APPLIED PHYSICS

Fast and reliable probabilistic reflectometry inversion with prior-amortized neural posterior estimation

Vladimir Starostin[1]*, Maximilian Dax[2]†‡, Alexander Gerlach[3], Alexander Hinderhofer[3], Álvaro Tejero-Cantero[1], Frank Schreiber[3]*

Reconstructing the structure of thin films and multilayers from measurements of scattered x-rays or neutrons is key to progress in physics, chemistry, and biology. However, finding all structures compatible with reflectometry data is computationally prohibitive for standard algorithms, which typically results in unreliable analysis with only a single potential solution identified. We address this lack of reliability with a probabilistic deep learning method that identifies all realistic structures in seconds, redefining standards in reflectometry. Our method, prior-amortized neural posterior estimation (PANPE), combines simulation-based inference with adaptive priors that inform the inference network about known structural properties and controllable experimental conditions. PANPE networks support key scenarios such as high-throughput sample characterization, real-time monitoring of evolving structures, or the corefinement of several experimental datasets and can be adapted to provide fast, reliable, and flexible inference across many other inverse problems.

INTRODUCTION

Scattering techniques enable the reconstruction of object structures through the analysis of scattered radiation (*1*, *2*). At the nanoscale, this requires radiation with short wavelengths, such as x-rays and thermal neutrons. While for the reconstruction of images from visible scattered light there are more established tools including optical lenses, using these tools for x-rays and neutrons frequently poses substantial challenges, leading to the use of algorithms for the reconstruction process (*3*). These algorithms, however, receive incomplete information, as detectors capture intensities but not the phase information of the scattered waves. This gives rise to the phaseless inverse problem in scattering physics. While physical models can simulate scattered intensities from a given structure, reconstructing the structure from actual measurements is analytically intractable, and experimental data can be consistent with multiple physical structures (*4*). This ambiguity can be then resolved through complementary measurements or physical knowledge, but it is crucial to first acknowledge the existence of multiple solutions to avoid costly misinterpretations of the data. Together with advances in experimental methods enabling time-resolved online experiments and high-throughput pipelines (*5*, *6*), this creates a pressing need for algorithms that are fast, capable of reliably identifying all possible solutions, and flexible enough to integrate additional data and physics-informed constraints.

The need of fast and reliable algorithms is especially evident for neutron and x-ray reflectometry (XRR) (*7*–*9*). The reflected intensity R in specular geometry as a function of momentum transfer q (see Fig. 1A) can inform about the scattering length density (SLD) profile for a broad range of thin films and layered structures, ranging from solar cells (*10*) to biological membranes (*11*, *12*). The SLD profile is typically modeled by parameters θ that include layer thicknesses d_l, densities ρ_l, and interface roughnesses σ_l. Obtaining these parameters θ from a reflectivity curve $R(q)$ in a fast and reliable way is a long-standing inverse problem (Fig. 1B) due to the phase loss, measurement noise, and limited range and resolution of q. For a long time, a common approach was to search for the "best" single set of parameters θ^* maximizing the likelihood of the measured data. However, maximum likelihood estimation remains fundamentally unreliable as it overlooks other potential physical solutions arising from ambiguity in the inverse problem. To address ambiguity, we need to embark on a principled probabilistic approach and estimate the posterior probability density of the parameters θ given the measured data $\mathbf{R}$. In such a Bayesian posterior $p(\theta \mid \mathbf{R})$, different probable structures appear as distributional modes (Fig. 1C). In practice, the inference of a high-dimensional posterior is inherently challenging and particularly so in reflectometry where multiple narrow distributional modes are common. Conventional Bayesian likelihood-based techniques such as Markov Chain Monte Carlo (MCMC) (*13*) are neither fast nor reliable, as they generally miss distributional modes.

Here, we present a machine learning solution for Bayesian reflectometry analysis that provides fast, reliable, and accurate inference along with the flexibility that online experiments demand. Speed is achieved by pretraining a neural network across large amounts of representative data, an amortization procedure that then allows for real-time inference on new samples. Reliability stems from the use of recent simulation-based inference (SBI) which, in contrast to likelihood-based modes, provides comprehensive coverage of the search space. Accuracy in the identified solutions is achieved via a subsequent likelihood-based step, which refines machine learning–based estimates. Fast likelihood evaluations are possible thanks to our PyTorch implementation of transfer-matrix simulator. Last, we enable great flexibility for a wide range of standard scenarios in experimental setups by extending neural posterior estimation (NPE) with prior amortization, which we term prior-amortized neural posterior estimation (PANPE). Our method PANPE can use dynamically set, adaptive prior distributions, allowing to track online expriments, leverages equivariance transformations to enable amortization over different q ranges, and can combine evidence from multiple measurements for efficient inference. Below, we describe in

[1] Cluster of Excellence Machine Learning for Science, University of Tübingen, Tübingen, Germany. [2] Max Planck Institute for Intelligent Systems, Tübingen, Germany. [3] Institute of Applied Physics, University of Tübingen, Tübingen, Germany.
*Corresponding author. Email: vladimir.starostin@uni-tuebingen.de (V.S.); frank.schreiber@uni-tuebingen.de (F.S.)
†Present address: ETH Zurich, Zurich, Switzerland.
‡Present address: ELLIS Institute Tübingen, Tübingen, Germany.

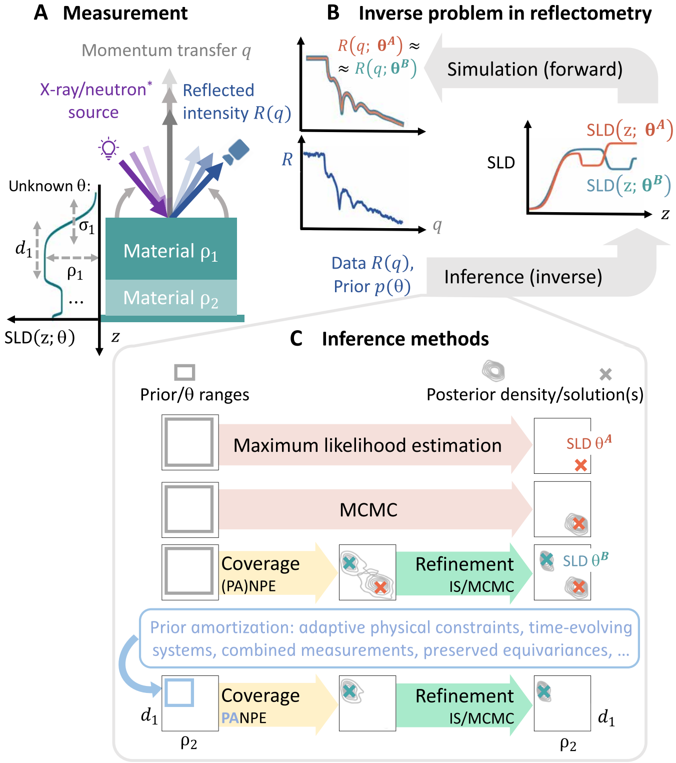

Fig. 1. The ill-posed inverse problem in reflectometry analysis. (**A**) A schematic experimental setup for reflectometry measurements. The reflected intensity $R(q)$ from a studied layered structure as a function of momentum transfer q contains information about parameters θ of the studied sample. The momentum transfer is typically controlled by the geometry in XRR or by the energy in time-of-flight neutron measurements. (**B**) Inverse problem in reflectometry: The forward simulations provided by the scattering theory should be inverted during inference, which is generally ambiguous. (**C**) Inference methods commonly used for reflectometry analysis, as well as our proposed approach. The standard maximum likelihood estimation approach provides a single solution by design. MCMC locally explores the parameter space and can overlook distributional modes. In contrast, (PA)NPE posterior estimate is guaranteed to cover all the solutions, with further refinement based on likelihood evaluation improving accuracy. Our prior amortization method, PANPE, enables the analysis of multiple experimental scenarios using a single neural network.

detail how PANPE works and benchmark its performance on real and synthetic reflectometry data.

RESULTS

Overview of PANPE

Bayesian framework for inverse problems

Reflectometry analysis aims to infer physical parameters θ from measured data **R**. Each parameter set θ describes a hypothetical SLD profile of the studied structure (SLD parameterization is discussed in Materials and Methods). For a given reflectometry measurement **R**, the Bayesian posterior distribution (*14*)

$$p(\theta \,|\, \mathbf{R}) \propto p(\mathbf{R} \,|\, \theta) p(\theta) \tag{1}$$

offers a probabilistic estimate of θ, characterized by the likelihood $p(\mathbf{R} \,|\, \theta)$ provided by scattering theory and a prior $p(\theta)$ provided by experimentalists.

The prior physical knowledge about the studied structure, formulated as a prior distribution $p(\theta)$ over parameters θ in Bayesian framework, serves as a crucial tool for resolving ambiguity in reflectometry and, more broadly, in scattering physics. It facilitates physics-informed analysis by integrating knowledge about the materials used and other properties of the system under study. What may be unphysical SLD profiles in one context can be considered legitimate solutions in another, depending on the known properties of the system being investigated.

With Bayes's theorem, we can use likelihood and prior to calculate the (unnormalized) posterior density for any given parameter set θ_i. However, such density evaluation does not directly provide most practically relevant estimates, such as mean values or confidence intervals. To compute them, we need to draw samples $\{\theta_i\}_{i=1}^N \sim p(\theta \,|\, \mathbf{R})$ from the posterior distribution, i.e., randomly selecting parameter sets θ_i in proportion to their posterior density $p\left(\theta_i \,|\, \mathbf{R}\right)$. The complexity of this operation for high-dimensional parameters θ renders Bayesian inference a highly challenging task, traditionally limiting the applicability of the Bayesian approach to simple cases.

Coverage and refinement of NPE

To reliably sample from the Bayesian posterior, we adopt a recent NPE method (*15*, *16*). It uses a normalizing flow architecture (*17*–*18*) as the neural network–based density estimator. Normalizing flows can learn complex high-dimensional conditional distributions and have been used for Bayesian inference in multiple applications. Once trained on a broad range of simulated data, the flow-based model $p_{\mathrm{NN}}(\theta \,|\, \mathbf{R})$ can efficiently generate samples $\{\theta_i\}_{i=1}^N \sim p_{\mathrm{NN}}(\theta \,|\, \mathbf{R})$ and evaluate densities $p_i = p_{\mathrm{NN}}\left(\theta_i \,|\, \mathbf{R}\right)$ for different measurements **R**.

A key property of normalizing flows is that the exact density evaluation enables training the model by minimizing the forward Kullback-Leibler (KL) divergence (see more details in Materials and Methods). This ensures the coverage property of NPE, i.e., it contains the full support of the true unknown posterior $p(\theta \,|\, \mathbf{R})$ and does not miss distributional modes (*19*). The coverage is guaranteed when the model is sufficiently expressive and the training data capture all relevant modes, as ensured by sampling from the joint distribution $\theta_i, \mathbf{R}_i \sim p(\theta, \mathbf{R})$.

Despite the coverage property, the "shape" of the NN-based density estimate might deviate from the target posterior. To address this, likelihood-based methods such as importance sampling (IS) and MCMC can refine the NPE results for more accurate estimates. Thus, NPE ensures the coverage property, while likelihood-based refinement enhances accuracy.

Our custom-made GPU-accelerated transfer-matrix reflectometry simulator (*20*) implemented using PyTorch (*21*) accelerates both the training and inference stages. This allows us to simulate new curves directly during the training process without reusing simulations, thereby preventing overfitting. Furthermore, as we show below, our reflectometry analysis with IS refinement—which requires likelihood evaluations—typically takes just seconds to less than a minute on a single graphics card, the NVIDIA RTX 2080 Ti.

Amortization across various experimental scenarios

Along with reliability due to the coverage property, NPE provides fast, amortized inference by shifting its computational cost to the training phase. However, the amortization also introduces a key practical limitation, as a trained model can only operate within some predefined training ranges of parameters. In the case of reflectometry, this limitation includes not only the parameter ranges but also experimental settings such as the discretization of the momentum transfer axis q or the measurement uncertainties, all of which may substantially affect the resulting posterior estimate. We address these limitations by implementing intensive amortization, taking into account the diverse varying aspects of the experiment, such as q discretization and range, measurement uncertainties, and prior information.

Measurements of the reflected intensity $R(q)$ are taken at different discrete values of momentum transfer q (see Fig. 1A). Together, these measurements form an input dataset $\mathbf{R} = \{q_p, R(q_p), s_p\}_{p=1}^{n_q}$ consisting of n_q measured points (here s_p represents the uncertainty of each data point). Our approach accommodates experiments featuring arbitrarily spaced q points and varying numbers of measurements n_q, by using an efficient embedding neural network equipped with trainable interpolation kernels (see more details in Materials and Methods).

However, in the context of reflectometry analysis, the most notable variable component is the prior information. Known constraints on physical properties vary substantially across different structures under study. Furthermore, in the online experiments discussed below, the experimentalists may modify the structure by changing control parameters—and accordingly, the respective priors. Adjustments to the priors are also necessary when combining multiple measurements in neutron reflectometry (NR), as discussed below. In these and other scenarios, the analysis must adapt to the changing prior distribution. Standard machine learning solutions like NPE, which typically assume a fixed prior distribution, fall short under these experimental conditions. To overcome this limitation, we introduce PANPE that accommodates a variety of prior distributions within a single model.

Prior-amortized neural posterior estimation

We incorporate dynamic prior information into a neural network by choosing a class of distributions $p(\theta \mid \phi)$ parameterized by ϕ. The newly introduced parameters ϕ reflect prior information about the system. They are supplied as an additional input to the flow-based neural network $p_{\mathrm{NN}}(\theta \mid \mathbf{R}, \phi)$ alongside the measured data. This allows us to train a single neural network and amortize inference across both measurements $\mathbf{R}$ and priors $p(\theta \mid \phi)$

$$p_{\mathrm{NN}}(\theta \mid \mathbf{R}, \phi) \approx p(\theta \mid \mathbf{R}, \phi) \propto p(\mathbf{R} \mid \theta) p(\theta \mid \phi) \tag{2}$$

In reflectometry analysis, it is typically sufficient to use uniform prior distributions $p(\theta) = \prod_{j=1}^{n} U\left(\theta_j^{\min}, \theta_j^{\max}\right)$, where n is the number of parameters θ. Thus, we define ϕ as a set of corresponding parameter ranges: $\phi = \{\theta_j^{\min}, \theta_j^{\max}\}_{j=1}^{n}$. This results in $2n = 20$ additional input values for a task with $n = 10$ parameters θ. In this manner, the model is trained to approximate posterior distribution for a continuous set $p(\phi)$ of (uniform) prior distributions $p(\theta \mid \phi)$ within a larger parameter space. Our approach can be extended to other classes of distributions by providing suitable parameterization. We illustrate why the likelihood-based refinement (or rejection sampling in the case of uniform prior distribution) is not a practical alternative to the prior amortization in our case on simulated data below. Our prior amortization approach is discussed in detail in Materials and Methods.

The inference pipeline is illustrated in fig. S1. Given the data $\mathbf{R}$ and the prior distribution characterized by parameters ϕ, we sample from the trained PANPE model and apply likelihood-based refinement either using IS (*19*) (PANPE-IS) or MCMC (PANPE-MCMC) (see more details in Materials and Methods).

Parameter-conditioned posterior estimation

In certain cases, it is required to estimate parameter-conditioned posterior estimation, where instead of providing narrow priors for a parameter, it is fixed. Here, we show one such case in the context of NR where parameter-conditioned posterior estimation is necessary for combining multiple measurements with partially shared parameters.

Another scenario relevant for future reflectometry applications considers choosing the appropriate physical model: By setting the thicknesses of a subset of layers to zero, one can effectively change the number of layers in the physical model. Consequently, several physical models represented by different numbers of free parameters can be compared via standard criteria such as the Bayes factor $\frac{p(\mathbf{R}|\phi_1)}{p(\mathbf{R}|\phi_2)}$ using the same neural network.

Fixing some parameters changes the dimensionality of the remaining free parameters, which is not supported by standard implementations of the normalizing flow. To circumvent this limitation, we introduce a reparameterization procedure of the parameter space that enables us to sample from the parameter-conditioned posterior estimation by providing "zero-width" prior. We discuss this approach in Materials and Methods and use it to analyze NR data below.

Preserved equivariances in the density estimator

The reflectometry simulator features a number of simple deterministic functional relationships between the input $\{q, \theta\}$ and the simulated reflectivity curve $R(q, \theta)$. We systematically review them in Materials and Methods. For instance, these include the unit scaling invariance: Rescaling the momentum transfer axis $\mathbf{q} \to u \cdot \mathbf{q}$ ($u \in \mathbb{R}_{>0}$) together with a certain parameter rescaling transformation $\theta \to T_u(\theta)$ does not alter the result: $R(u \cdot \mathbf{q}, T_u(\theta)) = R(\mathbf{q}, \theta)$. Conventional reflectometry analysis does not need to consider these relationships, but they become critical in amortized machine learning solutions: The trained model must reflect these relationships, ensuring that specific changes in the input data to the density estimator result in corresponding changes in the posterior distribution. To enhance the performance of the model, we directly incorporate these relationships into the inference pipeline (see more details in Materials and Methods), rather than having the model learn them from data. We note that the prior amortization is generally required for this operation, since the considered transformations alter the prior distribution.

Related work

Sequential NPE (*22*–*25*) effectively applies data-informed prior updates, but such methods require simulation and neural network training at inference time. The Simformer (*26*) framework enables prior updates by leveraging diffusion guidance. Equivariances between parameter and data spaces can be integrated with group-equivariant NPE (*27*, *28*), but this method requires iterative inference and is thereby slower than NPE. The Dingo-BNS framework (*29*) for gravitational-wave inference combines adaptive priors with amortized NPE to achieve improved data compression. A recent work (*30*) uses adaptive priors for sensitivity-aware amortized Bayesian inference. Our study builds on and extends upon these works, demonstrating how adaptive SBI priors enable zooming into arbitrary parts of the parameter space in a high-profile science application.

Performance on simulated data

Multiple modes in the posterior

Figure 2 showcases the inference results obtained with PANPE-IS for a curve simulated from a two-layer structure. For reflectometry,

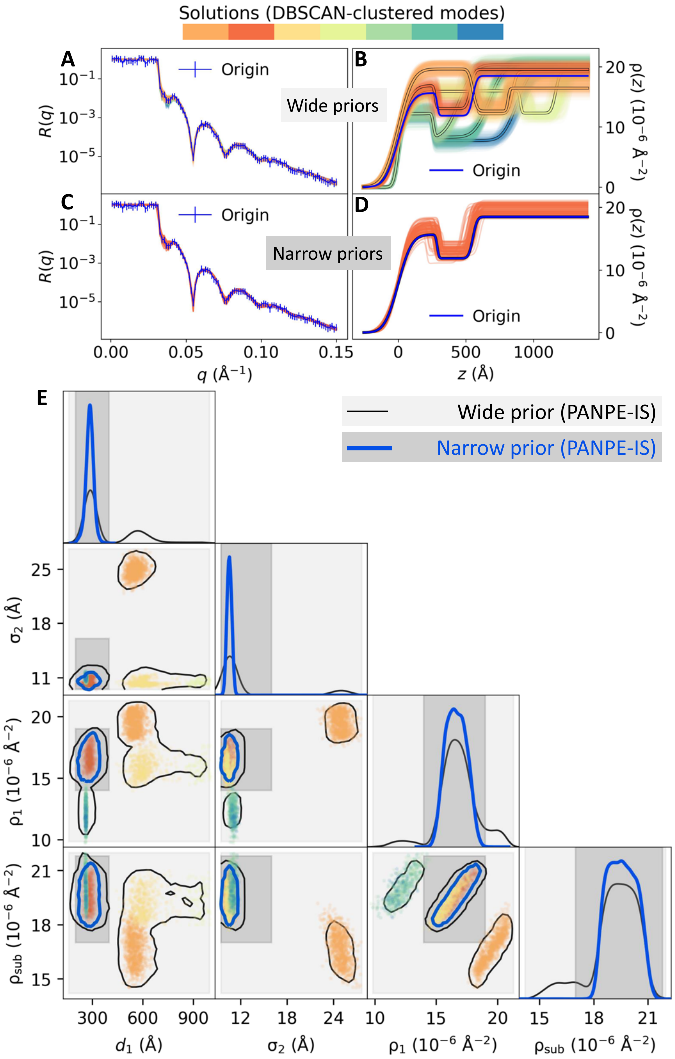

Fig. 2. Multimodal posterior distribution obtained by PANPE-IS on a simulated reflectivity curve with 10 free parameters for a two-layer structure. The neural network produces results in accordance with the provided prior information, identifying (**A**) multiple solutions for a "wide" prior distribution and (**C**) a single distributional mode for a "narrow" prior (gray dashed lines). Colors denote distinct distributional modes obtained by clustering samples. The corresponding reflectivity curves (**B**) and (**D**) enable real-time likelihood-based refinement, resulting in accurate posterior estimation. The corner plot (**E**) shows the resulting marginalized 4d distributions obtained for both priors along with the colored samples related to the colored profiles in (A).

generated samples $\{\theta_i\}_{i=1}^{N}$ represent SLD profiles that could potentially produce the measured data according to the neural network. Figure 2 (A and B) shows the reflectometry curve colored in blue on the left and $N = 1000$ colored SLD profiles obtained from our model on the right with a wide prior distribution. The colors indicate seven distinct solutions (distributional modes) separated via DBSCAN clustering for better visualization. The blue SLD profile represents the "true" structure used to simulate the reflectometry curve. The forward reflectometry simulations enable immediate visual validation of the result. In Fig. 2A, the studied blue reflectometry curve is superimposed with the colored curves simulated from the NN-produced SLD profiles. The colored reflectometry curves are mostly invisible since they overlap very well with the original curve and each other, despite a diverse variety of corresponding SLD profiles. Figure 2E provides an alternative visualization of samples and the estimated posterior on a corner plot. For visual purposes, only 4 of 10 parameters are shown here.

Efficiency gain from prior amortization

Figure 2 (C and D) shows the inference result for the same reflectivity curve but with narrower prior distribution, resulting in a single mode. In this case, the prior amortization allows excluding all the samples outside the domain of a more informative prior distribution.

The necessity for prior amortization might not be immediately obvious when likelihood evaluation is fast. A viable alternative could seem to be training a standard NPE model across a wide parameter range without prior amortization and refining results via likelihood-based methods later. For uniform priors, calculations of likelihood are not even required: Samples outside the prior domain can be simply rejected without the need of likelihood evaluation. However, the main issue with rejection sampling is its low acceptance rate. In practice, this quantity can be exceedingly small. In our case, an acceptance rate of less than one in a million applies to about 70% of the synthetic test data. Consequently, to obtain a single sample within the prior domain, an immense number of samples would need to be generated through neural network evaluations, making this approach essentially inapplicable. We also illustrate this problem on an experimental XRR data below.

Sample efficiencies on the simulated dataset

We evaluate PANPE-IS on a set of 1000 simulated test samples. These samples are generated following the same procedure as outlined for the training data in Materials and Methods. Each curve has different q discretization and is accompanied by its own prior distribution $p(\theta\,|\,\phi)$.

We assess the performance of the model on each test sample by evaluating its sample efficiency ϵ_{eff}

$$\epsilon_{\text{eff}} = \frac{\left(\overline{w_i}\right)^2}{\left(\overline{w_i^2}\right)}, \quad w_i = \frac{p(\mathbf{R}\,|\,\theta_i)\,p(\theta_i\,|\,\phi)}{p_{\text{NN}}(\theta_i\,|\,\mathbf{R},\phi)}, \quad \theta_i \sim p_{\text{NN}}(\theta\,|\,\mathbf{R},\phi) \quad (3)$$

where w_i are the importance weights. In practice, during inference, importance weights can be used for Monte Carlo estimates $\mathbb{E}_{\theta\sim p(\theta|\mathbf{R},\phi)}[f(\theta)] \approx \left(\sum_{i=1}^{N} f(\theta_i)\,w_i\right) / \left(\sum_{i=1}^{N} w_i\right)$. The sample efficiency ϵ_{eff} effectively quantifies the efficient sample size (ESS) $= N \cdot \epsilon_{\text{eff}}$ as a share of the total number of samples and, therefore, determines the time required for obtaining a desired ESS. By using our efficient reflectometry simulator, we are able to obtain ESS $= 100$ for $\epsilon_{\text{eff}} = 10^{-4}$ for less than a minute, but the same ESS would take more than a month for $\epsilon_{\text{eff}} = 10^{-10}$. In practice, only solutions within high probability mass regions are typically relevant for analysis, so several hundred efficient samples are generally sufficient to identify them.

Each test sample is characterized by its prior distribution, which results in different complexity of the inference task: Wider prior distributions that simulate the cases of higher uncertainty about the studied structure are generally more complex to analyze using conventional methods. We quantify this complexity through the sample efficiency of the conventional IS method with prior acting as a proposal distribution. In this case, it replaces the "neural" proposal distribution in Eq. 3, and the respective importance weights simply equal to the likelihood $w_i^{\mathrm{IS}} = p\left(\mathbf{R} \mid \theta_i\right)$, where $\theta_i \sim p(\theta \mid \phi)$. We discuss how we estimate low sample efficiencies for IS in Materials and Methods.

Figure 3 shows sample efficiency distribution of conventional IS on the left-hand side and our PANPE-IS on the right hand side, with lines in the middle connecting the same test samples and indicating the difference in sample efficiency between the two methods. Blue color corresponds to the synthetic test dataset (we discuss the experimental data in the next section). The axis on the right shows an average time required to obtain ESS = 100 for different sample efficiencies on a single graphics card, NVIDIA GeForce RTX 2080 Ti, with our GPU-accelerated reflectometry simulator. It shows that our PANPE model can perform inference in under a minute where the conventional IS approach would require days and even months of computation.

Adjustments to the prior distribution for the same data can influence sample efficiency of PANPE in certain scenarios. Preliminary tests indicate that modifying the width of the prior does not substantially affect efficiency as long as high-density regions of the parameter space remain included. However, narrowing the prior to focus on a single mode by excluding other high-density regions can enhance sample efficiency, as normalizing flows more effectively approximate unimodal distributions. This behavior reflects the model's ability to efficiently learn parameter dependencies across a range of prior specifications, maintaining robustness and adaptability across different experimental setups.

Evaluating PANPE performance without refinement

In addition, we evaluate the performance of raw PANPE estimates without likelihood-based refinement. Specifically, we evaluate the quality of marginal distributions, i.e., one-dimensional distributions $p_{\mathrm{NN}}\left(\theta_j \mid \mathbf{R}, \phi\right)$ for each jth parameter. For that, we perform standard Kolmogorov-Smirnov tests that use the true parameters θ used for simulating the test data (*27*). These tests determine whether the true parameters could realistically be sampled from PANPE-generated marginal distributions by checking if their percentile scores are uniformly distributed. Figure S1 shows the p-p plots, and the obtained *P* values demonstrate satisfactory performance on simulated data. These findings imply that one can rely on raw PANPE estimates derived from marginal distributions like means and variances for analysis. However, in the context of reflectometry, we always apply likelihood-based refinement as it is cost-effective and enhances the accuracy of our estimates while also providing a means to evaluate their quality.

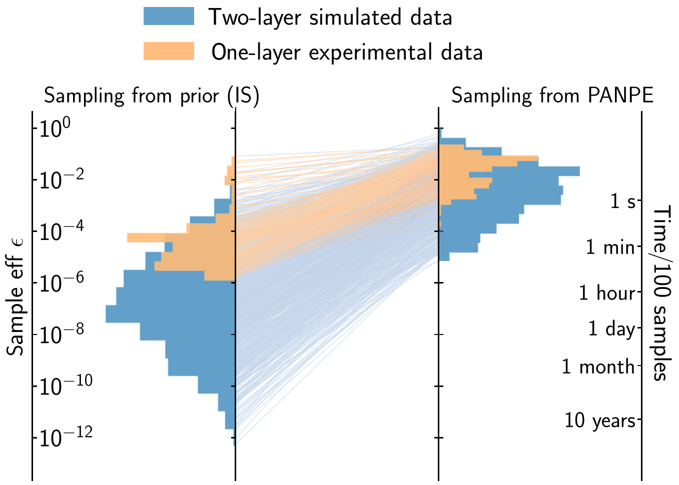

Fig. 3. Efficiency comparison between PANPE-IS and conventional IS. Sample efficiencies for conventional IS (**left**) and our PANPE-IS model (**right**) on a test dataset of 1000 simulated curves (blue) and a experimental dataset of 208 x-ray reflectometry curves (orange). An additional axis on the right-hand side indicates the estimated time it takes to generate 100 effective samples (ESS) on our hardware with the efficient GPU-accelerated reflectometry simulator (see the main text). Both the simulated and experimental data consist of two-layered structures with 10 parameters, but in the experimental data, only the top layer is unknown, as the parameters of the silicon substrate and the silicon oxide layer are largely constrained through their respective priors.

Performance on experimental XRR data

In this section, we evaluate the performance of PANPE-IS on the largest publicly available reflectometry dataset (*31*) that has been previously used for evaluating the performance of machine learning–based regression models (*32*–*34*).

Experimental data

As experimental XRR data, we use 208 curves, accompanied by a manual analysis using a conventional fitting procedure via maximum likelihood estimation. These data originate from three online in situ experiments conducted at different synchrotron facilities. Each experiment recorded in real time a process of growing an organic layer, specifically diindenoperylene (DIP), on a silicon substrate. DIP, an organic semiconductor, has gained interest due to its prospective uses in the field of electronics and photovoltaics (*35*). Real-time XRR measurements can provide insights into growth processes of such thin films. Furthermore, this type of experiment can benefit from rapid analysis. In this way, a machine learning–based solution was recently deployed for the first closed-loop XRR experiment (*36*). However, the ambiguity problem presented limitations to the regression-based model, rendering our probabilistic method a potential successor in such closed-loop systems (see fig. S7).

For each experimental curve, we set uniform priors based on a physical understanding of the experiment, aligning with conventional analysis. The parameters of the known silicon substrate and the oxide layer are essentially fixed by designating narrow ranges. In contrast, the parameters for the thickness d_1, roughness σ_1, and density ρ_1 of the growing organic layer have broader prior ranges due to uncertainty. Furthermore, as the film thickness d_1 increases, its ranges are set to increase linearly with time, in line with the anticipated growth rate. Although the physical model contains 10 parameters, in this case, the physics-informed prior information about the structure allows us to effectively constrain most of the parameters. Nonetheless, prior amortization allows us to use the same PANPE model that was applied to simulated data featuring two-layer structures.

We also note that the datasets under consideration feature different q ranges and resolutions. Nevertheless, a single model can process them due to the prior amortization that exploits the scaling invariance of the reflectometry data, as well as due to the amortized discretization of our model.

Comparison with conventional data analysis

The defined prior distributions are narrow enough to enable conventional IS for validation purposes. Figure 4 displays the time-dependent marginal distributions for three parameters obtained by both our PANPE-IS model (on the left) and the conventional IS method (on the right), showing equivalent solutions. Figure S3 demonstrates a corner plot with posterior estimates obtained for an

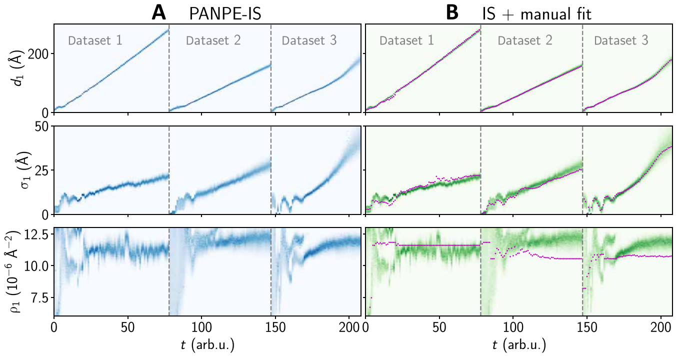

Fig. 4. Consistent results between PANPE and conventional sampling on experimental data, contrasting with inconsistencies in previously reported manual fits. Marginal distributions of the thickness d_1, roughness σ_1, and density ρ_1 of the DIP layer growing on a silicon substrate are shown for three in situ experimental XRR datasets. The distributions obtained by our model (**A**) are compared with those obtained via conventional IS from prior distribution (**B**). The colors designate normalized probability densities. Purple dots correspond to manual fits performed using differential evolution, as reported in (*31*). Figure S5 shows time-dependent sample efficiencies and log evidence estimations for both methods. arb. u., arbitrary units.

experimental curve using PANPE, PANPE-IS, and PANPE-MCMC, where two refinement methods lead to equivalent distributions.

Despite the relative simplicity of the dataset due to a small number of free parameters, the resulting distributions feature two solution branches for the density parameter ρ_1 visible in Fig. 4 (see also fig. S4). The upper branches vanish beyond a certain time for each dataset, suggesting that the correct solution corresponds to the lower branches. Conventional fits, indicated by purple dots in Fig. 4B, deviate from the maximum likelihood, underscoring the relevance of probabilistic methods even in such straightforward cases.

Figure 3 (orange color) demonstrates sample efficiencies on the experimental data for the conventional IS and our PANPE solution. Notably, in the case where most parameters are effectively known and constrained, conventional IS can be a practical solution, unless a real-time analysis is required. For most of the considered experimental data, our model performs the analysis in under a second, where IS may require tens of minutes per sample. Several curves where IS is almost as efficient as PANPE-IS correspond to the beginning of the growth process, when there is essentially still no organic layer, and the other parameters are known. Therefore, the axis for IS sample efficiency in Fig. 3 can be approximately divided into ranges: $\epsilon_{\text{eff}}^{\text{IS}} > 10^{-3}$ for "zero-layer" structures, $\epsilon_{\text{eff}}^{\text{IS}} \in \left[10^{-6}, 10^{-3}\right]$ for "one-layer" structures that constitute the rest of the experimental curves, and $\epsilon_{\text{eff}}^{\text{IS}} < 10^{-6}$ for more complex, simulated two-layer structures. The respective estimated inference time axis in Fig. 3 suggests that pure conventional likelihood-based methods become largely impractical for two-layer structures. On the other hand, PANPE-IS delivers accurate and reliable solutions in under a minute for most of these cases.

Beyond two-layered structures

Figure S6 displays the inference results for a simulated four-layer structure with 16 parameters, revealing a highly ambiguous outcome (the result is obtained via PANPE-IS using an additional model trained on four-layer structures). This example underscores the increasing complexity and ambiguity in reflectometry analysis as the number of free parameters grows. To maintain the high sample efficiency of PANPE in these more challenging scenarios, it is necessary to enhance the capacity of the density estimator. This enhancement could be achieved not only by enlarging the size of the neural network but also through the use of more sophisticated density estimators, such as continuous normalizing flows (*37–39*), without necessitating substantial changes to the overall framework presented in this paper. Improving performance for more complex scenarios may also involve customizing the hyperprior distribution and other solution aspects, such as q discretization, to better suit the specific application and experimental conditions.

Role of physics-informed priors

Figure 5 illustrates the essential role of prior amortization in the analysis of the presented experimental data. Without providing the physics-informed prior distribution to the neural network, the resulting samples span all possible solutions (represented by the red SLD profiles in Fig. 5A) within the expansive prior distribution that covers the complete training range. Yet, all of these solutions are unphysical, due to factors such as a too thick oxide layer. The physical solution, depicted by the blue SLD profiles, is practically unattainable without prior amortization given that the share of samples within the corresponding narrow prior is less than 10^{-6}. This scenario emphasizes once more the essential role of incorporating additional physical information into inverse scattering problems with missing phase.

Adaptive q discretization

The ability of our model to support arbitrary q discretization substantially broadens its applicability. In this way, the number of q points in the analyzed x-ray data ranges from 25 to 52, yet it is processed by the same model. We note that an alternative approach involving interpolation to conform to a fixed q axis can generally lead to missed solutions. For instance, if a model trained on 52 q points is subsequently tested on experimental data comprising only 25 q points, then interpolating these data to 52 points could create a falsely narrow distribution. This occurs because the interpolation process artificially adds "information" that the original experimental data does not have, compromising the guarantee of the coverage property.

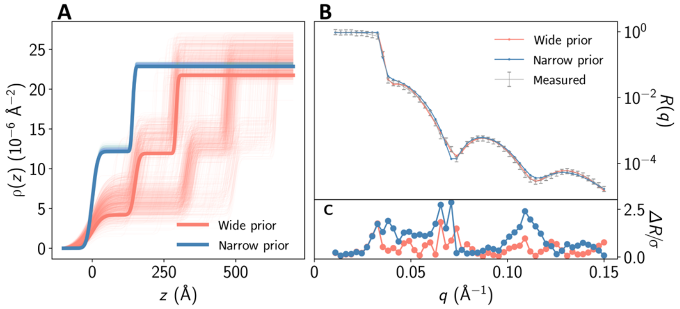

Fig. 5. Experimental XRR curve analyzed using both a wide prior distribution that encompasses the entire training range (red) and a narrow, physics-informed prior distribution (blue). (**A**) SLD profiles associated with the PANPE-IS samples. Profiles with the highest likelihood are highlighted with bold lines. (**B**) The observed reflectivity curve (in gray) is compared with simulated curves that correspond to the maximum likelihood from both the narrow and wide prior distributions. While both fits are satisfactory, the unphysical solution (in red) has a likelihood that is more than 10^6 times greater than its physical counterpart due to larger residuals (**C**). In this case, when trained solely with a wide prior distribution, the NPE network mainly samples unphysical solutions, which appear much more probable without the physics-informed prior. The use of prior amortization addresses this issue.

Figure S7 demonstrates another relevant experimental scenario that requires adaptive q discretization: While performing an XRR measurement by sequentially measuring intensities at distinct points, one can use the model to analyze the data at any particular moment. This analysis can inform whether additional data are needed to reduce uncertainty and can even guide the selection of the next most informative q point to measure. Obtaining such points is straightforward using reflectometry curves simulated using parameters sampled via PANPE-IS.

Combination of multiple NR measurements

In this section, we demonstrate that PANPE can be successfully used for the simultaneous analysis of several combined NR datasets, which is an indispensable tool in the context of contrast variation [e.g., using different levels of deuteration (*40*, *41*)].

Corefinement of neutron data

A common method to resolve ambiguity involves combining measurements taken under controlled variation of experimental conditions. In the context of reflectometry, such conditions can be determined by the different energies of the x-ray beam (i.e., anomalous scattering near an absorption edge), polarizations of neutrons for magnetic materials, or by using different contrasts via changing materials adjacent to the sample.

We demonstrate this corefinement procedure using publicly available neutron data (*42*). Specifically, we consider two NR measurements of a polymer on a silicon substrate performed separately with H_2O ($\rho = -0.56 \cdot 10^{-6}\ \text{Å}^{-2}$) and D_2O ($\rho = 6.36 \cdot 10^{-6}\ \text{Å}^{-2}$) solvents. Neutron reflectometry differs from x-ray data in several aspects, such as instrumental resolution, high scattering background, and negative SLD. Therefore, we trained an additional PANPE model for neutron data that incorporates these features and has 11 free parameters that now also include scattering background.

Constructing proposal distribution from several measurements

The likelihood for two measurements $\mathbf{R}_{H_2O}$ and $\mathbf{R}_{D_2O}$ is a product $p(\mathbf{R}_{H_2O} | \theta) p(\mathbf{R}_{D_2O} | \theta)$. The corresponding posterior distribution cannot be directly estimated by (PA)NPE model unless specifically trained on such combined measurements. However, using our model trained on single curves, one can combine two (or more) sets of samples generated independently for each measurement for constructing a proposal distribution $\frac{1}{2}\left(p_{\text{NN}}(\theta | \mathbf{R}_{H_2O}) + p_{\text{NN}}(\theta | \mathbf{R}_{D_2O})\right)$. Such a proposal distribution exhibits the probability mass coverage property and can be further refined via likelihood-based methods for obtaining a reliable and accurate posterior distribution.

However, additional complications arise when only a subset of the estimated parameters—namely, the unchanged parameters of the studied sample, θ^{shared}—are shared among multiple measurements. Other parameters, θ^{unique}, such as background, misalignment, and different contrast densities, are unique to each measurement. Thus, the estimated parameters are expressed as $\theta = \left[\theta^{\text{shared}}, \theta^{\text{unique}}_{H_2O}, \theta^{\text{unique}}_{D_2O}\right]$. As a result, a subset of parameters generated by the model for the first measurement, $\left[\theta^{\text{shared}}, \theta^{\text{unique}}_{H_2O}\right] \sim p(\theta | \mathbf{R}_{H_2O})$, is incomplete as it lacks the subset of parameters unique to the second (other) measurement(s) $\theta^{\text{unique}}_{D_2O}$ and vise versa. The solution involves sampling the remaining parameters from the parameter-conditioned posterior distribution: $\theta^{\text{unique}}_{D_2O} \sim p(\theta | \mathbf{R}_{D_2O}, \theta^{\text{shared}})$. This conditional distribution is not provided by NPE and typically necessitates training additional models.

The reparameterization operation that we introduce as a part of our prior-amortized approach offers a means to approximate such conditional probability densities with the same model by setting very narrow priors, essentially fixing the required parameters. However, this approach provides only samples and not density evaluation required for IS refinement (see more details in Materials and Methods). Therefore, we use PANPE-MCMC for likelihood refinement of the combined posterior in this case.

Inference results for combined measurements

Figure 6 demonstrates the results of the PANPE-MCMC analysis of a single neutron reflectivity curve measured with H_2O contrast (Fig. 6, A and B), as well as the joint analysis of two measurements incorporating both H_2O and D_2O contrasts (Fig. 6, C and D). The single measurement yields two distinct solutions (Fig. 6A), one of which is (implicitly) ruled out when performing a corefinement of two measurements (Fig. 6C), thereby resolving ambiguity in data interpretation. Nonzero ambient density is processed as discussed in Materials and Methods.

In the case of the neutron data analyzed in this work, the parameters unique for each measurement θ^{unique} include scaling misalignment, background, and densities of contrasts. The shared parameters correspond to the constant parameters (i.e., those that do not change in-between these measurements) of the system under study, such as thicknesses, roughnesses, and densities of Si and SiO_2. Notably, following the parameterization described in the refnx package, we account for the volume fraction $v_{\text{solv}} \in (0, 1)$ of the solvent that modifies the SLD of the polymer layer according to $\rho = (1 - v_{\text{solv}})\rho_{\text{polymer}} + v_{\text{solv}}\rho_{solv}$. We regard these densities ρ as parameters unique to each measurement, using them to compute v_{solv} and ρ_{polymer}. The mixture of solvent with polymer results in a small difference of polymer SLDs for measurements with different contrasts in Fig. 6.

It is worth noting that in certain cases of application-specific parameterization on an SLD profile, it might be more practical to retrain the model using a more suitable parameterization, but it is not necessary in this case. In general, the demonstrated approach can be equally well applied to other cases of parameter corefinement such as polarized NR, XRR measurements with different energies, or other similar applications that require combining several measurements.

DISCUSSION

In reconstructing physical systems from scattered intensities, the phase problem poses a fundamental challenge. This problem is encountered in numerous scattering techniques, including XRR and NR. The prevailing standard in reflectometry analysis is maximum likelihood estimation, which uses a differential evolution-based search over the parameter space. This method, by design, produces a single system reconstruction, even when multiple solutions exist due to phase loss, making it inherently unreliable. In contrast, Bayesian inference provides a foundational pathway to a more reliable analysis by inherently accounting for all possible solutions as a distribution over the considered physical parameters. However, despite its conceptual advantages, conventional likelihood-based Bayesian methods struggle with high-dimensional parameter spaces and multimodality, often falling short in real-world experimental

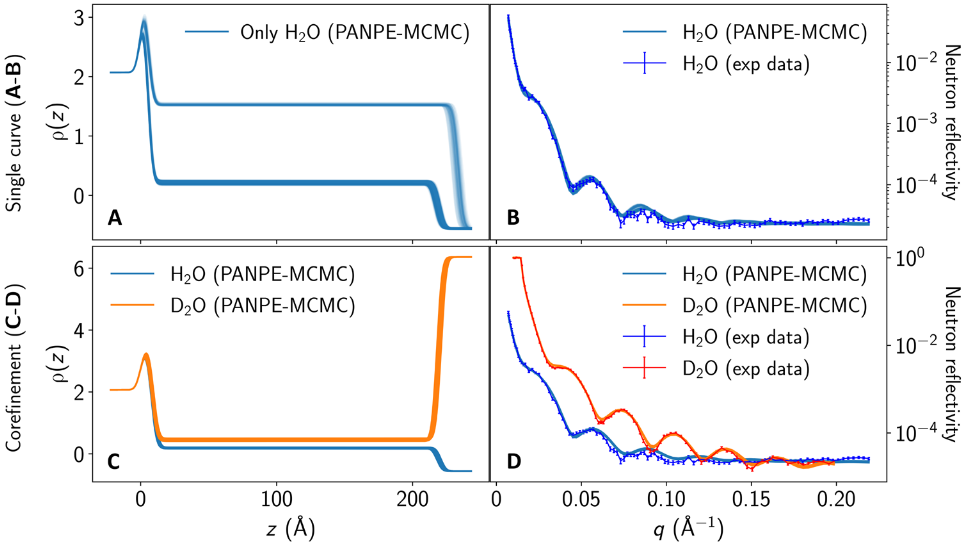

Fig. 6. Corefinement of two neutron reflectometry measurements of a polymer on a silicon substrate using H_2O and D_2O contrasts, conducted via PANPE-MCMC. (**A** and **B**) An analysis of a single measurement with H_2O contrast (B) yields two distinct solutions and their corresponding SLD profiles (A). (**C** and **D**) Corefinement of the measurements using H_2O contrast (blue) and D_2O contrast (orange) resolves ambiguity by eliminating one of the solutions. Data and priors are sourced from the refnx package (*42*). The combination of several measurements is enabled in this case by the use of parameter-conditioned posterior estimation.

contexts. This underscores the pressing need for more efficient and reliable Bayesian methods in reflectometry analysis.

In this work, we present an approach that enables reliable, accurate, and fast Bayesian reflectometry analysis. The high inference speed essential for various experimental contexts is achieved via neural network–based amortization. The training strategy reliably provides an approximate posterior distribution that fully covers the true posterior, while likelihood-based inference is subsequently used to obtain the accurate distribution.

Amortized inference proves to be a highly practical solution by allowing a model to be trained in advance, before any data are measured. A set of such models, once trained, can cover a broad range of applications, substantially accelerating analysis. This is particularly beneficial for experiments at large facilities. Given the high operational costs, experiments at neutron and synchrotron facilities require strict inference-time optimization to maintain cost-effectiveness. Furthermore, fast analysis enables more informed experiment design and opens possibilities for new data-driven experiments crucial for material discovery.

As the complexity of the task increases, such as with a higher number of layers in a multilayer structure and increased uncertainty, traditional methods become less feasible, with estimated inference times reaching months or longer for even moderately complex two-layer structures. In contrast, our method achieves accurate inference in under a minute for the same samples. This advancement allows fast and reliable analysis of such complex structures.

Our method enables reliable reflectometry experiments: It can guide experimentalists by indicating whether the remaining ambiguity requires additional measurements, thereby directing the experiment until the singular physical solution is determined. Furthermore, preparing experiments in advance by investigating ambiguities in simulated settings becomes possible.

Prior amortization crucially broadens the applicability of NPE to multiple experimental settings. In the context of reflectometry, constraining the parameter space based on individual characteristics of a studied structure is essential to resolve ambiguity. As shown, prior amortization allows to infer a physical solution that can feature a million times smaller likelihood and filter out an unphysical solution that is yet legitimate in other experimental settings or for other systems. This, in particular, underscores the importance of prior amortization when applying SBI methods to scattering problems with phase loss. Complex parameterizations of prior distributions can be used and should be investigated in future research aiming at adapting PANPE to specific applications.

Real-world benchmarks are valuable for the developing field of SBI. In this context, reflectometry analysis offers a notable benchmark, characterized by challenging multimodal distributions and bolstered by an efficient simulator. It can be straightforwardly scaled up by increasing the number of layers in the physical model. SBI is often seen as advantageous for applications where likelihood is costly or intractable to evaluate. Reflectometry does not fit into this category, and it exemplifies the broader utility of SBI methods beyond intractable likelihoods due to the coverage property and acceleration of inference through amortization. Moreover, the combination of SBI with likelihood-based methods presents as the optimal way to both preserve the coverage and achieve high accuracy.

Prior amortization can be beneficial for multiple applications, especially in experimental science like scattering where various experimental scenarios require adaptive prior distributions. In this manner, our method is suitable for a wide range of scientific experiments that permit SBI.

MATERIALS AND METHODS

Parameterization of the SLD profile

We consider the standard parameterization of the SLD profile of a layered structure with n_l layers through parameters $\theta = \{\mathbf{d}, \rho, \sigma, \Delta R, \Delta q, \log_{10}(R_0)\}$, and we primarily consider two-layer structures $n_l = 2$. Here, $\mathbf{d} = \{d_1, d_2\}$ are layer thicknesses in the top-bottom order, $\rho = \{\rho_1, \rho_2, \rho_{\text{sub}}\}$ are densities of two layers and the substrate, and $\sigma = \{\sigma_1, \sigma_2, \sigma_{\text{sub}}\}$ are roughnesses of three interfaces modeled via Névot-Croce factors. We exclude absorption in this work due to our focus on organic materials, but it can be straightforwardly included into our framework. In addition, we consider standard misalignment parameters: normalization misalignment ΔR and systematic misalignent of the q axis Δq. We only consider the parameter for scattering background $\log_{10}(R_0)$ for neutron data, which results in 11 parameters. Consequently, the model for x-ray data has 10 free parameters.

Equivariant transformations in reflectometry

Reflectometry features several equivariant transformations that can be considered to improve the performance of an amortized machine learning solution. In this section, we discuss these transformations and how we incorporate them into our PANPE model.

Unit-based scaling equivariance

Reflectometry simulation $R(q, \theta) = R(q, \mathbf{d}, \sigma, \rho)$ features an invariant scaling transformation that represents the change in used parameter units u

$$T_u(R(q, \mathbf{d}, \sigma, \rho)) = R(q, \mathbf{d}, \sigma, \rho) \tag{4}$$

where

$$T_u(R(q, \mathbf{d}, \sigma, \rho)) \equiv R(q \cdot u, \mathbf{d}/u, \sigma/u, \rho \cdot u^2) \tag{5}$$

and $u \in \mathbb{R}_{>0}$ is a positive value that defines the used units. The standard units are inverse angstroms (Å^{-1}) for q values, angstroms (Å) for layer thicknesses $\mathbf{d}$ and roughnesses σ, and inverse squared angstroms (Å^{-2}) for (scattering length) densities ρ. If we set $u = 1$ for these standard units, for instance, then the transformation with u = 10 would correspond to the change of units from angstroms to nanometers, which does not alter the resulting reflectivity curve. Similarly, $u = 2$ doubles the q range, halves thicknesses and roughnesses, and increases densities by a factor of $u^2 = 4$, leaving the reflectivity curve unchanged.

This invariance of the reflectometry simulator leads to the equivariance of the density estimator under the joint unit transformation of input and parameters. Specifically, stretching or squeezing the q axis in the input data and adjusting the input prior parameters ϕ accordingly should result in respective transformations of the parameters θ as per Eq. 4. We note that the transformation of prior distribution is commonly required to preserve equivariance in the density estimator. Applying this transformation requires prior amortization.

To incorporate this equivariance into our model, we can standardize the "pose" of the data (*28*) (the terminology is adopted from computer vision) to simplify the problem for the neural network. We do so by fixing the q range, on which our model is trained. During inference, we first preprocess the data by applying the transformation from Eq. 4 so that the measured q range matches the standard one. The corresponding scaling factor is the ratio of two ranges $u = q_{\max} / q_{\exp}$. We use this scaling factor to apply the respective transformation on the prior parameters ϕ. After obtaining samples from the PANPE model, we rescale the parameters back using u^{-1}. Respectively, the probability densities are corrected by a constant Jacobian determinant of the transformation, which equals u^{-2} in our case.

We note that this property should also be taken into account when considering parameter ranges for training. For instance, some unreasonably large parameter ranges that might seem unphysical, such as density values that do not correspond to any known materials, can be practically justified since they correspond to smaller densities when scaling the q axis. This relation is illustrated in fig. S9.

Density shifting equivariance

Reflectometry is sensitive to density contrasts rather than absolute density values. As a result, shifting all the densities ρ (SLDs) in the system, including the ambient and the substrate, is an invariant operation that does not change the resulting reflectivity curve. In the context of the density estimator, this leads to the equivariant operation: Shifting the respective prior parameters ϕ should result in the shift of the parameters θ (specifically, layer densities ρ).

We use this property by defining a natural standard pose in a form of the zero ambient density, on which the model is trained. During inference, the data with nonzero ambient density are first preprocessed by shifting the densities (i.e., the respective prior parameters) so that the ambient density becomes zero. We apply this transformation for NR. The Jacobian determinant of this transformation is equal to 1.

Misalignment shifting equivariances

The misalignment parameter ΔR results from the incorrect normalization when calculating reflected intensities as per $R(q, \theta) \cdot (1 + \Delta R)$, which effectively shifts the reflectometry curve in the "vertical" direction in the log space, resulting in an equivariant shifting transformation. A natural "standard pose" in this case corresponds to $\Delta\text{R} = 0$. We note that, in this case, the standard pose depends on the (unknown) parameter ΔR rather than the data and cannot be performed as a one-step preprocessing. This scenario is similar to the one considered previously (*28*), where an iterative inference scheme is proposed that allows converging to the standard pose. In our case, the range of the misalignment parameter is already very limited, making a direct application of the iterative scheme impractical. Our preliminary tests suggest that this does not lead to improved performance, so we do not use this equivariance in our solution. The same applies to the other misalignment parameter Δq.

PANPE training

We amortize Bayesian inference for a class of prior distributions $p(\theta \mid \phi)$, parameterized by ϕ. To specify the range of priors for which the model is trained, we introduce the hyperprior distribution—a distribution over prior parameters $p(\phi)$, which generally depends on the range of anticipated applications and can reflect various physical and practical parameter constraints. We set it to cover a broad range of practical scenarios where some of the parameters are known better (with lower uncertainty) than others.

The training process of the PANPE model involves adjusting the trainable parameters, denoted as $\mathbf{w}$, of the neural network to minimize a forward KL divergence between the true posterior distribution $p(\theta \mid \phi)$ and the flow-based density estimator $p_{\mathbf{w}}(\theta \mid \phi)$

$$\begin{aligned} L(w) &= \mathbb{E}_{p(\phi)} \mathbb{E}_{p(\theta|\phi)p(R|\theta)} \left[\log\left(\frac{p(\theta \mid R, \phi)}{p_w(\theta \mid R, \phi)} \right) \right] \\ &= \mathbb{E}_{p(\phi)} \left[D_{\text{KL}}(p(\theta \mid R, \phi) \| p_w(\theta \mid R, \phi)) \right] \end{aligned} \tag{6}$$

The loss in Eq. 6 is approximated by the Monte Carlo estimation

$$L(w) \approx \sum_i^N \left[-\log(p_w(\theta_i \mid \mathbf{R}_i, \phi_i)) \right] + \text{const} \tag{7}$$

where the constant does not depend on the model parameters $\mathbf{w}$. Evaluating Eq. 7 requires evaluating log density $\log\left(p_w\left(\theta_i \mid \mathbf{R}_i, \phi_i\right)\right)$ and drawing samples $\{\phi_i, \theta_i, \mathbf{R}_i\}_i^N$ from the training distribu tion $p(\phi, \theta, \mathbf{R}) \propto p(\mathbf{R} \mid \theta) p(\theta \mid \phi) p(\phi)$.

Exact density evaluation and consequently the utilization of the forward KL divergence are facilitated by normalizing flows, distinguishing them from many variational architectures. The forward KL divergence is a mass-covering loss, meaning that the optimized density density covers the whole support of the target distribution (otherwise the loss diverges), thereby ensuring no distributional modes are missed in the posterior estimation. Hence, although the training scheme of PANPE is independent of the specific architecture of the density estimator, the selection of normalizing flows as the density estimator and the corresponding loss function is crucial for the method's reliability. It is also noteworthy that some other recent density estimators exhibit the mass-covering property and can thus be integrated into the PANPE framework for future reflectometry applications.

Sampling from the training distribution can be performed in two principled ways. The first one is used in this paper and it goes as follows:

1) First, the prior parameters are sampled from the hyperprior distribution $[\phi_i \sim p(\phi)]$ defining the corresponding prior $p\left(\theta \mid \phi_i\right)$.

2) Then, the parameter θ_i is sampled from this specific prior distribution, $\theta_i \sim p\left(\theta \mid \phi_i\right)$.

3) Last, the corresponding reflectometry curve, $\mathbf{R}_i \sim p\left(\mathbf{R} \mid \theta_i\right)$, is sampled from the likelihood.

We note that a similar training scheme, involving sampling from a hyperparameter distribution, has been previously used in group-equivariant neural posterior estimation (*28*).

The second possible way of sampling from the training distribution relies on the relation $p(\phi)p(\theta \mid \phi) = p(\phi \mid \theta)p(\theta)$. Therefore, instead of first sampling prior parameters ϕ and then parameters θ, this order can be reversed to sample (potentially, multiple sets of) prior parameters ϕ that correspond to the same parameters θ, hence the same simulations. The gain in this case comes from the opportunity to reuse the same simulations by coupling them with different priors and potentially reduce the required number of simulations, which is critical for certain applications. Since this is not applicable to reflectometry, we do not investigate this method any further. We only note that its implementation would involve additional Bayesian inference $p(\phi \mid \theta) \propto p(\phi)p(\theta \mid \phi)$, the complexity of which depends on the chosen prior parameterization $p(\theta \mid \phi)$. For instance, in the case of the parameterization used in our work, the inference can even be performed analytically (via inverse transform sampling).

As an optional improvement of our method, we introduce a reparameterization transformation $\tilde{\theta} = T_\phi(\theta)$ of the parameters θ to effectively "rescale" the parameters according to the respective prior. The reparameterization is chosen to ensure that the prior for the rescaled parameters $p\left(\tilde{\theta}\right)$ does not depend on the prior parameters ϕ. The flow-based model then is trained to perform inference on these rescaled parameters

$$
\begin{aligned}
p_{\mathrm{NN}}\left(\tilde{\theta} \mid \mathbf{R}, \phi\right) &\approx p\left(\tilde{\theta} \mid \mathbf{R}, \phi\right) \propto p\left(R \mid \tilde{\theta}, \phi\right) p\left(\tilde{\theta}\right), \\
p_{\mathrm{NN}}(\theta \mid \mathbf{R}, \phi) &= p_{\mathrm{NN}}\left(\tilde{\theta} \mid \mathbf{R}, \phi\right) \ \left|\det J_{T_\phi}\right|
\end{aligned} \tag{8}
$$

This reparameterization effectively reframes the problem as a standard neural posterior estimation, only now the likelihood depends on both the parameters $\tilde{\theta}$ and the prior parameters ϕ. By doing so, we can now apply narrow priors without running into numerical issues. This approach accelerates the training process and decreases the number of samples generated outside the prior support.

When setting the prior width of some parameter θ_j to zero, we effectively fix it. The respective parameter $\tilde{\theta}_j$ estimated by the model does not influence the likelihood and is essentially trained to match the reparameterized prior distribution (uniform in our case). Consequently, the connection between the reparameterized space $\tilde{\theta}$ and the parameter space θ becomes surjective. To sample from the lower-dimensional distribution conditioned on the fixed parameter θ_j, we need to marginalize over the parameter $\tilde{\theta}_j$ in the reparameterized space. It is not directly possible to evaluate density of a marginalized distribution in normalizing flows. We note that in the ideal scenario when the parameter $\tilde{\theta}_j$ is uniformly distributed, density evaluation becomes straightforward. Our tests suggest that marginal distributions $p\left(\tilde{\theta}_j\right)$ can deviate from a uniform distribution in practice. However, sampling from a distribution marginalized over $\tilde{\theta}_j$ is straightforward as it simply requires omitting the marginalized parameters. In this way, our reparameterization scheme enables us to sample from the parameter-conditioned posterior estimation.

Trained models and parameter ranges

Trained models

We present the main results for XRR and NR. Because of certain differences in the underlying physics and subsequent differences in the simulator, we have trained two PANPE models: one for XRR and another for NR. The results on the simulated data are presented for the XRR model. Most properties are shared between these models, except for the ranges of density parameters used (the SLD for neutrons can be negative), the instrumental resolution (more pronounced in NR), and the presence of strong constant background scattering. Although we focus on these two models in the paper, we also show some examples from other models, such as those with four-layer structures.

Parameter ranges

The training parameters are constrained by the predefined ranges. Here, we use the following ranges shared by all the layers. Densities range within $\left[0,\ 60 \cdot 10^{-6}\ \text{Å}^{-2}\right]$ for the XRR model and $\rho \in \left[-20 \cdot 10^{-6}\ \ \text{Å}^{-2},\ 60 \cdot 10^{-6}\ \ \text{Å}^{-2}\right]$ for the NR model. Thicknesses and roughnesses range within $\left[0,\ 500\ \text{Å}\right]$ and $\left[0,\ 50\ \text{Å}\right]$, respectively. In addition, we limit the maximum roughness of the interface by the half thickness of the thickest adjacent layer. For the misalignment parameters, the ranges are $\left[-2 \cdot 10^{-3}\ \ \text{Å}^{-1}, 2 \cdot 10^{-3}\ \ \text{Å}^{-1}\right]$ for Δq and $[-5\%, 5\%]$ for ΔI. In addition, for the NR model, we introduce the (log) background parameter $\log_{10}\left(R_0\right)$, $R_0 \in \left[10^{-9}, 10^{-4}\right]$, which is set to 10^{-10} for XRR.

We note that because of the applied equivariant transformations discussed above, during inference, the model can also operate outside these parameter ranges. This also means that one can set rather unphysical training ranges, such as large roughness or density, to cover certain realistic scenarios at different q ranges and ambient densities.

Hyperprior distribution

During training, the parameters $\phi = \{\theta_j^{\min}, \theta_j^{\max}\}_{j=1}^n$ are generated as follows. First, for each parameter θ_j, the width of the uniform prior

$\Delta\theta_j = \theta_j^{\max} - \theta_j^{\min}$ is sampled, which generally can range from 0 to the the total parameter ranges introduced above. Here, we use the weighted sum of the uniform and the truncated exponential distribution for sampling prior widths. The latter term is added to better represent narrow prior widths corresponding to higher certainty in the prior knowledge. Then, the "center" of a prior $c_j = \left(\theta_j^{\max} + \theta_j^{\min}\right) / 2$ is sampled uniformly within the allowed range. The parameters ϕ are then calculated from $\Delta\theta_j$ and c_j. Last, the upper bounds for interface roughnesses are rescaled to not exceed half the maximum thickness of adjacent layers (*34*). This overall sampling scheme effectively defines the hyperprior distribution $p(\phi)$.

The test simulated data were produced using the same sampling procedure as the training data. However, some of the curves (less than 5%) were manually excluded from the test dataset since they exhibited pathological properties such as nearly zero contrast between the layers. This scenario essentially reduces the number of physical layers in the studied structure and results in exact linear correlation between thicknesses of these layers. The proper way to reduce the number of layers in PANPE would be by setting the thicknesses of redundant layers to zero.

Data simulations

During training, we simulate reflectometry data using the training parameters θ_i. Each reflectometry simulation $\mathbf{R}_i \sim p\left(\mathbf{R} \mid \theta_i\right)$, $\mathbf{R} = \{q_p, R\left(q_p\right), s_p\}_{p=1}^{n_q}$ is performed in several steps discussed below.

Q discretization

First, q values are sampled uniformly from the range $q_p \sim U\left(0, q_{\max}\right)$ to enable arbitrary discretization. As discussed above, the q range corresponding to the standard pose is set equal to $q_{\max} = 0.15\ \mathring{\mathrm{A}}^{-1}$. However, we can vary this during inference by exploiting the unit-scaling equivariant transformation. The number of points n_q is also sampled uniformly $n_q \sim U(20, 64)$. In practice, it is implemented by masking out some of the input data from the model during training.

Amortized discretization is generally necessary because the posterior can be highly sensitive to it in reflectometry applications. Our tests show that by fixing the q discretization during training, we are able to considerably improve the model performance on the simulated data. However, different discretization of the experimental data necessitates an interpolation procedure, which can deteriorate the performance of the model and generally lifts the mass-probability coverage guarantees, especially for experimental data with the lower number of points.

Furthermore, amortized discretization is especially important in online XRR experiments, where time limitations constrain the number of measured q points. It in principle enables closed-loop AI-guided measurements that enable choosing the most informative q point to measure given the current data to speed up the overall process and be able to real-time phenomena with higher time resolution.

Nevertheless, fixing discretization might be beneficial for applications with more standardized experimental setup. Furthermore, we acknowledge possible ways to improve q simulations for neutron data to better reflect the physical nature of the process and possibly even tailor it for the use at certain neutron sources.

Measurement uncertainty

Next, we generate relative measurement uncertainties $s_p \sim U(5 \text{ and } 30\%)$, independently for each q point. We treat these uncertainties as error bars that correspond to SEs typically used in reflectometry analysis. As a noise model, we use the normal distribution as a common approximation of Poisson counting statistics for a high number of counts. We note that, generally, the use of the Poisson likelihood should be preferred in the case of low counts, which are especially frequent in neutron reflectometry. However, that requires reporting raw intensities, which are typically not included in the published data.

Reflectivity curves

Last, we simulate reflectivity curves, using the generated parameters $\theta = \{\mathbf{d}, \sigma, \rho, \Delta q, \Delta R, \log_{10}\left(R_0\right)\}$, q points, and measurement uncertainties s. We simulate curves in mini-batches using our parallelized GPU-accelerated PyTorch implementation of Abelès transfer-matrix method (*20*)

$$R_p = \left(R\left(q + \Delta q, \mathbf{d}, \sigma, \rho\right) \cdot (1 + \Delta R) + R_0\right) \cdot e_p \tag{9}$$

where $e_p \sim \mathcal{N}\left(1, s_p\right)$. For neutron reflectometry, we apply constant instrumental resolution $\frac{\delta q}{q} = 5\%$.

Training data

The models are trained on 300,000 mini-batches sampled according to the introduced training scheme. Each mini-batch contains 8192 reflectometry curves, resulting in $N \approx 2.5 \cdot 10^9$ training samples. Data generation is performed during the training for every batch without repetition. In this way, the model cannot overfit on a fixed training dataset, further increasing the reliability of the solution. The training process takes approximately 30 hours using a single NVIDIA V100 GPU.

Inference pipeline

During inference, the measured reflectometry data and the prior parameters are supplied to the PANPE model. Given the desired effective sample size ESS, the model provides the respective number of parameter samples $\{\theta_i\}_{i=1}^{N}$, refined by either providing importance weights r_i (PANPE-IS) or by running MCMC (PANPE-MCMC). In the following, we discuss the pre- and postprocessing stages of the inference, as well as the model architecture.

Input preprocessing

The input to the network is a measured reflectivity data $\mathbf{R} = \{q_p, R\left(q_p\right), s_p\}_{p=1}^{n_q}$ and the set of prior parameters $\phi = \{\theta_j^{\min}, \theta_j^{\max}\}_{j=1}^{n}$. The input data are therefore $3n_q + 2n$-dimensional, where n_q is arbitrary. We first preprocess it as follows.

First, we apply the equivariant transformations discussed above to standardize the data before inference. That includes calculating the scaling coefficient $u = q_{\max} / q_{\exp}$ to rescale the q axis of the measured data to match the training q range. The input prior parameters are transformed accordingly. After transforming the prior parameters and the q axis, both the reflectometry curve $R\left(q\right)$ and the measurement uncertainties $s\left(q\right)$ are preprocessed using a logarithmic transformation $0.1 \cdot \log_{10}\left(R_q + 10^{-10}\right) + 0.5$. Last, the prior parameters $\phi_{\mathbf{scaled}}$ are normalized with respect to the absolute parameter ranges.

Embedding network

We use an embedding network to convert the input data to a fixed-dimensional latent vector, which is then supplied to the normalizing flow model. We note that our embedding architecture should have an ability to handle input data of varying sizes. Our tests suggest that for a fixed discretization, convolutional neural networks (CNNs) provide the best performance on reflectometry data among different architectures. To this end, for arbitrary discretization, we implement a trainable neural kernel, which acts as an intermediary step, adapting the data before it reaches the CNN. The kernel is a neural network $K\left(q_k, q, R_q, s_q\right)$ that "interpolates" reflectivity levels and the measurement uncertainties to a set of predefined points $\{q_k\}_{k=1}^{n_k}$. We use three such kernels, featuring 16, 32, and 64 equidistant points, respectively. The spacing between these points defines the kernel's "window." The kernel outputs are averaged for q points falling within this window. Each kernel is a multilayer perceptron with Gaussian Error Linear Unit (GELU) activation functions and four-channel input, a hidden layer with 32 channels, and a two-channel output layer. Each of three kernels is coupled with a convolutional network discussed below. We note that this architecture is not supposed to be discretization invariant, as the posterior can be highly sensitive to the choice of q points in reflectometry.

Each CNN is a sequence of five blocks, each block containing a one-dimensional convolutional layer, followed by a batch normalization layer and a GELU activation function. Convolutional layers feature a kernel of size 3, stride = 2, and padding = 1. Consequently, the dimension of the processed data is (approximately) halved after each layer. The number of channels is doubled in each block, starting from 32 up to 512.

Outputs from three CNNs are concatenated with the preprocessed parameters ϕ and provided to a multilayer perceptron. The final 256-dimensional latent representation of the input is supplied to the flow-based model.

Flow-based model

A normalizing flow ([17]) uses a series of reversible and differentiable transformations on a simple, base distribution (in our case, the standard normal distribution). This approach generates a complex distribution from which samples can be efficiently drawn and evaluated. In this work, we use a series of 40 transformations, each transformation block being a composition of a coupling layer with monotonic rational-quadratic splines ([18]) and a batch normalization layer ([43]). After each transformation block, the parameters are randomly permuted.

Refinement by likelihood-based methods

During the inference, we sample parameters $\theta_i \sim p_{\text{NN}}(\theta\,|\,\mathbf{R})$ and generate parameters in batches with the corresponding log probabilities. The obtained curves are used for calculating importance sampling weights (PANPE-IS) and streaming estimation of sample efficiency ϵ_{eff}. We continue this procedure until the effective sample size ESS $= N \cdot \epsilon_{\text{eff}}$ reaches an adequate threshold which we set equal to 500. The same criterion is used for the traditional importance sampling, where the prior distribution $p(\theta\,|\,\phi)$ is used as the proposal distribution.

Alternatively, samples generated by PANPE are used as efficient initialization points for MCMC (PANPE-MCMC). In our work, we introduce GPU-accelerated PyTorch-based implementations of several affine-invariant MCMC algorithms ([44–46]) enabling near real-time MCMC-based refinement operation.

Low sample efficiency estimation

We consider two approaches for estimating sample efficiency for the conventional importance sampling method with prior acting as a proposal distribution. The first approach is the use of importance sampling weights via direct sampling from the prior distribution $p(\theta)$

$$\epsilon_{\text{eff}} = \frac{\langle w_i^{\text{IS}} \rangle^2}{\left\langle \left(w_i^{\text{IS}}\right)^2 \right\rangle} \tag{10}$$

where $w_i^{\text{IS}} = p\left(\mathbf{R}\,|\,\theta_i\right)$, $\theta_i \sim p(\theta)$, and $\langle\cdot\rangle$ is the average operation over all samples i.

An accurate estimation requires $N = \text{ESS}\,/\,\epsilon_{\text{eff}}$ samples, e.g., sample efficiency $\epsilon_{\text{eff}} = 10^{-12}$ requires more than 10^{12} samples, which is computationally unfeasible. An insufficient number of samples N only provide an upper bound $\epsilon_{\text{eff}} < 1\,/\,N$. Therefore, it can only be used in practice for sufficiently high sample efficiencies.

The alternative approach uses the analytical form ([47])

$$\epsilon_{\text{eff}} \overset{a.s.}{\to} \epsilon^*_{\text{eff}} = \left(\mathbb{E}_{\theta\sim p(\theta|\mathbf{R})}\left[\frac{p(\theta\,|\,\mathbf{R})}{p(\theta)}\right]\right)^{-1} = \frac{p(\mathbf{R})}{\mathbb{E}_{\theta\sim p(\theta|\mathbf{R})}\left[p(\mathbf{R}\,|\,\theta)\right]} \tag{11}$$

In the case of the uniform prior distribution $p(\theta)$, the equation simplifies to

$$\left(\mathbb{E}_{\theta\sim p(\theta|\mathbf{R})}\left[\frac{p(\theta\,|\,\mathbf{R})}{p(\theta)}\right]\right)^{-1} = \frac{v\left(p(\theta\,|\,\mathbf{R})\right)}{v\left(p(\theta)\right)} \tag{12}$$

where the quantity $v\left(p(\mathbf{x})\right) \equiv \left(\mathbb{E}_{p(\mathbf{x})}\left[p(\mathbf{x})\right]\right)^{-1}$ can be interpreted as an "efficient volume" of the distribution $p(\mathbf{x})$. In this way, $v\left(p(\theta)\right) = \prod_{j=1}^{n}\left(\theta_j^{\max} - \theta_j^{\min}\right) = \Theta$ is the volume of the prior distribution, and

$$v\left(p(\theta\,|\,\mathbf{R})\right) = \left(\int_{\Theta} p(\theta\,|\,\mathbf{R})^2 d\theta\right)^{-1} \tag{13}$$

characterizes the efficient volume of the posterior distribution. For instance, in the case of d-dimensional standard normal distribution $\mathcal{N}(0, 1\cdot\sigma)$, $v\left(p\right) = \left(2\sqrt{\pi}\sigma\right)^d$. Naturally, the sample efficiency in our case is the ratio between the defined volume of the target distribution and the volume of the (uniform) proposal distribution.

We estimate ϵ^*_{eff} using samples from our PANPE-IS model $\{\theta_i\}_{i=1}^{N}$

$$\epsilon^*_{\text{eff}} = \frac{p(\mathbf{R})}{\mathbb{E}_{\theta\sim p(\theta|\mathbf{R})}\left[p(\mathbf{R}\,|\,\theta)\right]} \approx \frac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} w_i p\left(\mathbf{R}\,|\,\theta_i\right)} \tag{14}$$

where the importance weights and samples provided by PANPE-IS should not be confused with the weights and samples from prior distribution in Eq. 10.

In this way, we obtain ϵ_{IS} estimations in the case of low sample efficiency of the IS method. However, when the prior distribution is narrow enough, we can estimate ϵ_{IS} using both methods independently. Figure S8 illustrates the consistency between these two approaches.

Supplementary Materials

This PDF file includes:

Figs. S1 to S9

REFERENCES AND NOTES

1. J. Als-Nielsen, D. McMorrow. *Elements of Modern X-ray Physics* (John Wiley & Sons Ltd, ed. 2, 2011).
2. R. Feidenhans'l, Surface structure determination by x-ray diffraction. *Surf. Sci. Rep.* **10**, 105–188 (1989).
3. J. R. Fienup, Phase retrieval algorithms: A comparison. *Appl. Opt.* **21**, 2758 (1982).
4. M. Bertero, P. Boccacci, C. De Mol, *Introduction to Inverse Problems in Imaging* (CRC Press, ed. 2, 2024).
5. C. Wang, U. Steiner, A. Sepe, Synchrotron big data science. *Small* **14**, 1802291 (2018).
6. D. Schumi-Mareček, F. Bertram, P. Mikulík, D. Varshney, J. Novák, S. Kowarik, Millisecond x-ray reflectometry and neural network analysis: Unveiling fast processes in spin coating. *J. Appl. Cryst.* **57**, 314–323 (2024).
7. L. G. Parratt, Surface studies of solids by total reflection of x-rays. *Phys. Rev.* **95**, 359–369 (1954).
8. S. K. Sinha, E. B. Sirota, S. Garoff, H. B. Stanley, X-ray and neutron scattering from rough surfaces. *Phys. Rev. B* **38**, 2297–2311 (1988).
9. A. Hinderhofer, A. Greco, V. Starostin, V. Munteanu, L. Pithan, A. Gerlach, F. Schreiber, Machine learning for scattering data: Strategies, perspectives and applications to surface scattering. *J. Appl. Cryst.* **56**, 3–11 (2023).
10. K. O. Brinkmann, T. Becker, F. Zimmermann, C. Kreusel, T. Gahlmann, M. Theisen, T. Haeger, S. Olthof, C. Tückmantel, M. Günster, T. Maschwitz, F. Göbelsmann, C. Koch, D. Hertel, P. Caprioglio, F. Peña-Camargo, L. Perdigón-Toro, A. Al-Ashouri, L. Merten, A. Hinderhofer, L. Gomell, S. Zhang, F. Schreiber, S. Albrecht, K. Meerholz, D. Neher, M. Stolterfoht, T. Riedl, Perovskite-organic tandem solar cells with indium oxide interconnect. *Nature* **604**, 280–286 (2022).
11. A. Armanious, Y. Gerelli, S. Micciulla, H. P. Pace, R. J. L. Welbourn, M. Sjöberg, B. Agnarsson, F. Höök, Probing the separation distance between biological nanoparticles and cell membrane mimics using neutron reflectometry with sub-nanometer accuracy. *J. Am. Chem. Soc.* **144**, 20726–20738 (2022).
12. L. Caselli, T. Nylander, M. Malmsten, Neutron reflectometry as a powerful tool to elucidate membrane interactions of drug delivery systems. *Adv. Colloid Interface Sci.* **325**, 103120 (2024).
13. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
14. T. Bayes, LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philos. Trans. R. Soc.* **53**, 370–418 (1763).
15. K. Cranmer, J. Brehmer, G. Louppe, The frontier of simulation-based inference. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30055–30062 (2020).
16. G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.* **22**, 2617–2680 (2021).
17. D. Rezende, S. Mohamed, Variational inference with normalizing flows. *Proc. Mach. Learn. Res.* **37**, 1530–1538 (2015).
18. C. Durkan, A. Bekasov, I. Murray, G. Papamakarios, "Neural spline flows" in *Advances in Neural Information Processing Systems*. H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, R. Garnett, Eds. (Curran Associates Inc., 2019), vol. 32; https://papers.nips.cc/paper_files/paper/2019/hash/7ac71d433f282034e088473244df8c02-Abstract.html.
19. M. Dax, S. R. Green, J. Gair, M. Pürrer, J. Wildberger, J. H. Macke, A. Buonanno, B. Schölkopf, Neural importance sampling for rapid and reliable gravitational-wave inference. *Phys. Rev. Lett.* **130**, 171403 (2023).
20. F. Abelès, La théorie générale des couches minces. *J. Phys. Radium* **11**, 307–309 (1950).
21. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, "PyTorch: An imperative style, high-performance deep learning library" in *Advances in Neural Information Processing Systems 32* (Curran Associates Inc., 2019), pp. 8024–8035; http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.
22. G. Papamakarios, I. Murray, "Fast ε-free Inference of simulation models with bayesian conditional density estimation" in *Proceedings of the 30th International Conference on Neural Information Processing Systems systems* (Curran Associates Inc., 2016), pp. 1036–1044.
23. J.-M. Lueckmann, P. J. Gonçalves, G. Bassetto, K. Öcal, M. Nonnenmacher, J. H. Macke, "Flexible statistical inference for mechanistic models of neural dynamics" in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (NIPS, 2017), pp. 1289–1299.
24. D. Greenberg, M. Nonnenmacher, J. Macke, Automatic posterior transformation for likelihood-free inference. *Proc. Mach. Learn. Res.* **97**, 2404–2414 (2019).
25. M. Deistler, P. J. Goncalves, J. H. Macke, "Truncated proposals for scalable and hassle-free simulation-based inference" in *Advances in Neural Information Processing Systems* (Curran Associates Inc., 2022), pp. 23135–23149.
26. M. Gloeckler, M. Deistler, C. Weilbach, F. Wood, J. H. Macke. All-in-one simulation-based inference. arXiv:2404.09636 [cs.LG] (2024).
27. M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, B. Schölkopf, Real-time gravitational wave science with neural posterior estimation. *Phys. Rev. Lett.* **127**, 241103 (2021).
28. M. Dax, S. R. Green, J. Gair, M. Deistler, B. Schölkopf, J. H. Macke, "Group equivariant neural posterior estimation" in *10th International Conference on Learning Representations* (ICLR, 2022); https://openreview.net/forum?id=u6s8dSporO8.
29. M. Dax, S. R. Green, J. Gair, N. Gupte, M. Pürrer, V. Raymond, J. Wildberger, J. H. Macke, A. Buonanno, B. Schölkopf. Real-time gravitational-wave inference for binary neutron stars using machine learning. arXiv:2407.09602 [gr-qc] (2024).
30. L. Elsemüller, H. Olischläger, M. Schmitt, P.-C. Bürkner, U. Koethe, S. T. Radev, "Sensitivity-aware amortized bayesian inference" in *Transactions on Machine Learning Research* (2024), https://openreview.net/forum?id=Kxtpa9rvM0.
31. L. Pithan, A. Greco, A. Hinderhofer, A. Gerlach, S. Kowarik, N. Rußegger, I. Dax, F. Schreiber, Reflectometry curves (XRR and NR) and corresponding fits for machine learning (Zenodo, 2022); https://zenodo.org/record/6497438.
32. A. Greco, V. Starostin, C. Karapanagiotis, A. Hinderhofer, A. Gerlach, L. Pithan, S. Liehr, F. Schreiber, S. Kowarik, Fast fitting of reflectivity data of growing thin films using neural networks. *J. Appl. Cryst.* **52**, 1342–1347 (2019).
33. A. Greco, V. Starostin, A. Hinderhofer, A. Gerlach, M. W. A. Skoda, S. Kowarik, F. Schreiber, Neural network analysis of neutron and x-ray reflectivity data: Pathological cases, performance and perspectives. *Mach. Learn. Sci. Technol.* **2**, 045003 (2021).
34. A. Greco, V. Starostin, E. Edel, V. Munteanu, N. Russegger, I. Dax, C. Shen, F. Bertram, A. Hinderhofer, A. Gerlach, F. Schreiber, Neural network analysis of neutron and x-ray reflectivity data: Automated analysis using mlreflect, experimental errors and feature engineering. *J. Appl. Cryst.* **55**, 362–369 (2022).
35. J. Wagner, M. Gruber, A. Hinderhofer, A. Wilke, B. Bröker, J. Frisch, P. Amsalem, A. Vollmer, A. Opitz, N. Koch, F. Schreiber, W. Brütting, High fill factor and open circuit voltage in organic photovoltaic cells with diindenoperylene as donor material. *Adv. Funct. Mater.* **20**, 4295–4303 (2010).
36. L. Pithan, V. Starostin, D. Mareček, L. Petersdorf, C. Völter, V. Munteanu, M. Jankowski, O. Konovalov, A. Gerlach, A. Hinderhofer, B. Murphy, S. Kowarik, F. Schreiber, Closing the loop: Autonomous experiments enabled by machine-learning-based online data analysis in synchrotron beamline environments. *J. Synchrotron Radiat.* **30**, 1064–1075 (2023).
37. R. T. Q. Chen, Y. Rubanova, J. Bettencourt, D. K. Duvenaud, "Neural ordinary differential equations" in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett, Eds. (Curran Associates Inc., 2018), vol. 31; https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.
38. Y. Lipman, R. T. Chen, H. Ben-Hamu, M. Nickel, M. Le. Flow matching for generative modeling. arXiv:2210.02747 [cs.LG] (2022).
39. J. Wildberger, M. Dax, S. Buchholz, S. Green, J. H. Macke, B. Schölkopf, "Flow matching for scalable simulation-based inference" in *Advances in Neural Information Processing Systems* (Curran Associates Inc., 2024), pp. 16837–16864.
40. F. Heinrich, T. Ng, D. J. Vanderah, P. Shekhar, M. Mihailescu, H. Nanda, M. Lösche, A new lipid anchor for sparsely tethered bilayer lipid membranes. *Langmuir* **25**, 4219–4229 (2009).
41. G. Cai, Y. Li, Y. Fu, H. Yang, L. Mei, Z. Nie, T. Li, H. Liu, Y. Ke, X.-L. Wang, J.-L. Brédas, M.-C. Tang, X. Chen, X. Zhan, X. Lu, Deuteration-enhanced neutron contrasts to probe amorphous domain sizes in organic photovoltaic bulk heterojunction films. *Nat. Commun.* **15**, 2784 (2024).
42. A. R. J. Nelson, S. W. Prescott, refnx: Neutron and x-ray reflectometry analysis in Python. *J. Appl. Cryst.* **52**, 193–200 (2019).
43. L. Dinh, J. Sohl-Dickstein, S. Bengio. Density estimation using real NVP. arXiv:1605.08803 [cs.LG] (2016).
44. C. J. T. Braak, A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: Easy Bayesian computing for real parameter spaces. *Stat. Comput.* **16**, 239–249 (2006).
45. J. Goodman, J. Weare, Ensemble samplers with affine invariance. *Comm. App. Math. Comp. Sci.* **5**, 65–80 (2010).
46. D. Foreman-Mackey, D. W. Hogg, D. Lang, J. Goodman, emcee: The MCMC Hammer. *Publ. Astron. Soc. Pac.* **125**, 306–312 (2013).
47. F. M. Polo, R. Vicente, Effective sample size, dimensionality, and generalization in covariate shift adaptation. *Neural Comput. Appl.* **35**, 18187–18199 (2022).

Acknowledgments: This research was supported in part by the Maxwell computational resources operated at DESY with the assistance of A. Rothkirch and F. Schlünzen. V.S. is grateful to S. Axen for insightful discussions. M.D. thanks J. H. Macke and B. Schölkopf for support. **Funding:** This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC number 2064/1—Project number 390727645. F.S. and Á.T.-C. are members of the Machine Learning Cluster of Excellence, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research

Foundation) under Germany's Excellence Strategy—EXC number 2064/1—Project number 390727645. This work was also funded by the BMBF and the DAPHNE4NFDI. **Author contributions:** V.S. conceived the concept for the work, implemented, trained, and tested neural networks, and implemented and conducted the conventional Bayesian analysis. M.D. proposed the training strategy for prior amortization. M.D. and Á.T.-C. contributed expertise in machine learning. A.H., A.G., and F.S. contributed expertise in scattering. F.S. supervised the research process. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the Supplementary Materials and published in Zenodo repository https://doi.org/10.5281/zenodo.14267737. The code is also publicly available at https://github.com/mlcolab/panpe.

Submitted 26 July 2024
Accepted 6 February 2025
Published 14 March 2025
10.1126/sciadv.adr9668