Extracting Structural Motifs from Pair Distribution Function Data of Nanostructures using Interpretable Machine Learning

Andy S. Anker[1], Emil T. S. Kjær[1], Mikkel Juelsholt[1], Troels Lindahl Christiansen[1], Susanne Linn Skjærvø[1], Nicolas P. L. Magnard[1], Mads Ry Jørgensen[2], Innokenty Kantor[2], Daniel Risskov Sørensen[2], Raghavendra Selvan[3] and Kirsten M. Ø. Jensen[1]

1: Nano-Science Center and Department of Chemistry, University of Copenhagen

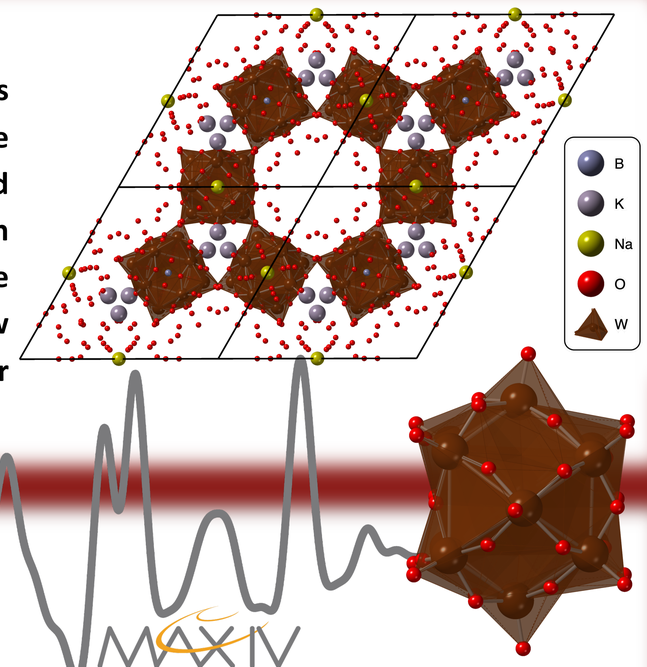
2: DanMAX, MAX-IV Laboratory, Lund, Sweden

3: Department of Computer Science and Department of Neuroscience, University of Copenhagen

*andy@chem.ku.dk

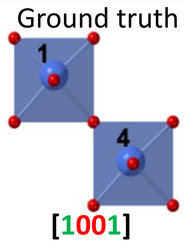
1: Introduction

Pair Distribution Function (PDF) is widely used as a method for characterization of materials without long-range order, such as nanomaterials[1] and disordered materials[2] where conventional crystallographic approaches fail.[3] A major challenge in structure analysis of these materials is to find a good starting model for the atomic structure. Currently, such starting models are often found by considering e.g., the structure of well-known, related bulk materials, based on chemical knowledge of the system. Such an approach can introduce human bias and limit the number of models to be tested. Recently, automated methods such as 'structure mining' and 'cluster mining' have therefore appeared in the literature to overcome this challenge.[4, 5] Here, we introduce a new approach using machine learning (ML) to evaluate results from automated modelling. Our Machine Learning based Motif Extractor (ML-MotEx), can automatically extract structural motifs from PDF data in semi-real experimental time without human bias.



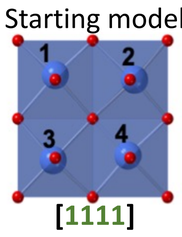
2: How does ML-MotEx work?

Ground truth

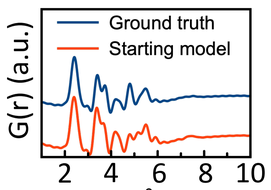


[1001]

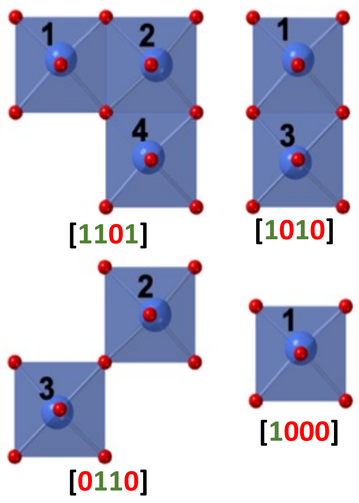
Starting model



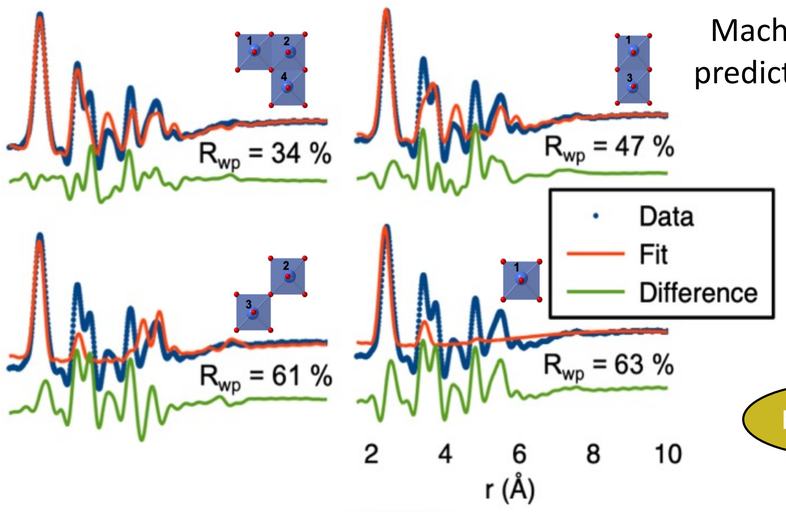
[1111]



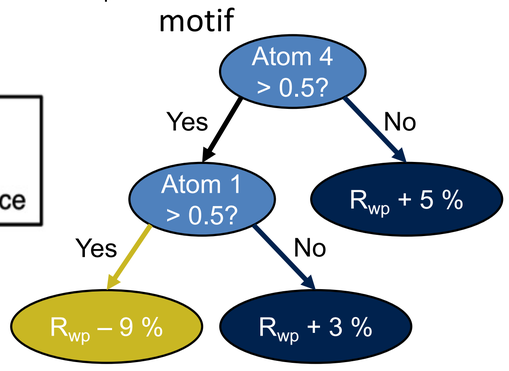
Step 1:
Catalogue of structural motifs



Step 2:
Fits of structural motifs to dataset



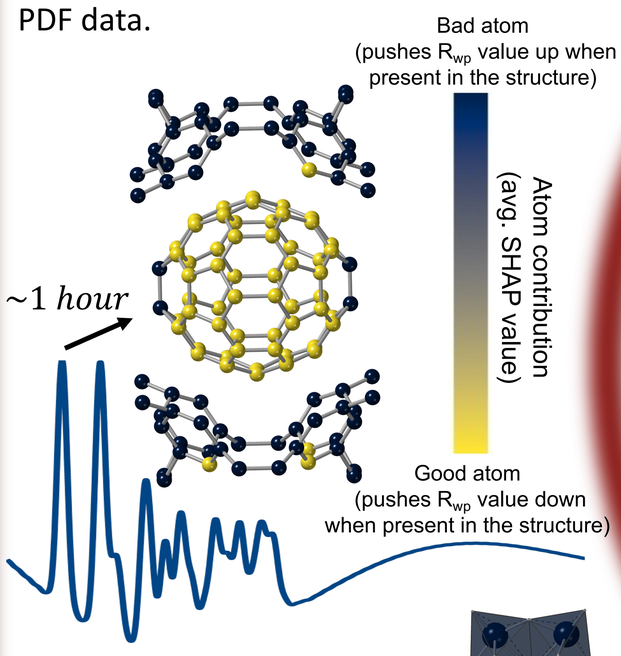
Step 3:
Machine Learning (XGBoost) to predict R_{wp} values from structural motif



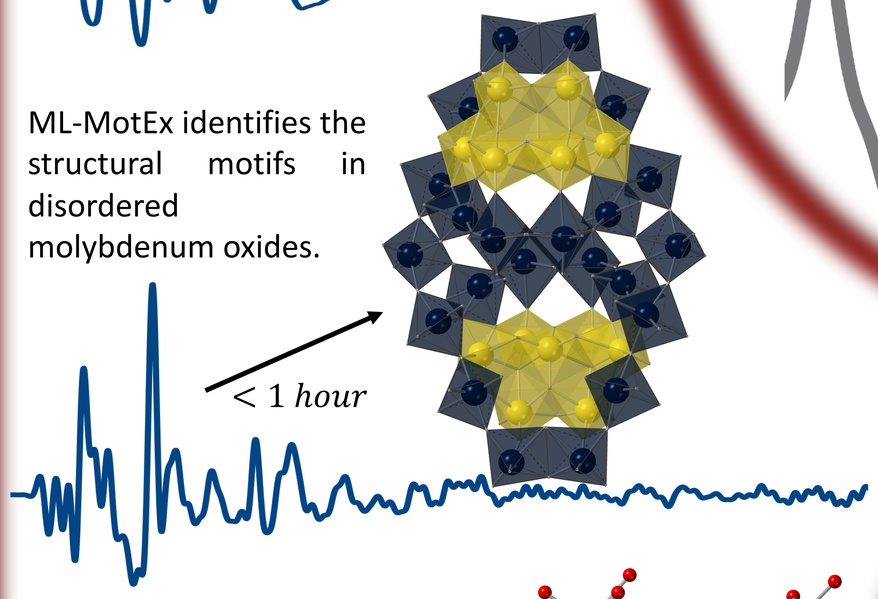
Step 4:
Calculate quantified (SHAP) values of atom importance for the fit quality

3: Extracting motifs with ML-MotEx

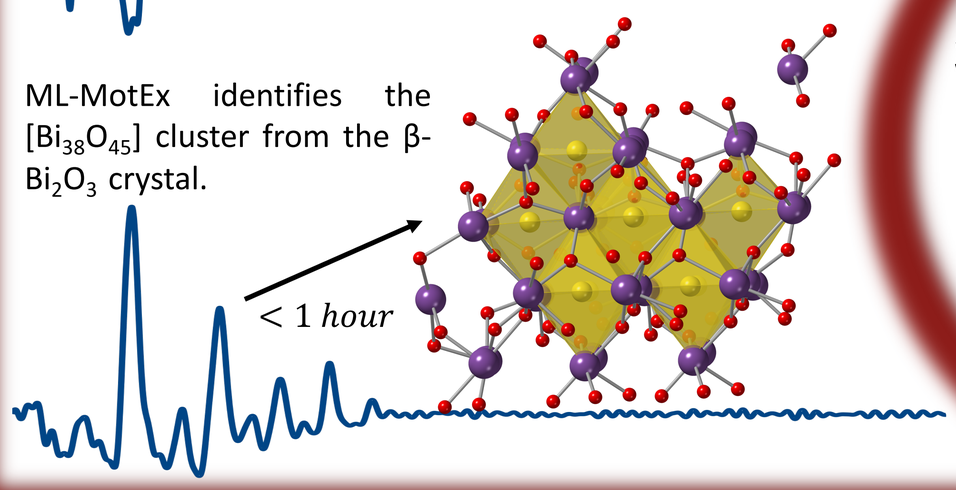
ML-MotEx extracts the C_{60} buckyball from a crystalline structure[6] based on a simulated PDF data.



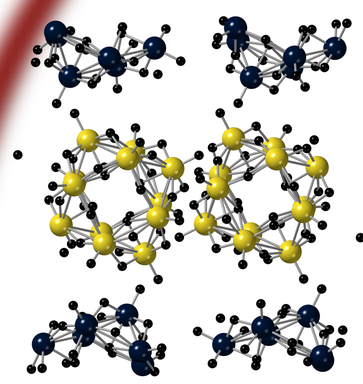
ML-MotEx identifies the structural motifs in disordered molybdenum oxides.



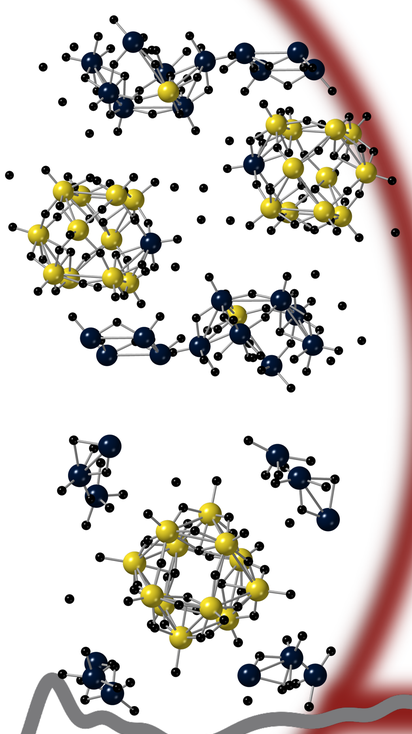
ML-MotEx identifies the $[Bi_{38}O_{45}]$ cluster from the β-Bi_2O_3 crystal.



4: PDF data from DanMAX, MAX IV

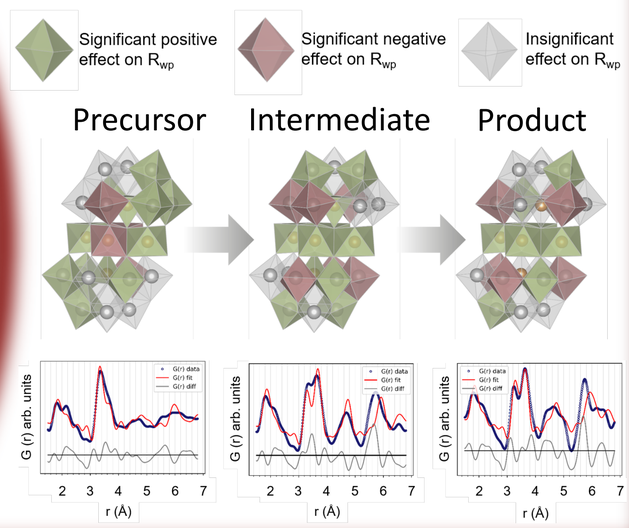


PDF data from DanMAX (MAX IV) was used to extract α-Keggin clusters from several different crystalline structures within 2 min.



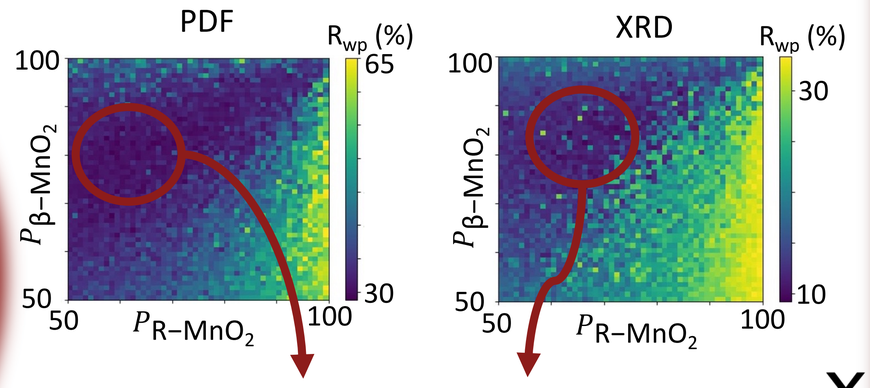
5: Using ML-MotEx to analyse *in situ PDF data*

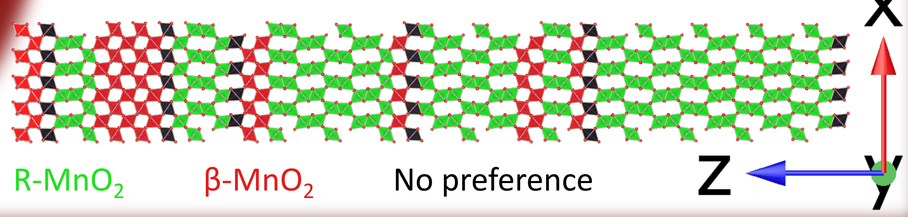
ML-MotEx can be used on each frame of an *in situ* dataset to follow the structural evolution of the reaction.



6: Using ML-MotEx to identify stacking faults from PXRD and PDF data

Both PDF and X-ray Powder-Diffraction data can be used with ML-MotEx to automatically identify stacking fault domains.





R-MnO_2 β-MnO_2 No preference

Acknowledgments

This work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, (Grant agreement no. 804066). We furthermore thank DANSCATT (supported by the Danish Agency for Science and Higher Education) for support.

References

[1] Billinge, S. J. L. et al., *The Problem with Determining Atomic Structure at the Nanoscale,* Science **2007**

[2] Keen, D. A. et al., *The crystallography of correlated disorder,* Nature **2015**

[3] Billinge, S. J. L. et al., *Beyond crystallography: the study of disorder, nanocrystallinity and crystallographically challenged materials with pair distribution functions,* Chem. Commun. **2014**

[4] Yang, L. et al, *Structure-mining: screening structure models by automated fitting to the atomic pair distribution function over large numbers of models,* Acta Cryst. **2020**

[5] Banerjee, S. et al., *Cluster-mining: an approach for determining core structures of metallic nanoparticles from atomic pair distribution function data,* Acta Cryst. **2020**

[6] Chen, X. & Yamanaka, S., *Single-crystal X-ray structural refinement of the `tetragonal' C_{60} polymer*. Chem. Phys. Lett. **2002**