Combining robotic deposition tools and advanced characterization to enable ML-guided material discovery

Tim Kodalle[1] Maher Alghalayini[1] Arian Gashi[1] Morgan Wall[1] Fabrice Roncoroni[1] Yi-Ru Liu[1] Aleksandr Razumtcev[1] Shaul Aloni[1] Adam Schwartzberg[1] Edward S. Barnard[1] Carolin M. Sutter-Fella[1]

[1]*[Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA-94608, USA]*

Correspondence to: [Tim Kodalle] timkodalle@lbl.gov

1. Introduction

Conventional material synthesis discovery processes typically consist of an iterative sequence of literature-informed synthesis, characterization, and consecutive adjustments to the synthesis parameters. These steps often involve large parameter spaces with numerous tuneable parameters, which makes the overall process slow, labour-intensive, and incremental. To be able to accelerate the synthesis discovery processes, the community has shifted their attention toward autonomous and machine learning (ML-)assisted high-throughput materials synthesis in recent years. Significant efforts have been made to use text-mining and high-throughput computation to identify and contextualize new material candidates.[1,2] In addition, many groups have started to use robotic, ML-driven experimental agents to perform high-throughput materials synthesis, often accompanied by in-line characterisation.[3-6]
However, in-line characterization often provides valuable but insufficient information about a new material and its properties, especially since researchers usually condense spectra and patterns to scalar figures of merit for optimization algorithms.[3,4]
To tackle that issue, this research aims to combine high-throughput material synthesis screening with advanced in-situ and ex-situ characterisation on two separate experimental platforms. We will discuss how the combination of Bayesian Optimization and in-situ characterization can be leveraged to efficiently and effectively explore (rather than exploit) large process parameter spaces as well as how synchronised high-throughput robotic synthesis and characterisation can be used to provide large datasets with high-quality metadata to be analysed by ML-models extracting synthesis-structure-property-relationships.

2. Results

The first platform discussed in this work is an in-house developed robotic set-up for plasma-enhanced atomic layer deposition (PE-ALD), referred to as 'ALDBot' in the following. ALDBot is controlled by the ScopeFoundry Python package [7] and has access to twelve process parameters, including precursor dose time, process pressure, plasma duration, etc., guiding the PE-ALD process. Additionally, ALDBot is equipped with an in-situ spectral ellipsometry (SE) setup measuring the thickness of the deposited thin-film in real time. Hereby, the growth rate per ALD-cycle (GPC) as determined from SE functions as the figure of merit, with a GPC equal to one monolayer of the target material representing optimal growth conditions. Using ALD-TiO_2 on Si as a model system, we demonstrate how robotic experimentation guided by Bayesian Optimization (BO) algorithms is able to efficiently explore a four-dimensional sub-set of the 12D parameter space fully modelling the GPC landscape across the four parameters with just 100 experiments (fully automated these were performed within less than one day). To do so, we start with a set of 5 random parameter combinations to provide the model with initial data. For each subsequent experiment, the BO determines the parameter combination with the highest uncertainty in order to minimize this uncertainty by providing GPC data. We show how this BO-approach outperforms random sampling of the parameter space both in terms of process efficiency and model accuracy (see Fig.1). We then expand the parameter space to the full 12 dimensions and use the same BO approach to explore this larger space. We similarly find that the BO outperforms random sampling and use feature importance analysis to identify the most influential process parameters. We furthermore discuss limitations of the model exploring the 12D space: the curse of dimensionality limits the accuracy even of the BO-approach when using very sparse data, requiring substantially more than 100 experiments to reach satisfactory model accuracy.

The second platform discussed in this work is a robotic tool for thin-film deposition from liquid precursors using a combination of spin-coating

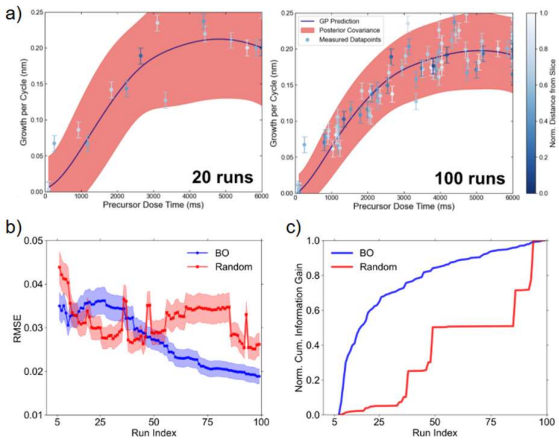

Fig. 1: a) Example of the evolution of the BO model during experimentation. Already after just 20 runs (experiments), the model accurately describes the underlying relationship between the model parameter and GPC. b) and c) show a comparison of the accuracy and efficiency of the BO-approach and random parameter sampling.

and annealing on a substrate, 'Spinbot'.[4] Combining Spinbot with robotic high-throughput optical characterisation as well as synchrotron based high-throughput (and partially automated) structural and chemical characterisation, we develop a "human-in-the-loop" concept to screen large parameter spaces in a short amount of time providing large datasets for ML-assisted discovery of synthesis-property relationships. In a first step, we developed workflows and data infrastructure to accurately track, annotate, and archive synthesis parameters, process and measurement datasets, as well as the physical thin-film samples themselves. In the next step, we aim to use Spinbot and corresponding characterization pipelines to fabricate and analyse about 2000 thin-film samples per month over a period of several months. We will discuss data management using an in-house developed platform ('Crucible'), as well as preliminary results derived for a model system from the growing database. We will also showcase how the combination of synthesis and characterisation robots is a powerful tool providing high-quality metadata and therefore enabling automated data analysis using customized ML-models.

Acknowledgments

T.K., M. A., and A. R. acknowledge
support from the Laboratory Directed Research & Development program (LDRD) of Lawrence Berkeley National
Laboratory funded under U.S. Department of Energy contract number DE-AC02-05CH11231.

References

[1] A. Jain, et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Mater. 1, 011002, 2013.
[2] M. Abdelsamie, et al. Combining text mining, in situ characterization, and ab initio calculations to rationalize BiFeO3 crystallization pathways. Matter 6 (12), 4291-4305, 2023.
[3] N. J. Szymanski, et al. An autonomous laboratory for the accelerated synthesis of inorganic materials. Nature 624, 86-91, 2023.
[4] A. Halder, et al. AI-Driven Robot Enables Synthesis-Property Relation Prediction for Metal Halide Perovskites in Humid Atmosphere. Adv. Ener. Mat. 15 (34), 2502294. 2025.
[5] N. Meftahi, et al. Machine Learning Enhanced High-Throughput Fabrication and Optimization of Quasi-2D Ruddlesden–Popper Perovskite Solar Cells. Adv. Ener. Mat. 13 (38), 2203859, 2023.
[6] J. Zhang, et al. Advancing perovskite photovoltaic technology through machine learning-driven automation. InfoMat 7 (5), e70005, 2025.
[7] D. B. Durham, et al. Scanning Auger spectromicroscopy using the ScopeFoundry software platform. ECASIA 50 (11), 1174-1179, 2018.