

---

# Amortized Bayesian Inference of GISAXS Data with Normalizing Flows

---

**Maksim Zhdanov**

Helmholtz-Zentrum Dresden-Rossendorf  
maxxxzdn@gmail.com

**Lisa Randolph**

European XFEL

**Thomas Kluge**

Helmholtz-Zentrum Dresden-Rossendorf

**Motoaki Nakatsutsumi**

European XFEL

**Christian Gutt**

University of Siegen

**Marina Ganeva**

Forschungszentrum Jülich

**Nico Hoffmann**

Helmholtz-Zentrum Dresden-Rossendorf

## Abstract

Grazing-Incidence Small-Angle X-ray Scattering (GISAXS) is a modern imaging technique used in material research to study nanoscale materials. Reconstruction of the parameters of an imaged object imposes an ill-posed inverse problem that is further complicated when only an in-plane GISAXS signal is available. Traditionally used inference algorithms such as Approximate Bayesian Computation (ABC) rely on computationally expensive scattering simulation software, rendering analysis highly time-consuming. We propose a simulation-based framework that combines variational auto-encoders and normalizing flows to estimate the posterior distribution of object parameters given its GISAXS data. We apply the inference pipeline to experimental data and demonstrate that our method reduces the inference cost by orders of magnitude while producing consistent results with ABC.

## 1 Introduction

**X-ray scattering** Structural and morphological properties of surfaces and multi-layer thin films can be investigated by grazing-incidence small angle X-ray scattering (GISAXS). An incident X-ray beam hits the sample under grazing incidence, after which its scattering pattern is recorded by a 2D area detector. The analysis of GISAXS patterns allows one to observe in-situ transient surface and subsurface density profiles with nanometer depth resolution. It is important for studying the growth of nanomaterials, as demonstrated by Metwalli et al. [1], and for investigating the (sub-)picosecond surface dynamics of laser-irradiated solids, as shown by Randolph et al. [2]. The reconstruction of the structural properties is complicated by multiple scattering contributions and the missing phase information. As the number of layers grows, the scattering patterns get highly complex, rendering the inference computationally infeasible without prior knowledge about imaged objects.

**Related work** Traditionally, a simulated GISAXS pattern of a sample model is generated and compared with the experimentally measured pattern. Iterative fitting processes adjust the model parameters by minimizing  $\chi^2$  until they match the experimental GISAXS pattern [2, 3]. Executing a forward model iteratively makes a comprehensive analysis of GISAXS images highly time-consuming, which in turn limits the throughput of large-scale neutron and X-ray facilities. The need for accelerated inference motivated the spread of machine learning-based algorithms [13, 14]. Cherukara et al. [4] demonstrated the real-time inversion of Wide-Angle X-ray scattering using generative neural networks yielding 500 times faster inference compared to standard iterative algorithms. Ikemoto et al. [5] and

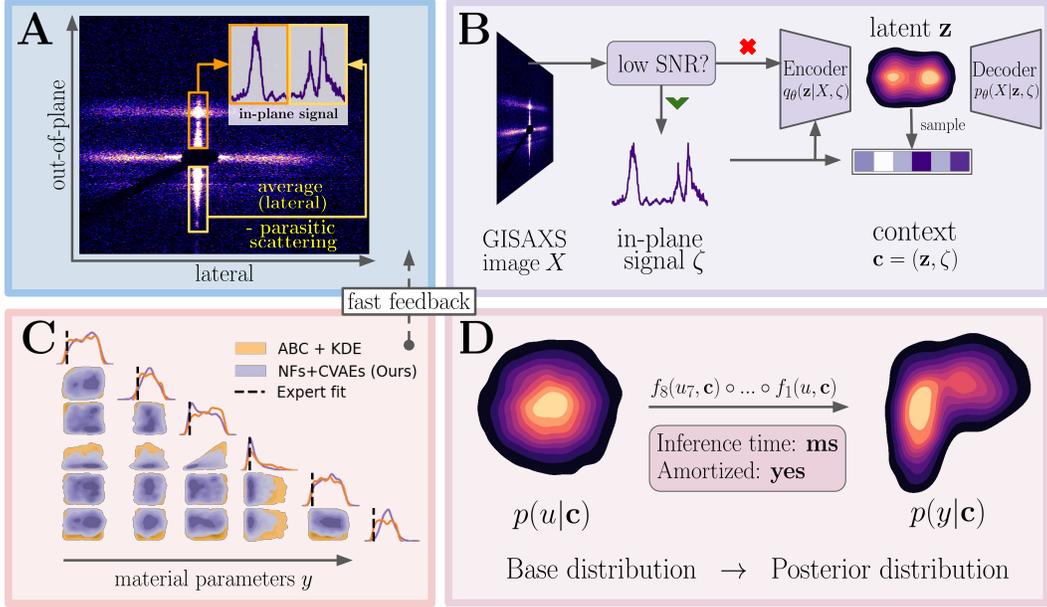


Figure 1: A schematic overview of the proposed inference pipeline for GISAXS data (A). In the first step, conditional probability  $p(X, z|\zeta)$  over GISAXS data  $X$  and latent variables  $z$  given the in-plane GISAXS signal  $\zeta$  is approximated by conditional VAEs (B). The probabilistic model allows us to compute robust representation  $\mathbf{c}$  that can be obtained even when only  $\zeta$  is given as input. Afterwards, we approximate posterior distribution  $p(y|\mathbf{c})$  over object parameters with normalizing flows (D), yielding fast inference that allows accelerating feedback during experiments (C).

Liu et al. [6] use convolutional neural networks for the one-step classification of experimental images. At the same time, Van Herck et al. [7] infer the rotation distribution of nanoparticle arrangements. Similar to our use case, Mironov et al. [8] use convolutional neural networks with uncertainty quantification to estimate film parameters from neutron reflectivity curves.

**Our contribution** In this paper, we develop an inference framework allowing for fast and robust reconstruction of GISAXS data to accelerate GISAXS data analysis. As the signal-to-noise ratio (SNR) of experimental images might suffer from distortions caused by grazing-incidence geometry [9], we mainly focus on using an in-plane scattering signal<sup>1</sup> as input. Despite the recent success of discriminative neural networks, such models do not account for the inherent ambiguity of reconstruction and, as a rule, do not provide a researcher with uncertainty quantification. Instead of learning a function from images to parameters, we use the Bayesian approach and estimate the posterior distribution of object parameters given the GISAXS data. Our framework has a two-fold structure. In the first step, we learn a robust probabilistic representation of the GISAXS generative process with variational auto-encoders [10]. Second, we model the posterior distribution via likelihood-free inference [11] with normalizing flows [12].

## 2 Methods

**Conditional variational auto-encoders** Variational autoencoders (VAE) [10] is a framework to model the data generation process via approximation of joint probability  $p(x, z)$  over observed variables  $x$  and latent variables  $z$ . It combines a generative model  $p_\theta(x|z)$ , an inference model  $q_\phi(z|x)$  and a prior  $p(z)$ , allowing unconditional data generation from a learned distribution model. Conditional VAEs (CVAEs) [15] is an extension of the framework that models a conditional distribution  $p(x, z|y)$  including an encoder  $p(z|x, y)$  and a decoder  $p(x|z, y)$ . To learn the model parameters  $\theta, \phi$ , one maximizes the conditional log-likelihood  $\log p_\theta(x|y)$  via maximizing the evidence lower bound [10, 15]. Subsequently, one can sample from conditional distribution  $p_\theta(x|y, z)$  where random noise

<sup>1</sup>We define the *in-plane scattering signal (profile)* as the average of a central region (see Fig. 1A) of an image over the lateral dimension of a detector with subtracted parasitic scattering signal from the beamstop.

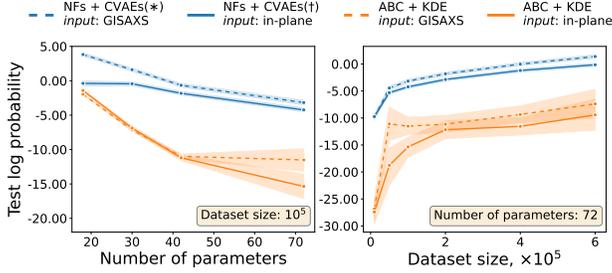


Figure 2: Log probability on test dataset for synthetic GISAXS data. Left: for multiple objects with different complexity. Right: with a varied number of seen data points during training/inference.

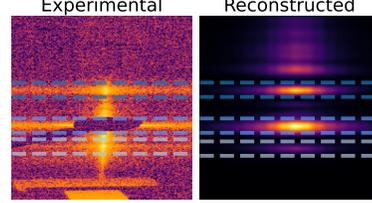


Figure 3: Reconstruction of a GISAXS image from its in-plane signal. Left: ground-truth experimental data. Right: reconstruction by CVAEs. Dashed boxes denote alignment intensity peaks.

$z$  attributes for the variance in reconstruction of  $x$  from  $y$ . We model each distribution as Normal distribution with learnable mean and variance.

**Normalizing flows** We are interested in estimating the intractable posterior distribution of unobserved variable  $y$  given an observation  $x$ . One particular class of neural density estimators that scales well to high-dimensional data is normalizing flows (NFs) [10] - the technique that was applied to inverse problems in various domains, including astronomy [17], lattice field theory [18], and high energy physics [19]. In the framework, a latent random variable  $u$  distributed according to a simple base distribution  $p(u)$  undergoes a sequence of  $K$  invertible parameterized transformations  $z_k = f_k(z_{k-1})$ , where  $z_0 = u$  and  $z_K = x$ . The framework can be further generalized to model the posterior distribution  $p(x|y)$  by conditioning each function  $f_k$  as well as the prior distribution  $p_\theta(u|y)$  to the context variable  $y$ . For a model density  $p_K(x|y)$  to approximate the true density  $p(x|y)$ , the negative log-likelihood of the observed data under the model is minimized.

**Proposed framework** We aim to estimate the posterior distribution  $p(y|X)$  of object parameters given corresponding GISAXS data  $X$ . As the likelihood of the data-generating process is intractable, we employ simulation-based inference [11] via NFs. Due to the scarcity of experimental data, we train the model on synthetic GISAXS data generated by state-of-the-art BornAgain [20] software. Although one can generate arbitrarily many simulated patterns, the distribution of these patterns can alter drastically from the distribution of experimental data. The discrepancy might render the model trained on simulated data inapplicable to real-world data, thus requiring more robust-to-noise data representations such as in-plane GISAXS signals  $\zeta$  derived from experimental images  $X$ . To generalize our framework to both modalities, we use CVAEs to learn conditional probability model  $p_{\theta_2}(X, z|\zeta)$ . Combined, the framework’s generative models allow us to estimate the posterior distribution over parameters  $y$  given  $X$  and  $\zeta$  factorized as follows:

$$p_{\theta_1, \theta_2}(y|\zeta, z, X) \stackrel{*}{=} p_{\theta_1}(y|\zeta, z) \cdot p_{\theta_2}(z|X, \zeta) \cdot p(X) \cdot p(\zeta) \stackrel{\dagger}{\approx} p_{\theta_1}(y|\zeta, z) \cdot p(z) \cdot p(\zeta) \quad (1)$$

where we use the chain rule and assume that  $y$  and  $X$  are conditionally independent given  $\zeta, z$ . Therefore, when only the in-plane signal  $\zeta$  is available, latent variables are sampled from a prior  $p(z)$  (denoted by \*). Otherwise, one uses GISAXS data to sample  $z \sim q_{\theta_2}(z|X, \zeta)$  (denoted by †). Parameters  $\theta_1$  and  $\theta_2$  are optimized by minimizing the sum of loss objectives of NFs and CVAEs. More details on the implementation and optimization procedure are provided in Appendix A.1, A.2.

### 3 Experiments

**Synthetic data** In this paper, we focus on metallic multi-layer materials consisting of multiple repetitions of Tantalum (Ta) and Copper nitride ( $\text{Cu}_3\text{N}$ ). The final goal of our research is to apply the inference pipeline to experimental GISAXS data obtained by irradiating a 12-layer sample ( $\text{TaO} - 5 \times (\text{Ta} - \text{Cu}_3\text{N}) - \text{Ta}$ ) (see Appendix A.3 for experimental details). We use BornAgain [20] to simulate synthetic labelled GISAXS patterns. We generate  $10^5$  images with varied parameters of each layer for 3-layer samples ( $\text{TaO-Ta-Cu}_3\text{N}$ ), 5-layer samples, 7-layer samples ( $\text{TaO-Ta-Cu}_3\text{N} \dots$ ) and  $6 \cdot 10^5$  images for 12-layer samples. Precisely, each layer is described by six parameters: dispersion,

Table 1: Application of our framework to experimental data. We quantify the deviation of the posterior distribution estimated by our framework from the one estimated by ABC+KDE with Wasserstein distance between marginal 1D distributions averaged across all parameters. We also provide log probability estimated by each framework for parameters obtained by Randolph et al. [2] and MAE between these parameters and values with the highest probability attributed by the corresponding framework. We also compute the inference time and dataset preparation time (500k data points) required for both frameworks for a single CPU\GPU.

	$W_p(\cdot, \text{ABC})$	Log probability	MAE	Inference time, s	Dataset prep-n, s	Amortized
ABC + KDE	-	-22.1*	0.34	$18 \cdot 10^6$ **	-	no
NFs + CVAEs	$0.06 \pm 0.04$	-14.9	<b>0.33</b>	<b>4.56</b>	$18 \cdot 10^6$	<b>yes</b>

\* The average value over multiple bandwidth values (0.15-0.3) of KDE is calculated.

\*\* If the same dataset is re-used each time, the inference time is 45 s.

absorption, thickness, roughness, Hurst parameter and lateral correlation length. During dataset generation, we uniformly sample the value of each parameter from a predefined range and simulate a GISAXS image via BornAgain. For each image, we compute the in-plane scattering signal (see Appendix A.4 for details on parameter ranges, simulation parameters and preprocessing).

**Model description** In this work, we employ two NFs+CVAEs models for inference on GISAXS data, each corresponding to a specific scenario. When a 2D GISAXS scattering signal is given, we use the first part of the equation 1 denoted by \* to reconstruct the posterior over sample parameters  $y$ . Otherwise, when only an in-plane signal is provided, we employ the second part of the equation for inference and denote the model by †.

**Baselines and evaluation** As a baseline approach, we use the combination of kernel density estimation (KDE) (Gaussian kernel, 0.2 bandwidth) and approximate Bayesian computation (ABC) [11] with Euclidean distance as distance measure and uniform distribution over predefined parameter range as prior. We set the acceptance threshold such that the acceptance rate is equal to 0.002 on average. To perform inference with ABC, we use precomputed training datasets. As an evaluation metric, we use log (natural) probability and normalized mean absolute error (MAE), i.e. scaled to a dimension range, between object parameters obtained by Randolph et al. [2] and parameters with the highest log probability assigned by an inference algorithm.

**Material complexity & dataset size** We here analyze the influence of material complexity (and hence the number of parameters) on the performance of our inference framework (see Fig. 2, left). It can be seen that neural density estimators consistently outperform ABC even with an in-plane scattering signal given as input (NFs+CVAEs(†)). Each algorithm’s performance worsens as the number of parameters increases as each layer contributes to the scattering pattern rendering highly complex GISAXS images. Besides, the deeper the layer, the weaker its contribution to the GISAXS image and hence the higher the uncertainty of an estimator. All estimators yield better performance as the number of training data points grows (see Fig. 2, right), with NFs + CVAEs demonstrating significantly higher log probability on test data than ABC for any dataset size.

**Experimental results** We apply our framework to perform inference on experimental data (see Table 1, Fig. 1C). The low value of Wasserstein distance between 1D marginal distributions indicates significant overlap between posteriors estimated by NFs+CVAEs(†) and ABC+KDE. Besides, MAE between expert fit and parameters with the highest log probability is slightly smaller for our framework. While the performance of both approaches is similar, the inference of our model is significantly faster. As it is amortized, inference time scales efficiently with the number of observations, while all the steps of ABC have to be repeated for a new observation. Therefore our model can be re-used for multiple observations on large-scale facilities, significantly accelerating data analysis. It is also possible to use CVAEs to reconstruct GISAXS image  $X$  from its in-plane scattering signal  $\zeta$  (see Fig. 3). Here, we sample latent variables  $z \sim N(0; \mathbf{1})$  and use decoder to predict image  $X$  given  $z$  and  $\zeta$ . One can see that Bragg and Yoneda peaks in both experimental and reconstructed images match well.

## 4 Conclusions

In this work, we utilized a neural density estimator to perform rapid parameter inference on GISAXS data, focusing on data with low SNR when only an in-plane scattering signal is available for inference. We propose a framework based on NFs and CVAEs that estimates object parameters' posterior distribution, given an experimental GISAXS image. We validate our approach on synthetic GISAXS data generated by BornAgain with different pattern complexity and apply it to an experimental GISAXS image obtained from complex multi-layer material. The method demonstrates inference speed-up by over six orders of magnitude compared to the traditional ABC framework enjoying GPU parallelization while yielding a consistent estimated posterior distribution. Furthermore, our method is amortized, meaning it scales efficiently when applied to many observations once trained. We expect the model to be applied to experiments on large-scale facilities that generate many independent and identically distributed observations and ultimately accelerate the analysis of GISAXS data.

### Broader impact statement

In this work, we applied a simulation-based inference framework to GISAXS data. We hope to continue developing the method to achieve the robust and reliable reconstruction of scattering data. We plan to eventually incorporate the framework into the inference pipeline on facilities where it can substantially accelerate experiments. Overall, the proposed approach might save resources and indirectly reduce the carbon footprint of scattering facilities. However, one must be careful as relying on learned density estimators requires rigorous cross-checking.

### Acknowledgments and Disclosure of Funding

This work was supported through the project *Artificial Intelligence for Neutron and X-Ray Scattering* (AINX) funded by the Helmholtz AI unit of the German Helmholtz Association. The XFEL experiments were performed at the BL2 of SACLA with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (Proposal No. 2018B8049).

### References

- [1] Metwalli, E., Körstgens, V., Schlage, K., Meier, R.J., Kaune, G., Buffet, A., Couet, S., Roth, S.V., Röhlberger, R., & Müller-Buschbaum, P. (2013). Cobalt nanoparticles growth on a block copolymer thin film: a time-resolved GISAXS study. *Langmuir : the ACS journal of surfaces and colloids*, 29 21, 6331-40 .
- [2] Randolph, L., Banjafar, M., Preston, T.R., Yabuuchi, T., Makita, M., Dover, N.P., Rödel, C., Göde, S., Inubushi, Y., Jakob, G., Kaa, J., Kon, A., Koga, J.K., Ksenzov, D., Matsuoka, T., Nishiuchi, M., Paulus, M., Schon, F., Sueda, K., Sentoku, Y., Togashi, T., Bussmann, M., Cowan, T.E., Kläui, M., Fortmann-Grote, C., Huang, L., Mancuso, A.P., Kluge, T., Gutt, C., & Nakatsutsumi, M. (2022). Nanoscale subsurface dynamics of solids upon high-intensity femtosecond laser irradiation observed by grazing-incidence x-ray scattering. *Physical Review Research* 4, 033038.
- [3] Ukleev, V., Khassanov, A., Snigireva, I., Konovalov, O.V., & Vorobiev, A.A. (2016). X-ray scattering characterization of iron oxide nanoparticles Langmuir film on water surface and on a solid substrate. *Thin Solid Films*, 616, 43-47.
- [4] Cherukara, M.J., Nashed, Y.S., & Harder, R.J. (2018). Real-time coherent diffraction inversion using deep generative networks. *Scientific Reports*, 8.
- [5] Ikemoto, H., Yamamoto, K., Touyama, H., Yamashita, D., Nakamura, M., & Okuda, H. (2020). Classification of grazing-incidence small-angle X-ray scattering patterns by convolutional neural network. *Journal of synchrotron radiation*, 27 Pt 4, 1069-1073 .
- [6] Liu, S., Melton, C.N., Venkatakrishnan, S., Pandolfi, R.J., Freychet, G., Kumar, D., Tang, H., Hexemer, A., & Ushizima, D.M. (2019). Convolutional neural networks for grazing incidence x-ray scattering patterns: thin film structure identification. *MRS Communications*.

- [7] Van Herck, W., Fisher, J., & Ganeva, M. (2021). Deep learning for x-ray or neutron scattering under grazing-incidence: extraction of distributions. *Materials Research Express*, 8.
- [8] Mironov, D., Durant, J.H., Mackenzie, R.B., & Cooper, J.F. (2021). Towards automated analysis for neutron reflectivity. *Machine Learning: Science and Technology*, 2.
- [9] Liu, J., & Yager, K.G. (2018). Unwarping GISAXS data. *IUCrJ*, 5, 737 - 752.
- [10] Kingma, D.P., & Welling, M. (2014). Auto-Encoding Variational Bayes. *CoRR*, abs/1312.6114.
- [11] Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117, 30055 - 30062.
- [12] Jimenez Rezende, D., & Mohamed, S. (2015). Variational Inference with Normalizing Flows. *ICML*.
- [13] Starostin, V., Munteanu, V., Greco, A., Kneschaurek, E., Pleli, A., Bertram, F., Gerlach, A., Hinderhofer, A., & Schreiber, F. (2022). Tracking perovskite crystallization via deep learning-based feature detection on 2D X-ray scattering data. *npj Computational Materials*, 8, 1-9.
- [14] Starostin, V., Pithan, L., Greco, A., Munteanu, V., Gerlach, A., Hinderhofer, A., & Schreiber, F. (2022). End-to-End Deep Learning Pipeline for Real-Time Processing of Surface Scattering Data at Synchrotron Facilities. *Synchrotron Radiation News*, 35, 21 - 27.
- [15] Yan, X., Yang, J., Sohn, K., & Lee, H. (2016). Attribute2Image: Conditional Image Generation from Visual Attributes. *ArXiv*, abs/1512.00570.
- [16] Kullback, S., & Leibler, R.A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, 22, 79-86.
- [17] Mishra-Sharma, S., & Cranmer, K. (2022). A neural simulation-based inference approach for characterizing the Galactic Center  $\gamma$ -ray excess. *ArXiv*, abs/2110.06931.
- [18] Haan, P.D., Rainone, C., Cheng, M.C., & Bondesan, R. (2021). Scaling Up Machine Learning For Quantum Field Theory with Equivariant Continuous Flows. *ArXiv*, abs/2110.02673.
- [19] Gao, C., Hoeche, S., Isaacson, J., Krause, C., & Schulz, H. (2020). Event generation with normalizing flows. *Physical Review D*.
- [20] Pospelov, G., Van Herck, W., Burle, J., Carmona Loaiza, J.M., Durniak, C., Fisher, J., Ganeva, M., Yurov, D., & Wuttke, J. (2020). BornAgain: software for simulating and fitting grazing-incidence small-angle scattering. *Journal of Applied Crystallography*, 53, 262 - 276.
- [21] Loshchilov, I., & Hutter, F. (2019). Decoupled Weight Decay Regularization. *ICLR*.
- [22] Papamakarios, G., Murray, I., & Pavlakou, T. (2017). Masked Autoregressive Flow for Density Estimation. *ArXiv*, abs/1705.07057.
- [23] Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. (2020). nflows: normalizing flows in PyTorch.
- [24] L. Biewald, "Experiment Tracking with Weights and Biases," *Weights & Biases*. [Online]. Available: [link](#), Accessed: 09/2022.
- [25] Yabuuchi, T., Kon, A., Inubushi, Y., Togahi, T., Sueda, K., Itoga, T., Nakajima, K., Habara, H., Kodama, R., Tomizawa, H., & Yabashi, M. (2019). An experimental platform using high-power, high-intensity optical lasers with the hard X-ray free-electron laser at SACLA1. *Journal of Synchrotron Radiation*, 26, 585 - 594.

## Checklist

1. For all authors...
  - (a) Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope? Yes
  - (b) Did you describe the limitations of your work? No
  - (c) Did you discuss any potential negative societal impacts of your work? Yes
  - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. If you are including theoretical results...
  - (a) Did you state the full set of assumptions of all theoretical results? No theoretical results are included.
  - (b) Did you include complete proofs of all theoretical results? No theoretical results are included.
3. If you ran experiments...
  - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes, we provide a link to our GitHub repository.
  - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, see Appendix A.1, A.2.
  - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes
  - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes, see Appendix A.2.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
  - (a) If your work uses existing assets, did you cite the creators? Yes
  - (b) Did you mention the license of the assets? Yes
  - (c) Did you include any new assets either in the supplemental material or as a URL? No
  - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? Yes
  - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes
5. If you used crowdsourcing or conducted research with human subjects...
  - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? Not applicable
  - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? Not applicable
  - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? Not applicable

## A Appendix

### A.1 Model implementation

Similarly to the model of Mishra-Sharma et al. [17], we employ masked autoregressive flows [22] as the transformation flow with 8 transformations made of a 2-layer masked autoregressive MLP with a hidden layer size of 128. The model was implemented with *nflows* library [23]. Model parameters are optimized by minimizing the following objective:

$$\mathcal{L}_{NFs}(x, y; \theta) = -\log p_K(x|y) = -\log p(f^{-1}(x, y)) - \sum_{k=1}^K \log \left| \det \frac{dz_k}{dz_{k-1}} \right| \quad (2)$$

To build CVAEs, we use a convolution neural network-based encoder and decoder that are reversed (transposed) versions of one another. Encoder/decoder has 6 blocks where each block consists of a

convolutional operator with kernel size 3, batch normalization and SiLU activation. Additionally, we have a conditional encoder - a 2-layer MLP with 128 hidden units. The loss function of CVAEs is defined as follows:

$$\mathcal{L}_{CVAEs}(x, y; \theta, \phi) = -KL(q_\phi(z|x, y) || p(z)) + \mathbb{E}_{q_\phi(z|x, y)} [\log p_\theta(x|y, z)] \quad (3)$$

where  $KL$  denotes the Kullback–Leibler divergence [16]. Thus, we minimize the sum of loss objectives to train the parameters of the framework:

$$\mathcal{L}_{NFs+CVAEs}(y, \zeta, z, X; \theta_1, \theta_2) = \mathcal{L}_{NFs}(y, (\zeta, z); \theta_1) + \mathcal{L}_{CVAEs}(X, \zeta; \theta_2) \quad (4)$$

Hyperparameters of our framework, such as the number of hidden units and learning rate, were optimized by grid search using *wandb* [24].

## A.2 Optimization procedure

We optimize parameters of each model separately from one another by minimizing the corresponding loss objective (Eq. 3, 2). We reserve 20% of training data for validation and 1000 data points for testing. For each model, AdamW [21] optimizer is used with initial learning rate  $10^{-3}$  and weight decay  $10^{-2}$ . CVAEs are trained for 30 epochs with early stopping and learning rate decaying by the factor of 0.1 every 10 epochs. We train normalizing flows for 100 epochs with early stopping and decreasing learning rate by 0.1 every 30 epochs. For each model, the training was performed in mini-batches of size 32. We use NVIDIA Tesla V100 GPU from the Hemera HPC system of HZDR.

## A.3 Experimental data

We use the same experimental data as Randolph et al. [2]. The experimental multi-layer material was grown on a 700  $\mu\text{m}$  thick silicon wafer carrying a Ta seed layer on a 100 nm thick layer of thermal silicon. The experimental GISAXS images have been measured at the EH6 station at the SACLA XFEL facility in Japan [25] by irradiating the multi-layer sample with 8.81 keV X-ray photons under the grazing incidence angle of  $0.64^\circ$  and recording the scattered intensity with a 2D-area detector (MPCCD). A beamstop has blocked the strong specular peak as it is orders of magnitude more intense than the diffusely scattered signal.

## A.4 BornAgain simulation

BornAgain software uses the DWBA formalism to synthesize GISAXS patterns for a given object and experimental setup. We use multi-layer objects, describing each layer by characteristic properties (see Table 3). We access BornAgain via Python script, which can be found in our GitHub repository. We indicate the experimental parameters used in each simulation in Table 2. Table 3 presents the sample parameters with corresponding ranges from which we sampled uniformly when generating datasets.

Table 2: Experimental setup

Experimental parameters	
pixel size	50 $\mu\text{m}$
pixel y-direction	512
distance in y-direction	25.6 mm
pixel z-direction	1024
distance in z-direction	51.2 mm
sample detector distance	1277 mm
y-position of specular reflection	10.75 mm
z-position of specular reflection	20.65 mm
wavelength	0.14073 nm
incident angle	$0.64^\circ$
azimuthal angle	$0.0^\circ$
beam intensity	$10^{13}$
constant background	60

Table 3: Parameter ranges

Layer parameters	range (Ta)	range (Cu <sub>3</sub> N)
dispersion, $\cdot 10^{-5}$	1 - 4	0.8 - 3
absorption, $\cdot 10^{-7}$	0.1 - 90	0.1 - 90
thickness, nm	0.3 - 4.7	0.3 - 12.0
roughness, nm	0.3 - 5	0.3 - 5
Hurst parameter	0.1 - 0.9	0.1 - 0.9
lateral correlation length, nm	3 - 30	3 - 30

## A.5 In-plane GISAXS signal computation

To obtain an in-plane signal, we first take only the center cut of 30 pixels along the lateral dimension of a detector around the specular beam as done for the experimental data [2]. We then average over

the lateral dimension of the detector and split the resulting curve into two parts: before and after the beam stop. Next, we normalize each part and concatenate them yielding an array of length 359. We subtract an underlying background signal related to parasitic scattering from the beamstop for experimental data. The parasitic scattering model was taken from [2].