**Towards AI-assisted high-throughput workflow for structural analysis of thin films**

*Dmitry Lapkin1, Ainur Abukaev1, Constantin Völter1, Mikhail Romodin1, Maik Hylinski1,   
Valentin Munteanu1, Ekaterina Kneschaurek1, Vladimir Starostin1, Alexander Hinderhofer1,   
Frank Schreiber1*

1. Institut of Applied Physics, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen

2. Cluster of Excellence – Machine Learning for Science, Universität Tübingen, Maria-von-Linden-Str. 6, 72076 Tübingen

The compositional optimization of new materials necessitates the high-throughput screening of a multitude of compositions, which must be investigated to elucidate the non-linear and non-monotonic structure-property-composition dependencies [1]. In this regard, data-driven material science enables researchers to accelerate the identification of new materials with desired properties for specific applications by efficiently exploring vast material spaces. Such high-throughput data-driven studies comprise two key elements: the combinatorial preparation of suitable sample libraries spanning wide compositional ranges, and the high-throughput screening of the structure and properties of the synthesized samples [2].

Surface-sensitive X-ray scattering methods available at modern X-ray sources, e.g. Grazing-Incidence Wide-Angle Scattering (GIWAXS), comprise a convenient tool for structural investigations of thin films with high screening rate and ultimate resolution. At the same time, the determination of the crystalline structure from the collected GIWAXS patterns represents a significant bottleneck of this approach as it is time- and resource-consuming, and frequently requires additional input from other methods. This makes it an ideal use case within TA3 in DAPHNE4NFDI.

In this work, we present a data analysis workflow for high-throughput and time-resolved *in situ* structural studies of thin films using GIWAXS. The workflow includes the following steps: data conversion and correction, metadata incorporation, Bragg peak detection, fitting and indexing [3-5]. We also present a universal data format used at all stages and a GUI software for convenient visualization of the final and intermediate results at any stage.

**References:**

1. A. Ludwig, npj Comput. Mater. 5 (2019), 70, <https://doi.org/10.1038/s41524-019-0205-0>

2. J. M. Gregoire, L. Zhou, J. A. Haber, Nat. Synth. 2 (2023), 493, <https://doi.org/10.1038/s44160-023-00251-4>

3. C. Völter et al., J. Appl. Cryst. (2025), in print

4. V. Starostin et al., Synchrotron Radiat. News 35 (2022), 21, <https://doi.org/10.1080/08940886.2022.2112499>

5. V. Starostin et al., npj Comput. Mater. 8 (2022), 101, <https://doi.org/10.1038/s41524-022-00778-8>