White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/245592/

Version: Published Version

Article:
Champneys, M.D., Parnell, A.J., Gutfreund, P. et al. (2026) Towards real-time surrogate-free Bayesian inversion for neutron reflectometry. Communications Materials. ISSN: 2662-4443

https://doi.org/10.1038/s43246-026-01341-4

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

communications materials

A Nature Portfolio journal

Article

Article in Press

https://doi.org/10.1038/s43246-026-01341-4

Towards real-time surrogate-free Bayesian inversion for neutron reflectometry

Check for updates

Max D. Champneys [1] ✉, Andrew J. Parnell [2], Philipp Gutfreund [3], Maximilian W. A. Skoda[4], J. Patrick A. Fairclough [5], Timothy J. Rogers[1] & Stephanie L. Burg [6]

Neutron reflectometry (NR) is a key enabling technology for many areas of scientific development. Although the forward reflectivity model is well-known, inferring the physical properties of a sample from NR data requires the solution of an inverse problem. Increasingly, beamline scientists are using NR in fast kinetic configurations and probing highly-complex structures and interfaces, introducing significant uncertainty. Existing uncertainty quantification (UQ) approaches in NR, such as Markov-Chain Monte-Carlo (MCMC), suffer from poor sample efficiency and slow convergence times. Recently, surrogate machine learning models have been proposed as an alternative. However, physical intuition is lost when replacing governing equations with fast surrogates. Instead, we propose a rapid, surrogate-free Bayesian inversion for NR. Our approach offers a step-change in inference speed and efficiency. For the first time in NR, exact gradients through the reflectivity are computed, enabling highly performant gradient-based inference schemes: Hamiltonian Monte-Carlo offers significant advances in sample efficiency compared to MCMC. Variational inference enables approximate UQ on the order of seconds rather than hours. We demonstrate state-of-the-art performance on a thick oxide quartz film, and robust co-fitting performance in the high complexity regime of organic LED multilayer devices. Additionally, we provide an open-source library of reflectometry kernels in the python language.

Neutron reflectometry (NR) is a widely used technique for the nanometre scale characterisation of soft matter[1], thin films[2] and bio-materials[3]. Increasingly, this has involved the study of more complex layered structures and ones that are evolving in time (dynamic). The application of NR has lead to many high-profile breakthroughs in the physical and biological sciences. Polarized neutron reflectivity (PNR) was crucial for understanding the oscillatory exchange coupling behaviour so fundamental to giant magnetoresistance (GMR)[4], a pivotal modern technology. Ongoing work has investigated how surfactant mixtures perform at varying temperatures using surface tension and neutron reflectivity[5], which has been crucial for developing more efficient, low-temperature detergents, vital for reducing the carbon footprint of washing laundry. Ongoing NR studies have also been important for the study of biological membranes[6], their interaction with drugs[7], such as antimicrobial peptides and other antibiotics.

NR is a scattering technique that illuminates a sample of interest using thermal neutrons produced and moderated from either a reactor or a spallation neutron source. NR measures the specular reflectivity R as a function of the momentum transfer Q from counting neutrons using a detector[8]. Theoretically, the forward reflectivity model of complex materials is well understood. Given the physical and apparatus parameters θ, one can appeal to Fresnel reflectivity to compute the predicted specular reflectivity $\hat{R}$. NR is however, inherently an indirect technique, which requires the reflectivity data to be modelled, as it is the absolute square of Fourier transform of the scattering length density, which gives rise to a loss in phase information. As an indirect measurement technique, access to the physical quantities of interest (layer thickness, scattering length density, roughness), necessitates the solution of an inverse modelling problem, that is inefficient for large amounts of data or complex multilayer structures (e.g. lithium batteries/electrodes). This inefficiency is particularly pronounced when

[1]Dynamics Research Group, University of Sheffield, Sheffield, UK. [2]School of Mathematical and Physical Sciences, University of Sheffield, Sheffield, UK. [3]Institut Laue-Langevin, Grenoble, France. [4]ISIS Pulsed Neutron and Muon Source, Rutherford Appleton Laboratory, Didcot, UK. [5]School of Mechanical, Aerospace and Civil Engineering, University of Sheffield, Sheffield, UK. [6]School of Chemical, Material and Biological Engineering, University of Sheffield, Sheffield, UK.
✉e-mail: max.champneys@sheffield.ac.uk

uncertainty quantification is desired. Sampling schemes such as Markov-chain Monte-Carlo (MCMC) can suffer from slow convergence and poor sample efficiency as the number of unknown parameters increases.

Fitting a parametric physical model of the sample is the inverse problem at the core of NR measurements and data analysis. In the four decades since the widespread adoption of NR as a measurement technique, many methods have been proposed to solve this inverse modelling problem. Until recently, advances in NR analysis and fitting have kept close pace with the progress in machine learning (ML). The rapid rise of deep learning, automatic differentiation and specialised computation libraries has revolutionised the scalability of ML in many other fields. Yet, these analysis routes have so far been very slow to translate to NR fitting. Among the leading software packages for NR and thin-film analysis[9,10], the underlying fitting routines are often based on traditional data optimisation techniques such as Levenberg-Marquardt (LM)[11,12], Broyden-Fletcher-Goldfarb-Shanno (BFGS)[13], or heuristic techniques such as differential evolution[14] or particle swarms[15]. Other approaches follow Bayesian methodologies such as Markov-chain Monte-Carlo (MCMC)[16,17] or nested sampling[18]. The authors are not currently aware of any widely-used NR software that fully utilises the power of modern automatic-differentiation and gradient-based optimisation techniques to directly solve the inverse problem.

In place of physically derived models, surrogate approaches have very recently been proposed to replace the inverse problem[19–23] with ML. These ML models supplant the forward reflectivity model and are trained to predict the parameters of interest directly from the measured R. Although such approaches are able to leverage advanced ML technologies and have shown success in some areas, it is inevitable that physical intuition is lost when replacing governing equations with fast machine learning.

In this paper, we seek to address the shortcomings of existing approaches to uncertainty quantification (UQ) in NR. The key innovation is that we access the gradients of the forward specular reflectivity of a thin multilayer film w.r.t the unknown parameters of interest, unlocking a host of powerful optimisation and inference techniques that remain thus-far unexploited in the context of neutron reflectometry. The principal advantages of such an approach are:

- Order of magnitude gains in sample-efficiency (using Hamiltonian Monte-Carlo) and inference speed (using a variational approach) compared to gradient-free MCMC approaches.
- The exact forward reflectivity laws are used to perform the inversion, no approximate pre-trained surrogate model is required.
- Full Bayesian posterior that can accommodate arbitrary prior knowledge.

The proposed approach is demonstrated in two NR case study examples: A benchmark Quartz film and a co-fit to four LED devices. Additionally, an open-source software library is provided for fast computation of forward reflectivity and gradients in the python language at https://github.com/MDCHAMP/refjax.

Regarding the existing literature, the authors are not aware of previous contributions directly applying automatic differentiation to the forward reflectivity model to perform NR inversion in the literature. Although no direct applications of automatic-differentiation for NR inversion are to be found in the scientific literature, the authors are aware of at least two library implementations that support this functionality in addition to our own contribution. The jax backend in refnx[9] which supports automatic differentiation, although this functionality is not at the time of writing exposed to the user through the user interface and differentiation with respect to refnx model parameters is not possible due to architectural constraints. Very recently, the reflectorch package[24] has been published with a pytorch backend that naturally supports automatic differentiation and tensor computation. This library has been used to generate training examples of NR data for neural networks in ref. 25. However, recent significant interest has been shown in simulation-based surrogate modelling in both the optimisation and inference settings.

In several works, neural-network surrogates have been proposed as an alternative to optimisation for the NR inverse problem in order to extract unknown parameters directly from data. Methods have been proposed in both NR[19,23] and X-ray diffraction[21], with the latter motivated as a way to provide real-time process control based on surrogate inversions. A recent review of ML technology applied to scattering problems can be found in ref. 20. In such cases, the principal idea is to pre-train a neural-network surrogate on indicative reflectometry curves (either simulated or experimental) such that it is able to directly extract the parameters of interest. The key advantage is that for a new unseen testing reflectometry curve, the parameters can be extracted in 'one-shot' without the need for potentially computationally expensive algorithms when solving the NR inverse problem. While such approaches have merit, it is apparent that they are weakened by the requirement to train the methods on large quantities of appropriately labelled training data (experimental or otherwise). While there are certainly situations in which this will be plausible, such methods require training data representative of the sample under investigation, which can be difficult to guarantee when the structure is unknown.

Recently, surrogate modelling has been proposed in an inference setting for NR. In[22], Starostin et al. propose to apply techniques from neural posterior estimation and prior amortisation to train a neural network to produce credible prior distributions based on reflectivity curves. The idea is that the network is trained to rapidly identify credible regions of the parameter space in the form of prior distributions that can be supplied to traditional Bayesian inference techniques such as Markov-chain Monte-Carlo (MCMC) or importance sampling (IS). The result is a significant reduction in the cost of evaluating a posterior over the NR parameters. The results of the paper are impressive and show that credible prior distributions can be found that include the multimodality that occurs due to the phase invariance problem when using very wide priors. As with all simulation based modelling approaches however, the requirement to pre-train a model (in this case on simulated data from a fixed number of layers) remains a limitation of the approach.

In summary, we show that exact gradients of the forward reflectivity model with respect to both the sample and instrument parameters can be obtained by automatic differentiation, and that access to these gradients materially changes what is achievable in uncertainty-aware NR inversion. On a benchmark crystalline quartz film, gradient-based optimisation recovers a quality of fit exceeding that previously reported for these data. For Bayesian inversion, Hamiltonian Monte-Carlo is shown to be highly sample efficient compared to a gradient-free scheme on the same data. On a co-fit example to four organic LED devices with 59 unknown parameters, variational inference returns an approximate posterior on the order of seconds per run while retaining direct contact with the physically meaningful slab parameters. We conclude that surrogate-free Bayesian inversion, at speeds compatible with in-situ decision-making on the beamline, is achievable without approximating the forward physics.

Methods

The NR inverse problem

At the very core of NR data analysis, is the inverse problem of obtaining estimates of unknown parameters of interest pertaining to both the slab model of the sample θ_{slab} (layer thicknesses, scattering length densities and roughnesses) and the instrument θ_{NR} (scaling/alignment parameters and the background detection level). For notational simplicity, in the description that follows these unknown parameters shall be referred to jointly as $\theta = \{\theta_{\text{NR}}, \theta_{\text{slab}}\}$. The Abeles matrix formalism[26] provides a physical model of the reflectivity $\hat{R}(Q, \theta)$ in terms of these parameters and the momentum transfer. Predictions from the forward model can be compared to the observed reflectivity in order to determine the values of the unknown parameters. A key challenge in NR, is the non-uniqueness of the inverse problem because the phase information is lost when measuring the reflected neutron intensity. For a particular measured neutron reflectivity measurement $R(Q)$, there may be many possible slab model solutions and therefore many sets of parameters that describe the data equally well. In practice, there

are a number of ways to mitigate the lack of phase information. One approach (as discussed in ref. 2) is to use multiple independent datasets to perform the inversion. Another approach uses reference layers with known density profiles.

In the absence of additional measurements, optimisation routines can be initialised in the vicinity of optima that are physically plausible given prior knowledge of the sample under investigation. In this work we will consider the problem of inversion in the vicinity of physically plausible solutions. Such local optimisation is an area in which gradient-based methods excel, particularly in high-dimensions.

Methods for solving the NR inverse problem can be broadly categorised into two groups that will be discussed in this section. The first group comprises *Optimisation methods*, these seek to find an optimal set of parameters θ^* that optimally satisfy an error metric e,

$$\theta^*|R = \underset{\theta}{\operatorname{argmin}}\ e\big(R(Q,\theta), \hat{R}(Q,\theta)\big) \tag{1}$$

Alternatively, *Inference methods* seek to describe the unknown parameters probabilistically, accounting for various sources of uncertainty in the analysis. Such approaches commonly apply Bayes rule to enumerate the posterior distribution,

$$p(\theta|R) = \frac{p(\theta)p(R|\theta)}{p(R)} \tag{2}$$

where $p(\theta)$ is the prior over parameters, $p(R|\theta)$ is the likelihood model that encompasses both the forward reflection model and the instrument parameters and $p(R)$ is the (usually intractable) marginal data likelihood. Since the observed $R(Q)$ is derived from count data, it may be possible to derive an exact likelihood function. However, the authors note that in practice it is often sufficient to approximate $p(R|\theta)$ as a Gaussian with variances derived from the count data directly.

In both the optimisation and inference setting, powerful techniques based on automatic differentiation exist in the ML literature for solving inverse problems. Access to the gradients of the error function (from this work, Fig. 1) means that these approaches can be readily applied to NR. The remainder of this section will be dedicated to introducing approaches from both categories that are applied in the case studies outlined in this paper.

Optimisation for NR by gradient descent

Although several approaches that use approximate gradient information[11–13] are common in the NR literature, we are not aware of any method that explicitly harnesses the analytical gradient with respect to the parameters of interest using automatic differentiation techniques. Methods in this class have seen intensive development since the rise of deep neural networks, particularly stochastic gradient descent methods. Access to gradient information as in this paper will make the application of these schemes very straightforward to the NR community. Efficient implementations of many popular algorithms are readily available in a number of programming languages. The most popular methods (e.g. Adam[27], RMSPROP[28]) make use of gradient averages between iterations and momentum terms that help to overcome local minima. Such approaches help to alleviate some of the inherent shortcomings of these popular gradient descent approaches when applied to NR data analysis, such as becoming trapped in local minima and computational efficiency in the case of high dimensions of very many parameters.

Inference for NR by Hamiltonian Monte-Carlo

In the interest of keeping this paper self-contained and in order to appeal to practitioners of NR who may not be familiar with gradient-based sampling and approximate inference schemes, brief overviews of Hamiltonian Monte-Carlo (HMC) and variational inference (VI) for NR practitioners are included here. The interested reader is directed to[29] (HMC) and[30] (VI) for more comprehensive tutorial overviews of the method.

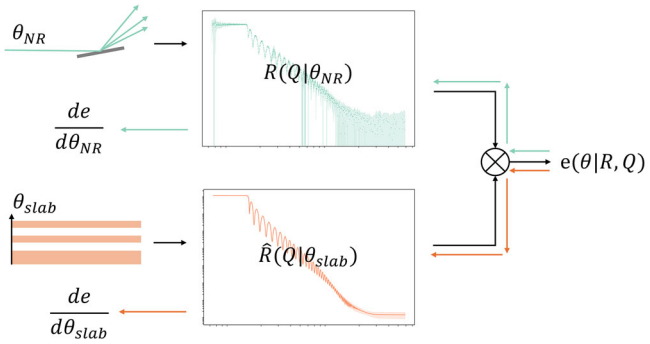


Fig. 1 | Automatic-differentiation back propagates the gradient of the error function through the forward reflectivity model by successive application of the chain rule. The result is the gradient of the error function with respect to the parameters of interest.

The essential idea behind HMC is to construct a pseudo-dynamical system in the space of the unknown NR and slab parameters, designed such that the stationary density of the resulting position vector is the target posterior distribution of interest,

$$\theta(t) \sim p(\theta|R) \tag{3}$$

Thus, one can simply simulate the dynamic system and interpret the time series as samples from the posterior in the same way as a Markov-chain approach. By requiring that the proposed dynamics of $\theta(t)$ are Hamiltonian, one is able to express the joint distribution of position θ and momentum r in terms of the potential and kinetic energies.

$$H(\theta, r) = U(\theta) + K(r) \tag{4}$$

$$p(\theta, r) \propto \exp(-H(\theta, r)) = \exp(-U(\theta))\exp(-K(r)) \tag{5}$$

From the above, it can be straightforwardly shown that the choice $U(\theta) = -\log p(\theta|R)$ leads to the required stationary density for θ. A common choice for the momentum distribution is to choose a zero-mean Gaussian distribution with 'mass' matrix M as a hyperparameter of the method $K(r) = \frac{1}{2} r^T M^{-1} r$. The resulting Hamiltonian system is governed by the ordinary differential equation (ODE),

$$\begin{aligned} \frac{d\theta}{dt} &= \frac{\partial H}{\partial r} = M^{-1} r \\ \frac{dr}{dt} &= -\frac{\partial H}{\partial \theta} = -\nabla_\theta U(\theta) = \nabla_\theta \log\big[p(\theta)p(R|\theta)\big] \end{aligned} \tag{6}$$

Where the differentiation has removed the unknown normalisation constant in the kinetic energy and the ODE now depends on the hyperparameters, likelihood and prior only. As with any other nonlinear ODE it is necessary to integrate forward in time numerically in order to obtain trajectories (and therefore samples from the target posterior distribution over the slab and NR parameters). Symplectic integrators are a class of integrator specifically designed to simulate Hamiltonian dynamics, by attempting to conserve (or bound the drift in) the systems total energy through the Hamiltonian. A popular solver in the context of HMC is the leapfrog algorithm[31]. For a given integration period Δt, the dynamics are updated sequentially as,

$$\begin{aligned} r_{t+\frac{1}{2}} &= r_t - \tfrac{\Delta t}{2} \nabla_\theta U(\theta_t) \\ \theta_{t+1} &= \theta_t + \Delta t\, M^{-1} r_{t+\frac{1}{2}} \\ r_{t+1} &= r_{t+\frac{1}{2}} - \tfrac{\Delta t}{2} \nabla_\theta U(\theta_{t+1}) \end{aligned} \tag{7}$$

Thus, the overall sampling scheme consists of a Markov chain with the following kernel. In each iteration, an initial momentum is drawn, and the integration is run for a number of steps L with a step size of Δt. Although the

position at the end of the integration window should theoretically be a valid sample from the target distribution, in practice, numerical errors in the integration can lead to bias. To overcome this, a Metropolis-Hastings step is typically taken after integration. At the end of the integration window, the last value of the parameter is accepted as the next state of the Markov Chain with probability α. It can be shown (in[29] for example) that the correct choice of α to ensure detailed balance in the Markov Chain is given by,

$$\alpha = \min\{1, \exp(H(\theta, r) - H(\theta', r'))\} \tag{8}$$

HMC for NR practitioners. In practice, the performance of HMC comes to depend strongly on the choice of the hyperparameters L, Δt and M. The optimal setting of these parameters strongly affects the sampling efficiency (how many samples must be drawn in order to robustly explore the posterior), and is affected by the geometry of the posterior, such as by parameters that vary on vastly different scales. For NR practitioners working with HMC, it is advisable to pre-scale parameters such that the parameter variances have the same order of magnitude. An example would be fitting the thickness of a wide layer, or a very small background term subject to a fixed scaling value to improve numerical conditioning.

A further factor that can lower sample efficiency in HMC is sharp curvature in the posterior geometry. A typical example in NR is parameters with hard constraints e.g. layer thicknesses or SLDs constrained to be positive. Close to a zero value, the posterior has infinite curvature and thus the leapfrog integrator presented above can quickly accumulate errors. A common mitigation that can work well in the context of NR inversion is to run the HMC sampler in log space and then exponentiate for the physical parameters once the sampling is completed.

Fortunately, the adaptive specification of HMC hyperparameters is a well studied problem. Variants of the standard HMC algorithm such as the no U-turn sampler (NUTS)[32] have shown incredible robustness to a wide range of posterior geometries. NUTS is now a standard approach in Bayesian inference and implementations can be found in many programming languages.

With regard to evaluation of the performance of HMC in the context of NR, all methods that are appropriate for MCMC can be applied to HMC. A common statistic is the *effective sample size*[33], a measure of representation power that is lower for chains that are non-stationary and poorly converged. The authors note that although acceptance rate can be a useful measurement of sampler performance, in HMC it is not unusual to see rates above 80% in well converged chains.

Approximate inference for NR by variational inference

HMC has shown good performance across a range of inference scenarios. However, in the limit of very large data and many unknown parameters, even advanced sampling schemes can suffer from slow convergence and high computational complexity. In such settings, variational approaches (collectively known as variation inference (VI)) have shown significant promise by approximating the true posterior with a parametric distribution[34]. VI essentially transforms the inference problem back into an optimisation problem that can be solved by gradient descent methods.

Variational inference methods differ from sampling approaches in that they attempt to directly *approximate* the posterior distribution with some parametric form that is straightforward to sample from,

$$p(\theta|R) \approx q(\theta; \phi) \tag{9}$$

Where ϕ are some deterministic unknown parameters that control the shape of a surrogate distribution q. It is important here that the choice of q is a valid distribution (that it integrates to unity and covers the support of the posterior) regardless of the parameters ϕ. For this reason it is common to choose families of probabilities distributions or mixture models for q[35,36].

With q selected, all that remains is to find the optimal setting of ϕ such that the surrogate distribution forms the best possible approximation to the true (but unknown) posterior distribution $p(\theta|R)$ over the unknown slab and NR parameters. To achieve this, we first require a notion of how far our approximation is from the true posterior. To measure divergence between probability distributions, we can use the Kullback-Leibler divergence (KL divergence), defined as,

$$KL(q||p) = \mathbb{E}_q\left[\log \frac{q(\theta; \phi)}{p(\theta|R)}\right] \tag{10}$$

thus, the optimal choice for the variational parameters ϕ is,

$$\phi^* = \arg\min_{\phi} KL(q(\theta; \phi)||p(\theta|R)) \tag{11}$$

Note that this form still contains the unknown posterior $p(\theta|R)$, and so some rearrangement is required. Recalling that from Bayes rule,

$$p(\theta|R) = \frac{p(\theta)p(R|\theta)}{p(R)} \tag{12}$$

$$KL(q||p) = \mathbb{E}_q\left[\log \frac{q(\theta; \phi)p(R)}{p(\theta)p(R|\theta)}\right] \tag{13}$$

$$= \mathbb{E}_q\left[\log q(\theta; \phi) - \log p(\theta)p(R|\theta) + \log p(R)\right] \tag{14}$$

Since the log evidence is constant we can bring it outside the expectation,

$$KL(q||p) = \mathbb{E}_q\left[\log q(\theta; \phi) - \log p(\theta)p(R|\theta)\right] + \log p(R) \tag{15}$$

Given that we want to minimise the KL divergence between q and p and that it is invariant to changes in ϕ we can essentially ignore this term in the optimisation and focus on the expectation in the above. This expectation term is often called the 'evidence lower bound' (ELBO)[30] because it provides a lower bound on the unknown evidence term. This can be proved with Jensen's inequality for probability distributions[34]. For a concave function f,

$$f(\mathbb{E}[\theta]) \geq \mathbb{E}[f(\theta)] \tag{16}$$

Now rewriting the log evidence term as,

$$\log p(R) = \log \int p(\theta)p(R|\theta)d\theta \tag{17}$$

$$= \log \int p(\theta)p(R|\theta)\frac{q(\theta; \phi)}{q(\theta; \phi)}d\theta \tag{18}$$

$$= \log \mathbb{E}_q\left[\frac{p(\theta)p(R|\theta)}{q(\theta; \phi)}\right] \tag{19}$$

since the logarithm is a concave function, Jensen's inequality gives that,

$$\log p(R) = \log \mathbb{E}_q\left[\frac{p(\theta)p(R|\theta)}{q(\theta; \phi)}\right] \geq \underbrace{\mathbb{E}_q\left[\log \frac{p(\theta)p(R|\theta)}{q(\theta; \phi)}\right]}_{\text{ELBO}} \tag{20}$$

Therefore, maximising the ELBO is equivalent to minimising the KL divergence,

$$\phi^* = \arg\max_{\phi} \text{ELBO} \tag{21}$$

Thus, the appropriate loss function is,

$$\mathcal{L}(\phi) = -\text{ELBO} = \mathbb{E}_q\left[\log q(\theta|\phi) - \log p(\theta)p(R|\theta)\right] \tag{22}$$

Note that these expectations are rarely available analytically. In practice, they are approximated by Monte-Carlo. A common choice in the VI literature is to make this approximation using a single sample[37].

$$\mathcal{L}(\phi) \approx \frac{1}{N}\sum_{i=1}^{N}\left[\log q(\theta^{(i)}|\phi) - \log p(\theta^{(i)})p(R|\theta^{(i)})\right] \tag{23}$$

where the $\theta^{(i)}$ are the N samples drawn from q to approximate the expectation. Taking the gradient of the above with respect to the unknown parameters ϕ,

$$\nabla_\phi\mathcal{L}(\phi) \approx \frac{1}{N}\sum_{i=1}^{N}\nabla_\phi\left[\log q(\theta^{(i)}|\phi) - \log p(\theta^{(i)})p(R|\theta^{(i)})\right] \tag{24}$$

It has been observed empirically that single sample Monte-Carlo approximations to the ELBO have low variance and give good performance[37]. However, care must be taken to ensure that the gradients remain available through the sampling procedure. A common approach is to reparametrise q such that the sampling procedure does not depend on ϕ.For more detail on the reparameterisation trick see ref. 37.

VI for NR practitioners. In the context of NR, the loss gradient above once again requires access to the gradient of the forward reflectivity model, made possible by the contribution of this work. With access to the gradient of the ELBO, all that remains is to solve the optimisation problem with a gradient descent scheme such as one of those described in 'Optimisation for NR by gradient descent' above.

It is important to note here that VI based on minimisation of the KL divergence can display 'mode-seeking' behaviour that emphasises high probability regions of the posterior at the expense of lower probability areas leading to underprediction of variance. For a discussion of why this occurs and advice in the context of NR see Supplementary Information Section S3.

In the limit of very large data (>10k samples), computing the expectations above (even with a single sample approximation) may become prohibitively expensive. In such contexts, a particularly powerful extension is *Stochastic variational inference* (SVI)[35]. The results of this paper indicate that SVI is particularly promising in the large data, complex model regime of NR having been successfully applied to datasets numbering millions of examples in other fields[35]. However, because most static NR datasets comprise <1k samples, only a standard VI approach is considered in this work.

An important choice in VI is the selection of a surrogate distribution q that will approximate the true posterior. In an NR context it is important that the support of this distribution covers the expected support of the parameters of interest. For example, if two parameters are expected to be correlated (e.g. thickness and SLD in a kinetic heating experiment) then a q must be selected that can represent this correlation. For example, independent Gaussian distributions would be unable to reproduce this correlation, whereas a multivariate Gaussian would (at the cost of additional parameters to be optimised).

refjax: an open-source library for gradient-based NR

Rather than implement any single bespoke gradient-based optimisation or inference technique for NR, the authors present here an open-source software library for fast computation of forward reflectivity and gradients in the python language. The library, titled refjax, is available at github.com/MDCHAMP/refjax. The library is built on top of the python JAX ecosystem from Google[38] that provides flexible automatic differentiation, parallelism and GPU acceleration. refjax provides simple access to the forward specular reflectivity calculation (including smearing) for an arbitrary slab model. In refjax, all input parameters are collected in standard python dictionaries with a standardised format.

These parameters can then be provided to a kernel to compute the forward reflectivity.

Because the library is built upon the JAX ecosystem, just in time compilation, parallelism and gradient computation are handled naturally.

The module also provides a number of smearing kernels including constant and pointwise Gaussian smearing.

A key advantage of the provided library is the freedom to apply any optimisation routine developed in the JAX ecosystem and beyond. To date, highly performant packages are available for traditional optimisers (LM, BFGS)[39], (stochastic) gradient descent[40], HMC (including the no U-turn sampler[32])[41] and others. Several examples of how to use the library in conjunction with popular ML optimisation suites are provided at https://github.com/MDCHAMP/refjax/tree/main/examples.

```
theta = {
    "thick": jnp.array([    1500, 10.0     ]), # thicknesses (n,)   n layers (A)
    "SLD":   jnp.array([0.0, 5.0, 4.2, 2.07]), # SLD          (n+2,) fronting + n layers + backing (A^-2)
    "rough": jnp.array([   5.0, 5.0, 2.07  ]), # roughnesses (n+1,) interfaces between media
    "scale": 1.0,
    "bkg":   2e-7,
}
```

```
import refjax
kernel = refjax.kernel(refjax.model.abeles)
R_hat = kernel(theta, Q)
```

```
kernel_jit = jax.jit(kernel) # just-in-time compilation

def mean_squared_error(theta): # define a loss function
  R_hat = kernel_jit(theta, Q)
  return ((R-R_hat)**2).sum()

de_dtheta_fn = jax.grad(mean_squared_error)(theta) # returns gradient of loss

jax.pmap(kernel_jit)(thetas, Qs) # parallel evaluations of the kernel+gradient are possible on GPU/TPU
    etc.
```

Results

Crystalline quartz

The effectiveness of gradient-based approaches to NR inverse problems and the software we have developed is demonstrated here for two case-study examples. In each case the forward reflectivity computation and automatic differentiation is implemented using the refjax software.

In the first case study, the performance of gradient based techniques are highlighted on an existing dataset. To this end, the parameters of a highly homogeneous crystal quartz film deposited on a flat silicon wafer with a thin native oxide layer are considered. The reflectivity data were collected on the D17 neutron reflectivity instrument at the ILL (Grenoble, France)[42]. As part of the data collection, pointwise full-width at half maximum (FWHM) smearing kernel estimates were estimated. These were used to perform smearing in the forward reflectivity model as fixed known parameters. The slab model considered in this case study is provided in Table 1 and the NR parameters are displayed in Table 2. All parameters are considered unknown and are fit using both the optimisation and inference methods.

First, an optimisation approach is considered. To perform the model fit, the widely-used stochastic gradient descent technique ADAM[27] is employed. Starting from an approximate initial values for the relevant parameters (listed in Tables 1 and Table 2), the optimiser descends the gradient of a χ^2 error metric,

$$\chi^2 = \sum \frac{\left(\hat{R}(Q) - R(Q)\right)^2}{\sigma^2_{R(Q)}} \tag{25}$$

for 2000 iterations with a learning rate of 0.05. The resulting fit to the data is given in Fig. 2. As can be seen in the Figure, the fit to the data is excellent,

```
kernel = refjax.kernel(abeles, refjax.smear.constant_Gaussian)
smeared_R_hat = kernel_pointwise(theta, Q, dQ)
```

achieving a χ^2 error of 1.302. This is marginally lower than the value of $\chi^2 = 1.32$ reported in ref. 42, indicating that the approach has found optimal parameters. Note that because the optimisation is unconstrained, and the data are not very sensitive to the natural oxide layer, the fitted values for this layer should be interpreted with care.

With the effectiveness of gradient based optimisation established, a further study considering gradient-powered inference for NR fitting is presented here. The same quartz data is used to perform Bayesian inference of the unknown parameters by HMC. The NUTS sampler[32] (as implemented in the 'numpyro' library[41]) is used to sample from the posterior distribution over the parameters. NUTS has the advantage that the hyperparameter L is set automatically in each iteration leading to efficient exploration of the typical set. The other HMC parameters are adapted using the default strategy in the numpyro library for a burn-in period of 1000 samples.

Table 1 | The Slab model for the Quartz sample including initial parameter guesses for optimisation, priors used for inference and identified posterior mean and standard deviations

Layer	Parameter	Initial value	Prior distribution	Posterior(mean ± stdev.)
Air	SLD ($Å^{-2}$) Thickness (Å)	0.0 ∞		
	Interfacial roughness (Å)	5.0	$\mathcal{U}(2.0, 20.0)$	11.402 ± 0.0827
Quartz (Si02)	SLD ($Å^{-2}$) Thickness (Å)	5.0 1500	$\mathcal{U}(0.0, 5.0)$ $\mathcal{U}(1400.0, 1500.0)$	4.240 ± 0.0015 1470 ± 0.487
	Interfacial roughness (Å)	5.0	$\mathcal{U}(2.0, 20.0)$	10.905 ± 0.43
Native oxide (Si02)	SLD ($Å^{-2}$) Thickness (Å)	4.2 10	$\mathcal{U}(0.0, 5.0)$ $\mathcal{U}(0.0, 50.0)$	0.323 ± 0.216 2.122 ± 0.415
	Interfacial roughness (Å)	5.0	$\mathcal{U}(2.0, 20.0)$	3.163 ± 0.733
Silicon backing (Si)	SLD ($Å^{-2}$) Thickness (Å)	2.07 ∞	$\mathcal{U}(0.0, 5.0)$	2.195 ± 0.0236

Note the large range in prior values corresponding to very weak prior knowledge.

The prior distributions were set to be uniform in the vicinity of the initial parameters used by the gradient scheme (see Tables 1, 2). Note that the Bayesian approach enables us to incorporate prior knowledge and constrain these parameters to physically-plausible values. The chain was initialised prior to the burn-in at the optimal value found by gradient descent. Overall, 10 independent chains were run, each comprising 2000 samples drawn after burn-in. Posterior means and standard deviations are recorded in Tables 1, 2. The resulting predictive posterior $p(\hat{R}|R)$ is plotted in Fig. 3, wherein an excellent fit to the data is seen taking the posterior mean parameters and performing further optimisation using a BFGS optimiser (SciPy implementation) results in a χ^2 score reduction of only 0.004 indicating that the inference has found optimal parameters. The mean prediction of the posterior $E[\hat{R}]$ has a χ^2 error of 1.22, an improvement over previously published results ($\chi^2 = 1.32$,[42]) although we note that in our work, all parameters are left free to fit. In Fig. 4, the marginal posterior distributions of the NR parameters of interest are shown alongside the distribution over the SLD profile of the quartz film.

Table 2 | NR parameters for the quartz data including initial parameter guesses for optimisation, priors used for inference and identified posterior mean and standard deviations

Parameter	Initial value	Prior distribution	Posterior (± σ)
Background	10^{-7}	$\mathcal{U}(e^{-20}, 1)$	$(1.05 \pm 0.201) \times 10^{-7}$
Scale parameter	1.0	$\mathcal{U}(0.9, 1.5)$	1.048 ± 0.00363

It is noteworthy that the obtained results are obtained by drawing only 2000 samples from the NUTS algorithm per chain, a fraction of the number of samples typically required by MCMC methods for good convergence. This phenomenon is in line with the efficiency gains expected from HMC over traditional MCMC methods, as demonstrated in other fields[43,44]. However, it is of interest to quantify this phenomenon in the context of NR inversion.

In order to compare the efficiency of HMC (here NUTS) against gradient-free MCMC methods, Fig. 5 plots the effective sample size (ESS) against iteration for both HMC and a Sample-adaptive MCMC (SA-MCMC) approach[45] implemented in the numpyro library[41]. ESS gives a measure of how many independent samples a correlated Markov-chain is

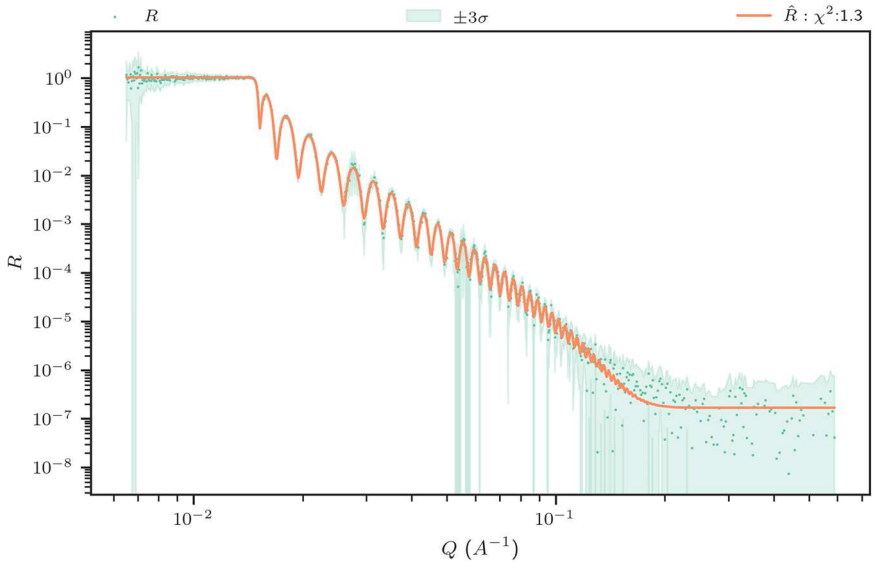


Fig. 2 | Optimisation of quartz thin film NR dataset fit using gradient decent with the ADAM optimiser[27]. All parameters are floating and so free to fit. The identified χ^2 error is lower than the value of $\chi^2 = 1.32$ reported previously in ref. 42 for the same data.

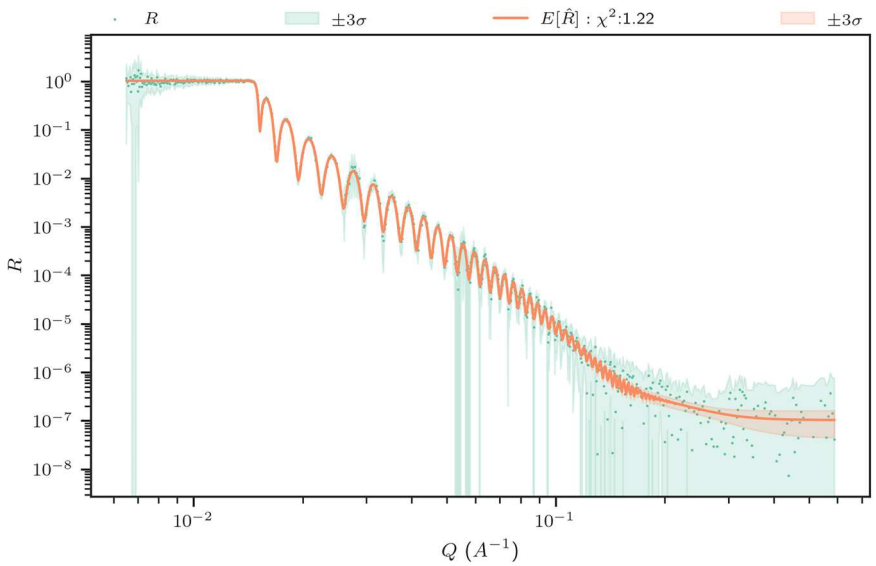


Fig. 3 | Inference of quartz NR model by Hamiltonian Monte-Carlo. All parameters remain free to fit. It is notable that the mean of the posterior distribution has a lower error than values identified by the deterministic optimisation routine. Posterior uncertainty is given by the width of the orange shaded region, corresponding to a ± 3σ interval.

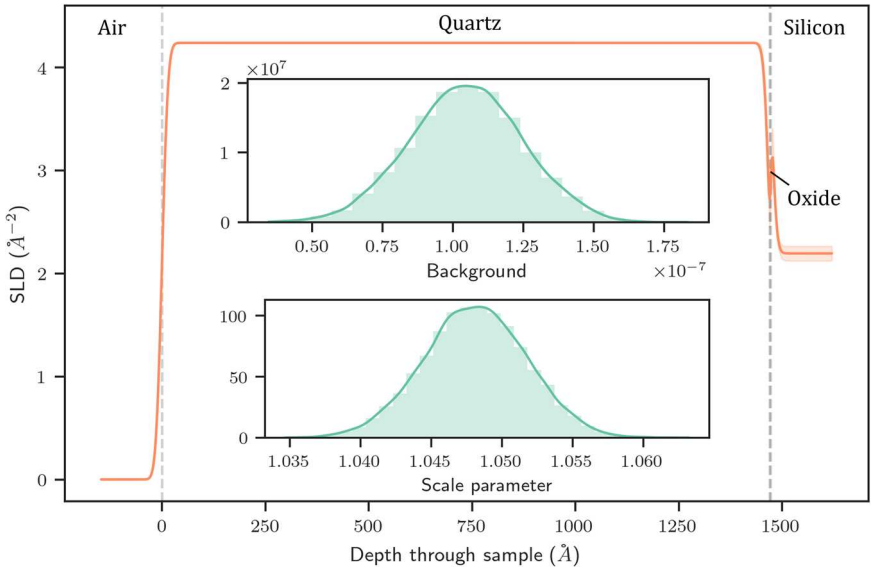


Fig. 4 | Posterior SLD profile and posterior marginal parameter distributions for the NR parameters of the quartz film, obtained by Hamiltonian Monte-Carlo. Error bars on the SLD profile represent $\pm 3\sigma$ intervals, these are visible on the right for the oxide and Silicon.

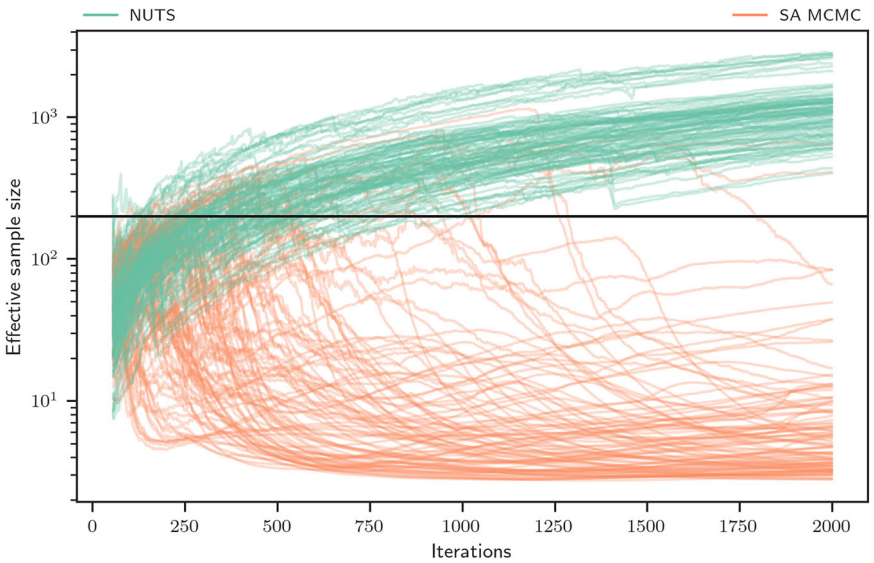


Fig. 5 | Comparison of effective sample sizes per iteration between a Hamiltonian Monte-Carlo (NUTS) scheme (green) and a gradient-free sample adaptive MCMC approach (orange). Each line is the effective sample size of a single parameter in a single chain. The horizontal line corresponds to a sample size of 200 which is a common rule of thumb for well converged chains in the literature[49].

equivalent to. A common rule of thumb in the Bayesian inference literature is that an ESS value of 200 per parameter represents a good coverage of the posterior for well-converged chains, whereas low values can correspond to poor exploration and autocorrelation.

Samples from both NUTS and SA-MCMC are collected from 10 independent chains for 2000 iterations, starting from the same initialisation point at the optimal parameter values found by gradient descent. At each iteration the ESS for each parameter in each chain is plotted in Fig. 5.

As can be seen in Fig. 5, the NUTS samples increase in ESS broadly monotonically with iteration, indicating good convergence and efficient exploration of the typical set even after only 1000 samples. SA-MCMC appears to start well (by exploring a local region of the typical set efficiently) but slow mixing and high auto-correlation lead to a decrease in ESS. It is interesting to note that although the per-sample computational complexity for NUTS is several times higher than SA-MCMC, better exploration of the typical set is seen with the gradient-based scheme, in line with results from other fields[43,44].

Organic LED

With good performance shown on data from a relatively simple slab model, it is next of interest to consider how gradient-based approaches perform on more complex samples. Here, data from several organic light-emitting diode devices architectures are presented that correspond to different stages in the processing of these devices.

The organic (OLED) devices are fabricated using spin coating. They comprise three main layers. The first being Poly(3,4-ethylenedioxythiophene):poly(styrene sulphonate) (PEDOT:PSS): This serves as the hole injection layer (HIL), facilitating charge transfer from the anode. Then Poly[(9,9-dioctylfluorenyl-2,7-diyl)-co-(4,4'-(N-(4-sec-butylphenyl)diphenylamine))] (TFB): This layer functions as both a hole transport layer (HTL) and an electron blocking layer (EBL), ensuring efficient charge recombination in the emissive zone. Finally, Poly[(9,9-di-n-octylfluorenyl-2,7-diyl)-alt-(benzo[2,1,3]thiadiazol-4,8-diyl)] (F8BT), this is the emissive layer, responsible for green light emission upon charge recombination. The TFB was deuterated to provide suitable neutron scattering length density contrast to accurately probe the interfacial mixing of TFB and F8BT.

Table 3 | The Slab model for the LED data including initial parameter guesses for optimisation and priors used for inference

Layer	Parameter	Initial value	Prior distribution
Air	SLD ($Å^{-2}$) Thickness (Å)	0.0 ∞	
	Interfacial roughness (Å)	10.0	$\mathcal{U}(0.0, 30.0)$
F8BT	SLD ($Å^{-2}$) Thickness (Å)	1.0 700.0	$\mathcal{U}(0.0, 5.0)$ $\mathcal{U}(0.0, 800.0)$
	Interfacial roughness (Å)	10.0	$\mathcal{U}(0.0, 30.0)$
TFB	SLD ($Å^{-2}$) Thickness (Å)	4.0 400.0	$\mathcal{U}(0.0, 5.0)$ $\mathcal{U}(0.0, 800.0)$
	Interfacial roughness (Å)	10.0	$\mathcal{U}(0.0, 30.0)$
PEDOT:PSS	SLD ($Å^{-2}$) Thickness (Å)	1.5 600	$\mathcal{U}(0.0, 5.0)$ $\mathcal{U}(0.0, 800.0)$
	Interfacial roughness (Å)	5.0	$\mathcal{U}(0.0, 8.0)$
Native oxide (Si02)	SLD ($Å^{-2}$) Thickness (Å)	3.0 50	$\mathcal{U}(0.0, 5.0)$ $\mathcal{U}(0.0, 800.0)$
	Interfacial roughness (Å)	5.0	$\mathcal{U}(0.0, 8.0)$
Silicon backing (Si)	SLD ($Å^{-2}$) Thickness (Å)	2.07 ∞	$\mathcal{U}(0.0, 5.0)$

Table 4 | NR parameters for the LED data including initial parameter guesses for optimisation and priors used for inference

Parameter	Initial value	Prior distribution
Scale parameter (per device)	1.0	$\mathcal{U}(0.9, 1.5)$
Background (shared)	5.0	$\mathcal{U}(e^{-20}, 1 \times 10^{-7})$
Smearing parameter (%) (shared)	1.0	$\mathcal{U}(0.0, 10.0)$

A slab model of the LED device is given in Table 3. Overall, four devices are considered. The first is not annealed after manufacture and was tested as cast. The other three were annealed at temperatures of 80 °C, 140 °C, and 180 °C respectively. Because smearing kernel sizes are not available for this dataset, a constant Gaussian smear is adopted with its size left as a further unknown parameter (see Table 4). The four devices were measured on the same instrument and were fabricated on nominally identical silicon substrates. It is therefore physically reasonable to treat only the instrument-related parameters (the background level and the resolution smearing) and the substrate scattering length density as being shared between devices. We emphasise that the physically meaningful slab parameters of the polymer layers (their thicknesses, SLDs and interfacial roughnesses) are left entirely free and are inferred independently for each device. The co-fit therefore does not assume that the samples are identical; it encodes only the fact that they share a common measurement apparatus and substrate, which reduces the number of instrument nuisance parameters that must be inferred. For this reason (and to demonstrate the effectiveness of variational inference for NR) the parameters of all four devices are inferred jointly, with the background and smearing parameters being common to all four devices. In addition, the SLD value of the silicon backing (which is expected to be very close to the literature value of 2.07 for all four devices) is treated as a co-fit parameter. In general, wider priors are adopted that apply to all layers, representing weak prior knowledge. However, in order to encourage physically plausible roughness values around the native silicon oxide layer, the prior over the roughnesses either side of the oxide are constrained to values less than 8.0. In total this results in a posterior distribution over 59 unknown parameters. Given the large parameter size, approximate inference with VI will be applied here.

To produce an approximate inference over the co-fit posterior of all four devices a VI scheme is adopted. The surrogate posterior q is selected to be a multivariate normal distribution defined in an unconstrained space. That is, all parameters are mapped from their physical bounded space to one that is unconstrained by an appropriate invertible transformation. The resulting ELBO is evaluated by automatic differentiation (using the numpyro library) and the refjax software for reflectivity kernel calculations. The mean of the surrogate posterior is initialised at identical values for each device and the variance is initialised as a diagonal matrix $\sigma_0 I$ with $\sigma_0 = 0.1$. The ELBO is optimised using the ADAM[27] optimiser for 2000 iterations with a learning rate of 0.05. To ensure that the optimisation does not become trapped in local minima, the optimisation procedure is run 100 times and the best (largest) value of the ELBO is retained as surrogate posterior. VI has significant computational advantages over sampling based approaches, each run of the inference is complete in <20 s on our hardware (AMD Ryzen Threadripper PRO 5995WX). Successive restarts can be run entirely in parallel and so the VI approach is very attractive with regard to modern CPU and GPU acceleration.

The predictive results of the inference are plotted in Fig. 6 (for full results see Supplementary Information Section S1). As can be seen in the Figure, the fit to data is very good with $\chi^2 < 3$ in all cases. Examining the SLD profiles, it appears that the samples annealed for the longest have higher SLD in the TFB layer due to extended annealing above the glass transition temperature T_g, this will help remove any trapped processing solvent as well as enable the layer to achieve an equilibrium density. We also see changes in the interfacial width between the two polyfluorenes which we are able to more rigorously quantify using this fitting approach. It is of note that the identified SLD of the Silicon backing (2.09 ± 0.0178) is fit very close to the literature value of 2.07 indicating convergence to a physically plausible slab model. The posterior distribution over the thickness of the native oxide layer for the device annealed at 140 °C is markedly larger than for the other three devices. Because the reflectivity is only very weakly sensitive to this thin, buried layer, we caution that this difference is almost certainly a fitting artefact rather than a genuine physical change between the devices, and the oxide-layer parameters should be interpreted with corresponding care. We further note that, for the as-cast device, the measured reflectivity displays slightly higher-amplitude fringes at high Q than the fitted model (Fig. 6a); we attribute this to the single resolution-smearing parameter that is shared across all four devices, which may slightly over-smear the as-cast curve. Device-specific sample non-uniformity or layer polydispersity, which are not captured by the shared smearing and slab model, may also contribute. A demonstration of both the HMC and VI approaches on a standard, previously published lipid bilayer dataset[9] is provided in Supplementary Information Section S2, together with a discussion of the variance under-prediction inherent to VI in Supplementary Information Section S3.

Discussion

The results of our work demonstrate the potential of automatic differentiation and gradient-based optimisation techniques in the context of neutron reflectometry data analysis. Excellent fits to both simple and complex NR data are demonstrated, in one case exceeding the quality of fit compared to previous literature results despite wide priors.

In the optimisation setting, access to cutting edge gradient descent algorithms will enable practitioners to quickly and efficiently fit NR data, particularly as samples become more complex and traditional approaches (LM, BFGS) become prone to computational complexity and become stuck in local minima in the large data limit.

In the inference setting, the application of Hamiltonian Monte Carlo to NR data has shown considerable sample efficiency gains over traditional gradient-free MCMC methods. HMC is a promising replacement for MCMC in NR in scenarios where accurate uncertainty quantification is required and complex posterior distribution geometries (multimodality, highly correlated variables) make NR inference challenging.

In the second case-study, techniques from variational inference show remarkable performance in a complex NR fitting task. VI is shown to be a potentially revolutionary technique for several NR applications particularly

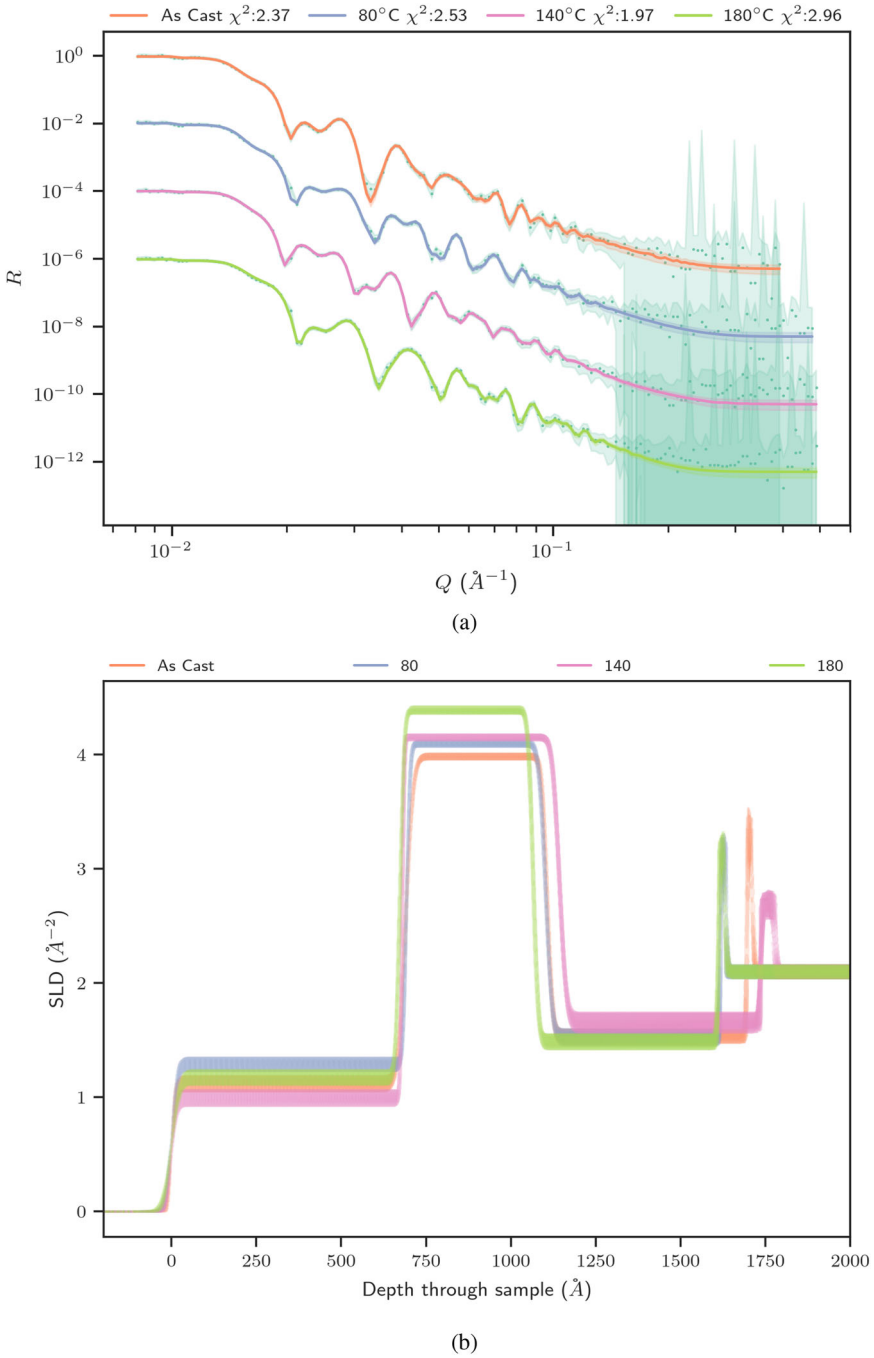


Fig. 6 | . a Results of the inference for the LED devices. Note that the reflectivity curves are spaced for visibility on the y-axis. **b** Samples from the distributions of the posterior SLD profiles.

in settings whereby uncertainty quantification of high-throughput data is critical. An example is fast kinetic NR experiments whereby large quantities of NR data are collected dynamically and in-situ inference of model parameters is required. A key application of VI in NR might be in real-time parameter inference for closed-loop process control, something that has thus far required fast model surrogates and has been unattainable with traditional inverse modelling techniques. Another application could be in iterative experimental design whereby the parameter posterior distributions of initial experiments are used to inform the design of subsequent experiments during a single beamtime. Here, techniques from Bayesian optimisation[46] may provide a principled way to select the most informative experiments sequentially. A related opportunity is to use the proposed approach to provide fast inversion to be used during active learning of sequential NR experiments as in ref. 47.

Alongside the case-studies developed here, the authors make available an open-source software library for fast computation of forward reflectivity and gradients in the python language. It is the intention of the authors to use this platform as a springboard to develop further forward reflectometry models for other indirect measurement techniques such as small-angle neutron scattering, ellipsometry and grazing-incidence X-ray scattering (where HMC has previously been shown to be a successful approach[48]). In conclusion, we have demonstrated that the availability of fast-forward reflectometry models is critical to ensure that NR fitting keeps pace with high-throughput experimentation and the rapid advances in ML and optimisation techniques. This paper presents a promising step in this direction by ensuring that gradient-based inverse methods are readily available for NR data analysis.

Data availability

Regarding the Quartz film data used in this manuscript, for users of the ILL, the data logbook and reduction parameters can be found under cycle 162/D17/InternalUse/processed at http://data.ill.fr/. However, to ensure all readers can access this dataset we additionally make the data accessible as part of the refjax library in the example data folder at https://github.com/MDCHAMP/refjax/tree/main/examples. The LED device data used in this manuscript is made available as part of the refjax library in the example data folder at https://github.com/MDCHAMP/refjax/tree/main/examples. The Lipid bilayer system example used in the appendices of this paper is freely available from freely available at https://github.com/reflectivity/reproducibility/tree/master/examples/refnx

Code availability

The refjax software library is available as open-source code at https://github.com/MDCHAMP/refjax, including several examples of fitting routines.

Received: 13 November 2025; Accepted: 14 August 2026;
Published online: 03 September 2026

References

1. Skoda, M. W. Recent developments in the application of x-ray and neutron reflectivity to soft-matter systems. *Curr. Opin. Colloid Interface Sci.* **42**, 41–54 (2019).
2. Kirby, B. J. et al. Phase-sensitive specular neutron reflectometry for imaging the nanometer scale composition depth profile of thin-film materials. *Curr. Opin. Colloid Interface Sci.* **17**, 44–53 (2012).
3. Junghans, A. et al. Analysis of biosurfaces by neutron reflectometry: From simple to complex interfaces. *Biointerphases* **10** (2015).
4. Majkrzak, C. F. et al. Neutron diffraction studies of fe/cr superlattices. *J. Appl. Phys.* **63**, 4048–4050 (1988).
5. Liley, J. et al. The performance of surfactant mixtures at low temperatures. *J. colloid interface Sci.* **534**, 64–71 (2019).
6. Pabst, G., Kučerka, N., Nieh, M.-P., Rheinstädter, M. & Katsaras, J. Applications of neutron and x-ray scattering to the study of biologically relevant model membranes. *Chem. Phys. Lipids* **163**, 460–479 (2010).
7. Clifton, L. A. et al. An accurate in vitro model of the e. coli envelope. *Angew. Chem. Int. Ed.* **54**, 11952–11955 (2015).
8. Penfold, J. & Thomas, R. The application of the specular reflection of neutrons to the study of surfaces and interfaces. *J. Phys. Condens. Matter* **2**, 1369 (1990).
9. Nelson, A. R. & Prescott, S. W. refnx: neutron and x-ray reflectometry analysis in python. *Appl. Crystallogr.* **52**, 193–200 (2019).
10. Woollam, J. Completeease (2017).
11. Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **2**, 164–168 (1944).
12. Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **11**, 431–441 (1963).
13. Nocedal, J. & Wright, S. J.Numerical Optimization. Springer Series in Operations Research and Financial Engineering, 2nd edn (Springer, 2006).
14. Storn, R. & Price, K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* **11**, 341–359 (1997).
15. Eberhart, R. & Kennedy, J. A new optimizer using particle swarm theory. In *MHS'95. Proceedings of the sixth international symposium on micro machine and human science*, 39–43 (IEEE, 1995).
16. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
17. Hastings, W. K. Monte carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (1970).
18. Skilling J. Nested sampling. *Bayesian inference and maximum entropy methods in science and engineering*, 735:395–405 (2004).
19. Doucet, M., Archibald, R. K. & Heller, W. T. Machine learning for neutron reflectometry data analysis of two-layer thin films. *Mach. Learn.: Sci. Technol.* **2**, 035001 (2021).
20. Hinderhofer, A. et al. Machine learning for scattering data: strategies, perspectives and applications to surface scattering. *Appl. Crystallogr.* **56**, 3–11 (2023).
21. Pithan, L. et al. Closing the loop: autonomous experiments enabled by machine-learning-based online data analysis in synchrotron beamline environments. *Synchrotron Radiat.* **30**, 1064–1075 (2023).
22. Starostin, V. et al. Fast and reliable probabilistic reflectometry inversion with prior-amortized neural posterior estimation. *Sci. Adv.* **11**, eadr9668 (2025).
23. Rentzsch, A. et al. Towards machine-learning-based on-the-fly analysis of neutron reflectometry. *Appl. Crystallogr.* **59** (2026).
24. Munteanu, V. et al. reflectorch: a deep learning package for x-ray and neutron reflectometry. *J. Open Source Softw.* **10**, 8169 (2025).
25. Munteanu, V. et al. Neural network analysis of neutron and X-ray reflectivity data incorporating prior knowledge. *J. Appl. Crystallogr.* **57**, 456–469 (2024).
26. Abelès, F.La théorie générale des couches minces. Ph.D. thesis, Université de Paris. https://www.worldcat.org/title/theorie-generale-des-couches-minces/oclc/1001507736 (1949).
27. Kingma, D. P. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
28. Hinton, G. Lecture 6e rmsprop: Divide the gradient by a running average of its recent magnitude. https://www.cs.toronto.edu/tijmen/csc321/slides/lecture_slides_lec6.pdf. Coursera: Neural Networks for Machine Learning (2012).
29. Betancourt, M. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434* (2017).
30. Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. Variational inference: A review for statisticians. *J. Am. Stat. Assoc.* **112**, 859–877 (2017).
31. Hairer, E., Hochbruck, M., Iserles, A. & Lubich, C. Geometric numerical integration. *Oberwolfach Rep.* **3**, 805–882 (2006).
32. Hoffman, M. D., Gelman, A. et al. The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).
33. Gelman, A. et al. Bayesian data analysis, 3 edn (Chapman and Hall/CRC, Boca Raton, FL, 2013).
34. Jordan, M. I., Ghahramani, Z., Jaakkola, T. S. & Saul, L. K. An introduction to variational methods for graphical models. *Mach. Learn.* **37**, 183–233 (1999).
35. Hoffman, M. D., Blei, D. M., Wang, C. & Paisley, J. Stochastic variational inference. *J. Mach. Learn. Res.* **14**, 1303–1347 (2013).
36. Yin, M. & Zhou, M. Semi-implicit variational inference. In *International conference on machine learning*, 5660–5669 (PMLR, 2018).
37. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. In *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1312.6114 (2014).
38. Bradbury, J. et al. JAX: composable transformations of Python +NumPy programs. http://github.com/jax-ml/jax (2018).
39. Rader, J., Lyons, T. & Kidger, P. Optimistix: modular optimisation in jax and equinox. arXiv preprint arXiv:2402.09983 (2024).
40. DeepMind et al. The DeepMind JAX ecosystem. https://github.com/google-deepmind (2020).
41. Phan, D., Pradhan, N. & Jankowiak, M. Composable effects for flexible and accelerated probabilistic programming in numpyro. *arXiv preprint arXiv:1912.11554* (2019).
42. Gutfreund, P. et al. Towards generalized data reduction on a chopper-based time-of-flight neutron reflectometer. *J. Appl. Crystallogr.* **51**, 606–615 (2018).
43. Hajian, A. Efficient cosmological parameter estimation with hamiltonian monte carlo technique. *Phys. Rev. D. -Part., Fields, Gravit., Cosmol.* **75**, 083525 (2007).

44. Tinto, M. Gravitational wave searches with pulsar timing arrays: Cancellation of clock and ephemeris noises. *Phys. Rev. D.* **97**, 084047 (2018).
45. Zhu, M. Sample Adaptive MCMC. *Advances in Neural Information Processing Systems* 32 (2019).
46. Ru, B., Osborne, M. A., McLeod, M. & Granziol, D. Fast information-theoretic Bayesian optimisation. In *International Conference on Machine Learning*, 4384–4392 (PMLR, 2018).
47. Hoogerheide, D. P. & Heinrich, F. Autorefl: Active learning in neutron reflectometry for fast data acquisition. *Appl. Crystallogr.* **57**, 1192–1204 (2024).
48. Zhang, J., Wang, J., Tirrell, M. V., de Pablo, J. J. & Chen, W. Parameter estimation for X-ray scattering analysis with Hamiltonian Markov Chain Monte Carlo. *J. Synchrotron Radiat.* **29**, 721–731 (2022).
49. Lanfear, R., Hua, X. & Warren, D. L. Estimating the effective sample size of tree topologies from Bayesian phylogenetic analyses. *Genome Biol. Evolut.* **8**, 2319–2332 (2016).

Acknowledgements

We acknowledge Dr Hiroshi Hamamatsu from Sumitomo Chemicals Corporation for producing the OLED device samples whilst on secondment to the University of Sheffield. We are also indebted to Cambridge Display Technologies (CDT) for synthesising and supplying the deuterated conjugated polymers used in the OLED devices. This work was (partially) supported by the Centre for Machine Intelligence at the University of Sheffield.

Author contributions

M.D.C.: Conceptualization, Data curation, Writing—Original Draft, Visualization, Software; A.J.P.: Conceptualization, Data curation, Writing—Review & Editing, Funding acquisition; P.G.: Data curation, Writing—Review & Editing; M.A.S.: Writing—Review & Editing; J.P.A.F.: Funding acquisition, Writing—Review & Editing; T.J.R.: Conceptualization, Writing—Review & Editing, Funding acquisition; S.L.B.: Conceptualization, Funding acquisition, Writing—Review & Editing;

Funding

S.B., P.F., and A.J.P. were supported via an EPSRC prosperity partnership, called Sustainable Coatings by Rational Design, SusCoRD, EP/S004963/1, this was jointly funded by the company AkzoNobel. The OLED data was collected on the ISIS neutron reflectometer instrument SURF, and we are grateful to ISIS for beamtime at ISIS under beamtime application RB 1120359 (DOI 10.5286/ISIS.E.24088355). We are grateful for their support and encouragement of this work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43246-026-01341-4.

Correspondence and requests for materials should be addressed to Max D. Champneys.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.