Integrating Experimental Expertise with Adaptive Bayesian Optimization for Perovskite Synthesis

Linhao Liu*
University of Ottawa
Department of Mechanical Engineering

Michel Nganbe
University of Ottawa
Department of Mechanical Engineering

Ericsson Tetteh Chenebuah
University of Ottawa
National Research Council of Canada

Alain Tchagang
University of Ottawa
National Research Council of Canada

Abstract

The ball milling synthesis of perovskite materials involves exploring a complex, high-dimensional parameter space, where conventional trial-and-error approaches are inefficient. For novel systems, large-scale prior datasets are often unavailable and may introduce bias if over-relied upon. In perovskite synthesis, the representation of process parameters and precursor descriptors plays a decisive role in the performance of predictive models and optimization strategies. To address this challenge, we propose a machine learning (ML)-guided Bayesian optimization (BO) framework for adaptive experimental design to accelerate the optimization of perovskite synthesis parameters. The framework integrates physicochemical descriptors of precursor elements with ball-milling and heat-treatment variables to construct a crystallite-size prediction model, which is embedded into a BO loop to dynamically guide experiments toward target crystallite sizes. This enables systematic control and fine-tuning of crystallite size, representing a performance-oriented reverse design paradigm. Preliminary results show that the framework converges to optimal conditions within only a few experimental iterations, significantly outperforming traditional trial-and-error methods. Combining ML with BO effectively reduces the experimental search space, lowers costs, and accelerates materials synthesis. The proposed framework provides a promising pathway for intelligent, data-driven synthesis of perovskites and complex inorganic materials, laying the methodological foundation for future self-driving experimental work.

1 Introduction

Recently, machine learning (ML) has emerged as a powerful tool for materials synthesis optimization, enabling complex numerical modelling that substantially accelerates the discovery of novel perovskite catalysts, their performance prediction, and process optimization. Bayesian optimization (BO), as an efficient strategy for exploring high-dimensional parameter spaces under limited experimental budgets, has been increasingly applied to molecular screening and self-driving laboratory work, and is now regarded as one of the core approaches for accelerating materials discovery and synthesis [1–3]. However, the efficiency of BO critically depends on the representation of features, which must be sufficiently informative to capture essential material and process information while remaining compact enough to avoid the curse of dimensionality [4, 5]. In the synthesis of complex inorganic materials, constructing suitable feature vectors and achieving

*All correspondence should be directed to lliu173@uottawa.ca

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

dynamic optimization under data-scarce conditions remain pressing challenges. ABX_3 perovskites, owing to their tunable crystal structures and broad energy-related applications, represent an ideal platform for validating AI-driven synthesis optimization strategies. Their performance is highly sensitive to crystallite size [6, 7], which is itself governed by the multidimensional coupling of ball milling parameters (e.g., ball-to-powder ratio, rotational speed, milling duration, and ball type) together with subsequent heat treatments [8, 9]. Conventional trial-and-error optimization not only suffers from inefficiency but also fails to enable goal-oriented inverse design. Notably, previous studies have explicitly argued that crystallite size is not strictly controllable during synthesis and therefore should not be considered as a suitable feature or optimization target [10]. In contrast, the present study deliberately sets crystallite size as the optimization objective and employs an ML-BO framework to iteratively approach the desired value.

2 Adaptive Learning Bayesian Optimization Framework for Perovskite Synthesis

The Adaptive Learning Bayesian Optimization (ALBO) workflow is designed to efficiently predict and optimize the synthesis parameters of ABX_3 perovskites with crystallite size as the target property, while minimizing costly experiments or simulations, as shown in Figure 1. The closed-loop optimization cycle includes data updating and labelling, surrogate model training and refinement, and the selection of subsequent experiments through an acquisition function. Similar to conventional Bayesian optimization, this framework relies on surrogate models capable of uncertainty-aware predictions and acquisition functions that balance exploration and exploitation.

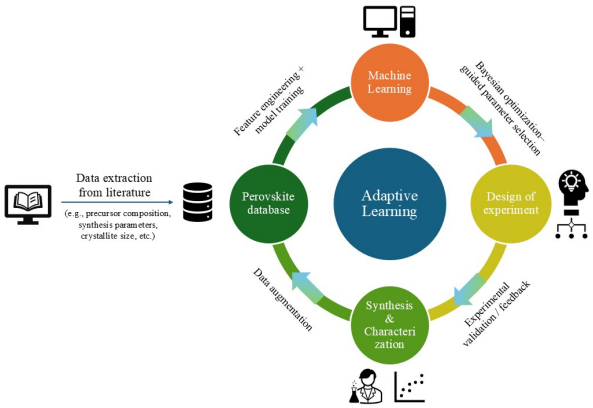


Figure 1: Closed-loop workflow for perovskite synthesis optimization. Literature-derived data and elemental descriptors are used to train ML models for crystallite size prediction, while Bayesian optimization proposes candidate conditions that are experimentally validated and fed back to expand the dataset.

In the model construction, bootstrap resampling was employed to repeatedly sample and evaluate the training data, thereby mitigating the impact of data scarcity and heterogeneity while enabling robust comparison across multiple regression models. The best-performing model identified through this procedure was then applied to imput missing crystallite size values in the database, ensuring completeness and reliability of the dataset. This model was subsequently retrained on the completed dataset to establish a stable mapping between synthesis parameters and crystallite size. To support model training, this study systematically curated approximately 330 experimental records from more than 94 peer-reviewed publications over the past two decades. Each record includes detailed synthesis information such as precursor composition, ball milling parameters (e.g., ball-to-powder ratio, rotational speed, and milling time, etc.), subsequent heat-treatment conditions, as well as crystalline structural outcomes including crystallite size, lattice parameters, space group, and phase. To further enrich the chemical representation, this dataset was integrated with a self-constructed library of physicochemical descriptors of elements, categorized into five groups: (i) periodic table position, (ii) atomic and ionic radii, (iii) electronic properties, (iv) thermochemical properties, and

(v) molar volume and ionic size–related descriptors [11]. By combining process-level parameters with element-level descriptors, a multidimensional feature space was established that simultaneously captures experimental conditions and the intrinsic chemistry of precursors. Importantly, all databases employed in this study are openly accessible, ensuring transparency and reproducibility in scientific research.

For feature representation, the ALBO framework incorporates the physicochemical descriptors of precursor elements (expressed as mean and standard deviation) together with key process variables [11]. This combined representation not only reflects the intrinsic chemical nature of the precursors but also accounts for the synthesis conditions. During the optimization stage, the trained model was coupled with Bayesian Optimization, employing the Expected Improvement (EI) acquisition function to inversely predict synthesis parameters that approach the predefined crystallite size target.

2.1 Experimental Validation Material

In this study, lanthanum ferrite ($LaFeO_3$) was selected as the validation material for the Adaptive Learning Bayesian Optimization (ALBO) framework. $LaFeO_3$ is an orthorhombic perovskite oxide that has attracted significant attention due to its broad functional applications. It has demonstrated excellent performance as a cathode material in solid oxide fuel cells, as a sensing material, and across numerous catalytic domains [12, 13]. Its structural stability, ease of achieving high phase purity, strong dependence of catalytic activity on crystallite size, and substantial representation in the dataset make it an ideal candidate for validating synthesis optimization strategies. These characteristics originate from its crystal structure, where La^{3+} ions occupy the A-site, Fe^{3+} ions reside at the B-site, and oxygen forms an interconnected octahedral network. Moreover, the structural flexibility of $LaFeO_3$, together with the possibility of tuning its properties through A-site or B-site substitution (or both), further enhances its multifunctionality in diverse applications [13]. Its broad relevance in energy conversion and catalysis further underscores its suitability and representativeness as a benchmark compound for validating AI-driven synthesis frameworks.

3 Results

Through bootstrap resampling, multiple regression models were evaluated for their ability to predict crystallite size. The comparison revealed clear differences in performance, with the Random Forest Regressor model achieving an initial prediction accuracy of approximately 54%, making it the most stable candidate among the tested models. This model was subsequently employed to imput the missing crystallite size values in the dataset and retrained on the completed dataset. After retraining, the model reached an accuracy of about 75% on the testing set (shown in Figure 2(a)), representing a significant improvement over the initial stage and demonstrating its capability to reliably capture the relationship between synthesis parameters and crystallite size.

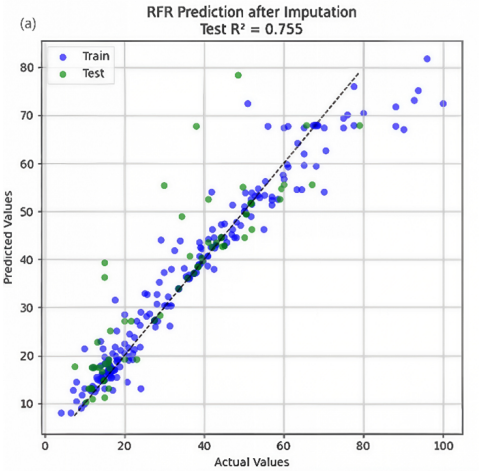
As shown in Figure 2(b), the XRD patterns compare the samples obtained under identical synthesis conditions but with different milling durations (2 h, 6 h, and 19 h). It can be observed that, under the Bayesian optimization-recommended synthesis parameters, all three conditions successfully yielded a high-purity orthorhombic (Pnma) $LaFeO_3$ perovskite phase, with diffraction peaks in good agreement with the Materials Project ID mp-22590 [14]. The corresponding lattice parameters are summarized in Table 1.

Table 1: Lattice parameters and crystallite sizes of synthesized $LaFeO_3$ samples. Experimental crystallite size was calculated using the Scherrer equation.

Sample	a (Å)	b (Å)	c (Å)	Experimental size (nm)	Target size (nm)
1	5.424	7.807	5.445	21.02	31
2	5.424	7.807	5.445	18.95	31
3	5.424	7.807	5.445	17.85	31

A closer comparison reveals that the sample milled for 6 h prior to calcination exhibited the strongest diffraction peak intensities, indicating the highest crystallinity, with a crystallite size slightly smaller than the target value. In contrast, the 19 h-milled sample did not display any fundamental structural

difference, but its crystallite size further decreased without a significant improvement in crystallinity. On the other hand, the 2 h-milled sample produced a crystallite size closer to the predefined target, yet its weaker diffraction peaks suggested a lower degree of crystallinity compared to the 6 h-milled sample. Overall, the first set of Bayesian optimization-recommended parameters consistently



(a) Performance of the Random Forest Regressor (RFR) model. Training (blue) vs testing (green).

(b) X-ray diffraction (XRD) patterns of $LaFeO_3$ perovskite at different milling times.

Figure 2: (a) RFR model performance; (b) XRD patterns of $LaFeO_3$ perovskite.

enabled the synthesis of high-purity $LaFeO_3$ perovskite across different milling times, with the most favourable balance achieved under the 6 h milling and 1h calcination conditions. It is worth noting that, although the long milling duration (19 h) resulted in the largest deviation from the target crystallite size, feeding the entire set of experimental results back into the original database and retraining the model to refine the mapping between synthesis parameters and structural features is a critical step of adaptive learning. This iterative feedback strategy ensures continuous model improvement and further validates the effectiveness of the ALBO workflow in guiding crystallite size optimization within a limited number of experimental iterations.

4 Conclusion

In this work, we propose an Adaptive Learning Bayesian Optimization (ALBO) workflow designed to achieve closed-loop optimization of ABX_3 perovskite synthesis parameters with crystallite size as the target property. The framework employs bootstrap resampling to enable robust model selection and mitigate biases arising from data scarcity and heterogeneity. At the same time, it integrates Bayesian Optimization (BO) to realize inverse design, allowing the crystallite size to be systematically controlled and precisely tuned to approach predefined targets. Under limited data conditions, this method establishes a reliable mapping between synthesis parameters and crystallite size. By combining the physicochemical descriptors of precursor elements with key process variables, the feature representation captures both the intrinsic chemical nature of the precursors and the experimental synthesis conditions, ensuring that the model learns physically and chemically meaningful relationships rather than mere statistical correlations. Experimental validation demonstrated that the BO-recommended synthesis parameters successfully produced a high-purity perovskite phase with a crystallite size close to the predefined target, fully confirming the feasibility and accuracy of the workflow. The results indicate that the ALBO framework can efficiently guide solid-state synthesis under limited experimental budgets and achieve directional control of crystallite size to meet specific performance requirements. This performance-oriented reverse design strategy breaks through the traditional forward-thinking paradigm of predicting performance from process parameters, establishing a new framework for intelligent, data-driven materials synthesis. Ongoing iterative experiments and model refinement aim to further enhance the workflow and validate its applicability and generalizability to more complex perovskite and alternative synthesis routes.

Data Availability

All data and source code supporting this study are openly available on GitHub at https://github.com/Linhao-Liu/ALBO.git

Acknowledgements

This research was supported by the National Research Council of Canada (NRC) through its Artificial Intelligence for Design Program led by the Digital Technologies Research Center and by the Natural Sciences and Engineering Research Council of Canada [NSERC Discovery Grant number: GR000882]. The authors also gratefully acknowledge Ernest Acheampong and Adeola Bamigbaiye for their contributions to the ball-milling experiments, and Ziqi Tang for assistance with the XRD pattern analysis.

References

[1] Frendy Jaya Kusuma, Eri Widianto, Wahyono, Iman Santoso, Sholihun, Moh. Adhib Ulil Absor, Setyawan Purnomo Sakti, and Kuwat Triyana. Multi-Properties prediction of perovskite materials using Machine learning and Meta-Heuristic feature selection. *Solar Energy*, 286:113189, January 2025.

[2] Fuchun Gou, Zhu Ma, Qiang Yang, Hao Du, Yixian Li, Qian Zhang, Wei You, Yi Chen, Zhuowei Du, Junbo Yang, Nan He, Junxin Luo, Zichen Liu, Zilu Tian, Maozhu Mao, Kai Liu, Jian Yu, Anan Zhang, Fan Min, Kuan Sun, and Ningqiang Xuan. Machine Learning-Assisted Prediction and Control of Bandgap for Organic–Inorganic Metal Halide Perovskites. *ACS Applied Materials & Interfaces*, 17(12):18383–18393, March 2025. Publisher: American Chemical Society.

[3] Ruiqi Zhang. Machine Learning Prediction of Organic–Inorganic Halide Perovskite Solar Cell Performance from Optical Properties | ACS Energy Letters.

[4] Mahyar Kochi, Negareh Mahboubi, Aseem Gill, and Seyed Moosavi. Adaptive Representation of MOFs in Bayesian Optimization. *38th Conference on Neural Information Processing Systems (NeurIPS 2024)*, October 2024.

[5] Optimizing the Synthesis Parameters of Double Perovskites with Machine Learning Using a Multioutput Regression Model | The Journal of Physical Chemistry C.

[6] Md. Sofiul Alom, Chandana C.W. Kananke-Gamage, and Farshid Ramezanipour. Perovskite Oxides as Electrocatalysts for Hydrogen Evolution Reaction. *ACS Omega*, 7(9):7444–7451, March 2022. Publisher: American Chemical Society.

[7] Ardra S. Darsan and A. Pandikumar. Recent research progress on metal halide perovskite based visible light active photoanode for photoelectrochemical water splitting. *Materials Science in Semiconductor Processing*, 174:108203, May 2024.

[8] Mohammad Ghasdi and Houshang Alamdari. CO sensitive nanocrystalline LaCoO3 perovskite sensor prepared by high energy ball milling. *Sensors and Actuators B: Chemical*, 148(2):478–485, July 2010.

[9] M. Imani, M. Farajnezhad, and A. Tadjarodi. 3D hierarchical flower-like nanostructure of Bi2MoO6: Mechanochemical synthesis, the effect of synthesis parameters and photocatalytic activity. *Materials Research Bulletin*, 87:92–101, March 2017.

[10] Synthesis of Machine Learning-Predicted Cs2PbSnI6 Double Perovskite Nanocrystals | ACS Nano.

[11] Ericsson Tetteh Chenebuah, Michel Nganbe, and Alain Beaudelaire Tchagang. Comparative analysis of machine learning approaches on the prediction of the electronic properties of perovskites: A case study of ABX3 and A2BB'X6. *Materials Today Communications*, 27:102462, June 2021.

[12] Roberto Köferstein, Lothar Jäger, and Stefan G. Ebbinghaus. Magnetic and optical investigations on LaFeO3 powders with different particle sizes and corresponding ceramics. *Solid State Ionics*, 249-250:1–5, November 2013.

[13] Nadir Ouldhamadouche, Ahmed Belguenoune, Salah Bassaid, Abdelkader Dehbi, and Abdallah Ben Rhaiem. Synthesis and characterization of LaFeO3 perovskite: Insights into dielectric properties and conduction mechanism. *Physica B: Condensed Matter*, 714:417513, October 2025.

[14] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013.