



Cite this article: Parkhurst JM, Dumoux M, Basham M, Clare D, Siebert CA, Varslot T, Kirkland A, Naismith JH, Evans G. 2021 Parakeet: a digital twin software pipeline to assess the impact of experimental parameters on tomographic reconstructions for cryo-electron tomography. *Open Biol.* **11**: 210160. <https://doi.org/10.1098/rsob.210160>

Received: 28 May 2021

Accepted: 23 September 2021

Subject Area:

structural biology/molecular biology

Keywords:

digital twin, electron microscopy, tomography, multislice simulation

Author for correspondence:

Gwyndaf Evans

e-mail: gwyndaf.evans@diamond.ac.uk

Parakeet: a digital twin software pipeline to assess the impact of experimental parameters on tomographic reconstructions for cryo-electron tomography

James M. Parkhurst^{1,2}, Maud Dumoux¹, Mark Basham^{1,2}, Daniel Clare², C. Alistair Siebert², Trond Varslot⁴, Angus Kirkland^{1,3,5}, James H. Naismith^{1,6} and Gwyndaf Evans^{1,2}

¹Rosalind Franklin Institute, Harwell Science and Innovation Campus, Didcot OX11 0FA, UK

²Diamond Light Source, and ³Electron Physical Science Imaging Centre, Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK

⁴Thermo Fisher Scientific, Vlastimila Pecha, Brno, Czech Republic

⁵Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, UK

⁶Division of Structural Biology, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

JMP, 0000-0002-9120-8354; MD, 0000-0002-1732-1041; MB, 0000-0002-8438-1415; DC, 0000-0002-3372-8378; CAS, 0000-0002-8126-1979; TV, 0000-0002-7301-6404; AK, 0000-0001-8068-1990; JHN, 0000-0001-6744-5061; GE, 0000-0002-6079-2201

In cryo-electron tomography (cryo-ET) of biological samples, the quality of tomographic reconstructions can vary depending on the transmission electron microscope (TEM) instrument and data acquisition parameters. In this paper, we present Parakeet, a ‘digital twin’ software pipeline for the assessment of the impact of various TEM experiment parameters on the quality of three-dimensional tomographic reconstructions. The Parakeet digital twin is a digital model that can be used to optimize the performance and utilization of a physical instrument to enable *in silico* optimization of sample geometries, data acquisition schemes and instrument parameters. The digital twin performs virtual sample generation, TEM image simulation, and tilt series reconstruction and analysis within a convenient software framework. As well as being able to produce physically realistic simulated cryo-ET datasets to aid the development of tomographic reconstruction and subtomogram averaging programs, Parakeet aims to enable convenient assessment of the effects of different microscope parameters and data acquisition parameters on reconstruction quality. To illustrate the use of the software, we present the example of a quantitative analysis of missing wedge artefacts on simulated planar and cylindrical biological samples and discuss how data collection parameters can be modified for cylindrical samples where a full 180° tilt range might be measured.

1. Introduction

In cryo-electron tomography (cryo-ET), a series of projection images of a sample obtained using a transmission electron microscope (TEM) are recorded at different angles through the sample. From the Fourier projection slice theorem [1,2], a single projection image in real space is equivalent to a single slice through Fourier space. Therefore, by filling three-dimensional Fourier space using projections from many angles, a tomographic reconstruction can be recovered. The quality of the tomographic reconstructions in cryo-ET of biological samples can vary

greatly depending on the TEM instrument specifications and experimental data acquisition parameters used; consequently, a significant amount of development time is dedicated to increasing the quality and efficiency of imaging acquisition [3–6], improving data acquisition software [7], and refining data analysis and reconstruction software [8]. For this reason, it can be helpful for developers to first implement new ideas and proposed developments within a ‘digital twin’ of the TEM to enable offline optimization of sample characteristics, data acquisition schemes and instrument parameters. For example, in a digital twin of the TEM, it is trivial to vary sample characteristics (such as the sample thickness, composition and sensitivity to beam damage), instrument parameters (such as electron energy, energy spread, microscope aberrations, and the presence of a phase plate) and data acquisition parameters (such as electron dose and sample tilt range). While some of these experimental parameters may also be easily varied in a physical experiment, modifying or controlling for others may be somewhat more challenging or time consuming.

In its most general form, a digital twin is a digital model representing a physical object that can be used to optimize the performance and utilization of the physical object. In the specific case of cryo-ET, a digital twin can be used to determine optimal data acquisition schemes and instrument parameters and provide phantom test datasets that can be used in the development of data processing software. This is increasingly important when the instrument is of a new design and representative data is not available from a real instrument. First the digital twin must implement a physically realistic model of the sample used in the TEM. In the case of cryo-ET of biological specimens, this sample may be represented by a collection of biological macromolecules embedded in a large volume of amorphous ice [9], typically with a planar geometry. Next, the digital twin must be able to accurately generate physically realistic TEM images of the sample and offer the same range of instrument and data acquisition parameter values. In this case, TEM simulations are a key component in the creation of the digital twin modelling a physically realistic microscope, sample and detector. In EM, the physics of image formation are well understood, and realistic images are routinely simulated using the multislice algorithm [10,11]. As well as assisting in the interpretation of experimentally acquired cryo-EM images, simulations have been used to provide insight into possible experimental limitations of different data acquisition schemes by performing *in silico* experiments that sample the available space of data acquisition parameters. For example, Zhang *et al.* [12] used simulated EM images with an ideal phase plate to show that the structure of hen egg white lysozyme could in principle be solved via cryo-electron microscopy (cryo-EM), and Palmer & Löwe [13] used simulations to assess the effects of missing wedge artefacts in cryo-ET reconstructions.

Finally, in order to use the simulated EM images to evaluate the effects of varying different instrument parameters, the digital twin must be able to produce three-dimensional reconstructions of the object of interest from the simulated images using standard algorithms. The digital twin must then have the ability to compare the reconstructed object with the reference object to determine the overall quality of reconstruction. In cryo-ET, the reconstruction quality is most commonly assessed using the Fourier shell correlation

(FSC) [14]. Comparing the reconstructed volume, derived from the simulated images, with the original reference model ‘closes the loop’ and enables optimization of sample preparation, instrument and data acquisition parameters for the physical instrument.

In this paper, we present a digital twin software pipeline, Parakeet (Program for Analysis and Reconstruction of Artificial data for Kryo ElEctron Tomography), to analyse the impact of varying data acquisition and analysis parameters on reconstruction quality for cryo-ET. A fundamental problem in tomography is the missing wedge resulting from the use of a limited range of tilt angles. Due to physical constraints, such as the limited space between the objective lens pole pieces and the dimensions of the specimen holder [15], tilt stages in current generations of electron microscopes tend to have a limited rotation range of $\pm 70^\circ$ [16–18]; indeed, in practice, cryo-ET datasets are collected with a tilt range of around $\pm 60^\circ$ [19]. New sample preparation, hardware and software developments in the field, such as the use of cylindrical specimens [13], dual-axis data acquisition schemes [4,20–23] and advanced reconstruction algorithms [24–28], have focussed on trying to alleviate this problem. To illustrate the use of Parakeet, we apply it to the problem of quantifying the impact of the missing wedge on reconstruction quality.

2. Digital twin

2.1. Design overview

The core aim of Parakeet is to enable convenient assessment of the effects of varying data acquisition and analysis parameters on the quality of tomographic reconstructions of biological samples. The workflow in the digital twin can be decomposed into discrete tasks that exchange information through data files as shown in figure 1. The first step is to describe the microscope model and the desired data acquisition strategy. A virtual sample is then generated with the desired geometry which contains a given number of particles with either specified or random positions and orientations embedded in amorphous ice. To simulate a tilt series of images, multislice TEM simulations [10,11] of biological macromolecules are performed. These simulations model the propagation and scattering of an electron wave through the representation of the atomic model of a specimen. The microscope optics and detector response are then introduced to yield the final simulated image. By rotating the virtual sample, a tilt series can be simulated from the atomic model. Once a simulated tilt series has been acquired, the data analysis component of the digital twin is used to apply a CTF correction to each of the projections in the tilt series. Finally, the projections are reconstructed using standard tomographic reconstruction algorithms and, to ‘close the loop’, the reconstructions are compared with the original known reference atomic model, providing an objective and quantitative assessment of the quality of reconstruction under different simulated data collection strategies. The specimen model is stored using HDF5 (<https://www.hdfgroup.org/hdf5>), and the simulated images can be exported into either HDF5 format or MRC format [29] using the `mrcfile` python library [30] with the correct FEI extended header information present in order to provide

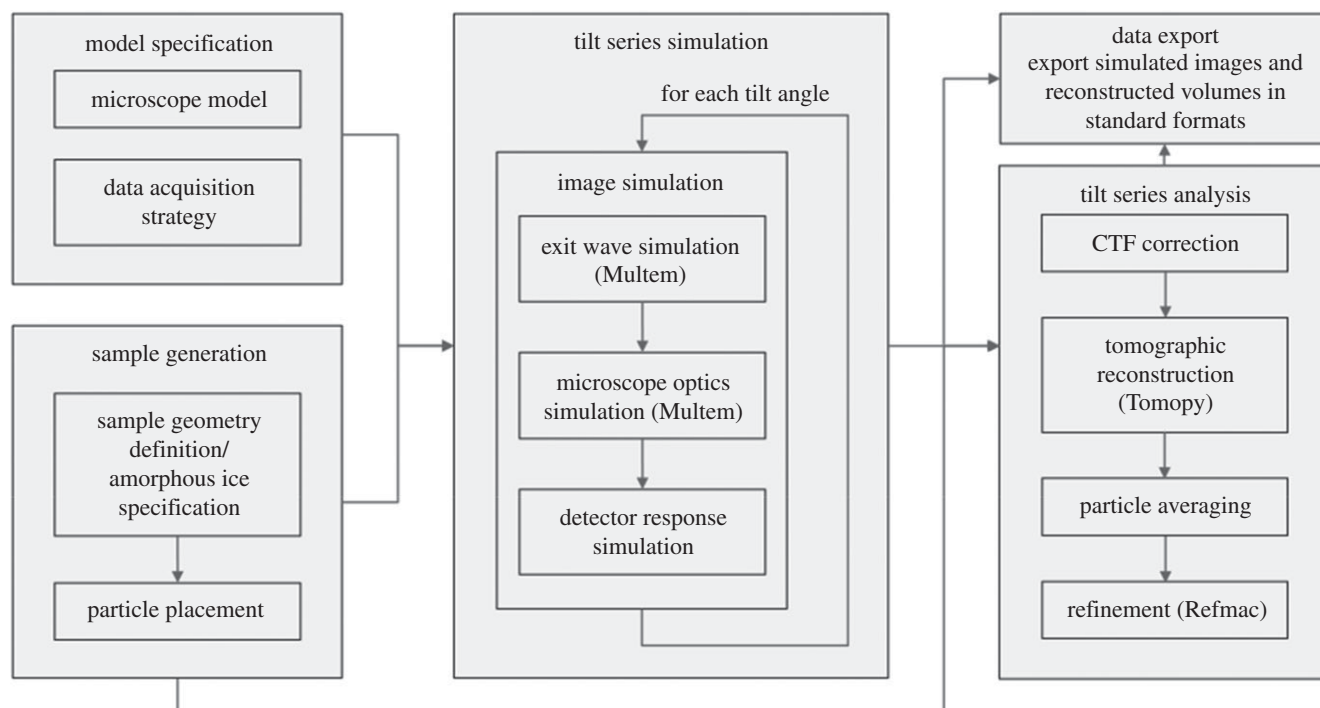


Figure 1. Flow chart describing the Parakeet digital twin software pipeline.

metadata compatible with other data analysis software. Therefore, Parakeet can be used to provide simulated test data that can be readily imported into external software programs to enable, for example, optimization of tomographic reconstruction and subtomogram averaging (STA) software. Reconstructed volumes are exported in MRC format. The software is open source, written in a combination of C++ and Python, and can be obtained online from the Rosalind Franklin Institute GitHub repository [31].

2.2. Model specification

The microscope model is built from a set of models representing the beam, detector and objective lens, the data acquisition scheme is represented by a scan model, and the specimen is represented by a sample model. These models are specified via a YAML (<https://yaml.org/>) configuration file which provides a convenient metadata serialization format that is both human readable and supported by a wide variety of programming languages.

- *Beam model.* This configures the beam energy, energy spread, source spread and electron flux. It also allows the beam drift to be configured as a function of tilt angle using either a random or sinusoidal model.
- *Detector model.* This configures the number and size of the pixels and configures the detective quantum efficiency (DQE) model.
- *Lens model.* This specifies the objective lens aberrations such as the defocus, spherical aberration and chromatic aberration. It also allows a phase plate to be configured via an arbitrary phase shift.
- *Scan model.* This specifies the axis of rotation, the exposure time for each image, the number of images, and the angle and translation between each image in the tilt series.

- *Sample model.* This specifies the geometry of the sample which can be either a plane or a cylinder, configures the amorphous ice model and defines the number, type, positions and orientations of particles in the sample.

2.3. Sample generation

The Protein Data Bank (PDB) [32] contains a repository of thousands of biological macromolecules which can be readily accessed to act as a foundation for building a digital twin of the sample. After first defining the overall shape and size of the sample—either planar or cylindrical—the number and type of particles (given by their PDB IDs) to include in the sample volume are then specified. If a single particle is specified, then by default the particle is positioned in the centre of the sample volume and field of view. If multiple particles are specified, they can either be positioned with predefined positions and orientations (provided these positions do not overlap), or they can be assigned random positions and orientations within the sample volume. Particles are only allowed within the defined sample volume and are not allowed to overlap with the surface of the sample volume. Once the sample volume is defined, the atomic model (which may be composed of many particles of different types and orientations) can be ‘milled’ into arbitrary shapes by removing atoms outside defined regions.

An important aspect of the atomic model of the sample is the amorphous ice component, which by volume may account for a substantial portion of the atoms in the model depending on the thickness of the sample. The digital twin implements two methods for the generation of the amorphous ice within the sample. The first is to simply generate water molecules with random positions and orientations within the sample volume. In this model, water molecules may be placed at unphysical distances relative to one another. An alternative, improved fully atomic model of the amorphous ice, would require molecular dynamics simulations to relax the water molecules to physically realistic distances;

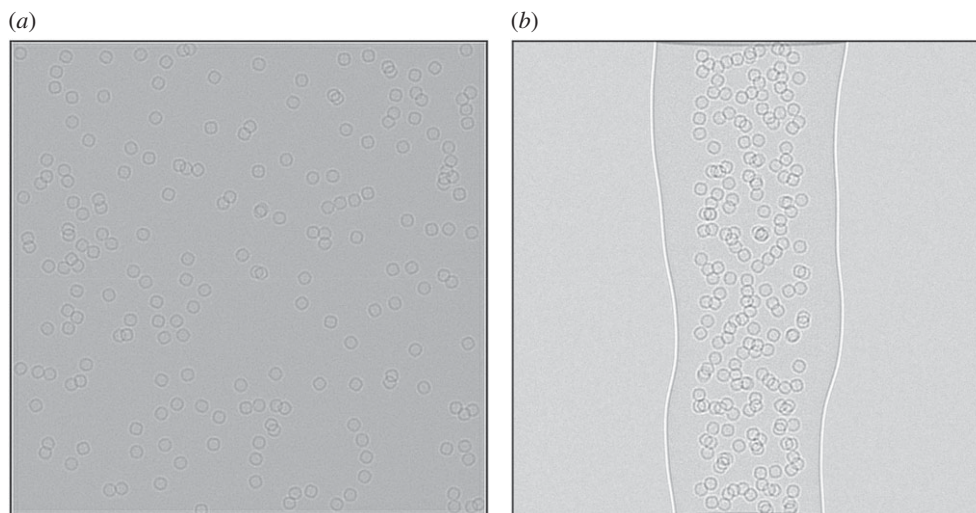


Figure 2. (a) A large planar lamella sample of size $4000 \text{ \AA} \times 4000 \text{ \AA} \times 1500 \text{ \AA}$ containing 200 apoferritin particles. (b) A cylindrical volume with diameter 1500 \AA with 'natural' edges containing 200 apoferritin particles.

however, this is very computationally intensive for the sample volumes considered here and is not discussed. The digital twin also implements a continuum approach using a Gaussian random field (GRF) model for the atomic potential of the amorphous ice in which the ice component is modelled as Fourier filtered noise with a given power spectrum to match the expected correlations for a physically realistic ice model. The GRF approach is more computationally efficient than the atom-based approach. Figure 2 shows a planar and cylindrical sample generated using the GRF approach. The model can also be defined to have 'natural' shapes which deviate from the ideal shape of a plane or cylinder. For example, in the case of the cylinder, the radius is parameterized along the length of the cylinder to allow different radii and offsets which are then interpolated using cubic splines.

2.4. Tilt series simulation

In order to simulate the TEM images, Parakeet uses the MULTTEM library [33] which provides a GPU accelerated implementation of the multislice algorithm [10,11] and a model for the microscope optics. These algorithms were extended and wrapped using the *Pybind11* C++/Python binding package [34] to create a simple Python API (*python-multem*). The python bindings are open source and can be obtained from the Rosalind Franklin Institute GitHub repository (<https://github.com/rosalindfranklininstitute/python-multem>).

The multislice algorithm operates by taking the atomic model and dividing the sample into slices of a given thickness along the direction of the electron beam. The thickness of each slice is typically a few angstroms such that each slice can be considered as a weak phase object; in the simulations used here, the slice thickness is 5 \AA . Each sample slice is then treated as infinitesimally thin, and for each slice the atomic potential is calculated as the sum of the atomic potentials of the constituent atoms projected onto the infinitesimally thin slice. The calculation for each slice is performed on a discrete grid with a pixel size of $1 \text{ \AA} \times 1 \text{ \AA}$. The wave function is then transmitted through the slice and propagated to the next slice via Fresnel diffraction.

Propagating the wave function through the whole sample gives a complex wave function at the exit surface of the sample.

As described in appendix B, beam damage to the sample is implemented by convolving the electrostatic potential of the specimen with a Gaussian function [12] whose variance, σ_B^2 , can be related to an isotropic B factor by $B = 8\pi^2\sigma_B^2$ [35]. This convolution is conveniently performed in Fourier space where the B factor filter is expressed as a function of spatial frequency, q . The potential from a damaged specimen, V , can then be related to the undamaged potential, V_0 by $V = F^{-1}[F[V_0] \exp(-q^2 B/4)]$. The isotropic B factor is parameterized as a linear function of the total accumulated incident electron dose, D_E , and is given by $B = 8\pi^2 D_E S_E$, where S_E is the sensitivity coefficient. The value of this coefficient is sample dependent but can be calibrated from the results of X-ray diffraction experiments [36–38], electron diffraction experiments [39] or single-particle cryo-EM experiments [40] and typically takes values between 0.020 and 0.090.

The effects of the microscope optics are modelled by application of a contrast transfer function (CTF), an oscillating complex function of the aberrations in Fourier space, which applies a frequency-dependent phase shift to the exit wave. Spatial and temporal coherence envelopes, which are applicable in the case of the linear imaging approximation, are included which damp the CTF at high resolution. Finally, the detector response, in the form of a frequency-dependent DQE and Poisson counting noise for the number of expected electrons per pixel are added to generate the final simulated images. Typical values for the parameters used in the simulations are summarized in table 1; our choices were guided by what is practical using a Thermo Fisher Scientific Titan Krios instrument [41].

2.5. Tilt series analysis

To obtain high-resolution reconstructions, CTF correction [43] or exit wave reconstruction from multiple images [44] must be performed. This is typically done either by phase-flipping, multiplying the Fourier transform of the image by the CTF, or through the use of a Wiener filter [43]. When a single

Table 1. Typical parameters used in the TEM simulations, chosen to be consistent with parameters for a Thermo Fisher Scientific Titan Krios TEM [41].

parameter	description	value
E	energy	300 keV
Δf	defocus	2.5 μm
C_s	spherical aberration	2.7 mm
C_c	chromatic aberration	2.7 mm
$\Delta I/I$	current spread	0.33 ppm
$\Delta V/V$	voltage spread	0.80 ppm
ΔE	energy spread	0.8 eV
θ_c	source spread	0.1 mrad
d_{px}	pixel size	1 \AA
t_s	multislice z-slice thickness	5 \AA
	potential approximation	Lobato <i>et al.</i> [42]
S_E	beam damage sensitivity coefficient	0.022 $\text{\AA}^2/\text{e}^-$

defocus is used, only the contrast inversions can be corrected for. In this case, a Wiener filter may amplify noise at spatial frequencies close to the zero crossings of the CTF and the Fourier components at the zero crossings themselves cannot be corrected at all. In phase-flipping, the contrast inversions are corrected by multiplying the Fourier transform of the image by the sign of the CTF. Phase-flipping has been shown to result in marginally better reconstructions than simply multiplying by the CTF [43] and is therefore implemented in Parakeet.

For two-dimensional single-particle analysis (SPA), where untilted images of thin specimens are used, a single CTF correction can be applied to each image. However, in the case of tomography, when a planar sample is tilted, different parts of the sample will appear at different defoci [45] and this defocus gradient can become significant at high tilt angles. Additionally, for thick samples, different voxels within a sample have different defocus values. In this case, CTF correction using a single defocus value is insufficient and a three-dimensional CTF correction is required. This can be implemented by computing multiple CTF corrected images for each tilt angle and then using the appropriate corrected image pixel for each voxel in the reconstruction according to the location of the voxel within the volume [46]. However, in order to use multiple CTF corrected images per tilt rather than a single image per tilt, this approach requires the core of the standard tomographic reconstruction algorithms to be extended. Alternatively, if the dataset is to be used for STA [47], CTF correction can be applied to reconstruct each particle independently before averaging the individual subtomograms; this is the approach used here, as it permits the use of standard tomographic reconstruction algorithms.

In Parakeet, the CTF corrected tomographic tilt series are reconstructed using a GPU accelerated weighted back projection (WBP) algorithm implemented using the Astra toolbox [48] through the Tomopy python package [49,50]. This is a more robust variant of the filtered back projection algorithm which applies a unique filter to each projection based on the total number of projections and their rotational distribution,

rather than applying the same ramp filter to all projections [51,52]. Perhaps the most widely used of the alternatives to the back projection family of algorithms is the simultaneous iterative reconstruction technique (SIRT) algorithm which has been shown to give better-quality reconstructions than WBP in the presence of missing and low SNR data (both features of cryo-ET) [25,53,54]. However, SIRT requires careful selection of the number of iterations; although each iteration increases the agreement between the forward projection of the reconstruction with the observed images, this comes with a concomitant increase in noise features and thus a compromise must be found [25]. Other advanced algorithms using expectation maximization [55] and other iterative approaches [56] have also been reported. The WBP algorithm is implemented here since it is used in many cryo-ET reconstructions and requires no additional parameter optimization for each tilt series, making it appropriate for use in an automated pipeline.

Tomopy is used because it gives access to various reconstruction algorithms that may not be implemented in standard cryo-ET packages; it therefore allows greater flexibility in evaluating different reconstruction strategies. Additionally, since the package is written in Python, it can be seamlessly incorporated into data processing scripts for automation and analysis in a way that is more difficult for packages lacking an accessible API [57]. Finally, individual particle reconstructions are manually picked from the reconstructed tomograms using the known simulated particle positions and orientations; the final reconstruction is obtained by taking the Fourier transforms of the individual particle reconstructions and averaging them while applying a missing wedge mask to each particle in Fourier space.

As the images are simulated from a known atomic model, it is straightforward to fit the original atomic model back onto the reconstructed map. In order to do this, refinement of the atomic model is performed using REFMAC5 [58] with rigid body restraints. By measuring the fit of model to map, REFMAC5 provides an objective analysis of the quality of different reconstructions. We use the FSC average [14,59] which is defined by

$$\text{FSC}_{\text{average}} = \frac{\sum_{i=1}^{N_{\text{shell}}} N_i \text{FSC}_i}{\sum_{i=1}^{N_{\text{shell}}} N_i},$$

where N_i is the number of elements in a shell and FSC_i is the FSC within a shell, i .

3. Application example

The missing wedge problem has been the topic of numerous studies and efforts aimed at minimizing or correcting its deleterious effects on the quality of the tomographic reconstruction [13,60–62]. Here, Parakeet is used to quantify the impact of the missing wedge with reference to data without a missing wedge and consider the hypothetical scenario of using a cylindrical sample with a tilt stage that offers full rotation capability.

3.1. The missing wedge problem

The missing wedge problem is used as an example because it is one of the most pressing issues in the field of cryo-ET with complex data collection strategies and new sample

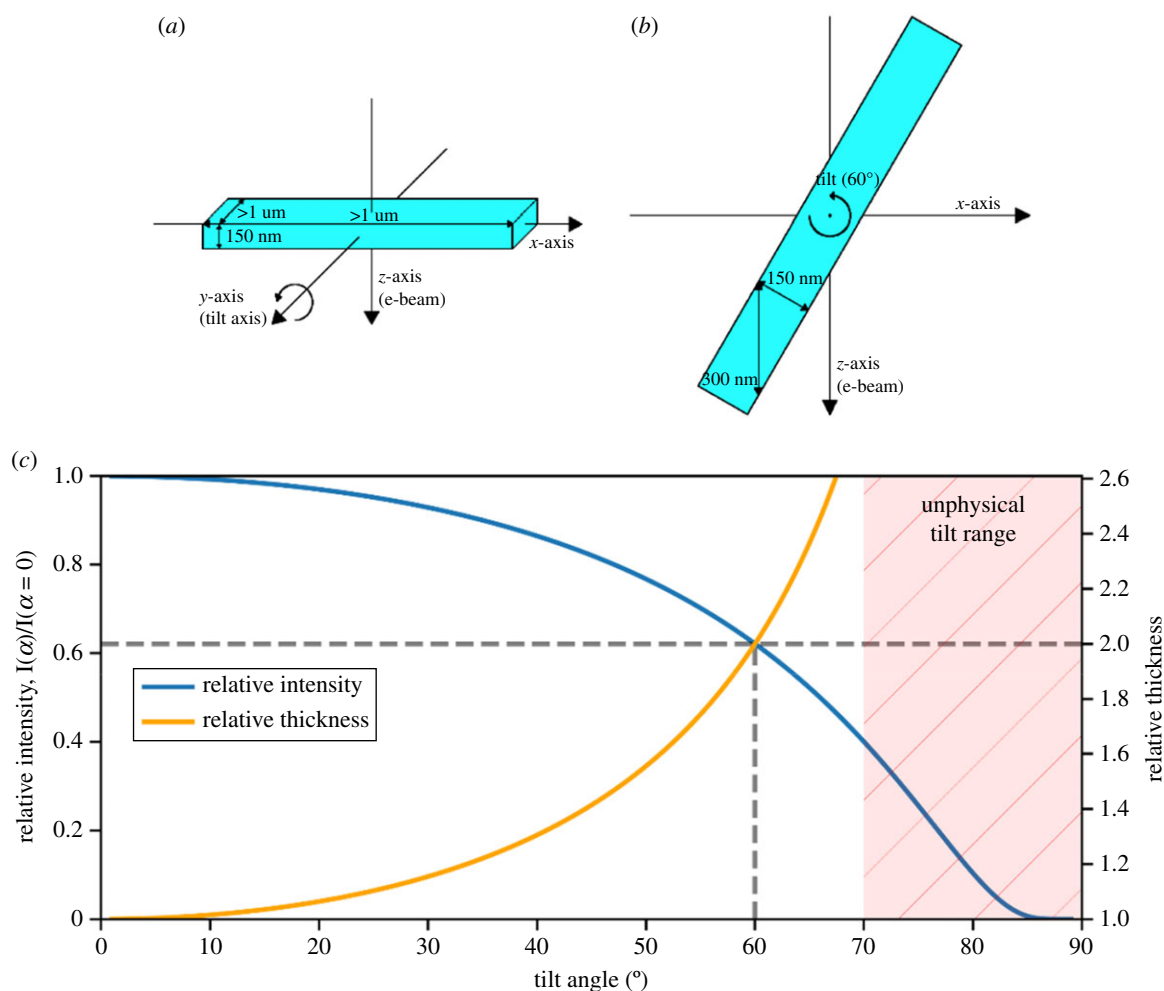


Figure 3. The lamella geometry. (a) Coordinate system and typical sizes of the lamella along with each axis at zero tilt. (b) Lamella tilted to 60° with an effective thickness along with the beam direction of twice the true thickness. (c) Relative thickness (orange, right axis) increasing for a planar sample at high tilt angle as $D/\lambda = 1/\cos(\alpha)$; the relative intensity (blue, left axis) as a function of tilt angle follows $I/I_0 = \exp(-(1 - 1/\cos(\alpha)) D_0/\lambda)$, plotted here for $D_0 = 150$ nm and $\lambda = 314$ nm [39]. The unphysical tilt region (for a lamella) is shown by the red hatched area. Relative thickness and intensity at a tilt angle of 60° are shown by the dashed lines.

preparation, hardware and software developments focused on alleviating it. In order to avoid artefacts relating to missing data, when reconstructing a three-dimensional object, the amplitudes across all spatial frequencies (to the desired resolution) need to be sampled. For objects with no symmetry, this requires a set of projections to be acquired over a full 180° rotation range. Where fewer data are recorded, spatial frequencies in some regions of Fourier space will be absent and the real space reconstruction will contain artefacts. In other words, the missing wedge problem arises when an entire region of Fourier space has not been sampled due to the use of a restricted rotation range. The missing wedge problem results in the introduction of artefacts, such as halos and streaking around objects with high contrast [13], anisotropic resolution in the reconstructed volume [16,63] and an incomplete three-dimensional representation of the specimen [47].

In the typical cryo-ET experiment, frozen-hydrated samples are prepared as planar lamellae which are 'thin' along the incident beam direction (at zero tilt) and 'thick' orthogonal to this [19], as shown in figure 3. Typically, cryo-ET datasets are collected with a tilt range of around $\pm 60^\circ$ [19], corresponding to a missing wedge of 60° . However, planar lamellae introduce additional complications because, as the planar lamella is tilted, its effective thickness along

with the beam direction increases at high tilt angles resulting in chromatic blurring due to the increased inelastic scattering [60]. An energy filter can be used to increase the signal to noise by removing the inelastically scattered electrons from the image; however, this results in a reduction in the intensity of the overall signal which becomes increasingly severe as more electrons are scattered inelastically and no longer contribute to the signal. A proposed solution to this problem is to use cylindrical samples which allow a full 180° tilt range. Such samples have been used in soft X-ray tomography [62], and in EM in materials science [60], with reconstructions showing improved reconstruction quality and isotropic resolution [61]. Palmer & Löwe [13] describe the use of a cylindrical specimen holder for cryo-ET of biological samples; however, the preparation of such samples has so far proved to be difficult and in order to become a useful technique, problems associated with the sample preparation must be addressed. For large cellular samples, high-pressure freezing can be used in the sample preparation to ensure vitrification [64,65] and a promising approach would be to then use a focussed ion beam to 'mill' these samples to achieve the desired cylindrical geometry [66]; this has already been demonstrated for radiation-resistant samples [23,26,61,67,68]. Once this becomes routinely possible, it is likely that on-axis tomography with cylindrical samples

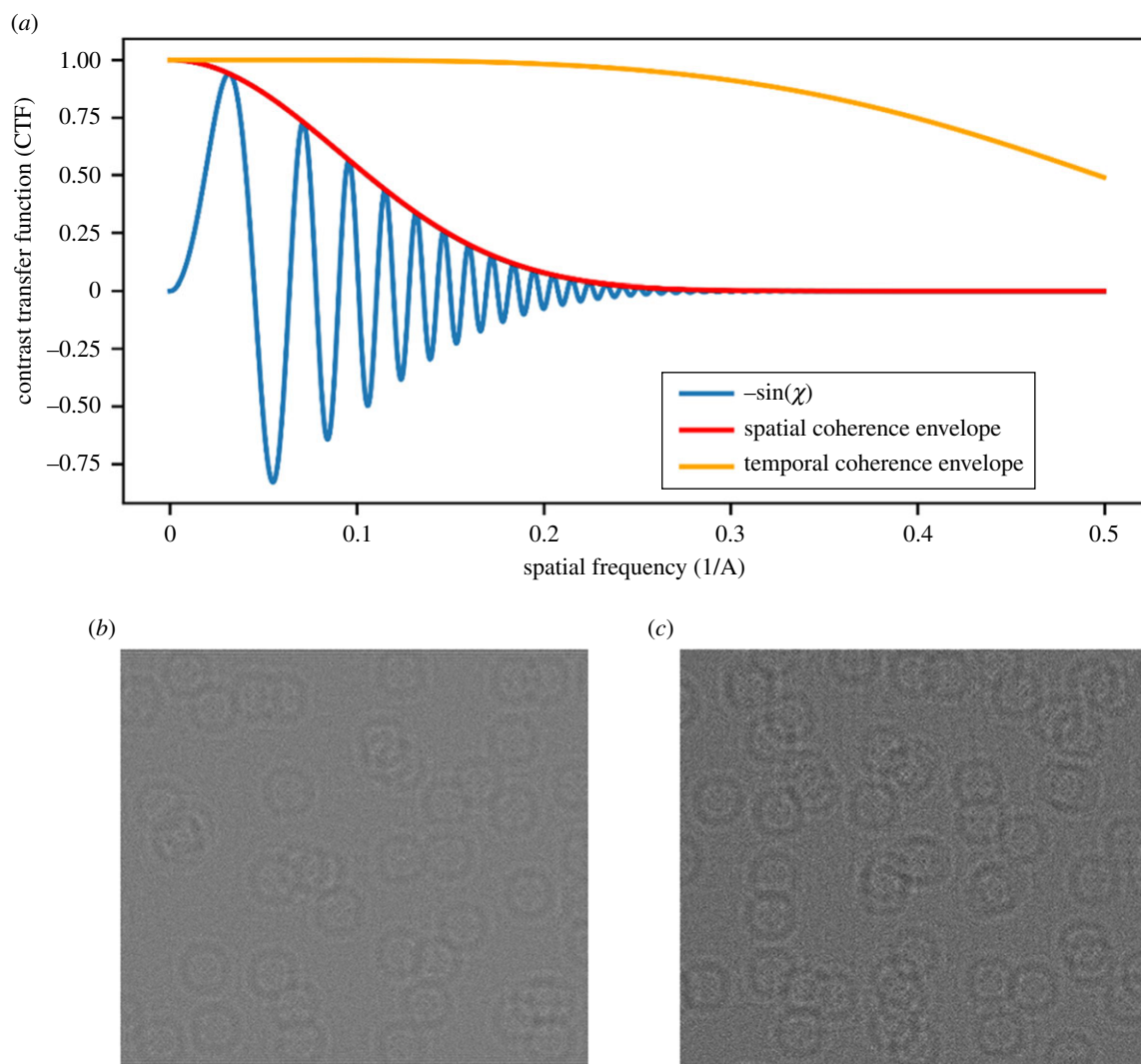


Figure 4. (a) Imaginary component of the CTF using the parameters in table 1. (b,c) Cropped 100×100 nm sections of the zero-tilt image for (b) a planar sample and (c) a cylinder sample containing apoferritin molecules where every particle has a random position and the same known preferred orientation.

will quickly become the method of choice for high-resolution cryo-ET data acquisition. Therefore, it is instructive to use the digital twin to determine what gains can be obtained in the quality of reconstruction for biological macromolecules using cylindrical samples rather than planar samples in the absence of a missing wedge artefact in order to provide motivation for the development of new advanced sample preparation techniques.

3.2. Simulated data

For the simulations reported here, a model of apoferritin and a ribosome available from the PDB [32] were used. The PDB entry for apoferritin, 6Z6U [69], is resolved to 1.25 \AA and was determined by cryo-EM SPA. It has a total structural weight of 511.09 kDa and 38 846 atoms in the model. The PDB entry for the ribosome, 4V5D [70], is resolved to 3.50 \AA and was determined by X-ray diffraction. It has a total structural weight of 4516.21 kDa and 296 042 atoms in the model.

In general, for single-particle samples, given the time-consuming sample preparation and data acquisition required for tomography, it is often more convenient to use standard SPA approaches rather than tomography. Cryo-ET with cylindrical samples is hence likely to provide the greatest benefit to the analysis of more complex cellular samples. However,

single-particle samples are convenient for the purpose of evaluating different data acquisition methodologies in cryo-ET, both experimentally and through simulation, since they provide a straightforward method to assess the quality of reconstructions from different schemes through the correlation between the reconstruction and the known structure. Apoferritin was chosen because it has a high degree of symmetry which allows the quality of the reconstructions along with different directions to be compared. Second, its alpha helices can be identified even in low-resolution maps which aids in the qualitative assessment of the reconstruction quality. Apoferritin is also commonly used as a test sample for cryo-EM applications [71]. Ribosome was used because it has characteristics complementary to apoferritin in that it is a larger structure with no structural symmetry.

The TEM parameters used in the simulations are shown in table 1. The imaginary component of the CTF given these parameters, including the spatial and temporal coherence envelopes, is shown in figure 4 along with a cropped 100×100 nm section of a simulated image for both the planar and cylindrical samples. The simulated planar samples had x -widths of 1000 nm and z -depths of 100 nm, 150 nm and 200 nm. The simulated cylindrical samples had diameters of 100 nm, 150 nm and 200 nm. Each samples contained 15 k randomly positioned molecules each with the same pre-determined orientation; for the apoferritin

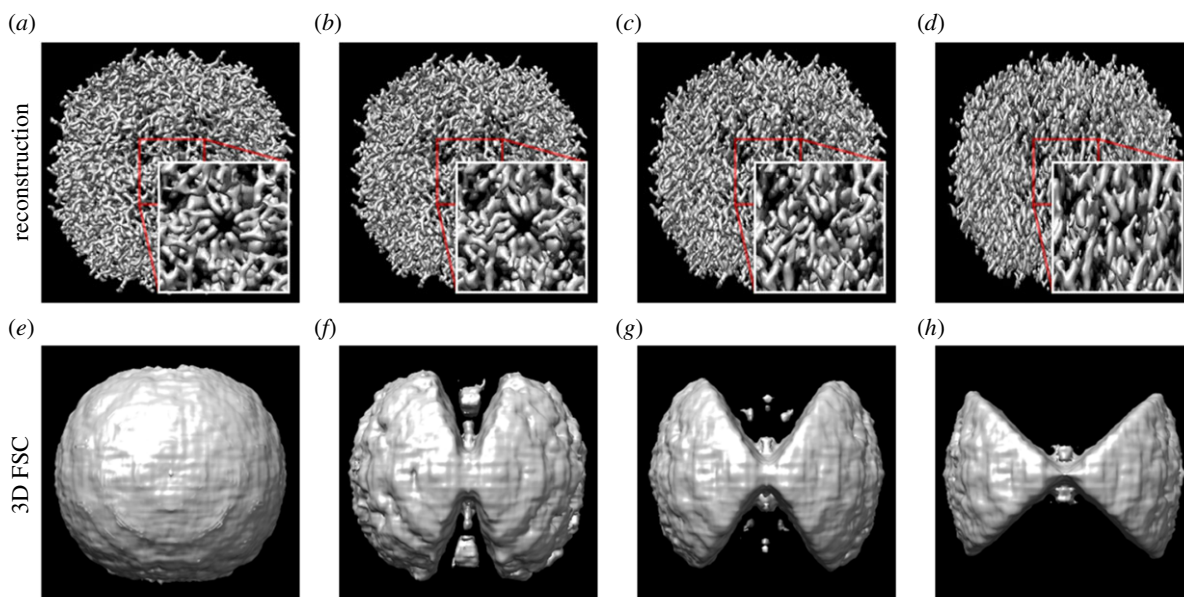


Figure 5. Reconstructions of a simulated apoferritin molecule from tilt series with a maximum tilt angle of (a) $\pm 90^\circ$, (b) $\pm 75^\circ$, (c) $\pm 60^\circ$ and (d) $\pm 45^\circ$. The features along with the z-axis are progressively blurred as the size of the missing wedge increases. The local 3D FSC for the reconstruction with the known structure at a FSC threshold level of 0.85 for (e) $\pm 90^\circ$, (f) $\pm 75^\circ$, (g) $\pm 60^\circ$ and (h) $\pm 45^\circ$.

simulations, the particles were oriented such that a symmetry axis was aligned with the sample to highlight the effect of the missing wedge on reconstructions. Although having only one orientation is not necessarily representative of all physical samples, some of which may have a preferred particle orientation but most of which will include particles with a variety of orientations, the inclusion of random orientations would effectively reduce the effects of the missing wedge on the reconstructions. The simulated projections have an isotropic pixel size of 1 Å.

3.3. Variation of reconstruction quality for different simulated data collection regimes

Apoferritin reconstructions were performed using particles embedded within a cylindrical sample with no missing wedge and missing wedges of 30° , 60° and 90° (figure 5). As expected, the quality of the reconstruction decreases as the missing wedge increases and most notably features along with the z-axis are blurred. The local three-dimensional FSC shows that the correlation between the true map and reconstructed map also decreases with an increasing missing wedge, giving an objective and quantitative assessment of the extent of degradation in the reconstruction.

Keeping the total simulated electron dose constant, we analysed planar and cylindrical samples with 90 projections per tilt series and separately with fixed 2° tilt increments. The data were simulated and analysed for a range of maximum tilt angles from 30° to 90° for sample thicknesses of 100 nm, 150 nm and 200 nm. In each case, the simulated data acquisition was performed using a standard dose symmetric data acquisition scheme where the projections were acquired in order of absolute tilt angle with projections at low tilt angles being acquired first [3]. Keeping the number of projections constant over different tilt ranges allows the tilt increment to vary. In the second approach, the tilt increment is kept constant, allowing the number of projections to vary depending on the overall range of tilt angles. Hence

these experiments interrogate which of two differential samplings of the three-dimensional object provides the highest quality reconstruction as assessed using the FSC average, FSC in high- and low-resolution bins and anisotropic FSC average in the three reciprocal space planes.

The FSC average and FSC in high- and low-resolution bins were plotted as a function of the maximum tilt angle for the planar and cylindrical apoferritin and ribosome samples for a constant number of projections and constant tilt increment (figure 6). In addition to the overall FSC average, the FSC average along the x/y , x/z and y/z planes (figure 7) was evaluated in order to probe the anisotropic effect of the missing wedge artefacts. As expected, the quality of the reconstruction, as assessed by the FSC average, improves as the maximum tilt angle increases. However, for a planar sample, images at a high tilt angle would be expected to have poorer SNR than images at a low tilt angle due to the increased projected thickness at a high tilt angle. With poorer SNR, the images taken at high tilt angles will contribute less information thus amplifying the missing wedge problem [13]. We observed that the quality of the reconstructions for planar samples did not increase beyond a maximum tilt angle of around 70° , corresponding to the typical physical limits of most instruments. For a planar sample, over the range of sample thicknesses considered here, at a 70° tilt, the deleterious effect of the increased apparent sample thickness (figure 1) outweighed the benefit of reducing the missing wedge. Our simulations in fact understate the problem, since the effects of the sample holder in a physical experiment will further degrade very high tilt measurements by shadowing.

As expected for the cylindrical sample, with a constant apparent thickness as a function of rotation, the quality of the reconstruction increases as the maximum tilt angle is increased to 90° and as expected outperforms the planar sample. For the cylindrical sample, the FSC average in the reciprocal x/y plane is relatively insensitive to the size of the missing wedge; however, the reciprocal x/z and reciprocal y/z planes are very sensitive to the extent of the

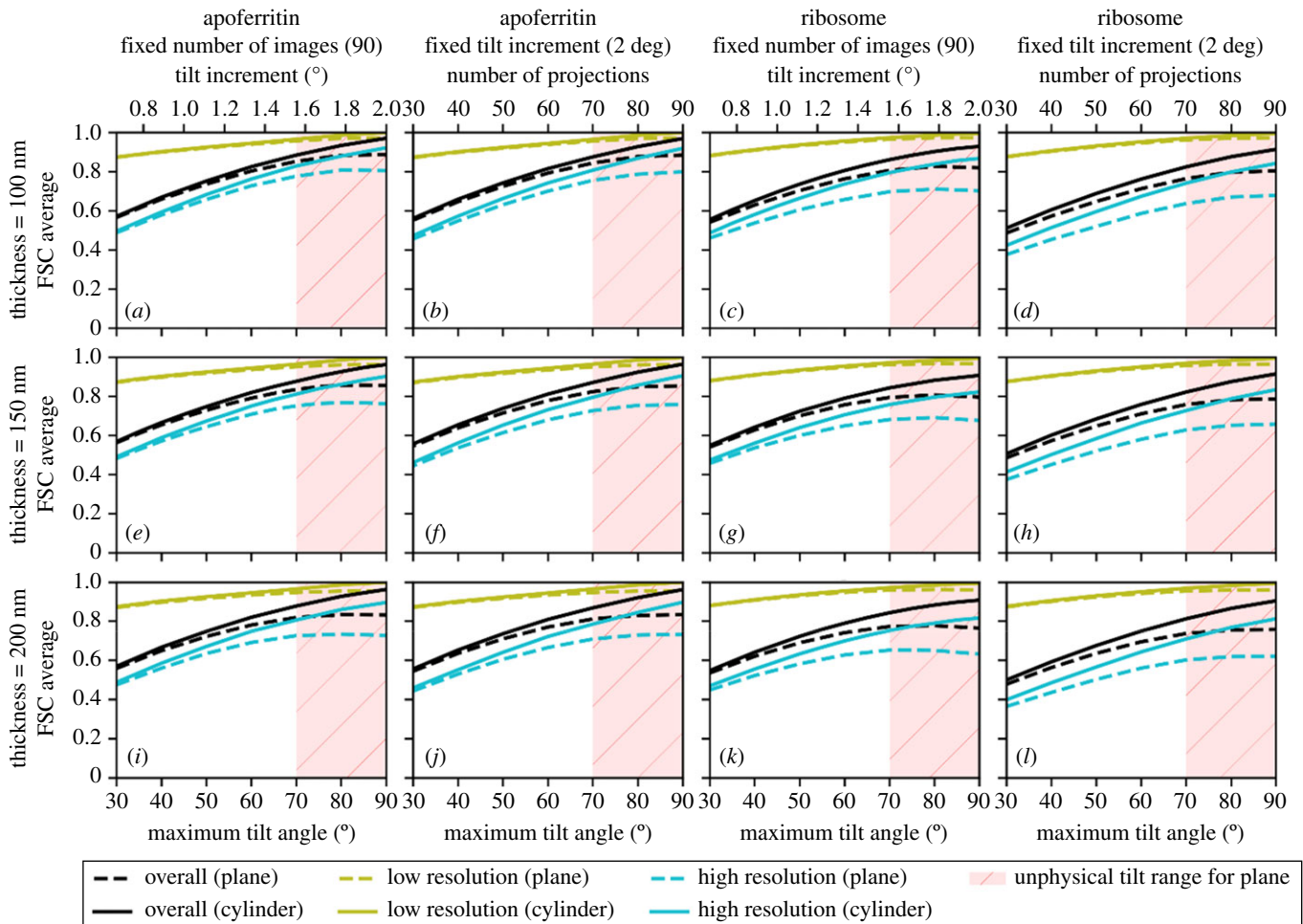


Figure 6. The effect of increasing the maximum tilt angle on the quality of reconstructed data for planar samples (dashed lines) and cylindrical samples (solid lines), for apoferritin (*a,b,e,f,i,j*) and ribosome samples (*c,d,g,h,k,l*), with a fixed number of projections and a fixed tilt increment, for sample, thicknesses of 100 nm (*a–d*), 150 nm (*e–h*) and 200 nm (*i–l*). The quality of the reconstruction is assessed by the overall FSC average and by the FSC in high- and low-resolution bins. As the maximum tilt angle increases, the quality of the reconstruction generally improves in each case. For the planar samples, high maximum tilt angles are considered to be unphysical due to sample and mechanical limitations and are shaded in red. In all cases, the cylindrical samples show marginally better reconstruction quality than the planar sample.

missing wedge. Even at more limited tilt angles, the cylindrical sample gives a slightly higher quality reconstruction. Tilt series with a fixed number of projections for a given tilt range produced marginally better-quality reconstructions than the equivalent tilt series with the same maximum tilt angle and a fixed tilt increment of 2° (figure 8). This is to be expected since the number of projections is closely related to the quality of the reconstructions and reducing the number of projections as well as the range of tilt angles will reduce the quality of the reconstruction even further. The difference in reconstruction quality is especially evident at high resolution.

3.4. Variation of reconstruction quality with number of projections

The reconstruction quality, as assessed by the FSC average, of the 180° dataset with an increasing number of projections was evaluated for the 150 nm thick cylindrical apoferritin and ribosome samples assuming perfect image alignment as shown in figure 9. As the number of projections is increased, the quality of the reconstruction improves to a maximum value which then remains approximately constant as other

factors affecting the reconstruction quality, such as the discrete sampling of the detector, begin to dominate over the rotational sampling of the volume. It is well known that the achievable resolution of a tomographic reconstruction is given by the Crowther criterion [2] which, ignoring any symmetry considerations or potential gains from averaging, is given by $d = \pi D/N$, where d is the achievable resolution, D is the diameter of the sample and N is the number of projections. This equation assumes that the resolution is not limited by the Nyquist sampling frequency of the detector, the sample is spherical with no additional symmetry, and that N projections completely fill Fourier space. In practice, when a tilt range $(-\alpha, \alpha)$ is used, and $\alpha < \pi/2$, the resolution is anisotropic such that, typically, $d_y < d_x < d_z$ [26] as previously illustrated in figure 7.

We have explored whether, given no missing wedge, fewer projections are needed to give the same reconstruction quality as data with a missing wedge. As shown in figure 10, for the 150 nm thick cylindrical apoferritin and ribosome samples, the number of projections required to achieve the same reconstruction quality is reduced by more than half when compared to corresponding planar samples with a 120° tilt range. This effect is approximately linear until a maximum rotation of around 70° where the increased apparent thickness of the planar sample at high

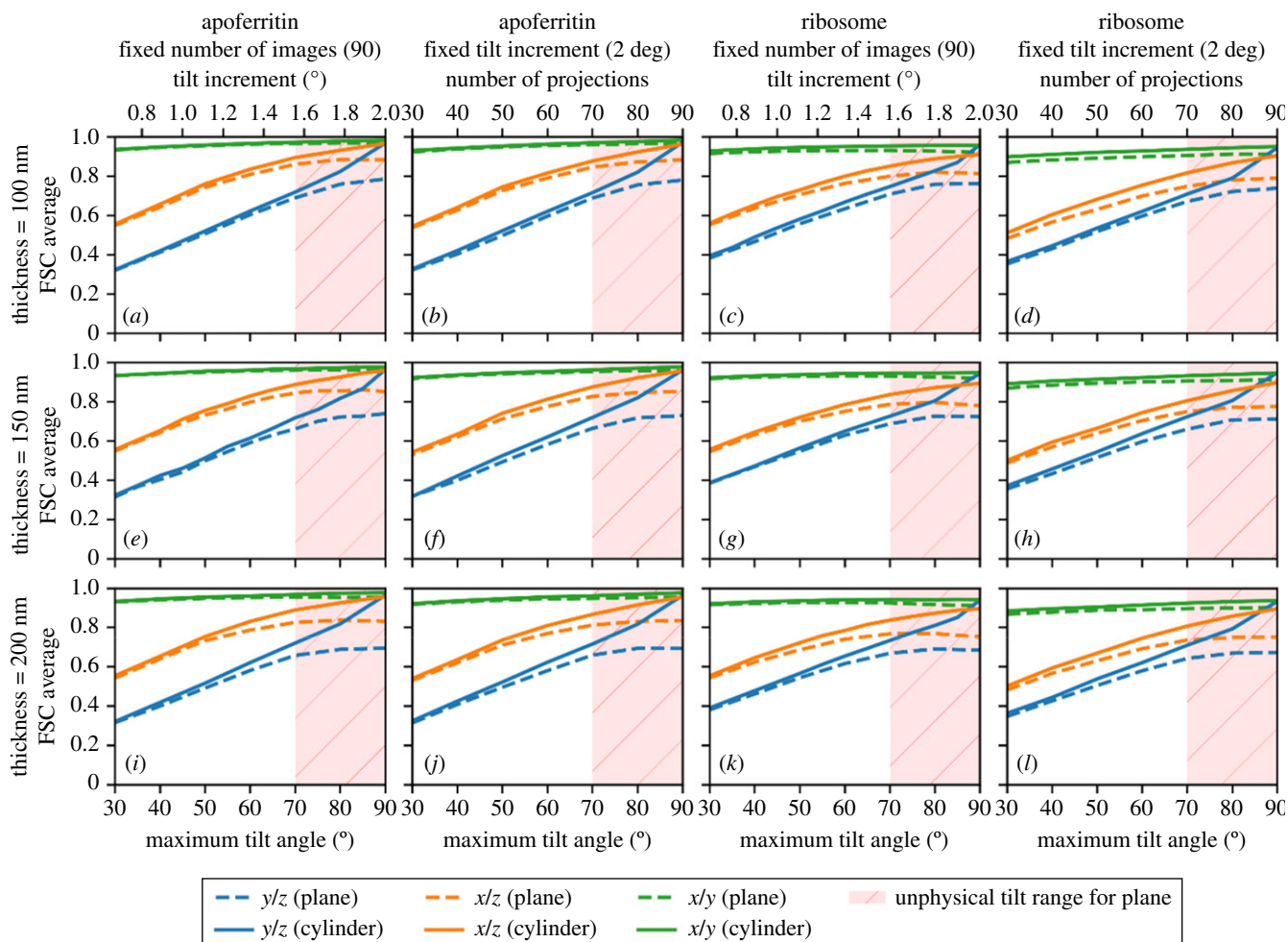


Figure 7. The effect of increasing the maximum tilt angle on the anisotropic FSC average of reconstructed data in the reciprocal x/y , x/z and y/z planes for planar samples (dashed lines) and cylindrical samples (solid lines), for apoferritin (a,b,e,f,i,j) and ribosome samples (c,d,g,h,k,l), with a fixed number of projections and a fixed tilt increment, for sample thicknesses of 100 nm ($a-d$), 150 nm ($e-h$) and 200 nm ($i-l$). For the planar samples, high maximum tilt angles are considered to be unphysical due to sample and mechanical limitations and are shaded in red.

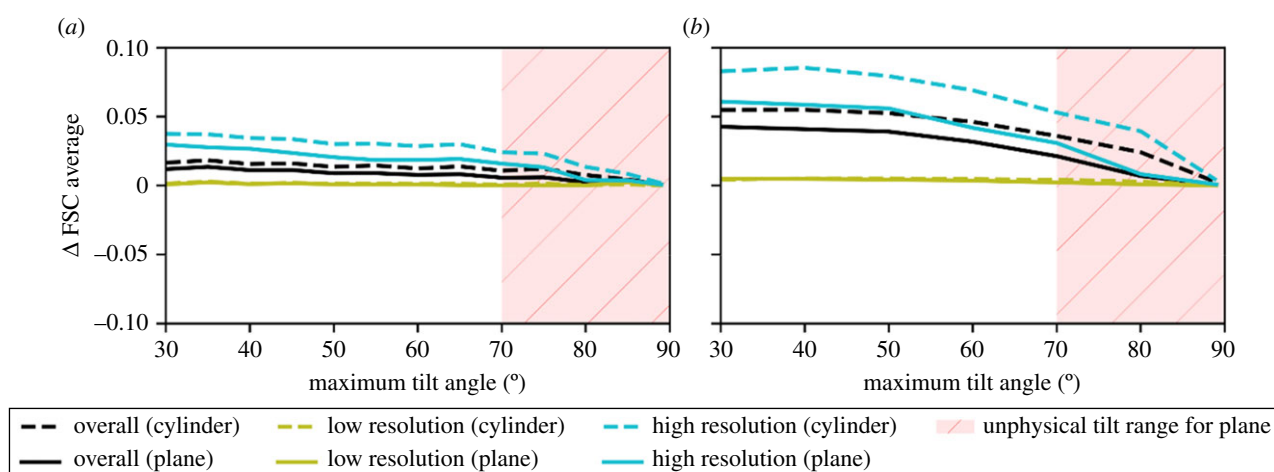


Figure 8. The difference in FSC average between the fixed number of projections and fixed tilt increment data acquisition schemes as a function of the maximum tilt angle for planar samples (dashed lines) and cylindrical samples (solid lines), for apoferritin (a) and ribosome samples (b). The plots show the overall FSC average and the FSC in high- and low-resolution bins. For the planar samples, high maximum tilt angles are considered to be unphysical due to sample and mechanical limitations and are shaded in red. In each case, the data acquisition scheme with a fixed number of images shows better reconstruction quality than the scheme with a fixed tilt increment.

tilt angles begins to reduce the benefit of including higher tilt angle data. The effect is even more pronounced in the reciprocal y/z plane where the reduction is 10-fold. This is highly beneficial since the limited dose tolerance of a

biological sample needs to be fractionated over the entire tilt series and hence a larger number of projections means a lower SNR per projection. The SNR is critical for alignment of the tilt series before reconstruction, thus although

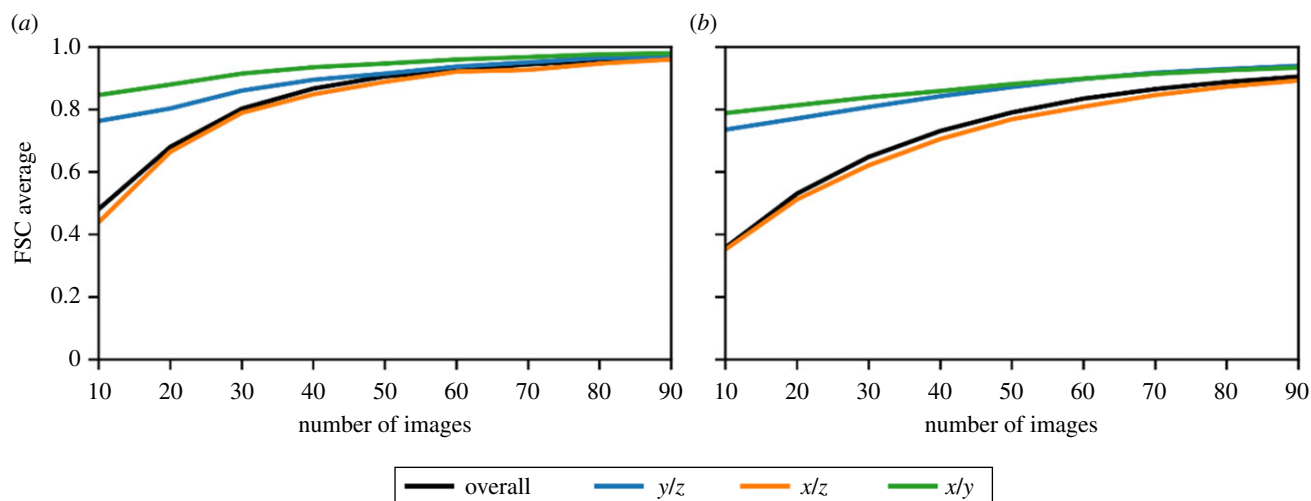


Figure 9. The quality of the reconstruction as a function of number of projections for a cylindrical specimen with no missing wedge. As the number of projections increases, the quality of the reconstruction increases until it begins to plateau. The overall FSC average is shown in black with the FSC average in the reciprocal x/y , x/z and y/z planes also shown.

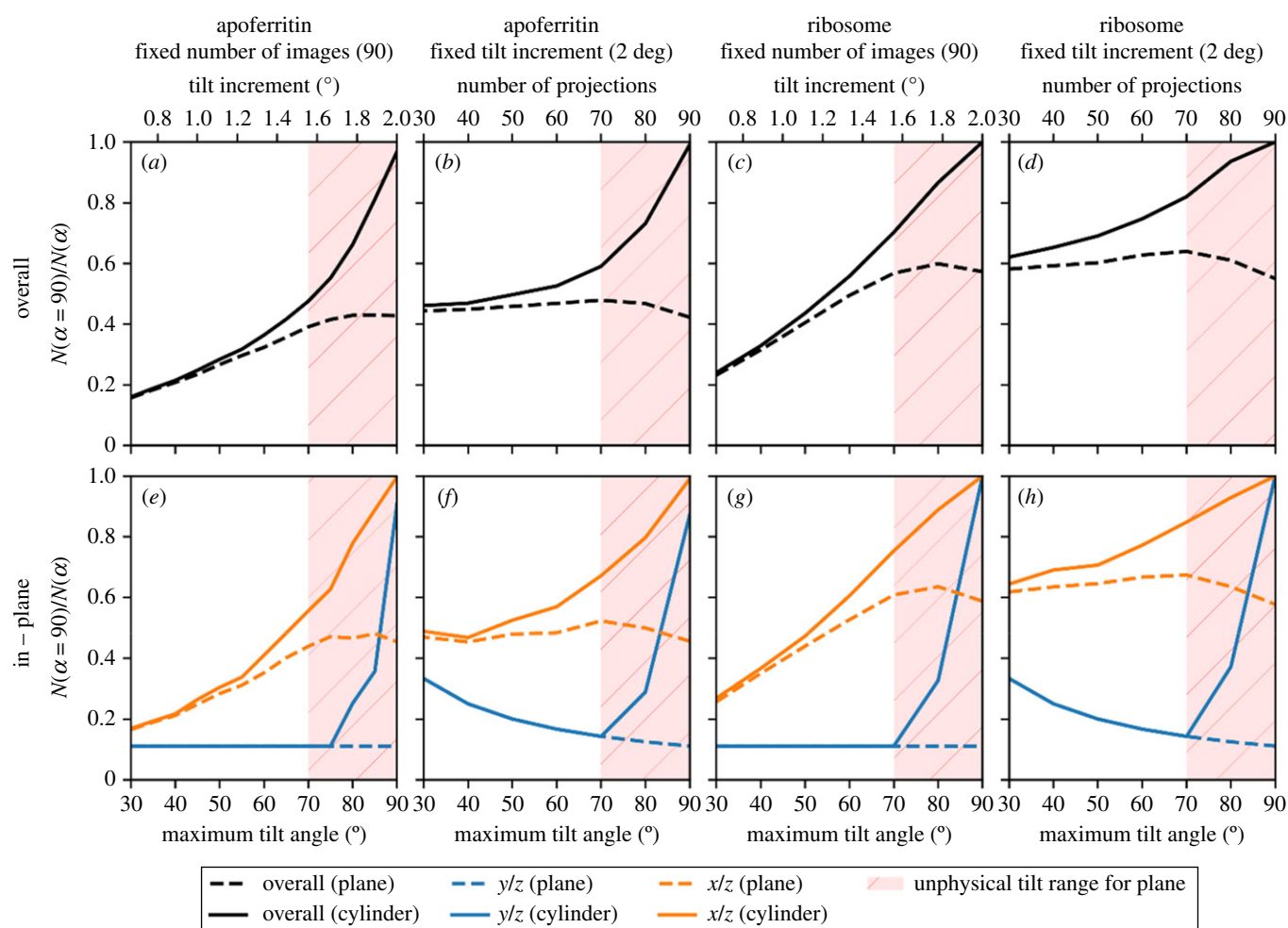


Figure 10. The fraction of projections required for a cylindrical sample with no missing wedge to achieve the same quality of reconstruction as the missing wedge datasets for a planar sample (dashed lines) and cylindrical sample (solid lines), for apoferritin (*a,b,e,f*) and ribosomes (*c,d,g,h*), with a fixed number of projections and a fixed tilt increment using the FSC average overall (*a–d*) and in the reciprocal x/z and y/z planes (*e–h*). For the planar samples, high maximum tilt angles considered to be unphysical due to sample and mechanical limitations are shaded in red.

increasing the number of projections increases the quality of reconstruction, the gain may be offset by errors in alignment. When fewer projections are needed to obtain the same quality of reconstruction, these projections can individually have higher SNR and thus reduced alignment errors.

4. Conclusion

Cryo-ET is revolutionizing cell biology, and there is significant interest in improving the technique. Current experimental approaches suffer from a missing wedge of data. Since the potential for the redesign of hardware is

limited, it is important to quantify the gains of any redesign. We report a digital twin software pipeline, Parakeet, which is capable of simulating tomographic tilt series of TEM projection images, reconstructing and analysing them. This has allowed us to quantify the effects of the missing wedge and sample geometry. For a planar lamella, the best reconstruction quality for a fixed number of projections was obtained with a maximum tilt angle of less than 90° since projections at higher tilt angles contributed little to the reconstruction and added noise. For a cylindrical sample, however, the quality of the reconstruction increased as the tilt range approached the full hemisphere. In addition, a cylindrical sample with the same thickness as the planar sample was shown to require fewer projections, spaced over the hemisphere, to achieve the same quality of reconstruction as a planar sample with a missing wedge. The software is open source and available to download from the Rosalind Franklin Institute GitHub page [31].

Data accessibility. 1. The atomic models of the apoferritin and ribosome used are publicly available and can be downloaded from the Protein Data Bank. 2. All the software described is free and open source and can be downloaded from the GitHub repository at <https://github.com/rosalindfranklininstitute/amplus-digital-twin>.

Authors' contributions. J.M.P. was involved in software, methodology, investigation and writing the original draft. M.D. was involved in writing the review and editing. M.B. was involved in writing the review and editing. D.C. was involved in writing the review and editing. C.A.S. was involved in writing the review and editing. T.V. was involved in writing the review and editing. A.K. was involved in supervision and writing the review and editing. J.H.N. was involved in supervision, conceptualization, writing, reviewing and editing. G.E. was involved in supervision, writing, reviewing and editing.

Competing interests. We declare we have no competing interests.

Funding. The Rosalind Franklin Institute is an EPSRC core funded Institute. Funding was also provided by the Wellcome Trust through the Electrifying Life Science grant.

Acknowledgements. Computing resources were provided by both the Rosalind Franklin Institute and Diamond Light Source.

Appendix A. Command line programs

The main command line programs that constitute the Parakeet software pipeline are shown below. These command line programs roughly correspond to the steps of the flow chart in figure 2.

— *parakeet.config.show*

This program is used to generate a new YAML configuration file (*config.yaml*) defining the microscope model and data acquisition strategy. This config file can then be used in all the subsequent command line steps. However, before proceeding, the config file may need to be manually modified by the user to customize some aspects of the simulation.

— *parakeet.sample.new* [-c *config.yaml*]

This program initializes the sample with the desired geometry and amorphous ice specification. The program takes a configuration file (by default assumed to be named *config.yaml*) and generates a skeleton HDF5 file (by default named *sample.h5*). The sample geometry can either be cuboid or cylindrical; the size of the cuboid or cylinder and its location and orientation with respect to the beam are specified here.

— *parakeet.sample.add_molecules* [-c *config.yaml*]

This program adds a number of particles to the sample atomic model with either given or random position and orientation. The program uses the same configuration file and takes as input the sample HDF5 file generated by the *parakeet.sample.new* program which is modified in place (i.e. multiple invocations of the command will modify the same file rather than overwriting the results of the previous function call).

— *parakeet.simulate.exit_wave* [-c *config.yaml*]

This program uses the multislice algorithm to simulate the exit wave of the electron through the sample. The program takes as input the config file and the sample HDF5 file generated in the previous steps. The exit wave is written out to either a HDF5 file (by default named *exit_wave.h5*) or an MRC file containing the simulated images for the whole tilt series.

— *parakeet.simulate.optics* [-c *config.yaml*]

This program simulates the effects of the microscope optics *via* the CTF. The program takes as input the config file and the exit wave file generated by the *parakeet.simulate.exit_wave* program. The tilt series with the microscope optics applied is then written out to either a HDF5 file (by default named *optics.h5*) or an MRC file containing the simulated images for the whole tilt series.

— *parakeet.simulate.image* [-c *config.yaml*]

This program simulates the detector response by adding the effect of the DQE and applying Poisson noise. The program takes as input the config file and the optics file generated by the *parakeet.simulate.optics* program. The tilt series images are then written out to either a HDF5 file (by default named *image.h5*) or an MRC file containing the simulated images for the whole tilt series.

— *parakeet.analyse.reconstruct* [-c *config.yaml*]

This program performs CTF correction and reconstructs the simulated tilt series. The program takes as input the same config file as used throughout the simulation process and the tilt images in MRC format. After reconstructing the tomogram, the program outputs an MRC file (by default named *rec.mrc*).

— *parakeet.analyse.average_particles* [-c *config.yaml*]

This program uses the given particle positions and orientations and averages the particle reconstructions. The program takes as input the configuration file, the sample HDF5 file and the reconstruction file output from the *parakeet.analyse.reconstruct* program. The particle reconstructions are extracted using the given particle positions and orientations and averaged in two halves. The program outputs these two half reconstructions in two files by default named *half1.mrc* and *half2.mrc*.

Appendix B. Beam damage model

The first-order effects of beam damage to the sample can be implemented by convolving the electrostatic potential with a Gaussian function [12] whose variance, σ_B^2 , can be related to an isotropic *B* factor by $B = 8\pi^2\sigma_B^2$ [35]. This convolution is most conveniently performed in Fourier space where the *B* factor filter is expressed as a function of spatial frequency, *q*. The potential from a damaged specimen, *V*, can then be

given in terms of the undamaged potential, V_0 , by

$$V = F^{-1}[F[V_0]e^{-q^2B/4}].$$

The isotropic B factor for protein crystals has been shown to vary linearly with the absorbed X-ray dose [36,37]. This dependence can be encapsulated in a coefficient of sensitivity to absorbed dose which gives the relative change in the isotropic B factor to the change in dose [37]:

$$S_{AD} = \frac{\Delta B}{8\pi^2\Delta D}.$$

This coefficient, which gives the slope of the linear relationship between the isotropic B factor and dose, is thought to take similar values for most protein crystals [37,72]. A typical value for this coefficient for cryo-cooled crystals at a temperature of 100 K is $S_{AD} = 0.012 \text{ A}^2 \text{ MGy}^{-1}$ [38]. The isotropic B factor for a given absorbed dose is then

$$B = 8\pi^2 D_{MGy} S_{AD}.$$

In Parakeet, rather than the sensitivity to absorbed dose in MGy, the sensitivity to the incident dose in e^-/A^2 is required. The dose in Gy for an incident number of electrons can be defined as follows, where D_e is the dose in electrons per unit area, ρ is the specimen density (kg m^{-3}), E_m is the mean energy loss (J), and λ_i is the total mean free path (m) for the inelastic scattering of electrons [73]:

$$D_{Gy} = \frac{D_e E_m}{\rho \lambda_i}.$$

The inelastic mean free path is given by $\lambda_i = 1/(N\sigma_i)$, where $N = \rho N_A/A$ is the number of particles per unit area

given in terms of the sample density, ρ , Avogadro's number, N_A and the atomic mass, A ; and σ_i is the total inelastic scattering cross-section which can be written in terms of the atomic number, Z , the velocity of the electrons, β , the electron acceleration voltage, V_0 , and the rest energy of the electron, mc^2 , as follows [74]:

$$\sigma_i = \frac{1.5 \times 10^{-6} Z^{1/2}}{\beta^2} \ln \left[2 / \left(\frac{E_m}{\beta^2 (V_0 + mc^2)} \right) \right] \text{ nm}^2.$$

Given typical values for biological specimens of $\rho = 1.35 \times 10^3 \text{ kg m}^{-3}$ and $E_m = 35 \text{ eV} = 5.6 \times 10^{-18} \text{ J}$ [73], and using carbon, which has an atomic number $Z = 6$ and atomic mass of $A = 12.01$, the mean free path at an electron energy of 300 keV is approximately $\lambda_i \approx 237 \text{ nm}$.

The isotropic B factor can then be written in terms of the electron dose as $B = 8\pi^2 D_e S_E$, where $S_E = 1.75 S_{AD} = 0.021 \text{ A}^2/\text{e}^-$. This value is broadly consistent with experimental observations [39] determined the beam damage using purple membrane crystals and reported that the B factor increased by 7 A^2 for every $1 \text{ e}^-/\text{A}^2$ of exposure at 300 keV. This would give a sensitivity coefficient of $S_E = 0.089 \text{ A}^2/\text{e}^-$. The model is also broadly consistent with the B factors observed and estimated in the Bayesian polishing algorithm implemented in RELION [40] where, for a Ribosome dataset, the total dose of $16 \text{ e}^-/\text{A}^2$ resulted in a final B factor of approximately 90, a β -Galactosidase dataset with a total dose of $45 \text{ e}^-/\text{A}^2$ resulted in a final B factor of approximately 160, and a γ -Secretase dataset with a total dose of $40 \text{ e}^-/\text{A}^2$ resulted in a final B factor of approximately 130. This would correspond to sensitivity coefficients of $S_E = 0.071 \text{ A}^2/\text{e}^-$, $S_E = 0.048 \text{ A}^2/\text{e}^-$ and $S_E = 0.041 \text{ A}^2/\text{e}^-$ respectively.

References

1. Bracewell RN. 1956 Strip integration in radio astronomy. *Aust. J. Phys.* **9**, 198–217. (doi:10.1071/PH560198)
2. Crowther RA, DeRosier DJ, Klug A. 1970 The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proc. R. Soc. A* **317**, 319–340. (doi:10.1098/rspa.1970.0119)
3. Hagen WJH, Wan W, Briggs JAG. 2017 Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198. (doi:10.1016/j.jsb.2016.06.007)
4. Eisenstein F, Danev R, Pilhofer M. 2019 Improved applicability and robustness of fast cryo-electron tomography data acquisition. *J. Struct. Biol.* **208**, 107–114. (doi:10.1016/j.jsb.2019.08.006)
5. Weis F, Hagen WJH. 2020 Combining high throughput and high quality for cryo-electron microscopy data collection. *Acta Crystallogr. Section D* **76**, 724–728. (doi:10.1107/S2059798320008347)
6. Bouvette J et al. 2021 Beam image-shift accelerated data acquisition for near-atomic resolution single-particle cryo-electron tomography. *Nat. Commun.* **12**, 1957. (doi:10.1038/s41467-021-22251-8)
7. Schorb M, Habersbosch I, Hagen WJ, Schwab Y, Mastrorade DN. 2020 Europe PMC funders group software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477. (doi:10.1038/s41592-019-0396-9. Software)
8. Zivanov J, Nakane T, Forsberg B, Kimanius D, Hagen WJH, Lindahl E, Scheres SHW. 2018 RELION-3: new tools for automated high-resolution cryo-EM structure determination. *BioRxiv*, 1–22. (doi:10.1101/421123)
9. Dubochet J, Adrian M, Chang J-J, Homo J-C, Lepault J, McDowell AW, Schultz P. 1988 Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics*, **21**.
10. Cowley JM, Moodie AF. 1957 The scattering of electrons by atoms and crystals. I. A new theoretical approach. *Acta Crystallogr.* **10**, 609–619. (doi:10.1107/s0365110X57002194)
11. Goodman P, Moodie AF. 1974 Numerical evaluations of N-beam wave functions in electron scattering by the multi-slice method. *Acta Crystallogr. Section A* **30**, 280–290. (doi:10.1107/S056773947400057X)
12. Zhang Y, Tammaro R, Peters PJ, Ravelli RBG. 2020 Could egg white lysozyme be solved by single particle cryo-EM? *J. Chem. Information Model.* **60**, 2605–2613. (doi:10.1021/acs.jcim.9b01176)
13. Palmer CM, Löwe J. 2014 A cylindrical specimen holder for electron cryo-tomography. *Ultramicroscopy* **137**, 20–29. (doi:10.1016/j.ultramic.2013.10.016)
14. Harauz G, van Heel M. 1985 Exact filters for general geometry three dimensional reconstruction. *Optik* **78**, 1986.
15. Midgley PA, Weyland M. 2003 3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography. *Ultramicroscopy* **96**, 413–431. (doi:10.1016/S0304-3991(03)00105-0)
16. Penczek P, Marko M, Buttle K, Frank J. 1995 Double-tilt electron tomography. *Ultramicroscopy* **60**, 393–410. (doi:10.1016/0304-3991(95)00078-X)
17. Tong JR, Weyland M, Midgley PA. 2003 Benefits and drawbacks of dual-axis STEM tomography. See www.rafaledb.com/papers/C-2004-08-Antwerp-EMC-dual-axis-STEM-tomography.pdf.
18. Tong J, Arslan I, Midgley P. 2006 A novel dual-axis iterative algorithm for electron tomography. *J. Struct. Biol.* **153**, 55–63. (doi:10.1016/j.jsb.2005.10.005)
19. Turoňová B et al. 2020 Benchmarking tomographic acquisition schemes for high-resolution structural biology. *Nat. Commun.* **11**, 1–9. (doi:10.1038/s41467-020-14535-2)

20. Iancu CV *et al.* 2005 A 'flip-flop' rotation stage for routine dual-axis electron cryotomography. *J. Struct. Biol.* **151**, 288–297. (doi:10.1016/j.jsb.2005.07.004)
21. Arslan I, Tong JR, Midgley PA. 2006 Reducing the missing wedge: high-resolution dual axis tomography of inorganic materials. *Ultramicroscopy* **106**(11–12 SPEC. ISS.), 994–1000. (doi:10.1016/j.ultramic.2006.05.010)
22. Cantele F, Paccagnini E, Pigino G, Lupetti P, Lanzavecchia S. 2010 Simultaneous alignment of dual-axis tilt series. *J. Struct. Biol.* **169**, 192–199. (doi:10.1016/j.jsb.2009.10.003)
23. Bals S, Goris B, Altantzis T, Heidari H, Van Aert S, Van Tendeloo G. 2014 Seeing and measuring in 3D with electrons. *Comptes Rendus Physique* **15**, 140–150. (doi:10.1016/j.crhy.2013.09.015)
24. Batenburg KJ *et al.* 2009 3D imaging of nanomaterials by discrete tomography. *Ultramicroscopy* **109**, 730–740. (doi:10.1016/j.ultramic.2009.01.009)
25. Goris B, Van den Broek W, Batenburg KJ, Heidari Mezerji H, & Bals S. 2012 Electron tomography based on a total variation minimization reconstruction technique. *Ultramicroscopy* **113**, 120–130. (doi:10.1016/j.ultramic.2011.11.004)
26. Bals S, Van Aert S, Van Tendeloo G. 2013 High resolution electron tomography. *Curr. Opin. Solid State Materials Sci.* **17**, 107–114. (doi:10.1016/j.cossms.2013.03.001)
27. Saghi Z, Benning M, Leary R, Macias-Montero M, Borrás A, Midgley PA. 2015 Reduced-dose and high-speed acquisition strategies for multi-dimensional electron microscopy. *Adv. Struct. Chem. Imag.* **1**, 1–10. (doi:10.1186/s40679-015-0007-5)
28. Liu J *et al.* 2018 A modified discrete tomography for improving the reconstruction of unknown multi-gray-level material in the 'missing wedge' situation. *J. Synchrotron Radiation* **25**, 1847–1859. (doi:10.1107/S1600577518013681)
29. Cheng A *et al.* 2015 MRC2014: extensions to the MRC format header for electron cryo-microscopy and tomography. *J. Struct. Biol.* **192**, 146–150. (doi:10.1016/j.jsb.2015.04.002)
30. Burnley T, Palmer CM, Winn M. 2017 Recent developments in the CCP-EM software suite. *Acta Crystallogr. Section D* **73**, 469–477. (doi:10.1107/S2059798317007859)
31. Parkhurst J. 2021 Software for parakeet (Program for Analysis and Reconstruction of Artificial data for Kryo Electron Tomography): a digital twin for cryo electron tomography. (doi:10.5281/zenodo.5185910Peet)
32. Berman HM *et al.* 2000 The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242. (doi:10.1093/nar/28.1.235)
33. Lobato I, Van Dyck D. 2015 MULTEM: a new multislice program to perform accurate and fast electron diffraction and imaging simulations using Graphics Processing Units with CUDA. *Ultramicroscopy* **156**, 9–17. (doi:10.1016/j.ultramic.2015.04.016)
34. Jakob W, Rhinelander J, Moldovan D. 2017 pybind11—Seamless operability between C++ and Python. Retrieved from <https://github.com/pybind/pybind11>.
35. Carugo O. 2019 Maximal B-factors in protein crystal structures. *Zeitschrift Fur Kristallographie—Crystalline Materials* **234**, 73–77. (doi:10.1515/zkri-2018-2057)
36. Owen RL, Rudiño-Piñera E, Garman EF. 2006 Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proc. Natl Acad. Sci. USA* **103**, 4912–4917. (doi:10.1073/pnas.0600973103)
37. Garman EF. 2010 Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr. Section D* **66**, 339–351. (doi:10.1107/S0907444910008656)
38. Shelley KL, Dixon TPE, Brooks-Bartlett JC, Garman EF. 2018 RABDAM: quantifying specific radiation damage in individual protein crystal structures. *J. Appl. Crystallogr.* **51**, 552–559. (doi:10.1107/S1600576718002509)
39. Peet MJ, Henderson R, Russo CJ. 2019 The energy dependence of contrast and damage in electron cryomicroscopy of biological molecules. *Ultramicroscopy* **203**, 125–131. (doi:10.1016/j.ultramic.2019.02.007)
40. Zivanov J, Nakane T, Scheres SHW. 2019 A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCr*, **6**, 5–17. (doi:10.1107/S205225251801463X)
41. Thermo Fisher Scientific. 2020 *Krios G4 pre-installation manual*. Waltham, MA: Thermo Fisher Scientific.
42. Lobato I, Van Dyck D. 2014 An accurate parameterization for scattering factors, electron densities and electrostatic potentials for neutral atoms that obey all physical constraints. *Acta Crystallogr. Section A* **70**, 636–649. (doi:10.1107/S205327331401643X)
43. Downing KH, Glaeser RM. 2008 Restoration of weak phase-contrast images recorded with a high degree of defocus: the 'twin image' problem associated with CTF correction. *Ultramicroscopy* **108**, 921–928. (doi:10.1016/j.ultramic.2008.03.004)
44. Haigh SJ, Kirkland AI. 2012 High resolution exit wave restoration. In *Nanostructure science and technology* (ed. DJ Lockwood), pp. 1–362. Berlin, Germany: Springer.
45. Zanetti G, Riches JD, Fuller SD, Briggs JAG. 2009 Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging. *J. Struct. Biol.* **168**, 305–312. (doi:10.1016/j.jsb.2009.08.002)
46. Turoňová B, Schur FKM, Wan W, Briggs JAG. 2017 Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *J. Struct. Biol.* **199**, 187–195. (doi:10.1016/j.jsb.2017.07.007)
47. Wan W, Briggs JAGG. 2016 *Cryo-electron tomography and subtomogram averaging*. Amsterdam, Netherlands: Elsevier.
48. van Aarle W, Palenstijn WJ, De Beenhouwer J, Altantzis T, Bals S, Batenburg KJ, Sijbers J. 2015 The ASTRA Toolbox: a platform for advanced algorithm development in electron tomography. *Ultramicroscopy* **157**, 35–47. (doi:10.1016/j.ultramic.2015.05.002)
49. Gürsoy D, De Carlo F, Xiao X, Jacobsen C. 2014 TomoPy: a framework for the analysis of synchrotron tomographic data. *J. Synchrotron Radiation* **21**, 1188–1193. (doi:10.1107/S1600577514013939)
50. Pelt DM, Gürsoy D, Palenstijn WJ, Sijbers J, De Carlo F, Batenburg KJ. 2016 Integration of TomoPy and the ASTRA toolbox for advanced processing and reconstruction of tomographic synchrotron data. *J. Synchrotron Radiation* **23**, 842–849. (doi:10.1107/S1600577516005658)
51. Radermacher M. 2005 Weighted back projection methods. In *Electron tomography: methods for three-dimensional visualization of structures in the cell* (ed. J Frank), pp. 245–274. Berlin, Germany: Springer.
52. Penczek PA. 2010 Fundamentals of three-dimensional reconstruction from projections. *Methods Enzymol.* **482**, 1–33. (doi:10.1016/S0076-6879(10)82001-4)
53. Heidari Mezerji H, Van den Broek W, Bals S. 2011 A practical method to determine the effective resolution in incoherent experimental electron tomography. *Ultramicroscopy* **111**, 330–336. (doi:10.1016/j.ultramic.2011.01.021)
54. Chen D *et al.* 2014 The properties of SIRT, TVM, and DART for 3D imaging of tubular domains in nanocomposite thin-films and sections. *Ultramicroscopy* **147**, 137–148. (doi:10.1016/j.ultramic.2014.08.005)
55. Paavolainen L *et al.* 2014 Compensation of missing wedge effects with sequential statistical reconstruction in electron tomography. *PLoS ONE* **9**, 10. (doi:10.1371/journal.pone.0108978)
56. Kupsch A *et al.* 2016 Missing wedge computed tomography by iterative algorithm DIRECTT. *J. Microsc.* **261**, 36–45. (doi:10.1111/jmi.12313)
57. Wadson N, Basham M. 2016 Savu: a Python-based, MPI framework for simultaneous processing of multiple, N-dimensional, large tomography datasets. *ArXiv:1610.08015 [CS]*. Retrieved from <http://arxiv.org/abs/1610.08015%0Ainternal-pdf//0.0.0.96/1610.html>.
58. Murshudov GN *et al.* 2011 REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Section D* **67**, 355–367. (doi:10.1107/S0907444911001314)
59. Brown A, Long F, Nicholls RA, Toots J, Emsley P, Murshudov G. 2015 Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. Section D* **71**, 136–153. (doi:10.1107/S1399004714021683)
60. Midgley PA, Ward EPW, Hungria AB, Thomas JM. 2007 Nanotomography in the chemical, biological and materials sciences. *Chem. Soc. Rev.* **36**, 1477–1494. (doi:10.1039/b701569k)
61. Kawase N, Kato M, Nishioka H, Jinnai H. 2007 Transmission electron microtomography without the 'missing wedge' for quantitative structural analysis. *Ultramicroscopy* **107**, 8–15. (doi:10.1016/j.ultramic.2006.04.007)

62. Le Gros MA *et al.* 2014 Biological soft X-ray tomography on beamline 2.1 at the advanced light source. *J. Synchrotron Radiation* **21**, 1370–1377. (doi:10.1107/S1600577514015033)
63. Baumeister W, Grimm R, Walz J. 1999 Electron tomography of molecules and cells. *Trends Cell Biol.* **9**, 81–85. (doi:10.1016/S0962-8924(98)01423-8)
64. Moor H. 1987 Theory and practice of high pressure freezing. In *Cryotechniques in biological electron microscopy* (eds RA Steinbrecht, K Zierold), pp. 175–191. Berlin, Germany: Springer.
65. Studer D, Humbel BM, Chiquet M. 2008 Electron microscopy of high-pressure frozen samples: bridging the gap between cellular ultrastructure and atomic resolution. *Histochem. Cell Biol.* **130**, 877–889. (doi:10.1007/s00418-008-0500-1)
66. Thompson RF, Walker M, Siebert CA, Muench SP, Ranson NA. 2016 An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* **100**, 3–15. (doi:10.1016/j.ymeth.2016.02.017)
67. Van Doren EAF, De Temmerman PJRH, Francisco MAD, Mast J. 2011 Determination of the volume-specific surface area by using transmission electron tomography for characterization and definition of nanomaterials. *J. Nanobiotechnol.* **9**, 1–8. (doi:10.1186/1477-3155-9-17)
68. Huang J, Wang X, Grandfield K. 2014 FIB preparation of bone-implant interfaces for correlative on-axis rotation electron tomography and atom probe tomography. *Microsc. Microanal.* **20**, 352–353. (doi:10.1017/S1431927614003481)
69. Yip KM, Fischer N, Paknia E, Chari A, Stark H. 2020 A structure of human apoferritin obtained from Titan Mono-BCOR microscope. See www.wwpdb.org/pdb?id=pdb_00006z6u. (doi:10.2210/pdb6z6u/pdb)
70. Voorhees RM, Weixlbaumer A, Loakes D, Kelley AC, Ramakrishnan V. 2009 Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat. Struct. Mol. Biol.* **16**, 528–533. (doi:10.1038/nsmb.1577)
71. Nakane T *et al.* 2020 Single-particle cryo-EM at atomic resolution. *bioRxiv*. (doi:10.1101/2020.05.22.110189)
72. Kmetko J, Husseini NS, Naides M, Kalinin Y, Thorne RE. 2006 Quantifying X-ray radiation damage in protein crystals at cryogenic temperatures. *Acta Crystallogr. Section D* **62**, 1030–1038. (doi:10.1107/S0907444906023869)
73. Egerton RF. 2015 Outrun radiation damage with electrons? *Adv. Struct. Chem. Imaging* **1**, 1–11. (doi:10.1186/s40679-014-0001-3)
74. Langmore JP, Smith MF. 1992 Quantitative energy-filtered electron microscopy of biological molecules in ice. *Ultramicroscopy* **46**, 349–373. (doi:10.1016/0304-3991(92)90024-E)