Machine-learning-based online data analysis enables autonomous closed-loop experiments

Real-time data analysis based on machine-learning (ML) presents an important opportunity to establish closed-loop feedback systems, enables live-monitoring physical parameters beyond observables and allows for real-time decision-making during synchrotron experiments. Here, an artificial neural network, capable of considering prior knowledge, was used to extract physical thin film parameters during an X-ray reflectometry (XRR) experiment.

X-ray user facilities rank amongst the largest scientific data producers in the world, and recent advances in accelerator development and detector technology are resulting in an increasing volume of data generated in experiments. This is driving a surge in interest regarding the application of machine-learning (ML) techniques to automate data analysis. In order to prepare beamlines for ML-driven experiments, specific solutions to manage acquisition, analysis and storage have been developed at research facilities or in data-driven national and international collaborations such as DAPHNE4NFDI **[1]**, PaNOSC and ExPaNDS.

Using an X-ray reflectivity (XRR) experiment as a case study, this work presents the seamless integration of user-developed ML code with beamline control infrastructure, enabling real-time data analysis and integrated archiving of the analysed results with respect to FAIR (findable, accessible, interoperable, reusable) principles. It also demonstrates the accuracy and robustness of ML methods when applied to the analysis of XRR curves and Bragg reflections of thin film structures **[2]** through the ability to autonomously control a vacuum deposition setup.

User-developed ML code can be integrated into beamline control and data acquisition software such as BLISS **[3]** through the underlying TANGO layer **[4]** that is commonly used in beamline environments. This approach ensures high portability of the user-developed code between multiple synchrotron sources and demonstrates the interoperability of ML codes and TANGO to access entire ML models. Unlike beamline control processes, ML data analysis can run on compute resources in central computing facilities. Using VISA **[5]** – a solution for remote access to IT infrastructure for data processing – users can prepare and use IT infrastructure exclusively available to the experimental team shortly before and during specific experiments that can be customised to their needs.

For the case-study, a combined one-dimensional convolutional neural network (CNN) with subsequent multilayer perceptron was trained to extract physical thin-film parameters (thickness, density, roughness). The ML-model was used to reconstruct scattering length density (SLD) profiles in an XRR experiment on beamline **ID10**.

It is important to note that for a given SLD profile, the corresponding theoretical XRR curve can be swiftly calculated. However, reversing this operation presents a challenge because of the inherent ambiguity that often

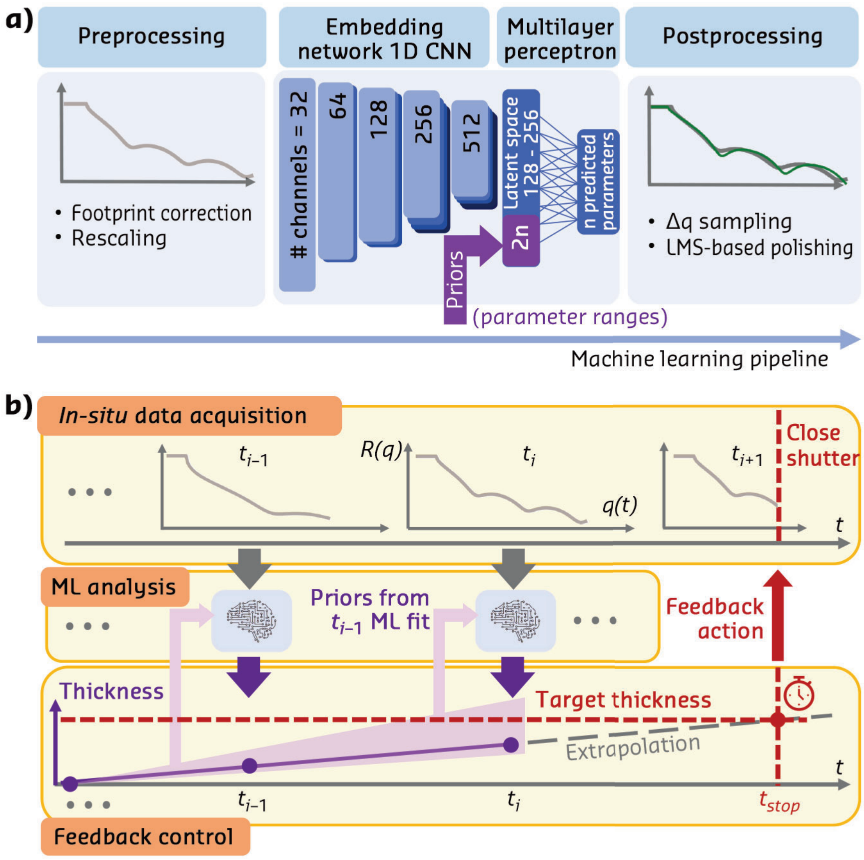

Fig. 136: a) The machine learning pipeline with special emphasis on the injection of priors at inference time. **b)** Sketch of the autonomous acquisition and feedback scheme.

allows for multiple, different SLD profiles to correspond to the same curve within the bounds of measurement uncertainty. Fundamentally, this is related to the famous phase problem of scattering. Consequently, it is vital in reflectivity analysis to make use of the physical understanding of the investigated system in order to reduce the number of potential solutions and to identify the correct one. In this work, two methods to integrate existing physical knowledge into the ML model at runtime are highlighted: physics-based parameterisation, and the including of boundaries through open parameters as additional input to the neural network (**Figure 136a**).

Molecular thin films of AlQ_3 were chosen for demonstration purposes. With the aim to grow molecular thin films of predefined thickness, an ML-based autonomous experiment took control over the growth process and terminated it once the target thickness was reached (**Figures 136b** and **137**). Prior knowledge from preceding measurements (e.g. a plausible film thickness range) was provided as input of the ML model to achieve robust fitting for a large number of consecutive scans. **Figure 137b** shows the result of the closed-loop deposition control for several target thicknesses between 80 Å and 640 Å. As expected for well-functioning closed-loop control, the target thicknesses closely matched the reached thicknesses, except for one outlier. Overall, the chosen target thicknesses could be reached within ±2 Å average accuracy. The control software BLISS was used to store the ML analysis results together with the original raw data in one NeXus-compliant hdf5 file and to interact with the facility-provided electronic notebook.

This use case convincingly demonstrates the main advantages of using ML in this context. The ML approach gives reliable fit results both for simple two- to three-layer models as well as for complex multilayer models in the millisecond regime. The combined speed and reliability of the ML approach could not be achieved by simple fitting scripts or with reliance on human supervision. More widely, ML-based online data analysis has enormous potential to make publicly available datasets FAIR through enriching the raw archived data with scientifically relevant, real-time data analysis results (data + metadata).

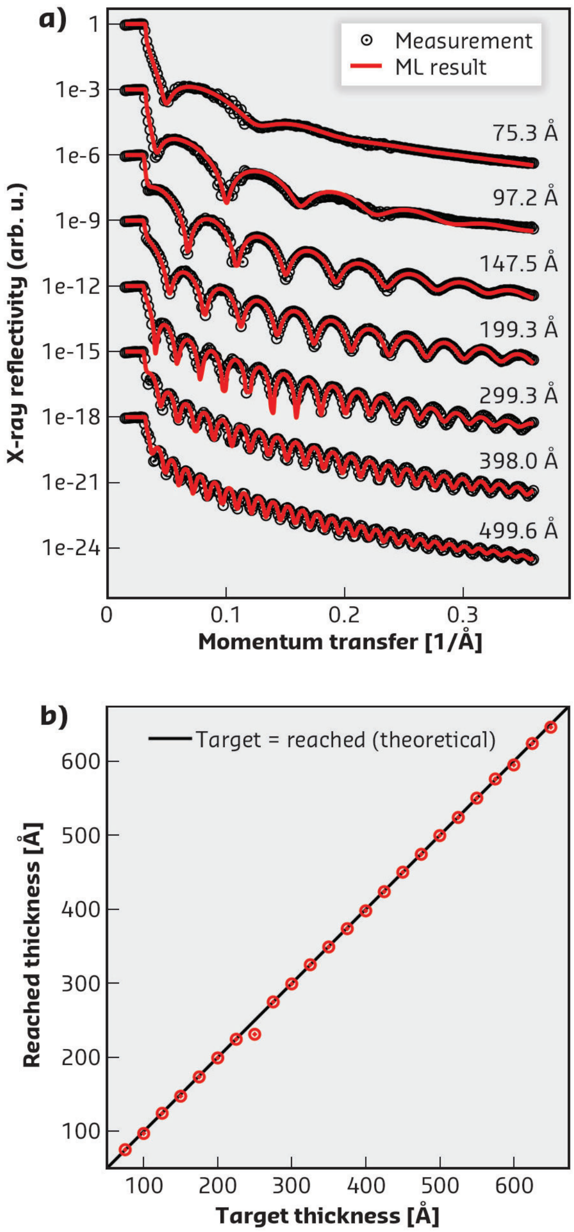

Fig. 137: a) X-ray reflectivity measurement results and corresponding fits performed on-the-fly. **b)** The target thicknesses are plotted on the x-axis, while the truly reached film thicknesses at which the deposition was terminated are given on the y-axis. In this representation, data on the diagonal line illustrates the well-functioning closed-loop experiment.

PRINCIPAL PUBLICATION AND AUTHORS

Closing the loop: autonomous experiments enabled by machine-learning-based online data analysis in synchrotron beamline environments , L. Pithan (a,e), V. Starostin (a), D. Mareček (b), L. Petersdorf (c), C. Völter (a), V. Munteanu (a), M. Jankowski (d), O. Konovalov (d), A. Gerlach (a), A. Hinderhofer (a), B. Murphy (c), S. Kowarik (b), F. Schreiber (a), *J. Synchrotron Rad.* **30**, 1064-1075 (2023); https://doi.org/10.1107/S160057752300749X
(a) University Tübingen, Tübingen (Germany)
(b) University Graz, Graz (Austria)
(c) University Kiel, Kiel (Germany)
(d) ESRF, (e) DESY

REFERENCES

[1] https://www.daphne4nfdi.de
[2] A. Hinderhofer *et al., J. Appl. Cryst.* **56**, 3-11 (2023).
[3] https://www.tango-controls.org,
[4] https://bliss.gitlab-pages.esrf.fr/bliss
[5] J.-F. Perrin, *ESRF Highlights 2022*, 158-159 (2022).