NIST Artificial Intelligence for Materials Science (AIMS) 2026

June 16-17, 2026
Gaithersburg, MD

As part of the JARVIS workshop series, the 7th Artificial Intelligence for Materials Science (AIMS) is a workshop aimed at getting together experts from industry, academia, and government to facilitate highly technical dialogue on the intersection of AI and materials science. Some of the key research areas for materials AI that will be discussed at the meeting are: developing well-curated and diverse datasets, choosing effective representations for materials, inverse materials design, integrating autonomous experiments and theory, challenges and advantages of self-driving laboratories, merging physics-based models with AI models, and choosing appropriate algorithms/workflows. Lastly, uncertainty quantification in AI-based predictions for material properties and issues related to building infrastructure for disseminating AI knowledge are of immense importance for making AI-based materials investigation successful. This workshop is intended to cover all of these challenges.

Schedule

Day 1: June 16		
Session I: 9:00-12:00		
Time	**Speaker**	**Title**
9:00-9:10	Kathryn Beers	Opening remarks
9:10-9:25	Francesca Tavazza	Overview and Logistics
9:25-9:45	Boris Kozinsky	Physics based learning of foundation models of microscopic interactions
9:45-10:05	Samarjeet Prasad	Accelerating Chemistry and Materials Innovation with NVIDIA ALCHEMI
10:05-10:25	Michael Taylor	Bringing AI to f-element Chemistry
10:25-10:45	Danish Khan	Physics-aware electronic-structure models for reliable materials AI
10:45-11:05	Break: Group Picture	
11:05-11:25	Stefano Falletta	Machine Learning Material Response to Electric Fields
11:25-11:45	Sanjeev Raja	Generative Models as Flexible Priors for Accelerating Long Timescale Molecular Simulation
Lunch: 12:00-1:00		
Session 2: 1:00-3:20		
1:00-1:20	Peter Beaucage	Deploying Autonomous Experimentation in Government and Industry
1:20-1:40	Jie Xu	Polymer Innovation Through AI-driven Autonomous Labs
1:40-2:00	Keith Task	Data Analytics in Research at BASF: Guiding Decision Making Through Efficient Experiments, Optimization, and ML & AI
2:00-2:20	Break	
2:20-2:40	David Elbert	Unlocking Autonomy: Event Driven Linked Data for Scientific Workflows
2:40-3:00	Linda Hung	Addressing the Crystal Structure Bottleneck in Materials Models
3:00-3:20	Tyler Martin	Integrating Physics, Chemistry, and Formulation Intuition into the Autonomous Formulation Lab
Poster Session 3:30-5:00		

Day 2: June 17		
Session 3: 8:45-12:00		
Time	**Speaker**	**Title**
8:45-9:05	Dung-Yi Wu	Applications of materials informatics to Aerospace-Relevant Problems
9:05-9:25	Jason Hattrick-Simpers	Small Smart Data Big Dumb Data and Models Don't Extrapolate
9:25-9:45	William Ratcliff	Attention is Not All You Need: For Powder Diffraction
9:45-10:05	Derek Ahneman	Building datasets and ML models to predict stepwise organic reactivity
10:05-10:25	Break	
10:25-10:45	Nina Andrejevic	Decoding Materials Dynamics with Physics-Aware Machine Learning
10:45-11:05	Wesley Reinhart	Towards Human-on-the-Loop Materials Discovery at the 2DCC and Beyond
11:05-11:25	Akshay Talekar	Data-Centric AI for Materials Discovery: Infrastructure, Integration, and Iteration at Scale
11:25-11:45	Fei Zhou	Scaling materials modeling beyond machine-learned potentials
Lunch: 12:00-1:00		
Session 4: 1:00-3:20		
1:00-1:20	Catherine Brinson	From Data to Discovery: accelerated materials data curation and interpretable AI for metamaterials
1:20-1:40	Jeffery Ting	ChainSpace: Searching Synthetically Accessible Sequence Space
1:40-2:00	Deyu Lu	Physics-Informed AI Pipeline for Real-Time Interpretation of X-ray Absorption Spectra
2:00-2:20	Break	
2:20-2:40	Joshua Schrier	Large language models for materials synthesis and extraction
2:40-3:00	Remi Dingreville	Shortening the learning curve: Schooling machine learning to power materials digital twins and predictions
3:00-3:20	Mitra Taheri	TBD
NCNR Tour: 3:30pm		

Organizing Committee

Daniel Wines, Debra Audus, Brian DeCost, Katelyn Jones, Wonseok Jeong, Austin McDannald, Kamal Choudhary (Johns Hopkins University), Francesca Tavazza

Website

https://www.nist.gov/news-events/events/2026/06/artificial-intelligence-materials-science-aims-2026

Questions or comments? Contact aims@nist.gov

Poster Information

Poster number	Name	Title
1	Abhishek Sose	To Be Determined
2	Alexander K Landauer	Uncertainty Estimate Mapping for Digital Image Correlation via Convolutional Neural Networks
3	Alexandru Bogdan Georgescu	Quantum Materials with Novel Quantum Building Blocks
4	Andy Shufer	Identifying the Mental Models of Scientists for the Use of AI in Accelerating Research
5	Bahador Bahmani	To Be Determined
6	Chowdhury Mohammad Abid Rahman	Supervised Pretraining for Material Property Prediction
7	Derek Juba	Uncertainty Quantification for Texture Directionality
8	Eva Natinsky	Machine-learning image reconstruction overcomes data scarcity in atomic force microscopy with domain specific image corruption
9	Feng Zhang	To Be Determined
10	Frank Abel	Developing Methods, Models, and Datasets for a Self-Driving Magnetic Nanomaterials Laboratory: Application for Thermal Magnetic Particle Imaging
11	Haochen Yang	Robotic Automation Discovery of Biodegradable Electronics via Multimodal Active Learning and AI-Guided Design
12	Jacob Horne	Predicting semicrystalline polymer properties with physics-informed machine learning models
13	Jaehyung Lee	SlakoNet DB: A Unified Tight-Binding Database for Electronic Structure and Bandgaps
14	John Head	Library of Geopolymer Chemistry & Applications
15	Joshua Young	Navigating High Entropy Alloy Compositional Space for Nitrate Reduction to Ammonia with a Universal Machine Learning Potential
16	Ju Sun	Accelerating Materials Discovery via Physics-Informed Constraints
17	Kai Wagoner-oshima	High-throughput search for topological materials for interconnects using first-principles transport calculations and machine learning
18	Katarina Goodge	Supporting rapid, reliable fiber identification
19	Katelyn Jones	Using NexusLIMs to Create Benchmark Datasets and Pretrained Microscopy Models
20	Lian Xiang	Machine learning-enabled multiplex biosensing using 2D materials
21	Ming-Chiang Chang	Evaluating Vision Transformer Architectures for High-Throughput X-Ray Diffraction Analysis
22	Mohamed Salem	MetPFN: The Metrological Prior Fitted Network
23	Noah Francis	A Two-Scale Finite Element Coupling Using a Machine Learning Accelerated Stochastic Micromechanics Solver for Thermal Conductivity
24	Qingjie Li	To Be Determined
25	Quinn Gallagher	Improving Machine Learning Extrapolation for Molecular Discovery
26	Ricardo Mathison Fuenmayor	A case study in the development of improved promoted Pt catalysts for propane dehydrogenation through Bayesian optimization with uncertainty quantification

Poster number	Name	Title
27	Rosa Diaz Rivas	Building Quantitative Descriptors for Heterogeneous Semiconductor Interfaces
28	Sarala Padi	STAMP: Species- and Topic-aware Representation Learning for Antimicrobial Peptide Discovery
29	Shuaijun Li	Predicting Properties from Near-Infrared Spectra with Machine Learning for Improved Polyolefin Differentiation
30	Snehi Shrestha	Machine Intelligence Accelerated Design of Conductive MXene Aerogels with Programmable Properties
31	Wonseok Jeong	Atomistic Dynamics as Sequential Decision-Making: Toward Experiment-Realistic Simulation
32	Yuhao Zhong	Uncertainty-Aware Explainable AI for Process-Structure-Property Discovery

Talk Abstracts

Physics based learning of foundation models of microscopic interactions

Boris Kozinsky, Harvard University

Discovery and understanding of next-generation materials requires a challenging combination of the high accuracy of first-principles calculations with the ability to reach large size and time scales. We pursue a multi-tier development strategy in which machine learning (ML) algorithms are combined with exact physical symmetries and constraints to significantly accelerate computations of electronic structure and atomistic dynamics. First, density functional theory (DFT) is the cornerstone of modern computational materials science, but its current approximations fall short of the required accuracy and efficiency for predictive calculations of defect properties, band gaps, stability and electrochemical potentials of materials for energy storage and conversion. To advance the capability of DFT we introduce non-local charge density descriptors that satisfy exact scaling constraints and learn exchange functionals called CIDER [1]. These models are orders of magnitude faster in self-consistent calculations for solids than hybrid functionals but similar in accuracy. On the atomistic level, we advance molecular dynamics simulations by developing machine learned interatomic potentials – NequIP and Allegro. Leveraging cutting edge kernel compilation techniques, we have increased the speed of these equivariant models by and order of magnitude compared to previous implementations and unlocked the possibility of training large foundation models in a matter of hours. We also introduced a way to train generalized potential and free energy functions with arbitrary nonlinear dependence on external fields and temperature. This framework enables learning and prediction of dielectric and vibrational response properties [2] and coarse-grained free energies [3]. We demonstrate these methods via first principles ML MD simulations of dynamics of phase transformations, heterogeneous reactions, ferroelectric transitions, nuclear quantum effects, and soft materials.

[1] K. Bystrom, B. Kozinsky, arXiv:2403.17002 (2024)
[2] S. Falletta et al, arXiv:2403.17207 (2024)
[3] B. Duschatko et al, arXiv:2405.19386 (2024)
[4] C.W. Tan et al, arXiv:2504.16068 (2025)

Accelerating Chemistry and Materials Innovation with NVIDIA ALCHEMI

Samarjeet Prasad, NVIDIA

Atomistic simulation is undergoing a rapid transformation as machine-learned interatomic potentials push the boundaries of accuracy and scale, but realizing their full impact in chemistry and materials discovery demands software built from the ground up for modern GPUs. In this talk, we introduce NVIDIA ALCHEMI, an open-source, two-layer stack designed to accelerate atomic simulation workflows end-to-end. At the foundation, nvalchemi-toolkit-ops delivers GPU-optimized, batched primitives for the core building blocks of atomistic simulation — neighbor enumeration, molecular dynamics integration, geometry optimization, and long-range physical interactions — with differentiable, stress-aware support for training the next generation of interatomic potentials. The kernels scale to systems of hundreds of thousands of atoms, and ship with JIT-compatible bindings for the major deep-learning frameworks. Built on top, nvalchemi-toolkit provides a GPU-first Python framework for assembling these primitives into production workflows: a standard interface for plugging in any interatomic potential, GPU-resident graph data structures with built-in serialization, composable dynamics that fuse stages on a single GPU or distribute them across many, an extensible hook system for observability and control, and inflight batching that keeps GPUs saturated across long-running pipelines. We will walk through the architecture, demonstrate the developer experience with short code examples, and discuss how researchers, model developers, and application engineers can plug into the stack today to accelerate chemistry and materials innovation on NVIDIA GPUs.

Bringing AI to f-element Chemistry

Michael Taylor, Los Alamos National Laboratory

Artificial intelligence is rapidly transforming chemical and materials discovery by accelerating the prediction, design, and optimization of molecules and materials across broad regions of chemical space. However, f-element chemistry, despite its central importance to energy, defense, and quantum information science remains comparatively underexplored by modern AI. In this talk, I will describe our efforts to bring AI to this challenging frontier through Architector, a software platform designed to generate, organize, and analyze chemically realistic three-dimensional structures for coordination complexes, including those containing f-elements. Architector addresses key barriers that have limited AI in this area, including sparse experimental data, uneven coverage of metal-ligand chemistry, structural diversity, and the difficulty of constructing plausible coordination geometries at scale. Further, I will highlight examples of how Architector-enabled data, AI, and machine learning models can already help with understanding and design of critical f-element chemistries.

Physics-aware electronic-structure models for reliable materials AI

Danish Khan, Caltech

Density functional theory (DFT) supplies much of the training data used by modern materials AI, so systematic errors in approximate functionals can propagate into downstream models. This talk presents a physics-aware strategy for using machine learning to improve electronic structure without replacing its theoretical foundation. First, I describe adaptive PBE0 (aPBE0), in which an ML model predicts a system-specific fraction of exact exchange. An uncertainty gate suppresses the learned correction outside its predictive domain, recovering the standard PBE0 functional as a controlled fallback. This single adaptive parameter improves molecular energies, electron densities, frontier orbital eigenvalues, and spin-state energetics, while allowing sparse high-level coupled-cluster data to be propagated into improved labels for larger molecular datasets. I then discuss a complementary constraint-guided approach based on the adiabatic connection. The nonlinear, nonempirical (nlane) functional combines exact exchange, perturbative correlation, and semilocal constraints without imposing a linear exchange-correlation mixture, improving the treatment of delocalization and static-correlation errors. Finally, I outline ongoing efforts to replace the expensive orbital-dependent exact-exchange and RPA-correlation terms in these methods with learned local density functionals and multiplicative potentials, targeting scalable self-consistent calculations for periodic materials. Together, these approaches show how machine learning can improve electronic-structure approximations, accelerate more accurate methods, and generate higher-quality data for downstream materials models.

Machine Learning Material Response to Electric Fields

Stefano Falletta, Harvard

Material response properties to external fields fundamentally dictate the performance of modern technologies involving dielectric, ferroelectric, and piezoelectric devices. However, predicting the dynamic response of realistic material systems to external stimuli with atomistic quantum accuracy remains a long-standing challenge in computational materials physics. To address this, I introduce a machine-learning framework in which response properties emerge from exact differential relationships between a generalized potential function and applied external fields. Focusing on electric fields, the method predicts response properties within a unified model that rigorously enforces exact physical constraints, symmetries, and conservation laws. I demonstrate that this physics-based approach accurately captures the vibrational, dielectric, and ferroelectric behavior of materials under arbitrary electric fields, including temperature and frequency dependence. This enables large-scale simulations of ferroelectric hysteresis and reveals the intrinsic nucleation and growth mechanisms governing ferroelectric domain switching. In addition, I present a closed-loop framework for dielectric materials discovery that combines forward prediction via direct models for band gaps and dielectric constants with inverse design via an encoder–decoder transformer.

Generative Models as Flexible Priors for Accelerating Long Timescale Molecular Simulation

Sanjeev Raja, UC Berkeley

Molecular simulation is a powerful tool for studying physical systems, but is hindered by two fundamental challenges: the difficulty of accessing long-timescales and the discrepancy between simulation and experiment. Generative models offer a promising path forward by learning expressive approximations of high-dimensional distributions from large datasets. In this talk, we present recent progress on how pre-trained generative models can accelerate molecular simulation by serving as adaptable priors for downstream scientific tasks. First, we show how score-based generative models can be repurposed for transition path sampling in a zero-shot manner. By interpreting candidate trajectories as stochastic dynamics induced by the force field of a learned generative model, the problem of identifying physically plausible transition pathways reduces to minimizing the Onsager–Machlup action functional. This enables the discovery of diverse, high-likelihood molecular transition paths without expensive task-specific training. Next, we address the simulation-to-experiment gap by developing a framework for aligning generative models trained on fully observed simulation data with partially observed experimental measurements. We provide theoretical guarantees for recovering target observable distributions and demonstrate alignment across a range of molecular and protein systems. Together, these results illustrate how generative models can function as flexible scientific priors that leverage large simulation datasets while remaining adaptable to new objectives, constraints, and observations.

Deploying Autonomous Experimentation in Government and Industry

Peter Beaucage, Lila Sciences, Inc

Soft materials—from pharmaceuticals to coatings—rely on complex formulations containing tens to hundreds of components. Traditional experimental approaches built on foundational structural characterization methods such as SAXS/SANS struggle to efficiently explore these vast parameter spaces, leading to long development cycles driven by performance heuristics and suboptimal outcomes. The Autonomous Formulation Laboratory (AFL) at NIST addresses this challenge by integrating automation, AI-driven decision-making, and high-throughput X-ray and neutron scattering to accelerate formulation discovery, optimization, and characterization.

This talk will provide an overview of AFL's open-source hardware and software ecosystem, including pipetting robotics, pneumatic sample transfer, in-line analytical techniques such as UV-Vis-NIR, SAXS/WAXS, and SANS, and AI-driven experimental design. Case studies will illustrate AFL's role in mapping phase boundaries in complex formulations and accelerating the synthesis of industrial nanomaterials, with a closer look at one bioformulation campaign in which AI-guided experimentation was deployed against a real industrial problem—optimizing protein stability across a high-dimensional excipient space with a 25 x reduction in experimental budget.

Finally, I will examine the broader impact of autonomous laboratories combined with the LLM-driven societal revolution on the pace of materials discovery, with specific focus on Lila Sciences, whose scientific superintelligence platform extends these principles into scaled, industrial approaches to materials AI. Lila’s platform pairs frontier models with robotic AI Science Factories to run every step of the scientific method at scale across domains in materials, chemistry, and life science.

Polymer Innovation Through AI-driven Autonomous Labs

Jie Xu, University of Chicago

Polymer science faces persistent challenges in navigating large molecular, formulation, and processing design spaces where material performance emerges from coupled chemistry–morphology–property relationships. In this talk, I will introduce Polybot, an AI-driven robotic laboratory that integrates automated synthesis, processing, characterization, and adaptive experimental design to accelerate polymer research. This talk will highlight research conducted in our self-driving lab, Polybot, covering topics from the inverse discovery of electrochromic polymer structures, the controlled assembly of conducting polymers through solution processing, and the discovery of design principles for mixed-conducting polymers in electrochemical transistors. We will also discuss ongoing efforts to evolve Polybot into a more adaptive system with enhanced human-machine interfaces and as a community resource.

Data Analytics in Research at BASF: Guiding Decision Making Through Efficient Experiments, Optimization, and ML & AI

Keith Task, BASF Corporation

Data-driven technologies are becoming ever more integral in industry to optimize systems and to guide decisions. The chemical industry is no different, as data science is being used in many sectors of the industry, from customer service to production. In this talk, we will discuss how BASF is utilizing digitalization in one key area: R&D. After introducing BASF, its global R&D landscape, and its digital focus, examples of technological developments in data analytics, and associated applications, will be highlighted. Recent advances in experimental design and developed software will be discussed, including nonlinear optimal design, large-language-model-assisted experiment planning, and Bayesian optimization. We will then highlight BASF's efforts in autonomous research, which connects these digital tools of machine learning, optimization, and experimental design to lab equipment and analytics. These integrated systems, having been demonstrated on numerous examples including photo and thermal reactions, allow for iterative closed-loop experimentation, leading to efficient and accelerated achievement of goals.

Unlocking Autonomy: Event Driven Linked Data for Scientific Workflows

David Elbert, Johns Hopkins University

A central gap facing laboratories of the future is a lack of data provenance to drive autonomous workflows. In the AI for Materials Discovery Laboratory (AIMD-L) we approach this challenge through novel use of typed events enriched with persistent identifiers, schema validation, uncertainty descriptions, and semantic context. AIMD-L's central, event-driven, linked-data layer converts data from a passive record into an active control surface to empower workflow choreography; on-demand materialization of state; and human or agentic reasoning over linked evidence. This creates opportunities for distributed laboratory coordination without brittle, centralized control while turning laboratory events into shared scientific intelligence infrastructure.

Addressing the Crystal Structure Bottleneck in Materials Models

Linda Hung, Toyota Research Institute

Many large-scale materials models rely on atomic-resolution crystal structure, which can be difficult or impossible to obtain for experimental samples. We present a multimodal model that instead requires only composition and XRD as input, and explore its use as a materials foundation model. We compare several self-supervised pretraining objectives, including masked XRD modeling (MXM) and masked element modeling (MEM), and find that MEM most effectively couples the two modalities. We benchmark the model on experimentally relevant tasks: transfer to small datasets, zero-shot prediction on experimental XRD, and prediction from incomplete composition — a setting inaccessible to structure-based models.

Integrating Physics, Chemistry, and Formulation Intuition into the Autonomous Formulation Lab

Tyler Martin, NIST

Biotherapeutic formulations present unique challenges for autonomous experimentation platforms. These complex colloidal systems require optimization across dozens of components with wildly varying purposes and design requirements, where small perturbations can cause large changes in stability and functionality. While machine learning-driven autonomous techniques have shown promise for materials discovery, purely data-driven approaches struggle with the highly constrained, nonlinear design spaces characteristic of biologic drug formulation. In this talk, I will describe how the Autonomous Formulation Laboratory (AFL) at NIST addresses these challenges by incorporating smart temperature sampling that accounts for asymmetric heating/cooling costs and knowledge of irreversible degradation boundaries, enabling joint optimization across compositional and environmental parameters. Further, I will highlight our efforts towards embedding physics-based priors from integral equation theory (RISM/PRISM) and molecular dynamics simulations to predict protein-excipient interactions, reducing reliance on costly physical measurements. Using a model bioformulation system, we experimentally demonstrate how this physics-informed approach achieves superior efficiency compared to traditional grid-based methods while balancing theoretical predictions with real-time stability measurements from multimodal characterization tools. This work illustrates how injection of domain expertise can dramatically improve the performance of autonomous systems on real-world formulation challenges.

Applications of materials informatics to Aerospace-Relevant Problems

Dung-Yi Wu, GE Aerospace

Small Smart Data Big Dumb Data and Models Don’t Extrapolate

Jason Hattrick-Simpers, University of Toronto

The dominance of artificial intelligence in materials discovery has reinforced a prevailing paradigm: ever larger datasets, ever larger models, and ever greater computational scale. This talk systematically interrogates that paradigm by examining how model complexity, dataset size, and dataset construction influence predictive performance, using DFT datasets as a case study. We show that smaller, carefully designed datasets can support strongly predictive models, while out-of-distribution prediction remains a fundamental challenge when training and test data do not share the same distribution. We further demonstrate that even carefully calibrated uncertainty estimates often fail to provide meaningful guidance in these out-of-distribution settings. Finally, we present a training-free active learning framework that uses large language models as surrogates and achieves state-of-the-art performance.

Building datasets and ML models to predict stepwise organic reactivity

Derek Ahneman, IBM

Mechanistic understanding is essential for reaction development and optimization. Reaction networks (RNs) encode intermediates as nodes and elementary steps as edges, but their construction faces challenges: physics-based simulations are computationally expensive, while machine learning approaches lack comprehensive elementary step datasets. Machine learning could accelerate RN construction and reveal promising reaction space, but the field lacks open, diverse datasets. Here, we present efforts toward building an elementary step dataset spanning diverse organic reactions. Predicting reaction outcomes requires discriminating between competing pathways, so we describe a workflow for computing barrier heights and reaction energies, emphasizing robustness across reaction types for solution-phase main group reactivity. Finally, we discuss progress toward training relevant ML models.

Attention is Not All You Need: For Powder Diffraction

William Ratcliff, NIST

Determining crystal symmetry from powder diffraction is a central problem in materials characterization, yet multiple space groups can produce indistinguishable patterns. We show that attention-based architectures, while superior to convolutional networks for this task, are insufficient on their own: reliable symmetry extraction requires encoding crystallographic knowledge into both the network architecture and the training curriculum. We introduce a physics-informed transformer that classifies powder patterns into 99 extinction groups — the information-theoretically natural targets for diffraction — using an explicit $\sin^2\theta$ coordinate channel, physics-aware positional encoding, and a structured multi-task decoder that separates geometric rule learning from holistic pattern recognition. A three-stage curriculum of balanced synthetic pretraining, realistic fine-tuning, and Bayesian prior injection at inference time proves essential for bridging the synthetic-to-real domain gap. We further show that post-hoc calibration, rather than additional training, is the key remaining ingredient for robust real-data transfer. We identify a “catastrophic paradox” in which patterns with the worst Rietveld fits are among the easiest to classify, explained by lower label-space entropy rather than data quality. These results demonstrate that physics-informed architecture, data design, and calibrated inference matter as much as model capacity for scientific machine learning on diffraction data.

Decoding Materials Dynamics with Physics-Aware Machine Learning

Nina Andrejevic, Argonne National Laboratory

Recent advances in materials characterization techniques and instrumentation are enabling the collection of increasingly sophisticated time-resolved datasets, creating the possibility for richer understanding of dynamic phenomena across multiple scales. However, traditional analysis pipelines that require computationally expensive reconstruction or inversion from measurements remain a bottleneck for extracting mechanistic insights from these data. Here, I discuss projects in which we approach this challenge by integrating machine learning with differentiable X-ray scattering forward models, enabling direct learning of governing equations and dynamic behavior in the observation space without intermediate reconstruction steps. First, I will present our work using neural differential equations combined with coherent X-ray scattering forward models to recover dynamic phenomena such as synchronization and self-organization and to extrapolate well beyond measurement windows. I will then discuss how symbolic regression can be applied to these trained neural differential equation models to distill compact, interpretable mathematical expressions that expose the underlying physical mechanisms governing the observed dynamics, bridging the gap between data-driven flexibility and the mechanistic clarity of first-principles models.

Towards Human-on-the-Loop Materials Discovery at the 2DCC and Beyond

Wesley Reinhart, Pennsylvania State University

The 2D Crystal Consortium Materials Innovation Platform (2DCC-MIP) has spent a decade building a comprehensive database of 2D materials synthesis and characterization. Spanning nearly 20,000 samples with dozens of characterization modalities from real experimental campaigns, the Lifetime Sample Tracking (LiST) database is now a one-of-a-kind informatics resource. Despite this volume, these data are challenging to model with data-driven surrogates; these data were not derived from high-throughput screening or a self-driving lab but from manual, expert-driven discovery campaigns. The result is a dataset where nearly every sample is novel: unique process conditions, instruments, morphologies, heterostructures, and phases. These properties make LiST a poor fit for off-the-shelf ML methods, but an exceptional stress test for AI systems aimed at solving real materials discovery problems in real laboratories.

In this talk, I present case studies along our path from expert-driven to human-on-the-loop, AI-driven discovery. First, we used Bayesian optimization within the 2DCC to discover phase-pure γ-In2Se3 Molecular Beam Epitaxy synthesis conditions that directly contradicted expert assumptions, demonstrating the use of data-driven models to overcome bias. In the same system, Large Language Models (LLMs) were used to perform meta-optimization, designing more robust surrogate models without hand-tuned hyperparameters. Second, to study the growth of SnSe thin films, we used counterfactual ExAI to describe film morphology in the original image modality that experimental synthesis scientists could evaluate mechanistically. To assess the generality of these ideas, I finally turn to molecular design challenges outside the 2DCC: small molecules and sequence-defined copolymers, which are discrete optimization problems in large action spaces. There, we used LLM-based agents to directly perform evolutionary optimization. We showed how these agents can perform search more efficiently than even model-driven active learning, and how the behavior of pre-trained models can be modified at inference time through instruction. In each case study, I consider both model performance and utility in closed-loop scientific discovery – aspects which appear increasingly orthogonal.

Data-Centric AI for Materials Discovery: Infrastructure, Integration, and Iteration at Scale

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. Progress is often 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.

Scaling materials modeling beyond machine-learned potentials

Fei Zhou, LLNL

MLIP has been the one of the most successful applications of AI/ML for materials science. To increase the spatial and temporal scales of computational materials science even further, one may look at a few different directions. This talk will present our latest works in faster data- and physics-driven time-integration in MD, unified foundation models of crystal structure classification and order parameter extraction, and mesoscale free-energy models that bridge MD simulation and surrogate dynamical models.

From Data to Discovery: accelerated materials data curation and interpretable AI for metamaterials

Catherine Brinson, Duke University

The materials genome initiative (MGI) sparked a revolution in how materials research and design is done, built on materials data and deploying new information science tools. This lecture will present some recent methods for accelerated data curation and novel AI approaches for understanding and designing complex hierarchical architected materials. First, we will discuss some of the challenges and promises associated with data curation, ontology and vocabulary development, standardization and interoperability, and data visualization and analysis tools. In this context, we will present recent work on use of LLM pipelines for rapid data curation from published literature and reports. This work systematically compares four extraction strategies of increasing complexity, from single-pass extraction to multi-step extraction with chain-of-verification. We also introduce a human-annotated benchmark and an evaluation framework. We will highlight the methods, successes and challenges remaining in turning pdfs into JSONs. Then, we will present several examples of an integrated AI-driven framework for metamaterials that leverages interpretable machine learning, hybrid optimization, and neural operator-based simulation. We introduce interpretable learning approaches that extract rule-based relationships between unit-cell geometry and target wave properties, enabling the discovery of metamaterial architectures with prescribed band gaps while maintaining design flexibility and generalizability across resolutions. We then present a hybrid computational design framework for three-dimensional acoustic metamaterials with stiffness gradients for impedance matching. This approach combines data-free evolutionary optimization with differentiable machine learning models, allowing efficient global exploration alongside high-fidelity local refinement under manufacturing constraints. In another example, we employ Fourier Neural Operators with wavelet-based encodings to directly solve eigenvalue partial differential equations governing wave propagation in complex metamaterial geometries, enabling rapid iterative design of material geometries to meet target responses. We will highlight challenges in application of AI specific to materials problems, the role of interpretability, and the value of data.

ChainSpace: Searching Synthetically Accessible Sequence Space

Jeffrey Ting, Nanite

When equipped with proper synthetic tools, polymer scientists can combinatorically explore important properties that direct function, such as molar mass, dispersity, monomeric selection/composition, or architecture. This combinatorial design space is immense, and therefore machine learning and AI can help tackle this search. By computationally searching such design spaces, the time and cost of developing new materials can be drastically reduced in principle. However, there is no comprehensive method for considering full statistical sequence distributions in industrially relevant polymers. The majority of works that focus on sequence effects frequently rely on manually defined patterns or canonical arrangements, such as random, alternating, or blocky representations; this approach neglects rich molecular details of multimonomeric systems used in industrial adhesives, coatings, personal care items, and drug/gene delivery excipients. In this work, we present a practical computational framework that describes synthetically accessible sequence space using 3 orthogonal descriptors across multiple length scales: the second eigenvalue (local scale correlations from Markov process modeling), Shannon entropy (mesoscale measure of pattern diversity), and mean run length (global scale blockiness tendency). We characterized sequence statistics across experimentally achievable syntheses in chain polymerization and post-polymerization modification. Parameter space exploration with our unified three-descriptor framework resulted in a curved 3D manifold that govern polymer sequences as a function of kinetic, temporal, or spatial control. By capturing this statistical plasticity, we demonstrate the versatility of using these levers for sequence analysis with proof-of-concepts that enable more exhaustive active learning campaigns to algorithmically navigate sequence space. This sequence first approach can accelerate the discovery and development of tailored polymers with highly unique properties over their random or block counterparts.

Physics-Informed AI Pipeline for Real-Time Interpretation of X-ray Absorption Spectra

Deyu Lu, Brookhaven National Laboratory

Fast and accurate interpretation of complex scientific measurement data, such as spectral functions, is a major bottleneck in scientific discovery that heavily relies on domain expertise and consumes substantial human time and computational resources. By integrating physics-informed models with artificial intelligence (AI) and machine learning (ML), we can transform scientific data analysis and accelerate the research cycle. Here, I demonstrate this paradigm through an AI-driven pipeline for X-ray absorption spectroscopy (XAS). XAS is a premier element-specific experimental technique for materials characterization and mechanistic understanding. Specifically, X-ray absorption near edge structure (XANES) encodes rich local structural and chemical information around absorber sites, making it a powerful tool for probing physical and chemical processes in broad domain fields. However, the correlation between XANES spectral features and the underlying local structural motifs or electronic descriptors is obscure. Recent advances in materials discovery driven by smart automation and in situ / operando experiments further underscore emerging challenges and opportunities of spectral analysis in real time. I discuss recent progress of the AI-driven spectral analysis pipeline that integrates first-principles theory, software development, high-throughput simulation, and data analytics tools to decipher the structure-spectrum relationship. The impact of this approach is demonstrated by its ability to extract physical insights from complex materials while enabling robust, real-time analysis.

Large language models for materials synthesis and extraction

Joshua Schrier, Fordham University

This talk will present several case studies of using and adapting pre-trained large language models for materials research tasks. First, I will describe how fine-tuning— perturbing the neural network weights of a pretrained model to match a dataset—can be used to predict the synthesizability (can it be made?) and select precursors (how to make it?) for solid state inorganic synthesis, both at the level of compounds and specific structural polymorphs. Second, I will show ways to use an LLM-based workflow to generate human-readable explanations of why certain compounds are synthesizable, by automating the extraction of underlying rules and assessment of the veracity and internal consistency of those rules. Third, I will show how an LLM-based agentic workflow can use software “tools” that include machine learning models and experimental databases to discover novel organic molecules for critical minerals extraction. These patterns are applicable to a wide variety of applications in materials science.

Shortening the learning curve: Schooling machine learning to power materials digital twins and predictions

Remi Dingreville, Sandia National Laboratories

Advanced computing is transforming how complex materials are designed, qualified, and deployed by enabling detailed predictions of their aging and performance under realistic operating conditions. Yet the underlying process–structure–property–aging relationships are high-dimensional, nonlinear, and increasingly beyond the reach of traditional brute-force simulations. In this talk, I will discuss how machine learning (ML) strategies confront this complexity by “schooling” algorithms with materials science—encoding physics and domain expertise directly into their architectures and training data. Through case studies ranging from thermomechanical simulations of composites and alloys, to simulations of manufacturing processes, I will show how these “educated” models speed up simulation predictions, enable performance-by-design, and form the computational foundation for digital materials twins.

To Be Determined

Mitra Taheri, Johns Hopkins University

Poster Abstracts

1 To Be Determined

Abhishek Sose, University of Maryland College Park

2 Uncertainty Estimate Mapping for Digital Image Correlation via Convolutional Neural Networks

Alexander K Landauer, NIST

Digital image correlation (DIC) is a technique for measuring displacement and strain fields on a test artifact surface throughout the field of view of an image, often for material characterization tasks. Here, we present a machine learning-based technique to provide a lower-bound estimate of the pointwise uncertainty for 2D-DIC displacement fields. This is of particular interest when using full-field data for inverse identification, as high-uncertainty regions may need to be handled differently than the bulk, or in situations where highly localized displacements are the measurand of interest, such as in fracture or impact. In this new workflow, a test image is manipulated using synthetic image warping with a range of analytical or user-defined displacements and noise profiles to generate a library of supervised training images for DIC error. A convolutional neural network (CNN) is trained on these synthetic, multi-channel RBGA-like 8-bit images that map displacements and errors at each pixel location. The model is a multi-input single-output framework where several input maps (image intensity and displacements) are used to predict a single output (uncertainty) map. This poster will highlight the framework implementation and training results. The complete toolchain is intended to be released as free and open-source on the usnistgov GitHub.

3 Quantum Materials with Novel Quantum Building Blocks

Alexandru Bogdan Georgescu, Indiana University

Correlated electron materials with electronic building blocks (e.g. spin, orbital) in the form of molecular orbitals spread on transition metal clusters - as opposed to d and f orbitals - form a novel materials family with a wide variety of possible applications. Nb3Br8 has recently attracted attention due to its magnetic state spread across Nb3 trimers - presenting the possibility of a spin liquid state - and its coupling of magnetic and ferroelectric degrees of freedom has led to the first field-free Josephson diode. Using novel graph representations, our group has recently built a database of 3000+ materials that may display such electronic states (https://cluster-explorer.me/), have identified symmetry based criteria that allow trimers to form in kagome materials, and - using electron localization function and electron basin analysis - discovered that the bonding patterns underlying this class of materials' electronic structure is different from that of most correlated electron materials. Additionally, a subset of these materials can be electrochemically doped dynamically - similar to battery materials - leading to unique electronic states.

4 Identifying the Mental Models of Scientists for the Use of AI in Accelerating Research

Andy Shufer, Carnegie Mellon University

Despite growing public and private investment in the use of AI for accelerating scientific research, scientists remain skeptical or unwilling to adopt the technology. Fundamental questions about AI's impact and limitations, its effect on research direction, and what barriers constrain its ability to have greater value limit policymaking aimed at incentivizing AI for science adoption. Characterizing the mental models of how scientists evaluate and adopt AI requires understanding their individual, institutional, and research context and background. We employ a qualitative mixed-methods approach: semi-structured interviews with US-based principal investigators in Biology and Material Science, followed by a national survey of scientists across all career levels and disciplines. Using individual research applications as an analytical framework, we examine variation in perceptions of AI by field, research question type, and career stage, as well as the structural and cultural barriers that constrain uptake. Findings can inform where policy interventions are most likely to succeed in aligning AI investment with how scientists actually evaluate and adopt new innovation methods.

5 To Be Determined

Bahador Bahmani, Northwestern University

6 Supervised Pretraining for Material Property Prediction

Chowdhury Mohammad Abid Rahman, West Virginia University

Accurate prediction of material properties facilitates the discovery of novel materials with tailored functionalities. Deep learning models have recently shown superior accuracy and flexibility in capturing structure-property relationships. However, these models often rely on supervised learning, which requires large, well-annotated datasets—an expensive and time-consuming process. Self-supervised learning (SSL) offers a promising alternative by pretraining on large, unlabeled datasets to develop foundation models that can be fine-tuned for material property prediction. In this work, we propose supervised pretraining, where available class information serves as surrogate labels to guide learning, even when downstream tasks involve unrelated material properties. We evaluate this strategy on two state-of-the-art SSL models and introduce a novel framework for supervised pretraining. To further enhance representation learning, we propose a graph-based augmentation technique that injects noise to improve robustness without structurally deforming material graphs. The resulting foundation models are fine-tuned for six challenging material property predictions, achieving significant performance gains over baselines, ranging from 2% to 6.67% improvement in mean absolute error (MAE)—and establishing a new benchmark in material property prediction. This study represents the first exploration of supervised pertaining with surrogate labels in material property prediction, advancing methodology and application in the field.

7 Uncertainty Quantification for Texture Directionality

Derek Juba, NIST

This project involves assigning probabilities to classes predicted by a Neural Network. Our group has developed a classifier Neural Network to determine the directionality of the texture in an image. The output of this network is a vector of 180 logit values, each corresponding to a possible direction. Higher logit values are thought to represent more likely directions, and the direction corresponding to the max logit value is taken as the network's classification; however, the logit values do not directly represent probabilities. In this work, we develop a calibration curve to convert the logit values into actual probabilities.

8 Machine-learning image reconstruction overcomes data scarcity in atomic force microscopy with domain specific image corruption

Eva Natinsky, Sandia National Laboratories

The need for accelerated nanometrology becomes more pressing with the rise of advanced manufacturing technologies, and bottlenecks exist in both data acquisition and the subsequent processing. While various instrumentation approaches to this issue exist, this work utilizes machine learning (ML)-driven image reconstruction to accelerate data processing and eliminate imaging artifacts. Current ML methodologies for nanometrology frequently encounter the challenge of training a data-hungry ML model in a data-scarce domain. Here, a novel training workflow called AFM-net is presented for an image reconstruction ML model, applied to atomic force microscopy (AFM), that enables data augmentation far exceeding experimental capabilities. In this workflow, undesirable corruptions are extracted from experiments and transplanted onto natural images, creating a large-volume, domain-specific training database comprising more than two million unique examples. The results show AFM-net produces high quality reconstructions and achieves accurate height predictions from AFM measurements. Critically, AFM-net accelerates data processing by 75x to 500x over traditional methods. Furthermore, it enables shortened acquisition times by producing high-quality reconstructions from faster scan speeds, with measurable quality tradeoffs that would still be acceptable for many applications. The modularity of AFM-net also makes it adaptable to new data types, and even to new domains with minimal tuning. SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525.

9 To Be Determined

Feng Zhang, Ames National Laboratory

10 Developing Methods, Models, and Datasets for a Self-Driving Magnetic Nanomaterials Laboratory: Application for Thermal Magnetic Particle Imaging

Frank Abel, NIST

Self-driving laboratories (SDLs) are rapidly emerging as a powerful approach for discovering new materials, mapping complex material synthesis spaces, and improving experimental reproducibility. Realizing SDL workflows requires automating synthesis processes and characterization by modifying protocols to support robotic execution. In addition, it is necessary to develop robust, scriptable, and machine-learning-based methods for extracting quantitative information from characterization datasets and for linking it back to synthesis conditions. Despite significant progress in autonomous materials and nanoparticle synthesis platforms, such as WANDA [1] and A-Lab [2], comparatively little work has focused on SDL-compatible workflows for magnetic nanoparticles, where synthesis outcomes for functional magnetic properties must be linked to size, phase, defects, and morphology.

Nearly single-phase, single-crystal ferrite nanoparticles are critical across several application areas, from biomedical imaging and sensing to newly emerging thermal-imaging applications in microelectronics packaging based on magnetic particle imaging (MPI). However, a significant limitation for both biological and thermal applications of MPI is the signal amplitude, which relies on the nanoparticle tracers. Recently, it has been demonstrated that highly phase-pure, single-crystal mostly magnetite nanoparticles with a critical size between approximately 15 nm and 25 nm can achieve super-resolution and > 10x improvement in MPI signals [3,4]. However, the reproducibility of these particles is poor, with only two methods reported in the literature demonstrating these effects, both of which suffer from batch-to-batch particle size control.

In this work, we have developed a new combinatorial, automation-friendly synthesis method for highly phase-pure ferrite nanoparticles. Along with the development of a machine learning model for predicting the magnetization of iron oxide nanoparticles from X-ray diffraction (a high-throughput method) [5], a scriptable Python-based XRD refinement tool, and a phase impurity detection algorithm, our approach enables integration of synthesis-structure-property datasets into a flexible format for future model development. When coupled, these tools can accurately predict if a given sample will exhibit a super-resolution MPI signal, which we validate through direct magnetic particle spectroscopy (MPS) measurements and preliminary solid-state MPI images in semiconductor-relevant imaging phantoms.

11 Robotic Automation Discovery of Biodegradable Electronics via Multimodal Active Learning and AI-Guided Design

Haochen Yang, University of Maryland College Park

Designing biodegradable electronic substrates that possess both mechanical robustness and tunable dielectric properties is a complex challenge, characterized by high-dimensional formulation spaces and nonlinear coupling between material properties. To accelerate discovery, we introduce a multimodal dynamic active learning (AL) framework integrated with multiple robotic automation platforms to navigate a vast design space with a 19-component library comprising biopolymers, layered clays, and metal oxides. Unlike static acquisition strategies, this AL framework employs a stage-wise approach that dynamically adapts sampling objectives, shifting from diversity-driven exploration to uncertainty-aware exploitation, to navigate the design space efficiently. This predictive modeling engine is tightly coupled with a fully automated experimental pipeline, featuring robotic liquid handling, automated tensile testing, and dielectric characterization, enabling closed-loop experimentation. Over eight AL rounds, we developed a prediction model leveraging an ensemble of hierarchical multimodal artificial neural networks for simultaneous feasibility screening and property prediction. Benchmarking demonstrates that this dynamic approach yields superior learning efficiency and target discovery compared to baselines under limited experimental budgets. We also successfully identified samples with mechanical robustness and an application-relevant range of dielectric properties demonstrating the practical utility of the closed-loop robotics--ML workflow under realistic experimental constraints.

12 Predicting semicrystalline polymer properties with physics-informed machine learning models

Jacob Horne, NIST

The application of machine learning (ML) to polymer design has advanced significantly in recent years. However, these methods typically rely on large datasets, which are often unavailable. To address this data scarcity, we develop physics-informed ML models capable of predicting properties of semicrystalline polymers from small datasets. These physics-informed models enable better model performance with limited training data and provide a direct path to obtain physical insight from model predictions. In addition, our approach accounts for the dispersity inherent in synthetic polymer samples by parameterizing theoretical models in terms of probability distributions describing key features of molecular architecture such as the molar mass and the spacing between comonomers. To validate this methodology, we use in silico data to train ML models to predict the melting temperature of semicrystalline polymers with varying concentrations of non-crystallizable comonomers. We compare the predictions of both physics-informed and uninformed models to demonstrate the gain in model performance and interpretability afforded by the physics-based models.

13 SlaKoNet DB: A Unified Tight-Binding Database of Electronic Structure

Jaehyung Lee, Johns Hopkins University

Recent advances in machine learning for materials science have led to the development of foundational models that achieve strong performance on classical atomistic properties such as formation energy, stability, and structural relaxation. However, extending these approaches to electronic quantities such as band structures and bandgaps remains challenging. Unlike classical observables, electronic properties are not solely determined by local structural descriptors and instead require explicit information about electronic orbitals and their interactions. To address this limitation, we introduce SlakoNet DB, a large-scale electronic structure framework built upon a unified Slater-Koster tight binding representation that explicitly incorporates orbital-resolved electronic degrees of freedom. SlakoNet enables scalable and transferable modeling of electronic properties across diverse materials classes, bridging the gap between machine learned interatomic potentials and quantum electronic structure methods. Using this framework, we construct a comprehensive database spanning multiple materials domains, including bulk hull structures, molecules, interfaces, surfaces, vacancy structures, superconducting candidates, and low dimensional 1D and 2D materials. The resulting dataset includes a broad range of pristine materials and defect containing configurations, all computed within a consistent tight binding formalism. This unified dataset enables high throughput prediction of bandgaps and related electronic properties across both ordered and defective systems, establishing a foundation for electronic structure aware machine learning. The SlakoNet model and database are available for real time inference and exploration at https://atomgpt.org/slakonet, providing an accessible platform for data driven electronic materials discovery.

14 Library of Geopolymer Chemistry & Applications

John Head, 5854 Limited

1. Geopolymer Chemistry and Applications Geopolymer chemistry generally refers to an emerging science surrounding a substance made from materials rich in aluminum oxide and silicon dioxide, that, when combined in an alkaline liquid, form a paste of long-chain molecules. These long-chain molecules, essentially repeating radicals of aluminum, silicon and oxygen atoms, harden into a molecular backbone that serve as a binder for a variety of composites. For example, when mixed with fillers such as sand and aggregate, it was poured to make concrete runways and the air terminal in Brisbane, Australia. With certain fillers and additives, it can be used with a 3D printer to print the walls of a residential structure, or made into a lining for water conduits, used to encase nuclear waste, and a myriad of other applications. Geopolymer composites with fibers of basalt, steel, fiberglass, or carbon have physical strength equal or superior to high performance concrete. Geopolymer binders have superior acid resistance when compared with conventional concrete, have better microbial compatibility, and are suitable for introducing electrically conductive additives to improve electron and ion transfer properties. The feedstocks that supply the essential ionic components include kaolin, red clay, granulated blast furnace slag, flume from coal fired power plants, and sodium silicate as well as other materials which have the building blocks of aluminum oxide and silicon dioxide. The science is attracting much attention for obvious reasons. It is friendlier to the environment than Portland cement. (The manufacture and use of Portland cement accounts for some 7% of the world’s annual carbon dioxide emissions. In stark contrast, geopolymer concrete produces very little.) It has performance characteristics superior to ordinary Portland concrete. Also, because it does not burn, it is an attractive alternative to organic polymers. 2. The Problem 2.1 The volume of written information is staggering. Google Scholar counts over 160,000 papers on geopolymers. About 9,500 papers are published on the subject every year which averages 26 each day. 2.2 Academic research papers are scattered among a plethora of academic journals owned by dozen or so publishers. 2.3 Sorting relevant, current, and authoritative articles requires time 2.4 Expensive, with 60 – 80% of academic research papers are behind paywalls of for-profit publishers. The annual subscriptions at one major publisher range from $1,300 to over $4,000. Individual papers range in price of between $30 to $55. 2.5 Time consuming to digest 3. The Solution 3.1 Build a digital library stored in the cloud with open access that contains everything presently known about geopolymer chemistry and applications of the technology. 3.2 Build a machine learning artificial intelligence tool to aid users in extracting knowledge from the library 4. The Library 4.1 When fully operational and curated daily, the Library will include approximately 30,000 – 35,000 documents to include: 4.1.1 Academic papers published during the past 20 years reflecting original research 4.1.2 Academic papers published during the past five years reflecting reviews of the state of the technology 4.1.3 Books 4.1.4 Patents

15 Navigating High Entropy Alloy Compositional Space for Nitrate Reduction to Ammonia with a Universal Machine Learning Potential

Joshua Young, Matlantis Inc.

The electrocatalytic nitrate reduction reaction (NO3RR) offers a potentially more facile and sustainable alternative to the production of ammonia than the traditional energy-intensive Haber-Bosch process used industrially. However, efficient catalysts for this reaction are still an open area of investigation. In this work, we perform a large-scale screening of high entropy alloy catalysts using our universal machine learning interatomic potential (PreFerred Potential, PFP) simulations. We investigate 5 HEA compositions using a high-throughput statistical approach to uncover the spectrum of active-site environments controlling nitrate and ammonia adsorption. We then perform Monte Carlo simulations to model realistic surface segregation, determining the probability of high-performance sites appearing. Finally, we extract latent space embeddings from the last layer of the MLIP's graph neural network architecture and use them as features to train supervised machine learning models; from there, we identify high-performance chemical motifs and propose optimized, non-equimolar HEA compositions. Overall, this PFP-enabled framework can dramatically speed up the discovery of complex catalysts for challenging reactions such as the NO3RR.

16 Accelerating Materials Discovery via Physics-Informed Constraints

Ju Sun, University of Minnesota

I'll describe our ongoing efforts to build principled computational tools to perform property prediction and inverse design with hard physics-informed constraints.

17 High-throughput search for topological materials for interconnects using first-principles transport calculations and machine learning

Kai Wagoner-oshima, Rensselaer Polytechnic Institute

The performance of semiconductor-based computing technologies is increasingly hindered by the resistivities of back-end-of-line (BEOL) interconnect materials as dimensions continue to scale down. Relative surface imperfections become more intense in narrow wires, and common interconnect materials such as copper exhibit increased resistivities at the nanoscale because of their sensitivity to surface scattering. Topological materials may address this critical BEOL interconnect resistance bottleneck through their surface states that are topologically protected against scattering from surfaces and other defects. Yet the large chemical and structural design space makes the discovery of viable candidates challenging. Here, we combine first-principles electron transport simulations with machine learning to accelerate the search for topological materials relevant to BEOL interconnect applications. Using calculated bulk and surface conductance data, we develop machine learning models to learn structure–property relationships governing transport behavior. We further integrate these models into an active learning workflow driven by Bayesian search, which balances predicted performance and model uncertainty to identify the most informative materials for follow-up simulation. By iteratively expanding the training set in this way, our framework aims to more efficiently guide the discovery of promising topological interconnect candidates for nanoscale devices.

18 Supporting rapid, reliable fiber identification

Katarina Goodge, NIST

Material recovery systems in the textile industry face significant challenges in rapid and reliable feedstock identification. Textile sorters are working to integrate Near Infrared (NIR) spectroscopy into their sorting process to increase their throughput, accuracy, and cost competitiveness. However, the lack of well-characterized, known-provenance materials limits the validation of complex sorting algorithms and hinders the development and deployment of efficient sorting technologies.

To bridge this gap, we have released our curated, machine-actionable Near-Infrared Spectra of Origin-defined and Real-world Textiles (NIR-SORT) dataset, comprising NIR spectral fingerprints of over 100 diverse fabric types. NIR-SORT provides a critical resource for developers of AI-based sorting systems to test and validate their algorithms, enhancing the performance and reliability of their technologies. As a demonstration of NIR-SORT utility and to further understand the material identification challenges that textile sorters are experiencing, we have built ML test models to probe underlying mechanisms for fiber misrecognition such as fabric additives and spectral quality.

19 Using NexusLIMs to Create Benchmark Datasets and Pretrained Microscopy Models

Katelyn Jones, NIST

This project focuses on the creation of a benchmark dataset and guidelines for building future material science datasets for convolutional neural networks (CNNs) and image-focused ML models, using data labeled by NexusLIMS. NexusLIMS is a laboratory information management system (LIMS) developed by NIST that automatically collects data and metadata from 10+ electron microscopes across OUs and campuses. This resulted in collecting thousands of electron micrographs, spectra, and diffraction data, creating a multi-material and multimodal dataset. A selection of image data from NexusLIMS has been manually labeled to develop classification models that identify the type of material and microscope used to capture the image. This demo will show a selection of images from the NexusLIMS page and the results of them being labeled by the trained models. The goal of this project is to enable more rapid and accurate development of electron microscopy-based model via the sharing of pre-trained vision-based models.

20 Machine learning-enabled multiplex biosensing using 2D materials

Lian Xiang, Rensselaer Polytechnic Institute

Multiplexed electrochemical detection of biomolecules holds great promise for transforming disease diagnostics, wearable health monitoring, and environmental sensing. Monoamine neurotransmitters like dopamine, serotonin, epinephrine and norepinephrine are of particular interest. However, the multiplexed detection of monoamine neurotransmitters is challenging due to the insufficient selectivity and sensitivity of existing sensors, as well as interference between analytes and other molecules in real biofluid samples. This work overcomes these challenges by creating a machine learning (ML)-enabled biosensing platform using 2D materials to perform the multiplexed electrochemical detection of serotonin, dopamine, epinephrine, and norepinephrin. Our study uses laser-induced graphene (LIG) as the sensor, with human urine serving as the target sample to be measured. Electrochemical fingerprints extracted from square wave voltammetry and differential pulse voltammetry serve as features for ML models. This work introduces a framework that transforms electrochemical biosensing from single-analyte measurement into multiplexed biomarker profiling. Utilizing our setup with LIG, we achieve an accuracy of 97% in analyte identification tasks for the monoamine neurotransmitters. This study paves the way for the reliable simultaneous classification and quantification of bioanalytes across the chemically heterogeneous and complex biofluids of real humans. Consequently, it lays the groundwork for next-generation electrochemical diagnostics.

21 Evaluating Vision Transformer Architectures for High-Throughput X-Ray Diffraction Analysis

Ming-Chiang Chang, NIST

X-ray diffraction (XRD) is essential for materials characterization, but rapid and reliable interpretation remains a bottleneck for high-throughput and autonomous experimentation. Deep learning can accelerate XRD analysis, yet many existing approaches rely on convolutional neural networks, whose translation-invariant inductive bias is not fully aligned with diffraction data, where peak positions carry direct crystallographic meaning.

Here, we investigate Vision Transformer-inspired architectures for one-dimensional XRD analysis. We ask whether vanilla ViT implementation is sufficient for diffraction patterns, or whether choices such as positional encoding and token fusion must be adapted to the physics and semantics of XRD data. Using the MP-20 dataset, we benchmark models on crystal system, extinction group, and space group prediction, systematically comparing positional embedding strategies and fusion mechanisms, including alternatives such as spatial mixing. We further evaluate the models on real experimental time-series XRD data.

This work identifies architectural design choices that improve transformer-based XRD interpretation and supports the development of fast, scalable analysis tools for AI-guided materials discovery and self-driving laboratories.

22 MetPFN: The Metrological Prior Fitted Network

Mohamed Salem, NIST

Modern measurement science depends on large, heterogeneous datasets generated across instruments, laboratories, simulations, and experimental workflows. Yet valuable scientific knowledge often remains dispersed across calibration procedures and individual researchers, making it difficult to reuse systematically in new metrological problems. We propose MetPFN, a metrological Prior Fitted Network framework that learns from synthetic data-generating processes designed to encode expert knowledge, physical constraints, experimental uncertainty, and partial causal structure.

MetPFN aims to transform metrological inference from a workflow of isolated, single-use estimators into a reusable foundation model for measurement science. By combining synthetic priors, historical data, physical models, and researcher input, the proposed approach supports fast, uncertainty-aware estimation and distributional prediction across related measurement tasks. This offers a path toward centralized, machine-readable metrological knowledge that can be shared, refined, and reused across laboratories and measurement domains.

23 A Two-Scale Finite Element Coupling Using a Machine Learning Accelerated Stochastic Micromechanics Solver for Thermal Conductivity

Noah Francis, Sandia National Laboratories

Classical deterministic micromechanics numerical solvers which homogenize (or course-grain) spatially distributed local properties into effective properties always require the computation of the local fields. Examples include the Finite Element (FE) Method which calculates the local fields at each node, and the Fast Fourier Transform (FFT)-based method which calculates them at each pixel. This fact is natural, but becomes a computational hindrance when scaling to problems of practical interest, such as in the two-scale FE^2 (or the FE-FFT) approach where there is a macroscopic FE simulation that has a micromechanical FE (or FFT-based) solver at every material point. Inspired by the recent rise of machine learning based acceleration techniques, a framework called the Deep Material Network (DMN) was proposed to solve this computational issue. The DMN is essentially an analytical surrogate model for a given micromechanics problem. Its expressivity comes from hierarchically stacking (into a binary tree) analytical homogenization formulas for binary laminate microstructures. Then with homogenization training data generated via FEM or FFT using a single microstructure, the geometric parameters of the DMN are learned, producing a fast surrogate micromechanics solver for the problem of interest. The fact that DMN learns a single microstructure means that it shines in an FE-DMN scheme when the microstructure chosen is a Representative Volume Element (RVE) and can be used at every material point in the macroscopic body. RVEs are often difficult to find, and in reality microstructures will have unavoidable manufacturing variations. This fact lead to the recent development of the Variational Deep Material Network (VDMN), which makes the geometric parameters of the original DMN into random variables, so that the output is a random variable over the effective properties of nearby non-RVE microstructures. The VDMN enables us to sample the effective property distribution, assign the samples spatially in a macroscopic FEM, and simulate under arbitrary boundary conditions which an experiment may require. Here we describe the VDMN for the thermal conduction problem as well as its FE coupling. SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525. SAND2026-18466A.

24 To Be Determined

Qingjie Li, Matlantis, Inc.

25 Improving Machine Learning Extrapolation for Molecular Discovery

Quinn Gallagher, Princeton University

The objective of data-driven molecular discovery is to leverage insights from data to more effectively identify molecules with state-of-the-art properties. Commonly employed approaches include screening molecular databases with machine learning property surrogates, Bayesian optimization, or steered generative modeling. However, the classical and deep machine learning models used by these approaches struggle to accurately make predictions for molecules with extreme property values [1], especially when relevant labeled datasets are small. In short, the three most prevalent algorithms for ML-driven materials discovery each deploy methods that are misaligned with the extrapolative demands of the problem.

To address this gap, we benchmark, develop, and apply extrapolative ML methods for molecular discovery. First, we identify whether simple modifications to classical ML models, such as Δ-learning or synthetic data generation, can improve out-of-support property prediction relative to standard baselines. Second, we analyze how large-scale pretraining on unlabeled molecular data modifies extrapolative performance, using the geometry of latent representations to rationalize model behavior. After identifying best practices for extrapolation, we adapt leading materials discovery methods accordingly by evaluating whether robust extrapolation meaningfully improves the efficacy of screening, adaptive experimentation, and steered generative modeling.

Overall, we provide a systematic benchmarking and development of extrapolative ML methods for molecular discovery. Our results identify best practices for extrapolation across common ML model classes and discovery algorithms, with particular implications for the discovery of molecules in sparsely labeled chemical spaces—such as therapeutic candidates or functional materials—where extrapolative prediction is most useful. We recommend concrete strategies for future ML-driven molecular design campaigns and suggest avenues for further study.

[1] Segal, N. Known Unknowns: Out-of-Distribution Property Prediction in Materials and Molecules. npj Comput Mater 2025, 11 (1), 345.

26 A case study in the development of improved promoted Pt catalysts for propane dehydrogenation through Bayesian optimization with uncertainty quantification

Ricardo Mathison Fuenmayor, Johns Hopkins University

Exploring the vast design space of multimetallic heterogeneous catalysts using traditional lab-scale reactors is often time-consuming and labor-intensive. This is due to the complexity of catalyst formulation and challenges with data quality arising from imprecise synthesis and reactivity testing. In this study, we present a strategy that leverages Bayesian optimization to efficiently identify promising catalysts within a large design space—over 1 million possible formulations—of Pt/γ-Al2O3 catalysts promoted with five different metals (Sn, Ga, Fe, Cu, and Ca). Our approach employs a two-tiered optimization process: initially, we utilize a coarser grid of catalyst formulations to develop a broader model. Subsequently, we use a refined search incorporating finer variations in metal loadings. We also explore the impact of different surrogate models on the performance of the Bayesian optimization algorithm. By integrating uncertainty into the surrogate model, we reduce overfitting and improve our ability to predict catalyst performance. This combined approach allows us to rapidly identify high-reactivity, high-selectivity multimetallic catalysts comparable to industrial materials. Our method demonstrates a practical, iterative framework for experimentalists to efficiently explore novel chemistries, experimental conditions, or mechanistic hypotheses.

27 Building Quantitative Descriptors for Heterogeneous Semiconductor Interfaces

Rosa Diaz Rivas, Purdue University

Heterogeneous interfaces strongly influence the behavior of semiconductor devices. In many materials systems, small variations at the atomic scale can affect carrier transport, strain relaxation, and defect formation. These structural variations are often visible in atomic resolution microscopy, yet they are usually described qualitatively. As a result, it is difficult to compare interfaces across different growth conditions or materials systems, and the connection between structure and device behavior remains unclear.

In this work I focus on developing quantitative descriptors of heterogeneous semiconductor interfaces using atomic resolution high resolution scanning transmission electron microscopy. HRSTEM imaging provides direct access to the atomic arrangement at the interface region and allows us to observe how the lattice evolves from one material to the next. These images reveal features such as local lattice distortions, interface roughness, structural disorder, and variations in atomic column positions that occur as the two materials meet.

Rather than relying only on visual inspection, the atomic resolution images are digitally analyzed using signal processing approaches that allow structural variations to be measured directly from the data. From these measurements we extract parameters that describe the interface region, including interface abruptness, roughness, and the spatial distribution of disorder along the interface. These quantities provide a practical way to describe interfaces in terms that can be measured and compared.

The goal of this effort is to move toward a more consistent way of describing heterogeneous semiconductor interfaces using atomic resolution microscopy. By defining measurable descriptors of interface structure, it becomes possible to compare interfaces produced under different growth conditions and across different materials systems. Establishing these descriptors is an important step toward connecting atomic scale structure with the properties that ultimately determine semiconductor device performance.

28 STAMP: Species- and Topic-aware Representation Learning for Antimicrobial Peptide Discovery

Sarala Padi, NIST

The discovery of potent antimicrobial peptides (AMPs) is critical to addressing global antibiotic resistance. However, due to vast sequence space makes exhaustive wet-lab validation of generative candidates computationally and economically prohibitive. Current predictive models for Minimum Inhibitory Concentration (MIC) often fail to generalize across diverse microbial targets due to the highly context-dependent, species-specific nature of peptide-membrane interactions. We present STAMP (Species- and Topic-Aware Multimodal Predictor), a unified framework that reformulates MIC prediction as a multimodal representation learning task. STAMP integrates high-dimensional protein language model embeddings (ESM-2) with discrete taxonomic conditioning and latent topic-aware representations. This architecture enables the model to capture sequence-level biophysical motifs and align them with species-specific inductive priors, facilitating robust cross-species generalization. To support rigorous evaluation, we introduce a newly curated, standardized benchmark dataset derived from DBAASP, addressing systemic data inconsistencies and redundancies. STAMP achieves state-of-the-art performance, yielding a Pearson correlation of 0.837 and an R2 of 0.70. We demonstrate the real-world utility of STAMP through experimental validation against E. coli and S. epidermidis, where predicted activity strongly correlates with in vitro results. Finally, mechanistic interpretability analysis via residue-level attribution reveals that STAMP recovers known biophysical determinants of antimicrobial efficacy, establishing it as a scalable and interpretable foundation for accelerated drug discovery.

29 Predicting Properties from Near-Infrared Spectra with Machine Learning for Improved Polyolefin Differentiation

Shuaijun Li, NIST

The increasing volume of plastic waste presents a critical challenge for recycling systems, where fast and accurate material sorting determines recycled product quality. Polyolefins, the largest class of commodity plastics, are particularly difficult to differentiate using current industrial sorting technologies such as near-infrared (NIR) spectroscopy due to their highly similar chemical structure, limiting the ability to perform property-based sorting needed for high-value recycling. In this work, we develop a machine learning framework that uses NIR spectra to directly predict key material properties, including density, crystallinity, and short-chain branching. By focusing on property prediction rather than classification, this approach enables a shift toward property-based sorting strategies. To improve interpretability, we identify the most important spectral features and link them to known polyolefin vibrational bands, establishing a direct connection between spectral signatures, polymer structure, and macroscopic properties. Ongoing efforts focus on extending the framework to post-consumer polyolefins to better capture real-world heterogeneity and enhance model robustness. This work advances data-driven sorting technologies that can improve recycling efficiency and support the transition toward a circular plastics economy.

30 Machine Intelligence Accelerated Design of Conductive MXene Aerogels with Programmable Properties

Snehi Shrestha, University of Maryland, College Park

Designing ultralight conductive aerogels with tailored electrical and mechanical properties is critical for various applications. Conventional approaches rely on iterative, time-consuming experiments across a vast parameter space. Herein, an integrated workflow is developed to combine collaborative robotics with machine learning to accelerate the design of conductive aerogels with programmable properties. An automated pipetting robot is operated to prepare 264 mixtures of Ti3C2Tx MXene, cellulose, gelatin, and glutaraldehyde at different ratios/loadings. After freeze-drying, the aerogels' structural integrity is evaluated to train a support vector machine classifier. Through 8 active learning cycles with data augmentation, 162 unique conductive aerogels are fabricated/characterized via robotics-automated platforms, enabling the construction of an artificial neural network prediction model. The prediction model conducts two-way design tasks: (1) predicting the aerogels' physicochemical properties from fabrication parameters and (2) automating the inverse design of aerogels for specific property requirements. The combined use of model interpretation and finite element simulations validates a pronounced correlation between aerogel density and compressive strength. The model-suggested aerogels with high conductivity, customized strength, and pressure insensitivity allow for compression-stable Joule heating for wearable thermal management.

31 Atomistic Dynamics as Sequential Decision-Making: Toward Experiment-Realistic Simulation

Wonseok Jeong, NIST

Atomistic simulation has advanced rapidly through machine learning, enabling increasingly accurate interatomic potentials and large-scale exploration of materials structures. However, a major gap remains between evaluating a material and simulating how that material is realized during processing. In realistic synthesis and processing environments, structural evolution is often controlled by rare, history-dependent events occurring in disordered and continuously evolving local environments. These events are difficult to capture with direct molecular dynamics, while many enhanced-sampling approaches require predefined reaction coordinates, endpoint guesses, or event catalogs that may not be available in complex interfaces or amorphous systems.

Here, we introduce REALIZE, a framework that reformulates rare-event atomistic pathway discovery as symmetry-constrained sequential decision-making in the full atomistic environment. Rather than relying on hand-crafted collective variables, REALIZE treats molecular motion as a sequence of physically constrained decisions, with each pathway evaluated by atomistic energetics and chemical verification. We apply REALIZE to dry silicon oxidation, where O_2 transport and dissociation through amorphous SiO_2 control interfacial growth but are inaccessible to straightforward molecular dynamics. REALIZE discovers O_2 transport and dissociation pathways, improves verified dissociation success, lowers effective activation barriers toward the experimental diffusion-limited regime, and generates evolved Si/a-SiO_2 interfacial structures.

These results suggest a route toward atomistic simulation of process pathways that are rare, history-dependent, and difficult to define in advance. More broadly, this work points toward future simulation frameworks that combine symmetry-aware atomistic models with pathway search, latent representations, and planning-based exploration for complex processing and environments.

32 Uncertainty-Aware Explainable AI for Process-Structure-Property Discovery

Yuhao Zhong, Virginia Polytechnic Institute and State University

Data-driven materials research increasingly seeks not only accurate prediction, but also trustworthy scientific discovery of how processing governs microstructure and, in turn, material properties. This poster presents an uncertainty-aware explainable AI framework for process-structure-property (PSP) discovery in advanced manufacturing. Our approach integrates heterogeneous manufacturing data—including process signals, microstructural characterization, and measured material responses—into predictive models with uncertainty (confidence) quantification. Beyond forward prediction, we develop explainable AI tools to identify which process and structure features most strongly influence target properties, enabling physically meaningful interpretation and discovery rather than black-box correlation.