

## The potential for two-dimensional crystallography of membrane proteins at future x-ray free-electron laser sources

Cameron M Kewish<sup>1,3</sup>, Pierre Thibault<sup>1</sup>, Oliver Bunk<sup>1</sup>  
and Franz Pfeiffer<sup>2,3,4</sup>

<sup>1</sup> Swiss Light Source, Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland

<sup>2</sup> Physik-Department (E17), Technische Universität München,  
James-Franck-Strausse, D-85748 Garching, Germany

E-mail: [cameron.kewish@psi.ch](mailto:cameron.kewish@psi.ch) and [franz.pfeiffer@ph.tum.de](mailto:franz.pfeiffer@ph.tum.de)

*New Journal of Physics* **12** (2010) 035005 (9pp)

Received 18 September 2009

Published 31 March 2010

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/12/3/035005

**Abstract.** Ultrashort pulses from x-ray free-electron laser (XFEL) sources promise to assist in obtaining the structures of membrane proteins at high resolution. We have reconstructed the electron density distribution of a two-dimensional (2D) aquaporin crystal from simulated XFEL data using ptychography, a diffractive imaging technique based on multiple exposures. Increasing the number of exposures compensates for Poisson noise, indicating that the achievable resolution is limited by the reproducibility of the crystals. This technique should therefore be applicable at all future ultrashort-pulsed hard x-ray sources.

### Contents

<b>1. Introduction</b>	<b>2</b>
<b>2. Simulation of two-dimensional membrane protein crystallography</b>	<b>2</b>
<b>3. Results and discussion</b>	<b>6</b>
<b>4. Conclusion</b>	<b>8</b>
<b>Acknowledgments</b>	<b>8</b>
<b>References</b>	<b>9</b>

<sup>3</sup> Authors to whom any correspondence should be addressed.

<sup>4</sup> Previous address: Swiss Light Source, Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland and École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.

## 1. Introduction

The most successful technique for obtaining protein structures to date, x-ray crystallography, has revealed the three-dimensional (3D) structures of a great many proteins at or near atomic resolution. However, the highest resolution is attainable only for proteins that can be crystallized, and is limited by radiation damage [1]. Membrane proteins, which include attractive drug target families for developing novel therapeutics, are notoriously reluctant candidates for crystallization and x-ray structure determination. Furthermore, their solubilization in detergents, which is necessary to substitute them for the lipid bilayer, may significantly perturb their stability and structure. 2D crystals of membrane proteins embedded in a lipid bilayer therefore represent a more natural state for these biomolecules [2] and it is of great interest to develop powerful tools for investigating such structures.

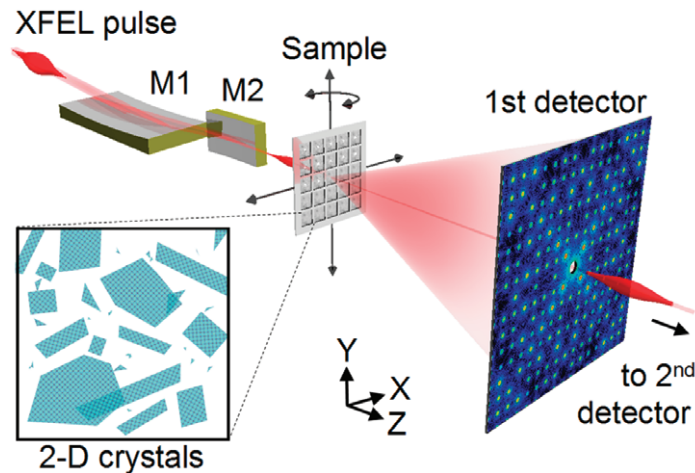
Atomic resolution imaging of individual biological macromolecules is perhaps one of the most challenging and ambitious science cases for the development of x-ray free-electron laser (XFEL) sources [3]–[5]. Provided useful information can be obtained prior to the radiation-induced explosion of a sample, theory [6, 7] and experiments with soft x-rays [8] predict that high-resolution images could be reconstructed from XFEL diffraction data. The scattering intensity from a single molecule will, however, be extremely low, demanding sophisticated algorithms to reconstruct the diffracted intensity distribution in reciprocal space [9, 10].

As an intermediate step toward single particles, XFEL imaging of 2D crystals of membrane proteins was proposed [4, 5, 11]. To date, 2D crystals of membrane proteins have only been investigated by cryoelectron crystallography up to atomic or near-atomic resolution, with aquaporin representing perhaps the most prominent example [2, 12]. The 2D aquaporin crystals, once spread out on to a membrane, are of micron-scale dimensions and of sufficient quality to allow the structure to be refined to within a few angstroms uncertainty. However, at large tilt angles, the collection of high-resolution electron diffraction data becomes difficult due to the increase in projected thickness and multiple scattering effects. This is not the case with x-ray diffraction from samples of this size, and therefore such structures are well suited to XFEL experiments. The symmetry and size of the crystals may also provide convenient workarounds to some of the difficulties expected with single molecule imaging, e.g. boosting the signal-to-noise ratio by having many copies of the sample in the beam [13] and reducing the classification problem of orienting the sample to translation and rotation in a single plane.

## 2. Simulation of two-dimensional membrane protein crystallography

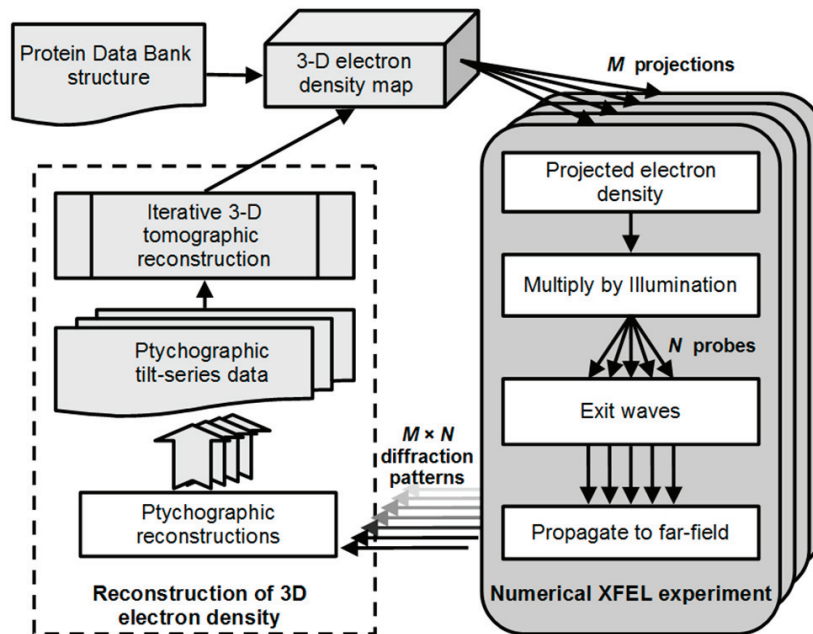
We present the results of a numerical experiment designed to assess the feasibility of ‘flash’ imaging of 2D membrane protein structures at planned XFEL facilities. The basis of this experiment, shown schematically in figure 1, was to simulate XFEL diffraction patterns with realistic photon counting statistics, using a known protein structure. From the simulated diffraction data, we reconstructed the exit wave leaving the sample using a ptychographic coherent diffractive imaging algorithm [14], thereby obtaining an estimate of the projected electron density. A tomographic reconstruction technique was then used to reconstruct the volume electron density from a tilt series. The procedure is summarized in figure 2.

Ptychography is a phase-retrieval method, introduced in electron microscopy, in which the interference between neighboring Bragg peaks and the redundancy introduced by measurements from partially overlapping regions of a sample are used to solve the phase problem [15]. The



**Figure 1.** Schematic diagram of the experiment geometry for 2D crystallography of membrane proteins. The XFEL beam is incident from the left, and is focused by a low-absorbing grazing-incidence mirror pair (M1, M2) on to the sample. The sample is an electron microscopy grid coated with 2D membrane protein crystals (illustrated in the inset), which can be translated perpendicular to the beam, and tilted over a range of angles. The diffracted beam is incident on a gated detector, which has a hole for the intense primary beam to pass through, to be measured downstream with an alternate detector having lower sensitivity.

iterative realization of this technique uses the illumination function, or ‘probe’, as a real-space support constraint and allows imaging of extended objects that would otherwise not be considered band limited in the sense of Shannon sampling for a given detector distance and pixel pitch [16]. It has been shown that both the exit wave [17] and the probe [14, 18] can be extracted from a ptychography dataset. The key feature of ptychography for XFEL-based 2D crystallography is that a relative movement of the sample and probe (also called diversity [19]) is a requirement for reconstruction. Focused XFEL pulses would destroy part of the sample, but the crystalline periodicity ensures that diffraction patterns measured from random undamaged regions on the same crystal, or from other crystals, contain redundant information that corresponds mathematically to regions of the sample that partially coincide. The secondary electrons released during radiation damage, which we have assumed move significantly only after the diffraction patterns have been measured from the undamaged 2D crystal, may proceed to cause further damage to the sample. It is known that the secondary electrons will be primarily produced in a direction perpendicular to the incident beam. Therefore, it may be possible that the electrons escape the 2D crystal, provided that there is a certain tilt angle. For most of the projections in the present simulations, the target is not hit under normal incidence, and therefore in the majority of cases the radiation damage to other parts of a given 2D crystal is not so severe. However, this could cause issues under normal incidence. Nevertheless, illumination of destroyed sample regions should produce a markedly different diffraction pattern than an undamaged sample, so in principle these affected data could be vetoed from the usable data. While the quality of the reconstruction may be affected by the accuracy with which the experiment parameters are known [19]–[21], we show that the exit wave can be obtained from randomly placed probes, which can be registered within a unit cell,



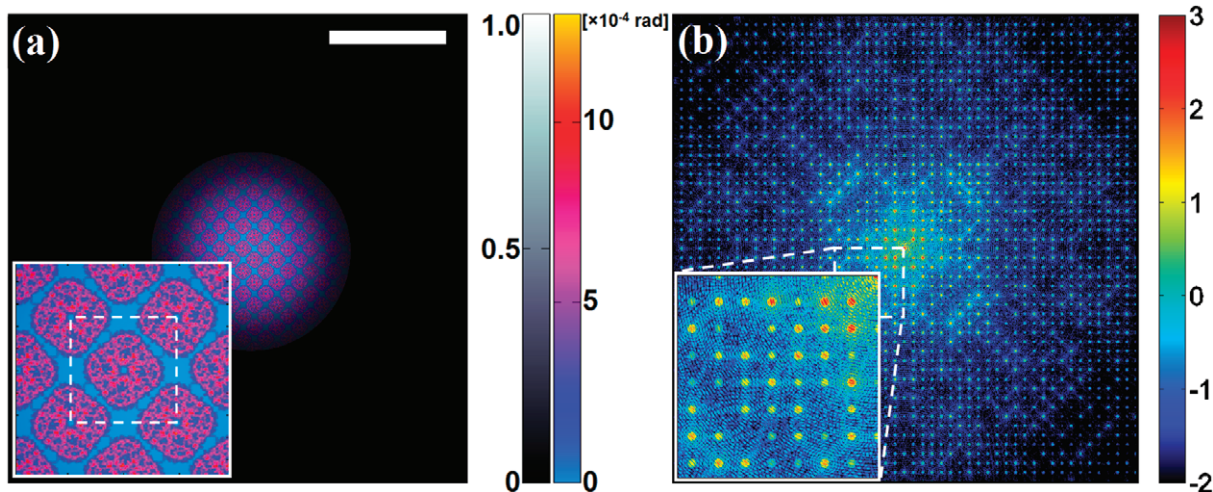
**Figure 2.** Flowchart representing the simulation procedures. From a PDB file, the electron density map was generated for a 2D crystal, from which diffraction data were generated. In an experiment, the right side of this diagram would be replaced by the measurement and classification of the XFEL diffraction patterns to be given as input to the ptychography reconstruction scheme.

and that the quality of the reconstruction improves with the number of probes and the XFEL pulse intensity.

As a demonstration structure, we selected the well-known human red cell aquaporin-1 (AQP1) membrane protein (Protein Data Bank: 1FQY). The atomic positions in the PDB file were used to synthesize an electron density map of the AQP1 unit cell with voxels of volume  $1 \text{ \AA}^3$ . In the membrane, AQP1 forms a tetramer structure, and the unit-cell effectively contains eight units of the monomer since there are two tetramers in opposite orientations with respect to the membrane [22, 23]. A 2D crystal was simulated by adding translated unit cells to form a square lattice with period  $a = b = 96 \text{ \AA}$ . This is somewhat idealized since we include no conformational flexibility or variation of long-range crystalline order; however, this method of imaging requires only that the crystals be uniform within the illuminated region.

To simulate a tilt series, projections of the rotated electron density,  $\rho_e^\theta$ , were calculated over a range of angles  $\theta = \pm 60^\circ$  around the  $Y$ -axis, in  $1^\circ$  intervals. The transmission function,  $O$ , was calculated assuming that the crystal is a weak phase object for radiation of wavelength  $\lambda = 1 \text{ \AA}$ , e.g.  $O^\theta(\mathbf{r}) = \exp[-ir_e\lambda\rho_e^\theta(\mathbf{r})]$ , where  $r_e$  is the classical electron radius and  $\mathbf{r}$  is the position perpendicular to the optical axis. This assumes that  $\lambda$  is far from any absorption edges, and short enough that the scattering factor in the forward direction can be approximated by the number of electrons.

The probe,  $P$ , was modeled as a coherent and monochromatic illumination function at the sample ( $XY$ ) plane with a symmetric 2D Gaussian amplitude envelope. Focusing optics, such as the mirrors shown in figure 1, produce an asymmetric and structured beam resulting



**Figure 3.** (a) Complex exit wave showing the Gaussian pulse amplitude and phase shift introduced by the 2D crystal. Inset: enlargement of the unit cell, outlined by a square with  $96 \text{ \AA}$  sides. (b) Difference between the diffraction patterns of the sample and probe at a pulse intensity of  $5 \times 10^{12}$  photons, shown on a log scale without noise. Inset: enlarged region showing Bragg peaks and interference structure which varies with probe position. The scale bar in (a) indicates  $50 \text{ nm}$ , and the main diffraction pattern in (b) is a subset  $[q_x, q_y] \in \pm 1.64 \text{ \AA}^{-1}$  of the full far-field intensity dataset which extends to  $\pm 3.14 \text{ \AA}^{-1}$ .

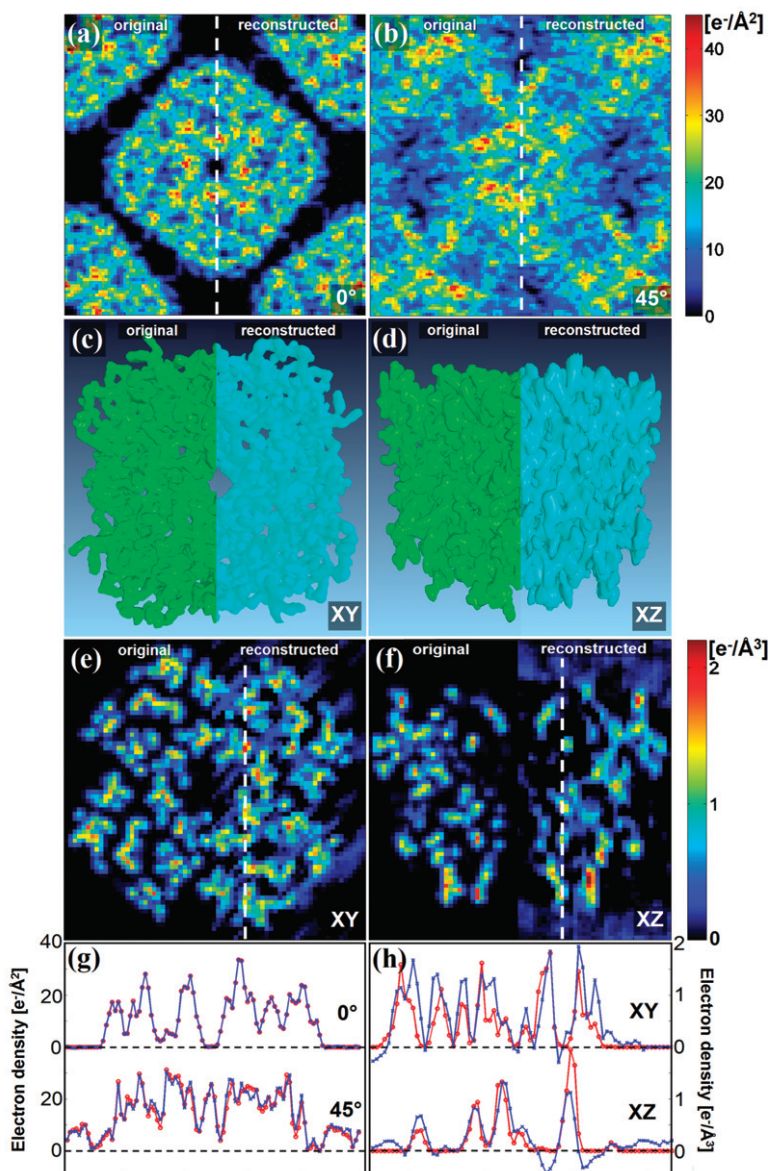
from the imperfections in the mirror surfaces [24]–[26]. Such a structure is not detrimental to ptychography reconstruction, provided it is constant during the measurements, because the structure in the probe can be refined from the data. The full width at half-maximum (FWHM) of the probe was  $50 \text{ nm}$ , a value achievable with state-of-the-art hard x-ray optics [27]–[30]. The probe was truncated to zero in the region where amplitude is less than  $1/e^2$  of the maximum value. Numerically, this allowed us to avoid spectral leakage when calculating the diffraction patterns, which results from the tails of the Gaussian overlapping the edges of the array. However, we observe that the presence of such an aperture in the immediate vicinity of the sample (created by e.g. a nanostructured substrate containing fiducial markers and windows of size  $0.1\text{--}0.2 \mu\text{m}$  on to which the sample is deposited) includes a high spatial frequency content in the wavefield that assists reconstruction due to the improved cross-talk between the Bragg peaks in the diffraction pattern.

We make use of a multiplicative approximation to obtain the exit waves,  $\psi$ , e.g.  $\psi_j^\theta(\mathbf{r}) = P(\mathbf{r} - \mathbf{r}_j)O^\theta(\mathbf{r})$ , where  $\psi_j^\theta$  is the exit wave for the  $j$ th probe at tilt  $\theta$ , and the displacements  $\mathbf{r}_j$  are randomly generated. The phase shifts induced by the AQP1 crystal are around  $5 \times 10^{-4} \text{ rad}$  (figure 3(a)). Consequently, the far field is dominated by the diffraction pattern of the probe. The difference between the diffraction patterns of the sample and the probe reveals the Bragg peaks from the crystal structure, and coherent interference between the peaks, which varies with the position of the probe (figure 3(b)). This structure manifests at the level of around  $10^2\text{--}10^{-2}$  photons per pixel, indicating that multiple shots are required to collect enough information to recover the protein structure to high resolution. The XFEL pulse intensity was assumed to be  $5 \times 10^{12}$  photons, implying a power density sufficient to destroy a sample within

a few tens of femtoseconds [3, 6, 7]. Poisson noise was included, but no readout noise or systematic background was applied to the digitized diffraction data. The diffraction patterns were calculated without including the flatness of the detector or a central hole from which the missing Fourier components would need to be measured with an alternate detector. With noise included, the Bragg peaks in the low-frequency region contain sufficient intensity to allow the angle of the lattice relative to the beam to be calculated. The diffraction patterns are not processed prior to ptychographic reconstruction. The probe function was kept constant, and although we included a spatial ‘jitter’ in the probe by random translations, these positions were saved for use during reconstruction. As the probe moves with respect to the crystal lattice, we can see changes in the patterns of diffracted intensity between the Bragg peaks, which are due to coherent interference of the probe function with the illuminated object region (shown in figure 3). In addition, these variations are accompanied by the systematic disappearance and reappearance of some of the Bragg peaks, due to symmetry reasons, as the circular probe explores different regions of the square lattice beneath. We believe that these changes, which can be appreciable even with movement of the probe by only angstroms, will allow us to refine the probe positions from even very noisy diffraction patterns. We suggest that the probe positions relative to the reconstructed unit cell could then be refined from the experiment data by adapting existing likelihood optimization techniques [9, 10, 31] or using a refinement method [19]. In the case where one could obtain pulses of coherent radiation having identical amplitude and coherence, it will be possible to refine an initial estimated probe from the data, as shown by Thibault *et al* [14, 32]. One could account for intensity variations from pulse to pulse by scaling the diffraction patterns to the same average intensity, and it may also be possible to veto changes in the probe function from the data, since the diffraction pattern of the illumination function is superimposed on to each Bragg peak. Nonetheless, for efficient data collection, we would place a high priority on obtaining pulses from future XFEL sources that are (i) similar in amplitude and phase distribution, and (ii) short enough so that the wave exiting the object represents the true structure prior to radiation-induced movement of the charges.

### 3. Results and discussion

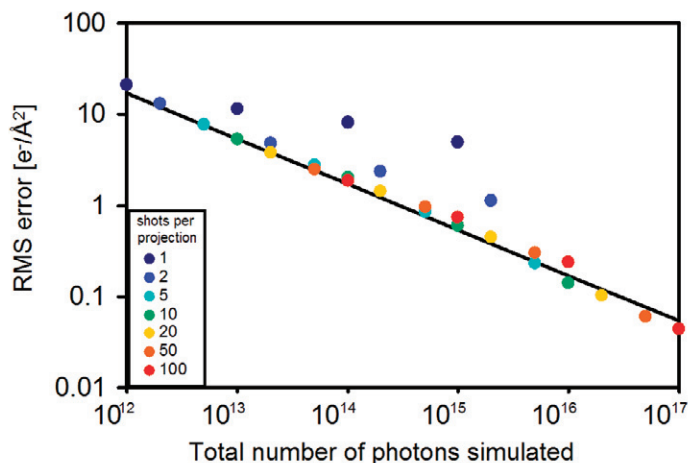
To reconstruct the exit wave at each projection angle  $\theta$ , all of the diffraction patterns were given as input to the ptychography algorithm [32], based on the difference map method [33]. After 100 iterations, assuming a constant illumination, the difference between iterations was negligible. Figures 4(a) and (b) show a comparison of the original and reconstructed projections. An algebraic reconstruction technique (ART) was used for the 3D reconstructions shown in figures 4(c) and (d) [34]. ART uses weighting factors to determine the contribution of each voxel in the reconstruction space to the measured data, and so ignores the ‘missing-wedge’. The solution process is complicated by noise, and the approximations used to efficiently estimate the weighting factors, so under-relaxation was employed to assist convergence. The relaxation parameter,  $\gamma$ , was assigned a value of the order of the inverse of the number of projections [35]. Excellent agreement can be seen in the projections, and the reconstructed slices show the major features of the original electron density map (figures 4(e)–(h)). The fidelity of the 3D map is slightly lower along the beam direction, due to the limited tilt angle range. The number of photons simulated for each projection in figure 4 corresponds to approximately 1000 shots, or approximately  $6 \times 10^{17}$  photons for the 3D structure. The reconstruction quality scales with the Poisson noise (see figure 5). One can conclude therefore that it is equivalent, under the



**Figure 4.** Original and reconstructed electron densities projected (a) at  $0^\circ$  and (b) at  $45^\circ$  to the  $Y$ -axis. The 3D electron density of the AQP1 tetramer is illustrated with a  $0.5e^- \text{ \AA}^{-3}$  isosurface, looking at (c) the  $XY$ -plane and (d) the  $XZ$ -plane. Corresponding slices from the 3D map are shown in (e) and (f). Dashed lines indicate the locations of line profiles shown in (g) and (h) for the projections and slices, respectively (red represents the original and blue the reconstructed values). The dimensions of the pixels and voxels in (a–f) and the data points in (g, h) are  $1 \text{ \AA}$ .

assumptions made, to increasing the photon flux per pulse, or the number of pulses used to create diffraction patterns for a given projection.

We have shown in figure 4, for simplicity, a single representative tetramer from the reconstruction. One can also exploit the periodicity of the sample by including a loop in the



**Figure 5.** The rms error between the reconstructed and original projections as a function of the total number of photons simulated. The log–log trend indicated by the black line is as expected, as the effect of Poisson noise decreases for a larger number of shots per projection or increased pulse intensity. The points well above the line indicate that the reconstruction is sub-optimal when only 1 or 2 shots per projection angle are used.

algorithm that averages over all periodic translations of the reconstruction grid. This assists convergence and decreases the rms error in the reconstruction by approximately the square root of the number of illuminated unit cells. For the above simulation, this is equivalent to using 20–30 times as many photons per projection, because we incorporate the information from many unit cells into a single reconstruction.

#### 4. Conclusion

In summary, we have demonstrated the feasibility of obtaining the structure of membrane proteins from XFEL ptychography data. We focused on 2D crystals that confine the proteins into an array, boosting the diffraction signal and reducing the orientational freedom. The reconstruction quality can be improved either by increasing the pulse intensity or by taking more diffraction patterns per projection angle. Our method is based on repeated measurements of a periodic sample, so the experimental implementation will require trains of identical ultrashort XFEL pulses, and high-quality 2D membrane protein crystals that are at least as large as the focused x-ray beam.

#### Acknowledgments

We acknowledge the discussions with our colleagues at Paul Scherrer Institut: Andreas Menzel, Martin Dierolf, Clemens Schulze-Briese, Fritz Winkler and Xiao-Dan Li. This work was carried out at the Swiss Light Source, Paul Scherrer Institut, CH-5200 Villigen PSI, Switzerland. FP acknowledges support through the German DFG excellence cluster ‘Munich Advanced Photonics’.



## References

- [1] Henderson R 1995 *Q. Rev. Biophys.* **28** 171–93
- [2] Hite R K, Raunser S and Walz T 2007 *Curr. Opin. Struct. Biol.* **17** 389–95
- [3] Neutze R, Wouts R, van der Spoel D, Weckert E and Hajdu J 2000 *Nature* **406** 752–7
- [4] Shenoy G K and Stöhr J 2000 LCLS—the first experiments *Technical Report* SLAC-R-611
- [5] Materlik G and Tschentscher T 2001 TESLA—technical design report part V: x-ray free electron laser facility *Technical Report* DESY 2001-11; ECFA 2001-9
- [6] Jurek Z, Oszlányi G and Faigel G 2004 *Europhys. Lett.* **65** 491–7
- [7] Hau-Riege S P, London R A and Szoke A 2004 *Phys. Rev. E* **69** 051906
- [8] Chapman H N *et al* 2006 *Nat. Phys.* **2** 839–43
- [9] Elser V 2008 [arXiv:0709.3858v1](https://arxiv.org/abs/0709.3858v1)
- [10] Fung R, Shneerson V, Saldin D K and Ourmazd A 2009 *Nat. Phys.* **5** 64–7
- [11] Becker M and Weckert E 2004 *On the Possibility of Determining Structures of Membrane Proteins in Two-Dimensional Crystals using X-Ray Free Electron Lasers (Conformational Proteomics of Macromolecular Architecture vol 1)* (Singapore: World Scientific), chapter 6, pp 133–47
- [12] Walz T, Smith B L, Zeidel M L, Engel A and Agre P 1994 *J. Biol. Chem.* **269** 1583–6
- [13] Mancuso A P *et al* 2009 *Phys. Rev. Lett.* **102** 035502
- [14] Thibault P, Dierolf M, Menzel A, Bunk O, David C and Pfeiffer F 2008 *Science* **321** 379–83
- [15] Hegerl R and Hoppe W 1970 *Ber. Bunsen—Ges. Phys. Chem.* **74** 1148–54
- [16] Rodenburg J M and Faulkner H M L 2004 *Appl. Phys. Lett.* **85** 4795–7
- [17] Rodenburg J M, Hurst A C, Cullis A G, Dobson B R, Pfeiffer F, Bunk O, David C, Jefimovs K and Johnson I 2007 *Phys. Rev. Lett.* **98** 034801
- [18] Maiden A M and Rodenburg J M 2009 *Ultramicroscopy* **109** 1256–62
- [19] Guizar-Sicairos M and Fienup J R 2008 *Opt. Express* **16** 7264–78
- [20] Faulkner H and Rodenburg J 2005 *Ultramicroscopy* **103** 153–64
- [21] Bunk O, Dierolf M, Kynde S, Johnson I, Marti O and Pfeiffer F 2008 *Ultramicroscopy* **108** 481–7
- [22] Walz T, Hirai T, Murata K, Heymann J B, Mitsuoka K, Fujiyoshi Y, Smith B L, Agre P and Engel A 1997 *Nature* **387** 624–7
- [23] Murata K, Mitsuoka K, Hirai T, Walz T, Agre P, Heymann J B, Engel A and Fujiyoshi Y 2000 *Nature* **407** 599–605
- [24] Yamauchi K *et al* 2005 *Appl. Opt.* **44** 6927–32
- [25] Kewish C M, Assoufid L, Macrander A T and Qian J 2007 *Appl. Opt.* **46** 2010–21
- [26] Barty A, Soufli R, McCarville T, Baker S L, Pivovarov M J, Stefan P and Bionta R 2009 *Opt. Exp.* **17** 15508–19
- [27] Schroer C G *et al* 2005 *Appl. Phys. Lett.* **87** 124103
- [28] Mimura H *et al* 2007 *Appl. Phys. Lett.* **90** 051903
- [29] Hignette O, Cloetens P, Rostaing G, Bernard P and Morawe C 2005 *Rev. Sci. Instrum.* **76** 063709
- [30] Kang H C, Maser J, Stephenson G B, Liu C, Conley R, Macrander A T and Vogt S 2006 *Phys. Rev. Lett.* **96** 127401
- [31] Scheres S H W, Gao H, Valle M, Herman G T, Eggermont P P B, Frank J and Carazo J M 2007 *Nat. Methods* **4** 27–9
- [32] Thibault P, Dierolf M, Bunk O, Menzel A and Pfeiffer F 2009 *Ultramicroscopy* **109** 338–43
- [33] Elser V 2003 *J. Opt. Soc. Am. A* **20** 40–55
- [34] Andersen A H 1989 *IEEE Trans. Med. Imaging* **8** 50–5
- [35] Herman G T, Lent A and Lutz P H 1978 *Commun. ACM* **21** 152–8