DIS-CSP: DISORDERED CRYSTAL STRUCTURE PREDICTIONS

Martin Hoffmann Petersen[a], **Ruiming Zhu** [b, c], **Haiwen Dai**[b,c], **Savyasanchi Aggarwal** [c,d], **Nong Wei** [b], **Andy Paul Chen** [b], **Arghya Bhowmik** [a], **Juan Maria Garcia Lastra** [a], **Kedar Hippalgaonkar** [b,c]

[a] *Technical University of Denmark, Department of Energy Conversion and Storage, Lyngby, 2800, Denmark*
mahpe@dtu.dk

[b] *School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore*
zhur0007@e.ntu.edu.sg, kedar@ntu.edu.sg

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

[d] *Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, U.K.*

1. Introduction

Most synthesized crystalline inorganic materials are compositionally disordered, meaning that multiple atoms occupy the same lattice site with partial occupancy. Moreover, the computed physical properties of disordered inorganic crystals are configuration dependent, because of this partial occupancy, making it extremely challenging to solve purely by computational methods: this makes property-oriented search impractical. Crystal structure prediction (CSP), for such crystals is crucial for the eventual development of highly efficient and stable functional materials. However, existing generative models cannot handle the complexities of disordered inorganic crystals. To address this gap, we introduce an equivariant representation, based on theoretical crystallography, along with a generative model capable of generating valid structures that allow for compositional disorder and vacancies, which we call Dis-CSP. We train Dis-CSP on experimental inorganic structures from the Inorganic Crystal Structure Database (ICSD), which is the world's largest database of identified inorganic crystal structures. We show that Dis-CSP can effectively generate disordered inorganic crystal materials while preserving the inherent symmetry of the crystals throughout the generation process.

2. Results and Discussion

2.1 Crystal Description

Crystals are highly structured materials defined by a repeated arrangement of atoms in space. This periodicity is captured within a unit cell, a fundamental building block that defines the crystal's structure [1]. Theoretical crystallography systematically describes crystals using lattice parameters, space groups, and Wyckoff sites. Space groups categorize the symmetry operations that preserve the crystal's structure, with 230 distinct groups identified in three-dimensional space. Within these groups, atomic positions are mapped through symmetry operations, enabling the reconstruction of the entire crystal from a single unit cell. Wyckoff sites further refine this description by defining equivalent atomic positions within a space group, classified based on symmetry and occupancy.

In the context of disordered inorganic crystals, structural information can be extracted from Crystallographic Information Files (CIFs) and represented in a matrix-vector format. The matrix encodes Wyckoff site properties, including partial occupancy, multiplicity, fractional coordinates, and symmetry labels, while the vector captures the lattice parameters and space group information. Previous approaches in materials representation, particularly in Variational Autoencoders (VAEs) [2] and transformer models [3], have leveraged Wyckoff sites and space groups but have not adequately addressed disordered structures with partial occupancy. In contrast, our representation, combined with the VAE model, provides a novel and previously unexplored framework for CSP, enabling a more comprehensive analysis of disordered crystals.

2.2 Model Performance and Materials Generation

We develop a VAE generative model for disordered inorganic crystals, trained on the ICSD dataset. The model encodes structural features using a convolution neural network (CNN)for atomic properties and an artificial neural network (ANN)for crystal parameters, learning a probabilistic distribution of disordered structures. The decoder reconstructs key crystallographic attributes, enabling the generation of novel disordered crystals for crystal structure prediction (CSP). The generative model achieves high accuracy in predicting disordered structure for both the ordered and disordered Wyckoff sites (Table 1). Using KDE estimation in the latent space, we generate disordered crystal structures, with four representative generations illustrated in Figure 1. The generated compositions align with known functional materials, including battery electrodes and catalytic compounds, while also suggesting new doping strategies.

2.3 Discussion and Conclusion

We introduce Dis-CSP, a generative framework that lays the foundation for discovery of stable, unique, novel and functional disordered inorganic materials, using the world's largest experimental inorganic crystal structure database (ICSD). Based on symmetry principles, our equivariant representation of the disordered crystal system explicitly accounts for partial occupancy, including dopants, defects and compositional disorder. While we use a VAE to demonstrate symmetry-compliant crystal generation, we can expand on this framework in future work towards property-oriented search.

2.4 Figures and tables

DIS-CSP: DISORDERED CRYSTAL STRUCTURE PREDICTIONS

Martin Hoffmann Petersen[a], **Ruiming Zhu** [b, c], **Haiwen Dai**[b,c], **Savyasanchi Aggarwal** [c,d], **Nong Wei** [b], **Andy Paul Chen** [b], **Arghya Bhowmik** [a], **Juan Maria Garcia Lastra** [a], **Kedar Hippalgaonkar** [b,c]

[a] *Technical University of Denmark, Department of Energy Conversion and Storage, Lyngby, 2800, Denmark*
mahpe@dtu.dk

[b] *School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore*
zhur0007@e.ntu.edu.sg, kedar@ntu.edu.sg

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

[d] *Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, U.K.*

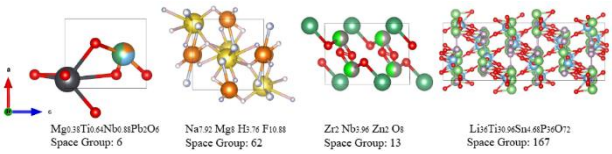

Fig. 1: Four generated disordered inorganic crystal structures, using Dis-CSP. All crystals are viewed along the b-axis of the crystal.

Table 1: Reconstruction errors of the test set for the parameters directly related to the disordered and ordered Wyckoff sites. The partial occupancy is presented with a MAE for the disordered sites, and an accuracy for the ordered sites. Accuracy is also used for the Wyckoff letter and Wyckoff multiplicity, while MAE is used for the fractional coordinate.

Parameters	Disordered	Ordered
Partial Occupancy	0.05 (MAE)	99.1%
Wyckoff Multiplicity	99.6%	99.8%
Wyckoff Letter	99.3%	99.6%
Frac. Corrdinate	0.07	0.08

Acknowledgments

A.B., J.M.G.L., and M.H.P. acknowledge support from the Det Frie Forskningsråd under Project ``Data-driven quest for TWh scalable \ce{Na}-ion battery (TeraBatt)'' (Ref. Number 2035-00232B). 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 NRFNRFF13-2021-0011.

References

[1] A. R. West, Solid State Chemistry and its Applications. 2014. doi: 10.1016/B978-0-444-51436-3.X5000-7.

[2] R. Zhu, W. Nong, S. Yamazaki, and K. Hippalgaonkar, “WyCryst: Wyckoff inorganic crystal generator framework,” Matter, pp. 1–20, 2024, doi: 10.1016/j.matt.2024.05.042.

[3] N. Kazeev et al., “Wyckoff Transformer: Generation of Symmetric Crystals,” 2025, [Online]. Available: http://arxiv.org/abs/2503.02407