APS/CNM 2020 Workshop on Autonomous Control of Experiments
Description

The APS and CNM are positioned to help solve some of the most challenging and novel scientific questions facing the energy needs of the nation. The design of new materials to manipulate classical and quantum information with high fidelity and ultralow power consumption and the enabling of systems for efficient energy storage, transportation, and conversion that will drive the emerging economy based on renewable energy are just a few examples. Addressing these scientific opportunities will be aided by the intrinsic capabilities of APS-U era facilities along with new measurement techniques and technological advances in detectors.

These advances in sources and detectors (x-ray and electron) will result in orders of magnitude higher data rates, and increased complexity from multi-modal data streams. Human-in-the-loop experiments become infeasible in the face of such large and varied data streams. As experiments progress to speeds where humans are too slow to make control decisions, adaptive control becomes imperative. This workshop is organized to discuss the state-of-the art and potential of autonomous control of experiments. It provides an opportunity for academics, laboratory and facility staff, researchers, and students from both x-ray and electron characterization communities to exchange ideas and think creatively about new avenues for collaborations and advance autonomous characterization and experimentation.

APS/CNM 2020 Workshop on Autonomous Control of Experiments
Speaker Abstracts

Marrying AI and Physics Towards Accelerating Scientific Discovery
Payel Das, IBM

Scientific discovery is one of primary factors underlying advancement of human race. However, traditional discovery process is slow compared to the growing need of new inventions, for example, new antibiotic discovery or design of next-generation energy material. Data-driven approaches such as machine learning and especially deep learning have achieved remarkable performance in many domains including computer vision, speech recognition, audio synthesis, and natural language processing and generation in recent years. Those methods have also infiltrated other fields of science including physics, chemistry, and medicine. Despite these successes and the potential to make huge societal impact, machine learning models are still at infancy in terms of driving and transforming scientific discovery.

In this talk, I will talk about a closed-loop paradigm to accelerate scientific discovery, which can seamlessly integrate machine learning, physics-based simulations, and wet lab experiments and enable new hypothesis and/or artefact generation and validation thereof. Development and use of deep generative models and reinforcement learning-based methods for designing novel peptides, small molecule drug candidates, and metamaterials with desired functionality will be discussed. Finally, I will discuss the importance of adding crucial aspects, e.g. creativity, robustness, and interpretability, to the machine learning models in order to enable and add value to AI-driven discovery.

Autonomous control at X-ray sources from accelerator to detector
Daniel Ratner, SLAC National Accelerator Laboratory

As the dimensionality and resolution of photon science experiments increases, it is no longer tenable for users to simply scan parameters during a beam time. Similarly on the machine side, the large number of control parameters needed to optimize advanced operating modes precludes an exhaustive search during setup. Instead “smart” scans are needed for both, adaptively probing the sample to focus on the highest value measurements or searching through parameter space to find the most likely regions to optimize performance. Here I will give an example on both the user and machine sides: In X-ray fluorescence, we use a reinforcement learning approach to control beam size and position to focus on the highest value regions of a sample. On the machine side, we use Bayesian optimization incorporating knowledge of physics to tune up the machine.

Autonomous X-ray Scattering
Kevin Yager, Brookhaven National Laboratory

This talk will discuss the ongoing development of autonomous experimentation at a synchrotron x-ray scattering beamline. Deep learning (convolutional neural networks) is used to classify x-ray detector images, with performance improving when domain-specific data transformations are exploited. These methods can be combined with customized data healing algorithms. To close the autonomous loop, we deploy a general-purpose algorithm that selects high-value experiments to conduct, attempting to minimize both uncertainty and experimental cost. Examples from recent autonomous experiments will be presented, including measuring nanoparticle ordering, combinatorial libraries of block copolymer materials, and realtime photo-thermal processing.

Autonomous Materials Discovery Under Uncertainty, Driven by Gaussian Processes
Marcus Noack, Lawrence Berkeley National Laboratory

Autonomous experimentation is an emerging paradigm for scientific discovery, wherein measurement instruments are augmented with decision-making algorithms, allowing them to autonomously explore parameter spaces of interest. Materials sciences, as well as many other experimental disciplines, suffer from large and high-dimensional parameter spaces, which have to be explored in search of new science. The vastness of these underlying spaces renders brute-force methods unusable. However, methods for autonomous experimentation have become more sophisticated in the recent past, allowing for multi-dimensional parameter spaces to be explored efficiently and entirely without human interaction. The scientist, in the meantime, is free to focus on interpretations and big-picture decisions. Gaussian process regression (GPR) techniques have emerged as the method of choice for steering many classes of experiments. GPR allows for uncertainty quantification of the model, given data, and can provide suggestions where future data should be collected in order to maximally decrease uncertainty. We have developed a software tool targeted at experimentalists, especially experimentalists at light sources, which utilizes the power of GPR. We tested our methodology at x-ray scattering beam lines at the ALS and the NSLS II. At some of those beam lines, the way experiments are conducted has transformed entirely; from scientists spending days at a time controlling experiments to pressing a button and focusing on high-level tasks.

Online, Quantitative Data Analysis for Coherent X-ray Imaging with the PyNX toolkit
Vincent Favre-Nicolin, European Synchrotron Radiation Facility

4th generation synchrotron sources provide two orders of magnitude more coherent photons, and thus the ability to collect coherent X-ray imaging datasets faster and/or with a higher resolution. Consequently, the increased volume of data requires dedicated tools to fully take advantage of the improved coherent flux. In this presentation we will present the PyNX toolkit[1], which is developed at ESRF to provide fast (GPU-accelerated) and accessible (using simple command-line scripts or notebook) data analysis for various of experimental techniques:

- Coherent Diffraction Imaging (CDI) and Ptychography (far field and near field) for two and three-dimensional imaging
- 3D CDI and Ptychography in the Bragg geometry to also provide strain information in nano-crystals

We will present examples for various types of samples, and show how the available tools aim to remove the need to master coherent X-ray imaging techniques to exploit them, including improved statistical methods for unsupervised analysis. This will ultimately enable a wider community to take advantage of the increased coherent photon flux, and focus on working on an extended range of applications and samples.

[1] - http://ftp.esrf.fr/pub/scisoft/PyNX/doc/
[2] – Favre-Nicolin, Leake & Chushkin, Sci. Rep. 10, 2664 (2020)

Machine Learning, Reinforcement Learning and Classical statistical physics: Opportunities for automated experimentation for non-equilibrium states and physics discovery
Rama Vasudevan, Oak Ridge National Laboratory

Deep learning and reinforcement learning have attracted increasing attention of late, buoyed by successes in traditionally difficult areas such as computer vision, machine translation and decision making in complex environments. However, their application for the 'labs of the future' will require close connection with existing physical models and be constrained by physical realities in order for them to both improve their predictive capabilities, as well as increase their data efficiency.

In this talk, I will explain our recent efforts at deploying deep learning and reinforcement learning in the area of materials imaging and materials synthesis, respectively. Deep learning, and machine learning approaches more generally when used correctly can greatly expand the amount of possible information that can be obtained from single (or multiple) datasets. Once the descriptors are formed, however, classical statistical physics methods offer a route towards understanding and predicting a system's behavior for arbitrary thermodynamic conditions. Merging these two areas together can rapidly improve convergence and offers an alternative to purely information-theory based searches, in contrast to most active-learning approaches in the literature.

Similarly, of note in automated experimentation settings is the dynamic capability to guide a system's state towards desired targets in stochastic environments. Such a task is well-suited for reinforcement learning. I will explain our recent work in this area, particularly of training agents via the Stein variational policy method, for a materials synthesis problem in a simulated environment. Deploying such agents on real systems can pave the way for atomic-scale fabrication and stabilization of non-equilibrium phases, a long-sought after goal of the materials community. This work was conducted at the Center for Nanophase Materials Sciences, a US DOE Office of Science User Facility.

Smart Data Acquisition and Automated Data Curation for Electron Microscopy
Charudatta Phatak, Argonne National Laboratory

Modern electron microscopy is no longer only driven by instrumentation but is increasingly linked with computational and data-driven algorithms and methods for acquisition and analysis. Harnessing this data for scientific research specially to enable machine learning approaches necessitates development of a supportive data infrastructure of large and well-curated datasets that can be used reliably. This talk will comprise of two parts focused on deep learning for data acquisition and automated data curation for electron microscopy.

In the first part, we will present a dynamic sampling method based on supervised learning algorithm and convolutional neural networks for data acquisition in a SEM. In conventional point-based scanning modalities for imaging or spectroscopy, each pixel measurement can take up to a few seconds, which can translate into several hours of data acquisition time for large image sizes (e.g. 2048 x 2048 pixels). This is often true for energy-dispersive X-ray spectroscopy (EDX) in a scanning electron microscope (SEM) which is widely utilized in materials science for determining elemental compositions. Furthermore, due to longer dwell times, the sample is exposed to high energy electrons which can result in radiation damage. We will demonstrate the results for two modalities: (1) secondary electron imaging, and (2) EDX mapping. For SE imaging, we have developed a method using deep neural networks to predict the optimal sampling locations based on a set of training images, and then reconstruct the final image. The network can be pre-trained using generic images available online and can perform satisfactorily by reducing the samples to as low as 40% of available pixels. For EDX mapping, we have developed method using convolutional neural networks that uses a dictionary for training and classification of the EDX spectra. We show that this method can achieve high quality elemental maps with as low as 20% sampling. We will discuss the experimental implementation of these algorithms for smarter data acquisition resulting in reduced time and radiation exposure of the sample.

In the second part of the talk, we will demonstrate an automated data curation workflow for electron microscopy that imposes minimal burden on users for additional information yet collects data in a form amenable to automated analysis and machine learning. This workflow is developed using the Materials Data Facility (MDF). We will discuss the implementation of the workflow for a multi-user transmission electron microscope facility. Our approach allows the end-user to create a record entry for their datasets, search through their data easily, and streamlines the final publication ready data to be shared. Future work will involve consolidation of records from multiple instruments.

Accelerating X-ray absorption spectroscopy characterization by high-throughput computations and machine learning

Chi Chen, The University of California, San Diego

X-ray absorption spectroscopy (XAS) is a powerful technique for characterizing atomic environment in materials. The strong signal component of XAS, namely the X-ray absorption near edge structure (XANES), is challenging to analyze, partly due to the lack of large reference databases and quantitative tools. Using the FEFF9 XAS calculation software, we have constructed the largest XANES database containing more than 800,000 site-wise K-edge XANES spectra. The calculated materials cover most materials in the Materials Project. Leveraging this database as reference, we build similarity-based models to rapidly find compounds given a spectrum. In addition, supervised machine learning models are constructed to predict atomic coordination environments directly from the spectral information. The models show >80% accuracy in both computational and experimental XANES identifications. Finally, the model interpretability study shows the importance of having full XANES information in identifying the atomic coordination environment.