Report on

SIXNS Workshop III:

Theoretical software, analysis tools, and software integration for scattering science

Talaris Conference Center, University of Washington

January 17-18, 2014

John J. Rehr

Brent F. Fultz

Simon Billinge

Kevin Jorissen

Disclaimer

This report was prepared as an account of a workshop funded by the National Science Foundation under NSF Award Number SI2 OCI-1216716.

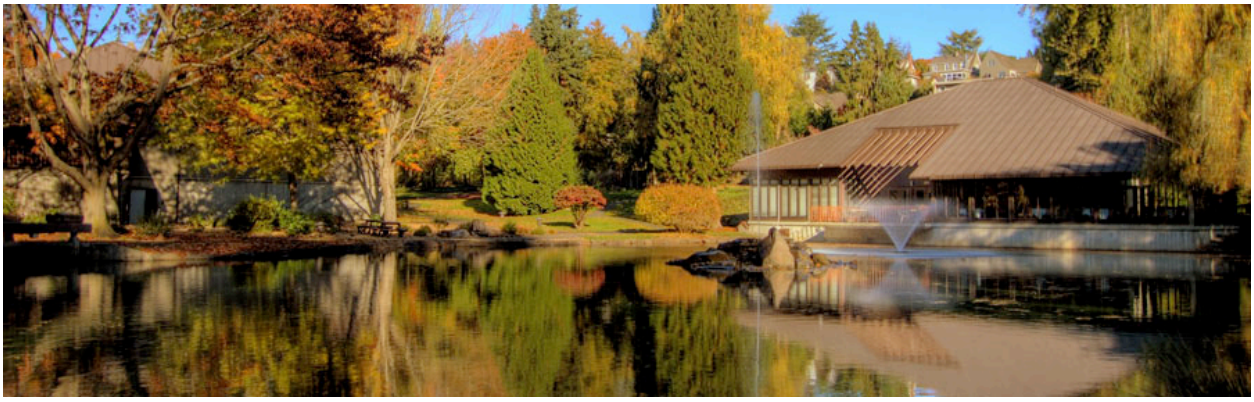
Table of Contents

1. Introduction

The 3rd SIXNS workshop, "Theoretical software, analysis tools, and software integration for scattering science," was held at the Talaris Conference Center at the University of Washington on Jan 17-18, 2014. These workshops are part of an NSF funded S2I2 Conceptualization Proposal to develop a sustainable Scientific Software Innovation Institute for Advanced Analysis of X-Ray and Neutron Scattering Data (SIXNS). The aim of this institute is to bring modeling and analysis of x-ray, electron, and neutron scattering experiments in alignment with the current and future needs of the scattering science community. This entails transforming existing codes into workflows, while modernizing them for current developments in computing and instrumentation.

The preceding workshops SIXNS I and SIXNS II focused respectively on scattering science workflows and data analysis, and described the benefits of the SIXNS Institute for the scattering science community. A complete report of SIXNS-I was published in June 2014 []. Its introduction clearly states the rationale for building a Sustainable Software Innovation Institute for X-Ray and Neutron Scattering, and may be of particular interest to readers of this report. The 3rd workshop SIXNS III focused on Theoretical codes, Analysis software, and Software integration tools that are needed in a sustainable SIXNS Institute. In particular, the workshop aimed to identify which theoretical and analysis tools are needed to model and interpret scattering experiments, and how to integrate these codes to efficient, user-friendly workflows, ready to produce innovative science in the hands of scattering scientists at national facilities and worldwide.

The SIXNS-III workshop brought together 50 experts for informal and creative discussions of these topics. The schedule was for a two-day meeting, all day Friday and Saturday till mid-afternoon on Jan 17-18. The workshop format combined plenary lectures with breakout sessions on a variety of topics. The body of this report consists of chapters summarizing the activity of each breakout group. The breakout groups were: Ground-state properties; Hardware and infrastructure; Neutron scattering theory codes and analysis tools; Excited states theory codes; Integration of software workflows; and Theoretical software institute.



2. Ground-State Properties

Group Leader: Gian-Marco Rignanese
Group members: Towfiq Ahmed, Michal Bajdich, Eric Bylaska, Wei Chen, Stephen Cramer, Tom Devereaux, Paolo Fornasini, Yan Li, Lars Pettersson, Fernando Vila, Tsu-Chien Weng

One of the goals of the institute is to provide its users with an access to theoretical codes for the calculation of ground-state properties as well as with some kind of support. As a minimal requirement, the institute website should include links to the original websites for the different codes supported. But, ideally, the institute should also provide some optimization, tutorials, machines for testing these codes, and training programs to teach these together with the related theories.

The institute should help to develop GUIs in order to facilitate the use of these codes in workflows. The main idea is to provide support to both inexperienced and expert users through the use of an established working environment. This will be a flexible, database-oriented evolutionary system guiding the user through the workflow on a high, task-oriented level. Initially common tasks will be identified and proper defaults and procedures will be established for the relevant calculators and tasks, including validation of the final results. The database orientation consists in recording the workflow and the outcome and adjusting the workflow and parameters to eliminate mistakes. The database will include validated reference results from earlier studies on other systems within the same class that can be used to validate future results by similarity (see discussion below). The database can also be used to set up new calculations within the same class and for the same properties. With time, the database will grow and provide more predictive support, facilitating the construction of new studies.

Furthermore, the institute should create synergistic collaborative programs in which theorists could help experimentalists to interpret their data, and experimentalists could also help theorists to improve the accuracy of their calculations, stimulate new calculations, and validate theorists' predictions.

A sound comparison of theory with experiment requires a reliable quantitative assessment of the accuracy of experimental data. Thus, another goal of the institute is to stimulate scientists (experimentalists and theorists alike) to be careful about uncertainty evaluation and the review panels of international facilities to support quantitative experiments.
The partial comparison of theory with experiment (e.g. the reproduction of the temperature dependence of Debye-Waller (DW) factors or of the 3rd cumulant in EXAFS) should be stimulated as a calibration tool in view of global approaches.
Multiple levels of quality checking need to be distinguished:
i) soft errors: at the machine level, some hardware errors (e.g. bit flip etc.) that do not make the code crash but give a wrong result;
ii) within a given approximation: errors may be generated by inappropriate input parameters;
iii) between approximations: errors may result from a given approximation (functional, single configuration, harmonic approximation, Born-Oppenheimer) ; (see 2013 report, fi
iv) between experiment and theory: experimental errors may propagate to the theoretical interpretation.

The ground-state properties targeted in the beginning will be:

A. Phonons, lattice dynamics in emerging materials

The calculation of phonon dispersion relations, phonon density of states, elastic properties and other higher-order derivative properties provides a direct way to check the validity of the many solid-state codes that calculate properties related to atomic and electronic structure as well as thermodynamics. These properties would show significant changes in lower dimensions and at nanoscale. In addition, as new energy-related materials like pnictide superconductors and skutterudites, multiferroics and other correlated electron systems are being discovered, a complete characterization is required.

With the availability of several inelastic x-ray scattering instruments around the world, it has become possible to measure phonon frequency and distribution with sub-meV resolution. Thus, solid-state codes that can calculate these properties are not only useful for planning the experiments, but also they can be tested rigorously against the experiments.

B. Mössbauer Spectroscopy

Mössbauer spectroscopy is a well-established technique that measures hyperfine interaction parameters that are related to solid-state effects. In particular, in modern era, synchrotron radiation is used to measure Mössbauer spectra in the time domain that enables measurements under extreme conditions of pressure and temperature. Complementing these measurements by calculations of isomer shifts, quadrupole splitting and magnetic hyperfine fields would provide a handle on spin and valence state, and local symmetry distortions that are hard to measure with other methods. The current codes that calculate the spectra are quite adequate but difficult to use. This would be a low-lying fruit where a GUI will enable many young scientists to exploit this field. Furthermore, prediction of hyperfine parameters would stimulate to research less understood subjects like spin transitions under pressure.

At later stage, other properties will also be added such as EXAFS, magnetic scattering (including improved stability for spin-polarized calculations, e.g. low spin/high spin transitions); X-rays and neutrons spectroscopy, ARPES, electron spectroscopy, as well as other advanced spectra (RIXS, Compton).

3. Hardware and Infrastructure

Group Leader: Simon Billinge
Group members: Dan Katz, Nick Draper, Jon Taylor, Karol Kowalski, Pavol Juhas, Kevin Jorissen, Bill Shelton, Stuart Campbell, Simon Billinge

Goals: An important goal is to organize the Hardware and Infrastructure of the SIXNS to focus on getting scientific impact out of the data. This broad target will require a large number of tasks:
-Compilation of available software
-Setting up the codes for error control, and establishing the reliability of results for non-expert users.
-Codes should fit inside a framework that allow for integration of different computation steps, e.g., standardized ground state codes in/out.
-Connecting databases between types of experiments
-Repository of workflows, data analysis pipeline
    - New people come and fork the workflow
-Defining interfaces and standards
    - For experimental and theoretical data
    - Scattering hub a la GitHub
-Custodial responsibilities
    - Management via institute
    - Validation, verification
    - Unit testing, datasets, test repo
-Creating and promoting semantic data formats
    - i.e. processing conditions matter
    - Metadata tools for comparing various materials
    - Provide good APIs such that it can be easily done automatically
    - Incentivizing
    - Get funding agencies to enforce it

Clearly, the SIXNS institute will serve a wide range of needs, from developing to maintaining to teaching and (end-user) supporting scattering codes; each of these roles requires resources. This asks the question of what hardware and infrastructure the SIXNS institute needs to fulfill these functions, and if it should own and maintain or contract those resources:
- HPC resources?
- Data servers?
- Networks and Databases?
- Experimental hardware and interfacing with instruments?

One approach is that of virtual infrastructure — which allows e.g. linking together databases between several federated areas (universities, labs). Cloud storage, compute clouds, and grid computing are now established technologies (though concerns exist about scaling up). For data storage SIXNS could piggyback on the user-facilities' infrastructure. However questions around long-term storage commitment and sustainability of databases must be addressed. The work group feels that a proposal to draw on national compute facilities and cloud resources may be

received more favorably than that to purchase yet another small or medium size cluster.

The infrastructure must accommodate the need to be able to easily pull together experiment and theory data:

- Analyze in place, like the proteomics-genomics pipeline at PNNL
- Use open source tools
- Run the analysis where the data is and move only the slices between
- Data standardization
- Dictionaries for cross-references
- Create interfaces for users to bring tools into, e.g. HubZero/Emsl-hub w $500k budget
- 55 instruments and 25 petabytes storage and 3.5 PF Intel Phi machine
- Smaller breakouts for real-time access
- ScalaBLAST, NWChem, Felix NMR pipeline

Real-time data crunching and feedback is necessary. This can be performed locally for small runs. If the workflow is too large, it would have to go to outside compute facilities. In order to effectively maintain codes and prepare to support their usage on cutting edge hardware, SIXNS must foster connections with the hardware industry, i.e., learn Intel, NVIDIA roadmap from reps:

- PNNL, Oak Ridge are models for developing vendor relationships
- How to pitch collaboration with SIXNS to vendors
- Try and bring in lab experts who already communicate with vendors
- Keep code development in the path that future hardware will take

All users who draw upon SIXNS' expertise for setting up their calculations will also require computational resources to perform these calculations. SIXNS could be an active provider of resources; but it could also take a more facilitating role and help users apply for securing these resources elsewhere. Facilities could be encouraged to provide compute resources in tandem with beam time. Another interesting model is that of providing "Computation as a Service" as offered by the successful platforms Nanohub and HubZero.

Other topics may involve Interfacing back to experimental hardware

- Should this be the Responsibility of labs?
- Education, tools for planning, planning compute requirements
- Need a good API
- Use of DOIs on data (cf. dryad.org — data dryad)

4. Neutron Scattering Theory Codes and Analysis Tools

Group Leader: Thomas Proffen, Garrett Granroth
Group members:
Stuart Campbell, Nicholas Draper, Daryl Hess, Dan Katz, Yonbin Lee, Jiao Lin, Gregory Schenter, Jonathan Taylor, Xiahao Yang, Arun Bansil, Keith Gilmore, Brent Fultz, Noa Marom, Yildrim Taner

Goal – There are three categories of analysis tools needed for Neutron Scattering experiments 1) Novel Theory work where there is no established analysis path for the specific problem, 2) Established theory, but no standard tool, and 3) Standard tools. The challenge to the scientific community is to integrate the use of all of these kinds of tools in regular analysis.

Novel Theory work 1)

One of the prime examples is the interpretation of off specular reflectometery data. Here some information is known for a few cases like magnetic multilayers, but fundamental theoretical work is needed to describe the underlying observed information specifically in polymers and other systems containing hydrogen.

Another example is first principle codes for magnetism. Currently the effect on the electronic structure can be calculated if a magnetic structure is assumed, however the first principle calculation of a magnetic ground state is not as simple as the structure of a material.

Many practitioners of filed theory techniques benefit from comparison to neutron scattering data. Examples are dynamical mean field theory and strongly correlated electrons or self consistent field theories and polymer chain interactions.

Extending the pair distribution function method into the time domain is another task that is seeing significant effort, but still has many problems to solve.

Established Theory but no Established codes 2)

Magnetic Spin waves is one area where the theory has been established for many years[1],[2], but a well established code does not yet exist. McPhase[3] provides some of these capabilities, but has a steep learning curve and so many choose to work their specific problems by hand. Very recently SpinW[4], Spinwave Genie[5], and Spinal[6] are all prospective codes for a simpler interface, but none have received wide acceptance. Neutron scattering studies of 1-D magnetic chains can also benefit from exact diagonalization codes many of these experiments are performed via home written codes[7]. Albeit there is now some use of the ALPS[8] package for doing these calculations[9].

Incommensurate crystal structure solution is another technique where the procedure is straightforward, but such features are not available in any of the standard analysis codes.

Specular reflectometery codes like multifit[10] and Refl1D[11] are close to moving in to the next category. However their interfaces are not well suited towards their user communities and thus there is still insufficient testing to know if they are as complete as something like GSAS[12].

3) Standard Tools

There are many standard tools that are now available. Prime examples are PDFgui[13] for working with pair distribution function data, GSAS[12], Fulprof[14], and Topas[15], for working with

powder diffraction data, McPhase[3] for fitting crystal field levels, Discus[16] and ZODS[17] for modeling diffuse scattering data, Sasview[18] for working with small angle scattering data, Unisoft[19] or VNF[20] for BvK models of Phonons, Mcstas[21] and McVine[22] for ray tracing simulation of neutron instruments, Restrax[23] and Reslib[24] for simulating instrument resolution of triple axis spectrometers, and Tobyfit[25] for simulating instrument resolution of direct geometry chopper spectrometers. More broadly, first principle codes like Castep[26], VASP[27], Quantum Espresso[28], CP2K[29], and Gulp[30] are being used to interpret neutron data. Here programs like aclimax[31], symphonies[32], and VNF[20] (the virtual neutron facility) are providing an interface between the codes and the measurements, however these codes are limited to either one code or one type of instrument. Ideally we want to use the first principal codes together with the more instrument tied codes. For this work we need to develop building blocks that close the gap between experimental data and the simulation output. To solve this challenge, a well-defined workflow is required. Two extreme cases are: 1) to bring the simulation to the detector; or: 2) take the experimental data from the detector to the simulation. The optimal framework should allow someone to approach their specific scientific problem from either direction to find their optimal solution. Such a design will enhance creativity and innovation for both the theoretical and experimental user communities. This framework will need to include standardized interfaces between applications and data formats. Novel approaches to documentation, training and outreach should be considered. Having this framework available to the user community will allow more users to fully evaluate the validity of their beam time proposals: a great need in an increasingly competitive beam time review process.

We envision significant scientific impact if neutron, X-ray, and data from other techniques can all be used together to constrain the theoretical models. This task will have the specific challenge of quantifying the uncertainty in a theoretical model based on the data. Significant foundational work will be needed to complete this task. Validation tools for the user's output should be available to reduce the errant results.

---

[1] S. W. Lovesey, *Theory of Neutron Scattering from Condensed Matter* Vol2. Page 57ff ,Clarendon Press, Oxford (1984).
[2] J.T. Haraldsen and R.S. Fishman, J. Phys.- Condens. Mat. **21** , 216001 (2009).
[3] https://kenai.com/projects/mcphase
[4] https://wiki.helmholtz-berlin.de/spinw/index.php5/SpinW
[5] S. E. Hahn, Private Communication, S. Hahn, *et al.* , Phys. Rev. B **89** , 014420 (2014)
[6] https://code.google.com/p/spinwaves/
[7] For example B. C. Watson *et al.,* Phys. Rev. Lett. **86**, 5168 (2001)
[8] http://alps.comp-phys.org/mediawiki/index.php/Main_Page
[9] M. B. Stone *et al.*, Phys. Rev. Lett. **99**, 087204 (2007).
[10] http://www.is.mpg.de/8137185/fitting
[11] http://www.reflectometry.org/danse/software.html
[12] https://subversion.xor.aps.anl.gov/trac/EXPGUI
[13] http://www.diffpy.org/products/pdfgui.html
[14] http://www.ill.eu/sites/fullprof/index.html; J. Rodriguez-Carvajal, Physica B, **55** 192 (1993)

[15] http://www.topas-academic.net
[16] http://discus.sourceforge.net
[17] T. M. Michels-Clark, Ph.D. Dissertation, The University of Tennessee, (2014).; http://trace.tennessee.edu/utk_graddiss/2770/
[18] http://www.sasview.org
[19] http://www.uni-pc.gwdg.de/eckold/unisoft.html; P. Elter and G. Eckold, Physica B **276-278**, 268 (2000).
[20] https://vnf.caltech.edu/vnf/1/
[21] http://www.mcstas.org;
[22] http://docs.danse.us/MCViNE
[23] http://neutron.ujf.cas.cz/restrax/; J. Šaroun and J. Kulda, Physica B **234** , 1102 (1997).
[24] http://www.neutron.ethz.ch/research/resources/reslib
[25] http://tobyfit.isis.rl.ac.uk/Main_Page
[26] http://www.castep.org
[27] http://www.vasp.at
[28] http://www.quantum-espresso.org
[29] http://www.cp2k.org
[30] https://nanochemistry.curtin.edu.au/gulp/ ; J.D. Gale, *Z. Krist.*, **220**, 552-554 (2005).
[31] A. J. Ramirez-Cuesta, Computer Physics Communications, **157**, 226 (2004).
[32] J. D. Budai *et al.*, Nature 13865 (2014); C. W. Li *et al.*, Phys. Rev. Lett **112**, 175501 (2014).

5. Excited States Theory Codes

Group Leader: Claudia Draxl, John Rehr
Group members:
Michal Bajdich, Arun Bansil, George Bertsch, Eric Bylaska, Alan Dozier, Paolo Fornasini, Cecile Hebert, Josh Kas, Steven Louie, Noa Marom, Micah Prange, Ritimukta Sarangi, Brian Moritz, Matt Newville, Chaitanya Das Pemmaraju, Niri Govind, Karol Kowalski

Goals - Theory and calculations of excited states in chemical and materials sciences have seen much progress in recent years. Several promising codes have emerged that are building a stable code base and proving their accuracy and applicability to a wide range of systems in complex environments. However, it will take a concerted effort to stimulate rapid adoption by a wide user community. Indeed, in computational chemistry and materials science theoretical innovations have traditionally "trickled down" to the people operating the instruments and maintaining the labs. Density Functional Theory (DFT), a ground-state framework that has long been adopted as a standard by the theoretical community years ago, still hasn't assumed the position of a benchmark method in applied research. It would be desirable to speed up this naturally slow process for the case of excited states codes. Many of these codes, especially those having implemented Time-Dependent Density Functional Theory (TDDFT), the Bethe-Salpeter Equation (BSE), and Equation of Motion Coupled-Cluster Theory (EOMCC) are now state-of-the-art, and the field of scattering science would benefit as a whole if these codes could be rapidly deployed throughout the entire field. Therefore, the SIXNS institute should serve as a hub to bring together experimentalists and theoreticians developing excited state theories applicable to scattering science. The experimentalists are faced with many questions: which theoretical frameworks are available for solving their research questions; how they differ or where they are applicable; which codes exist for a given framework; which code to choose for one or the other material or for one or the other question; or, once the right code is identified, how to install it and use it. Beyond inventory and support, many of these newer codes require development such as GUIs or interfacing to other codes and frameworks in order to build advanced workflows. The SIXNS institute can play a crucial role here.

Repository
The institute should function as a repository of software, knowledge, and expertise for theories of *excited* states, including:
DFT: Although Density Functional Theory is not formally valid beyond the ground state, its use is nevertheless widespread. Some DFT codes have, through decades of development, accumulated a large user base and a wide range of features, including scattering properties such as X-ray spectra and electron energy loss spectra for valence and core loss, … Important DFT codes include WIEN2k, ABINIT, VASP, CP2K, NWChem, …
GW+BSE: Over the past decade, codes implementing many-body perturbation theory (more specific GW+BSE) have emerged as the new state-of-the-art approach for calculating excitation spectra of materials, although some of the details are not yet as established as for DFT. These codes incorporate a self-energy in the GW formalism, and solve for the 2-particle Green's function in a 2-particle basis using the Bethe-Salpeter Equation. In contrast to DFT, the latter

can treat excitonic effects (particle-hole interactions) typically important near the onset of absorption edges, and are typically more accurate than the supercell core-hole approach within DFT. Some important codes include **`exciting`**, OCEAN/AI2NBSE, …

TDDFT is a time-dependent generalization of DFT that is theoretically also valid for excited states and formally equivalent to GW+BSE. It is typically implemented in the same codes that also offer GW+BSE as well as codes widely used in computational chemistry. However, TDDFT is constrained by the same exchange-correlation issues as DFT. Nevertheless there has been a lot of recent and ongoing research, especially in the computational chemistry community, to mitigate these problems at various levels with very promising results. Within the last five years, there have been several chemical and materials applications of TDDFT for XANES with very promising results. These methods are widely deployed in several computational chemistry user codes.

Multiplet: Multiplet codes are typically very fast and relatively small codes that calculate atomic properties with very high accuracy. As such they often describe the many-body aspects of excitations, which are so elusive to solid state theories, very well, while missing all or most of the material characteristics (sometimes summarized in a fitted parameter, e.g. crystal field splitting). The most important of these codes is CMT4XAS (along with its new relative, CMT4RIXS). "Hybrid" multiplet codes, which combine some of the traditional multiplet algorithms with "proper" solid-state calculations, have also been used, e.g. [Tanaka].

XAFS and XANES

RSMS: The real-space multiple-scattering formalism is based on calculations of the (1-particle) Green's function of the photoelectron, from which spectral properties and other properties can be derived in a straightforward way. This approach is very efficient as it avoids calculating a large number of states that may not contribute to the cross-section or may get broadened beyond recognition. In its real-space formulation it also has the advantage of not requiring symmetry or periodicity, making it useful for molecules and nanosystems. In accuracy it ranks above DFT and below GW+BSE, as it incorporates some many-body physics through a GW self-energy, but does not evaluate it explicitly. The most important code in this class is FEFF9. Among its reciprocal-space equivalents in the KKR approach, the SPRKKR code is most popular.

CC: Coupled-cluster approaches are, in a way, the gold standard, as they treat all interactions explicitly and therefore give unequaled accuracy. Despite the computational demands due to high numerical complexity, there has been significant progress in extending these methods to larger systems with state-of-the-art algorithms. However, these methods are mostly used in computational chemistry.

The institute should encourage codes that are open source and easy to use. The institute will – in collaboration with developers - define standards for spectroscopy data/file formats and infrastructure for interfaces between codes. Installers and binaries can be created if needed. The relative reliability and inherent uncertainty of methods for excited states will be documented by the institute through the development and maintenance of a set of benchmarks.

Outreach The institute will collect (links to) tutorials, webinars, example input files, etc. A database of published applications of excited state theories to scattering experiments will be developed and maintained. The institute will develop materials on how to design and perform simulations for complex, realistic systems. Theoretical beamline scientists will be available to advise novice users on the capabilities and limitations of available methods. These beamline

scientists can also advise established users with subtleties of existing codes and methods and inform about new features and conceptual developments. Travel funding could be provided to facilitate summer schools for potential or existing users.

Collaboration Advanced projects requiring new development can be completed collaboratively using expertise from code developers and the institute. The institute can serve as a conduit for new ideas for code development.

6. Integration of Software Workflows

Group Leader: Simon Billinge
Group members: John Rehr, Fernando Vila, Yonbin Lee, Brian Moritz, Tom Devereaux, Lars GM Petterson, Wei Chen, Xiaohao Yang, Das Pemmarajo, Pavol Juhas, Gian-Marco Rignanese, Noa Marom, Tsu-Chien Weng

Goals – Efficient Workflows: The SIXNS institute should support its users by capturing workflows all the way from data collection to publication. Its software design and development strategies should reflect this all-encompassing process. It will need to decide what pieces are needed and how to link them together. Its approach must offer maximal flexibility so that software elements can be combined to serve a large number of different workflows. Thus the institute will develop toolkits rather than prescribed procedures. Data should preferentially be passed between applications as a physics quantity.

The workflows may look different on the theoretical and experimental side of the same materials science problem. They may need to be handled differently. Users often just want to know the outputs of the calculations, and also how trustworthy the results are, not what happened in between. Ideally, the user can supply a list of parameters that they want calculated from a material and the framework figures out the best way to compute them and does the computation.

In an example implemented at the Molecular Foundry [], a workflow consists of a user who inputs a structure and gets back its NEXAFS spectrum. This user is uninterested in what happens on the back-end that processes the structure and calculates the spectrum. Some features of this implementation at the Molecular Foundry are that:
- the user can click on a peak in the spectrum and see a plot of the isosurfaces of the wave functions giving that feature;
- the forward problem calculation takes hours to days, depending on the complexity;
- thanks to the automation of the entire workflow behind a front-end, users are allowed to do calculations that an expert would consider routine, without the expert's direct input.

To inform the user of the trustworthiness of her analysis or simulation, the institute will need to increase scientific rigor of the workflow elements and quantify uncertainty in the final result. Additionally, workflows will need to be protected by a "custodian" program that can detect improper results due to a soft failure: the program runs to completion, no error, but the results are wrong.

Different types of users may require different workflows. Expert users may wish to have more control over the algorithmic details of the computational sequence, while novice users may prefer an abstracted, black-boxed interface. One could envision "expert features" (as well as perhaps increased computational resources) becoming available through an interface once a user reaches a certain experience level. Like computer games, you can progress from level to level and get new powers! Another classification of workflows, related to the matter of expertise level, can be conceived by thinking of user requests as falling in one of three categories:
(1) You can do it yourself (with the right tools).
(2) Give it to me and I will get you the results tomorrow morning (i.e., needs minimal support).

(3) I don't know the answer but I know someone who might be able to help.

2. Frameworks. To facilitate building flexible workflows, SIXNS must standardize data formats/descriptors; parameters can be stored in SQL databases or in a NOSQL database like mongoDB (Materials Project example). Also, SIXNS can reuse Materials Project tools such as "fireworks" and other open-source software like Avogadro, VESTA, etc. SIXNS must decide on the quantities that are stored in a database: what is the materials proteome? In Chemistry there is a start through CML (Chemistry Markup Language), JsonML (JSON Markup Language) and others.

SIXNS should consider carefully to what extent it will develop its own software frameworks and elements, or rather implement solutions that already exist. For example, if a framework such as ASE proves to be a good back end for SIXNS, the institute can focus on adding value through expertise, training and quality control (as opposed to placing more focus on the development of a new back end). Other platforms that could be wholly or partially adopted include the Materials Project, NanoHub, …

3. Community. Can the Institute foster collaboration as well as education? SIXNS could help users find the right people for collaboration. It would create an online community to connect users, e.g. SIXNS could host a StackOverflow or Diffpy-users type community help forum. Initially it could implement a tickets mode if there is not a big enough community to respond to all questions. A social network aspect to the site encourages participants to share best practices and help each other. The site can include a DOI or database of workflows that can be used as examples and adapted by rookies, modeled on open source software practices. User contributions can grow this database of workflows over time. A better way is needed to share workflows, give credit for contributions, share data and share scripts for making a publication.

The institute could play a role in providing feedback to experimental hardware and controls for computer intelligent high throughput studies.
Questions to study: i) Example workflows (Noa): use theory to guide experiment and conversely, use experiment to validate theory; ii) Inverse problem framework. [Check: look up Lars' Pettersson's SPECSWAP framework on GitHub. Lars P says BASF already did this successfully to find new polymers

7. Theoretical Software Institute

Group Leader: Brent Fultz
Group members: John Vinson, Kevin Jorissen, Keith Gilmore, Ercan Alp, Stephen Cramer, Yan Li, Matt Newville, William Shelton, Taner Yildrim

Benefits to scientists. Making scattering codes more usable for experimentalists is a primary goal of SIXNS. For example, designing codes and computational workflows that follow the structure of an experiment or the layout of an instrument is one approach to help experimentalists simulate measurements.

Scientists need a software infrastructure that is more professional than what is available today. All supported codes will be catalogued in terms of capability and accuracy and have benchmarks and unit testing. Codes will have similar interface structures to facilitate users switching between them. Clear, accessible tutorials, especially for inter-code workflows will be provided. These tutorials will be modular to compliment code modularity. A workflow for a research investigation should be able to smoothly span optimal platforms for each step, laptop to supercomputer to GPU/Intel MIC cluster to laptop, much the way an experimentalist uses a variety of specially built tools.

SIXNS will benefit theorists by offering easier paths to trying new approximations, functionals, and new applications. Theorists will have access to a collection of benchmark results and methods for comparison, including input files and measurements. SIXNS will tighten the feedback loop of interesting systems and problems that require better theory and/or better measurements. Importantly, theorists will receive support for the software design of GUIs, parallelization, and accelerators, and they will be assisted in creating sustainable code.

Using computing to accelerate the development of new materials is at the core of SIXNS, and this is the mission of the Materials Genome Initiative. The MGI will benefit from the connection to experiments at national user facilities for x-ray and neutron scattering.

Activities. A major activity of SIXNS will be to educate users about the functionalities and capabilities of codes and workflows. Often different types of codes are needed for the same project. Efficient, computationally cheap methods can initially screen materials for a given scientific problem (progressive selection). These methods can be calibrated against higher-precision methods.

SIXNS will foster collaborations between computational scientists in SIXNS and scattering scientists who use national user facilities. It will take effort to identify good scientific collaborations, and personnel at SIXNS will work with instrument scientists at national laboratories to help make these connections. A regular call for proposals for SIXNS projects could be set up in much the same way that computational proposals are solicited for nanoscience centers at national laboratories. In addition, SIXNS could establish visitor programs, internship programs, and exchange programs to facilitate interaction between the national labs and academia. A variety of workshops and webinars are possible. A conference series on

computational scattering science, hosting meetings such as the one that produced this report, is a high priority for keeping the community engaged.

The first software workflows from a scientific collaboration are likely to suffer from design flaws and inflexibility. SIXNS staff and management must make key decisions to rebuild important parts of a workflow, making them more useful for other research projects. Other research groups, and other staff at SIXNS, can then branch their own workflows, and later merge the improvements or features to the master branch. New tools such as Git make this practical, and help to promote a community effort.

Organization. For success, SIXNS must be a community-based activity, with a major role for the community in management. This requires a high level of responsibility in the community, since valuable resources are at stake. An executive board, staffed with members from national laboratories and universities, is needed. This board should have the authority to launch or terminate projects in SIXNS, and select the P.I. for a federal grant proposal. When SIXNS is underway, these board members should be elected. Voting may be restricted to regular attendees of the annual workshops, for example, since these persons would understand the challenge of matching the staff to the scientific mission of SIXNS.

Personnel and Career Advancement. It is likely that some of the most productive staff of SIXNS would be young computational scientists and software developers. Postdoctoral positions would be attractive to persons fresh with the Ph.D. degree when there are good opportunities to publish cutting-edge science. Mentoring these postdoctoral fellows must be a priority for the management team of P.I. and co-P.Is. Some of the postdoctoral fellows should have the opportunity to advance to technical leaders of SIXNS projects. Likewise, the software developers should have the opportunity to become project managers or software architects. It is anticipated, however, that for the postdoctoral fellows the priority will be on publications that allow careers outside SIXNS in more conventional academic disciplines, for example. Young software developers should have the opportunity to see and manage full project lifecycles that will give them an edge for outside employment, too.

Broader Education and Outreach. SIXNS can organize summer schools targeting graduate students who will be doing experiments at synchrotrons or neutron sources. Students can take advantage of SIXNS to receive training in the available codes for simulation and can immediately begin to apply it to their work. They may be able to simulate their experiment in advance, giving them an edge. Additionally SIXNS can offer periodic webinars (as NERSC does) teaching the codes through specific examples. SIXNS can also support internship programs at national labs for high school students. A fun and engaging variation on that theme would be a high school or college level competition: SIXNS would provide experimental data for an unidentified material and the students use computational tools to identify the material, and predict some of its properties. In working with young scientists on such short-term projects, start-up time is severely constrained; so ease-of-use is one of the primary requirements of any code. Undergraduate or summer student researchers can do exploratory work using lower-order computational methods that give useful results.

National attention is growing on the lack of women and under-represented minorities in careers of computer science or software development. SIXNS could reach out to women in the high school or the earliest years of college. For example, in the summer of 2014 a Caltech project in advanced visualization organized several student teams that each included a programmer, designer, and an artist. The resulting software applications were innovative, and valuable for both the students in the teams and seasoned researchers in neutron scattering. Visualization of rich datasets is appealing to a broader range of persons than is pure physics, and might provide an entry portal into computational physical sciences.

8. List of Participants.

The following experts participated in the SIXNS-3 workshop:

Kas	Josh	U Washington
Car	Roberto	Princeton
Draxl	Claudia	Humboldt U Berlin
Devereaux	Tom	SLAC/Stanford
Louie	Steven	Berkeley
Hebert	Cecile	EPFL-CIME
Alp	Ercan	ORNL
Proffen	Thomas	ORNL
Katz	Dan	NSF
Prange	Micah	PNNL
Vinson	John	NIST
Gilmore	Keith	ESRF
Ahmed	Towfiq	Los Alamos
Bansil	Arun	Northeastern U
Granroth	Garrett	ORNL
Li	Yan	BNL
Lin	Jiao	Caltech
Pemmaraju	Chaitanya Das	LBL
Govind	Niri	PNNL
Moritz	Brian	SSRL Stanford
Cramer	Stephen	UC Davis/LBNL
Chen	Wei	LBL
Schenter	Gregory	PNNL

Bajdich	Michal	SLAC Stanford
Draper	Nicholas	STFC
Taylor	Jonathan	STFC
Lee	Yonbin	Ames lab, DOE
Shelton	William	Louisiana State U
Newville	Matt	U Chicago
Fornasini	Paolo	U Trento
Sarangi	Ritimukta	SSRL
Weng	Tsu-Chien	SSRL
Rignanese	Gian-Marco	U Catholique de Louvain
Kowalski	Karol	PNNL
Dozier	Alan	U Kentucky
Roche	Ken	PNNL / UW
Bylaska	Eric	PNNL
Fultz	Brent	Caltech
Jorissen	Kevin	UW
Rehr	John	UW
Marom	Noa	Tulane U
Pettersson	Lars	Stockholm U
Campbell	Stuart	ORNL
Yang	Xiahao	columbia
Juhas	Pavol	BNL
Billinge	Simon	Columbia/BNL
Yildrim	Taner	NIST
Hess	Daryl	NSF
Vila	Fernando	UW
Bertsch	George	UW

9. Conference Program

Friday, 1-17-2014

Time	Topic	Speaker or Chair
7:45-8:30	*breakfast*	
8:30-8:50	Opening remarks; Scientific Software Innovation Institutes as part of NSF's SI2 program	J. Rehr, D. Katz
8:50-9:05	Outcomes of SIXNS-I	Brent Fultz
9:05-9:20	Outcomes of SIXNS-II	Simon Billinge
9:20-9:30	Goals for SIXNS-III	John Rehr
9:30-9:50	XAS analysis and modeling of x-ray absorption experiments	Matt Newville
9:50-10:10	Analysis and modeling of electron scattering experiments	Cecile Hebert
10:10-10:30	*break*	
10:30-10:50	The Materials Project	Wei Chen
10:50-11:10	The NWChem Platform	Eric Bylaska
11:10-11:30	Experiences w/ the UNEDF and PSI DOE SciDAC Computational Science Projects	Kenneth Roche
11:30-12:00	Advanced theory codes	S. Louie, C. Draxl
12:00-13:30	*lunch*	
13:30-15:00	Breakout Sessions 1:	
	1.A Theory codes for Ground-State properties	Gian-Marco Rignanese
	1.B Hardware, infrastructure, and data	Simon Billinge

	1.C Theory codes for neutron spectroscopy	Thomas Proffen
	1.D Theory codes for excited states, x-ray, and electron spectroscopy	Claudia Draxl
15:00-15:30	*break*	
15:30-17:00	Breakout Sessions 2:	
	2.A Analysis tools for x-ray and electron experiments	John Rehr
	2.B Analysis tools for neutron experiments	Garrett Granroth
	2.C Integration of software workflows	Karol Kowalski
	2.D SIXNS institute and proposal	Brent Fultz
19:00-21:00	*conference dinner*	

Saturday, 1-18-2014

Time	Topic	Speaker or Chair
7:45-8:30	*breakfast*	
8:30-10:00	Prepare reports for breakout sessions	(see Friday schedule)
10:00-10:30	*break*	
10:30-11:20	Presentation and discussion of breakout reports 1.A-D (4x12 mins.)	
11:20-12:10	Presentation and discussion of breakout reports 2.A-D (4x12 mins.)	
12:10-12:20	Brief participant contributions (10x1 min.)	
12:20-13:30	*lunch*	
13:30-14:50	Conclusions and SIXNS proposal	
14:50-15:00	Closing remarks	John Rehr

Breakout topic descriptions:

SESSION 1

- **1.A Theory codes for ground state properties.** It is usually necessary to refine a model by optimizing atom positions, lattice constants, and defect structures. Ground-state densities and potentials are frequently a necessary first step for calculating excited-states properties. The SIXNS institute therefore needs a repository of robust ground-state engines that together are capable of describing all relevant material classes.
- **1.B Hardware, infrastructure, and data.** What will be required of the SIXNS code in terms of handling data and properly utilizing the institute's hardware and infrastructure?
- **1.C Theory codes for neutron spectroscopy.** Inelastic neutron scattering is used to measure the dynamics of nuclei or spins. Phonon studies can benefit today from DFT methods because the electrons remain in the ground state throughout the scattering process. Here the question is why computational scattering science is not further along. For magnetic scattering studies, calculations on systems with spin disorder presents a challenge. A second issue is that much of the scattering research is on quantum magnetism in low dimensional materials, where ab initio methods are not available.
- **1.D Theory tools for excited states, x-ray and electron spectroscopy.** Generally, modeling of spectra requires theoretical tools that go beyond the approaches used for calculating ground state properties. The response of a material to a perturbation must be modeled carefully to the specifics of the excitation responsible for a given x-ray or electron spectroscopy. As no "perfect codes" exist that can reliably simulate all spectra of all materials, the SIXNS institute requires a portfolio of complementary codes that cover as much as possible of the range of spectroscopy experiments of interest.

SESSION 2

- **2.A Analysis tools for x-ray and electron experiments.** Experimentally measured data generally needs to be treated to remove undesirable contributions. Analysis requires comparison to theoretical reference data while fitting or extracting quantities of interest, while accounting for experimental conditions. The software tools for such analysis must be versatile yet user-friendly.
- **2.B Analysis tools for neutron experiments.** Analysis tools for neutron experiments. We will assume that the data are already "reduced," and corrected for most characteristics of the instrument. Further analysis today is usually done without reference to materials theory. How can our present analysis methods be adapted to better include the capabilities of computational materials science? What new workflows can be built? How should these be prioritized?
- **2.C Integration of software workflows.** The plethora of software elements must be combined into a versatile and intuitive workflow. This involves program control, data management, and user interfacing. Design decisions must be taken on interfacing codes and streamlining I/O.
- **2.D SIXNS institute and proposal.** The goal of the SIXNS workshops is a blueprint of a theoretical institute, and a matching proposal for this institute to be funded. In this workgroup we will bring together the work done at the three workshops and begin to integrate it into a blueprint and proposal.