William A. Brookshire
Department of Chemical
and Biomolecular Engineering

Cullen College of Engineering

ChBE Seminar Series

Bayesian Optimization-Driven Materials Discovery and Design: Hype, Power and Challenges

Paulette Clancy

Edward J. Schaefer Professor in Engineering, Department of Chemical and Biomolecular Engineering, Johns Hopkins University

The Materials Genome Initiative was started in 2011 with the prospect of using machine learning to make materials "faster, cheaper, better." A decade or so later, we are finally starting to have the tools to exploit that prospect more effectively. A major factor contributing to the inefficiency of material discovery and design is the large combinatorial space of materials candidates as well as processing conditions, which are often sparsely observed for a given application. Searches of this space are often formed by expert knowledge and clustered close to known materials. Exhaustive experimental characterization or first principles calculations are also expensive, invariably leading to small available data sets (<100 data points). As a result, there is a need to develop algorithms that can efficiently search this large parameter space capable of dealing with "tiny" data sets. In this talk, we will introduce our approach to mitigate some of these issues using our in-house PAL2.0, Bayesian optimization codebase, that features a chemistry-informed belief model.

Paulette Clancy is the Edward J. Schaefer Professor of chemical and biomolecular engineering at the Johns Hopkins University. She is currently the director of research for the Data Science and AI Institute at Hopkins. She is a Fellow of the Royal Society of Chemistry and a Fellow of the American Institute of Chemical Engineers. She spent over 30 years teaching at Cornell, including as the department chair for eight years. She moved to Johns Hopkins in 2018 to become the inaugural department Head of ChemBE (2018-2023). Clancy has one of the leading computational and AI/ML groups in the country studying atomic-scale modeling of semiconductor materials and sustainable energy applications. Her projects range from traditional silicon-based compounds to all-organic materials including flexible wearable organic electronics; machine learning algorithm development; electronic materials processing (III-V semiconducting materials and polymer-based thermoelectrics); and nucleation and crystal growth (hybrid organic/inorganic perovskites and quantum dot nanocrystals). Her lab focuses on studies of advanced materials processing and nucleation to link processing, structure and function. Her group develops new Bayesian optimization methods that encode chemistry and physics knowledge and intuition. Current machine learning projects creating optimal conditions for making energy-efficient solar cells, close-to-perfect quantum dots, and discovering new high-entropy alloys for shape memory applications.

10/24/2025, 10:30 – 11:30am, L2D2