

# Crystallographic data processing for free-electron laser sources

Thomas White  
Center for Free-Electron Laser Science  
DESY Hamburg, Germany



# Collaborators

Henry N. Chapman<sup>1,2</sup>, Petra Fromme<sup>3</sup>, Anton Barty<sup>1</sup>, Thomas A. White<sup>1</sup>, Richard A. Kirian<sup>4</sup>, Andrew Aquila<sup>1</sup>, Mark S. Hunter<sup>3</sup>, Joachim Schulz<sup>1</sup>, Daniel P. DePonte<sup>1</sup>, Uwe Weierstall<sup>4</sup>, R. Bruce Doak<sup>4</sup>, Filipe R.N.C. Maia<sup>5</sup>, Andrew V. Martin<sup>1</sup>, Ilme Schlichting<sup>6,7</sup>, Lukas Lomb<sup>7</sup>, Nicola Coppola<sup>1</sup>, Robert L. Shoeman<sup>7</sup>, Sascha W. Epp<sup>6,8</sup>, Robert Hartmann<sup>9</sup>, Daniel Rolles<sup>6,7</sup>, Artem Rudenko<sup>6,8</sup>, Lutz Foucar<sup>6,7</sup>, Nils Kimmel<sup>10</sup>, Georg Weidenspointner<sup>11,10</sup>, Peter Holl<sup>9</sup>, Mengning Liang<sup>1</sup>, Miriam Barthelmess<sup>12</sup>, Carl Caleman<sup>1</sup>, Sébastien Boutet<sup>13</sup>, Michael J. Bogan<sup>14</sup>, Jacek Krzywinski<sup>13</sup>, Christoph Bostedt<sup>13</sup>, Saša Bajt<sup>12</sup>, Lars Gumprecht<sup>1</sup>, Benedikt Rudek<sup>6,8</sup>, Benjamin Erk<sup>6,8</sup>, Carlo Schmidt<sup>6,8</sup>, André Hömke<sup>6,8</sup>, Christian Reich<sup>9</sup>, Daniel Pietschner<sup>10</sup>, Lothar Strüder<sup>6,10</sup>, Günter Hauser<sup>10</sup>, Hubert Gorke<sup>15</sup>, Joachim Ullrich<sup>6,8</sup>, Sven Herrmann<sup>10</sup>, Gerhard Schaller<sup>10</sup>, Florian Schopper<sup>10</sup>, Heike Soltau<sup>9</sup>, Kai-Uwe Kühnel<sup>8</sup>, Marc Messerschmidt<sup>13</sup>, John D. Bozek<sup>13</sup>, Stefan P. Hau-Riege<sup>16</sup>, Matthias Frank<sup>16</sup>, Christina Y. Hampton<sup>14</sup>, Raymond Sierra<sup>14</sup>, Dmitri Starodub<sup>14</sup>, Garth J. Williams<sup>13</sup>, Janos Hajdu<sup>5</sup>, Nicusor Timneanu<sup>5</sup>, M. Marvin Seibert<sup>5</sup>, Jakob Andreasson<sup>5</sup>, Andrea Rucker<sup>5</sup>, Olof Jönsson<sup>5</sup>, Stephan Stern<sup>1</sup>, Francesco Stellato<sup>1</sup>, Karol Nass<sup>2</sup>, Robert Andritschke<sup>10</sup>, Claus-Dieter Schröter<sup>8</sup>, Faton Krasniqi<sup>6,7</sup>, Mario Bott<sup>7</sup>, Kevin E. Schmidt<sup>4</sup>, Xiaoyu Wang<sup>4</sup>, Ingo Grotjohann<sup>3</sup>, James Holton<sup>17</sup>, Stefano Marchesini<sup>17</sup>, Raimund Fromme<sup>3</sup>, Sebastian Schorb<sup>18</sup>, Daniela Rupp<sup>18</sup>, Marcus Adolph<sup>18</sup>, Tais Gorkhover<sup>18</sup>, Martin Svenda<sup>5</sup>, Helmut Hirsemann<sup>12</sup>, Guillaume Potdevin<sup>12</sup>, Heinz Graafsma<sup>12</sup>, Björn Nilsson<sup>12</sup> and John C. H. Spence<sup>4</sup>.

**1.** Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany. **2.** University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. **3.** Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604 USA. **4.** Department of Physics, Arizona State University, Tempe, Arizona 85287 USA. **5.** Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 (Box 596), SE-751 24 Uppsala, Sweden. **6.** Max Planck Advanced Study Group, Center for Free Electron Laser Science (CFEL), Notkestrasse 85, 22607 Hamburg, Germany. **7.** Max-Planck-Institut für medizinische Forschung, Jahnstr. 29, 69120 Heidelberg, Germany. **8.** Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, 69117 Heidelberg, Germany. **9.** PNSensor GmbH, Otto-Hahn-Ring 6, 81739 München, Germany. **10.** Max-Planck-Institut Halbleiterlabor, Otto-Hahn-Ring 6, 81739 München, Germany. **11.** Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, 85741 Garching, Germany. **12.** Photon Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany. **13.** LCLS, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. **14.** PULSE Institute and National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. **15.** Forschungszentrum Jülich, Institut ZEL, 52425 Jülich, Germany. **16.** Lawrence Livermore National Laboratory, 7000 East Avenue, Mail Stop L-211, Livermore, CA 94551, USA. **17.** Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. **18.** Institut für Optik und Atomare Physik, Technische Universität Berlin, Hardenbergstrasse 36, 10623 Berlin, Germany.

# The PIs

- ▶ **Henry Chapman**

Center for Free-Electron Laser Science, DESY Hamburg.

- ▶ **John Spence**

Department of Physics, Arizona State University.

- ▶ **Petra Fromme**

Department of Chemistry and Biochemistry, Arizona State University.

- ▶ **Ilme Schlichting**

Max-Planck-Institut für medizinische Forschung, Heidelberg.

- ▶ **Janos Hadju**

Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University.

- ▶ **Richard Neutze**

Department of Chemistry, Biochemistry and Biophysics, University of Gothenburg.

# The CAMP Team

The CAMP instrument was designed and commissioned by the Max Planck  
CFEL Advanced Study Group

Sascha Epp<sup>1</sup>, Robert Hartmann<sup>1,2</sup>, Daniel Rolles<sup>1</sup>, Artem Rudenko<sup>1</sup>, Lutz Foucar<sup>1</sup>,  
Benedikt Rudek<sup>1</sup>, Benjamin Erk<sup>1</sup>, Carlo Schmidt<sup>1</sup>, André Hömke<sup>1</sup>, Nils Kimmel<sup>2</sup>, Christian Reich<sup>2</sup>,  
Günther Hauser<sup>2</sup>, Daniel Pietschner<sup>2</sup>, Peter Holl<sup>2</sup>, Hubert Gorke<sup>3</sup>, Helmut Hirsemann<sup>4</sup>,  
Guillaume Potdevin<sup>4</sup>, Tim Erke<sup>4</sup>, Jan-Henrik Mayer<sup>4</sup>, Heinz Graafsma<sup>4</sup>, Michael Matysek<sup>5</sup>,  
Sebastian Schorb<sup>6</sup>, Daniela Rupp<sup>6</sup>, Marcus Adolph<sup>6</sup>, Tais Gorkhover<sup>6</sup>, Christoph Bostedt<sup>7</sup>,  
John Bozek<sup>7</sup>, Marc Messerschmidt<sup>7</sup>, Joachim Schulz<sup>4</sup>, Lars Gumprecht<sup>4</sup>, Andrew Aquila<sup>4</sup>,  
Nicola Coppola<sup>4</sup>, Frank Filsinger<sup>8</sup>, Kai-Uwe Kühnel<sup>9</sup>, Christian Kaiser<sup>9</sup>, Claus-Dieter Schröter<sup>9</sup>,  
Robert Moshhammer<sup>9</sup>, Faton Krasniqi<sup>1</sup>, Simone Techert<sup>1,10</sup>, Georg Weidenspointer<sup>2</sup>,  
Robert L. Shoeman<sup>11</sup>, Ilme Schlichting<sup>1,11</sup>, Lothar Strüder<sup>1,2</sup> and Joachim Ullrich<sup>1,9</sup>

**1** Max Planck Advanced Study Group at CFEL, 22761 Hamburg, Germany. **2** Max Planck Halbleiterlabor, 81739 München, Germany.

**3** FZ Jülich, 52428 Jülich, Germany. **4** Deutsches Elektronen Synchrotron, 22607 Hamburg, Germany.

**5** Universität Hamburg, 22607 Hamburg, Germany. **6** Technische Universität Berlin, 10623 Berlin, Germany. **7** LCLS, Menlo Park, USA.

**8** Fritz-Haber-Institut der MPG, Berlin, Germany. **9** Max-Planck-Institut für Kernphysik, 69117 Heidelberg, Germany.

**10** Max-Planck-Institut für biophysikalische Chemie, 37077 Göttingen, Germany.

**11** Max-Planck-Institut für medizinische Forschung, 69120 Heidelberg, Germany

# Acknowledgements

- ▶ Rick Kirian, Nadia Zatsepin, Mark Hunter (ASU)
- ▶ Anton Barty (CFEL)
- ▶ Karol Nass, Francesco Stellato, Andy Aquila, Andrew Martin (CFEL)
- ▶ James Holton (LBL)
- ▶ ... many others ...

# X-Ray Sources

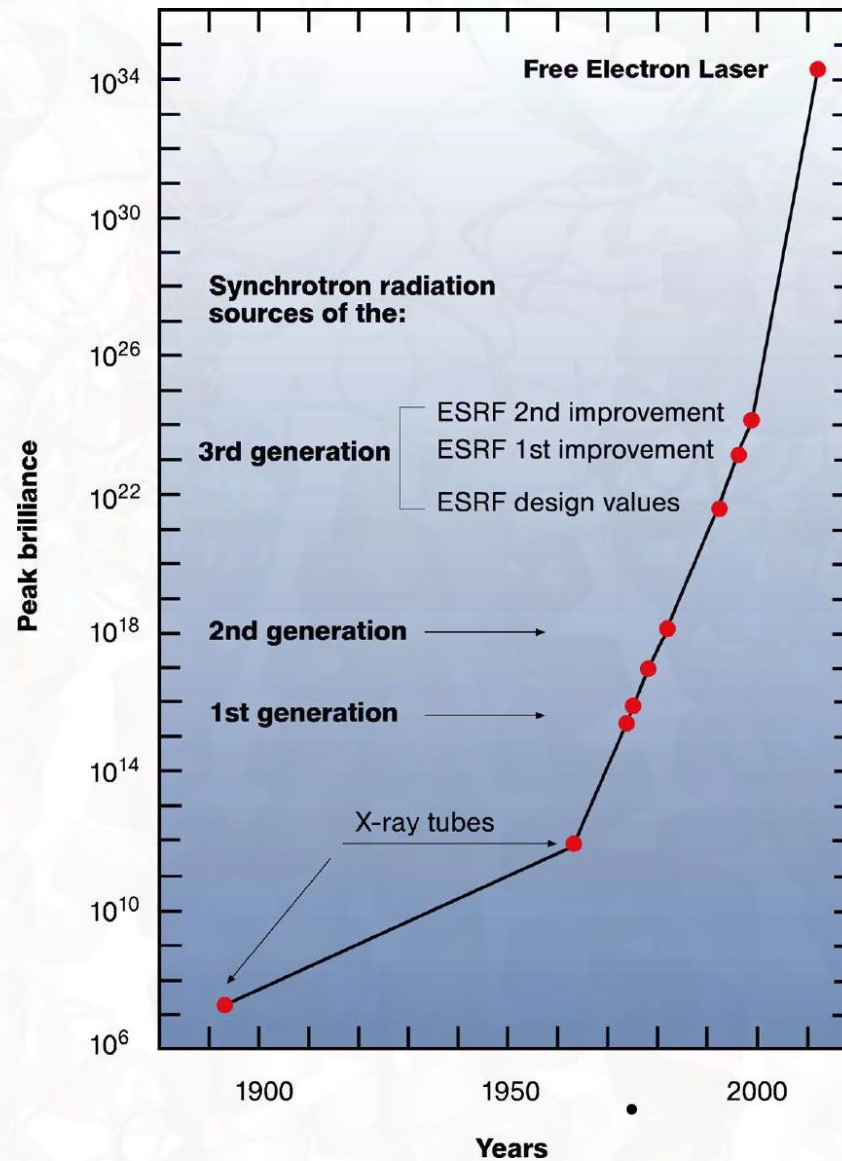


Image: European XFEL

# The Free-Electron Laser

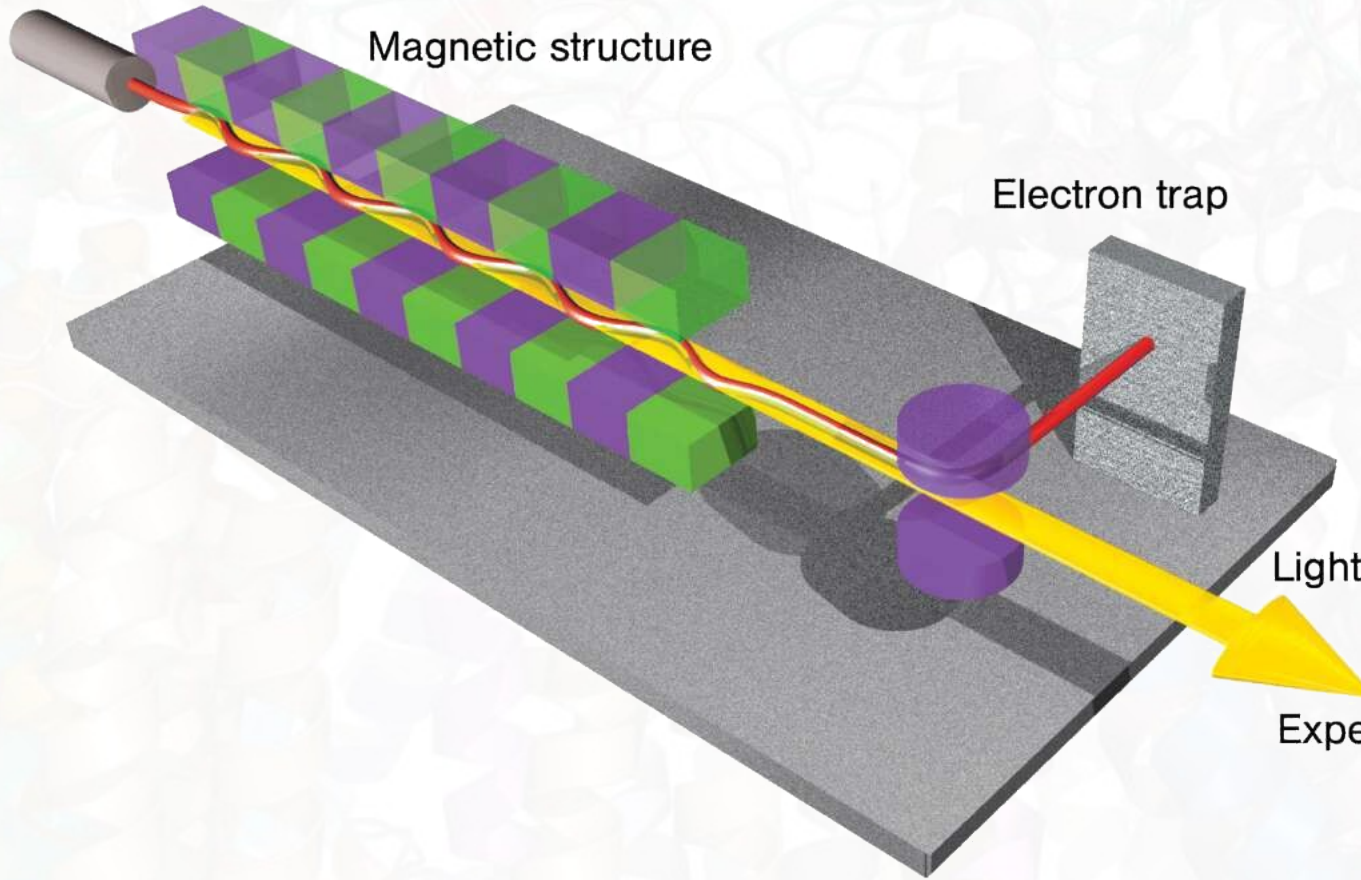
Electron source  
and accelerator

Magnetic structure

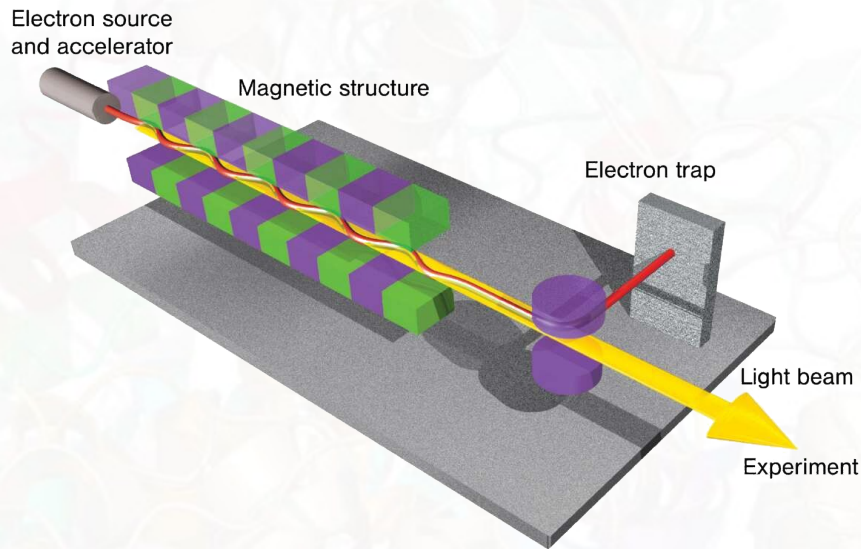
Electron trap

Light beam

Experiment

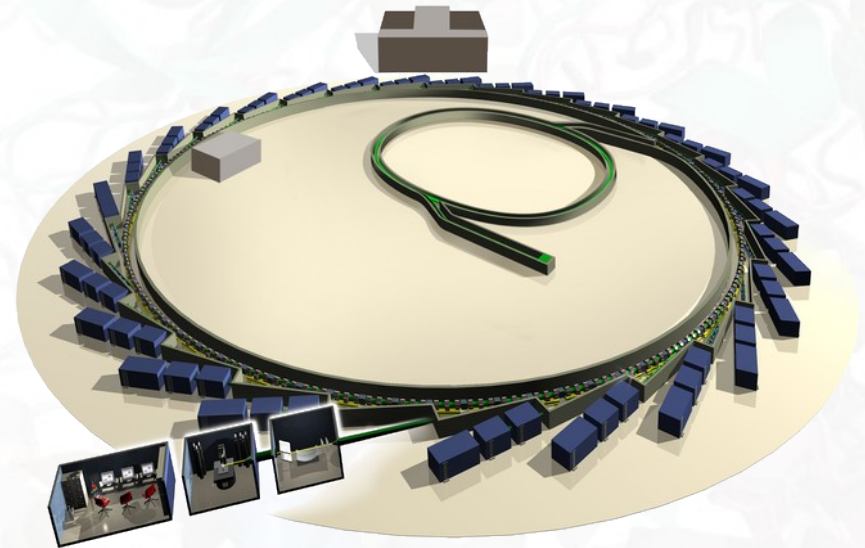


# The Free-Electron Laser



## X-ray Laser

Electrons pass through once.  
Extremely short pulses.  
Extremely high brightness.  
Few (1-3) experiments at a time.



## Synchrotron Light Source

Electrons circulate millions of times.  
Relatively long pulses.  
Relatively low brightness.  
Many experiments at a time.



# LCLS — Linac Coherent Light Source

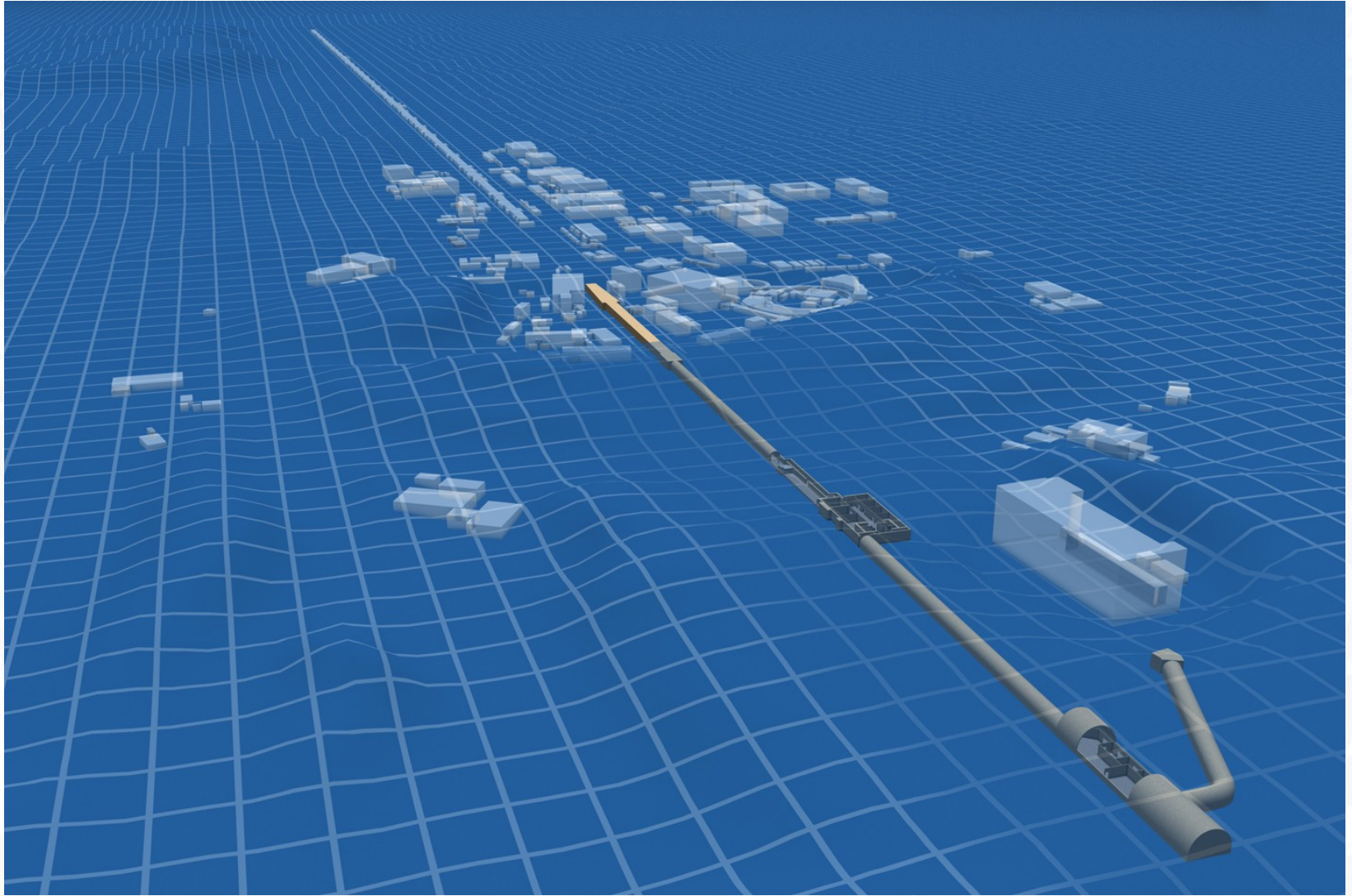


Image: SLAC National Accelerator Laboratory

# LCLS — Linac Coherent Light Source

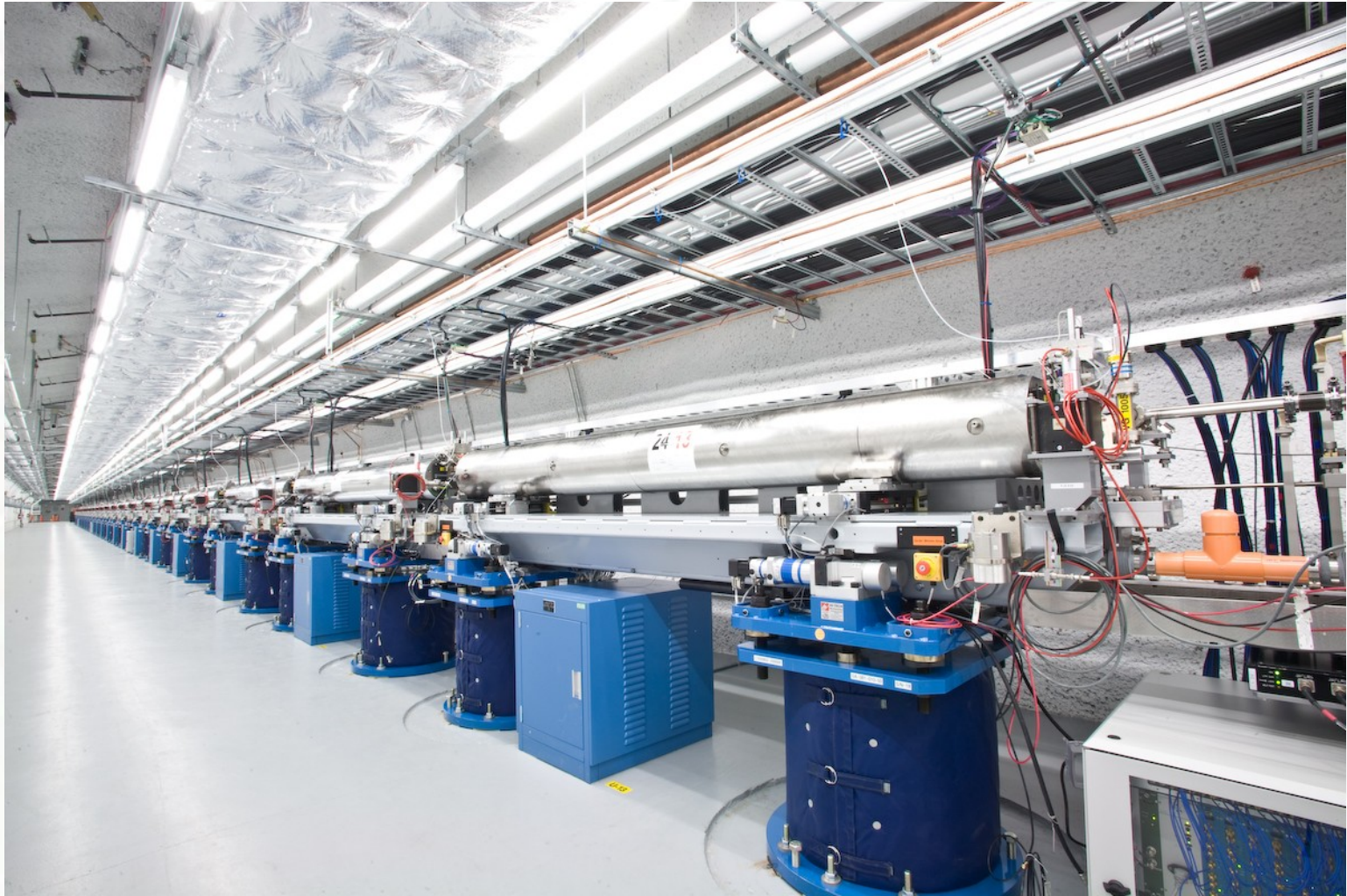


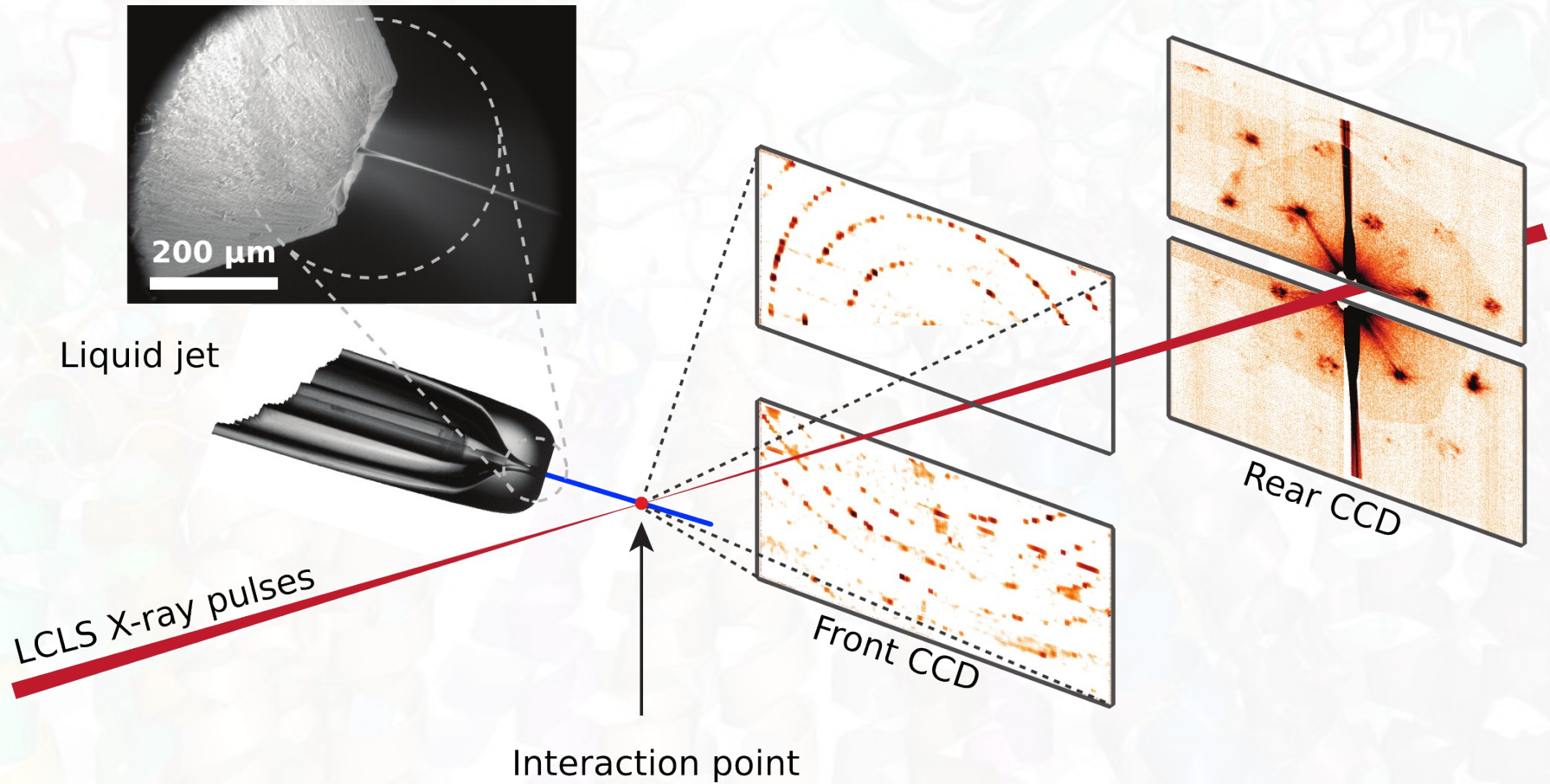
Image: SLAC National Accelerator Laboratory

# LCLS — Linac Coherent Light Source



Image: SLAC National Accelerator Laboratory

# Experimental Setup



# Big proteins...

- ▶ Transmembrane proteins control (amongst many other things) what may cross a membrane.

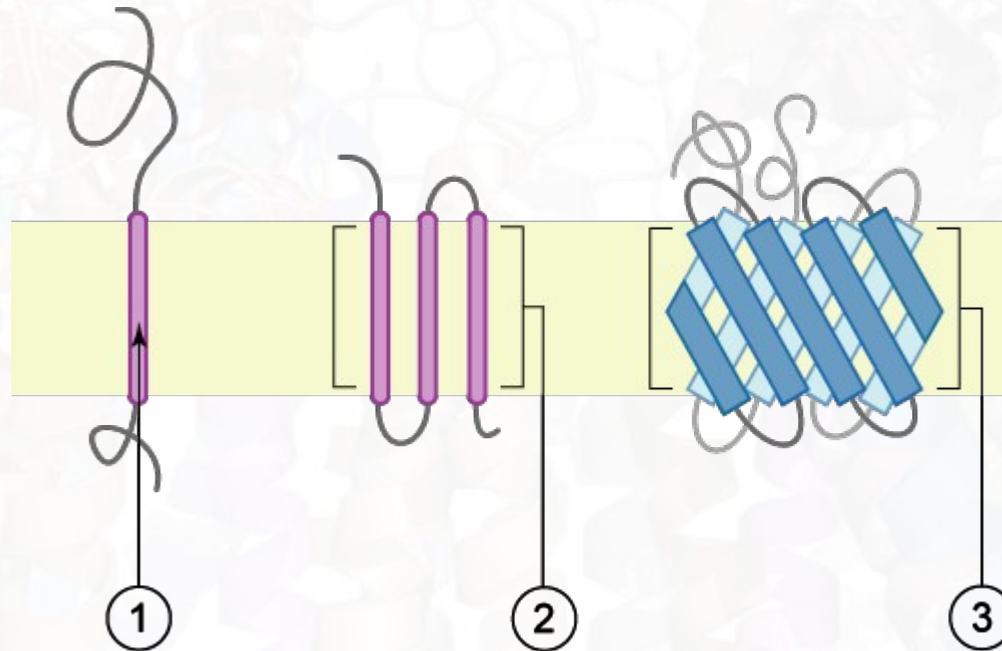


Image: [http://en.wikipedia.org/wiki/File:Polytopic\\_membrane\\_protein.png](http://en.wikipedia.org/wiki/File:Polytopic_membrane_protein.png) CC-BY-SA

# Characteristics of Data...

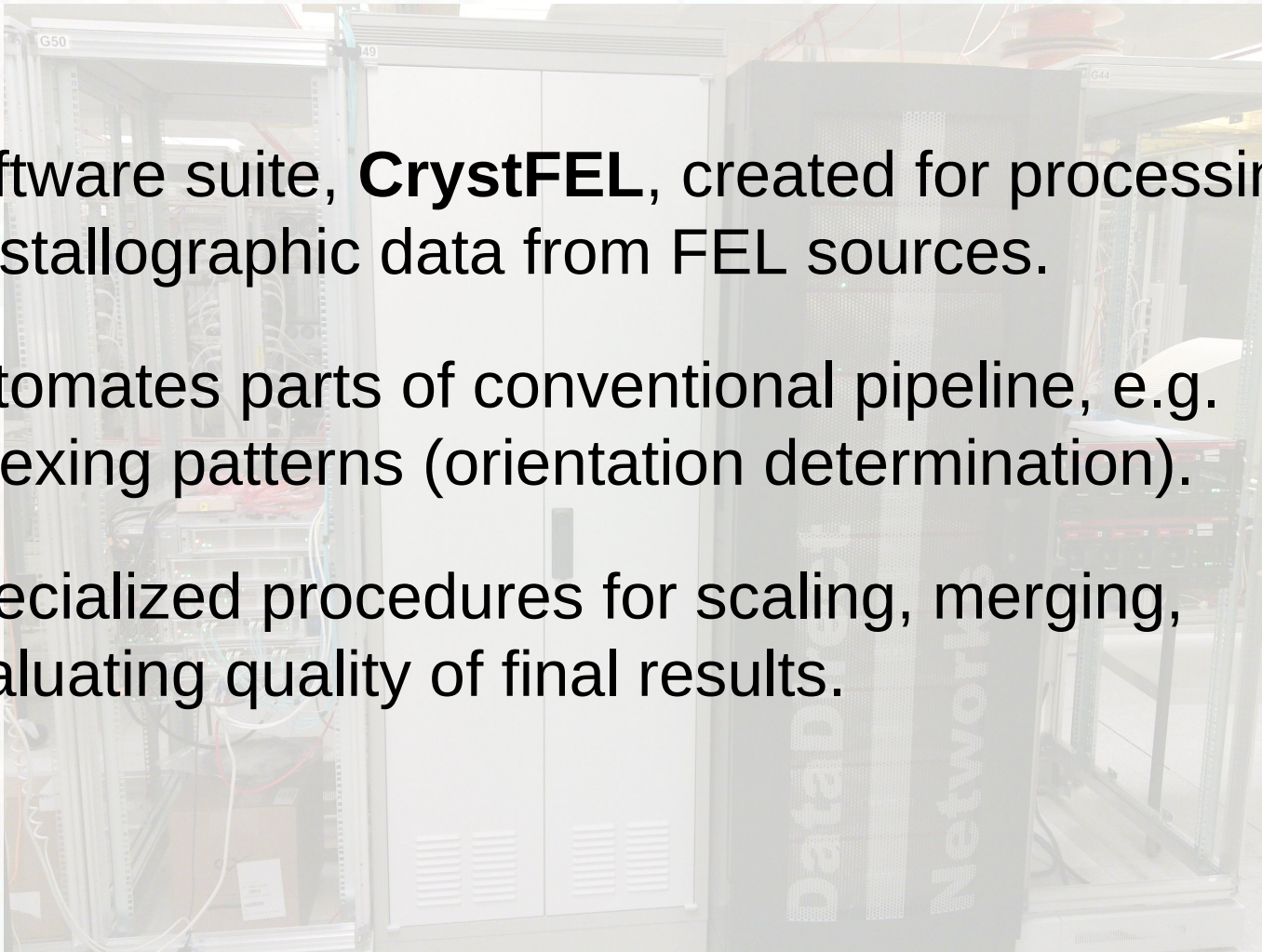
- ▶ **Lots** of patterns: millions of images, reduced to hundreds of thousands of patterns after “hit finding”.
- ▶ Patterns must be indexed individually.
- ▶ Large number of sets of intensity data to merge.

# Number crunching...



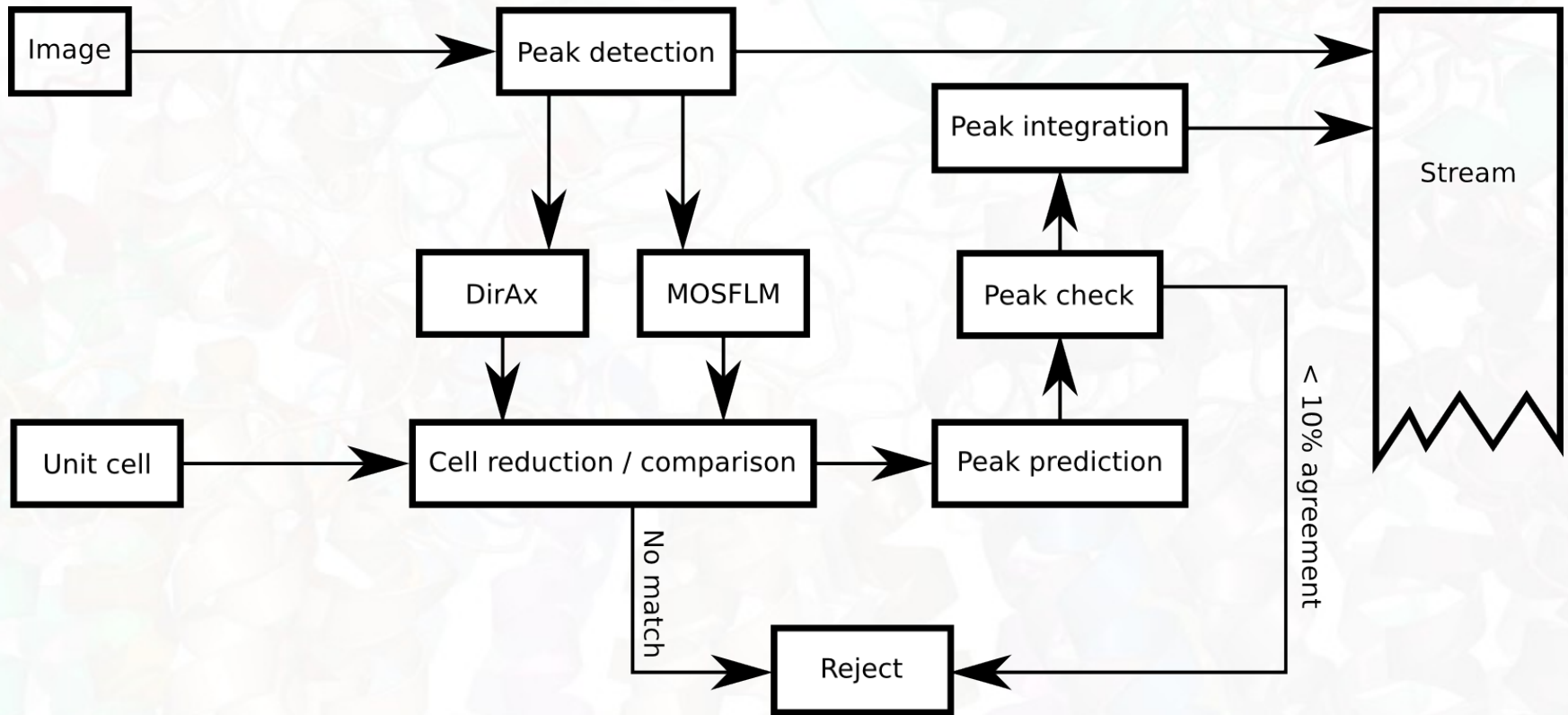
# Number crunching...

- ▶ Software suite, **CrystFEL**, created for processing crystallographic data from FEL sources.
- ▶ Automates parts of conventional pipeline, e.g. indexing patterns (orientation determination).
- ▶ Specialized procedures for scaling, merging, evaluating quality of final results.



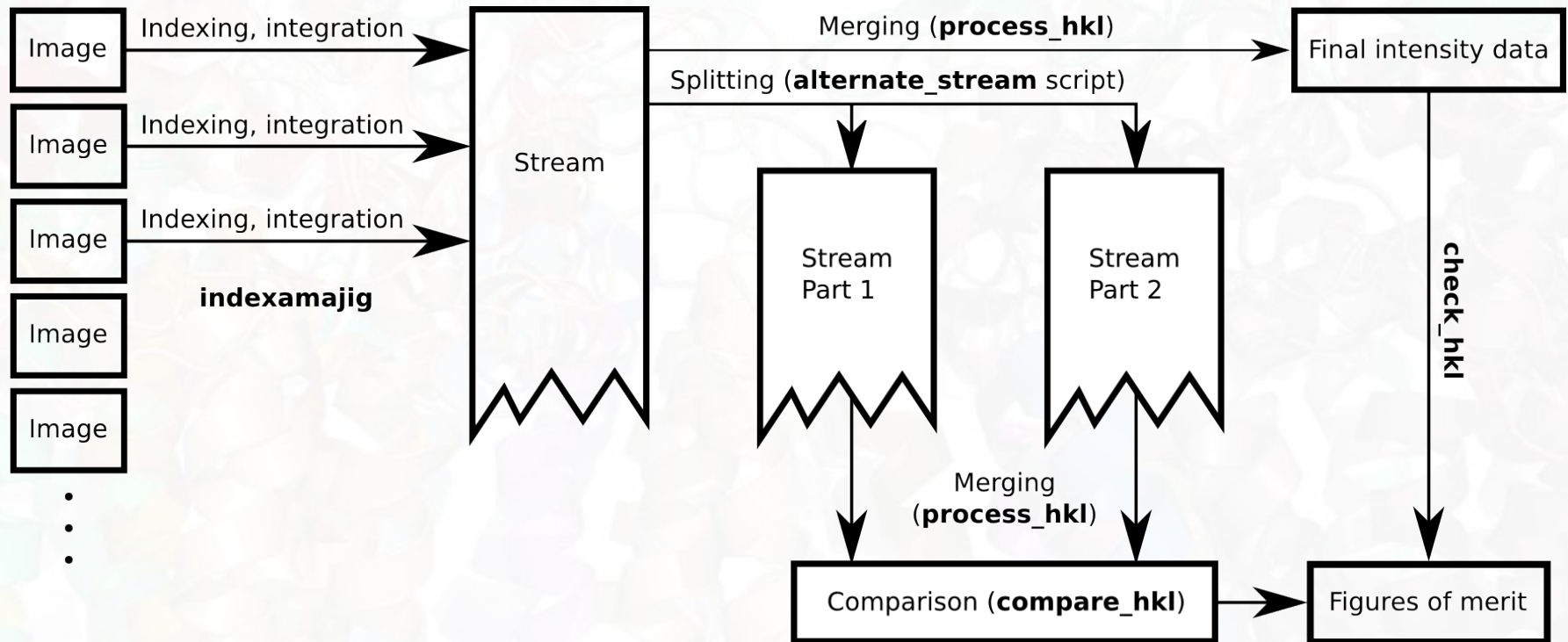


# Number crunching...

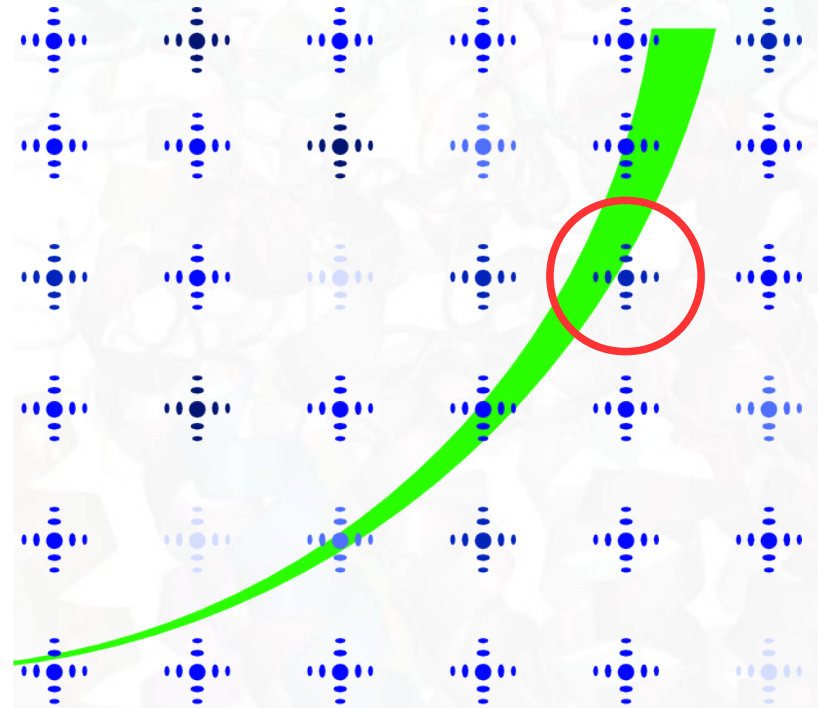
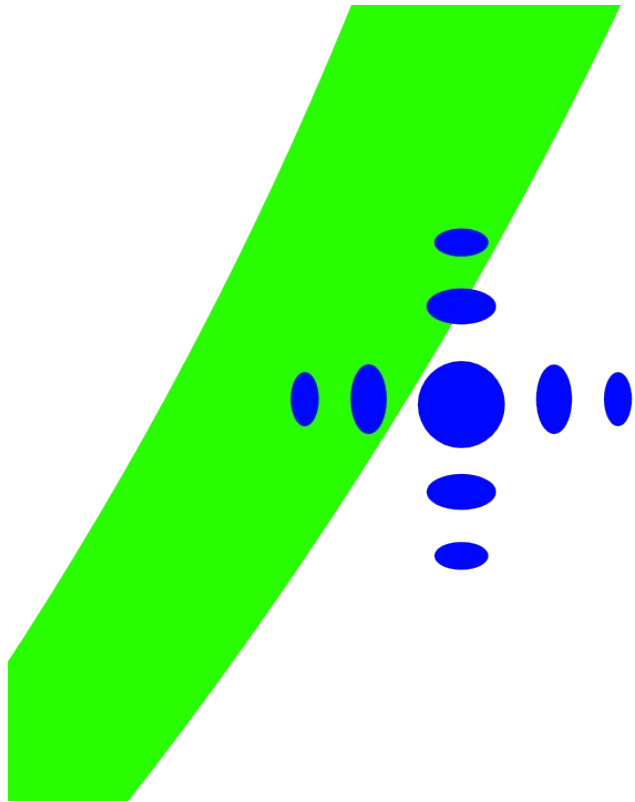


- ▶ If an image can't be processed at any stage, just skip over it. Define “yield” of processing pipeline and try to maximize it.

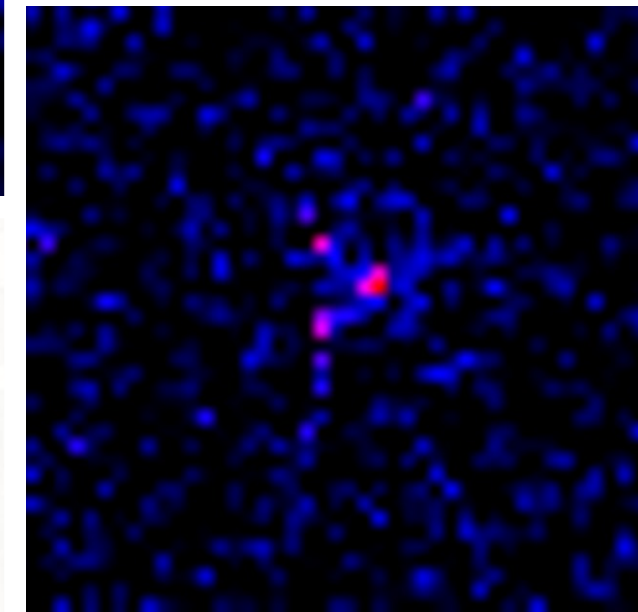
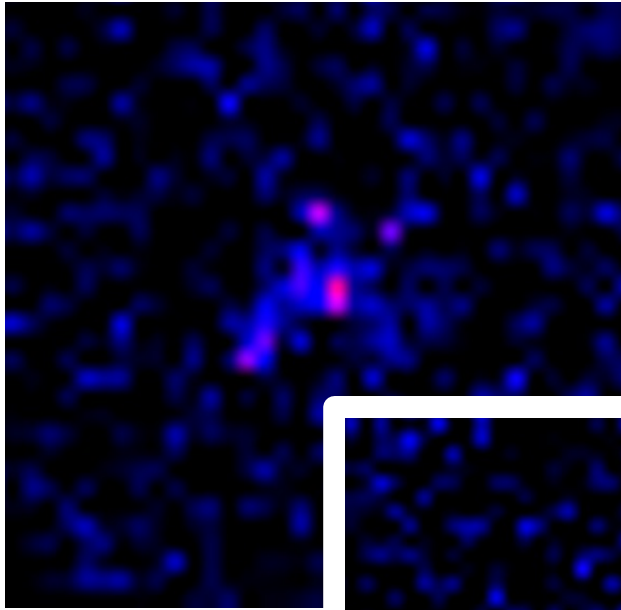
# Number crunching...



# Reciprocal Space



# Peak Shapes - Front Detector



Reflection is partially recorded

# The Rotation Method

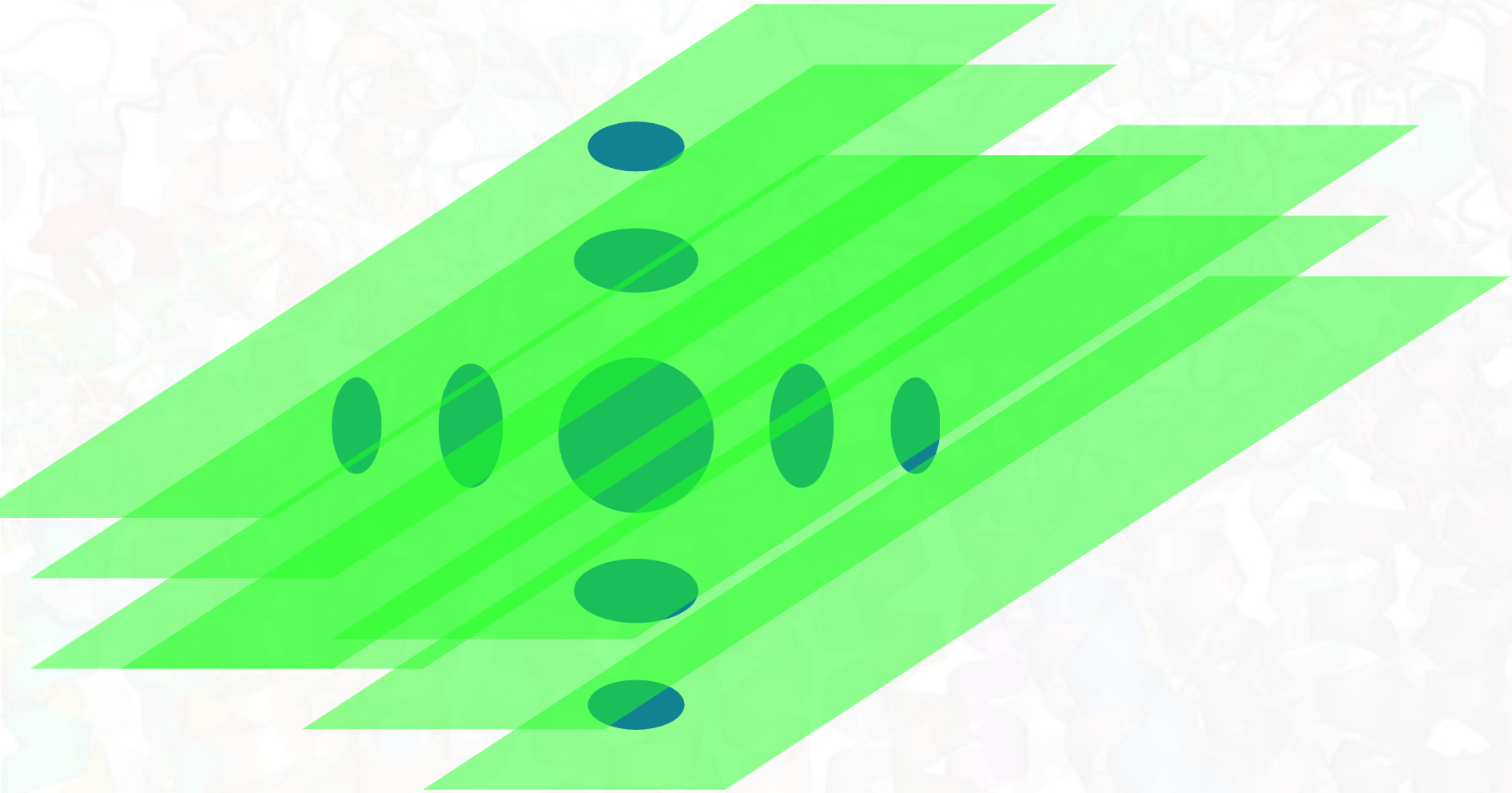
Exposure time: 100 fs

Rotation step: 0.5 degrees

Required rotation speed:

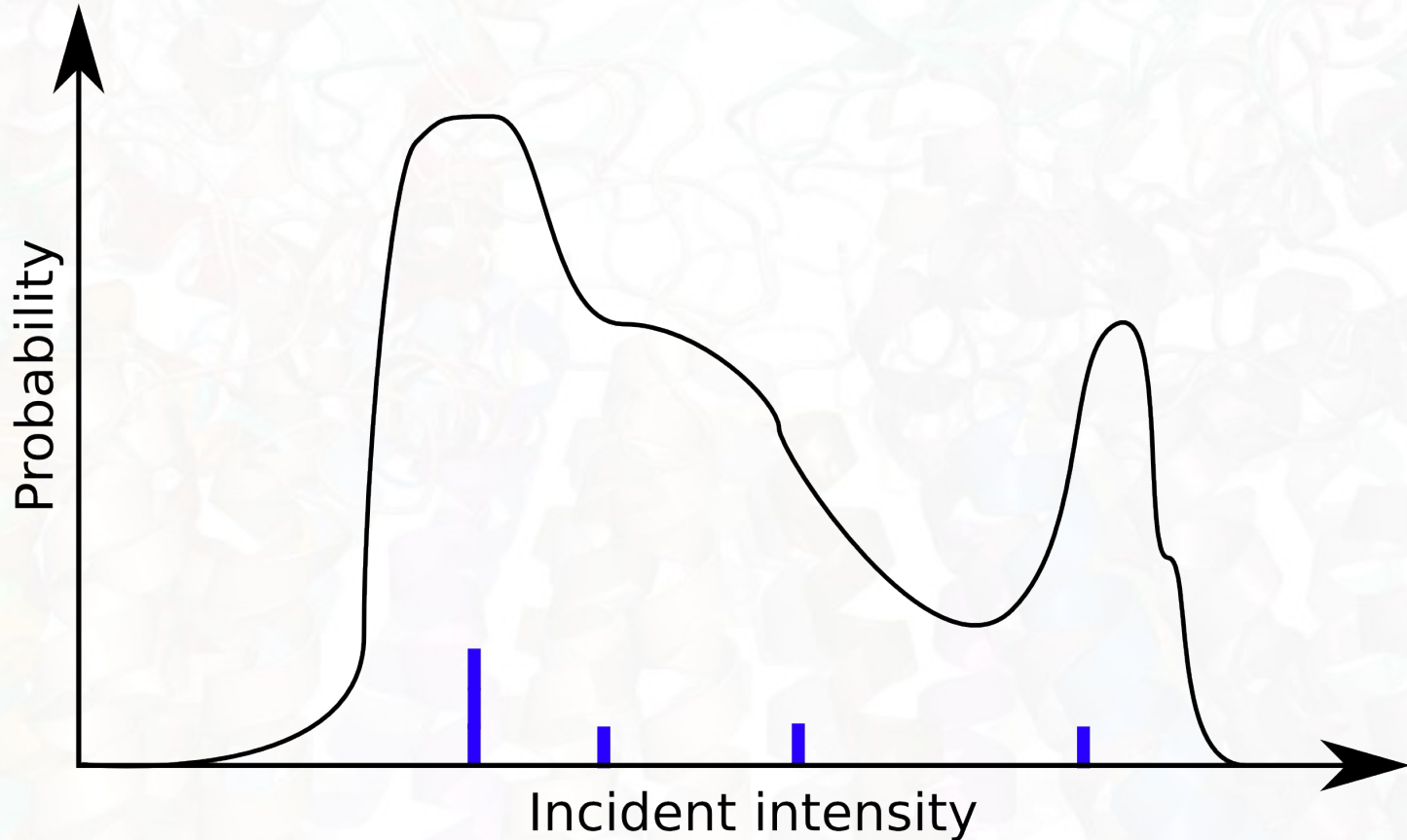
**13.9 billion revolutions  
per second**

# The “Monte Carlo” Method



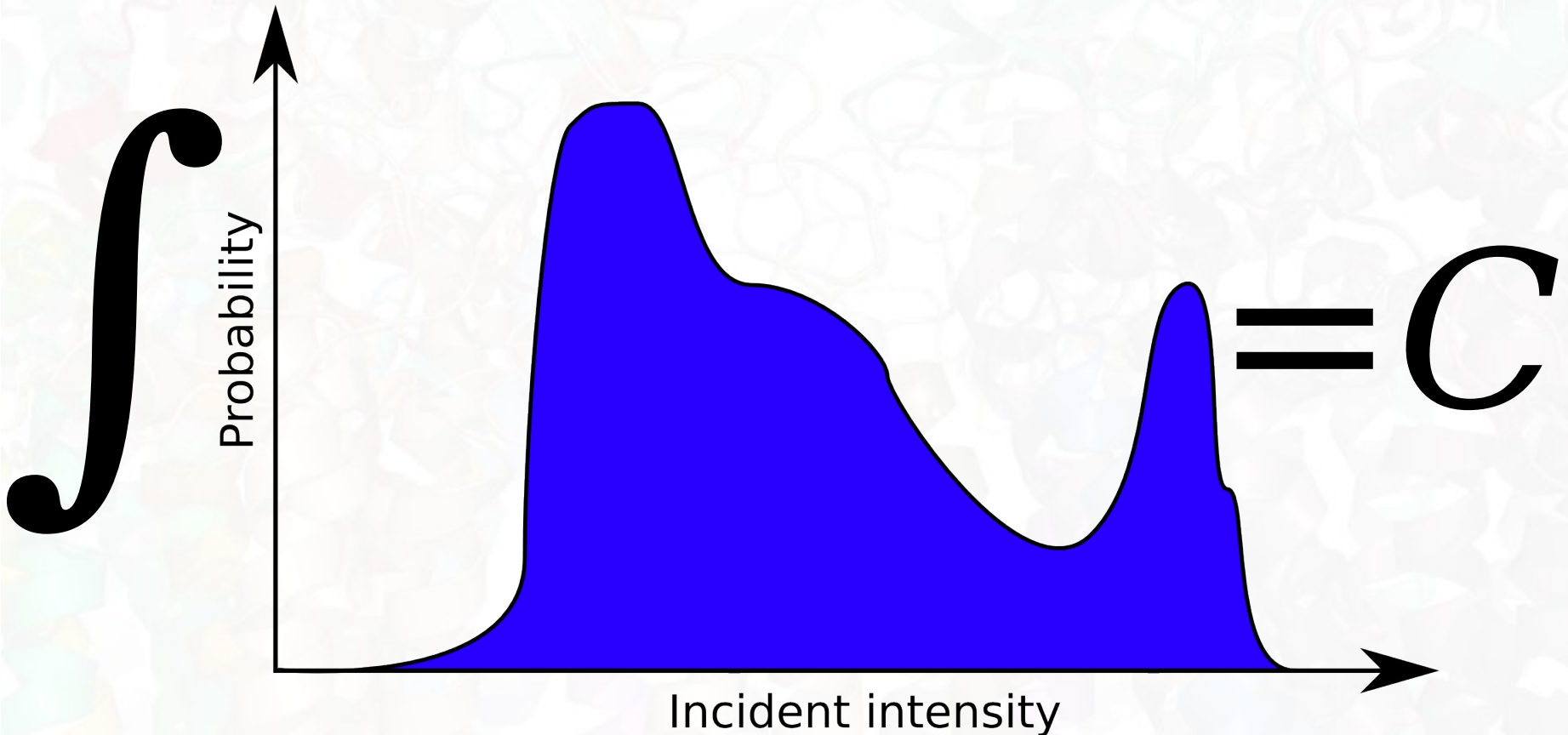
The sum of a large number of samples converges to the integrated intensity.

# The “Monte Carlo” Method



The sum of a large number of samples converges to **a constant**.

# The “Monte Carlo” Method



The sum of a large number of samples converges to **a constant.**



# The “Monte Carlo” Method

- ▶ The volume of the reciprocal lattice point.
- ▶ Crystal size (and shape).
- ▶ Incident intensity (SASE fluctuations).
- ▶ “Targeting accuracy”.
- ▶ Beam fluctuations (different bandwidth each shot).
- ▶ Crystal quality – mosaicity?

... all of the above are “integrated out” to constants.

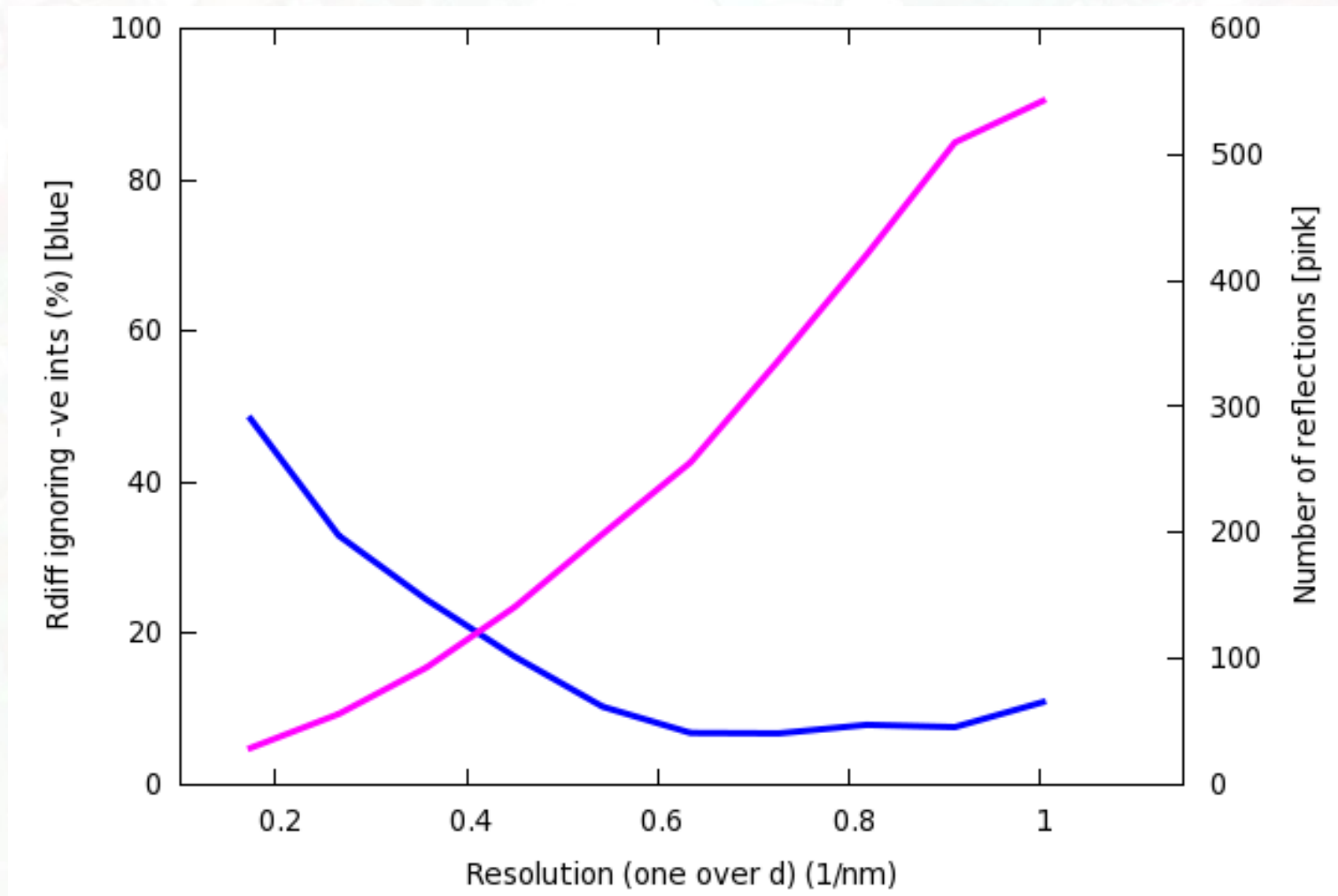
# The “Monte Carlo” Method

This **does actually work**, both in simulation and experiment:

Kirian et al., Optics Express **18** (2010) p5713-5723

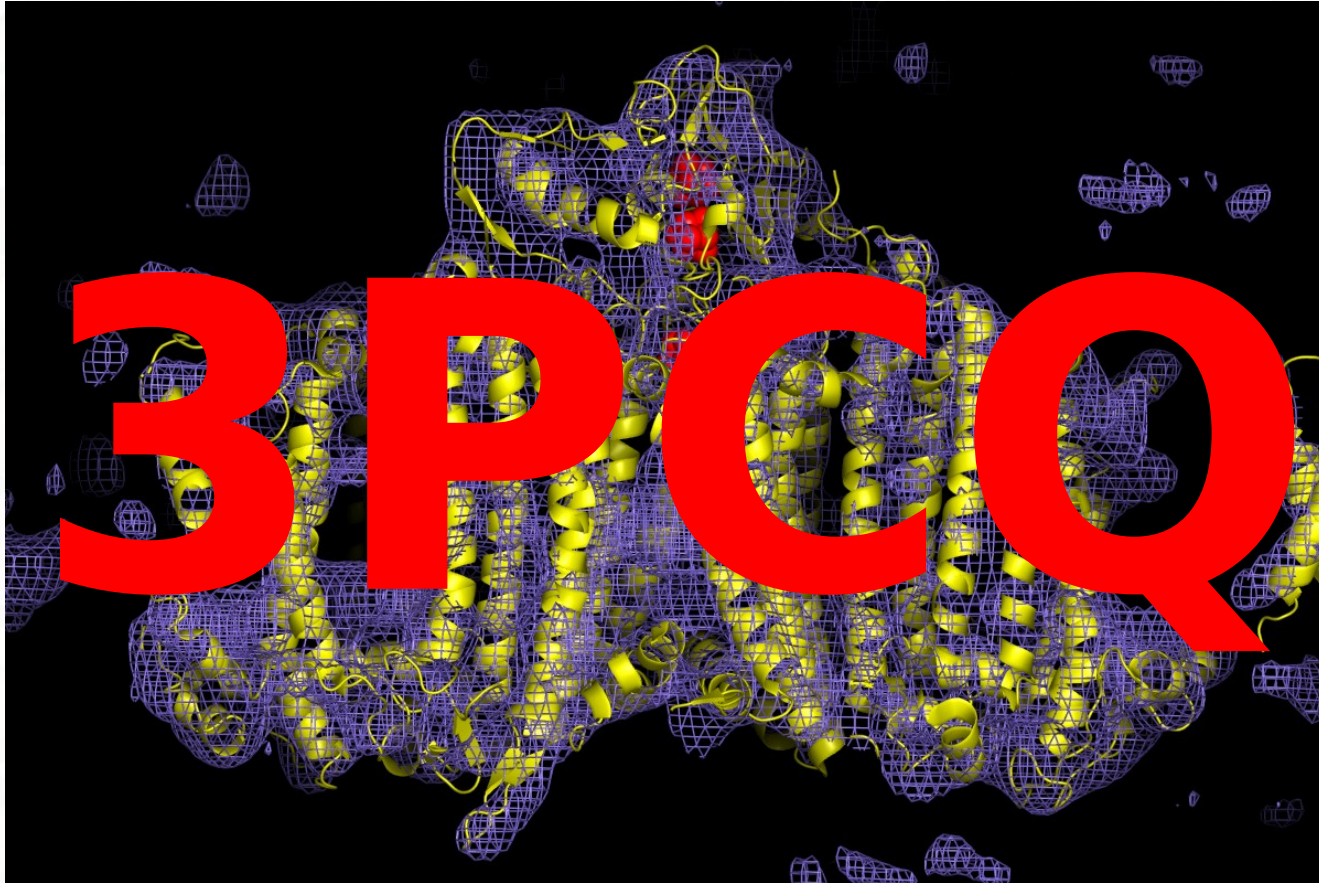
But you need a lot of patterns...

# The “Monte Carlo” Method



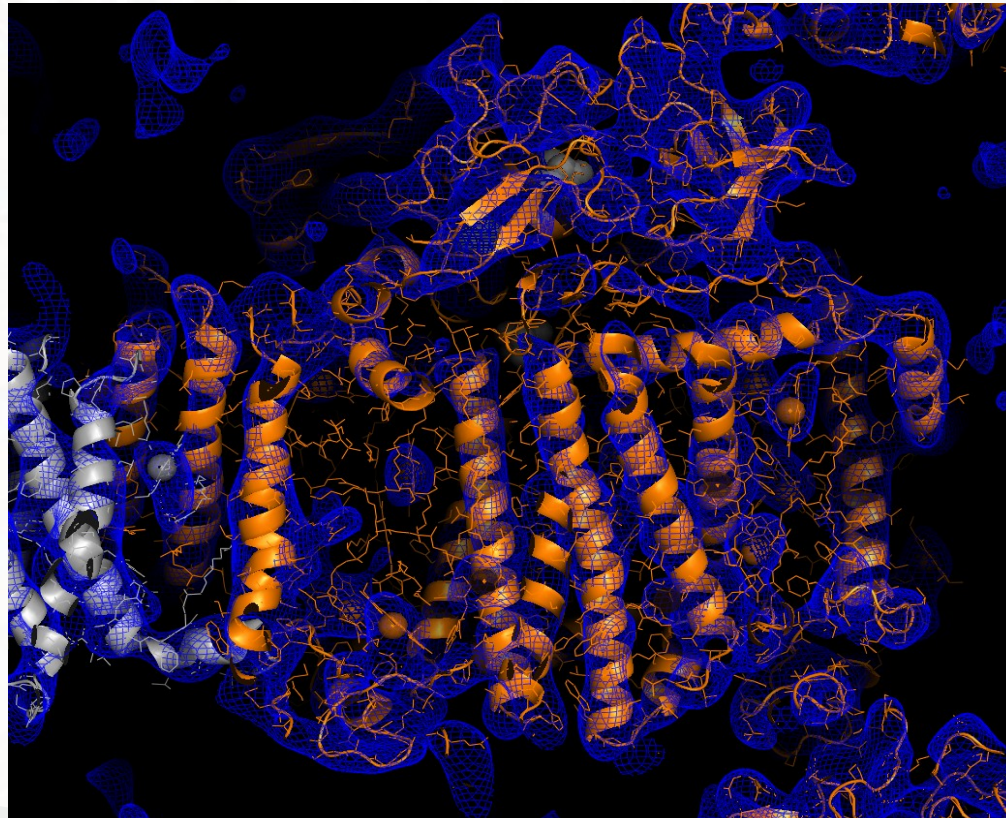
# Phasing the Data

- ▶ At this point, our data goes into the conventional MX analysis pipeline.

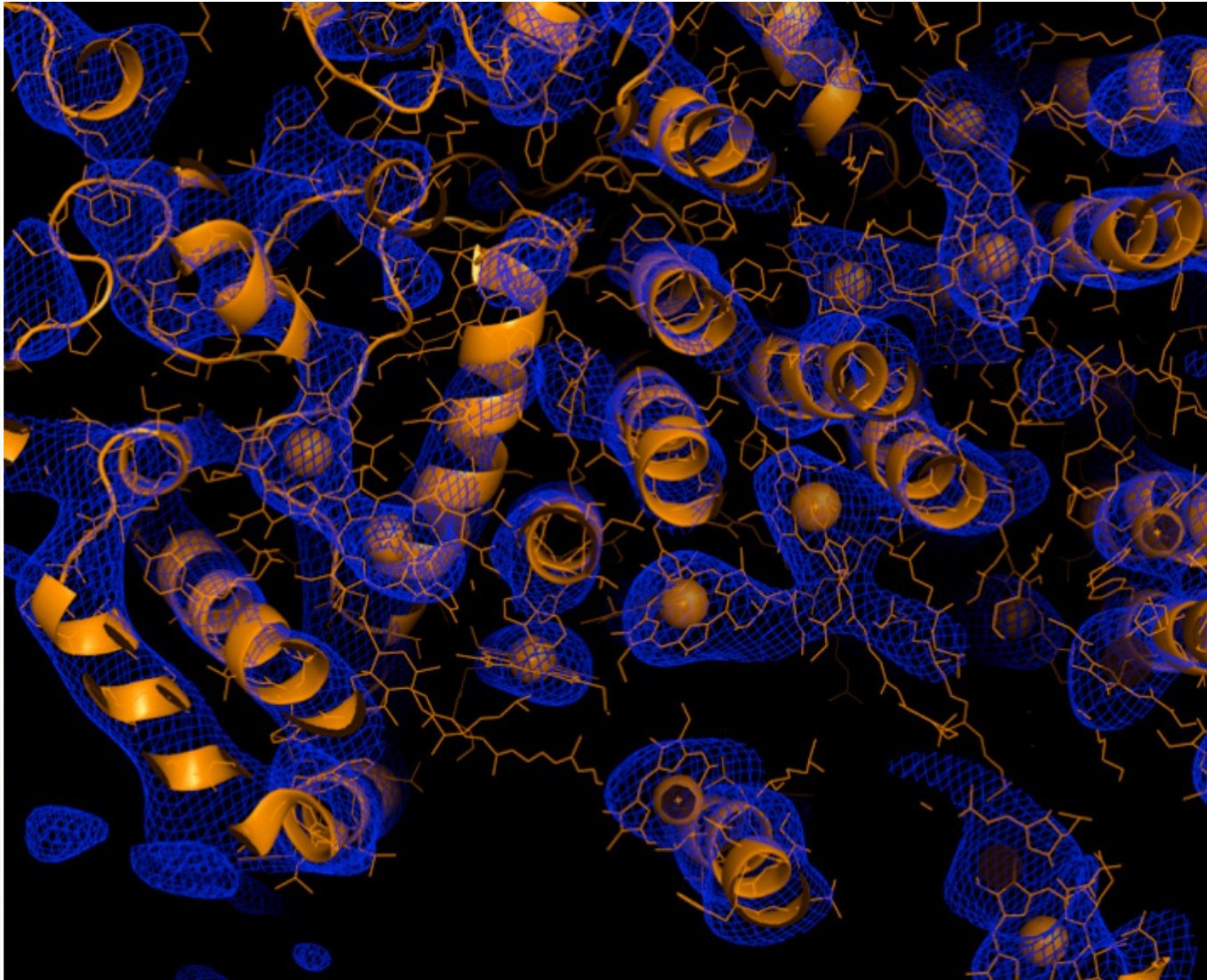


# PS1 at higher resolution (June 2010)

- ▶ Resolution increased from 8.5Å to 7.4Å due to slightly higher photon energy.
- ▶ Improved data processing and detector behaviour.

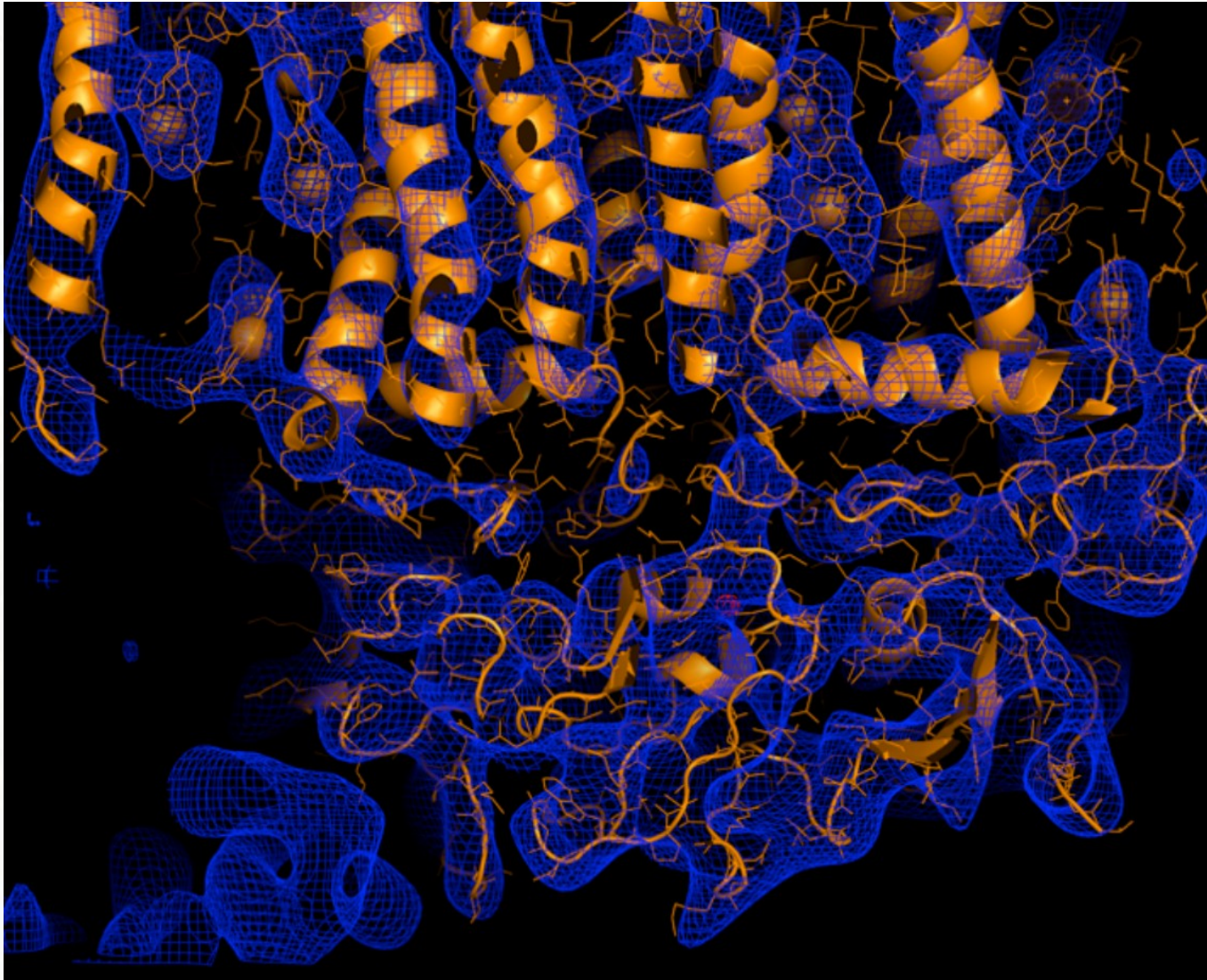


# PS1 at higher resolution (June 2010)



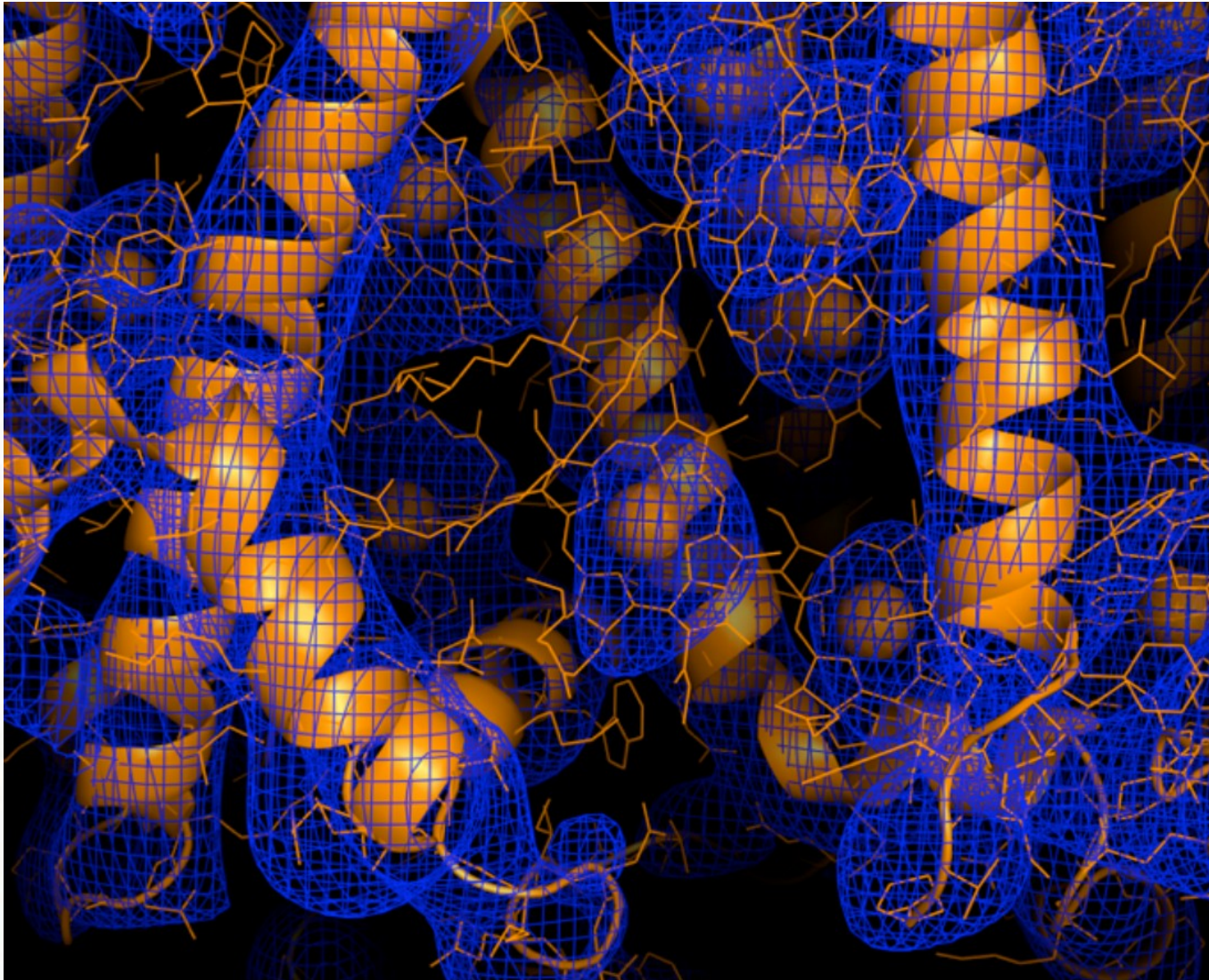
DEN refinement and graphics by Axel Brunger

# PS1 at higher resolution (June 2010)



DEN refinement and graphics by Axel Brunger

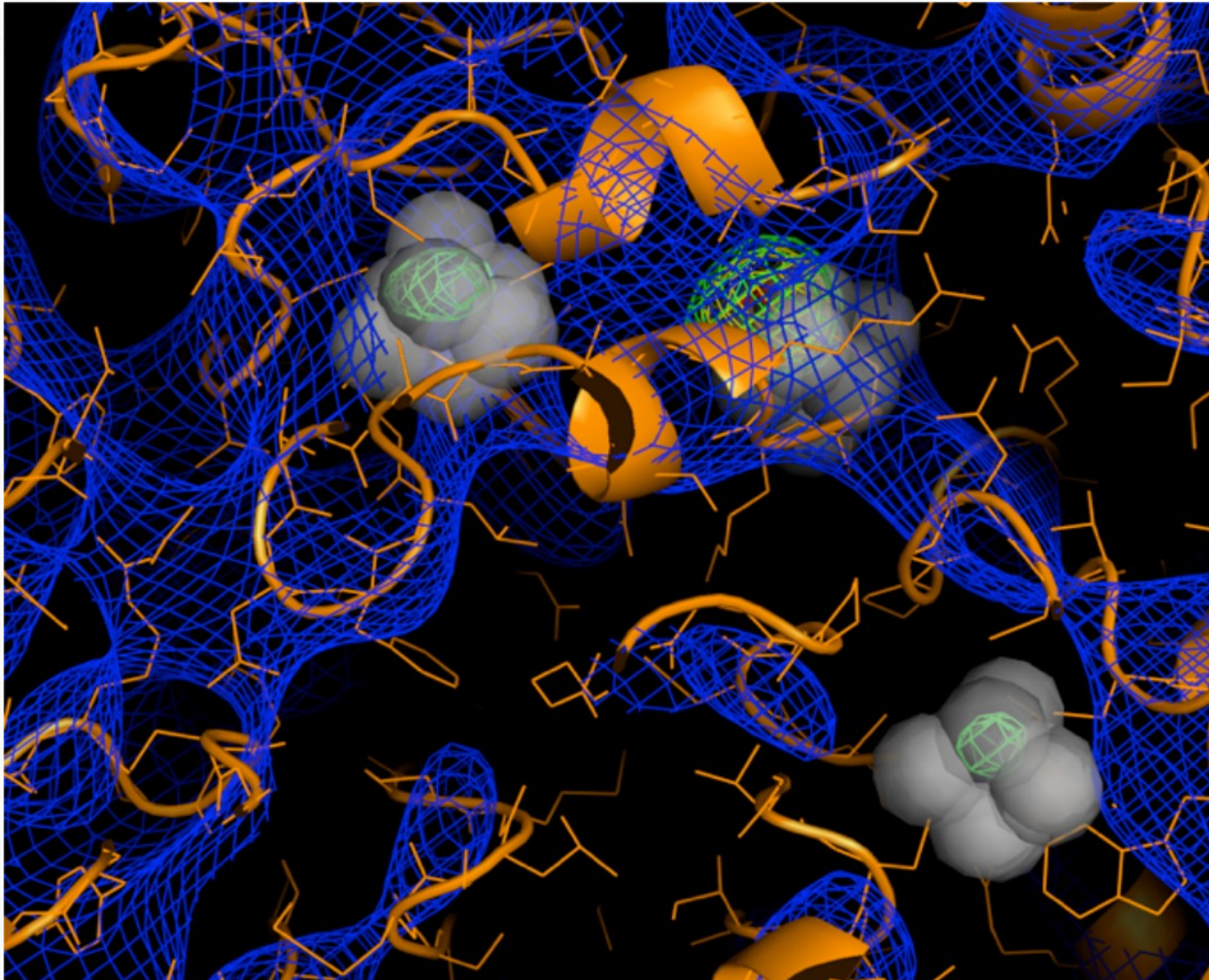
# PS1 at higher resolution (June 2010)



DEN refinement and graphics by Axel Brunger



# PS1 at higher resolution (June 2010)

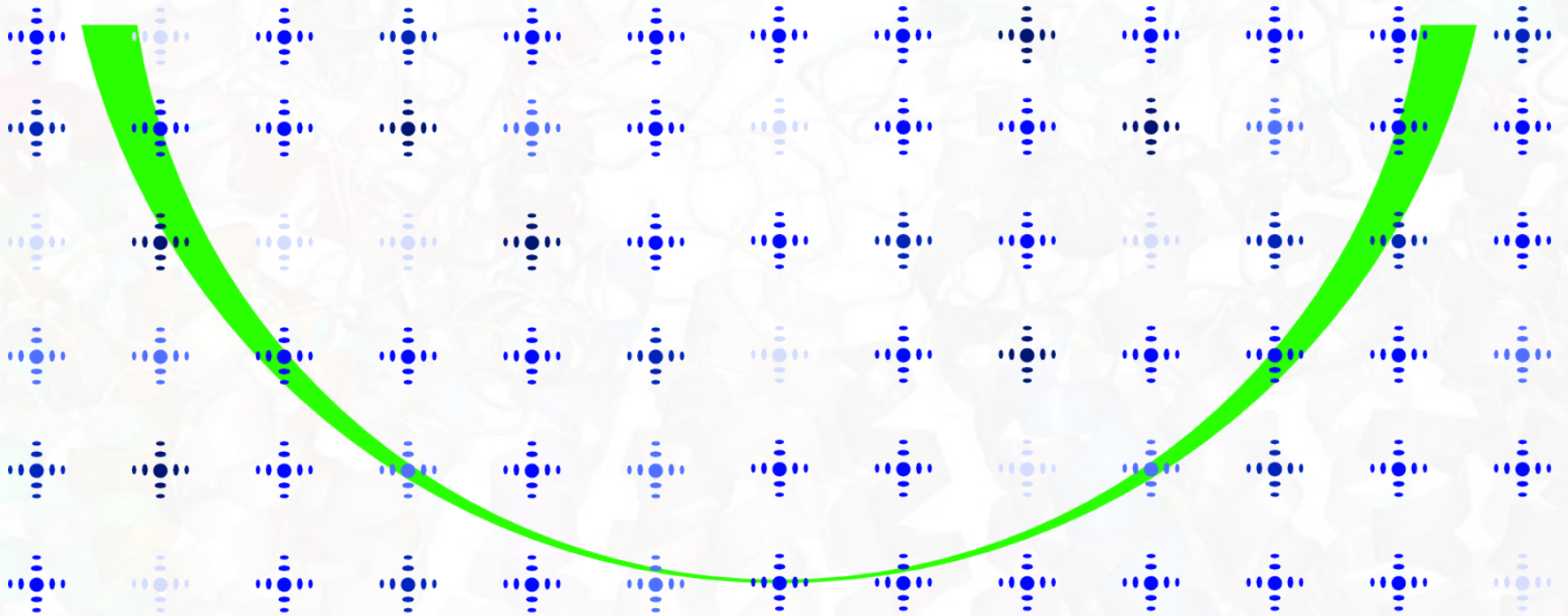


DEN refinement and graphics by Axel Brunger

# Beyond Monte Carlo: Scaling of Partial

- ▶ Fox and Holmes (1966) method:  
Set up full non-linear least squares for the scale factors and solve to get the best ones.
- ▶ Computational time: **number of patterns squared.**
  
- ▶ Kabsch (2010) method:  
Combine everything with equal scale factors, then scale everything to the result, and repeat.
- ▶ Computational time: **number of patterns.**

# Beyond Monte Carlo: Scaling of Partials



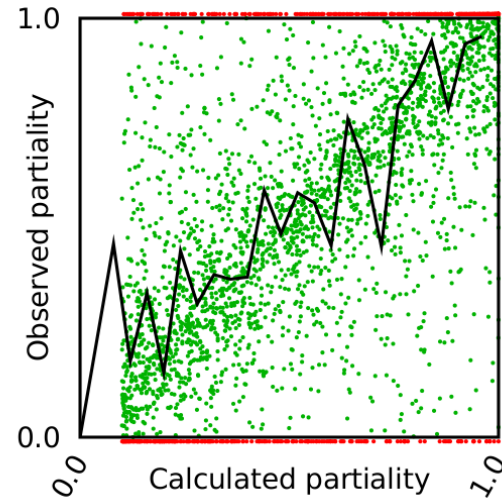
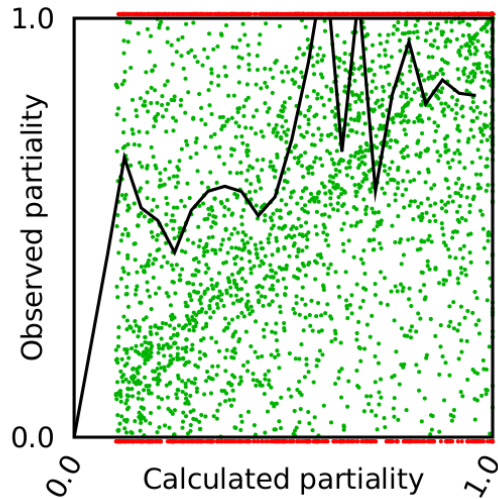
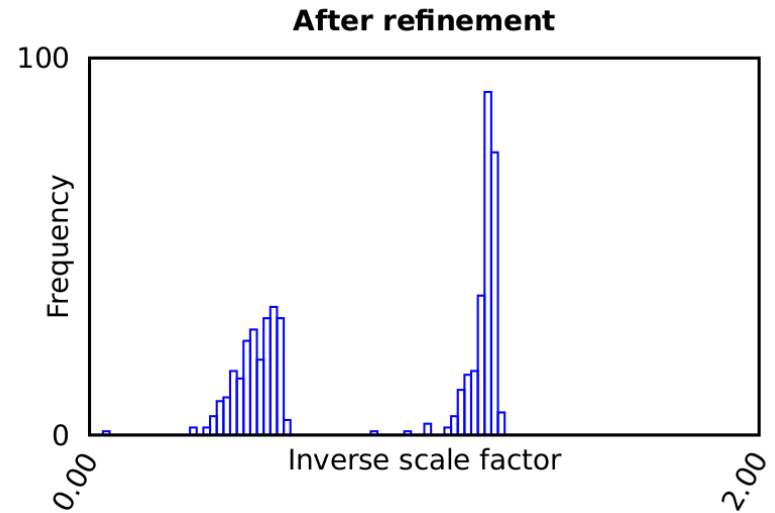
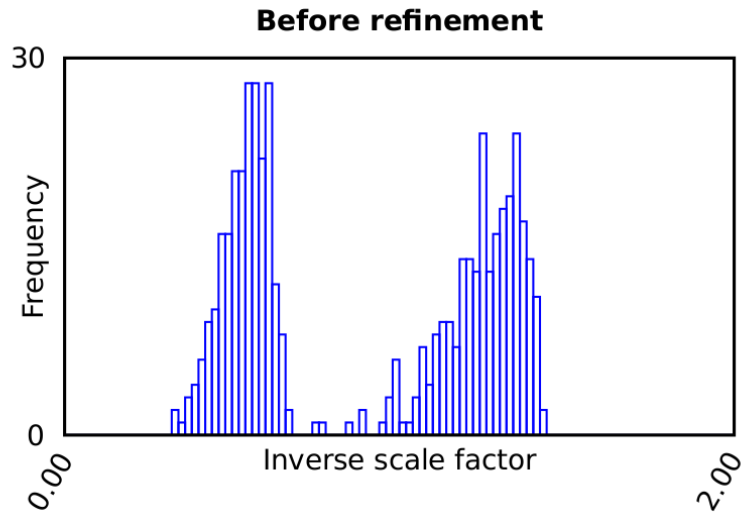
# Beyond Monte Carlo: Scaling of Partial

- ▶ Least squares target for post refinement:

$$E = \sum_h w_h \Delta I_h^2$$
$$= \sum_h w_h \left[ I_{h,\text{partial}} - p_h(a_0, a_1, a_2, \dots, a_m) I_{h,\text{full}} \right]^2.$$

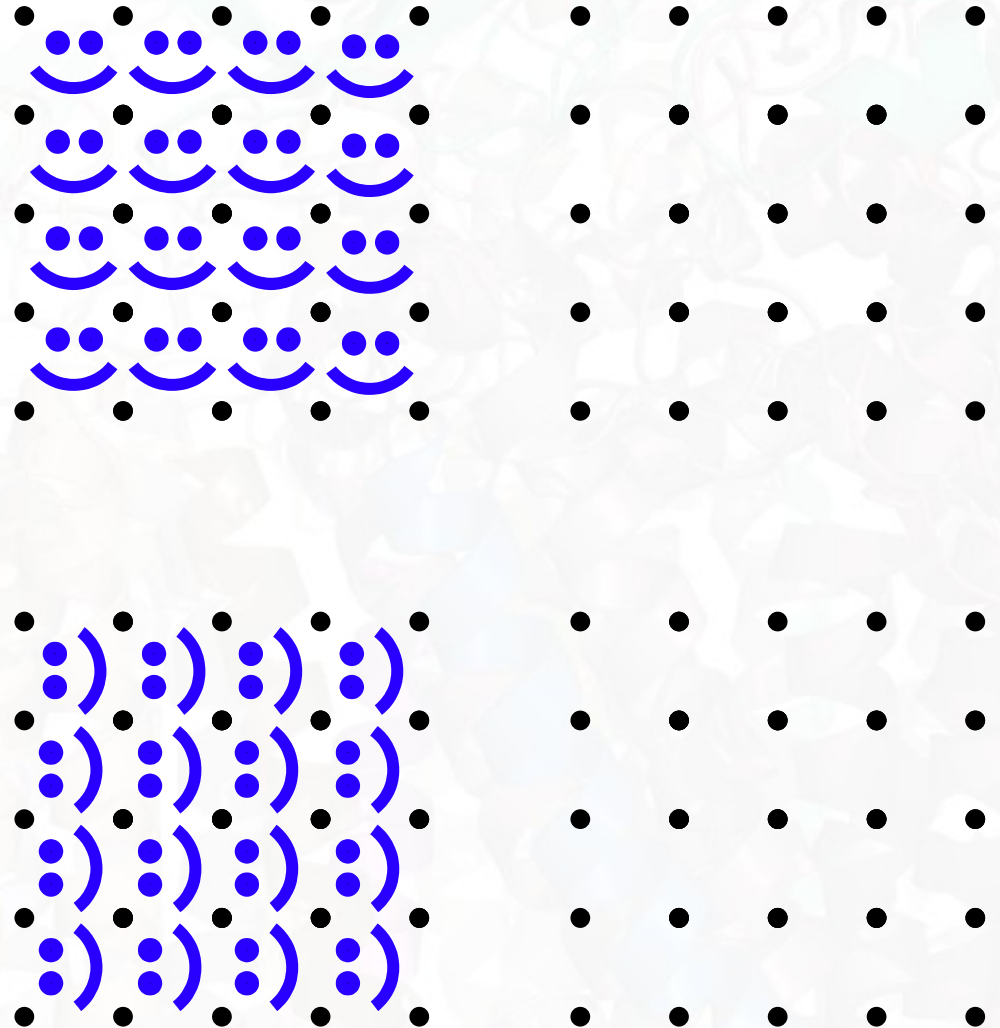
- ▶ Optimise the fit between this pattern and the others, by varying the parameters which affect partiality:
- ▶ Orientation, bandwidth, ...

# Beyond Monte Carlo: Scaling of Partialials



# Indexing Ambiguities

Some crystal lattices have geometry which looks the same in two or more different orientations, despite the actual crystal not having this symmetry.



# Beyond Monte Carlo: Scaling of Partialials

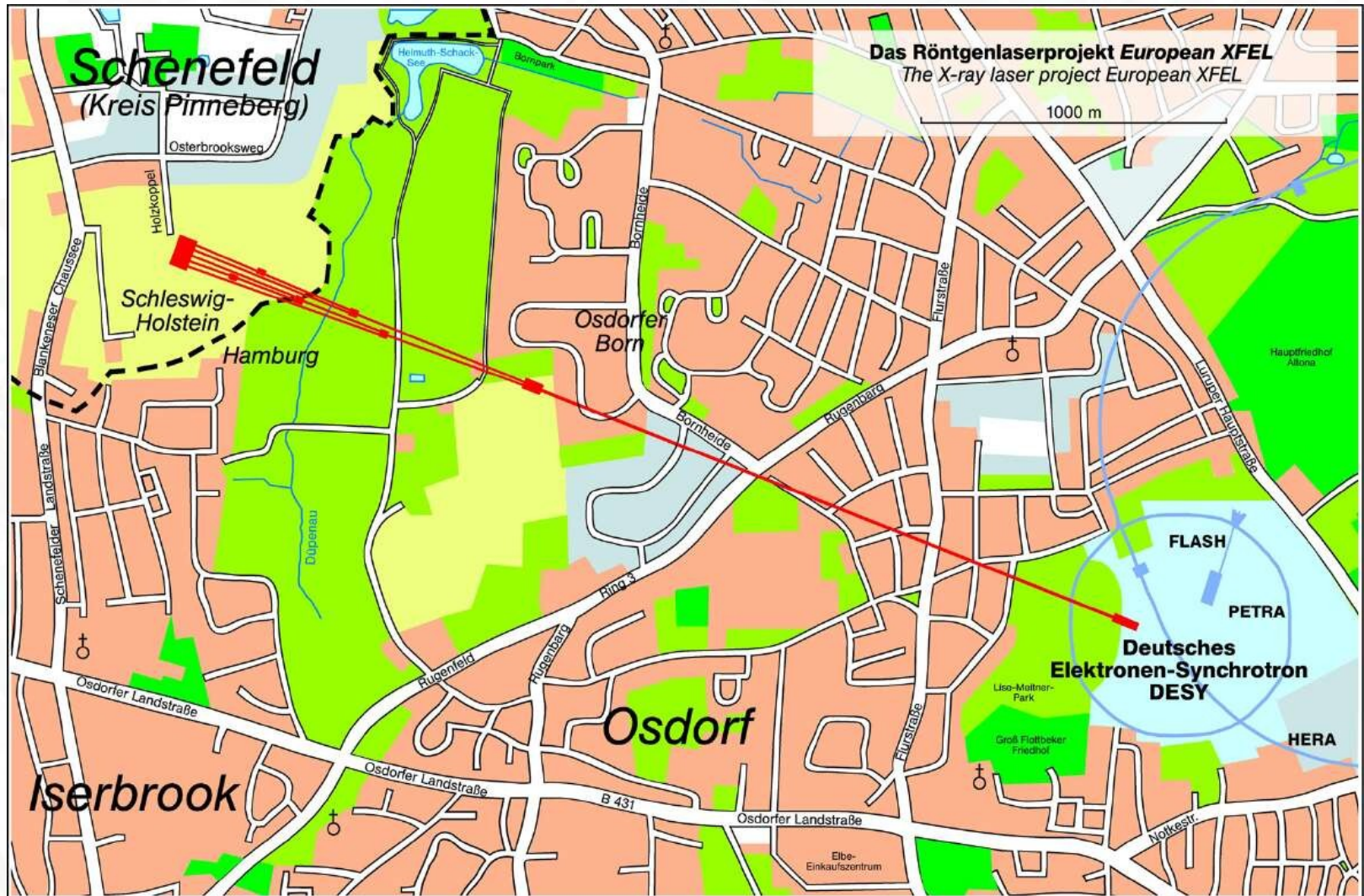
- ▶ Our situation is a mixture model: each measurement comes from one of two (or more) possible distributions, one for each of the “twin partners”.
- ▶ A mixture model is a classic example of the use of the expectation-maximisation (EM) algorithm.
- ▶ Simply put, we start with random twin assignments, refine the scale factors, then maximise the fit by varying the assignments.

# Scaling and Post Refinement

- ▶ First, **scale** the intensities with partiality estimates (from the **initial geometry estimate**).
- ▶ For each pattern, **refine** the geometry, wavelength, bandwidth etc to get the best match with the model.
- ▶ **Re-scale** the intensities.
- ▶ **Maximise the fit** by varying the twin assignments, using the current estimate as a lower bound.
- ▶ **Repeat** until converged.



# The European XFEL



# Sounds Interesting?

Other positions  
available besides  
this one!

Accelerators | Photon Science | Particle Physics

Deutsches Elektronen-Synchrotron  
A Research Centre of the Helmholtz Association



## X-RAY FEL CRYSTALLOGRAPHY .

**DESY, Hamburg location, is seeking:  
Senior Scientist or Postdoc (m/f)**

### DESY

DESY is one of the world's leading centres for the investigation of the structure of matter. DESY develops, runs and uses accelerators and detectors for photon science and particle physics.

The Center for Free-Electron Laser Science (CFEL) is a jointly operated research cooperation between DESY, Max Planck Society and the University of Hamburg. DESY supports three CFEL divisions with leaders jointly appointed with the University of Hamburg. For a joint EU Project with the European XFEL, the Coherent Imaging Division offers the following position. The European X-Ray Free Electron Laser Facility (XFEL) is a multi-national non-profit company. It will make available X-rays of unique quality for studies in physics, chemistry, life sciences, materials research and others. Located in the Hamburg area, Germany, it will comprise scientific instruments for a wide range of

Main publication

Chapman et al., Nature **470** (2011) p74.

Injector

DePonte et al., J. Phys. D **41**, (2008) 195505

CAMP instrument

Strüder et al., Nuclear Instruments and Methods in  
Physics Research A **614** (2010) p483.

Monte Carlo integration

Kirian et al., Optics Express **18** (2010) p5713.

Phasing by shape transforms

Spence et al., Optics Express **19** (2011) p2866.

PDB entry 3PCQ