New Challenges to Reproducible Data and Analysis Focus Topic
Room A211 - Session RA+AS+NS+SS-MoA

Quantitative Surface Analysis II/Big Data, Theory and Reproducibility

Moderators: Kateryna Artyushkova, Physical Electronics, Donald Baer, Pacific Northwest National Laboratory

1:40pm **RA+AS+NS+SS-MoA1 A Data-Centric View of Reproducibility**, *Anne Plant,* National Institute of Standards and Technology (NIST)*; J. Elliott,* NIST*; R. Hanisch,* National Institute of Standards and Technology (NIST) **INVITED**

Ideally, data should be shareable, interpretable, and understandable within the scientific community. There are many challenges to achieving this, including the need for high quality documentation and a shared vocabulary. In addition, there is a push for rigor and reproducibility that is driven by a desire for confidence in research results. We suggest a framework for a systematic process, based on consensus principles of measurement science, to guide researchers and reviewers in assessing, documenting, and mitigating the sources of uncertainty in a study. All study results have associated ambiguities that are not always clarified by simply establishing reproducibility. By explicitly considering sources of uncertainty, noting aspects of the experimental system that are difficult to characterize quantitatively, and proposing alternative interpretations, the researcher provides information that enhances comparability and reproducibility.

2:20pm **RA+AS+NS+SS-MoA3 Enhancing Data Reliability, Accessibility and Sharing using Stealthy Approaches for Metadata Capture**, *Steven Wiley,* Pacific Northwest National Laboratory **INVITED**

Science is entering a data-driven era that promises to accelerate scientific advances to meet pressing societal needs in medicine, manufacturing, clean energy and environmental management. However, to be usable in big data applications, scientific data must be linked to sufficient metadata (data about the data) to establish its identity, source, quality and reliability. This has also driven funding agencies to require projects to use community-based data standards that support the FAIR principles: Findable, Accessible, Interoperable, and Reusable. Current concerns about data reproducibility and reliability have further reinforced these requirements. Truly reusable data, however, requires an enormous amount of associated metadata, some which is very discipline and sample-specific. In addition, this metadata is typically distributed across multiple data storage modalities (e.g. lab notebooks, electronic spreadsheets, instrumentation software) and is frequently generated by different people. Assessing and consolidating all of the relevant metadata has traditionally been extremely complex and laborious, requiring highly trained and motivated investigators as well as specialized curators and data management systems. This high price has led to poorly documented datasets that can rarely be reused. To simplify metadata capture and thus increase the probability it will indeed be captured, EMSL (Environmental Molecular Sciences Laboratory) has developed a general-purpose metadata capture and management system built around the popular ISA-Tab standard (Investigation-Study-Assay Tables). We have modified this framework by mapping it onto the EMSL workflow, organized as a series of "transactions". These transactions are natural points where metadata is generated, include specifying how samples will be generated and shipped, instrument scheduling, sample storage, and data analysis. Software tools have been built to facilitate these transactions, automatically capture the associated metadata and link it to the relevant primary data. This metadata capture system works in concert with automated instrument data downloaders and is compatible with commercial sample tracking and inventory management systems. By creating value-added tools that are naturally integrated into the normal scientific workflow, our system enhances scientific productivity, thus incentivizing adoption and use. The entire system is designed to be general purpose and extensible and thus should be a useful paradigm for other scientific projects that can be organized around a transactional model.

3:00pm **RA+AS+NS+SS-MoA5 From Electrons to X-rays: Tackling Big Data Problems through AI**, *Mathew Cherukara,* Y. Liu, M.V. Holt, H. Liu, T.E. Gage, J.G. Wen, I. Arslan, Argonne National Laboratory **INVITED**

As microscopy methods and detectors have advanced, the rates of data acquisition and the complexity of the acquired data have increased, and these are projected to increase several hundred-fold in the near future. The unique electron and X-ray imaging capabilities at the Center for Nanoscale Materials (CNM) are in a position to shed light on some of the most challenging and pressing scientific problems we face today. To fully leverage the capability of these advanced instruments, we need to design and develop effective strategies to tackle the problem of analyzing the data generated by these imaging tools, especially following facility upgrades such as the upgrade to the Advanced Photon Source (APS-U) and the commissioning of the ultrafast electron microscope (UEM).

The data problem is especially acute in the context of coherent imaging methods, ultra-fast imaging and multi-modal imaging techniques. However, analysis methods have not kept pace. It is infeasible for a human to sort through the large, complex data sets being generated from imaging experiments today. At the CNM, we apply machine learning algorithms to our suite of electron and X-ray microscopy tools. Machine learning workflows are being developed to sort through data in real-time to retain only relevant information, to invert coherently scattered data to real-space structure and strain, to automatically identify features of interest such as the presence of defects, and even to automate decision making during an imaging experiment. Such methods have the potential not only to decrease the analysis burden on the scientist, but to also increase the effectiveness of the instruments, for instance by providing real-time experimental feedback to help guide the experiment.

4:00pm **RA+AS+NS+SS-MoA8 Quantifying Shell Thicknesses of Core-Shell Nanoparticles by means of X-ray Photoelectron Spectroscopy**, *Wolfgang Werner,* Vienna University of Technology, Austria **INVITED**

Determining shell thicknesses and chemistry of Core-Shell Nanoparticles (CSNPs) presently constitutes one of the most important challenges related to characterisation of nanoparticles. While for particlae number concentration various routine analysis techniques as well as methods providing reference measurements have been or are in the process of being developed, one of the most promising candidates for shell thickness determination is x-ray photoelectron spectroscopy (XPS).

Different approaches to quantify shell thicknesses will be presented and compared. These comprise: (1) The infinitesimal columns model (IC), (2) Shard's empirical formula (TNP-model) and (3) SESSA (Simulation of Electron Spectra for Surface Analysis) simulations with and (4) without elastic scattering.

CSNP XPS intensities simulated with SESSA for different combinations of core/shell-material combinations for a wide range of core and shell thicknesses have been evaluated with the TNP-model and the retrieved thicknesses are in good agreement with the nominal thickness, even when elastic scattering is turned on during the simulation, except for pathological cases. For organic shell materials these simulations fully confirm the validity of the (much simpler) TNP-method, which also coincides with the IC model.

Experimental data on of a round robin experiment of PMMA@PTFE CSNPs involving three research institutions were analysed with the aforementioned approaches and show a good consistency in that evaluations of the shell thicknesses among the institutions agree within 10% (and are in good agreement with the nominal shell thickness). This consistency is promising since it suggests that the error due to sample preparation can be controlled by following a strict protocol.

Use of the F1s signal leads to significant deviations in the retrieved shell thickness. Independent measurements using Transmission Electron Microscopy were also performed, which revealed that the core-shell structure is non-ideal, i.e. the particles are aspherical and the cores are acentric within the particles. SESSA simulations were employed to estimate the effect of various types of deviations of ideal NPs on the outcome of shell thickness determination.

The usefulness and importance of different kind of electron beam techniques for CSNP analysis and in particular shell thickness determination is discussed.

4:40pm RA+AS+NS+SS-MoA10 Modeling the Inelastic Background in X-ray Photoemission Spectra for Finite Thickness Films, *Alberto Herrera-Gomez*, CINVESTAV-Unidad Queretaro, México

The background signal in photoemission spectra caused by inelastic scattering is usually calculated by convolving the total signal with the electron-energy loss-function. This method, which was proposed by Tougaard and Sigmund in their classic 1982 paper [1], only works (as clearly indicated in [1]) for homogeneous materials. However, the method is commonly applied to finite thickness films. In this paper it is going to be described the proper way to remove the inelastic background signal of spectra from thin-conformal layers including buried layers and delta-doping [2]. The method is based on the straight-line inelastic scattering path, which is expected to be a very good approximation for low energy losses (near-peak regime). It is also a common practice to use the parametric Tougaard Universal Cross Section [3] with the provision that, instead of using the theoretical values for the parameters valid for homogeneous materials, the B-parameter is allowed to vary until the experimental background signal ~ 50 to 100 eV below the peak is reproduced. This is equivalent to scale the loss-function, which partially compensates the error from using the convolution method [1]. The error compensation on the modeling of the background of finite-thickness layers by scaling the loss-function will be quantitatively described.

[1] S. Tougaard, P. Sigmund, Influence of elastic and inelastic scattering on energy spectra of electrons emitted from solids, Phys. Rev. B. 25 (1982) 4452–4466. doi:10.1103/PhysRevB.25.4452.

[2] A. Herrera-Gomez, The photoemission background signal due to inelastic scattering in conformal thin layers (Internal Report), 2019. http://www.qro.cinvestav.mx/~aherrera/reportesInternos/inelastic_backgr ound_thin_film.pdf.

[3] S. Tougaard, Universality Classes of Inelastic Electron Scattering Cross-sections, Surf. Interface Anal. 25 (1997) 137–154. doi:10.1002/(SICI)1096-9918(199703)25:3<137::AID-SIA230>3.0.CO;2-L.

5:00pm RA+AS+NS+SS-MoA11 R2R(Raw-to-Repository) Characterization Data Conversion for Reproducible and Repeatable Measurements, *Mineharu Suzuki, H. Nagao, H. Shinotsuka,* National Institute for Materials Science (NIMS), Japan*; K. Watanabe,* ULVAC-PHI Inc., Japan*; A. Sasaki,* Rigaku Corp., Japan*; A. Matsuda, K. Kimoto, H. Yoshikawa,* National Institute for Materials Science (NIMS), Japan

NIMS, Japan, has been developing a materials data platform linked with a materials data repository system for rapid new material searching by materials informatics. The data conversion from raw data to human-legible/machine-readable data file is one of the key preparation techniques prior to data analysis, where the converted data file should include meta-information. Our tools can convert raw data to a structured data package that consists of (1) characterization measurement metadata, (2) primary parameters which we will not call "metadata" to distinguish from (1), (3) raw parameters as written in original raw data, and (4) formatted numerical data. The formatted numerical data are expressed as matrix type with robust flexibility, not obeying a rigid definition. This flexibility can be realized by applying the data conversion style of Schema-on-Read type, not Schema-on-Write type based on *de jure* standards such as ISO documents. The primary parameters are carefully selected from raw parameters and their vocabularies are replaced from instrument-dependent terms to general ones that everyone can readily understand. These primary parameters with linked specimen information are useful for reproducible and repeatable instrument setup. By this R2R conversion flow, we have verified that we can generate and store interoperable data files of XPS spectra and depth profiles, powder XRD patterns, (S)TEM images, TED patterns, EELS spectra, AES spectra, EPMA spectra and elemental mapping, and theoretical electron IMFP data. We have also developed a system to allow semi-automatic data transfer from an instrument-controlling PC isolated from the network, by adopting a Wi･Fi-capable SD card's scripting capability, while keeping the PC offline. We are working on further software development for on-demand data manipulation after R2R data conversion. So far it has been possible to perform XPS peak separation using automated information compression technique. Using these components, high-throughput data conversion/accumulation and data analyses are realized, where human interaction is minimized. Using metadata extracted from raw data, other users can reproduce or repeat measurements even if they did not carry out the original measurement. Human-legible and machine-readable numerical data is utilized for statistical analyses in informatics.

Author Index

Bold page numbers indicate presenter

— A —
Arslan, I.: RA+AS+NS+SS-MoA5, 1
— C —
Cherukara, M.: RA+AS+NS+SS-MoA5, **1**
— E —
Elliott, J.: RA+AS+NS+SS-MoA1, 1
— G —
Gage, T.E.: RA+AS+NS+SS-MoA5, 1
— H —
Hanisch, R.: RA+AS+NS+SS-MoA1, 1
Herrera-Gomez, A.: RA+AS+NS+SS-MoA10, **2**
Holt, M.V.: RA+AS+NS+SS-MoA5, 1

— K —
Kimoto, K.: RA+AS+NS+SS-MoA11, 2
— L —
Liu, H.: RA+AS+NS+SS-MoA5, 1
Liu, Y.: RA+AS+NS+SS-MoA5, 1
— M —
Matsuda, A.: RA+AS+NS+SS-MoA11, 2
— N —
Nagao, H.: RA+AS+NS+SS-MoA11, 2
— P —
Plant, A.: RA+AS+NS+SS-MoA1, **1**

— S —
Sasaki, A.: RA+AS+NS+SS-MoA11, 2
Shinotsuka, H.: RA+AS+NS+SS-MoA11, 2
Suzuki, M.: RA+AS+NS+SS-MoA11, **2**
— W —
Watanabe, K.: RA+AS+NS+SS-MoA11, 2
Wen, J.G.: RA+AS+NS+SS-MoA5, 1
Werner, W.S.M.: RA+AS+NS+SS-MoA8, **1**
Wiley, H.S.: RA+AS+NS+SS-MoA3, **1**
— Y —
Yoshikawa, H.: RA+AS+NS+SS-MoA11, 2