***SLAC National Accelerator Laboratory***

2575 Sand Hill Rd,

Menlo Park, CA 94025

**Project Title: Smart Autonomous Sampling for Neutrons and X-rays**

**PAMS Proposal Number: TBD**

Applicant/Institution: **SLAC National Accelerator Laboratory**

Street Address: **2575 Sand Hill Rd., Menlo Park, CA 94025**

PI name: **Christopher Tassone**

Position title for the PI: **Staff Scientist**

PI telephone number: **650-926-3124**

PI email: **tassone@slac.stanford.edu**

DOE National Laboratory Announcement Number: **LAB 20-2261**

DOE/SC Program Office: **BES**

Program Manager: **Dr. Eliane Lessner**

**Collaborators**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Institution Name** | **Name** | **Year 1 Budget** | **Year 2 Budget** | **Year 3 Budget** | **Total Budget** |
| SLAC | Daniel Ratner  Kevin H. Stone  Sam Webb | $717,300 | $703,827 | $689,952 | $2,111,079 |
| ANL | Steve Heald | $400,000 | $400,000 | $400,000 | $1,200,000 |
| ORNL | Stuart Calder | $382,700 | $396,173 | $410,048 | $1,188,921 |
| **Grand Total** | | **$1,500,000** | **$1,500,000** | **$1,500,000** | **$4,500,000** |

Table of Contents

[I. Introduction and Background 3](#_Toc34771855)

[I.1. Introduction and Background 3](#_Toc34771856)

[I.2 Proposal Objectives 4](#_Toc34771857)

[II. Proposed Research 4](#_Toc34771858)

[II.1 Isolating Signal 5](#_Toc34771859)

[5](#_Toc34771860)

[6](#_Toc34771861)

[II.2 Sparse Representation 8](#_Toc34771862)

[13](#_Toc34771865)

[II.3 Associating Structure 13](#_Toc34771866)

[13](#_Toc34771867)

[IV. Timetable of Activities 16](#_Toc34771870)

[IV. 1 Year one Milestones 16](#_Toc34771871)

[16](#_Toc34771872)

[16](#_Toc34771873)

[IV.2 Year two Milestones 17](#_Toc34771874)

[Installation 17](#_Toc34771875)

[Working Prototype 17](#_Toc34771876)

[IV.3 Year three Milestones 17](#_Toc34771877)

[Full Integration 17](#_Toc34771878)

[V. Project Management Plan 16](#_Toc34771870)

[Appendix 1: Biographical Sketch 18](#_Toc34771886)

[Appendix 2: Current and Pending Support 24](#_Toc34771887)

[Appendix 3: Bibliography & References Cited 25](#_Toc34771888)

[Appendix 4: Facilities & Other Resources 26](#_Toc34771889)

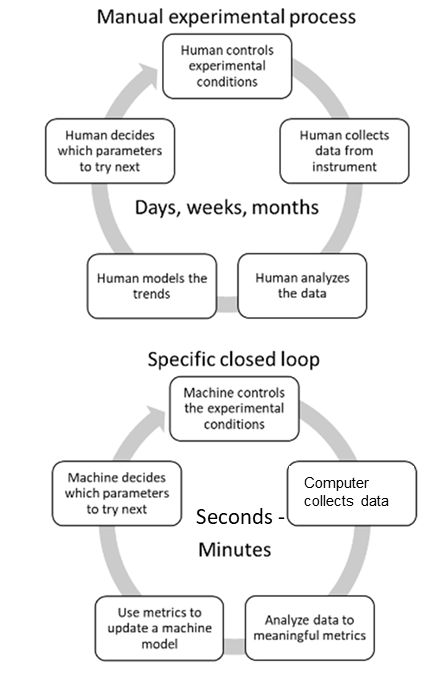
[Appendix 5: Equipment 29](#_Toc34771890)

[Appendix 6: Data Management Plan 30](#_Toc34771891)

[Appendix 7: Other Attachment 32](#_Toc34771892)

1. **Background**

As data rates at the DOE’s scientific user facilities continue to grow, the recent BES roundtable workshop “Producing and Managing Large Scientific Data with Artificial Intelligence and Machine Learning” identified efficient collection of *information* as a critical opportunity.1  Specifically, rather than seeking to maximize the data collected, the facilities should enable experiments to maximize the information content of each measurement.  By continuous, online analysis of data during the acquisition, users can change experimental parameters based on previous acquisitions to target high value measurements, and thereby do more science with less data.2–8 To take full advantage of the increasing data throughput, *adaptive* experiment guidance that leverages emerging machine learning and artificial intelligence techniques must become routine at BES Scientific User Facilities. The proposal here addresses directly the Priority Research Opportunity (PRO) “challenges of autonomous control and experimentation” as a vital need for future science across facilities.



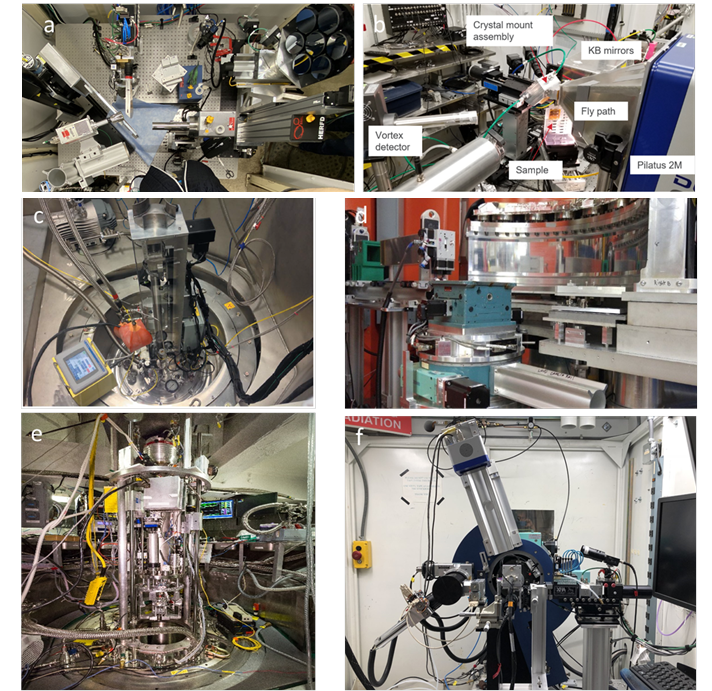
**Figure 1**. Schematic of the experimental process at SUFs. (Top) Current state of experimental decision making at SUF. Users collect data in the absence of actionable information, leading to sub-optimal experimental outcomes. (Bottom) Closed loop smart data collection approaches enables intelligent control algorithms to drive the experimental process informed by prior art, accelerating the acquisition of information rather than data.

**The overarching objective of the proposed work is to address the challenges of autonomous control and experimentation, and thereby demonstrate improved efficiency in the collection of information by automating scientific user facilities (SUF) instrumentation with intelligent control algorithms.** We will accomplish this objective by developing and deploying the components of an automated system. The generalized schematic of a closed loop automated instrument is shown in figure 1. The components necessary to build a closed automated loop are:

* Machine controllable instrument hardware with algorithm interface
* Real time extraction of meaningful metrics to assess information content
* Control algorithms which, given a history of measurements and results (i.e prior art), determine the design of future measurements

In order to demonstrate the applicability across the instrumentation at the SUFs we have chosen to automate several instruments with different design constraints, use cases, access to prior knowledge, and task complexity. With these instruments we will demonstrate and assess the improvement in efficiency at generating information when employing smart data collection strategies. We define smart data collection, as the design of experiments informed by the available prior art in the instrument and/or application space in real time. The efficiency of the data collection is then the quotient of the amount of information generated over the amount of data collected. This metric is flexible enough that it can be evaluated across instruments, and benchmarked against the current data collection strategies at the respective SUFs.

The first prerequisite to automate the SUF instruments is the implementation of machine controllable instrument hardware such that sample positioning motors, detectors, and beam size can be controlled through a programmatic interface; this step has largely been accomplished across the instruments in this proposal shown in figure 2. The second component of an autonomous system is developing a robust metric of success to feed-back upon. In traditional experiments at SUFs this is done in two ways. During an experiment a user will typically assess a superficial metric, such as signal to noise or peak amplitude, as data are being collected. Then, after returning to their home institutions an assessment of data quality is made as users interpret their data set from individual runs using traditional fitting software appropriate for their technique of interest. This approach has a large failing in that the user is not informed at any point during data collection if they have collected data of sufficient quality to answer the scientific questions of interest. While this proposal does not focus on the development of machine learning to automate the interpretation of data, we must develop metrics that enable the control algorithms to assess both whether the data quality is sufficient, and whether enough breadth of data have been collected to address the science questions the users came to the SUF to answer. This is a challenging task as these metrics or functions need to be robust to spurious signals, computationally inexpensive, and contain pertinent information to feedback upon. However, even in cases where superficial metrics have been used in feedback loops for autonomous experiments, significant efficiency improvements have been demonstrated.6,7 Ultimately, these metrics are likely to be replaced by automated knowledge extraction algorithms being developed in response to other priority research directions; however, this work will focus on the simpler task of developing metrics and functions to enable feedback for intelligent instrument control.



**Figure 2**. Photographs of the instruments to be automated for smart data collection. The instruments are (a) BL 10-2 scanning XRF imaging at SSRL (b) multimodal scanning x-ray imaging 25-ID at APS (c) POWGEN at SNS (d) WAND2 and HB2A (not shown) at HFIR (e) NOMAD at SNS and (f) BL 2-1 at SSRL.

Machine learning control algorithms leverage prior knowledge through the training process, which reveals the relationships between the experimental configuration, the subsequent measurement, and the change in information metric. The final prerequisite for autonomous control, is the development and assessment of control algorithms. Despite the relative popularity of active learning in industry, advances are needed to port solutions to scientific problems. Challenges include scaling to large images, handling uncertainties and low signal-to-noise, or incorporating physical information and simulations. SUF experiments can also be multi-modal and hyperspectral, increasing the dimensionality of the information manifold significantly. Consequently, smart data collection at the SUFs will require specific algorithmic development. The control algorithm needs to be tailored to the instrument and the science case in order to make intelligent decisions about each subsequent measurement using the content of the previous measurement. The control algorithm also must do this as rapidly as data are acquired, placing constraints on algorithm architectures depending upon the timescale of measurements. The scope of the proposed work is to develop these individual components, integrate them to exert autonomous control, and evaluate the efficiency enhancement at instruments across three SUFs.

The proposed work will demonstrate and quantify the benefit of intelligent automation of instruments across the SUFs. *In order to ensure that these solutions are generalizable across the facilities we have chosen instruments at the Advanced Photon Source (APS), the High Flux Isotope Reactor (HFIR), the Spallation Neutron Source (SNS), and the Stanford Synchrotron Radiation Lightsource (SSRL), shown in figure 2. Collectively, autonomous control of these instruments will demonstrate applicability across the majority of characterization techniques available at the x-ray and neutron user facilities including x-ray scattering, neutron scattering, x-ray absorption spectroscopy, x-ray emission spectroscopy, and scanning x-ray imaging.*

**II. Proposed Research and Methods**

**II.a. Scientific Motivation**

Data collection with x-rays and neutrons requires optimization of measurement conditions tailored to each unique experiment and dictated by the sample, instrument and interpretation approach. This optimization involves control of dose, motor movement, balance of flux and resolution, background measurements, changing instrumental configurations, and sample environment

parameters. Moreover, each control decision requires analysis and feedback from previous data points. Current data collection strategies are *ad hoc*, inefficient and expensive. Autonomous control to optimize scattering, imaging, and spectroscopic measurements has the potential to provide a paradigm shift in the collection of data at SUFs, by prioritizing the generation of information over data.

This proposal, by developing and deploying smart automation, will answer two fundamental questions that are routinely encountered in x-ray and neutron measurements.

* Firstly, for a single scan what is the optimal measurement condition for a sample? This includes identifying both the sample characteristics and determining additional phases or scattering (e.g. from sample environment) which will influence the measurement thereby allowing an optimized configuration of the instrument.
* Secondly, can smart automation increase the information contained in each measurement by selectively navigating salient regions in a high-dimensional experimental measurement space?

Autonomous experimental control has perhaps the most power when used to sort through a large experimental parametric space.2,5,6,8 For example, a user must select position, path, and spatial resolution in imaging, energy scales, dwell time, and step size in spectroscopy, the reciprocal space dimension in scattering, and the complex scales of experimental variables across all experimental modalities within in situ or operando experiments. In the last case, the goal is to obtain as much information as possible about how the material is changing with time or other external parameters. Autonomous parametric exploration at the SUF instruments has the potential to positively impact nearly every national research priority outlined by the Basic Research Needs reports. For this proposal we focus on three scientific areas to evaluate the improvement in efficiency at generating information. These topics have been chosen because current measurement paradigms exist to use as a baseline to directly compare the information collection efficiency against as part of task 2. Further, improvements in the efficiency with which information is generated can create immediate step changes in the rate of scientific discovery, accelerating the rate at which breakthroughs are made for these applications.





**Figure 3**. (top) Taken from reference [10]. Phase diagram, and associated diffraction signatures, of de-lithiation for LiyMn2-xO4 for compositions used for Li-ion battery electrodes. (bottom)Taken from reference [14]. Phase diagram of the cuprate superconductors as a function of temperature and hole doping. Automation of measurements in multi-sample changers would allow detailed and strategic probing of complex phase diagrams.

**Energy Storage**: Energy storage is of significance for transportation electrification and deployment of renewables on the energy grid. Improvements in batteries require an understanding of structural and chemical evolution during battery cycling. X-ray diffraction, with its rapid data collection, and neutron diffraction, with its high penetration and sensitivity to light elements, are commonly used for structure-property studies of energy storage solutions such as Li- and Na-ion batteries. Advances in this research area will positively impact future energy storage needs. Understanding the structural evolution as the stoichiometry is tuned or altered through charging cycles can be greatly enhanced by autonomous measurements that pick out changes associated with phase transitions, figure 3. For example, in a battery cycling experiment, there may be regions of the diffraction pattern which show the formation of new crystalline phases or diffraction peaks which show characteristic lattice changes in the form of peak shifts or splitting; these areas may be unknown before the measurements, but contain the most information.9,10 In this example, there are two levels of optimization possible, the measurement at a given point of the electrochemical cycle, and how to distribute the collection of measurements over the full cycle. Generally the operating conditions of this type of experiment are predetermined and fixed, for example the cycling rate of a battery. The development of smart control algorithms which adapt the measurement strategies in real time to identify phases of interest will accelerate they key understanding of the relationship between the electrode chemistry, the chemical evolution under operation, and the key performance metrics. The efficiency improvement will be evaluated by contrasting smart automated data collection strategies with previous operando measurements by quantifying the information content as a function of data quantity, with particular interest paid to the identification of any novel phase identification made by the smart control algorithm.

**Quantum Matter**: Quantum materials for future energy applications provide challenges where autonomous data collection would accelerate the rate of discovery. Magnetism and superconductivity are principal manifestations of quantum mechanics in materials.11 Notable here are magnetic frustrated and topological materials which provide diverse and complex magnetic phases that are promising routes for quantum information storage and processing as well as means of controlling macroscopic properties. Examples are magnetic materials on pyrochlore lattices and layered triangular-based compounds having functional properties that are highly sensitive to temperature, electric and magnetic fields, stoichiometry, impurities, and pressure.12,13 Autonomous optimization of the instrument configuration and measurement strategies by leveraging prior models and live data feedback of peak intensities, width and positions will allow the location of important phases, extraction of their properties, and detailed scanning of parameter and measurement space in an efficient way. An example of this is the long-standing investigations of superconductivity which show complex phase diagrams as a function of stoichiometry, temperature, pressure, and magnetic field, see Fig. 3.14 The use of optimized automation combined with available multi-sample changers would allow studies throughout a stoichiometric series to focus on the concentrations and regimes of interest in a systematic and efficient manner.

**Forensic Detection**: Improved methodology for detection and characterization of forensics particles is a common application for x-ray fluorescence imaging and spectroscopy. The term “forensics” can be applied across a large number of fields, from environmental studies to nuclear non-proliferation, where the goal is to determine the origin and fate of trace components in the sample. The first problem is often the detection of these components in a complex matrix, where traditional raster techniques at high-resolution are time-prohibitive. Additionally, the chemical behavior of these components and the areas around them are critical in understanding both past and future transformations. Ideally, probes should cover both spatial and energy dimensions to reveal chemical information. Machine learning and autonomous control have the potential to dramatically increase the efficiency of exploring these high multi-dimensional parameter spaces. These approaches become even more critical as multi-modal techniques, including visible reflectance, x-ray emission spectroscopy (XES), and micro-diffraction, are utilized.

These three science cases highlight the diversity of needs for accelerating the rate at which information, and not simply data, need be generated in order to address pressing science challenges. As part of the proposed work materials in these application spaces will be measured under control of the smart algorithm. We will characterize previously measured systems as a benchmarking activity, as well as novel material systems in collaboration with user groups. The automated characterization of the novel systems will be essential to demonstrate the applicability of this approach to systems which by definition have little prior art. Furthermore, this activity will serve to socialize these practices within the user communities at the respective SUFs.

**II.b. Current Need – Smart Control Systems**

To a large extent the instruments at the SUFs are already machine controllable, with programmatic interfaces to automate data collection. This has been boosted by the increased adoption of the experimental physics and industrial control system (EPICS), that provides a powerful set of software tools to control and program distributed systems common to large experiments. In this proposal we seek to develop “smart” control algorithms which can “react” to the observables present in the data to alter the measurement strategy on the fly. In the absence of such smart control, automation of even straight forward experiments will inevitably result in over- or under- sampling of the measurement space. Even in the case of a static sample probed by XRF or diffraction, the optimal measurement conditions are sample dependent even within a single instrument. For example, the current state of the art for powder diffraction automation is the mail-in programs at 11-BM-B (APS), and POWGEN (SNS), which use a standardized scan strategy for each and every sample governed by a fixed cost that is largely decoupled from the amount of information collected during the measurement. Smart adaptive control algorithms offer the opportunity to couple sample throughput, information content, and instrumental variables, such that every measurement extracts the optimal possible information content from a minimum of data. Even in the case where infinite collection time is available from an instrumental or institutional perspective; the sample itself may not be able to withstand prolonged periods of irradiation or bombardment. Photo-damage of samples is a well-documented phenomenon, and in extreme cases can require careful monitoring of every sample to ensure significant damage is not occurring between successive scans.15,16

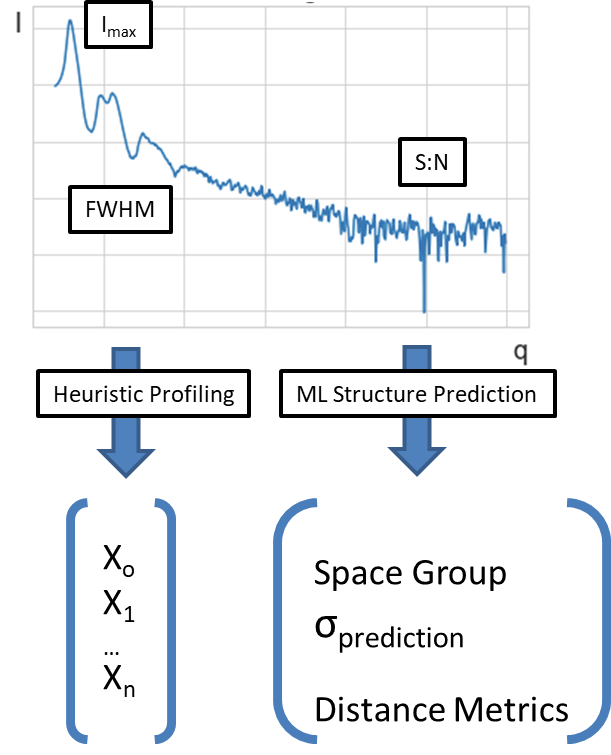
For sequential measurements such as scanning imaging or parametric studies of materials under in-situ or operando conditions it is generally impossible to scan all the potential parameter space in a given experimental time frame. In both of these types of measurements the best case scenario is that the user has performed a careful design of the experiment, informed by the literature, to pre-determine a strategy to scan the parametric space, either real space scan of the sample or in-situ alteration of the sample state, with a pre-determined step size for each parameter. Typically this results in either over- or under- sampling of the parameter space as described above. For example, XRF imaging is typically a raster scan technique, where the sample is scanned pixel-by-pixel in front of the incident x-ray probe to collect data. A sample composition is determined by examining the emission x-ray fluorescence spectrum at each pixel. Generally, information content is sparse due to the area of interest being localized. Raster strategies with such pre-set sampling conditions then record highly redundant data. For sequential multimodal measurements the problem becomes even more complex. It is not only impossible to collect complete data for every technique, but the likelihood of damage induced by the probe becomes more significant.16 In these cases it is imperative that only measurements which result in novel information are performed. Automating these techniques is not possible in the absence of smart control algorithms that are informed by the history of measurements. We will develop autonomous control approaches for two scanning imaging instruments, BL10-2 at SSRL and BL 25-ID at APS. These two instruments offer the opportunity to increase the complexity of the decision-making algorithm, as we move from unimodal XRF mapping to multimodal simultaneous XRF, XES and XRD scanning imaging.

For any of these techniques the question of how to make an optimized measurement becomes drastically more complex when studying materials undergoing phase transitions while varying parameters including magnetic and electric fields, pressure, and temperature. Often, there are subtle changes in the data at specific regions in complex instrument spaces (energy, reciprocal space, sample position) requiring particular focus to reveal the underlying behavior. As a consequence, experiments require the exploration of a highly complex phase space involving a diverse range of measurement signatures. Despite their intrinsic difficulty, these measurements are essential to provide unparalleled understanding of materials for fundamental insights and applications.

The functional properties of materials are generally closely linked to their responses to stimuli, which usually are strongest in the vicinity of phase transitions or depend on the realization of specific phases among a number of subtly balanced and competing phases. The phase diagrams shown in figure 3 demonstrate the inherent challenge. The regions of interest are often the phase boundaries present in the material. A pre-determined scan strategy may under sample these sparse points in the phase diagram. Alternatively, if the steps through the parametric space are fine enough to resolve the phase transition, the scan will collect an enormous amount of redundant information. In order to induce a significant advance in the exploration of high dimensional parametric spaces we will develop smart control algorithms which enable optimal exploration trajectories through these complex parametric spaces at SSRL’s BL 2-1, SNS’s NOMAD, POWGEN, CORELLI instruments, and HFIR’s HB2A and WAND2 instruments.17 In all cases described there is a clear need to develop smart algorithms for automated instrument control to enable the efficient collection of information and not data. To complete this objective the following tasks will be pursued.

**Task 1 – Developing metrics for success:** Autonomous experimental control relies on having robust, information rich, and actionable metrics for feed-back. These are the foundations of the smart autonomous control techniques being developed, as they are the basis by which the machine learning algorithm evaluates the successful outcome of a measurement. The goal of this task is to develop metrics for each of the experimental techniques considered here. The general strategy employed here is to develop a host of potential metrics with increasing complexity, and evaluate the tradeoff in computational expense and efficiency of the automated control algorithm. A general approach will be to assess an experiment as currently run, taking into account any and all data or metadata that is collected or collectable. These raw features will be correlated with what domain experts consider a “successful” experiment, with close communication between domain experts and experts in feature engineering. From these raw data points, more intelligent metrics will be derived which will include feature extraction from completed datasets using statistical approaches as well as more specialized fitting software currently available or developed as part of other efforts.

For the case of x-ray scanning imaging experiments, an experiment might begin by scanning with a relatively large beam size to rapidly gather compositions across the entire sample, and then focus with higher resolution on regions with higher information content. As the duration of a raster scan for XRF scales as the 4th power of pixel size when using an aperture to govern resolution, the ability to scan only the highest information containing areas as determined by unsupervised ML at the best resolutions is important. In XRF scanning, the feedback metric will be given by the reconstructed spatial image, figure 5, determined by the totality of previous measurements at all resolutions. The reconstruction, required for every measurement, is itself an optimization problem, and we have investigated several heuristics with tradeoff in speed and accuracy.18 Future studies will consider additional approaches such as training neural network reconstructions.



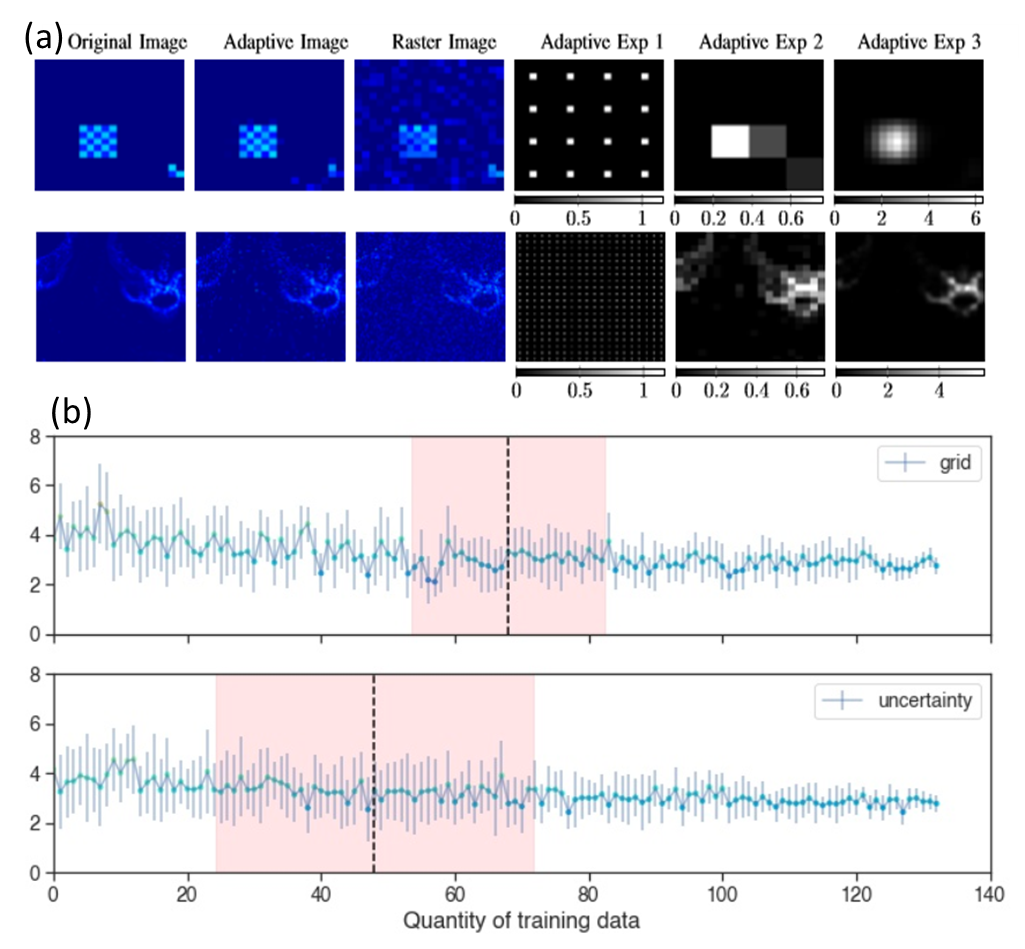
**Figure 4**. Schematic of the metric extraction task showing a sample x-ray scattering pattern (top), which can be profiled into a set of heuristic descriptors (left) or interpreted using existing machine learning approaches to determine the space group, associated prediction uncertainty or distance from prior patterns (right). The computational complexity and risk increase significantly moving from the left to the right.

Defining actionable metrics in XES as well as XRD is more complex since the information content of a measurement is inferred from characteristics such as peak positions, shapes and amplitudes of the XES spectrum or XRD patterns. Different regions of a spectrum or pattern can have peak intensities which differ by orders of magnitude from the maximum. For scanning XRD or XRF measurements, there may be phases present just above the detection threshold which require longer counting times to get good enough signal to noise for a fit. Thus, a sophisticated metric is needed to ensure these lower intensity signals are collected with reasonable statistics. In both such cases, the best such metric ultimately would be the result of the analysis of the experiment, judged by how precisely quantities of interest can be determined, such as atomic positions, lattice parameters, elemental composition or phase transitions and which regions of data best inform them.19 The further development of such sophisticated machine learning knowledge extraction metrics is beyond the scope of the proposed work. We will instead initially focus our efforts on expert guided metrics which can act as heuristics for how informative the measurement will be in the eventual extraction of the quantities of interest in subsequent detailed simulations. This will involve the optimization of the standard observables in all neutron and x-ray diffraction measurements of peak widths, intensities, range covered, background and non-sample scattering that in turn are influenced by specific motor settings in each instrument, figure 4. In the later stage of the project we will move to metrics with more predictive power to guide the control algorithm; for example, metrics will make use of peaks positions, amplitude, width and integrated intensity as identified by existing peak finding algorithms20 or apply convolutional neural networks (CNNs) to predict the space group of a material from a powder diffraction pattern.21–24 Similarly, for XAS/XES, oxidation and spin states predicted from the spectra using neural networks will be used as metrics.25 The uncertainty generated during these predictions should strongly correlate to the information content of the image, as well as the measurement conditions, and therefore provides a simple scalar value for feed-back.

For parametric studies, across both imaging and scattering platforms, it is common for users to view the totality of data overlaid or in a waterfall plot where changes are easily visible. This “by eye” approach can be quantified using distance metrics between images or scattering patterns.26,27 We will compare the use of several approaches, including mean squared difference, optimal transport28, and autoencoders29, to compare distances between scattering patterns or spectra. These distance metrics will then be used for hierarchical clustering of diffraction data. We will also compare the use of visual similarity metrics, including structural similarity index (SSIM)30, visual information fidelity (VIF)31, and features extracted from NNs trained for image retrieval, as distance metrics between images. The use of an easily calculable distance metric can not only show where in a large parameter space changes occur, but also in what regions of the fluorescence or scattering pattern. These will factor in prior knowledge through ancillary modeling as well as include timings for sweeps, control, and configuration changes. On the fly analytics to assess data quality and parameter extraction will be integrated as an essential part of the feedback. This includes statistical processing of the measured signal to identify new features and principle component analysis (PCA) to identify and predict changes. Further, neural networks trained on experimental data and models can discriminate subtle changes in peak-shapes, phases, and chemical compositions, as well as provide fast predictive trajectories for execution by the automated control sequence. The ability to autonomously control the experiment to focus on the changes in the data that are the signatures of interest will greatly enhance the ability to perform parametric studies in a routine way to achieve scientific insights.

The predictive power of these metrics will be evaluated initially in simulated experimental environments, and subsequently during live tests at the instruments. Examining the weights associated with the acquisition function during simulated experiments will vet the predictive power of the different metrics. The tradeoff between metric robustness, predictive power, and computation time will be used to guide the development and refinement of metrics throughout the duration of the project. This represents a standard feature engineering loop, but in this case we are leveraging the domain expertise of the scientists to code in features by which scientists typically evaluate data quality at SUFs.

**Task 2 – Maximizing success metrics:** We will develop two styles of machine learning control algorithms: Bayesian optimization (BO) and reinforcement learning (RL). For either Bayesian Optimization (BO) or Reinforcement Learning (RL), the basic assumption is that there is a real function, F(X), where X is a vector of input parameters, and F(X) yields a vector of metrics or a single metric (e.g. the goodness of fit for a simple peak finding algorithm). Our ability to measure F(X) is limited, thus it is not possible to infinitely probe the parameter space of X. Also, X can be multi-dimensional which can quickly make a thorough exploration of all possible parameters extremely cost prohibitive. Therefore it is necessary to make a “guess” as to how to best improve F(X). Bayesian optimization solves this by employing Gaussian regression to make a guess at the shape of F(X) as well as provide some uncertainty around that guess. BO then uses these guesses and uncertainties to determine which new values of X are likely to improve F(X) (either looking for a minimum or maximum value depending on the metric involved). A trade-off is made between maximizing likelihood of improvement and the uncertainty. Reinforcement learning, in contrast, does not attempt to guess at the shape of F(X) but instead looks at the current state and makes a decision as to the next step to take based on a policy learned from prior experience. The algorithm takes in the current state (still represented by F(X)), and possible actions (how to change X), and uses learned patterns to determine how best to change X. The new state is then evaluated and if F(X) improves this is fed back into the algorithm to better refine the policy. As the model is trained, it will be able to reach an optimal value of F(X) in fewer steps. The downside is this requires a great deal of data to train, while BO is best used in situations where very few measurements are available.



**Figure 5**. (a) This plot shows the reconstructed images for a learned adaptive policy trained using reinforcement learning with the highest resolution raster baseline set to scan for the same amount of time as the adaptive method. The ﬁrst row shows the performance on textured synthetic data. The second row shows the performance of the same policy applied to a real XRF image from SSRL beamline 2-3. Progressing from left to right, the columns depict the original image, the adaptive reconstruction, the highest resolution raster reconstruction, and the exposure times probe size decreasing in each successive experiment adaptive experiment. The colorbars for the exposure time plots indicate seconds per pixel spent in each region. (b) Data taken from a closed loop automated exploration of synthetic parameter space. The data plots the distance from a heuristic which describes the design rule for a reagent concentration vs. the quantity of data used to train the algorithm which determined the heuristic. The top panel used a strategy which sampled the data using a grid search, while the bottom panel uses a Bayesian strategy which prioritizes searching the regions of highest uncertainty, reaching the heuristic more efficiently.

We separate the target tasks described in the previous sections into two styles of problems: experiments that have strong priors, accurate simulations, or large training sets can make use of RL approaches, whereas tasks with weaker priors and limited training sets will likely use BO and related methods. For example, in scanning imaging with XRF, we have a wealth of both synthetic and experimental data: it is feasible to generate realistic images, and even a single SSRL beamline has 100s of mega-pixel scans that can be resampled to tens of thousands of images. XRF imaging also typically has well-defined, consistent goals that can be represented by image quality metrics. As a result, XRF imaging is well suited to reinforcement learning. By contrast, in X-ray scattering, where the datasets are typically smaller and there are profound variations that can arise from factors such as the precise microstructural makeup of the sample, experimental geometry, and instrumental contributions it may only be possible to have general knowledge of priors. In this case, Bayesian optimization will be better suited. Neutron diffraction falls in between these extremes. Direct modeling of data is possible which would allow a RL approach, however data sets can be limited, particularly in the case of magnetic scattering, which would favor a BO approach. In multi-modal studies, additional challenges will be tackled in terms of multi-dimensionality of the objective, F, for Gaussian Process Modeling or reinforcement learning. In general, the boundaries between the methods may not always be clear; we will begin by applying RL and BO approaches to the SUF tasks, and compare the performance and applicability of the two approaches. While some problems may appear more suited to RL or BO, to date there have been no comprehensive comparisons between RL and BO for scientific data acquisition. We expect a direct comparison over a range of different SUF applications to be one of the valuable outcomes of this work.

For RL, initial algorithmic studies have been performed which adapted existing policy-based methods to the context of scientific imaging, figure 5.a. We find that even simplistic models already outperform existing raster scan methods, achieving lower error given the same scan time or reaching the same error level with an order of magnitude shorter scan time. Next steps include scaling to more complex models, higher-dimensional space, and larger training sets.

For BO, we have also taken initial steps to evaluate the suitability for search tasks in high dimensional parametric spaces. Figure 5.b shows the results gathered from autonomous closed loop discovery of syntheses for nanomaterials performed at SSRL. This figure contrasts the number of experiments it takes to determine the heuristic design rule for the effect of one of the reagent concentration using either grid search of the parameter space, or a Bayesian search which selects the next position in phase space by probing the region with the highest uncertainty, clearly illustrating that the Bayesian control algorithm collects the desired information significantly faster. We will adapt these existing algorithms and extend them to the identified instruments. A separate effort focused on machine side has also found physics-informed BO can tune the accelerator up to a factor of 10-fold faster than the existing state-of-the-art.23,32

The example applications for this proposal were chosen in part because of the availability of both historical data and simulations to train and test ML algorithms against current state-of-art approaches. In addition, both Bayesian optimization and reinforcement learning require a simulation environment to develop and test solutions. While it is possible to learn hyperparameters for a Bayesian optimization model directly on training data (e.g. maximum likelihood estimation for Gaussian process kernel), direct comparison to the state-of-art requires a simulation of the full data acquisition process. For deep reinforcement learning, such environments are imperative.

At SSRL, a simulated XRF environment has been written specifically for autonomous control algorithms (Fig. 5). During each training step, the algorithm selects new parameters, and the environment returns a new measurement, experimental state, and evaluation of the desired information metric. When the iteration is complete, the environment provides an information score (quotient of information to data acquired) as described earlier. The simulation environment can take either existing data or can generate synthetic data on the fly. We expect to extend this environment to the additional experimental modes (i.e XAS/XES signals for specific systems), and algorithm architectures covered in the proposal.

We will evaluate the success of the research by the pareto front that balances the experimental metric and the time/dose used to acquire data; the ideal algorithm will outperform the state-of-art for any constraint on time/dose. While research and training will primarily use the simulation environment, performance will ultimately be evaluated during live experiments for the science cases discussed above.

For the three science cases highlighted (energy storage, quantum materials, and forensic detection) the principal investigators have access to previous collected data sets using standard strategies to search parametric spaces. In order to facilitate a direct comparison of the rate of information generation we will prepare samples for these three cases which are meant to be identical to historically measured samples. These samples will then be measured using identical scan strategies to those previously reported, as well as under the control of the smart autonomous control algorithm. The information quotient described previously will be calculated as a direct metric to compare the two strategies. Additionally, we will highlight the identification of any novel phases which were characterized by the smart data collection algorithms, but were previously “dark”. Lastly, we will characterize novel material systems in these application spaces. The specifics of the material systems will be determined in year 2 of the proposal activities. The specific material systems to be characterized will be selected from the available user groups with consideration for the potential impact of accelerated information collection to create breakthrough discovery in the application space.

**Task 3 – Extending Autonomous Smart Data Collection**

While the scope of this proposal is ambitious in developing smart automation capabilities for seven instruments at four SUFs, we recognize that this represents a small fraction of the total instruments available. Ultimately the goal of the proposed work is to extend the foundational tools, and methods developed here across the entirety of the SUFs and instrumentation available therein. Such a task goes well beyond the scope of the proposal; however, as part of this scope of work we will develop a roadmap for extending the automated approaches developed here to the rest of the SUF instrumentation. This will begin with deploying a survey to the SUF instrument scientists designed to assess readiness, suitability and potential gains from integrating smart automation. The results of this survey will be used to code the instruments to a stoplight (red, yellow, green) coding scheme. For each of the of the stoplight levels we will determine cost, and value bands. Additionally, each band will describe the implementation strategy and developments needed to be successful. This will be developed into a comprehensive roadmap for the deployment of smart automated data acquisition across the DOE SUFs.

1. **Timetable of activities (1 page)**

The following table lists the milestones, their completion date and the task that they are associated with. The milestones are interlinked across tasks. The general framework is to develop a feedback metric, code a data pipeline to extract it, simulate the instrument optimization task using that metric, and when successful move to a commissioning run at the appropriate instrument.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Task 1 | Task 2 | Task 3 |
| Q1Y1 | Initial heuristic scattering metric extraction pipeline coded | Extend RL simulation environment to include APS XRF Instrument costs |  |
| Q2Y1 | Computation costs evaluated across proposed scattering metrics | Extend Bayesian simulation environment to BL 2-1 and POWGEN |  |
| Q3Y1 | scattering metric predictive power evaluated in Bayesian sim environment | Extend RL simulation environment to include costs for multimodal XRF and XES |  |
| Q4Y1 | ML based scattering metric pipelines developed | Commissioning run of RL control of XRF and XES instrument |  |
| Q1Y2 | Simulated and Existing data aggregated for parametric mapping of science cases | Commissioning of Bayesian control of scattering measurement |  |
| Q2Y2 | Success metrics prototyped for simultaneous XRF, XES and XRD | Simulated comparison between RL and Bayesian approaches for XRF & XES performed |  |
| Q3Y2 | Multimodal success metrics evaluated in simulated environment | Second generation control algorithms tested for all instruments tested in simulation environments | Survey to SUF instrument scientists developed |
| Q4Y2 | Simulated test of success metrics for Bayesian search through parametric space performed | Commissioning of second generation algorithm architectures performed on scattering, XRF mapping, and multimodal mapping instruments | Survey to SUF instrument scientists received by project team |
| Q1Y3 | Tabulated metric performance and computational costs generated | Commissioning of smart parametric mapping | Inventory of instruments and control systems across SUFs completed |
| Q2Y3 | Final success metric pipelines coded | Efficiency of smart parametric mapping quantitatively evaluated | SUF Instruments coded with stoplight evaluation |
| Q3Y3 |  | Final control algorithm pipelines coded, and documented | Cost and value bands for stoplight evaluation developed |
| Q4Y3 | Smart automation of seven SUF instruments, with quantitative evaluation of efficiency improvements reported | | Roadmap report for SUF smart automation delivered |

1. **Project Management Plan**

To achieve the proposed objectives we assembled a strong, diverse team of complementary researchers consisting of instrument scientists, computational physicists, and computer scientists. The management plan described below was developed to ensure that the investigators can safely achieve these scientific goals while ensuring full compliance with DOE requirements. The objective of the management plan is to ensure that the PIs are not burdened with administrative tasks, foster an environment for collaborative development, and are empowered to focus on conducting leading edge science.

The management structure for the proposal reflects the primary tasks which must be managed and integrated across facilities to successfully complete the project objectives. The overall project will be managed by the principal investigator, Tassone (SLAC). The administration of the code repository for the project will be managed by Ratner (SLAC). This task will include setting up and maintaining the code repository, including reviewing commits, setting unit test schedules and requirements, and reviewing code sprint schedules across institutions. This will also include the contribution of trained models and software to the repositories specified in the data management plan. The development and maintenance of a structured repository of data to train control algorithms, while not the focus of this work, is essential to complete the requisite tasks. Chan (ANL) will oversee the creation of the data repository and the aggregation of data across the participating laboratories. This activity is anticipated to be on-going throughout the proposal period with data being contributed during live runs of the autonomous instruments. This task will include the identification of data required to train the desired control algorithms and lead the aggregation of existing data, or determine pathways to simulate the necessary data, as well as aggregate data generated during trial runs of the control systems, and finally to contribute the datasets to the repositories identified in the data management plan. The proposed work aims to automate a disparate set of instruments with a wide range of data acquisition and control systems, and different end uses. In order to ensure that the development needs are met for each instrument, and that synergies are exploited during development of autonomous control systems Sun (ANL), Stone (SLAC) and Calder (ORNL) will lead the Instrument Automation task. Together they will oversee the aggregation of technical specifications for instrument control systems, coordinate testing schedules across facilities, and identify development task synergies across instruments and institutions. The decision to have co-leaders for this task was made in acknowledgement of the differences between facility infrastructures. Together Calder, Chan, Ratner, Stone, Sun, and Tassone will form the management team for the project. They will meet Monthly to review project progress, resource allocation, and develop tasks, and maintain task coordination.

In addition to the management team, the proposed work will require coordination between staff, postdocs and students at several laboratories. The collection of PIs will directly manage and oversee the work of the postdocs and students hired to perform project activities. A large component of the proposed work is the development of algorithms and software, in addition to implementing those algorithms at SUF instruments. This presents a challenge to managing the project as coordination of effort needs to occur with high frequency, while not overburdening the team with meetings. We will exploit modern collaborative development tools, reinforced with a frequent meeting schedule. First, the proposal team will hold annual meetings that will rotate among the participant laboratories. This will provide an opportunity to frame and disseminate results to laboratory staff who are not participating, and to observe directly the automated instruments at each facility. Second, the proposal team will hold monthly coordination meetings, staggered by two weeks from the management meeting to review results, plan work, and discuss risks and their mitigations. Lastly, a collection of tools will be used to facilitate frequent communication among the project participants. Github will be used as a code repository. This tool enables visibility of the code base by all participants, facilitates task assignments, and provides version control across the labs. Weekly virtual huddles will be held among the postdocs and co-PIs. The purpose of the huddle is to report progress on associated tasks, and to define code sprints for the coming week. Collectively this structure will build cohesion among the project participants, ensure that deliverables are met in a timely manner, and enable mitigation of risks. Results will be disseminated at the annual meetings, as well as through peer reviewed publication and presentation at conferences.

**IV. Project Objectives**

The primary objective of the proposed work is to develop the approaches to enable smart automation of instruments across the neutron and x-ray scientific user facilities. Specific tasks in support of the overall are:

* The development and validation of metrics to drive autonomous decision making by control algorithms for x-ray scattering, neutron scattering, x-ray fluorescence, x-ray emission, and scanning x-ray imaging.
* The development of algorithm architectures which can control instruments on the timescales with which data is acquired
* Quantitative evaluation of the efficiency improvement for information generation when smart data collection is implemented versus the current state of the art autonomous data collection strategies for three science cases
* Develop fully autonomous smart instrument control for seven instruments at SUFs
* Roadmap the further implementation of smart automation across the neutron and x-ray user facilities instruments, including estimated ROI for each technique.

Accomplishing these tasks will not only demonstrate the potential of smart automation to impact SUF operation, but will also provide the foundational approaches to extend these tools to instruments not covered under the proposed work. Additionally, we will build the workforce within the National Lab complex to implement these methods across the facilities following this foundational work.

# Appendix 1: Biographical Sketch

**Christopher J. Tassone, PhD**

**Education/Training**:

2000-2004 Santa Clara University, Bachelor of Science in Chemistry

2005-2011 University of California Los Angeles. PhD. Physical Chemistry. Advisor: Sarah Tolbert

2011-2013 SLAC National Accelerator Laboratory. Postdoctoral Scholar Advisor: Michael F. Toney

**Research and Professional Experience:**

2013-Present SLAC National Accelerator Laboratory, Staff Scientist.

**Selected Publications:**

1. **Tassone, CJ**. Ayzner, A. Kennedy, R. Wanger, D.W. Day, C. Halim, M. Craig, I. Clavaguera, S. Clark, A. De Villers, B.T. Tolbert, S.H. Schwartz, B.J. Rubin, Y. “*Improved Photovoltaic Efficiency in Semiconducting Polymer/Fullerene Solar Cells through Control of Fullerene Self-Assembly and Stacking*”, provisional patent application Case Number 2008-662,
2. Ayzner, AL; **Tassone, CJ**; Tolbert, SH; Schwartz, BJ; . “*Reappraising the Need for Bulk Heterojunctions in Polymer-Fullerene Photovoltaics: The Role of Carrier Transport in All-Solution-Processed p3HT/PCBM Bilayer Solar Cells.”* *J. Phys Chem C.* 113 (46). **2009**
3. Schmidt, K; **Tassone, CJ;** Niskala, JR; Yiu, AT; Lee, OP; Weiss, TM; Weng, C; Freschet, JM; Beaujuge, PM; Toney, MF. “*A Mechanistic Understanding of Processing Additive-Induced Efficiency Enhancement in Bulk Heterojunction Organic Solar Cells”,* *Adv. Mater.* 26 (2). **2013**.
4. Unger, RL; Bowring, AR; **Tassone, CJ**; Pool, VL, Gold-Parker, A; Cheacharoen, R; Stone, KH; Hoke, ET; Toney, MF; McGehee, MD. “Chlorine in Lead Chlorine-Derived Organo-Metal Halides for Perovskite-Absorber Solar Cells”, *Chem. Mater.* 26 (24). **2014**.
5. Bartelt, JA; Douglas, JD; Mateker, WR; El Labban, A; **Tassone, CJ**; Toney, MF; Freschet, JMJ; Beaujuge, PM; McGehee, MD. “Controlling Solution-Phase Polymer Aggregation with Molecular Weight and Solvent Additives to Optimize Polymer-Fullerne Bulk Heterojunction Solar Cells”, *Adv. Ener. Mater.* 4 (9). **2014**
6. Rivnay, J; Inal, S; Collins, BA; Sessolo, M; Stavrinidou, E; Strakosas, X; **Tassone, CJ**; Delongchamp. DM; Malliaras, GG. “Structural Control Of Mixed Ionic and Electronic Transport in Conducting Polymers”, *Nature Comm*. 7. **2016**
7. Reinspach, JA; Diao, Y; Giri, G; Sachse, T; England, K; Zhuo, Y; **Tassone, CJ**; Worfolk, BJ; Pressselt, M; Toney, MF; Mannsfeld, S; Bao, Z. “Tuning the morphology of solution-sheared P3HT:PCBM films” *ACS Appl. Mater. Inter.* 8(3). **2016**.
8. Santra, PK; Palmstrom, AF; Tassone, CJ; Bent, SF. “Molecular ligands control superlattice structure and crystallite orientation in colloidal quantum dot solids” *Chem Mater*. 120(43). **2016**.
9. Gu, X; Zhou, Y; Gu, K; Kurosawa, T; Guo, Y; Li, Y; Schroeder, BC; Yan, H; Molina-Lopez, F; **Tassone, CJ**; Wang, C; Mannsfeld, SCB; yan, He; Zhao, D; Toney, MF; Bao, Z. “Roll-to-Roll Printed Large-Area All-Polymer Solar Cells with 5% Efficiency Based on a Low Crystallinity Conjugated Polymer Blend” *ACS Ener. Mater.* 7(14). **2016**.
10. Wu, L; Willis, JJ; McKay, IS; Diroll, BT; Qin, J; Cargnello, M; **Tassone, CJ**. “High-Temperature Crystallization of Nanocrystals into Three-dimensional Superlattices” *Nature*. 548 (7666). **2017**.

**Synergistic Activities:**

1. Lead PI for Advanced Manufacturing Office funded FWP “Aggregation and Structuring of Materials and Chemicals Data from Diverse Sources”. The project focus is on building foundational machinery to enable machine learning approaches to data analysis and experimental planning. The project fully automated the discovery of syntheses for nanocrystals in a closed loop as guided by a custom developed artificial intelligence.
2. co-PI for Advanced Manufacturing Office funded FWP, “In-Situ Data Acquisition and Tool Development for Additive Manufacturing Metal Powder Systems” Within the project my focus has been on developing the in-situ x-ray characterization approaches for additive manufacturing and data analysis pipelines to enable real time data handling to provide actionable information to researchers.
3. Lead-PI LDRD “Accelerating Nanocrystal Synthetic Development”. Focus of the project was on using in-situ simultaneous SAXS/WAXS during the development of nanocrystal syntheses in order to deepen mechanistic understanding and accelerate the synthesis of size controlled libraries of nanocrystals to elucidate structure-property relationships for catalysis.
4. Group leader for the SSRL materials science hard x-ray group. Managing operations for a suite of six beamlines at SSRL.

**Identification of Potential Conflicts of Interest or Bias in Selection of Reviewer:**

**Collaborators and Co-editors:** Frank Abild-Pedersen (Stanford), Hassan Ajiama (Stanford), PaulBalding (Molecular Vista), Zhenan Bao (Stanford), Stacey Bent (Stanford), David Bergsman (Stanford), Joseph Berry (NREL), Karsten Bruening (Stanford), Matteo Cargnello (Stanford), Jeffrey Christian (NREL), Ping-Hsun Chu (Georgia Tech), Bruce Clemens (Stanford), Richard Closser (Stanford), Brian Collins (University of Washington), Reinhold Dauskardt (Stanford), DeanDelongchamp (NIST), Ying Diao (University of Illinois Urbana-Champagne), David Diercks (Colorado School of Mines), Benjamin Diroll (Argonne), Kemar England (Stanford), Alessandro Gallo (SLAC), Gaurav Giri (University of Virginia), Emmet Goodman (Stanford), Andrew Gorman (Georgia Tech), Brian Gorman (Colorado School of Mines), Kevin Gu (Stanford), Xiaodan Gu (University of Southern Mississippi), Yikun Guo (Peking University), Michael Hawkridge (PANanlytical), Jeff Hernandez (Georgia Tech), Alexander Hexemer (ALS), Florian Hilt (Stanford), Michael Hovish (Stanford), Sahika Inal (KAUST), Thomas Jaramillo (Stanford), Lydia-Marie Joubert (Stanford), Tadanori Kurosawa (Stanford), Brian Larsen (Colorado School of Mines), Jeremy Leong (Bioenergy Technologies Office, DOE), Yunke Li (Hong Kong University of Science and Technology), Haoran Lin (Stanford), Joseph Luther (NREL), George Maliaras (University of Cambridge), StefanMannsfeld (Dresden University of Technology), Katherine Margulis (Stanford), Pedro Martins (Stanford), Ian McKay (Stanford), Francisco Molina-Lopez (Stanford), Ioan Negulescu (Louisiana State University Baton Rouge), Kenneth Neyerlin (Colorado School of Mines), Dennis Nordlund (SLAC), Stanislaw Nowak (SLAC), Stefan Oosterhout (Utrect University), Axel Palmstrom (Stanford), Jung Park (Georgia Tech), Katherine Park (Molecular Vista), Bryan Pivovar (Colorado School of Mines), Martin Presselt (Friedrich-Schiller-University Jena), Adam Printz (Stanford), Svitlana Pylypenko (Colorado School of Mines), Jian Qin (Stanford), Elsa Reichmanis (Georgia Tech), Julia Reinspach (Amazon), Ryan Richards (Colorado School of Mines), Andrew Riscoe (Stanford), Jonathan Rivnay (Northwestern University), Nicholas Rolston (Stanford), Cornelia Rosu (Georgia Tech), Anirban Roy (Anasys Instruments), Yves Rubin (UCLA), Paul Russo (Georgia Tech), Torsten Sachse (Friedrich-Schiller-University Jena), Pralay Santra (Center for Nano and Soft Materials Sciences, Bangalore), Victoria Savikhin (Stanford), Laura Schelhas (SLAC), Bob Schroeder (Stanford), Benjamin Schwartz (UCLA), Michele Sessolo (University of Valencia), Jamie Sethian (UC Berkeley), Dimosthenis Sokaras (SLAC), Mohan Srinivasarao (Georgia Tech), Eleni Stavrindou (Linkoping University), Xenofon Strakosas (UCSC), Matthew Strand (Colorado School of Mines), Sarah Tolbert (UCLA), Michael Toney (SLAC), Cheng Wang (ALS), Robert Waymouth (Stanford), Joshua Willis (Stanford), Brian Worfolk (Phillips 66), Liheng Wu (Stanford), He Yan (Hong Kong University of Science and Technology), Hongping Yan (Stanford), Richard Zare (Stanford), Xiangyi Zhang (Stanford), Dahui Zhoa (Peking University), Yan Zhou (Stanford)

**Graduate and Postdoctoral Advisors and Advisees:** Advisors: Sarah Tolbert (UCLA), Benjamin Schwartz (UCLA), Michael McGehee (Stanford), Michael F. Toney (SLAC).

Advisees: Karsten Bruening (Konica-Minolta), Lenson Pellouchoud (Independent Contract Developer), Amanda Fournier (Unemployed), Anthony Fong (SLAC), and Liheng Wu (SABIC).

**Advisory Committees**

None

**Daniel Ratner**

**Education and Training**

Ph.D. Applied Physics, Stanford University, 2011

B.S., Physics, Harvard University, 2002

**Research and Professional Experience:**

Staff scientist , SLAC, 2013 - Present

Description: Head of SLAC machine learning strategic initiative

Associate staff scientist, SLAC, 2011 - 2013

Description: LCLS laser group

**Awards:**

Outstanding Doctoral Thesis Research in Beam Physics, American Physical Society

FEL Young Scientist Award

**Conference activity, Refereeing, and Talks:**

Founder and member of international organizing committee, ICFA workshop on machine learning for accelerators

Scientific program committee, International FEL conference

Steering committee for DOE AI XLab

More than 35 invited and plenary talks

Guest editor and Referee for more than 30 publications

**Publications:**

Bayesian optimization of a free-electron laser

J. Duris, D. Kennedy, A. Hanuka, J. Shtalenkova, A. Edelen, P. Baxevanis, A. Egger, T. Cope, M. McIntire, S. Ermon, and D. Ratner, Phys. Rev. Lett., accepted (2020)

Laguerre-Gaussian mode laser heater for microbunching instability suppression in free-electron lasers

J. Tang et al., Phys. Rev. Lett., accepted (2020)

What are the advantages of ghost imaging? Multiplexing for x-ray and electron imaging

T.J. Lane and D. Ratner, Optics Express, 28, 5898 (2020)

Mapping photocathode quantum efficiency with ghost imaging

K. Kabra, S. Li, F. Cropp, T.J. Lane, P. Musumeci, and D. Ratner, Phys. Rev. AB, 22, 022803 (2020)

J. Betterton, S. Webb, M. Kochenderfer, D. Ratner, Adaptive Illumination at X-ray Sources with Reinforcement Learning, ICRA (2020).

Attosecond transient absorption spooktroscopy: a ghost imaging approach to ultrafast absorption spectroscopy

T. Driver et al., PCCP, 5 (2020)

Experimental observations of seed growth and accompanying pedestal contamination in a self-seeded, soft x-ray free-electron laser

G. Marcus et al., Phys. Rev. AB, 22, 080702 (2019)

D. Ratner, J. Cryan, T.J. Lane, S. Li, G. Stupakov, Pump-probe ghost imaging with SASE FELs, Phys. Rev. X, 9 011045 (2019).

A. Edelen, D. Bowring, J. Snuverink, G. Valentino, I. Bazarov, R. Ischebeck, I. Agapov, A. Adelmann, D. Ratner, C. Mayes, J. Edelen, J. Wenninger, R. Kammering, Opportunities in Machine Learning for Particle Accelerators, arXiv:1811.03172 (2018).

S. Li, F. Cropp, K. Kabra, T.J. Lane, G. Wetzstein, P. Musumeci, D. Ratner, Electron ghost imaging, Phys. Rev. Lett., 121, 114801 (2018).

M. McIntire, D. Ratner, S. Ermon, Weighted KL divergence for adaptive GP set selection, UAI16 (2016).

M. McIntire, T. Cope, S. Ermon, D. Ratner, Bayesian optimization of FEL performance at LCLS, UAI16 (2016).

**Synergistic Activities:**

Head of SLAC ML Initiative: Leading a lab-wide strategic initiative to coordinate machine learning activities at SLAC.

Founder of ICFA Workshop on ML for Accelerators: Co-leader of international ICFA workshop series on machine learning for particle accelerators.

**Collaborators and Co-editors:**

Ilya Agapov (DESY), Andre Al-Haddad (PSI), Sean Alverson (SLAC), Karl Bane (SLAC), Panos Baxevanis (SLAC), Uwe Bergmann (SLAC), Dorian Bohler (SLAC), James Budarz (Brown), Sergio Carbajo (SLAC), Alex Chao (SLAC), Florian Christie (DESY), Ryan Coffee (SLAC), William Colocho (SLAC), Tyler Cope (SLAC), Franz-Josef Decker (SLAC), Yuantao Ding (SLAC), Sebastian Doniach (Stanford), Taran Driver (SLAC), Joseph Duris (SLAC), Auralee Edelen (SLAC), Steve Edstrom (SLAC), Adam Egger (SLAC), Paul Emma (SLAC), Stefano Ermon (Stanford), William Fawley (SLAC), Yiping Feng (SLAC), Jo Frisch (SLAC), Alan Fry (SLAC), Franklin Fuller (SLAC), Siegfried Glenzer (SLAC), James Glownia (SLAC), Eduardo Granados (CERN), Marc Guetg (SLAC), Adi Hanuka (SLAC), Erik Hemsing (SLAC), Zhirong Huang (SLAC), Krish Kabra (UCLA), Dylan Kennedy (SLAC), Kwang-Je Kim (ANL), Adam Kirrander (Edinburg), Thomas Kroll (SLAC), Jacek Krzywinski (SLAC), TJ Lane (SLAC), Siqi Li (SLAC), Alberto Lutman (SLAC), Gabriel Marcus (SLAC), Agostino Marinelli (SLAC), Tim Maxwell (SLAC), Christopher Mayes (SLAC), Mitchell McIntire (Google), Derek Mendez (LBNL), Despina Milathianaki (SLAC), Michael Minitti (SLAC), Pietro Musumeci (UCLA), Tor Raubenheimer (SLAC), Jane Shtalenkova (SLAC), Gennady Stupakov (SLAC), Jingyi Tang (SLAC), Franz Tavella (SLAC), Peter Weber (Brown), Gordon Wetzstein (Stanford), Feng Zhou (SLAC)

**Graduate Advisors:** Alex Chao (SLAC), John Galayda (SLAC)

**Kevin Stone**

**Education and Training**

|  |  |  |
| --- | --- | --- |
| University of California, Berkeley | Physics and Astronomy | B.A. 2004 |
| Stony Brook University | Physics and Astronomy | Ph.D. 2009 |
| Lawrence Berkeley National Laboratory | Material Science Division | Post-Doc 2010-2013 |
| SLAC National Laboratory | SSRL | Post-Doc 2013-2014 |

**Research and Professional Experience**

|  |  |  |
| --- | --- | --- |
| Staff Scientist, SSRL | SLAC National Laboratory | 2016-Present |
| Associate Staff Scientist, SSRL | SLAC National Laboratory | 2014-2016 |

**Selected Publications**

1. V. Thampy, A.Y. Fong, N.P. Calta, J. Wang, A.A. Martin, P.J. Depond, A.M. Kiss, G. Guss, Q. Xing, R.T. Ott, A. van Buuren, M.F. Toney, J. Nelson Weker, M.J. Kramer, M.J. Matthews, C.J. Tassone, and **K.H. Stone**, “Subsurface Cooling Rates and Microstructural Response during Laser Based Metal Additive Manufacturing”, *Sci. Rep.*, vol. 10, no. 1, pp. 1-9, 2020. https://doi.org /10.1038/s41598-020-58598-z
2. L.T. Schelhas, Z. Li, J.A. Christians, A. Goyal, P. Kairys, S.P. Harvey, D.H. Kim, **K.H. Stone**, J.M. Luther, K. Zhu, V. Stevanovic, and J.J. Berry, “Insights into operational stability and processing of halide perovskite active layers”, *Energy Environ. Sci.*, vol. 12, pp. 1341-1348, 2019. https://doi.org /10.1039/C8EE03051K
3. **K.H. Stone**, A. Gold-Parker, V.L. Pool, E.L. Unger, A.R. Bowring, M.D. McGehee, M.F. Toney, and C.J. Tassone, “Transformation from crystalline precursor to perovskite in PbCl2-derived MAPbI3”, *Nat. Commun.*, vol. 9, p. 3458, 2018. <https://doi.org/10.1038/s41467-018-05937-4>
4. B.-R. Chen, W. Sun, D.A. Kitchaev, J.S. Mangum, V. Thampy, L.M. Garten, D.S. Ginley, B.P. Gorman, **K.H. Stone**, G. Ceder, M.F. Toney, and L.T. Schelhas, “Understanding crystallization pathways leading to manganese oxide polymorph formation”, *Nat. Commun.*, vol. 9, p. 2553, 2018. https://doi.org /10.1038/s41467-018-04917-y
5. W.E. Gent, K. Lim, Y. Liang, Q. Li, T. Barnes, S. Ahn, **K.H. Stone**, M. McIntire, J. Hong, J.H. Song, Y. Li, A. Mehta, S. Ermon, T. Tyliszczak, D. Kilcoyne, D. Vine, J. Park, S. Doo, M.F. Toney, W. Yang, D. Prendergast, and W.C. Chueh, “Coupling between oxygen redox and cation migration explains unusual electrochemistry in lithium-rich layered oxides,” *Nat. Commun.*, vol. 8, no. 1, p. 2091, Dec. 2017. <https://doi.org/10.1038/s41467-017-02041-x>
6. L.T. Schelhas, **K.H. Stone**, S.P. Harvey, D. Zakhidov, A. Salleo, G. Teeter, I.L. Repins, and M.F. Toney, “Point defects in Cu2ZnSnSe4 (CZTSe): Resonant X-ray diffraction study of the low-temperature order/disorder transition,” *Phys. status solidi*, vol. 254, no. 9, p. 1700156, Sep. 2017. <https://doi.org/10.1002/pssb.201700156>
7. C. Cao, H.-G. Steinrück, B. Shyam, **K.H. Stone**, and M.F. Toney, “In situ study of silicon electrode lithiation with X-ray reflectivity”, *Nano Lett.*, vol. 16, pp. 7394-7401, 2016. https://doi.org /10.1021/acs.nanolett.6b02926
8. L.T. Schelhas, J.A. Christians, J.J. Berry, M.F. Toney, C.J. Tassone, J.M. Luther, and **K.H. Stone**, “Monitoring a Silent Phase Transition in CH3NH3PbI3 Solar Cells via Operando X-ray Diffraction”, *ACS Energy Lett.*, vol. 1, no. 5, p. 1007, Oct. 2016. <https://doi.org/10.1021/acsenergylett.6b00441>
9. **K.H. Stone** and J.B. Kortright, “Molecular anisotropy effects in carbon K-edge scattering: Depolarized diffuse scattering and optical anisotropy,” *Phys. Rev. B*, vol. 90, no. 10, p. 104201, Sep. 2014. <https://doi.org/10.1103/PhysRevB.90.104201>
10. **K.H. Stone**, S.M. Valvidares, and J.B. Kortright, “Kramers-Kronig constrained modeling of soft x-ray reflectivity spectra: Obtaining depth resolution of electronic and chemical structure,” *Phys. Rev. B*, vol. 86, no. 2, p. 024102, Jul. 2012. <https://doi.org/10.1103/PhysRevB.86.024102>

**Synergistic Activities**

Panelist – Proposal Review Panel for “Structural Science” at the Advanced Photon Source, Argonne National Lab, 2019-present

Panelist – Proposal Review Panel for “Scattering-Applied Materials” at the Advanced Photon Source, Argonne National Lab, 2017-2019

Organizer – Workshop on Additive Manufacturing, 2017 SSRL/LCLS User Meeting

**Identification of Potential Conflicts of Interest or Bias in Selection of Reviewers**

***Graduate and Post-Doctoral Advisors:*** Peter W. Stephens (Stony Brook University), Jeffrey B. Kortright (LBNL), Michael F. Toney (SLAC), William Chueh (Stanford/SLAC)

***Post-Doctoral Advisee:*** Vivek Thampy (SLAC)

***Collaborators and Co-editors:***

Elke Arenholz (Cornell University) , Finn Babbe (University of Luxembourg) , Ji Hyun Baek (Stanford) , Riccardo Bassiri (Stanford) , Stacey F. Bent (Stanford) , Dominik M. Berg (Delaware State University) , Joseph J. Berry (NREL) , Andrea R. Bowring (Stanford) , Tonio Buonassisi (MIT) , Nicholas P. Calta (LLNL) , Chuntian Cao (SLAC) , Gerbrand Ceder (UC Berkeley) , Bor-Rong Chen (SLAC) , Ching-Chin Chen (Delaware State University) , In Sun Cho (Ajou University) , Steven T. Christensen (NREL) , Jeffrey A. Christians (NREL) , William C. Chueh (Stanford) , Pete Collins (Ames Lab) , Jeff Corbett (SLAC) , Peter M. Csernica (Stanford) , Hongjie Dai (Stanford) , Philip J. Depond (LLNL) , Kevin D. Dobson (University of Delaware) , Farid El Gabaly (Sandia National Lab) , Stefano Ermon (Stanford) , Martin M. Fejer (Stanford) , Ian R. Fisher (Stanford) , Anthony Y. Fong (SLAC) , Guss Gabe (LLNL) , Alex Garachtchenko (SLAC) , Lauren M. Garten (NREL) , William E. Gent (Stanford) , David S. Ginley (NREL) , Aryeh Gold-Parker (Stanford) , Brian P. Gorman (Colorado School of Mines) , Zixuan Guan (Stanford) , Jingha Guo (LBNL) , James E. S. Haggerty (Oregon State University) , Steven P. Harvey (NREL) , Aaron Holder (NREL) , Jihyun Hong (Korea Institute of Science and Technology) , Robert L. Z. Hoye (MIT) , Bryan D. Huey (University of Connecticut) , Po-Yu Hwang (Delaware State University) , Bart Johnson (SLAC) , Hyun Suk Jung (Sungkyunkwan University) , Dong Hoe Kim (NREL) , Andrew M. Kiss (Brookhaven) , Daniil A. Kitchaev (MIT) , Ai Leen Koh (Stanford) , Yuriy Kolotovsky (SLAC) , Matthew J. Kramer (Ames Lab) , Cheng-Yu Lai (Delaware State University) , Stephan Lany (NREL) , Saul H. Lapidus (Argonne National Lab) , Jae Myeong Lee (Sungkyunkwan University) , Ben Levy-Wendt (Stanford) , Chunlei Li (East China University of Science and Technology) , Kipil Lim (Stanford) , Yijin Liu (SLAC) , Mimi Liu (Delaware State University) , Xinyan Liu (Stanford) , Joseph M. Luther (NREL) , John S. Mangum (Colorado School of Mines) , Jamie L. Manson (Eastern Washington University) , Aiden A. Martin (LLNL) , Manyalibo J. Matthews (LLNL) , Michael D. McGehee (University of Colorado Boulder) , Apurva Mehta (SLAC) , Celeste Melamed (Colorado School of Mines) , Brent C. Melot (University of Southern California) , Matteo Monti (TVP Solar SA Geneva) , Paul F. Ndione (NREL) , Johanna Nelson Weker (SLAC) , Dennis Nordlund (SLAC) , Brendan Ortiz (Colorado School of Mines) , Ryan T. Ott (Ames Lab) , Philip A. Parilla (NREL) , Sangwook Park (Stanford) , Donata Passarello (SLAC) , John D. Perkins (NREL) , Vanessa L. Pool (SLAC) , David Prendergast (LBNL) , Xing Qingfeng (Ames Lab) , Daniela R. Radu (Florida International University) , Ingrid Repins (NREL) , Alberto Salleo (Stanford) , David O. Scanlon (University College London) , Laura Schelhas (SLAC) , Rekha Schnepf (Colorado School of Mines) , Philip Schulz (NREL) , Padraic Shafer (LBNL) , Yezhou Shi (McKinsey & Co Shanghai) , Badri Shyam (University of Dayton Research Institute) , Joseph A. Singh (Stanford) , Sebastian Siol (NREL) , Ian C. Smith (Stanford) , James Steffes (University of Connecticut) , Hans-Georg Steinrueck (SLAC) , Peter W. Stephens (Stony Brook University) , Vladan Stevanovic (Colorado School of Mines) , Wenhao Sun (University of Michigan) , Yuri Suzuki (Stanford) , Christopher J. Takacs (SLAC) , Adele Tamboli (NREL) , Christopher J. Tassone (SLAC) , Janet Tate (Oregon State University) , Glenn Teeter (NREL) , Brooks Tellekamp (NREL) , Vivek Thampy (SLAC) , Eric Toberer (Colorado School of Mines) , Michael F. Toney (SLAC) , Charlie Tsai (Stanford) , William Tumas (NREL) , Eva L. Unger (Lund University) , Arturas Vailionis (Stanford) , Anthony van Buuren (LLNL) , Douglas Van Campen (SLAC) , Jenny Wang (LLNL) , Yahong Xu (Donghua University) , Wanli Yang (LBNL) , Nuoya Yang (Stanford) , Di Yi (Stanford) , Dante Zakhidov (Stanford) , Andriy Zakutayev (NREL) , Liang Zhang (LBNL) , Wenjun Zhang (University of Saskatchewan) , Xiaolin Zheng (Stanford) , Kai Zhu (NREL)

***Samuel M. Webb***

Staff Scientist

Chemistry and Catalysis Division

Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory

Menlo Park, CA 94025

Phone: 650-926-3734. Email: [samwebb@slac.stanford.edu](mailto:samwebb@slac.stanford.edu)

**A. Education**

California Institute of Technology Pasadena, CA Geology B.S.*,* 1996

California Institute of Technology Pasadena, CA Environ. Engin. Sciences M.S.,1996

Northwestern University Evanston, IL Civil and Environ. Engin. Ph.D, 2001

SLAC National Accelerator Lab Menlo Park, CA Postdoctoral Fellow 2001

**B. Professional Experience**

Staff Scientist, Stanford Synchrotron Radiation Lightsource, SLAC 2008-present

Beam Line Scientist, Stanford Synchrotron Radiation Lightsource, SLAC 2004-2008

**C. Publications**

Zeyen N, Benzerara K, Menguy N, Brest J, Templeton AS, **Webb, SM**, Gérard E, Moreira D, López-García P, Tavera R. (2019). Fe-bearing phases in modern lacustrine microbialites from mexico. *Geochimica et Cosmochimica Acta* **253**, 201-230.

Stetten L, Blanchart P, Mangeret A, Lefebvre P, Le Pape P, Brest J, Merrot P, Julien A, Proux O, **Webb SM** (2018). Redox fluctuations and organic complexation govern uranium redistribution from u (iv)-phosphate minerals in a mining-polluted wetland soil, Brittany, France. *Environmental science & technology* **52**, 13099-13109.

Rose CV, **Webb SM**, Newville M, Lanzirotti A, Richardson JA, Tosca NJ, Catalano JG, Bradley AS, Fike DA (2019). Insights into past ocean proxies from micron-scale mapping of sulfur species in carbonates. *Geology*

Marnocha CL, Sabanayagam CR, Modla S, Powell DH, Henri PA, Steele A, Hanson TE, **Webb SM**, Chan CS. (2019) Insights into the mineralogy and surface chemistry of extracellular biogenic S (0) globules produced by Chlorobaculum tepidum. *Frontiers in Microbiology* **10**, 271.

Raven MR, Fike DA, Gomes ML, **Webb SM** (2019). Chemical and isotopic evidence for organic matter sulfurization in redox gradients around mangrove roots. *Frontiers in Earth Science* **7**, 98.

Edwards NP, **Webb SM**, Krest CM, van Campen D, Manning PL, Wogelius RA, Bergmann U (2018) A new synchrotron rapid-scanning X-ray fluorescence (SRS-XRF) imaging station at SSRL beamline 6-2. *J. Synchrotron Rad.* **25** 1565-1573 DOI: 10.1107/S1600577518010202

Raven MR, Fike DA, Gomes ML, **Webb SM**, Bradley AS, McClelland HLO (2018) Organic carbon burial during OAE2 driven by changes in the locus of organic matter sulfurization. *Nature Communications* **9** DOI: 10.1038/s41467-018-05943-6

McClain CN, Fendorf S, **Webb SM**, Maher, K (2017) Quantifying Cr(VI) Production and Export from Serpentine Soil of the California Coast Range. *Environ. Sci. Technol.* **51** 141-149. DOI: 10.1021/acs.est.6b03484

Kraal P, Burton ED, Rose AL, Kocar BD, Lockhart RS, Grice K, Bush RT, Tan E, **Webb SM** (2015) Sedimentary iron-phosphorus cycling under contrasting redox conditions in a eutrophic estuary. *Chem. Geol.* **392** p. 19-31.

Ying SC, Masue-Slowey Y, Kocar BD, Griffis SD, **Webb SM**, Marcus MA, Francis CA, Fendorf S. (2013) Distributed microbially- and chemically-mediated redox processes controlling arsenic dynamics within Mn-/Fe-oxide constructed aggregates. *Geochim. Cosmochim. Acta* **104** 29-41.

**D. Synergetic Activities**

Half of my professional time is allocated to educating and assisting visiting scientists with synchrotron techniques and developing new techniques for use in cultural heritage, environmental and geological sciences, and medicine at synchrotron sources. This paradigm insures a continuous flow of ideas from research projects into a large user community.

1. User Support: I provide hands-on assistance to visiting scientists (mostly post docs and graduate students) regarding experimental techniques at SSRL beam lines.

2. Education and Outreach: Organized and participated in 15 workshops and school symposia since 2000. Presented lectures to the general public on research.

3. Development of Data Collection Tools: I have helped to establish and construct the microprobe efforts at beam lines BL2-3, BL 6-2, BL10-2 and BL14-3. Part of these efforts has included developing and writing the software used to control the beam line hardware and collect microprobe fluorescence, micro-XAS, and micro-XRD data. These programs are used lab-wide at SSRL.

4. Development of Analysis Tools: I have developed software used in the analysis of EXAFS, XRD, and microXAS (microprobe fluorescence and spectroscopy) data. The programs are available freely on the web ([www.sams-xrays.com](http://www.sams-xrays.com)) and are used to analyze data from many major synchrotrons world-wide.

5. Development of Searchable, Archive Data Tools: I am currently collaborating with researchers at Dartmouth College to form a database of XRF data that satisfies the FAIR (Findable, Accessible, Interoperable, Reusable) data requirements. These will satisfy requirements for web-based browsing of integrated genomics data with X-ray data and will need to collate data sets from a wide variety of different synchrotron sources.

**Collaborators**

Jess Adkins, Caltech

Yuji Arai, U of Illinois

Loic Bertrand, Synchrotron SOLEIL

Alex Bradley, Washington University St. Louis

Lucile Brunell, C2MRF, FR

Jeff Catalano, Washington University St. Louis

Clara Chan, University of Delaware

Greg Dick, University of Michigan

James Evans, Utah State

Gabriella Farfan, Smithsonian Institute

Scott Fendorf, Stanford

David Fike, Washington University St. Louis

Woodward Fischer, Caltech

Andrea Foster, USGS

Amy Gartman, USGS

Graham George, U Saskatchewan

Simon George, LBL

Colleen Hansel, Wood Hole Oceangraphic Institute

Sunita Ho, UCSF

Dean Hesterberg, North Carolina State University

Jena Johnson, University of Michigan

Andreas Kappler, University of Tubingen

Katrien Keune, Rijksmuseum, NL

Chris Kim, Chapman University

Ann Lam, Green Neuroscience Institute

Christina Lopano, NETL

George Luther, University of Delaware

Maria McNamara, University of Cork, Ireland

John Magyar, Caltech

Phillip Manning, University of Manchester

Lisa Mayhew, U of Colorado

Apurva Mehta, SLAC National Accelerator Laboratory

Ron Oremland, USGS

Victoria Orphan, Caltech

Ingrid Pickering, U Saskatchewan

Matthew Polizzotto, University of Oregon

Bogdan Popescu, U Saskatchewan

Emeline Pouyet, Northwestern University

Tracy Punshon, Dartmouth

Morgan Raven, University of California, Santa Barbara

Catherine Rose, St. Andrews University, Scotland

George Rossman, Caltech

Cara Santelli, University of Minnesota

Phillipe Sarrazin, NASA Ames

Phillipe Sciau, CNRS-Orsay

Angie Seyfferth, University of Delaware

Josh Sharp, Colorado School of Mines

Donald Sparks, U of Delaware

Stephen Sturzenbaum, Kings College, London

Elizabeth Swanner, Iowa State University

Martial Taillefert, Georgia Tech

Brad Tebo, Oregon Health Science University

Alexis Templeton, U of Colorado

Nick Tosca, Oxford University, UK

Marc Walton, Northwestern University

Anna Wegorzewski, Bundesanstalt fur Geowissenschaften und Rohstoffe

Roy Wogelius, University of Manchester

Samantha Ying, University of California, Riverside

Michael Zeineh, Stanford School of Medicine

Teng Zeng, Syracuse University

**Advisors:**

Jean-Francois Gaillard, Northwestern Univsersity

John Bargar, SLAC National Accelerator Laboratory

**Advisees:**

Ben Kocar, MIT

Courtney Krest, Roach & Associates, LLC

Sharon Bone, SLAC National Accelerator Laboratory

Nick Edwards, SLAC National Accelerator Laboratory

ANL & ORNL BIOS

# Appendix 2: Current and Pending Support

**Chris Tassone**

|  |  |
| --- | --- |
| **Investigator: Chris Tassone (McIntyre)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Current** | |
| Project/Proposal Title and grant number, if appropriate: **SSRL User Facility Operations**  **(SCW0010)** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC** | |
| Annual Award Amount**: $N/A** Total Award Period Covered: **10/01/2019-09/30/2020** | |
| Annual Award Amount to PI’s Research: **$N/A** | |
| Person-Months Per Year Committed to Project: **12.0** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application.** | |
| **Investigator: Chris Tassone (Mehta)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Pending** | |
| Project/Proposal Title and grant number, if appropriate: **Collaborative Machine Learning Platform for Scientific Discovery** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC/Stanford** | |
| Annual Award Amount: **$555K** Total Award Period Covered: **9/01/2020-08/31/2023** | |
| Annual Award Amount to PI’s Research: **$555K** | |
| Person-Months Per Year Committed to Project: **0.6** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. Synergies include data handling capabilities which are applicable to the proposed project scope of work.** | |
| **Investigator: Chris Tassone** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Pending** | |
| Project/Proposal Title and grant number, if appropriate: **Autonomous Control of Experimental Stations at BES Lightsources** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC/Stanford** | |
| Annual Award Amount: **$4.5M** Total Award Period Covered: **9/01/2020-08/31/2023** | |
| Annual Award Amount to PI’s Research: **$4.5M** | |
| Person-Months Per Year Committed to Project: **0.6** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. Synergies include data handling capabilities which are applicable to the proposed project scope of work.** | |
| **Investigator: Chris Tassone (Bare)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Pending** | |
| Project/Proposal Title and grant number, if appropriate: **Center for Integrated Upcycling of Mixed Polymers (CUPs)** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC/Stanford** | |
| Annual Award Amount: **$920K** Total Award Period Covered: **8/01/2020-08/31/2024** | |
| Annual Award Amount to PI’s Research: **$460K** | |
| Person-Months Per Year Committed to Project: **0.6** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. Synergies include data handling capabilities which are applicable to the proposed project scope of work.** | |

**Daniel Ratner**

|  |  |
| --- | --- |
| **Investigator: Daniel Ratner** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Current** | |
| Project/Proposal Title and grant number, if appropriate: **SLAC Machine Learning Strategic Growth** | |
| Source of Support: **SLAC Indirect** Location of Project: **SLAC** | |
| Annual Award Amount**: $N/A** Total Award Period Covered: **10/01/2019-09/30/2020** | |
| Annual Award Amount to PI’s Research: **$N/A** | |
| Person-Months Per Year Committed to Project: **12.0** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. Synergies include role in connecting machine learning efforts across the lab to leverage relevant expertise.** | |
| **Investigator: Daniel Ratner (Tassone)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Pending** | |
| Project/Proposal Title and grant number, if appropriate: **Autonomous Control of Experimental Stations at BES Lightsources** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC/Stanford** | |
| Annual Award Amount: **$4.5M** Total Award Period Covered: **9/01/2020-08/31/2023** | |
| Annual Award Amount to PI’s Research: **$4.5M** | |
| Person-Months Per Year Committed to Project: **0.6** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. Synergies include data handling capabilities which are applicable to the proposed project scope of work.** | |
| **Investigator: Daniel Ratner (Sakdinawat)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Pending** | |
| Project/Proposal Title and grant number, if appropriate: **X-ray Diffractive Optics for Accelerator Science** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC/Stanford** | |
| Annual Award Amount: **$900K** Total Award Period Covered: **8/01/2020-07/31/2022** | |
| Annual Award Amount to PI’s Research: **$50K** | |
| Person-Months Per Year Committed to Project: **0.12** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. Synergies include the general study of machine learning applications at light sources.** | |
| **Investigator: Daniel Ratner (Huang)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Pending** | |
| Project/Proposal Title and grant number, if appropriate: **Model-based advanced tuning and control for scientific user facilities.** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC/Stanford** | |
| Annual Award Amount: **$4.5M** Total Award Period Covered: **9/01/2020-08/31/2023** | |
| Annual Award Amount to PI’s Research: **$4.5M** | |
| Person-Months Per Year Committed to Project: **.24** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap.  Synergies include the application of machine learning optimization strategies at the BES scientific user facilities.** | |

**Kevin Stone**

|  |  |
| --- | --- |
| **Investigator: Kevin Stone (McIntyre)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Current** | |
| Project/Proposal Title and grant number, if appropriate: **SSRL User Facility Operations**  **(SCW0010)** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC** | |
| Annual Award Amount: **$N/A** Total Award Period Covered: **10/01/2017-09/30/2020** | |
| Annual Award Amount to PI’s Research: **$N/A** | |
| Person-Months Per Year Committed to Project: **12.00** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. Synergies include characterization of materials with hard X-ray scattering.** | |
| **Investigator: Kevin Stone (Mehta)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Pending** | |
| Project/Proposal Title and grant number, if appropriate: **Closing the Measurement-Discovery Gap at DOE Lightsources** | |
| Source of Support: **DOE -BES**  Location of Project: **SLAC/Stanford** | |
| Annual Award Amount: **$4.5M**  Total Award Period Covered: **9/01/2020-08/31/2023** | |
| Annual Award Amount to PI’s Research: **$4.5M** | |
| Person-Months Per Year Committed to Project: **.6** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. Synergies include data handling capabilities which are applicable to the proposed project scope of work.** | |
| **Investigator: Kevin Stone** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Pending** | |
| Project/Proposal Title and grant number, if appropriate: **Advanced Diffraction Approaches for Beyond-Ideal Materials** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC/Stanford** | |
| Annual Award Amount: **$2.5M** Total Award Period Covered: **9/01/2020-08/31/2025** | |
| Annual Award Amount to PI’s Research: **$2.5M** | |
| Person-Months Per Year Committed to Project: **6.0** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. Synergies include data handling capabilities which are applicable to the proposed project scope of work and general characterization of materials with hard X-ray scattering.** | |
| **Investigator: Kevin Stone (Tassone)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Pending** | |
| Project/Proposal Title and grant number, if appropriate: **Autonomous Control of Experimental Stations at BES Lightsources** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC/Stanford** | |
| Annual Award Amount: **$4.5M** Total Award Period Covered: **9/01/2020-08/31/2023** | |
| Annual Award Amount to PI’s Research: **$4.5M** | |
| Person-Months Per Year Committed to Project: **0.6** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. Synergies include data handling capabilities which are applicable to the proposed project scope of work.** | |

**Sam Webb**

|  |  |
| --- | --- |
| **Investigator: Sam Webb (Hedman)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Current** | |
| Project/Proposal Title and grant number, if appropriate **Department of Energy, BER, SSRL Operations and Research, FWP 10061** | |
| Source of Support: **DOE BER** Location of Project: **SLAC** | |
| Annual Award Amount: **$3.1M** Total Award Period Covered: **10/01/2019-9/30/2020** | |
| Annual Award Amount to PI’s Research: **$130K** | |
| Person-Months Per Year Committed to Project: **5.88** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. SSRL produces extremely bright x-rays as a resource for researchers to study our world at the atomic and molecular level. Research at SSRL benefits many sectors of the American economy and leads to major advances in energy production, environmental remediation, nanotechnology, new materials and medicine. SSRL provides unique educational experiences and serves as a vital training ground for future generations of scientists and engineers. BER funding provides for the user operations and research/ development for biological environmental research.** | |
| **Investigator: Sam Webb (McIntyre)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Current** | |
| Project/Proposal Title and grant number, if appropriate: **SSRL User Facility Operations**  **(SCW0010)** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC** | |
| Annual Award Amount**: $N/A** Total Award Period Covered: **10/01/2019-09/30/2020** | |
| Annual Award Amount to PI’s Research: **$N/A** | |
| Person-Months Per Year Committed to Project: **6.0** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application.** | |
| **Investigator: Sam Webb (Tassone)** | Other Agencies to which this proposal has been/will be submitted:  None |
| Support (Current, Pending, Submission Planned in Future or Transfer of Support): **Pending** | |
| Project/Proposal Title and grant number, if appropriate: **Autonomous Control of Experimental Stations at BES Lightsources** | |
| Source of Support: **DOE-BES** Location of Project: **SLAC/Stanford** | |
| Annual Award Amount: **$4.5M** Total Award Period Covered: **9/01/2020-08/31/2023** | |
| Annual Award Amount to PI’s Research: **$4.5M** | |
| Person-Months Per Year Committed to Project: **0.6** Pers. Months; Specify: Cal., Acad., or Sumr: | |
| Describe Research Including Synergies and Delineation with Respect to this Proposal/Award: **No overlap with application. Synergies include data handling capabilities which are applicable to the proposed project scope of work.** | |

ANL & ORNL BIOS

# Appendix 3: Bibliography and References Cited

1. Office of Science, D. of E. *Roundtable on Producing and Managing Large scientific data with Artifical Intelligence and Machine Learning*. https://science.osti.gov/-/media/bes/pdf/reports/2020/AIML\_Roundtable\_Brochure.pdf?la=en&hash=F770596E0D861B48836A7D221A567B46DF455E3A (2019).

2. Granda, J. M., Donina, L., Dragone, V., Long, D. L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* (2018) doi:10.1038/s41586-018-0307-8.

3. Houben, C., Peremezhney, N., Zubov, A., Kosek, J. & Lapkin, A. A. Closed-Loop Multitarget Optimization for Discovery of New Emulsion Polymerization Recipes. *Org. Process Res. Dev.* (2015) doi:10.1021/acs.oprd.5b00210.

4. Krishnadasan, S., Brown, R. J. C., DeMello, A. J. & DeMello, J. C. Intelligent routes to the controlled synthesis of nanoparticles. *Lab Chip* (2007) doi:10.1039/b711412e.

5. Gongora, A. E. *et al.* A Bayesian experimental autonomous researcher for mechanical design. *Sci. Adv.* (2020) doi:10.1126/sciadv.aaz1708.

6. Ren, F. *et al.* Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. *Sci. Adv.* (2018) doi:10.1126/sciadv.aaq1566.

7. Noack, M. M. *et al.* A Kriging-Based Approach to Autonomous Experimentation with Applications to X-Ray Scattering. *Sci. Rep.* (2019) doi:10.1038/s41598-019-48114-3.

8. Ling, J., Hutchinson, M., Antono, E., Paradiso, S. & Meredig, B. High-Dimensional Materials and Process Optimization Using Data-Driven Experimental Design with Well-Calibrated Uncertainty Estimates. *Integr. Mater. Manuf. Innov.* (2017) doi:10.1007/s40192-017-0098-z.

9. Gent, W. E. *et al.* Coupling between oxygen redox and cation migration explains unusual electrochemistry in lithium-rich layered oxides. *Nat. Commun.* (2017) doi:10.1038/s41467-017-02041-x.

10. Bianchini, M. *et al.* Spinel materials for Li-ion batteries: New insights obtained by operando neutron and synchrotron X-ray diffraction. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* (2015) doi:10.1107/S2052520615017199.

11. Keimer, B. & Moore, J. E. The physics of quantum materials. *Nature Physics* (2017) doi:10.1038/nphys4302.

12. Balents, L. Spin liquids in frustrated magnets. *Nature* (2010) doi:10.1038/nature08917.

13. Rau, J. G. & Gingras, M. J. P. Frustrated Quantum Rare-Earth Pyrochlores. *Annu. Rev. Condens. Matter Phys.* (2019) doi:10.1146/annurev-conmatphys-022317-110520.

14. Keimer, B., Kivelson, S. A., Norman, M. R., Uchida, S. & Zaanen, J. From quantum matter to high-temperature superconductivity in copper oxides. *Nature* (2015) doi:10.1038/nature14165.

15. Christensen, J. *et al.* Radiation damage in small-molecule crystallography: Fact not fiction. *IUCrJ* (2019) doi:10.1107/S2052252519006948.

16. Hoye, R. L. Z. *et al.* Perovskite-Inspired Photovoltaic Materials: Toward Best Practices in Materials Characterization and Calculations. *Chem. Mater.* (2017) doi:10.1021/acs.chemmater.6b03852.

17. Calder, S. *et al.* A suite-level review of the neutron powder diffraction instruments at Oak Ridge National Laboratory. *Review of Scientific Instruments* (2018) doi:10.1063/1.5033906.

18. Betterton, J., Webb, S., Ratner, D. & Kochenderfer, M. Adaptive illumination at x-ray sources with reinforcement learning. in *ICRA* (2020).

19. High resolution neutron powder diffraction: a case study of the structure of C 60 . *Proc. R. Soc. London. Ser. A Math. Phys. Sci.* (1993) doi:10.1098/rspa.1993.0095.

20. Hauschild, A. C. *et al.* Peak detection method evaluation for ion mobility spectrometry by using machine learning approaches. *Metabolites* (2013) doi:10.3390/metabo3020277.

21. Park, W. B. *et al.* Classification of crystal structure using a convolutional neural network. *IUCrJ* (2017) doi:10.1107/S205225251700714X.

22. Lee, J. W., Park, W. B., Lee, J. H., Singh, S. P. & Sohn, K. S. A deep-learning technique for phase identification in multiphase inorganic compounds using synthetic XRD powder patterns. *Nat. Commun.* (2020) doi:10.1038/s41467-019-13749-3.

23. Duris, J. *et al.* Bayesian Optimization of a Free-Electron Laser. *Phys. Rev. Lett.* (2020) doi:10.1103/physrevlett.124.124801.

24. Vecsei, P. M., Choo, K., Chang, J. & Neupert, T. Neural network based classification of crystal symmetries from x-ray diffraction patterns. *Phys. Rev. B* (2019) doi:10.1103/PhysRevB.99.245120.

25. Li, L., Lu, M. & Chan, M. K. Y. A Deep Learning Model for Atomic Structures Prediction Using X-ray Absorption Spectroscopic Data. *arXiv* (2019).

26. De Gelder, R., Wehrens, R. & Hageman, J. A. A Generalized Expression for the Similarity of Spectra: Application to Powder Diffraction Pattern Classification. *J. Comput. Chem.* (2001) doi:10.1002/1096-987X(200102)22:3<273::AID-JCC1001>3.0.CO;2-0.

27. Iwasaki, Y., Kusne, A. G. & Takeuchi, I. Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries. *npj Comput. Mater.* (2017) doi:10.1038/s41524-017-0006-2.

28. Kolouri, S., Park, S. R., Thorpe, M., Slepcev, D. & Rohde, G. K. Optimal Mass Transport: Signal processing and machine-learning applications. *IEEE Signal Process. Mag.* (2017) doi:10.1109/MSP.2017.2695801.

29. Kramer, M. A. Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.* (1991) doi:10.1002/aic.690370209.

30. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* (2004) doi:10.1109/TIP.2003.819861.

31. Sheikh, H. R. & Bovik, A. C. Image information and visual quality. *IEEE Trans. Image Process.* (2006) doi:10.1109/TIP.2005.859378.

32. Edelen, A., Neveu, N., Emma, C., Ratner, D. & Mayes, C. Machine Learning Models for Optimization and Control of X-ray Free Electron Lasers. in *Second Workshop on Machine Learning and the Physical Sciences* (2019).

# Appendix 4: Facilities and Other Resources

**I. Beamlines**

The proposed automation of scanning XRF imaging, and powder x-ray diffraction will be carried out at the following SSRL end stations. We will apply for beam time, which will be supplemented with staff priority time, and facilities commissioning time. Additional end stations with similar configurations to the target end stations are listed, and will be utilized as needed for prototyping prior to commissioning. The details of the end stations are provided below:

* **Beamline 2-1 at SSRL:** Excels in high resolution X-ray diffraction. This is a bending magnet beamline with an energy range from 5.5 keV to 17 keV allowing for energy dependent X-ray scattering measurements (resonant elastic X-ray diffraction, REXD). High resolution measurements are done using a crystal analyzer/photo-multiplier tube setup. Time-resolved experiments are also possible using the Pilatus 100k detector. This beamline also has capillary mounting capabilities ideal for high resolution powder diffraction.
* **10-2 at SSRL:** This beamline is a multi-pole wiggler endstation, that has two sections in the experimental hutch, housing a scattering station in the rear station, and a spectroscopy imaging station in the front. The front hitch will be primary used in this project. The beamline has a Si(111) and Si(220) LN2 cooled monochromator options, and used a toroidal focusing mirror to achieve a focus of 200 um x 500 um. A Newport hexapod can place and position optical apertures to provide beam sizes on the sample from 200um down to 20um. The large size sample stage achieves rapid motion of 600mm horizontally and 300mm vertically. The station uses a series of 4-element and 1-element Vortex Si-drift detectors for fluorescence detection, equipped with Xspress3 counting electronics. A 7 crystal HERFD analyzer is used for high resolution fluorescence detection experiments.
* **Beamline 17-2:** This is an undulator beamline, currently being commissioned, and equipped with a 6-circle diffractometer for scattering measurements with an energy range from 5keV to 20keV. The optics of this beamline will allow for focused or unfocused operation (producing beam spots from roughly 10μm x~20μm up to several hundred μm) and includes multiple monochromator crystal sets.

The proposed automation and associated commissioning measurements for multimodal x-ray scanning imaging will be carried out at the Advanced Photon Source using existing instruments at the following end station.

* **APS Beamline 25-ID:** At the end of 2020 the spectroscopy facilities at 20-ID will be moving to a canted undulator beamline at 25-ID. One branch includes an x-ray microprobe with variable resolution down to 1 μm along with a variety of high-resolution detectors for microspectroscopy. On the other branch stations are equipped for x-ray emission spectroscopy, x-ray Raman, and XAFS experiments requiring a high-brilliance source. Also, an Eiger 500K detector is available for microdiffraction. Combining these capabilities with a new high-flux multilayer monochromator will allow multi-modal imaging. Both sectors will also support time-resolved pump-probe measurements.

The proposed neutron diffraction experiments will be carried at the SNS and the HFIR at ORNL using existing instrumentation at the following beamlines:

* **BL-1B (NOMAD) at SNS**: The Nanoscale Ordered Materials Diffractometer (NOMAD) is a high-flux, low-to-medium-resolution diffractometer for structural characterization of both local order in liquids, solutions, glasses, polymers, nanocrystalline, and disordered bulk materials, as well as time-resolved and sample size limited structure determination in long-range ordered polycrystalline materials.
* **BL-9 (CORELLI) at SNS**: CORELLI is a statistical chopper spectrometer with energy discrimination. The momentum transfer ranges from 0.5 to 12 Å-1, and the energy of incident neutrons ranges from 10 to 200 meV. This instrument combines the high efficiency of white-beam Laue diffraction with energy discrimination by modulating the beam with a statistical chopper. A cross-correlation method is used to reconstruct the elastic signal from the modulated data.
* **BL-11A (POWGEN) at SNS**: POWGEN is the highest resolution neutron powder diffractometer in the USA. The geometric construct of the instrument allows for all detected scattered neutrons to be focused onto a single diffraction profile yielding a high count rate with varying resolution as a function of d-spacing, preserving good resolution of Δ*d*/*d* = 0.0015 at *d* = 1 Å. POWGEN is optimized for both parametric studies of materials under a wide range of conditions and *ab initio* crystal structure determinations of complex solid-state materials with asymmetric unit-cells of the order ∼1500 Å3. The instrument is equipped with an automatic 24-sample changer to cover the temperature range 10 – 300 K.
* **HB-2A (POWDER) at HFIR**: The HB-2A powder diffractometer is primarily utilized for magnetic structure determination, with an emphasis on ultra-low temperatures combined with high field or pressure. The constant wavelength neutron beam, with options of 1.12 Å (Ge117), 1.54 Å (Ge115) and 2.141 Å (Ge113), flat background, simple beam profile, and open instrument layout make HB-2A well-suited to a variety of interchangeable sample environments with minimal instrument calibration.
* **HB-2C (WAND2) at HFIR**: The WAND2 instrument is a high intensity dual-purpose powder/single-crystal diffractometer with a continuous wavelength selected from a Ge monochromator with a low take-off angle. WAND2 recently completed its second and final upgrade phase, which amounts to a complete replacement of the whole instrument after the monochromator with the deployment of a large 2D position sensitive detector with nearly 2 × 106 pixels. The data rate achievable from the new detector enables kinetic studies over time scales of seconds.

**II. Conventional Laboratory Space**

SSRL maintains several laboratories for general-user experiments and for SSRL staff that are located near the beamline stations. Laboratory equipment includes fume–hoods, furnaces, controlled-atmosphere glove boxes, a spin-coater, UV-VIS spectrophotometers, a Jeol JES-RE1X X-band EPR spectrometer with an Oxford cryostat, as well as a wide variety of common glassware, lab supplies, and chemicals and other equipment necessary for sample preparation, handling, and characterization on-site. All laboratories provide personal protective equipment, laboratory safety training, and hazardous waste disposal.

One smaller laboratory (≈800 sq ft) is in the SSRL building and has close proximity to SSRL beamlines. This lab is mainly used for preparing samples for immediate use at these beamlines. There are two gloveboxes available for air sensitive samples and a fume hood in an adjacent laboratory.

HFIR and SNS operations are supported by sample environment teams and dedicated laboratories at each facility that maintain and develop cryomagnets and cryostats with dilution and 3He inserts, CCRs, vacuum and controlled atmoshpere furnaces and sample-changers for use on the instruments. Several laboratories for general-user and staff experiments are in operation to support neutron beamline experiments. Laboratory equipment includes controlled fume–hoods, furnaces, controlled-atmosphere glove boxes, microscopes, lab x-ray diffractometers. These facilities allow the preparation of both non-radiated and irradiated samples.

**III. Computing**

**SLAC Computing**

ML algorithm development will leverage both local clusters for early development and super-computers for large-scale learning. For local computing, SLAC has approximately 5000 high-performance cores, 3500 high-throughput cores, and 350 GPU cards available for lab-wide use, including 100 GPU cards devoted to ML development. The project will have use the National Energy Research Scientific Computing Center (NERSC), as well as the other high performance computing facilities described below.

**Argonne National Laboratory Computing**

The Advanced Photon Source operates a **High-Performance Computing (HPC)** system designed for near real-time data processing of beamline data. The system is a distributed-memory Linux- based computing cluster consisting of 74 nodes with a total of 1,224 CPU cores. There is 5 TB of total RAM across the cluster. Eight NVidia GPUs are available for processing CUDA analysis jobs. A 10 Gbps InfiniBand backplane connects the nodes and storage system. Approximately 600 TB of short-term data storage is available via a Lustre parallel-distributed file system.

**The APS Data Management System** are a set of tools and services built using Globus that provides data transfer and data ownership management capabilities for APS experiment data. Data sets can be moved from beamline detector systems to a centralized storage system over multiple 10 Gbps network links. Presently, the system consists of a Data Direct Networks (DDN) GRIDScaler storage device with 3.6 PB of usable storage (expandable to approximately 15 PB of usable storage) running the IBM GPFS file system. The system is connected to APS beamlines via 4 x 10 Gbps network links, expandable as needed. Eight Data Transfer Nodes (DTNs) connect the system to external networks.

**APS Network Connectivity:** High-bandwidth data networks help underpin the Advanced Photon Source’s. Within the Advanced Photon Source, a 40 Gbps network backbone established high-bandwidth data connections between beamline instrumentation and the facility’s computing and data storage systems. The Advanced Photon Source is connected to Argonne National Laboratory’s campus network via a 100 Gbps network uplink. Data transfers outside of the Argonne site and to the Internet2 are over the Energy Sciences Network (ESnet), an 8.8 Terabit network. The APS and ANL are connected to ESnet via a 100 Gbps link. Using ScienceDMZs, data transfers can be optimized with fixed paths to computing centers and other destinations for more guaranteed bandwidth and lower latencies.

The **Argonne Leadership Computing Facility (ALCF)** houses a range of high-performance computing capabilities, currently spearheaded by its **11-petaflop Theta** system based on the second-generation Intel Xeon Phi processor, which enables breakthrough computational science and engineering research. Theta is a massively parallel, many-core system based on Intel processors and interconnect technology, a new memory architecture, and a Lustre-based parallel file system, all integrated by Cray’s HPC software stack. The system is equipped with 4,392 nodes, each containing a 64-core processor with 16 gigabytes (GB) of high-bandwidth in-package memory (MCDRAM), 192 GB of DDR4 RAM, and a 128 GB SSD. Theta’s initial parallel file system is 10 petabytes.

Designed in collaboration with Intel and Cray, Theta serves as a steppingstone to the ALCF’s next leadership- class supercomputer, the **Aurora exascale supercomputer**, to become available in 2021. Aurora is designed to support numerical simulation, data analysis, and deep learning applications. To this end, it is architected with a mix of Intel CPUs and GPUs to deliver sustained performance of greater than one exaflop/s (1018 full-precision floating point operations per second) and substantially higher compute rates at reduced precision. It will have aggregate system memory of more than 10 petabytes and more than 230 petabytes of high-performance storage accessible at 25 terabyte/s.

**ORNL Computing**

***Summit*** is a 4,608 node IBM AC922 system that provides 9,216 IBM Power9 processors and 27,648 Nvidia Volta graphics processing units (GPUs). The Volta GPUs each boast a theoretical peak performance of 7 petaflops (PF), bringing the total peak performance of Summit to 200PF. With total system memory of 10 petabytes (PB) and a Red Hat Enterprise Linux operating system, Summit will enable unprecedented capabilities for machine learning and artificial intelligence. Summit is connected to a 250PB center-wide file system at a peak speed of 2.5 gigabits per second (Gb/s). The system is connected by a Mellanox EDR 100Gb/s InfiniBand interconnect that includes approximately 10,000 fiber-optic cables which are an aggregate of 180 miles in length.

The **CADES** (Compute and Data Environment for Science) facility originated from a desire to build upon ORNL’s key strengths in data system infrastructure and delivery of new capabilities through data intensive science to meet the mission needs of R&D projects at ORNL and beyond, while addressing big data analytics and science needs. The technical objective of the CADES facility is to provide a data intensive infrastructure that supports the mission needs of key internal and external projects at ORNL. The hardware infrastructure is comprised of a multi-petabyte data storage environment coupled with a multi-teraflop data intensive HPC compute environment and a multi-node cloud compute infrastructure. This environment includes the necessary software to apply the system to important data intensive problems at ORNL.

**The ORNL Data Management System**

ORNL uses the ADARA (Accelerating Data Acquisition Reduction and Analysis) project to communicate files from the beamlines to central storage. Data reduction is performed with the Mantid framework, and for most instruments data is reduced automatically. Data is processed using the analysis.sns.gov cluster and stored. Data is available to users by using a thinlinc server run in the analysis cluster. ORNL facilities use the NeXus data format to store data. The NeXus format was developed by an international collaboration of laboratories to define a common archival storage for neutron, x-ray, and muon experiments. It provides a standard and clear structure to capture the data and meta-data that describe a scattering experiment to facilitate their use by third-party software. At SNS and HFIR, the instrument data acquisition system produces HDF5 files using the NeXus format to capture all the data necessary to fully describe a measurement. In addition to the neutron data itself, it also contains meta-data describing the parameters of the measurement and time series of all process variables generated by the instrument. Such process variables include motor positions, information about the neutron source, beam chopper information, and any sample environment that may have been controlled by the system during the measurement. In both formats, extraction of relevant meta-data, both for the experiment as well as subsequent analysis, is simple.

# Appendix 5: Equipment

The required equipment for this proposal include dedicated computers (desktop or laptop with enhanced GPU capability) for the participating researchers for code development, testing, and data analysis. Additionally, access to the facilities described in Appendix 4. We do not anticipate the need for any additional equipment to complete the scope of work proposed beyond this.

# Appendix 6: Data Management Plan

In this data management plan we address the Office of Science’s statement on Digital Data Management (http://science.energy.gov/funding-opportunities/digital-data-management/) regarding the sharing and preservation of digital research data. The purpose of our data management plan is to publically release information that documents and validates our results and facilitates subsequent research in the broader community.

The primary data products generated through this project will generally fall into two categories of data: (1) consisting of mathematical algorithms, computer codes and machine-learning (ML) models and (2) consisting of experimental and simulated data used to train and validate those models. Both categories are discussed below:

1. **Algorithms, codes, and ML models**

The ML models with all its constituents would be packaged in containers in DLHub and published. DLHubs are designed to make publishing, sharing, verifying, reproducing, modifying and reusing models, especially ML model, easy and use-friendly. We will also deposit software associated with models and other data processing tasks on Github, and share software and other algorithmic and computational products with CAMERA for Lightsource wide distribution. To further widen access and usability of the ML software, we will publish a pre-configured virtual machine containing trained models to WholeTale (<http://wholetale.org/>).

1. **Experimental and Simulated Data**

The experimental raw data generated through this project would be saved in its native format on servers at Scientific User Facilities. The processed and data associated with new discoveries would be published alongside publications and made available on open databases and upon request. Experimental data, associated meta-data and curated results will be also published and made widely available through the open data management platforms of the Materials Data Facility (<http://Materialsdatafacility.org>) and the Materials Data Repository (<http://Materialsdata.NIST.gov>). Specific training and testing data for various ML models will be archived alongside those models in appropriate code repositories.

# Appendix 7: Letters of Commitment

