



# Determining the maximum information gain and optimizing experimental design in neutron reflectometry using the Fisher information

James H. Durant,<sup>a</sup> Lucas Wilkins,<sup>b</sup> Keith Butler<sup>c</sup> and Joshaniel F. K. Cooper<sup>a\*</sup>

<sup>a</sup>ISIS Neutron and Muon Source, Rutherford Appleton Laboratory, Harwell, Didcot, Oxfordshire OX11 0QX, United Kingdom, <sup>b</sup>Department of Zoology, University of Oxford, Mansfield Road, Oxford OX1 3SZ, United Kingdom, and <sup>c</sup>SciML, Scientific Computing Division, Rutherford Appleton Laboratory, Harwell Campus, Didcot, Oxfordshire OX11 0QX, United Kingdom. \*Correspondence e-mail: jos.cooper@stfc.ac.uk

Received 4 March 2021

Accepted 1 June 2021

Edited by E. P. Gilbert, ANSTO, Kirrawee DC, Australia

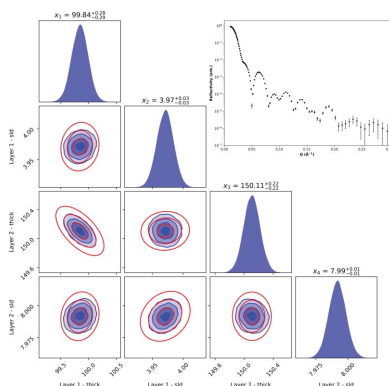
**Keywords:** neutron reflectivity; neutron reflectometry; Fisher information; information theory; analysis; experimental design.

**Supporting information:** this article has supporting information at journals.iucr.org/j

An approach based on the Fisher information (FI) is developed to quantify the maximum information gain and optimal experimental design in neutron reflectometry experiments. In these experiments, the FI can be calculated analytically and used to provide sub-second predictions of parameter uncertainties. This approach can be used to influence real-time decisions about measurement angle, measurement time, contrast choice and other experimental conditions based on parameters of interest. The FI provides a lower bound on parameter estimation uncertainties, and these are shown to decrease with the square root of the measurement time, providing useful information for the planning and scheduling of experimental work. As the FI is computationally inexpensive to calculate, it can be computed repeatedly during the course of an experiment, saving costly beam time by signalling that sufficient data have been obtained or saving experimental data sets by signalling that an experiment needs to continue. The approach's predictions are validated through the introduction of an experiment simulation framework that incorporates instrument-specific incident flux profiles, and through the investigation of measuring the structural properties of a phospholipid bilayer.

## 1. Introduction

The Fisher information (FI) (Fisher, 1925) has been applied across many fields, from information theory and communications (Wang & Yin, 2010; Barnes *et al.*, 2019) to quantum mechanics (Barndorff-Nielsen & Gill, 2000; Petz, 2002), quantitative finance (Taylor, 2019) and volcanology (Telesca *et al.*, 2009). The FI provides a way of measuring the amount of information an observable variable carries about an unknown parameter of a distribution that models the observable. For certain situations it is possible to calculate the FI analytically, giving a measure of parameter uncertainty and inter-parameter covariances from which correlations can be derived. Neutron reflectometry allows one to model a measured reflectivity curve in order to determine the properties of the thin-film layer structure that produced the curve. Most reflectometry analyses use sampling methods to extract parameter uncertainties, though this is expensive and cannot be performed in real time with current software (Nelson & Prescott, 2019; Kienzle *et al.*, 2017; Hughes, 2017). In this work, we describe an application of the FI to neutron reflectometry in enabling real-time estimation of parameter uncertainties, as well as a projection of these with time. We compare the results with established sampling methods and demonstrate the FI's use for experimental design, and for potentially enabling early stopping of experiments based on counting statistics.



OPEN ACCESS

In reflectivity, a thin film is described by a thickness, a scattering length density (SLD), which is the product of the neutron scattering length and the film density, and an interfacial roughness; thin-film heterostructures are composed of multiple thin films on top of each other. In the analysis of reflectometry data, we are presented with data points of reflectivity  $r$  as a function of momentum transfer  $Q$  and wish to infer the SLD profile from the top surface to the substrate. For a single interface, *i.e.* a semi-infinite substrate, the neutron reflectivity decays as  $\sim Q^{-4}$ , also called the Fresnel reflectivity. A single layer on a substrate is analytically solvable (Sivia, 2013), but for more layers multiple reflections are possible, and the inversion of the curve to an SLD profile is non-trivial. In fact, the loss of phase information upon reflection makes inversion of the SLD profile from the reflectivity profile an inverse problem (Majkrzak & Berk, 1995) and approximations are required.

Typically, reflectometry analysis is model dependent, where a model is defined using a series of contiguous layers and the model reflectivity is calculated using the Abelès matrix formalism for stratified media (Abelès, 1948) or the Parratt recursive method (Parratt, 1954). However, the solution to this analysis is not necessarily unique and often requires *a priori* knowledge such as details of the system or the underlying science. Such prior knowledge helps to limit the dimensionality of the optimization space by reducing the number of structures that agree with the experimental data within some tolerance. Methods have been devised to estimate interface properties, using this prior knowledge, that describe the data while adhering to a given set of constraints. Such methods include optimizers applying gradient projection (Byrd *et al.*, 1995), annealing processes (Xiang *et al.*, 1997) and evolutionary algorithms (Storn & Price, 1997).

Another approach to optimization is the use of sampling methods, of which two are discussed in this work, namely the Metropolis–Hastings Markov chain Monte Carlo (MCMC) method (Metropolis *et al.*, 1953; Hastings, 1970) and nested sampling (Skilling, 2004, 2006), both of which are Bayesian and sample the parameter posterior distribution. Due to the typically high dimensionality of the parameter space in reflectometry, Bayesian sampling methods tend to be computationally expensive and impractical for obtaining results, such as parameter estimates and covariances, in real time. We use *refnx* for MCMC sampling (Nelson & Prescott, 2019) and *dynesty* for nested sampling (Speagle, 2019) to sample our data and compare the results with those derived from the FI; *refnx* uses the *emcee* package (Foreman-Mackey *et al.*, 2012) to provide an implementation of an invariant MCMC ensemble sampler (Goodman & Weare, 2010).

Much work has been undertaken on quantifying the information content of a reflectivity data set, with most applying Bayesian statistics, where probability represents a degree of belief or plausibility based on the evidence at hand (Sivia & Skillings, 2012). One such approach looked at experimental optimization by determining the information gain from a given experiment using the entropies of the posterior and prior probability density functions (Treece *et al.*, 2019). Similarly,

work has been done on quantifying the information gain from scattering experiments as a function of the SLD of molecular components (Heinrich *et al.*, 2020). Many other Bayesian information-based approaches have been applied to reflectometry, including the use of Bayesian evidence to determine the set of free parameters that maximize model information density (McCluskey *et al.*, 2020), and using maximum entropy to reconstruct an SLD profile from a reflectivity curve (Sivia *et al.*, 1991).

Similarly to previous work, we propose a methodology for quantifying the information content of a reflectivity data set for use in determining the maximum information gain and experimental design optimization. However, we attempt to solve a slightly different problem to previous work, and the calculations that are made using our framework are different from those of Bayesian techniques. Although the goal of these estimation procedures remains the same, we derive the maximum information that the data set contains, given the current data point uncertainties, not the information content that can be readily extracted, for example, by sampling the posterior distribution.

For our application, the uncertainties on our reflectivity points are defined as the square root of the number of neutron counts and these counts are governed by Poisson statistics. Under these assumptions, we can analytically calculate the FI and apply the Cramér–Rao bound (Cramér, 1946; Rao, 1994). The bound states that the inverse of the FI provides a strict lower bound on the variance of an observable variable and, as a consequence, the FI provides us with a strict upper bound on the amount of information extractable from the observable. In practice, using this analytical derivation, we can achieve sub-second calculations of parameter uncertainties.

To evaluate the FI approach, we developed an experiment simulation framework based on the underlying assumptions of Poisson statistics for neutron counts. Since the FI is calculated using neutron counts, such a framework is necessary in calculating the FI in the general case where any model is given. Furthermore, this framework allows us to calculate the information content of any experimental conditions without costly beamtime to acquire the same data. The simulation framework is general in that it can simulate any beamline, given the incident flux profile of the instrument in question, and is shown to be accurate without requiring computationally expensive Monte Carlo methods.

## 2. Methods

### 2.1. The Fisher information matrix

The FI is a fundamental quantity in statistical theory and quantifies parametric changes in probability distributions in terms of information. It is related to various kinds of information ‘distance’, most notably the Kullback–Leibler (KL) divergence, a building block for many familiar information theoretic measures (Kullback, 1997). The KL divergence is the standard way of measuring a difference between distributions in terms of information and, when applied to parametric

distributions, provides the ‘distance’ between parameter values. As a rule of thumb, we can understand a one sigma (one standard deviation) difference in parameters as a KL divergence of one nat (natural unit of information). To a first approximation, this distance is calculated from the FI.

More formally, the FI matrix,  $\mathbf{g}$ , is the first nonzero term in the series expansion of the KL divergence from one (vector) parameter  $\xi$  to  $\xi + \Delta\xi$  [written here as  $D(\xi \parallel \xi + \Delta\xi)$ ]:

$$D(\xi \parallel \xi + \Delta\xi) = \frac{1}{2} \Delta\xi^T \mathbf{g} \Delta\xi + O(|\Delta\xi|^3). \quad (1)$$

Thus, we can think of the FI matrix as a way of scaling the parameters so that, for sufficiently small changes in parameters, the square Euclidean distance is the informational change. The FI is therefore relative to the specified parameters and measured in nats per parameter unit squared; it is local and not dimensionless.

The one nat per sigma relationship is exact for many widely used distributions, such as for a multivariate normal with constant covariance. In this case, the FI is the inverse of the associated covariance matrix. Correspondingly, a practical way this is used is to set an information threshold, *e.g.* one nat (one sigma), and find out by how much the parameters must change to reach this threshold, thereby specifying an acceptance and/or confidence region, as discussed in Section 2.4 and the supporting information.

### 2.2. Derivation

To derive the equations for information content quantification using the FI, we must first provide a structure for given reflectivity data of  $N$  points (equivalently, histogram bins). This structure consists of contiguous layers representing a physical sample, with each layer being defined by its thickness, SLD and interfacial roughness. A model is then described by this structure and given measurement background noise, experimental scale factor and instrument resolution; we need only vary the  $M$  unknown parameters of this model. Such a model describes the reflectance at a given neutron momentum transfer, for example, using the Abelès matrix formalism implemented in *refnx*. The reader is referred to the supporting information for the full derivation but, in summary, the FI matrix  $\mathbf{g}^\xi$  for the  $M$  parameters  $\xi$  of a model of  $N$  reflectivity points is given by

$$\mathbf{g}^\xi = \mathbf{J}^T \mathbf{M} \mathbf{J}, \quad (2)$$

where  $\mathbf{J}$  is the Jacobian of the reflectances  $r_i$  with respect to the parameters  $\xi$ , and  $\mathbf{M}$  is a diagonal matrix of incident counts  $s_i$  divided by model reflectances  $r_i$ . In the FI matrix, the FI for an individual parameter  $\xi_i$  corresponds to the diagonal element  $\mathbf{g}_{i,i}^\xi$ ,

$$\mathbf{J} = \begin{bmatrix} \partial r_1 / \partial \xi_1 & \partial r_1 / \partial \xi_2 & \cdots & \partial r_1 / \partial \xi_M \\ \partial r_2 / \partial \xi_1 & \partial r_2 / \partial \xi_2 & \cdots & \partial r_2 / \partial \xi_M \\ \vdots & \vdots & \ddots & \vdots \\ \partial r_N / \partial \xi_1 & \partial r_N / \partial \xi_2 & \cdots & \partial r_N / \partial \xi_M \end{bmatrix}, \quad (3)$$

$$\mathbf{M} = \begin{bmatrix} s_1/r_1 & 0 & \cdots & 0 \\ 0 & s_2/r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_N/r_N \end{bmatrix}. \quad (4)$$

Using equation (2), we can calculate the FI for a model describing a single data set. However, more complicated experiments often involve multiple data sets, such as measuring multiple experimental contrasts. For such cases,  $n$  models, potentially containing inter-model parameter constraints, are required as input. The calculation is the same as in equations (2), (3) and (4), except the parameters  $\xi$  are the union of all of the (potentially shared) parameters of the  $n$  models. Additionally, the Jacobian  $\mathbf{J}$  and matrix  $\mathbf{M}$  are calculated over the concatenated incident counts and model reflectances.

### 2.3. Experiment simulation

To simulate an experiment, we require both a model and knowledge of the flux of incident neutrons as a function of wavelength. In our case, this was taken on the OFFSPEC reflectometer (Dalglish *et al.*, 2011). We can multiply this incident neutron flux by a constant in order to change the experimental counting time and give us the number of incident neutrons for a simulated experiment; this approach was developed from the ideas presented by Mironov *et al.* (2021). To account for different measurement angles, we multiply this incident flux by a factor to compensate for different collimating slit openings. Since both of the slits that are used to define the beam footprint scale linearly with the angle, we scale the intensity as the square of the angle.

We calculate the momentum transfer  $Q$  for each wavelength  $\lambda$  in the file using the measurement angle  $\theta$  and the equation

$$Q = (4\pi \sin \theta) / \lambda. \quad (5)$$

By default, these  $Q$  values are assigned to geometrically spaced bins, with the number of bins being set to the desired number of points for the simulated data set. Following this, we calculate each bin’s centre,  $Q_i$ . Alternatively, a given set of  $Q$  bin centres can be used. The model reflectivity for each bin,  $r_i$ , is calculated using *refnx* and additive instrument background noise  $\alpha$  is added (optionally, this can be accounted for in the model reflectivity calculation). Next, this reflectivity is multiplied by the bin’s incident flux  $\mu_i$  to obtain the reflected flux, and then multiplied by the simulated measurement time  $\tau$  to get the reflected counts for the bin. We use the reflected counts as the mean rate parameter of a Poisson distribution, from which we obtain a random value giving us an appropriately randomized number of reflected counts  $N_i$ ,

$$N_i \sim \text{Poisson}[(r_i + \alpha)\mu_i\tau]. \quad (6)$$

The bin’s uncertainty in count space is then the square root of this value,  $(N_i)^{1/2}$ . To obtain the reflectivity and associated uncertainty, we simply divide the ‘noisy’ counts and uncertainty by the number of incident neutrons for the bin  $s_i$  (*i.e.* the product of the bin’s incident flux and measurement time,  $\mu_i\tau$ ).

The above-described process will generate a single reflectivity data set for a given model. For more complicated experiments involving multiple data sets,  $n$  models are required as input, and the above process is repeated for each model, yielding  $n$  simulated data sets.

#### 2.4. Applying the Fisher information

The FI can be applied to both experimentally measured and simulated data. As the FI measures the instrumental uncertainty relative to changes in model parameters, a parameterized model must be provided and parameter values input. These parameter values can be obtained as estimates based on data, or specified manually if the task is simply the verification of a particular structure.

We use *refnx* to define a model and to load the model's associated measured or simulated reflectivity data. The parameters of the model are then optimized using a fitting algorithm of choice, in our case differential evolution (Storn & Price, 1997), according to the  $\chi^2$  distance or, equivalently, the negative log-likelihood. The likelihood provides a measure of difference between a given data set and model and, for this work, is defined as its implementation in *refnx*,

$$-\ln \mathcal{L} = \frac{1}{2} \sum_{i=1}^N \left\{ \left( \frac{r_i - r_{i,m}}{\delta r_i} \right)^2 + \ln [2\pi(\delta r_i)^2] \right\}, \quad (7)$$

where  $N$  is the number of measured data points,  $r_i$  is the experimental reflectivity at the  $i$ th  $Q$  point,  $\delta r_i$  is the uncertainty in the experimental reflectivity at the  $i$ th  $Q$  point and  $r_{i,m}$  is the model reflectivity at the  $i$ th  $Q$  point calculated using the Abelès matrix formalism.

From here on, we no longer need the data, since the model and Poisson statistics describe the data sufficiently. Next, we calculate the Jacobian  $\mathbf{J}$ , whose entries are the gradient of the model reflectivity  $r_i$  with respect to each of the model parameters  $\xi_j$ . We estimate this using a finite difference approximation based on the reflectance for parameter values 0.5% either side of the input value. Using this and the diagonal matrix  $\mathbf{M}$  of incident counts divided by model reflectances, we can calculate the FI matrix using equation (2). Since the calculation of this matrix is relatively simple, implementation for use in other fitting software (Kienzle *et al.*, 2017; Hughes, 2017; Björck & Andersson, 2007) should be straightforward.

The FI matrix contains all of the information about parameter variances and covariances but these values require extraction. The variance of a single parameter is simply given by the inverse of the FI and so its uncertainty is given by the square root. In the general case of multiple parameters, the uncertainty  $\epsilon_i$  for a parameter  $\xi_i$  is obtained from the square root of the inverse of the diagonal elements of the FI matrix,

$$\epsilon_i = (1/\mathbf{g}_{i,i}^{\xi})^{1/2}. \quad (8)$$

Finally, to extract the covariance between any two parameters, we can calculate a confidence ellipse of given size  $k$  standard deviations (see the supporting information for details).

#### 2.5. Application to soft matter

To illustrate the utility of the FI, we applied our framework to an experiment measuring a common model system for structural biology: a 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) bilayer deposited onto a silicon surface. The lipids were measured against two water contrasts, H<sub>2</sub>O and D<sub>2</sub>O. The data were taken using the CRISP neutron reflectometer (Penfold *et al.*, 1987) as part of the ISIS neutron training course and simultaneously fitted using *RasCAL* (Hughes, 2017). This fitting was constrained against measured data for a bare Si/D<sub>2</sub>O interface including a native SiO<sub>2</sub> layer.

Our model for the bilayer was defined by two lipid leaflets with fixed surface coverage. The model was fitted by area per molecule rather than volume fractions, to avoid ambiguity arising from differing total molar quantities of headgroup and tailgroup components. The model also accounted for the headgroups and tailgroups containing water through defects across their surfaces, and for the water bound to the hydrophilic headgroups. After fitting the experimental data, we reparameterized the bilayer model as a function of contrast SLD and, using this new model, were able to simulate the DMPC bilayer experiment on the OFFSPEC reflectometer (using our instrument flux profile) with arbitrary contrast SLD. We then investigated the change in the FI for each model parameter with contrast SLD.

For our parameterization, we have assumed that the molecular volumes of the headgroups and tailgroups are known and constant, and that any changes in molecule surface area are inversely proportional to the headgroup and tailgroup thicknesses. Structural biology is a large field of research with varying values used for these molecular volumes. Furthermore, these molecular volumes may vary with measurement conditions and may not necessarily be constant in practice (Campbell *et al.*, 2018). However, so as to not overcomplicate our model we have fixed them. The full details of the bilayer model parameterization and fitting can be found in the supporting information.

### 3. Results and discussion

#### 3.1. Measured versus simulated data

To demonstrate the robustness of our experiment simulation, we compare a data set measured using the OFFSPEC neutron reflectometer with its simulated counterpart. The data were measured experimentally using angles of 0.3, 0.4, 0.5, 0.6, 0.7, 2.0 and 3.0° with measurement times of 7.5, 7.5, 7.5, 15, 15, 60 and 120 min, respectively. The data from these angles were stitched together to produce a single data file. To obtain a 'ground truth' model for simulation, we fitted these stitched data using *refnx* to get Table 1. The background, experimental scale factor and resolution used for fitting were  $8 \times 10^{-7}$ , 0.783 and 2.5%  $dQ/Q$ , respectively.

To facilitate a measurable difference in the noise characteristics of the experimentally measured data and the data generated by our simulation framework, we took 1.5 min time slices from the measured data associated with each individual



**Table 1**

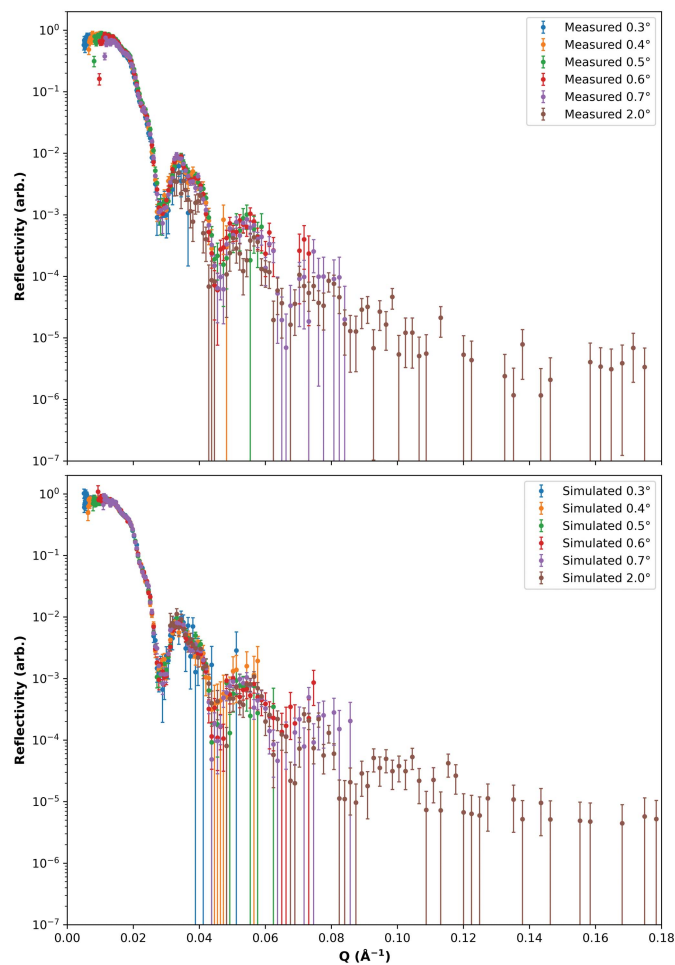
Fitted SLD, thickness and roughness values for each layer of the model corresponding to the measured data set.

	SLD ( $10^{-6} \text{ \AA}^{-2}$ )	Thickness ( $\text{\AA}$ )	Roughness ( $\text{\AA}$ )
Layer 1 (Si)	1.795	790.7	24.5
Layer 2 (Cu)	6.385	297.9	3.50
Substrate (quartz)	3.354	N/A	12.9

angle. For each measurement angle, we used the same angle, counting time and number of points for the simulation. As can be seen in Fig. 1, the noise characteristics of the time-sliced measured data and the simulated data are very similar. Statistically, comparing the data sets we find, using the Hotelling  $t^2$  test,  $p = 0.874$  and  $t^2 = 0.159$ , Anscombe transformed (Hotelling, 1931; Anscombe, 1948), implying no significant differences between the measured and simulated data.

### 3.2. Benchmarking

As mentioned above, the FI approach has notable performance upsides. To demonstrate this, we compared the time to



**Figure 1**

Experimentally measured reflectivity (top) and simulated reflectivity (bottom) versus momentum transfer  $Q$  for each measurement angle of the Table 1 sample.

**Table 2**

Calculation time of parameter uncertainties, in seconds, for MCMC sampling, nested sampling and the FI approach.

For each number of layers, ten samples were randomly generated using that number of layers, with the mean and standard deviation of the calculation time recorded for each approach.

No. of layers	No. of parameters	Calculation time (s)					
		MCMC sampling		Nested sampling		FI approach	
		Mean	SD	Mean	SD	Mean	SD
1	3	197.829	3.344	53.310	8.947	0.015	0.005
2	6	229.641	5.032	155.480	35.920	0.024	0.004
3	9	262.568	5.334	363.318	120.075	0.036	0.004
4	12	292.382	3.244	19680.574	124.743	0.047	0.004
5	15	330.579	9.531	2967.707	561.529	0.060	0.004
6	18	372.116	5.667	3862.186	700.430	0.076	0.005

obtain parameter uncertainties using the new approach and using established Bayesian methods, given a correct and fitted model. Fitting times have been excluded from these results since the time to fit would dominate the computation time of the FI approach, and thus they would provide little insight into the computational advantage of the FI calculation. Additionally, since fitting is typically required for MCMC sampling but not for nested sampling, a fair comparison between all methods becomes difficult. We therefore focus on the time taken for each method to reach completion, given optimal starting values.

The benchmark was run on a CPU with no methods having multiprocessing explicitly enabled. MCMC sampling was run with a 400 step burn-in period followed by a 30 step sample, with each sample being separated by 100 iterations. Nested sampling was run using the default *dynesty* stopping criteria which are optimized for evidence estimation (Speagle, 2019). Uniform priors were used with a 25% bound above and below the ground truth for each parameter. Following this, we ran our FI approach on the same samples. Table 2 compares the mean processing times of ten samples for each number of layers and, as can be clearly seen, the FI approach is significantly faster. Note that our implementation is not particularly optimized and we believe further performance gains could be obtained if they were required.

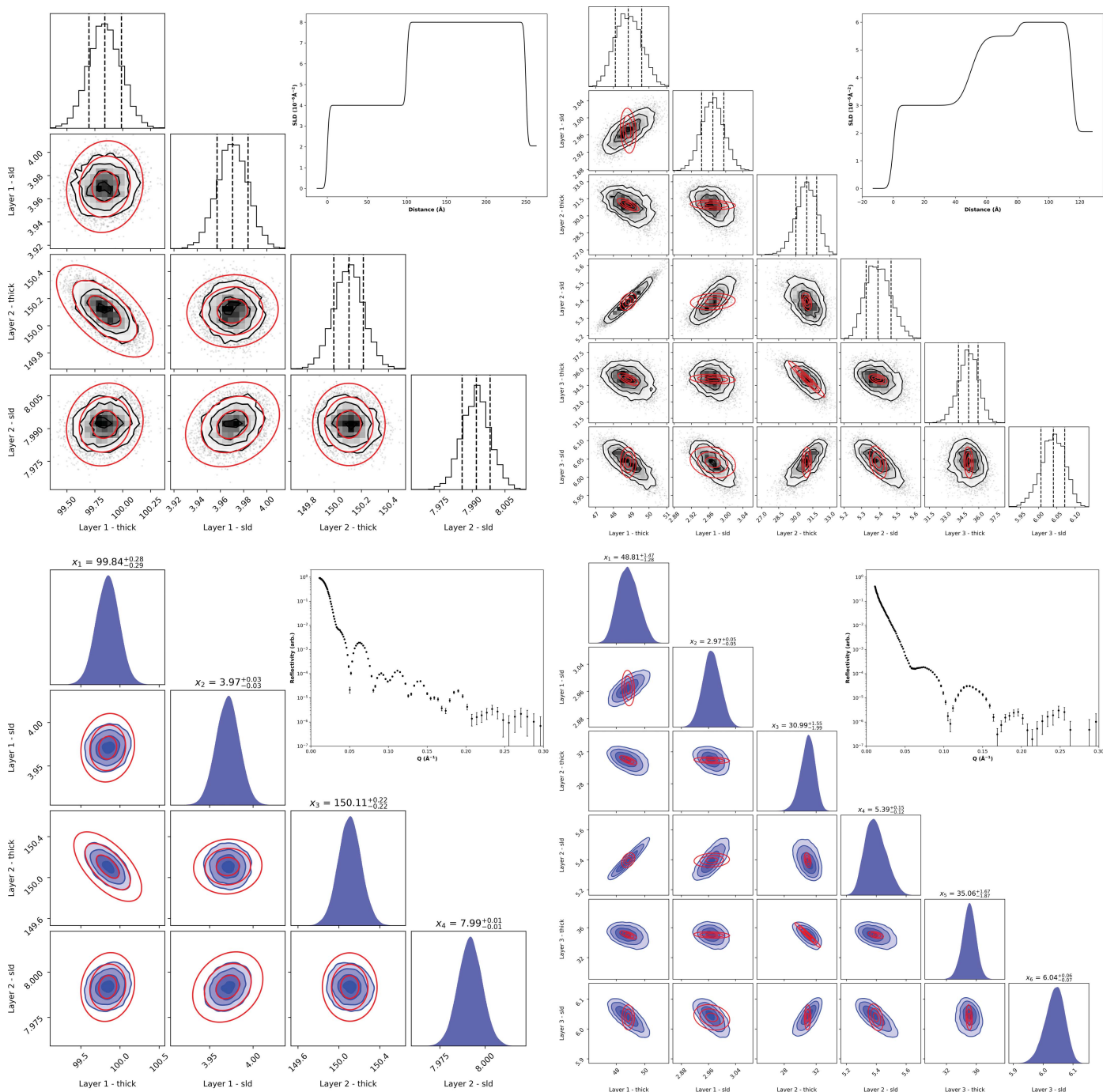
For each number of layers in the interval  $[1, 6]$ , we randomly generated ten samples and varied the SLD, thickness and interfacial roughness of each layer in each sample. Each sample used a silicon substrate of  $\text{SLD } 2.047 \times 10^{-6} \text{ \AA}^{-2}$ , and the random SLD, thickness and roughness of each layer were sampled from uniform distributions of intervals  $[-1, 10] \times 10^{-6} \text{ \AA}^{-2}$ ,  $[20, 1000] \text{ \AA}$  and  $[2, 8] \text{ \AA}$ , respectively. Using our experiment simulation, we synthesized data for each of these samples and ran both MCMC and nested sampling to obtain parameter uncertainties. Each experiment simulation consisted of 140 points obtained from two angles, 0.7 and 2.0°, using simulated measurement times of 7.5 and 30 min, respectively. Background noise of  $10^{-6}$ , instrument resolution of 2%  $dQ/Q$  and an experimental scale factor of 1.0 were used.

### 3.3. Corner plots and confidence ellipses

In *refnx* and *dynesty*, the results of MCMC and nested sampling, respectively, can provide a corner plot which is ‘an illustrative representation of different projections of samples in high dimensional spaces’ (Foreman-Mackey, 2016). These Bayesian sampling methods sample the parameter posterior distribution, allowing contours to be drawn through samples that are equally probable. The FI, however, is developed from

a frequentist view and the confidence ellipses bound regions where we have at least a  $k\sigma$  confidence in the value. Despite these fundamental differences, the sampling corner plots do still often agree very closely with the FI confidence ellipses.

For samples with mostly uncorrelated parameters, we found that corner plots show strong agreement with confidence ellipses. However, when more parameter correlation is present, the sampling uncertainties are much larger and we



**Figure 2** The FI confidence ellipses for  $k = 1, 2, 3$  (red) overlaid on the corner plots of MCMC (black) and nested sampling (blue) for the mostly uncorrelated parameter sample (left) and correlated parameter sample (right). Insets are the SLD profiles (top) and rebinned simulated reflectivity curves (bottom) of the two samples.

**Table 3**

Fitting biases in layer SLDs and thicknesses for the uncorrelated and correlated parameter samples of Fig. 2.

Sample	Fitting bias					
	SLD ( $10^{-6} \text{ \AA}^{-2}$ )			Thickness ( $\text{\AA}$ )		
	Layer 1	Layer 2	Layer 3	Layer 1	Layer 2	Layer 3
Uncorrelated	-0.003	-0.002	N/A	-0.011	0.010	N/A
Correlated	-0.012	-0.054	0.012	-0.556	0.286	0.264

reach a point at which the FI still represents the maximum obtainable information, but this seemingly cannot be extracted from the experimental data. As a consequence, the confidence ellipses do not match the corner plots as closely. This discrepancy is shown in Fig. 2, which compares two samples: a simple sample with mostly uncorrelated parameters and a more complicated sample with more parameter correlation due to similar layer SLDs. The data sets of Fig. 2 were both simulated with the same run condition as detailed in Section 3.2.

One potential source of deviation between corner plots and confidence ellipses may come from our fitting algorithm of choice. For our application of the FI, our estimator is a fitting algorithm and, so far, we have assumed that this estimator is unbiased. Thus, the Cramér–Rao bound implies that the inverse of the FI is a lower bound on the variance of this estimator. However, in practice, we found that our fitting algorithm of choice, differential evolution in *refnx*, may exhibit bias in some cases. To measure this bias we simulated 1000 experiments, using the same simulation conditions as used previously, for a number of different samples of varying complexity, and calculated the difference between the ground truth and mean fitted parameter values. Table 3 shows the fitting biases in the parameters of the Fig. 2 samples. As can readily be seen, the fitting bias is greater in the sample with larger inter-parameter correlations, particularly in the layer thicknesses; these biases are model dependent and potentially fitting-package dependent.

The Cramér–Rao bound may be modified for a biased estimator. However, for a real measurement, there is no way to tell if such a bias exists. As such, we leave our approach with the stricter limit (since any bias always increases the variance), and remind ourselves that the maximum possible information contained in the data is not always going to be the maximum extractable.

### 3.4. Time dependence

One potential use of the FI in reflectometry is enabling early stopping of experiments based on counting statistics. To determine the feasibility of this application, and to validate our implementation, we investigated how parameter uncertainties change with measurement time. As derived in the supporting information, we should expect the uncertainty of a parameter  $\epsilon$  to be inversely proportional to the square root of the experiment measurement time  $\tau$ . By using the fact that  $\epsilon \propto$

$1/\tau^{1/2}$ , introducing a nonzero proportionality constant  $\alpha$  and taking the natural logarithm of both sides we see

$$\begin{aligned} \ln \epsilon &= \ln(\alpha/\tau^{1/2}) = \ln \alpha - \ln(\tau^{1/2}) \\ &= -\frac{1}{2} \ln \tau + \ln \alpha. \end{aligned} \tag{9}$$

Using this result, we should expect the gradient of the plot of log parameter uncertainty  $\ln \epsilon$  versus log time  $\ln \tau$  to be  $-\frac{1}{2}$ . To confirm this is the case, we compared established fitting uncertainty measures and uncertainties derived from the FI with increasing time using our experiment simulation framework; the fitting uncertainties were calculated using differential evolution in *refnx* and the FI uncertainties using equation (8).

Using simple linear regression, we found that the time dependence for any parameter’s uncertainty was indeed determined by the square root of the measurement time, as shown in Fig. 3. We used the same samples and simulation parameters as used for Fig. 2 except for the simulated measurement time. Both samples were initially simulated using the same times as before and then these times were multiplied by an increasing ‘time factor’ from 1 to 1000. This essentially split a fixed time budget between the simulated angles  $0.7$  and  $2.0^\circ$  with a ratio of 1:4.

For the simple sample, the relationship is perfectly exhibited. However, for the more complicated sample, the results are slightly noisier due to the increased difficulty of fitting. This is particularly noticeable when the counting time is low and the data being fitted are impacted by our added noise to a greater degree. With low counting statistics, differential evolution may terminate in a minimum of the  $\chi^2$  parameter space that does not represent the ground truth model (*i.e.* the simulated data no longer uniquely describe the true parameter set), resulting in uncertainties that deviate from the time-dependence relationship previously derived. This difficulty in fitting is shown in Fig. 3, where the mean absolute error between the ground truth and fitted parameter values is plotted against time. As can be seen, the fitting errors at lower counting times are approximately an order of magnitude larger than those at higher counting times.

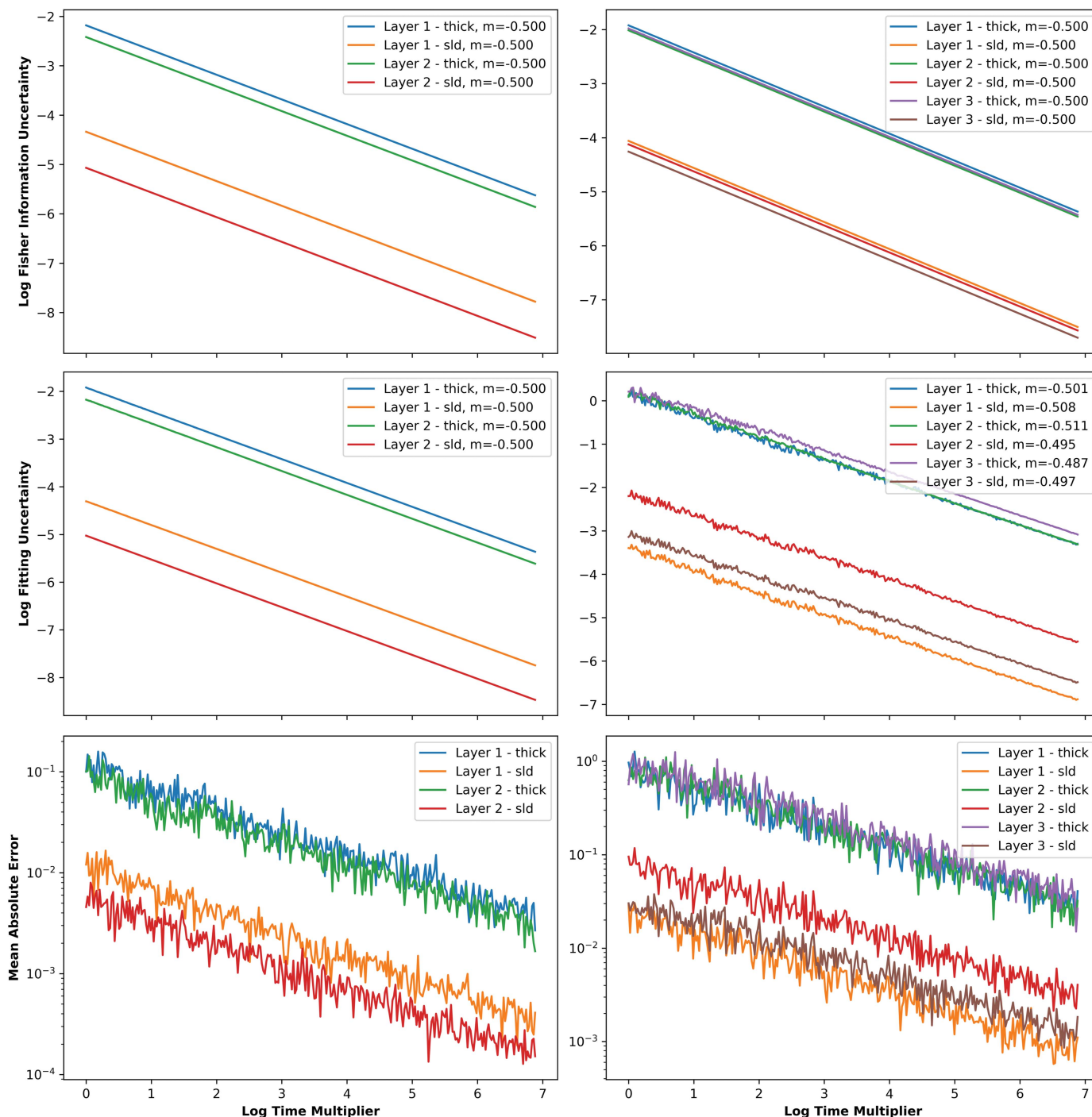
Since we now know that parameter uncertainties decrease as the square root of the measurement time, we are easily able to project the evolution of these uncertainties and can predict when some desired threshold will be reached, at which time we may want to cease the measurement. It is for the experimenter to decide on such a threshold, with the choice probably weighing up factors including time to change angle, time to change sample, total time budget and number of samples being measured. Such choices are not necessarily easy to make prior to starting an experiment and so automating this process may warrant further investigation.

### 3.5. Application to soft matter

As detailed in Section 2.5, we applied our framework to a soft matter experiment by taking experimentally measured data, fitting a DMPC bilayer model and reparameterizing the model as a function of bulk water contrast SLD. The fitted

SLD profiles and experimental reflectivity data are shown in Fig. 4. Using the model, data were simulated for each contrast SLD from  $-0.56 \times 10^{-6}$  to  $6.35 \times 10^{-6} \text{ \AA}^{-2}$  (pure H<sub>2</sub>O to pure D<sub>2</sub>O) and the FI calculated for each model parameter, obtained from the diagonal elements of the FI matrix of equation (2). These results are shown in Fig. 4 for an initial contrast choice and for a second contrast choice, assuming

D<sub>2</sub>O was measured first. For the simulation, angles of 0.7 and 2.0° and times of 15 and 60 min, respectively, were used (typical measurement times). Shown also are the nested sampling corner plots from sampling simulated data of solely D<sub>2</sub>O and of D<sub>2</sub>O and H<sub>2</sub>O contrasts using the reparameterized model. Included in these corner plots are the sampling uncertainties associated with each parameter.



**Figure 3** Log FI uncertainty (top), log fitting uncertainty (middle) and mean absolute error (bottom) versus log measurement time multiplier for each parameter of the mostly uncorrelated parameter sample (left) and correlated parameter sample (right). The uncertainties are taken as the mean from ten simulated experiments for a given time multiplier. Included in the legends of the uncertainty time-dependence plots are approximations of the gradients of the lines  $m$ , as given by linear regression.

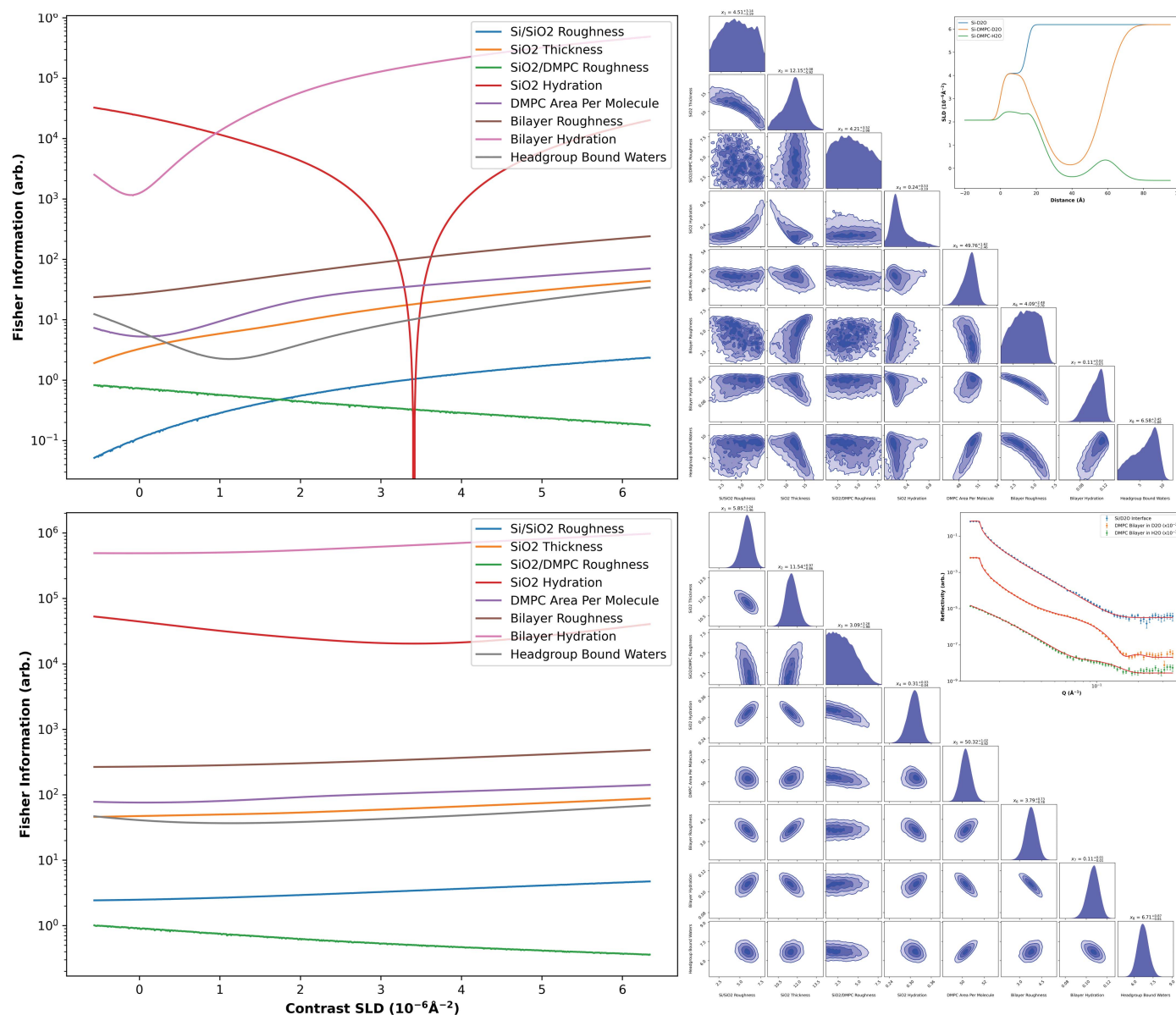


Since the units of the FI are nats per parameter unit squared, it is not technically correct to compare the FI directly between parameters. However, we can still compare the information content of parameters of the same unit. As might naïvely be expected, we show that it is possible to extract more information about some parameters than others. This result is certainly no surprise for researchers in the field who have experience fitting this system, but it does allow us to quantify it.

We show that the information of a parameter as a function of contrast is non-monotonic, almost certainly due to hydration of various components, leading to them becoming indistinguishable from neighbouring components for some bulk water deuterium concentrations. This is particularly noticeable

with the SiO<sub>2</sub> hydration parameter for the initial contrast choice, where the large drop in information is due to the matching of contrasts. Since most of the other model parameters describe multiple interfaces, with only one interface being able to become ‘invisible’ through contrast matching at a time, there are no zeroes in the FI for these plots.

Fig. 4 indicates that the most information is almost always obtainable from the highest SLD water, D<sub>2</sub>O. For the initial measurement contrast, the difference in information between the H<sub>2</sub>O and D<sub>2</sub>O extremes is significant. However, when considering the second measurement contrast, the information gain between the two contrast SLDs is less. It is well established that measuring multiple contrasts of different SLDs will



**Figure 4** FI versus bulk water contrast SLD for each model parameter of the reparameterised DMPC bilayer model, for an initial contrast choice (top left) and second contrast choice (bottom left), assuming D<sub>2</sub>O was measured first. Also shown are the nested sampling corner plots from sampling simulated data of solely D<sub>2</sub>O (top right), and D<sub>2</sub>O and H<sub>2</sub>O contrasts (bottom right). Insets are the experimentally fitted SLD profiles (top) and reflectivity curves (bottom) for the Si/D<sub>2</sub>O interface and the two solution isotopic contrast data sets, H<sub>2</sub>O and D<sub>2</sub>O, offset by factors of 10<sup>-2</sup> and 10<sup>-3</sup>, respectively, for clarity.

reduce parameter uncertainties and so this result may seem unusual; it could suggest that measuring D<sub>2</sub>O for twice as long would reduce the parameter uncertainties so that they are similar to those when measuring D<sub>2</sub>O and H<sub>2</sub>O. However, we believe the aim of measuring multiple contrasts is not only to lower parameter uncertainties but also to decouple parameter correlations (*i.e.* lowering inter-parameter covariances). Therefore, only considering which contrasts maximize the FI will not account for this covariance reduction. This is illustrated in the corner plots of Fig. 4, where the estimated posterior distributions from the D<sub>2</sub>O and H<sub>2</sub>O data are clearly much better defined (*i.e.* more Gaussian) than those from just D<sub>2</sub>O. For example, the model roughness parameters are very poorly defined with just a single D<sub>2</sub>O measurement.

While difficult to display due to the number of parameter pairs, if one has a model, it would be possible to calculate the optimal contrast to measure in order to minimize both parameter variances and inter-parameter covariances. The optimal solution is almost certainly model dependent, but given the broad features found here, having a slightly incorrect model is unlikely to be an issue.

### 3.6. Limitations

Our framework is essentially frequentist and, much like one does in hypothesis testing, it proceeds by calculating probabilities based on hypothetical, assumed or estimated parameter values. The sizes of uncertainties, for example, are those that would exist if the estimated parameter values were correct. Determining uncertainties in this way may appear to be an issue, particularly in reflectometry, since the determination of the globally 'correct' model is non-trivial. As ever, it is the choice of the experimenter to decide whether their model, guided by their underlying knowledge of the system, will accurately represent the true system. However, even if the model is not exactly correct, the FI still provides value. Since our calculations effectively perform a sensitivity analysis of the parameters (more sensitive means a larger gradient in the Jacobian **J**, and therefore a larger FI), similar models with differing values are still very likely to have the same behaviour and give the same trends. For example, for our DMPC bilayer data, there are many more model parameterizations that have been argued in the literature, and differing values for the parameters we have chosen to fix. However, our simplifying assumptions describe the measured data to a satisfactory level and we are confident that equivalent parameterizations would not change the trends found from our calculations.

Usually, to optimize the parameter values, the model is fitted before calculating the FI; in our case, we have used differential evolution. Fitting algorithms can often provide estimates of parameter variances and covariances and in some cases these values may be very similar to those provided by our framework. However, the fit-derived (co)variances have no deeper statistical underpinning and do not allow further investigation of the system. Our framework, being based on the counting statistics and parameter sensitivity of the model, enables almost instant calculation of uncertainties measured

at any point in the future, or for different contrasts/conditions. Determining the (co)variances from a fit several hundreds of times to create a 'phase diagram', similar to those described in Section 3.5, is not feasible in anything close to real time, as would be useful during an experiment.

### 4. Future work

The presented framework has many potential applications in neutron reflectometry and in other scattering techniques based on counting statistics. As demonstrated by our soft-matter application, experimental design is one such use where the FI could be used to influence real-time decisions regarding measurement angle and/or contrast choice; a similar Bayesian approach would be unfeasible for real-time application due to computational overhead. We would also like to apply our framework to more complex real-world systems, such as magnetic structures. This could provide answers to common questions posed in the literature and give insight as to why particular experimental design choices have found popularity.

The FI framework has the possibility of being extended to quantify additional factors of an experiment, such as the time to change the sample or angle. Additionally, work on quantifying and incorporating fitting biases and inter-parameter correlations into the FI calculation and subsequent analysis could bridge the gap between our framework and established methods. Since the FI uncertainties do not always match those obtained from established methods, it could also be possible to provide experimenters with a metric detailing how closely the FI results would be expected to match established methods. Since the largest variations were found to occur when parameters are strongly correlated, the Pearson correlation coefficient applied to the FI matrix (or similar) could be indicative here.

### 5. Conclusions

In this work, we have presented a framework for determining the maximum information gain and experimental design optimization of neutron reflectometry experiments using the Fisher information. We have demonstrated how the FI allows us to quantify the information content of a measured data point in relation to given model parameters. To illustrate this point, we have developed a robust framework for simulating experimental data with realistic noise characteristics, and then compared the FI-derived results with Bayesian sampling methods. The FI describes the maximum possible extractable information and therefore can be significantly different from sampling methods. However, this approach has significant upsides in its run time and its ability to project uncertainties, as well as the ability to run experiments *in silico*. Finally, we have demonstrated a practical application of the approach in determining the information content of the parameters of a DMPC bilayer sample parameterized as a function of the bulk water contrast, allowing us to ascertain optimal measurement conditions.

The code for this work is open source and freely available on GitHub (Durant *et al.*, 2021).

## Acknowledgements

The authors thank Luke Clifton for his assistance and expertise in fitting the DMPC data.

## Funding information

This work was partially supported by the STFC Facilities Programme Fund through the ISIS Neutron and Muon Source, by the Scientific Computing Department of Rutherford Appleton Laboratory, Science and Technology Facilities Council, and by Wave 1 of the UKRI Strategic Priorities Fund under the EPSRC grant EP/T001569/1, particularly the ‘AI for Science’ theme within that grant and The Alan Turing Institute.

## References

Abelès, F. (1948). *Ann. Phys.* **12**, 504–520.  
 Anscombe, F. J. (1948). *Biometrika*, **35**, 246–254.  
 Barndorff-Nielsen, O. E. & Gill, R. D. (2000). *J. Phys. A Math. Gen.* **33**, 4481–4490.  
 Barnes, L. P., Han, Y. & Özgür, A. (2019). *IEEE International Symposium on Information Theory*, 7–12 July 2019, Paris, France, pp. 2704–2708. New York: Institute of Electrical and Electronics Engineers.  
 Björck, M. & Andersson, G. (2007). *J. Appl. Cryst.* **40**, 1174–1178.  
 Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. (1995). *SIAM J. Sci. Comput.* **16**, 1190–1208.  
 Campbell, R. A., Saaka, Y., Shao, Y., Gerelli, Y., Cubitt, R., Nazaruk, E., Matyszevska, E. & Lawrence, M. J. (2018). *J. Colloid Interface Sci.* **531**, 98–108.  
 Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.  
 Dalglish, R. M., Langridge, S., Plomp, J., de Haan, V. O. & van Well, A. A. (2011). *Physica B*, **406**, 2346–2349.  
 Durant, J. H., Wilkins, L., Butler, K. & Cooper, J. F. K. (2021). *fisher-information*, <https://github.com/James-Durant/fisher-information>.  
 Fisher, R. A. (1925). *Math. Proc. Camb. Phil. Soc.* **22**, 700–725.  
 Foreman-Mackey, D. (2016). *J. Open Source Software*, **1**, 24.  
 Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. (2013). *Publ. Astron. Soc. Pac.* **125**, 306–312.

Goodman, J. & Weare, J. (2010). *Commun. Appl. Math. Comput. Sci.* **5**, 65–80.  
 Hastings, W. K. (1970). *Biometrika*, **57**, 97–109.  
 Heinrich, F., Kienzle, P. A., Hoogerheide, D. P. & Lösche, M. (2020). *J. Appl. Cryst.* **53**, 800–810.  
 Hotelling, H. (1931). *Ann. Math. Stat.* **2**, 360–378.  
 Hughes, A. V. (2017). *RasCAL*, <https://sourceforge.net/projects/rscl>.  
 Kienzle, P. A., Maranville, B. B., O’Donovan, K. V., Ankner, J. F., Berk, N. K. & Majkrzak, C. F. (2017). *NCNR Reflectometry Software*, <https://www.nist.gov/ncnr/data-reduction-analysis/reflectometry-software>.  
 Kullback, S. (1997). *Information Theory and Statistics*. Mineola: Dover.  
 Majkrzak, C. F. & Berk, N. F. (1995). *Phys. Rev. B*, **52**, 10827–10830.  
 McCluskey, A. R., Cooper, J. F. K., Arnold, T. & Snow, T. (2020). *Mach. Learn. Sci. Technol.* **1**, 035002.  
 Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). *J. Chem. Phys.* **21**, 1087–1092.  
 Mironov, D., Durant, J. H., Mackenzie, R. & Cooper, J. F. K. (2021). *Mach. Learn. Sci. Technol.* **2**, 035006.  
 Nelson, A. R. J. & Prescott, S. W. (2019). *J. Appl. Cryst.* **52**, 193–200.  
 Parratt, L. G. (1954). *Phys. Rev.* **95**, 359–369.  
 Penfold, J., Ward, R. C. & Williams, W. G. (1987). *J. Phys. E Sci. Instrum.* **20**, 1411–1417.  
 Petz, D. (2002). *J. Phys. A Math. Gen.* **35**, 929–939.  
 Rao, C. R. (1994). *Selected Papers of C. R. Rao*. Chichester: Wiley.  
 Sivia, D. S. (2013). *Elementary Scattering Theory*. Oxford University Press.  
 Sivia, D. S., Hamilton, W. A. & Smith, G. S. (1991). *Physica B*, **173**, 121–138.  
 Sivia, D. S. & Skilling, J. H. (2012). *Data Analysis: A Bayesian Tutorial*. Oxford University Press.  
 Skilling, J. (2004). *AIP Conf. Proc.* **735**, 395–405.  
 Skilling, J. (2006). *Bayesian Analysis*, **1**, 833–860.  
 Speagle, J. S. (2020). *Mon. Not. R. Astron. Soc.* **493**, 3132–3158.  
 Storn, R. & Price, K. (1997). *J. Glob. Optim.* **11**, 341–359.  
 Taylor, S. (2019). *Entropy*, **21**, 110.  
 Telesca, L., Lovallo, M., Ramirez-Rojas, A. & Angulo-Brown, F. (2009). *Physica A*, **388**, 2036–2040.  
 Treece, B. W., Kienzle, P. A., Hoogerheide, D. P., Majkrzak, C. F., Lösche, M. & Heinrich, F. (2019). *J. Appl. Cryst.* **52**, 47–59.  
 Wang, L. Y. & Yin, G. G. (2010). *IEEE Trans. Autom. Contr.* **55**, 674–690.  
 Xiang, Y., Sun, D. Y., Fan, W. & Gong, X. G. (1997). *Phys. Lett. A*, **233**, 216–220.