

Applied Surface Science

Room 101B - Session AS+SS-TuA

Data Analytics in Surface Science and Nanoscience

Moderators: Anders Mikkelsen, Lund University, Sweden, Petra Reinke, University of Virginia

2:20pm AS+SS-TuA1 Fast Strain Mapping of Nanowire Light-Emitting Diodes Using Nanofocused X-ray Beams, T. Stankevicius, Copenhagen University, Denmark; *U. Johansson, L. Samuelson*, Lund University, Sweden; *G. Falkenberg*, DESY, Hamburg, Germany; *R. Feidenhans'l*, Copenhagen University, Denmark; **Anders Mikkelsen**, Lund University, Sweden

Nanofocused X-ray beams are nondestructive probes that uniquely allow direct measurements of the nanoscale strain distribution and composition found at the interfaces and surfaces inside the micrometer thick layered structures of many electronic device architectures [1]. While the method has generally been considered time consuming, we demonstrate that by special design of X-ray nanobeam diffraction experiment we can (in a single 2D scan with no sample rotation) measure the individual strain and composition profiles of many structures in an array of upright standing nanowires [2]. We make use of the observation that in the generic nanowire device configuration, which is found in high-speed transistors, solar cells, and light-emitting diodes, each wire exhibits very small degrees of random tilts and twists toward the substrate. Although the tilt and twist are very small, they give a new contrast mechanism between different wires. In the present case, we image complex nanowires for nanoLED fabrication and compare to theoretical simulations, demonstrating that this fast method is suitable for real nanostructured devices.

We then go on to discuss the complications of data analysis as the amount of data available is dramatically increased with the advent of new highly coherent synchrotrons such as MAX IV in Lund Sweden [3] and improved experimental setups [2,4,5]. Using several detectors that give both real space fluorescence and 2D diffraction information combined with scanning both translational, rotational and time coordinates for in operando and in-situ studies in 3D - an enormous multidimensional dataset can be created in a few days. To fully retrieve all the information inside such dataset and pushing resolution and sensitivity limits new computational methods are needed in combination with advanced modelling.

[1] E. Lind et al., IEEE J. El. Dev. Soc. 3, 96 (2015); J. Wallentin et al., Science 339, 1057 (2013).

[2] T. Stankevicius et al, ACS Nano 9 (2015) 6978

[3] "Ultimate upgrade for US synchrotron", Nature 501 (2013) 148

[4] U. Johansson, U. Vogt, A. Mikkelsen, Proc. SPIE 8851, X-Ray Nanoimaging: Instruments and Methods, 88510L (September 26, 2013); doi:10.1117/12.2026609

[5] T. Stankevicius et al. Appl. Phys. Lett. 107 (2015) 103101

2:40pm AS+SS-TuA2 Bellerophon Environment for Analysis of Materials (BEAM), A High Performance Computing Workflow Platform for Materials Research, E.J. Lingerfelt, A. Belianinov, E. Endeve, Oak Ridge National Laboratory; *O.S. Ovchinnikov*, Vanderbilt University; *S. Somnath, R.K. Archiblad, S.V. Kalinin, Stephen Jesse*, Oak Ridge National Laboratory

Improvements in scientific instrumentation allow imaging at mesoscopic to atomic length scales, many spectroscopic modes, and now—with the rise of multimodal acquisition systems and the associated processing capability—the era of multidimensional, informationally dense data sets has arrived. Technical issues in these combinatorial scientific fields are exacerbated by computational challenges best summarized as a necessity for drastic improvement in the capability to transfer, store, and analyze large volumes of data. The Bellerophon Environment for Analysis of Materials (BEAM) platform provides material scientists the capability to directly leverage the integrated computational and analytical power of High Performance Computing (HPC) to perform scalable data analysis and simulation via an intuitive, cross-platform client user interface. This framework delivers authenticated, "push-button" execution of complex user workflows that deploy data analysis algorithms and computational simulations in HPC environments like Titan at the Oak Ridge Leadership Computing Facility (OLCF).

Here, we address the underlying HPC needs for characterization in the material science community, elaborate how BEAM's design and infrastructure tackle those needs, and present a small sub-set of user cases where scientists utilized BEAM across a broad range of analytical Tuesday Afternoon, November 8, 2016

techniques and analysis modes. BEAM system will be demonstrated for 4D Ronchigram analysis and property extraction of atomically resolved STEM (Scanning Transmission Electron Microscopy) data, parallel spectroscopic curve fitting in SPM (Scanning Probe Microscopy) data, and image segmentation.

Acknowledgements

This work is partially supported by the Laboratory Directed Research and Development (LDRD) program at ORNL, which is managed by UT-Battelle, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC05-00OR22725 (E.J.L., A.B., E.E., O.O., S.S., C.T.S., S.V.K., M.S., and S.J.). This research was conducted at the Center for Nanophase Materials Sciences and the Spallation Neutron Source, which are DOE Office of Science User Facilities. Research by J.M.B. is supported by the Center for Accelerating Materials Modeling (CMMM), which is funded by DOE Basic Energy Sciences under FWP-3ERKCSNL. This research used resources of ORNL's Compute and Data Environment for Science (CADES) and the Oak Ridge Leadership Computing Facility (OLCF), which are supported by the Office of Science of the U.S. Department of Energy under Contract No. DEAC05-00OR22725. The mathematical aspects were sponsored by the applied mathematics program at the DOE by the ACUMEN project.

3:00pm AS+SS-TuA3 The Center for Advanced Methods for Energy Research Applications (CAMERA): Mathematical Methods for Data Science from Experimental Facilities, James Sethian, University of California at Berkeley

INVITED

The Center for Advanced Methods for Energy Research Applications (CAMERA), jointly funded by the U.S. Department of Energy Offices of Advanced Scientific Research (ASCR) and Basic Energy Sciences (BES), focuses on mathematical models, algorithms, and codes that analyze, interpret, and understand the information contained within experimental data, particularly arising from light sources and nanoscale facilities. Initial focus areas include ptychography, tomography, grazing incidence small-angle scattering, image analysis and reconstruction methods, fluctuation scattering, single particle imaging, fast electronic structure methods, and automatic materials characterization and design. In this talk, we will describe the structure of CAMERA, and summarize some of the major projects. In particular, we will discuss work on: (1) Algorithms for real-time streaming ptychography. Ptychographical phase retrieval is a non-linear optimization problem, made tractable through exploiting redundancy inherent in obtaining diffraction patterns from overlapping regions of the sample. Here, we describe SHARP: our "Scalable Heterogeneous Adaptive Real-time Ptychography" framework that enables high-throughput streaming analysis. (2) New algorithms for fluctuation scattering and single particle imaging: In single particle diffraction (SPD) imaging, a large number of X-ray diffraction images are collected from individual particles, which are delivered to an ultrabright X-ray beam at random and unknown orientations through either a liquid droplet or aerosol delivery system. Recently, a new mathematical and algorithmic procedure has been introduced, known as "Multi-tiered Iterative Phasing" (MTIP), which simultaneously determines the orientations, 3D intensity function, complex phases, and the underlying molecular structure together in a single iterative process. (3) Machine learning methods for classification and characterization of scattering patterns. Grazing Incidence Small Angle X-ray Scattering (GISAXS) is an important reciprocal-space imaging modality which provides statistical information about a sample in 3-D. GISAXS is widely used for studying thin films that play a vital role as building blocks for the next generation of renewable energy technology. One challenge in GISAXS imaging is to be able to accurately infer properties of the material such as the crystal lattice corresponding to the sample from a single 2-D diffraction/scatter patterns. We will discuss our work using machine learning algorithms and convolution neural net classifiers to automatically provide structural details about the sample by analyzing the measured GISAXS diffraction patterns.

4:20pm AS+SS-TuA7 New Data Analysis Tools for X-ray Photoelectron Spectroscopy (XPS) and Spectroscopic Ellipsometry (SE), Matthew Linford, B. Singh, J. Bagley, Brigham Young University; *J. Terry*, Illinois Institute of Technology; *A. Herrera-Gomez*, CINVESTAV-Unidad, Mexico

Here we discuss a series of new data analysis tools for X-ray photoelectron spectroscopy (XPS) and spectroscopic ellipsometry (SE). For XPS, these include uniqueness plots, and the equivalent and autocorrelation widths. For SE, they include distance, principal component, and cluster analyses. Uniqueness plots are widely used in the SE community for identifying correlation between fit parameters. They are easily interpreted. However, they appear not to have been employed for XPS data analysis. And

certainly better tools are needed to identify inappropriate peak fits to XPS narrow scans because (i) XPS is now receiving in excess of 10,000 mentions in the literature each year, and (ii) with the proliferation of the technique, the number of untrained users that are collecting and fitting data has significantly increased. In a number of reported peak fits, too many fit parameters have been introduced into the data modeling, which has reduced or eliminated the statistical meaning of these parameters. Uniqueness plots show the error of a fit as a function of one of the variables in that fit, where the values of a specified variable are systematically fixed to quantities about its optimal value. If the same, low error can be obtained for all the values of the variable in question, a horizontal line is obtained, which signals fit parameter correlation. Here, the same error is obtained because other variables in the fit can compensate for the systematic change to the variable in question. In contrast, if the error in the fit rises as the variable in question is systematically changed about its optimal value, the fit has uniqueness. Uniqueness plots that indicate the absence of fit parameter correlation are often parabolic in shape. We have applied uniqueness plots to the peak fitting of XPS C 1s narrow scans of ozone-treated carbon nanotube (CNT) forests that were obtained as part of a study on CNT-templated thin layer chromatography plates, and Si 2p narrow scans of oxidized silicon. In both cases, uniqueness plots showed that unconstrained fits had poor uniqueness, while more reasonably constrained fits had better uniqueness. These results indicate that uniqueness plots may be a valuable tool for identifying inappropriate peak fits in XPS. In this presentation, I will also briefly mention the use of the equivalent and autocorrelation widths in analyzing XPS narrow scans, and then focus on distance, principal component, and cluster analyses in SE data analysis. Our recent (2016) paper on this topic appears to be only the second example of the application of chemometrics to SE data analysis in the literature.

4:40pm AS+SS-TuA8 A Surface Investigation of Parchments using ToF-SIMS and Principle Component Analysis, Marie-Laure Abel, J.F. Watts, V. Vilde, University of Surrey, UK

Parchments are an historical writing support mostly used during the Middle Ages. Their popularity dates from the second century before Christ (BC) in Pergame, Turkey, from which the name originates. Unlike paper, parchment is made of animal skin with a process similar to that used to produce leather. The products used in the fabrication vary and any animal species can be used, although most historical parchments are made from sheep, goat and calf. Information of species recognition on parchments is currently provided either using proteomics or DNA analysis. However each technique presents difficulties and sometimes it is not possible to obtain an unambiguous result. Many valuables manuscripts are written on parchment such as the Magna Carta or the Codex Sinaiticus, which justifies the effort put towards the study of this material in order to improve the conservation process and to learn more about its history.

In this work, a new technique was used in order to assess if any information may be gleaned and help in the process of recognition or even providing any further information to conservators to be used for preservation of historical parchment. Time of flight secondary ion mass spectrometry (ToF-SIMS) has been applied to the analysis of parchment specimens. Indeed while ToF-SIMS has been previously applied to a variety of samples of some significance in the cultural heritage field such as paintings or mummies, it has not been applied to parchments. To facilitate the data treatment process, this has been coupled with data analysis using chemometrics, namely principle component analysis (PCA).

A series of specimens of various ages and species were analysed on both sides, "skin" and "flesh". These samples included sheep, goat and calf. In addition, an unknown sample was also introduced to ascertain if its characteristics could be shown to be close to any species. Results indicate that it is fairly straightforward to distinguish between goat and sheep while calf is more difficult to separate from other species which is unexpected as biologically goat and sheep are considered the closest species within the selection. Furthermore the unknown specimen exhibits data which would classify it as a goat specimen. Considering the sides examined separations are seen within one particular species but the direction of the variation is not the same from one species to another. More work is needed to ascertain which side is being analysed for any unknown materials as the behaviour varies amongst the species examined in this work.

5:00pm AS+SS-TuA9 Multivariate Analysis of Very Large Hyperspectral SIMS Datasets: What Can We Do, and What Would We Like to Do?, Henrik Arlinghaus, ION-TOF GmbH, Germany

INVITED

Advances in instrumentation capabilities, as well as increases in the complexity of modern materials have resulted in a corresponding increase in the size and complexity of data acquired during sample analysis. The increase in the spatial and spectral resolution of the instrumentation is nominally a boon to the analyst, as the measured data more accurately depicts the sample. However, the resulting hyperspectral images routinely consist of upwards of ten thousand pixel spectra for 2D analyses (e.g. a 128x128 pixel image), or millions of voxel spectra for 3D analyses, each of which may consist of hundreds or thousands of ion peaks. Because of the sheer amount of information contained within such an image, it is often no longer feasible to conduct a full manual analysis of the data. An additional factor exacerbating this issue is the fact that many studies necessitate the analysis of a series of spatially resolved replicate measurements of a single sample, or of multiple similar samples. In these studies the aim is not only to characterize the contents of each individual measurement, but also to determine the similarities and differences between the measurements, while ignoring subtle differences caused by changes in analysis conditions between the individual measurements.

A solution to the problem of information overload is the use of multivariate analysis techniques to help guide the analyst, in order to reduce the time needed for determining the chemical make-up of the analyzed samples. These techniques use different approaches in order to reduce the dimensionality of the measured data, resulting in a small set of factors which recreate a simplified model of the data.

The use of MVA approaches, such as Principal Component Analysis (PCA) and Maximum Autocorrelation Factors (MAF), has become an established method of simplifying the analysis of SIMS data arising from a single measurement. We will discuss alternatives to these commonly used methods, including new variations of Multivariate Curve Resolution (MCR) which use additional optimization criteria, as well as MVA approaches not commonly used in SIMS data analysis. Additionally, we will discuss the unique challenges which may arise when applying MVA techniques to the full hyperspectral data contents of a series of measurements.

5:40pm AS+SS-TuA11 High mass-resolution 3D ToF-SIMS: PCA and visualization in seconds using Graphical Processor Units (GPUs), Peter Cumpson, I.W. Fletcher, N. Sano, A.J. Barlow, Newcastle University, UK

Multivariate analysis offers the exciting prospect of unlocking the information content of 3D SIMS of complex organic and biological samples with sub-micron resolution. However applying principal component analysis (PCA) to large images or 3D imaging depth-profiles has been difficult until now because of the Gb to Tb size of the matrices of data involved. The result has always been an "out of memory" error.

Recently[1] we applied two algorithms, RV1 and RV2, originally developed by Halko *et al*[2] that improve the speed of PCA and allow datasets of unlimited size respectively, even on ordinary personal computers. In this presentation we show results of applying these algorithms to perform PCA on full 3D ToF-SIMS data of several examples of plant and small animal tissue. The datasets we process in this way are typically 128x128 or 256x256 pixel depth-profiles of around 100 layers, each voxel having a 70,000 value mass spectrum associated with it, giving datasets of at least 1Tb in size when uncompressed. These data were acquired using our Ionoptika J105 and IonTOF IV instruments, with Helium Ion Microscope images of particular key features.

Even for such large datasets a rapid PCA calculation is often needed during analysis sessions to inform decisions on the next analytical step. We have therefore implemented the RV1 algorithm on a PC having a Graphical Processor Unit (GPU) card containing 2,880 individual processor cores[3]. This increases the speed of calculation by a factor of around 4 compared to what is possible using the fastest commercially-available desktop PCs, and full PCA is now performed in less than 7 seconds.

We then use the GPU to allow real-time interactive visualization of the principal components in 3D. This leads to some spectacular and information-rich tomographic images that can be an excellent basis for discussion between analysts and the biologists and medics who understand the morphology and anatomy of their tissue samples.

[1] P J Cumpson *et al*, *Surf. and Interface Anal.* **47** (2015) 986-993.

[2] N P Halko *et al*, *SIAM Review, Survey Rev. Sec. 53* (2011) 217-288.

[3] P J Cumpson *et al*, *Surf. and Interface Anal.*, onlinelibrary.wiley.com/doi/10.1002/sia.6042/full

Tuesday Afternoon, November 8, 2016

6:00pm **AS+SS-TuA12 Mass Spectrometry Image Fusion, Bonnie June Tyler**, Universität Münster, Germany; *H.F. Arlinghaus*, University of Muenster, Germany

As mass spectrometry imaging (MSI) has moved from the technique development stage into real world biological studies, the need to combine mass spectrometry images with other biologically relevant imaging techniques has become important. Techniques as diverse as electron microscopy, scanning probe microscopy, XPS imaging, H&E staining, and fluorescent labeling can provide important information that is complementary to the mass spectral images. Combining the information from these complementary measurements is often necessary for accurate understanding of biological samples. Within the field of mass spectrometry imaging alone, combining different imaging modes, such as MALDI/ToF-SIMS, or GCIB ToF-SIMS/LMIG ToF-SIMS, can enhance understanding of the specimens being studied.

In theory, more data should enable more confident conclusions. In practice, however, the challenges of handling and reducing very large imaging data sets, that have disparities in spatial resolution and contrast mechanisms, can result in biased or misleading conclusions. In order to facilitate more consistent, accurate and useful descriptions of real world samples, advanced data exploration tools are needed. Image fusion is an approach to combining data from different sources that is receiving increasing attention within the field of mass spectrometry imaging.

Although many algorithms for image fusion have been developed for applications in remote sensing, medical imaging and photography, the distinctive features of mass spectrometry make many of these techniques inappropriate for use in this field. We have tested algorithms from two major classes of image fusion, those that operate in the spatial domain and those that operate in the frequency domain. Common artefacts caused by the different algorithms have been identified. Two modified algorithms have been developed which can be used to produce satisfactory fused images using mass spectrometry data. The first approach combines multivariate analysis (MVA) and discrete cosine transform (DCT) and is useful for combining MSI images with monochromatic images. The second algorithm, which uses a combination of multivariate methods, is useful for fusing MSI data with a second spectral image. Both of these new image fusion approaches have been tested on simulations, model systems and real tissue samples. We have shown that MVA image fusion can be a valuable technique for reducing noise, improving image contrast and enhancing the sharpness of mass spectrometry images. With appropriate attention to the distinctive features of each imaging method, image fusion can be done without significant artefacts or distortion of the spectral detail.

Author Index

Bold page numbers indicate presenter

— A —

Abel, M.L.: AS+SS-TuA8, **2**

Archiblad, R.K.: AS+SS-TuA2, **1**

Arlinghaus, H.F.: AS+SS-TuA12, **3**; AS+SS-TuA9, **2**

— B —

Bagley, J.: AS+SS-TuA7, **1**

Barlow, A.J.: AS+SS-TuA11, **2**

Belianinov, A.: AS+SS-TuA2, **1**

— C —

Cumpson, P.J.: AS+SS-TuA11, **2**

— E —

Endeve, E.: AS+SS-TuA2, **1**

— F —

Falkenberg, G.: AS+SS-TuA1, **1**

Feidenhans'l, R.: AS+SS-TuA1, **1**

Fletcher, I.W.: AS+SS-TuA11, **2**

— H —

Herrera-Gomez, A.: AS+SS-TuA7, **1**

— J —

Jesse, S.: AS+SS-TuA2, **1**

Johansson, U.: AS+SS-TuA1, **1**

— K —

Kalinin, S.V.: AS+SS-TuA2, **1**

— L —

Linford, M.R.: AS+SS-TuA7, **1**

Lingerfelt, E.J.: AS+SS-TuA2, **1**

— M —

Mikkelsen, A.: AS+SS-TuA1, **1**

— O —

Ovchinnikov, O.S.: AS+SS-TuA2, **1**

— S —

Samuelson, L.: AS+SS-TuA1, **1**

Sano, N.: AS+SS-TuA11, **2**

Sethian, J.A.: AS+SS-TuA3, **1**

Singh, B.: AS+SS-TuA7, **1**

Somnath, S.: AS+SS-TuA2, **1**

Stankevic, T.: AS+SS-TuA1, **1**

— T —

Terry, J.: AS+SS-TuA7, **1**

Tyler, B.J.: AS+SS-TuA12, **3**

— V —

Vilde, V.: AS+SS-TuA8, **2**

— W —

Watts, J.F.: AS+SS-TuA8, **2**