Development of an autonomous synthesis framework for redox-active organic materials

Statement of Work

Period of Performance: Until September 30th, 2024

Technical Point of Contact:
Professor Shijing Sun
Department of Mechanical Engineering
University of Washington
3900 E Stevens Way NE, Seattle, WA 98195
Phone number: 1 857-242-9606
Email: shijing@uw.edu

Background

Aqueous flow batteries that utilize organic active materials offer an eco-friendly, adaptable, and secure solution for large-scale energy storage. They are particularly well-suited for long-term grid storage, which is needed to address the challenges posed by global climate change. However, finding and synthesizing the right redox molecules for flow battery applications remains a formidable challenge, demanding extensive molecular engineering.[1] The current state-of-the-art synthetic techniques encounter two primary limitations: 1) the vast and unexplored parameter space renders trial-and-error approaches highly inefficient; 2) the slow manual handling of synthesis procedures and data analysis hampers the pace of learning.[2] To tackle these challenges, there has been a growing focus on new data-driven frameworks aimed at expediting the materials discovery process. Recently, substantial interest has arisen in autonomous synthesis frameworks, where high-throughput synthesis and AI-driven data analytics combine to accelerate the journey from the laboratory to the market for new energy materials in replacement of the labor-intensive bench chemistry.[3]In this work, we propose the development of an **AI agent** capable of orchestrating the automated synthesis platforms at PNNL. This AI agent aims to function as the 'brain' for the high-throughput experimentation systems (HTESs) and therefore significantly accelerate the discovery of redox-active organic materials suitable for flow battery applications.

Scope of Work

Our proposed work comprises three work packages: WP1 and WP2 focus on developing machine learning models for property prediction (Model 1) and decision-making (Model 2), respectively. WP3 involves developing a data pipeline for creating a closed-loop research framework that enables automated materials discovery without human intervention. In addition to these three work packages, we will collaborate with PNNL (or a third-party collaborator), contributing to the development of hardware-software interface. This collaboration aims to enable the AI agent to

communicate directly with HTESs hardware, facilitating a fully autonomous synthesis platform.

WP1: Initial Sampling and the Prediction Model

We will develop a machine learning model (Model 1) based on supervised learning methods such as Gaussian process regression. This model will take synthetic parameters as input, conduct necessary feature engineering, and correlate these inputs with the properties of interest for materials. Simultaneously, it will provide uncertainty estimates for the posterior means. In collaboration with experimentalists at PNNL, we will design the initial sampling strategy, considering methods such as grid search, Latin-hypercube sampling, and random sampling, to generate a representative high-fidelity dataset for training the machine learning models. Additionally, we will explore the possibility of initiating model training using existing experimental data.

WP2: Sequential Decision-making Model

We will develop an active learning model (Model 2) determining the subsequent synthetic conditions based on observations from previous experimental observations. Depending on the throughput of the HTESs, we will build batch Bayesian optimization methods for organic material synthesis using Model 1 as a surrogate model.[4] We will determine acquisition functions to direct the search for the next evaluation point, striking a balance between exploration (probing areas with high uncertainty) and exploitation (probing areas with high anticipated objective values).

WP3: Prototype Closed-loop Experimentation Campaign

In this work package, we will integrate Model 1 and Model 2 into an AI agent and apply it to guide experimental data collection at PNNL. We will demonstrate a campaign consisting of three rounds of data collection, with each round using experimental parameters suggested by the AI agent. We will benchmark the acceleration in materials search achieved by the AI agent against random and grid sampling. In each round of organic synthesis, the AI agent will perform the following tasks:

I. Verify the validity of the new data collected.
II. Update the surrogate model with the new data points and propose the next set of synthetic conditions or decide to end the campaign.
III. Generate the synthetic protocol, upload it to share with the experimentalist, or directly communicate with the HTES robots if available, to initiate the subsequent round of sampling or end the campaign.

Deliverables:

0 - 3 months: Development of Model 1, the prediction unit
3 - 6 months: Development of Model 2, the decision-making unit and the integration of the two units
6 - 9 months: Execution of the experimental campaign guided by the AI agent **9 - 12 months:** Results analysis and preparation of the manuscript

Quarterly research summary and an annual report (or manuscript for peer review publication)

will be submitted to Dr. Liang at PNNL.

Reference:

1. Feng, R. *et al.* Reversible ketone hydrogenation and dehydrogenation for aqueous organic redox flow batteries. *Science* **372**, 836–840 (2021).
2. Liang, Y. *et al.* High-throughput solubility determination for data-driven materials design and discovery in redox flow battery research. Preprint at https://doi.org/10.26434/chemrxiv-2023-985h7 (2023).
3. Montoya, J. H. *et al.* Toward autonomous materials research: Recent progress and future challenges. *Appl. Phys. Rev.* **9**, (2022).
4. Sun, S. *et al.* A data fusion approach to optimize compositional stability of halide perovskites. *Matter* **4**, 1305–1322 (2021).