## Electronic Materials and Photonics
## Room 207 A W - Session EM2+AIML+AP+CPS+MS+SM-TuA

## Advances in AI and Machine Learning within the Semiconducting Industry

**Moderator: Erica Douglas**, Sandia National Laboratories

4:00pm **EM2+AIML+AP+CPS+MS+SM-TuA-8 Improved Design-of-Experiments and Process Modeling with Generative AI***, Somilkumar Rathi, Muthiah Annamalai,* Panmo LLC

Small volume semiconductor, photonic and materials manufacturing largely uses One-Factor at-a time (OFAT) to discover process window instead Design of Experiments (DOE). We demonstrate, *Panmo Confab,* a Generative AI based DOE and process-flow-design platform to accelerate process window discovery. Large volume semiconductor, photonic and materials automation tools have relied on statistical process control (SPC), design of experiments (DOE) and yield modeling techniques which are fairly manual and depend on specialized tools and deep knowledge [1,2]when such tools are not used we get a sub-optimal outcomes for process development teams through using one-factor at a time (OFAT). In this article we report, and demonstrate, *Panmo Confab* a Generative AI based process flow tracking and design of experiments platform to accelerate flow designs and generating DOEs. Previously our tool was used without Generative AI, features to show improvement in process discovery for plasmonic nanocavity fabrication [4]. The unique innovation of our tool is to use the emerging technology of large language models (LLM), like BERT or ChatGPT [5,6] and science of causality [3] to enable generation of process flows with a description. Our tool is presented in both on-premises and Software-as-a-Service (SaaS) formats.

References:

1. Montgomery, D. C. Design and analysis of experiments. (John Wiley & sons, 2017).

2. May, G. S., & Spanos, C. J. Fundamentals of semiconductor manufacturing and process control. (John Wiley & Sons, 2006).

3. Pearl, Judea, and Dana Mackenzie. The book of why: the new science of cause and effect. (Basic books, 2018).

4. Annamalai, M., Rathi, S., Methodology for robust process window discovery in plasmonic nanostructures", Proc. SPIE 13111, Plasmonics: Design, Materials, Fabrication, Characterization, and Applications XXII, 131110A (2024).

4:15pm **EM2+AIML+AP+CPS+MS+SM-TuA-9 Foundation Models in Semiconductor R&D: A Study on Segment Anything***, Fei Zhou,* Sandisk Corporation

Quantitative analysis of scanning and tunneling electron images is crucial in semiconductor manufacturing, particularly for defect detection, process margin checking, and morphology quantification. Traditional AI/ML approaches, such as using recurrent neural networks, require large labeled datasets and extensive transfer learning to generalize across different imaging conditions. Developing a usable AI tool for proof-of-concept demonstrations demands significant engineering effort and GPU resources, making these methods costly and time-consuming. These challenges are especially pronounced in semiconductor R&D, where fast turnaround, high accuracy, and efficient use of engineering resources are essential.

The Segment Anything Model (SAM) introduces a novel training free segmentation approach, eliminating the need for task-specific retraining while providing robust and efficient segmentation across diverse semiconductor imaging requirements. This paper explores SAM's application in semiconductor image analysis, demonstrating its ability to segment complex nanoscale features without prior dataset exposure. We assess SAM's performance in automated defect detection, where challenges such as varying defect morphology, background noise, and process-induced variations exist. With appropriate prompting and post-processing techniques, SAM adapts to different imaging conditions, offering a rapid, low-cost, and high-accuracy solution.

Additionally, we examine SAM's limitations, particularly in scenarios where the region of interest is small and contains limited useful pixel data. By employing image enhancement techniques, we demonstrate how SAM can effectively segment defects even in low-information conditions. Furthermore, we explore how integrating grounding techniques with SAM can expedite segmentation post-processing, further improving efficiency in real-world applications.

Our case studies show that SAM significantly reduces resource overhead and enables semiconductor image analysis automation, achieving saving of >100 engineering hours and >20 GPU hours per project. Its foundation model architecture allows it to generalize across different defect types, backgrounds, and imaging techniques without additional data labeling or fine-tuning. These findings suggest that integrating SAM into semiconductor workflows enhances efficiency, lowers costs, and accelerates R&D decision-making by providing a scalable and cost-effective solution for high-precision image segmentation. This study highlights the transformative potential of foundation models in semiconductor engineering, paving the way for broader adoption of AI-driven automation across the industry.

4:30pm **EM2+AIML+AP+CPS+MS+SM-TuA-10 MOFCreatioNN: A Novel Modular Machine Learning Approach for Designing 'Undesignable' Metal-Organic Frameworks.** *, Satya Kokonda,* Charter School of Wilmington

Many critical material discovery processes remain too complex for traditional computational modeling, necessitating costly and time-intensive experimentation. Here, we present a generalizable, application-driven methodology for material design, demonstrated through a case study in photocatalysis. Using a reinforcement learning ensemble, we generated 120,000 novel metal-organic frameworks (MOFs) optimized for $CO_2$ heat of adsorption and $CO_2/H_2O$ selectivity. A multi-objective fitness function—incorporating stability, catalytic potential, cost, sustainability, and adsorption properties—enabled computational modeling of photocatalytic performance aligned with industrial criteria. To enhance efficiency and prevent feature overfitting, a predictor funnel system iteratively filtered low-scoring candidates, narrowing the search space to 17,315 MOFs and improving computational efficiency by 313%. Our system, MOFCreatioNN, designed two high-performing, de novo MOFs: a Cr-based MOF with a photocatalyst score 239% higher than the control, and a Mn-based MOF that outperformed all baselines across every evaluated metric, demonstrating robustness against imperfect fitness functions. The proposed MOFs meet key synthesis and operational thresholds—including X-ray diffraction consistency with known structures, predicted synthesizability, temperature stability >300°F, and viable water stability—making them practical for real-world applications. Furthermore, we identify actionable design heuristics, such as the significant impact of the $N_2$62 metal cluster on photocatalytic performance. By integrating industrial considerations such as cost, stability, and environmental viability into the modeling process, this work showcases a scalable framework for the AI-driven design of industrially relevant materials in domains previously considered computationally intractable.

## Plasma Science and Technology
## Room 201 ABCD W - Session PS+AIML-ThA

## Plasma Modelling AI/ML

**Moderators: Kenji Ishikawa**, Nagoya University, Japan, **Angelique Raley**, TEL Technology Center, America, LLC

2:15pm **PS+AIML-ThA-1 Machine Learning for Low Temperature Plasma Applications**, ***Abhishek Verma***, *Kallol Bera, Shahid Rauf,* Applied Materials, Inc. **INVITED**

Low temperature plasmas are used for numerous depositions and etch applications in the semiconductor industry. The field is rapidly advancing driven by volumes of multimodal and complex spatiotemporal data from both experiments and simulations. Machine learning in combination with plasma modeling and simulation offers a wealth of techniques that could play key role in plasma source discovery, design and decision making. These techniques can also augment domain knowledge for plasma reactor control and process development. In this talk, we present our work on machine learning applications to modeling, control, and optimization of plasma chambers. To overcome the challenge of high computational cost associated with high fidelity plasma models for rapid and many-query analyses, we present a deep learning based non-linear surrogate modeling method. Our numerical experiments on capacitively coupled plasmas show that deep learning-based model can learn an efficient latent space representation of spatiotemporal features of plasma characteristics. Moreover, we extended this approach with physics informed neural networks to improve predictive accuracy and generalization while being data efficient. Physics informed approaches could also effectively incorporate expert knowledge while learning physics implicitly. Furthermore, we present application of regression methods for circuit estimation of collisional sheath in moderate pressure capacitively couple plasmas. The novel sheath model which includes collisional effects, ion current responses to sheath voltage and harmonics based resistive elements, builds on parametric flexibility using machine learning while maintaining interpretability. The talk outlines machine learning methodologies for modeling, optimizing, and controlling plasmas for semiconductor applications.

2:45pm **PS+AIML-ThA-3 Contour-Based Objectives for Robust Etch Model Selection**, ***Chad M. Huard***, *Prem Panneerchelvam, Shuo Huang,* KLA*; Lewis Hill, Janet Hopkins,* KLA UK*; Mark D. Smith,* KLA

As device scaling increasingly relies on 3D rather than CD scaling, etch has become a critical and challenging step, often limiting further scaling. The demand for high-quality, predictive etch models is growing, yet our understanding of surface mechanisms during dry etching remains limited. Techniques like XPS, SIMS, and AES provide clues to surface reactions, but the pathways are not immediately clear. First-principles computational methods such as DFT, quantum MD, and classical MD offer insights but are constrained by computational resources and turnaround times.We present a Monte Carlo profile model that bridges the gap between first-principles and empirical models by using simplified chemistry mechanisms calibrated with experimental data. Traditional models often rely on 'best-effort' mechanisms, risking calibration issues due to high dimensionality or model errors due to omission of critical pathways. We propose a unified method for evaluating etch mechanisms using rigorous contour-based objectives, which maximizes entitlement from metrology data and results in better model development/selection compared to gauge-based metrics. This approach identifies the simplest model that fits the data, addresses degeneracy in models and calibration objectives, and enhances model predictiveness.

3:00pm **PS+AIML-ThA-4 NAND Pillar Etch: Plasma and Feature Profile Modeling in Dry Etch Process**, ***Harutyun Melikyan***, *Ebony Mays,* NAND Pathfinding - Micron Technologies*; Ali Bhuiyan, Sumeet Pandey,* Advanced Modeling - Micron Technologies*; Jagannath Mahapatra,* Micron Technologies, USA

In this work we developed a model to study the Feature Profile Modeling (FPM) in the dry etch plasma process for NAND pillar etch. The model developed takes in process parameters, that is process knobs such as temperature, pressure, flowrates, Power, Frequency, Voltage as inputs. The output from the model is Feature profile information such as Etch rate, Etch Depth, Variation of CD with height, Twisting, Ellipticity, Necking (HM), Bowing (ONO) etc. This methodology makes possible the ability to correlate process knobs on an equipment directly to the feature profile. This can enable us to get a detailed sensitivity analysis of feature profile with respect to process knob on the equipment (like constructing a sort of digital twin for that equipment). In addition, the feature profile (for HAR) for the future nodes can be inferred from process knobs and recipe information even before running the experiments.

3:15pm **PS+AIML-ThA-5 Machine Learning of Simulated Nanosecond UV Laser Ablation Plumes**, *Jacob Paiste,* University of Alabama at Birmingham*; Sumner Harris,* Oak Ridge National Laboratory*; Shiva Gupta,* University of Alabama at Birmingham*; Eric Remington,* Samford University*; Robert Arslanbekov,* CFDRC Research Corporation*;* ***Renato Camata***, University of Alabama at Birmingham

Laser-generated plasmas are a rich laboratory of complex spatiotemporal phenomena emerging from coupled thermodynamic, electromagnetic, and quantum mechanical processes. The strength of laser-solid and laser-plasma interactions can vary over multiple orders of magnitude while gradients of density, temperature, and flow velocity give rise to shocks, instabilities, and turbulence in multiphase flows. Deep learning can be used to encode these complex spatiotemporal dynamics to discover correlations between the conditions under which a laser-generated plasma is produced—including the wide chemical and thermophysical diversity of ablation targets—and the resulting plasma flows. Predictive models can then be built to infer the fundamental properties of irradiated solids and plasmas, enabling a new experimental modality for measuring material properties like thermal conductivity or critical temperature. However, no databases of experimental or simulated laser-generated plasmas currently exist, so proof-of-concept for the efficacy of deep learning for this task is difficult to obtain.

In this work, we carry out a deep learning study on synthetic laser-generated plasma data. The synthetic data sets are produced using a combined laser ablation-fluid dynamics simulation based on the Hertz-Knudsen model, including phase explosion when a target temperature exceeds the thermodynamic critical temperature. The model is implemented on an open-source Adaptive Cartesian Mesh framework that enables laser ablation plume simulations out to centimeter distances over tens of microseconds for any elemental material with well-defined thermophysical parameters.

We generate a training dataset by simulating UV nanosecond pulsed laser ablation of elemental targets of Be, B, Na, Mg, Al, Sc, Ti, V, Fe, Co, Cu, Zn, Rb, Cs, Ta, W, and Pt with systematic variation of laser fluence (1–10 J/cm$^2$) and laser spot area (0.8–13 mm$^2$). We use (2+1)D convolutional neural networks (CNNs) to encode spatiotemporal plume dynamics for regression and classification problems using our simulated data. Results indicate that given a plume image sequence and associated laser parameters, we can not only predict which element the plasma was generated from with high confidence but also predict the set of thermophysical properties of the material. These results serve as proof-of-principle for plasma plume dynamics as strong predictors of fundamental material properties and motivate new experimental measurement techniques using laser ablation.

4:00pm **PS+AIML-ThA-8 PSTD Business Meeting & Awards Ceremony**,

## AI/ML Mini-Symposium
## Room Ballroom BC - Session AIML-ThP
## AI/ML for Scientific Discovery Poster Session

**AIML-ThP-1 AI Agents for Semiconductor Processing: A New Benchmark for Autonomous Materials Synthesis**, ***Angel Yanguas-Gil,*** Argonne National Laboratory

Over the past year there has been an increasing interest in leveraging foundation and large language models to design AI agents that can interact withexperiments to solve materials science and synthesis problems. One of the challenges of this approach is that testing the performance of these agents require access to automated labs. In contrast to benchmarks testing abilities such as knowledge, math skills, or reasoning, there is a lack of benchmarks that can help both design and evaluate agents without the access to dedicated experimental facilities.

In this work, we introduce Semibench, a benchmark to evaluate AI agents' ability to operate and solve synthesis challenges in the context of semiconductor processing. This benchmark introduces two core ideas: first, it introduces virtual tools that simulate the output of real life experiments. This allows us to test an agent's ability to solve a wide range of challenges involving different tool configurations, amount and nature of information that is accessible, and process complexity. Second, it focuses on the concept of microtasks, challenges designed to have a unique solution. This allows us to define quantitative performance metrics for the agent based on how far the proposed solution is to the ground truth. For Semibench, we have focused on three different techniques that are commonly used in the context of microelectronics: atomic layer deposition, sputtering, and reactive ion etching. For each challenge in the benchmark, agents are exposed to a collection of virtual tools and asked to solve specific questions by providing a sequence of synthesis steps. These steps involve selecting the right configurations for each of the tools, such as the precursor channels in the case of ALD, or the targets and power for sputtering, or the etching recipe for RIE.

We have applied this benchmark to agents based on leading large language models, such as OpenAI's o1 an o3 family of reasoning models. The results show that these agents can correctly identify the sequence of steps in a wide range of conditions. However, they struggle when they need to use quantitative data that is not provided explicitly to solve these challenges. These results provide useful information about how to design useful models and their limitations for thin film applications.

**AIML-ThP-2 Domain Knowledge + AI for Chemically Accurate Potentials: Application to Diamond Surfaces**, ***John Isaac Enriquez,*** Princeton University Plasma Physics Lab*; Yoshitada Morikawa,* Osaka University, Japan*; Igor Kaganovich,* Princeton University Plasma Physics Lab

Machine learning interatomic potentials (MLIPs) are powerful tools for accelerating atomistic simulations, but their reliability depends critically on training set construction. A common strategy is to build universal MLIPs from open-source databases, offering transferability but often sacrificing accuracy particularly in surfaces and interfaces with highly diverse chemical environments. These databases are dominated by equilibrium structures, leaving reaction pathways undersampled, forcing potentials to extrapolate in chemically critical regions—a limitation in catalysis, surface chemistry, and defect dynamics where reactive events dominate. Specialized MLIPs built via active learning can achieve higher accuracy but typically rely on molecular dynamics (MD) and committee models to sample configuration space. Because rare reactions are unlikely to appear within accessible timescales, discovery is left to chance, often requiring long simulations or many iterations. As a result, such MLIPs may fit training data well but fail to capture the most chemically relevant regions.

To address this limitation, we introduce DIAL (Domain-Informed Active Learning), a chemically targeted strategy that augments conventional active learning. Rather than relying solely on MD and uncertainty-driven sampling, DIAL incorporates both established reaction pathways and those identified via nudged elastic band (NEB) calculations. Training datasets are enriched with configurations along these pathways, particularly near transition states. By integrating data-driven active learning with domain knowledge of chemical processes, this approach ensures that the potential is trained on the chemically important regions of configuration space.

Using DIAL, we developed a specialized MLIP for diamond surfaces and interface reactions. The potential enabled large-scale molecular dynamics simulations that reproduced graphitized and oxidized surface morphologies and reaction products, while providing new insights relevant to diamond-based electronics and quantum technologies. In particular, the model captured thermal degradation mechanisms and suppression, facet-dependent oxidative etching, and suggested strategies for controlling surface termination to preserve quantum-relevant color centers. Although demonstrated on diamond, the DIAL framework is general and applicable to other reactive materials systems, including catalysis and battery interfaces.

These results demonstrate how DIAL bridges data-driven methods with domain expertise, highlighting the value of collaboration between materials scientists and AI specialists in advancing the next generation of materials discovery.

**AIML-ThP-3 AI Operating System for Accelerating Semiconductor R&D Process Development**, ***Suresh Ayyalsamy,*** *Manish Sharma,* Elucida Labs

Advanced plasma etch and deposition process development in semiconductor R&D requires the simultaneous optimization of dozens of interdependent parameters against stringent nanometer-scale metrics. Today, process engineers face a fragmented workflow characterized by siloed data, inefficient experimentation, manual analysis, and limited integration between process settings and physical outcomes. These bottlenecks slow discovery, drive up costs, and hinder knowledge transfer across teams.

We present Elucida Labs, an AI-native operating system designed to transform semiconductor R&D environments across both etch and deposition. Our system enables process teams to reduce experimental burden, shorten learning curves, and converge to target specifications faster. By embedding AI-driven intelligence directly into R&D workflows, Elucida Labs demonstrates how AI can amplify human expertise, accelerate innovation, and reshape the economics of semiconductor process development.

**AIML-ThP-4 Physics-Informed Neural Networks for One-Dimensional Capacitively Coupled Plasma Physics Problems**, ***Uvini Balasuriya Mudiyanselage,*** *Jesse Jing,* Arizona State University*; Abhishek Verma, Kallol Bera, Shahid Rauf,* Applied Materials Inc.*; Kookjin Lee,* Arizona State University

Physics-Informed Neural Networks (PINNs) offer a flexible framework for solving coupled partial differential equations by embedding physical laws directly into thetraining process. In this work, we develop and evaluate a PINN approach for modeling one-dimensional capacitively coupled plasma (CCP) discharges, governed by electron continuity equation under the drift-diffusion approximation and uniform ion density assumption, coupled with Poisson's equation for self-consistent electrostatic plasma description. The governing equations are non-dimensionalized to improve numerical stability and facilitate learning across disparate physical scales. The model consists of two separate fully connected networks—one for electron density and one for electric potential—augmented with Fourier Feature Mapping to capture multi-scale spatial variations and trained with exact Dirichlet boundary conditions enforced for both electron density and potential. Collocation points are sampled throughout the spatio-temporal domain to compute physics-based residuals directly. Our PINN approach successfully approximates the finite difference method (FDM) solution, achieving an average $L^2$ relative error of 3.55% for electron density and 3.89% for electric potential over spatio-temporal domain. To address training stiffness and gradient flow issues commonly observed in multi-equation PINNs, we are currently exploring adaptive loss balancing via gradient-based reweighting, as well as Neural Tangent Kernel (NTK)analysis. Preliminary results reveal a significant imbalance in the convergence rates of the two governing equations: the continuity equation loss decreases much faster than that of the Poisson's equation, necessitating disproportionately higher loss weights for the Poisson term to achieve balanced convergence. The model is currently being extended to include ion continuity and momentum conservation equations.

# Author Index

## Bold page numbers indicate presenter