APS Scientific Computing Strategy

15 January 2016

The Advanced Photon Source is a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

Table of Contents

0 Executive Summary

The Advanced Photon Source (APS) at Argonne National Laboratory (ANL) is a U.S. Department of Energy (DOE) Office of Science-Basic Energy Sciences (BES) scientific user facility. The core mission of the APS is to serve the scientific community by providing experiment facilities utilizing x-rays (beamlines) to allow users to address the most important basic and applied research challenges facing our nation.

All aspects of APS operation depend on computation, but data analysis software and computing infrastructure are of particular importance for facility productivity. Demands for increased computing at the APS are driven by new scientific opportunities enabled by new measurement techniques, technological advances in detectors, multi-modal data utilization, and advances in data analysis algorithms. The priority for the APS is to further improve our world-class programs that benefit from high-energy, high-brightness and coherent x-rays, all of which require advanced computing. The revolutionized high-energy synchrotron facility proposed in the APS Upgrade project will increase brightness and coherence, leading to further increases in data rates and experiment complexity, creating further demands for advanced scientific computation.

The APS and ANL are poised well to leverage advanced computing to maintain a world-leading position in the synchrotron community. The APS has a world-class photon science program with a large and diverse user base, but ANL is also home to world-leading supercomputing infrastructure and computer science expertise in the Computing, Environment, and Life Sciences (CLS) directorate. This colocation provides an unprecedented opportunity for collaboration. The APS will bring together expertise from within the APS, and across ANL and other facilities in order to maximize impact and leverage existing efforts.

The APS has created a facility-wide strategy, as described below, to adapt to the changing demands of experimental techniques and changes in computing technology. All techniques at the APS will benefit from advances in scientific computing, analysis, and data management. Many techniques enabled by the APS Upgrade, including microscopy, coherence imaging, and time-dependent research will only realize their full potential with advances in computing. This document identifies the core challenges facing the APS in this area, and presents a strategy for addressing these challenges over the coming years within the anticipated resource environment at the APS, and by leveraging synergies with other parts of ANL, and across the BES and DOE complex.

The APS will address algorithm research and development in the priority areas of high-energy scattering, microscopy, coherence imaging, time-dependent, and multi-modal techniques; the implementation of generally applicable, scalable high-performance computing (HPC) applications; and the application of workflow tools to APS beamlines. The APS will not internally have needed resources in this area, but will collaborate with world-leading computer science expertise at ANL. The APS will deploy data management and distribution tools for integration into the operations of APS beamlines. The APS will continue to work closely with world leading experts in data management to leverage best-in-class resources and tools (e.g. Globus Services). The computing and storage resources needed by the APS in the coming years are anticipated to grow by at least one to two orders of magnitude. In order to facilitate these needs, the APS will work to find synergies with existing ANL efforts and other BES and DOE sponsored facilities. The APS is expected to budget new resources for state-of-the-art operational software for beamlines as part of projects to create new or upgrade beamlines with the APS Upgrade.

Appendix A provides a prioritized list of individual software and infrastructure projects required to achieve these goals, along with anticipated completion timeframes, up-front and ongoing effort estimates, and funding sources. Appendix B details a SWOT analysis, and Appendix C describes APS operational constraints and requirements. The complementary document, *APS Scientific Computing Assessment*, provides detailed accounts of current and future computing needs of the APS on a technique-by-technique and beamline-by-beamline basis.

1 Drivers for Scientific Computing

Computing plays a role in all aspects of beamline science. While outside the scope of this plan, computation also drives the designs for the APS Upgrade's MBA lattice. Advanced software, in particular, is playing an increasingly important role in the operation and productivity of beamlines. As data rates increase, it becomes of greater importance that state-of-the-art computation be deployed to provide timely results. New or improved data analysis software and computing infrastructure, as well as development of tools to better assist the process of data collection, are consistently identified as a priority by facility users and reviewers.

New scientific opportunities are created by advances in beamline technology, such as superconducting undulators, improved detectors, high-throughput instrumentation, and multi-modal instruments that can acquire several measurements in a single experiment, but these advances are concomitant with increased computational demands. As well, computational strategies for data analysis are improving, offering better results through use of more extensive and rapid calculations. As measurements become more complex, less direct feedback is available to users to determine if the experiment is progressing properly; performing at least preliminary computations on a near real-time basis is required to ensure that optimum results are obtained.

In addition, the APS Upgrade is currently in the detailed planning stages. The APS Upgrade will deploy a high-energy diffraction-limited storage ring, producing a $10^2 – 10^3$ increase in brightness and major gains in coherence. These gains will enable novel experiments and drastically increase data rates while further multiplying the demands for computing and storage. While all users gain from investments in data collection and analysis, it must be stressed that for some techniques, it may not be possible to conduct and analyze experiments without improved scientific computing.

An example demonstrating where advances in computation are critical is found in the x-ray photon correlation spectroscopy (XPCS) technique, which is a unique tool to study dynamics in materials from micrometer to atomic length scales and time scales ranging from seconds to, at present, milliseconds. XPCS uses the coherence properties of the x-ray beam to probe the spatial and temporal fluctuations of the speckle pattern produced by the scattering of a fully or partially coherent beam from disorder in the sample. XPCS is a highly photon starved technique, as the current storage ring sources only contain a very small coherent fraction. The much more highly coherent beams from the APS Upgrade will extend XPCS to access time scales of nanoseconds. This will enable novel science in a wide range of areas such as soft and hard matter physics, biology and life sciences. In order to take advantage of the enhancements in the source, 2D pixel array detectors are being developed that can capture the fluctuation in speckles with a time resolution of microseconds to nanoseconds. These improved detectors and the improved source creates the challenge of providing real-time feedback at data rates that will increase by factors of 100 – 1,000.

Full-field x-ray imaging can be performed at present with sub-microsecond measurements with micron-scale resolution. The upgraded APS will allow 5-nanometer resolution and lensless imaging approaches, such as ptychography, which will allow resolution to drop to *circa* 1 nanometer, each offering much greater detail for study of specimens important in variety of fields, from microelectronics through biophysics. This imaging can be augmented through combined use of multiple techniques, such as fluorescence, phase-contrast and tomographic reconstructions. The rapid speed of the measurements will allow dynamic studies, where external parameters are varied or a reaction occurs, but will also create currently unmanageable volumes of data. The greater spatial resolution creates much greater sensitivity to experimental artifacts, such as sample displacement, which will require correction using advanced algorithms. Improved data handling and analysis techniques are needed to handle instrument advancements occurring at present; the APS Upgrade requires even more development. Where more than one imaging technique is applied (multi-modal data collection), it becomes possible to improve

corrections for experimental artifacts, since all images apply to a single object. State-of-the-art computational research is needed to address the challenges of: how to develop algorithms to apply corrections within such combined reconstructions; how to perform these computations on extremely large datasets in reasonable timeframes; and how to provide tools for interpreting the resulting >>3 dimensional results.

The ability to resolve inter- and intra-granular information in polycrystalline materials at larger levels of plastic deformation than what is currently achievable is another example of an exciting new capability offered by the APS Upgrade, but only possible through advances in computing. High-Energy Diffraction Microscopy (HEDM) allows non-destructive microstructural imaging of crystalline materials in three dimensions. The technique draws on one of the strengths of the APS, high-energy x-ray brilliance, to generate forward-scattering diffraction spots, which are collected using up to five area detectors simultaneously. The resulting datasets are used to derive crystallographic information for each grain present in a polycrystalline aggregate. Absorption-based tomographic imaging is often performed in concert to assess the overall shape of the polycrystalline aggregate, void content and related geometric features, thus complementing and enhancing the crystallographic information that one acquires from HEDM. With the higher brilliance and smaller beam sizes of the upgraded MBA lattice, this technique may be merged with coherence-based techniques to image more complex materials. Current data rates from this technique can exceed terabytes per day; with advanced detectors and the MBA lattice, increases up to 1,000-fold are anticipated. Advanced computational efforts are essential to allow users to incorporate real-time experiment feedback into experiment data collection, especially for *in situ* measurements. Data reconstructed quickly will allow, for example, the ability to rapidly scan the sample to find rare events such as crack initiation, and adjust applied variables to follow the event with the desired resolution. Significant effort in algorithm and workflow development will be required to full exploit the experimental capabilities enabled by the APS Upgrade, which will allow HEDM, high-energy tomography, and high-energy coherent diffraction imaging to be merged.

The current and future computing requirements required to fully enable these and many other techniques at the APS has been assessed beamline-by-beamline in the complementary document, *APS Scientific Computing Assessment*. All beamlines require investment and advances in scientific software, algorithm and data analysis research, software engineering, workflow tools, data management and distribution infrastructure, and computing and storage resources. To allow identification of resources that could be utilized, XSD science groups have been asked to describe desired software projects, which are compiled in Appendix A.

2 Software and Infrastructure Needs

Advances in computing are required for the APS to realize its potential. Improvements that would allow the full use of existing APS capabilities and the analysis of novel experiments enabled by the high-energy, and increased brightness and coherence provided by the APS Upgrade are described and prioritized in Appendix A. Below they are discussed broken down by area.

2.1 Scientific Software & Data Analysis

Every area of the APS has identified data analysis software needs. Increased data rates require scalable HPC-enabled software of general applicability in areas including time correlations, reconstructions, fluorescence mapping, diffraction space mapping, and coherent imaging. This software will provide results in near real-time to facilitate experiment feedback on the increasingly large and complex data sets. Workflows are needed that automate data analysis, can be integrated into workstation and HPC environments, and that feed metadata and provenance information into a data management system.

In alignment with planned first experiments for the APS Upgrade, new efforts on algorithm research and development are needed in the areas of microscopy, fluorescence, scattering, coherence imaging, time-dependent, and multi-modal acquisition. Novel and advanced algorithms are required to focus precious beam time on getting the most from samples, and to enable greater scientific discovery.

Of equal importance is to put analysis tools into the hands of users after they leave the APS. In cases where smaller-scale implementations, preferably implementing multi-core processing, are able to operate on their APS data sets, the APS can ease the process of software installation onto users' computers through development of software installation kits and instructional materials. Packaging of software for installation on HPC and GPU-based systems is even more demanding, but would also be of value where users wish to perform these computations outside APS. Certain applications may be provided to users for remote access, such as via a web portal.

2.2 Data Management & Distribution

Solutions for data management and distribution are currently being delivered to APS beamlines. They are intended to be integrated into beamline workflows where it simplifies migration and staging of data and automates release of data to users. Without them staff must perform many of these tasks manually. These solutions provide basic capabilities for automated data transfer with ownership control, but the envisioned goal will be facility-wide tools that will provide a full set of capabilities:

- *Automate data transfer:* A mechanism to automatically transfer data from acquisition systems to computing resources, and then to data distribution end-points and any archival systems. Additionally, any policies that may apply to data are enforced automatically, such as data retention or backup policies, or data migration and other features found in hierarchical storage management tools.
- *Data ownership:* Ownership of data is preserved and maintained throughout the entire data management system. In order to help automate the set up of data permissions, the system integrates with the facility's administrative systems for account access and experiment proposals.
- *Metadata cataloging:* Relevant experiment and facility metadata is cataloged along with raw and reduced/reconstructed/analyzed measurement data, and is easily searchable.
- *Provenance tracking:* Maintains information regarding data provenance including transformations performed on the data and information about the software packages and parameters used to perform those transformations.
- *Electronic notebook:* Integration of an existing package would provide a mechanism for electronic capture of what are normally hand-written logs, i.e. users record observations, annotate notes, and attach images and other information just like a paper logbook.

2.3 Computing & Storage Infrastructure

The computing and storage hardware infrastructure is critical not only to maintain basic operations at the APS, but also to provide needed resources to execute reduction, reconstruction, and analysis applications, provide experiment feedback, and to store user collected experiment data. Continued and wise investment and improvement in this area is a high-priority for the facility.

- Real-time data reduction and analysis allows users to monitor the quality of the experimental measurements; without it they run "in the dark" with a concomitant loss of effectiveness. Due to increases in data and new algorithms, faster detectors, and the feasibility of more fast feedback, an order of magnitude increase in the number of computing cores for peak use is anticipated between now and the APS Upgrade, but even at present several XSD groups are seeking more computing resources. Experimental data are usually acquired in bursts and need immediate analysis, creating computing demands that are far from continuous. ANL and the DOE maintain facilities and expertise in this area that should be leveraged to realize APS needs.
- As detector speeds grow, so does the need for low latency data storage. The APS must implement high-performance and high-availability data staging with capacity and performance dictated by anticipated detectors.
- The APS collects approximately 1 – 2 PB of experiment data per year across the entire facility. This will increase significantly as detectors improve. Further, it is estimated that the MBA lattice, when completed, will result in further increases in data rates, leading to a post-APS Upgrade data rate estimated to be at least one to two orders of magnitude greater across the facility. As an example, preliminary evaluations of one cutting-edge experiment enabled by the APS Upgrade indicate the production of data on the order of 1,000 PB per year. Similar trends are observed in other facilities and in other parts of ANL. Partnerships between the APS and ANL and DOE resources will help address this growing need.
- At present, the APS only provides short-term data curation, where users assume the responsibility of long-term data archiving. However, the APS should be prepared to assume at least part of this burden, should sponsor requirements change, or for users who are unprepared to assume archival responsibilities, as well as for data collected for internal analysis. Likewise, data storage will be needed for any beamlines that create data portals, where data are analyzed using facility computing and only results are communicated to the user. A pilot collaboration that leverages ANL and DOE resources for data archiving will allow the facility to provide initial archival services and better understand the costs for potential extension.
- Networking is a key consideration when planning computing and storage resources. Sufficient networking bandwidth will be required within the APS, and within and outside of ANL, and these demands will evolve with the architecture and utilization of computing and storage resources. Likewise, latency must also be considered, since the remote control of APS experiments is only possible when control interactions are implemented nearly instantaneously. Every effort will be made to reduce latency for remote steering applications while maintaining an appropriate level of security. Strong engagement with ANL networking expertise and ESnet is critical to the development of future network infrastructure at the APS.

2.4 Beamline Operation Software

When the software used to operate beamlines is designed with the needs of users in mind, productivity rises. This means that beamlines need to operate intuitively, but flexibly, since some of the most valuable experiments are ones that introduce new experimental approaches. Further, development of a common core of beamline alignment, calibration and experiment operation interface tools will encourage greater homogeneity in design, allowing experience at one beamline to be transferred to operation of others. Likewise, introduction of automation elements, which may be alignment scripts, data reduction workflows

or extended robotic operation, can reduce mundane demands on scientists' time, allowing them more time to innovate and provide scientific support.

Mail-in automation has proven to catalyze very high levels of productivity by lowering the barriers for scientists who are not routine facility users. A standardized facility-wide mail-in support system would allow many more beamlines to offer this service where user participation in data collection is not needed. Where user participation is appropriate, but only to oversee semi-automatic operations, remote operation has also shown to be productive and cost-effective, though demanding to engineer.

Beamline operation software needs to be designed through collaboration between software engineers and beamline staff, with input from both experienced and novice users. Since beamlines operate flexibly and with ongoing innovation and optimization, codes must be configurable and modifiable by beamline staff.

The next generation of instruments should utilize adaptive data collection technologies. Strategies may automatically increase or decrease data ranges, setting densities, measurement trajectories or velocities based on properties of the initial observations. In the new APS project review process, development of operating software for new and upgraded beamlines can be evaluated and resourced.

2.5 Staffing

The communicated list of needs with assigned priorities and effort estimates, both for initial development and for ongoing support and maintenance, are listed in Appendix A. In order to meet only the mission critical needs over the next five years, approximately 30 FTE of fixed-cost development effort (or approximately 6 FTE per year for five years), plus approximately 10 FTE per year of ongoing effort to maintain and support this work are required. A more ideal staffing situation that provides effort to address additional important projects, over the next five years increases demands to 42 FTE of fixed-cost development effort (or approximately 8.5 FTE per year for five years), plus approximately 14 FTE per year of ongoing effort.

However, with foreseeable funding no major expansions in staffing are anticipated. At present the XSD has approximately 6 FTE available for these tasks, augmented by approximately 1.3 FTE per year over the next three years of LDRD-funded staff, a partial FTE from an industrial partner, and approximately 1 FTE coming from CLS staff. The APS has benefitted significantly from joint projects with CLS that have supplied HPC software, innovative data management tools and new infrastructure. ANL management as well as organizations within the DOE have expressed interest in expanding these collaborations.

3 Strategy

The strategy laid out below describes how the scientific computing needs of the APS will be achieved. In order to make optimal use of available resources within the anticipated resource environment at the APS, scientific computing work, as listed in Appendix A, has been prioritized according to areas of most important interest to the APS. XSD management together with scientific and computational staff will review these projects periodically to ensure that they match current needs and to look for synergies across multiple beamlines. APS management will select the tasks to receive effort.

The Computational X-ray Science (CXS) and Scientific Software Engineering & Data Management (SDM) groups are charged with implementing the core strategy. In order to make optimal use of the world-class expertise and resources within the APS, scientific domain experts will be integrated into development teams with software engineers so that each can contribute to the effort with their respective strengths. Likewise, expertise and resources from CLS will be integrated into the same teams with APS domain experts and engineers, in order to best leverage the world-class knowledge and infrastructure available at ANL.

Where possible, the APS will actively seek out further collaborations to leverage effort and expertise. There have been great successes in working with organizations within CLS, principally through LDRD funding. ASCR has expressed an interest in SUF computational problems, so it may now be possible to bootstrap LDRD projects into larger-scale activities. Discussions have been initiated on ways that the APS can collaborate with LBL's CAMERA project. Another organization is the BES Facilities Computing Working Group, which has been invigorated by the SUF directors. While each participating laboratory has its own goals, opportunities for collaboration towards priorities of the APS will be vigorously pursued.

3.1 Scientific Software & Data Analysis

The APS will focus on technique development in support of science enabled by the planned MBA lattice upgrade, which will provide the benefits of a smaller beam profile, increased coherence, and increased brightness particularly at higher x-ray energies. These techniques already number amongst the most data intensive techniques performed at the APS and data rates are expected to grow by multiple orders of magnitude due to improved detectors and the upgraded source. Data reduction and analysis will rely heavily on use of HPC to obtain results with near real-time completion. Where possible, LDRDs and other collaboration methods will be used to explore algorithm development and the creation and porting of codes to HPC platforms. Beamlines not directly driving the APS Upgrade will benefit from the reuse of tools developed for priority applications.

Most software will largely be developed as open source and will be made available on an "as is" basis with user community code contributions encouraged, and work will be performed with a graded engineering approach according to impact and priority. Where possible, development efforts will be split between computational professionals and beamline staff. Active partners from other user facilities and occasionally user groups will be sought to share in development efforts, which may allow for a larger number of supported packages. Packaging and active support can be provided for only a select number of software systems that have been deemed to be most important for the success of APS users.

3.2 Data Management & Distribution

It is expected that the current roll-out of data management solutions for APS beamlines, which integrate with beamline data collection and analysis workflows, will be completed within two to three years. Priority will be assigned to beamlines based on current and anticipated data collection rates and volumes. Solutions will be implemented with a graded approach determined by the operational requirements for robustness, uptime, etc. of individual beamlines. Experience gained from this roll-out will determine

priorities for additional identified needs, such as metadata cataloging, provenance tracking, and an electronic logbook. The APS team will place great emphasis on leveraging best-in-class tools, rather than develop new systems. For example, they will continue to work closely with the Globus Services team in order to not duplicate effort and best leverage DOE and ANL resources, such as Globus Transfer and Globus Catalog. Collaborations will be pursued with other BES and DOE facilities, and organizations outside the DOE complex, and open source tools will be used in order to best meet the needs of the APS in an efficient and cost effective manner.

3.3 Computing & Storage Infrastructure

In the coming few years, the computing and storage resources required by the APS are anticipated to grow by one to two orders of magnitude. To satisfy these needs the APS will adopt a graded approach to resource utilization. Local small-scale resources will be used when sufficient. The APS must explore using shared resources to establish better mechanisms for larger-scale beamline computation, subject to the operational requirements set out in Appendix C, which set constraints on scheduling, data latency and hardware support.

- **Computing:** The APS will seek out synergies between ANL, BES, and DOE funded computing centers for use of shared hardware. Collaborations are underway to prototype mid-scale virtualized and containerized systems with other areas at ANL, and where appropriate explore commercial cloud-based computing options. A longer-range goal will be prototyping preemptive scheduling to allow long-running Laboratory computing tasks to utilize unneeded compute cycles between short-duration beamline data reduction and experiment validation tasks. We are working with other areas of ANL on addressing issues related to HPC and APS needs; to understand how HPC can support APS workloads. This includes investigating large-scale allocations via INCITE and ALCC on leadership resources, and utilization of NERSC. The APS is discussing with CLS how next generation machines might be configured to benefit APS beamline utilization patterns.
- **Storage:** In the mid-term, the APS will design and implement a high-performance and high-availability data staging area with capacity and performance dictated by anticipated detectors. Over the next four years the APS will continue deployment of its data management system, and leverage ANL-wide storage resources. The APS will explore archival storage options and develop a cost effective plan for data archival. For example, FNAL has capacity to archive approximately 600 PB of data and makes it available at cost via inter-laboratory agreements.
- **Networking:** ANL has completed a direct, high-bandwidth network connection between the APS and ANL's major data center in the Theory and Computing Sciences building. Working in conjunction with ESnet and ANL the APS has implemented a *Science DMZ*. These new connections will need further testing and augmentation as demands increase.

In coming years, costs for use of internal vs. externally managed computing resources will be reviewed to ensure that the APS is most wisely using laboratory funding.

3.4 Beamline Operation Software

With foreseeable levels of staffing, the APS will not be able to devote significant resources to improving beamline operation software for existing beamlines. The APS is expected to budget new resources for development of state-of-the-art operational software for beamlines as part of development projects to create new or upgrade beamlines, so that with time a growing percentage of APS beamlines will have user interfaces and data management systems of the same superb quality as the source and experimental hardware. This strategy complements the efforts underway in the XSD Beamline Controls group.

3.5 Staffing

Two groups are charged with software development tasks: The CXS group has been formed in order to meet the challenges presented by advanced algorithm development and to coordinate with external collaborators. The group consists of x-ray scientists with extensive computational experience as well as computational scientists. The SDM group has considerable expertise in application development for HPC platforms, and data management systems. The SDM group consists of full time professional software engineers focused on software creation, HPC development and deployment and data management systems. The two groups work together closely with other APS support groups, such as Beamline Controls and Information Technology.

The APS will provide small increases to staffing levels in the CXS and SDM groups in the coming years, which will meet the effort requirements for only some of the facility's mission critical needs. The APS will continue to seek out collaborations within ANL and the BES and DOE complex and matrix staff in order to offset costs where appropriate. In order to make optimal use of available resources within the anticipated resource environment at the APS, work will be conducted following the priorities listed in Appendix A. Additional synergies may be possible by aligning software projects so that new code may be deployed for multiple groups. CXS and SDM will work with XSD line management to explore this model. Software development will gain considerably when beamline scientists are integrated into code development teams with software professionals. Likewise, development efforts will incorporate appropriate expertise from CLS when possible, to further leverage resources. Beamline staff will continue to develop software for their immediate needs, but beamline scientists are encouraged to bring well-formed software enhancement needs to SDM and CXS.

Appendix A – Projects and Priorities

Table A.1 Scientific Software & Data Analysis

Project	Completion	Funding	Priority	Details	Effort	Ongoing Effort / Year
HPC enabled real-time correlation toolkit	2017	APS Operations / CLS	Mission Critical	Development of an HPC enabled application for real-time photon correlation analysis of time-resolved datasets (e.g. XPCS, surface-XPCS, etc.).	2 FTE	0.25 FTE
HPC enabled real-time XAS analysis toolkit	2018	APS Operations	Mission Critical	Development of an HPC enabled application for real-time analysis of x-ray absorption spectroscopy datasets (XANES, EXAFS, XFM).	3 FTE	0.25 FTE
Ultrafast time-resolved imaging with large-scale MD modeling	2018	LDRD	Mission Critical	Integrated ultrafast time-resolved imaging with large-scale molecular dynamics modeling for in situ data analysis and visualization of energy transport.	2 FTE	
Multimodal imaging of materials for energy storage	2018	LDRD	Mission Critical	Integration of multimodal data from x-ray and electron microscopies in order to understand the interaction of materials at multiple length scales from nano to micro.	2 FTE	
HPC enabled real-time CDI analysis toolkit	2019	APS Operations	Mission Critical	Development of an HPC enabled application for real-time x-ray coherent diffraction imaging datasets (CDI, Ptychography).	2 FTE	0.25 FTE
HPC enabled real-time x-ray scattering analysis toolkit	2019	APS Operations	Mission Critical	Development of an HPC enabled application for real-time reciprocal space analysis of x-ray scattering datasets (SAXS, WAXS, HEDM, XPD).	2 FTE	0.25 FTE
Microstructural Imaging using Diffraction Analysis Software (MIDAS)	Ongoing	APS Operations / Industrial Partner / AFRL	Mission Critical	Development and maintenance of a suite of analysis tools for different modalities (near-field, far-field, and very-far-field) of high energy diffraction microscopy (HEDM)		1 FTE
TomoPy	Ongoing	APS Operations	Mission Critical	Package, maintain, and support the TomoPy reconstruction toolkit.		0.5 FTE
General purpose analysis workflow toolkit	Ongoing	APS Operations / CLS	Mission Critical	Leverage best-in-class workflow tools for use at APS beamlines.	2 FTE	0.5 FTE
General purpose multimodal x-ray analysis tools	Ongoing	APS Operations	Mission Critical	Development of algorithms for integrated analysis of multimodal datasets.	2 FTE	0.5 FTE
Tools for scientific visualization	Ongoing	APS Operations / CLS	Mission Critical	Development and leveraging best-in-class scientific visualization tools for multimodal datasets.	1 FTE	1 FTE
Dynamic tomographic reconstruction development	2018	APS Operations	Strategic	Implementation of dynamic tomographic algorithms on next generation GPGPU processors.	0.5 FTE	
Real-time image processing for area detectors	2018	APS Operations	Strategic	Develop a real-time streaming system for basic image processing algorithms (i.e. background subtraction, etc.) that operates on high data rate detectors.	1 FTE	0.5 FTE
Consolidate small-angle scattering toolkits	2019	APS Operations	Strategic	Consolidate multitude of individual small-angle scattering toolkits at the APS into one general purpose toolkit	2 FTE	0.25 FTE

GSAS-II	N/A	APS Operations	High Impact	Only US-developed comprehensive materials structure analysis package. Replaces GSAS/EXPGUI. Requires supersymmetry and magnetic scattering analysis for completion.	3 FTE	1 FTE
Scedasticity	2017	APS Operations	High Impact	Model free real time statistical comparison data analysis for *in operando* measurements, using diffractograms and images.	2 FTE	0.25 FTE
fullrmc	2017	APS Operations	High Impact	Molecular reverse Monte Carlo modeling package enabled with machine learning and artificial intelligence.	1 FTE	0.5 FTE
GSAS/EXPGUI	Ongoing	None	Low	Comprehensive materials structure analysis package system. >500 citations/year in 2015. Discontinued for lack of resources to continue support.		1 FTE
CMPR	Ongoing	None	Low	Powder diffraction data manipulation and visualization. Widely used at beamlines. Discontinued for lack of resources to update to run in new computing environments.		0.25 FTE

Table A.2 Data Management & Distribution

Project	Completion	Funding	Priority	Details	Effort	Ongoing Effort / Year
Data Management and Distribution System	Ongoing	APS Operations / CLS	Mission Critical	Leverage/integrate tools for managing and distributing APS beamline datasets.		1.5 FTE
Metadata Catalog and Provenance Tracking System	Ongoing	APS Operations / CLS	Mission Critical	Leverage/integrate metadata catalog and provenance tracking tools at APS beamlines.	2 FTE	1 FTE
Prototype electronic notebook system	Ongoing	APS Operations / CLS	Mission Critical	Leverage/integrate an electronic notebook system for APS beamlines.	2 FTE	1 FTE

Table A.3 Computing & Storage Infrastructure

Project	Completion	Funding	Priority	Details	Effort	Ongoing Effort / Year
Prototype high-bandwidth network connection	2015	APS Operations / CLS / CIS	Mission Critical	Complete high-bandwidth network connection between APS and ANL central computing resources, and put into prototype use.	0.5 FTE	
Explore ALCC allocations for using ALCF resources	2016	APS Operations / CLS	Mission Critical	Explore using ALCC allocations in order to gain access to ALCF resources for APS beam time usage.	0.25 FTE	
Reliable high-bandwidth network connection	2017	APS Operations / CLS / CIS	Mission Critical	High-bandwidth network connection between APS and ANL central computing resources meet APS operations requirements for reliability, quality, support, etc.	0.5 FTE	

Utilize ANL central storage for APS data management and distribution	2017	APS Operations / CLS / CIS	Mission Critical	Move APS on-site storage system for data management and distribution to ANL centrally supported storage system.	1 FTE	0.5 FTE
Prototype ANL cloud-computing resource	2017	APS Operations / CLS	Mission Critical	Put cloud-computing system operated by CLS into prototype production use by one or two APS beamlines to determine feasibility.	1 FTE	
Campus-wide shared computing	Ongoing	APS Operations / CLS / CIS	Mission Critical	Through deployment of preemptive scheduling and virtualized computing data analysis, computing integrated with data collection is demonstrated on shared ANL compute resources.	1 FTE	1 FTE
Implement archival data storage system	2020	APS Operations / CLS / CIS	Strategic	Integrate the APS data management and distribution system with a long-term data archival and retrieval system supported by ANL or another DOE facility.	2 FTE	0.5 FTE

Table A.4 Beamline Operation Software

Project	Completion	Funding	Priority	Details	Effort	Ongoing Effort / Year
Generic mail-in automation system	2018	APS Operations	Strategic	Mail-in tools are developed and placed in production at appropriate beamlines.	2 FTE	0.25 FTE
Experiment Control Update	Ongoing	APS Operations	Strategic	Replace spec experiment control software with a modern system, where desired.	2 FTE	1 FTE
Beamline Efficiency Toolkit	Ongoing	APS Operations	Strategic	Continually develop alignment, calibration, and other beamline efficiency tools as a single, reusable toolkit.	2 FTE	1 FTE

Appendix B – SWOT Analysis

Figure B.1 Scientific Software and Data Analysis

Strengths	Weaknesses
• World-leading software efforts in a number of scientific areas, including tomography, ptychography, materials crystallography, small-angle scattering, and x-ray photon correlation spectroscopy. • The APS currently develops and maintains approximately 30 data analysis packages; beamline scientists at the APS have demonstrated leadership in most areas of x-ray data analysis software. • World-class user groups contribute new algorithms and software that expand the scientific productivity of the APS. • APS maintains a modest internal group of professional scientific software engineers and algorithm developers. • ANL possesses world-class expertise in computer science, HPC and data analysis. • Pilot projects with ALCF/CLS help to create new HPC enabled data analysis software.	• Current funding situation does not allow for the APS to meet its entire mission-critical data analysis software needs. • ANL typically funds proof-of-concept via LDRD; APS does not have enough funding to bring proof-of-concept tools into production. • No current prioritized and coordinated software and data analysis effort. • Many scientist-developed packages lack professional software engineering needed to make them more productive; scientist programmers have little modern and HPC experience, limiting their ability to develop needed code. • Lower facility productivity due to lack of data analysis tools.
Opportunities	**Threats**
• A prioritized and coordinated algorithm and data analysis effort will more cost effectively produce/leverage needed software. • Collaborations with ANL expertise will help bring state-of-the-art HPC applications to the APS. • Collaborations with the BESFCWG and CAMERA could amplify development efforts, and provide needed software in a cost effective manner for the entire DOE complex. • The projects in Appendix A could each contribute significantly to improved effectiveness and efficiency at the APS; some may produce science otherwise impossible. • The APS Upgrade-enabled techniques may be fully realized, answering new scientific questions; the APS maintains its position as the most productive light source.	• Without further investment and collaboration in this area, the APS will not fully realize the scientific potential of the APS Upgrade. • With current funding levels the APS will not be able to meet its mission critical needs if it does not seek out collaborations. • Near-term APS productivity will continue to suffer, decreasing as the scientific complexity of new questions require more sophisticated data analysis. • User groups may seek to perform cutting-edge experiments at other light sources where better software support is available. • Other domestic and international light sources have considerably larger and more active software and algorithm development programs that can leapfrog APS leadership.

Figure B.2 Data Management and Distribution

Strengths	Weaknesses
• World-leading expertise at ANL in data sciences, data management and transfer (e.g. Globus Services team). • APS is the DOE's largest data collecting user facility, producing a wealth of scientifically valuable data. • Good understanding of the complexity of the problem, and resources required to address it. • Collaborative efforts continue to form between the APS and expertise elsewhere at ANL.	• No cost-effective facility-wide tools for automating storage, cataloging, metadata management, and provenance tracking. • Preponderance of existing unique solutions at beamlines involving manual, inefficient management steps; no common user experience. • Current manual methods cannot keep pace with increasing data rates. • Lowered productivity due to time taken away from staff and users to address tasks that may be automated.
Opportunities	**Threats**
• Leverage expertise from CLS, UoC, and the Globus Services team. • Reduce cost by leveraging outside software resources and expertise. • Provide a consistent user experience related to data management. • Automatic management of APS data at pre- and post-APS Upgrade data rates. • Increase scientific productivity through automation of data management tasks. • APS will continue to be the world-leading, most productive light source.	• The full potential of the APS Upgrade cannot be realized without managed data workflows. • Potential changes to funding agency mandates for data management and retention could not be met. • Current APS efforts will be quickly outpaced by expertise outside of the APS due to funding limitations. • Lowered scientific productivity due to an inability to keep up with increases in data volumes and rates. • International light sources that have invested heavily in data management software may overtake the APS in terms of scientific productivity.

Figure B.3 Computing and Storage Infrastructure

Strengths	Weaknesses
• ANL has world leading computing and storage resources and expertise in the ALCF and CLS, and resources available for internal use in the LCRC. • APS has core knowledge of beamline computing and storage requirements derived from science use cases. • Pilot projects with ALCF/CLS are helping to determine requirements, create mutual understanding of needs and limitations, and prototype possible solutions. • Support levels are matched to beamline needs; the APS delivers 24/7 support for critical beamline computing resources so that infrastructure failures have minimal effect on beamlines.	• Current computing and storage resources at the APS will not meet the facility's near- and long-term needs. • Expected budgets prevent the APS from maintaining an internal computing environment sufficient for beamline needs. • Duplication of HPC environments at the APS and CLS/CIS is not cost effective. • The APS has no long-term, centralized data archival system or policy. Some beamlines maintain copies of all collected data, but in a heterogeneous, non-cost effective fashion. • Non-APS (e.g. ALCF/LCRC) support policies are not 24/7, which prevents incorporation of those resources in beamline operation workflows. • Preemptive scheduling may not be realized soon enough to meet APS needs.

Opportunities	Threats
• Utilization of ALCF/CLS and LCRC computing resources can provide access to resources on the scale needed by the APS in the near- and long-term. • World-leading infrastructure expertise in ALCF/CLS can help design platforms optimal for APS needs. • Reduced costs due to facility consolidation. • Provides sufficient infrastructure to fully take advantage of the benefits provided by the APS Upgrade; APS will continue to be the world-leading, most productive light source. • Shared use of facilities can allow compute-intensive lab projects to benefit from time between beamline tasks.	• The benefits of the APS Upgrade will not be realized due to the multiple order of magnitude increase in data rates with the current level of computation and storage infrastructure. • New scientific questions won't be answered because the computing and storage infrastructure required are beyond what the APS can currently resource. • International light sources that have invested heavily in infrastructure will quickly leapfrog the APS in productivity.

Figure B.4 Beamline Operation Software

Strengths	Weaknesses
• World-leading beamline operation software in some areas, for example USAXS, tomography, XPCS, and mail-in powder diffraction. • APS leads the worldwide EPICS collaboration and the EPICS V4 initiative. • Dedicated group of beamline operation software engineers that provide 24/7 support.	• Expected budget cannot support a comprehensive strategy related to beamline operations software; no improvements to the XSD beamlines not funded by the APS Upgrade. • Patchwork of solutions to critical problems exists due to a lack of funding in this area; not a cost-effective use of beamline resources. • Lowered beamline productivity due to a lack of proper software for beamline operation.
Opportunities	**Threats**
• Development effort from new beamlines that are developed in the APS Upgrade can supply reusable software modules to be employed in existing beamlines. • Collaboration with other DOE light sources is a cost-effective way to maintain cutting-edge beamline operations software. • Easier to use instruments will make users more productive and lower barriers to research groups that are not currently synchrotron users. • Cutting-edge user groups are drawn to the APS, and further deeper collaborations with beamline staff. • Benefits of the APS Upgrade, which require beamline operation software, will be fully realized, maintaining the APS as the world-leading synchrotron.	• The benefits of the APS Upgrade, which require advanced in data collection and beamline operation, are not fully realized. • APS beamline capabilities improve but usability stagnates, and beamlines are perceived as no longer world-class. • Cutting-edge user groups are drawn to other facilities where experiments are easier. • Experimental techniques and data collection advances at other facilities outpace advances at the APS.

Appendix C – Operational Standards

The data management policy of the APS requires the facility to provide "users with their data in a timely and convenient fashion. Users of the APS, however, are responsible for meeting their data management obligations to their home institutions and funding agencies. The APS does not provide any long-term data archiving or management service. Once data have been provided to each APS experimental group, the user is responsible for managing the long-term retention of his/her data and should not rely on the APS for this service." (https://www1.aps.anl.gov/Users-Information/Help-Reference/Data-Management-Retrieval-Practices)

Likewise, the APS maintains a commitment to operate beamlines reliably and consistently. This requires that scientific computation at the APS be considered in three general classes:

- **1st Stage Computing / Measurement Quality Assurance:** Such computing provides direct feedback to users and beamline staff indicating that their experiment is functioning normally and is providing useful results. It may require analysis of only part of the collected data (such as reconstruction from only selected projections in tomography), or it may require that all data be processed and reduced, such as in the case of x-ray photon correlation spectroscopy (XPCS), or it may be for the purpose of data visualization. Immediate computation is needed to ensure the most productive use of beamtime. At present these computations are not routinely done at the time of data collection for some beamlines. Computing equipment for this work needs to be available at best-effort levels. Maintenance must be available on-call during normal facility operations to address problems.
- **2nd Stage Computing / Data Processing:** This computation converts all collected data into an intermediate result that is often independent of the data collection strategy and that can be interpreted to provide experimental conclusions. Examples of such intermediate results are reduced data such as structure factors for single-crystal diffraction, an atomic pair distribution function, or a tomographic reconstruction computed from the full set of images. This 2nd stage computation is commonly done by beamline staff, but is not always completed in near real-time. In some cases, the 1st stage computation completes this step, such as in XPCS. Computing equipment for this stage can tolerate some downtime, but extended outages (more than several hours) need to be scheduled for times when the APS is not in operation. Maintenance must be available on-call during normal facility operations to address problems.
- **3rd Stage Computing / Data Analysis and Interpretation:** This class of computation is typically done to obtain final results from a measurement, and may include derivation of atomic models to match experimental data, or for techniques such as diffraction microscopy, tomography or ptychography, may be the process of drawing conclusions from analysis of composite images/tomographs. Depending on the beamline and the experiment, this may be done exclusively by users, largely by beamline staff, or in some collaborative fashion. It may involve extensive or very minimal computation. There are no special requirements for maintenance of computing equipment beyond what is routinely done for ANL facilities.

Each of the classes of computation has an influence on requirements for scientific software, data management and analysis workflow tools, and computing and storage infrastructure. These include quality and robustness, availability, scalability, performance, and maintenance and support constraints. With regard to robustness, hardware utilized for computation falls into different tiers depending on the implications for how it can affect facility operation.

- **Tier 0 / Facility-required**: Equipment required for operation of a large numbers of beamlines, such as supporting network services, must be highly fault-tolerant usually through redundancy and automatic recovery capabilities. A mechanism is needed for out-of-hours response by support staff to minimize downtime and manufacturer support contracts must provide for rapid response.

- **Tier 1 / Beamline-required:** Hardware needed for operation of beamlines includes equipment run at beamlines, such as detectors, workstations and IOCs, but can also include centralized hardware that supply network and file storage services. Where possible, this centralized equipment should be high-availability class, which allow beamlines to remain operational even after a hardware failure. For some very expensive devices, such as detectors, this may not be possible. Support staff must be accessible to beamline staff at all hours in order implement fallback operations.
- **Tier 2 / Beamline-ancillary**: Hardware needed for data reduction tasks that are not directly integrated with data collection, such as data staging facilities and data reduction/interpretation steps, can allow for longer outages before impact becomes severe. Typically outages of 6 to 24 hours can be tolerated.
- **Tier 3 / Desktop computing**: All scientists use commodity computers for tasks such as e-mail communication, ANL and APS management systems and their own research. Such equipment should not be integrated into beamline operations because no special efforts will be supplied to provide around the clock support.