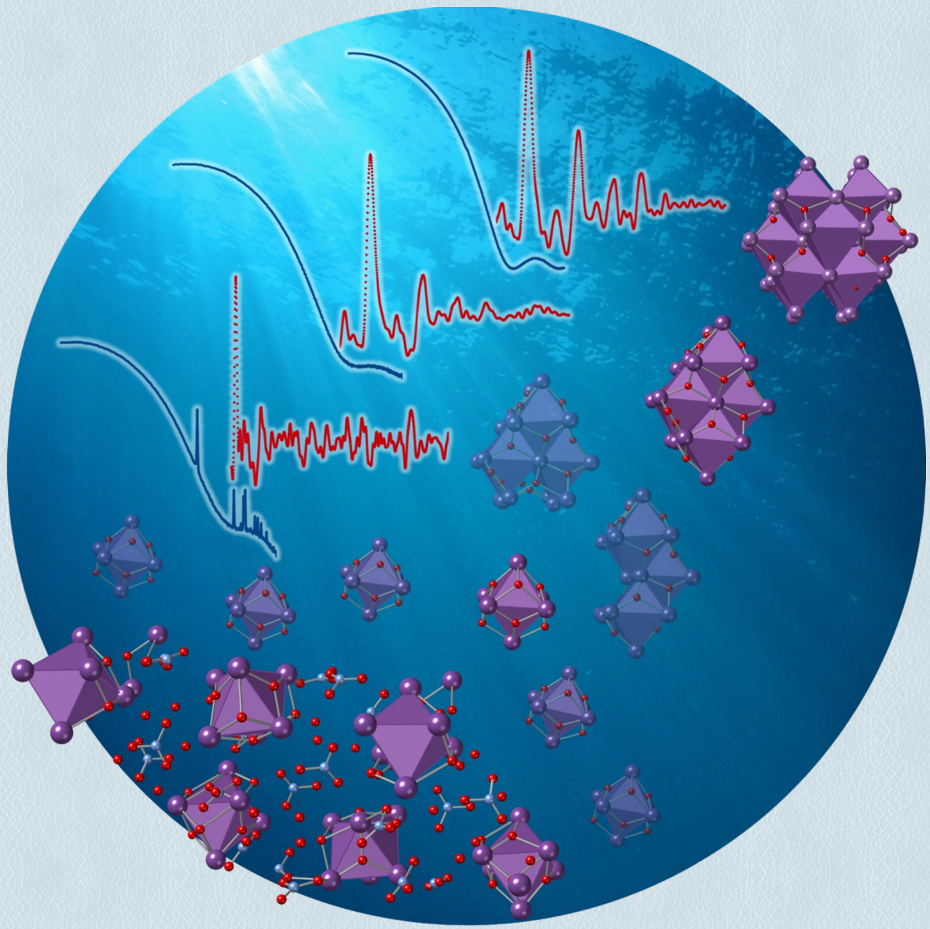
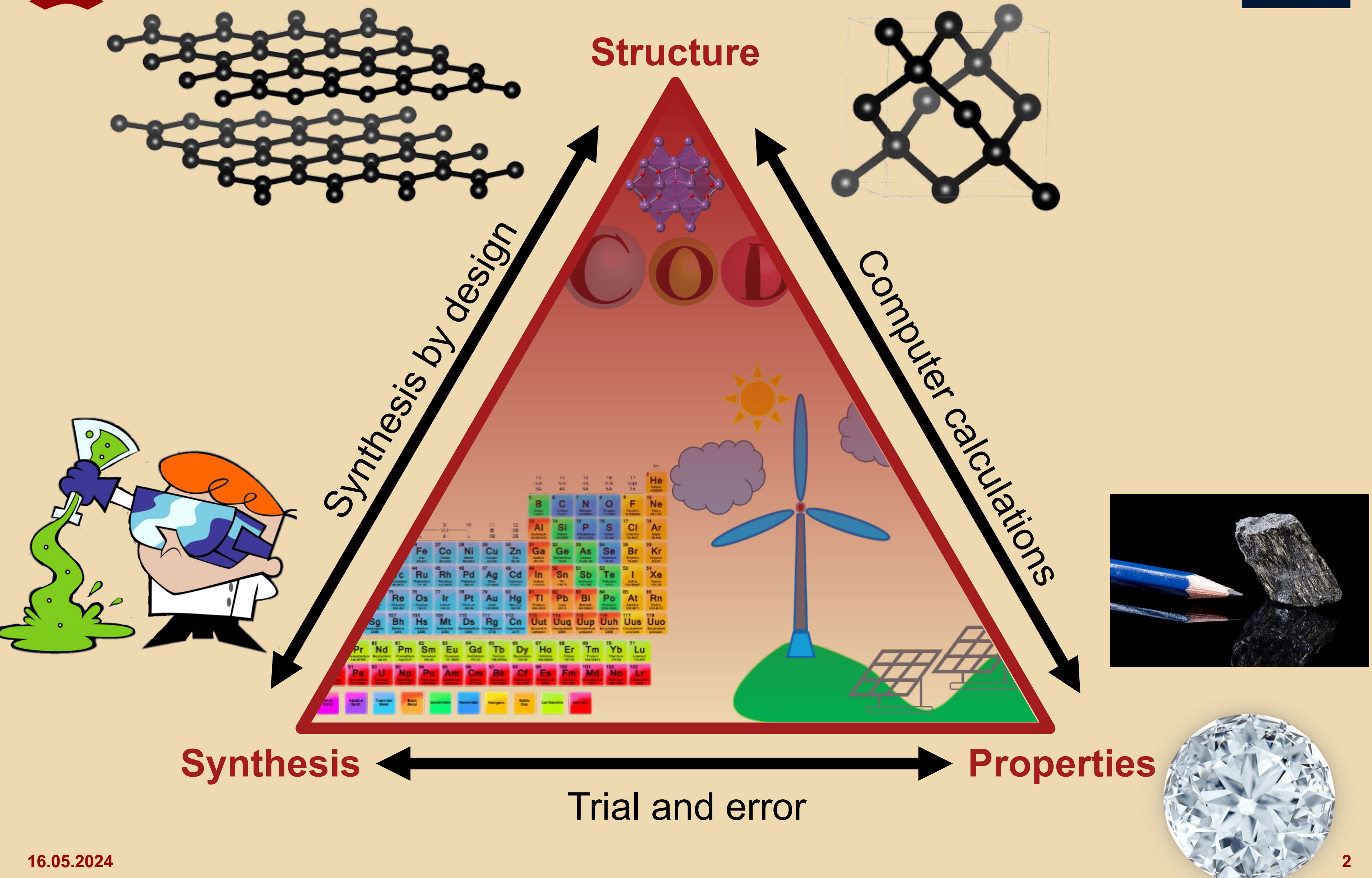
Machine learning for analysis of experimental scattering data in materials chemistry



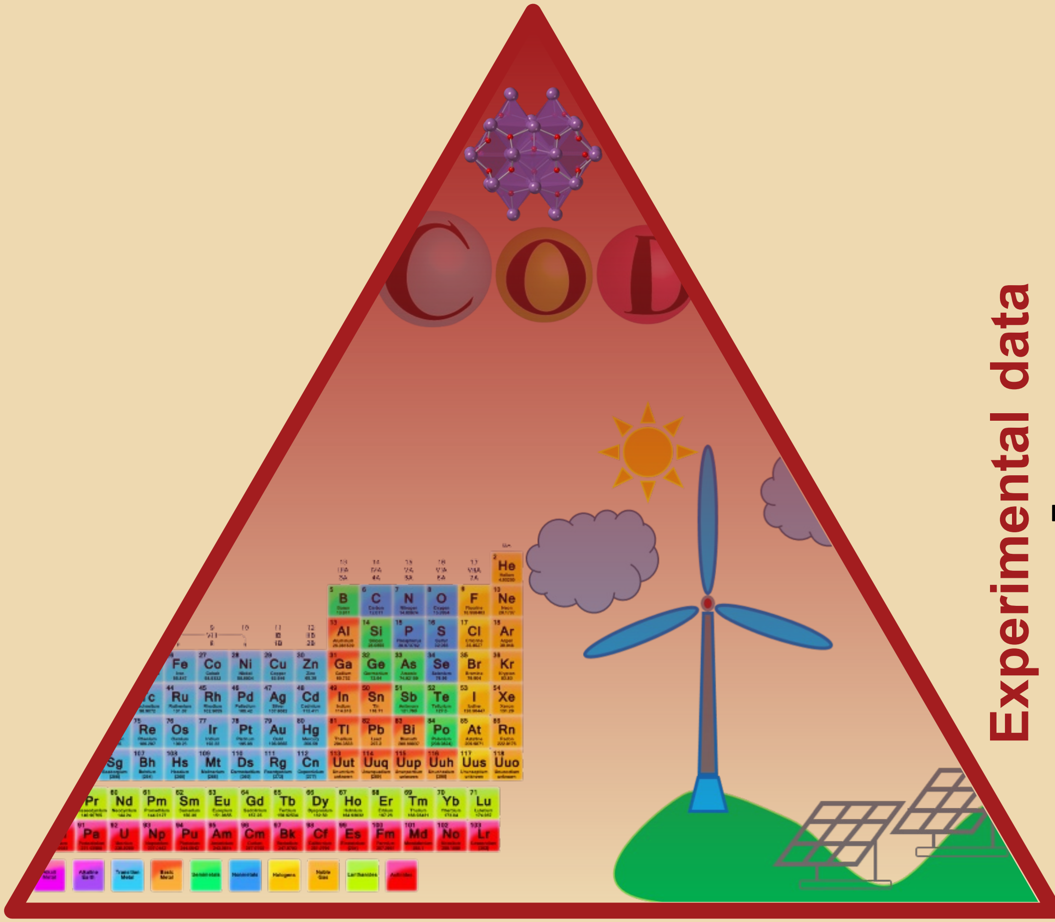
Andy Sode Anker

The Holy Grail of Materials Chemistry



The Holy Grail of Materials Chemistry

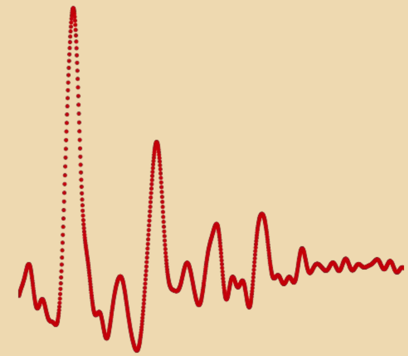
Structure



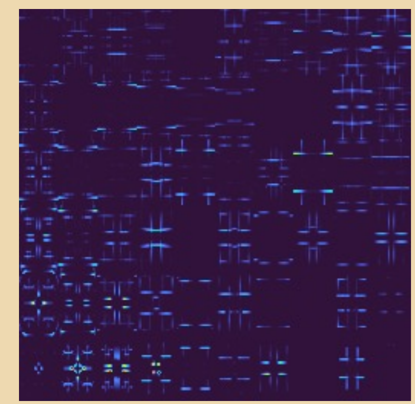
Synthesis

Properties

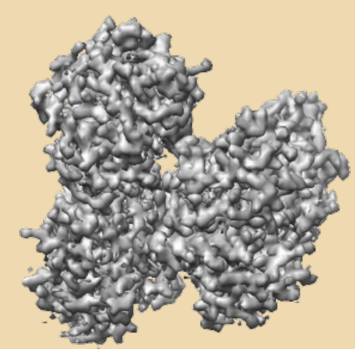
Experimental data



Scattering

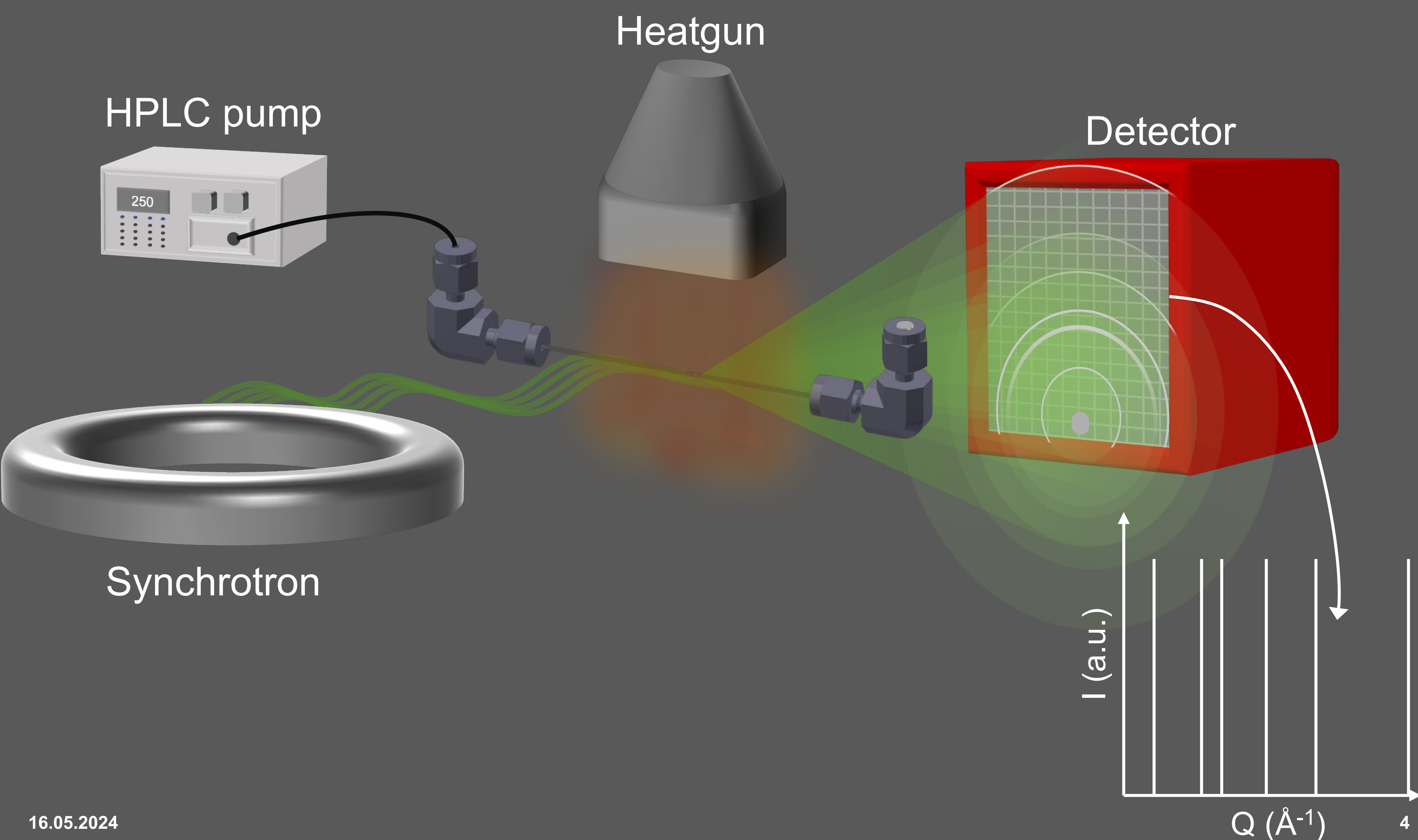


Spectroscopy



Imaging

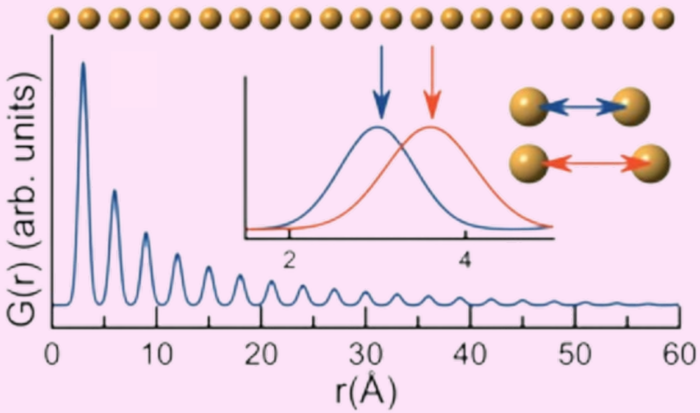
How to characterise the atomic structure



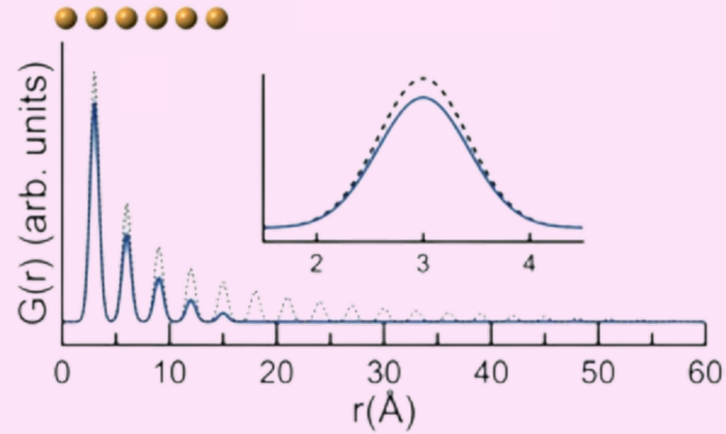
What is the PDF?

Pair distribution function (PDF) represents a histogram of interatomic distances

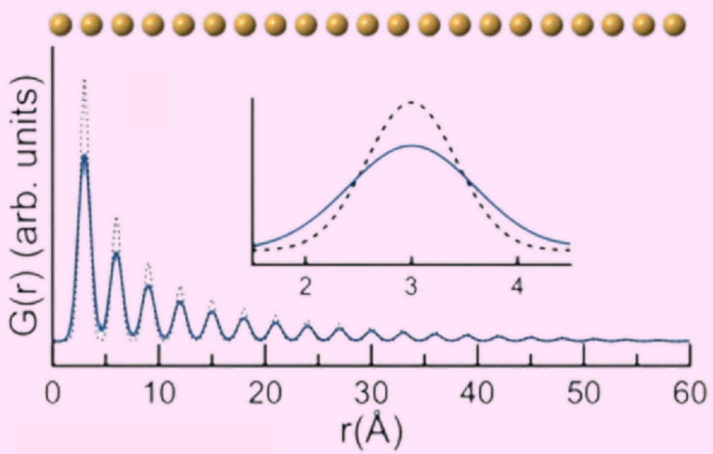
Atomic coordinates



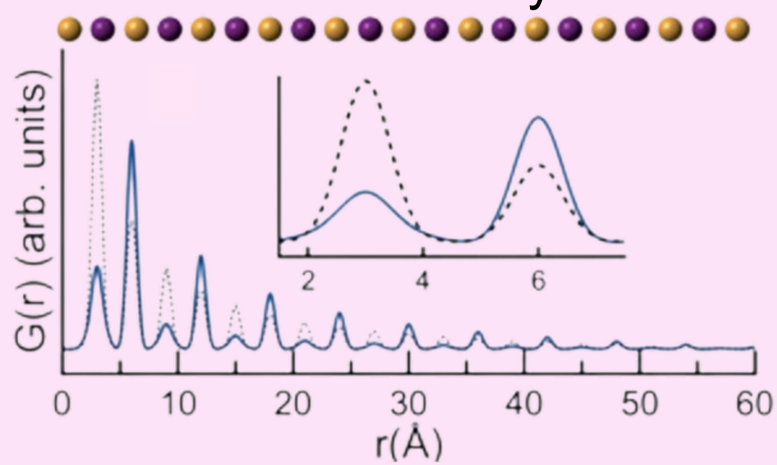
Size of material



Atomic vibrations

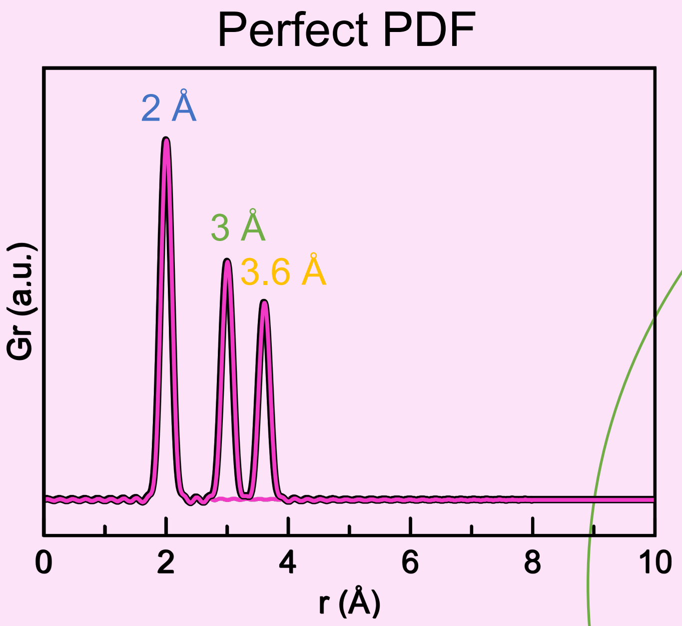


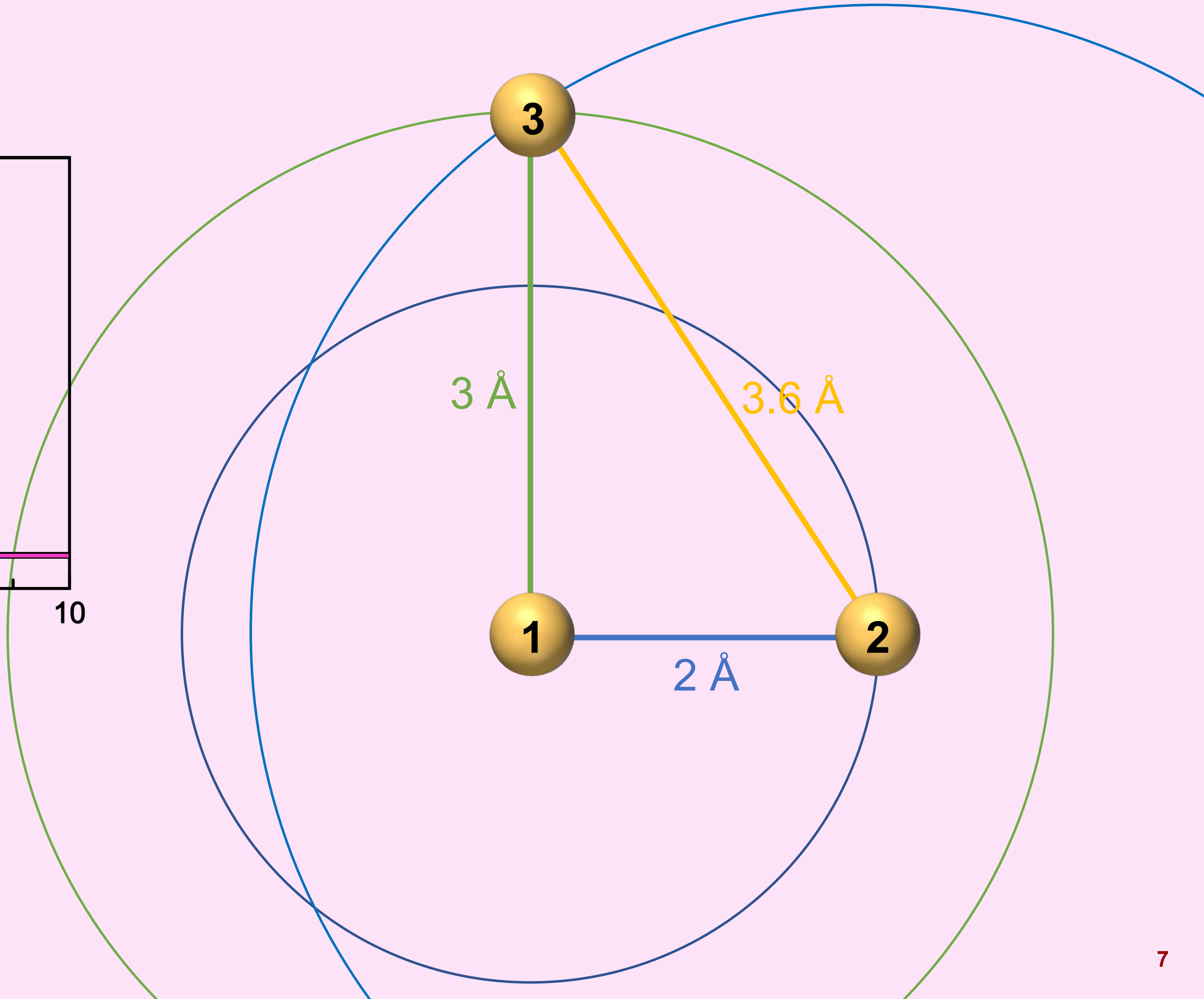
Atomic identity



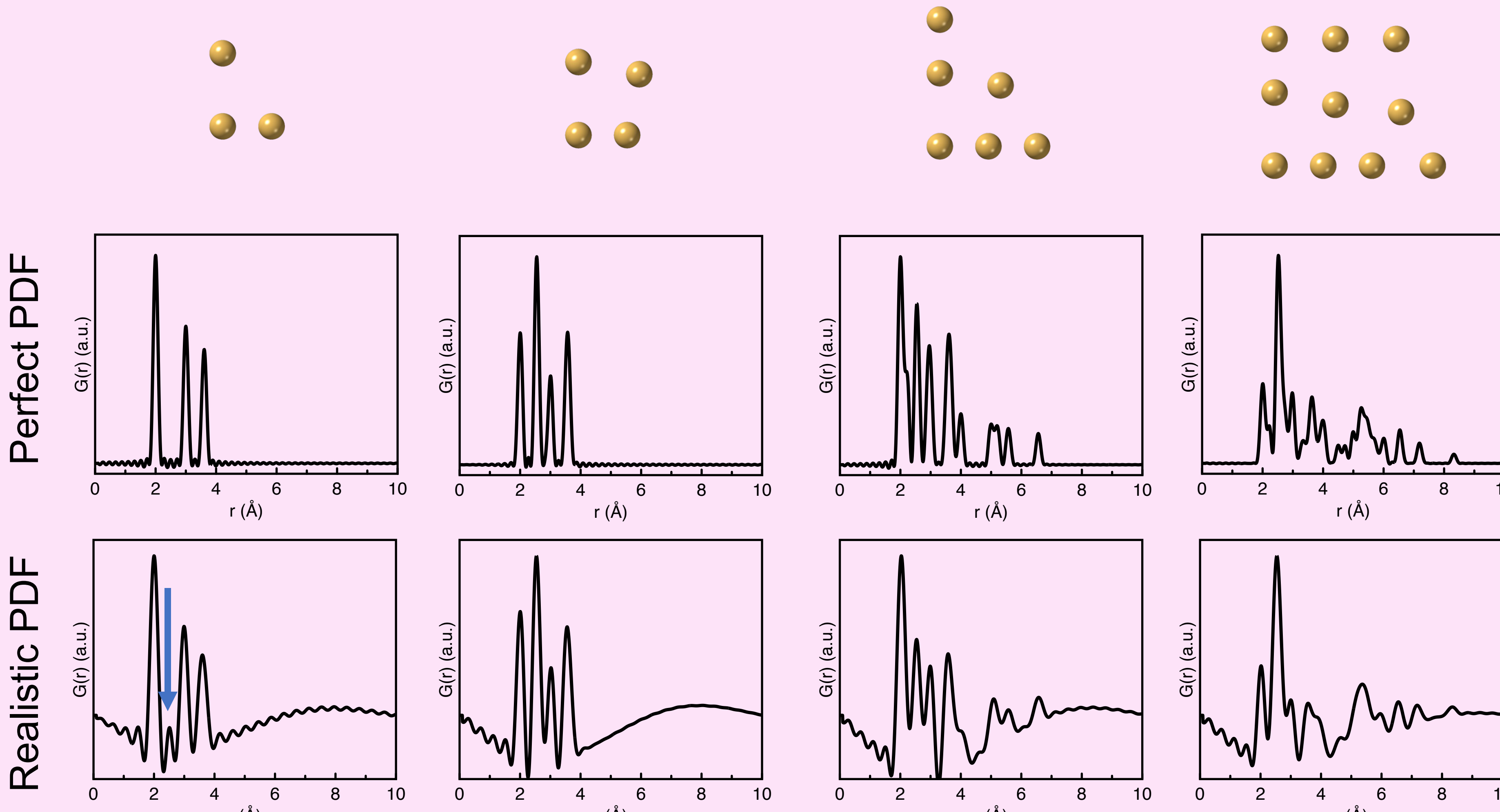
- There's no place like real-space: elucidating size-dependent atomic structure of nanomaterials using pair distribution function analysis, Christiansen T. L. & Cooper S. R., Jensen K. M. Ø., Nanoscale Adv. (2020)

Solving the PDF using triangulation

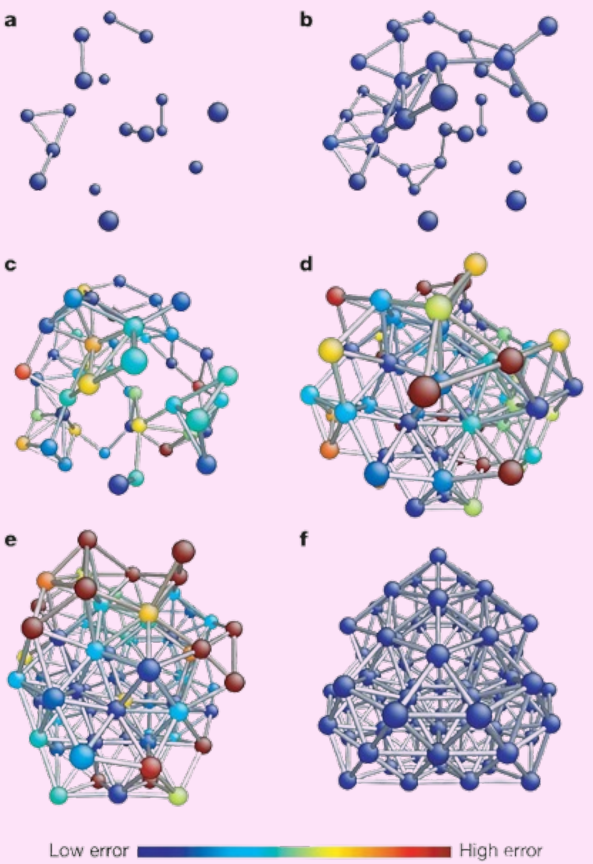


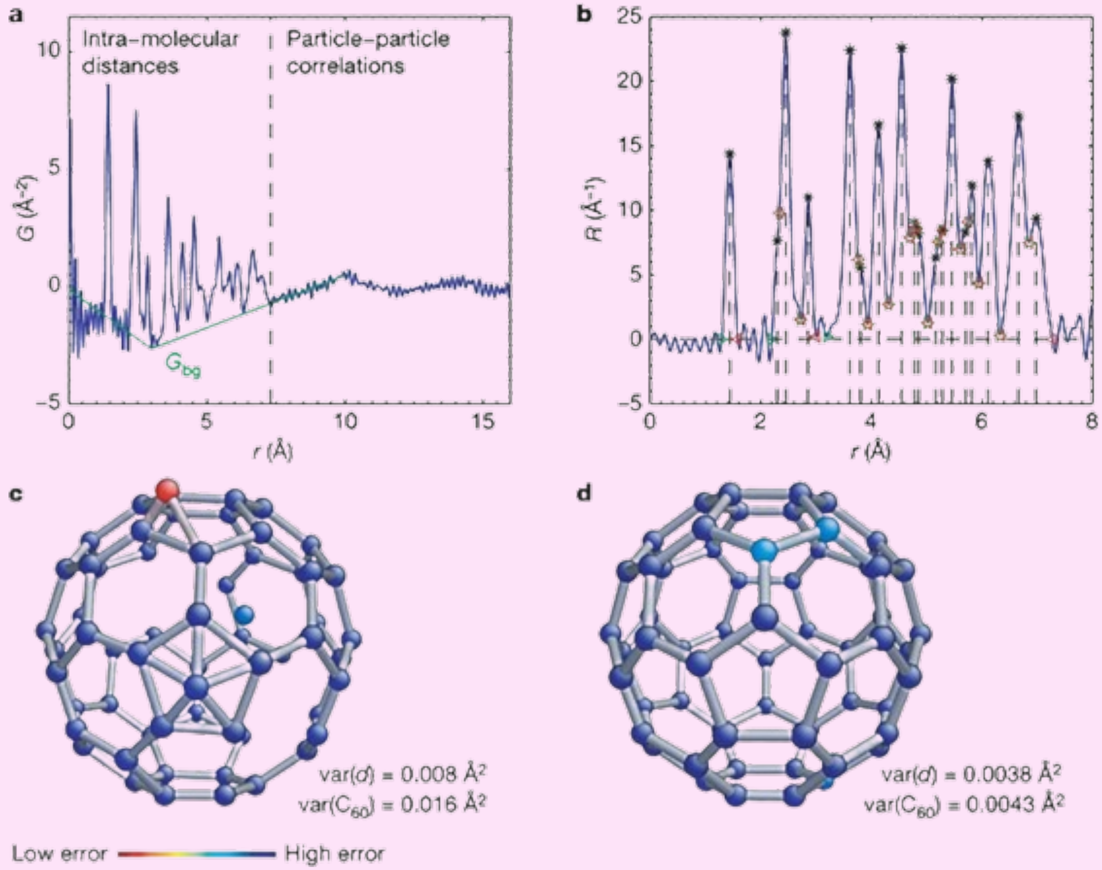


Solving the PDF using triangulation

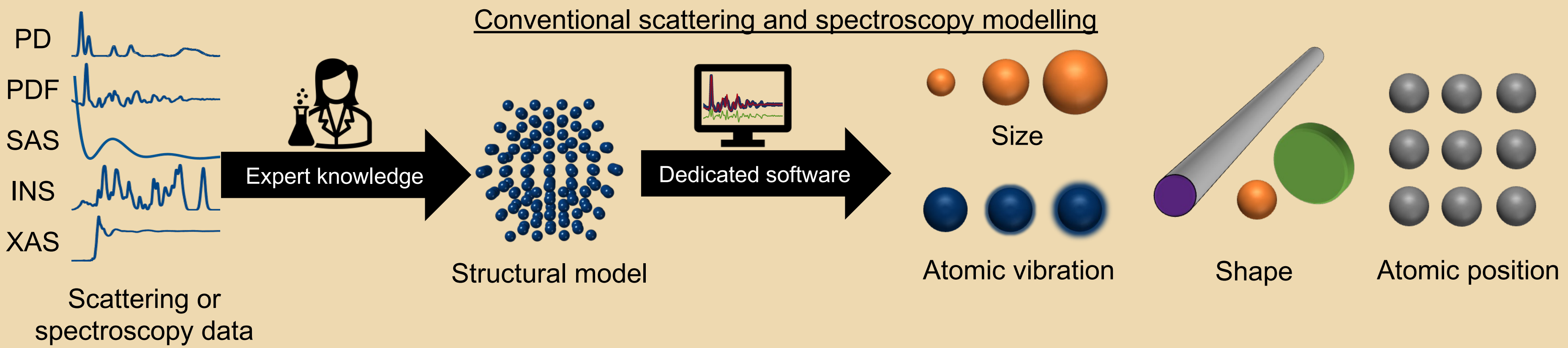


Ab initio structure solution with the LIGA algorithm

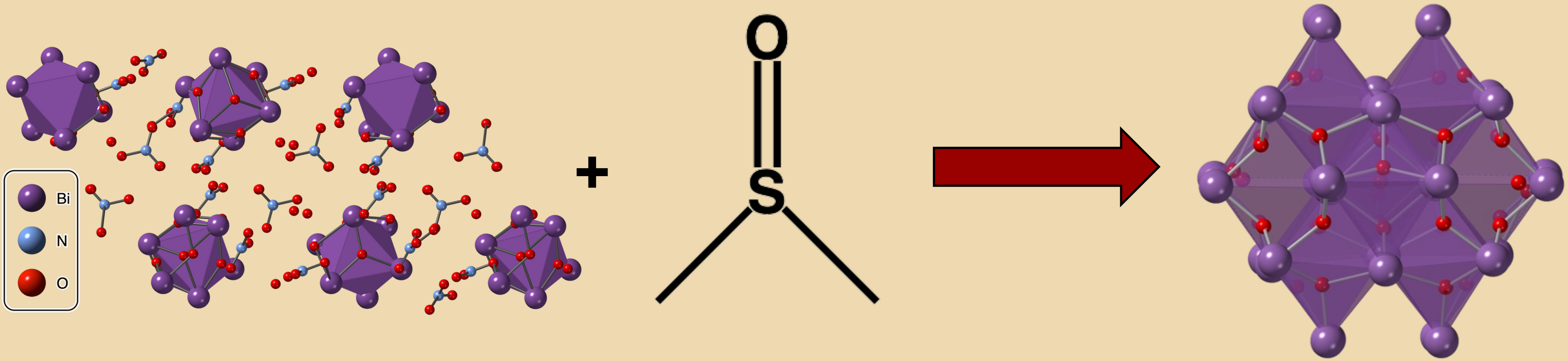




Conventional structure refinement



Structural changes during the growth of atomically precise metal oxido nanoclusters



$[Bi_6O_5(OH)_3(NO_3)_5]\cdot(H_2O)_3$ crystals

DMSO

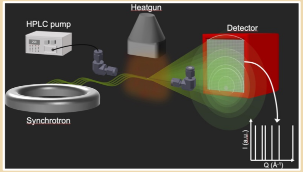
$[Bi_{38}O_{45}]$ cluster

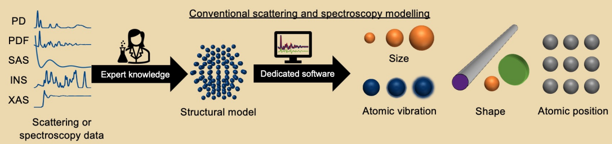
- Structural Changes during the Growth of Atomically Precise Metal Oxido Nanoclusters from Combined Pair Distribution Function and Small-Angle X-ray Scattering Analysis, Anker A. S., et al., *Angewandte Chemie* (2021) **(Back cover)**

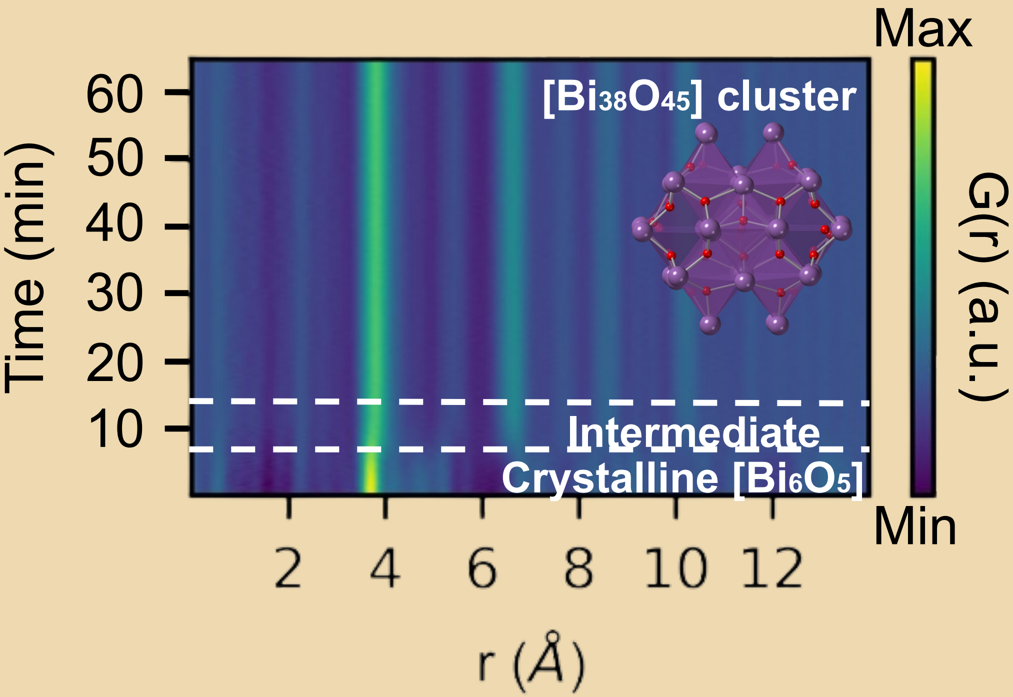
In situ PDF of crystalline $[Bi_6O_5]$ in DMSO solution

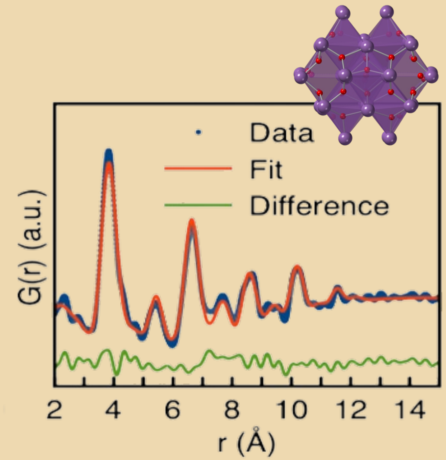
DTU

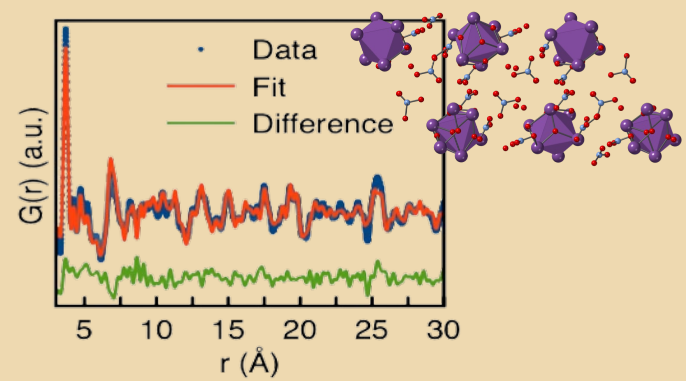
UNIVERSITY OF OXFORD





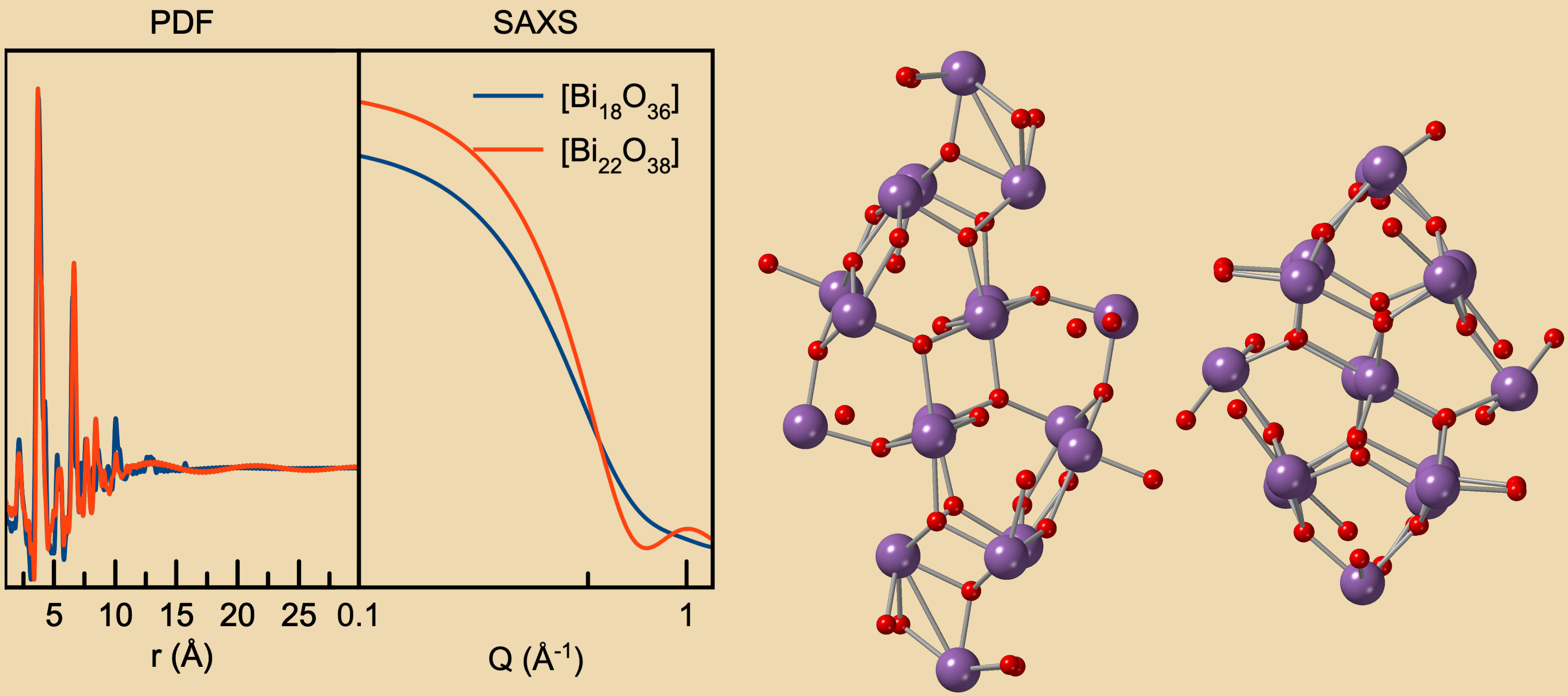




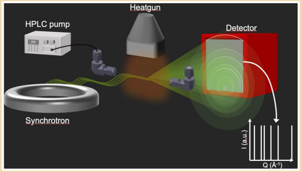


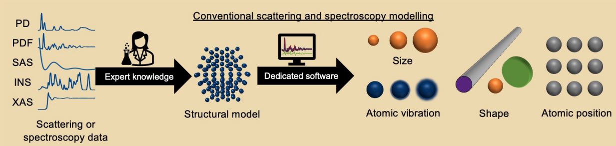
➢ Structural Changes during the Growth of Atomically Precise Metal Oxido Nanoclusters from Combined Pair Distribution Function and Small-Angle X-ray Scattering Analysis, Anker A. S., et al., *Angewandte Chemie* (2021) **(Back cover)**

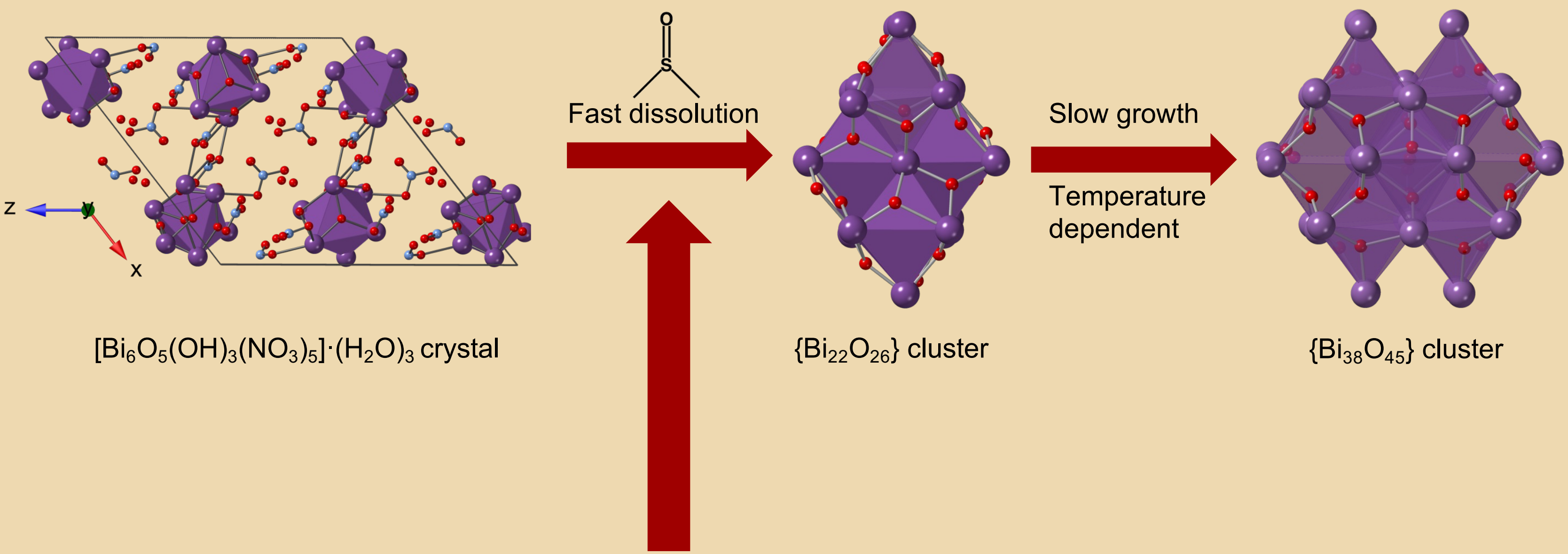
In situ PDF of crystalline $[Bi_6O_5]$ in DMSO solution



In situ PDF of crystalline $[Bi_6O_5]$ in DMSO solution







$[Bi_6O_5(OH)_3(NO_3)_5]\cdot(H_2O)_3$ crystal

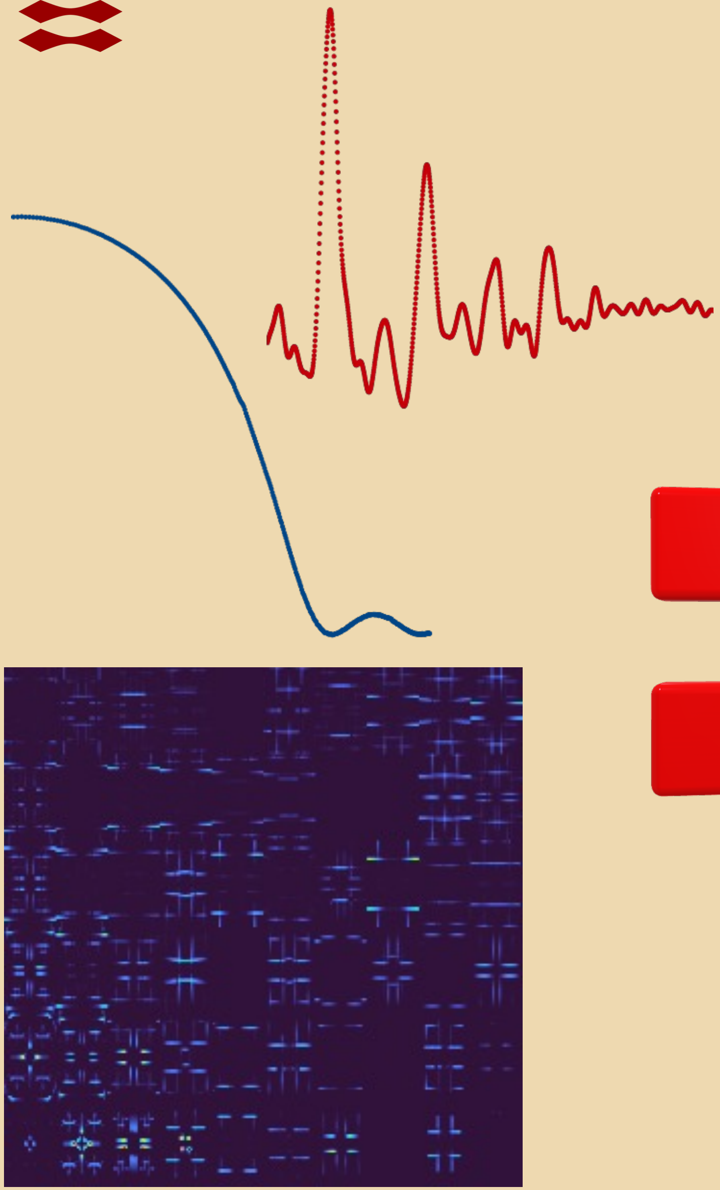
$\{Bi_{22}O_{26}\}$ cluster

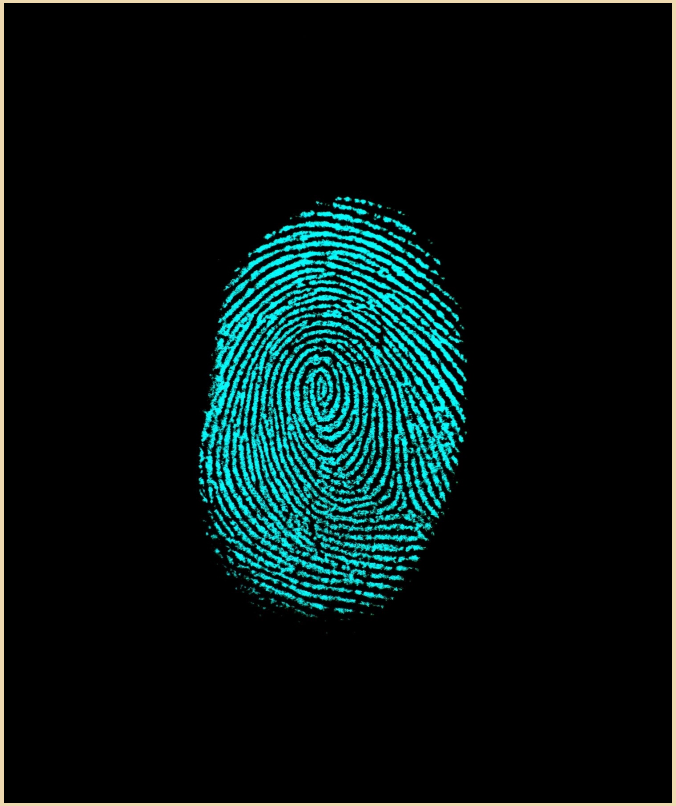
$\{Bi_{38}O_{45}\}$ cluster

What is happening here??

- Structural Changes during the Growth of Atomically Precise Metal Oxido Nanoclusters from Combined Pair Distribution Function and Small-Angle X-ray Scattering Analysis, Anker A. S., et al., *Angewandte Chemie* (2021) **(Back cover)**

Data is like a fingerprint



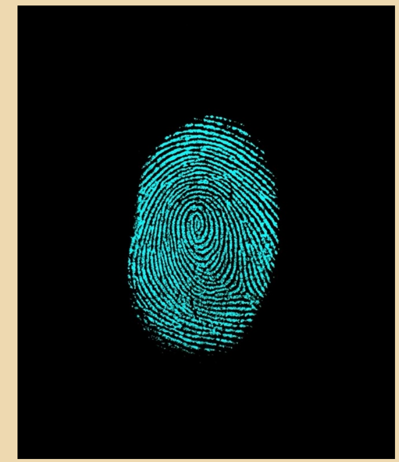


Fingerprint matching

The brute-force approach

Fingerprint database

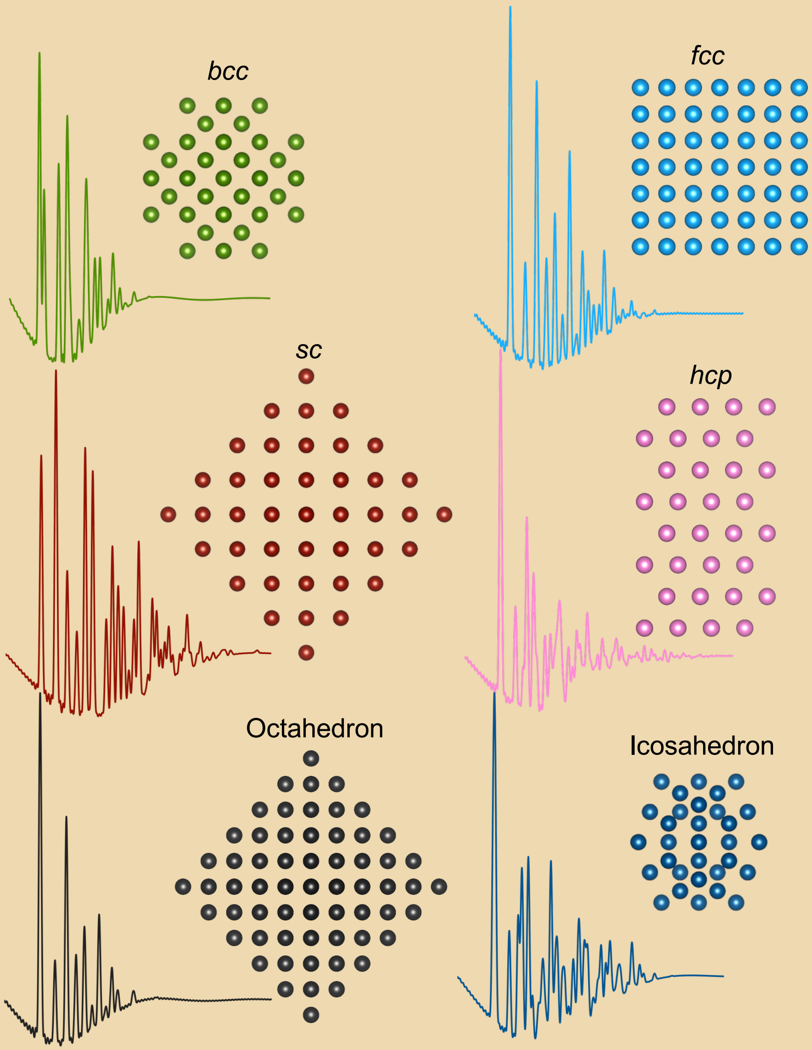
Measured fingerprint



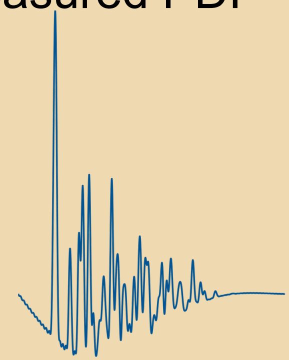
Scattering pattern matching

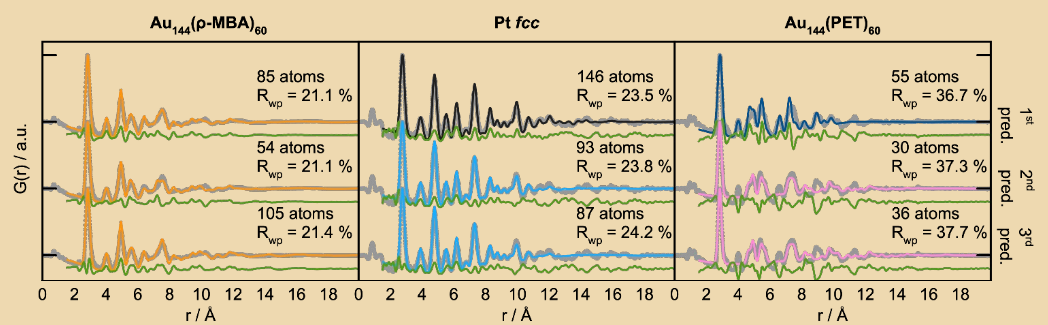
The brute-force approach

Structure database



Measured PDF





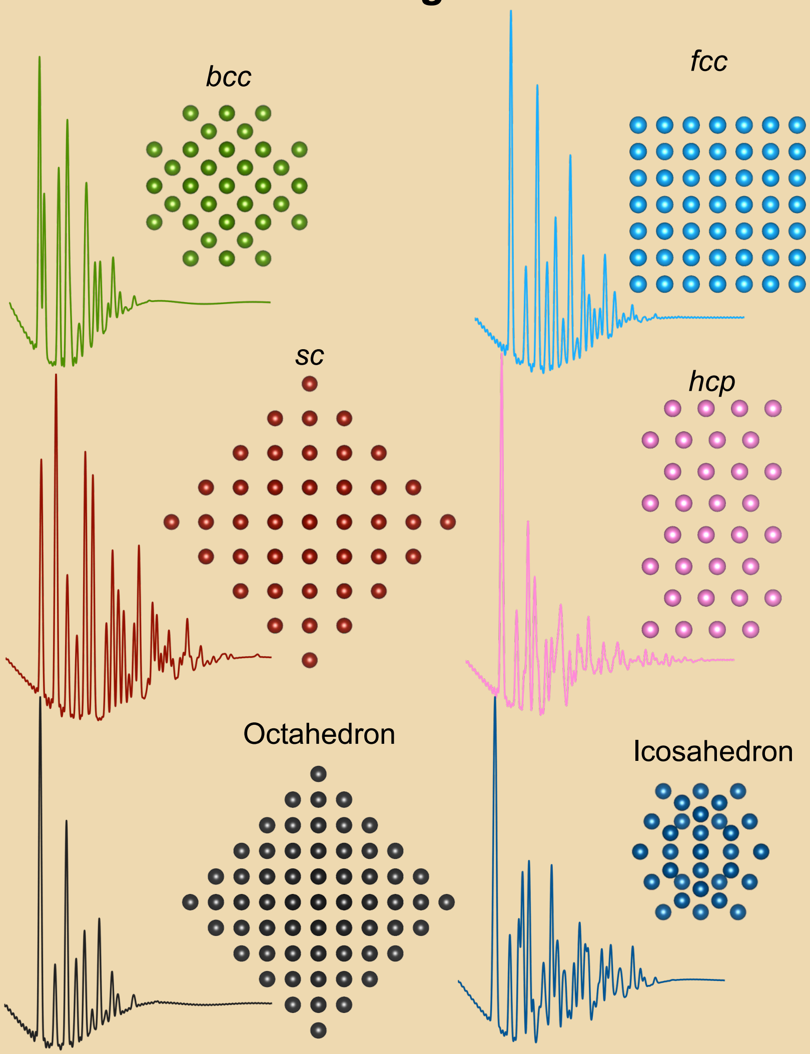
~14.5 hrs. / 3742 structures

- Cluster-mining: an approach for determining core structures of metallic nanoparticles from atomic pair distribution function data, BanerJee S., et al., *Acta Cryst. A* (2020)
- DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., *Digital Discovery* (2023) **(Front cover)**

Scattering pattern matching

The brute-force approach

Training set

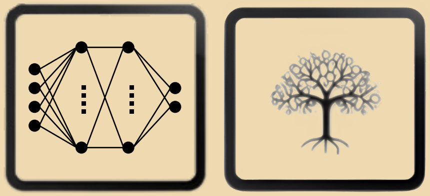


~1 sec

96 % on top 3

~20.5 hrs. / 3742 structures

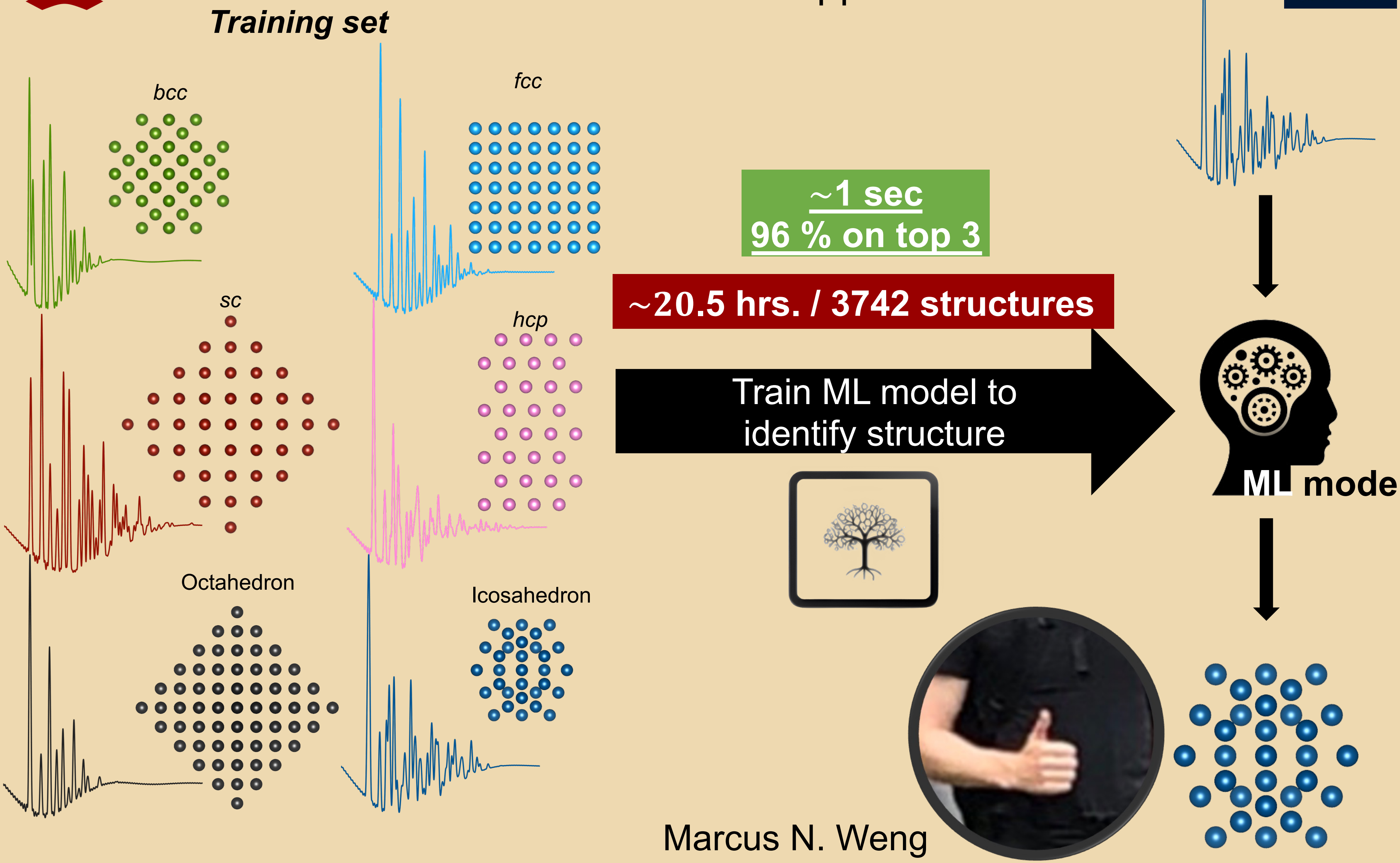
Train ML model to identify structure



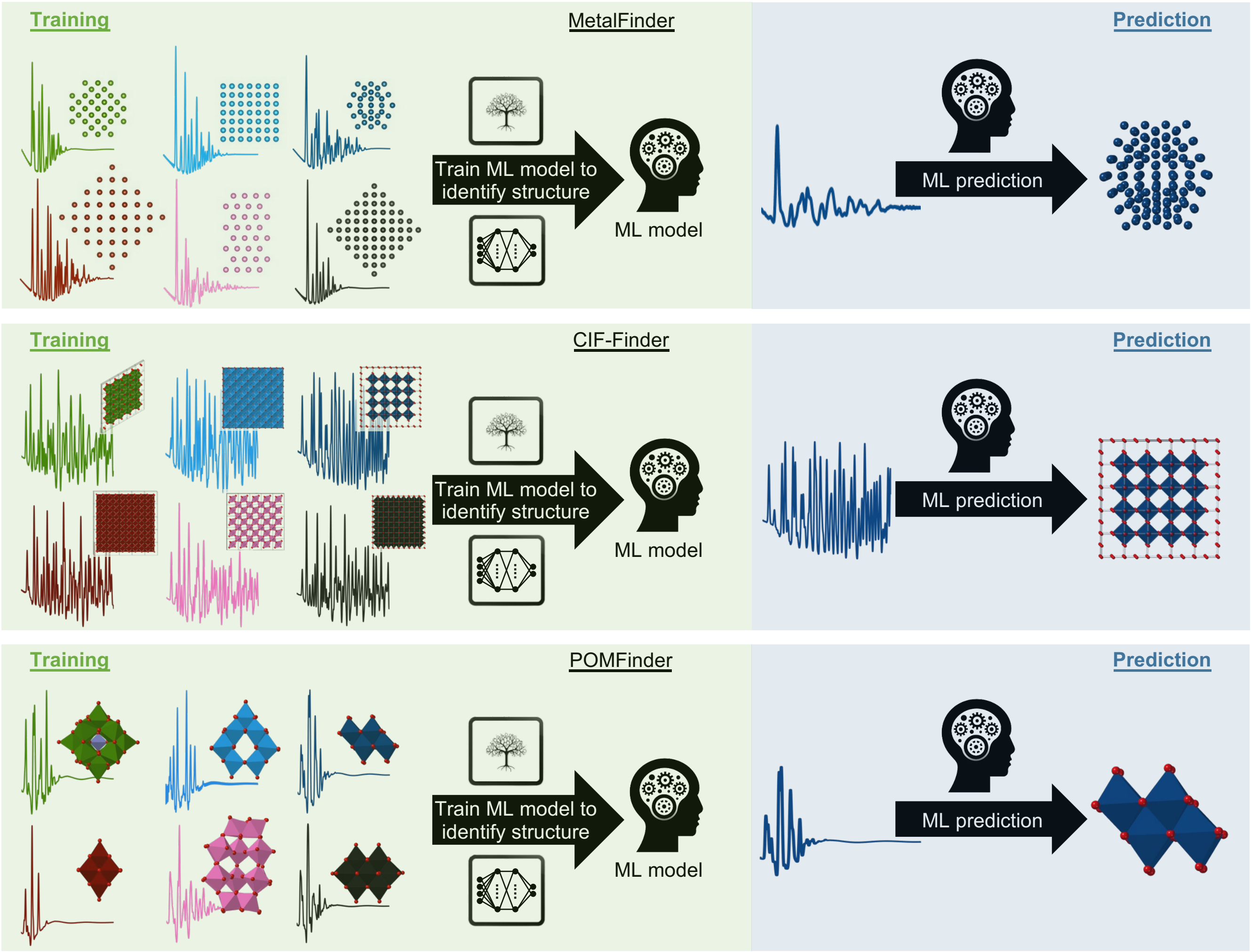
Marcus N. Weng

Scattering pattern matching

The brute-force approach

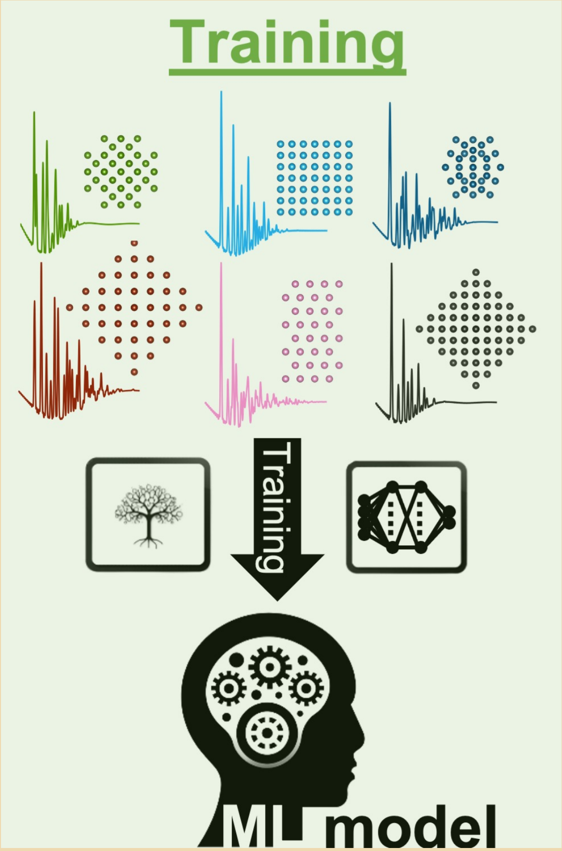


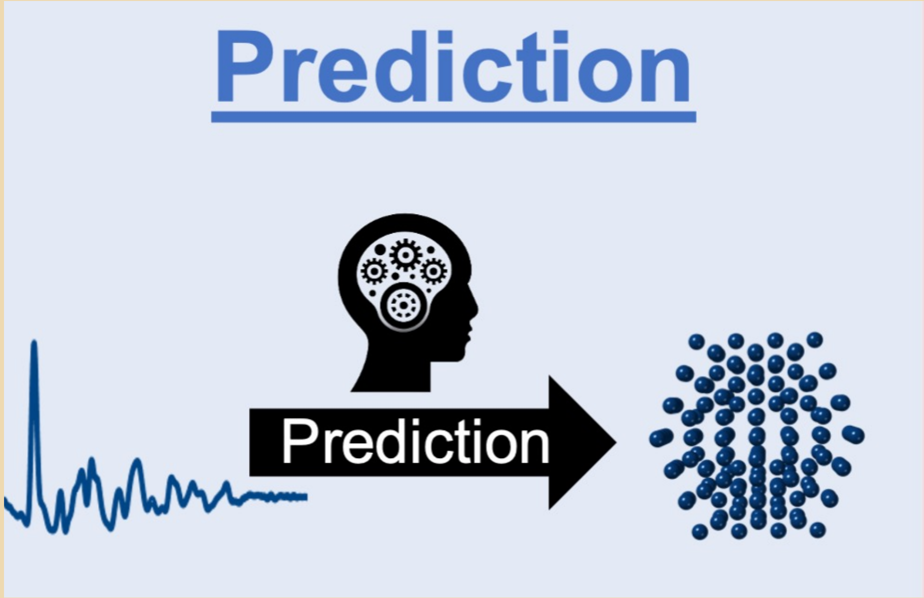
Using ML to analyse scattering data



- Machine Learning for Analysis of Experimental Scattering and Spectroscopy Data in Materials Chemistry, Anker S. A. et al., Chemical Science (2023)
- **MetalFinder** – DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., Digital Discovery (2023)
- **CIF-Finder** – MLstructureMining: A machine learning tool for structure identification from X-ray pair distribution functions, E. T. S. Kjær, et al. Digital Discovery (2024)
- **POMFinder** – Identifying polyoxometalate cluster structures from pair distribution function data using explainable machine learning, Anker A. S., et al., J. Appl. Crystallogr. (2024)

ML analysis of scattering and spectroscopy data

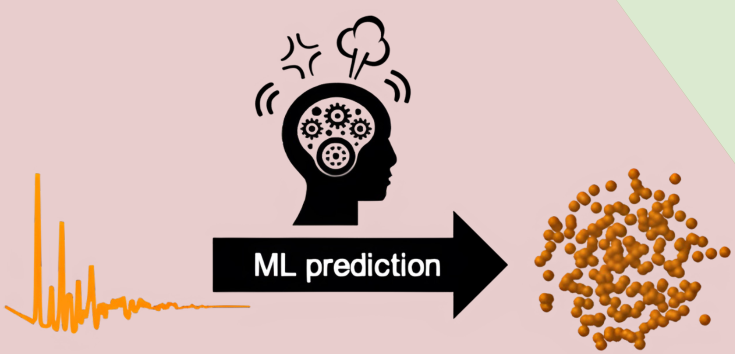




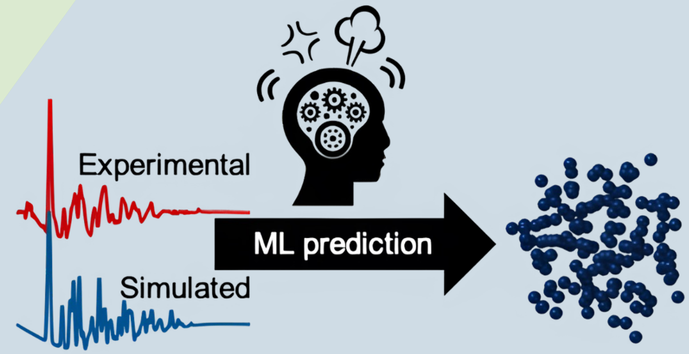
Machine learning challenges

Training process

Structure is not in structure database



Simulated data does not resemble experimental data

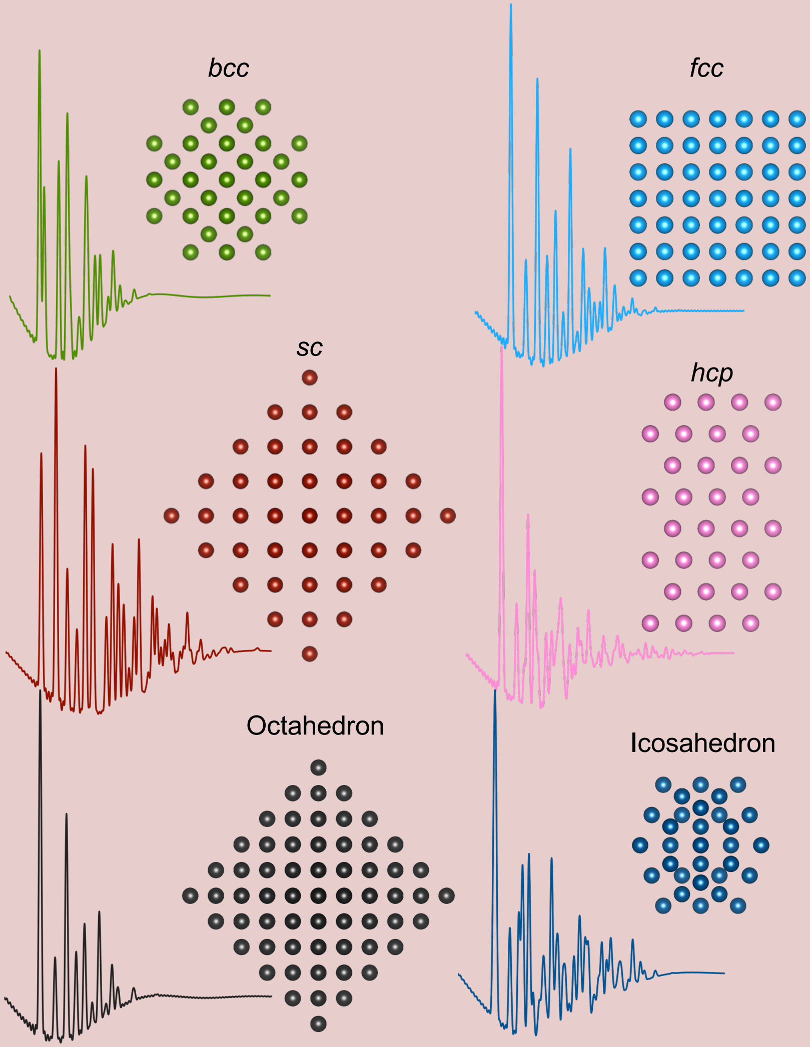


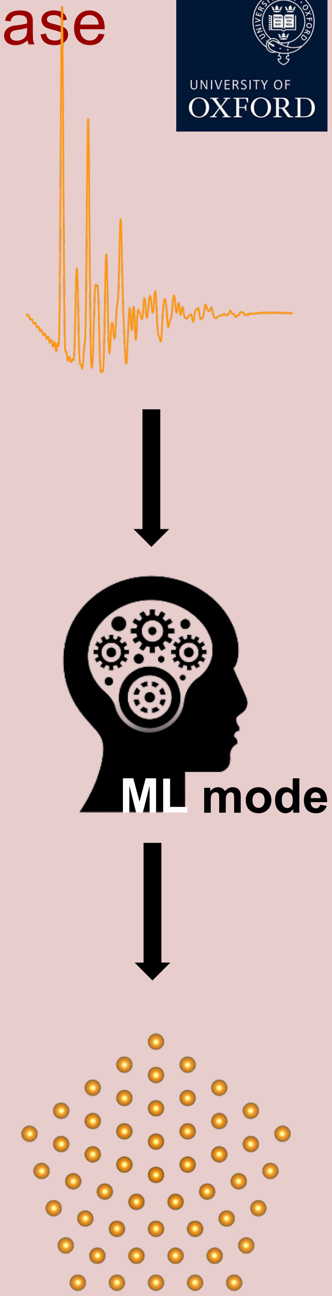
- Machine Learning for Analysis of Experimental Scattering and Spectroscopy Data in Materials Chemistry, Anker S. A. et al., Chemical Science (2023)
- DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., Digital Discovery (2023) **(Front cover)**
- Using generative adversarial networks to match experimental and simulated inelastic neutron scattering data, Anker A. S., et al., Digital Discovery (2023) **(Front cover)**

The challenge of structures not in the database

Using generative modelling

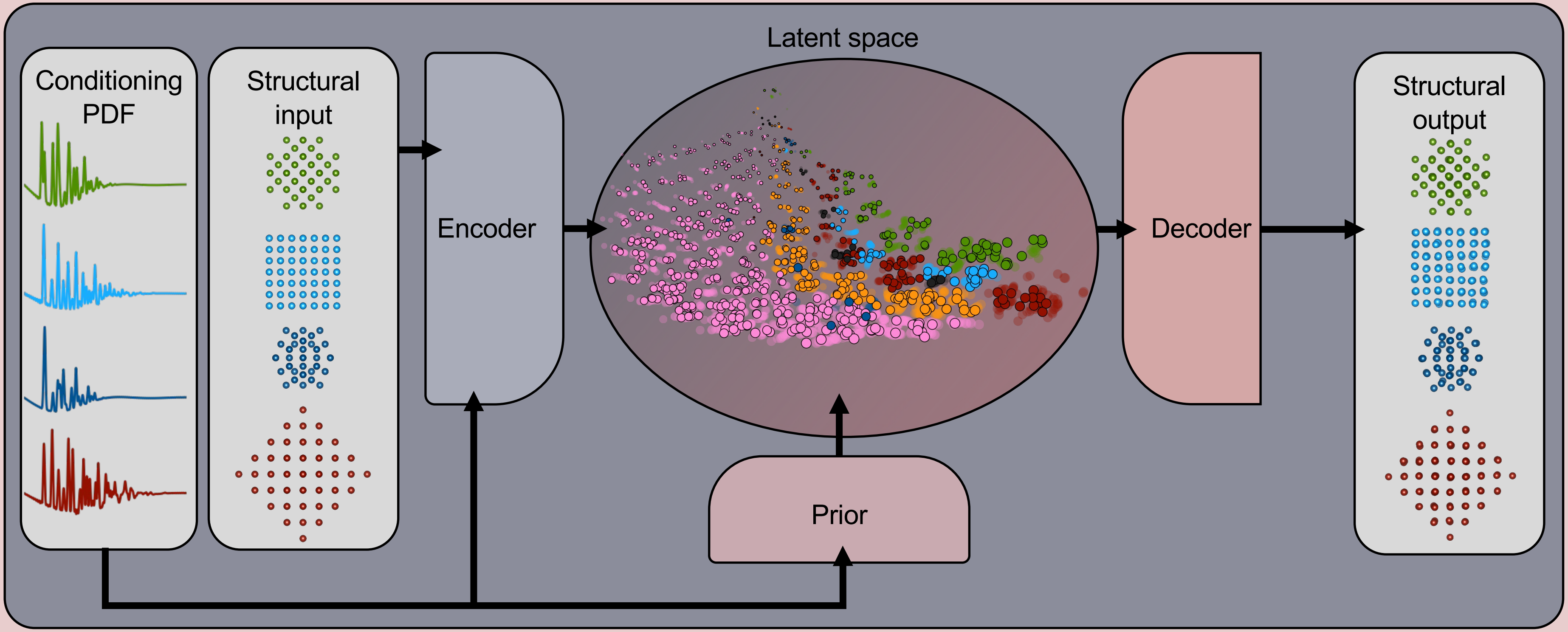
Training set





The challenge of structures not in the database

DeepStruc



- Characterising the atomic structure of mono-metallic nanoparticles from x-ray scattering data using conditional generative models, Anker A. S. & Kjær E. T. S., et al., 16th international workshop on mining and learning with graphs under KDD2020 conference (2020)
- DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., Digital Discovery (2023) **(Front cover)**

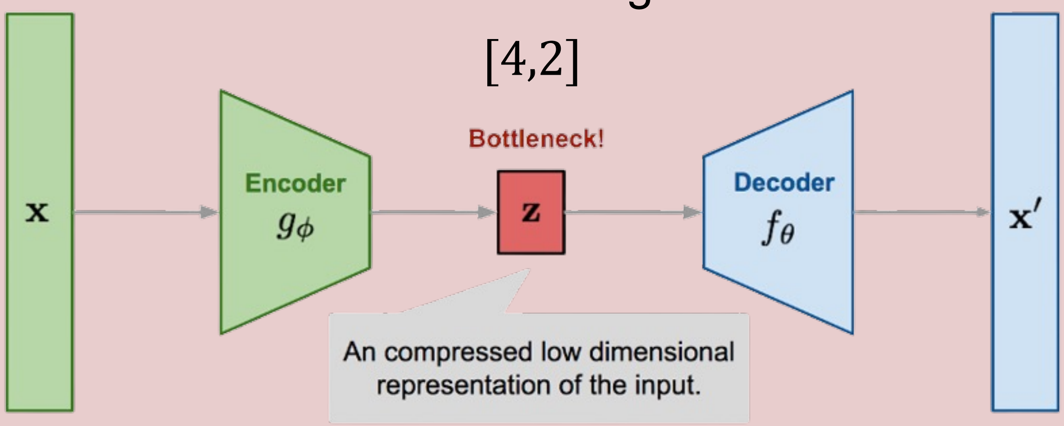
From point to probability
Autoencoder

Input

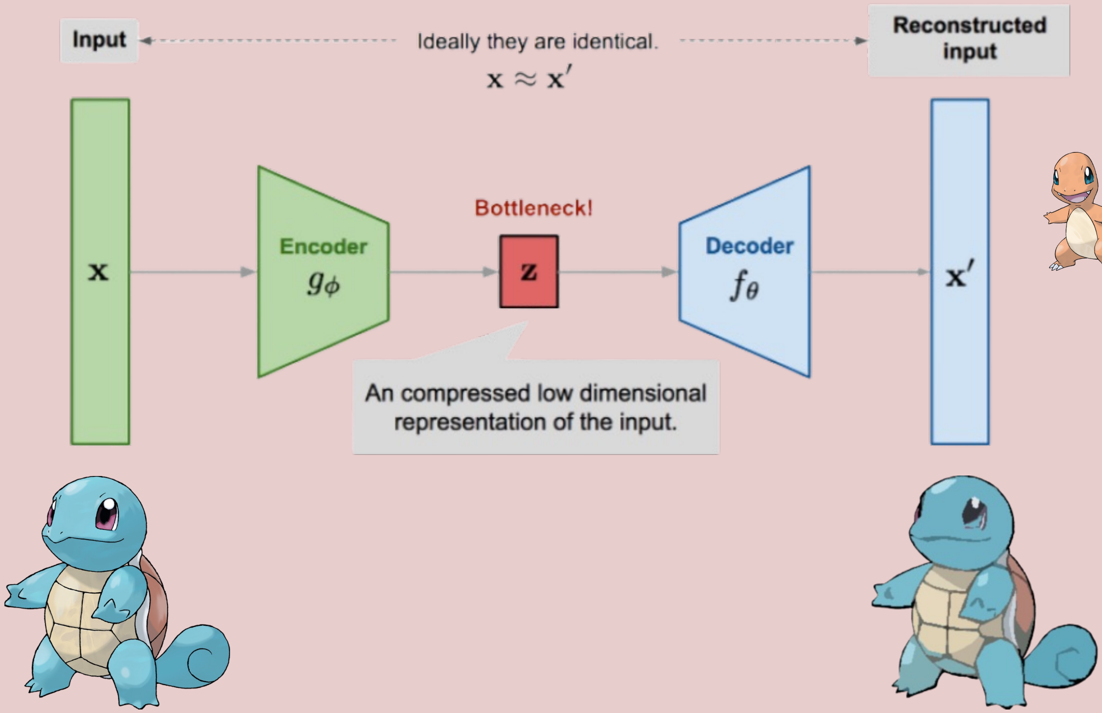
Output

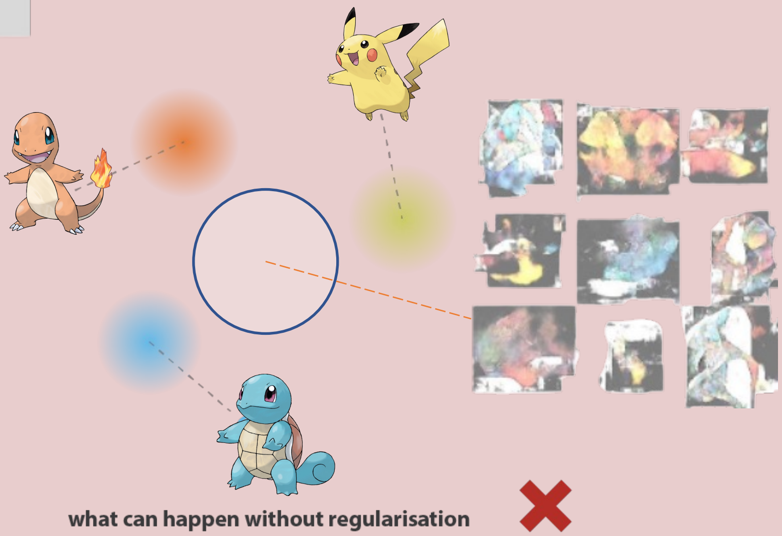
Embedding

[4,2]

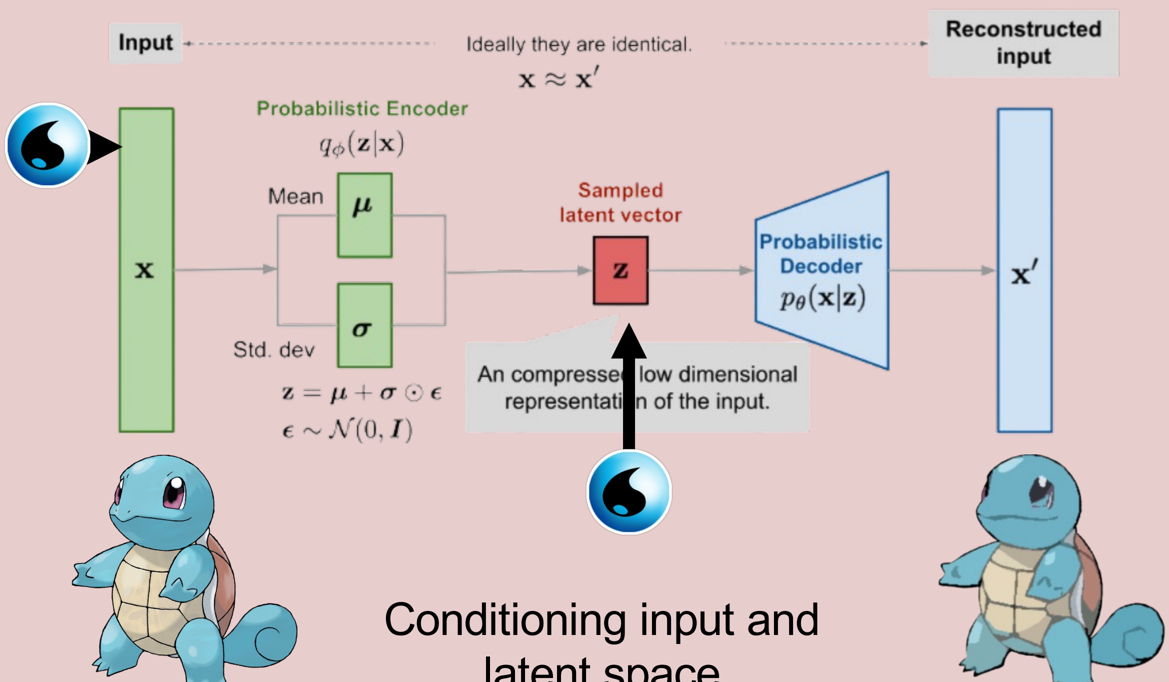


From point to probability
Autoencoder

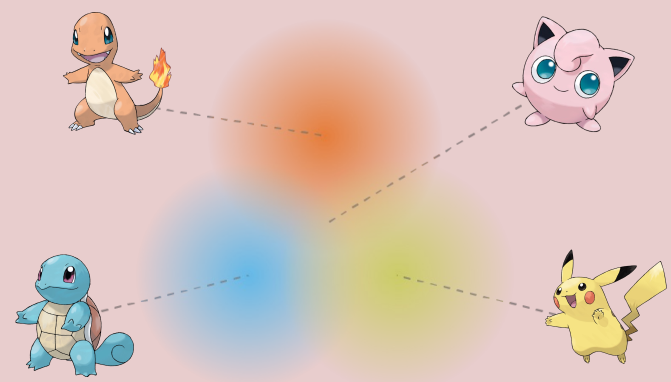




From point to probability
Variational autoencoder

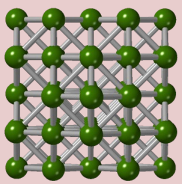


Conditioning input and latent space

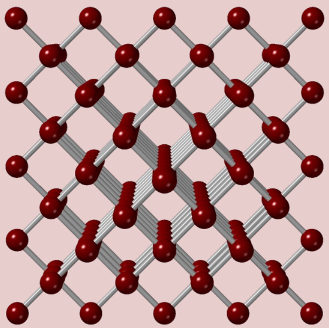


what we want to obtain with regularisation

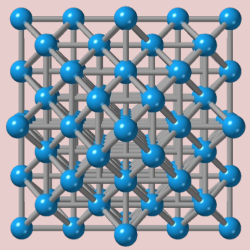
DeepStruc
Training set



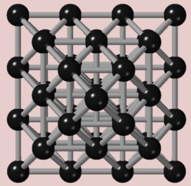
Body-centered cubic (BCC)



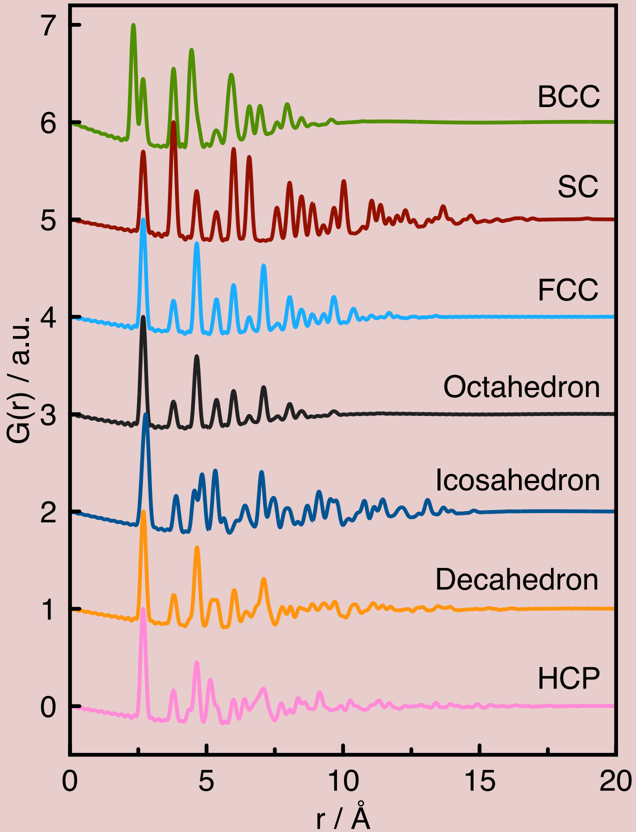
Simple cubic (SC)

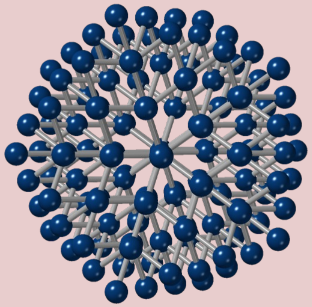


Face-centered Cubic (FCC)

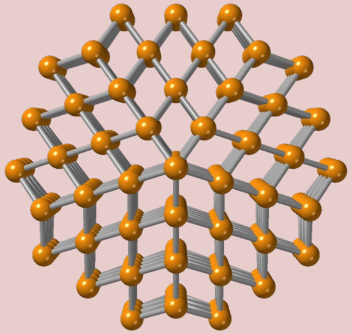


Octahedron

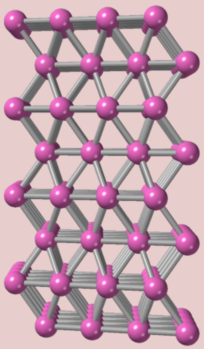




Icosahedron



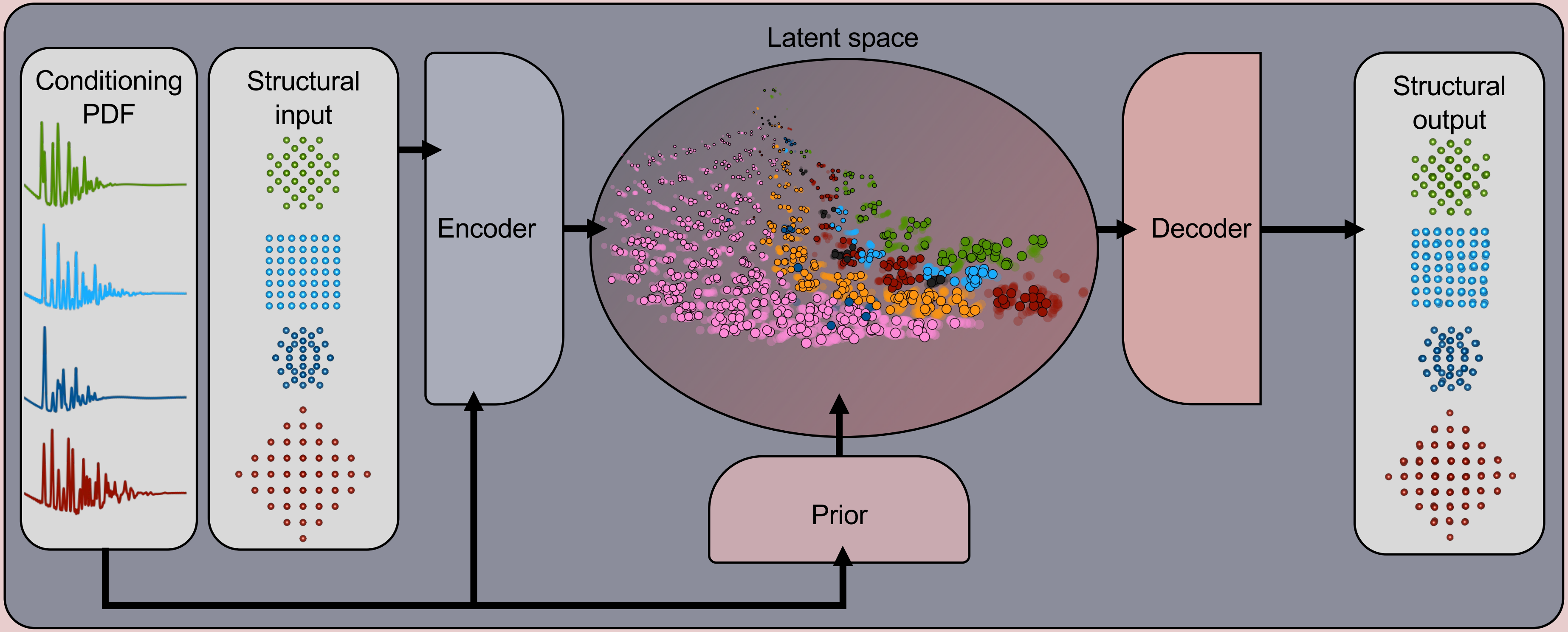
Decahedron



Hexagonal close-packed (HCP)

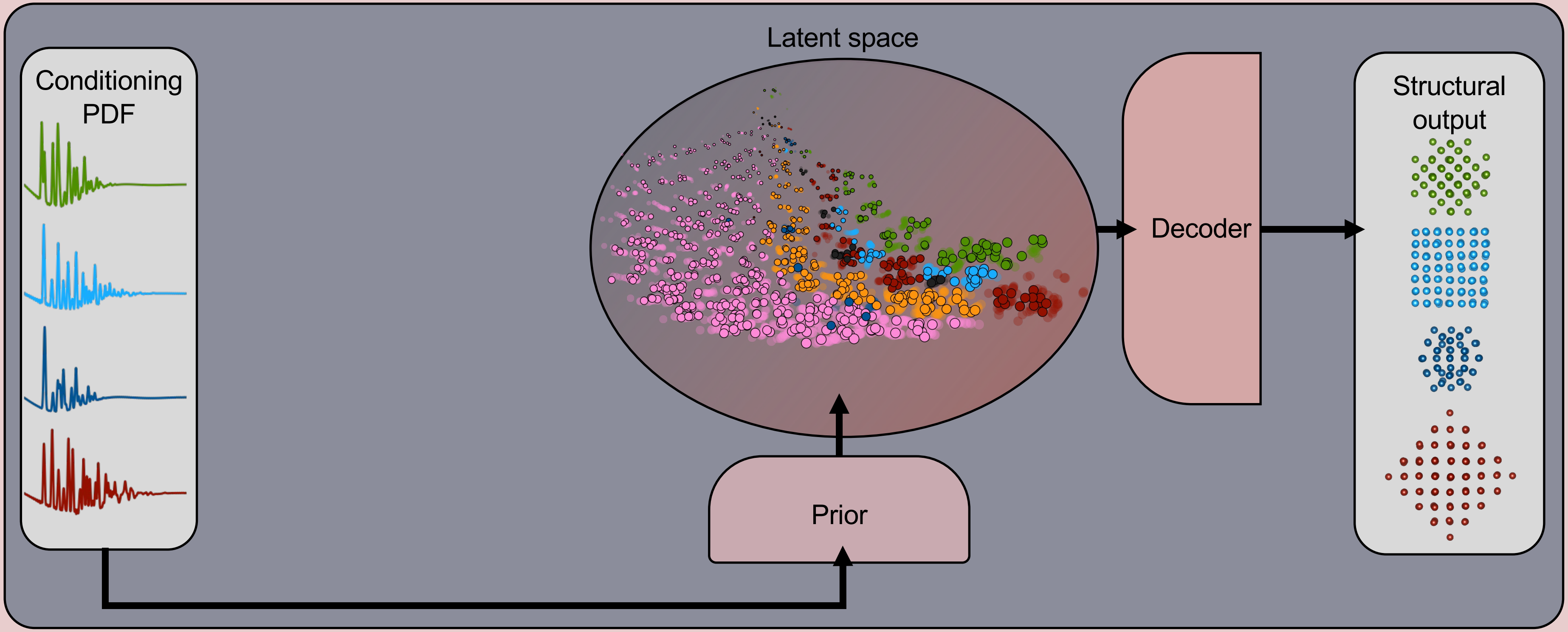
- Characterising the atomic structure of mono-metallic nanoparticles from x-ray scattering data using conditional generative models, Anker A. S. & Kjær E. T. S., et al., 16th international workshop on mining and learning with graphs under KDD2020 conference (2020)
- DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., Digital Discovery (2023) **(Front cover)**

DeepStruc
Training



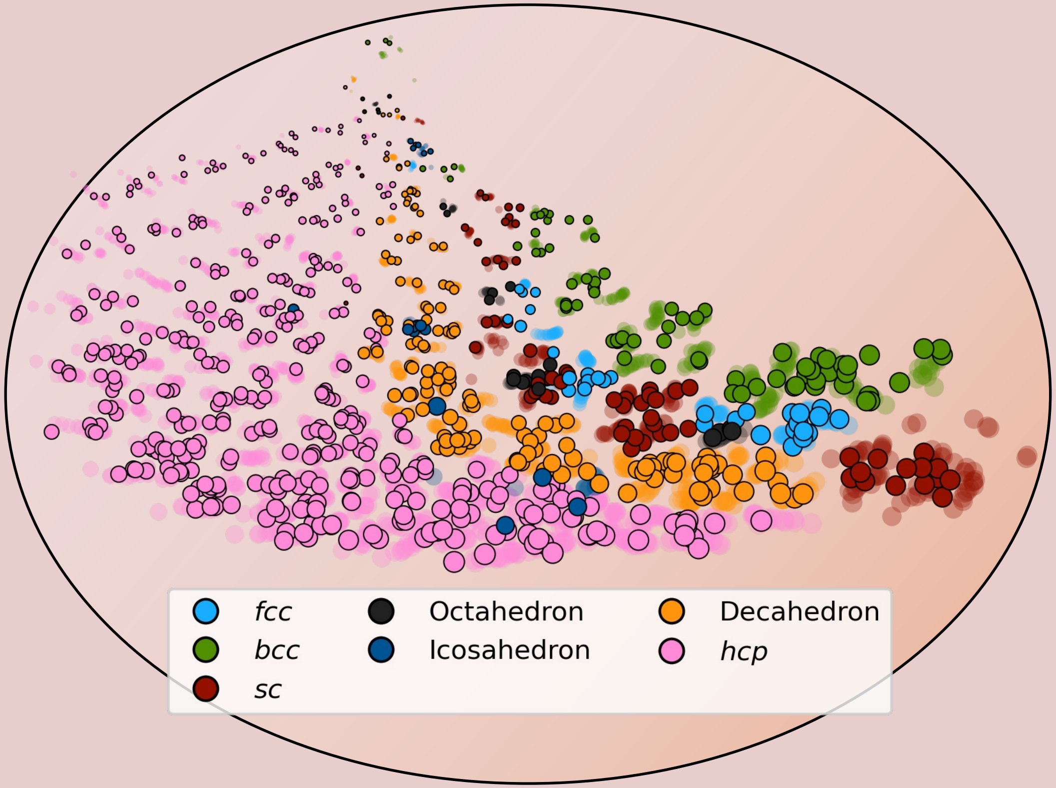
- Characterising the atomic structure of mono-metallic nanoparticles from x-ray scattering data using conditional generative models, Anker A. S. & Kjær E. T. S., et al., 16th international workshop on mining and learning with graphs under KDD2020 conference (2020)
- DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., Digital Discovery (2023) **(Front cover)**

DeepStruc
Inference



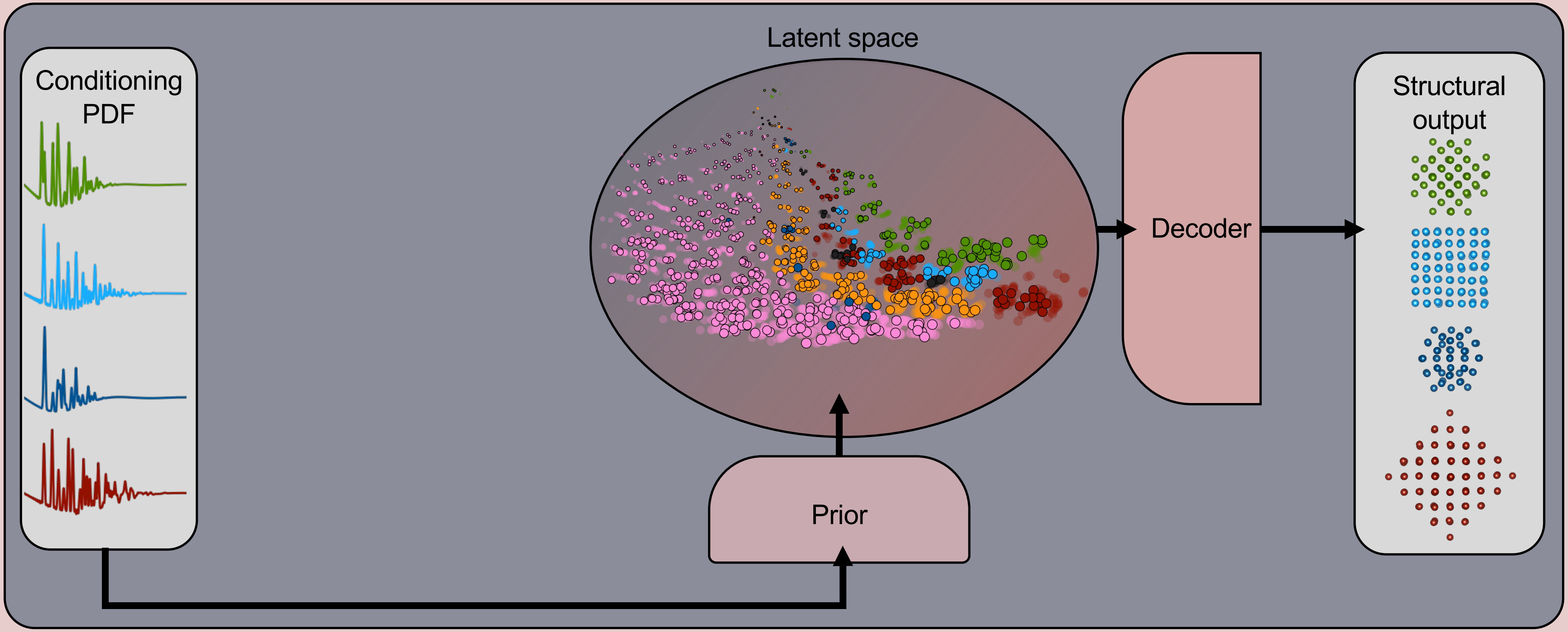
- Characterising the atomic structure of mono-metallic nanoparticles from x-ray scattering data using conditional generative models, Anker A. S. & Kjær E. T. S., et al., 16th international workshop on mining and learning with graphs under KDD2020 conference (2020)
- DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., Digital Discovery (2023) **(Front cover)**

DeepStruc
Latent space



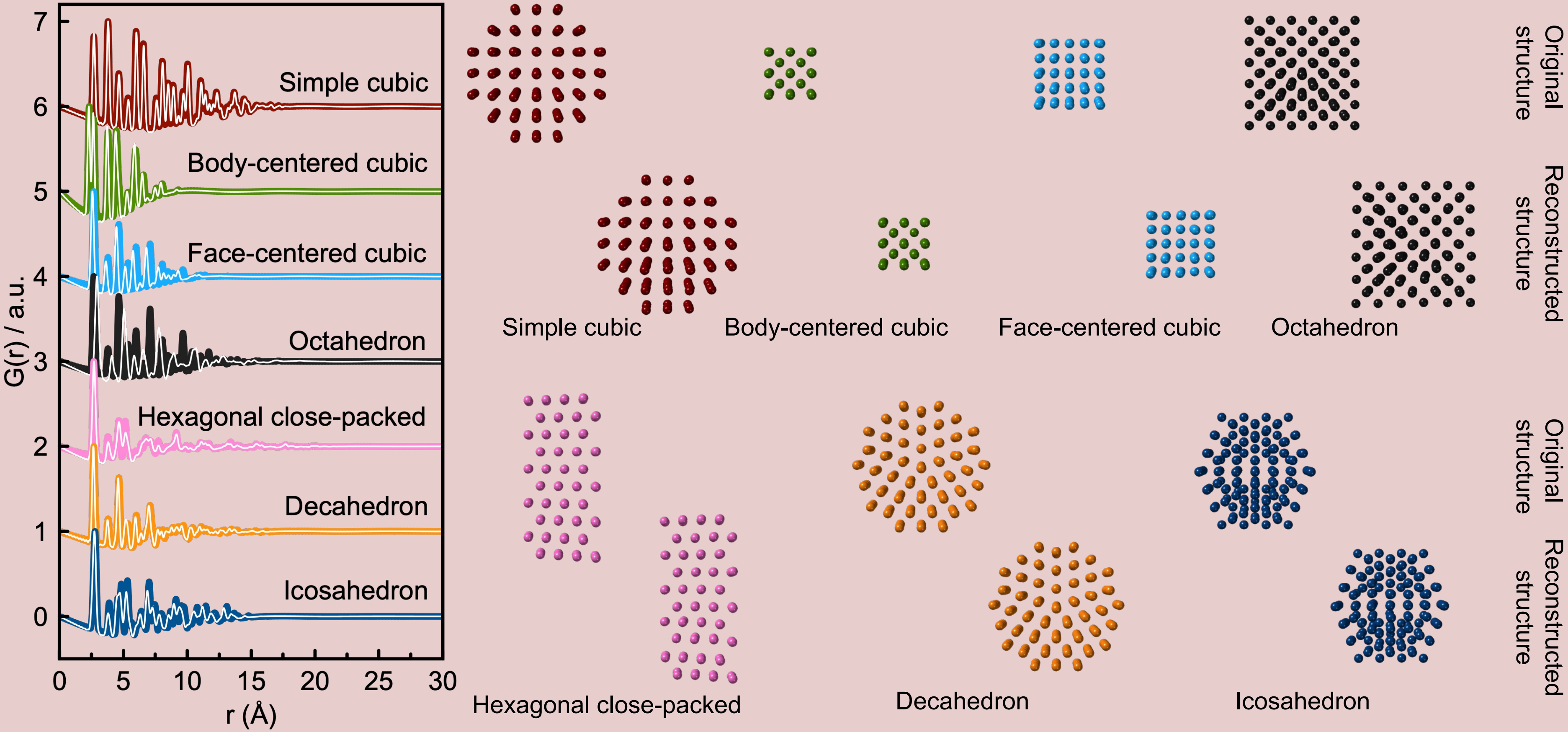
- Characterising the atomic structure of mono-metallic nanoparticles from x-ray scattering data using conditional generative models, Anker A. S. & Kjær E. T. S., et al., 16th international workshop on mining and learning with graphs under KDD2020 conference (2020)
- DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., Digital Discovery (2023) **(Front cover)**

DeepStruc
Reconstructions



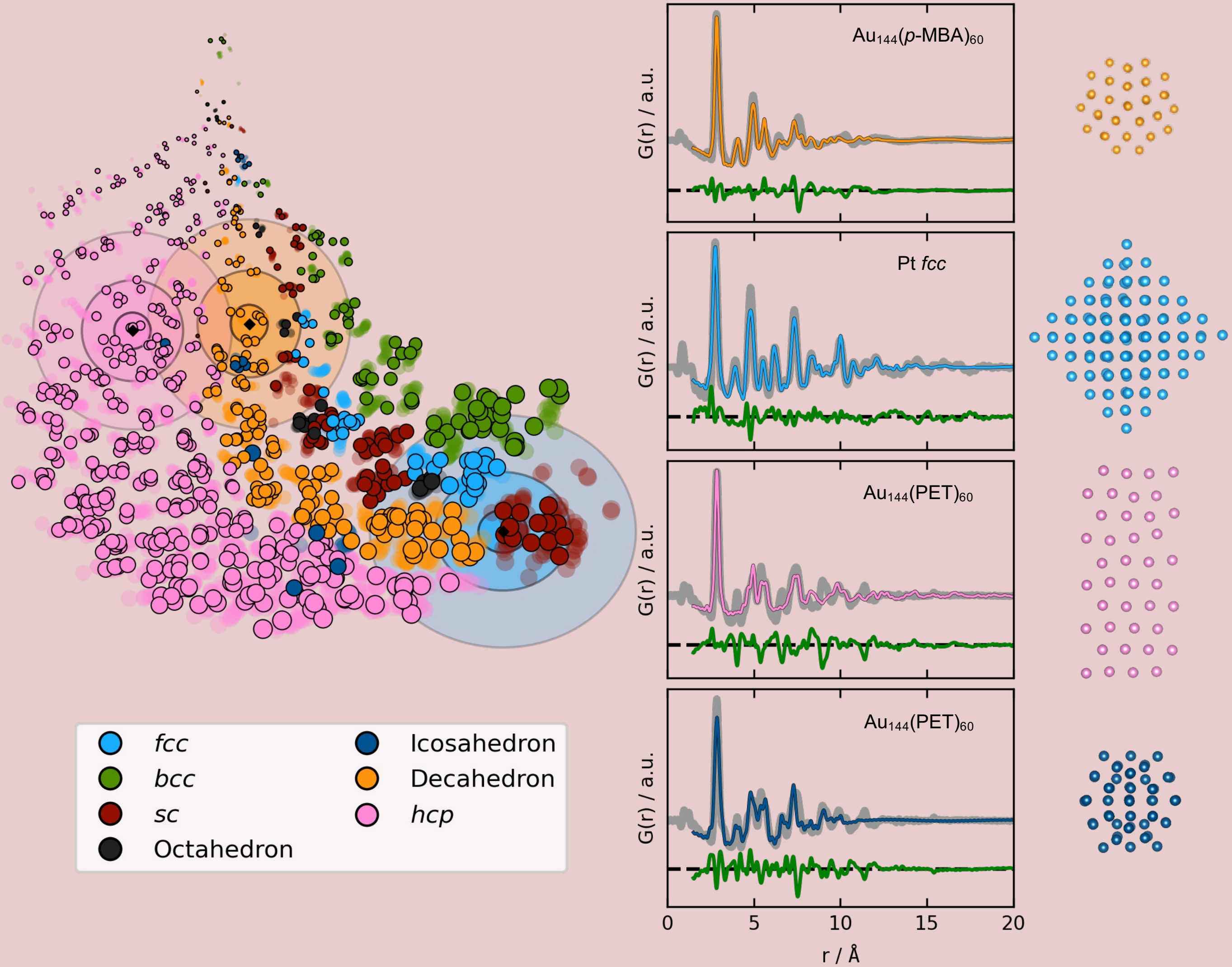
➢ Characterising the atomic structure of mono-metallic nanoparticles from x-ray scattering data using conditional generative models, Anker A. S. & Kjær E. T. S., et al., 16th international workshop on mining and learning with graphs under KDD2020 conference (2020)
➢ DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., Digital Discovery (2023) **(Front cover)**

DeepStruc
Reconstructions

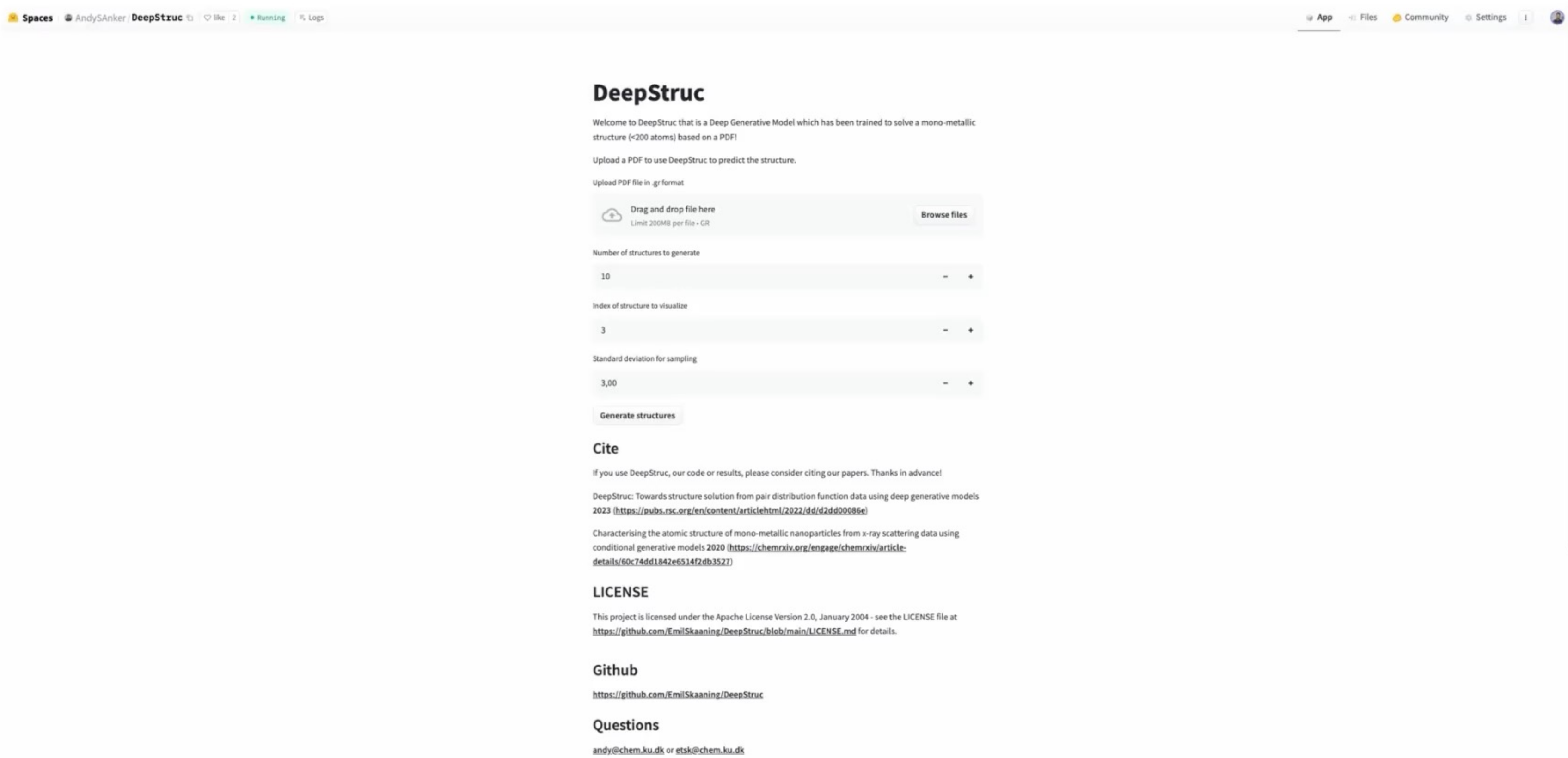


Mean squared error of atomic positions = 0.128 ± 0.073 Å

DeepStruc
Experimental PDFs



DeepStruc

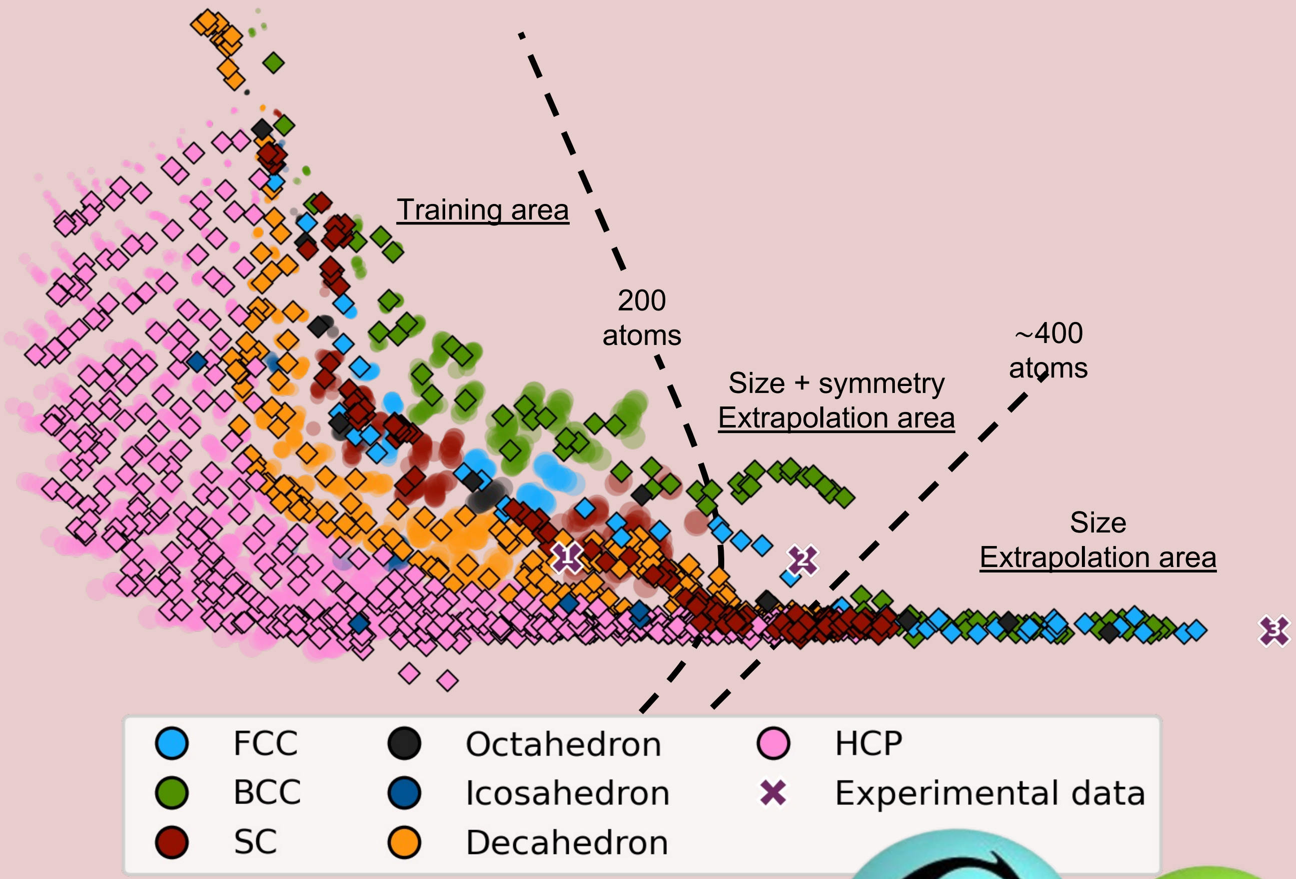


150 % speed

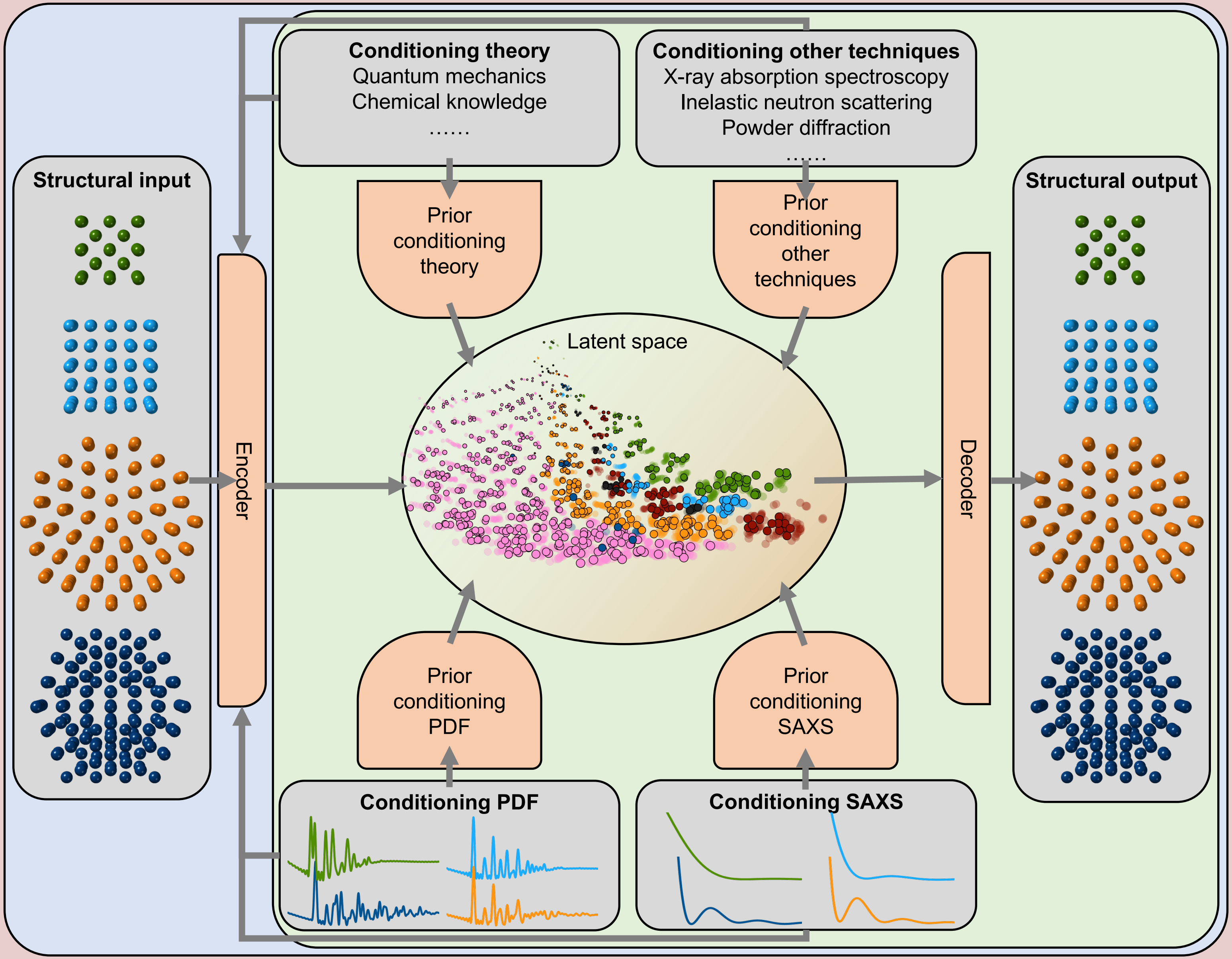
- Characterising the atomic structure of mono-metallic nanoparticles from x-ray scattering data using conditional generative models, Anker A. S. & Kjær E. T. S., et al., 16th international workshop on mining and learning with graphs under KDD2020 conference (2020)
- DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., Digital Discovery (2023) **(Front cover)**

GitHub: https://github.com/EmilSkaaning/DeepStruc
Hugging Face: https://huggingface.co/spaces/AndySAnker/DeepStruc

DeepStruc
Extrapolation

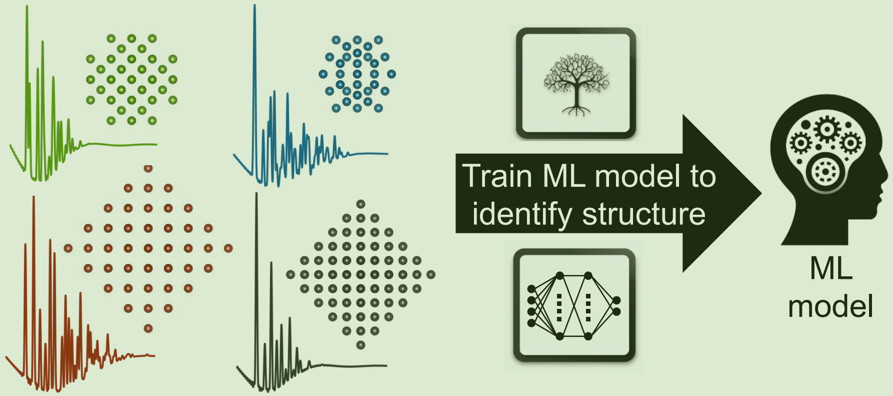


Automated structure solution

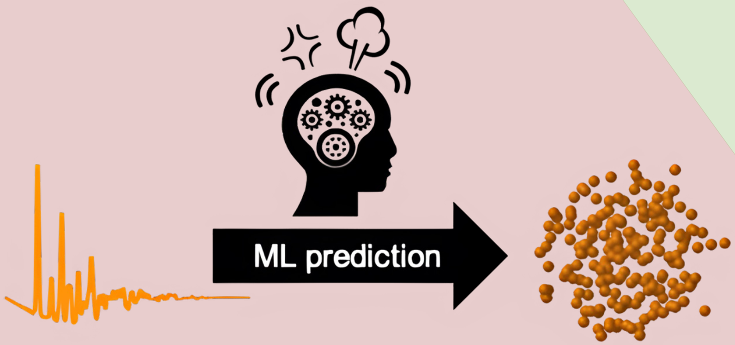


Machine learning challenges

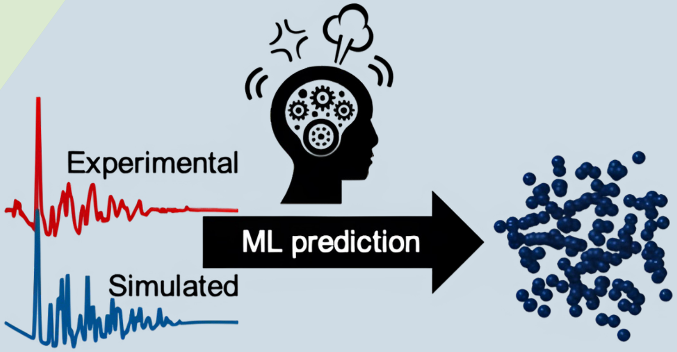
Training process



Structure is not in structure database

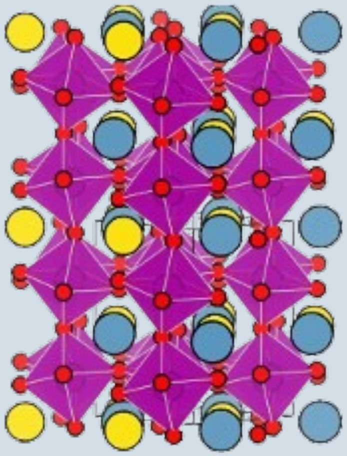


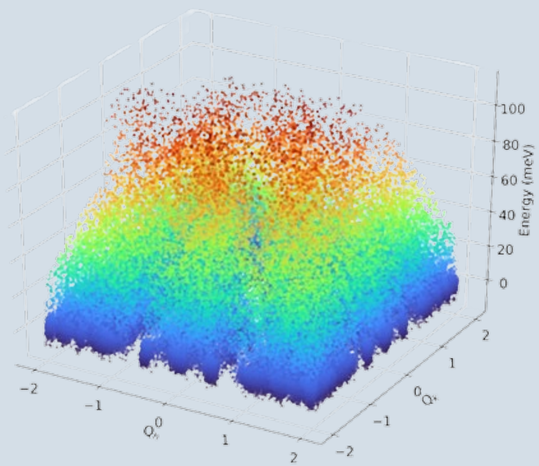
Simulated data does not resemble experimental data

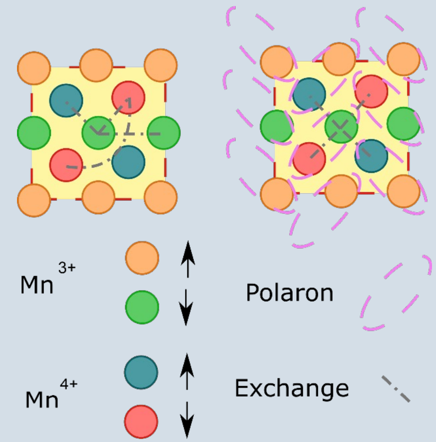


- Machine Learning for Analysis of Experimental Scattering and Spectroscopy Data in Materials Chemistry, Anker S. A. et al., Chemical Science (2023)
- DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., Digital Discovery (2023) **(Front cover)**
- Using generative adversarial networks to match experimental and simulated inelastic neutron scattering data, Anker A. S., et al., Digital Discovery (2023) **(Front cover)**

An Deep Learning model for identifying magnetic Hamiltonians







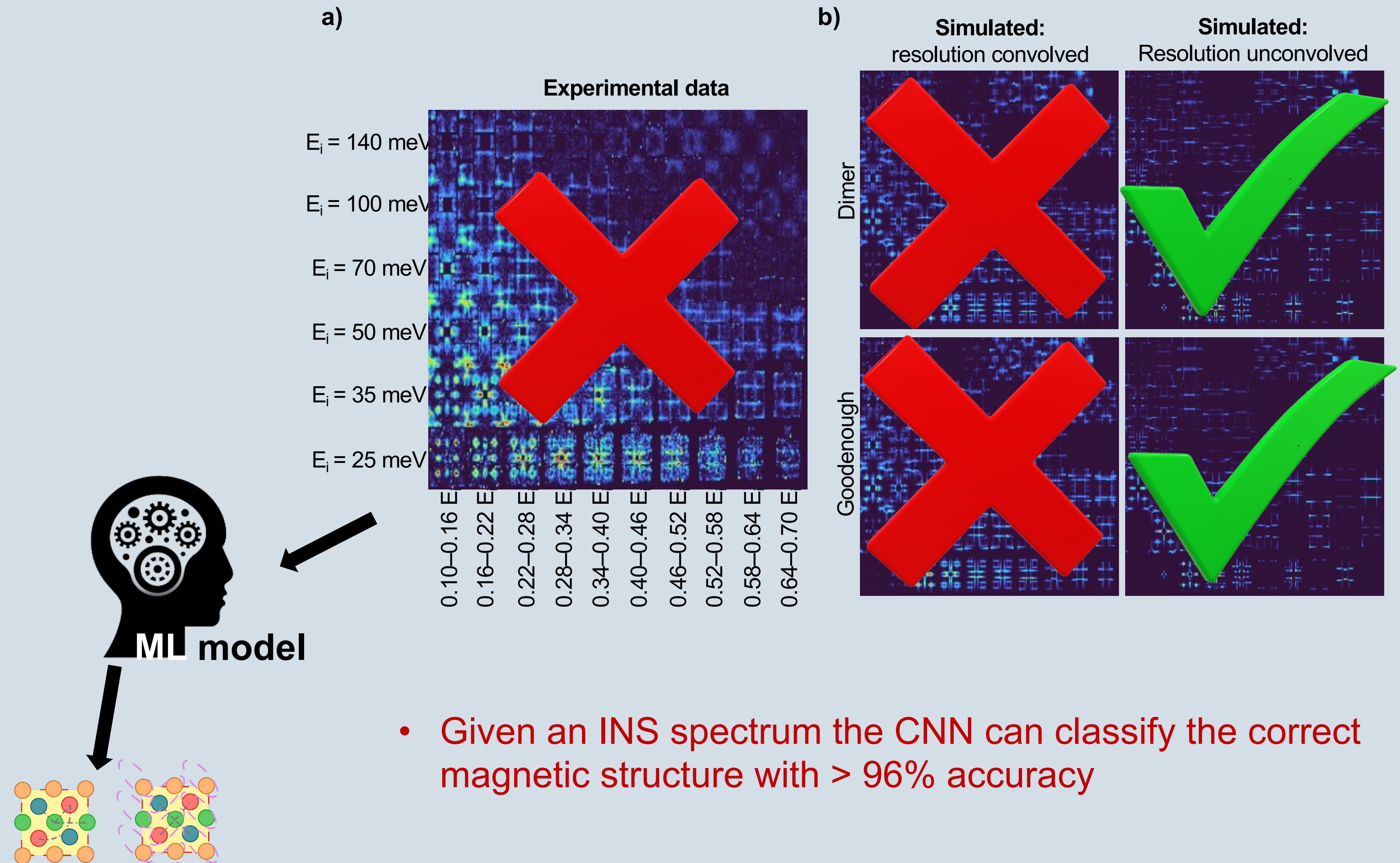
PCSMO is a half-doped manganite double perovskite
The nature of the magnetism has been probed with inelastic neutron scattering
Can a neural network infer the magnetic structure from the INS data?

Keith T. Butler

Manh D. Le

➢ Interpretable, calibrated neural networks for analysis and understanding of inelastic neutron scattering data, Butler T. K. et al., Journal of Physics: Condensed Matter (2021)

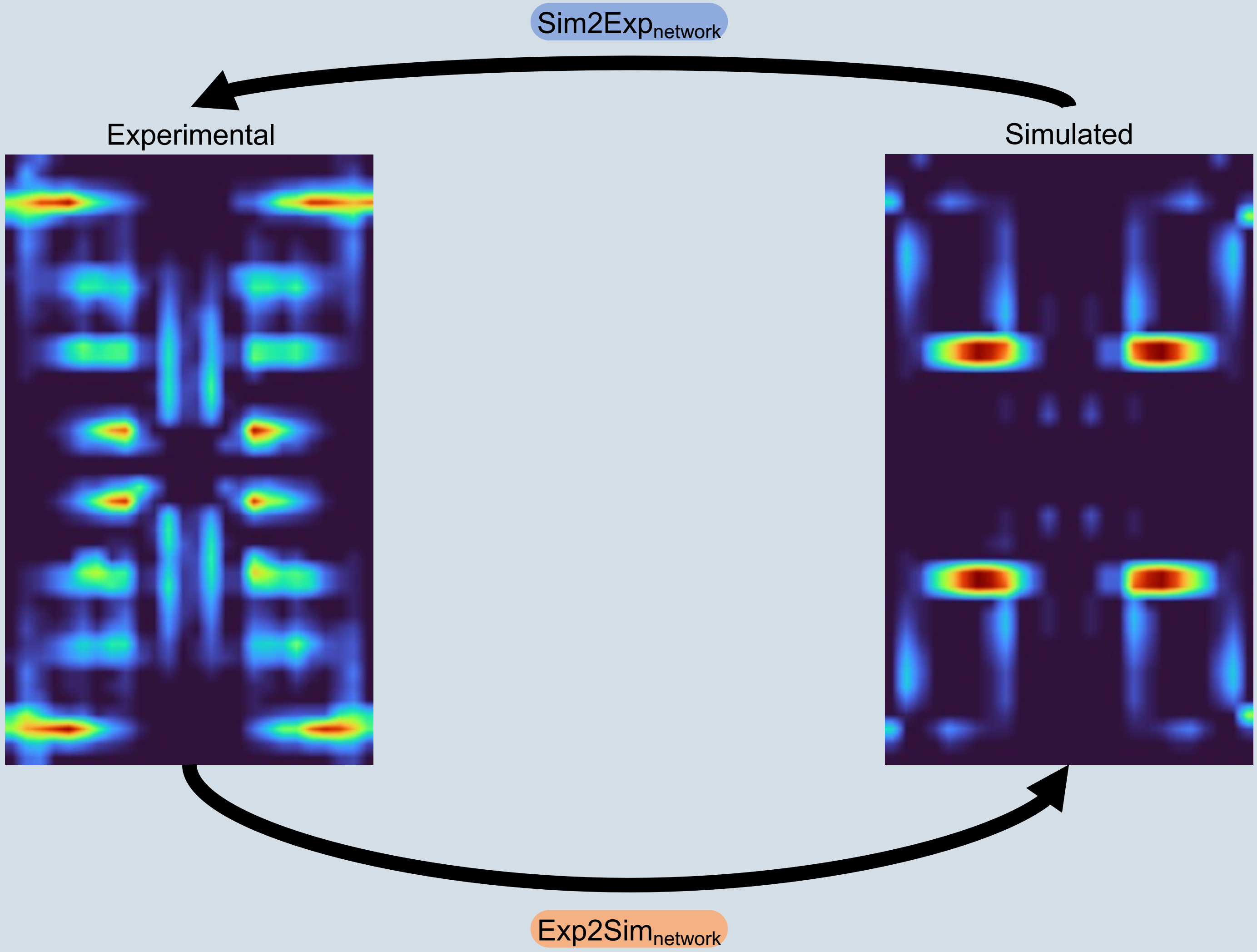
An Deep Learning model for identifying magnetic Hamiltonians



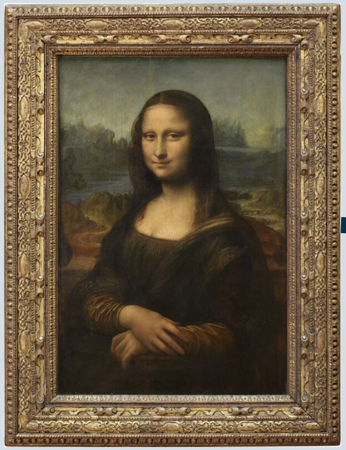
- Given an INS spectrum the CNN can classify the correct magnetic structure with > 96% accuracy

Unpaired image-to-image translation

Exp2SimGAN



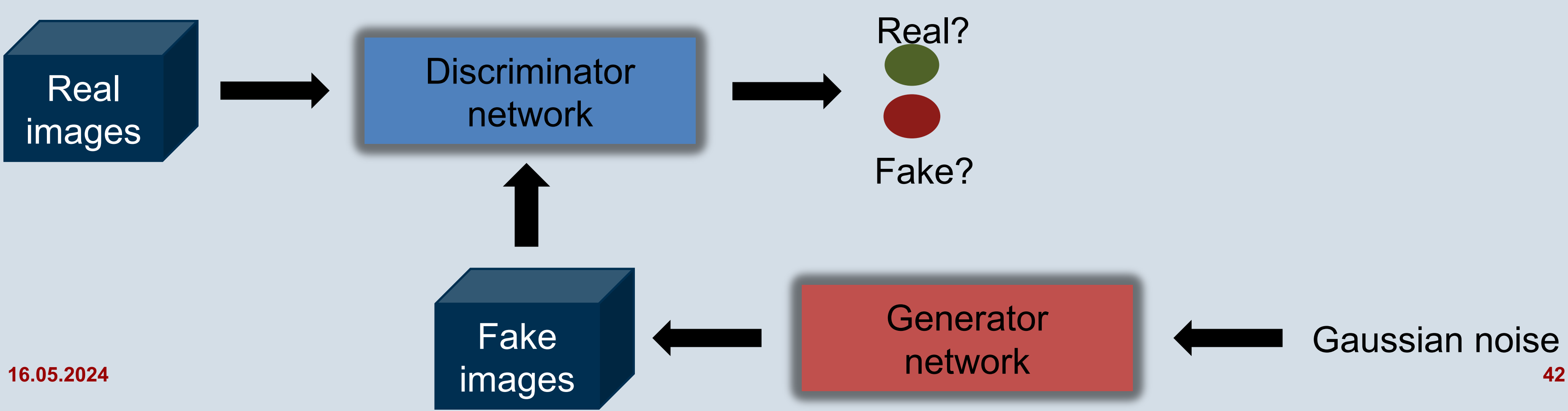
Generative adversarial networks (GANs)



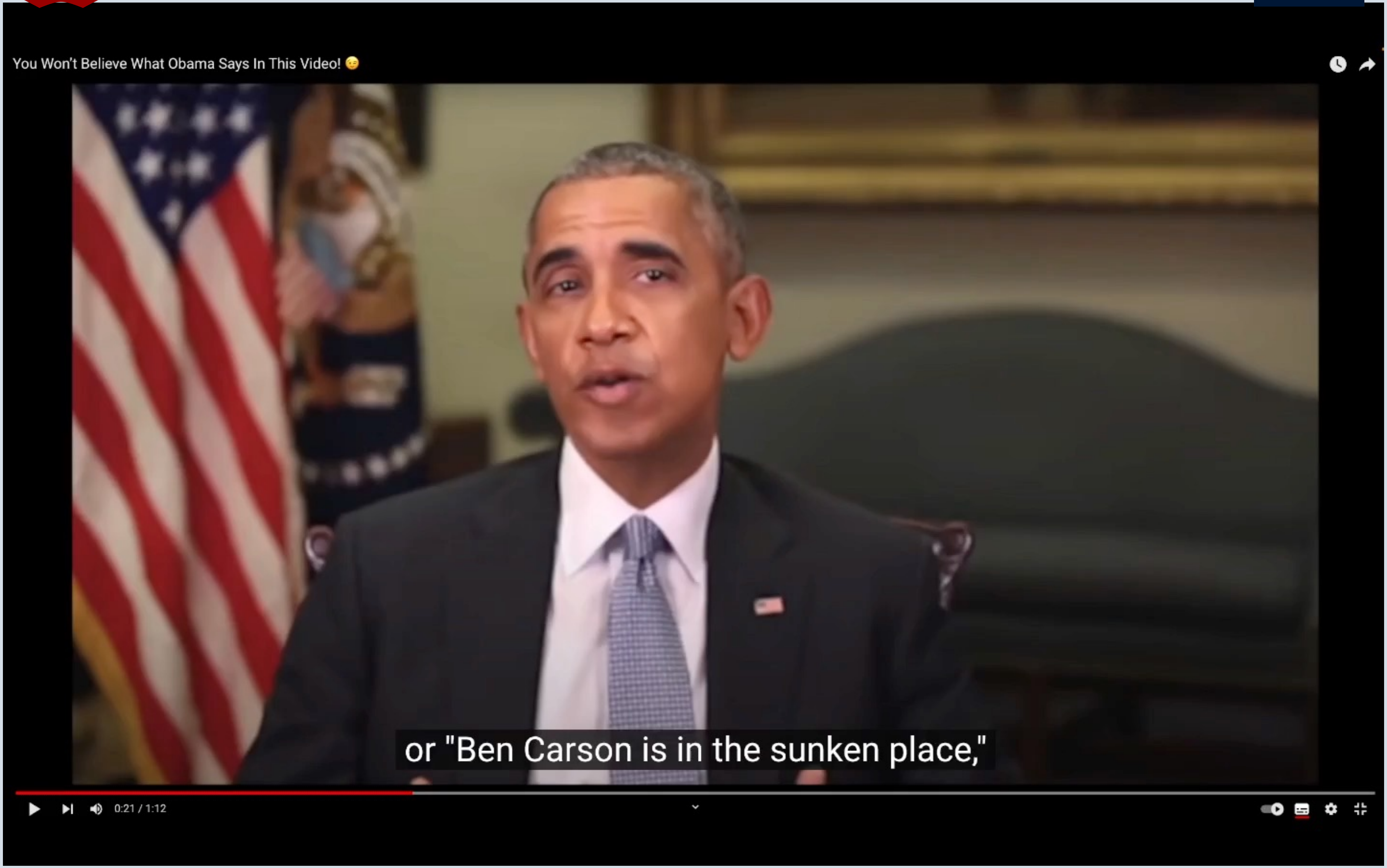
Real data (Paintings)

Discriminator
(Detective)

Fake data (painting) by
generator (Forger)

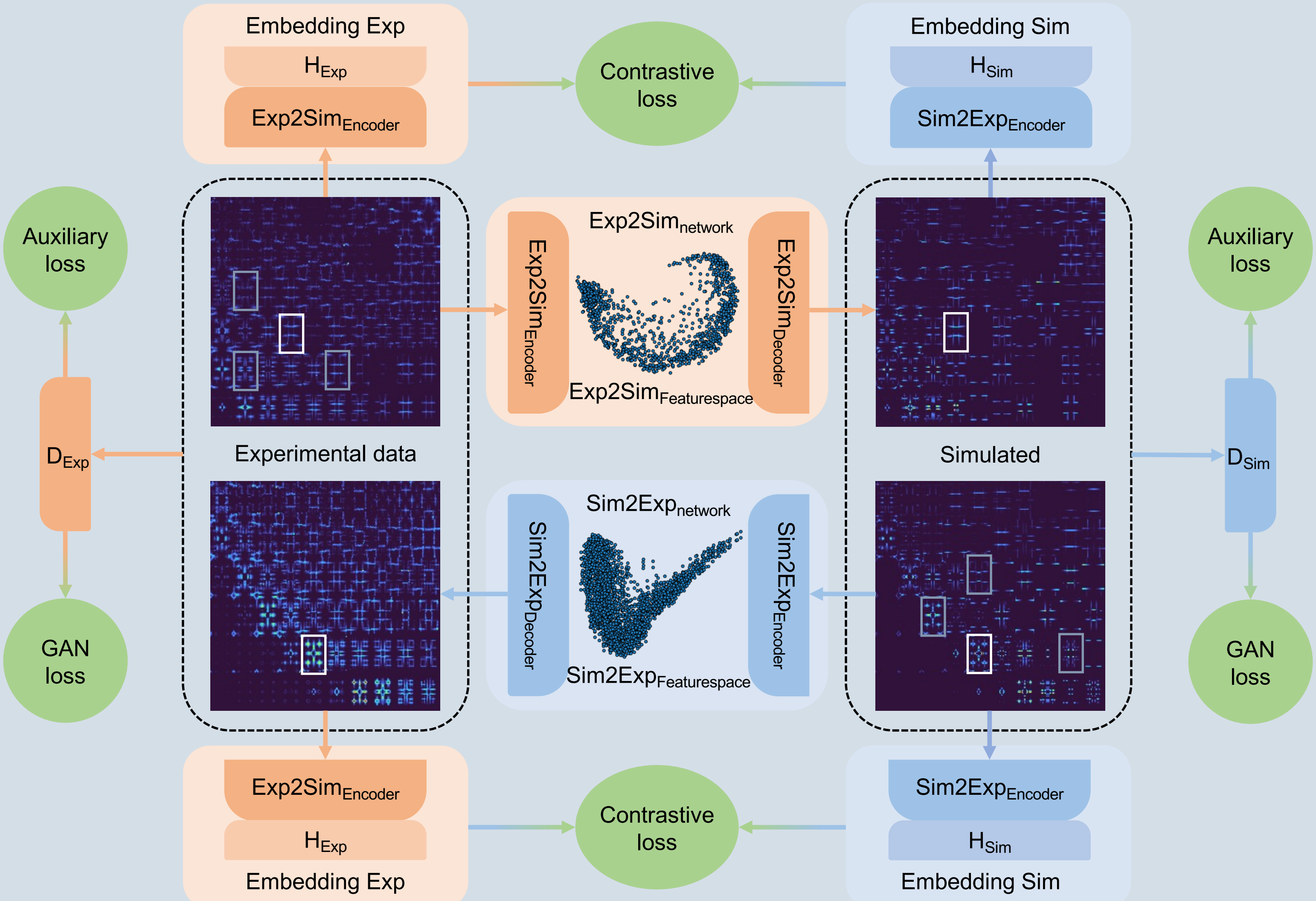


Deepfake



➢ https://www.youtube.com/watch?v=cQ54GDm1eL0&ab_channel=BuzzFeedVideo

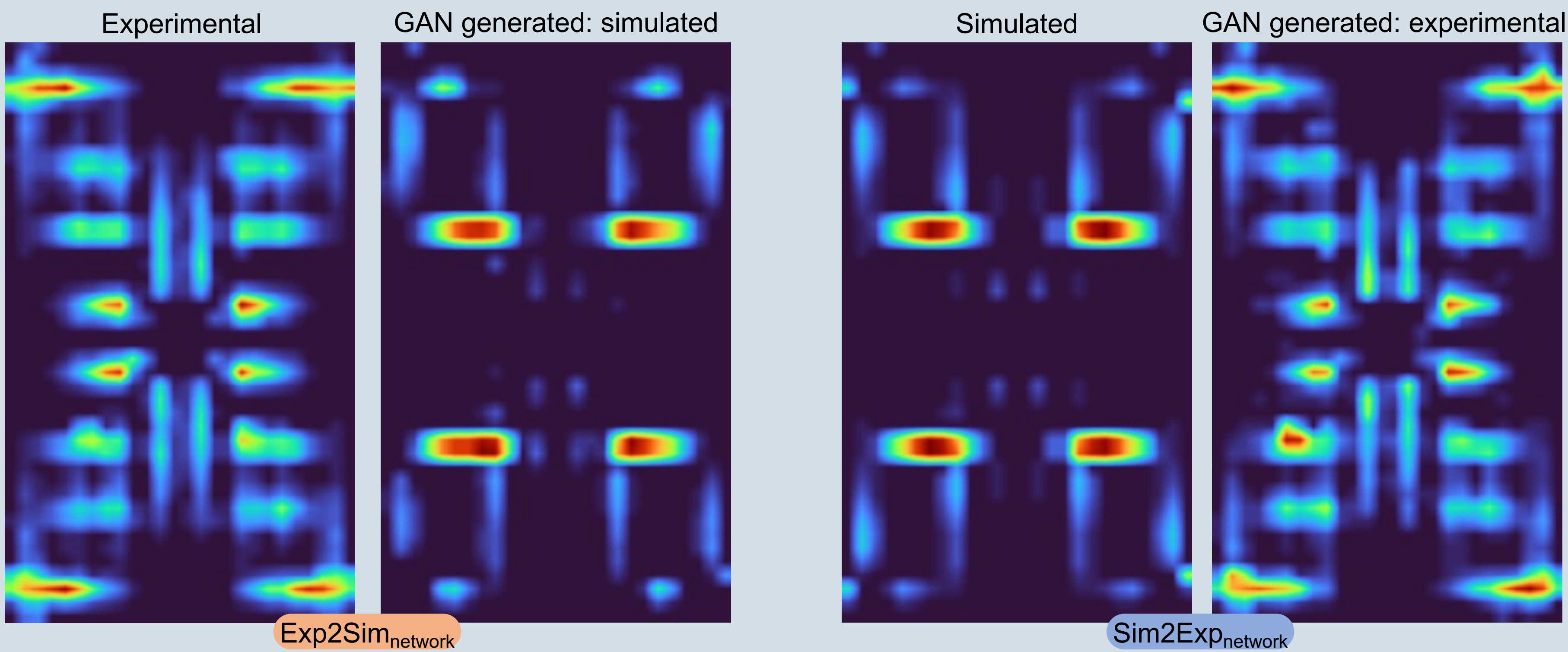
Unpaired image-to-image translation
Exp2SimGAN



➢ Using generative adversarial networks to match experimental and simulated inelastic neutron scattering data, Anker A. S., et al., Digital Discovery (2023) **(Front cover)**

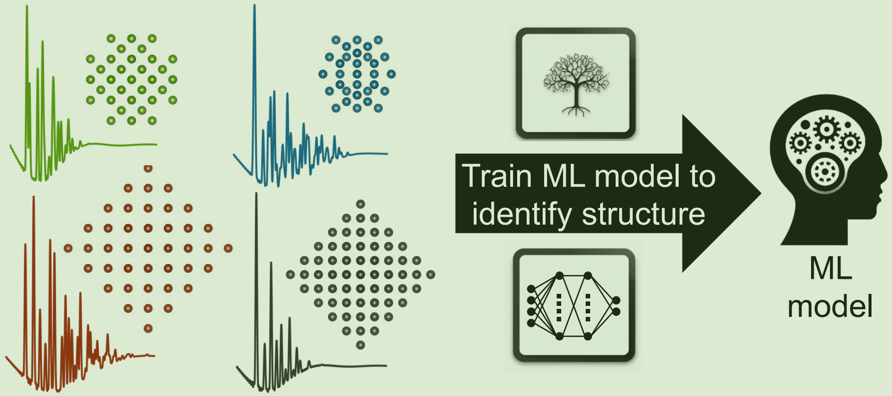
Unpaired image-to-image translation

Exp2SimGAN

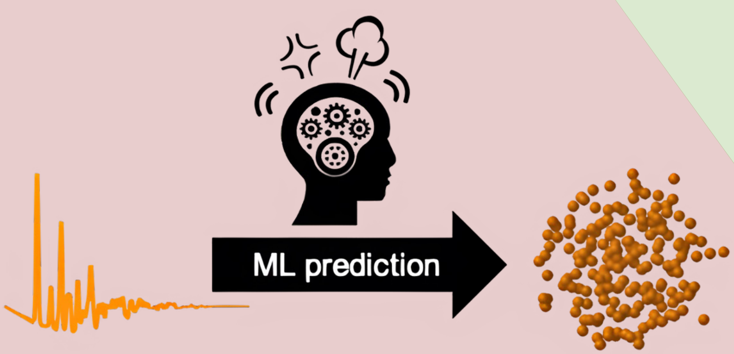


Machine learning challenges

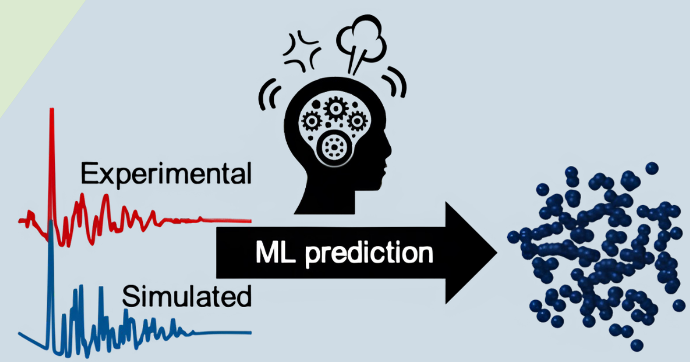
<u>Training process</u>



<u>Structure is not in structure database</u>



<u>Simulated data does not resemble experimental data</u>

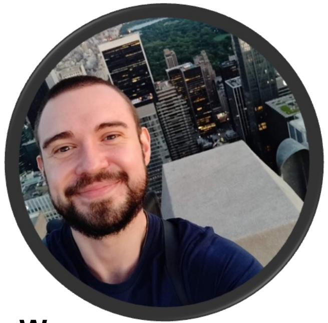


- Machine Learning for Analysis of Experimental Scattering and Spectroscopy Data in Materials Chemistry, Anker S. A. et al., Chemical Science (2023)
- DeepStruc: Towards structure solution from pair distribution function data using deep generative models, Kjær E. T. S. & Anker A. S., et al., Digital Discovery (2023) **(Front cover)**
- Using generative adversarial networks to match experimental and simulated inelastic neutron scattering data, Anker A. S., et al., Digital Discovery (2023) **(Front cover)**

Acknowledgements

Kirsten M. Ø. Jensen
University of Copenhagen

Emil T. S. Kjær



Marcus N. Weng

Scientific Machine Learning Group
Jeyan Thiyagalingam

Science & Technology Facilities Council
ISIS Neutron and Muon Source
Manh Duc Le
Toby Perring

Keith T. Butler
University College London

Simon J. L. Billinge
Columbia University

Raghavendra Selvan
University of Copenhagen