AI/ML/Autonomous Experimentation for Thin Films Processing
Room 317 - Session AIML2-ThM

AI/ML Across Materials Experimentation: Design, Characterization, Process Control, and Data Infrastructure

Moderators: Michael Nolan, Tyndall National Institute, University College Cork, **Campbell Sweet**, University of Missouri

11:00am **AIML2-ThM-13 From Noisy Data to Quantitative Structure: AI-Enabled Electron Microscopy and Spectroscopy**, *Judith Yang,* ***Brian Lee,*** *Meng Li, Dmitri Zakharov, Deyu Lu, Xiaohui Qu,* Brookhaven National Laboratory

Electron microscopy and X-ray spectroscopy increasingly generate complex datasets whose quantitative interpretation can become a bottleneck in extracting physical and chemical information. Here, we present recent AI/ML developments that move beyond conventional post-processing toward automated extraction of structural and chemical information directly from experimental measurements.

For high-resolution transmission electron microscopy (HRTEM), we demonstrate self-supervised denoising for low-dose, time-resolved imaging and introduce a new hybrid approach combining machine learning with tensor nuclear norm minimization to reconstruct atomic features under severe noise conditions. These approaches enable recovery of quantitative structural information.

For spectroscopy, we use a transformer-based model to predict Cu oxidation state from EELS and XAS L-edge spectra, enabling rapid and quantitative chemical-state analysis across simulated and experimental datasets. We are further extending AI-enabled spectroscopy from property prediction toward inverse structure determination.

Together, these developments illustrate a progression from AI-assisted data processing toward physics-informed extraction of quantitative chemical and structural information. Such approaches provide a foundation for faster interpretation of microscopy and spectroscopy measurements and, ultimately, for integrating them with real-time and autonomous experiments.

11:15am **AIML2-ThM-14 Machine Learning-Assisted Pulsed Laser Deposition Based on Plasma Optical Emission Signatures**, ***Dorien Carpenter,*** *Roman Luckett, Zahra Nasiri, Shiva Gupta, Rodion Mayatskiy,* University of Alabama at Birmingham*; Sumner Harris,* Center for Nanophase Materials Sciences, Oak Ridge National Laboratory*; Renato Camata,* University of Alabama at Birmingham

Growth of thin films by pulsed laser deposition (PLD) is governed by the formation, transport, and condensation of a transient laser-generated plasma plume. During ablation, the target is converted into a complex mixture of atoms, ions, molecules, clusters, droplets, and particulates, whose relative populations depend on target chemistry, constituent vapor pressures, and laser-material coupling. As the plume expands, its ionization state, temperature, particle flux, and kinetic energy distribution are further shaped by laser-plasma interactions, ablation geometry, and collisions with background gas. These coupled processes determine the composition, arrival rate, and energy of growth precursors reaching the substrate, thereby strongly influencing nucleation, phase formation, crystallinity, morphology, and ultimately thin film properties.

Optical emission spectra (OES) acquired during PLD encode rich information about the plasma processes that mediate thin film growth. Because OES can provide real-time, noninvasive measurements of plume composition, excitation, and plasma evolution, it offers actionable data streams that machine-learning workflows can harness to guide thin film growth on synthesis-relevant timescales. This could enable autonomous control of PLD and open new pathways for fabricating nonequilibrium materials that remain difficult to realize using conventional approaches.

In this work, we explore how representations of plasma temperature (*T*) and ionization fraction (*Xe*) can be constructed using Gaussian Process Bayesian Optimization (GPBO) from a limited number of experiments. These quantities are extracted from OES of Fe-rich plumes. A GP regression model is trained on progressively accumulating experimental data to generate surrogate models of the underlying *T* and *Xe* as a function of laser fluence and spot size. We assess the impact of different acquisition functions and GP kernels against baseline random sampling. Model performance is evaluated using synthetic datasets from coupled laser ablation-fluid dynamics simulations, where *T* and *Xe* serve as ground-truth quantities.

We discuss results for PLD of iron, over a fluence range of 2-10 J/cm^2 and spot area of 0.2-13 mm^2 obtained from 1000 independent GPBO trials initialized with random three-point seed pairs of fluence, spot area, and *T* or *Xe*. We quantify the rate of convergence of the optimization to regions of interest in the (*T*, *Xe*) space, with respect to the choice of kernel and acquisition function. We then show how the surrogate models for *T* and *Xe* evolve in time and how they can be used to guide PLD to little explored plume ionization and temperature conditions for thin film growth.

11:30am **AIML2-ThM-15 Human-in-the-loop Agentic Design of Experiments for Molecular Beam Epitaxy Synthesis of In(Ga)Sb Self-Assembled Quantum Dots in an InAs Matrix**, ***Molly McDonough***[1,2], Pennsylvania State University*; Priyanka Petluru, Aaron Muhowski,* Sandia National Laboratories*; Wesley Reinhart, Stephanie Law,* Pennsylvania State University

In(Ga)Sb self-assembled quantum dots (QDs) are a class of III-V nanostructures that have been studied primarily due to their broken-gap band alignment and emission wavelengths in the mid-wave (MWIR) and long-wave infrared (LWIR). The MWIR and LWIR regimes are crucial to many different technological areas including biosensing, trace gas analysis, environmental monitoring, and defense. However, options for compact, high efficiency emitters in this wavelength range are nascent. InSb and In(Ga)Sb self-assembled quantum dots (SAQDs) in an InAs matrix are a unique option for this wavelength regime due to their tunable emission wavelength by tailoring the lateral size and height of the SAQDs. Previous work on InSb SAQDs has yielded strong photoluminescence (PL) emission in the 3µm - 5µm range. For In(Ga)Sb SAQDs on InAs, emission out to 8µm has been successfully demonstrated.

Our work builds on latent knowledge by modeling the relationship between growth parameters, SAQD morphology and properties, and PL emission wavelength to create designer MWIR and LWIR emitters. We implement a human-in-the-loop agentic design of experiments process via tool-using large-language models to suggest growth conditions for maximal information gain per trial. The parameters explored here include how V/III beam equivalent pressure ratios, growth rate, deposition thickness, Ostwald ripening time, composition, and the use of bismuth surfactant influence our nanostructure morphology and emission characteristics. By this method, we achieve PL emission from In(Ga)Sb nanostructures in the 5µm to 7µm range. We observe the formation of In(Ga)Sb quantum dashes, which contrasts with existing studies on In(Ga)Sb quantum dots. The results presented here provide in-depth mapping of the synthesis-structure-property relationships in this system and more broadly demonstrate an effective and practical deployment of agentic hypothesis generation for synthesis exploration and device development.

SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525.

11:45am **AIML2-ThM-16 Enabling AI-Driven Materials Discovery Through Data Infrastructure**, *Akshay Talekar,* ***Harman Farwah,*** UL Research Institutes

Advances in machine learning have expanded the role of AI in materials discovery, but model development is rarely the primary constraint in practice. Instead, progress is limited by fragmented data ecosystems, inconsistent metadata, and weak integration between computational and experimental workflows.

This talk frames materials discovery as a data-centric systems problem, where the ability to ingest, standardize, and operationalize heterogeneous data ultimately determines the effectiveness of downstream learning. Emphasis is placed on the role of data architecture, provenance, and interoperability in enabling reproducible and scalable experimentation.

More broadly, this perspective highlights the importance of tightly integrated data and learning pipelines that support continuous iteration under real-world constraints. Addressing these challenges is essential for transitioning AI from isolated modeling efforts to robust, production-scale capabilities within experimental science.

[1] **AVS Nellie Yeoh Whetten Awardee**
[2] **JVST Highlighted Talk**

Author Index

Bold page numbers indicate presenter

— C —
Camata, Renato: AIML2-ThM-14, 1
Carpenter, Dorien: AIML2-ThM-14, **1**
— F —
Farwah, Harman: AIML2-ThM-16, **1**
— G —
Gupta, Shiva: AIML2-ThM-14, 1
— H —
Harris, Sumner: AIML2-ThM-14, 1
— L —
Law, Stephanie: AIML2-ThM-15, 1
Lee, Brian: AIML2-ThM-13, **1**
Li, Meng: AIML2-ThM-13, 1
Lu, Deyu: AIML2-ThM-13, 1
Luckett, Roman: AIML2-ThM-14, 1
— M —
Mayatskiy, Rodion: AIML2-ThM-14, 1
McDonough, Molly: AIML2-ThM-15, **1**
Muhowski, Aaron: AIML2-ThM-15, 1
— N —
Nasiri, Zahra: AIML2-ThM-14, 1
— P —
Petluru, Priyanka: AIML2-ThM-15, 1
— Q —
Qu, Xiaohui: AIML2-ThM-13, 1
— R —
Reinhart, Wesley: AIML2-ThM-15, 1
— T —
Talekar, Akshay: AIML2-ThM-16, 1
— Y —
Yang, Judith: AIML2-ThM-13, 1
— Z —
Zakharov, Dmitri: AIML2-ThM-13, 1