AI/ML for Scientific Discovery
Room 125 - Session AIML-WeM

AI/ML for Scientific Discovery

Moderators: Alain Diebold, University at Albany-SUNY, **Erica Douglas**, Sandia National Laboratories

8:00am **AIML-WeM-1 "Beyond Fingerprinting": Rapid Process Exploration and Optimization via High-Throughput and Machine Learning**, ***Brad Boyce***, Sandia National Laboratories, USA; *R. Dingreville, J. Coleman, E. Fowler, C. Martinez,* Sandia National Labs; *D. Adams,* Sandia National Laboratories **INVITED**

Material properties are governed by composition and associated microstructure dictated by the thermodynamics and kinetics of manufacturing processes.Often, the connectivity between process conditions and the resulting structure and properties is complex, evading full predictivity via high-fidelity modeling.In this work, we are exploring three manufacturing processes where material properties are difficult to predict directly from process settings: physical vapor deposition, electroplating and additive manufacturing (laser powder bed fusion).Each of the three processes offer unique challenges and opportunities.Across these three exemplars, we are augmenting traditional process-structure-property investigations with an accelerated workflow to detect material structure/composition, prognose associated properties, and adapt the associated process to achieve improved product outcomes.This accelerated detect-prognose-adapt cycle is aided by four key elements: (1) automated combinatorial synthesis to enable rapid parameter sweeps, (2) high-throughput evaluation of both conventional and surrogate indicators of material chemistry, structure, and properties, (3) unsupervised learning algorithms to unravel correlations beyond expert cognition, and (4) Bayesian optimization strategies to efficiently explore and exploit high-dimensional process parameter space.In each of these four domains, we take advantage of previously developed capabilities, or where such capabilities are insufficient, we develop novel techniques.For example, in the domain of electroplating synthesis, we have employed an existing robotic pipetting system for formulation of solution chemistries while developing a custom 12-cell parallel electroplating system that enables hundreds of unique conditions to be explored in about a day.While we consider purely data-driven ML algorithms for some correlation analysis, a more interpretable and robust solution includes physical models based on established governing equations.In this regard, we have developed a physics-informed multimodal autoencoder that fuses data from multiple characterization modalities alongside physical models to provide a deeper fingerprint of material state, enabling cluster disentanglement and cross-modal inference.SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525.

8:30am **AIML-WeM-3 Simulations of Epitaxial Inorganic Interfaces Using DFT with Machine-Learned Hubbard U Corrections**, ***Noa Marom,*** Carnegie Mellon University **INVITED**

Epitaxial inorganic interfaces lie at the heart of semiconductor, spintronic, and quantum devices. At an interface between two dissimilar materials physical properties and functionalities may arise, which do not exist in any of the isolated components in the bulk. To predict the structure of domain-matched epitaxial interfaces, we use a combination of lattice matching and surface matching algorithms implemented in the Ogre Python package [J. Chem. Phys., 155, 034111 (2021); J. Phys. Condensed Matter, 34, 233002 (2022)]. To study the electronic and magnetic properties of interfaces we use density functional theory (DFT). Within DFT, the many-body interactions between electrons are described by approximate exchange-correlation functionals. The accuracy of the results hinges on an appropriate choice of functional. We have developed a method of machine learning the Hubbard U correction added to a DFT functional by Bayesian optimization (BO) [npj Computational Materials 6, 180 (2020)]. The DFT+U(BO) method balances accuracy with computational cost, enabling unprecedented simulations of large models of surfaces and interfaces of interest for applications in quantum computing. Examples include InAs and InSb surfaces [Advanced Quantum Technologies, 5, 2100033 (2022)], which are the substrates of choice for superconductor/semiconductor Majorana devices; the HgTe/CdTe and InAs/GaSb interfaces [Phys. Rev. Mater. 5, 084204 (2021)], in which a 2D topological insulator phase may arise; the EuS/InAs interface [Phys. Rev. Mater. 5, 064606 (2021)], proposed as a candidate for the realization of a ferromagnet-semiconductor-superconductor Majorana devices without an external magnetic field; and CdTe as a tunnel barrier to control the coupling strength at the interface between InSb and α-Sn [ACS Applied Materials & Interfaces 15, 16288 (2023)].

9:00am **AIML-WeM-5 On-The-Fly Analysis of RHEED Images During Deposition Using Artificial Intelligence**, ***Tiffany Kaspar***, *J. Pope, S. Akers, H. Sprueill, A. Ter-Petrosyan, D. Hopkins, E. King, J. Drgona,* Pacific Northwest National Laboratory

Modern synthesis methods enable the fabrication of an ever-expanding array of novel, non-equilibrium, and/or metastable materials and composites that may possess unique and desirable functionality. Thin film deposition by molecular beam epitaxy (MBE) can produce atomically precise (or nearly so) materials with a wide range of functional electronic, magnetic, ferroelectric/multiferroic, optical, and/or ion-conducting properties.We are working to employ artificial intelligence (AI)-accelerated analysis of in situ and ex situ data streams for on-the-fly feedback control of the MBE deposition process that will enable targeted synthesis of novel materials with desired structure, chemical stability, and functional properties.Here we present a machine-learning-enabled framework for analysis of reflection high energy electron diffraction (RHEED) pattern images in real time (one image per second).Our approach utilizes pre-trained neural networks for image preprocessing, statistical analysis to identify changepoints in the images over time, and network graph analysis methods to precisely identify and classify changes.We demonstrate this framework using RHEED images collected from the deposition of epitaxial oxide thin films such as anatase TiO_2 on $SrTiO_3$(001).Advantages and disadvantages of our approach will be discussed, as well as its potential use as the basis for on-the-fly feedback control of deposition parameters.

9:15am **AIML-WeM-6 An Unsupervised Machine Learning Approach for the Identification of Adsorbates on a Semiconductor Surface: BCl_3 Adsorption on Si(100)**, ***Azadeh Farzaneh***, University of Maryland, College Park; *C. Wang, S. Kalinin,* University of Tennessee Knoxville; *R. Butera,* Laboratory for Physical Sciences

A more thorough understanding of the reaction of molecular precursors on crystalline and amorphous surfaces will play a significant role in the optimization of industrially relevant processes, such as chemical vapor deposition and atomic layer deposition. Here, we explore an unsupervised machine learning approach to identify reaction products related to molecular precursor adsorption on a semiconductor surface and provide a general framework for analyzing surface species. In particular, we focus our investigations on the adsorption of BCl_3 on Si(100) using scanning tunneling microscopy (STM). We designed an unsupervised workflow that results in the identification of distinct surface moieties and their relative concentrations following the initial adsorption of BCl_3 and subsequent decomposition reactions on the surface. While previous methods have relied on manual cropping of STM images based on defect coordinates, our workflow isolates surface features from the base lattice to generate a training dataset. Two key components of the Si(100) surface are taken into account for isolating surface features: (1) steps and (2) orientation of Si dimer rows. This unsupervised method eliminates the need for manual labeling an untenable amount of surface features, thereby removing any label bias introduced by the operator. It circumvents the bottleneck of machine learning workflows when experimental conditions change and new labeled data is required. We optimize the performance of the unsupervised neural networks by selecting the proper number of feature classes that minimize the image-to-image identification error of distinct surface features in a given experimental data set. This methodology can be generalized and extended to other material systems to provide insight into reactions on surfaces.

9:30am **AIML-WeM-7 Quantum and Classical Supervised Learning Analysis of Synthesis–Structure Relationships in Epitaxially–Grown Semiconductors**, ***Andrew Messecar***, Western Michigan University; *S. Durbin,* University of Hawai'i at Mānoa; *R. Makin,* Western Michigan University

The design of material synthesis experiments occurs within highly multidimensional processing spaces that are defined by many design parameters. Identifying the optimal values for each synthesis parameter is often performed through an expensive, Edisonian, trial–and–error approach to experiment design. Considerable interest exists in the development of machine learning–based approaches for the rapid and accurate identification of optimal materials designs and synthesis conditions yielding material samples that display target properties of interest. In this work, data detailing hundreds of plasma–assisted molecular beam epitaxy

(PAMBE) growth trials each of ZnO and various nitride semiconductors have been organized into separate, composition–specific data sets. For each growth record, the complete set of experiment parameters (substrate temperature, effusion cell temperatures, growth duration, etc.) are associated with binary measures of crystallinity as well as surface morphology as determined by *in–situ* reflection high–energy electron diffraction (RHEED) patterns. A Bragg–Williams measure of lattice disorder (S^2) is included as an additional, continuous figure of merit for investigation. Quantum and conventional machine learning algorithms – including logistic regression, tree–based algorithms, and quantum support vector machines – are trained on the data to investigate which growth parameters are most statistically significant for influencing crystallinity, surface morphology, and S^2. When predicting the occurrence of monocrystalline GaN PAMBE, we show that supervised machine learning algorithms for quantum computers can display significant advantage over their classical machine learning counterparts. The class conditional probabilities of obtaining single crystalline and atomically flat thin film crystals are predicted across processing spaces of the two PAMBE operating parameters determined to be most statistically important. S^2 is also forecasted across the same growth spaces. These predictions are compared to conventional experimental wisdom as well as the results described within published literature regarding the PAMBE growth of these materials. The predictions indicate that different growth conditions are of interest depending on whether a single crystalline sample, a flat surface, or a well–ordered lattice is desired. The superior generalization performance displayed by the quantum machine learning algorithms when predicting GaN crystallinity implies possible advantage gained via quantum algorithms when studying synthesis–structure relationships in other material systems.

9:45am **AIML-WeM-8 'DECIEDD with CARE' - Building an Autonomous Ecosystem for the Discovery and Optimization of Metal Nanoparticle Inks**, ***J. Elliott Fowler,*** Sandia National Laboratories*; N. Trask,* University of Pennsylvania*; M. Kottwitz, N. Bell, A. Hesu, A. Roth,* Sandia National Laboratories*; J. Hanna,* University of Wisconsin - Madison*; J. Foster,* University of Texas at Austin*; J. Boissiere,* Sandia National Laboratories

The end-to-end design and manufacturing of printed circuit boards substrates, a ubiquitous and critical technology in energy storage, communication, and defense systems, is poised to undergo a transformation following developments in additive manufacturing within the last decade. These advancements include droplet-on-demand inkjet printing of conductive inks—suspensions of metallic nanoparticles, graphene, carbon nanotubes, etc.—onto dielectric substrates. Despite extensive research, few printed commercial inks possess the conductivity and robustness desired by high-reliability design agencies. A major contributor to the limited availability of viable inks is the enormous parameter space of processing conditions and material structure, property, and performance criteria that must be balanced during development.

For this reason, Sandia National Laboratories, together with university partners, has engaged in the design and implementation of an autonomous materials discovery platform to efficiently (1) synthesize Cu, Ag, and Au nanoparticles, (2) formulate those nanoparticles into inks, and (3) print those inks to form devices. At each step of the process, characterization data of the structure, properties, and performance is provided to a machine learning algorithm utilizing a self-consistent and scalable/tunable data schema and data management application. Initial campaigns have utilized off-the-shelf machine learning methods to autonomously optimize the size and dispersity of silver nanoparticle via manipulating the stoichiometric ratio of mono-, di- and tri- functionalized carboxylic acid ligands, amongst other variables. Concurrently, development of bespoke solutions such as multifidelity reinforcement learning and scientific machine learning continues to address the challenges of relatively sparse data sets, multimodality and fidelity, and the need identify underlying process-structure-property-performance relationships.

SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525. SAND2024-06034A.

11:00am **AIML-WeM-13 Machine Learning for Accelerating Atomic Layer Deposition Process Optimization**, ***Minjong Lee****, D. Kim, T. Chu, D. Le, J. Kim, D. Narayan, J. Kim,* University of Texas at Dallas

In the era of Industry 4.0, artificial intelligence (AI) and machine learning (ML) are revolutionizing manufacturing processes, enhancing flexibility and efficiency while minimizing process errors [1]. ML integration is particularly valuable in semiconductor fabrication, where current techniques often require over 30 steps to deposit a single high-quality film layer. Out of semiconductor fabrication techniques, atomic layer deposition (ALD) stands out as a promising deposition method due to its precision in atomic-scale film engineering and compatibility with advanced 3D architectures. However, ALD involves numerous critical parameters affecting film quality, and comprehensive exploration of all conditions is economically unfeasible, with each experimental run costing over a thousand dollars per wafer, plus additional labor and analysis costs [1]. This limitation constrains inspection ranges of ALD conditions, hindering the development of accurate, atomic-scale process models. The incorporation of AI/ML technologies into ALD processes is thus beneficial in improving cost-efficiency for process optimization by potentially decreasing the volume of data necessary.

This study presents a deep neural network (DNN) framework for ML-driven ALD processes, focusing on optimizing hafnium oxide film deposition. While previous ML applications in ALD processes have primarily focused on monitoring film thickness [2], we investigate the film quality, particularly density, which is crucial property for film performance. We incorporate film density using the wet etch rate (WER) test into our DNN framework, leveraging the principle that higher density films have more tightly packed atoms (i.e., slow WER). In this presentation, we report that the DNN case studies involve four ALD process parameters (deposition temperatures, precursor temperatures, and sample locations (x, y)) and three film properties (thickness (T_{ox}), refractive index (RI), and WER). The DNN system demonstrates significant advantages in terms of high prediction accuracy, exceeding 90% for all film properties within prediction ranges of ±1.0 nm, ±0.04, and ±0.9 Å/min for T_{ox}, RI, and WER, respectively. We also investigate prediction maps of film properties for further efficiency toward“digital twin” of ALD process.

[1] K. J. Kanarik et al., Nature, 616, 707 (2023).

[2] A. Arunachalam et al., JVST A 40, 012405 (2022).

11:15am **AIML-WeM-14 Utilizing a Machine Learning Potential for Investigating Defects in Hexagonal Boron Nitride**, ***John Janisch****, D. Le, T. Rahman,* University of Central Florida

Defects in materials can bring rise to many unique properties and behaviors, and engineering these defects to appear in desirable ways is of utmost importance in maximizing these properties and behaviors. We need atomistic simulations with high accuracy to study not only the defect itself, but also the area around the defects that are affected by its presence. This influence can range from small, local areas near single atom defects, such as a vacancy, to much larger, sprawling areas near large defects, such as grain boundaries. It is not feasible to study large systems with typical high-accuracy methods such as Denstiy Functional Theory (DFT) due to prohibitively long calculation times but utilizing Machine Learning we can construct Machine Learning Potentials (MLP) and overcome these issues. Here, we will present the performance of a MLP for Hexagonal Boron Nitride with and without defects, demonstrating the DFT-level accuracy prediction of material properties such as phonon dispersion as well as defect properties such as vacancy structures and the motion of grain boundaries.

11:30am **AIML-WeM-15 Accelerating Innovation: Using AI for Process Pathfinding**, *L. Medina,* SandBox Semiconductor*;* ***Mokbel Karam****, S. Sirard, M. Chopra,* Sandbox Semiconductor

While AI is becoming more common in high volume manufacturing, it remains underleveraged in R&D and technology development settings. In these research environments, where data sets are often sparse and the process requirements constantly changing, it is challenging to establish robust data pipelines to take advantage of traditional AI/ML approaches. In this work, we show how AI can be used to provide key process and pathfinding insights for even small datasets, using a gate-all-around etch (GAA) as a case study. Using the software platform SandBox Studio™ AI, we demonstrate how physics-enabled AI can be used to (1) improve process metrology, (2) generate predictive models of the process space, (3) quickly rule out insufficient process regimes and target more viable spaces, and (4) evolve with pathfinding development cycles with novel process parameters permutations. We first collect a limited set of metrology data from disparate sources and use it to generate a high accuracy predictive model of the process space for the GAA etch. We specifically target common metrology sources which are non-destructive, and cost-effective, including Optical Critical Dimension (OCD) scatterometry, ellipsometry, and CD-SEM. We then illustrate how an AI-based model can be used to capture the experimental process space accurately and efficiently. Next, we demonstrate a search strategy for identifying an optimal set of process conditions subject to a defined set of constraints. We highlight that entire process regimes can be visualized, searched, and/or ruled out using the predictive model. Lastly, representative of an R&D environment, we

illustrate how the model can be updated to predict outcomes for new process parameters.

11:45am **AIML-WeM-16 AI-Driven Synthesis of Thin Films with Pulsed Laser Deposition***, **Sumner Harris,*** Oak Ridge National Laboratory*; A. Biswas,* University of Tennessee, Oak Ridge National Laboratory*; C. Rouleau, A. Puretzky, S. Yun, R. Vasudevan, D. Geohegan, K. Xiao,* Oak Ridge National Laboratory

Traditional methods for synthesizing thin films have typically involved slow, sequential processes requiring significant human intervention, with material optimization often relying on a mix of expertise and chance discoveries. Recent technological progress in autonomous synthesis experiments which combine automated synthesis and characterization with artificial intelligence (AI) has enabled rapid exploration of large parameter spaces, promising to greatly accelerate and advance our understanding of synthesis science. In this presentation, I will highlight the development of two flexible, autonomy-enabled pulsed laser deposition (PLD) platforms: one incorporating real-time, in situ gas-phase and optical diagnostics, and the other featuring in vacuo robotic transfer for subsequent characterization. I will detail how we merged in situ, real-time diagnostics and characterization with high-throughput methodologies and cloud connectivity to execute an autonomous synthesis experiment using PLD. We synthesized ultrathin WSe_2 films via co-ablation of two targets, employing real-time laser reflectivity for precise thickness control, and achieved a tenfold increase in throughput over conventional PLD workflows. Bayesian optimization with Gaussian process regression, utilizing in situ Raman spectroscopy, directed the synthesis process and autonomously identified the optimal growth windows after sampling 0.25% of a 4D parameter space. Furthermore, the Gaussian process surrogate model predicted process-property relationships, revealing two distinct growth regimes and providing a broader understanding of the synthesis space than would be feasible in traditional PLD workflows. Our platforms and methodologies enable the autonomous synthesis of any material that can be grown by PLD, promising to greatly accelerate thin film synthesis with automated, AI-driven workflows.

12:00pm **AIML-WeM-17 Active-Learning Based Structure Discovery in STM***, **Ganesh Narasimha,*** Oak Ridge National Laboratory*; S. Hus,* Oak Ridge National Lab (ORNL)*; A. Biswas,* Oak Ridge National Laboratory, USA*; D. Kong,* University of Virginia, USA*; Z. Gai, R. Vasudevan,* Oak Ridge National Laboratory, USA*; M. Ziatdinov,* Pacific Northwest National Laboratory

Scanning tunneling microscopy (STM) is a widely used tool for atomically-resolved imaging of materials and their surface energetics. However, the optimization of the imaging conditions is a time-consuming process due to the extremely sensitive tip-surface interaction. Additionally, conventional experimentation involves sequential imaging procedures, and the material-property correlations are usually deciphered by an operator based on auxiliary spectroscopic information. This limits the experimental throughput. Here we show a Bayesian optimization-based framework to improve imaging conditions in real time [1]. Further, we demonstrate a characterization technique using a probabilistic deep learning framework to automatically correlate structure-property relationships in a Europium-based semimetal, $EuZn_2As_2$ [2]. The data-driven inference is dynamically incorporated to drive STM exploration in regions that optimize a given material property. This framework employs a sparse sampling approach to efficiently construct the property space using minimal measurements, as little as 1 % of the data required in conventional hyperspectral imaging methods. We further demonstrate property-guided sample exploration using a multiscale-process implementation for the autonomous discovery of structural origins of an observed material property. Our findings reveal correlations of the electronic properties unique to local defect density, surface terminations, and point defects [3]. The deep learning framework offers future implications to study and induce dynamic processes such as molecular manipulations to assemble artificial quantum structures.

References:

1. Narasimha, G., Hus, S., Biswas, A., Vasudevan, R., & Ziatdinov, M. (2024). Autonomous convergence of STM control parameters using Bayesian optimization. *APL Machine Learning*, *2*(1).
2. Blawat, J. *et al.* Unusual Electrical and Magnetic Properties in Layered EuZn2As2. *Advanced Quantum Technologies* **5**, 2200012 (2022).
3. Narasimha, G., Kong, D., Regmi, P., Jin, R., Gai, Z., Vasudevan, R., & Ziatdinov, M. (2024). Multiscale structure-property discovery via active learning in scanning tunneling microscopy. *arXiv preprint arXiv:2404.07074*.

Author Index

Bold page numbers indicate presenter