Report of the Snowmass T7 Working Group on High Performance Computing*

Conveners: K. Ko (SLAC), R. Ryne (LBNL), P. Spentzouris (FNAL)
August 10, 2001

Charge to the group: Computers have played a larger and larger role in the theory, design and development of accelerators and the associated technologies. Some examples are calculations of beam optics, simulation of instabilities, electromagnetic field calculations, simulation of space-charge dominated beams and halo formation, beam-beam simulations, start-to-end simulations of systems, real-time modeling of accelerators, and simulations of new accelerator ideas such as those involving lasers and plasmas. This group should explore the impact that advanced computational techniques using the most powerful computers would have on research and development in particle beams and accelerator technology. The group should document past success and look at the immediate and long term future of high performance computing (HPC) as applied to particle beams and accelerator technology. In particular the group should outline a program of proposed research which will bring the world's most powerful computers, and the hardware and software technologies associated with them, to bear on the most challenging and important problems in our field.

Outline

Executive summary
1. Introduction
2. HPC requirements for next-generation accelerators
    2.1 High intensity proton driver
    2.2 Next-generation linear colliders
    2.3 Very large hadron colliders
    2.4 Neutrino sources/muon colliders
    2.5 Laser- and plasma-based accelerators
3. State-of-the-art in HPC simulation of accelerator systems
    3.1 Electromagnetic modeling
    3.2 Beam simulation
    3.3 Laser- and plasma-based systems
    3.4 Particle production codes
    3.5 Ionization cooling simulation codes
4. Plans related to the Scientific Discovery through Advanced Computing (SciDAC) project
    4.1 Project overview
    4.2 Collaboration with the applied mathematics and computer science communities
5. Conclusion
6. List of participants
7. List of talks

* Website: http://snowmassserver.snowmass2001.org/Working_Group_T7/

Executive Summary

Particle accelerators are among the largest, most complex, and most important scientific instruments in the world. They have enabled a wealth of advances in applied science and technology, many of which have huge economic consequences and many of which are greatly beneficial to society. They are also critical to research in the basic sciences (such as high energy physics, nuclear physics, materials science, chemistry, and biology). In particular, accelerators are the most versatile and powerful tools for exploring the elementary particles and fields of the universe. Experiments associated with high energy accelerators led to some of the most remarkable discoveries of the 20th century. Near-term experiments are likely to be just as exciting, if not more so, with the possible discovery of new physics beyond the Standard Model, such as supersymmetry and its associated implications for a radical new geometry of space-time, which will fundamentally change our view of the universe.

Given the great importance of particle accelerators, it is imperative that the most advanced computing technologies be used for their design, optimization, commissioning, and operation. The objective of the High Performance Computing (HPC) Working Group is to understand the modeling needs for current and future accelerator technology, identify the HPC hardware and software technologies required for such modeling, and outline a plan for the development of these technologies. The following summarizes the HPC requirements for next-generation accelerators and describes an action plan that responds to the identified needs.

HPC requirements for next-generation accelerators
All near- and far-future accelerator designs have very challenging modeling requirements that require HPC:

High intensity proton drivers needed for conventional neutrino "Superbeams", neutrino factories, and muon colliders require precise predictions of the effects of space charge. This need is shared by currently operating proton drivers, like the FNAL Booster and the BNL AGS, which are experiencing significant losses, currently attributed to space charge effects at injection. The losses at the FNAL Booster are currently the biggest issue for the success of the near future FNAL program (RunII+neutrino program). Due to the nature of this type of problem – which involves long (high aspect ratio) bunches propagating for thousands of turns including space-charge effects and wakefield effects – a full 3D simulation is prohibitive using the current algorithms and existing multi-processor hardware. Simulations using roughly 100 processors have been estimated to require 1 year of computer time.

Next-generation linear colliders require demanding computer simulations in regard to both electromagnetic and beam dynamics modeling. For example, extremely complicated 3D electromagnetic structures for the NLC must be modeled and analyzed with greater speed, accuracy, and confidence than has previously been possible. Presently popular serial electromagnetics codes are inefficient in handling complex geometric shapes, or are limited in their ability to solve large-scale problems. However, the recent development of parallel eigenmode and time-domain codes has already increased our modeling capabilities by roughly three orders of magnitude. In addition to modeling electromagnetic components, HPC capabilities are needed to model beam dynamics in linear colliders. For example, in both the NLC and TESLA designs, the accurate treatment of space-charge effects and other collective effects is important to predicting the beam’s behavior in the damping rings. In order to validate the basic operational characteristics of these machines, the linac and beam delivery systems need to be modeled including component fluctuations, tuning, and feedback systems. Such simulations are impossible on serial computers, where the execution time to run one such code with the desired accuracy has been estimated to be 1 year per processor.

Very large hadron colliders like the VLHC require HPC capabilities in areas such as long-term tracking to predict dynamic aperture, self-consistent simulations of beam-beam effects in the strong-strong regime, predicting the thresholds for instabilities (such as the electron-cloud, resistive wall, and transverse mode coupling instability), and the simulation of beam/material interactions (e.g. energy deposition from collision byproducts) that address safety and environmental issues. In addition to these “conventional” requirements, there are also “operational” ones involving the use of HPC to develop orbit correction algorithms, alignment procedures, etc., that are challenging due to the size of the machine, the large amount of diagnostic data, and the short period of time in which the analysis has to be performed. Here accelerator simulation is used in a similar way to HEP experiment simulation: accurate modeling of the machine and diagnostics are used to develop and optimize analysis algorithms, which are then used to optimize machine operation. Like a linear collider, full system simulations of the VLHC including beam dynamics and feedback systems are needed to verify operational characteristics of the proposed design.

Neutrino source/muon colliders present unique modeling challenges due to the fact that they involve ionization cooling. Ionization cooling requires accurate modeling of muon/matter interactions, especially energy loss and multiple scattering. There are a few codes that share the physics description of the above processes borrowed (or directly implemented) from HEP modeling packages. These codes are very slow, prohibiting accurate simultaneous optimization of the sub-systems of the design, although in many cases both performance and cost of these sub-systems are dependent on each other. In addition, for high intensity muon colliders space charge effects are crucial at the final stages of cooling. In both cases HPC is needed. An initial effort to embed cooling simulation capability in an HPC beam dynamics code has been successful, providing a good base for further development.

Besides the design of next-generation accelerator complexes, HPC is also needed, in concert with theory and experiment, to explore and develop novel methods of acceleration like *plasma-based and laser-based acceleration* techniques. Using these techniques, extremely high gradients (up to 100 GV/m) have been measured over short distances in the laboratory. The challenge is to control and stage high-gradient sections so that one can produce high quality, high energy beams in a less costly, more compact configuration that would be impossible using conventional technology. Beyond applications to HEP, such compact accelerators would have huge consequences in others areas of basic and applied science, industry, and medicine. However, modeling these complex systems requires solving the 3D coupled Maxwell/Vlasov equations. Given that the phenomena involve multiple length and time scales (a situation that is particularly challenging when the laser wavelength must be resolved), 3D simulations can only be performed using HPC resources. As an example, the simulation of a 1 GeV plasma accelerator stage using a fully explicit PIC code has been estimated to require 10,000 to 100,000 CPU hours for a single run.

Action Plan

Recognizing the challenges posed by these and other projects, a SciDAC (Scientific Discovery Through Advanced Computing) project on 21st Century Accelerator Simulation was approved in mid-2001. The primary objective of this national R&D effort is to establish a comprehensive terascale simulation capability for the US Particle Accelerator Community. The success of this effort, which is supported by both HENP and ASCR, will involve close collaboration of accelerator physicists with applied mathematicians, numerical analysts, and computer scientists to develop new theoretical formulations and new algorithms capable of high performance and scalability on massively parallel systems. In particular, the accelerator community will utilize HPC tools for mesh generation, mesh refinement, particle/mesh methods, multi-level PDE solvers, eigensolvers, performance optimization, software component integration, and visualization. Many of these tools will be developed in the SciDAC Integrated Software Infrastructure Centers. Code verification and validation will require collaboration of code developers working with researchers performing controlled, well-instrumented experiments.

As a result of the Snowmass meeting, a plan defining the necessary first steps needed to respond to the design needs of the next generation machines was formulated. This plan includes further development of HPC space charge codes for circular machines, a 3-month code comparison effort to test the accuracy and validity of the various models, and the simulation of existing proton drivers such as the FNAL Booster and the BNL AGS. The plan also includes continued code development needed to treat, on parallel computers, physical effects such as the beam-beam interaction, collisions, wakes, and coherent synchrotron radiation. In regard to electromagnetic modeling, the plan includes the development of a parallel statics solver, the treatment of lossy structures, surface effects, and the direct calculation of wakefields. In regard to laser- and plasma-based accelerators, the plan includes the development of a family of codes (fluid and particle) of varying complexity and capabilities, the most demanding of which are fully 3D parallel PIC codes, with moving windows and dynamic mesh capabilities, that have packages to include physical effects such as ionization of multiple species and the simultaneous treatment of laser and particle beams. The Working Group also addressed the issue of code integration, including the need to develop reusable software components and the need to adopt standards for exchange of data and interoperability between those components.

Conclusion

The accelerator community is well positioned to develop a comprehensive terascale capability that will utilize the latest advances in HPC technologies. Such a capability will help insure the success of future accelerators, by facilitating design decisions aimed at controlling and reducing cost, reducing risk, and optimizing performance. The use of terascale simulation, combined with theory and experiment, will provide greater understanding of the complex, nonlinear, multi-scale, and many-body phenomena encountered at the frontier of accelerator technology.

1. Introduction

The T7 Working Group on High Performance Computing (HPC) had more than 30 participants (listed in Section 6). During the three weeks at Snowmass there were about 30 presentations (listed in Section 7). This working group also had joint sessions with a number of other working groups, including E1 (Neutrino Factories and Muon Colliders), M1 (Muon Based Systems), M6 (High Intensity Proton Sources), T4 (Particle Sources), T5 (Beam dynamics), and T8 (Advanced Accelerators). The topics that were discussed fall naturally into three areas: (1) HPC requirements for next-generation accelerator design, (2) state-of-the-art in HPC simulation of accelerator systems, and (3) applied mathematics and computer science activities related to the development of HPC tools that will be of use to the accelerator community (as well as other communities). This document summarizes the material mentioned above and includes recommendations for future HPC activities in the accelerator community. The relationship of those activities to the HENP/SciDAC project on 21st century accelerator simulation is also discussed.

2. HPC Requirements for next-generation accelerators

During the Snowmass meeting there were several presentations and discussions related to HPC requirements for next-generation accelerators. These dealt with high intensity proton drivers, next-generation linear colliders, very large hadron colliders, and neutrino sources & muon colliders. In addition to the discussion of requirements for major facilities, a joint session was held with Working Group T8 to discuss HPC needs for advanced accelerator concepts. These discussions are summarized in the following subsections.

2.1 High Intensity Proton Drivers and Muon-Based Accelerators

High intensity proton drivers needed for conventional neutrino "Superbeams", neutrino factories, and muon colliders require precise predictions of the effects of space charge. This need is shared by currently operating proton drivers, like the FNAL Booster and the BNL AGS, which are experiencing significant losses, currently attributed to space charge effects at injection. The losses at the FNAL Booster are currently the biggest issue for the success of the near future FNAL program (RunII+neutrino program). Due to the nature of this type of problem – which involves long (high aspect ratio) bunches propagating for thousands of turns including space-charge effects and wakefield effects – a full 3D simulation is prohibitive using the current algorithms and existing multi-processor hardware. Simulations using roughly 100 processors have been estimated to require 1 year of computer time. Use of a “2.5-D” code (with full transverse phase space and projections for the longitudinal phase space) is possible. The need for comparisons of the different existing implementations was emphasized during discussions with M6 (High Intensity Proton Sources) and test problems were defined for this purpose. More work is needed to define experiments suitable for validation of the simulation packages. A comprehensive program of simulation studies coupled to experimental measurements is needed in order to understand these effects and improve the performance of the machines.

The simulation requirements for muon-based systems present additional challenges. In order to successfully generate a muon beam, many different subsystems have to be designed and optimized. The machine needs a high power proton driver, a pion production target and pion capture channel, a channel to capture and cool muons, and finally a muon accelerator. For muon colliders there are stringent requirements on controlling the longitudinal phase space, while for muon storage rings (neutrino factories) these requirements are relaxed. All of these subsystems have challenging modeling requirements in both the beam dynamics and electromagnetic simulation capabilities. The proton driver has very high beam power requirements, 1MW for the neutrino factory and 4MW for the muon collider, having $3\text{-}10 \times 10^{13}$ protons per cycle and very short (~1ns) bunch length. Modeling requirements for the front end of the machine involve accurate simulation of particle production, beam transport, and muon-matter interactions. In addition, for high-intensity muon colliders, space charge effects are also important and require accurate modeling. Because of the need to model muon-matter interactions for ionization cooling, the large size of the beam phase space and the very strong focusing fields with complicated spatial dependence, low order approximations cannot be used in these simulations. As a result, the simulation packages used in the design studies are single particle tracking codes that are very slow (~10k particles/day on a commodity PC) and prohibit simultaneous optimization of more than one subsystem. This simultaneous optimization is necessary for cost minimization and maximal performance, since the optimal solution for each component depends on the beam characteristics produced by the components upstream. The development of a simulation toolkit, with HPC capabilities, which integrates muon-matter interaction codes with conventional beam dynamics codes is necessary in order to obtain this optimization functionality and a complete physics description.

2.2 Next-Generation Linear Colliders

Next-generation linear colliders require demanding computer simulations in regard to both electromagnetic and beam dynamics modeling. For example, extremely complicated 3D electromagnetic structures for the NLC must be modeled and analyzed with greater speed, accuracy, and confidence than has previously been possible. Presently popular serial electromagnetics codes are inefficient in handling complex geometric shapes, or are limited in their ability to solve large-scale problems. However, the recent development of parallel eigenmode and time-domain codes has already increased our modeling capabilities by roughly three orders of magnitude. It is now possible to compute the fundamental mode frequency of complex 3D accelerating structures with an accuracy of 1 part in 10^5, as is needed for NLC structure design. In fact, for the first time, we are approaching a situation where 3D cavities can be modeled with a precision approaching that of fabrication tolerance. Besides high precision studies of individual structures, system-scale simulations of multi-cell structures are also needed to verify wakefield suppression strategies. Important new capabilities are needed for future studies, including the development of a parallel 3D eigenmode code that can model lossy structures and a parallel time-domain electromagnetics code with a rigid beam model for the direct calculation of wakefields. In addition to modeling electromagnetic components, HPC capabilities are needed to model beam dynamics in linear colliders. For example, in both the NLC and TESLA designs, the accurate treatment of space-charge effects and other collective effects is important to predicting the beam’s behavior in the damping rings. In order to validate the basic operational characteristics of these machines, the linac and beam delivery systems need to be modeled including component fluctuations, tuning, and feedback systems. Such simulations are impossible on serial computers, where the execution time to run one such code with the desired accuracy has been estimated to be 1 year per processor.

2.3 Very Large Hadron Colliders

The very large size of the VHLC (233 km circumference, 3136 dipoles, 3296 quadrupoles and sextupoles, with a damping time of 100 hrs) requires HPC capabilities not only for conventional beam dynamics design problems but also for problems that arise from the challenging operational requirements of the machine. Existing closed orbit correction techniques running on serial computers are not sufficient, since, due to the large number of beam position monitors and correctors, these algorithms will take a prohibitively long time to execute: the simulations will be 10^6 times slower than those associated with existing accelerators. Both parallelization and the development of new algorithms are necessary. New algorithm development is also needed to tackle alignment and other installation problems. The misalignment due to ground motion, which has been measured in a mine with similar geology, is expected to result to an orbit distortion of ~9mm/year. Accurate modeling of the machine taking into account these effects is required to understand the impact on luminosity and the orbit correction requirements. Since the dynamic aperture is a major cost driver, accurate modeling of the beam for multiple turns, including lattice imperfections, is necessary. HPC capabilities are needed to perform realistic simulations and to provide complete phase space coverage in the dynamic aperture calculation. Since the VLHC operates in the strong-strong regime, self-consistent modeling of the beam-beam interaction is necessary. This can be achieved by either PIC or Vlasov codes, but the problem is CPU intensive thus requiring HPC capabilities. Since the energy stored in the machine is large (3 GJ/beam for Stage I), with large debris from collisions (3KW/interaction point for Stage I), extensive simulations are necessary to design passive and active protection systems for vital components of the machine and the detectors. Finally, modeling and predicting the thresholds for instabilities, such as the resistive wall, electron cloud, and transverse mode coupling instabilities, is necessary to evaluate and further develop the feedback system designs.

2.4 Laser- and Plasma-Based Accelerators

Besides the design of next-generation accelerator complexes, HPC is also needed, in concert with theory and experiment, to explore and develop novel methods of acceleration like plasma-based and laser-based acceleration techniques. Using these techniques, extremely high gradients (up to 100 GV/m) have been measured over short distances in the laboratory. The challenge is to control and stage high-gradient sections so that one can produce high quality, high energy beams in a less costly, more compact configuration that would be impossible using conventional technology. Beyond applications to HEP, such compact accelerators would have huge consequences in others areas of basic and applied science, industry, and medicine. However, modeling these complex systems requires solving the 3D coupled Maxwell/Vlasov equations. Given that the phenomena involve multiple length and time scales (a situation that is particularly challenging when the laser wavelength must be resolved), 3D PIC simulations can only be performed using HPC resources. As an example, the simulation of a 1 GeV plasma accelerator stage using a fully explicit PIC code has been estimated to require 10,000 to 100,000 CPU hours for a single run. In addition to a parallel PIC capability, there is also a need to develop high-fidelity reduced description models. The goal for the advanced accelerator simulation community for the next several years will be to build a suite of parallel electromagnetic modeling codes in 2D and 3D that can support and guide further theoretical and experimental efforts to make plasma-based accelerator technology practical for real-world applications.

3. State-of-the-art in HPC Simulation of Accelerator Systems

Working Group T7 had three sessions devoted to the status of HPC codes in the accelerator community. The topics were organized around three areas: HPC simulation of electromagnetic systems, beam systems, and laser/plasma accelerators. The following includes highlights from those sessions and a description of some of the issues motivating the development of HPC codes in these three areas.

3.1 Electromagnetics

The acceleration of charged particle beams invariably involves electromagnetic fields, whether they are the external RF fields that provide the acceleration, or the magnetic fields that control the particle orbit, or the wakefields generated through the interaction of the beam with the accelerator environment. In present-day and future planned facilities, accelerating systems are reaching a new level of complexity as researchers continue to strive for improved performance and increased functionality so that 3D electromagnetic field solvers have become absolutely essential for their design and analysis. Some of these solvers have been in use for over a decade and have served the accelerator community well in providing a much needed capability for analyzing electromagnetics in a wide range of accelerator structures and beamline components.

During the development of the accelerating structure for the NLC, it soon became evident that existing codes could not meet the accuracy requirement that results from the tight tolerances needed to maintain structure efficiency and beam stability. The specific requirement is accuracy in the accelerating mode frequency of 1 part in 10,000 or 0.01% which is of the order of machining error. The challenge was to achieve this accuracy in a structure geometry that was greatly modified from the conventional 2D disk-loaded waveguide configuration in order to incorporate several important features (see Fig. 1). First, the outer wall of the cavity was contoured to increase the shunt impedance and therefore, the structure efficiency. Second, four external manifolds were added along the length of the structure and connected to the cells through coupling slots to damp out the dipole wakefields and also to increase the vacuum conductance. The resulting Round, Damped, Detuned Structure or RDDS is highly complex in shape and has 11 cell dimensions to vary. Because of the detuning, there is also variation from cell-to-cell along the 206-cell section on the order of microns. To arrive at an optimal design under these multiple specifications using the trial-and-error approach by cutting metal would have been impossible. Also, the NLC linacs will consist of millions of such cells so that any post tuning will be impractical. The only solution is a highly accurate field solver that can reliably generate the cell dimensions for direct input into computerized machining of the structure.

At the Snowmass T7 session on electromagnetic modeling, Cho Ng (SLAC) described two HPC codes developed under the Accelerator Grand Challenge, Omega3P and Tau3P. In order to be able to deliver the required accuracy on a highly complex structure like the RDDS, the desired field solver must use conformal meshes to preserve the geometry fidelity and must be able to run on multiple processors to perform the large-scale simulations needed for high-resolution studies. Existing codes possess either one of the two attributes but not both. The parallel 3D eigensolver Omega3P was built with these two attributes as central to its design. It uses a second-order finite element formulation based on unstructured tetrahedral meshes and it runs on distributed memory computers using MPI. The 3D parallel time-domain solver Tau3P was developed based on the same premise by using an unstructured finite volume grid. Both codes are object oriented, MPI based, share the same geometry data structure, and reuse existing parallel libraries.

The codes Omega3P and Tau3P have played a crucial role in the successful development of the Round, Damped, Detuned Structure (RDDS) for the NLC. Fig. 1a shows the partitioned RDDS mesh which was used in the Omega3P calculation for the accelerating mode. Fig. 1b is the computed frequency as a function of mesh size, showing that the required accuracy of 0.01% is reached. Fig.1c is the actual prototype built from dimensions generated from Omega3P calculations. There was excellent agreement between the calculation and measurement. This cavity has 14% higher shunt impedance than the original design so that it could potentially lead to savings in machine cost of $100M or more. But more importantly, a new tool is now available which has been shown to be capable of modeling complex cavities to machining tolerance and without which, the RDDS design would not have been realized. The Tau3P code also contributed to the RDDS design by modeling external loading effects such as couplers. Fig. 2a shows the geometry simulated while Fig. 2b shows the dipole mode spectrum calculated in a single run. The numerical results compared well with measured data.

It is important to point out that large-scale, parallel electromagnetics codes are needed, not only for the design of a next-generation linear colliders, but for other accelerators too, both existing and proposed. For example, Figure 3 shows how Omega3P is being used to study anomalous heating due to trapped modes in the PEP-II B-factory interaction region.

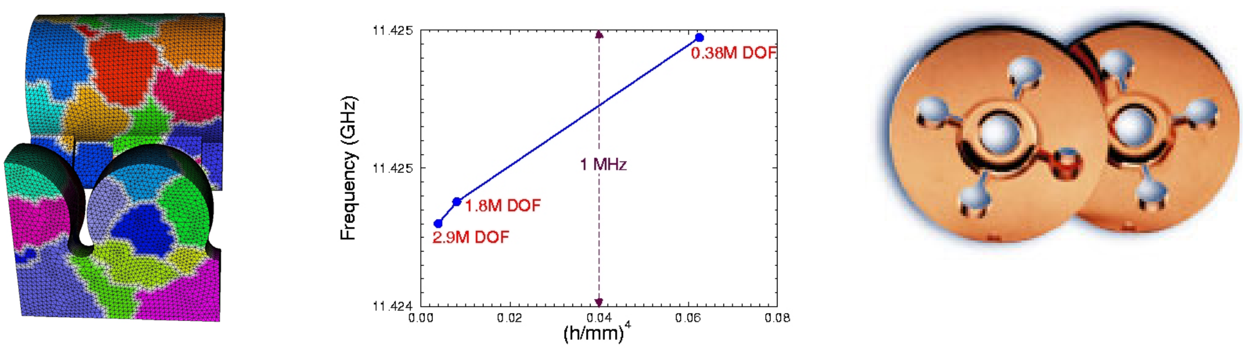

Figure 1a. An octant of NLC RDDS mesh showing domain decomposition; **1b.** Accelerating mode frequency calculated with **Omega3P** versus mesh size; **1c** RDDS prototype cells fabricated from numerically generated dimensions.

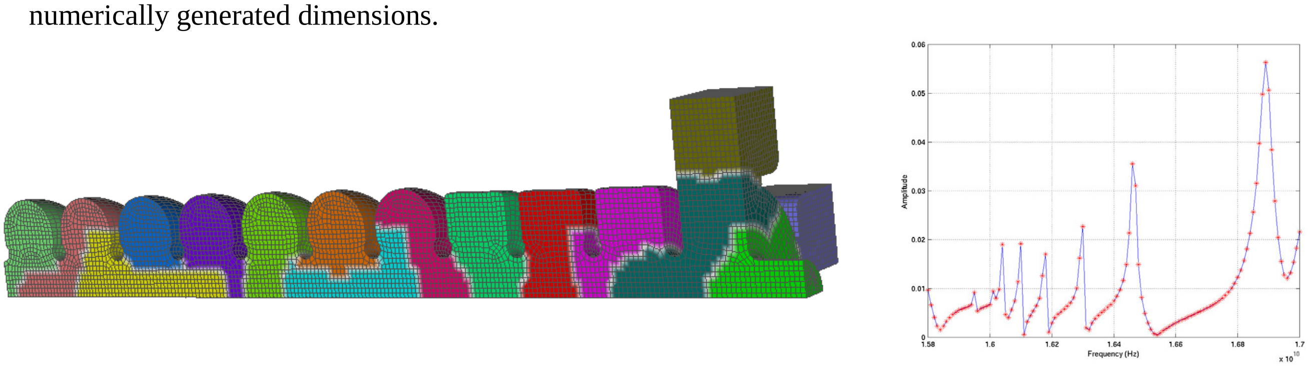

Figure 2a **Tau3P** simulation of an NLC RDDS 10-cell stack including output couplers; **2b** Dipole mode spectrum calculated with **Tau3P**.

.

T7001

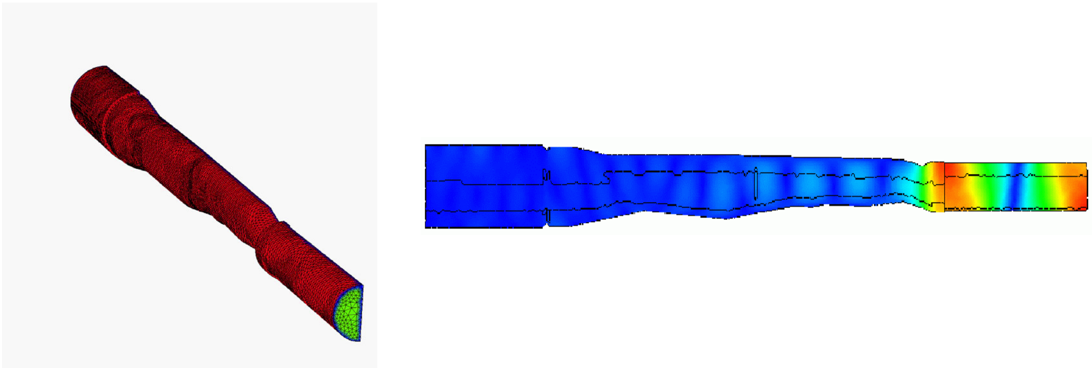

Figure 3. A mesh model (left) of a section of the PEP-II beamline from the IP out in one direction; **Omega3P** calculation (right) of a trapped mode showing electric field contours within the section.

3.2 Beam Dynamics

Understanding and predicting the dynamics of charged particle beams in accelerators can have huge consequences for their cost, performance, and, ultimately, for their success in achieving the goals for which they are developed. For example, a 1cm increase in the design aperture of the Superconducting Supercollider led to an estimated $1 billion increase in cost. In the area of high intensity accelerators, future machines will have to operate with *ultra-low* losses in order to limit radioactivity that would hinder or prevent hands-on maintenance. Quantitative prediction of such losses is a major challenge, since they are generally associated with modeling the very-low density tails of the beam distribution (the beam halo). The development of HPC codes makes it possible to model beam systems with unprecedented speed, resolution, and accuracy. Furthermore, the availability of HPC resources makes it possible to include multiple physical effects, such as collisions, ionization, beam/matter interactions, and coherent synchrotron radiation (CSR). Furthermore, a variety of algorithmic approaches have been used to model beam systems, such as particle-based methods, direct solvers (e.g. direct Vlasov solvers), and delta-f methods. All these topics were discussed in the T7 session on HPC simulation of beam systems.

Salman Habib (LANL) provided an introductory talk describing methods for the large-scale simulation of classical and quantum evolution equations. This included a description of particle-based methods and direct methods. Particle-based methods for charged particle beam simulation share much in common with simulation techniques for modeling astrophysical and cosmological systems. In fact, portions of the IMPACT beam dynamics code (described below) have been incorporated into a code that is being used to study galactic dynamics. At the other (microscopic) extreme, spectral techniques that are useful in modeling the Schrodinger equation have also been employed in spectral codes for solving the Vlasov equation on a grid in phase space. So far, 4-dimensional (2 degree of freedom) simulations have been performed. However, the advent of very large-scale parallel supercomputers, with several terabytes of memory, are bringing full 6D Vlasov simulations close to the realm of possibility. Lastly, Habib spoke about simulations of the quantum/classical interface. He pointed out that classical chaos has been shown to emerge from quantum mechanics when the measurement process is properly included, and furthermore that the computations associated with this system could only have been performed using a parallel supercomputer.

Ji Qiang (LANL) spoke on the IMPACT parallel Particle-In-Cell (PIC) code. IMPACT (Integrated Map and Particle Accelerator Tracking) was developed to enable very large-scale 3D simulations of high intensity beams in linacs using parallel supercomputers. The code is based on the split-operator methods, which make it possible to combine the capabilities of two major areas, magnetic optics and parallel particle simulation methods. The code uses parallel "particle managers" and "field managers" with dynamic domain decomposition to achieve high performance and maintain load balance. As with most modern beam dynamics codes, IMPACT uses canonical variables. In regard to space charge, a variety of 3D space-charge solvers are currently available corresponding to different boundary conditions. IMPACT has been used to model the beam halo in the Spallation Neutron Source (SNS) linac (Figure 4), the CERN Superconducting Proton Linac (SPL), and the Accelerator Production of Tritium (APT) linac. Simulations with up to 500M particles have been performed, a number that is within a factor of two of the real-world number of particles for some applications.

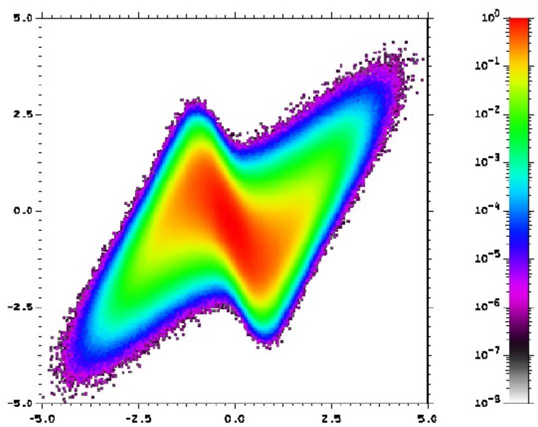

Figure 4. Horizontal phase space (x and px on the horizontal and vertical axes, respectively) at the end of a 500 million particle simulation of the Spallation Neutron Source linac using IMPACT.

In addition to enabling larger simulations that have greater resolution and accuracy, the availability of HPC resources also enables the inclusion of additional physical phenomena, the treatment of which may be highly compute-intensive, in accelerator simulations. An example is provided by the treatment of stochastic effects, or phenomena that can be modeled as such. For example, stochastic contributions to Vlasov/Poisson evolution can occur due to particle collisions and noise in external fields. External noise is easy to model via standard stochastic techniques applied to the particle equations of motion. Soft collisional effects may be included by generalizing the Vlasov/Poisson equations to a Fokker-Planck form. Fokker-Planck collisions can then be included in the PIC method via the addition of friction and (multiplicative) stochastic forces (i.e. damping and noise terms) in the equations of motion for the simulation particles: This is the Langevin approach to incorporating soft collisions.

As an example, LANGEVIN3D is the first Fokker-Planck solver to treat damping and diffusion from first principles, an approach that was, until recently, said to be impossible. However, it has now been demonstrated that such calculations are indeed possible, using terascale computers and algorithms targeted to such platforms. Figure 5 shows the self-consistent diffusion coefficients corresponding to a Gaussian beam, computed using LANGEVIN3D running on the LANL/ACL SGI Origin 2000. The horizontal line (the Spitzer approximation) shows the most widely used approximation for the (diagonal) diffusion coefficient. The ability to perform self-consistent Langevin simulations will have applications in modeling long-term behavior in storage rings, and applications in modeling astrophysical systems.

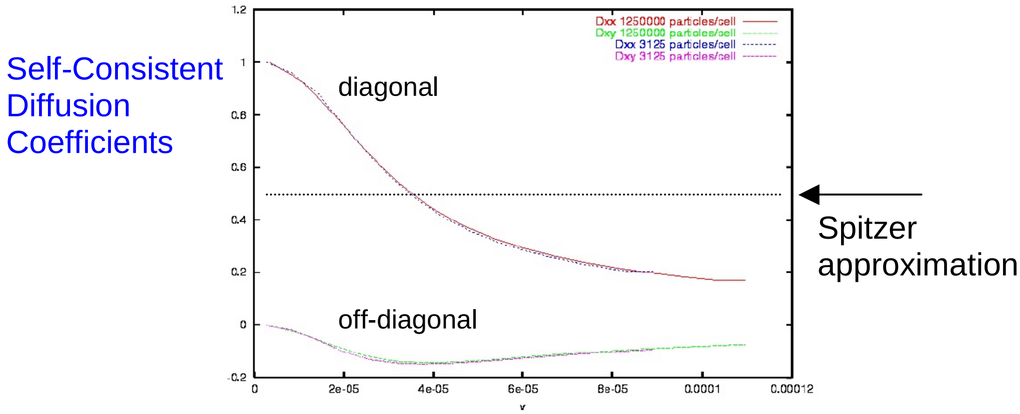

Figure 5. Diagonal and off-diagonal self-consistent diffusion coefficients (for a Maxwellian distribution) computed using the parallel code LANGEVIN3D. The computation of these coefficients is carried out by coarse-graining the spatial computational grid into "super-cells" and calculating the Rosenbluth potentials in each super-cell. Two results are shown, using 3000 and 1.25M particles/super-cell. The excellent agreement demonstrates the success of the computational strategy, even at modest particle densities.

David Grote (LBNL) provided an overview of the WARP parallel simulation suite of codes used to model beams for applications in Heavy Ion Fusion (HIF). WARP offers 3-D, axisymmetric, and "transverse-slice" 2-1/2D (x,y,p_x,p_y,p_z) geometries, and is used extensively throughout the HIF program for studies of beams in the accelerator, pulse-compression line, and final focusing system. Among its unique features is the ability to perform 3D injector simulations including time-dependent effects. The code suite runs on a variety of platforms, including machines such as NERSC's Cray T3E-900 and IBM SP, where message passing is implemented using MPI. WARP is organized around a set of physics packages that contain executable routines and data, and is written in a superset of Fortran that offers features such as dynamic memory allocation and a run-time database that is accessible by all code routines or by the user interactively. It runs under the control of the Python scripting language. This gives the user control over how the simulation proceeds, facilitating, for example, the use of iterative methods for steady problems, as well as fully time-dependent simulations. The user's input file is, in effect, a computer program written in the Python language extended by the capabilities of the code. WARP also includes a prototype web interface so that, through a browser, the user can query the code, or ask it to generate certain plots, etc., even when the code is running in batch mode. The WARP model uses time as the independent variable, in contrast with most beam optics codes that use the arc length. The self-consistent field used by WARP is assumed to be electrostatic, with simple semi-analytic corrections to handle multi-beam inductive effects. Poisson's equation is solved on a Cartesian mesh that moves with the beam, either steadily, or as a "treadmill" so that zone boundaries do not vary from step to step. When bends are used, the field solution is altered to include the curvature of the "warped" coordinates. Electrostatic elements can be described from first principles by inclusion of conductor geometry as boundary conditions in the solution of the self-fields at subgrid-scale resolution, using a "cut-cell" method in 3D. WARP has been validated against a number of small experiments in the HIF program.

Andreas Kabel (SLAC) provided an overview of the TraFiC[4] code for modeling coherent synchrotron radiation (CSR). CSR is an important effect for short, low-emittance bunches of charged particles traveling on strongly bent trajectories. Its action on a bunch will in general lead to emittance degradation and an increase of energy spread. The effect of CSR on beam dynamics involves integration over the past history of the bunch. As such, the accurate simulation of CSR is extremely time-consuming. Therefore a multi-processor version of the code has now been developed. TraFiC[4]/MP uses MPI for interprocessor communication. The code has been tested on Pentium cluster machines running FreeBSD 4.3 or Linux 2.4 and the freely available MPICH package. It has also been run on the IBM/SP at NERSC. It has been possible to assign a single node to each tracked macroparticle, thus reducing running times from 300 hours CPU time (for a long-beamline problem involving shielding, on a PC running Linux) to 5 hours of real time.

Hong Qin (PPPL) presented an overview of the beam equilibrium, stability, and transport (BEST) code. A 3D multispecies nonlinear delta-f formalism has been implemented in the BEST code, which has been applied to a wide range of important collective processes in intense beams. For example, the code has been used to simulate the linear and nonlinear properties of the electron-proton (e-p) two-stream instability observed in the LANL Proton Storage Ring (PSR) experiment for a long coasting beam. The delta-f method is a low noise method that has significant advantages over conventional PIC codes in those circumstances where one is interested in the evolution of a perturbed distribution around a known background distribution. However, the method is fully nonlinear and simulates completely the original nonlinear Vlasov/Maxwell equations. The code has been parallelized and runs on the IBM/SP at NERSC.

Robert Ryne (LBNL) presented, for Andreas Adelmann (PSI/ETH), a description of MAD9P. This code is a version of the C++ code MAD9 that has been parallelized using the POOMA framework. The code uses a split-operator approach and parallel Poisson solver to include the effects of space charge. It has both an FFT-based solver (with the required zero-padding to treat open boundary conditions) and a modified Barnes-Hut parallel tree solver. A novel feature is the use of multiple reference particles to treat situations where the perturbative approach of a standard magnetic optics code with a single reference particle would not be appropriate (e.g. when modeling the bunching of a beam which initially fills a large portion of the rf bucket.) MAD9P is currently being used for studies of high intensity beams in cyclotrons (see Figure 6), for which large simulations of beams in complex structures are needed to predict uncontrolled beam loss. The code has been run on the ETH/Zurich cluster, the LANL/ACL Nirvana system, and the IBM/SP at NERSC.

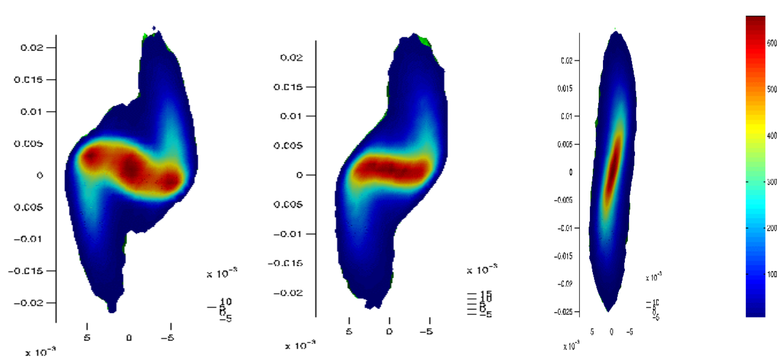

Figure 6. MAD9P simulated charged density distribution of a beam coasting (moving from right to left) in a cyclotron.

3.3 Laser- and Plasma-Based Accelerators

The quest to understand the fundamental nature of matter requires ever higher energy particle collisions, which in turn leads to ever larger and more expensive particle accelerators. Plasma-based accelerators can sustain electron plasma waves with longitudinal electric fields on the order of the nonrelativistic wave breaking field, $E_0=cm_e\omega_p/e$, where $\omega_p=(4\pi n_e e^2/m_e)^{1/2}$ is the plasma frequency at an electron density n_e. Laser wakefield accelerators (LWFA) have demonstrated accelerating gradients of 100 GV/m, which is three orders of magnitude higher than for conventional structures. Beam-driven plasma wakefield accelerators (PWFA) will likely be able to generate comparable gradients. Plasma lenses, which share much of their physics with the PWFA concept, can provide focusing strengths 1000 times greater than conventional quadrupole magnets.

The study of plasma-based accelerators is a critical piece of the R&D portfolio within the DOE/SC/HENP, because these advanced concepts offer a promising path toward the next generation of accelerator technology for achieving electron-positron and electron-hadron collisions in new energy regimes at reduced size and cost. Very exciting experimental work has demonstrated that ultra-high gradients can be achieved and that copious numbers of electrons can be accelerated. Recent PWFA experiments at SLAC have indicated that coherent wakes can be excited over meter distances. Also, innovative theoretical work suggests how to extract high-quality electron bunches from LWFA systems.

Besides their obvious importance in regard to reaching the high energy frontier, laser- and plasma-based accelerator systems have many other potential applications of great economic and societal importance. This is because such a drastic increase in gradient would allow a reduction in accelerator size, enabling tabletop accelerators to have beam energies approaching those now found only at a few national facilities. The ability to place such compact accelerators in university departments, government research organizations, high-technology businesses, and hospitals would have staggering consequences for science, industry, and medicine.

Three-dimensional PIC simulations of plasma-based accelerators require effective use of parallel processing on teraflop scale computers, with flexible domain decomposition and, for particular problems, dynamic load balancing. Other critical features include a moving window to follow the driving laser pulse or electron bunch, electromagnetic wave launchers and particle beam emitters, and absorbing boundary conditions. Laser ionization models are required to accurately model propagation of the drive laser pulse through a neutral gas jet. Similarly, electron impact ionization models are required to accurately model plasma lens experiments, and may play a role in the plasma afterburner application as well. The use of such ionization algorithms is nontrivial and requires the ability to handle multiple species, with algorithms for sub-cycling the advance of ion macroparticles, which requires damping of the resulting high-frequency field noise.

At the Snowmass meeting a joint session was held that involved the HPC Working Group (T7) and the Advanced Accelerator Working Group (T8). A full summary of that session and the status of HPC codes for plasma-based acceleration can be found in the T8 summary. Here we will focus on two parallel PIC codes (OSIRIS and XOOPIC) that were discussed at the session and which have been used for modeling laser/plasma accelerators. OSIRIS is at present the only fully 3D, parallel electromagnetic PIC code being used to model plasma-based accelerators. The code has a dynamic mesh and moving window. A typical simulation on the NERSC T3E uses 64 processors and follows ~50 million macroparticles using a 350x200x200 cell grid for 10,000 time steps; simulations using 256 processors have been run with up to 500 million macroparticles. The code was recently used to model the SLAC E-157 experiment (see Figure 7), in which an electron beam was passed through a partially ionized plasma to produce an intense plasma wake with gradients approaching 1 GeV/m. Using OSIRIS, it was shown that hosing is not a problem for the present experimental parameters, and that positron wakes are qualitatively different than electron wakes. Another code, XOOPIC, is also being used to model laser/plasma accelerators. An example simulation is shown in Figure 8. This code has been widely used in the plasma physics community, and has recently been modified to simulate plasma-based accelerators. The code has extensive capabilities to model ionization effects and surface physics effects. This code was also used to model E-157, and the results were found to be in agreement with the predictions of OSIRIS. Also, the XOOPIC simulations showed that electron-neutral scattering is insignificant for the current experiment, although it could be important in future experiments with a high density of neutral lithium.

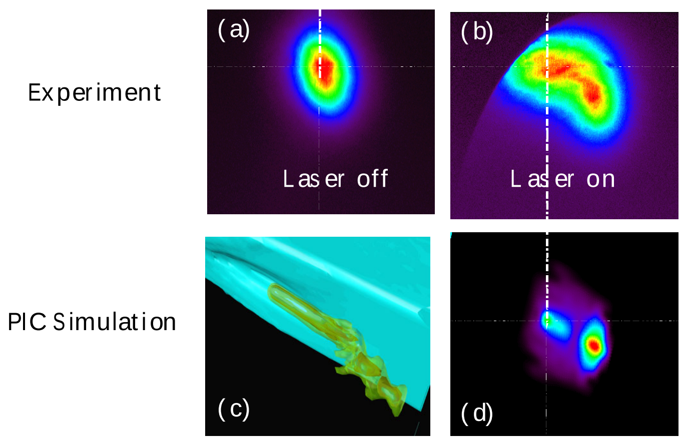

Figure 7. Comparison of simulation (using OSIRIS) and experimental data for the SLAC E-157 experiment. The simulation results are in agreement with the experimental observation that it is possible to reflect or refract a particle beam from a dilute plasma gas.

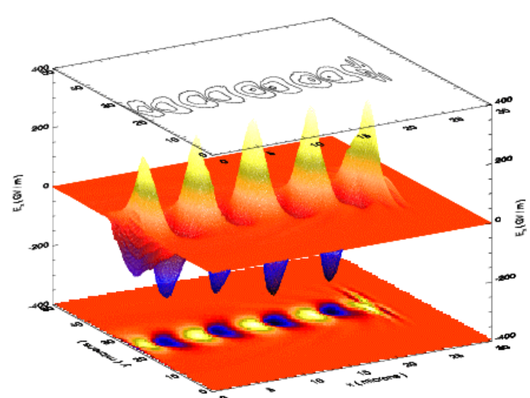

Figure 8. Surface plot of the longitudinal electric field generated by a $3x10^{18}$ W/cm^2 laser pulse (smaller, partially hidden peaks to the far right) and the resulting plasma wake (larger peaks). The structure of the laser pulse is seen in the contour plot (above) and the surface plot projection (below). E_x is shown in GV/m, while the coordinates x, y are shown in µm. The data were generated using XOOPIC.

In conclusion, high-performance computing has already been shown to be accurate and useful for interpreting and guiding these experiments. However, three-dimensional modeling of the full range of plasma-based accelerator experiments currently being planned will require advances in software engineering and the development of high-fidelity reduced description models. To model 100 GeV plasma accelerator stages will require ensuring that the high parallel efficiency currently obtained on hundreds of processors be extended to thousands or even 10's of thousands of processors. As such, it is important to build a suite of high fidelity electromagnetic PIC simulations in two and three dimensions that can support and guide further theoretical and experimental efforts to make plasma-based accelerator technology practical for real-world applications. Eventually it will be necessary to integrate the plasma accelerator code suite with a "conventional" accelerator HPC modeling capability so that one can seamlessly model an RF linac with a plasma lens or plasma afterburner incorporated into its design.

3.4 Particle Production Codes

Nicolai Mokhov discussed the status of the MARS code, a package developed to model hadronic and electromagnetic cascades in shielding, accelerator and detector components for energies from a few eV to ~100TeV. The hadron production models included in this package are also used to model secondary and tertiary beam fluxes from the interactions of a primary hadron beam with a target. The package includes a Graphical User Interface, utilities with histograming and diagnostics and a Beam Line Builder which uses the MAD language as its input. The hadron production includes both inclusive and exclusive models (with full differential distributions for all produced particles). The exclusive model is the default in the energy range of 1 MeV to 5 GeV, while at higher energies the inclusive model is used because the exclusive generator (DPMGET) is very CPU intensive (of order 1000 times slower). The package also includes neutrino (weighted), slow neutron (MCNP4), and electromagnetic interaction generators. The possibility of parallelizing MARS in order to avoid the use of the inferior inclusive model in hadron production modeling, and the design of the MAD-MARS parser (Beam Line Builder) were discussed in great detail. The parallelization of the MARS code seems straightforward, since the problem is ideally scalable to the number of primary interactions (N particles on M processors, N/M particles per processor) because in the evolution of the simulation primary interactions are independent of each other. It is desirable to integrate a code like MARS to an HPC accelerator modeling toolbox. The MAD-MARS parser is a very good starting point for the input module of an integrated HPC accelerator simulation toolbox. The parser, which uses object-oriented techniques, creates data structures (objects) that can be easily translated into the native language of other simulation programs. The Standard Input Format adopted by MAD and other programs is the most commonly used lattice description language. Tools such as the MAD-MARS parser and the LIBXSIF library provide a simple means of adding a common input language capability to accelerator codes and should greatly facilitate collaboration among accelerator modelers.

Roman Samulyak discussed the hydro- and magnetohydrodynamic analysis of the muon collider target using techniques of the FronTier package. This package uses the method of front tracking, which allows the numerical solution of problems that can be described by a system of conservation laws. The method often does not require highly refined grids and has no numerical diffusion; it is ideally suited for problems in which discontinuities are an important feature. The muon collider target is a mercury jet interacting with a high-intensity proton beam in a strong

(20T) magnetic field. The problem of the pressure field evolution and free surface instabilities is the major concern in the study of this target. The code has been used to study both the dynamics of liquid metal (mercury) jets in strong nonuniform magnetic fields and the interaction of the mercury target with high energy proton pulses. Numerical simulation for the target evolution, driven by strong pressure waves, is important for the optimal target design. The FronTier-MHD code has been parallelized using MPI. Simulations have been performed on a cluster of Pentium processors.

To model the interaction of the mercury target with proton pulses, a tabulated equation of state for mercury was created in a wide temperature - pressure domain which includes the liquid - vapor phase transition and the critical point. A two-phase isentropic analytic eos was also developed. Numerical simulation of the mercury target - proton pulse interaction using the FronTier code and the two phase mercury equation of state was performed during 120 microsecond time interval. According to the simulation results, the mercury target brakes into droplets due to the proton energy deposition and the radial velocity of the jet surface before the droplet formation is in the range 10 - 50 m/sec.

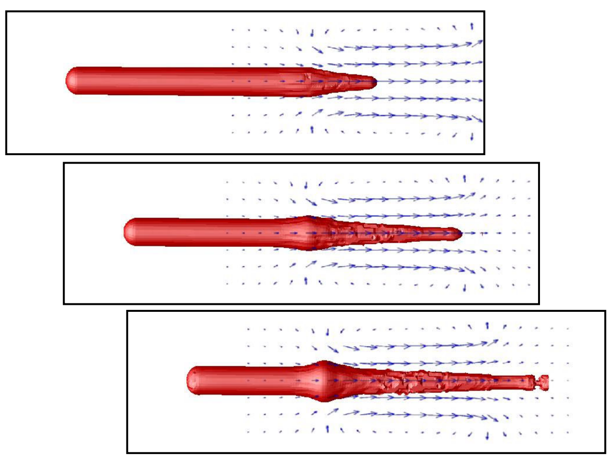

Figure 9. Simulation using FronTier-MHD of a liquid metal jet entering a 20 T solenoid.

3.5 Ionization Cooling Simulation Codes

Panagiotis Spentzouris described a hybrid simulation package that combines the 3-D space charge beam dynamics code IMPACT (see previous section) with DPGeant used in ionization cooling simulations (see following discussion). The space charge effects in a Lithium lens based channel have been demonstrated in the case of a muon machine with 10^{13} muons/bunch. The motivation for the hybrid simulation package is the study of space-charge effects in the material-free sections of ionization cooling channels. It is also a proof of principle that integrated simulation packages of diverse codes can be developed. In this case, the serial DPGeant simulations involved 500 simulation particles. By coupling the code to the parallel beam dynamics code IMPACT and running the problem on 128 processors of the T3E at NERSC (with a separate copy of DPGeant running on each processor), it was found that simulations with 500,000 particles required only about ten times the execution time of serial jobs with 500 particles, indicating good scalability for these test runs.

Paul LeBrun described Geant4 and DPGeant cooling simulations. Both DPGeant (a double precision version of the HEP simulation code Geant3), written in Fortran, and Geant4, written in C++, provide accurate modeling of muon-matter interactions needed for ionization cooling. These are single particle tracking codes that track particles in the lab frame using 3rd order Runge-Kutta numerical integration. The main issues in these packages are the representation of electromagnetic fields in the equations of motion (assuming no time dependence) and the fact that beam element models are not implemented in the toolkit. For ionization cooling simulations these codes have been expanded to include time dependent electromagnetic fields, and a limited list of beam element models has been implemented. These modifications are maintained and distributed by FNAL members of the muon collider collaboration. There is a need to integrate this functionality to the "official" versions of these packages, especially Geant4, which is currently under development.

Rick Fernow spoke on the ICOOL simulation package. ICOOL is a 3-D tracking code that was developed specifically for ionization cooling studies. Analytic models are used to describe beam elements. The physics models used to describe muon-matter interactions are adapted from the Geant3 package described above. ICOOL uses beam coordinates to describe particles and beamline regions. The package has a comprehensive beamline description language, standardized diagnostics, and provides some analysis utilities (e.g. calculation of beam moments and emittances). The biggest advantage of this package is that the beamline elements implemented are those needed for specific ionization cooling channel designs. The performance is similar to the performance of DPGeant, thus large-scale simulations of complete designs are very slow.

Vladimir Grichine summarized the Geant4 simulation package. This package employs a full object-oriented design to implement detailed particle-matter interaction physics processes and provides a flexible toolkit to describe complicated geometries. It is a single particle tracking code using a 3rd order Runge-Kutta integration. There were some concerns about the accuracy of the integrator in beam physics applications involving multi-turn simulations of circular accelerators. The modeling of electromagnetic fields in the equations of motion is still under development (no time dependence); this work was motivated by the ionization cooling applications developed at FNAL. The code has well defined interfaces which allow the use of the overall framework with any user implemented model of particle tracking. Furthermore, because of its object-oriented design, it is easy to integrate in more general frameworks that control a variety of diverse codes.

Martin Berz discussed the capabilities of the COSY INFINITY simulation package. COSY is an arbitrary order code that uses differential algebraic methods (DA) to allow the handling of high order maps for beam transport. The code takes full advantage of these techniques by providing a scripting language that is fully adapted to the DA methods. The DA approach is well suited to the exploration of high-order effects in beam dynamics. Various examples of systems very sensitive to these higher order effects were presented. The package provides a sophisticated 6th order Runge-Kutta integrator and three different optimizers that can be invoked from the COSY language environment. Normal form methods were also presented that included the effect of damping in the analysis.

Kyoko Makino presented ionization cooling channel simulation studies based on map methods using COSY INFINITY. The objective of these studies is to provide a fast and reliable environment where the cooling lattices can be analyzed and optimized. There is no attempt to model the ionization cooling process itself, since the main focus is to understand the optics. Examples of studies and optimization involving existing cooling lattice designs were shown. The functionality of this package is very valuable for the fast convergence of ionization cooling lattice design efforts.

4. Plans related to the Scientific Discovery through Advanced Computing (SciDAC) project

4.1 Project Overview

Recognizing the challenges posed by nest-generation accelerator projects, a SciDAC (Scientific Discovery Through Advanced Computing) project on 21st Century Accelerator Simulation was approved in mid-2001. The primary objective of this national R&D effort is to establish a comprehensive terascale simulation capability for the US Particle Accelerator Community.

The scientific software to be developed under this project falls naturally into six areas. Three of these correspond to accelerator physics application areas: Electromagnetic Systems Simulation (ESS), Beam Systems Simulation (BSS), and Beam/Electromagnetic Systems Integration (BESI). These will be supported by strong collaborations with researchers from the *Scientific Computing Software Infrastructure* component of SciDAC. For the purposes of the project, the activities associated with those collaborations have been grouped into three more areas: Applied Mathematics and Numerical Analysis (AMNA), Computer Science and Software Engineering (CSSE), and Visualization and Model Evaluation (VME). A description of these six areas follows.

A. Electromagnetic Systems Simulation (ESS): This focus area concentrates on developing a set of parallel tools for the design, analysis, and optimization of complex electromagnetic components and systems in accelerators. The tool set will consist of solvers of Maxwell's equations that encompass electrostatics and magnetics, low frequency responses such as eddy currents, and high frequency responses in the frequency and time domains. These solvers must have the ability to model complicated structures accurately using conformal meshes for high accuracy. They must also be able to simulate a variety of physical conditions pertinent to accelerator design such as cavities with lossy material or external coupling that require complex eigensolvers. They must also contain physics modules such as a rigid beam capability for wakefield calculations and the inclusion of surface properties needed in dark current simulations.

B. Beam Systems Simulation (BSS): This focus area covers the dynamics of intense charged particle beams propagating through a variety of accelerator structures that make up an accelerator complex. Methods used to solve the governing equations (Vlasov/Poisson or Fokker-Planck/Poisson) include particle simulation methods and, with the advent of multi-terascale computers, direct solvers. Activities in this area will involve the development of beam dynamics codes for a variety of accelerator systems (including high intensity beams in rings), capabilities for self-consistently modeling the beam-beam effect, software components to calculate space charge fields subject to a variety of boundary conditions, and physics modules to treat effects such as collisions, ionization cooling, wakefields, and CSR.

C. Beam/Electromagnetic Systems Integration (BESI): This focus area includes the development of Maxwell/Vlasov solvers that are needed to model laser- and plasma-based accelerators. Specific needs include moving windows, the ability to treat multiple speciies, and the ability to model multiple background gases and ionization. Full-scale three-dimensional modeling of plasma accelerators will require effective use of parallel processing on teraflop scale computers, with flexible domain decomposition and for particular problems dynamic load balancing. Also, the disparity in time-scales inherent in plasma-based acceleration will require the development of reduced description models while maintaining high parallel efficiency.

D. Applied mathematics and Numerical Analysis (AMNA): This focus area covers three main topics: (1) scalable algorithms, (2) gridding and meshing tools, and (3) optimization software and methods. In regard to scalable algorithms, the main requirement is for parallel PDE solvers, linear algebra libraries, preconditioners, and eigensolver algorithms for better convergence and accuracy. In regard to meshing, it involves techniques for handling structured grids, unstructured grids, and hybrid grids (the latter being particularly important to the BESI focus area). Parallel partitioning is essential to minimize communication and optimize load balancing on unstructured and hybrid grids. Also, parallel adaptive mesh refinement is essential to make optimum use of computing resources by putting the grid points where they are needed in the problem space. In regard to optimization software, this is typically needed at a high level where several terascale modules are integrated.

E. Computer Science and Software Engineering (CSSE): This focus area covers three main topics: (1) software integration tools, (2) parallel programming tools, and (3) performance analysis, monitoring, and enhancement tools. Software integration tools are needed to link terascale modules, e.g. linking linac and ring modules for proton driver simulations or linking linac simulations with plasma-based simulations to study the afterburner concept. Parallel programming tools are needed to facilitate the efficient development of high performance parallel modules. Performance analysis, monitoring, and enhancement tools are needed to observe and optimize the performance of parallel applications. Attention must be paid to single-processor performance (e.g. cache utilization), multi-processor performance, domain decomposition, and load balancing.

F. Visualization and Model Evaluation (VME): This focus area covers methods for large-scale visualization, run-time visualization, remote visualization, and methods for sharing and deploying data and programs. It also includes the development and application of statistical methods for the evaluation of complex computer models where computational and physical experimentation are expensive and the input space (boundary conditions, input variables, model parameters, etc.) and output space are large. In particular these methods will focus on the development of quantitative methods for the evaluation of prediction uncertainty for specific model outputs deemed important to support accelerator design decisions.

4.2 Collaboration with the Applied Mathematics and Computer Science Communities

As should be clear from the descriptions of the preceding three focus areas (D, E, and F), the successful development of terascale applications codes for accelerator design will require close collaboration of accelerator physicists with applied mathematicians, numerical analysts, computer scientists, geometry builders, visualization experts, etc., to develop new theoretical formulations and new algorithms capable of high performance and scalability on massively parallel systems. To this end, representatives of the SciDAC Integrated Software Infrastructure Centers (ISICs) were invited to Snowmass to describe to the community the HPC tools that are under development. The following centers/projects were represented:

<u>Terascale Optimal PDE Simulations (TOPS) Center</u>: David Keyes (Old Dominion University) presented an overview of the TOPS Center. This center will focus on developing, implementing, and supporting optimal or near optimal schemes for PDE simulations and closely related tasks. It will be concerned primarily with five PDE simulation capabilities: adaptive time integrators for stiff systems, nonlinear implicit solvers, optimization, linear solvers, and eigenanalysis. The center will research, develop, and deploy an integrated toolkit of open source, (nearly) optimal complexity solvers for the nonlinear partial differential equations that arise in many DOE/SC application areas, including fusion, accelerator design, global climate change, and reactive chemistry. These algorithms – primarily multilevel methods – aim to reduce computational bottlenecks by one to tree order of magnitude on terascale computers.

Terascale Simulation Tools and Technologies (TSTT) Center: Lori Freitag (ANL) presented an overview of the TSTT Center. A central goal of this center will be to address the technical and human barriers preventing the effective use of powerful adaptive, composite, and hybrid methods that are needed for many applications to obtain optimal simulation from terascale hardware. To achieve this goal the center will develop and deploy the mechanisms needed to enable the use of multiple strategies within a single simulation on terascale computers. To this end, the center will focus its effort in the areas of *high-quality, hybrid mesh generation* for representing complex and possibly evolving domains, *high order discretization techniques* for improved numerical solutions, and *adaptive strategies* for automatically optimizing the mesh to follow fronts or to capture important solution features. The center will work with SciDAC-supported projects in plasma physics, accelerator simulation, climate modeling, and other applications.

Algorithmic and Software Framework for Applied PDEs: Phillip Colella (LBNL) presented an overview of a SciDAC project whose goal is to develop a high-performance algorithmic and software framework for solving problems in PDEs arising in three important mission areas for DOE/SC: fusion, accelerator design, and combustion. This framework will provide investigators in these areas with a new set of simulation capabilities based on locally structured grid methods, including adaptive meshes for problems with multiple length scales; embedded boundary and overset grid methods for complex geometries; efficient and accurate methods for particle and hybrid particle/mesh simulations; and high-performance implementations on distributed-memory multiprocessors.

Centers that were not represented at the Snowmass meeting, but for which the accelerator community plans to collaborate, cover the areas of visualization, code integration, and performance optimization.

In addition to the above presentations describing specific ISICs, there were other applied mathematics and computer science talks presented on the T7 Working Group. The PLIB suite of software was described by Viktor Decyk. PLIB is a library of parallel high-level abstractions that was developed (under the Numerical Tokamak Transport Project) to aid in the development of high-performance parallel PIC simulations. PLIB has already played a significant role in the particle-manager capabilities currently implemented in IMPACT code and in a reduced-description plasma-accelerator code. Ongoing PLIB development activities include improved load balancing, more asynchronous operation, and optimized operation in a shared-distributed memory environment.

In other presentations, Esmond Ng (LBNL/NERSC) described parallel software for solving large, sparse eigenvalue problems; and Richard Gerber (LBNL/NERSC) gave an overview of the resources available at NERSC, which is the flagship scientific computing center for DOE/SC. The director of NERSC, Horst Simon, provided a glimpse of what lies ahead in a talk on the future of supercomputing.

5. Conclusion

Given the great importance of particle accelerators – especially the critical role they play in enabling great science – it is imperative that the most advanced computing technologies be used for their design, optimization, commissioning, and operation. The accelerator community has made remarkable progress in the last 10 years in embracing HPC. While its presence was barely noticeable to the broader HPC community in the early 1990's, it is now recognized as one of several areas – along with climate modeling, combustion modeling, materials modeling, quantum chemistry simulation, bioinformatic and biosystems simulation, and many others – where HPC is having an impact on addressing problems of national importance.

In order that HPC simulation of accelerators have maximum impact on DOE programs, we recommend that the following be adopted:

1. To ensure that the DOE gets that most science possible out of its current facilities, we recommend that HPC capabilities be developed to model existing machines with greater accuracy and realism than has been possible up to now, and that large-scale simulations be carried it to help improve the performance and expand the operational envelopes of those facilities. Examples include modeling losses in high intensity machines such as the FNAL booster and the BNL AGS, and modeling the strong-strong beam-beam interaction in machines such as the PEP-II B-factory.
2. We recommend that HPC capabilities be developed with a view toward reducing the cost and risk associated with next-generation accelerators, such as NLC, VLHC, muon-based systems, and proton drivers.
3. We recommend that HPC capabilities be developed in order to study the feasibility of proposed systems, especially those for which experimental data are currently lacking. An example of this is the simulation of an ionization cooling channel for a neutrino factory or muon collider. (A high-fidelity simulation of a channel that achieved the desired cooling would provide a strong incentive to move forward with an experimental program.)
4. We recommend that HPC capabilities be developed to help advance the technology of laser- and plasma-based accelerators.
5. In order to have a powerful HPC capability within the HEP community that can be used, in concert with theory and experiment, to advance the state-of-the-art in accelerator science and technology, and help make next-generation machines a reality, we recommend a target of 1% as the fraction of the HEP budget supporting base funding in HPC accelerator simulation. This support should include not only National Laboratories but also University groups, with the objective of creating a strong program to further develop and use HPC accelerator simulation tools. Given the possible impact of simulation on design optimization and decision making – which can have financial consequences ranging from a few million dollars to as great as $1 billion in the most expensive proposed machines – the likely return far surpasses this modest investment.

6. List of participants

Name	Institution	E-Mail
Andreas Adelmann	PSI/ETH	andreas.adelmann@psi.ch
Thomas Antonsen	U. of Maryland	antonsen@glue.umd.edu
Martin Berz	MSU	berz@msu.edu
Phil Colella	LBNL	pcolella@lbl.gov
Viktor Decyk	UCLA	decyk@physics.ucla.edu
Rich Fernow	BNL	fernow1@bnl.gov, fernow@bnl.gov
Lori Freitag	ANL	freitag@mcs.anl.gov
Richard Gerber	NERSC	RAGerber@lbl.gov
Dan Gordon	NRL	gordon@ppdmail.nrl.navy.mil
Vladimir Grichine	CERN	Vladimir.Grichine@cern.ch
David Grote	LBNL	DPGrote@lbl.gov
Salman Habib	LANL	habib@lanl.gov
Helmut Haseroth	CERN	Helmut.Haseroth@cern.ch
Andreas Kabel	SLAC	akabel@slac.stanford.edu
David Keyes	LLNL	keyes3@llnl.gov
Kwok Ko	SLAC	kwok@slac.stanford.edu
Paul Lebrun	FNAL	lebrun@fnal.gov
Zenghai Li	SLAC	lizh@slac.stanford.edu
Kyoko Makino	UIUC	makino@uiuc.edu
Edward May	ANL	may@anl.gov
Warren Mori	UCLA	mori@physics.ucla.edu
Esmond Ng	NERSC	EGNg@lbl.gov
Cho Ng	SLAC	cho@SLAC.Stanford.EDU
Chet Nieter	U. of Colorado	nieter@mail-beams.colorado.edu
Jean-François Ostiguy	FNL	ostiguy@fnal.gov
Ji Qiang	LANL	jiqiang@lanl.gov
Hong Qin	P.P.P.L.	hqin@pppl.gov
Rob Ryne	LBNL	RDRyne@lbl.gov
Roman Samulyak	BNL	rosamu@bnl.gov
Horst Simon	LBNL	HDSimon@lbl.gov
Panagiotis Spentzouris	FNAL	spentz@fnal.gov
Frank Tsung	UCLA	tsung@physics.ucla.edu

7. List of talks

A. HPC Tools

1. Robert Ryne, Issues and Goals for the T7 Working Group on High Performance Computing
2. Phil Colella, An Algorithmic and Software Framework for Applied Partial Differential Equations
3. Lori Freitag, The Terascale Simulation Tools and Technologies (TSTT) Center
4. David Keyes, Terascale Optimal PDE Solvers (TOPS)
5. Viktor Decyk, Parallel Tools for Building High-Performance Particle Simulation Codes
6. Richard Gerber, NERSC Overview: Hardware, Software, and Intellectual Services

B. HPC Simulation of Beam Systems

1. Salman Habib, Overview of Methods for the Large-Scale Simulation of Classical and Quantum Evolution Equations
2. Ji Qiang, Status of the IMPACT Parallel Beam Dynamics Code
3. David Grote, Overview of the Parallel Particle-In-Cell Code WARP for Modeling High Current Particle Beams
4. Andreas Kabel, Parallel Simulation of Coherent Synchrotron Radiation Effects
5. Robert Ryne (for Andreas Adelmann), Status of MAD9P, a Parallel Version of MAD with Space Charge
6. Hong Qin, Overview of the BEST Code
7. François Ostiguy, VLHC Modeling Needs

C. HPC Simulation of Electromagnetic Systems

1. Horst D. Simon, The Future of Supercomputers
3. Cho Ng, High Resolution Cavity Design
4. Cho Ng, System Scale Wakefield Computations
5. Esmond Ng, Computational Issues in Parallel Electromagnetic Simulations

E. HPC Simulations for Muon Based Systems (Joint with M1)

1. Robert Ryne, Overview of HPC Beam Dynamics Simulations
2. Cho Ng, Overview of HPC Electromagnetics Simulations
3. Panagiotis Spentzouris, HPC Ionization Cooling Simulations
4. Panagiotis Spentzouris, Future Computing Needs Associated with Muon-Based Accelerator Systems
5. Paul Lebrun, GEANT4 and DPGeant Ionization Cooling Packages
6. Richard Fernow, The ICOOL Ionization Cooling Package
7. Helmut Haseroth (for Alessandra Lombardi), CERN Cooling Packages (PATH)
8. Martin Berz, Normal Form Methods and Optimization for Nonlinear Properties of Cooling Channels
9. K. Makino, Cooling Channel Simulation based on Map Methods