuclear Security Administration, Washington, D.C. Available at http://www.sc.doe.gov/ofes/FESAC-HEDLP-REPORT%20(2).pdf. Accessed March 26, 2009.

idealized results will extend to large-scale three-dimensional systems. Even with extreme-scale computing, a first-principles, three-dimensional kinetic treatment of reconnection in hydrogen plasmas will be limited to fairly small systems. Progress in modeling realistic applications will require understanding the key physics sufficiently well to be able to capture them within reduced descriptions and to infer reliable scaling.

ADVANCED SCIENTIFIC COMPUTING RESEARCH CROSSCUTTING CHALLENGES

ASCR crosscutting challenges, which impact the five fusion energy science grand challenge areas, were identified by members of the four ASCR panels during the workshop. These findings are summarized in the following paragraphs.

ASCR Panel 1 - Algorithms for Fusion Energy Sciences at Extreme Scale

There are many motivations for scaling simulations in fusion energy science to the expanding architectural extremes of the coming decade. These include increasing efforts to resolve the full ranges of length and/or time scales in a model; accommodating physical effects with greater fidelity; allowing the model degrees of freedom in all relevant dimensions; optimizing or controlling plasma scenarios (inverse problem) that are adequately predicted by forward models; and quantifying uncertainty. However, as applications broaden to take full advantage of extreme architectures, the complexity of algorithms may grow super linearly in problem size, making it impossible to weak scale, even though memory capacity would seem to allow it. Extreme scales put a premium on finding "optimal" algorithms, whose complexity is at worst log-linear in problem size, because (by Amdahl's Law) any suboptimal component will ultimately dominate the execution profile. The availability of high-capability architectures makes algorithms more—not less—important. Fortunately, algorithms such as linear solvers have kept pace with extreme scales to date, and optimal versions are known for systems arising from some popular formulations of the plasma physics.

ASCR Panel 2 - Data Analysis, Management, and Visualization in Fusion Energy Science

Fusion energy scientists considering the computational needs and science questions that can be answered with extreme-scale computational power should also consider the implications of data requirements at the extreme scale. Managing fusion simulation data already has proven to be a problem in terms of volume, bandwidth, and complexity. Some codes (e.g., VPIC, OSIRIS, M3D-K) will model 1 billion cells and 1 trillion particles. Based on mean-time-between-failure concerns when running on a million cores, these codes will need to output 2 gigabytes/second per core or 2 petabytes/second of checkpoint data every 10 minutes. This amounts to an unprecedented input/output rate of 3.5 terabytes/second. The data questions to consider at the extreme scale fall into two main categories: data generated and collected during the production phase, and data that need to be accessed during the analysis phase.

ASCR Panel 3 - Mathematical Formulations

Panel members focused on identifying scientific challenges in the five topical areas in plasma physics separately studied in the FES panels for which there are "bottlenecks" that may be ameliorated with the development of new models and discretization methods. Many formulations for plasma physics exist, including partial differential equation (PDE)-based and particle-based models for kinetic approaches and PDE-based models for moment closures, and many discretizations that are customized to asymptotic physical regimes in different devices or different subdomains of the same device. Members of the physics panels that participated skewed heavily in the direction of kinetic modeling in plasma physics; therefore, the results of this session should not be taken as fully comprehensive. However, clearly there are central problems in each of these areas for which new models and discretizations could play an important role.

ASCR Panel 4 - Programming Models, Frameworks, and Tools

The coming transition in computer architectures as peak capability approaches the exascale offers challenges along with obvious opportunities for fusion energy science. Challenges include a paradigm shift in programming methodologies. Existing technologies for writing parallel scientific applications have sustained high-performance computing application software development for the past decade, and have been successful for petascale computing. However, these technologies were designed for coarse-grained concurrency largely dominated by bulk-synchronous algorithms. Future hardware constraints and growth in explicit on-chip parallelism will likely require a mass migration to new algorithms and software architecture that is as broad and disruptive as the migration from vector to parallel computing systems that occurred 15 years ago. Applications and algorithms will need to rely increasingly on fine-grained parallelism, strong scaling, and support fault resilience. Addressing these challenges creates a renewed opportunity to introduce a higher level of software engineering into current fusion application subsystems that will enhance the modularity, portability, and performance of codes while extending their capabilities to new levels. At the same time, past sound investments must be protected, and a migration path from current to future environments must be elaborated.

PRIORITY RESEARCH DIRECTIONS

PRDs that the five FES and four ASCR panel members identified in their workshop discussions are summarized in the following paragraphs.

FES Panel 1 - Burning Plasma/ITER Science Challenges

This panel identified five PRDs for which significant advances in understanding are needed to achieve targeted levels of controlled magnetic fusion power. These include the following topics:

Development of a new generation of magnetohydrodynamic (MHD) codes capable of accurately modeling the onset of plasma disruptions and their effects on the device components. The driving goal is to develop an improved macroscopic-simulation capability for ITER-class experiments. This is a critical goal because nonlinear macroscopic events play a central role in defining the operational space of these devices, and many details of the nonlinear processes and interactions are poorly understood.

Greater understanding of plasma transport and turbulence. This is a key physics requirement for enabling achievement of the required energy confinement time in fusion plasmas. A critically important challenge is associated with the recognition that realistic transport simulations for burning plasmas demand the development of a) electromagnetic simulation capabilities; and b) the ability to address the coupling of global, nonlocal transport on an equal-footing with MHD phenomena.

Realistic capability for simulating the physics of the edge barrier region in high-performance burning plasmas. Understanding the dynamics in this region, which are characterized by strong pressure gradients, is critical for optimizing performance in burning plasmas. The goal is to be able to conduct a comprehensive analysis across a wide range of overlapping spatio-temporal scales that include both the relevant small-scale kinetic/gyrokinetic dynamics and the large-scale MHD physics.

Experimentally validated predictive simulations of energetic particle dynamics in burning plasmas. This involves the development of realistic, self-consistent modeling capabilities for fusion alpha particle profiles in the presence of multiple Alfvénic and MHD instabilities.

Radio frequency wave heating and current drive for burning plasma scenarios. This involves the development of reliable simulations for the larger configuration dimensions of systems (such as the ITER project) of the following: a) wave propagation and coupling efficiency in the high-temperature pedestal region; and b) radio frequency interactions with fusion alpha particles.

Achieving significant progress in a timely manner for all of these grand challenge areas will require development of advanced simulation capabilities using computing at the extreme scale.

FES Panel 2 - Advanced Physics Integration Challenges

This panel identified five PRDs in which computing at the extreme scale would make a significant impact. These include the following topics:

Transport modeling with embedded turbulence. Computation offers the highest-fidelity path to the calculation of plasma profiles. Approaches include the following: a) integration of well-parallelized local computations of turbulent fluxes within a code that advances plasma profiles in response to sources of heat, momentum, current and particles; and b) coupling of global turbulence with transport over the same region—probably a necessary approach for dealing with the plasma edge. Challenges include verification and validation (with associated uncertainty quantification), formulating new mathematical algorithms, and addressing the lack of data alignment between the calculation of sources and transport.

Coupling disparate regions of the plasma. This capability is needed for a whole-device model that includes core, edge, and plasma-facing materials. Associated research areas of focus include the following: a) developing reduced models for edge dynamics that are closer to first-principles calculations; and b) addressing the coupling of sources in both the plasma edge and core.

Macroscopic stability control using radio frequency power. This is a well-known capability important for fusion devices. A classic example is the use of electron cyclotron waves to drive plasma currents that suppress key instabilities (such as neoclassical tearing modes). Associated focused research topics include reformulation and new code implementation when the non-inductively driven current is an integral part of the MHD equilibrium and stability evolution.

Recoverable non-axisymmetric macroscopic dynamics. These processes include periodic instabilities, such as internal sawtooth reconnections in the central part of the plasma and edge localized modes. Transport leads to thermal and particle profiles that are unstable. These instabilities then transiently alter the plasma profiles. Focused associated research needs here include development of periodic temporal coupling of computations involving brief intervals of rapid macroscopic dynamics and longer intervals of axisymmetric transport. Such couplings also have application to the key area of disruption mitigation, which involves ideal and resistive MHD, runaway electron dynamics and transport, pellet and gas fueling, and plasma-wall interactions.

Performance optimization of burning plasmas. This brings all of the preceding four PRDs together, but with even greater computational requirements to run with different parameter sets to optimize plasma profiles over control parameters, such as external energy and current drive sources.

Progress in these five areas will require teams of fusion energy scientists, applied mathematicians, and computer scientists to address problems across the range of physics, algorithms, data management, dynamic load balancing, and code modernization.

FES Panel 3 – Plasma-Material Interactions Science Challenges

This panel identified three PRDs with the common goal to develop comprehensive computational models for predictive, self-consistent, integrated, validated, full-process, time-dependent, plasma/material interactions. All three of these areas are expected to benefit significantly from the impetus provided by extreme-scale computing.

Modeling of the edge and scrape-off layer plasmas. This includes modeling of turbulent transport and full coupling of plasma ions and electrons, neutrals, photons, and electromagnetic fields. In addition, plasma contamination from near-surface transport of sputtered or vaporized material and quantification of plasma facing component particle and photon fluxes (with predictions of instability regimes) should be considered.

Predicting the near-surface material response to the extreme plasma fluxes of photons and particles under normal and transient operation. This includes predicting sputtering erosion/re-deposition and other time-integrated plasma facing component processes (e.g., dust formation and transport; helium- or deuterium-tritium-induced microstructure formation and flaking) and the resultant impurity transport, core plasma contamination, mixed-material formation, and tritium co-deposition in redeposited materials. The material and edge plasma response to transient processes such as high-powered edge localized modes vertical displacement events, plasma disruptions, and runaway electrons represent an important component of this effort.

Modeling the underlying structural materials response. This involves understanding the fundamental microstructure evolution and performance limits of structural materials in the fusion radiation environment that involve extreme cyclic thermo-mechanical stresses and simultaneous intense fusion neutron bombardment.

An overarching grand challenge will involve efficient integration of these three coupled PRDs to develop a comprehensive model. The associated collective impact on FES includes enabling a) effective operation of the ITER and proper design of DEMO; b) improved understanding of present experiments; and c) a plasma-material interaction code package for the macro-type code packages needed by the proposed Fusion Simulation Program.

FES Panel 4 - Laser-Plasma Interactions and High-Energy Density Laboratory Physics

Four PRDs were identified for HEDLP and IFES for which extreme-scale computing could make a transformative impact.

Nonlinear optics of plasmas. The goal is to understand how an ensemble of overlapping Gaussian beamlets (speckles) mutually interact in HEDLP. This understanding is critical to successful development of inertial fusion energy (IFE) concepts using laser drivers. It requires fully kinetic modeling because subtle changes to the electron distribution function can lead to substantial differences. On extreme-scale computers, the goal of simulating an ensemble of speckles using fully kinetic modeling could be achieved. This could—in turn—lead to ideas on how to tame these interactions and the development of high-fidelity reduced models for mesoscale simulations.

Relativistic high-energy density plasma and intense beam physics. The goal is to understand how lasers at the intensity and power frontier interact with and are absorbed in HEDLP. Because the associated physics requires detailed understanding of single-particle trajectories and how the complex patterns of large currents of relativistic particles form in plasmas and collectively interact, fully kinetic and relativistic modeling are required. On extreme-scale computers, fully kinetic simulations using true time and length scales of fast ignition targets could be possible for the first time. This will also require development of coupled microscale and mesoscale models.

Integrated fast ignition simulations. The goal is to provide full integrated modeling of high-gain, fast ignition IFE concepts where the timing of the intense ignition pulse, the compression of the pellet, and survival of an inserted cone tip can be important. On extreme-scale computers, the coupling of fully kinetic simulations of HEDLP with parameters obtained from macroscale hydrodynamic compression models may be possible, thereby enabling simulations representing the true time and space scales.

Magnetized high-energy density plasmas. The goal is to understand how spontaneous or induced magnetic fields can affect burning HEDLP. The physics spans a wide parameter space, from the dense compressed core of a traditional IFE target, as well as the more tenuous plasmas in reversed field configurations. Extreme-scale computers will enable high-fidelity simulations of dense collisional plasmas that are inertially confined and in which heat flux is limited by magnetic fields. The development of mesoscale models, coupled with extreme computing, should enable breakthroughs in the understanding of magnetized plasmas under compression.

FES Panel 5 - Basic Plasma Science/Magnetic Reconnection Physics

Looking to the future, significant progress on four PRDs in this basic plasma science grand challenge area were identified for which computing at the extreme scale could enable higher physics-fidelity simulations of magnetic reconnection physics for most applications of interest.

Influence of the electron and ion kinetic scales on the large-scale evolution. Currently, there are significant differences between fully kinetic and two-fluid simulations in weakly collisional regimes. Thus, there is no clear consensus on the minimal physics required to accurately capture the large-scale evolution. First-principles kinetic simulations, including Coulomb collisions, can provide a guidepost for developing reduced fluid descriptions that better capture the structure and dynamics. Other approaches may include reduced kinetic descriptions such as the following: a) the gyrokinetic model, and b) the hybrid model that embeds a kinetic description within a larger fluid simulation.

Reconnection and magnetic island dynamics in three-dimensional geometries. Evidence exists that a single reconnection layer may divide into multiple reconnection sites due to the formation of secondary magnetic islands or other secondary instabilities (such as ballooning modes) that may control the relaxation of current and pressure profiles in tokamaks. Evolution of reconnection dynamics on both fast- and long-transport time scales, including kinetic effects, is of great interest for fusion as well as space and astrophysical applications. Addressing these issues will require highly scalable fluid and kinetic algorithms, along with a realistic treatment of boundary conditions.

Energy partition and particle acceleration that results from reconnection. Thermal energy gained by ions and electron, as well as the formation of nonthermal tails, is of significant theoretical and observational interest. For the highly energetic tails, it is difficult to explain the observations with a single steady-state reconnection site. One critical question is whether most nonthermal particles are directly associated with reconnection sites and magnetic islands, or with other processes associated with the global relaxation (such as waves and shocks).

Reconnection in relativistic plasmas. In many astrophysical applications (pulsars, accretion near black holes, gamma-ray bursts), reconnection is thought to occur in highly relativistic regimes with both hydrogen and electron-positron plasmas. These regimes are well suited for relativistic kinetic simulations that are now feasible in three-dimensions at the petascale for electron-positron plasmas. These advancements in reconnection physics have the potential to impact fusion energy science through the following: a) more realistic modeling of tearing modes and sawteeth oscillations in tokamaks; b) understanding magnetic relaxation in reversed field pinches, stellarators, and field-reversed configurations; and c) higher physics-fidelity modeling of relativistic electrons for fast ignition.

ASCR Panel 1 – Algorithms for Fusion Energy Sciences at Extreme Scale

Six PRDs emerged for scalable algorithms that are relevant to accelerating progress in fusion energy science simulations.

Optimal representations. Full adaptivity in the sense of *h* (mesh refinement), *p* (discretization order), and *r* (mesh relocation) should be employed in space and time, according to the local smoothness of fields to be represented, to get the most "science per watt" out of a fusion energy science modeling simulation. This requires estimating and equi-distributing truncation errors, dynamic in-place load balancing, and managing and converting between different representations.

Multiphysics and multiscale algorithms. Algorithms that allow self-consistent coupling of multiphysics models across all relevant scales allow better focus on physical questions, free of concern about numerical instabilities and splitting errors, and longer windows of integration due to suppression of stability-limiting fast scales with greater accuracy. This requires scalable implicit methods and high-order interpolations between representations (e.g., from fields to particles and vice versa).

Real-time algorithms. Armed with first-principles models, reduced-order models can be parameterized for sufficiently narrow regimes to provide detection and control capabilities in real time.

This requires physics-based developments beyond current models based on principal component analysis or proper orthogonal decomposition.

Optimization. Robust (error-tolerant) optimization algorithms are needed for high-dimensional multiphysics models for optimal design, control, parameter estimation and the mapping of stability boundaries. Required are deterministic and stochastic techniques for derivative-free methods, adjoint-based derivative methods, and preconditioners for saddle-point systems.

Uncertainty quantification (UQ) and reduction. Models contain uncertainties in initial conditions, boundary conditions, coefficients, and/or forcings, coming from observations or other simulations. Incorporation of observations can improve uncertain models, balancing models, and numerical errors for more efficient computation. Needs include deterministic UQ tools based on sensitivity and adjoint techniques, probabilistic approaches based on sampling methods, and direct propagation of probability density functions from inputs to outputs.

Lower threshold of expertise required to use optimal algorithms on extreme architectures. Software for extreme-scale environments must offer multilevel ("incremental adoption") user interfaces. With proper interfaces to widely used (and therefore thoroughly debugged) modules, software will perform as closely as possible to expert reliability while auto-tuning or being tunable for high performance by expert users. With such tools, fusion energy physicists will work more productively and better understand the performance of their software tools, thus focusing more on physics and less on software issues.

ASCR Panel 2 - Data Analysis, Management, and Visualization in Fusion Energy Science

Five PRDs will require extensive research and development effort to support the data requirements at the extreme scale for fusion energy science.

Managing large-scale input/output volume and data movement. Techniques need to be developed that optimize input/output performance automatically based on hardware characteristics. Such techniques are crucial to avoid slowdown of computations because of insufficient input/output rates. Furthermore, future FES codes should be as independent as possible of input/output tuning, where all such details are processed automatically by the underlying input/output system. Parallel file systems and data movement tools need to be scaled to support these extreme volumes of data.

Real-time monitoring of simulations and run-time metadata generation. Having run-time monitoring capability on all supercomputing resources is essential to avoid computational waste. This capability will prevent runs that do not converge or progress correctly from continuing. Workflow technology already used for such purposes in fusion energy science applications need to be scaled and become part of the simulation system that supports summarization of results in real time, and/or permit the monitoring software to automatically manage simulations that do not progress correctly. Additionally, provenance and metadata information needs to be automatically collected (also at run time) for effective run-time and post-run data analysis.

Data analysis at extreme scale. The data analysis challenges in fusion energy science applications at the extreme scale stem not only from the large size of the data, but also from data complexity. First, areas of interest— such as coherent structures and fronts—are likely to be spread across many processors, making it difficult to extract poorly defined structures or track fronts over time. Second, techniques to process these data to reduce overall size before these data are output by the simulation require algorithms that are robust enough to process data correctly.

Visualization of very large datasets. Visualization is often a key technology for understanding data such as electron-temperature profiles. However, reducing and mapping terabytes or petabytes of data into meaningful visualization is a challenge that will require processing near to where the data are stored, as well as effective indexing techniques for real-time data exploration.

Experiment-simulation data comparison. Such tools are essential for validation of FES simulations and diagnostics, and for comparing shot data to reduced models for ITER runs. Experimental data are expected to grow to terabyte sizes, and therefore robust synthetic diagnostic tools need to be developed that are cross-platform scalable and based on forthcoming community standard data formats.

ASCR Panel 3 - Mathematical Formulations

Inheriting structure from the topics of the five FES panels, panel members identified one or more PRDs in each.

Burning Plasma/ITER Science Challenges. The main priority is the need for high-fidelity kinetics calculations, both in the core and in the edge region. Additional priorities also include more accurate gyrokinetic approximations, systematic methods for constructing nearly field-aligned coordinates, fundamental new numerical algorithms for particle-in-cell, the need (or lack thereof) for symplectic integrators for both particle-based and continuum-based methods, and treatments of kinetic electrons.

Advanced Physics Integration Challenges. There is a need for a mathematically systematic treatment of coupled systems with vastly different spatial and/or temporal scales, including well-posedness, stability, and accuracy. A classic example is the coupled treatment of turbulence and transport.

Plasma-Material Interaction Science Challenges. The main priority is the design of materials to withstand tokamak operating conditions, a topic outside the scope of numerical plasma physics. A second priority is interaction of the plasma environment with material boundaries. In the latter area, topics include the improvement of the fidelity of edge models with respect to the interaction with the boundary; the effects of impurities on the overall plasma; and the impact of liquid walls.

High-Energy Density Laboratory Plasma/Laser-Plasma Interactions. This priority includes understanding the interaction of the laser with plasma heterogeneities, known as speckles. Mathematically, this is a homogenization problem: scientists want to understand and represent the collective effect of thousands of speckles, while currently it is only possible to compute the interaction of the laser with one such speckle. This leads to the development of reduced/meso-scale models derived from large-scale Hydrologic Engineering Center calculations.

Basic Plasma Science/Magnetic Reconnection Physics. This is primarily a multiscale problem, exhibiting kinetic behavior in highly localized regions in space, combined with fluid behavior on larger scales. The traditional approach of using two-fluid extended MHD is questionable physically (particularly for larger scale problems), and difficult numerically while the kinetic models that are correct in reconnection zones are too expensive to use globally. This is an opportunity to introduce hybrid fluid-kinetic models that have been used successfully in other areas of fluid dynamics.

ASCR Panel 4 - Programming Models, Frameworks and Tools

To tackle challenges in programming models for FES, this panel identified six PRDs.

Find efficient algorithms and implementations that exploit new multicore, heterogeneous, massively parallel architectures. This research is directed primarily at languages, libraries, and runtime systems that allow fusion energy science programmers to use massive on-chip concurrency in a portable, cross-architecture manner while cooperating with interprocessor parallelism.

Find new, productive approaches to writing, integrating, validating, and tuning complex fusion energy science application programs. This involves development of programming models and systems for massive numbers of processors.

Develop tools for understanding complex application program behavior at scale and for optimizing application performance. This requires the evolution of existing tools and development of new ones to address heterogeneous processors and greater integration of model-based approaches in fusion energy science.

Ensure a migration path from current fusion energy science programming approaches to new ones. Existing Fortran + message-passing interface codes will continue to be used and extended as architectures scale up. Research into message passing interface interoperability and extreme scalability will be required, together with a new software development ecosystem that spans all scales of systems, from midrange to the exascale, to facilitate a viable migration path from development to large-scale production computing systems.

Define common framework tools or components that can be reused in multiple fusion energy science application domains. Frameworks that organize existing and future fusion energy science codes into coherent tools for scientific investigations are currently in an ad-hoc stage of development; research into general abstractions and tools for constructing components and frameworks are needed.

Establish methods and systems that enable pervasive fault resilience. At the exascale, faults of various kinds in both hardware and software components are expected to become commonplace in the execution environment. Fault recovery mechanisms will need to be integrated at every level of the system design—in hardware, software, and the programming model for fusion energy science applications.

CONTEXT

This executive summary highlights the main findings of the FES panels and the ASCR crosscutting panels. This summary also presents the FES and ASCR PRDs identified by the FES Grand Challenges workshop participants.[1] The main body of the report presents details that elaborate on the exciting and formidable nature of these challenges, as well as in-depth discussions of the new collaborative research opportunities for the FES and ASCR scientific communities.

---

[1] "Scientific Grand Challenges in Fusion Energy Sciences and the Role of Computing at the Extreme Scale," March 18-20, Washington, D.C. Workshop sponsored by DOE's Office of Advanced Scientific Computing Research and the Office of Fusion Energy Sciences.

TABLE OF CONTENTS

EXECUTIVE SUMMARY v
TABLE OF CONTENTS xvii
INTRODUCTION 1
PANEL REPORTS 3
BURNING PLASMA/ITER SCIENCE CHALLENGES 5
Current Status 5
Basic Science Challenges and Research Needs 6
Priority Research Directions 7
Conclusions 24
ADVANCED PHYSICS INTEGRATION CHALLENGES 25
Current Status 25
Basic Science Challenges and Research Needs 26
Priority Research Directions 27
Conclusions 41
PLASMA-MATERIAL INTERACTION SCIENCE CHALLENGES 42
Current Status 42
Basic Science Challenges and Research Needs 43
Conclusions 53
LASER-PLASMA INTERACTIONS AND HIGH-ENERGY DENSITY LABORATORY PHYSICS 55
Current Status 55
Basic Science Challenges and Research Needs 56
Priority Research Directions 58
Crosscutting Research Directions 71
BASIC PLASMA SCIENCE/MAGNETIC RECONNECTION PHYSICS 77
Current Status 77
Basic Science Challenges and Research Needs 77
Priority Research Directions 81
Conclusions 89
CROSSCUTTING CHALLENGES 91
ALGORITHMS FOR FUSION ENERGY SCIENCES AT EXTREME SCALE 93
Summary 93
Scientific Challenges and Research Approaches 97
Conclusions 107
DATA ANALYSIS, MANAGEMENT, AND VISUALIZATION IN FUSION ENERGY SCIENCE 109
Summary 109

Scientific Challenges and Research Approaches 110
Conclusions 133
MATHEMATICAL FORMULATIONS 137
Summary 137
Scientific Challenges and Research Approaches 138
Conclusions and Summary 147
PROGRAMMING MODELS, FRAMEWORKS, AND TOOLS 149
Summary 149
Scientific Challenges and Research Directions 150
Conclusions 163
PRIORITY RESEARCH DIRECTIONS 165
BURNING PLASMA/ITER SCIENCE CHALLENGES 167
ADVANCED PHYSICS INTEGRATION CHALLENGES 169
PLASMA-MATERIAL INTERACTION SCIENCE CHALLENGES 171
LASER PLASMA INTERACTIONS AND HIGH-ENERGY DENSITY LABORATORY PHYSICS 173
BASIC PLASMA SCIENCE/MAGNETIC RECONNECTION PHYSICS 175
ALGORITHMS FOR FUSION ENERGY SCIENCES AT EXTREME SCALE 177
DATA ANALYSIS, MANAGEMENT, AND VISUALIZATION IN FUSION ENERGY SCIENCE 179
MATHEMATICAL FORMULATIONS 181
PROGRAMMING MODELS, FRAMEWORKS, AND TOOLS 183
CONCLUSIONS AND RECOMMENDATIONS 185
Fusion Energy Science Panels 185
Advanced Scientific Computing Research Crosscutting Challenges 188
REFERENCES 193
APPENDIX 1: WORKSHOP AGENDA .... APPENDIX 1-1
APPENDIX 2: WORKSHOP PARTICIPANTS .... APPENDIX 2-1
APPENDIX 3: ACRONYMS AND ABBREVIATIONS .... APPENDIX 3-1

INTRODUCTION

Energy and matter are basic components of our physical world. These components are manifested in various ways under different physical conditions and are affected by various processes. The fusion of light nuclides forms the basis of energy release in the universe, which can potentially be harnessed and used as a clean and sustainable supply of energy.

The mission of the U.S. Department of Energy's (DOE) Office of Fusion Energy Science Program is to develop the predictive scientific understanding needed to create a sustainable fusion energy source. Additional objectives are to pursue scientific opportunities and grand challenges in plasma science—including high-energy density plasma science—to better understand the universe and to enhance national security and economic competitiveness (Eckstrand 2009).

A technical workshop to discuss forefront questions in fusion energy science and the role of high-performance computing was held March 18-20, 2009, at the Hilton Washington D.C. North/Gaithersburg hotel in Maryland. DOE's Office of Fusion Energy Sciences and the Office of Advanced Scientific Computing Research co-sponsored this collaborative workshop to identify leading scientific problems in fusion energy science with a further focus on those problems that require scientific computing capabilities at the extreme scale.

This workshop report is one of a series resulting from the Scientific Grand Challenges Workshops hosted by the DOE Office of Advanced Scientific Computing Research in partnership with other DOE Office of Science programs. The workshop series focuses on the grand challenges of specific scientific domains and the role of extreme-scale computing in addressing those challenges. Dr. Paul Messina, interim director of science at the Argonne Leadership Computing Facility, is overseeing the workshop series.

Over the next decade, significant advances will be made in computing technology that will allow increases in computing performance by orders of magnitude. These advances will have a major impact on the ability to solve critical scientific and technological problems. The purpose of this technical workshop was to accomplish the following:

1. Identify forefront scientific problems in fusion energy science that could be aided by computing at the extreme scale over the next decade.
2. Establish specifics of how and why new high-performance computing capabilities will address issues at the frontiers of fusion energy science.
3. Provide fusion energy scientists with the opportunity to influence the development of high-performance computing.
4. Provide the fusion energy science community with plans for development of future high-performance computing capability in collaboration with the DOE Office of Advanced Scientific Computing Research.

The workshop provided a forum for the exchange of ideas from multiple scientific disciplines, including fusion energy scientists, computer scientists, and applied mathematicians. These participants provided the interdisciplinary expertise required to identify and address challenges in fusion energy science and high-performance computing with an emphasis on the use of extreme-scale computing to accelerate progress in fusion energy science research toward significant advances and discoveries.

Technical panel discussions focused on five major fusion energy science scientific grand challenges areas where extreme-scale computing can significantly accelerate progress. These included the following:

- Burning Plasma/ITER Science Challenges
- Advanced Physics Integration Challenges
- Plasma-Material Interaction Science Challenges
- High-Energy Density Laboratory Plasma/Laser-Plasma Interactions Challenges
- Basic Plasma Science/Magnetic Reconnection Physics.

This report gives a description of key science issues and opportunities for each of these areas. The current status is illustrated with examples from active research projects that have effectively used large-scale computational resources at the national computing centers.

Workshop participants also provided the multidisciplinary expertise required to identify and address crosscutting challenges in high-performance computing with an emphasis on the use of extreme-scale computing for scientific research to enable advances and discoveries. Advanced scientific computing research crosscutting challenges, which impact the five major fusion energy science areas, are also discussed in this report:

- Algorithms for Fusion Energy Sciences at Extreme Scale
- Data Analysis, Management, and Visualization in Fusion Energy Science
- Mathematical Formulations
- Programming Models, Frameworks, and Tools.

The report also provides descriptions of priority research directions that identify specific opportunities for collaboration with scientists from multidisciplinary fields. The report concludes with a conclusions and recommendations section, followed by the references cited in this report. The appendices provide a list of workshop participants, the workshop agenda, and acronyms used within the report.

PANEL REPORTS

- **BURNING PLASMA/ITER SCIENCE CHALLENGES**
- **ADVANCED PHYSICS INTEGRATION CHALLENGES**
- **PLASMA-MATERIAL INTERACTION SCIENCE CHALLENGES**
- **LASER-PLASMA INTERACTIONS AND HIGH-ENERGY DENSITY LABORATORY PHYSICS**
- **BASIC PLASMA SCIENCE/MAGNETIC RECONNECTION PHYSICS**

BURNING PLASMA/ITER SCIENCE CHALLENGES

Lead: Ned Sauthoff, Oak Ridge National Laboratory
Co-Lead: Nikolai Gorelenkov, Princeton Plasma Physics Laboratory

Panel Members: Herbert Berk, University of Texas at Austin; Yasuhiro Idomura, Japan Atomic Energy Agency; Steve Jardin, Princeton Plasma Physics Laboratory; Guo-Yong Fu, Princeton Plasma Physics Laboratory; Raffi Nazikian, Princeton Plasma Physics Laboratory; William Nevins, Lawrence Livermore National Laboratory; Scott Parker, University of Colorado; Sergei Putvinski, ITER; Hong Qin, Princeton Plasma Physics Laboratory; Ned Sauthoff, Oak Ridge National Laboratory; Linda Sugiyama, Massachusetts Institute of Technology; Philip Snyder, General Atomics; Bruce Scott, Max-Planck-Institut Für Plasmaphysik; Xianzhu Tang, Los Alamos National Laboratory; Weixing Wang, Princeton Plasma Physics Laboratory; John Wright, Massachusetts Institute of Technology

CURRENT STATUS

The Burning Plasma/ ITER panel that collaborated at the March 2009 workshop titled, “Scientific Grand Challenges in Fusion Energy Sciences and the Role of Computing at the Extreme Scale,” has identified key research areas where significant advances are needed to achieve controlled magnetic fusion power.[1] Surmounting these physics and computational challenges will lead to new information from burning plasma experiments such as the ITER. This will involve developing modeling capabilities for dealing with multiscale physics in thermonuclear plasmas—especially those extreme-scale computational capabilities that will help accelerate progress toward resolving major burning plasma/ITER science challenges.

Panel members examined the needs for advanced multiscale physics theory and algorithm development in research involving burning plasmas and focused discussions on the physics and computational challenges associated with the following five topics, each of which was identified as a priority research direction (PRD):

- Plasma transport in large-scale burning plasmas on a confinement time scale—one of the most difficult and least understood problems in fusion energy research
- Magnetohydrodynamic (MHD) events and three-dimensional resonant and nonresonant field effects relevant to the onset of plasma disruptions, such as sawteeth, edge-localized modes (ELMS), tearing modes, vertical displacement events, and the generation of the runaway electrons during the disruptions
- Energetic particle dynamics in burning plasmas, including predictions for self-consistent fusion alpha particle profiles in the presence of multiple Alfvénic and MHD instabilities
- Radio-frequency wave heating and current drive for burning plasma scenarios
- Physics of the edge barrier region in high-performance burning plasmas.

[1] “Scientific Grand Challenges in Fusion Energy Sciences and the Role of Computing at the Extreme Scale,” March 18-20, Washington, D.C. Workshop sponsored by DOE’s Office of Advanced Scientific Computing Research and the Office of Fusion Energy Sciences.

BASIC SCIENCE CHALLENGES AND RESEARCH NEEDS

Fusion research is entering a new era of burning-plasma experimental studies on the reactor scale. A burning-plasma facility, such as the international device targeted for construction by the ITER project near the Cadarache research facility in France, is of unprecedented scale and cost (multibillion U.S. dollars). To mitigate the risk for achieving the desired performance of the plasmas in such novel experimental regimes, a compelling science-based research plan featuring experimentally validated predictive modeling is needed.

Two main concerns involve the larger spatial scales and longer energy-confinement time scales. Fundamental first-principles physics considerations—which require accounting for scales that span the small gyroradius of the ions to the radial dimension of the plasmas radius—indicate that an order-of-magnitude higher spatial resolution is needed. In addition, the major increase expected in the plasma-energy confinement time (approximately 1 second in the reactor used in the ITER project), coupled with the longer pulse of the discharges in such superconducting systems, will demand simulations of an unprecedented duration. For example, electromagnetic microturbulence and certain classes of tokamak MHD phenomena, such as tearing or kink modes, are often treated with the same equations even though there can be significant differences in the geometry or range of spatial scales of interest.

While there has been awareness of the shortcomings of conventional approaches for dealing with phenomena driven by current and/or the thermal gradients, properly addressing such issues usually introduces more complex dynamics into the computation. This is evident, for example, in attempts to consider the effects of the current gradient on drift wave turbulence (microturbulence) and/or the influence of thermal gradients on tearing modes. Because the larger-spatial scale range associated with burning plasma/ITER science challenges introduces higher resolution requirements, the need to treat both types of dynamics becomes unavoidable. Traditionally, tearing phenomena such as magnetic islands have been addressed separately from microturbulence. However, the associated electron/ion scale range is not actually separable, and the two types of dynamics in fact interact self-consistently. Specifically, while the linear growth rates of these phenomena are widely disparate, the nonlinear dynamics governing the energy transfer are not. The concept proposed to resolve this problem is known as mesoscale MHD, which takes into account electromagnetic responses that fall between the global MHD and microturbulence scales. The long wavelength component of the microturbulence exhibits the most prominent electromagnetic characteristics due to the slower damping rates of the associated shear Alfvén waves. In numerical studies of present-day tokamaks, the short wavelength end of a nonlinear MHD computation will overlap with the long wavelength end of a microturbulence simulation (typically falling in the range of toroidal mode numbers of order 10 to 20). For simulations of ITER-like plasmas, there is more separation—but only by a factor of approximately four. Hence, an MHD dynamic range of 20 to 30 modes will still overlap the microturbulence range. Because of cascade tendencies, there is an opportunity for self-consistent nonlinear interaction. In basic MHD turbulence, the inverse transfer of magnetic energy, and particularly magnetic helicity, are well known. Although the scenarios are different, the underlying nonlinear processes associated with MHD turbulence are likely relevant to both MHD and electromagnetic microturbulence in tokamaks.

In general, the problem of incorporating the magnetic islands coupled with turbulence in tokamaks is just starting to be explored with existing computational resources applied primarily to fluid-based models. The next generation of more-powerful computational resources might be sufficient to enable the implementation of advanced kinetic models for cases where unrealistic approximations for the separation of length scale or for species mass ratios could be avoided.

Panel members carefully assessed burning plasma/ITER-relevant scientific challenges with special attention focused on key problems that would require extreme-scale computational capabilities to produce advances in a time-critical manner. Specifically, any significant breakthrough achieved with such advanced computer simulation capabilities would greatly benefit the ITER project in particular and the burning plasma knowledge base in general. Several important fusion energy tasks addressed by the panel require the development of modern experimentally validated models incorporating robust numerical algorithms that are capable of addressing multiscale challenges in space and time associated with such tasks.

A common part of the research identified for all PRDs is the development of advanced theory and algorithms to address the multiscale physics challenges in burning-plasma simulations. Due to a magnetically confined plasma's multiscale nature, its nonlinear dynamics are quite complex. Accordingly, a comprehensive understanding of such plasmas is only possible through large-scale computer simulations. As the floating-point operation power of massively paralleled computers increases to exaflops (10^{18} floating point operations per second) and beyond, the integrated modeling of the long-term dynamic behavior of magnetically confined fusion plasmas—or even a complete discharge cycle—becomes feasible. It is also becoming increasingly important to maintain the long-term accuracy and fidelity of the algorithms implemented in advanced research codes through the duration of an entire simulation. It clearly cannot be assumed that standard, off-the-shelf algorithms for dynamic systems with familiar local accuracy properties are suitable for long-time simulations without modifications. It is therefore important to recognize and address the challenge of developing advanced theoretical formalisms and structure-conserving algorithms—such as the symplectic algorithms for individual particle motion—that can enable simulation capabilities with long-term accuracy and fidelity for multiscale physics in burning plasmas.

The following five topics were identified as PRDs for enabling significant advances in fusion energy research. The associated goal is to achieve experimentally validated predictive capabilities for burning plasmas in general and the ITER project in particular. A comprehensive summary of ITER project needs, which is also relevant for other burning plasma scientific challenges, is provided in ITER (1999a). Further progress in fusion energy science studies of burning plasmas have also been highlighted in a presentation by Shimada et al. (2007).

PRIORITY RESEARCH DIRECTIONS

Transport Under Large-Scale Burning Plasma Conditions

Plasma transport and turbulence is one of the most difficult and least understood problems in fusion research. A comprehensive review of the physics, including experimental and theoretical issues, is provided in ITER (1999b) and Doyle et al. (2007).

One of the biggest challenges in this PRD includes realistic predictive transport modeling, which requires electromagnetic simulation capabilities. While these simulation capabilities presently exist, much work remains to be done in both improving and exploiting them to gain improved understanding of turbulence and transport of plasmas with finite values of beta (ratio of plasma to magnetic pressure in a magnetically confined plasma device). Some of these codes have incorporated algorithms expected to be applicable for high-beta plasmas.

General Status of Research on Tokamak Transport and Turbulence

Transport by turbulence in magnetized plasma is characterized by a basic scale separation. Macroscopic events—such as disruptions—involve MHD motion on scales comparable to the plasma cross-section, which in an ITER-scale plasma would be on the order of 1 m. Theoretical and computational studies of turbulence and associated thermal transport in magnetically confined plasmas have been actively pursued. Connections to experimentally observed behavior have also been well documented. Such phenomena typically occur on scales of order 10 ion gyroradii and are accordingly categorized as microturbulence.

A magnetized plasma is characterized by spatial and temporal scale orderings resulting, for example, from the small sizes of the ratio of the thermal ion gyroradius to the toroidal minor radius. From the temporal perspective, this concerns the ratio of the ion gyrofrequency to the fastest scale of motion in microturbulence—with the ions chosen because their characteristic gyroradius is the largest and their gyrofrequency is the slowest. These ratios are closely related; i.e., if the temperature-scale length is the minor radius, they are actually the same. In current magnetic-confinement devices, these ratios are smaller than 1%, but for ITER plasmas, they are expected to be about four times smaller.

The spatial dimension along the equilibrium magnetic field does not involve such disparate scales—a great advantage when analyzing a magnetically confined plasma. In the direction parallel to the magnetic field, the electron kinetic motion and the Alfvén wave motion are rapid. The range of parallel scales is limited by the strength of the perpendicular dynamics and is set by its ability to compete with the constraints given by the parallel transit frequencies (electron or Alfvén speed divided by the parallel length scale associated with the frequency spectrum of the turbulence). Generally, only one decade in the parallel spatial spectrum, limited by the background parallel scales, is relevant and the smallest scale of interest is about 1 m (extending possibly to several meters for an ITER plasma). Self-consistently solving the field equations for the electric and magnetic potentials associated with this single active scale ensures the electromagnetic wave and fluid plasma responses are carried self-consistently. However, due to the interactions with the background flows and currents, the temporal spectrum is about one decade wider than the spatial spectrum. This problem is expected to persist (and possibly be even stronger) in an ITER plasma.

Overall, there is an increased awareness that all perpendicular spatial scales of a magnetically confined plasmas system—down to the smaller of the ion gyroradius or collisionless skin-depth covering both ion and electron dynamics—should be simultaneously and self-consistently captured in a single computation. The separation of scales adopted establishes the essential scale of computational complexity for the turbulence and transport problem. Microturbulence becomes nonlinear at small amplitudes because it is driven by the thermal gradient in the plasma—with the gradient of the fluctuation becoming comparable to the gradient of the background when the amplitude ratio reaches the scale ratio. Several different nonlinear energy-transfer processes are active (such as wave/wave, wave/particle, electromagnetic, etc.) and tend to transfer toward both larger and smaller scales. It is well known that the turbulence interacts self-consistently with background-scale flows. Research into the self-consistent interaction between turbulence and background-scale currents is still in a relatively early stage of development despite the existence of the basic ideas for several decades. This is one of many examples of how research in the field is limited by access to sufficiently powerful computational resources. There are some essential difficulties in the turbulent transport problem due to the many degrees of freedom and/or the wide range of time and space scales. Specifically, an electromagnetic computation is several tens of times more intensive than an electrostatic treatment of a given problem, and a gyrokinetic computation is several tens of times more intensive than a gyrofluid one. Consequently on the global scale, electrostatic gyrofluid computations are presently conducted in a relatively routine manner, and electromagnetic gyrofluid and electrostatic gyrokinetic computations have become increasingly common. However, global electromagnetic gyrokinetic computations of current tokamaks are barely possible—even if the full capabilities of the largest computational centers were engaged. To carry out ITER-relevant computations of this kind, more powerful next-generation computational resources at the exascale will likely be necessary. As fresh insights from newer investigations become available, the necessary limits on spatial resolution and temporal scale separation may be more stringent. Only with the availability of and access to the next generation of more powerful computational capabilities can such questions be properly addressed.

Scales of Basic Instabilities

The traditional view of microturbulence is that it is caused by and shares the same scaling properties as some underlying basic instabilities—with the linear and weakly nonlinear analysis of these instabilities being reasonably well established. Regarding the scales of these instabilities, the ion gyroradius is associated with the drift wave and ion temperature gradient (ITG) modes, the collisionless skin-depth with the microtearing modes and/or collisionless reconnection dynamics, and the electron gyroradius with kinetic MHD and electron temperature gradient (ETG) modes. Trapped electron effects lead to new instabilities and modify the other instabilities as a result of mirror-trapping motion along the field lines that largely average the parallel force responses. Such trapped-electron mode instabilities largely fall within the range of scales given by ITG mode—though for various combinations of gradient scale lengths among density and the electron and ion temperatures, the scaling properties can be different with the trapped-electron instabilities possibly falling outside the spectral range of ITG modes. The spatial scale is set by the most dominant of these instabilities, and the time scale is determined by their characteristic frequencies.

A global simulation of the plasma down to the collisionless skin depth carries all of the relevant scales self-consistently—with the exception of the kinetic MHD and ETG case. A currently active research topic is the assessment of the interaction between ITG and ETG dynamics, as the separation between the low ETG and high ITG spectral ranges is usually less than one decade. Only computations that can process the relevant scales simultaneously can properly address this issue. Because such computations are well beyond current capabilities, the problem is often addressed with reduced ion to electron mass ratios, perhaps using 900 or 400 rather than the deuterium/electron mass ratio value of 3670. However, compelling systematic assessments of possible qualitative changes from mass ratio variations can only be determined by computations able to deal with the real mass ratio. With the next generation of more-powerful computing capabilities, it may be possible to address this problem—at least for local flux-tube simulations that can treat scales up to tens of ion gyroradii and that can exclude global-scale interactions. If gyrofluid computations are applicable, then carrying out global case studies of this type could be feasible.

Spectrum Broadening and Shifts

Turbulence is not generally a case of interaction between the set of linearly unstable degrees of freedom (i.e., "modes"). Two types of new interactions are possible. The first type may include weakly growing or damped modes that are maintained at finite amplitude by nonlinear interactions. The second type leads to a much less intuitive picture and will be discussed in the next subsection, "Nonlinear Instability, Weak Instability Suppression." For the first category, the main drive is still caused by instabilities, but nonlinear processes transfer fluctuation energy to other scales. Interscale nonlinear interactions are usually cascade interactions (i.e., local in wave number space). Simulations show that the cascades involving the flow due to electric field drift or magnetic field energy are inverse—i.e., the transfer is preferentially to larger scales, while those involving thermal or parallel kinetic energy are direct, or preferentially towards smaller scales. Hence, the transfer goes both ways, and the spectrum tends to act as a unit—the fundamental reason that computations of this type should be self-consistent. The effective range of free energy in an instability-dominated system is about one decade above and below the scale of the instability. Current resources are adequate for any one instability scale for local cases, or for ITG and drift wave instabilities in global cases for moderate device sizes up to 200 ion gyroradii. However, they are not adequate for gyrokinetic computations for the ITER project or for several instability types with simultaneous disparate scales. The second nonlinear interaction is described in the next subsection, "Nonlinear Instability, Weak Instability Suppression."

Nonlinear Instability, Weak Instability Suppression

For the second (less intuitive) category of nonlinear interactions, turbulence is not solely due to interacting instabilities. In fact, in extreme cases turbulence can result from dynamics not involving instabilities. In self-sustained drift wave turbulence, linear instabilities are entirely absent; however, a set of weakly damped modes self-interacts to change the mode structure such that all the modes collectively destabilize each other. Scaling of this turbulence is similar to toroidal drift wave instabilities. The differences are qualitative—a smaller role for the instabilities and a much greater one for parallel Alfvén dynamics with respect to both free energy input from the background gradient and parallel dissipation mechanisms (resistivity, thermal conduction, Landau damping, etc.).

For nonlinearly interacting instabilities, the dynamics governing the temporal evolution of absolute modes (eigenmodes) and the scaling properties of self-sustained turbulence can be quite challenging to understand. Complications here include identifying not only the strong role possibly played by damped modes but also how the ultimate gradient drive can be self-consistently maintained by modes at scales that are unimportant to the instability. Hence, either larger or smaller scales might be especially important. For example, for edge turbulence, the growth rate of toroidal instabilities is bounded by the ideal interchange frequency. The native vorticity of drift wave turbulence at any scale larger than the ion gyroradius is comparable to the diamagnetic frequency. Because the gradient scale length is much shorter than the toroidal major radius, the diamagnetic frequency is larger than the corresponding linear growth rate (with the exception of the longest-wavelength MHD instabilities). As a result, the edge turbulence is largely insensitive to the details of the instability properties. The transport-scaling properties of edge turbulence can be either similar to or different from those of instabilities at the same spatial scale, depending on variables such as dimensionless parameters involving the mass ratio, plasma beta, transit frequencies, and ratios among the three main gradient scales.

Skin-Depth Phenomena

The collisionless skin depth is the scale above which the response to parallel forces on the electrons is either Alfvénic (electromagnetic response mediated by magnetic induction) or fluid-like (electrostatic response mediated by electron inertia). This is a topic more commonly covered by MHD research than by investigations of microinstabilities or microturbulence. For example, reconnection studies involve determining the role of the collisionless skin depth in establishing the character of the parallel response. The importance here is not so much the possible relationship of the skin depth with instability drive mechanisms, but rather this can provide a better understanding of how turbulence is saturated.

Tokamak turbulence studies have been criticized in the past for ignoring the skin depth—but within the last decade, they have been incorporated into electromagnetic studies. Nevertheless, this does make the computation of microturbulence in the tokamak core more difficult because the ratio of the skin depth to the ion gyroradius scales as an inverse square of the plasma beta. For standard cases, the ratio is about a factor of four, but for the ITER project, this might become larger. It should also be noted that the higher magnetic field strength will compensate for higher thermal energy density. Even if the electron gyroradius is ignored—because associated research is still in progress—a proper global electromagnetic microturbulence computation must resolve the skin-depth scale, which accordingly sets the spatial resolution demanded.

Transport Simulation on a Confinement Time Scale

Although transport models for approximating "steady-state" (confinement time scale) turbulent transport conditions have been quite actively pursued and developed, the resultant capabilities have not matured enough to realistically deal with transient transport phenomena (e.g., cold pulse propagation in modulation experiments and the formation of transport barriers). It is important to emphasize that significant progress in understanding such transient transport properties is needed for the development of control methods in future burning plasmas. In addition, the degree of stiffness in the turbulent transport models is not at all similar among the different models. Together with uncertainties in the modeling of pedestal physics, this leads to a wide range of predicted results, none of which carry a high level of confidence for predicting the fusion power gain in a burning plasma/ITER experiment. These open issues require more systematic studies on the interaction between turbulent transport and equilibrium profiles for the longer confinement time scale range of interest. Two possible approaches to address this challenging issue are described here. Each approach begins with the perturbed particle distribution function governing electromagnetic fluctuations in a plasma that can be expressed as $f = f_0 + \delta f$ with the first term being the slowly varying "adiabatic" part. One approach involves the coupling of a simplified low dimensionality (with respect to coordinate and velocity space) transport code with turbulence simulations using the nonadiabatic part of the perturbed distribution function (δf). Here, fixed gradients (associated with f_0) are adjusted based on those gradients evolved in transport codes. The other approach, which is far more challenging, is to develop "full-f" turbulence simulations in which δf and f_0 are evolved on the same basis. In this latter approach, it is necessary to take into account the collisional dynamics governing dissipation and the neoclassical physics responsible for mean flows (or associated radial electric fields). This is essential for self-consistent long time scale simulations that maintain the relevant entropy balance. Compared with the time scales (typically for ~ 1 ms) in most present day microturbulence simulations, future simulations (on confinement times of ~ 1 second) will require factors of thousands of larger time steps. In addition, the velocity-space grid resolution requirement, which is determined by the collisionality, will be far more demanding in the more "collisonless" regime appropriate for burning plasma/ITER conditions. Accordingly, even in the electrostatic limit, such long time scale simulations will require next-generation computational capabilities at the extreme scale.

Associated Computational Needs

In general, the computational requirements to treat the class of problems just described are set by spatial and temporal scale separation demands. These are severe enough to greatly increase the difficulty of simulations but still tractable enough to allow for the necessary self-consistent treatment of important phenomena of interest. For example, a global computation that resolves the collisionless skin depth in an ITER-scale burning plasma is estimated to have a spatial grid size of roughly 4096 x 16,384 x (16 or 32) points with the latter number for the parallel direction. For "gyrofluid" simulations, which have six moment variables per species plus two field variables (or three, if magnetic compressibility is treated), the moment variables are advanced with time-dependent partial differential equations (PDEs) with largely hyperbolic character. The field variables here are solved simultaneously with static PDEs that are largely elliptic in character. This can be strongly complicated by the level of sophistication in the treatment of finite gyroradius effects. The associated run time of such codes can exceed one million time steps, especially if new long time scale phenomena of importance should emerge. For gyrokinetic computations using a velocity space grid, the moment variables are replaced by a distribution function on a domain increased by two dimensions. While earlier treatments used coarse resolution, some newer results indicate the need for as many as 100 or more grid points in these dimensions. Particle-in-cell (PIC) models represent the distribution function with an ensemble of markers—with particles per grid cell replacing the velocity space resolution. The resolution requirements here lead to "noise" reduction challenges.

Expected New Findings

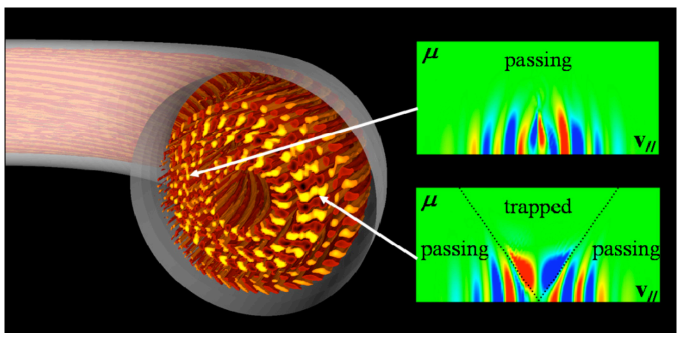
Experience with investigations of nonlinear phenomena in plasma physics indicate that any new study combining ingredients that were previously treated separately will result in unanticipated findings. Examples of this include nonlinear self-sustained turbulence and self-regulation of turbulence by zonal E-cross-B flows—in which the flux surface average component of the E-cross-B vorticity, generated by self-interaction of turbulent flow eddies, mediates the turbulence and changes the gradient threshold.

Usually, an interesting computational result emerges that is not understood (at least initially) and whose physical nature is then debated. This involves examination of self-consistency issues, which emerge when one or more theoretical mechanisms are assessed, when other kinds of numerical studies are carried out, and when more complete demonstrations of the targeted phenomena by well resolved computations are performed. Along the way, the qualitative nature of the phenomena in day-to-day, state-of-the-art computations changes as combinations of more complete physics, more resolution in space (including more dynamic range covering longer and shorter wavelengths) and time (as longer runs become possible). For example, in current studies of the possible relationship between microturbulence driven by the ITG and ETG, a reduction of the chosen mass ratio produces a corresponding change in the ratio of the parallel electron to ion transit frequencies—which then impacts the turbulence dynamics associated with the short-wavelength ETG instabilities and the longer-wavelength ITG modes.

Modeling Plasma Turbulence

The eventual size and cost of a fusion reactor will be determined in large measure by the balance between the self-heating rate in burning plasma and loss processes such as those associated with microturbulence. There has accordingly been much attention focused on advanced simulations to gain a better understanding of the confinement properties of a turbulent plasma in a burning plasma environment. In the figure below, an example of a modern nonlinear kinetic simulation is given using the GT5D numerical code (Idomura et al. 2008). The code resolves the spatial scale and the velocity space scales of the thermal ions, which drive the turbulence for the case considered here.

This example illustrates the need for currently available "extreme computing" capabilities as the associated simulations of around a 3-millisecond duration are performed using computational resources with around 60 teraflops of computing power. For the future, it is estimated that 100 petaflops or more will be needed to carry out an ITER-scale burning plasma simulation of a 1-second duration (i.e., for a discharge of 300 times greater duration) while maintaining the same spatial resolution as in this case. To resolve the collisional skin depth (as described in the text), computing capabilities at the exascale (exaflops) may well be required.



This figure illustrates the plasma cross section and the turbulent structures due to thermal ion temperature gradient driven modes together with the direction of induced plasma transport. In the insert, density perturbations associated with both trapped and passing particles in the presence of turbulence are depicted. These two classes of particles are present in sufficiently "collisionless" toroidal plasmas where some particles can be "trapped" in the magnetic well along the field lines. Images courtesy of Yasuhiro Idomura (Japan Atomic Energy Agency).

Importance of Improved Understanding to Studies of Burning Plasmas/ITER

Tokamak microturbulence studies have been actively pursued for many decades, beginning with simplified lower dimensionality models. Computations with increasing realism and associated complexity have continued to progress at an impressive rate and remain a "hot topical area" of current research because the eventual size and cost of a fusion reactor will be determined in large measure by the balance between the self-heating rate in a burning plasma and loss processes such as those associated with microturbulence. There has accordingly been much attention focused on advanced simulations to gain a better understanding of the confinement properties of a turbulent plasma in a burning plasma environment. For example, global kinetic electromagnetic simulations are currently developing capabilities to address key effects and instabilities extending from the usual microscales to macroscales in MHD-relevant regimes.

Kinetic electromagnetic studies have been pursued using both gyrokinetic and gyrofluid approaches. For significant progress in global gyrokinetic simulations, continuing advances in petascale resources and associated improvements in algorithms will be needed over the next 5 years. However, for applicability to burning plasma/ITER issues, hardware and software advances beyond the petascale will likely be demanded.

Significant advances in understanding the physics of turbulence and transport in burning plasmas are needed to help optimize operating scenarios in the ITER and to address the design challenges for DEMO – the follow-on reactor demonstration device. In particular, the behavior of the pedestal depends on the turbulence and transport in the edge region (outermost 10 cm of the minor radius) of the ITER. The associated multiscale challenges are formidable. Specifically, it will be necessary to account for the dynamical interactions between microscale turbulence and the pedestal mesoscale MHD physics. In general, confinement in the ITER device will be determined by self-consistent interactions between all the phenomena noted. It is likely that extreme-scale computational capabilities will be required to deal with this complex task.

Challenges in the Physics and Simulations of Plasma Disruptions

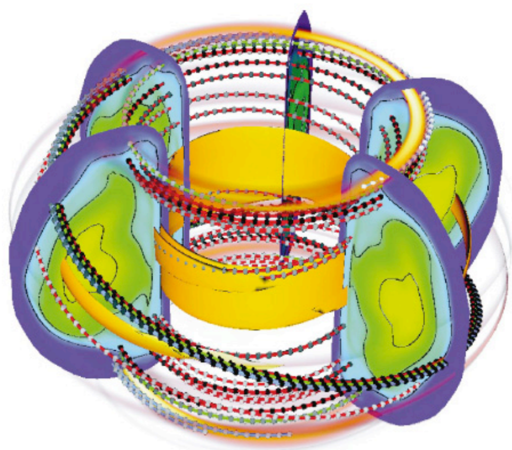
The driving physics goal for simulations of plasma disruptions is to develop an experimentally validated predictive capability for ITER-class experiments (see ITER 1999c and Hender et al. 2007). This is arguably the most mission-critical challenge because nonlinear macroscopic events play a central role in defining the operational space of these devices and because many of the details of the nonlinear dynamics and interactions remain poorly understood. While the onset conditions for macroscopic instabilities in tokamaks are reasonably well known, it is also the case that such linear stability thresholds are often crossed in fusion experiments. This is manifested, for example, with the appearance of current-peaking phenomena (such as sawtooth oscillations and tearing modes) and in the formation of the edge pedestal with the associated presence of ELMs. The subsequent nonlinear evolution of these events determines whether the instabilities manifest themselves as small amplitude repetitive oscillations or as large amplitude disturbances that can release significant amounts of thermal energy or couple with other modes and lead to plasma disruptions.

Disruption Avoidance

A disruption is basically the rapid termination of a plasma discharge with the accompanying loss of the stored thermal and magnetic energy. It is well known that all tokamak devices can exhibit disruptions under certain circumstances. The threat of a disruption sets limits on the maximum values of the plasma current, pressure, and density. A plasma disruption will cause the plasma current to decay at a rate of up to 109 A/second. Inductive effects cause this current to then be transferred to the surrounding metallic structures, such as the vacuum vessel, with accompanying large forces. These forces scale as the square of the plasma current—and because the plasma current in the ITER device is about three times as large as in any existing tokamak, the associated forces will be about an order of magnitude larger than those encountered today. Accordingly, the ITER can only withstand a few of these full-strength disruptions, and a fusion reactor must be designed to be virtually disruption free. In addition to the forces associated with the current quench, the disruption causes a sudden dump of plasma-stored energy to the walls and divertor plates—which would cause unacceptable erosion. In addition, a large collimated beam of multi-MeV electrons can be produced. The loss of these "runaway" electrons would damage the vessel.

Disruptions Pose a Grand Challenge for Burning Plasma Experiments

During tokamak experimental operation, events that rapidly terminate plasma discharges occasionally occur. The complete and rapid loss of thermal and magnetic energy in these disruptions results in large thermal and magnetic loads on the material wall. For proposed next-step experiments or projects such as ITER, the stored energy will be approximately 100 times greater than present-day devices, greatly increasing the potential damage of these events' numerical simulations. Because of the nonaxisymmetric properties of the disruption events, the heat load can be very localized.



Certain field lines (blue) open to strike the top and bottom divertor plates during the disruption. Source: Kruger et al. (2005). Reprinted with permission from SE Kruger, DD Schnack, CR Sovinec, *Dynamics of the Major Disruption of a DIII-D Plasma*, Vol. 112, Page 056113, 2005, Copyright 2005, American Institute of Physics.

The emergence of fully three-dimensional MHD dynamic modeling capabilities offers encouraging prospects for making more reliable quantitative predictions, and for helping assess disruption scenarios and their consequences for ITER operation. The application of these models to interpret data in present tokamaks is indispensible for evaluating some of the internal cause-and-effect dynamics that lead to disruptions. The primary goal of such computational studies is to identify parameter regimes and operational modes for the tokamak that are virtually disruption free. This is a necessary requirement for a viable fusion reactor. Accordingly, three-dimensional MHD simulations are being conducted to help clarify the causes for the wide range of scatter in existing databases that tabulate the effects of disruptions, such as the current quench rates and the forces imparted to the surrounding vacuum vessel. Overall, the challenge is to ensure higher physics fidelity in improved models that incorporate realistic boundary conditions and are systematically validated with experimental data. Realistic simulations are the only reliable means to project this data to ITER-scale devices.

Disruption Mitigation

As noted in the preceding discussion, the production of high-energy runaway electrons during disruptions continues to be a serious concern for the ITER project. In the largest present-day tokamak, the Joint European Torus, up to half of the thermal predisruption current can be converted into such runaway currents. It is currently estimated that, if not mitigated, the disruption produced runaway currents in the ITER reactor will be much higher—with some simulations showing that 70% to 80% of the equilibrium plasma current will be converted into runaway current following a disruption. Other simulations of the effect of the runaway energy deposition expected on the ITER's first-wall or divertor surfaces indicate that an uncontrolled runaway current interaction of this magnitude has the potential to produce severe damage to the plasma-facing component (PFC) surface and cause substrate melting and erosion.

Several mitigation techniques have been proposed for the ITER device with some currently being tested on existing tokamaks. These include injecting pre-emptive "killer pellets" (Jardin et al. 2000) consisting of hydrogen and impurities. Another mitigation technique involves "massive gas injection," whereby an intense stream of neutral gas is injected into the tokamak at the onset of the disruption. Both of these techniques have been demonstrated experimentally on smaller tokamaks, but the highly nonlinear processes involved make it difficult to simply extrapolate the results from smaller tokamaks to ITER-sized devices. The limitation on direct extrapolation is due in part to the different combinations of plasma geometry and surrounding passive structures in the ITER device (as contrasted with geometry and structures in existing machines). Limitations are also related to the differences in energy levels and underlying physics processes that arise from the increase in the ITER device's plasma size, current level, and energy content. Researchers have had some success in simulating these techniques in existing experiments, but extrapolating these for use in the ITER device requires a large increase in computational power as described below.

Edge-Localized Modes

Another physics mechanism that must be controlled or eliminated in the ITER device and in future reactors is the ELM. For the "H-mode" (high-confinement mode) plasmas, ELMs can dump large amounts of plasma energy as heat and particles on the surrounding wall in a short time. All existing high-performance tokamaks exhibit ELMs, and they are generally regarded as benign events associated with the H-mode. However, extrapolating ELMs to power levels used by the ITER device indicates that they are clearly unacceptable events. Specifically, they can quickly cause melting and severe erosion of the divertor plates. It has been demonstrated experimentally that if nonaxisymmetric magnetic fields of a particular form are deliberately applied to an H-mode plasma, ELMs can be eliminated, but the good confinement can be maintained. There has been some success in simulating these results on existing experiments using three-dimensional MHD codes. However, such simulations of resonant magnetic perturbation dynamics are very demanding because of the high resolution needed. Simulations applicable to ITER plasma are even more challenging because of the more-extreme parameters. Improvements in these simulation capabilities are accordingly required to optimize the application of the nonaxisymmetric fields.

Computational Requirements

Today's MHD codes can realistically simulate the events associated with the required time and space scales in a small tokamak such as the Current Drive Experiment-Upgrade (CDX-U) device. However, the computational requirements for the ITER are much more formidable (see Table 1).

Table 1. A Factor of 108 More Space-Time Points is Needed to Simulate ITER Compared with CDX-U for an Explicit Simulation with Uniform Zoning

Name	Symbol	Units	CDX-U	DIII-D	ITER
Field	B0	T	0.22	1	5.3
Minor Radius	a	m	.22	.67	2
Temp.	Te	keV	0.1	2.0	8.
Lundquist	S		$1´10^{4}$	$7´10^{6}$	$5´10^{8}$
Growth Time	$t_A S^{1/2}$	s	$2´10^{-4}$	$9´10^{-3}$	$7´10^{-2}$
Larmor Radius	r_i	m	$4.5x10^{-3}$	$4.4x10^{-3}$	$1.7x10^{-3}$
Zones	$N_R´N_q´N_f$		$3´10^{6}$	$5´10^{10}$	$3´10^{13}$
CFL Time Step	DX/V_A (Explicit)	s	$2´10^{-9}$	$8´10^{-11}$	$7´10^{-12}$
Space-Time Pts.			$6´10^{12}$	$1´10^{16}$	$6´10^{20}$

Table 1 demonstrates that if uniform zones and a fully explicit time-stepping algorithm were used, a calculation for the ITER reactor would require 10^8 more space-time points than in existing calculation of CDX-U. Much of this disparity is illustrative of the major challenge that will need to be addressed by formulating and using innovative approaches such as novel implicit algorithms and nonuniform adaptive zoning methods. Computers with much more substantial power will clearly also be needed (Sipics 2006).

Code performance for implicit algorithms is primarily limited by the solution of the associated large sparse matrixes. Several numerical algorithms for enabling massive parallelization of codes are currently being developed. These codes are expected to be able to execute in a much more efficient manner MHD disruption and ELM simulations. These will likely involve techniques such as algebraic multigrid and three-dimensional domain decomposition. Such capabilities must be further developed and applied for improved codes to take full advantage of emerging massively parallel computing hardware to address ITER-relevant scientific challenges.

Burning Plasma/ITER Energetic Particle Dynamics in the Presence of Multiple Alfvénic and MHD Instabilities

A key topic of interest in burning plasma experiments and projects, such as the ITER, is the confinement of fusion alpha particles in the presence of multiple alpha-driven Alfvén mode instabilities (as well as the interplay between alpha particles and the more global MHD instabilities [ITER 1999d and Fasoli et al. 2007]). The basic theory, supported by experimental validation studies, has shown that alpha particles and other super-Alfvénic energetic particles can resonantly excite Alfvén instabilities such as the toroidal Alfvén eigenmodes and other energetic particle modes. Investigations for plasma conditions that are expected to be encountered in the ITER indicate these instabilities are likely to be present, can cause significant alpha-particle losses, and may damage the reactor wall. In addressing the challenge posed by such multiscale nonlinear problems, researchers must develop efficient and scalable numerical methods that can be applied to the most powerful computers available.

Proper understanding of energetic particle physics is a critical issue for burning plasmas such as those encountered in the ITER project. These include fusion-product alpha particles and energetic ions from neutral beam and radio frequency sources, which are used to heat the bulk plasma via collisions with electrons as well as drive plasma current and rotation. However, because these energetic particles typically fall in the super-Alfvénic frequency range, they can resonantly destabilize shear-Alfvén waves, such as the toroidal Alfvén eigenmodes. As already noted, such Alfvénic instabilities can cause anomalous transport of the energetic ions and lead to damage to the reactor wall.

An important new topic in studies of burning plasmas is the nonlinear interaction between the energetic particle component and the bulk thermal background plasma. The associated dynamics can have favorable and deleterious consequences. Specifically, in addition to plasma heating, energetic particles can influence the bulk thermal plasma in other important ways. Neutral beams in the ITER can drive plasma rotation to stabilize resistive wall modes and also serve to drive the plasma current, which is important for the hybrid operational mode and other advanced operating scenarios. Alfvén instabilities can broaden the beam ion distribution and influence the profile of beam-driven current and toroidal rotation. The energetic particle-driven Alfvén instabilities can induce zonal flow, which may suppress core plasma turbulence. Furthermore, energetic particles can impact MHD modes significantly. In ITER plasmas, fusion alpha particles can be expected to stabilize the internal kink mode, leading to very large ("monster") sawteeth oscillations. Alpha particles can also stabilize resistive wall modes. Conversely, deleterious influences include significant "anomalous" enhancement of energetic particle transport and losses that can occur in the presence of unstable MHD modes in the bulk thermal background plasma. Such negative effects on energetic particles can also be induced by the presence of thermal plasma turbulence.

At present, the state-of-the-art kinetic/MHD hybrid codes can routinely simulate one cycle of growth, saturation, and decay of energetic particle-driven Alfvén modes with moderate toroidal mode numbers. However, this can only be conducted for parameters of present tokamak experiments. Much further development will be required to produce simulation capabilities capable of predicting the alpha-particle confinement in burning plasmas.

Nonlinear simulations of alpha-particle driven instability and alpha-particle transport in burning plasmas are extremely challenging for a number of reasons including:

- Spatial resolution requirements. High-spatial resolution is required to resolve multiple high toroidal mode number (n) Alfvén modes that are expected to be destabilized by alpha particles. The mode numbers of energetic particle-driven Alfvén instabilities scale with the ratio of the tokamak size (minor radius) to the energetic particle gyroradius. Because the magnetic field strength in burning plasma/ITER-scale experiments is stronger and the size is larger than those of the present machines, the alpha particle-driven Alfvén instabilities will have much higher mode numbers. Specifically, when carrying out simulations, up to $n = 40$ modes are needed to properly model the nonlinear evolution of the multiple Alfvén modes in a burning plasma, with the most unstable mode number expected to fall in a range from 5 - 30 (Fasoli et al. 2007).
- Velocity-space resolution requirements. High velocity-space resolution for Alfvén modes is required to resolve the fine hole-clump structures caused by wave- particle resonant interaction and particle distribution relaxation in the presence of multiple high-N modes. This translates to a demand for a large number of particles (up to 1 million) to be used for PIC simulations or for a large number of grid points in phase space be deployed for the Vlasov/continuum calculations. These requirements

are similar to those discussed previously in the context of resolving the thermal ion gyroradius (finite Larmor radius) scale in kinetic turbulence computations.

- Temporal-scale requirements. To accurately determine the mode saturation and bursting, as well as possible mode avalanche for Alfvén modes, it is necessary to resolve the time scales associated with the fast Alfvén wave oscillations, the slow mode growth, and the slower collisional relaxation of the energetic ions. This requires accurate long-time simulations with appropriate specification of sources and sinks for the energetic particles.
- Integration requirements. Integrated simulations with thermal plasmas are even more challenging than simulations of toroidal Alfvén eigenmode-like nonlinear effects. For example, it has been shown recently that plasma background turbulence may significantly affect alpha-particle transport. To model this properly, researchers must perform integrated simulations that include the Alfvén instability dynamics and the thermal plasma turbulence behavior. This requires even higher spatial resolution than when only simulating Alfvén instability dynamics.

In summary, nonlinear simulations of alpha-particle confinement are very formidable tasks due to multiple temporal and spatial scale resolution challenges. Scientists must develop efficient and scalable numerical methods capable of solving this multiscale nonlinear problem. Such advanced methods, together with extreme-scale computing resources, are expected to be required to properly model alpha-particle transport in the presence of multiple high-n Alfvén instabilities.

Radio Frequency Heating and Current Drive Simulation Challenges

The planning for radio frequency heating scenarios in the ITER project generally encompass three well-known frequency ranges: ion cyclotron radio frequency (ICRF), lower hybrid radio frequency (LHRF) and electron cyclotron radio frequency (ECRF). ICRF provides the bulk heating and central current drive. LHRF provides the edge current drive and control of the edge value of q (the MHD safety factor). ECRF is used for precise current generation to suppress some MHD instabilities. Of these three, ICRF and ECRF are included in day-one operations in the ITER program, and LHRF decisions are left open as possible upgrade paths (ITER 1999e).

The issues introduced by the ITER project for radio frequency physics fall broadly into three categories: 1) radio frequency interactions with alpha particles; 2) larger configuration dimensions to simulate; and 3) linear and nonlinear coupling effects in the high-temperature pedestal and scrape-off layer of ITER plasmas. Additional issues relevant to present-day devices that are also of interest to the ITER program include radio-frequency induced momentum (flow drive) for stabilization of resistive wall modes, radio-frequency interactions with plasma sheaths on PFCs, and refinement of the plasma dielectric model for RFs.

Alpha-particle interactions are an issue for LHRF and ICRF as well as for MHD Alfvén eigenmodes. Simulating LHRF-alpha interactions have involved the use of ray-tracing and Monte-Carlo codes (Barbato and Saveliev 2004). Results from these investigations can be followed up with full-wave calculations of LHRF in ITER plasmas to evaluate the acceptability of geometric optics in low damping cases. The present approach—using a mixed spectral basis in the poloidal direction and radial finite elements—is computationally intensive (Wright et al. 2008), as it results in large, dense matrices. For lower-hybrid full-wave simulations in present-day tokamaks, about 5000 CPU hours are needed on current computational platforms—simulations that have been readily carried out on a terascale system.

However, processor requirements can be expected to increase a factor of 10^6 if a single toroidal antenna mode was sufficient for lower-hybrid modeling of ITER plasmas. New algorithms for the problem formulation and matrix solution would likely be required, but the present algorithm could be applied to the ITER on an exascale system. A similar degree of work would be involved in evaluating geometric optics approximations in ECRF, though this has been done to some degree with beam-tracing codes and found not to be critical (Prater et al. 2008). ICRF-alpha interactions are more computationally intensive because ICRF requires full-wave modeling for accuracy. Initial studies modeling non-Maxwellian distributions and advanced by Monte-Carlo code in the ITER have been carried out (Wright et al. 2005; Jaeger et al. 2006b, 2008) but have not been coupled to Monte-Carlo codes to assess the importance of finite orbit (finite banana width, etc.) effects on any parasitic losses. This will require full-wave solutions, preferably three-dimensional solutions in ITER plasma geometry retaining all cyclotron harmonics iterated with a Monte-Carlo code using tens of millions of particles. Each three-dimensional simulation requires about 2 hours on 2048 processors on the Jaguar XT3/XT4.

The second issue is full-wave simulation required for ICRF. Dimensions of the ITER are 3 to 10 times larger than present-day tokamaks, while the scale lengths of ICRF waves are about the same. Existing full-wave codes scale well in terms of processor usage for a given problem. There is ongoing development to improve the absolute performance of the codes. For same-day results from simulation for ICRF analysis, processor counts of the order of 10^4 will be required. Possible mode-conversion scenarios in ITER plasmas would also require increased resolution and the use of more processors.

Coupling analysis, especially in ICRF and LHRF, is an issue because of the requirement for greater antenna-plasma separation to protect the antenna from the high-edge plasma temperatures and heat fluxes in the ITER. Antenna designs can be refined to avoid large surface potentials under plasma load. Nonlinear simulations of three-wave coupling processes and radio frequency plasma-sheath interactions are needed to understand how to mitigate these important parasitic edge losses. Important coupling effects depend on local nonaxisymmetric geometry, and resolving these effects will require giving up the axisymmetric assumption and effectively increasing the dimensionality of the problem, thereby raising the order of the computational work needed. Near field sheaths formed at the end of field lines magnetically connected to antennas (D'Ippolito and Myra 2006) and sheaths on the antennas and waveguides themselves—as well as direct acceleration of ions or electrons by high voltages on the antennas—provide sinks for radio frequency energy. The location and strength of the energy loss depends on the details of the edge-geometry for the plasma and its facing components. The sheath rectification is nonlinear. Taking into account the nonlinear and three-dimensional aspects of sheath effects will require significantly more computational resources than currently engaged in carrying out simulations with present models.

While not specific to the ITER program or burning plasma physics, it is important to consider issues relevant to improving the overall validity of radio frequency models. Exploring more-sophisticated models of wave particle interactions needed to more accurately calculate the plasma dielectric response and the deposition of wave energy leads to progressively more computationally intensive challenges with respect to the computational resources demanded and to dealing with the data generated. Approximations commonly invoked in present-day radio frequency codes include using simple Maxwellian distributions, which are treated analytically and depend only on the temperature and density of a flux surface; ignoring the banana-width of particle orbits in Fokker-Planck calculations of the distribution function evolution; and treating the wave-particle interactions linearly—the orbits of the particles are those that are unperturbed by radio frequency waves. Addressing each of these approximations raises the order of dimensions of the plasma representation in phase space and the associated computational requirements needed to determine the plasma dielectric and the plasma response. For example, when non-Maxwellian distributions are used, velocity space integrals must be computed numerically to determine the dielectric response, and the distribution function must be stored (Jaeger et al. 2006a). However, with finite banana width effects taken into account, two-dimensional midplane velocity space coordinates can be used on each flux surface (Figure 1). With finite orbit effects, the poloidal location must be stored as well (Choi et al. 2006). Nonlinear orbit effects and steep power deposition gradients are currently being modeled with gyrokinetics (Park and Chang 2007).

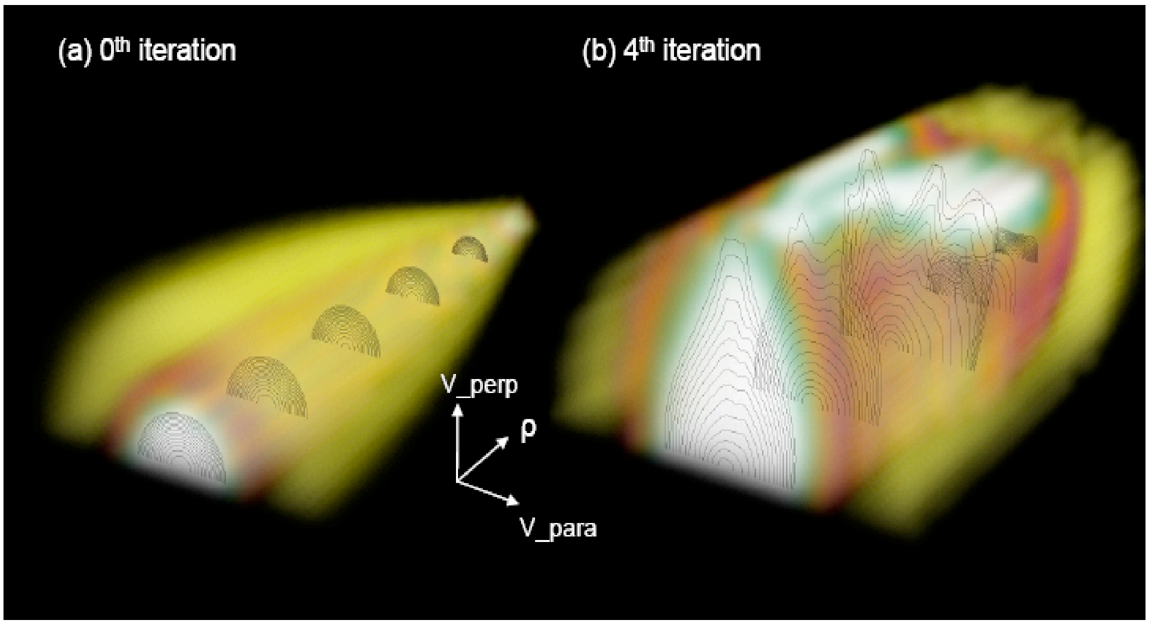


Figure 1. Minority hydrogen heating-induced evolution of the distribution function at the midplane in Massachusetts Institute of Technology's Alcator C-Mod tokamak experiment. Image courtesy of EF Jaeger (Oak Ridge National Laboratory). Source: Jaeger et al. (2006a).

In the realm of integrated modeling, the coupling of ECRF, Fokker-Planck, and MHD codes will be needed to more comprehensively study neoclassical tearing modes and stabilization of sawtooth oscillations. Also, the coupling with turbulence and transport codes in a self-consistent manner will be valuable for systematic studies addressing the creation of flow drive or the effects of current profile control (i.e., at the reversal surface of the MHD safety factor [q]) on transport. Any multiphysics coupling efforts will place greater demands on the execution time and efficiency of radio frequency codes.

Physics of the Edge Barrier Region in High-Performance Fusion Plasmas

High-performance (H-mode) operation in tokamaks is achieved via the spontaneous formation of a transport barrier in the outermost region of the confined plasma. This edge barrier is characterized by very sharp gradients in the density and temperature profiles, which gives the appearance of lifting the core plasma up onto a step or "pedestal." The pressure at the top of this barrier region, or "pedestal height," has a strong impact on overall performance of the plasma, and a high pedestal is likely to be crucial for optimizing fusion properties in the ITER and other tokamak burning plasma devices (ITER 1999b and Doyle et al. 2007). However, the strong pressure gradient and resulting bootstrap current in the pedestal region provide free energy that can drive ELMs. Large ELMs can limit the lifetime of PFCs, and mitigating or eliminating ELMs is an important goal of the ITER research program. Use of imposed three-dimensional magnetic fields to enhance edge particle transport, pellet triggering of small ELMs, and operation in small-ELM or ELM-free regimes are considered to be possible approaches to address this issue. Operation with a high pedestal and small or no ELMs is key to the ITER project achieving its performance objectives, and continued optimization of the pedestal should allow high performance in planned fusion reactors.

Predicting and optimizing the performance of the ITER reactor therefore requires a quantitative understanding of the physics controlling the edge barrier and driving ELMs. While substantial progress has been made in this area recently (Doyle et al. 2007), quantitative understanding continues to lag that of the core plasma. This is largely because direct numerical simulation of the edge plasma is far more challenging. Although core turbulence simulations have saturated the capabilities of the largest present-generation computers and will continue to be extended as discussed in earlier discussions, edge turbulence simulations with similar levels of realism are expected to stretch the capabilities of even the most advanced next-generation machines. However, performing such simulations is critical for the ITER project to reliably achieve its objectives, and such simulations are expected to be a high priority as the magnetic fusion energy community moves forward. Simulations of the plasma in the edge barrier (or pedestal) region are highly challenging for a number reasons, including the following:

- Invoking the separation of spatiotemporal scales, which in the core plasma research area is used to allow separate simulations of stability and transport physics, is not generally justifiable in the edge barrier region. Specifically, the equilibrium scale lengths and time scales are not sufficiently different from those associated with turbulence. Hence, there is a strong need for simulations capable of including both the relevant gyrokinetic and MHD physics as well as operating across a wide range of overlapping spatiotemporal scales. For example, taking into account the properties of atomic physics, electron drift waves, Alfvén waves, electron transit motion, ion drift waves, ion transit motion, ion collisions, macroscopic evolution, etc. that are characteristic of a typical pedestal plasma in existing tokamaks can encompass temporal scales that span up to seven orders of magnitude. Perturbations in the edge barrier—both those associated with L-mode turbulence and those associated with ELMs—are not necessarily small compared to the background equilibrium. Hence, the usual "delta-*f*" approximate techniques that simulate small perturbations (relative to the equilibrium) are not applicable to the edge for all circumstances of interest.
- The electrostatic approximation, which has greatly facilitated efficient simulations in the core, is generally not valid in the edge barrier plasma because the strong gradients in the barrier result in the plasma being close to the ideal ballooning mode critical gradient—a regime in which magnetic perturbations play an important role. The associated electromagnetic simulations are challenging because they introduce additional spatiotemporal scales and numerically challenging operators, and they cause the magnetic topology itself to change. In some situations, magnetic perturbations can become large enough that much of the advantage of field-aligned coordinates (which take advantage of the strong anisotropy of magnetized plasmas) is lost.
- The pedestal in a high-performance (H-mode) plasma typically crosses from a highly collisionless regime near the top of the pedestal to a collisional zone near the separatrix and extending into the scrape-off layer. Most simulation techniques are tailored either to collisional or collisionless regimes. Simultaneously treating both accurately requires a fully kinetic treatment that includes an accurate

collision operator. In dealing with such complexities, the numerical advantages of further optimization of a particular regime are often lost.

- The edge barrier region is associated with strong current and particle sources via the bootstrap current and fueling by neutrals from the unconfined plasma region. Hence, the source physics cannot be separated in general from the stability and turbulence physics, as is often done in core plasma calculations.
- Equilibrium currents and flows are expected to be important, as are the presence of impurity species, strongly shaped geometry, and complex topology associated with the X-point.

The above challenges require enormous computational capabilities to span the wide range of scales, and these challenges also require researchers to rethink traditional paradigms for developing physics insight and choosing appropriate approximations to allow practical simulations that can take optimal advantage of the available extreme-scale computing resources.

Development of a broad range of computational tools is necessary to develop and test physics understanding. In addition to the existing extended fluid and five-dimensional "delta-*f*" gyrokinetic codes and the emerging "full-*f*" electrostatic gyrokinetic codes, fully electromagnetic gyrokinetic codes capable of addressing dealing with realistic edge plasma conditions will be needed. In the long run, six-dimensional codes might even be needed to rigorously assess the limitations of gyrokinetic theory in the edge barrier region. Such six-dimensional codes would present an opportunity for truly massive-scale parallelization together with the deployment of very large numbers of processors.

In summary, it is currently expected that the ITER will require the establishment of a strong edge barrier region with no ELMs (or at most small ELMs) to achieve its performance goals. Key issues include significant improvements in understanding of pedestal structure, ELMs and ELM control, and pedestal formation (i.e., the L- to H-mode transition). Some near-term goals include the ability to simulate between ELM transport; three-dimensional field penetration and transport; barrier formation; pellet triggering of ELMs; between-ELM heat loads on PFCs; and ELM heat loads on PFCs. Simulations of the edge barrier are extremely challenging and computationally intensive due to the wide range of coupled scales, complex geometry, wide range of collisionality, electromagnetic perturbations, and coupling to atomic and materials physics. Innovative new approaches will need to be developed because standard paradigms and traditional numerical techniques developed for the core-plasma research area are not generally applicable. A full range of tools—including three-dimensional kinetic-fluid, four- or five-dimensional gyrokinetic, and possibly even six-dimensional kinetic methods—could be needed. Even with exascale computing, substantial work is needed to develop more efficient formulations as well as advanced algorithms and solvers. If successfully conducted, the associated research progress would potentially have enormous impact on accelerating advances in magnetic fusion energy development. In particular, achieving high-pedestal regimes is expected to improve fusion performance, and very high pedestals can possibly provide access to extreme performance regimes. Control and mitigation of ELMs, as well as heat and particle flows to material components in general, are essential to currently envisioned fusion concepts. They would also stimulate development of strong ties to materials science and have important implications for overall optimization of magnetic fusion concepts.

CONCLUSIONS

Productive discussion during the workshop resulted in the identification of a number of important challenges relevant to the physics of burning plasma/ITER experiments. A common overarching aspect of these scientific challenges involves the need to develop new plasma simulation capabilities that can effectively address the larger spatial scales extending to ITER-like plasma dimensions and longer temporal scales extending to actual transport times.

Any major breakthrough in the five PRDs outlined in this panel report is expected to have an immediate impact on the fusion program and the effective design of a fusion reactor. The actual delivery of such scientific advances is becoming increasingly urgent in view of the current construction phase of the ITER and international discussions of next-step fusion reactors.

Because of the multibillion-dollar (U.S.) construction cost and the complexity of a future fusion reactor, advanced numerical simulation capabilities that are validated against experiments will clearly need to be in place before a reactor prototype can be properly designed and built. As the fusion program moves into the future, it is expected the required optimization of plasma scenarios under burning plasma conditions will demand major advances in scientific understanding. Associated research campaigns, greatly aided by computing at the extreme scale, will likely result in key discoveries of new conditions and important new physics insights to accelerate progress toward resolving burning plasma/ITER scientific grand challenges.

ADVANCED PHYSICS INTEGRATION CHALLENGES

Co-Leads: John R. Cary, Tech-X Corporation and the University of Colorado
Arnold Kritz, Lehigh University

Panel Members: Donald Batchelor, Oak Ridge National Laboratory; Glenn Bateman, Lehigh University; Jeff Candy, General Atomics; Vincent Chan, General Atomics; C.S. Chang, New York University; Ronald Cohen, Lawrence Livermore National Laboratory; Patrick Diamond, University of California, San Diego; William Dorland, University of Maryland; Atsushi Fukuyama, Kyoto University; Steve Jardin, Princeton Plasma Physics Laboratory; Scott Kruger, Tech-X Corporation; Wayne Houlberg, ITER; Wei-li Lee, Princeton Plasma Physics Laboratory; Andrew Siegel, Argonne National Laboratory; and George Tynan, University of California, San Diego

CURRENT STATUS

The Advanced Physics Integration Challenges panel identified key topics of magnetically confined plasmas for fusion energy that could be significantly advanced by use of extreme-scale computing. Traditionally, computational fusion energy science has addressed separate areas such as macroscopic stability, radio frequency (RF) sources, energetic particles (injected from fusion), microturbulence, transport, and edge plasma physics (where atomic processes are important). Each of these areas can currently use the capabilities of existing leadership-class facilities (LCFs). With extreme-scale computational power, it will be possible to couple these large-scale simulations to produce an experimentally validated integrated modeling capability for scenario modeling of the whole device.

The panel report titled, "Burning Plasma/ITER Science Challenges," already introduced the importance of accurate and full computations. As noted in that panel report, the worldwide fusion program is embarking on the ITER experiment, which is of unprecedented scale and expense. Consequently, it is important that each experimental be modeled in detail to plan experimental discharge scenarios and to analyze data after a discharge is completed so the knowledge gained is maximized.

Even with extreme-scale (10^{18} floating point operations per second) computing resources, simultaneous direct simulations of all of the physical processes in a tokamak plasma are not possible. As noted by Cary et al. (2009), such fundamental simulations would require on the order of $10^{6\text{-}12}$ times the age of the universe to complete on petascale hardware, which still implies $10^{3\text{-}9}$ times the age of the universe to complete on extreme hardware.

Coupled simulations allow one to take advantage of proven approximations for each region, thus reducing the total computational time. For example, turbulence computations in the core can use the gyrokinetic approximation so that only time scales on the order of the drift frequency are resolved, resulting in an increase in the time step of several orders of magnitude. Even when the gyrokinetic approximation is used, a well-resolved simulation of core turbulence can require 1000 seconds on 512 processors to simulate one millisecond of experimental time. However, further savings in computational time can be obtained by noting that one need not run the turbulence computation continuously, but only often enough to recompute fluxes as needed for profile evolution; i.e., a few times each second of experimental time in ITER simulations. This leads to computational savings of another factor of 103.

Similar computational savings through coupling are possible in other areas. For example, even though the magnetohydrodynamic (MHD) modes in a magnetically confined plasma can evolve on microsecond time scales, users do not always need to follow that time scale. Instead, one can use a slowly varying equilibrium calculation and couple that to a stability calculation. When instability is indicated, a computation can be invoked to follow the relatively fast dynamics through the short time period during which the instability grows, enters the nonlinear phase, and reorganizes the plasma.

Integrated modeling has a long history in the fusion community for profile evolution. U.S. computational applications include BALDUR, CORSICA, ONETWO, PTRANSP, and TRANSP. However, such computations have for the most part been serial. Consequently, only minimally computationally intensive calculations were generally used, such as highly reduced models of turbulence-induced cross-surface fluxes. In other cases, the dimensionality of the computations was reduced, potentially leaving out important physics. For example, fast particles have been modeled in two-dimensional context only, thus leaving out the effects of magnetic ripple, which can lead to significant losses.

In recognition of the need to move to integrated models on parallel platforms, which would enable modeling of fusion systems with greater fidelity, the U.S. Department of Energy's (DOE) Office of Fusion Energy Sciences (FES) funded three coupling projects: the Center for Plasma Edge Simulation (CPES), Framework for Core-Edge Transport Simulations (FACETS), and Simulation of Wave Interactions with MHD (SWIM). Each of these projects is addressing different aspects of coupling, with differences in both the physics and the methods. These three coupling projects are co-funded by DOE's Office of Advanced Scientific Computing Research (ASCR) so that expertise in advanced computation can be used to address these problems.

BASIC SCIENCE CHALLENGES AND RESEARCH NEEDS

The progression to extreme-scale computing will make possible the use of even more sophisticated integrated models. The fusion community is taking steps toward full device modeling; it recognizes there are multiple computational couplings that must be performed to obtain the understanding needed for a more complete integrated model, in which all of the relevant physics is considered. The discussion identified the following coupling problems to initially be addressed:

- transport modeling with embedded local turbulence
- transport modeling with embedded global turbulence
- coupling disparate regions of the plasma
- macroscopic stability control using RF power
- edge transport with recoverable macroscopic dynamics
- performance optimization of burning plasmas.

PRIORITY RESEARCH DIRECTIONS

Transport Modeling with Embedded Local Turbulence

Summary of Research Direction

In the core, plasma densities and temperatures are nearly constant on flux surfaces because parallel transport causes rapid poloidal equilibration. As a result of this effect and establishment of distribution functions on time scales much shorter than the confinement time, it follows that one can describe the plasma evolution by a set of one-dimensional conservation equations relating the change of the densities and temperatures to the sources and the divergence of the flux.

The embedded local turbulence approach relies on locality. That is, the fluxes of particles and heat from turbulence are assumed to depend only on the local plasma parameters; e.g., the plasma rotation, densities and temperatures and the geometric quantities describing a flux surface, such as averages of the magnetic field strength. (This is the approach used by all of the integrated modeling computational applications.)

Computations of the fluxes can have various levels of fidelity. Higher fidelity tends to come with greater computational intensity. The simplest models express the fluxes in terms of algebraic formulas. Such calculations can easily be performed in serial with rapid turnaround (computational times of order the experimental time). Moving up the scale, one has more complicated, physics-based models (e.g., MMM95, GLF23, TGLF) that use quasilinear approach to compute the anomalous diffusivities, with the parameters calibrated by fitting experimental results or full turbulence computations. These calculations have traditionally also been performed in serial. Parallelizing these computations over roughly 1000 processors (not yet routine) can significantly increase the speed of computation. Beyond that, users can compute the fluxes using full, local turbulence calculations. Because fully resolved turbulence computations covering 1 ms of experimental time require roughly an hour on hundreds of processors, having 100 of such simulations running simultaneously—as needed to update the plasma profiles across the full core—will then require petascale resources. However, to speed up these calculations so that the execution time for an entire global simulation is reduced to a few hours, exascale computing resources will be required.

Scientific Challenges

Of course, it is never as simple as throwing a larger computer at the problem. With each new architecture, new computational software methodologies are needed. Just as the move to message passing parallelism required a nearly complete redesign of computational applications, so will the move to exascale because it will likely rely on new types of parallelism. Even when using the present message-passing machines with modest numbers of cores per node, challenges must be met as one moves embedded turbulence transport computations to the LCFs. These challenges span the domains of physics (ensuring the appropriate models for fluxes and sources), algorithms (implicit equation integration with noisy and stiff fluxes that are found from distributed computations), and computer science (hierarchical parallelism, load balancing). In practice, these challenges are all interdisciplinary. For example, when dealing with load balancing, it helps to understand how the computational components have varying computational intensities, which require a knowledge of the physics of turbulent transport. The following paragraphs discuss a few of these complexities.

Of fundamental importance is getting the physics correct; this is a consideration that remains at all levels, regardless of computational power. However, as one achieves greater fidelity through increased computational capability, it is no longer correct to leave out physics effects that could be left out when one is going after a first-cut answer at low fidelity. Thus, an important physics research area is to consider what new physics must be included in exascale computations.

Algorithms must be updated for exascale computations. Part of this task simply entails finding appropriate parallel algorithms that are able to make use of massively parallel computers. For example, the serial algorithms used in transport applications would need to be updated for use on parallel computers. In particular, in serial applications, users can use flux conservation and integrate from the magnetic axis to solve steady-state profiles. However, in parallel applications, where users want to compute all cross-surface fluxes simultaneously, an iterative solution is more appropriate. As another example, it has been recently shown that nested iterations can improve the convergence of the time advancement of the transport equations. The nested iterations require that a reduced grid be used at lower resolution. Thus, fewer turbulence computations are launched because there is one turbulence computation per grid cell. To continue to make use of the available processors, users should increase the number of processors utilized by each turbulence computation as the number of grid cells is reduced.

While moving to future exascale computing platforms could potentially enable direct coupling of turbulent fluxes, numerous other complex challenges will likely be encountered. For example, in addition to constructing turbulence codes to be compatible with exascale architectures, enabling software tools such as parallel solvers will need to be appropriately robust.

An example of a computer science problem that arises in integrated modeling is load balancing. In more traditional domain-decomposed parallelism, strategies for moving domain boundaries to obtain equal loads on all processors have been developed. The present problem, with multiple physics components spread across different processor sets, leads to the difficulty of reassigning the number of processors allocated to any component. This adds more difficulty because existing physics components are generally unable to recompose themselves on a different number of processors.

Potential Scientific Impacts and Outcomes

Success in this area will provide the fusion community with a high-fidelity, nearly first-principles predictive capability. Upon validation against the next large experiment (ITER), there will be greater confidence in determining the parameters of an eventual fusion reactor.

Transport Modeling with Embedded Global Turbulence

Summary of Research Direction

Evidence from experiments and computer simulations prove that turbulence and transport in one part of the plasma have an effect on all the rest of the plasma. In some cases, the turbulence itself is coupled across the plasma in a way that causes transient pulses to propagate much more rapidly than they would by diffusion. Thus, there is controversy over whether local turbulence calculations are sufficient, or whether global turbulence computations capable of linking meso- to micro-scale dynamics must be performed.

For example, global gyrokinetic simulations of turbulent transport in tokamaks from a particle-in-cell (PIC) code (Lee et al. 2008) have recently reported the observation of the formation of long wavelength (global) zonal flow (m=0, n=0) modes as well as the generation of global ion current during the nonlinear stage of ion temperature gradient (ITG) drift turbulence simulations. Without allowing for the presence of these global modes, the associated turbulence simulations failed to saturate. The simulations were carried out on the Jaguar (Cray XT3/4) computer at Oak Ridge National Laboratory (ORNL) using 2 billion particles with 32K processors using the GTC code (Lin98) for a TFTR-size tokamak. Most recently, simulations using the GTS code (Wang et al. 2006) for studying turbulent fluctuations driven by electron temperature gradient (ETG) drift modes have been conducted for NSTX plasmas showing the existence of radial streamers (Mazzucatto et al. 2009). These nonlocal properties of the microturbulence are in agreement with those reported earlier associated with turbulence spreading for ITG modes, as shown in Figure 2 (Lee et al. 2008) where the global properties are evident.

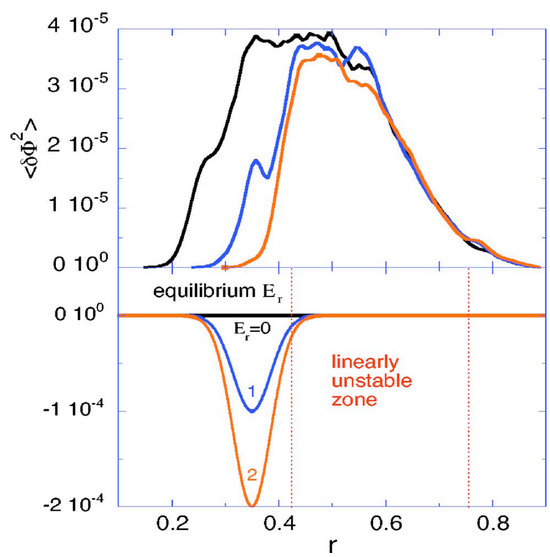


Figure 2. Example of nonlocal turbulence coupled across the plasma from a global gyrokinetic simulation. Image courtesy of Weixing Wang (Princeton Plasma Physics Laboratory).

Scientific Challenges

The PIC codes have excellent scaling on LCF computers such as the quad-core Jaguar at ORNL. As shown in Figure 3, 10 billion particles can be pushed on 100,000 cores per time step in just 1 second of the wall clock time using the GTS code. Therefore, based on the present model of gyrokinetic ions and adiabatic electrons for GTS, it would take half a day to simulate 1 millisecond of the ITER discharge using these resources; i.e., 10 billion particles on 100,000 processors. With more realistic physics models including electron inertia and electromagnetic effects, another order of magnitude of computing power will likely be needed to accomplish this type of simulation. The advantage of the global approach is that it allows radial interactions between different filaments (or flux tubes) of the simulated plasmas. Moreover, low (m, n) modes can be accurately simulated with a comprehensive electromagnetic global code that provides the natural coupling between microturbulence and MHD modes. To accomplish this goal, better numerical algorithms for PIC simulations are still needed along with the development of larger and faster supercomputers.

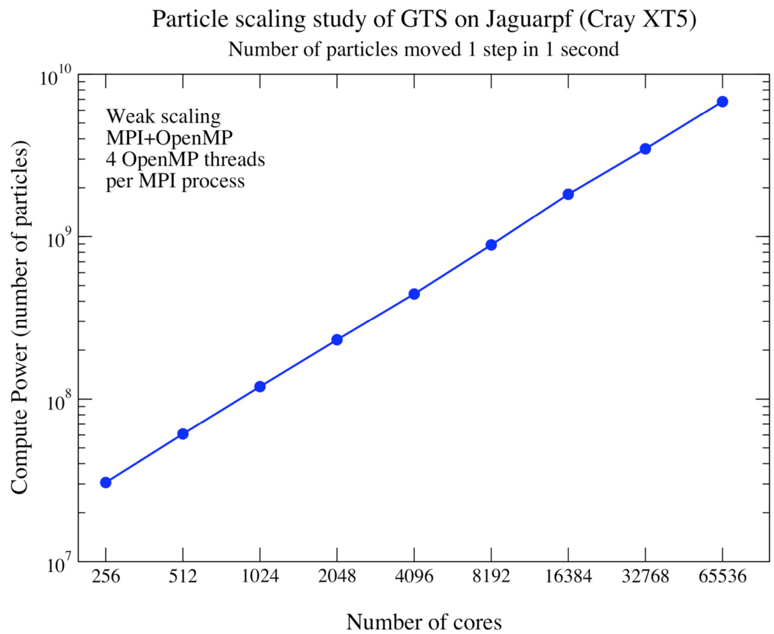


Figure 3. Scaling of a particle-in-cell code, GTS, on Jaguar. Image courtesy of Stephane Ethier (Princeton Plasma Physics Laboratory).

Potential Scientific Impacts and Outcomes

In the future, studies of this kind using extreme-scale computing capability will not only help accelerate progress in understanding the physics of microturbulence in tokamak experiments such as ITER, but can also be expected to help determine the conditions under which local approximations for microturbulence are valid. If carefully validated on present-day tokamaks, this kind of the study would provide important insights into the "local vs. global" nature of plasma microturbulence. Moreover, such investigations would also help support: 1) current experimental campaigns to gain the improved predictive understanding needed to better address ITER performance challenges; and 2) verification efforts involving systematic comparison with microturbulence codes based on different methodologies.

Coupling Disparate Regions of the Plasma

Summary of Research Direction

A basic requirement in fusion research is to reach core densities and temperatures sufficient to sustain fusion reactions, and yet keep plasma confined in a vacuum vessel that must maintain its structural integrity throughout the lifetime of the plasma and ultimately the lifetime of a fusion reactor. Tokamak experiments have achieved high performance by using magnetic field shaping to insulate the hot plasma from the material wall and thereby create a region of open field lines in which there is cold plasma. This region of cold plasma, plus the outer layer of the core, is known as the edge region. The three regions are thus the core region, the edge region, and the material wall, as illustrated in Figure 4. The problem of properly conducting coupled core-edge-wall transport simulations exemplifies the kind of multiphysics integration challenge faced by the fusion program. The core and scrape-off-layer (SOL) regions are very different in their spatial and temporal scales. Transport in the plasma core is dominated by turbulence with relatively short spatial scales. This can be represented in terms of surface fluxes for the basic moments (density, temperature, and momentum) and so is essentially a one-dimensional (radial) description. On the open field lines, which contact material walls, perpendicular transport competes with parallel transport so that, in its simplest description, edge transport is two-dimensional and essentially kinetic in nature. Because of these different spatial and temporal scales, computational modeling capabilities for the three regions have traditionally been developed and run separately.

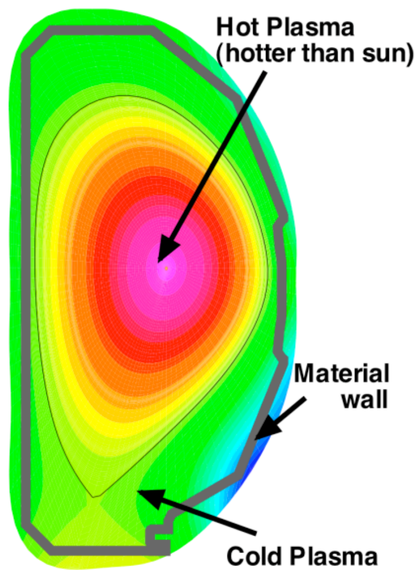


Figure 4. Three regions of a tokamak plasma shown in cross section. Image courtesy of Scott Kruger (Tech-X Corporation).

As discussed in this report, higher physics-fidelity simulations of individual regions in themselves will require exascale resources. In the core region, the gyrokinetic turbulence codes have demonstrated significant success in modeling the transport in a large number of tokamak experiments, and the recent integration of flux-tube simulations with transport codes has produced a new computational capability of enabling long time-scale simulations with high accuracy. The open-field lines and sharp gradients of the edge region offer new challenges for properly modeling kinetic dynamics. The computational fusion physics community is meeting this challenge by developing new kinetic codes for modeling the edge plasma region. As exemplified by the XGC1 code, this demands the utilization of large-scale computational resources. The material wall region offers arguably the most formidable computational challenges because the atomic scales are one of the fastest time scales in the system, and yet the retention of gases in the material wall is experimentally observed to depend on the time history of the discharges extending over hours, days, and weeks. Researchers focused on dealing with material wall issues are accordingly working on a range of models to handle these disparate scales.

Realistic simulations in each region will require extreme-scale resources. The integrated modeling of all three regions is limited in the fidelity of physics that can be studied. Even when the accuracy of the simulations in each individual region is significantly improved, integration of the different regions will require extreme-scale resources. As an example, consider a kinetic edge turbulence simulation that is running near the limit of available computational resources of 100,000 processor elements (PEs). If a scientist wishes to extend this simulation to include the nonlinear effects of core turbulent transport on edge turbulence, then having other 100,000 PEs available to perform the core simulation is not feasible at present. In this case, one approach to studying the sensitivity of including self-consistent transport simulations would be to use a simpler model for the core simulation. This simplified modeling approach is routinely chosen for addressing the material wall interactions where trivial models are typically adopted. The plasma-material wall interface (PMI) problem remains perhaps the greatest simulation challenge going forward. Because the ITER will have three different materials for its wall and divertor region, and there are concerns about tritium retention issues, improved PMI modeling is clearly important. While future access to extreme-scale computing resources can clearly be expected to enable improvements in the fidelity of the associated plasma simulations, many challenges will undoubtedly remain. This subject is a featured topic in the panel report titled, “Plasma-Material Interaction Science Challenges.”

Scientific Challenges

Integrating codes that model disparate temporal and spatial scales with different discretizations demands significant mathematical research advances. The separation of spatial scales will generally require implicit time advances, and, for coupled components, this involves the development of new solvers that can work with an implicit time advance. Also, understanding the convergence and accuracy of complex simulations is an active research topic.

As mentioned in the preceding discussions, flexibility in choosing the appropriate models is important in using the appropriate computational resources for the given physics simulation. When integrating different components that cover different physical effects, the simulations can become very load-unbalanced. A preliminary analysis for understanding this problem was undertaken by the FACETS project and the results are shown in Figure 5. As expected, the codes with greater physics fidelity (BOUT, GYRO) are orders of magnitude slower than the simpler codes (UEDGE, GLF23) addressing the same physics questions. One of the positive results of this analysis is that collections of components can be well balanced; e.g., NUBEAM and UEDGE, or GYRO and BOUT. As a greater collection of data becomes available for this type of analysis of other fusion codes, it is expected other collections of components will be well balanced. Unfortunately, it is likely some physics problems will best be served by matching unbalanced components. Understanding ways of ameliorating this situation requires collaborative research involving physicists, applied mathematicians, and computer scientists.

Potential Scientific Impacts and Outcomes

The grand challenge of proper integration of fusion codes offers great opportunity for engagement of computing at the extreme scale to enable producing simulations capable of delivering an “extreme level” of improved physics fidelity. By integrating the knowledge contained within different codes into a single simulation, the computational fusion community has a path toward matching the complexity found in the experimental results. The major obstacle to meaningful and timely progress in code integration is that there are many challenges to simultaneously overcome—a formidable overall challenge that is much more difficult than those of standalone codes, as befits the complexity of coupled simulations. Computational resources at the petascale and extending to the exascale and beyond, together with the associated advances in enabling software, are expected to be essential for producing experimentally validated, integrated predictive physics capabilities needed to accelerate progress toward the delivery of magnetic fusion energy.

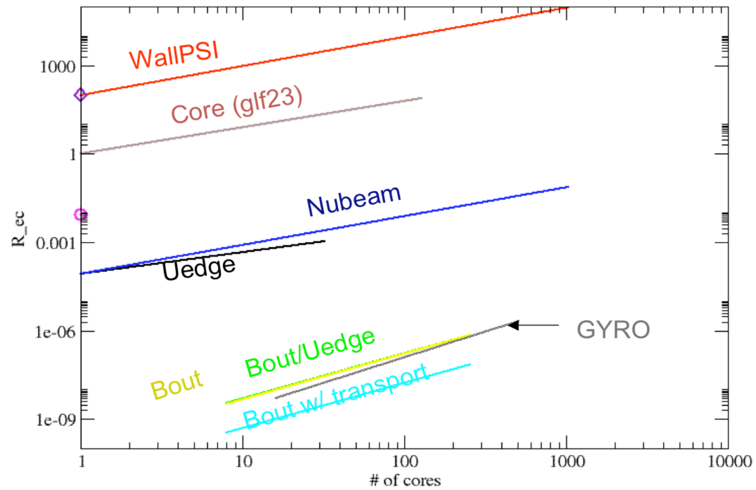


Figure 5. Ratios of computational time to experimental time, *R_ec*, for components over their ranges of parallelizability. From the Framework Application for Core-Edge Transport Simulations Team Meeting (2007). Image courtesy of John Cary (University of Colorado and Tech-X Corporation).

Macroscopic Stability Control Using Radio Frequency Power

Summary of Research Direction

Tokamak devices in the burning plasma regime are subject to a wide variety of instabilities—the strongest of which are the large-scale magnetohydrodynamc instabilities (MHD modes) that can produce 1) a localized loss of plasma confinement; 2) an extensive redistribution of plasma energy and magnetic field; and 3) in extreme cases, the complete termination of the discharge accompanied by a possibly damaging dump of plasma and magnetic energy to the vessel wall. Much is known about the physics of these instabilities from experiments, analytic theory, and sophisticated computer codes. Over the years, experimental techniques have been developed to optimize tokamak operation with respect to the important instabilities through careful control of the magnetics, heating, current drive, and fueling systems. Often this optimization is achieved by avoiding plasma states (distributions of pressure, current, magnetic field, and velocity) that are subject to instability. Sometimes specialized systems are used, such as feedback controls or special magnetic coils, which can directly suppress the instability in plasma states that would otherwise be unstable. Occasionally, to lessen the impact of a larger disturbance that might be built up in the plasma, it is desirable to actually produce instability to reduce a local reservoir of free energy, such as a pressure gradient. In any case, the ability to accurately model the interaction between the tokamak control actuators and plasma instabilities is key to the following: 1) understanding the coupled physics of burning plasmas; 2) optimizing the device's control systems; 3) planning optimal operating strategies; and 4) setting realistic requirements for new fusion devices such as the future Demonstration Reactor (DEMO).

The study of the interaction between plasma control systems and stability by bringing together the most advanced plasma modeling and stability computer codes is just beginning. For a long time, 1-1/2-dimensional transport codes have used reduced (semi-empirical) stability models to indicate when the simulated plasma state might be expected to be unstable. In many cases, similar kinds of reduced models for representing the effect that nonlinear growth of the instability might have on the plasma state have been included in the 1-1/2-dimensional simulations. An example of this is the use of the Porcelli model to estimate when a sawtooth collapse might occur, followed by a somewhat “ad hoc” redistribution of plasma energy, current, and energetic particle populations to model the plasma state after the magnetic reconnection event.

One of the Scientific Discovery through Advanced Computing (SciDAC) Proto-Fusion Simulation Program projects has the task of bringing massively parallel transport and source modeling codes together with massively parallel MHD codes to study these issues. In this pilot SciDAC project (supported jointly by DOE’s Office of Science, DOE’s Office of FES, and DOE’s Office of ASCR), a distinction is made between two different time scales for the development of an MHD instability relative to the time scale for control systems to modify the plasma state. In particular, the following considerations are addressed:

- For fast MHD phenomena, the unstable plasma motions are much more rapid than the rate of energy deposition, current or pressure profile evolution, and changes in energetic particle populations driven by external coils, RF waves, or other sources. To control fast MHD phenomena, the plasma control systems act over a longer time scale to drive the slow evolution of the plasma equilibrium and profiles, effectively setting the initial conditions for the fast MHD event. An example of such a phenomenon is the crash phase of the sawtooth oscillation.
- In slow MHD phenomena, the energy deposition, currents, or flows driven by RF waves or other sources and transport phenomena operate on time scales that directly influence the dynamics of the unstable motion. In this case, the perturbations of the plasma state by the instability influence the energy, currents, or flows driven by sources and the time evolution of these due to transport. In particular, transport driven by plasma microturbulence was discussed in some detail earlier in this report. An example of kinetic dynamics influencing slow MHD phenomena is the presence of the neoclassical tearing mode (NTM) and its stabilization by high-power electromagnetic waves in the electron cyclotron frequency range. This is an important consideration for the performance of burning plasma experiments such as the ITER where NTMs are predicted to be particularly troublesome. Treatment of these phenomena requires extension of the MHD equations by higher-order “closure relations,” which can provide a way of computing the plasma pressure and current response to the MHD fields, including the effects of RF and transport, which are essentially kinetic processes.

Scientific Challenges

At present, the SWIM project does not deal with the development of new codes that are designed 1) specifically for parallelism at the petascale or exascale; or 2) to interoperate as parts of an integrated system. Instead, the components are being implemented using existing, well-tested codes that are wrapped with adapter code to provide generic interfaces, allowing them to interoperate within an integrating framework. The limitations of this approach are those inherent in the component physics codes themselves, as well as those imposed by the tightness of coupling required between the multiphysics processes involved. While the codes are constantly being improved for efficiency and parallelism, none is designed to take advantage of the multilevels of memory and processing featured by the new architectures. Most do not scale beyond a few tens of thousands of processors, and some important component codes are limited to much smaller processor counts, and consequently, require very long running times. As a result, many calculations that are important for larger tokamaks or for the ITER are not feasible at this time.

As a specific example of the type of simulation that could possibly be enabled with access to exascale computational resources, the edge-localized modes (ELMs)—which are almost always present when the plasma is in a high-confinement state known as H-mode—would be a worthy objective. The high-confinement state is the result of a narrow radial region of extremely low transport, called a transport barrier, which can spontaneously appear at the plasma edge. This low transport produces a very steep rise in plasma pressure within the barrier, referred to as a pedestal. The ELM instability is driven by the free energy associated with the steep pressure and current gradients that develop at the plasma edge when the plasma is in H-mode. These instabilities play a pivotal role in that they limit the height of the pressure pedestal, which strongly affects the fusion reaction yield. Conversely, ELMs also tend to control the rise of plasma density and impurity density that—if uncontrolled—could result in a degradation of the fusion yield or termination of the discharge. In addition, the rapid growth of these instabilities results in bursts of plasma being expelled to the exterior power and particle-handling systems. If the ELM frequency is sufficiently high and the amplitudes of the bursts are sufficiently small, there are no apparent problems. However, if the bursts are infrequent so that large quantities of plasma are expelled all at once, ELMs can be a serious problem for the divertor systems. The phenomenology of ELMs is quite complicated. There are several types of ELMs, some of which are beneficial to certain operational strategies, while others are deleterious to both plasma performance and to the power and particle-handling systems. There is much experience in using the plasma control systems to affect the onset of ELMs and their nature once they appear. For the success of fusion, it is critical to understand the physics of these modes as they are coupled to both the core and SOL plasmas, and to develop methods to control them for optimum plasma performance and safety. To perform simulations at the sufficient level of detail will require a significant scale-up in computational capability.

Potential Scientific Impacts and Outcomes

In the future, the separation of time scales between fast MHD events and transport phenomena can continue to be exploited by maintaining the modeling of the slow equilibrium evolution punctuated by independent treatment of the fast, nonlinear MHD evolution. Computers at the exascale would permit far more realistic modeling of the dynamics at both of these scales. During the slow scale evolution, it would be possible to include a simulation of the turbulent transport at levels much truer to first-principles physics than can presently be carried out, including the proper treatment of electron dynamics, momentum transport and the effect of rotation on equilibrium and transport. Ideally, the calculation of particle and energy sources, such as RF, neutral beams, fusion reactions, and fueling should be unified so that a single self-consistent model is used to evolve the velocity space distributions of the plasma species. Far more accurate models of the edge plasma would be possible using the powerful new computers to carry out kinetic modeling at the high dimensionality required. It would also be possible to calculate the interaction between the burning plasma core and the SOL. At this scale of computation, it would be possible to account for the fact that tokamak plasmas are not really axisymmetric. Often after an MHD event, the plasma is left in a non-axisymmetric state containing helical filaments called magnetic islands. Furthermore, the plasma may transition to a non-axisymmetric state for a considerable period of time during the buildup to a fast MHD event. This is another example of the overlap of transport and stability time scales. In addition, the edge region of the plasma is non-axisymmetric due to the following: 1) mechanical structures at the plasma edge; and 2) toroidally localized heating and fueling effects. These can be quite important for stability control—with two prominent examples being when 1) non-axisymmetric coils are used for feedback control of resistive wall modes; and 2) pellet "pacing" or triggering of ELMs is used to control the build-up of the amplitude of the ELM heat pulse in the ITER.

Edge Transport With Recoverable Macroscopic Dynamics

The SOL in a tokamak plasma is a complex nonlinear multiscale system characterized by dynamic interactions involving plasma profiles, flows, microturbulence, MHD-type macroscopic events, neutral particles, atomic physics, and the material wall. The plasma includes multiple species with strong impurity radiation. At present, gyro-kinetic simulations, which are based on a "full-*f*" or full-particle-distribution-function approach, can best describe the edge plasma dynamics but without MHD events. The geometry adds another major element of complexity to this problem by requiring inclusion of a magnetic separatrix and material wall. A magnetic separatrix is a singular surface for the conventional (and convenient) flux-following coordinate system that is used in the core plasma simulation codes. The material wall absorbs plasma particles and heat. The absorbed plasma particles are recycled back to the plasma in the form of neutral particles, becoming a volumetric source of electrons and ions for the plasma. The full-*f* gyrokinetic scheme, which is capable of describing the edge physics, makes the simulation of the edge plasma much more compute-intensive than the conserved core plasma, in which the simplified perturbed distribution function (delta-*f*) method is usually used. A full-*f* gyrokinetic description requires several hundred times more marker particles than the delta-*f* gyrokinetic description.

At the expense of a major increase in computational resources, a full-*f* kinetic simulation can compute the multiscale plasma dynamics in the open magnetic field region with sources and sinks. It can encompass the meso- and micro-scale dynamics associated with the plasma profiles, radial electric field, flows, microturbulence, and neutral particles. This is performed without invoking scale-separation while using atomic and material interaction data. However, the macroscopic MHD events need to be simulated separately and integrated into the kinetic simulation by means of computer science tools. At the present time, there are two full-*f* gyrokinetic edge code development efforts in the U.S. fusion program. The ESL (Edge Simulation Laboratory) project at Lawrence Livermore National Laboratory is developing a five-dimensional (three-dimensional in real space and two-dimensional in velocity space) full-*f* edge gyrokinetic code TEMPEST-ESL [Xu07] using a PDE method (continuum method), while the CPES project (the SciDAC Proto-FSP Center for Plasma Edge Simulation) is developing a five-dimensional full-*f* edge gyrokinetic code XGC1 using a PIC approach.

The XGC1 PIC code has produced multiscale turbulence/neoclassical solutions in five-dimensional space for realistic edge geometry (Chang et al. 2009) using the high-performance computer system—Jaguar—at ORNL's National Center for Computational Sciences.[1] A particle, momentum, and energy conserving collision operator is used to ensure self-consistent results for the neoclassical/turbulence dynamics calculated. For example, Figure 6 shows electrostatic potential fluctuation results from ITG turbulence simulations that take into account the realistic DIII-D edge geometry. A typical full-*f* electrostatic edge-turbulence simulation for 3 milliseconds of experimental time takes about 10 hours on 29,952 Jaguar processor cores using 3.2 billion marker particles.

[1] See www.nccs.gov.

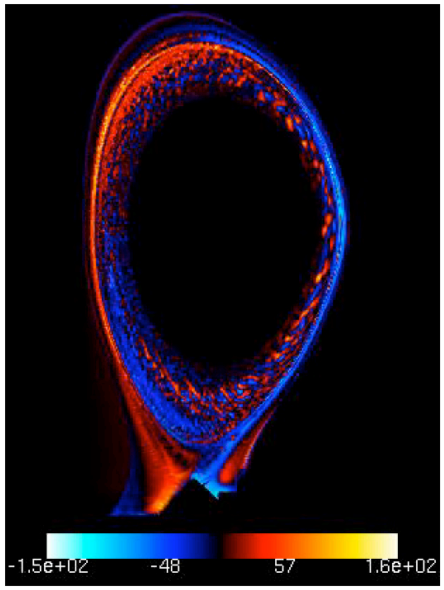


Figure 6. Ion temperature gradient turbulence at the edge of a DIII-D plasma computed using the XGC1 kinetic code. Image courtesy of Choong-Seock Chang (New York University).

Current estimates for extreme-scale computing needs of this class of full-*f* PIC codes are based on the experience of XGC1 on the Jaguar Cray XT5. The new physics capability of a PIC code on a high-performance computing (HPC) system is measured more accurately by weak scaling on the number of processor cores in proportion to the grid size. The speed-up achieved in a run is usually depicted by strong scaling on a fixed grid. XGC1 shows excellent scaling behavior on Jaguar up to the maximum number of available processor cores. An example of this scaling property is shown in Figure 7.

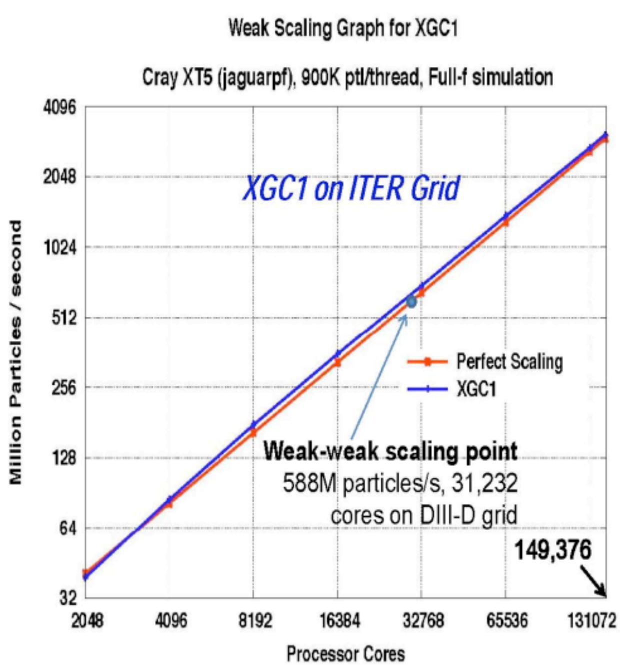


Figure 7. Scaling of XGC1 particle-in-cell kinetic code up to the maximum available number of processor cores in the Jaguar computer. Image courtesy of Choong-Seock Chang (New York University).

The next targeted problem size for electrostatic turbulence studies using XGC1 is the ITER-scale simulation of the ITG. This is expected to use 67.5B particles on 149,760 Jaguar XT5 processor cores for approximately 7 days. To reduce the simulation wall-clock to 1 day, it is estimated about 10 petaflops of HPC capability will be needed.

Basic Scientific and Computational Challenges

The XGC1 edge turbulence transport studies performed to date have been carried out in the electrostatic regime with adiabatic electrons. To improve the physics fidelity of XGC1, it will be necessary for XGC1 to develop an edge electromagnetic turbulence simulation capability that properly includes kinetic electron dynamics. It is estimated that such simulations require extreme-scale computing support. Specifically, the improved simulation capabilities targeted would require several teraflops computing power for a DIII-D size problem and several tens of teraflops for simulating the ITER plasma edge region.

As mentioned earlier, the macroscopic edge-localized MHD-type events (ELMs) presently cannot be addressed with the current gyrokinetic capability. In the CPES project, the simulations of ELM crashes are performed via an integrated system of an MHD code and an edge kinetic code developed using modern computer science methodology (Chang et al. 2008). Specifically, the CPES project has developed an automated integrated simulation framework called “EFFIS” (End-to-End Framework for Fusion Integrated Simulation) for this as well as other purposes (Cummings et al. 2008). Figure 8 shows typical results from pedestal build-up kinetic simulation using the EFFIS framework.

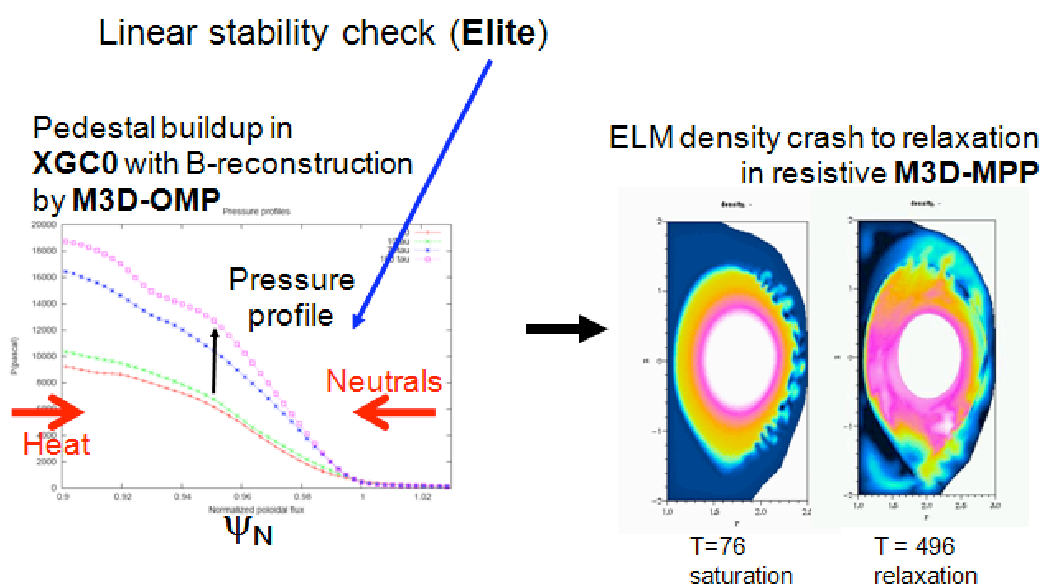


Figure 8. Pedestal buildup followed by an ELM crash computed using the XGC0 kinetic code, the Elite linear ideal MHD code, and the M3D nonlinear extended MHD code. Image courtesy of Choong-Seock Chang (New York University).

This system allows coupling with two distinct MHD codes: the ELITE code for carrying out edge profile (pedestal) stability analysis and the M3D code for tracking the nonlinear edge-localized mode crash of the pedestal. The computing requirement for the MHD codes used for this MHD study of edge-localized dynamics is quite small compared to that of the edge gyrokinetic code XGC1. The integrated transport-MHD modeling in the future will require at least 20 simulations of XGC1 for turbulence transport evaluations during a kinetic-MHD cycle study. To do so will likely require HPC resources at the exascale – especially for addressing the ITER physics associated with electromagnetic turbulence.

Another scientific and computational challenge faced by edge simulation studies involves predicting the influence of edge dynamics on confinement in the plasma core. First-principles-based edge-core transport simulations that are applicable to the whole tokamak volume will need to be developed before a higher physics fidelity core-edge coupled modeling capability can be produced. It has been discussed that turbulence at the plasma edge might actually spread into the core region to generate a new globally self-organized profile at a speed much faster than the radial heat transport rate. Nonlocal effects on core confinement could be especially relevant if recoverable macroscopic edge-localized mode events were to occur —whether caused by MHD instabilities or kinetic limit cycle events. With the availability of significantly more powerful HPC resource, more realistic modeling of nonlocal turbulence and profile effects can be accomplished in a straightforward way by just extending the inner boundary of an edge simulation all the way to the magnetic axis. Current HPC capability allows XGC1 to perform simulations of ITG-driven turbulence together with neoclassical physics in a whole-volume DIII-D plasma within 1 day using a total of 13.5B particles on 119,808 Jaguar cores for 20 hours (Chang et al. 2009).

As discussed at the March 2009 workshop titled, "Scientific Grand Challenges in Fusion Energy Sciences and the Role of Computing at the Extreme Scale," there are challenging applied math and computer science problems encountered in enabling current edge simulation capabilities to take advantage of HPC resources at the exascale for the purpose of accelerating progress in addressing complex physics integration issues. Problems that have already been encountered and solved to facilitate current usage of petascale hardware (e.g., quad-core XT5 processors) (Adams et al. 2009) have included solutions such as the replacement of MPI with a "hybrid" MPI-Open-MP approach. Major obstacles on the path to exascale that face global gyrokinetic PIC codes and extended MHD codes include serious issues such as load balancing and geometry hashing in an unstructured triangular mesh. Thus, while using the direct-matrix inversion solver method is costly at present, it might become a viably economical way to proceed in dealing with the parallelism expected to be encountered at the exascale.

The current generation of MHD codes used in the study of the recoverable macroscopic edge-localized instabilities will also obviously need major performance and scalability improvements to be able to make use of extreme-scale HPC resources. Parallelization at such a massive scale will be an unavoidable element that will likely require development of fully implicit algorithms, as well as innovative newer schemes. If successful in dealing with such formidable challenges, MHD codes will in the future be able to deal with realistic plasma resistivity and more complete two-fluid effects.

Potential Scientific Impacts and Outcomes

Understanding the edge pedestal dynamics and its impact on fusion performance in the plasma core and on material wall damage is one of the most critical issues in the toroidal fusion program. Empirical extrapolations from present-day experimental results indicate that a high-edge pedestal will likely be a necessary requirement for ITER operation, as well as for helping to ensure future efficient and economical fusion reactor performance. However, it is also evident that such a high edge pedestal can easily trigger a macroscopic collapse of the pedestal plasma, bombarding the material wall with the plasma energy. Overall, a reliable predictive simulation capability that can 1) successfully deal with the multiscale edge transport physics; 2) include the recoverable macroscopic edge-localized instabilities; and 3) allow integration with the multiscale transport physics in the plasma core would be a critically valuable new tool for addressing issues of the ITER and of the fusion reactor program in the future.

Performance Optimization of Burning Plasmas

Summary of Research Direction

Optimizing the performance of burning plasma discharges requires accurate computations for the evolution of all of the plasma profiles, including electron and ion temperatures, angular momentum, particle densities, current density, fast ions, and impurities. Many simulations must be performed to optimize the entire plasma discharge scenario from startup through plasma burn to shutdown. The simulation codes used for performance optimization must include all of the relevant physical processes, including sources, sinks, transport, effects of large-scale instabilities, and the interaction between the plasma and the rest of the tokamak (see Figure 9). Because the ITER will be expensive to operate, careful scenario modeling and planning must be conducted before each discharge, followed by detailed computational analysis of the experimental data after each discharge. Performance optimization roles include the following: 1) maximizing the fusion power produced by ITER discharges, and 2) reliably predicting and avoiding the conditions that are likely to produce damaging plasma disruptions or transient power excursions capable of destroying parts of the plasma-facing wall or divertor. Scenario modeling is used to design and set up the feedback systems that are used to control plasma discharges. Finally, integrated modeling is used to test and validate the theoretical models that embody the sum of our knowledge of fusion plasmas.

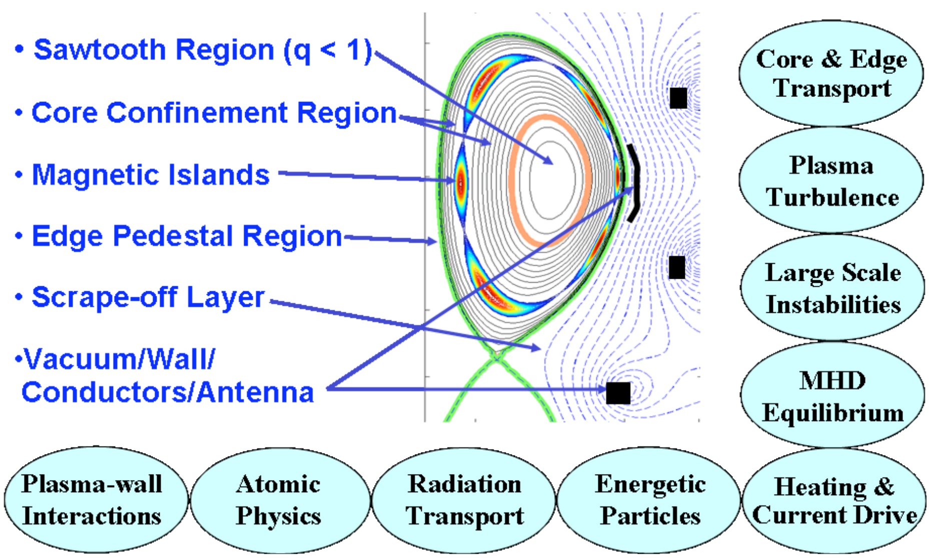


Figure 9. Illustration of the interacting physical processes within a tokamak discharge. Image courtesy of Arnold Kritz (Lehigh University) and David Keyes (Columbia University). Source: DOE (2008).

Scientific Challenges

Whole-device integrated modeling, which is required for performance optimization, involves computer simulations that combine all of the components described in the previous sections. Furthermore, many of these simulations must be run for a sufficiently long time to span different stages of the plasma discharge, from startup through the transitions that lead to fusion burn, and finally to plasma shutdown. Hence, integrated modeling involves a heterogeneous mix of different computational components representing very different physical phenomena that occur on different time and length scales. The load balancing and strongly nonlinear interactions among these heterogeneous components are further complicated by the fact that some of the physical phenomena are episodic while others are continuous. Periodic instabilities, such as sawtooth oscillations or ELM cycles, produce transient spikes that strongly influence the sources and transport of heat, momentum, current drive, and plasma particles. The load balancing on a massively parallel computer has to respond dynamically to rapidly changing events during each sawtooth crash or ELM crash, which have the effect of rapidly mixing different parts of the plasma.

Potential Scientific Impacts and Outcomes

Performance optimization entails performing many whole-device integrated modeling simulations to explore a wide range of plasma conditions and the parametric dependence of the associated results. Because hundreds or thousands of simulations must be performed, computational speed and efficiency are essential. Results must be obtained in a timely manner to influence the decisions that have to be made concerning the construction and operation of large experiments such as the ITER. High-performance computer simulations are becoming increasingly important for scenario modeling and for performance optimization, even as the simulations push against the limits of computer capability.

CONCLUSIONS

A fusion reactor is far too complex to be simulated with first-principles computations alone, even with the availability of exascale high-performance computer resources. However, such powerful hardware coupled with the enabling software, will make it possible to use the most advanced models to deliver experimentally validated predictive simulation results with much higher physics fidelity. Still, there are many challenges associated with making effective use of exascale resources. These issues range from advancing beyond the present state where multiple two-way couplings are being used to a fully integrated model. Meeting these challenges will require application of the best available physics, applied math, and computer science methods to accelerate progress. This goal can be effectively accomplished through productive interdisciplinary collaborative alliances.

PLASMA-MATERIAL INTERACTION SCIENCE CHALLENGES

Co-Leads: Jeffrey N. Brooks, Purdue University
Steven Zinkle, Oak Ridge National Laboratory

Panel Members: Sergei Dudarev, Culham Centre for Fusion Energy; David Schultz, Oak Ridge National Laboratory; Daren Stotler, Princeton Plasma Physics Laboratory; Arthur Voter, Los Alamos National Laboratory; and Brian Wirth, University of California, Berkeley

CURRENT STATUS

As one of the most critical scientific issues for fusion power, plasma-material interaction (PMI) affects the following:

1. Lifetime of plasma-facing components (PFCs) as a result of sputter and transient erosion
2. Plasma contamination by eroded material
3. Tritium codeposition in eroded and redeposited material
4. Operating limits on core plasma (beta, confinement, edge temperature/density, duty factor, etc.) as a result of the above-listed factors.

A related critical issue is PFC bulk material performance and optimization. Gaining a thorough understanding of and the predictive capabilities in this critical area will require simultaneously addressing rich and diverse physics occurring over a wide range of length (angstroms to meters) and time (femtoseconds to minutes) scales (see Figure 10).

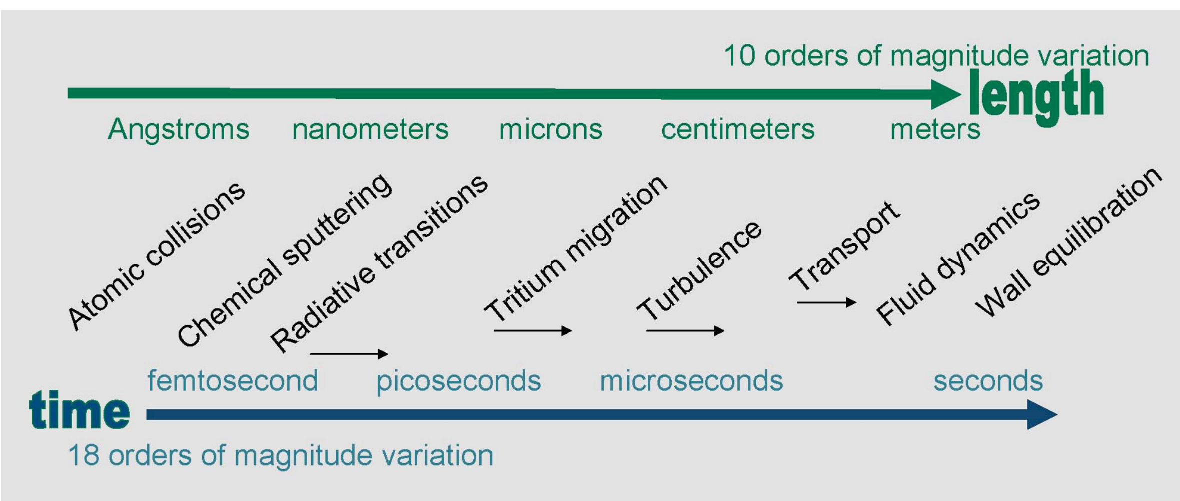


Figure 10. Relevant length and time scales required for integrated edge-wall modeling. Modeling is only possible with extreme-scale computing. Image courtesy of David Shultz (Oak Ridge National Laboratory).

This will necessitate further development of detailed physics models and computational strategies at each of these scales, as well as revolutionary algorithms and methods to intimately couple them in a way that can be robustly validated. While current research confined to each of these scales—or pioneering methods to couple two or more of them—already pushes the state of the art in technique and available computational power, simulations spanning multiple scales needed for the ITER, Demonstration Reactor, and future devices will require unprecedented extreme-scale computing platforms and integrated physics, and computer science advances.

Plasma-material interactions primarily involve the interaction of the edge and scrape-off layer (SOL) plasma with the following:

- Deeper layers for tritium permeation
- Deeper layers generally extending to the coolant for heat transport and material response because of plasma transients (edge-localized modes [ELMs], disruptions, etc.)
- First approximately 10 to 1000 nanometers of surface for particle surface interactions (primarily sputtering and tritium codeposition) but also dust formation and/or microstructure feature formation. The plasma particles are ions and atoms of D, tungsten, helium, impurities from surface material (i.e., carbon, beryllium, tungsten, etc.), and oxygen and other trace impurities (e.g., neon).

The focus of the PMI panel was on edge and SOL plasma properties, as well as material properties connected with plasma interactions involving the surfaces of the plasma-facing material. Bulk material issues, such as neutron damage, are important in their own right and may affect plasma interaction properties, such as sputtering, tritium permeation and/or retention, dust formation, and flaking. The bulk material performance and optimization also affect thermal and power management of the fusion reactor and control the operating and/or replacement lifetime of the vacuum vessel. Surface and bulk issues are included in the assessment of scientific issues and priority research direction needs. The discussion of the PMI issue is divided into three broad categories: 1) understanding the edge and SOL plasma; 2) understanding the PMI interactions; and 3) understanding bulk-material response.

BASIC SCIENCE CHALLENGES AND RESEARCH NEEDS

Plasma Edge and Scrape-Off Layer

The edge plasma is defined as the region of a confinement device inside the last closed flux surface that is dominated by steep plasma parameter gradients across flux surfaces and significant variations along them. For example, in H-mode tokamak discharges, the edge region extends from the top of the pedestal to the last closed flux surface or separatrix in diverted configurations. The SOL plasma then extends from the last closed flux surface to the material boundaries. Locating the inner boundary of the edge plasma at the top of the pedestal simplifies integration with models of the core plasma by rendering the boundary conditions one dimensional (assuming axisymmetry) and by minimizing neutral penetration into the core. At the same time, the interaction of pedestal plasma and magnetohydrodynamics (MHD) phenomena with the SOL and material surfaces can be considered simultaneously in the edge and SOL model. Fundamental progress requires a thorough understanding of the interactions of plasmas, neutral atoms, molecules, materials, and photons as well as the ability to self-consistently treat all of the associated tightly coupled processes encompassing phenomena varying over 18 orders of magnitude in time and 10 orders of magnitude in length (Figure 10). Table 2 summarizes existing and required capabilities for plasma edge and SOL codes.

Table 2. Comparison of Typical Present Edge and SOL Modeling Capabilities with Needs for a Predictive Model

Current Capability	Needed Capability
Plasma and neutral particles tightly coupled; coupling to photons and materials less developed	Plasma, neutral particles, photons, and materials tightly coupled
Fluid plasma turbulence and transport; low dimension, local kinetic transport	Full SOL two-dimensional and three-dimensional kinetic, turbulence, and transport
Axisymmetric geometry bounded by well-defined flux surfaces	Toroidal variation, localized ergodic volumes; extended to wall
Time averaged	Time-resolved behavior of turbulent structures
Immutable, idealized surfaces	Evolving material temperature, composition, structure, and hydrogen content; microscopic and macroscopic changes

The key need for edge and SOL plasma models is to characterize the heat and particle fluxes to the material surfaces and propagate the response from those surfaces to the core plasma. This response is critical to the overall confinement performance of the device because the edge plasma sets the boundary conditions for the core. For example, ITER predictions made with existing core plasma transport models show the fusion gain is strongly correlated with plasma parameters at the top of the pedestal. The edge plasma is also crucial in determining the viability of future devices because it governs the distribution of plasma and particle exhaust to the surrounding material surfaces. The associated heating and erosion of wall materials can generate impurities that migrate into the core plasma and reduce fusion gain. Furthermore, if the energy and particle fluxes to the wall are greater than anticipated, the operational lifetime of the plasma-facing surfaces will be correspondingly shortened. Just as serious are the safety concerns associated with the potential buildup of the tritium inventory in plasma-facing surfaces. In all of these instances, present understanding is insufficient to allow scientists to make definitive statements about the edge plasma in future devices.

The most widely used software models for edge and SOL plasmas (such as SOLPS [Schneider et al. 2006] and UEDGE [Rognlien et al. 1994]) self-consistently simulate plasma and neutral transport. Models of photon transport in tokamak diverters exist (Reiter et al. 2007), but such calculations are neither common nor completely self-consistent. Coupled calculations of the SOL plasma and material behavior are even less well developed (Coster et al. 2007). Instead, as has been the practice since the earliest SOL plasma simulations, materials are typically modeled as immutable, idealized surfaces.

The aforementioned plasma transport models are based on Braginskii plasma fluid equations for the plasma density, flow velocity, and temperature, with classical transport along field lines and anomalous transport across flux surfaces. The diffusion coefficients and convective velocities associated with the latter must be specified on input (usually calibrated against experimental data), limiting the utility of these simulations and making predictions problematic. Kinetic modifications to classical parallel transport are emulated by flux limiters, although these treatments have recently been called into question (Tskhakaya et al. 2009). These fluid plasma transport models are capable of simulating all relevant species in the plasma (Bonnin et al. 2009), including multiple-charge state impurity ions, and they can process both single and double null tokamak geometries. However, the outer radial boundary of these simulations is usually taken to be the outermost contiguous flux surface and therefore does not conform to the shape of the vacuum vessel.

Multiple-fluid plasma software codes for simulating turbulence in the plasma edge and SOL now exist. The three-dimensional BOUT electromagnetic turbulence code is the most sophisticated of these and is able to work with single null and double null diverted configurations. BOUT and UEDGE were successfully coupled to obtain a self-consistent, turbulent transport simulation (Rognlien et al. 2005), but this capability is not widely used.

Most kinetic models of SOL plasma transport are of low dimensionality (e.g., one-dimensional in space) or limited in spatial extent (Tskhakaya et al. 2007). However, three-dimensional (Monte Carlo) kinetic models (Brooks 2002; Stangeby and Elder 1992) of trace impurity transport in the plasma edge and SOL have been used productively for more than a decade.

SOL transport is now known to be dominated by intermittent, order unity turbulence (Umansky et al. 1999), and thus is not well characterized by radial diffusion. Consequently, using fluid plasma transport models to explain heat and particle fluxes in existing tokamaks is challenging (Pigarov et al. 2007). Because the transport coefficients are calibrated empirically, their extrapolation to the ITER is problematic. An equally important shortcoming of these simulations is their inability to replicate observed Mach flows in the tokamak SOL (Chankin et al. 2007). Properly simulating these phenomena will likely require the plasma transport model to be coupled in some way to a first-principles, electromagnetic plasma turbulence calculation (Scott 2007). The roughly thousand-fold difference in time scales (turbulence: microseconds; transport: milliseconds) prohibits a brute-force computational approach in which the radial fluxes at each step of the transport code are directly determined by a turbulence model.

A thorough understanding of the edge and SOL plasmas must also account for a variety of kinetic effects. In addition to the kinetic effects on parallel electron transport noted above, others are associated with steep radial gradients, large ion orbits, and non-Maxwellian distribution functions caused by low collisionality. The kinetic character of neutral species caused by recycling at material surfaces, uncollided dissociation products, and long mean-free paths have been acknowledged for decades and are accounted for in Monte Carlo simulations. This kinetic treatment must extend to the surrounding material surfaces so that a physically realistic boundary condition can be established there. Although most tokamak simulations for the foreseeable future will be axisymmetric, the capability for incorporating three-dimensional effects and chaotic magnetic fields, especially those imposed by external magnetic field coils, is highly desirable.

The heat and particle fluxes to material surfaces resulting from transient phenomena, such as ELMs and disruptions, by definition vary rapidly over time. However, this is also the case during quiescent operation because of the intermittent character of edge plasma turbulence. Insofar as the response of the material surfaces is a nonlinear function of these fluxes, the long-term behavior of the interactions can only be reproduced by integrating over the time-resolved response. A detailed, first-principles kinetic plasma turbulence simulation will provide the spatial and temporal resolution required for this task and is therefore viewed as the ultimate development objective in the area of edge and SOL plasma transport.

Other specific areas that are currently under investigation are detached plasmas and plasma behavior in the presence of low recycling walls. Simulating the former requires self-consistent treatment of plasma and neutral (and wall) behavior and of photons resulting from recombination that are trapped in the high-density plasma. The modeling to date has not been completely satisfactory (Wischmeier et al. 2009); a more thorough understanding may lead to improved modes of diverter operation for the ITER.

The improved performance of tokamaks that operate with low recycling walls (e.g., using lithium coatings [Mansfield et al. 2001]) is not understood but is of significant interest. Explaining this phenomena will require a materials model that is capable of replicating the low recycling behavior and a core plasma transport model that responds with higher confinement and/or improved stability. The edge and SOL plasma transport model must be able to mediate the exchange of information between the material and core plasma models.

Plasma Interaction with Material Surfaces

Fusion device PFCs consist of the major elements of the first wall, diverter, and additional components such as a startup limiter, diagnostic ports/modules, and radio frequency antennas. The PFC surfaces are subject to intense steady-state and transient particle and heat loading. Impinging particles are D-T and helium ions and neutral particles from edge plasma transport, including surface recycling and atomic and molecular processes, impurities from surface material erosion, and trace atoms and/or ions, such as background oxygen or, for example, neon added as a plasma radiative material. Steady-state surface heat loading is due to convected, conducted, and radiated power from the various plasma regions (edge, SOL, and core). Transient power loading arises from ELMs, plasma disruptions, vertical displacement events, and/or runaway electrons.

Key concerns for plasma-surface interactions are the impacts on the lifetime of the PFC surfaces from sputtering and other particle-induced changes, transient erosion, core plasma contamination by sputtered and/or transported impurities and dust formation and/or transport, and tritium codeposition in redeposited material.

PFC performance is a critical challenge for fusion energy. It may limit the permissible plasma operation in the ITER and appears to present major design challenges for the DEMO device. PFC response also plays a major role in the performance of present-day tokamaks. This is a scientific area that suffers from a major lack of data and predictive modeling tools. For example, a recent analysis (Brooks et al. 2009) concludes that the performance of an all-metal ITER PFC system (tungsten diverter with beryllium-coated first wall) may be marginally acceptable with acceptable sputtering and transient erosion, tungsten microstructure evolution, and high but manageable tritium codeposition. While this analysis is encouraging, the study and other analyses have numerous uncertainties due to a lack of certain critical plasma and material response data, and a general lack of validated, integrated, self-consistent codes. In any event, many aspects of the ITER PFC approach (e.g., with a beryllium-coated first wall) would not extrapolate to post-ITER high duty-factor devices.

Analysis has shown there are surprisingly few choices for PFC surface materials for DEMO and commercial devices because of concerns, including the previously mentioned sputter and transient erosion, with tritium codeposition and trapping, as well as activation issues. For solid materials, tungsten is a prime candidate material. For liquid surfaces, the feasible choices appear to be three flowing liquid metals: lithium, tin, and gallium (Brooks 2002).

A critical concern for the future is the operating limitations placed on the core and edge plasma because of the need to have viable plasma/surface interaction performance. ELMs are a good example because a moderate ELM frequency may be needed to achieve good core plasma energy confinement and acceptable core plasma impurity levels (e.g., with ELM-free operation, impurities may excessively accumulate in the plasma center). However, for much of the expected future device ELM parameter window (frequency, duration, and energy content) including so-called giant ELMs, the PFC erosion would be unacceptable. Likewise, the PFC response may severely limit acceptable disruption frequency and the related issue of operation at or near beta limits. For example, it is unclear if the ITER PFCs could survive even one "high-power" disruption (Hassanein et al. 2009). Accordingly, disruption mitigation by methods such as impurity pellet injection would be needed.

Codes and code packages generally exist for discrete parts of PMI science. Two major PMI code efforts used in the United States are the OMEGA Collaboration and HEIGHTS code package. OMEGA consists of the following:

- U.S. codes and code packages for plasma parameter computation in the plasma edge and SOL (UEDGE plasma and DEGAS neutral particle transport codes, and kinetic plasma add-on packages to the same)
- Full-kinetic, three-dimensional sputtering erosion/redeposition code package (REDEP/WBC)
- Related codes for sputter yield computation (e.g., TRIM-SP, molecular dynamics surface response code [several] with potential generation codes), tokamak-type sheath analysis (e.g., three-dimensional, full-kinetic BPHI-three-dimensional code), and material sputtering/D-T and helium recycle, mixed-material (e.g., beryllium-tungsten) formation and/or response codes.

OMEGA is not real-time coupled—instead, it uses coupling via individual efforts and coordination.

The HEIGHTS code package consists of coupled codes for the following:

- computing plasma transient deposition on surfaces
- vapor formation
- radiation transport
- atomic data
- MHD
- surface thermal conduction and hydraulics.

The HEIGHTS code package also computes the surface and plasma response to the above-mentioned ELMs, disruptions, and other plasma transients.

In addition, there are U.S. models and/or codes for such areas as dust formation and/or transport, atomic and molecular processes, liquid metal surface properties, and the like.

There have been successful efforts at code validation such as with the REDEP package showing good code-to-data comparisons with PISCES-B and DIII-D results for physically sputtered carbon and tungsten (Brooks 2002), and HEIGHTS code package comparisons for eroded material after plasma gun and fusion device surface irradiation (Hassanein 2002; Hassanein and Konkashbaev 2003).

Extensive gaps remain in understanding and modeling the coupled edge plasma, plasma surface interaction, and surface material evolution. Major needs in this area include detailed experiments with substantially improved diagnostics for the edge plasma, material response data, and extreme-scale predictive code packages.

In particular, major science gaps exist in modeling the following:

- Turbulent and/or convective (blob) edge and SOL plasma transport, including forces on impurities
- Mixed-material and/or alloy formation and response (sputtering, plasma transients, helium, and neon irradiation)
- Dust formation and/or transport, and various associated issues (e.g., ion cyclotron range of frequencies, sheath rectification, and atomic and molecular data).

Numerical gaps are numerous, including the overriding goal of extreme-scale, real-time coupling between plasma edge, SOL, and sheath codes; material response, evolution, and trapping codes; and impurity transport codes; and with specific needs, such as common incorporation of PFC computer-aided design (CAD) geometry and three-dimensional plasma and B-field mesh generation and use.

Transient plasma/surface interaction modeling has its own set of requirements including the following:

- Model development and coupling of transient plasma energy transport from core to SOL turbulent transport to PFCs (diverter, walls)
- Mixed-materials effects (beryllium, carbon, and tungsten) on plasma vapor shield induced formation and response
- Melt layer formation and splashing
- PFC structural changes due to impact of instabilities, etc.
- Droplet and dust formation and transport
- Transient effects on resulting core-plasma operating limitations in the ITER and DEMO devices, and solutions to the same
- Dynamic coupling between core, SOL, and PFC surface during instabilities
- Detailed analysis of various mitigation methods in full three-dimensional tokamak geometry; for example, this includes liquid metal (flow, splashing, contamination, etc.) and pellet injection (dynamic behavior of plasma during injection, radiation losses, radiation deposition on nearby components, etc).

The complexity of material evolution during plasma bombardment and the subsequent effect on plasma behavior has been less explored. A large gap in understanding the materials science of the plasma and particle-modified material can be subdivided into three major areas:

1. Synergistic effects of multiparticle interaction with materials that affect the material properties across multiple-spatial scales (e.g., neutrons that penetrate several millimeters into the material bulk versus energetic heavy ions that implant in the top several nanometers)
2. Time-dependent surface evolution (across multiple time scales) and its effects on intrinsic material properties between the bulk and surface of a PFC, including morphological and phase transformations affecting both emission properties (e.g., sputtering and evaporation) of the material surface and mechanical properties (e.g., ductility and strength) of the material bulk
3. Reactive erosion effects on the interaction of hydrogen in fusion plasmas.

Atomic, molecular, and near-surface particle/material interaction data have been produced, collected, evaluated, and disseminated relevant to the plasma edge through an ongoing community of processes, often coordinated by the Atomic and Molecular Data Unit of the International Atomic Energy Agency, the Controlled Fusion Atomic Data Center (United States), the Atomic Data and Analysis Structure at Joint European Torus (JET)/University of Strathclyde (Europe), and the Data and Planning Center at the National Institute for Fusion Science (Japan). However, a continual need exists to improve the quality of theoretical and experimental results, upon which the present knowledge of these processes is based. Other continual needs include the remediation of gaps in knowledge (e.g., those involving excited states for which data have not been available or for new materials such as lithium, beryllium, and tungsten; diagnostic species; and those involving radiation enhancers, such as nitrogen or heavy noble gases). Extreme-scale computing will play a central and crucial role in making these new results achievable.

Fundamental particle-surface interactions (PSIs) data for processes—such as chemical sputtering—that are complementary to more complex measurements from plasma devices have been produced, collected, and evaluated by a number of groups and the data centers listed in this panel report. As in the case of the required atomic and molecular data for the edge, emerging species (e.g., beryllium and tungsten wall materials and the inevitable mixed materials, such as beryllium-tungsten-carbon compounds) drive an urgent need for new benchmarking experiments and a wide variety of completely novel theoretical calculations. Molecular dynamics, largely enabled by powerful quantum chemistry techniques and teraflop- to petaflop-class supercomputers, is becoming the tool of choice to complete these calculations, particularly for the low-temperature regime where chemical reactions take place (e.g., hydrocarbon formation and chemical sputtering). Additionally, molecular dynamics are complementary to binary collision approaches that work reasonably well at high-impact velocities. These codes share a significant synergy with those modeling deeper layers and require longer processing periods. They are highly reliant on the steady advancement of computer capability to improve, elaborate, and test the interaction potentials that they require, and to run larger simulation volumes for longer times.

The molecular dynamics method is a powerful tool for studying surface and bulk properties of materials under the irradiation conditions expected for fusion reactors (e.g., Voter 2006). Assuming only classical mechanics and a definition for the dependence of the interatomic forces on geometry, molecular dynamics simulations evolve the dynamics with full atomistic detail. While direct molecular dynamics is limited to a time scale of nanoseconds, recently developed accelerated molecular dynamics (AMD) (Uberuaga and Voter 2006; Uberuaga et al. 2005) and adaptive kinetic Monte Carlo methods (Henkelman and Jonsson 2001) exploit the infrequent event nature of activated processes in materials to achieve time scales that are orders of magnitude longer. Performed carefully, AMD methods provide long time scale results that are as accurate as direct molecular dynamics. The crucial limitation that prevents using molecular dynamics methods effectively for the design of fusion materials is the lack of accuracy in existing interatomic potentials. This is especially true for multicom systems (more than one atom type) for which potentials often do not exist.

Exascale computing offers a path to solving this problem. If, instead of using empirical interatomic potentials, the forces for the molecular dynamics simulation are obtained from electronic structure calculations (e.g., density functional theory [DFT]), the accuracy of molecular dynamics predictions is dramatically improved. Defect energies and activation barriers calculated with the best DFT methods are now accurate enough to compare favorably with experiments. While DFT calculations are many orders of magnitude more computationally expensive than empirical potentials, there is reason to estimate that DFT algorithms could be developed to perform efficiently on millions of processors. Rather than using standard parallelism to extend the spatial scale of the molecular dynamics simulation, this approach would harness the massive concurrency to achieve high accuracy in the interatomic forces for systems with perhaps thousands of atoms. If this challenge can be met, the wall-clock cost of a DFT force call would become comparable to a present-day empirical potential force call. Direct simulations using this molecular dynamics-DFT capability would offer access to time scales approaching a microsecond, and DFT-AMD or DFT-adaptive kinetic Monte Carlo could be used to access longer time scales. For processes with high-activation barriers, time scales of milliseconds and well beyond are accessible. This exascale molecular dynamics-DFT or DFT-AMD capability would represent a quantum leap in atomistic simulation methodology, providing a predictive tool for addressing several problems critical to understanding materials in a fusion environment—problems that are currently out of reach. Examples include implantation, diffusion, and desorption of tritium, which are critical for understanding the tritium retention problem; helium diffusion and coalescence, which are critical for understanding helium bubble formation and embrittlement; sputtering in the presence of complex surface chemistry (e.g., adsorbed/incorporated carbon, beryllium, and tritium); and collision cascades and their long-time annealing response in realistic environments involving segregated impurities and alloying elements. Such simulations will directly answer several pressing questions and will serve as benchmarks for validating and improving higher-level models such as kinetic Monte Carlo.

Improved, and ultimately predictive, understanding of the properties, performance, and feedback of the fusion plasma with wall materials is highly dependent on developing robust and coupled models spanning phenomena on many time and length scales. These include the small and large spatial, and short and long temporal, scale behaviors of the plasma near the wall, and the interaction of this plasma with the wall surface and with deeper layers of the wall.

One regime within this range contains the atomic, molecular, and near-surface particle-surface interactions (AM&PSI) characterized by femtosecond to picosecond times and lengths on the order of nanometers. For example, at this scale electrons collide and excite ions, molecules are sputtered via ion bombardment of the wall, and photons are emitted from radiating mantle ions.

Models at this level challenge present computational capabilities and will require extreme-scale computing to treat with greater fidelity the effects presently identified and to enable—for the first time—consideration of a larger number of particles reaching even fully quantum simulations, thereby elucidating more uncertain or unknown effects. Specifically, present terascale to petascale computers allow treatment of one- or two-electron atomic collisions involving two heavy particles (e.g., $D^+ + H_2$) or many electrons in a single center (e.g., $e + W^{9+}$). Similarly, molecular dynamics simulations are possible for up to hundreds or a few thousand atoms for thousands of time steps. Extreme-scale computing and emerging new computational techniques would enable the required extension of these codes to treat multielectron transitions in complex atomic and molecular collisions (e.g., $e + Be_2C$, $D + W_2C$) and perhaps tens of thousands of atoms in novel molecular dynamics simulations including millions of electrons interacting at the quantum mechanical level.

Because of the short time and small scale of these interactions and the necessity of using the largest computational facilities to describe them, the coupling of physics resulting from this level with other levels of the multiscale modeling is best completed through a particular workflow taking advantage of this scale separation (Figure 11). That is, these physics results should be computed and then stored within an appropriate file structure separately from higher level, multiscale runs. This approach is logical because the AM&PSI physics possess similarities. In particular, each time an electron approaches a molecule, or an ion reaches a material surface, the multiscale code should not digress into a separate supercomputing solution of this AM&PSI-scale interaction. Rather, the required outcome should be retrieved from storage after being generated and verified. These AM&PSI physics should be successively improved and expanded beginning from an initial, less complete, less verified, and less detailed version.

The AM&PSI physics results therefore flow to the next-higher level of simulation (e.g., particle transport and plasma species, charge, and energy evolution) and separately to other levels that require them. In addition, this workflow allows for the use of the AM&PSI in other applications, such as experimental analysis suites or plasma diagnostics.

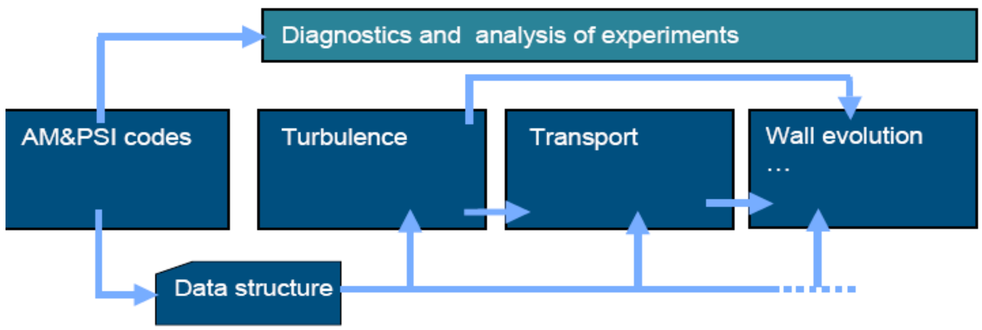


Figure 11. AM&PSI workflow-codes treating the atomic, molecular, and near-surface particle-surface interaction events require the largest computing resources available and possess similarity, which allows them to be run separately and then used in higher levels of the multiscale edge codes. Image courtesy of David Shultz (Oak Ridge National Laboratory).

Bulk Plasma-Facing Materials

The structural materials that define the engineering viability of fusion energy will be exposed to arguably the most hostile environment ever experienced by a large-scale structural engineering system. The anticipated thermo-mechanical stresses associated with cyclic operation between offline and full operational parameters, combined with the projected intense steady-state and transient thermal heat fluxes (>1 MW/m^2), approach or exceed the thermo-mechanical stresses experienced by structural materials during the re-entry to earth orbit of space flight systems. There is a strong desire to operate these structural materials at unprecedented high temperatures to maximize the thermodynamic efficiency of the operating fusion energy system (more electricity per unit fusion event). Many of the coolant systems under consideration for advanced fusion energy systems would corrode existing structural materials, thereby necessitating the science-based development of improved corrosion-resistant materials systems or specially tailored coatings that will remain intact during extended thermal and mechanical cycling. In addition to these daunting demands, a further operational requirement for the structural materials is satisfactory dimensional stability and retention of mechanical and physical properties during prolonged exposure to unprecedented high fluxes of 14-MeV neutrons. The potential for ballistic dissolution of nanoscale strengthening phases and property degradation from cumulative displacement damage in future fusion energy devices is at least an order of magnitude more severe than in existing fission reactor cladding and internal structures. In addition, the high levels of hydrogen and helium isotopes produced by nuclear transmutation reactions in the fusion neutron environment greatly enhance radiation defect accumulation compared to the fission neutron (low transmutant hydrogen and helium) case. These immense challenges for bulk structural materials will require new extreme-scale computational simulation tools to guide the development of a new suite of high-performance radiation-resistant materials for fusion energy systems.

Gaining understanding and predictive capabilities in this critical area will require addressing the simultaneously complex and diverse physics occurring over a wide range of lengths (angstroms to meters) and times (femtoseconds to days). The shorter time and length scales correspond to individual displacement cascade events, which serve as the source term for all subsequent radiation damage processes in the structural material. At intermediate length and time scales, many physical processes are initiated—including phonon scattering by radiation defects, crack formation, nucleation of radiation damage defect clusters, and creation or dissolution of corrosion-resistant passivation layers at surfaces and interfaces. At longer length and time scales, additional phenomena—such as induced stresses from dissimilar material systems, macroscopic swelling from radiation-induced void formation, irradiation creep deformation, thermal creep deformation, grain boundary embrittlement due to accumulation of helium-filled cavities, and uniform and localized corrosion—become important. This broad palette of physical phenomena will require development of detailed physics models and computational strategies at each of these scales, as well as algorithms and methods to strongly couple them in a way that can be robustly validated. While present research confined to each of these scales, or pioneering ways to couple two or more of them, already pushes the state of the art in technique and available computational power, simulations spanning multiple scales needed for the ITER, DEMO, and other devices will require extreme-scale computing platforms and integrated physics and computer science advances.

The current computational science approach for modeling bulk structural materials in neutron radiation environments involves a collection of state-of-the-art—with limited capability—standalone computational codes that are weakly linked to adjoining shorter- and longer-scale standalone codes. Numerous simplifying approximations are made at each length and time scale to pass tractable amounts of information between the standalone computer codes. The multiscale codes are not truly interactive in typical state-of-the-art, multiscale models; instead, the static results obtained by one set of codes are used as input parameters for a greater length or time scale code. The net effect is an increasing degree of uncertainty in the quantitative accuracy of the computational model predictions. At the greater length and time scales, some of these models begin to approach semiempirical, macroscopic data correlations, rather than truly predictive physics-based models.

The science basis is needed to develop high-performance alloys and ceramics that approach the ideal strength limit while retaining adequate ductility and fracture toughness (including large-scale fabrication and joining capabilities) and answer the following questions:

- What are the maximum practical limits in strength and toughness for materials?
- How are the laws of materials science altered under nanoscale and/or nonequilibrium conditions?
- What is the effect of crystal structure and atomic order and/or disorder (or noncrystallinity) on the properties of matter?

A fundamental understanding is needed of the high-temperature deformation mechanisms (i.e., the scientific basis for establishing robust structural design criteria for materials at high temperatures, including thermal creep, irradiation creep, cyclic thermomechanical fatigue, and creep-fatigue mechanisms).

The scientific mechanisms that control chemical compatibility and corrosion of materials must be elucidated.

CONCLUSIONS

The key challenge is to develop comprehensive exascale computational models for predictive, self-consistent, integrated, validated, full-process, time-dependent plasma material interaction. This would first encompass modeling of the edge and SOL plasma, including treatment of kinetic effects, three-dimensional geometry, turbulent transport, and full time-dependent coupling of plasma ions and electrons, neutral particles, photons, and electromagnetic fields. Next, plasma contamination from near-surface transport of sputtered or vaporized material and quantification of PFC particle and photon fluxes (and prediction of instability regimes) would be included. A critically related issue is predicting the near-surface material response to the extreme plasma fluxes of photons and particles under both normal and transient operations. This involves modeling of sputtering erosion and redeposition and other time-integrated PFC processes (e.g., dust formation and transport, helium, D-T induced microstructure formation and/or flaking) and the resultant impurity transport, core plasma contamination, mixed-material formation, and tritium codeposition in redeposited materials. The material and edge plasma response to transient processes, such as high-powered ELMs, vertical displacement events, plasma disruptions, and runaway electrons, would be an important component of this effort.

Predictive, multiscale models of bulk, as well as surface materials behaviors in fusion-relevant environments, are needed with the following elements:

- Fully predictive understanding of phase stability of materials in nonradiation and radiation environments, in particular correct modeling of magnetism effects in iron-chromium steels, and the mobility and consequences of helium complexes in structural materials (including the effects on grain boundary cohesive energy, etc.). Improved physical models are needed to describe the following phenomena:
  - cohesion (elastic constants, defect-free energies, and interaction of these defects with solutes including hydrogen, helium, and carbon)
  - phase stability under equilibrium and nonequilibrium conditions
  - energetic displacement collision cascade simulations
  - hydrogen, helium, and displacement damage accumulation
  - kinetics for atom transport and microstructural evolution
  - plasticity and fracture in defect-free and irradiated (radiation-hardened and/or embrittled) materials
  - thermal and irradiation creep.
- Void swelling expansion of the current disparate theory and modeling capabilities to develop well-integrated, experimentally validated, multiscale physical models that can predict the behavior, failure paths, and lifetimes of materials in the fusion environment.

Finally, an integrated materials-structure design development must be developed, including testing approaches for engineering structures with specific application to the harsh thermomechanical and irradiation environment of fusion energy systems.

Extensive computational resources are needed at all phases of fusion materials development to support model development, including large-scale structural damage mechanics computational codes to guide and interpret data obtained from component-level test facilities.

LASER-PLASMA INTERACTIONS AND HIGH-ENERGY DENSITY LABORATORY PHYSICS

Co-Leads: Riccardo Betti, University of Rochester
Warren Mori, University of California-Los Angeles

Panel Members: Brian J. Albright, Los Alamos National Laboratory; Kevin J. Bowers, Los Alamos National Security; Viktor Decyk, University of California, Los Angeles; Ronald Kirkpatrick, Los Alamos National Laboratory; Mike Key, Lawrence Livermore National Laboratory; Chuang Ren, University of Rochester; Xianzhu Tang, Los Alamos National Laboratory; and Richard Town, Lawrence Livermore National Laboratory

CURRENT STATUS

Controlled fusion offers the promise of a clean and sustainable supply of energy. One path towards this promise is inertial fusion energy (IFE), which relies on inertial confinement fusion (ICF). It is anticipated that ICF will soon be demonstrated by the achievement of ignition at the National Ignition Facility (NIF). The physics behind ICF is multiscale, both temporally and spatially, and it is also fundamentally multidimensional and nonlinear. Much of this physics is also part of the field of high-energy density laboratory plasma (HEDLP) science. Research in ICF has also been a driving force behind high-performance computational science, and ICF simulations are at the forefront of applications capable of effective utilization of petascale computing.

Recently, the U.S. Department of Energy's (DOE's) Office of Fusion Energy Sciences (FES) in partnership with the National Nuclear Security Administration (NNSA), established a Joint Program for HEDLP science. As part of this effort, a panel convened at the request of the Fusion Energy Sciences Advisory Committee (FESAC) to identify compelling research issues and opportunities in fundamental HEDLP science that need to be addressed to make the case for IFE. The committee wrote a report that was entitled *Advancing the Science of High-Energy Density Laboratory Plasmas* (FESAC 2009). It is clear from this report that both HEDLP and IFE are core areas within the DOE mission.

Objectives

The objective of the workshop was to identify how computing at the extreme scale could assist in overcoming forefront scientific challenges for developing IFE within the ensuing decade. The FESAC report (FESAC 2009) identified compelling science issues that need to be addressed to make the case for IFE to be a viable energy source. Because the issues are very broad, the present panel has focused on how computing at the extreme scale could help achieve major advances on a subset of these issues. We emphasized topics that are currently using petascale high-performance computing (HPC) resources and for which a case can be made that algorithms will scale to extreme-scale computers. Specifically, the panel members focused primarily on laser-plasma interactions (LPI) as they pertain to both compressing an IFE target using laser drivers and on the high-gain fast ignition (FI) IFE concept. The panel also considered how extreme-scale computing could impact science issues related to shock ignition, magneto-inertial fusion, and the compressed core. This panel report begins with the identification of four grand challenge science areas. The report then assesses current computing capabilities, discusses what transformative simulations could be possible on extreme-scale computers, and describes the algorithmic and computer science obstacles that need to be addressed to enable efficient use of these computers for accelerating scientific progress.

BASIC SCIENCE CHALLENGES AND RESEARCH NEEDS

Recent technological advances in lasers, particle beams, and Z-pinches have made it possible to generate plasmas with unprecedented energy densities in the laboratory. Understanding the properties and behavior of such plasmas constitutes the science area called HEDLP. This rapidly emerging subject is extremely rich in basic science phenomena and the foundation for IFE science, which is one possible path towards producing a clean and sustainable supply of energy. FESAC (2009) described compelling science opportunities in basic HEDLP and a prioritized list of issues for IFE science. Furthermore, the report also identified the science areas with opportunities for advanced computing to make a major impact. Many of the compelling science opportunities in HEDLP involve processes that can only be accurately understood through fully kinetic models and involve science that crosses from micro to meso and to macro time and space scales. For example, in some IFE science experiments a mm-scale pellet of deuterium and tritium is compressed to 1000 times solid density over ns time scales, and lasers with wavelengths of microns or smaller propagate through cm-scale plasmas.

IFE science currently involves examining the physics issues of the following IFE concepts:

1. High-gain direct drive
2. Z-pinch IFE
3. FI
4. Shock ignition
5. Heavy ion fusion
6. Magneto-inertial fusion.

The key physics and science areas that need to be understood to enable these concepts were identified in FESAC (2009); they include the following:

- <u>Laser-plasma instabilities and hot-electron generation.</u> Science question: How do laser-plasma instabilities in high-energy density (HED) plasmas scatter and reflect light, how to they generate energetic electrons, and how can they be controlled?
- <u>Implosion hydrodynamics for high gains.</u> Science question: How can HED plasmas be assembled to the densities and pressures required for maximizing the fusion-energy output?
- <u>Intense particle-beam generation by intense lasers.</u> Science question: What are the energy spectra of energetic charged particles generated by intense lasers interacting with plasmas, and how does intense light affect plasma dynamics?
- <u>Transport and energy coupling of intense particle beams in high-energy-density plasmas.</u> Science question: How are intense particle beams transported in and how does their energy couple to HED plasmas?
- <u>Influence of magnetic field on high-energy-density fusion plasmas.</u> Science question: How do magnetic fields, either spontaneous or induced, affect the behavior of fusing plasmas, and might the effects make improved prospects possible for IFE?

- Integrated target physics for IFE. Science question: What are the optimal target designs to achieve high gains with good stability and efficient driver coupling?

Identification

Four priority research directions (PRDs) were identified for HEDLP and IFE science for which extreme-scale computing could make a transformative impact. The first PRD is **the nonlinear optics of plasmas**, where the goal is to understand how an ensemble of overlapping Gaussian beamlets (speckles) mutually interacts in HED plasmas. This understanding is critical to successful development of IFE concepts using laser drivers. It requires fully kinetic modeling because subtle changes to the electron distribution function can lead to substantial differences. On extreme-scale computers, the goal of simulating an ensemble of speckles using fully kinetic modeling could be achieved, which could in turn lead to ideas on how to tame these interactions and to the development of high-fidelity reduced models for meso-scale simulations. The second PRD is **relativistic HED plasma and intense beam physics,** where one goal is to understand how lasers at the intensity and power frontier interact with and are absorbed in HED plasmas. This physics requires detailed understanding of single-particle trajectories and how the complex patterns of large currents of relativistic particles form in plasmas and collectively interact. Fully kinetic and relativistic modeling is required. On extreme-scale computers, fully kinetic simulations using true time and length scales of FI targets could be possible for the first time. This will also require the development of coupling micro- and meso-scale models. The third PRD is **integrated FI simulations,** where the goal is to provide fully integrated modeling of high-gain FI IFE concepts where the timing of the intense ignition pulse, the compression of the pellet, and the survival of an inserted cone tip can be important. On extreme-scale computers, the coupling of fully kinetic simulations of HED plasmas with parameters obtained from macro-scale hydrodynamic compression models may be possible for the true time and space scales. The fourth PRD is **magnetized HED plasmas,** where the goal is to understand how spontaneous or induced magnetic fields can affect burning HED plasmas. The physics spans a wide parameter space, from the dense compressed core of a traditional IFE target to the more tenuous plasmas in reversed field configurations. Extreme-scale computers will enable high-fidelity simulations of dense collisional plasmas that are inertially confined and in which heat flux is limited by magnetic fields. The development of meso-scale models, coupled with extreme-scale computing capabilities, should enable breakthroughs in the understanding of magnetized plasmas under compression.

Current Capabilities

LPI, relativistic higher-energy density, and magnetized HED plasmas are currently modeled using particle-in-cell (PIC), Vlasov, hybrid PIC/fluid, and fluid codes using full Maxwell, nonradiative Darwin, paraxial wave, and implicit, fluid field solvers. Substantial HEDLP simulation capabilities exist at DOE national laboratories and within a few key universities and small businesses. These codes have been run using more than 10,000 processors and have been applied to investigations of some of the relevant physics and science issues. They represent a significant starting point for scaling upward to computing at the extreme scale.

The present state of the art in PIC modeling of the nonlinear optics of plasmas has been conducted on Roadrunner, the first supercomputer to achieve a petaflop/s (Dongarra 2009), using the VPIC PIC code (Bowers 2008a). To date, multiple calculations have been performed of a solitary f/8 laser speckle under National Ignition Facility laser and plasma conditions. These calculations advanced 3-5x10^{11} macro-particles over ~10^5-10^6 time steps; each particle advance requires ~250 floating point operations leading to 10^{19}-10^{20} operations in total. With exascale computational resources, it will be possible to model one or more laser beams with on the order 1000 laser speckles total (10^{23} floating point operations). In addition, OSIRIS (Fonseca et al. 2002) and Z3 have been used to study single-speckle physics in detail in lower dimensions. OSIRIS has scaled to over 30,000 processors.

In present state-of-the-art two-fluid calculations, researchers at the Lawrence Livermore National Laboratory (LLNL) used the pF3D meso-scale model to simulate one full long-pulse laser beam of the NIF through the plasma conditions in a NIF ignition hohlraum (Hinkel et al. 2008). The spatial extent of the computational box was 2 x 2 x 5 mm^3, and the simulation advanced for 100-ps of the 2-ns peak power of the ignition laser pulse. The simulation took 33 petaflop-days to run on the IBM-Blue-Gene/L system.

For relativistic HED science and integrated FI simulations, PIC simulations are currently used to model the laser-generated energetic particles and their transport to the core, and hybrid PIC/fluid models are being used to model the transport near the core. Using the PIC code OSIRIS (Tonge et al. 2009), a laser interacting with an isolated target at 100 times critical density with a 50-μm radius for at least 2.5 ps has been simulated. These simulations take the equivalent of 2 days on 4096 processors of the Franklin CRAY XT-4 computer at the National Energy Research Scientific Computing Center (NERSC). Other PIC codes, such as PSC (Kemp and Ruhl 2005) and PICLS (Sentoku and Kemp 2008), have also been used to model similar physics.

Conducting integrated FI simulations will require coupling single-fluid hydro simulations to categories of models ranging from reduced and hybrid PIC-fluid models to full PIC models. Currently, all of the simulations for the specific models noted are carried out separately—not as an integrated system. Using the radiation hydro code HYDRA to simulate a 9 x 9 wedge of a NIF ignition capsule requires 160M zones to resolve mode 100. This simulation was advanced for 15-ns and took 2 weeks on 4096 central processing units (CPUs). The implicit PIC code, LSP, is used to model the transport of the relativistic electron beam (REB) in two dimensions from the cone tip to the compressed fuel. A typical computational box currently is 150 x 250 mm^2, and the associated simulation, which is advanced for 7-ps with a maximum density of 300 g/cm^3, took 2 days on 64 CPUs.

PRIORITY RESEARCH DIRECTIONS

As discussed above, four research directions were identified for which extreme-scale computing has the potential for transformative impact to HEDLP and IFE science. These are the nonlinear optics of plasmas, relativistic HED plasma, and intense beam physics, integrated FI simulations, and magnetized HED plasmas. In the following, the associated information presented for each of these subjects will include a summary of the research direction, a description of scientific challenges, and discussions of 1) the potential scientific impacts and outcomes, and 2) the potential impact on fusion energy science.

Nonlinear Optics of Plasmas

Summary of Research Direction

In laser-driven IFE, high-energy lasers compress a capsule containing deuterium-tritium (DT) fuel to conditions needed to ignite a hot spot; i.e., 1000 times solid density and 100 million Kelvin. To be practical as an energy source, IFE must achieve *high gain*, defined as the ratio of fusion energy yield to laser energy. Consequently, IFE is traditionally conceived to involve so-called “direct drive,” where the laser beams are focused directly onto the outer walls of the capsule. However, indirect drive ICF, where the lasers focus onto the inner walls of a hohlraum (that heat and re-radiate x-rays that are absorbed by the capsule), is also a possibility. If feasible, direct drive is believed to be promising because there is no loss of efficiency in converting energy from laser to hohlraum. In laser-driven IFE, absorption of laser light launches one or more inwardly directed shocks that converge spherically and create a hot spot in the DT fuel at the center of the capsule. To achieve the extreme conditions required for fusion ignition, laser energy must be absorbed in a carefully prescribed manner.

One of the primary challenges to IFE is to control the nonlinear optics (NLO) of plasmas. The NLO of plasmas includes laser-plasma instabilities, which involve resonant coupling of laser light to plasma waves. LPI have three deleterious effects on IFE: 1) LPI scatter laser light, so less energy is available for compression of the capsule; 2) LPI deflect laser light, which degrades implosion symmetry; and 3) collisionless damping of plasma waves associated with LPI leads to the production of hot electrons, which can pre-heat the fuel and make it harder to compress.

In IFE experiments, uniform irradiation of a pellet requires sophisticated techniques for smoothing the laser beams by the use of random phase plates to break each beam into an ensemble of laser "speckles" or coherent elements within a laser beam. The physical dimension of a laser speckle depends upon the optic f/# of the laser beam, the design of the random phase plates, and the laser wavelength. Predictive modeling of LPI requires an ability to model LPI not only within a solitary laser speckle (the "building block" of a beam), but also to account for *interaction* between speckles through the exchange of particles and waves.

Scientific Challenges

Modeling this complex interaction is a computational grand challenge that requires computing at the petascale to exascale and beyond because one must simultaneously capture small-scale-length (of order the Debye length), short-time-scale (of order the laser period) dynamics in simulation domains large enough to contain one or several laser speckles. Moreover, the physics of LPI can be very complex, involving intricate nonlinear and nonlocal wave-wave and wave-particle processes (see Figure 12 [Bowers et al. 2009]). These processes depend sensitively on the distribution functions of electrons and ions, so high-fidelity kinetic modeling is required. While much can be learned from lower dimensions, simulations in three spatial dimensions (three-dimensional) are required to obtain an accurate, quantitative representation of the physics. These constraints set the size of the problem: for the laser and plasma conditions encountered in a single IFE speckle, one must evolve phase space in of order 10^9 computational spatial cells for of order 10^6 time steps. However, the dynamics in one laser speckle can influence neighboring speckles, so one ideally would model an ensemble of interacting laser speckles over a range of laser and plasma conditions.

With present petaflop-scale supercomputers, and over the next decade as exaflop-scale computing becomes available, it is probable that the PIC method will be the only practical technique for *ab initio* three-dimensional modeling of LPI in an ensemble of laser speckles. However, Vlasov simulations on extreme-scale computers in two spatial dimensions could provide a useful comparison platform that could lead to a better understanding of the necessary fine-scale phase space resolution. In the PIC algorithm, the plasma is represented by a collection of computational macro-particles, each of which represents several physical electrons or ions. These computational particles propagate on a spatial mesh on which electric and magnetic fields are computed. Depending on plasma conditions and dominant LPI processes, tens to thousands of computational macro-particles must be used within each computational cell to adequately represent the distribution functions of particles. In a PIC time step, macro-particles are advanced according to Newton's laws, and electric and magnetic fields are interpolated from the computational mesh. These fields, in turn, evolve self-consistently according to Maxwell's equations using the electrical charges and currents of the macro-particles as sources. PIC codes with reduced field equations (e.g., electrostatic and Darwin) could also be used to understand deeply the coupling between particle-particle and wave-particle interactions on electron plasma and ion acoustic waves. For example, comparison between simulations of Stimulated Raman Scattering (SRS) using the fully electromagnetic PIC code OSIRIS (Hemker et al. 2000; Fonseca et al. 2002) together with the electrostatic PIC code UPIC (Fahlen et al. 2009; Decyk 2007) is already providing insight into the role of subtle kinetic effects on the saturation and recurrence of SRS (Winjum et al. 2009).

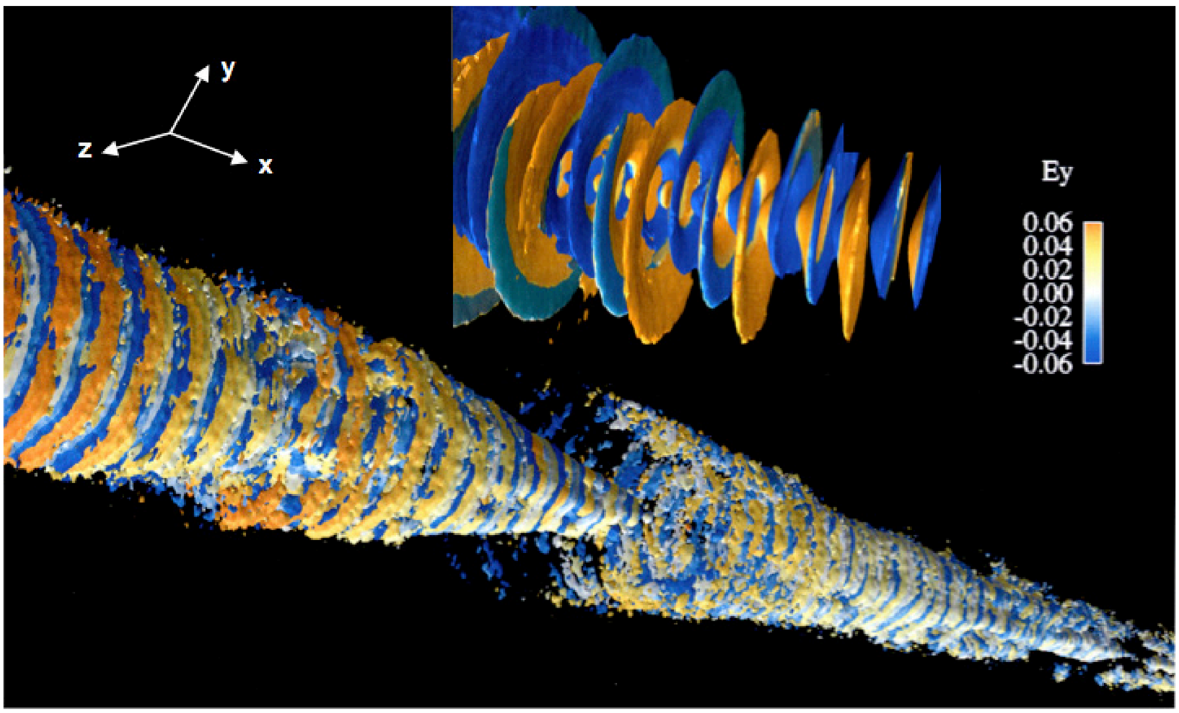


Figure 12. Three-dimensional PIC LPI simulation under NIF hohlraum plasma conditions. Isosurfaces of the constant electron plasma wave electric field, colored by the laser electric field, are shown during nonlinear saturation of the instability (Source: Bowers et al. 2009). Image courtesy of Lin Yin (Los Alamos National Laboratory).

An important complement to *ab initio* PIC modeling of LPI is mesoscale modeling, where the laser-plasma dynamics are treated in a reduced fashion. These models, exemplified by the PF3D code (Still et al. 2000), evolve envelope equations for the laser and plasma waves to track the growth of LPI without needing to resolve individual wavelengths or periods. Coupled to a hydrodynamic description of the plasma, they sacrifice fidelity of the kinetic physics for an orders-of-magnitude increase in speed and an ability to simulate much larger volumes than can be done with PIC. Presently, mesoscale models have proved extraordinarily useful in settings where the growth of LPI stays low and the effects of particle trapping and other kinetic processes (not modeled or contained within a reduced model in pF3D) are minimal. However, as the laser intensity is increased to achieve higher gain it is inevitable that kinetic processes are important.

Even with exascale computing, PIC modeling of LPI will be impractical in the experimental design cycle, where typically a large number of rapid-turn-around calculations must be done to refine the design of an experiment. Therefore, mesoscale modeling will be key to the design of IFE experiments in coming years. One of the goals of petaflop-scale supercomputing will be to explore the use of high-resolution PIC modeling at the petascale and beyond to obtain reduced descriptions of the complex, nonlinear LPI physics that can be implemented economically into codes such as PF3D.

Potential Scientific Impacts and Outcomes

As noted earlier in the "current capabilities" discussion in this document, the VPIC PIC code (Bowers et al. 2008a) has been utilized to carry out the largest PIC modeling of LPI to date at LANL on Roadrunner, the first supercomputer to achieve a petaflop/s capability (Dongarra 2009). Multiple calculations have been performed of a solitary f/8 laser speckle under laser and plasma conditions corresponding to those present in the NIF. These calculations advanced 3-5 x 10^{11} macro-particles over ~10^5-10^6 time steps with each particle advance requiring about 250 floating point operations – thereby leading to 10^{19}-10^{20} operations in total. With exascale computing, it will be possible to model one or more laser beams with on the order of 1000 laser speckles total (10^{23} floating point operations).

Running on the entirety of the IBM Blue-Gene-P petaflop-class supercomputer at LLNL, the present state-of-the-art pF3D code (Hinkel et al. 2008) can currently model an entire NIF laser beam for a duration of 100 ps. It can accordingly be projected that as supercomputing capability transitions to the exascale, multiple-beam modeling for ns time-scales will be practicable, ushering in a new era of high-fidelity design of inertial fusion experiments.

All of these tools will lead to major improvements in understanding how single and multiple laser beams interact with plasmas. They will also provide detailed insights of plasma waves with respect to both weakly nonlinear and strongly nonlinear dynamics and to how they interact with each other, with trapped electrons, and with lasers. Such advances in understanding could lead to the ability to control instabilities and to shut off one process by turning on another. Moreover, investment in exascale computing for IFE can be expected to lead to advances in PIC simulation techniques, including new optimizations, extreme (100-million-way) parallelization, de-noising, data analysis, visualization, and fault tolerance. Advances in Vlasov and Vlasov Fokker-Planck (Bell et al. 2006; Duclous et al. 2009) capability will include adaptive mesh refinement in phase space with promise of leading to improved understanding of LPI near thresholds. Sine significant progress in meso-scale modeling will require improved hydrodynamic, parabolic, and elliptic solvers on massively parallel platforms, these fundamental areas of advanced scientific computing research will also be invigorated by this effort in the LPI application domain.

Potential Impact on Fusion Energy Science

Over the next 5 to 10 years, extreme-scale supercomputing capability has the potential to be a transformative technology for IFE that can enable a host of scientific breakthroughs. Foremost is the prospect of *predictive* modeling of LPI, which will enhance the exploration of IFE scientific concepts. This capability may lead to an ability to control LPI from a first-principles physics basis. These advances not only underpin the high-gain, laser-driven IFE research, but they also cross over into other areas of inertial fusion (e.g., fast-ignition inertial fusion and magnetized target fusion).

As a final point, it is worth noting that some of this work overlaps with activities that are presently part of the National Ignition Campaign (NIC). This is fortuitous, because it means that many of the tools and techniques developed for the NIC can be leveraged for this effort, thereby increasing its chance of success and decreasing the investment that must be made by the DOE Office of Science. However, it should also be recognized that laser-driven IFE must address a different set of challenges. For instance, achievement of high gain will likely entail the use of higher laser intensity (with correspondingly higher levels of LPI) than is currently employed within point designs for NIF ignition experiments. Consequently, LPI may be a more acute concern for laser-driven IFE.

Relativistic High-Energy Density Laboratory Plasmas and Intense Beam Physics

Summary of Research Direction

Intense particle-beam generation through high-intensity LPIs and the transport of the particle beam in dense plasmas are at the frontier of HEDLP research. Understanding these processes is essential to many HEDLP applications, including new ICF schemes, laboratory astrophysics experiments, and new radiation sources. A primary example is a promising new ICF concept: FI (Tabak et al. 1994). Unlike conventional hot-spot ignition, this approach separates the compression of the fusion fuel from the ignition step. First, laser pulses drive the compression of a spherical shell of DT ice to high density at low temperature, without a central hot spot. Then, a second very high-intensity laser delivers a pulse of energy that ignites the compressed fuel. Compared to hot-spot ignition, the FI concept promises much higher gain for the same driver energy (Tabak and Callahan 2005). In addition, the driver energy necessary to achieve ignition may also be significantly reduced from the 1 MJ value that appears to be necessary for hot-spot ignition (Atzeni 1999).

The FI concept has three steps—compression, ignition, and burn—each with disparate scale lengths and physics requirements. In the compression phase, the compression laser beams deposit their energy onto the outside of the spherical shell, which rapidly heats up and ablates outwards. The ablation plasma acts like the exhaust in a rocket driving the remaining shell inward and compressing the fuel to form a compact, dense fuel mass with an areal density ρR of 1.5-3 g/cm^2 (ρ is the mass density of the core and R its radius). The purpose of the burn phase is to propagate a burn wave across the fuel. These two phases are usually simulated with radiation hydrodynamics codes used in conventional ICF with fusion burn packages.

The ignition phase in FI is a distinctive step and involves a separate ignition laser. Because the ignition laser cannot directly reach the dense core region, the laser energy needs to be converted into an energetic electron beam, which then propagates to the core and deposits its energy there. The location for the relativistic electron generation needs to be as close to the core as possible to reduce energy loss along its path to the core. Currently, there are two major schemes in FI research to bring the ignition pulse close to the core. One is to use fuel pellets with a hollow gold cone attached to their sides. The compression beams converge onto the pellet from all directions except those within the opening of the cone. The hollow cone keeps a clear path for the ignition pulse to propagate and generate the energetic electrons at the cone tip (Kodama et al. 2001, 2002). In the so-called hole-boring scheme, the ignition pulse is preceded by a channeling pulse to generate a channel through the underdense corona and into a critical density surface, beyond which it cannot propagate (n_c=10^{21} cm^{-3} for 1 μm light). The ignition pulse is then sent in tandem to reach the critical surface and may continue to push forward into the overdense plasma with its ponderomotive pressure ("hole-boring"). In the meantime, it heats the plasma to generate the energetic electrons (Figure 13 [a]). The ignition phase determines the coupling of the ignition laser to the target core and thus the viability of FI. It is also the least understood phase in FI.

In both approaches, the fast electron current significantly exceeds the Alfvén limit and is neutralized by a return current drawn from the cold background plasma. The resulting electron distribution can be susceptible to filamentary instabilities and generates large magnetic fields. The electrons deposit their energy in the compressed fuel and launch a burn wave across the fuel. The presence of the cone (which may have introduced large asymmetries in the imploded configuration) and the large magnetic fields that have been generated by the transport of relativistic electrons could significantly modify the burn.

There are other high-gain IFE concepts that rely on separating the fuel assembly from ignition. In proton-driven FI, a thin converter foil is inserted near the entrance of the cone and hence away from the cone tip. Relativistic electrons transit across the foil, and the space charge separation accelerates protons (or ions) toward the fuel from the rear surface of the foil. A proton beam is easier to focus relativistic electrons, but the efficiency of converting electron energy into proton energy is an issue. In shock ignition (Betti et al. 2007), an intensity spike is included at the tail of the compression pulses, causing a shock to be launched that coincides with the reflection of the compression shock.

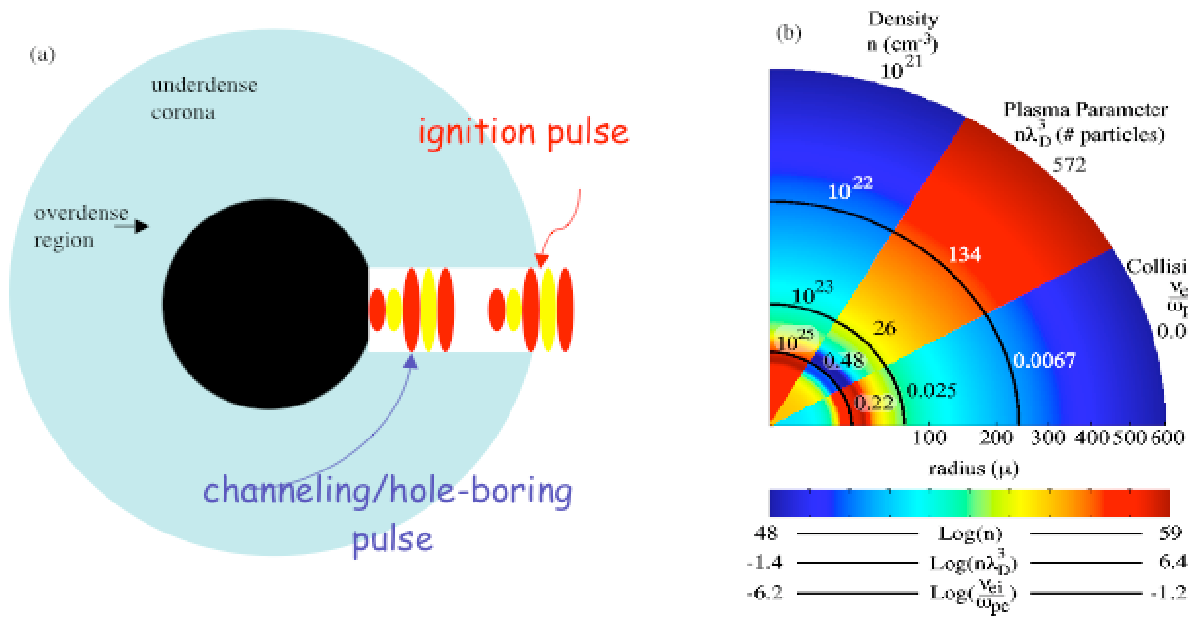


Figure 13. (a) A sketch of the hole-boring scheme for FI; (b) the density and temperature profiles above n=1021cm^{-3} of an FI target from a one-dimensional hydrodynamics simulation. Left image courtesy of Chuang Ren (University of Rochester); right image courtesy of John Tonge (University of California-Los Angeles).

Scientific Challenges

A key challenge is integrating laser plasma coupling at low densities and electron transport to higher densities while including complex physics across disparate time and length scales. To illustrate the complex physics and how exascale computation could lead to transformative progress, the "hole-boring" scenario is described.

The ignition phase can be approximately divided into two relatively independent processes: 1) channeling, and 2) relativistic electron generation and transport. Close to peak compression, the under-dense corona outside the critical (nonrelativistic) density surface has a radial length scale of ~1 mm and may significantly scatter and absorb a potential ignition pulse through numerous parametric instabilities. A channeling pulse of intensity 10^{18}-10^{-19} W/cm^2 is sent through the corona to form a ~20-μm-wide channel by expelling the plasma with its ponderomotive pressure. The duration of the channeling pulse is estimated to be ~100 ps. Scientists need to know whether the channeling pulse can reach the critical surface in this time, whether the channel is straight, and the density of the residual plasma remaining in the channel. After the channel has been formed, a short ignition pulse is sent down the channel. The ignition pulse currently estimated to require 100 kJ in 10 ps, which corresponds to a peak power of 10 PW. Scientists need to know how the ignition pulse interacts with the residual plasma. At the end of the channel, the pulse may continue to push to densities much higher than the nominal critical density through relativistic effects and laser hole boring. Eventually, the pulse deposits its energy in the form of relativistic electrons. Here, we need to know the laser-to-hot-electron conversion efficiency, the hot electron distribution, and how these electrons transport in the overdense plasma to reach the core.

Channeling occurs entirely in a collisionless (weakly collisional) plasma. The relativistic electrons are also generated in a collisionless plasma and must propagate up a density gradient, from a collisionless to a collisional plasma. The density and temperature profiles above $n=10^{21}$ cm^{-3} of a typical ICF target at its peak compression are used to generate the plot in Figure 13 (b). Rather than plotting the temperature, two important dimensionless parameters $n\lambda_D^3$ (number of electrons in a Debye cube) and ν_{ei}/ω_{pe} (ratio of the electron-ion collision frequency to the plasma frequency) are plotted along with the density. In the region where $n<10^{23}$ cm^{-3}, $\nu_{ei}/\omega_{pe} <<1$, which illustrates that the plasma is weakly collisional in this region (in some recent designs the plasma temperature is lower). The laser energy is most likely to be absorbed somewhere between $n=10^{22}$ to 10^{23} cm^{-3}. Since $\nu_{ei}/\omega_{pe} \sim n^{1/2}$, collisional effects are even less important in the mm-scale corona outside the critical surface. For both the channeling and hot electron generation/transport processes, the PIC model is the most accurate description. The PIC simulations for both processes require extreme-scale computing resources.

To simulate the channeling process in three-dimensional and full-scale with explicit PIC codes, scientists need a simulation box size of 1000 μm × (200 μm)^2 and duration up to 100 ps. To resolve the wavelength and frequency of a λ=1μm laser, these translate to a grid of 20000 × 4000 × 4000 and 10^6 time steps. Assuming 20 particles per cell and 400 FLOPs/particle-step (assuming higher order particle shapes), such a simulation requires 6.4 trillion particles, 360 TB memory, and a total of 2.6×10^{21} FLOPs.

Using the PIC code OSIRIS, full-scale simulations for channeling were performed two-dimensionally up to 15 ps (Li et al. 2008). They showed for the first time that channeling in mm-scale plasmas indeed has many new phenomena that were not present in previous short-scale experiments and simulations, including plasma buildup to above critical density in front of the laser, laser hosing/refraction, and channel bifurcation and self-correction. Reduced-size three-dimensional simulations with plasmas up to 540 μm long were also performed up to 3 ps, and three-dimensional laser and channel hosing were observed for the first time, which qualitatively confirmed the channeling process and predicted a channeling speed twice as large as in two-dimensions. Typical simulations took several days using 10^4 processors. However, none of these simulations ran long enough for the laser to channel the entire 1-mm underdense plasma. Scaling for the channeling time and energy was obtained by extrapolation. An exascale machine would enable high-fidelity simulations to obtain the scaling for a point design. As described in the following discussions, ultra-intense lasers show the promise of achieving FI with the hole-boring scheme. Any proposal for the construction of such lasers needs a convincing design based on these high-fidelity simulations.

To simulate hot electron generation and transport 2.5D PIC simulations of isolated targets have been performed with OSIRIS (Tonge et al. 2009). These targets have a 50 microns radius of over-dense plasma and a dense core 20 microns radius. The hot electrons are self-consistently generated with a 1m ignition laser with a spot size of 20 μm and intensities of 5 x 10^{19} W/cm^2, 2 x 10^{20} W/cm^2, and 8 x 10^{20} W/cm^2 intensities. The integrated line density from the target edge to the core is equal to the line density of a target with an exponentially rising density with a 100-μm scale length per decade from the target edge to 10^{23}cm^{-3} density. This mass distribution is more compact for simulation purposes. These simulations are run for at least 2.5 ps. These simulations take the equivalent of ~2 days on 4096 processors of the Franklin supercomputer at NERSC. Note that large-scale simulations of electron generation and transport for cone-guided targets have been performed as well using PIC codes with a collision model included (Sentoku 2008).

With the capability provided by an exaflop class machine, simulations of relativistic electron generation and transport in fast-ignition targets can potentially be performed at full scale in 2.5D and scaled targets in three-dimensions. Such simulations would enable answers to key questions regarding the feasibility of FI. Specifically, it would be possible in 2.5D to model a 550 μm radius pellet on a 1.2 mm x 1.2 mm simulation grid with a cell size that resolves half a skin depth at 10^{26} /cm^3 density. The target includes a 50 μm radius core at 10^{26}/cm^3 and an exponential density ramp down to 10^{21}/ cm^3 with a decade density drop per 100 μms. The simulation runs for 10 ps. The computation will require 1.7 x 10^7 time steps on a simulation grid with 2.3 x 10^{13} grid points. There are about 3 x 10^{14} electrons in the pellet with fixed weight particles with 4 electrons per grid at 10^{23} /cm^3 and 4000 particles per cell at 10^{26}/cm^3. The flop count for this simulation is 80 flops per particle per push giving 4.0 x 10^{23} flops for electrons or about 10 days on an exaflop machine.

The three-dimensional target will resolve half a skin depth at 10^{24}/cm^3 density with a 200-μm radius pellet in a 500 μm x 500 μm x 500 μm simulation grid. Above a density of 10^{24}/cm^3, an absorbing core will be used. This simulation is run for 2 ps. The computation will require 4.2 x 10^5 time steps on a simulation grid with 7.8 x 10^{15} grid points. There are about 3 x 10^{16} electrons in the pellet with 8 electrons per grid. The flop count for this simulation is 120 flops per particle per push giving 8.04 x 10^{23} flops for electrons or about 20 days on an exaflop machine.

Future algorithms that efficiently merge the physics of the weakly collisional lower density plasmas with the more collisional plasma surrounding the core will need to be developed. This includes coupling reduced PIC models to full PIC models. The reduced models might include hybrid PIC/fluid codes, PIC with Darwin field solvers, PIC with mesh refinement, PIC with Monte Carlo collisions, and Vlasov with Fokker Plank operators. Recently, Cohen et al. (2010) proposed a framework for merging a full PIC method with Monte Carlo collisions with a reduced PIC method that solves for the electric field using Ohms's law. This framework appears to have much potential.

Potential Scientific Impacts and Outcomes

The scientific impact of this research direction will be significant. Extreme-scale computing coupled with improved models could potentially lead to a quantitative and predictive understanding of how a laser is absorbed at a dense plasma boundary. It could provide the long-sought answers to what is energy spectra of electrons and/or ions generated by intense lasers and what happens to the energy of these electrons as they propagate through high-energy density laboratory plasma. Other outcomes include advances in PIC, collisional PIC, hybrid full and reduced PIC, hybrid PIC, and fluid models.

Potential Impact on Fusion Energy Science

The potential impact to fusion energy science of this research direction would be a simulation tool for determining the laser and plasma conditions that are needed to make fusion energy through FI. This would accelerate progress in the successful development of the FI concept.

Integrated Fast Ignition Simulations

Summary of Research Direction

The recent FESAC report (FESAC 2009) on HEDLP identified integrated HEDLP physics as one of the overarching issues that links the six fundamental HEDLP issues of IFE target physics that need to be addressed to enable the successful advancement of IFE sciences. The report identified the overarching science question for integrated simulations to be: *What are the optimal target designs to achieve high gain with good stability and efficient driver coupling?* Although this science question applies to all ICF concepts, the panel felt that the most scientifically and computationally challenging integration question related to the development and use of an integrated FI simulation capability. In particular, the key scientific challenge was identified as understanding the interdependence of hydrodynamics, intense laser propagation, energetic particle production and transport, and ignition. Furthermore, in the cone-guided approach to FI, issues related to nonspherically symmetric compression will be an issue. To address this scientific challenge requires the achievement of the computational challenge to couple microscale, fully kinetic models at low density and reduced meso-scale hybrid-kinetic models at high density with macro-scale models. The development and application of this integrated FI simulation capability would enable quantitative predictions of HEDLP experiments and accelerate the optimization of the FI concept.

The FI concept (Tabak et al. 1994) was described earlier in the summary of priority research directions for relativistic high-energy density laboratory plasmas and intense beam physics. As noted, this involves two distinct phases (compression and ignition) with disparate length and time scales and different physics requirements. In the compression phase of a direct-drive implosion, the long-pulse compression laser deposits its energy onto the outside of a spherical shell, which rapidly heats up and ablates outwards. The ablated plasma causes the remainder of the shell to be driven inwards, compressing the enclosed fuel to form a compact, high-density (~ 300 g/cm^3) assembly. This process involves the compression of 1-mm scale targets over tens of nanoseconds.

When the fuel has been assembled, the high-intensity, short-pulse laser delivers the tens of kJ of beam energy necessary to heat up a small volume of the compressed fuel to initiate a burn wave that then propagates through the remaining fuel. The ignition energy must be delivered in a time short compared to the dense core disassembly time (<100 ps). The ignition laser beam cannot directly deposit its energy in the dense fuel because laser light cannot propagate through electron densities above the relativistic critical density. The laser energy is first converted to a relativistic electrons, which can either propagate directly to the core in the electron-driven FI scheme or can be converted to a proton (or ion) beam that then propagates to the core (called proton-driven FI). The relativistic electrons are sometimes referred to as REB, but in reality the distribution function is a monotonically decreasing function of energy so a true beam is not formed.

Scientists have observed that FI has a range of spatial and temporal scales and different physics requirements that require a number of simulation capabilities. The relevant spatial scales range over seven orders of magnitude from the capsule size ($\sim10^{-1}$ cm) to the compressed core's Debye length ($\sim10^{-8}$ cm), while the temporal scales range over ten orders of magnitude from the implosion time (~20 ns) to the inverse plasma frequency of the core ($\sim10^{-6}$ ps). Integrated modeling therefore requires coupling micro to meso to macro-scale models.

Scientific Challenges

The compression phase can be modeled using multidimensional radiation-hydrodynamics codes that have been applied to conventional hotspot ICF, such as the three-dimensional HYDRA code (Marinak et al. 2001) or the two-dimensional DRACO code (Radha et al. 2005). The hydrocode needs to be multidimensional, not only because the cone introduces a large-scale asymmetry but also because ICF implosion is unstable to fluid instabilities such as the Richtmyer-Meshkov and Rayleigh-Taylor (Rayleigh 1883; Taylor 1950) instabilities. If, during the implosion phase, the laser intensity of the compression laser is too high, then a more detailed model of LPIs than hydrocodes typically contain may be required, such as the meso-scale pF3D model or micro-scale-explicit PIC codes such as OSIRIS.

During the ignition phase, the simulation must accurately model the propagation of the channeling laser, the propagation of the short-pulse ignition laser, and the generation of the REB. To date, these steps have been modeled using multidimensional explicit PIC codes such as OSIRIS, Z3, PSC, and PICILS. Subsequently, the REB propagates through high-density plasmas that preclude the use of standard explicit PIC codes, and so meso-scale models such as the implicit-hybrid PIC LSP code (Welch et al. 2001) will need to be fully coupled into them and demonstrate favorable scalability. As noted previously, other models such as PIC with Monte-Carlo collisions, Darwin field solvers, and mesh refinement may also be useful.

Although there has been considerable work on understanding the physics of hydrodynamics, LPI, and REB transport in isolation, there has only been a limited effort in understanding the interdependence of hydrodynamics, laser propagation, energetic particle production, transport, and ignition in an integrated sense. Work to date includes the FI Integrated Interconnecting (FI^3) code project in Japan (Sakagami and Mima 2004), LLNL's linking of the two-dimensional radiation hydro-code LASNEX to two-dimensional explicit PIC, and the two-dimensional version of the hybrid-implicit PIC code LSP using a Python interface (illustrated in Figure 14), and the University of Rochester's linking the two-dimensional radiation hydro-code DRACO to the two-dimensional version of LSP (Solodov et al. 2008). All of these efforts have made substantial approximations (e.g., reduced dimensionalities, smaller spatial extents) and use only limited data exchange between the various physics models. Developing and understanding how to couple micro-scale, fully kinetic models at low density and reduced meso-scale hybrid-kinetic models at high density with macro-scale models over long-time scales is the key computational challenge for which substantive progress must be achieved.

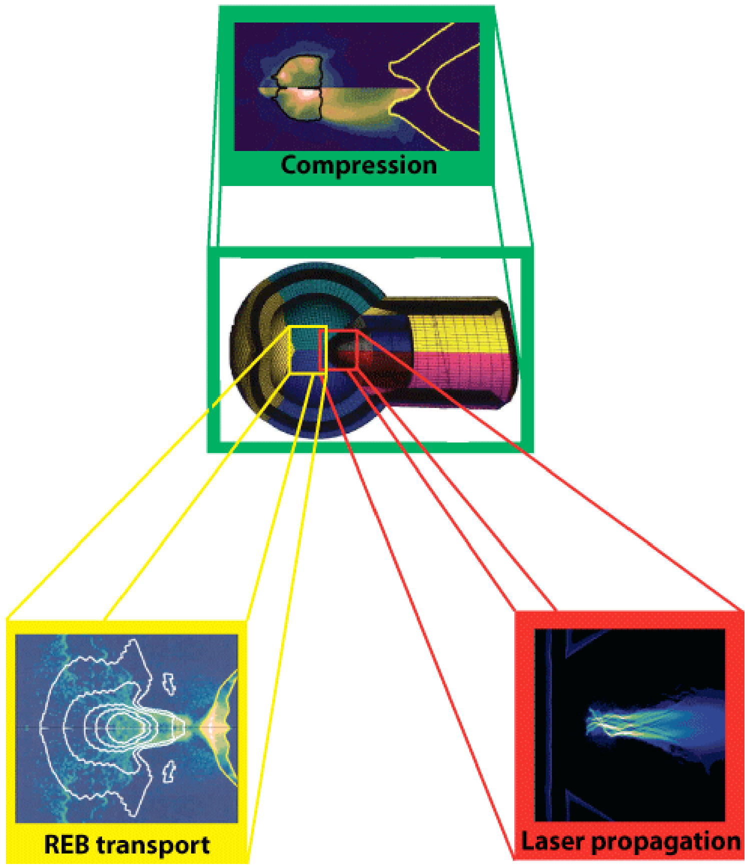


Figure 14. An illustration of how the physics of different regimes of an FI target is currently separated and modeled. Image courtesy of Riccardo Betti (University of Rochester) and Warren Mori (University of California – Los Angeles).

To develop an integrated FI simulation capability, advances in computer sciences are required to ensure that parallel scalability and load balancing are maintained when disparate physics modules are coupled. Computational physics challenges that need to be addressed include developing accurate representations of collisions, equation of state, and initialization techniques for micro- and meso-scale models.

Potential Scientific Impacts and Outcomes

The potential impact and outcome would be a simulation tool for designing an FI target and for quantitative predictions with integrated experiments. It would also lead to advances in techniques for coupling microscale, mesoscale, and macroscale models. These advances would include dynamic load balancing with disparate numerical methods running simultaneously, and the initialization of macro- and mesoscale simulations from data obtained from microscale models.

Potential Impact on Fusion Energy Science

The impact of this research direction to FES would be a simulation tool for designing integrated FI experiments that include modeling the nonspherical target compression from long pulse lasers and the ignition of the compressed core by the short pulse lasers. The simulations would model the macro scales of the long pulse laser and the target and the microscale physics of the energetic electrons and their transport. This would accelerate progress in and help to optimize the FI concept.

Magnetized High-Energy Density Plasmas

Summary of Research Direction

Fusion occurs when nuclei are maintained at sufficient temperature for sufficiently long times. Magnetic fields, both applied and self-generated, can aid confinement because they reduce the thermal conductivity perpendicular to the magnetic field. Using magnetic fields, together with ICF approaches, can potentially be attractive. For example, magneto-inertial fusion (MIF), also known as the Magnetized Target Fusion (MTF) (Lindemuth and Kirkpatrick 1983), is a concept that combines the key features of the two traditional fusion approaches (Lindemuth and Widner 1981). A schematic for such a system is shown in Figure 15.

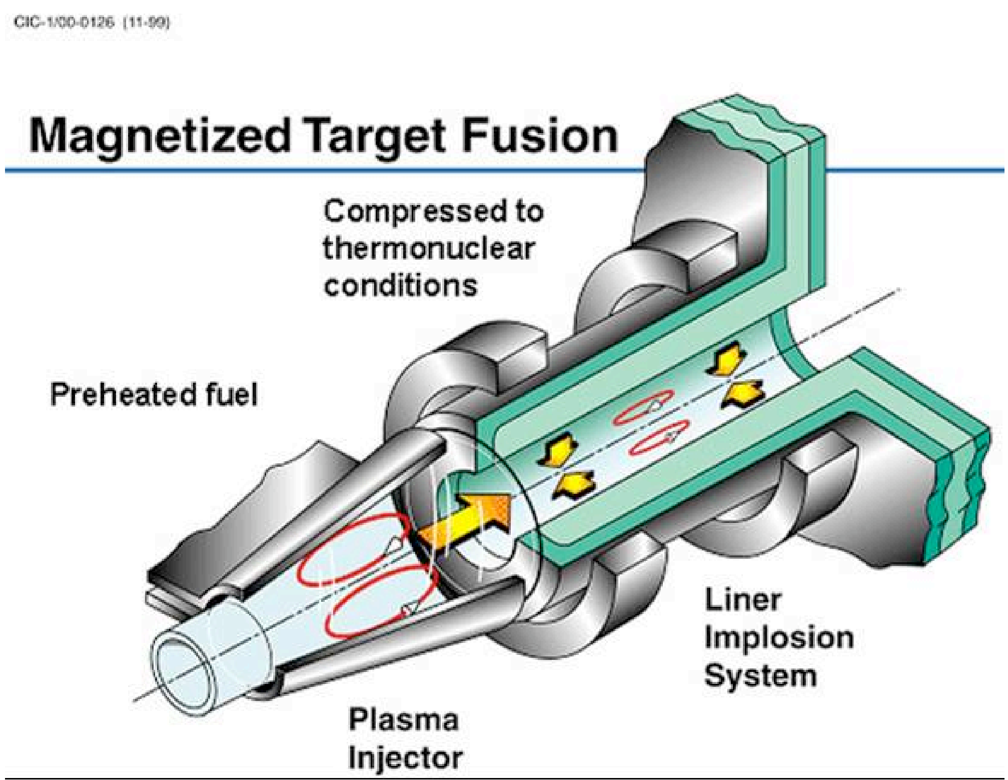


Figure 15. A schematic for magnetized target fusion. Image courtesy of Richard E. Siemon (University of Nevada, Reno).

Specifically, the DT fuel in a preheated magnetized target plasma is rapidly compressed to thermonuclear temperatures by imploding solid or plasma liners. The magnetic field with closed flux surfaces provides the thermal insulation. MIF operates in a fuel density and time scale regime intermediate between magnetic confinement fusion (MCF, also known as magnetic fusion energy or MFE and ICF), and therefore has an advantage of allowing orders of magnitude smaller system size compared with conventional MCF.

Scientific Challenges

The critical physics issues for MIF are related to the selection and optimization of the magnetized target and the inertial pusher. The magnetized target must have simply connected magnetic topology such as the Field Reversed Configuration (FRC) and the Spherical Torus or Spheromak, while the inertial pusher can be a solid liner or a standoff plasma liner. An optimal MIF target is to provide low 1) outward thermal and particle transport; and 2) inward impurity particle transport. Understanding the transport physics requires first-principle-based kinetic simulations, which share many of the challenges encountered in MCF but also have unique challenges. An interesting case is the FRC, which is the target plasma for the upcoming FRX-L experiment. The magnetic field nulls in an FRC challenge the validity of the conventional drift-kinetic and gyrokinetic formulation and likely require the fully resolved kinetic simulations. To this end, fully kinetic PIC codes including external collision models being developed for nonlinear optics of plasmas and FI will be very useful. As noted earlier, some of these codes already run on petascale platforms and with appropriate modifications, should be able to run on extreme-scale computing platforms with proper investment. The conventional gyrokinetic PIC and continuum codes from the MCF program can be applied to address the transport in a Spheromak target.

The impurity generation and inward transport are generally important issues to MCF but have particular significance for MIF. This is due to the much-enhanced radiation losses at higher plasma density in MIF. In the solid liner scheme, the target plasma core particle transport and the magnetized sheath physics determine the ion bombardment flux to the liner. The impurity flux from the wall by sputtering needs to be evaluated, and its inward penetration needs to be quantified. The MIF scheme has the mega-gauss magnetic field mostly parallel to the liner surface, leading to an unusual scenario in which the liner is positively charged by the ions as opposed to the usual plasma sheath. This is another area in which six-dimensional phase space kinetic simulation is essential to understand the physics (including potential instabilities), and where the state-of-the-art kinetic simulation codes have demonstrated potential readiness for extreme-scale computing. The impurity flux production at the liner surface is a plasma/materials interaction problem. The liner erosion can be quantified by molecular dynamics simulations. The longer-time-scale liner response to plasma particle and heat flux might benefit from the accelerated molecular dynamic simulation method (Voter et al. 2002), which is another tool that has demonstrated readiness for extreme-scale computing. Integrating the plasma transport and materials response is critical not only for MIF but also for MCF. If this can be achieved on extreme-scale computing platforms, significant progress could result.

The MIF target is heated to thermonuclear temperature by compression of an imploding liner. The target compression physics is where MIF and ICF share similar challenges. While radiation-hydrodynamics codes are the workhorse for ICF compression studies, the presence of a strong magnetic field in MIF requires at least the radiation-magnetohydrodynamics model. For the planned experiments, the final compressed target reaches a collisionality in which the mean free path is comparable to the target plasma size, but the target plasma during the long period of compression remains collisional. Accordingly, a collisional two-fluid model supplemented by a radiation transport model, would be appropriate. Computational tools of this kind would be also extremely useful and timely for the development of the magnetized ICF program.

Several issues should be addressed in more detail than has previously been possible with existing computers. One is the dynamic nature and interplay between the magnetized plasma and the fusion self-heating (Kirkpatrick 2007). This requires detailed transport calculations for the energetic fusion reaction products, calculations that preserve the correlation in space energy and direction for the products. This can be conducted using dynamic particle tracking, or perhaps more efficiently by some more subtle computational schemes, but diffusion-like approaches that resort to simplified characterization of the distribution functions (e.g., Fokker-Planck) do not suffice. Whether by particle tracking or by more subtle methods, proper treatment of fusion product transport in MIF systems requires a substantial increase in computer performance.

While several computer codes include some of the physics relevant to MTF, none fully suffice and the development of a comprehensive code for future application to the wide range of possible magneto-inertial fusion schemes must address the anticipated continuing advance in computing technology.

Potential Scientific Impacts and Outcomes

The experimental effort in MIF is still in its early stage. Predictive simulation using extreme-scale computing could play a critical role in enabling magneto-inertial fusion. In addition, self-generated or applied magnetic fields could improve the performance of traditional hot-spot initiated ICF. High-fidelity simulations of collisions and transport of a compressed target is another opportunity for extreme-scale computing using PIC techniques.

Potential Impact on Fusion Energy Science

Magnetic insulation provides MIF with the potential for orders of magnitude reduction in power requirements compared with conventional ICF, which is shown in Figure 16 (Siemon et al. 1999). If this avenue to low-cost energy-producing plasma is successful, MIF will permit fusion development without billion-dollar facilities, thus circumventing one of the most serious obstacles for the conventional fusion development.

CROSSCUTTING RESEARCH DIRECTIONS

There is an overlapping need for fully kinetic simulation needs within the research directions described above. The PIC technique (Birdsall and Langdon 1985; Hockney and Eastwood 1988; Dawson 1983) is one method that has already been used successfully on petascale platforms for some of these research areas. It is anticipated that PIC simulations will continue to be a critical tool in these areas, and there are clear paths for scaling it upward to extreme-scale computing platforms. In addition, other kinetic approaches, such as Vlasov simulations, could enable a deep understanding of both the physics and the necessary resolution in phase space. Extreme-scale computing will permit Vlasov simulations of two spatial and three velocity components. Extreme-scale computing will also allow the coupling of micro-, meso-, and macroscale models. The following sections discuss the necessary applied math and computer science support for advancing high-fidelity PIC to extreme-scale and coupling models.

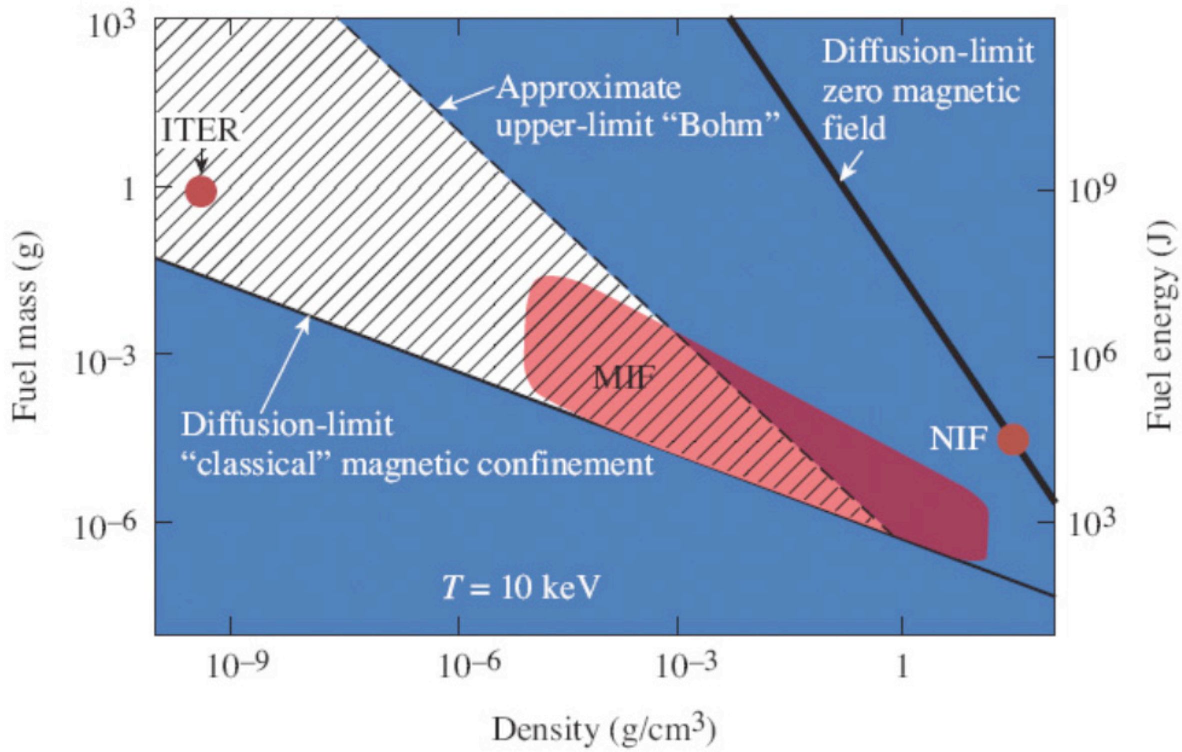


Figure 16. A parameter space diagram of fusion concepts. Image courtesy of Richard E. Siemon (University of Nevada, Reno).

Applied Mathematics and Computer Science Support

In early 2009, ATI began selling a teraflop-capable consumer graphics card for as low as $99 (~10 Gflops/dollar and ~12 Gflops/Watt) (AMD 2009). While the 40-nm graphics processing unit (GPU) used in this card is not ideal for scientific applications, it provocatively suggests computing resources that require less than 80 kW/petaflop will be available for under $100,000/petaflop in the near future. Thus, it is possible that petascale supercomputing will be routine by 2019, and we focus on scientific questions that would require a true exascale resource as opposed to a large number of petascale resources. The extreme-scale simulations identified in the preceding paragraphs are in this category.

Much of the extreme-scale computational needs for the four PRDs require first-principles three-dimensional relativistic electromagnetic kinetic simulations. One option is the PIC techniques (Birdsall and Langdon1985; Hockney and Eastwood 1988). PIC divides the simulation volume into a mesh and represents plasma as many particles. These particles are "pushed" by mesh-sampled electric and magnetic fields. The fields in turn are updated using electrical currents "accumulated" from the particle motions. While the PIC algorithm has been around for over 50 years, current PIC codes are complex and sophisticated. Ensuring that such codes run effectively on future compute nodes, scale to millions of compute cores, and provide high-fidelity simulations will require applied mathematics and computer science support.

Such simulations are ideal candidates for extreme-scale computing. Because particles cannot move faster than the speed of light and a time step times the speed of light is typically less than a cell size, few particles leave a cell within a time step. Therefore, internode communications are naturally optimized by a spatial domain decomposition (all information needed by a node in a simulation step is either already local to that node or on a nearest neighbor node) and map very well onto common, large supercomputer network topologies (e.g., hypertorus networks). As such, internode communication is seldom a limiting factor. Load balancing is not likely to be an issue for the interaction of multiple speckles; however, it is likely to be an issue for fast-ignition simulations.

Indeed, relativistic kinetic PIC simulations at the petascale with billions of mesh elements, trillions of particles, and hundreds of thousands of heterogeneous processing cores have already been performed (Bowers et al. 2008b) to model laser-plasma instabilities in a single speckle in the NIF laser beam (Yin et al. 2007, 2009). Given this success and the characteristics of the simulations discussed in the preceding subsections, it is highly likely that if sufficient resources are provided, the current state-of-the-art codes will be able to exploit fully the potential of extreme-scale computers when applied to extreme HEDLP simulations requiring trillions of mesh elements, hundreds of trillions of particles, and millions of heterogeneous processing cores (e.g., modeling a large ensemble of speckles in the NIF laser beam, including coupling between speckles induced by hot electrons, scattered light, and plasma waves).

Data analysis is a more daunting challenge; the I/O requirements for such should *not* be an afterthought in exascale supercomputer procurement. The I/O storage needs are dictated primarily by code checkpoint/restart needs. A single checkpoint may require petabytes of storage (essentially all supercomputer memory is written to nonvolatile storage in these extreme-scale simulations). Further, some strategies for data analysis, described in the following paragraph, can involve storing (and reusing several times) several checkpoints produced by a given simulation.

I/O performance needs are dictated by a single checkpoint size and the “short-run time scale” (the smaller of the queuing system time slice and the typical hardware uptime). That is, for check pointing to be viable at all, a checkpoint must to be written to nonvolatile storage much faster than the short-run time scale. Assuming a few hours for the short-run time scale, this requires several terabytes per second sustained aggregate I/O bandwidth during a checkpoint read or write. Fortunately, provided each node can access its checkpoint data independently, the bandwidth required per node is achievable with existing commodity technology.

The sheer volume of data generated by these simulations and I/O limitations can make traditional post-mortem analysis techniques impractical; i.e., in-situ simulation data reduction and analysis are a necessity. Because it is often unclear which simulation diagnostics will be most illuminating *a priori*, extreme-scale simulations can be more usefully thought of as experimental “shots.” Redoing shots with refined diagnostics is often useful. One such mode of operation is to perform a long, fast exploratory simulation with limited diagnostics but preserving several checkpoints. Interesting events can then be flagged and rerun from the closest checkpoint with more extensive custom diagnostics. Another such mode of operation is to run a long exploratory simulation interactively with extensive diagnostics, with the user indicating which diagnostics should be preserved as the simulation progresses. Both modes of operation work best when the job queuing system has provisions for interactive (low-latency) jobs. Likewise, rigorously deterministic parallel simulations codes are very beneficial. In addition, particle tracking will be a critical diagnostic. Options range from dumping data for every particle to flagging interesting particles after a run and rerunning it. The former will require dumping a checkpoint amount of data very often while the latter will require rerunning significant portions of a simulation. The latter seems most practical for extreme-scale simulations.

Mathematical Formulations

As described above, current state-of-the-art PIC codes are sufficient to model many of the phenomena found in these extreme-scale simulations. Nevertheless, several mathematical developments would greatly aid improving the simulation fidelity and analysis:

- Improved noise-reduction methods. Efficient higher-order methods, optimal particle weighting schemes, and mathematical theorems providing the tradeoff between particle count and accuracy in general situations would be extremely beneficial to all PIC simulations. Furthermore, the use of spectral methods should be investigated. The challenge of spectral solvers is parallel scalability as they involve global fields solves. However, Fast Fourier Transform (FFT)-based PIC codes, such as PARSEC and UPIC1, have shown parallel scalability out to ~10,000 processors for strong scaling studies.
- Improved boundary conditions. In open-simulation volumes, existing time-domain electromagnetic absorbing boundary conditions typically assume simple dispersion relations (e.g., vacuum) outside the simulation volume. Absorbing boundary conditions that assume a homogeneous warm, weakly collisional, magnetized kinetic plasma (a much more complex electromagnetic dispersion relation) might allow simulations to focus more tightly on regions of interest without loss of accuracy. Spurious surface waves when particles are adjacent to the boundaries are another potential issue. Finding the proper boundary conditions for particles such that the proper moments of the distribution function are maintained is also a challenge.
- Improved removal of numerical Cherenkov radiation. Common discretizations of the Maxwell equations alter the dispersion relation of light at short wavelengths such that ultra-relativistic particle motion can produce nonphysical Cherenkov radiation. Current methods for finite-difference time-domain solvers for addressing this are difficult to set up robustly and/or inefficiently. Spectral solvers have many advantages here.
- Improved short-range particle-particle collision models. Many commonly used models do not model all the collisional processes and relativistic effects found in high-energy density plasmas. Methods that get short-range forces correct within a mesh while using the standard PIC method for the forces for particles outside the mesh will be useful for studying collisions from first principles. Such methods combined with electrostatic or Darwin field solvers will be extremely useful. On extreme-scale computers, simulations of meaningful spatial domains with 10,000 or more particles per Debye length may be possible. In addition, collisional PIC models in which two-body collision models are added into the standard PIC model for particles within a collision cell also need to be developed further. This includes understanding the behavior of PIC models for particle sizes greater than 10 Debye lengths.

[1] Viktor K. Decyk. 2008. On the Atlas computer, a strong scaling study on a 512 x 512 x 126 grid with 16 particles per cell was carried out using the three-dimensional electromagnetic FFT-based PIC algorithm UPIC. For this relatively small problem size, the parallel efficiency was high until more than 8192 processors.

- Development of Vlasov codes that employ mesh refinement in velocity space will be a complementary technique for comparing against PIC codes for certain problems. In addition, it has been proposed that coupling a Vlasov code in which the distribution is represented as an expansion in spherical harmonics to a Fokker-Plank collision operator would be a useful for modeling fast-ignition physics. More generally, with algorithmic improvements, Vlasov simulations in two-spatial and three-velocity dimensions for spatial domains of interest may be a demonstrably useful approach on extreme-scale computers.
- Mesh refinement for electromagnetic solvers could provide large benefits to FI simulations. This is very challenging as scientists need to eliminate self-forces as particles cross from meshes of one resolution to another and spurious wave reflections for regions with different resolutions.
- Improved coupling between reduced models (e.g., fluid codes or kinetic models with reduced field equations) and kinetic models (e.g., PIC codes). This includes both hybrid fluid/kinetic codes to improve computational efficiency and analysis methods to parameterize reduced models (e.g., equations of state) from kinetic models.

Programming Models

Current state-of-the-art PIC codes exploit all forms of parallelism available to them on current resources—data parallel single-instruction, multiple data (SIMD), shared memory threads, distributed memory message passing, and embarrassingly parallel independent runs. These codes are written largely in C and FORTRAN programming languages for performance critical operations. The C++ and/or scripting languages like Python are frequently used for higher-level code (e.g., the outermost "physics loop") in a "simulation driving" approach. Libraries like MPI for distributed memory parallelism and OpenMP or "POSIX threads" (Pthreads) for shared memory parallelism are typically used. These libraries are frequently wrapped to allow quick "to-the-metal" replacements to be implemented and deployed on various platforms when desirable. Because of the complexity of the inner loops in these codes, data parallel SIMD is currently best exploited by hand-coding these loops in terms of SIMD compiler primitives (either platform-specific or wrapped to allow portability across different SIMD architectures). Although these languages and libraries do a very poor job of exposing communications and memory access limitations to the developer, state-of-the-art PIC codes, such as VPIC, were designed to optimize data flows to accommodate limitations of modern supercomputers. Hence, existing programming models, together with expected improvements due to ongoing research and development (R&D), appear to be adequate for most purposes on exascale supercomputers in the future.

An outstanding issue that must be addressed is how to best adapt to the unique architectural capabilities and limitations found in current GPUs. Specifically, current state-of-the-art codes exploit multicore processors efficiently using a multiple-instructions-multiple-data (MIMD) programming style with each core having its own independent instruction stream operating on its own dataset. Though current GPUs can be programmed MIMD style in principle, these GPUs are far more efficient when programmed as a generalized wide-vector SIMD processor. With sufficient resource investments, it is expected this issue can be resolved. Historically, many PIC codes were developed for traditional wide-vector SIMD processors, and this experience will be helpful with GPUs. GPUs and future multicore architectures may also require reorganization of data structures within the PIC codes, but such efforts will likely be common to all foreseeable architectures. Furthermore, current GPU architectural trends strongly suggest increased architectural flexibility in the future so that with the proper level of R&D support, the current limitations of GPUs may cease to be a significant issue for exascale supercomputers.

Like the mathematical foundations above, there are many barriers that—if overcome—would make exascale resources more amenable to applications deployment:

- Data motion limitations and parallelization issues make it difficult to tightly integrate third-party code efficiently; third-party libraries need to have well defined and documented nonuniform memory access (NUMA) characteristics. The lack of a standard deployment model and robust tools for cross-platform management further impedes software reuse.
- Domain-specific languages may be required to overcome intrinsic limitations of existing programming languages in making efficient portable code. Practically speaking, computational scientists are poor language designers, and tightly coupled collaborations with computer scientists is therefore necessary. Domain-specific automated code generation tools would make it possible to more rapidly and efficiently develop novel improved computational methods.
- Institutional machine-queuing policies impede development at scale. Reliable massively parallel software development requires high-priority, interactive access to large numbers of nodes for an hour of two at a time (i.e., the compile-run-debug cycle).
- Owing to both the complexity and opaqueness of data flow issues (e.g., NUMA/cache hierarchy, memory access) in current software tools and hardware, performance issues can be very hard to diagnose. Exposed architectures and detailed static and dynamic performance analysis tools would be especially useful in this regard.

BASIC PLASMA SCIENCE/MAGNETIC RECONNECTION PHYSICS

Co-Leads: Amitava Bhattacharjee, University of New Hampshire
William Daughton, Los Alamos National Laboratory

Panel Members: Hui Li, Los Alamos National Laboratory; Kai Germaschewski, University of New Hampshire; Luis Chaćon, Oak Ridge National Laboratory; Masahiro Hoshino, University of Tokyo; and Masaaki Yamada, Princeton Plasma Physics Laboratory

CURRENT STATUS

The liberation of magnetic field energy through the process of magnetic reconnection is one of the most far-reaching problems in basic plasma physics. Reconnection is at the core of a diverse range of plasma phenomena including solar flares, geomagnetic substorms, sawteeth oscillations, disruptions in tokamaks, extragalactic jets, and a wide variety of astrophysical phenomena. In the past decade, significant progress has been made in reconnection physics by means of theory, simulations, laboratory experiments, and spacecraft observations. With the advent of petaflop scale computers, large-scale fluid and kinetic simulations are playing an increasingly important role in addressing a variety of problems in reconnection physics. However, the scientific and computational challenges are immense and cannot be fully addressed even with petaflop scale computing. Looking toward the future of exascale computing, this panel report outlines some of the forefront scientific challenges in reconnection physics, identifies the required computational and algorithmic advancements needed to address these questions, and summarizes the potential impact on fusion energy science as well as broader space and astrophysical applications. Included are examples from recent efforts to use petaflop scale computers to study reconnection physics.

BASIC SCIENCE CHALLENGES AND RESEARCH NEEDS

Scale Separation in Reconnection Physics

Many of the scientific and computational challenges in reconnection physics are related to the vast separation of spatial and temporal scales. For reconnection to proceed, it is necessary to break the frozen-flux condition within so-called diffusion or non-ideal regions (Biskamp 2000). In most applications, these regions are extremely small in comparison to the macroscopic scales. However, without these non-ideal regions, there would be no reconnection. Thus, properly coupling the physics of the diffusion region to the dynamical evolution of the much larger system is a central difficulty in all reconnection problems. This scale separation largely determines which problems are computationally feasible. With more powerful computers, a growing class of problems will become accessible, but many other problems will remain beyond reach for the foreseeable future. Because this basic issue underlies all of the priority research directions (PRDs) addressed in this panel report, included is a review of the present understanding regarding the structure of the diffusion region within collisional and kinetic parameter regimes, followed by a discussion of the computational requirements imposed by resolving these features in large-scale simulations.

Assessing the multiscale magnetic reconnection problem usually begins with the magnetohydrodynamic (MHD) model, which is thought to provide a reasonably accurate description of reconnection physics in parameter regimes where the resistive layers remain well above the ion kinetic scales. In the classic Sweet-Parker solution (Parker 1957), the thickness of the diffusion region scales as $\delta_{sp} \sim (\eta L_{sp})^{1/2}$ where η is the resistivity and L_{sp} is the length of layer. The reconnection rate is determined by the aspect ratio of the layer $\delta_{sp} / L_{sp} \sim S^{-1/2}$ where $S = 4\pi V_A L_{sp}/(\eta c^2)$ is the Lundquist number. Assuming that the layer length L_{sp} scales with the macroscopic system size implies extremely large Lundquist numbers $S \sim 10^6 - 10^{14}$ for many applications. While the Sweet-Parker solution is well established for low Lundquist numbers $S < 10^4$, the scaling of collisional reconnection in the high S regime remains uncertain for various reasons. Even within idealized two-dimensional geometries, the Sweet-Parker solution is structurally unstable to plasmoid (secondary-island) formation beyond $S > 10^4$ (Biskamp 1986; Yan et al. 1992; Malara et al. 1992; Lapenta 2008). Recent linear theory (Loureiro et al. 2007) predicts a growth rate that scales as $S^{1/4} V_A/L_{sp}$ with the number of plasmoids increasing with $S^{3/8}$, indicating an increasingly vigorous instability for large S. Furthermore, the high S scaling in more complicated three-dimensional systems remains largely unexplored as a result of the computational challenge of resolving both the resistive layers, along with the macroscopic scales. Thus, even within the framework of resistive MHD, many basic questions regarding the dynamics and scaling of reconnection remain poorly understood.

Extreme-scale computing will provide an invaluable tool for addressing the high S regime. However, the computational cost to resolve the resistive layers and follow the macroscopic evolution on the global Alfvén time increases as $\sim S^{5/2}$ for three-dimensional explicit simulations. For Lundquist numbers in the range $S \sim 10^6$, these requirements can quickly surpass the capabilities of a petascale computer. These limitations suggest the central focus must be directed toward obtaining reliable scalings in the high S regime, which can then be used to better extrapolate to extreme parameter regimes $S > 10^{12}$ relevant to much of astrophysics.

In weakly collisional or collisionless parameter regimes, the structure of reconnection layers involves both ion and electron kinetic scales. As summarized in Figure 17, this imposes a daunting level of scale separation into the problem. The kinetic time scales are separated by the ion-to-electron mass ratio m_i/m_e while the spatial scales are separated by $(m_i/m_e)^{1/2}$. Furthermore, the macroscopic dimension L in most applications is vastly ($10^3 - 10^{10}$) larger than the ion kinetic scale, and it is necessary to follow the evolution on the global Alfvén time scale $\tau_A = L/V_A$ to understand the reconnection dynamics. The computational cost of explicitly resolving these kinetic scales three-dimensionally increases in the same steep manner $\sim (m_i/m_e)^{5/2} (L/d_i)^4$ for both two-fluid and fully kinetic particle simulations (although the kinetic simulations have a much larger coefficient). To reduce the separation between the ion and electron scales, most researchers presently employ an artificial ion to electron mass ratio $m_i/m_e \sim 25 - 400$. At the petascale, three-dimensional kinetic simulations for hydrogen plasmas $m_i/m_e = 1836$ will be possible for systems on the order of $\sim 10 d_i$. A factor of 10^3 increase in computing power will only increase the feasible three-dimensional system size by a factor of ~ 5.6. For most problems with a hydrogen mass ratio, the computational requirements are truly intractable on any computer for the foreseeable future. Instead, the focus must continue to be directed toward understanding the essential physics to eliminate (or skip) spatial and temporal scales of less relevance and ultimately, to obtain reliable scalings by a combination of numerical computation and analytical theory. Although significant progress was made over the past decade, a number of fundamental issues remain that must be resolved to confidently extrapolate to large-scale systems.

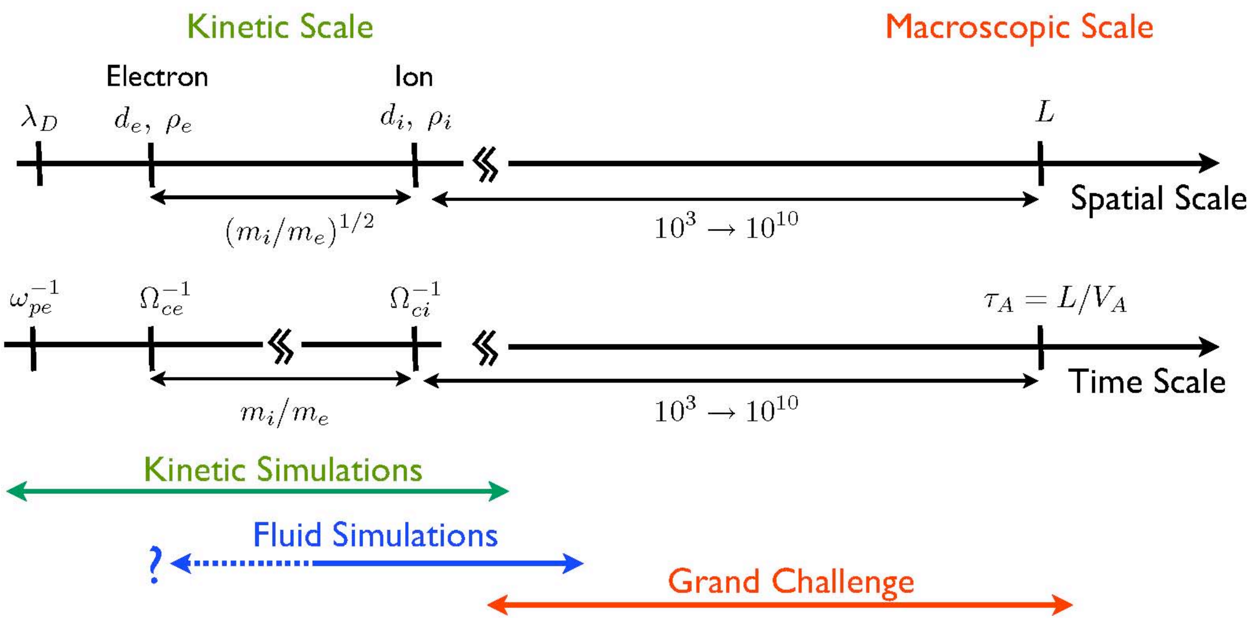


Figure 17. Overview of the spatial and temporal scales in collisionless reconnection. The fastest time scale is associated with the electron plasma frequency ω_{pe} followed by the electron cyclotron frequency Ω_{ce}. Electron spatial scales include the Debye length λ_D, the electron inertial length $d_e = c/\omega_{pe}$ and the electron gyroradius ρ_e. The same notation is used for the corresponding ion kinetic scales. Collisional dissipation introduces a resistive length scale which is usually estimated from Sweet-Parker theory $\delta_{sp} \sim (\eta L_{sp})^{1/2}$ where typically it is assumed that $L_{sp} \sim L$. Image courtesy of William Daughton (Los Alamos National Laboratory).

Current understanding of the weakly collisional regime is based primarily on two-fluid and kinetic simulations for relatively small, idealized two-dimensional geometries. From this body of work, there is clear evidence that MHD breaks down when the thickness of a resistive layer δ_{sp} approaches the ion kinetic scale. In neutral sheet geometry, two-fluid theory and simulations predict an abrupt transition to faster reconnection (Ma and Bhattacharjee 1996; Cassak et al. 2005; Simakov and Chaćon 2008) when $\delta_{sp} \leq d_i$, where d_i is an ion inertial length. Experimental validation of this transition was performed in the MRX laboratory device (Yamada 2006). This is often referred to as the kinetic or fast reconnection regime because a variety of models predict reconnection rates that are weakly dependent on the system size or dissipation mechanism (Ma and Bhattacharjee 1996; Birn et al. 2001; Hesse et al. 1999; Pritchett 2001; Shay et al. 2001), although the precise scalings are controversial (Bhattacharjee et al. 2005; Daughton et al. 2006). In the kinetic regime, the structure of the diffusion region consists of an outer layer in which the ions decouple from the magnetic field, along with an inner region where the frozen-in condition is violated for the electrons. For neutral sheet geometry, both scaling arguments and simulations predict a thickness on the order of d_i for the ion diffusion region and d_e for the electron diffusion region. However, the physics responsible for controlling the length of these layers remains poorly understood. Early simulations indicated the length of the electron layer should remain on the electron scale (Hesse et al. 1999; Shay et al. 2001) while the length of the ion layers was of order $\sim 10 d_i$. In contrast, more recent large-scale, two-dimensional kinetic simulations have demonstrated highly elongated electron layers with a complicated two-scale structure (Daughton et al. 2006; Karimabadi et al. 2007; Shay et al. 2007) that can extend well beyond $\sim 10 d_i$. In many kinetic simulations, these elongated electron layers are unstable to the formation of plasmoids (Daughton et al. 2006; Karimabadi et al. 2007;

Klimas et al. 2008; Wan and Lapenta 2008) leading to a time-dependent reconnection process, while other researchers report steady reconnection with minimal influence from plasmoids (Shay et al. 2007).

In the presence of a guide field, the collisionless kinetic simulations indicate the thickness of the electron layer is comparable to the electron gyroradius (Ricci et al. 2004; Hesse et al. 2004). In large systems, these layers are also highly elongated and unstable to plasmoid formation (Drake et al. 2006b). Currently, the structure and time dependence of the electron layers are quite different within two-fluid models of reconnection than observed in the kinetic simulations. In particular, the length of the electron layer remains quite short $< d_i$ within the two-fluid models; therefore, secondary islands are not observed and the reconnection process remains steady. It is crucial to understand the origin of these differences between the kinetic and fluid models (see PRD, "Influence of Kinetic Scales on Macroscopic Evolution").

During the past decade, laboratory experiments and space observations have played an important role in validating several important aspects of these simulation results. In weakly collisional regimes, there is now overwhelming experimental evidence that electron diffusion regions are embedded within larger ion scale layers. For anti-parallel reconnection, this scale separation gives rise to an out-of-plane quadrupole magnetic field structure (Sonnerup 1979), which has been confirmed in space observations (Nagai et al. 2001; Deng and Matsumoto 2001; Øieroset et al. 2001; Mozer et al. 2002) and laboratory experiments (Ren et al. 2005; Brown et al. 2006). Laboratory experiments have also confirmed the onset of faster reconnection when the thickness of the resistive layers approach kinetic scales. In the case of anti-parallel reconnection, this is observed (Yamada 2006) when the resistive layer approaches $\delta_{sp}{\sim}d_i$ while for a strong guide field is observed (Egedal et al. 2006) near $\delta_{sp}{\sim}\rho_s$ (where ρ_s is the ion-sound Larmour radius). While it is difficult to address the length and stability of reconnection layers in laboratory experiments, there is clear evidence the driver and boundary condition play an important role (Kuritsyn et al. 2007). In space observations, there is growing evidence that electron layers that develop during reconnection can become highly elongated. Multipoint observations from the Cluster satellites show electron diffusion regions that extend $\sim 60d_i$ downstream from the x-line (Phan et al. 2007). Furthermore, recent Cluster observations have also reported secondary magnetic islands during fast reconnection (Eastwood et al. 2007) and in association with electron current sheets (Chen et al. 2009). During the next decade, researchers expect that laboratory experiments and space observations will continue to play an essential role, both in validating new ideas from simulations and suggesting entirely new directions to consider.

Some key discoveries, as mentioned above, have their antecedents in fusion physics, where the role of electron inertia and the electron pressure gradient in triggering a sawtooth crash was recognized in reduced two-fluid models of tokamak plasmas (Aydemir 1992; Ottaviani and Porcelli 1993; Wang and Bhattacharjee 1993; Kleva et al. 1995; Rogers and Zakharov 1996; Grasso et al. 1999). These reduced models take advantage of the strong guide field in a tokamak to obtain asymptotic reduction of the two-fluid equations that are computationally much more efficient in following reconnection dynamics in the semicollisional and collisional regimes. In the era of extreme-scale computing, these reduced equations are likely to continue playing an important role in the description of nonlinear reconnection dynamics in hot fusion plasmas where the issues of scale separation (previously discussed) are as important as they are in space and astrophysical plasmas.

PRIORITY RESEARCH DIRECTIONS

Much of the current understanding developed over the past decade was obtained from relatively small two-dimensional systems using both fluid and kinetic descriptions. Presently, it remains unclear how these idealized results will extend to large-scale, three-dimensional systems. Even with extreme-scale computing, a first-principles three-dimensional kinetic treatment of reconnection in hydrogen plasmas will be limited to fairly small systems. Progress in modeling most real applications will require understanding the key physics sufficiently well to capture within reduced descriptions and to infer reliable scalings. With this goal in focus, this panel identified the following four PRDs:

- influence of the kinetic scales on the large-scale evolution
- reconnection and magnetic island dynamics in three-dimensional geometries
- energy partition and particle acceleration
- reconnection in relativistic plasmas.

In the following sections, each of these PRDs is discussed in more detail, along with computational and algorithmic developments needed to effectively achieve progress.

Influence of Kinetic Scales on Macroscopic Evolution

Summary of Research Direction

In a large fusion machine or in the earth's magnetosphere, typical macroscopic scales are $\sim 10^3$ times larger than the ion kinetic scale. Macroscopic scales in the solar corona and many astrophysical applications are upward to $\sim 10^9$ times larger than the ion scale. In contrast, collisionless kinetic simulations have established that electron scale layers are expected within the ion scale reconnection layers. Given this huge range of scales, *what is the interaction between the local kinetic scales and the global MHD evolution?* The coupling between these disparate scales could potentially go in both directions, with the small-scale features influencing the macroscopic dissipation and time dependence, and/or the large-scale magnetic geometry influencing the structure and development of new kinetic layers. To make progress on these questions, researchers must first resolve the outstanding discrepancies between kinetic and two-fluid simulations in weakly collisional regimes. As discussed earlier, the kinetic simulations feature elongated electron layers that are generally unstable to secondary magnetic islands, while the reconnection layers in two-fluid simulations remain in a steady open x-point configuration. Thus, even within these modest-sized $\sim 100 d_i$ simulation domains, there is no clear consensus on the minimal physics required to capture the structure and dynamics of reconnection layers in weakly collisional regimes.

Scientific Challenges

Making progress on these questions will require ambitious new research efforts using both kinetic and fluid approaches. One obvious direction is to critically reconsider the closure approximations and generalized Ohm's law. Most two-fluid models typically include three nonideal terms: electron inertia, electrical resistivity, and a simple electron viscosity. In contrast, collisionless kinetic simulations have demonstrated that the dominant nonideal terms involve off-diagonal terms in the electron pressure tensor (Hesse et al. 1999; Pritchett 2001), which are not well described by any of the existing terms. More work is needed to develop new fluid closures that can capture these physics terms and bring the structural evolution within two-fluid simulations in better alignment with kinetic results.

Working from the kinetic description, it is possible to include Coulomb collisions within kinetic particle-in-cell (PIC) simulations using standard Monte Carlo techniques (Takizuka and Abe 1977). This approach permits a direct solution of the full plasma kinetic equation, which forms the theoretical basis for all fluid models. Although computationally expensive, sufficient computing power is now available to employ this method for two-dimensional reconnection studies, thus allowing a fully rigorous and self-consistent treatment of Coulomb collisions in all parameter regimes (Daughton et al. 2009). This approach is particularly useful for understanding parameter regimes where the reconnection electric field is comparable to or larger than the runaway limit, as this condition is widely prevalent and the collisional resistivity from transport theory is no longer valid. These types of first-principles calculations may serve as a valuable guidepost for developing and testing better fluid closures in weakly collisional regimes.

Potential Scientific Impacts and Outcomes

Advancements in this research direction should be validated against well-diagnosed laboratory reconnection experiments (Yamada 2007). These experiments typically involve driven reconnection in parameter regimes where weak Coulomb collisions are important. To make meaningful comparisons between experiment and simulation, it is crucial to properly treat the experimental boundary conditions. For example, boundary conditions relevant to the MRX were recently employed in two-dimensional collisionless kinetic simulations (Dorfman et al. 2008). In these simulations, the observed electron layers was a factor of $\sim(3-5)\times$ thinner than measured in the actual experiment (Ji et al. 2008). The two leading candidates for explaining this discrepancy are weak Coulomb collisions and plasma instabilities. There is presently an active effort to address both of these possibilities by exploiting the newest petaflop-scale machines. Figure 18 shows a preliminary three-dimensional simulation of this configuration performed using the kinetic plasma simulation code VPIC (Bowers et al. 2008a) running on LANL's Roadrunner, the first petaflop supercomputer. In these calculations, reconnection is driven by reducing the flux core current in a manner similar to the actual experiment. Shown are the magnetic field structure, ion reconnection outflow jets, and electron scale current layer. In the next few years, these kinetic simulations will allow direct comparison of these features with the actual experiment to understand the influence of collisionality and plasma instabilities on both the ion and electron layers. With this type of detailed experimental validation, this problem could be an attractive test case for developing new reduced fluid descriptions.

After better structural agreement is obtained between the two-fluid and kinetic calculations in relatively modest-sized systems $\sim 100d_i$, it is crucial to perform scaling studies over a broad range of systems $> 1000d_i$ to understand the scaling and time dependence of reconnection in large systems. These types of studies should be performed with a variety of macroscopic drivers, including open boundary conditions (Daughton et al. 2006; Klimas et al. 2008) that attempt to mimic much larger systems by allowing plasma and magnetic flux to exit the system. Certain aspects of the reconnection physics may not be possible to treat accurately with fluid models. In this case, progress may require reduced kinetic descriptions such as gyrokinetic or possibility embedding a kinetic simulation within a larger fluid simulation. In both fluid and kinetic approaches, there is ample room for better algorithms capable of treating the multiscale problem (adaptive mesh refinement [AMR], implicit, etc.).

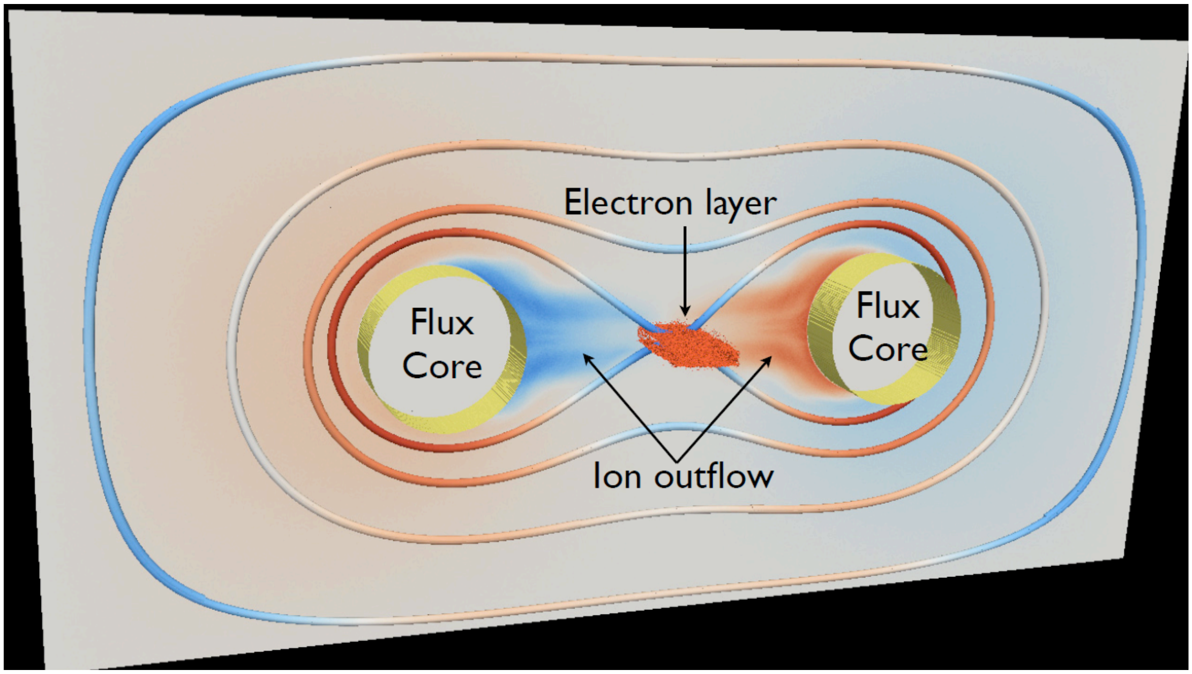


Figure 18. Fully kinetic modeling of laboratory reconnection experiments such as the Magnetic Reconnection Experiment is now feasible using realistic boundary conditions. This preliminary example was performed using the VPIC code running on Roadrunner, the new petaflop supercomputer at Los Alamos National Laboratory. The reconnection process is driven by reducing currents within the flux cores in a manner similar to the actual experiment. This preliminary three-dimensional simulation was performed with artificial mass ratio $m_i/m_e = 100$, while full-scale calculations are expected to permit $m_i/m_e \sim 400$ in the near future. Image courtesy of William Daughton (Los Alamos National Laboratory).

Potential Impact on Fusion Energy Science

Progress from these research efforts is expected to have significant practical importance to fusion energy science. A better understanding of how to accurately model reconnection with two-fluid closures could have direct relevance in modeling of tearing modes and sawteeth oscillations in tokamaks, as well as magnetic relaxation in reversed-field pinches (RFPs) and compact tori. Comparison of reconnection studies by gyrokinetic and fully electromagnetic PIC methods will be useful in delineating the domain of validity of predictions by the gyrokinetic model, which has already proved to be a very useful tool in studies of tokamak turbulence. Furthermore, the development of extended fluid equations that incorporate essential kinetic effects is of great interest in heliophysics applications, where global simulations based on fluid equations are often used as predictive tools for developing space missions and comparison with data obtained from satellites.

Reconnection and Magnetic Island Dynamics in Three-Dimensions

Summary of Research Direction

Geometry, boundary conditions, and intrinsically three-dimensional effects all play an important role in magnetic reconnection. In toroidal fusion plasmas, magnetic islands tend to develop near closed field lines, and their nonlinear stability and dynamics depend critically on two-fluid, kinetic, and transport effects that are not sufficiently understood. In three-dimensional space and astrophysical plasmas, the important role of null points and null-null lines or quasi-separatrix layers in controlling reconnection dynamics remain areas of active research with many open questions. To address these questions, computer simulations are particularly important tools of discovery because the primary source of experimental knowledge are in-situ measurements from a sparse group or multiple satellites in the magnetosphere or large-scale visual images of the sun, which do not provide any information on the small scales.

Scientific Challenges

As mentioned in the "Introduction and Current Status" section of this panel report, there is growing evidence that in large-scale systems a single reconnection layer may spontaneously break up into multiple interacting reconnection sites through the formation of secondary magnetic islands. This same basic expectation now appears to hold for collisional high Lundquist number regimes in which MHD is valid, as well as weakly collisional kinetic regimes. In three-dimensions, secondary magnetic islands should correspond to extended flux ropes that may interact in a variety of complicated ways (Yin et al. 2008). Furthermore, an assortment of additional secondary instabilities may be possible depending on the global configuration and boundary conditions (ballooning, current driven instabilities, kinetic instabilities, etc.).

Extreme-scale computing will permit many of these issues to be examined at some level using both fluid and kinetic simulations. To start, these efforts should focus on problems in which a limited subset of these issues can be isolated and systematically studied. In large systems, even simple initial configurations can give rise to very complicated magnetic island interactions. To illustrate this complexity, it is instructive to consider simple current sheet geometry. With a guide field, the tearing mode is unstable at a discrete number of resonant $\mathbf{k} \cdot \mathbf{B} = 0$ surfaces across the layer, depending on the system size. For the limit of electron-positron plasmas $m_i = m_e$, this problem is now possible to study in fairly large systems using fully kinetic simulations on petaflop-scale computers. As illustrated in Figure 19, the tearing modes at different resonant surfaces across the layer leads to formation of flux ropes across the layer that interact and coalesce in complex ways. During this evolution, new current sheets are formed that are in turn unstable to new tearing-like modes at ranges of different angles (corresponding to difference resonant surfaces).

To simulate the three-dimensional formation and interaction of magnetic islands, it is crucial to develop fluid and kinetic algorithms that scale well on new petaflop computers. In this regard, fully kinetic simulations are well positioned to use the new petascale architectures (Bowers et al. 2008b), but at high-mass ratio, the feasible problem sizes will still be limited. Fluid simulations should be capable of modeling larger domains, but may prove even more challenging to efficiently use extreme-scale computers. Modern implicit algorithms have demonstrated scaling up to thousands of cores (Chaćon 2008), but most fluid simulations of reconnection are presently performed at much smaller scales with larger dissipation (resistivity or viscosity) than realistic problems. Moving toward more realistic parameters, implicit methods have a significant advantage. For example, with three-dimensional resistive MHD, the computational cost scales as $\sim S^{5/2}$ with explicit methods in comparison to $\sim S^2$ for implicit simulations (Chaćon 2008). Even with implicit techniques, the main focus is expected to change from weak to strong scaling properties for the simulation codes, and this is still a formidable challenge. Furthermore, it is important to show that spatial adaptivity can be combined effectively with implicit techniques, which has been done in the context of reduced MHD (Philip et al. 2008), but needs to be demonstrated in more general settings. There are also several practical implementation concerns for using adaptive techniques for extreme-scale machines because of the increased overhead and unstructured nature of the data. Finally, realistic treatment of boundary conditions is important for modeling both three-dimensional laboratory reconnection and larger open systems in space and astrophysics; therefore, algorithms must be chosen that allow considerable flexibility.

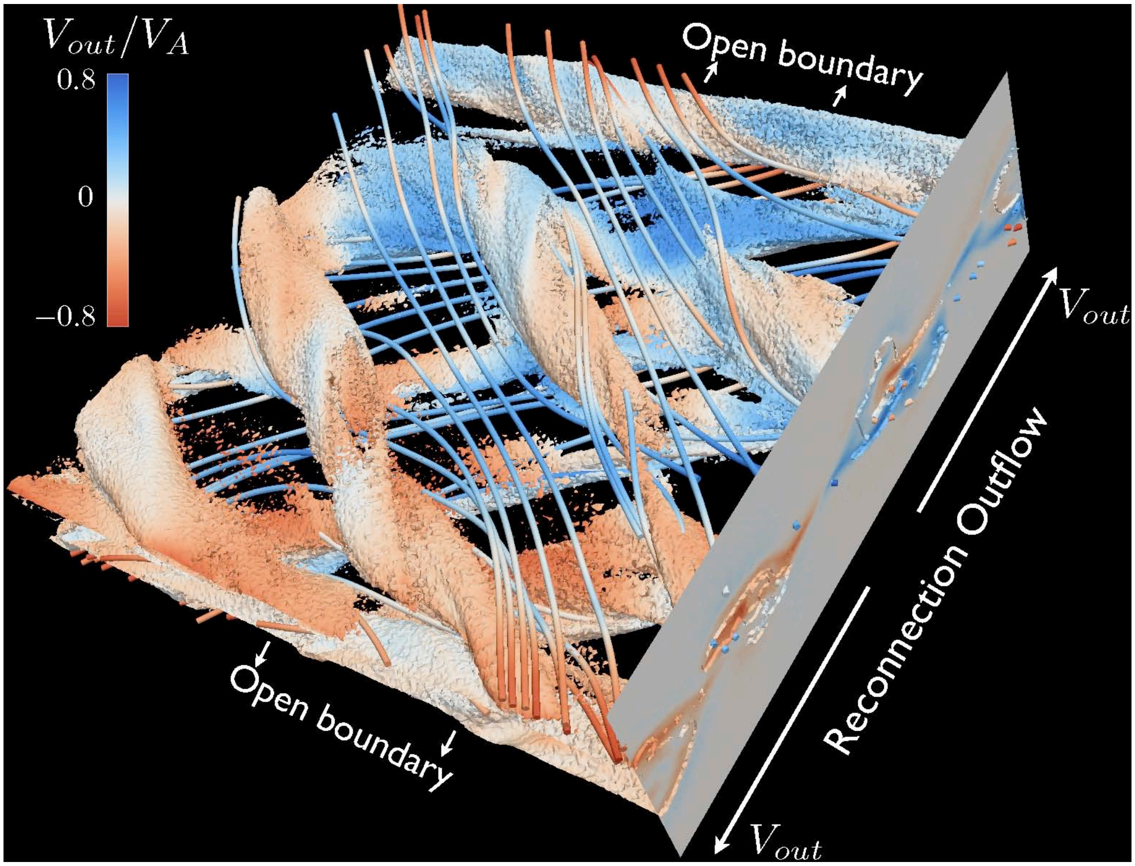


Figure 19. Three-dimensional kinetic simulation of magnetic reconnection in a large-scale electron-positron plasma with a guide field equal to the reconnecting field. This simulation was performed on the Roadrunner supercomputer at Los Alamos National Laboratory using the VPIC code and employing open boundary conditions (Daughton et al. 2006). Shown are density isosurfaces colored by the reconnection outflow velcocity. Magnetic islands develop at resonant surfaces across the layer leading to complex interactions of flux ropes over a range of different angles and spatial scales. Image courtesy of William Daughton (Los Alamos National Laboratory).

Potential Scientific Impacts and Outcomes

Given the uncertainties and controversies already discussed within two-dimensional approximations, it is clear that scientists' understanding of three-dimensional reconnection is still in its infancy. Most kinetic and two-fluid simulation studies of reconnection have been limited to simplified two-dimensional geometries. While three-dimensional resistive MHD simulations are quite common, the range of feasible Lundquist numbers for resolved simulations are many orders of magnitude smaller than applications in space and astrophysics or in a large tokamak. Given these limitations, it is not even clear whether scientists know the right questions to ask regarding three-dimensional reconnection in large-scale systems. However, exascale computing should permit significant steps forward by finally allowing well-resolved high Lundquist MHD simulations and sufficiently large two-fluid and kinetic studies, to permit a range of interesting dynamical processes to compete and interact. These simulations are expected to have a direct impact on a variety issues that are inherently three-dimensional. For example, can microscopic kinetic plasma instabilities modify the structure or influence the reconnection rate? What is the net influence of larger scale secondary instabilities, such as kink, ballooning, or secondary-island formation? Can the complex interactions between these various modes play a role in controlling the average dissipation rate or in accelerating nonthermal particles? Answers to these questions have direct relevance to understanding reconnection in a wide range of physical systems ranging from the solar corona to the earth's magnetosphere. In particular, understanding the three-dimensional formation and interaction of magnetic islands is expected to have important scientific impacts across a broad range of problems in space, astrophysical, and fusion relevant plasmas.

Potential Impact on Fusion Energy Science

Progress in understanding magnetic islands dynamics has strong relevance to fusion energy science, where much remains to be understood about the sawtooth oscillations observed in tokamaks and RFPs. In cylindrical geometry, two-fluid simulations show an impulsive nonlinear phase of the $m = 1$ resistive kink mode, which is in better agreement with experimental observations than a purely resistive model. In this regard, the cylindrical $m = 1$ tearing mode represents an attractive benchmark case to verify existing computational models—such as NIMROD and MRC—using profiles of relevance to experiments such as the Tokamak Experiment for Technology Oriented Research (TEXTOR). However, many questions remain open that require fully resolved three-dimensional toroidal simulations to improve current understanding and validate models against experiments. One particularly important question is the q_0 problem. While numerical simulations show complete reconnection, flattening out the q profile to 1, in most tokamaks (Tokamak Fusion Test Reactor [TFTR], JET, and TEXTOR) observations indicate that q remains nearly fixed and below 1 during the entire cycle. Various mechanisms have been proposed to explain this discrepancy; e.g., by the excitation of secondary ballooning instabilities, an intrinsically three-dimensional process by which the pressure profile is relaxed while the current profile is not. The role of heat conduction on the pressure profile is also very important in this context. Interactions of nonlinear tearing modes at multiple rational surfaces can modify substantially the current profile, destabilizing kink instabilities that can cause major disruptions in a tokamak. To resolve these questions, scientists need to develop three-dimensional toroidal simulations with sufficient spatial and temporal resolution (probably with AMR).

Energy Partition and Particle Acceleration

Summary of Research Direction

As reconnection proceeds, a portion of the stored magnetic energy in the system is converted to plasma kinetic energy, including bulk flows, ion and electron thermal heating, and highly energetic nonthermal tails. While a tremendous amount of effort has been directed towards understanding the overall dissipation rate in various systems, the detailed partition of kinetic energy is equally interesting and perhaps even more challenging. In particular, it is beginning to appear that a variety of different reconnection scenarios may yield similar fast Alvénic reconnection rates in large-scale systems. If this conjecture is confirmed, understanding the energy partition in these different scenarios may be a better measure of progress than focusing solely on the dissipation rate. Furthermore, the energy partition is of great practical importance in understanding the global evolution for a wide range of physical applications.

Scientific Challenges

In examining the partition of kinetic energy, conversion to bulk flow is fairly well understood because the maximum outflow from a reconnection site is energetically constrained by the upstream Alfvén speed. However, this maximum outflow can vary significantly depending on the downstream pressure gradients (Priest and Forbes 2000) that in turn depend on both the equation of state and the downstream boundary conditions. For incompressible fluid calculations, observed outflows are very near the Alfvén velocity, but can be 40% to 60% less in kinetic simulations (Pritchett 2001; Karimabadi et al. 2007). Both ions and electrons are heated in the downstream region, with ions generally gaining more thermal energy. Strong ion heating has been clearly observed in laboratory reconnection experiments (Hsu et al. 2000). Understanding the ion heating mechanisms may also have strong relevance to reversed field pinches, where anomalous ion heating is observed in conjunction reconnection events (Gangadhara et al. 2008). However, detailed comparisons between various fluid and kinetic simulations on this issue are still needed. Some of the potential heating mechanisms can be studied two-dimensionally, but in general three-dimensional simulations will be required to address the full range of possibilities.

Perhaps the most outstanding issue in this PRD is the formation of nonthermal tails during reconnection. Satellite observations in the earth's magnetosphere have long reported (Terasawa and Nishida 1976; Baker and Stone 1976) highly energetic electrons; more recently, these have been clearly observed in the vicinity of reconnection sites (Øieroset et al. 2002). Reconnection in solar flares can result in upwards to 50% of the total energy release in energetic electrons (Lin and Hudson 1971; Lin et al. 2003). In these regimes, the energy density of nonthermal particles may be sufficient to cause a non-negligible feedback on the overall reconnection dynamics. A large number of different acceleration mechanisms have been proposed (see Aschwanden [2002] for a recent review), including direct acceleration from the reconnection parallel electric fields, indirect acceleration via standing shocks, Fermi acceleration by turbulence, resonant wave-particle interactions, or magnetic islands. Kinetic simulations indicate that significant electron energization can occur during the coalescence of large magnetic islands (Pritchett 2008) or in the magnetic field pile-up regions where the reconnection outflow jet collides with pre-existing plasma (Hoshino et al. 2001; Pritchett 2008). Other researchers have proposed that a Fermi acceleration process involving multiple magnetic islands may play a crucial role (Drake et al. 2006a). All of these possibilities connect well with the PRD, "Reconnection and Magnetic Island Dynamics in Three-Dimensions." Thus, the type of simulation illustrated in Figure 19 is expected to be of interest for particle acceleration studies as well as the issue of island dynamics. Recent satellite observations have reported that energetic electrons are indeed correlated with magnetic islands (Chen et al. 2008), but more work is needed to clarify the precise acceleration mechanism.

For certain acceleration mechanisms, progress can be made using test particle treatments within fluid simulations. However, many of the promising ideas will require self-consistent kinetic simulations for adequate testing. These simulations are generally complex, and usually it is only feasible to save the full particle data at limited time intervals. When nonthermal tails are observed in these types of simulations, it is essential to understand the physics of the acceleration process to confidently extrapolate to real applications. Thus, there is a tremendous need for innovative new diagnostic and data analysis techniques.

Potential Scientific Impacts and Outcomes

A better understanding of the energy partition resulting from large-scale reconnection processes is likely to have scientific impacts for problems ranging from the earth's magnetosphere, the solar corona, and fusion experiments. In particular, the generation of highly energetic nonthermal particle distributions cut across a wide range of applications in space, astrophysical, and laboratory plasmas. While it is quite likely there are different acceleration mechanisms operative in these various regimes, it is also possible that similarities and common themes may emerge. In addition, advancements in the computational tools and diagnostics to study acceleration mechanisms are expected to be broadly applicable.

Potential Impact on Fusion Energy Science

The partitioning of energy released by reconnection between electrons and ions is not well understood in most fusion experiments. What complicates the interpretation of experimental results is the difficulty in separating the effects of turbulent transport from those due to reconnection. However, there are instances where reconnection is observed to be strongly correlated with ion heating, as has been seen in RFPs. Furthermore, this research could potentially lead to a better understanding of high-energy electrons generated during large sawteeth crashes or disruptions when super-Dreicer electric fields are produced.

Reconnection in Relativistic Plasmas

Summary of Research Direction

Hot plasmas and nonthermal highly relativistic particles are known to exist throughout the universe. For many applications, plasmas are often highly magnetized and include electron-positron as well as electron-proton plasmas. As recent observational measurements proceed to progressively higher resolution, there is growing recognition that magnetic reconnection may be a key scientific issue in high-energy astrophysics. For example, the plasma temperature in an accretion disk around a black hole reaches up to 10^{11} Kelvin and has been attributed to reconnection, while the radiation from such hot plasma is known to be important in the dynamics of the accretion disk. Diverse phenomena such as gamma-ray bursts, relativistic winds and shocks around radio pulsars, soft gamma-ray repeaters and plerionic nebulae (Kirk and Skjaeraasen 2003) also invoke similar concepts involving magnetic reconnection. In these various applications, reconnection is estimated to play an important role in accelerating and heating particles.

Scientific Challenges

The issue of highly relativistic electrons is important not only in astrophysical plasmas, but also in tokamaks where sawteeth and major disruptions can accelerate electrons up to ~100 MeV (Jarvis et al. 1988; Gill 1993). Collisions between these relativistic electrons and background ions may give rise to copious electron-position pair production (Helander et al. 2002). Clearly, a variety of relativistic plasma phenomena may emerge in these regimes that are completely different than the nonrelativistic limit. However, current understanding of relativistic plasmas is still quite limited. How does reconnection proceed and what mechanisms determine the reconnection rate? What happens when both the flow speed and the Alfvén speed are relativistic? How are particles accelerated in such a system?

More theoretical and simulation work is needed to address these questions. Relativistic kinetic simulations are especially well suited for understanding the physics of magnetic reconnection in these regimes. Recent large-scale, two-dimensional PIC simulations have addressed the microphysical processes mediating the particle energization in relativistic current sheets on the inertial scale (Zenitani and Hoshino 2001; Jaroschek et al. 2004). From these two-dimensional simulations, researchers found that relativistic reconnection is a powerful engine producing nonthermal particles, leading to well developed, power-law energy distributions. However, the physics of relativistic reconnection is inherently multidimensional, and a large-scale, computationally expensive three-dimensional kinetic simulation is required to fully address the problem (Zenitani and Hoshino 2005; Zenitani and Hoshino 2007). For electron-positron plasmas, it is now feasible to perform relativistic kinetic simulations with fairly large simulation volumes (~500× the inertial scale) and directly study the influence of kinetic scale dynamics on the macroscopic evolution (Yin et al. 2008).

Potential Scientific Impacts and Outcomes

The study of reconnection processes in relativistic plasmas has received far less attention than in the non-relativistic limit. However, this regime is crucial to understand because magnetic reconnection in highly relativistic plasmas is estimated to occur in a large variety of astrophysical problems. Currently, magnetic reconnection is often invoked at almost the "cartoon level" when it is required. Extreme-scale computing is expected to establish a much better understanding of some basic issues within relativistic regimes, and thus provide additional constraints regarding where and how reconnection may be invoked as new astrophysical models are developed.

Potential Impact on Fusion Energy Science

As already mentioned, a better understanding of the formation and dynamics of relativistic electrons has direct application to modeling runaway populations that can occur during large sawteeth and major disruptions in tokamaks. As this research proceeds, it will be interesting to compare the relativistic kinetic simulations with relativistic fluid approaches. These efforts are important for developing better relativistic fluid closures, which could be useful for a variety of applications. In particular, these techniques and tools may overlap with problems in high-energy density physics, such as electron fast ignition where highly relativistic electron populations are generated along with intense self-generated magnetic fields.

CONCLUSIONS

The scientific and computational advancements resulting from these research efforts are expected to impact fusion energy science in several ways. First, this research will help clarify the essential physics needed to properly model reconnection in fusion-relevant plasmas, and how to best incorporate these physical effects into reduced fluid models. This is important because the realization of high-performance regimes with superior energy confinement in fusion plasmas—such as the ITER—require their operation in a stable, quasi-steady state in which the size and dynamics of magnetic islands can be controlled by manipulating the background current and pressure profiles. For instance, sawtooth crashes, which represent an important paradigm for fast reconnection in tokamak plasmas, can trigger the formation of neoclassical tearing mode islands to produce major disruptions. Major disruptions can also occur in tokamak plasmas due to the coupling of tearing islands on multiple rational surfaces that can modify the background current profile and trigger kink modes that can potentially terminate a discharge. Understanding the behavior of these magnetic islands, and resolving the separate physics of ions and electrons, is critical to controlling them. The computational challenge of predicting the time evolution in realistic toroidal geometry, while resolving ion and electron dynamics within, as well as outside, of the islands will need the resources of extreme-scale computing. In this regard, the computational and algorithmic advances needed to make progress in reconnection physics may directly benefit a wide range of problems in fusion energy science.

Magnetic reconnection remains one of the most fundamental and widespread processes in basic plasma physics. Many theoretical and computational challenges arise from the immense separation of spatial and temporal scales that result from coupling nonideal diffusion regions to the larger-scale dynamics. Over the past 50 years, progress in reconnection research has benefited greatly from numerical simulations. This trend is accelerating with the advent of petascale computers and is expected to continue as exascale computers become available in the next decade. In this panel report, four PRDs were identified in which extreme-scale computing is expected to play a crucial role in scientific advancement of the field:

- influence of the kinetic scales on the large-scale evolution
- reconnection and magnetic island dynamics in three-dimensional geometries
- energy partition and particle acceleration
- reconnection in relativistic plasmas.

For some of the scientific challenges, exascale computing has the potential to provide a definitive resolution. However, in many applications the scale separations are so enormous, both fluid and kinetic simulations will remain beyond reach for the foreseeable future. To make progress, researchers will have to carefully select simulation parameters within fluid and kinetic simulations and work toward understanding the essential physics so that reliable scalings can be developed. Testing the predictions of such simulations with controlled, dedicated laboratory experiments is crucial in developing confidence in the predictive capabilities of the computational models. In this regard, analytic theory, laboratory experiments, and space observations all have critical roles in advancing scientific understanding of magnetic reconnection.

Advancements in reconnection physics are expected to impact both scientific understanding and practical modeling capabilities for a wide range of problems in space, astrophysical, and fusion-relevant plasmas. While this panel report focused on PRDs that are within reconnection physics, it is important to emphasize the study of reconnection in large-scale systems is increasingly interconnected with other forefront areas in basic plasma physics, including nonlinear waves, collisionless shocks, turbulence, and transport. The authors of this report expect many computational and algorithmic challenges discussed in this report will also be applicable to these broader issues. Thus, the efforts described in this report to use extreme-scale computing for reconnection problems will likely benefit a range of topics in basic plasma physics.

CROSSCUTTING CHALLENGES

- **ALGORITHMS FOR FUSION ENERGY SCIENCES AT EXTREME SCALE**
- **DATA ANALYSIS, MANAGEMENT, AND VISUALIZATION IN FUSION ENERGY SCIENCE**
- **MATHEMATICAL FORMULATIONS**
- **PROGRAMMING MODELS, FRAMEWORKS, AND TOOLS**

ALGORITHMS FOR FUSION ENERGY SCIENCES AT EXTREME SCALE

Lead: David Keyes, Columbia University

Panel Members: Mark Adams, Columbia University; David Brown, Lawrence Livermore National Laboratory; Lori Diachin, Lawrence Livermore National Laboratory; Stephane Ethier, Princeton Plasma Physics Laboratory; Xiaoye Li, Lawrence Berkeley National Laboratory; Lois Curfman McInnes, Argonne National Laboratory; John Shadid, Sandia National Laboratories; Mayya Tokman, University of California, Merced; and Lois Curfman McInnes, Argonne National Laboratory

SUMMARY

There are many motivations for scaling simulations in fusion energy sciences to the expanding architectural extremes of the coming decade. These include better resolving the full ranges of length or time scales in a model, accommodating physical effects with greater fidelity, allowing the model degrees of freedom in all relevant dimensions, optimizing or controlling (by solving inverse problems) plasma scenarios that are adequately predicted by forward models, and quantifying uncertainty. In a "game changing" comprehensive sense for fusion energy sciences, advances in all of these areas are needed to enable scientific discovery of new plasma phenomena with associated understanding that emerges only upon integration. However, as applications stretch to take full advantage of extreme architectures, the superlinear complexity of typical current algorithms does not allow them to scale indefinitely, even though memory capacity would allow it.

Extreme scales therefore put a premium on finding "optimal" algorithms with complexities that are—at worst—log-linear in problem size because (by Amdahl's Law) any suboptimal component will ultimately dominate the execution profile. The availability of high-capability architectures makes algorithms more, not less, important. It is fortunate that algorithms such as linear solvers have kept pace with extreme scales to date, and optimal versions are known for systems arising from some popular field and particle formulations of plasma physics. This provides feedback from the algorithms community to the physics community. It behooves modelers to attempt to cast their simulations in terms of formulations of known scalability; for instance, in terms of preconditioners for various second- and fourth-order formulations of magnetohydrodynamics (MHD) built up from scalar Poisson solves or in terms of domain decompositions that slice a physical domain in all dimensions rather than in only toroidal angle or radial coordinates. Overall, however, it should be kept in mind that the expected architectural complexity of exascale systems will likely pose formidable new challenges.

Fusion energy physicists want to scale simulations in a variety of formulations such as particles, conservation equations in primitive variables, or evolution equations for probability distribution functions for a number of reasons, including the following:

- <u>Better resolve the full, natural range of length or time scales in a model.</u> Often, analysis, asymptotics, or heuristics can be used to present to the computer a model resolving only a subrange of the physical scales and this should be exploited when possible. Even in such cases, however, a more fully resolved spectrum of scales may be needed to tune or validate the limited resolution model.

- Accommodate physical effects with greater fidelity. Often, physical effects are omitted or homogenized due to their computational cost. However, such lower fidelity models may be more interpolative than predictive and need to be compared with "first-principles" simulations that retain all of the known physical details that are feasible to incorporate.
- Allow the model degrees of freedom in all relevant dimensions. Often, symmetry is invoked to remove one or more dimensions from a model to substantially reduce computational cost. However, repressing nonsymmetric behavior as the plasma evolves in time may suppress relevant physical mechanisms, relocate stability boundaries, and even (e.g., in the case of turbulent flows) invert energy cascades.
- Better isolate artificial boundary conditions. Models are often run with reduced physical domains but imposition of artificial boundary conditions too close to the phenomena of interest may contaminate the reliability of prediction from the outset or after waves arrive at the artificial boundary and are partially reflected back. A succession of larger domains is required to properly interpret the results of such simulations and bound the associated uncertainty.
- Solve an inverse problem, or perform data assimilation. Models often contain unknown parameters that can be estimated if model output is available from experimental or observational data. This is known as an inverse problem, and it typically requires many runs of a suitably augmented forward problem inside an optimization loop. A related concept is that of data assimilation, in which an imperfectly known model can be nudged towards predictivity by the incorporation of data with a similar cost of looping around the forward model.
- Perform optimization or control. Models often contain geometric parameters (such as boundary definition) or control parameters (such as boundary conditions) that are at the disposal of the engineer, who wishes to choose them to produce a specific behavior; e.g., to move away from instabilities. Such computational optimization can be performed at a cost of multiple runs of a suitably augmented forward problem.
- Quantify uncertainty. Models often contain uncertain parameters, whose effects on model outputs need to be estimated and bounded. Sensitivity analysis and other forms of uncertainty quantification (UQ) typically require many runs of a suitably augmented forward problem.

For MHD problems in fusion energy sciences, the ranges of physical scales in space and time that must be represented in important applications grow as powers of the Lundquist number, which in turn grows with device size, magnetic field strength, and operating temperature. Without advances in algorithms, the brute force space-time resolution requirements for a uniform grid and explicit time treatment of relevant scales in a tokamak grow 12 orders of magnitude from a small device such as CDX-U (which can be modeled today) to one of the ITER scale. Over the coming decade, Moore's law provides at most three orders of magnitude of help in conquering this gulf, and almost all is in the form of extra concurrency. Mathematical advances in algorithms and computer science advances in allowing the algorithms to scale gracefully to a thousand-fold greater thread concurrency are essential to the goal of predictive simulation of ITER-scale plasmas and many other systems of interest in FES of both high- and low-energy density.

Significant research challenges remain in developing and scaling multiscale and multiphysics codes to the performance levels needed for fundamental studies, as well as engineering and design use. These research challenges provide good crosscutting collaborative opportunities in that they are similar (but not identical) to those faced in other science and engineering domains; thus, many techniques can be leveraged from other disciplines through close collaborations between those with a deep understanding of the physics and those with state-of-the-art expertise in algorithms and the hardware and software architectures of the computing platforms.

Addressing the scientific priorities of the U.S. Department of Energy's (DOE) Fusion Energy Sciences Program via fusion simulations at extreme scales will require investment in new codes and software tools at all levels—from physics to algorithms to data management and understanding. This is a consequence of Amdahl's Law, which allows no component to be left serialized or replicated *and* also because of the shift from pure distributed memory programming to mixed locally shared/globally distributed memory programming.

DOE has the needed expertise and is capable of deploying computational infrastructure to begin building a new generation of fusion simulation codes that could shift paradigms in scientists' ability to harvest new scientific insights from the ITER and other advanced fusion experiments. To reap the benefits of an approach that uses computation to reduce the scope, cost, and time latency of required experimentation will, in turn, require a sustained software development activity and a corresponding buildup of human resources to accompany the extreme-scale facilities.

While fusion modelers approach the challenges of scaling up to accomplish the seven motivations listed in the preceding discussion, and others that will emerge beyond the visible horizon, the architectural landscape is evolving away from the *Pax Romana* of distributed memory programming from the past two decades. Moving beyond the current ability to program at the chip level, it will be necessary to employ many core architectures efficiently and cost effectively, simulation codes and the libraries that support them must be ported to a hybrid architectural environment that is shared-memory at the lowest levels. It is not yet clear that emerging architectures are sufficiently well balanced in memory bandwidth relative to processing power to justify the substantial effort to recode current algorithms; however, new algorithms may be better suited to the new architectures. If the temporal urgency associated with the need to achieve major progress on the fusion energy sciences priority research directions (PRDs) require staying on the extrapolative path of Moore's Law, then the co-design of architecture, software libraries, and applications will ultimately be necessary. This will represent a fundamental infringement on the doctrine of "separation of concerns" that has governed aesthetics in computer programming for decades. It is also often characterized as being an expensive and high-risk approach. Nevertheless, for applications as important as fusion energy simulations, every path forward must be carefully considered.

In view of these challenges, the following five chief PRDs emerge for scalable algorithms.

Optimal Algorithms for Optimal Representations

Full adaptivity in the sense of *h* (mesh refinement), *p* (discretization order), and *r* (mesh relocation) should be employed in space and time, according to the local smoothness of fields to be represented, to get the most "science per Watt" out of a simulation. This requires estimating and equidistributing truncation errors, balancing loads dynamically and in place, and managing and converting between different representations.

Current optimal complexity algorithms for nonadaptive discretizations of the underlying physics must be extended to such optimal representations.[1]

Multiphysics, Multiscale Algorithms

Algorithms that allow self-consistent coupling/integration of multiphysics models across all relevant scales enable better focus on physical questions— with greater accuracy and free of concern about numerical instabilities, splitting errors, and longer windows of integration due to suppression of stability-limiting fast scales. This requires scalable implicit methods and high-order interpolations between representations (e.g., from fields to particles and vice versa). Current state-of-the-art nonlinearly implicit solvers for individual physical systems over narrow ranges of scale must be extended to tightly couple/integrate all relevant interacting physics and appropriate scales.

Optimization and Reduced-Order Modeling

Robust (noise- and error-tolerant) optimization algorithms are needed for high-dimensional multiphysics models for optimal design, control, parameter estimation, and the mapping of stability boundaries. These optimization tools are often effectively "outer loops" around analysis codes, which puts a premium on efficient implementations of the analyses, as described in the first three PRDs in this section. Required are deterministic and stochastic techniques for derivative-free methods, adjoint-based derivative methods, and preconditioners for saddle-point systems. Calibrated with first-principles simulations, reduced-order models can be parameterized for sufficiently narrow regimes to provide detection and control capabilities in real time for understanding and control. This requires physics-based developments beyond current general-purpose models based on principal component analysis or proper orthogonal decomposition.

Uncertainty Quantification and Reduction

Models contain uncertainties in initial conditions, boundary conditions, coefficients, and/or forcings, coming from observations or other simulations that feed into the model. Sometimes the mathematical forms of the models themselves are uncertain. Incorporation of observations can improve uncertain models, with observational errors balancing model and numerical errors for more efficient computation. Needed are deterministic UQ tools based on sensitivity and adjoint techniques, probabilistic approaches based on sampling methods, and direct propagation of probability density functions from inputs to outputs.

Lower Threshold of Expertise for Using Optimal Algorithms on Extreme Architectures

Software for extreme-scale environments must offer multilevel ("incremental adoption") user interfaces. With proper interfaces to widely used (and therefore thoroughly debugged) modules, software will work as close as possible to expert reliability, while auto-tuning (or being tunable by expert users) for high performance. With such software components available, fusion physicists will work more productively

[1]Throughout this discussion, "complexity" should be considered in the parallel context, with respect to both its computation and communication components, as measured along the critical path of the parallel execution. It is understood that the balance of resources in terms of processing rates in CPUs and auxiliary processors, data movement and staging rates in main and replicated memories, and communication rates (bandwidths and reciprocal latencies) is never perfectly matched to application requirements and that resources that are in excess at any given instant of an execution may reasonably be employed in a way that might be "suboptimal" relative to sequential execution but that reduces overall parallel execution cost.

and understand the performance of their software tools, therefore focusing more on the physics challenges and less on software issues.

SCIENTIFIC CHALLENGES AND RESEARCH APPROACHES

The five PRDs summarized above require advances in mathematical modeling, algorithms, numerical analysis, parallel programming paradigms, and scientific software engineering. The different PRDs emphasize advances in these fundamental disciplines found within the DOE Advanced Scientific Computing Research portfolio in differing degrees but each provides fresh motivation for advancing each discipline. In this section, the authors of this panel report further elucidate the nature of each challenge and map each one to four to six specific approaches for research and software development.

Optimal Algorithms for Optimal Representations

There are many reasons why large-scale simulations of plasma fusion energy systems may need solvers (linear, nonlinear, temporal integration, eigen, etc.) that scale to millions of processor elements (or nodes), and ultimately, thousands of threads per node. As further elaborated in the PRD titled, "Multiphysics, Multiscale Algorithms," simulations must usually be followed over time intervals that are long compared to the shortest time scales in the system; e.g., plasma discharge versus Alfvén time scales in tokamaks. Often, the phenomena associated with the shortest time scales may acceptably be filtered out relative to dynamics of interest; however, these phenomena control the computational time step if an explicit method is used, with the result that even weak scaling cannot be achieved. Often, scientists would ideally employ a high-order time stepping scheme and take relatively large time steps for computational economy. However, if operator-splitting techniques are used, the low-order splitting error thwarts this objective. Moreover, computational challenges on the immediate horizon— UQ, optimization for design or control, inverse problems for parameter identification, multiphysics coupling, etc. —are most naturally tackled with fully implicit formulations for contributing simulation components already established. Thus, this first PRD is foundational to all of the others. Computational physicists have historically made strategic use of implicit solvers, usually modularly and for linear problems only; e.g., for a Poisson field solve on a grid between particle-push steps, or for the components of a Helmholtz-decomposed magnetic field, etc. Because lack of an appropriate solver can be one of the principle "bottlenecks" to scaling, it is important both to inventory and assess the effectiveness of the types of solvers that fusion scientists want to scale today, and also to probe obstacles in employing more fully (nonlinearly) implicit solvers needed for tomorrow.

Evaluation of the conservation law residuals that lead to corrections in the typical inner loop of a fusion energy code typically costs O(N) operations in the size, N, of the discretization. The implicit solver is often superlinear, O(N^a), for a >1. Adaptive discretizations work on reducing N, while optimal solvers work on reducing the exponent a. Both are important to fitting more physics onto a limited computational resource.

In simulations at extreme scales, no data structure whose size scales with the system can be relegated to just one processor-memory element or replicated on each. All such must be distributed. Solvers are therefore just one of many algorithms that must scale. Tools for managing meshes, fields, and particles (e.g., their generation, partitioning, adaptation, interpolation and for constructing of the discrete equations from the underlying models) must all be scalable as well, or Amdahl's Law will impose a limit to scalability that is asymptotically independent of processor granularity.

The algorithmic techniques required to support fusion at extreme scales include computer-aided design to mesh geometric adaptivity, solution-based adaptivity, mesh partitioning, discretizations of virtually all types (with attention to advanced high-order discretizations), optimal implicit solvers, stiff method-of-lines integrators, kinetic and particle methods, unconstrained and constrained optimization (for parameter identification, control, design, etc.), sensitivity analysis (both statistics-based and derivatives-based), and UQ. Extreme-scale fusion simulation represents an opportunity for developers of the enabling technologies in applied mathematics and computer science to demonstrate a paradigmatic shift that they have envisioned for years. However, the connective and control code and the majority of the means of interchange of data between code components will have to be written to take advantage of modern software practices and high-performance parallel architectures. Virtually all large-scale data structures in existing codes will have to be replaced with distributed versions. As the software infrastructure is rebuilt, due attention can be given to extensibility, reusability, object orientation, componentization, portability, performance portability and tuning, code self-description and self-monitoring, and the construction of multilayered interfaces that enforce correct usage (see PRD titled, “Lower Threshold of Expertise for Using Optimal Algorithms on Extreme Architectures”).

Full adaptivity in the sense of *h* (mesh refinement), *p* (discretization order), and *r* (mesh relocation), should be employed in space and time, according to the local smoothness of fields to be represented. This requires estimating and equi-distributing truncation errors, dynamic in-place load balancing, and managing and converting between different representations. Current optimal complexity algorithms for nonadaptive discretizations of the underlying physics must be extended to such optimal representations.

To achieve optimality in the vast part of a typical fusion energy sciences code that lies outside of the basic physics loops, many advances are required:

1. Research on optimal discretization schemes. This is ultimately physics-specific and must be measured ultimately in accuracy per Joule or accuracy per time-to-solution, not by incomplete metrics like mathematical rate of convergence within an inner loop. Traditional measures of quality from numerical analysis include order of accuracy and preservation of properties of the continuum in the discretization, such as positivity, conservation, and gauge invariants, but the ultimate measure for optimality in the application context relates these to the productivity of the physicist in making scientific judgments for a given formulation of a computational problem within limited resources of hardware, power, and time.
2. Research on algebraic multigrid. Multigrid preconditioning provides provable optimal complexity for many problems (beginning with the symmetric, positive, isotropic scalar Laplacian) and needs to be extended to more general problems. Currently, algebraic multigrid is very effective for generalizations—including anisotropy and inhomogeneity—and reasonably effective for problems including asymmetry and indefiniteness. Extensions to singular and near-singular cases with known null spaces are also effective, and the adaptive version is effective at finding null spaces as part of the solution process. Extensions to the nonscalar case (multiple fields) are difficult. The scaling of the method is tied to the ability to coarsen aggressively, which can also be difficult. There are several important kernels interior to multigrid methods: smoothers, prolongators, restrictors, and sparse direct solvers. The first three are challenging because they offer little work in proportion to the data motion required, and the methods of choice depend upon the data (the values of the coefficients), as well as the sparsity structure of the system. The last is challenging because it offers only bounded concurrency. All four kernels deserve intense, ongoing research in fusion energy sciences simulation contexts.

3. Research on Krylov methods. Krylov acceleration techniques can be applied in a matrix-free manner for methods that require Jacobian-vector products only, but methods that require products with the transpose of the Jacobian remain challenging to employ at the large-scale because of the cost of Jacobian computations. Transpose-free methods typically require storage of many Krylov space basis vectors, so research is needed that cuts down on the memory requirements of Krylov solvers so that they can migrate to the memory-thin hardware of the future.
4. Research on globalization techniques for Newton-like methods. Newton-like methods are locally superlinearly convergent or better but may spend many iterations before entering the convergence domain or may diverge if started far from the desired root. Many robustification methods are known, ranging from general purpose, such as line search or trust region, to physically motivated, such as grid sequencing, pseudo-transient continuation, or parameter continuation. Means of transforming the original nonlinear system to a related system with the same root but with a larger domain of convergence are known for many problems, but need to be extended. Additional physically motivated globalization techniques must also be researched.
5. Efficient implementations of optimal methods. As problem size grows linearly with the number of processor-memory elements, codes must maintain constant surface-scaling of communication with volume-scaling of computational work, avoid duplicating any data structures that scale with problem size, and avoid any steps with synchronization-induced idleness except on data sets of constant (therefore asymptotically small) size. As concurrency of threads increases beyond about a million, codes must switch from weak to strong scaling within a shared many-core element of fixed memory. This puts a premium on memory per core, and memory bandwidth per core. Important synchronization-reducing algorithms tend to bloat memory requirements, so this delicate trade-off must be tuned for specific applications.

Multiphysics, Multiscale Algorithms

The physical mechanisms that are active, coupled, and strongly interacting in fusion plasma physics systems are many and challenging to model computationally. These include wave phenomena (magnetosonic, Alfvén, Whistler, etc.), material transport (mass, momentum, energy, charged species), diffusion processes, atomic physics, radiation transport, and interacting electromagnetic fields. The mathematical models that are used to describe such multiple-time-scale multiphysics plasma systems are varied and include, continuum MHD approximations, kinetic descriptions, particle-in-cell type methods, integral-equations, and others.

Multiphysics

Truly predictive simulations of multiple-time-scale multiphysics systems must include time integration methods with high-order accuracy and allow efficient and reliable estimation and control of long-time integration error. The use of fully-implicit methods with advanced coupled nonlinear solvers and scalable linear solvers have shown significant promise in this context. These methods can provide stable, variable- and higher-order techniques with local and global error control. They can also be stable and accurate when run at the dynamical time scale of interest in multiple-time-scale systems. Recently, progress has been made in developing fully implicit formulations that are intended to robustly and accurately integrate these systems and follow the dynamical time scales of interest. These initial studies have been based on iterated fixed point or nonlinear decoupled Gauss-Siedel type iterations and on more modern Newton-Krylov type methods. While some progress has been made with parallel domain decomposition methods with incomplete factorization subdomain solvers, the most successful to date have been preconditioned Newton-Krylov type methods that employ physics-based preconditioning techniques and can provide scalable solution methods. However, significant algorithmic advances and software development is required (as described below) as more complex MHD models are developed for more predictive modeling of fusion devices.

Looking beyond the need to efficiently time-integrate multiple-time-scale systems at the dynamical time scale of interest, there are additional computational challenges on the immediate horizon: UQ, inverse problems for parameter identification, and optimization for design and/or control. These challenges are most naturally addressed with fully implicit formulations.

Multiscale

Developing the infrastructure for predictive fusion device simulation will require multiscale plasma system modeling capabilities. In this context, the coupling of kinetic/particle and continuum scales introduces several additional challenges. First, the concepts of scale separation and interscale communication become intertwined. Mappings are needed that quantitatively pass information between the different types of scales and/or subregions in a domain decomposed hybrid simulation. A significant issue is that fluctuations at the kinetic/particle scale can introduce stochastic forcing at the continuum scale, which introduces significant discretization issues at the continuum MHD model scale. Additionally, accurate modeling and simulation of complex plasma physical systems requires multiscale multiphysics capabilities.

For this reason, predictive fusion device simulations will require not only accurate and efficient computational strategies that bridge continuum MHD and kinetic/particle scales, but also a detailed understanding of the uncertainty in the single-scale simulations and the propagation of uncertainty and numerical error between scales and between the disparate physical models. This requires new developments in mathematics and computational algorithms for accurately and robustly linking component simulations of individual physics-scales. New methods will be required to determine active scales in the space-time domain that are necessary to include, the strength of the interscale coupling, the required interscale transfer operators, and the appropriate model for each scale that accurately captures the relevant physical behavior. Advanced algorithms for component-scale simulations must be adaptive with multiresolution and have UQ capability and internal error control. Appropriate interscale coupling algorithms must be developed that allow error, stability control, and UQ of data at the interface between component-scale models. Advanced multiscale methods must necessarily integrate these component-scale models and the interscale coupling algorithms to allow UQ, sensitivity analysis, and allow for error estimation and control for the entire multiscale simulation.

To achieve accurate, stable, efficient and scalable predictive simulations for multiscale multiphysics systems with fully implicit methods, many advances in numerical methods and computational science are required, which would include the following:

1. Research and demonstration in a few high-risk, high-payoff, strongly coupled, fully implicit, extended MHD simulation codes of error estimation and error control capability. This type of code would be a tool for high-confidence scientific discovery with sensitivity analysis and design optimization capabilities. It could also be used as verification code for cross-benchmarking the more traditional loosely coupled code architectures and can also be used in offline high-resolution simulations for the development of reduced order models. (Codes that fulfill this aspiration exist in

fields like continuum mechanics where it is easier to relate error estimates to mesh adaptation strategies, but not so far in fields like MHD).

2. Research on formulations and solution method capabilities for multiscale hybrid methods that combine kinetic, particle, and MHD fluid models (e.g., subgrid or domain decomposed kinetic and particle models).
3. Development of accurate and efficient numerical methods for stochastic partial differential equations that arise from incorporating source terms representing the coupling of kinetic and particle models into continuum MHD models. The development of numerical methods for stochastic partial differential equations (PDEs) has lagged behind developments in deterministic PDEs.
4. Research on scalable and efficient physics-based preconditioners with subcomponent multilevel solvers that focus on difficult aspects of fluid MHD models. These include resistive and extended MHD, anisotropic transport, and two-temperature MHD systems for both transient and equilibrium calculations.

Optimization and Reduced-Order Modeling

As described elsewhere in this report, it will be essential to use comprehensive whole-device experimentally validated, computer simulations to plan and optimize discharge scenarios. This is necessary because each ITER discharge is estimated to cost approximately a million dollars, with a simple amortization of capital and operational costs over the lifetime of the ITER's experimental program. Invariably, to optimize the performance of burning plasma experiments, a variety of physical stability limits or boundaries will be encountered. Because the ITER can sustain only a limited number of violent disruptions (tens or perhaps hundreds depending upon magnitude), accurate knowledge of the complex stable nonlinear operational parameter space is critically important. For this reason, the ability to computationally optimize the performance of a complex fusion system and accurately and efficiently map the stable operation limits of the device is essential—especially in the presence of uncertainty in model and data input, and accurately and efficiently map the stable operation limits of the device is essential. The ability to implement an effective feedback control system to maintain operation safely within the desired stability boundaries is also critical to the long-term success of a production fusion energy production facility. Computational optimization, combined with bifurcation and stability techniques, can aid in efforts to improve experimental design, maximize fusion power, increase the envelope of parameters that produce stable operation, and help to minimize the occurrence, severity, and consequences of disruption events.

Optimization

In the context of optimization for sufficiently smooth solutions, a hierarchy of algorithms and software exist. This includes black-box, direct sensitivity, adjoint sensitivity, reduced-space, and full-space formulations. Black-box methods require only function evaluations and can use finite difference approximated derivatives. Direct sensitivity methods use information on the sensitivity of the constraints to the design parameters and solve multiple linear systems with the Jacobian matrix for differing right-hand side (RHS) vectors. The adjoint sensitivity methods additionally require a solve of an adjoint system. Reduced space adjoint formulations, which are applicable for a small number of design variables, require solution of repeated Jacobian and adjoint systems with differing RHS vectors. Finally, full-space formulations require approximations to higher-order derivatives (Hessians), formation of the full optimality conditions (KKT systems), and efficient iterative solution of the resulting saddle point system that allows for very large design spaces.

To take advantage of the power of intrusive PDE-constrained optimization techniques such as full-space methods, efficient solution of the KKT system is critical to allow convergence to optimal solutions for very large constraint spaces and design spaces. Scientists often begin with a scalable PDE solver and add optimization capabilities within a Lagrangian or augmented Lagrangian framework. This generally leads to a system matrix of saddle-point type in which the existing PDE Jacobian and its transpose appear as large blocks. The efficient and scalable solution of the KKT system is very challenging and is an active area of research. In general, PDE-constrained design optimization, inverse problems, optimization under uncertainty, and optimal feedback control systems are currently some of the most vigorous and fruitful areas of research in computational mathematics. With the abundance of important and complex applications that the currently proposed Fusion Simulation Project (FSP) contains, there is the potential for significant impact in various aspects of the ITER and FSP programs from optimization techniques.

Feedback Control

Briefly, to optimize the performance of burning plasma experiments, feedback controls are designed to operate near facility limits. To maintain the plasma discharge for sufficient duration, near these limits, real-time feedback control is essential. As described in "Burning Plasma/ITER Science Challenges," the burning plasma regime is fundamentally new with stronger coupling and weaker external control than ever before. For this reason, it is necessary to design feedback control more precisely than in present-day tokamaks.

For simulations aimed at either discharge optimization or plasma feedback control, it will be necessary to have the flexibility to employ reduced order modeling so that repeated calculations can be run in the context of an optimization process or used as the basis of the parameterization of a control strategy for a planned discharge in an experiment. In this context, the efficient scalable solution of the required multiphysics plasma system models must be available to produce high-resolution simulations that can be used to develop and/or evaluate reduced order modeling techniques. These reduced order modeling in turn can be used as the basis for feedback controls systems.

Bifurcation Analysis

As described earlier, it is crucial for the stable and sustained operation of the ITER that an understanding be developed of the stability boundaries associated with the nonlinear dynamics that lead to macroscopic instabilities (disruption events). To attempt to identify the stable operating regime and analyze simulation results to predict disruptions, analysts will need to understand the complex nonlinear space associated with the relevant ITER parameters that control the location and sensitivity of stability boundaries. This process would be time-consuming (possibly prohibitively) if the simulation is run in a "forward" mode; i.e., performing thousands of runs with different parameter sets to map the instabilities. Analysis tools must be incorporated that can efficiently and automatically traverse and map the stability boundaries of the parameter space. Mathematically, these disruption events represent an exchange of stability called a bifurcation. Large-scale stability and bifurcation analysis tools exist that can directly map out unstable regions in parameter space without running initial value computations to steady state. For explicit codes, the simulation can be treated as a black-box, via a recursive projection method that requires only a sampling of time steps instead of a full transient solution. Such techniques have been demonstrated to be scalable to large systems. For Newton-based implicit codes, a direct solution to steady-state is possible, allowing an efficient localization of the bifurcation point.

The development of comprehensive stability and bifurcation results would potentially allow fusion devices to be run closer to a bifurcation (stability boundary) without triggering a disruption. Because there will be uncertainty in the mathematical model, input data, source terms and fluctuations in the reactor, a safety window should be built into the bifurcation diagram. Research into incorporating sensitivity information and UQ techniques into numerical algorithms for constrained eigenvalue computations could be required to determine the size of the safety window. Of course, the actual viability of such an approach must be validated against actual experimental results.

To achieve accurate, stable, efficient and scalable predictive simulations for multiscale multiphysics systems with optimization, bifurcation, and stability methods a number of advances in numerical methods and computational science are required. The list includes the following:

1. Scalable and efficient solution of the optimality conditions for full-space methods (KKT system). This requires advances in specialized saddle-point preconditioners. Promising directions might include the use of reduced space methods as preconditioners for full-space methods and approximate block decomposition methods with various Schur complement approximations combined with the physics-based preconditioners described above for the resistive and extended MHD systems.
2. Optimization algorithms for deterministic/stochastic systems that can be used for hybrid kinetic and particle + continuum MHD type solution methods that can introduce stochastic character to the system.
3. Development of very fast reduced order modeling techniques with computable error estimates to implement model-based scenario planning, experimental design optimization studies, and feedback control systems.
4. Research and development for inverse modeling of various models to allow scientifically based inference of the internal structure and dynamics of fusion reactors from exterior sensor measurements. These are critical for instabilities in the experiments that need to be understood. Also, these are critical capabilities for helping to understand the ability to design reduced order modeling for active control purposes.
5. Efficient optimization based feedback control methods with provable accuracy. Specialized methods, when coupled with an underlying reduced order modeling, should be developed.
6. Research and development for determination of global equilibria modes in real geometries for ideal, resistive, and extended MHD models. In addition, work on the implementation and evaluation of continuation, bifurcation and stability algorithms for MHD PDE codes that simulate modeling various instabilities for steady and periodic equilibria in fusion reactor geometries.

Uncertainty Quantification and Reduction

Truly predictive simulations of the multiphysics plasma systems must include the characterization of numerical errors along with uncertainties in input data, mathematical and physical models as well as procedures for bounding and/or estimating, and reducing these errors and uncertainties. Uncertainty is formally classified as aleatory uncertainty and epistemic uncertainty.

Aleatory uncertainty (or variability, stochastic or irreducible uncertainty) describes the inherent randomness in the behavior of the system under study. Aleatory uncertainty is irreducible, except through design changes to the system under study. *Epistemic uncertainty* (or knowledge, subjective or reducible uncertainty) is used to characterize the lack of knowledge about the appropriate value to use for a quantity that is assumed to have a fixed value in the context of a specific application. Epistemic uncertainties are reducible by increased understanding (research) and increased higher fidelity data collection.

Clearly, the computational models that are used to study fusion devices contain many sources of numerical errors and uncertainty. These include discretization error (temporal and spatial); incomplete convergence error (nonlinear, linear, etc.); uncertainties in input data (initial conditions, boundary conditions, coefficients, thermo-physical properties, source terms, etc.) and the component models (the specific form and parameter values) that are used in the mathematical system model. The ability to estimate and control numerical error and uncertainty would allow simulations to effectively balance numerical error, input data and model uncertainties to produce more accurate and efficient large-scale computational simulations of the physics in future fusion devices.

Methods that attempt to estimate integration errors include feature-based techniques, energy norm methods, Richardson extrapolation, and adjoint-based methods. Based on the energy norm and Richardson techniques, global error can be controlled. Adjoint methods allow control over global average error, as well as specific subregions of the computational domain, and also derived specific QoI in the scientific solution.

Aleatory uncertainty is commonly estimated and propagated through computational models by probabilistic methods such as sampling methods (Monte Carlo, quasi-Monte Carlo, Latin hypercube techniques, etc.) and by direct propagation methods (stochastic Galerkin, stochastic collocation, Neumann and Taylor expansion, etc.). The sampling methods are black-box type techniques and have general applicability but can often converge very slowly for high-dimensional spaces. The direct propagation methods can often converge much faster but these techniques are more intrusive in the scientific application. In addition, recent work in adjoint-based formulations produces an output distribution of a QoI determined from the solution of a differential equation in which data and/or parameters are random variables. Epistemic uncertainty can be estimated by subjective probably methods, fuzzy sets, evidence theory, and Bayesian inference.

To achieve predictive simulations with high-physics fidelity for complex fusion devices, a number of advances in numerical methods and computational science are required. This list includes the following:

1. Research on efficient error estimation and control, sensitivity analysis, and UQ methods for combined deterministic and stochastic plasma physics models. Hybrid deterministic and probabilistic UQ approaches need to be developed and studied in the context of complex applications.
2. Probabilistic approaches based on sampling methods (e.g., Monte Carlo) and direct methods (e.g., polynomial chaos). These need algorithms and software tool development to support computationally efficient implementations for transient simulations.
3. Deterministic UQ tools based on sensitivity and adjoint-based techniques for data, integration, and model error estimation and control. Such tools need to be developed and demonstrated for complex multiphysics applications. To develop efficient, transient adjoint-based techniques, work is required to limit solution storage requirements, memory usage, parallel communication, and cost of the adjoint solve. In addition, adjoint techniques for hyperbolic systems are critically required.
4. Research on error estimation and UQ for multiphysics, multiscale, multimodel simulations. This includes methods for loosely coupled multiphysics multiscale solvers that would involve data

handoffs between multiple codes. Research is needed on propagating uncertainty between heterogeneous models through the component models at each scale and through the interscale transfer operators. Methods for tightly coupled multiphysics and multiscale solution methods are required as well.

Lower Threshold of Expertise Required for Using Optimal Algorithms on Extreme Architectures

For computer simulation to achieve its potential, close collaboration will be required between tool developers in DOE's Office of Advanced Scientific Computing Research, the DOE Fusion Energy Sciences Program areas, as well as with the applications scientists/users. At extreme scales, both the challenges and the tools are too complex for a transfer of technology except through shared personnel. However, as simulation expands as a preferred modality of scientific discovery and engineering design, in response to its advantages in cost and speed over some experimental programs, the intensely collaborative approach adopted in today's Scientific Discovery through Advanced Computing Proto-Fusion Simulation Program centers will not indefinitely scale on the human resource side. It is important to place increasingly advanced simulation capabilities into the hands of end users with all levels of computational expertise.

The DOE Fusion Energy Sciences Program research agenda embraces two different and equally important directions: harnessing the growing capabilities of computation to improve the fidelity of complex geometry, multiscale, multirate, and multiphysics models, and lowering the threshold of expertise required to employ "best practices" and scalable software. This PRD concerns the latter direction.

Standards from commercial software engineering are becoming established in scientific software libraries. The benefits of this canonization of "scientific cyber-infrastructure" are many and include confidence of users in quality and persistence of software, greater recognition for developers, and better amortization of their efforts in economies of scale, propagation of best algorithmic practices into user applications, propagation of library quality standards into user applications, and a better trained computer science and engineering workforce. Most importantly, computer science and mathematical expertise is reliably packaged for transmission in a way that does not require every practicing computational physicist to become an expert in high-performance computing. How best to deliver computer science and mathematical expertise in software is a long-standing field of research.

Today's most successful scientific software toolkits show the importance of abstraction to this endeavor. Three themes of modern scientific software are as follows: 1) encapsulation, through abstraction of data; 2) polymorphism, through abstraction of operations; and 3) composability ("plug'n'play") through abstraction of interfaces. Other themes include extensibility, which is the ease of adding functionality underneath the abstractions; and portability, which is the ease of moving code across architectures while maintaining correctness and performance.

Successful software is multilayered in its accessibility. The outer layer, which is accessible to all users, is an abstract interface featuring the language of the application domain, hiding implementation details, with conservative parameter defaults. For instance, a vector should be represented and manipulated as an element of a Hilbert space, without regard for how it is implemented in an array data structure partitioned across multiple memory systems. The goals of accessing the software at this layer are robustness, correctness, and ease of use. The middle layers, which can be opened up to experienced users through handles to the underlying objects, provide a rich collection of state-of-the-art methods and data structures, exposed upon demand, and variously configurable. The goals of access are capability, algorithmic efficiency, extensibility, composability, and comprehensibility of performance and resource use through profiling. The inner layer is intended for developer access only. It includes support for a variety of hardware and software execution environments. The goals of access are portability and implementation efficiency.

The ecosystem of extreme-scale computing is rich and consists of many types of software of varying degrees of maturity. On the modeling side, the required toolkits include the following: geometric modelers, meshers, discretizers, partitioners, solvers and integrators, systems for mesh and discretization adaptivity, random number generators for stochastic models, libraries of subgrid-scale physics models and constitutive relations, UQ, dynamic load balancing, graphical and combinatorial algorithms, and compression.

Code development relies on another fleet of software toolkits, beyond those related to the modeling itself, such as configuration systems, compilers and source-to-source translators, messaging systems, debuggers, and profilers. Finally, production use of extreme-scale applications relies on the following: dynamic resource management, dynamic performance optimization, authentication systems for remote access, input/output systems, visualization systems, data miners, workflow controllers, and frameworks. These toolkits will all need to be supported on emerging architectures, where they will need to be supplemented beyond their counterparts today by additional tools such as fault monitoring, fault reporting to the application, and fault recovery.

To propagate best practices in extreme-scale simulations into the fusion community, the DOE Office of Fusion Energy Sciences Program scientists should seek to influence scientific software engineering by being early adopters, drivers, and even co-developers of the software ecosystem described above. Research priorities in scientific software engineering would include the following:

1. Raising standards for the design of user interfaces. Interface design should be multilayered for incremental adoption, presenting as few concepts as possible at the highest level with implementation details hidden at the top but exposable upon request to experts. (The DOE-maintained, internationally popular solver toolkit PETSc [Portable, Extensible Toolkit for Scientific Computation] is an example of such a design.) The interface should allow calling from all relevant languages and be portable across all relevant architectures. It should be extensible in the sense that advanced users should be able to register routines that offer new functionality and have them accessible without changes to the interface.
2. Raising standards for reliability. New software should come with test harnesses that explore all paths of execution that are likely to be encountered during production use.
3. Raising standards for performance. Future architectures will strongly penalize data motion, which costs time and power. Data motion includes “visible” copying of data between different structures in user space and “invisible” replication of data at different levels of a hierarchical memory system by the runtime system. Software should attempt to minimize user space data copying, should restrict sharing to only data that require it, and should strive for spatial and temporal locality of data by means of architecture-specific tuning.
4. Raising standards for documentation, support, and training. Software intended for community use should have integrated documentation, parts of which are automatically generated from the source to

remain updated. Software under development should undergo nightly regressions. Long after the "research" aspects of the software are established, the software must be maintained through active, automatically tracked monitoring of bug reports, bug fixes, feature requests, and feature additions. Periodic training opportunities should be available on the assumption that the user community for simulation will continue to expand among users who are not primarily computer scientists.

CONCLUSIONS

Five PRDs in algorithms for fusion energy sciences at extreme scale have been developed from physical requirements and mapped onto required advances in mathematical modeling, algorithms, numerical analysis, parallel programming paradigms, and scientific software engineering. The portfolio of advances required: in discretizations, multigrid solvers, Krylov solvers, Newton solvers, coupling of multiple models, coupling of different formulations, coupling of deterministic and stochastic approaches, physics-based preconditioning, optimization algorithms, solvers and preconditioners for the special linear systems that arise in optimization, reduced-order methods, inverse methods, optimal control algorithms, methods for mapping out stability regions and bifurcations in operating regimes, error estimation and control, UQ for deterministic and stochastic methods and their combination, sensitivity analysis, adjoint methods, design of user interfaces, software reliability, flop/s performance relative to bandwidth and power, and software user-friendliness, are familiar within the overall extreme-scale applications portfolio. Many of these topics named as requiring advances might be deemed to be in relatively satisfactory states with respect to today's scales of analysis, but they do not migrate to the extreme scales towards which the fusion energy sciences agenda drives us. The principal advances required in these cases are to deliver similar quality of performance at larger scales with less memory and power per processor.

The magnetically confined fusion energy research agenda is a thorough driver for advances in algorithms and scalable software technology that are needed for other areas and are attractive and timely from the viewpoint of computational and applied mathematicians and computer scientists.

DATA ANALYSIS, MANAGEMENT, AND VISUALIZATION IN FUSION ENERGY SCIENCE

Lead: Arie Shoshani, Lawrence Berkeley National Laboratory

Panel Members: Hank Childs, Lawrence Berkeley National Laboratory; Alok Choudhary, Northwestern University; Chandrika Kamath, Lawrence Livermore National Laboratory; Scott Klasky, Oak Ridge National Laboratory; Kwan-Liu Ma, University of California, Davis; Robert Ross, Argonne National Laboratory; Nagiza Samatova, Oak Ridge National Laboratory and North Carolina State University; Allen Sanderson, University of Utah; David Schissel, General Atomics; Svetlana Shasharina, Tech-X Corporation; and Mladen Vouk, North Carolina State University

SUMMARY

Fusion energy scientists who are considering computational needs and science questions that can be answered with extreme-scale computational power should also consider the implications of data requirements at the extreme scale. Managing fusion simulation data already has proven to be a problem in terms of volume, bandwidth, and complexity. Some codes (e.g., VPIC, OSIRIS, M3D-K) will model 1 billion cells and 1 trillion particles. Based on mean-time-between-failure (MTBF) concerns when running on a million cores, these codes will need to output 2 gigabytes/second per core or 2 petabytes/second of checkpoint data every 10 minutes. This amounts to an unprecedented input/output (I/O) rate of 3.3 terabytes/second. The data questions to consider at the extreme scale are divided into two main categories: 1) data generated and collected during the production phase, and 2) data that need to be accessed during the analysis phase.

Summarized below are the main challenges identified in each of these areas at the extreme scale. Details about these challenges are provides in the next section.

Managing Large-Scale Input/Output Volume and Data Movement

Techniques need to be developed that optimize I/O performance automatically based on hardware characteristics. Such techniques are crucial to avoid slowdown of computations because of insufficient I/O rates. Furthermore, future fusion energy science codes should be as independent as possible of I/O tuning, so all such details should be handled automatically by the underlying I/O system. Parallel file systems and data-movement tools need to be scaled to support these extreme volumes of data. Furthermore, datasets are often distributed across facilities and must be consolidated for certain analyses.

Real-Time Monitoring of Simulations and Run-Time Metadata Generation

Having a run-time monitoring capability on all supercomputing resources is essential to avoid wasting valuable time on computational resources. This capability stops runs that do not converge or progress correctly. Workflow technology already used for such purposes in fusion energy science applications needs to be scaled and become part of the simulation system that supports summarization of results in real time, and/or permit the monitoring software to automatically manage simulations and identify and halt those that do not progress correctly. Also, as demonstrated in previous projects in other scientific fields, provenance and metadata information needs to be automatically collected (also at run time) for effective run-time and post-run data analysis.

Data Analysis at the Extreme Scale

The data-analysis challenges in fusion energy science applications at the extreme scale stem not only from the large size of the data, but also from the complexity of the data. First, areas of interest, such as coherent structures and fronts, are likely to be distributed among many processors, making it difficult to extract poorly defined structures or track fronts over time. Second, techniques to reduce overall data size before these data are output by the simulation require algorithms that are robust, operate with limited memory resources, and be highly scalable. Third, the diversity of data formats and representations provides a significant challenge when comparing and integrating—in particular—experimental and simulation data. Fourth, data required are unlikely to reside in the same physical location, requiring large data movement or being distributed, but integrated analysis applications.

Visualization of Very Large Datasets

Visualization is often a key approach used for understanding data such as electron-temperature profiles. However, reducing and mapping terabytes or petabytes of data into meaningful visualizations are challenges that will require processing capabilities near the data storage location, as well as effective indexing techniques for real-time data exploration. Additional challenges arise where the data to be visualized are distributed.

Experiment-Simulation Data Comparison

Experiment-simulation data comparison tools are essential for validation of fusion energy science simulations and diagnostics, and for comparing shot data to reduced models for runs at the ITER and similar devices. Experimental data are already reaching terabyte sizes at leading experimental facilities and future facilities such as XFEL or Hyper, and are expected to produce 10s of petabytes, outstripping Large Hadron Collider production rates; therefore, robust synthetic diagnostic tools need to be developed. These tools must be cross-platform-scalable and based on forthcoming community standard data formats for both experimental and computational communities with the ability to integrate across these semantically diverse representations.

SCIENTIFIC CHALLENGES AND RESEARCH APPROACHES

Managing Large-scale Input/Output Volume and Data Movement

As data continue to grow, I/O is continuing to be of major concern as it slows down computations. Several considerations should be examined to enable new approaches to this challenge. First, storage architectures must take greater advantage of the cost benefits of commodity storage, allowing for greater "raw" bandwidth given a fixed budget. Second, new autonomic storage software must be developed to operate at this scale and provide the level of performance and reliability necessary for these less reliable parts; however, this must be achieved in a transparent manner. Third, performing analysis while the data are generated can help both in accelerating knowledge discovery and in reducing demands on the storage system. Techniques, such as in situ analysis and online data reduction and transformation for reducing the demands on the storage system, must be pursued in conjunction with storage system improvements.

Current data-storage formats and interfaces, such as the HDF5 (Hierarchical Data Format version 5) and Parallel netCDF (Network Common Data Form) software, are only the beginning of a larger effort to improve the usability of storage by computational science applications. New interfaces and storage models must be developed to bridge this gap, and the community supported in the adoption and exploitation of these types of formats. This would provide interfaces that are appropriate for codes that do not use uniform meshes, such as adaptive and unstructured meshes. These high-level storage models need to be mapped onto the underlying physical storage (such as distributing blocks on a parallel file system) in such a way as to provide the most efficient access to the data for subsequent analysis.

As researchers move forward through petascale computing to extreme-scale computing, the opportunities for groundbreaking science will continue to grow, yet the complexities of gaining insight will increase proportionally. Challenges that will be encountered include the complexity of performing high-performance I/O for both simulations and analytics and the lack of flexibility in the interfaces. File systems, parallel I/O, and high-level application programming interfaces (APIs) are evaluated next.

Parallel File Systems

Leadership-class computing has brought a new era into high-performance computing. In the past, applications generally had smaller allocations and were unable to run very large applications for long periods of time. With the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program and the creation of leadership-class facilities (LCFs), application teams can allocate as many as 100 million central processing unit (CPU) hours in 1 year. This advance presents opportunities to take on new and exciting challenges in many areas, including the area of high-performance I/O. This area centers around three major pieces that involve determining 1) what file system will the LCF place on the system; 2) how to achieve efficient parallel I/O on the system; and 3) which high-level APIs will be used to provide the desired application performance.

Currently, there are only a small number of good parallel file systems that are available to users. The National Energy Research Scientific Computing (NERSC) Center at LBNL uses a combination of the Lustre file system and IBM's General Parallel File System (GPFS) across the major NERSC resources and on its Cray XT4. Oak Ridge National Laboratory (ORNL) uses Lustre, both shared and local, on many of its computers. Argonne National Laboratory (ANL) uses GPFS as well as the Parallel Virtual File System (PVFS).

Network File System (NFS) version 4.1 adds the parallel NFS (pNFS) capability, which enables data-access parallelism. The NFS v4.1 protocol defines a method of separating the file system metadata from the location of the file data; it goes beyond the simple name/data separation by striping the data amongst a set of data servers. This approach is different from that used by traditional NFS servers, which hold the names of files and their data in a single server. Multinode NFS products exist, but the participation of the client in the separation of metadata and data is limited. The NFS v4.1 pNFS server is a collection of server resources or components; these are assumed to be controlled by the metadata server. The NFS v4.1 client can be enabled to directly access file data distributed across many pNFS servers and avoid solitary interaction with the single NFS server when moving data.

PVFS is an open-source parallel file system used in a variety of laboratory, university, and industry settings. A parallel file system is a type of file system that distributes file data across multiple servers and provides concurrent access by multiple tasks of a parallel application. PVFS was designed for use in large-scale cluster computing and is especially tailored to scientific computing applications, eliminating overheads common in commercial products when applied to high-performance computing (HPC) workloads. PVFS consists of a server process and a client library, both of which are written entirely of user-level code. A Linux kernel module and PVFS-client process allow the file system to be mounted and used with standard utilities. The client library provides high-performance access via the message passing interface (MPI).

Lustre was initiated and funded almost a decade ago by the U.S. Department of Energy Office of Science and National Nuclear Security Administration laboratories to address the need for an open-source, highly scalable, high-performance parallel file system on existing and future supercomputing platforms. Throughout the last decade, Lustre was deployed on numerous medium- to large-scale supercomputing platforms and clusters, and it met the expectations of its user community. According to the Top500 list[1] at the time this workshop report was written, 15 of the top 30 supercomputers in the world use the Lustre file system.

GPFS is a high-performance, shared-disk clustered file system developed by IBM. It is used by many of the supercomputers on the Top500 list. Like some other cluster file systems, GPFS provides concurrent high-speed file access to applications executing on multiple nodes of clusters. In addition to providing file-system storage capabilities, GPFS provides tools for management and administration of the GPFS cluster and allows shared access to file systems from remote GPFS clusters.

State of the Art: MPI-I/O and I/O Rates in Fusion Codes

ROMIO, a high-performance portable MPI-IO implementation, is the standard way that many applications perform parallel I/O. This is the backbone of Parallel netCDF (PnetCDF), parallel HDF5, and ADIOS (Adaptable I/O System), which are described later in this section. ROMIO is optimized for noncontiguous access patterns, which are common in parallel applications. It also has an optimized implementation of collective I/O, an important optimization in parallel I/O.

Through extensive testing, researchers have seen that parallel I/O helps many applications scale to unprecedented numbers of processors. One file per process output simply will not work on many of the LCFs after researchers begin to get 100,000 files or more every time data are written out.

There are two classifications of fusion codes. The first class is for codes like VPIC. The restart data from VPIC is typically of the order of 2 gigabytes per core on a Cray XT; these are typically written every 4 hours. The high-dimensional data are typically hundreds of gigabytes to several terabytes written out for every simulation run. These data are processed and used in subsequent analysis and visualization. Clearly, there is a problem writing a total of 200 terabytes out of 100,000 cores if the user can only get 25% of the peak I/O performance. Specifically, on a 200 gigabyte/second file system, it will take 1000 seconds at peak I/O, and over 4000 seconds (> 1 hour) at 25% utilization to write out the restarts. This is one of the reasons why VPIC researchers only write their restart every 4 hours, spending 4000 seconds to write out data, which means that 28% of their time is spent in I/O rather than in computing. Of course, file systems are shared among all the users, and the performance of the I/O can be greatly affected, making I/O even slower at busy periods. This demonstrates the need for I/O methods

[1] http://www.top500.org/

that are performed concurrently with computation, such as dedicating some of the nodes on a cluster (called I/O nodes) to perform I/O while computations are continuing to run.

The second class of fusion codes are particle gyrokinetic simulations, such as the XGC-1, GTC, and GTS codes. These codes already produce unprecedented amounts of data. These codes generally write one-tenth the amount of memory for their restarts, which they choose to write about once per hour. These simulations are also producing very large amounts of analysis data because the computations are too expensive to reproduce. For early access runs at ORNL, both the GTC and GTS codes wrote over 100 terabytes of data per wall clock day, 25 terabytes of which was for analysis.

Another reason for potentially large I/O is in code coupling. One of many ways to couple codes is to write data to a file and then have another component read in the data. Another way is to keep data in memory, and perform in-memory transformations so that a second component can use the data. This approach allows codes to be coupled both on the same platform and on different platforms, or for components to be executed in sequence rather than in parallel, when compute resources are limited. Code coupling must incorporate both memory-to-memory methods, as well as file-based coupling approaches to accommodate different coupling rates on the same platform or multiple platforms that may be at different physical sites. Because code coupling can eventually involve many codes run over days of runtime, it is clear that large amounts of data can be generated.

State of the Art: File Formats

Currently, there are three file formats used for large-scale fusion simulations: pHDF5, pNetCDF, and ADIOS-BP (binary packed). All three formats are metadata rich, and have been used for very large fusion simulations.

HDF5

HDF5 is a versatile data model that can represent complex data objects and a wide variety of metadata. It is a portable file format, and there is no effective limit on the number or size of data objects in the collection. HDF5 was originally developed at the National Center for Supercomputing Applications (NCSA), and is currently supported by the HDF Group, whose mission is to continue the development of HDF5 technologies and provide accessibility to data currently stored in HDF. The HDF5 format is designed to address the current and anticipated requirements of modern systems and applications, especially in supporting parallel reads and writes.

HDF5 simplifies the file structure to include only two major types of objects: 1) datasets that are multidimensional arrays of a homogenous type, and 2) groups that are container structures created to hold datasets and other groups. This approach results in a hierarchical file-system-like data format. In fact, resources in an HDF5 file are even accessed using syntax similar to the portable operating system interface for Unix (POSIX). Metadata is stored in the form of user-defined and user-named attributes attached to groups and datasets. More complex storage APIs representing images and tables can then be built using datasets, groups, and attributes.

In addition to these advances in the file format, HDF5 includes an improved type system and enhanced data-space objects that represent selections over data-set regions. The API is object-oriented with respect to datasets, groups, attributes, types, data spaces, and property lists.

HDF5 is the underlying software for NetCDF version 4, and therefore it is one way that allows NetCDF to be used in parallel computing applications. As discussed in the following section, "NetCDF and Parallel Access," Parallel NetCDF is another.

Because HDF5 uses B-trees to index table objects, it works well for time series data such as stock market ticks or network monitoring data. The bulk of these data go into straightforward arrays (table objects) that can be accessed much more quickly than the rows of a relational database system.

NetCDF and Parallel Access

NetCDF is a set of software libraries and "self-describing" machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data. Self-describing means that a header describes the layout of the rest of the file, particularly the data arrays and arbitrary file metadata in the form of name/value attributes. The format is platform independent, with issues such as endianness addressed in the software libraries. The data arrays are rectangular, not ragged, and are stored in a simple and regular fashion that allows efficient subsetting.

There are two implementations of parallel access through netCDF interfaces:

1. The new 4.0 version of the API that allows the use of the HDF5 data format. NetCDF users can create HDF5 files with features not available with the netCDF format, such as variables with multiple unlimited dimensions, and much larger files. The netCDF library uses the classic netCDF binary format by default. Full backward compatibility in accessing old netCDF files is supported.
2. An extension of netCDF for parallel computing—called Parallel netCDF (or PnetCDF)—has been developed by ANL and Northwestern University. This extension is built on top of MPI-IO, the I/O extension to MPI communications. By using the high-level netCDF data structures, the PnetCDF libraries can make use of optimizations to efficiently distribute the file's read and write operations between multiple processors. The PnetCDF package can read/write the classic and 64-bit offset formats, but is not designed to read or write the HDF5-based format available with netCDF4.

ADIOS

ADIOS is a componentization of the I/O layer. It provides an easy-to-use programming interface, which can be as simple as Fortran file I/O statements. ADIOS abstracts I/O metadata information and data structures from the source code into an external extensible markup language (XML) file, which can reduce code pollution and create the connection between high-level APIs and the underlying I/O implementation details, as well as other technical descriptions such as buffering and scheduling. By separating the detailed I/O implementation from the APIs, ADIOS also allows users to simply change the declaration of the transport methods in the XML file without any source-code modification. Figure 20 illustrates the architecture of the ADIOS framework.

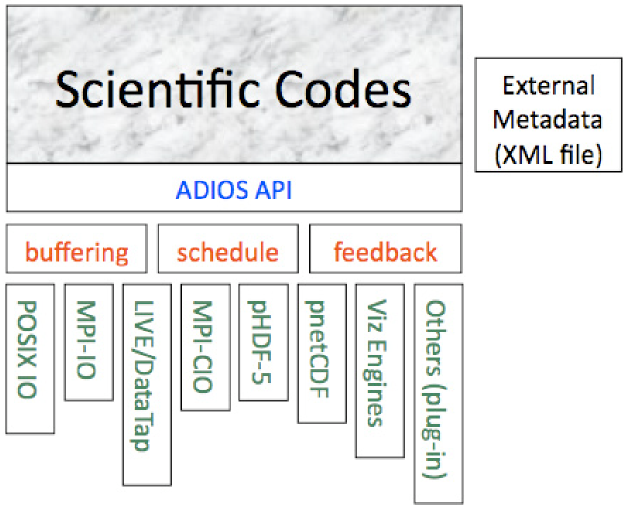


Figure 20. ADIOS framework.[1] Image courtesy of Scott Klasky (Oak Ridge National Laboratory).

The ADIOS architecture uses the XML metadata file to describe all the I/O-related variables used in the code. These collections of data are described in terms of groups that correspond to the individual subroutines. The file registers for each element the following collections: the element name, data type, element path (similar to HDF5 path), and static or dynamic array size, as well as any annotations. If the array represents a mesh, information about the global bounds of this array and ghost regions used in real-time visualization is encoded in the XML group level. For each data collection/group, it describes the selected transport mechanism and parameters, as well as timing information for the data transmissions. Using this information, the ADIOS I/O implementation can then control when, how, and how much data are written or transferred at a time, thereby enabling efficient overlapping with computation phases of the scientific applications and proper pacing to optimize the writing or transmission throughput.

So far, ADIOS has been integrated into several fusion codes (XGC0, XGC1, GTC, and GTS), a combustion-simulation code (S3D), and the astrophysics applications Chimera and Flash, on the Cray XT and on Infiniband and Gigabit Ethernet clusters. I/O performance results using ADIOS are very impressive. In addition, ADIOS permits I/O to be performed using both standard synchronous methods or new asynchronous techniques being developed in the High-End Computing University Research Activity (HECURA), jointly funded by the National Science Foundation and DOE, and conducted by researchers at Georgia Institute of Technology and Rutgers University.

ADIOS and its binary-packed (BP) intermediate file format not only support the flexible conversion to standard file formats, but they also facilitate the summary inspection of data to determine if it contains features of interest to end users. The BP file format is an implementation of a log-file format, one that is becoming very popular in middleware systems such as PLFS (Bent et al. 2009). One way to provide such functionality is to fully index the data, as done by multiple projects that have developed content indices for HDF5 files. To achieve a similar goal but with less overhead in space and time, ADIOS supports the notion of data characteristics that includes local, simple statistical and/or analytical data during the output operation (or later) for use in identifying desired datasets. Simple characteristics, such as local process array minimum and maximum values, can be collected almost at no cost as part of the I/O operation.

[1] http://www.nccs.gov/user-support/center-projects/adios/

More complex analytical measures, such as standard deviations or specialized measures particular to the science being performed, may require processing that can be done in a variety of ways, including before or after the data have been written to disk. In all such cases, the BP format offers efficient, compact ways of storing these characteristics.

Implementation of the BP file format uses a footer index to avoid the known limitation of header-based formats like NetCDF where any change to the length of file data will require restructuring the file. Further, by placing a version identifier and an offset to the beginning of the index as the last few bytes of a BP file, finding the index information and adding new and different data to the file without affecting any already-written data becomes trivial. Finally, data characteristics are incorporated into the index so the index can be separated for use as a table of contents for the file on a tape storage system such as the High-Performance Storage System (HPSS). At ORNL, ADIOS, writing BP files, has demonstrated that it can help codes such as GTC write over 100 GB/second on the petascale Cray XT.

Current State of Fusion Code, and Estimation of I/O Requirements at the Extreme Scale

There are several currently available codes that scale to the petascale level. These codes will be used as examples to present the challenges involved in I/O and storage when scaling to extreme-scale systems. Currently, there are two ways to consider these codes—individual codes and integrated codes. Individual codes, such as GTC, Gyro, M3DK, etc., demonstrate different I/O characteristics. Integrated codes, such as the Simulation of Wave Interactions with MHD (SWIM) framework, Center for Plasma Edge Simulation (CPES), and Framework for Core-Edge Transport Simulations (FACETS), couple multiple codes into one infrastructure.

There are several groupings of data that are output to the storage systems: 1) checkpoint-restart data, 2) high-dimensional analysis data, 3) visualization data, and 4) low-dimensional data for typical monitoring of applications.

1. GTC. The restart data are typically of the order of a few hundred megabytes per core. These data are typically written every hour. The high-dimensional data are typically hundreds of gigabytes written out every few minutes. Visualization data are approximately few hundred megabytes output every few minutes. Low-dimensional data are less than a megabyte and are output approximately every 30 seconds.

   Extreme Scale. 10 M cores × 100 MB/per core = 1 PB data for restart. Restart frequency depends on the MTBF of a system. Thus, the I/O frequency could be as high as one restart per 30 minutes to output 1 petabyte of data. The analysis data will increase by several orders of magnitude for both the high-dimensional and visualization data because of an increase in the number of variables and an increase in the number of dimensions of that data.

2. Gyro. The restart data are typically of the order of a few megabytes per core. These data are typically written every several hours. Visualization data are approximately a few gigabytes output every few hours. Low-dimensional data are less than a megabyte and are output approximately every 30 seconds.

   Extreme Scale. The number of simulations will increase by three orders of magnitude, while each simulation is expected to remain at the same level/size as it is currently. Therefore, per-simulation I/O requirements are not expected to change dramatically.

3. XGC1. The restart data are typically of the order of a few hundred megabytes per core. These data are typically written every hour. The high-dimensional data are typically hundreds of gigabytes and

are written out every few minutes. Visualization data are approximately a few hundred megabytes, and are output every few minutes. Low-dimensional data are less than a megabyte and are output approximately every 30 seconds.

Extreme Scale. Similar to GTC, 10 M cores × 100 MB/per core = 1 PB data for restart. Restart frequency depends on the MTBF of a system. Thus, the I/O frequency could be as high as one restart per 30 minutes to output 1 petabyte of data. The analysis data will increase by several orders of magnitude for both the high-dimensional and visualization data because of an increase in the number of variables and an increase in the number of dimensions of that data.

4. M3DK. The restart data are typically of the order of tens of megabytes per core. These data are typically written every hour. The high-dimensional data are typically hundreds of gigabytes and are written out every simulation. The same data are used for visualization applications. Low-dimensional data are less than a megabyte and are output approximately every 30 seconds.

   Extreme Scale. The extreme-scale version of M3DK is expected to output data approximately three orders of magnitude greater than that output by a petascale system.

5. VPIC. The restart data are typically of the order of 2 gigabytes per core. These data are typically written every 4 hours. These high-dimensional data typically range from hundreds of gigabytes to several terabytes that are written out for every simulation. These data are also used for visualization. Low-dimensional data are less than a megabyte and are output approximately every 30 seconds.

   Extreme Scale. The extreme scale is expected to write 2 gigabytes per core every 10 minutes. The visualization data are expected to be a few terabytes per dump.

Current Fusion Simulation Projects and Codes They Use

CPES

The current mechanism for coupling the codes in this framework is through ADIOS using both parallel-file I/O and in-memory code coupling. The amount of data transferred during coupling is currently small, but in the near future, the amount will become very large. The main codes in this framework are XGC0, XGC1, GTC, GEM, DEGAS2, M3D, and ELITE. I/O is dominated by either XGC1 code I/O as described above or by the GEM code, which demonstrates characteristics similar to those of the Gyro code.

SWIM

In this framework, coupling is performed through serial-file I/O. The amount of data transferred during coupling is in the order of a megabyte. The main codes in this framework are TRANSP, TSC, AORSA, M3D, and NIMROD. None of the individual codes currently generate massive amounts of data.

FACETS

In this framework, coupling is performed via memory. The main codes included in this framework are GYRO, UEDGE, etc. Codes such as GYRO dominate the I/O.

Data Movement and Storage Requirements

Increases in computational power have created the opportunity for new, more precise and more complex scientific simulations that are leading to new scientific insights. Consequently, large experiments generate ever-increasing volumes of data. At the data-generation phase, large volumes of storage have to be allocated for data collection and archiving. At the data-analysis phase, storage needs to be allocated to bring a subset of the data to the scientists' site for exploration, and to store the subsequently generated data products. Furthermore, following the experiences in other scientific communities, such as particle physics or environmental science, storage systems shared by a community of scientists need a common data-access mechanism that allocates storage space dynamically, manages stored content, and removes unused data automatically to avoid overloading data stores. Finally, scientists will need to analyze data generated or archived at different sites, requiring transfers of very large datasets. In the previous section, challenges in developing storage systems that can support expected I/O rates are discussed. In this section, the authors of this panel report consider the technology necessary to use such storage systems effectively, as well as moving data between storage systems.

A typical mode of operation today is to run fusion simulations on a supercomputer, and then move the data to the scientists' site for further exploration. This approach cannot scale as scientists move to extreme-scale computing. The expectation is that much of these data will be stored and archived close to where the data are generated, and that some analysis and data reduction steps will take place in facilities near the data. A dedicated data center, with high-network connectivity to the computational centers, could support long-term data storage, data analysis, and community dissemination. However, it is likely that the requirement of moving fairly large subsets of reduced (or even raw) data to the scientists' site will continue. Furthermore, some data of general interest, such as experimental data from large devices—such as the ITER—will be replicated worldwide. This activity requires capabilities that can guarantee successful transfer of large data volume efficiently (i.e., taking advantage of the available bandwidth) and without introducing errors in the data being replicated.

When dealing with storage, one of the problems facing scientists today is the need to interact with a variety of storage systems, and to preallocate storage to ensure data generation and analysis tasks can take place without failure resulting from a lack of storage. Typically, different interfaces and security mechanisms are employed on different storage systems. An urgent need exists to standardize and streamline the access interface, the dynamic storage allocation, and the content management of these systems (following large international collaborations such as the Large Hadron Collider Experiment or the Earth Systems Grid). The goal is to present to scientists the same interface regardless of the type of system being used. Ideally, the management of storage allocation and access to storage systems should become transparent to scientists.

Networking capacity is expected to increase over time. Capabilities already have been added to network providers, such as ESnet and semi-private science networks such as GÉANT, to "provision" network bandwidth. Provisioning in this context means that network bandwidth can be reserved ahead of time based on expected needs. However, a challenge in provisioning resources at the extreme scale is coordinating provisioning of computing, storage, and network resources. Furthermore, replicating petabytes of data reliably from end-to-end is a challenge that must contend with transient failures of storage systems, networks, and computers at both ends. It would seem beneficial for the community to collaborate in the future with other science domains, which are facing similar challenges and have already embarked on developing solutions.

Real-Time Simulation Monitoring and Run-Time Metadata Generation

Extreme-scale computing starts in the data-generation phase and with the original computations. In the future, fusion simulations will not only run on extreme-scale machines, but they also will generate petabytes to exabytes of data that will need to be efficiently written to disk, organized, and then made searchable through complex queries. An interesting—but not unexpected—outcome of the discussion of the fusion PRDs is that in all areas, scientists have needs for both run-time monitoring of the codes and computations they use, and/or for coupling of those codes with predecessor and successor codes. Some of the coupling is relatively loose (requiring only megabytes of data to be exchanged between codes every 100 to 1000s of time steps), while other couplings need to be tight (i.e., requiring exchange of large volumes of data between the codes during every time step).

For example, extreme-scale computing offers a substantial opportunity in the area of atomistic materials modeling (i.e., plasma/matter interactions). Currently, modeling in this area is very limited in its predictive value for fusion-material problems by the variable quality, or complete lack of, inter-atomic potentials relevant to fusion materials. Modern electronic structure theory methods and particularly density functional theory (DFT) can now provide atomic forces that are approaching chemical accuracy, but their applicability is extremely limited by their high computational cost. If DFT algorithms can be developed that efficiently exploit millions of processors, predictive-quality molecular dynamics simulations on time scales approaching a microsecond would be possible for the first time. Combined with recently developed accelerated molecular dynamics (AMD) methods, time scales of microseconds and beyond will be accessible. This would represent a quantum leap in the predictive power of atomistic simulations, and would significantly impact the ability to understand critical processes at the plasma-surface interface, as well as in the bulk of structural materials; processes controlling tritium retention, cascade annealing, bubble formation, impurity precipitation; etc. Real-time monitoring of codes at this scale will be essential.

Another extreme example is when fusion scientists start to couple the core and edge of the plasma, along with magnetohydrodynamics (MHD) codes. Looking at the particle-in-cell (PIC) codes, a case could be envisioned where a delta-*f* PIC code running in the core of the plasma is coupled to a full-*f* PIC code running at the edge. These codes can be coupled to MHD codes to make sure the simulations are MHD-stable. Finally, to analyze the wall of the reactor accurately, these codes can be coupled to an atomics materials code. Finally, these codes can be coupled to a transport code to evolve the simulations to very late time. These couplings are very complex, and are as much a challenge to the applied mathematics community as to the computer science community. Advanced technologies that facilitate such coupling of codes tightly through memory, or loosely by exchanging files between the codes, are essential at extreme scales.

Monitoring

The emergence of petascale computing is already creating a tsunami of data from petascale simulations. Typically, results are analyzed by dozens of scientists, often working as teams. It is very important to help these teams by facilitating management, analysis, sharing, and visualization of the data produced by their simulations, and by the auxiliary programs and activities used in the scientific process. One aspect of this is leveraging of their collective knowledge and experiences through a scientific social network, such as MyExperiment. This can be achieved through a combination of back-end information technology services, provenance capturing, and front-end tools that are easy to use. The Scientific Data Management (SDM) Center's dashboard is one such tool (Barreto et al. 2009); see Figure 21.

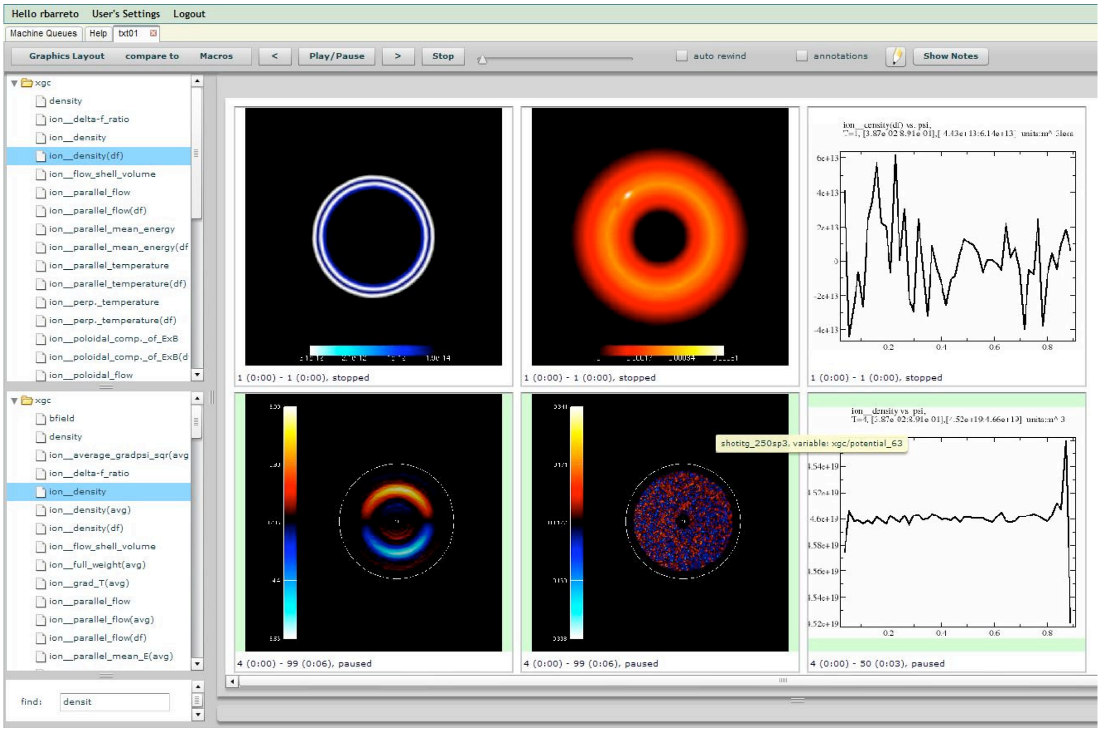


Figure 21. The Scientific Data Management Center dashboard showing pictures and graphs to monitor a fusion simulation while the simulation is running. Image courtesy of Scott Klasky and Roselyne Tchoua (Oak Ridge National Laboratory).

In all cases, simulation runs will need to deliver enough metadata to support efficient collaborations among scientists and rapid advanced analytics. For scientists to understand the complexities hidden in petabytes of data, information will need to be analyzed (in different ways that a scientist may choose), indexed, and visualized at runtime. As the number of processors per simulation grows, the required secondary calculations will need to be conducted outside of the main calculation. This means that metadata and analytics computers will be needed. This requirement comes from several simulations showing that computing the post-processing information as part of the simulation does not scale as well as the primary part of the calculation, and it adds extra complexity. Thus, computing the post-processing information could be done in the I/O pipeline (using staging nodes [Abbasi et al. 2009]), or on computers that are relatively close to the main supercomputer.

One open question is how to enable scientists to easily install their own analytics environments on the post-processing resources. Virtualization and clean on-demand reloading of these peripheral computers by end users may be one solution. Technology for that approach exists today in the "cloud computing" domain. A new paradigm similar to advanced web searching probably will emerge. A collaborative knowledge discovery model (akin to social scientific networks) will have to be developed. Such knowledge discovery mechanism will require the data to be indexed, and its metadata to be generated and available for preliminary analyses before the original raw data would be again accessed (similar to Salje et al. 2009 and Flannery et al. 2009). Run-time reduction of the raw simulation data before visualization and interactive discovery will be routine. Metadata will become as rich as content, will probably be hierarchical and tiered, and will become as valuable (perhaps even more so) than the original simulation data. In this context, the security and privacy of the metadata must be considered as much as the I/O of the simulations.

Workflow automation and collection of provenance data permits users to track lineage of the data and movement of the data from one file system to another. Current workflow systems, such as Kepler (Altintas et al. 2004), are already able to monitor complex simulations, although active steering is still an open problem. Typically, a simulation starts on a large HPC resource, and then the data may move from the HPC resource to a smaller cluster where the workflow system launches many parallel jobs that analyze and then visualize the data. Results are finally sent to a web portal where scientists can see the results, and further analyze the data. Workflow automation can also be used to run parameter surveys of the output data. With petabyte amounts of output and without automation, this could take weeks to accomplish. After a scientist identifies analyses to perform on the output from a simulation, the process can be automated.

One successful example of this is the work that the SDM Center has been doing with the CPES Project. When the users build the code, all of the provenance information about the code, along with the computers on which the code is being built, is archived into a database. When a user runs the XGC-1 simulation, they run the Kepler monitoring workflow, which moves data from the main computing platform to a smaller cluster. On this cluster, data are archived to tape, and in parallel, data are analyzed by a series of analysis programs, and then visualized. All of the metadata information about these analysis codes is also archived. After the simulation is complete, movies are created from the images, and information about the simulation is also archived into a database. Users can log onto the dashboard, and then look at the variables from the simulation and the analysis.

After the simulation is finished, users can collaborate on the dashboard. They can write notes, annotate the graphs, and share information with their collaborators. They also can perform complex analyses on backend servers, which can be Fortran code, written by the users, Matlab code, IDL code, R code, or any other analysis or visualization code. To help fusion scientists with their research, significant effort is focused on reducing the level of complexity of running fusion codes on large supercomputers.

Coupling

Gaining deep understanding and predictive capabilities in fusion science requires addressing rich and diverse physical phenomena that occur over a wide range of length and time scales spanning descriptions of different atomic and other processes and plasma behavior. Therefore, the predictive capabilities of a code can be improved through dynamic coupling with its predecessor codes. This becomes particularly important when data from predecessor codes are generated faster than it can be archived. In this case, an additional workflow step can be inserted.

An approach that employs scientific workflows and a dashboard-monitoring tool to automate and simplify the processes and the coupling is suitable for weak code-coupling scenarios. Workflow automation can start multiple jobs, provided there is an authenticated automated mechanism. Automation also can stop jobs when certain criteria are met. Only some of the metadata will be used by the workflow system; the rest can be stored in appropriate metadata repositories. For strong code coupling, an I/O system, such as ADIOS (Lofstead et al. 2009), needs to provide the user with the ability to write out data quickly, send the metadata to a repository, and if needed, send the data to another code running on the same or a different machine, or perhaps do so memory-to-memory (NxM processors). Users must be allowed to try multiple methods of coupling codes, and these methods should be pluggable to the I/O system. The workflow system needs to track all data movement, and it must allow for visual monitoring of the data from the coupled simulations.

For example, a complex production workflow is in use for coupling other codes with the XGC1 predecessor code, XGC0, where a parameter study has to be executed for each time step of the simulation. This study is then executed in the parallel mode. Inclusion of the high-frequency MHD phenomena into the XGC1 gyrokinetic framework is a longer-term goal. In the interim, while XGC1 establishes this MHD capability, simulations of edge-localized mode (ELM) crashes on the MHD time scale are being conducted using the nonlinear MHD code M3D, and then coupling it to the kinetic code XGC0 to incorporate kinetic information. The M3D code can model the diverted magnetic separatrix geometry. NIMROD is another nonlinear extended MHD code that can also be coupled to XGC0 for cross verification. The pedestal growth in XGC0 is monitored for ELM instability by a linear ideal MHD code ELITE. ELITE has been validated against many experimental datasets for the large-scale Type I ELM instability onset. A coupled simulation containing a kinetic model of edge pedestal build up and an extended MHD model of ELM behavior is a necessary component in the understanding of the pedestal-ELM cycle physics. Such a comprehensive model of this physics will yield the predicted pedestal height, which is related to the core confinement, and the expected wall load, which is related to material lifetime.

Projections to Extreme Scale

As we move to extreme-scale computing, one of the biggest obstacles that will be encountered is the much larger collaborations that will begin to occur in the fusion simulation HPC community. This increase in the number of collaborators will challenge current systems, and will move dashboard interaction much closer to a social network. Proper user interfaces become even more necessary when a wider range of users begin to use such systems. Provenance-capturing systems must also work with emerging architectures. Preproduction workflows, which test all of a code's components before the code is run, also need to be employed.

Workflow engines will need to become even more robust, and capable of running for months with very little human intervention. Fault tolerance will become a major challenge for all aspects of extreme-scale computing, including all of the technologies discussed above.

An example of coupling and monitoring that will be needed at extreme-scale computational levels is offered by plasma/material-interaction codes. Improved—and ultimately predictive—understanding of the properties, performance, and feedback of the fusion plasma with wall materials is highly dependent on developing robust and coupled models spanning phenomena on many time and length scales. These include small (short) and large (long) scale behavior of the plasma near the wall, and the interaction of this plasma with the surface of the wall and with deeper layers of the wall. One regime within this range contains the atomic, molecular, and near-surface particle-surface interaction (AM&PSI), which is characterized by femto- to pico-second times and lengths on the order of nanometers. At this scale, electrons collide and excite ions, molecules are sputtered from ion bombardment of the wall, and photons are emitted from diagnostic ions. Models at this level of challenge present computational capabilities and will require extreme-scale computing to 1) treat effects presently identified with greater fidelity, and 2) enable, for the first time, consideration of a larger number of particles, reaching even full quantum simulations that elucidate more uncertain or unknown effects.

Because of the short time and small scale of these interactions, and the need to use the largest computational facilities to describe them, the coupling of physics results from this level with other levels of multiscale modeling is best performed through a particular workflow taking advantage of this. That is, these physics results should be computed, and then stored within an appropriate file structure separately from higher level, multiscale runs. This approach makes sense because the AM&PSI physics possesses "similarity." In particular, each time an electron approaches a molecule, or an ion reaches a material surface, the multiscale code should not digress into a separate supercomputing solution of this AM&PSI scale interaction. Rather, the required outcome should be retrieved from storage because it had been previously generated and verified. This AM&PSI physics should be successively improved and expanded beginning from an initial, less-complete, less-verified, less-detailed version. Thus, the AM&PSI physics results flow to the next higher level of simulation (e.g., particle transport and plasma species, charge, and energy evolution, and separately to other levels that require them). In addition, this workflow allows for use of the AM&PSI in other applications such as flow to experimental analysis suites or plasma diagnostics. While the current suite of AM&PSI codes are not implemented with much or any dynamic coupling (although, of course, results from precedent codes are used as inputs), an effort to implement such functionality should be one of the major activities in the extreme-scale context.

Data Analysis at the Extreme Scale

Motivation

There are two types of analysis problems that need to be solved as movement continues towards extreme–scale computing in the fusion sciences. The first problem is analysis of data from the simulations themselves. These data are likely to be petabyte-sized and unstructured, high-dimensional, and spatial-temporal in nature. They may arise from multiphysics simulations as well as more focused studies (e.g., understanding how the properties of materials used in plasma-facing surfaces would evolve over time). The second problem is analysis of data from experiments that are used to validate the simulations or to refine the theories underlying the simulations. These data are likely to be in the form of images or image sequences. The images, though currently smaller in size compared to the simulation data, are usually noisy. Experimental data also may result from sensors used to monitor experiments such as DIII-D[1] or the ITER. These data are in the form of time series, and are likely to have noise, outliers, and missing values. While the experimental data are not directly the result of extreme-scale computing, and are currently smaller in size, they are essential to building the right codes—and with their exponentially growing data volumes, they will in turn require extreme-scale computing for their real time analysis. Further, these data are quite complex, and their analysis is far from a solved problem.

In the context of fusion applications, analysis techniques can be used to 1) understand and gain insights into the phenomena being simulated (e.g., through the extraction of coherent structures and their interactions); 2) build code surrogates to complement expensive simulations; 3) validate simulations by comparing them with experiments; and 4) provide insights into experimental data.

Given the increasing volume of data that will need to be processed in the analysis phase, various analysis tasks are expected to require facilities and codes that can run in parallel. Examples in the fusion domain

[1] The mission of the DIII-D Research Program is to establish the scientific basis for the optimization of the tokamak approach to fusion energy production. The DIII-D Program is a cornerstone element in the national fusion program strategy.

that will require parallel analysis facilities are 1) quantify energy distribution among zonal flows, geodesic acoustic modes (GAMs), and mean flows; 2) characterize how elongation affects the ion temperature gradient (ITG) linear stability; 3) validate statistical significance of the front-tracking boundary that separates the noise from the signal as the simulation progresses; and 4) quantitatively compare zonal flow behaviors without collisions and with collisions in the density slope after nonlinear saturation.

In addition, the sufficient annotation and accessibility of both simulation and experimental data need to be addressed to effectively support the various analysis capabilities listed above—in particular, community support for metadata standard development, data quality assessment, and the establishment of designated data centers connected to high-speed networks or co-located with data-intensive computing capabilities will be crucial.

State of the Art

There are a wide range of data analysis problems that are encountered in fusion science applications. Based on their size and complexity, they can be categorized as follows.

Small-to-Moderate-Sized, Relatively Simple Problems

These include problems such as the identification of key sensors for edge harmonic oscillations in DIII-D data. While these problems are nontrivial as they require preprocessing of the data to appropriately handle outliers and missing values, there are several algorithms available in the dimension-reduction literature that can be used to provide a solution. Further, the availability of several algorithms and the well-defined nature of the problem make it easy to gain confidence in the results of the analysis.

Small-to-Moderate-Sized, Complex Problems

These problems range in size from kilobytes to gigabytes and are challenging because of their complexity. Several factors can contribute to the complexity. The data could be noisy, making it difficult to identify the structures of interest in the data. The problem could be poorly understood, resulting in the lack of a definition of these structures. Alternatively, the type of data could be unique (e.g., as points in space tracing out a shape [as in Poincaré plots], simulation data on an unstructured mesh, making existing algorithms inapplicable, etc.). The current state of the art in analysis algorithms provides only partial solutions to these problems. Often, existing solution techniques have to be extended, or new techniques developed before a satisfactory solution can be found. Problems where discovery is the end goal, or where the problem itself is poorly understood, pose an additional challenge. Any analysis algorithm must be robust enough that the results obtained reflect the data rather than the algorithm or the choice of parameters. This is currently an active research area in analysis.

Large-to-Massive-Sized, Complex Problems

When the problem is complex, and the data are terabyte-sized or more, the analysis problem becomes much harder. For example, one may need to identify coherent structures in data on an unstructured mesh, where there is no definition of a coherent structure, a poor understanding of the science, and the data are distributed among multiple files for a single time step. The problem becomes more complex if, in addition, we need to understand the nonlinear interactions among coherent structures in fluid and particle data, or track a front, which spans multiple processors, over time. Further, if the results from the analysis must be obtained in a short time—for example, between two shots of the ITER—it would indicate extensive testing of the analysis algorithms is required to ensure smooth operation when deployed.

Analysis at the Extreme Scale

As the data from the simulations and experiments increase in size, the current analysis problems become aggravated. The challenges associated with small-to-moderate-sized complex data, which may arise in code validation or in improving current understanding of the theories that underlie simulations, remain. These challenges, though not driven by the extreme-scale nature of the data, must be addressed if scientists are to benefit from enhanced computing capabilities.

For problems where the size of the data is measured at terabytes and beyond levels, several analysis challenges must be faced. First, areas of interest in the data, such as coherent structures and fronts, are likely to be spread across many processors (and files output by these processors), making it difficult to extract the poorly defined structures or track the fronts over time. Second, techniques to process the data to reduce their size before they are output by the simulation will not work for the case where the analysis is done to understand the science in the data, or the algorithms and associated parameters are not robust enough to process the data correctly. Finally, if a fast turnaround of the analysis is required (e.g., to design the next shot of the ITER), then extensive testing of the analysis algorithms and parameters must be done to ensure they can handle data with all the possible statistical characteristics generated as a result of the experiments.

Analyzing fusion simulation data is particularly challenging because of the multiscale, nonlinear turbulence-dynamics nature of such simulations. The next section describes approaches to such simulations and the challenges they pose at the extreme scale.

Challenges in Analyzing Multiscale Nonlinear Turbulence Data

Analyzing fusion simulation data is particularly challenging because of the multiscale, nonlinear turbulence dynamics nature of such simulations. Understanding multiscale, nonlinear turbulence dynamics is the key driver for many emerging extreme-scale fusion plasma applications. FES is facing the formidable challenge of developing a predictive capability for plasma performance in the ITER using simulation codes based on first principles. There are two known approaches to first-principles fusion simulation in solving the Fokker-Planck particle differential equation (PDE) in five dimensions in realistic fusion device geometry. One approach is the continuum approach, solving the Fokker-Planck equation on a five-dimensional grid. The other approach is the PIC approach that involves time advancing maker particles along the characteristics and solving the field equation in three-dimensional real space. Both approaches are time varying. Additionally, there are hybrid approaches that use fluid techniques along with the PIC techniques. An ITER-size simulation could produce over a terabyte of fluid data and several petabytes of particle data. The PIC approach demands extreme-scale data analysis because of the high-velocity space resolution capability created by almost a trillion marker particles in extreme-scale computing of ITER plasma.

Development of proper mathematical and computer science tools for extreme-scale data analysis is essential in the particle simulation of fusion plasmas represented by these three Scientific Discovery through Advanced Computing Program (SciDAC) centers. Such tools can greatly enhance scientific understanding of complicated multiscale dynamics in turbulence data, and provide a quantifiable and detailed stress test of both theory and simulation for verification and validation work and for advancing physics. This problem is exacerbated for the science questions that cannot be easily addressed by simultaneous spatial-temporal turbulence measurements in laboratory experiments. A number of analysis requirements for characterization of multidimensional turbulent dynamics at the extreme scale have been recently articulated (Kritz and Keyes 2007) and were emphasized further during this workshop. Next, we highlight some of the key analysis needs for properly advancing FES and articulate mathematical and computer science challenges in addressing those needs.

Challenge: To Enable Structural, Compositional and Functional Characterization of Multiresolution, Multivariate Spatial-Temporal Signals in Five-Dimensional Phase Space (Three-Dimensional Real and Two-Dimensional Velocity)

The signals produced by the extreme-scale simulations are inherently complex because of the nonlinear, multiscale, and dynamic nature of the physical phenomena under study. In plasma fusion, critical gradient transport dynamics akin to self-organized criticality, breaking of gyro-Bohm transport scaling by intermittent extended eddys, transport bifurcation, and transport-barrier formation are all generated and regulated by the nonlinear interaction between gradient-driven small scaled drift turbulence and larger meso-scaled structures such as zonal flows, avalanches, fronts, and streamers. Yet, techniques for analyzing these complex signals are in their infancy. In particular, the phase-space dynamics of meso-scale structures has not been seriously addressed. These issues are also surely relevant to stellar and galactic dynamics simulations, so they are of interest beyond the magnetic fusion energy field. In particular, the following data-analysis capabilities currently are lacking and demanding the development of proper analysis tools:

Quantification of Fluctuation Energy Distribution. The observed composite energy signal is distributed in spatial and temporal scales across zonal flows (induced by transfer of turbulent kinetic energy from micro-scales to meso-scales), GAMs (similar to zonal flows), and mean flows with complex profiles, etc. Different energy distribution profiles need to be quantified.

Mapping and Tracking the Evolution of the Population of Structures in the Spatial and Temporal Scales. There is a need for automatically identifying phase-space structures, such as blobs, or clumps and holes. Akin to blobs, a front-tracking capability needs to be developed to establish the speed and dynamics of turbulence-front propagation that is critical to understanding nonlocality and nonlocal response in tokamak turbulence. The challenge is that different parts of turbulence front most assuredly will have different properties and scales, such as precursor oscillations, small-scale eddies, meso-scale (scale between ion gyro-radius and system size) zonal flows, and GAMs. The analysis should track both the fluctuation envelope and the thermal pulse.

Challenge: To Enable Statistically Quantifiable Noise Diagnostics in 5D Phase Space

Noise removal from the signal is a critical step. Complete output data from turbulent-dynamics simulations contains noise; the complexity necessitates advanced noise-analysis tools. Specifically, there is a need to develop approaches for detecting and classifying the color of the noise (e.g., white, red, pink, blue, or none of the above) as well as noise separation and smoothing techniques. Existing methods are designed primarily for low-dimensional, stationary time-series data; they also are not quite suitable for multiresolution, both in space and in time, data.

Challenge: To Enable Sensitivity Analysis for Validation of Simulation Codes

Individual simulation runs at the extreme scale are both computationally and data-intensive. Typical approaches to sensitivity analysis are based on performing multiple runs and then checking the sensitivity of the results to the input parameters. Because individual runs are expensive, there is a need for developing robust sensitivity-analysis techniques that could operate with much fewer available runs and to guide the designs of simulation experiments to proper inputs for a given physics study. Such analyses should be coupled with experimental data and take into considerations the uncertainties present at both the simulation code level and the experiment level. The reliable determination of these uncertainties is another challenge that needs to be addressed to conserve scientific validity.

Challenge: To Scale Analysis Codes to the Extreme-Scale

Many components of fusion data analysis pipelines scale nonlinearly with the problem sizes. For example, a routinely used technique in almost any data-analysis experiment is the computation of a two-point correlation function, including cross correlations of different fields. Existing analysis tools such as GKB and Vugyro are written in scripting languages such as IDL or Matlab. They do not scale to multidimensional datasets (e.g., four dimensions) and/or to trillions of particles. There is a need for accurately estimating different correlation functions and spectral densities. Possible approaches in this direction could deploy bi-spectral analysis techniques (Holland et al. 2003), along with parallel optimized implementations of such analysis codes. If petascale computing capacity is required to analyze the data, new communication libraries are required to support the exploitation of effective data tiling and locality in memory during analysis, such as is provided by the Global Arrays.

Visualization of Very Large Datasets

The challenges for visualizing extreme-scale fusion simulations will be many. Computational fusion scientists estimate that their grid sizes will reach 1 trillion to 10 trillion cells. Further, they indicate that as many as 120 trillion particles may be used in their simulations in a single time step. This incredible scale will worsen a litany of existing visualization and analysis problems. The most significant problems will be 1) managing the scale of this data (i.e., how to process petabyte amounts of data); 2) how to address the distributed nature of the data; and 3) how to gain scientific insight from these data.

Managing the Scale of Data

The scale issue will be particularly prevalent in the future. The current paradigm for processing large datasets is "pure parallelism," where parallel resources are used to read in the data, perform data mapping algorithms (like contouring and slicing), and then render the results (see Figure 22). Extrapolating for data size, visualization clusters of thousands of processors to tens of thousands of processors would be required to look at this large scale. Worse, this "small" estimate of the number of processors assumes that each processor contains multiple gigabytes of memory and has high-bandwidth access to the file system. Restated, this quote of several thousand processors will likely translate to a much higher amount if the processors are from an extreme-scale machine. To overcome this, the visualization community will have to deliver alternative approaches to pure parallelism. There are no panaceas; use cases and scalable approaches to solving these use cases must be identified. For example, for interactive data exploration and visual debugging use cases, multiresolution methods will need to be employed. For these methods to be successful, multiresolution hierarchies need to be generated when the simulation code writes data.

Then, a visualization tool can load reduced resolutions quickly, while maintaining the ability to "dive down" on high-resolution data where needed. Further, for data-analysis use cases, the visualization community will need to deliver *in situ* solutions, with the solutions being highly scalable and also having a low memory footprint. This approach allows data to be processed at full resolution (a requirement for analysis and comparative use cases) and quickly, and also avoids all writing of data to disk.

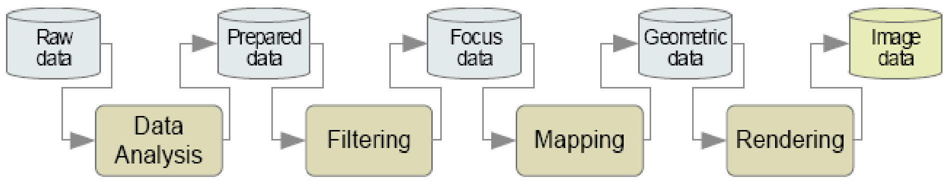


Figure 22. The visualization pipeline. Reprinted from *Computers and Graphics*, Vol. 28, Issue 3, S. dos Santos and K Brodlie, "Gaining Understanding of Multivariate and Multidimensional Data Through Visualization," page 15, Copyright 2004, with permission from Elsevier.

Descriptions of the main processing steps in the visualization pipeline follow:

- Data analysis: Data are prepared for visualization (e.g., by applying a smoothing filter, interpolating missing values, or correcting erroneous measurements).
- Filtering: Data portions to be visualized are selected.
- Mapping: Focus data are mapped to geometric primitives (e.g., points, lines) and their attributes (e.g., color, position, size); most critical step for achieving expressiveness and effectiveness.
- Rendering: Geometric data are transformed to image data.

Gaining Scientific Insight

The scientific insight issue will also be a major challenge. With trillions of data points and only millions of pixels on a display, it is obvious that not all of these data can be represented fully in a single image. There are multiple approaches to this problem. One approach is to emphasize interactivity, so that a user may navigate quickly around the dataset and explore the regions of interest at its full resolution. Another approach is to incorporate statistical summarizations to ensure that all data is represented in some way. Automated feature identification is yet another approach. Likely, a combination of these, and more, approaches will be needed to enable scientific insight on these extreme datasets.

Currently, the two visualization packages ParaView (Law et al. 2001; Cedilnik et al. 2006) and VisIt (Childs et al. 2005) have some underlying rendering foundation for addressing large-scale datasets by using remote distributed memory resources for visualization. However, many challenges still exist. Many of these challenges are similar to those currently faced by computational scientists including parallelization, load balancing, data distribution, I/O optimization, and network latency. The need for the final images to be generated in real time further compounds this challenge. This is especially important during the data exploration phase when a scientist wishes to see the data in real time.

While traditionally fusion scientists have viewed visualization as being a post-simulation process, as the size of datasets grows, they will rely more on it during the simulation monitoring process. For example, if a simulation is not exhibiting a desired phenomena or trait, letting the simulation run to completion may possibly waste expensive resources. In this situation, the scientist may choose to terminate the simulation or perhaps drive it in a different direction. To allow visualization tools to be used with web-based dashboard systems for monitoring large-scale fusion simulations, they must have robust interfaces that allow their integration into the dashboard systems.

Current visualization tools such as ElVis (Feibush et al. 2006), which use web-based tools and Java, have been incorporated into the Fusion Grid Monitoring system (Schissel 2004), and have some of the necessary underlying visualization capabilities required by fusion scientists. However, current focus of ElVis has been for monitoring two-dimensional codes such as TRANSP (Ongena et al. 1998), which have modest data visualization needs. Integrating more complex visualization tools for large-scale, three-dimensional codes into dashboard monitoring systems will require using remote resources with the resulting images being streamed over the network. This approach will face many of the same challenges described above (i.e., parallelization, load balancing, data distribution, I/O optimization, and network latency).

Another challenge lies in the ability for fusion scientists to collaboratively share their visualizations. Fusion simulation codes are a collective effort by groups of scientists who are rarely geographically co-located. As more codes, each representing specific physics, are integrated, it will be necessary to have tools that allow scientists to work collaboratively (Cummings and Kiesler 2005). Today's visualization tools are rarely designed around a collaborative environment and have instead relied on third-party applications such as Chromium and VNC.

Scientific Insight at the Extreme Scale

The other prevalent challenge, which has had much less attention, is how best to gain insight into extreme-scale datasets. Fundamentally speaking, how would scientific data be transformed into an image that provides insight to the fusion scientist? This is a critical step, as it dictates how effective the visualization will be to the scientist. The transformation process may be as simple as creating a series of iso-temperature surfaces of the plasma to more complex processes such as those that form flux surfaces from the magnetic field data.

Even simple transformations such as creating iso-contours, which are often used to gain initial insight into the data volume, will pose challenges as the dataset size increases to the terabyte-exabyte scale because it will be impractical to view all of these data at once. Even with the size of some of today's datasets, it has become impractical. One strategy is to use Level-of-Detail (Clark 1976) techniques where smaller features that would otherwise not be seen are merged into larger features. As the data size increases, the number of levels will increase and become more complex as the levels must be managed both in and out of core memory. Another strategy is to cull features that would otherwise be hidden from view by other features (Livnat and Hansen 1998). However, because the culling is view dependent, it must be done on a view-by-view basis; while it has the potential to greatly reduce the rendering time, it is dependent on not only the overhead of the culling process but the complexity of the data. Both strategies will face challenges on distributed memory computing resources, especially as fusion codes that use multiple grids are integrated.

Preserving Salient Features of Large-Scale Data

Another challenge is the development of algorithms that preserve salient features of the data while reducing the amount of geometry that must be rendered. These algorithms may be purely analysis algorithms such as those that search for topological features (Sanderson et al. 2006), or purely mapping algorithms such as query-based tools (Ruebel et al. 2008) or a combination. An example of this challenge is the visualization of fusion simulations that contain thousands to billions of particles, each of which hold multivariate data. To create an effective visualization, only a few hundred particles can be viewed at any one time (Jones et al. 2007, 2008). While seeing particles alone can provide some insight, visualizing the particles in the context of other variables, such as the electric potential, can provide even greater insight. One promising approach is for scientific visualization to adopt techniques from the information visual analytics community to help address the growing data heterogeneity and complexity.

In-context visualizations rely not only on query-base tools that are able to quickly search through the data, but must be able to also correlate the results across other data fields. With larger data volumes, it will be necessary to have even more powerful query tools that can use distributed systems and search between multiple types of data while returning the information in such a way that the fusion scientist can quickly assimilate the results.

A similar challenge, but at the opposite scale, is rather than seeing individual data points, fusion scientists will need to be able to visualize trends within all of the data while preserving anomalous features. This becomes even more challenging when multiple variables are included, especially as fusion codes are integrated over multiple spatial and temporal scales. Both of these examples highlight the challenge of transforming a particular phenomenon and visualizing the salient features.

Experiment-Simulation Data Comparison

Motivation

The ITER is an international burning plasma magnetic confinement experiment under construction in Europe.[1] It is the next major step toward proving the scientific viability of controlled fusion as an energy source. Current estimates for data volume and rates during the ITER's initial operation about a decade from now are 1 to 10 terabytes per pulse approximately once per hour. Thus, the total data collected will be on the order of 1 to 10 petabytes per year. The needs for network connectivity are estimated to be multiples of 10 gigabytes/second.

While this data quantity is low compared to the volume of data expected from full-device simulations, these data need to be analyzed and compared with extreme-scale data from simulations. The use of large-scale simulations for supporting design decisions, experiment planning, and performance optimization are critical for the ITER as they will increase the likelihood of success. However, this cannot be achieved without first demonstrating the validity of computational models in the ITER's experimental regime. Assessment of these models is divided into two distinct activities: 1) verification, which assesses the degree to which a code correctly implements the chosen physical model; and 2) validation, which assesses the degree to which a code describes the real world. These issues of verification and validation is of general concern throughout the computational physics community, as evidenced in the special issue of *Computers in Science and Engineering* (Trucano and Post 2004), the number of grant solicitations requiring a verification and validation plan, and in the recent editorial board statement of *Physics of Plasmas* (Davidson 2007), and several other papers (e.g., Terry et al. 2008).

[1] For more information on the ITER project, see http://www.iter.org/default.aspx.

Because confidence in the ability to predict is ultimately based on code performance measured against experimental data, a vigorous and ongoing validation activity is critical to the success of an extreme-scale computing mission. Synthetic diagnostics are a fundamental method of conducting model validation as they are supposed to produce a direct and/or statistical comparison of experimental and simulated data. The challenge of designing synthetic diagnostics comes from the fact that fusion experiments and simulations produce data using different variables, generated in different frames, and existing in different spatial and temporal basis.

What is needed is the ability to classify the types of diagnostics and provide a means to transform extreme-scale data from simulations and experiments to a common denominator. Ideally, this comparison needs to be done on the time scale that the experimental data are generated. A unique feature in the operation of fusion energy experiments is the requirement to access, analyze, visualize, and assimilate data between shots in near-real-time to support decision making during operations. This contrasts with large experiments in other fields, such as high energy or nuclear physics, which operate primarily in a batch mode. Fusion experiments put a particular premium on near real-time interactions with data and among members of the team; thus tools supporting real-time, distributed, collaboration and analysis will be especially valued.

A key part of the validation process is performing sensitivity analysis (i.e., performing multiple runs to see how sensitive results are to a code's input). This entails multiple simulations for a given physics study. Furthermore, it is necessary to relay experimental uncertainties to simulations so that they could run sensitivity studies and optimization runs based on the experimental uncertainty.

Although potentially powerful, synthetic diagnostics are not routinely used within the fusion community today. While not difficult in principle, synthetic diagnostics typically have been developed by the authors of individual codes using many hard-coded parameters. This style of development has prevented reuse of synthetic diagnostic across simulation codes, and makes it more difficult to implement new diagnostics. The process of synthetic diagnostic development and use needs to be abstracted, making the building blocks more readable and available to the community and thereby lowering the barrier of creation and increasing their usage. This more open and generalized approach will benefit all fusion code projects, and further the "culture of validation."

The SciDAC program in fostering the development of high-fidelity fusion computer codes, which leverage leadership-class facilities and the simulations, are rapidly moving to these platforms. As they move to the extreme scale, they will need to be validated on the extreme-scale platforms. This means that the synthetic diagnostics need to be developed as cross-platform-scalable modules.

State of the Art

Synthetic diagnostics in the fusion program can be placed into two categories: 1) coherent-mode diagnostics and 2) turbulence diagnostics. In the first category, the goal is to produce a direct comparison of an experimental signal from simulation data. An example is the reproduction of soft x-ray data from extended MHD codes (Park et al. 1995; Brennan et al. 2003; Kruger et al. 2005) to compare the structure of the internal kink mode using data from soft x-ray diagnostics and Electron Cyclotron Emission diagnostics.

For turbulence diagnostics, the chaotic nature of the turbulence itself makes direct comparisons of time traces meaningless, leaving only comparisons of statistical properties of the turbulence as being relevant. Thus, the goal of a synthetic turbulence diagnostic is not to reproduce a particular observed turbulent eddy, but rather to predict the power spectrum and other relevant statistical measures the modeled diagnostic is expected to observe. An example is work performed at the University of California-San Diego (Holland et al. 2008) that compared GYRO (Candy and Waltz 2003) simulation data with beam emission spectroscopy (BES) (McKee et al. 2007) and correlation electron cyclotron emission (CECE) (White et al. 2008) radiometry data from the DIII-D tokamak (Luxon 2002) (see Figure 23).

Existing synthetic diagnostics have the following commonalities:

- Common data transformations
  - Mapping between variables and units
  - Transformation of the code frame (ion rest frame) to laboratory frame
  - Interpolation of data
  - Transformation of simulation data into synthetic signal based on type of the diagnostic
  - Common post-processing steps.
- Feature extractions
  - Data reduction and sub-selection
  - Statistical analysis
  - Visualization.

On different stages of these steps, one might require movement of unprocessed or processed data from a remote system to a local system.

Projection to Extreme-Scale Challenges in Developing Synthetic Diagnostics

Moving to the extreme-scale computing will require the following:

1. Standardization on data formats, metadata formats, and creating common APIs for accessing simulation and experimental data
2. Adding “Remoting” capabilities to these APIs to minimize data movements
3. Automation and semi-automation of the steps discussed in the previous session (data transformation, analysis, visualization and movement)
4. Introducing good software engineering methods such as version control, automated build system, regression tests, performance tuning, and using standard external packages. Furthermore, such tools should be adaptable to different exascale architectures, and support cross-compilations on such architectures.

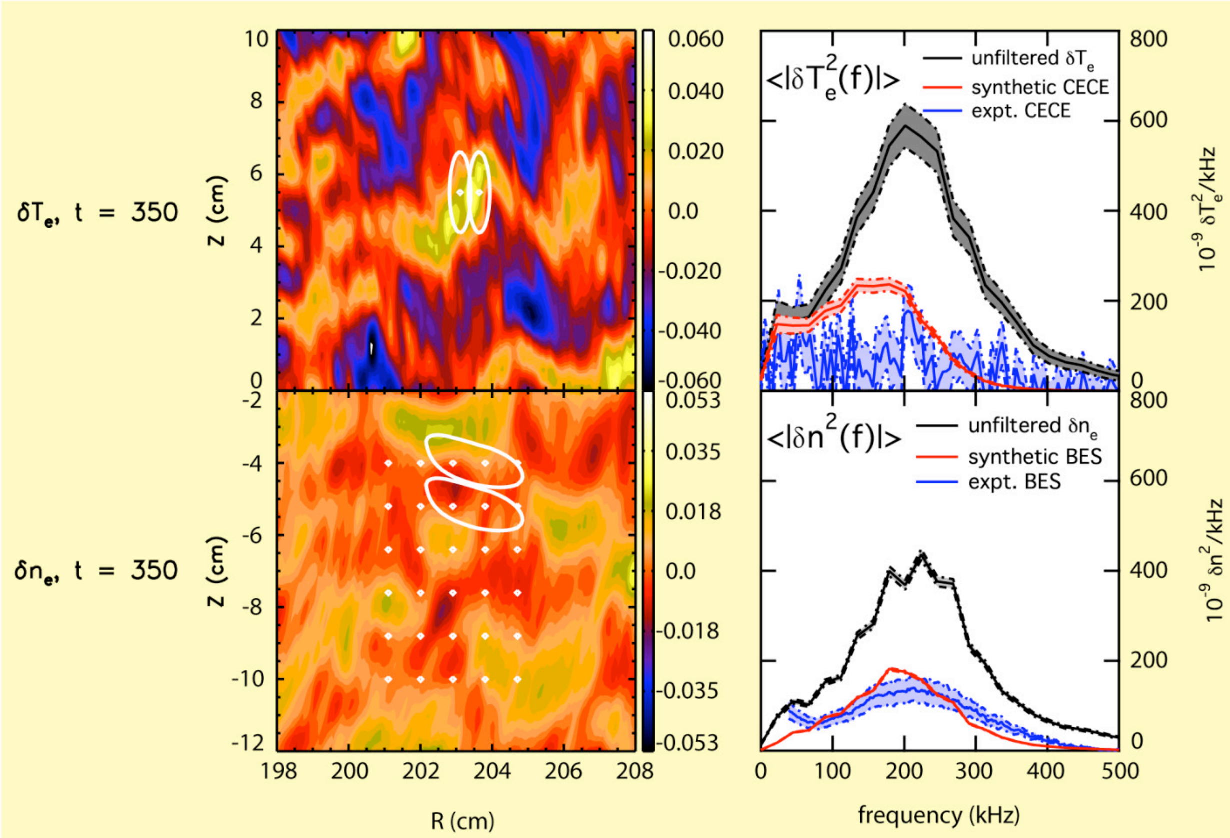


Figure 23. The left-side images are BES (bottom left) and CECE (top left) point-spread functions (PSFs) in the (R,Z) plane at rho=0.5 for a GYRO simulation. The right-side graphs show a comparison of the unfiltered (black), synthetic (red), and experimentally measured from the DIII-D tokamak (blue) lab-frame spectra of density and electron temperature fluctuations. In white in the left-side images, the 50% contours of the BES and CECE PSFs are shown with the diamonds indicating the center locations of the individual diagnostics. This figure shows that the synthetic diagnostics are essential for quantitative simulation-experiment comparisons, and with their application, good agreement is found between the simulation and experimental results. Image courtesy of C. Holland (University of California-San Diego).

CONCLUSIONS

This panel report covers five main areas related to data management, analysis, and visualization of fusion data, and describes the state-of-the-art in terms of requirements, and projections of the effects of extreme-scale computing in these areas. These areas are as follows:

- managing large-scale I/O volume and data movement
- real-time simulation monitoring and run-time metadata generation
- data analysis at the extreme scale
- visualization of very large datasets
- experiment-simulation data comparison.

Summarized below are the conclusions in each of these areas and projections at the extreme scale.

1. Managing Large-Scale Input/Output Volume and Data Movement. As system compute capabilities continue to scale toward the extreme scale, storage systems must adapt to address the increasing

bandwidth and storage challenges. In many simulation codes, the volume of data generated per core per time step is in the order of 2 gigabytes. Thus, the size of dataset produced by a single simulation on a 100,000 core machine is already about 200 terabytes. At this extreme scale, I/O storage and its use may have to focus on new approaches that consider architectures, the role of I/O systems, data formats, and I/O and storage systems as a vehicle for knowledge discovery. Three areas will have to have to be greatly enhanced to prevent I/O from becoming the main bottleneck that will slow down extreme-scale computations: 1) the file system associated with the extreme-scale computing facility; 2) providing efficient high-level I/O libraries; and 3) high-level APIs to support multiscale models that can achieve the desired application performance. Having a data center where midscale computational and extreme-scale storage resources are co-located may address some of the challenges that arise when petabyte datasets are generated at multiple locations.

2. <u>Real-Time Simulation Monitoring and Run-Time Metadata Generation.</u> Real-time simulation monitoring is essential to fusion simulations— for example to stop simulations that do not proceed as expected, to prevent wasting precious computation resources, or to allow spontaneous generation of tiered metadata about the computations and the results. Furthermore, the ability to couple multiple codes in real time will become essential to take advantage of the computational power and to increase predictive fidelity of the more complex codes that are emerging. Provenance capture becomes another critical piece during the data-generation phase because scientists need to be able to link their data from the original simulation code to the analysis pieces, and then all of the way through complex analysis and visualization. Tools that will automate the monitoring, provenance collection, and code-coupling will be essential as the volume of data grows in the extreme scale, especially when modeling and running simulations that involve multiscale physics.
3. <u>Data Analysis at the Extreme Scale.</u> The data analysis challenges in fusion energy science applications at the extreme scale stem not only from the large size of the data, but also the complexity of the data. Several tasks that support extreme-scale simulations, such as code validation and refinement of theories, involve small-to-moderate-sized data. Analysis of these data can be difficult as the task may be poorly defined, the science not well understood, and the data quite complex in the form of multiscale, noisy, time-varying images, or points in three-dimensional space.

   For problems where the data sizes approach terabytes and beyond, there is an added dimension to the challenges faced in the analysis. Specifically, analysis algorithms must not only handle all the issues arising from the complexity of the data, but must do so for data that will be distributed across many files and for analysis that may require a fast turnaround to keep pace with experiments. Techniques for data reduction that preserve the essential features of the data will have to be developed to allow for timely analysis capabilities. Furthermore, when code is parallelized, it partitions the problem on multiple cores, thus requiring the coordination of calculations across the boundaries of the partitioning. This can cause inaccuracies. Consequently, current analysis methods will need to be parallelized to the extent possible without compromising their accuracy.

   To successfully solve these analysis problems, current techniques from image and video processing, machine learning, statistics, and pattern recognition must be enhanced. In addition, new approaches must be found that are more robust and that can handle the diversity of data types, the variations within a dataset, the distributed nature of the data, and any physics-driven challenges to the analysis. Such developments should be done in close collaboration with physicists.
4. <u>Visualization of Very Large Datasets.</u> There are three fundamental challenges for visualizing extreme-scale data volumes: the need to effectively utilize remote distributed resources,

dissemination of the results, and algorithms to extract salient features. These challenges are interrelated and will require a close collaboration between those designing visualization systems and fusion scientists who will use these systems. As more visualization tools such as VisIt use a remote client architecture, creating collaborative environments will be easier to implement as part of their infrastructure. However, the challenge, which is true today, is compatibility across multiple architectures. As extreme scale-computing becomes more prevalent, a variety of diverse architectures will appear. It will be a great challenge for these tools to operate and perform well on these architectures as they rely more heavily on I/O and less heavily on central processing unit usage, which is the opposite of simulation codes. Visualization of large-scale data rely on tools that can quickly search through the data and also correlate the results from multiple data fields. At the extreme scale, it will be necessary to have even more powerful query tools that can run on parallel platforms, search multiple types of data, and generate visual presentations that help fusion scientists understand the salient features of the data.

5. <u>Experiment-Simulation Data Comparison.</u> The advances in plasma diagnostics combined with the advances in computational models and available computing power have created new opportunities for simulation validation. These advancing trends will continue to accelerate and, at the extreme-scale computing level, will represent a significantly more demanding challenge for model validation. The computer science community can greatly assist the fusion energy science community by creating cross-platform-reusable, interoperable, and scalable components for common elements required to perform comparison of experimental and simulation data, by automating some workflows in this process, and by introducing new software engineering practices into developing commonly available validation software.

Progress in these five areas will require teams of fusion energy scientists, applied mathematicians, and computer scientists to address problems across the range of physics, algorithms, data management, dynamic load balancing, and code modernization. Detailed discussions of the research needed to make substantive progress in all of these key priority research directions are provided in the preceding main section of this panel report.

MATHEMATICAL FORMULATIONS

Co-Leads: Phil Colella, Lawrence Berkeley National Laboratory
Ravi Samtaney, Princeton Plasma Physics Laboratory

Panel Members: Herbert Berk, University of Texas at Austin; Kevin Bowers, Los Alamos National Laboratory; Jeff Candy, General Atomics; Luis Chacon, Oak Ridge National Laboratory; C.S. Chang, New York University; William Daughton, Los Alamos National Laboratory; Bill Dorland, University of Maryland; Milo Dorr, Lawrence Livermore National Laboratory; Leslie Greengard, New York University; Wei-Li Lee, Princeton Plasma Physics Laboratory; Daniel I. Meiron, California Institute of Technology; Hong Qin, Princeton Physics Plasma Laboratory; Daniel Reynolds, Southern Methodist University; Bruce Scott, Institute of Plasma Physics; and Xianzhu Tang, Los Alamos National Laboratory

SUMMARY

The purpose of this panel was to identify the scientific challenges in the five topical areas of plasma physics, the principal bottlenecks for which lie within the development of new models and discretization methods. The members of the physics panels who attended the Mathematical Formulations panel skewed heavily in the direction of kinetic modeling in plasma physics, so the results should be viewed as indicative of the range of possibilities. However, it is clear that there are central problems in each of these areas for which possible new models and algorithms could play an important role.

Burning Plasma/ITER

The principal challenge in the area of burning plasma/ITER is the need for high-fidelity kinetics calculations, both in the core and edge regions. Issues raised during the panel included the following: more accurate gyrokinetic approximations, systematic methods for constructing nearly field-aligned coordinates, fundamental new numerical algorithms for particle-in-cell (PIC), the need (or lack thereof) for symplectic integrators for both particle- and continuum-based methods, and treatments of kinetic electrons.

Integrated Modeling

The principal challenge in integrated modeling is the need for a mathematically systematic treatment of coupled systems with vastly different spatial and/or temporal scales, including well-posedness, stability, and accuracy. A classic example of this is the coupled treatment of turbulence and transport.

Plasma-Material Interaction

The principal challenge in plasma-material interaction is the design of materials to withstand tokamak operating conditions, a topic that is out of the scope of numerical plasma physics. A second issue is the interaction of the plasma environment with material boundaries. In the latter area, topics include the improvement of the fidelity of edge models with respect to the interaction with the boundary, the effects of impurities on the overall plasma, and the impact of liquid walls.

Laser-Plasma Interaction

The main challenge in laser-plasma interaction is understanding the interaction of the laser with plasma heterogeneities known as speckles. Mathematically, this is a homogenization problem: one wants to understand and represent the collective effect of thousands of speckles, while currently it is possible to compute the interaction of the laser with one such speckle. This leads to the development of reduced/mesoscale models derived from very large-scale, high-end computing (HEC) calculations.

Magnetic Reconnection

Magnetic reconnection is a multiscale problem, exhibiting kinetic behavior in highly localized regions in space and fluid behavior on larger scales. The traditional approach of using two-fluid extended magnetohydrodynamic (MHD) is questionable physically (particularly for larger-scale problems) and difficult numerically, while the kinetic models that are correct in reconnection zones are too expensive to use globally. This is an opportunity to introduce hybrid fluid-kinetic models that have been used successfully in other areas of plasma dynamics.

SCIENTIFIC CHALLENGES AND RESEARCH APPROACHES

This section presents the applied mathematics research challenges as identified by the Mathematical Formulations panel members in a breakout session that focused on the five physics areas of interest to the fusion physics community.

Each of the physics areas has its own set of unique mathematical and computer science challenges. One unfortunate consequence of the complexity of the physics and mathematics involved in each of the areas was that the panel discussions were focused on the general mathematical challenges in each of the above areas, with little time being spent discussing how to effectively deal with these challenges on extreme-scale computing platforms.

Burning Plasma/ITER

The main challenge posed for the simulations of burning plasma/ITER was in the area of high-fidelity calculations both in the core and edge. The most difficult scientific and computational challenge was performing full kinetic calculations in the core and in the edge of the tokamak. The following were identified as the main research directions needed to be undertaken in this context:

- more accurate gyrokinetic approximations
- field-aligned coordinates
- fundamental numerical algorithms for PIC
- time integrators
- kinetic treatment of electrons.

Gyrokinetic Approximations

The starting point for a kinetic description of plasmas is generally the Vlasov equation or the Fokker-Planck equation (in the presence of collisions). For a tokamak with a strong guide field (e.g., the ITER), the fastest time scales are associated with the electron gyro-frequency and the ion gyro-frequency.

Gyrokinetics is essentially the formalism that averages over the gyro-radius of the charged particles with a resulting equation for the distribution function that is defined in a five-dimensional phase-space. A number of underlying assumptions are made in the derivation of gyrokinetics: small frequency compared with the gyro-frequency, and large spatial scales compared with the gyro-radius, etc. In the tokamak core, further assumptions underlie existing gyrokinetic codes: smallness of the drift speed (i.e., ExB speed) compared with the thermal speed and time-independent electrostatic potential. These assumptions break down at the plasma edge. Furthermore, at the plasma edge, gyrokinetic simulations must also contend with complex geometry, open magnetic field lines impinging on divertors, large magnetic shear near the separatrix, and interaction with material walls that increases the impact of neutral and impurities (Cohen and Xu 2008). A more accurate generalization of gyrokinetic models that includes the strong spatial variation and finer meshes is desirable. The following question naturally arises: given the complexity of the gyrokinetic models, does it make more sense to resort back to the six-dimensional kinetic models that may be more appropriate for extreme-scale-computing platforms and do not suffer from the limitations of the gyrokinetic models? The answer to this question is not easy, and further research is required in this area.

Field-Aligned Coordinates

The physical scales generally exhibit a clear separation because of the presence of a strong guide field in tokamaks. Thus, scales along the magnetic field lines tend to be much larger that those perpendicular to it. Existing gyrokinetic codes (gyrokinetic toroidal code [GTC], and a PIC gyrokinetic edge code XGC [XGC]) often make use of this physical separation of scales and align the mesh along the field lines. This alignment of one coordinate direction with the field results in more efficient computations because of the reduction in resolution requirements along one direction. This idea of field-aligned coordinates works especially well for relatively simpler poloidal cross-sections of a tokamak. For example, GTC employs a circular poloidal cross-section and twists the mesh points on radial locations to achieve near field alignment. For more complicated cross-sections, such as those encountered in the ITER with a separatrix and a divertor, field alignment is more complicated and generally treated on a case-by-case basis. An important research direction identified is that of a robust field-aligned coordinate system that can deal with separatrices, branch cuts, and differential shearing. An additional complication that must be addressed is that of electromagnetic simulations—the magnetic field is evolving, which requires a moving field-aligned mesh. Furthermore, topology changes must be considered when secondary magnetic islands appear and when the field becomes stochastic in certain regions. A thorough analysis is necessary with respect to the trade-off between computational cost of field alignment and accuracy loss when misalignment occurs.

Fundamental Numerical Algorithms for PIC

An often-used formalism in the area of PIC simulations is that of delta-*f*; i.e., the distribution function is split into its equilibrium Maxwellian component and a perturbation (noting that no linearization takes place). However, the physics community is concerned about the accuracy and validity of the delta-*f* formulation under certain conditions such as those prevalent in the plasma edge region of a tokamak. An alternative formulation is to use what is referred to as the full-*f* method. A particular difficulty associated with PIC simulations has been a lack of rigorous mathematical analysis of error, consistency, and convergence. An illustrative example is the work of Vay et al. (2004), which identified spurious forcing that appears because of abrupt changes in the grid size in adaptive mesh PIC simulations. Such spurious forces also occur because of the lack of cancellation of error when the mesh spacing is variable or non-uniform. Clearly, more research is required in this area. Convergence tests of PIC methods typically focus on increasing the number of particles per cell while keeping the mesh size constant. Generally, mesh sizes in gyrokinetic simulations tend to be of the same order as the ion gyro-radius because of the choice, for example, of using four-point averaging techniques (Lee 1987). This blending of the numerical method with physics intuition leads to insufficient progress and lack of rigor in mathematical analysis of gyrokinetic PIC algorithms. This needs to be rectified.

Time Integrators

The panel identified a research opportunity in comparisons between symplectic integrators versus higher-order time integration methods traditionally employed in kinetic simulations (Qin et al. 2009). While symplectic integrators soundly preserve certain conserved quantities, it is not yet clear if these methods are absolutely essential in kinetic simulations. Another important consideration is that the discretization for the partial differential equations (PDEs) preserves the Hamiltonian structure. This is akin to a generalization of Arakawa's method, which preserves the discrete Poisson bracket.

Stiff Electron

In kinetic simulations that go beyond treating the electrons' response as simply adiabatic, one needs to take into account the more than one order of magnitude increase in the spatial and temporal stiffness results from the electrons—typically, the increase in stiffness scales as the square root of the ion to the electron masses (~40). This increased factor, which is encountered in realistic mass ratio electromagnetic kinetic simulations, will pose a serious threat to how far kinetic simulations will progress in terms of physical times even on extreme scale (100s to 1000s of petaflops) computers. Algorithmic research in this area is required to overcome the stiffness induced by the electrons.

Integrated Modeling

A thorough examination of integrated modeling required for magnetic fusion was conducted by Kritz and Keyes (2007). As multiphysics simulations become more prevalent, a systematic approach to coupling codes is required. Examples of integrated modeling include the following:

- coupling between gyrokinetic edge turbulence codes with nonlinear MHD codes and atomic physics codes (Park et al. 2007)
- coupling between edge and core gyrokinetic codes (Cary et al. 2007)
- coupling between flux-tube gyrokinetic codes with transport codes
- coupling between nonlinear MHD and radio-frequency codes (Batchelor et al. 2007)
- coupling between nonlinear MHD codes and kinetic codes in the context of energetic particles physics investigations (Fu et al. 2006).

These are but a few existing examples of integrated models. Broadly, one may distinguish between three types of coupling approaches within the two sub-systems that are coupled: explicit, in which two codes perform their own time step and exchange data at the end of each time step, lagging the information supplied from one subsystem to another; semi-implicit, in which some measure of implicitness is introduced; and fully implicitly coupled systems, in which the entire system is treated implicitly (for instance, using a Newton-Krylov approach). Furthermore, distinction can be made between the overlap of the domain shared by the two codes or subsystems. For example, in energetic particles physics, the entire physical domain is under consideration by the two codes while in the core-edge coupling scenario, each code operates on a different portion of the physical domain. Development of a systematic coupling framework is desirable.

There exists anecdotal evidence of subsystems that are perfectly stable and accurately computable on their own, but when coupled in an explicit fashion may lead to numerical instability. The panel identified the following as the principal research goal in the coupling of codes for integrated modeling: *a mathematical infrastructure for understanding well-posedness, stability, and accuracy for coupling (both pre- and post-discretization).*

As an example, we considered coupled treatment of turbulence and transport. In this example, there is integration of well-parallelized flux-tube turbulence codes with transport computations and a profile advance code that would include core sources. The coupled turbulence-transport simulations appear to have significant promise in that they will effectively use resources at the extreme scale. The exchange between the codes includes a few parameters describing the local plasma state and the turbulent fluxes. One issue is that of boundary conditions in the flux-tube turbulence simulations, which presently use periodic boundary conditions. It is important to investigate other types of boundary conditions and to analyze accuracy, consistency, and convergence in these areas.

Plasma-Material Interaction

The most pressing issue in the area of plasma-material interaction appears to be the design of materials to withstand tokamak operating conditions—a topic that is out of the scope of numerical plasma physics. Hence, this subject will not be addressed in this panel report. A second scientific and computational challenge is in the area of interaction of the plasma environment with material boundaries. This area includes the following research directions:

- improved fidelity of edge models with respect to interaction with boundaries
- effect of impurities
- effect of having liquid walls (e.g., having lithium flowing along the walls of the tokamak).

Improved Fidelity of Edge Models

It is well known that plasma-material interaction is a complicated multiscale, multiphysics problem. The general treatment of atomic physics processes is with Monte-Carlo techniques (see the DEGAS 2 code).[1] These codes usually interact with other turbulent codes operating at the plasma edge. Analysis of such coupling is an important research issue. Another important issue, given the nondeterministic nature of the interaction, is the quantification of uncertainty in the results and providing error bars for predictive science.

Effect of Impurities

As the plasma interacts with the material surface, sputtering and other physical phenomena can cause contamination of the plasma with impurity elements (both in the form of neutrals and ions). Again, the multiscale nature of the interaction, along with the intrinsic nondeterminism in the results, poses a severe challenge.

[1] See Degas 2 at http://w3.pppl.gov/degas2/.

Effect of Liquid Wall

Recent developments have been made in having a liquid (lithium) flowing along the walls of the tokamak. Simulating tokamak liquid walls is still an open research area. Further challenges include simulating the coupling with the plasma edge and dealing with the onslaught of energetic neutrons.

Laser-Plasma Interaction

The ability to predict the interaction of intense lasers with plasmas is critical for laser-driven fusion. When a laser interacts with plasma, heterogeneities known as speckles arise within the plasma. Filamentation, an instability that occurs during laser-plasma interactions, causes an initially smooth beam to degrade into string-like structures called filaments. Parametric instabilities occur from the resonant coupling of the light wave with ion and electron waves. To adequately model filamentation and parametric instabilities requires resolution at the level of the wavelength of light. Understanding these phenomenas are the central scientific and computational challenge in laser-plasma interaction. The main research areas identified are the following:

- homogenization of speckles
- electromagnetic (EM) PIC
- adaptive mesh refinement (AMR) for PIC.

Homogenization of Speckles

Homogenization is the main mathematical technique proposed to deal with the effect of thousands of speckles. This somewhat standard technique needs to be applied to get the effect of many speckles in the spirit of “sub-grid-scale” modeling for laser-plasma interaction. With current state-of-the-art computing technology, only one such speckle can be computed. Extreme-scale computing can be used in the spirit of direct numerical simulations to compute the effect of many speckles and drive the development of reduced and mesoscale models.

Electromagnetic PIC

EM PIC simulations apparently suffer from unphysical features at the grid scale. It is recommended that higher-order methods or implicit methods for Maxwell’s equations be explored as possible solutions to overcome the problem of unphysical features. It is also important not to overlook boundary conditions when computing with higher-order or implicit methods. Better representations of collision effects are also required for further study.

Adaptive Mesh Refinement for EM PIC

AMR was identified as a means to overcome some of the spatial stiffness in EM PIC simulations (e.g., see Fujimoto and Sydora 2008). In AMR, the mesh is refined to have a uniform error in the domain, or refined based on heuristics such as large gradient regions where one expects to have a large error. One must be aware of the spurious forcing because of changes in mesh sizes and use techniques that cancel these spurious forcing terms.

Magnetic Reconnection

Magnetic reconnection is somewhat ubiquitous in many physical systems described by plasma physics. It is seen in a diverse range of phenomena including coronal mass ejections, substorms in the earth's magnetotail, sawteeth and disruptions in tokamaks, and a variety of astrophysical scenarios (e.g., Yamada 2007). The main scientific and computational challenge stems from the large range of spatial and temporal scales, as well as the fact that some portions of the domain are amenable to a fluid description while other regions necessarily require a kinetic description. There is general agreement that for high-collisional regimes, a fluid MHD description is sufficient in which the smallest scale—i.e., the thickness of the reconnection layer where diffusion effects dominate—depends upon the Lundquist number (a nondimensional ratio of Alfvén time to diffusion time). When the thickness of the resistive layer is smaller than the ion kinetic scale, a kinetic description is more appropriate. Brute force computations wherein the smallest scale governed by the electron kinetic scale is well resolved seem beyond the scope of even the largest computational facilities for the foreseeable future, although progress has been made in two-dimensional simulations modeling of real experiments (Dorfman et al. 2008). The following are the main research directions in magnetic reconnection:

- kinetic and/or hybrid models
- AMR/adaptive mesh and algorithm refinement (AMAR)
- curse of dimensionality
- treatment of stiff electrons.

Kinetic and/or Hybrid Models

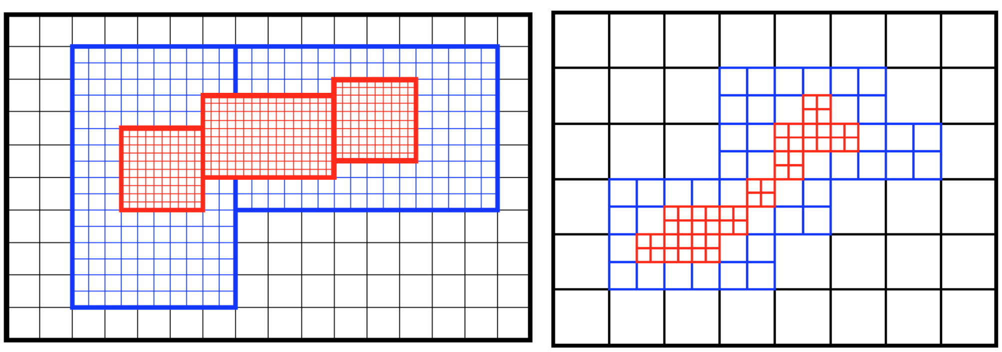
In magnetic reconnection, one can generally distinguish between an "outer" layer and an "inner" layer. It is within the inner layer that the change in magnetic topology takes place. When the thickness of the inner layer is comparable to the gyro-radius of the charged particles, it becomes imperative to use a kinetic model for the inner layer. The outer macroscopic region may still be adequately described by fluid models. Coupling the kinetic models with the fluid models is an extremely important research direction in magnetic reconnection. Applied mathematics can play the role of determining the type of interfacial, flux, and compatibility conditions required by the PDEs governing both the fluid and kinetic models. Mathematicians are also equipped to make important determinations with respect to the consistency and accuracy of coupling approaches in numerical implementations of such hybrid models. Hybrid models also require a decision as to determining the location of the interface between the two disparate regions. In some cases, instead of a sharp interface, an overlap region may be required to ensure smooth transition from one model to the other. Further study in these areas is recommended.

AMR/AMAR

AMR methods are a clear choice to decrease computational costs associated with employing a uniform mesh over the entire domain. AMR is akin to a numerical microscope in that finer resolution is used to resolve fine features. In magnetic reconnection, intuition dictates using a finer resolution in the inner region where the magnetic topology takes place. However, research is required to ensure this is appropriate because there are instances when physical intuition can be misleading and errors propagating from coarser mesh resolution portions of the domain can contaminate the solution accuracy in the finer mesh resolution regions. In addition to AMR, AMAR is identified as an important applied math research direction for magnetic reconnection. In AMAR, not only is the mesh refined, but also the algorithm employed is refined or adapted in response to the changes in the model describing the physical system. As mentioned earlier in the context of hybrid models for reconnection, it is appropriate to change the algorithm (or numerical method) as the reconnection layer is refined. Advances in numerical analysis will be required to assess the stability of the free boundary value problem arising in this context, which is akin to the analysis for stability in nested hierarchical mesh refinement for hyperbolic conservation laws (Berger 1985).

Adaptive Mesh Refinement

In many problems in partial differential equations, one is confronted with problems having multiple length scales and strong spatial localizations. Examples include nonlinear systems of hyperbolic partial differential equations containing complex combinations of discontinuities and smooth flow. Also included are combustion problems in which—at any given instant—burning is taking place in a small subset of the problem domain and problems with complex geometries in which localized geometric features can generate strong, localized solution gradients. Finite difference calculation using block-structured adaptive mesh refinement (AMR) is a powerful tool for computing solutions to partial differential equations involving such multiple scales. In this approach, the underlying problem domain is discretized using a rectangular grid and a solution is computed on that grid. Computing some local measure of the original error identifies regions requiring additional resolution, and the individual cells refined (as in-cell-based AMR) or covered by a disjoint union of rectangles in the domain, which are then refined by some integer factor as in block-structured AMR (see Sidebar Figure 1). The solution is then computed on the composite grid. This process may be applied recursively, and for time-dependent problems, the error estimation and regridding can be integrated with the time evolution and refinement applied in time as well as in space. Such an approach was first introduced by Berger and Oliger (1984) for computing time-dependent solutions to hyperbolic partial differential equations in multiple space dimensions. Since then, the approach has been extended to a variety of problems in applied partial differential equations.



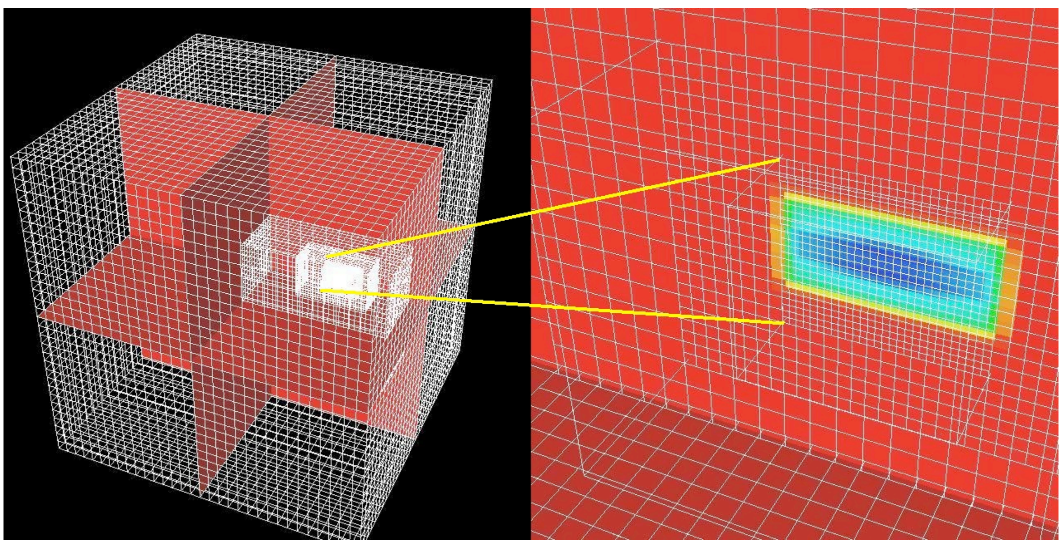
Sidebar Figure 1. Schematic depicting commonly used "h-refinement" block-structured AMR (left) in which there exists a hierarchy of nested meshes, and cell-based AMR (right) in which all computational zones are at one level. Images courtesy of Phil Colella (Lawrence Berkeley National Laboratory) and Ravi Samtaney (Princeton Plasma Physics Laboratory).

Block-structured AMR has been used in fusion applications such as pellet injection and magnetic reconnection. In pellet injection (a viable method to refuel tokamaks), the physical processes span several decades of spatio-temporal scales, which has prevented effective simulations of these processes. Naive estimates indicate that the number of space-time points required for simulating pellet injection in the ITER can exceed 10^{19}. AMR is a viable technology to handle the large disparity between pellet size and device size ($\sim 10^3$). Sidebar Figure 2 shows the mesh structure in computational coordinates wherein the pellet is buried within the finest mesh that occupies less than 0.015% of the volume of the coarsest mesh—a visual illustration of the resolving power afforded by the AMR technology.

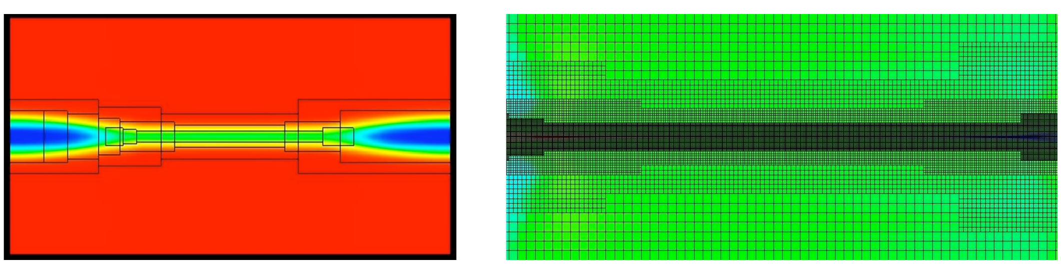
Adaptive Mesh Refinement (cont.)

Magnetic reconnection is a fundamental plasma process of rapid rearrangement of magnetic field topology, accompanied by a violent release of magnetically stored energy and its conversion into heat and into non-thermal particle energy. Understanding magnetic reconnection is of importance in physical phenomena such as coronal mass ejections, magnetic storms in the earth's magnetosphere and sawtooth crashes in tokamaks. Generally, a thin resistive layer wherein the topology change of the magnetic field takes place and an outer "ideal" region characterizes single-fluid resistive magnetic reconnection.

The inner layer is essentially a current layer whose thickness scales as inverse square root of the Lundquist number (a ratio of Alfvén wave crossing time scale over diffusion time scales and is typically very large in most physical systems of interest). AMR provides a viable method to compute reconnection effectively by providing resolution of the thin current layer. Sidebar Figure 3 shows an example simulation of magnetic reconnection with AMR.



Sidebar Figure 2. Density field in pellet injection in computational space with the mesh superimposed. The high-density region; i.e., the pellet ablation cloud, shown on the right resides in the finest mesh. The corresponding uniform mesh simulation would require over a billion mesh points, and would be approximately 200 times slower than the AMR simulation. (Image source: Applied Partial Differential Equations Center for Enabling Technologies SciDAC Center).



Sidebar Figure 3. AMR simulation of magnetic reconnection at S=105. The pressure field is shown on the left with outlines of the mesh blocks; and the x-direction velocity on the right. The mesh adaptation to resolve the thin current layer is apparent. Images courtesy of Phil Colella (Lawrence Berkeley National Laboratory) and Ravi Samtaney (Princeton Plasma Physics Laboratory).

Dimensionality

While a number of magnetic reconnection studies have been in two dimensions, it is of considerable interest in the physics community to investigate the reconnection and magnetic island dynamics in three dimensions. This adds an enormous computational burden, especially if the details of the inner layers have to be well resolved. It is apparent that AMR/AMAR methods will be indispensable for three-dimensional investigations of magnetic reconnection.

Treatment of Stiff Electrons

In kinetic simulations the stiffness increases by more than one order of magnitude due to the electrons (the increase in stiffness scales as the square root of the ion to the electron masses ~40). The outstanding issues in the treatment of stiff electrons are similar to those presented in the panel report titled, "Burning Plasma/ITER."

Other Vital Areas

Dwarfs

A dwarf is defined as an algorithmic method that captures a pattern of computation and communication. Colella (2004)[1] has identified algorithms, which are essentially the building blocks of high-end simulations in the physical sciences:

- structured grids (including locally structured grids; e.g., AMR)
- unstructured grids
- Fast Fourier Transform
- dense linear algebra
- sparse linear algebra
- particles
- Monte Carlo.

Each algorithm has its own distinctive combination of computational operations, data representation, and data choreography. It is recommended that a "dwarf identification" exercise be performed on each application to help guide its further development on extreme-scale architectures.

Implicit Methods

The wide spectrum of temporal scales and the stiffness of the systems arising in plasma applications also pose a question of identifying optimal time integrators. Development of state-of-art time-integration methods and efficient adaptive time-stepping techniques for extreme-scale computers can provide significant computational savings and allow simulations in previously inaccessible parameter regimes. Studies are needed that compare performance of explicit, implicit, and exponential methods of low and high orders, identify optimal methods for specific problems, and adapt them for extreme-scale computers. Fully nonlinearly implicit methods (e.g., the ones using Jacobian-Free Newton-Krylov approaches) in the

[1] Colella P. 2004. "Defining Software Requirements for Scientific Computing." Defense Advanced Research Projects Agency High Productivity Computing Systems Workshop, November 8, 2004, Pittsburgh, Pennsyvania.

area of MHD have been recently developed (e.g., Reynolds et al. 2010). As mentioned earlier, designing operator-based or physics-based pre-conditioners will be a challenge because the use of black-box pre-conditioners is limited in the context of extreme-scale computing.

Unit Testing

It is a responsibility of the mathematical formulation efforts to help establish the "unit tests" for simulation and algorithm work. In particular, we believe we can and should develop better (physically rich) model problems that can be used as benchmarks throughout the code development process. This is not a minor task.

Uncertainty Quantification

There are two basic approaches to integrated modeling (Bécoulet 2009): 1) building up from first-principle equations taking advantage of massive computing, and 2) building up from experimentally based models. By definition, first-principles will have no free parameters and will be oriented towards scientific discovery through high-performance computing. However, in the context of a real tokamak, there are many operational constraints and uncertainties. Furthermore, some physical phenomena (e.g., plasma-material interaction, sputtering, etc.) have an inherent stochastic component to them. First-principles-integrated tokamak simulations (such as direct numerical simulations) seem to be unreachable for any conceivable computer. Integrated tokamak simulations will necessarily use "knobs" in the simulation codes. Some of these knobs will be well calibrated against experiments or first-principle computations, while others will rely on physical intuition and experience. This poses a problem because integrated simulations cannot strictly be seen as doing predictive science. There has been a recent surge of interest in uncertainty quantification and predictive science. Uncertainty quantification is the characterization and reduction of uncertainty in applications. It can be aleatory (i.e., because of uncertainty of model and input parameters) or epistemic (because of uncontrollable processes) (see "Introduction to Uncertainty Quantification").[1] Future extreme-scale computing will have a real cost associated with it, and it will be useful to perform uncertainty quantification on computational results so that we understand their range of validity. It will be the task of applied mathematicians to assist scientists, perhaps by deriving stochastic models to take the place of current deterministic models, to help the uncertainty quantification process.

CONCLUSIONS

The mathematical formulations panel was charged with determining the challenging areas of applied mathematics research in each of the five physics areas for computing at extreme scales in fusion. The dominant themes centered around kinetic (and gyrokinetic) simulations, as well as integrated and hybrid simulations. For each of the five physics areas, the applied mathematics research areas are summarized below:

- Burning plasma/ITER: High-fidelity kinetic simulations, more robust gyrokinetic approximations, new algorithms for PIC, error analysis of gyrokinetic PIC methods, and treatment of kinetic electrons.

[1] See "Introduction to Uncertainty Quantification – Mini Tutorial." SIAM Conference on Computational Science and Engineering, March 2-6, 2009, Miami, Florida. Last accessed April 1, 2010, at http://www.stanford.edu/~jops/UQsiam09.html.

- Integrated modeling: Systematic frameworks for coupling, analysis of well-posedness, convergence and accuracy properties of coupled systems, and fully implicit methods.
- Plasma-material interaction: Coupling of edge plasma physics code with atomic processes codes.
- Laser-plasma interaction: Homogenization over multiple speckles and development of mesoscale and reduced models to deal with filamentation and parametric instabilities.
- Magnetic reconnection: Development of hybrid (kinetic-fluid) algorithms, treatment of stiff electrons, and adaptive mesh and algorithmic refinement.

Several underlying mathematical techniques were identified, including adaptive mesh refinement, implicit methods, and robust coupling of codes. These techniques crosscut several of the physics areas, and they merit further research and resources.

PROGRAMMING MODELS, FRAMEWORKS, AND TOOLS

Lead: Ewing Lusk, Argonne National Laboratory

Panel Members: Katherine Yelick, University of California, Berkeley and Lawrence Berkeley National Laboratory; Allen Malony, University of Oregon; Robert Harrison, Oak Ridge National Laboratory; V. Balaji, Princeton University and National Oceanic and Atmospheric Administration; Eduardo D'Azevedo, Oak Ridge National Laboratory; John Shalf, Lawrence Berkeley National Laboratory; Jack Dongarra, University of Tennessee and Oak Ridge National Laboratory

SUMMARY

The coming transition in computer architectures, as peak capability approaches the exascale, offers both challenges and opportunities. The challenges involve a paradigm shift in programming methodologies. Existing technologies for writing parallel scientific applications have sustained high-performance computing (HPC) application software development for the past decade and have been successful for petascale computing, but were architected for coarse-grained concurrency largely dominated by bulk-synchronous algorithms. Future hardware constraints and growth in explicit on-chip parallelism will likely require a mass migration to new algorithms and software architectures that is as broad and disruptive as the migration from vector to parallel computing systems that occurred 15 years ago. The software and algorithms will need to rely increasingly on fine-grained parallelism and strong scaling; they must also support fault resilience. Addressing these challenges renews the opportunity to introduce a higher level of software engineering into current fusion application subsystems. This software engineering will enhance the modularity, portability, and performance of codes and will extend their capabilities to new levels. At the same time, past sound investments must be protected, and a migration path from current to future environments must be defined.

To confront these issues directly, the panel on programming models, frameworks, and tools at the Workshop on Modeling and Simulation at the Exascale for Energy and the Environment identified six priority research directions (PRDs). The first is **to find and implement efficient algorithms that exploit new multicore, heterogeneous, massively parallel architectures**. This research thrust is directed primarily at the languages, libraries, and run time systems that allow programmers to use massive on-chip concurrency in a portable/cross-architecture manner while cooperating with interprocessor parallelism. The second PRD is **to find new, productive approaches to write, integrate, validate, and tune complex application programs**. Here, the development of programming models and systems for massive numbers of processors is addressed. The third PRD is **to develop tools for understanding complex application program behavior at scale and for optimizing application performance**. This will require the evolution of existing tools and the development of new ones to deal with heterogeneous processors and greater integration of model-based approaches. The fourth PRD is **to promote a migration path from current programming approaches to new ones.** Existing Fortran + message passing interface (MPI) codes will continue to be used and extended as architectures scale. Research into MPI interoperability and extreme scalability will be required, and a new software ecosystem that spans all scales of systems—from midrange to the exascale—must be developed to facilitate a viable migration path from development to large-scale production computing systems. The fifth PRD is **to define common framework tools or components that can be reused in multiple application domains**. Frameworks that organize existing and future fusion codes into coherent tools for scientific investigations

are currently in an ad hoc stage of development; research into general abstractions and tools for constructing components and frameworks is needed. The sixth PRD is **to establish methods and systems that enable pervasive fault resilience**. At the exascale, faults of various kinds in both hardware and software components are expected to become commonplace in the execution environment. Fault-recovery mechanisms must be integrated at every level of the system design: hardware, software, and even in the programming model.

SCIENTIFIC CHALLENGES AND RESEARCH DIRECTIONS

Find Efficient Algorithms and Implementations that Exploit New Multicore, Heterogeneous, Massively Parallel Architectures

When processor clock speeds flatlined in 2004, after more than 15 years of exponential increases, the era of near-automatic performance improvements enjoyed by the HPC application community came to an abrupt end. The air of crisis that followed in the wake of these new hardware trends continues to hang over computational science. To develop software that will perform well on petascale systems with thousands of nodes and millions of cores, the list of major challenges that must be confronted is formidable: 1) dramatic escalation in the costs of communication between processors and/or levels of memory hierarchy; 2) increased hybridization of processor architectures (mixing CPUs, GPUs, etc.) in varying and unexpected design combinations; 3) cooperating processes, to address high levels of parallelism and more complex constraints, that must be dynamically and unpredictably scheduled for asynchronous execution; 4) software that does not run at scale without more robustness and better resilience to faults; and 5) new levels of self-adaptivity that is required to enable software to satisfy limited energy budgets and transient load-imbalances caused by algorithms or fault-resilience strategies.

Researchers' current stable of algorithms were never designed with these emerging hardware constraints in mind. Therefore, much of their software infrastructure and numerical libraries must be redeveloped to enable research through the next decade of system scaling. The following key areas could facilitate development of new algorithms that will meet the challenges outlined above.

Communication Optimal Algorithms

Algorithmic complexity is usually expressed in terms of the number of operations performed rather than the quantity of data movement to memory. This is antithetical to the true costs of computation where memory movement is very expensive and operations are nearly free. To address the critical issue of communication costs, there is a need to investigate algorithms that reduce communication to a minimum. First, the well-known bandwidth lower bounds for dense matrix-multiplication must be extended to get bandwidth and latency lower bounds onto parallel and sequential machines' basic matrix operations for dense and sparse matrices. Researchers hope to discover new algorithms that attain these lower bounds in many cases. Second, for Krylov subspace methods like generalized minimal residual (GMRES), conjugate gradient (CG), and Lanczos, investigation should focus on taking k steps of these methods for the same communication costs as a single step.

Multi-targeted Autotuning for GPU and Hybrid Architectures

The objective is to provide a consistent library interface that remains the same for users independent of scale and processor heterogeneity, and achieves good performance and efficiency by binding to different underlying code, depending on the configuration. The diversity and rapid evolution of these platforms mean that autotuning of library functions such as basic linear algebra subprograms (BLAS) and Fast Fourier Transform (FFT) will be indispensable to achieving good performance, energy efficiency, load balancing, etc. across this range of systems. In addition, the community will need frameworks that go beyond library limitations and that are able to optimize data layout (Berkeley Lab Checkpoint/Restart [BLCR] blocking strategies for sparse matrix vector [SpMV] multiplication kernels), stencil autotuners (because stencils kernels are diverse and not amenable to library calls), and even tuning of optimization strategy for multigrid solvers (optimizing the transition between the multigrid coarsening cycle and bottom-solver to minimize run time). As early and successful pioneers in the area of autotuning, the DOE Office of Advanced Scientific Computing Research computer science research community can draw on a large base of experience available to achieve this result.

Scheduling and Memory Management for Heterogeneity and Scale

Extracting the desired performance from environments that offer massive parallelism—especially where additional constraints (e.g., limits on memory bandwidth and energy) are in play—requires more sophisticated scheduling and memory management techniques than have heretofore been applied to linear algebra libraries. Another form of heterogeneity comes from confronting the limits of domain decomposition in the face of massive explicit parallelism. Feed-forward pipeline parallelism can be used to extract additional parallelism without forcing additional domain decomposition but exposes the user to dataflow hazards. Ideas relating to a data flow-like model—expressing parallelism explicitly in directed acyclic graphs (DAGs), so that scheduling tasks dynamically—support massive parallelism and apply common optimization techniques to increase throughput. Isolating side effects include explicit approaches that annotate the input arguments to explicitly identify their scope of reference (e.g., IVY), or implicit methods such as using language semantics or strongly typed elements to render code easier to analyze for side effects by compiler technology (e.g., single-assignment arrays in Ct or annotated subroutines in Cilk). New primitives that enable diverse memory management systems to be managed efficiently and in coordination with the execution schedule are needed.

Adaptive Response to Load Imbalance

Adaptive multiscale algorithms are an important part of the DOE portfolio because they apply computational power precisely where it is needed. However, they introduce challenging computational requirements because they introduce dynamically changing computations that result in load imbalances. As researchers move toward systems with billions of processors, even naturally load-balanced algorithms on homogeneous hardware will present many of the same daunting problems with adaptive load balancing that are observed in today's adaptive codes. For example, software-based recovery mechanisms for fault-tolerance or energy-management features will create substantial load imbalances as tasks are delayed by rollback to a previous state or correction of detected errors. DAG-based scheduling also requires new approaches to optimize for resource use without compromising spatial locality. These challenges require development and deployment of sophisticated software approaches to dynamically rebalance computation in response to changing workloads and conditions of the operating environment.

Fault Tolerance and Robustness for Large-Scale Systems

As more researchers target emerging large-scale systems and as the number of cores per system escalates into the millions, issues of fault tolerance and robustness will inevitably come to the fore. Fault tolerance requires computationally efficient building upon previous successes in algorithm-based fault tolerance. Fault-tolerant, DAG-based algorithms offer a compelling approach to both dense and sparse linear algebra designed to run at scale. Exploration of mixed-precision algorithms, sparse hybrid (direct and iterative) solvers, and memory-aware algorithms to increase both performance and robustness at scale are also desired.

Building Energy Efficiency into Software Foundations

It is widely recognized that emerging constraints on energy consumption will have pervasive effects on HPC; power and energy consumption must now be added to the traditional goals of algorithm design—that is, correctness and performance. The emerging metric of merit becomes performance per watt. Consequently, researchers believe it is essential to build control, efficiency, and awareness of power and energy into the foundations of their numerical libraries. This will require the development of standardized interfaces and application programming interfaces (APIs) for collecting energy consumption data, just as the Performance Application Programming Interface (PAPI) has done for hardware performance counter data. Accurate and fine-grained measurement of power consumption underpins all tools that seek to improve such metrics. (Anything that cannot be measured cannot be improved.) Researchers must use these standardized interfaces and APIs to better understand the effects of energy-saving hardware features on the performance of linear algebra codes. Additionally, researchers must identify parameters and alternative execution strategies for each numerical library that can be tuned for energy efficient executions. Finally, researchers must enhance schedulers for better low-energy execution.

Find New, Productive Approaches to Write, Integrate, Validate, and Tune Complex Application Programs

A programming model[1] effort is a critical component of a program to build effective exascale computing systems, because with clock speeds projected to be flat or even dropping to save energy, all performance improvements within a chip will come from increased parallelism. While doubling traditional microprocessor cores is likely to be the industry response to this challenge for some markets, this evolutionary approach is unlikely to produce an exascale platform in the next 10 years. Instead, the architectures will either need lightweight cores with slower clocks and/or data parallel hardware (single instruction, multiple data [SIMD], quad-hummer, vectors, or accelerators). In any case, the amount of memory per arithmetic functional unit will drop significantly, implying the need for fine-grained parallelism and a programming model other than message passing or coarse-grained threads (e.g., [PThreads] or [OpenMP]). It is premature to rule out any of the architectural models for increasing on-chip parallelism, yet history suggests that a programming model specialized to a single architecture will fail. Even if architectures become somewhat specialized to a class of applications, the programming

[1] "Programming model" is used in a very general sense as a set of languages, libraries, and tools by which application programs are developed.

model must be portable across all viable architectures. Thus, an exascale software program must allow architectures to pursue multiple hardware solutions while programming models support a range of possible solutions.

The timing of the programming model effort is important: if the machines arrive with no viable programming model, it will significantly delay the science impact or result in limited domains of impact. If the programming model is developed prior to the machine design and without regard to their features, it may not be suitable to the hardware. Thus, the programming model effort must be tightly coupled with multiple architecture efforts. Programming model developers must have intimate knowledge of the proposed hardware designs and must be able to influence those designs. At the same time, the programming model development must be responsive to the needs of applications and must support the kinds of algorithms that will be used on exascale machines.

There are many challenges for programming model designers, including support for multiphysics applications, support for both fine- and coarse-grained parallelism, interoperability, portability, scalability, latency tolerance, support for energy-efficient programming, performance feedback requirements, and reduction of concurrency errors such as race conditions. Some of these challenges are specific to scientific programming or large-scale parallelism, and some will cross over into more general programming areas. However, a program with too many goals and lack of prioritization is unlikely to succeed. Therefore, the DOE exascale effort should focus on the two most critical problems: support for fined-grained parallelism within a chip (including data parallel hardware, heterogeneous processors, and locality control) and support for fault tolerance programming between chips (i.e., allowing programmers to write applications that tolerate hardware failures at the level of a single chip). The parallelism and scaling problems between chips are important, but the message-passing model as realized by the MPI interface and its implementations provide a viable solution that, with some investments, can be made to scale, whereas a viable programming model for massive on-chip parallelism does not exist. Similarly, there are only preliminary research efforts on fault tolerance programming, while current checkpoint approaches are likely to be inadequate. This is because the frequency of component failures in an exascale system may be close to the time to checkpoint an application to disk, in which case no forward progress is possible. Techniques to build redundancy into algorithms and software will become increasingly important and must be supported in the programming model.

Figure 24 shows a diagram of the multiple application and architecture efforts, with a common programming model effort. Some application teams may work more closely with one or more teams, but the programming model must be responsive to the needs of all applications and hardware efforts. The programming model effort should include representatives from vendor teams, who may not have access to other teams' hardware requirements but will see their effect on the programming model design process. Application and algorithm experts should be represented as well, if not on a day-to-day basis then at regular intervals to influence the programming model team. While the programming model team should respond to requirements above and below, they may make conscious decisions against fully supporting a particular class of algorithms or hardware; i.e., some things may not run optimally if it would affect the viability of a coherent model for the more common applications.

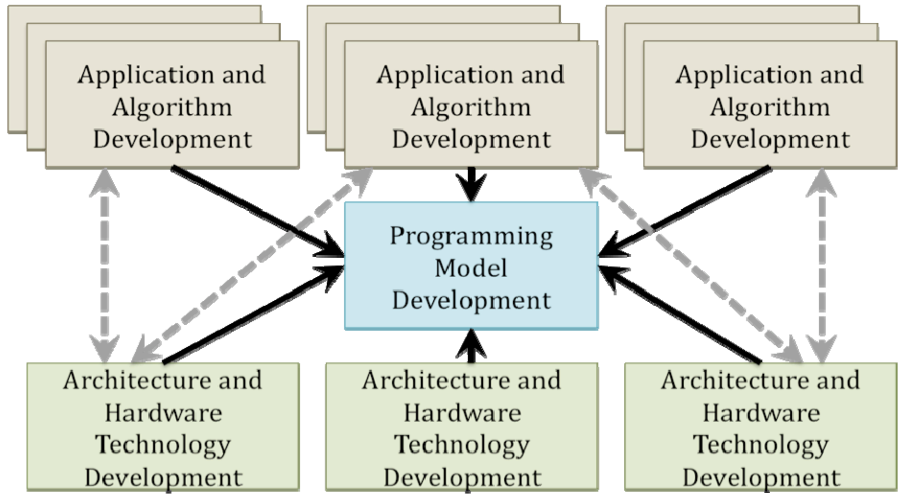


Figure 24. Structure of programming model activity relative to hardware, application, and algorithms developments. The programming model must be responsive to hardware, application, and algorithm needs. Image courtesy of Katherine Yelick (Lawrence Berkeley National Laboratory, University of California at Berkeley).

Develop Tools for Understanding Complex Application Program Behavior at Scale and for Optimizing Application Performance

The promise of extreme-scale computing systems for advancing fusion energy sciences lies in the delivery of scaled fusion simulations that optimize the performance potential of these machines. The processes of parallel program understanding and performance engineering are at the heart of achieving optimization goals. However, these processes and the tools applied to support them will need to change, along with the programming models and frameworks in response to architectural and system evolution. Traditional techniques for parallel debugging, performance diagnosis, and tuning will become intractable as the factors of scale, software complexity, application sophistication, and hardware integration continue to increase. While interactions among these factors create the need to observe application performance across the whole system hierarchy, it is the model-oriented knowledge of the computational semantics of the application and of performance expectations (with respect to petascale/exascale systems architectures and capabilities) that ultimately must be incorporated into tools to better focus and automate correctness/performance problem-identification and to guide tuning decisions.

Observation of Scalable, Complex Execution

To understand application program behavior and performance, researchers must observe parallel application execution in the target system environment. In general, the more that can be learned from the execution via measurements, the higher value of information a tool will have for assessment of behavior and problems. However, scale complicates observation and analysis by amplifying the amount and complexity of measured data, as well as the effects of measurement on the execution. These concerns can be addressed in part by enhancing current petascale tools with more scalable infrastructure and analysis methods, but extreme-scale computing will ultimately force a more intelligent methodology to optimize observation value versus measurement cost. It is important that such an approach must be rooted in knowledge about the application—its structure, computational model/domain, algorithms, etc.—so that what is observed can be related back at a higher level to application-specific concerns.

The problem is more than just a matter of scale. Application behavior, and in particular, performance will be determined by a complex interplay of the program code, processor, memory, interconnection network, and input/output (I/O) operation. Achieving extreme-scale performance requires an optimized orchestration of these components and a whole-system view to understand root causes of inefficiencies. While measurement of code execution will be more difficult due to increased processing heterogeneity, there has been very little heretofore in tools to observe other system aspects. Multilevel observations will be needed to understand behavior and performance *in toto*, and this support must be integrated with the run-time environment, operating system, I/O system, and even the machine hardware.

Greater levels of multicore parallelism and heterogeneous processors will additionally constrain observation because certain events of interest are less (not) visible and harder to measure. More highly integrated accelerator devices might not allow any access to internal parallel operation. These limitations, combined with the sheer massive parallelism, will make it impossible to observe all concurrent operations in the system. Furthermore, the increased importance of the memory system for extreme-scale computers will expand observation focus to understanding data transfer behavior. Measurement infrastructure is woefully lacking for these tasks.

A Model-Oriented Approach

To deal with the observational complexities and limitations in extreme-scale systems, tools for understanding behavior and performance must augment the information that can be reliably captured with knowledge of the application and the execution environment (Huck et al. 2008b). Knowledge of the computational semantics of the application (Li and Malony 2007), the scalability bounds of the algorithms used, the operational characteristics of the system architecture, and so on can all be applied to the design of effective experiments and the interpretation of data they produce. In fact, it will be necessary for the tools and practice for behavior understanding and performance engineering to incorporate a knowledge-discovery process to support a higher-level abstraction for debugging and tuning investigations.

A model-oriented approach can provide a framework for knowledge-based processes because it gives an abstract context for interpretation, testing, and exploration. With respect to performance engineering, models could be used to give focus to performance experiments, intelligently search the performance space, infer performance problems, and guide optimization decisions. Models could be represented at different levels of detail and used to generate performance expectations for testing against empirical data. Figure 25 provides a high-level diagram of an extreme-scale performance engineering approach based on integrated modeling, expectation evaluation, knowledge-based discovery, and performance measurement.

Traditional parallel-performance analysis methods force users to reason from the perspective of absolute performance for every experiment and every application operation, with peak measures providing an absolute upper bound. There is little context for determining whether the operations under consideration are performing well or performing poorly. In effect, performance methods implicitly require an expectation, but empirical data alone are insufficient to provide it. Scale and complexity exaggerate the problem such that the entire process breaks down and cannot support effective performance engineering.

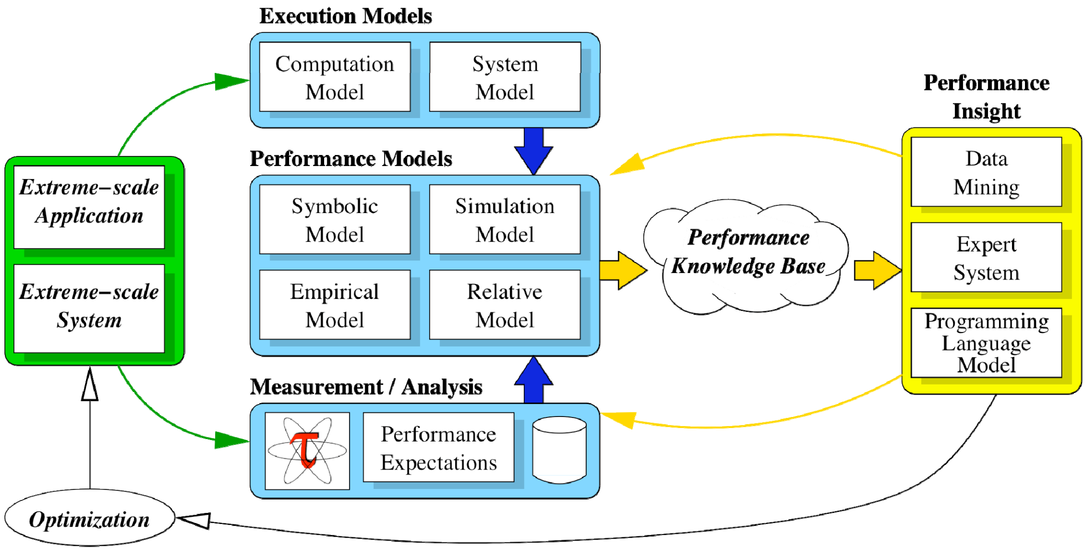


Figure 25. Extreme-scale performance engineering methodology based on integrated modeling, expectation evaluation, parallel performance measurement, and knowledge-based discovery. Image courtesy of Allen Malony (University of Oregon).

In contrast, sources for performance expectations can come from the following:

- *Computation model*—operational semantics of a parallel application, as specified by its algorithms and structure, define a space of relevant performance behavior
- *Algorithmic mode* —symbolic or analytical expressions of relevant performance parameters generate templates for constructing expectations (Alam and Vetter 2006)
- *Historical performance data—empirical* models can be mined from a multidimensional performance experiment space and used for expectation creation
- *Relative performance data—comparisons* of similar application operations across architectural components can test expectation factors
- *Architectural models—knowledge* of processor, memory, and communications performance can define bounds for expectation evaluation.

A framework for performance engineering based on performance models and expectations would be able to address scale and complexity issues by applying appropriate levels of abstraction for experimentation and analysis. It would also allow knowledge enhancement for creating more-powerful methods to understand behavior, to identify performance problems, to guide tuning decisions, and to predict performance. In addition, application developers can be more directly involved in the performance-engineering process because it can incorporate application semantics.

Integrated Tools Framework

Development of tools to understand complex application program behavior and to optimize application performance at the extreme scale must follow an integration strategy where multiple robust tool technologies can be targeted for specific objectives. Creating tool capabilities as components of an integrated tools framework allows for more-productive tool application and reuse. Existing components—such as those for performance data storage (Huck et al. 2005), data mining (Huck et al. 2008a), and visualization—should be leveraged where possible and enhanced to address extreme-scale concerns. The community of parallel tool technologists is too small to develop unused tools or tools with incompatible features.

An equally important aspect of integration concerns the incorporation of tools in the extreme-scale programming environment. A strong need exists for tightly coupled feedback of information as knowledge-based decision support in program-transformation systems. Current projects in automatic performance tuning are good examples. This need will become more acute as the optimization complexity increases. New parallel language systems will also require support for tools, ideally at the level of parallel programming semantics. In general, tools must be designed to facilitate broader environment integration.

Summary

The move to extreme-scale computing will require tools for understanding complex behavior and for performance optimization to be based on a knowledge-oriented process. Performance models and expectations will be used to drive knowledge-based investigation and reasoning. Extreme-scale computing will raise the level at which tools interoperate and can be integrated with the application development and execution environment. The challenges for performance analysis and tuning will grow because performance interactions and factor analysis must involve a whole-system perspective.

Promote a Migration Path from Current Programming Approaches to New Ones

Computational scientists have a large investment in currently successful codes. As new architectures require new programming approaches to fully exploit the codes' power, a problem arises. If the new architectures only support a programming model that is radically different from current approaches, applications will not be able to convert their large codes in a timely manner. However, if only current models are supported on the new machines, much of their capabilities will be wasted.

The solution is for the new architectures, new programming models, and applications to go forward together: large application codes need an incremental migration path into the future. The new architectures must be capable of running existing scalable applications from the beginning of their deployment, even if at significantly less than peak performance, with only modest modifications to the code. Then, tuning for performance can begin. For any radically new approach requiring significant code redesign and redevelopment, it is critical that an environment that adequately hosts the new programming model be available on current highly parallel machines.

What does this mean in the current situation? All scalable codes rely heavily on the message-passing model, expressed in the syntax and semantics of MPI, for their high-level structure. This is not the highest-level structure because some codes use libraries supporting a higher level of abstraction, but those libraries are in turn implemented in MPI. Some codes rely on multithreading within each MPI process to save memory; these codes are the ones most ready for the next generation of architectures. The multithreading is usually expressed using OpenMP, because the OpenMP and MPI standards work well together by explicit design.

While MPI is a stable and (almost) adequate mechanism for expressing parallelism among separate address spaces, the choice of programming model for expressing parallelism within a single address space (shared-memory parallelism), whether expressed in a language or a library, is far less clear. OpenMP has the disadvantage that it is difficult to program for peak performance because of its lack of mechanism for expressing locality. Other shared-memory approaches, such as the partitioned global address space (PGAS) languages, may do better but need development in the area of interactions with MPI. Similarly, brand-new approaches for programming heterogeneous processors will be of most use if they are designed for use in the context of a larger distributed-memory computation.

Very abstract parallel languages, such as those developed as part of the Defense Advanced Research Projects Agency's (DARPA's) High-Productivity Computing Systems (HPCS) project, may also play a role in a migration of existing applications to advanced architectures. Here, the critical components are 1) robust, scalable implementations on current large machines so that applications may begin experimenting with them; 2) good implementations on new architectures to demonstrate the performance potential of the new approaches; and 3) the capability of combining the new languages with existing ones so that progress can be made incrementally.

Finally, MPI itself needs to be continually refreshed to respond to the challenges of new, very large machines and to interoperate with new languages and libraries because MPI is likely to remain a critical component of large-scale codes. The MPI-3 Forum is currently addressing issues related to scalability and interoperability.

Define Common Framework Tools or Components that Can Be Reused in Multiple Application Domains

Software framework tools or libraries are desperately needed to confront the challenges of programming at the exascale, especially in effective use of multiple cores, management of dynamic load, understanding of application performance, and integration of fault resilience into large simulation. The goal is to develop framework abstractions that can be reused and tailored for multiple frameworks. Application-specific frameworks are valuable integration tools—but without software reuse, they can otherwise be enormously expensive to develop. Currently, frameworks that organize existing and future fusion codes into coherent tools for scientific investigations are in an ad hoc stage of development. Development is needed to enable component-level parallelism in conjunction with other parallel approaches. Moreover, further research into general abstractions and tools for constructing components and frameworks is needed. Such research will bring forth a new understanding of the nature of frameworks, new libraries, and tools for the construction of application-specific frameworks. This technology in software frameworks will facilitate the coupling of application subsystems to improve fidelity of simulations and will impact the design and implementation of the Fusion Simulation Project (FSP).

Run time discovery and reallocation of computing resources will be necessary for management of dynamic load imbalances and integration of fault recovery at the exascale on a million cores. The current batch-oriented computing environment enforces a static allocation of computing resources with an emphasis on keeping all processors busy. A different paradigm at the exascale is to keep some processors in reserve to enable dynamic dispatch of processors as new physics modules are activated, or as model refinement requires more resources. Intermittent node failures may further require unforeseen dynamic reconfiguration.

Tuple space is an interesting conceptual framework of communication middleware for hiding complexity associated with dynamic discovery of computing resources. Tuple space concepts have already appeared in the Linda parallel language, PVM mailbox, JavaSpaces, and IBM TSpaces. The middleware acts as associative array or a consistent database for atomic operations in writing, searching, and reading. Tuple space technology has diverse uses. It can be used to discover MPI contexts or message tags for efficient communication. Worker tasks can use tuple space to register their capability in the database, or tuple space can be used as a work queue for load leveling in a distributed manager-worker programming pattern. Critical-state information may also be stored in tuple space for fault resilience.

However, a fault-tolerant and scalable implementation of a tuple space database may be a significant implementation challenge. Disciplined use of threads with a thread-safe communication library will greatly simplify programming of tuple space by allowing background listener threads to respond to message requests for tuple lookup and computation. Remote function invocation or method dispatch, active messages, and emulated global shared memory with consistent updates can all be easily implemented. Moreover, access to tuple space, or emulated global shared memory can be implemented using library calls without inventing new language features. This feature may be attractive in leveraging the large base of legacy application codes written for MPI in Fortran or C.

The tuple space concept has been implemented in the Linda parallel programming language developed by David Gelernter and Nicholas Carriero at Yale University (Gelernter 1992). Tuple space is viewed as a distributed associative array. The language supports operations to write a tuple object into tuple space, search and retrieve a tuple, or spawn a new process to perform computation on tuples and write the resulting tuple object back into tuple space.

The tuple space concept has inspired similar capabilities in other systems such as matching of character string to MPI context [such as MPI_lookup_name(), MPI_publish_name()], or expression matching of mailbox entries [such as pvm_recv_info(), pvm_put_info()] in PVM. JavaSpaces has further developed a notification capability that signals a blocked task when a matching tuple object has been written.

Establish Methods and Systems that Enable Pervasive Fault Resilience

As several other sections of this document demonstrate, the next step in extreme-scale computing is not a mere scaling up of solutions based on lessons learned at terascale and petascale. Some of the components from which such systems will be constructed are operating at or near their limits (e.g., processor size and clock speed), and it is clear researchers can expect a dramatic increase in the number of constituent components in an exascale system. Traditional programming methods may require radical rethinking as well.

It is becoming clear (Kogge et al. 2008) that an exascale computational system will comprise a massively parallel fabric connecting roughly a million computational nodes. Each will be able to handle 100 to 1000 concurrent functional units, and each will contain a multilevel memory hierarchy globally consisting of a million or so memory modules. The corresponding exascale storage system will consist of file systems striped across millions of spinning disks.

To underline the need for a pervasive fault-resilience approach encompassing every stage, imagine a test computation on such an exascale platform, with a known result. The computation is executed on the many nodes of this massively parallel system, touching millions of memory modules, communicating across a million sockets, executing on hundreds to thousands of functional units at each node, and finally writing the result to exascale storage striped across a million rotating disks. If an unexpected result is reported by the time the computation is written, it will be too late for any remedial measures. A pervasive approach intervenes at every stage in this process and must seamlessly handle faults in hardware and software.

The following questions are posited from the perspective of a computational fusion researcher. What should one expect from an exascale system? What sorts of errors are to be expected? How will they be detected? Will they be corrected in hardware, by the run time system, or by user code? What happens if a fault is detected but cannot be corrected—how will graceful recovery of such a large system from failure be handled?

Explored below are the limits of fault resilience in exascale hardware and run time systems, and what algorithms, programming models, and software engineering can do to extend those limits. The key message is to promote pervasive fault resilience and to understand the limits of inbuilt fault tolerance and how intelligent software can extend it at every stage.

Fault Resilience in the Memory Subsystem

Memory—whether it is the dynamic random access memory (DRAM) typically used for main memory, the static random access memory (SRAM) used for the fastest operations such as L1 and L2 cache, or the more recent memory incarnations such as "NAND flash" memory—is subject to both hard errors (failure to read or write a word) and soft errors (a read/write is completed but the wrong bits are read). These errors are induced at transistor voltage barriers by the pervasive radiation field in which the earth is embedded. Cosmic rays can cause anomalous electron transmission/nontransmission events.

Soft errors can generally be corrected with error-correcting codes (ECCs). The customary ECC approach, known as single-bit-error correction, double-bit-error detection (SECDED), uses redundant bits at the resolution of a memory word to detect cosmic-ray-induced random bit flips. Single-bit errors are corrected and the computation may continue; double-bit errors throw an exception. Burst errors that cause three or more bits in a single word to be incorrect and that are undetected in SECDED are fortunately quite rare. More expensive ECCs, such as Reed-Solomon, can be constructed to tolerate more erroneous bits with a concomitant loss in memory performance.

While supercomputing pioneers such as Seymour Cray initially scoffed at ECC, later supercomputers have almost universally used it and non-ECC supercomputers have been shown to be no more or less than expensive cosmic-ray detectors. SECDED should be regarded as a minimum requirement for memory subsystems.

Failure rates of memory have improved dramatically over time, and error rates of order of single failures per billion hours of operation can be expected for a projected exascale system. This system will consist of a million such components, which translates into a global failure rate of 1 per 1000 hours.

Fault Resilience in Storage Media

Magnetic storage remains the predominant storage technology and is a likely subsystem of any exascale system, rotating disks for nearline storage and tape for offline storage. Once again, the sheer scale of an exascale system is a significant challenge: an exabyte of storage on disk is likely to be spread across somewhere between 10^5 to 10^6 drives.

Drive failure remains the most common mode of error in storage. Redundant disks are now commonplace, with one or even two RAID-6 redundant disks on every spindle for every four primary disks.

More insidious is the issue of erroneous bits being read to or written from disk. Disk reads and writes also commonly use parity bits for ECC, as in SECDED, for memory. With ECC enabled, soft error rates during disk operations are about 1 per 10^{14} bits and can perhaps be brought as low as 1 per 10^{21} bits. Exascale poses a challenge by its sheer size: an exabyte is 10^{19} bits, yielding an uncomfortably low margin for error. It is worth noting that vendor-generated performance claims could well be within ECC for error rates, and without ECC for bandwidth estimates—using ECC for bandwidth estimates should be avoided.

While there are promising emergent storage media, such as holographic memory, these media are still relatively immature. Iterative decoding serves as a relatively expensive method for ECC in holographic memory. As previously noted, caution is warranted when assessing claims of bandwidth to storage of such systems to verify that reported measurements include ECC or not.

To improve fault resilience in storage, researchers recommend that checksums be computed and stored for every key input and output file associated with an exascale computation. This yields the possibility of verifying every file before it is used, and retransmitting or regenerating the file if needed.

There is of course a significant cost associated with checksum computation and verification. Application-aware checksums that ignore unused bits could reduce this burden; another intriguing possibility under consideration at some sites is the use of FPGAs (field-programmable gate arrays) to build specialized checksum units capable of performing checksums at “hardware” rather than “software” speeds.

As a last resort in guarding against data loss, it must be possible to regenerate any dataset exactly (thus treating the generating program as a compressed and exact representation of the bits on disk). This necessitates capturing and recording the precise sequence of events (or “workflow”) involved in the execution. Workflow languages such as Kepler are becoming more common in the scientific workplace and are sometimes coupled with archives capable of replicating their contents (“data curators”). Of course, for the replication to be exact to the bit level, the program must run on the same hardware as before or must be coded to a virtual machine guaranteed to reproduce answers exactly even on different hardware.

Fault Resilience in the Communication Fabric

The exascale system under consideration is likely to have a vast communication fabric consisting of order 10^5 sockets. Extrapolations to the exascale era (Schroeder and Gibson 2007) lead one to expect a socket failure every 24 minutes, or every 4 hours based on an optimistic expectation of 10-fold improvement in socket-level reliability. It is still shorter than a typical job sitting in the queue of any of today’s large systems.

What happens in the event of a socket failure? Today’s operating systems are resilient to node failure: nodes can be swapped in and out of racks without a system-wide interrupt. It can therefore be expected that jobs not directly on the failed node continue to operate normally.

The first level of response to a node failure (at least one that is noncatastrophic and trapped by an error handler) should be a system-initiated checkpoint/restart (CPR) allowing the node to save state, so that it can be restarted exactly where it was interrupted. System-level CPR is likely to be fraught with difficulty as the memory per socket on an exascale system will be measured in the terabytes, and simply streaming it to disk is likely to take minutes even on extremely high-performance disks.

Application-level checkpointing can significantly reduce this burden. Modern frameworks such as the Earth System Modeling Framework (ESMF) (Collins et al. 2005) in the climate domain contain built in data structures that serve as state vector container classes. These state variables contain exactly the information needed to restart the model and can be considerably more compact than the full memory per socket. Such frameworks are set up to trap signals from the operating system to initiate application-level CPR.

Another approach being considered is for the communication library (e.g., MPI) to build in redundancy, so that each MPI process is fully replicated on another node. In the event of a node failure, the MPI shepherd process will switch to the second, redundant copy of the process. Of course, there are open questions on this approach, such as *how often do the replicant processes copy their state?* Another open question is *what is the expected behavior should neither of the nodes fail but their states differ?* A new MPI standard encompassing these features is expected to be available in 2010.

Fault Tolerance on the Computational Node

The computational node on the exascale system is expected to hold 100 to 1000 cores, or perhaps a "sea of functional units" to use Kathy Yelick's phrase.[1] A complete failure of any of these units would place the user in the situation described above where the system fabric must cope with a failed node.

More subtle is the problem of irreproducible computation on a multicore chip. While chips on the scale proposed for exascale computing are still on the drawing board, researchers have some experience with the problem on smaller scale even on today's 2-way and 4-way nodes.

Irreproducibility has been a pernicious problem on some recent large systems. Dual runs of some applications result in intermittently producing different results, requiring a third run to break the tie—as in Byzantine fault-tolerance approaches. These were caused by subtle, hard-to-detect, and rarely encountered hardware race conditions. A recommended practice for managing large systems is to maintain a continuous background level of dual-running—for example, 5% of the system load randomly sampling and dual-running jobs on the system.

More fundamental still is the problem of chips not being designed with reproducibility in mind. Graphics processing units (GPUs), for instance, have massive thread concurrency on-chip but do not have an execution consistency model for threads. Because floating-point operations are non-associative, two different executions of the same operand on a GPU are not guaranteed to produce the same result to round off. For graphics applications, small errors in pixel-shading can be tolerated, but when the same chip is proposed for computational science applications, this becomes a serious problem. There is no good solution to this issue. Perhaps the 1000-core systems that will be the basis of the exaflop system

[1] Katherine Yelick (Lawrence Berkeley National Laboratory, University of California-Berkeley).

will indeed embrace execution consistency, but there is a significant likelihood that they will not. It may then be up to the scientific community to develop working methods on such systems.

Certain classes of problems can be constructed to be tolerant of irreproducibility. Initial-value problems with uncertain initial conditions, for instance, are often solved using ensemble methods that sample the uncertainty distribution of the initial conditions. There, the statistically significant answer only attaches to the ensemble mean and has error bounds associated with it. Elliptic problems are also often solved only to within a specified tolerance.

In view of this, there are hardware approaches that actively embrace irreproducibility, such as the "probabilistic CMOS" being pioneered by Palem and colleagues (see Korkmaz et al. 2006). The scientific community is now well on its way to seeing computation more like experimentation, with no two realizations likely to be exactly the same. Large-scale computational systems become more like biological systems, with many cells as well as many molecular and signaling pathways, and the system as a whole is fault-resilient but never twice the same.

Summary

Exascale systems are likely to challenge traditional approaches to fault tolerance at the hardware and run time system level. This section examines approaches to fault tolerance in storage, in system memory, in communication fabrics, and on the computational node. Key findings include the cost of error correction and redundant design, which must be taken into account in assessing system performance and specifications. Application software can significantly enhance fault resilience such as in reducing the burden of checkpointing. Researchers recommend a pervasive fault resilience approach seamlessly spanning hardware, system software, and applications. Finally, reproducibility of computation at the bit level is increasingly at risk, and the community should consider how to operate in a world where the distinction between approximate *in vitro* approaches and exact *in silico* approaches is increasingly blurred.

CONCLUSIONS

Computer architectures are now being designed that take scientific computing beyond the current petascale regime into the exascale regime. In the rest of this document, advances in fusion science that will be enabled by such new computational capabilities are described. This section has described the challenges this transition provides for the most fundamental aspects of the code development process, the programming model, which describes the way researchers think about the machine controlled by researcher-developed programs. This section identifies six PRDs in computer science, where progress will be necessary to enable fusion science to exploit the most advanced computational tools of the 21^{st} century.

PRIORITY RESEARCH DIRECTIONS

- **BURNING PLASMA/ITER SCIENCE CHALLENGES**
- **ADVANCED PHYSICS INTEGRATION CHALLENGES**
- **PLASMA-MATERIAL INTERACTION SCIENCE CHALLENGES**
- **HIGH-ENERGY DENSITY LABORATORY PHYSICS/LASER-PLASMA INTERACTIONS**
- **BASIC PLASMA SCIENCE/MAGNETIC RECONNECTION PHYSICS**

BURNING PLASMA/ITER SCIENCE CHALLENGES

This panel identified five priority research directions for which significant advances in understanding are needed to achieve targeted levels of controlled magnetic fusion power. These include the following topics:

Development of a new generation of magnetohydrodynamic (MHD) codes capable of accurately modeling the onset of plasma disruptions and their effects on the device components. The driving goal is to develop an improved macroscopic-simulation capability for ITER-class experiments. This is a critical goal because nonlinear macroscopic events play a central role in defining the operational space of these devices, and many details of the nonlinear processes and interactions are poorly understood.

Greater understanding of plasma transport and turbulence. This is a key physics requirement for enabling achievement of the required energy confinement time in fusion plasmas. A critically important challenge is associated with the recognition that realistic transport simulations for burning plasmas demand the development of a) electromagnetic simulation capabilities; and b) the ability to address the coupling of global, nonlocal transport on an equal-footing with MHD phenomena.

Realistic capability for simulating the physics of the edge barrier region in high-performance burning plasmas. Understanding the dynamics in this region, which are characterized by strong pressure gradients, is critical for optimizing performance in burning plasmas. The goal is to be able to conduct a comprehensive analysis across a wide range of overlapping spatio-temporal scales that include both the relevant small-scale kinetic/gyrokinetic dynamics and the large-scale MHD physics.

Experimentally validated predictive simulations of energetic particle dynamics in burning plasmas. This involves the development of realistic, self-consistent modeling capabilities for fusion alpha particle profiles in the presence of multiple Alfvénic and MHD instabilities.

Radio frequency wave heating and current drive for burning plasma scenarios. This involves the development of reliable simulations for the larger configuration dimensions of systems (such as the ITER project) of the following: a) wave propagation and coupling efficiency in the high-temperature pedestal region; and b) radio frequency interactions with fusion alpha particles.

Achieving significant progress in a timely manner for all of these grand challenge areas will require development of advanced simulation capabilities using computing at the extreme scale.

Progress in these five areas will require teams of fusion energy scientists, applied mathematicians, and computer scientists to address problems across the range of physics, algorithms, data management, dynamic load balancing, and code modernization. Detailed discussions of the research needed to make substantive progress in all of these key priority research directions are provided in the preceding main panel report titled, “Burning Plasma/ITER Science Challenges.”

ADVANCED PHYSICS INTEGRATION CHALLENGES

This panel identified five priority research directions in which computing at the extreme scale would make a significant impact. These include the following topics:

Transport modeling with embedded turbulence. Computation offers the highest-fidelity path to the calculation of plasma profiles. Approaches include the following: a) integration of well-parallelized local computations of turbulent fluxes within a code that advances plasma profiles in response to sources of heat, momentum, current and particles; and b) coupling of global turbulence with transport over the same region—probably a necessary approach for dealing with the plasma edge. Challenges include verification and validation (with associated uncertainty quantification), formulating new mathematical algorithms, and addressing the lack of data alignment between the calculation of sources and transport.

Coupling disparate regions of the plasma. This capability is needed for a whole-device model that includes core, edge, and plasma-facing materials. Associated research areas of focus include the following: a) developing reduced models for edge dynamics that are closer to first-principles calculations; and b) addressing the coupling of sources in both the plasma edge and core.

Macroscopic stability control using radio frequency power. This is a well-known capability important for fusion devices. A classic example is the use of electron cyclotron waves to drive plasma currents that suppress key instabilities (such as neoclassical tearing modes). Associated focused research topics include reformulation and new code implementation when the non-inductively driven current is an integral part of the MHD equilibrium and stability evolution.

Recoverable non-axisymmetric macroscopic dynamics. These processes include periodic instabilities, such as internal sawtooth reconnections in the central part of the plasma and edge localized modes. Transport leads to thermal and particle profiles that are unstable. These instabilities then transiently alter the plasma profiles. Focused associated research needs here include development of periodic temporal coupling of computations involving brief intervals of rapid macroscopic dynamics and longer intervals of axisymmetric transport. Such couplings also have application to the key area of disruption mitigation, which involves ideal and resistive MHD, runaway electron dynamics and transport, pellet and gas fueling, and plasma-wall interactions.

Performance optimization of burning plasmas. This brings all of the preceding four PRDs together, but with even greater computational requirements to run with different parameter sets to optimize plasma profiles over control parameters, such as external energy and current drive sources.

Progress in these five areas will require teams of fusion energy scientists, applied mathematicians, and computer scientists to address problems across the range of physics, algorithms, data management, dynamic load balancing, and code modernization. Detailed discussions of the research needed to make substantive progress in all of these key priority research directions are provided in the preceding main panel report titled, "Advanced Physics Integration Challenges."

PLASMA-MATERIAL INTERACTION SCIENCE CHALLENGES

This panel identified three priority research directions with the common goal to develop comprehensive computational models for predictive, self-consistent, integrated, validated, full-process, time-dependent, plasma/material interactions. All three of these areas are expected to benefit significantly from the impetus provided by extreme-scale computing.

Modeling of the edge and scrape-off layer plasmas. This includes modeling of turbulent transport and full coupling of plasma ions and electrons, neutrals, photons, and electromagnetic fields. In addition, plasma contamination from near-surface transport of sputtered or vaporized material and quantification of plasma facing component particle and photon fluxes (with predictions of instability regimes) should be considered.

Predicting the near-surface material response to the extreme plasma fluxes of photons and particles under normal and transient operation. This includes predicting sputtering erosion/re-deposition and other time-integrated plasma facing component processes (e.g., dust formation and transport; helium- or deuterium-tritium-induced microstructure formation and flaking) and the resultant impurity transport, core plasma contamination, mixed-material formation, and tritium co-deposition in redeposited materials. The material and edge plasma response to transient processes such as high-powered edge localized modes vertical displacement events, plasma disruptions, and runaway electrons represent an important component of this effort.

Modeling the underlying structural materials response. This involves understanding the fundamental microstructure evolution and performance limits of structural materials in the fusion radiation environment that involve extreme cyclic thermo-mechanical stresses and simultaneous intense fusion neutron bombardment.

An overarching grand challenge will involve efficient integration of these three coupled PRDs to develop a comprehensive model. The associated collective impact on FES includes enabling a) effective operation of the ITER and proper design of DEMO; b) improved understanding of present experiments; and c) a plasma-material interaction code package for the macro-type code packages needed by the proposed Fusion Simulation Program.

Progress in these three areas will require teams of fusion energy scientists, applied mathematicians, and computer scientists to address problems across the range of physics, algorithms, data management, dynamic load balancing, and code modernization. Detailed discussions of the research needed to make substantive progress in all of these key priority research directions are provided in the preceding main panel report titled, "Plasma-Material Interactions Science Challenges."

LASER PLASMA INTERACTIONS AND HIGH-ENERGY DENSITY LABORATORY PHYSICS

Four priority research directions were identified for high-energy density laboratory plasmas (HEDLP) and inertial fusion energy science (IFES) for which extreme-scale computing could make a transformative impact.

Nonlinear optics of plasmas. The goal is to understand how an ensemble of overlapping Gaussian beamlets (speckles) mutually interact in HEDLP. This understanding is critical to successful development of inertial fusion energy (IFE) concepts using laser drivers. It requires fully kinetic modeling because subtle changes to the electron distribution function can lead to substantial differences. On extreme-scale computers, the goal of simulating an ensemble of speckles using fully kinetic modeling could be achieved. This could—in turn—lead to ideas on how to tame these interactions and the development of high-fidelity reduced models for mesoscale simulations.

Relativistic high-energy density plasma and intense beam physics. The goal is to understand how lasers at the intensity and power frontier interact with and are absorbed in HEDLP. Because the associated physics requires detailed understanding of single-particle trajectories and how the complex patterns of large currents of relativistic particles form in plasmas and collectively interact, fully kinetic and relativistic modeling are required. On extreme-scale computers, fully kinetic simulations using true time and length scales of fast ignition targets could be possible for the first time. This will also require development of coupled microscale and mesoscale models.

Integrated fast ignition simulations. The goal is to provide full integrated modeling of high-gain, fast ignition IFE concepts where the timing of the intense ignition pulse, the compression of the pellet, and survival of an inserted cone tip can be important. On extreme-scale computers, the coupling of fully kinetic simulations of HEDLP with parameters obtained from macroscale hydrodynamic compression models may be possible, thereby enabling simulations representing the true time and space scales.

Magnetized high-energy density plasmas. The goal is to understand how spontaneous or induced magnetic fields can affect burning HEDLP. The physics spans a wide parameter space, from the dense compressed core of a traditional IFE target, as well as the more tenuous plasmas in reversed field configurations. Extreme-scale computers will enable high-fidelity simulations of dense collisional plasmas that are inertially confined and in which heat flux is limited by magnetic fields. The development of mesoscale models, coupled with extreme computing, should enable breakthroughs in the understanding of magnetized plasmas under compression.

Progress in these four areas will require teams of fusion energy scientists, applied mathematicians, and computer scientists to address problems across the range of physics, algorithms, data management, dynamic load balancing, and code modernization. Detailed discussions of the research needed to make substantive progress in all of these key priority research directions are provided in the preceding main panel report titled, "Laser-Plasma Interactions and High-Energy Density Laboratory Physics."

BASIC PLASMA SCIENCE/MAGNETIC RECONNECTION PHYSICS

Looking to the future, significant progress on four priority research directions in this basic plasma science grand challenge area were identified for which computing at the extreme scale could enable higher physics-fidelity simulations of magnetic reconnection physics for most applications of interest.

Influence of the electron and ion kinetic scales on the large-scale evolution. Currently, there are significant differences between fully kinetic and two-fluid simulations in weakly collisional regimes. Thus, there is no clear consensus on the minimal physics required to accurately capture the large-scale evolution. First-principles kinetic simulations, including Coulomb collisions, can provide a guidepost for developing reduced fluid descriptions that better capture the structure and dynamics. Other approaches may include reduced kinetic descriptions such as the following: a) the gyrokinetic model, and b) the hybrid model that embeds a kinetic description within a larger fluid simulation.

Reconnection and magnetic island dynamics in three-dimensional geometries. Evidence exists that a single reconnection layer may divide into multiple reconnection sites due to the formation of secondary magnetic islands or other secondary instabilities (such as ballooning modes) that may control the relaxation of current and pressure profiles in tokamaks. Evolution of reconnection dynamics on both fast- and long-transport time scales, including kinetic effects, is of great interest for fusion as well as space and astrophysical applications. Addressing these issues will require highly scalable fluid and kinetic algorithms, along with a realistic treatment of boundary conditions.

Energy partition and particle acceleration that results from reconnection. Thermal energy gained by ions and electron, as well as the formation of nonthermal tails, is of significant theoretical and observational interest. For the highly energetic tails, it is difficult to explain the observations with a single steady-state reconnection site. One critical question is whether most nonthermal particles are directly associated with reconnection sites and magnetic islands, or with other processes associated with the global relaxation (such as waves and shocks).

Reconnection in relativistic plasmas. In many astrophysical applications (pulsars, accretion near black holes, gamma-ray bursts), reconnection is thought to occur in highly relativistic regimes with both hydrogen and electron-positron plasmas. These regimes are well suited for relativistic kinetic simulations that are now feasible in three-dimensions at the petascale for electron-positron plasmas. These advancements in reconnection physics have the potential to impact fusion energy science through the following: a) more realistic modeling of tearing modes and sawteeth oscillations in tokamaks; b) understanding magnetic relaxation in reversed field pinches, stellarators, and field-reversed configurations; and c) higher physics-fidelity modeling of relativistic electrons for fast ignition.

Progress in these four areas will require teams of fusion energy scientists, applied mathematicians, and computer scientists to address problems across the range of physics, algorithms, data management, dynamic load balancing, and code modernization. Detailed discussions of the research needed to make substantive progress in all of these key priority research directions are provided in the preceding main panel report titled, "Basic Plasma Science/Magnetic Reconnection Physics."

ALGORITHMS FOR FUSION ENERGY SCIENCES AT EXTREME SCALE

Six priority research directions emerged for scalable algorithms that are relevant to accelerating progress in fusion energy science simulations.

Optimal representations. Full adaptivity in the sense of *h* (mesh refinement), *p* (discretization order), and *r* (mesh relocation) should be employed in space and time, according to the local smoothness of fields to be represented, to get the most "science per watt" out of a fusion energy science modeling simulation. This requires estimating and equi-distributing truncation errors, dynamic in-place load balancing, and managing and converting between different representations.

Multiphysics and multiscale algorithms. Algorithms that allow self-consistent coupling of multiphysics models across all relevant scales allow better focus on physical questions, free of concern about numerical instabilities and splitting errors, and longer windows of integration due to suppression of stability-limiting fast scales with greater accuracy. This requires scalable implicit methods and high-order interpolations between representations (e.g., from fields to particles and vice versa).

Real-time algorithms. Armed with first-principles models, reduced-order models can be parameterized for sufficiently narrow regimes to provide detection and control capabilities in real time. This requires physics-based developments beyond current models based on principal component analysis or proper orthogonal decomposition.

Optimization. Robust (error-tolerant) optimization algorithms are needed for high-dimensional multiphysics models for optimal design, control, parameter estimation and the mapping of stability boundaries. Required are deterministic and stochastic techniques for derivative-free methods, adjoint-based derivative methods, and preconditioners for saddle-point systems.

Uncertainty quantification (UQ) and reduction. Models contain uncertainties in initial conditions, boundary conditions, coefficients, and/or forcings, coming from observations or other simulations. Incorporation of observations can improve uncertain models, balancing models, and numerical errors for more efficient computation. Needs include deterministic UQ tools based on sensitivity and adjoint techniques, probabilistic approaches based on sampling methods, and direct propagation of probability density functions from inputs to outputs.

Lower threshold of expertise required to use optimal algorithms on extreme architectures. Software for extreme-scale environments must offer multilevel ("incremental adoption") user interfaces. With proper interfaces to widely used (and therefore thoroughly debugged) modules, software will perform as closely as possible to expert reliability while auto-tuning or being tunable for high performance by expert users. With such tools, fusion energy physicists will work more productively and better understand the performance of their software tools, thus focusing more on physics and less on software issues.

Progress in these six areas will require teams of fusion energy scientists, applied mathematicians, and computer scientists to address problems across the range of physics, algorithms, data management, dynamic load balancing, and code modernization. Detailed discussions of the research needed to make substantive progress in all of these key priority research directions are provided in the preceding main panel report titled, "Algorithms for Fusion Energy Sciences at Extreme Scale."

DATA ANALYSIS, MANAGEMENT, AND VISUALIZATION IN FUSION ENERGY SCIENCE

Five priority research directions will require extensive research and development effort to support the data requirements at the extreme scale for fusion energy science.

Managing large-scale input/output volume and data movement. Techniques need to be developed that optimize input/output performance automatically based on hardware characteristics. Such techniques are crucial to avoid slowdown of computations because of insufficient input/output rates. Furthermore, future fusion energy science codes should be as independent as possible of input/output tuning, where all such details are processed automatically by the underlying input/output system. Parallel file systems and data movement tools need to be scaled to support these extreme volumes of data.

Real-time monitoring of simulations and run-time metadata generation. Having run-time monitoring capability on all supercomputing resources is essential to avoid computational waste. This capability will prevent runs that do not converge or progress correctly from continuing. Workflow technology already used for such purposes in fusion energy science applications need to be scaled and become part of the simulation system that supports summarization of results in real time, and/or permit the monitoring software to automatically manage simulations that do not progress correctly. Additionally, provenance and metadata information needs to be automatically collected (also at run time) for effective run-time and post-run data analysis.

Data analysis at extreme scale. The data analysis challenges in fusion energy science applications at the extreme scale stem not only from the large size of the data, but also from data complexity. First, areas of interest— such as coherent structures and fronts—are likely to be spread across many processors, making it difficult to extract poorly defined structures or track fronts over time. Second, techniques to process these data to reduce overall size before these data are output by the simulation require algorithms that are robust enough to process data correctly.

Visualization of very large datasets. Visualization is often a key technology for understanding data such as electron-temperature profiles. However, reducing and mapping terabytes or petabytes of data into meaningful visualization is a challenge that will require processing near to where the data are stored, as well as effective indexing techniques for real-time data exploration.

Experiment-simulation data comparison. Such tools are essential for validation of FES simulations and diagnostics, and for comparing shot data to reduced models for ITER runs. Experimental data are expected to grow to terabyte sizes, and therefore robust synthetic diagnostic tools need to be developed that are cross-platform scalable and based on forthcoming community standard data formats.

Progress in these five areas will require teams of fusion energy scientists, applied mathematicians, and computer scientists to address problems across the range of physics, algorithms, data management, dynamic load balancing, and code modernization. Detailed discussions of the research needed to make substantive progress in all of these key priority research directions are provided in the preceding main panel report titled, "Data Analysis, Management, and Visualization in Fusion Energy Science."

MATHEMATICAL FORMULATIONS

Inheriting structure from the topics of the five fusion energy science panels, panel members identified one or more priority research directions in each.

Burning Plasma/ITER Science Challenges. The main priority is the need for high-fidelity kinetics calculations, both in the core and in the edge region. Additional priorities also include more accurate gyrokinetic approximations, systematic methods for constructing nearly field-aligned coordinates, fundamental new numerical algorithms for particle-in-cell, the need (or lack thereof) for symplectic integrators for both particle-based and continuum-based methods, and treatments of kinetic electrons.

Advanced Physics Integration Challenges. There is a need for a mathematically systematic treatment of coupled systems with vastly different spatial and/or temporal scales, including well-posedness, stability, and accuracy. A classic example is the coupled treatment of turbulence and transport.

Plasma-Material Interaction Science Challenges. The main priority is the design of materials to withstand tokamak operating conditions, a topic outside the scope of numerical plasma physics. A second priority is interaction of the plasma environment with material boundaries. In the latter area, topics include the improvement of the fidelity of edge models with respect to the interaction with the boundary; the effects of impurities on the overall plasma; and the impact of liquid walls.

High-Energy Density Laboratory Plasma/Laser-Plasma Interactions. This priority includes understanding the interaction of the laser with plasma heterogeneities, known as speckles. Mathematically, this is a homogenization problem: scientists want to understand and represent the collective effect of thousands of speckles, while currently it is only possible to compute the interaction of the laser with one such speckle. This leads to the development of reduced/meso-scale models derived from large-scale Hydrologic Engineering Center calculations.

Basic Plasma Science/Magnetic Reconnection Physics. This is primarily a multiscale problem, exhibiting kinetic behavior in highly localized regions in space, combined with fluid behavior on larger scales. The traditional approach of using two-fluid extended magnetohydrodynamic is questionable physically (particularly for larger scale problems), and difficult numerically while the kinetic models that are correct in reconnection zones are too expensive to use globally. This is an opportunity to introduce hybrid fluid-kinetic models that have been used successfully in other areas of fluid dynamics.

Progress in these five areas will require teams of fusion energy scientists, applied mathematicians, and computer scientists to address problems across the range of physics, algorithms, data management, dynamic load balancing, and code modernization. Detailed discussions of the research needed to make substantive progress in all of these key priority research directions are provided in the preceding main panel report titled, “Mathematical Formulations.”

PROGRAMMING MODELS, FRAMEWORKS, AND TOOLS

To address challenges in programming models for fusion energy science, this panel identified six priority research directions.

Find efficient algorithms and implementations that exploit new multicore, heterogeneous, massively parallel architectures. This research is directed primarily at languages, libraries, and runtime systems that allow fusion energy science programmers to use massive on-chip concurrency in a portable, cross-architecture manner while cooperating with interprocessor parallelism.

Find new, productive approaches to writing, integrating, validating, and tuning complex fusion energy science application programs. This involves development of programming models and systems for massive numbers of processors.

Develop tools for understanding complex application program behavior at scale and for optimizing application performance. This requires the evolution of existing tools and development of new ones to address heterogeneous processors and greater integration of model-based approaches in fusion energy science.

Ensure a migration path from current fusion energy science programming approaches to new ones. Existing Fortran + message-passing interface codes will continue to be used and extended as architectures scale up. Research into message passing interface interoperability and extreme scalability will be required, together with a new software development ecosystem that spans all scales of systems, from midrange to the exascale, to facilitate a viable migration path from development to large-scale production computing systems.

Define common framework tools or components that can be reused in multiple fusion energy science application domains. Frameworks that organize existing and future fusion energy science codes into coherent tools for scientific investigations are currently in an ad-hoc stage of development; research into general abstractions and tools for constructing components and frameworks are needed.

Establish methods and systems that enable pervasive fault resilience. At the exascale, faults of various kinds in both hardware and software components are expected to become commonplace in the execution environment. Fault recovery mechanisms will need to be integrated at every level of the system design—in hardware, software, and the programming model for fusion energy science applications.

Progress in these six areas will require teams of fusion energy scientists, applied mathematicians, and computer scientists to address problems across the range of physics, algorithms, data management, dynamic load balancing, and code modernization. Detailed discussions of the research needed to make substantive progress in all of these key priority research directions are provided in the preceding main panel report titled, "Programming Models, Frameworks and Tools."

CONCLUSIONS AND RECOMMENDATIONS

Five major areas of fusion energy science are discussed in this report:

- Burning Plasma/ITER Science Challenges
- Advanced Physics Integration Challenges
- Plasma-Material Interaction Science Challenges
- Basic Plasma/Magnetic Reconnection Physics.

Within each area, extreme-scale computational resources are required to accelerate progress on the priority research directions (PRDs) that are crucial to the advancement of fusion energy science.

Workshop participants also provided the multidisciplinary expertise required to identify and address crosscutting challenges in high-performance computing with an emphasis on the use of extreme-scale computing for scientific research to enable the needed advances and discoveries. Advanced scientific computing research crosscutting challenges, which impact the five major fusion energy science areas, are discussed in this report:

- Algorithms for Fusion Energy Sciences at Extreme Scale
- Data Analysis, Management, and Visualization in Fusion Energy Science
- Mathematical Formulations
- Programming Models, Frameworks, and Tools.

FUSION ENERGY SCIENCE PANELS

Burning Plasma/ITER Science Challenges

As fusion research enters a new era of burning plasma experiments on the reactor scale, it becomes increasingly urgent to develop experimentally validated predictive capabilities that can produce accurate and robust simulations. This is particularly important for mitigating the risk associated with achieving—in a timely manner—the desired plasma performance in major investments such as the ITER project. At the highest level, two main concerns in producing the required capabilities involve addressing the larger spatial and longer energy-confinement time scales. Assessments based on fundamental, first-principles physics considerations indicate that scales spanning the small gyro-radius of the ions to the radial dimension of the plasmas will need to be addressed when properly simulating the dynamics in a magnetically confined burning plasma. Compared to present-day experiments, an order of magnitude greater spatial resolution is needed to account for the larger plasmas of interest, and the major increase expected in the plasma energy confinement time (~1 second in the ITER device), together with the longer pulse of the discharges in these superconducting systems, will demand simulations of unprecedented aggregate floating point operations.

Productive discussion during the workshop resulted in the identification of a number of important challenges relevant to the physics of burning plasma/ ITER experiments. Any major breakthrough in the PRD's outlined in the "Burning Plasma/ITER Science Challenges" panel report will have an immediate impact on the magnetic fusion energy research program and on the effective design of a fusion reactor. The actual delivery of such scientific advances is becoming increasingly urgent in view of the current construction phase of the ITER and the international discussions of the design for the next-step magnetic fusion reactor.

Because of the multibillion-dollar construction cost and the complexity of a future fusion reactor, advanced numerical simulation capabilities that are validated against experiments will clearly need to be in place before a reactor prototype or Demonstration Reactor (DEMO) can be properly designed and constructed. As the fusion program progresses, it is expected the required optimization of plasma scenarios under burning plasma conditions will demand major advances in scientific understanding. Associated research campaigns, greatly aided by computing at the extreme scale, will likely result in key discoveries of new conditions and important new physics insights to accelerate progress toward resolving burning plasma/ITER scientific grand challenges.

Advanced Physics Integration Challenges

Computational modeling is expected to have a major impact on the fusion plasma science program. Because of the high cost of each discharge in burning plasma experiments (such as the ITER project), planning experimental campaigns and analysis of data demand simulations with unprecedented physics fidelity. Traditionally, computational FES has addressed separate areas such as macroscopic stability, energetic particles (from auxiliary heating sources including radio frequency waves and neutral-beam injection, and also as products from the fusion reactions), microturbulence and associated transport, and edge plasma physics (where atomic processes are important). Each of these areas has currently demonstrated at varying levels of efficiency the capability of productively using existing leadership class computing facilities. With extreme scale computational power, it will be possible to couple improved versions of these large-scale simulations to produce an experimentally validated *integrated* simulation capability for scenario modeling of the whole device.

In general, a magnetically confined fusion-grade plasma is too complex to be simulated with first-principles computations alone—even with the availability of extreme-scale, high-performance computing resources. However, such powerful hardware coupled with the enabling software will make it possible to use the most advanced models to deliver experimentally validated predictive simulation results with much higher physics fidelity. Still, there are several challenges associated with making effective use of exascale resources. These issues range from advancing beyond the present state where multiple two-way couplings are being used to a fully integrated model. Meeting these challenges will require application of the best available physics, applied mathematics and computer science methods to accelerate progress. This goal can be accomplished through productive interdisciplinary collaborative alliances.

Plasma-Material Interaction Science Challenges

Plasma and material interactions are among the most critical scientific issues for fusion power, affecting the following: 1) lifetime of plasma-facing components due to sputter and transient erosion; 2) plasma contamination by eroded material; 3) tritium co-deposition in eroded/re-deposited material; and 4) operating limits on core plasma (beta, confinement, edge temperature/density, duty factor, etc.) as a result of the above factors. A related critical topic is bulk material performance and optimization. Gaining understanding and predictive capabilities in this vitally important area requires addressing simultaneously complex and diverse physics occurring over a wide range of lengths (angstroms to meters) and times (femtoseconds to days). This will require further development of not only detailed physics models and computational strategies at each of these scales but also algorithms and methods to strongly couple them in a way that can be robustly validated. While present research confined to each of these scales, or pioneering approaches to couple two or more of them, already push the state of the art in technique and available computational power, simulations spanning multiple scales needed for major future fusion energy projects (e.g., ITER and DEMO) will require extreme-scale computing platforms and integrated physics and computer science advances.

The key challenge is to develop comprehensive simulation models for predictive, self-consistent, integrated, validated, full process, time-dependent plasma material interaction that can effectively utilize the needed computational resources at the exascale. This would first encompass modeling of the edge and scrape-off layer plasma, including treatment of kinetic effects, three-dimensional geometry, turbulent transport, and full time-dependent coupling of plasma ions and electrons, neutral particles, photons, and electromagnetic fields. Next, plasma contamination from near-surface transport of sputtered or vaporized material and quantification of plasma-facing components particle and photon fluxes (and prediction of instability regimes) would be included. A related issue is predicting the near-surface material response to the extreme plasma fluxes of photons and particles under both normal and transient operations. This involves modeling of sputtering erosion and re-deposition and other time-integrated plasma-facing components processes and the resultant impurity transport, core plasma contamination, mixed-material formation, and tritium co-deposition in re-deposited materials. The material and edge plasma response to transient processes, such as high-powered edge-localized modes, vertical displacement events, plasma disruptions, and runaway electrons, would be an important component of this effort.

High-Energy Density Laboratory Physics/Laser-Plasma Interactions

Recent technological advances in lasers, particle beams, and Z-pinches have made it possible to generate plasmas with unprecedented energy densities in the laboratory. Understanding the properties and behavior of such plasmas constitutes the science area called high-energy density laboratory plasmas (HEDLP). This rapidly emerging science area is extremely rich in basic science phenomena as well as potential applications such as inertial fusion energy science—one possible approach towards producing a clean and sustainable energy supply. A recent U.S. Department of Energy (DOE) Office of Science and National Nuclear Security Administration panel (Advancing the Science of High-Energy Density Laboratory Plasmas) produced compelling science opportunities in basic HEDLP, issues for inertial fusion energy science, and related opportunities for advanced computing to make a major impact. Many of these opportunities include processes that demand fully kinetic models involving multiscale science issues spanning micro- to meso-time and space scales. As an illustration, millimeter-scale pellets of deuterium and tritium in some IFES experiments can be compressed to 1000 times solid density over nanosecond-time scales, and lasers with wavelengths of microns or smaller can propagate through centimeter-scale plasmas.

Basic Plasma Science/Magnetic Reconnection Physics

The liberation of magnetic field energy through the process of magnetic reconnection is at the core of a diverse range of plasma phenomena including solar flares, geomagnetic substorms, sawteeth oscillations and disruptions in tokamaks, extragalactic jets, and a wide variety of astrophysical settings. In the past decade, most of the theoretical and simulation efforts have been directed at relatively small two-dimensional systems using both fluid and kinetic descriptions. Presently, it remains unclear how these idealized results will extend to large-scale three-dimensional systems. Even with extreme scale computing, a first-principles, three-dimensional kinetic treatment of reconnection in hydrogen plasmas will be limited to fairly small systems. Progress in modeling realistic applications will require understanding the key physics sufficiently well to be able capture them within reduced descriptions and to infer reliable scaling.

With regard to advances in magnetic fusion energy science, the scientific and computational advancements resulting from research efforts detailed in the "Basic Plasma/Magnetic Reconnection Physics" portion of the panel report can impact progress in several ways. First, this research will help clarify the essential physics needed to properly model reconnection in fusion-relevant plasmas and incorporate these physics into reduced fluid models. This is important because the realization of high-performance regimes with superior energy confinement in fusion plasmas—such as the ITER—require their operation in a stable, quasi-steady state in which the size and dynamics of magnetic islands are controlled by manipulating the background current and pressure profiles. For example, sawtooth crashes, which represent an important paradigm for fast reconnection in tokamak plasmas, can trigger the formation of neoclassical tearing mode islands to produce major disruptions. Major disruptions can also occur in tokamak plasmas due to the coupling of tearing islands on multiple rational surfaces that can modify the background current profile and trigger kink modes that can potentially terminate a discharge. Understanding the behavior of these magnetic islands, and resolving the separate physics of ions and electrons, is critical to controlling them. The computational challenge of predicting the time evolution in realistic toroidal geometry, while resolving ion and electron dynamics within, as well as outside, of the islands will require extreme-scale computing resources. In this regard, the computational and algorithmic advances needed to make progress in reconnection physics may directly benefit a wide range of problems in fusion energy sciences.

In general, magnetic reconnection remains one of the most fundamental and widespread processes in basic plasma physics. Many theoretical and computational challenges arise from the immense separation of spatial and temporal scales that result from coupling nonideal diffusion regions to the larger-scale dynamics. Over the past 50 years, progress in reconnection research has benefited greatly from numerical simulations. This trend is accelerating with the advent of petascale computers and is expected to continue as exascale computers become available in the next decade.

ADVANCED SCIENTIFIC COMPUTING RESEARCH CROSSCUTTING CHALLENGES

DOE's Office of Advanced Scientific Computing Research (ASCR) crosscutting challenges, which impact the five DOE Office of Fusion Energy Sciences grand challenge areas, were identified by members of the four ASCR panels during the workshop. These findings are summarized in the following paragraphs.

Algorithms for Fusion Energy Sciences at Extreme Scale

This panel report covers the key challenges of scalable algorithms for scaling simulations in fusion energy science to the expanding architectural extremes of the coming decade. These include increasing efforts to resolve the full ranges of length and/or time scales in a model; accommodating physical effects with greater fidelity; allowing the model degrees of freedom in all relevant dimensions; optimizing or controlling plasma scenarios (inverse problem) that are adequately predicted by forward models; and quantifying uncertainty. However, as applications broaden to take full advantage of extreme architectures, the complexity of algorithms may grow superlinearly in problem size, making it impossible to weak scale, even though memory capacity would seem to allow it. Extreme scales put a premium on finding "optimal" algorithms, whose complexity is at worst log-linear in problem size because any suboptimal component will ultimately dominate the execution profile. The availability of high-capability architectures makes algorithms more—not less—important. Fortunately, algorithms have kept pace with extreme scales, and optimal versions are known for systems arising from some popular formulations of the plasma physics.

Data Analysis, Management, and Visualization in Fusion Energy Sciences

This panel report covers five main areas related to data management, analysis, and visualization of fusion data, and describes the state of the art in terms of requirements, and projections of the effects of extreme-scale computing in these areas. These areas are as follows:

- managing large-scale input/output volume and data movement
- real-time simulation monitoring and run-time metadata generation
- data analysis at extreme scale
- visualization of very large datasets
- experiment-simulation data comparison.

Summarized below are the conclusions in each of these areas and projections at the extreme scale.

Managing Large-scale Input/Output Volume and Data Movement

As system compute capabilities continue to scale toward the extreme scale, storage systems must adapt to address the increasing bandwidth and storage challenges. In many simulation codes, the volume of data generated per core per time step is in the order of 2 gigabytes. Thus, the size of dataset produced by a single simulation on a 100,000-core machine is already about 200 terabytes. At this extreme scale, input/output (I/O) storage and its use may have to focus on new approaches that consider architectures, the role of I/O systems, data formats, and I/O and storage systems as a vehicle for knowledge discovery. Three areas will have to have to be greatly enhanced to prevent I/O from becoming the main bottleneck that will slow down extreme-scale computations: 1) the file system associated with the extreme-scale computing facility; 2) providing efficient high-level I/O libraries; and 3) high-level application programming interfaces to support multiscale models that can achieve the desired application performance. Having a data center where midscale computational and extreme-scale storage resources are co-located may address some of the challenges that arise when PB datasets are generated at multiple locations.

Real-Time Simulation Monitoring and Run-Time Metadata Generation

Real-time simulation monitoring is essential to fusion simulations; for example, to stop simulations that do not proceed as expected, to prevent wasting precious computation resources, or to allow instant generation of tiered metadata about the computations and the results. Furthermore, the ability to couple multiple codes in real time will become essential to take advantage of the computational power and to increase predictive fidelity of the more complex codes that are emerging. Provenance capture becomes another critical piece during the data-generation phase because scientists need to be able to link their data from the original simulation code to the analysis pieces, and then through complex analysis and visualization. Tools that will automate the monitoring, provenance collection, and code-coupling will be essential as the volume of data grows in the extreme scale, especially when modeling and running simulations that involve multiscale physics.

Data Analysis at Extreme Scale

The data analysis challenges in fusion energy science applications at the extreme scale stem not only from the large size of the data, but also the complexity of the data. Several tasks that support extreme-scale simulations—such as code validation and theory refinement—involve small-to-moderate sized data. Analysis of these data can be difficult as the task may be poorly defined, the science not well understood, and the data quite complex in the form of multiscale, noisy, time-varying images, or points in three-dimensional space.

For problems where the data sizes approach terabytes and beyond, there is an added dimension to the challenges faced in the analysis. Specifically, analysis algorithms must not only handle all the issues arising from the complexity of the data, but must do so for data that will be distributed across many files and for analysis that may require a fast turnaround to keep pace with experiments. Techniques for data reduction that preserve the essential features of the data will have to be developed to allow for timely analysis capabilities. Furthermore, when code is parallelized, it partitions the problem on multiple cores, thus requiring the coordination of calculations across the boundaries of the partitioning. This can cause inaccuracies. Consequently, current analysis methods will need to be parallelized to the extent possible without compromising their accuracy.

To successfully solve these analysis problems, current techniques from image and video processing, machine learning, statistics, and pattern recognition must be enhanced. In addition, new approaches must be developed that are more robust and can handle the diversity of data types, the variations within a data set, the distributed nature of the data, and any physics-driven challenges to the analysis.

Visualization of Very Large Datasets

There are three fundamental challenges for visualizing extreme data volumes: the need to effectively use remote distributed resources, dissemination of results, and algorithms to extract salient features. These challenges are interrelated and require close collaboration between those designing visualization systems and fusion scientists who will use these systems. As more visualization tools use a remote client architecture, creating collaborative environments will be easier to implement as part of their infrastructure. However, the challenge is compatibility across multiple architectures. As extreme-scale computing becomes more prevalent, a variety of diverse architectures will appear. It will be a great challenge for these tools to operate and perform well on these architectures as they rely more heavily on I/O and less heavily on central processing unit usage, which is the opposite of simulation codes. Visualization of large-scale data rely on tools that can quickly search through data and correlate the results from multiple data fields. At the extreme scale, it will be necessary to have even more powerful query tools that can run on parallel platforms, search multiple types of data, and generate visual presentations that help fusion scientists understand the salient features of the data.

Experiment-Simulation Data Comparison

Advances in plasma diagnostics, combined with the advances in computational models and available computing power, have created new opportunities for simulation validation. These advancing trends will continue to accelerate, and at the extreme-scale computing level, will represent a significantly more demanding challenge for model validation. The computer science community can greatly assist the fusion energy science community by creating cross-platform reusable, interoperable, and scalable components for common elements required to perform comparison of experimental and simulation data, by automating some workflows in this process, and by introducing new software engineering practices for developing commonly available validation software.

Mathematical Formulations

This panel identified the challenging areas of applied mathematics research in each of the five physics areas for computing at extreme scales in fusion. The dominant themes centered around kinetic (and gyrokinetic) simulations as well as integrated and hybrid simulations. For each of the five physics areas, the applied mathematics research areas are as follows:

- Burning plasma/ITER: high-fidelity kinetic simulations, more robust gyrokinetic approximations, new algorithms for particle-in-cell, error analysis of gyrokinetic particle-in-cell methods, and treatment of kinetic electrons
- Integrated modeling: systematic frameworks for coupling, convergence and accuracy properties of coupled systems, and fully implicit methods
- Plasma-material interaction: coupling of edge plasma physics code with atomic processes codes
- Laser-plasma interaction: homogenization over multiple speckles and development of mesoscale and reduced models to address filamentation and parametric instabilities
- Magnetic reconnection: development of hybrid (kinetic-fluid) algorithms, treatment of stiff electrons, and adaptive mesh and algorithmic refinement.

Several underlying mathematical techniques were identified, including adaptive mesh refinement, implicit methods, and robust coupling of codes. These techniques crosscut several physics areas, and they merit further research and resources.

Programming Models, Frameworks, and Tools

Computer architectures are now being designed that take scientific computing beyond the current petascale regime into the exascale regime. Advances in fusion science will be enabled by such new computational capabilities. This transition provides for the most fundamental aspects of the code development process in the programming model, which describes the way researchers think about the machine controlled by researcher-developed programs. Progress in the PRDs in computer science, as described in the panel report, "Programming Models, Frameworks, and Tools," will be necessary to enable fusion science to exploit the most advanced computational tools of the 21st century.

REFERENCES

Abbasi H, M Wolf, G Eisenhauer, S Klasky, K Schwan, and F Zheng. 2009. "DataStager: Scalable Data Staging Services for Petascale Applications." In *High Performance Distributed Computing (HPDC), Proceedings of the 18th ACM International Symposium on High Performance Distributed Computing*, pp. 39-48. Garching, Germany.

Adams MF, S Ku, P Worley, E D'Azevedo, J Cummings, and CS Chang. 2009. "Scaling to 150K Cores: Recent Algorithm and Performance Engineering Developments Enabling XGC1 to Run at Scale." *Journal of Physics: Conference Series* 180(1).

Advanced Micro Devices (AMD). 2009 "AMD Launches the ATI Radeon HD 4770, World's First Desktop Graphics Card Utilizing Industry-Leading 40nm Process Technology." Accessed October 20, 2009, at http://www.amd.com/us-en/Corporate/VirtualPressRoom/0,,51_104_543~131125,00.html.

Alam SR and JS Vetter. 2006. "Hierarchical Model Validation of Symbolic Performance Models of Scientific Applications." In *Euro-Par 2006 Parallel Processing, Lecture Notes in Computer Science* 4128:65-77. Accessed October 20, 2009, at http://www.springerlink.com/content/r34401x051102372/.

Altintas I, E Jaeger, C Berkley, M Jones, B Ludaescher, and S Mock. 2004. "Kepler: "An Extensible System for Design and Execution of Scientific Workflows." In *The 16th International Conference on Scientific and Statistical Database Management (SSDBM)*, June 21-23, Santorini Island, Greece.

Aschwanden MJ. 2002. "Particle Acceleration and Kinematics in Solar Flares - A Synthesis of Recent Observations and Theoretical Concepts." *Space Science Reviews* 101(1-2):1-227.

Atzeni S. 1999. "Inertial Fusion Fast Ignitor: Igniting Pulse Parameter Window vs. the Penetration Depth of the Heating Particles and the Density of the Precompressed Fuel." *Physics of Plasmas* 6(8):3316:3326.

Aydemir AY. 1992. "Nonlinear Studies of m=1 Modes in High-Temperature Plasmas." *Physics of Fluids B* 4(11):3469-3472.

Baker DN and EC Stone. 1976. "Energetic Electron Anisotropies in Magnetotail Identification of Open and Closed Field Lines." *Geophysics Research Letters* 3(9):557-560.

Barbato E and A Saveliev. 2004. "Absorption of Lower Hybrid Wave Power by α-Particles in ITER-FEAT Scenarios." *Plasma Physics and Controlled Fusion* 46(8):1283-1297.

Barreto R, S Klasky, P Mouallem, N Podhorszki, and M Vouk. 2009. "Collaboration Portal for Petascale Studies." In *The International Symposium on Collaborative Technologies and Systems (CTS)*, May 18-22, Baltimore, Maryland.

Batchelor D, E D'Azevedo, G Bateman, DE Bernholdt, LA Berry, PT Bonoli, R Bramley, J Breslau, M Chance, J Chen, M Choi, W Elwasif, G-Y Fu, RW Harvey, WA Houlberg, EF Jaeger, SC Jardin, D Keyes, S Klasky, S Kruger, LP Lu, D McCune, J Ramos, DP Schissel, D Schnack, and JC Wright. 2007. "Integrated Physics Advances in Simulation of Wave Interactions with Extended MHD Phenomena." *Journal of Physics: Conference Series* 78: Art. No. 012003. Accessed October 20, 2009, at http://www.iop.org/EJ/abstract/1742-6596/78/1/012003.

Bécoulet A. 2009. *Advanced Physics Integration Challenges in the Fusion Energy Sciences*. Accessed October 20, 2009, at http://extremecomputing.labworks.org/fusion/plenaries/becoulet.pdf.

Bell AR, APL Robinson, M Sherlock, RJ Kingham, and W Rozmus. 2006. "Fast Electron Transport in Laser-Produced Plasmas and the KALOS Code for Solution of the Vlasov–Fokker–Planck Equation." *Plasma Physics and Controlled Fusion* 48:R37–R57.

Bent J, G Gibson, G Grider, B McClelland, P Nowoczynski, J Nunez, M Polte, and M Wingate. 2009. "PLFS: A Checkpoint Filesystem for Parallel Applications." Los Alamos National Laboratory (LANL) Technical Release LA-UR 09-02117, Los Alamos National Laboratory, Los Alamos, New Mexico.

Berger M and J Oliger. 1984. "Adaptive Mesh Refinement for Hyperbolic Partial Differential Equations." *Journal of Computational Physics* 484–512.

Berger MJ. 1985. "Stability of Interfaces with Mesh Refinement." *Mathematics of Computation* 45:301-318.

Betti R, CD Zhou, KS Anderson, LJ Perkins, W Theobald, and AA Solodov. 2007. "Shock Ignition of Thermonuclear Fuel with High Areal Density." *Physical Review Letters* 98: Art. No.155001.

Bhattacharjee A, K Germaschewski, amd CS Ng. 2005. "Current Singularities: Drivers of Impulsive Reconnection." *Physics of Plasmas* 12:042305.

Birdsall CK and AB Langdon. 1985. *Plasma Physics via Computer Simulation* McGraw-Hill Inc., New York.

Birn J, JF Drake, MA Shay, BN Rogers, RE Denton, M Hesse, M Kuznetsova, ZW Ma, A Bhattacharjee, A Otto, and PL Pritchett. 2001. "Geospace Environmental Modeling (GEM) Magnetic Reconnection Challenge." *Journal of Geophysical Research* 106:3715.

Biskamp D. 1986. "Magnetic Reconnection via Current Sheets." *Physics of Fluids* 29:1520. DOI:10.1063/1.865670.

Biskamp D. 2000. *Magnetic Reconnection in Plasmas*. Cambridge University Press, Cambridge, New York.

Bonnin X, AS Kukushkin, and DP Coster. 2009. "Code Development for ITER Edge Modelling−SOLPS5.1." *Journal of Nuclear Materials* 390-391:274.

Bowers KJ, BJ Albright L Yin B Bergen, and TJT Kwan. 2008a. "Ultrahigh Performance Three-Dimensional Electromagnetic Relativistic Kinetic Plasma Simulation." *Physics of Plasmas* 15(5): 055703.

Bowers KJ, BJ Albright, B Bergen L Yin, KJ Barker, and DJ Kerbyson. 2008b. "0.374 pflop/s Trillion-Particle Kinetic Modeling of Laser Plasma Interaction on Roadrunner." In *SC'08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*, pp 1-11. IEEE Press, New Jersey.

Bowers KJ, BJ Albright, L Yin, W Daughton, V Roytershteyn, B Bergen, and TJT Kwan. 2009. "Advances in Petascale Kinetic Plasma Simulation with VPIC and Roadrunner." *Journal of Physics: Conference Series* 180 (2009) 012055.

Brennan DP, RJ La Haye, AD Turnbull, MS Chu, TH Jensen, LL Lao, TC Luce, PA Politzer, EJ Strait, SE Kruger, and DD Schnack. 2003. "A Mechanism for Tearing Onset Near Ideal Stability Boundaries." *Physics of Plasmas* 10(5):1643-1652.

Brooks JN. 2002. "Modeling of Sputtering Erosion/Redeposition−Status and Implications for Fusion Design." *Fusion Engineering and Design* 60(4):515-526.

Brooks JN, JP Allain, RP Doerner, A Hassanein, R Nygren, TD Rognlien, and D Whyte. 2009. "Plasma-Surface Interaction Issues of an All-Metal ITER." *Nuclear Fusion* 49(3): Art. No. 035007. DOI:10.1088/0029-5515/49/3/035007.

Brown MR, CD Cothran, and J Fung. 2006. "Two-Fluid Effects on Three-Dimensional Reconnection in the Swarthmore Spheromak Experiment with Comparisons to Space Data." *Physics of Plasmas* 13(5):056503.

Candy J and RE Waltz. 2003. "Anomalous Transport in the DIII-D Tokamak Matched by Supercomputer Simulations." *Physical Review Letters* 91(4):045001-1-045001-4.

Cary J, J Candy, RH Cohen, S Krasheninnikov, DC McCune, D J Estep, J Larson, AD Malony, PH Worley, JA Carlsson, AH Hakim, P Hamill, S Kruger, S Muzsala, A Pletzer, S Shasharina, D Wade-Stein, N Wang, L McInnes, T Wildey, T Casper, L Diachin, T Epperly, TD Rognlien, MR Fahey, JA Kuehn, A Morris, S Shende, E Feibush, GW Hammett, K Indireshkumar, C Ludescher, L Randerson, D Stotler, A Yu Pigarov, P Bonoli, CS Chang, DA D'Ippolito, P Colella, DE Keyes, R Bramley, and JR Myra. 2007. "Introducing FACETS, the Framework Application for Code-Edge Transport Simulations." *Journal of Physics: Conference Series* 28: Art. No. 012086.

Cary JR, J Candy, J Cobb, RH Cohen, T Epperly, DJ Estep, S Krasheninnikov, AD Malony, DC McCune, L McInnes, A Pankin, S Balay, JA Carlsson, MR Fahey, RJ Groebner, AH Hakim, SE Kruger, M Miah, A Pletzer, S Shasharina, S Vadlamani, D Wade-Stein, TD Rognlien, A Morris, S Shende, GW Hammett, K Indireshkumar, AYu Pigarov, and H Zhang. 2009. "Concurrent, Parallel, Multiphysics Coupling in the FACETS Project." In *Proceedings of the Scientific Discovery for Advanced Computation*, San Diego, California.

Cassak PA, MA Shay, and JF Drake. 2005. "Catastrophe Model for Fast Magnetic Reconnection Onset." *Physical Review Letters* 95:235002.

Cedilnik A, B Geveci, K Moreland, J Ahrens, and J Favre. 2006. "Remote Large Data Visualization in the ParaView Framework." In *Eurographis Parallel Graphics and Visualization 2006*, pp. 163-170.

Chang CS, S Klasky, J Cummings, R Samtaney, A Shoshani, L Sugiyama, D Keyes, S Ku, G Park, S Parker, N Podhorszki, H Strauss, H Abbasi, M Adams, R Barreto, G Bateman, K Bennett, Y Chen, E D' Azevedo, C Docan, S Ethier, E Feibush, L Greengard, T Hahm, F Hinton, C Jin, A Khan, A Kritz, P Krsti, T Lao, W Lee, Z Lin, J Lofstead, P Mouallem, M Nagappan, A Pankin, M Parashar, M Pindzola, C Reinhold, D Schultz, K Schwan, D Silver, A Sim, D Stotler, M Vouk, M Wolf, H Weitzner, P Worley, Y Xiao, E Yooon, and D Zorin. 2008. "Toward a First-Principles Integrated Simulation of Tokamak Edge Plasmas." *Journal of Physics: Conference Series* 125: Art. No. 012042.

Chang CS, S Ku, PH Diamond, Z Lin, S Parker, TS Hahm, and N Samatova. 2009. "Compressed Ion Temperature Gradient Turbulence in Diverted Tokamak Edge." *Physics of Plasmas* 16: Art. No. 056108.

Chankin AV, DP Coster, N Asakura, G Corrigan, SK Eents, W Fundamenski, HW Muller, RA Pitts, PC Strangeby, and M Wischmeier. 2007. "A Possible Role of Radial Electric Field in Driving Parallel Ion Flow in Scrape-Off Layer of Divertor Tokamaks." *Nuclear Fusion* 47(8):762-772.

Chen LJ, A Bhattacharjee, PA Puhl-Quinn, H Yang, N Bessho, S Imada, S Muehlbachler, PW Daly, B Lefebvre, Y Khotyaintsev, A Vaivads, A Fazakerley, and E Georgescu. 2008. "Observation of Energetic Electrons Within Magnetic Islands." *Nature Physics* 4(1):19-23.

Chen LJ, N Bessho, B Lefebvre, H Vaith, A Asnes, O Santolik, A Fazakerley, P Puhl-Quinn, A Bhattacharjee, Y Khotyaintsev, P Daly, and R Torbert. 2009. "Multispacecraft Observations of the Electron Current Sheet, Neighboring Magnetic Islands, and Electron Acceleration During Magnetotail Reconnection." *Physics of Plasmas* 16: Art. No. 056501. DOI: 10.1063/1.3112744.

Childs H, E Brugger, K Bonnell, J Meredith, M Miller, B Whitlock, and N Max. 2005. "A Contract Based System for Large Data Visualization." In *Proceedings of the Conference on Visualization '05, Minneapolis, Minnesota*. Accessed October 20, 2009, at http://doi.ieeecomputersociety.org/10.1109/VIS.2005.3.

Choi M, VS Chan, RI Pinsker, CC Petty, SC Chiu, J Wright, P Bonoli, and M Porkolab. 2006. "Simulation of Fast Alfvén Wave Interaction with beam Ions Over a Range of Cyclotron Harmonics in DIII-D Tokamak." *Nuclear Fusion* 46(7):S409-S415.

Clark JH. 1976. "Hierarchical Geometric Models for Visible Surface Algorithms." *Communications of the ACM* 19(10):547-554.

Cohen BI, AJ Kemp, and L Divol. 2010. "Simulation of Laser-Plasma Interactions and Fast-Electron Transport in Inhomogeneous Plasma." *Journal of Computational Physics* 229(12): 4591-4612.

Cohen RH and XQ Xu. 2008. "Progress in Kinetic Simulation of Edge Plasmas." *Contributions to Plasma Physics* 48:212-223.

Collins N, G Theurich, C DeLuca, M Suarez, A Trayanov, V Balaji, P Li, W Yang, C Hill, and A da Silva. 2005. "Design and Implementation of Components in the Earth System Modeling Framework." *International Journal of High-Performance Computing Applications* 19(3):341-350. Accessed October 20, 2009, at http://hpc.sagepub.com/cgi/content/abstract/19/3/341.

Coster DP, X Bonnin, A Mutzke, R Schneider, and M Warrier. 2007. "Integrated Modeling of the Edge Plasma and Plasma Facing Components." *Journal of Nuclear Materials* 363-365:136-139.

Cummings JN and S Kiesler. 2005. "Collaborative Research Across Disciplinary and Organizational Boundaries." *Social Studies of Science* 35(5):703–722.

Cummings J, A Pankin, N Podhorszki, G Park, S Ku, R Barreto, S Klasky, CS Chang, H Strauss, L Sugiuyama, P Snyder, D Pearlstein, B Ludäscher, G Bateman, and A Kritz. 2008. "Plasma Edge Kinetic-MHD Modeling in Tokamaks Using Kepler Workflow for Code Coupling, Data Management and Visualization." *Communications in Computational Physics* 4(3):675-702.

D'Ippolito DA and JR Myra. 2006. "A Radio-Frequency Sheath Boundary Condition and Its Effect on Slow Wave Propagation." *Physics of Plasmas* 13(10): Art. No.: 102508.

Daughton W, J Scudder, and H Karimabadi. 2006. "Fully Kinetic Simulations of Undriven Magnetic Reconnection with Open Boundary Conditions." *Physics of Plasmas* 13:072101.

Daughton W, V Roytershteyn, BJ Albright, H Karimabadi, L Yin, and KJ Bowers. 2009. "Influence of Coulomb Collisions on the Structure of Reconnection Layers." *Physics of Plasmas* 16(7):072117-01−072117-16.

Davidson RC. 2007. "Editorial: Announcement of Editorial Policy Statement on Verification and Validation." *Physics of Plasmas* 14: Art. No. 050401.

Dawson JM. 1983. "Particle Simulation of Plasmas." *Reviews of Modern Physics* 55:403-447.

Decyk VK. 2007. "UPIC: A Framework for Massively Parallel Particle-in-Cell Codes." *Computer Physics Communications* 177(1-2):95-97. Accessed October 20, 2009, at http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6TJ5-4N5CXJM-J&_user=2741876&_rdoc=1&_fmt=&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1056887500&_rerunOrigin=google&_acct=C000058656&_version=1&_urlVersion=0&_userid=2741876&md5=5c4764dee4a4e5d71047ab2f354a2009.

Deng XH and H Matsumoto. 2001. "Rapid Magnetic Reconnection in the Earth's Magnetosphere Generated by Whistler Waves." *Nature* 410:557-559.

DOE (U.S. Department of Energy). 2008. *Fusion Simulation Project Workshop Report*. "Fusion Simulation Project Workshop," May 16-18, 2007, Rockville, Maryland. Chairs: A Kritz, Lehigh University, Pennsylvania, and D Keyes, Columbia University, New York. Report prepared by Fusion Simulation Project Workshop members for the DOE Office of Science, Washington, D.C. Accessed October 20, 2009, at http://www.science.doe.gov/ofes/programdocuments/reports/FSPWorkshopReport.pdf.

Dongarra J. 2009. *Top 500 Supercomputer Sites*. Accessed October 20, 2009, at http://www.top500.org.

Dorfman S, W Daughton, V Roytershteyn, H Ji, Y Ren, and M Yamada. 2008. "Two-Dimensional Fully Kinetic Simulations of Driven Magnetic Reconnection with Boundary Conditions Relevant to the Magnetic Reconnection Experiment." *Physics of Plasmas* 15(10):102107-1−102107-14. DOI:10.1063/1.2991361.

dos Santos S and K Brodlie. 2004. "Gaining Understanding of Multivariate and Multidimensional Data Through Visualization." *Computers and Graphics* 28 (3).

Doyle EJ, WA Houlberg, Y Kamada, V Mukhovatov, TH Osborne, A Polevoi, G Bateman, JW Connor, JG Cordey, T Fujita, X Garbet, TS Hahm, LD Horton, AE Hubbard, F Imbeaux, F Jenko, JE Kinsey, Y Kishimoto, J Li, TC Luce, Y Martin, M Ossipenko, V Parail, A Peeters, TL Rhodes, JE Rice, CM Roach, V Rozhansky, F Ryter, G Saibene, R Sartori, ACC Sips, JA Snipes, M Sugihara, EJ Synakowski, H Takenaga, T Takizuka, K Thomsen, MR Wade, and HR Wilson. 2007. "Chapter 2: Plasma Confinement and Transport." ITPA Transport Physics Topical Group, ITPA Confinement Database and Modeling Topical Group and ITPA Pedestal and Edge Topical Group. *Nuclear Fusion* 47(6):S18-S27.

DP Stotler. 2008. "Validation in Fusion Research: Towards Guidelines and Best Practices." *Phys. Plasmas* 15: 062503.

Drake JF, M Swisdak, H Che, and MA Shay. 2006a. "Electron Acceleration from Contracting Magnetic Islands During Reconnection." *Nature* 443(7111):553-556.

Drake JF, M Swisdak, KM Schoeffler, BN Rogers, and S Kobayashi. 2006b. "Formation of Secondary Islands During Magnetic Reconnection." *Geophysics Research Letters* 33: Art. No. L13105. DOI:10.1029/2006GL025957.

Duclous R, B Dubroca, F Filbet, and V Tikhonchuk. 2009. "High-Order Resolution of the Maxwell-Fokker-Planck-Landau Model Intended for ICF Applications." *Journal of Computational Physics* 228(14):5072-5100. Accessed October 20, 2009, at http://portal.acm.org/citation.cfm?id=1552584.1553083.

Eastwood, JP, TD Phan, FS Mozer, MA Shay, M Fujimoto, A Retinò, M Hesse, A Balogh, EA Lucek, and I Dandouras. 2007. "Multi-Point Observations of the Hall Electromagnetic Field and Secondary Island Formation During Magnetic Reconnection." *Journal of Geophysical Research-Space Physics*, 112: Art. No. A06235. DOI:10.1029/2006JA012158.

Eckstrand S. 2009. "Fusion Energy Sciences Program." Presented at the *Scientific Grand Challenges in Fusion Energy Sciences and the Role of Computing at the Extreme Scale* workshop, March 18-20, Gaithersburg, Maryland. U.S. Department of Energy's Office of Fusion Energy Sciences, Washington, D.C. Accessed October 20, 2009, at http://www.ofes.fusion.doe.gov/more_html/FESAC/ProgPrioritiesReport.pdf.

Egedal J, W Fox, N Katz, M Porkolab, K Reim, and E Zhang. 2006. "Laboratory Observations of Spontaneous Magnetic Reconnection." *Physical Review Letters* 98(1):015003.

Fahlen JE, BJ Winjum, T Grismayer, and WB Mori. 2009. "Propagation and Damping of Nonlinear Plasma Wave Packets." *Physical Review Letters* 102(24): Art. No. 245002. DOI: 10.1103/PhysRevLett.102.245002. Accessed October 20, 2009, at http://adsabs.harvard.edu/abs/2009PhRvL.102x5002F.

Fasoli A, C Gormenzano, HL Berk, B Breizman, S Briguglio, DS Darrow, N Gorelenkov, WW Heidbrink, A Jaun, SV Konovalov, R. Nazikian, J-M Noterdaeme, S Sharapov, K Shinohara, D Testa, K Tobita, Y Todo, G Vlad, and F Zonca. 2007. "Chapter 5: Physics of Energetic Ions." *Nuclear Fusion* 47(6):S264-S284.

Feibush E, D McCune, R Nazikian, G Kramer, and E Valeo. 2006. "Web Interfaces to Fusion Codes." In *The 48th Annual Meeting of the Division of Plasma Physics*, October 30-November 3, 2006, Philadelphia, Pennsylvania.

Flannery D, B Matthews, T Griffin, J Bicarregui, M Gleaves. L Lerusse, R Downing, A Ashton, S Sufi, G Drinkwater, and K Kleese. 2009. "ICAT: "Integrating Data Infrastructure for Facilities-Based Science." In *Proceedings of the 5th IEEE International Conference on e-Science (e-science 2009)*, December 9-11, 2009, Oxford, United Kingdom.

Fonseca RA, JT Mendonca, AL Brinca, J Loureiro, LO Silva, and I Viieira. 2002. "Physical Problems of Artificial Magnetospheric Propulsion." In *Lecture Notes in Computer Science* Vol. 29, III-342, Springer-Verlag, Heidelberg, Germany.

Fu G-Y, W Park, HR Strauss, J Breslau, J Chen, S Jardin, and LE Sugiyama. 2006. "Global Hybrid Simulations of Energetic Particle Effects on the n=1 Mode in Tokamaks: Internal Kink and Fishbone Instability." *Physics of Plasmas* 13(5):052517-1–052517-10. DOI:10.1063/1.2203604.

Fusion Energy Science Advisory Committee (FESAC). 2009. *U.S. Department of Energy Small Business Innovation Research and Small Business Technology Transfer Programs: Technical Topic Descriptions*. Prepared by the Fusion Energy Science Advisory Committee for the U.S. Department of Energy, Washington, D.C.

Gangadhara S, DA Ennis, DJ den Hartog, G Fiksel, and SC Prager. 2008. "Ion Heating During Reconnection in the Madison Symmetric Torus Reversed Field Pinch." *Physics of Plasmas* 15(5):056121.1-056121.9.

Gelernter D. 1992. "Current Research on Linda." In *Research Directions in High-Level Parallel Programming Languages*, *Lecture Notes in Computer Science* 574:73-76. Springer, Berlin.

Gill RD. 1993. "Generation and Loss of Runaway Electrons Following Disruptions in JET." *Nuclear Fusion* 33(11):1613-1625.

Grasso D, F Pegoraro, F Porcelli, and F Califano. 1999. "Hamiltonian Hagnetic Reconnection." *Plasma Physics and Controlled Fusion* 41(12):1497-1515. DOI:10.1088/0741-3335/41/12/306.

Hassanein A. 2002. "Prediction of Material Erosion and Lifetime During Major Plasma Instabilities in Tokamak Devices." *Fusion Engineering and Design* 60:527-546. DOI: 10.1016/S0920-3796(02)00008-X .

Hassanein A and I Konkashbaev. 2003. "Comprehensive Modeling of ELMS and Their Effect on Plasma-Facing Surfaces During Normal Tokamak Operation." *Journal of Nuclear Materials* 313-316:664-669. DOI:10.1016/S0022-3115(02)01376-4 .

Hassanein A, T Sizyuk, and I Konkashbaev. 2009. "Integrated Simulation of Plasma Surface Interaction During Edge Localized Modes and Disruptions: Self-Consistent Approach." *Journal of Nuclear Materials* 390-391:777-780.

Helander P, L-G Eriksson, and F Andersson. 2002. "Runaway Acceleration During Magnetic Reconnection in Tokamaks." *Plasma Physics and Controlled Fusion* 44(12B):B247-B262.

Hemker RG, FS Tsung, VK Decyk, WB Mori, S Lee, and T Katsouleas. 2000. "Development of a Parallel Code for Modeling Plasma Based Accelerators." In *Proceedings of the 1999 Particle Accelerator Conference*, 5:3672-3674, March 27-April 2, New York. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Hender TC, JC Wesley, J Bialek, A Bondeson, AH Boozer, RJ Buttery, A Garofalo, TP Goodman, RS Granetz, Y Gribov, O Gruber, M Gryaznevich, G Giruzzi, S Günter, N Hayashi, P Helander, CC Hegna, DF Howell, DA Humphreys, GTA Huysmans, AW Hyatt, A Isayama, SC Jardin, Y Kawano, A Kellman2, C Kessel, HR Koslowski, RJ La Haye, E Lazzaro, YQ Liu, V Lukash, J Manickam, S Medvedev, V Mertens, SV Mirnov, Y Nakamura, G Navratil, M Okabayashi, T Ozeki, R Paccagnella, G Pautasso, F Porcelli, VD Pustovitov, V Riccardo, M Sato, O Sauter, MJ Schaffer, M Shimada, P Sonato, EJ Strait, M Sugihara, M Takechi, AD Turnbull, E Westerhof, DG Whyte, R Yoshino, and H Zohm. 2007. "Chapter 3: MHD Stability, Operational Limits and Disruptions." The ITPA MHD, Disruption and Magnetic Control Topical Group. *Nuclear Fusion* 47(6):S128-S202.

Henkelman G and H Jonsson. 2001. "Long Time Scale Kinetic Monte Carlo Simulations Without Lattice Approximation and Predefined Event Table." *Journal of Chemical Physics* 115(21):9657-9666. DOI:10.1063/1.1415500.

Hesse M, K Schindler, J Birn, and M Kuznetsova. 1999. "The Diffusion Region in Collisionless Magnetic Reconnection." *Physics of Plasmas* 6:1781.

Hesse M, M Kuznetsova, and J Birn. 2004. "Suprathermal Electron Acceleration in Magnetic Reconnection." *Physics of Plasmas* 11:5387.

Hinkel DE, DA Callahan, AB Langdon, SH Langer, CH Still, and EA Williams. 2008. "Analyses of Laser-Plasma Interactions in National Ignition Facility Ignition Targets." *Physics of Plasmas* 15: Art. No. 056314. DOI:10.1063/1.2901127.

Hockney RW and JW Eastwood. 1988. *Computer Simulation Using Particles*. Institute of Physics Publishing, Philadelphia, Pennsylvania.

Holland C, P Diamond, S Champeaux, E Kim, O Gurcan, M Rosenbluth, G Tynan, N Rocker, W Nevins, and J Candy. 2003. "Investigations of the Role of Nonlinear Couplings in Structure Formation and Transport Regulation: Experiment, Simulation, and Theory." *Nuclear Fusion* 43:761-780.

Holland C, J Candy, RE Waltz, AE White, GR McKee, MW Shafer, L Schmitz, and GR Tynan. 2008. "Validating Simulations of Core Tokamak Turbulence: Current Status and Future Directions." SciDAC 2008. *Journal of Physics: Conference Series* 125: Art. No. 012043.

Hoshino M, T Mukai, T Terasawa, and I Shinohara. 2001. "Suprathermal Electron Acceleration in Magnetic Reconnection." *Journal of Geophysical Research* 106(A11):25979-25997.

Hsu SC, G Fiksel, TA Carter, H Ji, RM Kulsrud, and M Yamada. 2000. "Local Measurement of Nonclassical Ion Heating During Magnetic Reconnection." *Physical Review Letters* 84(17):3859-3862.

Huck K, A Malony, R Bell, and A Morris. 2005. "Design and Implementation of a Parallel Performance Data Management Framework." In *Proceedings of the 2005 International Conference on Parallel Processing*, pp. 473-482. Accessed October 20, 2009, at http://portal.acm.org/citation.cfm?id=1079422.

Huck K, O Hernandez, V Bui, S Chandrasekaran, B Chapman, A Malony, LC McInnes, and B Norris. 2008a. "Capturing Performance Knowledge for Automated Analysis." In *Conference on High-Performance Networking and Computing*, Article No. 49. Accessed October 20, 2009, at http://portal.acm.org/citation.cfm?id=1413420.

Huck K, A Malony, S Shende, and A Morris. 2008b. "Knowledge Support and Automation for Performance Analysis with PerfExplorer 2.0." *The Journal of Scientific Programming* 16(2-3):123-134 (Special Issue on Large-Scale Programming Tools and Environments).

Iaccarino G, MS Eldred, A Doostan, and O Ghattas. 2009. "Uncertainty Quantification Minitutorial: Introduction to Uncertainty Quantification." *Society for Industrial and Applied Mathematics (SIAM)* Conference on Computational Science and Engineering, March 2-6, Miami, Florida. Accessed October 20, 2009, at http://www.stanford.edu/~jops/UQsiam09.html.

ITER 1999a. "Chapter 1: Overview and Summary." ITER Physics Expert Group Chairs and Co-Chairs and ITER Joint Central Team and Physics Integration Unit. *Nuclear Fusion* 39(12):2137-2174.

ITER 1999b. "Chapter 2: Plasma Confinement and Transport." ITER Physics Expert Group on Confinement and Transport, ITER Physics Expert Group on Confinement Modeling and Database and ITER Physics Basis Editors. *Nuclear Fusion* 39(12):2175-2249.

ITER 1999c. "Chapter 3: MHD Stability, Operational Limits and Disruptions." Physics Expert Group on Disruptions, Plasma Control, and MHD and ITER Physics Basis Editors. *Nuclear Fusion* 39(12):2251-2389.

ITER 1999d. "Chapter 5: Physics of Energetic Ions." Physics Expert Group on Energetic Particles, Heating and Current Drive and ITER Physics Basis Editors. *Nuclear Fusion* 39(12):2471-2495.

ITER 1999e. "Chapter 6: Plasma Auxiliary Heating and Current Drive." Physics Expert Group on Energetic Particles, Heating and Current Drive and ITER Physics Basis Editors. *Nuclear Fusion* 39(12):2495-2539.

Jaeger EF, RW Harvey, VE Lynch, N Ershov, LA Berry, PT Bonoli, V Tang, and RD Moore and the members of the SciDAC Center for Simulation of Wave-Plasma Interactions (CSWPI). 2006a. "Quasilinear Evolution of Non-Thermal Distributions in Ion Cyclotron Resonance Heating of Tokamak Plasmas." *Journal of Physics: Conference Series* 46:82-86.

Jaeger EF, RW Harvey, LA Berry, JR Myra, RJ Dumont, CK Phillips, DN Smithe, RF Barrett, DB Batchelor, PT Bonoli, MD Carter, EF D'Azevedo, DA D'Ippolito, RD Moore, and JC Wright. 2006b. "Global-Wave Solutions with Self-Consistent Velocity Distributions in Ion Cyclotron Heated Plasmas." *Nuclear Fusion* 46(7):S397-S408. DOI: 10.1088/0029-5515/46/7/S02.

Jaeger EF, LA Berry, EF D'Azevedo, RF Barrett, SD Ahern, D W Swain, DB Batchelor, RW Harvey, JR Myra, DA D'Ippolito, CK Phillips, E Valeo, DN Smithe, PT Bonoli, JC Wright, and M Choi. 2008. "Simulation of High-Power Electromagnetic Wave Heating in the ITER Burning Plasma." *Physics of Plasmas* 15(7): Art. No. 072513.

Jardin SC, GL Schmidt, ED Fredrickson, KW Hill, J Hyun, BJ Merrill, and R Sayer. 2000. "A Fast Shutdown Technique for Large Tokamaks." *Nuclear Fusion* 40(5)923-933.

Jaroschek CH, RA Treumann, H Lesch, and M Scholer. 2004. "Fast Reconnection in Relativistic Pair Plasmas: Analysis of Particle Acceleration in Self-Consistent Full Particle Simulations." *Physics of Plasmas* 11(3):1151-1163.

Jarvis ON, G Sadler, and JL Thompson. 1988. "Photoneutron Production Accompanying Plasma Disruptions in JET." *Nuclear Fusion* 28(11):1981-1993.

Ji H, Y Ren, M Yamada, S Dorfman, W Daughton, and SP Gerhardt. 2008. "New Insights into Dissipation in the Electron Layer During Magnetic Reconnection." *Geophysical Research Letters* 35(13): Art. No. L13106.

Jones C, K-L Ma, A Sanderson, and L Myers. 2007. "Visual Interrogation of Gyrokinetic Particle Simulations." *Journal of Physics: Conference Series* 78: Art. No. 012033. Accessed October 20, 2009, at http://www.iop.org/EJ/abstract/1742-6596/78/1/012033/.

Jones C, K-L Ma, S Ethier, and W Lee. 2008. "An Integrated Exploration Approach to Visualizing Multivariate Particle Data." *IEEE Computing in Science and Engineering* 10(4):20-29.

Karimabadi H, W Daughton, and J Scudder. 2007. "Multi-Scale Structure of the Electron: The Electron Diffusion Region." *Geophysical Research Letters* 34(13): Art. No. L13104.

Kemp AJ and H Ruhl. 2005. "Multispecies Ion Acceleration Off Laser-Irradiated Water Droplets." *Physics of Plasmas* 12: Art. No. 033105 (and references therein). DOI: 10.1063/1.1856933.

Kirk JG and O Skjaeraasen. 2003. "Dissipation in Poynting-Flux-Dominated Flows: The Sigma-Problem of the Crab Pulsar Wind." *The Astrophysical Journal* 591(1):366-379.

Kirkpatrick RC. 2007. "Fusion Reaction Product Transport for Magnetized Target Fusion." *Fusion Science and Technology* 52(4):1075-1078.

Kleva RG, JF Drake, and FL Waelbroeck. 1995. "Fast Reconnection in High-Temperature Plasmas." *Physics of Plasmas* 2(1):23-34. DOI:10.1063/1.871095.

Klimas A, M Hesse, and S Zenitani. 2008. "Particle-In-Cell Simulations of Collisionless Reconnection with Open Outflow Boundaries." *Physics of Plasmas* 15(8):082102-1−082102-9.

Kodama R, PA Norreys, K Mima, AE Dangor, RG Evans, H Fujita, Y Kitagawa, K Krushelnick, T Miyakoshi, N Miyanaga, T Norimatsu, SJ Rose, T Shozaki, K Shigemori, A Sunahara, M Pampo, KA Tanaka, Y Toyama, T Yamanaka, and M Zepf. 2001. "Fast Heating of Ultrahigh-Density Plasma as a Step Towards Laser Fusion Ignition." *Nature* 412:798-802.

Kodama R, H Shiraga, K Shigemori, Y Toyama, S Fujioka, H Azechi, H Fujita, H Habara, T Hall, Y Izawa, T Jitsuno, Y Kitagawa, KM Krushelnick, KL Lancaster, K Mima, K Nagai, M Nakai, H Nishimura, T Norimatsu, PA Norreys, S Sakabe, KA Tanaka, A Youssef, M Zepf, and T Yamanaka. 2002. "Nuclear Fusion: Fast Heating Scalable to Laser Fusion Ignition." *Nature* 418:933-934.

Kogge P, K Bergman, S Borkar, D Campbell, W Carlson, W Dally, M Denneau, P Franzon, W Harrod, K Hill, J Hiller, S Karp, S Keckler, D Klein, R Lucas, M Richards, A Scarpelli, S Scott, A Snavely, T Sterling, RS Williams, and K Yelick. 2008. *Exascale Computing Study: Technology Challenges in Achieving Exascale Systems.* DARPA Information Processing Techniques Office, Vol. 705. Accessed October 20, 2009, at http://users.ece.gatech.edu/mrichard/ExascaleComputingStudyReports/ECS_reports.htm.

Korkmaz P, BES Akgul, KV Palem, and LN Chakrapani. 2006. "Advocating Noise as an Agent for Ultra-Low Energy Computing: Probabilistic CMOS Devices and their Characteristics." *Japanese Journal of Applied Physics* 45(4B):3307-3316. Accessed October 20, 2009, at http://sciencelinks.jp/j-east/article/200614/000020061406A0421152.php.

Kritz A and D Keyes. 2007. *Fusion Simulation Project - Workshop Sponsored by the U.S. Department of Energy, Rockville, Maryland, May 16-18, 2007*. Accessed October 20, 2009, at http://www.ofes.fusion.doe.gov/ProgramDocuments/reports/FSPWorkshopReport.pdf.

Kruger SE, DD Schnack, and CR Sovinec. 2005. "Dynamics of the Major Disruption of a DIII-D Plasma." *Physics of Plasmas* 12(5):056113-1-056113-10. Accessed October 20, 2009, at http://adsabs.harvard.edu/abs/2005PhPl...12e6113K.

Kuritsyn A, H Ji, SP Gerhardt, Y Ren, and M Yamada. 2007. "Effects of Global Boundary and Local Collisionality on Magnetic Reconnection in a Laboratory Plasma." *Geophysical Research Letters* 34(16):L16105.

Lapenta G. 2008. "Self-Feeding Turbulent Reconnection on Macroscopic Scales." *Physical Review Letters* 100(23): Art. No. 235001.

Law CC, A Henderson, and J Ahrens. 2001. "An Application Architecture for Large Data Visualization: A Case Study." In *Proceedings of the IEEE 2001 Symposium on Parallel and Large-Data Visualization and Graphics (PVG '01)*, pp. 125-128, San Diego, California.

Lee WW. 1987. "Gyrokinetic Particle Simulation Model." *Journal of Computational Physics* 72(1):243-269. DOI: 10.1016/0021-9991(87)90080-5.

Lee WW, S Ethier, R Kolesnikov, WX Wang, and WM Tang. 2008. "Nonlinear Turbulent Transport in Magnetic Fusion Plasmas." *Computational Science & Discovery* 1(1): Art. No. 015010. DOI:10.1088/1749-4699/1/1/015010.

Li L and A Malony. 2007. "Knowledge Engineering for Automatic Parallel Performance Diagnosis." *Concurrency and Computation: Practice and Experience* 19(11):1497-1515.

Li G, R Yan, C Ren, T-L Wang. J Tonge, and WB Mori. 2008. "Laser Channeling in Millimeter-Scale Underdense Plasma of Fast-Ignition Targets." *Physical Review Letters* 100(12):125002.1-125002.44.

Lin RP and HS Hudson. 1971. "10-100 keV Electron Acceleration and Emission from Solar Flares." *Solar Physics* 17(2):412-435.

Lin RP, S Krucker, GJ Hurford, DM., Smith, HS Hudson, GD Holman, RA Schwartz, BR Dennis, GH Share, RJ Murphy, AG Emslie, C Johns-Krull, and N Vilmer. 2003. "RHESSI Observations of Particle Acceleration and Energy Release in an Intense Solar Gamma-Ray Line Flare." *The Astrophysical Journal* 595(2):L69-L76.

Lindemuth IR and MM Widner. 1981. "Magneto-Inertial Fusion: An Emerging Concept for Inertial Fusion." *Physics of Fluids* 24:753.

Lindemuth IR and Kirkpatrick RC. 1983. "Progress with Developing a Target for Magnetized Target Fusion." *Nuclear Fusion* 23:263.

Livnat Y and CD Hansen. 1998. "View Dependent Isosurface Extraction." In *IEEE Visualization '98*, pp. 175-180.

Lofstead J, F Zheng, S Klasky, and K Schwan. 2009. "Adaptable, Metadata-Rich IO Methods for Portable High Performance IO." In *IEEE International Parallel and Distributed Processing Symposium* (IPDPS).

Loureiro NF, AA Schekochihin, and SC Cowley. 2007. "Instability of Current Sheets and Formation of Plasmoid Chains." *Physics of Plasmas* 14(10): Art. No. 100703.

Luxon JL. 2002. "A Design Retrospective of the DIII-D Tokamak." *Nuclear Fusion* 42(5): 614-633. Accessed October 20, 2009, at http://www.iop.org/EJ/toc/0029-5515/42/5.

Ma ZW and A Bhattacharjee. 1996. "Fast Impulsive Reconnection and Current Sheet Intensification due to Electron Pressure Gradients in Semi-Collisional Plasmas." *Geophysical Research Letters* 23:1673.

Malara F, P Veltri, and V Carbone. 1992. "Competition Among Nonlinear Effects in Tearing Instability Saturation." *Physics of Fluids B* 4(10):3070-3086. DOI:10.1063/1.860477.

Mansfield DK, DW Johnson, B Grek, HW Kugel, MG Bell, RE Bell, RV Budny, CE Bush, ED Fredrickson, KW Hill, DL Jassby, RJ Maqueda RJ, HK Park, AT Ramsey, EJ Synakowski, G Taylor, and GA Wurden. 2001. "Observations Concerning the Injection of a Lithium Aerosol into the Edge of TFTR Discharges." *Nuclear Fusion* 41(12):1823-1834.

Marinak MM, GD Kerbel, NA Gentile, O Jones, D Munro, S Pollaine, TR Dittrich, and SW Haan. 2001. "Three-Dimensional HYDRA Simulations of National Ignition Facility Targets." *Physics of Plasma* 8:2275. DOI:10.10631/1.1356740.

Mazzucato E, RE Bell, S Ethier, JC Hosea, SM Kaye, BP LeBlanc, WW Lee, PM Ryan, DR Smith, WX Wang, JR Wilson, and H Yuh. 2009. "Study of Turbulent Fluctuations Driven by the Electron Temperature Gradient in the National Spherical Torus Experiment." *Nuclear Fusion* 49(5): Art No. 055001. DOI:10.1088/0029-5515/49/5/055001.

McKee GR, RJ Fonck, DK Gupta, DJ Schlossberg, MW Shafer, RL Boivin, and W Solomon. 2007. "Plasma Turbulence Imaging via Beam Emission Spectroscopy in the Core of the DIII-D Tokamak." *Plasma and Fusion Research* 2: Art. No. S1025.

Mozer FS, SD Bale, and TD Phan. 2002. "Evidence of Diffusion Regions at a Subsolar Magnetopause Crossing." *Physical Review Letters* 89:015002.

Nagai T, I Shinohara, M Fujimoto, M Hoshino, Y Saito, S Machida, and T Mukai. 2001. "Geotail Observations of the Hall Current System: Evidence of Magnetic Reconnection in the Magnetotail." *Journal of Geophysical Research* 106: Art. No. 25929.

Øieroset M, TD Phan, M Fukimoto, RP Lin, and RP Lepping. 2001. "*In-Situ* Detection of Collisionless Reconnection in the Earth's Magnetotail." *Nature* 412:414-417. DOI:10.1038/35086520.

Øieroset M, RP Lin, TD Phan, DE Larson, and SD Bale. 2002. "Evidence for Electron Acceleration up to 300 keV in the Magnetic Reconnection Diffusion Region of Earth's Magnetotail." *Physical Review Letters* 89(19):195001/1−195001/4.

Ongena J, M Evrard, and D McCune. 1998. "Numerical Transport Codes." In *Proceedings of the Third Carolus Magnus Summer School on Plasma Physics*, September 1997, Spa, Belgium; and published in *Transactions of Fusion Technology* 33(2T):181-191.

Ottaviani M and F Porcelli. 1993. "Nonlinear Collisionless Magnetic Reconnection." *Physical Review Letters* 71(23):3802-3805.

Palaji P, A Chan, R Thakur, W Gropp, and E Lusk. 2009. "Toward Message Passing for a Million Processes: Characterizing MPI on a Massive Scale Blue Gene/P." In *Proceedings of the International Supercomputing Conference*, Hamburg, Germany, June 23-26, 2009.

Park G and CS Chang. 2007. "A 5-1/2-Dimensional Theory for Fast and Accurate Evaluation of the Cyclotron Resonance Heating Using a Real-Space Wave Representation." *Physics of Plasmas* 14(5): Art. No. 052503.

Park G, J Cummings, CS Chang, N Podhorszki, S Klasky, S Ku, A Pankin, R Samtaney, A Shoshani, P Snyder, H Strauss, L Sugiyama, and the CPES Team. 2007. "Coupled Simulation of Kinetic Pedestal Growth and MHD ELM Crash." *Journal of Physics: Conference Series* 78: Art. No. 012087. DOI: 10.1088/1742-6596/78/1/012087.

Park W, ED Fredrickson, A Janos, J Manickam, and WM Tang. 1995. "High-β Disruption in Tokamaks." *Physics Review Letters* 75:1763. Accessed October 20, 2009, at http://prola.aps.org/abstract/PRL/v75/i9/p1763_1.

Parker EN. 1957. "Sweet's Mechanism for Merging Magnetic Fields in Conducting Fluids." *Journal of Geophysical Research* 62(4):509-520.

Phan TD, JF Drake, MA Shay, FS Mozer, and JP Eastwood. 2007. "Evidence for an Elongated (>60 Ion Skin Depths) Electron Diffusion Region During Fast Magnetic Reconnection." *Physical Review Letters* 99(25):255002.

Philip B, L Chaćon, and M Pernice. 2008. "Implicit Adaptive Mesh Refinement for 2D Reduced Resistive Magnetohydrodynamics." *Journal of Computational Physics* 227(20):8855-8874.

Pigarov SI, B Krasheninnikov, T LaBombard, and TD Rognlien. 2007. "Simulation of Large Parallel Plasma Flows in the Tokamak SOL Driven by Cross-Field Transport Asymmetries." *Journal of Nuclear Materials* 363-365:643-648.

Prater R, D Farina, Y Gribov, RW Harvey, AK Ram, Y-R Lin-Liu, E Poli, AP Smirnov, F Volpe, E Westerhof, A Zvonkov, and the ITPA Steady State Operation Topical Group. 2008. "Benchmarking of Codes for Electron Cyclotron Heating and Electron Cyclotron Current Drive Under ITER Conditions." *Nuclear Fusion* 48(3): Art. No. 035006.

Priest E and T Forbes. 2000. *Magnetic Reconnection*. Cambridge: Cambridge University Press, Cambridge, Massachusetts.

Pritchett PL. 2001. "Geospace Environmental Modeling Magnetic Reconnection Challenge: Simulations with a Full Particle Electromagnetic Code." *Journal of Geophysical Research* 106(A3): 3783-3798.

Pritchett PL. 2008. "Energetic Electron Acceleration During Multi-Island Coalescence." *Physics of Plasmas* 15(10):102105-1−102105-9.

Qin H, X Guan, and WM Tang. 2009. "Variational Symplectic Algorithm for Guiding Center Dynamics and its Application in Tokamak Geometry." *Physics of Plasmas* 16(4):042510-1-042510-7. DOI:10.1063/1.3099055.

Radha PB, VN Goncharov, TJB Collins, JA Delettrez, Y Elbaz, VY Glebov, RL Keck, DE Keller, JP Knauer, JA Marozas, FJ Marshall, PW McKenty, DD Meyerhofer, SP Regan, TC Sangster, D Shvarts, S Skupsky, Y Srebro, R Town, and C Stoeckl. 2005. "Two-Dimensional Simulations of Plastic-Shell, Direct-Drive Implosions on OMEGA." *Physics of Plasmas* 12: Art. No. 032702.

Rayleigh L. 1883. "Investigation of the Character of the Equilibrium of an Incompressible Heavy Fluid of Variable Density." In *Proceedings of the London Mathematics Society* 14(1):170-177.

Reiter D, V Kotov, P Börner, K Sawada, RK Janev, and B Küppers. 2007. "Detailed Atomic, Molecular and Radiation Kinetics in Current 2D and 3D Edge Plasma Fluid Codes." *Journal of Nuclear Materials* 363-365:649-657.

Ren Y, M Yamada, S Gerhardt, H Ji, R Kulsrud, and A Kuritayn. 2005. "Experimental Verification of the Hall Effect during Magnetic Reconnection in a Laboratory Plasma." *Physical Review Letters* 95:055003-1−055003-4.

Reynolds DR, R Samtaney, and CS Woodward. 2010. "Operator-Based Preconditioning of Stiff Hyperbolic Magnetohydrodynamics Systems." *Society for Industrial and Applied Mathematics (SIAM) Journal of Scientific Computing* 32(1):150-170. DOI:10.1137/080713331.

Ricci P, JU Brackbill, W Daughton, and G Lapenta. 2004. "Collisionless Magnetic Reconnection in the Presence of a Guide Field." *Physics of Plasmas* 11:4102-4114.

Rogers B and L Zakharov. 1996. "Collisionless m=1 Reconnection in Tokamaks." *Physics of Plasmas* 3(6):2411–2422.

RognlienTD, PN Brown, RB Campbell, TB Kaiser, DA Knoll, PR McHugh, GD Porter, ME Rensink, and GR Smith. 1994. "2-D Fluid Transport Simulations of Gaseous/Radiative Divertors." *Contributions to Plasma Physics* 34(2-3)362-367.

Rognlien TD, MV Umansky, XQ Xu, RH Cohen, and LL LoDestro. 2005. "Simulation of Plasma Fluxes to Material Surfaces with Self-Consistent Edge Turbulence and Transport for Tokamaks." *Journal of Nuclear Materials* 337-339.

Ruebel O, W Prabhat, K Wu, H Childs, J Meredith, C Geddes, E Cormier-Michel, S Ahern, GH Weber, P Messmer, H Hagen, B Hamann, and EW Bethel. 2008. "High Performance Multivariate Visual Data Exploration for Extremely Large Data." In *SuperComputing 2008*, Austin, Texas.

Sakagami H and K Mima. 2004. "Integrated Simulation of Fast Ignition." *Laser Particle Beams* 22:41-44.

Salje EKH; E Artacho, KF; Austen, RP; Bruin, M; Calleja, HF Chappell, G-T Chiang, MT Dove, I Frame, AL Goodwin, KK van Dam, A Marmier, SC Parker, JM Pruneda, JT Todorov, K Trachenko, RP Tyer, AM Walker, and TOH White. 2009. "eScience for Molecular-Scale Simulations and the eMinerals Project." In *Philosophical Transactions of The Royal Society A* 367(1890):967-985.

Sanderson A, X Tricoche, C Garth, S Kruger, C Sovinec, E Held, and J Breslau. 2006. "Detection of Magnetic Nulls in Toroidal Geometry." In *The 48th Annual Meeting of the Division of Plasma Physics*, October 30-November 3, 2006, Philadelphia, Pennsylvania.

Schissel DP. 2004. "Grid Computing and Collaboration Technology in Support of Fusion Energy Sciences." In *The 46th Annual Meeting of the Division of Plasma Physics*, November 15-19, 2004, Savannah, Georgia.

Schneider R, X Bonnin, K Borrass, DP Coster, H Kastelewicz, D Reiter, VA Rozhansky, and BJ Braams. 2006. "Plasma Edge Physics with B2-Eirene." *Contributions to Plasma Physics* 46(1-2):3-191. DOI:10.1002/ctpp.200610001.

Schroeder B and GA Gibson. 2007. "Understanding Failures in Petascale Computers." *Journal of Physics: Conference Series* 78: Art. No. 012022.

Scott BD. 2007. "Tokamak Edge Turbulence: Background Theory and Computation." *Plasma Physics and Controlled Fusion* 49:S25-S41.

Sentoku Y and A Kemp. 2008. "Numerical Methods for Particle Simulations at Extreme Densities and Temperatures: Weighted Particles, Relativistic Collisions and Reduced Currents." *Journal of Computational Physics* 227(14):6846-6861 (and references therein).

Shay MA, JF Drake, BN Rogers, and RE Denton. 2001. "Alfvénic Collisionless Magnetic Reconnection and the Hall Term." *Journal of Geophysical Research* 106:3759.

Shay MA, JF Drake, and M Swisdak. 2007. "Two-Scale Structure of the Electron Dissipation Region during Collisionless Magnetic Reconnection." *Physical Review Letters* 99:15502-1−15002-4.

Shimada M, DJ Campbelle, V Mukhovatov, M Fujiwara, N Kirneva, K Lackner, M Nagamie, VD Pustovitov, N Uckan, J Wesley, N Asakura, AE Costley, AJH Donné, EJ Doyle, A Fasoli, C Gormezano, Y Gribov, O Gruber, TC Hender, W Houlberg, S Ide, Y Kamada, A Leonard, B Lipschultz, A Loarte, K Miyamoto, V Mukhovatov, TH Osborne, A Polevoi and ACC Sips. 2007. "Chapter 1: Overview and Summary." *Nuclear Fusion* 47(6):S1-S7.

Siemon RE, IR Lindemuth, and KF Schoenberg. 1999. "Why Magnetized Target Fusion Offers a Low-Cost Development Path for Fusion Energy." *Comments on Plasma Physics and Controlled Fusion* 18(6):363-386.

Simakov AN and L Chaćon. 2008. "Quantitative, Comprehensive, Analytical Model for Magnetic Reconnection in Hall Magnetohydrodynamics." *Physical Review Letters* 101(10):105003.1−105003.4.

Sipics M. 2006. "Taking on the ITER Challenge, Scientists Look to Innovative Algorithms, Petascale Computers." *Society for Industrial and Applied Mathematics (SIAM)* 39(7). Accessed October 20, 2009, at http://www.siam.org/news/news.php?id=1005.

Solodov A, K Anderson, R Betti, V Gotcheva, J Myatt, JA Delettrez, S Skupsky, W Theobald, and C Stoeckl. 2008. "Simulations of Electron Transport and Ignition for Direct-Drive Fast-Ignition Targets." *Physics of Plasmas* 15: Art. No. 112702.

Sonnerup BUO. 1979. "Magnetic Field Reconnection." *Solar System Plasma Physics*, LJ Lanzerotti, CF Kennel, and EN Parker, Eds., pp 46. North Holland Publishing, Amsterdam, The Netherlands.

Stangeby PC and JD Elder. 1992. "Calculation of Observable Quantities Using a Divertor Impurity Interpretive Code, DIVIMP." *Journal of Nuclear Materials* 196-198:258-263.

Still CH, RL Berger, AB Langdon, DE Hinkel, LJ Suter, and EA Williams. 2000. "Filamentation and Forward Brillouin Scatter of Entire Smoothed and Aberrated Laser Beams." *Physics of Plasmas* 7:2023.

Tabak M, J Hammer, ME Glinsky, WL Kruer, SC Wilks, J Woodworth, EM Campbell, and MD Perry, and RJ Mason. 1994. "Ignition and High Gain with Ultrapowerful Lasers." *Physics of Plasmas* 1(5):1626-1634.

Tabak M and DA Callahan. 2005. "Models of Gain Curves for Fast Ignition." *Nuclear Instruments and Methods in Physics Research, Section A* 544(1):48:54.

Takizuka T and H Abe. 1977. "A Binary Collision Model for Plasma Simulation with a Particle Code." *Journal of Computational Physics* 25:205-210. DOI: 10.1016/0021-9991(77)90099-7.

Taylor G. 1950. "The Instability of Liquid Surfaces when Accelerated in a Direction Perpendicular to their Planes." In *Proceedings of the Royal Society of London*, Series A 201(1065):192-196.

Terasawa T and A Nishida. 1976. "Simultaneous Observations of Relativistic Electron Burst and Neutral-Line Signatures in Magnetotail." *Planetary and Space Science* 24(9):855-866.

Terry PW, M Greenwald, J-N Leboeuf, GR McKee, DR Mikkelsen, WM Nevins, DE Newman, and DP Stotler. 2008. "Validation in Fusion Research: Towards Guidelines and Best Practices." *Physics of Plasmas* 15 (6) Art. No. 062503. DOI:10.1063/1.2928909.

Tonge J, J May, WB Mori, F Fiuza, SF Martins, RA Fonseca, LO Silva, and C Ren. 2009. "A Simulation Study of Fast Ignition with Ultrahigh Intensity Lasers." *Physics of Plasmas* 16: Art. No. 056311.

Trucano T and D Post, Guest Editors. 2004. "Verification and Validation in Computational Science and Engineering." *Computing in Science and Engineering* 6(5): 8-9.

Tskhakaya D, K Matyash, R Schneider, F Taccogna, and Contributors. 2007. "The Particle-In-Cell Method." *Plasma Physics* 47(8-9):563-594.

Tskhakaya D, RA Pitts, W Fundamenski, T Eich, S Kuhn, and JET EFDA Contributors. 2009. "Kinetic Simulations of the Parallel Transport in the JET Scrape-Off Layer." *Journal of Nuclear Materials* 390-391(0):335-338.

Uberuaga BP, F Montalenti, TC Germann, and AF Voter. 2005. *Handbook of Materials Modeling, Part A - Methods*, S Yip, Editor, pp. 629. Springer, New York.

Uberuaga BP and AF Voter. 2006. Pages 25-43 in: *Radiation Effects in Solid*s, KE Sickafus, EA Kotomin, and BP Uberuaga, Editors. Springer, NATO Publishing Unit, Dordrecht, The Netherlands.

Umansky MV, SI Krasheninnikov, B LaBombard, B Lipschultz, and JL Terry. 1999. "Modeling of Particle and Energy Transport in the Edge Polasma of Alcator C-Mod." *Physics of Plasmas* 6:2791. Accessed October 20, 2009, at http://scitation.aip.org/getabs/servlet/GetabsServlet?prog=normal&id=PHPAEN000006000007002791000001&idtype=cvips&gifs=yes.

Vay JL, P Colella, JW Kwan, P McCorquodale, DB Serafini, A Friedman, DP Grote, G Westenskow, JC Adam, A Heron, and I Haber. 2004. “Application of Adaptive Mesh Refinement to Particle-In-Cell Simulations of Plasmas and Beams.” *Physics of Plasmas* 11(5):2928-2934. DOI:10.1063/1.1689669.

Voter AF, F Montalenti and TC Germann, 2002. “Extending the Time Scale in Atomistic Simulation of Materials.” *Annual Review of Materials Research* 32:321.

Voter AF. 2006. *Radiation Effects in Solids*, KE Sickafus, EA Kotomin, and BP Uberuaga, Eds., pp 1-23. Springer, NATO Publishing Unit, Dordrecht, The Netherlands.

Wan W and G Lapenta. 2008. “Electron Self-Reinforcing Process of Magnetic Reconnection.” *Physical Review Letters* 101:015001.

Wang WX, Z Lin, WM Tang, WW Lee, S Ethier, JLV Lewandowski, G Rewoldt, TS Hahm, and J Manickam. 2006. “Gyrokinetic Simulation of Global Turbulent Transport Properties in Tokamak Experiments.” *Physics of Plasmas* 13(9): Art. No. 092505.

Wang X and A Bhattacharjee. 1993. “Nonlinear Dynamics of the m=1 Instability and Fast Sawtooth Collapse in High-Temperature Plasmas.” *Physical Review Letters* 70(11):1627-1630.

Welch DR, DV Rose, BV Oliver, and RE Clark. 2001. “Simulation Techniques for Heavy Ion Fusion Chamber Transport.” *Nuclear Instruments and Methods in Physics Research Section A* 464(1-3):134-139.

White AE, L Schmitz, GR McKee, C Holland, WA Peebles, TA Carter, MW Shafer, ME Austin, KH Burrell, J Candy, JC Deboo, EJ Doyle, MA Makowski, R Prater, TL Rhodes, GM Staebler, GR Tynan, RE Waltz, and G Wang. 2008. “Measurements of Core Electron Temperature and Density Fluctuations in DIII-D and Comparison to Nonlinear Gyrokinetic Simulations.” *Physics of Plasmas* 15(5): 056116-1-056116-11.

Winjum BJ, JE Fahlen, FS Tsung, and WB Mori. 2009. “The Effects of Plasma Packets and Local Pump Depletion in Stimulated Raman Scattering.” In *Preparation for the 51st Annual Meeting of the APS Division of Plasma Physics*, November 2-6, Atlanta, Georgia.

Wischmeier M, M Groth, A Kallenbach, AV Chankin, DP Coster, and R Dux. 2009. “Current Understanding of Divertor Detachment: Experiments and Modeling.” *Journal of Nuclear Materials* 390-391:250-254.

Wright JC, EJ Valeo, CK Phillips, PT Bonoli and M Brambilla. 2008. “Full Wave Simulations of Lower Hybrid Waves in Toroidal Geometry with Non-Maxwellian Electrons.” *Communications in Computational Physics* 4(3):545-555.

Wright JC, LA Berry, PT Bonoli, DB Batchelor, EF Jaeger, MD Carter, E D'Azevedo, CK Phillips, H Okuda, RW Harvey, DN Smithe, JR Myra, DA D'Ippolito, M Brambilla, and RJ Dumont. 2005. “Nonthermal Particle and Full-Wave Diffraction Effects on Heating and Current Drive in the ICRF and LHRF Regimes.” *Nuclear Fusion* 45(11):1411-1418.

Yamada M. 2006. "Experimental Study of Two-Fluid Effects on Magnetic Reconnection in a Laboratory Plasma with Variable Collisionality." *Physics of Plasmas* 13:052119.

Yamada M. 2007. "Progress in Understanding Magnetic Reconnection in Laboratory and Space Astrophysical Plasmas." *Physics of Plasmas* 14(5):058102-1−058102-16. DOI:10.1063/1.2740595.

Yan M, LC Lee, and ER Priest. 1992. "Fast Magnetic Reconnection with Small Shock Angles." *Journal of Geophysical Research Part A, Space Physics* 92(66):8277-8293.

Yin L, BJ Albright, KJ Bowers, W Daughton, and HA Rose. 2007. "Saturation of Backward Stimulated Scattering of a Laser Beam in the Kinetic Regime." *Physical Review Letters* 99(26): Art. No. 265004.

Yin L, W Daughton, H Karimabadi, BJ Albright, KJ Bowers, and J Margulies. 2008. "Three-Dimensional Dynamics of Collisionless Magnetic Reconnection in Large-Scale Pair Plasmas." *Physical Review Letters* 101: Art. No. 125001.

Yin L, BJ Albright, KJ Bowers, W Daughton, and HA Rose. 2009. "Saturation of Backward Stimulated Scattering of Laser in Kinetic Regime: Wavefront Bowing, Trapped Particle Modulational Instability, and Trapped Particle Self-Focusing of Plasma Waves." *Physics of Plasmas* 15(1): Art. No. 013109.

Zenitani S and M Hoshino. 2001. "The Generation of Nonthermal Particles in the Relativistic Magnetic Reconnection of Pair Plasmas." *The Astrophysical Journal* 562(1, 2):L63-L66.

Zenitani S and M Hoshino. 2005. "Three-Dimensional Evolution of a Relativistic Current Sheet: Triggering of Magnetic Reconnection by the Guide Field." *Physical Review Letters* 95(9):095001.

Zenitani S and M Hoshino. 2007. "Particle Acceleration and Magnetic Dissipation in Relativistic Current Sheet of Pair Plasmas." *The Astrophysical Journal* 670:702-726.

APPENDICES

APPENDIX 1: WORKSHOP AGENDA

APPENDIX 2: WORKSHOP PARTICIPANTS

APPENDIX 3: ACRONYMS

APPENDIX 1: WORKSHOP AGENDA

Wednesday, March 18, 2009

Time	**Session**	**Lead**	**Room**
7:30-8:30 a.m.	Registration/Working Breakfast: Panel Chair Meetings		Montgomery Ballroom
8:30-8:40 a.m.	Introduction: Charge/Expectations/Logistics	William Tang	Montgomery Ballroom
8:40-8:50 a.m.	Welcome Address	Stephen Eckstrand	
8:50-9:00 a.m.	Welcome Address	Michael Strayer	
9:00-9:30 a.m.	Burning Plasma/ITER Science Challenges Presented by: James Roberto	Ned Sauthoff	
9:30-9:50 a.m.	General Discussion		
9:50-10:25 a.m.	Advanced Physics Integration Challenges in Fusion Energy Sciences Presented by: Arnold H. Kritz	Alain Becoulet	
10:25-11:00 a.m.	Plasma-Material Interaction Science Challenges	Steven Zinkle	
11:00-11:35 a.m.	HEDP Science Challenges/Laser-Plasma Interactions	Riccardo Betti	
11:35-12:10 p.m.	Basic Plasma Science/Magnetic Reconnection Physics Challenges	Amitava Bhattacharjee	
12:10-12:45 p.m.	ASCR Development Plans for Future High Performance Computing Capability	Rick Stevens	
12:45-1:30 p.m.	Working Lunch		
1:30-5:00 p.m.	**Breakout Sessions: Plasma & Fusion Energy Science**		
	Burning Plasma/ITER Science	Nikolai Gorelenkov	Montgomery Ballroom
	Advanced Physics Integration	John R. Cary	Potomac
	Plasma Materials Interaction Science Challenges	Jeffrey N. Brooks	Rockville
	High-Energy Density Laboratory Physics/Laser-Plasma Interactions	Warren Mori	Bethesda
	Basic Plasma Science/Magnetic Reconnection Physics	William Daughton	Frederick
2:30-3:00 p.m.	General Discussion		Montgomery Reception
5:00-7:00 p.m.	Working Dinner: Summary and Presentations of Each Breakout		Darnestown/ Gaithersburg
7:00-7:30 p.m.	Wrap Up and Agenda for Next Day		Darnestown/ Gaithersburg

Thursday, March 19, 2009

Time	Session	Lead	Room
7:30-8:30 a.m.	Working Breakfast: Summary of Day 1 and Expectations for Day 2	WilliamTang/ David Keyes	Montgomery Ballroom
8:30-Noon	**Breakout Sessions: Plasma & Fusion Energy Science**		
	Burning Plasma/ITER Science	Nikolai Gorelenkov	Montgomery Ballroom
	Advanced Physics Integration	John R. Cary	Potomac
	Plasma Materials Interaction Science	Jeffrey N. Brooks	Rockville
	High-Energy Density Laboratory Physics/Laser-Plasma Interactions	Warren Mori	Bethesda
	Basic Plasma Science/Magnetic Reconnection Physics	William Daughton	Frederick
10:00-10:30 a.m.	General Discussion		Montgomery Reception
12:00-1:00 p.m.	Working Lunch: Outcomes and Changes to Summary of P&FES Breakouts		Montgomery Ballroom
1:00-5:30 p.m.	**Breakout Sessions: Crosscutting Areas**		
	Scalable Algorithms	David Keyes	Potomac
	Data Analysis, Management, and Visualization	Arie Shoshani	Rockville
	Mathematical Formulations	Phil Colella	Bethesda
	Programming Models, Frameworks, & Tools (incl. Languages, Optimization, etc.)	Ewing Lusk	Frederick
2:30-3:00 p.m.	General Discussion		Montgomery Reception
5:30-7:30 p.m.	Working Dinner: Summary and Presentations of Crosscutting Breakouts		Darnestown/ Gaithersburg

Friday, March 20, 2009

Time	Plenary Session	Lead	Room
7:30-8:30 a.m.	Working Breakfast (TBD): Expectations for Day 3	William Tang/David Keyes	Montgomery Ballroom
8:30-9:00 a.m.	Burning Plasma/ITER Science Summary	Nikolai Gorelenkov	
9:00-9:30 a.m.	Advanced Integration Challenges Summary	John R. Cary	
9:30-9:50 a.m.	Plasma Materials Interaction Summary	Jeffrey N. Brooks	
9:50-10:10 a.m.	High-Energy Density Laboratory Physics Summary	Warren Mori	
10:10 a.m.	General Discussion		Montgomery Reception
10:10-10:30 a.m.	Basic Plasma Sciences/Magnetic Reconnection Challenges Summary	William Daughton	Montgomery Ballroom
10:30-10:45 a.m.	Scalable Algorithms Summary	David Keyes	
10:45-11:00 a.m.	Data Analysis & Management &Visualization Summary	Arie Shoshani	
11:00-11:15 a.m.	Mathematical Formulations Summary	Phil Colella	
11:15-11:30 a.m.	Programming Models, Frameworks, & Tools (incl. Languages, Optimization, etc.) Summary	Ewing Lusk	
11:30-12:00 p.m.	Closing Remarks	William Tang/David Keyes	Montgomery Ballroom
12:00-1:00 p.m.	Working Lunch: Plenary Session Closing Remarks	TBD	
Report Writing			
1:00-1:30 p.m.	Discussion and writing of letter report and consolidation of materials (including art work) for draft report	Panel members/scribes in attendance	Montgomery Ballroom
1:30-4:00 p.m.	Finish documentation	Panel leads, organizers, conveners, editors	

APPENDIX 2: WORKSHOP PARTICIPANTS

Name	E-mail Address	Affiliation
Adams, Mark	mark.adams@columbia.edu	Columbia University
Aivazis, Michael	aivazis@cacr.caltech.edu	CalTech
Albright, Brian	balbright@lanl.gov	LANL
Balaji, V	v.balaji@noaa.gov	NOAA
Baxevanidis, Kyriakos	Kyriakos.Baxevanidis@ec.europa.eu	European Commission
Beckman, Peter	beckman@mcs.anl.gov	ANL
Berk, Herb	hberk@mail.utexas.edu	University of Texas
Betti, Ricardo	betti@lle.rochester.edu	University of Rochester
Bhattacharje, Amitava	amitava.bhattacharjee@unh.edu	University of New Hampshire
Bolton, Curt	curt.bolton@science.doe.gov	DOE
Bowers, Kevin	kevin.j.bowers@gmail.com	LANL, LLC
Brooks, Jeffrey N.	brooksjn@purdue.edu	Purdue University
Brown, David	dlb@llnl.gov	LLNL
Buntinas, Darius	buntinas@mcs.anl.gov	ANL
Campolargo, Mario	Mario.Campolargo@ec.europa.eu	European Commission
Candy, Jeff	candy@fusion.gat.com	GA
Carlson, Bill	wwc@super.org	IDA CCS
Cary, John	cary@txcorp.com	University of Colorado
Chaćon, Luis	chaconl@ornl.gov	ORNL
Chan, Vincent	chanv@fusion.gat.com	GA
Chang, C.S.	cschang@cims.nyu.edu	NYU
Chatterjee, Lali	Lali.Chatterjee@science.doe.gov	DOE
Childs, Hank	childs3@llnl.gov	LLNL
Choudhary, Alok	a-choudhary@northwestern.edu	Northwestern University
Cohen, Ron	rcohen@llnl.gov	LLNL
Collela, Phil	pcolella@lbl.gov	LBNL
Cottam, Jean	Jean_CottamAllen@ostp.eop.gov	OSTP
Crawford, Glen	glen.crawford@science.doe.gov	DOE
D'Azevedo, Ed	dazevedoef@ornl.gov	ORNL
Dagazian, Rostom	Rostom.Dagazian@science.doe.gov	DOE
Daughton, William	daughton@lanl.gov	LANL
Decyk, Victor	decyk@physics.ucla.edu	UCLA
Dehmer, Pat	patricia.dehmer@science.doe.gov	DOE
Demmel, James	demmel@eecs.berkeley.edu	UC Berkeley
Diachin, Lori	diachin2@llnl.gov	LLNL
Diamond, Patrick	pdiamond@ucsd.edu	UCSD

Name	E-mail Address	Affiliation
Dongarra, Jack	dongarra@cs.utk.edu	ORNL
Dorland, Bill	bDorland@umd.edu	University of Maryland
Dorr, Milo	dorr1@llnl.gov	LLNL
Drake, James	drake@plasma.umd.edu	University of Maryland
Dudarev, Sergei	sergei.dudarev@ukaea.org.uk	Culham Centre
Eckstrand, Steve	Steve.Eckstrand@science.doe.gov	DOE
Ethier, Stephani	sethier@pppl.gov	PPPL
Fay, Dan	Dan.Fay@microsoft.com	Microsoft
Foster, Ian	itf@mcs.anl.gov	ANL
Fu, Guo-Yong	fu@pppl.gov	PPPL
Fukuyama, Atsushi	fukuyama@nucleng.kyoto-u.ac.jp	University of Japan
Gibson, Garth	Garth.Gibson@cs.cmu.edu	CMU
Gorelenkov, Nikolai	ngorelen@pppl.gov	PPPL
Grama, Ananth	ayg@cs.purdue.edu	Purdue University
Greengard, Leslie	greengard@cims.nyu.edu	NYU
Harrison, Robert	harrisonrj@ornl.gov	ORNL
Harrod, Bill	William.Harrod@darpa.mil	DARPA
Hatton, Angela	angela.hatton@pnl.gov	PNNL
Hesse, Michael	michael.hesse@nasa.gov	NASA
Hey, Tony	Tony.Hey@microsoft.com	Microsoft
Hoang, Thuc	Thuc.Hoang@nnsa.doe.gov	NNSA
Holland, Michael	Michael_J._Holland@omb.eop.gov	OMB
Hoshino, Masahiro	hoshino@eps.s.u-tokyo.ac.jp	University of Tokyo
Houlberg, Wayne	wayne.houlberg@iter.org	ITER
Idomura, Yasuhiro	idomura.yasuhiro@jaea.go.jp	JAERI
Jameson, Leland	ljameson@nsf.gov	NSF
Jardin, Steve	sjardin@pppl.gov	PPPL
Johnson, Gary	garymichaeljohnson@gmail.com	Consultant
Kamath, Chandrika	kamath2@llnl.gov	LLNL
Key, Mike	key1@llnl.gov	LLNL
Keyes, David	david.keyes@columbia.edu	Columbia University
Khaleel, Moe	moe.khaleel@pnl.gov	PNNL
Kirkpatrick, Ronald	rck@lanl.gov	LANL
Klasky, Scott	klasky@ornl.gov	ORNL
Kothe, Douglas	kothe@ornl.gov	ORNL
Kritz, Arnold	kritz@lehigh.edu	Lehigh University
Kruger, Scott	kruger@txcorp.com	Tech-X
Kung, Harriet	harriet.kung@science.doe.gov	DOE
Kusnezov, Dimitri	dimitri.kusnezov@nnsa.doe.gov	NNSA

Name	E-mail Address	Affiliation
Lee, Steven	steven.lee@ee.doe.gov	DOE
Lee, Tsengdar	tsengdar.j.lee@nasa.gov	NASA
Lee, Wei-li	wwlee@pppl.gov	PPPL
Levedahl, W. Kirk	kirk.levedahl@nnsa.doe.gov	NNSA
Li, Hui	hli@lanl.gov	LANL
Li, Sherry	xsli@lbl.gov	LBNL
Livny, Miron	miron@cs.wisc.edu	University of Wisconsin
Lusk, Ewing	lusk@mcs.anl.gov	ANL
Ma, Kwan-Liu	ma@cs.ucdavis.edu	UC Davis
MacCabe, Barney	maccabe@unm.edu	ORNL
Malony, Allen	malony@cs.uoregon.edu	University of Oregon
Mandrekas, John	john.mandrekas@science.doe.gov	DOE
Matthews, Hope	hope.matthews@pnl.gov	PNNL
McInnes, Lori Curfman	mcinnes@mcs.anl.gov	ANL
Meiron, Dan	dim@caltech.edu	CalTech
Meisner, Bob	bob.meisner@nnsa.doe.gov	NNSA
Messina, Paul	messina@mcs.anl.gov	ANL
Meza, Juan	jcmeza@lbl.gov	LBNL
Miller, Bart	bart@cs.wisc.edu	University of Wisconsin
Mori, Warren	mori@physics.ucla.edu	UCLA
Nazikian, Raffi	rnazikian@pppl.gov	PPPL
Nevins, William	nevins@llnl.gov	LLNL
Parker, Scott	sparker@colorado.edu	University of Colorado
Patra, Abani	apatra@nsf.gov	NSF-OCI
Pennington, Rob	robp@ncsa.uiuc.edu	NCSA, NSF
Polansky, Walt	walt.polansky@science.doe.gov	DOE
Post, Douglass	post@hpcmo.hpc.mil	DoD
Putvinski, Sergei	sergei.putvinski@iter.org	ITER
Qin, Hong	hqin@pppl.gov	PPPL
Ren, Chuang	cren@lle.rochester.edu	University of Rochester
Reynolds, Dan	reynolds@smu.edu	Southern Methodist University
Riley, Katherine	riley@mcs.anl.gov	ANL
Roberto, James	robertojb@ornl.gov	ORNL
Romine, Charles	Charles_H._Romine@ostp.eop.gov	OSTP
Rosner, Robert	rrosner@anl.gov	ANL
Ross, Robert	rross@mcs.anl.gov	ANL
Salinger, Andy	agsalin@sandia.gov	SNL
Samatova, Nagiza	samatovan@ornl.gov	ORNL
Samtaney, Ravi	samtaney@pppl.gov	PPPL

APPENDIX 2: WORKSHOP PARTICIPANTS

Name	E-mail Address	Affiliation
Sanderson, Allen	allen@sci.utah.edu	University of Utah
Schissel, Dave	schissel@fusion.gat.com	GA
Schultz, David	schultzd@ornl.gov	ORNL
Scott, Bruce	bds@ipp-garching.mpg.de	IPP, Germany
Sekine, Yukiko	yukiko.sekine@science.doe.gov	DOE
Shadid, John	jnshadi@sandia.gov	SNL
Shalf, John	jshalf@lbl.gov	LBNL
Shasharina, Svetlana	sveta@txcorp.com	Tech-X
Sheehey, Peter	pete@lanl.gov	LANL
Shoshani, Arie	shoshani@lbl.gov	LBNL
Siegel, Andrew	siegela@uchicago.edu	University of Chicago, ANL
Smith, Barry	bsmith@mcs.anl.gov	ANL
Snyder, Philip	snyder@fusion.gat.com	GA
Stevens, Rick	stevens@anl.gov	ANL
Stotler, Daren	dstotler@pppl.gov	PPPL
Straatsma, TP	tps@pnl.gov	PNNL
Strayer, Michael	michael.strayer@science.doe.gov	DOE
Sugiyama, Linda	sugiyama@mit.edu	MIT
Tang, William	tang@pppl.gov	PPPL
Tang, Xianzhu	xtang@lanl.gov	LANL
Thakur, Rajeev	thakur@mcs.anl.gov	ANL
Thio, Francis	francis.thio@science.doe.gov	DOE
Tokman, Mayya	mtokman@ucmerced.edu	University of CA Merced
Town, Richard	town2@llnl.gov	LLNL
Turnbull, Bill	william.t.turnbull@noaa.gov	NOAA
Tynan, George	gtynan@AD.UCSD.EDU	UCSD
Voter, Arthur	afv@lanl.gov	LANL
Vouk, Mladen	vouk@ncsu.edu	North Carolina State Univ.
Walkup, Bob	walkup@us.ibm.com	Microsoft
Wang, Weixing	wwang@pppl.gov	PPPL
Warren, Mike	msw@lanl.gov	LANL
Williams, Roy	roy@cacr.caltech.edu	CalTech
Wirth, Brian	bdwirth@nuc.berkeley.edu	UC Berkeley
Wright, John	jwright@psfc.mit.edu	MIT
Yamada, Masaaki	myamada@pppl.gov	PPPL
Yelick, Kathy	kayelick@lbl.gov	LBNL
Zacharia, Thomas	zac@ornl.gov	ORNL
Zarstorff, Michael	mzarnstorff@pppl.gov	PPPL
Zinkle, Steven	zinklesj@ornl.gov	ORNL

APPENDIX 3: ACRONYMS AND ABBREVIATIONS

AM&PSI	atomic, molecular, and near-surface particle-surface interactions
AMAR	AMR/adaptive mesh and algorithm refinement
AMD	accelerated molecular dynamics
AMR	adaptive mesh refinement
ANL	Argonne National Laboratory
API	application programming interface
ASCR	U.S. Department of Energy Office of Advanced Scientific Computing Research
BES	beam emission spectroscopy
BLAS	basic linear algebra subprograms
BLCR	Berkeley Lab Checkpoint/Restart
BP	binary-packed
CAD	computer-aided design
CDX-U	Current Drive Experiment-Upgrade
CECE	correlation electron cyclotron emission
CG	conjugate gradient
CPES	Center for Plasma Edge Simulation
CPR	checkpoint/restart
CPU	central processing unit
DAG	directed acyclic graph
DARPA	Defense Advanced Research Projects Agency
DEMO	Demonstration Reactor
DFT	density functional theory
DOE	U.S. Department of Energy
DRAM	dynamic random access memory
DT	deuterium-tritium
ECC	error-correcting code
ECRF	electron cyclotron radio frequency
ELM	edge-localized mode
EM	electromagnetic
EPM	energetic particle mode
ESL	Edge Simulation Library
ESMF	Earth System Modeling Framework

ETG	electron temperature gradient (mode)
FACETS	Framework for Core-Edge Transport Simulations
FES	U.S. Department of Energy Office of Fusion Energy Sciences
FESAC	Fusion Energy Sciences Advisory Committee
FFT	Fast Fourier Transform
FI	fast ignition
FPGA	field-programmable gate array
FRC	Field Reversed Configuration
FSP	Fusion Simulation Project
GAM	geodesic acoustic mode
GMRES	generalized minimal residual
GPU	graphics processing unit
GTC	gyrokinetic toroidal code
HEC	high-end computing
HECURA	High-End Computing University Research Activity
HED	high-energy density
HEDLP	high-energy density laboratory plasma
H-mode	high-confinement mode
HPC	high-performance computing
HPCS	High-Productivity Computing Systems
HPSS	High-Performance Storage System
I/O	input/output
ICF	inertial confinement fusion
ICRF	ion cyclotron radio frequency
IFE	inertial fusion energy
IFES	inertial fusion energy science
ITG	ion temperature gradient
JAERI	Japan Atomic Energy Research Institute
JET	Joint European Torus
LANL	Los Alamos National Laboratory
LCF	leadership-class facilities
LHRF	lower hybrid radio frequency

LLNL	Lawrence Livermore National Laboratory
LPI	laser-plasma interactions
MCF	magnetic confinement fusion
MHD	magnetohydrodynamic
MIF	magneto-inertial fusion
MIMD	multiple-instructions-multiple-data
MPI	message passing interface
MTBF	mean time between failures
MTF	Magnetized Target Fusion
NERSC	National Energy Research Scientific Computing
NFS	Network File System
NIC	National Ignition Campaign
NIF	National Ignition Facility
NLO	nonlinear optics
NNSA	National Nuclear Security Administration
NSTX	National Spherical Torus Experiment
NTM	neoclassical tearing mode
NUMA	nonuniform memory access
ORNL	Oak Ridge National Laboratory
PAPI	Performance Application Programming Interface
PDE	partial differential equation
PE	processor elements
PFC	plasma-facing components
PGAS	partitioned global address space
PIC	particle-in-cell
PMI	plasma-material interaction
pNFS	parallel Network File System
POSIX	portable operating system interface for Unix
PPPL	Princeton Plasma Physics Laboratory
PRD	priority research direction
PSF	point spread function
PSI	particle-surface interactions
PVFS	Parallel Virtual File System

REB	relativistic electron beam
RF	radio frequency
RFP	reversed-field pinches
RHS	right-hand side vectors
SciDAC	Scientific Discovery through Advanced Computing
SDM	Scientific Data Management (Center)
SECDED	single-bit-error correction, double-bit-error detection
SIMD	single instruction, multiple data
SOL	scrape-off layer
SpMV	sparse matrix vector
SRAM	static random access memory
SRS	Stimulated Raman Scattering
SWIM	Simulation of Wave Interactions with MHD
TEXTOR	Tokamak Experiment for Technology Oriented Research
TFTR	Tokamak Fusion Test Reactor
UQ	uncertainty quantification
XGC	gyrokinetic edge code
XML	extensible markup language

U.S. DEPARTMENT OF ENERGY

Office of Science

Office of Advanced Scientific Computing Research
Office of Fusion Energy Sciences

Prepared for the U.S. Department of Energy under Contract DE-AC05-76RLO1830

Production support provided by Pacific Northwest National Laboratory,
Fundamental & Computational Sciences Directorate

PNNL-19404