Toward unsupervised single-shot diffractive imaging of heterogeneous particles using X-ray free-electron lasers

Hyung Joo Park,[1,2,*] N. Duane Loh,[3] Raymond G. Sierra,[3] Christina Y. Hampton,[3] Dmitri Starodub,[3] Andrew V. Martin,[4,5] Anton Barty,[5] Andrew Aquila,[6] Joachim Schulz,[6] Jan Steinbrener,[7] Robert L. Shoeman,[7] Lukas Lomb,[7] Stephan Kassemeyer,[7] Christoph Bostedt,[8] John Bozek,[8] Sascha W. Epp,[9,10] Benjamin Erk,[9,11] Robert Hartmann,[12] Daniel Rolles,[9,11] Artem Rudenko,[9,13] Benedikt Rudek,[9,14] Lutz Foucar,[7,9] Nils Kimmel,[15] Georg Weidenspointner,[15,16] Guenter Hauser,[15,16] Peter Holl,[12] Emanuele Pedersoli,[17] Mengning Liang,[5] Mark S. Hunter,[18] Lars Gumprecht,[5] Nicola Coppola,[6] Cornelia Wunderer,[11] Heinz Graafsma,[11] Filipe R.N.C. Maia,[19] Tomas Ekeberg,[20] Max Hantke,[20] Holger Fleckenstein,[5] Helmut Hirsemann,[11] Karol Nass,[7,21] Herbert J. Tobias,[22] George R. Farquar,[18] W. Henry Benner,[18] Stefan Hau-Riege,[18] Christian Reich,[12] Andreas Hartmann,[12] Heike Soltau,[12] Stefano Marchesini,[19] Sasa Bajt,[11] Miriam Barthelmess,[11] Lothar Strueder,[15] Joachim Ullrich,[14] Philip Bucksbaum,[3] Matthias Frank,[18] Ilme Schlichting,[7] Henry N. Chapman,[5,21] Michael J. Bogan,[3] and Veit Elser[1]

[1]*Cornell University, Laboratory of Atomic and Solid State Physics, Ithaca, NY 14853, USA*
[2]*Cornell University, Center for Applied Mathematics, Ithaca, NY 14853, USA*
[3]*PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025 USA*
[4]*ARC Centre of Excellence for Coherent X-ray Science, School of Physics, University of Melbourne, Parkville, Victoria 3010, Australia*
[5]*Center for Free-Electron Laser Science, DESY, Notkestraße 85, 22607 Hamburg, Germany.*
[6]*European X-Ray Free-Electron Laser Facility GmbH, Albert-Einstein-Ring 19, 22761 Hamburg, Germany*
[7]*Max-Planck-Institut für medizinische Forschung, Jahnstr. 29, 69120 Heidelberg, Germany*
[8]*SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025 USA*
[9]*Max Planck Advanced Study Group, CFEL, Notkestraße 85, 22607 Hamburg, Germany*
[10]*Max Planck Institute for Structural Dynamics, 22607 Hamburg, Germany*
[11]*Deutsches Elektronen Synchrotron (DESY), Notkestraße 85, 22607 Hamburg, Germany*
[12]*PNSensor GmbH, Römerstr. 28, 80803 München, Germany*
[13]*J.R. MacDonald Laboratory, Department of Physics, Kansas State University, Manhattan, KS 66506, USA*
[14]*Physikalisch-Technische Bundesanstalt, Bundesallee 100, D-38116 Braunschweig, Germany*
[15]*Max-Planck-Institut Halbleiterlabor, Otto-Hahn-Ring 6, 81739 München, Germany*
[16]*Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, 85741 Garching, Germany*
[17]*Elettra-Sincrotrone Trieste, SS 14 - km 163.5 in AREA Science Park, 34149 Basovizza, Trieste, Italy*
[18]*Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94551 USA*
[19]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[20]*Laboratory of Molecular Biophysics, Uppsala University, Husargatan 3 (Box 596), SE-751 24 Uppsala, Sweden*
[21]*Department of Physics, University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany*
[22]*Cornell University, Division of Nutritional Sciences, Savage Hall, Ithaca, NY 14853, USA*

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

*hp255@cornell.edu

Abstract: Single shot diffraction imaging experiments via X-ray free-electron lasers can generate as many as hundreds of thousands of diffraction patterns of scattering objects. Recovering the real space contrast of a scattering object from these patterns currently requires a reconstruction process with user guidance in a number of steps, introducing severe bottlenecks in data processing. We present a series of measures that replace user guidance with algorithms that reconstruct contrasts in an unsupervised fashion. We demonstrate the feasibility of automating the reconstruction process by generating hundreds of contrasts obtained from soot particle diffraction experiments.

OCIS codes: (140.2600) Free-electron lasers; (110.7440) X-ray imaging; (030.6600) Statistical optics; (050.1940) Diffraction; (110.1650) Coherence imaging; (110.1758) Computational imaging.

References and links

1. M. J. Bogan, W. H. Benner, S. Boutet, U. Rohner, M. Frank, A. Barty, M. M. Seibert, F. R. Maia, S. Marchesini, S. Bajt, B. W. 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).
2. D. Shapiro, P. Thibault, T. Beetz, V. Elser, M. Howells, C. Jacobsen, J. Kirz, E. Lima, H. Miao, A. M. Neiman, and D. Sayre, "Biologial imaging by soft x-ray diffraction microscopy," Proc. Natl. Acad. Sci. USA **102**(43), 15343–15346 (2005).
3. M. M. Seibert, T. Ekeberg, F. R. Maia, M. Svenda, J. Andreasson, O. Jönsson, D. Odić, B. Iwan, A. Rocker, D. Westphal, M. Hantke, D. P. DePonte, A. Barty, J. Schulz, L. Gumprecht, N. Coppola, A. Aquila, M. Liang, T. A. White, A. Martin, C. Caleman, S. Stern, C. Abergel, V. Seltzer, J. M. Claverie, C. Bostedt, J. D. Bozek, S. Boutet, A. A. Miahnahri, M. Messerschmidt, J. Krzywinski, G. Williams, K. O. Hodgson, M. J. Bogan, C. Y. Hampton, R. G. Sierra, D. Starodub, I. Andersson, S. Bajt, M. Barthelmess, J. C. H. Spence, P. Fromme, U. Weierstall, R. Kirian, M. Hunter, R. B. Doak, S. Marchesini, S. P. Hau-Riege, M. Frank, R. L. Shoeman, L. Lomb, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, C. Schmidt, L. Foucar, N. Kimmel, P. Holl, B. Rudek, B. Erk, A. Hömke, C. Reich, D. Pietschner, G. Weidenspointner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, I. Schlichting, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. U. Khnel, R. Andritschke, C. D. Schröter, F. Krasniqi, M. Bott, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, H. N. Chapman, and J. Hajdu, "Single mimivirus particles intercepted and imaged with an X-ray laser," Nature (London) **470**, 78–81 (2011).
4. H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz1, D. P. DePonte, U. Weierstall, R. B. Doak, F. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Hömke, C. Reich, D. Pietschner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. Kühnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. Seibert, J. Andreasson, A. Rocker, O. Jönsson, S. Stern, K. Nass, R. Andritschke, C. Schröter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. Holton, S. Marchesini, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, M. Svenda, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, and J. C. H. Spence, "Femtosecond X-ray protein nanocrystallography," Nature (London) **470**, 73–77 (2011).
5. S. Marchesini, "A unified evaluation of iterative projection algorithms for phase retrieval," Rev. Sci. Instrum. **78**, 011301 (2007).
6. M. J. Bogan, D. Starodub, C. Y. Hampton, and R. Sierra, "Single-particle coherent diffractive imaging with a soft X-ray free electron laser: towards soot aerosol morphology," J. Phys. B: At. Mol. Opt. Phys. **43**, 194013 (2010).
7. N. D. Loh, C. Y. Hampton, A. V. Martin, D. Starodub, R. G. Sierra, A. Barty, A. Aquila, J. Schulz, L. Lomb, J. Steinbrener, R. L. Shoeman, S. Kassemeyer, C. Bostedt, J. Bozek, S. W. Epp, B. Erk, R. Hartmann, D. Rolles, A. Rudenko, B. Rudek, L. Foucar, N. Kimmel, G. Weidenspointner, G. Hauser, P. Holl, E. Pedersoli, M. Liang, M. S. Hunter, L. Gumprecht, N. Coppola, C. Wunderer, H. Graafsma, F. R. Maia, T. Ekeberg, M. Hantke, H.

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

Fleckenstein, H. Hirsemann, K. Nass, T. A. White, H. J. Tobias, G. R. Farquar, W. H. Benner, S. P. Hau-Riege, C. Reich, A. Hartmann, H. Soltau, S. Marchesini, S. Bajt, M. Barthelmess, P. Bucksbaum, K. O. Hodgson, L. Strüder, J. Ullrich, M. Frank, I. Schlichting, H. N. Chapman and M. J. Bogan, "Fractal morphology, imaging and mass spectrometry of single aerosol particles in flight," Nature (London) **486**, 513–517 (2012).
8. M. Wentzel, H. Gorzawski, K.-H. Naumann, H. Saathoff, and S. Weinbruch, "Transmission electron microscopical and aerosol dynamical characterization of soot aerosols," J. Aerosol Sci. **34**, 1347–1370 (2003).
9. S. Schwyn, E. Garwin, and A. Schmidt-Ott, "Aerosol generation by spark discharge," J. Aerosol Sci. **19**, 639–642 (1988).
10. Standard Reference Material 2975, Diesel particulate matter (industrial forklift), certificate of analysis; National Institute of Standards & Technology, Gaithersburg, MD 20899, November 7, 2000.
11. M. J. Bogan, W. H. Benner, S. P. Hau-Riege, H. N. Chapman, and M. Frank, "Aerosol sample preparation methods for X-ray diffraction imaging," J. Aerosol Sci. **38**, 1119–1128 (2007).
12. 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 multi purpose chamber for experiments at 4th generation light sources," Nucl. Instrum. Meth. A **614**, 483–496 (2009).
13. C. H. Yoon, P. Schwander, C. Abergel, I. Andersson, J. Andreasson, A. Aquila, S. Bajt, M. Barthelmess, A. Barty, M. J. Bogan, C. Bostedt, J. Bozek, H. N. Chapman, J. M. Claverie, N. Coppola, D. P. DePonte, T. Ekeberg, S. W. Epp, B. Erk, H. Fleckenstein, L. Foucar, H. Graafsma, L. Gumprecht, J. Hajdu, C. Y. Hampton, A. Hartmann, E. Hartmann, R. Hartmann, G. Hauser, H. Hirsemann, P. Holl, S. Kassemeyer, N. Kimmel, M. Kiskinova, M. Liang, N. D. Loh, L. Lomb, F. R. Maia, A. V. Martin, K. Nass, E. Pedersoli, C. Reich, D. Rolles, B. Rudek, A. Rudenko, I. Schlichting, J. Schulz, M. Seibert, V. Seltzer, R. L. Shoeman, R. G. Sierra, H. Soltau, D. Starodub, J. Steinbrener, G. Stier, L. Strder, M. Svenda, J. Ullrich, G. Weidenspointner, T. A. White, C. Wunderer, and A. Ourmazd, "Unsupervised classification of single-particle X-ray diffraction snapshots by spectral clustering," Opt. Express **19**(17), 16542–16549 (2011).
14. N. D. Loh, D. Starodub, L. Lomb, C. Y. Hampton, A. V. Martin, R. G. Sierra, A. Barty, A. Aquila, J. Schulz, J. Steinbrener, R. L. Shoeman, S. Kassemeyer, C. Bostedt, J. Bozek, S. W. Epp, B. Erk, R. Hartmann, D. Rolles, A. Rudenko, B. Rudek, L. Foucar, N. Kimmel, G. Weidenspointner, G. Hauser, P. Holl, E. Pedersoli, M. Liang, M. S. Hunter, L. Gumprecht, N. Coppola, C. Wunderer, H. Graafsma, F. R. Maia, T. Ekeberg, M. Hantke, H. Fleckenstein, H. Hirsemann, K. Nass, T. A. White, H. J. Tobias, G. R. Farquar, W. H. Benner, S. Hau-Riege, C. Reich, A. Hartmann, H. Soltau, S. Marchesini, S. Bajt, M. Barthelmess, L. Strueder, J. Ullrich, P. Bucksbaum, M. Frank, I. Schlichting, H. N. Chapman, and M. J. Bogan, "Sensing the wavefront of X-ray free-electron lasers using aerosol spheres," Opt. Express **21**, 12385–12394 (2013).
15. V. Elser, "Phase retrieval by iterated projections," J. Opt. Soc. Am. A **20**, 40–55 (2003).
16. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
17. N. D. Loh, S. Eisebitt, S. Flewett, and V. Elser, "Recovering magnetization distributions from their noisy diffraction data," Phys. Rev. E **82**, 061128 (2010).
18. A. V. Martin, F. Wang, N. D. Loh, T. Ekeberg, F. R. Maia, M. Hantke, G. van der Schot, C. Y. Hampton, R. G. Sierra, A. Aquila, S. Bajt, M. Barthelmess, C. Bostedt, J. D. Bozek, N. Coppola, S. W. Epp, B. Erk, H. Fleckenstein, L. Foucar, M. Frank, H. Graafsma, L. Gumprecht, A. Hartmann, R. Hartmann, G. Hauser, H. Hirsemann, P. Holl, S. Kassemeyer, N. Kimmel, M. Liang, L. Lomb, S. Marchesini, K. Nass, E. Pedersoli, C. Reich, D. Rolles, B. Rudek, A. Rudenko, J. Schulz, R. L. Shoeman, H. Soltau, D. Starodub, J. Steinbrener, F. Stellato, L. Strüder, J. Ullrich, G. Weidenspointner, T. A. White, C. B. Wunderer, A. Barty, I. Schlichting, M. J. Bogan, and H. N. Chapman, "Noise-robust coherent diffractive imaging with a single diffraction pattern, " Opt. Express **20**(15), 16650–16661 (2012).
19. D. R. Luke, "Relaxed averaged alternating reflections for diffraction imaging," Inverse Probl. **21**, 37–50 (2005).
20. S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. H. Spence, "X-ray image reconstruction from a diffraction pattern alone," Phys. Rev. B **68**(14), 140101 (2003).
21. P. Thibault, V. Elser, C. Jacobsen, D. Shapiro and D. Sayre, "Reconstruction of a yeast cell from X-ray diffraction data," Acta Crystallogr. A **62**, 248–261 (2006).
22. H. N. Chapman, A. Barty, M. J. Bogan, S. Boutet, M. Frank, S. P. Hau-Riege, S. Marchesini, B. W. Woods, S. Bajt, W. H. Benner, R. A. London, E. Plonjes, M. Kuhlmann, R. Treusch, S. Düsterer, T. Tschentscher, J. R. Schneider, E. Spiller, T. Möller, C. Bostedt, M. Hoener, D. A. Shapiro, K. O. Hodgson, D. van der Spoel, F. Burmeister, M. Bergh, C. Caleman, G. Huldt, M. M. Seibert, F. R. Maia, R. W. Lee, A. Szoke, N. Timneanu, and J. Hajdu, "Femtosecond diffractive imaging with a soft-X-ray free-electron laser," Nat. Phys. **2**, 839–843 (2006).
23. F. R. Maia, T. Ekeberg, D. van der Spoel, J. Hajdu, "Hawk: the image reconstruction package for coherent X-ray diffractive imaging," J. Appl. Crystallogr. **43**, 1535–1539 (2010).

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

1. Introduction

Single-shot diffraction imaging via X-ray free-electron lasers [1] has emerged as a potentially significant tool for studying particles in the nanometer regime, from biological samples [2, 3] to nanocrystals [4]. As particles of interest are propelled into the path of short X-ray pulses of high fluence such as those generated at the Linac Coherent Light Source (LCLS), their interaction diffracts a small fraction of the photons off the particle before the onset of significant radiation damage. The resulting far-field diffraction patterns, recorded on X-ray detectors, can be used to reconstruct real space contrasts of the diffracting particles via iterative phase retrieval methods [5]. Since the pulse-particle interactions occur mid-flight, imaging individual particulate matter such as soot *in situ* at nanometer resolution is made possible, allowing for morphological studies of in-flight particles [6, 7] that have in the past relied on other imaging techniques such as transmission electron microscopy [8], where substrate deposition could potentially alter particles' morphologies.

A typical imaging experiment could generate hundreds of thousands of usable diffraction patterns in a single day, compelling the need for an unsupervised contrast reconstruction process requiring minimal user guidance. Due to experimental realities such as variable pulse profiles and detector noise, however, there are significant differences in the quality of the data from pattern to pattern. Simply automating existing tools for reconstruction would be problematic, and in some cases insufficient, especially when these tools rely heavily on human supervision. Any scalable protocol ought to replace user guidance and visual inspection with efficient unsupervised algorithms.

In this paper, we present a series of measures that aim to facilitate the unsupervised contrast reconstruction of a large collection of single shot diffraction patterns. We identified steps during the reconstruction process that require user guidance and replaced them with reasonable algorithms. Through these measures, we were able to successfully reconstruct hundreds of contrasts with minimal guidance.

2. Experiment and data set

In this study we worked with highly variable diffraction patterns of soot particles of multiple length scales. Data was collected at the Atomic, Molecular and Optical Science beam line at the LCLS. Two different kinds of soot particles were considered for imaging: particles created by a Palas GFG100 spark source generator [9] and NIST 2975 diesel soot particles [10]. In separate runs, the Palas and NIST soot were propelled into the path of X-ray pulses by a differentially pumped aerodynamic focusing inlet [11]. Some of the particles, when they reached the interaction region with a velocity of $100-200$ m/s, were intercepted by a single X-ray pulse focused to an area of about $10\mu\text{m}^2$ with an average fluence of 4×10^{12} photons, each with 1.24 keV of energy, per pulse, assuming a transmission efficiency of 20%. The scattered photons were recorded on a pair of pn-junction charge coupled device (pnCCD) panels installed in the CFEL ASG Multi-Purpose (CAMP) instrument [12]. Each panel contained 512×1024 pixels, each of area $75\times75\mu\text{m}^2$. A gap of 1.6mm between the panels and semicircular cutouts of 1.2 mm diameter allowed for the passage of the pulse into a beam dump. Further details can be found in Loh *et al.* [7].

Pulse generation and the detector readout rate coincided at 60 Hz, allowing for a theoretically maximum data collection rate of 2.2×10^5 patterns per hour. In practice, however, the sample hit rate was lower than the pulse generation rate because of random particle injection. In the soot experiments, the hit rate was observed, on average, to be 0.09 Hz. A total of 953 successful hits were identified and considered for reconstruction.

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

3. Practical considerations

Any collection of diffraction patterns, even when sorted and classified [13], is bound to show differences in quality from pattern to pattern. The pulse-particle interaction contributes to much of this variability, as the profile of individual pulses can differ from each other. Wavefront aberrations in pulses can result in randomly shifted diffraction patterns which need to be corrected [14]. Also, as each pulse's transverse profile cannot be assumed to be a planar wave of constant intensity, the position of each randomly injected particle during the pulse-particle interaction, relative to the focus, has a noticeable effect on the signal-to-noise ratio of the resulting pattern, a value that is already affected by the variability in the pulse fluence.

Given the incomplete nature of phasing algorithms, successful contrast reconstruction is not guaranteed within a set number of iterations, even for ideal, noiseless patterns. As a result, reconstructions from the same pattern can exhibit significant differences. Noise can further frustrate the phasing process and encourage significant variability between reconstructions. A reliable check for the confidence in a reconstruction is how often the algorithm will arrive at the same, or similar solution, beginning from different initial conditions. This step, often done visually, can be a major bottleneck in the reconstruction process.

In this section, we introduce techniques to address these various issues. In the first part, we discuss ways of exploiting centrosymmetry to correctly center each pattern. In the next part, we highlight a noise robust phasing algorithm that can handle patterns with low signal-to-noise ratios. Then, we propose a technique for assessing the reliability of a reconstruction algorithmically. Lastly, we suggest a strategy to check the degree to which the missing data region could affect the final reconstruction.

3.1. Centering the diffraction pattern

While the detailed form of each X-ray pulse is lost as soon as it is absorbed into the beam dump, the pulse variability is often noticeable in diffraction patterns. Random phase tilts in the pulses is one such detail, and they were observed to translate diffraction patterns of polystyrene nanospheres by as much as six pixels [14]. These translations, when uncorrected, could potentially decrease the overall resolution of the reconstructed contrasts, especially when the speckle features are roughly on the same scale as the translations. Thus, correctly centering diffraction patterns before reconstruction is attempted is crucial.

The diffraction pattern of a real (*i.e.*, not complex-valued) object can be approximated as centrosymmetric when the extent of the Ewald sphere's curvature is less than half a speckle diameter, $d/2$, at the edge of the detector,

$$k_0(1-\cos\theta_{max}) \ll d/2, \tag{1}$$

where $k_0 = 2\pi/\lambda$ is the magnitude of the wave vector and θ_{max} is the maximum scattering angle. Equation (1) can be simplified to

$$\theta_{max} \ll 1/N, \tag{2}$$

where $N = k_0\theta_{max}/d$ is approximately the number of speckles that can be counted on a ray from the origin to the edge of the detector.

When unshifted, the Fourier transform of the intensity $I(q)$ is the autocorrelation of the particle contrast Ψ. Mathematically, $\mathscr{F}[I(q)]$ should be real and centrosymmetric if absorption effects are negligible. When shifted by an unknown amount, $q_{unknown}$,

$$\mathscr{F}[I(q-q_{unknown})] = \mathscr{F}[I(q)]\exp(-iq_{unknown}\cdot x) \tag{3}$$

the Fourier transform of the shifted intensity is equal to the Fourier transfer of the unshifted intensity multiplied by a linear phase ramp. To identify the shift that best approximates $q_{unknown}$,

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

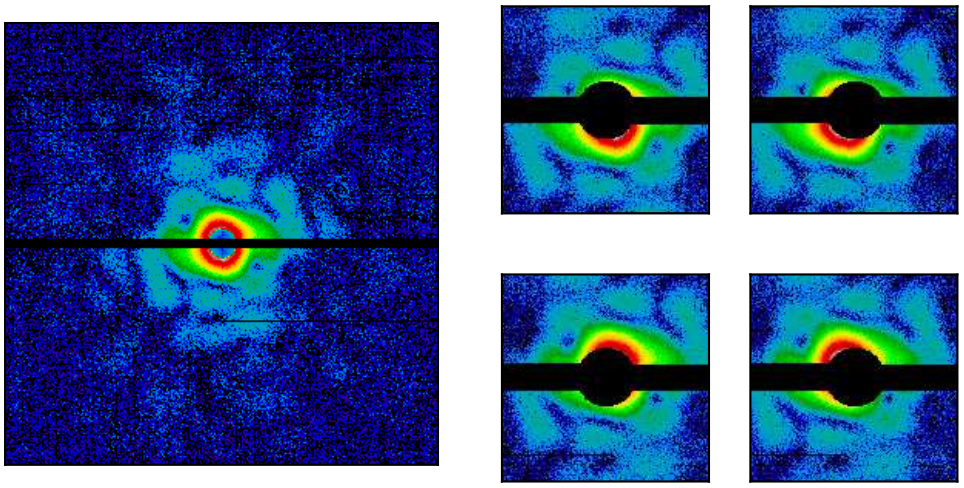

Fig. 1. To find the center of a diffraction pattern (left), square regions, translated by a set of candidate shifts (exaggerated on the right), are tested for centrosymmetry. An identical mask is applied to each of these square regions to mask out the missing central intensities. The shifted square region that is most centrosymmetric (see text for details) is presumed to be properly centered. This diffraction pattern was found to be shifted to left by two pixels.

the intensity $I(q-q_{unknown})$ is shifted by a known amount q_C. Once the Fourier transform is then computed, the sum of the absolute values of its imaginary components

$$\sum_x |\mathrm{Im}[\mathscr{F}[I(q)]\exp(-iq_{unknown}\cdot x)\exp(-iq_C\cdot x)]| \tag{4}$$

will equal zero if

$$q_C = -q_{unknown} \tag{5}$$

and the shifts are restricted to small values. Because actual recorded intensities are not perfectly centrosymmetric due to noise and missing data regions, Eq. (4) will most often equal a nontrivial value even when the proper shift is found. Still, the sum should be smaller for Eq. (5) than for other shifts.

To implement the shift locator, square regions of the pattern centered at different offsets q_C are cut out as shown in Fig. 1. A mask, fixed with respect to the square cutout, is applied to cover up the CCD gap and the central, circular region of unreliable photon count. The mask is made thicker so that it can be applied uniformly on any square cutout and still cover up the appropriate regions. The Fourier transform of these square cutouts is computed, and the offset with the lowest sum of the absolute value of the imaginary components is identified as the correct shift.

3.2. Noise robust difference map

In far-field diffraction theory, the contrast Ψ is the scattered wavefront immediately past the scattering particle. It can be characterized by I via the inverse Fourier transform relation,

$$\Psi = \mathscr{F}^{-1}[\sqrt{I}\exp(i\phi)], \tag{6}$$

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

where ϕ is the associated phase of the complex exit wave. To recover Ψ, both the magnitude and the phase of the observed wavefront must be known. However, since the phase is not recorded, it is recovered using other information, such as the size and shape of the scattering particle.

Phase retrieval methods in use today employ iterative schemes to find a ϕ that best reflects all available information regarding the scattering particle. This is done by finding a Ψ that satisfies the measured I as well as an additional constraint based on the shape S of the particle. Two projection operators, P_F and P_S, known as the Fourier and support projections, respectively, are defined as follows,

$$P_F[\Psi] = \mathscr{F}^{-1} \circ M_F \circ \mathscr{F}[\Psi] \tag{7}$$

where

$$M_F[\hat{\Psi}] = \begin{cases} \sqrt{I(q)}\dfrac{\hat{\Psi}(q)}{|\hat{\Psi}(q)|} & \text{if } I(q) \text{ is known and } |\hat{\Psi}(q)| \neq 0 \\ \hat{\Psi}(q) & \text{otherwise} \end{cases} \tag{8}$$

rescales the Fourier magnitude of the input to match that of the square root of the measured intensity and

$$P_S[\Psi] = \begin{cases} \Psi(r) & r \in S \text{ and } \Psi(r) \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

sets to zero any region that lies outside of S and imposes positivity. Beginning with a random initial contrast Ψ_0, phase retrieval methods search for solutions by projecting iterates onto the constraint sets through a combination of the projection operations and follow the form

$$\Psi_{n+1} = \Psi_n + \varepsilon[\Psi_n], \tag{10}$$

where $\varepsilon[\Psi_n]$ is an additive update to the iterate which depends on the choice of the phasing method. When the error metric $||\varepsilon[\Psi_n]||^2$ falls below some fixed tolerance, the iterations are stopped and the reconstruction is defined by the estimate $\Psi \approx P_F[\Psi_{n+1}]$.

We use the $\beta = 1$ form of the difference map [15],

$$\Psi_{n+1} = \Psi_n + P_S[2P_F[\Psi_n] - \Psi_n] - P_F[\Psi_n] = \Psi_n + \varepsilon_D[\Psi_n], \tag{11}$$

which is equivalent to the $\beta = 1$ form of Fienup's hybrid input-output rule [16]. The difference map is best suited for finding a true common element in the constraint sets, on the assumption one exists. When dealing with measured diffraction patterns, however, the presence of noise could shift the Fourier constraint set such that it does not exactly intersect the support constraint set. As a result, the difference map's propensity for guiding iterates away from near-intersections to avoid stagnation works to its disadvantage as iterates are sent elsewhere to look for solutions. Practically speaking, this results in a frustrated search where reconstructions associated with each iterate fluctuate significantly in shape and size.

In the face of variable signal-to-noise ratios, the phase retrieval process should be robust so that the search does not so easily stray from near-intersections [17, 18, 19]. Loh *et al*. [17] propose an intermediate step where Ψ_n is updated by the formula

$$\Psi'_n = \alpha\Psi_n + (1-\alpha)P_F[\Psi_n], \tag{12}$$

where $0 \leq \alpha \leq 1$ is a "leash" parameter that reins in the iterate such that it is brought closer to the Fourier constraint set before it is run through the difference map:

$$\Psi_{n+1} = \Psi'_n + \varepsilon_D[\Psi'_n]. \tag{13}$$

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

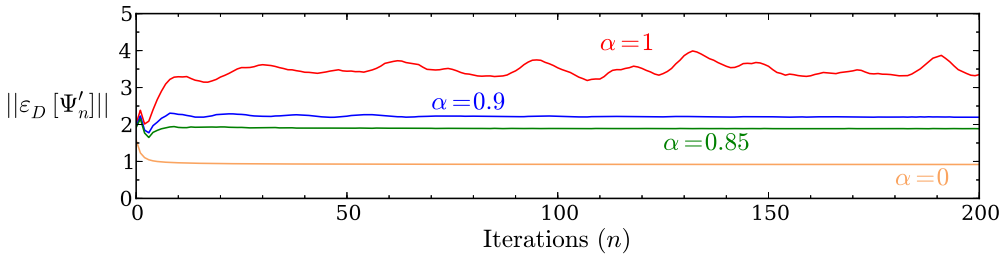

Fig. 2. The stability of the modified difference map for various α's around a solution can be measured by the square root of the error metric, $||\varepsilon_D[\Psi'_n]||$. Starting with a final reconstruction (*i.e.* solution) as the initial contrast and using a fixed support previously generated with Shrinkwrap [20], the modified difference map continues on its search in the neighborhood of the solution. An α slightly decreased from unity will significantly tighten the scope of the search and improve the stability of the difference map around a solution.

In our trials, $\alpha = 0.85$ was used. The choice for α reflects a desire to balance out the need for preventing the search from deviating from a near-intersection too much while also preventing it from settling too easily near a point which may not necessarily best reflect the near-intersection. There is some latitude in the choice of α as even a slight decrease from unity will significantly tighten up the search neighborhood around a near-intersection, as Fig. 2 suggests.

3.3. Unsupervised reconstruction assessment

In iterative phase retrieval methods, convergence to a unique solution within a set number of steps is not necessarily guaranteed, so the iterations are stopped when the difference between iterates is small enough or a large number of iterations, t_{max}, is reached. Consequently, an individual reconstruction will, at times, seem like it has not converged or perhaps even converged to a seemingly different point when compared to a different reconstruction. When presented with a collection of dissimilar reconstructions, visual inspection usually aids in assessing which reconstructions are successful, but this would be time consuming when processing thousands of diffraction patterns.

Since the particle's contrast is not known beforehand, assessing the success of reconstructions presents another challenge as there are no training examples to aid in assessing. In practice, given a set of m reconstructions, the largest subset that contain similar looking reconstructions is usually deemed to be a successful collection. This practice, however, relies on the assumption that the phasing algorithm can guide iterates to the correct near-intersection most of the time.

The steps taken in visually assessing and rejecting reconstructions are used to devise an algorithm that can perform the same task. Beginning with a set of m individual reconstructions, their corresponding supports, as found by Shrinkwrap [20], are compared. The supports are preferred over the reconstructions as that will emphasize during comparison the overall low-resolution shapes of the reconstructed contrasts as opposed to their subtle high-resolution features which could vary greatly. Let y_i be an $n \times n$ pixel array of binary values representing an individual support appropriately translated and inverted and $\bar{y} = \sum_i y_i/m$ be the mean of the supports. Translations and inversions can be identified relative to a reference support, which can be any of the m supports, by maximizing the cross-correlation of the individual and reference supports.

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

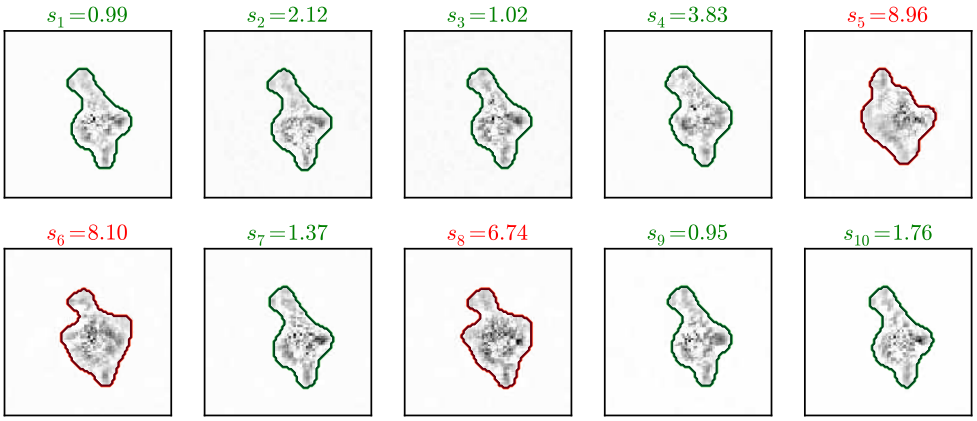

Fig. 3. Ten individual reconstructed contrasts with overlaid outlines of their supports, as found by Shrinkwrap, and their corresponding s_i values. The reconstructions whose s_i's exceed the threshold $s_{max} = 5\%$ are marked in red and were deemed failures.

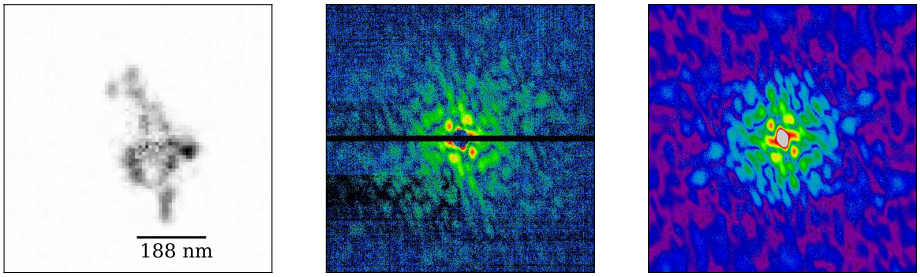

Fig. 4. A final reconstruction Ψ (on the left) obtained from averaging ten acceptable individual reconstructions. The measured diffraction pattern I (in the middle) and reconstructed intensity $|\hat{\Psi}|^2$ (on the right) demonstrate similar speckle structures in the low scattering angle regions, but differ considerably in the higher scattering angle regions.

The % deviation from the mean,

$$s_i = \frac{||y_i - \bar{y}||}{||\bar{y}||} \times 100, \tag{14}$$

where $||\cdot|| = \sum_{j,k} |(\cdot)_{jk}|$ is summed over pixels, also known as the L_1 norm, is obtained for each reconstruction and ordered from least to greatest. In the event all m supports are similar, all s_i's will generally be small. When some supports differ greatly from the majority, however, the % deviations will increase across all i's due to the inclusion of those dissimilar supports in computing the mean. To mitigate the effects of these inflated % deviations, a new mean $\bar{y}_{new}$ is computed based on the $m/2$ individual reconstructions with the lowest s_i's, which we consider to be the similar reconstructions. New % deviation values s_i's are then computed for each reconstruction. An absolute rejection criterion, $s_{max} = 5\%$, is set such that all reconstructions with $s_i > s_{max}$ are rejected, as shown in Fig. 3.

As subtle differences in shape and density can exist between similar individual reconstructions, these subtleties can be averaged away by adding the reconstructions. The approach taken

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

in this paper averages $m = 10$ reconstructions from different phasing runs with random initial contrasts and different initial circular supports of varying radii. For dissimilar reconstructions in the collection, new reconstructions are attempted using the same initial circular support with which the reconstruction began but with a different initial contrast. After the new reconstructions are obtained, the s_i's are computed again for the set of ten individual reconstructions consisting of the new as well as the previously unrejected reconstructions, and those failing to meet the rejection criterion are again discarded. This method is repeated until all the s_i's fall below s_{max}. Once ten acceptable reconstructions are obtained, they are then averaged to obtain a final reconstruction, as shown in Fig. 4.

3.4. Missing data

Diffraction patterns will often have significant regions of missing data mainly due to the gap between the CCDs and pixel saturation. When using information about the particle's support in phase retrieval, these missing data regions can be problematic as they could give rise to unconstrained modes, which are spurious features with enough power to exist in the support in real space and the missing data region in Fourier space [21]. Depending on the size of the missing data regions relative to the speckles, unconstrained modes could become problematic as they superimpose themselves over the true particle contrasts and result in inaccurate reconstructions.

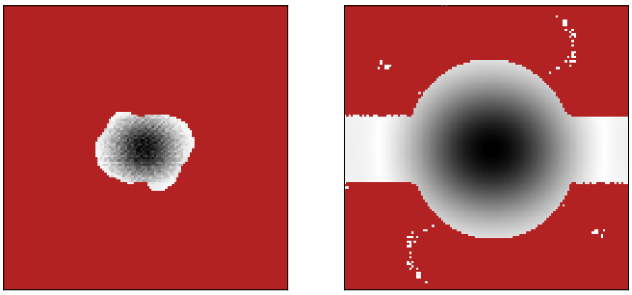

Fig. 5. A weakly constrained feature f in real space, shown in greyscale, with most of its power contained within the support, regions not colored in red (left). In Fourier space, the same feature, again shown in grayscale, has most of its power contained within the missing data region, again regions not colored in red (right).

The degree to which modes may be unconstrained can be measured by the rate at which they lose power during the phase retrieval process. Given some unconstrained feature in real space f we define its unconstrained power to be

$$W[f] = \frac{1}{2}\left(\int_S |f|^2\,dr + \int_M |\hat{f}|^2\,dq\right) \tag{15}$$

where S is the (previously defined) particle's support in real space and M is the missing data region in Fourier space.

Measuring the degree to which these features are constrained can be done by separately running a variation of the phase retrieval process. We define the missing data projection,

$$P_M[\Psi] = \mathscr{F}^{-1} \circ S_M \circ \mathscr{F}[\Psi] \tag{16}$$

where

$$S_M[\hat{\Psi}] = \begin{cases} \mathrm{Re}[\hat{\Psi}(q)] & q \in M \\ 0 & q \notin M \end{cases} \tag{17}$$

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

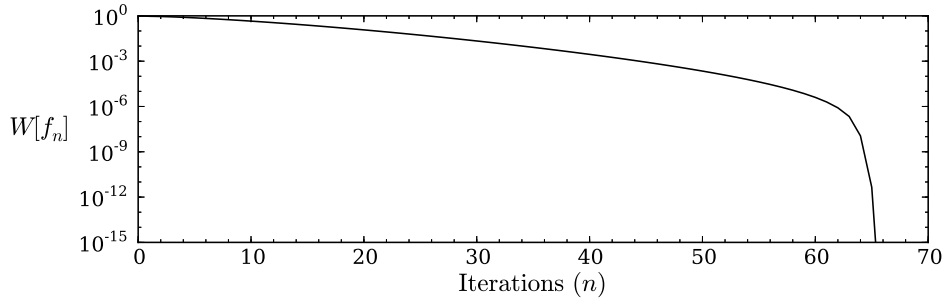

Fig. 6. The power of an unconstrained feature as it is iteratively updated via the variation of the modified difference map with *S* and *M* from Fig. 5. The feature's power decreases by six decades in ~ 60 iterations before it abruptly falls effectively to zero. This suggests any unconstrained features that arise during the reconstruction process will effectively be suppressed if the time scales of their decay are much less than the time scales of the overall reconstruction process.

and substitute the Fourier projection P_F with P_M in the modified difference map, while keeping everything else, such as the α parameter, the same. Beginning with a random initial contrast f_0, the modified difference map with the missing data projection will search for features whose power are not constrained within the *S* and *M* in real and Fourier space, respectively. When there are no modes with significant unconstrained power, we expect any initial contrast to decay quickly when the above scheme is iterated.

The rate of power loss gives a sense of how constrained the features are during the phase retrieval process and of how severely they could distort the final reconstruction. For power loss as shown in Fig. 6, the decrease of six decades in about sixty iterations followed by the abrupt drop to zero in total power suggests that the support and missing data regions are too restrictive in allowing any significant unconstrained features to persist. For phase retrieval runs consisting of thousands of iterations, it can be expected that features as those described in Fig. 5 will not contribute significantly to the final reconstruction's total power.

4. Results

The reconstruction process consists of the following steps: centering the diffraction patterns, generating ten acceptable individual reconstructions, and checking whether unconstrained modes and features could exist in the reconstructions. All computations were performed on a standard desktop computer equipped with a quad-core Intel i7-2600 with a clock cycle of 3.4 GHz and 8 GB RAM. Each individual reconstruction run with $t_{max} = 2000$ iterations took approximately 15 minutes on a single thread. By taking advantage of multithreading, up to five threads ran simultaneous individual reconstructions, shortening the computation of ten individual reconstructions to a minimum of 30 minutes. A maximum of four attempts were made with each initial support. In the event an individual reconstruction attempt was rejected a fourth time, the whole reconstruction process was deemed a failure.

The diffraction images recorded at the LCLS underwent preprocessing where the running background was subtracted. They were then subjected to an intensity-based thresholding routine to identify those that contained sufficient photon signal likely to result from particle-pulse diffraction events and did not exhibit pixel saturation effects. A collection of 953 patterns was generated, and 309 of those patterns were chosen for phasing through visual inspection based on the size of the speckles and good signal-to-noise ratio. An investigation on unbiased pattern

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

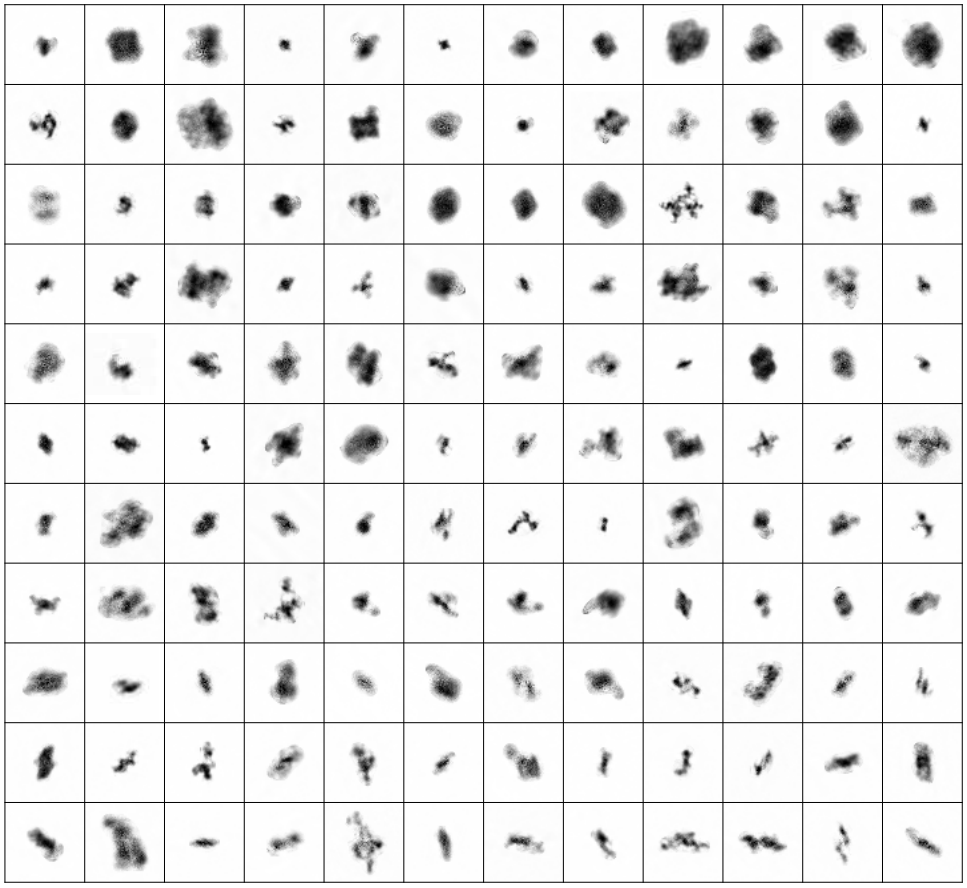

Fig. 7. A selection of reconstructed soot contrasts, arranged by increasing shape eccentricity. The length of each square box is 573 nm.

selection is underway.

The top and bottom halves of the patterns were added after centering to increase the signal-to-noise ratio and to constrain the contrasts to be real, assuming the patterns largely obeyed centrosymmetry. Given the maximum scattering angle, $\theta_{max} = 0.075$ rads, the condition for centrosymmetry as described in Eq. (2) to hold requires that the distance from the center of the detector to the edge not exceed 13 speckles. A number of patterns did exceed that count by a couple of speckles, but in most of those cases, noise made it difficult to clearly discern any speckles close to the edge, making the effective maximal scattering angle less than what the detector allows for.

Of those chosen, 36 patterns failed to produce 10 similar individual reconstructions. In 30 out of those 36 instances, at least 8 individual reconstructions were deemed similar. A total of 273 patterns yielded averaged reconstructions, and some these reconstructions are shown in Fig. 7. The quality of the reconstructions was assessed by computing the phase retrieval transfer function (PRTF) and an effective resolution was characterized by where the PRTF drops to $1/e$ [22]. There was great variability in the quality of the reconstruction, as shown in Fig. 8, with a resolution range of 18 nm to 89 nm.

Many of the diffraction patterns were shifted by various amounts as shown in Fig. 9. A considerable number of them demonstrated shifts as much as 4 pixels and only 19 patterns

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

were unshifted by the pulse. None of the averaged reconstructions had significant missing data problems as unconstrained features all experienced power decay to zero when run through the procedure outlined in Sec. 3.4 using averaged supports. Only in 3 cases did the decay take over 100 iterations. Even then, the longest time it took for complete power loss was 258 iterations. Since the overwhelming majority of patterns chosen had speckles larger than the missing data region, which was indirectly a consequence of Eq. (1), this was to be expected.

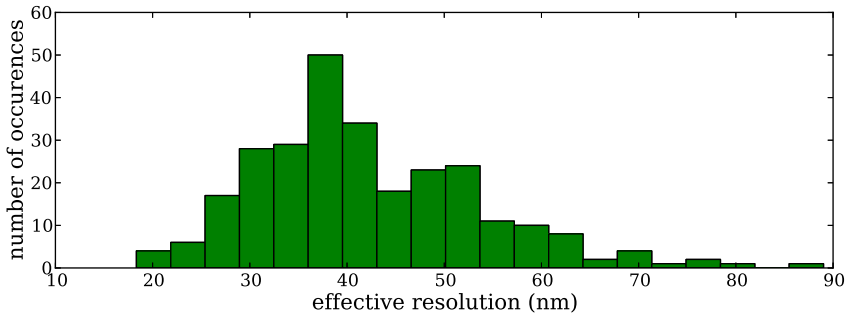

Fig. 8. Histogram of the effective resolution of the 273 reconstructions, quantified by where the phase retrieval transfer function dips below $1/e$. The smallest effective resolution was determined to be 18 nm, and the largest was 89 nm.

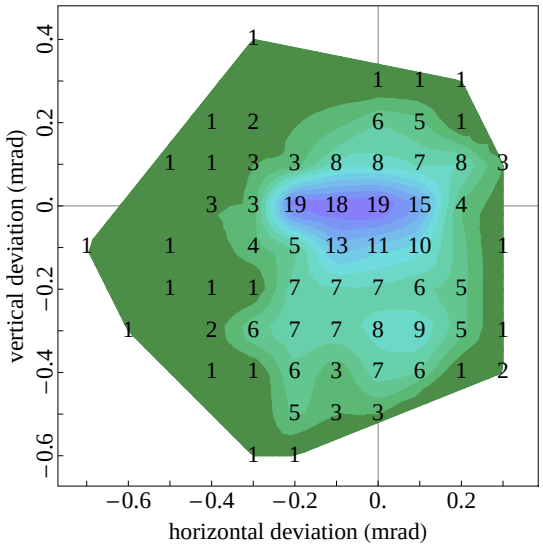

Fig. 9. 2D histogram of the offsets, measured as outlined in Sec. 3.1, in the 309 patterns due to random phase tilts in the X-ray wavefront. The distribution of offsets displays a strong spread in horizontal deviations, particularly those with no vertical deviations.

5. Conclusion

Data collection rate at facilities such as the LCLS makes it infeasible to carry out a user guided reconstruction for each diffraction pattern. The ability to analyze and extract meaningful results from single shot diffraction imaging experiments will invariably require a speedy and reliable

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013

contrast reconstruction process, ideally with no supervision. We presented measures aimed at facilitating data processing and the contrast reconstruction steps, and they have shown that high throughput, unsupervised reconstructions are possible. A desktop implementation of our methods quickly reaches its computational limits, however, and orders of magnitude speedup is necessary for the data collection and processing rates to reach parity. Graphical processing units (GPUs) [23] and other "multicore" computing solutions show promise in providing the necessary speedup.

The ability to generate a large collection of images via single shot diffraction imaging enables the possibility for morphological studies analogous to those performed on collections of images obtained through other imaging techniques such as transmission electron microscopy. Diffraction imaging has an advantage over those imaging techniques as it allows for observation *in situ* of airborne particles such as soot. A whole host of other aerosols, such as medicinal nanoparticles to cloud seeds, could benefit from study via single shot diffraction imaging as their airborne structures could yield new insight into their function.

Acknowledgments

Experiments were carried out at the LCLS, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy (DOE), Office of Basic Energy Sciences. We acknowledge support by the following: DOE grant DE-FG02-11ER16210 (H. J. Park, V. Elser); Human Frontier Science Program (N. D. Loh, M. J. Bogan); AMOS program within the Chemical Sciences, Geosciences, and Biosciences Division of the Office of Basic Energy Sciences, Office of Science, U.S. DOE (N. D. Loh, R. G. Sierra, C. Y. Hampton, D. Starodub, and M. J. Bogan); the Max Planck Society for funding the development and operation of the CAMP instrument within the ASG at CFEL; the Hamburg Ministry of Science and Research and Joachim Herz Stiftung as part of the Hamburg Initiative for Excellence in Research (LEXI); the Hamburg School for Structure and Dynamics in Infection; the Swedish Research Council, the European Research Council, and Knut och Alice Wallenbergs Stiftelse. Part of this work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Support for M. Frank, G. Farquar, W.H. Benner, S. Hau-Riege was provided by the UCOP Lab Fee Program (award no. 118036). The Max Planck Advanced Study Group at CFEL acknowledges technical support by R. Andritschke, K. Gärtner, O. Hälker, S. Herrmann, A. Hömke, Ch. Kaiser, K.-U. Kühnel, W. Leitenberger, D. Miessner, D. Pietschner, M. Porro, R. Richter, G. Schaller, C. Schmidt, F. Schopper, C.-D. Schröter, Ch. Thamm, A. Walenta, A. Ziegler, and H. Gorke. We thank the staff of the LCLS for their support in the experiments that provided the data for this study.

#195785 - $15.00 USD Received 19 Aug 2013; revised 26 Oct 2013; accepted 27 Oct 2013; published 14 Nov 2013