# Analysis pipeline for crystallographic data from FELs

Thomas A. White

19th June 2012

taw@physics.org

# Acknowledgements (first experiments)

Henry N. Chapman, Petra Fromme, Anton Barty, Richard A. Kirian, Andrew Aquila,
Mark S. Hunter, Joachim Schulz, Daniel P. DePonte, Uwe Weierstall,
R. Bruce Doak, Filipe R. N. C. Maia, Andrew V. Martin. Ilme Schlichting,
Lukas Lomb, Nicol Coppola, Robert L. Shoeman, Sascha W. Epp,
Robert Hartmann, Danial Rolles, Artem Rudenko, Lutz Foucar, Nils Kimmel,
Georg Weidenspointner, Peter Holl, Mengning Liang, Miriam Barthelmess,
Carl Caleman, Sebastien Boutet, Michael J. Bogan, Jacek Krzywinski,
Christoph Bostedt, Sasa Bajt, Lars Gumprecht, Benedikt Rudek, Benjamin Erk,
Carlo Schmidt, Andre Homke, Christian Reich, Daniel Pietschner, Lothar Struder,
Gunter Hauser, Hubert Gorke, Joachim Ullrich, Sven Herrmann, Gerhard Schaller,
Florian Schopper, Heike Soltau, Kau-Uwe Kuhnel, Marc Messerschmidt,
John D. Bozek, Stefan P. Hau-Riege, Matthias Frank, Christina Y. Hampton,
Raymond Sierra, Dmitri Starodub, Garth J. Williams, Janos Hajdu,
Nicusor Timneanu, M. Marvin Seibert, Jakob Andreasson, Andrea Rocker,
Olof Jonsson, Stephan Stern, Francesco Stellato, Karol Nass, Robert Andritschke,
Claus-Dieter Schroter, Faton Krasniqi, Mario Bott, Kevin E. Schmidt, Xiaoyu Wang,
Ingo Grotjohann, James Holton, Stefano Marchesini, Raimund Fromme,
Sebastian Schorb, Daniela Rupp, Marcus Adolph, Tais Gorkhover, Martin Svenda,
Helmut Hirsemann, Guillaume Potdevin, Heinz Graafsma, Bjorn Nilsson
and John C. H. Spence

# Acknowledgements (later experiments)

Sebastien Boutet, Lukas Lomb, Garth J. Williams, Thomas R. M. Barends,
Andrew Aquila, R. Bruce Doak, Uwe Weierstall, Danial P. DePonte,
Jan Steinbrene, Robert L. Shoeman, Marc Messerschmidt, Anton Barty,
Stephan Kassemeyer, Richard A. Kirian, M. Marvin Seibert, Paul A. Montanez,
Chris Kenney, Ryan Herbst, Philip Hart, Jack Pines, Gunther Haller,
Sol M. Gruner, Hugh T. Philipp, Mark W. Tate, Marianne Hromalik,
Lucas J. Koerner, Niels van Bakel, John Morse, Wilfred Ghonsalves,
David Arnlund, Michael J. Bogan, Carl Caleman, Raimund Fromme,
Christina Y. Hampton, Mark S. Hunter, Linda Johansson, Gergely Katona,
Christopher Kupitz, Mengning Liang, Andrew V. Martin, Karol Nass,
Lars Redecke, Francesco Stellato, Nicusor Timneanu, Dingjie Wang,
Nadia A. Zatsepin, Donald Schafer, James Defever, Richard Neutze,
Petra Fromme, John C. H. Spence. Henry N. Chapman and Ilme Schlichting

# Acknowledgements

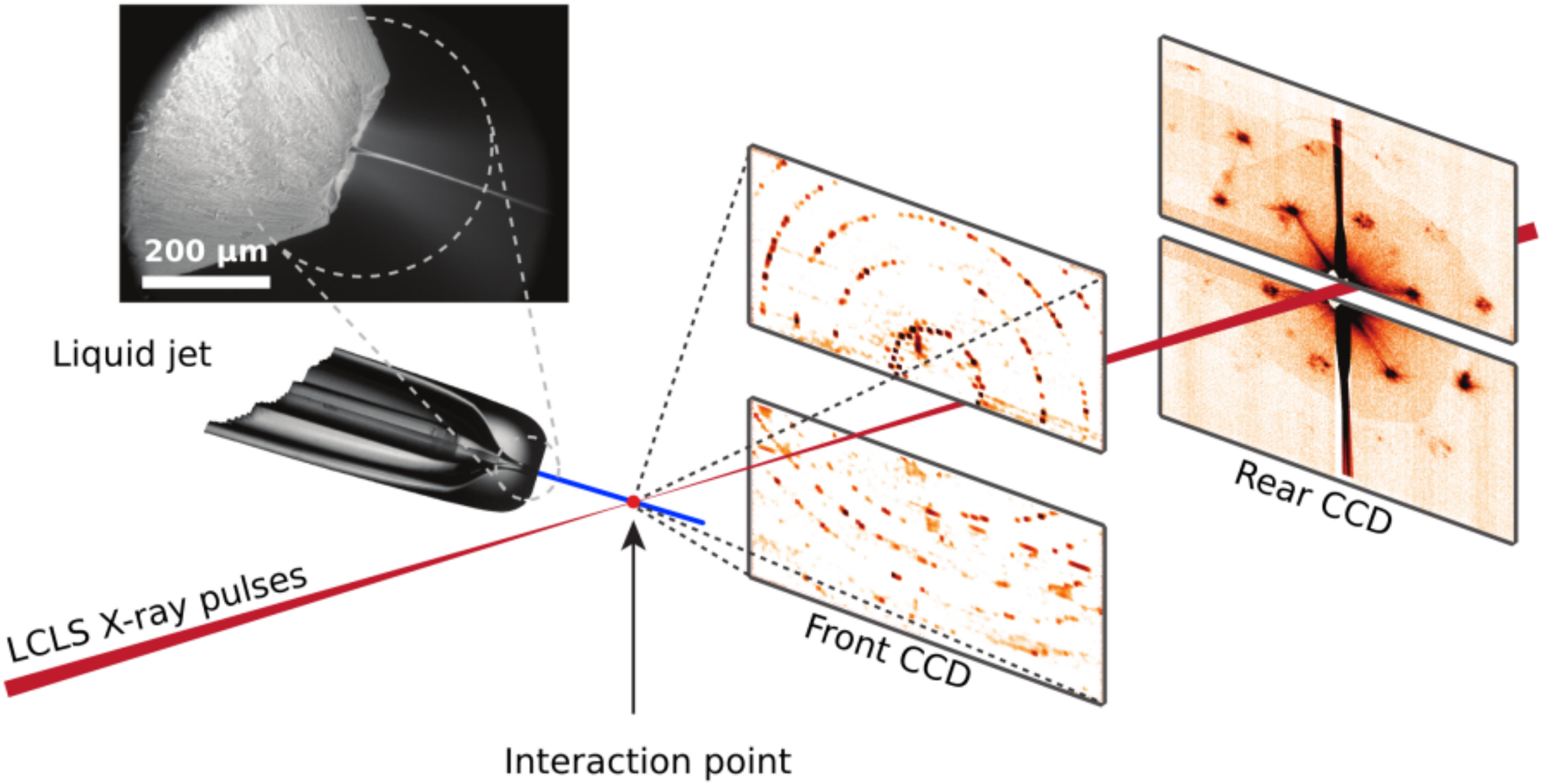
Anton Barty, Kenneth Beyerlein, Lorenzo Galli, Richard Kirian,
Andrew Martin, Karol Nass, Francesco Stellato and Henry Chapman
CFEL / University of Hamburg, Germany

Andrew Aquila and Chun Hong Yoon
European XFEL, Hamburg, Germany

Nadia Zatsepin, Shibom Basu, Raimund Fromme, Petra Fromme
Deptartment of Physics, Arizona State University, USA

James Holton
Biochemistry and Biophysics, UCSF, USA
Physical Biosciences, LBNL, USA

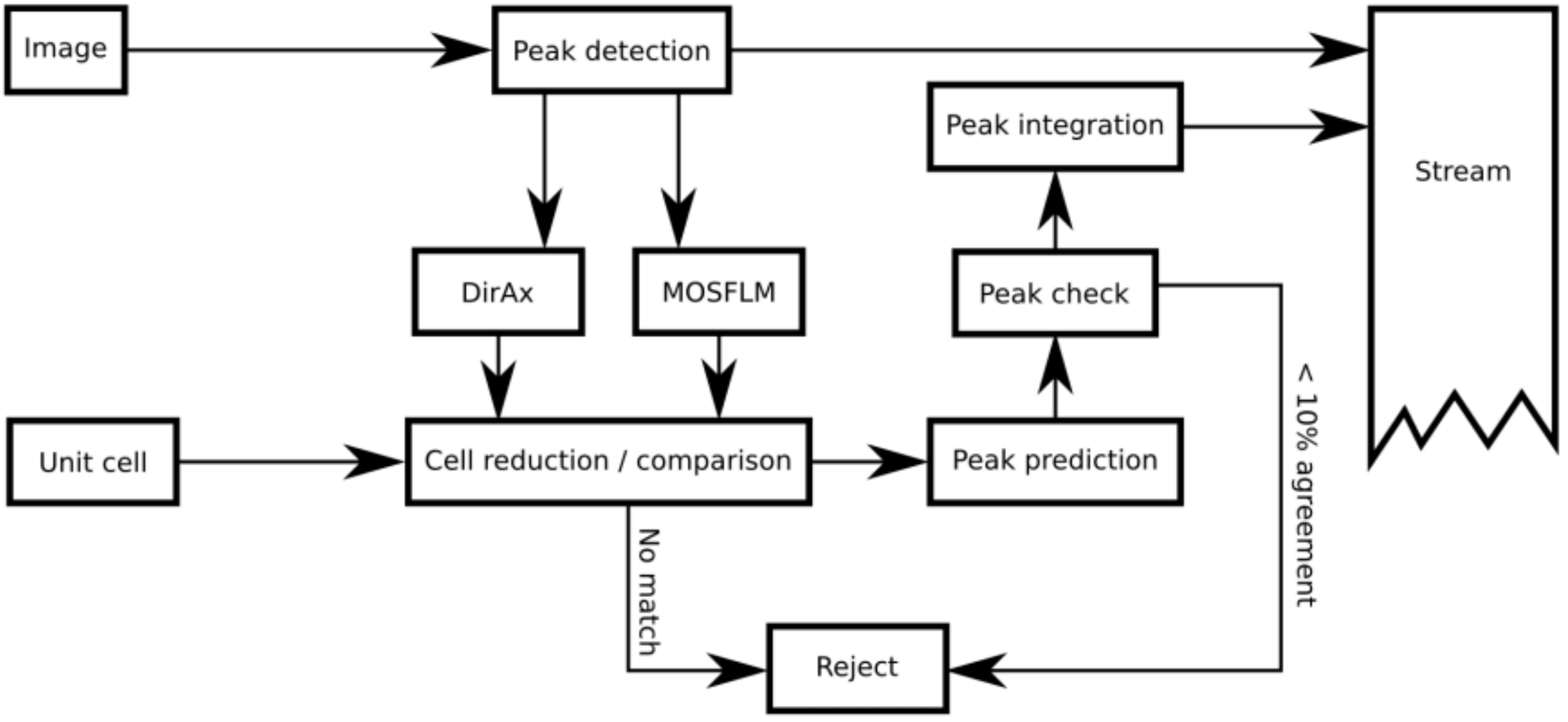
# Experimental setup for serial crystallography




One X-ray pulse, one detector readout, (hopefully) one crystal.

First published in Nature 470 (2011) p73-77

# CrystFEL indexing pipeline




Fully automated indexing and integration pipeline.

Parameters, e.g. for peak picking or unit cell tolerance, can be optimised to get the highest "yield".

# Data quantities in the "ideal" experiment

With current data processing methods, we need around 10,000 to 20,000 indexed diffraction patterns to solve the structure.

The indexing yield varies between 10% and 50%.

A good number of patterns to aim for is 100,000
- these will vary in quality and "indexability".

"Hit rate" varies between 1% and 10%, so we need about one million "shots".

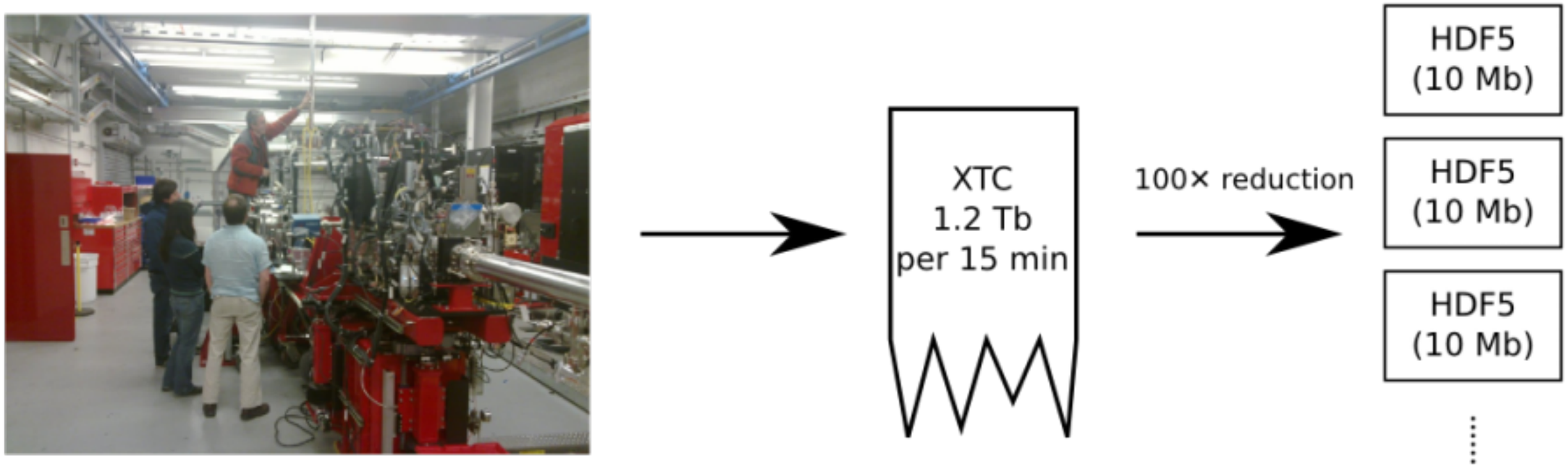
Two 2048 by 2048 pixel detectors, 16 bit readout = 16 Mb / shot.

At 120 Hz, this is about 2 Gb / second.

16 Tb for the whole dataset, acquired in just over two hours.

# Pre-processing software: Cheetah

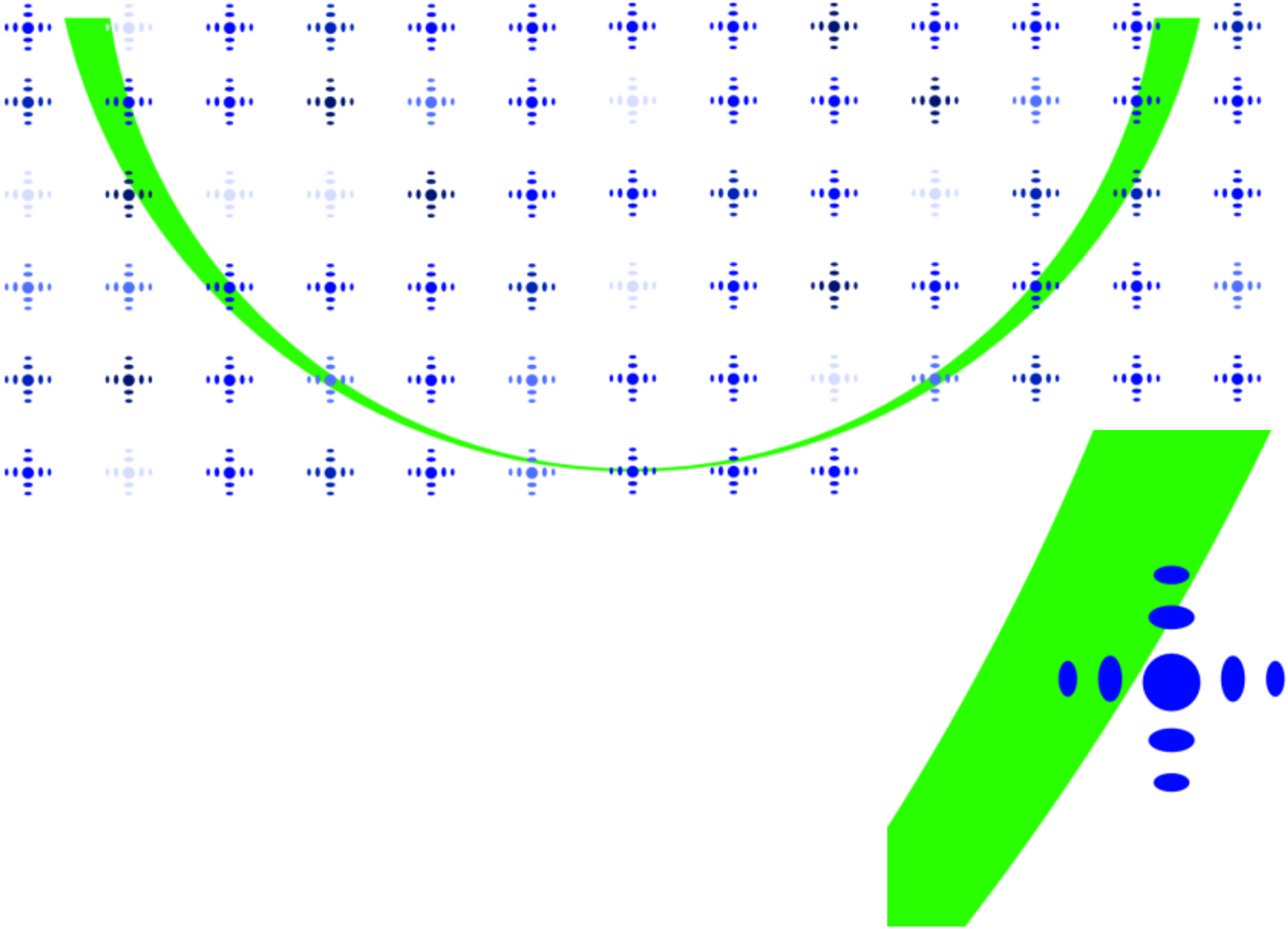
Cheetah performs detector corrections and finds the "hits".
Each "hit" is written to an individual HDF5 file for later processing.




Processing speed is almost the same as acquisition speed.

http://www.desy.de/~barty/cheetah/

# Snapshots are made up of partial reflections



# Every pattern is different...

Wavelength jitter

Crystal size

Quality of crystal

Bandwidth of X-ray pulse

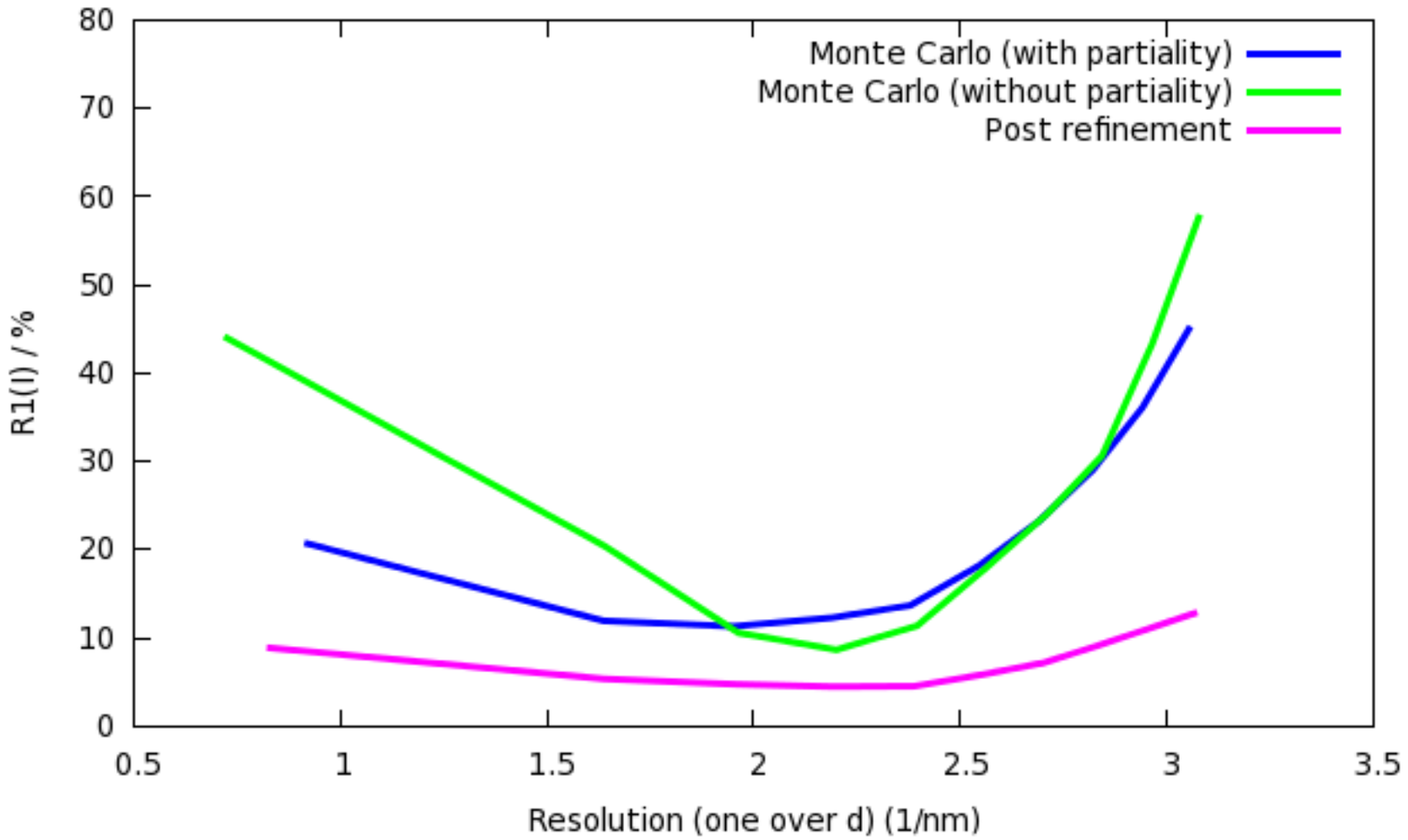
Spectrum of X-ray pulse

Incident intensity

Position of crystal in beam

....

# Improved analysis may reduce the amount of data



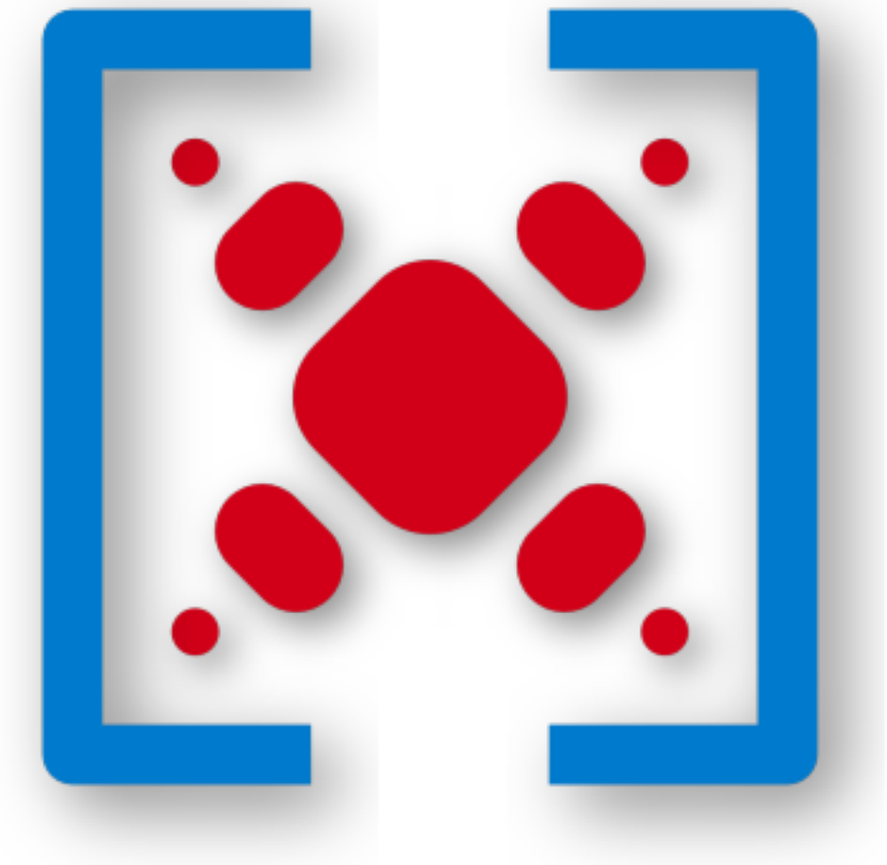

By modelling and reconstructing some of the things which vary between shots, we should be able to reduce the number of patterns required, perhaps significantly.

# CrystFEL software suite

Software suite for processing diffraction data from XFEL sources, and perhaps other experiments of a "serial crystallography" type.

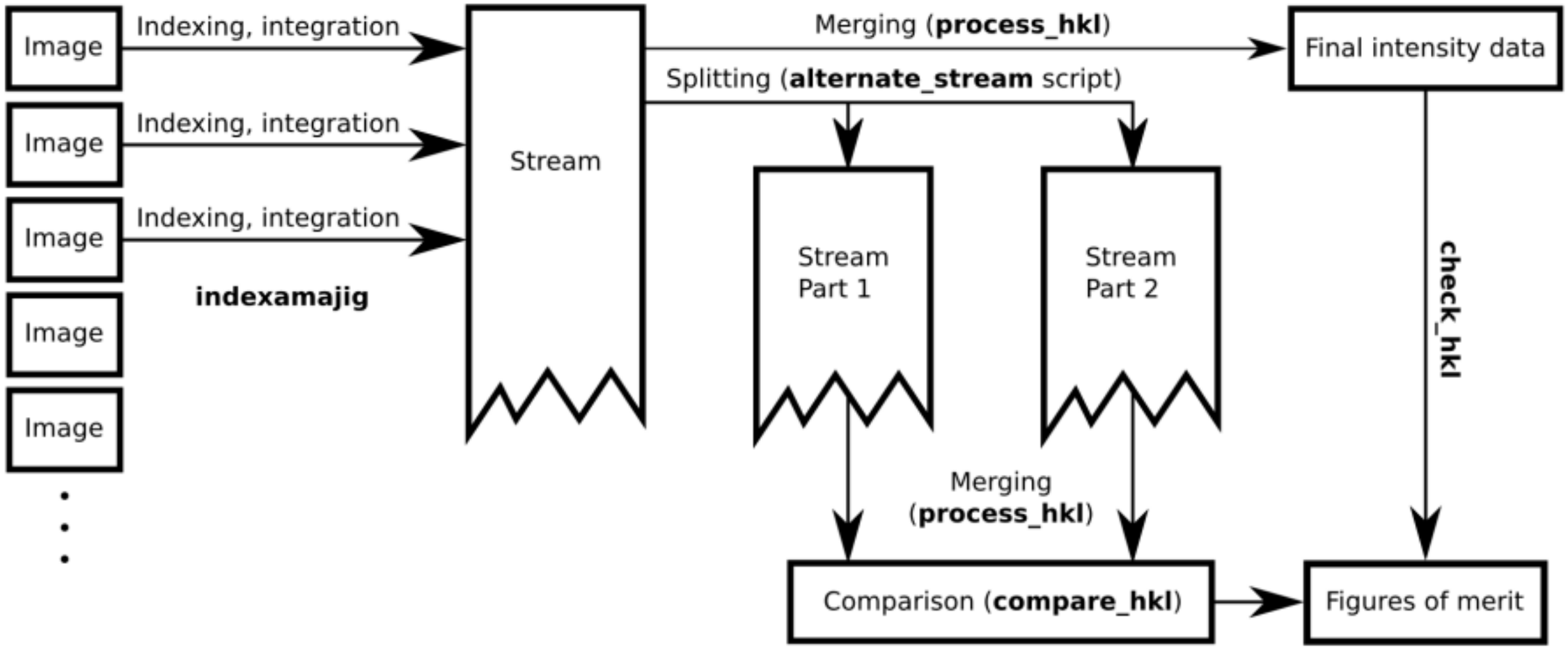
"Clean diffraction patterns to merged intensities."

Free and open-source software: GPLv3



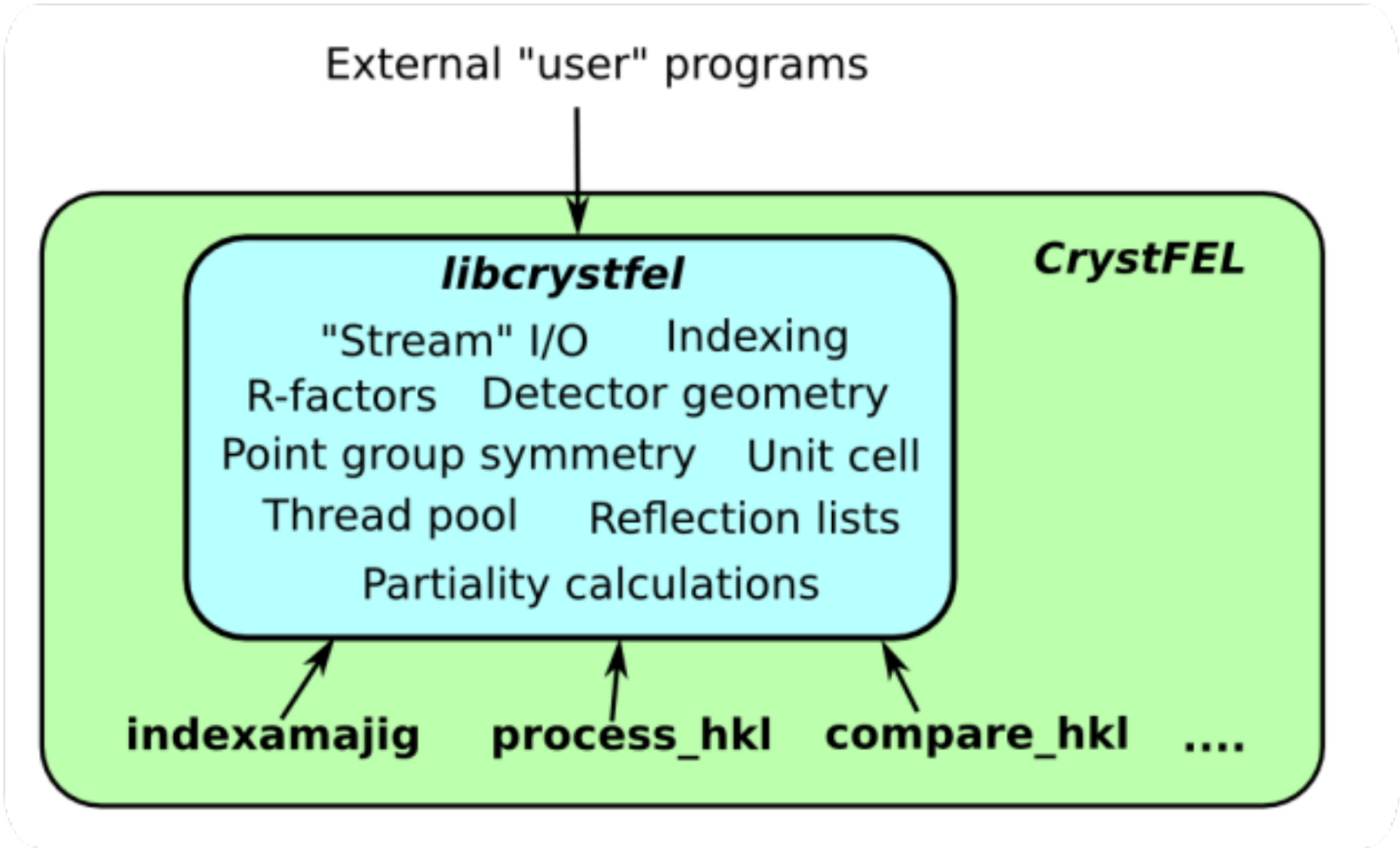
http://www.desy.de/~twhite/crystfel/

# CrystFEL software suite



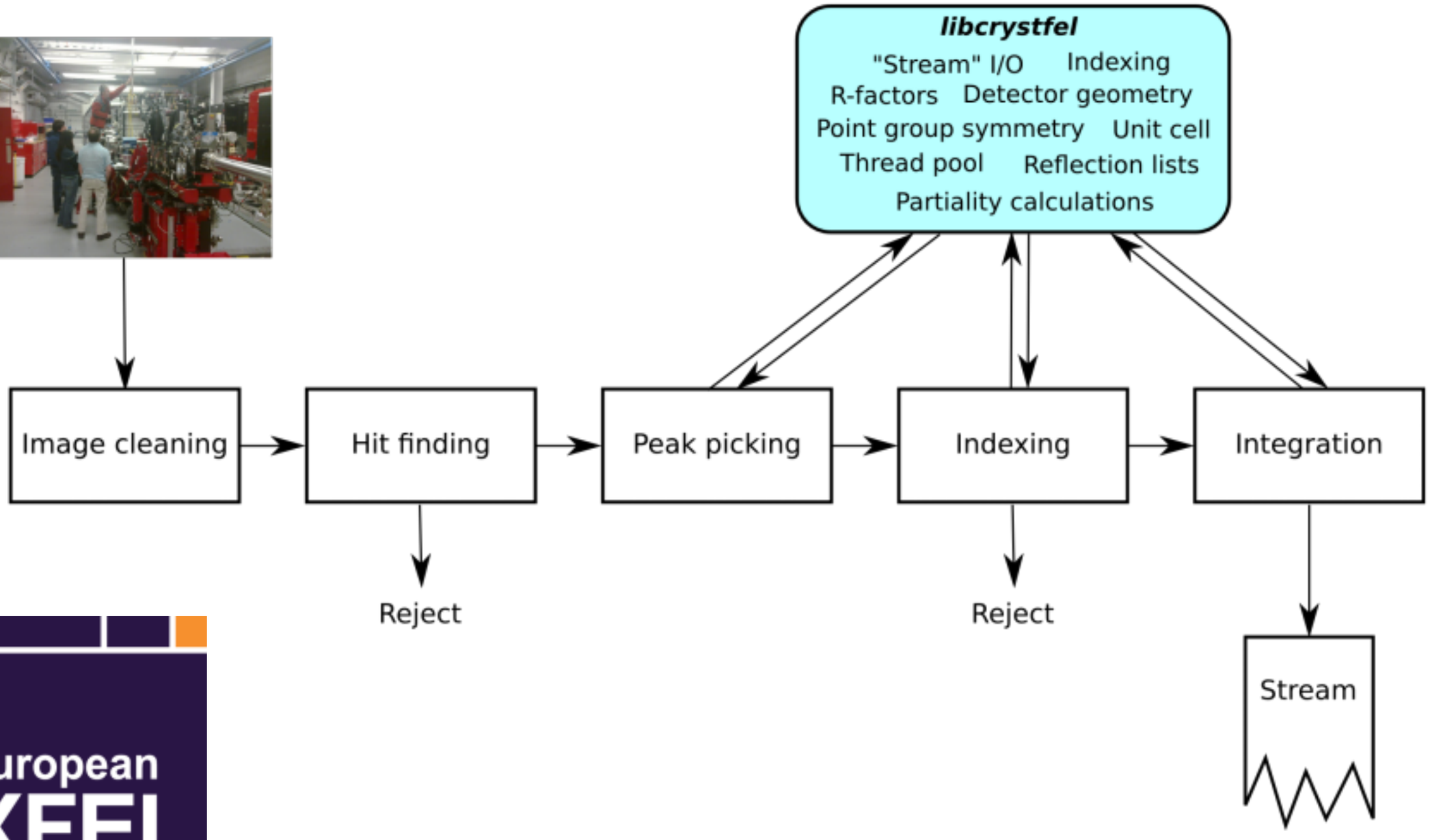

"Stream" contains all the relevant information in a series of "chunks". One chunk per pattern (whether indexed successfully or not).

# CrystFEL software suite



# The future ... ?






European XFEL

# References

T. A. White et al., "CrystFEL: A software suite for snapshot serial crystallography". J. Applied Crystallography 45 (2012) p334-341.

H. N. Chapman et al., "Femtosecond X-ray protein nanocrystallography", Nature 73 (2011) p73-77.

R. A. Kirian et al., "Structure-factor analysis of femtosecond micro-diffraction patterns from protein nanocrystals". Acta Crystallographica A68 (2011) p131-140.

http://www.desy.de/~twhite/crystfel

taw@physics.org