AI/ML/Autonomous Experimentation for Thin Films Processing
Room 317 - Session AIML1-ThM
Linking AI/ML Tools with Diagnostics and Film Growth: PLD/CVD/ALD

Moderators: Prof. Renato Camata, University of Alabama at Birmingham, **Dr. Gregory Doerk**, Brookhaven National Laboratory

8:00am **AIML1-ThM-1 AI-Driven Synthesis by Pulsed Laser Deposition: Autonomous Experimentation and Human-AI Collaboration***, **Sumner B. Harris,*** Center for Nanophase Materials Sciences, Oak Ridge National Laboratory **INVITED**

Thin-film synthesis with pulsed laser deposition (PLD) presents a vast design space of chemistries, structures, and non-equilibrium processing pathways that is challenging to navigate when targeting specific film properties. By combining automated synthesis platforms equipped with multimodal in situ diagnostics and a range of AI methods, we can generate testable hypotheses, autonomously explore large parameter spaces, uncover non-trivial correlations between diagnostics and film properties, and control the growth trajectory in real time. In this talk, I will discuss human–AI collaborative autonomous synthesis, which employs an LLM as a co-scientist to help generate testable hypotheses for oxide remote epitaxy and assists with additional LLM-based data analysis between batches of fully autonomous PLD experiments. I will further discuss how this example sets the stage for Agentic AI-driven synthesis and characterization platforms, incorporating cross-modal knowledge to make decisions and analyze results.This work was supported by the Center for Nanophase Materials Sciences (CNMS), which is a US Department of Energy, Office of Science User Facility at Oak Ridge National Laboratory. PLD synthesis was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division

8:30am **AIML1-ThM-3 Physics-Informed Gaussian Process Active Learning for Modeling of PLD Plume Dynamics***, **Zahra Nasiri**, Dorien Carpenter, Jacob H Paiste,* University of Alabama at Birmingham*; Sumner B Harris,* Center for Nanophase Materials Sciences, Oak Ridge National Laboratory*; Renato P Camata,* University of Alabama at Birmingham

In pulsed laser deposition (PLD), accurate models of plume particle flux and kinetic energy as functions of laser fluence and spot area are essential for autonomous thin film growth optimization. It is well established that Gaussian Process (GP) active learning can construct high-fidelity representations of these functions from a small number of strategically selected experiments. Physics-inspired structured mean functions have been identified as a promising direction for improving model performance and sample efficiency. Here, we present a fully developed physics-informed GP framework that embeds this physical structure directly into the surrogate modeling process.

We replace the conventional zero-mean GP prior with physically motivated baselines derived from an isentropic expansion model. For kinetic energy, the structured mean expresses the kinetic energy of ablated species in terms of a stagnation temperature at the beginning of expansion and the terminal Mach number of the plume. The terminal Mach number is derived from the Knudsen number for each combination of fluence and spot area through a known analytical relation involving the heat capacity ratio. The stagnation temperature is modeled through a secondary GP trained on fluid dynamics simulation data. For particle flux, the structured mean captures the scaling with the initial vapor density and the terminal Mach number, where the initial thermodynamic state is determined through the laser-target interaction. The primary GP learns only residuals from these baselines, reducing model complexity and improving sample efficiency.

Performance is evaluated via root mean square error (RMSE) and compared against zero-mean GPs and random sampling. Results from 100 independent runs confirm that embedding isentropic expansion physics dramatically reduces model error, with RMSE dropping from approximately 3.5% to 0.1%. We carry out experimental validation through real-time ion probe measurements of KrF excimer laser ablation of copper in vacuum. We will discuss how to interpret the time-dependent Mach number extracted from the surrogate model and its use to predict the angular distribution of the PLD plume, which is critical for film uniformity and thickness control. Finally, we generalize our approach to complex compound materials, with the iron-based superconductor FeSe as an initial target where stoichiometry control presents additional modeling challenges.

8:45am **AIML1-ThM-4 Multi-Objective Bayesian Optimization for Aligning Carbon Nanotube Growth Simulations with In-Situ SEM Observations***, Ramakrishna Surya, Brendan Young, Brendan Alvey, James Keller, Matthias Young, **Matthew Maschmann,*** University of Missouri-Columbia

Carbon nanotube (CNT) forests exhibit exceptional potential for lightweight conductors, interconnects, thermal interfaces, and multifunctional materials, yet their effective properties remain far below those of individual CNTs because forest-scale behavior is governed by stochastic growth, CNT–CNT interactions, delamination, density loss, and evolving microscale morphology. This work presents an experiment-informed computational framework for matching numerical simulations of CNT forest growth with in-situ environmental SEM observations and simultaneous electrical measurements. CNT forests are synthesized across pre-patterned electrodes, enabling real-time imaging of morphological evolution while measuring conductance during growth. Image analysis and machine-learning-based feature tracking are used to extract CNT density and growth kinematics from SEM image sequences, providing quantitative structural targets for model calibration.

A three-dimensional finite element model captures CNT growth, mechanical interactions, van der Waals contact formation, delamination, and electrical network evolution. Because key model parameters, including growth-rate variability, CNT–CNT contact resistance, intrinsic CNT resistance, critical delamination stress, and initial CNT areal density, are difficult to measure directly, we use a multi-objective Bayesian optimization to efficiently calibrate the simulation. Gaussian process surrogate models guide active learning toward parameter combinations that minimize both density error and conductance error relative to experiment. This approach identifies Pareto-optimal parameter sets using fewer than 50 simulations from a search space exceeding 50,000 combinations. The resulting calibrated simulations reproduce experimentally observed density decay and conductance evolution, capturing the coupled effects of CNT junction formation, network degradation, and delamination. This work establishes a pathway for calibrating digital twins with experimental observations by combining in-situ measurement, physics-based simulation, and multi-objective active learning to infer latent model parameters and improve predictive fidelity.

9:00am **AIML1-ThM-5 QCMPy: Open-Source Python Codebase for in-Situ QCM Kinetics Monitoring***, **Eden Goodwin**, Maram Bakiro, Victoria Velez, Derek Reid, Seán Barry,* Carleton University, Canada

It is well known that not all chemical systems are equally prepared for modelling with Artificial Intelligence (AI). Well documented chemical systems with predictable structures and invariable synthesis have yielded phenomenal modelling results, as is best exemplified by the Nobel protein modelling AI software AlphaFold.[1] Small molecule systems are more difficult to model as they have inconsistent documentation, more dimensions of structural diversity, and significant variability in synthetic approach. Structural modifications can introduce substantial changes to the synthetic procedure, new decomposition pathways, intermolecular interactions, or even unpredicted functionality at the surface. These factors make iteration on and testing of models challenging. As such a different approach may yield better results: extraction of critical information from under-analyzed data.

A prime target for this data collection is an under utilized, robust, affordable, in-situ metrology: Quartz Crystal Microbalance (QCM). QCMs are used within Atomic Layer Processing (ALP) to measure in-situ the growth, etch, and modification of surfaces as they are exposed to highly reactive precursors. The relative mass gain due to each step of a sequential ALPs can provide valuable chemical insights, especially when coupled with other in-situ techniques. Careful kinetics analysis of sequential adsorptions of N-Heterocyclic Carbenes on surfaces revealed a previously unreported phenomenon: Atomic Layer Restructuring.[2] A reduction in the maximal coverage of the physiosorbed component revealed a significant reduction in the number of available surface sites and consequently surface roughness.

To translate kinetics modelling to a broader audience we have developed an open-source tool "*QCMPy*" that (1) separates a QCM trace of into cycle by cycle traces for each precursor then (2) aligns cycle traces using pulse detection algorithms and finally (3) conducts comprehensive cycle by cycle adsorption modelling. To demonstrate QCMpy we probe the growth kinetics of trimethyl aluminum using three different co-reactants: Water, Fumaric Acid and Salicylic acid. Analysis of the modeled parameters (ka, θmax) over each individual Molecular Layer Deposition cycle reveal pulsing

drift and variability in our ALD-reactor, predict the transition from interfacial to bulk growth (Figure 1), predict the ideal growth per cycle from sub-saturative recipes and quantify the chemisorption/physisorption ratio of TMA and all co-reactants.

References

1- Jumper, J. et. al. *Nature* **2021**, *596* (7873), 583–589.

2- Goodwin, E. et al. *ACS Nano* **2025**, *19* (16), 15617–15626.

9:15am **AIML1-ThM-6 Accelerating Peald Process Development with Explainable Machine Learning***,* ***Hamidur Rahman,*** University of Michigan, Ann Arbor*; Ian Campbell,* IMEC USA*; Paula Arellano-Vasquez, Joshua Kammeraad, Paul Zimmerman, Ageeth Bol,* University of Michigan, Ann Arbor

Plasma-enhanced atomic layer deposition (PEALD) has emerged as a promising route to synthesize emerging 2D materials, yet rationalizing its highly nonlinear, multi-parameter process space remains an open challenge. Machine Learning (ML) promises to address this challenge, but requires consensus on which models, validation schemes, and feature sets are appropriate for the relatively high-noise, small data sets that are typically generated in PEALD studies. We address this issue by a systematic, end-to-end methodology study to optimize the crystallinity and surface morphology of PEALD MoS_2. Building on a critical survey of prior approaches, we systematically evaluate ML practices on our own controlled experimental data set, collected through experiments designed to optimize material properties, to clarify which ML models reliably translate to the small-sample, high-noise regime characteristic of PEALD.

A PEALD process to synthesize MoS_2 is developed using metal $Mo(N^tBu)_2(NMe_2)_2$ precursor and combinations of Ar/H_2 plasmas and a tert-butyl disulfide co-reactant. Various ALD parameters, such as process pressure, plasma power, gas composition, etc., were systematically varied to guide the film growth towards more crystalline and minimal out-of-plane growth using Atomic Force Microscopy (AFM). The data were subsequently modeled using eight algorithms encompassing linear regressors, k-nearest neighbors, random forest variants, and gradient boosting methods. A comparative evaluation of these models is conducted to explore the relationships inherent in PEALD process parameters and forecast material properties. Incorporating Raman Spectroscopy features as auxiliary inputs was also explored, highlighting the potential of multimodal data in low-data ALD studies. Model interpretability techniques were applied to examine the contributions of individual steps in the PEALD cycle to material properties, providing physically meaningful insights that complement human interpretation of the underlying chemistry. The trained models drive inverse recipe design, yielding candidate process windows for target film properties. We further explore in-situ spectroscopic ellipsometry as a real-time surrogate for ex-situ characterization, paving the way toward closed-loop control. Finally, we outline a roadmap for a transferable PEALD digital twin, where MoS_2-trained models can be adapted via transfer learning to analogous precursor systems, laying the foundation for autonomous, ML-guided PEALD.

9:30am **AIML1-ThM-7 Non-Negative Matrix Factorization for in Situ Gas Analysis During Atomic Layer Deposition***,* ***Andreas Werbrouck,*** *Aditya Chalishazar, Dirk Poelman, Philippe F. Smet, Jolien Dendooven, Christophe Detavernier,* Ghent University, Belgium

In situ characterization of atomic layer deposition (ALD) processes enables the study of reaction mechanisms, kinetics, and industrial process control. An established way to monitor gas species in a reactor is quadrupole mass spectrometry (QMS). In past work, we have shown that time-resolved, full-range mass spectrometry data with very good sensitivity can be collected provided that steady-state reaction conditions are maintained. Another method for gas-phase characterization is capturing the optical emission from a cold cathode discharge (optical emission spectroscopy or OES). OES could quickly detect and characterize gas-phase constituents under non-steady-state conditions. Both methods yield dense, time-resolved series of spectra, showing multiple correlated and potentially overlapping peaks. Therefore, we adopt non-negative matrix factorization (NMF), a machine learning technique, as a holistic way to analyze the obtained datasets. NMF is able to identify fingerprints of gas phase elements, and their time evolution. We evaluated the use of NMF for QMS and OES analysis by studying the trimethylaluminum (TMA)-H_2O and diethylzinc (DEZ)-H_2O ALD chemistries.

In the case of OES, we demonstrate that a naive analysis method of tracking channel intensities corresponding to specific emission lines is not sufficient to identify and separate reaction products. Using NMF, we were able to identify spectra and time signatures from precursor fragments and reaction products, confirming the established reaction mechanism between TMA and H_2O yielding CH_4. The NMF analysis also suggests that the discharge generates spurious H signals that can give rise to wrong interpretation of the H emission lines.

For QMS under steady state conditions, NMF easily and cleanly separates precursors and reaction products. Interpretation of mass spectra is easier, compared to OES, and the clean spectra and time traces NMF yields allow for faster identification of novel reaction products. This way, by using NMF on OES and QMS data, we reveal autocatalytic alkene formation during the DEZ-H_2O process, in addition to the generation of ethane.

With this work, we aim to draw attention to a relatively simple machine learning method that can be straightforwardly adopted by experimentalists. While in situ experiments are capable of yielding (very) large datasets, NMF helps to draw a complete picture of the gas phase during ALD processes.

9:45am **AIML1-ThM-8 Machine Learning Assisted Multiscale Simulation of Metal Deposition***,* ***Michael Nolan,*** Tyndall Institute, Ireland*; Karl Ronnby, Samuel Delgado,* Tyndall National Institute, University College Cork, Ireland

The deposition of metals onto a multitude of substrates and the resulting interfaces are critical elements of many technologies, e.g. semiconductor devices, catalysts, sensors, etc. Atomistic simulations provide a powerful tool to select between different metal-substrate combinations for target technologies. Semiconductor devices using 2D materials require good quality interfaces between metal contacts and the 2D channel, back end of line interconnects require smooth conducting metal films that are immune to metal diffusion and catalysts using support metals benefit from metal islands with high density of active sites. In this contribution we present atomistic multiscale simulations of metal deposition and morphology evolution that use machine learning potentials to accelerate simulations and reach time and length scales that are out of reach with density functional theory (DFT). We use DFT to explore how the modification of TaN can promote a change in morphology of deposited copper from islands to lateral growth, computing activation barriers for critical metal migration processes using neural network ML potentials, which reduces the time for barrier calculation to a few minutes. With these barriers we employ kinetic Monte Carlo simulations to predict key film properties including roughness, covered surface area and density of defects. A similar approach is employed for Ru, which is of high interest for metal interconnects. Cu-Ru alloys are of interest and we present genetic algorithm studies of global minimum energies structures with an alloy ML potential and apply this to atom migrations. Finally, we explore Ru-TMD interfaces and dynamics to understand formation of Ohmic contacts found experimentally.

AI/ML/Autonomous Experimentation for Thin Films Processing

Room 317 - Session AIML2-ThM

AI/ML Tools in Molecular Beam Epitaxy & Flash Session

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

11:00am **AIML2-ThM-13 AI/ML Techniques for Molecular Beam Epitaxy***,* ***Stephanie Law,*** Pennsylvania State University **INVITED**

MBE synthesis offers precise control over film thickness, composition, and crystallinity by tuning a range of growth parameters. While this tunability enables tailored synthesis, it also introduces significant complexity, resulting in multidimensional growth spaces. Identifying optimal growth parameters within these complex spaces typically involves trial-and-error experimentation, in which growers use a combination of intuition and grid search methods to guide the choice of growth parameters. An alternative way to navigate complex growth spaces is through the use of Bayesian Optimization (BO) or AI-driven hypothesis generation, which can explore parameter space more efficiently and find global optimal growth parameters. By leveraging probabilistic models such as Gaussian Process Regression (GPR), BO can statistically select the most informative experiments, balancing exploration of new growth parameter spaces with exploitation of known good parameters. AI-guided hypothesis generation can be used to read the extant literature and generate experimental campaigns. In this work, we apply BO to efficiently explore the growth space for In_2Se_3 thin films on Al_2O_3 substrates using MBE. By integrating prior experimental data and employing GPR, we employed a data-driven synthesis campaign that significantly reduced the number of required

experiments. Leveraging BO resulted in an increase in the fraction of γ-In_2Se_3 from ~50% to >90% in fewer than 10 samples. To further understand the influence of individual growth parameters on model predictions, we used SHapley Additive exPlanations (SHAP), w hich indicated that the indium flux is the most important parameter in predicting the film polymorph fraction, pointing toward defects as a method of stabilizing the polymorph. We also employed AI-guided experimental design to explore growth parameter space for In(Ga)Sb quantum dots. This approach led us to a clear understanding of how QD morphology and density depends on growth parameters, and how QD photoluminescence is correlated with both. These findings highlight the potential of data-driven optimization in complex materials synthesis and further solidify the growing body of evidence supporting the use of ML for efficient thin film experimental design.

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***, 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.

AI/ML/Autonomous Experimentation for Thin Films Processing

Room 317 - Session AIML1-ThA

AI/ML in Materials Characterization

Moderators: Mr. Lucas Kuehnel, University of Missouri, **Dr. Andreas Werbrouck**, Ghent University, Belgium

2:15pm **AIML1-ThA-1 Combining Optimization and Agentic Control in Automated Scanning Probe Microscopy**, ***Boris Slautin***, *Sheryl Sanchez, Jordan Marshall, Mahshid Ahmadi, Sergei Kalinin,* The University of Tennessee **INVITED**

Closed-loop optimization is increasingly used in automated microscopy for tasks ranging from experimental parameter tuning to mapping unknown structure–property relationships. The core idea is to guide the experiment by selecting the next most informative measurement location or parameter set while accounting for uncertainty, noise, and limited experimental budget. However, when the objective, decision space, and constraints must be defined in advance, the automation remains largely local to a specific task. This becomes limiting for hierarchical or evolving experiments that require coordinated measurement selection, quality control, parameter adjustment, mode switching, and refinement of the experimental question.

The emergence of agentic AI provides a complementary route for automated microscopy. Many experiments are not a single optimization loop, but a set of coupled steps involving data acquisition, image analysis, quality control, parameter adjustment, and physical interpretation. Closed-loop methods are powerful for well-defined local optimization or exploration tasks, but the connection between these steps often still requires expert judgment. Agentic systems can help provide this connecting layer by supporting experiment-level orchestration, propagating information between local loops, incorporating physical context, and dynamically planning the next experimental step. In this sense, agentic control should not replace classical optimization, but rather coordinate optimization modules, data-analysis routines, physical constraints, and microscope-control actions within a broader automated experiment.

In this work, we develop an agentic system for automated scanning probe microscopy (SPM), demonstrated on ferroelectrics using piezoresponse force microscopy as the example. The agent provides global orchestration of the experiment by connecting microscope control with image and spectroscopy analysis, incorporating physical context, and making higher-level decisions about the experimental trajectory. Classical optimization are integrated as local modules for tasks that can be formulated quantitatively, such as tuning measurement parameters or selecting informative measurement locations.

We deploy this approach in a real experimental setting, where the agent coordinates microscope actions, analysis routines, and local optimization loops during PFM measurements. Although demonstrated for ferroelectric PFM, the same architecture can be extended to broader automated experimental problems, including thin-film processing, synthesis–characterization loops, and structure–property mapping.

2:45pm **AIML1-ThA-3 Multi-Agent System for Optical Microscopy of 2D Materials Characterization**, ***Haozhe "Harry" Wang***, Duke University

Optical microscopy remains an indispensable tool for the rapid characterization of two-dimensional materials, yet conventional workflows depend heavily on trained operators to navigate samples, identify regions of interest, and interpret complex image features — creating a persistent bottleneck in experimental throughput. Here, we present a multi-agent system built on foundation models that fully automates the optical microscopy pipeline for 2D material characterization, from stage navigation to image acquisition to quantitative analysis.

Our architecture decomposes the microscopy workflow into specialized, cooperating agents: a navigation agent that performs intelligent search over the substrate, a perception agent that segments and classifies material features in real time, and an analysis agent that extracts quantitative descriptors such as layer number, domain morphology, and spatial coverage. These agents communicate through a structured orchestration layer, enabling adaptive decision-making — for example, dynamically adjusting scan density in response to local material heterogeneity or prioritizing high-value regions for higher-magnification imaging. Built on the ATOMIC (Autonomous Technology for Optical Microscopy and Intelligent Classification) framework, the system leverages large vision models to achieve zero-shot generalization across material systems and substrate types without task-specific fine-tuning.

We demonstrate the system on chemical vapor deposition-grown 2D semiconductors, where it achieves classification accuracy exceeding 99% while autonomously generating statistically representative maps of growth outcomes across centimeter-scale substrates. By eliminating manual intervention and enabling continuous, unsupervised operation, the multi-agent architecture transforms optical microscopy from a serial, operator-dependent technique into a scalable, intelligent characterization platform — laying the groundwork for closed-loop integration with upstream synthesis and downstream property measurements in autonomous materials laboratories.

3:00pm **AIML1-ThA-4 A Reference Architecture for Complex Scientific Chatbots Based on ChASE, an AI Assistant for SIMS**, ***Jordan Barrette***, *Jerry Hunter, Prateek Maheshwari,* Beamline Incorporated

Secondary Ion Mass Spectrometry analysis planning requires synthesizing multiple knowledge sources: textbook physics governing sputter rate, ion yields, depth resolution, and matrix effects; institutional memory encoded in years of past setups; and analyst judgment about instrument state and sample characteristics. We present ChASE (Chat-bot Augmented for SIMS Expertise), a chatbot-style AI assistant for SIMS laboratories that integrates theoretical, historical, literature, and tribal knowledge into a unified decision-support tool. Rather than treating the underlying language model as an oracle trained to know everything a priori, we instruct it as a research expert—one that synthesizes information from many sources, much as a lab scientist draws on tribal knowledge, reference literature, recent work, and customer expectations.

The presentation outlines the architecture of ChASE, composed of three primary components: technical “skills,” tribal knowledge, and a reference literature corpus. Using this architecture, we demonstrate automated setup guidance that consults and reconciles historical setups, theoretical predictions, laboratory best practices, and published experimental methods. Building on the suggested setup guidance, we demonstrate automated scheduling across the current job queue. The setup-related predictions produced—sputter rate, required depth, estimated tuning overhead—roll up into time estimates that are compared against similar historical runs to produce expected and worst-case total analysis times, giving lab managers a preliminary schedule with built-in margin.

Central to this research-expert approach is the skill framework, which mirrors the way an experienced analyst decomposes a complex problem into prerequisite questions before synthesizing an answer. Skills are applied recursively: complex tasks decompose into collections of prerequisite subtasks that themselves constitute skills. For instance, a “SIMS Setup Predictor” skill leverages “Primary Ion Predictor,” “Mass Resolution Predictor,” and “Depth Resolution Predictor” as constituent skills. Each is a sophisticated model on its own, built on the same reference architecture—hence the recursive nature—supporting prediction models for ion yield, mass interference, and cascade mixing, respectively. We compare the architecture’s response accuracy for complex scientific tasks with a standalone language model and more common adaptation approaches such as pre-training, fine-tuning, prompt engineering, and retrieval-augmented generation. Finally, we discuss strengths and limitations of the approach, within the context of the ultimate goal—an AI framework that can be expanded beyond SIMS.

3:15pm **AIML1-ThA-5 Leveraging Automated Characterization and Machine Learning to Optimize Defects for Quantum Sensing**, ***Dane deQuilettes***, Princeton University

Spins in solid-state materials, molecules, and other chemical systems have the potential to impact the fields of quantum sensing, communication, simulation, and computing. In particular, color centers in diamond, such as negatively charged nitrogen vacancy (NV−) and silicon vacancy centers (SiV−), are emerging as quantum platforms poised for transition to commercial devices. A key enabler stems from the semiconductor-like platform that can be tailored at the time of growth. The large growth parameter space makes it challenging to use intuition to optimize growth conditions for quantum performance. In this work, we use supervised machine learning to train regression models using different synthesis parameters in over 100 quantum diamond samples. We train models to optimize NV−defects in diamond for high sensitivity magnetometry. Importantly, we utilize a magnetic-field sensitivity figure of merit (FOM) for NV magnetometry and use Bayesian optimization to identify critical growth parameters that lead to a 300% improvement over an average sample and a 55% improvement over the previous champion sample. Furthermore, using Shapley importance rankings, we gain new physical insights into the most impactful growth and post-processing parameters, namely electron

irradiation dose, diamond seed depth relative to the plasma, seed miscut angle, and reactor nitrogen concentration. As various quantum devices can have significantly different material requirements, advanced growth techniques such as plasma-enhanced chemical vapor deposition (PE-CVD) can provide the ability to tailor material development specifically for quantum applications.

AI/ML/Autonomous Experimentation for Thin Films Processing

Room 317 - Session AIML2-ThA

AI/ML-enhanced RHEED in Thin Film Growth

Moderators: Mr. Dorien Carpenter, University of Alabama at Birmingham, **Prof. Martin Seifrid**, North Carolina State University

3:30pm **AIML2-ThA-6 Turning RHEED into a Tool for Stoichiometry Analysis and Real-time Control of Epitaxial Film Growth Using Machine Learning**, ***Ryan Comes,*** University of Delaware **INVITED**

Progress in materials synthesis increasingly depends on combining real-time characterization with data-driven analysis and feedback. Reflection High–Energy Electron Diffraction (RHEED) offers a unique capability for monitoring thin film growth during deposition, yet its broader utility is constrained by limited dynamic range, signal saturation, and challenges in extracting quantitative information from diffraction images. To address this, we have developed a robust and extensible methodology to improve RHEED image quality, enable quantitative analysis, and support feedback-aware synthesis workflows. Weshowed that deep convolutional neural networks trained on RHEED patterns could accurately predict the cation stoichiometry of $SrTiO_3$ thin films during growth, with explainable AI tools revealing previously unrecognized correlations between diffraction streak intensity ratios and reciprocal-space spacing.[1] These results establish RHEED as a viable surrogate measurement for film stoichiometry.

To identify and track structural evolution during growth, we have also applied unsupervised machine learning approaches, including principal component analysis and k–means clustering, to time-resolved RHEED image sequences. Image drift correction and precise alignment of the specular reflection enable reliable frame-to-frame comparisons for $LaFeO_3$ and $SrIrO_3$ films synthesized by molecular beam epitaxy. These preprocessing and analysis steps establish a foundation for automated recognition of growth modes and cross-material comparisons.[2]

Our ongoing work focuses on the development of software for real time analysis of videos to provide ML-driven feedback to the human operator of the MBE system using reinforcement learning. This framework incorporates a pathway toward real-time synthesis feedback by integrating enhanced RHEED video streams with growth metadata such as fluxes, shutter states, chamber pressure, and substrate information. By combining unsupervised learning, deep neural networks, and real-time diagnostics, this work lays the groundwork for closed-loop thin-film synthesis capable of quantitative interpretation, adaptive control, and accelerated materials discovery.

[1] S.B. Harris, et al. *Nano Letters*, **25**, 5867–5874. (2025)

[2] P.T. Gemperline, et al. *JVSTA,* **43**, 032701 (2025)

4:00pm **AIML2-ThA-8 AI-Driven Recognition of RHEED Pattern Features for MBE Monitoring**, ***Vadim Gorshanov***, DCA Instruments Oy, University of Turku, Finland*; Markku Rajala,* DCA Instruments Oy, Finland*; Milica Todorović,* University of Turku, Finland

Reflection High-Energy Electron Diffraction (RHEED) is a standard in situ characterization technique used to monitor surface crystal structure during thin film deposition methods such as Molecular Beam Epitaxy (MBE). RHEED patterns contain rich information about the surface ordering, but their interpretation is often manual or restricted to simple intensity tracking, which fails to fully capture the evolution of the surface. This work presents an approach to automating RHEED pattern analysis using semantic segmentation (pixel-wise labeling) to detect and classify diffraction features, including spots, streaks, and Kikuchi lines.

To enable such analysis, we compiled and annotated a ground-truth dataset of diverse RHEED patterns. This dataset was utilized to conduct a comparative study of modern deep learning architectures, benchmarking Convolutional Neural Networks (CNNs) and Transformer-based architectures paired with various encoder backbones. The goal of this comparison is to identify and optimize the architecture that maximizes accuracy to ensure reliable pattern monitoring. Results demonstrate that the optimized models robustly isolate diffraction features from complex backgrounds. Furthermore, the method exhibits strong generalization capabilities, successfully recognizing RHEED patterns of different material systems and imaging conditions not present in the training data. In addition, the image processing achieves frame rates suitable for real-time monitoring. The procedure enables the extraction of quantitative diffraction features from raw RHEED images and videos, paving the way for automated control in industrial RHEED workflows, as well as facilitating post-growth analysis.

4:15pm **AIML2-ThA-9 Toward LLM Driven Agentic Automation of RHEED Diagnostics in Thin Film Deposition**, ***Asraful Haque***, *Rama Vasudevan, Christopher Rouleau, Sumner Harris,* ORNL

Real-time interpretation of reflection high-energy electron diffraction (RHEED) patterns during thin film growth remains a bottleneck for fully autonomous deposition systems. While individual machine learning tools have demonstrated promise for isolated RHEED analysis tasks, integrating them into a coherent, decision-capable framework has been challenging. Here we present an agentic automation architecture that orchestrates multiple AI/ML modules for RHEED diagnostics using large language model (LLM)-based agents as the reasoning backbone. The system integrates: (i) a YOLO (You Only Look Once) based object detection and classification pipeline for real time identification of RHEED pattern features including spot geometry, streak morphology; (ii) a structural similarity (SSIM) based algorithm for automated sample alignment and beam registration in RHEED; and (iii) an FPGA-DAC (Field Programmable Gate Array and Digital to Analog Converter) controlled electron beam deflection subsystem enabling automated rocking curve acquisition. The LLM agent layer interprets outputs from each module, coordinates decision logic across the diagnostic pipeline, and interfaces with the deposition control system to enable closed-loop feedback.

We demonstrate proof of concept autonomous operation where the agent sequences alignment, real-time pattern classification, and real-time rocking curve measurements without human intervention. This modular, agent-orchestrated approach represents a scalable path toward self-driving thin film growth platforms where natural language reasoning and computer vision together manage the complexity of in-situ diagnostics. This research was conducted at the Center for Nanophase Materials Sciences (CNMS), which is a DOE Office of Science User Facility at Oak Ridge National Laboratory.

AI/ML/Autonomous Experimentation for Thin Films Processing
Room Ballroom A - Session AIML-ThP
Poster Session

AIML-ThP-1 Ai-Based Real-Time Ashing Rate Estimation and Anomaly Detection Using Equipment Sensor and OES Data in Semiconductor Manufacturing, *Taekyung Ha,* ***You Tack Suh****,* PSK, Republic of Korea

The semiconductor industry is currently undergoing rapid transformation driven by the expansion of artificial intelligence (AI) technologies. AI is not only increasing the demand for advanced semiconductor devices, but is also becoming a key enabler of innovation in semiconductor manufacturing processes. As semiconductor devices become more sophisticated with higher integration density and smaller process nodes, wafer fabrication costs continue to rise significantly. Consequently, unexpected process failures or abnormal equipment conditions can result in substantial financial losses and reduced production yield.

Despite the increasing complexity of semiconductor manufacturing, many plasma-based processes such as etching and ashing still lack direct in-situ methods to determine whether the process is proceeding normally in real time. In current manufacturing environments, abnormal process conditions are typically monitored through post-process sampling inspections and offline metrology techniques. However, these approaches have inherent limitations, including delayed feedback, increased inspection costs, and the risk of missing transient process anomalies.

To address these challenges, this study proposes an AI-based process monitoring framework for semiconductor ashing equipment. The proposed approach utilizes both equipment sensor signals and Optical Emission Spectroscopy (OES) measurement data collected during the ashing process to estimate the ashing rate and detect abnormal process conditions in real time. By integrating multiple process-related data sources, the model aims to improve process visibility and enable early detection of equipment or process deviations.

For the development of the predictive models, machine learning methodologies including XGBoost and Long Short-Term Memory (LSTM)-based neural network algorithms were investigated and compared. Experimental results demonstrated that the XGBoost-based models achieved the best overall performance among the evaluated approaches. The proposed model achieved an ashing rate estimation accuracy of 89% and a process anomaly detection rate of 95%, indicating strong potential for practical deployment in semiconductor manufacturing environments.

The results of this study suggest that AI-driven virtual metrology and anomaly detection technologies can significantly enhance process stability, reduce inspection dependency, and minimize yield loss in advanced semiconductor fabrication processes.

AIML-ThP-2 ASSD Student Award Finalist Poster: Leveraging Convolutional Neural Networks (CNN) for the Real-Time Classification of Melt-Fraction of Phase Change Materials (PCMs) Using Infra-Red (I.R.) Imaging, ***Nishit Pachpande****, Anusree Sen, Debjyoti Banerjee,* Texas A&M University

Phase Change Materials (PCMs) are attractive candidates for incorporation in the Latent Heat Thermal Energy Storage (TES/ LHTES) Systems due to their high latent-heat capacity. Despite their attractive performance, PCMs (especially salt hydrates) often pose reliability issues. Salt hydrates (as PCMs) suffer from debilitating complications due to supercooling issues, i.e. the tendency of PCMs to stay in liquid phase at a temperature lower than the melting temperature (during the solidification portions of thermal cycles involving repeated melting and freezing). Also, the traditional methods for analyzing the thermal properties of PCMs are often time-consuming, expensive and require specialized protocols (where the measured properties are dependent on the measurement techniques).

Machine Learning (ML) techniques can be leveraged to address these issues by enhancing the reliability of PCMs with minimal effect on their performance.The objective of this study is to predict the amount of energy stored in a PCM-based TES using ML techniques (CNN model). An experimental apparatus which was developed in this investigation for performing experimental validations of the computational predictions. These model predictions were also compared with that of a pre-trained model ("EfficientNetB0"/ TensorFlow). The PCM (PureTemp 29^{TM}) was melted in a vertically graduated cylinder using a nichrome coil. Digital image acquisition (GoPro HERO8) of the melting PCM was performed for obtaining the corresponding melt-fraction values based on the height of meniscus in the melt pool. For training and testing the CNN model ~1660 I.R. images were acquired at 1-minute intervals (using FLIR One camera) during melting. The classification model was trained for classification classes corresponding to 10 melt-fraction bands (0-10%, 11-20%, ..., 91-100%). The CNN model predictions yielded a test accuracy of 93.98%, thus performing better than the pre-trained model (93.73%). These results demonstrate the efficacy of implementing deep learning based image analyses models as a fast, low-cost, reliable, accurate, robust, resilient and scalable tool for the real-time monitoring, prediction and characterization of PCM based TES systems.

KEYWORDS: latent heat thermal energy storage, TES, LHTES, machine learning, ML, artificial intelligence, AI, thermal management

AIML-ThP-3 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.

AIML-ThP-4 Role of Oxide Phase and Surface Facets on Self-Limiting Thermal Atomic Layer Etch in High-k Oxides, ***Michael Nolan****, Rita Mullins,* Tyndall National Institute, University College Cork, Ireland

Thermal Atomic Layer Etching (ALE) is of high interest for its potential to deliver atomic level control over the etch of many materials and shows potential for use in future CMOS nodes with requirements for sub-nm levels of control on complex structures. It is performed using sequential surface modification and volatile release reactions. For metal oxides, HF fluorinates the initial surface to form a MF4 layer (M = metal) which undergoes ligand-exchange with precursors such as TiCl4 or SiCl4, which volatilizes the MF4 layer. The question of the role of the phase and surface facets in a deposited high-k metal oxide film has received little attention but can be addressed with first principles atomistic simulations. In this contribution we use density functional theory and molecular dynamics simulations to explore the effect of the phase and surfaces of HfO2 and

ZrO2 on the HF modification half-cycle of ALE. The models used in this study representing polycrystalline materials are the (111) and (001) surface facets of monoclinic, orthorhombic and tetragonal HfO2 and ZrO2. Our thermodynamic analysis shows that for polycrystalline HfO2 and ZrO2, the HF pulse reacts in a self-limiting manner, and is preferred up to processing temperatures that are sensitive to the phase and surface. Models of HF coverage are used to compute calculated theoretical etch rates for the different oxide phases and surface facets and these show a strong dependence on both the crystal phase and the surface so that if different phases and facets are present an uneven etch profile will be seen. The stability, geometry and surface atomic coordination environments drive this dependence.

AIML-ThP-5 Automated Experimentation Reveals Co-Spacer Design Rules for Targeted Phase Selection in Quasi-2D Halide Perovskites*, **Elham Foadian**, Mahshid Ahmadi,* University of Tennessee Knoxville

Targeted phase control in quasi-2D halide perovskites (HPs) is essential for tailoring their optoelectronic properties, yet remains challenging since phase formation is governed by the coupled effects of spacer chemistry, precursor assembly, and crystallization kinetics. Despite rapid advances in quasi-2D HPs, the identification of spacer design rules still relies largely on manual trial-and-error synthesis, limiting mechanistic understanding and reproducible phase selection. Here, we present a multimodal, closed-loop autonomous experimentation workflow for targeted phase selection in quasi-2D HPs using co-space engineering. We investigate a ternary 3D:2D compositional space composed of 3D $FAPbI_3$ and two Dion–Jacobson co-spacers, 1,4-butanediammonium (BDA) and 3-(aminomethyl)piperidinium (3AMP), to identify compositions that converge toward a stable 3D-like phase. In situ photoluminescence (PL) spectroscopy reveals that BDA promotes rapid crystallization and early formation of 3D-like emissive domains, whereas 3AMP slows crystallization and favors the persistence of lower-n quasi-2D phases. By systematically regulating the BDA:3AMP co-spacer ratio, the crystallization pathway can be tuned to balance fast framework formation with sufficient structural relaxation, enabling convergence toward the target phase. These kinetic insights are integrated with automated high-throughput synthesis, time-dependent PL characterization, X-ray diffraction similarity scoring, and Gaussian process–Bayesian optimization to navigate phase homogeneity and stability. This work establishes co-spacer engineering as a design strategy for targeted phase selection in kinetically governed quasi-2D HPs.

AIML-ThP-6 Toward Autonomous Discovery of UV-Activated Small-Molecule Inhibitors for Area-Selective Atomic Layer Deposition*, **Lucas R. Kuehnel**, Campbell A. Sweet, Anthony A. Khoury, Erick A. Gutierrez-Monje, Matthias J. Young,* University of Missouri-Columbia

Recent work from our group has demonstrated functional group lithography to nucleate atomic layer deposition (ALD) for patterned deposition on non-growth surfaces like MoS_2 and highly oriented pyrolytic graphite. An analogous approach could enable patterned deposition on growth surfaces through UV-activated inhibition to further develop the dry functional group lithography tool set. However, it is unclear what chemistries might enable UV-activated inhibition. To efficiently search chemical space for UV-activated inhibitors, we developed a high-throughput screening framework for machine-learning-guided discovery. As a model system, we focus here on identifying inhibitors for ZnO ALD from diethylzinc and water on alumina surfaces. Quartz crystal microbalance measurements are used to monitor adsorption, desorption, and film growth dynamics, enabling quantitative assessment of nucleation delay and related mass change metrics. These experimentally derived metrics are used as targets in an initial exploratory active learning phase to generate a diverse seed dataset. We employ a suite of machine learning models for predicting vapor pressure, melting point, boiling point, pK_a, and oxidation potential to screen candidate molecules for physical processing feasibility within the constraints of the experimental system. These experimental constraints, imposed by a high-throughput design, place inherent biases on the searchable chemical space. For example, choosing to eliminate precursor heating enables more rapid testing, but favors smaller, more volatile molecules. Practically speaking, careful consideration of the imposed constraints can favor the discovery of molecules that are more likely to be industrially viable. The success of this discovery framework would provide a first step toward establishing dry functional group lithography as a tool for nanofabrication.

AIML-ThP-7 Toward Autonomous Screening of oMLD Thin Films for Next-Generation Lithium-ion Battery Materials*, **Campbell A. Sweet**, Kim A. Gerhard, Blaine D. Kelly, Lucas R. Kuehnel, Matthias J. Young,* University of Missouri

Organic conductive polymers with high lithium-ion redox capacities are promising as sustainable positive electrode materials and as performance-enhancing coatings for established cathodes such as lithium iron phosphate. Oxidative Molecular Layer Deposition (oMLD) enables the delivery of conformal organic conductive polymer thin films on porous powder substrates with precise thickness control, making it a promising platform for this application domain. However, exploration of oMLD chemistries remains limited because the experimental workflows are time-intensive and require substantial manual intervention. Given the large chemical design space of organic precursors, traditional oMLD workflows are not readily scalable for rapid materials discovery. Here, we discuss progress to date on establishing a rapid and autonomous workflow for screening oMLD thin films as organic positive electrode coating materials for lithium-ion batteries. We specifically describe our efforts to accelerate deposition, automate precursor handling, and streamline electrochemical characterization and analysis. Workflow analysis revealed key bottlenecks including low precursor volatility, slow performance characterization, and inefficient chemical data mapping and search. We report progress toward autonomous workflow operation using LLM agents to manage repetitive and cumbersome tasks. These advances provide a framework for accelerating oMLD-driven materials discovery and may extend to other challenging-to-accelerate workflows.

AIML-ThP-8 Small Science Agentic Models for Automated Perovskite Thin Film Exploration via Monte Carlo Decision Trees and Delayed Bayesian Optimization Feedback*, Elham Foadian, **Sheryl Sanchez**, Mahshid Ahmadi,* University of Tennessee Knoxville

Autonomous materials discovery requires decision frameworks capable of handling process-dependent outcomes, delayed feedback, and sparse experimental data. We introduce Small Science Agentic Models (SSAMs), a distributed architecture for self-driving laboratories in which scientific reasoning is embedded across narrow, physically interpretable models tied directly to the experimental workflow. Implemented as an event-driven, agent-based closed loop, the SSAM framework integrates hypothesis generation, protocol construction, characterization-derived state representation, machine-learning decision-making, and orchestration, treating each experiment as an evolving process trajectory rather than an isolated parameter evaluation. We apply this framework to discover an effective Dion–Jacobson (DJ) organic spacer for stabilizing the photoactive 3D-like phase of $FAPbI_3$. Through agent-guided hypothesis formation employing Socratic reasoning and Tree-of-Thought branching, furan-2,5-diyldimethanammonium (FuDMA) was identified as a previously unexplored DJ spacer compatible with the perovskite lattice. The binary $FAPbI_3$/$FuDMAPbI_4$ compositional system was then explored autonomously using a Monte Carlo decision tree with delayed-reward updating. Over ten iterations, the system achieved 95% predictive agreement between predicted and measured film quality, identified the highest-performing sample within three iterations, and converged toward focused exploitation by approximately seven iterations. Optimal film formation was localized to FA based compositions of 60–80% within a tightly constrained spin-coating and annealing regime. Complementary Nuclear Magnetic Resonance Spectroscopy (NMR), Fourier Transform Infrared Spectroscopy (FTIR), and time-resolved photoluminescence (TRPL) analysis revealed two composition-dependent FuDMA interaction regimes. At intermediate spacer concentrations, FuDMA acts as a lattice-stabilizing additive that extends carrier lifetimes and suppresses non-radiative recombination, while at higher fractions it drives quasi-2D phase formation accompanied by accelerated fast-decay dynamics. These results establish SSAMs as a transparent, workflow-grounded approach to trajectory-aware autonomous experimentation in dynamic materials systems.

AIML-ThP-9 Revolutionizing Sem Image Analysis: Deep Learning vs. Classical Restoration for Enhancement and Segmentation*, **Amanda Georgina Nieto Sánchez**,* Universidad Anáhuac México*; Leon Hamui,* School of Engineering, Universidad Anáhuac México

Scanning electron microscopy (SEM) images frequently suffer from noise, blur, low contrast, and acquisition artifacts that complicate automated microstructure analysis and segmentation. While classical restoration approaches such as Wiener filtering, Richardson–Lucy deconvolution, CLAHE enhancement, and unsharp masking are commonly used, their impact on downstream AI-based segmentation tasks remains insufficiently explored.In this work, we investigate a deep learning-based image

restoration framework for SEM post-processing using convolutional autoencoders trained with synthetically degraded SEM datasets. The degradation pipeline incorporates Gaussian blur, additive and multiplicative noise, and contrast perturbations designed to emulate realistic SEM acquisition conditions. The restored images are subsequently evaluated both through image-quality metrics and automated segmentation performance.A comparative study was performed between the proposed convolutional autoencoder and four classical restoration techniques: Wiener filtering, Richardson–Lucy deconvolution, CLAHE, and unsharp masking. Restoration quality was quantified using PSNR and SSIM metrics. In addition, segmentation performance was evaluated using U-Net architectures trained separately on degraded and restored SEM images.Although classical metrics showed moderate variations between restoration methods, segmentation-oriented evaluation revealed significant differences in model behavior. Quantitative analysis demonstrated that restoration-assisted segmentation reduced false positive regions, decreased over-segmentation, and lowered fragmentation of detected microstructures. Specifically, restored-image segmentation reduced the false-positive component ratio from 0.36 to 0.29 and decreased fragmentation metrics relative to models trained on degraded data.The results suggest that deep learning-based SEM restoration may provide advantages not fully captured by traditional image-quality metrics alone, particularly when evaluated within automated materials characterization workflows. This approach demonstrates the potential of AI-assisted SEM post-processing for improving robustness and reliability in microstructure segmentation tasks.

AIML-ThP-10 Enabling AI-Driven Materials Discovery Through Data Infrastructure*, **Akshay Talekar**,* 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.

AIML-ThP-11 Statistical Analysis of Pressure-Growth Correlations in Temporal ALD and MLD toward the Identification of Process Deviations*, Andrew Ball, Joelle Scott, Jay Werner, **David Bergsman**,* University of Washington

In the development of many new vacuum deposition-based processes, such as in atomic layer deposition (ALD) and molecular layer deposition (MLD), process conditions are optimized through correlation with deposition outputs, such as film thickness and growth per cycle (GPC). Other data, such as reactant dose pressure and exposure, can also be used in process optimization, either quantitatively (e.g., by correlating growth per cycle with reactant exposure) or qualitatively (e.g., by identifying erroneous dose conditions or reactant depletion). However, while there are many robust tools informing process optimization, there remains a need for tools that can identify deviations within a deposition run.

In this work, we introduce a diagnostic approach for processing and analyzing reactor parameters in temporal ALD/MLD processes. By linking reactor pressure data (via LabVIEW JSON output) with a Python-based processing pipeline, we are able to correlate steady-state and maximum chamber pressure of a given step in the deposition process with the overall process growth-per-cycle (GPC). Through this correlation, we identify processing steps and pressures that most influence the GPC of the process, which can help identify areas for process improvement. False positives were filtered from the list of correlations using a perturbation stress test that involved repeatedly injecting 15% Gaussian noise across 1,000 iterations of a given step. To test this pipeline, we analyzed the pressure and GPC data from several processes in development in our lab as case studies. Results of these analyses suggest that this approach is able to identify subtle pressure drifts as well as processing steps that may be contributing to effects like CVD. This work provides a framework for cleaning data for potential use in interpretable machine learning models like Random Forest and Gradient Boosting, enabling autonomous process optimization with suitable, high-training data, while maintaining the transparency needed for physical reactor troubleshooting.

AI/ML/Autonomous Experimentation for Thin Films Processing
Room 317 - Session AIML1-FrM

Materials Design and Autonomous Discovery

Moderators: Eden Goodwin, Carleton University, Canada, **Prof. Stephanie Law**, Pennsylvania State University

8:15am **AIML1-FrM-1 Accelerating Discovery and Design for Thin Film Polymer Self-Assembly through Machine Learning**, ***Gregory Doerk,*** Brookhaven National Laboratory **INVITED**

Thin film polymer self-assembly is a powerful approach to create nanoscale patterns across large areas with low cost for diverse applications in electronics and materials science. The confluence of compositional and architectural variation, entropic and surface energy constraints arising from thin film confinement, and process-dependent assembly pathways yields of panoply of nanoscale morphologies of which only a fraction can be discovered, catalogued, and engineered through manual experimentation. This talk will elaborate examples where artificial intelligence and machine learning (AI/ML) can accelerate thin film polymer self-assembly research. We describe the application of Gaussian process regression to automate x-ray scattering characterization for the discovery of new morphologies in template-directed self-assembly of block copolymer blends. We then illustrate the application of an active learning approach that incorporates foundation model-supported "few shot" image classification for determining processing-morphology relationships in spray-deposited block copolymer/homopolymer blend films using minimal quantities of data. Complementary in situ x-ray scattering measurements during spraying provide key mechanistic insights into trends in self-assembled structural organization across the process design space.

8:45am **AIML1-FrM-3 Autonomous Discovery of PFAS-Free Low Surface Energy Thin Films**, ***Stanley Lo,*** *Abraham Herzog-Arbeitman, Alan Aspuru-Guzik, Helen Tran, Nipun Gupta, Harrison Mills,* University of Toronto, Canada

Per- and polyfluoroalkyl substances (PFAS) have long dominated low-surface-energy coatings due to their exceptional water and oil repellency. However, regulatory pressure and environmental persistence concerns have created urgent demand for fluorine-free alternatives. Rational design of such materials requires systematic exploration of polymer composition–wettability relationships, a challenge ideally suited to autonomous experimentation.

We present a self-driving laboratory for accelerated discovery of PFAS-free polymer thin films with low surface energy. The platform integrates a continuous-flow reactor, an automated fraction collector for composition-resolved sample handling, a spin coater for reproducible thin-film deposition, and a multi-liquid dynamic contact angle goniometer for comprehensive wettability characterization. Closed-loop control enables autonomous navigation of copolymer composition space.

The chemical system employs ring-opening metathesis polymerization (ROMP) of norbornene-based monomers to produce statistical polynorbornene copolymers. Monomer design is grounded in the Hansen solubility parameter (HSP) framework, which decomposes intermolecular interactions into three orthogonal dimensions: dispersion (δD), polarity (δP), and hydrogen bonding (δH). Three monomers were synthesized with side groups systematically representing each dimension: an apolar aliphatic chain (low δP, low δH), a polar non-protic group (high δP), and a polar hydrogen-bonding group (high δH). ROMP copolymerization of these monomers yields materials with continuously tunable surface chemistry across the HSP space.

Dynamic contact angle measurements using multiple probe liquids (water, ethylene glycol, and hexadecane) enable surface free energy decomposition and discrimination between hydrophobic and oleophobic character. The platform iteratively selects compositions, deposits films, measures wettability, and updates a surrogate model to identify low-surface-energy optima without fluorinated components.

This work establishes a generalizable, data-driven framework for discovering sustainable high-performance surface coatings and provides mechanistic insight into how monomer polarity and hydrogen-bonding capacity govern macroscopic wettability.

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

AI/ML in Semiconductor Processing & Manufacturing

Moderators: Prof. Dane de Quilettes, Princeton University, **Ms. Zahra Nasiri**, University of Alabama at Birmingham

9:00am **AIML2-FrM-4 Process Informatics Framework for Optimizing Thin-Film Deposition in Advanced Semiconductor Manufacturing**, ***Shota Oda,*** *Seitaro Sanai, Takeshi Hashishin, Ichiro Akai, Takeshi Momose,* Kumamoto University, Japan

We developed a framework based on physics-informed process informatics for supercritical fluid deposition (SCFD) that enables efficient optimization of high deposition rates and conformality in high-aspect-ratio (HAR) structures. As semiconductor devices transition toward three-dimensional integration, high-throughput conformal deposition in HAR features such as through-silicon vias is increasingly required. SCFD, which involves chemical reactions of metal organics in supercritical CO_2($scCO_2$) to allow thin film deposition, is promising because the high diffusivity and solubility of precursors in $scCO_2$ enable concentrated precursor transport deep into microstructures. However, fluid properties and reaction kinetics are highly sensitive to process conditions, and their relationships with deposition behavior have not been quantitatively established. Consequently, conventional process development has relied largely on empirical trial-and-error.

To address this issue, we propose a causal decomposition framework that separates process-condition from physicochemical-response effects, introducing physicochemical parameters as intermediate variables between process conditions and deposition outcomes. Specifically, the SCFD process is mapped onto a causal chain: process conditions → physicochemical parameters → deposition outcomes. The physicochemical parameters are linked to deposition outcomes through mass balance analysis of precursors in HAR features, clarifying their relationship with deposition rate and conformality. To model the relationship between process conditions and physicochemical parameters, we applied Physics-Informed Bayesian Optimization (PIBO). In PIBO, a prior physical model is incorporated as the mean function of Gaussian process regression. The predictive model is sequentially updated using newly acquired data to recommend subsequent process conditions in a closed-loop manner. In this contribution, PIBO was evaluated through sequential sampling from synthetic process-response surfaces designed to emulate SCFD behavior and compared with conventional Bayesian Optimization.

Rather than relying solely on data-driven learning, the proposed framework integrates physically describable behavior with data-driven residual learning. This hybrid approach enables optimal deposition conditions to be derived from sparse datasets while maintaining physical consistency and extrapolation reliability, and can be extended to different target structures without re-optimizing the entire process from scratch. This framework provides a scalable and physically interpretable strategy for optimizing next-generation semiconductor thin-film deposition processes.

9:15am **AIML2-FrM-5 Physics-Informed AI for Semiconductor Equipment Status Estimation and Process Quality Prediction: Equipment Manufacturer's Perspective**, ***Hyunho Yun,*** *Yongjo Kim, Dain Ham, Yun Hyeong Lee, Seon Hye Kim, Young Hoon Kim, Da Gyung Lee, Yoonsun Lee, Hanbee Lee, Hyemin Song, Youjin Jeong, Hongyeon Kim,* Wonik IPS, Republic of Korea

The rapid expansion of AI services has significantly increased demand for advanced semiconductor chips, accelerating the need for equipment strategies that maximize yield, process stability, and tool availability. In fabs, even a single instance of unscheduled downtime can cause substantial losses in production throughput. However, conventional rule-based process management systems such as Fault Detection and Classification (FDC) and Statistical Process Control (SPC) have inherent limitations. They require extensive manual effort to define process-specific rules across diverse process conditions and hardware configurations. In addition, high-cost of metrology equipment typically limits wafer sampling to approximately 1% of mass production wafers, making real-time quality assurance difficult.

To address these challenges, Wonik IPS has developed an intelligent monitoring framework that combines AI-based equipment health diagnosis with wafer quality prediction. The core of the framework is a Condition-Based Maintenance (CBM) model. In deposition equipment, complex

physical interactions among sensors often make direct analysis of raw sensor data unreliable. Wonik's CBM model overcomes this limitation by integrating process log data with a sensor relationship network that captures process role and physical location of each sensor, thereby enabling equipment health diagnosis from the system level down to individual components. Unlike conventional Time-Based Maintenance (TBM), CBM provides a more effective approach for maintaining optimal equipment conditions by reducing downtime and enabling early fault diagnosis.

Wafer quality prediction is further enhanced by integrating CBM-based equipment health information with real-time Virtual Metrology (VM). Wonik's VM solution adopts a hybrid architecture that combines a neural network optimized for subtle pattern recognition and a physics-based model capable of robust inference based on physical principles. Using process log data collected from Wonik's deposition equipment, GEMINI HQ, thickness prediction results demonstrated strong performance even under data-constrained conditions, achieving average error rates of approximately 1% for the neural network model and 5% for the physics-based model.

Going forward, Wonik plans to integrate CBM and VM into a unified predictive framework to further advance autonomous process control. By combining equipment data, health diagnosis, and recipe conditions, this data-driven optimization strategy is expected to enhance real-time monitoring across all process stages and contribute to the development of an intelligent semiconductor equipment ecosystem.

9:30am **AIML2-FrM-6 Multi-Objective Optimization of the Operational Durability of Meniscus-coated Perovskite Solar Cells***, **Eric Oberholtz**, Jongbeom Kim, Maimur Hossain, Ayush Tiwary, Lucas Mar, Yube Ostos, Divenaa Madan, David Fenning,* University of California San Diego

Perovskite solar cells are strong contenders for commercialization of renewable photovoltaics, but making the leap from lab-scale fabrication to commercial devices requires switching to industrial-scale thin film fabrication techniques such as meniscus coating while maintaining high quality thin films. Transitioning from spin coating to meniscus coating introduces a highly complex, multi-variate parameter space we must navigate while balancing thin film quality alongside operational stability of the finished devices. To bridge the lab-to-fab gap, we integrate MEITNER (Meniscus Integration for Thin film Nano Electronics Robot)- an autonomous platform for meniscus depositions- within our robotic experimentation ecosystem paired with a multi-objective Bayesian optimization framework. This framework explores the multi-dimensional parameter space comprised of fabrication and chemistry design choices aimed to maximize (1) the initial quality of the perovskite thin film, (2) the initial performance of the pristine perovskite devices, and (3) the operational stability under accelerated degradation of the thin film devices. Our workflow accelerates the discovery of promising candidates for industrially relevant fabrication of operationally durable semiconductor thin film devices.

AI/ML/Autonomous Experimentation for Thin Films Processing

Room 317 - Session AIML3-FrM

Connecting Datasets, Models, and Characterization

Moderators: Ms. Molly McDonough, Pennsylvania State University, **Dr. Amanda Sanchez**, Universidad Anahuac

9:45am **AIML3-FrM-7 In the Design of Thin Charge-Transport Films for Perovskite Solar Cells, Predictions from Machine Learning and Physics Only Partially Overlap***, Joseph Oti, Elvis Twumasi,* SUNY Polytechnic Institute*; Nana Derkyi,* SUNY Polytechnic Institute, Ghana*; Tabiri Asumadu, Desmond Klenam, Bernice Abraham,* SUNY Polytechnic Institute*; Rodica Neamtu, **Nancy Burnham**,* Worcester Polytechnic Institute*; Iulian Gheraisou, Winston Soboyejo,* SUNY Polytechnic Institute

We present a hybrid machine learning (ML) and physics-based computational modeling framework that informs experimentally realizable thin charge-transport-layers for perovskite solar cells (PSCs). Using supervised ML models, 12294 previously published curated datapoints predicted device efficiency. Ten high-efficiency descriptor combinations with greater than 21% power conversion efficiency were mapped into SCAPS-1D (a physics-based simulation software) employing a $FASnI_3$ absorber to assess their physical plausibility. In contrast with the ML predictions, the SCAPS configurations yielded PCEs in the range of 0.03–22.7%. Discrepancies between ML and SCAPS predictions arose from interfacial recombination, optical losses, and physical constraints not explicitly encoded in the ML dataset, describing a tightly focused set of fabrication parameters. Thus, the complementary use of both machine learning and physical simulations offers a time- and material-saving scalable pathway for the development and fabrication of low-cost photovoltaic devices.

10:00am **AIML3-FrM-8 Closed-Loop Ai Workflow for Literature-Guided Perovskite Thin-Film Discovery***, **Jordan Marshall**,* University of Tennessee Knoxville

In this project, we are developing an AI-assisted thin-film materials discovery workflow for halide perovskite solar cells. The goal is to make materials discovery more connected, adaptive, and efficient by bringing together steps that are often handled separately: literature knowledge, predictive modeling, robotic synthesis, and high-throughput characterization. Instead of treating these as independent parts of the research process, the workflow is being designed so that information from one stage can help guide the next. A large part of this work has focused on building a literature-mining agent for perovskite solar-cell papers. The perovskite literature contains a large amount of useful information, including device structures, perovskite compositions, processing conditions, interface layers, power conversion efficiencies, and stability measurements. However, this information is often reported in different ways across papers, which makes it difficult to use directly for machine-learning studies. The literature agent is designed to read papers, extract the relevant information, and organize it into a structured format compatible with the Perovskite Database. The workflow also aims to preserve where the information came from, separate evidence from model-ready features, flag incomplete records, and identify possible duplicate device entries. This makes the dataset update process more transparent and easier to check. The updated dataset can then be used to train sequential machine-learning models for device performance and stability. First, power conversion efficiency is predicted from design, processing, interface, and device-context features while avoiding target leakage from already measured performance values. Stability is then modeled using a similar feature space. We compare design-only stability prediction, stability prediction using model-predicted PCE, and post-fabrication prediction using measured initial device performance. This structure helps separate early design guidance from later experimental decision-making. This work is an important step toward a closed-loop discovery platform for perovskite and related semiconductor thin films. Literature knowledge can update the database, models can suggest promising directions, robotic synthesis can test those ideas, and high-throughput characterization can provide new information that feeds back into the system. While the workflow is still being developed, it provides a practical foundation for more efficient and autonomous thin-film experimentation.

10:30am **AIML3-FrM-10 Multimodal AI Meets Materials Characterization***, **Linda Hung**, Weike Ye, Jithendaraa Subramanian, Daniel Schweigert, Amalya Johnson, Flora Chen,* Toyota Research Institute **INVITED**

Many large-scale materials AI models rely on atomic-resolution crystal structure as a required input. Such models leverage the availability of large computational datasets but can be difficult to apply to experimental workflows. We present multimodal models that are instead built on inputs more commonly available in the lab, such as XRD and composition, and show benchmarks on use cases including noisy XRD and incomplete composition. We then discuss what it takes to build the datasets these models need, and share progress toward that goal.

11:00am **AIML3-FrM-12 Building Informative Materials Datasets Beyond Targeted Objectives***, **Rafael Espinosa Castañeda**,* University of Toronto, Canada, Mexico*; Jason Hattrick-Simpers,* University of Toronto, Canada

Materials science data collection can be expensive, making the reuse and long-term utility ofdatasets critically important for future discovery campaigns. In practice, researchers prioritize a subsetof properties due to research interests. However, ignoring a subset of outcomes in data collectioncampaigns potentially generates datasets poorly suited for future learning tasks. Here, we present a framework for dataset construction that maximizes informativeness for target properties of interest while preserving performance on untargeted ones. Our approach uses diversity-aware selection to ensure broad coverage of the materials space. In noisy experimental dataset construction, we find that without our diversity-aware framework, prediction performance on untargeted properties can degrade by up to sim40% relative to random sampling, whereas applying our framework yields improvements of up to 10\%. For targeted properties, performance can degrade with respect to random sampling by up to 12.5%

without diversity, while our framework achieves gains of up to sim25%. Incorporating diversity into dataset construction not only preserves informativenessfor the targeted properties, but also improves materials coverage for potential future objectives. As aresult, the constructed datasets remain broadly informative across considered and unconsidered outcomes,ensuring unbiased quality entries and mitigating cold-start limitations in subsequent modeling anddiscovery campaigns.

AI/ML/Autonomous Experimentation for Thin Films Processing
Room 317 - Session AIML4-FrM

Panel: Funding Agencies

Moderators: Andreas Werbrouck, Ghent University, Belgium, **Prof. Matthias Young**, University of Missouri

11:15am **AIML4-FrM-13 Panel: Funding Agencies**

Author Index

Bold page numbers indicate presenter