Accelerator Simulation for the 21st Century

submitted to the Basic Science Component of the
Scientific Simulation Initiative

May 10, 1999

Contents

Chapter 1

Executive Summary

Chapter 2

Introduction

2.1 Early History of Accelerators

At the end of the last century, Crookes, Thomson, Roentgen, and others built what today we would call the first particle accelerators. They consisted of evacuated glass tubes about a foot long equipped with various internal electrodes. High voltages, as high as could then be measured and applied given their rudimentary electrical technology, were applied to some of these electrodes (and perhaps lower control voltages were applied to others) with various results.

In carefully planned experiments with his tubes, Thomson demonstrated the particle-like nature of cathode rays, and measured their mass and charge. In so doing, he discovered the electron, the first example of an elementary particle.

By pure serendipity (but very careful observation) with his tubes, Roentgen discovered X-rays. In so doing, he provided key information for the modern understanding of light and the vast extent of the electromagnetic spectrum.

Today almost every home and every office has one or several direct descendants of Thomson's tubes in the form of TV tubes and computer monitors. Moreover, in many ways, the work of Thomson laid the foundation for all of modern electronics.

Similarly, every hospital and every doctor's and dentist's office has one or several descendants of Roentgen's tubes in the form of X-ray machines. And again, his work too has had a vast impact on both our current technology and our fundamental understanding of the nature of matter.

2.2 Importance of Accelerator Physics Theory and Technology

In our century large Particle Accelerators and their associated Particle Detectors are among the most vast and complicated projects ever built by human kind; and experiments based on Accelerator Physics theory and technology have made remarkable discoveries about the basic nature of matter. These discoveries include nuclear structure, the behavior of nuclear matter, quark dynamics, the nature of elementary particles and the fundamental forces, unified field theories, and the origin and development of the universe. Future possible discoveries include quark plasmas, the origin of the asymmetry between matter and antimatter, supersymmetric counterparts to the known existing particles, the fundamental origin of mass, and a unified understanding of the ultilmate nature of matter.

The US Department of Energy (DOE) High Energy and Nuclear Physics (HENP) science programs are world leaders in this fundamental science. These programs have a distinguished history of important discoveries, and a large number of Nobel Prizes bear testimony to their recognition.

At the same time, Accelerator Physics theory and technology have made substantial contributions to other branches of science and technology. Major existing applications include electron microscopy, microprobes, charged-particle-beam lithography, ion implantation, isotope production, particle beams for precision irradiation medical therapy, superconducting magnets and medical magnetic resonance imaging, neutral-beam heating of plasmas, synchrotron light sources, X-ray lithography, and free-electron lasers.

Just this past year synchrotron light was used in the field of Biology to determine how meters of DNA can be coiled and managed in cells, to determine the long-sought structure of bacteriorhodopsin, and to determine the largest X-ray crystal structure to date, that of the bluetongue virus made of more that 1000 separate proteins. Indeed, because of their great utility, some 26 new synchrotron light sources are anticipated worldwide.

The principal new facility in the FY1999 DOE budget is the $1.3B Spallation Neutron Source. This collaborative project, involving several DOE laboratories, will make fundamental contributions to both the material and biological sciences.

Possible new applications of Accelerator Physics theory and technology include intense beams for inertial fusion; the production of tritium; the production of nuclear fuels; the transmutation, treatment, and disposal of nuclear waste; and high-speed and high-resolution proton radiography for nuclear stockpile stewardship.

Research efforts in Accelerator and Particle Physics have also had remarkable indirect benefits. For example, to meet their needs, scientists working in these areas have also made fundamental contributions to the development of Information Technology including the founding of the World Wide Web, the ability to handle large-scale distributed data bases, and electronic journals.

Particle accelerators are fundamental tools in all four of the DOE mission areas of Science, Energy, National Security, and Environmental Restoration. The total capital investment in accelerator facilities and laboratories over the past 25 years has been approximately $XXX. The operating budget for these facilities and their associated scientists (including those in National Laboratories, Universities, and Industry) is approximately $YYY per year.

2.3 Advances in Computation, Understanding, and Accelerator Design

(One or a few brief motivating paragraphs about what we hope to do plus a brief summary of what appears in the remaining sections of the proposal.)

(HERE IT IS WITHOUT THE MOTIVATING PARAGRAPHS.)

The goals of this proposal are three-fold:

1. To have a major impact on DOE accelerator projects by using large-scale simulation to significantly reduce cost and risk in the design of next-generation accelerators,

to improve performance of exisiting facilities, and to improve our understanding of accelerator and beam physics by combining theory and large-scale simulation with experimental results from existing facilities.

2. To develop new, and extend current, tera-scale capabilities to revolutionize the state-of-the-art in accelerator simulations on parallel and distributed platforms.

3. To use large-scale simulation as a tool of discovery to advance the frontiers of accelerator science and technology.

Each of these areas is described in more detail in the following sections.

Chapter 3

Project Goals

3.1 Impact to DOE Projects

A major goal of our initiative is to have a dramatic impact on the design of next-generation accelerators. Major projects, planned for the near term and the more-distant future, include the Next Linear Collider (NLC), the Spallation Neutron Source (SNS), a muon collider, and a 4th Generation Light Source. All of these projects will benefit greatly from large-scale modeling, and for some projects such modeling is essential. For example, the NLC damped detuned structure (DDS) is a highly complex, multi-cell, three-dimensional electromagnetic strucure that is a critical component of the NLC. The purpose of the structure is to decohere undesired deflecting fields inside the DDS and transport any such residual fields out of the structure to high order mode external loads. The design of such a structure, which involves modeling 206 cells simultaneously, where each cell involves an unstructured grid of up to 1 million grid points, is impossible on all but high performance computing platforms. Such modeling is essential to the success of the project. Furthermore, the optimization of such a structure can lead to power efficiency improvements that may result in a cost savings of 100's of millions of dollars. Tera-scale modeling issues for the NLC and other projects are described below.

Besides impacting next-generation accelerators, large-scale simulation is an important tool in maximizing the DOE's multi-billion dollar investment in existing facilities. Such simulations, coupled with theory and experiment, are essential to understanding certain phenomena observed in existing facilities and improving the operation of those facilities, as well as the design of future facilities. Collective effects, in particular, can lead to instabilities that limit the intensity of high-average current rings and the luminosity of colliders. Multi-species instabilities, in particular, are extremely important phenomena leading to intensity limitations. This includes, for example, beam-beam instabilities that can occur in high-intensity colliders and which were first considered by Chao and Ruth [**?**]. Considerably more difficult to understand are the instabilities due to background particles in machines. The Aladdin light source was plagued with background ion instabilities until a complete set of clearing electrodes was installed. The CERN ISR had a persistent electron-proton instability that limited the luminosity and was dealt with by greatly improved vacuum to eliminate the electrons. The Proton Storage Ring (PSR) (a spallation neutron source) also seems to have an electron-proton instability due to background electrons (perhaps from multipactor). Progress is being made in the ability to model these systems, as evidenced by recent work of Galambos *et al.* which predicted the longitudinal tails in the PSR in agreement with measurements to an accuracy of order 1% [**?**]. Still, the ability to predict small beam fractions on the order of 1-100 parts per million, as is required for applications such as APT and SNS, will require at least a 100- to 10000-fold increase in computing capability compared with serial computations on commodity microprocessors. Finally, it is expected that the B-factories being completed may also have instabilities due to background species (ions for the electron rings, electrons for the positron rings). When this effect is caused by a run-away build up of electrons it is referred to as the electron-cloud effect. This effect is an important issue not only for B-factories (in the positron ring) but also in some hadron machines such as the Large Hadron Collider.

The ability to realistically model collective effects is essential to advancing the frontiers of accelerator science, and such modeling constitutes a major deliverable for this project. Consider, as a case in point, that no one has ever performed fully 3D, self-consistent beam-beam calculation in the strong-strong regime (i.e. in the regime where both colliding beams are treated self-consistently). The accelerator community is now in an excellent position to do this. Three-dimensional, parallel space-charge solvers have been developed under the DOE Grand Challenge in Computational Accelerator Physics; this represents the foundation upon which a strong-strong beam-beam capability can be developed, working closely with accelerator applications experts in collider design and simulation. As another example, the ability to model instabilities due to background particles, such as the electron-cloud effect, will require knowing with high precision the electromagnetic fields in cavities having extremely complex 3D geometries. These fields are available through the development of the parallel eigenmode code $\Omega 3P$ developed under the Grand Challenge. Such a capability, combined with an ability to track particles through massive hybrid grids (i.e. grids containing 100's of millions of mesh points and containing structured and unstructured portions), combined with an ability to accurately treat surface effects, will bring electron-cloud simulations to a completely new level of accuracy and make realistic simulations of this effect possible.

Another aspect of collective effects is the exciting possibility of using laser- and plasma-based techniques to obtain ultra-high accelerating gradients of order 1-100 GeV/m, orders of magnitude greater than with conventional techniques. Tera-scale simulations, combined with theory and experiment, will be crucial to advancing the state-of-the-art in plasma-based acceleration schemes from research programs to realistic devices based on a well-understood and controllable technology.

At a minimum, modeling wakefield induced instabilities will require accurate models of the impedances of structures, which will be available through the parallel electromagnetics capabilities described elsewhere in this proposal. Furthermore, like modern fighter aircraft that are intrinsically unstable, accelerators operating beyond the limits of stability can only operate with active feedback control. While feedback damps the center-of-mass motion of a beam, it does not damp the rms moments of the beam, nor does is damp the higher order moments. Understanding the behavior of intense beams and the efficacy of feedback systems beyond centroid motion requires time-consuming simulations that involve tight coupling between the beam dynamics (using parallel particle codes) and models of the feedback control system itself. Such simulations will make it possible to increase the dynamic aperture of machines operating beyond the regime of stability through continuous feedback control.

In the latter half of this century the accelerator physics community brought the field of magnetic optics from its infancy to a level of maturity that makes this development one of the great successes in computational accelerator physics. (In the process, the development of sophisticated mathematical techniques even led the resolution of some problems that stymied Hamilton himself [**?**]). These advances, coupled with advances in parallel particle simulation techniques and the availability of SSI-scale resources means that the accelerator community is ready to do now for collective effects modeling what it did so successfully for single particle modeling earlier this century. The development of a tera-scale accelerator simulation capability will have a profound effect on the design of next-generation accelerators. The following sections provide an overview of several planned projects and describe the need for tera-scale simulation in those projects.

3.1.1 Next Linear Collider

Roughly a paragraph briefly describing the project and its importance,
and a paragraph on tera-scale simulation needs and how they will be addressed in the proposed work

3.1.2 Spallation Neutron Source

Roughly a paragraph briefly describing the project and its importance,
and a paragraph on tera-scale simulation needs and how they will be addressed in the proposed work

3.1.3 Muon Collider

Overview: The muon collider represents a very compelling option for the next generation lepton accelerator, both because of its physics discovery potential [6] and because of technical advantages such as its compact size, and precise beam energy control. With the possibility of building a Neutrino Factory based on Muon Storage Ring gaining reality, an unprecedented era of exciting Neutrino physics may be in the waiting as well. Of course, such a machine requires the development of techniques never applied before to either accelerator or detector design. Large scale computer simulation and optimiation are essential for the success of this enterprise.

Tera-scale Needs: HPC modeling is essential to three design issues: (1) the muon collider source, which is a high intensity proton ring, (2) the ionization cooling channel, and (3) detector design in the presence of beam induced backgrounds. The muon collider source consists of... The design of an efficient and practical cooling system is one of the major challenges for the muon collider project... Another challenging simulation issue involves the design of the collider detector taking into account the large backgrounds generated by the muon beam. For example, with 4×10^{12} muons per bunch in a 2 + 2 TeV collider ring there are approximately 4×10^5 muon decays per meter giving rise to high energy electrons. These off-energy, off-axis electrons undergo bremsstrahlung when they traverse magnetic fields, and when they exit the beam pipe they interact and produce electromagnetic showers, hadrons and muons. Although much of this debris can be locally shielded, the muon decays close to the interaction point constitute a significant background to physics measurements. In general, the design of shielding either at the interaction point or further upstream, is a problem which couples the beam optics to the detector design. The optimal parameters have to be chosen in such a way that the physics measurements are not compromised neither from the background rate nor from the lack of acceptance. Furthermore, such a design depends on the capabilities of the pattern recognition and reconstruction software (for example, the ability to to identify and reconstruct secondary vertices such as those associated with b-quark decays). Detailed Monte Carlo simulations are used to optimize these designs. NEED TERA-ESTIMATES.

3.1.4 4th Generation Light Source

Roughly a paragraph briefly describing the project and its importance,
and a paragraph on tera-scale simulation needs and how they will be addressed in the proposed work

3.2 Tera-Scale Capability Development

A second major goal of this SSI proposal is to develop new, and enhance current, tera-scale capabilities to perform state-of-the-art accelerator simulations on parallel and distributed platforms. As mentioned previously, such a capability is needed to design the next generation of accelerators, which are advancing the frontiers of beam energy, intensity, quality, and system complexity.

The tera-scale capability development of our effort can be described both in terms of the phenomena to be modeled and in terms of the computer science and software engineering needed to model these phenomena on tera-scale systems.

3.2.1 Accelerator Physics Issues

The phenomena that we intend to address fall naturally into 3 categories: multi-particle beam dynamics (BD), electromagnetics (EM), and problems involving coupled BD/EM. Furthermore, one can identify accelerator physics issues running through the efforts:

- Multi-Particle Dynamics:
  - **self-field effects** including the general issue of space charge and specific issues such as beam-beam effects
  - **beam-environment effects** including the general issue of wakefields and the specific issue of impedance-based simulations
  - **collisional effects** including hard and soft collisions
  - **radiation effects** including coherent synchrotron radiation (CSR)
- Electromagnetics:
  - **Eigenmode analysis**
  - **Time-domain analysis** including broadband impedance boundary conditions)

  - **Electrostatics and Magnetostatics**

- Coupled Beam Dynamics and Electromagnetics:
  - **Plasma-Based Accelerators**
  - **Tightly coupled BD/EM problems (NEED MORE EXPLANATION, e.g. small rings)**

3.2.2 Accelerator Systems Issues

The above issues are common to many parts of accelerator systems such as injectors, linacs, rings, and transport systems. A major goal of this initiative is to develop an HPC-based accelerator systems-scale capability allowing, eventually, a source-to-target simulation by coupling HPC modules, developed at different sites, within a scripting environment. Beyond the purely tera-scale aspect of this proposal, our effort aims to bring to the accelerator community a coherent vision of, and paradigm for, accelerator simulation. Through the adoption of standards and the use of widely accepted tools and problem solving environments, the community will develop a flexible, extensible framework for accelerator modeling. Such a common vision and framework will greatly increase the productivity of the community, by enabling sharing and integration of software modules developed at numerous sites in the accelerator community.

3.2.3 Computer Science and Software Engineering Issues

Accelerator simulations rank among the most challenging scientific computational tasks, not only at the accelerator systems level but even at the level of individual accelerator components. In order to achieve a system-scale HPC capability, it is essential to make extensive use of CSET-related technologies. This is crucial to achieving high performance on distributed systems (through the use of parallel frameworks and libraries such as POOMA, PETSc, Aztec, and ParMETIS), enabling interoperability of HPC code modules and codes themselves (through the use of tools such as PAWS, SILOON, and TULIP), allowing remote visualization (through the use of Globus), and dealing with the huge amounts of data generated in large-scale simulations (using, for example, parallel visualization resources). Many of the CSET tools mentioned above are already in extensive use by the Computational Accelerator Physics Grand Challenge effort to be described later in this proposal.

The approach to software development in this proposal is based on a three-level concept (see Table 1). The top level (Level III) is the software description of a major accelerator component such as a linac or a ring. Eventually, at this level, we would aim to encompass a complete accelerator system by coupling separate sections such as linacs, rings, and colliders (including feedback and control sections). At the most basic level are HPC modules (Level I) such as Poisson solvers, eigenmode and time-domain electromagnetics solvers, and Langevin/Fokker-Planck solvers. Integration of these modules to form identifiable accelerator sections/applications at an intermediate level constitutes Level II.

Each Level II area constitutes almost a separate situation, with different time and length scales, different important physics, etc. Within each of these sections multiple Level I components must work together efficiently – in a linear accelerator segment, for example, one has the initial particle distribution algorithms, different beamline elements, different Poisson solvers, and so on. These Level I components will be internally tightly-coupled and as large as possible, but only to the point that they perform one specific task, e.g., distribution generation, a single beamline element, etc. The components will be chosen so that using one in a simulation represents a specific choice among several similar options (e.g., choosing one Poisson solver over another depending on boundary conditions). The components at Level I will be designed to be composed with some scripting language (e.g., Python), within a single parallel job, so that they can exchange data very efficiently. A full linac simulation made up of the selected distribution scheme, beamline elements, and Poisson solver method is such an example. This can be done for each section of a full acclerator simulation, where each section is formed by composing components from a set relevant to that section. The components for one section (say, the linac portion) do not have to be completely reuseable with the components for other sections (although clearly there will be some commonality due to the ubiquitous nature of low-level modules such as parallel Fast Fourier Transforms and convolution routines, sparse linear systems solvers; and at a more complex level, space charge, beam-beam, and radiation modules). There will thus be a set of "component pools," one for each Level II section of the whole system. Such an approach allows each section to be developed by different teams, and allows for different choices of source code language, parallel run-time system, and scripting interface.

NOTE: THIS TABLE NEEDS TO BE REDONE DUE TO THE CHANGES IN LEVEL II/III DESCRIPTIONS

Level I: HPC Modules	Space Charge (inc. beam-beam), Impedance, Collisions, Ionization, CSR, Single particle dynamics on hybrid grids, Poisson solvers on hybrid grids, Vlasov/Poisson PIC/Direct, Vlasov/Maxwell PIC, Linear systems solvers, AMR, Mesh Generation, Statics solvers, Eigenmode solvers, Time-domain solvers
Level II: HPC Integration/Applications	Ionization Cooling Systems, Plasma-Based Acceleration Sections, Rings via Vlasov/Maxwell
Level III: Accelerator system integration	Intense Rings, Colliders, High Current Linacs

Table 1: Three-Level HPC Software Hierarchy

Once each section of a complete accelerator using sets of tightly-coupled components has been put together with some scripting language glue, all the sections need to be able to work together. This constitutes Level III. At this level, each of the sections represents a separate program, perhaps even running on different computers and architectures. Also, the different sections can have very different timescales and resource requirements. To couple these "macro-scale" components, we will use PAWS to exchange simulation or control data at specific points between the applications. (The data-centric nature of PAWS is well-suited to HPC applications.) Within this hierarchy, one first creates each section job by writing scripts to assemble the various section components together, and then tests them separately.

This is followed by the use of one more script to start them all up and link their shareable data structures together. A great advantage of this system is that it allows completely separate, testable large units, and great flexibility in the language/parallelism/scripting methodology for each section. The only thing that needs to be agreed upon by different development teams is the tool used to share data among the separate applications. (We will use PAWS but the system will be flexible enough to allow other choices to be made in the future.) The ability to integrate Level II sections as described above would lead eventually to a complete, high resolution end-to-end simulation capability on high performance computers.

At this point it should be mentioned that, while a prime target of our effort is SSI-specific hardware, there is agreement in the community that this effort must also support the mid-range, i.e. activities on PC clusters, on which much experience has already been gained. As such, versions of the software would be supported that are written entirely with explicit message passing. This would be straightforward to implement within our software plan outlined above, since only the appropriate low-level routines would have to be modified.

3.3 Enabling Scientific Discovery

A third goal of this proposal is to use large-scale simulation as a tool of discovery to advance the frontiers of Accelerator Science and Technology. Like Moore's Law in the computer industry, the near-exponential growth in accelerator capability over time is governed by the Livingston Curve [**?**]. And, as in the computer industry, it is now perceived that such growth will not continue without breakthroughs in new technologies, collectively referred to as Advanced Accelerator Concepts. One example of such a technology is that of plasma-based accelerators, which provide a means to reach new regimes in acceleration gradients.

3.3.1 Ultra-High Gradient Accelerators

Present accelerators operate with gradients of below 50 MeV/m, due to limitations imposed by rf breakdown effects. In contrast, plasma-based systems hold the potential to reach ultra-high gradients in the range of 1-100 GeV/m. In fact, gradients of 100 GeV/m have already been achieved in the laboratory over very short distances. However, much work remains to be done, in terms of assuring beam quality and sequentially staging ultra-high gradient sections. High resolution simulations are crucial to advancing this field. Fortunately, the existence of experimental research groups at several sites means that there is ample opportunity to compare theory and simulation with experiment. In one especially intriguing experiment, known as E-157 and which will be held at SLAC in 1999, it is expected that a beam will be accelerated to 1 GeV over 1 meter in a plasma.

Besides the obvious possibility of using ultra-high gradients to reach TeV energies in a geographically small space, advanced accelerator concepts also hold the key to developing compact devices to accelerate beams to moderately high energies in very small spaces, e.g. on a table-top. Analogous to the impact of table-top terawatt lasers, the development of compact accelerators based on advanced accelerator concepts has a potentially huge payoff in basic and applied research, industry, and medicine.

3.3.2 Extreme Regimes of Particle Beams

There is a fundamental need, with tremendous discovery potential, to explore the physics of particle and light beams under extreme conditions of matter and energy densities, temperatures and pressures. Such conditions naturally prevail at, for example, the : (i) collision points of high energy colliders;(ii) storage and accelerating devices involving super-cooled electron, positron, muon, proton, antiproton and ion beams, including crystalline beams; (iii)the interaction region of ultrashort and ultra-dense laser-matter, laser-beam and particle beam-matter interactions e.g in ionization cooling for muons, plasma focussing lenses and other laser-plasma interaction aspects; (iv) situations involving various phase-space cooling mechanisms and (v) facilities supporting low emittance beams.

New physical states of particle and light beams can be anticipated under these extreme conditions e.g. frozen beams; fermionic beam-condensates; squeezed states of light in free electron lasers; Compton-wavelength dominated particle beam phase-space analogous to diffraction-limited laser beams, etc. Discovery and laboratory formation and control of any of these anticipated exotic beam states would be a much treasured prize. However, a proper theoretical and experimental investigation of such states leading to “discovery experiments” must necessarily rely on high-fidelity simulations of a complex system that involves many degrees of freedom in complicated mutual dynamical interaction. The various degrees of freedom differ from each other both quantitatively and qualitatively e.g. self-interacting charged relativistic particle beams, relativistic high density plasmas, high intensity laser fields and superposed microwave radiation. The light and matter fields influence each other mutually in such a strongly coupled system. A fully developed moving-frame particle-in-cell code together with high order Maxwell solvers on field grids, including quantum and radiative fluctuational, diffusive and dissipative effects is called for. A simple system could involve a stored high brightness electron beam at the quantum limit of phase-space density interacting with a diffraction-limited laser beam at high energy densities of 10^{18} watts/sq. cm. in presence of mm and THz microwaves in a cavity of volume 1 cm^3 containing a pre-ionized gas or plasma in the path of the beam. A typical computation for such a system would already require tera-scale computing, when the relevant time-scales of interest span all the way from a few femtoseconds to seconds. Such studies are crucial to pushing the capabilities of beam-based science and technology to its furthest limits – be it in ultrafast phenomona, or ultracold coherent condensates, or limits of coherent and incoherent synchrotron radiation, or limits of particle phyics interaction at beam collision points, or in optical and quantum computing.

Chapter 4

R&D Approach: Accelerator Science Issues and Computational Methods

4.1 Multi-Particle Beam Dynamics

4.1.1 Introduction

Simulation is an essential part of accelerator physics and technology. The variety of beam dynamics problems encountered is immense, ranging from controlling the quality of a high brightness electron beam in the few nanoseconds after its being born in an rf photocathode, to predicting the dynamic aperture (i.e. the region of long-term stability) of a high energy hadron collider where the beam may undergo billions of revolutions. Simulation is also essential to advancing our understanding of the physics of beams and their interaction with the accelerator environment, and includes such current problems as the electron cloud effect, the stability of non-equipartitioned beams, and intensity limitations in high current rings.

Current and future accelerators are pushing the frontiers of intensity, energy, beam quality and beam control. Tera-scale computing will bring simulation to the level of a tool of discovery that will help make dramatic scientific advances possible in accelerator science and technology. In the Multi-Particle Dynamics component of this proposal, we are emphasizing areas that have potentially high pay-off and broad applicability. They fall naturally into three areas: (1) Issues involving high intensity, such as phenomena in high current beams (e.g. space charge, beam halo), collective effects leading to intensity limitations and instabilities in high current rings, and the beam-beam interaction; (2) collisional effects, such as those affecting the beam lifetime in colliders and storage rings, and "short-term" effects such as those related to ionization cooling a would be found, for example, in a muon collider (OTHER COOLONG EXAMPLES?); and (3) radiation phenomena, such as coherent synchrotron radiation (CSR) which is a major issue in high brightness electron beams and 4th generation light sources. These topics are discussed briefly below, before describing the tera-scale simulation of these topics in later sections. NOTE: ARE WE GOING TO DO (START?) CSR IN THE FIRST YEAR? IF NOT, THERE IS TOO MUCH EMPHASIS IN THE ABOVE.

NOTE: SOME OF THE FOLLOWING MAY BE REPITITIVE AND SHOULD BE

REMOVED IF IT HAS BEEN COPIED TO AN EARLIER SECTION OF THE PROPOSAL.

Intensity Limitations and Instabilities in High Current Rings

Multiparticle effects pose severe difficulties for the present and proposed generation of circular accelerators. There are two classes of problems that occur: (1) intensity limitations due to space charge and instabilities that limit the intensity and/or luminosity regardless of the average power of the machine; and (2) space charge induced beam loss effects at a very low level that are important when the average power of the machine is high.

Single-Beam Tune Shift and other Collective Effects

Intensity limitations due to collective effects come in a variety of forms. The simplest is the usual limitation on space charge tune shift which results when one of the single particle tunes (number of transverse oscillations in one revolution) crosses a resonance due to modified oscillations from space charge and either single particle loss or beam envelope oscillations become severe. The Brookhaven booster has had much success in correcting and avoiding these resonances (but hasn't managed to cross integer resonances) by eliminating the driving terms of particular resonances throughout the acceleration cycle by carefully powering corrector magnets. Many machines would greatly benefit from the ability to run with a higher tune shift. Examples include the rapid cycling synchrotrons and/or accumulator rings for both spallation neutron source (SNS) and for muon colliders), and the machines in the Heavy Ion Fusion (HIF) program require it. The LLNL/LBNL recirculating HIF linac would have a tune shift of several dozen and the HIF accumulator ring proposed by GSI would have a tune shift greater than one. Accurate modeling of intense beams in rings will require tera-scale resources in order to investigate correction strategies that include the complex interplay of magnet nonlinearities and space-charge nonlinearities. Such modeling will be crucial in enabling present and future accelerators to operate in regimes of increased intensity and tune shift.

In addition to these straightforward intensity limitations, another class of intensity limitations occurs because of collective instabilities. The basic cause of all these instabilities is that an accelerator beam has a very large source of free energy available in the translational motion, and nature will take every opportunity to degrade large sources of free energy. There is now a large list of such instabilites that have been observed and that place strong constraints on the design of circular machines and often require the inclusion of complicated feed-back systems in the design. Even the classification of all such instabilities must take into account a number of different aspects of the circular accelerator beam.

There are single bunch instabilities in the longitudinal dimension, transverse dimension, and in multiple dimensions at once. The longitudinal resistive wall instability and microwave instabilities are purely longitudinal modes that have been observed and are important. Transverse resistive wall and more generally transverse impedance instabilities have been found and must be accomodated. There are also single bunch instabilities that involve both transverse and longitudinal degrees of freedom simultaneously; these include the head-tail and coupled-mode instabilities. The single bunch instabilites may be particularly important for the new high-intensity rapid cycling synchrotrons and/or accumulator rings for neutron sources and muon colliders.

Most (but not all) rings have multiple particle bunches and there are a host of instabilities that involve multi-bunch effects. In fact, many of the single bunch instabilities, both longitudinal and transverse, have multi-bunch mode analogues. The principal difference between single bunch and multi-bunch instabilities is that the frequency ranges are different and the resultant impedances can be quite different. The B-factories under construction are generally unstable to multi-bunch instabilities and must stabilize them with feed-back systems. In addition, high luminosity hadron colliders such as LHC, will be beset with these instabilities.

The Beam-Beam Interaction

An important limitation in high luminosity colliders is the beam-beam tune shift limitation. In this case the tune shift is due to the space charge of the colliding beam and the tune shift limit is considerably smaller than the single beam tune shift limit because of the strongly nonlinear nature of the beam-beam force. This limitation affects and severely limits the luminosity of all high-energy and nuclear physics colliders, present, under construction and proposed. These include the Tevatron, PEP-II, KEK B-factory, RHIC, LEP, muon colliders, RHIC, and the LHC at CERN.

The beam-beam effects have most often been studied using the "weak-strong" approximation, in which the dyanamics of the "weak" beam are calculated assuming that the dynamics of the "strong" beam are known *apriori*. Strong-strong simulations have been performed, but they are not performed self-consistently, instead assuming that the beams have a pre-determined density profile, such as a Gaussian. With the advent of large-scale and tera-scale resources, fully self-consistent strong-strong simulations (i.e. beam-beam Particle-In-Cell codes) are finally possible. Already, glimpses of new coherent phenomena have been observed in low resolution simulations. Tera-scale resources will have a major impact in our ability to study the beam-beam interaction and predict the impact of beam-beam phenomena. Such resources will enable fundamental studies of the beam-beam interaction in the strong-strong regime. Furthermore, it will enable studies of specific designs, in particular those related to multi-bunch operation in colliders. In such a situation, not only are local bunch collisions at the interaction point (IP) important, but so are distant interactions among bunches, namely those taking place over several meters within the interaction region. For example, if bunches cross at the IP at an angle of a few tens of milliradians, then such bunches will also interact with and disrupt other bunches in the train separated by roughly one meter. This scenario, leading to the so-called Pac-Man effect, can lead to significant beam disruption after a few million crossings, occuring in a time measured in minutes. As such, this is a serious concern for hadron colliders which must store beams for many (~ 20) hours. Tera-scale resources are crucial for modeling multi-bunch beam-beam phenomena in the strong-strong regime.

Ultra-Low Loss Requirements in Next Generation Accelerators

All of the above effects limit the intensity or luminosity regardless of the average power of the machine; in other words the above effects are catastrophic, when they occur most of the beam is lost almost immediately. High average power machines are also affected by phenomena that lead to small (noncatastrophic) losses of the beam because of the need to have low activation of a high average power machine. The machines that fall into this category are the rapid cycling synchrotron (RCS) and/or accumulator rings for spallation neutron sources (SNS and ESS) and the muon generator rings for muon colliders. Of course, these machines will also be limited by all the above effects, but the designers must also be concerned with more subtle effects such as slow leakage out of the RF buckets and with transverse halo formation due to a variety of effects. The understanding of these effects is not as advanced as the catastrophic effects because they are harder to observe, but they are crucial for these high average power machines. In recent years, significant progress in understanding effects such as the formation of beam halo. This has involved a combination of theoretical work and large scale (parallel) simulations. For example, large scale simulations performed on the CM-5 at LANL, the Origin 2000 at LANL, and the T3E at NERSC have provided validation to particle-core models that have greatly improved our understanding of the beam halo issue. Such simulations have also been used to study instabilities of intense charged particle beams (that can help populate the beam halo), the existence of beam equilibria and their evolution when subject to beam mismatch, the existence of a maximum particle amplitude for the transverse beam halo, the effects of rf nonlinearites on longitudinal halo formation, and coherent coupling phenomena in 3D bunched beams. NEED CITATIONS. These simulations have also been used to assess the performance of specific designs for the APT and SNS projects. NEED CITATIONS.

Collisional Phenomena

Ionization Cooling

THIS SECTION NEEDS TO BE WRITTEN

Radiation Phenomena

Coherent Synchrotron Radiation

THIS SECTION NEEDS TO BE WRITTEN

4.1.2 Mathematical Formulation and Computational Methods

Single Particle Nonlinear Dynamics (Magnetic Optics)

As a prelude to discussing the simulation of multi-particle phenomena, we provide a brief overview of the methods used to treat single-particle dynamics in Accelerator Physics. In the absence of radiation, the dynamics of individual particles are governed by a Hamiltonian, $H(\vec{q},\vec{p})$, which depends on the electromagnetic potentials, $(\vec{A},\phi)$, associated with the externally applied fields in the accelerator. The first step in any calculation is the specification of these fields. For example, in the case of a magnetic quadrupole (with its axis oriented along the z-axis of a Cartesian coordinated system), this would involve the specification of the quadrupole gradient $g(z)$ and some number of its derivatives, $g'(z), g'', \ldots$, depending on the order other calculation. Similarly, in the case of an RF accelerating gap, this would involve the specification of the electric field on-axis at fixed time, $E(z)$, and its derivatives. In the case of a parallel-faced bending with midplane symmetry, this would involve the specification of the midplane magnetic field, $B_0 = B(x, y = 0, z)$, and its derivaties with respect to x and z. In general, the high order treatment of beamline elements requires higher order derivatives of the functions used to specify the field.

The next step in the calculation of single particle optics is usually the specification of the Hamiltonian with a coordinate as an independent variable. The appropriate choice depends on the coordinate system that is most natural to describe the element itself. For example, it is the z coordinate in rectilinear elements such as quadrupoles, RF gaps, high order multipoles, and parallel-faced bending magnets. In the case of a sector bend it is more natural to use polar coordinates and treat the angle θ as the independent variable. It is possible to express the Hamiltonian with z or θ as the independent variable, so long as a particle does not "change direction" inside an element. The resulting Hamiltonians are

$$K(x,p_x,y,p_y,t,p_t;z) = -\left[(p_t+q\psi)^2/c^2 - m^2c^2 - (p_x-qA_x)^2 - (p_y-qA_y)^2\right]^{1/2} - qA_z. \tag{4.1}$$

and

$$K(r,p_r,z,p_z,t,p_t;\theta) = -r\left[(p_t+q\psi)^2/c^2 - m^2c^2 - (p_r-qA_r)^2 - (p_z-qA_z)^2\right]^{1/2} - qrA_\theta. \tag{4.2}$$

Notice that, in both of the above, the variables (z,p_z) no longer appear as dependent variables; instead, they have been replaced by (t,p_t) which correspond to arrival time (at the location z or θ) and the particle energy at that location, respectively.

Often one is interested in the motion of particles around some reference trajectory. This leads to the final step in setting up a beam optics calculation, namely, re-expressing the Hamiltonian in deviation variables around the reference trajectory, and expanding the Hamiltonian as a power series in these new (small) deviations. The result is a Hamiltonian with *hundreds* or *thousands* of terms in the six variables (x,p_x,y,p_y,t,p_t). For example, a third order calculation involves 209 terms, and a 5th order calculation involves 923 terms. If one is interested in higher order nonlinear effects, or if one is interested in high order representations of maps in order to combine maps without significant loss in accuracy, the number of coefficients increases rapidly with order. For example, calculations of order 7, 9, and 1involve 3002 terms, 8007 terms, and 18563 terms, respectively. It is worth noting that these denote the theoretical *minimum* number of terms needed to treat the dynamics at a specified order. Of course, if one includes other degrees of freedom such as spin, the number of required parameters increases.

The fact that accelerators consist of localized beamline elements, and that, globally, there is a complicated variation of the electromagnetic potentials with z and θ has led to accelerators being described as "hopelessly complicated" Hamiltonain dynamical systems [?]. In summary, we are faced with dealing with a Hamilonian containing a huge number of terms and which, if viewed globally, has a hopelessly complicated dependence on the independent variable.

The treatment of such systems has proceeded over nearly three decades, and has led to the development of (high order) Magnetic Optics. This field is, without doubt, one of the great achievements in Computational Accelerator Physics. It is now possible to design magnetic optical systems to very high order, and in fact computations can be performed, in principle, to arbitrary order in deviations from the reference trajectory. For example, electron microscopes have been designed that have no 5th order spherical aberrations. ARE THERE BETTER EXAMPLES? 7TH ORDER? Not only can systems be designed to meet certain criteria, they can also be designed to do so while optimizing other effects; for example, beam expanding telescopes have been designed to eliminate third order aberrations while minimizing fifth order effects. Such a capability is crucial since merely targeting effects at a given order can cause harmful effects at a higher order. Another example of the success of Magnetic Optics is in the area of achromatic beamlines. During the 1970's achromats were only designed through 2nd order; in the 1980's the first 3rd order achromat was designed, and it is now possible to design 5th order achromats. ALSO MENTION FINAL FOCUS IN LINEAR COLLIDER.

The equations governing the motion of individual particles, whether in a beamline element, a string of elements, and entire revolution (or multiple revolutions) in a circular accelerator are Hamilton's Equations,

$$\frac{d\vec{q}}{dt} = \frac{\partial \vec{H}}{\partial \vec{p}}, \qquad \frac{d\vec{p}}{dt} = -\frac{\partial \vec{H}}{\partial \vec{q}}. \tag{4.3}$$

If we let the six-vector $\zeta = (x, p_x, y, p_y, t, p_t)$ denote the canonical coordinates and momenta, then Hamilton's equations can be written

$$\frac{d\zeta}{dt} = -[H, \zeta], \tag{4.4}$$

where [,] denotes the Poisson bracket.

In the language of *mappings*, which are a major theme in modern Computational Accelerator Physics, we would say that there is a (generally nonlinear) map, $\mathcal{M}$, corresponding to the Hamiltonian H, which maps initial phase space variables, ζ^i, into final variables, ζ^f, and we write

$$\zeta^f = \mathcal{M}\zeta^i. \tag{4.5}$$

Sophisticated techniques now exist to compute maps to essentially any order, to combine and manipulate maps, and apply maps to phase space coordinates (or functions of the coordinates), and to analyze maps. Map analysis techniques, in particular, are essential in the design of beamlines in order that they have certain properties (usually related to the vanishing of certain monomials in the Lie polynomial representing the transport system).

Vlasov/Poisson Systems

The previous section dealt with the dynamics of single particles subject to Hamilton's equations, and the concept of a mapping $\mathcal{M}$ associated with the dynamics. The corresponding equation governing the beam distribution function, $f(\zeta, t)$, is simply the Liouville equation,

$$\frac{df}{dt} = -[H, f] = 0, \tag{4.6}$$

from which we obtain

$$\frac{\partial f}{\partial t} = [H, f]. \tag{4.7}$$

It is straightforward to show that the evolution of a distribution function (i.e. the solution of Eq. (4.7) is also contained in $\mathcal{M}$. Namely, a distribution function $f(\zeta, t)$ whose initial value is $f^0(\zeta) = f(\zeta, 0)$ evolves according to

$$f(\zeta, t) = f^0(\mathcal{M}^{-1}\zeta). \tag{4.8}$$

In the previous section we dealt with individual particles subject to externally applied fields. We can treat systems of particles by assuming that they are subject to both external fields and the mean field associated with the beam's own self fields (space charge). In many cases one can write

$$H = H_{ext} + H_{sc}, \tag{4.9}$$

where H_{ext} denotes the Hamiltonian in the absence of self-fields (i.e. it includes only externally applied fields along with purely kinematic terms), and H_{sc} denotes the Hamiltonian associated with the space charge fields. As mentioned earlier, H_{ext} is a hopelessly complicated function of the independent variable, and furthermore it is usually treated in perturbation theory around a reference trajectory. In contrast with the treatment of external fields, self-fields governed by H_{sc} are not normally treated as a power series in the canonical variables because the variation is too great over the domain of interest. In many cases H_{sc} is simply proportional to the scalar potential ϕ, which satisfies the Poisson equation:

$$\nabla^2\phi(\vec{q}) = -\rho(\vec{q}). \tag{4.10}$$

The combination of Eq. 4.7 and Eq. 4.10 constitute the Vlasov/Poisson system of equations.

Split-Operator Methods and the Treatment of Space Charge

Thanks to the DOE Grand Challenge in Computational Accelerator Physics, as well as advances in accelerator computations on serial computers, we are already making great progress in the area of using parallel particle simulation codes to model Vlasov/Poisson systems. One of the keys to our success is the use of split-operator methods. As already mentioned, beam dynamics calculations often involve a Hamiltonian that can be written as a sum of two parts, $H = H_{ext} + H_{sc}$. Such a form is ideally suited for the application of symplectic split-operator methods [1]. More generally, consider a Hamiltonian that can be written as a sum of two parts $H = H_1 + H_2$, where each part, separately, can be solved exactly or to some desired accuracy or order. In other words, suppose that we can compute the mapping $\mathcal{M}_1$ corresponding to H_1 and the mapping $\mathcal{M}_2$ corresponding to H_2. In our case H_1 includes the external fields, and $\mathcal{M}_1$ can be computed to any order using the techniques of Magnetic Optics; the second term, H_2, corresponds to the space-charge fields, and can be dealt with using parallel particle simulation techniques. Given $\mathcal{M}_1$ and $\mathcal{M}_2$, the following algorithm is accurate through 2nd order in τ,

$$\mathcal{M}(\tau) = \mathcal{M}_1(\tau/2)\ \mathcal{M}_2(\tau)\ \mathcal{M}_1(\tau/2)\ , \tag{4.11}$$

where τ denotes the time step (really, a step in the independent variable z or θ). As a side note, if we were to use a differnt splitting in which H_1 depends only on momenta and H_2 depends only on positions, then this algorithm is the same as the well-known leap-frog algorithm. However, the split-operator approach provides a powerful framework capable of dealing with the far more complicated Hamiltonians often encountered in accelerator physics. The method is easily generalized to more terms (i.e. more splittings) if necessary. Furthermore, symplectic split-operator methods are easily generalized to higher order accuracy in time. A well-known fourth-order algorithm is due to Forest and Ruth, and an arbitrary order scheme was derived by Yoshida [2][3]. There are also implicit symplectic methods based on this approach that do not require the Hamiltonian to be split into a sum of exactly solvable pieces [1].

To summarize, extending a magnetic optics code to treat propagation through a beamline element in the presence of space charge, using the above algorithm, involves the following: (1) transport of a numerical disribution of particles through half the element based on $\mathcal{M}_{ext}$, (2) solving Poisson's equation based on the particle positions and performing a space-charge "kick" (corresponding to $\mathcal{M}_{sc}$), and (3) performing transport through the remaining half of the element based on $\mathcal{M}_{ext}$. If the space charge is intense, this can be performed repeatedly on successive pieces of a beamline element; if the space charge is weak, it may be possible to achieve good accuracy by computing the space charge kicks infrequently, in which case $\mathcal{M}_{ext}$ would correspond to a string of elements (or perhaps even the map for one turn of a circular machine). Thus, an important feature of this approach is that it enables one to use large time steps (i.e. large spatial steps) in the regime of weak or moderate space charge. Essentially, it enables one to decouple the rapid variation of the externally applied fields from the more slowly varying space charge fields. If more accuracy is required, one can use the 4th order method of Forest and Ruth, although this requires 3 space charge calculations per full step instead of just one, and since this dominates the execution time it is costly.

At this point it should be apparent that a requirement of this approach is the ability to solve Poisson's equation efficiently. This can be accomplished using any parallel particle-based method (Particle-In-Cell, Fast Multipole, etc.) Since this calculation normally dominates the execution time, parallelization is essential. In addition, it should be pointed out that even the single particle dynamics in external fields (i.e. the application of $\mathcal{M}_1$) can be parallelized, since such calculations, if done to high order, involve Lie or Taylor series with hundreds or thousands of terms. Interestingly, because of inefficient cache usage when performing simple transformations, it is likely that high order maps (which are readily available in beam optics codes) can be used, with no performance degradation compared with low order maps.

Finally, a subtle point is that, while most beam dynamics codes use z or θ as the independent variable, Poisson's equation has to be solved *at fixed time.* Thus, prior to every space charge calculation it is necessary to convert from (x, p_x, y, p_y, t, p_t) to (x, p_x, y, p_y, z, p_z) (actually, all that is needed is (x, y, z)). Such a calculation makes sense and is easily accomplished if the particle motion is essentially ballistic over a distance corresponding to the bunch length, since a bunch contains a distribution of arrival times, and they must all be moved to a fixed time. For example, in a rectilinear beamline element, the transformation for the jth particle is given by

$$\begin{aligned} x^j &\rightarrow x^j + p_x^j \Delta z \\ y^j &\rightarrow y^j + p_y^j \Delta z \\ \Delta z &= t^j v_z^j. \end{aligned} \tag{4.12}$$

A similar tranformation holds for beamline elements described in polar coordinates (e.g. sector bending magnets), with the added complication that all the coordinates are deviations from the reference orbit (the arc of a circle) that have to be converted to absolute position in space.

In summary, split-operator methods provide the "glue" to join two major fields, Magnetic Optics and parallel particle simulation techniques. All that is required is (1) the ability to compute maps corresponding to external fields, (2) the ability to compute the space charge fields (normally accomplished using a parallel Poisson solver), and (3) a knowledge of the particle positions at fixed times, or the ability to compute it.

Particle-based and Direct Solvers

The Vlasov/Poisson equations are normally solved by one of two methods: particle simulation techniques and direct solvers. In the particle simulation approach, the beam distribution function is represented indirectly by macroparticles that evolve according to the underlying (single particle) equations of motion. The Poisson equation is solved at every time-step and may or may not involve the use of a spatial grid, though grid-based schemes are the most popular in the beam dynamics community at the present time. The particle simulation approach has been the workhorse of the beam dynamics community for many years. Typically a beam is represented by a collection of particles initially generated using Monte Carlo techniques, which are propagated through the accelerator under study using time advance algorithms. This approach has two major advantages: (1) in a grid-based scheme, the memory requirement scales as the cube of the system size for three-dimensional problems, not to the sixth power as is the case for direct solvers (see below), and (2) the method does not break down even when subgrid scales are generated by the nonlinear evolution. Disadvantages of this approach include numerical collisionality and sampling errors associated with using a finite number of macroparticles much smaller than the physical number of particles in the beam bunch. When one is interested in high spatial accuracy, this necessitates the use of a fine grid and a correspondingly large number of macroparticles. The systematics underlying this procedure is well-known [**?**] and for a typical simulation of an intense charged-particle beam requires roughly ten particles per grid cell. For example, a simulation with a 512^3 grid in (x,y,z)-space has 134 million grid points. A beam dynamics simulation would therefore require approximately 1 billion particles.

At this point it is worth commenting on the difference between beam dynamics simulations and plasma and cosmology simulations. In the case of beams, the number of charged particles in a bunch ranges from $10^9 - 10^{12}$ for typical applications (CHECK NUMBERS). As such, we are approaching a simulation regime under SSI where simulations can be run with close to or exactly the physical number of particles. This is a totally different situation compared to many plasma physics applications (typically 10^{20} particles) and cosmology (typically 10^{70} in dark matter simulation codes) (CHECK NUMBERS). This distinction is important because it allows for the almost complete elimination of numerical collisionality in beam physics codes and the reintroduction of the physical collisionality, if desired, using Langevin techniques without a crippling memory overhead.

An alternative approach is to use a direct method in which the beam distribution function is defined explicitly by its values on a numerical grid in phase space. The dimensionality of the grid is $2N_{dim}$, corresponding to the space spanned by $(\vec{q}, \vec{p})$, where N_{dim} denotes the dimensionality of the problem, e.g. $N_{dim} = 3$ for a three-dimensional simulation. For this reason, direct Vlasov solvers in two dimensions and higher can only be performed with reasonable accuracy on very large memory (tera-scale) parallel supercomputers. For example, a 2D problem performed on a 256^4 grid in 64-bit precision would require 32 GB to store a single copy of the main array, 64 GB if stored as a complex array as is common in spectral methods. In the 3D case, a 64^6 problem would require approximately 500 GB to store the array as a real array, and 1 TB for the complex case. For the first time, such simulations may be within reach on SSI hardware. Advantages of direct solvers include (1) No particle sampling errors and direct access to the phase space distribution function and (2) no load balance issues on parallel platforms. Disadvantages include (1) very high memory cost, (2) need for sub-grid smoothing in phase space if such structures develop due to the dynamics. The appropriate applications for direct solvers are in studies of beam equilibria and their stability, in fundamental studies of beam physics that require the distribution function, and when good statistical information is desired in the tails of the distribution.

Particle-In-Cell, Fast Multipole, and Hybrid Techniques

As already mentioned, single-particle optics is a very mature field. Furthermore, the need for high performance computing in single-particle optics calculations is restricted mainly to long-term tracking in circular machines. Since such calculations involve particles that do not interact with one another, these calculations are embarassingly parallel and can be performed easily on loosely coupled platforms such as PC farms. On the other hand, multi-particle dynamics, such as that governed by the Vlasov/Poisson equations, represents a great challenge to high performance computers. Fortunately, the parallel solution of the Poisson equation has been studied in many fields, including computational accelerator physics, plasma physics, astrophysics, and cosmology. In the widely-used Particle-In-Cell (PIC) approach, this involves 3 steps: (1) charge deposition on a numerical grid; (2) field solution on the grid; and (3) interpolation of the fields to the particles' positions.

It is important to note that, in the parallel PIC approach, there are three main bottlenecks to parallel performance: (1) the fact that particles, and the grid points that they contribute to, may reside on different processors, necessitating irregular interprocessor communication; (2) the fact that some processors may be depositing far more or far fewer particles on the grid than other processors (load imbalance); and (3) the fact that, traditionally, a field-solve is performed at each step, involving global communication, e.g. a Fast Fourier Transform. The first issue can by dealt with by using an optimized "particle manager," which ensures that, at every time step, particles and their associated grid points are on the same processor. A particle manager can also be used to balance the load after a specified number of time steps (or automatically, when certain conditions are satisfied). The third problem may be dealt with on an SMP cluster by reducing the number of processors involved in the field solution; another approach is to avoid a global field solve at every time step by instead using an algorithm in which the fields themselves are dynamical quantities. An example of the latter will be provided below.

NEED A DESCRIPTION OF FAST MULTIPOLE METHODS.

NEED A DESCRIPTION OF HYBRID METHODS, AND WHY THEY MAKE SENSE.

Direct Solvers

As mentioned previously, unlike particle-based methods, direct solvers represent the distribution function on a phase space grid. A Vlasov/Poisson code solves the equation:

$$\frac{\partial f}{\partial t} + (\mathbf{p} \cdot \partial_{\mathbf{q}})f - (\nabla V \cdot \partial_{\mathbf{p}})f = 0, \tag{4.13}$$

where $f(\zeta, t)$ is the phase space distribution function, and where ζ denotes the phase space variables. The potential V is a sum of external and space charge contributions. To solve this equation, one approach is to use a spectral method combined with the above-mentioned split-operator techniques. For example, a second-order algorithm for evolving the distribution function is:

$$f(\zeta, \tau) = \mathrm{e}^{-\frac{\tau}{2}\mathbf{p}\cdot\partial_{\mathbf{x}}}\ \mathrm{e}^{\tau\nabla V\cdot\partial_{\mathbf{p}}}\mathrm{e}^{-\frac{\tau}{2}\mathbf{p}\cdot\partial_{\mathbf{x}}} f^0(x,p). \tag{4.14}$$

To evaluate the above using a spectral method, one first performs a forward Fourier transform of $f^0(x,p)$ in position space, to which the right-most operator $\mathrm{e}^{-\frac{\tau}{2}\mathbf{p}\cdot\partial_{\mathbf{x}}}$ can be applied trivially; next one performs a backward transform in position and a forward transform in momentum, to which the middle operator can be applied; lastly one performs a backward transform in momentum and a forward transform in position, to which the third operator can be applied; a final backward transform in position returns f to the physical basis, and completes the step. In complete analogy to the situation in particle simulations, this can be generalized to arbitrary-order accuracy in time, if desired, using the algorithm of Forest and Ruth or the method of Yoshida. In summary, all that is required to perform such a calculation is a computer with very large memory (due to the use of a direct method), and an ability to efficiently compute Fast Fourier Transforms. Unfortunately at the present time it is difficult to perform large, cross-box FFTs on some SMP clusters (such as the SGI Origin 2000 using certain HIPPI bypass hardware). On other systems such as the IBM SP this may be more tractable.

Collisional Effects

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 (Langevin) techniques applied to the particle equations of motion. The treatment of particle collisions is more subtle and requires the distinction between weak and strong encounters. Weak or soft collisions refer to interactions in which $\delta v \ll v$ where δv is the change of velocity during the collision. Strong or hard collisions have $\delta v \sim v$. Soft collisional effects may be included by generalizing the Vlasov/Poisson equations to a Fokker-Planck form (usually the assumption of locality, i.e. impact parameter small compared to system size, is also made). The corresponding change in the dynamics of the macroparticles is reflected in the appearance of noise and damping terms in the original particle equations of motion as a result of a systematic expansion in inverse powers of the Coulomb logarithm [**?**]. Hard collisions can be included within the Boltzmann approximation. In particle codes, these sorts of collisions may be treated using Direct Simulation Monte Carlo (DSMC) techniques.

Depending on the problem at hand, the Fokker-Planck equation to be solved may be more or less complicated. The simplest situations arise when the damping and diffusion coefficients are both constants (coupling to external environment) or when simplifying assumptions such as an isotropic velocity distribution turn out to be valid. Simple Langevin models corresponding to some of these cases have already been added to our codes. We are now developing methods to solve the full Fokker-Planck equation using techniques very similar to those used in parallel Vlasov/Poisson solvers. Consider, for example, the Landau form of the Fokker-Planck equation,

$$\frac{\partial}{\partial t}f + \sum_{i=1}^{3} v_i \frac{\partial}{\partial x_i} f - \sum_{i=1}^{3}\left(\frac{\partial}{\partial x_i}\Phi\right)\frac{\partial}{\partial v_i} f + \sum_{i=1}^{3}\frac{\partial}{\partial v_i}F_i f - \sum_{i,j=1}^{3}\frac{1}{2}\frac{\partial^2}{\partial v_i \partial v_j}D_{ij} f = 0 \tag{4.15}$$

where the damping and diffusion terms can be cast in terms of the Rosenbluth potentials

$$F_i \propto \frac{\partial}{\partial v_i} H, \quad D_{ij} \propto \frac{\partial^2}{\partial v_i \partial v_j} G, \tag{4.16}$$

which in turn are given by

$$H(\vec{x}, \vec{v}) = \int d\vec{v}' \frac{f(\vec{x}, \vec{v}')}{|\vec{v} - \vec{v}'|}; \quad G(\vec{x}, \vec{v}) = \int d\vec{v}' f(\vec{x}, \vec{v}')|\vec{v} - \vec{v}'|. \tag{4.17}$$

Such a situation is clearly nonlinear since, in analogy to the computation of the self-fields in the Vlasov/Poisson case, the calculation of the quantities F and D (or equivalently H and G) depends on the distribution function. Furthermore, it is clear from (4.17) that by introducing a grid in velocity space and then depositing macroparticle velocities onto the grid, we may determine the damping and diffusion coefficients numerically using exact analogs of the techniques used to solve the Poisson equation in PIC algorithms. (It should be kept in mind that Fokker-Planck techniques can be correctly applied only when numerical collisionality effects are already unimportant in the corresponding collisionless simulation.)

Even in the local approximation that we have discussed here, the noise and damping terms in the Langevin equations that correspond to the above Fokker-Planck description are functions of both space and velocity. Thus, in principle, a six-dimensional space is required to hold the information relating to the values of the transport coefficients with a potentially enormous memory cost. Fortunately, the corresponding phase space resolution needed for acceptable accuracy is far lower than that required for acceptable force accuracy in the parent PIC simulation (since the quantity being calculated is a modest correction treated as a stochastic force). For this reason, it has just now become possible to carry out these simulations on state-of-the-art HPC platforms. Nevertheless, the calculation will still push the envelope is what is possible, even on SSI-class platforms. For example, even a crude 16^3 grid in position space would involve the equivalent of 4096 space charge calculations. Though one may use symmetry or near-symmetry to reduce this to perhaps one eighth this value, or 512, the prospect of performing 512 convolutions instead of one per step is clearly a task requiring very high performance. Also, the fact that the calculation of the damping and diffusion coefficients will take place on a small grid (e.g. $16^3 \times 16^3$ in $\vec{x} - \vec{v}$ space) means that 3D domain decomposition will likely be required, unlike the 1D or 2D decomposition that can be used for most Vlasov/Poisson problems. Lastly, it should be noted that the treatment of such a system using Langevin techniques would involve treating systems with multiplicative noise, since the stochastic force associated with F (appearing in the equation for $d\vec{p}/dt$) is velocity dependent. Fortunately, we have already developed a 2nd order algorithm for treating systems with multiplicative noise [**?**].

NOTE: SALMAN WILL WRITE SOMETHING ON LANGEVIN/DSMC

NOTE: PANAGIOTIS: PLEASE CONDENSE THE WEB INPUT YOU PROVIDED INTO 1/2 TO 3/4 OF A PAGE ON THE MATHEMATICAL AND COMPUTATIONAL ASPECTS OF MODELING IONIZATION PHENOMENA

Radiation Phenomena

Coherent Synchrotron Radiation

4.2 Electromagnetics

4.2.1 Introduction

4.2.2 Mathematical Formulation and Computational Methods

4.3 Plasma-Based Accelerators

THIS SECTION WILL INVOLVE MELDING TOGETHER VARIOUS PARTS OF THE INPUT PROVIDED BY THE UCLA/USC GROUP AND THE LBNL/UCB GROUP

4.3.1 Introduction

4.3.2 Mathematical Formulation and Computational Methods

Chapter 5

Past Accomplishments and Tera-Readiness

The Accelerator Physics community is already performing large scale simulations on parallel computers at DOE High Performance Computing Centers, namely, on the T3E at NERSC and on the Origin 2000 system at the LANL Advanced Computing Laboratory. The codes, most of which are written using explicit message passing, will be ready to be run on SSI hardware as soon as it becomes available. The community's HPC experience includes that associated with the DOE Grand Challenge in computational accelerator physics, as well as experience in parallel plasma-based accelerator modeling and parallel modeling for Heavy Ion Fusion research. The community also has a long and successful history of serial software development, particularly in terms of algorithms and frameworks, related to single particle dynamics (SPD). Each of the preceding areas is described below. The following described some of the community's current High Performance Computing activities.

5.1 High Performance Computing Experience

5.1.1 The DOE Grand Challenge in Computational Accelerator Physics

In 1997 the U.S. Department of Energy initiated a Grand Challenge in Computational Accelerator Physics to support a collaborative effort involving LANL, SLAC, UCLA, and Stanford, together with two HPC centers, NERSC and the ACL. The primary goal of this project is to develop a new generation of accelerator modeling tools for HPC platforms, and to apply them to large complex problems of importance in future accelerators, notably the NLC and future high current accelerators such as that associated with the Accelerator Production of Tritium (APT) project. The following provides a brief overview of progress-to-date in the Grand Challenge.

Beam Dynamics

A new 3D parallel PIC code called IMPACT has been developed as part of the Grand Challenge in Computational Accelerator Physics. IMPACT (Integrated-Map and Particle Accelerator Tracking code) is a state-of-the-art parallel PIC code for modeling high current linacs. The code has a novel treatment of RF accelerating gaps, based on similar techniques used in the Magnetic Optics community, obtained by numerical integration of the gap transfer map. Such an approach avoids having to perform fine-scale integration of single particle trajectories through the rapidly varying fields of the RF gaps. IMPACT uses a parallel 3D PIC routine, also developed as part of the Grand Challenge, for the treatment of space charge. Several versions of the code exist, including the original version written in HPF, and new, higher-performance versions written using the C++ objected-oreinted framwork POOMA [?] developed at LANL, a Fortran90-based parallel particle simulation library PLIB [?] developed at UCLA as part of Numerical Tokamak Project [?], and an object-oriented parallel particle simulation library developed in a collaboration involving LANL and UCLA [?]. All versions of IMPACT use dynamic load balancing.

An example output from IMPACT is shown in Figure 1. The system being modeled was a version of the superconducting linac for the APT project. The run was performed on 4 boxes of the Origin 2000 at the ACL, using 128 processors/box. The simulation was performed using 300 million simulation particles and a 256^3 physical grid (corresponding to a 512^3 grid internally). The total time for this run was X hours, and involved a total of Y space-charge kicks. It was visualized on a Silicon Graphics Reality Monster system with 4 Infinite Reality pipes. The color scale is logarithmic, indicating excellent resolution of the beam halo in the outer regions, where the beam density is more than 10,000 times smaller than the central density. A plot of the reference beam energy is shown in Figure 2. Two curves are shown in the expanded region: the dashed curve is based on a split-operator approach, with $H = H_{ext} + H_{sc}$, and uses only 5 steps per superconducting structure; the solid curve is based on 100 steps per structure. Note that, even with only 5 steps/structure, the points on the dashed curve coincide with the solid curve. This is because the split-operator method "knows" about rapid changes in the external field, which are shown superimposed on the Figure. For the parameters of this problem, split-operator methods enable one to take large steps through the accelerating structure without any noticable loss in accuracy. Parallel space charge calculations, which dominate the execution time, are performed only at the 5 points shown. Thus, the intelligent use of split-operator methods drastically reduces the execution time, i.e. to use fine-step integration of the particle trajectories through this region (presumably to follow the rapid changes in the external accelerating field) would be enormously wasteful.

Besides running on the ACL Origin 2000, IMPACT also runs on the NERSC T3E, and runs 200 million particles have been performed on the ACL PC cluster [?].

The previous example showed a visualization of a 300 million particle data set, using a high-end Graphics supercomputer. While our effort obviously targets high-end applications, we also intend to support mid-range users and mid-range resources (such as few hundred node PC clusters that are becoming extremely popular). For this reason, and because even parallel I/O can be a bottleneck in some applications, it is important to be able to perform internal data analysis so that one has a choice of whether or not to write what could amount to several hundred terabytes of data. We have developed parallel data analysis subroutines that enable us to extract important features at run time. An example is shown in Figure 3a. This Figure is from the same simulation as Figures 1 and 2. However, it is a 2D z-pz phase space plot using a feature known as a "low density cut," which is needed to be able to see very low density features in high-fidelity particle simulations. Figure 2a contains approximately 100,000 particles grouped in four regions. The density varies from 1 at the beam center to 1/10 at the edge of the central region. Similarly, the density equals 1/100, 1/1000, and 1/10000 at the edge of the remaining three regions. If we had simply plotted the same number of particles chosen randomly from the 300 million in the simulation, the halo would barely be visible and show almost no structure, as is evident if Figure 3b.

Parallel Langevin simulation codes have reached the first stage in our planned approach to develop, for the first time, a first-principles, self-consistent Fokker-Planck capability to treat collisions (see Section Z). In the current version of LANGEVIN3D the damping and diffusion coefficients are treated as constants. Such a model is useful, for example, to drive a beam to thermal equilibrium which can then be used as the input for a simulation to study beam evolution in an accelerator from an initially stationary state. This capability, with the diffusion term alone, is also useful for modeling the effects of noise in an accelerator. However, the present capability is one in which the beam energy changes (even in a constant focusing channel), with the final energy governed by a fluctuation/dissipation relationship. In an approach where the damping and diffusion coefficients are computed self-consistently from their defining relationships, the beam would be driven to a thermal equilibrium where the final temperature is that associated with the initial beam energy. Such a Fokker/Planck model, treated from a Langevin point of view, involves multiplicative noise. For this reason, a new 2nd order accurate stochastic integration algorithm has been developed [**?**].

An example from from LANGEVIN3D is shown in Figure X... Note that, while the beam does indeed equilibrate, the system energy is not constant, for the reason just discussed. Note also that, when the damping and diffusion terms are "turned off" at z=XXX, there is virtually no change in the beam emittance, as expected for a distribution that has reached thermal equilibrium.

Electromagnetics

The development of new electromagnetic tools for the Grand Challenge project originated from advanced accelerator structure research for the NLC. The main thrust of the effort is aimed towards large-scale simulations of realistic 3-D structures. Such a capability can be applied to system-scale analysis such as finding the wakefields in the entire DDS as described in Fig. 5.1, or to individual component design such as one cell in the DDS Fig. 5.2, by modeling with an accuracy approaching fabrication tolerance level. The new set of tools incorporates the following features to enable the high-resolution modeling required: (i) the use of unstructured grids to capture realistic geometries, (ii) the development of refinement algorithms to improve accuracy and optimize computing resources, and (iii) the implementation on parallel platforms to take advantage of the latest in HPC resources for large-scale simulations.

Figure 5.1: One end of the DDS 206-cell section including the input coupler and HOM load termination.

Figure 5.2: One-eighth of the DDS cell geometry from a solid model.

Figure 5.3: A non-orthogonal cell for the generalized Yee algorithm used in $\tau3$.

Presently, there are two types of solvers being developed for the tool set. The first type is formulated in the frequency domain using linear and quadratic finite elements on an irregular grid, and includes eigenmode solvers in two and three dimensions, named $\Omega 2$ and $\Omega 3$ respectively. The parallel versions, Ω2P and Ω3P, use MPI and currently run on the SGI/Cray T3E at NERSC. In the case of $\Omega 3P$, the calculation proceeds in three steps: First, there is the mesh distribution step with a module called DistMesh that uses the parallel library ParMETIS for partitioning unstructured graphs. This is followed by the matrix assembly step which consists of the finite element formulation that provides the mass and stiffness matrices for the generalized eigenvalue problem. The final step is a solution step, with Lanczos or Jacobi-Davidson algorithms, that utilizes the parallel iterative library AZTEC to solve the sparse linear systems.

The other type of solver is a three-dimensional time-domain code, called $\tau 3$, that uses a generalized Yee algorithm on an unstructured grid [**?**][**?**], Fig. 5.3. A leapfrog time advancement scheme with filtering is implemented as well as a broadband termination at the waveguide ports. Therefore it is able to handle pulse transmission for S-parameter evaluations of RF components over a wide frequency range in a single run. Dipole excitation is also possible to calculate external Q's of waveguide-loaded cavities. A parallel version, τ3P, currently runs on a shared memory machine, like the 4-node Intel Xeon server, using threads while the distributed memory version for the T3E is presently under development. The inclusion of a transiting rigid beam to compute wakefields is planned. NOTE: DON'T FORGET TO MENTION 34 NODE CLUSTER AT SLAC.

We present here some recent results from Ω3P and $\tau 3$. First, we report the modeling of a single DDS cell using Ω3P on the T3E. Fig. 5.4 shows the cell geometry partitioning into 8 pieces using DistMesh/ParMETIS. Fig. 5.5 is the mass matrix distribution over 16 processors. Together they demonstrate the importance of domain decomposition to achieve load balancing. Currently, Ω3P is the tool we use to determine the new dimensions for improved versions of the DDS design. Another Ω3P result comes from the Spallation Neutron Source project for which we modeled the RFQ cavity as shown in Fig. 5.6. This cavity is difficult to simulate accurately due to the disparate lengthscales between the focusing vanes, the stabilizing rods, and the cavity proper. Fig. 5.7 shows the convergence of the calculated frequencies to measured data for the two modes of interest when the resolution is increased to require several million degrees of freedom in the simulation.

The $\tau 3$ capability is demonstrated by the matching and tuning of the input power coupler for the NLC accelerator section. Fig. 5.8 shows the mesh used in modeling a pair of couplers separated by two regular cells. $\tau 3$ finds the match by either single frequency or pulse excitation. The reflection is determined to be 0.005 at 11.424 GHz. The match is confirmed by the accelerating field amplitude and phase along the beam axis. Fig. 5.9 indicates the field that is due to a travelling wave with the correct phase advance of 120 degrees per cell from coupler to coupler. The advantage of $\tau 3$ over other commercial packages will be the ability to model much larger problems when the parallel version, τ3P, is completed.

Figure 5.4: Domain decomposition of the DDS cell.

Figure 5.5: Mass matrix distribution over 16 processors.

Figure 5.6: RFQ cavity for the SNS.

Figure 5.7: Calculated frequencies and measured data for two modes of the SNS RFQ cavity.

Figure 5.8: Input power coupler for the NLC linac modeled by $\tau 3$.

Figure 5.9: Field amplitude and phase of accelerating field at matched condition.

5.1.2 Plasma-Based Accelerator Simulation

5.1.3 Simulation for Heavy Ion Fusion Applications

5.2 Single Particle Dynamics

AT MOST ONE PAGE.

Chapter 6

Implementation on Tera-Scale Platforms and Integration with CSET

BRIEFLY REVIEW TYPES OF CODES AND TYPES OF COMMUNICATION IN CODES.

6.1 Libraries and Frameworks

6.2 Problem Solving Environments

6.3 Visualization

6.4 Collaboration Tools

DON'T FORGET SECTION ON PERFORMANCE OPTIMIZATION WITH STRONG INPUT FROM ACL AND NERSC.

DON'T FORGET TO MENTION SOMEPLACE VISIBLE THE SUPPORT FOR THE MID-RANGE.

Chapter 7

Proposed Work

7.1 Specific Tasks

7.2 Personnel, Budget, and Management Plan

Bibliography

[1] E. Forest *et al.*, *Phys. Lett. A* **158**, p. 99 (1991).

[2] E. Forest and R. Ruth, *Physica D* **43**, p. 105 (1990).

[3] H. Yoshida, *Phys. Lett. A* **150**, p. 262 (1990).

[4] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, (Adam Hilger, NY, 1988).

[5] R. Ferrell and E. Bertschinger, *Int. J. Mod. Phys. C* **5**, p. 933 (1994).

[6] Muon collider Status report in preparation (need better reference here)

[7] G. I. Budker and A. N. Skrinsky, *Electron cooling and new possibilities in elementary particle physics*, Sov. Phys. Usp. **21**, p. 277 (1978).

[8] A. N. Skrinsky and V. V. Parkhomchuk, *Methods of cooling beams of charged particles*, Sov. J. Part. Nucl. **12**, p. 223 (1981).

[9] The $\mu^+\mu^-$ Collider Collaboration, *$\mu^+\mu^-$ Collider Feasibility Study*, BNL-52503, FERMILAB-Conf-96/092, LBNL-38946 (July 1996)