

ML-XLPA for HCP Nanostructure Characterization

Extending Machine Learning-Based X-ray Line Profile Analysis
from Cubic to Hexagonal Close-Packed Crystals

Nguyen Khac Huy

MSc Applied Mathematics / ELTE

Semester 2025/26 II

Key Terms & Abbreviations

Methods

- **XRD** — X-ray Diffraction
- **XLPA** — X-ray Line Profile Analysis
- **ML-XLPA** — Machine Learning-based XLPA
- **CMWP** — Convolutional Multiple Whole Profile (traditional fitting)
- **ARPLS** — baseline subtraction algorithm

Crystal structures

- **FCC** — Face-Centred Cubic
- **BCC** — Body-Centred Cubic
- **HCP** — Hexagonal Close-Packed

Microstructural parameters (model outputs)

- m — median crystallite size [nm]
- σ — width of the size distribution
- ρ — dislocation density [nm^{-2}]
- R^* — how clustered the dislocations are
- α — stacking fault probability
- β — twin fault probability

ML models

- **XGBoost** — gradient-boosted decision trees
- **CNN** — Convolutional Neural Network

Problem & Motivation

The bottleneck:

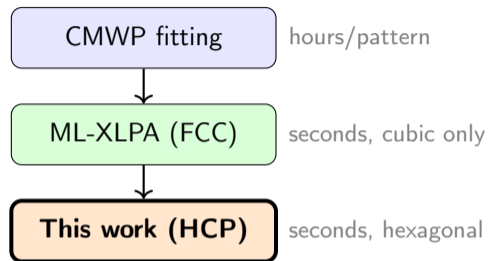
- Synchrotron XRD produces **hundreds to thousands** of diffractograms per sample
- Traditional CMWP fitting: accurate but **hours per pattern**
- ML-XLPA (Nagy et al. 2022): seconds per pattern — but only for **FCC (cubic)** crystals

The gap:

Many important materials (Ti, Zr, Mg alloys) are **HCP** — no ML-XLPA exists for them.

This project:

Build a complete ML-XLPA pipeline for HCP structures.



What Is Reused vs. What Is New

Important: all physics formulas come from the **literature** — none were derived by me. There is no automatic tool for hexagonal shapes, but the equations exist and physicists evaluate them by hand. My work is **implementing and adapting** them in code for HCP.

Reused from DifFault [Kaszás 2024]

github.com/balintkaszas/DifFault

- Peak-broadening framework $A(L) = A_S A_d A_f$
- Size term A_S
- Dislocation term A_d — Wilkens (1970)
- Cubic forward-model code structure

⇒ *The physics engine for cubic crystals*

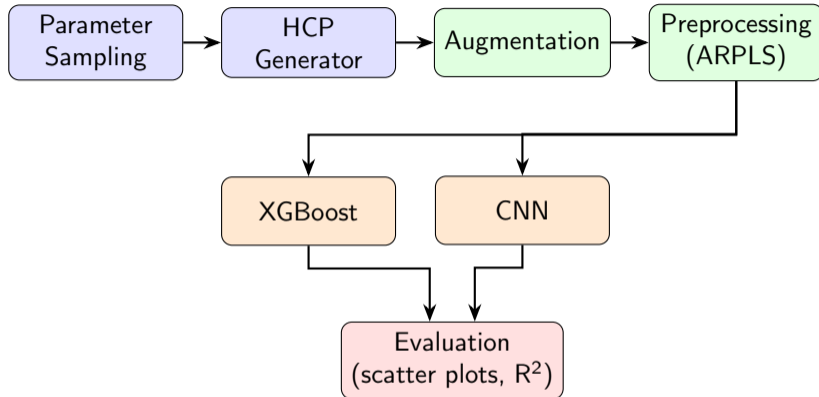
Adapted / newly implemented (HCP)

- HCP d -spacing & reflection rules
- HCP contrast factors — Dragomir & Ungár (2002)
- HCP fault model — Warren (1969)
- Full ML pipeline (12 modules)
- 6 predicted parameters (vs. 4)

⇒ *HCP forward model + end-to-end ML*

Roughly: the **cubic forward model** is inherited; the **HCP geometry, contrast factors, fault model, and the entire ML pipeline** are my implementation.

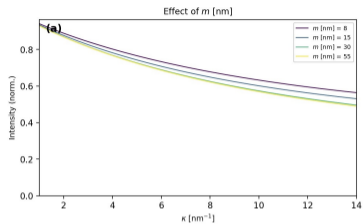
Pipeline Architecture



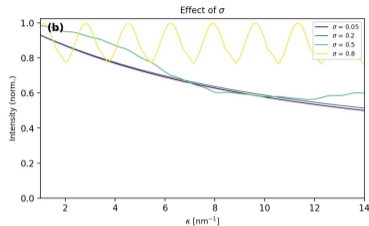
Data flows through generation → preprocessing, then into **both** XGBoost **and** CNN in parallel; both are scored by the same evaluation step.

Full code (12 modules): github.com/khachuyz/ML_XLPA

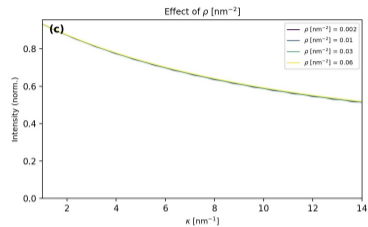
Sensitivity Analysis: What Can the Model See?



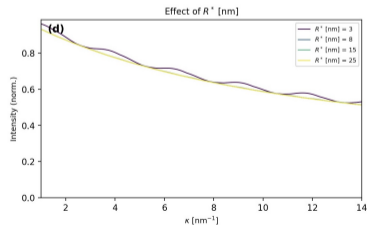
(a) Effect of m



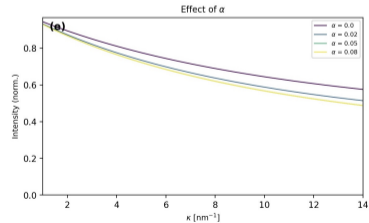
(b) Effect of σ



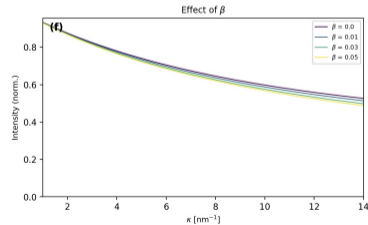
(c) Effect of ρ



(d) Effect of R^*

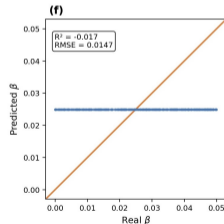
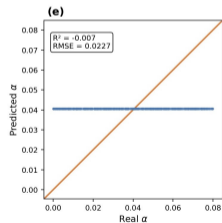
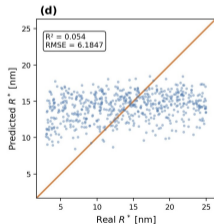
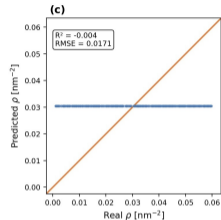
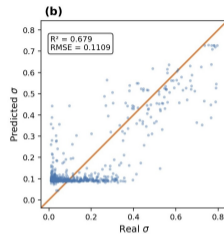
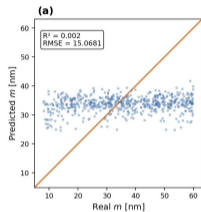


(e) Effect of α



(f) Effect of β

Results: $N = 100\,000$ Training Patterns



Results Analysis

Parameter	What it measures	Learned?	R^2
σ (size spread)	how varied the grain sizes are	Yes	0.76
α (stacking fault)	a type of planar defect	Partly	0.15
m (size median)	typical grain size	Weak	0.11
R^* (disloc. spread)	how clustered dislocations are	Weak	0.09
ρ (disloc. density)	how many dislocations	Weak	0.08
β (twin fault)	another planar defect	No	0.01

Why σ works:

- It leaves a unique fingerprint (oscillations) in the pattern

Why the rest are limited:

- m , ρ , R^* all just change the curve's slope — the model can't separate them
- HCP peaks overlap into smooth envelopes, hiding fine detail

($R^2 = 1$ is perfect; $R^2 = 0$ means no better than guessing the average. XGBoost, 300 trees.)

Current State of the Project

Delivered this semester — a complete, working pipeline:

- HCP forward model implemented & tested
- Two ML architectures: XGBoost + CNN
- Trained and evaluated on 100 000 synthetic patterns
- Best result: σ with $R^2 = 0.76$
- Publication-quality evaluation figures

12 Python modules

100K patterns trained

6 parameters predicted

Code repository:

github.com/khachuyz/ML_XLPA

Key references: [Nagy 2022] *Nanomaterials* 12, 4407 | [Kaszás 2024] *SoftwareX* 27, 101860 | [Dragomir & Ungár 2002] *J. Appl. Cryst.* 35, 556

Shift from synthetic → experimental diffractograms

1 Real data acquisition

Synchrotron XRD from HCP materials (Ti, Zr) with CMWP reference parameters

2 Domain adaptation

Bridge synthetic–real gap: instrumental broadening, real backgrounds, noise

3 Validation against CMWP

Compare ML predictions with traditional fitting on the same patterns

4 Microstructural mapping

Composition–microstructure maps on spatially resolved samples

5 Model refinement

Physics-informed features, CNN tuning, hyperparameter search

Thank you

Questions?