



# Simple Bayesian method for improved analysis of quasi-two-dimensional scattering data

Alexander T. Holmes\*

*School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, United Kingdom*

(Received 3 April 2014; revised manuscript received 22 April 2014; published 16 July 2014)

A new method is presented for the analysis of small angle neutron scattering data from quasi-two-dimensional systems such as flux lattices, Skyrmion lattices, and aligned liquid crystals. A significant increase in signal to noise ratio and a natural application of the Lorentz factor can be achieved by taking advantage of the knowledge that all relevant scattering is centered on a plane in reciprocal space. The Bayesian form ensures that missing information is treated in a controlled way and can be subsequently included in the analysis. A simple algorithm based on Gaussian probability assumptions is provided which provides very satisfactory results. Finally, it is argued that a generalized model-independent Bayesian data analysis method would be highly advantageous for the processing of neutron and x-ray scattering data.

DOI: [10.1103/PhysRevB.90.024514](https://doi.org/10.1103/PhysRevB.90.024514)

PACS number(s): 61.05.fd, 61.05.fg

## I. INTRODUCTION

The study of flux line lattices (FLLs) in type-II superconductors is a major use of the small angle neutron scattering (SANS) technique [1–3]. In their so-called “mixed state” these materials contain a spatially periodic magnetic field pattern. Due to their wave nature and intrinsic magnetic moment, neutrons diffract from these magnetic structures in a manner which make them a powerful tool for the investigation of FLLs.

The mixed state occurs when type-II superconductors are subject to an external magnetic field  $H$  above their lower critical field  $H_{c1}$ . In this phase, the field penetrates the bulk superconductor, creating regions of suppressed order parameter, surrounded by screening supercurrents. Coherence of the superconducting pair wave-function means that each of these regions, known as vortices or flux lines, corresponds to a single flux quantum,  $\Phi_0 = h/2e$ . Repulsion between flux lines causes a two-dimensional (2D) lattice to form, which would be close-packed hexagonal in a perfectly isotropic system [4]. In real materials, however, the nature of the superconducting pairing and anisotropies in the Fermi surface can cause distortions and even phase transitions in the flux lattice. For example, in single crystal niobium, temperature and field-dependent transitions are found in the FLL [5], which even reflect topological considerations due to the so-called “hairy-ball theorem” [6].

SANS can give information about the structure of the FLL, and also about the field contrast and lattice perfection. The contrast is related to the London penetration depth, and hence superfluid density. The way this varies as the temperature approaches zero can tell us about the presence of nodes in the superconducting gap, indicating unconventional pairing symmetry. The lattice perfection is affected by pinning of the flux lines and by thermal fluctuations.

In a uniform field in most materials the FLL is essentially 2D in nature, consisting of parallel rods aligned with the magnetic field. The 2D array of rods gives a single plane of spots in reciprocal space, at positions  $\mathbf{q}$  equal to the reciprocal lattice points  $\mathbf{G}_{hk}$ . The width of the spots perpendicular to the reciprocal lattice plane is given by the correlation length along the

field direction in real space, and their integrated intensity is determined by the Fourier component of the field contrast  $B(\mathbf{q})$ .

Figure 1(a) shows the experimental geometry; Figs. 1(b) and 1(c) show the reciprocal lattice and Ewald sphere relevant to the FLL problem. Diffraction will only occur when the momentum transfer between the neutron and the lattice  $\hbar\mathbf{q} = \hbar(\mathbf{k}_f - \mathbf{k}_i)$  is equal to a point in the reciprocal lattice  $\mathbf{G}_{hk}$ , where  $\mathbf{k}_i$  and  $\mathbf{k}_f$  are, respectively, the incoming and outgoing neutron wave vectors. For elastic scattering (where  $|\mathbf{k}_f| = |\mathbf{k}_i|$ ) the Ewald sphere construction is often used to illustrate this, and in the case of the 2D FLL, it implies that the only diffraction signal to occur will be on the locus of points where the sphere determined by  $|\mathbf{k}_i|$  intersects the reciprocal lattice plane containing  $\mathbf{G}_{hk}$ . This is a circle passing through the origin at  $\mathbf{q} = 0$ , with a radius depending on  $|\mathbf{k}_i|$  and the angle between the lattice plane and the beam direction. The points where this circle passes through a reciprocal lattice point  $\mathbf{G}_{hk}$  correspond to the Bragg condition for a given lattice point. However, the present method makes no assumptions about the form of the lattice and works for any form of 2D scattering, for example, from an aligned nematic liquid crystal [7]. It would also work in non-2D cases where other points in reciprocal space are far enough away not to cut the Ewald sphere.

The effect of the finite correlation length along the field direction is to thicken the lattice plane, causing the spots to have a finite width in the direction normal to the plane. The wavelength spread, beam size, and divergence also serve to broaden the final diffracted beam, as does the mosaic spread of the sample [8]. If the correlation length is the dominant source of broadening then the spreading will have a Lorentzian form in reciprocal space; however if instrumental resolution is the main factor, it is likely to be more Gaussian. For this example we will assume a Lorentzian shape with full width half maximum (FWHM)  $\Gamma$ , but a different function is straightforward to implement.

During a diffraction experiment, the angle between the incoming neutron beam and the magnetic field is varied by rotating the entire sample environment, including the sample and the magnet, about an axis passing through the sample. This has the effect of bringing different reciprocal lattice spots in and out of the Bragg condition, as the Ewald sphere passes through them. In general the field and sample can be rotated about two perpendicular axes, by angles referred to as  $\omega$  and  $\phi$ ,

\*a.t.holmes@bham.ac.uk

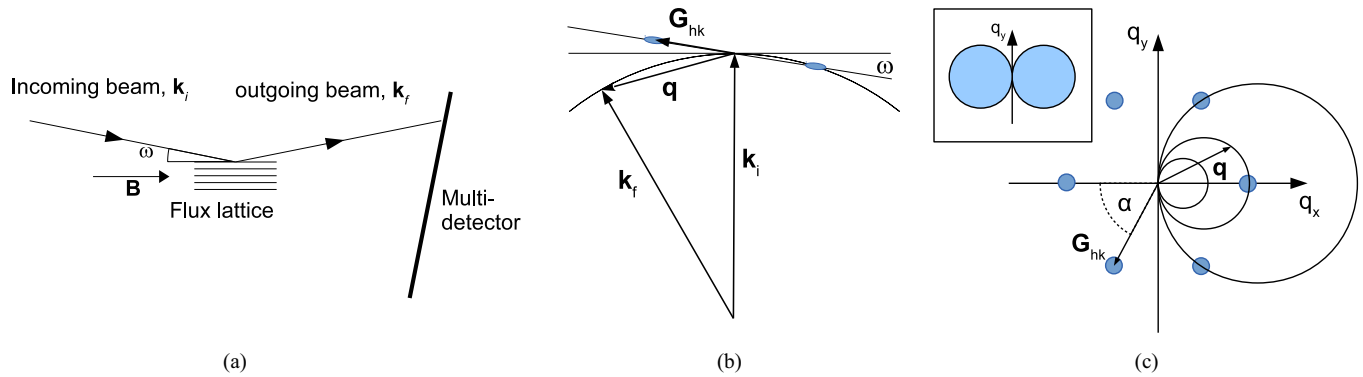


FIG. 1. (Color online) (a) Schematic diagram of a typical experimental setup. The incoming neutrons have a wave vector  $\mathbf{k}_i$ , outgoing neutrons have wave vector  $\mathbf{k}_f$ , where  $|\mathbf{k}_f| = |\mathbf{k}_i| = 2\pi/\lambda$ , and  $\lambda$  is the neutron de Broglie wavelength. Each pixel in the multidetector corresponds to a different direction of  $\mathbf{k}_f$ . (b) Section through the Ewald sphere and reciprocal lattice in the scattering plane. The field and hence reciprocal lattice is rotated by an angle,  $\omega$ , with respect to  $\mathbf{k}_i$ . (c) Reciprocal lattice plane showing circles where the Ewald sphere intersects for different rotations  $\omega$  around  $\mathbf{q}_y$ . If all scattered intensity is in this plane, then only pixels close to a circle will contain relevant data. The inset in panel (c) shows the region of reciprocal space for which relevant data is collected when the field is “rocked” from  $-\omega$  to  $+\omega$  about a single axis in the  $y$  direction.

to access different regions of reciprocal space. We will mostly refer to rotation about a single axis for clarity, but this method works for any combination of  $\omega$  and  $\phi$ . Strictly speaking we should specify an order in which the rotations occur, as they do not commute; however if the angles are small this will not be important. A 2D detector placed several meters away records the diffracted intensity at small angles ( $\sim 1^\circ$ ) to the beam axis. To produce an entire diffraction pattern, a series of measurements are taken at different rotation angles.

## II. DATA ANALYSIS

The traditional method of producing a diffraction pattern from such an experiment is to sum the counts on each pixel in the multidetector over frames (i.e., individual exposures at given rotations) taken over a range of angles which encompass the Bragg condition for all relevant diffraction spots. Backgrounds are subtracted using measurements taken either in zero field or above the superconducting transition temperature. This is necessary as the diffracted intensity from the FLL is relatively weak and superimposed on a  $|q|^{-4}$  background. The  $\mathbf{q} = 0$  point is determined by a measurement of the undiffracted “direct” beam, which is also used to determine the absolute diffracted intensity in subsequent processing. The counts are also usually normalized to reflect the integration time, or total neutron flux, giving a result in “counts per standard monitor” which accounts for different measuring times.

While simple to understand and implement, the simple “rocking sum” described has a number of disadvantages. One is that the rate at which the Ewald sphere will pass through a spot depends on the angle between the relevant reciprocal lattice vector and the rotation axis. As shown in Figs. 1(b) and 1(c), the reciprocal lattice plane is initially tangent to the Ewald sphere at  $\omega = 0$ . Then the plane is rotated by  $\omega$  about an axis, also within the plane, passing through the point where they touched. The intersection between the sphere and the plane is a circle, fixed at one point on its circumference, expanding in radius in a direction perpendicular to the axis

of rotation. The circle will take the smallest range of  $\omega$  to pass through a finite sized spot in this direction and will cross the other spots over an angular range larger by a factor of  $1/|\cos(\alpha)|$ , where  $\alpha$  is the angle between the  $\mathbf{q}$  vector of the spot and the perpendicular to the rotation axis. This is known as the Lorentz factor and means that spots closer to the rotation axis will have a correspondingly higher integrated intensity. One must therefore divide this by the Lorentz factor to calculate the total integrated intensity for individual spots in  $\mathbf{q}$  space.

The other problem is that an unweighted sum of detector counts will include many angles for which any individual pixel is not at or close to the Bragg condition. In that case noise will be accumulated, but no signal. This is an especially big problem when the quality of the lattice is very good, with a long correlation length, as each spot will only appear in a narrow range of angles (with a FWHM known as its “rocking curve width”).

One way of handling the noise issue described above would be to do a weighted sum of counts, instead of equally weighting the counts at each rocking angle. The intensity in a given spot will have a Lorentzian shape as a function of rocking angle, with a peak centered at the relevant Bragg angle, if the intrinsic longitudinal broadening of the flux lattice diffraction pattern is dominated by the longitudinal coherence length. The weights for each rocking angle can be optimized so as to reflect the fraction of the peak intensity captured at each pixel for a given frame, if the rocking curve width has a known value.

This method is effectively a fit to the peak intensity. However it runs into problems when the Bragg condition is never reached. This will happen close to the rocking axis, where the Ewald sphere never cuts the reciprocal lattice plane. In that case, every single weighting coefficient will be close to zero, and subsequent normalization will only serve to amplify the noise in these areas.

There are several ways this can be handled, but one of the most justifiable and logically consistent is to use Bayes’ theorem, which brings with it several advantages, as we discuss below.

### III. BAYESIAN INFERENCE OF DIFFRACTION INTENSITY

The problem can be stated in the following terms. We wish to infer the integrated diffracted intensity  $I(\mathbf{q})$  for a region of reciprocal space at a given momentum transfer  $\mathbf{q} = (q_x, q_y, 0)$  in the reciprocal lattice plane [i.e., the plane in reciprocal space passing through  $(0, 0, 0)$  perpendicular to the field direction, which is defined as the  $(0, 0, 1)$  direction]. This coordinate system is fixed with respect to the field and sample; the rocking angle  $\omega$  defines the beam direction  $\mathbf{k}_i$ , and the detector pixel position determines  $\mathbf{k}_f$  and hence the momentum transfer  $\mathbf{q} = (q_x, q_y, q_z)$ . In the simplest case we make one assumption: that the diffraction pattern is 2D, except for a constant longitudinal correlation length, over the entire  $q$  range, i.e., that all diffracted intensity can be described by the in-plane intensity and a single Lorentzian or Gaussian maximum perpendicular to the plane, such that

$$I_{3D}(q_x, q_y, q_z) = f(q_z)I_{2D}(q_x, q_y, 0), \quad (1)$$

where a Lorentzian rocking curve shape gives

$$f(q_z) = \frac{1}{\pi} \frac{\frac{1}{2}\Gamma}{(\frac{1}{2}\Gamma)^2 + q_z^2} \quad (2)$$

or a Gaussian gives

$$f(q_z) = \frac{1}{\sqrt{2\pi}\gamma} \exp\left(-\frac{q_z^2}{2\gamma^2}\right), \quad (3)$$

with  $\Gamma$  being the FWHM in Eq. (2) and  $\gamma$  being the rms width for Eq. (3). We refer to the former from now on, but all considerations apply equally to a Gaussian shape.  $f(q_z)$  is normalized so that  $\int_{-\infty}^{\infty} f(q_z) dq_z = 1$ . It should be noted that  $f(q_z)$  has units of  $[1/q_z]$ .

This model makes no assumptions about the form of the in-plane intensity, either the diffraction pattern itself or the resolution effects. It explicitly rules out any modulation other than the correlation length  $2/\Gamma$  along the  $z$  direction which is the same for all Bragg planes of the FLL, though it could easily be extended to account for anisotropic  $\Gamma$ .

We have a set of data consisting of an array of counts from a 2D detector taken at a set of rocking angles  $\{\omega_j\}$ , already normalized and background subtracted (N.B., sets of multiple values or data points are denoted by curly brackets). The intensity at pixel  $i$  in frame  $j$  is referred to as  $D_{ij}$ , with an error  $\sigma_{ij}$  determined from Poisson statistics.

In terms of an inference problem, we wish to know the probability, for a given set of points at  $\{\mathbf{q}_i\}$ , that the integrated intensity has a particular value  $\{I_i\}$ , given the experimental data  $\{D_{ij}\}$ , and a FWHM  $\Gamma$ . For convenience and clarity we select  $\{\mathbf{q}_i\}$  to be equivalent to the detector pixels, which for a given  $|k|$  and small rotation angles can be assumed to have the same  $q_x$  and  $q_y$  for each  $\omega_i$ . We consider the case in which the neutrons are selected to be (approximately) monochromatic by a velocity selector. The time of flight mode, when a large range of neutron wavelengths is present during a measurement, can also be treated in the same way, but in that case the choice of  $\mathbf{q}$  values at which to calculate the intensity is not as straightforward.

The expected measurement at pixel  $i$  in frame  $j$  is given by  $f_{ij}I_i$  according to Eq. (1). This is a function of  $(q_x, q_y)_i$ , the integrated intensity  $I_i$ , rocking angle  $\omega_j$ , and rocking curve width  $\Gamma$  in reciprocal space.  $\Gamma$  can initially be estimated from the angular rocking curve width  $\eta$  for a particular flux lattice diffraction spot (later we make this a parameter to be determined).

The FWHM  $\Gamma$  along the field direction in  $\text{\AA}^{-1}$  for a spot centered on  $\mathbf{q} = (q_x, q_y, 0)$  and rocking axis parallel to the vector  $\mathbf{r}$  is given by

$$\Gamma = 2 \frac{|\mathbf{q} \times \mathbf{r}|}{|\mathbf{r}|} \tan(\eta/2), \quad (4)$$

where  $\eta$  is the rocking curve FWHM in degrees. Conversely  $\eta$  is given by

$$\eta = 2 \tan^{-1} \left( \frac{\Gamma |\mathbf{r}|}{2|\mathbf{q} \times \mathbf{r}|} \right). \quad (5)$$

In the usual case of a rock about a vertical or horizontal axis,  $\frac{|\mathbf{q} \times \mathbf{r}|}{|\mathbf{r}|}$  is equal to  $q_x$  or  $q_y$ , respectively.

For a set of measurements  $\{D_{ij}\}$  at a given pixel  $i$ , we want to know the best estimate of  $I_i$ ; i.e., what is the conditional probability  $P(I_i|\{D_{ij}\}, \mathcal{H})$  where  $\mathcal{H}$  represents our model and background knowledge and includes things like  $\Gamma$  and the measurement errors? Items to the right of the  $|$  sign indicate known information which may affect the probability of  $I_i$ . Including  $\mathcal{H}$  indicates there may be additional relevant parameters which can be explicitly taken into account if need be. We take the approach of introducing parameters when it becomes apparent they are necessary, which is intended to minimize the amount of computation needed.

#### A. Single pixel

Bayes' theorem gives, for a single frame and pixel,

$$P(I_i|\{D_{ij}\}, \mathcal{H}) = \frac{P(\{D_{ij}\}|I_i, \mathcal{H})P(I_i|\mathcal{H})}{P(\{D_{ij}\}|\mathcal{H})}. \quad (6)$$

$P(\{D_{ij}\}|I_i, \mathcal{H})$  is known as the likelihood and  $P(I_i|\mathcal{H})$  is the prior, representing our starting knowledge in the form of a probability distribution.  $P(\{D_{ij}\}|\mathcal{H})$  can usually be treated as a normalizing constant. This does not depend on  $I_i$  and ensures  $\int P(I_i|\{D_{ij}\}, \mathcal{H}) dI_i = 1$ .

The data will have a probability distribution which is a convolution of two Poisson distributions from the foreground and the background. If the number of counts is  $\gtrsim 10$ , this will approximate closely a Gaussian of mean  $f_{ij}I_i$  and a variance  $\sigma_{ij}^2$  equal to the sum of foreground and background variances.

For a single frame  $j$ , the likelihood is given by

$$P(D_{ij}|I_i, \mathcal{H}) = (2\pi\sigma_{ij}^2)^{-1/2} \exp\left[-\frac{(D_{ij} - f_{ij}I_i)^2}{2\sigma_{ij}^2}\right]. \quad (7)$$

Let's assume the prior  $P(I_i|\mathcal{H})$  is a Gaussian with mean  $\mu$  and variance  $\xi^2$ , i.e.,

$$P(I_i|\mathcal{H}) = (2\pi\xi^2)^{-1/2} \exp\left[-\frac{(I_i - \mu)^2}{2\xi^2}\right]. \quad (8)$$

The justification for this is discussed below, but note that  $\xi$  has different units from the sample error  $\sigma_{ij}$  and represents

the prior uncertainty on the total integrated intensity over the entire rock, not the measured intensity in any frame.

It can be shown that because each measurement is conditionally independent, i.e., given  $I_i$ , any measurement gives no further information about the next, then likelihoods can be multiplied [9]:

$$\begin{aligned} P(\{D_{ij}\}|I_i, \mathcal{H}) &= \prod_j P(D_{ij}|I_i, \mathcal{H}) \\ &= \text{const.} \times \prod_j \exp\left[-\frac{(D_{ij} - f_{ij}I_i)^2}{2\sigma_{ij}^2}\right] \\ &= \text{const.} \times \exp\left[-\frac{1}{2} \sum_j \left(\frac{D_{ij} - f_{ij}I_i}{\sigma_{ij}}\right)^2\right]. \end{aligned} \quad (9)$$

Substituting Eqs. (8) and (9) into Eq. (6), remembering that the denominator is a normalizing constant, gives

$$\begin{aligned} P(I_i|\{D_{ij}\}, \mathcal{H}) &= \text{const.} \times \exp\left\{-\frac{1}{2} \left[\left(\frac{I_i - \mu}{\xi}\right)^2\right.\right. \\ &\quad \left.\left.+ \sum_j \left(\frac{D_{ij} - f_{ij}I_i}{\sigma_{ij}}\right)^2\right]\right\}. \end{aligned} \quad (10)$$

To find the optimum value of  $I_i$ , we maximize Eq. (10), or equivalently its logarithm:

$$\begin{aligned} -\frac{1}{2} \frac{\partial}{\partial I_i} \left[ \left(\frac{I_i - \mu}{\xi}\right)^2 + \sum_j \left(\frac{D_{ij} - f_{ij}I_i}{\sigma_{ij}}\right)^2 \right] \\ = \frac{I_i - \mu}{\xi^2} + \sum_j \left[ \frac{-f_{ij}(D_{ij} - f_{ij}I_i)}{\sigma_{ij}^2} \right] = 0. \end{aligned} \quad (11)$$

Collecting terms in  $I_i$ ,

$$I_i \left( \frac{1}{\xi^2} + \sum_j \frac{(f_{ij})^2}{\sigma_{ij}^2} \right) = \frac{\mu}{\xi^2} + \sum_j \frac{f_{ij}D_{ij}}{\sigma_{ij}^2}, \quad (12)$$

i.e., the value of  $I_i$  which maximizes  $P(I_i|\{D_{ij}\}, \mathcal{H})$  is

$$I_i = \frac{\mu/\xi^2 + \sum_j f_{ij}D_{ij}/\sigma_{ij}^2}{1/\xi^2 + \sum_j f_{ij}^2/\sigma_{ij}^2}. \quad (13)$$

This can be written as

$$I_i = I_0 + \sum_j w_{ij} D_{ij}, \quad (14)$$

where  $I_0$  is proportional to the prior mean and uniform across all pixels, and the weights  $w_{ij}$  are given by

$$w_{ij} = \frac{f_{ij}/\sigma_{ij}^2}{1/\xi^2 + \sum_j f_{ij}^2/\sigma_{ij}^2}. \quad (15)$$

The error on  $I_i$  can be found by assuming that the posterior probability is Gaussian, which is true in this case as it is formed from the product of Gaussian functions. For a Gaussian function  $f(x)$  with mean  $x_0$ , the variance is given by

$f(x_0)/f''(x_0)$ , i.e.,

$$\sigma_{\text{total}}^2 = \left( \frac{1}{\xi^2} + \sum_j \frac{f_{ij}^2}{\sigma_{ij}^2} \right)^{-1}. \quad (16)$$

So the best estimate of  $I_i$ , including error, is given by

$$I_i = \frac{\mu/\xi^2 + \sum_j f_{ij}D_{ij}/\sigma_{ij}^2}{1/\xi^2 + \sum_j f_{ij}^2/\sigma_{ij}^2} \pm \left( \frac{1}{\xi^2} + \sum_j \frac{f_{ij}^2}{\sigma_{ij}^2} \right)^{-1/2}. \quad (17)$$

## B. Multiple pixels

When dealing with the entire detector, we can do the same thing for each pixel. The values of  $f_{ij}$  will be different for each position and will depend on the geometry of the setup, the neutron wavelength, and the value of  $\Gamma$ . The GRASP data analysis package [10], which was used to display and process the data (i.e., subtract backgrounds, normalize by total monitor counts, etc.), provides information about the values of  $\mathbf{q}$  at each pixel position and information about the Ewald sphere. This can be used to calculate  $q_z$  for any pixel and hence  $f_{ij}$ , given  $\Gamma$ ,  $\delta_\omega$ , and  $\delta_\phi$ , where we introduce the latter to represent any misalignment of the field with the coordinate axes defined in the experimental setup. These are included by subtraction from the diffractometer angles reported in the measurement file.

Usually we wish to know the optimum values of  $\{I_i\}$ ,  $\Gamma$ ,  $\delta_\omega$ , and  $\delta_\phi$ , i.e., those which maximize the posterior probability. To find these we make the rather strong simplifying assumption that all pixels are independent, implying their individual probabilities can be multiplied to form a joint posterior probability for all pixels. First, the optimum integrated intensity  $\{I_i\}$  for each pixel is calculated analytically as a function of  $\Gamma$ ,  $\delta_\omega$ , and  $\delta_\phi$  using Eq. (17), and the probability for each pixel is calculated using Eq. (10). It is quite straightforward to then numerically maximize the (log) joint posterior probability with respect to  $\Gamma$ ,  $\delta_\omega$ , and  $\delta_\phi$ , using priors which are broad Gaussians encompassing all likely values for all three.

When diffraction spots are visible in the raw data, which is usually the case, it is useful to mask the regions without diffracted intensity and exclude them from the probability calculation. When the optimum values of  $\Gamma$ ,  $\delta_\omega$ , and  $\delta_\phi$  have been determined, these can be used to calculate the entire diffraction pattern.

The independent pixel assumption does not take account of the finite in-plane instrumental resolution, which dominates the spot shape. The present method therefore preserves in-plane correlations. This should not affect the values of, for example, total integrated intensities over a whole diffraction spot, but may mean that errors are underestimated.

The misalignment  $\delta_\omega, \delta_\phi$  depends on the experimental setup, not the sample, except in cases where pinning to twin planes is extremely strong. It can therefore be determined using a reference sample which gives a strong signal, such as niobium, and regarded as fixed for subsequent measurements.

The errors  $\{\sigma_{ij}\}$  should be averaged over all frames for each pixel. This reduces noise and avoids unwanted correlations between weights and frame number.

### C. Choice of prior

In general we do not want our prior to significantly influence the result in regions where the data is informative, i.e., where the Ewald sphere has crossed the reciprocal lattice plane. This requirement means that the prior variance should be significantly larger than that of data and preferably should have a significant probability at the “correct” intensity. We also wish to avoid introducing any extra structure, so it must be applied uniformly across the detector. This can be accomplished while still leaving some freedom in areas where the data are not very informative. There is more than one approach we can take, each with different advantages; the choice is largely determined by how we would like to present the result, ideally in a pleasing, yet accurate, manner.

Beyond its 2D nature, for a FLL diffraction pattern the prior information we have is this: most pixels to have zero intensity, with a finite number of spots at  $q$  vectors with a simple symmetry and a positive intensity corresponding to the highest count value found in any individual frame. To reflect this accurately would require a complicated prior which assumes rather too much about the form of the diffraction pattern. Therefore for reasons of transparency and ease of calculation we use the Gaussian prior shown in Eq. (8), with mean  $\mu = 0$ , and variance  $\xi^2$  chosen large enough to encompass the range of intensities expected. So long as the prior is much broader than the likelihood function of the data, relevant data will quickly dominate the posterior probability density function. Setting  $\mu = 0$  can lead to negative as well as positive values, but it has the advantage that in the absence of relevant information it gives zero average intensity.

One option to determine  $\xi$  is to use a value equal to the maximum intensity recorded in any pixel (excluding the direct beam), normalized by  $\pi\Gamma/2$  to give an integrated intensity. This is a fairly good representation of our true state of knowledge, as we do not know if there are any spots in the unmeasured regions. The main problem with this prior is that it still produces rather large amounts of unsightly noise in these areas. This can be dealt with by masking these regions as described below. Alternatively the prior variance can be reduced until the noise is at an acceptable level, though care should be taken that the measured regions are not significantly affected.

Another possibility is to use a separate prior variance for each pixel, determined in the same way but using (a few times) the maximum number of counts in the individual pixels over the whole range of rocking angles. This has the effect of reducing the noise and giving a nicer looking result, but it will also tend to suppress weak signals, or those where  $f_{ij}$  is small for all angles but there is some evidence of a signal.

Masking of unmeasured regions can be carried out by comparing the prior and posterior probabilities. In regions of the detector in which the Bragg criterion is never reached, the data collected carry little information about the in-plane intensity. In this case the posterior variance will be close to that of the prior. These areas can be found by taking the ratio of the prior to posterior variances and masking pixels where this ratio is above a certain threshold. Alternatively, the uncertainty on the final result could be encoded in a color mapped display of the results, for example, as a color saturation value.

One can also consider a prior based on the peak intensity, rather than the integrated intensity (this is equivalent to normalizing the scaling factors  $f_{ij}$  such that  $f(0) = 1$ , but keeping everything else the same). In practice this is very much less useful, due to strong correlations between the peak width and height, so the integrated intensity is a better choice.

As all of these methods are applied identically to each pixel, there is no chance of introducing any additional structure beyond that indicated by the data. So long as the priors are not too strong, which is satisfied by the methods described, the data will dominate in the “measured” regions and the effect of the prior is confined to the places where the inferred intensity is weakly constrained.

## IV. RESULTS

Figure 2 shows diffraction patterns from  $\text{BaFeAs}_{(1-x)}\text{P}_x$  ( $x = 0.3$ ) at 5 T, obtained by rocking about a vertical axis [11]. Note in Fig. 2(b) the difference in intensity of the on-axis and off-axis spots, and the absence of spots at the top and bottom. Figure 2(d) shows the summed intensities in the marked sector boxes as a function of angle  $\omega$ . Off-axis spots will have larger FWHM, by a factor of  $1/|\cos(\alpha)|$ , where  $\alpha$  is defined above. This means that they will accumulate more intensity in a summed rock.

Figure 3 shows a typical map of the weighting coefficients  $w_{ij}$  for a frame taken at a single angle from a rock about the  $y$  axis, calculated using Eq. (15). As expected, there is a maximum corresponding to the position where the Ewald sphere intersects the reciprocal lattice. More surprisingly, there is also a region around the rocking axis where a large weighting factor appears. This is due to a small value of the denominator in Eq. (15), when all values of  $f_{ij}$  are  $\ll 1$  but the prior does not quite dominate. If the rocking curve width is broad, there may be relevant data present from the tails of the peak and, remarkably, a diffraction spot can be recovered which was not actually measured. Another slightly counterintuitive effect that can occur, particularly with narrow rocking widths or rocks in more than one direction, is that maxima in the weighting factor of one frame produce minima in the others, again via the denominator of Eq. (15).

Figure 4 shows the result using original data from Fig. 2, with and without smoothing (applied after all processing) and masking. In Fig. 4(d) the cumulative estimate for the integrated intensity of two on-axis diffraction spots is shown, along with error bars corresponding to  $\pm 1\sigma$ . As the data arrives, the estimated intensity of each spot converges, within errors, to a single value. In this example only one, vertical, rocking axis has been used, partly to illustrate the results with incomplete reciprocal space coverage. These data can be combined with that from a rock about a horizontal axis to give a more complete set of information about all spots with useful intensity over the entire detector area.

It is evident from inspection that this method gives a significant improvement in the signal to noise ratio. However it is hard to give this an exact value, as it relates to the fraction of the rocking curve which contains relevant data, which varies across the detector. One can make an estimate for the data shown here of about a factor of 2, thanks to the narrow rocking curve width.

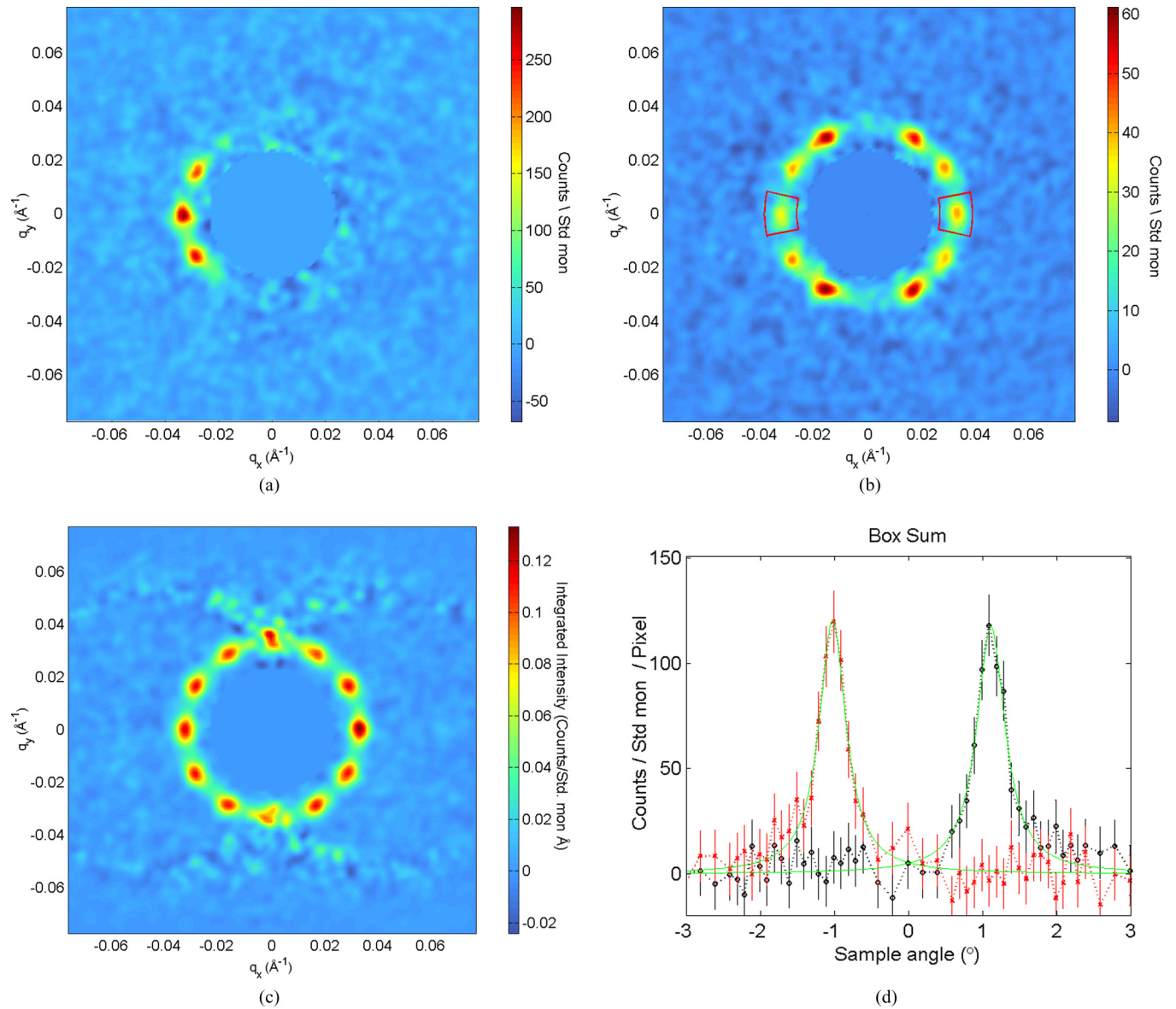


FIG. 2. (Color online) (a) Data taken at a single angle  $\omega$ . Only certain spots fulfill the Bragg condition to within the experimental resolution. (b) Sum of data taken over a range of angles encompassing all first-order Bragg reflections attainable by rotation about a vertical axis. (c) Result of Bayesian weighted method, using individual pixel priors, as described in Sec. III C. (d) Sum of intensity inside boxes shown in panel (b) plotted as a function of rocking angle. The width of the peaks is a measurement of the longitudinal coherence of the flux lattice, and the total integrated intensity under the peaks is proportional to the square of  $B(\mathbf{q})$ . The data in both panels (a) and (b) have backgrounds taken in the normal state subtracted from them. A small amount of smoothing (convolution with a  $3 \times 3$  pixel Gaussian) has been applied to panels (a)–(c) after all other processing. The direct beam in the center of the detector is masked.

An implementation of this method using MATLAB is available from the author.

## V. DISCUSSION

### A. Simplifying assumptions

The method shown above is deliberately as simple as possible, so as to be easily understood and reproduced as required. It is subject to the following approximations and limitations:

(i) The diffraction pattern is treated as strictly 2D, plus a single perpendicular coherence length. This could easily

be extended to different coherence lengths along different axes, but it will fail to correctly account for any out of plane scattering beyond this.

(ii) Probability distributions for the data are initially treated as Gaussian—in reality the foreground and background should be treated as two separate Poisson distributions, possibly with different scaling factors.

(iii) No account of finite in-plane resolution is taken. This could be incorporated into further modeling of the diffraction pattern or treated as a deconvolution problem. The presence of correlations between pixels means that the calculated errors on “box sums”, summed intensity over several pixels, may

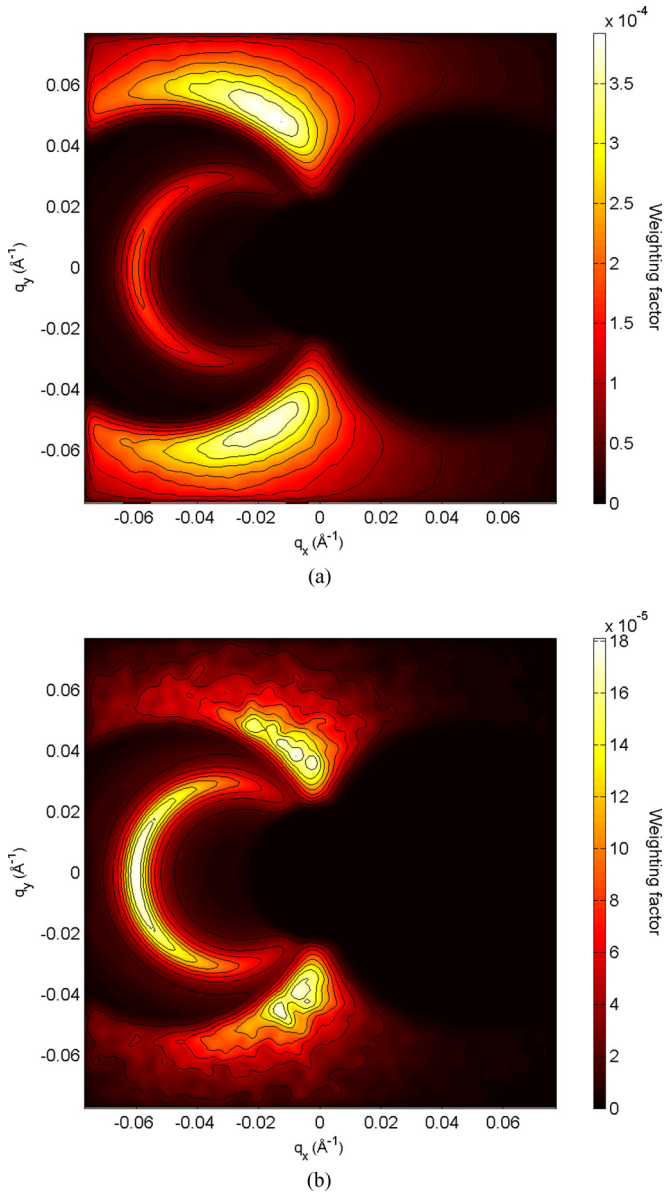


FIG. 3. (Color online) Weighting factors  $w_{ij}$  for a single frame using (a) uniform prior and (b) individual pixel priors, as described in Sec. III C. Note the crescent shape where the Ewald sphere cuts the lattice plane, but also the maxima at the top and bottom, which are the result of the denominator in Eq. (15). In panel (b) these are less prominent as the prior is suppressed by the low counts in these regions, though more noise is evident. A  $3 \times 3$  pixel Gaussian smoothing has been applied to both figures.

be incorrect; however, the optimum values should reflect the true integrated intensity (again assuming everything is Gaussian).

(iv) The maximum posterior probabilities for the rocking width and misalignment values are used. In principle with additional computational effort one could marginalize these out as nuisance parameters, i.e., integrate over all possible values weighted by the posterior probability.

A way in which the presented method can fail to improve on the usual unweighted sum is if there is a very broad rocking curve, much larger than the range of angles measured.

In this case all frames will contain relevant information, so there will not be a large improvement in the signal to noise ratio. The misalignment and rocking curve width will not be strongly constrained by the data, so unless there is information about this from other sources, the weighting factors may not be accurate, even for the best fit solution. This is actually somewhat reassuring, as the Bayesian method will not make a bad sample into a good one!

### B. Philosophical issues

Many objections to Bayesian methods relate to so-called “subjective” priors. The prior is, however, unavoidable, as any other approach is equivalent to an implicit, usually uniform prior. In the limit of very large amounts of high quality data, the choice of prior will be largely irrelevant; however in scattering experiments we are often limited by statistical noise, and we wish to extract the maximum information from data collected in a limited region of reciprocal space with finite resolution. In this case, the choice of a suitable prior can be extremely useful.

In the present case we can divide the prior into two components. The first is the prior as understood in the usual sense of a prior probability distribution for the parameters of interest, i.e.,  $P(\{I_i\}, \Gamma, \delta_\theta, \delta_\phi | \mathcal{H})$ , which can be separated as  $P(\{I_i\} | \mathcal{H})P(\Gamma | \mathcal{H})P(\delta_\theta | \mathcal{H})P(\delta_\phi | \mathcal{H})$  if all are independent. In general, so long as no zero values are included, and the sharpness or curvature (i.e., information content) of the prior distribution is much smaller than that of the likelihood, this part of the prior will make very little difference to the final result after accounting for the data. On the other hand, using a broad but informative prior encompassing a physically reasonable range of parameters can help the solution to converge numerically in a way that may not happen if a uniform prior is chosen and the data are not good enough in quality or quantity. Previous measurements carried out with the same or different techniques can also be used to produce an informative prior. This can be very useful, as often different measurement methods (x-ray vs neutron scattering, for example) can have different sensitivities or resolution in different parts of parameter space.

The second component to the prior is more subtle, it essentially comprises the choice of model (or set of models) to be considered and contains information about symmetries of the system. This is contained within the background information  $\mathcal{H}$ .

The combination of the two types of prior determines the parameter space of interest; the larger the parameter space, the more data required to come to a conclusion. This means in particular that prior information about symmetry can have a huge effect. In the current case the reduction from 3D to 2D reciprocal space means a large increase in the signal to noise ratio, as one can average over the irrelevant dimensions.

An interesting question remains as to whether or not the entire background information  $\mathcal{H}$  can be described in this way. That is to say, what information is necessary to fully define a statement of probability as a state of knowledge once a record of all experimental results has been taken into account? Is the choice of model space and symmetry enough, or is some other information required? This is a philosophical,

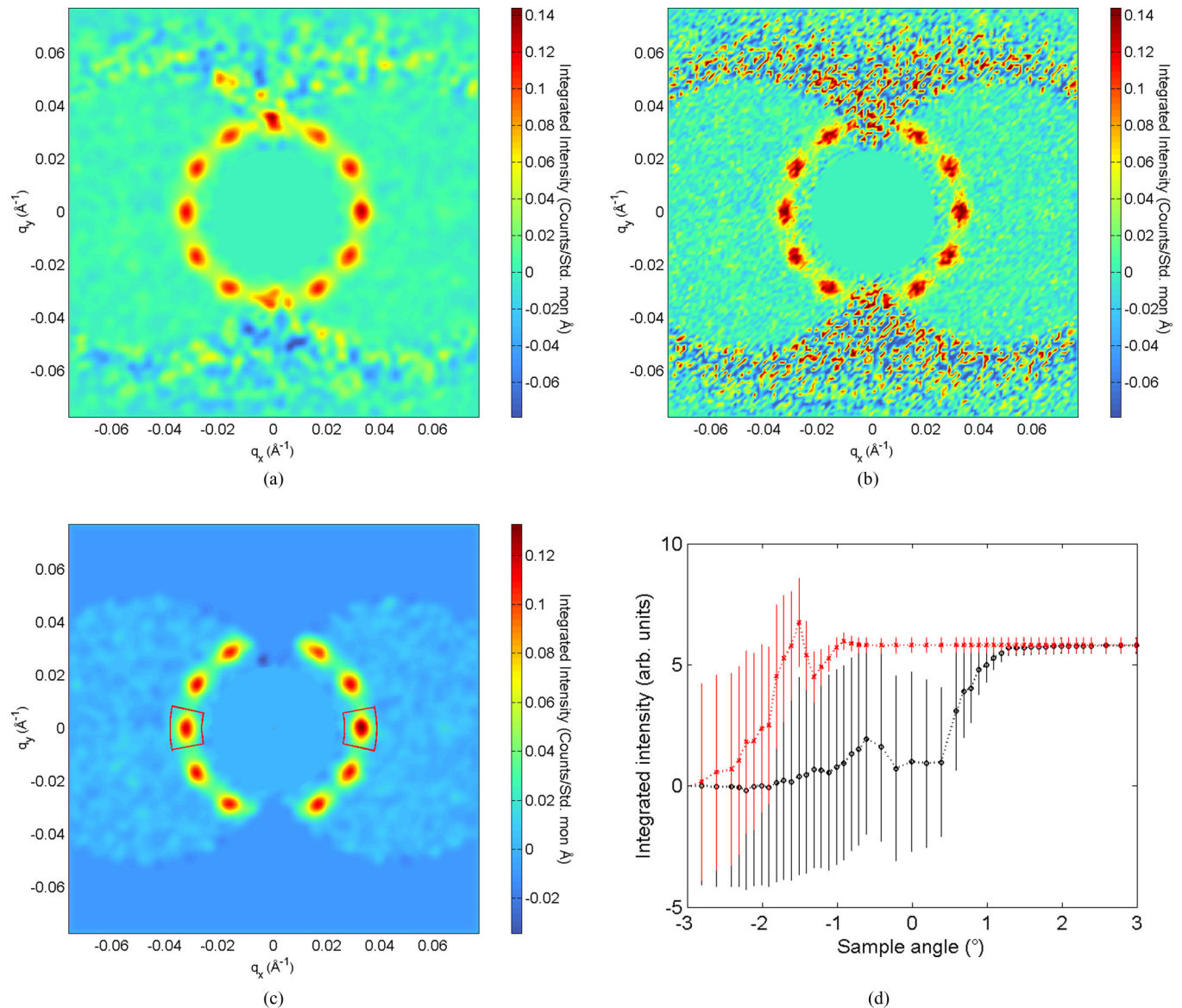


FIG. 4. (Color online) (a) Result of Bayesian weighted sum of data from Fig. 2, with prior mean  $\mu$  set to zero, and uniform s.d.  $\xi$  determined by the maximum spot intensity. Note that spots are recovered at the top and bottom, even though they never reach the Bragg criterion. Gaussian smoothing has been applied to the weighted sum with a  $3 \times 3$  pixel resolution. (b) The same as panel (a) without smoothing. (c) Areas with  $\sigma > 0.2\xi$  are masked, where  $\sigma$  is the error on the posterior intensity. (d) Sum of intensity inside boxes in panel (c) as a function of angle. These converge on the final result at different positions, depending on the angle at which the peak appears. Data taken at angles after passing through the peak of the rocking curve at the Bragg condition make very little difference to the final result.

or perhaps mathematical, point, but raises an important issue when considering how to represent actual probabilities, which in practice are always conditional, in an unambiguous way.

## VI. CONCLUSIONS

The results shown on flux lattices demonstrate that Bayesian techniques can provide very large improvements in the quality of data analysis. The resulting scattering patterns more accurately represent the integrated intensities than simple sums of the data. They can include the usual (Lorentz) corrections in a natural way, by working in a model space suited to the problem. A very simple Gaussian analytical treatment is

more than adequate. This has the advantage of transparency as well as speed of implementation. The maximum posterior probability solution was chosen, with rocking curve width and field misalignments with the coordinate axes left as free parameters.

The case of the FLL may be particularly suited to this problem, due to its 2D nature, but it is by no means the only such system, another example being Skyrmion lattices [12]. Often there is also a lot of accumulated data about instrument characteristics or accumulated experimental data from previous experiments. These can all be naturally taken into account using Bayesian methods, and there are many examples of such work over the years in a wide variety of fields [13–17]. Due to the relatively complex nature of the analysis, which



often gives similar if not identical results to more traditional methods (though the exceptions are of significant interest), this has mostly been carried out by motivated individuals with a model hard-coded into the analysis. However the methods themselves are completely general and model independent. As computers get cheaper and more powerful with respect to the cost of gathering data, it is starting to make sense to provide general tools for model-independent Bayesian data analysis. By separating the modeling from the calculations and using standardized methods for data reduction it would no longer be necessary to reinvent the wheel for each problem. A dataflow programming language, such as LABVIEW/G, would be the most natural way of encapsulating this process, particularly as many of the calculations are inherently parallel. This would require an agreed data format for conditional probability distributions, meeting the requirements detailed above to include all necessary background information. This would particularly suit facility-based techniques such as

neutron and synchrotron radiation scattering, which have institutional computing support and a well-characterized data archival procedure. We have shown how this method is highly advantageous for one particular system, and we look forward to its widespread adoption.

#### ACKNOWLEDGMENTS

Many thanks to Ted (E. M.) Forgan for the original suggestion of a weighted rocking sum and many thoughtful discussions, Charles Dewhurst for letting me butcher his GRASP software, Hazuki Kawano-Furukawa for generous use of her BFAP SANS data, and Jonny White, Sebastian Mühlbauer, and Elizabeth Blackburn for useful discussions and comments. This work was supported by the Engineering and Physical Sciences Research Council (Grant No. EP/J016977/1); and the JSPS BRIDGE fellowship program.

- 
- [1] C. Dewhurst and R. Cubitt, *Phys. B* **385–386**, Part 1, 176 (2006).
  - [2] M. Eskildsen, *Front. Phys.* **6**, 398 (2011).
  - [3] M. R. Eskildsen, E. M. Forgan, and H. Kawano-Furukawa, *Rep. Prog. Phys.* **74**, 124504 (2011).
  - [4] W. H. Kleiner, L. M. Roth, and S. H. Autler, *Phys. Rev.* **133**, A1226 (1964).
  - [5] M. Laver, E. M. Forgan, S. P. Brown, D. Charalambous, D. Fort, C. Bowell, S. Ramos, R. J. Lycett, D. K. Christen, J. Kohlbrecher, C. D. Dewhurst, and R. Cubitt, *Phys. Rev. Lett.* **96**, 167002 (2006).
  - [6] M. Laver and E. M. Forgan, *Nat. Commun.* **1**, 45 (2010).
  - [7] E. van den Pol, A. V. Petukhov, D. M. E. Thies-Weesie, D. V. Byelov, and G. J. Vroege, *Phys. Rev. Lett.* **103**, 258301 (2009).
  - [8] R. Cubitt, E. Forgan, D. Paul, S. Lee, J. Abell, H. Mook, and P. Timmins, *Phys. B* **180–181**, Part 1, 377 (1992).
  - [9] A. P. Dawid, *J. R. Statist. Soc. Ser. B* **41**, 1 (1979).
  - [10] GRASP User Manual, Technical Report No. ILL03DE01T, 2003, available at <http://www.ill.fr/lss/grasp>.
  - [11] R. Morisaki-Ishii, H. Kawano-Furukawa, A. S. Cameron, L. Lemberger, E. Blackburn, A. T. Holmes, E. M. Forgan, L. M. DeBeer-Schmitt, K. Littrell, M. Nakajima *et al.*, [arXiv:1405.7468](https://arxiv.org/abs/1405.7468).
  - [12] S. Mühlbauer, B. Binz, F. Jonietz, C. Pfeleiderer, A. Rosch, A. Neubauer, R. Georgii, and P. Böni, *Science* **323**, 915 (2009).
  - [13] D. S. Sivia and C. J. Carlile, *J. Chem. Phys.* **96**, 170 (1992).
  - [14] K. V. Klementev, *J. Phys. D: Appl. Phys.* **34**, 209 (2001).
  - [15] V. Dose, *Rep. Prog. Phys.* **66**, 1421 (2003).
  - [16] C. R. Hogg, J. B. Kadane, J. S. Lee, and S. A. Majetich, *Bayesian Anal.* **5**, 1 (2010).
  - [17] U. von Toussaint, *Rev. Mod. Phys.* **83**, 943 (2011).