Femtosecond protein nanocrystallography—data analysis methods

Richard A. Kirian[1], Xiaoyu Wang[1], Uwe Weierstall[1], Kevin E. Schmidt[1], John C. H. Spence[1]*, Mark Hunter[2], Petra Fromme[2], Thomas White[3], Henry N. Chapman[3,4], and James Holton[5,6]

[1]*Department of Physics, Arizona State University, Tempe, Arizona 85287 USA*
[2]*Department of Biochemistry, Arizona State Unviversity, Tempe, Arizona 85287 USA*
[3]*Center for Free Electron Laser Science DESY University of Hamburg, Notkestrasse 85, 22607, Hamburg, Germany*
[4]*University of Hamburg, Luruper Chaussee 149, Hamburg 22761, Germany*
[5]*Advanced Light Source, Lawrence Berkeley Laboratory, Berkeley, California 94720 USA*
[6]*Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158 USA*
**spence@asu.edu*

Abstract: X-ray diffraction patterns may be obtained from individual submicron protein nanocrystals using a femtosecond pulse from a free-electron X-ray laser. Many "single-shot" patterns are read out every second from a stream of nanocrystals lying in random orientations. The short pulse terminates before significant atomic (or electronic) motion commences, minimizing radiation damage. Simulated patterns for Photosystem I nanocrystals are used to develop a method for recovering structure factors from tens of thousands of snapshot patterns from nanocrystals varying in size, shape and orientation. We determine the number of shots needed for a required accuracy in structure factor measurement and resolution, and investigate the convergence of our Monte-Carlo integration method.

©2010 Optical Society of America

OCIS codes: (260.1960) Diffraction theory; (290.5840) Scattering, molecules; (320.7100) Ultrafast measurements.

References and links

1. H. N. Chapman, "X-ray imaging beyond the limits," Nat. Mater. **8**(4), 299–301 (2009).
2. M. R. Howells, T. Beetz, H. N. Chapman, C. Cui, J. M. Holton, C. J. Jacobsen, J. Kirz, E. Lima, S. Marchesini, H. Miao, D. Sayre, D. A. Shapiro, J. C. H. Spence, and D. Starodub, "An assessment of the resolution limitation due to radiation-damage in X-ray diffraction microscopy," J. Electron Spectrosc. Relat. Phenom. **170**(1-3), 4–12 (2009).
3. R. B. G. Ravelli, and E. F. Garman, "Radiation damage in macromolecular cryocrystallography," Curr. Opin. Struct. Biol. **16**(5), 624–629 (2006).
4. L. Strüder, S. Epp, D. Rolles, R. Hartmann, P. Holl, G. Lutz, H. Soltau, R. Eckart, C. Reich, K. Heinzinger, C. Thamm, A. Rudenko, F. Krasniqi, K.-U. Kühnel, C. Bauer, C.-D. Schröter, R. Moshammer, S. Techert, D. Miessner, M. Porro, O. Hälker, N. Meidinger, N. Kimmel, R. Andritschke, F. Schopper, G. Weidenspointner, A. Ziegler, D. Pietschner, S. Herrmann, U. Pietsch, A. Walenta, W. Leitenberger, C. Bostedt, T. Möller, D. Rupp, M. Adolph, H. Graafsma, H. Hirsemann, K. Gärtner, R. Richter, L. Foucar, R. L. Shoeman, I. Schlichting, and J. Ullrich, "Large-format, high-speed, X-ray pnCCDs combined with electron and ion imaging spectrometers in a multipurpose chamber for experiments at 4th generation light sources," Nucl. Instrum. Methods. in press.
5. P. Denes, D. Doering, H. A. Padmore, J. P. Walder, and J. Weizeorick, "A fast, direct x-ray detection charge-coupled device," Rev. Sci. Instrum. **80**(8), 083302 (2009).
6. M. J. Bogan, W. H. Benner, S. Boutet, U. Rohner, M. Frank, A. Barty, M. M. Seibert, F. Maia, S. Marchesini, S. Bajt, B. Woods, V. Riot, S. P. Hau-Riege, M. Svenda, E. Marklund, E. Spiller, J. Hajdu, and H. N. Chapman, "Single particle X-ray diffractive imaging," Nano Lett. **8**(1), 310–316 (2008).
7. D. P. DePonte, U. Weierstall, K. Schmidt, J. Warner, D. Starodub, J. C. H. Spence, and R. B. Doak, "Gas dynamic virtual nozzle for generation of microscopic droplet streams," J. Phys. D Appl. Phys. **41**(19), 195505 (2008).
8. U. Weierstall, R. B. Doak, J. C. H. Spence, D. Starodub, D. Shapiro, P. Kennedy, J. Warner, G. G. Hembree, P. Fromme, and H. N. Chapman, "Droplet streams for serial crystallography of proteins," Exp. Fluids **44**(5), 675–689 (2008).

#121942 - $15.00 USD Received 23 Dec 2009; revised 19 Feb 2010; accepted 20 Feb 2010; published 5 Mar 2010

9. D. A. Shapiro, H. N. Chapman, D. Deponte, R. B. Doak, P. Fromme, G. Hembree, M. Hunter, S. Marchesini, K. Schmidt, J. Spence, D. Starodub, and U. Weierstall, "Powder diffraction from a continuous microjet of submicrometer protein crystals," J. Synchrotron Radiat. **15**(Pt 6), 593–599 (2008).
10. S. Marchesini, S. Boutet, A. E. Sakdinawat, M. J. Bogan, S. Bajt, A. Barty, H. N. Chapman, M. Frank, S. P. Hau-Riege, A. Szöke, C. W. Cui, D. A. Shapiro, M. R. Howells, J. C. H. Spence, J. W. Shaevitz, J. Y. Lee, J. Hajdu, and M. M. Seibert, "Massively parallel X-ray holography," Nat. Photonics **2**(9), 560–563 (2008).
11. S. Boutet, M. J. Bogan, A. Barty, M. Frank, W. H. Benner, S. Marchesini, M. M. Seibert, J. Hajdu, and H. N. Chapman, "Ultrafast soft x-ray scattering and reference-enhanced diffractive imaging of weakly scattering nanoparticles," J. Electron Spectrosc. Relat. Phenom. **166–167**, 65–73 (2008).
12. F. Coulibaly, E. Chiu, K. Ikeda, S. Gutmann, P. W. Haebel, C. Schulze-Briese, H. Mori, and P. Metcalf, "The molecular organization of cypovirus polyhedra," Nature **446**(7131), 97–101 (2007).
13. T. Warne, M. J. Serrano-Vega, J. G. Baker, R. Moukhametzianov, P. C. Edwards, R. Henderson, A. G. W. Leslie, C. G. Tate, and G. F. X. Schertler, "Structure of a beta1-adrenergic G-protein-coupled receptor," Nature **454**(7203), 486–491 (2008).
14. C. Nave, and M. A. Hill, "Will reduced radiation damage occur with very small crystals?" J. Synchrotron Radiat. **12**(Pt 3), 299–303 (2005).
15. S. P. Hau-Riege, R. A. London, and A. Szoke, "Dynamics of biological molecules irradiated by short x-ray pulses," Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **69**(5 Pt 1), 051906 (2004).
16. H. T. Witt, I. Witt, N. Krauss, W. Hinrichs, P. Fromme, and W. Saenger, "Crystals and structure of photosystem-1," Biophys. J. **66**, A2–A3 (1994).
17. I. Witt, H. T. Witt, D. Difiore, M. Rogner, W. Hinrichs, W. Saenger, J. Granzin, C. Betzel, and Z. Dauter, ""X-ray characterization of single-crystals of the reaction center-I of water-splitting photosynthesis," Ber. Bunsen-Ges," Phys. Chem. Chem. Phys. **92**, 1503–1506 (1988).
18. N. Krauss, W. D. Schubert, O. Klukas, P. Fromme, H. T. Witt, and W. Saenger, "Photosystem I at 4 A resolution represents the first structural model of a joint photosynthetic reaction centre and core antenna system," Nat. Struct. Biol. **3**(11), 965–973 (1996).
19. P. Jordan, P. Fromme, H. T. Witt, O. Klukas, W. Saenger, and N. Krauss, "Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution," Nature **411**(6840), 909–917 (2001).
20. A. G. W. Leslie, "The integration of macromolecular diffraction data," Acta Crystallogr. D Biol. Crystallogr. **62**(Pt 1), 48–57 (2006).
21. E. J. W. Whittaker, "The polarization factor for inclined-beam photography using crystal-reflected radiation," Acta Crystallogr. **6**(2), 222–223 (1953).
22. S. Bailey; Collaborative Computational Project, Number 4, "The CCP4 suite: programs for protein crystallography," Acta Crystallogr. D Biol. Crystallogr. **50**(Pt 5), 760–763 (1994).
23. H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J. D. Westbrook, and C. Zardecki, "The protein data bank," Acta Crystallogr. D Biol. Crystallogr. **58**(Pt 6 No 1), 899–907 (2002).
24. D. E. Tronrud, "TNT refinement package," Methods Enzymol. **277**, 306–319 (1997).
25. G. Hura, J. M. Sorenson, R. M. Glaeser, and T. Head-Gordon, "A high-quality x-ray scattering experiment on liquid water at ambient conditions," J. Chem. Phys. **113**(20), 9140–9148 (2000).
26. G. Hura, Advanced Light Source, Lawrence Berkeley Laboratory, Berkeley, Ca., 94720 USA (personal communication, 2009).
27. N. T. D. Loh, and V. Elser, "Reconstruction algorithm for single-particle diffraction imaging experiments," Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **80**(2 Pt 2), 026705 (2009).
28. R. Fung, V. Shneerson, D. K. Saldin, and A. Ourmazd, "Structure from fleeting illumination of faint spinning obects in flight," Nat. Phys. **5**(1), 64–67 (2009).
29. J. C. H. Spence, "Diffractive (lensless) imaging," in *Science of Microscopy*, P. Hawkes and J. C. H. Spence eds. (Springer 2007), pp. 1196–1227.
30. I. Robinson, and R. Harder, "Coherent X-ray diffraction imaging of strain at the nanoscale," Nat. Mater. **8**(4), 291–298 (2009).

1. Introduction

The invention of the hard X-ray free-electron laser (FEL) has created the opportunity for experiments in coherent diffractive (lensless) imaging and holography at nanometer resolution [1]. With a transverse coherence width which exceeds the beam diameter, and a beam focus of a few microns diameter, it has become possible to record coherent hard-X-ray diffraction patterns from individual micron-sized particles with a resolution of a few tens of nanometer, while the achievement of atomic resolution may be possible in the future. This work extends previous work using soft X-rays [1,2] in which iterative digital methods are used to solve the phase problem for scattering from a small particle, so that a real-space image of a projection of the particle may be reconstructed.

#121942 - $15.00 USD Received 23 Dec 2009; revised 19 Feb 2010; accepted 20 Feb 2010; published 5 Mar 2010
(C) 2010 OSA

The resolution in diffractive imaging may be considerably improved if multiple copies of the particle or molecule are available in the form of nanocrystals, in which case the intensity of the Bragg scattering which results is proportional to the square of the number of molecules, greatly strengthening the high angle scattering. This makes possible a new form of protein nanocrystallography, which we discuss in this paper. Using a focused beam a few microns in diameter and an X-ray pulse duration of a few femtoseconds, the FEL X-ray source also allows diffraction patterns to be obtained from sub-micron particles under an entirely new regime of radiation damage considerations [2,3]. In this "diffract-and-destroy" mode [3], it is found that a sufficiently brief X-ray pulse will terminate before atomic (and possibly electron) motion commences. Thereby, it should be possible to outrun the associated radiation damage effects, each of which has an associated time-scale longer than the X-ray pulse. Each pulse delivers about 10^{12} photons, with a peak brightness many orders of magnitude higher than a modern third-generation synchrotron source, and will be sufficient to provide a useful diffraction pattern from protein crystallites of sub-micron dimensions. At the US Department of Energy's Linac Coherent Light Source (LCLS) at Stanford, these pulses are generated with a repetition rate of up to 120 Hz, which, with an efficient protein crystallite delivery system and new fast-readout area detectors [4,5], will allow the collection of hundreds of thousands of snapshot diffraction patterns in a matter of hours.

The development of a suitable device for the injection of hydrated proteins or protein crystallites has proven a difficult problem, however, at least two devices now show promise in preliminary trials, as described elsewhere in the literature [6–9]. These injectors deliver a beam of hydrated bioparticles or nanocrystals across the X-ray beam with hit rates varying between one hit in tens of seconds to many per second, and have been used to produce diffraction patterns from single viruses, bioparticles, and cells in developmental tests at the soft X-ray free-electron laser (FEL) facility FLASH at DESY in Hamburg, Germany [6]. Images have also been reconstructed from a fixed single-cell sample, using holographic [10] and phase-retrieval methods [11]. Data collection rates can be improved in the future by synchronization of the injected particles with the FEL pulses and further improvements in time resolution of the detector.

In this paper, we consider the application of this method to protein nanocrystallography, and in particular, discuss methods of analysis for this new kind of data. We assume that individual X-ray diffraction patterns will be collected from a stream of hydrated submicron protein crystallites with significant size and shape variation. We assume that each X-ray pulse produces a diffraction pattern from a single crystallite, resulting in a data set which consists of hundreds of thousands of diffraction patterns from randomly oriented crystallites recorded under "snap-shot" conditions (so-called "stills"). Each pattern is not angle-integrated across the Bragg reflections, but is affected by beam divergence, energy spread, and broadening by the small size and possible lattice imperfections of the crystals. We assume that the pulse duration is sufficiently short that no radiation damage effects occur, and that the crystallites are too small to be affected by extinction (multiple scattering). Our aim here is to demonstrate that a complete set of structure factors, equivalent to single-crystal structure factors, may be obtained from these snapshot diffraction patterns, which in turn will allow the reconstruction of an electron density map. If the data from these crystallites is sufficiently finely sampled (and if crystals of closely similar size and shape were used), we note that images of the entire crystallite could be reconstructed using iterative phasing techniques. By selecting the phases only on lattice sites, a phased density map of one molecule could then be obtained. This would provide a new and useful solution to the phase problem for protein nanocrystallography.

Since the beam diameter at the LCLS is of micron dimensions, the method of snapshot protein nanocrystallography discussed here will allow study of protein crystals much smaller than those which have thus far been analyzed at conventional synchrotrons. Recent work using X-ray micro diffraction has already demonstrated the opportunity this provides to

#121942 - $15.00 USD Received 23 Dec 2009; revised 19 Feb 2010; accepted 20 Feb 2010; published 5 Mar 2010
(C) 2010 OSA

expand the range of proteins for which X-ray structure determination is possible to molecules that can only be grown as small crystals [12]. Data have recently been collected with great difficulty to obtain the first structure of a medically important human membrane protein, the beta-androgenic receptor, grown in a lipid cubic phase [13]. The reduction in damage observed with small crystallites is attributed to the ejection of photoelectrons into the surrounding vacuum rather than into surrounding bulk protein crystal, where further damage may be caused, as predicted by Nave and Hill [14]. The use of femtosecond "snapshot" diffraction patterns from even smaller crystallites supplied continuously and automatically as a stream of hydrated nanocrystals, would circumvent problems associated with radiation damage [15] and the long data collection times associated with conventional non-automated searches for protein crystals within a cryo-loop sample holder.

The protein chosen for the first protein nanocrystallography work at the LCLS is Photosystem I (PSI) from the cyanobacterium *T. elongatus* (PDB code 1JBO). The protein has already been shown to be capable of producing crystals that can be as small as 100 nm on a side [9]. The protein remains the largest membrane protein to have its structure solved to atomic resolution. However, the urgent need for a nanocrystallographic method arises from the fact that it took 13 years from the generation of the first micron-sized crystals of PSI to arrive at the atomic-resolution structure [16–19]. Obtaining diffraction data from such small crystals could provide a method to increase the rate at which structure determination takes place, especially amongst the membrane proteins, of which only around 200 unique structures have been determined.

2. Data analysis for crystallite snapshots

The approach to data analysis we consider in this paper is as follows. The orientation of each nanocrystal is determined from its diffraction pattern using automated indexing software such as the DPS FFT-based autoindexing algorithm implemented in MOSFLM [20]. Because of the small crystal size and full spatial coherence of the FEL beam, each Bragg reflection from a crystallite is shape-transformed and appears in the diffraction pattern as the intersection of the Fourier transform of the crystallite shape and the Ewald sphere. Peak locating algorithms may require some optimization for this type of diffraction data. In this work we demonstrate how existing automated indexing software may be used under these conditions. Having determined crystal orientations from tens of thousands of patterns, the diffracted intensities that fall within a small distance from a reciprocal lattice point are then summed. This distance is determined from the crystal orientation and experimental geometry. As the crystals will differ in size, shape and orientation, this summation can be expected to perform a Monte-Carlo integration of the intensity over these quantities near the Bragg condition, and so (if complete) produce a quantity proportional to the square of the structure factor magnitude. In this paper we test this integration hypothesis using simulated data, to study the convergence of this process and so determine the number of snapshot patterns needed for a given accuracy.

To do this, we consider a simplified parallelepiped protein crystal with an asymmetric unit consisting of just twelve electrons. We assume the same large hexagonal space-group $P6_3$ and cell constants ($a = b = 28.8$nm, $c = 16.7$nm) as for the PSI crystallites which will be used in subsequent experimental work at the LCLS. Since PSI contains about 70,000 non-hydrogen atoms, this simplified model allows the necessary speed increase in our simulations for crystals of finite size. Since the reciprocal space lattice, symmetry and forbidden reflections are unchanged, only the relative Bragg intensities (and overall scale) will be affected by this simplification.

For plane-polarized monochromatic incident radiation with wave vector $\mathbf{k}_i$ ($|\mathbf{k}_i| = 1/\lambda$) and negligible beam divergence, the diffracted photon flux I (counts/pulse) at $\mathbf{\Delta k} = \mathbf{k}_i - \mathbf{k}_o$ produced by the *n*-th parallelepiped crystallite, consisting of $N = N_1 \times N_2 \times N_3$ unit cells, is given in the kinematic theory as

#121942 - $15.00 USD Received 23 Dec 2009; revised 19 Feb 2010; accepted 20 Feb 2010; published 5 Mar 2010
(C) 2010 OSA

$$I_n(\Delta\mathbf{k},\mathbf{k}_o,\alpha,\beta,\gamma,N_i) = J_o \mid F(\Delta\mathbf{k}) \mid^2 r_e^2 P(\mathbf{k}_o) \frac{\sin^2(N_1\Psi_1)}{\sin^2(\Psi_1)} \frac{\sin^2(N_2\Psi_2)}{\sin^2(\Psi_2)} \frac{\sin^2(N_3\Psi_3)}{\sin^2(\Psi_3)} \Delta\Omega, \tag{1}$$

where F($\Delta\mathbf{k}$) is the structure factor of the unit cell. J_o is the incident photon flux density (counts/pulse/area) and $\Delta\Omega$ is the solid angle subtended by a detector pixel. Here

$$\begin{aligned} \Psi_1 &= 2\pi a \sin(\theta)\cos(\alpha)/\lambda \\ \Psi_2 &= 2\pi b \sin(\theta)\cos(\beta)/\lambda \\ \Psi_3 &= 2\pi c \sin(\theta)\cos(\gamma)/\lambda, \end{aligned} \tag{2}$$

where θ is half the scattering angle, and α, β and γ define the crystal orientation as the angles which the scattering vector $\mathbf{\Delta k}$ makes with the directions of the real-space unit cell vectors **a**, **b** and **c**. $\mathbf{\Delta k}$ is defined by the position of the detector pixel and X-ray wavelength, and defines a point in reciprocal space where the Ewald sphere intersects the shape transform. r_e is the classical radius of the electron, equal to $2.82\text{x}10^{-5}$ Å. The X-ray radiation produced by the LCLS is plane polarized, so that the polarization factor for polarization along the unit vector $\hat{\mathbf{u}}$ becomes $P(\mathbf{k_o}) = 1 - |\hat{\mathbf{u}} \cdot \hat{\mathbf{k}}_\mathbf{o}|^2$ [21]. An angular integration over the triple product in Eq. (1) is proportional to $N_1N_2N_3$, and the volume of the crystal. At a Bragg condition, the triple product is equal to $N_1^2N_2^2N_3^2$ and the diffracted intensity is therefore proportional to the square of the number of electrons in the crystal.

In order to simulate realistic photon counts using our simplified twelve-electron point-scatterer protein model, we scale the structure factor F($\mathbf{\Delta k}$) calculated from our model so that it agrees on average with calculated structure factors of PSI. Our structure factor then becomes

$$F^{PSI}(\Delta\mathbf{k}) = A f_N(\sin\theta/\lambda)\exp(-B(\sin\theta/\lambda)^2)F(\Delta\mathbf{k}), \tag{3}$$

where

$$f_N(\sin\theta/\lambda) \approx 7\exp(-10.7(\sin\theta/\lambda)^2), \tag{4}$$

models the structure factor for nitrogen, which is a reasonable approximation for the average atomic scattering factor in proteins. Spot fading due to intrinsic crystal disorder is modeled by the Wilson *B* factor, which for PSI is 44.3 Å^2. The overall scale factor of A was determined by calculating PSI structure factors with the CCP4 Suite [22] program SFALL using the model deposited as 1jb0 in the Protein Data Bank [23] and a simple bulk solvent model [24]. A scale factor of A = 70 brings the 12-electron model structure factors into reasonable agreement with these PSI structure factors. This parameterization provides an approximately correct absolute scale for the diffracted intensity, allowing Poisson noise effects to be added, but does not include effects due to the water background. For scattering angles corresponding to distances larger than the interatomic distances in water, scattering from pure water is due to long-range fluctuations, and is nearly constant [25]. We model this water background scatter as

$$I_{bg}(\Delta\mathbf{k},\mathbf{k}_o) = J_o r_e^2 P(\mathbf{k}_o) N_W \mid f_W \mid^2 \Delta\Omega, \tag{5}$$

where N_w is the number of water molecules in the beam, and f_W = 2.57 electron equivalents is a scattering factor for water [26]. We assume that the crystallites are delivered to the beam in a 1μm^3 volume of water, and do not consider scattering from the detergent molecules, X-ray fluorescence or inelastic scattering.

#121942 - $15.00 USD Received 23 Dec 2009; revised 19 Feb 2010; accepted 20 Feb 2010; published 5 Mar 2010
(C) 2010 OSA

In order to merge intensities from equivalent reflections, we first determine the regions of each diffraction pattern which are in close proximity to a given Bragg reflection. To do this, we may first use autoindexing software to determine the matrix

$$\mathbf{A} = \begin{pmatrix} a^*_x & b^*_x & c^*_x \\ a^*_y & b^*_y & c^*_y \\ a^*_z & b^*_z & c^*_z \end{pmatrix}, \tag{6}$$

which specifies the crystal orientation (relative to the laboratory frame) through the reciprocal lattice vectors

$$\mathbf{a}^* = \frac{\mathbf{b}\times\mathbf{c}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}}, \quad \mathbf{b}^* = \frac{\mathbf{c}\times\mathbf{a}}{\mathbf{b}\cdot\mathbf{c}\times\mathbf{a}}, \quad \mathbf{c}^* = \frac{\mathbf{a}\times\mathbf{b}}{\mathbf{c}\cdot\mathbf{a}\times\mathbf{b}}. \tag{7}$$

Fractional Miller indices $\mathbf{h}_j^{frac}$ corresponding to each detector pixel *j* are then determined by the equation

$$\mathbf{h}_j^{frac} = \mathbf{A}^{-1}\Delta\mathbf{k}_j. \tag{8}$$

For a flat area detector, the wave vector transfer $\Delta\mathbf{k}_j$ corresponding to pixel *j* is

$$\begin{aligned} \Delta\mathbf{k}_{j,x} &= X / [\lambda\sqrt{X_j^2 + Y_j^2 + D^2}] \\ \Delta\mathbf{k}_{j,y} &= Y / [\lambda\sqrt{X_j^2 + Y_j^2 + D^2}] \\ \Delta\mathbf{k}_{j,z} &= D / [\lambda\sqrt{X_j^2 + Y_j^2 + D^2}] - 1/\lambda, \end{aligned} \tag{9}$$

where *D* is the sample-to-detector distance, X_j and Y_j are distances from the incident beam position on the detector plane, and λ is the photon wavelength. Corrections for detector tilts and rotation may also be applied. The nearest reciprocal lattice vector to each detector pixel is obtained by rounding the fractional Miller indices to the nearest integer values, $\mathbf{h}_j$, and inverting Eq. (8):

$$\mathbf{g}_j = \mathbf{A}\mathbf{h}_j. \tag{10}$$

The reciprocal-space distance δ_j between $\Delta\mathbf{k}_j$ and the nearest reciprocal lattice point is then

$$\delta_j = \mid \Delta\mathbf{k}_j - \mathbf{g}_j \mid. \tag{11}$$

For *m* crystallites, the integrated "experimental" Bragg reflected intensities are evaluated as

$$I_{hkl}^{\exp}(m,\delta_t) = \sum_{n=1}^{m} \sum_{\{j\}_{m,hkl,\delta_t}} I'_n(\Delta\mathbf{k}_j), \tag{12}$$

where $\{j\}_{m,hkl,\delta t}$ is the set of pixels in pattern m for which which $\mathbf{h}_j$ are the Miller indices *hkl*, and $\delta < \delta_t$. $I'_n(\Delta\mathbf{k}_i)$ is the diffracted intensity after background subtraction and correction for the polarization factor:

$$I'_n(\Delta\mathbf{k}_j) = \frac{I_n(\Delta\mathbf{k}_j) - I_{bg}(\Delta\mathbf{k}_j)}{P(\mathbf{k}_{o,j})\Delta\Omega_j}. \tag{13}$$

Equation (12) is equivalent to integrating the diffracted intensities which fall within a spherical volume centered about the reciprocal lattice vectors $\mathbf{g}_{hkl}$. Finally, we obtain our "experimental" structure factors $F_{hkl}^{exp}(m,\delta_t)$ by averaging over all diffracted intensities from equivalent reflections

#121942 - $15.00 USD Received 23 Dec 2009; revised 19 Feb 2010; accepted 20 Feb 2010; published 5 Mar 2010
(C) 2010 OSA

$$\left|F_{hkl}^{\exp}(m,\delta_t)\right|^2 = \frac{I_{hkl}^{\exp}(m,\delta_t)}{M_{hkl}(m,\delta_t)}, \quad M_{hkl}(m,\delta_t) = \sum_{n=1}^{m} \sum_{\{j\}_{m,hkl,\delta_t}} 1. \tag{14}$$

The quality of the merged data were assessed using a standard crystallographic R-factor. We evaluate

$$R(m,\delta_t) = \frac{\sum_{\{hkl\}} \left| \left|F_{hkl}^{calc}\right| - \eta \left|F_{hkl}^{\exp}(m,\delta_t)\right| \right|}{\sum_{\{hkl\}} \left|F_{hkl}^{calc}\right|}. \tag{15}$$

after merging of the m-th crystallite. Here $F_{hkl}^{exp}(m,\delta_t)$ is obtained from Eq. (13). The Bragg intensities calculated from our model are

$$| F_{hkl}^{calc} |^2 = J_o r_e^2 | F^{PSI}(\mathbf{g}_{hkl}) |^2, \tag{16}$$

and the scaling factor η is determined through least squares minimization.

3. Simulations

Diffracted intensities were simulated according to Eq. (1), assuming parameters expected in planned experiments at LCLS. Simulated patterns were carried out for crystals in random orientations with an isotropic distribution. The number of unit cells N_1, N_2, N_3 were randomly generated, with a Gaussian distribution corresponding to a mean length of 500 nm, and with a standard deviation of 10%. Since N_1, N_2, and N_3 were varied independently, the overall shape of the crystals varies in each pattern. The detector contains 1024x1024 80 μm pixels, a sample-to-detector distance of 5 cm, and we assume 100% quantum efficiency. The pulse fluence was chosen to be 10^{12} photons focused to a 3 μm spot, with a photon energy of 2keV (λ = 0.6 nm). The resolution at the side (not the corner) of the detector is 0.94 nm for these conditions. Figure 1(a) shows a typical simulation, clearly showing the shape transformed Bragg reflections and their asymmetrical character. Figure 1(b) provides an enlargement showing the shape transforms around each Bragg beam. We find that the DPS autoindexing algorithm, when called from MOSFLM crystallography data analysis software without modification, is capable of determining crystal cell constants with a 0.05% RMS error, and orientations with an RMS error of 0.06 degrees. The algorithm found the correct unit cell in 98% of cases after testing on 50,000 patterns. Although these results are remarkable and most likely sufficient for a successful Monte-Carlo integration, in our assessment of R-factors here we used the known crystal orientation matrices from our simulated patterns due to the substantial amount of time it takes for MOSFLM to index hundreds of thousands of patterns.

Figure 2 shows plots of the crystallographic R-factor plotted against number of crystallites in random orientations merged after indexing. Curves are also shown for crystallites of identical size and shape, and the effects of a Gaussian size distribution and of Poisson noise are also shown. In addition, curves are given for various values of the reciprocal-space integration volume defined by δ_t.

#121942 - $15.00 USD Received 23 Dec 2009; revised 19 Feb 2010; accepted 20 Feb 2010; published 5 Mar 2010
(C) 2010 OSA

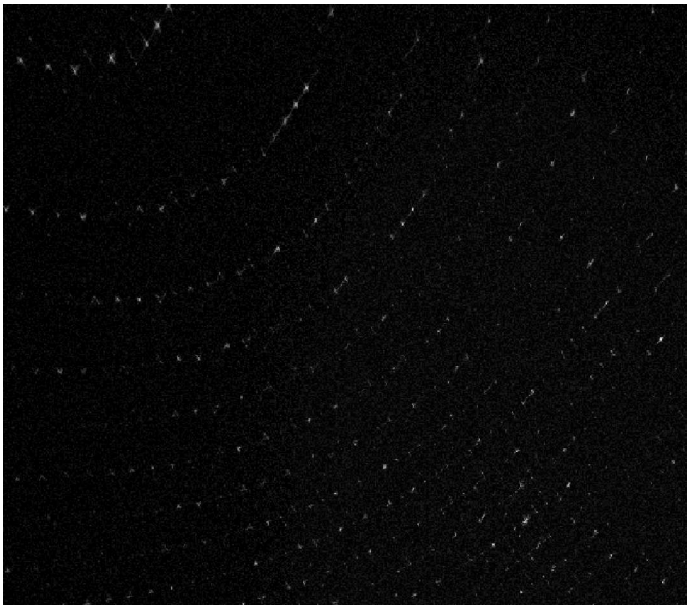

Fig. 1. (a). Log scaled simulation of a PS1 X-ray diffraction pattern for a 500nm crystallite at 2 keV, 1.5 mrad beam divergence, 0.1% FWHM energy spread, and background due to 1 μm^3 of water. Diffraction at edges corresponds to about 0.94 nm resolution. Polarization factor and detector pixel solid angles are included, and Poisson noise is included assuming that the detector counts single photons.

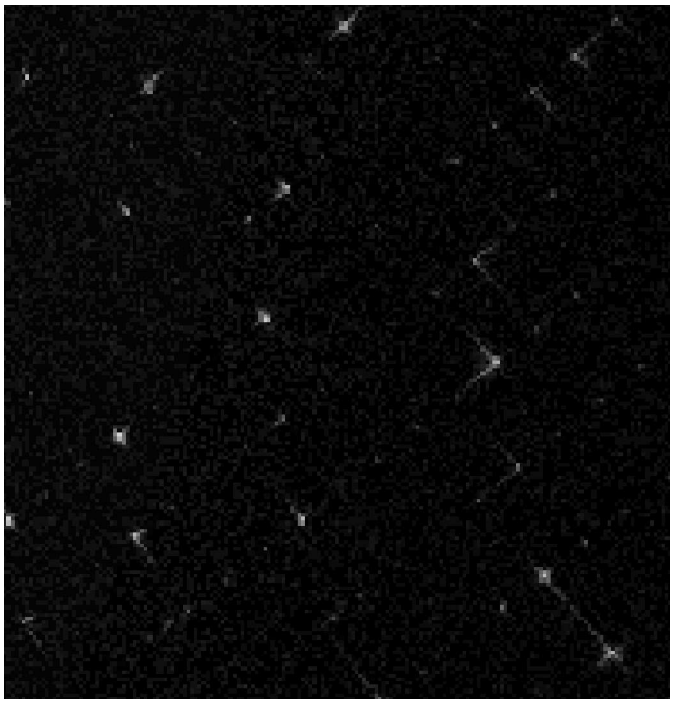

Fig. 1.(b). Enlargement of Fig. 1(a), showing shape-transforms around each lattice point.

#121942 - $15.00 USD Received 23 Dec 2009; revised 19 Feb 2010; accepted 20 Feb 2010; published 5 Mar 2010
(C) 2010 OSA

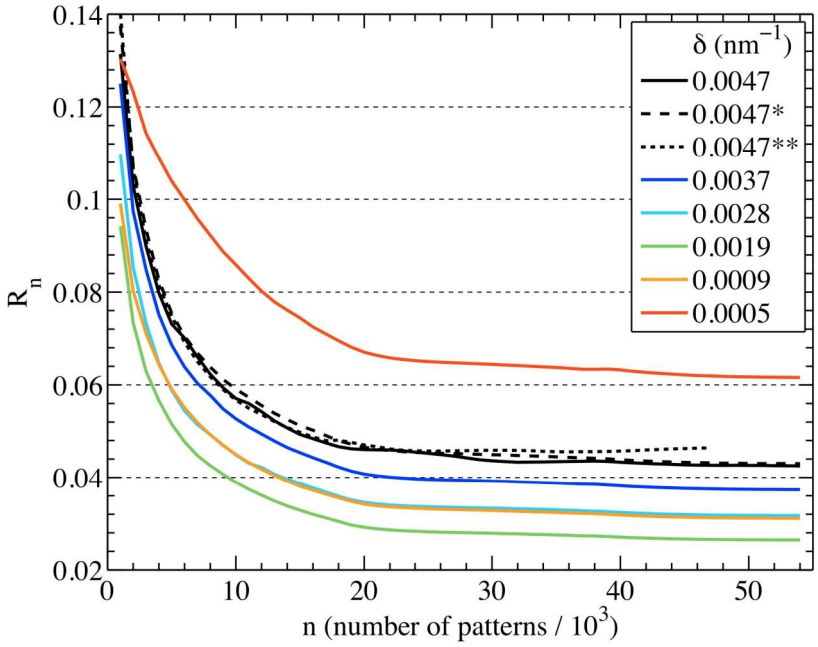

Fig. 2. Crystallographic R-factor plotted against number of crystallites in random orientations merged after indexing. Colors indicate the threshold value δ_t. Solid lines represent simulations including Gaussian size distribution, photon noise, and water background. Dashed lines are for comparison to simulations without photon noise (*) and without noise or size distribution (**).

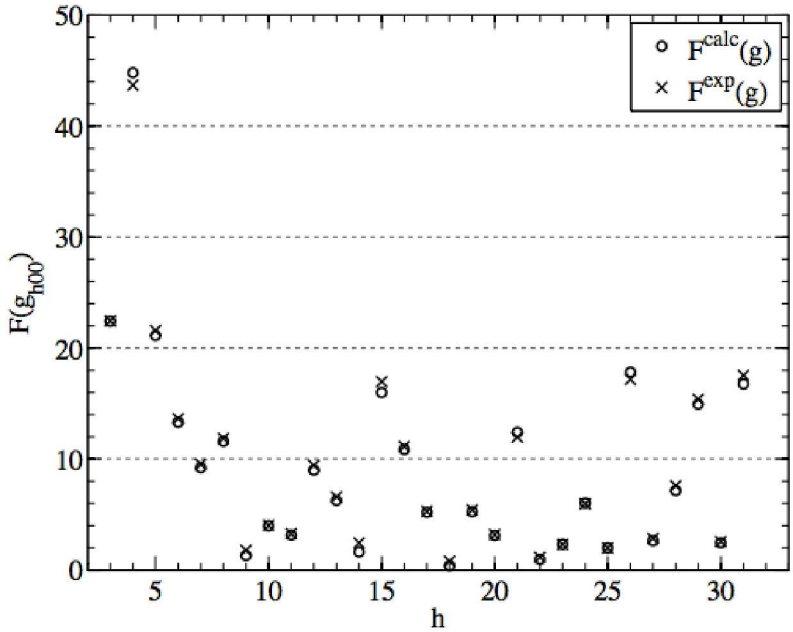

Fig. 3. Comparison of true structure factors (o) to the Monte-Carlo integrated structure factors (x) for a threshold value of $\delta_t = 0.0047$ nm^{-1}. Size variation and noise effects are included.

We see here that the optimal value for δ_t is ~0.0019 in our particular case; R factors get worse for deltas with larger or smaller values. For large δ_t, systematic errors due to counting near-zero valued pixels outside of the shape transform dominate, whereas small δ_t means

#121942 - $15.00 USD Received 23 Dec 2009; revised 19 Feb 2010; accepted 20 Feb 2010; published 5 Mar 2010
(C) 2010 OSA

slower convergence due to fewer pixels contributing to the average. In all but one case, the R-factor has fallen to less than 0.05 after 20,000 nanocrystals. The optimum value occurs when δ_t is matched to the mean size of the crystals $1/\delta_t$. We note that Poisson noise and the 10% crystal size distribution do not affect the convergence significantly because the shape transform dominates variance in diffracted intensities.

Figure 3 shows a comparison of true structure factors against those recovered from the Monte-Carlo integration over random orientations, with size variation and Poisson noise effects included. Figure 4 shows the number of unique reflections recorded in the simulations as a function of the number of crystallites for various values of delta. These simulations took about 36 hours for 50,000 crystallites using a Macintosh Pro, using a single 2.26 GHz CPU of the 8 available, in Matlab code. The time per pattern is about 3 seconds. By taking advantage of the multiple CPU and GPU processing available on these machines, this time could be greatly reduced. For the extraction of structure factors from experimental data, the time bottleneck is likely to be the indexing with MOSFLM, which takes a few seconds per pattern under the same conditions.

If the fall-off in scattering-factor and noise effects are ignored, we can estimate from geometry the number of crystals (orientations, diffraction patterns) needed to record all reflections out to 0.2 nm resolution under similar conditions. Since the number of reflections varies as the cube of the scattering angle, while the number of reflections per pattern varies as the square, we expect the number of patterns to vary approximately linearly with the scattering angle, or inversely as the resolution. Then about 10^5 crystallites would be needed for 0.2 nm resolution.

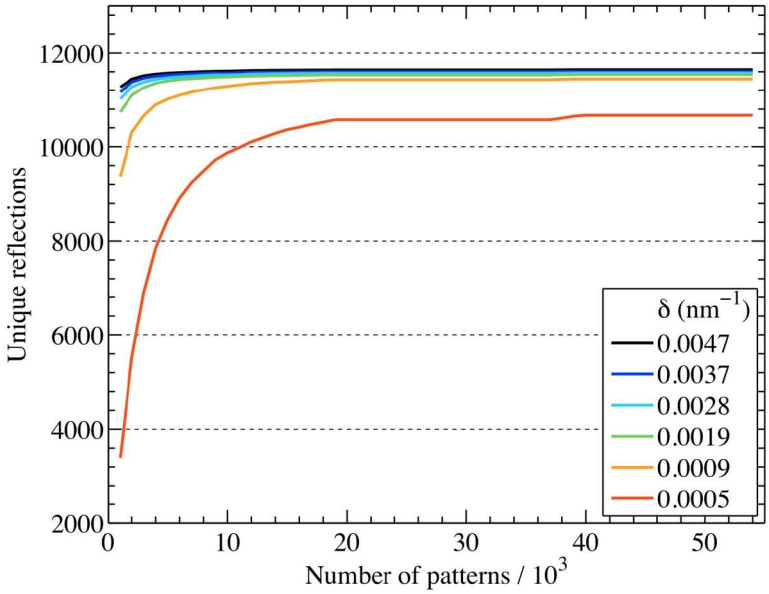

Fig. 4. The number of unique reflections recorded as a function of the number of crystallites. Colors indicate the threshold value δ.

4. Effects of beam divergence, energy spread, and a larger size distribution

The spectral width of the X-ray beam may be simulated using Eq. (1) by summing Gaussian-weighted diffracted intensities over a spread of wavelengths. Beam divergence can be approximated similarly by summing Gaussian-weighted intensities over a tilt series of the crystal (a more accurate model is to rock the detector about the crystal position). For our simulations of 500 nm crystals in this work, we neglect beam divergence and spectral width because, for small beam divergence, these effects do not result in a significant change in the

#121942 - $15.00 USD Received 23 Dec 2009; revised 19 Feb 2010; accepted 20 Feb 2010; published 5 Mar 2010
(C) 2010 OSA

diffraction patterns. For larger crystallites, these effects, along with crystal mosaicity, will become significant, and will likely result in faster convergence of the Monte-Carlo integration since these effects will partially integrate Bragg reflections.

Provided background is small and the reciprocal space volume of integration is larger than the central maximum of the shape transform of the smallest crystal, accurate results can be expected. However with a large background (for example arising from oxygen fluorescence from water) errors will arise since the integration volume is fixed, but the size of the shape transform varies. Pulse-height analysis (or an absorbing filter across the area detector) could be used to eliminate the lower energy oxygen fluorescence.

5. Conclusions

For crystallites with a mean size of half a micron immersed in a cubic micron of water, whose size distribution has a standard deviation of 10%, a total of about 20,000 diffraction patterns are needed in the presence of shot-noise to obtain an R-factor smaller than 0.05 and a resolution of 0.9nm. The number of crystallites increases approximately as the inverse of the resolution. We find that an optimum value of the integration volume in reciprocal space exists, equal to the reciprocal of the crystallite dimension, and that small R-values are obtained for a range around this value. (This corresponds to integration over the central maximum of the shape transform). A knowledge of the particle size distribution, obtained for example from dynamic light scattering measurements, or by powder diffraction analysis of the sum of all these nanocrystal patterns, would therefore be useful in setting the integration volume parameter.

For experimental data collected from PSI, the effects of twinning must be considered. The crystallites themselves are not physically twinned, but the possibility arises that data for two untwinned crystals may be merged in twin-related orientations. For the $P6_3$ space group of PSI, only one twin possibility arises, that of Merohedral twinning described by rotation of 180 degrees about the h = k axis. This operation brings the lattice into coincidence with itself, but not the associated structure factors. It converts index (h,k,l) in the hexagonal system into (k,h,-l). Since these reflections have different intensities, they may be identified prior to merging. In the simulations given here we have avoided this problem through the use of known crystal orientations. Alternatively, if a "twinned" data set is available, one can refine crystal orientation, size, etc against it. This puts all the observations on the same scale and creates a bimodal distribution for certain reflections, which will classify the crystals into "twin A" and "twin B". One then flips all the "twin B" indexing choices and re-merges data.

In the future, the LCLS is expected to produce shorter wavelength X-rays, and higher repetition rates. The LCLS will is expected to provide 1.5 Å wavelength X-rays within a year, which could open up the possibility for atomic resolution diffraction patterns to be recorded from submicron crystals, and, in addition, for pump-probe experiments on molecular dynamics. It remains to be determined whether the reconstruction of particular particle size classes will allow iterative phasing methods to be applied [27–30]. Higher repetition rates will allow for faster data collection, and for less sample loss. Under these conditions at shorter wavelength, the effects of diffraction from the water structure must also be considered.

Acknowledgements

Supported by Department of Energy (DOE) award DE-SC0002141, National Science Foundation (NSF) awards MCB 0919195 and 0417142. We are grateful to Drs. R. Doak, A. Barty, I. Schlichting, S. Marchesini, and F. Maia for useful discussions.

#121942 - $15.00 USD Received 23 Dec 2009; revised 19 Feb 2010; accepted 20 Feb 2010; published 5 Mar 2010
(C) 2010 OSA