

CCP-FEL: a collection of computer programs for free-electron laser research

Filipe R. N. C. Maia,^{a,*} Thomas A. White,^b N. Duane Loh^{c,d,e} and Janos Hajdu^a

^aLaboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 (Box 596), SE-751 24 Uppsala, Sweden, ^bCenter for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron DESY, Notkestrasse 85, 22607 Hamburg, Germany, ^cCentre for Bio-imaging Sciences, National University of Singapore, 14 Science Drive 4, 117557, Singapore, ^dDepartment of Physics, National University of Singapore, 2 Science Drive 3, 117551, Singapore, and ^eDepartment of Biological Sciences, National University of Singapore, 14 Science Drive 4, 117557, Singapore. *Correspondence e-mail: filipe@xray.bmc.uu.se

Keywords: free-electron lasers; computer programs.

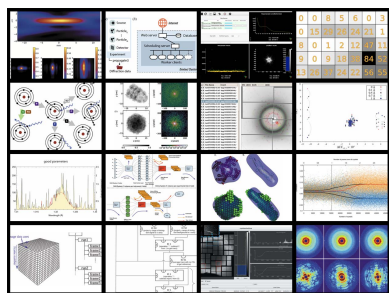
The latest virtual special issue of *Journal of Applied Crystallography* (http://journals.iucr.org/special_issues/2016/ccpfel) collects software for free-electron laser research and presents tools for a range of topics such as simulation of experiments, online monitoring of data collection, selection of hits, diagnostics of data quality, data management, data analysis and structure determination for both nanocrystallography and single-particle diffractive imaging. This article provides an introduction to the special issue.

1. Introduction

X-ray free-electron lasers (FELs) are transforming photon science and its various applications (Young *et al.*, 2010; Chapman *et al.*, 2011; Seibert *et al.*, 2011). Operational X-ray lasers include the Linac Coherent Light Source at Stanford, CA, USA (Emma *et al.*, 2010; <https://www6.slac.stanford.edu/facilities/lcls>), the SPring-8 Angstrom Compact Free Electron Laser in Harima, Japan (<http://xfel.riken.jp/eng/index.html>), the FERMI facility in Trieste, Italy (<http://www.elettra.eu/lightsources/fermi.html>), and the FLASH facility in Hamburg (<https://flash.desy.de/>). Further X-ray lasers are under construction, and these include the European XFEL in Hamburg (Altarelli, 2006; <http://www.xfel.eu/>), the Swiss Free-Electron Laser in Villigen, Switzerland (<http://www.psi.ch/swissfel/swissfel>), and the PAL X-ray Free-Electron Laser (PAL-XFEL) in Pohang, South Korea (<http://accelconf.web.cern.ch/accelconf/ipac2013/papers/weodb103.pdf>).

Compared to X-ray pulses from third-generation synchrotron light sources, FEL pulses are 10^{10} times higher in peak brightness, are 10^9 times higher in their coherence degeneracy parameters and provide femtosecond–attosecond time resolutions. These extraordinary parameters of FELs have stirred intense and broad development in ultra-fast imaging science.

Over the past decade, a diverse range of knowledge has been incorporated into computational tools that help us manage and interpret the statistically complex data from FELs. The latest virtual special issue of *Journal of Applied Crystallography* (http://journals.iucr.org/special_issues/2016/ccpfel) collects software for FEL research and presents tools for a range of topics such as simulation of experiments, online monitoring of data collection, selection of hits, diagnostics of data quality, data management, data analysis and structure determination for both nanocrystallography and single-particle diffractive imaging. We hope that this collection of software opens the doors of FEL research to a wider



community, not unlike what the early collection of software in the CCP4 suite of programs did for synchrotron-based crystallography (CCP4: Collaborative Computational Project, No. 4; Winn *et al.*, 2011).

Ultra-short and extremely intense FEL pulses offer extraordinary research opportunities, and the present *Collection of Computer Programs for FEL Research* in *Journal of Applied Crystallography* will aid users in performing diffraction experiments using ultra-short coherent FEL pulses. The two main areas covered in this special issue are single-particle imaging (SPI) (Seibert *et al.*, 2011; Ekeberg *et al.*, 2015) and serial femtosecond nanocrystallography (SFX) ([http://www-](http://www-ssl.slac.stanford.edu/lcls/papers/lcls_experiments_2.pdf)

[ssl.slac.stanford.edu/lcls/papers/lcls_experiments_2.pdf](http://www-ssl.slac.stanford.edu/lcls/papers/lcls_experiments_2.pdf); Chapman *et al.*, 2011). Both of these areas exploit the phenomenon of ‘diffraction before destruction’ (Neutze *et al.*, 2000) to outrun key damage processes and obtain diffraction patterns from virtually unperturbed samples. This principle was first demonstrated on 1 February 2006 in a flash-diffraction experiment (Chapman *et al.*, 2006) at the VUV-FEL facility in Hamburg, subsequently renamed as FLASH (http://www.desy.de/news/news/archive_before_2010/2006/photon_1004/index_eng.html).

The high repetition rate of X-ray pulses from FEL facilities leads to large volumes of experimental data. The noisy and incomplete experimental data collected at FELs necessitated the development of statistically robust approaches to both data classification and analysis (Maia *et al.*, 2010; Loh, 2014). While these challenges were progressively addressed in early proof-of-principle diffraction imaging experiments performed by large specialist groups, they remain problematic for a broader community of users wishing to take advantage of the revolutionary new capabilities of FELs. This collection of software aims to directly address this issue.

1.1. Imaging of non-periodic single objects with FELs (Fig. 1a)

The diffraction pattern of a non-periodic object is continuous, and in principle, there is a direct way to determine the phases from the pattern for image reconstruction by ‘over-sampling’ (Bernal *et al.*, 1938; Shannon, 1949; Sayre, 1952; Fienup, 1982). The resolution in single-particle experiments does not depend on sample quality in the same way as in conventional crystallography but is a function of radiation intensity, pulse duration and wavelength, which are factors controlling ionization and sample movement during exposure. A three-dimensional data set can be assembled from the diffraction patterns when copies of a reproducible sample are exposed to the beam one by one in random orientations (Huldt *et al.*, 2003; Ekeberg *et al.*, 2015). Damage can be distributed over many copies of the sample in a similar manner as in ‘single-molecule’ electron microscopy or in crystallography. The diffraction signal from reproducible objects can be enhanced by merging multiple patterns, and this may be possible even from very weak individual shots (Loh & Elser, 2009). Averaging in a redundant data set reduces the error in the signal by the square root of the number of patterns that sample the same pixel. The limiting factors here are the accuracy of the orientation information and

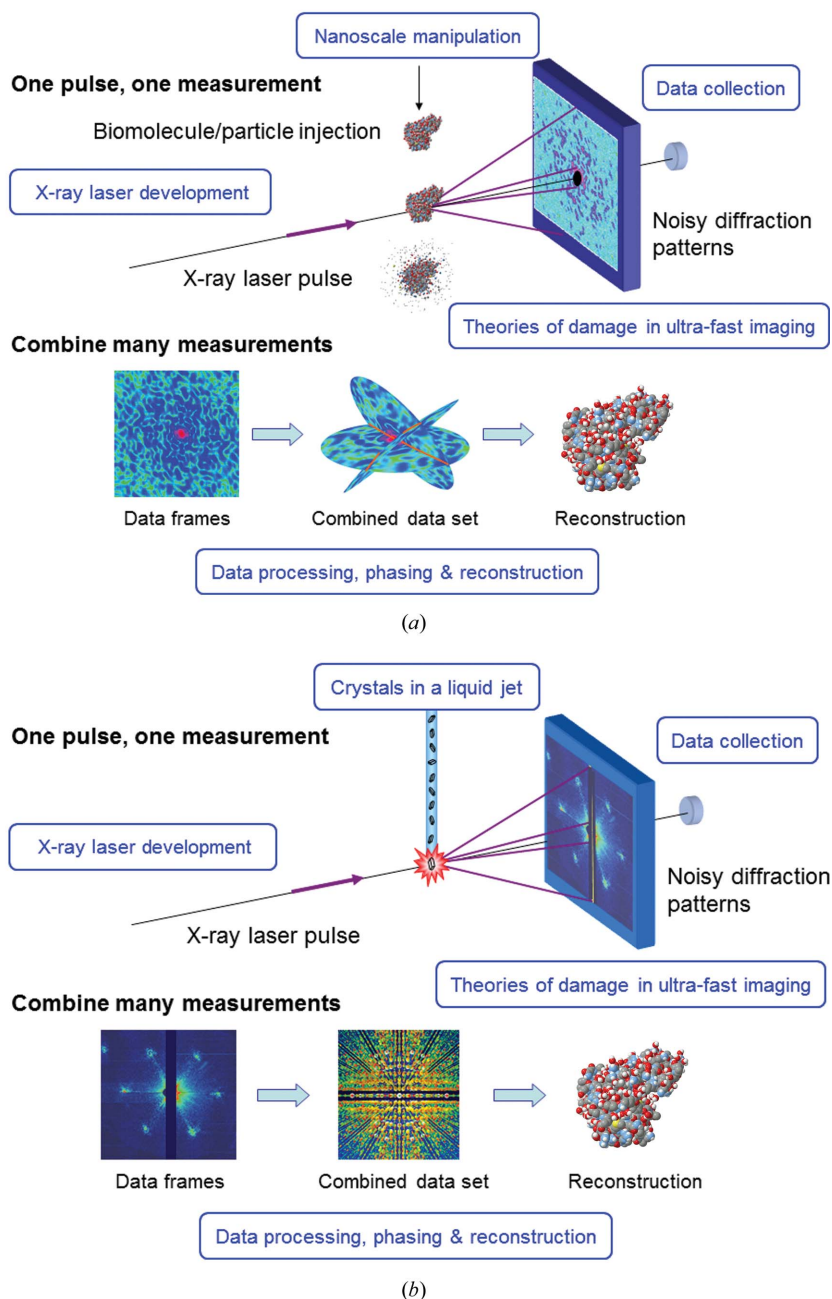


Figure 1 Experimental schemes for (a) single-particle imaging and (b) protein nanocrystallography with FELs.

heterogeneity in the sample population, both of which will blur the reconstructed image (Maia *et al.*, 2009; Quiney & Nugent, 2010). The highest possible resolution in a single shot is not necessarily at the shortest possible wavelength. For a detailed analysis, see Bergh *et al.* (2008).

With small non-reproducible objects, however, there is no chance of repeating the FEL experiment to improve the signal-to-noise ratio or of assembling a three-dimensional data set from multiple two-dimensional views. Nevertheless, serially imaging individual non-reproducible objects in a heterogeneous ensemble can lead to insights unavailable in ensemble-averaged measurements (Loh *et al.*, 2012). Experimental strategies are, therefore, different for high-resolution studies on reproducible and non-reproducible objects (*e.g.* random aggregates, cells).

1.2. Femtosecond protein nanocrystallography (Fig. 1b)

X-ray crystallography provides the vast majority of macromolecular structures, but the success of the method relies on growing crystals of sufficient size. It is particularly challenging to obtain large, well diffracting crystals of certain classes of proteins (as well as many other types of materials), including membrane proteins and glycoproteins, for which a high-resolution means of structure determination would be invaluable. Smaller crystals can be studied by focusing the X-ray beam more tightly, but pushing this too far leads to the crystal suffering extensive radiation damage before a diffraction signal can be recorded. Single-crystal X-ray diffraction ‘snapshots’ can be collected from a stream of nanocrystals using femtosecond hard-X-ray pulses from a FEL. This concept was proven with nanocrystals of photosystem I (Chapman *et al.*, 2011) at the LCLS, with nearly 200 000 diffraction patterns being found from over 3 000 000 detector frames, one frame per crystal and one frame per X-ray pulse. SFX offers a new approach to structure determination of macromolecules that do not yield crystals of sufficient size for studies using conventional radiation sources, and of those which are particularly sensitive to radiation damage, as well as for obtaining high time resolution in pump–probe experiments (Pande *et al.*, 2016). In a ground-breaking experiment in which nanocrystallography met single-particle imaging, Henry Chapman and colleagues demonstrated structure determination from the continuous diffraction pattern of disordered crystals to significantly higher resolutions than the resolution permitted by the Bragg peaks alone (Ayyer, Yefanov *et al.*, 2016). The continuous part of the diffraction pattern was used to phase the pattern directly.

2. Overview of the software collection

This collection of computer programs includes three software packages to aid experiment design. The *WavePropaGator* package (Samoylova *et al.*, 2016) simulates the efficiencies and aberrations that may arise during photon transport in a FEL. Downstream of this, the twin packages *XMDYN* and *XATOM* (Jurek *et al.*, 2016) cover key photon–matter interactions that

occur when biological samples are illuminated with intense X-ray pulses. Finally, the *Condor* simulation package (Hantke *et al.*, 2016) estimates the signal level in single-particle elastic scattering, which is useful in deciding the feasibility of structure determination.

This suite includes a number of online and offline diagnostic tools to give prompt feedback when faced with the FEL data ‘firehose’. The *Hummingbird* (Daurer *et al.*, 2016), *OnDA* (Mariani *et al.*, 2016) and *CASS* (Foucar, 2016) packages implement parallel socket-based frameworks to monitor key metrics in real time during a FEL experiment, with a different emphasis placed on SPI and crystallography between them. A combined online and ‘fast feedback’ pipeline for SFX experiments at SACLA has been created (Nakane *et al.*, 2016). Application-specific programs such as these are built upon facility-specific data analysis frameworks such as the *psana* framework at LCLS (Damiani *et al.*, 2016), which in turn are built upon large-scale data handling infrastructures (Thayer *et al.*, 2016). In both the SPI and SFX cases, this brings us to a reduced set of data containing only the detector frames that contain useful diffraction patterns.

In the case of SPI, two packages are presented for determining structures from diffraction patterns. For imaging faint reproducible samples, the *Dragonfly* package (Ayyer, Lan *et al.*, 2016) implements an expectation maximization recipe to recover the unmeasured orientations of very many single-particle diffraction patterns. For non-reproducible samples, Ma & Liu (2016) studied the structure determination by harvesting information from angular correlations similar to that in small-angle X-ray scattering experiments performed at synchrotrons.

For the case of SFX, the package *cpxfel* is introduced in this special issue (Ginn *et al.*, 2016). This package aims to provide a ‘toolkit’ and showcase new techniques for advanced data processing. Significant advances in the *CrystFEL* suite are described (White *et al.*, 2016), both of a technical nature and regarding the underlying methods of data processing. Two papers describe new analysis methods: a method for excluding diffraction snapshots which reduce the final data quality instead of increasing it (Assmann *et al.*, 2016), and a method for reconstructing the true values of pixels with negative intensities from certain detectors in the field of electron diffraction (Hattne *et al.*, 2016). Electron crystallography overlaps with free-electron laser crystallography in its recent applications to very small macromolecular crystals, and is facing many similar analysis problems. The type of problem solved in this paper will doubtless seem familiar to many scientists working with FEL data. Finally, *IOTA* (Lyubimov *et al.*, 2016) is introduced, which allows the optimum values of the many SFX data processing parameters to be found by an automated grid search technique.

This virtual special issue of *Journal of Applied Crystallography* on FEL software is published simultaneously with a collection of Data Descriptors in Nature’s *Scientific Data*, with articles describing diffraction data obtained with the LCLS. The theme of the collection is *Structural Biology Applications of X-ray Lasers* (Maia & Hajdu, 2016; *Scientific Data*, 2016).

The papers cover data on nanocrystals (White, Barty *et al.*, 2016; Zhou *et al.*, 2016), single virus particles (Ekeberg *et al.*, 2016; Munke *et al.*, 2016), isolated cell organelles (Hantke, Hasse *et al.*, 2016) and living cells (van der Schot *et al.*, 2016). All these data sets are available from the Coherent X-ray Imaging Data Bank (CXIDB; Maia, 2012) to test software and to develop ideas, tools and procedures to meet challenges with the expected torrents of data from X-ray lasers.

References

- Altarelli, M. *et al.* (2006). *The European X-ray Free-Electron Laser*, DESY Technical Design Report 2006-097. Hamburg: DESY XFEL Project Group.
- Assmann, G., Brehm, W. & Diederichs, K. (2016). *J. Appl. Cryst.* **49**, 1021–1028.
- Ayyer, K., Yefanov, O. M. *et al.* (2016). *Nature*, **530**, 202–206.
- Ayyer, K., Lan, T.-Y., Elser, V. & Loh, N. D. (2016). *J. Appl. Cryst.* **49**, 1320–1335.
- Bergh, M., Huldt, G., Timneanu, N., Maia, F. R. N. C. & Hajdu, J. (2008). *Q. Rev. Biophys.* **41**, 181–204.
- Bernal, J. D., Fankuchen, I. & Perutz, M. F. (1938). *Nature*, **141**, 523–524.
- Chapman, H. N., Barty, A. *et al.* (2006). *Nat. Phys.* **2**, 839–843.
- Chapman, H. N., Fromme, P. *et al.* (2011). *Nature*, **470**, 73–77.
- Damiani, D., Dubrovin, M., Gaponenko, I., Kroeger, W., Lane, T. J., Mitra, A., O'Grady, C. P., Salnikov, A., Sanchez-Gonzalez, A., Schneider, D. & Yoon, C. H. (2016). *J. Appl. Cryst.* **49**, 672–679.
- Daurer, B. J., Hantke, M. F., Nettelblad, C. & Maia, F. R. N. C. (2016). *J. Appl. Cryst.* **49**, 1042–1047.
- Ekeberg, T., Svenda, M., Abergel, C. *et al.* (2015). *Phys. Rev. Lett.* **114**, 098102.
- Ekeberg, T., Svenda, M., Seibert, M. M. *et al.* (2016). *Sci. Data* **3**, 160060.
- Emma, P. *et al.* (2010). *Nat. Photon.* **4**, 641–647.
- Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758–2769.
- Foucar, L. (2016). *J. Appl. Cryst.* **49**, 1336–1346.
- Ginn, H. M., Evans, G., Sauter, N. K. & Stuart, D. I. (2016). *J. Appl. Cryst.* **49**, 1065–1072.
- Hantke, M. F., Ekeberg, T. & Maia, F. R. N. C. (2016). *J. Appl. Cryst.* **49**, 1356–1362.
- Hantke, M. F., Hasse, D. *et al.* (2016). *Sci. Data* **3**, 160061.
- Hattne, J., Shi, D., de la Cruz, M. J., Reyes, F. E. & Gonen, T. (2016). *J. Appl. Cryst.* **49**, 1029–1034.
- Huldt, G., Szoke, A. & Hajdu, J. (2003). *J. Struct. Biol.* **144**, 219–227.
- Jurek, Z., Son, S.-K., Ziája, B. & Santra, R. (2016). *J. Appl. Cryst.* **49**, 1048–1056.
- Loh, N. D. (2014). *Philos. Trans. R. Soc. London Ser. B*, **369**, 20130328.
- Loh, N. T. D. & Elser, V. (2009). *Phys. Rev. E*, **80**, 026705.
- Loh, N. D., Hampton, C. Y. *et al.* (2012). *Nature* **486**, 513–517.
- Lyubimov, A. Y., Uervirojnangkoorn, M., Zeldin, O. B., Brewster, A. S., Murray, T. D., Sauter, N. K., Berger, J. M., Weis, W. I. & Brunger, A. T. (2016). *J. Appl. Cryst.* **49**, 1057–1064.
- Ma, S. & Liu, H. (2016). *J. Appl. Cryst.* **49**, 665–671.
- Maia, F. R. N. C. (2012). *Nat. Methods*, **9**, 854–855.
- Maia, F. R. N. C., Ekeberg, T., van der Spoel, D. & Hajdu, J. (2010). *J. Appl. Cryst.* **43**, 1535–1539.
- Maia, F. R. N. C., Ekeberg, T., Timneanu, N., van der Spoel, D. & Hajdu, J. (2009). *Phys. Rev. E*, **80**, 031905.
- Maia, F. R. N. C. & Hajdu, J. (2016). *Sci. Data* **3**, 160059.
- Mariani, V., Morgan, A., Yoon, C. H., Lane, T. J., White, T. A., O'Grady, C., Kuhn, M., Aplin, S., Koglin, J., Barty, A. & Chapman, H. N. (2016). *J. Appl. Cryst.* **49**, 1073–1080.
- Munke, A. *et al.* (2016). *Sci. Data* **3**, 160064.
- Nakane, T., Joti, Y., Tono, K., Yabashi, M., Nango, E., Iwata, S., Ishitani, R. & Nureki, O. (2016). *J. Appl. Cryst.* **49**, 1035–1041.
- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature*, **406**, 752–757.
- Pande, K. *et al.* (2016). *Science*, **352**, 725–729.
- Quiney, H. M. & Nugent, K. A. (2010). *Nat. Phys.* **7**, 142–146.
- Samoylova, L., Buzmakov, A., Chubar, O. & Sinn, H. (2016). *J. Appl. Cryst.* **49**, 1347–1355.
- Sayre, D. (1952). *Acta Cryst.* **5**, 843.
- Schot, G. van der *et al.* (2016). *Sci. Data* **3**, 160058.
- Scientific Data (2016). *Structural Biology Applications of X-ray Lasers*, <http://www.nature.com/sdata/collections/xfel-biodata>.
- Seibert, M. M., Ekeberg, T., Maia, F. R. N. C. *et al.* (2011). *Nature*, **470**, 78–81.
- Shannon, C. E. (1949). *Proc. Inst. Radio Eng.* **37**, 10–21.
- Thayer, J., Damiani, D., Ford, C., Gaponenko, I., Kroeger, W., O'Grady, C., Pines, J., Tookey, T., Weaver, M. & Perazzo, A. (2016). *J. Appl. Cryst.* **49**, 1363–1369.
- White, T. A., Barty, A. *et al.* (2016). *Sci. Data* **3**, 160057.
- White, T. A., Mariani, V., Brehm, W., Yefanov, O., Barty, A., Beyerlein, K. R., Chervinskii, F., Galli, L., Gati, C., Nakane, T., Tolstikova, A., Yamashita, K., Yoon, C. H., Diederichs, K. & Chapman, H. N. (2016). *J. Appl. Cryst.* **49**, 680–689.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Young, L. *et al.* (2010). *Nature*, **466**, 56–66.
- Zhou, X. E. *et al.* (2016). *Sci. Data* **3**, 160021.