Inverting unidentified X-ray Powder Diffraction Spectra through Machine Learning-Driven Prototype enumeration

Abhijith S Parackal[1], Rhys E. A. Goodall [2], Felix A. Faber[2], Rickard Armiento [1]
[1]Linköping University, Sweden
[2] University of Cambridge, United Kingdom

Contact: abhijith.s.parackal@liu.se

Abstract

X-ray Powder Diffraction (XRD) is a fundamental technique for determining crystal structures and is extensively used across a wide range of materials science disciplines [1]. Recovering atomic positions from a powder diffraction pattern presents significant challenges due to phase ambiguities and the averaging of signals from randomly oriented particles, leading to the existence of numerous synthesized materials for which the precise crystal structures remain unknown [2].
Most methods that aim to decode powder XRD data into crystal structures rely on refining atom positions within known structures to fit new data. While this method provides an opportunity to quickly explore and identify experimental XRD, the reliance on existing crystal structures limits the ability of such methods to discover structures that represent completely new prototypes yet to be observed in scientific literature. Thus, we propose a need to develop methods that do not rely on previously seen materials.
A practical approach to determining atomic positions from X-ray diffraction patterns involves navigating the space of all possible configurations of atoms in 3D. Techniques such as genetic algorithms or gradient descent are used to optimize the match between experimental XRD patterns and predicted stability from density functional theory (DFT). However, this process involves a complex search space with 3^N degrees of freedom, for N atoms in a unit cell
One can limit the search space by leveraging the inherent symmetry in crystal structures [3]. Identifying correct Wyckoff positions, which define specific symmetry-determined locations within a unit cell, can significantly reduce the number of free parameters. The challenge then becomes determining these Wyckoff positions for new materials without reference to previously known structures, effectively solving a crucial part of the puzzle.
In this work, we present a workflow to resolve powder diffraction data into crystal structures by systematically exploring the space of all possible atomic arrangements, i.e., structural prototypes, for a given space group using a pre-trained machine learning model named **Wren** [4]. Wren is trained to predict formation energies of a coarse-grained representation of crystal structure, represented using Wyckoff positions.
We first enumerate all Wyckoff assignments that can accommodate the given chemical composition within the selected spacegroup. Then, the elements assigned Wyckoff positions are ranked using Wren. Finally, rhe atoms assigned to Wyckoff positions are perturbed within their degrees of freedom to align the simulated X-ray diffraction (XRD) pattern with the experimental data. This iterative refinement continues until the simulated structure achieves an acceptable R-value fit. Structures that meet this criterion are further evaluated using density-functional theory (DFT) to confirm their thermodynamic stability.
The procedures for the workflow, named ***httk-symgen*** [5] are written in high-performance GPU-accelerated Python code, ensuring the process is scalable for high-throughput applications. We explore the workflow on the RRUFF dataset [6] and unidentified XRD spectra from the ICDD database and show that it can be used to find crystal structures with entirely new symmetric arrangements of atoms by identifying at least two structures with unseen prototype assignments.

References

[1] Massa, Werner, and Robert O. Gould. Crystal structure determination. Berlin: Springer, 2004.
[2] Harrison, Robert W. "Phase problem in crystallography." JOSA a 10.5 (1993): 1046-1055.
[3] Meredig, Bryce, and C. Wolverton. "A hybrid computational–experimental approach for automated crystal structure solution." Nature materials 12.2 (2013): 123-127.
[4] R. E. A. Goodall, A. S. Parackal, F. A. Faber, R. Armiento, and A. A. Lee, Sci. Adv. 8, eabn4117 (2022)
[5] Parackal, Abhijith S., et al. "Identifying Crystal Structures Beyond Known Prototypes from X-ray Powder Diffraction Spectra." arXiv e-prints (2023): arXiv-2309.
[6] Lafuente, B., et al. "The power of databases: The RRUFF project, 2016." B., Highlights in Mineralogical Crystallography: 1-30.

Figures

Figure 1. Experimental XRD resolved using *httk-symgen*. Blue line shows the experimentally obtained peaks fitted over a pseudo-voigt function. Yellow-dashed lines show the XRD profile generated using structures generated using our workflow

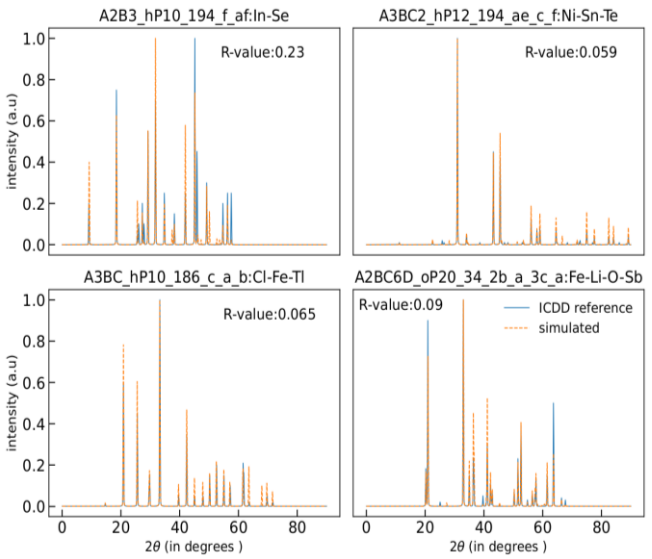

Figure 2. *httk-symgen* workflow for automated structure discovery from powder diffraction data.

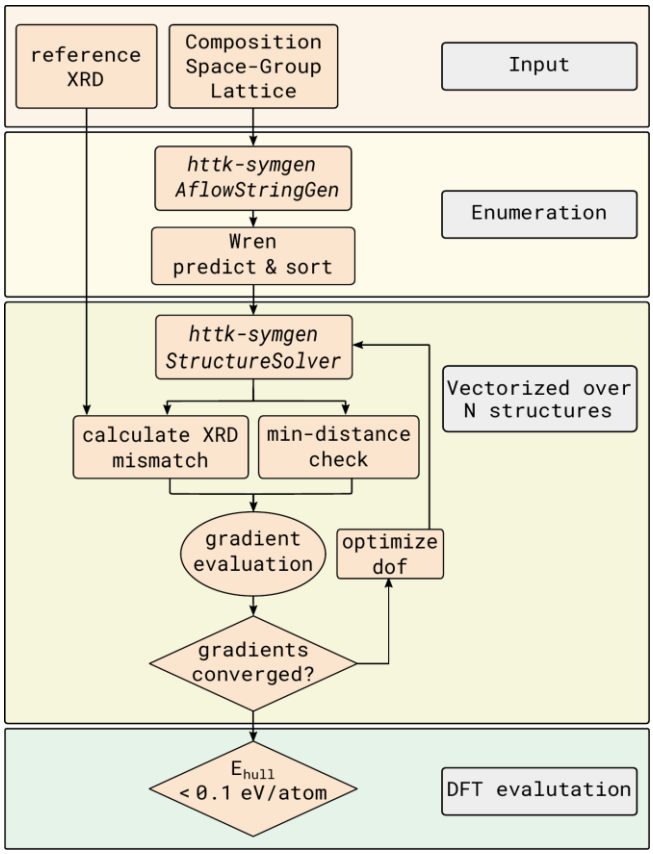

Figure 3. Each row corresponds to an entry in RRUFFF. First, all the candidate structures identified by our workflow are shown, then the crystal structure obtained from RRUFF project. Finally, a comparison of the synthetic XRD from our best match and the experimental XRD provided in RRUFF is shown.