# White Paper for Establishing a User Facility for Synchrotron-based High Throughput Materials Characterization

John M. Gregoire[1], Matthew J. Kramer[2], Apurva Mehta[3]
[1]Joint Center for Artificial Photosynthesis, California Institute of Technology, Pasadena, CA, gregoire@caltech.edu
[2]Critial Materials Institute, Ames Laboratory, Iowa State University, Ames IA, mjkramer@ameslab.gov
[3]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA, mehta@slac.stanford.edu

*This technical white paper summarizes the outcomes of the Dec 13th, 2013 workshop "Rapid Discovery of Functional Materials: A Workshop on Synchrotron-enabled High Throughput Approaches." The primary finding of the workshop was that High Throughput approaches will play a key role in rapid discovery of new functional materials. It will quickly identify composition-synthesis spaces, equilibrium and metastable, with interesting and technologically important properties. Also by generating a large number of structure-property relationships, especially for near-equilibrium and metastable structures, it will drive development of new theories and computational prediction of metastable materials, and hence it will have a large impact on Materials-by-Design efforts as well.*

*The High Throughput Research cycle is composed of three distinct steps: High Throughput synthesis of sample libraries, High Throughput characterization of structures and properties, and automated and machine-learned methods for datamining and analysis. A truly productive High Throughput materials discovery pipeline, not unlike a structural genomic pipeline, must have each of the three stages working at the highest level with seamless transitions among the material synthesis, characterization and computational processes. Each step requires very different types of human and material resources, and it is unrealistic that they will be performed at the same institute. But, it was recognized that a High Throughput Characterization User Facility can play the vital role of providing a gathering place where not only experimentalists and theorists but researcher from very different fields of functional materials come to exchange ideas and foster new collaborations.*

*The summary is divided into the following six sections:*
*1. Introduction*
*2. High Throughput Characterization Vital to the HiTp Research Cycle*
*3. A High Throughput XRD Characterization User Facility*
*4. Long-term and Short-term Challenges*
*5. Summary of Resource Requirements*
*6. Appendix: Workshop Program, Agenda and List of Attendees*

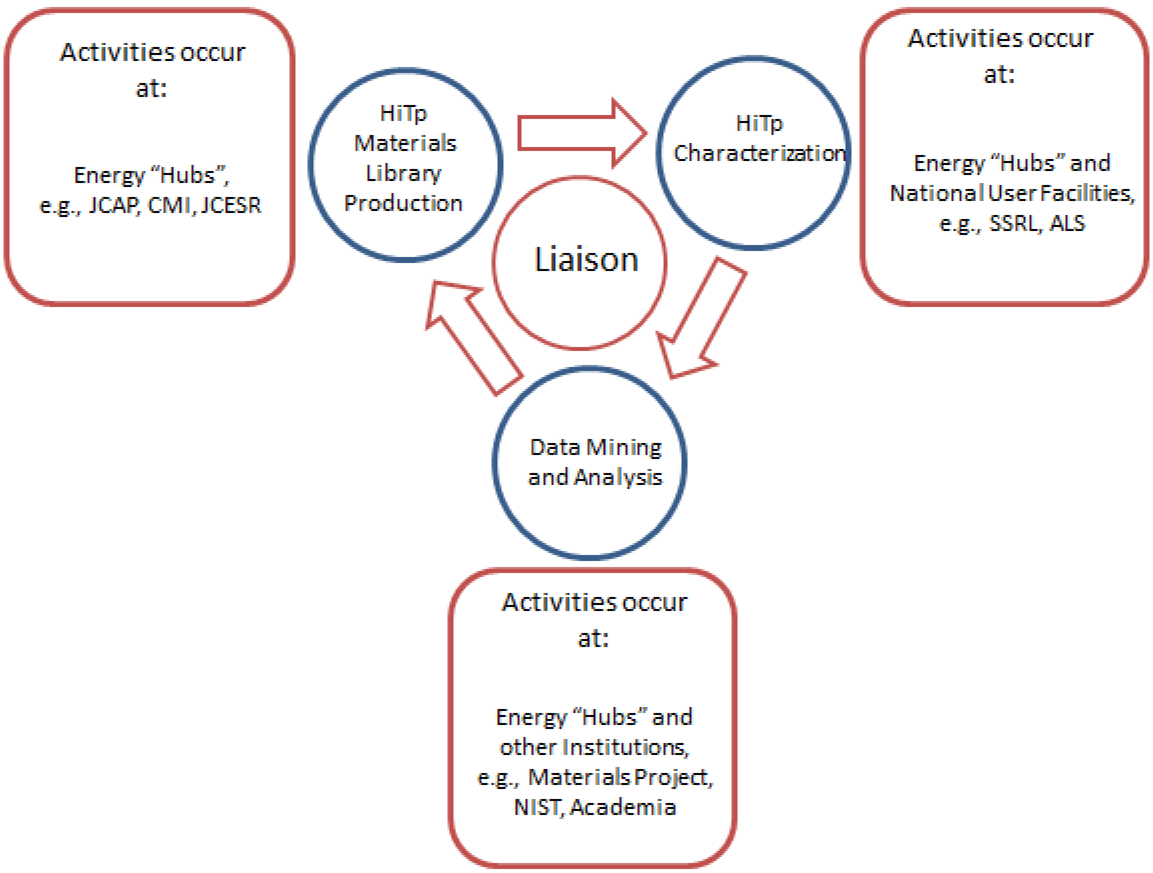
## 1. Introduction:

Rapid discovery of new functional materials is key for meeting many of the mission goals set by DOE for the national energy and technology needs. Investigators at several of the US-DOE Energy Innovation Hubs, national user facilities and academia, including Joint Center of Artificial Photosynthesis (JCAP), Joint Center for Energy Storage Research (JCESR), and Critical Materials Institute (CMI), National Institute of Standards and Technology (NIST), Advance Light Source (ALS) and Stanford Synchrotron Radiation Lightsource (SSRL) have proposed that high-throughput (HiTp) approaches to rapid materials discovery will result in many new functional materials, as well as many unforeseen new structure-property relationships. The rapid discovery of new structure-property relationships will impact many aspects of materials science from manufacturing to theory by creating new insights into metastable materials, and by discovering new methods for achieving the desired metastable states, and by providing a feedback loop for computational materials design. The HiTp is, therefore, both complementary to and a validation of computational approaches based on materials genomics and materials-by-design for discovery of new functional materials. To test some aspects of this approach, a proto-type HiTp diffraction setup was designed and commissioned at SSRL in an informal collaboration between JCAP, CMI, University of Maryland, and SSRL. The HiTp XRD-XRF prototype was commissioned to address specific scientific needs, but it also provided insights and strategies that could be applicable to many other HiTp characterizations systems and for HiTp research cycle in general.

To share and discuss these insights, identify major challenges and bottlenecks, and devise strategies for further improvement and extension of the existing prototype, a workshop was organized at SLAC National Accelerator Laboratory on Dec 13$^{th}$, 2013. Participants came from various DOE national user facilities (SSRL and ALS), DOE hubs (JCAP, CMI, JCESR), academia (University of Maryland, Cornell University, University of South Carolina), and other institutions (NIST, Materials Project). The workshop also served as a primer on high throughput x-ray diffraction (HiTp XRD), as the 26 attendees ranged from facility and program management to beam line scientists and from high throughput experimentalists to computer scientists.

## 2. HiTp Characterization Vital to the HiTp Research Cycle:

We consider the HiTp materials discovery approach as a three-step cycle, depicted in blue circles below: HiTp production of combinatorial materials libraries, HiTp characterizations of these libraries, and unsupervised, automated and machine-learned approaches to extraction of often hidden and unforeseen patterns and relationships from them. The new insights from data mining step drive the next generations of combinatorial synthesis, completing a feedback loop for rapid optimization of material function.




Activities at each of these stages generally take places at different institutions, but even when they occur in the same institution they are performed by different groups of people coming from different scientific and technical background. Institutions and groups where these activities take place are listed in the red boxes in the diagram above. A truly productive HiTp materials discovery pipeline, not unlike a structural genomic pipeline, must have each of the three stages working at the highest level with seamless transitions among the experimental and computational processes.

The workshop commenced with educational primers presented by experts from each of these disciplines. In the ensuing discussions, workshop attendees expressed a desire for a HiTp Characterization User Facility that will serve the characterization needs for many diverse HiTp materials discovery efforts, from better photoabsorbers to strong rare-earth free magnets. Also as such a facility lies between the synthesis and data mining institutes, it was recognized that it will become a gathering place where experimental and computational groups come to exchange ideas and develop collaborations, as they have at SSRL prototype XRD-XRF HiTp facility over the last two years. Several specific ideas for coordination with the Materials Genome Initiative and other efforts were discussed, with the workshop coordinators agreeing that this important topic merits additional attention that is beyond the scope of this document, and will be addressed in follow-up dedicated workshops, in the autumn of 2014 perhaps in conjunction with the SSRL and ALS users' meeting.

## 3. A HiTp XRD Characterization User Facility:

Because of limited time, but also because structural characterization is an integral aspect of most HiTp research cycle, at the workshop we focused on needs and challenges of a HiTp XRD characterization facility and rest of this summary will do likewise. However, though the challenges and goals discussed were specific to HiTp XRD characterization, it was felt that many of the strategies will have implication and influence on other types of HiTp characterization as well.

One of the challenges in establishing a user facility for HiTp characterization is that each research group addresses unique sets of materials ranging from metals to oxides, to organo-metallics with morphologies ranging from thin films to bulk. As such, a user facility will require some flexibility in data acquisition and data handling.

Below we outlined a configuration of the XRD and x-ray fluorescence (XRF) synchrotron instrumentation that will accommodate a variety of research programs and is suitable for a multi-user facility:

- Reflection scattering geometry, with a large 2D detector, and X-ray energies from 12-20 keV, to allow for phase identification even for textured materials.
- Concurrent XRD and XRF measurements, to allow for direct composition-structure cross-correlation.
- Translation of the combinatorial library over 200 mm in 2 directions with fixed vertical position with respect to the incident beam, which will allow deep composition coverage even for a quaternary system.
- Translation stage mounts with sufficient vertical translation to accommodate different types of stages, such as controlled environment, heater, etc., which will allow exploration of near-equilibrium and metastable phase diagrams.

There is also a commonality of the data collection and the initial data processing, motivating the integration of data capture and organization into the experimental station. Thoughtful consideration was given to the distinction between general data processing, which must be accomplished as nearly in real time as possible, and research-specific data interpretation, which will be mostly done elsewhere. However, because it was felt that, at present, development of data mining and machine learning algorithms are the weakest part of the HiTp research cycle, they were discussed with the conclusion that these research activities are beyond the scope of a user facility. But given the importance of these data tools to the HiTp Characterization User Facility workshop attendees proposed the organization of a future workshop focused on these efforts.

The proto-type HiTp XRD experiment at SSRL BL1-5, 7-2 and 10-2 provided a robust framework for considering future capabilities. The consensus was that the throughput of about 2,000 diffraction patterns a day is adequate for the near future, with the eventual goal of increasing throughput to 20,000 patterns/day. Representatives of the DOE hubs noted that dedicated beam time of approximately 4-6 weeks per year would meet the present research needs of the DOE hubs and that the establishment of this user facility would positively impact future research strategies directed at rapid discovery of functional materials. The ideal energy range for the measurements planned by the hubs is 12-20 keV. Preliminary measurements of HiTp libraries in transmission geometry at SSRL BLs 10-2 and 7-2, and other similar experiments at CHESS, indicated that diffuse scattering from the substrate posed a daunting background subtraction challenge. Collection of diffraction patterns in the reflection geometry with shallow incidence angle at the proposed energy range, through experience from measurements at the prototype facility at SSRL BLs 1-5 and 7-2, and extensive simulations, indicated that with the optimal placement of a 4-6 Mpixel diffraction detector, this will provide sufficient information content from even heavily textured samples. While working at energies higher than 20 keV may allow for collection of a large slice of the reciprocal space, reduction of the incidence angle and subsequent enlargement of the beam foot-print on the library did not merit this.

While a focused bending-magnet end station at SSRL is adequate for the near future, the establishment of the user facility on an insertion device beam line would be beneficial, particularly in reaching energies beyond 15 keV. Further determination of the beam line source and optics parameters requires both additional modelling of the experiment and identification of next-generation detectors, which will be part of the technology development effort noted below.

## 4. Long-term and Short-term Challenges:

Several present shortcomings and challenges were identified. They can be classified in three broad categories: facility access and support, data handling, and instrumentation development.

The biggest **access challenge** is that the beam line where the proto-type experiment was developed (BL1-5) is undergoing renovation and will not be available for further experiments until the late summer or fall of 2014. The future access to beam lines that might be more appropriate for HiTp will be determined by the growth in this activity and the support that can be brought in to to facilitate further development and sustain operation of this activity. .

Two major **software challenges** associated with the characterization facility were identified as further automation of data collection (see also equipment and hardware development) as well automated data reduction that is agnostic to the experimental configuration.

There was consensus that adaption of HDF5 data archiving format for large experiments at SSRL and ability for the SSRL developed 2-D diffraction analysis software (WXdiff) to handle this container format will not only accommodate some of the needs of HiTp experiments but can also help with other similarly structured large diffraction experiments (e.g., *operando* measurement of charging/discharging of an energy storage system) and hence would benefit significantly larger and growing user community.

Two *software action items* were discussed as a result:
1) Development of a GUI program for packaging a set of diffraction patterns from a combinatorial library (or any other experiment that generated correlated data - could be, for example, an *operando* measurement on charging-discharging of a battery) in to an HDF5 container (with compression turned on). Each data element would be a 2D image, but stored as a 3 or 4 D block to achieve maximum compression. The container would also contain additional diffraction patterns from known standards (e.g., $LaB_6$, Si, Ag behenite, etc.), XRF scans from standard NIST-supplied or custom elemental alloys, and pertinent meta-data (energy of the incidence beam, approximate position and tilt of the detector with respect to the incident X-rays, etc.).

2) Second recommended development was transformation of the diffraction and fluorescence data to become agnostic of the actual experimental conditions and thus could be used by any data mining and analysis group for pattern extraction. These transforms would convert XRF data into relative elemental concentrations. Pymca has most of the functionality for such a transformation. The transformation for diffraction data would convert a 2D diffraction image in pixel space into a 2D surface in reciprocal space

defined in Q and chi (azimuthal angle) space, and a further reduction into a 1D (I vs Q) powder diffraction pattern. Most of the current data mining and indexation efforts ignore texture, and they would use the 1D patterns. But as newer texture sensitive data-analysis approaches become more common use of 2D Q, chi 2D patterns grow. WXdiff (the software for analyzing 2D diffraction data, developed at SSRL) has most of these functionalities, but it needs some further development to read and write into large containers. With the needed development, either as part of WXdiff or as part of a stand-alone program, the raw data HDF5 container would be read, XRF data will be sent to a client Pymca process, and the diffraction data would be processed by WXdiff. The transformed XRD and XRF would be stored into another (compressed) HDF5 container. These transformed and compressed data containers, agnostic of the actual experimental conditions, would be the output of the characterization facility, and made easily accessible to any data mining and analysis group for pattern and relationship extraction.

The developments would require dedicated software engineering for these capabilities to be implemented and commissioned for wider use.

Near term and longer range **equipment and hardware challenges** were identified.

The near term *experimental action items* are:

1) Translate existing proto-type experiment to new end station
2) Improve and automate alignment of materials libraries, especially those produced on economical but less than perfect substrates (glass plates or warped Si wafers), with real time error correction from misalignment.
3) Incorporate *in-situ* rapid annealing capability for combinatorial libraries to enable characterization of metastable and near-equilibrium phases.
4) Build enclosure to enable detection of low-Z (to O, and N) elements in XRF measurements (ability to obtain chemistry of higher Z element already exists).

The primary long term need is a more optimal detector for the XRD measurement. Discussions at the workshop identified characteristic of an ideal detector for these measurements. They are, small pixel size (<100 microns), low dark current (a few photons/sec), large dynamic range (> 1E5), large area (> 4 Mpixels), and short readout time (< 10 msec). Discussion with detector experts at the workshop and follow-up discussions with the SLAC detector group indicated that no current detector meets all of the requirements for an ideal HiTp detector. However, several new detectors systems with characteristics similar or better than specified above, most of them based on hybrid pixel arrays, are being actively developed at several institutions, including SLAC. These detectors, when available, will not only fulfill the needs of HiTp diffraction measurements, but a very large and growing class of synchrotron scattering measurements.

## 5. Summary of Resource Requirements:

As noted above, additional effort is required to define the full resource requirements for the long-term goal of establishing a user facility with the noted performance. Delineation of these resources will occur through further workshops and related activities but also requires additional design and modeling. These activities are commensurate with the continued development and operation of the proto-type XRD/XRF experiment.

## 6. Appendix: Workshop Program, Agenda and List of Attendees

# Rapid discovery of functional materials: A workshop on synchrotron-enabled high throughput approaches

**13 December 2013**
**9:00 AM to 5:00 PM**
**SLAC National Accelerator Laboratory, Menlo Park, CA**
**Truckee River Conf. Room (Room 206, Bldg. 52)**

Organizers: John Gregoire, Matt Kramer and Apurva Mehta

Invited Attendees:

**SLAC**[1]: Apurva Mehta, Anders Nilsson, Marvin Weinstein, Gabriella Carini, Amber Boehnlein, Stefan Mannsfeld, Britt Hedman, Mike Toney

**ALS**[2]: Alex Hexemer, Alpha N'Diaye

**JCAP**[3]: John Gregoire, Santosh Suram, Misha Pesenson, Junko Yano, Frances Houle

**CMI**[4]: Matt Kramer

**JCESR**[5]: Venkat Srinivasan

Kristin Persson (**MP**[6]), Tieren Gao and Ichiro Takeuchi (**UMD**[7]), Ronan LeBras, Carla Gomes and R Bruce van Dover (**CU**[8]), Jason Hattrick-Simpers (**SC**[9]), Gilad Kusne (**NIST**[10])

[1]Stanford Synchrotron Radiation Laboratory, [2]Advanced Light Source, [3]Joint Center for Artificial Photosynthesis, [4]Critical Materials Institute, [5]Joint Center for Energy Storage Research, [6]Materials Project, [7]University of Maryland, [8]Cornell University, [9]University of South Carolina, [10]National Institute of Standards and Technology

## Agenda

| | | |
|---|---|---|
| Continental Breakfast | | 8:30 |
| Welcome and Introductions: HiTp research cycle | Apurva Mehta | 9:00 |
| JCAP, CMI and JCESR describe their hub's mission and need for HiTp x-ray science | Hub representatives (5min each) | 9:20 |
| Primer: Composition libraries and combinatorial materials science | John Gregoire | 9:35 |
| Primer: Synchrotron HiTp XRD | Matt Kramer | 9:45 |
| Primer: X-ray 2-D detectors | Gabriella Carini | 9:55 |
| Primer: Data formats and existing analysis tools | John Gregoire | 10:05 |
| Break | | 10:20 |
| Primer: Computational investigation of high order composition spaces | Kristin Persson | 10:30 |
| Primer: Existing big data analytics capabilities at the light sources | Alex Hexemer | 10:45 |
| Discussion: Enhancing data quality with experiment improvements | Apurva Mehta | 11:00 |
| Discussion: Real-time diagnostics, meta data organization and data archiving | John Gregoire | 11:30 |
| Working Lunch: Literature review | Santosh Suram | 12:00 |
| Discussion: Unsupervised data analysis | John Gregoire | 1:00 |
| Discussion: Data interpretation and coupling with theory/computational efforts | Kristin Persson | 1:30 |

| | | |
|---|---|---|
| Report: Discussion results and additional topics | Santosh Suram | 2:00 |
| Break | | 2:45 |
| Report: Define required resources, distinguish user facility vs. hubs vs. single PI efforts | Apurva Mehta | 3:00 |
| Report: Identify potential sources for required resources | Mike Toney | 3:30 |
| Concluding remarks | Apurva Mehta/John Gregoire/Matt Kramer | 4:00 |
| **Report Generation** | | |
| **Workshop notes:** Santosh Suram**. Report editors:** John Gregoire, Apurva Mehta, Matt Kramer | | |

**Workshop goals:**

1. By enabling open discussion among the gathered experts from each of the three stages of the HiTp cycle, **identify major challenges both within each step and in transitioning between steps; identify the aspects of high throughput synchrotron experiments and data analysis that can address these challenges.**
2. **Identify strategies, opportunities and resources for overcoming these challenges** and elevating high throughput synchrotron studies to a collaborative effort among experimentalists and theorists at DOE energy hubs, national labs (including user facilities), and academia. In order to help us attain these goals effectively, will we try to breakdown them down into challenges that can be overcome with immediate (and perhaps incremental) effort and resources, and challenges that would require a more comprehensive and inclusive vision and long term investment, but at the same time will have very high impact in accelerating the pace of material discovery.

**Discussion outlines:** In this workshop, due to time constrains, we will use synchrotron based HiTp XRD as a prototype example, with the hope that strategies, approaches and protocols developed for XRD, will be easily modified and expanded to include other HiTp synchrotron-based characterization techniques.

**Discussion: Enhancing data quality with experiment improvements**
Each high throughput experimentation effort must tailor composition libraries to the pertinent performance screening techniques. To establish a unified high throughput synchrotron experiment for user groups, the various relationships between library format (sample properties, substrate, etc.) and experiment data quality need be noted. With this information, we can discuss aspects of the synchrotron experiments that provide both versatility and excellent data quality. This session will provide important experimental background for the ensuing discussion sessions.

**Discussion: Real-time diagnostics, meta data organization and data archiving**
A HiTp experiment will produce large quantity of data very quickly. It is essential that the data is assessed for signal quality and accuracy as close to real time as possible and errors and artefacts corrected quickly as quickly as possible. In this session we will explore various ways of placing calibration references on the library and other benchmarking protocols to qualify the data in real time. To make the transition from data collection to data analysis stage transparent, it is essential that the data is universally readable and easy to store and transport. In this session, we will also, explore ways of transforming the data from a experiment specific format (e.g., set of pixels maps) to a more universal and self describing format (e.g., Q vs I matrices in HD5F structure), or alternatively, extracting and associating experiment specific metadata (e.g., x-ray wavelength), so that the data is easy to read, store and transport.

**Discussion: Unsupervised data analysis**
The real power of HiTp approaches will be realized if unexpected and hidden, and sometime subtle, new features (phases or structure-property relationships) are discovered. In this session will discuss strength and weaknesses of

the current unsupervised datamining strategies to find these hidden features, and explore other new approaches to enhance the discovery rate.

**Discussion: Data interpretation and coupling with theory/computational efforts**

Finally, in the last few years, materials theory is also developing rapidly. These new theoretical approaches are capable of predicting structures and properties of many material systems. In this session, we will discuss potential interactions between these theory predictions and HiTp synchrotron characterization. Ultimately, we envision that HiTp measurements can help test, validate and refine these theories, and that theory-derived composition-structure-property relationships can guide the HiTp search for new functional materials. To build toward this vision, we will establish some concrete modes of theory-experiment data interaction. For example, in the context of the algorithms noted in the previous session, we will discuss the minimum HiTp XRD data requirement for determining the presence of a theory-predicted phase.