Software tools for the analysis of serial femtosecond crystallography (SFX)
Nadia Zatsepin, ASU

Thank you

- **Tom White, CFEL/DESY**
- **Anton Barty, CFEL/DESY**
- **Rick Kirian, CFEL/DESY, ASU**
- **John Spence, ASU, LBNL**
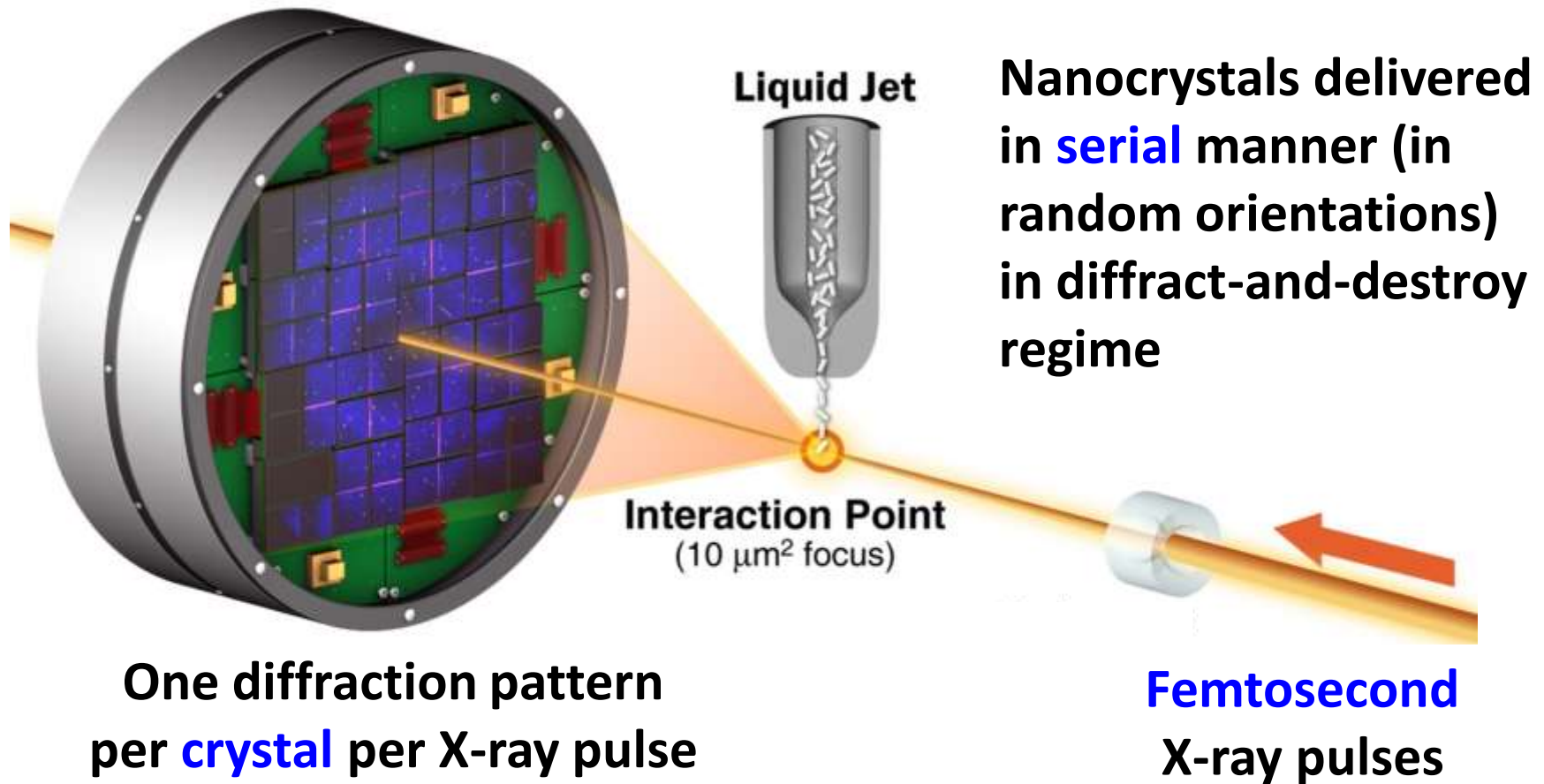- **Henry Chapman, CFEL/DESY, U.Hamburg**
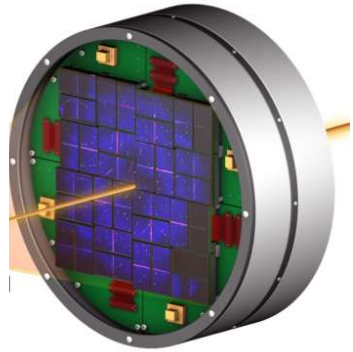
Crystallography at XFEL Sources
Workshop, August 5th, 2014

A National Science Foundation Science and Technology Center

bioxfel.org

Serial femtosecond crystallography (SFX)

Liquid Jet

Interaction Point
(10 µm² focus)

Nanocrystals delivered in serial manner (in random orientations) in diffract-and-destroy regime

One diffraction pattern per crystal per X-ray pulse

Femtosecond X-ray pulses

SFX data analysis pipeline

CSPAD

DAQ: XTC files, containing X-ray pulse parameters, diagnostics, motor positions, ..

Cheetah

1. Hit finding (data reduction)
2. Background estimation, removal
3. Clean diff. pattern & meta data → HDF5
4. Statistics & prelim. analysis

CrystFEL

1. Indexing
2. Integration
3. Merging
4. Post refinement
5. Indexing ambiguity removal

CCP4, Phenix et al.

Phasing, building, refinement, validation

Why new software?

- New type of data
- Large amount of data
- New, complicated detectors

LCLS fires at **120 Hz**

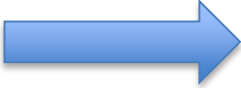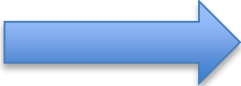CSPAD detector at CXI, LCLS: 2.3×10^6 pixels, 16 bit

4.6 MB / frame = 2 TB / hour

→ Up to ~ **120 TB from a single experiment** (5 shifts) from one detector

European XFEL will run at 27,000 Hz…

Data retention at LCLS

Storage Classes

Space	Size	Backup	Lifetime	Storage class	Comment
xtc	Unlimited	Tape archive	6 months	Short-term	Raw data
usr	Unlimited	Tape archive	6 months	Short-term	Raw data from users' DAQ systems
hdf5	Unlimited	Tape archive	6 months	Short-term	Data translated to HDF5
scratch	Unlimited	None	6 months	Short-term	Temporary data
xtc/hdf5	10TB	n/a	2 years	Medium-term	Selected XTC and HDF5 runs
ftc	10TB	None	2 years	Medium-term	Filtered, translated, compressed
res	1TB	Tape	2 years	Medium-term	Analysis results
User home	20GB	Disk + tape	Indefinite		User code
Tape archive	Unlimited	Two copies	10 years	Long-term	Raw data

Papers take ~2 years to publish:
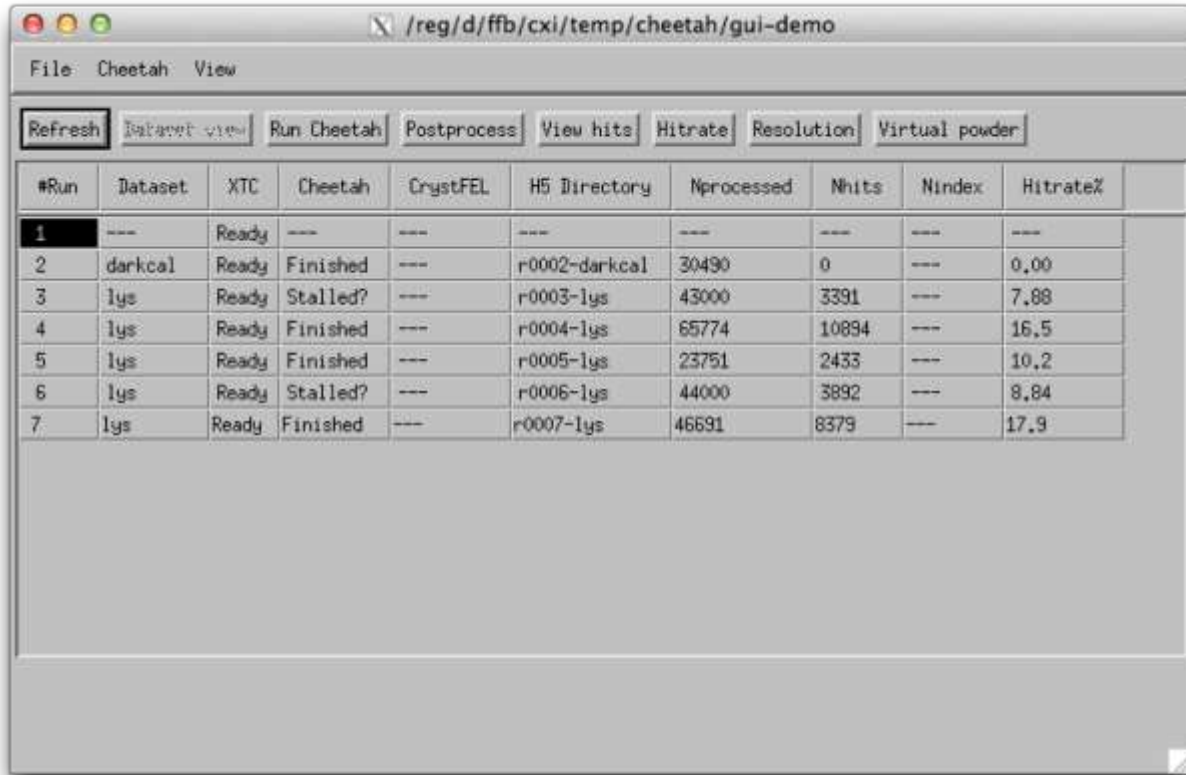
6 months is very little time for deletion of raw data

Need reliable, high throughput data reduction software

Cheetah

- Raw data →
  diffraction frames

- Parallelized
- High throughput
- Live feedback during
  experiment

- Preinstalled at SLAC

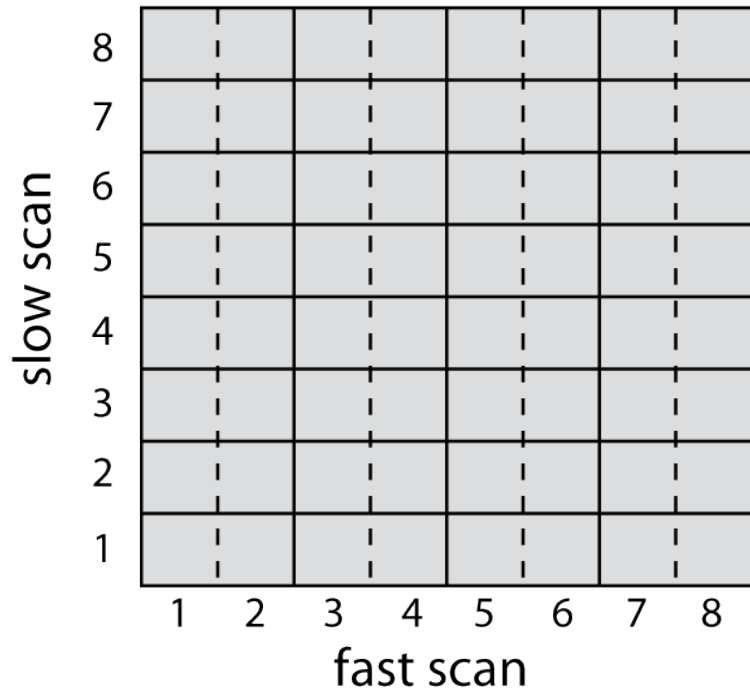A. Barty, et al. (2014) Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. J Appl Cryst. 47 1118.
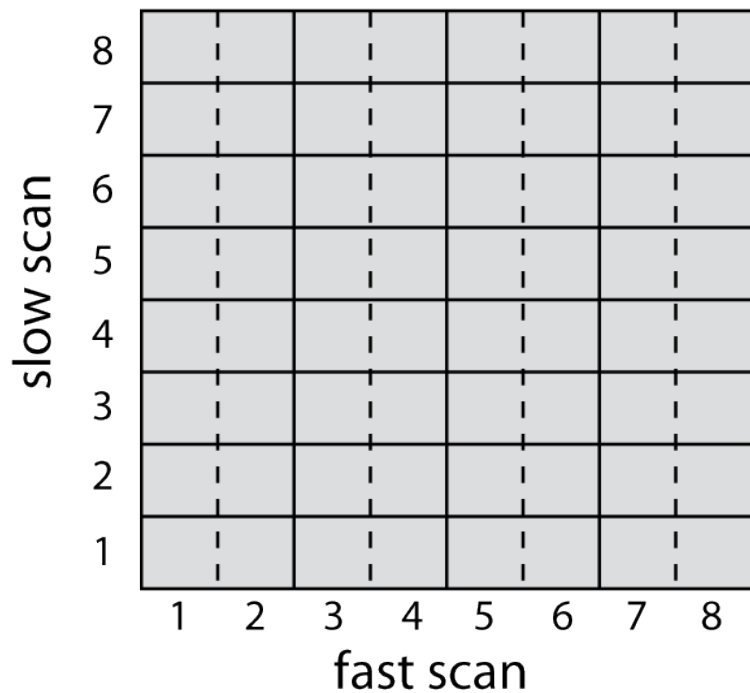
CSPAD geometry: non trivial
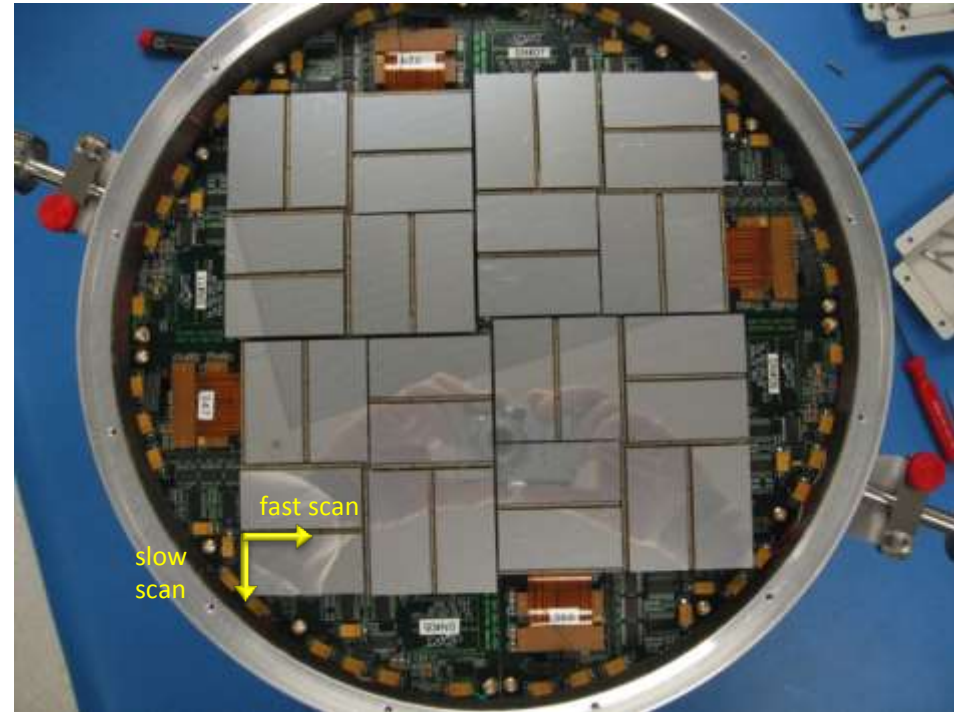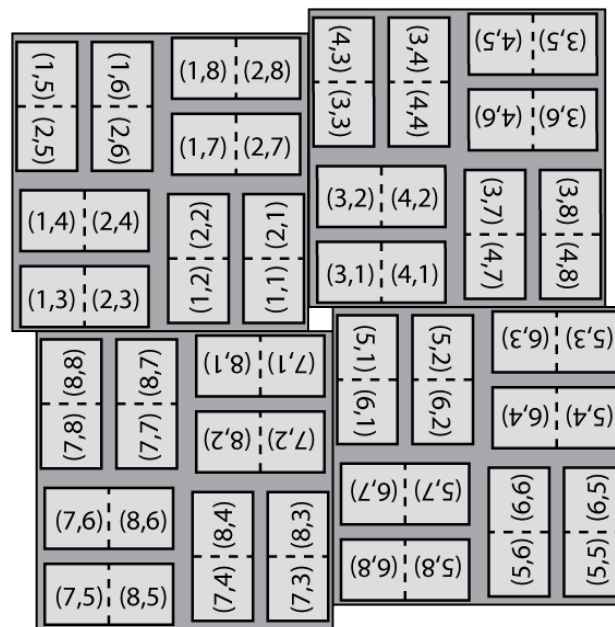

Data layout

Physical layout

H. T. Philipp et al. (2011) *Pixel array detector for x-ray free electron laser experiments.* Nucl Instrum. Methods A **649**, 67.

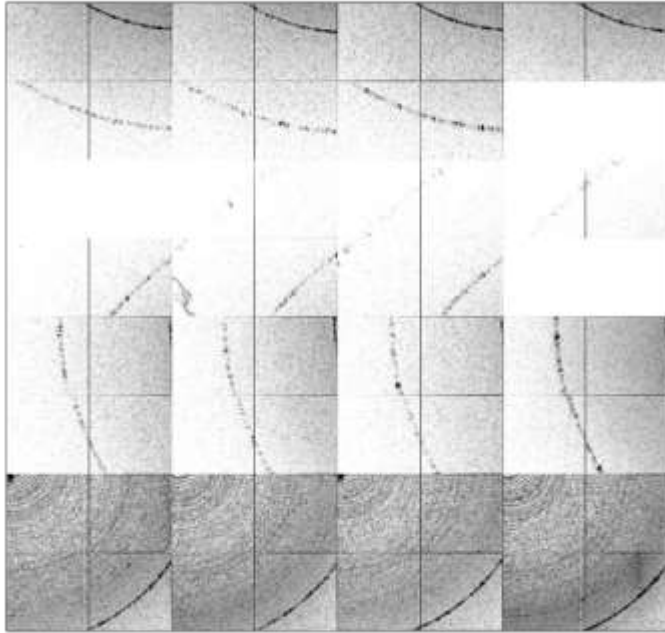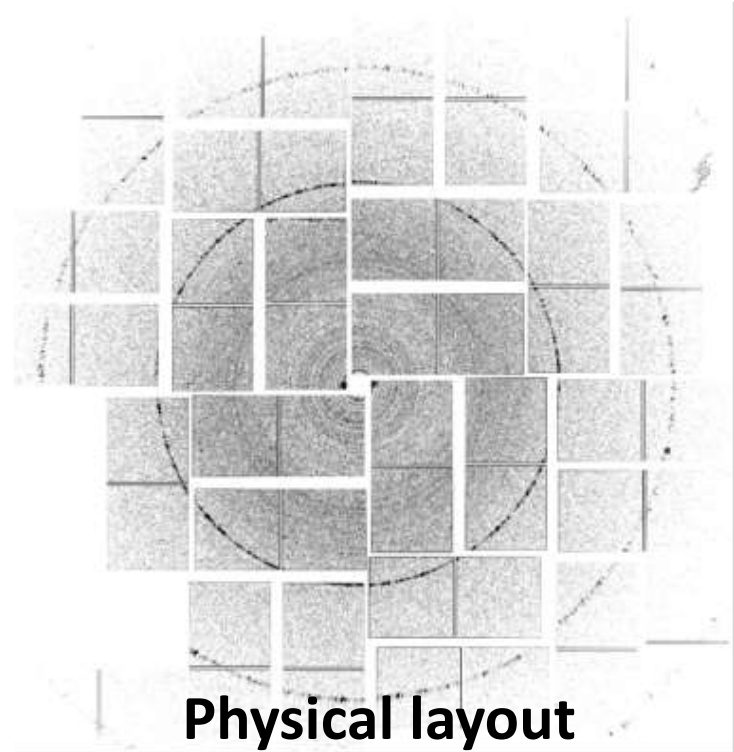# CSPAD geometry: non trivial

Data layout

Physical layout

CSPAD geometry: non trivial


Data layout
Virtual powder

Physical layout
Virtual powder

Data quality is very sensitive to precise geometry.
Comparison of predicted and found peak locations can be used to refine the detector geometry

Figures from
Anton Barty

BioXFEL
A National Science Foundation
Science and Technology Center

ASU

CFEL
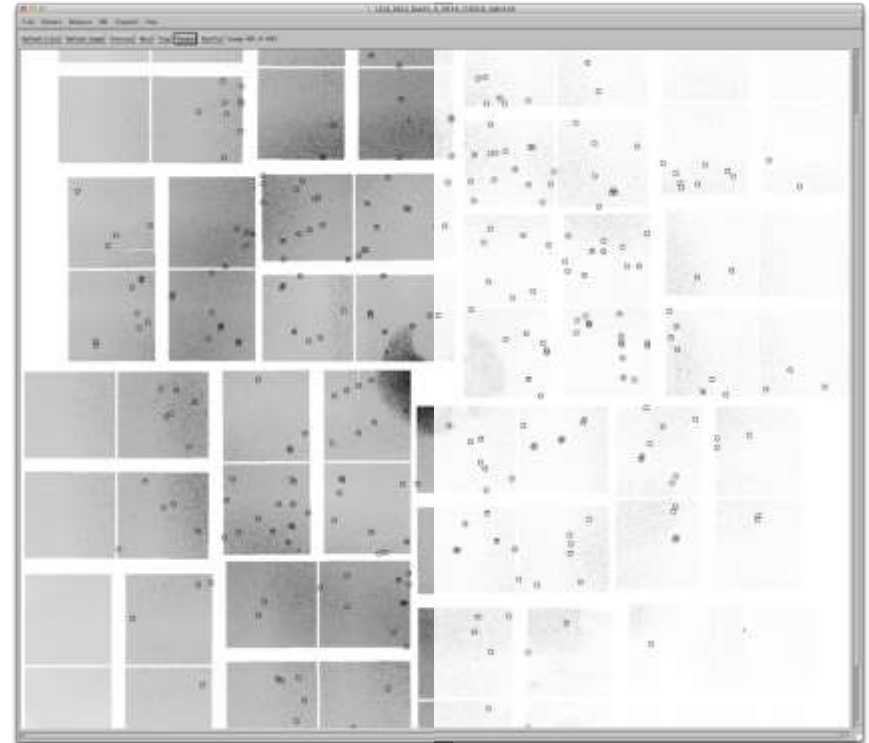SCIENCE

Cheetah: background subtraction

Use data from the current frame to estimate background in current frame
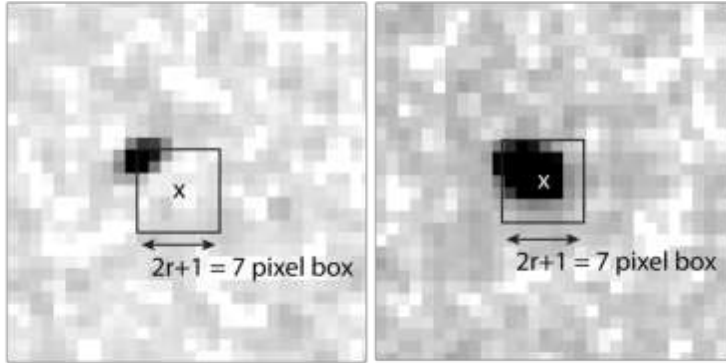
Background = median of pixels in the box
- Assumes peaks are small compared to background variations
- Requires more background pixels in box than peak-containing pixels

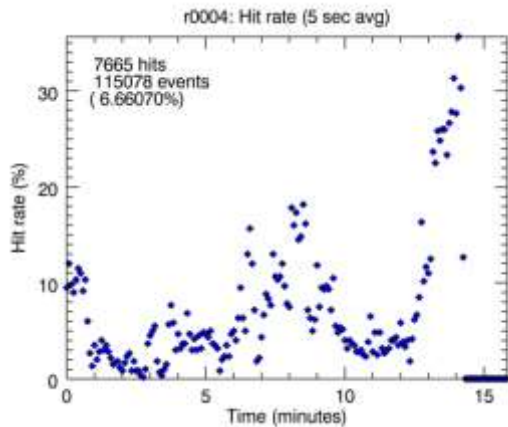**Selection of local background radius is important**

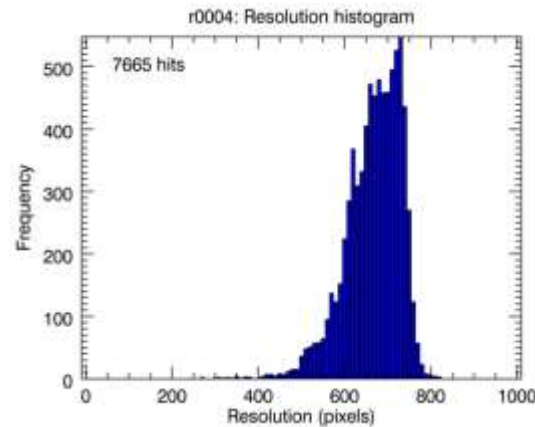**(right) After local background subtraction**

Figures adapted from Anton Barty
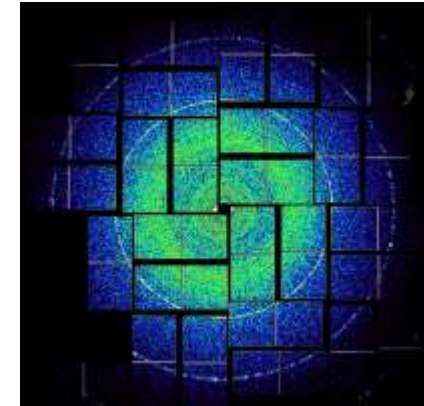
Cheetah: useful diagnostic tool

Hit rates

Resolution

Virtual powder

Radial stacks
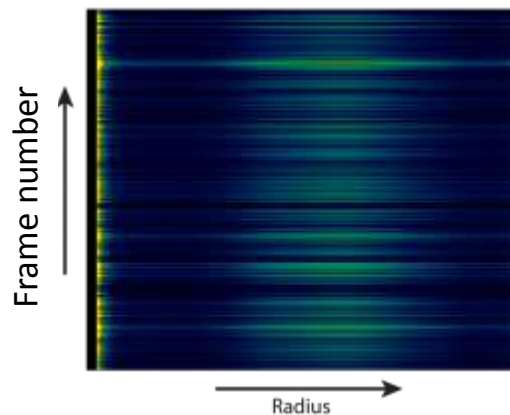
Spectral stack

Saturation plots

Software for serial crystallography
Nadia Zatsepin – IUCr 2014, Aug 5th

Figures from Anton
Barty, Tom Grant

BioXFEL
*A National Science Foundation
Science and Technology Center*

ASU

CFEL
SCIENCE
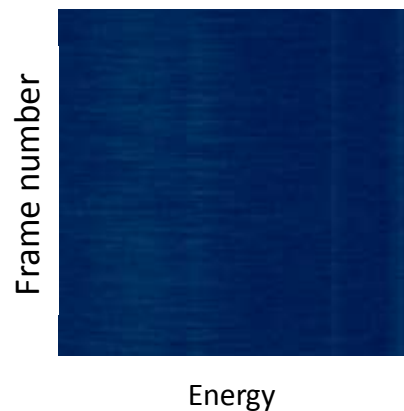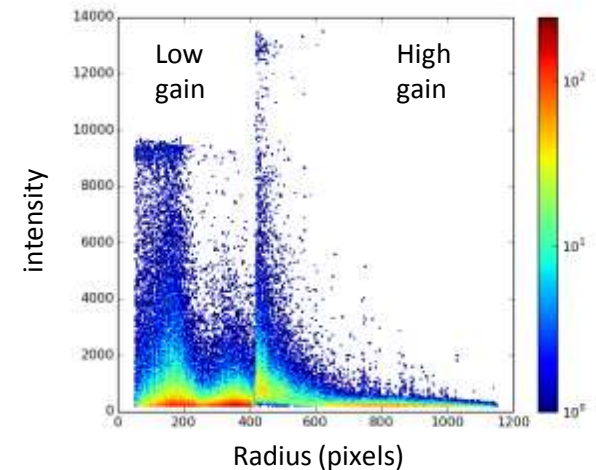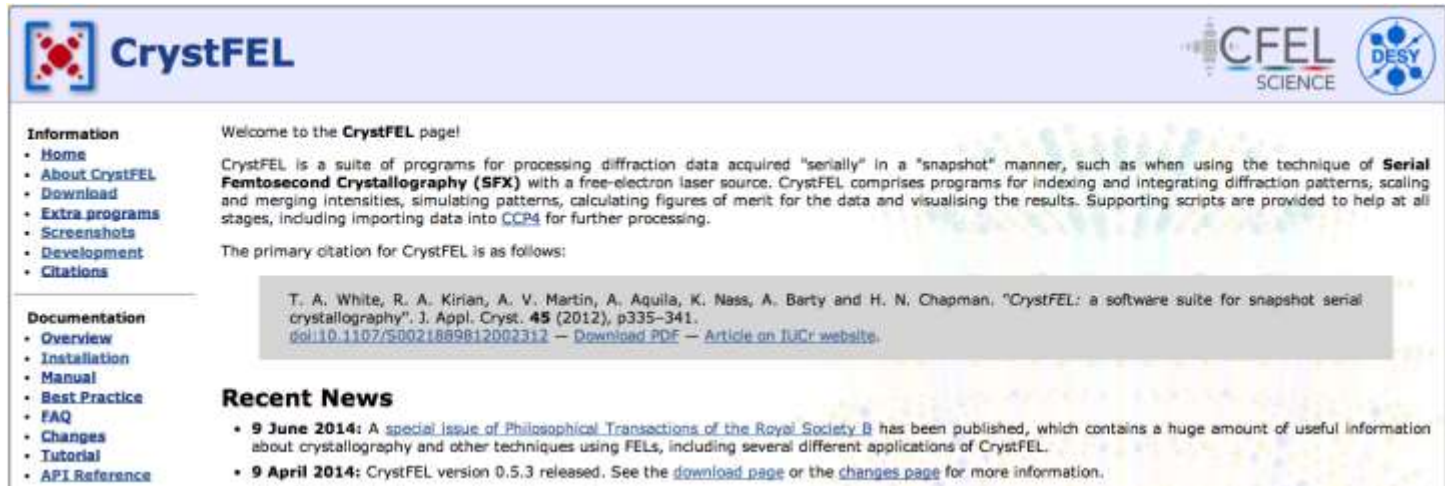
CrystFEL

- Software for serial crystallography data analysis
- Designed for, but not limited to, working with XFEL data
- Free and open-source (GNU GPL3)

www.desy.de/~twhite/crystfel

CrystFEL processing pipeline (each image)

Figure from Tom White

Once parameters have been optimized, processing is completely automatic, and parallelized.

Command line driven, easily called from scripts

Well documented; full tutorial online

CrystFEL's core programs

e.g. R split, CC ½, CC*, CCano, SNR, multiplicity
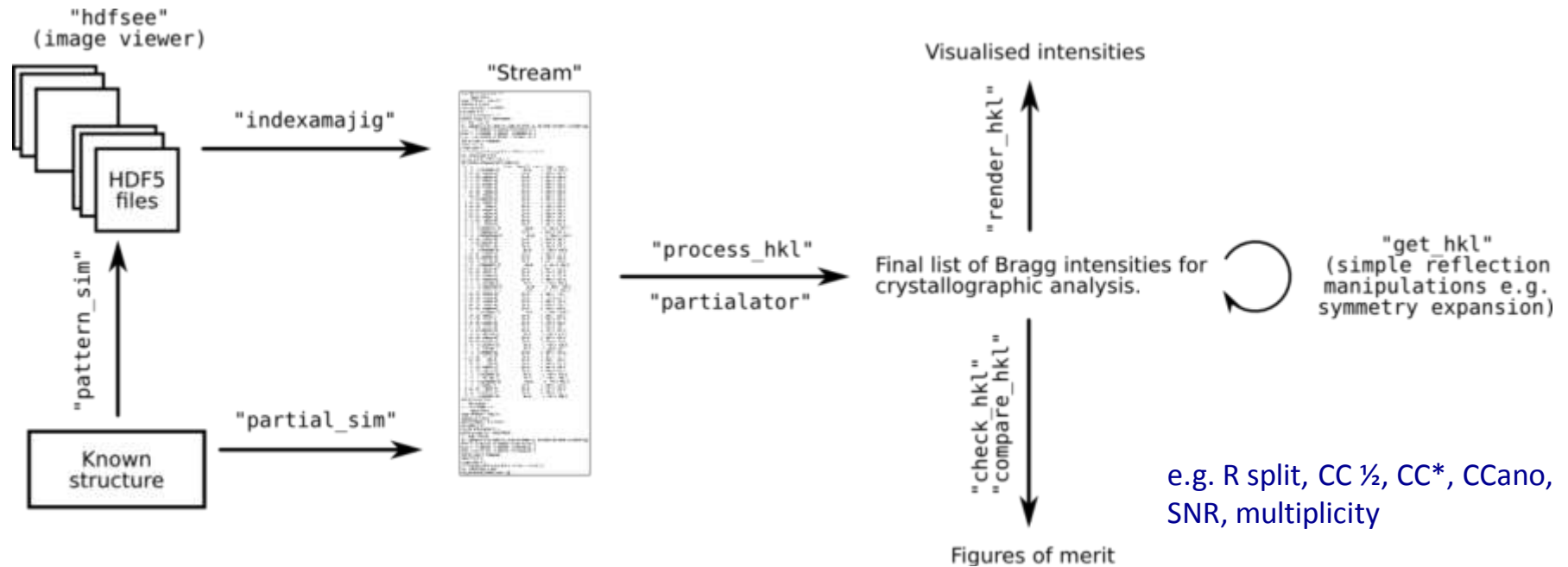
- File formats: HDF5's, everything else is text
- Stream file contains everything about each diffraction pattern: e.g. peaks positions (found; integrated), reflection intensities, cell parameters, indexing method

- Share library contains most of CrystFEL functionality
- External programs can use these functions too

Figure from
Tom White

Unknown unit cell?

CrystFEL's unit cell explorer
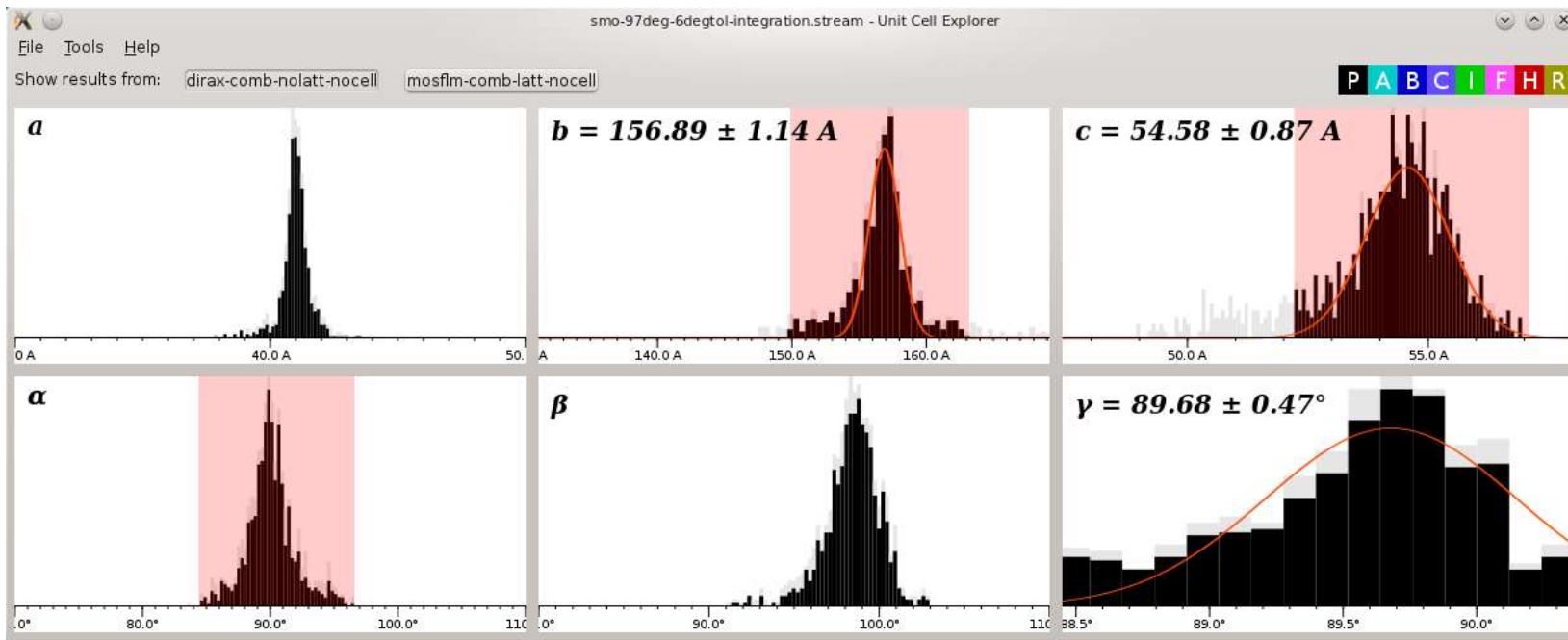
- Displays indexing results from different algorithms / space groups
- Fits Gaussians to calculate mean cell parameters

Indexing ambiguities

If the lattice has higher symmetry than the unit cell, the diffraction pattern can be indexed in more than one orientation since the lattice has the same geometry in various orientations of the crystal

Autoindexing algorithms can't tell these apart, so for SFX data are (initially) merged in higher symmetry
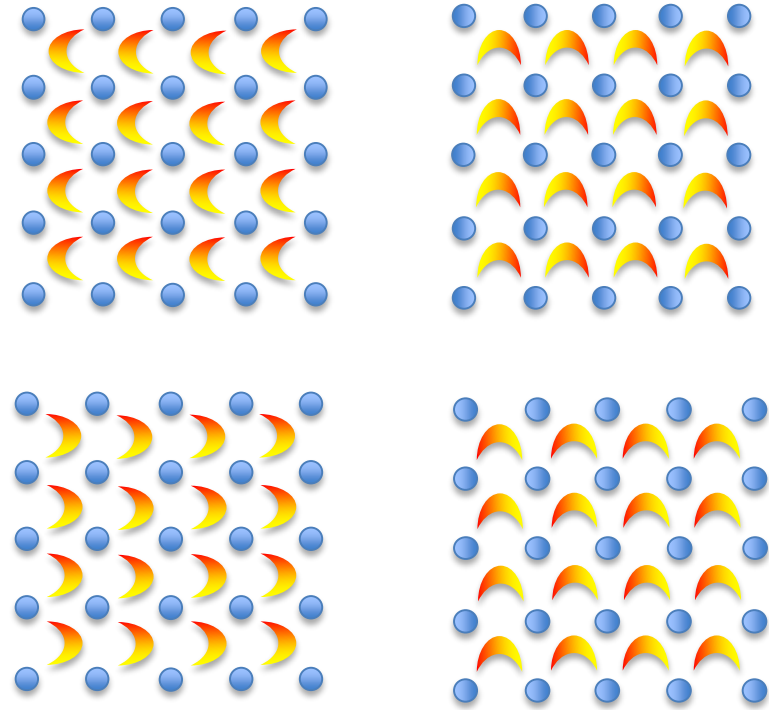
e.g. P63 will have ___ P6322 symmetry

Ambigator: removing the indexing ambiguity

- Large amounts of noise make it difficult to resolve the ambiguity by correlating intensities.
- Ambiguity can be resolved by a **clustering approach:**
- Brehm & Diederichs (2014) Breaking the indexing ambiguity in serial crystallography. Acta Cryst. D70, 101.
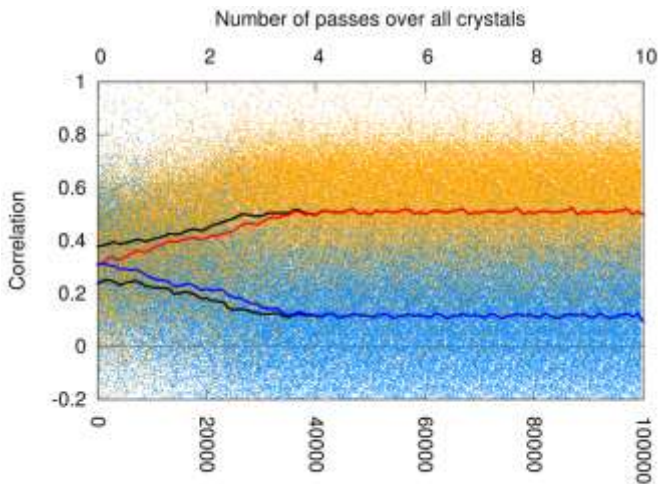
Mean correlation coefficients between individual crystals in two orientations, showing separation into clusters.
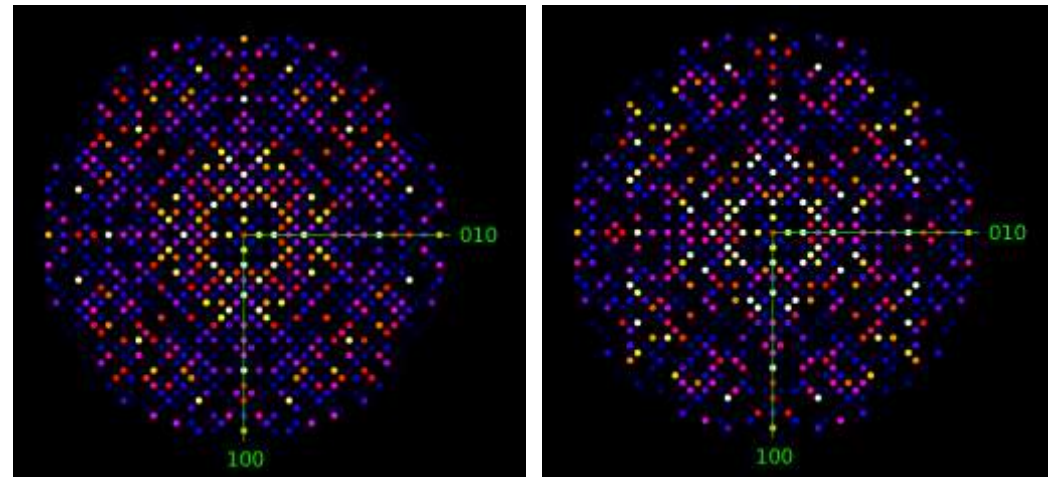
Reciprocal space sections before and after (right) the resolution procedure, showing that the artificial fourfold symmetry arising from the ambiguity has been eliminated.

BioXFEL
A National Science Foundation Science and Technology Center

ASU

CFEL SCIENCE

Why do we need so many patterns?

- Error sources in SFX are not well understood

- SASE: every pulse is like a new experiment

- Shot to shot variations need to be modeled or averaged over
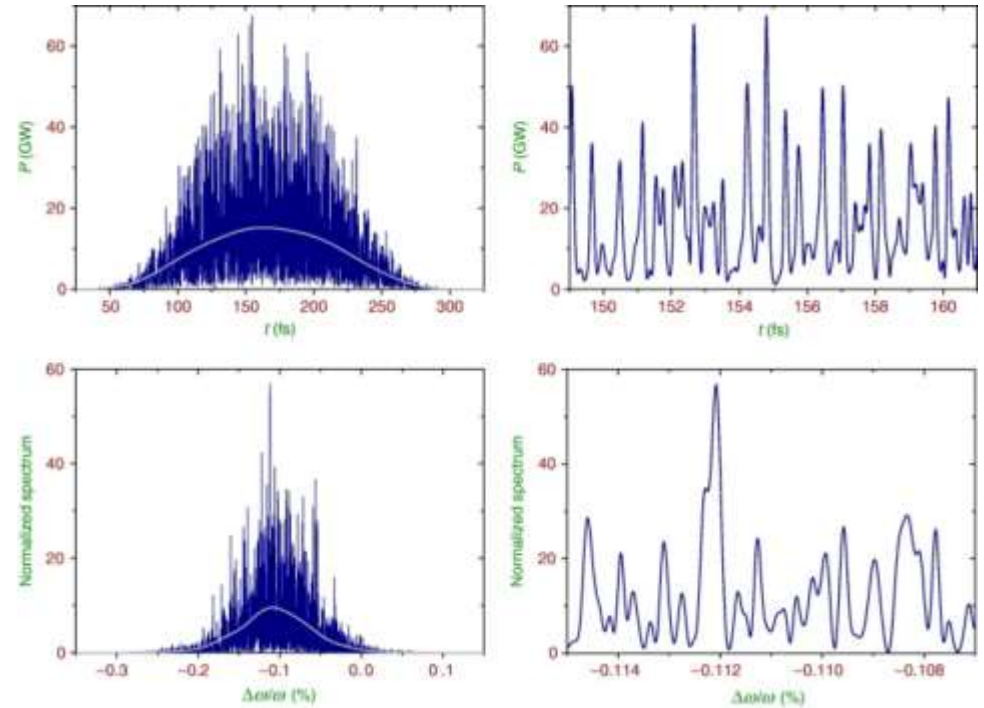
Images: Geloni et al., New J. Phys. 12 (2010) 035021

Why do we need so many patterns?

- Crystal size
- Crystal quality
- Crystal isomorphism
- X-ray pulse bandwidth
- X-ray pulse spectrum
- X-ray pulse intensity
- Position of crystal in the beam
- Partially recorded reflections

...

k_{in}

http://xray.bmc.uu.se/spb/
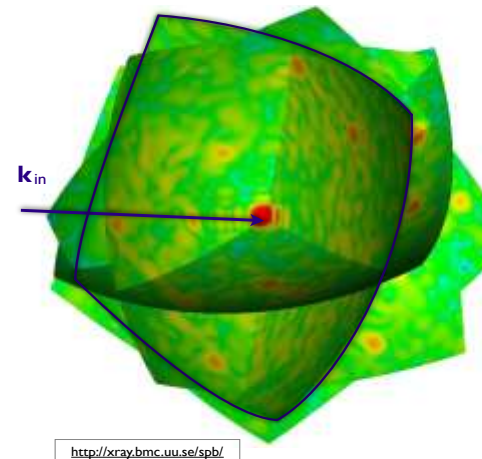
An angular integration over the Bragg spot is needed to obtain the structure factor.
One snapshot gives partial reflections.

R. A. Kirian et al. (2010) Femtosecond protein nanocrystallography - data analysis methods. Opt. Exp. 18, 5713

ASU

CFEL SCIENCE

partiality

Incident beam direction

Limiting Ewald spheres

Origin of reciprocal space

A

B C

D E
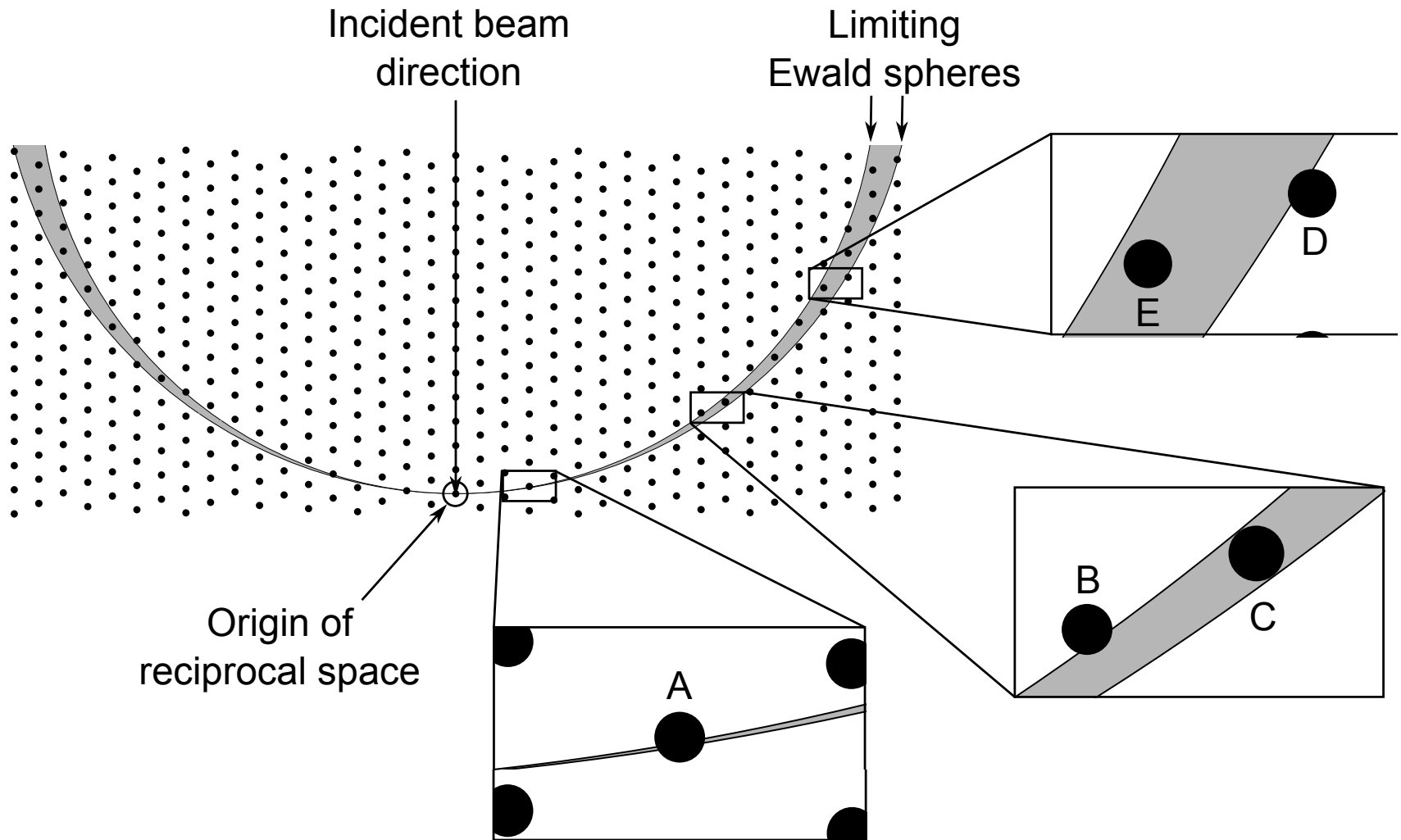
Figure from White, T. (2014) *Phil. Trans. R. Soc. B* 369, 20130330 .

BioXFEL
A National Science Foundation Science and Technology Center

ASU

CFEL SCIENCE

Solving the partiality problem

White, T. (2014) Post-refinement method for snapshot serial crystallography. *Phil. Trans. R. Soc. B* 369, 20130330

Algorithm similar to: Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. Acta Cryst D. 66, 133.

Solving the partiality problem
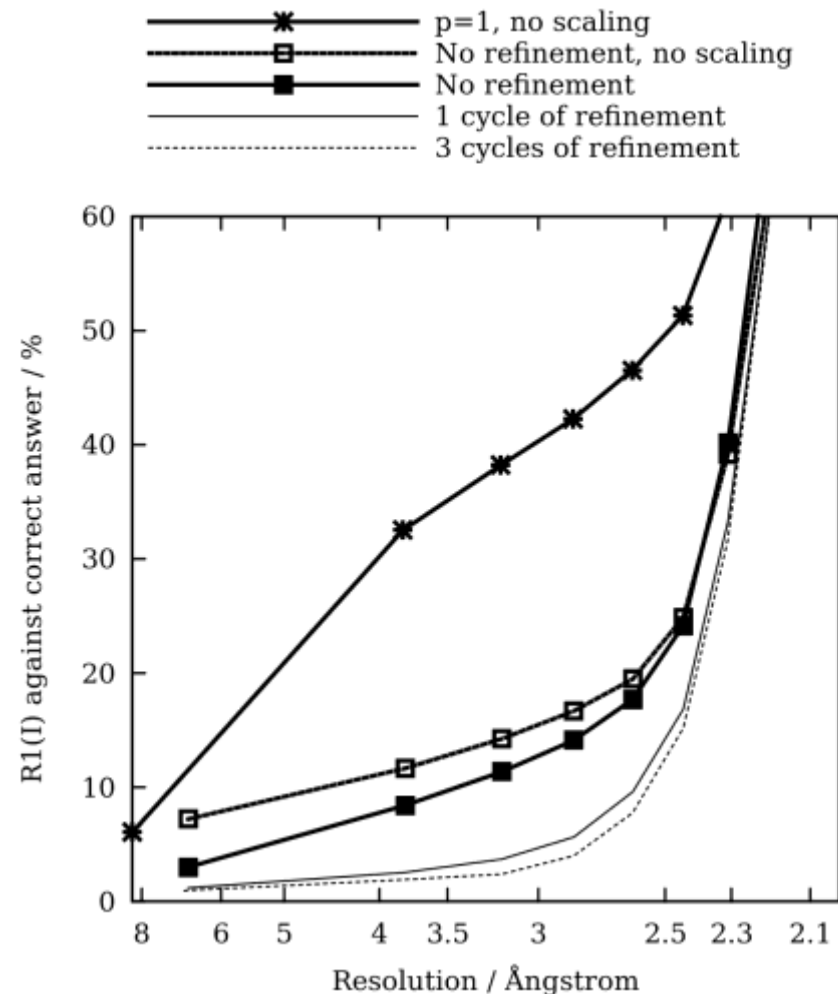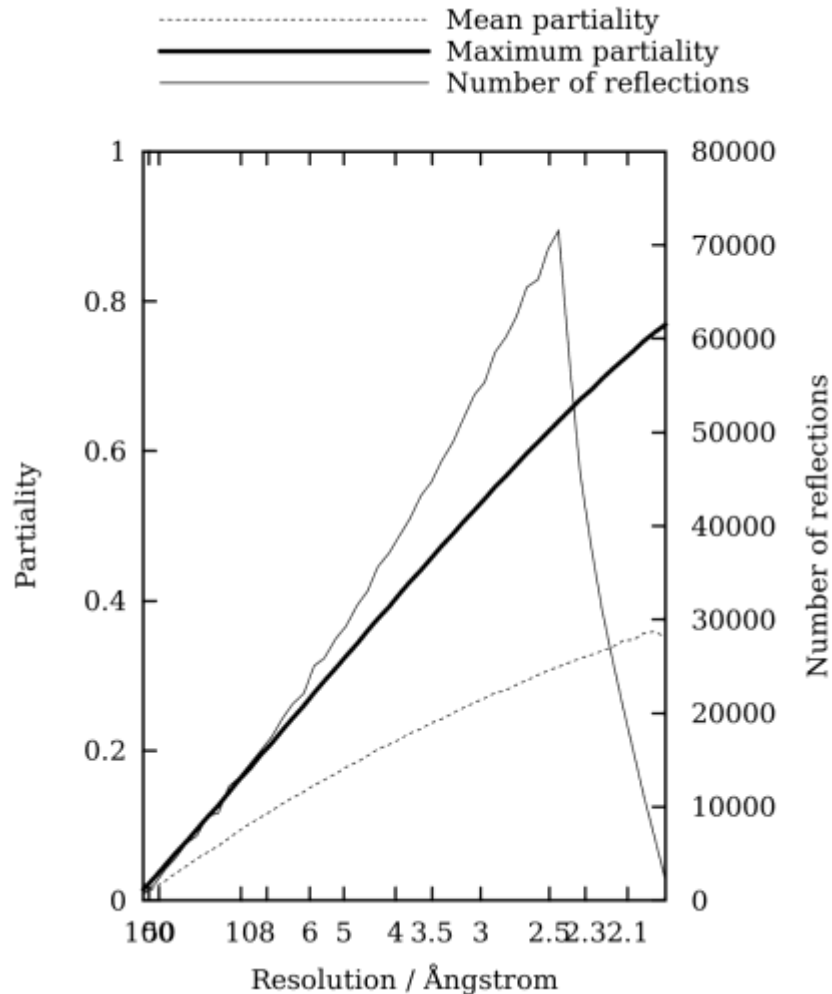
Figures from Tom White.
See: White, T. (2014) *Phil. Trans. R. Soc. B* 369, 20130330 .

Solving the partiality problem
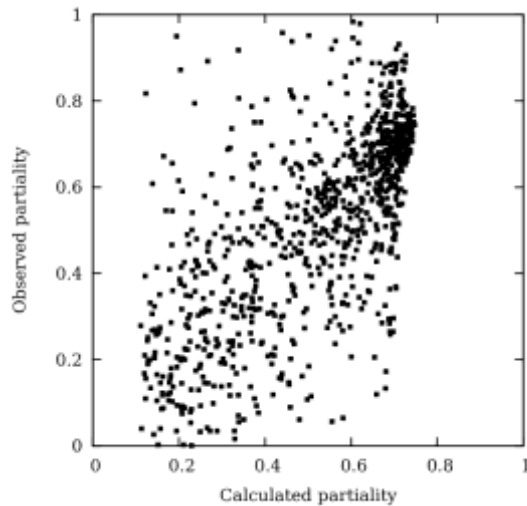
Initial condition
(0.1% cell error)
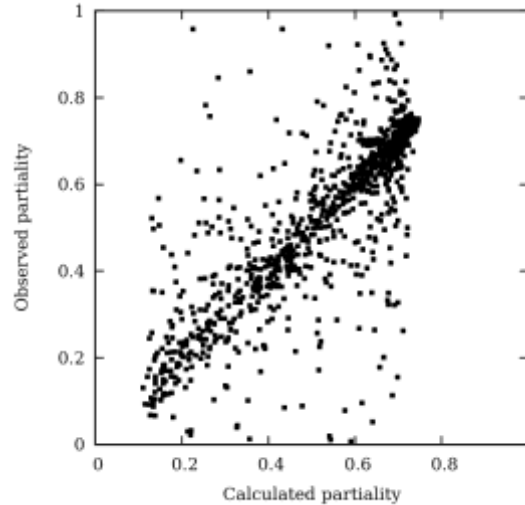
After 1 cycle of
post refinement
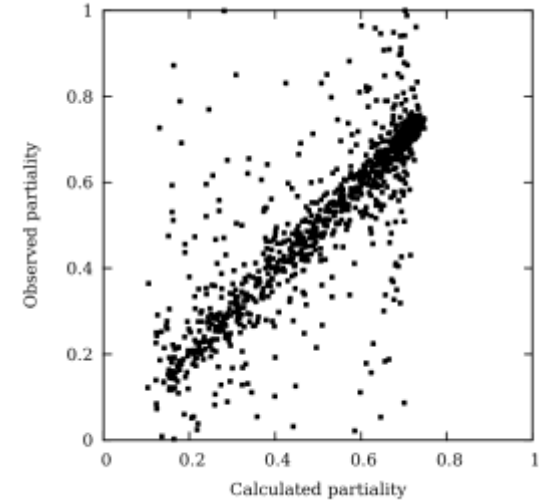
After 3 cycles of
post refinement

White, T. (2014) Post-refinement method for snapshot serial crystallography. *Phil. Trans. R. Soc. B* 369, 20130330 .

Algorithm similar to: Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. Acta Cryst D. 66, 133.

Figures from Tom White.
See: White, T. (2014) *Phil. Trans. R. Soc. B* 369, 20130330 .

- Serial crystallography has been succesfully carried out at synchrotrons
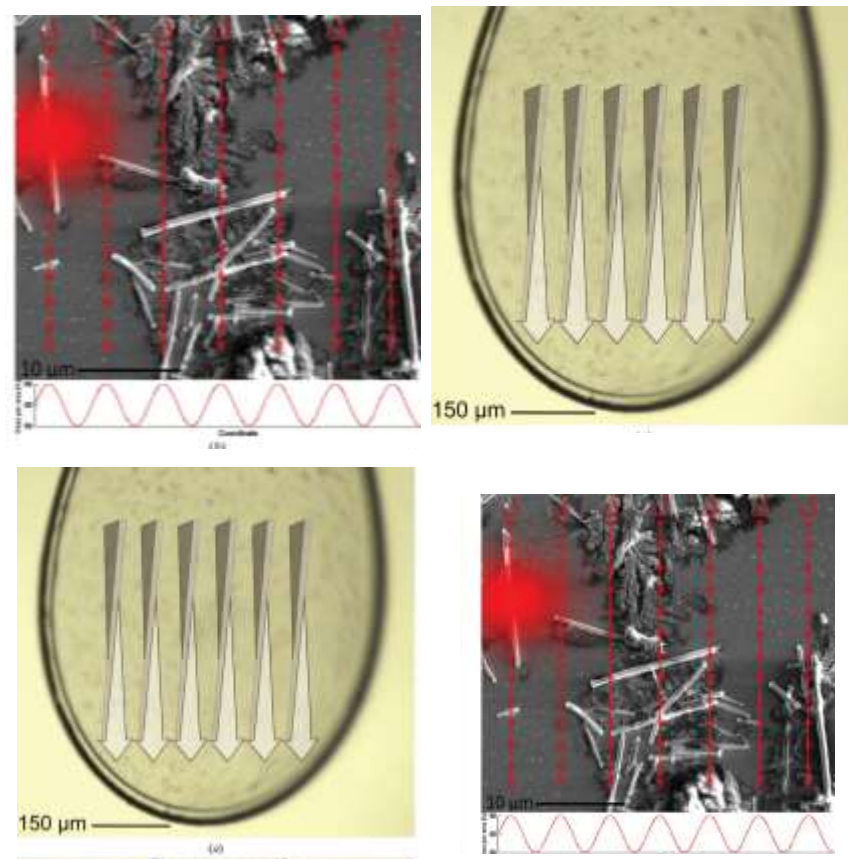
Figures from Gati, C. et al. Serial crysta-llography on in vivo grown microcrystals using synchrotron radiation. IUCrJ 2014.

CrystFEL in the PDB

Some results from LCLS data processed with Cheetah and CrystFEL
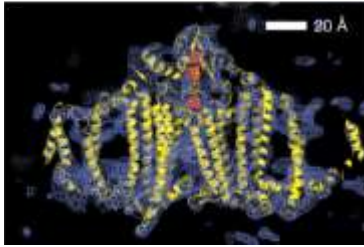
3PCQ
Photosystem I (the
first SFX experiment)
Chapman et al., Nature 2011

4HWY
Natively inhibitied
Cathepsin B
Redecke, Nass et al.,
Science 2013
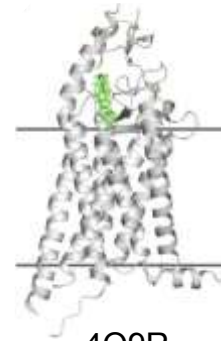
4O9R
Smoothened receptor
using LCP injector
Weierstall et al., Nature
Communications 2014
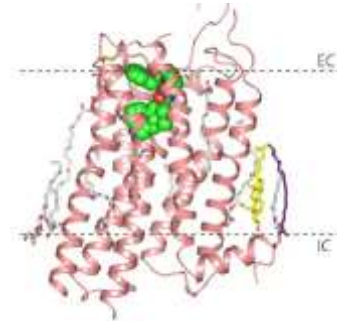
4NC3
Serotonin receptor
bound to ergotamine
Liu et al., Science 2013
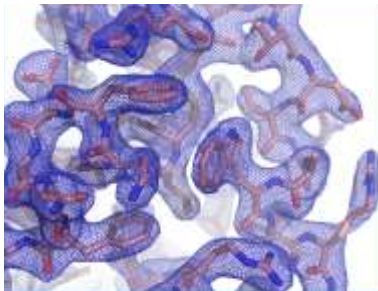
4O34
Serial crystallography using
a synchrotron beamline
Stellato et al., IUCrJ 2014
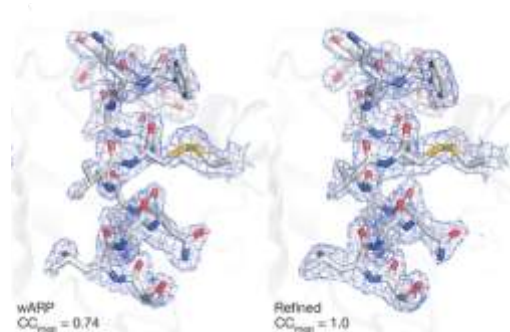
4N5R
Lysozyme (Gd derivative)
ab initio phasing using SAD
Barends et al., Nature 2013
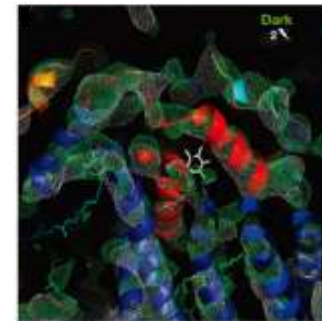
4Q54
Photosystem II in S_3
excited state
Kupitz, Basu et al., Nature 2014

& more

BioXFEL
A National Science Foundation
Science and Technology Center

ASU

CFEL
SCIENCE

More serial crystallography software

- Psana - LCLS analysis frameworks. Python interface
- CASS – data reduction (no indexing etc)
  – Foucar, L. et al. CASS—CFEL-ASG software suite. (2012) Comp. Phys. Comm. 183 (10) 2207-2213.
- cctbx.xfel – data reduction and SFX processing
  – Built on Computational Crystallography ToolBox: same as Phenix and LABELIT
  – Sauter N. et al. Acta Cryst. (2013). D69, 1274
  – Future: post refinement
- nXDS – Kabsch, Acta D. 2014
  – Not Monte Carlo
  – Future: model X-ray spectrum; consider shape transforms from nanocrystals

Workshops: where to learn more

- Bio-XFEL data analysis workshop August 21-22, 2014
  - Organized by BioXFEL STC, LBNL and SLAC, at Berkeley.
  - www.bioxfel.org/events/details/9
  - There will be a Bio-XFEL data analysis workshop in 2015

- **SSRL/LCLS User meeting 2014: Oct 10, 2014**
  - Software for Serial Crystallography (1 day workshop)
  - Spaces available – book now.
  - https://conf-slac.stanford.edu/ssrl-lcls-2014/welcome

- **BioXFEL 2nd International Conference – Puerto Rico, Jan 14-16, 2015**
  - www.bioxfel.org/events/details/6

- **ACA 2015 (July, PA): SFX data analysis workshop (possibly)**

www.desy.de/~barty/cheetah & www.desy.de/~twhite/crystfel/

CrystFEL citations

2014

- C. Kupitz, S. Basu, I. Grotjohann, R. Fromme et al. "Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser". Nature (2014). doi:10.1038/nature13453 and PDB entries 4PBU and 4Q54.

- M. Caffrey, D. Li, N. Howe and S. T. A. Shah. "'Hit and run' serial femtosecond crystallography of a membrane kinase in the lipid cubic phase". Phil Trans. Roy. Soc. B 369 (2014) 20130621. doi:10.1098/rstb.2013.0621 (**open access** - download PDF).

- O. Yefanov, C. Gati, G. Bourenkov, R. A. Kirian et al. "Mapping the continuous reciprocal space intensity distribution of X-ray serial crystallography". Phil. Trans. Roy. Soc. B 369 (2014) 20130333 doi:10.1098/rstb.2013.0333 (**open access** - download PDF).

- T. A. White. "Post-refinement method for snapshot serial crystallography". Phil. Trans. Roy. Soc. B 369 (2014) 20130330. doi:10.1098/rstb.2013.0330 (**open access** - download PDF).

- J. C. H. Spence, N. A. Zatsepin and C. Li. "Coherent convergent-beam time-resolved X-ray diffraction". Phil. Trans. Roy. Soc. B 369 (2014) 20130325 doi:10.1098/rstb.2013.0325.

- F. Stellato, D. Oberthür, M. Liang, R. Bean et al. "Room-temperature macromolecular serial crystallography using synchrotron radiation". IUCrJ 1 (2014). doi:10.1107/S2052252514010070 (**open access** - download PDF) and PDB entry 4O34.

- K. Qu, L. Zhou and Y.-H. Dong. "An improved integration method in serial femtosecond crystallography". Acta Cryst. D70 (2014) p1202. doi:10.1107/S1399004714002004.

- M. Frank, D. B. Carlson, M. S. Hunter, G. J. Williams et al. "Femtosecond X-ray diffraction from two-dimensional protein crystals". IUCrJ 1 (2014) p95. doi:10.1107/S2052252514001444 (**open access** - download PDF).

- C. Gati, G. Bourenkov, M. Klinge, D. Rehders et al. "Serial crystallography on in vivo grown microcrystals using synchrotron radiation". IUCrJ 1 (2014) p87. doi:10.1107/S2052252513033939 (**open access** - download PDF) and PDB entry 4N4Z.

- U. Weierstall, D. James, C. Wang, T. A. White et al. "Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography". Nature Communications 5:3309 (2014). doi:10.1038/ncomms4309 and PDB entry 4O9R.

- C. Song, K. Tono, J. Park, T. Ebisu et al. "Multiple application X-ray imaging chamber for single-shot diffraction experiments with femtosecond X-ray laser pulses". J. Appl. Cryst. 47 (2014) p18. doi:10.1107/S1600576713029944.

CrystFEL citations

2013

- W. Liu, D. Wacker, C. Gati, G. W. Han et al. "Serial Femtosecond Crystallography of G Protein-Coupled Receptors". Science 342 (2013) p1522. doi:10.1126/science.1244142, PDB entry 4NC3 and CXIDB ID 21.

- L. C. Johansson, D. Arnlund, G. Katona, T. A. White et al. "Structure of a photosynthetic reaction centre determined by serial femtosecond crystallography". Nature Communications 4:2911 (2013). doi:10.1038/ncomms3911 (**open access** - download PDF) and PDB entry 4CAS.

- T. R. M. Barends, L. Foucar, S. Botha, R. B. Doak et al. "De novo protein crystal structure determination from X-ray free-electron laser data". Nature (2013). doi:10.1038/nature12773, PDB entry 4N5R and CXIDB ID 20.

- H. Dermirci, R. G. Sierra, H. Laksmono, R. L. Shoeman et al. "Serial femtosecond X-ray diffraction of 30S ribosomal subunit microcrystals in liquid suspension at ambient temperature using an X-ray free-electron laser". Acta Cryst. F69 (2013) p1066. doi:10.1107/S174430911302099X (**open access** - download PDF).

- T. A. White, A. Barty, F. Stellato, J. M. Holton, R. A. Kirian, N. A. Zatsepin and H. N. Chapman. "Crystallographic data processing for free-electron laser sources". Acta Cryst. D69 (2013), p1231. doi:10.1107/S0907444913013620.

- T. R. M. Barends, L. Foucar, R. L. Shoeman, S. Bari et al. "Anomalous signal from S atoms in protein crystallographic data from an X-ray free-electron laser". Acta Cryst. D69 (2013) p838. doi:10.1107/S0907444913002448 (**open access** - download PDF).

- L. Redecke, K. Nass et al. "Natively Inhibited Trypanosoma brucei Cathepsin B Structure Determined by Using an X-ray Laser". Science 339 (2013) p227. doi:10.1126/science.1229663 and PDB entry 4HWY.

2012

- S. Boutet, L. Lomb, G. J. Williams, T. R. M. Barends et al. "High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography". Science 337 (2012) p362. doi:10.1126/science.1217737, PDB entries 4ET8 and 4ET9, and CXIDB ID 17.

- L. C. Johansson, D. Arnlund, T. A. White, G. Katona et al. "Lipidic phase membrane protein serial femtosecond crystallography". Nature Methods 9 (2012) p263. doi:10.1038/nmeth.1867 and PDB entry 4AC5.

- R. Koopmann, K. Cupelli, L. Redecke, K. Nass et al. "In vivo protein crystallization opens new routes in structural biology". Nature Methods 9 (2012) p259. doi:10.1038/nmeth.1859.

- A. Aquila, M. S. Hunter, R. B. Doak et al. "Time-resolved protein nanocrystallography using an X-ray free-electron laser". Optics Express 20 (2012) p2706. doi:10.1364/OE.20.002706 (**open access** - download PDF).

CrystFEL citations

- A. Barty, C. Caleman, A. Aquila, N. Timneanu et al. "Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements". Nature Photonics 6 (2012) p35. doi:10.1038/nphoton.2011.297.

2011

- H. N. Chapman, P. Fromme et al. "Femtosecond X-ray protein nanocrystallography". Nature 470 (2011) p73. doi:10.1038/nature09750 and PDB entry 3PCQ.

Useful references

Here are some papers describing CrystFEL or its algorithms:

- T. A. White. "Post-refinement method for snapshot serial crystallography". Phil. Trans. Roy. Soc. B **369** (2014) 20130330. doi:10.1098/rstb.2013.0330 — Download PDF

- T. A. White, A. Barty, F. Stellato, J. M. Holton, R. A. Kirian, N. A. Zatsepin and H. N. Chapman. "Crystallographic data processing for free-electron laser sources". Acta Cryst. **D69** (2013), p1231–1240. doi:10.1107/S0907444913013620 — Download PDF — Article on IUCr website

- T. A. White, R. A. Kirian, A. V. Martin, A. Aquila, K. Nass, A. Barty and H. N. Chapman. "*CrystFEL:* a software suite for snapshot serial crystallography". J. Appl. Cryst. **45** (2012), p335–341. doi:10.1107/S0021889812002312 — Download PDF — Article on IUCr website

- R. A. Kirian, X. Wang, U. Weierstall, K. E. Schmidt, J. C. H. Spence, M. Hunter, P. Fromme, T. White, H. N. Chapman and J. Holton. "Femtosecond protein nanocrystallography - data analysis methods". Optics Express **18** (2010) p5713. doi:10.1364/OE.18.005713 — Download PDF

Bonus: CSPAD

CSPAD

- 110 x 110 µm^2 pixels

- 185 x 194 pixels per ASIC

- 32 modules (2 ASICs)

- Quadrants are independently movable to change size of hole in center (instead of using a beam stop)

- 2 gain settings (each pixel)

- Signal-to-noise ~ 3.5 (high gain mode)

- Dynamic range of about 350 photons at 9.4 keV