Scientific Grand Challenges

CHALLENGES FOR UNDERSTANDING THE QUANTUM UNIVERSE AND THE ROLE OF COMPUTING AT THE EXTREME SCALE

December 9-11, 2008 • Menlo Park, CA

U.S. DEPARTMENT OF ENERGY

Sponsored by the Office of Advanced Scientific Computing Research and the Office of High Energy Physics

DISCLAIMER

This report was prepared as an account of a workshop sponsored by the U.S. Department of Energy. Neither the United States Government nor any agency thereof, nor any of their employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of document authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Copyrights to portions of this report (including graphics) are reserved by original copyright holders or their assignees, and are used by the Government's license and by permission. Requests to use any images must be made to the provider identified in the image credits.

On the cover: The IBM Blue Gene/P supercomputer at the U.S. Department of Energy's Argonne National Laboratory. The computer, dubbed the Intrepid, is the fastest supercomputer in the world available to open science and the third fastest among all supercomputers. Future reports in the Scientific Grand Challenges workshop series will feature different Office of Science computers on their covers.

SCIENTIFIC GRAND CHALLENGES: CHALLENGES FOR UNDERSTANDING THE QUANTUM UNIVERSE AND THE ROLE OF COMPUTING AT THE EXTREME SCALE

Report from the Workshop Held December 9-11, 2008

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

Chair, **Roger Blandford**

Stanford Linear Accelerator Center

Co-Chair, **Norman Christ**

Columbia University

Co-Chair, **Young-Kee Kim**

Fermi National Accelerator Laboratory

Breakout panel chair, Cosmology and Astrophysics Simulation, **Michael Norman**

University of California, San Diego

Breakout panel chair, Astrophysics Data Handling, Archiving, and Mining, **Alex Szalay**

Johns Hopkins University

Breakout panel chair, High Energy Theoretical Physics, **Steve Sharpe**

University of Washington

Breakout panel chair, Accelerator Simulation, **Panagiotis Spentzouris**

Fermi National Accelerator Laboratory

Breakout panel co-chair, Experimental Particle Physics, **Frank Wuerthwein**

University of California, San Diego

Breakout panel co-chair, Experimental Particle Physics, **Jim Shank**

Boston University

Representative, Office of High Energy Physics, ***John Kogut***

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

EXECUTIVE SUMMARY

A workshop titled "Scientific Challenges for Understanding the Quantum Universe" was held December 9-11, 2008, at the Kavli Institute for Particle Astrophysics and Cosmology at the Stanford Linear Accelerator Center-National Accelerator Laboratory. The primary purpose of the meeting was to examine how computing at the extreme scale can contribute to meeting forefront scientific challenges in particle physics, particle astrophysics and cosmology. The workshop was organized around five research areas with associated panels. Three of these, "High Energy Theoretical Physics," "Accelerator Simulation," and "Experimental Particle Physics," addressed research of the U.S. Department of Energy (DOE) Office of High Energy Physics' Energy and Intensity Frontiers, while the "Cosmology and Astrophysics Simulation" and "Astrophysics Data Handling, Archiving, and Mining" panels were associated with the Cosmic Frontier.

The scientific grand challenges for this field are described in several publications, including the High Energy Physics Advisory Panel's (HEPAP) Quantum Universe report and the P5 subcommittee's Long Range Plan for Particle Physics in the United States. From these, the following fundamental questions are posed:

- ***What are the main features of physics Beyond the Standard Model?*** These can be sought by exploring physics at high energy using the Large Hadron Collider (LHC), by seeking small departures from predictions made assuming Standard Model physics for interactions at lower energy, or by exploring physics with neutrinos.
- ***What is the identity of dark matter and how does large scale structure develop in the expanding universe?*** The prime dark matter candidate is a supersymmetric particle, but less massive particles called axions may also be involved. The primordial fluctuations out of which contemporary large-scale structure grew may provide a window on fundamental processes occurring during the epoch of inflation.
- ***What are the empirical properties of dark energy and what do they imply about its nature?*** Present evidence involving observations of supernova explosions and the growth of perturbations in the expanding universe is consistent with it having the properties of Einstein's cosmological constant. Departures from this behavior may signify the presence of new physics operating over cosmological scales.
- ***What are the reasons for and implications of neutrino masses?*** The measured properties of neutrinos provide an existence proof for important physics Beyond the Standard Model (BSM) but may also lead to an explanation of why there is a net preponderance of matter over anti-matter. Astrophysical arguments also contribute to the answer.
- ***Are there extra dimensions beyond the three spatial and single temporal dimensions familiar from everyday experience?*** These may have tiny length scales as suggested by string theory or may have macroscopic scales that can lead to a weakening of the law of gravity.
- ***How do Nature's particle accelerators operate in extreme environments?*** Ultra high energy cosmic rays are accelerated by cosmic sources possibly associated with massive black holes to energies as high as 1 ZeV. When these cosmic rays hit the Earth's atmosphere, they do so with center of momentum energies in excess of 100 TeV.

Discovering the answers to these questions requires an effort of historic proportions. Many of the world's best scientists, using research instruments of unprecedented scale and complexity, are vigorously pursuing these matters.

Critical to this research is massive computation. Spectacular research tools such as the LHC in Europe, Fermilab's Tevatron, or the recently launched Fermi Gamma Ray Space Telescope produce data in such quantity and of such complexity, that the fastest computers are essential to filter, analyze, and store petabytes of data. Our understanding of the sub-atomic world now permits first-principles calculations that test current theories and point the way to new ones. These calculations use the fastest petaflop computers available. While we can observe cataclysmic events such as a supernova explosion or the collapse of a massive star into a black hole from a vast distance, these involve phenomena that can never be reproduced in a terrestrial laboratory. However, it becomes increasingly possible to recreate these conditions in a digital universe where important questions can be asked and alternatives explored.

The planning and design of the International Linear Collider (ILC) is critical for the next step in high energy physics. How can we be certain that this expensive machine will work when it is turned on? Instead of building a series of increasingly costly prototypes, accelerator physicists turn with increasing sophistication to detailed numerical models. These provide proof of principle and also allow them to optimize a cost-effective design where both a series of important small efficiencies and dramatic new approaches can be explored and validated.

Thus, large-scale computation provides the framework in which research into the mysteries of the quantum universe takes place. It enables the present successes in the field and represents some of the most important challenges that must be overcome if that research is to advance. As was repeatedly demonstrated during this workshop, the requirements from computing are indeed extreme both in their difficulty and in the transformational opportunities which they offer. Some of the most important of these opportunities include:

- Compute the elements of the famous Cabbibo-Kobayashi-Maskawa (CKM) matrix with accuracies approaching 1% or better. This can reveal new physics BSM and will provide key tests of the theory of quantum chromodynamics, analogous to the development of atomic physics and quantum electrodynamics in the 20th century.
- Analyze the 100 petabyte datasets anticipated from future dark matter/dark energy optical surveys in order to reduce the level of systematic error. This will require intensive modeling of the Point Spread Function and the atmosphere. It must be followed by intensive analysis of the derived galaxy clustering datasets in order to derive cosmological parameters which can be used to describe the properties of dark matter and energy.
- Perform *ab initio* calculations of Type Ia and Type II supernova explosions to provide absolute calibration of the luminosity and to understand the constraints that are placed upon neutrino physics.
- Optimize the design of a high energy lepton collider linear particle accelerator (linac) module for cost and risk reduction.
- Perform multi-scale and multi-physics beam dynamics simulations for next generation Intensity Frontier accelerators.

- Create the seamless digital fabric required to filter, aggregate, store, retrieve, and analyze the exabytes of data that will be produced in LHC experiments and must be understood by collaborations of thousands of physicists distributed across the globe.

These are among the 20 priority research directions requiring extreme computing which are analyzed in the following sections of this report.

As the workshop and the discussion to follow clearly establish, meeting these computational challenges and realizing the enormous scientific advances which are possible requires three essential ingredients. The most obvious is providing the computational capability. Enormous aggregate processing speed, vast data storage and data manipulation capacity as well as high bandwidth links between the research instruments, computers, data storage and scientists must be provided. No single machine will be optimized for tackling all of these and a diversity of new capabilities will be needed. Exaflop "capability" computers are critical to exploiting present theoretical understanding of the Standard Model and teasing out the places where it fails. Such machines are also essential to the simulation of supernova collapse, or understanding the formation of large scale cosmological structure. However, enormous capacity, provided by many smaller and highly cost-effective computers, is needed to provide the event-by-event analysis which dominates accelerator-based experiments. Different architectures again are needed to analyze and synthesize astrophysical data. Providing such computing resources is a daunting challenge that is made much more difficult by the need to operate machines in an energy–efficient and sustainable manner. It lies outside our charge to discuss how this might happen, but we note the U.S. Department of Energy (DOE) is in a unique position to assess the prospect of "green" computing operating at the extreme scale. Given current technologies and practices, more effort will be required to meet the challenges of power, cooling, and space. Power consumption has steadily increased over the past several years, and faster central processing units (CPUs), larger memory chips, and smaller disk drives also continue to increase power demands. Approximately half of the power consumed by a data center is required for cooling. Innovative solutions must be addressed or extreme scale computing will not be realized.

The second essential ingredient is the new software paradigms, numerical algorithms, and programming tools needed to take advantage of extreme scale computer architectures. Even more difficult than designing and fabricating a computer chip with hundreds of cores each with a multitude of threads and joining hundreds of thousands of such chips together into a well integrated computer is actually using this multitude of resources efficiently to do forefront science. Such hardware advances will be of little use for most of the challenges described here without enormous advances in software. Questions of fault tolerance for both hardware and software and of establishing the correctness of code which embodies extreme application and system complexity must be solved.

The third and most important challenge is creating the highly trained cadre of scientists who will understand the challenging science and master the new level of software and system sophistication inherent in exascale computing. Of course, such research must be done by teams of specialists representing both application and computer science expertise. However, to take full advantage and possibly even to successfully exploit such extreme scale resources we will need a large and highly talented group of individuals whose expertise bridges the science of the application and the software/hardware environment. It is these individuals who will recognize the truly revolutionary opportunities for science that should be expected from the advance to exascale computing.

To develop strategies to create these exascale computational resources, to develop the critical algorithmic and software advances and to meet the essential level of scientific/computation training and personnel support is the objective of other workshops. However, our study leaves no doubt as to their importance.

The DOE has traditionally played a highly visible and important role in the development of advanced computing and its application to the frontiers of science. From the provision of cutting edge, high-performance computers, exemplified by DOE’s current leadership class machines, to the very successful algorithm/application development and training enabled by the Scientific Discovery through Advanced Computing (SciDAC) program, DOE is nurturing the computational infrastructure on which an advance to exascale computing must be based. Given the critical importance of computation in the future science mission of DOE as laid out in the following chapters, the case is very strong that this focus on computing is essential and that DOE’s efforts in this direction must be substantially intensified. Such investment in extreme computing will enable revolutionary advances in our understanding of the quantum universe.

PRIORITY RESEARCH DIRECTIONS

Cosmology and Astrophysics Simulation

- Conduct cosmic structure formation probes of dark universe
- Understand and calibrate supernovae as probes of dark energy
- Examine particle accelerators and gravity in extreme environments
- Exploit cosmic probes of neutrino physics

Accelerator Simulation

- Design and optimize a lepton collider linac module for cost and risk reduction
- Predict beam loss and resulting activation in intensity frontier accelerators and maximize performance of energy frontier accelerators
- Shorten design and build cycle of accelerator structure
- Develop and design a compact plasma-based collider
- Develop the techniques necessary to design muon-based accelerators

Astrophysics Data Handling, Archiving, and Mining

- Analyze cosmic microwave background temperature and polarization data to probe the epoch of inflation
- Analyze astrophysics data from extreme environments such as black holes and relativistic jets
- Analyze baryonic acoustic oscillations to study dark energy
- Analyze galaxy clusters and large-scale structure to study dark matter
- Analyze weak gravitational lensing to study dark energy

High Energy Theoretical Physics

- Search for physics BSM by accurate calculation of weak decay amplitudes
- Test quantum chromodynamics (QCD) at the sub-percent level
- Simulate possible theories of physics BSM

Experimental Particle Physics

- Understand the origin of mass for elementary particles
- Explore unification of forces
- Understand the nature of dark matter
- Examine the properties of neutrinos and ultra-rare processes
- Search for extra dimensions, micro-black holes, and other speculative new phenomena

TABLE OF CONTENTS

INTRODUCTION

THE SCIENCE

The relationship between high energy physics (HEP) and high-performance computing (HPC) is long and close. The goals of the U.S. Department of Energy (DOE) Office of High Energy Physics, namely understanding the fundamental character of the constituents of nature and the interactions between them, involves accelerators and telescopes of extraordinary sophistication and complexity to test theories whose equations defy simple approximation, let alone exact solution. Both aspects of this endeavor have historically been intensely computational, driving hardware and software, most famously with the World Wide Web, which derives from the research of particle physicists working at the European Organization for Nuclear Research (CERN).

Particle physics is widely believed to be on the threshold of making the most exciting discoveries in over 20 years. A beautiful theory, now called the "Standard Model," was developed to account for most features of the strong, weak, and electromagnetic interactions of elementary particles in terms of six quarks, six leptons, and the forces between them which are mediated by eight gluons, three gauge bosons, and the photon. To date all of these particles have been detected and every prediction of this model that has been tested experimentally has been successful with the conspicuous exception of the failure, so far, to discover the Higgs Boson. This omission is expected to be rectified using experiments to be conducted at the Large Hadron Collider (LHC) at CERN. This facility, operating at a higher energy than the Tevatron at Fermilab, is also anticipated to provide the confident, experimental observation of physics Beyond the Standard Model (BSM) and may verify the existence of a completely new family of supersymmetric elementary particles. If this happens, it is proposed to push forward with an electron-positron linear collider. (The discovery that neutrinos have finite masses may already be a demonstration that the Standard Model is incomplete and there is considerable optimism that corresponding discoveries will be made at the LHC using proton collisions). An alternative approach to this new physics is to study behaviors of neutrinos using much more intense neutrino beams than currently available and to seek tiny deviations from Standard Model predictions at low energies using either much more intense beams or much larger collision rates than were possible.

Even within the confines of the Standard Model, one of the great challenges is to solve the complex equations of quantum chromodynamics (QCD) to explain how fundamental quarks and gluons combine to make particles like protons and to compute properties of these particles such as their masses. The approach that is followed is to solve a discretised version of the problem on a lattice in spacetime, an approach known as lattice QCD (LQCD). The solution should become more accurate and new computational "measurements" become possible as the spacing of the lattice points decreases, which requires ever faster processing speed.

There is also a Standard Model of cosmology of which the principle components are dark matter, which is suspected to be composed of "supersymmetric" particles, and dark energy, which may be a new classical field like electromagnetism, normal, baryonic matter, photons, and neutrinos. The underlying equations are just as intractable analytically and detailed numerical simulations are necessary to compare observational data with theory and to uncover, for example, the empirical properties of dark energy and learn about neutrino masses. It is also possible to use observations of discrete cosmic sources, especially black holes, neutron stars, and supernovae, to prosecute independent investigations of fundamental physics. Simulating cosmological and astrophysical phenomena is mostly executed on highly interconnected clusters on the ~10 teraflops scale but these simulations do not now have the resolution and reach to match upcoming telescopes.

Turning to data, existing accelerators, such as the Tevatron at Fermilab and the B factory at the Stanford Linear Accelerator Center (SLAC), are near the cutting edge in terms of data acquisition and pipeline processing (several petaflops). The LHC data storage and mining challenge is even greater (more than 20 petabytes per year) and will grow further as the LHC is upgraded over the next decade. Astronomical datasets have typically lagged particle datasets by several years but appear to be catching up. The Large Synoptic Survey Telescope (LSST), for example, with its 3 GPx camera, is projected to archive and make publicly available several hundred petabytes of data. A future international radio telescope, the Square Kilometer Array (SKA), promises a pipeline processing challenge that far exceeds any conceivable capability and strong compromises will be necessary. In fact, the telescope design will be heavily influenced by the ability to process the data.

Finally, turning to the future, the next generation of particle accelerators, like the International Linear Collider (ILC) and new technology alternatives such as wakefield accelerators rely upon vast numerical simulations for their design. The fundamental physical principles, mainly those of electromagnetic theory, are very well understood and extremely complex configurations can be optimized with confidence, given adequate computing power. The astrophysical view of telescope design, epitomized by LSST, is interestingly to regard the influence of propagation across the universe, through the Earth's atmosphere and within the telescope itself as part of the same light path which must be simulated end to end in order to extract cosmological measurements. The challenge here is also considerable.

To date, the computational capacity has barely been able to keep up with the experimental and theoretical research programs. There is considerable evidence that the gap between scientific aspiration and the availability of computing resource is now widening and the case for this is examined in detail in what follows. What is clear on the basis of current usage is that the future needs are unlikely to be met by a single high-performance machine or even several installations of the same machine. The demands vary considerably across fields.

Another conclusion that can be drawn on the basis of existing practice is that it will not be possible to achieve Exaflops computing simply by scaling up contemporary clusters. The power and cooling requirements will be prohibitive—about 20 GW or roughly 10% of national electrical power usage today—for a single machine which will be prohibitive in a "green" future. Radical innovation in processors, memory, and algorithms will be necessary to get to the exascale.

There are many differences in the social organization of the HEP subfields. Experimental particle physics operates under a distributed analysis model that requires the archive to be mirrored at multiple sites. This generates large input/output (I/O) challenges and considerable network traffic, currently up to 10 $\mathrm{GB \cdot s^{-1}}$. This is already a large fraction of the DOE Office of Advanced Scientific Computing Research's (ASCR's) ESNet dataflow. The associated software effort has to be highly managed as hundreds of developers can be involved with a single utility. By contrast, the practice in handling astronomical data typically involves taking the process to the data. Likewise, much of the science in cosmological simulations emerges from heavy graphical display of large data volumes, a step that is essentially unnecessary in LQCD. As is discussed further below there are likely to be major challenges to the existing social organization of all subfields and, probably, much to be learned from and contributed to quite separate disciplines, such as climate change, which are grappling with similar challenges. All of this will, in turn, require a major revision of the training in computer science that high energy physicists are likely to need in future.

WORKSHOP STRUCTURE AND REPORT PREPARATION

In light of the scientific opportunities and technological challenges described above, DOE's Office of High Energy Physics in partnership with the ASCR held a workshop on scientific challenges for understanding the quantum universe and the role of computing at the extreme scale, December 9-11, 2008, at SLAC in Menlo Park, California.

The key goals of this workshop were to engage the national and international scientific leaders in identifying the crucial global scientific challenges and to provide them with an opportunity to shape future scientific computing at extreme scales. In support of these goals, this workshop:

- identified how computing at extreme scale could assist in overcoming forefront scientific challenges for understanding the quantum universe by the end of the next decade
- provided the quantum universe research community with an opportunity to identify the appropriate role for scientific computing at extreme scale in its quest to advance the frontiers
- provided the quantum universe research community with an opportunity to contribute to the ASCR's planning for acquiring computational resources at exascale
- provided the HPC community with an opportunity to understand the potential future needs of the high energy physics research community.

At the workshop, participants identified the scientific challenges facing the field of high energy physics and outlined the research directions of highest priority that should be pursued to meet these challenges. Representatives from the national and international high energy physics community as well as representatives from the HPC community attended the workshop. This group represented a broad mix of expertise. Of the 98 participants, 10 were from international institutions.

All attendees were welcomed by Walt Polansky of DOE's Office of Advanced Scientific Computing Research and John Kogut of DOE's Office of High Energy Physics. Workshop chair Roger Blandford of SLAC presented the charge, expectations, and logistics. The workshop agenda and the names of the plenary speakers are presented in Appendix 1. The workshop participants are listed in Appendix 2.

The activity of the workshop was directed toward five central areas in high energy physics where extreme scale computing plays a critical role. These five areas were:

- Cosmology and Astrophysics Simulation
- Astrophysics Data Handling, Archiving, and Mining
- Accelerator Simulation
- High Energy Theoretical Physics
- Experimental Particle Physics.

For each of these areas a panel chair was selected. The group of panel and workshop chairs/co-chairs then identified and recruited leading scientists both from the United States and abroad, highly qualified to address the questions posed above. In advance of the workshop, each of the five panels prepared a white paper, which provided a starting place and focus for the workshop discussions. Through a combination of plenary and breakout sessions, these white papers were refined into preliminary panel reports which were then further polished after the workshop ended into the five reports that make up the body of this document.

Research on each of these questions is already intensely computational. DOE resources are already used to provide essential access to petaflops machines and the mining of petabyte datasets. Perhaps because of this well-developed ability to exploit HPC, it was easy to recognize that the current research program in each of these areas is dramatically computation-limited. In each area there are crucial problems associated with the priority research directions that can only be answered by moving to computing at the extreme scale. Examples of transformational questions in each of these areas whose answers require extreme scale capability include:

- Can the hypothesis that weakly interacting massive particles are a major component of dark matter be ruled out by present and future gamma ray observations?
- Can future observations of extreme astrophysical environments reveal the character of strong field general relativity where the most revolutionary aspects of Einstein's theory become dominant?
- Can practical plasma or wakefield accelerators be built to deliver high intensity particle beams? Is it possible to begin with the existing proofs of concept for these radically new technologies and to design a robust and affordable machine that can deliver the required collision rate?
- Do presently unrecognized violations of the Standard Model exist in present or future intensity frontier experiments and hold the clues for new physics BSM, physics that may involve an energy scale so high that its direct discovery must wait for decades?
- What is the origin of the masses of the quarks and leptons? Do these masses require the new form of matter represented by the Higgs boson or point to never-before-seen supersymmetry or a new scale of strong interactions, mirroring the nuclear force only a million times stronger?

These five research areas represent highly promising directions for inquiry that when pursed together will yield answers to the frontier questions of high energy physics. They encompass a great diversity of computational methods, hardware and software requirements, and targeted, cost-effective computation platforms. However, there is also much in common beyond their cross-cutting scientific goals. Both areas of astrophysics data management and high energy experiment must begin to collect and manipulate data sets on the exascale. Accelerator physics, high energy theory and cosmological simulation can exploit the largest capability machines that can be created. All five areas must incorporate and exploit the substantial future advances in computer and computational science required to effectively use exascale resources to advance science. These needs cover a wide swath from the global networking technologies needed to support the large collaborations and geographically disbursed groups that must have unfettered access to the data of the next astrophysics observatories and the largest particle accelerators to the advanced coding paradigms and algorithms needed for high energy theory, accelerator design or cosmological simulation to exploit computers with a million active concurrent threads. Both the required advances in computational science and the training in their use must be a critical part of DOE's mission.

The five panel reports lay out the present scientific environment and explain the critical role played by HPC in on-going high energy physics research. They each then identify a series of priority research directions central to the scientific goals of high energy physics in which exascale computation promises transformational advances. The result is a compelling case to redouble our efforts to reach this threshold of computational capability.

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

PANEL REPORTS

COSMOLOGY AND ASTROPHYSICS SIMULATION

ACCELERATOR SIMULATION

ASTROPHYSICS DATA HANDLING, ARCHIVING, AND MINING

HIGH ENERGY THEORETICAL PHYSICS

EXPERIMENTAL PARTICLE PHYSICS

COSMOLOGY AND ASTROPHYSICS SIMULATION

Chair: Michael Norman, University of California, San Diego

1. INTRODUCTION

The universe is the ultimate high energy physics experiment. Astronomers, cosmologists, and astrophysicists are just beginning to decode its secrets. A universe dominated by dark matter on galaxy and galaxy cluster scales, and one by dark energy on the largest scales, are two of the most important and remarkable discoveries to emerge in recent decades. Dark energy also controls the fate of the universe. The nature of dark energy has been called the most fundamental problem in physics today. The recent discovery that neutrinos have mass by observing the universe has also been revolutionary. It is our first taste of physics Beyond the Standard Model (BSM). The controversy as to whether black holes really exist has been laid to rest by increasingly sophisticated X-ray and gamma ray observations from space. It is now legitimate to ask: Are the black holes we observe as predicted by Einstein's theory of gravity? How do they produce the universe's highest energy particles and photons?

The U.S. Department of Energy (DOE) has recognized the importance and relevance of these cosmic discoveries to the High Energy Physics program by introducing the Cosmic Frontier as a new and complementary thrust to be pursued, along with the traditional High Energy Frontier and High Intensity Frontier of accelerator-based science. Just as computational science has been vital to advance the accelerator based programs on both the experimental and theoretical sides, so too is it the case in cosmological and astrophysical research.

Numerical simulations are now the dominant tool in theoretical astrophysics and cosmology, where they play three important roles.

- Simulation by itself makes fundamental advances in basic science. The best simulations are the most realistic physical descriptions of the known universe.
- Simulations are the beneficiaries of new techniques in computation and data analysis, and they generate large "clean" data sets for use in developing new machine-learning tools.
- Simulations play a key role in mission optimization. Practicing on the outputs of simulations aids in mission design and matching simulations to observations validates the models while enhancing the scientific return from observations.

This last role is of particular relevance to the Quantum Universe program, where significant investments are planned for dark energy and dark matter research. In addition, it has been noteworthy over the past five years that simulations have been increasingly used in the direct analysis of observational data.

The panel on astrophysics and cosmology simulations has identified four scientific grand challenges of direct relevance to the Quantum Universe program for which extreme scale computing will be transformational. In priority order, they are:

- cosmic structure formation probes of the dark universe
- understanding and calibrating supernovae as probes of dark energy
- particle accelerators in extreme environments
- neutrinos and the universe.

2. BASIC SCIENCE CHALLENGES, OPPORTUNITIES, AND RESEARCH NEEDS

2.1. *Cosmic Structure Formation Probes of the Dark Universe*

Observations of cosmic structure are the primary means of investigating foundational questions relating to the dark sector. Structure formation probes of dark energy, possible modifications of general relativity, and the dark matter distribution on large scales are being rapidly developed, principally weak gravitational lensing, baryon acoustic oscillations, and galaxy clusters. These and other large-scale probes such as the Lyman-alpha forest, the spatial distribution of galaxies and 21cm (the hydrogen spin flip line) surveys provide ways to measure the primordial fluctuation power spectrum, yielding valuable information on inflation and neutrino masses. Tracking the formation of structure on smaller scales, down to the galaxy and the solar system, is essential for direct and indirect dark matter detection experiments. The cosmic structure probes are described below.

2.1.1 Baryon Acoustic Oscillations

The cosmological structures observed today arise from tiny density perturbations in the primordial fireball. In this very early phase, the universe was a hot plasma; acoustic oscillations arose from the competition between the attractive force of gravity and gas pressure. These oscillations left their imprint on structures at every epoch of the evolution of the universe. They provided a robust standard ruler from which the expansion history of the universe can be inferred and gravity on very large scales can be probed.

2.1.2 Weak Lensing

Dark matter, by its very nature, can only be astrophysically detected through its gravitational action. Since gravity feels “all” of the matter, luminous and dark, lensing provides a clean geometrical probe of the distribution of both luminous and dark matter in the cosmic web. The amplitude of the lensing signal depends on geometric “efficiency” factors and the amplitude of mass perturbations in the universe. This allows, in principle, a measurement of both geometry and structure growth. The two are related in Einstein’s theory, thus lensing provides simultaneous dark energy phenomenology and testing of general relativity.

2.1.3 Redshift Space Distortions

Our chief measure of distance is the redshifting of light resulting from the Universal expansion, but the redshift has an additional component from any motion of the object with respect to the universal reference frame. Since such motions are generated by gravitational potentials, redshift space distortions can probe the dark sector and test our theory of gravity in much the same way as lensing. In fact, redshift space distortions and lensing are complementary in that they employ non-relativistic and relativistic tracers (respectively), thereby probing perturbations to the metric in different ways.

2.1.4 Galaxy Clusters

The distribution and abundance of galaxy clusters at different epochs are strong probes of dark energy. They measure the growth of structure on much smaller scales than baryon oscillation measurements, and therefore provide complementary information. Ongoing and future surveys will provide rich catalogs of clusters. Several hurdles must be overcome to achieve the stringent scientific goals. The main task is to connect observations to simulations and to answer the following questions:

- How can the mass of a simulated dark matter halo be connected to observations in the optical, millimeter-wave, X-ray, and infrared?
- What is the importance of feedback effects?
- How well can we predict the mass function of clusters – exponentially sensitive to dark?

2.1.5 Lyman-alpha and 21cm

Most of the baryons in the universe are not in galaxies, but in intergalactic gas. This gas, observable in hydrogen Lyman alpha absorption after reionization and by hydrogen 21cm emission/absorption prior to reionization, traces the dark matter distribution, and thus provides a complementary probe to galaxies. The Lyman alpha absorption is well observed and modeled, and has been used to constrain cosmological parameters. The Baryon Oscillation Spectroscopic Survey (BOSS) survey of the Sloan Digital Sky Survey (SDSS)-3 project proposes to use both Lyman alpha absorption and galaxies for baryon acoustic oscillation (BAO) measurements. The 21cm observations must await the completion of new low frequency radio arrays, but the potential is enormous to trace the growth of structure in the redshift interval between recombination and reionization across a wide range of angular scales. Theoretical interpretation will require radiation hydrodynamical cosmological simulations spanning a commensurately wide range of physical scales.

2.2. *Key Role of Simulations*

Many of these observations are associated primarily with very large volume sky surveys, which source very large static as well as time-domain multispectral databases. Development of theoretical approaches that can make predictions for—and analyze and interpret—the data stream is the fundamental challenge for the next decade. On a similar timescale, direct and indirect dark matter search experiments—terrestrial and in space—require extreme scale simulations with sufficient space and mass resolution to make definitive predictions of the likely distribution of dark matter substructure in the Milky Way. Modeling requirements cover scales ranging from the entire observable universe down to the structure of galaxy clusters, individual galaxies, the solar system, and individual stars. The fundamental challenge is to perform simulations including gravity and hydrodynamics with the dynamic range necessary to design and interpret future experiments.

Depending on the observational technique being modeled, the underlying complex physical processes must be simulated with a high degree of fidelity and numerical accuracy. These include the three-dimensional, coupled gravitational evolution of dark matter and baryons while the latter is subjected to complex (magneto-) hydrodynamic, thermodynamic, chemical and radiative processes. An emerging concept is to use full-physics mesoscale simulations of galaxy and galaxy cluster evolution to develop semi-analytic models that would be applied to global dark matter-only simulations. Alternatively, with extreme scale computing, one could attempt full physics global scale simulations with subgrid models derived from local scale simulations. The complexity and richness of the information contained in the observational signatures of large-scale structure, combined with extreme accuracy requirements, have created a critical need for extreme scale modeling and analysis tools.

2.3. *Scientific Impact*

The scientific impact of discoveries in this research area is profound. It includes the determination of the equation of state of dark energy, distinguishing between dark energy and modifications of general relativity, measuring the masses and interactions of dark matter, measuring the sum of the neutrino masses, and probing the fields responsible for primordial fluctuations.

The current constraints on the dark energy equation of state are consistent with a cosmological constant at roughly 10% accuracy. Explanations of dark energy can be distinguished by their predictions for the time evolution of the dark energy equation of state. This will be probed only by next-generation observations ([BOSS], Dark Energy Survey [DES], Joint Dark Energy Mission [JDEM], Large Synoptic Survey Telescope [LSST], etc.) that target percent level accuracies for determining cosmological parameters. These observations will also target alternative explanations for the cosmic acceleration, such as modifications of general relativity on cosmological scales.

Cosmological surveys currently provide the tightest constraints on the sum of the neutrino masses complementing oscillation experiments that measure neutrino mass differences. The two measured mass differences ensure that at least one neutrino has mass greater than 0.05 eV. Current upper limits are within a factor of 10 of this, so—with the aid of simulations—upcoming surveys may well measure a non-zero signal.

The hypothesis that weakly interacting massive particles constitute the dark matter predicts that the current dark matter abundance is determined by its annihilation rate. When the abundance is fixed by current measurements, a fairly robust prediction is that the annihilation cross-section times velocity is of order 10^{-26} cm^3/sec. Gamma ray detectors (Fermi Gamma-ray Space telescope, Very Energetic Radiation Imaging Telescope Array System [VERITAS]), would detect many photons from dark matter particles in the center of our galaxy annihilating with this cross-section. Resolving astrophysical caveats is essential to this argument, but these can be addressed with high-resolution simulations. Certainly this cross section represents a target for upcoming indirect detection experiments.

The clustered mass distribution observed in the universe today is a consequence of the amplification of primordial density perturbations by the gravitational instability. These primordial fluctuations inform on high energy processes in the early universe, inaccessible to accelerators. Measurements of the fluctuation power spectrum amplitude, its slope (characterized by the spectral index), the running of the spectral index with wave number, and the possibility of detection of B-modes all provide crucial input to our understanding of the early universe and high energy physics.

2.4. *State-of-the-Art: N-Body*

On large-scales, gravity dominates the evolution of the mass distribution in the universe. Consequently, N-body codes form the backbone of all cosmological simulations. The current state of the art covers two types of applications:

- cosmological simulations of large volumes, from hundreds of Mpc^3 to tens of Gpc^3
- high-resolution simulations at the individual galaxy scale (Figure 1), needed to resolve the details of the mass distribution. The largest cosmological simulations run to date have employed approximately 70 billion particles in tens of Gpc^3 volumes with roughly 100 kpc force resolution.

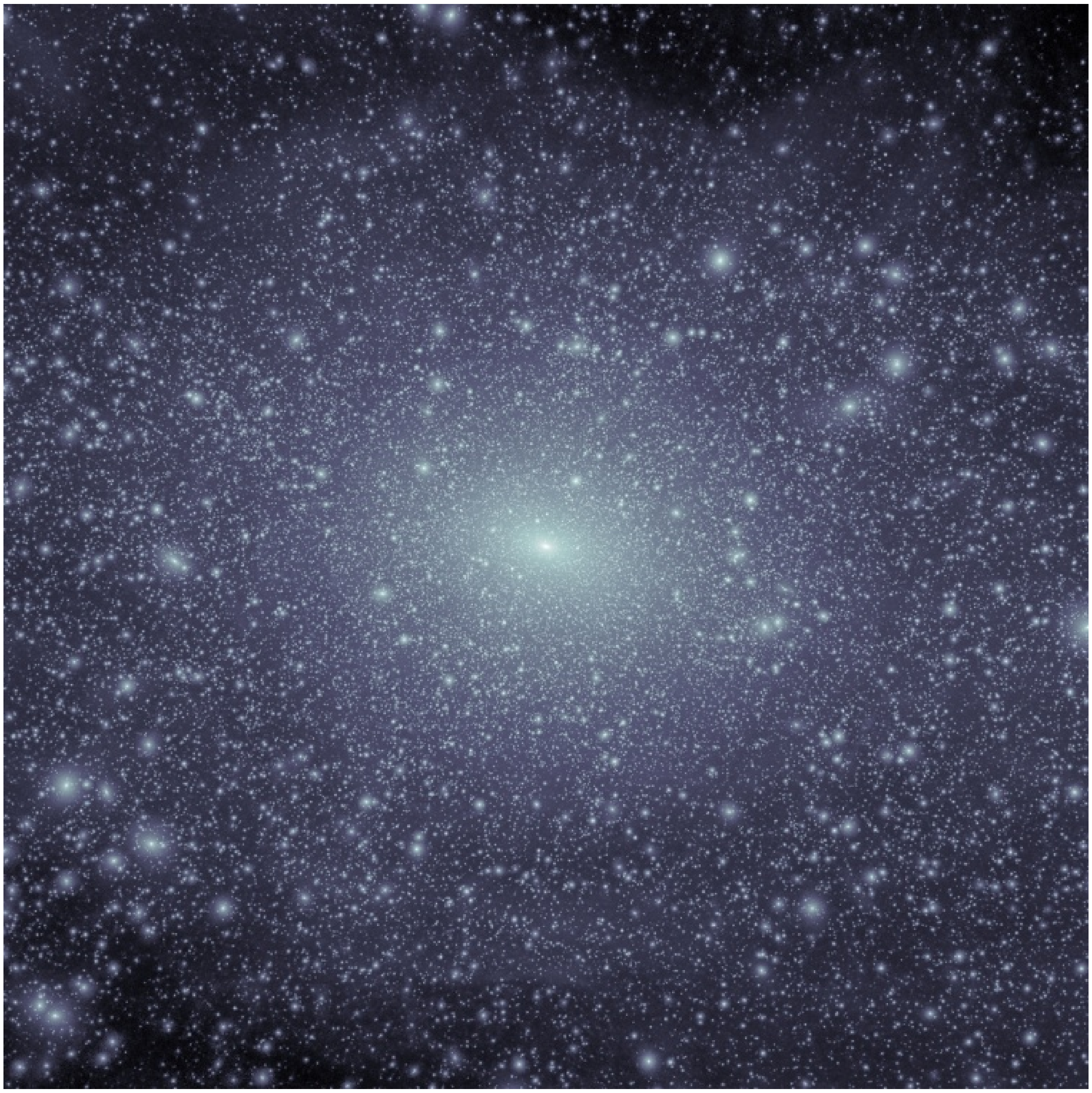


Figure 1. Results of a 1.1 billion particle N-body simulation carried out on DOE resources of the clustering of dark matter on the scale of the Milky Way galaxy. Such simulations predict that our galaxy is embedded in a dark matter halo with thousands of dark matter subhalos orbiting within the solar circle. If sufficiently dense, they could be detected by gamma ray emissions arising through dark matter particle annihilations (Diemand et al. 2008). Image courtesy of Michael Norman (University of California, San Diego).

2.5. *State-of-the-Art: Hydrodynamics and Radiative Transfer*

On smaller scales, non-gravitational forces become increasingly important and tracking the gas distribution becomes crucial for weak lensing, clusters, 21cm surveys, the Lyman-alpha forest, individual galaxies and stars. This requires the use of hydrodynamic simulations augmented by "local" physics, and including gravity, via coupling to an N-body solver. Adaptive mesh refinement (AMR) and smooth particle hydrodynamics (SPH) are the two most commonly used methods. The state of the art in cosmological hydrodynamics is represented by AMR simulations with billions of simulation cells (root grid) and up to seven levels of refinement, and by 10 billion-particle SPH simulations. Radiative transfer simulations are by comparison less well developed, with quasi-self-consistent simulations approaching two orders of magnitude in linear dynamic range.

2.6. *Understanding and Calibrating Supernovae as Probes of Dark Energy*

Supernovae of both types can be employed as standard candles in cosmology, and the earliest evidence for an accelerating expansion of the universe came from an analysis of Type Ia supernova light curves. In the next decade, major observatories and space missions plan to use supernovae to obtain distance estimates of unprecedented precision. Understanding how supernovae work is thus not only an inherently interesting question, but also one whose answer is essential if we are to better understand the possible systematic biases in their appearance at different times and in different places in the universe.

Type Ia supernova research stands poised on the threshold of a breakthrough that could be accomplished by one additional order of magnitude in resolution. Realistic calculations must be three-dimensional and treat the interaction of turbulence and nuclear burning carefully, but current simulations have barely broached the turbulent regime. That is they show evidence of a turbulent power spectrum in the vicinity of the burning that is not far above the grid scale. Consequently, the effective numerical Reynolds number is low, approximately 1000. With one additional order of magnitude in resolution, the integral scale for the turbulence could be well resolved and the effective Reynolds number increased by a factor of 100. This will reduce the dependence of the results on currently controversial subgrid models for the turbulent burning. The same computational power will also enable calculations of three-dimensional radiation transport to get the light curves and spectra of the models at all angles. Given a library of such models spanning the range of conditions that cosmic evolution might provide (metallicity, rotation rate, ignition density, carbon-oxygen ratio, etc.), robust combinations of luminosity indicators can be brought to bear that would increase the accuracy of distance determinations.

2.7. *Current Status of the Problem*

The "Type Ia Supernova Problem" is really four problems that, computationally at least, are separable, but with the outcome of each stage serving as the conditions for the next.

2.7.1 Ignition

Beginnings are important, especially for supernovae. An exploding carbon-oxygen white dwarf ignited at a single point slightly off center will be barely disrupted if that ignition is not followed by detonation, but off-center ignition can also set the stage for very powerful, bright explosions if detonation occurs later. The observed properties of the supernova are especially sensitive to how much burning occurs during the explosion and the density where that burning occurs. Current simulations show a century of convection prior to the final runaway and suggest a dipole pattern for the convective flow in the star's final moments. In this flow pattern, the hottest points that will eventually seed the explosion occur in a jet-like outflow from the center. Such a geometry leads naturally to lop-sized off-center ignition. However, it is known that the nature of the pre-ignition flow depends on the rotation rate of the white dwarf and it is suspected that it depends on the Reynolds number.

2.7.2 Burning

Once ignited, carbon and oxygen burn to elements in the iron group, raising the temperature and lowering the density of the ash. Since the hot ash lies beneath cold dense fuel, the flame front is Rayleigh-Taylor instability. Plumes of rising ash create shear and turbulence that cascades down to smaller scales, affecting the motion and burning rate of the fuel. Without such instabilities, the burning would be far too slow to give the violent explosion that is observed. Thus, the study of Type Ia supernovae is an exercise in turbulent combustion. The models that give the best current agreement with observations are of the "delayed detonation" variety (Khohklov 1991). The burning proceeds for a time as a subsonic flame, then makes a transition to a supersonic front. The energy and brightness of the brighter Type Ia explosions strongly suggest that detonation happens, but the physics of the transition remains poorly understood.

2.7.3 Transition to Detonation

One way a transition might happen is in colliding waves of matter rolling around the white dwarf as the burning breaks through the surface—the Chicago greatest common divisor model (Plewa et al. 2004). However, for a long time it has been suspected that detonation might happen in a more robust way, when a critical mass of fuel and ash mix as the flame enters the "distributed burning regime." In the distributed burning regime, which happens in a Type Ia supernova at a density of around 10^7 g·cm^{-3}, the laminar flame becomes thick enough to be disrupted by the turbulence that is naturally present. Hot ash and cold fuel can co-mingle, producing a mixture that is potentially explosive. This long sought transition to detonation has no direct analog in terrestrial combustion because the degree of turbulence needed would simply put out a flame on the Earth. In the supernova, with its large size and long time scale, new physics comes into play.

2.7.4 Light Curves and Spectra

Given a model that has reached a homologously coasting state, its spectrum and multi-band photometry can be calculated as a function of time (Figure 2). The radiation transport calculation is challenging because the supernova is inherently three dimensional and the light curve and spectra are known, from current simulations, to depend upon the angle at which it is observed. Additionally, the millions of atomic levels of over 100 relevant ions do not, in general, have populations given by the Saha ionization equation, so a huge array of rate equations must be solved.

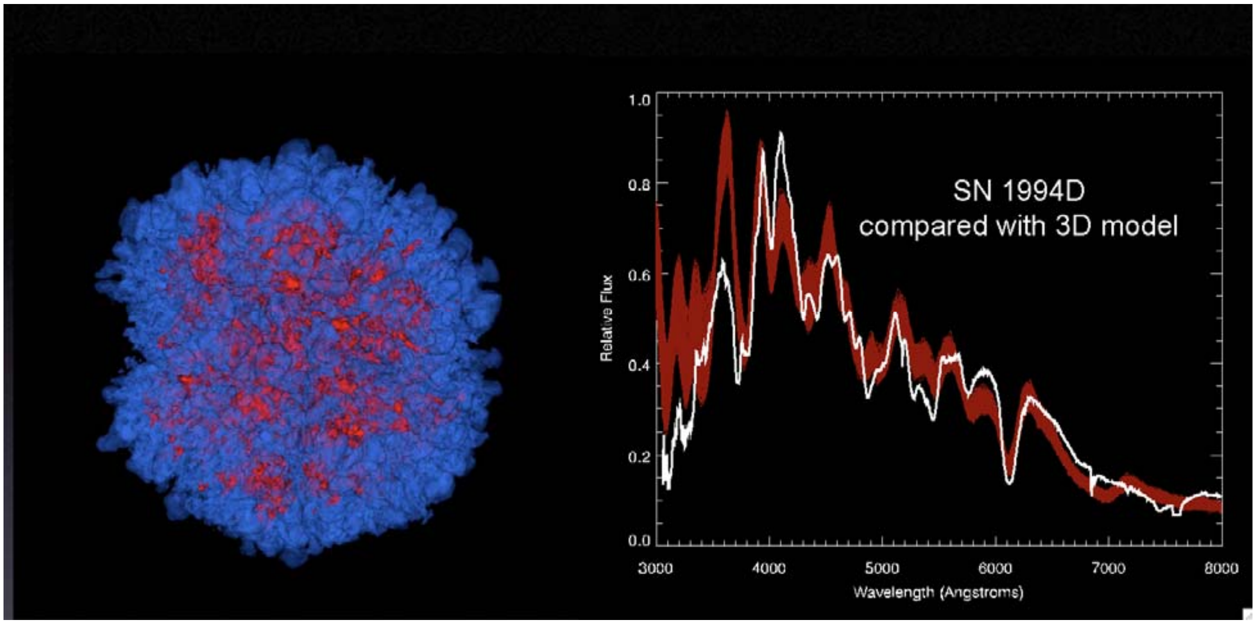


Figure 2. Left: Three-dimensional simulation of a carbon thermonuclear deflagation supernova (Type Ia). Right: Overlay of the simulated optical spectrum with an observed supernova of the sort used for cosmic acceleration measurements. The simulated spectrum required a three-dimensional solution of the radiative transfer equation coupled to a large ionization network (Kasen et al. 2008). Image courtesy of Stan Woosley (University of California, Santa Cruz).

2.8. *Potential Scientific Impact*

Understanding how a Type Ia supernova works is classic problem in astrophysics that has been studied computationally for at least 35 years. In recent years, there has been a convergence upon a “Standard Model” of a carbon-oxygen white dwarf of near the Chandrasekhar mass, but broad agreement on the details of ignition, flame propagation and detonation initiation is lacking. Finding a solution to this long standing problem in astrophysics will be a major accomplishment for simulation science.

Type Ia supernovae, along with other indicators, will also be used in the near future to study the expansion of the universe with unprecedented precision. A major concern in such applications is whether such supernovae might have different characteristics at moderate redshift and if those differences could mask subtle effects of expansion. Statistical errors can be reduced by observing a large number of supernovae, but systematic errors can only be controlled by having a sample that incorporates all the possible effects of evolution. Given a credible and sufficiently detailed model for Type Ia supernovae, one could explore its variation with physical input parameters such as metallicity, rotation, ignition density, and carbon-oxygen ratio. New and redundant indicators of brightness can be determined that, hopefully, will be robust when evolution is included.

Type Ia supernovae are also a laboratory for studying turbulent combustion under extreme conditions of turbulent energy and length scale not achievable on Earth.

2.9. *Particle Accelerators and Gravity in Extreme Environments*

There are two fundamental sources of energy in astrophysical objects, nuclear reactions and gravitational collapse. In its most extreme form, namely accretion onto a black hole, gravitational collapse can release 10 to 100 times more energy per unit mass than nuclear reactions. In this sense, gravity power is the most efficient form of energy production in the universe. Cataclysmic events such as core collapse and black hole merger may be the most luminous events in the universe. In many black hole systems, a significant fraction of that energy comes out in the form of highly relativistic particles. In fact, the Pierre Auger Observatory has recently provided evidence that the highest energy cosmic rays are accelerated in the environments of accreting massive black holes at the centers of galaxies (Pierre Auger Collaboration 2008).

Black holes are among the most remarkable predictions of general relativity because so much mass is compressed into such a small volume that nothing, not even light, can escape. The point of no return is known as the event horizon, the “radius” of a black hole. General relativity has successfully passed many tests of its validity, but this strong field limit has never been probed directly. Unambiguous observations of black holes would provide a validation of general relativity in its most extreme form and would complement studies of the expansion history of the universe that test the predictions of general relativity on cosmological scales.

Black holes occur in two forms. One for is stellar mass black holes that are believed to result from the core-collapse supernova explosion of the most massive stars. One form of gamma ray burst is believed to be just such an explosion. The second black hole form has masses equal to millions or billions times that of the sun. These holes are found in the cores of almost all galaxies, including the Milky Way galaxy. Their origin is uncertain, but it is likely that they play an important role in galaxy formation itself, and the subsequent evolution of the galaxy’s environment. In about 1% of galaxies in the local universe, these black holes manifest themselves a strong central radiation sources known as active galactic. These sources produce emission across a wide range of photon energies (from radio to X-rays) and are usually quite variable in their output. They are often associated with bidirectional outflows of relativistic particles, or *jets* (Figure 3).

Black holes can be detected from the light emitted by material as it falls into the hole or via the radiation produced when particles expelled from the vicinity of the black hole in jets collide with the surrounding material. Black holes can also produce gravitational radiation if they occur in binary systems: the orbit decays, ultimately leading to the merger of the system. Predicting the wave form of the resulting gravitational radiation requires direct solution of the dynamic space-time geometry itself by solving Einstein’s equations.

The cleanest tests of general relativity in the strong-field regime are likely to be produced by comparing theoretically predicted gravitational wave forms for neutron-star neutron star, neutron star-black hole, or black hole-black hole mergers with observations. The Laser Interferometer Gravitational-Wave Observatory, already operating at its initial design specifications, is sensitive to neutron star-neutron star mergers and neutron star-black hole mergers, while the space-based Laser Interferometer Space Antenna (LISA) mission in the coming decade will study the mergers of approximately one million solar mass black holes at cosmological distances (Hughes 2003). In the past five years, there has been a revolution in the implementation of numerical general relativity: numerical simulations are now able to self-consistently solve for the mergers of two neutron stars or black holes, and the collapse to the remnant black hole. As a precision test of general relativity, matched filtering of gravitational wave data with computer-generated waveform templates can help rule out alternative theories of gravity such as those proposed to explain dark matter or dark energy. Matched filtering is sensitive to relative phasing, and modifications to general relativity can significantly change the phase evolution of gravitational waves.

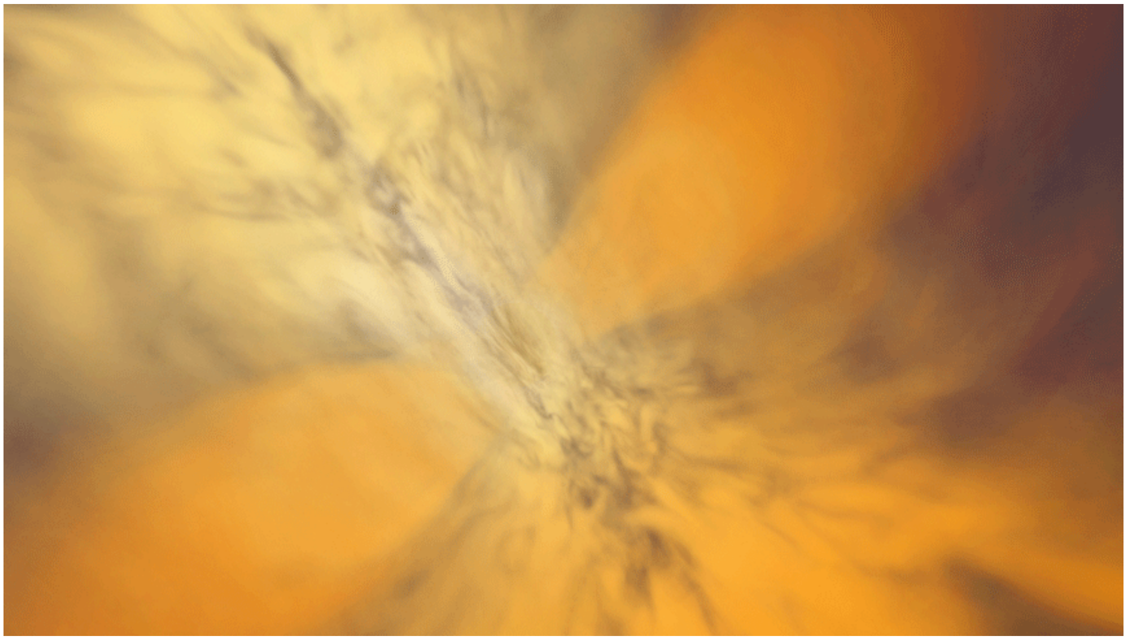


Figure 3. Rendering from a three-dimensional general relativistic magnetohydrodynamics (MHD) simulation of matter accretion and bipolar ejection in the environment of a rotating black hole. (Hawley and Krolik 2006). Image courtesy of John Hawley (University of Virginia).

In several situations, it is crucial to couple calculations of the time-dependent dynamics of plasma in the vicinity of a neutron star/black hole to calculations of the dynamic space-time during the mergers of two neutron stars or black holes. In particular, observations of electromagnetic counterparts to gravitational wave sources will greatly increase the science that can be accomplished with gravitational wave detections. To cite one example of particular relevance to high energy physics, if the redshift of the host galaxy containing merging black holes detected by LISA can be determined via an electromagnetic counterpart, gravitational wave sources can be used for precision studies of the expansion history of the universe and the properties of dark energy.

Electromagnetic observations of accreting black holes can also provide unique probes of general relativity in the strong-field regime. Most of the electromagnetic radiation is produced in the vicinity of the horizon, often by highly relativistic (and highly suprathermal) particles accelerated *in situ*. Understanding this particle acceleration is thus a key to the interpretation of the observations. Because the emission is produced so close to the horizon, the spectrum of radiation is sensitive to the spin of the black hole in addition to its mass. Since astrophysical black holes effectively have no charge, measurements of spin and mass uniquely characterize black holes according to general relativity.

An even more stringent probe of a black hole's space-time will be provided by high resolution images of matter near the horizon: such observations are possible using interferometry from the ground in the radio and infrared in the next decade, and in the X-rays from space on a longer timescale. Interpreting such observations will require high-fidelity simulations of plasma in the neighborhood of the black hole's horizon, together with calculations of the acceleration of particles, the emission of electromagnetic radiation by such particles, and the propagation of radiation through the hole's curved space-time. In the past five years, dramatic progress has been made in simulating the dynamics of plasma in the vicinity of a black hole's horizon, but the observational implications of are only now beginning to be investigated.

Predicting the observations of photons, particles, and gravitational radiation requires understanding the complex behavior of matter and space-time in the environment near the black hole. This is a complex multi-dimensional, time-dependent, relativistic magnetoplasma problem that will be fully accessible only through extreme scale computing.

The primary science goal is to develop models of the black hole environment that are sufficiently detailed and complete so as to provide a predictive theory for comparison with observations. These observations will include high energy photons, cosmic rays and gravitational radiation. This will test general relativity theory in the strong field limit, provide direct evidence for the existence and properties of black holes, elucidate the processes behind the highest energy photons and cosmic rays, and elucidate the properties of high energy discrete events such as hypernova and mergers that are visible across the observable universe and could be used as cosmological probes. The primary resource drivers required for this ambitious program of simulations include the wide range of spatial- and temporal-length scales and the complexity of the physics of the black hole environment, e.g., gravitational radiation and recoil, relativistic particle acceleration, magnetoplasma turbulence, electromagnetic radiative processes and radiation transfer, and photon tracing through the general relativistic space-time.

2.10. *Neutrinos and the Universe*

There is a direct, tangible, and symbiotic connection between key front-line issues in experimental and theoretical neutrino physics on one hand, and the potential fruits of an extreme scale computational astrophysics effort on the other. Extreme scale computing could provide a qualitative leap in understanding of the genesis and growth of large scale structures in the universe and of core collapse supernova physics, which in turn could provide us with insights into fundamental neutrino physics which are either unobtainable in terrestrial laboratories or complement those gained from ongoing and future experiments. Likewise, new insights into neutrino physics obtained in the laboratory will leverage our computational efforts to understand supernovae and cosmology.

Neutrino rest masses and vacuum mixing properties represent physics BSM, and the significant investment by the DOE in theoretical and experimental work has given us the neutrino mass-squared differences and two of the three vacuum mixing angles. This has been a signal achievement, but there are particle physics unknowns that are even more interesting: the third mixing angle θ_{13} and the CP-violating phase, as well as the absolute neutrino masses themselves and the neutrino mass hierarchy. Much of the ongoing and future experimental effort is directed at getting at some of these numbers, because they are key to basic issues in elementary particle physics, such as the origin of mass.

Tantalizingly, astrophysics provides a complementary window on this physics. And the advent of extreme scale computing could usher in a new regime in both particle physics and the subjects in astrophysics that overlap with it. Some examples are in order and described in the following paragraphs.

The universe has always been and will continue to be a powerful laboratory for fundamental neutrino physics. Observations of solar and atmospheric neutrinos, for example, led to the theory-altering discovery that neutrinos have mass, which in turn changed neutrino physics forever and demanded a rethinking of the Standard Model of elementary particles that until now described well the basic building blocks of our world. Currently, our best constraints on the absolute rest masses of neutrinos come from cosmology (the cosmic background radiation plus the matter power spectrum at large wave number). However, it must be said that these limits are dependent on many assumptions in cosmology. Extreme scale simulations of cosmological evolution hold out the promise of beating down the uncertainties associated with these limits. Likewise, core collapse supernovae may be the best way to probe neutrino mixing and neutrino-neutrino interaction physics. The stakes for particle physics are high.

Neutrinos play a surprisingly important role in the cosmos. This may seem unlikely given these particles have no electric charge and influence matter principally through the aptly named weak interaction. However, it turns out that neutrinos can more than make up for their feeble interactions with large numbers. In environments like core collapse supernovae, neutrinos can carry off up to 10% of the rest mass of the neutron star remnant. A typical core collapse event will emit some 10^{58} neutrinos in a time span of seconds. Neutrino fluxes in the supernova environment can be so large the local number density of neutrinos can greatly exceed the electron number density. It is therefore not surprising that neutrinos and their interactions are at the root of almost every aspect of core collapse supernova phenomena.

These events may be the ultimate neutrino physics laboratories. The early universe and gravitational collapse environments are the only venues in nature that we know of where neutrino-neutrino scattering can be significant. Astrophysicists have made a great deal of progress in the past decade in modeling these environments. The neutrino signatures from core collapse supernovae have been a particular focus of interest for astrophysicists, in part because of the exciting results coming from experimental and observational neutrino physics.

In fact, in the last decade there has been a revolution in experimental and observational neutrino physics. We now know that neutrinos have small, but non-vanishing rest masses and that the three “flavors” of neutrinos (electron, muon, and tau) can be transformed among themselves. For example, a neutrino leaving the surface of the neutron star as a tau neutrino could convert into an electron neutrino. This flavor transformation process can be important in core collapse supernovae because, for example, the neutrino interactions that allow neutrinos to drive these explosions and to set the neutron-to-proton ratio in neutrino-heated supernova ejecta that are critically important to supernova heavy element synthesis are flavor-dependent. In many stages of the supernova, the fluxes and/or energy spectra of the various neutrino flavors will be different. Therefore, neutrino flavor transformation could become important for the origin of the supernova explosion itself, the synthesis of the heaviest nuclei in the supernova, and certainly for the neutrino signal we might hope to measure on earth.

2.11. *The Grand Challenge Problem*

Until recently, researchers have not been able to calculate how this neutrino flavor conversion process would work in supernovae, in part because of the nonlinear self-coupling that neutrinos experience as a consequence of neutrino-neutrino forward scattering contributions to the neutrino effective mass. Worse, in the regime above the neutron star where neutrinos and antineutrinos propagate coherently, the flavor histories of neutrinos on different, but intersecting, world lines can be quantum mechanically coupled. New large-scale numerical simulations for the first time have explored this problem (Duan et al. 2006a, 2006b, and 2007; see also Raffelt and Smirnov 2007, Esteban-Pretel 2007, and Fogli et al. 2007). The results are surprising and in many ways completely unexpected.

Neutrino observations and experiments have given us the mass-squared differences between the neutrinos. These are small and, coupled with the conventional Mikheyev-Smirnov-Wolfenstein treatment of neutrino flavor transformation in matter, they seem to suggest there should be little flavor transformation deep in the supernova envelope. The numerical simulations mentioned above, which include the nonlinear neutrino self coupling, find the opposite: large-scale flavor conversion will occur relatively near the neutron star despite the small neutrino mass-squared differences. This alters dramatically our picture of neutrino emission in core collapse supernovae and likely will be important for models of the supernova explosion mechanism and nucleosynthesis, but especially for the neutrino signal. It is the neutrino signal that, if we could measure it for a Galactic supernova event, could tell us about neutrino properties in ways that are complementary to or even transcend planned or existing terrestrial experiments (e.g., accelerator-based long baseline neutrino oscillation experiments, reactor neutrino disappearance experiments, and neutrino-less double beta decay experiments).

Specifically, the nonlinear collective flavor transformation modes in the supernova revealed by the numerical calculations mentioned above could provide signals for key neutrino physics parameters that might never be measurable in a terrestrial laboratory. These parameters include, for example, the unknown vacuum mixing angle θ_{13} and the neutrino mass hierarchy (i.e., whether the solar mass-squared doublet lies above or below in mass the atmospheric neutrino mass-squared doublet).

The supercomputer simulations performed have uncovered a dramatic feature of nonlinear neutrino self-coupling in the region above the proto-neutron star (Figure 4). For the normal mass hierarchy, nearly all neutrinos below a “cut-off” energy have their flavors transformed, while the neutrinos above this energy remain in the flavor states in which they were emitted at the neutron star surface. For the inverted neutrino mass hierarchy, exactly the opposite occurs. This “spectral swap” could give a clear signature of the neutrino mass hierarchy if we could detect a neutrino burst from a core collapse supernova event.

The simulations also show that in the normal neutrino mass hierarchy the swap cut-off energy decreases with decreasing θ_{13}. Thus, measuring the swap cut-off energy could in principle provide a measurement of θ_{13}. This might be complementary to lab-determined values of this quantity. However, it is important to note that if $\theta_{13} < 0.01$ it is not likely that lab measurements will be able to determine it at all. In this case, supernova signal-based techniques, like the swap, might still have a chance and be our only hope of determining this important parameter. On the other hand, in the inverted neutrino mass hierarchy the above mentioned supercomputer simulations suggest that the swap phenomenon is quite insensitive to θ_{13}. Nonetheless, we should still be able to observe a swap signal in this case, and thereby determine the neutrino mass hierarchy, although not θ_{13}, even if θ_{13} is extremely small. By contrast, terrestrial experiments will only be able to resolve the neutrino mass hierarchy if θ_{13} is relatively large.

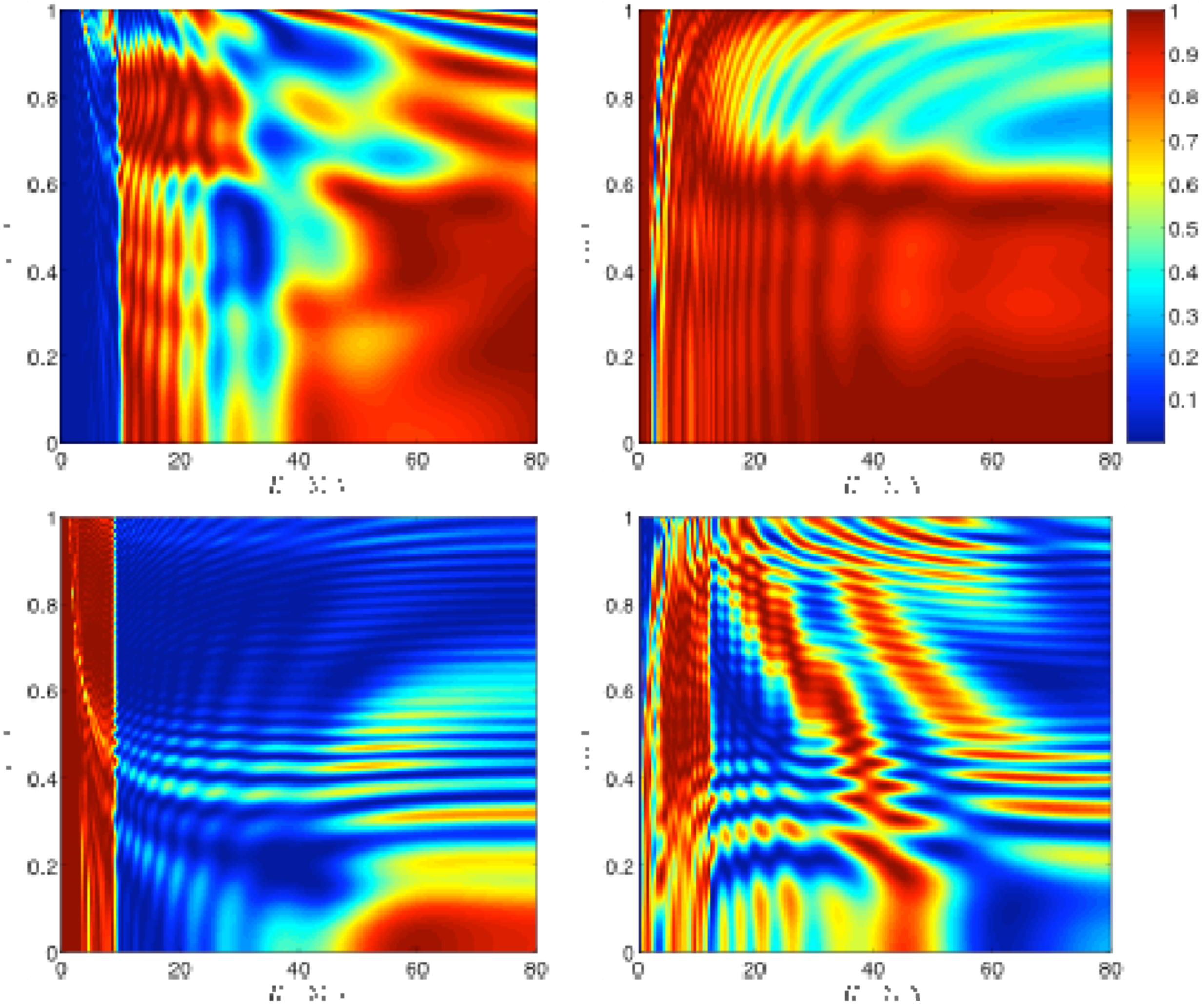


Figure 4. Shown are the results of large-scale numerical simulations of neutrino flavor transformation in supernovae. Color-coded (scale on upper right) survival probabilities for neutrinos are shown on the right-hand panels, for antineutrinos in the left-hand panels. Red indicates no flavor transformation, blue complete flavor conversion. The upper panels are for the normal neutrino mass hierarchy; the lower panels show results for the inverted hierarchy. The survival probabilities are given as functions of neutrino or antineutrino energy (horizontal axes) and propagation direction. (Vertical axes show the cosine of the propagation direction angle relative to the normal to the neutron-star surface at the emission point: e.g., $\cos\theta_0 = 1$ corresponds to a radially propagating neutrino; $\cos\theta_0 = 0$ to a tangentially propagating one.) Image courtesy of H. Duan et al. (2006b).

The pioneering supercomputer simulations mentioned above that led to the discovery of new modes of neutrino flavor transformation and opened up the door to new possibilities of using core collapse supernovae as probes of neutrino physics BSM were performed in spherical symmetry and in the coherent limit only. Three-dimensional, multiphysics simulations of core collapse supernovae with complete quantum kinetics that will self-consistently determine the turbulent environment within which the neutrinos are produced and propagate. They will also determine whether both the classical (collisional) and quantum mechanical (coherent) aspects of neutrino flavor evolution within this environment will need to be performed before the dream of using core collapse supernovae as cosmic laboratories can be realized.

3. GAP ANALYSIS OF TECHNICAL CAPABILITIES TO MEET SCIENCE CHALLENGES

3.1. *Cosmic Structure Formation Probes of the Dark Universe*

Baryon acoustic oscillation surveys target 0.1% accuracy or better for the power spectrum to scales k~0.3 h/Mpc. To achieve this, one must have

- sufficient mass resolution to resolve halos in which galaxies reside
- sufficient volume to reduce sample variance errors
- enough information to study and model galaxy bias. Multi-Gpc boxes are needed with mass resolutions of order 10^{10} solar masses (i.e., a total number of approximately 300 billion simulation particles).

These gravity-only simulations (albeit with a dynamic range of roughly a million) require petascale resources. The first such simulation will be carried out in 2009 as a part of Los Alamos National Laboratory's (LANL's) Roadrunner Universe project.

More generally, the ultimate N-body science goal is enough resolution to resolve sub-halos within a Hubble volume. Additionally, the key difficulty lies not only in carrying out individual simulations, but in implementing a suite of simulations that includes

- a large number of realizations for each cosmology to understand errors and their covariances
- a large number of runs over a set of cosmologies, required for MCMC (Markov Chain Monte Carlo) analysis. This is certainly an exascale enterprise. To resolve subhalos in a Hubble volume demands between 10^{13}-10^{14} particles, in turn requiring 10^{24} flop of computation. To resolve the dark matter halos hosting typical galaxies over a Hubble volume is a petascale project; using these simulations in Monte Carlo fashion translates to an exascale project.

Weak lensing requires sub-percent/percent modeling accuracy up to k~10 h/Mpc. At the required accuracy level, baryonic effects are known to be important beyond k~0.1-1 h/Mpc, and must be modeled. This can be done with AMR and SPH hydrodynamic cosmological simulations; petascale resources will be applied here in the near future. Modeling various systematics (intrinsic alignments, photo-z) is a bigger concern and will require observational input. The ultimate situation is not clear; these are nonlinear systems with feedback so the ultimate effect of changing parameters can only be determined *a posteriori*. The addition of hydrodynamics increases the run-time by an order of magnitude and requires many more exploratory runs to understand the effects of parameter settings in complicated sub-grid models.

In the case of galaxy clusters, the basic mass and dark matter distribution information is straightforward to obtain from N-body simulations. Accurate modeling of actual observations is much more difficult. Issues such as mass-observable relations, survey completeness, cluster morphology, magnetic fields, gasdynamics/turbulence, cool cores, cooling flows, cluster mergers, etc., all enter at various levels. Methods that combine observations with theory/phenomenology will be essential to reduce systematic errors. Hydrodynamic simulations of the integrated Sunyaev-Zeldovich effect in clusters have shown that it is less susceptible to astrophysical effects than other measures (Motl et al. 2005) and is less stringent on numerical resolution requirements relative to other observables. Single Hubble volume-scale simulations are feasible on petascale platforms. Again, using these simulations in a Monte Carlo fashion translates to an exascale project.

A full solution of the 21cm problem involves complex radiative transfer and the ability to resolve "mini-halos" in cosmologically relevant volumes. Current approaches are computation-limited, restricted to either every small volumes or relying on drastic simplifications. The simulation challenge is definitely petascale and could be significantly bigger. As a rough approximation, simulations including one radiation group are 2 to 4 times as expensive as a hydrodynamic run. These in turn are 5 to10 times as expensive as a pure N-body run. Using only 10 frequency groups to approximate the spectral energy distributions of the high redshift source population brings a Hubble volume simulation of the high redshift (greater than10) universe to 100 petaflops.

3.2. *Improving Predictions for Indirect Detection of Dark Matter in the Galaxy*

Multi-billion particle N-body simulations (e.g., Diemand et al. 2008) of Galactic dark matter substructure have been recently performed in order to describe in as much as possible detail the expected formation and accretion history of galaxy-sized halos, their inner density profiles, and especially the properties and survival of stripped subhalo cores as well as tidal debris orbiting within these systems. The main goal of these projects is to substantially refine the predictions that the Lambda Cold Dark Matter cosmology makes for direct and indirect dark matter detection and about the detailed properties of dark-matter dominated Milky Way's dwarf satellites. While producing many tens of terabytes of data and using million of hours of central processing unit (CPU) time, these simulations are not anywhere near resolving Galactic dark matter substructure on parsecs scales. The central high density regions of dark matter subhalos are artificially smoothed, and the simulation fails to accurately describe the clumpiness of the dark matter distribution in our galactic neighborhood. The largest numerical simulations use millions of hours of CPU time, yet have problems writing and analyzing even tens of terabytes of output. There is a huge loss of temporal information when one can only store a hundred snapshots of a simulation involving hundreds of thousands of timesteps.

The gravitational forces in region of the Milky Way enclosed in the sun's orbit has been dominated by baryons in the form of gas and stars over most of cosmic time. This is true for most all the galaxies that host most of the stars in the universe. Even the smallest dwarf galaxies found to be dominated by dark matter in their most central regions have most likely been gravitationally dominated by gas for short periods of time in the past. This simple fact that the baryons provide the dominant gravitational forces to the centers of galaxies has profound implications in our ability to reliably compute and predict the distribution of dark matter on the scales relevant for all our experiments that aim at direct as well as indirect detection. Whereas the latest billion particle simulations of dark matter halos clearly seem to converge it is also clear now that the biggest uncertainties in the predicted dark matter distributions form simulations comes from our inadequate treatment of baryonic physics.

Allowing for primordial gas chemistry, cooling physics, and hydrodynamics, cosmological simulations have now reached a level where they can follow the formation of individual protostars that are some billion times smaller than region from which their constituent gas is collected (Abel et al. 2002). At the same time, using three-dimensional radiation transport techniques have allowed to follow the buildup of early dwarf galaxies at unprecedented resolutions (Wise and Abel 2007). This chance of making galaxies one star at a time will be feasible for entire Milky Way galaxies using extreme scale computing. The Milky Way contains some 100 billion stars whose phase space positions and associated variables age and heavy element content could be stored using 40 terabytes of memory. About 10 times less in mass is the cold gas available to star formation, which at 100 resolution elements per Jeans mass storing, all magneto hydrodynamic state variables and a minimal chemical composition would take approximately another 80 terabytes of main memory. At any point in time, there are approximately one million massive stars dominating the ultraviolet radiation input and the transfer of radiation from them to the interstellar medium has to be computed. Similarly, hundreds of supernovae remnants will dominate the cosmic ray injection rate at any point in time. Given that any solver will need to keep old and new state variables at least for the gas adding up these requirements, we can reasonably expect to start being able to realistic models of the Milky Way galaxy once 300 terabytes of main memory are available for routine calculations. Here we have not yet addressed whether we would have sufficient flops to carry out this calculation. At the required resolution of one tenth of a parsec, it would take some 20 million time steps at the solar circle to compute the Milky Way over a Hubble time. For a solver with an average of 300 operations per state variable (averaged over hydro, gravity, radiation transfer, and chemistry) this calculation will take three months on a 100 petaflop machine. It would have followed the buildup of the Milky Way galaxy one star at a time from when its first star formed until the present. The dark matter profile in the galactic center, all the relevant dark matter substructure, dark matter spikes around early stellar mass black holes and neutron stars would all be captured down to scales of approximately one tenth of a parsec.

3.3. *Understanding and Calibrating Supernovae as Probes of Dark Energy*

Current calculations of models for Type Ia supernovae are resolution limited, and it is expected that fundamental breakthroughs could occur with grids that were one order of magnitude finer. In a three-dimensional Courant-limited study with adaptive mesh, this implies an increase of about 10^3 in CPU. Since current studies are already using all available resources, this implies that moving from the petascale to the exascale will be necessary to yield a profound improvement.

3.3.1 Ignition

Studies of ignition have a numerical Reynolds number (Re) that is, in the best of circumstances, about 1000, while the value in the supernova is 10^{14}. Clearly this is too large a range to bridge in any foreseeable future, but it is highly desirable to push the simulations at least one order of magnitude further into the turbulent regime, which has barely been broached. The characteristic dipole flow seen in all current three-dimensional studies may have a qualitatively different character even at moderately high Re. The location and extent of ignition "hot spots" is critical to determining the explosion properties and may be quite sensitive to Re.

3.3.2 Flame Propagation

While it is not necessary to resolve the laminar flame (10^{-4} cm -1 cm) in the supernova (10^9 cm) in order to calculate a credible first principles' model, it is necessary to fully resolve the instabilities and turbulence in the vicinity of the burning. All calculations to date require a "turbulent subgrid model", and it is different assumptions about the subgrid burning that make the results of the major groups doing the simulations discrepant. This sensitivity to the subgrid model will diminish when the turbulent cascade is everywhere well resolved on the grid. The integral scale for the turbulence currently calculated to be approximately10 km when the supernova has expanded to 10,000 km. Some advantage is obtained by using adaptive mesh in the vicinity of the flame, but still calculations with approximately10^{12} zones are required to see approximately 1.5 dex of the turbulent cascade. Current limits are approximately10^{9} zones.

3.3.3 Deflagration-Detonation Transition (DDT)

Even with sufficient resolution to follow the average flame speed correctly, full star calculations will, for some time to come, still inadequately resolve the small rare fluctuations that can cause a detonation. In fact, DDT in an unconfined medium remains a poorly understood phenomenon in chemical combustion as well as in astrophysics. Current scaling arguments and one-dimensional simulations suggest that DDT in a supernova happens when the flame enters the “stirred flame” regime in which the individual laminar flamelets are disrupted by turbulence. The scale of the phenomenon is still set by the turbulent integral scale, but it is necessary to achieve a higher numerical Reynolds number within that region in order to see the transient, well-mixed regions of fuel and ash necessary to initiate a detonation. For $Re \sim 10^4$ on a scale of 10 km with an effective Kolmogorov scale set by grid resolution, zoning in excess of 10^{13} is again required. In addition, the effect of turbulence created above the grid scale must be well represented on the grid.

3.3.4 3D Radiation Transport

Because of the instabilities that lie at the heart of its explosion, a Type Ia supernova is necessarily a three-dimensional object lacking any spherical or rotational symmetry. Because its appearance is determined by the energy deposited by radioactive decay diffusing through this asymmetrical structure, the supernova will look different, even with different brightnesses, when viewed at different angles. Three-dimensional local thermodynamic equilibrium (LTE) calculations of the spectrum that assume thermal equilibrium among the ground and excited states of the atoms are just beginning to be done, but apparently they can be completed with petascale resources. However, 0.5 to 1 million CPU hours will be required per spectrum for non-LTE calculations. Including time dependence and doing 100 sample points moves the calculation beyond resources that can be obtained today. It will also be necessary to explore more than 100 models to determine sensitivity to the metallicity, explosion energy, initial composition and density, etc., taking these calculations well into the exascale as well.

3.4. *Particle Accelerators and Gravity in Extreme Environments*

Carrying out realistic simulations of extreme black hole environments has long been a goal. Only recently, however, have improvements in computational power and algorithms enabled the first simulations of fully three-dimensional global accretion disk dynamics (Figure 3), and simulations of a binary black hole spiraling inward over several orbits and then merging.

Current resource requirements can be estimated from the product of the current number of discrete spatial computational elements (e.g., grid zones, particles, finite-elements), times the number of iterations (time steps), times the number of floating point operations necessary to advance from one time step to the next per spatial element. Present simulations typically use 10^{7} zones distributed over three spatial dimensions and are run for limited times (10^{6} time steps). Using only idealized physics (single fluid ideal magnetohydrodynamics [MHD], optically thin cooling, vacuum space-times) requires roughly 10^4 floating point operations per spatial element per timestep. The total number of floats for this sort of simulation is therefore roughly 10^{17} flop. This scale is accessible with current systems, but there are at least two additional factors driving resource demand. First, this is a problem that cannot be addressed through a few highly resolved models; instead we require a substantial ensemble of simulations to cover the wide range of environmental variables expected. Second, while the simulations to date have provided some preliminary explorations, they are inadequate as predictions of observables. For black hole accretion simulations, greater spatial and temporal resolution is required, but, more significantly, detailed thermodynamics, radiative processes, and particle acceleration have yet to be included self-consistently. For dynamical spacetime simulations, there are several factors driving demand for greater resolution. For example, in unequal mass binary systems, while the total mass of the system sets the distance scale to the wave-zone, the mass of the smaller black hole sets the scale of the finest grid spacing; thus resolution scales with mass ratio and large mass ratios are expected to be common. As another example, as the envelope is pushed on the initial separation of the simulated binary black holes, the rate of change of the energy in the early part of the simulation decreases, thus higher resolution is required for sufficiently accurate phase evolution. Dynamical spacetime simulations that include matter (e.g., neutron star collisions) face all of the above challenges.

The overall black hole model consists of sub-problems, each of which is itself an extreme scale computation. For example, the fundamental mechanism driving accretion is the transfer of orbital angular momentum by MHD turbulence driven by MHD instability; the subsequent dissipation of that turbulence accounts for the energy release. This is a much more challenging problem compared to hydrodynamic turbulence. The fluid is driven on many scales and the overall behavior depends on both the viscosity and the resistivity and their ratio (Prandtl number). The properties of this MHD turbulence must be understood to be incorporated as a sub-grid model in global simulations. The photon emission from the disk involves optically thick and thin regimes, Comptonization processes within a relativistic corona, high-Z element fluorescence, and substantial gravitational and Doppler shifts. Dynamical radiative transfer simulations and Monte Carlo modeling will be required for predictions of telescope observations. Another example is particle acceleration within the jet. Global simulations have shown how large-scale magnetic fields can drive powerful Poynting fluxes away from the black hole horizon. The mechanisms by which those fields subsequently accelerate particles to highly relativistic energies are largely unexplored, but they could be investigated by detailed particle-in-cell (PIC) simulations that use the global field configurations as their starting point.

The planned computer power promises to bring these simulations within the realm of possibility as algorithms improve. Together this portends a transformational improvement in modeling allowing the simulations to finally approach what is needed for the predictive modeling required to understand what the observations are telling us about gravity's ultimate systems.

3.5. *Neutrinos and the Universe*

Core collapse supernovae are general relativistic, neutrino radiation, MHD events. The production, transport, and interaction of neutrinos and antineutrinos of all flavors in stellar cores of massive stars play a central role in powering such supernovae. However, precisely how core collapse supernovae are generated remains unknown. Some combination of neutrino heating, aided by hydrodynamic phenomena such as convection, and magnetic field effects produces these explosions, but the precise mix and the many details of core collapse supernova explosion dynamics remain to be determined in the context of three-dimensional, multi-physics simulations that have yet to be performed. Our ability to use core collapse supernovae as laboratories for fundamental neutrino physics will depend on the successful completion of this computational program over the next decade as we march toward the sustained exascale supercomputing platforms that will be required.

To determine how neutrinos contribute to generating this class of supernova explosions, knowledge of their distributions in the two direction cosines that uniquely specify a direction in three spatial dimensions and of their distributions in energy, at each instant of time and three-dimensional spatial location, is required—i.e., core collapse supernovae are inherently seven-dimensional phenomena.

To solve the equations that govern the evolution of the neutrino distributions and the stellar core fluid flow, magnetic fields and self gravity will require, among other things, the solution of large, sparse linear systems of equations distributed over the millions of processors we can expect on exascale supercomputing platforms. Effective preconditioning and solution algorithms for these systems on such a daunting number of processors must be developed, and such algorithms will need to accommodate nontrivial computing and memory hierarchies—e.g., we can expect such platforms to organize the cores (processors) into sockets (nodes), requiring solution algorithms that will take advantage of cross-socket and cross-core parallelism. Other challenges include: 1) mitigating the impact of latency and bandwidth limitations associated with communication between sockets, 2) developing and maintaining good load balance across sockets, 3) developing fault-tolerant algorithms that will sustain a simulation as cores fail (the likelihood of this will increase as platforms move toward millions of cores), and 4) achieving efficient collective parallel I/O from millions of cores. Moreover, each exascale simulation can be expected to produce petabytes of data. Such data will require end-to-end solutions for data management, analysis, and visualization, and the management and automation of the workflows associated with the overall simulation execution.

4. PRIORITY RESEARCH DIRECTIONS

For each of the four grand challenges described above, we identify the following priority research directions (see Table 1 for a summary).

4.1. *Cosmic Structure Formation Probes of the Dark Universe*

The major research direction is the extraction of precise predictions of structure formation in the universe, from the largest scales that can be measured (Hubble volume), down to the scale of the solar system, to individual stars. Multi-physics, multi-scale simulations at extreme dynamic ranges form the essential component of this endeavor. Computation-relevant research and associated infrastructure development includes work on spatially and temporally adaptive codes, algorithms, and workflows for simulations and data on extreme-scale architectures. Three-, six-, and 10-year goals are given in Table 1.

Table 1. Summary of priority research directions

Grand Challenge	3-year goals	6-year goals	10-year goals
Structure formation probes	Ensembles of 10^{12} N-body simulations in Hubble volume single 10^{12} cell/particle AMR radiation hydrodynamic simulations in Hubble volume Workflow tools for analysis and mock surveys	Ensembles of 10^{13} N-body simulations in Hubble volume single 10^{13} cell/particle AMR radiation hydrodynamic simulations in Hubble volume. Convergent halo substructure predictions (N-body)	Simulations guided by dark energy surveys for precision tests Convergent halo substructure predictions (N-body+gas)
Supernova standard candles	Low-resolution 3D calculations of a grid of supernova Ia models with credible flame physics. LTE calculations of 3D spectra and light curves. Library of model SN II light curves and spectra. Assist in mission planning and optimization for DES, JDEM and LSST and PTF (Palomar Transient Factory).	Convergence on flame propagation physics. Continued studies of ignition and detonation. First 3D non-LTE light curves and spectra. Improve scientific yield of JDEM and LSST Improved cosmological measurements for DES 3 year dataset and 5-year PTF dataset.	Fully resolved 3D studies of all aspects of the explosion. Non-LTE, 3D radiation transport in all models. Complete library of models for comparison with data. Improved distances and cosmological parameters from JDEM and LSST.
Extreme environments	Inclusion of particle acceleration, transport, and photon ray tracing in black hole accretion disk simulations	Inclusion of dynamic spacetimes Engage observations	Suites of fully integrated simulations for observational comparison and hypothesis testing.
Neutrino universe	Inclusion of full quantum kinetics in spherically symmetric supernova models. Two-dimensional supernova simulations with Boltzmann (classical) kinetics.	Two-dimensional supernova simulations with full quantum kinetics.	Three-dimensional supernova simulations with full quantum kinetics.

4.2. *Understanding and Calibrating Supernovae as Probes of Dark Energy*

The major research direction is pushing the current three-dimensional simulations of degenerate carbon core ignition, deflagration, and detonation transition to higher spatial resolution, ultimately to a factor of ten relative to current models. This will require scaling the codes to much higher core counts and possibly incorporating adaptive mesh refinement. In addition, three-dimensional radiative transfer codes will need to be ported to extreme scale architectures, which may be highly non-trivial because of the large memory per node requirements of current implementations.

A second research direction is the validation and calibration of future simulation codes. Essentially, the over-reaching goal of the above computational exercises is to push the state of theory and simulations to the point that we are able to directly and correctly predict absolute brightness of Type Ia supernovae as function of other supernova observables. Careful calibration of the models using observations and scaled terrestrial combustion experiments, when feasible and available, will be crucial for developing and validating such supernova codes. A significant part of funding and computational resources should be dedicated to this effort if we have any hope to use theory for removing systematic errors in supernova calibration. Essentially, we must push the theory and simulations must be pushed to be able to directly predict absolute brightness as function of supernova observables and bypassing secondary brightness standards such as Cepheids. The gain will be much more difficult when achieving 1% accuracy.

4.3. *Particle Accelerators and Gravity in Extreme Environments*

The goal is the development of integrated three-dimensional relativistic MHD models of black accretion disks and binary black hole merger events, including all relevant radiative processes, high-energy particle processes, and accurate photon ray tracing in the curved space-time, to make detailed predictions of energetic processes near black holes. The Physics Near the Event Horizon Initiative would support the design, execution, and analysis of an ensemble of codes to address the unique, yet connected physics problems involved in this extreme environment. These would form the building blocks of a multi-scale, multi-physics simulation combining dynamical space-times and kinetic plasma models of particle acceleration, and photon production and transport. Most of these applications share commonalities with other initiatives, e.g., supernova physics. The outcome of this effort would be testing general relativity in the strong field regime, obtaining diagnostics that would reveal the properties of black holes, and provide a theoretical understanding of nature’s most powerful particle accelerator that generates the most energetic photons and particles that we observe.

4.4. *Exploiting Cosmic Probes of Neutrino Physics*

The above-mentioned sustained exascale simulations of core collapse supernovae will be the culmination of a series of increasingly sophisticated simulations on petascale- through exascale supercomputing platforms that will be performed over the near- and longer-term during the next decade, and these simulations stem from groundbreaking three-dimensional simulations that have already been performed. Recent three-dimensional hydrodynamics-only (uni-physics) simulations constructed to reflect the post-stellar core bounce supernova dynamics led to the discovery of the so-called stationary accretion shock instability that has become a central phenomenon in core collapse supernova theory (Blondin and Mezzacappa 2007). The stationary accretion shock instability is now known to play a major role in powering and defining the morphology of core collapse supernova explosions. Ongoing three-dimensional multi-physics simulations on the Oak Ridge National Laboratory (ORNL) Leadership Computing Facility using the CHIMERA code now include, in some cases with significant approximation, all known-to-be-important physics except magnetic fields. In particular, multi-frequency neutrino transport is included in the “ray-by-ray” approximation, which neglects lateral neutrino transport. General relativity is included as a spherical correction to the three-dimensional Newtonian self-gravitational field. While these ongoing simulations are pioneering in that they are the first to include multi-frequency neutrino transport and thus incorporate far greater realism, the approximations made will need to be eliminated and definitive simulations with full three-dimensional, multi-angle and multi-frequency neutrino quantum kinetics must replace them. Table 1 summarizes the timeline toward this end.

5. CROSS-CUTTING RESEARCH DIRECTIONS

Because of their reliance on particle methods, there is significant algorithmic and even possible code overlap between beam dynamics simulations, cosmology structure simulations, and particle acceleration around black holes.

Interpolation over a high-dimensional space (in cosmology, useful for application to Markov chain Monte Carolo (MCMC) simulation results from an ensemble of simulations with varying parameters) is a capability that we can offer the HEP experimentalists.

5.1. *Relevance to HEP*

Cosmological probes of high energy physics are unique; they can discover and measure BSM physics in ways that are inaccessible to conventional methods, such as accelerator experiments. The remarkable discovery of dark energy is a case in point, as is the cosmological constraint on the sum of neutrino masses, more stringent than that from current terrestrial experiments. Measuring the matter fluctuation power spectrum opens a window to test inflation (and alternatives), thereby accessing energy scales far higher than possible with terrestrial accelerators.

DOE-supported projects have led to a sea of change in observational cosmology (SDSS, supernova surveys). The tradition continues with BOSS, DES, and Fermi Gamma-ray Space Telescope among current missions, and with JDEM and LSST to follow. In the near-term, petascale resources are coming on line and these will already be necessary to fully exploit the scientific power of ongoing experiments. Within a decade, however, the qualitative change in data size, complexity, measurement cross-cuts, and modeling accuracy challenges represented by JDEM and LSST will require an extreme-scale computing capability. Planning for this next stage should be initiated now.

The principal programmatic impact of realistic models for Type Ia supernovae will be to assist in planning the next generation of dark energy missions, e.g., DES, JDEM, and LSST, and optimizing their scientific return. The original discovery of dark energy was based upon a simple one-parameter adjustment to the light curve shape—either width or template fitting. More recently, color and infrared-light curves have been used. With a library of model spectra at all wavelengths, angles, and times, these diagnostics can be employed on a data set that is free of reddening and other observational difficulties. New diagnostics can be discovered. The models will help plan what to observe and with what cadence.

Turbulent combustion in Type Ia supernovae occurs under conditions that are either unattainable or attained with great difficulty on the earth. In particular, the large length scales and high degree of turbulence—approaching an appreciable fraction of sonic—are novel. A better understanding of turbulent combustion and the physics of the deflagration-detonation in an unconfined medium will advance our understanding of chemical combustion in general.

The codes developed to study Type Ia supernovae also have broad applicability to other problems. The same codes used to calculate light curves and spectra for one kind of supernova can be applied to others, and the techniques for three-dimensional radiation transport will find broad applicability both in and outside of astrophysics. Low Mach Number solvers can be applied to other problems in tight hydrostatic equilibrium, e.g., convection. The adaptive mesh, compressible codes developed to study turbulent flames can be applied, with appropriate physics, to core-collapse supernovae and other problems.

5.2. *Cross-Cutting Issues*

Current petascale platforms are architecturally ill-suited to the task of massive data analysis. We would not expect this to change as we move to the exascale. The cosmology and astrophysics simulation panel agrees with their vision of a data-intensive supercomputer (DISC), and furthermore suggest that a computer center for extreme scale computational science be designed as shown in Figure 5. It consists of three components:

- a compute-intensive engine (e.g., exascale computer), for running simulations and inline data analysis of these simulations, typically in batch mode
- a data-intensive engine (something with the total memory of an exascale computer, but fewer processor nodes, and higher I/O bandwidth), for analyzing the results of simulations and observational data mining, doing interactive visualization and data reduction
- high performance online storage for integrated science collections (simulation + observation), stored in formats supporting interoperable and joint analysis.

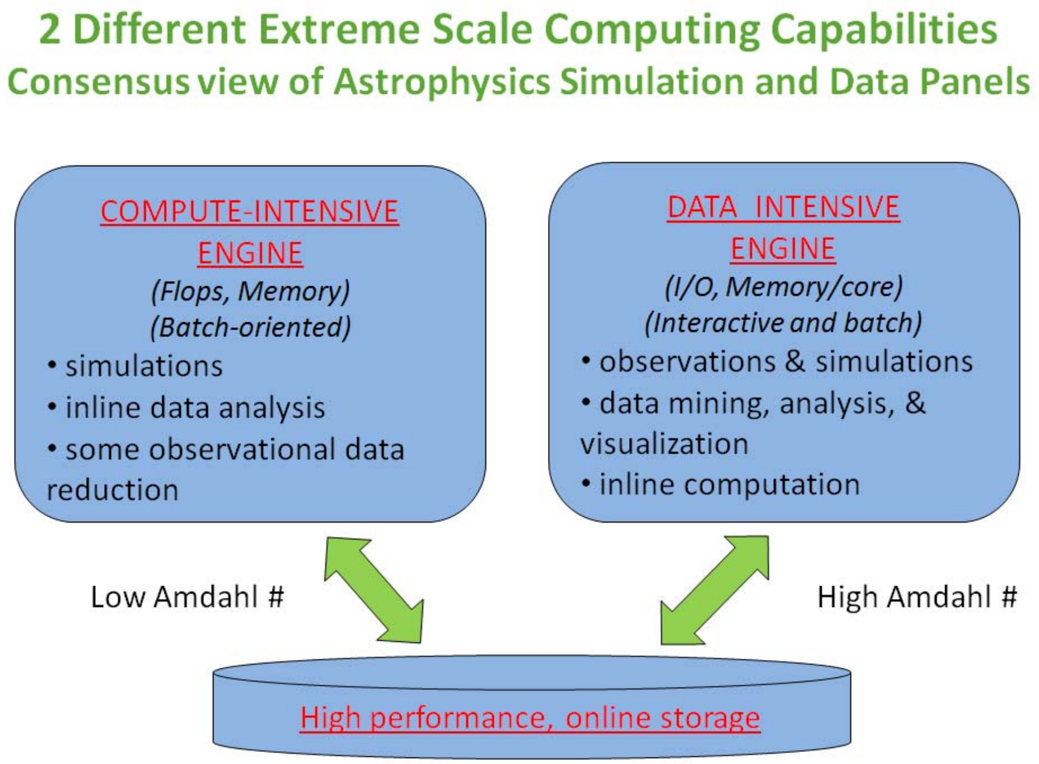


Figure 5. Two different extreme scale computing capabilities' needs for extreme scale astrophysics simulations, simulations data analysis, and observational data analysis.

6. SUMMARY OF TRANSFORMATIVE SIMULATIONS AND EXTREME SCALE COMPUTING REQUIREMENTS

We conclude by highlighting three applications of extreme scale computing in cosmology and astrophysics that would be transformative for their disciplines and fundamental to advancing the cosmic frontier goals of the HEP quantum universe program. We provide an estimate of their resource requirements by scaling from current numerical experiments running on near-petascale platforms.

(i) Perform high accuracy N-body dark matter simulations (augmented by fluid dynamics on small scales) that describe the growth of structure on linear scales smaller than galaxies so as to be able to interpret observations of weak lensing and baryon acoustic oscillations made by DES, JDEM, and LSST and thereby infer the equation of state of dark energy or, equivalently, describe how gravity must be modified from the predictions of the theory of general relativity. This will require being able to process N of order 10^{14} particles in roughly 100 realizations which requires 10^{26} flops. This mandates a highly parallelized machine operating at a level of ~10^{18} flops, RAM memory of ~100 petabytes and disk storage of ~10 exabytes.

(ii) Perform simulations of the type of stellar explosion associated with Type 1a supernovae. This will allow us to understand the dependence of the luminosity of supernovae upon factors like stellar mass, composition, and rotation which in turn will be crucial to calibrating their use as distance indicators and observing the kinematics of the expanding universe using DES, JDEM, and LSST. It will be necessary to solve the three spatial dimension equations of radiative hydrodynamics including flame propagation and this in turn will require a variable mesh with spacing one tenth that of contemporary "petascale" simulations. In other words, this will require an exascale machine similar to (i).

(iii) Perform numerical simulations of coalescing binary neutron stars such as are believed to be responsible for many of the gamma ray bursts observed by Fermi and likely to be observed simultaneously over the next decade in gravitational radiation using the Advanced Laser Inferometer Gravitational-Wave Observatory. Success in this endeavor will provide powerful tests of strong field general relativity and probe the cold equation of state of nuclear matter at super-nuclear densities. The computational requirements for a single simulation with a 10x increase in the number of cells per spatial dimension, number of timesteps, and flop/cell for adding radiation diagnostics or full GR is 10^{22} flop. Running 1000 such models to vary orbital and spin parameters, mass ratios, and other model parameters raises the operation count to 10^{25} flop. While this is an order of magnitude fewer flops than in (i), the memory requirement would be similar to (i) because of the large number of grid functions (>100) that must be evolved in numerical relativity.

ACCELERATOR SIMULATION

Chair: Panagiotis Spentzouris, Fermi National Accelerator Laboratory

1. INTRODUCTION

The U.S. Department of Energy (DOE)'s Office of High Energy Physics (HEP) promotes a broad, long-term particle physics program at three interrelated frontiers of particle physics (DOE/NSF 2008). The office supports current operations and experiments, and research and development for future facilities and experiments in three different frontiers:

- The Energy Frontier directly explores the fundamental constituents and architecture of the universe.
- The Intensity Frontier, accessed with a combination of intense particle beams and highly sensitive detectors, offers a second, unique investigation of fundamental interactions.
- The Cosmic Frontier reveals the nature of dark matter and dark energy by using particles from space to explore new phenomena.

These scientific frontiers form an interlocking framework that addresses fundamental questions about the laws of nature and the cosmos. The accelerator modeling panel will outline how the development of extreme scale computing capabilities can contribute to meeting these grand scientific HEP challenges by advancing accelerator science.

Particle accelerators are critical to scientific discovery in the DOE program in the United States and indeed the world (DOE 2003b). Of the 28 facilities listed in *Facilities for the Future of Science: A Twenty-Year Outlook* (DOE 2003a), 14 involve accelerators. The development and optimization of accelerators are essential for advancing our understanding of the fundamental properties of matter, energy, space, and time directly in two out of the three frontiers supported by the HEP program—the Energy and the Intensity Frontiers, and indirectly in the Cosmic Frontier. Modeling of accelerator components and simulation of beam dynamics are necessary for understanding and optimizing the performance of existing accelerators and for optimizing the design and cost effectiveness of future accelerators.

2. SCIENCE CHALLENGES, OPPORTUNITIES, AND RESEARCH NEEDS

In the next decade, the HEP community will explore the Energy Frontier by operating the Large Hadron Collider (LHC) and designing the LHC accelerator complex upgrades, and by developing novel concepts and technologies necessary for the design of the next lepton collider. The HEP community will also be exploring the Intensity Frontier by designing and possibly operating high intensity proton sources for neutrino physics and precision measurements in the framework of rare process searches, and designing high intensity muon sources for neutrino physics—neutrino factories. There is overlap between the accelerator science developments necessary to explore the two frontiers because high-intensity sources are required for lepton colliders, either an e^+e^- linear collider or a muon collider.

The accelerators and novel accelerator concepts that we will need to study and design include:

- a new, high intensity, versatile, proton source for the LHC

- a high-intensity proton source capable of producing multi-MW power beams at energies relevant for the creation of neutrino beams that could be used for long-baseline neutrino experiments and potentially muon beams for muon colliders and neutrino factories
- a linear electron-positron collider based on superconducting radio-frequency (rf) technology, capable of delivering 500 GeV–1 TeV center of mass energy
- a linear electron-positron collider based on high-gradient normal conducting rf technology and two beam acceleration techniques, capable of delivering 500 GeV–3 TeV center of mass energy
- ultra-high gradient laser and plasma wakefield accelerators, including acceleration stages that could be used either with the superconducting or the two beam acceleration linear collider options
- a muon collider facility that uses cooling of intense muon beams, capable of reaching 2-4 TeV center of mass energy.

To enable the realization of the above designs, we will need to support a world-class research and development program to develop new accelerator technologies and accelerator structures, such as

- superconducting technology and accelerator structure design
- accelerator techniques and beam dynamic concepts to control beam losses in high intensity proton sources
- high-gradient normal conducting rf technology
- advanced materials capable of revolutionary increases in breakdown voltages
- moon-based accelerator components and muon cooling concepts
- ultra-high gradient laser wakefield and plasma wakefield accelerator structures.

2.1. *Identification*

The design, cost optimization, and successful operation of modern accelerators, such as those described in the previous section, require the optimization of many parameters, and the understanding and control of many physics processes. This can only be accomplished by employing high fidelity computational accelerator models. In addition, the preceding discussion shows that there is a wide spectrum of particle acceleration methods that we need to study to enable the different design activities. Although the basic concepts that guide the design and operation of the different types of particle accelerators are similar, the technologies involved, and the critical design and operational parameters vary depending on the application, thus increasing the complexity of high fidelity modeling tools. Finally, massive computations, requiring tightly coupled parallel computing and advanced algorithms, are necessary to model many important accelerator physics processes, such as collective beam effects in beam dynamics, the properties of complex electromagnetic structures, and plasma and laser wakefield acceleration techniques. For example, high-fidelity modeling of space charge effects, beam-beam interactions, wakefield generation in electromagnetic structures and their interaction with the beam, electron cloud generation and dynamics, high precision computations of rf fields in structures with realistic geometry, and three-dimensional modeling of laser- and plasma-based systems, can only be accomplished by employing scalable high-performance computing (HPC) accelerator modeling tools.

The increased complexity, precision, beam intensity and energy requirements, and ultimately the increased cost of the next generation particle accelerators further increases the demands on computational accelerator science. Not only is it required to accurately model single-stage, single-physics, single-scale systems, but it is also required to enable end-to-end (multi-stage or complete system), multi-physics, multi-scale simulations. Such simulations can only be performed with the use of extreme scale computing resources and with algorithms that can effectively exploit these resources. This will allow designers to achieve high-model accuracy, to shorten the turnaround time in designing and optimizing accelerators, and to save cost significantly by decreasing the frequency in the trial-and-error procedure for producing accelerator component prototypes.

In our discussions, the members of the panel define extreme scale to mean the computational power of supercomputers that run at speeds of order 1 exaflop per second. This corresponds to a three order of magnitude increase over the present computing resources available for accelerator simulation (capable of petaflop performance), which define the current scope of HPC accelerator modeling tools development. Of course, only very few current applications are able to sustain this currently available throughput, because of algorithmic performance issues and because of the realities (load, reliability) of currently operating supercomputers.

2.2. *Past Experiences and Current Capabilities*

A representation of the current HPC accelerator modeling capabilities could be obtained by reviewing the capabilities developed under the Scientific Discovery through Advanced Computing (SciDAC1) Accelerator Science and Technology project, and the capabilities that are currently being developed and deployed by the SciDAC2 Community Petascale Project for Accelerator Science and Simulation (ComPASS) project (Spentzouris 2008). ComPASS codes obtain good scalability and parallel performance efficiency on many hundred to tens of thousands of processors and are routinely used to perform single-physics, single-scale simulations on a few thousand processors. These codes are already used for simulations of the accelerators and accelerator concepts described earlier in this section.

The research effort in HPC for accelerator applications under DOE support started over a decade ago with the award of the Accelerator Grand Challenge, followed by the SciDAC Accelerator Science and Technology and the SciDAC2 ComPASS projects, with the goal to develop and apply high-performance computational tools for the design, optimization and analysis of existing and future accelerator projects in the Offices of High Energy Physics, Basic Energy Sciences, and Nuclear Physics within DOE. During this period, such HPC tools have been applied to major DOE accelerator facilities and future accelerator projects including the Tevatron, PEP-II, LHC, Relativistic Heavy Ion Collider (RHIC), Next Linear Collider (NLC), International Linear Collider (ILC) design, Spallation Neutron Source, and Linac Coherent Light Source (LCLS) (Blumenfeld et al. 2007). They were also applied to advanced concepts such as plasma-based accelerators, where they played a key role in understanding the physics of doubling the energy of a 42 GeV beam at the Stanford Linear Accelerator Center and of low-energy-spread beam production in laser wakefield accelerators (Lee et al. 2005). A few representative examples of advances in accelerator science because of the successful use of HPC accelerator codes are discussed below.

As our first example, we consider the development and application evolution of the parallel finite-element electromagnetic modeling code Omega3P. Figure 6 shows the progress of the eigensolver Omega3P[9] for accelerator cavity simulation over the past 15 years. The first parallel application was to determine the

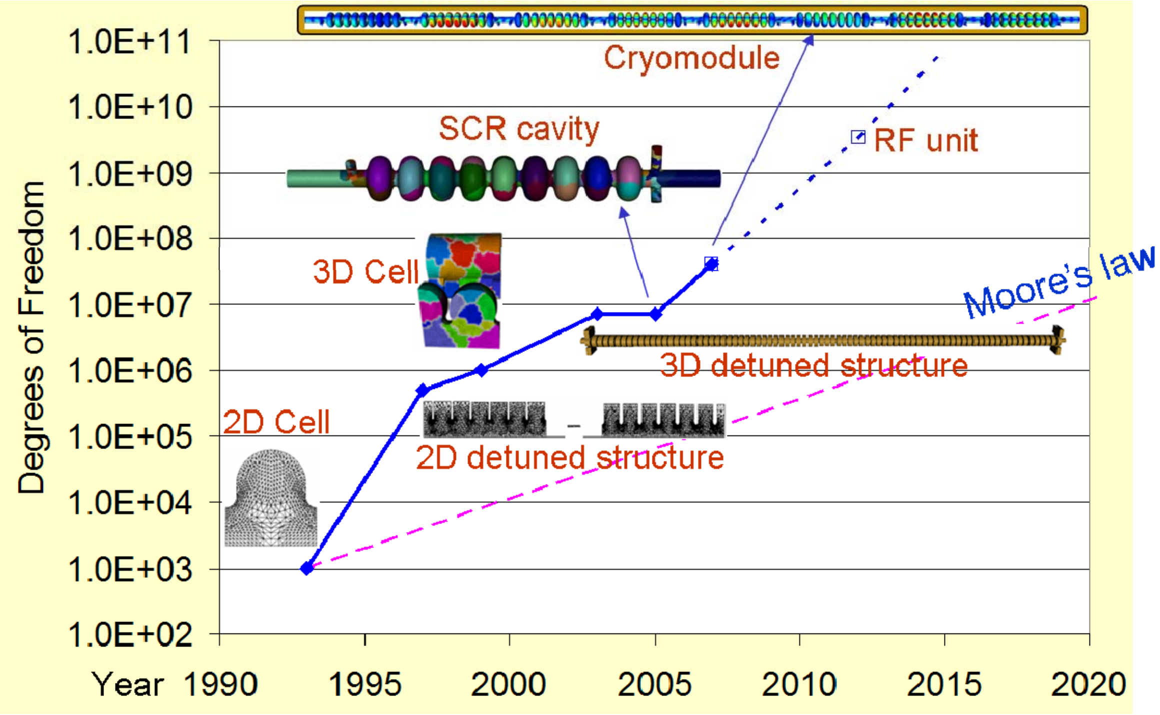
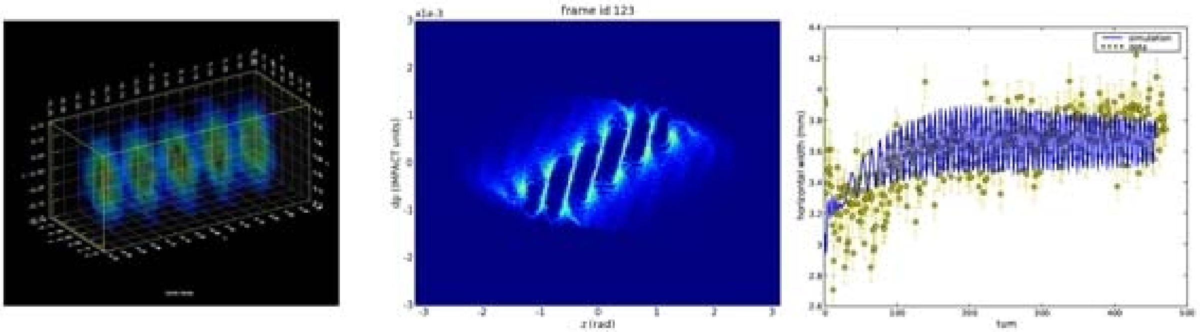


Figure 6. Evolution of computational requirements in electromagnetic structure modeling and use of Omega3P for such large-scale problems. Image courtesy of Cho-Kuen Ng (Stanford Linear Accelerator Center).

512 dipole modes in the first two dipole bands of a detuned structure for NLC. Subsequently (late 1990s), Omega3P was used for the three-dimensional cell design of a damped detuned structure to the required frequency accuracy of 0.01%, a result that was validated by subsequent quality control of the fabricated cells. By eliminating the need for cavity cell tuning, this would have saved over $100 million in the machine cost if the NLC were to proceed. The code was later used to carry out the first-ever three-dimensional end-to-end simulation of a prototype NLC 55-cell structure to verify the damping and detuning scheme for wakefield suppression. In mid-2000, Omega3P was used to calculate the dipole modes and their damping for the ILC linac superconducting cavity, and the results agreed well with measurements at the Deutsches Elektronen Synchrotron (the German Elektronen Synchrotron). In 2007, it was used for the first-ever calculation of dipole modes in the ILC linac cryomodule that consists of eight superconducting cavities. As shown in the figure, over a period of more than a decade, the problem size handled by the code has increased by five orders of magnitude while maintaining the solution accuracy at the same 0.001%. The success of Omega3P on large-scale electromagnetic modeling depends upon the coordinated research effort in accelerator science, parallel computing and applied mathematic techniques. The biggest problem that has been achieved is the evaluation of higher order modes (HOMs) in a superconducting linac cryomodule. The computational domain had 20 million degrees of freedom, and it took one hour per mode calculation using 1024 processors with 300 gigabyte memory on the National Energy Research Scientific Computing Center Seaborg supercomputer.

Another example is the development and deployment of beam dynamics frameworks with accurate three-dimensional space-charge and impedance modeling capabilities, used to predict emittance growth and generation of beam halo. Simulations performed with these codes (MaryLie/Impact, Orbit, Synergia) were used to help control losses in Fermi National Accelerator Laboratory's (FNAL's) Booster, Oak

Ridge National Laboratory's (ORNL's) Spallation Neutron Source, and to help choose the design for the damping ring of the proposed ILC. Figure 7 shows the results of the first-ever simulation of the process of linac microbunch capture (7a), debunching (7b), and acceleration, including beam position feedback models, all using a three-dimensional space-charge model, in the FNAL Booster, using Synergia. The results of our simulations were compared to beam measurements, both under normal machine operations and during machine studies (7c). These simulations provided guidance to machine operators to reduce losses and maximize Booster intensity and commission the Booster collimators (Spentzouris and Amundson 2005, Amundson et al. 2007).



(a)
Merging of 5 linac microbunches

(b)
Longitudinal phase space shows halo and space-charge "drag" during bunch merge

(c)
Beam profile evolution compared in the in the FNAL booster with Ionization Profile Monitor experimental data

Figure 7. Fermilab Booster Synergia Simulation, including three-dimensional space-charge, 2n second order maps, multi-turn injection, and rf ramping models matches experimental data. Image courtesy of Panagiots Spentzouris (Fermi National Accelerator Laboratory).

HPC simulations supported by the DOE Office of Nuclear Physics Small business Innovative Research program and now also by SciDAC were instrumental in design modifications of the proposed electron cooling system for the RHIC that led to cost reductions estimated in the tens of millions of dollars. Rather than using a technically challenging high-field solenoid magnet and 20 nC magnetized electron bunches (Ben-Zvi et al. 2005), the new design specified conventional undulator magnet technology with 5 nC electron bunches (Ben-Zvi et al. 2007). The initial effort established confidence in the new algorithms implemented within the parallel Versatile Object-oriented Relativistic Plasma Analysis with Lasers (VORPAL) framework, and contributed directly to the solenoid- based design. Subsequent algorithmic improvements and additional VORPAL simulations (Bell et al. 2008) confirmed the conjecture that the magnetic fields of the undulator would only reduce the friction force logarithmically and, thus, be a much cheaper yet viable alternative. Although detailed cost estimates are not available, it is agreed that the modified design for the RHIC electron cooler would be tens of millions of dollars less expensive than the previous design, with much lower technical risk.

The accomplishments of the HPC plasma and laser wakefield simulation efforts have led to the publication of nine Physical Review Letters (Blue et al. 2003, Hogan et al. 2000, Hogan et al. 2003, Hogan et al. 2005, Mugli et al. 2004, Mangles et al. 2004, Johnson et al. 2006, Mangles et al. 2006, Lu et al. 2006) being highlighted in the Fall 2006 volume of SciDAC Review, and in the publication of four articles, including one cover picture, in *Nature* (Blumenfeld et al. 2007, Geddess et al. 2004, Mangles et al. 2004, and Faure et al. 2004). Successful applications include modeling the major U.S. plasma-based accelerator experiments at the Stanford Linear Accelerator Center (SLAC) and Lawrence Berkeley

National Laboratory (LBNL), developing key physics understanding of plasma wakefield acceleration (PWFA) and laser wakefield acceleration (LWFA) in the nonlinear blowout regime, and applying plasma-based codes to e-cloud interactions in conventional accelerators. In some cases, these codes predicted important effects in plasma acceleration, later to be verified by experiment, and helped guide the design of the experiments.

3. OPPORTUNITIES OFFERED BY EXASCALE COMPUTING

To go beyond the current capabilities to meet the goals outlined above, the community needs to

1. improve the performance while ensuring the scalability to approximately 10^6 fold parallelism of the underlying algorithms of each individual physics capability code
2. create a Virtual Accelerator Modeling" environment for the realistic, inclusive simulation of all relevant beam dynamics effects (single and multi-particle dynamics, realistic geometry and parameters), and a "Virtual Prototyping" environment for realistic simulation of all relevant accelerator component effects (thermal, mechanical, and electromagnetic properties with accurate geometry description).

Both objectives will require close coupling between domain, applied math, and computer scientists. More specifically, they will require rethinking how data are organized and will depend on computer science advances (development of frameworks and framework components), and result in some non-trivial constraints on the software infrastructure and support that will be required on the future extreme scale computing facilities. To further underline the importance of this second objective and its appeal as an extreme scale computing application, it is worth noting that in effect it transforms the strong-scaling requirements of single-physics applications to weak-scaling requirements for the multi-physics applications (since the size of the problem can now be much bigger, with the inclusion of many physics processes), while allowing for much improved design cycle.

Exascale computing will enable us to

- design and optimize a lepton collider linac module for cost and risk reduction by using unscaled ebam structure simulations for the first time
- predict beam loss and resulting activation in Intensity Frontier accelerators by developing and employing multi-scale, multi-physics beam dynamics simulations to cover the full range of scales relevant to accelerator design from 10^{-3} m beams, to 10 m wakefields, to many 10^3 m propagation
- shorten the design and build cycle of accelerator structure by developing and employing multi-scale, multi-physics accelerator structure simulation capabilities that integrate thermal, mechanical, and electromagnetic effects
- develop and design an ultra-compact plasma-based collider using high fidelity, multi-scale, self-consistent models that simulate plasma experiments, non-linear wakefields, beam loading, and beam dynamics for full-scale collider parameters. Such simulations will cover a range of scales from 10 nm beams, to 10 μm wakefields, to m stages, to 100 m accelerator lengths.
- develop the techniques necessary to design neutrino factories and muon colliders by employing end-to-end simulations of extreme ionization cooling of ultra-intense muon beams and multi-physics simulations of re-bunching schemes required to produce ultra-short high-intensity proton bunches.

3.1. *Design and Optimize Lepton Collider Linac and Proton Driver Linac Modules for Cost and Risk Reduction*

Future lepton linear colliders have very stringent design tolerances. Today's HPC simulation capabilities do not allow modeling of heat deposition because of higher-order modes with realistic beam size parameters. Such effects are essential for the successful design of the linac. Extreme scale computing tools will allow building accurate, realistic models of these accelerators that will ensure performance and minimize risk and cost. The time scale for the design of such an accelerator is 10 to 15 years based on the physics results from LHC.

Unscaled beam-structure simulations for the first time result in unprecedented size of computational problems. For example, the simulation of heating because of the transit of a realistic bunch in a superconducting linac involves 10^{13} degrees of freedom (DOF) for 100K time steps, resulting in computing requirements that are 10^{6} times larger than current capabilities.

The success of this type of simulation activity on the new generation extreme scale computers requires advances in meshing, sparse-matrix algorithms, load balancing, higher-order embedded boundaries, optimization, data analysis, visualization, and fault tolerance.

The time scale for needing to deploy such capability is determined by the time scale for of the next generation lepton collider, which is 10 to 15 years. The decision for such machine depends on the outcome of the LHC physics program.

To illustrate the advances in the design of the next generation lepton colluder and proton driver linac modules that only extreme computing could enable, we discuss a few concrete examples of such numerical simulations. The parameters discussed are extrapolated from simulations of current experiments to the needs of future accelerators in the energy and intensity frontiers.

The electromagnetic simulation of HOM heating because of the transit of a realistic, short bunch through a cryomodule in a superconducting linac such as of lepton linear colliders or proton drivers involves solutions of a linear system with 10^{13} DOFs for 100K time steps, or solutions of a nonlinear eigenvalue problem with similar number of DOFs. The objective is to calculate the wakefields in a 3 cryomodule rf unit (26 cavities) with realistic three-dimensional dimensions and misalignments and the beam heating for realistic beam bunches. The estimated computational requirement is 10^{22} flops. With today's state of the art codes, the largest problems tackled for time-domain analysis involve 80 million-element mesh, approximately 500 million DOFs, 4096 central processing units (CPUs) (Jaguar), with 4 seconds per time-step (Figure 8), and for frequency domain 3 million-element mesh, approximately 20 million DOFs, 1024 CPUs (Seaborg), 300 gigabyte memory, requiring 1 hour per mode.

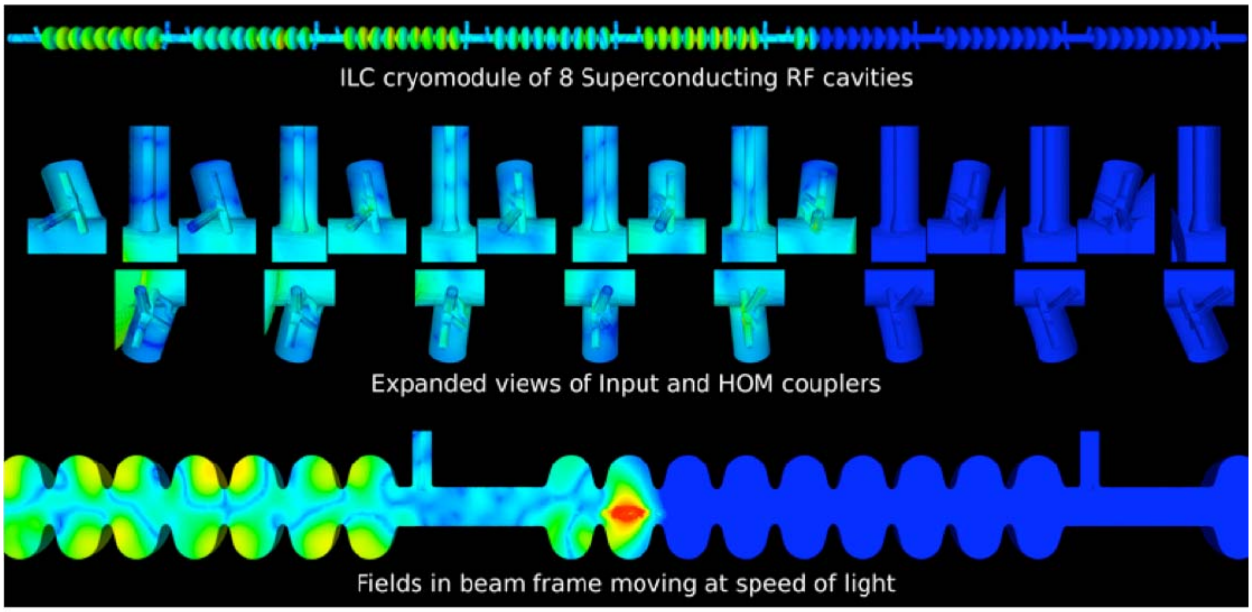


Figure 8. A snapshot of a wakefield generated by an electron bunch traversing an ILC cryomodule, using T3P, a code developed at SLAC in the context of the SciDAC2 ComPASS program. Image courtesy of Cho-Kuen Ng (Stanford Linear Accelerator Center).

3.2. *Predict Beam Loss and Resulting Activation in Intensity Frontier Accelerators and Maximize Performance of Energy and Intensity Frontier Accelerators*

In order to maximize the reach of HEP experiments at lepton and proton colliders (luminosity, beam lifetime) and proton drivers (beam intensity) beam losses need to be controlled. Such losses cause accelerator component activation and damage, thus limiting the performance of the machine. Current simulation capabilities prohibit the required detailed understanding of these effects, thus requiring large safety factors in the design, which result in increased cost. Even with the safety factors the risk is substantial because the requirements of the next generation accelerators are much more extreme than those in operating machines, so that simple scaling from experimental data is not sufficient.

Such simulations must include all physics effects spanning all relevant length scales in an accelerator: from the approximately 10^{-3} m beam size, to approximately 10 m span of wakefield propagation, to many 10^3 m length of beam propagation. They require the development and employment of multi-scale, multi-physics beam dynamics frameworks. Using these frameworks to compute beam losses requires extreme scale computing. For example, to enable the design of beam extraction for high intensity proton driver the computing requirements are 10^5 times larger than current capabilities.

The success of this development activity requires understanding of highly nonlinear collective beam effects which in turn require advances in particle-in-cell technologies, pipelining algorithms, multi-language software infrastructure, data analysis, visualization, fault tolerance, and optimization.

The timescale for using such capabilities with impact on the next generation high-intensity machines is 5 years (for the design) to 10 years (for the operation).

In order to illustrate the benefits in the design of the next generation energy and intensity frontier accelerators that extreme computing could enable, we discuss a few concrete examples of necessary numerical simulations. The parameters discussed are extrapolated from simulations of current experiments to the needs of future accelerators in the energy frontier.

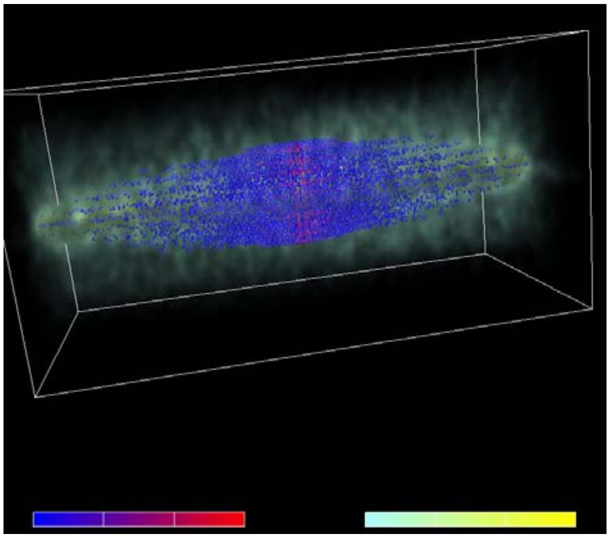


Figure 9. Electron cloud simulation of a ring section. Image courtesy of Panagiotis Spentzouris (Fermi National Accelerator Laboratory).

Currently, electron cloud generation models are decoupled from beam dynamic models (Figure 9). To perform high-fidelity simulations, we need to couple these models to accurate models of the accelerator and other beam dynamics effects, while treating the beam and the electrons of the cloud self-consistently. Just modeling the electron cloud dynamic effects is a petascale problem using a quasi-static particle-in-cell (PIC) code. For high-intensity proton driver, it is required to have a self-consistent model and also to include impedance and space-charge effects, thus requiring extreme scale capabilities. As another example consider the simulation of electron cloud effects in a damping ring for a future lepton collider that uses wiggler magnets. The beam pipe radius, wiggler period, and beam size, lead to a mesh of size 3000 x 3000 x 6000, or more than 50 billion cells. The simulation time step follows from the damping time and the time for an electron to traverse one simulation cell in one time step (roughly 10-13 seconds); the result is that 100 billion time steps are required.

Another requirement for predicting losses and understanding performance issues of a modern accelerator is the ability to perform long-term tracking with collective effects. For example, at the moment we are not able to perform a sufficient long simulation of nominal LHC bunch parameters (3000 bunches) considering coupled bunch effects, which can cause instabilities (coupled bunch instability). Studying these effects is essential for maximizing luminosity, predicting luminosity lifetime, and exploring new techniques to minimize disruption to reduce the risk to achieving our luminosity goals (high luminosity is essential for scientific discovery at energy frontier colliders). Hence we need long-term tracking with wakefields and possibly space charge for European Organization for Nuclear Research (CERN) Super Proton Synchrotron and LHC. So for any LHC upgrade scenario this is a problem that requires extreme scale computing. The requirements for such simulations are listed below:

- three-dimensional “strong-strong” (i.e., self-consistent) beam-beam model
- complex interaction pattern—multiple bunches, multiple collision points, head-on, long-range
- nonlinear external elements—conducting wire, crab cavity, electron lens, multipole magnets
- long-term tracking with large number of macroparticles
  - now: 10^5 turns; needs: at least 10^7 turns [100X]
  - now: 1M particles; future: 100M particles [100X]
  - now: 1282 grid; future: 5122 grid [16X]
  - now: 1-on-1 bunch; future: 3-on-3 bunches [9X]
  - now: ~1500 proc-hours on Cray-XT4; future: 106 X more challenging.

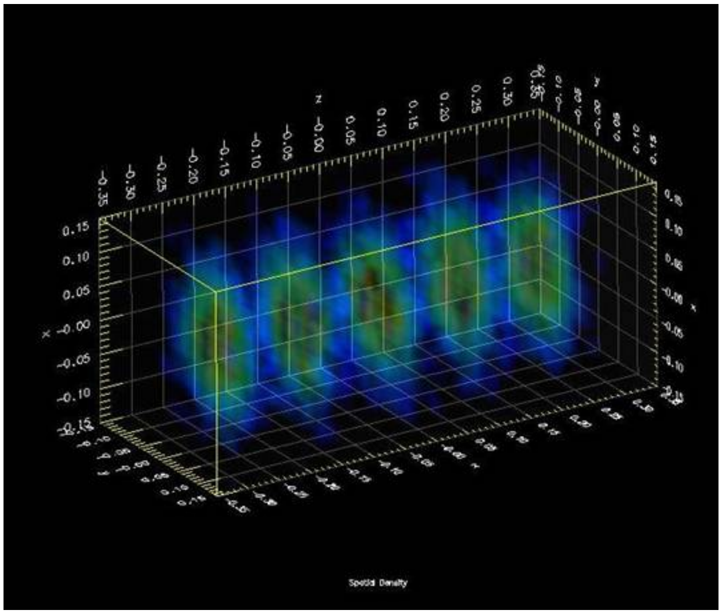


Figure 10. One million macroparticle simulation of the Fermilab Linac beam injected into the Fermilab Booster, using the Synergia framework, a ComPASS code developed at Fermilab. The five linac bunchlets will be captured into one Booster bunch. Image courtesy of Panagiotis Spentzouris (Fermi National Accelerator Laboratory).

In addition, it is necessary to be able to perform multi-objective optimization of accelerator components or complexes. Having the ability to perform large-scale optimizations to determine an optimal set of accelerator beam parameters is crucial for the development and operation of future facilities at maximum performance, luminosity for the Energy Frontier, and losses for the Intensity Frontier accelerators. As our final example, we consider the beam requirements for rare process searches at the Intensisty Frontier. The proposed Mu2e experiment at Fermilab will use a high-intensity 8 GeV proton beam. The beam will be circulated first through the Recycler, then the Accumulator and finally the Debuncher, from which it will be extracted via resonant extraction. A high-intensity beam such as this is subject to multiple collective (intensity-dependent) effects. The largest initial concern is the effect of space charge on the resonant extraction procedure in the Debuncher. The resonant extraction process occurs over the space of 10,000 turns. The limitation on the step size of such a simulation derives from the oscillations of the beam envelope, which requires multiple samples per period. In the Debuncher, the beam envelope oscillates 57 times. A minimal simulation of the resonant extraction process will therefore require $4 \times 57 \times 10000 = 2.3 \times 10^6$ steps. The simulations performed today (Figure 10) use one bunch of 1 million macro-particles and model only one machine (2 million simulation steps). To accurately model losses, we need 100 million macro-particles, model all accelerators in the accelerator chain, and simulate 100 bunches, including bunch-to-bunch effects. This results in computational requirements 10^5 larger than currently used.

3.3. *Shorten Design and Build Cycle of Accelerator Structure*

The design, optimization, and reliable operation of accelerator components and systems have to take into account electromagnetic, thermal and mechanical effects. The ability to shorten the time for the design and build cycle for engineering prototyping will substantially reduce the cost for an optimized design that satisfies beam quality preservation and machine operational reliability. The integrated capability that can deal with accelerator entities at the component level has already been developed (Figure 11), but for those at the subsystem or system levels such as the superconducting linac cryomodule or an entire rf units computing power at extreme scale is required, as illustrated in the single electromagnetic simulation described in Section 4.1.

Such simulations require the development of multi-scale, multi-physics accelerator structure simulation framework, which includes integrated electromagnetic, thermal and mechanical analysis. To develop such framework advances in meshing, load balance, solver, coupling technology (example, mesh to mesh), optimization, data analysis and visualization (*in situ*) are necessary.

The time scale deploying such capability is determined by the time scale for the next generation lepton collider, which is 10 to 15 years. The decision for such machine depends on the outcome of the LHC physics program.

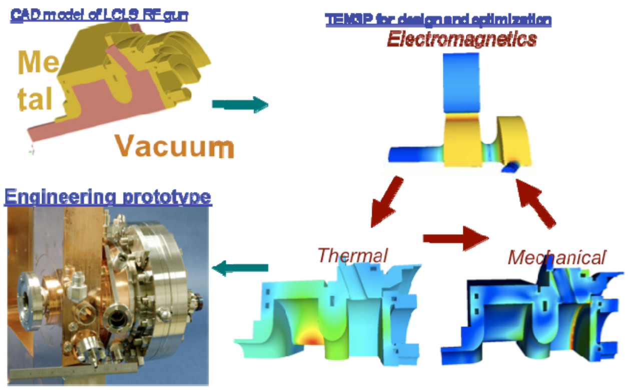


Figure 11. Integrated simulations of the LCLS rf gun using the ComPASS code TEM3P developed at SLAC start with a CAD model of the gun and include Electromagnetic, Thermal, and Mechanical models. Image courtesy of Cho-Kuen Ng (Stanford Linear Accelerator Center).

A two-beam module of the Compact Linear Collider (CLIC) accelerator consists of drive beam structures, Power Extraction and Transfer Structure (PETS) and main beam accelerator structures. Combined simulations of a two-beam module to study the wakefield coupling between PETS and accelerating structures, and the dark current capture in the main linac are necessary for a successful design. The electromagnetic simulation of wakefield coupling between the two types of structures involves solutions of a linear system of 10^{14} degrees of freedom for 10,000 time steps, or solutions of a nonlinear eigen value problem with similar number of DOFs (Figure 12) The estimated computational requirement is 10^{23} flops. Including coupled electromagnetic, thermal and mechanical optimization for this many-component system requires extreme scale computing capabilities.

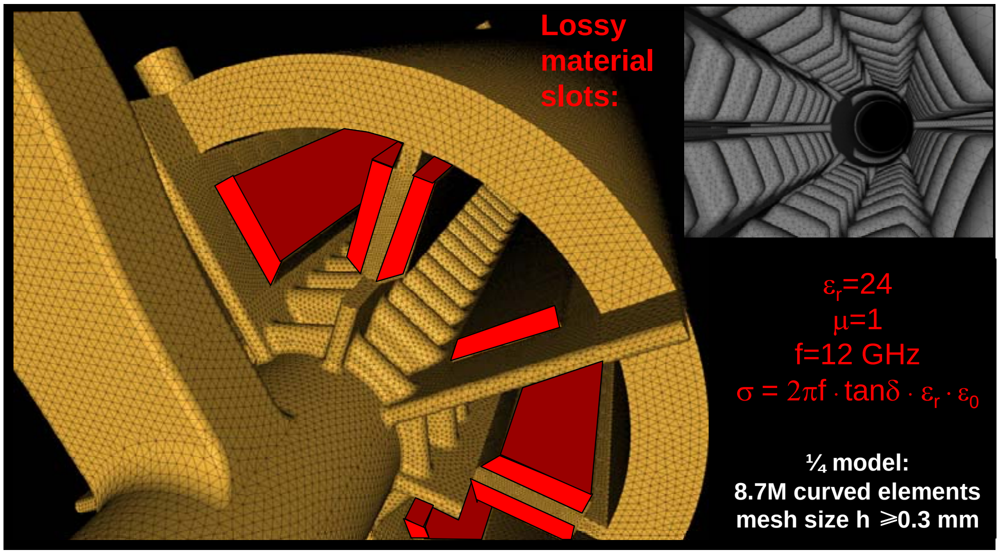


Figure 12. The PETS structure features a very complex geometry. Optimization of the CLIC structures requires extreme scale computing resources. Figure courtesy of Igor Syratchev (CERN).

3.4. *Developing and Designing a Compact Plasma Based Collider*

The limiting factors for designing next generation lepton colliders are size and cost. Plasma-based accelerators have the potential to provide acceleration gradients three orders of magnitude beyond existing technologies and to significantly reduce the cost. To realize this potential requires understanding the physics of plasma wakefields driven by particles and lasers. To achieve this understanding with the necessary high fidelity requires resolving the smallest beam dimension (approximately 10 nm), and the wakefield (approximately 10μm) while the driver propagates for meter distances. In addition, modeling an entire accelerator will require coupling 10 to 100 stages together. Simulations relevant to collider design can only be performed on extreme scale computing. For example, moving to matched beam spot sizes from the few microns scale (current simulation reach) to the 10 nm scale (required to design a future collider) results in five orders of magnitude more computational power than is employed today.

The success of this effort requires advances in PIC methods, pipelining algorithms for quasi-static PIC models, multi-language software infrastructure, performance, data analysis, visualization, fault tolerance, dynamic load balancing, and mesh refinement. Massively parallel reduced models of beam-plasma and laser-plasma interactions could also provide real-time feedback for future experiments and accelerator systems. Such models must be carefully benchmarked with the time-explicit PIC simulations. Improving the fidelity and scaling of these reduced models will also require new algorithm development.

The timescale for this development is dictated by the need to design the next generation accelerator beyond the LHC and a superconducting or normal conducting linear accelerator. Plasma-based acceleration could also be used as a booster to either option. Taking this possibility as setting the timing, the relevant time scale is approximately 2030.

To illustrate the advances in the design of the next generation plasma accelerators and concepts that extreme computing could enable, we discuss a few concrete examples of necessary numerical simulations. The parameters discussed are extrapolated from simulations of current experiments to the needs of future accelerators in the energy frontier.

With today's state-of-the-art-codes, simulations of a 1 GeV PWFA stage using an explicit PIC code and 10 μm size beams requires 1,500 processor hours. Modeling the Lasers, Optical Accelerator Systems Integrated Studies (LOASIS) 1 GeV experiment requires 2,500,000 processor hours with the same code because of the need to resolve the laser wavelength and the plasma channel. Without HPC hardware and software resources, such simulations and the analysis of the resulting data would not have been possible (Figure 13). As wakefield accelerator experiments and concepts move to the higher energies required to be relevant to lepton collider applications (approximately 10 GeV), and the beam specifications become more demanding (a few nanometers) the computational requirements increase.

More specifically, let us consider the experimental roadmap to a plasma-based collider in 2025 as conceived by the Facilities for Accelerator Science and Experimental Test (FACET) Beams collaboration at SLAC. The plan consists of three phases: FACET I from 2009-2017, FACET II from 2017-2025, and the final design determination in 2025. The FACET I facility will test the physics of one stage of a 500 GeV plasma collider based on 10 electron and 10 positron stages of 25 GeV energy gain each. It will demonstrate high efficiency and monoenergetic acceleration of a precisely phased second bunch from 25 to 50 GeV on the wake of a lead bunch. The transverse spot size is on the micron scale in FACET I and sets the cell sizes and scale of simulations. These simulations are being actively pursued at present using

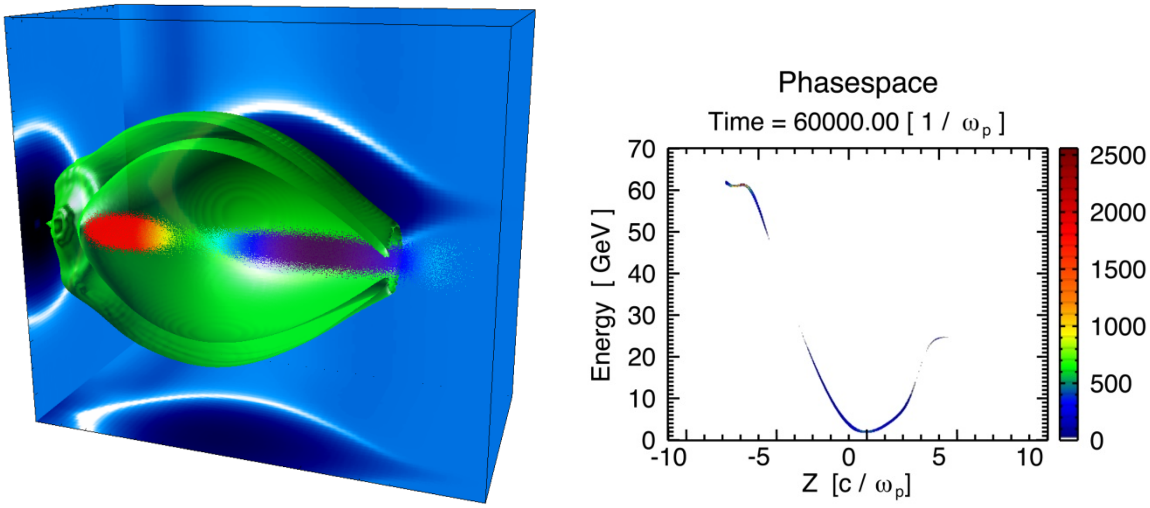


Figure 13. On the left are isosurface plots of the electron density for a 25 GeV PWFA stage. A drive (mostly blue) and trailing beam (mostly red) are propagating from left to right. The plasma wake structure is green. On the right, the energy vs. position of the drive and trailing beams after 1 meter of propagation is shown. The simulation was done using the quasi-static PIC code QuickPIC, developed at the University of California, Los Angeles, with partial support from the accelerator SciDAC programs. Depending on the resolution, such a simulation is typically performed on 250 to 2,000 processors. For the resolution needed to model 1 TeV collider parameters, the simulation would need to be run on approximately 1 million processors. Image courtesy of Warren Mori (University of California, Los Angeles).

a reduced, i.e., quasi-static, PIC code with typically 1,000 processors. FACET II will demonstrate the staging of two of these 25 GeV meter-long modules and will address the extreme transverse requirements on emittance and pointing. Such beams will have matched beam spot sizes at the 100-nm scale, require tracking of mobile ions and higher ionization states and consequently need one to two order of magnitude more cells and processors. By the end of FACET II, detailed simulations of the final collider will be needed. At the final energy of 250 GeV, the beams further pinch to dimensions of order 10 nm in the y-direction and the system size will need to be on the order of one million processors. Such resolution will most likely require quasi-static PIC. An example of a simulation of the first stage of a plasma based collider is shown in Figure 14.

As a second example, we consider the roadmap to develop a plasma-based collider driven by high power lasers now being pursued by the Berkeley Lab Laser Accelerator team at LBNL. A 500 GeV - 1 TeV collider is envisioned to consist of 50-100 stages, each with 10 GeV gain and 1 m plasma length. The first phase of the Berkeley Lab Laser Accelerator experimental program, proposed to begin in 2009, will study the physics of a single 10 GeV stage powered by a 1 petawatt laser and the staging of two multi-GeV modules. Single-stage physics issues to be studied included controlled injection of particles, preservation of low-energy spread and low emittance in the plasma structure, and efficiency. Staging physics issues include the in-coupling of fresh laser pulses and particle beam transport between stages. Because of the requirement to resolve the (micron scale) laser wavelength and bunch size over the wake volume and acceleration distance, present simulations of cm-scale GeV experiments require millions of processor hours on thousands of processors using explicit particle in cell models. Simulation of a m-scale 10 GeV stage will require order of 1 million processors. Reduced models that average out the laser and model the wake using quasi-static or explicit PIC explicit but still include much of the essential physics will still require thousands of processors or more per stage. Modeling of beam evolution and emittance through the tens to hundred stages of a collider places stringent demands on momentum accuracy, which is anticipated to require order of a million processors even with reduced models, and many million for explicit models.

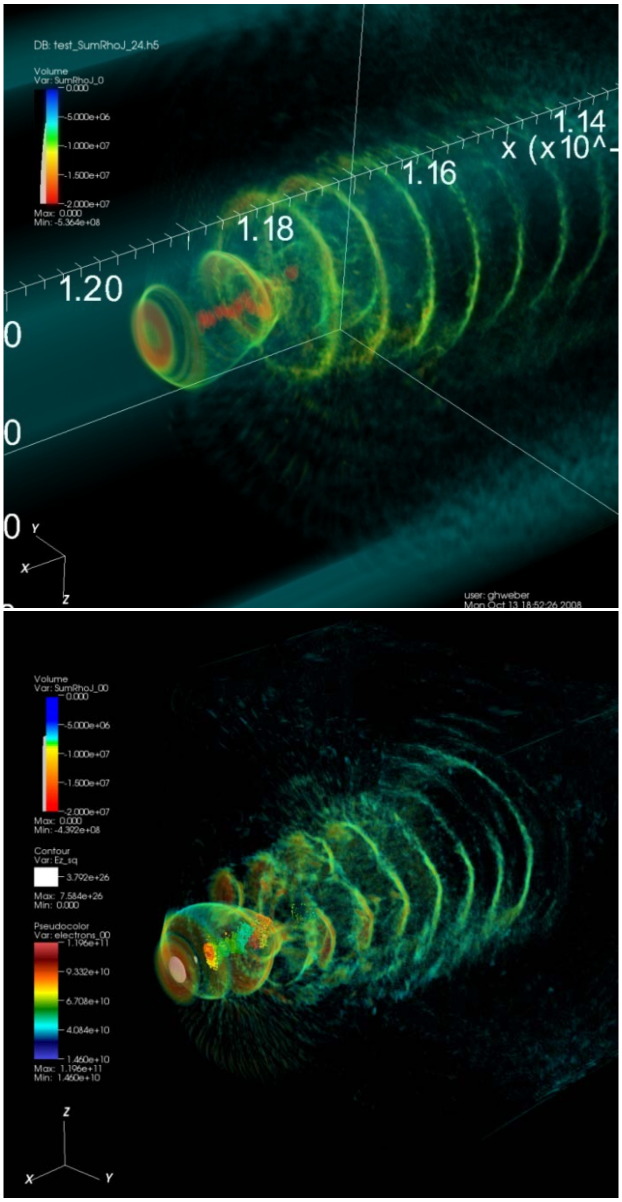
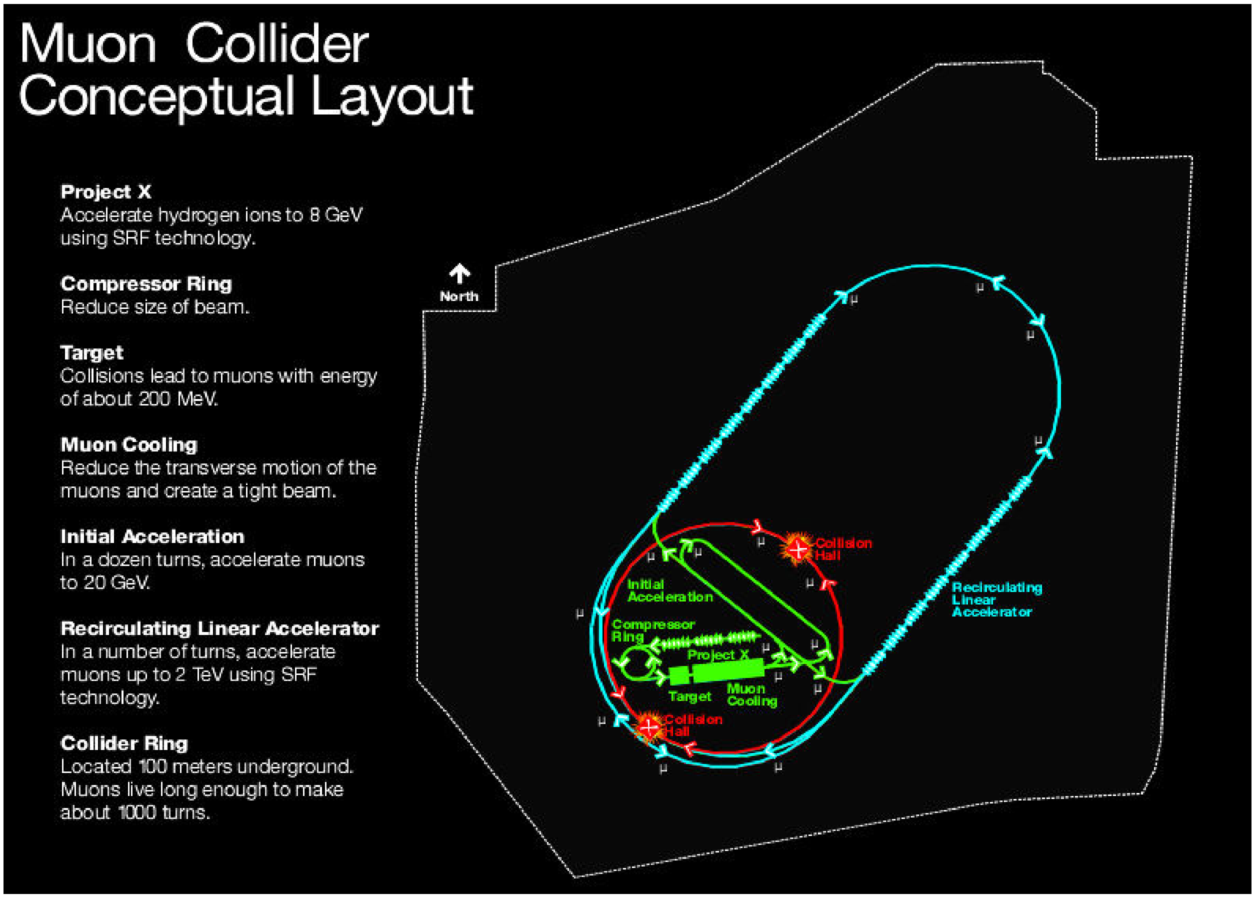


Figure 14. An intense laser pulse (traveling from right to left) produces a density wake in a hydrogen plasma, shown by colored density isosurfaces in VisIT parallel visualizations of VORPAL simulations modeling LOASIS (LBNL) laser wakefield accelerator experiments. On the right, high-energy particles are overlaid with the wake, colored by their momentum, facilitating understanding of how these experiments produced narrow energy spread bunches for the first time in an LWFA (red). The present simulations run on 3500 processors for 36 hours, and visualization of the 50 GB/time snapshot datasets runs on 32 processors taking tens of minutes/snapshot. Future experiments will increase these demands by orders of magnitude. VORPAL is developed by the Tech-X corporation, partially supported through the SciDAC accelerator modeling program (ComPASS). Image courtesy of David Bruhwiler (Tech-X Corporation).

3.5. *Developing the Techniques Necessary to Design Muon Based Accelerators*

A muon collider is a possible future path for the United States to get back to the Energy Frontier. Since muons are much heavier than electrons, they do not radiate as much and can therefore be accelerated to very high energies in relatively small circular accelerators, which would fit on the site of existing DOE laboratories (e.g., a 4 TeV Muon Collider could fit on the Fermilab site as shown in Figure 15). For the same reason, at a given energy, a high energy muon collider will also have a much smaller energy spread than an electron collider, enabling precision energy scans and efficient resonant production of e.g., Higgs-like particles.



Muon Collider Conceptual Layout

Project X
Accelerate hydrogen ions to 8 GeV using SRF technology.

Compressor Ring
Reduce size of beam.

Target
Collisions lead to muons with energy of about 200 MeV.

Muon Cooling
Reduce the transverse motion of the muons and create a tight beam.

Initial Acceleration
In a dozen turns, accelerate muons to 20 GeV.

Recirculating Linear Accelerator
In a number of turns, accelerate muons up to 2 TeV using SRF technology.

Collider Ring
Located 100 meters underground. Muons live long enough to make about 1000 turns.

Figure 15. A muon collider scenario for the Fermilab site. Image courtesy of Fermi National Accelerator Laboratory.

Muons are produced by an intense beam of protons impinging on a target. As a result, the muon beam starts out with a large emittance, and its phase-space must be reduced ("cooled") before it can be used in a collider. Furthermore, since muons have a short lifetime, the beam cooling and acceleration process must be rapid. The only cooling scheme that works on this timescale is ionization cooling, using a long sequence of alternating low atomic number (low-Z) absorbers and accelerating rf cavities (see Figure 16 for an example of a muon cooling channel design).

It is expected that muon ionization cooling concepts will be tested experimentally the first experiment is under commissioning at the Rutherford Appleton Laboratory, but for the selection of the optimal ionization cooling scheme the design of the Muon collider must rely on simulations. Because these are multi-parameter schemes, affected by many physics processes, the successful selection of the optimal extreme muon cooling scheme will depend on efficient use of extreme computing resources, both for cost and design time minimization. Also, because of the short muon lifetime, it is conceivable to simulate the entire accelerator chain, from muon production to collisions, with the use of extreme scale computing resources. Such an end-to-end simulation is an important milestone in the design of a Muon Collider, as can be seen in the 5-year muon collider research and development plan recently submitted to DOE (Fermilab 2008). In addition to the design of the optimal muon cooling scheme, muon colliders present many other unique challenges that could be met efficiently with the use of extreme computing resources: need for very intense proton sources with extremely short bunches, liquid production targets, low-frequency and high-gradient rf with absorbers able to operate in high-fields, muon matter interactions, space-charge effects, and collimation and shielding of the interaction region, to name a few.

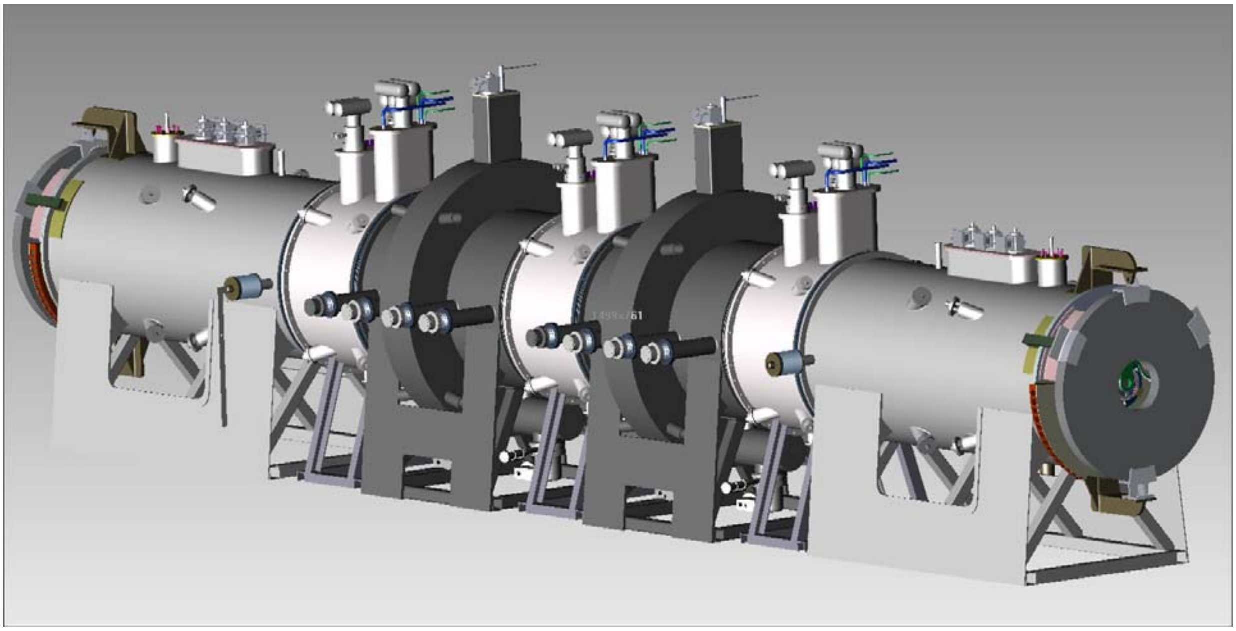


Figure 16. A Scheme for Muon Ionization Cooling Based on High-field Solenoids and rf Cavities with Absorbers. Image courtesy Alan Bross (Fermi National Accelerator Laboratory)

The timescale for developing muon accelerator technology is determined by the time scale for of the next generation lepton collider, which is 10 to 15 years. The decision for such machine depends on the outcome of the LHC physics program, and the progress of the design effort.

The design requirements for muon-based systems present unique simulation challenges. 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. All of these subsystems have extreme computing modeling requirements in both their beam dynamics and electromagnetic simulation needs.

- Muon Collider design requires a multi-megawatt proton driver. In addition to the computational challenges for high intensity proton drivers in general, a particular issue related to this machine is the need to repackage the beam to hit the muon production target in very short and intense single bunches at repetition rates of about 10 to 60Hz. Simulations of rf manipulations such as rebunching and bunch compression under high space charge condition will therefore be critical.

- The proposed technology for the muon production target is a liquid mercury jet immersed in a strong magnetic field. A prototype of such a target was recently successfully tested with beam. Extensive MHD simulations will be required to optimize the target design.
- The operation of rf cavities in high-field regions will require extensive studies of multipactoring and dark current effects.
- Wakefields during acceleration and beam-beam effects at collisions are expected to be significant in a muon collider, and will require extensive simulation study. Beam-related detector backgrounds and mitigation by collimation, for example, will also require a significant simulations effort.
- Modeling requirements for the front end of the machine involve accurate simulation of particle production, beam transport, and muon-matter interactions. In addition, for the high-intensities necessary for muon colliders, space charge effects are important and require accurate modeling. Because of the need to model muon-matter interactions, 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. Currently, the simulation packages used in the design studies are serial computing single particle tracking codes that are slow 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 use of a HPC simulation toolkit, which incorporates both muon-matter interactions and multi-particle effects, is necessary for obtaining this optimization capability and a complete physics description (Figure 17).

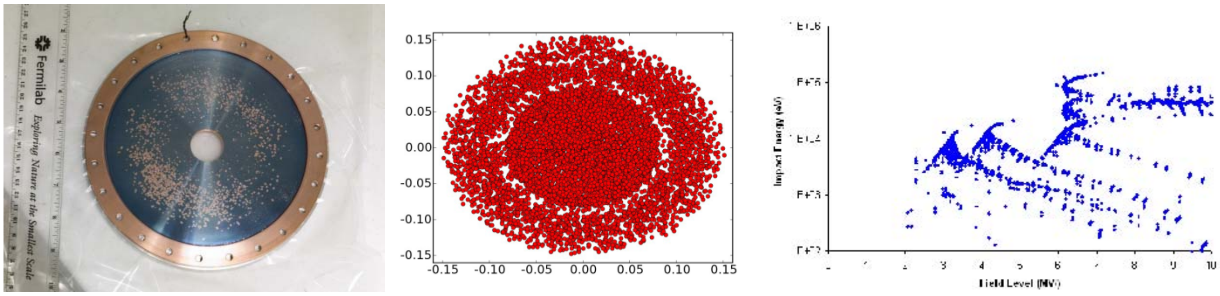


Figure 17. Observed multipactoring effects in running a FNAL 805 Mhz rf cavity with a Be absorber (left), compared to HPC simulation results: simulated pattern center (VORPAL), and parameter scan (Track3P). Image courtesy of Panagiots Spentzouris (Fermi National Accelerator Laboratory).

A detailed estimate of the HPC resources necessary for the successful simulation study of the above components of the muon accelerator is not possible, since currently there is limited use of HPC resources in the design effort. On the other hand, each one of the necessary ingredients for the end-to-end simulations that the muon collider community has identified as essential for the success of their program is a petascale size problem (this can be easily estimated from our experience in modeling similar problems for other accelerator designs). The deployment of multi-physics simulations of all these processes in order to perform multi-parameter optimization runs requires extreme scale HPC resources.

4. RESOURCES NEEDED

The development of HPC accelerator modeling capabilities would not have been possible without the Grand Challenge program, which set the foundation for such development, and the SciDAC1 and SciDAC2 programs that created the framework for collaboration with the math and computer science communities. Such collaboration has been essential for enabling the development of petascale and the development and deployment of terascale capable accelerator modeling codes. Since the evolution of HPC hardware to the extreme scale will most likely introduce significant changes to the current paradigm, a program similar to SciDAC would be necessary. The evolution of resources dedicated to HPC accelerator modeling development full-time equivalent/year, for each of the three 5-year programs is shown in Figure 18. The change in computing paradigm will require an increase in available resources at least as large as the increase between the Grand Challenge and SciDAC programs. In addition, it will require a significant increase in the investment on software infrastructure support on the extreme scale supercomputers, to enable the necessary software development.

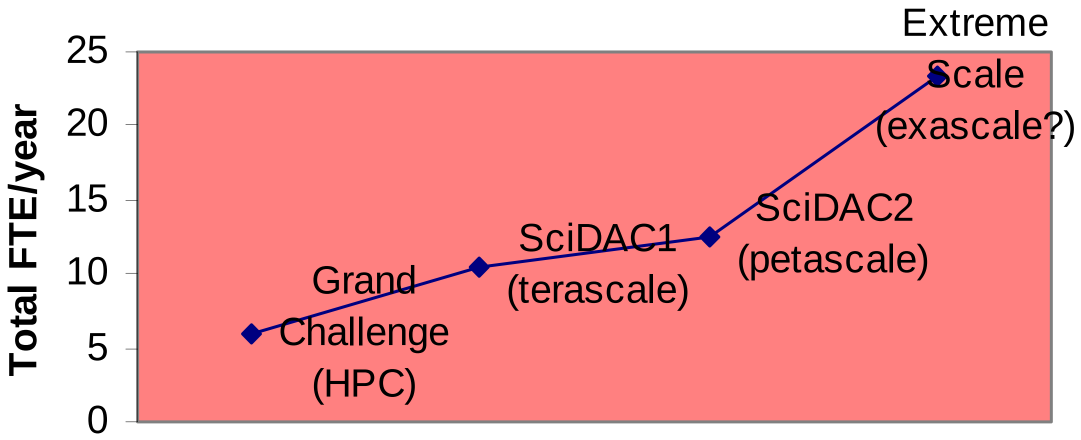


Figure 18. Evolution of resources dedicated to HPC accelerator modeling.

5. CROSS-CUTTING RESEARCH DIRECTIONS

To tackle the challenges of accelerator modeling at the extreme HPC scale, a qualitatively different level of applied mathematics and computer science capabilities will be required from what is available to today's petascale accelerator simulations. Here we discuss the required research and development in some of the most important topics for accelerator modeling in these areas.

5.1. *Applied Mathematics and Computer Science Support*

Mathematical techniques, advanced scalable numerical algorithms and computational tools continue to play an important role in future research and development in accelerator physics. HPC accelerator modeling application codes use finite difference, finite element, and PIC methods, employing either structured or unstructured meshes, with time-domain or frequency-domain analysis. Because of the expected sophistication and heterogeneity (because of the possible use of multicore, many-core, and/or graphics processing units) of future HPC architectures, in order to achieve extreme scale capabilities, much focus is needed on studying and optimizing the parallel performance and aspects of the single node performance of the algorithms and solver implementations, on which future accelerator simulation codes depend. In some cases, it may be necessary to take a fresh look at the applications. Perhaps new formulations of the applications and/or entirely new mathematical algorithms may be needed.

A partial list of relevant research activities is described in the following paragraphs.

5.1.1 Sparse Matrix Algorithms at Extreme Scale

The performance of the finite-element electromagnetic code implementations is strongly dependent upon the efficiency of sparse matrix algorithms such as sparse linear system solvers. Sparse linear systems also appear in large-scale eigenvalue calculations in electromagnetic modeling. The current petascale platforms present already unprecedented challenges to performance of these finite-element electromagnetic codes because of the increasing parallel scale and the architectural complexity of the hardware. Research toward optimizing performance of these codes for extreme scale resources should aim to achieve better performance and scalability for the finite-element simulation, especially for sparse matrix computations through more efficient distribution of work and memory usage across processors, reduced communication for high-order finite elements, improved per-core performance for sparse matrix-vector computations with better locality, and fine-grained parallelism on multi-core architectures. Since future HPC architectures are expected to have small per-node memory, research on robust, limited-memory linear system solvers will become particularly urgent and important. Increasing the number of CPUs will not be able to solve larger problems without scalability of per-node memory usage in solvers. To name a few of extreme-scale linear systems, the electromagnetic simulation of HOM heating because of the transit of a realistic, short bunch through a cryomodule in a superconducting linac such as of lepton linear colliders or proton drivers involves solutions of a linear system with 10^{13} degrees of freedom. Modeling a two-beam module of the CLIC consisting of drive beam structures and main beam accelerator structures involves solutions of a linear system of 10^{14} degrees of freedom.

5.1.2 Meshing at Extreme Scale

Meshing is crucial step in the simulation process. It can affect the accuracy of the overall accuracy. Finite element meshes for the modeling of the structures in future generations of accelerators are expected to have 10^{13} to 10^{14} elements. Therefore, there is a need to generate the mesh on the fly in parallel to avoid moving a huge volume of data from disk to memory. Improving parallel mesh generation and ensuring mesh quality for high-fidelity modeling of extreme-scale complex problems are important for high-confidence and high-fidelity simulations. Supporting local refinement without changing global data structures is required to avoid expensive and unnecessary re-computation of the whole matrix. Scalable mesh-to-mesh transfer interfaces and algorithms are needed for multiphysics coupled applications. Additional interface and framework support is also needed for different algorithms (analysis, load balancing, visualization) operating on local portions of the mesh, and for providing different “views” of the mesh when various algorithms in the workflow are operating on different local subsets of the mesh. Mesh data access must be optimized for the deep memory hierarchies found in extreme scale computers.

5.1.3 Large-Scale PDE-Constrained Optimization

The extreme-scale computing will not only enable large-scale analysis, but also enable design, control, and parameter optimization of accelerator structures through partial differential equation (PDE)—constrained optimization. Uncertainty quantification of shape of accelerator structures with measured observables will address physics issues that cannot be resolved experimentally. Design of accelerator structures through shape optimization will eliminate the expensive process of optimizing these structures manually. Such efforts are particularly critical for the design of future generations of accelerators and will require a concerted multidisciplinary effort in the areas of applied mathematics, computer science, and accelerator scientists to successfully develop tools that can be used to tackle challenging accelerator problems of national scientific significance clearly related to DOE missions.

5.1.4 Multiobjective Optimization

To determine an optimal set of machine control parameters would require exploring a wide range of the parameter or design space (x 100 parameters), which becomes prohibitive without using parallel computing facilities. We need simulation-based parallel optimization techniques to determine an optimal set of accelerator beam parameters, i.e., improving the theoretical. Developing a parallel optimization capability integrated in parallel beam dynamics simulation codes will help accelerator scientists to explore a wider range of operational space for optimal operation of current and future accelerators. This tool will also enable accelerator scientists to design next generation accelerators more effectively by shortening the design cycle needed to determine optimal parameters.

5.1.5 Load Balancing at Extreme Scale

Computational domain needs to be partitioned into a large number of pieces for simulations on extreme-scale architectures. Any small imbalance is expected to lead to severe scalability issues, much more sensitive than current computing status. There is a need to develop novel schemes to balance loads and to maintain communication volumes among hundreds of thousands of CPUs, which are expected on future generations of HPC architectures. New approaches exploit multi-core parallelism and provide efficient multilevel, potentially asynchronous load balancing. Significant rethinking of programming models and/or many established algorithms (not just for load balancing) may be necessary. Load balancing is crucial for the scalability PIC methods and Electromagnetic/PIC coupled approaches.

5.1.6 Massive Data Analysis at Extreme Scale

Current simulation of a cryomodule with a centimeter size bunch and a mesh of 80 million elements and 500 million DOFs generated more than 5 terabytes of data for a time-domain run of 4000 time steps. The size of data is expected to exceed by 10^6 orders of magnitude or more for extreme-scale problems. It is important to develop exploratory algorithms to examine the entire physical domain and display some of the interesting region from the large dataset. To facilitate this exploration, researchers require easy-to-use tools that include techniques for accessing and manipulating remote data.

5.1.7 Framework Development

Currently available multi-physics, multi-scale frameworks depend on a variety of libraries for the interface glue and component definitions. The current state of support at supercomputing centers makes porting and deployment of these codes very difficult and taxing for computational accelerator science researchers. The difficulty will increase as the complexity of these frameworks increases. We will need to define a set of requirements to enable development and deployment of such frameworks on the future extreme scale supercomputers. Extreme-scale scientific computing requires robust solutions to current challenges in the development and deployment of software, including the following:

- Large-scale multiphysics simulations and simulation-based optimization problems are based on multiple libraries or packages, each with a distinct build system.
- Maintaining portability across platforms is difficult because of the lack of robust tools for cross-platform package management.
- The lack of a standard deployment model for scientific applications impedes reuse of existing software.
- Existing tools do not provide adequate cross-compilation support on new parallel platforms.

To understand and improve the performance of large-scale coupled applications, we need scalable and usable tools for performance analysis and tuning, including performance data gathering, performance bottleneck detection and analysis, and tuning for different architecture without sacrificing application maintainability and portability (with a significantly higher degree of automation than is currently available).

5.1.8 Optimization of File I/O

High-fidelity simulations involving particles, fields, or a combination of particles and fields require processing and analyzing many terabytes of data per application run. Analysis and visualization of these data will require major changes in our current file I/O paradigm. We will need to develop the software infrastructure for analysis and visualization on the fly, both for physics analysis and checking, in addition to enhancing the performance of the post-processing and postmortem analysis tools used at the petascale level. It will be imperative to identify the problems and define the strategies to overcome I/O bottlenecks in applications running on many thousands to many hundreds of thousands of processors. In addition, a common data representation and well-defined interfaces are required to enable analysis and visualization.

5.1.9 Single-Node Performance

Optimizing single-node performance of PIC (and other) methods will also have broad impact. PIC optimization is challenging because of the mismatch of the particle structures and the field data structures. We will need to come up with strategies that minimize the effects of this misalignment, beyond what has already been done in this community, such as particle sorting. Of particular interest would be studies of different caching strategies and exploitation of the available vector instruction sets on the processors. Source code transformation strategies to generate tuned code (both data structures and computation) can enable automated tuning of the implementations without negatively impacting performance portability.

5.1.10 Fault Tolerance

Fault-tolerant algorithms and operating system support are of great importance in using extreme-scale architectures. It is no to be expected that the mean time between failure will be increase in the same order of the number of cores will, hence

- progress (task) migration
- detecting and recovering from soft errors (e.g., because of non-error correcting code memory) will be a major issue.

5.1.11 Other Areas

Several beam dynamics codes use an electrostatic particle-in-cell model. Research is needed to improve the performance of the different components of the codes and determining the best strategy for the different applications:

- Depending on the physics of the problem, the codes might use domain decomposition, particle decomposition, or hybrid decomposition. There may be communication of particle data, grid data, or both. Particle movement between Poisson solves may be slight or large, hence, some codes use a particle manager and some do not use a particle manager. Solvers. The codes use spectral based, finite-difference based, and hybrid discretizations with fast Fourier transform and iterative solvers (e.g., conjugate gradient, multi-grid).
- Identifying and resolving bottlenecks for pipelining implementations to allow scaling to many thousands of processors. Pipelining is also very important for the multiphysics applications, as it will be the best path toward good weak scaling.

Optimization of single-node performance of PIC (and other) methods will also have broad impact. PIC optimization is challenging because of the mismatch of the particle structures and the field data structures. We will need to come up with strategies that minimize the effects of this misalignment, beyond what has already been done in this community, such as particle sorting. Of particular interest would be studies of different caching strategies. Source code transformation strategies to generate tuned code (both data structures and computation) can enable automated tuning of the implementations without negatively affecting performance portability.

Emerging capabilities in finite-difference electromagnetic modeling include new techniques for calculating frequencies in the time domain. Improving these techniques for the extreme scale requires: improving the embedded boundary method to work with higher-order finite difference algorithms (for instance, fourth-order algorithms), improving PIC charge conservation properties for higher-order algorithms, and improving the dispersion properties of higher-order algorithms. Current hardware trends point towards heterogeneous processor architectures at the exascale, consisting of some general purpose processor supplemented with some accelerator hardware.

Exploiting such architectures will require modifications of the algorithms currently in use for accelerator modeling. Research areas include: sparse solvers on heterogeneous architectures, particle algorithms, especially scatter and gather operations in the absence (or at high cost) of atomic memory operations, characterization of the benefit and accuracy of mixed precision calculations, and portability among different architectures.

ASTROPHYSICS DATA HANDLING, ARCHIVING, AND MINING

Chair: Alex Szalay, Johns Hopkins University

1. INTRODUCTION: DATA INTENSIVE SCALABLE COMPUTING

The nature of high-performance computing (HPC) is rapidly changing. While a few years ago much of high-end computing involved maximizing central processing unit (CPU) cycles per second allocated for a given problem, today it revolves around performing computations over large data sets. This means that efficient data access from disks and data movement across servers is an essential part of the computation. Data sets are doubling every year, growing slightly faster than Moore's Law. This is not an accident. It reflects the fact that scientists are spending an approximately constant budget on more capable computational facilities and disks whose sizes have doubled annually for over a decade. The doubling of storage and associated data are changing the scientific process itself, leading to the emergence of eScience—as stated by Gray's Fourth Paradigm of Science based on Data Analytics (Szalay and Gray 2006).

Much data are observational because of the rapid emergence of successive generations of inexpensive electronic sensors. At the same time, large numerical simulations are also generating data sets with increasing resolutions, both in the spatial and temporal sense. These data sets are typically tens to hundreds of terabytes. The scientific community is in a dire need of a scalable solution for data-intensive computing. Data sets in high energy physics have already exceeded a petabyte (Becla and Wang 2005), with much more to come when the Large Hadron Collider accelerator begins operating. Astrophysics is going to cross the petabyte boundary over the next 12 months (Institute for Astronomy 2005). The commercial world has clearly recognized the need for processing and analyzing large amounts of data. Several of the world's largest companies are setting up their own facilities to store and analyze many petabytes of data.

1.1. *The Next Challenge: Massive Data*

Because of these data-intensive scientific problems, a new challenge is emerging as many groups in science, and also beyond, are facing analyses of data sets in tens of terabytes, eventually extending to a petabyte because disk access and data-rates have not grown with their size. There is no magic way to manage and analyze such data sets today. The problem exists both on the hardware and the software levels. The requirements for the data analysis environment are:

- scalability, including the ability to evolve over a long period
- performance, both for streaming and random access
- ease of use, including ease of efficient programming, multi-purpose data representations and interactive sharing of massive data
- fault tolerance, including efficient tolerance of failure during computation and high integrity and persistence of valuable data
- cost-effectiveness, including continued storage density improvements and memory hierarchies that decouple application performance from the least cost storage media.

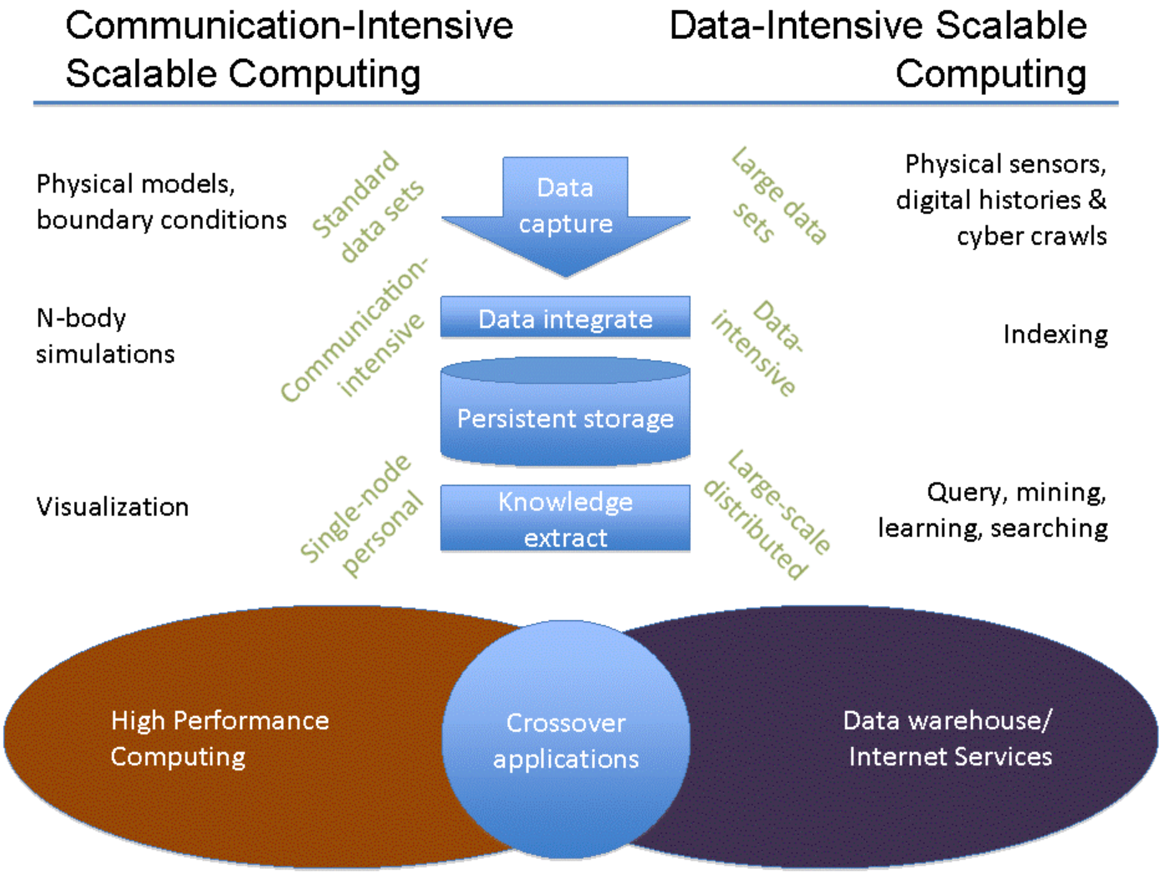


Figure 19. Data-Intensive Scalable Computing (DISC) differs from the Communication-Intensive Scalable Computing (CISC) architectures common in today's high-end computing. Data-intensive systems process changing datasets that are too large to move practically between computers. They have design parameters driven by the character and usage of the data rather than the computation. DISC systems assign relatively low importance to fast inter-processor communication. Data-intensive systems do not lend themselves to batch scheduling, common in today's high-end computing; rather they are characterized by interactive post-processing. Image courtesy of Alex Szalay (Johns Hopkins University).

1.2. *Massive Data Analysis is Different From Conventional Supercomputing*

The applications that drive today's supercomputer designs are inter-process-intensive. Distributed memory simulation begins with relatively small inputs and executes synchronized N-body models of complex processes, generating periodic checkpoints of systems state for visualization and fault tolerance. In contrast, the analysis of retained states for structure or abnormality, shares more with the characteristics of observation. Inputs are huge, computation is less intensive, inter-process communication rare, but data-driven storage access is common. Data-intensive computing uses large computing clusters to extract knowledge from data stores far larger than the combined memory size of any communication-intensive computer.

1.3. *Bring Analysis to the Data, Not Vice-Versa*

Many of the typical data access patterns in science require a first, rapid pass through the data, with relatively few CPU cycles carried out on each byte. These involve filtering by a simple search pattern or computing a statistical aggregate, very much in the spirit of a simple mapping step of MapReduce (Dean and Ghemawat 2004). Such operations are naturally performed within a relational database and expressed in Structured Query Language. So a traditional relational database fits these patterns extremely well.

The picture gets a little more complicated when one needs to run a more complex algorithm on the data, not necessarily easily expressed in a declarative language. Examples of such applications can include complex geospatial queries, processing time series data, or running the Basic Local Alignment Search Tool algorithm for gene sequence matching. The traditional approach of bringing the data to where there is an analysis facility is inherently not scalable, once the data sizes exceed a terabyte because of network bandwidth, latency, and cost. It has been suggested (Szalay and Gray 2006) that the best approach is to bring the analysis to the data.

1.4. *Building Balanced Systems: Amdahl's Laws*

Amdahl has established several laws for building a balanced computer system (Amdahl 2007). These were reviewed recently (Bell et al. 2006, Szalay et al. 2009) in the context of the explosion of data. Contemporary computer systems' input-output subsystems are lagging CPU cycles. In the discussion below, we will be concerned with two of Amdahl's laws:

A balanced system

- needs one bit/s of I/O for each CPU instruction/s
- has 1 byte of memory for each CPU instruction/s

These laws enumerate a rather obvious statement: To perform continued generic computations, data are delivered to the CPU through the memory. Amdahl observed that these ratios need to be close to unity and this need has stayed relatively constant (Table 2).

Table 2. The two Amdahl numbers characterizing a balanced system are shown for a variety of systems commonly used in scientific computing today. Amdahl numbers close to 1 indicate a balanced architecture.

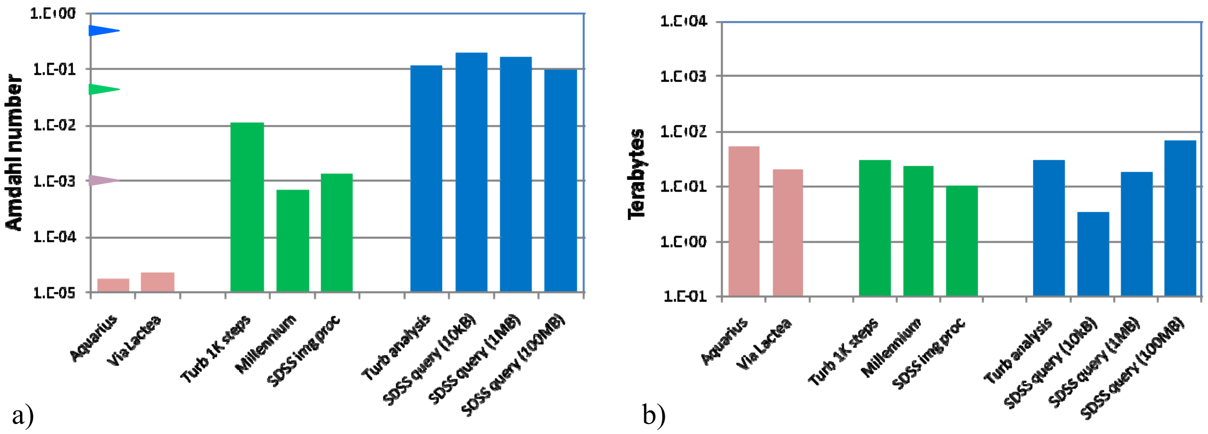
System	CPU count	GIPS [GHz]	RAM [GB]	diskI/O [MB/s]	Amdahl	
					RAM	IO
BeoWulf	100	300	200	3000	0.67	0.080
Desktop	2	6	4	150	0.67	0.200
Cloud VM	1	3	4	30	1.33	0.080
SC1	212992	150000	18600	16900	0.12	0.001
SC2	2090	5000	8260	4700	1.65	0.008
GrayWulf	416	1107	1152	70000	1.04	0.506

The emergence of multi-level caching led to several papers pointing out that a much lower I/O to million instructions per second ratio coupled with a large enough memory can still provide a satisfactory performance (Hsu and Smith 2003). While this is true for problems that mostly fit in memory, it fails to extend to computations that need to process so much data (petabytes) that they must reside on external disk storage. At that point, having a fast memory cache is not much help because the bottleneck is disk IO.

It is revealing to consider some recent simulations and data analysis projects. Figure 20 shows the Amdahl number, the data in bits divided by the number of CPU cycles, for a set of different projects and different hardware systems. The small, purple arrows on the left indicate typical supercomputer configurations ranging in an Amdahl number from 0.001 (BlueGene) to 0.01 (XT-3). The green arrow shows the Amdahl number of a typical BeoWulf cluster (0.08), while the blue arrow corresponds to the GrayWulf, our I/O oriented database cluster built at Johns Hopkins University (0.5).

Aquarius (Springel et al. 2008) GHALO (Diemand et al. 2008), and Via Lactea-II (Diemand et al. 2007a, Kuhlen et al. 2007, Kuhlen et al. 2008) are large cosmology simulations with several 10 billion particles, generated using millions of hours of CPU time on supercomputers. The BeoWulf-based computations are shown in green, with different numbers of snapshots in the output (128 for Millennium, 10 for Turbulence, 1024 for Turbulence 1K). SDSS represents the pipeline processing of the Sloan Digital Sky Survey images, resulting in the automated detection and measurement of 500 million astronomical objects. The last few columns show the actual and projected analysis usage of these large data sets at the Johns Hopkins' GrayWulf facility, the Turbulence 1K, and estimates derived from the actual SDSS query logs, corresponding to three different query sizes. It is important to note (Figure 20b) that all the relevant data sets are between 10 and 100 terabytes because of disk space and data transfer limitations.

There are some interesting things to note: The supercomputing generated projects have Amdahl numbers considerably below what the mainframe hardware could deliver. The main reason is the limitations on the data set sizes. If the simulations could have written a 500 terabyte output instead of 20 to 50 terebytes, their Amdahl number would have been 10 to 20 times larger, and the much higher Amdahl numbers on the right panel represent the quite different needs of the data analysis phase. It is clear that these require a different balance of CPU and I/O than the main computational facility.



a)

b)

Figure 20. The left panel displays the Amdahl number (data size in bits divided by the CPU cycles) for several large data sets. The pink color denotes computations on a supercomputer, the green shows data created on a BeoWulf cluster and the blue is posterior analyses done on database servers (like the GrayWulf). The right hand panel indicates how all data sizes fall between 10 and 100 terabytes. Image courtesy of Alex Szalay (Johns Hopkins University).

1.5. *Raw Sequential I/O*

For very large data sets, the only way we can even hope to accomplish the analysis if we follow a maximally sequential read pattern. Over the past 10 years, while disk sizes have increased by a factor of

1,000, the rotation speed of large disks used in disk arrays has only changed a factor of 2 from 5,400 rpm to 10,000 rpm. Thus random access times of disks have only improved about 7% per year.

The sequential I/O rate has grown somewhat faster as the density of the disks has increased by the square root of disk capacity. For commodity Serial ATA (Advanced Technology Attachment) drives, the sequential I/O performance is typically 60 megabyte/sec, compared with 20 megabyte/sec 10 years ago. Nevertheless, compared to the increase of the data volumes and the CPU speedups, this increase is not fast enough to conduct business as usual. Just loading a terabyte at this rate takes 4.5 hours. Given this sequential bottleneck, the only way to increase the disk throughput of the system is to add more disk drives and to eliminate obvious bottlenecks in the rest of the system.

1.6. *Scale-Up or Scale-Out?*

Today's supercomputers typically dedicate a few percent of the nodes as I/O nodes, restricting the bisection bandwidth between compute and storage to a small fraction of the computer's capability. Data-intensive scalable computers are primarily concerned with storage processing and need much higher storage bisection bandwidth. Therefore, traditional models attaching a few Fibre Channel disk arrays to a few nodes and requiring all data to flow through these nodes before any processing is ineffective. Storage needs to be more widely distributed in data-intensive scalable computers.

Today's data-intensive systems emphasize replicated commodity components modeled on the success of BeoWulf and Google clusters. Where simple filtering dominates data processing, for example, a cluster of inexpensive nodes with a few inexpensive disks each can deliver surprisingly high data rates inexpensively. Understanding the limitations of such distributed storage architectures for future generations of data-intensive applications is an important goal for data-intensive systems (Riedel et al. 1998) as well as understanding the complexities of data partitioning and management. Moreover, physical space efficiency and power efficiency may benefit from a deeper exploration of storage-in-node architectures exemplified by BeoWulf and Google experience.

1.7. *The Challenges of Building a Data-Intensive Exascale System*

Adding more disk spindles per CPU is not the only change required. Extremely large systems come with many other challenges not seen in smaller scale systems. To address these challenges, the software dealing with data-intensive applications must be rethought from ground up. The challenges are presented in the following paragraphs.

- Unavoidably, petabytes require thousands of disks. Even though a mean-time-to-failure of each individual disk might be measured in tens of years, when combined, the system will experience hardware failures daily, or often hourly. Graceful, transparent recovery from such failures is therefore a must to ensure uninterrupted job execution. Data-intensive system software will need to meet its performance requirements while simultaneously recovering from many concurrent multi-terabyte disk failures – this implies a revolution in the software design of the data storage system.
- At extremely large scales, administrative costs that increase with the size of the data are unsustainable. Systems must be self-sustaining, self-healing, self-load-balancing and self-adjusting to avoid requiring a large number of administrators.

- A large fraction of the computational challenges in astronomy in the next decade will be driven by complex analytics. An example of a modern analytics can be time series analysis, or complex spatial analysis involving densities. Such analysis often requires multi-joins of entire data sets, or cross-correlating entire catalogs. Performing such operations at extreme scale requires new approaches to analytics tuned for the scale.
- Disk I/O dominates the cost in data intensive systems. For this reason, aggressive data compression will start playing an increasing important role and will serve as a way to reduce I/O and use frequently under-utilized CPUs. Given that scientific data sets are almost always write-once-read many, compression techniques such as delta compression will likely become increasing important.
- In data intensive systems, where I/O cost dominates and compute cycles are cheap, it is often the case that reconstructing intermediate data is preferred over storing it. To perform this reconstruction, accurate recording and management of the lineage and provenance of each data item is required. Capturing provenance in large scale systems is also essential in order to capture entire lineage of produced data.
- At the extreme scale, it is essential to move computation to the data. This easy-to-understand paradigm often requires major rethinking of the entire architecture and must be combined with parallelization. Scanning a single petabyte of data at a speed delivered by a single disk would take several years, thus reading and processing data in parallel becomes de facto a must.
- At the extreme scale, it becomes worthwhile considering relaxing the accuracy of the answer - often obtaining a 100% precise answer can cost several orders of magnitude more than obtaining a 99.9% precise one. Taking into the account such precision, as well as accuracy of data measurements, can lead to huge performance gains and cost reduction. Lossy compression techniques can similarly reduce storage and I/O bandwidth costs.
- System flexibility is becoming increasing important. At smaller scales, it was relatively easy to change the schema of the underlying data. At extreme scale, a system must be flexible to allow schema changes without too much cost.

2. SCIENCE GRAND CHALLENGES

Several complementary probes of dark matter and dark energy are planned with the next generation surveys. Stage III surveys such as Dark Energy Survey (DES), PS1, and Baryon Oscillation Spectroscopic Survey (BOSS) will be carried out in the next five years or so, while the more ambitious Stage IV surveys (Large Synoptic Survey Telescope, Joint Dark Energy Commission, Square Kilometre Array) will be completed in the 10-plus year timescale. The science goals of these and other surveys can be roughly categorized as:

- cosmic microwave background (CMB)
- extreme environments
- large-scale structure (LSS), galaxy clusters and baryon acoustic oscillations (BAO)
- weak gravitational lensing.

2.1. *CMB Temperature and Polarization*

The computational challenges posed by the analysis of a CMB data set can be quantified by its numbers of observations in the time domain (N_t) and in the pixel (map) domain (N_p). The first is set by the duration of the mission and the numbers and sampling rates of its detectors at each of its observing frequencies, and largely drives the computational requirements of the data analysis. The second is set by the angular resolution of the detectors and the fractional sky coverage of the mission and primarily affects the communication requirements. The ongoing quest to measure ever fainter signals at ever higher angular resolution has driven a steady and continuing increase in both of these parameters.

2.2. *Current Satellite Missions*

The Wilkinson Microwave Anisotrophy Probe is a National Aeronautics and Space Administration (NASA) satellite mission that has been surveying the full sky with 20 detectors at five frequencies at moderate angular resolution. Launched in 2001, after its scheduled nine years of operation, the Wilkinson Microwave Anisotrophy Probe will have gathered 2 x 10^{11} time samples and mapped them into 5 x 10^7 sky pixels. Planck is a joint European Space Agency/NASA satellite mission to survey the full sky with 74 detectors at nine frequencies at high angular resolution. Scheduled to launch in spring 2009, over its two-year mission Planck will gather 6 x 10^{11} samples that will be mapped into 8 x 10^8 pixels.

2.3. *Current/Future Suborbital Missions*

With Planck set to make the definitive measurement of the CMB radiation temperature anisotropies, the focus of suborbital missions has shifted to CMB polarization, and especially the lensing-induced B-mode polarization signal at small angular scales. Since they can only observe part of the sky, suborbital experiments have much lower pixel counts. Examples of these missions include the E and B Experiment long-duration balloon experiment, intending to fly in Antarctica in late 2010 and gather 3 x 10^{11} samples with 1,400 detectors at three frequencies, and the Polarization of Background Radiation (PolarBear) and Q/U Imaging ExperimenT ground-based experiments to be deployed in the Atacama Desert in the next two to three years and each use several thousand detectors to gather 3 x 10^{13} and 10^{14} samples, respectively.

2.4. *Future Satellite Missions*

The large-scale CMB B-mode signal predicted to be induced by gravity waves from the inflationary epoch will require an exquisitely sensitive all-sky mission such as the CMBpol satellite called for in NASA's "Beyond Einstein" program. Although this will likely not launch before the end of the next decade, mission concept studies typically envisage several thousand detectors observing at a large number of frequencies for two years or more, gathering a data set of a few x 10^{14} samples over a few x 10^9 pixels.

2.5. *Computational Challenges in CMB Data Analysis*

Since CMB temperature and polarization signals are so faint, their precise measurement requires 10^3 to 10^5 observations of each point on the sky, so the analysis of a CMB data set is dominated by operations on the N_t time samples. Each such sample includes CMB signal, foreground contamination, and instrument noise, and these signals are correlated in the multipole-, pixel- and time-domains respectively.

Because we have to account precisely for each of these correlations (and indeed the CMB correlations are the fundamental measure of the data we are seeking), the entire data set has to be treated as a single data object, precluding the kind of divide-and-conquer analysis approaches used in other domains.

The most computationally challenging operation on a CMB data set is the mapping of the time-ordered data to the sky pixels, and in particular properly accounting for the temporal correlations in the noise. Using preconditioned conjugate gradient techniques, and exploiting the block band Toeplitz structure of the time-time noise correlation matrix, solving for the maximum likelihood map scales as $O(N_t)$ with a pre-factor that includes the number of iterations (typically around 50 for a white-noise pre-conditioner). Such analyses are not just used to make maps of the actual data, but also of the tens of thousands of simulated data sets required both for mission design and deployment and for the estimation of the CMB power spectra and their uncertainties using Markov chain Monte Carlo (MCMC) techniques. Finally, these analyses are inevitably iterative, with each analysis uncovering systematic errors in the model of the data and/or the instrument that have to be accounted for in the subsequent re-analysis; the need for these iterative Monte Carlo techniques will be even greater with the increased sensitivity to systematic effects in the forthcoming generation of multiplexed polarization experiments. Overall, we can expect a combined iteration/realization/reanalysis pre-factor of $O(10^5)$.

Over the next 15 years, we expect the size of CMB time-ordered data to grow by three orders of magnitude; coincidentally, this matches the projected growth in computing power over the same period assuming a continuation of Moore’s Law. Since today’s CMB data analyses are already pushing the limits of current HPC systems, this implies that our algorithms and their implementations will have to continue scaling on the leading edge of HPC technology for the next 10 Moore-foldings if we are to be able to support first the design and deployment of these missions and then the scientific exploitation of the data sets they gather.

2.6. *Extreme Environments*

A number of facilities over the coming decade and beyond are targeting extreme environment physics—black holes, pulsars, relativistic outflows, gamma-ray bursts, the origin of the highest energy cosmic rays, sources of gravitational waves, and cosmic neutrinos. The key current and medium term future facilities are probably the Chandra X-ray Observatory, the XMM-Newton Science Operations Centre, the Fermi National Accelerator Laboratory (FNAL), and the Pierre Auger Observatory; the most important future facilities under development are the International X-ray Observatory, Square Kilometre Array (SKA, radio), Laser Interferomter Space Antenna (LISA, gravitational waves), and IceCube Neutrino Observatory. Many other facilities with a time monitoring capability will play a role—notably the Panoramic Survey Telescope And Rapid Response System (Pan-STARRS) and the Large Synoptic Survey Telescope (LSST), but also many smaller telescopes with an alert-follow-up capability. As well as the intrinsic astrophysical topics, these studies will have direct bearing on fundamental physics — testing strong field general relativity, testing our understanding of black holes, making the first measurement of gravitational waves, and probing neutrino physics.

These projects all have significant data volume and processing requirements, which add to the general points made elsewhere in the report. However, the special feature in this area is the fact that the sky changes on every timescale from milliseconds to decades. There is a need to detect and recognize transients; to separate the rare interesting ones from the millions of daily false alarms (flaring M stars, moving asteroids, unexpected data “junk” etc.); to mine the evolving database for more subtle variables; and to construct time histories post hoc as required. These problems are with us now but become far more challenging in very high rate data flows, partly because of the processing requirement and partly because the contamination problems become much more severe with deeper data.

The prime need is for development of fast automated algorithms for detection and classification of transients. Also needed are improved methods for trawling the very large databases that will result from multiple time slices of the sky, with searches being run by many simultaneous science users, often returning to raw pixel data. This will probably require new, more efficient techniques for structuring and searching incremental databases, which will find commonalities with many other areas of science and commerce.

2.7. *Baryonic Acoustic Oscillations*

Oscillations of the coupled photon-baryon fluid in the early universe imprint a scale on the clustering of matter. This BAO scale, set by the sound horizon scale at the epoch of recombination, can be calculated from straightforward physics and calibrated by its imprint in the CMB. In the galaxy correlation function, the BAO signature is a sharp local peak at 150 Mpc. In the galaxy power spectrum, the Fourier transform of this peak appears as a series of oscillations, analogous to but much smaller than the acoustic oscillations in the CMB power spectrum. Measuring the BAO scale from galaxy clustering in the transverse and line-of-sight directions yields estimates of the angular diameter distance $D_A(z)$ and Hubble parameter $H(z)$, respectively. Determining these quantities with high precision requires enormous survey volumes.

BAO measurements can be carried out from both spectroscopic and multi-color imaging surveys. For the latter, accurate photometric redshifts must be estimated from multi-color information. This is an inversion of an inherently noisy system (from measured to physical properties) that results in non-Gaussian errors. Propagation of the full covariances (as a function of redshift and spectral type) is required to fully model the uncertainties on the cosmological parameters. This requires new computational and algorithmic approaches to inverse processes and error propagation. The measurement and analysis issues for BAO are addressed below since they overlap significantly with large-scale structure.

2.8. *Galaxy Clusters and Large-Scale Structure*

While BAO measurements use the standard ruler test to probe dark energy, the full distribution of galaxy groups, clusters, and large-scale filaments constrains a wealth of information on dark energy and dark matter. The structure of the halos of galaxies and galaxy clusters provide particularly good tests of dark matter.

Galaxy clusters with masses in the range 10^{14} to 10^{15} solar masses are useful probes of dark matter and dark energy. Clusters are the largest virialized objects in the mass distribution and are a particularly tractable target for observations of structure and its evolution over cosmic time. They are observed at redshifts between $0<z<1$ and are sensitive to the geometry of the universe and the amplitude of mass fluctuations. For any set of cosmological parameters, the growth of cluster-sized dark matter halos as a function of redshift and mass can be precisely predicted from N-body simulations. Comparing these predictions to observations of the real universe provides constraints on cosmology. This depends critically on understanding the mapping between predicted properties such as halo mass and the observed properties of clusters, such as galaxy counts/velocities, X-ray emission, Sunyaev-Zeldovich flux decrement, or weak lensing shear.

Correlation function (or Fourier space spectra) measurements of the galaxy distribution are sensitive to cosmology on large-scales and on the physics of clusters and galaxies on small scales. With appropriate parameterization, information on the cosmology and small-scale physics can be extracted (see below).

2.9. *Weak Gravitational Lensing*

Upcoming major weak lensing surveys will be performed by Pan-STARRS, the DES, and ultimately by LSST from the ground and by Joint Dark Energy Mission (JDEM) from space. Data from these surveys will accumulate over the next 20 years and will dwarf past current efforts in the field by many orders of magnitude.

These surveys will use the coherent distortion of images of distant galaxies to determine the structure in the dark matter, which is determined by the early universe and also the growth of this structure that is influenced by the current equation of state of the background, i.e., the dark energy, and is also affected by possible interactions in the dark sector. Additional tests of cosmological theory come from the dependence of the geometric path lengths on redshift.

With high precision and high-accuracy measurements of cosmic shear, it will be possible to tightly constrain cosmological parameters; measure evolution of the equation of state; test gravity on large scales; provide clues to the nature of the dark matter through the effect of self-interactions or interactions between the dark matter and dark energy; and test theories of inflation from evolution of dark matter structures.

Weak lensing, like other precision cosmology probes, is limited by the understanding and correction of systematic errors. Weak lensing measurements require exquisite understanding of the point spread function (PSF) as the effects of any PSF anisotropy must be corrected for high precision. The PSF can be measured from stars, but the devil is in the details of how this depends on the spectra of the galaxies, which are different from the stars. These effects are the limiting factor in current ground-based surveys. Tackling this problem will require the development of techniques for characterizing and modeling of the instrumental PSF via interaction with data and simulations that incorporate propagation of radiation through the atmosphere (for ground-based surveys) and through the instrument optics and the detectors. The next generation of surveys will also require the development of techniques for simultaneous multi-image object detection and shape analysis. This poses a significant computational challenge as the upcoming surveys will be generating peta-pixels of image data per year.

In addition to the challenge of making accurate measurements, there are challenges in development of the theoretical predictions with which the data will be confronted. Linear theory is not adequate on the scales of interest, so simulations will be needed and techniques must be developed for propagation of light to generate synthetic images and synthetic shear catalogs; this will also require simulations that accurately model the sources. Generating and analyzing such high precision and high-dynamic range simulations spanning the range of theoretical parameter space promises to be a greater computational challenge than making the measurements. A further computational challenge is developing techniques for high dimensional model parameter fitting.

3. DATA ANALYSIS CHALLENGES

3.1. *Systematic Error Issues in Survey Data Analysis*

Several of the science challenges discussed above are currently dominated by systematic errors in the measurements. The challenge for ground-based astronomy is because of the inherently uncontrollable interference of the atmosphere. While ground-based facilities are improving their characterization of the physical instrumentation, the time-varying nature of the atmosphere makes precision measurements extremely difficult. The stochastic contributions from the atmosphere affect both the spectral response of the system, thus the photometry and also the PSF.

3.2. *Reducing Systematics by Improved PSF Measurements*

Perhaps the most critical source of systematic errors is the accuracy of the determination of the PSF for real images. Poor PSF models affect the quality of photometric, astrometric, and morphological measurements of real objects in astronomical images. Current state-of-the-art PSF modeling consists of empirical measurements from individual images using the available stars, coupled with modeling of the effects of the telescope optics and atmosphere to constrain the range of likely PSF models. The future missions will require more advanced techniques that will be extremely data intensive. One promising technique currently in consideration improves the PSF model of individual images and simultaneously measures the shapes and brightness of images by performing a single global solution to all of these parameters for a collection of overlapping images. To perform this type of analysis on a reasonable timescale (months), on the PS4/LSST-scale datasets, will require 1 to 10 petaflops operating on approximately100 petabytes of data.

3.3. *Precision Photometry and Modeling of the Atmosphere*

The photometric precision ground-based observations will continue to be lacking because of our limited knowledge of the photometric response of the atmosphere. Current techniques using the empirical measurements of the impact of the atmosphere are reaching the limit of their effectiveness. Future improvements will require coupling the onsite monitoring of the atmospheric observables (temperature, pressure, humidity, temperature vs. altitude, etc.) to detailed modeling of the atmosphere. Observatories are already forming partnerships with climate research groups; future efforts will require closer interaction between local climate modeling and measurements from astronomical observatories. Modeling efforts needed by the astronomical community will include both the modeling of the spectral response of the atmosphere given the observable constraints, but also constraints on the seeing and scattered light because of the atmosphere.

3.4. *Computational Challenges in Clustering Statistics and Parameter Estimation*

The computational problems in several of the large-scale science probes discussed above relate to the estimation of two-point and higher correlations, inverting their covariance matrices, and in estimating parameters from them. Upcoming surveys will provide rich information in the redshift direction that spans a large fraction of the horizon volume, as well as the two dimensions on the sky. We consider here the generic analysis of data from imaging surveys that provide photometric redshifts and angular positions for millions of galaxies (exceeding a billion galaxies for Stage IV surveys) over a tenth to half the sky.

In the five-year time-scale, Stage III imaging surveys will provide auto- and cross-spectra (at least five) for lensing and angular clustering. These power spectra can be binned in at least 10 bins each in redshift and angular scale. Counting all cross-spectra in redshift bins, we have at least 5 x 50 x 10, or approximately 2500 binned spectra. Thus the calculation of covariances is feasible as it requires inversion of matrices of order 10,000 x 10,000. However, with Stage IV surveys, we must consider finer measurements in both redshift and angular scale. This problem rapidly reaches exascale proportions: with 100 bins each in redshift and length scale, one has 5 x 50,000 x 100, or approximately 2.5 million binned spectra. So calculation and inversion of covariances, common to the analysis of all clustering probes, will require major algorithmic and computational advances. For higher-order correlations this problem is significantly harder.

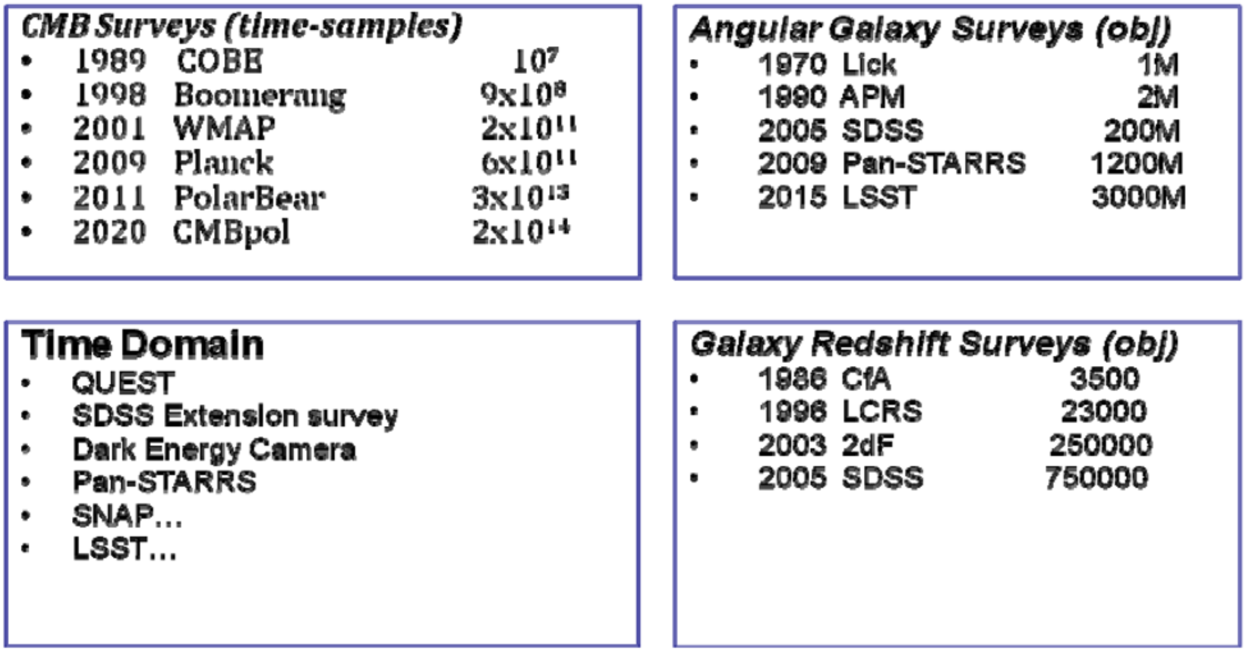
Beyond two-point statistics, higher order correlations scale naively as Nk where k is the order of the correlation function. For current data sets, 3-point statistics have been measured on 50-100,000 sources in 10,000 CPU hours (to estimate the covariance of these measures requires multiple realizations of these or simulated data sets). Scaling to the size of angular surveys in BAO and lensing surveys and to 4 and higher order statistics is out of the range of current computational facilities. To achieve these orders requires not just new computational resources but also the development of approximation techniques where the accuracy of the measure is limited by the desired accuracy on the physical cosmological rather than the underlying shot noise.

The simultaneous estimation of cosmological parameters, small-scale physics and systematic uncertainties in the measurement (such as shear calibration errors in lensing) or model (scale and redshift-dependent biases of galaxies) is essential in making rigorous dark energy measurements. So far, the estimation of systematic uncertainties has been done using approximate models, but upcoming surveys offer the opportunity to estimate them from the data with minimal assumptions. This however requires the ability to carry out the equivalent of MCMC parameter estimation over parameter spaces of very high dimension. Current analyses work with order five-dimensional parameter spaces. Even with Stage III surveys, the goal will be to fit for up to 10 cosmological parameters and order 100 “nuisance” parameters: a computational challenge that requires vast increases in computational power and new algorithms. The number of nuisance parameters would be even larger depending on one’s degree of conservatism and with Stage IV surveys.

4. EMERGING NEW SURVEYS AND COLLABORATIONS

Table 3 shows the trends in astrophysical data over the last few decades. The exponential growth in data is quite obvious. In the following section, we discuss in some detail the main data sets relevant to the current discussion that will emerge over the next decade and will provide unique science opportunities, but will also require extreme scale computational capabilities.

Table 3. Different astrophysical data sets and their cardinalities as a function of time. The exponential trends are apparent. Table courtesy of Alex Szalay (Johns Hopkins University).



CMB Surveys (time-samples)

1989	COBE	10^{7}
1998	Boomerang	9×10^{8}
2001	WMAP	2×10^{11}
2009	Planck	6×10^{11}
2011	PolarBear	3×10^{13}
2020	CMBpol	2×10^{14}

Angular Galaxy Surveys (obj)

1970	Lick	1M
1990	APM	2M
2005	SDSS	200M
2009	Pan-STARRS	1200M
2015	LSST	3000M

Time Domain

- QUEST
- SDSS Extension survey
- Dark Energy Camera
- Pan-STARRS
- SNAP…
- LSST…

Galaxy Redshift Surveys (obj)

1986	CfA	3500
1996	LCRS	23000
2003	2dF	250000
2005	SDSS	750000

4.1. *Upcoming Optical Surveys*

BOSS, a part of the Sloan Digital Sky Survey (SDSS-3), is a next generation spectroscopic redshift survey following up on the SDSS; it is intended to measure structure in the distribution of baryonic matter via direct redshift measurements of 1.5 million large red galaxies (LRG) to a redshift of 0.7 and via measurements of the “Lyman alpha forest” in front of 160,000 QSOs to obtain the structure between redshifts of 2.3 and 3.3. BOSS observations are expected to begin in mid 2009, following the completion of a new spectrograph for the APO/SDSS telescope. The survey will span 2009-2014. BOSS is an order of magnitude larger than existing surveys. The distribution of baryons, as seen in the LRG distribution, shows a feature in its “correlation function,” this sets a scale, which is directly correlated with the fluctuation power spectrum of the early universe. At low redshift, this scale was measured by SDSS and 2dF to be 110 Mpc. Changes in this scale as a function of redshift will directly show the growth of these fluctuations, leading to greatly improved estimates of the cosmological parameters.

The spectroscopic redshifts from BOSS will provide fundamental calibrators for the photometric redshifts used by imaging based dark energy experiments, such as the Panoramic Survey Telescope and Rapid Response System (PanSTARRS) and LSST.

4.1.1 The Pan-STARRS Project PS1 Survey

The Pan-STARRS Project PS1 survey is a single 1.8m telescope with a 7-degree field of view imaged by a 1.4 GPixel camera. This telescope will perform a near-all-sky survey covering a 30,000 degree field over 3.5 years, with 60 visits distributed over five filters *(grizy)*. Including data from several deep fields with frequent re-visits, the total raw data volume for this survey will be 1.5 petabytes. The Pan-STARRS Project longer-term goal is to replicate the PS1 optical system and camera multiple times. The project is on track to complete a four-telescope system (PS4) in the 2012 timeframe that will perform similar wide-area and deep surveys over a 10-year mission lifetime. The PS4 data volume will approach 20 petabytes and will result in approximately 500 repeated observations of the 30,000 degree region distributed across the five filters. The PS1 survey depths are *g*=24.1, r=23.6, i=23.5, z=22.5, y=21.0. The PS1 telescope is located on the summit of Haleakala on Maui and will begin survey operations in early 2009, while the PS4 telescope is planned for the summit of Mauna Kea on the island of Hawaii. The science goals of the survey cover the full range of science from cosmology to the structure of the solar system. Dark energy and dark matter studies will be enabled by the combinations of large-scale structure observations, weak-lensing observations on the large-area survey, and the observation of Type Ia supernovae.

4.1.2 The Dark Energy Survey

The Dark Energy Survey (DES) is a Type III dark energy survey comprising a wide field (5000 sq degrees) survey and narrow field (9 sq degree) time-domain survey in four passbands g, r, i and z passbands. With a 500 Mpixel camera, it will achieve a data rate approaching 1 terabyte per night and, over the 5 years of survey operations, will catalog and archive approximately 1 petabyte of data with a depth of g=24.6, r=24.1, i=24.3, and z=23.9. The science drivers are the study of dark energy and dark matter using four primary probes: the evolution of the galaxy cluster density with redshift (100,000 clusters to a redshift of z<1.3), Type Ia SNe as standard candles (1000 SNe at z~1), weak lensing at redshifts of z=1 and z=4 and BAO at z<1. Situated at the Blanco 4m telescope at Cerro Tololo Inter-American Observatory, operations are expected to begin in 2011 and run for 5 years.

4.1.3 The Large Synoptic Survey Telescope

The Large Synoptic Survey Telescope (LSST) is an 8.4-m telescope with a 3.5-degree field of view and a 3 Gigapixel camera that will generate a 6-band (0.32-1.1 microns) survey of 20,000 sq degrees of the sky over a period of 10 years. Returning to each patch of the sky approximately 1000 times in 10 years, it will reach a limiting magnitude of r wil be approximately 27.5. The data rate for the LSST will be approximately 30 terabytes of imagery per night that will be analyzed in real time and will result in an archive of 30 petabytes of data in the form of catalogs and 60 petabytes of imagery and 270 teraflops for archive operations (after 10 years). A Type IV dark energy survey the primary science goals of LSST are the study of dark energy and dark matter using gravitational weak lensing (with a density of 55 sources per square arcmin at a median redshift of z=1.2), Type Ia supernovae (10^6 SNe), BAO. The size of the LSST data set will, however, enable a broad range of ancillary science ranging from a census of the solar system, studies of Galactic structure and probes of the high energy transient universe. LSST is scheduled to begin early science operations in 2015 and enter its survey mode in 2016.

4.1.4 The Joint Dark Energy Mission

The Joint Dark Energy Mission (JDEM) is an Einstein probe that is being developed in a partnership between DOE and NASA. Currently in the process of defining a mission statement a number of proposals have been made for JDEM to study dark energy using Type Ia supernovae BAOs and gravitational weak lensing. These include a near-infrared slitless spectroscopic survey (targeting H-alpha) of 100 million galaxies at 1<z<2 covering 28,000 sq. degrees, optical and infra-red imaging over several thousand square degrees for weak lensing to a depth of approximately J=25 (30 galaxies per sq arcmin) and a slit or slitless spectroscopic survey of Type Ia SNe to z<1.7. The imaging data volumes and resulting archives present similar challenges to those from DES and LSST. As a space-based mission the telemetry data rate and on-board processing present significant computational and algorithmic challenges. JDEM is expected to begin operations around 2016

4.2. *Radio Surveys and the High Redshift Universe*

Two major international facilities—International X-ray Observatory and SKA—being developed over the next decade are aimed at bridging the gap between our knowledge of the early universe and the current epoch. The challenge is to find and study the earliest objects to condense out of the cosmic soup – first stars, protogalaxies and accreting black holes – and to map out the physical and magnetic structure and evolution of the intergalactic medium out of which these objects are forming. This will connect first principles physics to observable structures and so constrain that physics and the cosmic ingredient list – e.g., the type of dark matter and for the first time attempt to explain the origin of cosmic magnetism.

The biggest computational challenge comes from SKA, because of the requirement to collect a high bandwidth time series from each of many elements in the array and to produce a skymap by Fourier Transforming these "fringes." The challenge will build up gradually as SKA is being conceived as a sequence of pathfinders and intermediate projects (e.g., Low Frequency Array and PrepSKA) starting in 2009 and continuing through to 2021. By 2018, SKA estimates data volume of 100 petabytes/day and a processing requirement of 20 teraflops, figures that are only just plausible with Moore's Law type extraopolations. The processing requirement would be many times worse if, as is traditional in radio astronomy, users were given the facility to recompute on demand many different skymaps from the fringe data using different assumptions. It may even be more cost-effective to throw away the fringe data and re-observe the sky rather than re-compute. Therefore, there is an urgent need to understand how to make the FT calculations as fast as possible, either by algorithm or dedicated machine architecture, how to provide compute-on-demand, and where the cost-performance drivers are.

3 years	6 years	10 years
Bring more community codes to petaflop scale of 10^5 way parallelism for various statistical analysis scenarios	Provide better characterization of extreme scale programming models with simulators	Have data, tools, hardware and a community ready to exploit extreme-scale computing
Develop scalable analysis methods capable of using $10^{6\text{-}7}$ way parallelism	Implement $10^{6\text{-}7}$ way parallel analysis codes	
Develop best practices for petabyte-scale analysis systems, encourage diversity	Learn from best practices, develop, build and test solutions scalable to 10 petabytes	
Build actual implementations of existing petabyte-scale data sets with an interactive analysis workflow environment, and test usage	Build and deploy several geographically distributed, multi-petabyte-scale analysis workbenches	
Develop scalable statistical algorithms targeted for the upcoming astrophysical challenges (100D parameter estimation, clustering analysis)	Run these algorithms on existing extreme datasets	

5. PRIORITY RESEARCH DIRECTIONS/COMPUTATIONAL CHALLENGES

5.1. *Algorithms and Architectures*

Action Items

3 years	6 years	10 years
Expand DOE CSGF program. Engage Department of Education's Graduate Assistance in Area of National Need program and National Science Foundation in supporting graduate training grants for Computational Science	Develop Laboratory/University/ Industry partnerships for extreme scale era	Support a wide range of extreme-scale applications, tools
Expand Computational Science Summer schools with focus on extreme scale in participating labs	Develop National Extreme scale graduate program linking labs and universities	
(Re)start broad inter-agency program in Extreme scale computing similar to the High Performance Computing and Communications program of a decade ago	Build shared large DISC system(s) to be used for science analyses	

6. CROSS-CUTTING RESEARCH DIRECTIONS

6.1. *Developing a Common Data Access Platform: SCIDB*

Existing off-the-shelf data management systems no longer meet the extreme requirements of data-intensive scientific and industrial users. As a result, the majority of existing data intensive systems, both scientific and industrial, is relying on custom, home-grown solutions. Because of increasing complexity and cost, building such systems from the ground up per project is becoming impossible. Developing a common scalable data management system, defining common data models, standards and interfaces become essential. Involved tasks include

- exploring commonalities and improving collaboration
- understanding new technologies and emerging systems.

We should explore commonalities and differences between large-scale commercial systems and scientific systems from different domains. We should work closely with researchers and academics focusing of data memory systems issues, as well as with vendors including database vendors. This exploration and collaboration has already started through the series of Extremely Large Databases Workshop (SLAC 2007, SLAC 2008), organized by the Stanford Linear Accelerator Center.

We should understand and evaluate applicability of the new technologies and emerging systems, such as the SciDB project (Stonebraker et al. 2009). SciDB is a new project that aims to build a new open source science database with focus on complex analytics, scalability, new data model, improved interfaces, provenance, and data uncertainty.

6.2. *The Next Generation of Computational Scientists*

The success of computational science within DOE is critically dependent on the development of a workforce that can exploit the new generation of experiments and surveys. This new generation of scientists needs to be as comfortable in computer science, statistics, and applied mathematics as they are in physics or astrophysics. To accomplish this requires a concerted and sustained effort to develop and train a workforce that is computationally and mathematically literate. As an initial focus, graduate programs in cross-disciplinary research need to be created, supported, and expanded as well as the development of cross-disciplinary summer schools and workshops. This should be undertaken as a multi-agency initiative (including DOE, NASA, and the National Science Foundation) and should be in partnership with high-tech companies (Google, Microsoft, Yahoo, Amazon, eBay, etc.) that require a work force with data intensive and computer intensive skills.

It should be noted that this is not just a question of increasing the skills of the next generation of scientists. If they are taught to think about data and computing in new ways, they will likely uncover new physical relations within the data that were not envisaged when the experiments were defined. New discoveries often come at the edges.

An ongoing concern is the lack of career paths for scientists who work at the boundaries between fields. Departments are often reluctant to embrace scientists whose expertise is not easily compartmentalized (a physics department may view researchers as a computer scientists while a computer science department may regard them as statisticians and a statistics department may regard them as a physicists). Departments can be encouraged to view these computational cross-disciplinary scientists as integral to their departments if grant programs exist to explicitly support their research effort.

HIGH ENERGY THEORETICAL PHYSICS

Chair: Steve Sharpe, University of Washington

1. INTRODUCTION

The Standard Model of particle physics is one of the towering achievements of twentieth century science. It encompasses the strong, weak and electromagnetic interactions—three of the four established forces in nature—and has been enormously successful in explaining a wealth of data produced in accelerator and cosmic-ray experiments over the past 30 years. We now know that, down to a distance scale of at least 10^{-18} m, matter consists of point-like particles: quarks and leptons. Quarks come in three families: up-down, strange-charm, and bottom-top, each with their leptonic partners: electron, muon, and tau. In turn, each lepton has a neutrino partner. This family division is not perfect, for the weak interactions lead to mixing between families, encapsulated in the Cabbibo-Kobayashi-Maskawa (CKM) matrix for quarks, and a similar, though much less well-known, matrix for neutrinos. The success in explaining the properties of electroweak interactions in terms of the CKM matrix was rewarded with the 2008 Nobel Prize in physics.

Despite its enormous success to date, we know that the Standard Model is incomplete. We know this from several lines of evidence (DOE/NSF 2009). For example, although Standard Model particles compose the visible matter in the universe, this is but a small fraction of the total. The dominant component is dark matter, which we know cannot be composed of Standard Model particles. Another example concerns the violation of particle-antiparticle symmetry (C) and parity (P). Although both are violated in the Standard Model, this violation is too weak to explain the dominance of normal matter over anti-matter in the observed universe. There are also several fundamental issues that remain unanswered in the Standard Model, for example, the origin of mass (is it the Higgs particle?) and of the family structure itself.

The focus of present high energy physics (HEP) research at the U.S. Department of Energy (DOE) is the search for physics Beyond the Standard Model (BSM). This search is being undertaken at three frontiers: the Energy Frontier of the highest-energy accelerators, the Intensity Frontier at which one searches for extremely rare processes, and the Cosmic Frontier. The Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) will attain the energies needed to directly probe BSM physics, including the possibility of producing the particles that compose the dark matter. The next decade will likely produce revolutionary discoveries at all three frontiers.

The underlying mathematical language used to describe the particles and interactions that appear in the Standard Model is quantum field theory. Here the quantum variables are not the position and momentum of an atomic electron orbiting the nucleus of a hydrogen atom as in the original quantum theory of Schrödinger, but instead the amplitudes of continuously distributed *fields* such as the electric and magnetic fields. When such fields are described using quantum mechanics, the resulting quanta are the particles of which everything is made: the electron itself orbiting that hydrogen nucleus, the quarks confined inside that nucleus, the Z and W mesons correctly predicted by the Standard Model and even the elusive Higgs particle, which has yet to be discovered.

Numerical simulations of quantum field theories will play a crucial and increasing role in the search for new physics. This is because one sector of the Standard Model—quantum chromodynamics (QCD), the theory of strong interactions—as well as many of the theories proposed for BSM physics, cannot be studied using the analytic methods known today. In particular, perturbative expansions, based on the presence of a numerically small expansion parameter, are completely inadequate. Non-perturbative methods are needed, and, for most theories, the only known method is to study the theory using numerical simulations. For this, space-time must be replaced by a discrete lattice of points, so that this method is called "lattice field theory" (LFT). LFT, or more specifically lattice quantum chromodynamics (lattice QCD or LQCD), has been the dominant user of computational resources within theoretical high energy particle physics for the past three decades and is very likely to remain so for the foreseeable future.
The last decade has seen lattice simulations mature into a quantitative tool for calculations in QCD, setting the stage for a golden era of such calculations in QCD, as well as the extension of simulations to other theories. The theoretical HEP panel identified three priority research directions for numerical LFT in the upcoming decade:

- searching for physics BSM
- testing QCD at the sub-percent level
- simulating possible theories of physics BSM.

All three will be transformed by exascale computational resources. For the remainder of this section, we give a short overview of and motivation for each of these research directions, describe other possible directions in which exascale computing may also play an important role and provide a summary of our findings. Subsequent sections will provide a more detailed discussion of the challenges facing each direction and the role that exascale computing can play in overcoming these challenges.

1.1. *Searching for Physics Beyond the Standard Model*

The search for new physics at the intensity frontier requires performing ever more stringent tests of Standard Model predictions for rare processes. There are many rare processes in the decays of mesons containing strange, charm, and bottom quarks for which there are, or will be, very accurate experimental results. To determine whether these results are consistent with the Standard Model requires precise calculations in QCD, and for this numerical simulations of LQCD are needed. The ultimate aim is to reduce the error in the LQCD calculations to a level significantly below that in the corresponding experimental measurement. Searching for new physics in this indirect manner allows one to reach energy scales that are comparable to, or in some cases exceed, those that are directly accessible at the LHC. The required LQCD calculations can be viewed as the low-energy analogues of the perturbative QCD calculations needed to determine the Standard Model background rates at the Energy Frontier.

A related, though different, role that LQCD will play occurs once evidence for new physics is found at the Energy Frontier. This evidence will likely be consistent with many possible models, and the Intensity Frontier will play a crucial role in its interpretation. This is because new physics leads to contributions to rare Standard Model processes that must be consistent with experimental results. As for the Standard Model contributions themselves, one needs non-perturbative calculations in QCD to determine the contribution of a given model of new physics. Once again, LQCD is the only tool for many of these calculations.

Similar input will be needed in some searches for dark matter at the cosmic frontier. The coupling of candidate dark matter particles to ordinary matter depends on non-perturbative QCD physics. LQCD is thus needed as an integral component of direct dark-matter searches.

1.2. *Testing QCD at the Sub-Percent Level*

Two of the three components of the Standard Model have been verified with great precision. Tests of quantum electrodynamics attain the extraordinary precision of a few parts in 10^{10}, while the electroweak sector is tested at the part per 10^3 level. By contrast, QCD, the theory of the strong interactions, has been tested much less accurately. At both high energies, where one can use perturbative methods since the strong coupling constant is relatively small, and at low-energies, where one relies on non-perturbative LQCD simulations, verification at the 5% to 10% level is typical. A grand challenge for theoretical calculations is to improve this accuracy to the 1% level and below. This would be an extraordinary achievement—precise numerical control over a non-perturbative quantum field theory. It would not only solidify our confidence in the Standard Model, but also provide validation for the numerical methods of LQCD that will be used to search for new physics. Furthermore, the methods needed to obtain sub-percent level accuracy from LQCD calculations will, at least in part, carry over to simulations of other quantum field theories, such as candidates for BSM physics.

LQCD also plays an important, and growing, role in nuclear physics. The major applications are to the excited hadron spectrum, nucleon structure, and QCD at finite temperature and density. There is very extensive overlap in methods used for particle and nuclear applications, and a strong overlap of interests. This broad range of applications to nuclear physics provides further motivation for attaining highly accurate results from LQCD. These applications are discussed in a companion nuclear theory workshop.

1.3. *Simulating Theories of Physics Beyond the Standard Model*

The LHC will soon directly probe physics at the energy frontier. By the dawn of the era of extreme scale computing, the LHC will have run for several years and provided tantalizing evidence of physics BSM. Physicists exploring candidate models have developed an overwhelming array of possible theoretical models. To really understand the options and make a definite discrimination between experimental signatures requires investigations of lattice field theory beyond the realm of QCD (Fleming 2008). Two important classes of BSM theories are those involving supersymmetry and those that are nearly conformal (invariant under a rescaling of lengths). Computations in both classes of theories are significantly more challenging than in QCD, and there is, in addition, a large variety of models to search. The ultimate aim is to allow rapid screening of the viability of new models, or classes of models, as experimental hints from LHC and elsewhere accrue. When a particular model proves to be viable, one would then want to undertake more detailed calculations. In this way, the interplay between theory and experiment could allow one to zero in on the correct interpretation of BSM physics.

The discussion above targets specific questions that remain unanswered in present attempts to build a theory for the physics that must lie BSM. Here we hope to exploit known but perhaps not well understood theoretical mechanisms such as conformal symmetry or supersymmetry to solve the puzzle of constructing a viable BSM theory. However, the important ideas needed for this solution may lie outside what is presently known. It is important to recall that the mathematical phenomena of quark confinement and spontaneous chiral symmetry breaking were not deduced from the theory but instead induced (i.e., guessed) from experiment. Later LQCD calculations were needed to demonstrate that the theory indeed possessed these remarkable properties. While it is difficult to forecast, there is a real possibility that more open-ended studies of field theories beyond QCD will turn up even more amazing behaviors, one of which might be just the key needed 10 years from now to unlock the mysteries being recognized in LHC data. There is also great promise in a direct study of the next level of quantum theory: a theory of strings instead of particles. Such a direction may both provide important tests of the many conjectures that make up string theory and also permit the more fundamental character of string theory to guide the choice of the field theory most likely to describe physics BSM. While such truly exploratory work is not possible with today's computers, it is a worthy target for exascale computing.

1.4. *Additional Directions in High Energy Theoretical Physics*

This panel report focuses on the three priority research directions introduced above. However, it is important to recognize other areas in theoretical high energy physics where computation plays a critical role and where exascale computing offers exciting possibilities. The first is the computation of complex Feynman amplitudes needed to understand the effects of QCD in high energy scattering (Bern et al. 2007) and (Berger et al. 2009). This field has experienced remarkable progress over that last 10 years as deep mathematical insights have allowed previously intractable problems to be addressed. A similar, rapidly developing area is the use of light-front methods to describe QCD (de Teramond and Brodsky 2009), (Vary et al. 2008), and (Brodsky et al. 1998). While this approach addresses low-energy, non-perturbative phenomena, as does LQCD, the results of the light-front approach are more directly useful to high-energy scattering. Recent efforts to exploit the anti-de-Sitter space/conformal field theory correspondence (AdS/CFT) between QCD with a large number of colors and anti-de Sitter space promise an appealing basis for a first approximation in such light-front methods. Present research efforts in both the complex Feynman amplitude and light-front areas are directed at the theoretical foundations and the development of necessary tools. These two areas were the topic of special presentations to the High Energy Theoretical Physics panel made by Lance Dixon and Stanley Brodsky, respectively. Only modest (approximately 10 to 100 gigaflops) computational resources presently are used. This makes it difficult to forecast whether exascale computational resources will be essential for future work in either area. However, it is important to recognize that these and possibly other directions may well develop to a point where extreme scale computing could have an enormous, unanticipated effect on high energy physics.

1.5. *Summary of Panel Findings*

As presented in detail below, extreme scale computing will allow truly transformational advances to be made in each of these three priority research directions. The first two directions, searching for physics BSM and testing QCD at the sub percent level are well developed topics. Many important quantities have now been computed with accuracy well below 10%, and we can confidently predict the computing resources that will be needed to reach the important goal of accuracy at or below 1%. Since QCD is a fundamental theory there is no intrinsic limitation to the accuracy that can be achieved with sufficient computer resources. However, the scaling of the computer resources required for improving accuracy is indeed daunting. For example, even the most accessible quantities require computer resources growing as the sixth or seventh power of the inverse lattice spacing so that reducing finite grid size errors requires nearly exponential growth in computer power.

Fortunately Moore's law provides this exponential growth and the present quantitative success LQCD is directly driven by the seven orders of magnitude growth which started as sustained 1 megaflops performance on the DEC VAX computers of the early 1980s and has evolved to the 10 teraflops now sustained on the BG/P machine at the Argonne Leadership Computing Facility. It is precisely the two or three orders of magnitude further advance to extreme scale computing that is needed to drive the transformational changes in LQCD discussed here.

The third priority research direction, simulating theories BSM is more speculative and possibly even more exciting. This attempt to use the techniques of lattice field theory to explore conformal and supersymmetric field theory is just beginning. The search for new non-perturbative phenomena in non-QCD theories has barely been undertaken. This makes precise forecasting of the benefits of extreme scale computing difficult. However, present experience suggests that even the most accessible questions require terascale (sustained) resources to answer, for a single candidate theory. Providing computational resources on the order of hundreds of sustained teraflops to each of many aspects of this research direction would permit an appropriately broad range of theories to be studied. Added resources would be needed to track down and understand unexpected difficulties. Very important for this exploratory work is the need for rapid results. Such revolutionary use of large scale computation as a tool to support broad-ranging, speculative theoretical exploration will only be effective if questions can be answered in days or hours, not in the present many months that characterize state-of-the-art research in LQCD. This will only be possible with extreme scale computing hardware and substantially improved software resources.

A careful discussion of the required hardware and software infrastructure is beyond the mandate of this workshop. However, as is discussed in Section 4, we can easily deduce that each of these research directions can benefit from two types of computing platform. The most challenging is a large, tightly coupled computer with millions of processors/threads, relatively modest memory and a powerful, low-latency network. Such machines are the natural descendants of the present DOE leadership class computers. These are needed to create the long Markov chains that are central to the present Monte Carlo methods that have proven so successful in the study of quantum field theory numerically. However, at least the same scale of resources is required to compute quantum expectation values on the members of these Markov chains whose Monte Carlo average is the actual target of the calculation. These represent perhaps thousands of independent calculations, each of which can be efficiently performed on a few central processing units (CPUs) or graphics processors and are highly suited to small-cluster computing, albeit on a massive scale.

Of equal importance is the software infrastructure needed to exploit these dramatically advanced architectures. Substantial developments in both standards/tools/protocols and in the application algorithms will be required for efficient use of the fine processor/thread granularity that will characterize exascale computing. All of these developments, essential to transformational advances in science, must rely on a growing cadre of highly trained and motivated computer, computational, and application scientists upon whose ideas and implementations all of this must be built.

2. STATUS OF RESEARCH AND COMPUTATIONAL CAPABILITIES

In this section, we provide the necessary background on the science and computational methods to understand the opportunities presented by exascale computing, and we summarize the present status of research in the three priority research directions laid out above. LQCD simulations are integral to the first two directions, and it is convenient to reverse the order in which we discuss them in this section and the next. For each research direction, we illustrate the situation with one or more examples.

2.1. *Testing QCD at the Sub-Percent Level*

QCD is the theory of the strong interactions, describing the properties of quarks (*u*, *d*, *s*, *c*, *b,* and *t*) coupled to gluons. The coupling constant of quarks to gluons (and gluons to themselves) grows at low energies, causing the quarks and gluons to be absolutely confined into hadrons such as protons, neutrons, pions, and kaons. This peculiar property is non-perturbative, and the only method presently known for first-principles calculations in this regime is lattice QCD. Indeed, it was one of the early triumphs of LQCD to observe the confinement of heavy static quarks by a "string" of gluon flux running between them.

Confinement leads to a major challenge when attempting to determine the properties of quarks. If one quark decays into another because of the weak interactions, or because of some weaker BSM physics, this decay will be shielded from direct observation since the quark has a "wrapping" of gluons, quarks, and antiquarks inside the hadron. This is the sense in which non-perturbative QCD provides a "background" to the study of rare decays. LQCD aims to "strip away" this wrapping and make the connection between the experimentally observed process and the underlying quark interaction.

LQCD approximates continuous space-time with a grid of uniformly spaced points (Wilson 1974). The spacing between the points, or "lattice spacing," is denoted a and is typically less than 10^{-16}m. A surprising feature of the formulation is that one must use imaginary, or Euclidean, time, so that the quantum mechanical Feynman path integral has a real weight e^{-S} (with S the Euclidean action). If one places the system in a box (typically a few femtometers [10^{-15}m] across) it is reduced to a finite, albeit large, quantum mechanical system, suitable for numerical simulations. Quarks are placed on sites and gluons on links between them. The action is chosen to approach the continuum form as a goes to zero. This leaves much freedom for tuning the action to minimize discretization errors, and several choices have been made for the fermion action, each with different advantages and disadvantages. A crucial check of the methodology is that one obtains the same results using different fermion discretizations after one extrapolates to a=0. One must also increase the box size L until finite volume effects are negligible (or can be extrapolated away). For a recent review of these sources of error see Jansen (2008).

LQCD is studied numerically using Monte-Carlo methods to evaluate the Feynman path integral. Lattice quarks, being fermions, cannot be included (in a computationally viable way) as variables in the numerical integral. They can, however, be integrated analytically, leading to an effective action for the gluon fields of the form $S_{\text{eff}} = \log\det(D)$, with D the lattice Dirac operator and det taking its determinant. This effective action contains the effects of "quark loops": the virtual creation of quark-antiquark pairs that is the hallmark of a non-trivial quantum vacuum. It is the non-locality of this effective action that has slowed the development of LQCD. A decade ago, quark loops were largely ignored, leading to simulations of only a model of QCD — so called "quenched QCD." Major advances in algorithms, together with the speed-up of computers, have led to the present status in which *u*, *d,* and *s* quark loops are routinely included.

Present LQCD simulations use variants of the Rational Hybrid Monte Carlo (RHMC) algorithm, including many optimizations. The output is a large number (500 to 1000) of "gauge configurations" weighted by the desired action. They are picked from the Markov chain sufficiently infrequently that they are essentially independent. Observables are then measured by averaging over the configurations. Because these configurations are very expensive to create, taking up a quarter to a half of the total computational budget, they are stored and then reused repeatedly for calculation of different observables.

What is the present status of LQCD calculations? We are at the beginning of the "petascale" era, which in practice means that simulations of a scale of up to 10 teraflops-years (sustained speed) are being undertaken, employing as many as tens of thousands of cores in a single calculation. Such resources have allowed LQCD to reach the important milestone of calculating several simple quantities with all sources of error understood and with the total error at the few percent level (see, for example, Davies et al. 2004, and Durr et al. 2008). In addition, checks between different fermion methods are beginning. The success of these post-dictions, as well as that of a few predictions (Kronfeld 2006), provides quantitative validation of the lattice methodology. This is a highly non-trivial test, an essential rite of passage for a maturing field. An example of this success is shown in Figure 21, which displays the ratio of post-dictions to experimental results for a range of quantities. The left panel shows the failure of quenched QCD (no quark loops) to reproduce experiment, while the right panel shows how the inclusion of these loops leads to agreement with experiment.

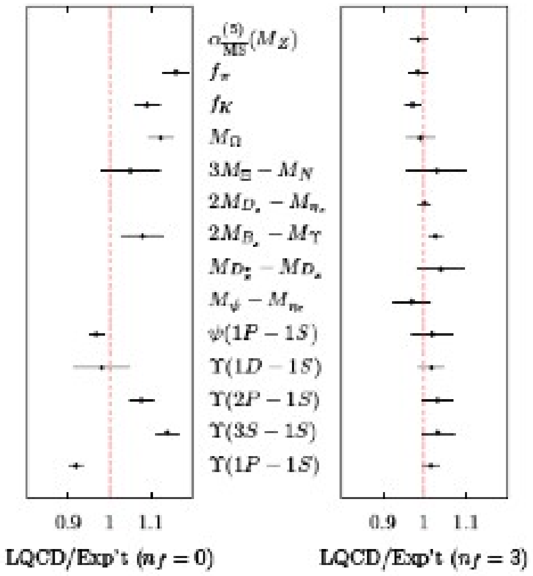


Figure 21. Results from the FNAL/HPQCD/MILC super-collaboration showing the ratio of results from LQCD calculations to experimental measurements. These results are for those quantities for which the LQCD calculations are most advanced. They include the properties of hadrons composed only of u, d, and s quarks, as well as those containing the heavier c and b quarks. The left panel shows results if quark loops are not included; the right panel includes u, d, and s loops. Image courtesy of Peter LePage (Cornell University).

Present calculations, however, have several major limitations. Most notable is the need to perform a "chiral" extrapolation from the simulated values of the light quark masses (m_u and m_d) to their physical values. The simulated values are typically a factor of 3 or more too high, and furthermore they are taken to be equal in simulations while they are significantly different in nature. The required extrapolation, while guided by theoretical considerations, leads to a significant systematic error. A contrasting second limitation arises from the large masses of the charm and bottom quarks. These quarks are sufficiently massive that the use of normal lattice discretization techniques would introduce large, finite lattice spacing errors and specialized effective field theory methods must be used in each case, methods which introduce their own uncertainties. In addition, for many quantities, the errors are considerably larger than those shown in Figure 21, while for others, e.g., the masses of resonances such as the D baryon, present resources are barely adequate to make a first controlled calculation. Finally, we note that to reduce errors to the sub-percent level will require understanding new systematic effects that are presently ignored. Examples of these will be discussed below.

All in all, while the present calculations have forged a path towards high precision calculations in the low-energy regime of QCD, present computer resources are insufficient to follow that path to the desired sub-percent level accuracy.

2.2. *Searching for Physics Beyond the Standard Model*

As described in the introduction, a major aim of this research direction is to determine the Standard Model "background" rates for rare processes so that contributions of BSM physics can be teased out. Rare processes in the Standard Model involve the electroweak interactions, and their rates are dependent on the CKM matrix that describes the transitions between quarks of different flavors. This is a unitary, 3×3 matrix, which turns out to contain only four physical parameters—three angles and a phase. In principle, each of the many electroweak processes in the Standard Model determines a combination of these four parameters. Turning this system of constraints around, one can use four standard processes to determine the CKM matrix, and then predict the Standard Model rates for all others.

Would that it were so simple! As described above, non-perturbative QCD obscures the connection between the quark-level where the CKM matrix enters and the hadron-level at which experimental measurements are made. The Standard Model prediction for each process is governed by a master equation whose schematic form is:

$$\boxed{(\mathit{Experimental\ quantity})_{SM} = (\mathit{known}) \times (\mathit{CKM}) \times (\mathit{QCD\ matrix\ element})}$$

For many experimental quantities, the only known method for calculating the "QCD matrix element" is LQCD. The priority research direction is to advance LQCD calculations to the point that they can calculate such QCD matrix elements with fractional errors that are significantly smaller than those in the corresponding experimental quantities. This would allow the search for new physics from the intensity frontier to reach its full potential.

Figure 22 illustrates the master equation for the process in which a D meson (composed of a *c* quark and an anti-*d* or anti-*s* quark) decays via the weak interactions into leptons (an electron or muon along with a neutrino). This is a timely example since these processes have been measured experimentally only recently, and the first accurate LQCD calculations have also been reported.

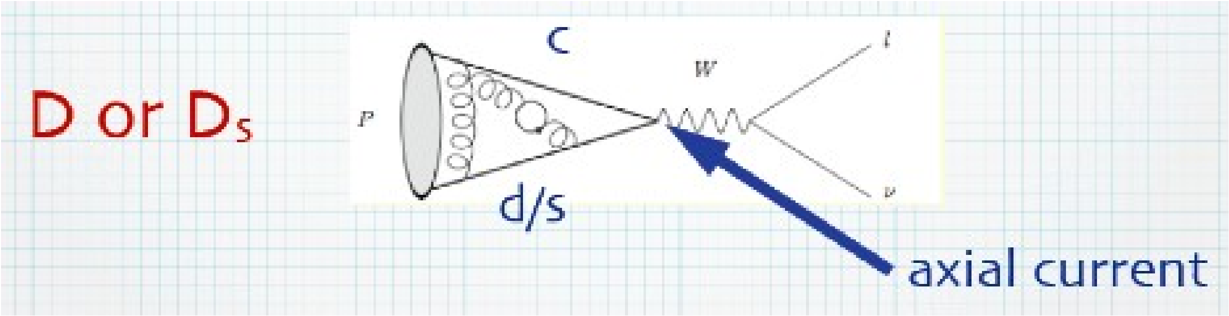


Figure 22. Example of "master equation" used to calculate QCD backgrounds to new physics contributions to rare decays. The leptonic decay of a D or D_s meson requires calculating the QCD matrix element of the axial current between the meson and the QCD vacuum. The loops and lines to the left of the axial current represent the multiple interactions of the quarks and gluons that make up the decaying D or D_s meson and which is the object of the LQCD calculation. The CKM element that enters is either V_{cd} or V_{cs}. Image courtesy of Steve Sharpe (University of Washington).

Figure 23 illustrates the range of experimental quantities for which LQCD calculations are essential. In all these cases, the methodology of LQCD calculations is well developed. The same also holds true for the additional process *B*➔*ln* as well as for CP-violation in kaon-antikaon mixing. These processes involve only a single particle, stable under strong interactions, in either or both initial and final states. This turns out to be the type of quantity for which numerical LQCD calculations are optimal. Thus the methodology is straightforward in principle—the critical issue is the need to reduce all errors to the desired level.

$$\begin{pmatrix} \mathbf{V_{ud}} & \mathbf{V_{us}} & \mathbf{V_{ub}} \\ \pi \to l\nu & K \to \pi l\nu & B \to \pi l\nu \\ & K \to l\nu & \\ \mathbf{V_{cd}} & \mathbf{V_{cs}} & \mathbf{V_{cb}} \\ D \to \pi l\nu & D \to K l\nu & B \to D^{(*)} l\nu \\ D \to l\nu & D_s \to l\nu & \\ \mathbf{V_{td}} & \mathbf{V_{ts}} & \mathbf{V_{tb}} \\ \langle B_d | \overline{B}_d \rangle & \langle B_s | \overline{B}_s \rangle & \end{pmatrix}$$

Figure 23. Some examples illustrating the connection between the underlying quark transition, labeled by the element of the CKM matrix which gives rise to the transition, and experimentally measured processes. For each process there is a master equation, and each requires a separate LQCD calculation of a hadronic matrix element.

The sources of error in the calculations of the required hadronic matrix elements include those mentioned above for the calculation of the spectrum of QCD. Statistical errors must be controlled by using a large enough ensemble of gauge configurations. We must do several extrapolations: of the *u* and *d* quark masses to their physical values, the lattice spacing to zero, and the box size to infinity and rely on effective field theory to include the massive charm and bottom quarks. One needs also to include the effects of electromagnetism, and at some level the effect of charm quark loops. A new source of error that enters for matrix elements arises from the need to relate the normalization of lattice operators to those in the continuum. And, finally, many of these processes involve initial and/or final state particles with non-vanishing spatial momentum, which increases the statistical errors. The net effect is that many of the required matrix elements are significantly more challenging to calculate than those needed to test QCD itself.

To give an indication of the present status, as well as the challenges involved, we show in Table 4 the present experimental and lattice errors for several processes, the likely improvement in the experimental result over the next decade (from LHCb and a super-B factory) as well as an estimate (described in the following section) of what a thousand-fold increase in resources could accomplish. We see that typical lattice errors are presently in the few to 10% range, while to exceed the future experimental accuracy one needs calculations that have percent or sub-percent level lattice errors. To illustrate this further, we discuss three examples in more detail.

Table 4. Important quantities related to bottom quarks in which LQCD calculations play a critical role. The contribution of experimental and LQCD errors to each of these quantities is shown, including the size of the LQCD errors that can be achieved with exascale resources as is discussed in Section 3. For a more detailed discussion of the basis of these estimates, see (Brower et al. 2007a). For recent reviews of the status of the lattice calculations see (Artuso et al. 2008, Gamiz 2008, Wingate 2006).

Process/quantity	Present experimental error	Present lattice error	Future experimental error	Exascale lattice error
$B \rightarrow \mathrm{l}\, n \ / \ f_B$	29%	6%	3-4%	1.5%
$B \rightarrow p\, l\, n$	6%	10%	2-3%	3%
$B \rightarrow D\, l\, n$	3.7%	2.2%	2%	0.5%
$B \rightarrow D^{*}\, l\, n$	2.2%	2.6%	1%	0.5%
$B - \overline{B} mixing$	0.5%	5-10%	--	1%
$B_S - \overline{B_S}\ mixing$	0.35%	4-8%	--	1%
ξ $(B_S - \overline{B_S} / B - \overline{B} mixing)$	0.6%	3-5%	< 0.5%	< 1%

2.2.1 Mixing of b-Quark Mesons and the ξ Parameter

The oscillations between B or B_S mesons and their antiparticles is an example of a process that is expected to be a sensitive probe of new physics, since the rate is highly suppressed in the Standard Model. The oscillation rates have, furthermore, been measured to high accuracy: for B mesons (which contain an anti-b and a d quark) an accuracy of 1% has been attained by “B-factories” at the Stanford Linear Accelerator Center (SLAC) and Japan’s High Energy Accelerator Research Organization; for B_S mesons (which contain an anti-b and an s quark) the Tevatron has attained 0.7% accuracy. Thus a LQCD calculation of the corresponding hadronic matrix elements with percent-level accuracy is of highest importance. Particularly promising in this regard is the ratio of the two oscillation rates, because some uncertainties cancel in the Standard Model prediction. In particular, what the lattice must provide is a ratio of matrix elements—traditionally called $(x)^2$—and a number of the lattice uncertainties are reduced in this ratio. As can be seen from the bottom row of the table, the error in present lattice calculations of x exceeds the experimental accuracy of the mixing measurements by almost an order of magnitude. This is an example, therefore, where there is a large potential payoff for improving the lattice calculation.

2.2.2 Kaon Physics and First-Row Unitarity

The last few years have seen a dramatic improvement in the accuracy of experimental measurements of certain kaon decays, thanks to a worldwide effort utilizing several experiments (BNL-E865, KLOE, KTeV ISTRA+ and NA48). Some transition rates are now known with uncertainties on the order of a few parts per thousand. The challenge to LQCD is to match this level of accuracy in the corresponding matrix elements. Most important in this regard are the ratio of kaon to pion decays constants, f_K/f_p, and the "form factor" relevant for $K \rightarrow pln$ decays. The former is presently calculated to 1% to 1.5% accuracy using several choices of lattice fermion, the latter to about 0.5%. (For a recent compilation of lattice results, see [Sciascia et al. 2008; Artuso et al. 2008]). These are in striking distance of corresponding experimental errors, which are presently 0.3% and 0.4%, respectively. The required accuracy in both cases is somewhat misleading because LQCD is really calculating the deviation from unity of, for example, f_K/f_p, and the desired error in these deviations are at the few percent (rather than few per mil) level.

The search for new physics in kaon decays can be phrased in a particularly simple and appealing way: it would lead to a violation of the unitarity of the first row of the CKM matrix:

$$|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 = 1 + \varepsilon_{NP}$$

Here ε_{NP} is the contribution from new physics, which will be suppressed by $(M_W/\Lambda_{NP})^2$, with $M_W \approx 100\, GeV$ the W boson mass, and Λ_{NP} the energy scale at which the new physics enters. Here one sees very directly the connection between precision measurements and the search for new physics. A deviation from unity of one per mil would correspond to $\Lambda_{NP} \approx 3\, TeV$. Conversely, if there is evidence for new physics at scales lower than this from the LHC, then the models explaining it must be consistent with the constraints on ε_{NP}. What are those constraints at present? One finds

$$|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 = 0.9999(5)(9)$$

where the first error is the uncertainty from $|V_{ud}|^2$ and the second from $|V_{us}|^2$ (the error from $|V_{ub}|^2$ being negligible (Sciascia 2008). Errors in lattice calculations are the largest contribution to the $|V_{us}|^2$ error, and the aim is to substantially reduce them.

2.2.3 CP Asymmetry and the Two Pion Decays of the Kaon

The K mesons (K^+, K^-, K_0 and $\overline{K_0}$) are spinless particles with a mass of roughly 500 MeV that decay because of the electroweak interactions. Their lifetimes vary between 10^{-8} and 10^{-10} seconds. This system has been studied for more than 50 years and has been the source of two famous discoveries: non-conservation of parity in 1956, and the violation of CP symmetry in 1964. The latter violation arises from two mechanisms, direct and indirect. Indirect CP violation is the larger and was measured first: it is characterized by a small parameter $|\varepsilon| = 2.229(10) \times 10^{-3}$. Direct CP violation is nearly a thousand times smaller and is described by a second parameter ε'. After enormous experiment effort, ingenuity and expense extending over three decades, the real part of their ratio was shown definitively to be different from zero:

$$\mathrm{Re}(\varepsilon'/\varepsilon) = 16.5(2.6) \times 10^{-6}.$$

Both ε and ε' are prime targets in the search for new physics, since, as for first-row unitary, the experimental precision reaches to a level where BSM effects might be expected to appear. Thus it is an important goal of lattice calculations to determine the corresponding matrix elements. In this regard, it is important to note that both ε and ε' are non-vanishing because of the single CP-violating phase in the

CKM matrix, the same phase which leads to (the by now very well measured) CP-violation in B-meson decays. This once again illustrates the highly constrained nature of the Standard Model contributions to different processes.

The hadronic matrix element needed to predict ε in the Standard Model is called B_K, and has been the target of LQCD calculations for more than 20 years. It is presently known to about 7%, an error more than an order of magnitude larger than the experimental error (Antonio et al. 2008). Nevertheless, this is a straightforward quantity to calculate. The calculation of the Standard Model prediction for ε' is, by contrast, much more difficult. While pioneering work has been done, no calculation with controlled errors has yet been possible. It is an example of a class of quantities for which exascale resources are needed.

2.3. *Simulating Theories of Physics Beyond the Standard Model*

Simulating BSM theories is an exciting new research direction that has been growing rapidly in the last few years, spurred by theoretical developments, the increase in computational resources, and the approach of the LHC era. For a recent review with extensive references see (Brower et al. 2007b).

2.3.1 TeV Scale Strong Coupling Models

Nearly conformal gauge theories provide a very interesting and promising class of BSM theories in which a new strong force similar to QCD but operating at the TeV scale (three orders of magnitude larger than the QCD scale) plays a central role, replacing the Higgs particle as the source of electroweak symmetry breaking. The theories are built on new fundamental particles, historically known as “techniquarks,” that are different from the QCD quarks. They are massless and occur in an unknown number of N_c colors and N_f flavors. Candidate models also differ in the choice of representation of the gauge group in the fundamental Lagrangian. The theories have a chiral symmetry at the fundamental, Lagrangian level that is dynamically broken in the vacuum — thus providing the replacement for the canonical Higgs mechanism of electroweak symmetry breaking. The techniquarks are bound by the confining force of the new theory into heavy and colorless composite particles with masses at the TeV scale. One of those composite particles could have similar properties to the Higgs boson of the Standard Model.

This new physics can affect low-energy precision measurements through virtual quantum fluctuations. A tight constraint on such theories is that the effects of such fluctuations are small enough to be consistent with the lack of evidence of deviations from Standard Model predictions in low-energy experimental measurements. Two such quantum effects are the contribution of the new interactions to the S parameter (Peskin and Takeuchi 1992) and to flavor changing neutral currents. There are theoretical arguments that nearly conformal gauge theories will naturally suppress contributions to flavor changing neutral currents and perhaps even the S parameter (Appelquist and Sannino 1999). This is the reason for the interest in searching for nearly conformal theories.

Finding a nearly conformal Yang-Mills theory requires that one understand how the properties of the theory depend on N_f. Theories with relatively few flavors, such as QCD, exhibit the necessary confinement and chiral symmetry breaking, but are not nearly conformal. However, a study of scaling behavior in perturbation theory implies that these properties are lost for large N_f, giving way to conformal behavior in the infrared. The transition between these two phases occurs at a critical value $N_f^c(N_c)$, just below which the theory may show approximately conformal infrared behavior, making it a good candidate for a viable model of walking Technicolor (Applequist and Wijewardhana 1987). Simulating a nearly conformal theory is intrinsically, and unavoidably, more challenging than simulating LQCD. This is because there is a larger range of scale between the short distance perturbative region in which the theory is unambiguously defined and the large distance regime in which near conformal behavior can be convincingly recognized. Present work has focused on mapping out the "theory space" as a function of N_f, N_c and the representation of the techniquarks, searching for candidates for more detailed studies. An example of results from this pioneering work is shown in Figure 24. Simulating QCD-like theories with N_c=3 but with more flavors indicates that the critical number of colors satisfies $8 < N_f^c < 12$. Zeroing in on the precise value of N_f^c is the next task.

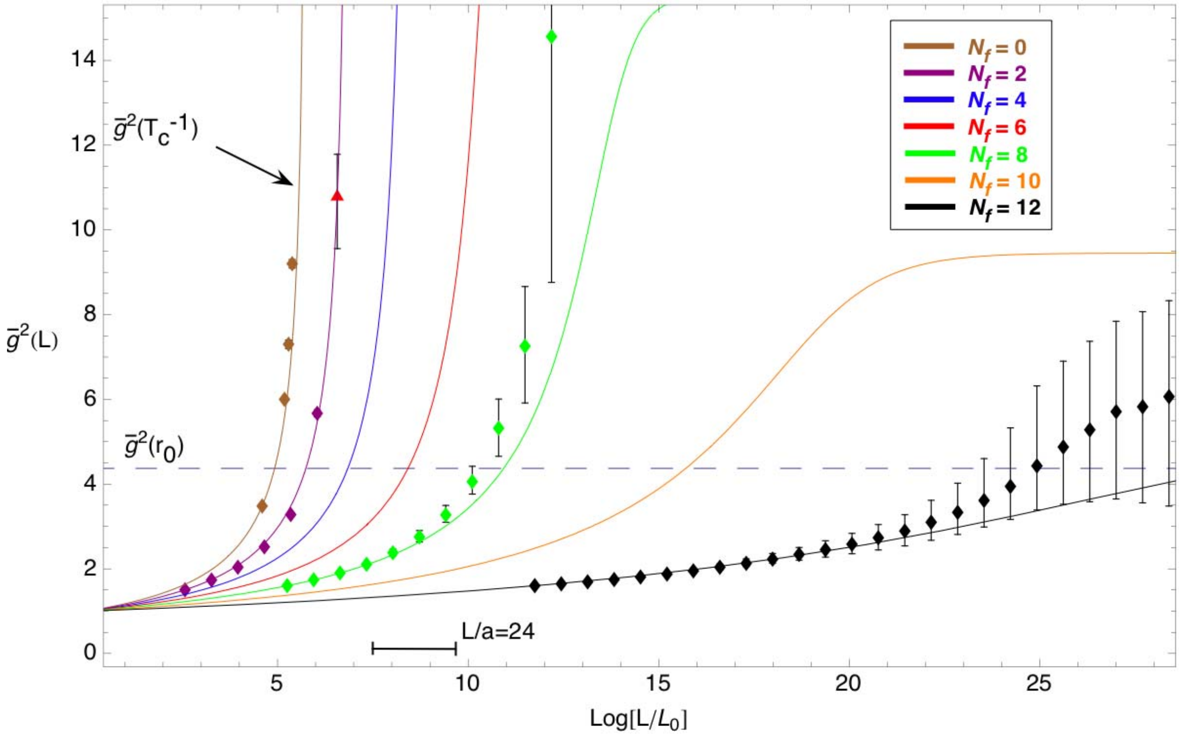


Figure 24. The running of the renormalized coupling in the Schrödinger functional scheme for QCD-like theories with various numbers of flavors (Bode et al. 2002, Della Morte et al. 2005, Appelquist et al. 2008). The horizontal axis gives the length scale, L, at which the coupling g^2 is measured (in arbitrary units). Curves are predictions from an analytic model. Conformal behavior is indicated by the coupling becoming constant at large L. Such constant behavior may be evident in the $N_f = 12$ points. Image courtesy of George Fleming (Yale University).

2.3.2 Lattice Supersymmetry

Supersymmetry was proposed as a possible symmetry of nature nearly 40 years ago. It is a natural extension of the usual symmetries of space and time such as translations and rotations. Its consequences for particle physics are dramatic—it predicts that every boson or particle of integer spin is partnered with a spin one-half fermion of equal mass and charge. From a phenomenological point of view, it can explain why the Higgs particle responsible for breaking electroweak symmetry is light, while from a theoretical perspective it plays a crucial role in constructing consistent string theories. For these and other reasons supersymmetric (SUSY) models have been the subject of intense study by theoretical particle physicists for several decades. Many of the properties of these theories are, however, non-perturbative, and despite significant advances, are often not tractable using analytic methods. Thus, there has been considerable recent effort at formulating lattice methods to study SUSY theories non-perturbatively.

Unlike for QCD or the nearly conformal theories discussed above, it is highly non-trivial to set up a lattice theory that becomes fully supersymmetric when the lattice spacing is extrapolated to zero. This problem has now been solved for a range of interesting theories (see, for example, Kaplan et al. 2003, Catterall 2008); an example of the exotic lattice structures that are needed is shown in Figure 25. The next stage is to simulate such theories. Preliminary work has begun (Catterall 2009), but the scale of the calculations is such that they will certainly require exascale resources.

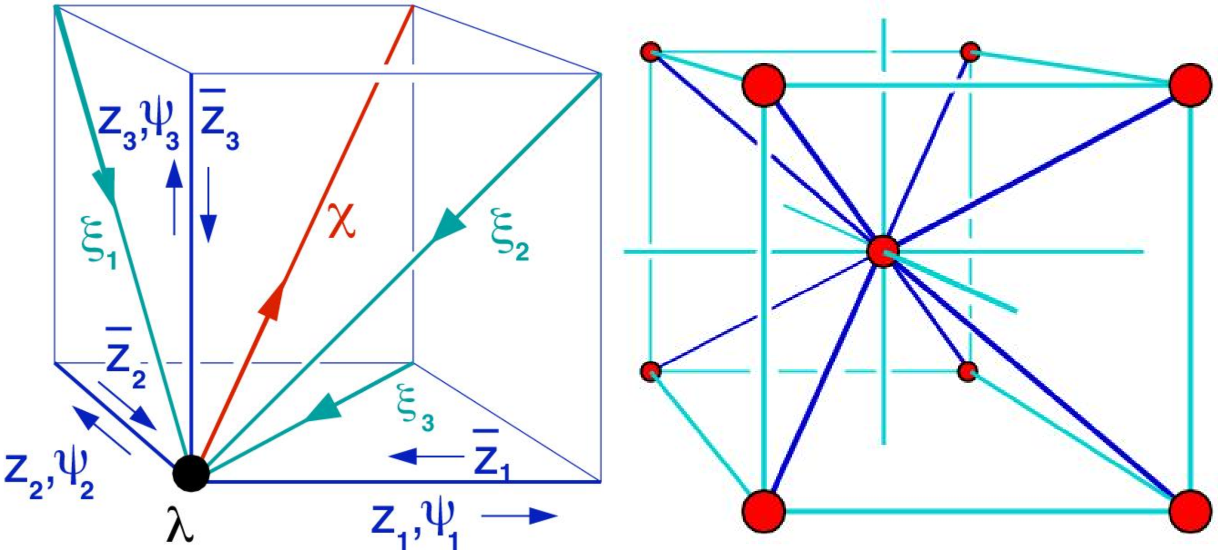


Figure 25. Lattice geometries needed for sample supersymmetric Yang-Mills theories. The left figure is for the three-dimensional theory with 8 supercharges (Cohen et al. 2003), the right for the three-dimensional theory with 16 supercharges (Kaplan and Unsal 2005). An analogous lattice exists in four dimensions for simulating N=4 super Yang-Mills theory, a gauge theory that has been related to quantum gravity in the limit of a large number of colors. Image courtesy of David Kaplan (University of Washington).

3. OPPORTUNITIES FOR PRIORITY RESEARCH DIRECTIONS OFFERED BY EXTREME SCALE COMPUTING

In the previous two sections, we have introduced and described in some detail the three priority research directions that offer enormous opportunities for extreme scale computing. In the present section, we will discuss the transformative results that might be expected when exascale resources are brought to bear on these three directions. It is important to recognize that the specific goals in each case are directly the results of large-scale computation. Thus, extreme computing resources will have an immediate, direct impact on the priority research directions.

While the emphasis in this section is on the advances that can be achieved with exascale resources using present numerical methods, it is important to recognize a second transformative effect that exascale resources will have on these research directions. Since these efforts are very much science driven, the initial exascale resources will be applied to solving important problems that yield results intended to extract new physics from experiments performed at the intensity frontier and to guide the interpretation of experiments at the energy frontier. However, between 20% to 30% of these resources should be immediately devoted to developing and testing new numerical methods. With pioneering ideas being developed, we expect a series of promising new developments that can only be verified and exploited when exascale resources become available. If these developments follow the pattern seen when LQCD studies moved to sub-petascale resources, e.g., the development of the Rational Hybrid Monte Carlo algorithm, we should expect major advances in methods that will further enhance the science opportunities that can be exploited. Of course, such synergy requires the human resources discussed in Section 4.

3.1. *Testing QCD at the Sub-Percent Level*

To attain the goal of precision testing of QCD in the non-perturbative regime, it will be essential to *work directly at the physical quark masses*, i.e., directly simulate the physical theory. This avoids the need for chiral extrapolation. To do so necessitates covering a large range of scales: from a box size of $\sim 4/m_\pi \approx 6\,fm$ (ensuring that corrections because the finite volume is very small and controllable, at least for single-particle states) down to a lattice spacing of $a \sim 0.03 - 0.06\,fm$ (ensuring that we are close enough to the continuum limit that discretization errors after continuum extrapolation are very small).

Extreme scale computing offers the opportunity to reach this milestone. We can estimate the necessary resource requirements based on conservative scaling laws for operation counts normalized to the performance of our present (highly optimized) algorithms. These extrapolations assume that lattice field theorists will be able to program future machines so as to maintain the 30% to 50% efficiency attained on present architectures. The extrapolation formulas depend on the type of fermion used and vary in their level of sophistication. We quote as an example that for improved Wilson fermions used by the PACS-CS collaboration:

$$\boxed{C_{op} = 0.024 \times \left(\frac{L^3 \cdot T}{(6\,fm)^4} \right)^{5/4} \left(\frac{135 MeV}{m_\pi} \right)^2 \left(\frac{0.1 fm}{a} \right)^6 \left(\frac{run\ length}{10^4 \tau} \right) Pflops \cdot yrs}$$

Using this formula, one finds that for $L=6$ fm, temporal length $T=2L$, physical pion masses ($m_p=135$MeV), and a lattice spacing $a=0.06$ fm, a run of length 10^4 time units (chosen so that statistical errors are likely to be sufficiently small) requires about 1.2 petaflops-years. Projections using domain-wall fermions, which have theoretical advantages, are estimated to require similar time within a factor of 2. Those using staggered fermions will require about a fifth of this time, while those using overlap fermions will need at least an order of magnitude more.

Scaling up from our present experience, we expect that when the peak speed of the fastest computer reaches 1 exaflops, simulations with a given type of fermion will have of order 10 petaflops-years (sustained speed) available. Given that a majority of this time will be needed for calculating observables, such as those required to search for BSM physics, the scale of resources available for generating gauge configurations is likely to be a few petaflops-years per fermion type. Based on the estimates given above, we conclude that exascale resources will allow lattice simulations to attain the goal of direct physical simulations of QCD.

A second and more specialized opportunity offered by exascale resources is a direct simulation of relativistic *b* quarks. The *b* quark presents a significant challenge to lattice methods because of its high mass (~5GeV—about 5 times the proton mass). Present, high-precision results for the masses and matrix elements of states involving *b* quarks all exploit the fact that *b* quarks move slowly inside hadrons because the energy scale of QCD, LQCD, is an order of magnitude smaller than m_b. For example, "heavy-quark effective theory" allows one to study *B*-mesons in an expansion in powers of LQCD/m_b in such a way that discretization errors are similar to those for hadrons composed of only the light *u*, *d* and *s* quarks (El-Khadra et al. 1997, Heitger and Sommer 2004). These approaches, however, necessarily introduce new errors from truncating the heavy quark expansions that are employed. These errors could be better understood and estimates of their size tested by a direct calculation of *relativistic b* quarks on the lattice. An unresolved question is how small the lattice spacing needs to be for such a direct calculation to have a major impact: Is $am_b<0.5$ sufficient (corresponding to $a=0.02$fm), or does one need $am_b\sim0.1$ (corresponding to $a=0.004$fm)? The former is attainable with exascale resources (by reverting to somewhat heavier *u* and *d* quark masses), while the latter is more challenging. We stress, however, that even if only exploratory simulations of relativistic *b*-quarks are possible, heavy-quark effective theory and related approaches will still provide accurate methods for simulation, since truncation errors can be estimated and will be small.

Even with direct simulations at physical quark masses, several new sources of error must be understood in order to attain sub-percent level precision. We describe two examples.

3.1.1 Electromagnetic and Isospin Breaking Effects

A transformational change in LQCD simulations would be the inclusion of electromagnetic and isospin breaking effects directly into the calculations. Quantum electrodynamics is the quantum field theory describing one of Nature's fundamental forces, that of electricity and magnetism. All charged particles interact through the exchange of photons. Quarks are charged—e.g., the *u* quark has charge +2/3*e* while the *d* has charge -1/3*e*—but current lattice calculations do not include these interactions. Isospin breaking refers to the fact that the *u* and *d* quarks differ not only in their charges, but also in their masses. Both masses are very small compared to that of the strange quark, which is about 27 times heavier than the average of the *u* and *d* quark masses. However, the positive up-down mass difference has important consequences that must be included when aiming for high accuracy, as we now explain.

What are the typical sizes of these effects? This can be seen from the masses of strongly interacting particles such as pions and nucleons. For example, $m_{\pi^+} - m_{\pi^0} = 4.5936(5)\,MeV$ (dominantly an EM effect), while $m_n - m_p = 1.2933317(5)\,MeV$ (because of isospin breaking dominating over the negative EM contribution). These are, respectively, effects of size 3% and 1%. Clearly, sub-percent level precision (which is certainly possible for the statistical errors alone) requires their inclusion.

What are the implications of these mass-splittings in nature? To see this, consider two alternative situations. Were the proton heavier rather than lighter than the neutron, the latter would be the stable particle and our world would look very different, with most charged nuclei and the familiar chemical elements built from them becoming unstable. If instead the neutron mass were larger than its present value, its lifetime would be shorter and consequently the evolution of the early universe through the era of nucleosynthesis would be dramatically changed. It would clearly be a great achievement of LQCD to determine the neutron-proton mass difference from first principles.

Another opportunity presented by the inclusion of electromagnetic and isospin breaking effects is a more accurate determination of the u and d masses themselves. These are fundamental parameters of the Standard Model, which should be determined in a more fundamental synthesis of nature. The better we can determine these parameters of the Standard Model, the more stringent will be our demands of more fundamental theories.

What will be required to include these two effects? Taking account of the u-d mass is relatively straightforward in present algorithms once one is able to attain physical values, and adds little computational cost (Aubin et al. 2008, Beane et al. 2007). Including electromagnetic effects is more costly, possibly doubling the total computational time, and the preferred method is not yet clear. A direct approach would include dynamical photons, and the quark loops that they introduce. An alternative would be to first leave out the (very small) electromagnetic-induced quark loops but keep the effect of charges on external quarks (see, for example, Blum et al. 2007). Because electromagnetic effects are small, it might also be necessary to artificially boost the quark charges to enhance these effects and then to extrapolate to the physical value. Whether this will be needed can be determined in less costly calculations on coarser grids. This and other issues are already the target of preliminary research, but the bottom line is that we expect all such issues can be addressed with exascale resources.

3.1.2 Including the Charmed Sea

To achieve our goal of sub-percent uncertainties, it will be necessary to include the effect of charmed sea quarks. Most of their effect is a renormalization of the bare gauge coupling and light quark masses and, hence, is included implicitly in all lattice calculations. The remaining effects can be separated into two categories. In the first, one expects changes in low energy QCD phenomena arising from rare fluctuations in which a virtual charmed-quark loop is long-lived. Such effects have not yet been thoroughly investigated, but it is possible that for some quantities they may be as large as a few percent. The second category of charm quark sea effects appears when one attempts to relate the normalization of lattice operators, defined at an energy scale given by the inverse lattice spacing, $1/a = 2 – 3$ GeV, with the normalization typically imposed on continuum operators at the much larger energy scales needed for QCD perturbation theory. At these energies, the charm, strange, up and down quarks all play similar roles and a consistent inclusion of charm quark loops is important. Once that has been done, an application of existing non-perturbative techniques should permit accurate operator normalization, including the effects of charmed sea quarks.

The extra computational burden for including charmed sea quarks is moderate and smaller than that for the strange sea. However, an important reason to omit the charmed sea is because the associated discretization errors, on lattices available until now, can potentially equal the small physical effect of the charmed sea quarks so that their inclusion may actually decrease the accuracy of the calculation. However, for sub-percent accuracy, these effects must be carefully studied and included. This will be feasible once the lattice spacing is small enough that one can control the associated discretization errors. For example, as described above we expect with exascale resources to be able to study lattice spacing in the range of approximately 0.02 – 0.06 fm. The corresponding product of lattice spacing times charm quark mass, the natural measure of finite lattice errors, lie in the range $m_c a \sim 0.13 – 0.4$, which should be sufficiently small to allow an accurate extrapolation to the continuum limit. By reducing the product $m_c a$ to a size where a standard extrapolation $m_c a \rightarrow 0$ can be performed, exascale resources will permit the effects of charm sea quarks to be included without degraded accuracy.

3.2. *Searching for Physics Beyond the Standard Model*

The search for BSM physics relies crucially on LQCD calculations of hadronic matrix elements. As explained in the previous section, percent (or even sub-percent) accuracy for such matrix elements is required to fully exploit the array of impressive experimental results on rare Standard Model processes. This is shown by the examples given in Table 4. The right-hand column gives estimates of the error with calculations requiring of order 10 petaflops-years. These estimates assume the same methodology—and in that sense are conservative. They do, however, assume that the additional systematic errors that arise at the approximately 1% level (two of which were discussed in the previous sub-section) have been controlled. Direct simulations at the physical quark masses—a possibility requiring exascale resources—are essential to reaching this accuracy. Attaining errors at the level of the right-hand column will transform the search for new physics using precision processes by essential removing the theoretical errors as a concern.

To achieve this accuracy, it is clearly essential to have verified the methodology at better than this level in calculations of purely strong-interaction quantities. Thus the priority research direction being discussed in this subsection is intimately tied with the previous one of testing QCD. In particular, the systematic errors discussed in the previous sub-section must also be controlled in the calculations of matrix elements. There are, however, additional sources of error that enter for many matrix elements—most notably in the normalization of operators and in the need for states involving more than a single particle. We discuss the latter below, in the context of our example of two pion decays of the kaon. We also mention quantities for which a methodology has not yet been developed, but for which LQCD calculations are of great interest in order to search for new physics.

3.2.1 Mixing of b-Quark Mesons and the ξ Parameter

We have called attention earlier to the good features of the ratio, x, of *B* mixing matrix elements. We stress here that we are in a better position to estimate what a factor of 1000 in computing power will do to the errors in x than for most other matrix elements. This is because it has an unusually simple error budget. Since it is a ratio, many of the usual uncertainties in lattice calculations—from heavy quark discretization, finite volume, and operator normalization—largely cancel. Indeed, in present calculations, only statistical errors and those from chiral extrapolations are above 1%. These two sources of error are those whose future reduction can be most reliably estimated. Indeed, the chiral extrapolation error can be entirely removed by working at the physical light quark masses. Thus, for x we can feel particularly confident that sub-percent errors are a reasonable, though ambitious, goal.

3.2.2 CP Asymmetry and the Two Pion Decays of the Kaon

As discussed in Section 2.2.3, calculation of direct CP violation in *K* → *p p* decay, specifically the quantity re(*e ¢/e*), with controlled errors is an extreme scale computing challenge with substantial scientific potential. However, the accurate calculation of this quantity must overcome four major difficulties (as reviewed, e.g., in Dawson et al. 1998):

1. operator normalization
2. quadratically divergent matrix elements
3. disconnected graphs
4. two particle final states.

The first two difficulties have been studied extensively in calculations of the easier K to vacuum and K to pion matrix elements (Blum et al. 2003, Noaki 2003). With sufficient statistics, these two obstacles are expected to contribute errors that can be reduced to the level of a few percent. Calculations involving disconnected diagrams are notoriously difficult and present the greatest challenge. However, this case is likely the most favorable of those that have been studied, benefitting from the largest signal to noise ratio for the disconnected diagrams that must be computed. In addition, early indications suggest that the disconnected diagrams may be less important than others that are more easily computed. Finally, the need to treat two-particle final states is also far from routine. However, both the third and fourth difficulties are presently the target of active research at the sub-petaflops scale. As a result the necessary techniques should be well understood during the next five years and will ready to be used when exascale computing resources become available.

Thus, we expect that the necessary groundwork will be carried out to permit break-through exaflops computation of these $K \rightarrow p\, p$ decays. With total errors of 10% or less, these calculations will determine the contribution of the Standard Model to these decays making it possible to recognize the contributions from new physics. We must determine if these decays will offer still deeper insights into the asymmetry between particles and their anti-particles, perhaps revealing a new source of CP violation of sufficient strength to explain why we and our universe is constructed from matter rather than anti-matter.

3.2.3 New Challenges at the Exascale

There are many other *B*-meson decay modes for which BSM physics may contribute significantly. For instance, there are already now indications that non-Standard Model effects are observed in $B \rightarrow pK$ decays. At present, lattice calculations cannot be directly applied to such non-leptonic decays in which only hadrons appear in the final state. There is, however, a theoretical framework in which to treat these decay modes systematically by separating perturbative and non-perturbative pieces. The non-perturbative piece, the so-called "light-cone distribution amplitude," is an obvious candidate for lattice calculations. Developing an appropriate methodology is an extreme-scale challenge.

3.3. ***Simulating Theories of Physics Beyond the Standard Model***

The transformational potential of exascale resources is particularly striking in our third priority research direction. The successful simulation of BSM theories would enormously broaden our knowledge of the variety of non-perturbative phenomena exhibited by field theories, knowledge that is likely to be essential to the search for new physics at the LHC and elsewhere. As explained above, however, the simulation of BSM theories is intrinsically more challenging than simulations of QCD: BSM theories often involve a larger range of scales that must be studied. There is typically a broad variety of theories that might be examined. In many cases the methodology for simulating the theories is immature. The pioneering efforts to date have begun by considering those theories that are closest to QCD—to move further from this familiar ground will undoubtedly require exascale resources.

3.3.1 TeV Scale Strong Coupling Models

A phenomenologically viable nearly conformal BSM gauge theory must satisfy three core constraints.
It must

- provide a replacement for the electroweak symmetry breaking mechanism provided in the Standard Model by the Higgs sector
- exhibit appropriate strong-coupling behavior over a large enough range of energy scales to allow the generation of quark masses
- remain consistent with precision tests of the electroweak theory.

While petascale resources allow us to begin scanning for candidate theories, the detailed study of candidates that are identified is beyond present capabilities. Such detailed study will be essential to determining whether models exist that satisfy all three constraints. It will require not only simulation of a wider range of energy scales than needed for QCD, but also the preservation of the chiral properties of the theories under consideration. This is likely to ultimately require the use of domain-wall or overlap fermions, the most computationally demanding of the fermion discretizations presently used for QCD. Nevertheless, our experience with QCD allows us to estimate the required resources, and thus we can confidently predict that exascale resources will produce major advances.

3.3.2 Lattice Supersymmetry

Lattice simulations of SUSY model lattice theories in the exascale era are likely to be in two primary areas: models directed at elucidating LHC phenomenology and those aimed at forging a connection with superstring theory. We discuss these in turn.

Dynamical SUSY Breaking and Soft Parameters in the “MSSM”

The Minimal Supersymmetric Standard Model (MSSM) is perhaps the most studied extension of the Standard Model and a great deal of effort has already been expended in predicting what kind of signals such a theory would produce at the LHC. Unfortunately, the MSSM contains a large number of so-called soft breaking parameters that are put in by hand to ensure that the superpartners of the usual Standard Model particles are sufficiently heavy to have escaped detection so far. In principle, these parameters should be determined by the strongly coupled dynamics that breaks SUSY at a high energy scale. Present candidates for such high-scale theories are essentially supersymmetric versions of QCD (SQCD) and so one might hope to use the tools and techniques of lattice gauge theory to understand and predict this dynamics, and hence constrain or even predict many of the parameters of the MSSM. Thus a primary goal is to develop the tools and experience to study SQCD on the lattice.

A possible roadmap for these calculations is as follows. (1) Conduct studies of N=1 super Yang-Mills, arguably the simplest SUSY theory—it contains only gauge bosons and fermions—but also the core of the MSSM. One can guarantee that supersymmetry is automatically restored in the continuum limit by using domain wall fermions – a feature not true of other fermion discretization schemes. The first studies of this theory have begun, optimized codes exist and the relevant range of parameter space has been determined. A computation of physical observables such as the gaugino condensate and low lying spectrum will be achieved with order 1 teraflops-years of resources. (2) Extend these calculations to SQCD. This requires the inclusion of scalar (spinless) particles, a major complication since one must tune their interactions in order to recover supersymmetry in the continuum limit. We expect that such tuning will be possible with exascale resources, using a series of runs over a grid of parameter values. Multicanonical reweighting techniques can then be used to tune to the supersymmetric point. At each point in this parameter space a computational effort will be needed which is comparable to the N=1 super

Yang-Mills case. For example, for the typical case of a theory with four such fine tunings we estimate the tuning would require 1 to 10 petaflops-years sustained.

Lattice Gauge Theory as a Tool in String Theory

String theory evolved out of attempts to understand the strong interactions. Such early efforts were discarded with the successful development of QCD, and string theory moved on to the more ambitious goal of unifying all four forces of nature, including gravity. In the last decade, however, there has been a huge resurgence of interest in the connections between string theory and gauge theories such as QCD. This was initiated by Maldacena's conjecture (1997) that N=4 super Yang-Mills theory is equivalent to type IIb string theory propagating on a five dimensional anti-de Sitter space. The number of these so-called "AdS/CFT dualities" connecting QCD-like theories with string theories is now vast. Examples exist in many dimensions, for many different space-time geometries and for many types of conformal and non-conformal gauge theories. However, in most cases these dualities are conjectural in nature and based on calculations in which the number of colors is taken large and the string theory is computed at low energy. Typically the dual Yang-Mills theories are strongly coupled and existing analytical techniques fail. This motivates the use of the lattice field theory methods—whose application here could provide a new, non-perturbative tool to study theories of quantum gravity.

This is the general focus of this second thread of lattice supersymmetry study. A major aim is to learn how to simulate the N=4 super Yang-Mills theory. Two possible strategies are as follows.

- N=4 super Yang-Mills is a special case of the super QCD theories discussed above. We can thus hope to study them using domain-wall fermions. In this case, exactly four quartic operators require fine tuning. The computational cost is high but the approach is somewhat conservative in the sense that the numerical algorithms are well understood and simple versions of the code for N=1 already exist. The scale is that given above, 1 to 10 petaflops-years.
- Over the last five years a number of exciting theoretical developments have taken place which have culminated in a lattice action for N=4 super Yang-Mills in which the supersymmetry is exact even for non-zero lattice spacing. This dramatically reduces the amount of fine tuning need to ensure the continuum limit is supersymmetric. To use this second approach, new codes need to be developed. However, single core codes already exist and exploratory calculations have begun with some promising results. This approach has the potential to reduce the computational costs by an order of magnitude.

These two approaches should allow us to make contact with the wealth of physics applications flowing from the AdS/CFT correspondence. It is also possible that certain mass deformations of N=4 super Yang-Mills could play a role in LHC type phenomenology.

4. REQUIRED COMPUTATIONAL AND INFRASTRUCTURE RESOURCES

In this section we discuss the hardware capabilities and software environment needed to realize the science opportunities presented by exascale computing for high energy theoretical physics. As with the 4 previous panel topics, a variety of hardware platforms are needed, or can be efficiently exploited, and very substantial software innovation and development is required. In addition, success depends critically on an increasingly sophisticated and computationally expert cohort of application scientists supported by computational and computer scientists focused on the challenges of exascale computing for these applications.

4.1. *Hardware Requirements*

Lattice field theory calculations can exploit a broad range of hardware platforms. Central to LFT are the extreme scale capability machines that are the focus of this report. Each of the ensembles of gauge configurations which are required for this work represents a series of Monte Carlo samples generated as a single, long Markov chain. Substantial computer resources are needed to evolve such a Markov chain to a point where the underlying equilibrium distribution is being sampled. Thus, a single, large capability machine is needed to generate one such ensemble. Alternatively a few large partitions of such a computer might be used to equilibrate and sample configurations corresponding to a series of lattice spacings, lattice volumes and boundary conditions.

Once these valuable ensembles of gauge configurations have been generated, “measurements” must be performed by evaluating the physical quantities of interest on each Monte Carlo sample and averaging the results over the entire ensemble. Since many independent calculations must be performed, these calculations can be efficiently carried out on much smaller partitions and are presently well suited to workstations clusters, or farms of graphics processing units (GPUs). This situation is likely to continue as technology advances to the exascale although one might imagine that the large-scale integration needed to create practical exascale computers may result in this platform becoming the most efficient means of delivering the extreme resources that are required for the measurements being discussed here. Note that while this measurement part of a typical LFT project does not require the few large partitions of a capability machine needed for the configuration generation, these measurements are likely to consume 2 to 4 times the total floating point operations needed to generate the gauge configurations.

The substantial requirements of lattice ensemble generation are expected to be well suited to future, exascale computers. These calculations require a challenging balance between floating point capability and communication bandwidth and latency. However, given the steep scaling of computation cost with lattice spacing discussed in Section 2.2, the total memory requirements grow slowly with computing performance that typically implies relatively small requirements for the memory per node. We anticipate that cache sizes in the exascale era will be sufficient to hold most critical data in an ensemble generation project. Current trends in the design of petaflops systems appear efficiently matched to these LFT calculations. For example the trend (typified by IBM’s BG/X series of now petaflops computers or Intel’s 80-core Polaris chips) of using increasingly powerful nodes with tens to hundreds of cores, allows fast, on-chip communication to make up the most challenging, first layer of a heterogeneous network. Even with economical mesh-like higher layers, a regular problem like lattice field theory is expected to achieve good performance, so the 20% to 30% efficiency obtained now on Cray and IBM platforms might be expected to continue to the exaflops scale. Thus, we expect that lattice field theory will continue to be one of the leading applications areas demonstrating the transformative potential for extreme scale computing capability.

4.2. *Software Infrastructure*

As the complexity of both the numerical algorithms and the computer hardware grows, accessible, well-documented and verified software becomes indispensible. It is now impossible for a small group to carry out a competitive, large-scale calculation without the support of community-designed and maintained software. While this offers a substantial challenge for community-wide cooperation and organization it also offer substantial opportunities to pursue difficult science objectives by pooling resources and expertise to create a software environment that enables individual creativity and initiative. The DOE's Scientific Discovery through Advanced Computing (SciDAC) program is an outstanding example of both what is needed and what can be accomplished. This program supports the collective design and maintenance of critical routines, which can be used in the software systems of many groups to achieve improved efficiencies on commercial platforms. This software both exploits the low-level features of industry standards such as the communication protocol MPI as well as targets particular processor architectures to achieve impressive degrees of both efficiency and portability.

A second opportunity provided by the need to develop community-wide code is the chance to incorporate the best algorithms developed by computational and computer scientists. This comes more slowly because of the already high level of existing expertise in the lattice field theory community and the special characteristics of the lattice field theory problem that are not common to well-studied problems in other fields. However, recent examples such as the new multi-grid/deflation methods suggest that the collaborative efforts of applied mathematicians and lattice physicists will pay substantial dividends. Such collaborative work must be an important part of the preparations needed for extreme scale computing.

4.3. *Training a New Generation of Computational Physicists*

The size and sophistication of the code base in LQCD, as in other computational fields discussed in this report, leads to a significant hurdle for young researchers wishing to enter the field. In addition, such researchers must become expert at both computational and scientific aspects of the problems. At the same time, the field relies on the efforts of graduate students and postdoctoral fellows, since it is they who create and run most of the code, albeit under oversight of more senior researchers. Thus, investments in training, and support for cross-cutting postdoctoral fellowships are essential.

The structure of LQCD simulations has proven to be an aid to creating an organization for the necessary training and support. Since much of the computer time is spent on generating gauge configurations, it is natural to form a national (or international) "umbrella collaboration" to decide on the parameters used. In the United States, this is the USQCD collaboration. This collaboration also allocates resources to smaller sub-groups to do the (usually smaller scale) calculation of observables. It also organizes training for young researchers and a variety of workshops. It is the SciDAC funding of the software activities of this USQCD collaboration that is described in the previous section.

At present this funding typically supports postdocs with a physics background together with senior computer scientists and applied mathematicians who work with the broader USQCD collaboration to develop both algorithms and software. An important future goal is the creation of a cadre of applied mathematics postdocs appropriately trained so that they too could directly participate in this development work. Finally, it is important to emphasize that the entry level physics Ph.Ds trained in this effort not only go on to lattice field theory research positions in universities and national laboratories. Some also move to other applications areas and to the computer industry making this lattice field theory activity one of the drivers of high-performance computing internationally.

EXPERIMENTAL PARTICLE PHYSICS

Co-chair: Jim Shank, Boston University
Co-chair: Frank Wuerthwein, University of California, San Diego

1. INTRODUCTION: SCIENCE GRAND CHALLENGES

Experimental Elementary Particle Physics strives to understand the fundamental building blocks of matter and its interactions, as well as the nature of space and time. It addresses scientific grand challenges in all three areas of the three frontiers of particle physics (Particle Physics Project Prioritization Panel, 2008):

- The Energy Frontier, using high-energy colliders to discover new particles and directly probe the architecture of the fundamental forces.
- The Intensity Frontier, using intense particle beams to uncover properties of neutrinos and observe rare processes that will tell us about new physics Beyond the Standard Model (BSM).
- The Cosmic Frontier*,* using underground experiments and telescopes, both ground and space based, to reveal the natures of dark matter and dark energy and using high-energy particles from space to probe new phenomena.

To probe the smallest distance and thus the most fundamental particles and their interactions, requires the largest energies and highest intensities. In the past and foreseeable future, the dominant computing challenges in this field originate with the accelerator-based science programs because they provide the largest volumes of data.

We understand nature at a given energy or distance scale by virtue of the kinds of fundamental particles that get produced at that energy, their mass and lifetime, and the rates and processes by which they get produced and decay. To push our knowledge frontier, we need to discover new rare, high-energy particles and measure fundamental properties of these new particles as precisely as possible. This requires accelerator facilities at the technological frontier: the highest possible energies and the highest beam luminosities. Facilities such as the Tevatron, a proton-antiproton collider at Fermilab that has been operating in the past 28 years, the Large Hadron Collider (LHC), a proton-proton collider at the European Organization for Nuclear Research (CERN) being commissioned now, and future possibilities like electron-positron linear colliders or a muon collider are the type that will produce these new discoveries.

On September 10, 2008, the LHC started circulating its first proton beams. The design center of mass energy of this machine is 14 TeV, a factor 7 larger than the currently running machine, the Tevatron. This is the first such increase our field has seen in about 20 years. The increase in energy alone results in four to five orders of magnitude increase in expected production rates for collisions at the high energy frontier. To understand this strong non-linearity between beam energy increase and physics reach, one must realize that protons are compound objects made out of quarks and gluons. It is these quarks and gluons inside the protons that actually collide to create the largest energies. The probability for two of these so-called partons to carry a significant fraction of a proton's momentum each is small. In other words, the probability for a gluon-gluon collision at an energy of 1 TeV is roughly 5 orders of magnitude larger at the 14 TeV center of mass energies of the LHC than at the 2 TeV proton anti-proton collision energies of the Tevatron. The LHC has an initial design luminosity roughly an order of magnitude above the Tevatron. By 2020, the luminosity is expected to increase another factor of 100, thus leading to seven to eight orders of magnitude increase in rates for particle collisions at today's high energy frontier.

There is reason to believe that the increase in energy provided by the LHC propels us into a regime that is qualitatively different, because it pushes us above the energy scale of electroweak symmetry breaking. Crudely speaking, one might think of this as being able to explore the other side of a phase transition for the very first time. What exactly we will find on the other side is the object of much theoretical speculation and very little experimental fact at this point.

The fact is that some sort of symmetry breaking mechanism must be discovered, the most likely of which is the so-called Higgs mechanism. The search for "supersymmetry" is somewhat more speculative but still well motivated by theoretical arguments and astronomical evidence of dark matter. More speculative scenarios include the possible observation of phenomena indicating additional dimensions of space, as well as a host of other new phenomena.

The next 10 years of exploration is an era with enormous discovery potential. We are at the start of a once in a lifetime scientific opportunity (Acosta 2006, CERN-OPEN-2008-020, 2008). In the foreseeable future, we envision a next-generation electron-positron collider, referred to as the International Linear Collider (ILC). This machine will be complementary to the LHC in that its fully controlled electron-positron initial state allows for precision physics that is very difficult to impossible to do with the LHC (Weiglein et al. 2006).

Alongside the compelling physics of terascale explorations at the energy frontier, another window on discovery at the intensity frontier has opened with the remarkable recent discoveries in neutrino science and with the ability to detect new physics phenomena in ultrarare events or in the small perturbations they induce in other processes. Experiments in symmetry-violating processes and ultrarare processes via experiments such as Mu2e (Prebys 2008) (at a sensitivity four orders of magnitude beyond the current state of the art) and SuperB factory (INFN/AE, 2007) can provide windows into new mass scales of many thousands of TeV. Neutrino experiments (Fermilab 2007) may tell us about physics at even higher energies near unification or about an entirely new source of charge conjunction and parity (CP) violation that may help explain the excess of matter over antimatter in the universe. In addition, there are a number of experiments addressing the cosmic frontier via the detection of cosmic dark matter and dark energy.

The scientific questions at the intensity frontier generally do not pose additional grand challenges in data and central processing unit (CPU) intensive computing but are rather well served by the solutions developed for the LHC and ILC programs. These solutions significantly benefit from the experience at programs such as the Tevatron's CDF and DZero experiments that require high-performance computing (HPC) and networking, distributed computing infrastructure, petascale scientific data management, and scientific visualization. The dramatic growth in computing power used at Fermilab to drive the scientific program with about 100 physics publications per year, principally CDF and DZero, is shown in Figure 26.

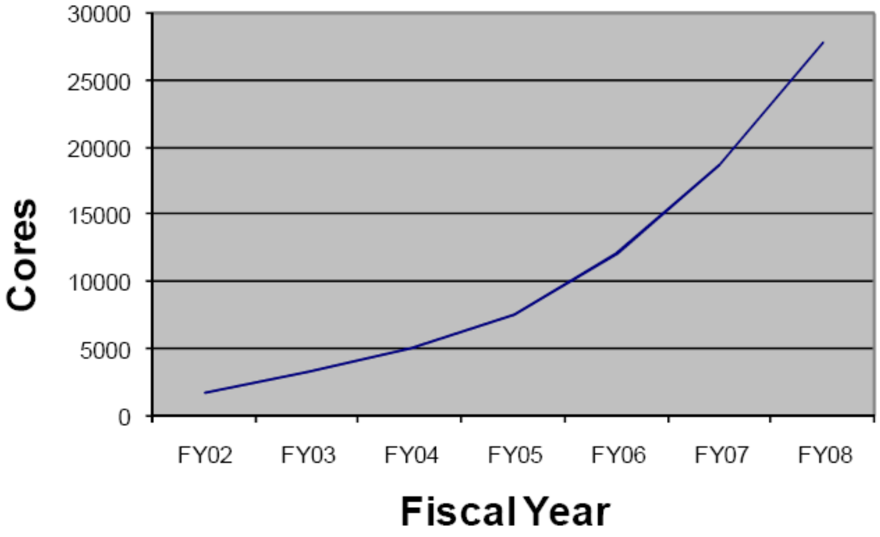


Figure 26. The growth in computing cores installed at Fermilab during the Tevatron collider era, and the build-up of LHC computing. Figure courtesy of Jim Shank (Boston University).

Harnessing this power to deliver the science required parallel growth in storage systems, now in excess of 15 petabytes at Fermilab, high-performance networking, and the emergence of powerful and robust grid computing illustrated below in Figure 27.

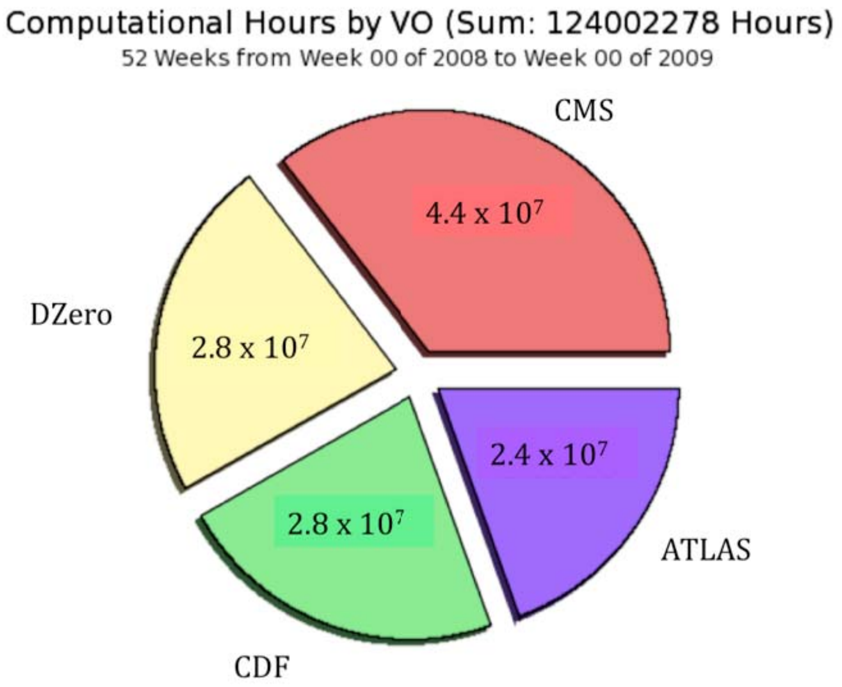


Figure 27. The current shares for the four hadron collider experiments of Open Science Grid which provides in excess of 100 M hours of computing in 2008. Figure courtesy of Jim Shank (Boston University).

Computing challenges for experiments at the cosmic frontier, accelerators at the Energy and Intensity Frontiers, and high energy theory will be addressed by the “Astrophysics Data” and “Cosmology and Astrophysics” panels, the “Accelerator Simulation” panel, and the “High Energy Theoretical Physics” panel, respectively.

2. TECHNICAL CAPABILITIES TO MEET SCIENCE CHALLENGES

The Science Grand Challenges presented above will require that we deal with extreme scale computing: exabytes of data, exascale networking, and approaching exaflops of CPU power. This move to the extreme scale will have to build on the existing cyberinfrastructure that is distributed worldwide. During the recent data challenges leading up to the start of the LHC, more than 30 petabyes of data have been stored, sustained dataflows among CERN, the Tier 1s and Tier 2s of up to 3 petabytes/month and peak transfer rates of more than 20 gigabits/second have been observed, and more than 100 sites worldwide have participated. Further increases are expected as the tools mature and the level of familiarity and expertise in transmitting data among the sites increases.

The buildup will not just be adiabatic and requires innovation in a number of areas such as:

- beyond state-of-the-art networking to replicate petabytes of data per day between computing centers around the world
- unprecedented data mining including managing and tracking an exabyte of experimental data, and providing access via globally distributed computing centers
- extraordinarily long-lived world-wide computing systems supporting a global community of thousands of experts and collaborators with 20-plus years of operations
- creative innovations in algorithms and methods to address increasing needs for simulation, and advances in hardware and software technologies.

These and other challenges are detailed in Section 3.

We first define the current working model for computing in the field as a background to defining and understanding the technical capabilities needed to address the challenges in the science. The overall workflow and system architecture is shown in Figure 28.

There are many distinct workflows in the computing for experimental high energy physics (HEP), but we may characterize them as follows: processing and reprocessing, simulation, and user analysis.

- *Reconstruction takes the digitized data from the detector systems and attempts to reconstruct particles* in terms of their trajectories, energies and momenta, and also the identity of the particles. This process requires a good understanding of the geometry and the calibration of the detector, which is typically derived from iterative partial reconstruction on a subset of the data with an initial estimate of the calibration, and the calibrations thus derived are stored in a relational conditions database. These are then read-back for the full reconstruction; part of the scheduling task for the work is to arrange the reprocessing of data in groups with the same calibrations, and also to minimize the number of files recalled from archive.

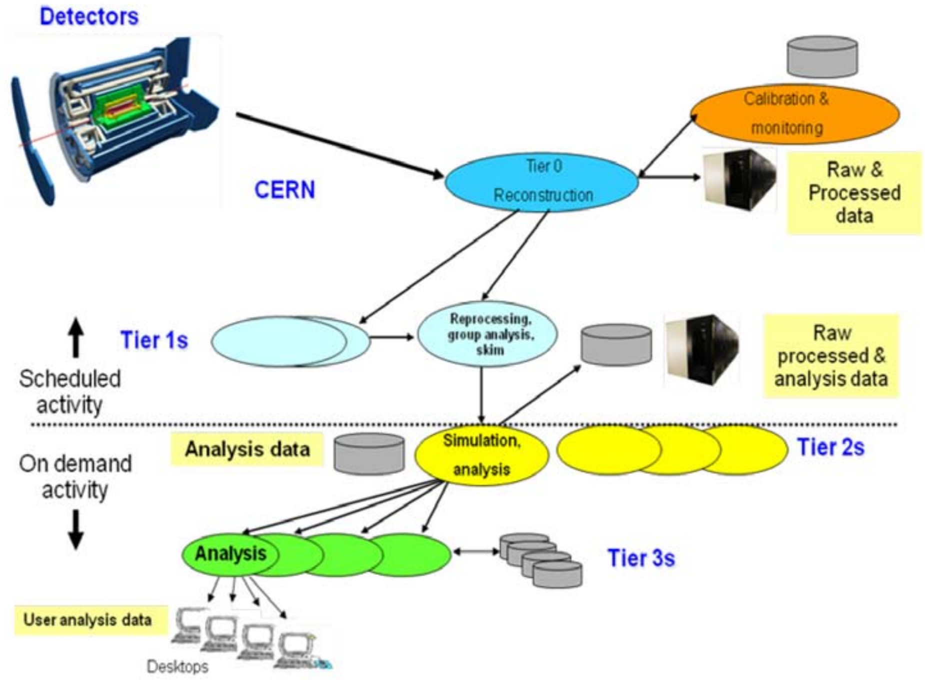


Figure 28. Experimental particle physics computing workflow and architecture.

- *Simulation attempts to model the underlying physics processes* in the collisions, and the physics of the detector response to the particles produced, and to produce the simulated digitized data that would be expected from the detector.
- *Analysis can have many workflows.* Groups distill subsets of data from the full sample by applying selections (‘skims’); they may also perform CPU-intensive specific reconstructions, and may select subsets of the data in individual events (“thins”). The end-user will typically analyze O(10) terabyte samples in the Tier 2s using 1000 cores, producing a 10-gigabyte reduced data format that fits conveniently on a laptop, and do all of this without any centralized planning. It is in the nature of the creativity of the physicist that end-user analysis workflows will be many and varied.

The LHC experiments have evolved very similar computing strategies (ATLAS TDR—017, CERN-LHCC-2005-023, 2005) driven by the same sociological and technological drivers. The problems in the simulation, the reconstruction of the real raw data from the detectors, and the end-user analysis are all open to coarse-grained parallel solutions, and so have all exploited multi-core processors in that way. The LHC experiments have developed a petascale distributed computing model, where data are distributed, processed and served by a dozen national Tier-1 centers, and more than 140 university and laboratory-based Tier-2 computing and storage facilities, supporting the efforts of physics groups and their local Tier-3 computing clusters at hundreds of sites around the world.

CERN will have a special role in the solutions, providing a Tier-0 facility to perform prompt monitoring, automated calibration, and a first-pass reconstruction a few days after data-taking. CERN plans to provide facilities with swift access to the new data in raw and poorly calibrated trial reconstructed formats to support prompt, human-driven calibration activities that allow a reasonably calibrated first pass full reconstruction on the required short timescale.

The longer-term reprocessing of the data and the custodial role for most of the data-sets reside with major national computing facilities, called Tier 1s. These also provide for scheduled access to the full data samples to the physics and detector performance groups. Summary data sets are distributed (in several copies) to smaller facilities, called Tier 2s, which is where most of the individual physicists analyze the data. These facilities also have the important role of providing simulation capacity for the experiments. Further reduced data sets are taken back to local facilities, called Tier 3s.

Current experiments such as the CDF and DZero experiments at the Tevatron and BaBar and Belle experiments at B factories have already been performing with globally distributed computing infrastructure. The future experimental high energy physics projects, such as the ILC, the neutrino experiments, the muon experiments, and the heavy flavor experiments, also have globally distributed communities and currently plan to adopt similar strategies.

3. PRIORITY RESEARCH DIRECTIONS

3.1. *Unification of Forces, Exploring the Higgs Phenomenon and the Electroweak Scale*

Precision measurements from present and past colliders such as the Tevatron, the Large Electron-Positron Collider, and the Stanford Linear Collider predict the existence of a Higgs particle below 200 GeV in the framework of the Standard Model, and thus well within the reach of the LHC. One of the primary goals for the first five years of data taking with the LHC is thus observation of Higgs boson production, and measurement of its couplings. Precision measurements of these couplings as well as spin and parity of any Higgs boson(s) found at the LHC will most likely require a future lepton collider (Weiglein et al. 2006).

3.2. *What is the Nature of Dark Matter?*

We have ample evidence from astronomical observation that dark matter exists and makes up a large fraction of the matter in the universe. To understand the nature of dark matter particles, we need to produce them in the laboratory to measure their mass, couplings, and quantum numbers. The abundance of dark matter measured provides some indication of an order 100 GeV mass scale for dark matter particles, and thus might be produced at significant rates at the LHC. The favorite candidate theory for a dark matter particle that might be observable at the LHC is supersymmetry. Superparticles would be detected at the LHC most likely via a variety of complex decay chains. Once discovered, gaining a detailed understanding of these particles and their decay chains will almost certainly require a future lepton collider (Weiglein et al. 2006) and flavor experiments with rare decays and precision measurements.

3.3. *What Are Neutrinos and Ultra-Rare Decay Processes Telling Us?*

The last decade saw a revolution in our understanding of neutrinos. We discovered that neutrinos have mass by observing neutrino flavor oscillations. This opens up the possibility that mixing in the lepton sector might be responsible for the striking matter anti-matter asymmetry observed in the universe today, and is required by cosmology. The next decade will focus on precision measurements of neutrino oscillations as part of a general assessment of the feasibility of a future accelerator based neutrino program to measure CP violation in the neutrino sector. While lepton flavor violation was discovered in neutrinos, we do not know why lepton flavor violating (LFV) occurs or if it is related to the flavor violation seen with quarks or to new phenomena at the terascale. A related key question is whether LFV also occurs with the charged leptons such as muons and tau leptons. Theoretical models that incorporate ideas such as unification, supersymmetry or heavy-neutrino mixing predict charged LFV at rates that could be within reach of new experiments. Combined with results from neutrinos and the LHC, these experiments could point the way to leptogenesis or unification.

3.4. *Search for Extra Dimensions, Micro-Black Holes, and Other Speculative New Phenomena*

Many of the revolutionary paradigm shifts in science are the result of discoveries based on speculative, rather than programmatic, research. Large accelerator programs such as the LHC and a lepton collider are multi-purpose facilities built with programmatic and speculative research in mind. It is the speculative part of the program that generally imposes the larger demands on computing infrastructures via the requirements of broad and open access to large data volumes and computing resources by the entirety of the globally distributed scientific collaborations.

4. FUTURE COMPUTING STRATEGY

The required computing capacity for the experiments is set by the precision required to achieve the physics goals, as motivated above, which in turn determines the total number of events to be collected. In the current phase, the number of events is determined by the live time, with a fixed rate of events selected by the trigger systems. At some point, the accepted rate of events from the trigger must rise to avoid throwing away important physics processes at the trigger level. This means the important quantity determining the rate of recorded events becomes the luminosity, which is a measure of the beam intensity. For the purpose of this report, we assumed the trigger accept rate to start increasing for luminosities in excess of 10^{34} cm^{-2} sec^{-1}. This is motivated by the observation that the rate for Z to dileptons will start to dominate the trigger at that level. Supersymmetry, as well as Higgs, has dilepton signatures at the trigger level. To pursue two of our priority research directions thus motivates our data volume extrapolations.

The projected luminosities are given in Table 5 where we assume a shutdown for detector and accelerator upgrades in 2017, after which a new phase of data taking would begin and reprocessing of earlier data largely cease. To broadly characterize the results, there is a good degree of agreement between the outcomes for the two experiments, and at CERN, in eleven major national centers (Tier 1s) and over 100 shared university facilities (Tier 2s) there is a growth of between two and three orders of magnitude in the required processing power, active and archival storage. While no explicit calculation was made for the non-shared (Tier 3s) resources required, a similar scaling to that seen in the Tier 2s is to be expected. It is worth noting that the assumption that the output rate from the detectors scales with the specific luminosity after 2013 affects the projections over a decade by only a factor of around three, although it would become more significant thereafter.

Table 5. The assumed specific and integrated luminosity profile from 2012-2020.

	2012	2013	2014	2015	2016	2017	2018	2019	2020
Specific luminosity ($10^{34}cm^{-2}s^{-1}$)	1	1.5	2	2.5	3	0	5	8	10
Integrated luminosity (fb^{-1})	108	198	318	468	648	648	948	1428	2028

The sum of the required resources (World Wide HC Computing Grid Management Board web pages) at CERN, the Tier 1s and the Tier 2s for A Toroidal LHC ApparatuS (ATLAS) (ATLAS TDR—017, 2005) and Compact Muon Solenoid (CMS) combined is given in Figure 29. An attempt has been made to convert the processing requirement into a number of cores.

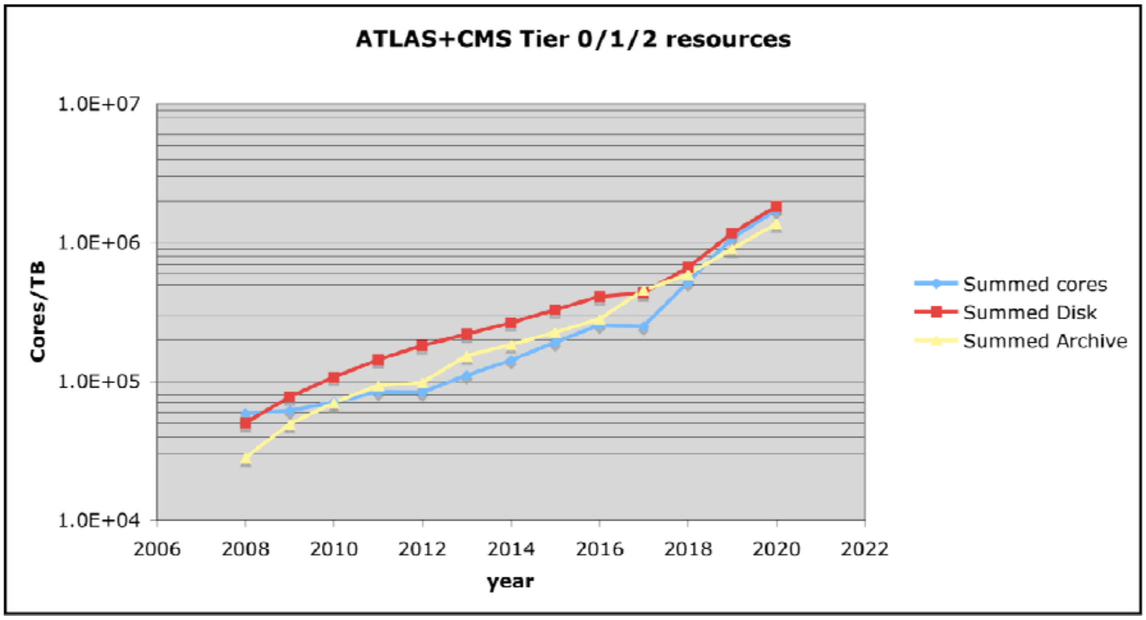


Figure 29. The projected resource needs for ATLAS and CMS combined at CERN and in the Tier 1s and the Tier 2s. The capacity in local Tier-3 facilities is not included. Figure courtesy of Jim Shank (Boston University).

To illustrate the extreme scale of the requirements, the Tier-1 capacity by 2020 of each of the two experiments, ATLAS and CMS, would be about 1 million cores and one thousand petabytes of active storage. The Tier-2 capacity numbers are similar. While these projections would be met by a Moore's Law growth with a doubling every 1.5-2 years, they present many technical challenges to allow the resources to be exploited, and will entail significant support from the host institutions.

Neutrino experiments, the Mu2e experiment, the Super-B factory experiment and the ILC experiments have also been considered in these resource projections. Their requirements will be less than 10% of those of ATLAS or CMS in 2020. These projects may pose particular challenges in the level of precision required in their simulation and algorithms, which is a different but equally real technical challenge.

Consideration has also been given to the networking and bandwidth requirements implied by the growth in resource requirements described above. Some of the issues will be discussed in later sections, but one obvious challenge that presents itself is the need to reprocess the raw data stored in archival media at the Tier 1s. This must be done to benefit from improving understanding of the detector calibration and alignment and improved reconstruction algorithms that will be obtained from study of the initial processing of the data. At a typical Tier 1, this will start by requiring O(100) megabyte/s average recall rate from archive, but after a decade this will be more like O(10) GB/s. This would seem to pose a challenge for any sequential access medium like tape, and indeed may be hard with any mechanical storage medium.

A similar consideration applies to the access to the data for on-demand user analysis at the Tier-2 facilities. In the present architecture, the data are typically accessed locally from a storage element by processes on the cores. Each process reads at between 2 and 30 megabyte/s. As the number of cores grows, the aggregate average rate will grow from Q(1) Gb/s per experiment today to O(20) Gb/s in a decade. This can face many potential bottlenecks, some of which will be addressed below.

4.1. *Research Priorities in Computing*

The rate of information flow off the large LHC detectors will approach hundreds of petabytes per second by 2020. Real-time data reduction and filtering on the detectors themselves leads to an expected 7 to 8 orders of magnitude reduction in data volume to a manageable 10 to 100 gigabytes per second to archival storage. The typical duty cycle of past accelerators is 30%, leading to exabyte scale yearly volumes of recorded data by 2020. This is a factor of approximately 100 greater than early LHC running at its initial design luminosity. The computing strategy described above enables scientists to develop their analyses on laptops and desktops anywhere in the world, and use these platforms to steer global data movement, both small and large scale, as well as submit and monitor workflows that operate transparently on a globally distributed infrastructure.

4.2. *Extreme Wide Area Networking*

HEP has undergone a revolutionary paradigm shift within the last five years, which depends completely on highly capable, scalable, high speed, highly interconnected, and very reliable networks. Within the United States, two U.S. Department of Energy (DOE) Office of Science laboratories—Fermi National Accelerator Laboratory (FNAL) and Brookhaven National Laboratory—are the Tier-1 repositories of the data from the LHC. The amount of data flowing first from CERN via USLHCNet, then via ESnet to the Tier-1 centers, and then from there to the mostly university-based data analysis (Tier 2 and Tier 3) sites, produces several orders of magnitude more network traffic than any past science use of Wide Area Networks networking. Even during the testing phase of the LHC data handling systems, network traffic was generated at the rate of 4.5-9 Gb/s, sustained 24 hours a day for several months, with peaks in the 15 to 20 Gb/s range. Networks supporting the distribution of the datasets produced to the Tier 2s have been of comparable size, often reaching full use of a 10 Gb/s link in the case of several of the U.S. Tier-2 sites.

HEP has engaged in the development of state-of-the-art tools for long-range data transfer over the last decade. Demonstrations in 2007-2008 among relatively small clusters of storage servers indicate that the next generation of 40 Gb/s and 100 Gb/s links can also be fully used.

To support this sort of next generation large-scale science, the Office of Advanced Scientific Computing Research in the Office of Science funded ESnet to design and build a completely new network with a new architecture specifically tailored for science like that of HEP. The new network, ESnet4, was based on use-cases and requirements that were identified from 2003 to 2005. The network took about 18 months to build and now provides about 20 Gb/s throughout the United States, connecting the laboratories to other

U.S. and international research and education networks. The current network is designed and funded to grow to about 50 Gb/s in late 2010.

The currently scoped networks will handle the load probably for the next several years (approximately 2010). Beyond this, the network capacity will have to increase considerably over the original plan. This expansion cannot happen by just adding many more 10 Gb/s optical circuits for two reasons. First, the cost would be prohibitive. Second, the optical network infrastructure that ESnet is built on is shared on a dedicated optical fiber with Internet2 (U.S. research and education network). By 2010, this optical infrastructure will be approaching its capacity and it is not practical to obtain a second complete set of fibers around the country.

In order to increase the capacity of the network until the next generation network is built (in the 2015 to 2017 timeframe), several new approaches are needed and these will require research, development, and deployment. The currently most promising approaches are

- the dynamic management of optical circuits thus allowing their integration with the user transport layers of the network
- increasing the current 10 Gb/s per optical circuit to 100 Gb/s per circuit capacity
- the transparent, selective, and dynamic re-routing of in-transit data flows from one part of the network to another.

All of these technologies are designed to maximize the use of the entire current optical fiber infrastructure. Another topic that is important for the effective use of the network by the science community is highly capable, “universally” deployed, end-to-end (user application to user application) network monitoring across all of the intervening network domains (e.g., ESnet, Internet2, GÉANT).

4.3. *Data Management at the Exabyte Scale*

A tremendous challenge within the HEP computing strategy is data management. Principal elements of the data management challenge are in cataloging, movement, storage systems, integrity strategies, automated flow, and fault tolerance.

- Data cataloging provides for data discovery based on selection criteria; higher level data organization such as data hierarchies (e.g., datasets aggregating files) and metadata tags; and location tracking and access information. Common services today consist only of a file catalog that relates logical file identifiers to physical location. Higher level services for data aggregation and categorization (datasets) and metadata systems have been implemented in the experiments. These services are presently file-centric; future developments such as the Large Synoptic Survey Telescope (LSST) initiative in a massively scalable scientific database, with a database-centric organization, may have greater scalability and utility for analysis at the exabyte scale.
- Data movement services provide for managed, scheduled, optimized, and fault-tolerant replication of data among sites and among storage services within sites. The common layer now consists of file replication tools such as file transfer service with some scheduling and throughput optimization capabilities. Full-fledged managed replication systems that offer the robustness, scalability, and manageability required by the LHC experiments do not yet exist in the common layer and have been implemented independently by the experiments.

- Data storage systems as seen by data management services consist of robust, fault-tolerant data repositories accessed via standardized interfaces that hide the heterogeneity in the underlying storage facilities. Present services such as Storage Resource Management storage interface and its underlying implementations in storage systems go some way towards creating the needed uniformity in storage services but without the robustness, homogeneity, and transparency that is required to scale above present deployments.
- End-to-end data integrity strategies need to be developed for exabyte data volumes. To set the scale, the expected media error rate is 10^{14} as compared to the exascale of 10^{18}. A 2007 CERN study found end-to-end byte error rates of as high as one error per 3 x 10^{7} bytes.
- Automated dataflow, addressing the data intensive nature of HEP processing, demands that data flow be closely coupled to processing workflow and carefully managed for overall processing efficiency. A high level of dataflow automation is required today and even more so as we approach exabyte scales in order to have an overall workflow that is maintainable and sustainable with a reasonably sized operations effort. Dataflow automation must adhere to and implement overall computing strategy while providing sufficient flexibility to rapidly adapt to contingencies and priorities.
- A high level of *fault tolerance* is required in order to sustain a robust, high-availability data service above a data management infrastructure made up of storage services at hundreds of sites, with the accompanying unavoidable service failures and downtimes.

The principal requirement of HEP data management today and in the future is effective support for distributed collaboration in extracting the science. The complexity of this global distribution must be hidden from the user, who must be provided location-independent transparency in their access to and use of the data. Intelligence and automation is needed in the distributed processing and data management systems to transparently mate the data analysis requests of users with collocated data and processing resources that can be used to fulfill the request, whether done by routing processing to the data or replicating data to the optimal processing site.

4.4. *Systems for Large-Scale Global Collaboratories*

LHC data analysis will go beyond 2020. More than 5,000 physicists work collaboratively to extract the discoveries from the data acquired. Our most valuable resources in this are the scientists and engineers making the science discoveries; developing and using the more than 10 million lines of algorithm and framework codes; accessing, integrating and supporting the data analysis, storage, and distribution systems. The World Wide LHC Computing Grid is the globally distributed system for distribution, processing and analysis of the data onslaught. The Open Science Grid today provides the underlying facility in the United States. This system works today sufficiently well for the goals of the first year of LHC data taking. Maintaining this model is essential not only for the success of the collaborative science, but also for the longer term buy-in of the many government and funding agencies involved.

The system must grow by a factor of 100 or more in size and performance over the next decade. Managing this change while sustaining the availability and efficiency is critical to meeting the needs and maintaining the commitment of the global community. During this same period, physicists in the experiments at the Tevatron will continue to analyze their multi-petabyte data sets, neutrino experiments will be delivering results, and developments for the ILC and experiments at the intensity frontier will be ramping up. New approaches, listed together with some example focus areas, are needed to:

- Support distributed development and integration of large complex code sets for 20-plus years, where of order 1,000 developers contribute to a common code base
- Scale and operate a highly available, efficient, secure global infrastructure across more than a decade of change
- Sustain system support and interoperation through independently evolving technologies and organizational structures in Europe, Asia and the United States and scaling of the distributed facilities in the United States by more than a factor of 100.

4.5. *Managing Change*

Over the 30-plus years of lifetime of a large modern particle physics experiment, computer hardware and software transition dramatically, and in ways that are impossible to predict when the software and computing model are initially developed. We rely on commodity computing hardware and have to adopt new technology trends as they become available for cost-effective handling of our data volume. To maximize our ability to extract physics from the data, we must also adapt to new software techniques as new generations of scientists join these experiments.

HEP detectors generate complex data that take millions of lines of code, produced by 1,000 authors, to simulate, reconstruct, and analyze. Efficient access to the large volume of data requires custom storage techniques. As a result, the process of migrating to new technology is often quite disruptive and requires a large amount of code to be re-engineered.

Handling this process of change could potentially be made less disruptive by designing for change from the outset, for example, by designing data storage formats less tightly coupled to specific technology choices. These issues could potentially benefit from cross-experiment/laboratory collaboration within HEP, as well as with wider collaboration with computer scientists, engineers, and other scientific disciplines which are, or will soon be, facing challenges of handling similar volumes of complex data.

4.6. *Planning for Long-Term Data Stewardship*

Historically, the accumulated data sample for particle physics experiments is actively analyzed for approximately five years after data-taking ends, then is archived for an additional 10 to 20 years to provide an opportunity for cross checks of discoveries from future detectors or validation of a future analysis hypothesis. The long time scales of data stewardship lead to interesting computing challenges in the field. The high data volume and long time scale dictate the need for extremely low data corruption rates to ensure the integrity of the data sample. There is interesting work left to do in this area. The data archive has traditionally benefited from continuing improvements in storage density and the accumulated data are migrated to newer technology and must be readily accessible with low human effort during its working life. The technology migration simplifies the data storage problem and extends the life expectancy of the archive into the range needed, but it comes with the cost of additionally maintaining data integrity through technology migrations.

Even more challenging than the archiving of the data sample is capturing the expertise and accumulated institutional memory needed to derive meaningful physics results many years after the initial data collection. The code used to reconstruct and analyze the data is the result of the effort of hundreds of individual developers. During data collection, the experiment relies on a similar-sized team for calibration, validation, and detector operations. While writing the data files, the conditions databases, and the software into an archive is a manageable technical problem, being able to record the complete provenance of a published analysis for independent execution at a future time or provide for a reliable new analysis on old data from the archive has not been achieved except in some specialized cases. Collaborative development effort is needed in managing the complexity, recording the activities, and capturing the activities of a large collaboration.

4.7. *Simulation and Algorithms*

High energy physics research as we know it today would not be possible without simulations and powerful analysis techniques. The massive production of event samples similar to those expected in the real experiments is an integral part of the process to design, build, commission, and operate the highly complex accelerators and detectors used in experimental particle physics. Simulations are also essential to develop the software tools, stress-test computing infrastructure, and analyze the data collected by the experiments. Progress on the three frontiers of particle physics relies critically on these techniques. A few notable examples are described here.

- *The Energy Frontier using present and future colliders:* As analyses are performed on larger and larger data sets, the uncertainty of finite statistics will decrease and the precise modeling of the stochastic processes that give rise to the measured detector response will become increasingly important. For example, precise modeling of the energy deposit of hadrons in detectors “hadron calorimetry” will limit our understanding of new particles produced at the primary collision point that subsequently decayed into jets of hadrons. Simulating the response of hadron calorimeters can be compute intensive, and the issue of generator validation against existing control data sets will become increasingly important. The detailed reconstruction of charged particles and energy flow from collision event data is computation intensive. The corresponding reconstruction latency per event strongly affects the computing strategy and the rate of progress in analyzing data and extracting science. These reconstruction algorithms were originally optimized against processor and memory constraints that do not exist today and which are rapidly evolving. Application of computer science and applied mathematics expertise to re-optimize and explore reformulating the problem could yield a substantial return in analysis throughput.
- *The Intensity Frontier for neutrino physics:* Design decisions regarding research facilities in the next decade will heavily rely on modeling and simulation tools that today are not state-of-the-art. The detailed modeling of hadronic interactions and detector response is important to the eventual experiment design and will benefit directly from concurrent progress on these same issues at the Energy Frontier.
- *The Intensity Frontier modeling ultra-rare decay processes:* An experimental program of searching for and measuring ultra-rare decays at future “flavor factories” and “intense muon beams” aims for sensitivities of one part per trillion (10^{-12} - 10^{-14}), and even less in some cases. Direct modeling of background processes would consequently require simulating in excess of one trillion particle decays that to date has been out of reach. This barrier has historically led to modeling backgrounds with factorization methods that are less robust than direct simulation. Emerging computing technologies and advanced algorithms could break through this barrier, which would dramatically impact the design and prospects of these ambitious experiments.

- *The Cosmic Frontier:* Progress at the Cosmic Frontier is intimately tied to the Energy Frontier and Intensity Frontier through, for example, the discovery and detailed study of dark matter candidate particles at the LHC, and ultimately the ILC, and directly elucidating the properties of neutrinos at the Intensity Frontier. The general theme is that we need to recreate, in the laboratory, the particles that drive cosmology to study their properties in detail and develop a consistent theoretical framework to understand them.

Experimental particle physics has standardized on the Geant4 toolkit for particle-collision and detector response simulation. Geant4 requires research and development to improve its precision and performance. Some of the most critical precision issues relate to improving models of hadronic processes – a crucial research and development topic that requires work by high-energy, nuclear and other physicists rather than computer scientists. However, performance issues also effectively limit precision, for example, high-precision simulation of the almost perfectly understood electromagnetic processes, responsible for the majority of the energy deposited in detectors, can easily become impossibly costly.

4.8. *Adjusting to Changing Technology: Multi-Core*

Experimental HEP data are organized in “events,” each corresponding to one collision between to incoming particles. Simulation, reconstruction and analysis programs process one event at a time. Event processing programs are composed of a number of algorithms selecting and transforming “raw” event data into “processed” (reconstructed) event data and statistics. Algorithms may require additional “detector conditions” data (calibrations, geometry, environmental parameters, etc.). Statistical data (histograms, distributions, etc.) are typically the final data processing results. All these algorithms are managed by a software framework that schedules them according to their dependencies and manages the data they consume and produce.

Events are independent of each other and therefore trivial to process in parallel. In this respect experimental HEP has cost-effectively exploited high data-throughput computing on approximately 100 thousand cores. On the current generation of quad-core machines we are still able to run at full efficiency at the cost of using about 2 gigabytes of memory for each process. Part of this memory (about half) is used to store “detector conditions” data that are common to all events acquired in a small time frame. A first way to alleviate the cost of memory will therefore be to share these common data among the event-processing tasks running on one node. This technique can be implemented by exploiting modern operating system features without major reengineering of application software, thus allowing event processing to scale to the near future generation of 16 to 32-core processors within an affordable memory footprint.

Scaling to processors with many more cores will, in contrast, require innovative software solutions to bring parallelism below the event level. Part of this parallelism can be achieved at the level of the algorithm scheduler in the “framework.” This has the advantage that modifications will be confined to an area of software mainly developed by experienced computing engineers without touching the core of the scientific computation developed by physicists. The major gain will nevertheless come from parallelization of the simulation, reconstruction and analysis of algorithms themselves. Pattern recognition, particle tracking, and simulation of the detector response are typical algorithms that present great opportunities for parallelization. To achieve a level of efficient parallelism at the scale of future, many-core processors may require changes in the way these algorithms are implemented. A more global redesign most likely will be required.

Traditional HEP approaches to physics analysis are based on the selection of a region in the multi-dimensional space of the data that maximizes the difference between signal and background and where the background itself and the acceptance of the detector are well understood. This procedure produces relatively small data samples that are easy to manage and allows the fast processing turnaround essential for the individual physicist to explore procedures and methodologies, and understand systematic errors. However, a large fraction of the data sample carrying valuable information about the physics under investigation is rejected. Advances in simulation will allow a better understanding and modeling of the backgrounds and of the detector response. This will open the opportunity to enlarge the region of multi-dimensional space used in physics analysis to essentially the whole data sample. To fully exploit the potential of such large datasets and the rich detailed experimental information that modern detectors provide requires the use of multivariate classifiers – neural networks, decision trees, likelihood functions based on matrix elements – sensitive to the small kinematic differences between signal and background processes. Already in present experiments, these techniques are currently limited by the turnaround time for each single iteration in the optimization (fitting). Prototypes attempting to parallelize the software implementing these multivariate classifiers and their optimization have been successful.

4.9. *Exploitation of Evolving Multi-Level Storage Hierarchy*

The technologies in the storage hierarchy relevant to exabyte particle physics datasets are changing in evolutionary and revolutionary ways. Within processors, the cache hierarchy is deepening in an attempt to hide latency and bandwidth limits in multi-core architectures. Magnetic tape, at the bottom of the particle physics data hierarchy for over four decades, has a changing and uncertain role. In the middle of the storage hierarchy, solid-state mass storage promises to have a transformational impact, while at the same time it poses widely recognized computer science challenges and lacks, as yet, tape's resistance to cyber attack.

Physics at the energy and intensity frontiers requires high-statistics understanding of measured physics backgrounds and of detailed detector performance. This understanding can be deepened when physicists have unfettered access to all relevant data in the exabyte-scale datasets. "All relevant data" are likely to be a small but constantly evolving subset of the complete set such that sufficiently intelligent management of data within the multi-level storage hierarchy could render access to this subset largely painless.

Effective exploitation of the entire cache hierarchy will be vital, but a customized particle physics approach to the use of each of the 6 to 10 levels of storage would be infeasible. The most promising and challenging new development is the emergence of solid-state storage as a financially and technically viable part of the storage hierarchy. Particle physics data analysis has the write-once-read-many characteristics that match the performance and reliability features of current solid-state devices. The full transformational effects of the new technologies will require an evolution of the particle physics approaches to data access.

5. CROSS-CUTTING RESEARCH DIRECTIONS

5.1. *Research Within HEP*

As never before, progress on the energy, intensity, and cosmic frontiers is fueled by advances in moderncomputing. Petascale computing is now the standard of the field, and full exploration of these interlocking frontiers will require reaching the exascale about 2020. The Energy Frontier is particularly data-intensive and is leading the way for enabling advances in future intensity and cosmic frontier facilities including

- advances in the comprehensive modeling of detector responses available now with the Geant4 toolkit permit precise modeling of next generation accelerator based neutrino experiments and rare-decay experiments and space-based telescope platforms
- advances in data acquisition and data management that enable consideration and design of billion-pixel array telescopes both on the ground and in space
- advances in high performance networks and grid technologies that enable access to world-wide computing resources now exploited by the gravitational wave and survey observatories.

5.1.1 Handling Exabytes

The impact of Moore's law on experimental science tends to be exponential growth of data volumes. It is thus not surprising that astro- and particle physics experiments share a concern about data management, storage, and transfer.

- **Networking:** The envisioned growth in network provisioning does not match the projected network bandwidth needs. Both experimental HEP and climate bandwidth consumption grow faster than the network provisioning.
- **Local input/output:** We foresee significant input/output challenges for several of our typical data access patterns, across the entire caching hierarchy from archival storage to chip caches. We note with concern that developments in the HPC area are going towards smaller Amdahl numbers, while for most of our applications the Amdahl numbers stay the same and maybe even increase for some workflows like skimming. Our most aggressive access patterns appear to be more like Ebay's operations than BlueGene's.
- **Data management:** The exabyte data volumes of the future require new more sophisticated software approaches to data management including new strategies for cataloguing, replication, automation, and fault tolerance.

5.1.2 Global Collaborative Science

Distribution of the intellectual enterprise, of which computing is a primary enabler, is critical to the success of these global collaborations of unprecedented size. The use of local resources enables leverage of the infrastructure and expertise at collaborating institutes and provides a sense of ownership. Cosmic Frontier experiments might encounter similar requirements in the future.

5.1.3 Change Management for Multi-Decade Long Projects

It is a common feature of both data and algorithms to have a lifetime of use that is far longer than the lifetime of any of the technologies that they are used on or with. This leads to a host of common challenges ranging from data stewardship to software porting. These challenges are particularly daunting looking into the future as the rate of technological change appears to accelerate.

5.1.4 Effective Use of Multi-Core

The change in Moore's Law scaling from increasing clock cycles to increasing number of cores at fixed clock speed is presently leading to a global research and development effort to parallelize software and algorithms in experimental HEP. This effort, as well as increased attention to algorithm performance in general, would undoubtedly benefit from increased collaboration with computer science, applied math, and engineering scientists.

5.2. *Non-HEP Science*

Table 6 gives a sample of non-HEP science showing where the computational research challenges for experimental particle physics have cross-cutting applicability. We mark (**) specific research directions where application of mathematical techniques and expertise has the potential for achieving fundamental breakthroughs and innovation.

Table 6. Cross-cutting applicability of the computational research challenges in experimental particle physics.

	Climate	Life Sciences	Nuclear Physics	Fusion	Chemical Science	Materials Science	Environ-ment
Capacity and throughput of Wide-Area Networking	Y		Y	Y			
Data management at the exascale**	Y		Y	Y			Y
Globally distributed collaborations	Y		Y	Y			Y
Change management for multi-decade long projects**	Y		Y	Y		Y	Y
Long-term data stewardship	Y	Y	Y	Y	Y	Y	Y
Massive simulations of physical processes bound by large-channel detector/environment descriptions**	Y		Y	Y	Y	Y	Y
Algorithm development to take advantage of to new hardware technologies (multi-core, GPU, etc.)**	Y	Y	Y	Y	Y	Y	Y
Exploitation of evolving multi-level storage hierarchy**	Y	Y	Y	Y		Y	Y

5.2.1 Mathematics, Software, Hardware

Most of the experimental particle physics computing challenges either already or in the future has commonalities with cross-cutting research in one or more of mathematics, computer science, computer hardware, and software research.

5.2.2 Exascale Networking

Beyond developing and deploying advanced technologies in the current infrastructure, it is important to look at the technology for the next generation of network that must be designed and deployed by 2010 to

2017 in order to support the anticipated large-scale use of the network by HEP as well as the other science communities of the Office of Science such as, climate science, coupling of supercomputers as they cooperate on problems too large for a single machine, the international fusion energy experiment (ITER), and others yet to emerge. All of this will require research and development, and the deployment of new network technology, and a highly cooperative relationship between the research and education network community that supports large-scale science, and between the research and education community and industry.

5.2.3 Data Management Technologies

Meeting the data management needs of today's LHC experiments has provided an experience base and a set of lessons that inform the far greater challenge of a 100-times scale-up to exabyte computing over the next decade. Middleware currently provides the basic service layer required to store, replicate, and access data. Higher-level functionality providing sophisticated and scalable data cataloging, fail-safe data movement, highly robust and transparent distributed data storage with sufficient ease of use, and dataflow automation remain largely at the drawing-board or prototype level in terms of the common middleware layer. These higher-level functionalities are critical for scaling up capacity, robustness, and usability as needed to reach exabyte scale capability in data management. They offer a rich opportunity for a continuation into the exabyte era of the HEP-computer science collaboration that has successfully brought HEP computing to the threshold of LHC operation.

5.2.4 Distributed Computing

Techniques to maintain expertise, software robustness, and correctness through multiple transitions in workforce; toolkits and processes to support software rebuilds from legacy releases for validation and confidence levels; ubiquitous frameworks for performance, security, and fault analysis; these are all areas where applied mathematics, computer science and engineering continue to make significant contributions to the lifecycle management of the distributed computing software used in HEP. Methods to integrate the users environment from the personal laptop, through the local and regional computing clusters, to the high end capability machines such as:

- virtualization techniques to ensure robustness of use across heterogeneous changing configurations of resources and services
- improvement in the effectiveness and fault-tolerance of job meta-schedulers and demand driven data replication and access
- full integration of federated, local identity management, models, policies and processes for international trust and policy exchange are being developed, implemented, and deployed in collaborations between computer scientists and particle physicists. All of this is done in a global context, engaging with and contributing to the change in organizational structure in Europe and Asia toward National Grid Infrastructures or Next Generation Internet (NGI).

5.2.5 Improving Simulation Frameworks, including Geant4

Research and development is required to optimize the algorithms that are used to represent the complex geometries of detectors, and to optimized Geant4's internal data and task management for the next generation of processor architectures. With the possible exception of hadronic physics models, all this research and development will benefit from a close collaboration with computer scientists and applied mathematicians.

5.2.6 Improving Algorithms and Methods

In experimental particle physics, data analysis algorithms are mainly developed by physicists who have deep knowledge of the problem but often lack professional computing training. The software design and implementation is often very sequential and in many cases poorly matches the architecture of modern CPUs. Collaboration with computing scientists with specific experience in efficiently implementing scientific algorithms on super-scalar and vector processors has the potential to improve the performance of HEP software, both the single-core and multi-code levels. Examples include developing common software libraries or just methodologies for basic algorithms to be used in parallel applications such as random number generators, linear algebras, and operations on data-structures. Parallel applications will stress even further the need for validation, verification and reproducibility. In this area, HEP could benefit enormously from experience in other fields that have long been using parallel applications (particularly for stochastic simulations).

5.2.7 The Use of Massive Parallel Computing Facilities

The use of massive parallel computing facilities has the potential to transform the way individual physicist run the optimization (fitting) of multivariate classifiers on very large datasets. The goal here would be to maintain today's turnaround time of a day or less to match the human work cycle, but for data samples much larger and fitting problems much more complex. This goal can only be reached by an extensive research and development program involving computer science and HEP scientists working together to implement efficient parallel algorithms able to work on very large datasets.

5.2.8 Exploitation of Multi-Level Storage Hierarchy

There will be clear benefits in research and development that is common to many or all particle physics experiments and non-HEP science as shown above. The commonality with a wider range of cache-management issues in science and commerce needs to be quantified. The computer science issues include predictive cache management in a read-dominated environment and combining externally provided policy and priority constraints with automated predictive management.

5.3. *Enabling Sustained Human Capital*

It is increasingly difficult to maintain a sufficiently well-trained workforce to tackle software, data, and computing challenges. While there is certainly a lot of intellectual interest, it is often difficult to find the people to turn ideas into products, and to maintain, support, and evolve them for the lifetime of our experiments or algorithms. We suspect this to be a problem across all panels, and most likely across many scientific domains.

REFERENCES

Abel,T; GL Bryan, and ML Norman. 2002. "The Formation of the First Star in the Universe," *Science*, 295, 93.

Acosta D, ed. 2006. *CMS Physics Technical Design Report Volume I: Detector Performance and Software.* CERN-LHCC-2006-001. The CMS Collaboration. Accessed February 2, 2009 at http://cdsweb.cern.ch/record/922757/files/lhcc-2006-001.pdf.

Amdahl GM. 2007. "Computer Architecture and Amdahl's Law." *IEEE Solid-State Circuits Society News* 12(3):4-9. http://www.ieee.org/portal/cms_docs_societies/sscs/PrintEditions/200707.pdf.

Amundson J, W Pellico, L Spentzouris, P Spentzouris, and T Sullivan. 2007. "An Experimentally Robust Technique for Halo Measurement using the IPM at the Fermilab Booster." *Nuclear Instruments & Methods in Physics Research A*. 570(1): 1-9.

Antonio DJ, PA Boyle, T Blum, NH Christ, SD Cohen, C Dawson, T Izubuchi, RD Kenway, C Jung, S Li, MF Lin, RD Mawhinney, J Noaki, S Ohta, BJ Pendleton, EE Scholz, A Soni, RJ Tweedie, and A Yamaguchi. 2008. "Neutral-kaon Mixing from (2+1)-Flavor Domain-Wall QCD." *Physical Review Letters* 100:032001 arXiv:hep-ph/0702042v4.

Appelquist T and F Sannino. 1999. "The Physical Spectrum of Conformal SU(N) Gauge Theories." *Physical Review D* 4(6): 067702.

Appelquist T and LCR Wijewardhana. 1987. "Chiral Hierarchies from Slowly Running Couplings in Technicolor Theories." *Physical Review D* 36(2):568-580.

Appelquist T, GT Fleming, and E Neil. 2008. "Lattice Study of the Conformal Window in QCD-Like Theories." *Physical Review Letters* 100 (17): 171607.

Artuso M, B Meadows and AA Petrov. 2008. "Charm Meson Decays." *Annual Review of Nuclear and Particle Science* 58:249-291. arXiv:0802.2934 [hep-ph].

ATLAS TDR—017, CERN-LHCC-2005-022. 2005. *The ATLAS Computing Technical Design Report*. Prepared by the ATLAS Computing Group. Accessed February 2, 2009 at http://cdsweb.cern.ch/record/837738/files/.

Aubin C, C Bernard, C DeTar, J Osborn, S Gottlieb, EB Gregory, D Toussaint, UM Heller, JE Hetrick, and R Sugar. 2008. "Light Pseudoscalar Decay Constants, Quark Masses, and Low Energy Constants from Three-Flavor Lattice QCD." *Physical Review D* 70(11):114501. arXiv:hep-lat/0407028v3.

Beane SR, K Orginos, and MJ Martin. 2007. "Savage Strong-Isospin Violation in the Neutron-Proton Mass Difference from Fully-Dynamical Lattice QCD and PQQCD." *Nuclear Physics B* 768:38-50. arXiv:hep-lat/0605014v1.

Becla J and D Wang. 2005. "Lessons Learned from Managing a Petabyte." In *Proceedings of the Second Biennial Conference on Innovative Data Systems Research, (CIDR 2005)*. University of Wisconsin-

Madison, Computer Sciences Dept., Madison, Wisconsin. http://www-db.cs.wisc.edu/cidr/cidr2005/papers/P06.pdf.

Bell G, J Gray and AS Szalay. 2006. "Petascale Computational Systems: Balanced Cyber-Infrastructure in a Data-Centric World." *IEEE Computer*, 39(1):110-112.

Bell GI, DL Bruhwiler, A Fedotav, A Sobol, RS Busby, P Stoltz, DT Abell, P Messmer, I Ben-Zvi, and V Letvinenko. 2008. "Simulating the Dynamical Friction Force on Ions Due to a Briefly Co-Propagating Electron Beam." *Journal of Computational Physics* 227(19): 8714-8735, DOI:10.1016/j.jcp.2008.06.01.

Ben-Zvi I, J Alduino, D Barton, D Beavis, M Blaskiewicz, J Brennan, A Burrill, R Calaga, P Cameron, X Chang, A Drees, AV Fedotov, W Fischer, G Ganetis, D Gassner, J Grimes, H Hahn, L Hammons, A Hershcovitch, HC Hseuh, D Kayran, J Kewisch, R Lambiase, D Lederle, V Litvinenko, C Longo, W MacKay, G Mahler, G Mclntyre, W Meng, B Oerter, C Pai, G Parzen, D Pate, D Phillips, S Plate, E Pozdeyev, T Rao, J Reich, T Roser, A Ruggiero, T Russo, C Schultheiss, Z Segalov, J Smedley, K Smith, T Tallerico, S Tepikian, R Than, R Todd, D Trbojevic, J Tuozzolo, P Wanderer, G Wang, D Weiss, Q Wu, K Yip, A Zaltsman, A Favale, D Holmes, J Sredniawski, H Bluem, M Cole, J Rathke, T Schultheiss, A Todd, V Parkhomchuk, V Reva, A Burov, S Nagaitsev, L Prost, A Sidorin, A Smirnov, Y Derbenev, P Kneisel, J Mammosser, L Phillips, J Preble, C Reece, R Rimmer, J Saunders, M Stirbet, H Wang, A Aleksandrov, D Douglas, Y Kang, D Abell, G Bell, D Bruhwiler, R Busby, J Cary, D Dimitrov, P Messmer, V Ranjbar, D Smithe, A Sobol, and P Stoltz. 2007. "Status of the R&D Towards Electron Cooling of RHIC." In *Proceedings of the Particle Accelerator Conference*, pp. 1938-1940. June 25-29, 2007, Albuquerque, New Mexico. IEEE Press, Piscataway, New Jersey.

Ben-Zvi I, V Litvinenko, D Barton, D Beavis, M Blaskiewicz, J Brennan, A Burrill, R Calaga, P Cameron, X Chang, R Connolly, Y Eidelman, A Fedotov, W Fischer, D Gassner, H Hahn, M Harrison, A Hershcovitch, HC Hseuh, A Jain, P Johnson, D Kayran, J Kewisch, R Lambiase, W Mackay, G Mahler, N Malitsky, G McIntyre, W Meng, K Mirabella, C Montag, T Nehring, A Nicoletti, B Oerter, G Parzen, D Pate, J Rank, T Roser, T Russo, J Scaduto, K Smith, T Rao, D Trbojevic, G Wang, J Wei, N Williams, KC Wu, V Yakimenko, A Zaltsman, Y Zhao, H Bluem, A Burger, M Cole, A Favale, D Holmes, J Rathke, T Schultheiss, A Todd, I Koop, V Parkhomchuk, Y Shatunov, A Skrinsky, J Sekutowicz, A Burov, S Nagaitsev, I Meshkov, A Sidorin, A Smirnov, G Troubnikov, D Abell, D Bruhwiler, J Delayen, Y Derbenev, W Funk, P Kneisel, L Merminga, L Phillips, and J Preble. 2005. "Electron Cooling of RHIC." In *Proceedings of the Particle Accelerator Conference*, pp. 2741-2743. May 16-20, 2005, Knoxville, Tennesee. IEEE Press, Piscataway, New Jersey.

Berger CF, Z Bern, LJ Dixon, F Febres Coredero, D Forde, T Gleisberg, H Ita, DA Kosower, and D Maitre. 2009. *Precise Predictions for W + 3 Jet Production at Hadron Colliders*. SLAC-PUB-13539. SLAC National Accelerator Laboratory, Menlo Park, California. Accessed April 30, 2009 at http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-pub-13539.pdf.

Bern Z, Dixon LJ, and DA Kosower. 2007. "On-Shell Methods in Perturbative QCD." *Annals of Physics* 322(7):1587-1634.

Blondin JM and A Mezzacappa. 2007. "Pulsar Spins from an Instability in the Accretion Shock of Supernovae." *Nature* 445(7123): 58-60.

Blue BE, CE Clayton, CL O'Connell, FJ Decker, MJ Hogan, C Huang, R Iverson, C Joshi, TC Katsouleas, W Lu, KA Marsh, WB Mori, P Muggli, R. Siemann, and D Walz. 2003. "Plasma-Wakefield Acceleration of an Intense Positron Beam." *Physical Review Letters* 90(21): 214801, DOI: 10.1103/PhysRevLett.90.214801.

Blum T, P Chen, N Christ, C Cristian, C Dawson, G Fleming, R Mawhinney, S Ohta, G iegert, A Soni, P Vranas, M Wingate, L Wu, and Y Zhestkov. 2003. "Kaon Matrix Elements and CP Violation from Quenched Lattice QCD: 3-Flavor Case." *Physical Review D* 68(11):114506. arXiv:hep-lat/0110075v2.

Blum T, T Doi, M Hayakawa, T Izubuchi, and N Yamada. 2007. "Determination of Light Quark Masses from the Electromagnetic Splitting of Pseudocalar Meson Masses Computed with Two Flavors of Domain Wall Fermions." *Physical Review D* 76(11):114508. arXiv:0708.0484 [hep-lat].

Blumenfeld I, CE Clayton, FJ Decker, MJ Hogan, CK Huang, R Ischenbeck, R Iverson, C Joshi, T Katsouleas, N Kirby, W Lu, KA Marsh, WB Mori, P Muggli, E Oz, RH Siemann, D Walz, and MM Zhou. 2007. "Energy Doubling of 42 GeV Electrons in a Metre Scale Plasma Wakefield Accelerator." *Nature* 445(7129): 741.744.

Bode A,R Frezzotti, B Gehrmann, M Hasenbusch, J Heitger, K Jansen, S Kurth, H Simma, S Sint, R Sommer, P Weisz, H Wittig, and U Wolff. 2002. "First results on the running coupling in QCD with two Massless flavors." *Nuclear Physics B – Proceedings Supplements* 106:792. arXiv:hep-lat/0105003v2.

Brodsky SJ, HC Pauli, and SS Pinsky. 1998. "Quantum Chromodynamics and Other Field Theories on the Light Cone." *Physics Reports* 301(4-6):299-486.

Brower R, N Christ, M Creutz, P Mackenzie, J Negele, C Rebbi, D Richards, S Sharpe, and R Sugar. "Lattice Gauge Theory for Physics Beyond the Standard Model," USQCD collaboration white paper, Accessed at: http://www.usqcd.org/documents/bsm.pdf.

Brower R, N Christ, M Creutz, P Mackenzie, J Negele, C Rebbi, D Richards, S Sharpe, and R Sugar. 2007. *Fundamental Parameters from Future Lattice Calculations*. USQCD collaboration white paper. Accessed February 5, 2009 at http://www.usqcd.org/documents/fundamental.pdf.

Bruhwiler DL. 2005. *ParSEC-Parallel Simulation of Electron Cooling.* Final Report: DE-FG03-01ER83313. Prepared by Tech-X Corporation for the U.S. Dept. of Energy, Office of Science, Washington, DC. Accessed January 8, 2009 at http://www.osti.gov/bridge/servlets/purl/850178-UQQd8X/850178.PDF.

Catterall S. 2008. "From Twisted Supersymmetry to Orbifold Lattices." *Journal of High Energy Physics* (1): 0801.

Catterall S. 2009. "First Results from Simulations of Supersymmetric Lattice." *Journal of High Energy Physics* 09(2008):040. arXiv:0811.1203 [hep-lat].

CERN-LHCC-2005-023. 2005. *CMS; the Computing Project Technical Design Report*. Accessed February 2, 2009 at http://cdsweb.cern.ch/record/838359/files/lhcc-2005-023.pdf.

CERN-OPEN-2008-020. 2008. *Expected Performance of the ATLAS Experiment: Detector, Trigger, and Physics.* The ATLAS Collaboration. Accessed February 2, 2009 at http://cdsweb.cern.ch/record/1125884.

Cohen A, DB Kaplan, E Katz, and M Unsal. 2003. "Supersymmetry on a Euclidean Space-Time Lattice. 2. Target Theories with Eight Supercharges." *Journal of High Energy Physics* (12): 031.

Davies CTH, E Follana, A Gray, GP Lepage, Q Mason, M Nobes, J Shigemitsu, HD Trottier, M Wingate, C Aubin, C Bernard, T Burch, C DeTar, S Gottlieb, EB Gregory, UM Heller, JE Hetrick, J Osborn, R Sugar, D Toussaint, M Di Pierro, A El-Khadra, AS Kronfeld, PB Mackenzie, D Menscher, and J Simon. 2004. "High precision Lattice QCD Confronts Experiment." *Physical Review Letters* 92(2): 022001.

Dawson C, G Martinelli, GC Rossi, CT Sachrajda, S Sharpe, M Talevi, and M Testa. 1998. "New Lattice Approaches to the Delta I=1/2 rule." *Nuclear Physics B* 514:313-335. arXiv:hep-lat/9707009v1.

Dean J and S Ghemawat. 2004. "MapReduce: Simplified Data Processing on Large Clusters." In *OSDI'04: 6th Symposium on Operating System Design and Implementation*, USENIX Association, Berkeley, California.

Della Morte M, R Frezzotti, J Heitger, J Rolf, R Sommer, and U Wolff. 2005. "Computation of the strong Coupling in QCD With Two Dynamical Flavors." *Nuclear Physics B* B713:378-406. arXiv:hep-lat/0411025v1.

De Roeck A, ed. 2006, *CMS Physics Technical Design Report Volume II: Physics Performance.* CERN/LHCC 2006-021. The CMS Collaboration. Accessed February 2, 2009 at http://cdsweb.cern.ch/record/942733/files/lhcc-2006-021.pdf.

de Téramond GF and SJ Brodsky. 2009. "Light-Front Holography: A First Approximation to QCD." *Physical Review Letters* 102(8): 081601.

Diemand J, M Kuhlen, and P Madau. 2007a. "Dark Matter Substructure and Gamma-Ray Annihilation in the Milky Way Halo." *Astrophysical Journal* 657(1): 262-270.

Diemand J, M Kuhlen, and P Madau. 2007b. "Formation and Evolution of Galaxy Dark Matter Halos and their Substructure." *Astrophysical Journal* 667(2): 859-877.

Diemand J, M Kuhlen, P Madau, M Zemp, B Moore, D Potter, and J Stadel. 2008. "Clumps and Streams in the Local Dark Matter Distribution." *Nature* 454(7205): 735-738.

DOE/NSF High Energy Physics Advisory Panel, Quantum Universe Committee. *Quantum Universe- The Revolution in 21st Century Particle Physics* Accessed January 8, 2009 at http://www.interactions.org/pdf/Quantum_Universe_GR_SP.pdf.

Duan H, GM Fuller, J Carlson, and YZ Qian. 2006a. "Coherent Development of Neutrino Flavor in the Supernova Environment: Correlated Neutrino Trajectories." *Physical Review D.* 74(10): 105014.

Duan H, GM Fuller, J Carlson, and YZ Qian. 2006b. "Coherent Development of Neutrino Flavor in the Supernova Environment." *Physical Review. Letters* 97(24): 241101.

Duan H, GM Fuller, J Carlson, and YZ Qian. 2007. "Neutrino Mass Hierarchy and Stepwise Spectral Swapping of Supernova Neutrino Flavors." *Physical Review Letters* 99(24): 241802.

Durr S, Z Fodor, J Frison, C Hoebling, R Hoffmann, SD Katz, S Krieg, T Kurth, L Lellouch, T Lippert, KK Szabo and G Vultert. 2008. "Ab Initio Determination of Light Hadron Masses." *Science* 322(5905):1224-1227.

El-Khadra AX, AS Kronfeld and PB Mackenzie. 1997. "Massive Fermions in Lattice Gauge Theory." *Physical Review D* 55:3933-3957. arXiv:hep-lat/9604004v1.

Esteban-Pretel A, S Pastor, R Tomas, GG Raffelt, and G Sigl. 2007. "Decoherence in Supernova Neutrino Transformations Suppressed by Deleptonization." *Physical Review D.* 76(12): 125018.

Faure J, Y Glinec, A Pukov, S Kiselev, S Gordienko, E Lefebvre, JP Rousseau, F Burgy and V Malka. 2004. "A Laser-Plasma Accelerator Producing Monoenergetic Electron Beams." *Nature* 431(7008): 541-544.

Fermilab, Muon Collier Coordinating Committee. 2008. *Muon Accelerator R&D: A proposal for the next 5 years*. Revision 0; November 18, 2008. Accessed January 8, 2009 at http://apc.fnal.gov/groups2/MCCC/Muon5yearplanFinalR0.pdf.

Fermilab. 2007. "Physics at the Intensity Frontier." Chapter 4 in *Fermilab Steering Group Report*. Accessed February 2, 2009 at http://www.fnal.gov/pub/directorate/steering/chapter_4b.shtml.

Fleming GT. 2008. "Strong Interactions for the LHC."In *Proceedings of Science*: arXiv:0812.2035v1. Accessed February 5, 2009 at http://arxiv.org/PS_cache/arxiv/pdf/0812/0812.2035v1.pdf.

Fogli G, E Lisi, A Marrone, and A Mirizzi. 2007. "Collective Neutrino Flavor Transitions in Supernovae and the Role of Trajectory Averaging." *Journal of Cosmology and Astroparticle Physics* (12): 010.

Gamiz E. 2008. "Heavy Flavour Phenomenology from Lattice QCD." In *Proceedings of Science:* arXiv:0811.4166. Accessed February 5, 2009 at http://arxiv.org/PS_cache/arxiv/pdf/0811/0811.4146v2.pdf.

Geddes CGR, C Toth, J van Tilborg, E Esarey, CB Schroeder, D Bruhwiler,C Nieter, J Cary, and WP Leemans. 2004. "High-Quality Electron Beams from a Laser Wakefield Accelerator Using Plasma-Channel Guiding." *Nature* 431(7008): 538-541.

Hawley JF and J Krolik. 2006. "Magnetically Driven Jets in the Kerr Metric." *Astrophysical Journal* 641(1): 103-116.

Heitger J and R Sommer. 2004. "Non-Perturbative Heavy Quark Effective Theory." *Journal of High Energy Physics* 02(2004):022. arXiv:hep-lat/0310035v2.

HEPAP – High Energy Physics Advisory Panel. 2004. *Quantum Universe: The Revolution in 21st Century Particle Physics*. Accessed February 5, 2009 at http://www.interactions.org/quantumuniverse/qu/.

Hogan MJ, CD Barnes, CE Clayton, FJ Decker, S Deng, P Emma, C Huang, RH Iverson, DK Johnson, C Joshi, T Katsouleas, P Krejcik, W Lu, KA Marsh, WB Mori, P Muggli, CL O'Connell, E Oz, RH Siemann, and D Walz. 2005. "Multi-GeV Energy Gain in a Plasma-Wakefield Accelerator." *Physical Review Letters* 95(5): 054802.

Hogan MJ, CE Clayton, C Huang, P Muggli, S. Wang, BE Blue, D Walz, KA Marsh, CL O'Connell, S Lee, R Iverson, FJ Decker, P Raimondi, WB Mori, TC Katsouleas, C Joshi, and RH Siemann. 2003. "Ultrarelativistic-Positron-Beam Transport through Meter-Scale Plasmas." *Physical Review Letters* 90(20): 205002.

Hogan MJ, R Assmann, FJ Decker, R Iverson, P Raimondi, S Rokni, RH Siemann, D Walz, D Whittum, B Blue, CE Clayton, E Dodd, R Hemker, C Joshi, KA Marsh, WB Mori, S Wang, T Katsouleas, S Lee, P Muggli, P Catravas, S Chattopadhyay, E Esarey, and WP Leemans. 2000. "E-157: A 1.4-m-Long Plasma Wake Field Acceleration Experiment Using a 30 GeV Electron Beam from the Stanford Linear Accelerator Center Linac." *Physics of Plasmas* 7(5): 2241-2248.

Hughes SA. 2003. "Listening to the Universe with Gravitational-Wave Astronomy." *Annals of Physics* 303(1): 142-178.

Hsu WW and AJ Smith. 2003. "Characteristics of IO traffic in personal computer and server workloads." *IBM Systems Journal*, 42(2):347-372.

INFN/AE - 07/2, SLAC-R-856, LAL 07-15. 2007. *SuperB: A High-Luminosity Asymmetric e+ e- Super Flavor Factory Conceptual Design Report*. Accessed February 2, 2009 at http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-r-856.pdf.

Institute for Astronomy, University of Hawaii. (2005) *Pan-STARRS: Panoramic Survey Telescope and Rapid Response System*, Accessed February 2, 2009 at http://pan-starrs.ifa.hawaii.edu/.

Jansen K. 2008. "Lattice QCD: A Critical Status Report." DESY 08-138. In *Proceedings of Science:* arXiv:0810.5634. Accessed February 5, 2009 at http://arxiv.org/PS_cache/arxiv/pdf/0810/0810.5634v2.pdf.

Johnson DK, D Auerbach, I Blumenfeld, CD Barnes, CE Clayton, FJ Decker, S Deng, P Emma, MJ Hogan, C Huang, R Ischebeck, R Iverson, C Joshi, TC Tatsouleas, N Kirby, P Krejcik, W Lu, KA Marsh, WB Mori, P Muggli, CL O'Connell, E Oz, RH Siemann, D Walz and M Zhou. 2006. "Positron Production by X Rays Emitted by Betatron Motion in a Plasma Wiggler." *Physical Review Letters* 97(17): 175003.

Kadanoff LP. 2001. "Turbulent Heat Flow: Structures and Scaling." Physics Today 54(8): 34-39.

Kaplan DB and M Unsal. 2005. "A Euclidean Lattice Construction of Supersymmetric Yang-Mills Theories with Sixteen Supercharges." *Journal of High Energy Physics* (9): 042.

Kaplan DB, E Katz, and M Unsal. 2003. "Supersymmetry on a Spatial Lattice." *Journal of High Energy Physics (5):* 037.

Kasen D, Thomas RC, Woosley SE and F.Ropke, “Multi-Dimensional Radiative Transfer Calculations of the Light Curves and Spectra of Type~Ia Supernovae”, Journal of Physics, Conference Series (2008), 125, 012001.

Khokhlov AM. 1991. “Delayed Detonation Model for Type-IA Supernovae.” *Astronomy and Astrophysics* 245(1): 114-128.

Kronfeld A. 2006. “Predictions with Lattice QCD.” *Journal of Physics: Conference Series* 46:147-151. Accessed February 5, 2009 at http://www.iop.org/EJ/article/1742-6596/46/1/020/jpconf6_46_020.pdf?request-id=e2ddb943-d77c-4f6e-9d5a-ac27070dd0bb.

Kuhlen M, J Diemand, and P Madau. 2007. “The Shapes, Orientation, and Alignment of Galactic Dark Matter Subhalos.” *Astrophysical Journal* 671(2): 1135-1146.

Kuhlen M, J Diemand, and P Madau. 2008. “The Dark Matter Annihilation Signal from Galactic Substructure: Predictions for GLAST.” *Astrophysical Journal* 686(1): 262-278.

Lee LQ, Z Bai, W Gao, L Ge, P Husbands, M Kowalski, X Li, Z Li, CK Ng, C Yang, E Ng, and K Ko. 2005. “Modeling RF Cavities with External Coupling.” In *SIAM Conference on Computational Science & Engineering,* Orlando, Florida, February 12-15, 2005.

Lu W, C Huang, M Zhou, W Mori, and T. Katsouleas. 2006. “Nonlinear Theory for Relativistic Plasma Wakefields in the Blowout Regime.” *Physical Review Letters* 96(16): 165002.

Maldacena J. 1997. “The Large N limit of Superconformal Field Theories and Supergravity.” *International Journal of Theoretical Physics* 38(4):1113-1133.

Mangles SPD, AGR Thomas, MC Kaluza, O Lundh, F Lindau, A Persson, FS Tsung, Z Najmudin, WB Mori, CG Wahlstrom, and K Krushelnick. 2006. “Laser-Wakefield Acceleration of Monoenergetic Electron Beams in the First Plasma-Wave Period.” *Physical Review Letters* 96(21): 215001.

Mangles SPD, CD Murphy, Z Najmudin, AGR Thomas, JL Collier, AE Dangor, EJ Divall, PS Foster, JG Gallacher, CJ Hooker, DA Jaroszynski, AJ Langley, WB Mori, PA Norreys, FS Tsung, R Viskup, BR Walton, and K Krushelnick. 2004. “Monoenergetic Beams of Relativistic Electrons from Intense Laser-Plasma Interactions.” *Nature* 431(7008): 535-538.

Motl PM, EJ Hallman, JO Burns, and ML Norman. 2005. “The Integrated Sunyaev-Zeldovich Effect as a Superior Method for Measuring the Mass of Clusters of Galaxies.” *Astrophysical Journal* 62(2): L63-L66.

Muggli P, BE Blue, CE Clayton, S Deng, FJ Decker, MJ Hogan, C Huang, R Iverson, C Joshi, TC Katsouleas, S Lee, W Lu, KA Marsh, WB Mori, CL O’Connell, P Raimondi, R Siemann, and D Walz. 2004. “Meter-Scale Plasma-Wakefield Accelerator Driven by a Matched Electron Beam.” *Physical Review Letters* 93(1): 014802.

National Research Council, Committee on the Physics of the Universe. 2003. *Connecting Quarks with the Cosmos: Eleven Science Questions for the New Century*. National Academies Press, Washington, DC.

Noaki JI. 2003. "Calculation of Nonleptonic Kaon Decay Amplitudes from K ->pi Matrix Elements in Quenched Domain-Wall QCD." *Physical Review D* 68:014501. arXiv:hep-lat/0108013v3.

Particle Physics Project Prioritization Panel. 2008. *US Particle Physics: Scientific Opportunities – A Strategic Plan for the Next Ten Years. Report of the Particle Physics Project Prioritization Panel.* Accessed February 2, 2009 at http://www.er.doe.gov/hep/files/pdfs/P5_Report%2006022008.pdf

Peskin ME and T Takeuchi. 1992. "Estimation of Oblique Electroweak Corrections." *Physical Review D* 46(1):381-409. Plewa T, AC Calder, and DQ Lamb. 2004. "Type IA Supernova Explosion: Gravitationally Confined Detonation." *Astrophysical Journal* 612(1): L37-L40.

The Pierre Auger Collaboration, J Abraham, P Abreub, M Aglietta, C Aguirre, D Allard, I Allekotte, J Allen, P Allison, J Alvarez-Muñiz, M Ambrosio, L Anchordoqui, S. Andringa, A Anzalone, C Aramo, S Argirò, K Arisaka, E Armengaud, F Arneodo, F Arqueros, T Asch, H Asorey, P Assis, BS Atulugama, J Aublin, M Ave, G Avila, T Bäcker, D Badagnani, AF Barbosa, D Barnhill, SLC Barroso, P Bauleo, JJ Beatty, T Beau, BR Becker, KH Becker, JA Bellido, S BenZvi, C Berat, T Bergmann, P Bernardini, X Bertou, PL Biermann, P Billoir, O Blanch-Bigas, F Blanco, P Blasi, C Bleve, H Blümer, M Boháčová, C Bonifazi, R Bonino, J Brack, P Brogueira WC Brown, P Buchholz, A Bueno, RE Burton, NG Busca, KS Caballero-Mora, B Cai, DV Camin, L Caramete, R Caruso, W Carvalho, A Castellina, O Catalano, G Cataldi, L Cazonck, R Cester, J Chauvin, A Chiavassa, JA Chinellato, A Chou, J Chye, RW Clay, E Colombo, R Conceição, B Connolly, F Contreras, J Coppens, A Cordier, U Cotti, S Coutu, CE Covault, A Creusot, A Criss, J Cronin, A Curutiu, S Dagoret-Campagne, K Daumiller, BR Dawson, RM de Almeida, C De Donato, SJ de Jong, G De La Vega, WJM de Mello Junior, JRT de Mello Neto, I De Mitri, V de Souza, L del Peral, O Deligny, A Della Selva, C Delle Fratte, H Dembinski, C Di Giulio, JC Diaz, PN Diep, C Dobrigkeit, JC D'Olivo, PN Dong, D Dornic, A Dorofeev, JC dos Anjos, MT Dova, D D'Urso, I Dutan, MA DuVernois, R Engel, L Epele, M Erdmann, CO Escobar, A Etchegoyen, P Facal San Luis, H Falcke, G Farrar, AC Fauth, N Fazzini, F Ferrer, S Ferry, B Fick, A Filevich, A Filipčič, I Fleck, CE Fracchiolla, W Fulgione, B García, D García Gámez, D Garcia-Pinto, X Garrido, H Geenen, G Gelmini, H Gemmeke, PL Ghia, M Giller, H Glass, MS Gold, G Golup, F Gomez Albarracin, M Gómez Berisso, R Gómez Herrero, P Gonçalves, M Gonçalves do Amaral, D Gonzalez, JG Gonzalez, M González, D Góra, A Gorgi, P Gouffon, V Grassi, AF Grillo, C Grunfeld, Y Guardincerri, F Guarino, GP Guedes, J Gutiérrez, JD Hague, JC Hamilton, P Hansen, D Harari, S Harmsma, JL Harton, A Haungs, T Hauschildt, MD Healy, T Hebbeker, G Hebrero, D Heck, C Hojvat, VC Holme, P Homola, J Hörandel, A Horneffer, M Horvat, M Hrabovský, T Huege, M Hussain, M Iarlori, A Insolia, F Ionita, A Italiano, M Kaducak, KH Kampert, T Karova, B Kégl, B Keilhauer, E Kemp, RM Kieckhafer, HO Klages, M Kleifges, J Kleinfeller, R Knapik, J Knapp, DH Koang, A Kriege, O Krömer, D Kuempel, N Kunka, A Kusenko, G La Rosa, C Lachaud, BL Lago, D Lebrun, P LeBrun, J Lee, MA Leigui de Oliveira, A Letessier-Selvon, M Leuthold, I Lhenry-Yvon, R López, A Lopez Agüera, J Lozano Bahilo, R Luna García, MC Maccarone, C Macolino, S Maldera, G Mancarella, ME Manceñid, D Mandat, P Mantsch, AG Mariazzi, IC Maris, HR Marquez Falcon, D Martello, J Martínez, O Martínez Bravo, HJ Mathes, J Matthews, JAJ. Matthews, G Matthiae, D Maurizio, PO Mazur, T McCauley, M McEwen, RR McNeil, MC Medina, G Medina-Tanco, A Meli, D Melo, E Menichetti, A Menschikov, Chr. Meurer, R Meyhandan, MI Micheletti, G Miele, W Miller, S Mollerac, M Monasor, D Monnier Ragaigne, F Montanet, B Morales, C Morello, JC Moreno, C Morris, M. Mostafá, M.A. Muller, R. Mussa, G. Navarra, J.L. Navarro, S. Navas, P. Necesal, L. Nellen, C. Newman-Holmes, D. Newton, PT Nhung, N Nierstenhoefer, D Nitz, D Nosek, L Nožka, J Oehlschläger, T Ohnuki, A Olinto, VM Olmos-Gilbaja, . Ortiz, F Ortolani, S Ostapchenko, L Otero, N Pacheco, D Pakk Selmi-Dei, M Palatka, J Pallotta, G

Parente, E Parizot, S Parlatiaz, S Pastor, M Patel, T Paul, V Pavlidou, K Payet, M Pech, J Pe¸kala, R Pelayo, IM Pepe, L Perrone, S Petrera, P Petrinca, Y Petrov, A Pichel, R Piegaia, T Pieroga, M Pimenta, T Pinto, V Pirronello, O Pisantia, M Platino, J Pochon, P Privitera, M Prouza, EJ Quel, J Rautenberg, A Redondo, S Reucroft, B Revenu, FAS Rezende, J Ridky, S Riggi, M Risse, C Rivière, V Rizi, M Roberts, C Robledo, G Rodriguez, J Rodriguez Martino, J Rodriguez Rojo, I Rodriguez-Cabo, MD Rodríguez-Frías, G Ros, J Rosado, M Roth, B Rouillé-d'Orfeuil, E Roulet, AC Rovero, F Salamida, H Salazar, G Salina, F Sánchez, M Santander, CE Santo, EM Santo, F Sarazin, S Sarkar, R Sato, V Scherini, H Schieler, A Schmidt, F Schmidt, T Schmidt, O Scholten, P Schovánek, F Schüssler, SJ Sciutto, M Scuderi, A Segreto, D Semikoz, M Settimo, RC Shellard, I Sidelnik, BB Siffert, G Sigl, N Smetniansky De Grande, A Smiałkowski, R Šmída, AGK Smith, BE Smith, GR Snow, P Sokolsky, P Sommers, J Sorokin, H Spinka, R Squartini, E Strazzeri, A Stutz, F Suarez, T Suomijärvi, AD Supanitsky, MS Sutherland, J Swain, Z Szadkowski, J Takahashi, A Tamashiro, A Tamburro, O Taşcău, R Tcaciuc, NT Thao, D Thomas, R Ticona, J Tiffenberg, C Timmermans, W Tkaczyk, CJ Todero Peixoto, B Tomé, A Tonachini, I Torres, P Travnicek, A Tripathi, G Tristram, D Tscherniakhovski, M Tueros, R Ulrich, M Unger, M Urban, JF Valdés Galicia, I Valiño, L Valore, AM van den Berg, V van Elewyck, RA Vázquez, D Veberič, A Veiga, A Velarde, T Venters, V Verzi, M Videla, L Villaseñor, S Vorobiovb, L Voyvodic, H Wahlberg, O Wainberg, D Warner, AA Watson, S Westerhoff, G Wieczorek, L Wienckeb, B Wilczyńska, H Wilczyński, C Wileman, MG Winnic, H Wu, B Wundheiler, T Yamamoto, P Younk, E Zas, D Zavrtanik, M Zavrtanik, A Zech, A Zepeda, and M Ziolkowski. 2008. "Correlation of the Highest-Energy Cosmic Rays with the Positions of Nearby Active Galactic Nuclei." *Astroparticle Physics* 29(3): 188-204.

Prebys E. 2008. *mu2e Collaboration*. Accessed February 2, 2009 at http://mu2e.fnal.gov/ (last updated September 6, 2008).

Raffelt GG and AY Smirnov. 2007. "Self-Induced Spectral Splits in Supernova Neutrino Fluxes." *Physical Review D* 76(8): 081301.

Ren C, M Tzoufras, FS Tsung, WB Mori, S Amorini, RA Fonseca, LO Silva, JC Adam, and A Heron. 2004. "Global Simulation for Laser-Driven MeV Electrons in Fast Ignition." *Physical Review Letters* 93(18): 185002.

Riedel E, G Gibson, and C Faloutsos. 1998. "Active Storage for Large-Scale Data Mining and Multimedia." In *Proceedings of the 24th International Conference on Very Large Databases (VLDB '98),* pp. 62-73. August 24-27, 1998, New York, NY. Morgan Kaufman, San Francisco, California.

Sciascia B and the FlaviaNet Kaon Working Group. 2008. "Precision tests of the Standard Model with leptonic and semileptonic kaon decays." *Nuclear Physics B – Proceedings Supplements* 181-182:83-88. arXiv:0801.1817 [hep-ph]

SLAC National Accelerator Laboratory. 2007. *Extremely Large Databases Workshop (XLDB07),* Accessed December 9, 2008 at http://www-conf.slac.stanford.edu/xldb07/ (last updated October 2007)

SLAC National Accelerator Laboratory. 2008. *2nd Extremely Large Databases Workshop (XLDB08)* Accessed December 9, 2008 at http://www-conf.slac.stanford.edu/xldb08/ (last updated October 2008)

Sloan Digital Sky Survey III. 2008. “BOSS: Dark Energy and the Geometry of Space.” Accessed January 6, 2009 at http://www.sdss3.org/cosmology.php.

Spentzouris P and J Amundson. 2005. Emittance Dilution and Halo Creation During the First Milliseconds After Injection at the Fermilab Booster.” In *AIP Conference Proceedings: 33rd ICFA Advanced Beam Dynamics Workshop on High Intensity and High Brightness Hadron Beams.* 773: 127–131, DOI: 10.1063/1.1949511.

Spentzouris P. 2006. “Accelerator Modeling Under SciDAC: Meeting the Challenges of Next-Generation Accelerator Design, Analysis, and Optimization.” In *9th International Computational Accelerator Physics Conference*, pp. 315-320. October 2-6, 2006, Chamonix Mont-Blanc, France. Accessed January 8, 2009 at http://bel.gsi.de/icap2006/PAPERS/THMPMP03.PDF.

Spentzouris, P. 2008. “Community Petascale Project for Accelerator Science and Simulation: Advancing Computational Science for Future Accelerators and Accelerator Technologies.” *SciDAC08*. Accessed January 8, 2009 at http://hpcrd.lbl.gov/SciDAC08/files/presentations/ComPASS_SciDAC08_spentzouris.ppt (last updated 8/27/2008).

Springel V, SDM White, CS Frenk, JF Navarro, A Jenkins, M Vogelsberger, J Wang, A Ludlow, and A Helmi. 2008. “Prospects for Detecting Supersymmetric Dark Matter in the Galactic Halo.” *Nature* 456(7218): 73-76.

Stonebraker, M, J Becla, D Dewitt, K-T Lim, D Maier, O Ratzesberger and SB Zdonik. 2009. “Requirements for Science Data Bases and SciDB.” In *Proceedings of the Fourth Biennial Conference on Innovative Data Systems Research,(CIDR 2009)*, University of Wisconsin-Madison, Computer Sciences Dept., Madison, Wisconsin. http://www-db.cs.wisc.edu/cidr/cidr2009/Paper_26.pdf.

Szalay AS, G Bell, J Vandenberg, A Wonders, R Burns, D Fay, J. Heasley, T Hey, M Nieto-SantiSteban, A Thakar, C van Ingen and R Wilton. 2009. “GrayWulf: Scalable Clustered Architecture for Data Intensive Computing.” In *Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS-42),* IEEE, Piscataway, New Jersey.

Szalay AS and J Gray. 2006. “Science in an Exponential World.” *Nature* 440(7083): 413-414.

DOE—U.S. Deptartment of Energy, Office of Science. 2003a. *Facilities for the Future of Science: A Twenty-Year Outlook*. DOE/SC-0078. Accessed January 8, 2009 at http://www.science.doe.gov/Scientific_User_Facilities/History/20-Year-Outlook-screen.pdf.

DOE—U.S. Deptartment of Energy, Office of Science. 2007. *Four Years Later: An Interim Report on Facilities for the Future of Science: A Twenty-Year Outlook*. Accessed January 8, 2009 at http://www.science.doe.gov/Scientific_User_Facilities/History/20-Year-Outlook-Interim%20Report(August).pdf.

U.S. Dept. Of Energy. 2003b. *Energy Secretary Spencer Abraham Announces Department of Energy 20-year Science Facility Plan*. Accessed January 8, 2009 at http://www.science.doe.gov/News_Information/News_Room/2003/Energy%2020-Year%20Science%20Facility%20Plan.htm.

US Dept. of Energy. Office of High Energy Physics. *Exploring the Quantum Universe: Vision for HEP*. Accessed January 8, 2009 at http://www.er.doe.gov/hep/vision/index.shtml.

Vary JP, H Honkanen, J Li, P Maris, and SJ Brodsky. 2008. "Recent Progress in Hamiltonian Light-Front QCD." *Proceedings of Science* PoS(LC2008)040. Accessed April 30, 2009 at http://pos.sissa.it//archive/conferences/061/040/LC2008_040.pdf.

Weiglein G, T Barklow, E Boos, A De Roeck, K Desch, F Gianotti, R Godbole, JF Gunion, HE Haber, S Heinemeyer, JL Hewettb, K Kawagoe, K Mönig, MM Nojiri, G Polesello, F Richard, S Riemann, WJ Stirling, AG Akeroyd, BC Allanach, D Asner, S Asztalos, H Baer, M Battaglia, U Baur, P Bechtle, G Bélanger, A Belyaev, EL Berger, T Binoth, GA Blair, S Boogert, F Boudjema, D Bourilkov, . Buchmüller, V Bunichev, G Cerminara, M Chiorboli, H Davoudiasl, S Dawson, S De Curtis, F Deppisch, M. Díaz, M Dittmar, A Djouadi, D Dominici, U Ellwanger, JL Feng, IF Ginzburg, A Giolo-Nicollerat, BK Gjelsten, S Godfrey, D Grellscheid, J Gronberg, E Gross, J Guasch, K Hamaguchi, T Han, J Hisano, W Hollik, C Hugonie, T Hurth, J Jiang, A Juste, J Kalinowski, W Kilian, R Kinnunen, S Kraml, M Krawczyk, A Krokhotine, T Krupovnickas, R Lafaye, S Lehti, HE Logan, E Lytken, V Martin, HU Martyn, DJ Miller, S Moretti, F Moortgat, G Moortgat-Pick, M Mühlleitner, P Nieżurawskia, A Nikitenkoa, LH Orr, P Osland, AF Osorio, H Päs, T Plehn, W Poroda, A Pukhov, F Quevedo, D Rainwater, M Ratz, A Redelbach, L Reina, T Rizzo, R Rückl, HJ Schreiber, M Schumacher, A Sherstnev, S Slabospitsky, J Solà, A Sopczaka, M Spira, M Spiropulu, Z Sullivan, M Szleper, TMP. Tait, X Tata, DR Tovey, A Tricomi, M Velasco, D Wackeroth, CEM. Wagner, S Weinzierl, P Wienemann, T Yanagida, AF Żarnecki, D Zerwas, PM Zerwas, L Živković and The LHC/ILC Study Group LHC/LC Study Group. 2006. "Physics Interplay of the LHC and the ILC." *Physics Reports* 426(2-6):47-358.

Willingale L, SPD Mangles, PM Nilson, RJ Clarke, AE Dangor, MC Kaluza, S Karsch, KL Lncaster, WB Mori, Z Najmundin, J Schreiber, AGR Thomas, MS Wei, and K Krushelnick. 2006. "Collimated Multi-Mev Ion Beams from High-Intensity Laser Interactions with Underdense Plasma." *Physical Review Letters* 94(24): 245001.

Wilson KG. 1974. "Confinement of Quarks." *Physical Review D* 10(8):2445-2459.

Wingate M. 2006. "B Physics on the Lattice: Present and Future." *Modern Physics Letters A* 21(15):1167-1182.

Wise, JH.and T. Abel. 2007. "Resolving the Formation of Protogalaxies. I. Virialization", Astrophys. J., 665, 899.

APPENDICES

WORKSHOP AGENDA

WORKSHOP PARTICIPANTS

ACRONYMS AND ABBREVIATIONS

APPENDIX 1: WORKSHOP AGENDA

Tuesday, December 9, 2008

7:30 a.m. – 8 a.m. Registration/Working Breakfast/Panel Chair Meetings

8:30 a.m. – 8:45 a.m. Roger Blandford: Introduction/Charge/Expectations/Logistics

8:45 a.m. – 9 a.m. Walt Polansky and John Kogut: Welcome Address

9:00 a.m. – 9:40 a.m. Mike Peskin: Plenary Speech 1

9:40 a.m. – 10:20 a.m. Mike Norman: Plenary Speech 2

10:20 a.m. – 10:40 a.m. General Discussion

10:40 a.m. – 11:20 a.m. Horst Simon: Plenary Speech 3

11:20 a.m. – noon Bob Rosner: Plenary Speech 4

Noon – 1 p.m. Working Lunch

Breakout Sessions

Cosmology and Astrophysics Simulation (Chair: Mike Norman)

High Energy Theoretical Physics (Chair: Steve Sharpe)

Accelerator Simulation (Chair: Panagiotis Spentzouris)

Astrophysics Data Handling (Chair: Alex Szalay)

Experimental Particle Physics (Chairs: Jim Shank and Frank Wuerthwein)

1:00 p.m. – 1:30 p.m. Opening Talk

1:30 p.m. – 3:30 p.m. Discussion on the White Paper

3:30 p.m. – 3:45 p.m. General Discussion

3:45 p.m. – 5:30 p.m. Continue Panel Work

6:00 p.m. – 8:00 p.m. Reception

Wednesday, December 10, 2008

7:30 – 8:30 p.m. Working Breakfast: Roger Blandford, Norman Christ, Young-Kee Kim: Summary of Day 1 and Expectations of Day 2.

8:30 a.m. – 10:45 a.m. Panel Work

10:45 a.m. – 11 a.m. General Discussion

11:00 a.m. – noon Continued Panel Work

Noon – 1 p.m. Working Lunch

1 p.m. – 2:30 p.m. Continued Panel Work

2:30 p.m. – 3:30 p.m. Presentation of Preliminary Findings by Each Panel

3:30 – 3:45 p.m. General Discussion

3:45 p.m. – 5:00 p.m. Presentation of Preliminary Findings by Each Panel

5:30 p.m. – 7:30 p.m. Working Dinner/Wrap-up of Day 2

Thursday, December 11, 2008

7:30 a.m. – 8:30 a.m. Roger Blandford/Norman Christ/Young-Kee Kim: Working Breakfast

8:30 a.m. – 10:45 a.m. Panel members: Discussion and Writing of Letter Report, and Consolidation of Materials Including Art Work for Draft paper

10:45 a.m. – 11:00 a.m. General Discussion

11:00 a.m. – noon Roger Blandford/Norman Christ/Young-Kee Kim: Discussion of Unresolved Issues; Validation of Draft Document

Noon-1:00 p.m. – Report Writing Team: Working Lunch

1:00 p.m. – 4:00 p.m. Panel Leads/Scribes: Finish Documentation

4:00 p.m. – 4:15 p.m. Conclusion and Next Steps

APPENDIX 2: WORKSHOP PARTICIPANTS

Name	**Affiliation**	**Email**
Workshop Chair		
Blandford, Roger	Stanford Linear Accelerator Center	rdb3@stanford.edu
Workshop Co-Chairs		
Christ, Norman	Columbia University	nhc@phys.columbia.edu
Kim, Young-kee	Fermi National Accelerator Laboratory	ykkim@fnal.gov
Panel Leads		
Norman, Mike	University of San Diego	michael.l.norman@gmail.com
Shank, Jim	Boston University	shank@bu.edu
Sharpe, Steve	University of Washington	srsharpe@u.washington.edu
Spentzouris, Panagiotis	Fermi National Accelerator Laboratory	spentz@fnal.gov
Szalay, Alex	Johns Hopkins University	szalay@pha.jhu.edu
Wuerthwein, Frank	University of San Diego	fkw@ucsd.edu
Foreign Attendees		
Adelmann, Andreas	Paul Scherrer Institut	andreas.adelmann@psi.ch
Aoki, Sinya	University of Tsukuba	saoki@het.ph.tsukuba.ac.jp
Bazevanidis, Kyriakos	European Commission	Kyriakos Baxevanidis@ec.europa.eu
Bird, Ian	CERN	ian.bird@cern.ch
Innocente, Vincenzo	CERN	vincenzo.innocente@cern.ch
Jones, Roger	Lancaster University	roger.jones@lancaster.ac.uk
Lake, George	University of Zurich	george@georgelake.org
Lellouch, Laurant	CERN	laurent.lellouch@cern.ch
Sommer, Rainer	DESY Zeuthen	sommer@ifh.de
Ukawa, Akira	University of Tsukuba	ukawa@ccs.tsukuba.ac.jp
Participants		
Abel, Tom	Stanford Linear Accelerator Center	tabel@stanford.edu
Amundson, James	Fermi National Accelerator Laboratory	Amundson@fnal.gov
Angerwala, Tilak	IBM	tilak@us.ibm.com
Banerjee, Bakul	Fermi National Accelerator Laboratory	bakulb@fnal.gov
Becla, Jacek	Stanford Linear Accelerator Center	becla@SLAC.Stanford.EDU
Bell, John	Lawrence Berkeley National Laboratory	jbbell@lbl.gov
Bethel, Wes	Lawrence Berkeley National Laboratory	ewbethel@lbl.gov
Bohnlein, Amber	DOE	amber.boehnlein@science.doe.gov
Borrill, Julian	Lawrence Berkeley National Laboratory	jdborrill@lbl.gov

Name	Affiliation	Email
Brower, Richard	Brown University	brower@bu.edu
Brown, David	Lawrence Livermore National Laboratory	dlb@llnl.gov
Cary, John	Tech-X Corporation	cary@txcorp.com
Catterall, Simon	Syracuse University	smc@physics.syr.edu
Centrella, Joan	NASA	joan.centrella@nasa.gov
Chan, Tony	National Science Foundation	tfchan@nsf.gov
Chatterjee, Lali	DOE	lali.chatterjee@science.doe.gov
Colella, Philip	Lawrence Berkeley National Laboratory	pcolella@lbl.gov
Connolly, Andrew	University of Washington	ajc@astro.washington.edu
Crawford, Matt	Fermi National Accelerator Laboratory	crawdad@fnal.gov
Diachin, Lori	Lawrence Livermore National Laboratory	diachin2@llnl.gov
Dodelson, Scott	Fermi National Accelerator Laboratory	dodelson@fnal.gov
Escala, Andres	Stanford Linear Accelerator Center	andres@SLAC.Stanford.EDU
Evrard, Gus	University of Michigan	evrard@umich.edu
Fisk, Ian	Fermi National Accelerator Laboratory	ifisk@fnal.gov
Fleming, George	Yale University	george.fleming@yale.edu
Gibson, Garth	Carnegie Mellon University	Garth.Gibson@cs.cmu.edu
Gottlieb, Steve	Indiana University	sg@denali.physics.indiana.edu
Grama, Ananth	Purdue University	ayg@cs.purdue.edu
Hawley, John	University of Virginia	jh8h@virginia.edu
Huang, Thuc	NNSA/DOE	Thuc.Hoang@nnsa.doe.gov
Huber, Melissa	Pacific Northwest National Laboratory	Melissa.huber@pnl.gov
Johnson, Gary	CSS	garymichaeljohnson@gmail.com
Johnson, Tony	Stanford Linear Accelerator Center	tony_johnson@slac.stanford.edu
Johnston, William	Lawrence Berkeley National Laboratory	WEJohnston@lbl.gov
Kaiser, Nick	University of Hawaii	kaiser@ifa.hawaii.edu
Khokhlov, Alexei	University of Chicago	ajk@oddjob.uchicago.edu
Ko, Kwok	Stanford Linear Accelerator Center	kwok@slac.stanford.edu
Kogut, John	DOE	john.kogut@science.doe.gov
Kronfeld, Andreas	Fermi National Accelerator Laboratory	ask@fnal.gov
Kurtz, Michael	Harvard University	kurtz@cfa.harvard.edu
Kuti, Julius	University of California, San Diego	jkuti@ucsd.edu
Lawrence, Andy	Stanford Linear Accelerator Center	axl@slac.stanford.edu
Lee, Rich (Lie-Quan)	Stanford Linear Accelerator Center	liequan@slac.stanford.edu
Lee, Steven	DOE	steven.lee@ee.doe.gov

Name	Affiliation	Email
Luth, Vera	Stanford Linear Accelerator Center	luth@slac.stanford.edu
MacCabe, Barney	University of New Mexico	maccabe@unm.edu
Mackenzie, Paul	Fermi National Accelerator Laboratory	mackenzie@fnal.gov
Madau, Piero	University of California, Santa Cruz	pmadau@ucolick.org
Magnier, Eugene	University of Hawaii	eugene@ifa.hawaii.edu
Messina, Paul	Argonne National Laboratory	messina@mcs.anl.gov
Mezzacappa, Tony	Oak Ridge National Laboratory	mezzacappaa@ornl.gov
Murgia, Simona	Stanford Linear Accelerator Center	murgia@SLAC.Stanford.EDU
Ng, Esmond	Lawrence Berkeley National Laboratory	EGNg@lbl.gov
Norris, Boyana	Argonne National Laboratory	norris@mcs.anl.gov
Peskin, Michael	Stanford Linear Accelerator Center	mpeskin@slac.stanford.edu
Polansky, Walt	DOE	walt.polansky@science.doe.gov
Pordes, Ruth	Fermi National Accelerator Laboratory	ruth@fnal.gov
Riley, Katherine	Argonne National Laboratory	riley@mcs.anl.gov
Rosner, Rob	University of Chicago	r-rosner@uchicago.edu
Ross, Rob	Argonne National Laboratory	rross@mcs.anl.gov
Rotem, Doron	Lawrence Berkeley National Laboratory	D_Rotem@lbl.gov
Sameh, Ahmed	Purdue University	sameh@cs.purdue.edu
Schreiber, Rob	Hewlett Packard Laboratories	rob.schreiber@hp.com
Scott, Steve	Cray Inc.	sscott@cray.com
Sekine, Yukiko	DOE	yukiko.sekine@science.doe.gov
Shigemitsu, Junko	Ohio State University	shige@mps.ohio-state.edu
Simon, Horst	Lawrence Berkeley National Laboratory	hdsimon@lbl.gov
Sterling, Thomas	Louisiana State University	tron@cct.lsu.edu
Stevens, Rick	Argonne National Laboratory	stevens@anl.gov
Straatsma, TP	Pacific Northwest National Laboratory	tps@pnl.gov
Thakur, Rajeev	Argonne national Laboratory	thakur@mcs.anl.gov
Tschirhart, Bob	Fermi National Accelerator Laboratory	tsch@fnal.gov
Turner, Michael	University of Chicago	mturner@kicp.uchicago.edu
Van Meter, James	NASA	james.r.vanmeter@nasa.gov
Wechsler, Risa	Stanford Linear Accelerator Center	rwechsler@stanford.edu
Wenaus, Torre	Brookhaven National Laboratory	wenaus@bnl.gov
White, Martin	University of California, Berkeley	mwhite@berkeley.edu
William, Roy	Caltech	roy@cacr.caltech.edu
Woosley, Stan	University of California, Santa Cruz	woosley@ucolick.org

APPENDIX 3: ACRONYMS AND ABBREVIATIONS

AdS/CFT	anti-de-Sitter space/conformal field correspondence
AMR	adaptive mesh refinement
ASCR	DOE Office of Advanced Scientific Computing Research
ATLAS	A Toroidal LHC ApparatuS
BAO	Baryon Acoustic Oscillations
BOSS	Baryon Oscillation Spectroscopic Survey
BSM	Beyond the Standard Model
CERN	European Organization for Nuclear Research
CKM	Cabbibo-Kobayashi-Maskawa
CLIC	Compact Linear Collider
CMB	cosmic microwave background
CMS	Compact Muon Solenoid
ComPASS	Community Petascale Project for Accelerator Science and Simulation
CP	charge conjugation (C) and parity (P)
CPU	central processing unit
DDT	deflagration-detonation transition
DES	Dark Energy Survey
DISC	data-intensive scalable computing
DOE	U.S. Department of Energy
DOF	degrees of freedom
FACET	Facilities for Accelerator Science and Experimental Test
FNAL	Fermi National Accelerator Laboratory
GCD	greatest common divisor
GPU	graphics processing unit
HEP	high energy physics
HQET	heavy-quark effective theory
HOM	higher order mode
HPC	high-performance computing
ILC	International Linear Collider
I/O	input/output
IXO	International X-ray Observatory
JDEM	Joint Dark Energy Mission
LANL	Los Alamos National Laboratory
LBNL	Lawrence Berkeley National Laboratory
LCLS	Linac Coherent Light Source
LFT	lattice field theory
LFV	lepton flavor violating
LHC	Large Hadron Collider
LISA	Laser Interferometer Space Antenna

LOASIS	Lasers, Optical Accelerator Systems Integrated Studies
LQCD	lattice quantum chromodynamics
LSST	Large Synoptic Survey Telescope
LTE	local thermodynamic equilibrium
LWFA	laser wakefield acceleration
MCMC	Markov Chain Monte Carlo analysis
MHD	magnetohydrodynamics
MSSM	Minimal Supersymmetric Standard Model
NLC	Next Linear Collider
ORNL	Oakridge National Laboratory
PanSTARRS	Panoramic Survey Telescope and Rapid Response System
Pc	parsec
PDE	partial differential equation
PETS	Power Extraction and Transfer Structure
PIC	particle in cell
PolarBear	Polarization of Background Radiation
PSF	point spread function
PTF	Palomar Transient Factory
PWFA	plasma wakefield acceleration
QCD	quantum chromodynamics
Rf	radio frequency
RHIC	Relativistic Heavy Ion Collider
SciDAC	Scientific Discovery through Advanced Computing
SDSS	Sloan Digital Sky Survey
SKA	Square Kilometre Array
SLAC	Stanford Linear Accelerator Center
SPH	smooth particle hydrodynamics
SQCD	supersymmetric quantum chromodynamics
SUSY	supersymmetry
VORPAL	Versatile Object-oriented Relativistic Plasma Analysis with Lasers

U.S. DEPARTMENT OF
ENERGY

Office of Science

Office of Advanced Scientific Computing Research
Office of High Energy Physics

Prepared for the U.S. Department of Energy under Contract DE-AC05-76RLO1830

Production support provided by Pacific Northwest National Laboratory,
Fundamental & Computational Sciences Directorate

PNNL-18775