Multi-property directed generative design of inorganic materials through Wyckoff augmented transfer learning

Shuya Yamazaki[a,c], **Wei Nong**[a], **Ruiming Zhu**[a,b], **Kostya S. Novoselov**[c], **Andrey Ustyuzhanin**[c], **Kedar Hippalgaonkar**[a,b,c]

[a] *School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore*
shuya001@e.ntu.edu.sg

[b] *Institute of Materials Research and Engineering, Agency for Science, Technology and Research (A*STAR), Singapore 138634, Singapore*

[c] *Institute for Functional Intelligent Materials, National University of Singapore, Singapore 117544, Singapore*

1. Introduction

Accelerated materials discovery is imperative for advancing technologies in energy conversion, storage, and catalysis. However, the search for new inorganic materials with multiple functional properties remains a formidable challenge due to data scarcity and the complexity of multi-objective optimization. In the past year, several attempts have been made to address these challenges, including MatterGen [1], DiffCSP++ [2], and CrystalFormer [3]. However, no existing model has successfully demonstrated multi-property-directed design while explicitly conserving space group symmetry across all space groups.

In this study, we present a generative inverse design framework that overcomes this limitation by integrating Wyckoff-position-based data augmentation and transfer learning, enabling the controlled generation of novel functional inorganic crystal structures while preserving site symmetry constraints. Our framework simultaneously optimizes key material properties, including space group, band gap, and formation energy, facilitating a systematic approach to property-driven materials discovery beyond P1 translational symmetry constraints.

2. Results and discussion

2.1. Wyckoff augmentation and transfer learning

Wyckoff augmentation applies space group-constrained E(3)-transformations to enrich crystallographic data by providing multiple, symmetry-equivalent views of the same structure, thereby expanding the effective training set (Fig 1a). This process embeds the necessary symmetry constraints directly into the model, enhancing its ability to predict functional properties and generate site symmetry-compliant materials.

Transfer learning employs a two-step procedure: first, the model is pre-trained on a large dataset to capture broad, fundamental symmetry–property relationships; then, it is fine-tuned on a smaller, target-specific dataset to refine the prediction of key material properties such as formation energy and band gap (Fig 1b).

Combined, these two strategies reduce the MAE for formation energy and band gap predictions by 10–30% while enhancing the reconstruction accuracy of Wyckoff positions—essential for preserving crystal symmetry—by more than 30%.

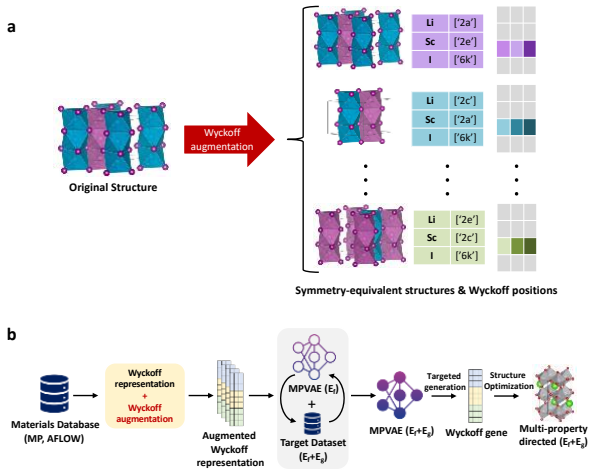

Fig. 1: (a) Wyckoff augmentation and symmetry equivalent structures; (b) An overview of the multi-property guided WyCryst generation framework [4]

2.2 MPVAE Generative Model Performance

Figure 2 illustrates the multi-property-structured latent space learned by the Multi-Property-directed Variational Autoencoder (MPVAE), where distinct regions correspond to specific target values or ranges for band gap, formation energy, and crystal systems. Conditional generation shifts the sampling distribution from the broad, unconditional baseline toward desired property targets, demonstrating the MPVAE's effective control over multiple properties simultaneously.

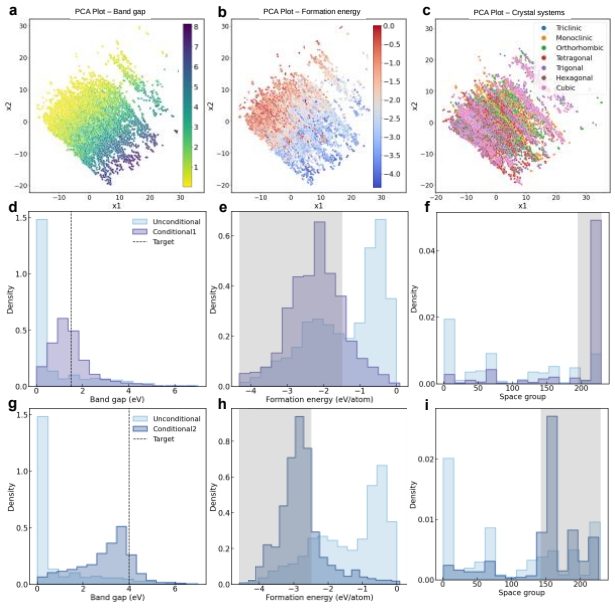

Fig. 2: Multi-property-structured latent space and conditional generation. (a-c) Property-structured latent space: (a) band gap (eV); (b) formation energy (eV/atom);

Multi-property directed generative design of inorganic materials through Wyckoff augmented transfer learning

Shuya Yamazaki[a,c], **Wei Nong**[a], **Ruiming Zhu**[a,b], **Kostya S. Novoselov**[c], **Andrey Ustyuzhanin**[c], **Kedar Hippalgaonkar**[a,b,c]

[a] *School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore*
shuya001@e.ntu.edu.sg

[b] *Institute of Materials Research and Engineering, Agency for Science, Technology and Research (A*STAR), Singapore 138634, Singapore*

[c] *Institute for Functional Intelligent Materials, National University of Singapore, Singapore 117544, Singapore*

(c) crystal systems; (d-i) Density of property values for conditional and unconditional generation using the MPVAE model for two different multi-property targets: Condition 1 (d) band gap = 1.5 eV, (e) formation energy < -1.5 eV/atom, and (f) space group ≥ 195; Condition 2 (g) band gap = 4.0 eV, (h) formation energy < -2.5 eV/atom, and (i) space group ≥ 143.

2.3 Multi-property-directed De Novo Generation Task

In our multi-property directed de novo generation task, we aimed to design high symmetry inorganic semiconductors with lower formation energies and band gaps ranging from 0.5 to 2.0 eV. The end-to-end process, illustrated in Figure 3, begins with conditionally generating thousands of Wyckoff representations that are rigorously screened for physicochemical validity, including charge neutrality, synthesizability [5], and thermodynamic stability, yielding 135 promising candidates. Subsequent DFT validations confirm our predictions through detailed comparisons of predicted versus DFT-calculated formation energies and band gaps, with further analysis identifying 8 dynamically stable novel semiconductors, as validated by Γ-phonon stability, and revealing favorable thermoelectric descriptors for potential applications.

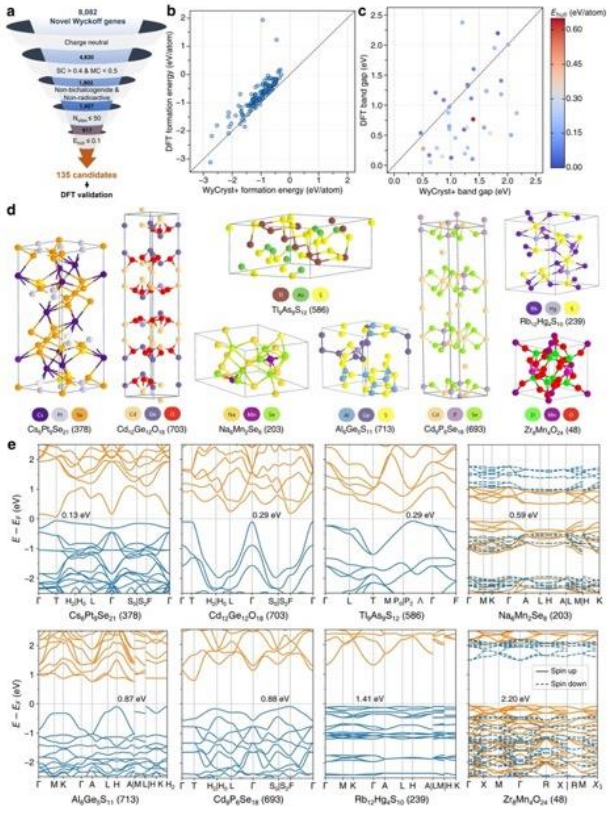

Fig. 3: Validation of property-directed design of semiconductor materials. (a) Screening steps for generated compounds. (b) WyCryst+ predicted and DFT-calculated formation energies for 135 candidates. (c) WyCryst+ predicted and DFT-calculated band gaps for 35 semiconductors, colored by the DFT-calculated E_{hull}. (d, e) DFT-relaxed crystal structures and corresponding electronic band structures of eight generated semiconductors with Γ-phonon stability.

3. Summary

In this work, we present a multi-property-directed generative model for the inverse design of inorganic materials by integrating Wyckoff-position-based data augmentation and transfer learning. Our framework addresses key challenges in materials discovery, particularly data scarcity and multi-objective optimization, improving both symmetry-preserving crystal generation and prediction accuracy for multiple target properties like formation energy, band gap, and space group. The integration of Wyckoff augmentation and transfer learning enhanced both forward property predictions and crystal structure generation by leveraging space group site symmetry and the two-step learning approach. Furthermore, we showcased the MPVAE's capability of controlling the distribution of multiple target properties in our multi-objective de novo generation tasks. Notably, this framework successfully generated 8 novel semiconductor materials with targeted functional properties, thermodynamic stability, and lattice-dynamic stability, offering a significant step forward in AI-driven inverse design of inorganic materials.

Acknowledgments

K.H. acknowledges funding from the MAT-GDT Program at A*STAR via the AME Programmatic Fund by the Agency for Science, Technology and Research under Grant No. M24N4b0034. K.H. also acknowledges funding from the NRF Fellowship NRF-NRFF13-2021-0011. The computational work of DFT calculations for this article was fully performed on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg) under the project No. 12003663.

References

[1] Zeni, C. *et al.* A generative model for inorganic materials design. *Nature* (2025) doi:10.1038/s41586-025-08628-5.

Multi-property directed generative design of inorganic materials through Wyckoff augmented transfer learning

Shuya Yamazaki[a,c], **Wei Nong**[a], **Ruiming Zhu**[a,b], **Kostya S. Novoselov**[c], **Andrey Ustyuzhanin**[c], **Kedar Hippalgaonkar**[a,b,c]

[a] *School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore*
shuya001@e.ntu.edu.sg

[b] *Institute of Materials Research and Engineering, Agency for Science, Technology and Research (A*STAR), Singapore 138634, Singapore*

[c] *Institute for Functional Intelligent Materials, National University of Singapore, Singapore 117544, Singapore*

[2] Jiao, R., Huang, W., Liu, Y., Zhao, D. & Liu, Y. Space Group Constrained Crystal Generation. Preprint at https://doi.org/10.48550/arXiv.2402.03992 (2024).

[3] Cao, Z., Luo, X., Lv, J. & Wang, L. Space Group Informed Transformer for Crystalline Materials Generation. Preprint at https://doi.org/10.48550/arXiv.2403.15734 (2024).

[4] Zhu, R., Nong, W., Yamazaki, S. & Hippalgaonkar, K. WyCryst: Wyckoff inorganic crystal generator framework. *Matter* **7**, 3469-3488 (2024).

[5] Zhu, R. *et al.* Predicting Synthesizability using Machine Learning on Databases of Existing Inorganic Materials. *ACS Omega* **8**, 8210–8218 (2023).