COMPUTING IN HIGH ENERGY PHYSICS

Report from the Topical Panel Meeting on Computing and Simulations in High Energy Physics

---

Sponsored by the U.S. Department of Energy,
Office of Science, High Energy Physics

December 9-11, 2013 Rockville Hilton Hotel, Rockville, MD

Computing in High Energy Physics

March 24, 2014
Report from the Topical Panel Meeting on Computing and Simulations in High Energy Physics

December 9-11, 2013, Rockville Hilton Hotel, Rockville, MD

Steering Committee

Paul Avery (co-Chair)	University of Florida
Salman Habib (co-Chair)	Argonne National Laboratory
Amber Boehnlein	Stanford Linear Accelerator Center
Robert Roser	Fermilab
Stephen Sharpe	University of Washington
Heidi Schellman	Northwestern University
Craig Tull	Lawrence Berkeley National Laboratory
Torre Wenaus	Brookhaven National Laboratory

Panel Membership

Cosmic Frontier
Anders Borgland (SLAC)
Andy Connolly (U of Washington)
Gus Evrard (U Michigan)
Cristiano Galbiati (Princeton)
Peter Nugent (LBNL)
Martin White (Berkeley)

Energy Frontier
Lothar Bauerdick (FNAL)
Chip Brock (Michigan State)
Kaushik De (UT Arlington)
Steven Gottlieb (Indiana)
Stefan Hoeche (SLAC)
Tom LeCompte (Argonne)
Dan Marlow (Princeton)
Harvey Newman (Caltech)

Technology
Paul Messina (ANL)
Inder Monga (ESnet)
Richard Mount (SLAC)
Don Petravick (NCSA)
David Skinner (NERSC)
Artur Barczyk (Caltech)

Intensity Frontier
Bob Bernstein (FNAL)
Milind Diwan (Brookhaven)
Paul Mackenzie (FNAL)
Mayly Sanchez (Iowa State)
Malachi Schram (PNNL)

Accelerators
Cho Ng (SLAC)
Robert Ryne (LBNL)
Tor Raubenheimer (SLAC/Stanford)
Panagiotis Spentzouris (FNAL)
Jean-Luc Vay (LBNL)

Table of Contents

Executive Summary

The Topical Panel on computing and simulation in high energy physics was tasked with (i) assessing the current status of essential enabling technologies in this area and (ii) suggesting paths forward to improve effectiveness as measured in terms of overall impact on the discipline. The role of computing in particle physics cuts across all HEP Frontiers and research efforts, with significant computational elements in almost all science areas. In view of rapidly evolving computer architectures and technology and the varied yet synergistic needs of computation and simulation across the field, this community-based study was deemed both essential and timely.

Participants represented a broad spectrum of scientific and computational expertise across the DOE-HEP community as well as external to it. A substantial amount of work preceded the meeting; significant input being derived from a number of previous studies, including the recent Snowmass reports. The Panel also gathered additional input from scientists in a number of specific areas. This material is accessible online at www.hepcomputing.org.

Panel activities centered on addressing a number of urgent issues that cut across project and frontier boundaries, with critical ramifications for the particle physics community. These include dealing with rapid evolution of technology, management and development of the software environment as the science program evolves, and resource management and overall organization of computing.

The panel meeting was held in Washington DC in December, 2013, opening with an introductory overview from DOE-HEP Associate Director Dr. James Siegrist. The primary mode of activity was a set of plenary discussions, each initiated by short summary presentations. Speakers from the Office of Science (DOE-SC) and Advanced Scientific Computing Research (DOE-ASCR) provided perspectives on data preservation and high performance computing and data-intensive science in a broader context.

The organization of the meeting consisted of five subpanels (Accelerators, Cosmic Frontier, Energy Frontier, Intensity Frontier, Technology), with several cross-cutting sessions. The panelists represent a cross-section of expertise spanning the entire domain of particle physics research. To address a number of potentially critical issues, experts from other disciplines were recruited to provide inputs in specific technology areas.

The meeting consisted of plenary sessions only, the importance and wide-ranging nature of the topic requiring that all areas be actively represented. Two special sessions were organized; with Laura Biven (DOE-SC) on data preservation and sharing policies, and with Barbara Helland (DOE-ASCR) on ongoing ASCR-HEP collaborations and future possibilities, particularly in the area of data-intensive science.

Following the specified charge (Appendix 1), the Panel arrived at a number of specific findings, grouped into five subsets: (**1**) challenges arising from the rapid evolution of hardware technology (computer architecture, storage and networking, high performance computing), (**2**) issues related to development and management of the software environment (code maintenance/modernization, managing code diversity), (**3**) resource management issues (demands of smaller-scale projects, role of simulations, data preservation), (**4**) the organization of computing and simulation (management of computational R&D programs, interactions external to DOE-HEP, training), and, finally, as asked for by the charge, (**5**) the opportunities presented by establishing a Center for Computational Excellence for particle physics research.

After careful consideration of the findings, and based on its discussions, the Panel identified the following areas of opportunity (no particular ordering) for increased impact on progress in HEP science:

1. **Code modernization, maintenance, and dissemination:** Well-defined mechanisms for the continued maintenance and development of a number of particle physics software frameworks and tools, especially those that cut across frontiers, can greatly benefit the HEP science effort.
2. **Common tools and coding standards; reduced software footprint:** Husbanding software resources in high energy physics, given future funding and manpower limits, strongly argues for increased emphasis on the use of common tools across and within frontiers, and the development of a shared set of 'high energy physics computing best practices' to optimize resources.
3. **Resource support models for smaller-scale projects:** Smaller-scale projects are resource-starved in their ability to exploit complex tools and to develop new ones. To address this need, a support model that cross-cuts across all frontiers would significantly aid these projects.
4. **Data preservation policy for the high energy physics community:** Establishment of data preservation policies within the community that are consistent with DOE–HEP and with broader DOE-SC Data Management Requirements would promote the long-term value and integrity of HEP science.
5. **Distributed Center for Computational Excellence (CCE):** The CCE would provide opportunities for a cross-cutting R&D program that includes compute and data intensive elements addressing the challenges of next-generation architectures, innovative approaches to the use of high performance computing for HEP science, and data-intensive collaborations with the ASCR community, including networking and data transport, as well as partnerships with industry when appropriate.
6. **Multi-level computer and computational science training program:** A program aimed at directly impacting the particle physics science agenda and advancing next-generation software development is expected to have significant scientific impact.
7. **Community-based expert group for high energy physics computing:** Such a group could provide opportunities to continue the community-based approach to address issues in computing and simulation, including addressing the technical details.
8. **Expansion of current interactions with researchers in external disciplines, particularly those in DOE-ASCR community**: New opportunities exist in key areas such as data-intensive science, with emphasis on ASCR computational and data facilities; in initiation of 'triangular' collaborations with researchers from other program offices and industry, in collaboration with the ASCR community; in establishing collaborations with research communities supported by other agencies (e.g., NASA, NSF), as relevant.

1 Introduction: Computing and High Energy Physics

The high energy physics community has played early and significant roles in computing, dating back to the 1950s and the early 1960s. Although by now computing plays a critical role in all of science, its role in high energy physics is particularly pervasive and unique. The uniqueness is expressed by the fact that the field has historically exploited all aspects of "at-scale" computing in a way that no other scientific area has so far approached, although the gap is narrowing. HEP has played leading roles in developing and using high throughput and distributed/grid computing, online (real-time and near-real-time) data processing, high performance computing, high performance networking, large-scale data storage, large-scale data management and analysis, as well as managing a global scale of operations. As a consequence of this scale of activity, there is a well-established history of HEP collaboration with several external organizations and agencies including CERN, DOE-ASCR, ESA, NASA, NNSA, and NSF in several of the above-mentioned areas.

Computing, for the purpose of the current activity, is defined as including all off-line computing and simulation. (On-line computing has a more natural place within the scope of detector technology.) Management, analysis, curation, and storage of data generated by experiment, simulation and theory are all included in this definition of 'computing'.

The HEP science program is divided into the Cosmic, Energy, and Intensity Frontiers. Computing is considered one of the three key HEP enabling technologies (with accelerators and detectors) and simulation one of the key HEP science paths (with experiment and theory). Thus computing and simulation constitute a core competency and capability within HEP. Nevertheless, how computing and simulation are managed and funded within the agency varies widely depending on the particular program. Much of DOE-HEP's computing effort is aligned 'vertically', i.e., largely stovepiped along project and experiment boundaries. In this language, cross-cutting activities would be seen as possessing a 'horizontal' component.

The scale and diversity of HEP computing has been noted above. To obtain some notion of scale of the effort, we note that distributed, high-throughput computing by experiments at the Large Hadron Collider (LHC) uses 250K processing cores and nearly 170 PB of disk, as well as multi-hundred PB capacity tape libraries. High-performance computing across the frontiers consumes in excess of a billion core-hours annually. In terms of global networking, HEP data transport across National Research and Education Networks (NRENs) has been the primary driver of NREN growth.

Computational needs for addressing HEP science requirements are rapidly increasing, especially in the experimental arena. In the Energy Frontier, the physics data that can be analyzed is already limited by the available computing. LHC upgrades in energy and luminosity will likely result in order of magnitude increases in events and event complexity. The raw annual data rate is expected to rise to 130 PB in 2021, further stressing LHC data management, storage, and networking resources. The computing requirements of Intensity Frontier experiments combined are at the scale of a single energy frontier experiment, but add a significant level of complexity due to their diversity. Identifying commonalities between the experiments and leveraging the Energy Frontier effort are key aspects of future efforts in Intensity Frontier computing. Cosmic Frontier experiments will greatly extend their data needs as new surveys and instruments come on line. The total data volume is expected to be about 50 PB in ten years, and in 10–20 years the rate is estimated at 400 PB/yr. The role of theoretical computations and simulations across all of HEP science will increase in importance, as high fidelity modeling of ever more complex experiments will be required.

The charge to the Panel consisted of two sorts of activities: (i) survey of current hardware and software practices in computation and the management and analysis of data across the DOE-HEP community and (ii) identification of gaps, targets for improvements, and new leadership opportunities. (The charge itself can be found in Appendix 1.) Because the scale of the HEP computational effort is so vast and distributed, involving literally thousands of researchers, a very large amount of work was involved in getting to grips with the basic tasks laid out in the charge.

A number of recent requirements workshops, other meetings, and the Snowmass Summer Study documents are a rich source of material for information relevant to the first (survey) aspect of the charge. In addition, members of the Topical Panel, aided by the larger community, collected more material to fill in some of the gaps. Overall, material from more than two hundred documents (some of which can be found in the Appendices) factored into the Panel discussions and, eventually, made its way into this report. Even so, our survey and recommendations can by no means be considered complete; one of our recommendations is that an expert group be set up to track progress in the HEP computational program and provide feedback on a continuing basis.

As may be expected, much of the panel's activity focused on the second aspect of the charge, that of identifying gaps and opportunities. The Panel also discussed a number of strategies, in particular the idea of a distributed Center for Computational Excellence (as mentioned in the charge), to address a number of significant issues and to best exploit future opportunities.

To address the elements of the charge, the co-Chairs were assisted by a 6-member Steering Committee, and five sub-panels in the areas of Accelerators, the three HEP Frontiers – Cosmic, Energy, and Intensity, and Technology.

2 Methodology, Resources, and Meeting Format

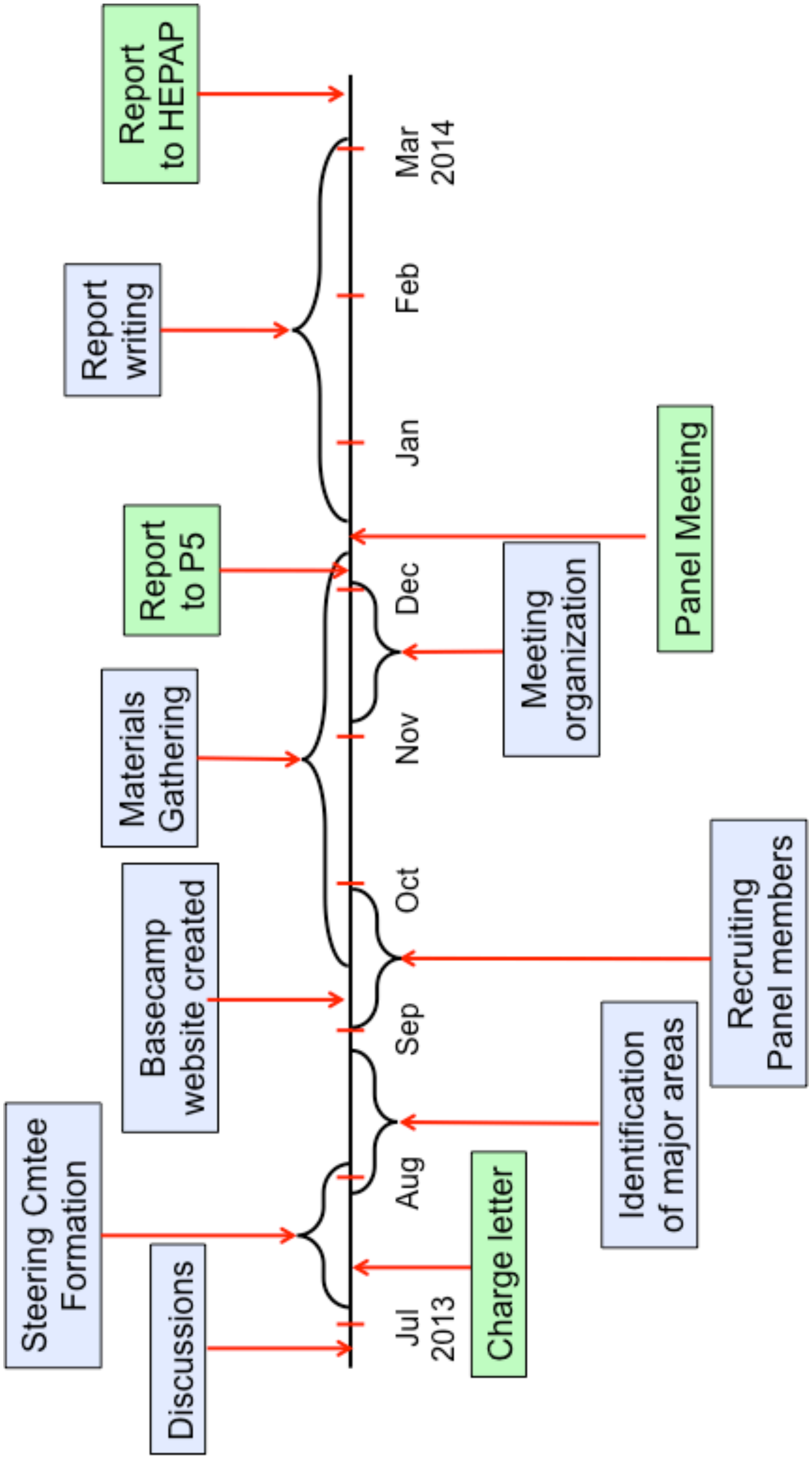
The information gathering process, as mentioned earlier, used a number of recent reports and white papers, including documents from the 2013 Snowmass Summer Study. A few representative examples are the Snowmass Computing Frontier report (August 2013), the HEP-ASCR Summit report (April 2013), and the ESnet HEP-NP Requirements report (August 2013). Details on the materials and how they were used will appear in the following Sections. The complete Snowmass summary can be found in Appendix 3. The website www.hepcomputing.org will contain links to all the publicly available documents as well as a considerable body of material specific to the charge that was gathered under the auspices of the Topical Panel.

The organization of the meeting consisted of five subpanels, with several cross-cutting sessions. The panelists represent a cross-section of expertise spanning the entire domain of HEP research. To address a number of potentially critical issues, experts outside of particle physics were recruited to provide inputs in specific technology areas.

The meeting consisted of plenary sessions only, as it was felt that such an important and wide-ranging topic requires that representatives from all areas be present, and be active participants. Two special sessions were organized; one with Laura Biven (DOE-SC) on policy issues with data preservation and sharing, and another with Barbara Helland (ASCR) on ongoing ASCR-HEP collaborations and future possibilities, particularly in the area of data-intensive science. A substantial amount of work was put in to prepare for the meeting, and a significant effort continued afterwards to complete the final report.

3 Timeline for Topical Panel Activities

The timeline below summarizes in a compact way the historical development of the significant events and activities of the Topical Panel.



4 Summary of Sessions – Main Themes

Science carried out within HEP naturally divides into accelerator technology (for which the office has a key stewardship role in DOE-SC) and into the three research frontiers: Cosmic (dark energy, dark matter, particle astrophysics), Energy (investigations of the fundamental interactions at the highest energies), and Intensity (rare processes with intense beams, neutrino physics). Lattice QCD is listed separately as a major computational activity because it spans two frontiers. Because of the key role of hardware, software, and networking in HEP's computing portfolio, a separate subpanel was dedicated to addressing the tasks in this area.

The main themes that emerged in the individual subpanels are summarized below. Cross-cutting aspects of these themes will be discussed in the next section.

4.1 Accelerators

The main computational activity is the support of accelerator design and technology advancement via electromagnetics and beam dynamics simulations, including those that combine the two. The use of high performance computing, combined with improved algorithms, has led to significant advances in the ability to model accelerating structures, substantially reducing the design cycle, in understanding beam dynamics in new regimes, such as those encountered in plasma acceleration, enabling the development of new technologies, and in accurately modeling multi-physics dynamics effects in high intensity and high energy accelerators, resulting in substantial improvements in performance.

The complexity of accelerator technologies and the large number of possible applications has led to a proliferation of codes, although there is movement in transitioning from individual codes to multi-physics frameworks. In terms of taking advantage of next-generation hardware, accelerator simulation is reasonably well placed. The demands on data movement and storage, in comparison with other areas, are modest. There is a recognition that (i) self-organization across the field needs to improve and be fostered, (ii) a more coherent approach towards code integration be adopted and (iii) dedicated support of the complete code lifecycle is needed, including deployment and maintenance, user support, and training in computational physics relevant to accelerator science.

Until now, the majority of the development of accelerator codes has been in the project context with neither mandate nor dedicated funding for coordination, distribution, or support. A more coordinated approach is needed for general frameworks targeting a user base that extends beyond individual projects, as well as for cross-cutting activities (e.g., exploiting new architectures). Such an approach could be based on fostering a collaboration of existing efforts in the form of a national consortium. The establishment of the consortium would provide a unique point-of-contact for users and enable a wider dissemination of codes. It would accelerate the pace of advances in accelerator science by enabling users to simulate complex problems involving multi-physics simulations that require the integration of modeling capabilities that currently only exist in separate packages. This would enable "virtual prototyping" of accelerator components on a larger scale than is currently possible, and ultimately lead to the modeling of virtual accelerators or experiments.

4.2 Cosmic Frontier

This is the youngest of the frontiers within HEP, possessing a significant diversity due to the merger of a number of historically distinct research areas. On the experimental front, many issues are shared with the Intensity Frontier (e.g., dark matter and high-energy astrophysics experiments), while others are shared with astronomy (e.g., dark energy surveys, cross-correlating multi-wavelength observations). Consequently, development of new tools and integration of tools with very different provenance is a key issue. Specific areas that were identified include data management, workflows, dealing with emerging hardware technologies, and improved data analytics (including new database technologies). Simulations have become an essential component of the work in this frontier, and developing a true 'end to end' simulation capability, particularly for cosmological surveys, is a key goal. Cosmological simulations have taken full advantage of supercomputing resources; this is fully expected to continue.

Data collection in the Cosmic Frontier is characterized by completeness and persistence – the aim is to keep every pixel, and never to throw away any information; additionally the collected data has a very long lifetime, and can remain scientifically relevant for decades. Large-area surveys in multiple wavebands allow for a number of powerful cross-correlation studies, essential for extracting science as well as in controlling systematic errors. For these reasons, heavy use is made of databases, with a need to investigate new database technologies relevant to large datasets. Observational data sizes are currently at the 1-10 PB, with simulations capable of producing much larger datasets. This is expected to increase by a factor of 10-100, with collaboration sizes of ~100-1000 researchers. Unlike the case of the Energy Frontier, however, the resources available to deal with data management and analytics have been inadequate. This gap is partly due to the relative youth of the Cosmic Frontier, but it also points to a need for increased inter-agency coordination among the principal sources of support for cosmic surveys (DOE, NSF and NASA).

Simulations play an ever more important role, functioning as the basic theory, as well as providing synthetic catalogs and modeling resources for fully characterizing observational platforms and systems. In this sense, the simulations play a role analogous to that of Lattice QCD in the Intensity Frontier. Like Lattice QCD, the cosmology simulation effort makes heavy use of supercomputing resources, in the 100's of million core-hours in 2013-14, and partners with ASCR scientists under SciDAC, and with industry (IBM, Kitware, NVIDIA) to optimize performance, and to extend analysis capabilities. The development of simulation capabilities (adding more physics, extending to new architectures) and of the associated data-intensive analytics (shared with observations) is a high priority.

4.3 Energy Frontier

In terms of scale, the computational activity within the Energy Frontier is by far the largest of the frontiers and will continue to be so in the near future. The combined datasets (raw data, data products, simulated data) of the LHC experiments currently totals over 350 PB. This data is managed and distributed by the physics collaborations over a global computing grid, for which the Open Science Grid provides access to US computing resources.

Work in this frontier will need to continue to push the limit of what is possible in order to extract the best science. Looking forward, the immediate task is to deal with an increase of roughly an order of magnitude in data rates and complexity. An important concern is how to attain this goal when, over the same time period of time, the underlying computational and data infrastructure is expected to undergo major changes that cannot be fully predicted (much higher levels of concurrency, lower memory/core, cloud technologies, compute/storage ratio, etc.). The current design, complexity, and size of the software stacks (many millions of lines of code) pose a significant challenge in managing this transition.

A number of ideas are being put forward and tried out in this area – some are more evolutionary, while others explore new directions. Examples include the use of leadership-class supercomputers, federated storage systems and direct remote access to data for more efficient utilization of disk resources, intelligent (dynamic, cached) data placement, new generations of software systems designed to support fine grained concurrency (GEANT4, experiment frameworks, eventually ROOT), seamless integration of cloud resources into processing infrastructure and workflows, expanded use of opportunistic and volunteer computing, leveraging intelligent networking to optimize workflows and dataflows, and using virtual data approaches to flexibly trade off CPU against storage. Many if not all of these directions continue to depend on the excellent networking that has been the foundation and enabler for Energy Frontier computing.

4.4 Intensity Frontier

HEP presently supports 19 experiments classified as being at the Intensity Frontier. These range from collider experiments, such as BaBar and Belle II, to reactor experiments. Some are located outside the US (Daya Bay, KOTO, T2K, BES) but a plurality are hosted at US National Labs, mainly Fermilab. We note that BaBar and Belle II have more in common with the LHC experiments in terms of data models and operations while Dark Matter experiments, classified in the Cosmic Frontier, share many of the attributes of small Intensity Frontier experiments.

The Intensity Frontier program in the US is growing, with most experiments in the planning and early running phases. Typical smaller experiments have 60-200 collaborators and data rates in the 100-1000's of TB per year per experiment. Their data models are significantly different from those for the LHC experiments. In accelerator based neutrino experiments, almost every beam spill contains useful information. For this reason, simulated data samples, in generation, reconstruction and storage, dominate over detector data. In the rare process experiments, such as KOTO and mu2e, simulation can also dominate the computing needs, as extremely rare background processes must be simulated while looking for extremely rare signals. One consensus among the Intensity Frontier experiments is the need for improvements in the physics modeling for low energy processes in GEANT4 and support for continued improvements in the efficiency and robustness of the GEANT4 code.

Over the past decade, the larger collider experiments have developed a sophisticated computational infrastructure to deal with problems in data handling, calibration and geometry management and world-wide job scheduling. These systems are well beyond the capability of any single 100-person collaboration to develop on their own. But the collective scale, in terms of US collaborators and data for the Intensity Frontier is comparable to a large modern collider experiment. Therefore, success depends on fostering inter-collaboration cooperation and the development of common infrastructure.

4.5 Lattice QCD

Lattice QCD calculations are the largest user of HPC resources within HEP's theory program. These calculations provide results for hadronic matrix elements that are an essential adjunct to several past, present, and future Intensity Frontier experiments, e.g. those testing the CKM paradigm, measuring muonic g-2, looking for mu to e conversion, measuring neutrino properties, and making highly precise determinations of Higgs partial widths. Calculations involve roughly equal capability and capacity components, the former running on leadership class resources (e.g. 800M core-hrs of INCITE time in 2013-14), the latter mainly on GPU-accelerated clusters (totaling about 1000M core-hrs in 2013-14). The capability component creates "gauge configurations" which are then used to calculate many physical quantities.

Lattice gauge theorists have long been involved in supercomputer development. Most notably, the Columbia University group designed and built machines (QCDSP in the late 1990's and QCDOC in the early 2000's) that were tuned for Lattice QCD performance and were highly cost-effective. The QCDOC was a forerunner to the Blue Gene/L and members of the group participated with IBM in the design of the Blue Gene/Q. The community has strong connections with GPU manufacturer NVIDIA.

Since 1999, the US lattice community has worked together in the umbrella "USQCD" collaboration, which has created an effective software and hardware infrastructure for US Lattice QCD calculations. USQCD works in the manner of a facility, obtaining support from SciDAC for software and algorithm development (the latter in part in collaboration with SciDAC institutes), from DOE-HEP and NP for hardware purchases, and from the DOE-ASCR INCITE program for HPC. These resources are essential to the success of USQCD. Computer time is allocated competitively to subgroups, an arrangement that maintains vibrancy and innovation within the overall priorities set by the needs of HEP experiments. Gauge configurations are shared (and publicly available). Software development is coordinated, with four packages maintained which all use common low-level routines that are optimized for new architectures. This allowed Lattice QCD to be an early and effective user of GPUs.

4.6 Data Preservation and Data Sharing Session

DOE-SC data management plans are soon expected to be associated with all research proposals. The most likely immediate impact is simply that all data associated with published figures must be made available; this was viewed as a very reasonable requirement. The question of what software should be released remains open. The relevant software should be discussed and fully explained.

Facilities are expected to be able to provide statements about what is available to help support data management plans, so that this information can be included as standard background in all proposals aiming to use these facilities. In a similar vein, large collaborations can and should prepare uniform data management plans, once details on how to do this become available.

Although some experiments in high energy physics have a data preservation policy (e.g., ATLAS), this has not so far been identified as a major goal within HEP experiments. With the new DOE-SC plans close to being announced, it is important that a uniform set of principles be established to enable experimental and theoretical collaborations to design and implement optimal data preservation and sharing strategies. For the Cosmic Frontier there is an additional need for coordinating data management across agencies involved in project support. Given the size and complexity of the datasets involved, and the fact that they cannot be interpreted without a complex software stack, the existence of an associated strategy for data management and analytics software immediately arises.

4.7 DOE ASCR Session

Historically, ASCR has focused on large-scale supercomputing, but there is now an increasing emphasis on data-intensive science. As data rates and volumes grow exponentially across the DOE-SC portfolio, there is increased pressure to develop new data-intensive computational strategies, both within the host science programs, but also within ASCR in its role of providing computational support to the science programs.

Within the traditional HPC role, HEP researchers have a strong history of using ASCR supercomputing resources, in areas such as Lattice QCD, accelerator simulation, and computational cosmology, supported by programs such as ALCC, INCITE, and also by joint ASCR-HEP partnerships such as SciDAC. Areas for continued collaborations in this sector include addressing challenges posed by next-generation architectures and the possibility of using leadership class supercomputers for simulation and data analysis for HEP experiments – an area where pilot studies are already in progress.

Returning to the issue of data, it was noted by Barbara Helland that HEP's experience with managing large-scale data provides an excellent basis for a new set of collaborations in this area. (A joint ASCR-HEP data summit was held in April 2013, identifying a number of concrete areas where joint work would be highly beneficial.) ASCR is studying the idea of a VDF (Virtual Data Facility), a facility that would store as well as allow analysis of the stored data, each VDF being associated with an existing ASCR facility. This was viewed as a significant opportunity for HEP experiments as well as large-scale computational efforts; an ASCR-HEP workshop on VDF concepts was supported as an excellent first step. The idea of 'triangular' collaborations was also supported; these are projects where ASCR and HEP communities would collaborate with a third program office community (e.g., BES in the context of light sources) on data-intensive projects. Other efforts suggested include HEP 'end stations' at ASCR facilities and increased interactions with ASCR SciDAC institutes.

5 Findings and Comments

Each of the areas discussed above have computational challenges that were to be identified under the charge; some of these challenges are unique to a given area, but many do cross area boundaries, with some of the most significant shared across all the frontiers. A large number of the discussion points fell under the first aspect of the charge, i.e., cross-cuts and common solutions, and several high-impact opportunities were also identified. The points of the charge relating to improved software practices and increased efficiency of hardware utilization were also addressed. Considerable attention was paid to the idea of establishing a Center for Computational Excellence. Our responses to the charge are encapsulated in a set of findings and future opportunities.

5.1 Hardware Evolution

These findings are characterized by dealing with challenges posed by the rapid evolution of hardware technology:

1. **Emerging computing architectures pose a major across-the-board challenge to HEP computing**: Commodity machines in the future will look quite different, both in hardware and organization, than those of the current generation. To avoid obsolescence and increased hardware expense, codes must adapt to these changes (increased concurrency, reduced memory/core, I/O bottlenecks, reduced resiliency), which will likely include the use of new programming models. Despite some pioneering activity such as the concurrency forum (established in 2011), the HEP community lacks sufficient expert manpower to carry out this task across the broad span of all of its activities, and a coherent large-scale strategy is currently absent. Not addressing this important issue runs the risk of going down a future path of niche hardware and software (expensive and inefficient, because it no longer exploits the 'commodity train'). The challenges occur across multiple levels such as, (i) re-architecting of workflows, (ii) new approaches for workflow 'nodes', and (iii) new low-level algorithms. The first issue relates to concurrency at a higher level of workflow software, whereas the latter two are particularly hardware-driven (low-level concurrency, memory/core, I/O). The subset of the HEP community that already exploits high performance computing resources is well placed to provide leadership in this area.
2. **Evolution of data archiving, data-intensive computing, and storage will drive new computational strategies:** Unit disk storage costs are expected to decline more slowly than previous historical rates. If data storage and I/O requirements scale faster than the hardware costs, as appear likely, economics will drive the evolution of computing and networking strategies in new directions, including exploiting the benefits of increased data locality by overlapping computing and data resources. New strategies will include the ability to recompute rather than store (virtual data) or to more efficiently analyze data with a smaller storage footprint (integration of networks). As data-intensive computing systems evolve, development of service offerings that better accommodate the needs of HEP experiments will become an important focal point. Other significant areas include overcoming shortcomings of distributed file systems, and investigations of possible non-file system approaches.
3. **High performance computing platforms are an important resource for HEP science:** Access to DOE high performance computing is largely made available through ASCR-managed resources (at ALCF, NERSC, and OLCF). These are already expertly exploited in areas such as Lattice QCD, accelerator physics, and computational cosmology, and nascent efforts are underway to use these resources more directly for data-intensive tasks across all of the science frontiers. Not only do HPC systems provide much needed computational power that would be otherwise inaccessible, they also provide early access to the technologies that underlie the next generation of commodity processing.

5.2 Software Environment

This set of findings is related to the management and development of the software environment:

4. **Code maintenance and distribution lacks well-defined guidelines and support mechanisms:** Outside of projects and facilities support, there is no base mechanism to maintain, update, and distribute codes within the HEP program. Maintenance is an important concern as the underlying computational hardware is rapidly evolving – in many cases, software updates will be necessary simply to maintain the capability at its current level, especially for key software components such as Geant4, Panda, Root and others. A number of other codebases in common use have very little FTE support devoted to their maintenance. Code distribution can be a complex issue in terms of setting policies that promote source code sharing while at the same time protecting commercial value, as needed.
5. **Common tools and coding standards are insufficient:** The stove-piping inherent to project-oriented software development and utilization promotes a lack of awareness of tools built outside of given experiments/frontiers that could otherwise provide broader benefits. Correspondingly, there are associated needs for more uniform data formats and coding standards across the HEP computing effort. Increasing commonality in the software base not only widens the scope of the benefits accrued due to new developments, but also allows researchers to move more easily from one frontier to another, a major advantage given the shortage of manpower and funding resources.
6. **Code diversity is a strategic issue:** Across the HEP frontiers, there is a large diversity of codes, with significant overlaps in capability. As hardware diversifies, the individual development routes taken by different codes will also diversify, leading to a significantly more complex and manpower-intensive software management situation, as already discussed above in the context of software maintenance. The lack of organization of codes within well-defined 'families' means that beneficial strategies are harder to identify, develop, and apply in a coherent fashion. The lack of a minimal set of coding standards, as mentioned above, is also a handicap. There is a need for raising awareness and fostering communication in this area.

5.3 Resource Management

These findings were associated with resource management issues that potentially cross-cut across all Frontiers:

7. **The scale of computational needs of smaller-scale projects is significant:** Smaller-scale projects within the HEP program (~10-100 researchers) typical of the Cosmic and Intensity Frontiers, while not at the scale of the LHC experiments ATLAS or CMS, can nevertheless require a significant amount of coordinated computational support. A robust resource support model for small projects is lacking with respect to providing needed expertise and manpower, of which the latter is the more pressing. There are insufficient mechanisms for collaborative transfer of technologies, codes, and managing FTE support amongst the HEP portfolio experiments; the current project-oriented

methodology does not sufficiently encourage, nor provide explicit pathways for, such collaborative work.

8. **The role of simulations will continue to grow in importance:** It was recognized that simulations play a major role in HEP science. This is true across all the frontiers – as an example from the Energy Frontier, two-thirds of the grid is used for simulations; other obvious examples can be found in the Cosmic and Intensity Frontiers, accelerator modeling, and in HEP theory. As experiments become ever more complex and computational resources become more powerful, this balance will shift towards an even further increased role for simulations, further increasing the demand for computational resources.

9. **Investment in data preservation and the formulation of an associated data policy is lacking:** There is increased pressure to maintain the capability of analyzing data from past experiments. While certain components of the HEP program (e.g., in the Cosmic Frontier) already work under this constraint, and there are efforts such as the BaBar LTDA (Long Term Data Access), a broader strategy is currently lacking. A coherent statement of such a policy would apply across the frontiers, and not be left to individual experiments. A related issue is the maintenance and updating of the associated data analysis software stacks. Commonality in data analysis frameworks will help avoid significant drains on manpower.

5.4 Organization of Computing and Simulation Tasks

These findings focused more on the organization of computing and simulation tasks for optimizing scientific success:

10. **Computational R&D programs are needed:** New R&D programs were discussed in several areas. These programs, broadly speaking, fall into two classes: (i) A more evolutionary set of developments, specified by clear needs from experiments, and (ii) possibly revolutionary approaches targeting large gains in science impact and productivity. From an organizational perspective, the first type of tasks are naturally the responsibility of individual experiments and cross-cutting collaborations; additional flexibility can allow projects of the second type to be undertaken outside of project or facility boundaries.

11. **More uniform interactions with external organizations are desirable:** It is considered highly desirable that there be a single entity functioning as the natural coherent point of contact for computation-related matters between the communities of HEP, ASCR, other DOE offices, and NSF, as well as industry. This could be a natural role for the Center for Computational Excellence discussed below. The aim is to promote broader cooperation, enable early exposure to new technologies, improve planning processes, and intensify the focus on developing transformative solutions. Such interactions will also promote joint responsibility of code development and data management across DOE-supported science.

12. **Training in computational science is lacking:** A shortage of computationally well-trained students and junior researchers was cited as a common problem. Lack of training at multiple levels, i.e., insufficient knowledge of programming paradigms, algo-

rithms, and software tools, is being observed to hold back progress in certain DOE-HEP science areas. Examples of targeted training programs include the existing European CERN School of Computing, active since the early 1970's. The existence of such programs would help considerably in exploiting new technologies and in promoting the use of common tools. At the same time, the need for increased user support was also stressed. Having better-trained users will increase efficiency in the usage of current resources (e.g., reducing so-called 'junk' analyses), as well as lead to improved awareness of what is possible in the future.

5.5 HEP Center for Computational Excellence

Finally, as part of the response to the charge, the setting up of a Center for Computational Excellence was considered:

13. **The formation of a Center for Computational Excellence is viewed as highly desirable:** A Center for Computational Excellence (tentatively, the CCE) would be of great utility in addressing several of the concerns and opportunities identified in the above findings. In particular, it would enable a quick and effective response to a number of identified gaps and well-defined tasks, as well as initiate collaborations and mechanisms for cross-frontier activities and interactions with external entities (ASCR, NSF).

6 Future Opportunities

The role of computing in high energy physics cuts across all Frontiers and research efforts as mentioned in the Introduction; almost all of the science possesses a significant computational element. It is recognized that the close coupling of computation and science is essential to the success of high energy physics campaigns. At the same time the findings above demonstrate that resources could be better utilized if computing is organized so that there is better alignment across multiple thrusts and frontiers, or, in other words, achieve a better balance between the vertical and horizontal nature of the organization of the computational effort.

Many of the findings above, as well as the themes that emerged from the individual frontier discussions, argue for addressing a number of pressing issues that cut across project and frontier boundaries. These include dealing with rapid evolution of technology, management and development of the HEP software environment, and resource management and overall organization of computing. *Consequently, a fundamental observation is that bringing major computing-related activities to a higher level of visibility within the high energy physics community would present opportunities for a more coordinated and optimized approach, especially as it pertains to the software and resource management aspects of the findings listed above*. Achieving and maintaining such a high level of coordination (where the Center would play a powerful supporting role) will require a cultural shift within the particle physics community, but will also offer many new opportunities. On the basis of the findings and discussions, the Panel identified several key opportunities:

1. **Code modernization, maintenance, and dissemination:** Well-defined mechanisms for the continued maintenance and development of major HEP software frameworks and tools, especially for the tools that cut across multiple frontiers, such as Geant4, ROOT, Pythia, and others, would greatly benefit the science effort.

2. **Common tools and coding standards, reduced software footprint:** Maintaining the current software footprint in the light of future funding limits and demands on manpower, strongly argues for an increased emphasis on the use of common tools across and within frontiers, and on the development of a shared set of 'high energy physics computing best practices' to optimize resources.

3. **Resource support models for smaller-scale projects:** Smaller-scale projects are resource-starved in their ability to exploit complex tools and to develop new ones. A support model that in principle cross-cuts across all frontiers, could address this need.

4. **Data preservation policies for the HEP community:** To promote long-term value and integrity of HEP science, data preservation policies that address the needs of DOE-HEP and are consistent with broader DOE-SC requirements should be established.

5. **Distributed Center for Computational Excellence (CCE):** Multiple opportunities could become available via the CCE. The Center could play an important role as a computing-related resource and a recognized collaborative space on computing technology issues for the disparate components of the field, including experiments in the three frontiers, accelerator design and modeling and theoretical calculations. It could also help to organize and coordinate computing training (face-to-face and web) for physicists and students. Very importantly, the Center presents an opportunity to provide key points of contact with computational activities in other programs/agencies (ASCR, NSF) as well as industry and other organizations and forums.

   Initial activities for the Center based on the above findings include (i) exploiting next-generation hardware, (ii) 'HPC for HEP', covering the innovative utilization of HPC platforms for HEP science, (iii) data-intensive science, based on topics already identified at the HEP/ASCR summit in April 2013, and including preliminary work for Virtual Data Facility concepts jointly with ASCR, (iv) networking and data transport, including programmable networks and co-scheduling of data motion and computation.

   The Center as envisaged here constitutes a specific initial response to the findings. It is not intended that the full set of opportunities fall under the umbrella of the CCE.

6. **Multi-level training activities to enable more efficient use of resources and optimize development of next-generation software:** Multi-level training activities in computer and computational science, both hands-on and web-based, present important opportunities for accelerating progress in HEP science. These activities would have separate tracks for theorists, accelerator experts and experimentalists, as appropriate. In addition, increased user support aimed at improving use of HEP software tools would also provide significant returns. The US Particle Accelerator School (USPAS) provides a good model of a successful training paradigm.

7. **Community-based expert group to address issues in computing and simulation:** Although a very substantial amount of work was put in by the Topical Panel in generating the information, findings and opportunities contained in this report, the results are necessarily incomplete. The scope of high energy physics computing is very wide, with significant interconnects across frontiers, experiments, and institutions. It is therefore important to continue this type of activity, by constituting a group of domain experts to

provide information and feedback on computing activities, particularly as it relates to the technical details of the implementations of the observations included here.

8. **Expansion of interactions with the ASCR community – an important next step for high energy physics computing:** New opportunities in data-intensive science require us to expand interactions with members of external organizations, particularly the DOE ASCR community. Following on the already close connections between HEP and ASCR researchers in high performance computing, as exemplified by the SciDAC program, there are ample possibilities for expanding into collaborations involving ASCR facilities as well as in data-intensive science, an area where HEP plays a leadership role. Helping to define the operational role of the proposed ASCR VDF centers, and participating in them, is a significant opportunity for HEP researchers, following on the joint ASCR-HEP data summit in April 2013. Increased interactions with ASCR facilities and research groups will provide important access to new avenues for high energy physics R&D in the areas of next-generation hardware, software, and networking. Finally, the interaction with ASCR-supported researchers also opens up the possibility of 'triangular' collaborations across the DOE science portfolio.

Acknowledgments

The Topical Panel Co-Chairs, Paul Avery and Salman Habib, wish to thank the many people who contributed time, energy and resources to the success of the meeting, including resource contributors not listed in the subpanels. We are also grateful to Lali Chatterjee and Larry Price of DOE-HEP for the critical support and stewardship they provided throughout the entire process.

Appendix 1 – Charge from the DOE Office of High Energy Physics

The Topical Panel Co-Chairs were tasked by the Associate Director of the DOE Office of High Energy Physics, Dr. James Siegrist, with the following charge:

The Department of Energy's (DOE) Office of High Energy Physics (HEP) funds scientific computing and simulations at laboratories and universities to make possible the experimental, theoretical, and advanced technology programs and projects necessary to understand how our universe works at its most fundamental level. HEP scientific computing includes the administration and operation of hardware facilities including storage and data management and software applications including simulations, analysis and reconstruction code, frameworks, and databases. Discovery science through computational physics techniques and advanced scientific simulations is also an integral part of computing for HEP. Since HEP is a data-intensive science, its computing addresses all areas of data management and the transformation of data to scientific knowledge.

DOE-HEP is sponsoring this meeting to identify opportunities and requirements for improving the effectiveness and efficiency of the fundamental reliance on computing and simulation. The meeting should focus on the role of computing and simulation as enabling technologies for the scientific mission of the High Energy Physics program. The meeting should explore models of collaborative efforts that include computer scientists and industry as well as HEP researchers.

We ask that you establish a program committee to organize the meeting which will consist of plenary and breakout sessions.

The goals of this meeting are to:

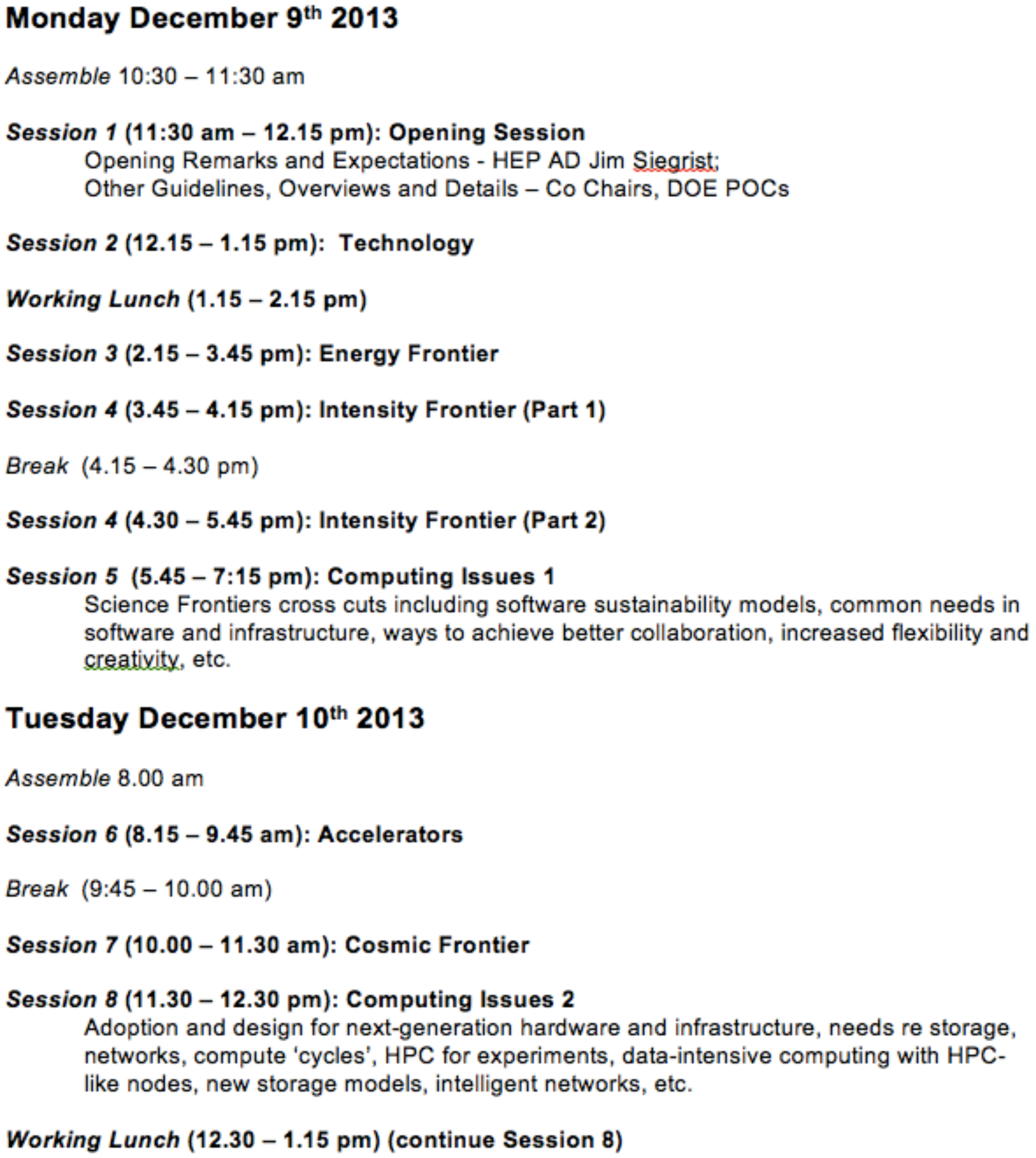
- *Identify the cross cutting aspects of computing and simulations across the different parts of the HEP program that have significant common features and could benefit from common solutions*
- *Identify opportunities for US computational R&D with high programmatic impact including potential leadership roles in the international context*
- *Survey widely used software, including frameworks, codes and data tools that are specific to HEP and*
    - *Identify those that need continued maintenance and updating*
    - *Identify gaps or "missing" tools that are needed by the community*
    - *Identify those that will benefit most from non–HEP partnerships, including more use of tools from other fields than we have traditionally used;*
    - *Identify ways of managing the lifecycle of these tools, including appropriate conditions for phasing them out.*
- *Survey computing and data management practices in different parts of the HEP program and determine whether improved structure would accelerate science*
- *Survey use of computing hardware across the program to identify increased efficiency, cost effectiveness, and best technologies applicable to HEP*
- *Identify the opportunities presented by establishing a Virtual Center for HEP Computing Excellence, consisting of distributed experts in different aspects of computing to promote/facilitate cross cut solutions to the HEP community*

The meeting should be held in the Washington, DC, metropolitan area in the November-December, 2013, time frame. We request that a written report representing the results of the meeting be prepared by you as meeting chairs, with inputs from panel leads, and other assigned writers. The report should specifically address all meeting goals. We would like a draft version of the Executive Summary, containing an overview of the major findings of the meeting, within 7 days after the meeting and a final report within 60 days after the meeting. The final report will be used by HEP to shape out-year program plans and to inform the Office of Science long-range budget planning process.

Dr. Larry Price (larry.price@science.doe.gov) and Dr. Lali Chatterjee (Lali.Chatterjee@ science.doe.gov) of HEP will be your primary DOE contacts for this meeting and will provide any support needed to organize and conduct a successful meeting.

This meeting is an important step toward developing and executing the strategic vision for extracting maximal value from the large investment made by HEP in these areas. Thank you again for agreeing to contribute to this effort.

Appendix 2 – Meeting Agenda



Monday December 9th 2013

Assemble 10:30 – 11:30 am

***Session 1* (11:30 am – 12.15 pm): Opening Session**
Opening Remarks and Expectations - HEP AD Jim Siegrist;
Other Guidelines, Overviews and Details – Co Chairs, DOE POCs

***Session 2* (12.15 – 1.15 pm): Technology**

***Working Lunch* (1.15 – 2.15 pm)**

***Session 3* (2.15 – 3.45 pm): Energy Frontier**

***Session 4* (3.45 – 4.15 pm): Intensity Frontier (Part 1)**

Break (4.15 – 4.30 pm)

***Session 4* (4.30 – 5.45 pm): Intensity Frontier (Part 2)**

***Session 5* (5.45 – 7:15 pm): Computing Issues 1**
Science Frontiers cross cuts including software sustainability models, common needs in software and infrastructure, ways to achieve better collaboration, increased flexibility and creativity, etc.

Tuesday December 10th 2013

Assemble 8.00 am

***Session 6* (8.15 – 9.45 am): Accelerators**

Break (9:45 – 10.00 am)

***Session 7* (10.00 – 11.30 am): Cosmic Frontier**

***Session 8* (11.30 – 12.30 pm): Computing Issues 2**
Adoption and design for next-generation hardware and infrastructure, needs re storage, networks, compute 'cycles', HPC for experiments, data-intensive computing with HPC-like nodes, new storage models, intelligent networks, etc.

***Working Lunch* (12.30 – 1.15 pm) (continue Session 8)**

***Session 9* (1:15 – 2:15 pm): Computing Issues 3**

Data issues across the three frontiers, what is common, and what not. Possibilities for common infrastructure usage, databases, new methods for large-scale data analytics, etc.

***Session 10* (2.15 – 3.15 pm): Computing Issues 4**

(Includes Data Overview by SC Senior Advisor Laura Biven.)
Data discussion continues in a broader context: new DOE rules (data management, long-term data storage and availability), how can HEP connect to other offices, esp. BES, BER, NP, etc.

Break (3:15 – 3.30 pm)

***Session 11* (3:30 – 5.00 pm): Discussion**

Partnerships [includes Overview by ASCR AD Steve Binkley] (Discussion for 'HEP meets ASCR' space – topics TBA)

***Wrap Up Day 2* (5.00 – 6.00 pm)**

Wednesday December 11th 2013

Assemble 8.00 am

***Session 12* (8.15 – 9:45 am): Computing Issues 5**

Previous night, scribes and others go over what has been discussed, and put everything together – this is then gone over in the morning session to make sure we have everything where we want it – what's been done, what we need to work on next, etc.

Break (9:45 – 10.00 am)

***Close Out Session* (10.00 – 11.30 am):**

Roadmap (Ideas for an action plan for the HEP program)

***Wrap Up discussions and Working lunch* (11:30 – 2:00 pm)**

Finalize draft report points and make assignments

Appendix 3 – Snowmass 2013: Conclusion of Computing Section

Below is the Conclusion statement of Snowmass 2013 for computing which can be found at arxiv.org/abs/1401.6117.

9.12: Conclusion

For the **Energy Frontier**, computing limitations already reduce the amount of physics data that can be analyzed. The planned upgrades to the LHC energy and luminosity are expected to result in a ten-fold increase in the number of events and a ten-fold increase in event complexity. Efforts have begun to increase code efficiency and parallelism in reconstruction software and to explore the potential of computational accelerators such as GPUs and Xeon Phi. Saving more raw events to tape and only reconstructing them selectively is under consideration. The LHC produces about 15 petabytes (PB) of raw data per year now, but in 2021 the rate may rise to 130 PB. Attention needs to be paid to data management and wide-area networking, to assure that network connectivity does not become a bottleneck for distributed event analysis. It is important to monitor storage cost and throughputs. More than half of the computing cost is now for storage, and in the future it may become cost-effective to recalculate certain derived quantities rather than storing them.

Intensity Frontier experiments have combined computing requirements on the scale of a single Energy Frontier experiment, but they are a more diverse set than those of the Energy Frontier. We conducted a survey and found that there is significant commonality in different experiments' needs. Sharing resources across experiments, as in the Open Science Grid, is a first step in addressing peak computing needs. Continued coordination of software development among these experiments will increase efficiency of the development effort. Leveraging the data handling experience and expertise of the Energy Frontier experiments for the diverse Intensity Frontier experiments would significantly improve their ability to reconstruct and analyze data.

Cosmic Frontier experiments will greatly expand their data volumes needs with the start of new surveys and the development of new instruments. Current data sets are about 1 PB, and the total data set is expected to be about 50 PB in ten years. Beyond that, in 10–20 years data will be collected at the rate of 400 PB/yr. On the astrophysics and cosmology theory side, some of the most challenging simulations are being run on supercomputers. Current allocations for this effort are approximately 200M core-hours annually. Very large simulations will require increasing computing power. Comparing simulations with observations will play a crucial role in interpretation of experiments, and simulations are needed to help design new instruments. There are very significant challenges in dealing with new computers' architectures and very large data sets, as described above. Growing archival storage, visualization of simulations, and allowing public access to data are also issues that need attention.

Accelerator science is called on to simulate new accelerator designs and to provide near-real-time simulations feedback for accelerator operation. Research into new algorithms and designs has the potential to bring new ideas and capabilities to the field. It will be necessary to include additional physics in codes and to improve algorithms to achieve these goals. Production runs can use from 10K to 100K cores. Considerable effort is being expended to port to new architectures, in particular to address the real-time requirements.

Lattice field theory calculations rely on national supercomputer centers and hardware purchased for the USQCD Computing Project. Allocations at supercomputer centers have exceeded 500 M core-hrs this year, and resource requests will go up by a factor of 50 by the end of this decade. This program provides essential input for interpretation of a number of experiments, and increased precision will be required in the future. For example, the b quark mass and the strong coupling α_s will need to be known at the 0.25% level, a factor of two better than now, to compare precision Higgs measurements at future colliders with Standard Model predictions. Advances in the calculation of hadronic contributions to muon g − 2 will be needed for interpretation of the planned experimental measurement.

Perturbative QCD is essential for theoretical understanding of collider physics rates. Codes were ported to the HPC centers at NERSC and OLCF, and also run on the Open Science Grid. They have also been benchmarking GPU codes and finding impressive speed up with respect to a single core. A computer at CERN was used to benchmark the Intel Xeon Phi chip. A repository of codes has been established at NERSC. A long term goal is to make it easy for experimentalists to use these codes to compute Standard Model rates for the processes they need.

The **Distributed computing and facilities infrastructures** subgroup looked at the growth trends in distributed resources as provided by the Open Science Grid, and the national high performance computing (HPC) centers. Most of the computing by experiments is of the HTC type, but HPC centers could be used for specific work flows. Using existing computing centers could save smaller experiments from large investments in hardware and personnel. Distributed HTC has become important in a number of science areas outside particle physics, but particle physics is still the biggest user and must continue to drive the future computing development. HPC computing needs for theoretical physics will require an order of magnitude increase in capacity and capability at the HPC centers in the next five years, and two orders of magnitude in the next ten years.

The **Networking** subgroup considered the implications of distributed computing on network needs, required R&D and engagement with the National Research and Education Networks (which carries most of our traffic). A number of research questions were formulated that need to be answered before 2020. Expectations of network performance should be raised so that planning for network needs is on par with that for computing and storage. The gap between peak bandwidth and delivered bandwidth should be narrowed. It was not felt that wide-area network performance will be an insurmountable bottleneck in the next five to ten years as long as investments in higher performance links continue. However, there is uncertainty as to whether network costs will drop at the same rate as they have done in the past.

The **Software development, personnel, and training** subgroup has a number of recommendations to implement three main goals. The first goal is to use software development strategies and staffing models that result in software more widely useful to the particle physics community. The second goal is to develop and support software that will run with optimal efficiency on future computer architectures. The third goal is to insure that developers and users have the training necessary to deal with the increasingly complex software environments and computing systems that will be used in the future.

The **Storage and data management** subgroup found that storage continues to be a cost driver for many experiments. It is necessary to manage the cost to optimize the science output from the experiment. Tape storage continues to be relatively inexpensive and should be more utilized within the storage hierarchy. Disk storage is likely to increase in capacity/cost relatively slowly due to a shrinking consumer market and technology barriers. It can be costly for experiments to operate their own distributed data management systems, thus continued R&D in this area would benefit a number of experiments.

To summarize, the challenging resource needs for the planned and proposed physics programs require efficient and flexible use of all resources. Particle physics needs both distributed HTC and HPC. Emerging experimental programs might consider a mix to fulfill demands. Programs to fund these resources need to continue. It may also be possible to use shared computer resources and opportunistic sources of computing to meet some needs. Commercial cloud providers may also provide a useful resource, particularly if prices are reduced. There is increasing need for data-intensive computing in traditionally computation-intensive fields, including at HPC centers.

In order to satisfy our increasing computational demands, the field needs to make better use of advanced computing architectures. With the need for more parallelization, the complexity of software and systems continues to increase, impacting architectures for application frameworks, workload management systems, and also the physics code. We must develop and maintain expertise across the field, and re-engineer frameworks, libraries, and physics codes. Unless corrective action is taken to enable us to take full advantage of the new hardware architectures, we could be frozen out of cost-effective computing solutions on a time scale of 10 years. There is a large code base that needs to be re-engineered, and we currently do not have enough people trained to do it.

The continuing huge growth in observational and simulation data drives the need for continued R&D investment in data management, data access methods, and networking. Continued evolution of the data management and storage systems will be needed in order to take advantage of new network capabilities, ensure efficiency and robustness of the global data federations, and contain the level of effort needed for operations. Significant challenges with data management and access remain, and research into these areas could continue to bring benefit across the Frontiers.

Network reliability is essential for data intensive distributed computing. Emerging network capabilities and data access technologies improve our ability to use resources independent of location. This will enable use of diverse computing resources including dedicated facilities, university computing centers, resources shared opportunistically between PIs, and potentially also commercial clouds. Leadership-class HPC centers may also become relevant for data-intensive computing. The computing models should treat networks as a resource that needs to be managed and planned for.

Computing will be essential for progress in theory and experiment over the next two decades. The advances in computer hardware that we have seen in the past may not continue at the same rate in the future. The issues identified in this report will require continuing attention from both the scientists who develop code and determine what resources best meet their needs, and from the funding agencies who will review plans and determine what shall be funded. Careful attention to the computational challenges in our field will increase efficiency and enable us to meet the experimental and theoretical physics goals identified through the Snowmass process.