S2I2 Conceptualization Proposal

Scientific Software Innovation Institute for
Advanced Analysis of
X-Ray and Neutron Scattering Data
SIXNS

Brent Fultz[1]
Professor of Materials Science and Applied Physics, PI

Simon Billinge[2]
Professor of Materials Science and of Applied Physics and Applied Mathematics, Co-PI

Houman Owhadi[1]
Professor of Applied & Computational Mathematics and Control & Dynamical Systems, Co-PI

John Rehr[3]
Professor of Physics, Co-PI and

Mark Stalzer[1]
Executive Director of the Center for Advanced Computing Research, Co-PI

[1] *California Institute of Technology, Pasadena, California*

[2] *Columbia University, New York, New York*

[3] *University of Washington, Seattle, Washington*

December 12, 2011

Table of Contents

A conceptualization effort is proposed for designing a Sustainable Software Innovation Institute to elevate the level of scientific computing in X-ray and neutron scattering science (SIXNS). The SIXNS Institute would adapt modern methods of computational materials science to predict scattering from materials. It would incorporate these software tools into workflows for scattering scientists, giving them new pathways to scientific discovery.

Since 1980, the performance per dollar of computer hardware has increased by a factor of 100 every decade. Over the same time period, this million-fold improvement has been closely matched by the increased brilliance of X-ray sources, and in the past decade the performance of neutron sources has increased by a factor of ten. These improvements should be multiplied by comparable factors to account for improvements in software and methods of computational science, and for major improvements in optics and detectors for X-rays and neutrons. These enormous advances in computing and in scattering have occurred independently. Today there are exciting opportunities for combining them to do new science, and there is a growing body of work in computational scattering science that does so. Today this is only a small fraction of the work done by users of the synchrotron and neutron sources in the U.S., but it accounts for a disproportionate fraction of high impact publications.

The intellectual merit of the proposed activity is based on developing new computational workflows that open channels for discovery in scattering science. Sometimes this is as direct as offering a common environment for comparing results from experiment to results from computational materials science. Computing also facilitates the combined analysis of information from different types of experiments, linked by an underlying model of the structure and dynamics of a material. Such a combined approach requires the assessment of uncertainties in the model using mathematical methods that are not yet standard practice in scattering science. For example, when measurement uncertainties are well understood, Bayesian methods can often be used to refine prior information as new results are added, and it is understood in principle how to quantify the uncertainties when combining similar scattering experiments. In many other cases, however, the models are not well specified, and new methods of uncertainty quantification are needed from computational and mathematical sciences. This is especially true when experimental results are used to optimize a computational model with parameters that relate indirectly to measured quantities.

For broader impact, the SIXNS Institute is expected to serve a significant fraction of the community of 14,000 annual users of X-ray and neutron scattering facilities in the U.S. Workflows that include calculations of the structure and dynamics of materials can allow experimental results to be interpreted on a more fundamental level, letting scientists explore properties that are not measured directly by experiment. Such workflows are needed by the Materials Genome Initiative, and by the interdisciplinary field of materials science. Although several workflows have been identified as priorities, a major part of the conceptualization effort would be interacting with the scattering science community to better define these priorities. A few collaborations with individual scientists on focused topics are appropriate if they also develop prototype workflows of value for related types of science. It seems likely that computational scattering science initiatives will develop with DOE support at the national user facilities, and representatives from these facilities will participate in the Conceptualization Phase and future governance of the SIXNS Institute. There is a large overlap in the needs for software by both scientists and students, so a SIXNS Institute could offer materials simulations that serve the needs of education. There are also opportunities for the Institute to bring a better gender balance to the field of computational science.

When the Conceptualization Phase is complete in two years, the priorities of users and facilities should be well defined. It will be a good time to set the directions of computational scattering science in the U.S. Finding these directions is important, but it is also important to do scientific discovery with new computational workflows now. Using new tools for discovery will advance scattering science, and will help define efficient processes for serving the scattering science community.

1. Results from Prior NSF Support

Brent Fultz was the P.I. on the project IMR-MIP: DANSE Distributed Data Analysis for Neutron Scattering Experiments, DMR-0520547, funded as a construction project from 6/1/2006 – 11/30/2011 [1]-[41]. The budget was M$ 11.9, with subcontracts to Columbia University (Simon Billinge), University of Tennessee (Paul Butler), University of Maryland (Robert Briber/Paul Kienzle), and Iowa State University (Ersan Ustundag). The different subcontracts addressed different subfields of neutron scattering research, and these different subfields have different computing needs. The software products were released from 2008-2011 through the web site http://danse.us.

The small angle scattering project under Paul Butler produced SansView, now in version 2.0.1. This standalone software package is deployed on computers running Windows or Macintosh operating systems, and offers modeling of 1D and 2D SANS data with a rich selection of models of shapes and orientation distributions. It also allows inversions of data from k-space to real space. This is all done with a consistent graphics-oriented user interface.

The diffraction subproject under Simon Billinge developed tools for atomic pair distribution function analysis of X-ray or neutron diffraction data [36]-[41]. The PDFfit2 package is an analysis engine that can be run from scripts or through a graphical user interface, PDFgui. It is deployed for Windows, Linux and Macintosh operating systems. SrFit is a prototype modular software framework that allows different models, scattering functions and regression schemes to be plugged in to a fit as needs dictate. SrRietveld is a gui wrapper simplifying workflows for high throughput Rietveld refinements.

The inelastic scattering subproject under Brent Fultz built the data reduction software that has been used at the ARCS spectrometer since its commissioning in 2008. The larger effort was developing vnf, the virtual neutron facility. It is deployed as a web service that offers high performance computing to multiple users. Vnf provides Monte Carlo simulations of neutron scattering experiments, including ab-initio and molecular dynamics calculations of phonon scattering from the sample. It was released in 2011 and is undergoing testing with users.

The reflectometry subproject built tools for off-specular scattering from 3-D objects, using the distorted wave Born approximation, magnetic scattering tools for reflectometry, and an interface to a global optimizer for multi-parameter model fits to specular scattering data. The user interface is not yet complete.

The engineering diffraction subproject did not go to completion owing to health issues with the Co-I. Project management tools identified the poor performance, and funding was reduced and eliminated midway through the DANSE project.

Today, DANSE software for small-angle scattering is being merged into the Mantid development trunk, and will be maintained by the Mantid team so long as they receive financial support from the SNS and ISIS. Other parts of the DANSE software are being sustained with support of the SNS for the next two of years to retain their present capabilities, and gauge community acceptance.

John Rehr is the P.I. on the project SI2-SSE: Cloud-Computing-Clusters for Scientific Research, OCI-1048052 funded from 9/15/10 – 8/31/2013 (k$ 489 to date). The aim of this research is to develop a complete scientific computing cloud environment that is robust, easy to use, and cost-effective, and which contains pre-optimized applications of significant interest to the scientific community. In prototype studies our research demonstrated the feasibility of such scientific cloud computing [42] using the parallelized FEFF X-ray spectroscopy code as a prototype. More recently they have developed a a more complete "virtual machine prototype" for the high-performance Amazon elastic compute cloud (EC2); this virtual image contains the operating system and selected scientific software including the X-ray code FEFF and electronic structure codes WIEN2k, and ABINIT. Technical details of this compute environment are described in a preprint [43] which has been submitted for publication (Sept. 2011). This virtual cloud platform will make high performance scientific computing widely available, especially for materials science applications, without the need

for expertise in high performance computing environments or the need to purchase and maintain sophisticated hardware.

Mark Stalzer received NSF support for the project "CDI-Type 1-Bringing a Bayesian perspective to the study of large earthquakes and their impacts on the built environment" award number 0941374 for the amount: k$ 700 from 1/1/10 – 12/31/12. The PIs (Mark Simons, James Beck and Mark Stalzer) are developing a Bayesian framework and computational tools for challenging ill-conditioned inverse problems in high dimensions related to earthquake fault slips [44]-[47]. This has included the development of CATMIP (Cascaded Adaptive Tempered Monte Carlo in Parallel), a multi-stage Markov chain Monte Carlo algorithm that uses local random walking with simulated annealing. It has been run on a multi-processor system at Caltech to infer the space-time function for the fault slip for the 11 March 2011 Tohoku-oki Earthquake (M9.0) [44]. Two graduate students and a post-doctoral fellow received training under this project.

Houman Owhadi received NSF support for the project DynSyst_Special_Topics: Dynamics and Control of Bio-molecular Systems using Geometric Model Reduction and Stochastic Variational Integrators award number CMMI-092600 for the amount: k$ 450 from 9/1/09 - 08/31/12. The PIs (Houman Owhadi and Jerry Marsden (deceased)) have developed robust structure preserving integrators for the simulation and control of (possibly stochastic and multi-scale) Hamiltonian systems. Results have included the development of FLAVORS (flow averaging integrators) and the sparse approximation of PDEs with stochastic inputs [48]-[55]. One graduate student and a post-doctoral fellow received training under this project.

2. Present Status: Computing in Scattering Science

2.1. Overview

The past decade has seen extraordinary advances in X-ray and neutron scattering research. The Spallation Neutron Source is reliably operating with an order-of-magnitude increase in neutron flux over its predecessors, and its instruments offer similar gains in efficiency. The Linac Coherent Light Source, with brightness six orders of magnitude higher than previous synchrotron sources, has begun operations. The NSLS2 project at Brookhaven National Lab is well underway, and upgrades at the Advanced Photon Source and the NIST Center for Neutron Research are keeping these facilities competitive for neutron and X-ray scattering science.

Advances in computing are equally spectacular. Since 1980, hardware performance per dollar has increased by more than a factor of a million. Software advances have led to a rich field of discovery in computational materials science, and computations that were unimaginable only 20 years ago are commonplace today.

There are many opportunities for combining results from scattering experiments with new computational methods. The data of Fig. 1a show results from a keyword search for publications that combine ab-initio theoretical calculations with scattering experiments. This is not the whole picture of computing and scattering science because computing is used for every publication by the 14,000 annual users of the synchrotron and neutron facilities in the U.S. A keyword search for papers on neutron scattering or X-ray scattering gives 60 times as many publications. Although almost all research with X-ray or neutron scattering involves studies of materials, today only few percent of these studies use the computational tools from modern materials science.

This proposal shows new paths to scientific discovery by combining computing and scattering science, and points out some natural combinations of computing and scattering measurements that have not been tried, or have been used far below their potential. Compared to 10 years ago, today it takes far less effort to learn and use today's tools for computational materials science. It is more difficult to include these computational tools into workflows for scattering science, however, and it is here where the scattering community needs help. This is the motivation for a Scientific Software

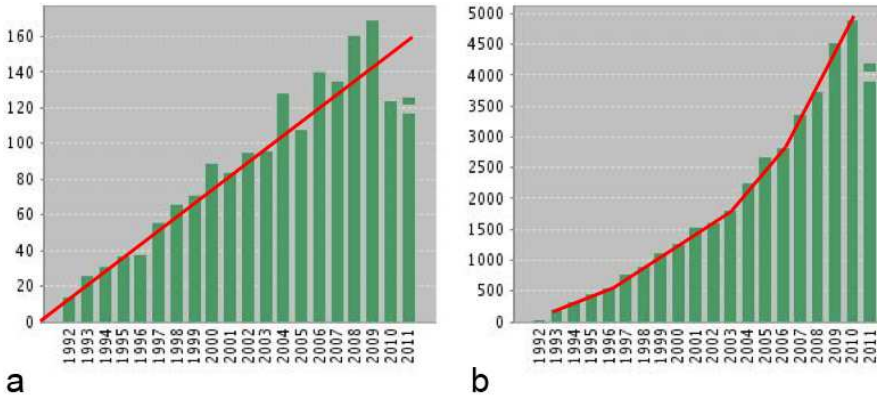

Figure 1. Results of Nov. 4, 2011 from a keyword search of Web of Knowledge, an electronic publications database of Thomson Reuters. The keyword entry was: "ab-initio AND scattering AND (X-ray OR neutron)" (**a**) Published papers each year, showing linear trend since 1990. (Total papers is 1800.) (**b**) Citations to the papers each year, showing an increasing rate of citation. (Average since 1992 is 21.5 citations/paper)

Innovation Institute. This Institute would not write new codes for computational materials science, but would develop flexible workflows for interpreting experimental data by scattering scientists. Direct comparisons of computed and experimental results are often appropriate and useful, and there are numerous other opportunities to mix or iterate between computing and measurements, or combine experimental results from different types of measurements. An Institute, with the placeholder name SIXNS, will build these workflows, adapt them for new science, and support scattering scientists who need them.

When sophisticated features are extracted from a workflow that combines computing and experiment, or when different types of experimental data are combined to develop an underlying materials model, we have little experience with the reliability of the results. Uncertainty quantification needs to be considered in building such new scientific workflows. Although underutilized in scattering science, Bayesian methods can be incorporated naturally into efforts that combine computation with experiment, using results from one as prior information for the other, for example. These methods are challenged when the models for obtaining conditional probabilities are not well known, and this is a topic for applied mathematics research as described below.

2.2. The Materials Genome Initiative for Global Competitiveness

The Materials Genome Initiative seeks to use computing to shorten the time from materials discovery to commercialization [56]. The proposed SIXNS Institute addresses key issues in the early phases of materials discovery and optimization. The proposed Institute addresses directly three key recommendations at the end of the Materials Genome Initiative description [57]:

- *Recommendation 1. NSF and DOE will work together to develop software for the next-generation design of matter.* The SIXNS Institute, if funded by the NSF, will provide modern software tools for interpreting the data acquired at DOE BES national user facilities. These computational methods will improve the chances of materials discovery, and strengthen the workflow of getting the results from these facilities into scientific publications. Representatives from these DOE user facilities will be on the Governing Board of the SIXNS Institute.

- *Recommendation 2. DOE and NSF will coordinate the development of materials characterization tools for validation of computational results.* X-ray and neutron scattering facilities are premier materials characterization tools. The SIXNS Institute would provide software tools to extract more detailed information on structure and dynamics measured with these experimental tools. The structure and dynamics of atoms are the *sine qua non* of understanding materials properties, and designing the properties that give them commercial value.

– *Recommendation 6. NSF and DOD will lead in addressing next-generation workforce goals.* The SIXNS Institute will be university-based, where undergraduate students, graduate students, postdoctoral fellows, and junior scientists pass at different stages of their careers. Our diversity strategic plan addresses directly the need to bring more women into computational science.

2.3. Some Opportunities for Computational Scattering Science

Nearly all scattering science research with X-ray and neutron facilities is on studies of materials, but the experiments cover an enormous range of science. Nevertheless, some computational materials science methods are important for broad classes of materials, such as methods for calculating electronic structure, molecular dynamics, and tools for modeling atomic structure and dynamics. Principles of materials at the atomic level, such as quantum mechanics and statistical mechanics, are directly relevant to a large fraction of scattering science research. It is our plan to develop specific scientific workflows that use modern tools of computational materials science to assist in interpreting specific scattering experiments, but we can select workflows that use computational methods that are applicable to wide ranges of materials and materials phenomena. In this way an Institute could soon offer tools of value for a significant fraction of this large user community. A goal for the Conceptualization Phase is to refine the goals for the community of scattering scientists, and to work with individuals in the community to better understand what is practical. Here we list some opportunities to show the existence of new directions in computational scattering science, but this listing cannot be considered finalized.

2.3.1. *Diffraction*

Elastic scattering research has advanced with new experimental capabilities for measuring diffraction to high Q, giving an unprecedented amount of detail on arrangements of atoms at short-, medium- and long-range length scales. Modern methods of Rietveld refinement and analyses of pair distribution functions employ the optimization of an underlying structural model. With complex models comes the technical difficulty of performing the optimizations, and questions about the uniqueness of the structural information. Bringing more experimental data into the modeling effort should improve the reliability of the model.[1] A SIXNS Institute could develop software to allow multiple X-ray wavelength analysis to become a routine part of structural analysis. The anomalous dispersion at an X-ray absorption edge changes the scattering cross-section of the resonant ion. By taking a difference between scattering at the edge and away from it, it is possible to get chemical-specific structural information. This is not new, but it faces difficulties in data analysis due to changing absorption and fluorescence signals. We will both develop software to make data collection and analysis routine, but also the software to include this information in a co-refinement scheme with high resolution non-resonant scattering of X-rays and neutrons. (This will also be done for the much less exploited PDF side of things. Differential anomalous PDFs provide very similar information to EXAFS, but include atomic correlations over longer distances which is highly complementary.) Combining different diffraction datasets, which have similar structures but different emphases on different atoms, is well suited for Bayesian analyses of uncertainties. We know of no such attempts at uncertainty quantification in diffraction patterns, but this is a natural opportunity to bring methods of applied mathematics to address the reliability the results. Such an effort promises reliable structural models of atom arrangements in nanostructured materials, for example.

2.3.2. *EXAFS and Diffraction*

EXAFS has matured into one of the most important techniques for local structure determination, due in large part to the development of accurate theoretical models to fit experimental measurements [58]. However, the analysis of complex systems is often more difficult, due to the limited

[1] Combining X-ray and neutron diffraction data, for example, has been used for decades, but this style of research is specialized and its benefits are not always easy to quantify.

information content in the EXAFS signal. Likewise despite its high signal to noise ratio, extracting quantitative information from the near edge structure (XANES) is constrained by the limited data range though this spectral regime contains important chemical information. To be applicable to many complex materials of current technological interest, quantification of the measurement uncertainties is essential. Such a quantification can be carried out using inverse methods based on Bayesian fitting techniques [59]-[62]. These methods generalize conventional least squares fitting methods through the addition of appropriate a priori data. This generalization naturally separates the fit parameters into relevant and irrelevant subspaces and stabilizes the fits. Similar techniques can be applied to DAFS (Diffraction Anomalous Fine Structure) which measures EXAFS-like structure in the Bragg peaks of X-ray diffraction signals [63].

The EXAFS method gives local structure centered at a single atom, whereas traditional diffraction methods offer a global view of the whole crystal, and PDF methods give non-chemically-resolved local structural information. Information from the three methods may be a powerful combination, especially for determining atom arrangements around dilute chemical species. For example, the local correlations obtained from EXAFS give prior information that can restrict the structural models used to fit pair distribution functions from diffraction data. For such a combined workflow, strategies to optimize the reliability of posterior models of atom structure have not been developed. We will explore possible strategies during the Conceptualization Phase. Earlier attempts to combine PDF and EXAFS showed that incompatible, improperly-handled systematic errors in each method led to uncertain results. The errors resulted in slightly different bond-lengths from the two methods. Combining methods allows us to find systematic errors, and then forces us to account for them properly. This should improve the methods themselves, to the benefit of a much broader community.

2.3.3. *Electronic Structure by Inelastic Scattering*
Non-resonant (NRIXS) and Resonant inelastic X-ray scattering (RIXS) are examples of important novel X-ray spectroscopies that have been made possible by high brilliance third generation X-ray sources. In particular NRIXS measures the dynamic structure factor $S(\vec{q}, \omega)$. In contrast to EXAFS, the NRIXS signal also includes the momentum transfer dependence $\vec{q}$ which yields a more complete description of the electronic density of states [64]. The resonant enhancement in RIXS yields data of very high resolution due to the suppression of broadening effects from short lifetimes of deep-core holes. On the other hand, RIXS simulations require a solution of the Kramer-Heisenberg equations rather than Fermi's golden rule, which considerably complicates the analysis. Recently, however, an approximation has been developed [65] which permits an interpretation similar to that for XAS and XES. Both RIXS and NRIXS can now be modeled by extensions incorporated in the FEFF9 code, and can be analyzed by a Bayesian inverse analysis similar to that described above.

2.3.4. *High-Resolution Inelastic Scattering*
Atomic motions are routinely studied by techniques such as *ab-initio* and classical molecular dynamics (MD). Diverse scientific fields, soft and hard condensed matter for example, have employed these methods. Although different phenomena are studied, they are frequently on similar scales of length and time as measured in scattering experiments. Today, however, it is still unusual for inelastic scattering measurements to be accompanied by simulations of atom dynamics, even by the classical MD methods that are routine in computational materials science. Inelastic scattering studies of atom dynamics at surfaces are emerging, as are methodologies to calculate vibrational dynamics of surface phenomena. It is not hard to look ahead to see the advantage of having some interoperability between these experimental and computational methods for surface studies.

Calculating lattice dynamics from first-principles electronic structure methods is now routine for simple materials. Capabilities to incorporate first-principles lattice dynamics calculations in inelastic neutron scattering experiments were developed as web services in the DANSE project. The tools

were released in 2011 as part of the "virtual neutron facility," vnf, and are being evaluated by the community. Vibrational dynamics also plays an important role in EXAFS analysis since the temperature dependence is dominated by Debye-Waller factors. Likewise a number of first principles methods have been developed to calculate these quantities [66] based on first principles electronic structure calculations. It is possible that the SIXNS Institute will build on these capabilities for calculating atom dynamics, making them more user friendly and adapting them to new scientific workflows. Accurate lattice dynamical information is needed to properly model PDF and EXAFS data, allowing the correct separation of peak broadening from phonons from that caused by static or quasi-static disorder. There is currently no example where this has been done properly. The dynamics are handled as a fitting parameter in a harmonic approximation, but proper methods should be routine.

3. Path Forward: Identify Important Workflows, and Implement Them

Every investigation has unique features in its workflow from data to publication. Expertise in adapting the computational tools in this workflow is learned, much like the expertise in mastering the experimental tools. A mission of the SIXNS Institute is to lower the barriers for learning computational tools and interconnecting them. This requires working with scientists in the scattering community, and careful consideration of their opinions on prototype software, for example.

Bringing more experimentalists to computational scattering science is an effort that requires careful planning. From our past experiences, we believe the best way to do this is through a collaboration where an experimentalist and a computational scientist work together on a problem with a clear scientific goal. This approach assures that both parties have a shared mission in the workflow, and will share in the scientific discovery.[2] This high-level type of user interaction would be a mission of the SIXNS Institute. It is considerably different from what is understood as user support in the commercial software industry. We need to test methods for providing such support during the Conceptualization Phase. We have had success with targeted workshops, and this approach is our first priority.

Adapting computational tools to new scientific workflows is most efficient when the software packages are modular, and have a consistent structure for I/O, for example. The rules of good object-oriented programming go a long way towards satisfying this goal, although compatibility still must be designed. Our approach will be to use the Python language at a high level, with bindings to C, C++, and FORTRAN codes that perform the heavy work. This is the approach taken by DANSE, SNS, ISIS (through the Mantid project), and the LCLS for large parts of their scientific software infrastructure. An early priority of SIXNS is to identify software packages that should be adapted early for inclusion in modular scientific workflows.

A central goal of the proposed workshops in the Conceptualization Phase in 2012-2014 is to meet with the scattering community to discuss and identify workflow needs. These needs will influence strongly, although not dictate, the first efforts of a Institute proposal.

3.1. Candidate Workflows

It will be a goal of the Conceptualization Phase to identify specific workflows for new scattering science, and it is expected that two or three of these will be implemented to the level of beta testing. With the resources of the Conceptualization Phase we may be limited to building workflows from pre-existing components. In a SIXNS Institute the choice will be driven in part by scientific opportunity, but also by the future flexibility that is allowed by the software packages integrated

[2] Not all investigations need to take this path – there are cases where the computational tools are already available and can be put to use with little adaptation or instruction. As the computational tools become more available and user friendly, we expect more of these easy cases.

into the workflow. Some candidate examples are listed here to give concrete examples of what is possible, but these should not be considered final decisions.

3.1.1. *Visualization of Results from Experiment and Computation*
Visual output is an important gauge of fit quality, and gives essential guidance when comparing theoretical and experimental data. Basic tools are widely used in the scattering community, but challenges arise when looking for trends in multiple sets of multi-dimensional data. An important shift of visualization software is underway as data and services move into the cloud. This shift not only allows for unbounded data storage, but enables a wider range of interactions with the data and the community, through a large pool of computing hardware and web based services. On the client side, increasingly capable browser-based applications drive the interaction seamlessly. Examples of such applications in the business world include IBM's Many Eyes [67], Tableau Dashboards [68], and Bime [69]. Data visualization on the cloud provide benefits unachievable by desktop software such as ease of deployment, ease of access to stored data, and process sharing and collaboration.

3.1.2. *Uncertainty Analysis when Combining EXAFS spectra with Diffraction Patterns*
It is straightforward to generalize the Bayesian analysis technique to handle multiple data sets. Thus the same analysis engine can be used to combine data from different methods, each weighted by its own statistical uncertainties. Such a combined approach can be useful, for example, when combining XAS and SNS or DAFS data. The big challenges, however, are the systematic errors that currently manifest themselves in aberrations to the refined bond-lengths and peak widths. It is crucial to understand how to remove these errors, or estimate their effect (and consequently build that in as a non statistical uncertainty in a refinement engine by a suitable weighting mechanism). This requires prototyping and empirical testing, which will be made possible with the modular code frameworks that we are developing and will continue to develop as part of SIXNS, in combination with an understanding of uncertainty quantification from applied mathematics. What is already apparent from our prior efforts is that combining data from complementary sources has a significant effect on the topology of the fitting surface and can often smooth it, facilitating the search. Changing the weight of a data contribution during a refinement thus can have an effect on the outcome, even if it is weighted to zero at the end of the refinement. Issues like these need to be explored in detail to develop best-practice workflows.

3.1.3. *Errors in Thermodynamic Quantities for Excitations*
It is straightforward to use an energy spectrum from inelastic scattering to calculate a partition function for the excitations in a material, and this is routine for single phonon excitations in harmonic or quasiharmonic crystals of pure elements. The two elements of a binary alloy generally have significant differences in their efficiencies for phonon scattering, however. Today we can compute the lattice dynamics of alloys by first principles electronic structure methods, and use these results to calculate a phonon partition function and all thermodynamic quantities. It is possible to take another step, however, to assess the reliability of workflows for analyzing experimental data. By simulating a scattering experiment on a neutron instrument, using vnf for example, we can obtain a set of simulated data with the same type of neutron-weighting that is present in experimental data. The neutron weight corrections can then be performed on these simulated data, and the "corrected spectra" can then be used to calculate partition functions for phonon thermodynamics. By varying the algorithms and parameters for these corrections, we can obtain a range of thermodynamic quantities, from which we can estimate bounds on the uncertainties.

3.2. Uncertainty Quantification

An uncertainty quantification (UQ) effort will develop a rigorous framework for the propagation and quantification of information and uncertainties in the analysis of scattering data. A traditional

way to deal with the missing information has been to generate (possibly probabilistic) models that are compatible with known aspects of the system and its governing equations. A key problem with this approach is that the space of such models typically has infinite dimensions while individual predictions are limited to a single element in that space. One should also be cautious with a direct/traditional application of Bayesian learning methods not only because the choice of priors involves some degree of arbitrariness that is incompatible with the estimation of rare events (such as failures) but also because the Bayesian method may fail to converge or may converge towards the wrong solution if the underlying probability mechanism allows an infinite number of possible outcomes [70, 71].

We will develop a method for computing optimal (upper and lower) bounds on quantities of interest with respect to the available information: *these bounds are optimal best case and worst case predictions on quantities of interest*. These correspond to infinite dimensional optimization problems over spaces of functions and measures that can be reduced to finite dimensional, and are computationally tractable. As demonstrated in the *Optimal Uncertainty Quantification* (OUQ) framework [24]: just as a linear program finds its extreme value at the extremal points of a convex domain in $\mathbb{R}^n$, OUQ problems reduce to searches over finite-dimensional families of extremal scenarios. Importantly, our proposed approach does not implicitly impose inappropriate assumptions (nor does it repudiate relevant information) by favoring a single model compatible with the given information [24]. The extremisers of these problems will identify the *key characteristic descriptors* of the system of interest.

In a parallel effort, we will investigate convergence conditions of a generalization of the Bayesian framework by allowing incomplete information on priors and likelihoods. To achieve this goal we will generalize the OUQ framework to the Bayesian framework based on the observation that the resulting optimization problems are infinite dimensional fractional problems (over measures) that are equivalent to infinite dimensional linear problems. The OUQ approach is neither Bayesian nor frequentist, but it allows for a hybrid Bayesian/frequentist approach (OUQ is a method of utilizing information, assessing its flow, and making intelligent, defensible decisions; this information may come in the forms of priors). In this work we compute optimal bounds on posteriors when priors are incompletely specified (through finite-dimensional marginals for instance).

3.3. Supported Software Packages

Even if the full SIXNS Institute is funded, we do not expect the main effort to be developing new software packages. Instead, the focus will be adapting and modifying packages from computational materials science that address structure and dynamics at the atomic level. In the Conceptualization Phase we propose to finalize the selection of a few software packages, and do some prototyping tests of how they would be integrated into workflows for computational scattering science. Those packages for which we have extensive experience include the following.

FEFF is an automated program for ab initio multiple scattering calculations of X-ray Absorption Fine Structure (XAFS), X-ray Absorption Near-Edge Structure (XANES) and various other spectra for clusters of atoms [72]. The code yields scattering amplitudes and phases used in many modern XAFS analysis codes, as well as various other properties. The latest version FEFF 9 includes improved treatments of many-body effects, inelastic losses and Debye-Waller factors and comes with an efficient Java-based GUI JFEFF.

Quantum ESPRESSO is an integrated suite of computer codes for calculating the electronic structure of materials and their properties with density-functional theory [73]. Several of its main modules were integrated into DANSE services and made available as web services for calculating the dynamics of atoms in crystalline solids.

GULP is a program for performing a variety of types of simulation on materials, including lattice dynamics and classical molecular dynamics [74, 75]. It offers a wide variety of potentials for interatomic interactions that give useful results for inelastic scattering experiments.

Mystic is a software package that provides algorithms and tools to solve optimization problems [76]. All optimization algorithms included in mystic provide workflow at the fitting layer, not just access to the algorithms as function calls.

DAKOTA (Design Analysis Kit for Optimization and Terascale Applications) is a toolkit developed and supported by the Sandia National Labs [77]. DAKOTA provides a flexible interface between analysis codes and iterative systems analysis methods, containing algorithms for optimization, uncertainty quantification, parameter estimation, and sensitivity/variance analysis.

SrFit is a modular optimization framework for structure refinement. It is written in Python and is currently fully operational, but in a prototype form and undergoing alpha testing. Different optimization algorithms can be used, for example scipy.otimize, or Mystic (see above) when parallel execution is desired. It is highly flexible allowing different model representations to be refined to different datasets and theoretical outputs with powerful control of constraints and restraints.

SrReal is a highly optimized pair-iterator engine that does the function calculations in SrFit for any pair-wise quantity. It can calculate the PDF, but also total energy in a pair-potential model or bond valence sums, for example. Kinematical scattering is limited to two-body interactions, so it can be extended to calculate any kinematical scattering function.

These will be discussed during the Conceptualization Phase, during which there should be prototype adaptations of some of them to demonstrate their capabilities to the user community. Other packages that will be brought for community discussion are VASP (Vienna Ab-Initio Simulation Package, a commercial but widely-used DFT code) [78], Casino (a quantum Monte-Carlo code for many-body electronic structure calculations) [79], ABINIT (a general purpose plane-wave pseudopotential electronic structure code) [80], AI2X (a set of ABINIT plug-ins developed by the Rehr group that extend the capabilities of and drive ABINIT. These include GW/Bethe-Salpeter codes AI2NBSE [81] and OCEAN [82] for optical and X-ray response, and DMDW an extension for calculating Debye-Waller factors [66]). Some software packages for diffraction and structure analysis that should be discussed are GSAS2 (a Python package being developed for Rietveld analysis), and a reverse Monte Carlo method package, of which several are available.

3.4. Development of New Software

The development of new software tools could occur on a 3-4 year time scale. It is not possible to develop any major applications during the Conceptualization Phase of SIXNS, but it is important for a full Institute to have an early experience of designing and building at least one new software package. There are several possible candidates, but selection should not be made today.

Optimal Estimation of a Quantity of Interest. The description and propagation of information at various levels of complexity brings non-trivial questions at the interface between computer science and optimization theory, probability theory and statistics. Consider, as a practical example, the estimation of the distance between two atoms from noisy measurements of the Fourier coefficients of the inter-atomic distance of a large group of atoms. If the distribution of the measurement is known, in absence of systemic and modeling errors it can be shown that the Bayesian estimate is optimal if it is derived from the prior associated with the true distribution of the measurement noise. If the distribution of the measurement noise is unknown, the proposed methodology will quantify the accuracy of the Bayesian posterior associated with the available (possibly arbitrary) prior. This quantification requires solving optimization problems in infinite-dimensional spaces of measures via reduction to finite-dimensional computationally tractable problems [24]. The proposed framework also allows to compute error bounds on the estimation of the quantity of interest in presence of epistemic or systemic uncertainties such as: incompleteness or inaccuracy of the model (the relation between the distance between two atoms and the measurements may not be a linear Fourier transform), incompleteness in the measured data (that data may not be sufficient to recover the distance between two atoms), arbitrariness in the choice of priors and unknown unknowns. By minimizing these error bounds over all possible statistical estimators, the proposed framework will

produce not only robust but also optimal estimates of the distance between two-atoms. This can be done within the Bayesian framework through the computation of optimal priors in relation to the available information. The dependence of these bounds on various model parameters constitute a form of non-linear sensitivity analysis that identifies the impact of each parameter on the accuracy of the estimation (this impact may be null in presence of incomplete information).

Interfaces Between Materials Computations and Scattering Cross-Sections. Many computational materials science software packages do not generate outputs that are directly usable for simulating results of scattering experiments. The methods for converting materials structure and dynamics into measurable cross-sections are well-known, and sometimes implemented in software. In some cases development is needed, however.

Refactoring Web Applications for More Modular Workflows. In a first offering of web services for computational science, we expect that standalone software packages would be provided intact, with appropriate changes to their interfaces. As workflows are developed that incorporate these software packages, it may prove important to allow users to rearrange the workflows. This will require some design effort, and some balance between effort and scientific benefit.

Tools for Visualizing Higher-Dimensional Datasets. We can build on our experience of providing cloud-based data management portals, and we are in a good position to generate visualization portals through web-based toolsets and data-visualization widgets. Example of such applications are: iVu [83], the VNF atomic structure viewer [84] and a phonon dispersion viewer [85].

GUI Development. One of the most important needed developments is the development of an enhanced GUI for handling the modeling and analysis. Such GUI control both simplifies and standardizes analysis techniques, and hence leads to a more definitive interpretation of experiment. Moreover such GUI control can enable a seamless modularization of the processes and hence facilitates combining disparate tools. One possible candidate is the Java-based JFEFF GUI already developed for the FEFF9 code [72]. Interfacing with existing EXAFS fitting tools like IFEFFIT may also be desirable [86]. We have also developed Luban [87], a generic user interface specification "language," that can serve as an integration tool for building cloud-enabled interfaces.

3.5. Scientific Support for Users

A primary mission of the SIXNS Institute will be to support users, especially new users, in their efforts to better utilize computational science for understanding their scattering experiments. We propose to begin this effort by collaborations, where computational scientists will play roles roughly analogous to instrument scientists at beamlines at national user facilities. By teaming together on mutually-satisfying common scientific goals, the computational scattering scientist and the experimental scattering scientist will ensure that the scientific workflow is relevant to producing new science and science publications. This arrangement would benefit junior computational scientists, as discussed in section 4.5. Other issues that will be addressed by an Institute, and require planning in the Conceptualization Phase include the following.

Adapting software packages to work as web services requires "componentizing" the software package with standard I/O, life cycle control, and error handling features. We propose to do this with the Python language, which will then allow a scientist to flexibly adapt the componentized package into a custom workflow.[3]

With a service for many users comes the opportunity to share information and input files for similar types of computational workflows. A relational database can manage metadata and pointers to output files from simulations, for example. This requires a data object model (DOM) that is reasonably stable over time, and selecting this DOM is an early requirement for the SIXNS Institute.

[3] A Python framework for supporting software components and building applications and web services is a reasonable choice in 2011. This situation needs to be monitored, however, and it is possible that better software architectures will evolve over the next five years, especially for web services.

Documentation allows software to be used effectively, and is critical for deciding if the software is appropriate for the problem at hand. Good documentation requires an effort comparable to that of writing the code itself. For software intended for community service, documentation is essential. The culture of software project management can elevate the documentation effort, bringing it towards the level of scientific writing. Using documentation as a key part of release reviews, and even design reviews (at the level of UML diagrams), can demonstrate the importance of documentation to the developers.

A user service requires a policy on feature requests and bug fixes. Feature requests can broaden the base of science that is served by the Institute, but as mentioned at the beginning of this section, feature requests must be managed carefully to avoid dissipation of developer resources. We propose to use the Conceptualization Phase to work with users of our existing software packages to identify the best policies that could later be formalized by the Institute, if appropriate.

3.6. Software Maintenance, Availability, Sustainability

"All that is human must retrograde if it does not advance."[88] Software is a human endeavor, so effort is needed to keep it working and available to the community of scattering scientists. The question of long-term sustainability is challenging because of our lack of experience with it in the scattering community. To date, software packages for analysis of scattering data were developed and maintained by individuals, and the availability of a package has followed the interest and activity of the developer. In a few cases other developers picked up where the original developer left off, but this was usually facilitated by the design of the software as a single application with single entry points. Such an approach cannot be expected for an architecture with modular components. As operating systems are updated, as new packages replace old, as bugs are uncovered, and as new usage patterns emerge, there will be problems with any static base of code. Fortunately, many problems will be minor ones. Although these can stop the software from working, making it unavailable to users, a competent staff can address these problems efficiently.

Sometimes the support will require adding new features to established scientific workflows to accommodate the uniqueness of a new material or experiment. Authorizing and managing software projects, even "small" ones, requires a process, so that the work is done efficiently and Institute resources are used effectively. This is the domain of project management, which can become restrictive for small efforts. Nevertheless, a clear plan, scope, and a risk watch list may be required for all feature requests that require substantial code development.

Without user support, good software cannot be used to its full potential. Disciplined human talent is the essential resource for software development and user support, and an Institute will require an ongoing commitment to retain talented persons. To keep excellent staff engaged with DANSE, there must be opportunities to advance both their science and their career, as discussed in Section 4.5.

The SIXNS Institute would itself maintain only a limited amount of hardware, largely for the development and testing of scientific workflows. Instead, SIXNS would negotiate on the behalf of users for access to computers, software, data, and networks, and develop working relationships with grid service providers, both public and private. It would also set rules for resource allocations to users. Some computationally intensive workflows could benefit from access to national level resources. The Institute will facilitate access to resources such as those offered by INCITE and XSEDE [89, 90]. INCITE provides access to DOE's Leadership Computing Facilities. For 2012, nearly 1.7 billion hours have been allocated to 60 projects. The DOE facilities are for well defined problems that are computationally challenging. The NSF supported Extreme Science and Engineering Discovery Environment (XSEDE), which supersedes the TeraGrid, is another option. In particular, SIXNS will develop an XSEDE "Gateway" that supports common workflows. This will allow users to gain quick access to large computer resources and the tools to use them. Some software [e.g., FEFF9 and AI2X] will also be available on virtual compute resources such as the high performance scientific

cloud platform being developed for the Amazon EC2. This resource will make many of these tools available to scientists who lack expertise or access to high performance computing facilities [42, 43].

The SIXNS Institute will need some software and hardware to do its work, including development tools, software standards, user interfaces, some simulation components, visualization components, a developers' repository, and an automated regression test system with some form of build system. Some of these are described in the Data Management Plan. It will administer a central server and some other hardware for development and testing. It is expected that the Institute will offer software components to the community that vary in maturity and validation. Some will be commercial packages, and others will be evolving as developers or scientists work on them. The SIXNS Governance Board will oversee the policies, license management, code quality, and authorization for access to these different codes.

4. Future: A Sustainable Software Innovation Institute and Its Management

4.1. Coordination with Users 2012-2014

A substantial budget in the Conceptualization Phase is proposed for a series of workshops and meetings to coordinate with the community of scattering scientists. Small workshops will focus on working with scientists to design and test workflows. The first workshop(s) will focus on determining how this is best accomplished. For the larger meetings, agenda items will include:

Workflows - What science opportunities are achievable, and offer the most impact for the least effort?

Computational Engines - Which materials simulation codes and optimizers should be supported? Are there issues with licenses or export control?

Visualization - What are the visualization needs; what is the state of the current tools; can we partner with other groups (such as the VATE project at ISIS) to develop the tools?

Data Management - What is the best interface to data at national user facilities; what privacy is expected for derived data products; should there be a repository for computational results; should the Institute provide provenance for computed results; what should be in the software bundles?

Computer Resources - What resources should the Institute provide; how can it help users access other resources such as XSEDE and INCITE; what should be included in the gateway?

Governance - How should the Institute be governed; how is the user community defined; who gets to vote?

The first meeting will engage the original participants who helped develop the strategic plan in the Report *Computational Scattering Science 2010* [22], plus another 30-40 persons. This meeting would be 2 full days at Caltech. Its focus will be on planning the next two years, with discussion and refinement of the topics in this proposal.

Another early workshop will be an educational one for obtaining user feedback on software developed under the DANSE and FEFF9 projects. It will be the responsibilities of the P.I. and Co-P.I.s to select 2-3 scientists each who have research problems appropriate for building workflows with our existing computational components. These interactions with users will help set priorities and guide the workflow development. This workshop would be 2 full days at the Univ. Washington.

We will make immediate contact with the executive committees of the user groups such of national user facilities such as the SHUG and APSUO to discuss how to include a software meeting or workshop at their annual meetings. These organizations control the agendas of their meetings, of course, but we will suggest a symposium on computational methods at their annual meetings (for which there are precedents). We could also plan a satellite meeting with tutorials and discussion. These activities would be in the first or second year, depending on coordination with the user groups.

In the first year we will have a four day meeting at Caltech that focuses on workflow prioritization, together with tutorial sessions. This will be a prototype for subsequent meetings on computa-

tional scattering science, where we will balance science presentations, management discussions, and separate days for focused tutorials for scientific collaborators.

In the second year of the Conceptualization Phase, we will organize two more combined meetings with both workshop and tutorial sessions. These will be organized by Columbia University and by Caltech.

4.2. Coordination with DOE Facilities

The DOE Office of Science, Office of Advanced Scientific Computing Research and Office of Basic Energy Sciences held the 2011 ASCR/BES Data Workshop in October 2011 to address the issue of advanced scientific computing and workflows from data to publication at national user facilities. The workshop report is still being written, but it will encourage DOE to increase its efforts in computational scattering science. While this DOE effort is getting organized, we will work with the scientific software development groups at the APS, SNS, and NSLS-2 to coordinate technical interfaces between their software for data reduction and our advanced analysis software. We have had experience with this interface with the SNS, and we have also had extensive contacts with the ISIS facility in the U.K., who are working with the SNS on the Mantid project for data reduction and visualization. We will also coordinate with parallel efforts in the E.U., e.g., through the European Theoretical Spectroscopy Facility, in which Rehr is an active participant.

4.3. Governance 2014-2019

Neutron and X-ray scattering research has always included a diversity of stakeholders, and the Institute must accommodate this reality in its governance. A possible governance would have a Board that plays a role analogous to a corporate board in a company. A chief executive officer will report to this board, for example. The first Governance Board could comprise 1) The five investigators (PI, Co-PI) of the SIXNS proposal, 2) One representative from each of the user facilities, especially the APS, SNS, and NSLS-2, and 3) Two elected members from the user community, likely chosen after the Execution Phase is well underway.

The Governance Board should have an advisory role during the Conceptualization Phase. The transition to full governance will be presented in the Project Execution Plan, which will be in draft form when the full S2I2 proposal is submitted. Completing a transition plan from P.I. management to Governance Board control will be a major effort during the first budget period of the Execution Phase.

4.4. Metrics of Success and Risk Mitigation

A successful SIXNS Institute will increase the number and impact of scientific discoveries in X-ray and neutron scattering. Measuring this is a challenge. Users, publications, and downloads are likely metrics, but methods of counting need definition during the Conceptualization Phase. They need to be chosen with care so they give long-term information about changes in the performance of an Institute.

There is a low risk for building the main software packages, modularizing them, and deploying them as web services. Such work has been done extensively by Billinge, Fultz, and Rehr, and many key modules are fully developed, tested, documented, and have been modified with input from user experience. We will assess the tools for developing web services that are emerging today, perhaps with a framework for object-oriented programming in Python that allows bindings to C++ code because this is the approach being used by the user facilities SNS, ISIS, LCLS, and was the approach taken by DANSE. If there is no appropriate alternative, we will use the Opal framework for web services, with the Luban tools for user interface development as has been done for deploying the DANSE software as web services. Although these tools are not complete, they are robust and are being considered seriously by the scientific Python community as platform-independent tools for building web services.

The risk of user acceptance of the software products is low; several of the main packages are already in use by members of the community. Improving access to these packages, making them easier to use, and allowing them to be used in a greater range of scientific workflows should expand their use by the scattering community, who will be engaged in these decisions.

4.5. Workforce Development

The mission of the SIXNS Institute will include two important directions for workforce development. First will be the university training of graduate students and postdocs. Many students in science have interests in computing, and all fields of science now offer in computational science opportunities at the graduate level. Largely as an outgrowth of computational science research projects, Caltech offers a subject minor in Computational Science and Engineering (CSE) to graduate students in all Ph.D. programs. At the postdoctoral level, many experimental research groups value individuals who can use the tools of modern computational materials science. The SIXNS Institute can offer excellent opportunities for junior computational scientists to do new science by collaborating with experimental groups. There will be a substantial expectation for these junior scientists to publish new science that will advance their careers. From our experience, most long-term career paths are through the established academic fields of science like chemistry, physics, and materials science. Computational science is setting new directions in these fields.

Another opportunity for workforce development is science education. With good design, such as separating user interfaces from computing engines, software can fulfill the needs of both research and education. With a user interface appropriate for the level of the student, software can help introduce concepts of X-ray, neutron, and electron scattering to undergraduate and graduate students, and scientists in other fields. Deployment of software as a web service is particularly convenient for course instructors, freeing them from issues of installation, maintenance, and resource allocation. Web-based tutorials can present concepts in scattering that are part of university courses in innovative ways, especially if SIXNS makes an investment in new visualization tools. The teaching of concepts such as dynamical diffraction would benefit from simulations that supplement the mathematical textbook approaches used today. Simulations could also show graduate students and scientists if scattering experiments are appropriate for their own research. There is a considerable overlap of requirements for user friendly but advanced software for education, and software for advanced analysis by scattering experts.

4.6. Diversity Plan

By bringing together scattering science, computational science, cyberinfrastructure, and modern software engineering, the SIXNS Institute will have high national and international visibility. It therefore would have responsibilities beyond research in scattering and materials sciences, including responsibility to recognize and reflect the diversity of people who make up American society. SIXNS must reach out to people of diverse ethnic, racial, economic and gender groups.

A rich source of statistical information on women, under-represented minorities, and persons with disabilities in the sciences and engineering has been compiled and maintained by the NSF [91]. Trends in computer science and physical sciences were evaluated in depth. In addition to the need to reach out to under-represented minority groups, the NSF statistics indicate a need for special attention to women in computational science. The percentage of female graduate students in computer science and in physical sciences is nearly the lowest of all science and engineering fields. For undergraduate students the NSF report shows that the gap in degrees earned by male and female students in computer science has been growing. The fraction of women with bachelors degrees in computer sciences dropped from 1985 to 2004 to 2008 from 37 to 25 to 18 percent [91].[4] In the Conceptualization Phase, and later with SIXNS, we propose to offer part-time employment opportunities to undergraduate women in the physical sciences and engineering. The experience of

[4] Curiously, the fraction of women earning Ph.D. degrees in computer sciences has been increasing [92].

working with an active computational science team can help develop technical skills, but it also gives perspective on how large projects are developed, how teams are formed, and how a national effort can be organized. A young woman can see more clearly the big picture of how her work makes a difference, such as by empowering others with new tools and capabilities.

Caltech's Minority Undergraduate Research Fellowship (MURF) program provides support for talented undergraduates to spend a summer working in a research laboratory on the Caltech campus. The MURF program is focused on underrepresented students (such as African American, Latino, and Native American) in science and engineering, helping to make Caltech's programs more visible to students not traditionally exposed to Caltech. For several years, Fultz has hosted one or two high school interns from the Institute for Educational Advancement. These talented youths have integrated well into software development projects, which were unique experiences for them. An important component of these high-school and undergraduate intern experiences is the co-mentorship by a graduate student or postdoctoral fellow, who offers a more informal and sympathetic source of support and advice that can be offered by a faculty member alone.

In the hiring of postdoctoral fellows and professional staff, efforts will be made to identify women and under-represented minority candidates, and these efforts will be documented. This will include: 1) efforts to announce the positions to a more diverse pool of potential applicants. The National Society of Black Engineers offers job posting services, for example, as does the Society of Hispanic Engineers. 2) Statistical information on the total number of women and under-represented minority applicants. 3) Information on number of qualified women and under-represented minority applicants who were identified, interviewed, and hired.

4.7. Deliverables from the Conceptualization Phase 2012-2014

Full proposal for an Execution Phase. A proposal with supporting documentation for a Sustainable Software Innovation Institute for Advanced Analysis of for an X-Ray and Neutron Scattering Data (SIXNS)

Strategic Plan. An early version of a strategic plan is the NSF-DOE workshop report *Computational Scattering Science 2010* [22]. Many community-based ideas are in this report, some suitable for a Sustainable Software Innovation Institute. A series of five meetings will be held over two years to address the different community expectations for the Institute, especially the goal state and a path to achieve it.

Workshop Findings Reports. These will be prepared to summarize community discussions on topics such as governance, data policy, and workflow priorities. This text may become part of the Strategic Plan.

Project Execution Plan. The PEP will lay out the S2I2 governance and its transition to a user organization, management structure, project controls, release management and quality assurance. (This draft would be unofficial, requiring future approval by the NSF and other stakeholders.)

References

1. E. Ustundag, R. A. Karnesky, M. R. Daymond, I. C. Noyan, "Dynamical Diffraction Peak Splitting in Time-of-Flight Neutron Diffraction," Applied Physics Letters **89**, 233515 (2006).
2. J. B. Keith, H. Wang, B. Fultz and J. Lewis, "Ab-Intio Free Energy of Vacancy Formation and Mass-action Kinetics in Vis-active TiO_2," Journal of Physics: Condensed Matter **20**, 022202 (2008).
3. J. B. Keith, J. R. Fennick, C. E. Junkermeier, E. R. Nelson and J. P. Lewis, "A Web-Deployed Interface for Performing Ab-Initio Molecular Dynamics, Optimization, and Electronic Structure in Fireball," Computer Physics Communications **180**, 418 (2009).
4. C. L. Farrow, P. Juhas, J. W. Liu, D. Bryndin, E. S. Bozin, J. Bloch, Th. Proffen and S. J. L. Billinge, "PDFfit2 and PDFgui: Computer programs for studying nanostructure in crystals," Journal of Physics: Condensed Matter **19**, 335219 (2007).
5. A. D. Christianson, M. D. Lumsden, O. Delaire, M. B. Stone, D. A. Abernathy, M. A. McGuire, A. S. Sefat, E. D. Mun, P. C. Canfield, J. Y. Y. Lin, M. S. Lucas, M. Kresch, J. B. Keith, B. Fultz, E. A. Boremychkin and R. J. McQueeney, "Phonon Density of States of $LaFeAsO_{1-x}F_x$," Physical Review Letters **101**, 157004 (2008).
6. R. B. McClurg and J. B. Keith, "Molecular crystal global phase diagrams II: Sufficient parameter space determination," Acta Crystollographica Section A **66**, 38 (2010).
7. J. B. Keith and R. B. McClurg, "Molecular crystal global phase diagrams III: Sufficient parameter space determination," Acta Crystollographica Section A **66**, 50 (2010).
8. O. Delaire, A. F. May, M. A. McGuire, W. D. Porter, M. S. Lucas, M. B. Stone, D. L. Abernathy, V. A. Ravi, S. A. Firdosy and G. J. Snyder, "Phonon density of states and heat capacity of $La_{3-x}Te_4$," Physical Review B **80**, 184302 (2009).
9. P. A. Kienzle, N. Patel and M. McKerns, "Parallel Kernels: An Architecture for Distributed Parallel Computing," Proceedings of the 8th Python in Science Conference **36** (2009). http://conference.scipy.org/proceedings/SciPy2009/
10. C. W. Li, M. M. McKerns and B. Fultz, "Raman spectroscopy study of phonon anharmonicity of hafnia at elevated temperatures," Physical Review B **80**, 054304 (2009).
11. T. J. Sullivan, U. Topcu, M. McKerns and H. Owhadi, "Uncertainty quantification via codimension-one partitioning," International Journal for Numerical Methods in Engineering **85**, 1499 (2011).
12. C. W. Li, M. M. McKerns, and B. Fultz, "A Raman Spectrometry Study of Phonon Anharmonicity of Zirconia at Elevated Temperatures," Journal of the American Ceramic Society **94**, 125 (2011).
13. X. L. Tang, C. W. Li and B. Fultz, "Anharmonicity-induced phonon broadening in aluminum at high temperatures," Physical Review B **82**, 184301 (2010).
14. M. S. Lucas, J. A. Munoz, O. Delaire, N. D. Markovskiy, M. B. Stone, D. L. Abernathy, I. Halevy, L. Mauger, J. B. Keith, M. L. Winterrose, Y. M. Xiao, M. Lerche and B. Fultz, "Effects of composition, temperature, and magnetism on phonons in bcc Fe-V alloys," Physical Review B **82**, 144306 (2010).
15. M. S. Lucas, J. A. Munoz, L. Mauger, C. W. Li, A. O. Sheets, Z. Turgut, J. Horwath, D. L. Abernathy, M .B. Stone, O. Delaire, Y. M. Xiao and B. Fultz, "Effects of chemical composition and B2 order on phonons in bcc Fe-Co alloys," Journal of Applied Physics **108**, 023519 (2010).
16. P. G. Evans and S. J. L. Billinge, "Advances in Scattering Probes for Materials," MRS Bulletin **35**, 495 (2010).
17. O. Delaire, "Studies of high-temperature electron-phonon interactions with inelastic neutron scattering and first-principles computations," Applied Physics A-Materials Science & Processing **99**, 523 (2010).
18. S. J. L. Billinge, "The nanostructure problem," Physics **3**, 25 (2010).

19. M. S. Lucas, "ARCS shows effect of ferromagnetism on phonons in FeV alloys," Oak Ridge National Lab Report, (2010). http://neutrons.ornl.gov/research/highlights/ARCS/ARCSLucasFerromagnetism_2010.pdf
20. X. Tang and B. Fultz "A First-principles Study of Phonon Linewidths in Noble Metals," Physical Review B **84**, 054303 (2011).
21. N. D. Markovskiy, J. A. Munoz, M. S. Lucas, C. W. Li, O. Delaire, M. B. Stone and D. L. Abernathy, "Non-harmonic Phonons in MgB_2 at Elevated Temperatures," Physical Review B **83**, 174301 (2011).
22. B. Fultz, K. W. Herwig and G. G. Long, "Computational Scattering Science" (2010). `http://www.its.caltech.edu/~matsci/Publish/CompScatWkshp_2010.html`
23. A. A. Kidane, A. Lasgari, B. Li, M. McKerns, M. Ortiz, G. Ravichandran, M. Stalzer and T. J. Sullivan, "Rigorous Model-based Uncertainty Quantification with Application to Terminal Ballistics: Systems with Controllable Inputs and Small Scatter," Reliability Engineering & System Safety, Accepted (2010).
24. H. Owhadi, C. Scovel, T. Sullivan, M. McKerns and M. Ortiz, "Optimal uncertainty quantification," SIAM Review, under review 2010. arXiv:1009.0679.
25. P. Juhas, L. Granlund, P. M. Duxbury, W. F. Punch and S. J. L. Billinge, "The Liga algorithm for ab initio determination of nanostructure," Acta Crystallography A **64**, 631 (2008).
26. J. A. Munoz, M. S. Lucas, O. Delaire, M. L . Winterrose, L. Mauger, C. W. Li, A. O. Sheets, M. B. Stone, D. L. Abernathy, Y. M. Xiao, P. Chow and B. Fultz, "Positive Vibrational Entropy of Chemical Ordering in FeV," Physical Review Letters **19**, 115501 (2011).
27. R. K. Dumas, Y. Y. Fang, B. J. Kirby, C. L. Zha, V. Bonanni, J. Nogues and J. Akerman "Probing vertically graded anisotropy in FePtCu films," Physical Review B **84**, 054434 (2011).
28. G. M. Newbloom, F. S. Kim, S .A. Jenekhe and D. C. Pozzo, "Mesoscale Morphology and Charge Transport in Colloidal Networks of Poly(3-hexylthiophene)," Macromolecules **44**, 3801 (2011).
29. M. L. Winterrose, L. Mauger, I. Halevy, A. F. Yue, M. S. Lucas, J. A. Munoz, H. Tan, Y. Xiao, P. Chow, W. Sturhahn, T.S. Toellner, E. E. Alp and B. Fultz, "Dynamics of iron atoms across the pressure-induced Invar transition in Pd_3Fe," Physical Review B **83**, 134304 (2011).
30. K. M. Weigandt, L. Porcarand D. C. Pozzo, "In situ neutron scattering study of structural transitions in fibrin networks under shear deformation," Soft Matter **7**, 9992 (2011).
31. J. R. Fennick, J. B. Keith, R. H. Leonard, T. H. Truong and J. P. Lewis, "A Cyberenvironment for Crystallography and Materials Science and an Integrated User Interface to the Crystallography Open Database and Predicted Crystallography Open Database," Applied Crystallography **41**, 471 (2008) .
32. C. Li, X. Tang, J. A. Munoz, J. B. Keith, S. J. Tracy, D. L. Abernathy and B. Fultz, "The Structural Relationship Between Negative Thermal Expansion and Quaratic Anharmonicity of Cubic ScF_3," Physical Review Letters **107**, 195504 (2011).
33. B. Fultz, T. Kelley, J. Y. Y. Lin, J.-D. Lee, O. Delaire, M. Kresch, M. McKerns, M. Aivazis, *Experimental Inelastic Neutron Scattering* (2010). Graduate-level e-textbook. http://docs.danse.us/DrChops/ExperimentalInelasticNeutronScattering.pdf
34. B. Fultz and J. J. Hoyt, "Phase Equilibria and Phase Transformations," in *Alloy Physics*, Editor: Wolfgang Pfeiler (Wiley-VCH, Weinheim, 2007) ISBN-10: 3-527-31321-4.
35. B. Fultz and J. M. Howe, *Transmission Electron Microscopy and Diffractometry of Materials* 3rd ed. (Springer-Verlag, Heidelberg, 2007) ISBN: 978-3-540-73885-5.
36. C. L. Farrow, P. Juhas, J. Liu, D. Bryndin, E. S. Bozin, J. Bloch, Th. Proffen and S. J. L. Billinge, "PDFfit2 and PDFgui: Computer programs for studying nanostructure in crystals," Journal of Physics: Condensed Matter **19**, 335219 (2007).
37. P. Juhas, L. Granlund, P. M. Duxbury, W. F. Punch and S. J. L. Billinge , "The Liga algorithm for ab initio determination of nanostructure," Acta Crystallogr. A **64**, 631 (2008).

38. V. A. Blagojevic, J. P. Carlo, L. E. Brus, M. L. Steigerwald, Y. J. Uemura, S. J. L. Billinge, W. Zhou, P. W. Stephens, A. A. Aczel and G. Luke, "Magnetic phase transition in V_2O_3 nanocrystals," Physical Review B,**82** 094453 (2010).
39. P. Juhas, L. Granlund, R. Gujarathi, P. M. Duxbury and S. J. L. Billinge, "Crystal structure solution from experimentally determined atomic pair distribution functions," Journal of Applied Crystallography **43**, 623 (2010).
40. P. Tian, W. Zhou, C. L. Farrow, P. Juhas and S. J. L. Billinge, "SrRietveld: A program for automating Rietveld refinements for high throughput studies," cond-mat ntrl-sci arXiv:1006 0435, (2010).
41. P. Tian and S. J. L. Billinge, "Testing different methods for estimating uncertainties on Rietveld refined parameters using SrRietveld," Z. Kristallography **226**, 898 (2011).
42. J. J. Rehr, F. Vila, J. P. Gardner, L. Svec and M. Prange, "Scientific Computing in the Cloud," CiSE **12**, 34 (2010).
43. K. Jorissen, F. D. Vila, J. J. Rehr, "A high performance scientific cloud computing environment for materials simulations," arXiv:1110.0543.
44. M. Simons, S. E. Minson, A. Sladen, F. Ortega, J. L. Jiang, S.E. Owen, L. S. Meng, J. P. Ampuero, S.J. Wei, R. S. Chu, D. V. Helmberger, H. Kanamori, E. Hetland, A. W. Moore, F. H. Webb, "The 2011 Magnitude 9.0 Tohoku-Oki Earthquake, Science **332**, 1421 (2011). doi:10.1126/science.1206731.
45. J. L. Beck and K. M. Zuev, Asymptotically Independent Markov Sampling: new MCMC scheme for Bayesian Inference, submitted to Statistics and Computing, http://arxiv.org/abs/1110.1880 (2011).
46. K. M. Zuev, J. L. Beck, S. K. Au, and L. S. Katafygiotis, "Bayesian postprocessor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions, Computers & Structures, accepted for publication, http://arxiv.org/abs/1110.3390 (2011).
47. Y. Huang, J.L. Beck, S. Wu and H. Li, "Robust Bayesian compressive sensing for signals in structural health monitoring: compression and reconstruction, to be submitted (2011).
48. L. Zhang, L. Berlyand, M. Fedorov and H. Owhadi, "Global Energy Matching Method for Atomistic to Continuum Modeling of Self-Assembling Biopolymer Aggregates," SIAM Multiscale Modeling & Simulation **8**, 1958 (2010).
49. M. Tao, H. Owhadi and J. E. Marsden, "Non-intrusive and structure preserving multiscale integration of stiff ODEs, SDEs and Hamiltonian systems with hidden slow dynamics via flow averaging," SIAM Multiscale Modeling & Simulation **8**, 1269 (2010).
50. B. Bayati, H. Owhadi and P. Koumoutsakos, "Cutoff Phenomenon in Accelerated Stochastic Simulations of Chemical Kinetics via Flow Averaging (FLAVOR-SSA)." Journal of Chemical Physics. **133**, Issue 24. (2010).
51. M. Tao, H. Owhadi and J. E. Marsden, "Symplectic, linearly-implicit and sta- ble integrators with applications to fast symplectic simulations of constrained dynamics," arXiv:1103.4645.
52. S. Ober-Blobaum, M. Tao, M. Cheng, H. Owhadi and J. E. Marsden, "Variational integrators for electric circuits" arXiv:1103.1859.
53. M. Tao, H. Owhadi and J. E. Marsden, "From efficient symplectic exponentiation of matrices to symplectic integration of high-dimensional Hamiltonian systems with slowly varying quadratic stiff potentials," Accepted for publication in Applied Mathematics Research Express. arXiv:1006.4659.
54. M. Tao, H. Owhadi and J. E. Marsden, "Space-time FLAVORS: finite difference, multisymplectic, and pseudospectral integrators for multiscale PDEs. Dynamics of Partial Differential Equations," **8**, (2011). arXiv:1104.0272.
55. A. Doostan and H. Owhadi, "A non-adapted sparse approximation of PDEs with stochastic inputs," Journal of Computational Physics **230**, 3015 (2011).

56. http://www.whitehouse.gov/blog/2011/06/24/materials-genome-initiative-renaissance-american-manufacturing
57. http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf
58. J. J. Rehr and R. C. Albers, "Theoretical approaches to x-ray absorption fine structure." Revs. Modern Physics **72**, 621 (2000).
59. H. J. Krappe and H. H. Rossner, "Error analysis of XAFS measurements," Phys. Rev. B **61**, 6596 (2000).
60. H. J. Krappe and H. H. Rossner, "Bayes-Turchin approach to x-ray absorption fine structure data analysis," Phys. Rev. B **66**, 184303 (2002).
61. H. J. Krappe and H. H. Rossner, "Bayesian approach to background subtraction for data from the extended x-ray-absorption fine structure," Phys. Rev. B **70**, 104102 (2004).
62. J. J. Rehr, K. Kozdon, J. Kas , H. J. Krappe and H. H. Rossner, "Bayes-Turchin approach to XAS analysis," J. Synchrotron Radiation **12**, 70 (2005).
63. H. Stragier, J. O. Cross, J. J. Rehr, L. B. Sorensen, C. E. Bouldin, J. C. Woicik, "Diffraction anomalous fine-structure – A new X-ray structural technique," Phys. Rev. Lett. bf 69, 3064 (1992).
64. J.A. Soininen, A. L. Ankudinov, and J. J. Rehr, "Inelastic Scattering from Core-electrons: a Multiple Scattering Approach," Phys. Rev. B **72**, 045136 (2005).
65. J. J. Kas, J. J. Rehr, J. A. Soininen, and P. Glatzel, "Real-space Green's function approach to resonant inelastic x-ray scattering," Phys. Rev. B **83**, 235114 (2011).
66. F. D. Vila, J. J. Rehr, H. H. Rossner, and H. J. Krappe, "Theoretical x-ray absorption Debye-Waller factors," Phys. Rev. B **76**, 014301 (2007).
67. http://www-958.ibm.com/software/data/cognos/manyeyes/
68. http://www.tableausoftware.com/
69. http://bimeanalytics.com/
70. P. Diaconis and D. Freedman, "On the consistency of Bayes estimates," Ann. Statist., **14**, 167 (1986). With a discussion and a rejoinder by the authors.
71. P. W. Diaconis and D. Freedman, "Consistency of Bayes estimates for nonparametric regression: normal theory," Bernoulli **4**, 411444 (1998).
72. John J. Rehr, Joshua J. Kas, Micah P. Prange, Adam P. Sorini, Yoshinari Takimoto, Fernando Vila, "Ab initio theory and calculations of X-ray spectra," Comptes Rendus Physique **10**, 548 (2009). http://leonardo.phys.washington.edu/feff/
73. http://www.quantum-espresso.org/
74. http://projects.ivec.org/gulp/
75. J.D. Gale, "GULP - a computer program for the symmetry adapted simulation of solids," JCS Faraday Trans. **93**, 629 (1997). J.D. Gale and A.L. Rohl, "The General Utility Lattice Program," Mol. Simul. **29**, 291 (2003).
76. http://docs.danse.us/mystic/current/mystic-module.html
77. http://dakota.sandia.gov
78. http://cms.mpi.univie.ac.at/vasp/vasp/
79. http://www.tcm.phy.cam.ac.uk/∼mdt26/casino2.html
80. http://www.abinit.org/
81. H. M. Lawler, J. J. Rehr, F. Vila, S. D. Dalosto, E. L. Shirley, Z. H. Levine, "Optical to UV spectra and birefringence of SiO_2 and TiO_2: First-principles calculations with excitonic effects," Phys. Rev. **78**, 205108 (2008).
82. J. Vinson, J. J. Rehr, J. J. Kas, E. L. Shirley, "Bethe-Salpeter equation calculations of core excitation spectra," Phys. Rev. B. **83**, 115106 (2011).
83. http://www.cacr.caltech.edu/ slombey/downloadme/psaap/iVu-pkg/
84. https://vnf.caltech.edu/vnf/1/atomicstructure.html

85. https://vnf.caltech.edu/vnf/1/phonons.html
86. M. Newville, "IFEFFIT: interactive EXAFS analysis and FEFF fitting," J. Synchrotron Rad. **8**, 322 (2001).
87. http://lubanui.org
88. Edward Gibbon, *The Decline and Fall of the Roman Empire* (1789).
89. http://www.doeleadershipcomputing.org
90. http://www.xsede.org
91. http://www.nsf.gov/statistics/wmpd/tables.cfm Table 5-1
92. http://www.nsf.gov/statistics/wmpd/tables.cfm Table 7-2