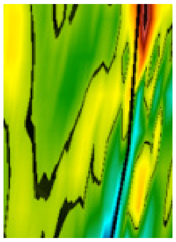
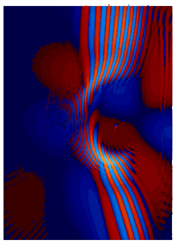
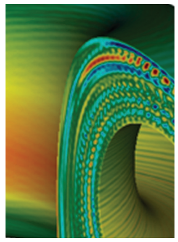
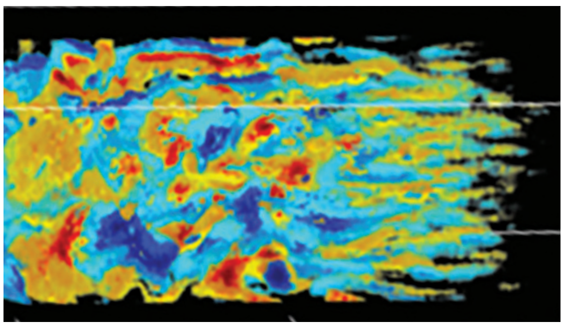
FES

FUSION ENERGY SCIENCES

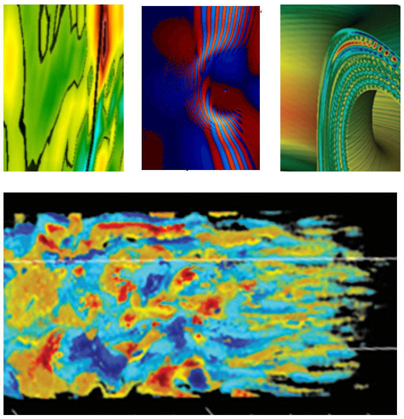
EXASCALE REQUIREMENTS REVIEW

An Office of Science review sponsored jointly by
Advanced Scientific Computing Research and Fusion Energy Sciences

JANUARY 27–29, 2016

GAITHERSBURG, MARYLAND

U.S. DEPARTMENT OF ENERGY



On the cover:

Top left: Relative density fluctuations in the edge region of a tokamak plasma from the XGC1 edge gyrokinetic code (Source: OLCF 2014).

Top middle: Electromagnetic field simulation using the TORIC solver showing incoming long wavelength magnetosonic waves (right), intermediate wavelength mode converted ion cyclotron waves (off the midplane), and short wavelength mode converted ion Bernstein waves (near the midplane) in the Alcator C-Mod tokamak. Reproduced from Bonoli, P.T., R. Parker, S.J. Wukitch, et al., *Fusion Science and Technology* 51, 401 (2007).

Top right: Turbulence spreading in from the edge to the core. Instabilities in the steep gradient edge region drive turbulence to propagate radially inward (Chowdhury et al. 2014).

Bottom: 3D OSIRIS PIC simulation of the formation of a collisionless shock for National Ignition Facility conditions. Two laser-driven counter-streaming plasma flows with velocity of 2,000 km/s and density of 1,020 cm^{-3} interact in the central region, leading to the development of a shock mediated by the Weibel instability. The strong B-fields (up to 300 T) thermalize and slow down the initial flows, leading to a density compression of 4, which is consistent with hydrodynamic jump conditions. Because of the outstanding computational challenges posed by the need to model the kinetic (electron skin depth) scales (submicron) and the system size (cm), these simulations are currently limited to reduced ion to electron mass ratios of ~100.

Disclaimer

FUSION ENERGY SCIENCES

U.S. DEPARTMENT OF ENERGY

An Office of Science review sponsored jointly by
Advanced Scientific Computing Research and Fusion Energy Sciences

Organizing Committee

Meeting Chairs
Choong-Seock Chang, Chair, Princeton Plasma Physics Laboratory
Martin Greenwald, Co-Chair, Massachusetts Institute of Technology

Meeting Organizers
Katherine Riley, Argonne Leadership Computing Facility
Richard Coffey, Argonne Leadership Computing Facility
Richard Gerber, National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory
Jack C. Wells, Oak Ridge Leadership Computing Facility

We want to thank the meeting attendees, who all contributed to the writing of this report. In addition, these individuals took leading roles on the following sections.

Report Section Leads

Fusion Energy Science

Turbulence and Transport
Scott Parker, lead, University of Colorado at Boulder
Frank Jenko, co-lead, University of California at Los Angeles

Energetic Particles and Magnetohydrodynamics
Zhihong Lin, lead, University of California at Irvine
Steve Jardin, co-lead, Princeton Plasma Physics Laboratory

Radio Frequency Wave Heating
Paul Bonoli, lead, Massachusetts Institute of Technology
David Green, co-lead, Oak Ridge National Laboratory

Whole Device Modeling
C.S. Chang, lead, Princeton Plasma Physics Laboratory

Plasma Surface Interactions and Structural Materials
Brian Wirth, lead, University of Tennessee at Knoxville

Discovery Plasma Science
General Plasma Science
Amitava Bhattacharjee, lead, Princeton Plasma Physics Laboratory

High-Energy-Density Laboratory Plasmas
Andreas Kemp, lead, Lawrence Livermore National Laboratory

Low-Temperature Plasmas
John Verboncoeur, lead, Michigan State University

Code Verification and Experimental Validation
Martin Greenwald, lead, Massachusetts Institute of Technology

Path Forward
Katie Antypas, National Energy Research Scientific Computing Center
Richard Gerber, National Energy Research Scientific Computing Center
Katherine Riley, Argonne Leadership Computing Facility
Tjerk Straatsma, Oak Ridge Leadership Computing Facility

Sponsors and Representatives

Program Managers
John Mandrekas, U.S. Department of Energy/Office of Fusion Energy Sciences
Carolyn Lauzon, U.S. Department of Energy/Office of Advanced Scientific Computing Research

Additional Contributors

Argonne National Laboratory Document Design and Development Team
Mary Fitzpatrick, Andrea Manning, Michele Nelson, and Vicki Skonicki, Communications, Education, and Public Affairs
Beth Cerny and Richard Coffey, Argonne Leadership Computing Facility

TABLE OF CONTENTS

FIGURES

TABLE

EXECUTIVE SUMMARY

Abstract

The additional computing power offered by the planned exascale facilities could be transformational across the spectrum of plasma and fusion research — provided that the new architectures can be efficiently applied to our problem space. The collaboration that will be required to succeed should be viewed as an opportunity to identify and exploit cross-disciplinary synergies.

ES.1 Summary and Key Findings

To assess the opportunities and requirements as part of the development of an overall strategy for computing in the exascale era, the Exascale Requirements Review meeting of the Fusion Energy Sciences (FES) community was convened January 27–29, 2016, with participation from a broad range of fusion and plasma scientists, specialists in applied mathematics and computer science, and representatives from the U.S. Department of Energy (DOE) and its major computing facilities. This report is a summary of that meeting and the preparatory activities for it and includes a wealth of detail to support the findings. Technical opportunities, requirements, and challenges are detailed in this report (and in the recent report on the Workshop on Integrated Simulation). Science applications are described, along with mathematical and computational enabling technologies. This review generated the following key findings:

- Progress in computation across the range of fusion and plasma topics in recent years has been dramatic. Advances can be attributed to coordinated improvements in theory, computational and mathematical algorithms, performance engineering, computing hardware and software platforms, and uncertainty quantification (UQ).
- Broader and deeper integration into multiphysics and multiscale domains is a critical next step and will be necessary to address many important questions. These are exascale-level problems. Dramatically increased computing needs are also driven by ensemble runs in support of uncertainty quantification.
- The technical implementations for practical and affordable exascale platforms will present a number of significant challenges to approaches and algorithms used in today's codes.
- Additional challenges are presented in the areas of fault tolerance, software engineering, workflows, data management, *in-situ* analytics, and visualization.
- Close collaboration among stakeholders in various communities will be crucial to overcoming these challenges and realizing the advantages afforded by the new platforms. To that end, a large and specific set of needs for improved computational techniques, programming models, tools, software libraries, and algorithms have been identified.
- Predictable and stable access to high-performance computing resources is essential if the returns from major programmatic investments in code development are to be realized. In general, closer integration of processes for funding people and projects on the one hand and provisioning computer time on the other could lead to more efficient and optimal outcomes.

ES.2 Fusion Energy Sciences Vision and Grand Challenges

ES.2.1 Background and Context

Plasmas, the fourth state of matter, are ubiquitous in nature, making up 90% of the visible universe. The study of plasmas is key to our basic understanding of the cosmos, including the sun and other stars, the interaction of the solar winds and the earth's magnetic field, and the cores of large planets. Plasmas are important for practical applications in modern civilization, notably in the achievement of plasma processing for semiconductors and new materials, lighting, biomedical applications, and — as a scientific grand challenge for the benefit of future civilizations — in the realization of fusion energy. A closely related topic is the interaction of burning plasmas with ordinary matter, including the effects of fusion products on the first wall and structural materials in a fusion reactor. The physics of plasmas is an application of nonequilibrium statistical mechanics governed by the Boltzmann equation (in three spatial dimensions, three velocity space dimensions, and time) coupled to Maxwell's equations for the evolution of magnetic and electric fields. In many practical problems, these also couple to equations for atomic, molecular, and nuclear processes. The theoretical challenges arise from the intrinsic nonlinearity, high dimensionality, and extreme range of mutually interacting temporal and spatial scales that interact with each other in a typical problem. In a strong magnetic field, as in the case with fusion energy applications, extreme anisotropy and sensitivity to geometric details are also present.

ES.2.2 The Role of Advanced Computing

Advanced computing has been an integral part of plasma science and fusion research from its beginnings and is motivated by the recognition that numerical methods would be a necessary complement to analytic theories. As a result, plasma researchers have been in the forefront of scientific computing for the last four decades. This expertise has been a particular strength of the U.S. research program, and advanced computing is a key element in the DOE-FES strategic plan (DOE-SC 2015). In developing the strategy, the program has enumerated a well-defined set of computational challenges that will require exascale computing. It is important to note that historically, improvements in algorithms and software have been as important as speedup in hardware in setting the pace of progress. At the same time, we note the requirement for expanded capabilities on computational platforms at all scales in support of program objectives. Speedy advanced computing for experimental operation planning and next-shot improvement has also been affecting fusion experiments and code validation. The ever-increasing speed and size of experimental data generation from a variety of physics diagnostic measurements have also benefitted from advanced computing. The fusion community has been developing synthetic diagnostics tools for high-fidelity comparison between experimental data and advanced computing data. As ITER will produce larger-than-ever amounts of data in the future exascale computing era, with research and operation participants collaborating all over the world, an advanced remote data management workflow for computational pre-operation planning, run-time experimental steering, and real-time scientific discovery and code validation will be an essential feature.

ES.3 Priority Research Topics and Computing Needs

Review participants focused on the following areas in which advancement in the transformative opportunities can be achieved through key and sustained efforts in computation, simulation, and advanced tool design.

ES.3.1 Fusion Energy Science

ES.3.1.1 Turbulence and Transport in a Fusion Reactor

Plasma turbulence and transport determine the viability of a fusion reactor. To obtain a high enough temperature and particle density for efficient fusion reactions, the plasma is self-heated by the fusion-born energetic alpha particles with assistance from external heating, such as radio frequency (RF) or neutral beam heating. If the plasma energy and particles are lost too quickly, which is mostly determined by plasma turbulence, the fusion burn cannot occur or be sustained. If the plasma is not confined at the edge, then core confinement is not achieved. Advances in high performance computing (HPC) have enabled enormous progress in direct numerical simulation of fusion plasma turbulence. Today, simulations of turbulence and transport routinely compare well with experimental measurement — enabling predictive capability to move forward. These simulations involve complex magnetic geometries, multiple temporal and spatial scales, and multiple physical effects. Comprehensive numerical simulations solve the nonlinear gyrokinetic equations that are rigorously derived from first principles using well-agreed-upon ordering parameters. Multiphysics in the complicated nonthermal edge plasma are not scale-separable and require extreme-scale computing. These simulations are already fully utilizing the capabilities of leadership-class supercomputers (up to the maximal Titan and Mira cores). Considering that the world's first burning plasma, ITER, currently under construction in southern France, will have a plasma volume of about an order of magnitude larger than today's largest existing tokamaks, well-resolved full-torus gyrokinetic simulations for ITER will definitely require exascale-class supercomputers. The ultimate goal is the reliable prediction of the radial temperature and density profiles in ITER and future fusion reactors in order to determine and optimize their fusion performance.

ES.3.1.2 Energetic Particles and Magnetohydrodynamic Instabilities in a Fusion Reactor

Confinement of energetic particles (EPs) is a critical issue for burning plasma experiments because the ignition in ITER relies on the self-heating by energetic fusion products (α-particles). Energetic particles can readily excite mesoscale instabilities that drive large EP transport, which, in turn, can degrade overall plasma confinement and threaten the machine's integrity. Because EPs constitute a significant fraction of the plasma energy density in ITER, energetic particles will also strongly influence the microturbulence responsible for turbulent transport and macroscopic magnetohydrodynamic (MHD) instabilities potentially leading to disruptions. In fact, plasma confinement properties in the ignition regime of self-heating by α-particles is one of the most uncertain issues when extrapolating from existing fusion devices to ITER. Predictive capability requires exascale-level, integrated first-principles simulation of nonlinear interactions of multiple kinetic-MHD processes. For example, the excitation, dynamics, and control of the neoclassical tearing mode (NTM), the most likely instability leading to disruption in a tokamak, depend on nonlinear interaction of MHD instability, microturbulence, collisional (neoclassical) transport, energetic particle effects, and RF waves.

The MHD studies may have different computational requirements than exascale can provide and can be accommodated by mid-range computing: Fusion energy science will continue to require many mid-scale computations in the exascale era. While exascale computing will enable a small number of "heroic" runs of the models in our community that are closest to first-principles, there are other classes of computations (e.g., those that take fluid moments of the kinetic distribution function) that play an essential role in interpreting and guiding experiments. However, the large timescale separation assumption between stability and transport phenomena necessitates very long time calculations to study the onset and eventual saturation or other termination of a global event, such as a disruption. Even though the codes use advanced and fully implicit time-stepping, a single initial value simulation can require hundreds of wall-clock hours and millions of CPU hours to perform a realistic (experimentally relevant) simulation. These jobs normally need to be carried out as a series of restarts, each taking 10–20 wall-clock hours. Because the National Energy Research Scientific Computing Center (NERSC) typically has hundreds of jobs waiting in the queue at any given time, each new restart has to get in line and wait again for its time to run, which alone can take typically several days or longer. This situation leads to periods of months to run a single job to completion. In addition, addressing a scientific or engineering objective often requires scanning parameters and hence many mid-scale calculations. Applying modern UQ techniques also requires performing many simulations of the same event where parameters are systematically varied. It would greatly improve productivity if more hardware were available for running many computations at the 10,000- to 50,000-processor level — an approach that is often called capacity computing. An additional consideration is that many codes (such as the implicit MHD codes) run more efficiently with larger amounts of memory per node than are available on the leadership systems. Dedicated capacity systems with larger amounts of memory may be very cost effective for these codes (which will also help free up the leadership machines for problems that really need their capability).

ES.3.1.3 RF Heating in a Fusion Reactor

The success of next-generation fusion devices and subsequent commercial power plants will rely critically on the robust and efficient application of high-power RF systems in the ion cyclotron, electron cyclotron, and lower hybrid ranges of frequencies. Achieving these goals will depend on the development of a predictive simulation capability of sufficient fidelity for how RF waves interact with the tenuous edge plasma of a fusion device where they are subject to a variety of deleterious interactions, including the formation of RF sheaths at plasma-material surfaces, parametric decay instability, and scattering from turbulence. Once RF power has been successfully coupled to the plasma core, the RF heated plasma species in the core can form nonthermal distributions of ions or electrons, thereby heating specific species or enabling control of the current and pressure profiles, and RF waves can also interact with energetic populations of fusion alpha-particles. A predictive simulation capability is therefore needed to study the stability of these wave-particle interactions and to understand how these energetic populations can be used most effectively to heat and control a burning plasma. A successful collaboration with the Advanced Scientific Computing Research (ASCR) community would make it possible to implement simulation models for coupled antenna-to-core wave-particle interactions on emerging exascale architectures that fully account for the multiscale and multiphysics nature of RF heating and current drive.

ES.3.1.4 Whole-Device Fusion Modeling

A special property of the hot fusion plasma in a toroidal geometry is that there are several multiphysics processes working together, and most of them are scale inseparable and interacting nonlinearly with each other in a self-organized manner at a fundamental physics level. Thus, the fusion reactor plasma must be understood and predicted in a whole-device modeling approach. The best way to simulate the whole-device plasma is to use the 6-dimensional (6-D) Boltzmann or 5-dimensional (5-D) gyrokinetic Boltzmann equation coupled to Maxwell equations for the electromagnetic fields. All of the multiphysics phenomena are included in the 6-D Maxwell-Boltzmann equation system. However, whole-device modeling with a 5-D or 6-D Boltzmann equation cannot be realized until an exascale (or beyond) computational capability is available.

For this reason, only kinetic physics addressing individual phenomena — or a combination of only a few multiphysics phenomena — has been studied in today's leadership-class computers, with corresponding scale separation assumptions. An exascale computer can truly enhance the whole-device modeling capability at high fidelity. An alternative method for whole-device modeling is to componentize the multiphysics using reduced transport equations and fluid models with scale separation assumptions. The fidelity of the reduced components can be improved from the knowledge obtained from kinetic simulations. The latter method has different computational requirements from the former in that it requires only a small to mid-size computer with a quick turnaround time, and thus is preferred by experimental modelers for quick experimental data analysis. A larger-scale computer is needed to run many jobs simultaneously.

ES.3.2 Plasma Surface Interactions and Structural Materials

The realization of fusion as a practical, twenty-first-century energy source requires improved knowledge of plasma-material interactions and the materials engineering design of component systems that can survive the incredibly extreme heat and particle flux exposure conditions of a fusion power plant. The traditional trial-and-error approach to developing first-wall material and component solutions for future fusion devices (ITER, DEMO) is becoming prohibitively costly because of the increasing device size, curved toroidal geometry, access restrictions, and complex programmatic priorities. This set of conditions requires changing from an engineering emphasis toward a more fundamental approach, grounded in a multiscale modeling methodology capable of simultaneously attacking the plasma-material interface problems from both a bottom-up and a top-down approach. The dynamic modeling of the kinetic processes occurring at the near-wall layer requires the coupling together of different physical models and codes, namely:

1. A multi-species kinetic model of the plasma sheath/presheath region, handling the evolution of the distribution function of electrons, ions, neutrals, and material impurities from the quasi-neutral region to the first surface layer; the target equations are the Boltzmann-Poisson and the Boltzmann-Maxwell.
2. A kinetic model of the material wall, handling ion-matter interaction and including relevant phenomena such as sputtering, backscattering, and implantation, on a material surface having dynamic composition and evolving morphology; the target equation is the classical multibody problem for a given (known) interaction potential.
3. A proper collision operator accounting for the interaction among species, handling the relevant atomic physics such as ionization, charge exchange, ion and impurity recycling, and more. The target equations are the Fokker-Planck and nonlinear collision operator.

We anticipate that exascale computing will enable the fusion and ASCR community to achieve an integrated and first-principles-based suite of advanced codes to predictively model the boundary plasma and material surface. Such codes will incorporate rigorous treatment of the turbulent transport, along with kinetic and sheath effects in the plasma, and will be efficiently coupled to a multiscale materials modeling framework. The codes will also predict the evolving plasma-facing components' (PFCs') performance, in terms of erosion, PFC lifetime, and tritium inventory, such that the plasma boundary models can provide feedback to the codes modeling the plasma pedestal and the burning plasma core performance.

ES.3.3 Discovery Plasma Science

ES.3.3.1 General Plasma Science

In this review, we focused on magnetic reconnection and turbulence and their role in particle acceleration and heating in space and astrophysical plasmas. Our understanding of these plasma processes is mature in two dimensions (2D), thanks to advances in theory and the availability of petascale computers; however, much remains to be understood in three dimensions (3D). In high-Lundquist-number (S) plasmas, the computational cost to resolve the reconnection layers and follow the macroscopic evolution on the global Alfvén time increases as $S^{5/2}$ for 3D explicit simulations. For S~10^6, these requirements can quickly surpass the capabilities of petascale computers, thereby requiring exascale-level resources. *The priority research directions in this area are: (1) the influence of the electron and ion kinetic scales on the large-scale evolution, (2) reconnection and magnetic island dynamics in 3D geometries, (3) energetic partition and particle acceleration, and (4) relativistic reconnection.* Because turbulence mediates the transport of energy, momentum, and particles through motions spanning many orders of magnitude in scale, the modeling of plasma turbulence is an inherently multiscale problem, formally beyond the reach of even today's most advanced computers and sophisticated algorithms, so exascale computing shows the path forward to making transformative progress in the field. In addition, the problem of space and astrophysical plasma turbulence is made yet more complex by the fact that, at the typically low densities and high temperatures of these plasmas, the turbulence dynamics are often weakly collisional, requiring the application of kinetic plasma theory to follow the evolution and dissipation of the turbulence. For turbulence research, the key question to answer is: *How does turbulence in a kinetic plasma mediate the conversion of the energy of plasma flows and magnetic fields at large scales to plasma heat, or some other form of particle energization?* Over the next decade, through a coordinated program of spacecraft measurements, theoretical calculations, and nonlinear kinetic numerical simulations, the scientific community is poised to make transformative progress on this problem. Exascale computing will play an essential role in this research effort, enabling direct numerical simulations of the high-dimensional, nonlinear turbulent dynamics.

ES.3.3.2 High-Energy-Density Laboratory Plasmas

High-energy-density laboratory plasmas (HEDLPs) are extreme states of matter characterized by pressures in excess of 1 Megabar. Such systems are routinely created by powerful lasers at many university-scale facilities around the world; they span a wide range of physical phenomena, from microscopic instabilities and laser-driven particle accelerators, to millimeter-scale inertial confinement fusion (ICF) capsule implosions at the National Ignition Facility, to cosmic ray acceleration. The physics of laser-plasma interactions and HEDLP is multiscale, highly nonlinear, and often needs to be described by a kinetic modeling approach. For these reasons, computer modeling of HEDLP experiments requires extreme HPC resources.

Opportunities for HEDLP physics modeling on exascale computer systems arise from several factors, including the ability to (1) increase the problem size to more closely approximate realistic systems than currently possible; (2) increase the grid resolution to improve credibility; (3) run ensembles of runs for error sensitivity/UQ or for providing trends; and (4) reduce turnover time for interactivity with experimental campaigns. To take full advantage of extreme-scale HPC systems, it is essential to have robust I/O tools and *in-situ* analysis/visualization capabilities, as well as the support of high-level languages (e.g., Python) as a front end to number-crunching modules.

HEDLPs hold the promise of leading to breakthrough discoveries in fundamental science, such as discoveries concerning the origin of cosmic rays to applications like inertial confinement fusion and compact X-ray sources for homeland security, provided that exascale computing resources can be leveraged to better understand the multiscale aspects involved in the underlying systems.

ES.3.3.3 Low-Temperature Plasmas

Low-temperature plasmas (LTPs) are partially ionized gases with electron temperatures in the range of 1–10 eV. Atomic and molecular processes play a key role in LTPs, as does interaction with solid state or liquid surfaces, where surface current and charge, particle fluxes, and many geometric effects on fields can play a dominant role. LTPs are involved in about 70% of the steps in the manufacture of the ubiquitous electronics components that drive modern civilization. One of the fastest-growing areas in LTPs is that of biomedical plasmas, which have current applications in surgery, wound healing, and sterilization, with the promise of many future applications yet to be discovered. LTPs are used to modify thin film material for packaging and solar panels, and the ozonizing processes used for water treatment and plasma-based physical vapor deposition coatings, with these markets together amounting to tens of billions of dollars annually. Driven low-temperature plasmas are typically in a strongly nonequilibrium state. To obtain a high-fidelity understanding of these nonequilibrium plasmas, large-scale computational research is key, even though few researchers are utilizing these resources for LTP study at the present time. Exascale resources will provide a capacity improvement of many orders of magnitude. The capability to provide high confidence models will transform the applied use of LTPs in industry, where capital investments and subsequent business success depend upon finding the correct answer and increasingly on knowing the error bars. Exascale-enabled verification, validation, and UQ techniques are one such game changer. Elimination or decreased reliance on ad hoc and simplified physics models is another such disruptor.

ES.3.4 Verification, Validation, and Uncertainty Quantification

A major program element in the FES strategy is to develop “validated predictive models,” so it is important to note that confidence in our models can only be earned through systematic confrontation with experimental data and a sharp focus on careful and quantitative estimates of errors and uncertainties. Fortunately, the disciplines of verification, validation, and uncertainty quantification (VVUQ) are rich areas of research in many technical fields. Our challenge is to find the methodologies and algorithms best suited to our problems; to identify gaps where additional research in applied mathematics and computer science is needed and to apply those techniques to specific codes and simulations; and to secure large-scale computational resources for UQ. Overall, these efforts will have a significant impact on future research directions and computational requirements. A key question, and perhaps the most important source of uncertainty in our domain, is the sensitivity of the models to assumptions and inputs used for any particular problem. Challenges are particularly acute for multiphysics integration, which presents mathematical obstacles, and for multiscale integration, which drives a need for large numbers of production runs of computationally expensive codes. The priority research directions compiled in the body of this report summarize the challenges, enabling us to recommend potential approaches to address those challenges. Notable among these are the need for improved methodologies for code verification, especially for coupled/integrated physics models and the extension of existing intrusive and nonintrusive methods for uncertainty quantification and sensitivity analysis to our particular codes.

ES.4 Path Forward

The support and development of our evolving computing ecosystem relies on continued collaboration between the FES and ASCR communities. Rooted in the discussions about the FES vision, research directions, and computing needs, four categories grew out of the review: methods development, computational environment, data, and communication and community involvement.

Regarding **methods development**, the advancing complexity of computer hardware requires FES researchers to have more scalable, performant algorithms and applications that are capable of efficient execution on future computing architectures fielded by ASCR facilities. Meeting participants discussed those computing ecosystem aspects that will accelerate or impede their progress in the next 5–10 years. Participants named application codes and verification and validation techniques, as well as models and algorithms, as key factors requiring significant methods development activity, as well as additional representative methods identified in Section 3 (and listed in Section 4).

Regarding the **computational environment**, requirements for the access, scheduling, and software ecosystem identify an evolving use-model. The "traditional" HPC model, defined as running a large simulation that generates data that are then post processed, is no longer the only primary use-model for many FES projects. Emerging demands, such as for complex workflows and near-real-time computing, are changing the landscape.

The scale of **data** generated from FES simulations and the requirements needed for verification and validation have created an opportunity and a challenge. ASCR and FES facilities must create more data-centric environments with highly effective data analytics tools for their users. Development of such environments and tools will require expertise from domain scientists, data scientists, and applied mathematicians. Continued collaboration will be required to assess proper deployment of the environments as computing resources evolve.

Activities related to **communication and community involvement** are ongoing today in multiple institutions; however, efforts to connect them to the larger science community have been attempted on an "ad hoc" basis to date. ASCR facilities can explore new or improved communication channels and activities. In addition, experience has shown some of the best impact from strong collaborations.

1 INTRODUCTION

1.1 The DOE Exascale Requirements Reviews Initiative

Throughout fiscal years (FYs) 2015 and 2016, the U.S. Department of Energy's (DOE's) Office of Science (SC) has conducted Exascale Requirements Reviews: one review has been held for each of DOE's six program offices:

- High-Energy Physics (HEP) in June 2015,
- Basic Energy Sciences (BES) in November 2015,
- Fusion Energy Sciences (FES) in January 2016,
- Biological and Environmental Science (BER) in March 2016,
- Nuclear Physics (NP) in June 2016, and
- Advanced Scientific Computing Research (ASCR) in September 2016.

The reviews have brought together key computational domain scientists, DOE planners and administrators, and experts in computer science and applied mathematics to help determine the requirements for an exascale ecosystem that includes computation, data analysis, software, workflows, high-performance computing (HPC) services, and other programmatic or technological elements that may be needed to support forefront scientific research.

A tangible outcome of each Exascale Requirements Review is a report prepared by DOE for wide distribution to subject matter experts and stakeholders at DOE's Office of Advanced Scientific Computing Research (ASCR) facilities, including the leadership computing facilities (LCFs), National Energy Research Scientific Computing Center (NERSC), and the Energy Sciences Network (ESnet).

1.1.1 Previous DOE Requirements-Gathering Efforts: "Lead with the Science"

DOE has experienced definite value in implementing its previous requirements-gathering efforts. As noted by Helland (2016), such review meetings have served to:

- Establish requirements, capabilities, and services.
- Enable scientists, programs offices, and the facilities to have the same conversation.
- Provide a solid, fact-based foundation for service and capability investments.
- Address DOE mission goals by ensuring that DOE science is supported effectively.

1.1.2 National Strategic Computing Initiative (NSCI)

The National Strategic Computing Initiative (NSCI) was established by Executive Order on July 30, 2015. Helland (2016) identified the following four guiding principles:

1. The United States must deploy and apply new HPC technologies broadly for economic competitiveness and scientific discovery.
2. The United States must foster public-private collaboration, relying on the respective strengths of government, industry, and academia to maximize the benefits of HPC.
3. The United States must adopt a whole-of-government approach that draws upon the strengths of and seeks cooperation among all executive departments and agencies with significant expertise or equities in HPC while also collaborating with industry and academia.

4. The United States must develop a comprehensive technical and scientific approach to transition HPC research on hardware, system software, development tools, and applications efficiently into development and, ultimately, operations.

Many of the NSCI's five objectives echo plans already under way in DOE's current exascale computing initiatives. In fact, DOE is among the NSCI's three lead agencies (along with the U.S. Department of Defense and the National Science Foundation), which recognizes these agencies' historical roles in pushing the frontiers of HPC and in helping to keep the United States at the forefront of this strategically important field (Helland 2016).

1.2 FES Exascale Requirements Review, Subsequent Report Preparation, and Purposes

DOE SC convened its programmatic Exascale Requirements Review for Fusion Energy Sciences on January 27–29, 2016, in Gaithersburg, Maryland. The review brought together nearly 100 participants to interact regarding areas of expertise, challenges faced, and possibilities for the future exascale computing environment (Appendix A contains the list of invited review participants).

After DOE and ASCR presenters highlighted tasks to be accomplished (or at least initiated) during the review, a number of FES domain scientists highlighted FES science drivers, focusing on scientific goals to be pursued over the next decade and how exascale computing would play a role in reaching these goals.

Review participants then assembled into breakout sessions for the remainder of the review to discuss key issues in their science domains and the challenges they will need to surmount to make use of exascale-level resources and to begin drafting content for this report concerning FES's HPC requirements (Appendix B contains the review agenda). Participants were tasked with:

- Identifying forefront scientific challenges and opportunities in fusion energy and plasma sciences whose resolution is (1) essential to meeting the FES mission and (2) could be aided by exascale computing over the next decade.
- Establishing the specifics of how and why new HPC capability will address issues at various FES science frontiers.
- Promoting the exchange of ideas among application scientists in the fusion energy and plasma sciences, computer scientists, and applied mathematicians to maximize the potential for use of exascale computing to advance discovery in FES sciences.

Discussions were also guided by input from white papers and case studies that, in many cases, had been authored by the participants themselves and submitted to the FES Organizing Committee chairs in advance of the meeting (Appendices C and D contain the FES white papers and case studies, respectively). This report therefore reflects extensive and varied forms of input from many voices in the FES community regarding HPC requirements for FES's world-class initiatives.

1.2.1 Post-Review Contributions of the FES Organizing Committee

Since the January 2016 review, members of the FES Organizing Committee have met regularly on the FES Exascale Requirements Review report. This effort — led by committee members C.S. Chang (Princeton Plasma Physics Laboratory) and Martin Greenwald (Massachusetts Institute of Technology) — has involved liaising with lead authors on their section drafts, soliciting further input and clarification from review participants when needed, and elaborating upon submitted material.

1.2.2 Exascale Requirements Reports Will Meet Multiple Needs

DOE managers will use the Exascale Requirements Review reports to guide investments and budgeting, complete their strategic planning, and respond to inquiries, including specifically in their efforts to:

- Articulate the case to DOE and SC management, the Office of Management and Budget, and Congress for future HPC upgrades.
- Identify emerging hardware and software needs for SC, including for research.
- Develop a strategic roadmap for the facilities based on scientific needs.

FES program managers may also use the reports to inform their work. The reports therefore need to balance varied end uses and are intended as an information tool to be used by many stakeholders.

1.3 Report Organization

The main sections of this Exascale Requirements Review are these:

- Section 2 provides an overview of the FES vision and grand challenges facing the field of fusion energy sciences.
- Section 3 addresses four areas of scientific challenge and opportunity, along with the priority and cross-cutting research directions and computing needs and requirements associated with each.
- Section 4 outlines a path forward for successful collaboration to occur between FES and the ASCR facilities (i.e., the LCFs and NERSC).

References and the acronyms/abbreviations used in the report are listed in Sections 5 and 6, respectively, followed by the appendices mentioned previously.

This page is intentionally left blank.

2 FUSION ENERGY SCIENCES: VISION AND GRAND CHALLENGES

2.1 Mission and Program Goals

The Fusion Energy Sciences (FES) program mission is to expand the fundamental understanding of matter at very high temperatures and densities and to build the scientific foundation needed to develop a fusion energy source (FES 2016). This mission is accomplished through the study of plasma, the fourth state of matter, and how it interacts with its surroundings.

FES has four strategic goals:

- Advance the fundamental science of magnetically confined plasmas to develop the predictive capability needed for a sustainable fusion energy source;
- Support the development of the scientific understanding required to design and deploy the materials needed to support a burning plasma environment;
- Pursue scientific opportunities and grand challenges in high-energy density plasma science to better understand our universe and to enhance national security and economic competitiveness; and
- Increase the fundamental understanding of basic plasma science, including both burning plasma and low-temperature plasma science and engineering, to enhance economic competiveness and to create opportunities for a broader range of science-based applications.

In response to a Congressional request, FES developed and submitted a report on a ten-year strategic plan perspective (DOE-SC 2015). The report highlights five areas of critical importance for the U.S. fusion sciences program over the next decade:

- Massively parallel computing with the goal of validated whole-fusion-device modeling will enable a transformation in predictive power, which is required to minimize risk in future fusion energy development steps.
- Materials science as it relates to plasma and fusion sciences will provide the scientific foundations for greatly improved plasma confinement and heat exhaust.
- Research in the prediction and control of transient events that can be deleterious to toroidal fusion plasma confinement will provide greater confidence in machine designs and operation with stable plasmas.
- Continued stewardship of discovery in plasma science that is not expressly driven by the energy goal will address frontier science issues underpinning great mysteries of the visible universe and help attract and retain a new generation of plasma/fusion science leaders.
- FES user facilities will be kept world-leading through robust operations support and regular upgrades.

In addition, in 2015, FES sought further community input about scientific challenges and opportunities in the critical areas identified in its strategic plan through a series of four technical reviews. These priorities are described in detail in the reports from these four reviews (FES 2015a–c, FES and ASCR 2015).

2.2 Technical Challenges and Opportunities

Fusion and plasma science researchers have long recognized the opportunities afforded by high-performance computing, establishing the first (unclassified) magnetic fusion energy national supercomputer center at the Lawrence Livermore National Laboratory in the mid-1970s. While plasmas are governed by well-known classical physics embodied in the Maxwell and Boltzmann equations, solution of those equations is among the greatest challenges in physics. The problems of interest exhibit intrinsic nonlinearity, extreme ranges of scales in time and space, extreme anisotropy, and sensitivity to geometric details. Real problems, particularly those that concern the plasma boundary, require additional physics to model atomic physics, neutral and radiation transport, and plasma-material interactions. The early reliance on computation grew from the recognition of these challenges, leading to the understanding that numerical methods would be a necessary complement to analytic theory.

In recent years, advances in theory, numerical algorithms and computer hardware/software have led to dramatic progress allowing meaningful, quantitative comparison between codes and experimental observations across a wide range of problem domains. Notable among many examples are: gyrokinetic modeling of turbulence and neoclassical physics for magnetic fusion energy (MFE) systems and for space and astrophysics applications; full wave calculations of the launching, propagation, and dissipation of short-wavelength radio frequency (RF) waves; nonlinear magneto-hydrodynamics (MHD), including self-consistent interaction with superthermal particles; molecular dynamics models that simulate plasma-material interactions and radiation damage in bulk materials; calculations of magnetic reconnection in plasmas of astrophysical and geophysical interest; and laser-plasma interactions and three-dimensional (3D) hydrodynamics in high-energy-density laboratory plasmas (HEDLPs). Interested readers can find more on these activities in Section 3 of this report and in the recent report on integrated simulations (FES and ASCR 2015).

Also notable, even amid these advances, is that on none of these problems are researchers close to reaching a complete, integrated first-principles solution due partly to the lack of computing power. Progress to date has largely been based on subdividing the problem into separate temporal, physics, or physical domains. It is understood that this separation is only approximate and breaks down in important instances, and thus integration into self-consistent multiphysics and multiscale approaches is critical to making further progress on many important problems. Advances in integrated solutions can be incremental and are already under way. The fusion program has identified a number of “focused integration” or “integrated science applications” that address important challenges with potentially large impact. A sample of significant examples, each of which individually require exascale-class computing to solve, include the following:

- Calculation of plasma profiles — self-consistent with turbulent and neoclassical transport of heat and particles — current drive, and momentum sources: these are crucial for prediction of plasma performance, stability, and sustainment.
- Prediction, avoidance, and mitigation of tokamak disruptions: the tokamak is the most developed magnetic fusion concept and the basis for the ITER device to be built in southern France. However, the tokamak’s use as a practical energy source depends on minimizing the occurrence and impact of disruptions — i.e., sudden collapse of the plasma current, which provides the confining magnetic field.
- Simulations of the boundary plasma on both closed and open magnetic field lines, including edge transport barriers, turbulence, neoclassical physics, MHD, and atomic physics: the boundary plasma provides the critical confinement barrier and boundary condition for the plasma core and the interactions with material surfaces including the divertor, vacuum vessel wall, and RF launching structures.

- Plasma-material interactions (PMI) including relevant plasma and material physics and the effects on materials of bombardment from energetic fusion products: these calculations are essential to predict the erosion and evolution of the material surface, long-term retention of the tritium fusion fuel, and fuel recycling and impurity sources that could affect core plasma performance.
- Coupled simulations of 3D hydrodynamics and radiation transport for HEDLPs.
- Extension of magnetic reconnection calculations from small 2D systems to larger 3D systems with kinetic treatments.

Greatly enhanced computational capability is needed to address these challenges. For example, the computation of plasma profiles will require extension of the validity of the turbulence simulation to timescales typical of the transport equilibrium, that is, roughly three orders of magnitude greater than today's capabilities and significantly more again, if electron-gyroradius-scale turbulence must be resolved simultaneously (which recent work suggests that it does). Similar extrapolations are needed for the other problems as well. Of course, perfect simulation is not the goal, but rather a sufficiently faithful simulation to validate the underlying theory and to predict all needed parameters with enough precision for any future design or operational decisions. We still have considerable progress to make before that distinction is important. It is also worth emphasizing that properly executed simulations at almost every level of physics fidelity have been and will be of immense practical value in the development of our field. Computing platforms at varied levels of capability will continue to be essential in the future.

For MFE, PMI, and low-temperature plasma research, the logical endpoint of integrated simulation is the creation of "whole device models" (WDMs), which self-consistently include components that calculate all elements of an existing or proposed machine. Such a model would need to perform all of the calculations outlined above and more, modeling, for example, the system's power supplies, the electromagnets that they drive along with eddy currents in the device's structure, the response of the plasma, and the feedback control system. For a fusion-burning device, nuclear physics and neutron transport need to be included. It seems that simulating WDMs at full fidelity for all detailed physics may not be practical, even on exascale platforms, and may require beyond-exascale platforms. However, achievable and reasonable goals on exascale computers are WDM with important high-fidelity multiphysics and WDM with a mix of high-fidelity and reduced-physics components. The latter WDM is a collection of model hierarchies, selected for particular applications, to enable utilization of smaller-scale computers than the former WDM requires. The formulation and solution of the strongly and weakly coupled multiphysics problems on extreme-scale computers in the former WDM approach, and the factorization of the problem, the appropriate choice of modules, and development of the required full-physics and reduced-model hierarchies in the latter WDM approach are all ambitious research problems.

Additional computer power offered by exascale facilities could be transformational if it can be applied to our problem set efficiently. Specific approaches to design and build exascale computers are being actively debated and tested, and the outcome of that debate may not be known for years; however, certain architectural features common to various approaches are likely to emerge and will present challenges to our applications. Challenges recognized today include these:

- The central role of power consumption in driving the new generation of machines will require that application codes perform less data movement, use more localized computation, incorporate more on-memory physics analysis, and have lower I/O requirements.
- The large increase in the number of processors with little change in processor speed will push applications toward levels of concurrency not possible for many of today's algorithms.

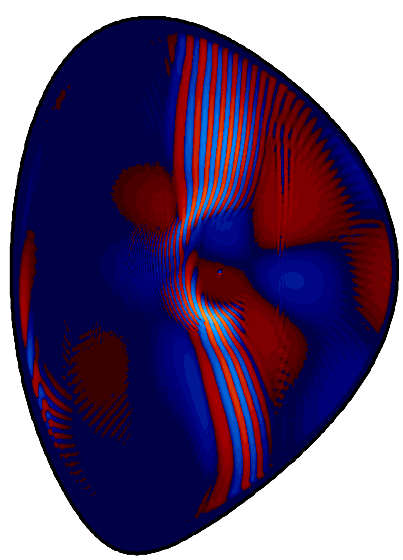
- More complex hardware with hierarchical and heterogeneous memory structure and with a mixture of processor types will require careful tuning of software to the hardware to realize the full capabilities of the platform.
- More frequent, fundamental changes, likely between generations of this class of computers, will require more timely evolution of software.
- Higher error and failure rates in hardware will need to be accommodated in applications and system software.

Close collaboration among the stakeholders will be crucial to overcoming these and other challenges — many of which will perhaps only become apparent as actual computing platforms are put into use since important features of the exascale hardware and software are unknown moving targets. Technologies developed by the applied math and computer science (AM/CS) communities will be essential for FES domain scientists as they develop new codes or adapt older ones. New programming models are needed that will abstract and encapsulate some of the complexity that will be presented by the exascale environment. A key will be to insulate application developers as much as practical from the inevitable changes in underlying implementations. Accompanying the needs for extreme-scale platforms and methods is a set of technologies and platforms that enrich the broader computational ecosystem. The I/O challenge will put a premium on development of *in-situ* and in-transit methods for data analytics and visualization. I/O challenges on current platforms already limit what can be saved for post-run analysis, leaving valuable physics on the table. Similarly, there will be a need for better methods of data management, metadata capture, and organization that span on- and off-HPC processing. Techniques for uncertainty quantification (UQ) that are practical in the exascale environment need to be developed and deployed. Finally, there is a need to develop a model for sustaining the exascale-oriented tools, libraries, and capabilities after they are developed. Widespread adoption of common tools can dramatically reduce costs and enhance productivity for individual research groups; however, this adoption will only come about if those groups are confident that the tools will be maintained and that their investment in adoption will be protected.

Although the road ahead will not be easy, the history of FES partnerships with the advanced computing facilities and collaborations with the AM/CS community, which have been crucial elements for advances made in the past, should provide a model for progress in the future.

.

A VIEW OF THE SCIENCE



Electromagnetic field simulation using the TORIC solver showing incoming long wavelength magnetosonic waves (right), intermediate wavelength mode converted ion cyclotron waves (off the midplane), and short wavelength mode converted ion Bernstein waves (near the midplane) in the Alcator C-Mod tokamak. Reproduced from Bonoli, P.T., R. Parker, S.J. Wukitch, et al., ***Fusion Science and Technology*** **51, 401 (2007).**

3 FES RESEARCH DIRECTIONS AND COMPUTING NEEDS/REQUIREMENTS

3.1 Fusion Energy Science

3.1.1 Turbulence and Transport

3.1.1.1 Scientific Challenges and Opportunities

Key physics challenges on the way to building and operating a fusion power plant are to understand, predict, and control the turbulent transport of heat, momentum, and particles across magnetic surfaces, which is directly related to the computing power. As it turns out, one of the most important figures of merit of a magnetic confinement fusion device is its energy confinement time. For ITER to be successful, this quantity needs to exceed a certain threshold (several seconds).

Advances in high-performance computing have enabled enormous progress in the direct numerical simulation of plasma turbulence in gyrokinetic codes. In the past, the energy confinement time was extrapolated to new (larger) experiments with the help of empirical scaling laws based on the performance of existing (smaller) devices. Today, we have had a breakthrough, where gyrokinetic simulations of turbulence and transport in the core plasma routinely compare well with experimental measurement, enabling predictive capability. These simulations involve complex magnetic geometries, multiple time and space scales, and multiple physical effects or components that are tightly interacting. One shortfall at present is the incompleteness in the simulation of turbulence and transport in the edge plasma, which forms the critical confinement barrier and the boundary condition to the core plasma simulation. Even though gyrokinetic simulations have reproduced the dominant edge turbulence in the low-pressure edge plasma known by experimentalists as nonlinear "blobby" turbulence in the electrostatic limit, many of the important turbulence physics in the steep edge pedestal need to be understood in the nonlinear electromagnetic limit in realistic diverted toroidal geometry. Simulation of the nonlinear electromagnetic turbulence in a gyrokinetic code in steep edge pedestal plasma is still in an incomplete state. Performing a more complete edge turbulence simulation has been a difficult task because of its requirement for more powerful leadership-class computers (exascale preferred) that results from the multiscale, nonlinear, and nonthermal multiphysics issues in the complicated magnetic separatrix and material wall geometry.

Comprehensive numerical simulations solve the nonlinear gyrokinetic equations, which are rigorously derived from first principles using well-agreed-upon ordering parameters. The gyrokinetic equations apply to low-frequency (in relation to the cyclotron frequency of the ions) phenomena in magnetized plasmas and evolve the particle distribution function for each species as a function of time, 3D configuration space, and 2D velocity space (the very fast cyclotron motion can be decoupled and removed analytically). These simulations already fully utilize the capability of leadership-class supercomputers in some codes but consume long wall-clock time on the order of months to include all of the multiscale physics down to the electron gyroradius ($\sim 10^{-5}$ m). Given that ITER will have a plasma volume about an order of magnitude larger than today's largest existing tokamak, performing well-resolved full-torus ("global") gyrokinetic simulations of all-scale multiphysics for ITER will definitely require exascale supercomputers.

To be somewhat more specific, the nature of this computational challenge can be illustrated by considering the vast range of active spatiotemporal scales, all coupled with each other nonlinearly. In the time domain, the relevant scales extend from the important physical timescales that need to be resolved in the turbulence simulations of tens of nanoseconds to transport timescales of up to several seconds and beyond. Meanwhile, in the space domain, the relevant scales span from a typical system size of several meters down to the electron gyroradius smaller than a tenth of a millimeter. While in today's simulations, the scale range is often truncated for practical reasons, clearly the development of a predictive capability over the next decade can be achieved only by leveraging advances in extreme computing resources. The need for extreme-scale computing is elevated to a much higher level for solving problems related to the edge plasma given that this scale-range truncation is not too physical and the geometry is complicated.

The ultimate goal is the reliable prediction of the radial temperature and density profiles, as well as the energy confinement time in the magnetized plasma of a fusion device like ITER or a future fusion power plant, which will help to interpret and guide the experiments. At this future point, there will be a gradual transition to a more computation-driven way of doing fusion science, exploring the enormously large parameter space with the help of virtual fusion devices. And while the spotlight is on ITER, which is a conventional tokamak, computer simulation of a variety of toroidal confinement schemes is prudent, including the stellarator and the spherical tokamak. Contributing to identifying the most promising magnetic field configurations, as well as robust and efficient operational regimes, is an ambitious, but realistic, goal that will have very high impact.

Developing a fundamental turbulent transport physics understanding and a reliable predictive capability for fusion devices is crucial to a successful research program utilizing present-day experiments. This research, in turn, is critical for ITER and the design of future commercial fusion power plants.

3.1.1.2 Priority Research Directions

Although there are many challenging problems in the area of turbulent transport involving gyrokinetic simulations, we focus on six general areas of particular importance that clearly require enormous computing resources:

1. **Multiscale turbulence effects in plasma transport.** A series of investigations has revealed that turbulence at electron gyroradius scales of below 0.1 mm may contribute in a significant way to the overall cross-field heat transport, even dominating it under certain conditions. Moreover, there can be nonlinear interactions with the turbulence on ion gyroradius scales of over millimeters. To retain these effects, conventional turbulence simulations resolving (only) ion scales need to be extended to electron scales, increasing the computational effort by more than three orders of magnitude.
2. **Turbulence and transport in the tokamak edge.** Because the energy confinement time of a tokamak is determined to a large degree by the transport in the outer few centimeters of the plasma (the so-called pedestal and scrape-off-layer [SOL] regions), its quantitative understanding is very important. This problem is extremely difficult and requires extreme-scale computing, however, owing to a lack of (spatial-temporal) scale separation between the profiles and the turbulence, as well as the need to include a large number of physical effects, like open field lines, orbit loss, neutrals, and accurate modeling of magnetic field fluctuations and collisions.
3. **Stellarators and spherical tokamaks.** Another key challenge and also an enormous opportunity for plasma turbulence simulation is the modeling of future fusion power plant confinement schemes other than conventional tokamaks, especially stellarators and spherical tokamaks. International teams are studying alternate fusion reactor concepts that provide a test-bed for direct numerical simulation. Quite a large extrapolation is required for modeling ITER and future fusion power plants. Validating models against alternates provides better and more reliable projections.

4. **Embedded six-dimensional kinetics.** With the advent of exascale computing, it will become possible, for the very first time, to carry out fully kinetic simulations in a 6D phase space, retaining the full Lorentz force in the equations of motion (at least for the ion species). This capability will provide a useful tool for the development and verification of more accurate gyrokinetic models in certain parameter regimes, particularly when there is no clear separation of spatiotemporal scales, like in the edge region of tokamaks.
5. **Coupling gyrokinetics and magnetohydrodynamics.** Although in recent years, gyrokinetic and magnetohydrodynamic studies have usually been carried out separately, there is a growing recognition that these two approaches need to be bridged consistently to enable scientists to address some key open issues, including the long-time evolution of 3D island structures (so-called neoclassical tearing modes) or the comprehensive modeling of the evolution of the equilibrium profiles on transport timescales (of several seconds in the case of ITER).
6. **Extension of the gyrokinetic physics to an experimental timescale.** Fusion scientists may not need the gyrokinetic simulation at all simulation times in order to extend the first-principles-based physics to an experimental transport timescale. A proper multiscale time integration technique should be identified and developed that prolongs the gyrokinetic turbulence and transport physics to an experimental timescale while using only a limited fraction of expensive simulation time. The coarse-grained system can be an axisymmetric kinetic transport system or a fluid/MHD transport system.

3.1.1.2.1 Multiscale Turbulence Effects in Plasma Transport

Nonlinear gyrokinetic simulations have shown that turbulence on both ion scales and electron scales can be important. Moreover, in recent years, several large-scale turbulence validation efforts have demonstrated that turbulence at both scales can interact synergistically. This finding implies that, in general, a reliable predictive capability for turbulent transport needs to be based on multiscale simulations, including electron gyroradius scales. The requirement of resolving both scales (i.e., the electron radius and ion gyroradius scales) increases the size of computation by more than three orders of magnitude. Consequently, a multiscale simulation of core plasma in a large tokamak in flux-tube geometry requires in the range of 15–60 million core-hours. Such flux-tube computations can be carried out on present-day petascale platforms but not the full-torus computations.

The real goal in this research area will be to predict the temperature and density profiles under steady-state conditions in actual fusion experiments by employing multiscale gyrokinetic simulations. One way to achieve this goal is via the technique of coupled flux-tubes. Here, a set of flux-tube computations is carried out in parallel at different radial positions, coupled through a transport code that translates the obtained radial transport fluxes into changes of the profiles and establishing a feedback loop that is reiterated until a steady state is reached (given the experimental heat and particle sources/sinks). Rough estimates for the computational cost of coupled multiscale flux-tube simulations, based on extrapolation from existing single-scale runs, suggest that each individual coupled run may require the equivalent of approximately **~10 B core-hours** on today's machines (~20 M core-hours per run, several parameter variations to characterize parameter sensitivities, ~10 radial positions, ~10 iterations), which does not take into account potential gains from algorithmic improvements.

An alternative route to predicting the steady-state temperature and density profiles, also considering electron-scale turbulence, is to carry out full-torus gyrokinetic simulations, resolving sub-ion scales (Figure 3-1). For such gradient-driven runs that utilize field-aligned coordinates, the overall computational effort scales like the square of the linear machine size (normalized to the thermal ion gyroradius). Thus, simulating an ITER plasma in this way is about 100 times more expensive than doing the same for a present-day smaller tokamak (given a linear size ratio of the order of 10), suggesting that individual full-torus multiscale simulations for ITER will require at least **~2 B core-hours**, again not taking into account potential gains from algorithmic improvements. Given that a few such runs might be necessary to adjust the temperature and density profiles until the experimental heat and particle fluxes are matched, the overall effort is expected to approximate that required for the coupled flux tubes. However, a major advantage of the latter approach is that it can also capture regions of rapid radial profile variations, like transport barriers. Although in principle we could also envision carrying out flux-driven full-torus simulations (including multiscale effects) in lieu of a small set of their gradient-driven counterparts, the computational cost would be prohibitive for ITER-size plasmas unless a spatially embedded grid technique is used, in collaboration with ASCR scientists, for the electron-scale turbulence by identifying spatial regions where such activity is possible.

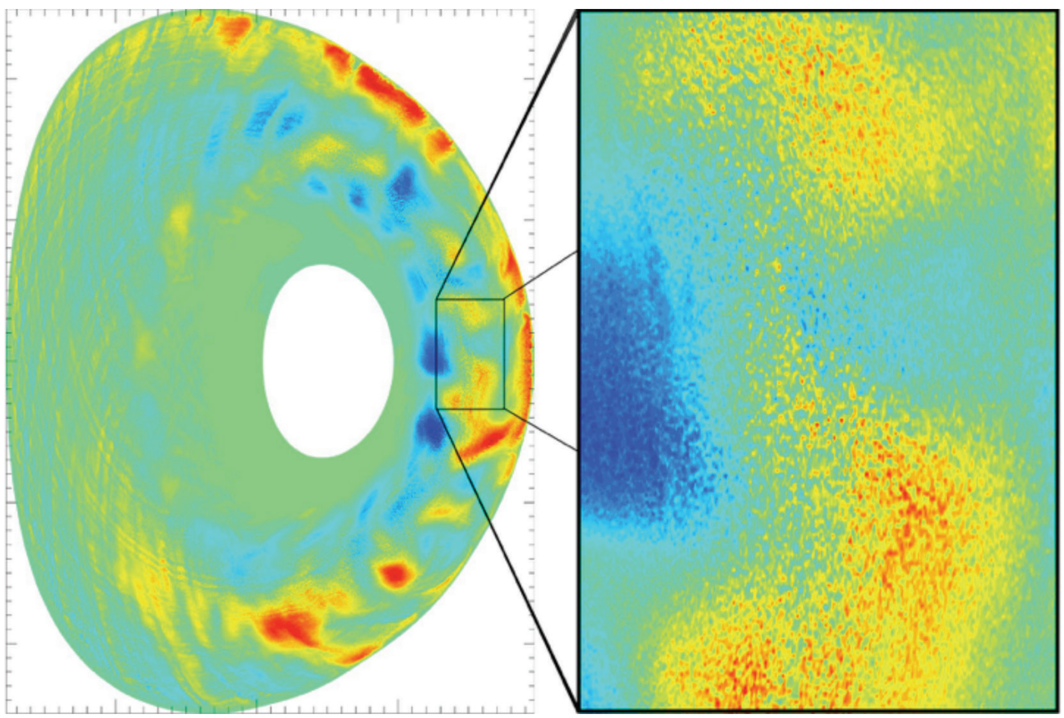


Figure 3-1. Full-torus gyrokinetic simulation of multiscale plasma turbulence in a small experimental tokamak (Jenko et al. 2013). For the first time, all spatial scales from the system size down to the electron gyroradius are treated self-consistently.

3.1.1.2.2 Turbulence and Transport in the Tokamak Edge

Simulating the tokamak edge is an enormously important computational challenge. Present-day tokamak experiments behave like a thermos bottle where a narrow edge boundary layer provides a large part of the plasma confinement. This boundary layer includes the so-called pedestal and adjacent SOL regions. Unlike in the core, there is a compression of the space scales due to the small radial width of the pedestal that closely couples the profile, neoclassical, turbulence, and neutral particle dynamics. The complexity of the edge is further compounded by low to high collisionality approaching the SOL from pedestal top; also important are neutrals, as well as the changes in the magnetic topology going from closed to open field lines that bombard the divertor plate with plasma energy.

The ratio of the ion gyroradius and the device size at low collisionality is the dimensionless plasma parameter whose ITER values cannot be accessed on present-day experiments. Predicting this dependence represents a critical challenge and opportunity for computational modeling in extrapolating the physics from the present machine to ITER. Recent work suggests that, because of the scaling of the pedestal velocity shear (helping to suppress ion-scale turbulence) with the pedestal width, ITER may lie in a different regime from present-day experiments and thus highlight the role of magnetic field fluctuations and electron-scale dynamics. Alternately, the enormity of the plasma's free-energy in the lower collisionality edge pedestal may produce a greater level of microturbulence than we observe from simulations on present machines and may reduce the pedestal gradient to a less violent level. In addition, the primary turbulence activity in the edge plasma, especially in the low beta (plasma energy/magnetic field energy) region in the pedestal foot and the scrape-off area, is in the form of large-scale coherent structures called "blobs" (see Figure 3-2). The blobby turbulence is very different from the turbulence seen in the core plasma and is considered to play an important role in the edge plasma transport. Thus, a primary task for gyrokinetic simulation is to be able to simulate pedestal physics with ever-greater fidelity in order to determine the degree of impact that low-velocity shear has on the expected ITER performance. Furthermore, another challenge for simulations is to help determine the optimal geometry to ensure robust pedestal structure and concomitant confinement, potentially identifying burning plasma devices whose size and cost are a fraction of current projections.

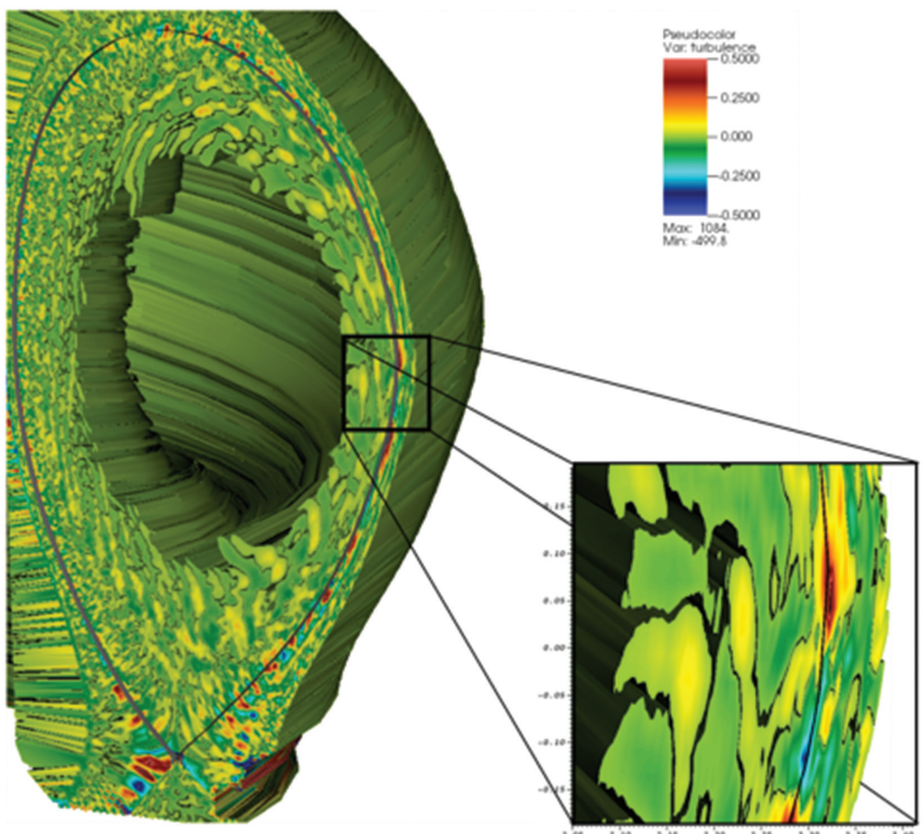


Figure 3-2. Relative density fluctuations in the edge region of a tokamak plasma from the XGC1 edge gyrokinetic code. The black line depicts the magnetic separatrix surface. The blobby nonlinear coherent turbulence in the low beta region around the magnetic separatrix and scrape-off layer can be seen in the enlarged box. Inside the magnetic separatrix surface, on the other hand, the streamer-type turbulence from trapped electron and ion temperature gradient modes can be seen, together with the self-regulating, sheared ExB-flow[1] flow activities (Source: OLCF 2014).

[1] ExB-flow: Flow of plasma in a magnetic field when there is an electric field perpendicular to the magnetic field vector.

In addition, it is also important to better understand and validate the transport bifurcation mechanisms in the tokamak edge by simulating the multiphysics self-organization phenomena that self-consistently include the turbulence, kinetic neoclassical physics, neutral particles, and effects of the separatrix geometry. The bifurcation mechanisms may be different for different tokamaks' operational regimes and transitions. Another important problem where simulation is needed is in determining the interactions between the turbulence and macroscopic MHD-type instabilities. In particular, it is important to find ways to avoid so-called edge-localized modes (ELMs). In ITER, the heat load width (the width of the SOL) is a critical problem because of the amount of localized energy flow to the divertor plate. We need a better understanding of impurity transport and radiation loss, as well. Contamination of the wall-generated impurities that diffuse into the core plasma is of serious concern. The nonlinear turbulence in the edge region is also nonlocally connected to the core region, synergistically affecting the core transport (Figures 3-2 and 3-3).

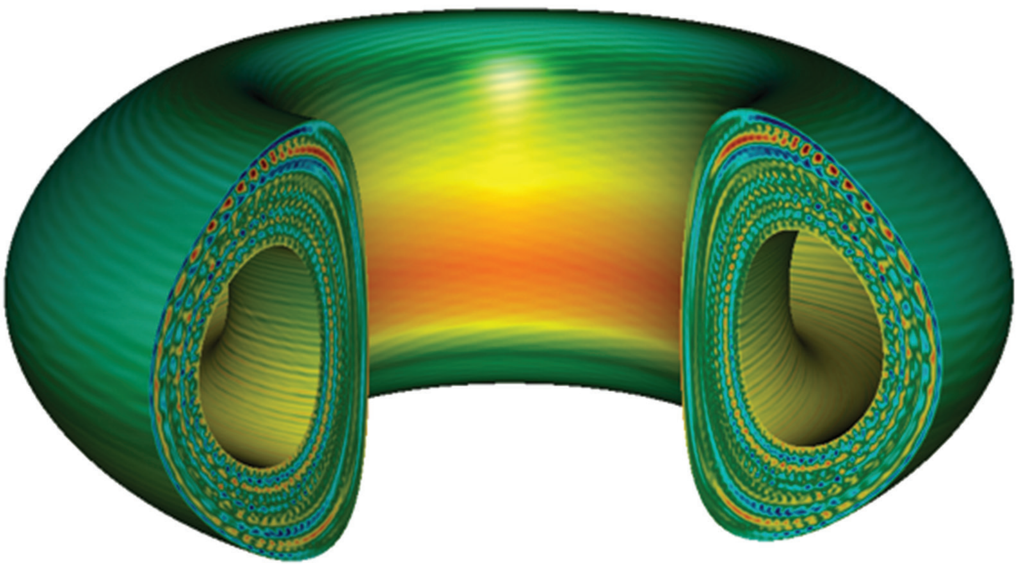


Figure 3-3. Turbulence spreading in from the edge to the core. Instabilities in the steep gradient edge region drive turbulence to propagate radially inward (Chowdhury et al. 2014).

3.1.1.2.3 Stellarators and Spherical Tokamaks

While the most advanced magnetic confinement scheme is the conventional ITER-like tokamak, international teams are studying several alternate fusion reactor concepts, especially stellarators and spherical tokamaks. Simulating these concepts present additional key challenges but also enormous opportunity for plasma turbulence simulation. First, these alternate concepts are of great interest in and of themselves and may turn out to lead to some attractive design features of fusion power plants and/or high-energy neutron sources. Second, they provide an excellent test-bed for direct numerical simulation. Given that quite a large extrapolation is required for modeling ITER and future fusion reactors, validating models against alternates provides better and much more reliable projections.

Very recently, a new large stellarator, Wendelstein 7-X, has started operation in Greifswald, Germany, and is supported by various U.S.-based efforts. There is also a large stellarator called the Large Helical Device (LHD), in Toki, Japan. Both of these stellarators utilize superconducting coils, as will ITER. Wendelstein 7-X is the first large stellarator that has been systematically optimized (with the help of computer-aided design) to minimize collision-induced (so-called neoclassical) transport, which has been the dominant transport mechanism before the optimization. As a consequence, *turbulent* transport will dominate the confinement properties of Wendelstein 7-X.

This provides a significant challenge to gyrokinetic simulation, given the fact that the lack of a geometrical continuous axisymmetry (Figure 3-4) calls for the development of advanced simulation codes that are able to retain fully 3D variations of the magnetic geometry. As it turns out, such tools can also be applied to understand and predict the behavior of modern tokamaks that employ symmetry-breaking external field coils to suppress large-scale instabilities near the plasma edge (so-called ELMs).

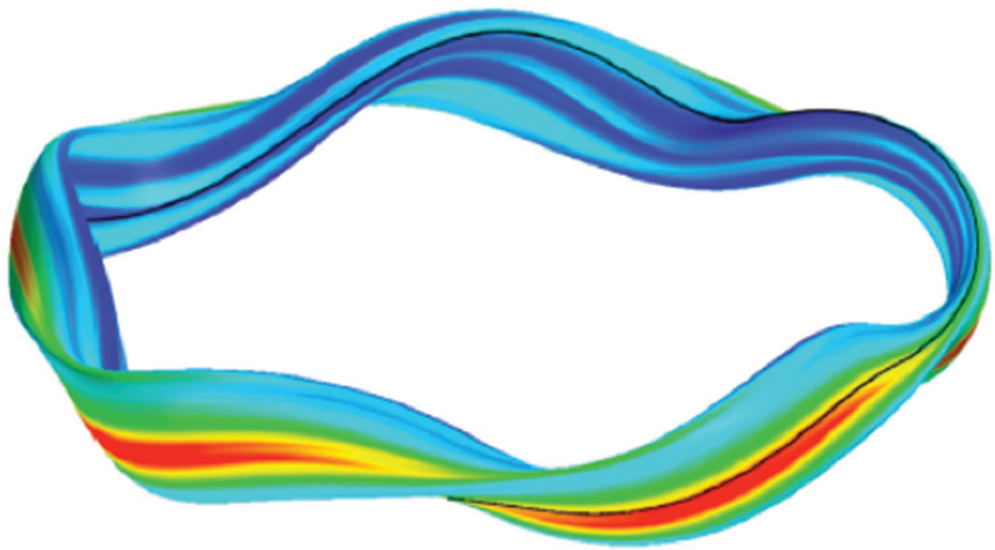


Figure 3-4. Snapshot of the first full-flux-surface gyrokinetic simulation of plasma turbulence in the Wendelstein 7-X stellarator. All variations of the confining non-axisymmetric magnetic field are retained, and the turbulent dynamics are very unevenly distributed across the flux surface (Xanthopoulos et al. 2014).

Another exciting new opportunity in the age of optimized stellarators is to extend the transport minimization strategy so that it also includes turbulent transport. Stellarators have an enormous number of design parameters (approximately 100) that define the confining magnetic field. Minimizing both collision- and turbulence-induced transport at the same time opens up completely new windows for reactor design. This transport model represents an ideal usage of exascale computing power and an excellent area for applied math/physics collaborations. There is ongoing work to try to optimize stellarator design by using fast proxy functions that approximate the turbulent transport. Exascale computing will allow replacing the proxy functions with running a nonlinear gyrokinetic code directly to calculate the turbulent transport. Pioneering studies along these lines have recently been carried out using gyrokinetic simulation (Mynick et al. 2010; Xanthopoulos et al. 2014). Extending these runs to fully global simulations with kinetic electrons and sub-ion-scale dynamics will require exascale resources, namely, **several billion core-hours** per high-resolution run.

Spherical tokamaks also lead development in gyrokinetic simulation in important ways. Their relative compactness, high-beta operational regime, and strong shear-flow suppression of ion-scale turbulence calls for full-torus (global), electromagnetic, multiscale runs, similar to the one shown in Figure 3-1. Such simulations can be carried out only on exascale computers, requiring **more than 1B in core-hours' time**.

3.1.1.2.4 Embedded 6D Kinetics

In order to support existing gyrokinetic models and facilitate the development and verification of more accurate gyrokinetic models in certain parameter regimes — particularly when there is no clear separation of spatiotemporal scales, like in the edge region of tokamaks — it will be useful to devise embedded 6D kinetic simulations based on the fully kinetic Lorentz ion dynamics. Modeling 6D kinetics for low-frequency phenomena presents an enormous challenge owing to the extreme computing requirements, as well as the need for implicit and multiscale techniques to handle the ion cyclotron dynamics. Advanced algorithms and efficient implementations on cutting-edge supercomputers will be required to render this task tractable.

Although in principle the gyrokinetic model can be made as accurate as desired, in practice only terms that are accurate to first order in specified small parameters are typically retained. Higher-order terms can be kept, but this treatment can lead to increased computational complexity. There are certain problems where further validation of the gyrokinetic model would be of significant value. For example, an important problem where higher-order gyrokinetic equations are needed involves modeling the evolution of very-long-wavelength radial electric fields. Another high-priority research area is the edge plasma, which has very strong profile variation and where the pressure gradient scale is only a few ion gyroradii in length. Third, in steep gradient regions where there is strong plasma flow with strong radial flow shear (such as in the transport barrier), electromagnetic gyrokinetic equations that are suitable for numerical implementation have not been validated yet. A typical embedded simulation will have a problem size of 10^8 grid cells, with at least 100 to 1,000 particles per cell.

3.1.1.2.5 Coupling Gyrokinetics and Magnetohydrodynamics

A major scientific challenge for ITER is gaining an understanding of the self-consistent interaction of magnetic islands with the surrounding turbulence. This research area is important because the evolution of a magnetic island, called a neoclassical tearing mode (NTM), can involve nonlinear instability and lead to a major plasma disruption, terminating the discharge. Experimental studies suggest that a large fraction of discharge disruptions could be triggered by an NTM. Predictive understanding of NTM physics needs to include the interaction with the background turbulence.

Performing extended two-fluid MHD stability computations based on solutions of the magnetic equilibrium equation is routine, and these computations are known to be both robust and accurate. However, tracking the evolution of the extended MHD system into the strongly nonlinear magnetic distortion regime, which is characterized by significant deviations from the MHD equilibrium, remains a scientific challenge. Some MHD codes have the capability of loosely coupling a "test particle" population of energetic particles (EPs) to the MHD part, and this approach is useful for studying the resonant interaction of energetic ions with MHD modes. However, present efforts to systematically incorporate kinetic effects in extended MHD codes are focused on obtaining closures for the parallel viscous stress and heat flow using numerical solutions of long-wavelength kinetic equations. The overarching goal is a self-consistent, high-fidelity, global simulation capability for slow (compared to gyromotion), long-wavelength (compared to gyroradii) phenomena in hot tokamak plasmas.

It is important to note that all such efforts currently are limited to the long-wavelength limit and thus rule out the faithful inclusion of drift-wave turbulence. Thus, with regard to kinetic extensions to MHD, a long-term challenge is to incorporate drift-wave physics into the formulation. However, looking at this problem from the perspective of transport theory provides a complementary picture. The separability of neoclassical and turbulent transport is a special result of the assumption in which equilibrium and fluctuation scales are widely separated. When equilibrium scales are short, the total radial flux cannot be separated into a simple sum of neoclassical and turbulent parts.

Thus, a true kinetic extension to MHD would actually represent a unified theory of low-n and high-n phenomena in which collisional and drift-wave radial fluxes are not separable. Solving this problem is a challenging and long-term goal.

3.1.1.2.6 Extension of the Gyrokinetic Physics to an Experimental Timescale

Fusion scientists may not need the gyrokinetic simulation at all simulation times in order to extend the first-principles-based physics to an experimental global transport timescale. A proper multiscale time-integration technique can be identified and developed to prolong the gyrokinetic turbulence and transport physics to an experimental timescale using only a limited fraction of simulation time. The fine-grained gyrokinetic simulation can be coupled to a coarse-grained system for the restricting and lifting operations. The coarse-grained system can be an axisymmetric kinetic transport system or a fluid/MHD transport system depending upon the problem. In the core plasma where the plasma is mostly Maxwellian, a fluid-restricting operation could be a choice, as described in Section 3.1.1.2.1, Multi-Scale Turbulence Effects in Plasma Transport. However, in the nonthermal edge plasma where the particle distribution function is highly distorted from that of Maxwellian plasma, a kinetic restriction operation will be needed in order to avoid losing the critical kinetic information in the restricted transport processes. The critical kinetic information can then be transferred back to the lifting operation. Collaboration with applied mathematicians and data management scientists is a necessary element in developing a kinetic restriction operation, as will devising an efficient usage of the heterogeneous processors and the hierarchical memory structures.

3.1.1.3 Cross-Cutting Research Directions

Computing effectively at the petascale is already a daunting task, and exascale promises to be much more challenging. This is why our community will need computer scientists and applied mathematicians to work hand in hand with the computational plasma physicists. Ideally, these experts should be active participants in the code development teams and come from all areas of large-scale computing, including performance optimization, parallel algorithms, *in-situ* data analysis and visualization, resiliency management, and heterogeneous computing and hierarchical memory management. FES physicists well recognize that these topics are fields of research in their own right. Exascale code projects will need well-balanced teams of dedicated scientists from both fusion plasmas physics and HPC research.

A significant part of all MHD (and some gyrokinetic solvers) is the direct and sparse matrix algebra in combination with semi-implicit time-advance methods. Thus, threaded and accelerated versions of these computational kernels are needed. The present bottleneck for "test-particle MHD" (in which test particles are pushed in low-n MHD fields) is the memory- and logic-intensive mapping between particles and mesh, and between physical mesh and logically rectangular discretization mesh. In the continuum simulation of method, drift-kinetic and gyro-kinetic solutions with two extra-velocity space dimensions are memory intensive with millions of unknowns (velocity-mesh quantities) even in modest simulations. The required implicit treatment of free-streaming and bouncing terms in the continuum model leads to large matrices requiring domain decomposition in velocity space, sophisticated preconditioning strategies, and robust linear algebra solvers. Global simulations often require extensive grid resolution near rational surfaces and divertor separatrices and will likely require coupling to neutral and atomic physics models from the edge out to plasma-facing components in unstructured triangular mesh where appropriate boundary conditions must be supplied. Matrix conditioning challenges arise for implicit MHD. Operator splitting is used to alleviate the conditioning problem. Here, the HPC challenges are all too familiar. The discretized MHD equations, which require implicit or semi-implicit time advance, are limited in their ability to scale efficiently to more than a few thousand cores on, say, the Edison system at NERSC, which has Intel Xeon processors and a Cray Aries node interconnect. Solvers almost universally make use of direct dense/sparse solves (SuperLU, Mumps, Scalapack). Producing new numerical algorithms that merge the capabilities of both low-n toroidal and high-n field-aligned schemes is an important area of research for cross-cutting research between MHD and turbulence.

Simulation of low-frequency turbulence with Lorentz ions can benefit from many areas of computational plasma research. 6D kinetics will evolve ions using Newton's second law directly with gyrokinetic equations for the electrons. Existing implicit multiscale orbit-averaging and sub-cycling techniques have been developed to allow modeling low-frequency plasma physics with Lorenz ions. The main computational challenge at this time is developing an efficient algorithm for Maxwell's equations in complex toroidal geometry. Implicit methods are needed that allow a time step to be a fraction of the inverse ion gyrofrequency. At the core of the field solvers is the need for a massively parallel solver for linear equations with dense matrices. This effort will greatly benefit from the expertise of applied mathematicians. Once developed, the field solver can be used in MHD or two-fluid modeling. The orbit-averaging and subcycling of the ion cyclotron motion is amenable to fine granularity in many CPU architectures (e.g., in GPUs or math co-processors) because this compute-intensive part of the calculation can be performed locally in memory. For some 6D applications, such as neoclassical transport, a noise-free delta-f collisional algorithm is needed. Non-Monte-Carlo collision methods have been developed already for edge turbulence using a phase-space grid distribution from the particle data. Such a simulation will require a very large number of particles per cell, on the order of 10^5, which results in simulations with as many as 10^{13} particles running for on the order of 10^5 time steps.

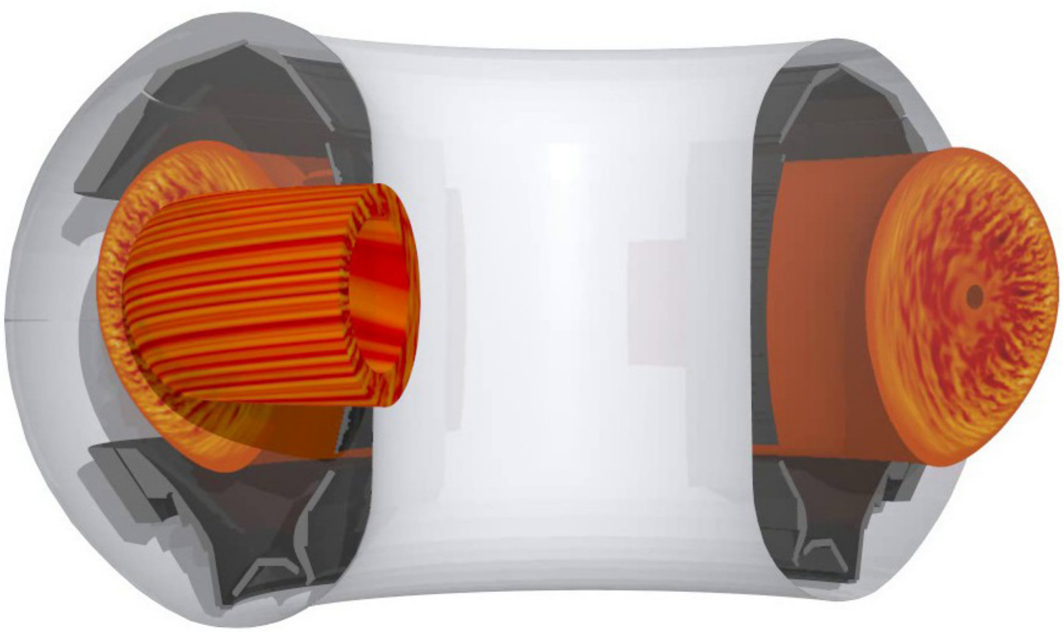
3.1.1.4 Computing Needs and Requirements

Multiscale core-region turbulence simulations of experiments by a continuum gyrokinetic code that have been completed to date required approximately 15 M CPU hours each on the NERSC Edison system and approximately 37 days for completion (using approximately 17 K cores). Clearly, access to larger and more capable computers and development of algorithms that can scale to processor counts in the 10^5–10^7 range are likely needed to use multiscale simulation for profile prediction and to reduce the time to solution to reasonable levels. Multiscale edge simulations (even without resolving the electron gyroradius-scale turbulence) by a particle-in-cell gyrokinetic code require runs lasting a few days on the entire heterogeneous 27 PF (peak) Titan system at the Oak Ridge Leadership Computing Facility to complete a one-case physics study, consuming more than 30 M CPU-GPU hours using a scalable gyrokinetic code, XGC1.

Increased physics fidelity and simulation resolution will inevitably increase requirements for data analysis and storage. Current physics output files can reach 0.5 TB per multiscale simulation for continuum codes and much more than 100 TB for particle codes. Expected increases in the simulation dimensions and inclusion of additional physics would result in at least an order-of-magnitude increase in storage needs. On-the-fly, on-HPC data analysis and reduction are necessary elements for these fusion codes. Computational requirements for particle codes will likely exceed the requirement for continuum codes. These estimates underscore the need for exascale resources for this challenge, as both approaches could conceivably require in the range of 10 B core-hours. However, it is reasonable to assume that improvements to algorithms over the next decade will likely reduce the requirements of such simulations, perhaps by an order of magnitude or more.

Edge plasma simulation is enormously challenging. It involves multiscale and multiphysics effects that are inherently kinetic and take place far from thermal equilibrium. Simulations with fundamental physics require an exascale computing ecosystem. Data from a 24-hour clock time run can reach exa-bytes assuming an exascale HPC using particle-cell methods. Dedicating this amount of storage may not be possible with future file storage systems. *In-situ*, on-HPC data analysis, visualization, and reduction are needed, in addition to a physics-/math-based coarse graining of the data. We will also need on-the-fly data analysis and reduction in the filesystem to make data archiving possible. Tools that support portability and accessibility among various extreme-scale HPCs will be important for productive workload sharing of edge physics research, depending upon the algorithms that dominate different edge physics topics.

A VIEW OF THE SCIENCE



GENE simulation of turbulent fluctuations in an actual tokamak discharge. Such studies reveal a complex nonlinear interplay between structures of various sizes and shapes, and they allow for qualitative and quantitative predictions that can be used to interpret and guide experimental measurements.

3.1.2 Energetic Particles and MHD

3.1.2.1 Scientific Challenges and Opportunities

The confinement and stability properties of burning plasmas depend on nonlinear interactions of multiple physical processes spanning a large range of spatial-temporal scales. Predictive capability requires integrated simulation of the nonlinear interactions of multiple kinetic-MHD processes. For example, the excitation and evolution of macroscopic electromagnetic instabilities often depend on kinetic effects at microscopic scales, as well as the nonlinear coupling of multiple physical processes (e.g., turbulent and neoclassical transport, EPs, heating and current drive) spanning disparate spatial and temporal scales. In fact, the excitation of an NTM, the most likely instability leading to disruption in a tokamak, depends on nonlinear interaction of MHD instability, microturbulence, collisional (neoclassical) transport, and EP effects. Controlling NTM requires radio frequency (RF) waves. For example, NTM islands flatten the local pressure profile and modify plasma flow, thus affecting microturbulence and the neoclassical bootstrap current. On the other hand, microturbulence can affect island dynamics by regulating plasma current and electron heat conductivity along and across the magnetic field and by driving sheared flows via Reynolds stress and Maxwell stress. Energetic particles also strongly affect the tearing modes. A fully self-consistent NTM simulation must therefore incorporate nonlinear interactions between resistive MHD tearing modes, neoclassical transport, microturbulence, EP effects, and RF waves (Figure 3-5).

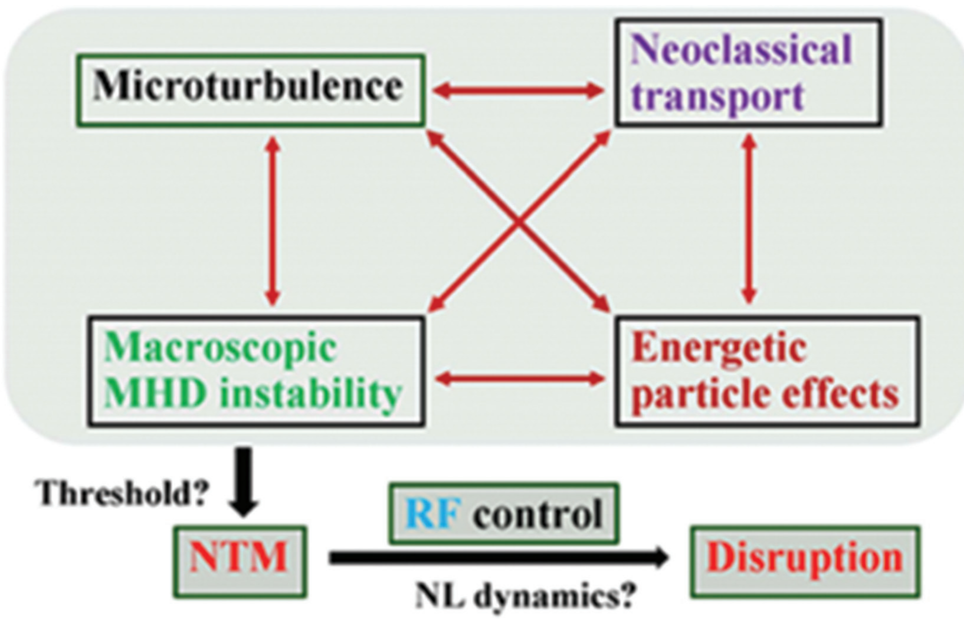


Figure 3-5. Nonlinear interactions in an NTM simulation between resistive MHD tearing modes, neoclassical transport, microturbulence, energetic particle effects, and RF waves.

Realizing a first-principles kinetic-MHD simulation of burning plasmas with multiphysics and multiscale dynamics is clearly a computational "grand challenge" problem that requires exascale computing. Currently, these processes are addressed separately by different topical fusion codes — which typically address one particular topic at a time. Thanks to productive collaborations in the framework of FES SciDAC centers and CAAR/ALCC/INCITE[2] projects, gyrokinetic particle-in-cell (PIC) simulation has recently been upgraded to incorporate all of these important physical

[2] SciDAC = Scientific Discovery through Advanced Computing; CAAR = Center for Accelerated Application Readiness Proposal; ALCC = ASCR Leadership Computing Challenge; INCITE = Innovative and Novel Computational Impact on Theory and Experiment.

processes (i.e., microturbulence, energetic particle, MHD, and neoclassical transport) in a single production version. Nonlinear simulation of RF waves in a tokamak has also been carried out for the first time by implementing the 6D Vlasov equation in the same version of the PIC code, which is currently being optimized for the next-generation SUMMIT supercomputer through the CAAR project at Oak Ridge National Laboratory (ORNL).

Such first-principles, integrated fusion simulation can readily tap the full power of exascale computers and enable kinetic simulations with a significant increase in both particle number and experimental duration as compared with the current largest production simulations on the entire Titan system. The larger number of particles (and associated spatial grids) will enable whole device simulations covering the region from the reactor core to the edge region and the material wall. The longer experimental time will enable integrated simulations of nonlinear interactions of multiple physical processes that determine the confinement properties of burning plasmas such as ITER. The integrated simulations targeted for exascale computing represent a paradigm shift and will resolve for the first time many critical issues in fusion energy research and development. These will include obtaining improved understanding of turbulent transport and associated confinement, energetic particle instability and associated losses, MHD stability with associated modifications for kinetic physics, and RF heating and current drive — as identified on many occasions previously and reaffirmed during the Exascale Requirements Review. For example, an advanced simulation goal could be the prediction of the onset of the NTM — the most likely candidate instability for disruptions. Another example is the promise of obtaining much-needed knowledge about confinement properties associated with ignited plasmas that rely on self-heating by energetic fusion products — one of the most uncertain issues when extrapolating from existing fusion experiments to future burning plasmas such as at ITER.

3.1.2.1.1 Energetic Particle Scientific Challenges and Opportunities

The confinement of EPs is a critical issue for burning plasma experiments given that high gain (or high Q) ITER operation relies on the self-heating by energetic fusion products (α-particles). Energetic particles exist in states far away from thermal equilibrium and thus readily excite mesoscale instabilities (such as Alfvén eigenmodes) through wave-particle interactions. The resulting electromagnetic turbulence can drive large EP transport, which, in turn, can degrade overall plasma confinement and threaten the machine's integrity. Because EPs constitute a significant fraction of the plasma energy density in ITER, energetic particles will also strongly influence the microturbulence responsible for turbulent transport and macroscopic MHD instabilities potentially leading to disruptions. In fact, plasma confinement properties in the ignition regime of self-heating by α-particles is one of the most uncertain issues when extrapolating from existing fusion devices to ITER.

The fusion community has made significant progress in developing comprehensive EP simulation codes and understanding key EP physics. Verification and validation have rapidly advanced thanks to close collaborations between simulation, theory, and experiment. Furthermore, productive collaborations with computational scientists (e.g., via the CAAR project) have enabled EP simulation codes to utilize current petascale computers and, hopefully, emerging exascale computers effectively. Nonetheless, more coordinated efforts and advanced computing hardware and software are urgently needed to develop first-principles, integrated simulations incorporating multiple physical processes and disparate spatial-temporal scales in order to build the EP predictive capability for ITER burning plasmas. EP transport and turbulence in ITER will also be influenced by complicated 3D effects from magnetic field ripple, ferritic materials in test blanket modules, and MHD modes. Beyond their use for fusion applications, EP excitations of electromagnetic instabilities through wave-particle interactions are ubiquitous in space and astrophysical plasmas.

3.1.2.1.2 MHD Scientific Challenges and Opportunities

Disruptions

Disruptions, the premature termination of tokamak plasma discharges through sudden loss of macroscopic stability and energy confinement, pose one of the most serious challenges to the tokamak concept for fusion energy. In ITER and future reactor-scale devices, disruptions will be capable of producing heat fluxes, relativistic beams of electrons, and forces sufficient to damage physical structures. Some of the root causes of disruptions include inadequate operations planning, failure of feedback control or other systems, and natural fluctuations that exceed the nonlinear metastability of a confinement state. Fundamental scientific questions remain about the onset and evolution of disruptions and how best to predict, avoid, and mitigate them. Many of these questions involve the interaction of diverse physical processes on multiple scales, which must be addressed through an integrated approach to modeling.

For circumstances in which disruption is not successfully avoided, ITER and future large tokamaks require mitigation to minimize damage from the following: extreme localized heating on surfaces, electromagnetic forces on structures, and the impact of relativistic "runaway" electrons (REs) that can be created during disruption. Development of effective mitigation strategies requires accurate characterization of disruptive transients. Disruptive evolution involves many effects, including nonlinear macroscopic dynamics, relativistic and nonrelativistic particle kinetics, electromagnetic responses of external structures, radiation, neutral dynamics, and plasma-surface interaction. Although numerical models of some of these processes exist, predictive simulation requires integration of the physical elements. Achieving this integration in a way that is accurate and makes use of future computational resources will require recent developments from applied mathematics and computer science.

Edge Plasma

Performance of tokamaks is strongly enhanced in the high confinement operational regime known as the H-mode. The H-mode of confinement is characterized by steep gradients in density and temperature at the plasma edge referred to as the pedestal, indicating the presence of an edge transport barrier that is associated with a drop in turbulence fluctuations. The transition from the low confinement regime, or L-mode, requires sufficient heating power that is applied to the plasma. It is generally agreed that L-H transition is caused by the suppression of turbulent transport due to ExB flow shear, and many models have been introduced for L-H transition. However, the mechanism of the L-H transition remains an open question, and a model for L-H transition capable of predicting quantitatively spatial and temporal observations together with their threshold parameters still remains to be demonstrated.

An edge-localized mode is a periodic process occurring in the edge region of tokamak plasma and arising from the quasi-periodic relaxation of the H-mode transport barrier. The significance of ELMs is in their release of short bursts of thermal energy that cause the erosion of plasma-facing components (PFCs) and can negatively affect the core plasma. A promising approach for actively controlling ELMs uses external coils to induce small perturbations, $\delta B/B_{tor} \sim 10^{-3}$ to 10^{-4}, in the edge of the tokamak. These small perturbations have a strong effect on ELM events, from mitigation to complete suppression, in some cases triggering them. However, the mechanism of ELM control in the application of resonant magnetic perturbations (RMPs) has not been consistently understood in general. The quiescent H-mode (QH mode) is an ELM-free regime that has been observed in various devices starting with DIII-D and followed by ASDEX U (Germany's Axially Symmetric Divertor Experiment-Upgrade), JET (the U.K.'s Joint European Torus), and JT-60U (Japan Torus-Upgrade). However, the mechanism of particle transport by edge harmonic oscillations (EHOs) has not been well understood, and extrapolation of the QH-mode regime to ITER remains uncertain.

3.1.2.2 Priority Research Directions

3.1.2.2.1 Energetic Particle Priority Research Directions

The following topics have been identified as having the highest priority:

1. Verification and validation of nonlinear simulations of energetic particle-driven instabilities and associated transport. Careful selection of an appropriate experimental regime is needed for nonlinear validation of simulation models.
2. Study of the interaction between energetic particles and thermal plasmas in order to develop predictive capability for burning plasma experiments. First-principles, integrated simulations at the exascale are essential to study the coupling of multiple physical processes including energetic particle instabilities, microturbulence, MHD modes, and collisional effects.
3. Development of efficient reduced models of energetic particle transport for whole device integrated modeling. First-principle simulations are essential in the development and verification of the reduced models.
4. Gyrokinetic analysis of EP physics in 3D configurations, ranging from tokamaks with 3D perturbed fields to stellarators.

Working with computational scientists and applied mathematicians is essential to developing first-principles, integrated simulations on next-generation supercomputers.

3.1.2.2.2 MHD Priority Research Directions

The following MHD-oriented topics have been identified as of the highest priority:

Specific Problems to Be Undertaken in the Disruption Area

The multitude of effects that influence the macroscopic stability of tokamak discharges and their evolution during disruptive transients leads to theoretical challenges in identifying and modeling all of the important contributions. There are challenges and opportunities in the following areas: describing the process of locking and the subsequent growth of magnetic islands, better describing the thermal and current quench phases, predicting relativistic runaway electron generation, and simulating disruption mitigation techniques.

Magnetic islands, plasma rotation, and locking. Mode locking is the process in which a non-axisymmetric magnetic field exerts torque on the plasma through interaction with external conducting structures or through an increase in viscous transport, ultimately stopping plasma rotation. Locking events generally exhibit a bifurcation in which the plasma rapidly transitions from a rotating state with a small static non-axisymmetric field to a stationary state with a large non-axisymmetric field. The transition is qualitatively described by the nonlinear theory of island penetration, which involves the balance of electromagnetic torque with viscous momentum diffusion and external sources of torque. However, a quantitative model for the onset of this bifurcation does not yet exist. The nonrotating state is highly prone to disruption for reasons that are not entirely understood. Because of the low rotation frequency expected in ITER and next-step devices, mode locking is expected to be one of the dominant causes of disruptions in these devices, as it has been in JET. The transport of angular momentum in the presence of magnetic asymmetry, how the plasma state evolves to a locked condition, and why this state leads to disruption are active research topics and represent gaps in current understanding.

Thermal quench. Apart from hot vertical displacement events (VDEs), disruptive transients typically start with the thermal quench (TQ), a rapid decrease of the plasma temperature from its pre-quench value down to several tens of eV. The timescale for this temperature decrease varies, but it can occur in as little as 1 ms in a large tokamak — hundreds or thousands of times faster than the pre-quench energy confinement time. The TQ may be caused by one or more 3D global instabilities that destroy magnetic surfaces and hence the confinement properties of the device. TQ also results when an accumulation of high-Z impurities causes radiative collapse. The rapid heat loss during the TQ produces damaging thermal loads on surrounding material surfaces. The sudden decrease in plasma temperature causes a sudden increase in resistivity that leads to the subsequent current quench (CQ) and associated large electric field that drives energetic electrons to relativistic speeds. At present, the detailed mechanism of how heat is lost during the TQ is poorly understood. It is certain that a large, free-streaming parallel heat transport along chaotic, temporally evolving magnetic lines plays a role, as does a concomitant impurity influx from the surrounding structures. Integrated simulation will be able to distinguish the effects underlying this apparently universal phenomenon.

Current quench. The magnetic energy associated with electrical current carried by the plasma is released during the CQ phase of a disruption. Low temperature following the TQ implies relatively fast resistive decay; however, the timescale is still much longer than Alfvénic times, as noted above. The TQ transient also upsets positioning control, and the plasma configuration drifts both radially and vertically (cold VDE). The motion induces eddy currents in external conducting structures. It also conducts current along open magnetic field lines into the external structures. This current has both symmetric and asymmetric components. Without mitigation, the magnetic forces associated with these currents may be sufficient to cause structural damage in ITER, especially with asymmetry driven by the external-kink instability induced by contact with the wall. Disruptions also produce toroidal rotation, and the possible resonance between oscillating wall forces and low-frequency harmonics of conducting structures would exacerbate damage. Models that reliably predict current paths and forces can help protect expensive experimental hardware by providing guidance for control, mitigation, and design.

Runaway electron generation and confinement. The TQ enhances the electric field significantly due to the large resistivity of the cooled plasma, and the enhanced field can generate an avalanche of relativistic runaway electrons. The electric field then decreases to a level near the avalanche threshold on a timescale that is comparable to the avalanche growth time. Understanding the processes of the formation and loss of REs requires continued theoretical study and improved numerical modeling to achieve the quantitative predictability needed for confidence in mitigation techniques. Areas of particular importance include relativistic kinetic effects on the avalanche growth mechanism, pitch-angle scattering and synchrotron losses of the runaways, and the stability and evolution of the runaway distribution function. There is broad scientific value in studying this topic, as it has applicability in other contexts including atmospheric events (lightning) and astrophysical and solar phenomena.

Disruption mitigation. Disruption mitigation strategies involve the injection of large quantities of impurities so that the thermal quench is dominated by radiative rather than conducted heat loss, although MHD activity has a significant role in the TQ evolution. Mitigation modeling therefore requires impurity radiation, ionization/recombination, neutral dynamics and transport, and pellet ablation. In some cases, opacity and radiation transport may be important. Like present-day experimental studies of mitigation, existing simulation results start with healthy-plasma conditions. Simulating the mitigation of disrupting conditions represents a gap in predictive capability and will require integrating the physical effects described above with those that influence the injected impurity radiation.

Specific Problems to Be Undertaken in the Edge Plasma Area

L-H transition. It is generally agreed that the L-H transition is caused by the suppression of turbulent transport due to ExB flow shear, and many models have been introduced. However, the mechanism of the L-H transition remains an open question, and a model for L-H transition capable of predicting quantitatively spatial and temporal observations together with their threshold parameters still remains to be demonstrated. Such a model would have to include a full X-point geometry, with the scrape-off layer region, as the L-H transition in the experiments is known to be strongly sensitive to the magnetic geometry.

Edge-localized modes. There has been significant progress in gaining a theoretical understanding of the linear instabilities driving ELMs; the nonlinear evolution of ELMs is a subject of ongoing computational studies. Demonstrating the full ELM cycle in simulations and reproducing quantitatively the heat and particle fluxes on plasma-facing material surfaces associated with an ELM pulse are of significant interest to the community.

Resonant magnetic perturbation suppression of ELMs. Understanding the mechanism of ELM suppression by RMPs is of great interest and importance. Most modeling attempts to date have been focused on understanding the linear plasma response to externally applied fields; some nonlinear calculations have been undertaken in the last few years, as well. There is no generally accepted model — even for the linear response. The RMP problem is probably going to be a significant computational challenge, because apparently its solution is needed to resolve small spatial structures associated with magnetic islands on resonant surfaces in the full geometry of edge plasma with the magnetic separatrix and X-point.

QH-mode (EHO). As the QH-mode is a promising ELM-free regime that may scale to ITER and other future tokamaks, it is of great interest to understand the underlying physics, the mechanism of the linear drive and nonlinear saturation of EHO, and the mechanism of enhanced radial transport associated with EHO. A physics-based model capable of quantitatively reproducing QH-mode characteristics and of making predictions for future machines would be of paramount importance.

Enhanced D-alpha (EDA) mode. Understanding the EDA operation mode is important for enhancing our knowledge of the basic tokamak plasma physics. Furthermore, the EDA regime may be of interest for future high-field compact ignition experiments. It is understood that the quasi-coherent (QC) mode is a key physical process there; however, presently there is not even a generally accepted understanding of the underlying linear instability — even less so for the QC mode's nonlinear saturation mechanism and enhanced radial transport.

Turbulence and transport in the edge plasma and the width of SOL. A model reproducing quantitatively the characteristics of transport of particles and energy in the edge plasma would be a real game changer for tokamak science. Such a model could potentially be used for predicting the width of the SOL and whether it is feasible to run it on long transport timescales or somehow couple it to a slow transport model. However, such a model would probably have to include a range of physics, from neoclassical to electron-scale turbulence and therefore involving enormous separation of spatial scales. Thus, producing such a model is going to be a computational grand challenge because of the multiscale and multiphysics nature of edge plasma turbulence and transport.

3.1.2.3 Cross-Cutting Research Directions

3.1.2.3.1 Verification and Validation/Uncertainty Quantification (VVUQ)

Verification and validation have recently become important activities for the energetic particle physics community. V&V has been carried out between several EP stability models for a well-diagnosed DIII-D experiment. In addition, a larger group of EP codes were verified for ITER high-n alpha particle-driven instabilities as part of a 2014 DOE theory milestone project. Validation for Alfvén instability models against experimental results presents unique challenges owing to the fact that the fast ion distribution function, which drives the instabilities, is not directly measured. Furthermore, there are many modes that can be destabilized and often exhibit strong sensitivities to the q-profile. These issues motivate UQ methods, which could be initiated now with reduced models; in the future, the availability of exascale computing will allow UQ techniques to be extended to the more complete models, as well. The most direct experimental measurements of Alfvénic instabilities are obtained through frequency spectrograms and associated ECEI (electron cyclotron emission imaging). Given that multiple instabilities are often present and that they persist for many oscillation periods, the fluctuation and imaging data can lead to the accumulation of large data sets. Making comparisons with modeling can be facilitated through data mining techniques, along with the addition of synthetic diagnostics to the simulation models. Alfvénic instabilities can also increase the transport of fast ions out of the plasma, leading to loss of heating efficiency and the potential for PFC damage; measurement of these losses is typically performed at only a few locations on the vacuum chamber wall and over limited regions of energy/pitch angle space. Simulation of and validation of these losses can lead to confidence in the prediction of losses and the potential for damage at other locations where measurements are not made. In order to further develop understanding and reduced models for EP transport, it is useful to map out the dominant resonant particle interaction regions in a 5D phase space. Efficient interpolation/fitting methods are needed to project these into either 3D (for the tokamak) or 4D (for the case of stellarators and 3D tokamaks) constant-of-motion spaces.

3.1.2.3.2 Multiple-Timescale Simulation, Multiple Physics Coupling

Integrated simulation methods need to be developed for EP-driven Alfvén instabilities and EP transport. We need to determine EP transport due to multiple 3D perturbations including MHD modes, Alfvén eigenmodes, and external 3D perturbations. In particular, we need to bridge multiple timescales ranging from the fast Alfvén time to the slow confinement time. It is very difficult to carry out long time simulations of EP transport using the first-principles-based gyrokinetic model owing to its prohibitive computational cost. One way to bridge the multiple timescales is by developing reduced simulation models that describe the slow evolution of mode amplitude and EP profile. One such model is the quasi-linear model recently developed for EP transport. Another way is combining appropriately the first-principles models and reduced models for the long time transport simulation. The initial fast onset and saturation phase can be modeled with a first-principles model, whereas the slow, quasi-steady-state phase can be treated with a reduced model. A third approach is to use gyrofluid closure models, which can be extended to include many of the same mechanisms as the more complete models but at a lower computational cost. These approaches have yet to be verified and validated. An important element of VV is studying the extent to which the reduced models are justified in applications to the experiments.

3.1.2.3.3 Linear Algebra Solvers

Both energetic particle and extended-MHD codes use external libraries that perform linear algebra operations such as matrix-free solves (e.g., PETSc or Trilinos) and associated sparse-matrix solver libraries that are commonly employed as part of a preconditioning strategy (e.g., SuperLU_DIST, HYPRE, MUMPS, PaStiX). Some of these libraries are supported by ASCR scientists. As the underlying computational architectures change and the parallelism model is diversified from MPI-only to MPI+X, it is essential that new implementations (and possibly new algorithms) are developed that provide the same functionality.

3.1.2.3.4 Domain-Specific Solvers

Present sparse-matrix solver libraries do not scale to exascale-size problems. As the computing architecture transitions to many-core machines, this limitation becomes more problematic for scaling. As sparse-matrix solvers are often used within block Jacobi preconditioners, parallel-scaling and memory usage improvements would improve overall computational efficiency. One potential avenue that can be exploited to make these advances is to tune solvers to the specific matrix properties of codes. For example, the solvers could exploit matrix properties related to structured dimensions of the computational mesh in addition to existing general-purpose implementations. For example, the 3D finite elements used in M3D-C1 are unstructured in two dimensions but structured in the third dimension. One could perhaps take advantage of this structure in a customized solver by applying a geometric multi-grid preconditioner in this direction only and using other techniques in the remaining directions. Implementing this approach would require an applied mathematician who is an algorithm expert to interact closely with the physics groups that developed the code to explore more effective algorithms.

3.1.2.4 Computing Needs and Requirements

Our computing needs and requirements can be placed into the following three categories.

3.1.2.4.1 Data / I/O

Fusion simulations can generate enormous amounts of data; for example, a simulation for MHD science can use 10^7–10^9 grid points, 5 to 10 variables, and 10^3 time steps, which would amount to ~100 GB in RAM and ~100 TB on disk. As computers become more powerful, increasing amounts of FLOPS can be processed; however, the amounts of data for the anticipated exascale MHD simulation challenges will not fit in memory so the storage systems need to keep pace with computing improvements — otherwise, the fast machines of the future will be wasting time waiting for the storage systems to deliver the data. The trend in HPC is use of multicore architectures, which increases the concurrent load that can be sent to the storage system. On the other hand, the availability of extra computing power can be used to optimize the I/O path and make I/O operations faster. Parallel file systems distributing the data of a single object or block of data across multiple storage nodes will probably continue to be used in the HPC environment in the future, with the degree of parallelization increasing to meet the increasing pace of computing.

3.1.2.4.2 Need for Support of Ensemble Computing and Parameter Studies

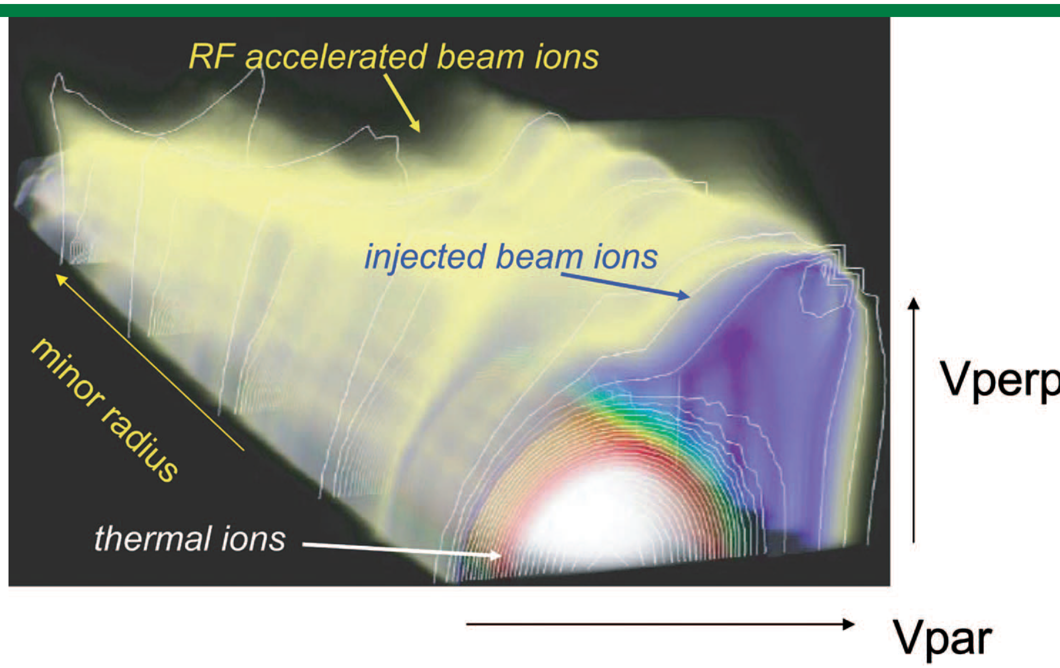
While exascale computing will enable the integrated kinetic simulations described in this report, 3D magnetohydrodynamic codes such as NIMROD and M3D-C1 scale well enough to 10,000 processors, which is normally enough processors to efficiently process the kinds of spatial grids we need to represent today's fusion experiments with sufficient resolution. However, as discussed in the case study, "Simulation of a Disrupting Tokamak Plasma" (Appendix D), the large timescale separation between stability and transport phenomena necessitates calculations spanning a very long transport time to study the onset and eventual saturation or other termination of a global event such as a disruption. Even though these MHD codes already use advanced and fully implicit time stepping, the simulation requires hundreds of wall-clock hours to perform a realistic simulation and at least 1 million CPU-hours. These runs are now normally performed as a series of restarts, each taking 10–20 wall-clock hours. For this workflow to be scientifically productive, we need the time between completion of one job and restart of the next to be on the order of hours or a day at most. Because of the high demand at centers like NERSC, wait times between jobs can be very long, which has a negative effect on our science.

It would greatly improve our productivity if there were resources available to run jobs at the 10,000- to 20,000-processor level that could support the needs of individual groups — an approach that is often called capacity computing. Because we are simulating a very complicated physical system and are unsure of the exact initial conditions of the physical system and the sensitivity to these conditions, we need to perform many simulations where we systematically vary parameters. This testing outlook is very similar to that of experimental facilities where they run numerous discharges in order to understand dependencies. Quick and seamless access to intermediate-scale resources can also improve the readiness of users for runs at extreme scales and facilitate code development. MHD codes also run more efficiently with larger memory per node than is typically available on leadership systems, so that access to architectures with larger amounts of memory may be very cost effective for these codes.

3.1.2.4.3 Awards Program Criteria

One aspect that precludes progress by implicit MHD codes is the emphasis by some computational awards programs on algorithmic scaling (e.g., scaling per time step or solver iteration). The ultimate metric may instead be resources used relative to the science goals achieved. If scaling data is required as a proof of concept for resource utilization, time-to-solution scaling on the proposed problem should be chosen, not algorithmic scaling. Algorithmic scaling, which is useful internally to a project to characterize code-kernel performance, does not provide a full picture of time to solution. For example, the time-step size typically decreases when explicit algorithms scale weakly with a fixed domain size in order to avoid numerical instability; but this decrease is not reflected in a computational cost per time-step plot. These arguments are not against the use of explicit (or other) algorithms, which are well suited for certain classes of problems; rather, we argue that the current system does not permit an apples-to-apples comparison of the capabilities of different codes.

A VIEW OF THE SCIENCE



Simulation of the interaction of high harmonic fast waves with fast ions from neutral beam injection in the DIII-D tokamak using the combined full-wave/Fokker-Planck model AORSA/CQL3D. Reproduced from Jaeger, E.J., L.A. Berry, S.D. Ahern, et al., *Physics of Plasmas* 13**, 056101 (2006).**

3.1.3 RF Heating and Current Drive

3.1.3.1 Scientific Challenges and Opportunities

The success of magnetically confined nuclear fusion as an economically attractive source of power largely depends on the success of ITER, the next step device currently under construction in France. ITER in turn relies on the successful operation of three plasma heating technologies (Figure 3-6). Of those, two are based on external application of radio-frequency (RF) power. These are in the electron- (GHz) and ion-cyclotron (MHz) range of frequencies and are here referred to as electron-cyclotron heating (ECH) and ion-cyclotron heating (ICH), respectively. The primary science driver in this area is the robust, reliable, and efficient operation of these systems to enable a successful ITER mission, and in the long term, steady-state operation and control of fusion-based reactors (e.g., DEMOnstration Fusion Power Plant [Stork 2009]). However, it has been observed on present devices that the operation of ICH correlates with the production of impurity ions from increased interactions between the plasma and its confining structures (i.e., plasma-material interactions or PMI), which can have deleterious effects like collapsing the plasma. As such, the scientific challenge here is to obtain a fundamental understanding of the coupling of externally applied RF power to fusion plasmas and to use that understanding to mitigate the PMI issue, thereby making available the required robust RF technology that a successful fusion mission requires. In a practical implementation, the fast timescales and extreme environment make diagnosing RF-related phenomena experimentally incredibly difficult, and as such, much of the understanding in this area relies on computer simulation. It is clear that this trend will continue, and that exascale computing resources will be one of the key tools in meeting this challenge.

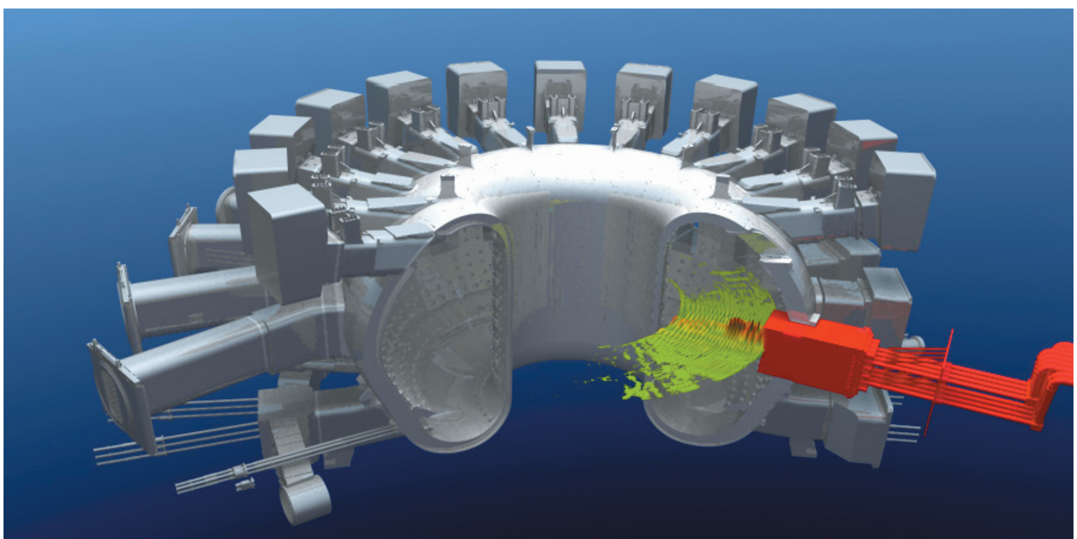


Figure 3-6. AORSA simulation of the ICH antenna system on ITER (Source: Jaeger 2008).

The physics basis for how the application of ICH power drives plasma waves and enhances the electric potential that exists between the plasma and any confining material structure (the sheath) is thought to be understood, as is the basis for how materials respond to the bombardment of ions accelerated by that sheath potential. However, implementing these understandings in predictive and reliable computational models that have the required fidelity and dimensionality to be directly validated with experiment is only now becoming possible. In the application of RF power, experimental observations (Wilson and Bonoli 2014) have made it clear that it is the details (both geometric and in the physics model) that are important in determining whether that power will heat the plasma to fusion or whether it will burn a hole in a wall tile and collapse the plasma. As such, the present and future state of the art focuses on building reliable simulations that couple the required pieces, all at the required fidelity. In the 5- to 10-year timeframe, exascale computing resources present an opportunity to utilize a linear, kinetic plasma-wave solver that resolves not only the tens of cubic meters of the core fusion plasma at the fastest (nano- through microsecond) timescale but also the millimeter-scale features and nonlinear field response near the launching antenna structures and confining material walls (Figures 3-7 and 3-8), together with models of the material response to plasma bombardments. Indeed the transport of sputtered impurities through the edge plasma that ultimately affects core fusion performance is a multi-time-scale and multiphysics process that couples the RF, PMI, SOL/divertor, pedestal, and core problems. At the material surfaces, many atoms are quickly ionized and are on gyro orbits that intersect the wall. These are promptly redeposited, but the material created as a result is often loosely bound and has different mechanical and thermal properties. For the ions that escape that fate, subsequent transport is highly dependent on local conditions and the 3D geometry of the machine. The fraction of sputtered atoms that end up in the core plasma is a strong function of source location through processes that are poorly understood. Once in the core, the impurities are subject to turbulent and collisional transport processes that we are only just beginning to understand. Understanding this multiphysics process will allow us to design strategies to mitigate the interaction with material surfaces, while maximizing the heating efficiency and reliability.

In parallel to the above challenge is the impact that the application of RF power has on, and how itself is affected by, plasma turbulence, MHD instabilities, and energetic particle populations, that is, the timescales that exist between the RF and transport scales. For the control of MHD instabilities, RF actuators have long been recognized as tools that will be essential for realizing a steady-state tokamak. How RF power interacts with turbulence and energetic particles is less clear. However, what is clear is that the proper design of reactor-grade, steady-state tokamaks involves coping with a complex interplay of the effects of transport, external current drive and heating profiles, MHD stability, and control of edge density and temperature pedestals and scrape-off plasma parameters. While great strides have been made in developing the modeling capability for most critical areas, very little progress has been made in modeling the whole device: thus, the exascale computing resources represent an opportunity to integrate the advances that have been made in transport, core and edge MHD, RF current drive, and SOL simulations in order to determine optimal reactor configurations and operating scenarios.

A stated top-level goal for DOE's Office of Fusion Energy Sciences (FESAC 2014) is the use of massively parallel computing for validated predictive simulation for magnetically confined fusion plasmas. This capability should ultimately enable, and minimize the risk in, future fusion energy development stages. A subset of this goal is the integration of independently developed computational tools that make up the DOE portfolio of legacy and state-of-the-art simulation codes. For many years, RF source modules have been employed as components within integrated simulation (e.g., within TRANSP [Hawryluk 1980]), and the RF SciDAC program has produced both high-fidelity and reduced models for many aspects of simulating the application of RF power, with the goal of creating a predictive and robust tool that will bring the coupled antenna-to-core system to within reach. This opportunity and the challenge of integration are dealt with more thoroughly in the whole-device-modeling section of this report (Section 3.1.4).

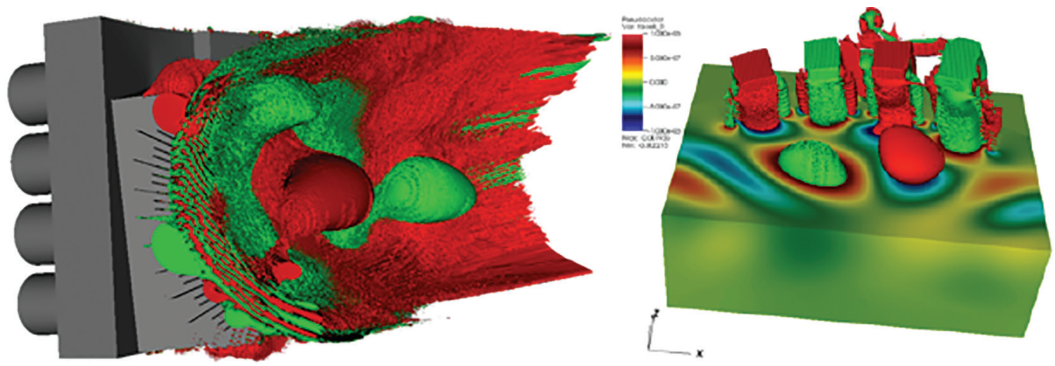


Figure 3-7. 3D Vorpal simulations of the Alcator-CMod field-aligned antenna (Figure 3-8). Slow (small-scale) and fast (large-scale) blob waves are evident, illustrating the multiple spatial scales present in one small volumetric piece of a tokamak. Left: low edge density; right: high edge density (Source: Jenkins and Smithe 2014).

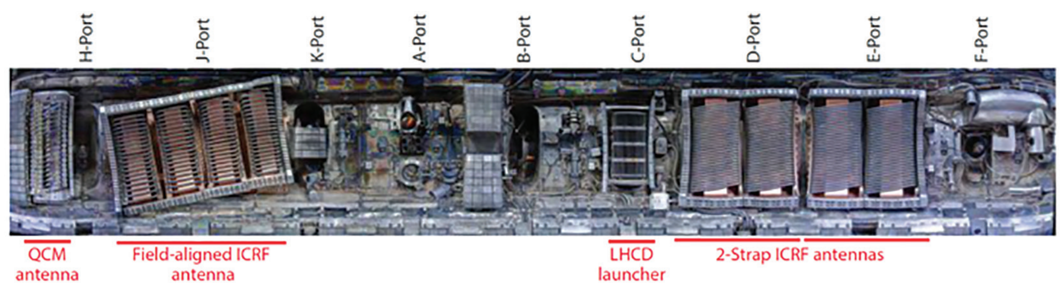


Figure 3-8. Panoramic view of the plasma-facing surfaces in the Alcator C-Mod tokamak. This illustrates the fine geometric detail required, as well as the alignment (magnetic and nonmagnetic) of two ICH antennas (J-Port [top left] and D/E-Port [top right]).

In principle, the RF challenge (and subsequently the entirety of the plasma physics occurring at longer timescales in fusion plasmas) may be met by solving the full 6D + time Vlasov-Maxwell coupled system (with an appropriate collision operator). However, with a naive, explicit time advance approach, this would require time steps of the order 10^{-11} seconds in a system where the resulting impurity transport occurs at milliseconds on a mesh that has to resolve the Debye length throughout the device, meaning many trillions of spatial points. Without suitable application of some (yet-to-be-created) multiscale time integration scheme or the like, such an approach likely exceeds even the exascale, and this still does not incorporate the auxiliary equation set for the material response to the incident plasma bombardment, or generated neutron flux, or interaction of these with any of the myriad of engineering components required in the operation of any fusion reactor. As such, we therefore expect a continued reliance on scale separation combined with integration within the 5- (and perhaps even 10-) year timeframe, although we continue to keep our eye on possible avenues of computer science and applied math that may enable this 6D approach to be incorporated into a comprehensive RF predictive capability.

3.1.3.2 Priority Research Directions (PRDs)

Following are the driving RF-focused physics research directions and descriptions of the computing landscape resources that will help aid this research in the 2020–2025 timeframe. Further discussion of these needs in terms of a global computing ecosystem follows in Section 3.1.3.4, "Computing Needs and Requirements."

3.1.3.2.1 Predict How Much RF Power Is Coupled to the Core Plasma

Here we note that this PRD matches "PRD-Boundary-4" from the FES report on integrated simulation (FES and ASCR 2015). It deals specifically with accounting for not only the wave physics in the well-confined core plasma but also including all possible mechanisms through which applied power may be lost before reaching its target.

Physics Model Requirements

- High geometric fidelity, including coupled antenna-to-core simulations resolving interactions of RF waves/antennas with the SOL plasma.
- Nonlinear RF interactions may be important, specifically, RF sheaths and decay of the applied single frequency power into multiple frequencies via nonlinear parametric decay.
- Self-consistency with the SOL plasma, which will require high fidelity for quantitative predictions (e.g., resolving plasma turbulence features and / or blobs) and should support 3D RF transport and impurity models.

Computational/Algorithmic Requirements

- High geometric fidelity here means that a domain of limited extent around the launching RF structure requires approximately 10 billion grid cells (for something like the National Spherical Torus Experiment Upgrade [NSTX-U] fast-wave antenna). A finite-difference time-domain simulation would typically be run over 11,520 nodes on the Titan at ORNL for 24 hours to simulate roughly 200 RF antenna cycles; the total charge for such high-fidelity simulations is 10 million CPU hours.
- Perhaps a full 6D linear delta-f type particle or continuum approach can be used for the entire SOL to capture parametric decay instabilities to the kinetic ion Bernstein wave (IBW). Even in 1-space/3-velocity types of particle-in-cell calculations (e.g., Jenkins et al. 2013), these waves have been difficult to resolve due to particle noise, so perhaps high dimensionality, time domain continuum approaches are better suited. Such continuum approaches require gridding a high dimensional space (likely at least 5D with 2-space/3-velocity for RF problems of interest), which even for small domains at modest resolution means more than 10 billion grid points.

3.1.3.2.2 Mitigate the RF-Induced Plasma-Material Interactions

The usefulness of externally applied RF power is limited if the plasma-material issue cannot be resolved. We expect that the state of the art is such that the pieces of the required simulation are now available (or at least the physics basis is), and that the goal should be to put all of the required physics into the same (perhaps coupled) simulation.

Physics Model Requirements

- Coupled edge (and possibly core), material response, and subsequent impurity transport.
- Both RF and material codes to use a (the same) rectified sheath model, probably calculated by the RF code, and used to determine fluxes to the surface material.
- Binary collision approximation (BCA) codes to give sputtered yield.

- The larger values of the RF sheath (perhaps in and around the antenna) may require further following of the evolution of the material structure leading toward erosion, flaking, etc. Investigate whether antenna material lifetime is an issue and quantify the difference between inboard and outboard launch locations.
- Multiple impurity species transport.
- Include both the near- and far-field regimes (which means a larger simulation volume).
- RF to provide heat source into SOL fluid transport solvers.
- RF breakdown, which has typically not been included in such models.
- All on real geometries (e.g., Figure 3-8).

Computational/Algorithmic Requirements

- We point out that computational requirements for this section are in addition to those required in Section 3.1.3.2.1, as this section would be extending the above simulation approach to include a material response and longer timescale transport.
- Additional demand may exist from the BCA code requirements.
- High fidelity (3D) production simulations of the RF-sheath will be employed to complete our understanding of how the sheath potential is configured for various design and operational choices of the RF systems. These calculations are either of finite-difference or finite-element in time or frequency domains, at the order of 10^9 (~1 TB) to 10^{12} (~230 TB) grid cells in order to represent the immediate area around the ion cyclotron resonant heating (ICRH) antenna and the larger area of the entire vacuum vessel interior, respectively. For the finite-element method approach, iterative sparse matrix solvers at this scale are an active area of development, as are preconditioners to aid that iterative process. Alternatively, the time-domain methods avoid these at-scale matrix inversion issues but must advance many timesteps while obeying the so-called Courant-Friedrichs-Lewy (CFL) stability constraints (so using timesteps of the order 10^{-12} s) to reach the desired steady-state response (tens of RF cycles: so 10^{-6} s, or 10^6 timesteps).

3.1.3.2.3 Understand RF Power Interaction with Core Energetic Ion and Electron Populations

Here one of the key questions to be answered is whether or not the application of RF power is likely to destabilize or stabilize energetic particle modes. It is also crucial to understand the extent to which the ICH power required for bulk plasma heating in burning plasmas will interact parasitically with fast particle populations already present in the plasma (e.g., fast ions from neutral beam injection [NBI] or fusion alpha particles).

Physics Model Requirements

- ICH and lower hybrid (LH) current drive simulations create, and depend on, a 5D plasma distribution.
- It will be necessary for these models to assess fast ion orbit width effects and non-diffusive velocity space effects, which can affect the loss of fast ions accelerated by the ion-cyclotron range of frequencies (ICRF) power.
- In the lower hybrid range of frequencies (LHRF), it will also be necessary for these models to account for full-wave effects, such as diffraction and focusing in the wave propagation.
- Synthetic diagnostics, which make use of the simulated nonthermal distribution function to validate combined wave propagation/Fokker-Planck models with measurements of hard X-ray emission, photon counts from a neutral particle analyzer, fast ion D-alpha emission, and RF wave fields detected with reflectometry and phase contrast imaging techniques.

Computational/Algorithmic Requirements

- Typical model simulations for ICH wave interactions with energetic particle populations employ a full-wave field solver coupled to a continuum Fokker-Planck code or a Monte Carlo orbit code. In the lower hybrid frequency range, wave interactions with energetic particles are described by a coupling of ray tracing, or more recently, full-wave field solvers to continuum Fokker-Planck codes. The connections between the wave codes and particle codes are established through either a diffusive or nondiffusive (non-Gaussian probability distribution function) RF operator. The dielectric response in the full-wave solver is evaluated using the nonthermal particle distribution. The wave solvers and particle codes are typically iterated in time.
- The core and wall-clock hour requirements for generic coupled full-wave/Fokker-Planck simulations are dominated by the 3D full-wave field reconstruction. For ICH, a typical breakdown is (5000 cores/toroidal Fourier mode) × 2 hours × (50 toroidal modes) × (10 iterations with a Fokker-Planck solver or Monte Carlo code) = 5,000,000 CPU hours per run (on today's HPC systems). In the helicon and lower hybrid regimes, typical 3D field reconstructions require (15,000 cores/toroidal mode) × 1 hour × (20 toroidal modes) × (20 iterations with Fokker-Planck solver) = 6,000,000 hours per run. The matrix to be inverted in an LHRF field solve can be as large as 10–20 TB. It is important to note that for both cases, each toroidal mode simulation is independent and can be executed concurrently with little penalty for tolling if, for example, 200,000–300,000 cores are available. Thus, the 3D field reconstruction is a problem that benefits enormously from capacity computing.
- It is also expected that more efficient 3D solutions (two velocity space dimensions and one configuration space dimension) of the Fokker-Planck equation will be realized during the 2020–2025 time period. A challenge in this area is the need to develop efficient algorithms (either direct or iterative) for inverting the sparse/ill-conditioned matrices produced by the 3D Fokker-Planck solver.

3.1.3.2.4 Use RF Waves for Plasma Control

Magnetically confined plasmas are subject to a range of unstable modes. Often these modes can be stabilized by driving current in the right place. This PRD focuses on developing the computational tools to simulate this process, ultimately being able to incorporate such RF actuators for stability into a larger whole device model.

Physics Model Requirements

- Sawtooth control via energetic particle population created via ICH.
- NTM control via localized control of the current profile via electron-cyclotron current-drive (ECCD).
- MHD-produced 3D effects in the equilibrium and profiles (including the fast-ion profile) need to be resolved in the RF calculation (i.e., "real 3D").
- Impact of ECH and high-harmonic fast-wave (HHFW) on reverse shear (fast particle driven) Alfvén eigenmodes (RSAE modes). RF alters the fast ion distribution function, or modifies the shear, which subsequently changes the driving of the mode. If accomplished via the shear, then the coupling is only through the equilibrium; but if it is via the distribution function, the coupling is more complex.

Computational/Algorithmic Requirements

- Codes exist for both MHD (slower timescales) and ICH. Coupling via passing the full 4D or even 5D distribution function (or some parametrization) is required.
- 5D Fokker-Planck solver or Monte-Carlo particle-based code.

- The MHD 3D effects would likely involve a time-dependent coupling of an RF code into the hybrid MHD calculation — something similar to the Heterogeneous Multiscale Method (HMM) (Weinan 2007), or some other multiscale time integration method.
- Coupled MHD-RF algorithms currently exist for simulating the stabilization of neoclassical tearing modes via ECCD; for example, the coupling of the nonlinear 3D MHD code NIMROD with the GENRAY ray-tracing code (Jenkins et al. 2016). The CPU requirements in these coupled calculations are dominated by the MHD solver, with a single simulation using 35,000 cores for 36 hours (in wall-clock time) and producing 6 TB of data.
- The computational requirements for a coupled nonlinear extended MHD/ICRF simulation will likely be determined by a number of factors: the timestep constraints for extended MHD modeling are slightly more stringent than what is needed for ICRF; however, the spatial resolution needed to ensure that the RF physics are correct will be comparable to what is used for very well-resolved poloidal planes in NIMROD, for example, but extended to 3D. This is because extended MHD does not care about the toroidal direction very much, but when the ICRF antenna is included, there is then a quasi-toroidally localized source introducing features with comparable toroidal and poloidal resolutions.

3.1.3.3 Cross-Cutting Research Directions

Here we present PRDs that, while important to RF, apply generally to all of fusion simulation. We also note that three of the four points in the above section also include a significant portion of cross-cutting research and that this is indicative of the importance of RF in a practical implementation of a fusion device.

3.1.3.3.1 Perform Rigorous Verification and Validation with Uncertainty Quantification

Within RF, as in all other areas of fusion simulation, there are many examples of verification, especially cross-code verification. For example, ray tracing codes can be compared with full-wave calculations, and continuum Fokker-Planck codes can be compared to Monte Carlo orbit codes. However, there is a far lesser amount of validation (with experiment), and that is typically limited to one-off or single representative (but not really) cases. The same is true for sensitivity and UQ analysis. Perhaps within RF exceptions are: the GENRAY-CQL3D study for a 3D space (T_e, n_e, I_p) where LHRF power density and current density profiles from 880 simulations were tabulated in a look-up-table for the EAST tokamak (see Section 3.4, Verification and Validation); and the validation studies of ICRF mode conversion simulations using AORSA-CQL3D and TORIC using a phase contrast imaging diagnostic (Tsujii, Porkolab, and Bonoli et al. 2015).

Physics Model Requirements

In terms of RF control of instabilities, it may come down to a need for look-up-table types of reduced models for real-time control systems (or neural nets). This means knowing which parameters are the most sensitive, an effort that can be nontrivial in distilling this information from models that have many inputs.

Computational/Algorithmic Requirements

How do we do UQ? Two approaches come to mind: (1) brute force as in ensemble calculations that are non-intrusive, or (2) adjoint-like methods that are intrusive (such as by calculating gradients of certain code quantities). Automatic differentiation lies between these two methods (which tries to calculate the Jacobian of your code), which also has limitations. A standard approach to fusion code UQ would be beneficial.

3.1.3.3.2 Acquire Ability to Perform Whole Device Simulations

While we mention whole device simulation here, we refer the reader to Section 3.1.4, Whole Device Modeling, for a more thorough discussion of the computational needs associated with fusion whole device modeling.

Physics Model Requirements

- Rigorous verification and validation of model hierarchies for wave propagation, absorption, and coupling through synthetic diagnostic comparisons with experiment.
- Understanding of the sensitivities of models that comprise the hierarchy, as well as the sensitivity of synthetic diagnostic results to variations in the equilibria, kinetic profiles, etc.

Computational/Algorithmic Requirements

While the coupling of many codes of all scales can be envisioned (and some are already accomplished), one of the foreseeable difficulties is having such simulations, with their many contributing models and each with its own sensitivities, produce robust and reliable results that have an associated (and rigorous) uncertainty. As such, given some set of black box type models/codes, a framework for propagating uncertainties and sensitivities throughout such couplings may be required.

In recent years, implementation of advanced RF modules has occurred in computational frameworks for integrated modeling such as the Integrated Plasma Simulator (IPS) (Batchelor 2009) and TRANSP. These simulations typically employ (at a minimum) 240 cores for electron cyclotron ray tracing (GENRAY), 240 cores for lower hybrid ray tracing (GENRAY), 128–256 cores for the ICRF solver (TORIC), and 48 cores for 3D Fokker-Planck code. The RF computational requirements are in addition to the 48–64 cores needed for the Monte Carlo beam orbit code (NUBEAM) and the 512–1024 cores needed for an advanced reduced model for thermal transport (the gyro Landau fluid code TGLF). The wall-clock times needed for a single time-dependent simulation can range from several hours to greater than 48 hours, depending on the time duration of the discharge that is to be simulated and the level of physics fidelity (resolution) required in the physics components.

3.1.3.4 Computing Needs and Requirements

In this section, we present identified computing ecosystem needs, grouped together in similar categories, with some examples linking back to the physics PRDs in the two previous sections.

3.1.3.4.1 ASCR Collaboration

Sometimes the embedding of applied math or computer science expertise within science projects reveals tools and methodologies that would otherwise go unnoticed. For example, the interaction of applied math experts with the physics community at the review generated several possible avenues of future research, and having this expertise available (or expanding it) at the project level would be invaluable. Several aspects of RF simulation development are listed below that lay outside the physics expertise and that we identify as needs where collaboration with ASCR may greatly assist.

- Accelerated inversion libraries/preconditioners for large ill-conditioned, sparse matrices.
  Example: At present, most (if not all) full-wave RF codes rely on direct solvers. Here there is overlap with MHD, where they have had success with preconditioning for iterative inversion of sparse matrices in 3D (the iterative solvers use direct solves on the matrices associated with each poloidal plane as a preconditioner).
- At-scale geometry/dispersion adapting meshing and domain-decomposition technologies.
  Example: Moving to "real-3D" for full-wave codes is not practical at fixed resolution. Nor is the existing fixed resolution a requirement, but rather an artifact of the chosen algorithm for ease of implementation of the particular physics.

3.1.3.4.2 Algorithm Development and Code Re-Engineering

There are two aspects here; the first is rethinking our problems and considering new algorithms that may be well suited to exascale platforms (but perhaps not for desktop platforms), and the second is re-engineering existing codes (i.e., refactoring for performance/new platforms).

- Either implement simpler methods/programming environments/software stacks to enable code reengineering issues raised by new computing ecosystems or expand support for expertise to assist in doing so.
  Example: Performance portability is likely to be required for codes that will form part of an integrated simulation (i.e., if the performance is not portable to the platform chosen for the integrated simulation, then the component is not applicable for use).
- Support to investigate development of new RF algorithms that exhibit improved scaling with resolution that will ultimately enable effective use of an exascale platform to do the required science. Perhaps a published review of the types of algorithms and/or best practices to keep in mind when targeting the exascale.
  Example: Trade FLOPS at low numbers of degrees of freedom for improved algorithmic scaling with high numbers/dimensionality (e.g., to move from 2D to 3D in kinetic full-wave codes).

3.1.3.4.3 Workflows, Data, and Visualization

- Improve support for HPC code coupling frameworks, perhaps ones that facilitate tight (in-memory as opposed to file based) as well as loose coupling of independently developed codes.
  Example: Perhaps documentation and examples on best practices for these, or some standardized approach to the coupling and managing of HPC codes on DOE compute systems.
- Continued support for workflow managers running on the service nodes for coordinated code execution on compute nodes.
  Example: Something similar to DAKOTA but for verification and validation.
- Continued/improved support for interpreted languages.
  Example: Python for the use of OMFIT and the IPS integration tools.
- Technologies for large data set exploration/debugging, including *in-situ* analysis to avoid large I/O and visualization data transfers that impede scaling.
  Example: Non-fatal bugs that appear only at scale are typically best tracked down via visualization of the results. In such situations, it would be helpful to explore data in an intuitive manner (as opposed to manual rendering at full resolution).

3.1.3.4.4 Software as a Service

In RF and fusion simulation in general, as more of the individual compute codes mature they become candidates for inclusion in a larger whole device model. The trend here is that larger and larger compute-capable component codes are being included in such models, with more of the community wanting access to such capability. One challenge is making this capability available while avoiding the overhead associated with having individuals build, set up, and maintain their own WDM software stack. The software-as-a-service type of model represents one approach to solving this problem, and it has proven successful in previous integrated modeling efforts (e.g., TRANSP). Issues to consider going forward are:

- Many users (perhaps tens to hundreds in the 2025 timeframe) running simultaneously at TF (TensorFlow) or less rate.
  Example: If this model is eventuated, it may be the case that there is always some portion of an HPC resource being utilized by this capability. This usage would have an impact on the possibility of allocating dedicated queues (mentioned below).
- User space/project space types of issues.
  Example: This capability means allowing a user of some user-facing submission system to submit a job that is ultimately submitted and run by the project account and did not require the actual user to have an account on the compute system.
- Dedicated queues on national facilities for software-as-a-service types of workflows.
  Example: This capability means establishing a real, or near real-time, turnaround on interpretive simulations between experimental shots such that HPC model-based interpretation of each shot can inform the next.

3.1.3.4.5 Computing and Programming Environment

- Continued or expanded access to fast job turnaround at moderate concurrency to support rigorous validation and uncertainty quantification of models (probably as a requirement of any contributions to a community whole device model). The use case of running a particular code many times for varied input parameters is likely to become more common.
- Training, specifically student training to build the next generation of leadership computing-capable science experts.
- Common programming model/environment across facilities. While we recognize the difficulty in providing this environment, we imagine that at a minimum, a set of best practices for programming in scientific codes could be made available that result in (1) minimal time invested when porting from one machine to another, (2) minimal time required for a computer scientist to learn a new code, and (3) minimal wasted efforts due to codes falling to legacy status. At more advanced levels, this common environment or approach may be in the form of offering advanced software engineering tools that perform code re-engineering or provide suggestions for doing so.

.

3.1.4 Whole Device Modeling

3.1.4.1 Scientific Challenges and Opportunities

Several factors — including the high cost of building future experiments and prototype fusion facilities, such as the ITER device now under construction, the complexity of the multiscale multiphysics in fusion plasmas, and the advances expected in extreme-scale computing over the coming decade — provide strong motivation for developing integrated simulation capabilities to predict key physical processes occurring in tandem in these devices. Such capabilities can ultimately enable predictive simulation of an entire fusion device, thus minimizing risk to the device and guaranteeing its successful operation. WDMs are required to assess reactor performance in order to minimize risk and qualify operating scenarios for next-step burning plasma experiments, as well as time-dependent or single-time-slice interpretive analysis of experimental discharges.

Figure 3-9, adapted from the Integrated Simulation Workshop report (FES and ASCR 2015), provides a high-level view of the WDM. Highlighted in the figure are the levels of complexity — or physics hierarchy — constituting a WDM that span reduced models to extreme-scale kinetic physics models. The former can be analytic or physics based, should typically be computationally fast, and can run on a reduced number of processors for fast *in-situ* jobs in direct connection with experimental runs. The latter extreme-scale models provide a deeper physics understanding, need to be first-principles-based codes, and typically require capability computing. Into this category fall gyrokinetic codes coupled with multispecies fluid/MHD, RF, and materials codes for fast timescale physics. In between the two are the so-called advanced reduced models. The fitting parameters in these models are typically derived from ensemble results of extreme-scale calculations over a finite range of plasma parameters and might be applicable to only a limited operational space. Advanced reduced models are not as fast as reduced models, but they can still be used in time-dependent simulations for higher-fidelity reduced calculations. A WDM plan should provide the flexibility to choose among levels of physics hierarchy depending on the needs, which can span from *in-situ* experimental planning and analysis (simulations requiring only a few minutes) to fundamental physics understanding and prediction (simulations requiring a few hours to a few days).

The choice of multiple components will facilitate the verification and validation of individual physics models and the verification of integrated physics. The VVUQ of a reduced-model WDM will rely mostly on the **availability of a large-scale computer ecosystem** owing to the need to carry out large ensemble calculations. VVUQ of a high-fidelity, extreme-scale WDM still remains an open research topic in the ASCR community. Multiphysics processes are coupled together in a nonlinear manner. Thus, studies on the individual components may not present much reflection on the integration result.

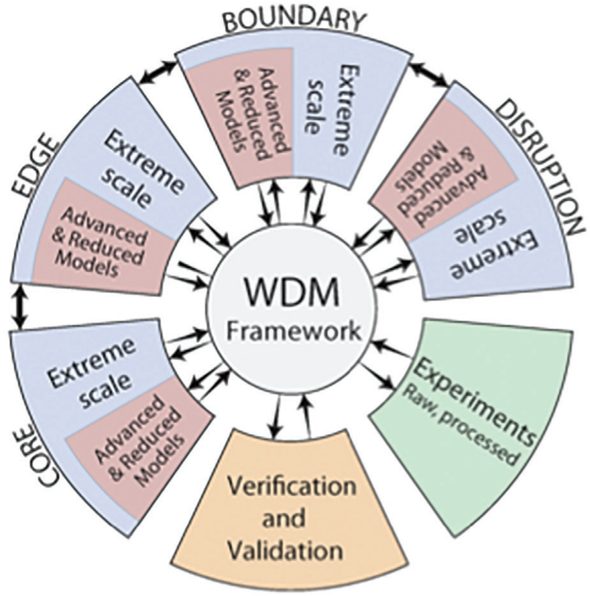


Figure 3-9. Graphic overview of the WDM adapted from the *Workshop Report on Integrated Simulations for Magnetic Fusion Energy Sciences* (FES and ASCR 2015). The multiphysics aspects of the WDM are exemplified by the plasma regions that include the core plasma, edge plasma, and boundary. The physics phenomena of interest in each area are described as a hierarchy of models ranging from the extreme scale to reduced models. Also shown is the envisioned interaction between topical areas, V&V technology, and experimental data, along with the essential connectivity provided by the WDM framework.

Special attention should be directed to the load balancing among each component, alternative coupling schemes, component interchangeability, regression analysis, and exceptions handling. Collaboration with ASCR is critical. Scalability to the future exascale computers must also be a significant part of WDM research activities.

A WDM should provide the interface between simulations and experiments. The WDM should be connected to an experimental database for two purposes. At one end of the process, the experimental data, once they are calibrated and processed, are an input to simulations; at the other end, they are used for validation of simulation outputs. Data should be prepared using a common format so that interfacing with codes is facilitated, as well as exchange of data among experimentalists and modelers. Along this line, some topical groups of the ITPA (International Tokamak Physics Activity framework for ITER) are considering adopting the ITER Data Structure (IDS) as a standard, as well as the Integrated Modeling and Analysis Suite (IMAS) as a framework to support all of the physics modeling needs. Access to a centralized experimental database would facilitate cross-checks among experiments, as well as submission of runs for Uncertainty Quantification studies. All of this interaction would require storage capabilities, large-scale computing and centralized data management, processing, and visualization.

The output from simulations should be reduced using centralized software capabilities to ensure that the metrics needed for V&V are defined consistently.

Assessing uncertainties in the physics models needs to be incorporated in the WDM codes so that the simulation results include the confidence intervals of the predictions. In predictive simulations, UQ tools can be used to evaluate the probability of events occurring (such as disruptions and edge-localized modes), for the computation of confidence intervals of predicted quantities, and for the optimization of plasma performance. There is very little work being performed in this direction at this moment, and action should be taken to ensure that interfaces for this scientific objective are ready in ten years.

The following have been identified in the white papers as scientific challenges that represent attractive scientific opportunities for the collaborative FES/ASCR community to tackle in the next five or ten years, and whose progress would benefit from the availability of increased computing capabilities.

- Predictive WDM core-edge simulations that are based on the kinetic models: First-principles-based kinetic codes that can treat the core and the edge self consistently are extremely valuable for validation against experiments and for verification of reduced models. The validation activity needs to be increased during this time period. Despite the significant progress in experimental diagnostics in the plasma edge region, the error bars for experimental data in this region remain among the largest in tokamaks.
- Efforts to bridge the gap between short and long timescales: A multiscale time-integration method needs to be developed in order to "telescope" the kinetic simulation to experimental timescale. Plasma turbulence correlation time, which determines the transport coefficients, is milliseconds, whereas the resulting plasma profile evolution can be over seconds in a large size tokamak.

3.1.4.2 Priority Research Directions

The following is a list of physics problems that have been identified in the white papers as high priority integrated simulation problems to be solved over the next five to ten years with the current and pre-exascale computer capabilities:

- *To determine the core plasma boundary conditions from plasma-wall interactions and pedestal structure through the scrape-off-layer and the H-mode pedestal area.* All of the core plasma profiles are strongly influenced by the evolution of the plasma boundary.
- *To predict the core plasma confinement and details of transport in tokamak core discharges.* Currently, a variety of reduced transport models yield different predictions for confinement and fusion power production in burning plasma tokamaks such as ITER. There must be a convergence in the transport predictions based on the high-fidelity turbulence and particle orbit computations. Various physics effects need to be considered including effects associated with the behavior of nonlocal transport. A reliable core-edge coupling is a prerequisite for a reliable prediction of the core plasma performance.
- *To predict the onset, frequency, and consequences of macroscopic instabilities.* Comparisons can be made with experimental data for the frequency of sawtooth oscillations, the effect that a sawtooth crash has on the plasma profiles, the onset of neoclassical tearing modes, and the resulting magnetic island widths. There is also a critical need to predict the onset of edge-localized modes and their frequency and effect on the divertor plates, as well as the onset of disruptive instabilities and their nonlinear evolution.
- *To compute the sources and sinks that drive all of the profiles in plasma discharges.* Sources such as neutral beam injection, fusion reaction products, and radio frequency heating and current drive all involve the computation of fast particle distributions and their interaction with the thermal plasma profiles. Predictions are needed for the effect of fast ions on macroscopic instabilities such as sawtooth oscillations and on microscopic turbulence.

Priority actions toward the achievement of the above physics goals include efforts to:

- *Engage the whole-device tokamak community with the SciDAC groups.* This engagement is a necessary and important step toward the development of verified and validated WDM model hierarchies and toward ensuring that a national WDM initiative includes the essential physics modules, with their synergy and interactions with reduced models.

- *Demonstrate robustness of the WDM as it develops and increases in complexity.* This action means balancing modularity with self-consistent evolution of the magnetic equilibrium as this is modified by plasma profile evolution, wave propagation, and MHD instabilities. Work should be progressed in parallel on the development of a framework, workflow, and inclusion of advanced physics. The framework and workflow should allow tighter coupling of different modules when needed. Loose coupling can be adopted if the coupled physics are scale-separable and are not nonlinearly multiscale coupled. At present time, for example, workflows use loose coupling to manipulate and exchange the RF and MHD data. This approach may be sufficient for some applications, but it does not provide a self-consistent treatment of the evolution of the magnetic equilibrium as a consequence of the interactions of the RF and MHD waves. A scientific challenge identified in the white paper, "Computational Needs: Coupling Extended MHD Simulations and RF Wave Codes," by Jenkins et al. (Appendix C) is the small size of the resonant region in which the RF modifies plasma dynamics, which needs high resolution (sub-millimeter), whereas the tearing mode size is of the order of the device. Substantial computing effort is needed in order to resolve this issue (hundreds of runs with at least ~10K cores each). Modern WDM workflows (such as FACETS, EFFIS, TGYRO, and TRINITY) that utilize novel computational approaches, rigorous regression tests, and advanced solvers still do not include many important physics components, synthetic diagnostics, and interfaces to various experimental data. Older codes, such as TRANSP, typically include sophisticated physics components but are outdated with respect to computational aspects including code portability requirements, regression analysis, parallel code execution, and parallel load balancing.

Engagement of the computer scientists and the applied mathematicians is needed to develop, improve, and maintain predictive integrated codes for carrying out whole device modeling of tokamak plasmas. This team should bring together into one framework the essential component codes and models that presently constitute separate disciplines within plasma science. Furthermore, as computing architectures evolve toward the exascale, it will be necessary to work with the applied mathematics and computer scientists in order to adapt existing physics components and the integrated WDM codes to these new architectures.

3.1.4.3 Cross-Cutting Research Directions

One of the main objectives of pursuing the integrated WDM is capturing the complex physics interactions over a wide range of space and time scales. This complexity includes (1) particle, heat, and momentum fluxes between the plasma core and the wall; (2) stiff plasma profile interaction between the core and edge plasma; (3) interaction of edge plasma with the material wall; (4) propagation of RF waves from the antenna to the core plasma; (5) interactions between RF waves and energetic particles; (6) interactions between RF waves and MHD instabilities; and (7) interactions of RF waves with the plasma-material interface.

Another cross-cutting effort involves engaging the integrated tokamak community with the SciDAC groups. This effort is a necessary and important step toward realizing improvement of reduced models and toward ensuring that a national WDM initiative includes essential physics components (as embodied in Figure 3-9). Note that in carrying out the "probabilistic WDM" studies, the validated reduced models that have been improved by the high-fidelity WDM will be useful. High-fidelity understanding and predictive modeling from the SciDAC groups can be used in the reduced models in the following synergistic ways:

- Development of reduced models for sawtooth instabilities and neoclassical tearing mode instabilities, their onset and interaction, and for how these instabilities are modified by interaction with RF waves.
- Inclusion of a reduced model for RF propagation that includes the damping of waves in the SOL.

- Inclusion of a reduced model to describe the interaction between RF waves and fast ions.
- Representation of flows between the core and the plasma wall.

3.1.4.4 Computing Needs and Requirements

Following are needs and requirements identified with respect to pursuit of WDMs:

Computing time required at scale. For the kinetic first-principles WDM, the whole exascale HPC is expected to be utilized at a time. One simulation of an ITER discharge may take one day of wall-clock time on an exascale computer at 100% capability. Twenty (20) predictive simulations of ITER could require 20 CPU days of an entire exascale HPC. The advanced reduced model WDM will require 10% of an exascale HPC at a time. Even in the reduced model WDM, most of the parallel computing will be performed by advanced (kinetic) modules. Thus, 200 simulations of advanced reduced WDM can consume 20 CPU days of an entire exascale HPC.

Solvers and algorithms. Development of more efficient parallel solvers is important for the WDM codes. These solvers can improve the utilization of available first-principle models and advanced reduced models at large scale. Improvements to dynamic load balancing are important because they will help to utilize the computational resources more efficiently between different physics modules of the WDM codes. The algorithms, visualization, and analysis systems need to scale. Verification of the current numerical algorithms involves comparing computational solutions with benchmark solutions, analytical solutions, manufactured solutions, and with heroically resolved numerical solutions.

Capacity computing for reduction of models from extreme scale simulations. Advanced reduced models are typically derived from an ensemble of extreme-scale calculations over a selected range of plasma parameters and therefore might be applicable to only a limited operational space. Capacity computing would allow running a few tens to hundreds of reduced-model simulations simultaneously in order to optimize the large-number parameter fit to the ensemble of extreme-scale results.

Storage capabilities. A WDM should be connected with a large-scale experimental and simulation database. The reasons for this are twofold. On the one hand, the experimental data, once they are calibrated and processed, are an input to simulations. On the other hand, they should be used for analysis and validation of the simulations themselves. Data from the extreme-scale, high-fidelity simulations can be large. Experimental and simulation data with their measurement uncertainties are used for UQ purposes. Having connection to many experimental and simulation databases implies (1) large-scale storage capability needs, and (2) short latency and large bandwidth.

These data have to be prepared using a common format, or through a universal interpreter, so that interfacing with codes is facilitated. One such data structure employed, for example, could be the IDS adopted by the ITER Organization. Using a common data structure would facilitate exchange of data among experimentalists and modelers. Some of the ITPA Topical Groups are considering adopting the IDS for exchange of data that modelers can use. A common experimental database of reduced data might be the input for millions of runs that can be submitted concurrently for UQ on several tokamaks at the same time.

Ideally, the V&V process connected to a WDM should be centralized. This approach means having a common interface that reduces both experimental data and simulation outputs to generate metrics for V&V purposes. Synthetic diagnostics are also part of the V&V process.

Data Management and Visualization. Data from the extreme-scale, high-fidelity codes are expected to be large. Because the I/O bandwidth and the storage size are not expected to increase as fast as the computer size, an on-memory, *in-situ* data analysis, visualization, and compression component will be a necessary part of high-fidelity WDM development. Utilization of the hierarchical memory structure and the heterogeneous computing, with multilevel programing models, will need to be optimized. Restart check-pointing and fault tolerance may need to utilize the node-level NVRAM (nonvolatile random access memory) memory.

Common interface to reduce output from WDM for more advanced codes. Self-consistent, time-dependent simulations are routinely used to prepare plasma profiles and equilibria for selected time-slices that are then used by extreme-scale codes, for example, gyrokinetic simulations.

Separation between physics development and code refactoring. Of high concern is the separation between physics advancement and code refactoring. Dedicated collaborations between ASCR and FES for code refactoring issues raised by new computing ecosystems would minimize losses in scientific productivity.

Common interface for V&V. Simulation output must typically be reduced in order to analyze it for the purposes of verification and validation. A common interface ensures that the metrics used are consistent among codes and experiments.

Application codes. Modern fusion WDM workflows that utilize novel computational approaches, rigorous regression tests, and advanced solvers still do not include many important physics components, synthetic diagnostics, and interfaces to various experimental data. Older codes, such as TRANSP, typically include sophisticated physics components but are outdated with respect to computational aspects including code portability requirements, regression analysis, parallel component execution, and parallel load balancing.

Community service software. Unlike gyrokinetic turbulence codes, reduced whole device modeling codes such as TRANSP require only a moderate number of processors (2000 to 5000). Timely progress of research from these codes requires that this capability be provided with fast turnaround time.

Workforce. A committed team of computer scientists, applied mathematicians, and plasma physicists is needed to develop, improve, and maintain predictive integrated codes for carrying out whole device modeling of tokamak plasmas. This team should bring together into one framework the essential codes and models that presently constitute separate disciplines within plasma science.

3.2 Plasma Surface Interactions and Structural Materials

3.2.1.1 Scientific Challenges and Opportunities

The realization of fusion as a practical, 21st-century energy source requires improved knowledge of plasma material interactions and the materials engineering design of component systems to survive the incredibly extreme heat and particle flux exposure conditions of a fusion power plant. In considering plasma-material interactions (PMI), it is evident that three coupled spatial regions influence PFC materials evolution and performance, as indicated in Figure 3-10. These regions consist of (1) the edge and scrape-off layer region of the plasma; (2) the near-surface material response to extreme thermal and particle fluxes under the influence of, and feedback to, the plasma sheath; and (3) the structural materials response to an intense, 14 MeV peaked neutron spectrum, which produces very high concentrations of transmuted elements through (n,p) and (n,α) reactions and structural material property degradation. The coupled nature of these spatial domains necessitates creating the interfacing between modeling approaches for each, in order to better evaluate the feedback between each region on the performance of the other. For example, the interface of the surface to the plasma edge/scrape-off layer is necessary to define the incident particle and thermal fluxes that are the driving force for PMI, as well as to account appropriately for the processes of excitation, ionization, and charge-exchange that can result in species re-deposition. Likewise, the interface between the surface and the bulk, where defect creation is no longer influenced by the presence of a free surface, is critical in determining the extent to which defect creation by high-energy neutrons impact retention and permeation of hydrogen isotopes, with a significant unknown regarding the tritium permeation behavior in metallic PFC at elevated temperatures.

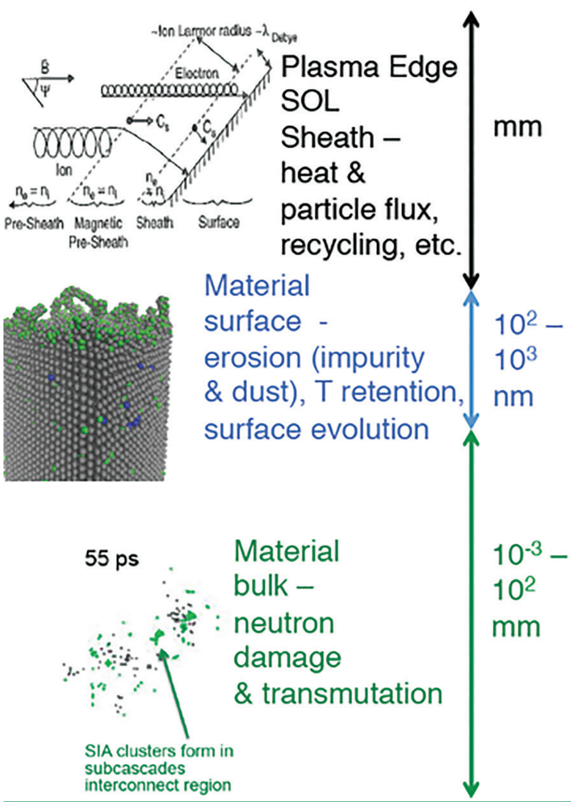


Figure 3-10. Important spatial domains of the plasma material interface, which control plasma performance through erosion, recycling, and retention and impurity generation.

The control of coupling between the plasma edge and the wall surface has been inhibited by a lack of fundamental understanding of their interface. PMI processes mix the materials of these two worlds, creating in between a new entity — a dynamical surface that communicates between the two — creating one of the most challenging areas of multidisciplinary science, which has many fundamental processes and synergies. We know that the edge plasma governs particle and energy exhaust, and impurities eroded from the surfaces may reduce the fusion gain if they are transported back to the confined plasma. The other critical effects of these interactions are (1) reduced lifetime of plasma-facing surfaces owing to erosion by transients, and (2) restrictions on duty cycle because of retention of tritium in re-deposited material and in dust created by plasma surface interactions. Furthermore, all choices for plasma-facing materials (PFMs) in a fusion reactor have known issues. There are good arguments for both low-Z vs. high-Z PFMs in fusion devices, which have been discussed many times in the fusion community.

The traditional trial-and-error approach to developing first-wall material and component solutions for future fusion devices (ITER, DEMO) is becoming prohibitively costly because of the increasing device size, curved toroidal geometry, access restrictions, and complex programmatic priorities. The complexity of this environment requires a change from an engineering emphasis toward a more fundamental approach, grounded in a multiscale modeling methodology capable of attacking the plasma-material interface problems simultaneously from both a bottom-up and a top-down approach. The experimentally validated atomistic theory/computation for studying the dynamics of the creation and evolution of the PMI under irradiation by heavy particles (atoms, molecules) at carbon, lithiated and boronised carbon (National Spherical Torus Experiment [NSTX]) and tungsten (ITER), as well as the emerging elastic and inelastic processes, in particular retention and sputtering chemistry, are some of the burning scientific challenges with which we are dealing. Quality validation of the simulations is the key for success, which requires well-designed modeling and simulation to enable modeling the available experimental conditions precisely. Another critical aspect of the modeling to be performed moving forward is a rigorous implementation of uncertainty quantification, which will further require a combination of capacity and leadership-scale computing driving toward the exascale.

Turning toward the plasma side across the PFCs of a fusion device, the electrostatic sheath and the collisional and magnetic pre-sheath act as an interface layer between the Pedestal/Scrape-Off-Layer plasma and the material surface. Such interfaces include a multitude of processes, highly kinetic in nature, involving multiple plasma species (electrons, ions, neutrals, material impurities) in a dynamically evolving environment tightly coupled to the surface. At the nominal conditions anticipated for a reactor, the majority of sputtered material ionizes close to the surface and is redeposited nearby. The redeposition process forms a new “reconstituted” surface layer with different and unknown thermo-mechanical properties that differ from the original ordered lattice. This continuously eroded and re-deposited surface can significantly alter the PFC lifetime, affect the retention of hydrogenic species (deuterium, tritium), and affect the mechanisms associated with microscopic erosion of the surface (both net and gross erosion). Under continuous plasma exposure, the near-wall plasma and the surface form a system, far from equilibrium, in which the wall is continuously eroded, redeposited, and reconstituted. The problem is intrinsically multiscale, both in space (from nanometers to centimeters) and time (from fractions of a picosecond to minutes and hours) and multiphysics. The dynamic modeling of the kinetic processes occurring at the near-wall layer requires the coupling of different physical models and codes together, namely:

1. A multi-species kinetic model of the plasma sheath/presheath region, handling the evolution of the distribution function of electrons, ions, neutrals, and material impurities from the quasi-neutral region to the first surface layer; the target equations are the Boltzmann-Poisson and the Boltzmann-Maxwell.
2. A kinetic model of the material wall, handling ion-matter interaction, and including relevant phenomena such as sputtering, backscattering, and implantation on a material surface having dynamic composition and evolving morphology; the target equation is the classical multi-body problem for given (known) interaction potential.
3. A proper collision operator accounting for the interaction among species, handling the relevant atomic physics such as ionization, charge exchange, ion and impurity recycling, and more. The target equations are the Fokker-Planck and nonlinear collision operator.

Turning toward the bulk region below the PFC and first wall surfaces, it is clear that the development of successful structural materials will ultimately require a fusion-relevant neutron source in order to code-qualify these materials for nuclear service; however, a purely experimental approach to understanding and mitigating radiation-induced degradation is not practical because the cost to design, perform, and examine materials from irradiation experiments is high and the available irradiation volumes are low. The lack of a fusion-relevant neutron source in which to conduct prototypical experiments reinforces the need for a robust theory and simulation program in order to understand experiments carried out in surrogate irradiation facilities. Furthermore, there is a combinatorial problem in that the broad range of materials, phenomena, and irradiation variables — and variable combinations — makes a purely experimental approach intractable. Physically based computational models of microstructure and property evolution are indispensable tools because they provide a means to re-evaluate existing data, optimize the design and execution of new experiments, and interpret the results from those experiments. Although multiscale models describing radiation and mechanical damage processes are under intense development, numerous details remain to be resolved before they can accurately predict material performance, because those models must simultaneously span length and time scales ranging from atomistic to the continuum and from sub-picoseconds to years, respectively. Herein we discuss the development status and computational needs of one class of simulation methods known as kinetic Monte Carlo (KMC), which is a mesoscale technique that bridges the gap between atomistic level tools, such as density functional theory and molecular dynamics (MD) simulations, to more continuum-level approaches, such as cluster dynamics and rate theory.

Three major challenges of using the KMC method to simulate radiation damage are (1) reaching experimentally relevant timescales for physically meaningful system sizes, (2) simulating experimentally relevant length scales, and (3) carrying out high-fidelity or realistic simulations incorporating all needed defect interaction physics. For radiation damage simulations, object KMC (OKMC) is the method of choice in which the objects of interest are defects and their reaction/diffusion mechanisms; therefore, all of the challenges mentioned above pertain to OKMC simulations. Computational time per KMC step increases slightly while the advance of simulation time per KMC step decreases significantly with an increasing number of mobile objects. The latter has the largest effect on the achievable simulation time. Therefore, the maximum simulation time achievable in a reasonable (or affordable) amount of real time depends on the number of mobile objects in the simulation. Due to both the inherently serial nature of the KMC algorithm and minimal computational cost per KMC step, exascale systems will not be helpful for extending the simulation time (timescale) on system sizes that a serial KMC code can handle (strong scaling). This issue is and will always remain a challenge. However, the use of exascale systems would be most beneficial in extending the length scale of KMC simulations even with existing parallel KMC algorithms, which will also indirectly extend the timescale of OKMC simulations (weak scaling). The fidelity of radiation damage evolution predictions obtained from OKMC simulations depend on the degree to which all possible/relevant reaction/diffusion mechanisms are included and on how rigorously they are treated in the simulations. Making OKMC simulations more realistic than they are presently rendered would increase the computational cost per KMC step, making them computationally more expensive. Exascale systems would be very beneficial in extending the length and timescales as well as the fidelity of OKMC simulations.

3.2.1.2 Priority Research Directions

The materials group identified two specific PRDs, which we feel can be accomplished in the 10-year time horizon and which would enable the scientific grand challenge associated with a coupled plasma–material model for magnetic fusion energy. The first of these is to build an integrated and first-principles-based suite of advanced codes to predictively model the boundary plasma and material surface. It is imperative that such codes incorporate rigorous treatment of the turbulent transport, along with kinetic and sheath effects in the plasma, and that they be efficiently coupled to a multiscale materials modeling framework to enable the prediction of evolving PFC performance in terms of erosion, PFC lifetime, and tritium inventory, such that the plasma boundary models can provide feedback to the codes modeling the plasma pedestal and the burning plasma core performance. The second PRD is focused on extending the materials modeling framework to accurately predict with high fidelity the properties and performance of W-alloy structural components. An important aspect of this second PRD is the continued rapid expansion of computational materials capabilities to accurately model the electronic structure interactions of numerous impurity and transmuted elements in tungsten and other possible PFC and structural materials.

PRD 1: Integrated, first-principles-based models of boundary plasma and materials surfaces in a burning plasma environment.

As mentioned earlier, the problem is inherently multiscale and multiphysics, with timescales spanning from sub-picosecond to minutes and hours; spatial scales that span the atomic spacing of solid surfaces to the millimeter scale of gyro-orbits in the plasma; and the extent of the electric and magnetic sheaths. The dynamic modeling of the kinetic processes occurring at the near-wall layer requires the coupling of different physical models and codes together, namely:

1. A plasma model capable of rigorously treating turbulent transport of both heat and particles through the pedestal and into the scrape-off layer and the plasma boundary.
2. A multi-species kinetic model of the plasma sheath/presheath region handling the evolution of the distribution function of electrons, ions, neutrals, and material impurities from the quasi-neutral region to the first surface layer.
3. A kinetic model predicting the evolution and transport of sputtered impurities within the plasma boundary, capable of detailed calculation of the full gyro-orbits over spatial scales relevant to tokamak plasma.
4. A kinetic model of the material wall, handling ion-solid interaction and including relevant phenomena such as sputtering, back-scattering, and implantation, as well as modeling the transport and fate of implanted gas atoms to assess the fuel recycling, permeation, and retention, along with the ability to predict the chemical and structural morphology evolution of the surface.
5. A proper collision operator accounting for the interaction and coupling among species and spatial scales, handling the relevant atomic physics such as ionization, charge exchange, ion and impurity recycling, and more.

PRD 2: Extend structural materials modeling to accurately predict the chemical and microstructural evolution of structural components under neutron irradiation beyond 10s of displacements per atom (dpa).

This problem is also inherently multiscale and multi-physics, especially when considering the transmutation of solid elements by a 14-MeV peaked neutron source. Being inherently multiscale in both time and space presents a crucial difficulty: timescales range from femtoseconds (to describe dynamics in the immediate aftermath of collision cascades) to hours and days (to capture slow microstructural changes) while length scales range from nm to describe individual defects to cm to fully describe microstructures. Adding to the difficulty is the fact that the relevant physics operating at each of these different scales and the synergetic effects of the couplings between scales are still known only in part. Being able to understand, predict, and ultimately design materials that can reliably operate in this kind of extreme environment is a grand challenge that has to be tackled by the community in order to make fusion energy a reality. Although multiscale models describing radiation and mechanical damage processes are under intense development, numerous details remain to be resolved before they can accurately predict material performance, because those models must simultaneously span length and time scales ranging from the atomistic to the continuum and from sub-picoseconds to years, respectively. Of particular interest are the development of KMC algorithms to more effectively utilize parallel computing when tracking time-dependent evolution and the ability of continuum-based cluster dynamics models to more efficiently and effectively model multi-component systems with evolving chemistry.

Note that, even though several algorithms exist, KMC simulations in almost all cases employ serial architecture. To the best of our knowledge, KMC codes have utilized neither terascale nor petascale systems to carry out parallel KMC simulations. This absence of utilization is mainly because parallelization of the KMC method is nontrivial and parallel efficiency is highly dependent upon the problem being studied as well as the simulation parameters used. Moreover, parallel efficiency of a KMC simulation varies over the course of the simulation. Therefore, a priori prediction of parallel efficiency on any system, let alone the extant computing environment, is very difficult. Considering the fact that the KMC method has never (or very rarely) been used on previous as well as existing HPC systems, it is hard to comment on what can or cannot be solved using parallel KMC simulations on exascale systems. Nevertheless, there is a lot of scope for improvement and great benefit in initiating the extensive use of HPC systems to carry out parallel KMC simulations. Because GPUs have excellent computing power (FLOPS) per dissipated watt, it is highly likely that exascale systems will be a heterogeneous computing environment with CPU-GPU systems. Accordingly, testing and implementation of parallel KMC algorithms on a GPU(s) would be the first logical step in porting KMC codes to exascale systems.

Again, a critical aspect of the multiscale materials modeling problem of spanning such disparate scales is the primary outstanding issue in the field. Even without major methodological advances, the range of size and length scales that can be directly simulated with high accuracy is expected to increase significantly. While we now consider growth of nanoscale He bubbles over timescales of tens of microseconds, the advent of exascale systems will enable millisecond simulations and a possible increase in the system sizes (and hence defect sizes) that can be handled. This capability will certainly improve the quality of the microscopy parameters than can be fed to higher-scale models that aim at modeling evolution at the microstructural (or even reactor) level. The robustness and predictive nature of these models will therefore significantly improve. Similar improvements can be expected in the understanding of other nanoscale defects.

3.2.1.3 Cross-Cutting Research Directions

A number of cross-cutting research directions were identified by FES scientists during this review, including some with obvious interest to researchers funded by ASCR and the Office of Basic Energy Sciences. In particular, it is widely recognized that improvements in computing capacity as we move toward exascale will continue to improve our ability to rigorously model the chemistry of multiple plasma/material impurities in PFCs, and multi-component solid materials in general, which are necessary to increase the fidelity of atomistic modeling for discovering controlling mechanisms for meso- to continuum-scale materials modeling, and will directly lead to improved interatomic potentials for atomistic modeling. Likewise, there is a clear need for algorithmic development to extend the timescale of atomistic modeling techniques (i.e., MD, accelerated MD, and kinetic Monte Carlo).

In addition, a number of cross-cutting opportunities exist with ASCR and FES-funded researchers associated with developing computationally efficient, 3D models of plasma fluid transport, including turbulent and kinetic effects, and treating impurities. This effort will most likely require the development of new computer codes, which are portable and easily able to adapt to exascale computing platforms under development, as well as improving the workflow techniques used to couple codes across pertinent spatial domains and interfaces. Another clear cross-cutting opportunity is related to radio-frequency and ICRF launchers, which feature similar severe challenges associated with plasma materials interactions.

The domain of low-energy plasma materials processing offers numerous synergies to capture the best practices of modeling surface evolution, and the approach to hierarchically integrating numerous timescale issues.

Finally, it is critically important to address two very important issues as the scale of scientific computing of plasma material interactions and structural materials performance in the fusion environment continues to increase. Namely, we recognize the need to address data management issues as the models grow larger in scope. We further recognize that we are beginning to reach a point of fidelity in the modeling such that it is now time to expand the focus on verification and validation, along with uncertainty quantification evaluations, ensuring the separate validation of plasma and materials models before attacking the coupled codes. Additional emphasis needs to be placed on impact UQ to gain confidence in the robustness of models.

3.2.1.4 Computing Needs and Requirements

Table 3-1 highlights factors that are accelerating or impeding progress toward exascale-level computing.

Table 3-1. Factors accelerating and impeding progress on exascale.

Accelerate	Why?
1. Hardware resources (at all scales)	Speed increases will offer the opportunity to simulate a larger volume of material and the possibility for higher-fidelity KMC algorithms.
2. Models and algorithms	Improvements here will allow greater access to the extant ecosystem.
3. Application codes	Improved codes will allow greater and more efficient usage of the ecosystem.

Impede	Why?
1. Visualization and analysis resources	Rendering enormous amounts of exascale data into images on screen may require a prohibitively long time.
2. Data workflow	Challenges handling exascale data files may affect simulation and analysis software stability.
3. Libraries/frameworks	Math libraries that are not optimized for exascale systems may impede KMC code optimization.
4. Programming models	Requirements that researchers possess extensive knowledge of the underlying hardware and data movement architectures will make programming, maintaining, and extracting code performance information quite tedious. This requirement will also affect the portability of the code from one HPC system to another.

3.3 Discovery Plasma Science

3.3.1 General Plasma Science

3.3.1.1 Magnetic Reconnection

3.3.1.1.1 Scientific Challenges and Opportunities

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, sawtooth 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 toward relatively small two-dimensional (2D) systems using both fluid and kinetic descriptions. Presently, it remains unclear how several of these idealized results will extend to large-scale 3D systems. Even with exascale computing, a first-principles 3D 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 it within reduced descriptions and to infer reliable scalings.

3.3.1.1.2 Priority Research Directions

Looking toward the future, we identify the following four PRDs.

In order to model reconnection for most applications, the first PRD is the influence of the electron and ion kinetic scales on the large-scale evolution. At present, 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 capture the large-scale evolution accurately. First-principles kinetic simulations including Coulomb collisions can provide guideposts for developing reduced fluid descriptions that better capture the structure and dynamics. Other approaches may include using reduced kinetic descriptions, such as gyrokinetic, or embedding a kinetic description within larger fluid simulations.

The second PRD is reconnection and magnetic island dynamics in 3D geometries. There is already evidence that a single reconnection layer may break up into multiple interacting reconnection sites due to the formation of secondary magnetic islands or other secondary instabilities. For many applications, this complex evolution is expected to depend on the global geometry and boundary conditions. Addressing these issues will require highly scalable fluid and kinetic algorithms, along with realistic treatment of boundary conditions.

The third PRD is the energy partition and particle acceleration that results from reconnection. The thermal energy gained by ions and electrons, as well as the formation of nonthermal tails, is of significant theoretical and observational interest. For the highly energetic tails, it appears very 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.

The fourth PRD is reconnection in relativistic plasmas. In many astrophysical applications (pulsars, active galactic nuclei [AGN], 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, which are now feasible in 3D at the petascale for electron-positron plasmas.

These advancements in reconnection physics have the potential to affect fusion energy science through better modeling of tearing modes and sawtooth oscillations in tokamaks and of relativistic electrons for fast ignition and increased understanding of magnetic relaxation in reversed field pinches and field-reversed configurations.

3.3.1.1.3 System Requirements

Extreme-scale computing will provide an invaluable tool for addressing the high-Lundquist-number (S) regime. However, the computational cost to resolve the resistive layers and follow the macroscopic evolution on the global Alfvén time increases as $S^{5/2}$ for 3D explicit simulations. For S~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 of S~10^{12}, which are 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 3-11, this structure imposes a daunting level of scale separation. The kinetic timescales are separated by the ion-to-electron mass ratio, (m_i /m_e) whereas 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 timescale of $\tau_A = L/V_A$ to understand reconnection dynamics.

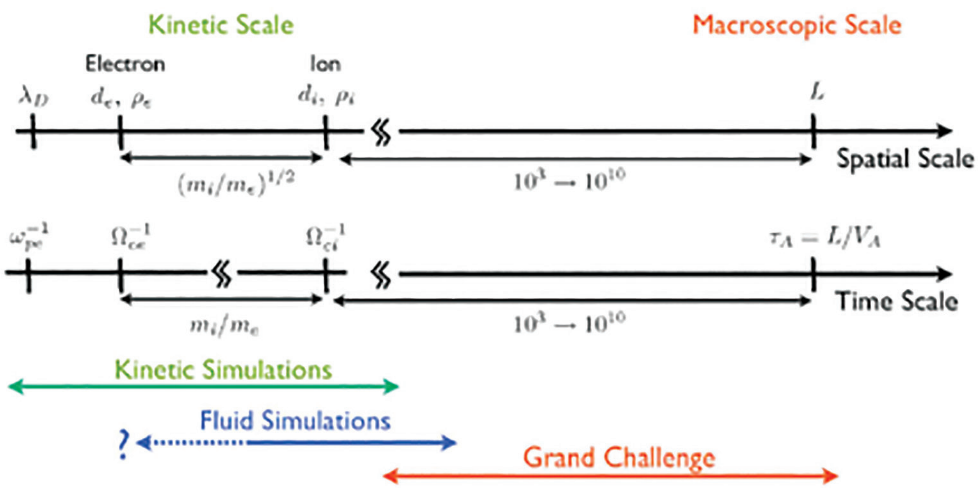


Figure 3-11. Overview of spatial and temporal scales in collisionless reconnection. The fastest timescale is associated with the plasma frequency ω_{pe} followed the electron cyclotron frequency Ω_{ce}. Electron spatial scales include the Debye length λ_D, the electron inertial length $d_e = c/\omega_{pe}$, and the electron gyroradius ρ_e. The same notation, with *e* replaced by *i*, is used for ion kinetic scales.

The computational cost of explicitly resolving these kinetic scales three-dimensionally increases in the same steep manner ~$(m_i/m_e)^{5/2}(L/d_i)^4$ 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 (~25–400). At the petascale, 3D kinetic simulations for hydrogen plasmas with $m_i/m_e =$ 1836 will be possible for systems on the order of ~$10d_i$. A factor of a 10^3 increase in computing power will only increase the feasible 3D 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 scaling by a combination of numerical computation and analytical theory.

During the past decade, laboratory experiments (such as the magnetic reconnection experiment [MRX] at PPPL; see Figure 3-12) and space observations have played an important role in validating several important aspects of these simulation results (from codes such as PSC at PPPL/University of New Hampshire and VPIC at Los Alamos National Laboratory). Renewed interest in the consequences of extreme weather events have spurred development of global multifluid models in which kinetic effects are parameterized through closure relations. The parameter range of relevance to reconnection in weakly collisional plasma has been augmented significantly by HEDLPs in facilities such as Omega and Omega EP at the Laboratory for Laser Energetics at the University of Rochester. Some key discoveries in fast reconnection 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 two-fluid models of tokamak plasmas. In the era of extreme-scale computing, such studies are likely to continue playing an important role in the description of nonlinear reconnection dynamics in ITER-grade plasmas where the issues of scale separation are as important as they are in space and astrophysical plasmas.

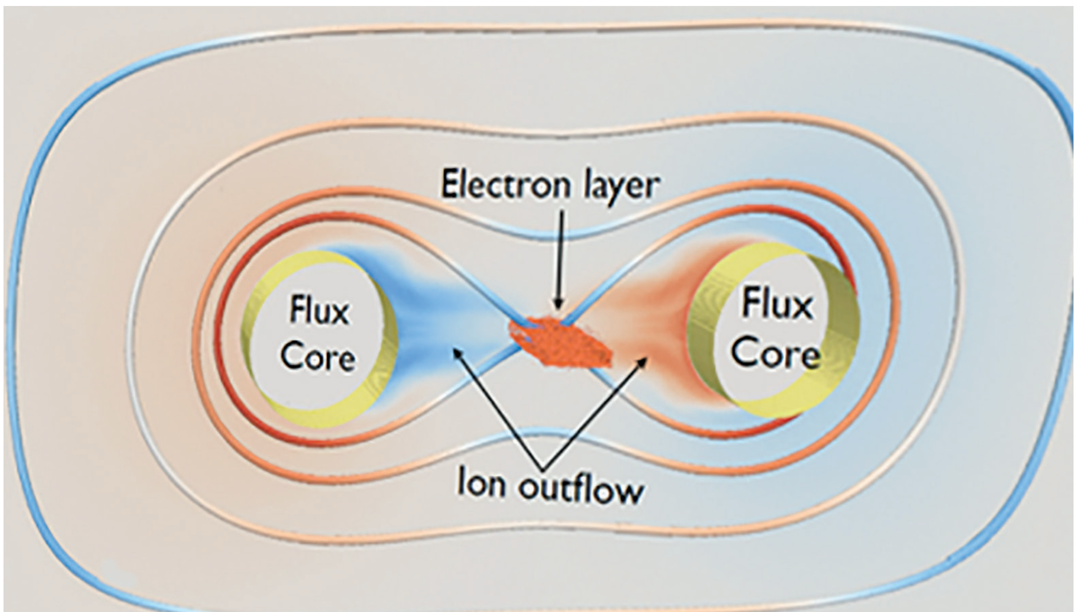


Figure 3-12. Fully kinetic modeling of laboratory reconnection experiments such as MRX.

3.3.1.2 Turbulence

Turbulence plays a central role in governing the flow of energy in the heliosphere from the sun, through interplanetary space, to the magnetospheres of the Earth and other planets, and to the outer boundary of the heliosphere. Yet our understanding of how turbulence governs energy transport and plasma heating remains incomplete, constituting a grand challenge problem in heliophysics. Beyond the heliosphere, in the most extreme astrophysical environments — such as the neighborhood of the supermassive black hole at the center of the Milky Way Galaxy or the powerful shocks emerging from supernova explosions — turbulence mediates the conversion of large-scale motions to plasma heat, or some other energization of particles.

Because turbulence mediates the transport of energy, momentum, and particles through motions spanning many orders of magnitude in scale, the modeling of plasma turbulence is an inherently multiscale problem, formally beyond the reach of even today's most advanced computers and sophisticated algorithms; thus, exascale computing promises the ability to make transformative progress in the field. In addition, the problem of space and astrophysical plasma turbulence is made yet more complex by the fact that, at the typically low densities and high temperatures of these plasmas, the turbulence dynamics is often weakly collisional, requiring the application of kinetic plasma theory to follow the evolution and dissipation of the turbulence. Kinetic plasma theory describes the evolution of six-dimensional (6-D) particle velocity distribution functions, so that in addition to tackling a large spatial dynamic range, astrophysical plasma turbulence problems demand efficient algorithms to handle nonlinear, high-dimensional simulations and also require novel approaches to the visualization and analysis of this high-dimensional data.

3.3.1.2.1 Priority Research Directions

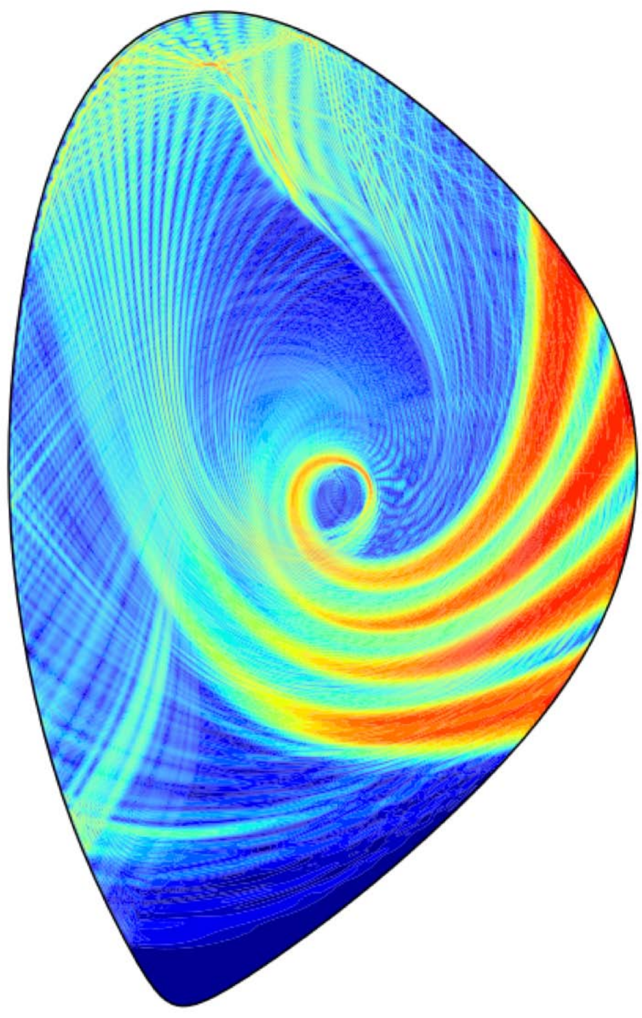
Fundamentally, the key question to answer is, *How does turbulence in a kinetic plasma mediate the conversion of the energy of plasma flows and magnetic fields at large scales to plasma heat, or some other form of particle energization*? The scientific community aims to understand in detail the nonlinear interactions that serve to transfer turbulent energy from large to small scales and to identify the physical mechanisms by which the turbulent fluctuations at these small scales are damped and their energy inevitably converted to plasma heat. Ultimately, this foundation of knowledge will facilitate the development of a predictive capability, enabling the determination of partitioning of energy deposited among protons, electrons, and minor ions as a function of the turbulence and plasma parameters. Determining this plasma heating from first principles is essential to explaining the macroscopic evolution of many important systems: it is the key to understanding how the solar corona is heated to a temperature nearly one thousand times hotter than the solar surface, or to explaining the unexpectedly low emission of radiation from the supermassive black hole at the center of our galaxy.

Over the next decade, through a coordinated program of spacecraft measurements, theoretical calculations, and nonlinear kinetic numerical simulations, the scientific community is poised to make transformative progress on the problem of how turbulent energy is dissipated and converted to plasma heat in a weakly collisional heliospheric plasma. Two National Aeronautics and Space Administration missions that will provide invaluable *in-situ* measurements of plasma turbulence at the physical scales of turbulent dissipation are the $850M *Magnetospheric Multiscale* (MMS) mission, just launched in 2015, and the $1.5B *Solar Probe Plus* mission, which is due to launch in 2018. In addition, novel theoretical work on high-dimensional analysis is driving the effort to identify correlations arising through collisionless interactions between the measured electromagnetic fields and the fluctuations in the particle velocity distributions, and to devise strategies to exploit those correlations as a probe of the transfer of energy from the turbulent electromagnetic fields to the particles. Exascale computing will play an essential role in this synergistic research effort on the dissipation of turbulence, enabling direct numerical simulations of the high-dimensional, nonlinear turbulent dynamics. Such exascale-based kinetic simulations are indispensable to test these new analysis methods and their theoretically predicted results under controlled conditions. Furthermore, the simulations are crucial to interpret the high-dimensional spacecraft measurements of the 3D velocity distribution and electromagnetic field fluctuations. Numerical challenges include the implementation of new algorithms to compute the evolution of the 5-D gyrokinetic or 6-D fully kinetic plasma efficiently over a large spatial dynamic range and the visualization and analysis of high-dimensional data sets generated by these exascale computations.

3.3.1.2.2 Cross-Cutting Research Directions

The visualization and analysis of high-dimensional data sets (5D gyrokinetic or 6D fully kinetic) is an upcoming computational science challenge that spans plasma science. For example, because the physics of the acceleration and propagation of runaway electrons in fusion devices is inherently kinetic, developing a predictive capability to complement upcoming fusion experiments also requires sophisticated hardware infrastructure and software tools to realize the full potential of the high-dimensional kinetic code data sets. Current and upcoming space missions, such as MMS and *Solar Probe Plus* generate high-dimensional data sets by full 3D measurements of particle velocity distribution functions. The development of the tools and knowledge to maximize the scientific return from these missions will strongly utilize the next generation of kinetic plasma simulation codes to develop the foundation of knowledge and experience needed to interpret the spacecraft measurements.

A VIEW OF THE SCIENCE



Electromagnetic field simulation using the TorLH solver showing lower hybrid waves coupled from four waveguide launchers (right) in the Alcator C-Mod tokamak. Reproduced from Wright, J.C., P.T. Bonoli, A.E. Schmidt, et al., *Physics of Plasmas* 16**, 072502 (2009).**

3.3.2 High-Energy-Density Laboratory Plasmas

3.3.2.1 Scientific Challenges and Opportunities

Intense lasers are the primary drivers for HEDLP experiments, which play a central role in the core missions of FES (Plasma 2010 Committee 2010; DOE-SC and NNSA 2009). In the last 15 years, laser-driven HEDLP experiments have brought many new discoveries in inertial confinement fusion (ICF); created new extreme states of matter that are otherwise not attainable on Earth; and started to enable exploration in the laboratory of fundamental properties of astrophysical plasmas, leading to the emerging field of laboratory astrophysics. HEDLP experiments have revealed equations of the state, radiative, and transport properties of plasmas, as well as collective phenomena like shocks, instabilities, and plasma photonics. The physics of laser-plasma interactions is multiscale, highly nonlinear, and often kinetic, that is, it cannot be captured by fluid models. Furthermore, its physical and spatial scales range from the experimental scale (i.e., in nanoseconds and millimeters) down to the laser- and plasma scales of sub-femtosecond and sub-micrometers. This range of scales means that computer modeling of HEDLP experiments often requires extreme HPC resources.

The common numerical challenge for many HEDLP problems is the solution of the Vlasov equation, often including collisions. Solving this equation can, in principle, be accomplished directly through a fluid-like approach in n+m dimensional phase space, where *n* is the dimensionality of the simulated configuration space, and *m* the dimensionality of momentum/velocity space. However, due to the complex structure of a particle's phase space, this approach is currently deemed to be too expensive for the full six-dimensional (6D)-phase space. Many HEDLP researchers today make use of a particle-in-cell approach that combines the explicit solution of Maxwell's equations on a grid with a particle-based electric current, and boundary conditions (antennas) to include a laser (Birdsall and Langdon 1991). Collisions can be included, for example, through a direct binary-collision operator based on the Landau equation (Takizuka and Abe 1977). The PIC method is conceptually simple, because its locality and the finite speed of light allow problems to be parallelized in a domain-based decomposition with relatively little communication overhead. However, PIC simulations are still inherently expensive because (1) most compute resources are consumed by depositing the charge-current deposition of particles: their number N_{part} determines noise fluctuations via $\sim 1/N_{part}^{1/2}$ so that a large number of particles is required for noise suppression; and (2) the grid resolution dx is ideally determined by the plasma's Debye length, which is one-tenth of a micron at critical density for a laser wavelength of 1 micrometer and a temperature of 1 keV, and much smaller for cold solid-density plasma. The time step, on the other hand, is linked to the spatial resolution through the so-called Courant-Friedrichs-Lewy condition (Birdsall and Langdon 1991), so that the cost of a PIC simulation scales with resolution as dx^4.

Opportunities from faster and larger computer systems arise from four potential avenues:

1. **Increased problem size.** On a larger computer system, one can simply increase the simulation volume and/or time (e.g., to model the interaction between several laser speckles), instead of a single or pair of speckles. The transition from 2D to 3D simulations at constant resolution typically involves a cost increase by a factor of a few thousand, depending on the original problem size. Doing so will allow for one-to-one modeling of experiments and a direct comparison to experimental diagnostics.
2. **Increased grid resolution** at constant problem size to include spatial scales/modes that had to be ignored in past work, improving credibility while reducing noise for the modes that are now better resolved.
3. **Ensembles/families of runs** for purposes of error sensitivity (UQ), or for providing trends for the formulation/verification of models.

4. **Enhanced inter-activity with experimental campaigns.** At constant problem size and resolution, faster computers deliver an enhanced inter-activity with experimental campaigns. While current runs take days, accelerated simulations could inform experimental campaigns in real time (i.e., minutes to hours), thus allowing for online decision-making.

3.3.2.2 Priority Research Directions

In the following discussion, we highlight four areas of plasma research where great opportunities for scientific discovery exist within the next ten years. The questions we are asking are these:

1. **Can we model the interaction of an ensemble of laser speckles in direct-drive ICF over a millimeter of under-critical-density plasma?** Laser-plasma interaction (LPI) with laser intensities of $<10^{14}$–10^{16} W/cm^2 and nanosecond pulse durations determines the efficiency and uniformity of the laser-target coupling in direct-drive ICF (NRC 2003), as well as the properties of the blow-off plasma plumes in laboratory astrophysics experiments (Huntington, Fiuza, and Ross, et al. 2015). LPI also determines the generation of hot electrons that can be a preheat threat during ICF implosions, as well as an aid to increase the laser-coupling efficiency in shock ignition, a new high-gain ignition scheme (Betti et al. 2007). Major LPI effects include stimulated Raman scattering (SRS), stimulated Brillouin scattering (SBS), and two-plasmon decay (TPD); they can cause laser backscattering (via SRS and SBS), crossbeam energy transfer (via SBS), and hot electron generation (via SRS and TPD). The computational challenge for modeling LPI physics is rooted in the vast range of temporal and spatial scales, from the hydrodynamic scales of ns and mm to the laser scales of sub-fs and sub-µm.

   Modern hydro-codes include LPI effects via empirical models of bremsstrahlung, heat transport, fast electron generation, etc.; when code results disagree with experimental findings, it is often unclear which part of the modeling is at fault. At that point, kinetic physics-based descriptions of the micro-scale are required to verify these macroscopic models. State-of-the-art computer simulations of LPI can provide a closure for hydrodynamic models in a very direct way: a hydro-simulation of an ICF experiment can be inter-leaved with a series of (much more costly) kinetic simulations of LPI that use hydro conditions as input and generate a description of heating and fast particle generation as output. A major tool for kinetic modeling of the LPI physics is particle-in-cell simulation. Two main challenges of PIC modeling of LPI are in accounting for the interactions of different LPI modes in a wide density region and in establishing realistic laser conditions such as laser speckles.

   One goal is to model SBS, SRS, and TPD in the entire corona region of a direct-drive ICF target. Within the millimeter scale of the corona, different LPIs dominate at different densities; however, they can couple to each other through pump depletion and density perturbations carried by the plasma and ion acoustic waves. Current state-of-the-art PIC simulations are 2D simulations of a narrow (~40–100-µm) region near the quarter-critical surface (Yan et al. 2012). Figure 3-13 shows the evolution of the SRS- and TPD-induced plasma wave amplitudes near the quarter-critical surface, located at x=1200 c/ω_0 (Yan et al. 2014). Two strong bursts originate near the quarter-critical surface and propagate toward the left, a result of the interaction of different LPI modes. The simulation box size is 100×20 µm^2 and only covers a density range of 0.17–0.33 n_c requiring 10^{18} FLOPS. Extending these 2D simulations into the entire corona region not only requires sizes about 30 times larger (~10^{20} FLOPS), but we also need to overcome the large intrinsic noise levels in the PIC simulations. This is because the LPI modes in the low-density region are convective, and their saturation levels are proportional to the seed level. Proper simulation of the convective modes requires controlling the seed level and may require a combination of PIC-, fluid-, and Vlasov simulations. Extending the simulations to 3D will enable the study of side scattering and interaction of laser speckles but would pose even greater computation challenges.

Another goal is to model an ensemble of laser speckles over a millimeter of under-critical density plasma in three dimensions. Currently, simulations can model two speckles interacting in 3D or an ensemble of 100 speckles in 2D (Yin et al. 2012). Modeling a beam of 1,000 speckles in 3D would require exascale computing and beyond (i.e., ~10^{23} FLOPS).

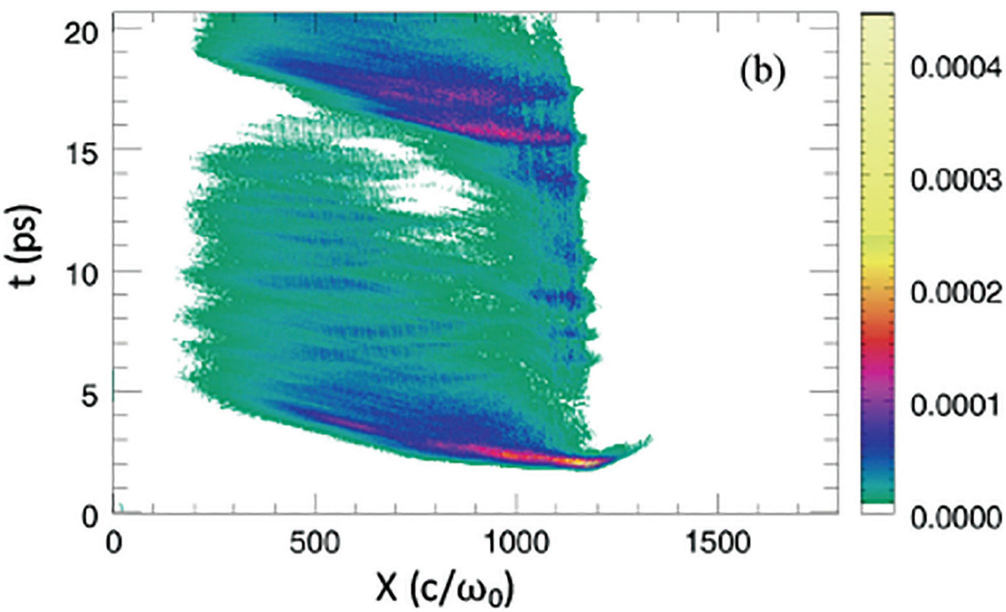


Figure 3-13. Evolution of the longitudinal electric field energy in a 2D PIC simulation for shock ignition (adapted from Yan et al. 2014). This image shows intermittent LPI activities due to interaction of different LPI modes near the quarter-critical surface. The simulation box is 100 µm × 20 µm and covers a density range of 0.17–0.33 nc. The density scale length is 170 µm.

2. **Can we model the interaction of an intense sub-picosecond laser pulse with a solid density target to scale?** Relativistic laser pulses at intensities of $>10^{18}$ W/cm^2 at a 1–µm laser wavelength are capable of accelerating electrons to relativistic energies within a single laser period, that is, about 3.3 fs; understanding the interaction of relativistic laser pulses with matter is a grand challenge that is motivated by applications such as the fast-ignition approach to inertial confinement fusion (Figure 3-14) (Kemp and Fiuza et al. 2014); the development of compact, inexpensive plasma-based particle accelerators for electrons and ions for medical applications; and sources of radiation for science, industry, and medicine (Plasma 2010 Committee 2010; DOE-SC and NNSA 2009; Macchi and Borghese 2013; Fiuza et al. 2012).

High-resolution modeling of laser-driven charged-particle acceleration (Leemans et al. 2015; Steinke et al. 2015), flying plasma mirrors (Leemans et al. 2015), high-intensity laser-matter interaction (Bulanov et al. 2015), laser-driven X-ray sources (Geddes et al. 2015), collisionless shocks (Huntington, Fiuza, and Ross et al. 2015; Fiuza et al. 2012), for example, are typically modeled with kinetic particle-in-cell simulations and necessitate a very large number of cells and macro-particles that stress computational resources. In various cases, such as the acceleration of ions through the interaction of intense lasers with plasmas, rendering high-resolution simulations that capture the phenomena with high accuracy is possible only in 2D. With that limitation, fairly high-resolution simulations will be possible in the 2020–2025 time frame.

As an example, a 2D simulation of a $(100\text{–}\mu m)^2$ volume of plasma at solid density at a numerical grid resolution of 5 cells per plasma skin depth (or 400 cells per micron) with 10^4 particles per cell requires about 2×10^{15} bytes of memory; in principle, such a system can be modeled on existing tier-1 machines such as LC-Vulcan at Lawrence Livermore National Laboratory (LLNL) or Mira at Argonne National Laboratory (Argonne). However, the same simulation in 3D would use 4×10^4 times more memory. Earlier simulations of multi-picosecond laser pulses interacting with over-critical density plasma, shown in Figure 3-14 (left/right), using a grid resolution of 50 cells per micron and ~100 particles per cell, were performed in 2012 on 10,000 cores of LC-Sierra and cost about 300,000 CPU-hours. Additional computational challenges arise from including radiation transport in particle simulations, an approach taken by Sentoku et al. (2014); including radiation transport enables modeling transport of intense X-ray beams in matter, as well as radiative losses of laser-driven plasmas. Depending on the number of photon groups and directions that need to be resolved in a multi-group diffusion scheme, the memory requirements for radiation transport can actually exceed those for particles.

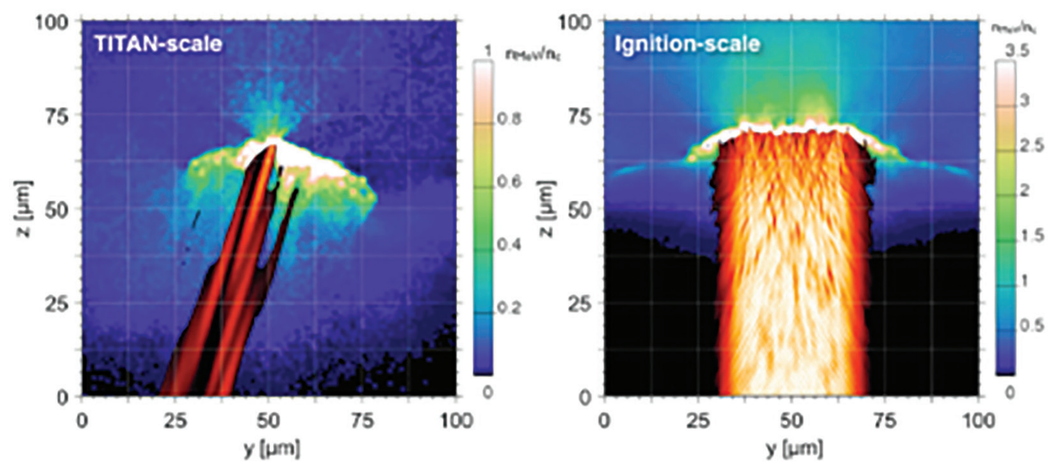


Figure 3-14. PIC simulations of intense laser-matter interaction; (left) showing the interaction of the TITAN laser pulse, and (right) a fast-ignition-scale laser pulse incident on a solid target; laser Poynting flux (red-black); energetic electron density (white-green-blue) (Kemp and Fiuza et al. 2014).

3. **Can we probe cosmic-ray ("Fermi") acceleration in the lab?** The origin of energetic cosmic rays remains one of the long-standing problems in astrophysics. It is believed that particle acceleration in collisionless shock waves and magnetic reconnection may hold the key to this problem. However, the direct study of this physics in astrophysical systems is limited. Laboratory astrophysics studies are seen as a possible route to study directly the fundamental plasma physics associated with these astrophysical environments. In particular, high-power laser-plasma interactions can generate high Mach number collisionless flows that would allow the study of particle acceleration in shocks and magnetic reconnection in the laboratory for the first time (Huntington, Fiuza, and Ross et al. 2015; Fiuza et al. 2012; Fiksel et al. 2014; Chen and Fiuza et al. 2015; Totorica et al. 2016). Numerical simulations play a critical role in the design and interpretation of these high-energy-density plasma experiments, where in many cases diagnostics are significantly limited. The study of particle acceleration in these systems requires capturing the kinetic processes occurring at microscopic scales. Thus, PIC codes are often used

to model these experiments from first principles. This poses great computational challenges owing to the need to model at the same time both the microscopic scales and the large scales associated with the macroscopic system evolution. Because of this challenge, studies often used different codes to capture different parts of the problem. Typically, hydrodynamic or MHD codes are used to model the laser-plasma interaction of formation of the plasma flows, and then PIC codes are used to model the interaction of the flows for the conditions predicted by hydrodynamic simulations. Even with this approach, just simulating the interaction of the plasma flows in two dimensions can be very demanding. A typical 2D simulation of shock formation uses 2×10^9 cells, 2×10^{11} particles, and runs for 2×10^5 time steps, requiring approximately 10 million CPU-hours. Despite the already tremendous challenges posed by 2D simulations, it is important to note that in many cases 3D simulations are mandatory, either because the physics is intrinsically 3D (e.g., turbulence in colliding laser-driven plasma plumes) or because the experimental diagnostics are 3D and the comparison with experimental data can only be made by means of 3D simulations (Huntington, Fiuza, and Ross, et al. 2015). In these cases, even using the largest supercomputers in the world to run simulations on the order of 10 million CPU-hours each, reduced ion-to-electron mass ratios must be employed, which therefore does not allow for a correct scaling of all physics processes. This is the best performance that is currently possible when designing ongoing laboratory astrophysics experiments; see Figure 3-15. By 2020–2025, it will be possible to approach realistic mass ratios in 3D for simulating some of the aspects of these systems, although it will likely still not be possible to run a full-3D simulation of the entire system with realistic mass ratios by then.

4. **Can we quantify the role of ion-acoustic turbulence in laser-driven nonthermal heat flows?** As one example of a kinetic effect in ICF-related physics, ion-acoustic turbulence has been identified as a potential mechanism that is responsible for anomalously increasing the electron-ion "collisionality" and thus absorption at near-critical density. Modeling of this instability with collisional particle-in-cell codes requires (1) resolving the plasma Debye screening length to capture kinetic effects; (2) particle statistics to accurately describe the energetic tail of the electron distribution function that carries the heat flux; and (3) performing many (on the order of tens of millions) time steps to bridge the gap between kinetic scales (i.e., sub-femtoseconds) and hydro timescales on the order of nanoseconds. This example is typical of a larger group of kinetic problems that have in the past been addressed with multi-fluid or fluid-kinetic simulation models, ignoring electron plasma waves and the complicated interplay between fluid and kinetic physics.

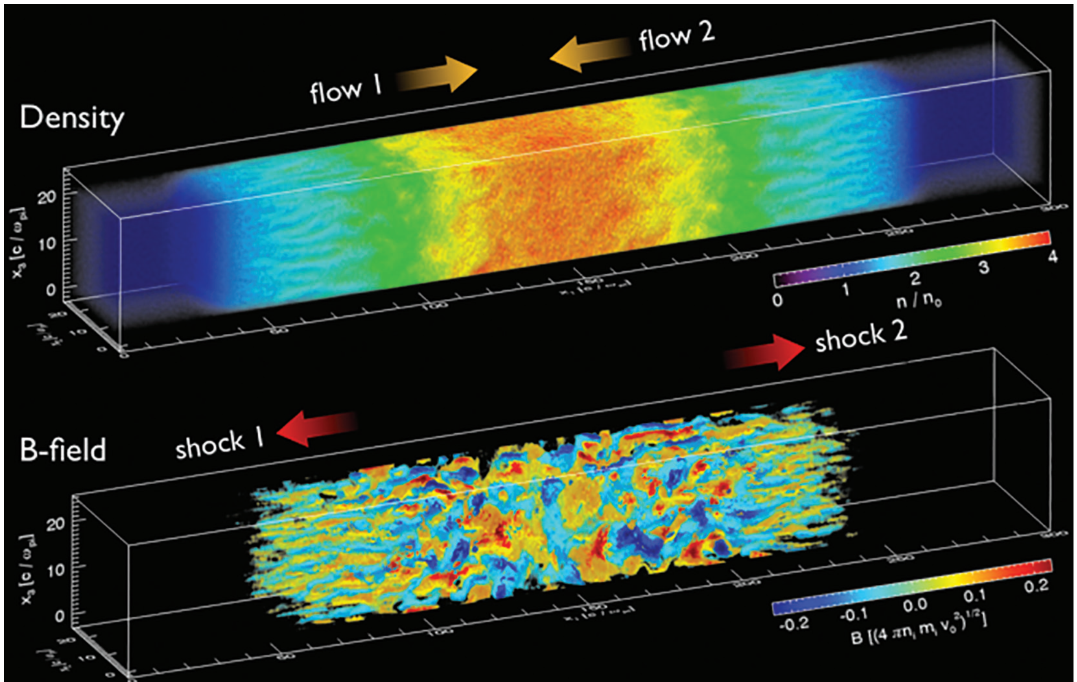


Figure 3-15. 3D OSIRIS PIC simulation of the formation of a collisionless shock for National Ignition Facility conditions. Two laser-driven counter-streaming plasma flows with velocity of 2,000 km/s and density of 1,020 cm^{-3} interact in the central region, leading to the development of a shock mediated by the Weibel instability. The strong B-fields (up to 300 T) thermalize and slow down the initial flows, leading to a density compression of 4, which is consistent with hydrodynamic jump conditions. Because of the outstanding computational challenges posed by the need to model the kinetic (electron skin depth) scales (sub micron) and the system size (cm), these simulations are currently limited to reduced ion to electron mass ratios of ~100.

3.3.2.3 Cross-Cutting Research Directions

High-energy-density laboratory plasma physics overlaps with other areas in Discovery Plasma Sciences, in particular with Low-Temperature Plasmas (LTPs) and General Plasma Physics, in terms of computational methods as well as structural aspects. All topic areas use particle-in-cell descriptions of plasma and involve laser-plasma interaction to set up experiments. Across all three topic areas, researchers will benefit from the following investments in:

1. Hybrid algorithms capable of capturing both kinetic *and* magneto-hydrodynamic features by either using MHD codes to initialize plasma conditions for much more expensive kinetic simulations or by creating workflows that consist of a loop between the kinetic model of laser energy deposition and MHD energy re-distribution and convection of plasma.
2. Open-source codes and modules, which will benefit the wider Discovery Plasma Science community.
3. Particle solvers with low-numerical noise and diffusion properties, and Maxwell solvers with low-numerical dispersions and relaxed CFL.
4. High-dimensional partial differential equation (PDE) solvers (i.e., 6D Vlasov).
5. Generation of accurate Atomic Molecular and Optical (AMO) databases as required by LTPs, which will benefit the other topic areas.
6. Code development for plasma-based particle accelerators for FES/High Energy Physics and other DOE agencies will benefit FES.

7. Adaptive Mesh Refinement combined with Adaptive Particle Refinement coupled to Lagrangian distribution of particles (Wang, Miller, and Colella 2011). See also Section 3.3.2.4, Computing Needs and Requirements.

Computer simulation codes for HEDLP and plasma physics in general are becoming increasingly complex. As the entire community needs to port its codes to new architectures, facing similar hurdles, it makes sense to encourage doing so collaboratively rather than independently. It is thus desirable to develop ecosystems of modules with common front-end, output formats, and visualization/analysis tools with kernels that can be developed collectively or separately, all integrated into a suite of interoperable components. There is a need for community software for optimal usage of community hardware.

3.3.2.4 Computing Needs and Requirements

Large-scale particle-in-cell simulations generate extreme amounts of numerical output; a three-dimensional simulation of a $(100\text{–}\mu m)^3$ plasma volume with 400 cells per micrometer and 10^4 particles per cell, where each particle is described by 10 double-precision numbers, generates $\sim(100\times400)^3 \times 10{,}000 \times 100 = 6\times10^{19}$ bytes of particle data per time step output. To work with these amounts of data, it is essential to have robust, scalable, and easy-to-use I/O tools, like HDF5 or ADIOS; optimizing I/O on extreme computational scale is “still too much of a black art.” It is expected that for detailed high-frequency data analysis, the required I/O performance levels will greatly surpass the expected availability, necessitating efficient data reduction and *in-situ* analysis/visualization capabilities. In this regard, common, versatile, high-quality, and easy-to-use tools for post-processing and *in-situ* visualization and analysis are needed. Efforts toward standardizing and optimizing parallel I/O for PIC codes are currently under way (e.g., OpenPMD at https://github.com/openPMD), which will facilitate data exchanges between codes, data processing, and visualization software. For some codes, the support of Python as front-end to number-crunching modules is essential.

The hardware and associated software of exascale supercomputers will involve highly specialized programming skills, requiring proper continuous training as they evolve over time, as well as close partnership between applications developers, users, computer scientists, and applied mathematicians. For maximum benefits to the community, the suites of tools will benefit from more intuitive interfaces, wide dissemination, extensive documentation, and resources for user support.

3.3.3 Low-Temperature Plasmas

3.3.3.1 Scientific Challenges and Opportunities

LTPs are partially ionized gases with electron temperatures in the range of 1–10 eV. Background gas pressures typically range from about 1 mTorr to 1 atm. LTPs are usually sustained with modest amplitude driving electric or electromagnetic fields with frequencies ranging from direct current (DC) to microwaves applied in a wide variety of configurations. Because of coupling efficiencies and power supply costs, RF plasmas at or around 13.56 MHz are most common. Applications of LTPs are manifold, including materials processing for tools, electronic materials, and surface treatments, as well as lighting, thrusters, biomedical interactions, and many more.

Low-temperature plasmas are typically strongly driven, partially ionized, and highly collisional. These processes typically take place at rates faster than relaxation to equilibrium, leading to nonequilibrium, with the electron temperature much higher than the ion temperature, which is in turn higher than the neutral gas temperature. Components of the distribution, especially in strong field regions, can be highly energetic. For example, in the strong fields of the sheath near the plasma-materials interface, charged particles can be accelerated strongly. At driven electrodes, this property is even more extreme. Most inelastic collisions in LTPs occur in the tail of the distribution, including ionization, generation of radicals, excited states, and so on. Small modifications of the tail of the distribution function can lead to large variations in the inelastic collisions rates, and dramatically change the nature of the plasma and its constituents. Indeed, manipulation of the distribution function enables optimization of LTPs for various applications.

The societal impact of low-temperature plasmas is significant and broad. LTPs are involved in about 70% of the steps in the manufacture of the ubiquitous electronics components that drive modern civilization, for example, and have enabled the continuation of Moore's law well beyond the feature sizes that can be manufactured using chemical techniques. One of the fastest-growing areas in plasmas is that of biomedical plasmas, which has current applications in surgery, wound healing, and sterilization, with the promise of many future applications yet to be discovered. A recent whitepaper on low-temperature plasmas (Kushner et al. 2014) described several key performance metrics on the economic impact of low-temperature plasmas that were based on scaling a detailed study of impact on the German economy to the U.S. economy. It found that about 600,000 jobs are directly attributable to plasma technologies, which indirectly enable about 4,000,000 jobs, with about $650 billion in current economic activity and growing at an annual rate of about 10%. Kushner et al. (2014) also estimated a worldwide market of about $70 B for plasma-treated flexible packaging material, a worldwide market of about $740 M for plasma-based ozonizers used in water treatment, and a global market of $20 B in plasma-based physical vapor deposition coatings.

Low-temperature plasmas play a key role in magnetic fusion energy. Neutral beams, a key current and heat injection system in tokamaks and other strongly magnetized configurations, begin life as a low-temperature plasma. Ions are extracted from an LTP discharge, accelerated, and then neutralized. The neutrals retain most of the directed momentum as a beam, and are able to cross strong field lines and deposit energy deep in dense core plasmas where they interact collisionally with the plasma. In addition, many of the theories and techniques for high-temperature plasmas were first developed and tested in LTPs, which are more amenable to laboratory measurements and modest cost experiments. Hence, LTPs remain a key tool for development of techniques and technology for fusion plasmas, both directly and indirectly.

Perhaps the chief science driver in LTPs is control of the velocity distribution function, $f(\mathbf{r},\mathbf{v},t)$. Surface wave microwave sources, hybrid RF/DC sources and direct e-beam injection have all emerged over the last decade. Control of the distribution function of electrons allows selectivity among the myriad atomic and molecular reactions occurring among all the plasma species, control of the energy fluxes, and hence the ability to perform specific applications. For example, by tuning the distribution function near a peak for a particular reaction, researchers can increase the probability of that reaction occurring over others. This capability can, in turn, be used to control the species in the plasmas, the radicals and excited states, photons, and heat flux to the wall. Likewise, localized control is also possible, enabling development of spatially varying profiles of species and emissions. In practice, control of the distribution function is challenging, as it involves balancing the energy-dependent deposition of power primarily to the electron distribution at a broad range of frequencies up to the electron plasma frequency, and the deposition of power to the ion species at frequencies below the ion plasma frequency. This frequency-dependent behavior can be exploited to affect the electrons and ions somewhat independently in order to achieve specific physics goals.

The processing of materials by plasmas is one application of LTPs which benefits from control of the distribution functions and consequently species profiles and fluxes to the substrate being processed, particularly radicals, energetic ions for deposition, etching, implantation or sputtering, and photons. An important application for plasma control is the world of “post 5 nm” semiconductor device manufacturing. Plasma processing for device manufacturing provides incredible opportunities for scientific discovery and paradigm changing impact, because with 5 nm and smaller devices, control of impinging on the surface fluxes needs to be made on the level of single atoms, electrons, and ions. Dopants must be inserted with atomic-scale precision; with films deposited with atomic-level uniformity over scales of nearly 1 m, atoms must be removed with the same level of precision. Plasmas must control everything, everywhere all the *time* on a surface. Controlling *everything, everywhere all the time* means *chemistry control* and chemistry control ultimately means c*ontrolling the fate of electrons* in the plasma and at surfaces. In the parlance of plasma science, this standard translates to electron energy distribution function control.

A number of other areas pose scientific challenges:

- Low-pressure 3D WDM, including engineering studies, small-scale features, validation, and uncertainty quantification.
- High-pressure discharges, where high density and high collision frequency constrain the time step, where large ratio of system length to Debye length constrains the space scale, and where particle models require ~100 particles per cell to reduce the effects of fluctuations.
- Power coupling mechanisms in helicon plasmas.
- Plasma formation and transport in ion thrusters.
- Cross-field transport in LTPs.
- Atmospheric pressure discharges, including microdischarges.
- Multiphase flows involving complex chemistry and interaction with liquids, including biomedical plasmas.
- Calculation of cross sections, rate constants, and surface interaction coefficients such as secondary emission and sputtering yields.

These challenges include both fundamental science questions as well as applied engineering issues of prediction and design. The scientific challenges are exemplars of the thrusts of current studies, with varying degrees of maturity. The scale of many of these computations is such that they can presently be performed with modeling restrictions. Restrictions might include scaling to modest densities, limited physical extent, lower pressures, and other simplified physical models to reduce the computational cost. Many LTP models, particularly at high pressure/density, are limited to two or even one dimension of variation, which precludes many more complicated plasma behaviors. Ad hoc approximations are also employed, with uncertain implications for the accuracy of the computations. Frequently, geometric fidelity is limited by the range of space scales needed to represent a full device, limiting the fine structure representation, which can be of increasing importance as materials with a micro-structured surface are developed to engineer enhanced surface properties. Many of these constraints would be relaxed or removed with increased computational power. This capability is even more crucial when studying the quality of computations, including verification and validation, and uncertainty quantification, which require many runs to bound errors and build statistical representations, which can increase computational cost by one to two orders of magnitude.

This discussion suggests that both the scope and quality of scientific computations in low-temperature plasma physics would benefit from exploiting enhanced HPC resources. Exascale resources will provide a capacity improvement of many orders of magnitude, suggesting that many of the aforementioned problems could be addressed. Although most low-temperature plasma research groups have at least limited access to high-performance computers of at least modest scale, few are making heavy use of these resources. A number of reasons seem to play a role:

1. Unstable programming models have discouraged the development of high-performance codes. The low-temperature plasma community is scattered and thinly resourced. Consequently, the development cycle for computer codes can extend over decades (e.g., HPEM [Kushner 2009], PDP1 [Verboncoeur et al. 1993]). However, disruptive changes in the HPC computing model have occurred several times since the development of these codes was initiated. The resources likely have not existed to follow these changes.
2. The community has been insular with respect to methodological improvements that increase the scale of the computing problem. Other communities (fluid dynamics, combustion science) have been energetic in developing techniques that improve the quality of their computations, in various senses. The low-temperature plasma science community has not attended much to these developments. Indeed, in many ways, basic concepts and approaches first seen in the early 1990s are still the foundation of most computational work, despite the pace of development in cognate fields.
3. Concerns about basic data undermine confidence in predictive power. In most low-temperature plasma physics contexts, prediction involves extensive use of information such as cross section data, which historically has been patchily available and of variable quality. This consideration has produced widespread anxiety about the consequences for accuracy of prediction, but with not much action taken to address the issue, with the apparent result that prediction has not been taken very seriously.

Opportunities for transformational understanding are possible by addressing these physical challenges. Exascale computing can make this possible by providing the ability to address the accuracy of computational results, in order to develop confidence in simulation for engineering analysis and design. The capability to provide high confidence models will transform the applied use of LTPs in industry, where capital investments and consequent business success depend upon getting the correct answer, and increasingly in knowing the error bars. Exascale-enabled verification, validation, and uncertainty quantification is one such game changer. Elimination or decreased reliance on ad hoc and simplified physics models is another such disruptor. In particle and Monte Carlo models, increased statistical representation can reduce statistical uncertainty, leading to better convergence and higher quality results. The ability to perform convergence studies to ensure a fully converged solution also required significant compute resources.

Another opportunity is the use of higher-fidelity physical models, from improved geometric representations including finer features and conformal meshes, to first-principles models of plasma-surface interactions and better resolution of high-energy tails in particle distributions.

Low-temperature plasma physics simulations must attempt to capture disparate length and time scales, typically within severely constrained computer resources. A widely adopted approach is to construct a modular simulation, in which modules operating at different scales exchange data in some iterative fashion (Kushner 2009). For example, a common approach is to combine a microscopic (or kinetic) treatment of electrons with a macroscopic (or fluid) description of neutral species. The verification procedure described in the previous paragraph cannot be applied to a simulation of this kind because the exchange of data between modules is informally specified, so there is no clear underlying mathematical model, and the convergence of the method has no well-defined order. Moreover, even if these problems could be addressed, the presence of optional or alternative modules in such codes is liable to lead to an impractical proliferation of verification test cases. In short, such codes are, in terms of present recommended procedure, unverifiable.

3.3.3.2 Priority Research Directions

Solution of the scientific challenges listed above set the research priorities. Unordered research priorities include the following:

- **Kinetic Treatment of Coupled Neutral-Plasma Systems.** Well-resolved solutions of nonlocal, inhomogeneous, strongly driven kinetic plasmas, including resolution of the high-energy tail. Enhanced particle density results in lower numerical heating, improving the convergence of the result especially for longer simulation times needed to include low-frequency effects, such as ion and neutral transport, surface modification, and collisional effects.
- **High-pressure Whole Device Models.** Atmospheric pressure plasmas are among the most important economically, especially in air. When in air, these plasmas may not require complicated pumps and high vacuum systems.
- **Hierarchy of Uncertainties in LTPs.** Understanding the fundamental hierarchy of uncertainties in LTPs, including the relative importance of statistical representation, atomic and molecular data, and accuracy of physical models including surface and bulk physics.
- **Verification, Validation, and Uncertainty Quantification to Engineering Analysis and Design Problems in LTPs.** Addressing these priorities will enable confidence in the degree of convergence of solutions, and especially in the effect of error bars in input parameters on the error bars in the final solution. This effort can be particularly challenging with complicated chemistry in the model.
- **Understanding the Plasma-Liquid Interface.** This challenging multiscale model is in its nascent stages with one-dimensional (1D) models at low density. Ultimately, a true challenge in this space is the modeling of atmospheric pressure bio-plasma systems, which are of increasing importance in plasma medicine.
- **First-Principles Coupled Plasma-Surface Modeling.** The ability to self consistently simulate the plasma-surface interaction with self-consistent models will enable the coupled calculation of the interaction. Present models are limited to weakly coupled and phenomenological models, which lack proper feedback between models.
- **First-Principles Calculation of Fundamental Atomic and Molecular Data.** The ability to compute cross sections, rates, transport coefficients, and plasma-surface interaction constants is a transformational one, which will enable the high-confidence calculations necessary for engineering modeling.

3.3.3.3 Cross-Cutting Research Directions

A number of cross-cutting research directions are key in low-temperature plasmas. The verification, validation, and uncertainty quantification methodology and requisite computational resources are common across a spectrum of plasma parameter regimes.

The understanding of plasma heating mechanisms, and the impact on kinetic velocity distribution functions, is an important cross-cutting area. This area is one in which model development and experiments and diagnostics are much simpler at the LTP scale — a finding that suggests that LTPs should again be used as the initial environment for developing understanding and modeling techniques, in addition to the experiments and diagnostics.

Another key cross-cutting area is in the plasma materials interface, including the interaction models for charged and neutral particle absorption, emission, reflection, and sputtering. Although low-temperature plasmas operate in a much lower energy regime, and usually in a lower flux regime as well, compared to magnetic fusion plasma facing material interactions, the fundamental physics is similar. In addition, the lower energy and flux environment of LTPs lends itself to simpler laboratory experiments and diagnostics that can be used to develop modeling and testing capabilities for later scaling to the fusion regime. The ability to rapidly test and model at the LTP scale may prove crucial to advancing science at the fusion scale.

3.3.3.4 Computing Needs and Requirements

For simulations of plasma devices for plasma processing and electric propulsion applications, we will be able to perform two-dimensional (2D), electrostatic, and electromagnetic particle-in-cell Monte Carlo collisions (PIC-MCC) simulations of typical low-temperature plasmas without any scaling of physical parameters. Such a simulation might use 10,000 cells to resolve 10 cm in each linear dimension with 100 to 1,000 particles per cell. Computation on the particles scales well for parallel codes, and by using multigrid for the Poisson solve, overall scaling to 10 or 100,000 cores can be achieved. The workflow would be to use capacity computing on a system with scaled-down parameters and then switch to capability computing for one, or a few, simulations with true, physical parameter values. Fully kinetic three-dimensional simulations of realistic plasma devices still would have to utilize scaling to reduce the linear system size by at least a factor of ten unless a much larger number of cores could be used for suitably designed and highly scalable code.

Random collisions are at the heart of PIC-MCC. Unfortunately, random particle trajectories scramble spatially ordered particle data in computer memory, which causes cache misses. Continuous sorting of the particle data is also costly and can, at best, only mitigate the problem. As cache misses counted in clock cycles becomes costlier for each new generation of computers, the inherent data scrambling makes PIC-MCC less and less efficient.

Estimates of computing requirements in LTP can vary over many orders of magnitude. For example, a PIC model of a modest crossed-field device at modest temperature can require $\sim10^{11}$ grid cells in 3D to resolve the Debye length, which implies $\sim10^{13}$ particles to maintain modest statistical noise levels. This amount corresponds to ~600 TB of total memory and requires on the order of 150,000 cpu-years. This need can be reduced to ~300 hrs using 10^7 cpus at 50% parallel efficiency. If scaling this problem to extend the range of physics that can be studied — for example, to resolve high-frequency effects and reduce statistical fluctuations — the memory requirements would increase to ~3 exabytes.

Atmospheric pressure plasmas can increase the computing requirements by at least four orders of magnitude, given that densities can reach 10^{21} m^{-3} and up. This level scales the space scales, timescales, memory, and core requirements. In 3D for practical device sizes, this problem can become one to be addressed by exascale resources.

3.4 Verification and Validation

3.4.1 Needs Driven by Experiments

3.4.1.1 Scientific Challenges and Opportunities

There are four main HPC drivers arising directly from fusion experiments:

1. **Experimental Operations Planning.** Experiments are usually motivated by or connected to tests of particular theoretical or modeling questions, but only a fraction of experimental proposals are currently qualified ahead of time by extensive HPC modeling. However, this sort of modeling can be particularly useful for operation in unfamiliar or challenging parts of parameter space. With improvements anticipated in capabilities and fidelity of WDM as an element in experimental planning, activity in this area is likely to increase and should lead to more successful operation and more efficient use of expensive run time. Looking further into the future, it is anticipated that proposals for run time on ITER or devices of that class would only be approved if accompanied by extensive discharge modeling (Greenwald et al. 2015).
2. **Run-Time Experiment Steering.** While remote HPC will likely not be used in real-time control loops in the present-day short-pulse tokamak experiments, its use for between-shot operations steering has already been demonstrated (e.g., White et al. 2013, where between-shot linear gyrokinetic modeling was employed using a local 500-core cluster at the Massachusetts Institute of Technology). This approach entails accelerated workflows that can provide preliminary analysis of experiments in minutes rather than days or weeks. The result is that experimental operations can be adjusted throughout a run day to more precisely meet the experimental goals and to increase overall productivity of the facility. Early experiments in queuing for near real-time applications are currently being tested at NERSC by the General Atomics group during its current experimental campaign. The trade-off for time spent in data transfer can be favorable for computationally intense calculations. In a steady or near-steady operation of fusion reactors, the technique can be extended to the real-time control of the experiments. Technical challenges remain, as discussed below.
3. **Code Validation.** A large fraction of run time is devoted to tests of particular theoretical ideas, often backed by simulation results, or for explicit comparisons with code predictions. The increasing rigor of validation activities will likely drive much larger computational requirements — especially for sensitivity analysis.
4. **Interpretation of Experimental Data.** Comparisons between simulations and experimental data can generate new hypotheses about the physical phenomena being observed. This activity is less formal than VVUQ but can be an essential element of scientific discovery.

3.4.1.2 Priority Research Directions

Each of the drivers listed above has implications for research directions:

1. **Experimental Operations Planning.** Fast turn-around and large numbers of runs characterize this application. Research will be required to create the reduced or surrogate models that are well tested against full-physics calculations.
2. **Run-Time Experiment Steering.** This capability requires exceptionally fast turn-around for submission, queueing, execution, and post-processing of HPC tasks. This application requires more complicated data workflows — for example, reading data stored externally to the HPC system — that are currently not utilized or possible due to firewall related issues. Research and development are likely required to arrive at widely accepted solutions for federated authentication and authorization, credential forwarding, etc.

3. **Code Validation and Uncertainty Quantification.** These requirements are covered in Section 3.4.2 of this report.
4. **Interpretation of Experimental Data.** This application would include both capacity and capability-type computing. Research needs would generally be aligned with those identified for the science topical areas (Sections 3.1–3.3), whole device modeling (Section 3.1.4) and the V&V discussion in Section 3.4.2.

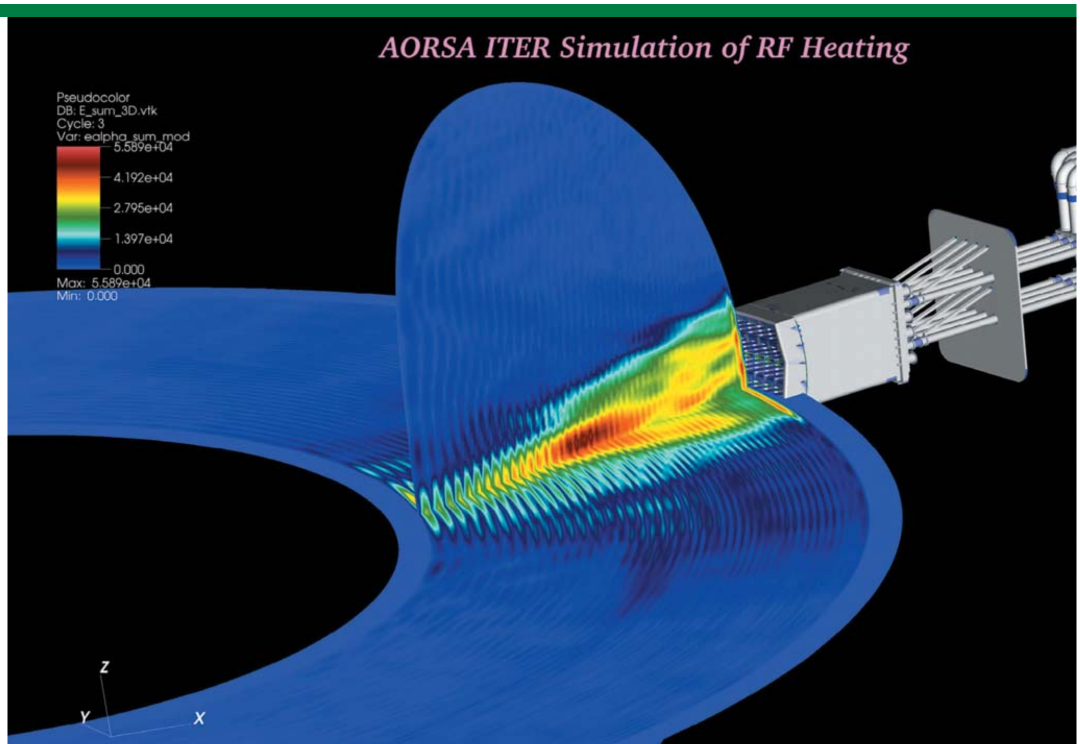
3.4.1.3 Cross-Cutting Research Directions

The requirements listed in all topical areas apply here.

3.4.1.4 Computing Needs and Requirements

In aggregate, these requirements are likely to affect the overall computational demands for plasma, that is, a significant fraction of computation in our field has a tight connection to experiments. Computational needs would generally be aligned with those identified for the science topical areas (Sections 3.1–3.3) and whole device modeling (Section 3.1.4). Needs beyond raw computational power, particularly the requirements for *in-situ* analytic techniques, improved metadata capture, and other data management issues, are mostly covered in Section 3.4.2 (V&V).

A VIEW OF THE SCIENCE



Three-dimensional simulation of ion cyclotron heating in the ITER burning plasma using the combined full-wave/Fokker-Planck model AORSA/CQL3D. Reproduced from Jaeger, E.J., L.A. Berry, E.D'Azevedo, et al., *Physics of Plasmas* 15, 072513 (2008).

3.4.2 Code Verification and Experimental Validation

3.4.2.1 Scientific Challenges and Opportunities

3.4.2.1.1 Motivation/Background

VVUQ can be thought of as no more than the scientific method extended into areas where the consequences of theory can only be understood and computed through complex simulations. Confidence in our models must be earned through careful and quantitative estimates of errors and uncertainties and systematic confrontation with experimental data.

These efforts are beginning to have an important impact on research directions and computational requirements in our field. Verification asks the question: Does the code solve the model equations correctly? Validation seeks to quantify the extent to which the equations and boundary conditions accurately describe the physical phenomena of interest, i.e., are we solving the right set of equations using the right set of boundary conditions? UQ is a closely related and overlapping set of methodologies — stressing the quantitative assessment of errors and uncertainties. Essentially, it seeks to provide meaningful error bars for simulations.

VVUQ can have somewhat different implications in the context of engineering vs. science. For engineering, the goals are entirely conditional — a calculation (not a code or model) is validated, for a specified set of outputs, in a particular regime, at a specified level of accuracy — a code is formally (or even legally) qualified for a particular application but no general guarantee is implied. There is a clear and direct application of V&V, in this case, to risk management and decision making. For a science goal, we are asking more generally about the correctness and adequacy of our physical models. In this case, the codes embody some aspect of our scientific understanding, and validation is intended to test that understanding and whether we can attribute a given discrepancy between code results and experiments to particular elements or the lack thereof in our models. Fusion plasma research and other areas of applied science mix these two approaches with a significance that should be taken into account in the design and interpretation of VVUQ activities. Prediction — outside the range of parameters for which data already exist — is a stated goal of the fusion program. This approach requires a high degree of confidence that we have a reliable and applicable theory.

In both cases, VVUQ is aimed at building confidence in models, to make the statement of confidence more mathematically sound and meaningful — especially for use in decision making. We note that a major program element in the FES strategy is to produce "validated predictive models" for the purpose of extrapolating to future machines operating beyond the regimes already achieved. This is clearly a long-term, ambitious, and somewhat ambiguous goal. Therefore, we should not underestimate how difficult it will be to realize it. To make the goal of a validated predictive capability meaningful and achievable, we will need to establish acceptance criteria for predictions (what do we need to predict and to what level of confidence?), define metrics against these criteria, adopt methodologies which ensure that our inferences are justified, and quantify our level of confidence — how do we measure success or failure? A shorter-term but crucial benefit of VVUQ is the ability to identify the most important shortcomings in computational models and use that information to improve them.

3.4.2.1.2 Challenges

VVUQ is a rich area of research in many technical fields, and while plasma and fusion science researchers are increasingly engaged in this set of problems, it is essential that they share and collaborate across disciplines. Our challenges are to find methodologies and algorithms suited to our problems, to identify gaps where additional research in applied math and computer science is needed, and to apply those techniques to specific codes and simulations.

Verification. Standard methods of code verification include various software engineering techniques, convergence tests, code-to-code benchmarking, the method of manufactured solutions (described in 3.4.2.2, priority research directions), and comparisons with analytic theory. This last method represents a gold standard for verification but is difficult to apply for many plasma physics problems. Even for single physics models, closed-form solutions are often only possible at extremes in parameter regimes — often where codes do not converge and in any case not in regimes of relevance to applications. Coupled physics models are typically beyond the reach of any analytic solutions. The method of manufactured solutions has been applied to plasma turbulence problems in the fluid regime. The extension of that methodology to coupled physics models is new and has apparently not been carried out in our domain. As a consequence, the popular tools used are detailed convergence studies and code-to-code comparison.

UQ. The most important sources of uncertainty in plasma and materials simulations are the fluctuating nature of the nonlinear dynamics and the propagation of uncertainties from assumed parameters and initial or boundary conditions. Errors arising from the discrete mathematics used in the computation can typically be made subdominant and are usually easier to quantify. Simulations must run with sufficient ensembles of particles or waves, or run long enough to allow meaningful averages to be taken and for the transient effects of initial conditions to die away. We note that although we generally believe these uncertainties to be subdominant, rigorous techniques for quantifying them in plasma simulations have not been widely pursued or routinely deployed by the community.

For plasma problems, the forward propagation of errors from input parameters or assumptions is probably the most important source of uncertainty in simulation and the most challenging to address. Notable examples include the profile gradient drives for turbulence or MHD and uncertainties in atomic physics cross sections or materials properties for PMI models. The approaches for addressing this set of problems come under the heading "Sensitivity Analysis." The goal is to identify and quantify the most important sources of uncertainty as measured by their impact on the output quantities of interest. The dimensionality of the input/parameter space presents a mathematical and computational challenge and makes brute force approaches unappealing if not prohibitively expensive for the numerous calculations. Another challenge is that the statistical/uncertainty properties of the input parameters are often poorly understood. In the so called "flux-driven" simulations, the number of input parameters for the sources and sinks are finite. The initial input plasma parameters are less important for the final self-organized state, which produces the experimental observables. Thus, UQ or the sensitivity analysis is of less challenging. However, the flux-driven simulations require much more computing resources than the profile-gradient driven simulations since the profile gradients are evolved together with the turbulence.

Validation. Once the uncertainties of the computations including their dependence on input parameters are understood, detailed validation studies can be carried out. The challenge here is to make a meaningful interpretation of the comparison between experiments and simulations. How strong a test is the comparison? How much does the result change our confidence in the model? This is a statistical inverse problem that may be best addressed with Bayesian inference techniques.

Obtaining rigorously derived experimental uncertainties can be challenging — particularly through the analysis workflow chains required. Although emerging statistical models offer promise, this is still an area of active research. Choosing appropriate "quantities of interest" for the comparison is crucial. These should represent quantities that we need to predict in a self-consistent final state and for which corresponding experimental measurements must be available. The challenges are heightened for nonlinear multi-physics, integrated models. Because of their tightly coupled physics, it is easy to be misled about what is incorrect in such a calculation. It is probably best to approach physics integration gradually, for example, through what has been called focused integration initiatives (FESAC 2002) or Integrated Science Applications.

3.4.2.2 Priority Research Directions

We identify four priority research directions in the VVUQ area:

1. **Develop/identify improved methodologies for code verification, especially for coupled/integrated physics models.** Software quality assurance and engineering techniques are often — but not always — applied. Some activity to identify and implement "best practices" could be useful. As noted above, despite the best efforts and intents of developers, verification typically depends on convergence studies and code-to-code benchmarking. The method of manufactured solutions (MMS) may be practical in our domain, and could be a target of opportunity for further research. This method functions by choosing, a priori, a manufactured solution to a modified set of equations that consists of the original equations plus an analytic source term. The task for developers is to make an intelligent choice for the manufactured solution and then to solve the inverse problem — find the source function that is required to make the manufactured solution correct. This method has been successfully applied to plasma fluid turbulence and computational fluid dynamics, including problems involving multi-physics couplings, suggesting that there may be a broad set of problems in the plasma physics domain that are amenable to this approach.
2. **Assess and extend existing methods (intrusive and nonintrusive) for UQ/Sensitivity Analysis (UQ/SA) on our codes.** As noted above, the major source of uncertainty in many plasma simulations is the forward propagation of errors in input parameters. Ideally, a probability distribution representing the uncertainties in all input variables would be known, and an efficient method would be available to transform these distributions through the code to distributions in the output parameters. A simple and common approach is to try to estimate this level of uncertainty by executing an ensemble of code runs, each with slightly modified inputs. However, for problems requiring extreme-scale computational resources, this sampling-based method can be expensive and limits the range and numbers of parameters that can be tested. For a problem with a large number (high dimension) of input parameters, brute force approaches may not be possible with available computational resources. There is a good deal of ongoing research on sensitivity analysis methodologies aimed at addressing this problem with the goal of improving their efficiency and statistical rigor. The methods can be classified as intrusive — that is, they require some lesser or greater degree of modification of the underlying code — or nonintrusive. Generally, there is a trade-off between intrusiveness, which puts a burden on code developers, and computational intensity. In some cases, the intrusive methods may not be applicable or practical, so the optimal approaches need to be determined on a case-by-case basis.

   The efficiency of random sampling-based methods can be improved through several approaches — deterministic sampling (importance sampling, pseudo-random) or non-sampling methods (e.g., pseudo-random, polynomial chaos, probabilistic collocation, etc.). While providing improvements, all typically suffer from the "curse of dimensionality": that is, they work efficiently for only a small number of input parameters. It is an open research question as to which approach would work best for each of our calculations.

   Rather than sampling via large ensemble of code runs, SA methods have been proposed that can perform the calculation (i.e., derivatives of outputs with respect to inputs) as part of the main simulation. These are more intrusive but may offer the prospect of greater efficiency — especially for higher-dimensional spaces. Additional computational overhead is typically a small multiple of a base case, so the win becomes bigger when the dimension of the input space exceeds a few. These methods are currently the subject of intense research, but in our problem space, there are few applications so far. It is an open question as to which of these methods are applicable to our problems, what their limitations are, and what their costs for implementation will be. The most intrusive methods require some significant rewriting of the underlying code.

For example, by solving an adjoint equation along with the originals; a single adjoint can provide the derivative of a single output with respect to all inputs. However, solvers for the original system of equation may have problems with the adjoint equation. This method has been developed and tested for CFD problems in recent years, and a research team is currently trying it out on a fluid plasma turbulence problem. Its general application needs to be demonstrated. Somewhat less intrusive is automatic differentiation. This method calculates local (linear) dependence of inputs to outputs by a symbolic differentiation through applying the chain rule to each elementary operation in the code. The approach is somewhat analogous to code compilation. The method might be extended to higher-order, nonlinear dependences — but at a cost. The applicability to our codes needs to be demonstrated. One limitation is that the method requires source code for the entire calculation — no opaque libraries or subroutines are allowed. Implementation of these approaches may be particularly challenging in the exascale era as codes may increasingly rely on independently optimized external libraries.

Another possible method is to explore the sensitivities of a model through the use of surrogate models or computations with reduced physics or resolution. These can be used to identify the most important inputs, which can then be studied with the full models or in some cases may themselves produce reasonable estimates of sensitivity — even if their absolute accuracy is unacceptable. (That is, for a quantity of interest Q and input variable x, they may produce an acceptable estimate of $\delta Q/ \delta x$ even if they cannot produce a useful value of Q.) All of the SA methods described above might benefit from "pre-screening" in this manner. An interesting example was shown at the Exascale Requirements Review meeting where the sensitivity of lower hybrid current drive (LHCD) models to assumed density and temperature profiles was studied. A very systematic exploration of this sensitivity was carried out using a combined ray tracing/Fokker-Planck model (see Figure 3-16). This approach may stand on its own, or help direct studies with the full wave/Fokker Planck model targeting more interesting or important parts of the parameter space. In any particular case, this approach may or may not be applicable and so must be studied for each problem. Studies of the sensitivities of the multiscale turbulent transport code XGC1 are being attempted using a similar method. This approach would also be extremely relevant in the context of integrated and whole-device modeling where a variety of physics models at various fidelity levels will be coupled together. In such a context, understanding whether individual reduced or "advanced reduced" models exhibit the same sensitivities as the "first-principles" models they seek to describe will be important in its own right, as well as understanding how the sensitivities of multiple coupled models interact with each other.

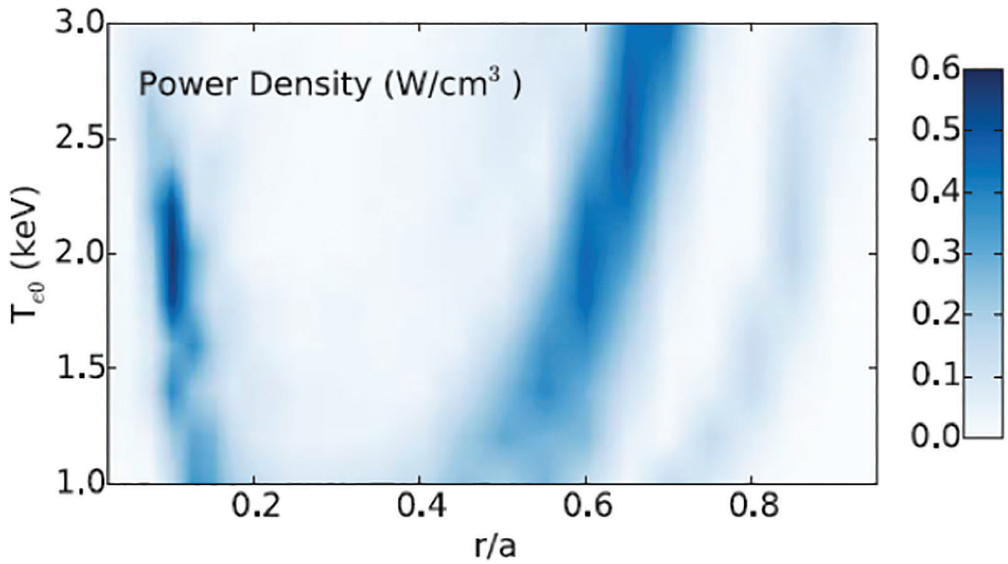


Figure 3-16. The sensitivity of RF power deposition to temperature is explored through a large ensemble of ray tracing calculations. These can serve as a surrogate for a more demanding full wave model.

3. **Develop and adopt tools to enhance data management, including automated capture of provenance and other metadata and improved data browsing of very large and complex data sets.** The data management challenge for VVUQ extends to processes occurring both on and off HPC hardware and will typically involve an extended collaboration that includes modelers and experimentalists. The scope of VVUQ activities includes raw experimental data and its reduction and analysis, preparation of inputs for code runs, post-processing of code outputs, and comparisons between experimental and simulation results including statistical analysis. Meaningful validation requires careful documentation of assumptions and parameters at each step. Overall, a much more systematic approach to metadata and provenance capture is required.

   The large datasets generated by simulations and experiments offer additional challenges. We need tools capable of more effective data browsing and exploration of these data — probably driven by more systematic provision of metadata and automated feature or event detection. The ability to explore 5D and 6D data effectively may depend on tags generated during analysis — akin to the process of geo-tagging. The overall challenges in data management are well covered in detail in the report from the recent Workshop on Integrated Simulation (FES and ASCR 2015). A particular challenge in the context of PRD2 is developing workflows for tracking ensembles of simulation results (particularly first-principles ones) used for UQ and sensitivity analysis.

4. **Develop and deploy new tools that enhance analytics and visualization for computations under emerging computing architectures.** While the new architectures present a challenge to compute, it is already the case that analytics and visualization are lagging. I/O challenges restrict the amount of data that can be stored, preventing the full exploitation of the simulations performed. In the future, the new computing architectures may make this situation worse if the present technologies are used. Post visualization of the huge quantities of data produced, including 5D and 6D arrays, is a daunting challenge. One solution is to perform more of the data analysis and visualization tasks during the computation itself while the data are still in memory (*in situ*) or being transferred between

processors (in-transit). *In-situ* and in-transit methods are particularly well suited to processor-rich, I/O-poor architectures. These methods are the subject of current research but are not yet widely deployed in production. Finding the best approaches for our domain and deploying them will take close collaboration between applied computer scientists and computational physicists. Ideally, toolkits would be made available that would ease the adoption of these approaches. Because full-resolution data are only available during the HPC run, decisions about analysis must be made ahead of time — thus, it will also be a challenge to develop strategies that maximize the productivity of each run. To that end, *in-situ* analysis could also be used to detect interesting events or features in the simulations, triggering particular analyses or higher-resolution I/O around the features for off-line analysis and visualization. For code validation, a particularly important subject of *in-situ* analysis may be the generation of synthetic diagnostic data. The report from the recent Workshop on Integrated Simulation covers the overall challenges for *in-situ* analytics well (FES and ASCR 2015).

3.4.2.3 Cross-Cutting Research Directions

VVUQ is intrinsically cross cutting, so that all of the research challenges listed above cut across topical areas within the fusion/plasma domain and into other scientific domains. Similarly, as captured in the research priorities, the need for development and deployment of tools for VVUQ exists across much of the Office of Science domain and beyond. Data management, analytics, and visualization needs are common — including the focus on metadata and provenance capture and *in-situ* and in-transit methods for analytics and visualization. Note, for example, the similarity of requirements as summarized for the fusion community (FES and ASCR 2015) and across the wider Office of Science domain (DOE-SC 2016).

3.4.2.4 Computing Needs and Requirements

VVUQ is a strong driver for computing needs. The sensitivity analysis described above is an essential activity — simulation results without an assessment of sensitivity to input parameters have severely limited utility. Quantities of interest typically depend on many parameters, and spanning that space can be computationally expensive — even using some of the techniques described above. This finding suggests that UQ drives 10–100 simulations for every “base case” for which a researcher needs to apply UQ. Extrapolating from the most demanding of fusion simulations, the computation of turbulent-driven transport, implies a requirement of more than 200 million core-hours for each base case of interest using today’s codes. Note that each of these is only for a snapshot in time at a particular location in the plasma. The aggregate impact is hard to estimate but will certainly be significant. One might expect that the amount of computational time dedicated to VVUQ will be at least as large as other activities combined.

It is also important to note that simulation cases run for validation will often be among the most demanding. That is, they typically require using the most realistic physics assumptions. As an example, a turbulent transport calculation might require including both electron and ion gyrokinetic dynamics with the real mass ratio, finite β (plasma kinetic energy/magnetic field energy) effects, and operation near critical gradients (i.e., at marginal stability). Even if agreement is achieved with simpler assumptions, it is critical to understand whether this agreement is fortuitous or whether it is maintained in the full physics calculation.

This page is intentionally left blank.

4 PATH FORWARD

For researchers to move forward in addressing the scientific challenges documented in this review, an evolving computing ecosystem must support them. This computing ecosystem includes computational and data capabilities and capacity; scientific applications and software; and the infrastructure for data transfer, sharing, access, and analysis — each of which must undergo further investment and development on the path to exascale. New advances required for large-scale verification and validation must be realized. The coupling of an exascale ecosystem, along with a convergence of theoretical, mathematical, computational, and experimental capabilities, will bring many opportunities for new scientific breakthroughs at an unprecedented scale.

Collaboration between FES and ASCR scientists and facilities staff will help ensure development and deployment of an effective, realistic computing ecosystem that enables revolutionary discoveries in areas described in this report. The recent FES and ASCR (2015) *Report of the Workshop on Integrated Simulations for Magnetic Fusion Energy Sciences* highlights collaboration on similar topics. The computing ecosystem requirements resulting from this review will form the basis to direct future investments of time and resources. These requirements fall into broad categories: methods development; computational environment; data and workflow; and communication and community involvement.

4.1 Methods Development

The advancing complexity of computer hardware requires FES researchers to have more scalable, performant algorithms and applications that are capable of efficient execution on future computing architectures fielded by ASCR facilities. Meeting participants discussed those computing ecosystem aspects that will accelerate or impede their progress in the next 5–10 years. Participants named application codes and verification and validation techniques, as well as models and algorithms, as key factors requiring significant methods development activity. A representative list of the methods development topics discussed by the review participants is as follows (see Section 3 for a more detailed overview of the methods development topics presented by the review participants):

- Threaded and accelerated versions of the direct and sparse matrix algebra in combination with semi-implicit time advance methods (3.1.1).
- Ameliorating the memory and logic intensity of particle/mesh mapping (3.1.1).
- Efficient algorithms for Maxwell's equations (3.1.1).
- Improved capabilities for verification, validation, and uncertainty quantification (3.1.2, 3.1.3, 3.1.4).
  - New mechanisms for managing, discovery, method exploration (brute force, adjoint-like, automatic differentiation, etc.).
- Improved multiple-timescale, multiple-physics coupling (3.1.2).
- More scalable and performance linear algebra solvers (3.1.2).
- Domain-specific solvers (3.1.2).
- Improved scaling and resolution in RF algorithms (3.1.3).
- Hybrid algorithms capable of kinetic and magneto-hydrodynamic features (3.3.2).
- Improved particle solvers (3.3.2).
- High-dimensional PDE solvers (3.3.2).
- Exascale-ready adaptive mesh algorithms (3.3.2).

In almost all discussions for new models and methods, integrating applied mathematicians and computer scientists with computational scientists is crucial for success. Programs like SciDAC are key examples of success.

A close dialogue between FES and ASCR researchers and facilities staff will streamline and promote research and development through the exchange of information about computing ecosystem roadmaps and application requirements and the availability of systems for simulation and testing.

4.2 Computational Environment

Requirements for the access, scheduling, and software ecosystem identify an evolving use-model. The "traditional" HPC model, defined as running a large simulation that generates data that are then post processed, is no longer the only primary use-model for many FES projects. Emerging demands, such as for complex workflows and near-real-time computing, are changing the landscape.

New requirements for the computing ecosystem include the following:

- Support a user-friendly development environment, with uniform environments among DOE HPC centers supporting portable, high performance across systems with improved and new runtime systems that mask HPC complexity from application programmers, and training aimed at all levels of HPC developers, including nontraditional HPC users (3.1.1, 3.1.3, etc.).
- Reconsider batch queuing methods and priorities. Queue wait times can limit the ability to push through mid-scale/capacity computing science needs (3.1.1, 3.2.1).
- Promote accessibility (3.1.1).
- Keep balance between storage, memory, and FLOPS (put the problem in memory) (3.1.2).
- Help support better software engineering efforts (programming environments, models, software stacks, etc.) (3.1.3).
- Identify/develop programming models and languages that can increase productivity and still provide performance (3.1.3). There will be a need to abstract and encapsulate some of the complexity that will be present in exascale environments. A key to success in the past has been — and will again be — to insulate application developers as much as practical from the inevitable changes in underlying implementations.
- Reduce the need for porting and maintaining code, perhaps through containers or software-as-service models (3.1.3).
- Sustain the capabilities — tools, libraries, etc. — that are developed in the process of moving to exascale computing; explore models for doing this by other agencies (e.g., National Nuclear Security Administration).

4.3 Data

The scale of data generated from FES simulations and the requirements needed for verification and validation have created an opportunity and a challenge. ASCR and FES facilities must create more data-centric environments with highly effective data analytics tools for their users. Development of such environments and tools will require expertise from domain scientists, data scientists, and applied mathematicians. Continued collaboration will be required to assess proper deployment of the environments as computing resources evolve.

Requirements related to data generation, storage, transport, curation, and exploration include the following:

- Because simulations for particle codes already generate more than ~100 TB of physics data, expect at least an order-of-magnitude growth (3.1.1).
- Develop advanced workflow managers (3.1.3).
- Support access to simulation and experimental databases at runtime for input, analysis, and verification (3.1.4).
- Support *in-situ* analysis (3.1.4).
- Improve capture and organization of metadata. including provenance.
- Develop community standards and federated databases.
- Develop data curation methods.

4.4 Communication and Community Involvement

To foster development of the requisite exascale-level skills and to disseminate this learning widely throughout the community, DOE (with the ASCR facilities) must seek to create or make use of existing initiatives that promote the following:

- Proposal/award processes to support the wider array of requirements, including flexible allocations mechanisms and metrics based on science goals.
- Expanded involvement of ASCR applied math and computer science (3.1.3).
- Deeper involvement in open source applications and libraries.
- Workforce development (education and training).

These activities are ongoing today in multiple institutions; however, efforts to connect them to the larger science community have been attempted on an "ad hoc" basis to date. ASCR facilities can explore new or improved communication channels and activities. In addition, experience has shown some of the best impact from strong collaborations. The previously identified structured collaborative efforts could focus more attention on this important mechanism for community involvement.

4.5 Conclusions

Requirements that are key to an evolving computing ecosystem have been identified in areas of methods development, computational environment, data, and communication and community involvement. These areas are a collaborative research opportunity across much of the computational ecosystem. Structured collaborative efforts between FES and ASCR are a path to address the growing complexity of the science and the computational resources.

This page is intentionally left blank.

5 REFERENCES

Batchelor, D., G. Abla, E. D'Azevedo *et al.*, 2009, "Advances in Simulation of Wave Interactions with Extended MHD Phenomena," *Journal of Physics: Conference Series* **180**, 012054.

Betti, R., C. Zhou, K. Anderson, L. Perkins, W. Theobald, and A. Solodov, 2007, "Shock Ignition of Thermonuclear Fuel with High Areal Density," *Physical Review Letters* 98(15), http://doi.org/10.1103/PhysRevLett.98.155001, accessed June 29, 2016.

Birdsall, C.K., and A.B. Langdon, 1991, *Plasma Physics via Computer Simulation*, New York: Adam Hilger.

Bulanov, S.S., et al., 2015, "High Intensity Laser Matter Interaction," DOE Frontier of Plasma Science Workshops, http://www.orau.gov/plasmawkshps2015/whitepapers/general-Bulanov_Stepan.pdf., accessed June 29, 2016.

Chen, H., F. Fiuza, et al., 2015, "Scaling the Yield of Laser-Driven Electron-Positron Jets to Laboratory Astrophysical Applications," *Phys. Rev. Lett.* 114, 215001.

Chowdhury, J., W. Wan, Y. Chen, S.E. Parker, R.J. Groebner, C. Holland, and N.T. Howard, 2014, "Study of the L-mode Tokamak Plasma 'Shortfall' with Local and Global Nonlinear Gyrokinetic δf Particle-in-Cell Simulation," *Physics of Plasmas* **21**, 112503; http://dx.doi.org/10.1063/1.4901031, accessed May 13, 2016.

DOE (U.S. Department of Energy), 2011, *Fusion Simulation Program Execution Plan*, Office of Science, September 30, http://w3.pppl.gov/fsp/FSP_Summary_FILES/FSP_Program_Execution_Plan.pdf, accessed April 12, 2016.

DOE-ASCAC (U.S. Department of Energy-Office of Advanced Scientific Computing Advisory Committee), 2014, *Top Ten ExaScale Research Challenges: Subcommittee Report*, R. Lucas (chair), February 10, http://science.energy.gov/~/media/ascr/ascac/pdf/meetings/20140210/Top10reportFEB14.pdf, accessed April 15, 2016.

DOE-SC and NNSA (U.S. Department of Energy-Office of Science and National Nuclear Security Administration), 2009, "Basic Research Needs for High-Energy-Density Laboratory Physics," Report of the Workshop on High Energy Density, Laboratory Physics Research Needs, November 15–18.

DOE-SC, 2015, *The Office of Science Fusion Energy Science Program: A Ten-Year Perspective (2015–2025)*, Office of Fusion Energy Sciences, December, http://science.energy.gov/~/media/fes/pdf/program-documents/FES_A_Ten-Year_Perspective_2015-2025.pdf, accessed October 14, 2016.

DOE-SC, 2016, *Report of the DOE Workshop on Management, Analysis, and Visualization of Experimental and Observational Data: The Convergence of Data and Computing*, Office of Advanced Scientific Computing Research, LBNL-1005155, Bethesda, Maryland, September 29–October 1, 2015, http://science.energy.gov/~/media/ascr/pdf/programdocuments/docs/ascr-eod-workshop-2015-report_160524.pdf, accessed June 21, 2016.

FES (Office of Fusion Energy Sciences), 2015a, *Fusion Energy Sciences Workshop on Plasma Materials Interactions: Report on Science Challenges and Research Opportunities in Plasma Materials Interactions*, U.S. Department of Energy-Office of Science, May 4–7, http://science.energy.gov/~/media/fes/pdf/workshop-reports/2016/PMI_fullreport_21Aug2015.pdf, accessed June 23, 2016.

FES, 2015b, *Fusion Energy Sciences Workshop on Transients in Tokamak Plasmas: Report on Scientific Challenges and Research Opportunities in Transient Research*, U.S. Department of Energy-Office of Science, June 8–11, http://science.energy.gov/~/media/fes/pdf/program-news/Transients_Report.pdf, accessed June 23, 2016.

FES, 2015c, FES, 2015c, *Frontiers of Plasma Science Workshops*, U.S. Department of Energy, Office of Science, https://www.orau.gov/plasmawkshps2015/, accessed October 13, 2016.

FES, 2016, *Fusion Energy Sciences (FES)*, U.S. Department of Energy, Office of Science, http://science.energy.gov/fes/, accessed June 23, 2016.

FES and ASCR (Office of Advanced Scientific Computing Research), 2015, *Report of the Workshop on Integrated Simulations for Magnetic Fusion Energy Sciences*, P.T. Bonoli (chair) and L.C. McInnes (co-chair), U.S. Department of Energy Office of Science, Rockville, Maryland, June 2–4, http://science.energy.gov/~/media/fes/pdf/workshop-reports/2016/ISFusionWorkshopReport_11-12-2015.pdf, accessed April 15, 2016.

FESAC (Fusion Energy Sciences Advisory Committee), 2002, *Panel on Integrated Simulation and Optimization of the Magnetic Fusion Systems*, prepared by the FESAC Integrated Simulation & Optimization of Fusion Systems Subcommittee, U.S. Department of Energy Office of Science, DOE/SC-0073, November, http://science.energy.gov/~/media/fes/fesac/pdf/2002/Fsp_report_dec_9_2002.pdf, accessed April 12, 2016.

FESAC, 2014, *Report on Strategic Planning: Priorities Assessment and Budget Scenarios*, M. Koepke (chair), December, https://www.burningplasma.org/resources/ref/fspp/FINAL_FESAC_123114-1.pdf, accessed April 15, 2016.

Fiksel, G., et al., 2014, "Magnetic Reconnection between Colliding Magnetized Laser-Produced Plasma Plumes," *Phys. Rev. Lett.* 113, 105003.

Fiuza, F., et al., 2012, "Laser-Driven Shock Acceleration of Monoenergetic Ion Beams," *Phys. Rev. Lett.* 109, 215001.

Geddes, C.G.R., et al., 2015, "Ultrafast Photon and Particle Probes for and via Plasma Science, white paper for FES Exascale Requirements Review.

Greenwald, M., D. Hillis, A. Hubbard, J. Hughes, S. Kaye, R. Maingi (coordinator), G. McKee, D. Thomas, M. Van Zeeland, and M. Walker, 2015, *Modes of Participation in ITER*, report of the U.S. Burning Plasma Organization Working Group for ITER Experimental Operation, U.S. Team Formation and Participation, April 24, https://burningplasma.org/resources/PDFS/taskgroups/BPO_ITER_Participation_FullReport_Final_23June2015.pdf, accessed April 12, 2016.

Hawryluk, R., 1980, "An Empirical Approach to Tokamak Transport," in *Physics of Plasmas Close to Thermonuclear Conditions*, B. Coppi, G.G. Leotta, D. Pfirsch, R. Pozzoli, and E. Sindoni (eds.), published for the Commission of the European Communities by Pergamon Press, Oxford, UK, 1981, Volume 1, p.19.

Helland, B., 2016, *FES Requirements Review*, presentation at the Fusion Energy Sciences Exascale Requirements Review Meeting, U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research, Gaithersburg, Maryland, January 27.

Huntington, C. M., F. Fiuza, S. Ross, et al., 2015, "Observation of Magnetic Field Generation via the Weibel Instability in Interpenetrating Plasma Flows," *Nature Physics* 11, 173.

Jaeger, E.F., L.A. Berry, and E.F. D'Azevedo, et al., 2008, "Simulation of High-power Electromagnetic Wave Heating in the ITER Burning Plasma," *Physics of Plasmas* **15**, 072513.

Jenkins, T.G., and D.N. Smithe, 2014, “3D Modeling of RF Antennas, Sheaths, and Slow Waves in Time Domain,” U.S. Burning Plasma Organization eNews, November 30 (Issue 90), https://www.burningplasma.org/newsandevents/?article=enews&issue=120114#highlight, accessed April 15, 2016.

Jenkins, T.G., S.E. Kruger, J.R. King, and E.D. Held, 2016, “Computational Needs: Coupling Extended MHD Simulations and RF Wave Codes,” white paper submitted to the Fusion Energy Sciences Exascale Requirements Review Workshop.

Jenkins, T.G., T.M. Austin, D.N. Smithe, J. Loverich, and A.H. Hakim, 2013, “Time-domain Simulation of Nonlinear Radiofrequency Phenomena,” *Physics of Plasmas* **20**, 012116.

Jenko, F., D. Told, T. Görler, J. Citrin, A. Bañón Navarro, C. Bourdelle, S. Brunner, G. Conway, T. Dannert, H. Doerk, D. R. Hatch, J. W. Haverkort, J. Hobirk, G. M. D. Hogeweij, P. Mantica, M. J. Pueschel, O. Sauter, L. Villard, E. Wolfrum and the ASDEX Upgrade Team, 2013, “Global and Local Gyrokinetic Simulations of High-Performance Discharges in View of ITER,” *Nuclear Fusion* **53**, 073003, http://www.physics.ucla.edu/~jenko/PAPERS/Jenko-NF13.pdf, accessed May 13, 2016.

Kemp, A., and F. Fiuza, et al., 2014, Laser–plasma Interactions for Fast Ignition, *Nuclear Fusion* **54** 054002.

Kushner, M.J., et al., 2014, “A Low Temperature Plasma Science and Engineering Program: Discovery Science for Societal Benefit,” Whitepaper to the NSF Engineering Directorate, September 29.

Kushner, Mark J., 2009, “Hybrid Modelling of Low Temperature Plasmas for Fundamental Investigations and Equipment Design,” *J. Phys. D: Appl. Phys.* **42**, 194013, October. ISSN 0022-3727, 1361-6463.

Leemans, W.P., et al., 2015, “BELLA-i PW Laser Facility for Intense Laser-matter Interactions,” DOE Frontier of Plasma Science Workshops. Available at http://www.orau.gov/plasmawkshps2015/whitepapers/general-Leemans_Wim.pdf, accessed May 13, 2016.

Macchi, A., and M. Borghesi, 2013, “Ion Acceleration by Superintense Laser-Plasma Interaction,” *Rev. Mod. Phys.* 85.

Mynick, H.E., N. Pomphrey, and P. Xanthopoulos, 2010, “Optimizing Stellarators for Turbulent Transport,” *Phys. Rev. Lett.* **105**, 095004, http://journals.aps.org/prl/pdf/10.1103/PhysRevLett.105.095004, accessed May 13, 2016.

National Research Council, 2003, Frontiers in High Energy Density Physics: *The X-Games of Contemporary Science*.

OLCF (Oak Ridge Leadership Computing Facility), 2014, “The Bleeding ‘Edge’ of Fusion Research,” Oak Ridge National Laboratory, February 14, https://www.olcf.ornl.gov/2014/02/14/the-bleeding-edge-of-fusion-research, accessed May 13, 2016.

Plasma 2010 Committee, 2010, “Plasma Science: Advancing Knowledge in the National Interest,” Plasma Science Committee, National Research Council.

Sentoku, Y., I. Paraschiv, R. Royle, and R.C. Mancini, 2014, “Kinetic Effects and Nonlinear Heating in Intense X-ray-laser-produced Carbon Plasmas,” *Phys. Rev. E* 90, 051102(R).

Steinke, S., et al., 2015, “Laser-matter Interaction and Ion Acceleration with PW Laser Pulses,” DOE Frontier of Plasma Science Workshops, http://www.orau.gov/plasmawkshps2015/whitepapers/general-Steinke_Sven.pdf., accessed June 29, 2016.

Stork, D., 2009, “DEMO and the Route to Fusion Power,” Third Karlsruhe International School on Fusion Energy, September, http://fire.pppl.gov/eu_demo_Stork_FZK%20.pdf, accessed April 15, 2016.

Takizuka, T., and H. Abe, 1977, “A Binary Collision Model for Plasma Simulations with a Particle Code,” *J. Comp Phys* 20, 205.

Totorica S., T. Abel, and F. Fiuza, 2016, “Nonthermal Electron Energization from Magnetic Reconnection in Laser-Driven Plasmas,” *Phys. Rev. Lett.* **116**, 095003.

Tsujii, N., M. Porkolab, and P.T. Bonoli, et al., 2015, “Validation of Full-wave Simulations for Mode Conversion of Waves in the Ion Cyclotron Range of Frequencies with Phase Contrast Imaging in Alcator C-Mod,” *Physics of Plasmas* **22**, 082502, http://scitation.aip.org/content/aip/journal/pop/22/8/10.1063/1.4927912, accessed April 15, 2016.

Verboncoeur, J.P., M.V. Alves, V. Vahedi, and C.K. Birdsall, 1993, “Simultaneous Potential and Circuit Solution for 1D Bounded Plasma Particle Simulation Codes,” *J. Comput. Phys.* **104**, 321–328, February. ISSN 0021-9991.

Wang, B., G.H. Miller, and P. Colella, 2011, “A Particle-in-cell Method with Adaptive Phase-space Remapping for Kinetic Plasmas,” *SIAM Journal on Scientific Computing* 33(6), 3509–3537, http://doi.org/10.1137/100811805, accessed May 18, 2016.

Weinan, E., et al., 2007, “Heterogeneous Multiscale Methods: A Review,” http://www.math.nus.edu.sg/~matrw/multiscale/publications/HMM_review.pdf, accessed April 15, 2016.

White, A.E., N.T. Howard, M. Greenwald, M.L. Reinke, C. Sung, S. Baek, M. Barnes, J. Candy, A. Dominguez, D. Ernst, C. Gao, A.E. Hubbard, J.W. Hughes, Y. Lin, D. Mikkelsen, F. Parra, M. Porkolab, J.E. Rice, J. Walk, S.J. Wukitch, and Alcator C-Mod Team, 2013, “Multi-channel Transport Experiments at Alcator C-Mod and Comparison with Gyrokinetic Simulations,” *Physics of Plasmas*, 20, 056106, DOI:http://dx.doi.org/10.1063/1.4803089.

Wilson, J.R., and P.T. Bonoli, 2014, “Progress on Ion Cyclotron Range of Frequencies Heating Physics and Technology in Support of the International Tokamak Experimental Reactor,” *Physics of Plasmas* **22**, 021801.

Xanthopoulos, P., H.E. Mynick, P. Helander, Y. Turkin, G.G. Plunk, F. Jenko, T. Görler, D. Told, T. Bird, and J.H.E. Proll, 2014, “Controlling Turbulence in Present and Future Stellarators,” *Phys. Rev. Lett.* **113**, 155001, http://journals.aps.org/prl/pdf/10.1103/PhysRevLett. 113.155001, accessed May 13, 2016.

Yan, R.,C. Ren, J. Li, A.V. Maximov, W.B. Mori, Z.M. Sheng, and F.S. Tsung, 2012, “Generating Energetic Electrons through Staged Acceleration in the Two-plasmon-decay Instability in Inertial Confinement Fusion,” *Physical Review Letters*, 108(17), 175002, http://doi.org/10.1103/PhysRevLett.108.175002., accessed June 29, 2016.

Yan, R., J. Li, and C. Ren, 2014, “Intermittent Laser-plasma Interactions and Hot Electron Generation in Shock Ignition,” *Physics of Plasmas* 21(6), 062705, http://doi.org/10.1063/1.4882682., accessed June 29, 2016.

Yin, L., B.J. Albright, H.A. Rose, K.J. Bowers, B. Bergen, R.K. Kirkwood, et al., 2012, “Trapping Induced Nonlinear Behavior of Backward Stimulated Raman Scattering in Multi-speckled Laser Beams,” *Physics of Plasmas* 19(5), 056304.

6 ACRONYMS AND ABBREVIATIONS

1D, 2D, 3D,...	one-, two-, three-dimension, etc.
ALCC	ASCR Leadership Computing Challenge
Argonne	Argonne National Laboratory
ASCR	(Office of) Advanced Scientific Computing Research (DOE)
BCA	binary collision approximation
CAAR	Center for Accelerated Application Readiness Proposal
CPU	central processing unit
CQ	current quench
DC	direct current
DOE	U.S. Department of Energy
ExB flow	Flow of plasma in a magnetic field when there is an electric field perpendicular to the magnetic field vector
ECCD	electron-cyclotron current drive
ECH	electron-cyclotron heating
EDA	enhanced D-alpha
EHO	edge harmonic oscillation
ELM	edge-localized mode
EP	energetic particle
FES	Fusion Energy Sciences (DOE)
FLOPS	floating-point operations per second
GPU	graphics processing unit
HEDLP	high-energy-density laboratory plasma
HPC	high-performance computing
I/O	input/output
ICF	inertial confinement fusion
ICH	ion-cyclotron heating
ICRF	ion cyclotron range of frequencies

ICRH	ion cyclotron resonant heating
IDS	ITER Data Structure
IMAS	Integrated Modeling and Analysis Suite
INCITE	Innovative and Novel Computational Impact on Theory and Experiment
ITER	Latin for “the way”
ITPA	International Tokamak Physics Activity
JET	Joint European Torus
KMC	kinetic Monte Carlo
LH	lower hybrid
LHRF	lower hybrid range of frequencies
LPI	laser-plasma interaction
LTP	low-temperature plasma
MD	molecular dynamics
MFE	magnetic fusion energy
MHD	magnetohydrodynamic(s)
MMS	magnetospheric multiscale, method of manufactured solutions
NERSC	National Energy Research Scientific Computing Center
NSTX	National Spherical Torus Experiment
NSTX-U	National Spherical Torus Experiment Upgrade
NTM	neoclassical tearing mode
OKMC	object KMC
ORNL	Oak Ridge National Laboratory
PDE	partial differential equation
PFC	plasma-facing component
PFM	plasma-facing material
PIC	particle-in-cell
PIC-MCC	particle-in-cell Monte Carlo collisions
PMI	plasma-material interaction
PRD	priority research direction
QC	quasi-coherent

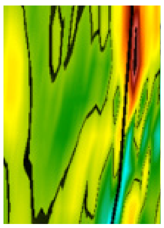
RAM	random access memory
RE	runaway electron
RF	radio frequency
RMP	resonant magnetic perturbation
SA	sensitivity analysis
SBS	stimulated Brillouin scattering
SC	Office of Science (DOE)
SciDAC	Scientific Discovery through Advanced Computing (DOE)
SOL	scrape-off layer
SRS	stimulated Raman scattering
TPD	two-plasmon decay
TQ	thermal quench
UQ	uncertainty quantification
VDE	vertical displacement event
VVUQ	Verification, validation, and uncertainty quantification
WDM	whole device modeling

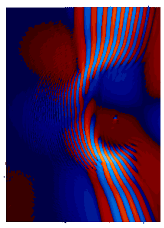
Units of Measure

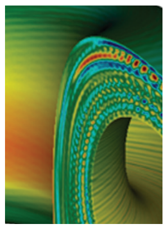
atm	atmosphere
B	billion
cm	centimeter(s)
eV	electron volt(s)
fs	femtosecond(s)
GB	gigabyte(s)
GHz	gigahertz
hrs	hours

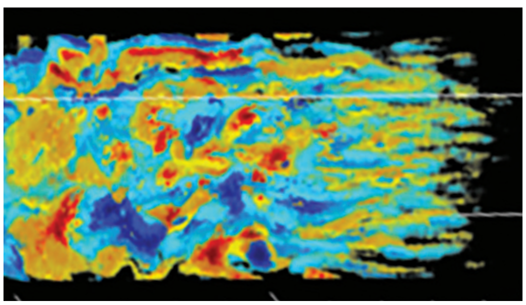
m	meter(s)
M	million
MHz	megahertz
mm	millimeter(s)
ms	millisecond(s)
mTorr	milli Torr
nm	nanometer(s)
ns	nanoscale
PF	petaflop
T	Tesla
TB	terabyte(s)
µm	micron(s)
W	watt(s)

FES









FUSION ENERGY SCIENCES

APPENDICES: MEETING MATERIALS

An Office of Science review sponsored jointly by
Advanced Scientific Computing Research and Fusion Energy Sciences

This page is intentionally left blank.

APPENDIX A: FUSION ENERGY SCIENCES ORGANIZING COMMITTEE AND MEETING PARTICIPANTS

A.1 FES ORGANIZING COMMITTEE

C.S. Chang[a] (cschang@pppl.gov)

Martin Greenwald (g@psfc.mit.edu)

Katherine Riley (riley@alcf.anl.gov)

Richard Coffey (richardc@alcf.anl.gov)

Richard Gerber (ragerber@lbl.gov)

Jack Wells (wellsjc@ornl.gov)

A.2 FES SPONSORS AND REPRESENTATIVES

John Mandrekas (john.mandrekas@science.doe.gov)

Carolyn Lauzon (carolyn.lauzon@science.doe.gov)

A.3 FES MEETING PARTICIPANTS

Katie Antypas	Lawrence Berkeley National Laboratory, National Energy Research Scientific Computing Center
Amitava Bhattacharjee	Princeton University, Princeton Plasma Physics Laboratory
Paul Bonoli[a]	Massachusetts Institute of Technology
Iain Boyd[a]	University of Michigan
David Brown	Lawrence Berkeley National Laboratory
Jeff Candy	General Atomics (GA)
John Canik	Oak Ridge National Laboratory
Rich Carlson	U.S. Department of Energy, Office of Advanced Scientific Computing Research
Louis Chacon	Los Alamos National Laboratory
Choong-Seock Chang[a]	Princeton Plasma Physics Laboratory
Yang Chen[a]	University of Colorado at Boulder
Daniel Clark	U.S. Department of Energy, Office of Fusion Energy Sciences
Richard Coffey	Argonne National Laboratory, Argonne Leadership Computing Facility
Scott Collis	Sandia National Laboratories
Davide Curreli[a]	University of Illinois at Urbana-Champaign
Eli Dart	Lawrence Berkeley National Laboratory, ESnet

[a]Authored or co-authored a White Paper (Appendix C).

[b]Authored or co-authored a Case Study (Appendix D).

Eduardo D'Azevedo	Oak Ridge National Laboratory, Oak Ridge Leadership Computing Facility
Lori Diachin	Lawrence Livermore National Laboratory
Sudip Dosanjh	Lawrence Berkeley National Laboratory, National Energy Research Scientific Computing Center
Stephane Ethier[a]	Princeton Plasma Physics Laboratory
Frederico Fiuza	Stanford University
Guoyong Fu[a]	Princeton Plasma Physics Laboratory
Richard Gerber	Lawrence Berkeley National Laboratory, National Energy Research Scientific Computing Center
Lisa Gerhardt	Lawrence Berkeley National Laboratory, National Energy Research Scientific Computing Center
Kai Germaschewski	University of New Hampshire
David Goodwin	U.S. Department of Energy, Office of Advanced Scientific Computing Research
David Green[a]	Oak Ridge National Laboratory
Martin Greenwald	Massachusetts Institute of Technology
James Hack	Oak Ridge National Laboratory
Greg Hammett	Princeton Plasma Physics Laboratory
David Hatch[a]	University of Texas at Austin
Barb Helland	U.S. Department of Energy, Office of Advanced Scientific Computing Research
Jeff Hittinger	Lawrence Livermore National Laboratory
Adolfy Hoisie	Pacific Northwest National Laboratory
Chris Holland	University of California, San Diego
Nathan Howard[a]	Massachusetts Institute of Technology
Gregory Howes	University of Iowa
Steve Jardin[b]	Princeton Plasma Physics Laboratory
Frank Jenko[a]	University of California, Los Angeles
Wayne Joubert	Oak Ridge National Laboratory, Oak Ridge Leadership Computing Facility
Igor Kaganovich	Princeton Plasma Physics Laboratory
Andreas Kemp	Lawrence Livermore National Laboratory
Jacob King[a]	Tech-X Corporation
Scott Klasky	Oak Ridge National Laboratory
Alice Koniges	Lawrence Berkeley National Laboratory, National Energy Research Scientific Computing Center

[a]Authored or co-authored a White Paper (Appendix C).
[b]Authored or co-authored a Case Study (Appendix D).

Tuomas Koskela	Lawrence Berkeley National Laboratory, National Energy Research Scientific Computing Center
Doug Kothe	Oak Ridge National Laboratory
Predrag Krstic[a]	Stony Brook University
Seung-Hoe Ku	Princeton Plasma Physics Laboratory
Rick Kurtz	Pacific Northwest National Laboratory
Carolyn Lauzon	U.S. Department of Energy, Office of Advanced Scientific Computing Research
Randall Laviolette	U.S. Department of Energy, Office of Advanced Scientific Computing Research
Steven Lee	U.S. Department of Energy, Office of Advanced Scientific Computing Research
Stephen Lee	Los Alamos National Laboratory
Zhihong Lin[a]	University of California, Irvine
Vyacheslav Lukin	National Science Foundation
Barney MacCabe	Oak Ridge National Laboratory
John Mandrekas	U.S. Department of Energy, Office of Fusion Energy Sciences
Andrea Manning	Argonne National Laboratory
Jaime Marian	University of California at Los Angeles
Paul Messina	Argonne National Laboratory, Argonne Leadership Computing Facility
Harry Mynick	Princeton Plasma Physics Laboratory
Thomas Ndousse-Fetter	U.S. Department of Energy, Office of Advanced Scientific Computing Research
Lucy Nowell	U.S. Department of Energy, Office of Advanced Scientific Computing Research
Scott Parker[a]	University of Colorado at Boulder
Robinson Pino	U.S. Department of Energy, Office of Advanced Scientific Computing Research
Nirmol Podder	U.S. Department of Energy, Office of Fusion Energy Sciences
Francesca Poli[a]	Princeton Plasma Physics Laboratory
Dave Pugmire	Oak Ridge National Laboratory, Oak Ridge Leadership Computing Facility
Chuang Ren	University of Rochester
Katherine Riley	Argonne National Laboratory, Argonne Leadership Computing Facility
Sonia Sachs	U.S. Department of Energy, Office of Advanced Scientific Computing Research

[a]Authored or co-authored a White Paper (Appendix C).
[b]Authored or co-authored a Case Study (Appendix D).

Yasuhiko Sentoku	University of Nevada-Reno
Marc Snir	Argonne National Laboratory, Argonne Leadership Computing Facility
Phil Snyder	General Atomics (GA)
Donald Spong	Oak Ridge National Laboratory
David Sprouster	Brookhaven National Laboratory
Bert Still	Lawrence Livermore National Laboratory
Tjerk Straatsma	Oak Ridge National Laboratory, Oak Ridge Leadership Computing Facility
Rangan Sukumar	Oak Ridge National Laboratory, Oak Ridge Leadership Computing Facility
Edmund Synakowski	U.S. Department of Energy, Office of Fusion Energy Sciences
Xianzhu Tang	Los Alamos National Laboratory
Rollin Thomas	Lawrence Berkeley National Laboratory, National Energy Research Scientific Computing Center
Miles Turner[a]	Dublin City University
Maxim Umansky[a]	Lawrence Livermore National Laboratory
Jean-Luc Vay[b]	Lawrence Berkeley National Laboratory
John Verboncoeur	Michigan State University
Art Voter[a]	Los Alamos National Laboratory
Weixing Wang[a]	Princeton Plasma Physics Laboratory
Jack Wells	Oak Ridge National Laboratory
Julia White	Oak Ridge National Laboratory
Timothy Williams	Argonne National Laboratory, Argonne Leadership Computing Facility
Brian Wirth[a]	University of Tennessee at Knoxville
Patrick Worley	Oak Ridge National Laboratory
John Wright[a]	Massachusetts Institute of Technology
Mike Zarnstorff	Princeton Plasma Physics Laboratory

[a]Authored or co-authored a White Paper (Appendix C).
[b]Authored or co-authored a Case Study (Appendix D).

APPENDIX B: FUSION ENERGY SCIENCES MEETING AGENDA

TUESDAY, JANUARY 26

Evening	**Chairs and Leads Pre-Meeting**

WEDNESDAY, JANUARY 27

7:30	**Registration, Refreshments**
8:30	**Welcome & Introductions**
8:40	**Genesis of this Meeting** Barb Helland, U.S. Department of Energy, Office of Advanced Scientific Computing Research (ASCR)
9:20	**View from Fusion Energy Sciences** John Mandrekas, Team Lead for Theory & Simulation, U.S. Department of Energy, Office of Fusion Energy Sciences (FES)
9:50	**Exascale Review Status Updates** Chairs and Session Leaders
10:30	**Break**
10:45	**ASCR Computing Facilities Presentation**
12:10	**Working Lunch** ASCR Computing Facilities Presentation Charge to Working Groups
1:00	**Breakout Sessions:** **Magnetic Fusion Energy Sciences** ▪ Turbulence & Transport ▪ MHD & Energetic Particles ▪ RF Heating & Current Drive ▪ Whole Device Modeling (cross-cutting) ▪ Verification and Validation (cross-cutting) **Materials Science** ▪ Plasma-Surface Interactions ▪ Structural Materials **Discovery Plasma Science** ▪ General Plasma Science ▪ High-Energy-Density Laboratory Plasmas ▪ Low-Temperature Plasmas
4:00	**Break**
4:15	**Q&A Session with the FES and ASCR Associate Directors**
5:30	**Preliminary Feedback: Breakout Leads, 15 minutes each**
6:30	**Dinner on your own**

THURSDAY, JANUARY 28

8:00	**Check-in, Refreshments**
8:30	**Plenary Talk**
9:10	**Summary by Chairs, Outline of Report**
9:45	**Break**
10:00	**Breakouts, Discuss and Start Outlining Sections**
12:10	**Working Lunch**
1:00	**Breakout Sessions: Developing Outlines**
4:00	**Break**
4:20	**Reports on Wednesday Breakouts, Breakout Leads, 15 minutes each**
5:20	**Summary and Thanks from Chairs**
	End of the Meeting for Most Participants

FRIDAY, JANUARY 29

All Day:	**Co-chairs, Leads, Writers meet to continue working on report**

APPENDIX C: FUSION ENERGY SCIENCES WHITE PAPERS

The following white papers were submitted by the authors listed below in advance of the Exascale Requirements Review to guide both the agenda and meeting discussions.

C.1 White Papers Addressing Fusion Energy Science

C.1.1 Turbulence and Transport

C.1.2 Energetic Particles and MHD

C.1.3 RF Heating and Current Drive

C.1.4 Whole Device Modeling

C.2 White Papers Addressing Plasma Surface Interactions and Structural Materials

C.2.1 Plasma Surface Interactions

C-72 Brian Wirth, University of Tennessee, and David Bernholdt, Oak Ridge National Laboratory
Fusion Plasma Material Interactions

C.2.2 Structural Materials

C-74 Richard Kurtz, Giridhar Nandipati, and Wahyu Setyawan, Pacific Northwest National Laboratory
Mesoscale Modeling of Radiation Damage in Fusion Structural Materials

C.3 White Papers Addressing Discovery Plasma Science

Low-Temperature Plasmas

C-77 Iain D. Boyd, University of Michigan
Cross-Field Transport in Low-Temperature Plasma Systems

C-80 Miles Turner, Dublin City University
Low-Temperature Plasma Physics and Exascale Computing

C.4 White Papers Addressing Verification and Validation

Needs Driven by Experiments

C-88 N.T. Howard, Massachusetts Institute of Technology
Comparison of Multi-Scale Gyrokinetic Simulation with Experiment:
Computational Requirements and Impact on Modeling of Tokamak Plasmas

C-91 John Wright, Massachusetts Institute of Technology
Data Management, Analysis, and Assimilation

This page is intentionally left blank.

C.1 White Papers Addressing Fusion Energy Science

C.1.1 Turbulence and Transport

Kinetic Simulation of Low-Frequency Phenomena with Lorentz Ions and Gyrokinetic Electrons

Y. Chen, Center for Integrated Plasma Studies, University of Colorado at Boulder

Many phenomena in tokamak plasmas are essentially kinetic. At present, kinetic simulation of the low-frequency phenomena, such as neoclassical and anomalous transport, and the energetic particle-driven MHD modes is largely based on the gyrokinetic model. In the past two decades much progress has been achieved in gyrokinetic simulations. This white paper proposes to move beyond the gyrokinetic ion model for modeling low-frequency phenomena.

1. **Please specify the current science drivers for your field of research.**

At the present time there are many gyrokinetic codes (e.g., GS2, GYRO, GENE, GEM, GTC, GTS, GKW, GYGLES, XGC) that can be used for microturbulence and anomalous transport simulation. Aside from numerical details, these codes all solve the same gyrokinetic Maxwell system of equations. Many of these codes have been verified with each other, and much validation study has been carried out. In many cases, simulation has been able to predict the experimentally measured transport level.

However, the validity of gyrokinetics has been questioned in many areas. Although in theory the gyrokinetic model can be made as accurate as possible, in practice only terms that are accurate to the first order in gyrokinetic ordering are consistently retained. Such equations are not sufficient for the following reasons. First, Parra and Catto have argued that such first-order-accurate gyrokinetic formulation is not suitable for neoclassical transport that depends on the long-wavelength radial electric field. Their argument also implies that, in general, the gyrokinetic quasineutrality equation used in practical simulations is not accurate enough for the long-wavelength fluctuations. Second, the gyrokinetic ordering is violated for some edge plasmas that have strong profile variation, with pressure-scale length of only a few ion Larmor radii. Third, in regions where there is strong plasma flow with strong radial flow shear (such as in the transport barrier), electromagnetic gyrokinetic equations that are suitable for numerical implementation have not been derived yet. These considerations suggest that kinetic modeling with the Lorentz ion model should be developed, as a modeling tool for problems where current gyrokinetic simulation is not accurate, and as an independent verification of the gyrokinetic model. The gyrokinetic model can be used for the electrons with confidence due to the much smaller electron Larmor radius.

A complete set of equations has been proposed [1]. The exact, generalized Ohm's law is used to obtain the electric potential. The model is quasineutral, but the gyrokinetic Poisson equation is not used.

2. **Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.**

A fully electromagnetic Lorentz ion/gyrokinetic electron code will be available in 5–10 years, for tokamak or stellarator. Ion subcycling allows the simulation to last for about the same physical time as typical present-day gyrokinetic ion simulations (~1 to 5 ms). Problems that are beyond current gyrokinetic simulation, such as the formation of transport barriers with sonic-level, strongly sheared flow and the evolution of the long-wavelength radial electric field, can be solved.

3. **Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.**

Direct kinetic simulation (i.e., without some form of multiple-scale technique) on the transport timescale will not be solved. Direct-transport time-scale simulation requires the use of physical energy and particle source/sink in the kinetic equation, and it requires the simulation to continue over a timescale that are orders of magnitude longer than present-day simulations. Numerical accuracy over such a timescale cannot be trusted.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Good parallel algorithm for the 6-scalar 3D field equations	The Lorentz ion model can be best pursued with the Particle-in-Cell method, since evolving the distribution in PIC is straightforward for Lorentz ions. The main computational challenge is an efficient algorithm for the generalized Ohm's equations coupled to Faraday's law. For a global simulation, the problem size (number of grid points and unknowns) will be large due to kinetic electrons. An advanced parallelization scheme is needed.
2. Close collaboration between applied mathematicians and computational physicists	At the core of the field solvers is a massively parallel solver for linear equations with dense matrices. This will greatly benefit from applied math.
3. Supercomputers with large memory	Because the problem is now 6-D in phase space, the number of particles will be large. Particle arrays, as well as field arrays, should reside in memory for computational efficiency.

Impede	Why?
1. Supercomputers	While the Lorentz ion model code can be developed on present computers, productive runs can only be done on future computers.
2. Lack of efficient collisional algorithm	For some applications of the Lorentz ion model, such as neoclassical transport, a noise-free delta-f collisional algorithm is needed. Lessons learned from the ongoing XGC collisional scheme will be very useful.
3. Lack of a theoretically solid multiple-scale technique for whole-device, transport-timescale modeling	Direct kinetic simulation is most useful for short time dynamics. It is ill suited for a long time simulation. Direct long-time simulation results are difficult to interpret. For a whole-device, transport-timescale simulation, it is crucial to have a solid multiple-scale framework, in which the short-time direct simulation is used as a component tool.

5. Reference

[1] Chen, Y., and S.E. Parker, Particle-in-cell simulation with Vlasov ions and drift kinetic electrons, *Physics of Plasmas* **16**, 052305 (2009), http://dx.doi.org/10.1063/1.3138743.

Continuum Gyrokinetic Turbulence and Transport in Magnetic Fusion Research on Exascale Computers

D. R. Ernst, Massachusetts Institute of Technology, Cambridge, Massachusetts

1. Please specify the current science drivers for your field of research.

Dedicated experiments on the DIII-D and Alcator C-Mod tokamaks study core turbulence and transport in conjunction with comprehensive gyrokinetic simulations, elucidating the fundamental mechanisms relevant to burning plasmas [1,2]. Validation and verification are essential steps, including simultaneous comparisons with measured fluctuation spectra using synthetic diagnostics. The larger goal is to develop a validated predictive understanding of the turbulence and underlying instabilities that control transport in all channels, particularly with strong electron heating. Simultaneous quantitative agreement has been obtained for all transport channels and with measured density fluctuation spectra, without adjustments to measured profiles, with and without strong electron heating (in H-Mode plasmas), in only one study so far [1,2]. Prior to this, only the heat flux, density fluctuation spectrum, and intensity profile in one plasma condition were matched in L-Mode plasmas [3]. A host of other studies are unable to simultaneously match transport fluxes and fluctuations, particularly with strong electron heating, or they resort to adjusting measured gradients to achieve agreement.

Fundamental improvements to simulations, including accurate gyrokinetic collision operators [4], and developing algorithms for their implementation [5] are also supported by the DOE SciDAC Center for the Study of Plasma Microturbulence. New global continuum simulations of neoclassical transport in the H---Mode pedestal [6] have revealed the importance of strong radial electric fields and finite orbit widths in determining the pedestal flows and bootstrap current.

2. Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.

During this period, Summit at ORNL (IBM Power 9 with NVIDIA Volta GPUs; 3,400 multiprocessor nodes), Cori II at NERSC (Knights Landing 2nd gen. Xeon Phi, 9,300 nodes), and Aurora at Argonne (Knights Hill 3rd gen. Xeon Phi, 50,000 nodes, 180 PFLOPS) will bring many core technologies with an important new development – access to system memory at CPU bandwidth, eliminating a major bottleneck. The Knights Landing (KNL) processor also has 64 GB of onboard memory accessible at 4.4 times this speed, enough to avoid using system memory in many applications. These technologies will lower barriers for application development and should reduce the performance penalty for codes not well-optimized. These well-known barriers have resulted from the adoption of commodity graphics hardware to claim high ideal FLOP rates, which unfortunately could not be effectively utilized. The recent, more scientifically oriented technologies appear to have resulted in part from DOE partnerships with industry

through the DesignForward, FastForward, and FastForward2 programs [7]. The 2016 Presidential budget doubles funding for exascale development with industry.

This new KNL architecture, for example, should make it possible for the entire velocity space (in ion-scale continuum gyrokinetic simulations), as well as the parallel spatial dimension, to be parallelized with OpenMP, each in a single node. The fast on-chip memory access, combined with excellent single-thread performance including 8-wide double precision vectorization, should result in major speedups. The large node counts on these systems should allow us to do well-resolved predictive transport simulations of the plasma core using gyrokinetic simulations rather than reduced models. This capability could be extended to stellarators and could include self-consistent kinetic equilibrium, sources, and kinetic neoclassical transport calculations. More efficient gyrofluid simulations of multiscale core turbulence (ETG/ITG), with improved treatment of zonal flows, are being developed and will be verified against full gyrokinetic multiscale simulations (TGLF, GryffinX). At least one of these codes (GryffinX) runs on GPUs and, for ion scales, runs three orders of magnitude faster than gyrokinetic simulations. Reducing the time needed to compute turbulent transport would make further integrated simulations possible, for example, disruption prediction, and the interaction of turbulence with tearing modes and RF and fast particles. Overall, exascale promises to bring 500 times more computational power and 100 times more memory (and 100 times faster memory) than offered by present facilities.

3. Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.

Developing a validated understanding of edge turbulence in H---Mode pedestals will require an improved and tractable formulation in addition to new code development and raw computing power.

4. What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Hardware resources	During this time frame, the Xeon Phi is coming into its own in systems of unprecedented scale (9,300-node Cori II at NERSC and 3,500-node Aurora at Argonne). The Knights Landing and Knights Hill Xeon Phi multicore architecture not only has enough on-chip memory per core to serve as the main memory in many applications, at 4.4 times the bandwidth of off-chip memory, but off-chip memory can also be accessed at the DDR4 bandwidth of 90 GB/s, a dramatic improvement over the previous generation Knights Corner co-processor PCIe connection. The next system at OLCF will be IBM Power 9 + NVIDIA GPU based with NVLINK, which also allows the GPU core to access system memory at a similar 80 GB/s bandwidth. The ability to access system memory at CPU speeds, though not alone sufficient to keep all cores busy, should simplify programming and greatly improve performance.

Impede	Why?
1. Process for allocation of computational resources	For new proposals, sufficient computational resources to meet the proposed milestones are not guaranteed. NERSC ERCAP awards appear to be based on prior use, and significant time increments for new projects are handled through the ALCC, INCITE, and NICE programs. Applicants must compete across disciplines and recompete each year. These proposals are lengthy and peer-reviewed. If legally possible, providing additional computational resources as part of new grant awards would leverage one review process and reduce overhead for researchers, DOE, and reviewers, while helping to ensure sufficient resources for funded research.
2. Application optimization/ development support	Support resources for application development and optimization on the new architectures are limited. NERSC often provides good webinars and holds hackathons. However, having an expert working closely on code improvement appears to come mainly with Scientific Application Partnerships (FASTMATH), sometimes possible through SciDAC grants or ad hoc collaborations. Exascale computers will have one-fifth the memory per core.

3. Persistent archival data storage	We have no long-term archival storage available with guaranteed persistence. NERSC provides long-term storage (without guarantees), but does not legally comply with the new Data Management Plan requirements. At OLCF, the data are deleted when the allocation award ends. In an era of exascale computing, output data may become so voluminous that it will be impractical to transfer them to local institutions for permanent storage. A multisite VDF (Virtual Data Facility) has been proposed to add data storage and analysis resources to the existing ASCR facilities for all SC programs.

6. References

[1] *Controlling H-Mode Particle Transport with Modulated Electron Heating in DIII-D and Alcator C-Mod via TEM Turbulence*, D. R. Ernst, K. H. Burrell, W. Guttenfelder, T.L. Rhodes, L. Schmitz, A.M. Dimits, E.J. Doyle, B.A. Grierson, M. Greenwald, C. Holland, G.R. McKee, R. Perkins, C.C. Petty, J.C. Rost, D. Truong, G. Wang, L. Zeng and the DIII-D and Alcator C-Mod Teams, 25th IAEA Fusion Energy Conference (FEC 2014, Saint Petersburg, Russia), invited oral paper CN-221-EX/2-3, MIT PSFC Report JA-14-27.

[2] *Role of Density Gradient Driven Trapped Electron Modes in the H-Mode Inner Core with Electron Heating*, D. R. Ernst, K.H. Burrell, W. Guttenfelder, T.L. Rhodes, A.M. Dimits, R. Bravenec, B.A. Grierson, C. Holland, A. Marinoni, G.R. McKee, C.C. Petty, J.C. Rost, L. Schmitz, G. Wang, S.E. Zemedkun, L. Zeng, 57th Annual Meeting of the APS Division of Plasma Physics, Savannah, GA, 2015 APS-DPP invited paper NI3, Bull. Am. Phys. Soc. **60** (19) 214 (2015). Submitted to Phys. Plasmas (Dec. 2015).

[3] *Turbulence in the TORE SUPRA Tokamak: Measurements and Validation of Nonlinear Simulations,* A. Casati, T. Gerbaud, P. Hennequin, C. Bourdelle, J. Candy, F. Clairet, X. Garbet, V. Grandgirard, O. D. Gürcan, S. Heuraux, G. T. Hoang, C. Honoré, F. Imbeaux, R. Sabot, Y. Sarazin, L. Vermare, and R. Waltz, Phys. Rev. Lett. **102**, 165005 (2009).

[4] *Gyrokinetic Fokker-Planck Collision Operator,* B. Li and D.R. Ernst, Phys. Rev. Lett. **106**, 195002 (2011).

[5] *New velocity-space discretization for continuum kinetic calculations and Fokker-Planck collisions*, M. Landreman and D. R. Ernst, J. Comp. Phys. **243** 130 (2013).

[6] *Local and global Fokker-Planck neoclassical calculations showing flow and bootstrap current modifications in a pedestal*, M. Landreman and D. R. Ernst, Plasma Phys. Contr. Fusion **54** 115006 (2012).

[7] *America's Next Generation Supercomputer: The Exascale Challenge*, Hearing before the Subcommittee on Energy, Committee on Science, Space, and Technology, House of Representatives, Wednesday May, 22, 2013. Serial 113-31. https://www.gpo.gov/fdsys/pkg/CHRG-113hhrg81195/pdf/CHRG-113hhrg81195.pdf.

FES White Paper

Mastering the Edge of Fusion Plasmas

D. R. Hatch, Institute for Fusion Studies, University of Texas at Austin

The achievement of a robust H-mode transport barrier (pedestal) is an inescapable precondition for achieving confinement sufficient to produce a burning plasma in tokamaks. Recent gyrokinetic simulation results suggest that empirical scaling laws — the primary tool used to date to predict fusion performance — may be seriously inadequate for predicting pedestal properties in a burning plasma regime. Hence, developing the capability of modeling transport in the pedestal region represents a computational challenge whose mastery is indispensable for confidently designing future devices. The capability of pedestal predictive modeling, moreover, may even empower the identification of burning plasma devices whose size and cost are a fraction of current projections.

1. **Science drivers.**

In recent years, gyrokinetic simulations have established the ability to reproduce experimental core transport levels with ever more fidelity. Recent work using the GENE code[1] has demonstrated encouraging progress in extending these capabilities to the challenging edge pedestal region; simulations of the JET-ILW (ITER-like wall) pedestal demonstrate that microtearing modes are the dominant transport mechanism and, in combination with electron temperature gradient (ETG) turbulence and neoclassical transport, reproduce experimental transport levels very accurately.[2]

This initial success lays a foundation for a first attempt at identifying key trends that may alter pedestal dynamics in large (low ρ^*) devices like ITER. Since ρ^* is the only dimensionless plasma parameter whose ITER values cannot be accessed on present-day experiments, ρ^* dependences represent a critical challenge and opportunity for computational modeling. In particular, ongoing work is examining the dynamics of shear suppression — the key mechanism that facilitates edge transport barriers — as ρ^* decreases toward values that are expected on ITER. The first results[3,4] from gyrokinetic simulations demonstrate that, consistent with basic scaling arguments, shear suppression erodes as ρ^* decreases. This suggests that ITER likely lies in a fundamentally different regime from present-day experiments.

Thus, a primary challenge for gyrokinetic simulation is to be able to simulate pedestal physics with ever greater accuracy, to determine the degree of impact that low-velocity shear has on the expected ITER performance. Furthermore, recent results indicate that other tokamak geometries may have improved confinement under burning plasma conditions. Hence, another challenge for simulations is to help determine the optimal burning plasma geometry to ensure robust pedestal structure and concomitant confinement.

Ongoing work is

1. Pursuing increasingly comprehensive gyrokinetic pedestal simulations and expanding the capabilities of the GENE code to model pedestal dynamics.
2. Continuing validation efforts with present-day experiments, including JET, C-Mod, and ASDEX Upgrade.
3. Interacting with experimentalists to identify targeted scenarios to explore the limits of shear suppression.
4. Further verifying ITER results and pursuing possible solutions.

The striking progress of core gyrokinetic simulations over the past decade fosters confidence that similar advances can be made, with sufficient computational resources, in the critical effort to model pedestal dynamics. Developing such capabilities will ensure that the critical ingredients of good confinement can be built in when designing next-generation devices.

2. Science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.

With high-end resources (typical of, e.g., INCITE), it is likely that comprehensive (gradient-driven) gyrokinetic simulations of pedestal transport will be feasible in the coming 5–10 years. This would include global operation with full electromagnetic effects and, potentially, multiscale (ion and electron) dynamics. This could represent the ability to faithfully reproduce transport dynamics on existing experiments, given input profiles and equilibria.

3. Science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.

It is unlikely that fully comprehensive (as described above) simulations in the flux-driven mode (i.e., evolving profiles) will be feasible with current resources. This is what is necessary for truly predictive capability.

4. Computing ecosystem aspects in the next 5–10 years.

Accelerate	Why?
1. Increased CPU time	Efforts to date have been limited by available allocation limits.
2. Support for grad students / postdocs	Increasing manpower (i.e., support for graduate students and postdocs) would accelerate progress on several fronts.
3. Models and algorithms	Better models are needed for pedestal gyrokinetics and corresponding algorithms to make them numerically tractable.

Impede	Why?
1. Porting codes to advanced architectures	Upgrading codes for advanced architectures can be time consuming and distracting from obtaining physics results.
2. Limited CPU time	Limitations on CPU time often enforce constraints on how comprehensive simulations can be.

5. References

1. genecode.org
2. D. R. Hatch, M. Kotschenreuther, S. Mahajan, P. Valanju, F. Jenko, D. Told, T. Goerler, and S. Saarelma, "Microtearing Turbulence Limiting the JET-ILW Pedestal," submitted to *Physical Review Letters*.
3. M. Kotschenreuther, "Gyrokinetic Simulations of the ITER Pedestal," invited talk at APS-DPP, November 2015.
4. M. Kotschenreuther et al., manuscript in preparation.

First Principles Integrated Simulation of Boundary

Multi-Physics Using the Particle-in-Cell Method

Scott Parker[1] and C.S. Chang[2]

[1]*University of Colorado, Boulder*
[2]*Princeton Plasma Physics Laboratory*

The study of tokamak boundary physics is a scientific grand challenge due to several interrelated physical complexities including: (I) the nonequilibrium state of boundary plasma, (II) the highly nonlinear scale-inseparable multiphysics that interacts both in velocity and configuration spaces, (III) the wall-born neutral particles that interact with the plasma through atomic collisions, and (IV) the complex geometry including the magnetic separatrix and material wall.

(I) Nonlocal non-equilibrium state of the boundary plasma: Strong plasma source and sink drive the boundary plasma into a non-equilibrium state. There is no heat reservoir for boundary plasma as often assumed in the non-equilibrium statistical mechanics. The environment – core plasma and wall – is strongly coupled non-locally to the boundary plasma. Core temperature responds in a stiff manner to the boundary plasma. There are also internal non-equilibrium drivers. Wall interacts with boundary plasma through plasma-material interaction (PMI). The pedestal gradient scale length is roughly the same as the neoclassical orbit excursion width and the blobby turbulence size. The kinetic physics information is continuously mixed between different pressure regions at the timescale of particle orbital motion, which is inseparable from the turbulence, edge instability, and neutral transport timescales. When the pedestal gradient becomes too steep, the strong free energy drives the boundary plasma into large-scale edge-localized mode (ELM) instabilities. Even in a quiescent state, the boundary plasma contains large-scale blobby turbulence, which may not be described by equilibrium thermodynamics. It will be highly difficult, if not impossible, to properly close the fluid equations under these situations. *A fully kinetic approach is needed.*

(II) Scale-inseparable nonlinear multi-physics: All the important boundary physics phenomena – turbulence, neoclassical particle dynamics, ELMs, and neutral particle transport – have significantly overlapping space-time scales. The pedestal profile may evolve more slowly than others in the absence of ELMs, but its radial scale length is similar to others. These physics components interact nonlinearly with each other to form the boundary plasma. The conventional modular theoretical and computational approach that assumes scale separation among the multi-physics phenomena has very limited validity applied to the boundary, and will face very difficult mathematical constraints at best. A common first-principles set of equations needs to be solved that contains the multiphysics without scale separation.

(III) Neutral particles: Plasma interaction with the material wall produces neutral particles that are an important particle, momentum, and energy source/sink to the whole plasma. Since the neutral particles are intrinsically in a non-thermal state as well, their transport needs to be studied kinetically. Thus, the plasma-neutral model must be kinetic-kinetic for reliable predictions.

(IV) Complicated geometry: The plasma boundary crosses the magnetic separatrix surface, which divides the closed and open (SOL) magnetic regions. The SOL plasma is in contact with the arbitrarily shaped material wall. The geometry effects on the boundary plasma are known to be important for all spatial regions, including the edge pedestal region, the scrape-off layer, and the divertor region. The numerical method that is used to study boundary plasma needs to be robust to be able to cope with the difficulty caused by the complicated topology and geometry.

After carefully analyzing available models and numerical methods for solving the gyrokinetic equations in the boundary region, we have chosen to use the particle method as the primary tool. An ODE particle code is much less susceptible to the show-stopping CFL stability condition in both configuration and velocity space under large amplitude fluctuations. For a particle code, the field part is separated from the grid, and only accuracy of particle dynamics is required, leading to an indirect CFL-like accuracy condition in configuration space. Particle methods are amenable to modeling arbitrary-shape recycling boundary conditions and neutral particle transport from first principles. This endeavor has already been selected as a joint OFES and OASCR project, currently the Center for Edge Physics Simulation (EPSI) in the SciDAC-3 program. A very strong collaboration between the OFES and the OASCR scientists is being fully utilized to solve the difficult boundary plasma problem. As a result, the edge gyrokinetic particle code XGC1 has emerged, fully utilizing the largest open-science computing platforms. XGC1 is the leading international code in the field of kinetic boundary simulation; it contains neoclassical physics, neutral particle recycling and transport, atomic cross-sections, blobby electrostatic turbulence, and edge-core interaction. XGC1 has been revealing the physics of pedestal, blobby turbulence, edge momentum source, and divertor heat-load footprint at a first-principles level for the first time. There are continuum edge gyrokinetic codes under development in the United States, including the ESL code at Livermore and the Gkeyll code at PPPL, which employs a new technique (discontinuous Galerkin method).

XGC1 has successfully acquired the electromagnetic turbulence capability using the gyrokinetic ions and fluid electrons. Various verification exercises have been performed that include the tearing modes. The linear and nonlinear onset of kinetic ballooning modes is presently being studied in the edge pedestal plasma. Work is underway to develop fully kinetic electron extensions to the electromagnetic model in XGC1.

Besides the electromagnetic turbulence, a few other challenging physics features need to be added to XGC1 in order to complete the boundary physics capability at a first-principles level. ELMs are not really scale-separable from turbulence, and their mutual interactions could be strong. In XGC1, in the future, ELMs will be simulated together with neoclassical and turbulence physics from the same set of gyrokinetic equations. The fluid or MHD codes are capable of studying only the large-scale Type-I ELMs. However, ITER may have to rely upon small-scale ELMs, which have not been seen from the fluid/MHD codes, nor have they been understood. XGC1 will investigate the small ELM physics, too.

Control of ELMs by external resonant magnetic perturbation (RMP) coils or molecular injection is another outstanding issue. XGC1 needs to include these capabilities in the future. With the XGC1's capability in combining MHD/fluid modes, electromagnetic turbulence, neoclassical physics, and neutral-atomic physics, a comprehensive study of ELM control could be possible at first-principles level. A reduced version of XGC0 already possesses the kinetic RMP penetration and plasma transport response capabilities.

Having a realistic PMI model is important for the fidelity of the boundary plasma simulation. XGC1 can evaluate the ion bombardment data that are necessary for accurate PMI modeling, which include the ion flux PDF in the incident angle and the incidence kinetic energy at each wall position. For a more accurate evaluation of the sputtered impurity recirculation at the material wall, a six-dimensional Debye sheath calculation could be desirable, instead of the "logical" sheath that XGC1 is presently calculating. XGC1 can use an embedded 6D simulation technique in front of the material wall for this purpose.

In returning to the easy and physics capabilities, computation in XGC1 is expensive due to the large number of particles required for Monte Carlo noise reduction, hence requiring an extreme-scale high-performance computer (HPC) with good code scalability. Throughout the development of the XGC1 particle code, a hand-in-hand partnership with the OASCR scientists in all four SciDAC Institute areas (Data Management, Applied Mathematics, Performance Engineering, and Uncertainty Quantification) has been required to overcome the challenges. As a result, XGC1 scales efficiently to the maximal hardware capability, with a high degree of portability, on the major leadership class computers; including the heterogeneous Titan, and the homogeneous Mira and Edison. Production runs utilize the maximal available capability of these HPCs. The more powerful the computers are, the more physics XGC1 can include.

With the new hardware and software architectures employed by the future leadership class computers, and with further development of XGC1 to include more complete boundary physics, the collaboration with OASCR scientists will continue to be highly important. The technical merit of XGC1 development into exascale has been proven by the recent selection into the main pre-exascale programs at OLCF CAAR and NERSC NESAP.

Fusion Energy Sciences White Paper

Massively Parallel Gyrokinetic Turbulence Simulations and Application to Alternative Fusion Reactor Concepts

M.J. Pueschel, University of Wisconsin-Madison, Madison, Wisconsin 53706

This white paper addresses ongoing and future developments in the area of alternative fusion reactor concepts – the stellarator and reversed-field pinch (RFP) – and, specifically, challenges for numerical simulations of plasma phenomena and turbulence in these devices.

1. Science Drivers

Despite the fact that present-day stellarators and RFPs tend to simultaneously be less well diagnosed experimentally than leading tokamak-type devices and offer specific challenges for numerical simulations. Ongoing research at the University of Wisconsin—Madison encompasses various avenues of research on the Helically Symmetric eXperiment (HSX) stellarator and the Madison Symmetric Torus (MST) RFP.

High-performance discharges on MST tend to utilize pulsed poloidal current drive in order to suppress tearing modes in the core plasma, which otherwise would strongly deteriorate confinement. With tearing modes reduced to significant lower levels, pressure gradients steepen in the outer radii, leading to the destabilization of trapped electron modes (TEMs). If tearing modes were to be stabilized completely, zonal flow activity would be able to push the nonlinear critical gradients to values much larger than those observed in the experiment—residual tearing mode activity and its impact on zonal flows[1,2] need to be considered to explain experimental heat flux levels.[3] Presently, tearing modes are modeled by adding a static, resonant magnetic perturbation to the field equation in gyrokinetic turbulence simulations with the GENE code.[4]

HSX plasmas are similarly producing TEM turbulence. Flux-tube-based simulations show a rich variety of physical effects, including the formation of a coherent structure at the rational surface, which increasingly affects transport at larger density gradients.[5] Furthermore, it has been shown that the linear and nonlinear excitation of subdominantly unstable and stable modes are essential for turbulence and transport.[6]

2. Solvable Science Challenges in the 2020−2025 Time Frame

Regarding MST, global multiscale simulations will be used to model core tearing modes self-consistently with TEM microturbulence at outer radii, thereby including all relevant physics to fully describe how zonal flows – which regulate TEM turbulence and, by extension, heat fluxes – react to the tearing modes. This will provide essential information on how to improve MST confinement, and ideally will help to make the RFP concept competitive as a fusion reactor design. In stellarators, the next decade will see advances in full-flux-surface simulations, as well as radially global studies. For HSX in particular, both types of investigations will be used to determine their effect on coherent structure formation and behavior. Furthermore, the flexible design of HSX stellarator will be utilized for transport optimization based on simulation predictions. Quasilinear modeling and nonlinear verification will play a major role in this context.

3. Unsolvable Science Challenges in the 2020−2025 Time Frame

The simultaneous inclusion of tearing modes, microturbulence, and fast particle effects in MST simulations is not within reach within the next ten years. On HSX, it will not be feasible to include the entire plasma – corresponding to a fully global domain – in a single simulation while retaining all other relevant physics, such as electromagnetic effects.

4. Computing Ecosystem Aspects Affecting Science Progress

Accelerate	Rationale
1. Availability of CPU hours	Presently, projects have to be scaled back due to limited high-performance computing resources.
2. Scientific personnel funding	With university and federal funding for professors, scientists, postdocs, and graduate students becoming increasingly scarce, long-term planning and high-risk/high-reward projects become nearly impossible.

Impede	Rationale
1. Intermittent personnel funding	Periods of strongly varying availability of project funding and well as short funding periods strongly impede scientific progress and cause loss of trained personnel.
2. Short-term computing resource planning	One-year periods for computing time proposals are too short for some projects and force modularity and a small-steps approach.

References

[1] M.J. Pueschel *et al.*, Phys. Plasmas **20**, 102301 (2013)
[2] P.W. Terry, M.J. Pueschel, D. Carmody, and W.M. Nevins, Phys. Plasmas **20**, 112502 (2013)
[3] D. Carmody, M.J. Pueschel, J.K. Anderson, and P.W. Terry, Phys. Plasmas **22**, 012504 (2015)
[4] *see* http://www.genecode.org *for code details and access*
[5] B.J. Faber *et al.*, Phys. Plasmas **22**, 072305 (2015)
[6] M.J. Pueschel *et al.*, *submitted to* Phys. Rev. Lett. (2015)

Turbulence, Transport, and Transients in Boundary Plasma Modeling

Maxim Umansky, Lawrence Livermore National Laboratory

1. **Please specify the current science drivers for your field of research.**

In the field of boundary plasma modeling, the main science driver is the need to develop predictive capability, based on computer simulations, for turbulent and collisional transport and various transient phenomena in the edge plasma: (i) radial transport of plasma particles and energy, (ii) radial transport of impurity ions, (iii) parallel heat transport in weakly collisional plasma, (iv) self-organized nonlinear phenomena (ELMs, EHO, QC mode, etc.) that in many cases dominate transport processes; (v) phase transitions between various transport regimes (L-, H-, I-modes); and (vi) plasma-material interactions. Present-day understanding of these phenomena (i–vi) is limited, in many cases semiqualitative at best. This is partially due to limitations of analytic theory and numerical models, but in many cases, it is also due to insufficient experimental measurements. The predictive capability for (i–vi), if acquired, would revolutionize the design of tokamak-based fusion reactors and would have a strong impact on the timescale of developing practical fusion-based energy sources, which is in the interest of the U.S. Department of Energy and beyond.

2. **Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.**

There is a good chance that in the next 5–10 years, significant advances will occur in the understanding of (a) SOL width, (b) ELMs, (c) EHO, QC-modes, (d) L-H transition, and (e) detached divertor regimes. This will bring us closer to predicting properties of edge plasma in tokamak experiments.

3. **Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.**

Self-consistent predictive modeling of plasma-material interactions, in its full complexity, is not likely to be developed within 10 years because the material physics has an enormous range of spatial and temporal scales. This would impede designing of a practical tokamak-based fusion reactor.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Application codes	In the edge plasma field, there are a number of existing application codes that could likely produce important advances if more work could be done on applications.
2. Data workflow	Developing better ways of communicating data should improve the throughput.
3. Hardware resources	Increasing the resolution of edge plasma simulations will open the possibility of including a larger range of spatial and temporal scales.

Impede	Why?
1. Workforce development	Insufficient application workforce would impede progress.
2. Internal/external libraries/frameworks	Application codes used in edge plasma community use standard mathematical libraries and frameworks (SUNDIALS, PETSc, SLEPc, Hypre, Chombo, etc.). If these libraries cannot successfully transition to new architectures coming in next 10 years, then the application codes would not be able to take advantage of advances in computer hardware.

Exascale Computing of Confinement Physics Coupling Turbulence, MHD, and Neoclassical Dynamics in Advanced Fusion Experiments

W.X. Wang, Princeton Plasma Physics Laboratory

As fusion research moves to the new era focusing on ITER, along with the development of other advanced experiments such as Fusion Nuclear Science Facilities, understanding global confinement phenomena in burning plasma regimes, which couple multiscale dynamics and multiphysics tightly together, is critical and presents a grand challenge in the next 5–10 years. Exascale computer simulations will play a key role in providing a solution to the challenge. Discussed in this white paper are a few examples.

The multiscale coupling of turbulence dynamics and macroscopic MHD evolution. Here, one of the most prominent issues is the slow, self-consistent interaction of magnetic islands with turbulence, in the context of neoclassical tearing mode (NTM) growth and associated major plasma disruption. Experimental observations suggest that a large fraction of discharge disruptions could be triggered by the incident of NTM. Comprehensive understanding of NTM physics needs to take into account its interaction with turbulence. There is a broad class of questions to be addressed concerning slow NTM evolution in the presence of turbulence. These include how kinetic effects modify the threshold condition of NTM; how turbulence-induced transport affects NTM evolution (e.g., through its effects on the ratio of parallel to perpendicular transport); how important nonlocal transport physics is (e.g., through turbulence spreading across the separatrix of islands); how pressure profiles flatten and saturate for finite island size; what the effects are of islands on rotation, in particular, intrinsic rotation; how NTM physics (e.g., threshold and saturation amplitude) depends on rotation shear; what type of confinement state or regime may result when an island is present; and so on. Answering these questions is extremely nontrivial and requires highly self-consistent modeling dealing with large separation of timescales associated with turbulence as well as macroscopic time evolution, which together will require exascale computing. The exascale simulations for such physics require the coupling of high-fidelity, first-principles-based gyrokinetic simulations of turbulence with transport models and MHD simulations, so as to develop an iterative approach to the solution. For this purpose, the kinetic simulations should be able to deal with global, flux-driven, electromagnetic turbulence. One of the central issues affecting NTM physics that we want to highlight is the generation of bootstrap current in the presence of turbulence and islands. Turbulent fluctuations may drive a non-inductive plasma current via nonlinear flow generation, in particular, in electrons [1, 2]. A self-consistent calculation of plasma self-driven current (both amplitude and profile structure) requires a multiscale gyrokinetic simulation integrating both turbulent and neoclassical physics consistently in island geometry, which, by itself, presents a challenge for HPC. Self-driven plasma currents have a generic but great impact on overall plasma confinement, in particular, for long-pulse magnetic fusion experiments, and the possible existence of fluctuation-induced plasma current may radically affect our understanding of tokamak physics in many aspects.

Global simulations of electromagnetic turbulence with focus on electron transport in high-beta, advanced spherical tokamak experiments. The low-aspect-ratio spherical tokamak (ST) experiments explore an alternative roadmap toward fusion energy production compared to that of conventional tokamaks. Energy transport in STs is usually dominated by the electron channel, which is always highly anomalous. There remains tremendous effort in the quest to understand electron transport and associated confinement properties in high-beta ST regimes. Gyrokinetic simulations carried out so far suggest that electromagnetic turbulence such as microtearing mode (MTM) turbulence could be important. However, the work toward fully resolving the problem could be a lot harder with respect to the computational size of the problem and the demand on HPC power. First, global, nonlocal physics is important in determining ST transport, which requires the use of global simulations for such a small-aspect-ratio device (local simulations of ion-scale fluctuations normally predict a much higher transport in ST regimes). Furthermore, for collisional MTM, the current layer near a rational surface could be of the scale of the electron skin depth, which is about (1/20 ~ 1/10) rho_i (in ST beta regimes). For collisionless MTM, the size of the current layer near a rational surface could even be close to the order of the electron gyroradius. Therefore, the fine spatial resolution needed to resolve the current layers may dramatically increase the size of the simulations. Moreover, the difficulties are more pronounced when the problem is being solved for edge parameters, for which high “q” values and high magnetic shear can dramatically increase the number of rational surfaces, and for future advanced ST experiments with a larger size. Taking all these considerations together, this may present an outstanding issue for exascale computing to solve.

Global ITER-size gyrokinetic simulations including both turbulent and neoclassical physics. This may present a case for exascale computing for the following reasons: first, the size of ITER simulations is at least 40 times larger than that of current DIII-D plasmas; second, simulations need to cover well-separated timescales from the turbulence timescale to the collisional timescale (for neoclassical dynamics). The science drivers for this include size (rho*) scaling of intrinsic rotation and prediction of intrinsic rotation (both amplitude and profile structure) in an ITER regime; plasma self-driven current (both neoclassical and fluctuation-driven) in a burning plasma regime; and non-neoclassical poloidal flow, which may make a considerable contribution to the mean radial electric field in low-torque ITER discharges.

The top three computing ecosystem aspects that may accelerate or impede fusion research progress in the next 5–10 years are application codes (implementation, development, and portability), models and algorithms, and hardware resources.

[1] W. X. Wang et al., *Proceedings of the 24th IAEA Fusion Energy Conference, San Diego, CA* (IAEA, Vienna, 2012), TH/P7-14.
[2] C. J. McDevitt et al., Phys. Rev. Lett. 111, 205002 (2013).

C.1.2 Energetic Particles and MHD

Computational Needs: Coupling Extended MHD Simulations and RF Wave Codes

T.G. Jenkins,[*] S.E. Kruger,[*] J.R. King,[*] E.D. Held[+]
[*]*Tech-X Corporation, 5621 Arapahoe Avenue Ste. A, Boulder, CO 80303*
[+]*Utah State University, Logan, UT 84322*

Current Science Drivers
The experimental use of radiofrequency (RF) waves to drive localized plasma currents that suppress deleterious tokamak instabilities (e.g., neoclassical tearing modes [NTMs]) has been remarkably successful [1]. Sophisticated control algorithms for the application, timing, and steering of injected RF power have been empirically developed; these have significantly reduced the likelihood of NTM-induced disruptions that could damage experimental hardware in existing experiments [2]. Avoidance of such disruptions will be crucial in the ITER device, because its stored energy is projected to exceed that of any present-day device by at least an order of magnitude. Numerical simulation can augment physics understanding of the coupled RF/extended-MHD interaction that is critical to mode stabilization and can also explore broader issues such as control strategies, RF power optimization, and so on. Parameter regimes for which the plasma disruptivity is not empirically known can be explored in numerical experiments without risk to the device.

The recent derivation of a self-consistent theoretical [3–5] framework in which the RF/extended-MHD interaction can be explored facilitates such predictive numerical analysis.[1] Loose coupling between the RF and MHD codes is sufficient for such analysis, permitting the use of vastly different data structures and representations (e.g., finite element, spectral, and ray-tracing characteristics) in these two classes of code. Robust numerical techniques [6] have been developed to map RF ray-tracing data onto more conventional extended-MHD representations such as finite elements. A python-based simulation framework [7], developed to manage the interactions of loosely coupled physics components, facilitates the RF and MHD data manipulation and exchange.

Science Challenges for the 2020–2025 Time frame
Research and progress in coupled RF/MHD simulation requires large-scale computing resources. Physics issues of interest include (a) quantification of the driven RF current efficiency in various operating regimes; (b) the influence of source width and position on island stabilization; (c) the detailed physics of RF effects on closure, and the role of this physics in Fisch-Boozer or Ohkawa stabilization mechanisms; (d) the relationship of detailed 3D models to 1D modified Rutherford equations; (e) development of optimal NTM control strategies for a fixed RF input power; and (f) optimal responses to mode locking scenarios that arise as NTMs interact with resonant magnetic perturbations (RMPs). The small size of the resonant region in which the RF modifies plasma dynamics necessitates very high resolution (possibly sub-millimeter), while the tearing mode size is on the order of the size of the device. Thus, substantial computing efforts (hundreds of runs using tens of thousands of cores) at or near the capacity of existing resources are required.

---

[1] Initial theoretical and computational work—carried out by the SciDAC Center for Simulation of RF Wave Interactions with MHD (SWIM)—has continued on a limited basis with support from the SciDAC Center for Extended MHD Modeling (CEMM). Reduced MHD simulations that use heuristic models for the RF-induced currents (methods used primarily by European research groups) have also begun to explore some of the basic physics imparted by RF.

Exacerbating the problem is the need for higher-fidelity physics models. Closure computations, which in their most general form require solutions of 5D drift-kinetic equations throughout the computational domain [8], impose additional computing and storage requirements; the phase space resolutions required to guarantee numerical convergence may not be attainable using present resources. Tighter coupling requirements between the RF and MHD aspects of the problem may also be imposed by neoclassical or closure physics. Existing computing ecosystems have already modeled rudimentary RF/MHD interaction at marginal resolution. However, the increased capability that larger-scale systems afford will enable more detailed models, including neoclassical effects and full closure computations, to be fruitfully compared with experiments.

Top Computing Ecosystem Aspects to Accelerate or Impede Progress

Accelerate	Why?
1. Dedicated consulting and financial support for code refactoring issues raised by new computing ecosystems.	Minimizes scientific productivity losses as the computing platforms supporting the scientific studies evolve, and ensures the optimal use of new computing platforms.
2. RF ray-tracing code development to make optimal use of GPU architectures	GPU-enabled RF computations are faster, enabling more tightly coupled RF/MHD modeling scenarios to be carried out efficiently.

Impede	Why?
1. Code refactoring requirements imposed by fundamental computing ecosystem changes.	Diverts time and effort from the physics studies to re-establish currently extant code capabilities.
2. Systems that prohibit the use of Python on the back-end nodes (although we only need the master node to run Python)	The problem is loosely coupled and the feedback systems require rapid prototyping. Python is perfectly suited to this problem.

References

[1] H. Zohm et al., Nucl. Fusion **39**, 577 (1999); G. Gantenbein et al., Phys. Rev. Lett. **85**, 1242 (2000).
[2] E. Kolemen et al., Nucl. Fusion **54**, 073020 (2014).
[3] C. C. Hegna and J. D. Callen, Phys. Plasmas **16**, 112501 (2009).
[4] J. J. Ramos, Phys. Plasmas **17**, 082502 (2010); Phys. Plasmas **18**, 102506 (2011).
[5] T. G. Jenkins and S. E. Kruger, Phys. Plasmas **19**, 122508 (2012).
[6] T. G. Jenkins and E. D. Held, J. Comp. Phys. **297**, 427 (2015).
[7] http://sourceforge.net/projects/ipsframework/.
[8] E. D. Held et al., Phys. Plasmas **22**, 032511 (2015).

Capability and Capacity Needs for Implicit, Nonlinear, Continuum Modeling

J R King, T G Jenkins, S E Kruger, A Y Pankin [Tech-X Corp.],
C R Sovinec [U. Wisc.], E D Held [Utah State U.], V A Izzo [UCSD]

Current science drivers

Computational support of magnetic fusion energy (MFE) enables better understanding of current fusion-plasma capabilities and provides projections for future devices. This white paper discusses the computational needs for a subset of codes for research in MFE that employ implicit, nonlinear, continuum methods. The sophistication of the model employed with this method varies from 3D MHD or extended-MHD to 5D drift-kinetic equations. Typically, these codes explore instability dynamics around a confined equilibrium state. For the diverted tokamak concept, this includes vertical-displacement and sawtooth events, neoclassical tearing, edge-localized and resistive-wall modes, disruption mitigation through gas and pellet injection, runaway electron beam generation, confinement and impact to the wall, and energetic particle effects. Details and further references on work on these applications by the NIMROD and M3D-C1 codes can be found on the NIMROD team [1] and CEMM websites [2].

Advances in computational power are moving simulations from those that interpret the physical dynamics of current experiments toward those that both predict dynamics and quantify the uncertainty in that prediction. The latter set of simulations not only require high fidelity (in terms of resolution and coupled physics models), but also must examine a large parameter space in order to determine the sensitivity of the dynamics to the underlying fusion-plasma discharge conditions (which may not be known exactly and may vary in time). This provides a need for both capability (runs that require extreme core counts for high-resolution modeling) and capacity (many independent runs that explore a parameter space) computing.

Science challenges for the 2020–2025 time frame

Progress on each science challenge listed above will be made in the next 5–10 years, even with current resources and extant computing systems. However, the degree of progress in terms of confidence in the result will be enhanced by next-generation systems.

In terms of capability applications, larger computational systems enable higher-resolution, nonlinear, continuum-kinetic (5D) modeling. This modeling represents the vanguard of the continuum method for MFE and enables modeling with high-fidelity neoclassical closures and/or with energetic particle populations. Relative to particle-in-cell methods, the continuum approach has the advantages of a straightforward implicit implementation and accurate collision operators without relying on extreme numbers of particles per cell. Another application for capability jobs is 3D extended-MHD modeling of the tokamak edge. Compared to applications that study the tokamak core, simulations require high toroidal and poloidal resolution to model the large-wave-number modes associated with the tokamak edge. Currently, these simulations are prohibitively expensive with existing computational resources, and this slows progress on edge studies.

Capacity applications are also critical. At present, nonlinear MHD cases are largely used to interpret experimental results, but are they not expected to produce an exact match. Discrepancies are expected for two reasons: (1) the initial condition is not known exactly and (2) simplifying model and algorithmic approximations may lead to error. Capacity computing permits a larger exploration of varied initial conditions and model approximations. Consider, for example, a capacity application, the highly successful ELITE code capability, which maps the 2D edge-stability parameter space in terms of normalized current and pressure gradient (known as a varyped). Although varyped plots have become ubiquitous in analysis of the tokamak pedestal and discharge conditions, building upon this model is possible. Currently, the linear ideal-MHD model with an analytic vacuum region outside the last closed flux surface (LCFS) is used to assess stability. Although this model is computationally prudent with existing resources, if resources were greater, similar parameter space modeling could be done with linear, extended-MHD calculations that include resistive and drift effects as well as plasma modeling outside the LCFS. Even greater resources still may enable nonlinear simulations that would predict ELM frequency and impulsive heat loads instead of stability; however, work on the algorithms to perform such a calculation remains to be done.

Top computing ecosystem aspects to accelerate or impede progress

One advantage of high-order finite elements is that the parallel implementation of the assembly operations is relatively straightforward. With the NIMROD code, the computational scaling of the assembly is very efficient for a variety of architectures (e.g., Intel CPUs, Intel MIC, BG/Q) and methods of parallelism (e.g., MPI-only or MPI+OpenMP). The largest barrier to scaling is preconditioning the ill-conditioned matrices required for implicit extended-MHD solvers. Present approaches with block-Jacobi poloidal-plane preconditioning using LU-decomposition sparse solvers exhibit limited scaling, but have been more successful than other approaches such as the multigrid method, in which scaling with the ill-conditioned matrices is difficult. Advances in both external libraries and internal code capabilities for this purpose would be highly beneficial, and the assembly implementation remains ready to conform to whatever parallelism model is chosen by the external libraries.

One aspect that precludes progress by implicit codes is the emphasis by some computational-award programs on algorithmic scaling (e.g., scaling per time step or solver iteration). The ultimate metric should instead be resources used relative to the science goals achieved. If scaling data are required as a proof of concept for resource utilization, time-to-solution scaling on the proposed problem should be chosen, not algorithmic scaling. Algorithmic scaling, which is useful internally to a project to characterize code-kernel performance, does not provide a full picture of time to solution. For example, the time-step size decreases when explicit algorithms weak scale with a fixed domain size, but this decrease is not reflected in a computational cost per time-step plot. These arguments are not against the use of explicit (or other) algorithms, which are well-suited for certain classes of problems, but rather we argue that the current system does not permit an apples-to-apples comparison of the capabilities of different codes.

Other aspects that impede scaling, such as I/O, memory, and visualization, are significant, but these have been tractable with current solutions (e.g., parallel HDF5, memory profiling, and the VisIt software). With high resolutions, it is likely that in situ analysis during the run will be required to limit the size of the output. Additionally, distributed memory versions of the current pre- and post-processor work flows will also be necessary. In order to make the best use of large machines, resources should be dedicated to these software development tasks.

References

[1] nimrodteam.org
[2] w3.pppl.gov/cemm

Capacity Computing for Macroscopic Plasma Dynamics

C.R. Sovinec, *University of Wisconsin-Madison*
J.R. King and S.E. Kruger, *Tech-X Corporation*

1. Science Driver

To realize self-heated conditions in the near future, experimental magnetic confinement efforts are focusing on devices of increasing size, where macroscopic plasma dynamics can cause material damage and significant programmatic setbacks. Nonlinear numerical modeling contributes to confinement studies by addressing questions of both practical and fundamental interest. The modeling needs for disruptions and edge-localized modes (ELMs) in tokamaks are described in the recent Fusion Energy Sciences/Advanced Scientific Computing Research community workshop report on integrated simulations.[1] They include nonlinear simulations to characterize the macroscopic dynamics of disruptions and ELMs and to aid the development of mitigation and suppression systems that will protect hardware. At present, nonlinear magnetohydrodynamic (MHD) and two-fluid computations struggle to achieve sufficient resolution to model macroscopic dynamics in realistic conditions. Some simulations can be accomplished as heroic efforts, consuming large fractions of annual computational allocations and/or many calendar days. Looking ahead, advances in computing and algorithms will be applied to make the models more comprehensive and to broaden studies with fluid models.

This white paper emphasizes the role of "capacity" computing in broadening our studies. It is seldom the case that a single large heroic computation provides more than a confirmation of less-resolved results or a demonstration of capability. Understanding the mechanisms that lead to different sequences of events during disruptions and ELM cycles in similar experimental conditions requires many nonlinear simulations to handle experimental uncertainties. While some effects will require new kinetic and neutral models, scans with existing fluid-based plasma models can address—or can provide an important step for addressing—the locking of magnetic islands to external structures and the magnetic forcing from disruptive events. Scans will also be critical for designing and improving disruption mitigation systems and when testing ELM suppression strategies.

The multiscale nature of plasma macroscopic dynamics favors implicit numerical methods for dynamically appropriate time steps. Parallel computing is communication intensive, even with methods that scale well, because each appropriate time step entails global physical coupling. In addition, many (10^4–10^6) large time steps are typically needed to examine nonlinear limit cycles and the consequences of dynamics over transport scales. In fluid-based models, the number of degrees of freedom in the state vector seldom reaches 10^9, which is modest in comparison with kinetic plasma computations that have additional dimensions for velocity coordinates. Nonetheless, the combination of problem size and communication requirements challenges computing efficiency in present-day node/interconnect computing architectures.

2. Contemporary Computing Ecosystems

Macroscopic plasma simulations can be made to scale reasonably well on contemporary computing ecosystems, meaning hardware nodes with modest core counts, if the interconnect hardware has low latency and if off-node communication is minimized in the most intensive part of the algebraic solves. The latter is tractable for codes like NIMROD[2] and M3D-C1[3] that employ more aggressive preconditioning strategies over two of the three spatial dimensions. Recent developments in algebraic solvers, time-stepping methods, and spatial representation for macroscopic plasma dynamics may hold promise for further reducing the turnaround time of nonlinear parameter and sensitivity studies.[4] As a complementary development, hybrid parallelization with on-node threading and off-node message passing for the contemporary ecosystems has proven effective in reducing memory requirements in NIMROD, and it is expected to be part of the strategy for future efforts.

If we are able to take advantage of new developments in multiscale algorithms and hybrid parallelization, contemporary ecosystems can provide some of the capacity to run nonlinear parameter scans. However, the resources presently available at NERSC are insufficient when they are divvied among many users or when they are dedicated to capability computations.

3. Future Computing Ecosystems

As the ITER experiment comes closer to first plasma, the U.S. and international computational communities need to become more responsive in answering technical questions in a timely fashion. The challenges of capacity computing must be resolved, which is unlikely with contemporary ecosystems. It is therefore imperative to devise future computing ecosystems for efficient capacity computing at large scale. To this end, it is important to recognize that the computing challenge for macroscopic modeling is off-node communication rather than in-core processing. If sufficient memory can be made available, slower many-core processing with less off-node communication, for example, would be appropriate for nonlinear parameter scans. On the other hand, architectures that rely on accelerators with restriction of data movement between processor and accelerator cores would compound already existing communication challenges.

4. Key Computing Ecosystem Recommendations

The following summarizes computing ecosystem aspects that are expected to accelerate progress in studying nonlinear macroscopic dynamics.

1. Hardware resources: computing centers for energy research have chosen their largest systems for capability computing. Shifting philosophy to recognize and value capacity computing that exceeds local-workstation/cluster capabilities is essential for computational studies of critical macroscopic dynamics in magnetic confinement.
2. Algorithms: changes in solvers, spatial representation, and time-advance methods have led to revolutionary improvements in solving macroscopic dynamics problems. Past experience fosters hope for similar improvements in the face of changing computing environments.
3. External libraries: much of the most computationally intensive aspects of fluid-based computation is handled by external solver libraries. Macroscopic modeling relies on their continued development for changing environments and hardware capabilities.

---

[1] Report of the Workshop on Integrated Simulations for Magnetic Fusion Energy Science, Offices of Fusion Energy Science and Advanced Scientific Computing Research, Rockville, MD, June 2–4, 2015. http://science.energy.gov/~/media/fes/pdf/workshop-reports/2016/ISFusionWorkshopReport_11-12-2015.pdf.

[2] C.R. Sovinec, A.H. Glasser, T.A. Gianakon, et al., J. Comput. Phys. **195**, 355 (2004); https://nimrodteam.org.

[3] J. Breslau, N. Ferraro, and S. Jardin, Phys. Plasmas **16**, 092503 (2009).

[4] For example, L. Chacón, Phys. Plasmas **15**, 056103 (2008); J.N. Shadid, R.P. Pawlowski, J.W. Banks, L. Chacón, P.T. Lin, and R.S. Tuminaro, J. Comput. Phys. **229**, 7649 (2010).

C.1.3 RF Heating and Current Drive

Core RF – Energetic Particle Simulation Needs

P. T. Bonoli[1], D. L. Green[2], J. C. Wright[1]
[1]Plasma Science and Fusion Center, MIT, Cambridge, MA 02139
[2]Oak Ridge National Laboratory, 1 Bethel Valley Rd, Oak Ridge, TN 37831

The basic motivation for this whitepaper can be found in the recent *Report on the Workshop on Integrated Simulations for Magnetic Fusion Energy Sciences*, where the interaction of fast particles with thermal plasma waves and instabilities was recognized as an area of new opportunity for integrated modeling and model hierarchy development [IS, 2015].

1. **Please specify the current science drivers for your field of research.**

Where we are today? Be sure to include broad impact, DOE interest, ties between experiment/theory, etc.
Fast magnetosonic wave in the ion cyclotron range of frequencies (ICRF) will be used for plasma heating in burning plasmas such as the ITER device and beyond. Furthermore, mode-converted ICRF waves have the potential for pressure profile control. Thus, the self-consistent interaction of ICRF waves with energetic particles due to neutral beam injection (NBI) heating, fast fusion alpha particles, or a minority ion population that is "self-generated" by the ICRF power is of great interest as it will impact these applications and is therefore the primary science driver for this whitepaper. Also of interest for this science driver is the generation of nonthermal electron distributions by waves in the lower hybrid range of frequencies (LHRF), as well as the use of the lower frequency electromagnetic polarization of the fast LH wave (the so-called "helicon"), both of which can be used for localized control of the plasma current profile. Typical model simulations for ICRF wave – energetic particle interactions employ a full-wave field solver coupled to a continuum Fokker Planck code [Jaeger, 2006] or a Monte Carlo orbit code [Choi, 2010], and LHRF wave – energetic particle interactions are described by the coupling of ray tracing or full-wave field solvers to continuum Fokker Planck codes [Harvey and McCoy, 1992; Wright et al., 2009; Shiraiwa, 2011]. The connections between the wave codes and particle codes are done through either a "diffusive" or "nondiffusive" RF operator [Harvey, 2005] or an RF "kick" operator. The dielectric response in the full-wave solver is evaluated using the nonthermal particle distribution. The wave solvers and particle codes are typically iterated in time. Comparisons with experiment are done using synthetic diagnostics, which make use of the simulated nonthermal distribution function to calculate diagnostic measurements of hard X-ray emission, photon counts from a neutral particle analyzer, fast ion D-alpha emission, and RF wave fields detected with reflectometry and phase contrast imaging techniques.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

What will probably be solved in the next 5-10 years? Why is this important to the field?
In the 2020–2025 timeframe, it is expected that a model hierarchy will be developed to describe the core ICRF wave – energetic ion and core LHRF–energetic electron interactions assuming the coupled power is known. These simulation capabilities would make it possible to determine to what extent ICRF power needed for bulk plasma heating in a burning plasma and how it will interact parasitically with energetic distributions of fast particles already present in the plasma such as fast ions from NBI and fusion alpha particles. Models will be developed to assess fast ion orbit width effects and nondiffusive velocity space effects, which can impact the loss of fast ions accelerated by the ICRF power. This simulation capability would also elucidate the importance of full-wave effects in lower hybrid wave propagation such as diffraction and focusing, thus establishing the regimes of validity for geometrical optics and ray tracing techniques.

The core and wall-clock hour requirements for generic coupled full-wave/Fokker Planck simulations in the ICRF and LHRF regimes are dominated by the 3D full-wave field reconstruction. In the ICRF, this implies, for example, ~ (5000 cores/toroidal mode) × 2 hours × (50 toroidal modes) × (10 iterations with a Fokker Planck solver or Monte Carlo code) = 5,000,000 CPU hours per run. About 10 runs per year are needed to facilitate synthetic diagnostic comparisons with fast ion diagnostics, yielding 50,000,000 CPU hours per year. In the helicon and LHRF regimes, typical 3D field reconstructions require ~ (15,000 cores/toroidal mode) × (20 toroidal modes) × (20 iterations with Fokker Planck solver) = 6,000,000 hours per run. About 10 runs per year are also needed to facilitate synthetic diagnostic comparisons with hard X-ray cameras and phase contrast imaging diagnostics, yielding a total MPP usage of 60,000,000 CPU hours. It is important to note that for the LHRF and ICRF regimes, each toroidal mode simulation is independent and takes about the same amount of compute time; thus, all modes can be executed concurrently with little penalty for tolling if, for example, 200,000–300,000 cores are available. Thus, the 3D field reconstruction is a problem that benefits enormously from capacity computing.

It is also expected that more efficient 3D solutions (two velocity space and one configuration space) of the Fokker Planck equation will be realized during the 2020–2025 time period. A challenge in this area is the need to develop efficient algorithms (either direct or iterative) for inverting the large sparse matrices produced by the 3D Fokker Planck solver.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

What might not necessarily be solved in the next 5–10 years? Again, what is the importance?
It is unlikely that within the 5–10 year timeframe it will be possible to simulate the self-consistent interaction of ICRF generated/accelerated fast ion distributions with models for energetic particle instability. This is an important outstanding issue for ICRF heating in burning plasmas since it is not known if ICRF waves will destabilize energetic particle instabilities. Also, it is unlikely that a simulation capability for sawtooth stabilization via ICRF-generated tails can be developed with existing computing ecosystems. Both of these problems are nearing the required theoretical formulations, i. e., closure schemes for the MHD equations that properly include nonthermal ion distributions. However, the numerical implementation of these closure schemes is at this time not well-developed.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?** *Suggested topics include the following.*

Accelerate	Why?
1. Improved algorithms and workflows for code couplings.	Primary approach in this area is to couple wave solvers and Fokker Planck solvers. Algorithms are needed to maintain/accelerate convergence between the codes as they are iterated in time.
2. Access to capacity computing resources.	The toroidal modes needed to reconstruct 3D ICRF and LHRF wave fields can all be done simultaneously since they are independent.
3. Development of new algorithms for evaluation of the plasma response.	Particle-based methods for simulating the plasma response [Green and Berry, 2014] make it possible to include 3D plasma geometry effects in the plasma response while taking advantage of emerging architectures such as GPU accelerators.

Impede	Why?
1. Lack of access to capacity computing resources.	See point #2 above.
2. Lack of scalable solvers on new architectures, such as GPU accelerators.	Most of the work in full-wave spectral solvers is dominated by the inversion of a dense matrix, where the work scales as (spectral resolution)^3.

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.**

The data ecosystems used to simulate and validate models for core RF wave – energetic particle interactions involve 3D nonthermal particle distributions (2 velocity space and one configuration space dimension), 4D RF diffusion operators (2 velocity and 2 configuration space), and experimental data from synthetic diagnostics (for example, time-dependent chord-integrated hard X-ray emission, photon counts, or scattered laser intensity or RF signal intensity).

The biggest challenge in comparisons of experimental data measurements of nonthermal particles distributions with synthetic diagnostic predictions is accurate incorporation of the diagnostic geometry, sensitivities, and etendue (for example).

6. **References** *(please keep the reference length to no more than 10 critical references)*

[Choi, 2010] M. Choi et al., "Iterated finite-orbit Monte Carlo simulations with full-wave fields for modeling tokamak ion cyclotron resonance frequency wave heating experiments," Physics of Plasmas **17**, 056102 (2010).

[Jaeger, 2006] E. F. Jaeger et al., "Self-consistent full-wave and Fokker-Planck calculations for ion cyclotron heating in non-Maxwellian plasmas," Physics of Plasmas **13**, 056101 (2006).

[Green, 2014] D. L. Green and L. Berry, "Iterative addition of parallel temperature effects to finite-difference simulation of radio-frequency wave propagation in plasmas," Computer Physics Communications **185** 736 (2014).

[Harvey, 1992] R. W. Harvey and M. McCoy, "The CQL3D Fokker Planck code," in Proceedings of the IAEA Technical Committee Meeting on Simulation and Modeling of Thermonuclear Plasmas, Montreal, Canada, 1992 (IAEA, Vienna, 1992), US DOC NTIS Document No. DE93002962.

[Harvey, 2009] R. W. Harvey et al., "Comparing Direct and QL Calculation of ICRF Diffusion Coefficients," Proc. of European Physical Soc. Mtg, Sofia, Bulgaria (2009).

[IS, 2015] Report of the Workshop on Integrated Simulations for Magnetic Fusion Energy Sciences, P. T. Bonoli, and L. Curfman-McInnes, pgs. 69–70 (http://science.energy.gov/~/media/fes/pdf/workshop-reports/2016/ISFusionWorkshopReport_11-12-2015.pdf).

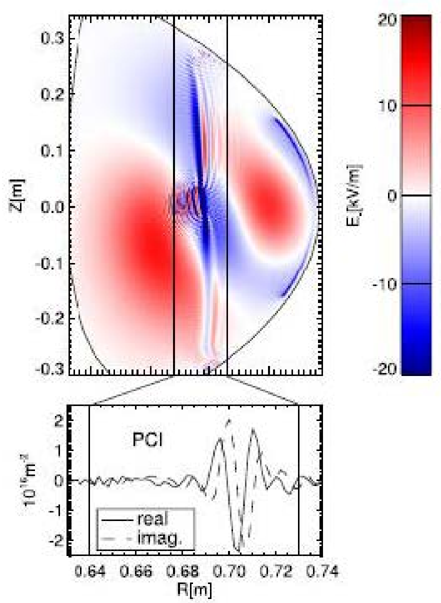
[Shiraiwa, 2011] S. Shiraiwa et al., "Full wave effects on the lower hybrid wave spectrum and driven current profile in tokamak plasmas," Physics of Plasmas **18**, 080705:1-4 (2011).

[Wright, 2009] J. C. Wright et al., "An assessment of full wave effects on the propagation and absorption of lower hybrid waves," Physics of Plasmas **16**, 072502 (2009).

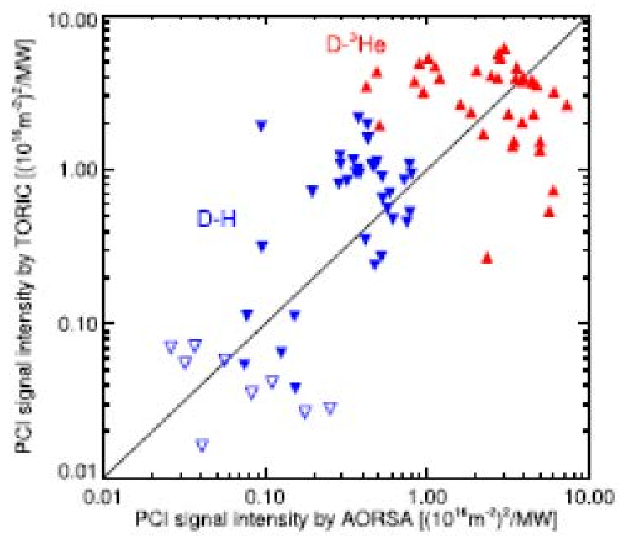
7. (Optional) Images

Consider submitting one or two already published high-resolution images suitable for inclusion in the report. Please provide the reference(s).

(a) Simulated ICRF wave fields used to reconstruct the intensity of mode converted ICRF waves detected by a Phase Contrast Imaging (PCI) diagnostic in the Alcator C-Mod tokamak. (b) Comparison of the simulated and measured mode converted ICRF wave intensity in Alcator C-Mod for varying hydrogen concentration in a deuterium plasma and varying Helium-3 concentration in a deuterium plasma. Reproduced from Figs. 2 and 11 of T. Tsujii et al., Physics of Plasmas **22**, 082502 (2015).



The left-hand circularly polarized component of the electric field simulated by the AORSA field solver (top) and the simulated PCI signal (bottom) in a D-(He-3) plasma.



Comparison of the PCI signal intensity simulated by AORSA and TORIC field solvers. Red upward triangles: D-(He-3) plasmas. Blue downward solid triangles: D-(H) plasmas. Blue downward open triangles: H minority heating.

Requirements for RF Antenna-to-Core Simulation

D.L. Green[1] and P.T. Bonoli[2]
[1] Oak Ridge National Laboratory, 1 Bethel Valley Rd, Oak Ridge, TN 37831
[2] Plasma Science and Fusion Center, MIT, Cambridge, MA 02139

1. Please specify the current science drivers for your field of research.

Our primary science driver is understanding how the details of the ITER edge plasma will impact the required performance and reliability of the RF heating and current drive systems. This means extending the present state of the art to resolving these edge plasma/plasma-facing component details at quantitative fidelities and dimensionality.

Historically, simulating RF heating and current drive in fusion devices has been broken down into tractable pieces appropriate to the computational tools available. This has meant (i) 2-D linear, time-harmonic, kinetic core plasma calculation; (ii) a nonlinear, time-dependent, dielectric based edge plasma calculation; and (iii) a quasilinear, slow-time varying diffusion calculation that forms the background of the kinetic calculation in (i). The linear core calculation is typically formulated as a double complex dense matrix factorization, and even with out-of-core type portings to GPU based heterogeneous architectures, which yield factors of two or more reductions in wall-clock time [D'Azevedo 2012], the work required to invert such systems scales is N^3. As with any dense factor, eventually the strong-scaling curve rolls over due to lack of work per node and the ensuing communication overhead. <u>The limitation here is not being able to address the problems of interest</u> (i.e., high fidelity in 3-D), due to large N, stemming from how the problem is formulated. As such, we have been investigating alternative formulations that do not rely on a dense matrix factor, but instead weigh additional computational cost per N against the scaling properties for large N. For reference, the 2-D problem size of interest is about 512 × 1024, which gives a 36-TB matrix, with an approximate 0.5 million CPU hours to factor; moving to 3-D at a (very) modest 256 × 512 × 128 domain resolution, given the same approach, yields a 36-petabyte matrix and more than 10^3 million CPU hours to factor; obviously we need an approach that scales better with problem size.

The dielectric based edge plasma calculation of (ii) needs to conform to the geometry of the device wall at mm resolutions, and include the physics of RF sheath formation at the plasma-material interface. This is typically approached via Finite-Difference Time-Domain (FDTD, e.g., Jenkins 2015) or finite-element frequency-domain (FEFD, e.g., Wright 2015) methods with a nonlinear sheath boundary condition to avoid resolving the Debye length scales of the plasma-material interface. The FDTD and FEFD have quite different computational needs (i.e., flops versus large memory requirements respectively), as well as ill-conditioned sparse matrix inversion for the FEFD approach.

2. Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.

We expect that in the 2020–2025 time frame, leadership class computing platforms will enable simulation of the ITER ion-cyclotron RF heating system in 3-D, accounting for all the edge plasma nonlinearities and RF sheath-related impurity production, plus resulting density modification and decay into other frequencies.

In addition to the device wall/antenna structure details, future work is likely to focus on resolving edge density and temperature variations at both large (e.g., blobs) and small (e.g., turbulent fluctuations) levels, to investigate the impact on robust coupling of power, and effectiveness of current drive at all frequencies (EC, LH, and IC). We also expect that calculations to determine how the application of RF power modifies the background density profiles of the edge plasma will become standard practice.

Need: Geometry/dispersion adapting meshing & domain-decomposition technologies

Meshing methods that adapt to the wavelengths present in the linear core kinetic plasma response, such that we do not oversample the problem and waste computing resources, which in turn, limits problem size; this is a problem with present dense matrix formulations that assume complete non-locality in the plasma response, whereas the locality is far from complete in reality.

Need: Large, accelerated, ill-conditioned sparse-matrix inversion libraries/preconditioners

Given the development of new 3-D linear kinetic algorithms that scale better than N^3 (e.g., Green 2014), accelerator versions of ODE integrators are particularly needed to support these efforts, beyond versions that run from the CPU (i.e., accelerate just the ODE); instead, they would need to be something useable as a subroutine in a custom accelerator kernel. In addition, efficient sparse matrix factoring methods would be necessary for large $O(10^9$ to $10^{12})$ degrees of freedom.

Need: Large dataset exploration/debugging visualization tools (and storage)

In the past we have noted that data exploration at scale can consume exorbitant man hours, for example, when trying to identify the originating location of an instability that only shows up in at-scale runs. Visualization tools exist, but these are focused on presenting an at-scale result, rather than streamlining exploration of a large dataset. The capability to move through a dataset in time and space with the visualization package intelligently choosing what to render such that it can be done in real time for distributed datasets would be invaluable to both debugging at-scale simulation and discovering the physics of interest.

3. **Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.**

Although a solution to the complete 6-D + time Maxwell-Vlasov system of equations for the full problem domain and steady state timescale is beyond even exascale resources (particle or continuum based), there is certainly opportunity to advance to a small subset of the problem to this level (i.e., that part of the edge plasma where the decay of power into kinetic waves at frequencies other than the antenna frequency). Such a 6-D + time solver, if using the particle-in-cell method, would require an intrusive uncertainty quantification implementation to ensure the simulation results are not overwhelmed by particle noise. Alternatively, it may also be possible to extend the FDTD method, which recovers multiple frequencies, to include linear kinetic response by a kinetic-J-like extension (Green 2014). This would numerically calculate the kinetic dielectric at each point in space and some subset of time steps; this would require considerably more resources than presently used by cold-plasma dielectric FDTD algorithms, which already consume a large fraction of the leadership-class machines.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Geometry/dispersion adapting, meshing and domain-decomposition technologies	Uniform resolution makes N so large as to preclude solving the problems of interest. At-scale meshing and domain decomposition libraries would mitigate this.
2. Large, accelerated, ill-conditioned sparse-matrix inversion libraries/preconditioners	RF produces large ill-conditioned (sparse and dense) matrices. Accelerated, production level, available by default on leadership machines is needed; at present this is custom.
3. Large dataset exploration/debugging visualization tools (and storage)	Bug tracking at-scale. Real discovery via high-fidelity datasets produced via simulation, and storage of those datasets. Often simply too large to move back to a local machine.

Impede	Why?
1.	
2.	
3.	

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.**

6. **References**

[D''Azevedo 2012] E. D'Azevedo et al., Proc. Comp. Sci., Vol. 9, pg. 67–75, 2012; doi:10.1016/j.procs.2012.04.008.
[Green 2014] D.L. Green et al., Comp. Phys. Comm. 185(3), 2014; doi:10.1016/j.cpc.2013.10.032.
[Jenkins 2015] T.G. Jenkins et al., Proc. of 21st Topical Conference on Radiofrequency Power in Plasmas, IO6, 2015.
[Wright 2015] J.C. Wright et al., Proc. of 21st Topical Conference on Radiofrequency Power in Plasmas, A42, 2015.

Requirements to Study RF Plasma-Material-Interactions

D.L. Green[1] and P.T. Bonoli[2]
[1] Oak Ridge National Laboratory, 1 Bethel Valley Rd, Oak Ridge, TN 37831
[2] Plasma Science and Fusion Center, MIT, Cambridge, MA 02139

1. **Please specify the current science drivers for your field of research.**

The success of magnetically confined nuclear fusion as an economically attractive source of power largely depends on the success of ITER, the next-step device currently under construction in France. ITER in turn relies on the successful operation of three plasma heating technologies. Of those, two are based on external application of radiofrequency power, and only ion-cyclotron resonant heating (ICRH) can directly heat ions. However, operating the ICRH system on devices available at present has been observed to correlate with the production of impurity ions from increased plasma–material interactions, which can have deleterious effects like collapsing the plasma. The basic physics mechanisms of how the application of ICRF power enhances the electric potential that exists between the plasma and any confining material structure (the sheath) are thought to be understood, as are the basic physics of how materials respond to the bombardment of ions accelerated by that sheath potential. However, implementing these understandings in predictive computational models of the required fidelity to be validated with experiments is only now becoming possible. Therefore, the present state of the art is the coupling/incorporating models of plasma-wave propagation and heating, with high-fidelity representations of the launching antenna structures and confining material walls, with models of the surface response to plasma bombardments, and the subsequent longer timescale models of how sputtered impurities transport through the edge plasma and ultimately affect the core performance. The driving science objective here is being able to predict how RF power produces impurities, in order to suggest strategies to mitigate that production on ITER.

2. **Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.**

Although the theory of plasma-wave enhancement of sheath potential is well developed [D'Ippolito 2013], application of these theories at the required high fidelity and on at-scale computing platforms to represent the Tokamak edge structures is not. We expect that in the 2020–2025 timeline, high-fidelity (3-D) production simulations of the RF sheath will be employed to complete our understanding of how the sheath potential is configured for various design and operational choices of the RF systems. These calculations are either of finite-difference or finite-element in time- or frequency-domains, at the order of 10^9 (~1 TB) to 10^{12} (~230 TB) grid cells in order to represent the immediate area around the ICRH antenna and the larger area of the entire vacuum vessel interior, respectively. For the finite-element method approach, iterative sparse matrix solvers at this scale are an active area of development, as are preconditioners to aid that iterative process. Alternatively, the time-domain methods avoid these at-scale matrix inversion issues, but must advance many time steps, while obeying the CFL stability constraints (so time steps of the order 10^{-12} s), to reach the desired steady state response (tens of RF cycles: so 1^{-6} s, or 10^6 time steps).

Need: An identifiable need here are sparse matrix solvers for the ill-conditioned matrices produced by the finite-element approach for O(10^{12}) degrees of freedom.

Although the sheath is a kinetic process, the recent development [Kohno 2013] of a reduced model for application to dielectric plasma full-wave solvers will enable this high-fidelity investigation. We expect this capability, combined with at-scale compute resources, to enable the study of issues such as the far-field versus near-field sheath issues (i.e., are those sheaths that form on material surfaces due to waves leaving the plasma important relative to those sheaths directly connected [magnetically] to the near field of the driving ICRH antenna?). Such a simulation requires the much larger domain sizes of realistic geometry, something we are only starting to do now with INCITE/ALCC-level allocations.

Another issue we expect to be resolved is the validation of the nonlinear sheath boundary condition (a reduced model) via fully kinetic simulation (e.g., PIC). This will require kinetic simulation that also resolves the plasma-material interface in 3-D, something that has yet to be done. This level of dimensionality and fidelity is required not only for a complete verification of the reduced model, but also for direct comparison with new diagnostic capabilities where the diagnostic results are sensitive to the geometric details of the plasma-material interface.

In addition, we expect a tight coupling to a hierarchy of models representing the sputtering response of the material to the bombardment of ions caused by the sheath potential. For example, utilizing an RF-calculated sheath potential as input to a coupled erosion/re-deposition/sputtering simulation (e.g., ERO [Kirschner 2000] + TRIM [Biersack 1984]), which would in turn inform the RF-induced impurity flux boundary conditions to edge fluid models like EDGE2D [Simonini 1994], SOLPS [Schneider 2006], and UEDGE [Ronglien 1998].

Need: Extension of present "loose-coupling" HPC coupling frameworks (e.g., the IPS), to allow tight(er) coupling.

3. **Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.**

The full nonlinear, 6-D plus time-domain kinetic problem (for ions and electrons)—as well as running some type of materials code to inform that impurity production self-consistently—that would track additional species produced from the wall (i.e., the impact of RF on impurity transport back across the separatrix). This would essentially be the type of simulation required to simulate fusion plasma physics in its entirety—no gyro-averaging (although perhaps on electrons), and extending beyond MHD timescales to the transport timescale. This means an additional dimension to the full-f PIC codes being run today, and we would also need kinetic electrons, perhaps not Lorentz orbits, but certainly parallel kinetics (i.e., electron Landau damping). Unfortunately, long-timescale simulations with present methods are not applicable; PIC gives particle noise, and the time step is *very* small relative to the transport timescales, on top of the additional grid size stability limitation of the Debye length (the dielectric based solvers do not have this), so we are stuck with coupling to fluid transport solvers to investigate the impact of impurities produced by RF, unless implicit PIC solvers enable resolution of RF timescale physics (as we cannot simply step over RF timescales), while still allowing stepping to transport timescales.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. At-scale sparse matrix inversion methods for ill-conditioned matrices.	Pre-conditioners for RF problems are difficult to construct in a robust manner, and direct solves push memory limitations.
2. Tighter coupling HPC framework.	Plasma physics and material physics really are separate areas of study, so it is unlikely that a single code/theory is going to be implemented. This means coupling.

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.**

N/A.

6. **References**

[Kohno 2013] H. Kohno, J.R. Myra, and D.A. D'Ippolito, "Radio-frequency sheath-plasma interactions with magnetic field tangency points along the sheath surface," Physics of Plasmas, **20** p. 082514 (2013).
[D'Ippolito 2013] D.A. D'Ippolito et al., "Modeling far-field radiofrequency sheaths in Alcator C-Mod," Plasma Phys. Control. Fusion **55**, 085001 (2013).
[Kirschner 2000] A. Kirschner et al., Nucl. Fusion, 40, 989, 2000.
[Biersack 1984] J.P. Biersack et al., Appl. Phys., A, 34, 73, 1984.
[Simonini 1994] R. Simonini et al., Contributions to Plasma Physics, 34 368–373, 1994.
[Schneider 2006] R. Schneider et al., Contributions to Plasma Physics, 46 (1-2) 3–191, 2006. **[Ronglien 1998]** T.D. Rognlien et al., Contr. Plasma Phys., 38, 152, 1998.

Requirements for Rigorous RF Validation Workflow Needs

D.L. Green[1], P.T. Bonoli[2]
[1] Oak Ridge National Laboratory, 1 Bethel Valley Rd, Oak Ridge, TN 37831
[2] Plasma Science and Fusion Center, MIT, Cambridge, MA 02139

1. **Please specify the current science drivers for your field of research.**

A stated top-level goal for DOE's OFES [Koepke 2014] is the use of massively parallel computing for validated predictive simulation for magnetically confined fusion plasmas. This capability should ultimately enable, and minimize the risk in, future fusion energy development stages. A subset of this goal is the integration of independently developed computational tools that makeup the DOE portfolio of legacy and state-of-the-art simulation codes. As such, the science driver addressed in this white paper is the creation of a validated model hierarchy. Specific examples for RF are validated single-component models (e.g., full-wave codes), with further validation of coupled simulations, (e.g., full-wave + Fokker-Planck iterations — see white paper to this panel by Bonoli et al.); however, the scope should ultimately be expanded to validation of fully integrated/whole-device-model simulations (e.g., see white paper to this panel by Poli et al.). While by necessity, and history, the model hierarchy encompasses models of all compute-scale sizes and physics fidelity, the need for verification and validation of these models is universal. Identification of where models are valid for prediction, and where they are not, thereby elucidates where further model development is required. To support the hierarchy of models being developed within fusion, and to ensure their continued availability and integrability into the larger goal of a whole device model, a rigorous validation program should be supported and encouraged as one aspect of the future DOE computing ecosystem. Here we suggest some properties of the upcoming extant computing ecosystem that will aid in this effort.

2. **Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.**

At present, rigorous validation efforts are rare. Codes are typically validated by the developer for a limited experimental data point. In recent community workshops, the idea of "analysts" has been gaining traction (e.g., [Holland 2015], [White 2015], [Green 2015]), whereby some community standards and dedicated personnel are assigned to applying available computational models to produce predictions that can be tested over a very large database of experimental observations. We expect that this idea will become far more prevalent within the 2020–2025 time frame, and that the computing needs for these rigorous validation efforts will need to be in place. In addition to the rigorous validation of the model hierarchy, is the development of the models themselves, which in this time frame is likely to consist of the construction of reduced models from the analysis of datasets produced by extreme-scale simulations and the subsequent validation of the reduced model with the extreme-scale simulation.

Need: Increased emphasis on "capacity" computing — The rigorous validation effort will by necessity mean executing many simulations each with only slightly different input parameters. Mapping parameter spaces, or running at all available experimental conditions, are likely workflows, which means capacity, rather than capability, use of the facilities being discussed. We suggest that an increased emphasis on the validity of utilizing the extant compute ecosystem in a "capacity" mode will aid in enabling the rigorous validation needs of fusion science. This emphasis may take the form of modifications to the allocation call for proposals, and more favorable queue/scheduling policies and capabilities (e.g., submission of jobs to part of a node).

3. **Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.**

While a rigorous validation effort may be put in place and executed, therefore telling us where any given model is valid for predicting experiment, it will not necessarily lead to filling the gaps where those models are identified as being invalid. For example, the overlap/combination of MHD and gyrokinetic models is not merely a computing-scale issue but rather a fundamental theory issue which will require development of new theoretical implementations. However, it should be pointed out that we think it important to develop a new generation of computational physicists who are trained in both the use of leadership class computing and who are familiar enough with the physics issues to see what new avenues for solving such problems may arise with the availability of an extant computing ecosystem. For example, this may take the form of postdoctoral science positions at the user facilities.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Continued/increased emphasis on, and ability to do "capacity" computing.	Validation workflows require the execution of many, slightly different input parameter variants of any given code. This should be supported by the allocation proposal process.
2. Access to remote databases from the compute nodes, or local data repositories where institutional privacy/access policies hold.	Institutional/experimental device data access policies must be worked around to enable HPC resources to be brought to bear on the validation problem.
3. Smart data/result archiving.	Storing validation results with enough metadata to know which need recomputing if, for example, a calibration error was found in a diagnostic.

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.**

A rigorous validation effort will mean automated access to databases of processed experimental results. The data will need to be obtained from the full range of U.S. experimental devices, including relevant validation platform devices at various universities. However, institutional policies likely mean that these experimental databases reside at the experimental facilities. Manual copying of the appropriate experimental data, or simply ignoring the experimental data and just comparing it a posteriori, is unsuitable for a vigorous effort.

Need: Either remote data access or local data stores with privacy — Either direct connections to remote experimental databases accessible from the compute nodes, or repositories at the compute facility where experimental facilities can store their data with their own privacy and access policy implemented. For present facilities, these data are low in volume and would not vary significantly, although this is likely to change with ITER.

Need: Store results with metadata — For validation studies, we envision the need to store the run results with enough metadata that if, for example, the calibration of an experimental diagnostic changes, we would know which of the validation runs needs to be redone. This issue is dealt with more thoroughly in the white paper by Wright et al. submitted to this panel.

6. References

[Green_2015] D. L. Green, et al., 2015, "Next Steps in Whole Device Modeling," White Paper submitted to the DOE Workshop on Integrated Simulations for Magnetic Fusion Energy Sciences, https://burningplasma.org/resources/ref/Workshops2015/IS/C_green1_d.pdf.

[Holland_2015] C. Holland, et al., 2015, "A National Validation Initiative for Guiding Predictive Model Development," White Paper submitted to the DOE Workshop on Integrated Simulations for Magnetic Fusion Energy Sciences, https://burningplasma.org/resources/ref/Workshops2015/IS/CABDEFG_holland_c.pdf.

[Koepke_2014] M. Koepke, et al., 2014, "Report on Strategic Planning December 2014 U.S. Department of Energy Office of Science Priorities Assessment and Budget Scenarios," https://www.burningplasma.org/resources/ref/fspp/FINAL_FESAC_123114-1.pdf.

[White_2015] A. White, et al., 2014, "A New Research Initiative for 'Validation Teams'," White Paper submitted to the DOE FESAC Strategic Planning Panel for Status and Priorities, and Initiatives, https://www.burningplasma.org/resources/ref/fspp/whitepapers/White-Paper-Validation-vf.pdf.

Capability Computations of RF Antenna Wave Propagation with the VORPAL Framework

J.R. King, J.R. Cary, T.G. Jenkins, D.N. Smithe

Current science drivers

State-of-the-art time-domain simulation of RF heating in tokamaks has an increasing focus on the surrounding fields and plasma physics of their complex antenna launchers. These simulations are able to model the non-axisymmetric nature of sheath potentials surrounding Faraday shields, antenna boxes, and limiters, for the purpose of determining potential sputtering sites. As the simulation domain is expanded to encompass the confined plasma, the efficacy of RF current drive and heating for a given antenna may be computed. Advances in RF heating and current drive expand the engineering design options for a non-inductive (steady-state) burning tokamak.

Science challenges for the 2020–2025 time frame

Progress on RF antenna modeling will be made in the next 5–10 years, even with current resources and extant computing systems. However, there are limitations to this progress as present-day computations are nearing the full resources of the Titan supercomputer. To illustrate some of the challenges with current systems, we describe three science challenges that are difficult, if not impossible, to simulate without enhanced computational capabilities.

The first of these is helicon antennas, as planned for installation on the DIII-D tokamak. The frequencies involved are an order of magnitude larger than strap-antenna systems such as those used on C-Mod and NSTX. Thus, the RF wavelengths are an order of magnitude smaller; higher spatial resolution such as would be available on a next-generation computing platform would be very beneficial.

The second is simulations that include the full tokamak poloidal (and potentially toroidal) geometry. Presently, absorbing boundary conditions are used with domains that only contain a partial poloidal cross-section. Larger domains will enable higher-fidelity predictions of the RF power deposition and could include the full scrape-off layer and divertor region.

The third application is RF simulation for the ITER tokamak. ITER is an international tokamak under construction in France that is planned to demonstrate the feasibility of fusion power ignition and power generation with the tokamak concept (the fusion power generated is expected to be a factor of 10 larger than the input power). The ITER volume will be approximately an order of magnitude larger than that of current devices; however, the RFs and wavelengths will be comparable to those of C-Mod or NSTX. This larger disparity of spatial scales requires greater computational resources than current simulations do, and this is compounded by the need to simulate the full poloidal cross section of the domain. Present-day simulations are performed at unsatisfactory resolution and full simulation is thus an exascale problem.

Top computing ecosystem aspects to accelerate or impede progress

The VORPAL framework is well positioned to take advantage of the raw computational power of advanced many-core architectures (e.g., GPU or Intel MIC). In addition to leveraging existing methods to avoid write conflicts that commonly arise in PIC computing, it is being developed to go much further to assure use of all compute capabilities available to the computation, and to develop code in a manner that is economical and maintainable. To develop code that assures use of all compute capabilities, runtime discovery software is employed that queries the system for the available devices and their compute capabilities and a flexible decomposition is used that allows load balancing on heterogeneous architectures. As such, we expect progress will be quickly enhanced with access to larger and faster machines.

Current scaling studies (Figure 1 from Ref [1]) show that VORPAL scales well on the Titan XK7 supercomputer. Other aspects that impede scaling, such as I/O and visualization, are significant, and while these have been tractable with existing solutions (e.g., parallel HDF5 and the VisIt software), it is unknown how they will affect scaling on next-generation architectures. For example, with higher resolutions, it is possible that in situ analysis during the run will be required to limit the size of the output. In order to best make use of large machines, resources should be dedicated to these software development tasks.

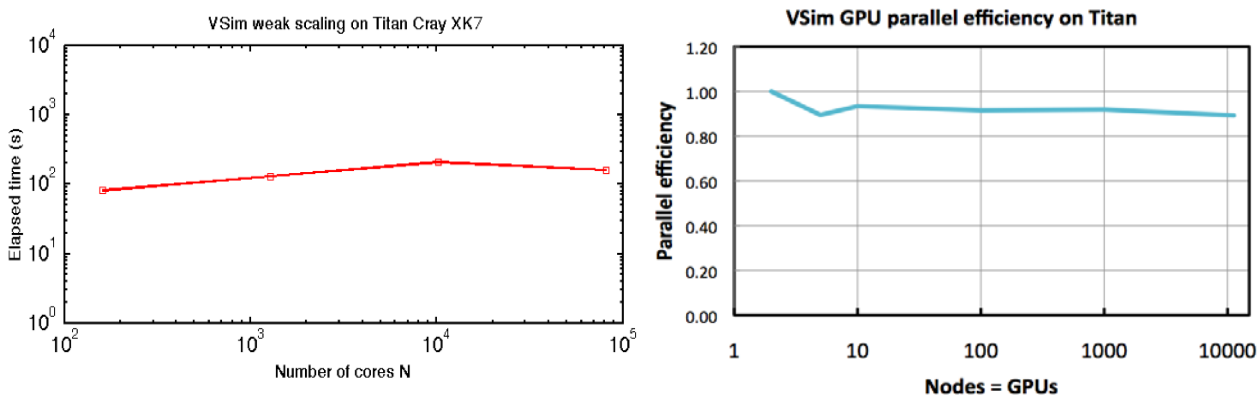


Figure 1: Weak parallel CPU (left) and GPU (right) scaling on the Titan XK7 at Oak Ridge National Laboratory. For the GPU plot, parallel efficiency is defined as the ratio of elapsed wall clock time for M simulation time steps on 3 GPUs divided by the elapsed wall clock time for M simulation time steps on N GPUs.

References

[1] T.G. Jenkins and D.N. Smithe, High-Performance Finite-Difference Time-Domain Simulations of C-Mod and ITER RF Antennas, in *Proceedings of the 21st Topical Conference on RF Power in Plasmas*, Lake Arrowhead, California, 2015.

Optional Published Figures

Figures attached from Ref [1].

Integration of RF Models in Integrated Simulations: Hardware and Software Needs

F.M. Poli, Princeton Plasma Physics Laboratory
P.T. Bonoli, Plasma Science and Fusion Center, MIT

The material presented in this document has been reproduced from the following white papers submitted to the FES Workshop on Integrated Modeling:

- S.J. Wukitch, "RF Sustainment Simulation Opportunities for Steady State Fusion Reactor Plasmas"
- R.R. Parker, G.M. Wallace, and S. Shiraiwa, "Whole Device Modeling with Novel Radio-frequency Actuator Schemes in Steady-State Reactor Designs"
- D.L. Green and J.M. Park, "The Role of HPC & First-Principles Simulation in Whole-Device-Modeling"
- D. Green, D. Batchelor, J.M. Canik, W.R. Elwasif, D.E. Bernholdt, N. Bertelli, C. Holland, and J.M. Park, "Next Steps in Whole Device Modeling"
- N. Bertelli, D. Green, C.K. Phillips, E.J. Valeo, and J.C. Wright, "The Role of RF Source Components in a Whole Device Model"

1. Please specify the current science drivers for your field of research.

RF (radiofrequency) actuators have long been recognized as essential tools for realizing a steady-state tokamak. The proper design of reactor-grade steady-state tokamaks involves coping with a complex interplay of the effects of transport, external CD and heating profiles, MHD stability, and control of edge pedestals and SOL parameters. While great strides have been made in developing modeling capability for most critical areas, very little progress has been made in modeling the whole device; that is, integrating the advances that have been made in transport, core and edge MHD, RF current drive, and scrape-off layer simulations in order to determine optimal reactor configurations and operating scenarios. For many years, RF source modules have been employed as components within integrated simulation (e.g., within TRANSP), and the RF SciDAC program has produced both first-principles and reduced models for many aspects of simulating the application of RF power, with the goal of a predictive and robust tool for the coupled antenna-to-core system within reach. Considerable progress has been made on the validation of these RF source modules with the experiments.

2. Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.

At present, the core wave physics is well described and simulated, but the integration with antenna interaction and RF power propagation in the scrape-off layer is less well developed. The emergence of open source FEM codes, combined with core simulation solvers and time domain simulations of antennas with detailed 3D geometry, have shown significant promise. The next challenge is to model interactions between RF and energetic particles (White Paper by Bonoli) and to include an accurate model of the scrape-off layer and PMI (White Papers by D. Green, J. King, and T. Jenkins); these should be given high priority.

Rigorous validation and uncertainty quantification (UQ) — necessary for a robust and predictive whole device model — require that available models be robust enough to test for some range of parameters, and a statistical ensemble of selected experimental data to be compared with. The use of **a common framework to facilitate modularity and extensibility** is advised; for any integrated or whole device model to be useful for a range of use cases, it has to be modular, that is, be able to interchange models of varying physics, fidelity, or compute capacity without doing any coding. A common framework would be flexible to allow connections to experimental databases for model validation. This requires the availability of storage space. However, it should be noted that loose coupling has its own limitations, and the interactions between components and the convergence of calculations need to be assessed (see White Papers by D. Green, J. Wright, and P. Bonoli).

De-institutionalizing the integration effort by setting standards will facilitate contribution and progress. Perhaps one metric of the next round of SciDACs could be the production of reduced models based on their large compute-scale investigations, or delivering codes that are "component-available" within the community framework, for the reasons of facilitating comparisons with reduced models, or benchmarking. A stepwise approach would (1) emphasize the delivery of reduced models derived from HPC discovery efforts; (2) emphasize HPC code availability for benchmarking and validation via inclusion in an integration framework; and (3) focus the utilization of ASCR tools and expertise in any future call for proposals targeting an eventual whole-device-model capability beyond that of just HPC performance and massive-scale computing. Such measures will facilitate progress toward a useful whole-device-model, as well as HPC code validation, and community-wide benchmarking.

While the RF source modules are intrinsically crucial to advance the whole device predictive capability, several integrations at different levels of accuracy with other specific topical areas, are still necessary to seriously move toward a whole device modeling. A few examples are the following: integration with turbulence codes to understand RF driven flows/currents and their impact on transport barriers and pedestal modification, plasma rotation, and impurity transport (in the core and edge).

These challenges are within reach in the next 5–10 years, provided increased accessibility to large capacity computing.

3. **Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.**

We envisage medium-fidelity time-dependent simulations with reduced models for core transport and the pedestal, and actuators are now within reach, with the limiting factor being the availability of robust, medium-fidelity models for all components of a WDM, for example, models for the edge, engineering components, and neutrons. A critical area is erosion and material lifetime. Simulation capability to assess impurity penetration would be very beneficial for evaluating different antenna concepts, magnetic geometry, and plasma confinement conditions.

Ideally, we should get to a point where a time-dependent simulation can treat all the physics problems at several levels of sophistication. Depending on the required physics, a hierarchy of fidelity exists to choose from within the RF suite. For example, the impact of the RF-driven sheath potentials at the plasma-material interface requires a 3D simulation that resolves the entire launching structure and at least some part of the plasma at appropriate resolutions (meaning multiple teraByte-level simulations). These types of problems, although not ready (or appropriate) for a whole device model, are being coupled with high-fidelity simulations of core plasma as the core and SOL responses are coupled, for example VORPAL and AORSA.

The availability of resources to undertake such challenging simulations over a large range of plasma and antenna parameters would allow the construction of a lookup table that — combined with neural networks — can be used to develop reduced models for WDMs.

Thus, although an all-inclusive simulation from the core to the wall will not be possible in the 2020–2025 time frame using high-fidelity models, it can be imagined that the same computing capabilities would allow parametric scans from these high-fidelity models, which in turn can be valuable toward the development of a complete, lower-fidelity simulation.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Capacity computing	Allow parameter scans to build up a lookup table and reduced models.

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.**

- Experimental database for model validation and benchmarking would require (a) access to selected, analyzed data, stored in a standard format and available to users, (b) a workflow to access, read, and use the data (like OMFIT).
- Model validation would require a workflow manager (like OMFIT) to access models.

C.1.4 Whole Device Modeling

Can We Use Exascale?

Stephane Ethier, Princeton Plasma Physics Laboratory

It is relatively straightforward to come up with an *exascale-worthy* problem in magnetic confinement fusion research. The physics of importance in a tokamak covers more than 10 orders of magnitude in length scale, from the size of an electron gyro radius all the way to the size of the whole device, and 15 orders of magnitude in timescale, from the fast electron gyro period to the time of the whole discharge. However, it is extremely difficult to devise an algorithm capable of numerically simulating a substantial part of a tokamak discharge in an efficient way on more than a billion computational threads, which is what exascale is on track to deliver. A *billion* computational threads! All needing to be fed data at a high enough rate to keep the floating-point units busy at all times. That's almost impossible with our current *bulk synchronous* codes, which make use of frequent, and often global, blocking communications and have several synchronization points at each time step or iteration (several stages that depend on each other and need to proceed in a given order). Exascale, we learn, will require asynchronous calculations to be efficient. We need to think in terms of well-defined computational tasks that can be executed in any order and without the need to synchronize.

However, a tremendous number of these tasks have to be ready for processing in order to keep the compute cores busy. We need to think *big*! Very *big*! And we need to think about overlapping—overlapping communication and computation; overlapping I/O—data analysis; and visualization as part of the running calculation. The latter two have traditionally been considered a distinct part of the main simulation or a separate post-processing stage. For exascale, they will all have to be considered essential tasks of the simulation work flow that can be scheduled at appropriate times to maintain the computational intensity (and efficiency). A concrete example could involve a large-scale particle-in-cell (PIC) code, where the number of particles exceeds the number of grid points by at least 1000 to 1 in a full distribution function calculation. It is clear that the particle data set contains a lot more parallelism and potential computations than the grid data set. During the simulation stages involving the particles, all the computational threads are busy, while not all of them are needed when the calculation enters grid-only stages. This is then a perfect time to carry out data analysis tasks over the particles at the same time as the main grid-based task.

Of the two types of scales mentioned above, the length scale is easier to deal with, although it is not *easy* by any means. Domain decomposition can take care of very large meshes and be fairly efficient as long as few global synchronization events are required. Most, if not all, grid-based algorithms use domain decomposition for dividing computations between processors, and there is extensive research on that topic. The tough problem, truly, lies in the incredibly wide range of timescales. Even at the exascale, one cannot hope to simulate a whole tokamak discharge while using a time step small enough to resolve the fastest timescale in the system. However, the advances in computer hardware and the availability of increasingly large supercomputers have allowed us to carry out ever longer simulations of ever bigger systems during the past 20 years or so. More emphasis has been put on the *bigger systems* though, essentially scaling the size of the numerical problem to be solved proportionally to the size of the supercomputer (also called *weak scaling*). This helps maintain the efficiency of the computation as more processors are being used but the overall time-to-solution remains (about) the same. In the case of a grid-based calculation in which the mesh resolution is increased, the Courant condition on the time step may end up increasing the time-to-solution by forcing the use of a smaller step. At the exascale, the effort has to be placed on reducing that time-to-solution for a large but fixed problem size. This is called *strong scaling*, and it is the key for tackling wider timescales and including more physics in simulations that will not have to run for months.

A concrete example of an exascale-worthy problem in tokamak physics is the pursuit of kinetic-MHD simulations, for which large-scale MHD physics is included as part of a fully global gyrokinetic calculation that includes core, edge, and device wall. This allows for the study of turbulence-driven instabilities and zonal flows along with MHD waves, profile changes, finite-beta effects, and so on. The preferred numerical approach, in my view, would be particle-based because it is easier to scale to a large number of threads and easier to cast as *asynchronous computational tasks*. Moving the particles is fairly straightforward once the fields at the particles' positions are known. Some PIC codes, such as VPIC for example, have already carried out simulations using a trillion particles with very high scalability and efficiency. The main difficulty is in the evaluation of the fields, which has traditionally been done using various grid-based solvers of the spectral or finite-element types (or even finite difference). The elliptic Poisson equation to be solved in gyrokinetic is global by nature and thus a problem when going to exascale. It may very well be that a more hierarchical numerical approach, such as the Fast Multipole Method (FMM), will end up being faster at very large scale. This is an active field of research that still needs to prove its worth though. Solving the nonlinear collision operator for the particles is also a difficult task that will require some optimization at the exascale.

I am confident that the fusion community will be able to use an exascale computer when it becomes available. However, much work will have to be done to bring our chosen applications up to the task. More urgent for the near future is the development of an efficient and scalable whole-device model, which can only be done at the moment by coupling various numerical components operating at different timescales. Several projects are already tackling this daunting task, and the whole community should participate in one way or another. There is a lot of work to be done.

Finally, in this era of Big Data we should not forget about the large amount of experimental data that have been accumulated during the past 50 years and continue to accumulate at an accelerated rate from the current experiments. The innovative statistical methods and machine learning algorithms that have recently been developed by the Big Data race can be used on these data to help us gain new insights in the physics at play in a tokamak discharge. This avenue has been largely untapped, and it could lead, for example, to better empirical models for use in a whole-device simulation model.

First-Principles Whole Device Modeling of Fusion Plasma on Extreme Scale HPCs

R. Hager, Princeton Plasma Physics Laboratory

1. **Please specify the current science drivers for your field of research.**

The hot plasma in a tokamak magnetic fusion device evolves in self-organization among many multi-physics phenomena in velocity, configuration, and time spaces. Whole device modeling will need to integrate the multi-physics. Two directions of research are being pursued in the fusion community: one is based on the coupling of many simple unit physics that are solved in the assumed scale-separated domains using reduced models, and the other is based on the fundamental first-principles kinetic equations without the scale-separation assumptions [1]. These two approaches have their own advantages in that the reduced-model WDM approach can take advantage of smaller computers for an experimental time-scale simulation and makes the parameter scan easier, while the first-principles approach requires leadership-class HPCs of today for a turbulence time-scale study and the exascale HPCs for experimental time-scale study; and that the first-principles approach yields more complete self-organization physics at all hierarchical levels, while the reduced-model approach will have to assume the scale separation and requires the lower hierarchical closure model that is based on first-principles solutions. Experimentalists are trying to rely upon both approaches. Even the first-principles kinetic WDM approach can take advantage of the code coupling when there is clear scale separation among some first-principles modeling, such as the gyrokinetic, rf, and molecular dynamics physics. The reduced-model code coupling approach can couple some first-principles kinetic modules. Findings from first-principles kinetic WDM can be used to strengthen the reduced-model approach.

Because the development of a fusion reactor is inherently expensive, difficult to understand, and of long duration, a high-fidelity WDM could be of great benefit to DOE and the world fusion program. This white paper is about the first-principles kinetic WDM approach. The United States has an edge in the first-principles approach compared to the European Union due to DOE's aggressive program in exascale computing.

2. **Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.**

Using the pre-exascale ecosystems (>100 peak PFlops/s), a first-principles WDM approach using 5D gyrokinetic equations can be achieved in 10 days wall-clock time; this contains the global device scale, the ion scale, and the spatially embedded electron and Debye scale multiphysics of ITER core-edge plasma. The electron gyroradius scale grid can be embedded using AMR. A multiscale time integration technique can be used to extend the first-principles WDM simulation to experimental time (>1 sec), without relying on the scale-separation assumption [2]. Applied mathematical tools are needed for both techniques. The clearly separable rf and material science simulations can be compiled together in the 5D gyrokinetic simulation using an on-memory, in situ computer science coupling tool. The fusion reaction and neutral recycling physics can be incorporated into the gyrokinetic framework.

3. **Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.**

What might not necessarily be solved in the next 5–10 years? Again, what is the importance?

a) The wall-clock time issue: occupying a pre-exascale HPC for 10 days can be an issue with other users, but can be handled over less-busy periods of time.
b) The electron gyroradius scale turbulence physics still needs to be confined to narrow, linearly unstable radial domains using the embedded grid and particle technique. This restriction is not as

severe as the other multiscale problems and may not need to be resolved within 5–10 years. However, the multiscale turbulence interaction may provoke a subcritical electron gyroradius-scale turbulence at unexpected places, and needs to be dealt with in post-exascale computers.

c) The gyrokinetic-rf coupling still needs to be on the scale-separation assumption, and may be justified at the lowest order level. However, there can be a non-negligible non-separable interaction between the 6D particle dynamics and the rf wave propagation. In the exascale and post-exascale HPCs, a phase-space embedded 6D kinetic simulation or a fully 6D kinetic simulation is desirable. Experimental time prolongation of the 6D simulation using a multiscale time integration technique without using the scale-separation assumption will also require exascale and post-exascale HPCs.

4. What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Multiscale time integration model	Enable experimental time study of first-principles WDM.
2. Embedded 6D solver algorithm in 5D grid	Enable embedded 6D kinetic simulation in 5D WDM. The field solver equations are different.
3. On-memory data analysis and visualization resources	Amount of data to be analyzed is too big for output to the file system and tape.

Impede	Why?
1. Hardware resources	Lack of sufficient computing time on pre-exascale HPCs will impede scientific progress.
2. Workforce development	First-principles WDM requires proficient HPC ecosystem knowledge as well as in-depth kinetic physics knowledge.
3. Fault tolerance and restart I/O	When the full pre-exascale HPCs are used for first-principles WDM, fault tolerance can be a serious issue. A restart file size can be large and can take a significant portion of the total computing time.

5. (Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.

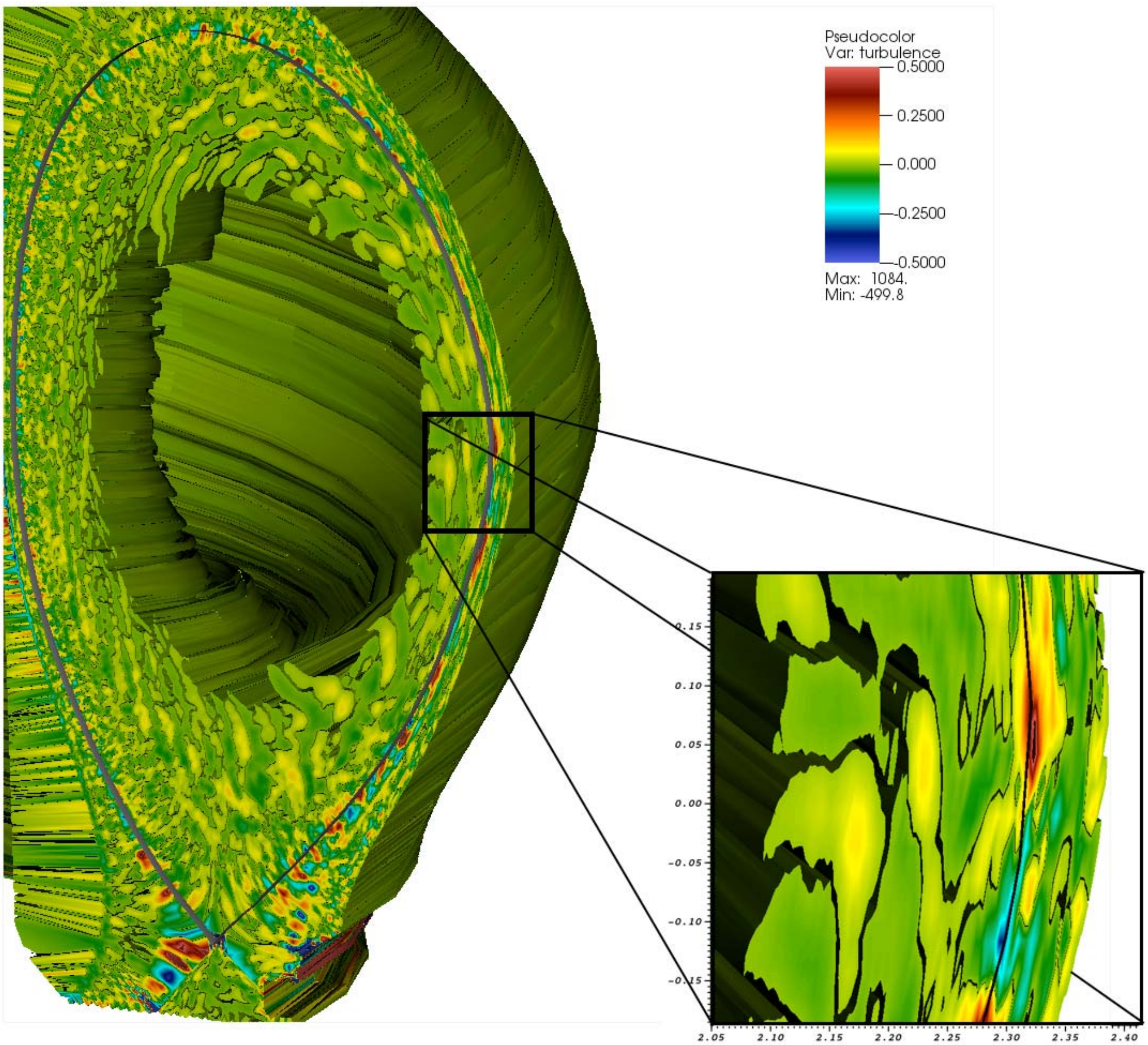
Even though the primary driver for the kinetic WDM research is not in the field of data transmission, analysis and processing, a kinetic WDM will produce extreme-scale data expected to be >0.5 PB per 2-minute time step in 5–10 years. This means that one restart file size will be >0.5 PB. The data *volume* is so large that conventional file systems may not be able to accept it; the data production *velocity* is so high that present-day I/O technology may not be able to handle it; and the *variety* of data contains time-dependent 5D kinetic particle dynamics information, 3D electrostatic potential fluctuation information, 3D magnetic field vector fluctuation information, 3D density and temperature fluctuation information, and tens of synthetic diagnostics data with different level of accuracy and confidence. We are addressing the big data problem using the in-memory in situ data analysis using the Data Spaces technology in the Adios framework while the simulation proceeds. In this way, we will be able to output the reduced physics data to the file system. While the data is flowing to the file system, we can do the secondary in situ data analysis and reduction before it is transferred to permanent storage.

6. **References** *(please keep the reference length to no more than 10 critical references)*

[1] R. Hager, 2015 FES Integrated Simulation Workshop Whitepaper
[2] S. Janhunen, et al., 2015 US-EU Transport Task Force Workshop, April 28–May 1, 2015, Salem, MA

7. (Optional) Images

Consider submitting one or two already published high-resolution images suitable for inclusion in the report. Please provide the reference(s). Submit these separately from the two-page report; they will not count against the page limit.



A whole volume XGC1 simulation of electrostatic turbulence, together with neoclassical and neutral particle physics, in diverted geometry. Turbulence is still moving inward at the time of observation. Turbulence is blobby in the edge and of ExB-flow sheared streamer type in the core.
Reference: OLCF Featured Highlight, February 2014; 2014 IAEA Fusion Energy Conference, Invited Talk.

Goals and Challenges Associated with Whole Device Modeling

Arnold Kritz and Tariq Rafiq, Lehigh University; and Alexei Pankin, Tech-X Corporation

The goals for Whole Device Modeling (WDM) are to provide a comprehensive predictive simulation capability for magnetically confined plasmas that integrates the knowledge from key multiscale physical processes to continually improve fidelity. This capability is needed to maximize exploitation of fusion experiments, especially ITER, and to establish the scientific basis for an economically and environmentally attractive source of energy. In particular, WDM software must be designed to meet the following needs:

- Model scenarios to plan new experimental campaigns in existing tokamaks or to extrapolate to planned future devices. Scenario modeling is used to optimize discharge parameters, such as maximizing fusion power production in burning plasmas, and to maximize the effectiveness of planning new experiments.
- Advances needed to compute core and edge turbulence and transport; large-scale instabilities, the sources and sinks of heat, momentum, plasma current and particles; and the plasma equilibrium.
- Develop code components that are well documented and that satisfy community-based standards.
- Allow for verification in order to eliminate sources of error and to access the degree to which the code correctly implements the physical models used.
- Validation and/or calibration of theoretical models using experimental data and synthetic diagnostics. WDM provides a platform where the validation of individual physics models can be performed in an environment that involves the interaction of the various physics components.
- Real-time modeling of tokamak experiments that utilize the predictive capabilities and feedback control techniques. The real-time modeling can be used to optimize discharge performance and to avoid disruptive events. Real-time modeling of tokamak plasmas can utilize advances in the uncertainty quantification techniques to include uncertainties in experimental data measurements.
- Production of self-consistent simulation results that are passed on to other more specialized codes.
- Mitigate disruption effects, predict pedestal formation and transient heat loads on the divertor, and compute tritium migration and impurity transport.
- Whole Device Modeling simulations will result in cost-effective harvesting of physics from national and international facilities and will accelerate progress to fusion power by stimulating innovations that will lead to better fusion device designs.

Within the past few years, advanced scientific computing has achieved a level where it is on par with laboratory experiments, enabling WDM to become a major tool for scientific discovery. WDM capability embodies the theoretical and experimental understanding of confined thermonuclear plasmas. The ultimate success of ITER will rely heavily on the development and use of whole device modeling. The kinds of physics problems that will be addressed with WDM codes by 2025, given the appropriate computational tools and facilities, will include the following:

- Predict the plasma confinement and details of transport in tokamak discharges. Currently, there are a variety of transport models that yield different predictions for confinement and fusion power production in burning plasma tokamaks such as ITER. There must be a convergence in the transport predictions based on high-fidelity turbulence and particle orbit computations. Various physics effects need to be considered, including effects associated with the behavior of non-local transport.
- Predict the onset, frequency, and consequences of macroscopic instabilities. Comparisons can be made with experimental data for the frequency of sawtooth oscillations, the effect that a sawtooth crash has on the plasma profiles, the onset of neoclassical tearing modes, and the resulting magnetic island widths. There is also a critical need to predict the onset of edge localized modes, their frequency and width, as well as the onset of disruptive instabilities and their nonlinear evolution.
- Determine the plasma boundary conditions from plasma-wall interactions through the scrape-off-layer and the H-mode pedestal. All of the plasma profiles are strongly influenced by the evolution of

the plasma boundary. Some WDM codes are also used to compute interactions between magnetic coil currents and plasma currents.

- Compute the sources and sinks that drive all of the profiles in plasma discharges. Sources such as neutral beam injection, fusion reaction products, and radio frequency heating and current drive, all involve the computation of fast particle distributions and their interaction with the thermal plasma profiles. Predictions are needed for the effect of fast ions on macroscopic instabilities such as sawtooth oscillations.

Scientific challenges which are not likely to be solved by 2025:

- Bridging the gap between short and long timescales, or between microscopic and macroscopic space scales. An example of this last kind of integration would be the simulation of turbulence, which grows on microsecond timescales and sub-millimeter space scales, resulting in transport across the plasma and the evolution of plasma profiles over tens of seconds in a tokamak with dimensions of several meters. Predictive WDM simulations that are exclusively based on the first principle models will still be unrealistic in the period between the present and 2025. The computational requirements for first-principle simulations of discharges that exceed 1,000 seconds in optimizations studies that require more than 100 runs in each optimization study are not likely to be met by 2025. The validation of all physics components needed for these runs are also not likely to be completed by 2025.
- The account of uncertainties in the physics models needs to be incorporated in the WDM codes so that the simulation results include the confidence intervals of the predictions. In predictive simulations, the UQ tools can be used for the evaluation of the probability of events, for the computation of confidence intervals of predicted quantities, and for the optimization of plasma performance. There is very little work in this direction at this moment, and it is unlikely that this scientific objective can be met by 2025.

Computing ecosystem aspects that can accelerate the progress in the next 5–10-year period include:

- The development of scalable algorithms, visualization, and analysis systems is required.
  - Removal of deficiencies in current numerical algorithms involves comparing computational solutions with benchmark solutions, with analytical solutions and with heroically resolved numerical solutions.
- Investments are needed in software design, repository management, release management, regression suites, and code documentation.
  - Software tools are needed to enhance the connection between simulations and experiments.
  - Experimental and simulation data need to be organized in a way that facilitates comparison.
  - Development of tools is needed to automate documentation of scientific workflows.
- Unlike gyrokinetic turbulence codes, whole device modeling codes, such as PTRANSP, often require 2,000 to 5,000 processors. Progress requires that NERSC provide this capability with fast turnaround time.
- A committed team of computer scientists, applied mathematicians, and plasmas physicists is needed to develop, improve, and maintain predictive integrated codes for carrying out whole device modeling of tokamak plasmas. This team should bring together into one framework a large number of codes and models that presently constitute separate disciplines within plasma science.

Computer Ecosystem Requirements for Coupled Core-SOL-Wall Whole-Device Modeling Simulations

A.Y. Pankin, S.E. Kruger, J.R. Cary, and J.R. King,
Tech-X Corporation, Boulder, Colorado
A. Hakim,
Princeton Plasma Physics Laboratory, Princeton, New Jersey
A.Y. Pigarov,
University of California San Diego, San Diego, California
A.H. Kritz and T. Rafiq,
Lehigh University, Bethlehem, Pennsylvania

Tokamak plasmas are complex nonlinear systems, and the evolution of these plasmas is determined by large numbers of interdependent physical effects that span vast time and spatial scales. The different regions of tokamak are described by a variety of models that require the tracking of dynamics that often span several regions. The tokamak boundary region includes the H-mode pedestal, the scrape-off-layer (SOL), and the tokamak wall. These three regions are often considered independently. However, the three regions are tightly coupled to each other as well as to the plasma core region. The particle, heat, and momentum fluxes from the plasma core contribute to the development of the H-mode pedestal. The H-mode pedestal sets the boundary conditions for the plasma core region and provides important sources to both the core and the SOL. Neutrals from the SOL supply particles to the edge and core plasmas. The physics of plasma wall recycling and sputtering are important in order to understand the dynamics of SOL profiles. Models that treat these regions as isolated cannot reliably pro duce predictive computations that describe the behavior of tokamak plasmas.

Importance of coupling of core-edge regions

The nonlinear coupling between different plasma regions can significantly alter the evolution of the plasma. The coupling of several components that have been previously validated independently of one another also requires a separate validation when the components are coupled. Coupling of plasma core and edge regions has been investigated using the FACETS simulation of the pedestal recovery after an ELM crash in a DIII-D discharge [1]. The two-dimensional UEDGE and one-dimensional FACETS:Core components have been coupled with a transition region that extends from the pedestal top towards the plasma core. In the initial simulation, the UEDGE component utilized the energy and particle fluxes found in the stand-alone interpretive UEDGE simulations. However, it is observed that the coupling of the pedestal region with the core region changes the H-mode pedestal profiles. In particular, the ion pedestal temperature has been found to be over-predicted relative to that computed in stand-alone UEDGE, indicating a greater flow of ion thermal energy into the edge region. The density buildup in the edge was under-predicted in the coupled core-edge simulation. To obtain the experimental density pedestal height and density level in the scrape-off-layer region, the neutral influx required in the coupled simulation was a factor of two greater than that obtained in the stand-alone interpretive UEDGE simulation.

Transient effects in the SOL region

The importance of SOL-wall coupling for the investigation of the dynamics of plasma boundary profiles after an ELM crash has been recently shown in Pigarov et al. [2]. The SOL profiles and wall-recycling parameters change on a very short timescale after an ELM crash. Depending on the type of ELM, the distribution of densities and energies in the SOL region can be very different after an ELM crash. Figure 1 shows the carbon charge states after an ELM crash represented by so called Marco-Blobs (MB) in the UEDGE-MB model [2]. The carbon species with lower charge states move toward the pedestal region, and the species with higher charge states move toward the wall. These changes in the distribution of charge species affect the plasma-wall interactions and the H-mode pedestal recovery dynamics. The profiles of these and other species in SOL as well as fluxes from the plasma core affect the pedestal recovery dynamics.

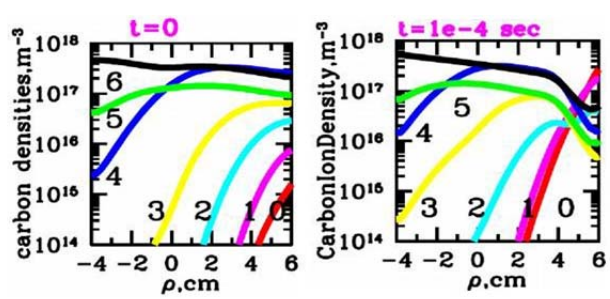


Figure. 1. Distribution of carbon charge state in the tokamak boundary region before (left panel) and after (right panel) an ELM crash.

Scientific challenges that can be solved by 2025

One of the main objectives of integrated whole-device modeling (WDM) is elucidating the tokamak physics that arise from complex interaction between different effects on a wide range of spacial and time scales. The prediction of plasma performance, discharge scenario optimization, and the investigation of transient effects are within the scope of WDM. Because of the nature of WDM research, the capacity of the computer ecosystem, which defines a number of independent runs with short turnaround times, is often more important than the computer ecosystem capability, which defines the hardware performance. Since we believe that development of computer ecosystems in the next 10 years will focus mostly on the development of hardware capabilities rather than on computer ecosystem capacity, most WDM improvements are expected in the area of higher fidelity components for WDM codes. The WDM codes will include a selection of theory-based models of different fidelity levels for all tokamak regions. This choice of multiple components will facilitate the verification and validation of individual physics models and the verification of coupled physics components. Modeling results obtained with stand-alone plasma boundary codes will need to be verified using codes that include the contributions from all the relevant regions. WDM verification and validation rely mostly on the capacity aspect of the computer ecosystem. Progress in the validation of high-fidelity components within WDM codes will depend upon the resources available for WDM. In the coupled core-SOL-wall WDM simulations, the complexity of coupled treatment of the boundary region (due to the wide range of spatial and time scales to be resolved, and the different dimensionality of the physics modules) imposes rigorous computational requirements for the coupling framework. Special attention will be given to load balancing, alternative coupling schemes, uncertainty quantification, components interchangeability, regression analysis, and exception handling. Collaboration with ASCR is critical to the large-scale computing needed for WDM.

Scientific challenges that are not likely to be solved by 2025

- The predictive WDM core-edge simulations that are exclusively based on first-principle models will still be unrealistic in 2025. The computational requirements for first-principle simulations of discharges that exceed 1,000 seconds in optimization studies that require more than 100 runs in each optimization study are not likely to be met by 2025. Validation of all physics components needed for these runs is also not likely to be completed by this time. Despite significant progress in experimental diagnostics in the plasma edge region, the error bars of experimental data in this region remain among the largest in tokamaks.
- The real-time control of tokamak discharges that is based on first-principle, gyrokinetic, self-consistent simulations is not likely to be achieved by 2025 due to stringent requirements for the computational resources and data-flow frameworks.
- The account of uncertainties in the physics models needs to be incorporated in the WDM codes so that the simulation results include the confidence intervals of the predictions. In predictive simulations, the UQ tools can be used for the evaluation of probability of events, for the computation of confidence intervals of predicted quantities, and for the optimization of plasma performance. There is very little work in this direction at this moment, and it is unlikely that this scientific objective can be met by 2025.

Computing ecosystem aspects that can accelerate the progress in the next 5–10 years

1. *Models and algorithms*: Development of more efficient parallel solvers is important for the WDM codes. These solvers can improve the utilization of available first-principle models such as GYRO and TGLF. Improvements of dynamic load balancing are important because they will help to utilize the computational resources more efficiently between different physics components of the WDM codes.
2. *Application codes*: The modern WDM codes (such as FACETS and TRINITY) that utilize novel computational approaches, rigorous regression tests, and advanced solvers still do not include many important physics compo- nents, synthetic diagnostics, and interfaces to various experimental data. Older codes, such as TRANSP, typically include sophisticated physics, but are outdated with respect to computational aspects including the code portability requirements, regression analysis, and parallel load balancing.
3. *Workforce development*: Many computational and scientific aspects cannot be addressed at this time because of insufficient workforce (for example, see the last item in the previous section).

[1] A. H. Hakim et al., “Coupled Core-Edge Simulations of H-Mode Buildup Using the Fusion Application for Core-Edge Transport Simulations (FACETS) Code,” *Phys. Plasmas* **19**, 032505 (2012).

[2] A. Yu. Pigarov et al., “Multi-fluid transport code modeling of time-dependent recycling in ELMy H-mode,” *Phys. Plasmas* **21**, 062514 (2014).

Computational Needs for a Community-wide Whole Device Model

F.M. Poli, R. Andre, and X. Yuan, TRANSP Group; *S. Ethier*, Princeton Plasma Physics Laboratory

1. Please specify the current science drivers for your field of research.

At the recent workshop on Integrated Simulations, the role of reduced models in a whole device model was emphasized.
This white paper illustrates the hardware and software challenges that the development of such a whole device model would likely encounter on a timescale of 5 to 10 years. Here the transport solver TRANSP, developed at PPPL, is used as an example of these challenges.
The choice of TRANSP follows from:
1 – It is a code widely used by the physics community both for interpretive analysis of experiments and for predictions, including ITER and power plant studies.
2 – It is a code in continuous evolution, where upgrades are mainly driven by user needs and requests.
3 – It is being used worldwide, and it is well suited for experimental validation and for verification of its individual modules.
The computational needs of TRANSP can therefore be understood as an example of typical requirements for a national transport solver.

2. Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.

TRANSP is historically a core transport solver. Over the years, implementation has focused on improving core transport and heating and current drive source codes, which has evolved the code from a serial to a parallel code. Over the past 2 years, a major implementation of the thermal transport has been the integration of TGLF, a reduced model for thermal transport, which scales well up to 512–1,024 cores. Another recent implementation is the 3D bounce-averaged Fokker-Planck solver CQL3D, combined with an MPI version of the ray-tracing code GENRAY (up to 240 cores). These capabilities, which are still under testing and not yet available to the average user, will increase the average number of cores for one predictive simulation up to typically 512 or more. This increase in compute core requirements will be in addition to the existing parallel code usage in TRANSP that consists of the Monte Carlo beam orbit module NUBEAM (~48–64 cores) and the TORIC ICRF solver (64–128 cores).
With a load of 30–40 simulations every day (average over the past 12 months), assuming one-third of them predictive, this would increase the request up to 30,000 processors every day from the present 800 average load. This additional demand is likely to happen within the next 12 months, when these new MPI capabilities are made available to all users.

3. Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.

What might not necessarily be solved in the next 5-10 years? Again, what is the importance?
Over a longer timescale, of the order of 3 to 5 years, TRANSP will make available to users simulation capabilities for coupled core-edge simulations. The reduced edge model will use Monte Carlo calculations (EIRENE), which scales well to processors numbering in the tens of thousands. Although it is expected that this code will evolve for efficiency, and perhaps use GPU, there is a clear shift in the computational needs, and, therefore, in the needed infrastructures.

Over a timescale of 5 years, a single TRANSP simulation complete with state-of-the-art thermal transport and particle transport model, high-fidelity RF codes coupled with Fokker-Planck solvers and with a plasma edge model, will need on average 10,000 to 20,000 or more processors and access to large memory.

TRANSP is a code used worldwide, but supported and run locally on the PPPL cluster. Soon local clusters, like the one at PPPL, will not be able to manage the workload and the requests from users any longer. It is unlikely that individual institutions will see their computer clusters increase in capacity. More likely, it is a situation where large-scale facilities, like NERSC, for example, will accommodate such large-scale needs. This would require not only funding for dedicated resources, but changes in the policy of such large-scale user resources, for example, dedicated queues, secure tunnel access, and file system. It also would require a change in the workflow of WDM-to-be codes like TRANSP, to provide adequate management of remote manual workflows.

Looking at the Exascale era, work is ongoing at PPPL on algorithms to adapt the computational time step during the discharge. This would allow us to interface TRANSP with first-principle codes — like gyrokinetic codes — at specific steps in the simulation. While capacity computing would be needed in preparation for that stage, these simulations in the Exascale era will allow direct replacement of reduced models (or their benchmarking) where these are most needed.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?** *Suggested topics include the following.*

Accelerate	Why?
1. Dedicated queues on national facilities (e.g., NERSC)	Take advantage of state-of-the-art computing capabilities, support, and competence.
2. Capacity computing	Allow more users to run with up to 5K cores and reduce wait time.
3. Efficient workflow managers	Optimal use of resources and wall clock.

Impede	Why?
1. Present queue policies	
2. Local cluster systems	Small size, inadequate to WDM needs on a 5-year timescale.

C.2 White Papers Addressing Plasma Surface Interactions and Structural Materials

C.2.1 Plasma Surface Interactions

Modeling Requirements for Incorporating Sheath and Near-Wall Plasma Physics in Coupled Plasma-Material Interaction Codes

Davide Curreli,
Department of Nuclear, Plasma, and Radiological Engineering,
University of Illinois at Urbana-Champaign, Illinois, Urbana 61801
dcurreli@illinois.edu
White Paper, OFES/ASCR Exascale Review Meeting
for Fusion Energy Science, 27–28 January 2016, Gaithersburg, MD

1. Science drivers

At the wall of fusion devices, the electrostatic sheath and the collisional and magnetic presheath act as an interface layer between the pedestal/scrape-off-layer (SOL) plasma and the material surface. Such interfaces include a multitude of processes, highly kinetic in nature, involving multiple plasma species (electrons, ions, neutrals, material impurities) in a dynamically evolving environment tightly coupled to the surface. At the nominal conditions anticipated for a reactor [1–8], the majority of sputtered material ionizes close to the surface and is redeposited nearby. The redeposition process forms a new *reconstituted* surface layer with different and unknown thermomechanical properties, different from the original ordered lattice. This continuously eroded and redeposited surface can significantly alter the PFC lifetime, affect the retention of hydrogenic species (deuterium, tritium), and affect the mechanisms associated with microscopic erosion of the surface (both net and gross erosion). Under continuous plasma exposure, the near-wall plasma and the surface form a system, far from equilibrium, in which the wall is continuously eroded, redeposited, and reconstituted. The problem is intrinsically multi-scale, both in space (from nanometers to centimeters) and in time (from fractions of a picosecond to minutes and hours), and multi-physics. The dynamic modeling of the kinetic processes occurring at the near-wall layer requires the coupling of different physical models and codes together, namely,

(1) A multi-species kinetic model of the plasma sheath/presheath region, handling the evolution of the distribution function of electrons, ions, neutrals, and material impurities from the quasineutral region to the first surface layer; the target equations are the Boltzmann-Poisson and the Boltzmann-Maxwell.

(2) A kinetic model of the material wall, handling ion-matter interaction, and including relevant phenomena such as sputtering, backscattering, implantation, on a material surface having dynamic composition and evolving morphology; the target equation is the classical multi-body problem for given (known) interaction potential.

(3) A proper collision operator accounting for the interaction among species, handling the relevant atomic physics such as ionization, charge exchange, ion and impurity recycling, and more. The target equations are the Fokker-Planck and non-linear collision operator.

2. Science challenges expected to be solved in the 2020–2025 time frame

Two technical challenges can be addressed with the existing computing ecosystem within the next decade:

(1) *Coupling of plasma sheath models to surface models.* Strong coupling of a kinetic plasma solver with a simplified material model can be achieved using the existing computing infrastructure. If limited to models of reduced dimensionality (less than 6D, up to maximum 5D) and to limited volumes (mm/cm maximum), production runs can be done using existing petascale machines. The same framework can be prepared for runs at full dimensionality (6D) on the next generation of exascale computers. A kinetic description of the magnetized plasma sheath can, for example, be achieved by using either discrete methods of intermediate dimensionality, such as 2D3V particle-in-cells [10], or continuum Boltzmann approaches of dimensionality ranging from 1D3V to 2D3V [11, 12], both cases in time-dependent conditions. The kinetics of the material response, limited to the physics of

implantation/sputtering/backscattering, can be described by using multiple methods, mainly divided into two classes:

- Monte Carlo codes using Binary Collision Approximations, such as TRIM, TRIDYN, SDtrim-SP, Fractal-TRIDYN. This class of methods reduces the multi-body problem to a Monte Carlo sequence of two-body problems. The advantage is mainly speed of execution and parallelization (perfectly parallel problem).
- Molecular dynamics codes for the inclusion of low-energy (multi-body) effects, such as LAMMPS. Long-time evolution including cluster dynamics and thermal effects can be added by an additional link to codes as PARASPACE or XOLOTL-PSI [9].

Strong coupling between the plasma module and the material module is required, since the timescales involved at the near-surface region in reactor-relevant conditions do not allow multi-scale separation. Coupling of plasma-sheath models to surface models can allow the characterization of the microscopic erosion occurring at the near-surface region, including the strong redeposition of the surface and its reconstitution, intended as dynamic evolution of the surface composition and morphology.

(2) *Coupling of sheath/material models to the pedestal/SOL.* Direct coupling of reduced-dimensionality sheath/material models to a pedestal/SOL code (e.g., XGC kinetic, SOLPS fluid) can also be prepared using the existing computing ecosystem. Production runs will likely require the next-generation computing infrastructure, to probe the long-time evolution of the plasma and surface. In the short term, parametrization of the sheath/material results can be done either via the generation of databases or the generation of proper parametric formulas. In this case large data sets might be needed locally on the machine at the node level.

3. Science challenges that cannot be solved in the 2020–2025 time frame using existing computing ecosystems

The final goal of achieving first-principle simulations of the long-time evolution of the plasma and the surface cannot be achieved using the existing computing infrastructure, including (1) fully coupled simulations of the pedestal/SOL kinetics with sheath/material kinetics, and (2) investigations of the long-time feedback loop between the pedestal/SOL and the wall. The exascale tool required for such investigations would probably be a gyrokinetic or full-f model of the pedestal/SOL coupled to a full-f 2D3V or 3D3V kinetic model of the PMI layer (sheath, presheath, and first surface layers interested by reconstitution processes) including a cluster dynamic model of the wall. Production runs of such a model will require multi-petascale or exascale facilities.

4. What top three computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Implementation and development of application codes, including code coupling	- Coupling between different codes will accelerate discovery science on PMI physics. - Monte Carlo BCA methods would greatly benefit from GPU acceleration (embarrassingly parallel problem similar to other 3D computer graphics problems).
2. Multi-code and multi-scale frameworks	- The need to couple more physical processes together would benefit from the generation of multi-code software frameworks handling data passage and interfaces. - Multi-scale methods can be implemented in a high-level framework independent from the client applications; logic can be abstracted from application-specific processes.
3. Visualization and analysis	- Visualization of large sets of simulation-produced data would significantly accelerate progress.

Impede	Why?
1. Models and algorithms	- Sustained load balancing, critical in plasma kinetic solvers, requires further development. - Scalability beyond ~200 nodes is critical for applications as MD and Poisson solvers.
2. Software resilience	- Software reliability and resilience at the largest scale will require further development or impede large-scale production runs on long-time evolution of plasma-surface interactions.
3. Hardware resources	- Multi-petascale facilities are required to solve for large-scale production PMI runs.

6. References

[1] R. Maingi, et al., *Report on Science Challenges and Research Opportunities for Plasma Materials Interactions*, FES Report, 2015.
[2] IAEA *Coordinated Research Project on Plasma-Wall Interaction with Irradiated Tungsten and Tungsten Alloys in Fusion Devices*, 2013-2018.
[3] M. Greenwald, et al., *Priorities, gaps, and opportunities: towards a long-range strategic plan for magnetic fusion energy*, 2007.
[4] R. Hazeltine, *Research needs for magnetic fusion energy sciences* (2009) http://burningplasma.org/web/ReNeW/ReNeW.report.press1.pdf.
[5] M. Greenwald, et al., *Opportunities for Fusion Materials Science and TechnologyResearch Now and During the ITER Era*, 2012.
[6] R. Betti, et al., *Report of the FESAC Subcommittee on the Priorities of the Magnetic Fusion Energy Science Program*, Technical Report, 2013.
[7] *Report on Strategic Planning. Priority Assessment and Budget Scenarios*, DOE FES Technical Report (2014).
[8] P. Bonoli et al., *Report of the Workshop on Integrated Simulations for Magnetic Fusion Energy Sciences*, 2015, Sections 4.2, 4.2.2, 4.2.4.
[9] Xolotl-PIS code at: http://sourceforge.net/projects/xolotl-psi/.
[10] R. Khaziev, D. Curreli, *Ion energy-angle distribution functions at the plasma-material interface in oblique magnetic fields*, Phys. Plasmas 22, 043503 (2015).
[11] D. Curreli, S. Keniley, R. Khaziev, S, Marcinko, *A Vlasov-BCA method for numerical simulations of the plasma sheath structure in presence of a material releasing wall,* Bull. Am. Phys. Soc. http://meetings.aps.org/link/BAPS.2015.GEC.KW3.2.
[12] S. Keniley, D. Curreli, *A Vlasov-BCA analysis on the wall erosion of a beryllium wall exposed to a high-density Helium plasma,* Bull. Am. Phys. Soc., http://meetings.aps.org/link/BAPS.2015.DPP.NO6.2.

FES White Paper

Efficient Self-consistent Integrated Simulations of the Response of Tokamak Plasma-facing Components to Transient Events on Exascale Computer Systems

A. Hassanein and V. Sizyuk

School of Nuclear Engineering, Purdue University, West Lafayette, IN 47906

1. Overview Description and Impact.

Understanding and controlling the integrated response of tokamak plasma-facing components (PFCs) to transient events is probably the single most serious obstacle to a successful concept of magnetic fusion energy production at the moment. Current tokamak operations show the complexity of the multi-processes nature of the divertor/edge plasma interaction with PFCs during various plasma instabilities. As a result, the need to control and mitigate the response of PFCs to plasma instabilities starting at the core to the SOL and to the divertor. Successful development of fusion reactors critically depends on the correct prediction of the heat and particle loads to reactor walls and the response/optimum materials choice for PFCs. A comprehensive computer package (i.e., HEIGHTS) has been successfully developed over the years that has unique capabilities to investigate several aspects of plasma-material interaction (PMI) phenomena during plasma transient events, including disruption, edge-localized models (ELMs), vertical displacement events (VDEs), and runaway electrons [1–4]. Any of these events, if not mitigated, could cause significant damage to PFCs and inhibit safe and reliable operation of the tokamaks. HEIGHTS simulates various stages of plasma interaction with facing materials and the subsequent vapor cloud evolution and secondary plasma and photon radiation to nearby components that are not directly exposed to plasma impact during these instabilities. Figure 1 shows the HEIGHTS model mapping to simulate the current ITER design using the adaptive mesh refinement approach (AMR) for multi-processor space discretization. The current version of HEIGHTS package integrates all processes during instabilities, starting from 3D Monte Carlo models of core plasma escaping events (Figure 2), transport and impact of particles into thedivertor surface, and including the effects of the secondary plasma formation during the wall/divertor response and erosion processes. The plasma heat conduction, radiation transport, and magnetic diffusion processes are also taken into account. Figure 3 shows a data exchange map across HEIGHTS blocks. The package is developed for the detail study of walls erosion and possible plasma contaminations during the transient events in ITER-like devices of various designs. Our recent calculations following this approach predicted unexpectedly significant damage risk for the hidden stainless steel “umbrella” of the dome structure in the current ITER design due to the intense secondary radiation during a disruption on the original tungsten divertor plate [1]. This can only be predicted by mapping the entire divertor area with detailed 3-D geometrical configuration. The potential of such damage to the hidden components is very serious and could end up causing reactor operation to be shut down for several months until repair of such hidden and critical components is achieved!

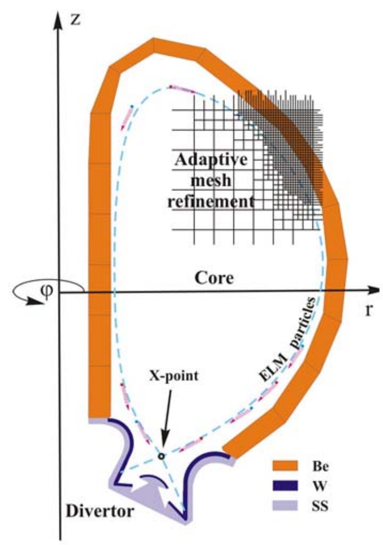


Fig. 1. HEIGHTS model mapping for the current ITER design

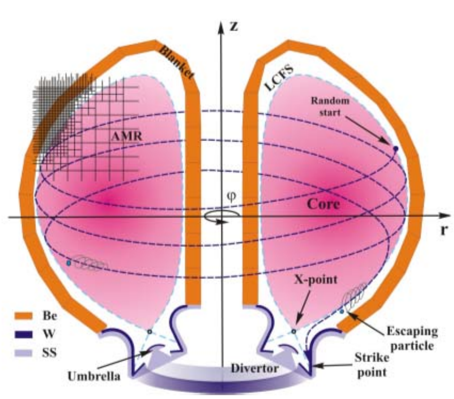


Fig. 2. 3D Model of core escaping particles

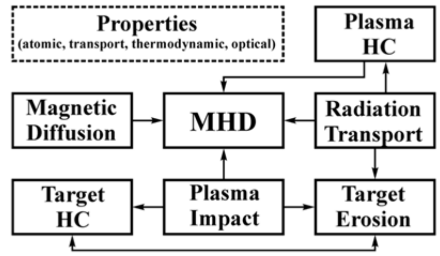


Fig. 3. Data exchange map of the HEIGHTS package

Integrated computer simulation of complex but realistic engineering systems is becoming important in the analysis and design optimization of such systems based on comprehensive integrated multiscale physical models. A reasonable amount of computer time is critical for any successful integrated computational package. Figure 4 shows that hydrodynamic and radiation transport blocks of the HEGHTS Package have very good parallel efficiency that assures fine domain decomposition (which is important for hydrodynamics block) and is explained by recently developed efficient and innovative Monte Carlo models (for the radiation transport) [5]. However, opposite to that, the thermal conduction block scalability is very inefficient and very slow. Moreover, this situation becomes a "bottleneck" of modern integrated simulations needed for accurate prediction of these events. At the same time, the heat conduction problem in edge plasma is very important, as described in various references [6], and cannot be ignored. Experimental evidence shows that high thermal conductivity in the electron fluid is the main cause for the electron losses from tokamaks, theoretical research and modeling are complicated with the extreme anisotropy of process, and practical use of supercomputers is limited due to the small scalability of appropriate numerical methods. We are limited in the application of large computers for such necessary integrated simulation of various transient events and overall divertor response due to both the current low scalability of the numerical solutions in HEIGHTS and the inefficient data communication between the integrated blocks of HEIGHTS.

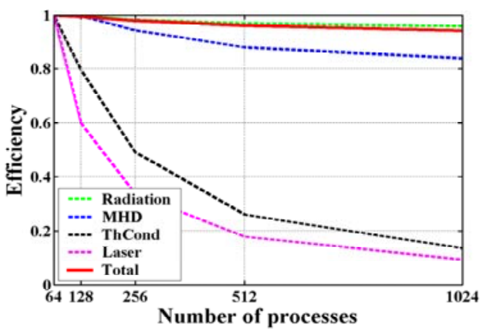


Fig. 4. Efficiency of various HEIGHTS modules as a function of number of processors

2. **The science challenges expected to be solved in the 2020–2025 time frame using extant computing exascale systems.**

The full 3-D simulation of tokamak design of PFCs should be done in this time frame by the fusion research community. These integrated models should include a wide range of the separate sub-processes that should be simulated in the large package. The dissipation processes based on the parabolic equations (heat conduction, magnetic diffusion, radiation transport) should be significantly reformulated and new numerical approaches should be developed. Some of the most important results needed are details about plasma-wall interaction and plasma/material contaminations. The integrated simulation needed for the wide range of multiscale physical processes requires intensive optimization for the use of an exascale system. The solutions developed during the optimization of numerical schemes and data transfer between blocks can also be useful for other fusion computer packages.

In this white paper, we propose to integrate the 3D tokamak geometry of SOL and divertor/wall areas and to direct significant efforts to overcoming the scalability limitations related to the numerical schemes. We previously demonstrated [7] that accurate simulations using structured meshes result in the composition of diagonal matrixes and *that* the final solution of thermal conduction and magnetic diffusion in the multidimensional MHD divertor/wall evolution problems can be performed using the sparse matrix methodology. Three-dimensional modeling with mesh sizes larger than 100 × 100 × 100 will be very problematic because of the large memory needed and the low scalability of implicit methods based on parallel linear equations solvers. Our simulations and benchmarking of various linear solvers (IMSL

Windows and HYPRE Linux) showed limitations in the total amount of the AMR domain cells that correspond approximately to the standard structured mesh in 220 × 220 cells. Larger mesh sizes have resulted in calculation breakdowns in both solvers. These limitations forced us to look for other alternative methods that might be more expensive computationally but that are more scalable, robust, and memory efficient [8]. We developed and tested a new direct Monte Carlo algorithm based on a new concept of introducing "thermal particles." The algorithm is similar to our solution of the radiation transport problem in the secondary plasma evolution of the divertor/wall material [9]. This new Monte Carlo procedure warrants high scalability (green curve in Figure 4) and minimum additional memory requirements for time-dependent and integrated multi-physics tasks. The algorithm can be expanded for fusion applications using the 3D curvilinear coordinate systems.

Our recent modeling [4] showed that the physical processes involved are highly intercorrelated and for accurate assessment and evaluation to these serious events, self-consistent integrated models for the entire SOL area should be implemented. It requires extensive communications among the separate blocks included in the HEIGHTS package (Fig. 5) that need scalability issues to be reduced significantly. The overall efficiency of the much-needed integrated package will be determined by the slowest block and by the time required for data transfer among these modules. Optimization of data exchange in the integrated package is the second goal of our project, because these large, separate physics modules critically depend on each other and on exchange large data at every time step.

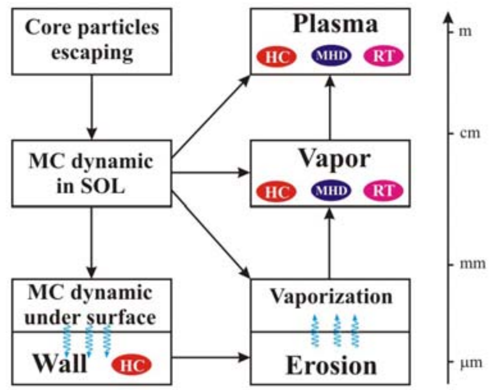


Fig. 5. Data exchange scheme of the HEIGHTS package. The multiscale codes (see right axis) are integrated for tokamak secondary plasma modeling

3. **Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.**

The numerical solutions based on implicit numerical schemes and linear equations solvers should be upgraded for extant computing ecosystems because of weak scalability. According to the standard numerical schemes, systems of partial differential equations are expressed as a closed system of the linear equations to be solved with pre-developed solvers. We do not anticipate a large amount of progress in the use of the standard linear equation solvers due to the required active data interchange between the processors. We propose to pursue the development of new probabilistic methods instead.

4. **What top three computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. New models and algorithms	Existing models and algorithms have very low scalability.
2. New efficient codes	Fusion modeling can be accelerated by balancing and optimizing the integrated code parts.
3. Data workflow	The calculated data should be efficiently available for common sharing and use among various modules.

5. References

1. V. Sizyuk and A. Hassanein, Heat loads to divertor nearby components from secondary radiation evolved during plasma instabilities, *Phys. Plasmas* **22**, 013301 (2015).
2. J.N. Brooks, A. Hassanein, A. Koniges, P. S. Krstic, T.D. Rognlien, T. Sizyuk, V. Sizyuk, and D.P. Stotler, Scientific and computational challenges in coupled plasma edge/plasma-material interactions for fusion tokamaks, *Contrib. Plasma. Phys*. **54**, 329 (2014).
3. V. Sizyuk and A. Hassanein, Kinetic Monte Carlo simulation of escaping core plasma particles to SOL for accurate response of plasma-facing components, *Nucl. Fusion* **53**, 073023 (2013).
4. V. Sizyuk and A. Hassanein, Damage to nearby divertor components of ITER-like devices during giant ELMs and disruptions, *Nucl. Fusion* **50**, 115004 (2010).
5. V. Sizyuk, A. Hassanein, V. Morozov, V. Tolkach, T. Sizyuk, Numerical simulation of laser-produced plasma devices for EUV lithography using the heights integrated model, *Numer. Heat Transfer, A*, vol. 49, pp. 215–236, 2006.
6. L.C. Woods, Theory of Tokamak Transport: New Aspects for Nuclear Fusion Reactor Design, Weinheim: WILEY-VCH, 2006.
7. G.V. Miloshevsky, V.A. Sizyuk, M.B. Partenskii, A. Hassanein, and P.C. Jordan, Application of finite-difference methods to membrane-mediated protein interactions and to heat and magnetic field diffusion in plasmas*, J. Comp. Phys*., vol. 212, pp. 25–51, 2006.
8. V. Sizyuk and A. Hassanein, Efficient Monte Carlo simulation of heat conduction problems for integrated multi-physics applications, *Numer. Heat Tr., B-Fund*. vol. 66, p. 381, 2014.
9. V. Sizyuk and A. Hassanein, Heat loads to divertor nearby components from secondary radiation evolved during plasma instabilities, *Phys. Plasmas*, vol. 22, p. 013301, 2015.

Multiscale Approach for Plasma-Material Interface

Predrag Krstic

Institute for Advanced Computational Science, Stony Brook University, Stony Brook, New York 11753-50501.

Where are we today?

The control of coupling between the plasma edge and the wall surface has been inhibited by a lack of fundamental understanding of their interface. The plasma-material interface (PMI) mixes materials of the two worlds, creating in between a new entity, a dynamical surface that communicates between the two, creating one of the most challenging areas of multidisciplinary science, which has many fundamental processes and synergies. We know that the edge plasma governs particle and energy exhaust, and impurities eroded from the surfaces may reduce the fusion gain if they are transported back to the confined plasma. The other critical effects of these interactions are reduced lifetime of plasma facing surfaces due to erosion by transients, and restrictions on duty cycle due to retention of tritium in redeposited material and in dust created by plasma surface interactions. Furthermore, all choices for plasma-facing materials (PFM) in a fusion reactor have known issues. There are good arguments for both low-Z versus high-Z PFMs in fusion devices and they have been discussed many times in the fusion community.

The traditional trial-and-error approach to developing first-wall material and component solutions for future fusion devices (ITER, DEMO) is becoming prohibitively costly because of the increasing device size, curved toroidal geometry, access restrictions, and complex programmatic priorities. This requires change from engineering toward a fundamental, bottom-up approach to the plasma-material interface. The experimentally validated atomistic theory/computation for studying the dynamics of the creation and evolution of the PMI under irradiation by heavy particles (atoms, molecules) at carbon, lithiated, and boronised carbon (NSTX) and tungsten (ITER), as well as the emerging elastic and inelastic processes, in particular, retention and sputtering chemistry, are some of the burning scientific challenges we are dealing with. Quality validation of the simulations is the key for the *right track*. Mimicking the experiments by simulation is the key for successful validation. High-quality experiments, well suited for the purpose, do exist (NSTX-U, UUIC). With these challenges in mind, we study the dynamic surface response and evolution for long-pulse PFCs in mixed material environments (D, Li, C, B, O, Mo, W, TZM, Fe) with impingement of plasma particles in the energy range below 100 eV, including microstructure changes, erosion, surface chemistry, deuterium implantation, and permeation. Our main research goal is to understand the changes in surface composition and chemistry at the nanoscopic temporal and spatial scales and link these to overall NSTX-U machine performance by both deciphering the nanoscale phenomenology of plasma-surface interaction and providing the parameters needed for correct treatments and understanding at the mezoscopic scales. Whenever lithium is present, the evolution of the material mixture has to be treated by quantum-classical molecular dynamics (QCMD), which is a technical challenge and a bottleneck for the fundamental PMI science, currently achievable, though with size-based difficulties, using approximate approaches to quantum mechanics and with the extensive use of the supercomputing facilities. Thus, with one set of structural and energy parameters, one run at the Cray XK7 computer of NCCS takes about 100 hours with 70,000 computing cores, in order to reach acceptable statistics and uncertainty quantification of the results, indicating a need for exascale computing power.

Plasma-material science has a big effect on the plasma performance, and we do not understand why. The answer can be found in plasma-PMI integration science. To date, predictions have been made via a combination of extrapolation of data from existing experiments and from simulations calibrated on those experiments. The predictive capability of both plasma and PMI approaches is limited since neither is yet based upon a first-principles theory that would provide the required degree of confidence in extrapolation beyond the existing database. Furthermore, the simulations utilize simplified models for the behavior of both the plasma and materials. Still, the main weight in the science of integration of fusion plasma and its interfacial surface boundaries is carried by PMI because (1) the basic phenomenology often evolves much faster than the plasma timescale and (2) it evolves through wider range of the scales, which partially overlap with the scale of plasmas. The PMI has to be understood and parameterized at nanoscale before it can be integrated with plasma at the same footing at microscale. Both the plasma and material systems are inherently multiscale in their interactions, requiring descriptions of phenomena over many decades of time and length scales. Bringing together the various scales of PMI and plasma is the fundamental multidisciplinary question covering plasma science, surface science, atomic physics computer science, and applied mathematics. Even within the divertor plasma itself, especially in the detached regime, one needs to couple self-consistently the partially ionized plasma, high-density neutrals, Debye sheath physics, and radiation fields. That complexity has prevented the assembly of a comprehensive description of the two systems and their interactions. At best, simulations to date have focused on describing either the plasma or material in detail, but not both. Moreover, the evolution of plasma profiles and observed material surfaces effects often happens on the same timescale; therefore it is a high priority to explore the direct coupling of plasma and plasma-facing materials in the time range they overlap, in the mesoscopic range. The most successful plasma simulation codes in use presently utilize approximate, not-first-principles descriptions of plasma transport and plasma material interactions; the inadequacy of this treatment is widely acknowledged.

2. The science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems

- The coordinate-dependent charging and polarization are correctly described only by electron dynamics. Since atomic electrons are necessarily quantum-mechanical entities, the QCMD of large and mixed systems present, in these cases, a first-principles alternative to classical MD, and a scientific discovery challenge in the time frame of the next 10 years.
- How to realistically treat the system beyond the timescales of classical molecular dynamics (ps- ns), when the events are not known in advance, is certainly a central computational challenge. An efficient and robust treatment of multiscale aspects of the developed models is at the center of mathematical contributions in the next 10 years.
- Creation, calibration, and validation of improved multibody, semiempirical potentials for CMD.
- Deriving the coefficients from the first principles for the mesoscopic source and sink terms using MD for various material interfaces with fusion plasma.
- A big computational challenge in the next 10 years will be how to realistically treat the system beyond the timescales of classical molecular dynamics (ps-ns), when the events are not known in advance, enabling a prediction.
- Developing an integrated model of divertors based on first-principles models that will lead to a solution consistent with experiments will acquire a fundamental understanding of two outstanding divertor and edge plasma phenomena: the radiative, detached plasma and the erosion, transport, and redeposition of impurities. This is a difficult but realizable task in the next 10 years.
- Development and application of the UQ and sensitivity analysis of complex fusion plasma and PMI systems. For that purpose, having in mind a noisy nature of the relevant data, a very large number of simulations will be performed using exascale computation (linear scaling with a number of parameters).

3. Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.

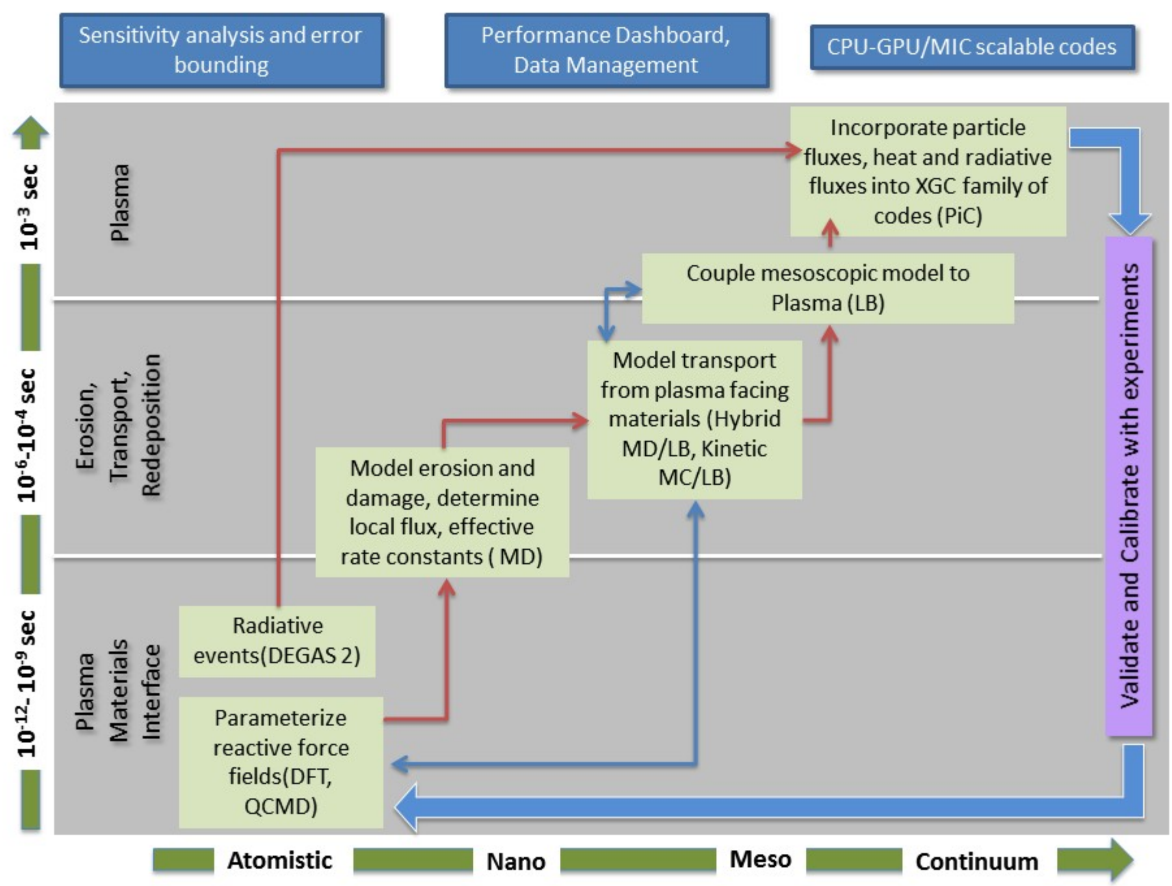
An integrated modeling system that will be based on first-principles models and will eventually provide a consistent solution of the whole device from the material surfaces to the plasma core.

4. What top three computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Models and algorithms	The mathematical research should focus on systematic development of coarse-grained potentials, stability and error control, and robustness of parameterized models.
2. Hardware resources (at all scales) including I/O, memory, and so forth	This is the driver.
3. Workforce development	One-to-one correspondence of workforce development and progress in science

Impede	Why?
1. Underdeveloped system software	Might cause bad scaling, which undermines any progress above.

5. Images



Multiscale Challenges in Modeling Plasma Facing Materials

Danny Perez and Arthur F. Voter, Los Alamos National Laboratory

1. Current science drivers for your field of research

Applications in fusion energy production place extraordinary demands on plasma facing materials because they have to withstand very high temperatures, high doses of radiation, and impinging He and H ions. Taken together, these factors can lead to extremely high concentrations of defects, whose interactions cause complex and deleterious changes in the microstructure and hence in the properties of the material. A crucial difficulty is that the problem is inherently multiscale in both time and space: timescales range from femtoseconds (to describe dynamics in the immediate aftermath of collision cascades) to hours and days (to capture slow microstructural changes), while length scales range from nanometers to describe individual defects to centimeters to fully describe microstructures. Adding to the difficulty is the fact that the relevant physics operating at each of these different scales and the synergetic effects of the couplings between scales are still only partly known. Being able to understand, predict, and ultimately design materials that can reliably operate in this kind of extreme environment is a grand challenge that has to be tackled by the community in order to make fusion energy a reality. However, experiments alone, because of their difficulty and cost, are likely to be insufficient to answer many of the key materials science questions; thus, support from computational and theoretical approaches is essential.

The problem of spanning such disparate scales is the primary outstanding issue in the field. We can now directly simulate the behavior of individual defects (clusters of He/H, of vacancies, and of self-interstitials; small He bubbles) over timescales that allow us to probe their mobility and kinetics. For example, using long-time-scale, fully atomistic methods such as Accelerated Molecular Dynamics (AMD) techniques [1,2] on petascale computers such as Titan at Oak Ridge National Laboratory, we can simulate the growth of nanoscale He bubbles in W under growth rates that approach these expected in ITER [3]. Because of their extremely high fidelity, such simulations provide unique insights into how such bubbles nucleate, grow, and interact with microstructural features such as surfaces following He intake. However, much less is known about the interactions between defects and between defects and the microstructure. For example, considerable experimental evidence points to the essential role of Hein the formation of the so-called fuzz on W surfaces. However, the mechanism by which this occurs is still the subject of intense discussion because it presumably stems from collective effects that occur only on long time and length scales. As discussed below, some of these questions can be addressed simply by directing more computing power at the problem with existing methods, but some will remain out of reach even on exascale computers unless novel methodologies are developed.

2. Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.

Even without major methodological advances, the range of size and length scales that can be directly simulated with high accuracy is expected to increase significantly. While we now consider growth of nanoscale He bubbles over timescales of tens of microseconds, the advent of exascale systems will enable millisecond simulations and a possible increase in the system sizes (and hence defect sizes) that can be handled. This will certainly improve the quality of the microscopy parameters that can be fed to higher scale models that aim at modeling evolution at the microstructural (or even reactor) level. The robustness and predictive nature of these models will therefore improve significantly. Similar improvements in the understanding of other nanoscale defects can be expected.

3. **Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.**

Even on an exascale system, direct simulation based on an atomistic description of the system cannot be expected to single-handedly reveal microstructural evolution mechanisms on relevant scales. For example, direct MD simulation using a computationally inexpensive semi-empirical potential can be expected to reach microseconds of simulation time on billions of atoms. Using advanced AMD methods such as Parallel Replica Dynamics, one will perhaps achieve tens or hundreds of milliseconds with 10^5 atoms. While such a dramatic improvement will be put to good use by improving the fidelity of higher-scale models, direct atomistic simulation is, and will remain, insufficient if predictions are to be made on time and length scales that are of direct relevance to the operation of fusion reactors. Upscaling by relying on intermediate-scale models will be essential. However, an increase in computing power does not in itself provide a solution to the upscaling problem, and progress in this direction will have to rely on significant methodological advances. In other words, parameterizing coarser models from microscale models will be made easier at the exascale, but developing these coarser models will remain a crucial challenge.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

As discussed above, an increase in the amount of available flops and memory will contribute immensely to the improvement of the fidelity of our models of materials evolution in two ways. First, it will greatly extend the length and timescales that are amenable to direct numerical simulation with MD or AMD. This will allow us to model the problem in its full complexity, albeit on time and length scales that still fall short of experiments. Second, exascale resources will provide tremendous capacity that can be leveraged to search for new physics using an extremely large number of smaller simulations. Taken together, this will put the basic physics of the problem on firmer ground by reducing the uncertainty inherent in identifying crucial physical processes that control nanoscale evolution.

Single-scale models, however, will be insufficient. Our ability to perform at the exascale will be contingent on the development of adapted methodologies that can span scales and whose fidelity can be systematically improved using lower scale models without human intervention. This will require significant theoretical advances as well as an acute understanding of computational realities. Playing against us will be the added complexity of the task at hand: we will need to efficiently generate, analyze, and integrate huge amounts of information in an automated fashion. This will probably be best carried out on systems with specialized nodes, some of which allow for large memory and fast I/O, while others concentrate on high flop counts. Successfully leveraging these resources will also require middleware, such as work-flow managers, that can integrate the on-the-fly analysis of the data in order to identify the next simulations to be performed. Given the amount of data that will be generated, it is imperative that such work flows be completely automated. Further, computational flexibility will be essential, as execution of the work flow might alternate between phases where millions of cores could be harnessed for individual simulations, while others might perform best by dividing the work into a large number of smaller tasks. A monolithic computational model that emphasizes only either capability or capacity at extreme scales is unlikely to be optimal.

Finally, given the amount of human time and effort involved in developing methods and software that can leverage extreme-scale computers, success will also rely on a certain level of predictability of the platform's evolution over a time frame of a few years. While unforeseen adjustments will be unavoidable, the cost associated with frequent architectural changes should be considered in the long-term planning.

5. References

[1] "The parallel replica dynamics method – Coming of age," Danny Perez, Blas P. Uberuaga, and Arthur F. Voter, *Computational Materials Science* **100**, 90 (2015).

[2] "Accelerated molecular dynamics methods: Introduction and Recent Developments," D. Perez, B.P. Uberuaga, Y. Shim, J.G. Amar, and A.F. Voter, *Annual Reviews of Computational Chemistry* **5**, 79 (2009).

[3] "Competing kinetics and He bubble morphology in W," L. Sandoval, D. Perez, B.P. Uberuaga, and A.F. Voter. *Physical Review Letters* **114**, 105502 (2015).

Fusion Plasma Material Interactions
Brian Wirth, University of Tennessee, and David Bernholdt, ORNL

The realization of fusion as a practical, 21st-century energy source requires improved knowledge of plasma material interactions and the materials engineering design of component systems to survive the incredibly extreme heat and particle flux exposure conditions of a fusion power plant. In considering plasma-material interactions (PMI), it is evident that three coupled spatial regions influence PFC materials evolution and performance. These regions consist of (1) the edge and scrape-off layer region of the plasma, (2) the near-surface-material response to extreme thermal and particle fluxes under the influence of, and feedback to, the plasma sheath, and (3) the structural materials response to an intense, 14 MeV peaked neutron spectrum, which produces very high concentrations of transmuted elements through (n,p) and (n,**α**) reactions and structural material property degradation. The coupled nature of these spatial domains necessitates the interfacing between modeling approaches for each, in order to better evaluate the feedback between each region on the performance of the other. For example, the interface of the surface to the plasma edge/scrape-off layer is necessary to define the incident particle and thermal fluxes that are the driving force for PMI, as well as to appropriately account for the processes of excitation, ionization, and charge-exchange that can result in species redeposition. Likewise, the interface between the surface and the bulk, where defect creation is no longer influenced by the presence of a free surface, is critical in determining the extent to which defect creation by high-energy neutrons affect retention and permeation of hydrogen isotopes; a significant unknown is the tritium permeation behavior in metallic PFC at elevated temperatures.

Gaining physical understanding and predictive modeling capability in this critical PMI area requires simultaneously addressing complex and diverse physics occurring over a wide range of lengths (Ångströms to meters) and times (femtoseconds to seconds, days to years) and integrating extensive physical processes across the plasma-surface-bulk materials boundaries. The objective of this proposal is to develop and deploy validated, high-performance simulation tools capable of predicting the performance of tungsten-based PFCs in a burning fusion plasma environment. This includes modeling surface morphology evolution in either erosion or redeposition regimes, and the recycling of hydrogenic species, as a function of plasma exposure conditions, temperature, and damaging 14 MeV neutron flux. This requires a leadership-scale computational code that is well integrated with a suite of multiscale modeling techniques to bridge the scales over which complex and synergistic PMI processes determine performance.

System Requirements: HPC system size in terms of available memory and processing speed limit the domain size and available simulation time for atomistic molecular dynamics simulations. Current capability allows for simulations of order 10^8 atoms for order 10^{-5} seconds using order 10^5 processors within a year. Extrapolation to exascale will allow millisecond simulation time only. Continuum reaction-diffusion cluster dynamics approaches are currently limited by available memory for direct numerical integration of the linear algebra based on the Jacobian size.

- **Atomistic code, molecular dynamics code** (LAMMPS, http://lammps.sandia.gov/): Predominantly run using MPI and OpenMP. GPUs can be used well where the potentials are implemented for them. Some of the potentials required for this work are already implemented on GPU; we are working toward implementing the Tersoff-style 3 body potentials for the tungsten-hydrogen interaction.

- **Continuum reaction-diffusion cluster dynamics code** (Xolotl, http://sourceforge.net/projects/xolotl-psi/): Currently under intensive development as part of the Plasma-Surface Interactions SciDAC project. The Xolotl code currently uses only MPI for node-level parallelism. The recent extension of the code to three dimensions greatly increases the computational intensity of these simulations, and will motivate a near-term emphasis on node-level parallelization, primarily through OpenMP and/or OpenACC directives, to take full advantage of both many -core and hybrid/accelerated node architectures. Time parallelization techniques will be of significant interest due to the desire to extend the modeling to multi-decade component lifetimes.

- **End-to-End Requirements:** An extensive database of atomistic simulation results, with sophisticated data management capabilities, will be required to characterize mechanisms and extract parameters for key reactions. Although available visualization and analysis techniques are adequate for current rates of data production, higher production rates enabled by exascale computing and the introduction of more extensive uncertainty quantification will undoubtedly motivate new developments in this area. Although cluster dynamics data requirements will be significantly smaller than those for the atomistic simulations on a per-time-step basis, the long simulation times and the large numbers of runs required for uncertainty quantification will result in huge data volumes to be managed. Innovation will undoubtedly be required to visualize and analyze the long-time simulation runs and make sense of the UQ results.

Related Research: Plasma material interactions influence the behavior of plasma thrusters for space propulsion, and share synergies with nuclear materials behavior in extreme-environment and low-energy plasma materials processing. Each of these technical areas would benefit from increased computational capability to predict behavior in extreme thermal and radiation environments, where experimental data are often limited or unattainable.

10-Year Problem Target: Fully parameterized continuum cluster dynamics model with radiation damage, incident plasma and neutral ions on the plasma facing materials and representative heat fluxes representing next generation fusion reactors beyond ITER. This requires the ability to simulate 20-year component lifetimes within hundreds of seconds of wall-clock time. It requires extensive input from atomistics and UQ, as well as the ability to treat finite-difference based ODEs with order 10^{11} degrees of freedom. It will enable predictive engineering scale modeling of fusion reactor performance.

Other Considerations/Issues: The research program described here will require a multidisciplinary and almost certainly multi-institutional team to work in close collaboration on different aspects of the problem. The ability to efficiently share data and analyses amongst the group will also be vital to its success.

C.2.2 Structural Materials

Mesoscale Modeling of Radiation Damage in Fusion Structural Materials

Richard Kurtz, Giridhar Nandipati, and Wahyu Setyawan
Pacific Northwest National Laboratory

1. Current Science Drivers

It has long been recognized that fusion materials must function in an extremely hostile environment, including various combinations of high temperatures, reactive chemicals, time-dependent thermal and mechanical stresses, and intense damaging neutron fluxes. Material properties depend on composition and microstructure at length scales ranging from smaller than a nanometer to centimeter dimensions. Exposure to the fusion environment causes the composition and microstructure of materials to evolve, frequently leading to property degradation and unanticipated failures.

Development of successful structural materials will ultimately require a fusion-relevant neutron source in order to code-qualify these materials for nuclear service. However, a purely experimental approach to understanding and mitigating radiation-induced degradation is not practical because of the high cost to design, perform, and examine materials from irradiation experiments and the low volumes of irradiation available. The lack of a fusion-relevant neutron source in which to conduct prototypic experiments reinforces the need for a robust theory and simulation program in order to understand experiments carried out in surrogate irradiation facilities. Furthermore, there is a combinatorial problem; that is, the broad range of materials, phenomena, and irradiation variables and variable combinations make a purely experimental approach intractable. Physically based computational models of microstructure and property evolution are indispensable tools because they provide a means to reevaluate existing data, optimize the design and execution of new experiments, and interpret the results from those experiments. Multiscale models describing radiation and mechanical damage processes are under intense development, but numerous details remain to be resolved before these models can accurately predict material performance, because these models must simultaneously span length scales ranging from atomistic to the continuum and timescales ranging from subpicosecond to years. Here we discuss the development status and computational needs of one class of simulation methods known as kinetic Monte Carlo (KMC), which is a mesoscale technique that bridges the gap between atomistic-level tools, such as density functional theory and molecular dynamics simulations, to more continuum-level approaches, such as cluster dynamics and rate theory.

Three major challenges of using the KMC method to simulate radiation damage are (1) reaching experimentally relevant time scales for physically meaningful system sizes, (2) simulating experimentally relevant length scales, and (3) carrying out high-fidelity or realistic simulations incorporating all needed defect interaction physics. For radiation damage simulations, object KMC (OKMC) is the method of choice in which the objects of interest are defects, and their reaction/diffusion mechanisms and therefore all the challenges mentioned above pertain to OKMC simulations. Computational time per KMC step increases slightly while the advance of simulation time per KMC step decreases significantly with an increasing number of mobile objects. The latter has the largest effect on the achievable simulation time. Therefore, the maximum simulation time achievable in a reasonable (or affordable) amount of real time depends on the number of mobile objects in the simulation. Due to the inherently serial nature of the KMC algorithm and the minimal computational cost per KMC step, exascale systems (ExS) will not be helpful for extending the simulation time (time scale) on system sizes that a serial KMC code can handle (strong scaling). This is and will always remain a challenge. However, the use of ExS would be most beneficial in extending the length scale of KMC simulations even with existing parallel KMC algorithms, which indirectly will also extend the time scale of OKMC simulations (weak scaling). The fidelity of radiation damage evolution predictions obtained from OKMC simulations depends on the degree to which all possible/relevant reaction/diffusion mechanisms are included, and on how rigorously they are treated in the simulations. Making OKMC simulations more realistic than they are currently would increase the computational cost per KMC step, making them computationally more expensive. ExS would be very beneficial in extending the length and time scales as well as the fidelity of OKMC simulations.

Note that, even though several algorithms exist, KMC simulations in almost all cases employ serial architecture. To the best of our knowledge, KMC codes have utilized neither megascale nor petascale systems to carry out parallel KMC simulations. This is mainly because parallelization of the KMC method is nontrivial and parallel efficiency is highly dependent upon the problem being studied as well as the simulation parameters used. Moreover, parallel efficiency of a KMC simulation varies over the course of the simulation. Therefore, an *a priori* prediction of parallel efficiency on any system, let alone in the extant computing environment, is very difficult. Considering the fact that the KMC method has never been (or very rarely) used on previous as well as existing high-performance computing (HPC) systems, it is hard to comment on what can or cannot be solved using parallel KMC simulations on exascale systems. Nevertheless, there is a great deal of room for improvement and great benefit in initiating the extensive use of HPC systems to carry out parallel KMC simulations. Since GPUs have excellent computing power (FLOPS) per dissipated watt, it is highly likely that ExS will be a heterogeneous computing environment with CPU-GPU systems. Accordingly, testing and implementation of parallel KMC algorithms on a GPU(s) would be the first logical step in porting KMC codes to ExS.

2. Science Challenges to be Solved in the 2020–2025 Time frame Using Extant Computing Ecosystems

One problem that can be addressed and partially solved is performance of radiation damage KMC simulations with microstructurally relevant system sizes to damage (or dose) levels of approximately 1 dpa incorporating high-fidelity defect physics. These high-fidelity models can then be directly compared to neutron-damaged materials using atom probe tomography, high-resolution electron microscopy, and other appropriate materials characterization tools. Depending on the irradiation temperature, this level of damage is sufficient to dramatically affect material properties, so that model validation will provide strong tests of our knowledge of defect physics and mesoscale damage accumulation simulation methods. This is an important step going forward with these models of defect interactions as the basis for understanding radiation damage in materials. This validation must occur in order to gain confidence in the simulation methods and approaches.

3. Science Challenges That Cannot be Solved in the 2020–2025 Time frame Using Extant Computing Ecosystems

Despite the opportunities the exascale computing environment will provide to perform KMC simulations of larger volumes of matter as well as higher fidelity defect physics, there are interesting and important problems that will remain intractable. Microstructure evolution of material subjected to significant temperature or strain gradients will likely not be solvable by KMC methods because the volume of material to be simulated is on the order of several cubic millimeters. Examples include microstructure evolution of nuclear fuel elements where substantial radial temperature gradients exist or at the tip of a growing crack in a structural component where significant stress and strain gradients are present. Furthermore, it is also unlikely that KMC simulations will be able to reach doses prototypical of next-step fusion devices such as the proposed Fusion Nuclear Science Facility or DEMO. KMC simulations will probably be limited to doses ≤10 dpa; this underscores the need to implement a multiscale modeling approach to bridge the gaps connecting atomistic-scale processes to continuum-level property predictions.

4. Top Three Computing Ecosystem Aspects That Will Accelerate or Impede Progress in the Next 5–10 years? Why?

Accelerate	Why?
1. Hardware resources (at all scales)	Speed increases will offer the opportunity to simulate a larger volume of material and the possibility for higher fidelity KMC algorithms.
2. Models and algorithms	Improvements here will allow greater access to the extant ecosystem.
3. Application codes	Improved codes will allow greater and more efficient usage of the ecosystem.

Impede	Why?
1. Visualization and analysis resources	Rendering enormous amounts of exascale data into images on screen may require a prohibitively long time.
2. Data workflow	Challenges handling exascale data files may affect simulation and analysis software stability.
3. Libraries/frameworks	Math libraries that are not optimized for exascale systems may impede KMC code optimization.
4. Other: Programming Models	Requirements for extensive knowledge of the underlying hardware and data movement architectures will make programming, maintaining, and extracting code performance information quite tedious. This will also affect the portability of the code from one HPC system to another.

C.3 White Papers Addressing Discovery Plasma Science

Low-Temperature Plasmas

Cross-Field Transport in Low-Temperature Plasma Systems

FES White Paper – Iain D. Boyd, University of Michigan

1. **Please specify the current science drivers for your field of research.**

The focus of this white paper is on cross-field (magnetic and electric) low-temperature-plasma (LTP) systems. While devices based on this technology have been developed for space propulsion and ion implantation, their basic physical operation is not fully understood. In particular, there is incomplete understanding of the electron transport across magnetic field lines, which is measured to be much higher than that predicted by theory. In state-of-the-art computational modeling, this situation requires use of a phenomenological approach for electron mobility that cannot be established from first principles. The only truly predictive computational path forward appears to be use of fully kinetic simulation approaches. Such calculations are extremely resource-intensive because (1) the geometry is three-dimensional and (2) the system is unsteady in time. Two different computational approaches appear viable: particle-in-cell (PIC) simulations [1] and direct kinetic (DK) methods [2]. Let us estimate the computational resources required for the less expensive PIC simulation for application to the problem of interest.

Space and velocity scales to determine required memory

We consider a typical cross-field device of volume = 10^{-4} m^3 with an average plasma density of 10^{17} m^{-3} and electron temperature of 10 eV, so that the Debye length ~ 10^{-5} m. Hence, spatial mesh resolution in 3D to the Debye length requires a total of

$$Nc = 10^{11} \text{ cells.}$$

For the PIC simulation, let us assume that we require an average of 25 particles per cell for each species and resolution of four main species (e.g., Xe, Xe+, Xe2+, e-); this gives a total number of particles:

$$Np = 100 \times Nc = 10^{13} \text{ particles.}$$

Assuming each particle carries seven pieces of information (3D space and velocity coordinates and a species identifier) and double precision is used throughout, the total memory required is

$$Mt = Np \times 7 \times 8 \text{ Bytes} = 56 \times 10^{13} \text{ Bytes} = 560 \text{ TB.}$$

Temporal scale to determine required computation time

The main oscillating frequency under these plasma conditions is typically about 10 kHz, and we assume it is necessary to resolve 10 oscillations for a total physical time of 1 msec.

PIC requires the time step to be less than the electron plasma frequency such that

$$\Delta t = \text{Debye} \div \text{thermal-velocity} \sim 10^{-12} \text{ sec}$$

However, the heavy particles can be advanced at a much larger time step, say, a factor of 1,000 larger, and electrons will be subcycled. So the total number of primary time steps required to evolve a total physical time of 1 msec is

$$Nt = 10^6$$

On current processors, the performance of the PIC algorithm including electron subcycling is about 5×10^{-7} seconds of CPU time per particle per time step, thus

$$\text{Total solution time} = Np \times Nt \times 5 \times 10^{-7} = 5 \times 10^{12} \text{ sec} \sim 150{,}000 \text{ years!}$$

Computing resource implications

- Memory: in PIC, it is typical to run with 10^6 particles per processor, so 10^7 cores are required.
- Wall time: on 10^7 cores, assuming 50% parallel efficiency, the required wall time is 10^6 sec ~ 300 hours.

2. **Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.**

It is likely that a computation of the size of the PIC problem described above can be accomplished by 2020. However, it is also anticipated that the numerical statistical noise associated with such PIC simulations, which employ 100 particles per cell, will be insufficient to resolve many of the plasma oscillation frequencies of interest. One option for addressing this issue is to use a significantly larger number of particles per cell (which will increase the required memory and run time linearly). Assuming an increase in the size of a PIC simulation that can be accomplished due to hardware improvements of less than a factor of 10 between 2020 and 2025, no more than 1,000 particles per cell will be possible, and that is likely still not enough to suppress the numerical noise sufficiently.

3. **Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.**

Beyond increasing the number of particles per cell, for example, to 10^6, a second option to the limitation described in (2) is to use the DK approach, which does not suffer from statistical fluctuations. However, this method is more memory intensive as it must discretize 3D velocity space for all species and is estimated to require 3 exa-Bytes of memory for the same problem described above. This would require significantly more processors with larger memory capacity.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Models and algorithms	We must be smart about with how we solve the problem; brute force alone will not work.
2. Hardware resources	Problem requires large memory, long run times.

5. References

[1] Birdsall, C.K., and A.B. Langdon, *Plasma Physics Via Computer Simulation*, Adam Hilger Press, 1991.
[2] Hara, K., I.D. Boyd, and V.I. Kolobov, “One-Dimensional Hybrid-Direct Kinetic Simulation of the Discharge Plasma in a Hall Thruster,” *Physics of Plasmas*, 19, 2012, Article 113508.

Low-Temperature Plasma Physics and Exascale Computing

Miles Turner
Dublin City University, Ireland

1 Please specify the current science drivers for your field of research.

1.1 Introduction

The low-temperature plasma community has made little use of computing at scales beyond the desktop. This is not because there are no challenging problems to address. Examples include:

1. Plasma formation and neutral gas flow in low-pressure processing tools
2. Power coupling mechanisms in helicon plasma sources
3. Plasma formation and transport in ion thrusters
4. Atmospheric pressure discharges with complex spatial structure
5. Multiphase flows involving complex chemistry and interaction with liquids, especially in the context of biomedical plasma applications
6. Calculations of cross sections and rate constants required to describe physical and chemical processes in such plasmas.

This list includes examples of both basic science questions and problems of engineering prediction. These challenges are being addressed today, but under constraints that are consequences of limited computer resources, such as reduced dimensionality and simplified physical models. These measures are typically introduced by *ad hoc* approximations, with uncertain implications for the accuracy of the associated computations. At least some of these approximations could be discarded, if consumption of greater computer resources was acceptable. Emerging concerns related to the quality of technical computations—which are addressed by techniques such as verification, validation, and uncertainty quantification—are, if seriously considered, likely to increase (by at least one or two orders of magnitude) the amount of computation required to reach any desired conclusion.

These observations suggest that both the scope and the quality of scientific computations in low-temperature plasma physics could be greatly expanded by exploiting enhanced computer resources. The ratio between the computer capacity (FLOPS, memory) available on the desktop and in an exascale facility is about 10^6, so there is great potential for such increases. Most low-temperature physics laboratories probably have at least limited access to large-scale computation (although not exascale as yet) through national programs but, on the evidence available, are not making much use of them. Why not?

Several reasons suggest themselves:

1. Unstable programming models have discouraged the development of high-performance codes. The low-temperature plasma community is scattered and thinly resourced. Consequently, the development cycle for computer codes can extend over decades (e.g., HPEM [6], PDP1 [17]). However, disruptive changes in the HPC computing model have occurred several times since the development of these codes was initiated. The resources to follow these changes likely have not existed.

2. The community has been insular with respect to methodological improvements that increase the scale of the computing problem. Other communities (fluid dynamics, combustion science) have been energetic in developing techniques that improve the quality of their computations, in various senses. The low-temperature plasma science community (and indeed the plasma community more generally) has not paid much attention to these developments. Indeed, in many ways, basic concepts and approaches first seen in the early 1990s are still the foundation of most computational work, despite the pace of development in cognate fields.
3. Concerns about basic data undermine confidence in predictive power. In most low-temperature plasma physics contexts, prediction involves extensive use of information such as cross-section data, which historically have been intermittently available and of variable quality. This consideration has produced widespread anxiety about the consequences for accuracy of prediction, but not much action to address the issue, with the apparent result that prediction has not been taken very seriously.

1.2 Programming Model

The programming model for high-performance computation has evolved rapidly over the last several decades. For example, there has been a progression from vector supercomputers to parallel networks, and within the last category, a wide variety of network topologies has been employed. There have been corresponding changes in the memory structure. Proper exploitation of these facilities involves code design decisions that occur early in the development cycle. Consequently, a change in the programming model can require code development (or redevelopment) on a large scale. This is practical for national laboratories or similar institutions, but a major obstacle to less intensively resourced communities, such as (but presumably not only) low-temperature plasma physics.

At present, the most common HPC programming model involves computing nodes linked through by message passing (MPI). Each node offers thread level parallelism (OpenMP), and each thread has simultaneous instruction multiple data (SIMD) parallelism. Because MPI and OpenMP are mature and well-maintained standards and SIMD parallelism is (in theory) a matter of data organization and compiler technology, this model seems to offer the prospect of a stable HPC environment. However, recent interest in GPU platforms shows that disruptive changes to the computing model may not be at an end.

In summary, the following questions arise:

1. HPC adoption is impeded by the instability of the computing model, and the mismatch between code development cycle time and the frequency of changes in computing model. Can we assume that the MPI-OpenMP-SIMD model is stable enough to reduce this impediment?
2. Should maintaining a stable computing model be a design criterion for (some) future HPC facilities?

1.3 Methodological Changes

From the late 1990s, or possibly earlier, concerns emerged relating to the integrity of scientific and technical computations, in particular with respect to the frequency of errors in complex codes [4, 3]. These concerns relate primarily not to gross errors leading to flagrantly incorrect results, but to subtle errors that are not readily detected by expert judgment. This is a serious matter, because complex computations are presumably only carried out when the desired conclusions cannot reached by simpler means, including the exercise of expert judgment. Consequently, the integrity of the whole exercise, in particular the conclusions, is undermined if the computational results can reasonably be suspected to be in error in ways that are undetected and unquantified. Investigations carried out in the 1990s showed that such suspicions were reasonable [4, 3], to a degree that tended to undermine the case for technical computation as a serious investigative and/or predictive tool. This alarming outcome motivated the development of techniques designed to prove that computer codes were free of error [9]. Whether this aim has been strictly achieved is disputable, but what is certainly the case is that techniques have been developed that encourage much greater confidence in the correctness of codes. The process of applying such techniques is called verification. In probably unfortunate terminology, a distinction is drawn between "verification of codes," which is concerned with the correctness of computer programs, and "verification of calculations," which is about ensuring that the choices of numerical parameters for a particular calculation are appropriate. Both of these procedures involve investigating the behavior of the simulation as the numerical parameters are varied, and this requires many individual cases to be computed, therefore increasing the computational resources required by a factor of perhaps 10 or 100 [10]. A common situation, however, is that the cases of practical interest are already consuming days to weeks of computer time on a desktop platform and also in all likelihood are being computed near the limit of what the researcher judges to be acceptable in terms of numerical parameters. There are dangerous temptations here, which concern some journals sufficiently that they require convergence data to be made available [12, 2, 15]. If such policies became universal, there would clearly follow either a move to more powerful computer facilities, or a retrogression in the scope of computer calculations.

1.4 Uncertainty Quantification

Even when both code and calculations have been verified, in the senses discussed above, uncertainties remain in simulation results, because the physical parameters used in the calculation are themselves not certain. The number of these parameters can be large, and the associated uncertainties may be significant. For instance, a complex chemistry model may contain hundreds of uncertain rate constants. There are many approaches to characterizing these uncertainties, but all involve evaluating many cases [18, 13, 14, 5, 8]. At best, the number of cases to be simulated will be a few times the number of uncertain coefficients [7, 1].

In practice, the effort devoted to uncertainty quantification must depend on the purpose of the computations. The practical point to be established is that the conclusions being asserted are robust in the presence of uncertainties in the simulation results on which the conclusions are based. If the conclusion is that certain kinds of phenomena are important under certain broadly defined circumstances, then not much attention to uncertainty quantification might be needed. If, on the other hand, the aim is to predict the density of a particular parameter under specified conditions, strenuous attention to uncertainty quantification may be essential. In general, as one progresses from scientific exploration through validation and toward engineering prediction, the effort devoted to certainty quantification should increase.

1.5 Physical Models and Approximations

As noted above, code verification has an important role in raising confidence in the correctness of computations. The most powerful technique for verifying a code is to establish convergence to an exact solution of the physical model that is supposed to be solved [11]. This entails both that such an exact solution is known and that the solution algorithm is sufficiently understood that the rate of convergence can be predicted. The exact solution need not be relevant to the intended domain of application of the code and indeed need not be physical at all. Consequently, the challenge of discovering or constructing an exact solution, while not simple, is not as difficult as it first appears. The position adopted by most writers on the subject of code verification is that this procedure represents the only compelling procedure [9]. If a code has optional or alternative configurations, then a verification test case is needed for each such configuration.

Low-temperature plasma physics simulations must attempt to capture disparate length and time scales, typically within severely constrained computer resources. A widely adopted approach is to construct a modular simulation, in which modules operating at different scales exchange data in some iterative fashion [6]. For example, a common approach is to combine a microscopic (or kinetic) treatment of electrons with a macroscopic (or fluid) description of neutral species. The verification procedure described in the previous paragraph cannot be applied to a simulation of this kind because the exchange of data between modules is informally specified, so there is no clear underlying mathematical model and the convergence of the method has no well-defined order. Moreover, even if these problems could be addressed, the presence of optional or alternative modules in such codes is liable to lead to an impractical proliferation of verification test cases. In short, such codes are, in terms of present recommended procedure, unverifiable.

The search for models that are computationally tractable has led in this case in the direction of models that are intractable to conventional verification techniques. The difficulties apply to both varieties of verification. Since each module typically has its own numerical parameters, the number of such parameters is likely to be large, and demonstrating convergence with respect to all these parameters is likely to be difficult.

In part, this difficult position has arisen because many modular codes originated in the 1990s, or earlier, when no consideration had been given to formal verification procedures. In a context in which computational resources are a less serious constraint and verification has become a concern, some of the model complexity could usefully be unwound in the interest of a more mathematically transparent approach that would facilitate verification. Such an approach would lead to greater consumption of computer resources, but not necessarily to larger wall-clock simulation time.[1]

1.6 Data

The casual handling of basic data by the low-temperature plasma physics community has probably contributed to a widely felt lack of confidence in the predictive utility of computer models [16]. There is a contrast with other fields where computer simulation is used routinely with an expectation of accurate prediction. This includes fields with complex data requirements, such as combustion science. However, some recent evidence suggests that, in plasma chemistry models, the number of critically uncertain parameters may be relatively small (for instance, a few rate constants in a model containing

[1] This is not a paradox. For example, a particle-in-cell implementation written by the author proved faster in wall-clock terms than a COMSOL implementation of a fluid model for some problems. This was achieved of course by efficient use of vector-parallel facilities in the particle-in-cell code.

hundreds). If this proves typical, then the number of parameters in need of critical attention may be tractably small, especially in a context where direct computation of rate constants and cross sections is increasingly practical. Consequently, a combination of uncertainty quantification to identify critical parameters with computation of refined values for those parameters could lead initially to the quantification of predictive uncertainty and then to the elimination of much of this uncertainty, with a consequent improvement in the quality of simulation.

2 Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.

1. Tools for uncertainty quantification when complex chemistry models are employed.
2. Understanding the hierarchy of uncertainty in low-temperature plasma simulations, for example, are we most in need of better atomic and molecular data, or more accurate physical models?
3. Understanding the plasma-liquid interface.

3 Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.

1. Kinetic treatment of coupled neutral and plasma flow in, for example, low-pressure plasma processing tools in three dimensions.
2. Mature tools for on-demand computational of rates and cross sections with accuracy comparable with experiments.
3. First-principles simulation of plasma-surface interaction.
4. First-principles simulation of plasma interaction with living tissue.

4 What top three computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Stable computing model	Facilitates thinly resourced communities
2. Larger computing capability	Ability to employ better models
3. Larger computing capacity	Verification of calculations, uncertainty quantification
Impede	**Why?**
Unstable computing model	Insufficient resources for code development
Access to "capacity computing"	Needed for UQ

References

[1] Francesca Campolongo, Jessica Cariboni, and Andrea Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10):1509–1518, October 2007. ISSN 1364-8152.

[2] Christopher J. Freitas. Editorial. *J. Fluid. Eng.*, 115(3):339, 1993. ISSN 00982202.

[3] L. Hatton. The T experiments: Errors in scientific software. *IEEE Comput. Sci. Eng. Mag.*, 4(2):27–38, 1997.

[4] L. Hatton and A. Roberts. How accurate is scientific software? *IEEE Trans. Softw. Eng.*, 20(10):785–797, 1994.

[5] Kamal Kumar and Chih-Jen Sung. Autoignition of methanol: Experiments and computations. *Int. J. Chem. Kinet.*, 43(4):175–184, April 2011. ISSN 05388066.

[6] Mark J Kushner. Hybrid modelling of low temperature plasmas for fundamental investigations and equipment design. *J. Phys. D: Appl. Phys.*, 42(19):194013, October 2009. ISSN 0022-3727, 1361-6463.

[7] Max D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[8] Sebastian Mosbach, Je Hyeong Hong, George P. E. Brownbridge, Markus Kraft, Soumya Gudiyella, and Kenneth Brezinsky. Bayesian error propagation for a kinetic model of n-propylbenzene oxidation in a shock tube. *Int. J. Chem. Kinet.*, 46(7):389–404, July 2014. ISSN 1097-4601.

[9] William L. Oberkampf and Christopher J. Roy. *Verification and Validation in Scientific Computing*. Cambridge University Press, October 2010. ISBN 9780521113601.

[10] Gregg A. Radtke, Keith L. Cartwright, and Lawrence C. Musson. Stochastic Richardson Extrapolation Based Numerical Error Estimation for Kinetic Plasma Simulations. Sandia Report SAND2015-8620, Sandia National Laboratories, Albuquerque, New Mexico 87185, U. S. A., October 2015.

[11] Patrick J. Roache. Code Verification by the Method of Manufactured Solutions. *J. Fluids Eng*, 124(1):4–10, November 2001. ISSN 0098-2202.

[12] Patrick J. Roache, Kirti N. Ghia, and Frank M. White. Editorial policy statement on the control of numerical accuracy. *J. Fluids Eng.*, 108(1): 2, 1986. ISSN 00982202.

[13] Andrea Saltelli, Marco Ratto, Stefano Tarantola, and Francesca Campolongo. Sensitivity analysis for chemical models. *Chem. Rev.*, 105(7): 2811–2828, July 2005. ISSN 0009-2665, 1520-6890.

[14] Rex T. Skodje, Alison S. Tomlin, Stephen J. Klippenstein, Lawrence B. Harding, and Michael J. Davis. Theoretical validation of chemical kinetic mechanisms: Combustion of methanol. *J. Phys. Chem. A*, 114(32):8286–8301, August 2010. ISSN 1089-5639, 1520-5215.

[15] The Editors. Editorial: Uncertainty Estimates. *Physical Review A*, 83 (4), April 2011. ISSN 1050-2947, 1094-1622.

[16] Miles M. Turner. Uncertainty and error in complex plasma chemistry models. *Plasma Sources Sci. Technol.*, 24(3):035027, 2015. ISSN 0963-0252.

[17] J.P. Verboncoeur, M.V. Alves, V. Vahedi, and C.K. Birdsall. Simultaneous potential and circuit solution for 1D bounded plasma particle simulation codes. *J. Comp. Phys.*, 104(2):321–328, February 1993. ISSN 0021-9991.

[18] Judit Zádor, István Gy. Zsély, Tamás Turányi, Marco Ratto, Stefano Tarantola, and Andrea Saltelli. Local and global uncertainty analyses of a methane flame model. *J. Phys. Chem. A*, 109(43):9795–9807, November 2005. ISSN 1089-5639, 1520-5215.

C.4 White Papers Addressing Verification and Validation

Needs Driven by Experiments

Comparison of Multi-Scale Gyrokinetic Simulation with Experiment: Computational Requirements and Impact on Modeling of Tokamak Plasmas

N.T. Howard (MIT–PSFC)

1. **Please specify the current science drivers for your field of research.**

The existence of a well-developed theory and years of model validation efforts on tokamaks worldwide has dramatically improved our confidence in predicting kinetic profiles of future fusion devices and has established core transport as amongst the most well developed area of tokamak modeling. At the present time, gyrokinetic model validation focuses almost exclusively on comparing ion-scale (long wavelength: $k_\theta\rho_s < 1.0$) simulation with experimental heat fluxes and turbulence measurements. Despite the success of this long-wavelength model, electron heat flux predictions that are robustly lower than those in experiments are not uncommon, and "anomalous" transport in the electron channel has remained poorly understood. Such disagreements have generally been swept aside, suggesting that the "missing" electron heat transport can be recovered by resolving short-wavelength, electron-scale (short wavelength: $k_\theta\rho_s < 60.0$) turbulence. However, due to the extreme computational requirements associated with coupled ion and electron-scale (multiscale) gyrokinetic simulation, this was not demonstrated. As a result, how (and if) ion and electron-scale turbulence couples in current experimental plasma conditions and future reactors remains effectively unknown.

Recent results obtained from high-fidelity, (three gyrokinetic species, collisions, ExB shear, realistic electron mass, etc.) coupled ion ($k_\theta\rho_s < 1.0$) and electron-scale ($k_\theta\rho_s < 60.0$) simulations demonstrate that electron-scale turbulence can play an important, even dominant, role in the core of standard Alcator C- Mod, L-mode plasmas, and that significant ion-electron cross-scale coupling exists [1]. In simulations of experimental plasma conditions, coupled interactions between ion and electron-scale turbulence have been observed, in some conditions, to increase the simulated electron heat flux by nearly a factor of 10 and to increase the ion heat flux by a factor of 2 relative to corresponding long-wavelength simulations [2]. These results call into question the validity of using a long-wavelength model for the prediction of kinetic profiles of ITER, as the parameter space where such cross-scale coupling is important is unknown. In order to address this open question, significant computing resources need to be dedicated to comparing, multiscale gyrokinetic simulations to experiments over a wide range of input parameters. The goal is to both understand the physical processes responsible for the coupling between ion and electron-scale turbulence and to identify regions of parameter space where cross-scale coupling is important. Ultimately, these results may greatly influence the prediction of plasma profiles on ITER and will result in improved confidence and accuracy in these predictions.

2. **Describe the science challenges expected to be solved using extant computing ecosystems.**

With significant dedication of computing resources, and further advances in supercomputing, (currently amounting to an approximate doubling of computing every year), the open questions regarding multiscale simulation of plasma turbulence can be addressed in the 2020–2025 time frame. The current computational requirements for performing multiscale gyrokinetic simulations are extreme, and until recently were effectively inaccessible. Multiscale simulations must simultaneously resolve both the ion and electron spatiotemporal scales. Resolving spatial scales spanning more than 2–3 orders of magnitude ($k_\theta\rho_s \sim 0.1$–60.0) and temporal scales associated with the linear growth rates of short wavelength turbulence (~60 times larger than long-wavelength turbulence) requires extremely high spatial and temporal resolutions. All particle species simulated must be fully gyrokinetic. To make direct comparisons with experiments, an impurity species, effects of rotation, ExB shear, realistic electron mass, and collisions all must generally be included, and all of these increase computational requirements. To date, eight simulations meet the requirements outlined above and been quantitatively compared with experiments. Each simulation required approximately 15 million CPU hours, while utilizing approximately 17,000 processors (~37 days on the NERSC Edison supercomputer). However, future simulations will likely need to be performed with even higher physics fidelity (finite-beta, larger simulation boxes, higher resolutions) and therefore will have increased computational requirements. With the development of more efficient algorithms and increased computing capabilities, it should be possible to rigorously validate the multiscale gyrokinetic model against a wide range of existing plasma conditions in the 2020–2025 time frame.

3. Describe the science challenges that cannot be solved using extant computing ecosystems.

Even with projected increases in computation in the next 5 to 10 years, it may still be the case that multiscale gyrokinetic simulation itself will not be used for **routine** prediction of tokamak profiles. Therefore, the goal for the next decade should be understand to validation the multiscale gyrokinetic model and use the physics obtained from these large multiscale simulations to guide the development of physics-based, reduced transport models. Such reduced models could address the need for nearly real-time profile prediction of fusion devices until routine use of multiscale gyrokinetic simulation becomes feasible.

4. What *three* computing aspects will accelerate/impede your progress in the next 5–10 years?

Accelerate	Why?
1. Upgrades to hardware resources	Increased number/faster cores and increased memory will reduce time required for computation.
2. Improved models and algorithms	Better scaling of continuum codes to larger processor counts reduces time required for computation.
3. Emphasis on a computing paradigm geared toward rigorous model validation	Prioritization of rigorous model validation will increase resources to this computing paradigm.

Impede	Why?
1. Prioritization of computing resources to only large (but short) capacity computations.	Placing such large computations as the highest priority removes resources for computations better suited for rigorous model validation.

In addition to the obvious benefits provided by improved algorithms and upgrades to hardware resources, increased emphasis on a computing paradigm that promotes rigorous model validation is crucial to addressing the open questions presented above. Single multiscale simulations are of only limited benefit because they do not allow for any investigation into model sensitivities within experimental uncertainties and therefore provide limited ability to extrapolate to even slightly different conditions. In reality, large dedicated computing grants in the range of 100 million CPU hours are required to perform a relatively rigorous assessment (including a handful of parameter scans) of a **single** plasma condition. Completion of such studies will provide more information on sensitivity of cross-scale turbulence coupling and will allow for a more complete validation of the multiscale gyrokinetic model by exploring results within experimental uncertainties. However, emphasis at many computing facilities is generally focused on large-capacity computations that require a majority of the site's capabilities, but 24 hours or less for completion. In contrast, a set of parameter scans (bundled to make reach capacity processor counts or individually submitted) composed of large, 17,000–50,000 processor jobs, running for longer time periods (as outlined in Section 2), are the only means to provide a rigorous test of the multiscale gyrokinetic model. In order to validate the gyrokinetic model and assess the importance of cross-scale turbulence coupling, increased emphasis and resources must be dedicated to a computing paradigm geared toward rigorous comparison with experiments and not based solely on short, large-capacity computations.

5. References

[1] N.T. Howard, et al., Physics of Plasmas **21**, (2014) 112510
[2] N.T. Howard, et al., Nuclear Fusion **56**, (2016) 014004

Data Management, Analysis, and Assimilation

John Wright, Massachusetts Institute of Technology

Data from simulations and experiments are the common currency that underlies scientific research. These data are useful only to the extent that their meaning can be conveyed and preserved. However, ultrascale computing, new computer architectures, the growing complexity of scientific processes, and the increasing importance of extended collaborations challenge traditional approaches to data assimilation, analysis, and visualization. These issues are discussed in depth in the recent DOE and ASCR sponsored report on Integrated Simulations for Magnetic Fusion Energy Sciences [Bonoli 2015].

1. **Please specify the current science drivers for your field of research.**

We have found that the issues of data management are generic to the scientific processes regardless of field, though we address these issues in this white paper in the context of our area of research, magnetic fusion energy. For experimental fusion data, the MDSplus system [mdsplus] has emerged as a *de facto* standard. It includes a level of metadata and self-description enabling it to drive a work-flow engine; however, in practice, a good deal of analysis is also carried out under manual control, leaving the provision of the more complete level of metadata up to individual users. For example, MDSplus may be used to extract experimental data for a simulation. At this point, the provenance of the data has been lost and the workflow proceeds manually. The simulation inputs may be prepared on a local desktop or workstation and then uploaded to a leadership compute platform where the simulation is executed with results or a portion of them being downloaded to a local workstation for further analysis and perhaps being included in a publication. To address this issue of tracking provenance through disjointed and physically separated experimental and simulation workflows, a project was initiated in 2012 [Schissel 2014, Greenwald 2015]. Called the MPO, for Metadata, Provenance and Ontology, this project has developed a web service with an API that allows users to instrument any analysis script with callouts that automatically populate a database, documenting their scientific workflows to their preferred level of detail [mpo].

2. **Describe the science challenges expected to be solved in the 2020–2025 time frame using extant computing ecosystems.**

Real-time remote control of experiments, a degree of work-flow data provenance recording connected to data persistence.

3. **Describe the science challenges that cannot be solved in the 2020–2025 time frame using extant computing ecosystems.**
   - Adoption of a data model and ontology standard to facilitate systemization of queries and collaboration.
   - In situ visualization and steering of large-scale simulations.
   - Integration of VV/UQ into simulation workflows in pre- and posteri analysis and in codesign code development.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Data workflow (including sharing, transmitting, archiving, and so on)	Simulations and data need reproducibility and the ability to be searched and mined later, particularly in collaborations.
2. Models and algorithms	Codesign to incorporate latest applied math algorithms into physics codes at the beginning to enable access to next-generation architectures.
3. Visualization and analysis resources	Enabling understanding of large data sets and ensemble simulations. UQ to help steer towards the next simulation.

Impede	Why?
1. Application codes (implementation, development, portability, and so on)	Porting to new accelerator architectures, dealing with concurrency.
2. Hardware resources (at all scales) including I/O, memory, and so on.	High-concurrency architectures, I/O limited may require rethinking of present coding approaching.
3. Capacity versus capability computing	Computing centers often emphasize capability computing, resulting in long queue times for workhorse capacity computation simulations.

5. **Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.**
Visualization, in situ analysis, and VV/UQ (verification, validation and uncertainty quantification) are aspects of data management and analysis that only become more acute at larger computational scales. Presently, large-scale simulations on leadership class hardware are beginning to be bandwidth limited and use specialized I/O libraries such as ADIOS [Liu 2014]. Movement of data from computational centers to local resources for additional analysis is already burdensome in many instances, and the same issues appear when working with remote experimental fusion facilities, especially in cases of remote driving of experiments where real-time analysis of data is needed before the next *shot*. Optimization of long network routes can lead to notable improvement in data transfers (e.g., U.S.–Korean KSTAR collaboration) as well as in work-flow software frameworks that are selective about data transferred such as ICEE [Wu 2015]. Fastbit [Wu 2006] allows fast index-based queries of high dimensional data from simulations.

6. **References**

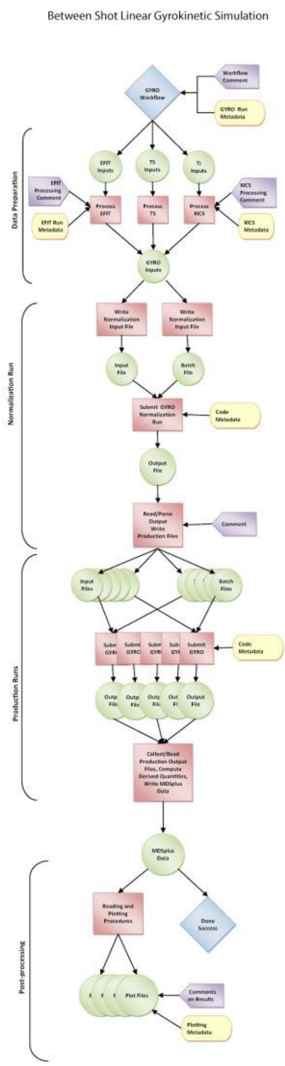
i. [Wu 2006] Kesheng Wu, Ekow J. Otoo, and Arie Shoshani, "Optimizing bitmap indices with efficient compression," ACM Transactions Database Systems, **31**, pp. 1–38 (2006).

ii. [Bonoli 2015] "Report of the Workshop on Integrated simulations for magnetic fusion energy sciences," Sec. 5.3, Chairs Paul Bonoli and Lois Curfman McInnes, 2015, (http://science.energy.gov/~/media/fes/pdf/workshop-reports/2016/ISFusionWorkshopReport_11-12-2015.pdf).

iii. [Greenwald 2015] M. Greenwald, D. Schissel, and J. Wright, "An unmet need: Documenting complex scientific workflows – end to end," Report of the Workshop on Integrated Simulations for Magnetic Fusion Energy Sciences, Chairs Paul Bonoli and Lois McInnes, 2015.

iv. [mdsplus] www.mdsplus.org.
v. [Schissel 2014] D.P. Schissel, et al., “Automated Metadata, Provenance Cataloging, and Navigable Interfaces: Ensuring the Usefulness of Extreme-Scale Data,” Fusion Engineering and Design **89** 745–749 (2014).
vi. [mpo] https://mpo.psfc.mit.edu
vii. [Liu 2014] Liu, Q., Logan, J., Tian, Y., Abbasi, H., Podhorszki, N., Choi, J. Y., Klasky, S., Tchoua, R., Lofstead, J., Oldfield, R., Parashar, M., Samatova, N., Schwan, K., Shoshani, A., Wolf, M., Wu, K. and Yu, W. (2014), Hello ADIOS: the challenges and lessons of developing leadership class I/O frameworks. Concurrency Computat.: Pract. Exper., 26: 1453–1473. doi:10.1002/cpe.3125.
viii. [Wu 2006] Kesheng Wu, Ekow J. Otoo, and Arie Shoshani, Optimizing bitmap indices with efficient compression, ACM Transactions Database Systems, 31 (2006), pp. 1–38.
ix. [Wu 2015] “White Paper on State of the Artin Distributed Area (DA) and In Situ Workflow Management ICEE –a Low-Latency Distributed Workflow System,” John Wu, Scott Klasky, CS Chang, Jong Choi, Alex Sim, and Michael Churchill. ASCR Workshop on the Future of Scientific Workflows, http://extremescaleresearch.labworks.org/events/workshop-future-scientific-workflows (2015).

7. (Optional) Images (next page)



Workflow for preparing inputs and running a gyrokinetic simulation [Figure 34 from Bonoli 2015], demonstrating the provenance chain between fits to experimental data to plots of simulations used in publications.

APPENDIX D: FUSION ENERGY SCIENCES CASE STUDIES

The following case studies were submitted by the authors listed below in advance of the Exascale Requirements Review to guide both the agenda and meeting discussions.

This page is intentionally left blank.

Simulation of a Disrupting Tokamak Plasma

S. C. Jardin, SciDAC Center for Extended-MHD Modeling (CEMM)

1. Description of Research

1.1 Overview and Context

The tokamak is a toroidal magnetic "bottle" that magnetically confines high-temperature plasma. Much of the magnetic field is produced by an electrical current induced in the plasma itself. In the next-step device ITER, this electrical current will be on the order of 10 million amperes. The tokamak is a pulsed device with a pulse length of 10s of minutes, so this current needs to be induced over a period of several minutes to start the discharge, and it normally takes several minutes to shut down this current in a controlled manner at the end of the discharge. If one tries to shut down the current too fast, it can create unstable conditions that lead to a disruption: a rapid termination of the discharge in which all confinement is lost and the plasma energy and current are rapidly transferred to the surrounding metallic vessel. This can produce large stresses, which potentially can damage the vessel. We are modeling this process by solving the 3D magnetohydrodynamic (MHD) equations in realistic tokamak geometry using the M3D-C1 code. Similar calculations are performed with the NIMROD code, which is also part of the CEMM SciDAC.

1.2 Research Objectives for the Next Decade

Our goals are to validate our model by performing many simulations and comparing them with existing tokamak experiments. Once the model is well validated, emphasis will shift to performing predictive runs for ITER. This is much more challenging because ITER is much larger than today's tokamaks, and the higher plasma temperature and stronger magnetic field exacerbate the range of spatial and temporal scales that need to be modeled.

2. Computational and Data Strategies

2.1 Approach

We solve the 3D MHD equations as an initial value problem using high-order C^1 finite elements in 3D. We initialize the simulation with a fully formed tokamak plasma in equilibrium, and then begin inductively reducing the plasma current to determine under what conditions a disruption occurs, and compare this with the corresponding experimental result.

2.2 Codes and Algorithms

We use implicit time-stepping to avoid the otherwise severe time step constraint based on the fast MHD waves and the smallest element size. In each time step, a large sparse matrix is formed by linearizing around the present solution, and this sparse matrix is solved to advance from one time step to the next.

3. Current and Future HPC Needs

3.1 Computational Hours

A typical run now on Edison or Cori-Phase1 uses 6,000 processors and takes about 150 hours. This is normally run in 15 restart segments of 10 hours each. Each restart segment may wait several days in the queue, so that the entire calculation will normally take several months to complete.

3.2 Parallelism

We presently use MPI and vector parallelism. In each time step, about half the time is spent in populating the matrices and the other half in solving the matrices using PETSc linear solver routines and their preconditioners.

3.3 Memory

Only about 4 GB of memory is required to store all the scalar fields for one time step. However, the PETSc preconditioners we use require a lot of memory, so that we are normally memory limited. Thus, if each Cori node has 128 GB and we are using 200 nodes, that is about 25 TB of memory.

3.4 Scratch Data and I/O

Our restart files are about 4 GB. We typically save about 50 time slices of 4 GB each for a total of 200 GB for the entire calculation. I/O is a small fraction of the run time, less than 1%.

3.5 Long-term and Shared Online Data

We need to save all the data from 2–3 calculations presently being worked on so we need about 0.5 TB of online data.

3.6 Archival Data Storage

If we generate 4 TB of data per year and keep it for 10 years, we will need 40 TB of archival storage.

3.7 Workflows

This year we were allocated 20 M hours for this project. We could use 80 M hours by 2020 and 400 M hours by 2025.

3.8 Many-Core and/or GPU Readiness

We presently do not use GPUs. The part of our code that populates the matrices uses MPI and vectorization. We are presently experimenting with introducing OpenMP into our code as a third level of parallelization, but it will likely require major restructuring to be effective. We presently use PETSc to solve our matrices and are not aware of any plans by the PETSc group to take advantage of many-core or GPUs. We are experimenting with adding the option to use Trilinos rather than PETSc to evaluate the relative efficiency of that library.

3.9 Software Applications, Libraries, and Tools

We are presently limited both in speed and in memory by the linear solvers and associated preconditioners in the PETSc library. Improvements in linear solvers and their efficiency would greatly benefit our project.

3.10 HPC Services

We could use more allocated hours and shorter queue wait times.

3.11 Additional Needs

Because our jobs require long run times and we must perform them in a number of restarts, it is important to us that the queue wait times be short. We often have to wait several days for a restart submission to start, and that severely affects our productivity.

4. Requirements Summary Worksheet

Table 1 shows our projected HPC requirements.

Table 1. M3D-C1 Requirements

Code: M3D-C1	Column 1: Current Usage	Future Usage: 2020 (As a factor of column 1)[d]	Future Usage: 2025 2 (As a factor of column 1)[d]
Computational core hours (Conventional)[a]	20 M/year	4×	16×
Computational node hours (Homogeneous many-core)[b]			
Computational node hours (w/GPU or accelerator)[c]			
Memory per node	128 GB	GB	GB
Aggregate memory	25 TB	TB	TB
Data read and written per run	0.2 TB	TB	TB
Maximum I/O bandwidth needed	NA	GB/sec	GB/sec
Percent of runtime for I/O	1	1×	1×
Scratch file system space needed	0.2 TB	TB	TB
Permanent online data storage	0.5 TB	TB	TB
Archival data storage needed	40 TB	TB	TB

[a] “Core hours” are used for “conventional” processors (i.e., node-hours * cores_per_node). Intel “Ivy Bridge” is an example conventional processor.

[b] “Node hours” are used for homogenous many-core architectures. A self-hosted Intel Xeon Phi “Knights Landing” is an example.

[c] “Node hours” are used for “GPU or accelerator” usage.

[d] For example, 32 × column 1.

Exascale Modeling of Compact Plasma-Based Ion Acceleration Concepts

J.-L. Vay, H. Vincenti, S. Bulanov, B. Loring, and O. Rübel

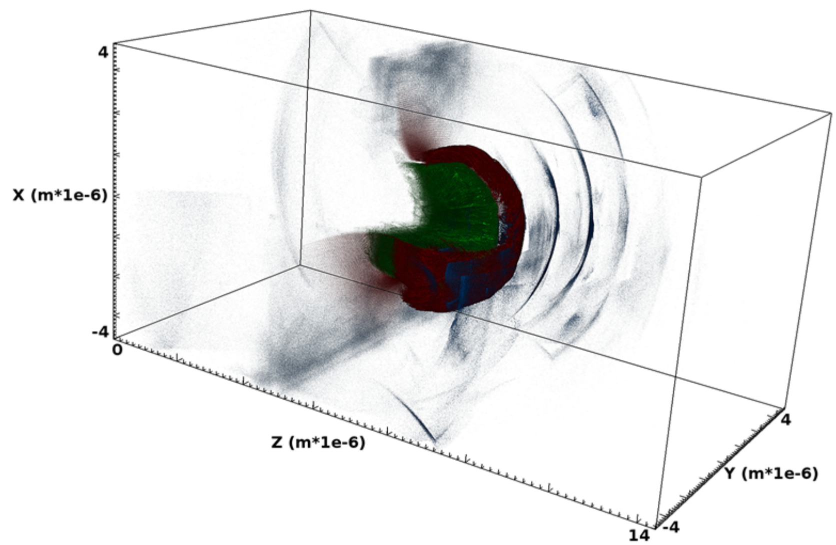


Figure 1: Visualization of the kinetic energy (opacity proportional to magnitude) of protons (red), carbon (green), and electrons (blue) from a 3D Warp simulation of ion acceleration driven by laser. The rendering was performed in situ, as the simulation was progressing, using the toolkit WarpIV (https://bitbucket.org/berkeleylab/warpiv), which combines the particle-in-cell and visualization frameworks Warp and VisIt.

1. Description of Research

1.1 Overview and Context

Understanding the interaction of relativistic laser pulses with matter is a grand challenge that is motivated by applications such as the development of compact, inexpensive plasma-based particle accelerators for medical applications (electrons- or ions-based), and sources of radiation for science, industry, and medicine. Large 3D numerical simulations offer the opportunity to quickly identify optimal regimes of acceleration at moderate cost and will be essential to ensure the success of current and upcoming accelerator facilities (e.g., Bella-i). In particular, exascale computing will bring the computational power that is needed to achieve the detailed three-dimensional numerical studies, including parametric optimizations, that are necessary but currently not possible on existing petascale supercomputers.

1.2 Research Objectives for the Next Decade

The modeling of laser-plasma ion acceleration is currently limited to 2D at high resolution on petascale supercomputers, which is insufficient to quantitatively and sometimes even qualitatively describe the physics at play. Exascale computers will bring the required computational power to enable the 3D simulations at high resolution that are needed to accurately represent the physical phenomena. In each case, parametric studies for tolerance to non-ideal effects (jitter, asymmetries, and so on) that are essential for the design of a production-level accelerator will require, and even stress, exascale resources.

2. Computational and Data Strategies

2.1 Approach

Computational and Data Problems. The codes use the particle-in-cell (PIC) method in which the plasmas and beams are described as collections of charged macroparticles that evolve self-consistently with their electrostatic or electromagnetic fields, which are resolved on an Eulerian mesh. The electromagnetic PIC approach has been shown to scale to hundreds of thousands of cores or beyond when the distribution of macroparticles is regular. Realistic distributions, however, can be highly irregular spatially, which eventually significantly limits the actual scaling for production runs.

Strategies Used to Solve Them. Efficient dynamic load-balancing algorithms using an internode MPI load balancer and an intranode OpenMP implementation, ultimately coupled to the use of Adaptive-Mesh-Refinement, that scale to very high concurrency will be needed. This may necessitate some code restructuring (especially for upcoming many-core-based architecture) with the introduction of new data structures grid/particles tiling (completed) that are chunks of work that fit in a core cache and can be efficiently handled independently by different OpenMP threads. Sorting per cell can also further increase memory locality and improve vectorization performance and is also under consideration.

2.2 Codes and Algorithms

The PIC code Warp and the PICSAR kernel from the Berkeley Lab Accelerator Simulation Toolkit will be advanced to conduct the simulations at the exascale level. Both codes use FORTRAN-77/95 for the most compute-intensive subroutines. PICSAR is a kernel of Particle-In-Cell functionalities (preexisting in Warp) that have been optimized for multicore and many-core CPUs at the intranode level (better memory locality, efficient and portable vectorization on SIMD architecture, good share memory implementation using OpenMP) and internode level (in progress: optimization of MPI exchanges and communicator topology, dynamic load balancing). Warp uses Python at the front end for code steering, user programmability, and fast prototyping and calls PICSAR FORTRAN routines via Python.

3. Current and Future HPC Needs

3.1 Computational Hours: N/A.

3.2 Parallelism:

Presently, parallel execution has relied mostly on domain decomposition with MPI communications. Hybrid MPI+OpenMP parallelization of PICSAR is under way within the NESAP program in preparation for NERSC's supercomputer Cori.

3.3 Memory: N/A.

3.4 Scratch Data and I/O: N/A.

3.5 Long-term and Shared Online Data: N/A.

3.6 Archival Data Storage: N/A.

3.7 Workflows: N/A.

3.8 Many-Core and/or GPU Readiness

Hybrid MPI+OpenMP parallelization is under way within the NESAP program in preparation for NERSC's supercomputer Cori, and several developments have been made to meet the requirements of future CPU multicore/many-core architectures and, in particular, to ensure memory locality. New data structures (particle/grid tiles) have been implemented in the code to ensure memory locality and excellent cache reuse. Tests on intel KNC and intel Haswell showed a cache reuse of 99% due to the new implementation. We also designed a new vector version of hotspots routines of the code that will be essential to achieving good performances on future many-core architectures that will have reduced clock speed and larger vector data registers.

Acceleration using GPU is also being explored with collaborators and may become more portable through advanced concepts being developed within the OpenMP standard such as the new "flexible" construct as well as "target" syntax.

To fully enable advanced algorithms with significant multithreading, restructuring of code and different data storage patterns may be required. Tuned algorithms or kernels for solving FFTs, finite-difference operations, multigrid, scatter/gather will also be key to developing particle-mesh codes that will use exascale supercomputers efficiently.

3.9 Software Applications, Libraries, and Tools: N/A.

3.10 HPC Services: N/A.

3.11 Additional Needs

Warp uses Python as the front end for code steering, user programmability, and fast prototyping. Producing a single executable that includes all the compiled and interpreted Python scripts is a complex task, and the loading of the various Python modules (shared libraries and Python scripts) at runtime is highly preferable. It is thus very important that the supercomputer and its environment support fast load of shared libraries at very high concurrency.

4. Requirements Summary Worksheet

Table 1 shows our projected HPC requirements.

Table 1. WARP/PICSAR Requirements

Code: ________Warp/PICSAR___________	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 2 (As a factor of column 1)[d]**
Computational core hours (Conventional)[a]	5 M	20×	100×
Computational node hours (Homogeneous many-core)[b]	200 k	20×	100×
Computational node hours (w/GPU or accelerator)[c]	N/A	TBD	TBD
Memory per node	128 GB	4×	20×
Aggregate memory	10 TB	10×	50×
Data read and written per run	20 TB	10×	50×
Maximum I/O bandwidth needed	0.5 GB/sec	4×	20×
Percent of runtime for I/O	80	80	80
Scratch file system space needed	50 TB	4×	20×
Permanent online data storage	200 TB	4×	20×
Archival data storage needed	500 TB	4×	20×

[a] "Core hours" are used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" are used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" are used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

This page is intentionally left blank.

U.S. DEPARTMENT OF
ENERGY