Strategic plan for SLAC PPA Scientific Computing

SLAC PPA has core expertise in HEP-related scientific computing including mid- and large-scale data management, community software tools like BlackHat, ACE3P, GEANT4 and xrootd and large-scale data base systems like SciDB/XLDB. Experience and capabilities developed for previous and ongoing experiments like *BaBar* and Fermi can be applied to upcoming mid-scale experiments like CDMS, EXO and CTA. The operating model contains both project funding and multi-laboratory proposals like CCI and LCSim.

SLAC has now established a lab-wide cost model for computer hardware and services which supports basic operations with additional hardware and operating costs being proposal driven and funded. The recharge costs are heavily subsidized and quite competitive with other laboratories and universities, and hence is not an impediment to siting new projects at SLAC.

The strategic plan document for Scientific Computing in SLAC PPA is divided into two parts. The first section contains the major issues facing Scientific Computing in SLAC PPA. The second part contains the strategic plans for individual parts of the SLAC PPA Scientific Computing program.

1 Issues facing Scientific Computing in SLAC PPA

1.1 Support for scientific computing hardware post *BaBar*

1.1.1 Efficiencies of shared resources

BaBar has literally been the elephant in the SLAC computing room for the last 10+ years. The capacity of the SLAC computing center was largely due to servicing *BaBar*'s needs and all other projects were able to piggyback off that installation. Now that *BaBar* is in the final period of science analysis, the hardware requirements are expected to begin contracting to a modest footprint, resulting in the release of much of the large compute base it had assembled. This would noticeably shrink the capabilities of the general batch system and limit the available capacity which has benefited many other small and mid-scale computing projects at SLAC. For example, Fermi data processing is synchronized to periodic data downloads, but easily handles the peak loads because of access to the general batch queues. In addition, a variety of small experiments, ranging from CDMS and EXO to physics performance studies for the ILC have all made use of spare cycles from the large installed base. While recharge center costs can be passed through to the three major users on the site (*BaBar*, ATLAS, and Fermi), a continuation of detector operations funding for access to general computing cycles would maintain the power of the current facility; in the current model this would be about $300k per year. We can probably stretch the lifetime of the existing cores, and we hope to absorb at least some of them in upcoming, growing projects like DES and LSST (and possibly CTA). New projects are strongly encouraged to share their resources in general pools as much as possible.

1.1.2 CCI proposal

The Cosmology Computing Initiative is a multi-lab initiative to provide new computing resources to make large simulation samples in Cosmology and will be critical to understanding upcoming experimental results in Dark Matter and Dark Energy detection. The concern is the effect that lack of funding will have on available computing resources for Computational Cosmology in the post *BaBar* era.

1.2 Support for community tools

SLAC has core expertise in community software tools like GEANT4, xrootd, LCSim and

ACE3P. These tools are proposal driven, multi-laboratory efforts to ensure good user support and stability for community adoption.

1.2.1 GEANT4

SLAC PPA will work with the other laboratories to streamline the GEANT4 management structure to improve the Collaboration's responsiveness to the major resource providers. There is also concern about the lack of EM expertise in the US, which would be remedied by an additional hire into the SLAC GEANT4 team.

1.2.2 xrootd

xrootd is being adopted by more and more projects, including all of the LHC experiments. The need for more dynamic data management systems, by for example the LHC experiments, means that there is also a need to deploy xrootd in more varied environments. This necessitates an increased level of support and development. While xrootd is supported by a collaborative team from SLAC and CERN, there is, however, only one core xrootd developer at SLAC. Additional manpower would enable a more appropriate level of support for this broader set of future applications.

1.2.3 LCSim

LCSim was originally developed to support detector concept evaluation for the International Linear Collider (ILC). The delays in the LHC startup have postponed the Linear Collider timetable. Physics discoveries in any event may suggest energies beyond the 1 TeV capability of the ILC. In the meantime, CERN has continued to develop the CLIC concept as a multi-TeV option. Fermilab is pursuing a planned expansion of Muon Collider R&D as another long-term energy frontier design. LCSim has been identified as an important tool for exploration of detector concepts for CLIC and the Muon Collider. The international community will need to be able to evaluate the physics capabilities of detectors for these collider designs on a common basis and LCSim is the ideal tool for these studies. However, additional development is required and an expansion of user support for it to be widely adopted for this purpose and other applications.

1.3 Data management support for small- and mid-scale experiments

SLAC has core expertise in data management from previous and ongoing large projects like *BaBar* and Fermi. These experiments were large enough to support professional software developers to work on their software infrastructure. This expertise can now be applied to mid-scale experiments like EXO, CDMS and CTA. Besides the advantage of not having to reinvent the wheel, this is also a necessity as these upcoming experiments do not have a sufficiently large size to support the same level of original professional software development. An additional advantage is that these experiments can hit the ground running and make rapid progress with a minimum level of development, planning and support. This does, however, involve a high level of coordination in PPA across different projects and is one of the reasons PPA/SCA was created in 2010.

1.4 Role and support models for SCA

PPA/Scientific Computing Applications is operated as a professional center and gets its funding by charging a fixed rate for services to the projects it supports. This entrepreneurial model encourages selling services as much as possible, but leaves PPA/SCA vulnerable to overall funding ebb and flow. There is no stability nor headroom for taking on new projects in this model. A better model would be to fund a core team from a single source to maintain a stable core capability, which then provides a foundation for expanded effort in support of individual projects. Furthermore, there is a need for scientific computing R&D – a similar model to the above would have 50% of a small core team funding R&D with the remaining 50% participating in PPA projects.

2 Detailed strategic plans for SLAC PPA program elements

2.1 Future plans for GEANT4

The SLAC GEANT4 group will continue to provide management, leadership and technical expertise. The plans for the next three years are:

- Streamline GEANT4 Collaboration management and increase responsiveness and transparent accountability to resource providers
- Position the GEANT4 architecture to meet the needs of future multicore architectures and propose a concrete work plan
- Continue improvements to hadronics to meet increasingly demanding LHC needs
- Complete planned upgrade of high precision neutron code
- Propose funding of a core developer of GEANT4 electromagnetic code (not currently a US responsibility but becoming urgent to meet US requirements).

2.2 Future plans for ATLAS computing

SLAC will continue to work collaboratively with the US ATLAS management to bring SLAC's particular computing strengths to bear in areas of US responsibility:

- The Tier 2 will be operated efficiently and will serve as a basis for support for evolving xrootd functions in ATLAS.
- Xrootd development will continue, responding to the ATLAS need for an increasingly dynamic data management system with particular importance for university based Tier 3 centers.
- PPA/SCA staff with outstanding skills and experience will continue to be available to ATLAS. The current evaluation of the ATLAS production system may lead to a substantial SLAC role in the evolution of this system.

2.3 Future plans xrootd

Xrootd's popularity is growing and it has been adopted by all LHC experiments. In the next years the emphasis will be on consolidation and expansion:

- Continue support for xrootd and make the already extensive documentation more accessible.
- Conduct regional workshops to substantially lower the “adoption barrier”.
- Enhance xrootd to apply to more varied environments.
- Enlarge the xrootd collaboration beyond CERN and SLAC.
-

2.4 Future plans for LCSim

LCSim will work to clarify the path forward while continuing the technical development. Areas of focus for the next years are:

- Position the LCSim architecture to better meet the needs of future users and propose a concrete work plan

- Complete the planned improvements to the silicon tracking persistency model.
- Complete the implementation of the Kalman track and vertex fitters.
- Improve and robustify the current track-finding strategy builder.
- Improve the CAD-to-Geant4 pathway for detector geometry description.
- Improve the LCIO event data model to provide a common basis for HEP data archiving.
- Continue outreach aimed at promoting the use of LCSim for scientific, medical and industrial (e.g. aerospace) simulations.
- Collaborate with Fermilab in providing support for the use of LCSim to explore conceptual designs for future lepton collider applications.

2.5 Future plans for accelerator modeling

With support from SLAC's Accelerator Research Division and the SciDAC program, the Advanced Computations effort benefits the national and international accelerator community by developing and supporting advanced electromagnetic simulation capabilities for the RF design of existing and future facilities in the US and beyond. The areas of focus include: Parallel code development (ACE3P), Computational science R&D under SciDAC, Accelerator modeling and simulations and High performance computing for large-scale simulations.

The plans for the next three years for accelerator modeling are as follows:

- Further advance ACE3P capabilities in multi-physics analysis including electromagnetic, thermal and structural effects, and in simulation of RF breakdown. Continue and expand the series of ACE3P Code Workshops to reach the broader accelerator community.
- Focus on computational science achievements in solvers and optimization to produce results that strengthen ACD's leading role in a SciDAC3 proposal in Accelerators.
- Apply ACE3P to ongoing OHEP projects including LARP/LHC upgrade, Project X, advanced accelerator concepts, and high gradient and superconducting cavity R&D, as well as projects of importance to accelerator facilities in the US and beyond.
- Increase effort to ensure ACE3P's capabilities on multi-core systems at NERSC and NCCS for large-scale simulations to support forefront DOE projects in Accelerator Science and Development.

2.6 Future plans for data archiving at *BaBar*

The *BaBar* Collaboration, the SLAC PPA directorate and PPA/SCA Department and the DOE have set as goal the preservation of the *BaBar* data at least until 2018 to allow the completion of the *BaBar* physics program. The total amount of data to be preserved is about 2 PB. The existing large *BaBar* Computing infrastructure provided by SLAC will be replaced by the *BaBar* Archival System as the primary facility for the analyses of *BaBar* data. The Archival System is an integrated cluster of computation and storage resources using virtualization technologies.

The Archival System is the key element of the *BaBar* Long Term Data Access (LTDA) project. The LTDA is a member of the ICFA (International Committee for Future Accelerators) Study Group on Data Preservation and Long Term Analysis in High Energy Physics. *BaBar* has a leading role in the DPHEP group, being the first large HEP experiment to start a full scale archival project.

2.7 Evolution of PPA/SCA

The PPA/SCA department has now been in existence for a year and will continue its consolidation-and-expansion phase with more emphasis on the latter part:

- Consolidate PPA/SCA as the focal point of scientific computing in PPA.
- Continue support of projects external to PPA like LCLS.
- Dedicated funding for scientific computing.
- Invest in new technologies for future experiments like clound and interactive web applications.
- Continue to encourage software reuse between experiments at SLAC.
- Build collaborations with the SLAC Computing Department, LCLS and other laboratories.

2.8 Future plans for KIPAC computing

KIPAC, a joint SLAC/Stanford Institute hosts a wide array of science programs. The SLAC based KIPAC computing department, with assistance and support from the SLAC Computing Department, develops and maintains the computing environment used for analysis, simulation, storage, archiving, visualization and data-access. Areas of focus for the next years are:

- Refresh existing MPI compute cluster and parallel storage
- Provide adequate data storage growth
- Integration of visualization and data access systems
- Exploit new many-core, GPU and data storage technologies
- Architect system that complements Cosmological Computing Collaboration and DES experiment resources.