Interpreting macromolecular diffraction through simulation

**Iris D. Young[a], Derek Mendez[a,b], Billy K. Poon[a],
Johannes P. Blaschke[c], Felix Wittwer[c], Michael E. Wall[d], and
Nicholas K. Sauter[a,*]**

[a]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory,
Berkeley, CA, United States
[b]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA,
United States
[c]National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory, Berkeley,
CA, United States
[d]Computer, Computational and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos,
NM, United States
*Corresponding author. e-mail address: nksauter@lbl.gov
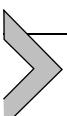
Contents

Abstract

This chapter discusses the use of diffraction simulators to improve experimental outcomes in macromolecular crystallography, in particular for future experiments aimed at diffuse scattering. Consequential decisions for upcoming data collection include the selection of either a synchrotron or free electron laser X-ray source, rotation geometry or serial crystallography, and fiber-coupled area detector technology vs. pixel-array detectors. The hope is that simulators will provide insights to make these choices with greater confidence. Simulation software, especially those packages focused on physics-based calculation of the diffraction, can help to predict the location, size, shape, and profile of Bragg spots and diffuse patterns in terms of an underlying physical model, including assumptions about the crystal's mosaic structure, and therefore can point to potential issues with data analysis in the early planning stages. Also, once the data are collected, simulation may offer a pathway to improve the measurement of diffraction, especially with weak data, and might help to treat problematic cases such as overlapping patterns.

1. Introduction

Most macromolecular crystallographers are familiar with one or more software packages for data reduction, such as *xds* (Kabsch, 2010b) or *dials* (Winter et al., 2018), which take the raw X-ray diffraction patterns, and convert the Bragg spot intensities to real-valued amplitudes of structure factors, which are the Fourier coefficients for calculating the electron density map. Diffraction simulation is the opposite concept—it is the idea of having a program that would accurately and quantitatively predict every pixel on the diffraction pattern, including the location, size, shape, and intensity profile of every Bragg spot, given a set of underlying conditions. There is a long history of software packages geared toward such simulation. Recent authors have used them for instructive purposes; for example, Diederichs illustrated the effects of mosaicity and spectral dispersion on the diffraction pattern (Diederichs, 2009), while Holton & Frankel calculated the minimum crystal size needed for a complete diffraction data set (Holton & Frankel, 2010). Given the current progress in diffuse scattering, where it seems that the analysis of correlated motion may be within reach (Wall, Wolff, & Fraser, 2018), it is appropriate to ask whether simulation programs can and should be extended to cover the diffuse pattern.

Let us first ask whether the purpose of simulation is solely educational, or does it play a critical role in data analysis? After all, if we have a powerful data reduction package, can we not simply interpret the data by direct

measurement? For example, to process Bragg spots, we could run a spot-finding program, and then integrate the signal by summing up all the photons on the Bragg-spot pixels. The reality, however, is that Bragg spot measurement is never direct. Rather, quantitative modeling has been an important part of data reduction for decades, both for locating the Bragg spots on the image, and for accurately finding their intensities. The critical thing to appreciate is that the diffraction pattern is a combination of Bragg reflections and background diffraction (which includes diffuse scattering). Measuring a given Bragg spot always means quantifying the total signal and then subtracting the background, all the while accounting for experimental noise, which itself consists of several parts (counting uncertainty, instrument readout, and systematic effects) and potential overlap of neighboring spots. It is easily shown (Bourenkov & Popov, 2006) that the Bragg signal is often weaker than the background, especially at higher resolutions. While many strong spots may be noticed on a diffraction pattern, the weaker ones are not visually apparent by simply looking at an image, and would not be detected by a naïve spotfinding algorithm. Therefore, in real data reduction, the strong spots are used to deduce the lattice (the unit cell and crystal orientation), then all the spots are predicted (both strong and weak), and Bragg intensity is summed on the pixels considered to be at the predicted positions (Leslie, 2006).

Therefore it is clear that quantitative modeling has an established role in processing the data *post*-experiment, but what about for planning purposes, *pre*-experiment? Here it needs to be acknowledged that the required modeling tools are so familiar that they are used routinely without discussion. For example, synchrotron beamline user interfaces like *Blu-Ice* allow the scientist to decide whether the detector position will allow for high-angle Bragg spots, or if instead the detector needs to be repositioned closer to the sample (McPhillips et al., 2002). However, with today's expanded focus on collecting data to reveal diffuse scattering, it is prudent to re-examine the toolset to make sure we are asking the right questions. Since the diffuse signal lies between Bragg spots, it may be sensible to focus on whether the unit cell parameters and detector distance permit sufficient pixels between the Bragg orders to distinguish the low-intensity diffuse signal from the Bragg spikes. Also, we may ask if the Bragg peaks might be broadened by other factors such as beam divergence or dispersion, or the crystal mosaic structure. If the Bragg peaks are too wide, neighboring spots may partially overlap and preclude the observation of diffuse signal. The software described herein is capable of illustrating these effects for a specific crystal form identified by the

program user, in the form of a PDB–format structure file. It thus serves as a mechanism for general education, with the added potential of driving the successful design and execution of future experiments.

To frame the discussion that follows, it is useful to draw a distinction between *empirical* and *physical* modeling approaches. In many cases, data reduction workflows are empirical in the sense of not necessarily breaking down the underlying physics of diffraction. A key example is the spot analysis performed by the program *xds* (Kabsch, 2010a). The popularity of that software is due in part to its approach to modeling of the size and shape of the Bragg spot, employing a clever transformation to a coordinate system in which all Bragg reflections share a common intensity profile described by two empirical parameters. Under these conditions, the structure factor (Fourier coefficient) intensity can be modeled by fitting a single scale factor to the data.

In contrast to the empirical approach, we will take a physical approach in view of our pedagogical purpose, to show how subtle properties of the diffraction pattern arise based on more fundamental building blocks. Familiar elements of the experiment are shown in Fig. 1, including the X-ray beam, detector, crystal, and rotating goniometer, along with parameters used to express the experimental geometry and sample properties. While the list is not comprehensive, it is at least a minimum

B X→ Beam	D Detector	C Crystal	G Goniometer
		P Profile	S Scan
• Wavelength	• Distance	• Space group	• Rotation range
• Bandwidth	• Position	• Cell parameters	• Scan angle
• Incident flux	• Panel metrology	• Cell variances	• Exposure time
• Pink spectrum	• Bad pixel mask	• Orientation	• Dose
• Polarization	• Gain	• Illuminated volume	• Spindle axis
• Divergence	• Pedestal	• Mosaic rotation	• Background scattering
	• Common mode	• Mosaic block size	
	• Quantum efficiency	• Absorption path	
	• Material/Thickness	• Radiation damage	
		• Structure factors	
		• Diffuse intensity	

Fig. 1 A general list of experimental elements for physical modeling. The first three columns list parameters relevant to all experiments, while the fourth column applies only to rotation experiments, and not to serial crystallographic data, where every shot is static with no known rotational relationship to any other shot.

starting point. We are now poised to construct a diffraction model from this set of parameters. There are indeed data reduction workflows that adhere to physical modeling principles, in particular *eval15* (Schreurs, Xian, & Kroon-Batenburg, 2009). Here the physics is represented by four statistical distributions, representing respectively the crystal volume, mosaic texture, divergent beam, and wavelength spread. Random drawings are performed from the four distributions, with the diffraction pattern built up by thousands of repeated drawings. Parameters describing the width of these distributions are fit interactively through a graphical interface. Ultimately the information carried forward to structure solution is derived from the simulation rather than directly measured from the raw data. While primarily meant for small molecules, *eval15* has also been applied to protein crystallography (Kroon-Batenburg, Schreurs, Ravelli, & Gros, 2015). Separately, James Holton described a similar ray-tracing approach in great mathematical detail (Holton, Classen, Frankel, & Tainer, 2014), and together we have collaborated to include his program *nanoBragg* (Sauter, Kern, Yano, & Holton, 2020) in the Computational Crystallography Toolbox (*cctbx*). We subsequently published the code *diffBragg* (Mendez et al., 2020), to calculate first derivatives of the *nanoBragg* simulation, for future use in parameter optimization calculations that will ultimately form a new data reduction pipeline. Other current authors are also using simulation to address issues in data reduction (Brehm, White, & Chapman, 2023).

This chapter is an attempt to highlight our physical modeling approach to a more general audience, while its eventual applications to data reduction (and particularly to diffuse scattering data) are still under development. We describe *simtbx.sim_view*, which supports the simulation functionality of *diffBragg* within an interactive image viewer, and which was used as a teaching tool at the 2022 School of Crystallography in Erice, Italy. Section 2 discusses current best practices for the experimental measurement of protein diffuse scattering, Section 3 presents the general principles of pixelwise image simulation, while Section 4 considers how to extend these pixel models to at least some aspects of diffuse scattering. Section 5 explains how *simtbx* models the complexity of real-world experiments, while Section 6 enumerates the model parameters that may be adjusted by the software user, and some that await implementation. Section 7 speculates on future directions for data interpretation, while Section 8 gives instructions for downloading and using the package.
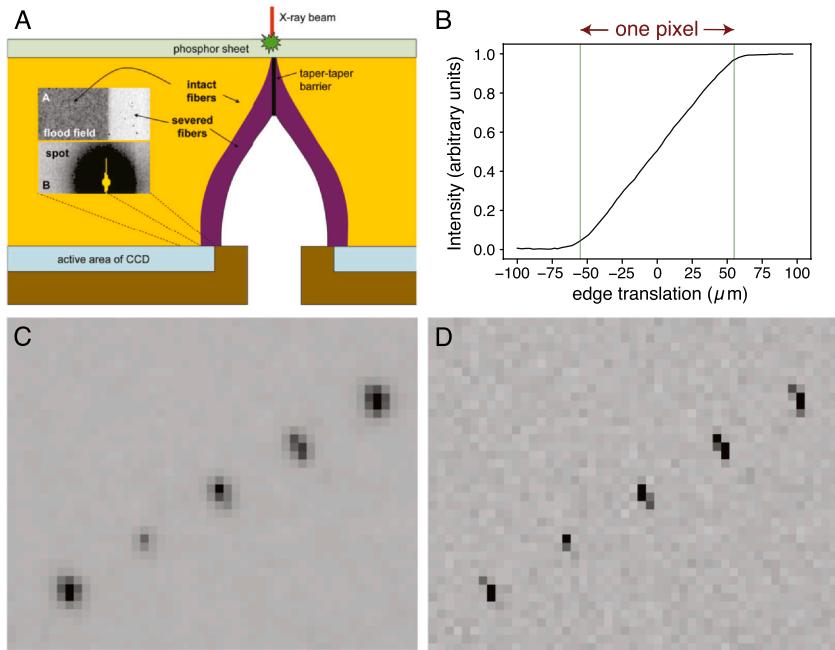
2. Present status and future prospects for diffuse scattering

The above discussion of macromolecular diffuse scattering empha-sizes the importance of purposeful planning, with careful optimization of the experimental geometry so that the diffuse pattern is well-measured and clearly distinguishable from the Bragg pattern. It is therefore useful to examine recent experience, noting how experimental choices can lead to successful data collection. We focus the discussion on MCA, the 2020 paper of Meisburger, Case, and Ando (2020). Not only does this paper present an exciting result dealing with correlated motion, it also has a noteworthy methods section, serving as a road map for future experiments by identifying critical best practices for measurement of diffuse scattering data. We examine these practices here, especially considering their relation to diffraction simulation.

2.1 Pixel array detectors

The MCA authors' first experimental choice was to use a Pilatus imaging detector, which uses pixel array detection (PAD), meaning the direct detection of X-ray photons in a silicon crystal via the production of photoelectrons. Critically, a knife-edge test (Fig. 2B) shows that PAD technology reads out the photon signal in essentially the same pixel where the photon is absorbed, thus producing a very sharp Bragg spot (Fig. 2D) absent other effects [in this test, the edge of a knife held against the detector at a slight angle is used to generate a sequence of pixels with smoothly varying X-ray exposure (Koerner, Philipp, Hromalik, Tate, & Gruner, 2009; Philipp, Koerner, Hromalik, Tate, & Gruner, 2010)]. This is in contrast to fiber-optic coupled detectors (Fig. 2A), where X-ray photons are first converted to visible photons within a thin phosphor sheet, which are in turn coupled through a fiber optic taper to a charge-coupled device (CCD) detector. Here, the secondary visible photons generated by the phosphor spread out (within the phosphor) in all directions (Fig. 2C), by one or two pixels (Holton, Nielsen, & Frankel, 2012). Mathematically this is described as a point-spread function (PSF) that is applied to the incident diffraction pattern. While fiber-optic coupled detectors are very popular for general structure solution, the PSF makes them less than ideal for diffuse scattering, as it leaves less room to record the diffuse scattering pattern in between Bragg spots. In addition, commercially available CCD detectors exhibit artifacts that interfere with measurement of diffuse scattering when

Fig. 2 A comparison of imaging detector types. (A) In a fiber-coupled area detector, X-rays are converted to visible photons in the phosphor layer (green), then coupled through a fiber-optic taper (orange) to a charge-coupled device (blue). (B) In a pixel array detector (PAD), the intensity response of one pixel as a knife edge is translated across the detector surface, exposed to an X-ray flood field. Photons are directly converted to detected charge in silicon, with most of the response confined to one pixel. (C) A simulated image models the point-spread function of a fiber-coupled detector, while (D) the corresponding simulation of a PAD shows much sharper Bragg spots. *Panel A: Reproduced from Holton et al. (2012), with permission from IUCr Journals. Panel B: Adapted from Koerner et al. (2009), with permission from IOP Publishing.*

the number of X-rays recorded on a pixel becomes too large, due to the spread of charge to neighboring pixels after saturation. While it is possible to simulate the PSF with *nanoBragg*, and therefore possible in theory to deconvolute its effects, we will only consider PAD technology for the discussion of simulation.

2.2 Space group and unit cell

Studying lysozyme was a best-case scenario, as it is a well-known, relatively small protein. Larger proteins to be studied in the future will have larger unit cells, and by the nature of reciprocal space the resulting Bragg spots will be closer together in image space. Crowded patterns will make it

harder to detect the diffuse signal in between Bragg spots, but this might be mitigated by simple measures such as moving the detector further from the sample. For this paper it was a judicious choice to use a crystalline lysozyme form in space group $P1$, since lysozyme normally crystallizes in space group $P4_32_12$, and thus produces a Bragg pattern eight times as crowded as the one studied here. Higher-symmetry crystals provide more multiplicity, which is advantageous for increasing signal-to-noise and improving characterization of data quality. In contrast, lower-symmetry crystals provide more independent measurements for diffuse scattering, which can be advantageous in modeling. These examples support the argument that there is a potential role for simulation in the planning of difficult future experiments. One can read a PDB file to set the symmetry and unit cell, set the detector type and geometric parameters, and use a physical model to predict the size and shape of spots. One can then inspect for overlap, either from spot to spot, or between the spots and the diffuse halo.

2.3 Ambient temperature

It has become clear over the last decade that proteins exist in a conformational ensemble that can be observed crystallographically (Woldeyes, Sivak, & Fraser, 2014), with the likelihood that enzyme function and regulation are determined by contact networks that can be modulated in response to the local environment (van den Bedem, Bhabha, Yang, Wright, & Fraser, 2013). Furthermore, cryopreservation techniques that have long guarded against radiation damage are best avoided, in order to sample conformation space under operating conditions (Keedy et al., 2015). Therefore the MCA authors were particularly concerned with keeping the radiation absorbed dose below 65 kGy, thus protecting against both general damage that reduces resolution, and specific damage that accrues to relevant chemical species like disulfide bonds and metallocenters in enzymes. Monitoring the electron density maps in real time during rotational data collection, they translated the crystal to a fresh position in the crystal at the first sign of damage. This was possible because large crystals were available in the 100–200 μm range.

If large crystals are not available in future studies, the best path forward may lie with serial crystallographic methods. Here the experiment is performed at ambient temperature, but only one shot is taken per crystal to avoid radiation damage. Thus there is a need to accumulate the full pattern by merging thousands of patterns from small crystals in random orientations. The serial approach is available at either synchrotron sources

(with millisecond up to second time scales), or at X-ray free electron lasers (XFELs, where pulses, tens of femtoseconds long, enable pump-probe experiments in the femtosecond to microsecond regimes). In both cases, there is a capability for performing time-domain studies under ambient conditions. However, serial crystallography also presents challenges such as the partiality of Bragg spots, as the data are collected with still shots rather than with goniometer rotation. XFEL X-ray pulses also have significant X-ray dispersion. Simulation programs offer tools for understanding (and therefore mitigating) these problems, as will be discussed.
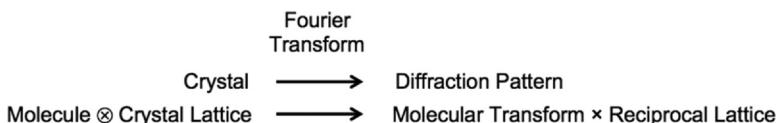
3. Image simulation methods
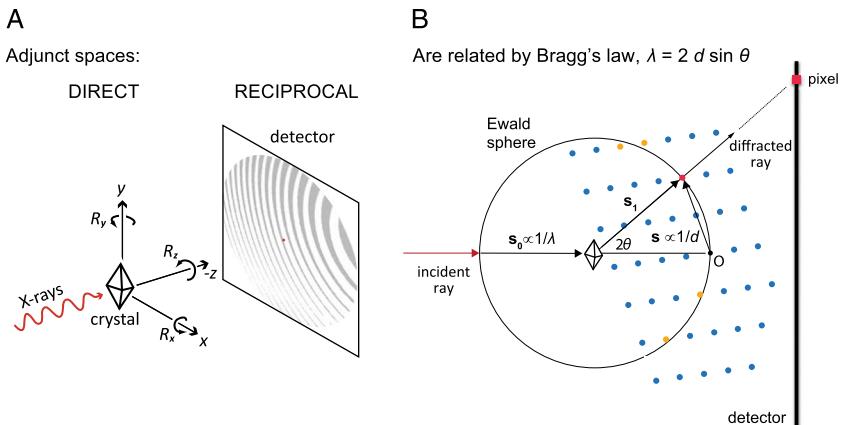
　　Before mentioning the details of image simulation, we will outline the mathematical framework, which is the same framework used to explain crystallography in a graduate-level class. The crystal is thought of as existing in *direct space*, and the diffraction pattern in *reciprocal space*, with the two adjunct spaces co-rotating, and related to each other by Fourier transformation (Scheme 1).

　　A copy of the molecule is found at each position in the crystal lattice. Mathematically this is the operation of convolution ($\otimes$), and its Fourier transform is a multiplication. Specifically, in the diffraction pattern, the molecular transform is sampled at each reciprocal lattice point. Experimentally, the Bragg spot intensity gives us the structure factor (squared) for the corresponding Miller index. Furthermore, the Ewald sphere construction (Fig. 3) reminds us that we only observe Bragg spots that are exactly in the diffraction condition.

　　We need a few proportionality factors to make sure our simulated image is calculated with sensible units (photons/pixel) and is quantitative on an absolute scale. Various formulae go back at least a hundred years (Darwin, 1914) but the modern equivalent may be summarized as (Eq. (1)):

Fourier
Transform

Crystal　$\longrightarrow$　Diffraction Pattern

Molecule $\otimes$ Crystal Lattice　$\longrightarrow$　Molecular Transform × Reciprocal Lattice

Scheme 1 Initial framework for simulation.

A

Adjunct spaces:

DIRECT   RECIPROCAL

B

Are related by Bragg's law, $\lambda = 2\,d\sin\theta$
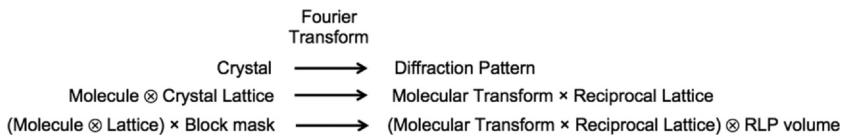
Fig. 3 Fundamental framework for image simulation. As implied by Scheme 1, (A) the adjunct spaces representing crystal (direct) and diffraction pattern (reciprocal) are (B) related by Bragg's law, relating the wavelength λ, the scattering angle 2θ, and the spatial resolution d. The red reciprocal lattice point (located a reciprocal distance $1/d$ from the reciprocal origin O) satisfies the diffraction condition and therefore produces a scattered ray, reaching the indicated detector pixel. Although the adjunct spaces co-rotate, their rotation origins differ in the Ewald sphere construction (B), with the crystal rotating around the sphere center, and the reciprocal lattice rotating around O. Adapted from Sauter et al. (2014).

$$I_{\mathrm{pixel}} = J_0\left(\lambda\right) r_e^2\, P\, \Delta\Omega_{\mathrm{pixel}}\, N_{\mathrm{cells}}\, |\,F_{\mathrm{cell}}\left(\mathbf{h_0}\right)\,|^2\, F_{\mathrm{latt}}^2\left(\mathbf{s}\right)$$

where $J_0\left(\lambda\right)$ is the number of photons incident on the crystal, r_e is the radius of the electron, P is a polarization factor (Kahn et al., 1982), $\Delta\Omega_{\mathrm{pixel}}$ is the solid angle subtended by the pixel, and N_{cells} is the number of illuminated unit cells. The factor $|\,F_{\mathrm{cell}}\left(\mathbf{h_0}\right)\,|^2$ is just the familiar structure factor intensity at a particular Miller index $\mathbf{h_0}$. Finally for our initial discussion, $F_{\mathrm{latt}}\left(\mathbf{s}\right)$ is a simple delta function, telling us if the reciprocal lattice point is on the Ewald sphere (in diffracting condition), given our unit cell and crystal orientation that determines the scattering vector $\mathbf{s}$.

From here we can add nuanced detail to our mathematical treatment. The statement that $\mathbf{F_{\mathrm{latt}}(s)}$ is just a delta function must be an over-simplification, since that would imply seeing only one Bragg spot at a time as the crystal rotates on the goniometer, rotating successive reciprocal lattice points through the Ewald sphere. In reality we know that diffraction patterns are crowded with Bragg spots. Let's re-examine the math implied by Scheme 1, and introduce further complexity, Scheme 2:

Fourier
Transform

Crystal $\longrightarrow$ Diffraction Pattern

Molecule ⊗ Crystal Lattice $\longrightarrow$ Molecular Transform × Reciprocal Lattice

(Molecule ⊗ Lattice) × Block mask $\longrightarrow$ (Molecular Transform × Reciprocal Lattice) ⊗ RLP volume

Scheme 2 Revised simulation framework taking into account the finite size of the mosaic block.
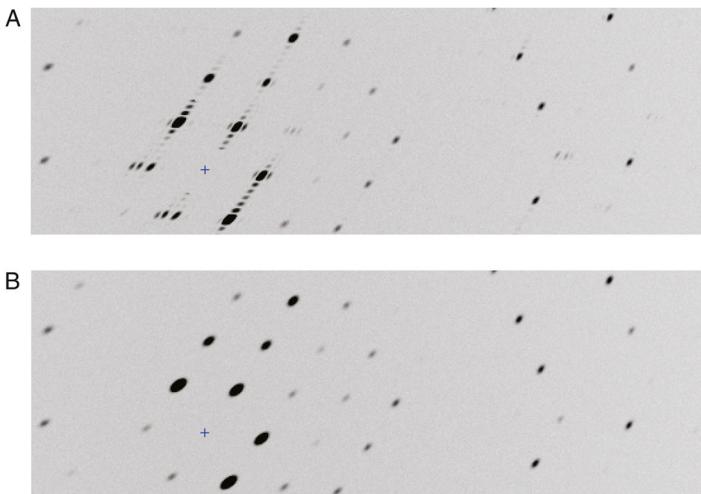
On reconsidering, the reciprocal lattice point is a delta function only if the crystal lattice is infinite in extent (due to the Fourier transform of an infinite sine wave being a single spatial frequency). We may think of a real crystal being constructed, mathematically, as an infinite crystal multiplied by a block mask that describes the volume of matter that diffracts coherently, i.e., where the scattering from unit cells constructively interferes to produce Bragg spots. As Scheme 2 shows, this has the effect of placing a reciprocal lattice point volume (RLP volume) at every reciprocal lattice position, again the operation of a convolution. This reciprocal object, the Fourier transform of the coherently-diffracting mosaic block, is finite in size and thus gives rise to the crowded pattern of Bragg spots.

Let's consider a possible mathematical form for the RLP volume. Suppose that the block mask (the coherently scattering mosaic domain) is a rectangular parallelepiped of $N^3 = 7 \times 7 \times 7$ unit cells. We get the familiar Laue interference pattern (Eq. (2)),

$$F_{latt}(\mathbf{s}) \propto \frac{\sin \pi N h}{\sin \pi h} \times \frac{\sin \pi N k}{\sin \pi k} \times \frac{\sin \pi N l}{\sin \pi l}$$

where h, k, and l are the reciprocal space coordinates (Miller indices), and specifically for $N = 7$, this gives rise to five rectangular interference fringes between every pair of Bragg spots (Fig. 4A). While it is difficult to observe this pattern in practice, there is one famous XFEL experiment where diffraction was collected from single mosaic blocks, giving rise to precisely this type of pattern (Chapman et al., 2011). In a more typical experiment, the crystal might consist of numerous mosaic blocks all of slightly different sizes and shapes, with the total diffraction being the incoherent sum of the coherent diffraction from each block (Fig. 4B). Therefore it no longer makes sense to speak of a specific value of N, but rather a Gaussian distribution, such that the overall RLP shape might reflect the average block size $\langle N \rangle$ while taking a Gaussian form (Eq. (3)),
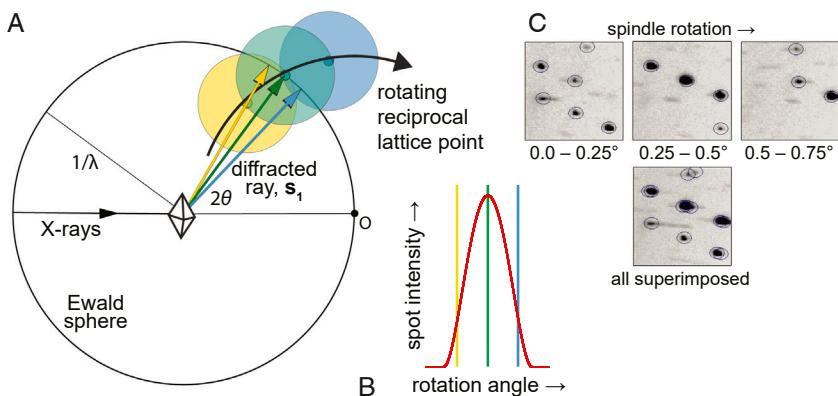
$$F_{latt}(\mathbf{s}) \propto e^{-(\langle N \rangle \Delta \mathbf{h})^2}$$

Fig. 4 Size and shape of the mosaic block affects the Bragg spot geometry. Each reciprocal lattice point may be thought of as a Fourier transform of the average coherently scattering block within the crystal (Scheme 2). If the block is a rectangular parallelepiped, seven unit cells on an edge (Eq. (2)), the simulated pattern is as shown in (A), with five fringes appearing between every Bragg spot center. In contrast, if the block geometry draws from a 3D Gaussian distribution with a mean width of $\langle N \rangle = 7$ cells (Eq. (3)), the generated pattern is as shown in (B). Crosshairs indicate the direct beam.

Importantly, this implies a Gaussian falloff of pixel intensity when the RLP is a reciprocal space distance $\Delta \mathbf{h}$ away from the Ewald sphere. The behavior may be summarized as follows: when the average block size $\langle N \rangle$ is small, many Bragg reflections are visible because RLPs are in diffracting condition relatively far from the Ewald sphere; and vice versa, with few reflections being visible if $\langle N \rangle$ is large (Sauter et al., 2014).

This gives us some fairly sophisticated grounding for understanding the finer points of diffraction (Pflugrath, 1999; Sauter & Adams, 2017). As the RLP volume passes through the Ewald sphere (due to a goniometer rotation), the intensity of a Bragg spot first increases, then reaches a maximum, and then decreases, eventually to zero (Fig. 5B). This is the famous "rocking curve". The rocking curve might be narrower (in rotation) than one X-ray frame, or it might extend over several consecutive frames in the rotation series. To correctly measure the structure factor, we must integrate the number of photons over the full rocking curve, and such a Bragg spot would be considered to be "fully measured". Interestingly, as the RLP passes through the Ewald sphere (Fig. 5A), the $\mathbf{s_1}$ vector representing the scattering direction (and which points from the Ewald sphere
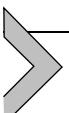
Fig. 5 Bragg spot position changes slightly with crystal rotation. (A) As the crystal rotates on the goniometer spindle, a reciprocal lattice point (RLP) co-rotates thru the diffraction condition (the Ewald sphere). However, the s_1 vector, indicating the direction of the corresponding Bragg spot, makes a slightly larger scattering angle 2θ when the RLP is inside the sphere, than when the RLP is outside. (B) The rocking curve indicates that the diffracted spot intensity reaches a maximum when the RLP intersects the Ewald sphere surface. (C) Experimental diffraction patterns collected at increasing spindle rotations reflect these effects. *Panels A, B: Reproduced from Sauter and Adams (2017) with permission from the Royal Society of Chemistry. Panel C: Adapted from Pflugrath (1999).*

center to the RLP center) changes slightly, just a fraction of a degree. The visible manifestation is that Bragg spots change their position very slightly (Fig. 5C), as the crystal rotates on an axis perpendicular to the beam. It is a small effect but it should be accurately reproduced by the simulator.

What happens with the rocking curve during serial crystallography? Here there is no goniometer rotation; the rocking curve is sampled at a single orientation only. Without the context of a model, we don't know if we are near the rocking curve peak or near the tails. In short, we are in the uncomfortable position that the thing we measure (the Bragg spot intensity) is not at all proportional to the thing we are seeking (the integrated area of the rocking curve). The initial approach in serial crystallography was to simply improve the outcome with overwhelming multiplicity of measure–ment. A more sophisticated approach is to use a model to determine a separate correction factor for each Bragg intensity, so our partial measurements can be converted to the full spot equivalent. Physics-based modeling is especially applicable here, especially if iterative parameter estimation (*postrefinement*) can be used to derive best estimates of the crystal, detector, and beam parameters, along with the mosaic block size $\langle N \rangle$. Many software authors addressed this

problem in the last decade, leading to a dramatic improvement in the electron density maps (Ginn et al., 2015; Kabsch, 2014; Kroon-Batenburg et al., 2015; Sauter, 2015; Uervirojnangkoorn et al., 2015; White, 2014).
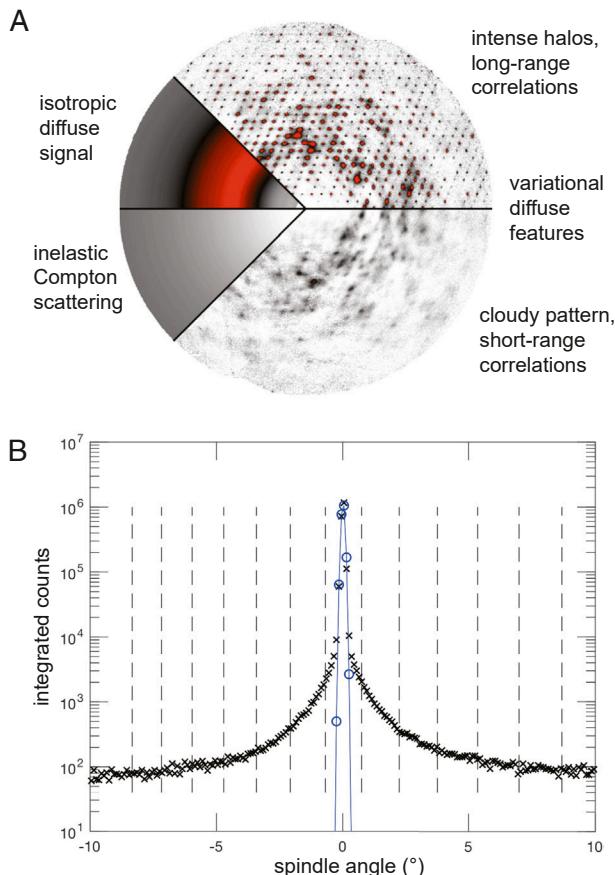
4. Extending simulators to diffuse scattering

We are now in a position to include some limited information about diffuse scattering in the simulator. First let's give a plausibility argument. We know that there are several contributing factors to the diffuse scattering (Fig. 6) among which the inelastic Compton scattering is the least interesting for molecular function, and will not be addressed here. The isotropic diffuse signal contains interesting information about internal motions, but varies on a reciprocal length scale longer than we are concerned with here. Of the remaining variational diffuse features, there is a cloudy pattern observed between lattice layers that informs us about short-range correlations in the molecular motion; this will also not be discussed here. Finally there is a pattern of intense halos that surround each Bragg spot, which are visible at the lattice positions even after Bragg spots have been removed (Fig. 6A). These halos reflect the long-range correlations on scales of a few unit cells. An alternative representation is shown in Fig. 6B taken from the MCA paper, showing the scattering intensity around one Bragg spot as a function of goniometer rotation. The largest signal (blue) is just the rocking curve for the Bragg spot. In this particular experiment, where the spindle was rotated in steps of $0.1°$, the full width of the Bragg spot was typically found to be on a single rotational step. The intense halo also has a characteristic footprint (black), with a width extending about 1 or $2°$ on either side of the central peak. The fact that we can clearly see the intense halo as a separate phenomenon, which mimics the well-understood rocking curve producing the $F_{latt}(s)$ factor in Eq. (1), suggests that we could simply add another term to our simulation that would represent the reciprocal space form of the intense halos (Eq. (4)):

$$I_{pixel} = J_0(\lambda)\, r_e^2\, P\, \Delta\Omega_{pixel}\, N_{cells}\, |\, F_{cell}(\mathbf{h_0})\, |^2\, \times$$
$$\left(F_{latt}^2(s) + F_{diffuse\ halo}^2(s) \right)$$

Possible functional forms for the $\mathbf{F}_{diffuse\ halo}^2(s)$ term for protein crystals have been discussed in the literature for at least 30 years. Caspar et al. expressed the idea of an autocorrelation function to express an exponentially
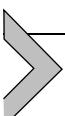
Fig. 6 Summary of the diffuse scattering pattern from macromolecular crystallography. (A) A section through reciprocal space, highlighting various contributions to the diffuse isotropic scattering (left) and diffuse anisotropic, or variational scattering (right). (B) Detail of the scattering focusing on one Bragg spot, as the goniometer spindle rotation brings the reciprocal point through the Ewald sphere. The blue curve is the rocking curve proper, with its sharp peak at the central position (note the logarithmic scale of intensity), while the black data originate from the long-range correlations, forming an intense halo around the tails of the Bragg peak. *Adapted from Meisburger et al. (2020)*.

decreasing effect with correlation length γ (Caspar, Clarage, Salunke, & Clarage, 1988). The Fourier transform of this function gives an expression for the diffuse halo density in reciprocal space, which models the black curve of Fig. 6B. Since we are discussing long-range correlations between unit cells, a first estimate of the length γ could be 50 Å. Anisotropic correlation can also be

represented, in which case the single parameter γ would be replaced with a tensor Λ. Finally, an additional parameter σ gives the vibrational amplitude of the motion, typically of magnitude <1 Å. We cannot go into mathematical detail here, but the reader may refer to Michael Wall's calmodulin paper (Wall, Clarage, & Phillips, 1997) where Wall Eq. (8) essentially gives the $F^2_{\text{diffuse halo}}(\mathbf{s})$ term, with Wall Eq. (9) used for the autocorrelation function. The reader is also referred to Meisburger Eqs. (138) to (141) in a recent review (Meisburger, Thomas, Watkins, & Ando, 2017).

Eq. (4), with its diffuse halo term, is incorporated into the *simtbx*. *sim_view* simulator associated with this chapter (Fig. 7).

5. Complexity of the simulation

It should be emphasized that the Eq. (4) model for pixel intensity is valid only for a still shot. If the experiment includes goniometer rotation, we take a simple summation over intermediate goniometer rotational positions (Eq. (5)),
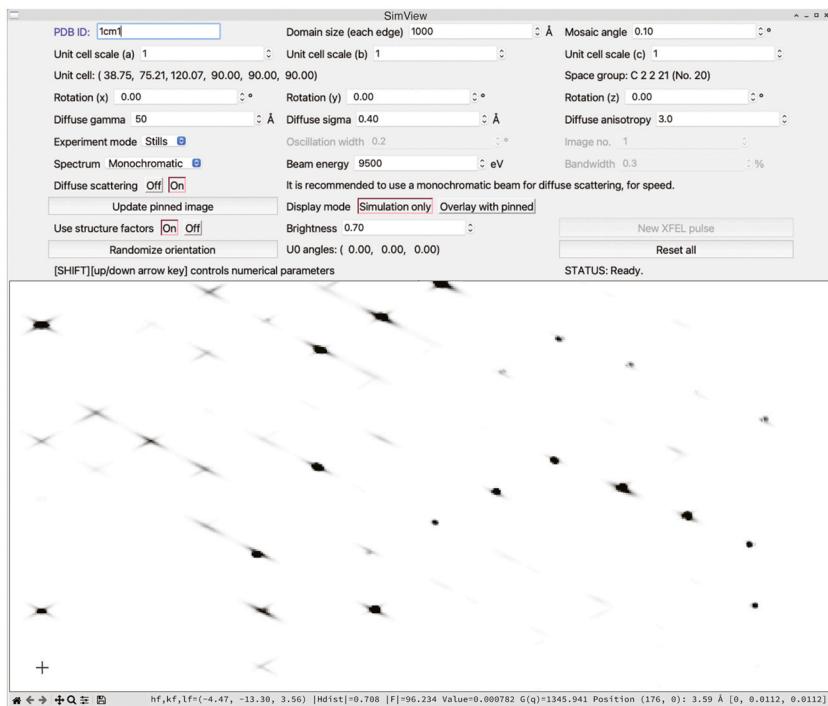
$$I_{\text{pixel}} = J_0(\lambda)\, r_e^2\, P\, \Delta\Omega_{\text{pixel}}\, N_{\text{cells}}\, |\, F_{\text{cell}}(\mathbf{h_0})\,|^2\; \times$$

$$\sum_{\substack{\text{spindle} \\ \text{rotation}}} \left(F^2_{\text{latt}}(\mathbf{s}) + F^2_{\text{diffuse halo}}(\mathbf{s}) \right).$$

For example, we might represent a 0.1° rotation as a sum over 10 still shots spaced 0.01° apart. The changing rotational position does not appear explicitly but is implicit in Eq. (5). The exact math is beyond the scope of this chapter, but we can show the basic idea (and the reader is invited to skip this paragraph if only interested in the broad framework). In simplest terms (Fig. 3B), let's define the incident beam vector as $\mathbf{s_0}$, and the scattered ray vector as $\mathbf{s_1}$, both with a length of inverse wavelength $(1/\lambda)$, and $\mathbf{s} = \mathbf{s_1} - \mathbf{s_0}$. Then the fractional Miller index $\mathbf{h}$ implicitly used for $F^2_{\text{latt}}(\mathbf{s}) + F^2_{\text{diffuse halo}}(\mathbf{s})$ is (Eq. (6)),

$$\mathbf{h} = \mathbf{R}_\phi \mathbf{A} \mathbf{s}$$

where $\mathbf{R}_\varphi$ is the rotation matrix corresponding to a spindle rotation of φ, and $\mathbf{A}$ is the direct space orientation matrix (Eq. (7)),

$$\mathbf{A} = \begin{pmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{pmatrix}$$

Fig. 7 Snapshot of the *simtbx.sim_view* image viewer, simulating the diffuse halo pattern from calmodulin crystals (PDB entry 1CM1), which reflects a long-range correlation between unit cells, according to the model of Wall et al. (1997), Eq. (8). Parameters for the correlation length γ and vibrational amplitude-squared σ^2 may be specified by the user to empirically fit experimental data. In this case, streaked diffuse features were produced by multiplying γ by an anisotropy factor along the *a–b* and *c* unit cell directions in direct space (leaving *a+b* alone), and multiplying σ^2 by the same anisotropy factor along the *a–b* direction (leaving *a+b* and *c* alone). The Laue symmetry was then applied, leading to streaks along both the $a^* - b^*$ and $a^* + b^*$ directions in reciprocal space. In the present case, this type of anisotropy was purely an *ad hoc* decision allowing us to generate a simulation that resembles the data in the 1997 paper, but this exercise suggests that correct models may require similar anisotropic features in the general case.
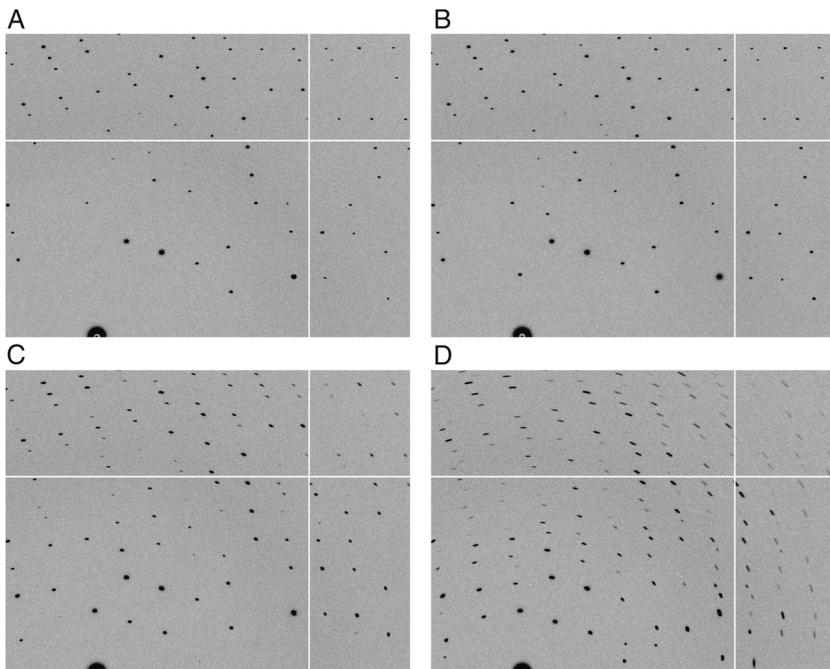
consisting of the x, y, and z components of the unit cell basis vectors **a**, **b**, and **c**.

Another source of computational complexity is the phenomenon of mosaic rotation. In Eq. (5) we are treating all the mosaic blocks of the crystal as having the same exact orientation represented by matrix **A**. This may be a reasonable approximation for work at ambient temperature.

However, from several decades experience with cryopreserved crystals, we know that the quick cooling process essentially shatters the crystal into many mosaic blocks of slightly different orientations, such that the Bragg spots can actually appear as tiny arcs (Fig. 8D) concentric on the direct beam (Nave, 2014). We handle this in our simulation by adding another large sum over individual mosaic rotations (Eq. (8)),

$$I_{\text{pixel}} = J_0(\lambda) r_e^2 P \Delta\Omega_{\text{pixel}} N_{\text{cells}} | F_{\text{cell}}(\mathbf{h_0}) |^2 \times$$

$$\sum_{\substack{\text{spindle} \\ \text{rotation}}} \sum_{\substack{\text{mosaic} \\ \text{rotations}}} \left(F_{\text{latt}}^2(\mathbf{s}) + F_{\text{diffuse halo}}^2(\mathbf{s}) \right)$$

where the rotations represent a Gaussian distribution over small angles with randomly distributed axes of rotation.
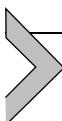
Fig. 8 Simulations that differ in the crystal's mosaic rotational disorder but that are otherwise identical. (A) Perfectly ordered mosaic blocks (rotational full width $\eta = 0°$). (B) Small mosaicity typical of ambient temperature crystals ($\eta = 0.1°$). (C) Medium-high mosaicity found with cryopreserved samples ($\eta = 0.6°$). (D) Large mosaicity, with arclets indicating high disorder ($\eta = 1.5°$). Courtesy of James Holton.

A final source of complexity is the beam dispersion. While synchrotron beamlines are often monochromatic (with a dispersion of ~0.01% in energy), XFEL sources have a broader dispersion of ~0.3%, with a unique mean energy for each shot, and a spiky spectrum with a stochastic shape. We incorporated 100 experimental XFEL spectra into the *simtbx.sim_view* simulator so that the user can sample the slight differences in the diffraction. These are accessed by the "New XFEL pulse" button in the GUI. Computationally we incorporate spectral shape into the calculation with (Eq. (9)):

$$I_{\text{pixel}} = \sum_{\text{wavelength } \lambda} J_0(\lambda) \, r_e^2 \, P \, \Delta\Omega_{\text{pixel}} N_{\text{cells}} \, | F_{\text{cell}} (\mathbf{h_0}) |^2 \times$$

$$\sum_{\substack{\text{spindle} \\ \text{rotation}}} \sum_{\substack{\text{mosaic} \\ \text{rotations}}} \left(F_{\text{latt}}^2 (\mathbf{s}) + F_{\text{diffuse halo}}^2 (\mathbf{s}) \right)$$

where the factor $J_0(\lambda)$ refers to the number of incident photons within each wavelength interval. The effect of spectral dispersion is to transform Bragg spots into radial streaks emanating from the direct beam position (Fig. 9).

Considering the triple summation, Eq. (9) may not be the most computationally efficient design for image simulation, and alternate ideas are welcome. To help compensate for the combinatorial footprint, we have implemented a GPU-accelerated version, presently available for users with Linux machines with Nvidia graphics cards.
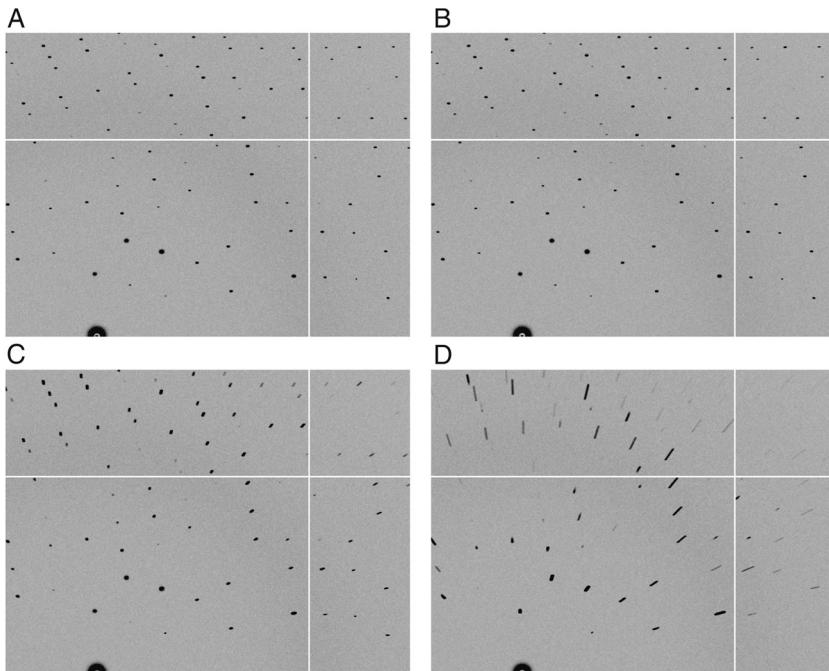
6. Parameter space and implications for simulation

Here we offer further discussion of each of the parameters thus far identified as relevant to simulating the diffraction experiment, with a brief overview of how we anticipate the user may interact with them via the *simtbx.sim_view* display.

6.1 Incident beam

The X-ray beam can be described in terms of its wavelength, flux, divergence, polarization, shape, and bandpass or spectrum. We ignore shape as the area exposed to the beam is bundled into illuminated volume, and the shape itself has no effect. We are also implicitly ignoring the pulsed nature of an XFEL beam as well as flux, which again do not matter if we are not addressing radiation damage or any other changes over the course of the exposure. Wavelength is clearly important—this determines diffracting conditions—and therefore bandpass also concerns us. Here we

Fig. 9 Simulations that differ in the dispersion of the incident beam but that are otherwise identical. (A) Low dispersion ($\Delta E/E = 0.014\%$) typically produced by a synchrotron monochromator equipped with a Si 111 crystal. (B) Medium dispersion ($\Delta E/E = 0.25\%$) found in an XFEL pulse. (C) Pink beam, $\Delta E/E = 1.6\%$ and (D) Laue, $\Delta E/E = 5.1\%$. Courtesy of James Holton.

must consider the possibilities case wise. If we are working with a monochromatic beam that we are content to model as perfectly monochromatic, we use a single wavelength. If we choose to model bandpass, our model currently approximates this as a histogram of contributions from different energy channels and sums these contributions. This approach can handle a pink beam or an XFEL spectrum equally well. We have three spectrum shapes available in *simtbx.sim_view*: monochromatic, Gaussian, or SASE (self-amplified spontaneous emission, applicable to XFEL experiments). As just mentioned, we include 100 measured SASE spectra from LCLS in our simulator, and the binning of this histogram was selected to preserve the different shapes of these spectra. The binning is not a parameter exposed to the user at present, but it would be trivial to make this change if the need arises. The effect of a noticeable bandpass or spectrum is a radial streaking of each Bragg spot (Fig. 9D). It is more noticeable when a

large mosaic rotation also spreads peaks over many pixels, and under the right conditions, the user can observe realistic shot-to-shot differences when simulating the exposure of the same crystal to different XFEL pulses. Finally, divergence and polarization are properties that are important to model correctly, but typically are fixed at a particular beamline, and are known quantities. We therefore incorporate these into the model but do not expose them as variables to adjust or refine.

6.2 Unit cell and symmetry

The reciprocal lattice itself scales with the unit cell of the crystal, producing a global expansion or contraction of the diffraction pattern on the detector. We include each unit cell parameter as an independently adjustable variable in the simulation, as we have often found ourselves in the position of observing a slightly different unit cell than expected during an experiment. Where applicable, crystal symmetry is enforced so that symmetry-related axes scale together. The user is encouraged to explore multiple orientations when adjusting unit cell parameters, since the effects of scaling each axis will be more or less visible in different orientations—when the beam is aligned with one of the unit cell axes, changes in that axis length will not alter the spot-to-spot distance. The "randomize orientation" button can be useful for exploring many orientations.

6.3 Crystal orientation

The orientation of the crystal with respect to the beam determines which Bragg reflections will be observed (Fig. 3B). An initial crystal orientation is acquired during indexing, but in practice, especially for serial crystal-lography, small errors may remain. We measure this misorientation as rotational offsets on the x-, y-, and z-axes, and denote these rot_x, rot_y and rot_z, respectively. Rot_z is easy to detect and correct, since this angle rotates spots about the origin in the plane of the detector. Rot_x and rot_y are more difficult. These latter two have almost no effect on the position of spots on the detector, but do impact which reflections meet the diffracting condi-tions. In rotation crystallography we have the opportunity to integrate over the entire rocking curve as the reciprocal lattice point rotates through the Ewald sphere, so a slight misorientation is tolerable. In contrast, recall that in serial crystallography we accumulate only a large number of slices on individual still shots. This renders us extremely sensitive to any mis-estimation of a reflection's position relative to the Ewald sphere: Although each point in the reciprocal lattice is smeared out into a small (reciprocal)

volume by crystal imperfections, allowing us to record a partial spot even when we are not perfectly in diffracting conditions, the resulting intensity changes dramatically with slight rotations of the crystal. All three (mis) orientation angles may be adjusted in the *simtbx* simulation. To aid the user in viewing the different effects of rotations in *x*, *y* and *z*, we also simulate a "reference" crystal, which begins as identical to the primary simulated crystal but does not change as the simulation parameters are updated. (It can be brought back into register with the "Update pinned image" button.) The user toggles whether the reference is displayed in red; the primary simulation is always displayed in cyan, and the two combine to produce white where pixel intensities match perfectly. A spectrum emerges at positions where intensities are only slightly different.

6.4 Mosaic character

Mosaicity describes variation among individual coherently scattering blocks within a crystal, and there are multiple approaches to parameterizing this effect (Nave et al., 2016). In *simtbx.sim_view* we focus on the distribution of mosaic block orientations. We presume this to be a three-dimensional normal distribution (over rot_x, rot_y, and rot_z) and thus sample a representative set of orientations (Eq. (8)). Interference between the mosaic blocks is ignored: the coherent diffraction from each block is computed separately and summed to produce the simulated diffraction pattern. These simplifications appear to be reasonable for both cryocooled microcrystals at synchrotrons and crystals at ambient temperature at XFELs, although when mosaic domains are individually, explicitly modeled, there may be further improvement in the agreement between measured and simulated images. Additional parameters of mosaicity not modeled here include variation in unit cell dimensions and variation in the mosaic block size, *N*. The former would produce uniform radial streaking of Bragg peaks and the latter would produce nonuniform distortion of the peaks. As we are primarily interested in applying these simulations to cases where other effects on spot shape are much stronger, we have omitted these variables for now. For simplicity we model the mosaicity as isotropic. It is possible to model anisotropic rotations and block sizes in *nanoBragg* and *diffBragg*, and this functionality could be extended to *simtbx.sim_view* in the future.

6.5 Diffuse scattering

As discussed earlier in this chapter, several types of diffuse scattering arise from different physical phenomena, and only the diffuse halos around Bragg peaks

produced by long-range correlations on the order of several unit cells are modeled in our simulation. By default diffuse scattering is disabled in *simtbx.sim_view* since it is an expensive calculation, but it can be enabled with a toggle switch. (We recommend selecting a monochromatic beam or turning off diffuse scattering while exploring other parameters.) Halos are parameterized by a correlation length γ in Ångstroms, and the amplitude σ of the correlated vibrations also in Ångstroms (Fig. 7). These also govern the strength of the diffuse signal relative to the Bragg signal. Diffuse scattering may be either isotropic or anisotropic, and we have incorporated a unitless anisotropy term in our simulation to reflect this, which at present is assigned a specific directionality. An increasingly thorough treatment of diffuse scattering is a high priority for ongoing work (Peck, Poitevin, & Lane, 2018; Wall et al., 2018).
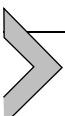
6.6 Oscillation width and rotation

Our simulator was designed with serial crystallography as the target use case, but we later added a rotation mode in order to explore rotation experiments as well. Note that this mode is much slower because each oscillation image is produced as the sum of a series of still shots, so we generally do not recommend adjusting other variables with this mode enabled, and we especially caution against enabling it at the same time as diffuse scattering. Rotation experiments are presumed to be carried out at synchrotron sources, so we fix the spectrum shape to monochromatic in this mode. The subset of parameters displayed in the GUI is adjusted accordingly when switching between modes.

6.7 Background and noise

The *simtbx.sim_view* viewer still awaits features that usefully depict background and noise. Air and water scattering are well-understood and in principle easily modeled. There may be other known sources of scattering or absorption as well, but these two typically dominate, producing familiar isotropic patterns. Several types of noise arise from the experiment. Shot noise is a consequence of the statistical likelihood of observing a particular photon given a ground-truth photon rate, and it follows a Poissonian distribution. The shorter the exposures (and fewer photons collected on each image), the greater the fractional contribution of shot noise to each image. There is also readout noise, a result of dark current, which is the signal registered by the detector in the absence of any photons. This is because at the stage where photons impinging on the detector are converted to electrical signal, the same process of charge separation into

electrons and holes very occasionally occurs spontaneously. The likelihood of these spontaneous events is governed by the Boltzmann distribution and increases with temperature. Readout noise, or simply read noise, follows a Gaussian distribution. Detectors that do not convert X-ray photons directly to electron–hole pairs—indirect detectors, as opposed to direct detectors— are also subject to a point-spread function (PSF; Fig. 2A). X-ray detectors also do not have perfect linear responses, and diffraction requires a very large dynamic range compared with direct imaging. Ideally each additional photon hitting the detector would register as one additional photon event, but at very high photon counts, this response drops off. This last effect can be surprisingly difficult to identify from a diffraction pattern directly, as it does not alter the expected Bragg peak shapes. Detectors usually have some variation in responsiveness from pixel to pixel even after calibration, manifesting as slightly different gains, and typically have a few "hot" and "cold" pixels that deviate more dramatically. Cosmic radiation occasionally interacts with the detector, producing a strong, sharp signal at a random location; this can be distinguished from a hot pixel by the fact that it appears on only one image. In *simtbx.sim_view*, we presume an ideal detector and avoid modeling most of these phenomena. *nanoBragg* and *diffBragg* do, however, have the functionality necessary to handle them. The relative contributions of various sources of noise are well worth exploring in a detailed simulation.
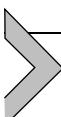
7. Conclusions

With the expectation that image simulation can serve as both an informative teaching reference and an aid for planning future experiments, this chapter provides a first implementation of the graphical program *simtbx.sim_view*. However, given its focus on pre-experiment modeling, the present software represents only a starting point. Ultimately we wish to include a mechanism for post-experiment data interpretation. At minimum, this would require the development of new code to simultaneously display the computational model alongside the experimental data. With this type of interactive display, the software user could then adjust the model parameters so that the model matches the data as closely as possible. This is essentially the data processing paradigm of *eval15* (Schreurs et al., 2009), except that the parameter set would be expanded to include aspects of the diffuse scattering such as correlation length, vibrational amplitude and anisotropy.

We certainly expect that most parameter optimization will be performed automatically, just as is currently done with conventional diffraction experiments. Indeed, the program *diffBragg*, which provides the foundation for *simtbx.sim_view,* is capable of computing the first-derivative of the simulated diffraction pattern with respect to each parameter. Therefore it is envisioned that it can be used for iterative inverse modeling, using gradient–descent methods to minimize the variance between model and experiment (Mendez et al., 2020). However, we caution that this prospect of automated optimization does not obviate the need for developing an intuitive under–standing of the underlying physics. We can easily imagine a situation where numerous physical parameters (such as the unit cell and mosaic structure) are refined within an automatic data reduction pipeline, only to find that the model has radically diverged. Interactive display of model and data is the final safeguard, to verify that the numerical metrics of variance are telling the truth, and that published model is comfortably within reason.

The Meisburger study of diffuse scattering from lysozyme (Meisburger et al., 2020) benefited from an uncrowded diffraction pattern where the separation between Bragg and diffuse features was readily accomplished. In future experiments, a combination of factors may contribute to more crowded and complex patterns; these factors include larger unit cells, smaller mosaic block sizes, and the use of disperse X-ray beams at XFEL sources, used in the pursuit of time-resolved data that are free from radiation damage. In this event, image simulation provides a potential avenue to deconvolute overlapping signals: either overlapping Bragg spots, or Bragg spots overlapping with diffuse halos. Deconvolution has been described in the literature (Bourgeois, 1999; Schreurs et al., 2009), but has not been common practice. However, in the intervening years computa–tional hardware has advanced such that GPU acceleration is widely avail–able. A fresh look at simulators is thus warranted, both to facilitate deconvolution and simply to provide the best physics–based profiles for data reduction to most effectively account for experimental uncertainty.

8. Software availability

Instructions and files for running *simtbx.sim_view* may be found at https://github.com/cctbx–xfel/erice_2022. Optimal performance requires a Linux machine accelerated with an Nvidia GPU; currently both Volta and Ampere architectures are supported. Instructions for installing and

running the software, plus a short program description, will be kept up to date and accessible in the aforementioned Github repository. Github makes provision for users to report bugs, request features, and contribute code.

Acknowledgements

References

Bourenkov, G. P., & Popov, A. N. (2006). A quantitative approach to data-collection strategies [Review]. *Acta Crystallographica. Section D, Biological Crystallography, 62*(1), 58–64. https://doi.org/10.1107/S0907444905033998

Bourgeois, D. (1999). New processing tools for weak and/or spatially overlapped macromolecular diffraction patterns. *Acta Crystallographica. Section D, Biological Crystallography, 55*(10), 1733–1741. https://doi.org/10.1107/s0907444999008355

Brehm, W., White, T., & Chapman, H. N. (2023). Crystal diffraction prediction and partiality estimation using Gaussian basis functions. *Acta Crystallographica Section A: Foundations and Advances, 79*(2), 145–162. https://doi.org/10.1107/S2053273323000682

Caspar, D. L. D., Clarage, J., Salunke, D. M., & Clarage, M. (1988). Liquid-like movements in crystalline insulin. *Nature, 332*, 659–662. https://doi.org/10.1038/332659a0

Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., ... Spence, J. C. (2011). Femtosecond X-ray protein nanocrystallography. *Nature, 470*(7332), 73–77. https://doi.org/10.1038/nature09750

Darwin, C. G. (1914). XXXIV. The theory of X-ray reflexion. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 27*(158), 315–333. https://doi.org/10.1080/14786440208635093

Diederichs, K. (2009). Simulation of X-ray frames from macromolecular crystals using a ray-tracing approach. *Acta Crystallographica. Section D, Biological Crystallography, 65*(6), 535–542. https://doi.org/10.1107/S0907444909010282

Ginn, H. M., Brewster, A. S., Hattne, J., Evans, G., Wagner, A., Grimes, J. M., ... Stuart, D. I. (2015). A revised partiality model and post-refinement algorithm for X-ray free-electron laser data. *Acta Crystallographica. Section D, Biological Crystallography, 71*(6), 1400–1410. https://doi.org/10.1107/s1399004715006902

Holton, J. M., Classen, S., Frankel, K. A., & Tainer, J. A. (2014). The R-factor gap in macromolecular crystallography: An untapped potential for insights on accurate structures. *The FEBS Journal, 281*(18), 4046–4060. https://doi.org/10.1111/febs.12922

Holton, J. M., & Frankel, K. A. (2010). The minimum crystal size needed for a complete diffraction data set. *Acta Crystallographica. Section D, Biological Crystallography, 66*(4), 393–408. https://doi.org/10.1107/S0907444910007262

Holton, J. M., Nielsen, C., & Frankel, K. A. (2012). The point-spread function of fiber-coupled area detectors. *Journal of Synchrotron Radiation, 19*(6), 1006–1011. https://doi.org/10.1107/S0909049512035571

Kabsch, W. (2010a). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica. Section D, Biological Crystallography, 66*(2), 133–144. https://doi.org/10.1107/S0907444909047374

Kabsch, W. (2010b). XDS. *Acta Crystallographica. Section D, Biological Crystallography, 66*(2), 125–132. https://doi.org/10.1107/S0907444909047337

Kabsch, W. (2014). Processing of X-ray snapshots from crystals in random orientations. *Acta Crystallographica. Section D, Biological Crystallography, 70*(8), 2204–2216. https://doi.org/10.1107/S1399004714013534

Kahn, R., Fourme, R., Gadet, A., Janin, J., Dumas, C., & André, D. (1982). Macromolecular crystallography with synchrotron radiation: Photographic data collection and polarization correction. *Journal of Applied Crystallography, 15*(3), 330–337. https://doi.org/10.1107/s0021889882012060

Keedy, D. A., Kenner, L. R., Warkentin, M., Woldeyes, R. A., Hopkins, J. B., Thompson, M. C., ... Fraser, J. S. (2015). Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *Elife, 4*, e07574. https://doi.org/10.7554/eLife.07574

Koerner, L. J., Philipp, H. T., Hromalik, M. S., Tate, M. W., & Gruner, S. M. (2009). X-ray tests of a pixel array detector for coherent x-ray imaging at the Linac Coherent Light Source. *Journal of Instrumentation, 4*(03), P03001. https://doi.org/10.1088/1748-0221/4/03/p03001

Kroon-Batenburg, L. M., Schreurs, A. M., Ravelli, R. B., & Gros, P. (2015). Accounting for partiality in serial crystallography using ray-tracing principles. *Acta Crystallographica. Section D, Biological Crystallography, 71*(9), 1799–1811. https://doi.org/10.1107/S1399004715011803

Leslie, A. G. W. (2006). The integration of macromolecular diffraction data. *Acta Crystallographica. Section D, Biological Crystallography, 62*(1), 48–57. https://doi.org/10.1107/S0907444905039107

McPhillips, T. M., McPhillips, S. E., Chiu, H. J., Cohen, A. E., Deacon, A. M., Ellis, P. J., ... Kuhn, P. (2002). Blu-Ice and the distributed control system: Software for data acquisition and instrument control at macromolecular crystallography beamlines. *Journal of Synchrotron Radiation, 9*(6), 401–406. https://doi.org/10.1107/s0909049502015170

Meisburger, S. P., Case, D. A., & Ando, N. (2020). Diffuse X-ray scattering from correlated motions in a protein crystal. *Nature Communications, 11*(1), 1271. https://doi.org/10.1038/s41467-020-14933-6

Meisburger, S. P., Thomas, W. C., Watkins, M. B., & Ando, N. (2017). X-ray scattering studies of protein structural dynamics. *Chemical Reviews, 117*(12), 7615–7672. https://doi.org/10.1021/acs.chemrev.6b00790

Mendez, D., Bolotovsky, R., Bhowmick, A., Brewster, A. S., Kern, J., Yano, J., ... Sauter, N. K. (2020). Beyond integration: Modeling every pixel to obtain better structure factors from stills. *IUCrJ, 7*(6), 1151–1167. https://doi.org/10.1107/S2052252520013007

Nave, C. (2014). Matching X-ray beam and detector properties to protein crystals of different perfection. *Journal of Synchrotron Radiation, 21*(3), 537–546. https://doi.org/10.1107/s1600577514003609

Nave, C., Sutton, G., Evans, G., Owen, R., Rau, C., Robinson, I., & Stuart, D. I. (2016). Imperfection and radiation damage in protein crystals studied with coherent radiation. *Journal of Synchrotron Radiation, 23*(1), 228–237. https://doi.org/10.1107/S1600577515019700

Peck, A., Poitevin, F., & Lane, T. J. (2018). Intermolecular correlations are necessary to explain diffuse scattering from protein crystals. *IUCrJ, 5*(2), 211–222. https://doi.org/10.1107/s2052252518001124

Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallographica. Section D, Biological Crystallography, 55*(10), 1718–1725. https://doi.org/10.1107/s090744499900935x

Philipp, H. T., Koerner, L. J., Hromalik, M. S., Tate, M. W., & Gruner, S. M. (2010). Femtosecond radiation experiment detector for X-ray free-electron laser (XFEL) coherent X-ray imaging. *IEEE Transactions on Nuclear Science, 57*(6), 3795–3799. https://doi.org/10.1109/TNS.2010.2085445

Sauter, N. K. (2015). XFEL diffraction: Developing processing methods to optimize data quality. *Journal of Synchrotron Radiation, 22*(2), 239–248. https://doi.org/10.1107/S1600577514028203

Sauter, N. K., & Adams, P. D. (2017). Overcoming data processing challenges for break-through crystallography. In U. Bergmann, V. K. Yachandra, & J. Yano (Eds.). *X-Ray free electron lasers: Applications in materials, chemistry and biology* (pp. 70–87). Royal Society of Chemistry. 〈https://doi.org/10.1039/9781782624097〉.

Sauter, N. K., Hattne, J., Brewster, A. S., Echols, N., Zwart, P. H., & Adams, P. D. (2014). Improved crystal orientation and physical properties from single-shot XFEL stills. *Acta Crystallographica. Section D, Biological Crystallography, 70*(12), 3299–3309. https://doi.org/10.1107/s1399004714024134

Sauter, N. K., Kern, J., Yano, J., & Holton, J. M. (2020). Towards the spatial resolution of metalloprotein charge states by detailed modeling of XFEL crystallographic diffraction. *Acta Crystallographica. Section D, Structural Biology, 76*(2), 176–192. https://doi.org/10.1107/S2059798320000418

Schreurs, A. M. M., Xian, X., & Kroon-Batenburg, L. M. J. (2009). EVAL15: A diffraction data integration method based on *ab initio* predicted profiles. *Journal of Applied Crystallography, 43*(1), 70–82. https://doi.org/10.1107/s0021889809043234

Uervirojnangkoorn, M., Zeldin, O. B., Lyubimov, A. Y., Hattne, J., Brewster, A. S., Sauter, N. K., ... Weis, W. I. (2015). Enabling X-ray free electron laser crystallography for challenging biological systems from a limited number of crystals. *Elife, 4*, e05421. https://doi.org/10.7554/eLife.05421

van den Bedem, H., Bhabha, G., Yang, K., Wright, P. E., & Fraser, J. S. (2013). Automated identification of functional dynamic contact networks from X-ray crystallography. *Nature Methods, 10*(9), 896–902. https://doi.org/10.1038/nmeth.2592

Wall, M. E., Clarage, J. B., & Phillips, G. N. J. (1997). Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering. *Structure (London, England: 1993), 5*(12), 1599–1612. https://doi.org/10.1016/s0969-2126(97)00308-0

Wall, M. E., Wolff, A. M., & Fraser, J. S. (2018). Bringing diffuse X-ray scattering into focus. *Current Opinion in Structural Biology, 50*, 109–116. https://doi.org/10.1016/j.sbi.2018.01.009

White, T. A. (2014). Post-refinement method for snapshot serial crystallography. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369*, 20130330. https://doi.org/10.1098/rstb.2013.0330

Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., ... Evans, G. (2018). DIALS: Implementation and evaluation of a new integration package. *Acta Crystallographica. Section D, Structural Biology, 74*(2), 85–97. https://doi.org/10.1107/S2059798317017235

Woldeyes, R. A., Sivak, D. A., & Fraser, J. S. (2014). *E pluribus unum*, no more: From one crystal, many conformations [Review]. *Current Opinion in Structural Biology, 28C*, 56–62. https://doi.org/10.1016/j.sbi.2014.07.005