Tackling real-world crystal structure prediction from powder X-ray diffraction data

A. F. Sapnik [1], F. L. Johansen[1], K. M. Ø. Jensen[1]

1 University of Copenhagen (Denmark).

afs@chem.ku.dk

Determining crystal structures directly from powder X-ray diffraction (PXRD) remains challenging because peak overlap, finite crystallite size, disorder, background, and instrumental artefacts can make the inverse problem underdetermined. Machine-learning models offer new routes to PXRD-based crystal structure prediction, but many are trained and benchmarked on idealised simulated data, leaving their reliability on real experimental patterns uncertain. Here, we evaluate the real-world behaviour of deCIFer, a PXRD-conditioned autoregressive transformer that generates candidate crystal structures in CIF format.[1] Using controlled robustness tests, we quantify how its predictions respond to realistic artefacts, including noise, background, peak asymmetry, unit-cell shifts, instrumental broadening, Scherrer broadening, and missing-*Q* regions, benchmarking performance against an unconditioned baseline.[2]

deCIFer improves structure prediction when the PXRD pattern retains sufficiently resolved and informative features, tracking lattice changes and recovering simple, highly crystalline structures such as Si and CeO_2. However, for lower-symmetry systems and nanocrystalline samples, where peak overlap and broadening reduce the structural information content, the benefit of PXRD conditioning decreases and the model expresses greater uncertainty through broader predicted lattice distributions. Tests using experimental PXRD data for Si, CeO_2, Fe_2O_3, and nanocrystalline CeO_2 reproduce these trends. Overall, our results show that ML-based structure prediction from PXRD is limited by the information content of the diffraction data but can serve as a valuable human-in-the-loop tool for rapidly generating plausible structural candidates, quantifying uncertainty, and guiding downstream refinement.

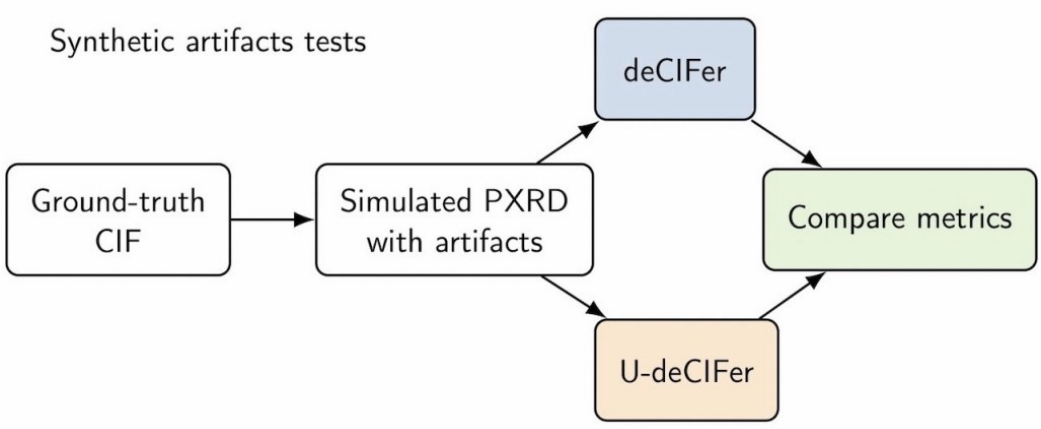


Figure 1. *Synthetic-artefact tests: controlled distortions (noise, asymmetry, peak broadening) are applied to simulated PXRD from known structures and evaluated before experimental PXRD tests are applied.*

[1] F. L. Johansen et al., arXiv, DOI: 10.48550/arXiv.2502.02189 (2025)
[2] F. L. Johansen & A. F. Sapnik et al., ChemRxiv, DOI: 10.26434/chemrxiv.15000236/v2 (2026).