Metastable Polymorph Stabilization with Physics-based Descriptor Engineering and Machine Learning

Dai Haiwen[a], **Andy Paul Chen**[a], **Shuya Yamazaki**[a], **Andre Low**[a], **Jose Recatala-Gomez**[a], **Meng Fanxu**[a], **Kedar Hippalgaonkar**[a]

[a] *School of Material Science and Engineering, 50 Nanyang Avenue, Nanyang Technological University, Singapore 639798.* haiwen.dai@ntu.edu.sg

1. Introduction

1.1 Background

Crystal polymorphism, observed both in nature and the laboratory, refers to the phenomenon where materials undergo structural transitions without altering their stoichiometry. These transitions are often accompanied by significant changes in the material's functional properties, making polymorphs valuable for a wide range of applications. Materials with rapid transition kinetics are useful in phase transition devices, while those with slower kinetics can be stabilized for use in ambient conditions.

Traditionally, it is understood that displacive transitions tend to occur more rapidly than reconstructive transitions, making the stabilization of metastable polymorphs more challenging. However, this generalization does not always hold. For instance, the metal-to-insulator transition in VO_2 occurs on a femtosecond timescale, while austinite, despite undergoing a displacive transition, can be stabilized through quenching. This illustrates that there is no straightforward rule for predicting which polymorphs will remain stable under ambient conditions and which will undergo further transitions. As a result, the trial-and-error approach remains the dominant method for discovering and utilizing polymorphs, which limits their potential for widespread application.

1.2 Related work

In recent years, material informatics has provided new insights into material stability. Aykol et al. [1] demonstrated that most materials, whether discovered or synthesized, lie on the hull surface (E_{hull} = 0), while others are only a few tens of meV/atom above the hull. This concept has since been widely adopted: materials with a large E_{hull} are considered thermodynamically difficult to synthesize, while those with a small E_{hull} are easier to produce. While this approach has proven effective, achieving approximately 70% accuracy, it does not offer insights into whether a polymorph can be stabilized in ambient conditions, as density functional theory (DFT) calculations are typically performed at 0 Kelvin.

Numerous techniques have been developed to stabilize metastable polymorphs in ambient conditions. Methods such as quenching[2], epitaxy[3], hydrothermal synthesis[4], precursor engineering [5], and surface mediation[6] aim to lower kinetic nucleation barriers and "trap" a phase in its metastable state. While these techniques have been successful in stabilizing metastable polymorphs experimentally, they do not provide reliable predictions regarding whether a polymorph can be stabilized in ambient conditions without further experimentation.

1.3 Hypothesis and this work

In this work, we aim to address the question: For a given metastable polymorph that is experimentally synthesizable, can it be stabilized under ambient conditions?

We hypothesize that the ability to kinetically trap a metastable polymorph is influenced by both structural descriptors and structure-transition-related descriptors. By combining these features in a physics-based model, we demonstrate that it is possible to predict at which condition a metastable polymorph can be stabilized. Using a six-descriptor tree-based model, we showcase the potential of this approach in providing predictions for the stabilization of polymorphs without the need for extensive experimentation.

2. Result and discussions

2.1 Model Training

Around 11,000 metastable phases exhibiting polymorphism, sourced from the Materials Project, were used for model training. These polymorphs were categorized into three target classes based on their characterization temperature (T) and pressure (P) from the ICSD database:

- Class 0: Ambient T&P;
- Class 1: Higher or lower than ambient P;
- Class 2: Ambient P and higher T.

Six structure-related descriptors were utilized, including unit cell size (normalized per atom), density, energy above the hull, number of elements, percentage change in unit cell size (normalized per atom), and percentage change in density. An 80%/20% train/test split was applied for model training. An XGBoost classifier was trained with a maximum depth of 6 to prevent overfitting. The model threshold was tuned to achieve a balanced precision and recall.

While additional material descriptors, such as band structure, atomic information, and global/local symmetry, could potentially improve the model, we chose not to incorporate them in this work. This decision was made to ensure the model's applicability even when a crystallographic information file (CIF) is incomplete, which is especially useful for unexplored material systems with limited composition or structural data.

2.2 Model performance

The XGBoost classifier achieved an overall accuracy of 83.58%, precision of 82.79%, and recall of 83.58% on the test set (Fig. 1). Specifically, the recall rates for classes 0, 1, and 2 were 85.23%, 75.52%, and 75.86%, respectively. These results suggest that the model performs well in predicting polymorphs, with most predicted candidates aligning with their ground truth. This makes the model particularly useful for

Metastable Polymorph Stabilization with Physics-based Descriptor Engineering and Machine Learning

Dai Haiwen[a], **Andy Paul Chen**[a], **Shuya Yamazaki**[a], **Andre Low**[a], **Jose Recatala-Gomez**[a], **Meng Fanxu**[a], **Kedar Hippalgaonkar**[a]

[a] *School of Material Science and Engineering, 50 Nanyang Avenue, Nanyang Technological University, Singapore 639798.* haiwen.dai@ntu.edu.sg

determining whether high temperature or pressure is necessary to stabilize a novel metastable polymorph within unexplored material system.

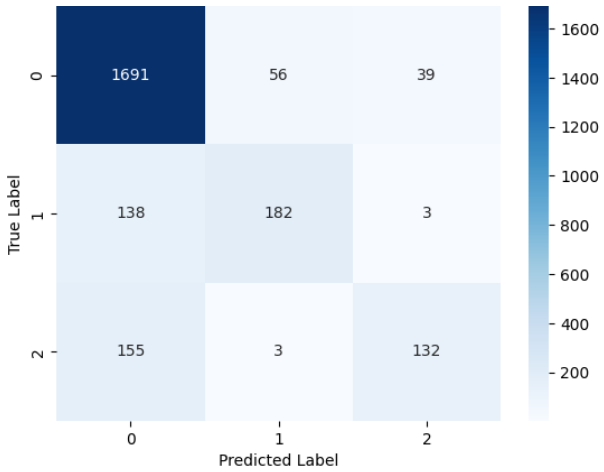

Fig. 1: Confusion matrix of the 3-class XGBoost classifier on the test set. (Class 0: Ambient T&P; Class 1: Higher/lower than ambient P; Class 2: Ambient P and higher than ambient T.)

To better understand the global feature importance, we conducted a SHAP analysis (Fig. 2), averaging feature importance by class to mitigate the dominance of class 0, which constitutes around 80% of the data. Notably, two self-constructed descriptors - percentage change in material density and percentage change in unit cell volume (normalized per atom) - ranked first and third in importance, respectively. The absolute unit cell size and density values ranked second and fourth. In contrast, the commonly used E_{hull} descriptor had much less influence, a result that challenges the common assumption that a metastable polymorph with low E_{hull} is easier to stabilize under ambient conditions.

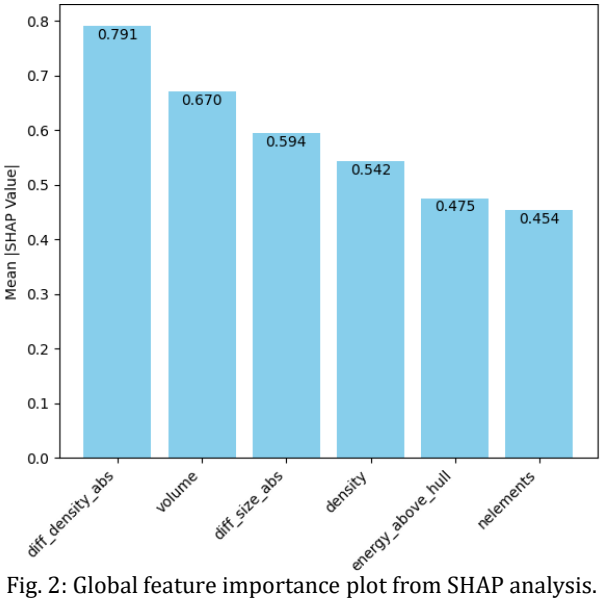

Fig. 2: Global feature importance plot from SHAP analysis.

These analysis underscores the significance of material volume and weight metrics, often overlooked in prior studies, in predicting the stabilization of metastable polymorphs.

3. Summary

In this work, we introduced two novel descriptors - percentage change in unit cell size and density - which, along with four other features, were used to train a machine learning classifier that predicts the conditions under which a metastable polymorph can be stabilized. The model achieved an accuracy of over 80%, demonstrating strong predictive performance. Feature importance analysis revealed that volume and weight metrics - often overlooked in previous studies - are key drivers in determining stabilization conditions. Due to the physics-based nature of the descriptors, the model does not require high-level symmetry or coordination information, making it useful for predicting the stabilization conditions of novel metastable polymorphs with incomplete structural or compositional data. This approach provides a guideline for metastable polymorph synthesis beyond conventional approaches.

Acknowledgments

The authors would like to acknowledge the funding support from National Research Foundation Singapore Grant No. NRF-NRFF13-2021-0011 and NRF-CRP25-2020-0002. ChatGPT is used to improve the readability of the work.

References

[1] M. Aykol, S. S. Dwaraknath, W. Sun, and K. A. Persson, "Thermodynamic limit for synthesis of metastable inorganic materials," *Sci. Adv.*, vol. 4, no. 4, p. eaaq0148, Apr. 2018, doi: 10.1126/sciadv.aaq0148.

[2] Y. Yao *et al.*, "Carbothermal shock synthesis of high-entropy-alloy nanoparticles," *Science*, vol. 359, no. 6383, pp. 1489–1494, Mar. 2018, doi: 10.1126/science.aan5412.

[3] H. Ding, S. S. Dwaraknath, L. Garten, P. Ndione, D. Ginley, and K. A. Persson, "Computational Approach for Epitaxial Polymorph Stabilization through Substrate Selection," *ACS Appl. Mater. Interfaces*, vol. 8, no. 20, pp. 13086–13093, May 2016, doi: 10.1021/acsami.6b01630.

[4] J. B. Felder, "Hydrothermal Synthesis: A Gateway To Metastable Crystals With Unusual Properties".

[5] Y. Zeng *et al.*, "Selective formation of metastable polymorphs in solid-state synthesis," *Science Advances*, vol. 10, no. 3, Jan. 2024, doi: 10.1126/sciadv.adj5431.

[6] X. Ma *et al.*, "Guiding Synthesis of Polymorphs of Materials Using Nanometric Phase Diagrams," *J. Am. Chem. Soc.*, vol. 140, no. 49, pp. 17290–17296, Dec. 2018, doi: 10.1021/jacs.8b11029.