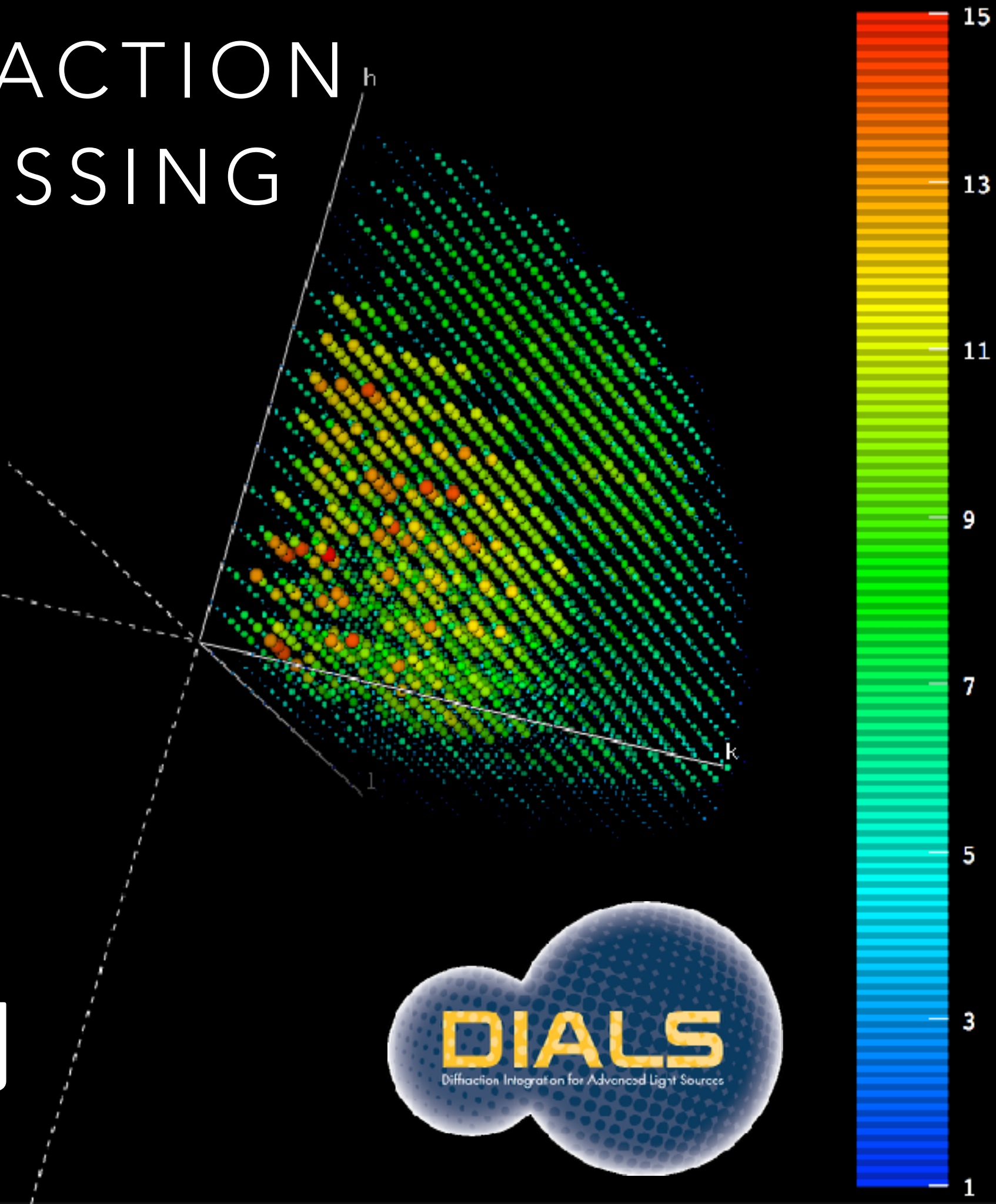
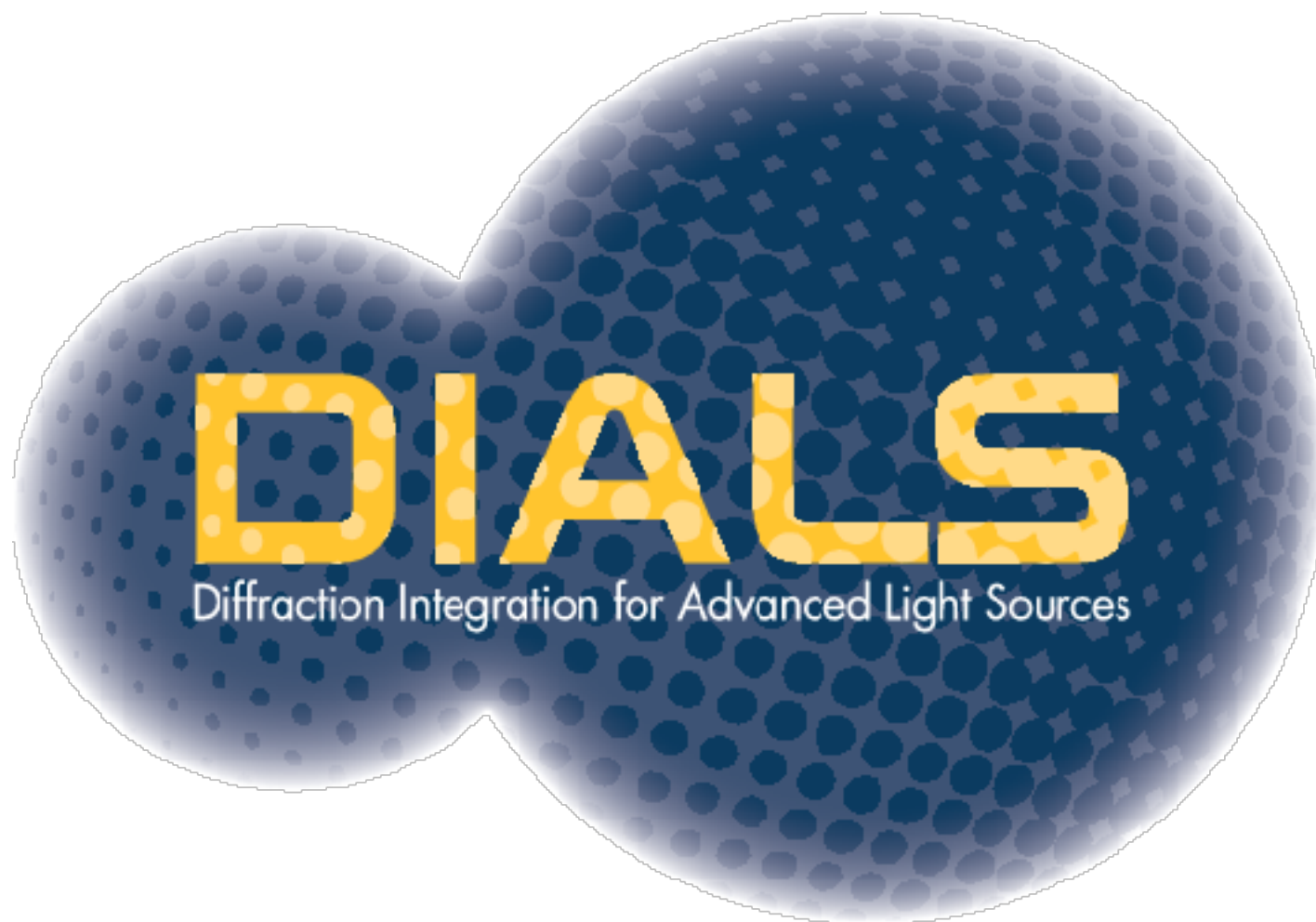


X-RAY DIFFRACTION DATA PROCESSING WITH DIALS & XIA2/DUI

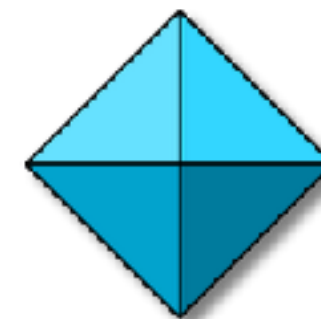
DLS /CCP4
JAMES BEILSTEN-EDMANDS
DIAMOND LIGHT SOURCE



PROJECT SUPPORT



Science & Technology
Facilities Council



CCP4

EU FP7: #283570 NIH: GM095887 & GM102520

ACKNOWLEDGEMENTS

research papers

Acta Crystallographica Section D
Biological
Crystallography
ISSN 0907-4449

XDS

Wolfgang Kabsch

Max-Planck-Institut für Medizinische Forschung,
Abteilung Biophysik, Jahnstrasse 29,
69120 Heidelberg, Germany

Correspondence e-mail:
wolfgang.kabsch@mpi-med.uni-heidelberg.de

The usage and control of recent modifications of the program package *XDS* for the processing of rotation images are described in the context of previous versions. New features include automatic determination of spot size and reflecting range and recognition and assignment of crystal symmetry. Moreover, the limitations of earlier package versions on the number of correction/scaling factors and the representation of pixel contents have been removed. Large program parts have

Received 19 August 2009
Accepted 9 November 2009

A version of this paper will be published as a chapter in the new edition of Volume F of International Tables for Crystallography.

research papers

Acta Crystallographica Section D
Biological
Crystallography
ISSN 0907-4449

The finer things in X-ray diffraction data collection

J. W. Pilgrath

Molecular Structure Corporation, 5009 New
Trails Drive, The Woodlands, TX 77381, USA

Correspondence e-mail: jwp@msc.com

X-ray diffraction images from two-dimensional position-sensitive detectors can be characterized as thick or thin, depending on whether the rotation-angle increment per image is greater than or less than the crystal mosaicity, respectively. The expectations and consequences of the processing of thick and thin images in terms of spatial overlap, saturated pixels, X-ray background and $\overline{R}_{int}$ are discussed. The *dSTAR* software suite for processing diffraction images is briefly introduced, and results from *dSTAR* are compared with those from another popular package.

Received 6 May 2009
Accepted 5 July 2009

computer programs

Journal of
Applied
Crystallography
ISSN 0021-8898

xia2: an expert system for macromolecular crystallography data reduction

G. Winter

Diamond Light Source, Hermal Science and Innovation Campus, Colindale, UK; STFC Daresbury Laboratory, Warrington, Cheshire, UK; and University of Manchester, Manchester, UK

An expert system for macromolecular crystallography data reduction is presented, which builds on existing software to automate the complete data reduction process from images to merged structure factor amplitudes. This can automatically identify multi-wedge, multi-pass and multiwavelength data sets and includes explicit procedures to test for crystallographic special cases. With the push towards high-throughput crystallography at synchrotron beamlines and automation of structure solution, the ability to reduce data with no user input

Received 3 September 2009
Accepted 30 October 2009

research papers

Acta Crystallographica Section D
Biological
Crystallography
ISSN 0907-4449

The integration of macromolecular diffraction data

Andrew G. W. Leslie

MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, England

Correspondence e-mail:
andrew@lmb.cam.ac.uk

The objective of any modern data-processing program is to produce from a set of diffraction images a set of indices (*hkl*s) with their associated intensities (and estimates of their uncertainties), together with an accurate estimate of the crystal unit-cell parameters. This procedure should not only be reliable, but should involve an absolute minimum of user intervention. The process can be conveniently divided into three stages. The first (autoindexing) determines the unit-cell parameters and the orientation of the crystal. The unit-cell parameters may indicate the likely Laue group of the crystal. The second step is to refine the initial estimate of the unit-cell parameters and also the crystal mosaicity using a procedure known as post-refinement. The third step is to integrate the images, which consists of predicting the positions of the Bragg reflections on each image and obtaining an estimate of the intensity of each reflection and its uncertainty. This is carried out while simultaneously refining various detector and crystal parameters. Basic features of the algorithms employed for

Received 19 May 2007
Accepted 24 November 2007

research papers

Acta Crystallographica Section D
Biological
Crystallography
ISSN 0907-4449

How good are my data and what is the resolution?

Philip R. Evans* and Garib N. Murshudov

MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 0QH, England

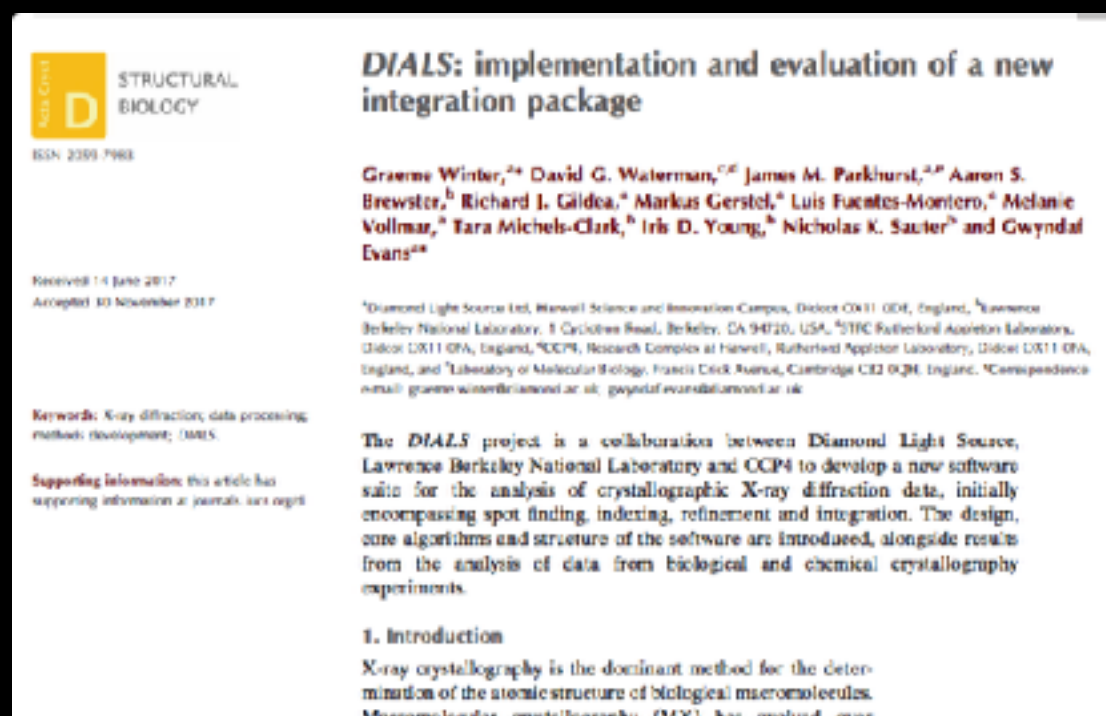
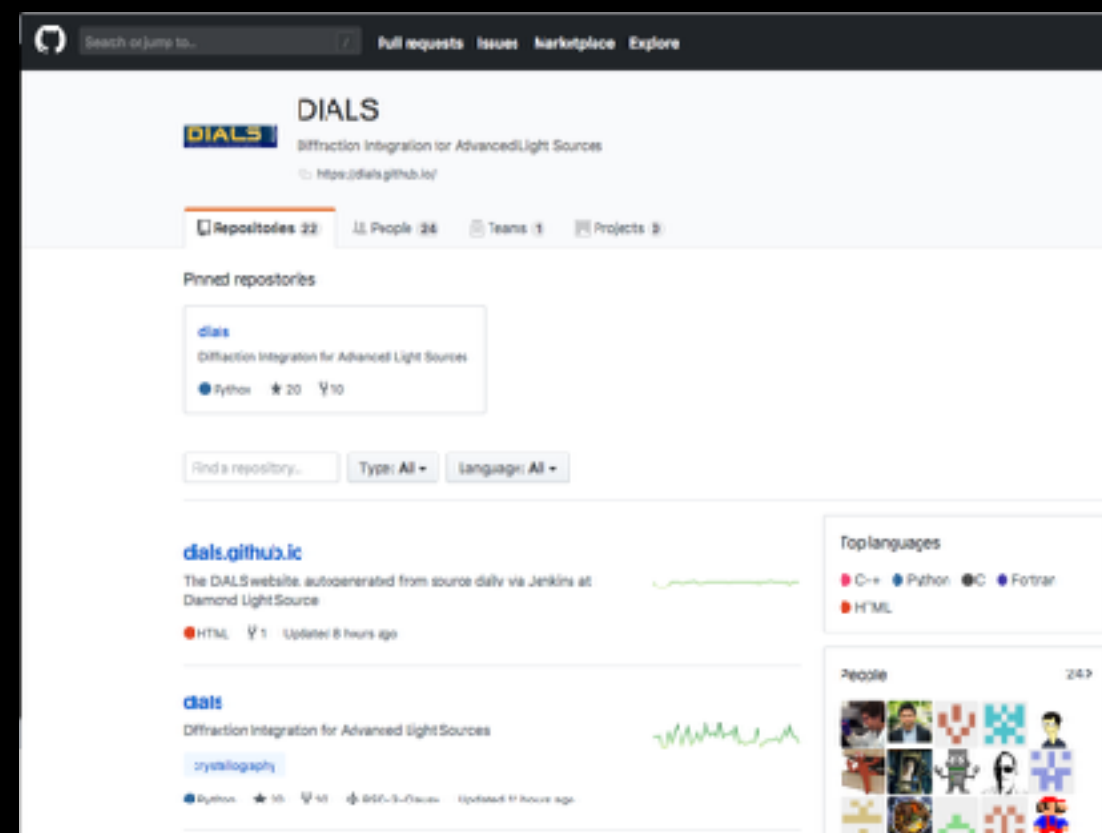
Correspondence e-mail:
pre@mrc-lmb.cam.ac.uk

Following integration of the observed diffraction spots, the process of 'data reduction' initially aims to determine the point-group symmetry of the data and the likely space group. This can be performed with the program *POINTLESS*. The scaling program then puts all the measurements on a common scale, averages measurements of symmetry-related reflections (using the symmetry determined previously) and produces many statistics that provide the first important measures of data quality. A new scaling program, *AIMLESS*, implements scaling models similar to those in *SCALA* but adds some additional analyses. From the analyses, a number of decisions can be made about the quality of the data and whether some measurements should be discarded. The effective 'resolution' of a data set is a difficult and possibly contentious question (particularly with referees of papers) and this is discussed in

Received 7 September 2012
Accepted 2 January 2013

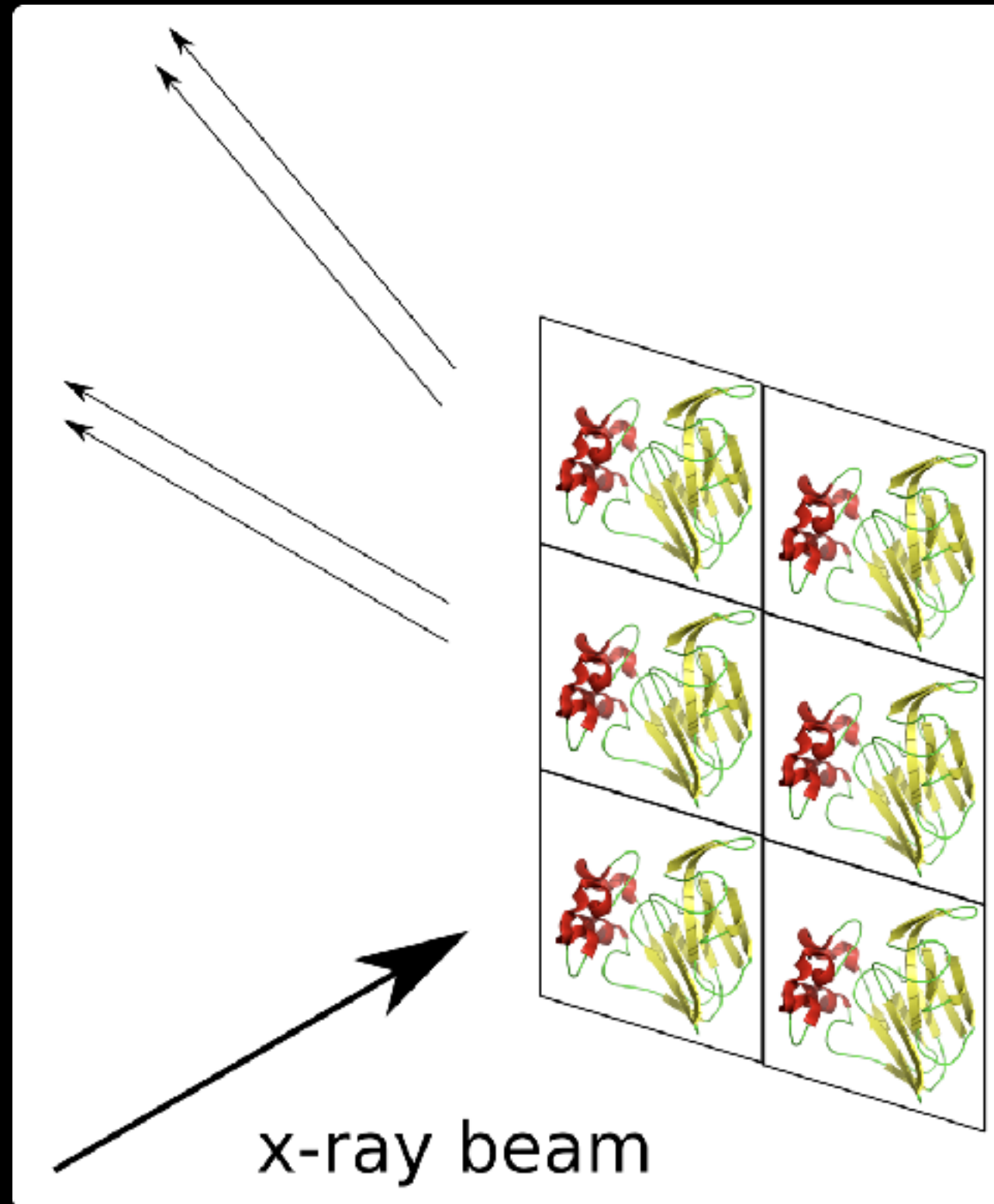
PHILOSOPHY

- DIALS is open source, so you can look at the code to find out exactly what it is doing
- Everything here is applied mathematics and nothing more!
- Happy to help!

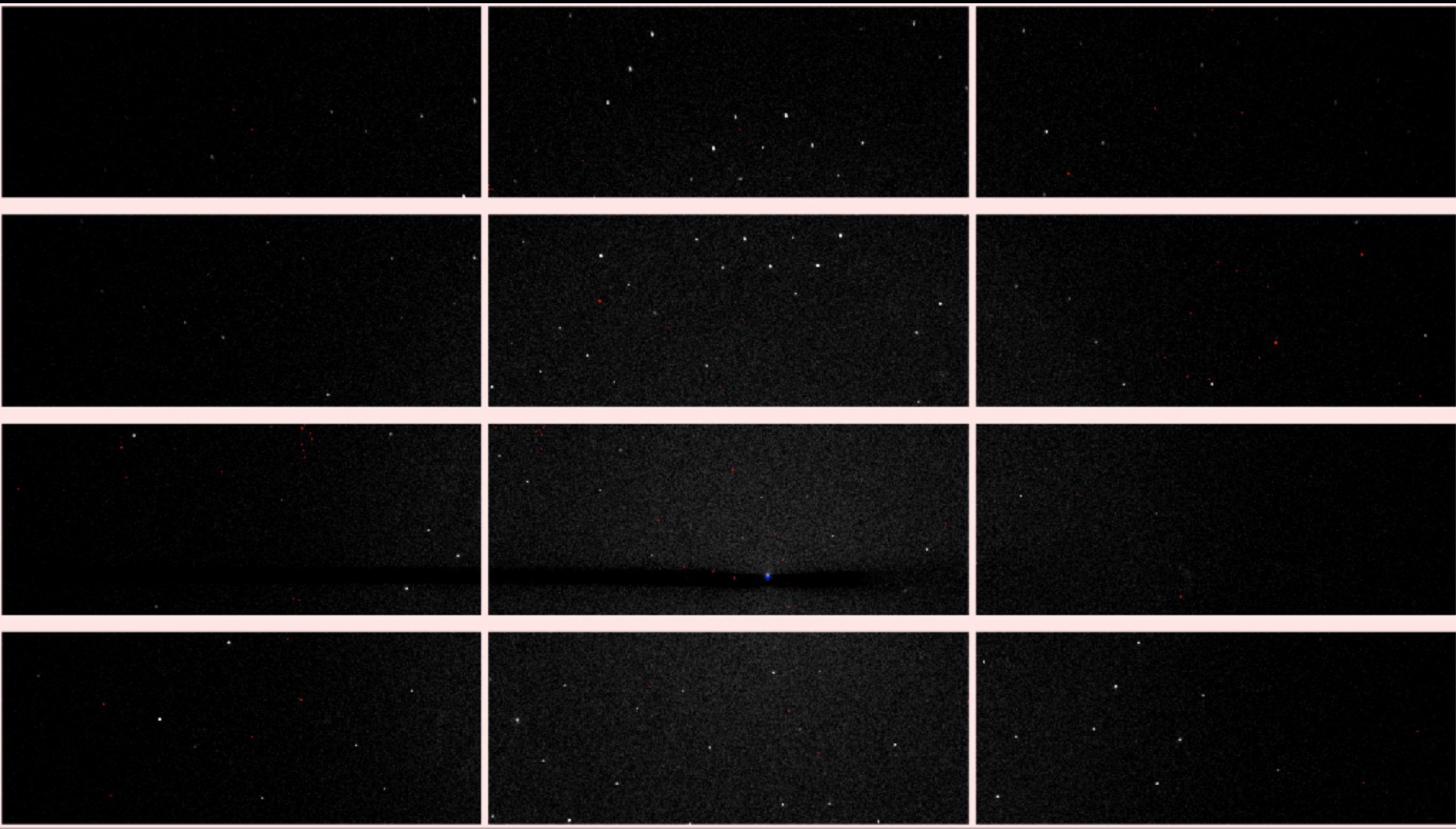


DATA INTEGRATION AND SCALING (IN SIMPLE TERMS)

AIM : INTENSITY DISTRIBUTION



START WITH DIFFRACTION SPOTS

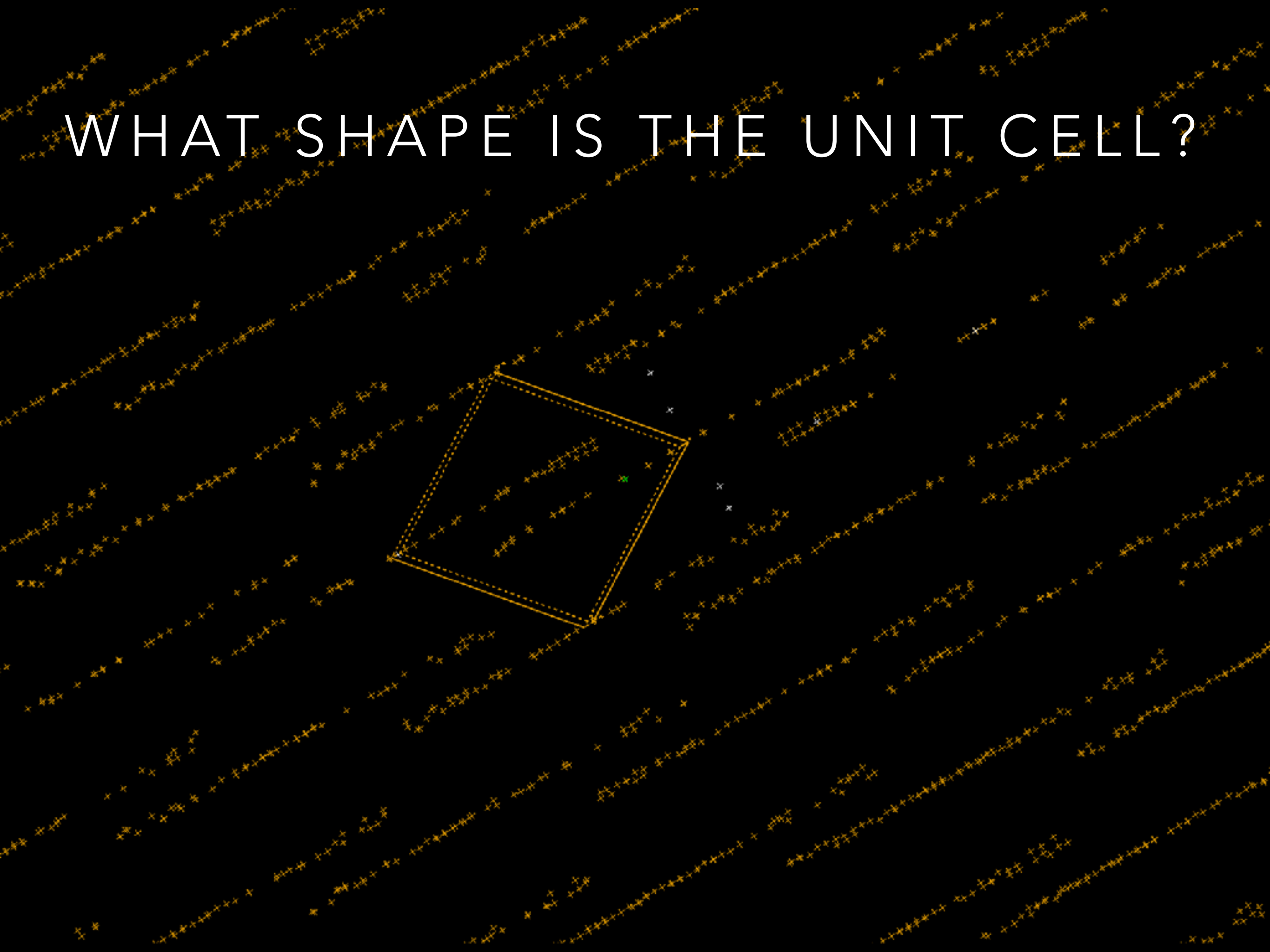


NAME THAT SPOT (MILLER INDICES)

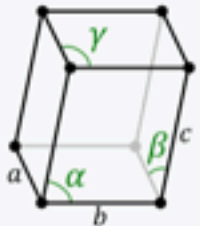
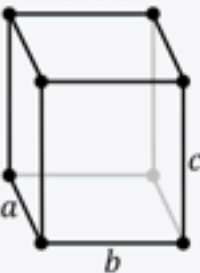
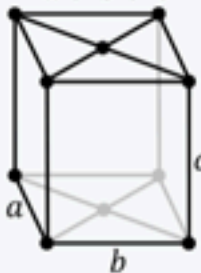
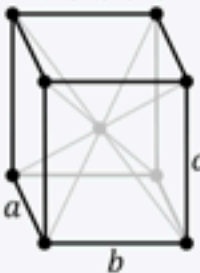
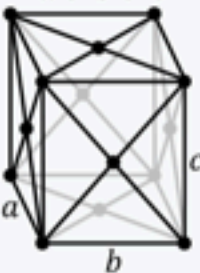
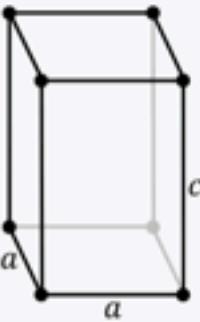
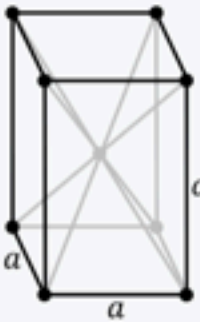
WHICH SPOT IS WHICH?



WHAT SHAPE IS THE UNIT CELL?



WHAT IS THE CORRECT BRAVAIS LATTICE?

Crystal family	Lattice system	Schönflies	14 Bravais Lattices			
			Primitive	Base-centered	Body-centered	Face-centered
triclinic		C_i				
monoclinic		C_{2h}	$\beta \neq 90^\circ$ $a \neq c$ 	$\beta \neq 90^\circ$ $a \neq c$ 		
orthorhombic		D_{2h}	$a \neq b \neq c$ 	$a \neq b \neq c$ 	$a \neq b \neq c$ 	$a \neq b \neq c$ 
tetragonal		D_{4h}	$a \neq c$ 		$a \neq c$ 	

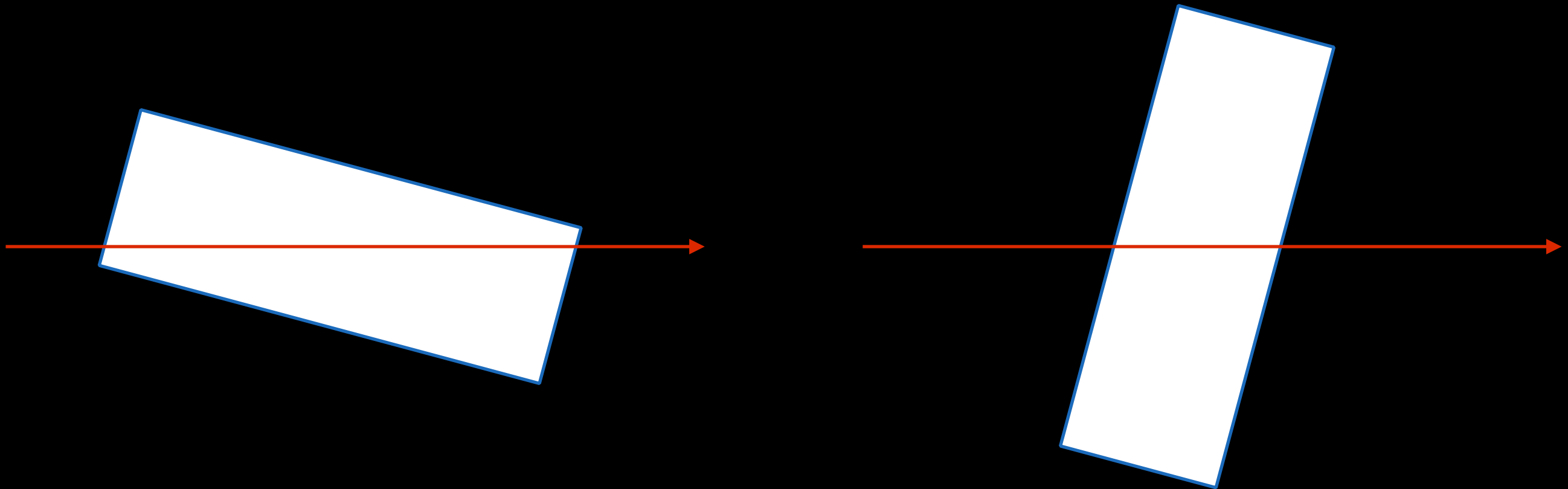
DATA INTEGRATION

HOW MANY COUNTS?



CORRECTIONS FOR EXPERIMENTAL EFFECTS

PLACE ON COMMON SCALE



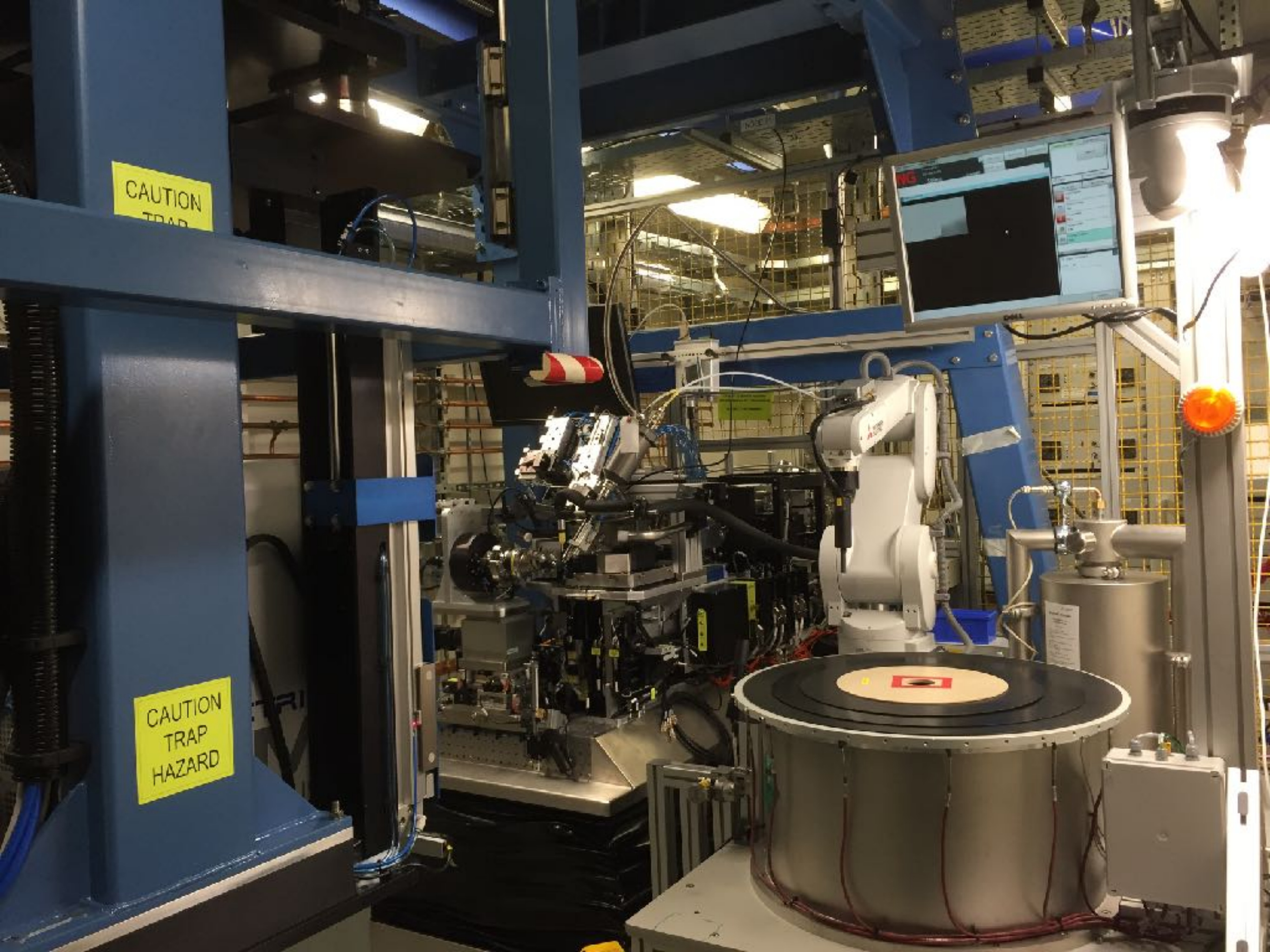
TERMINOLOGY (UNIVERSAL)

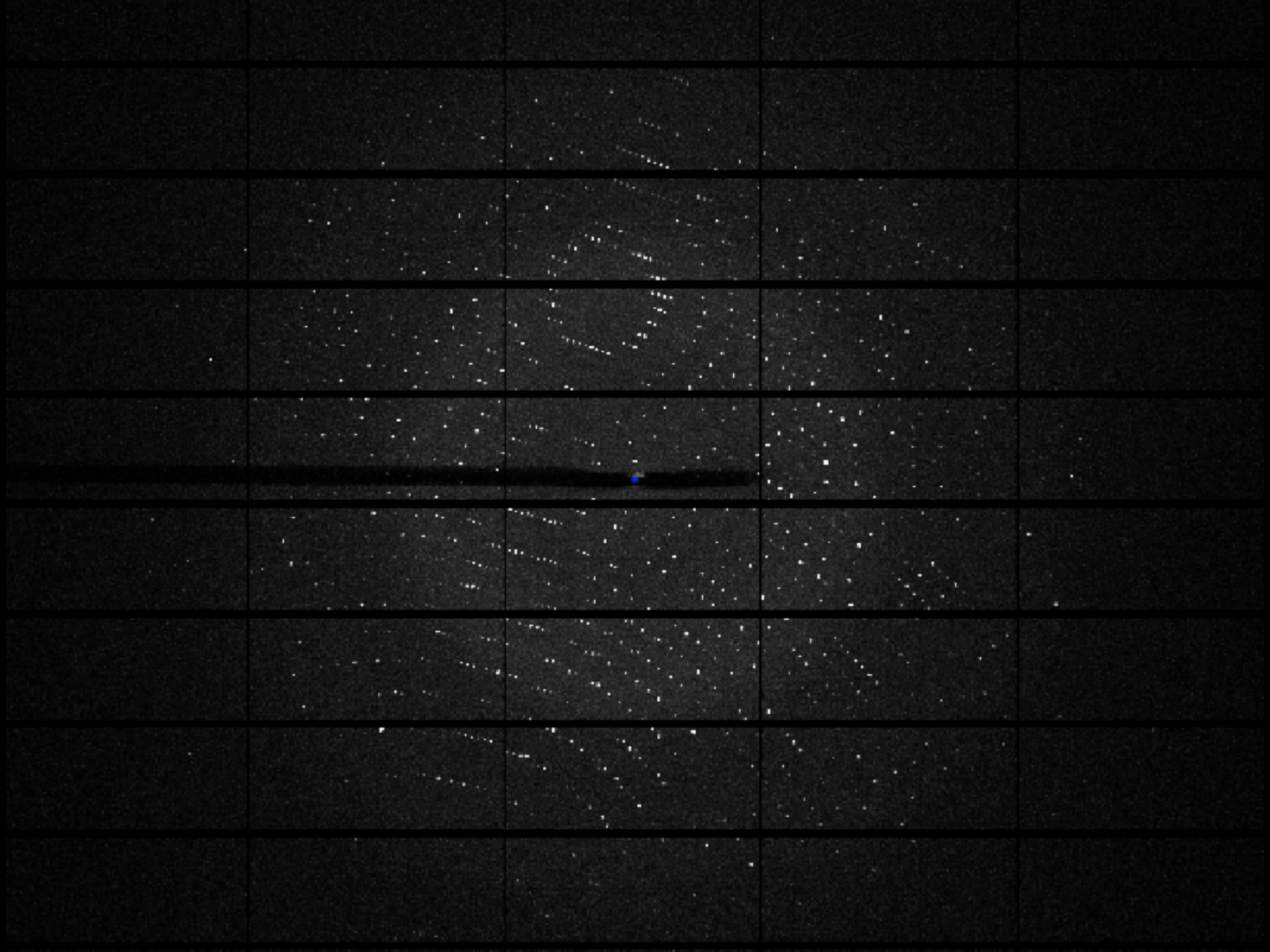
- Spot finding: locate spots on some or all of the data
- Indexing: looking at the positions of the spots and calculating a model which explains those positions (may involve choices about lattice or space group, *optional*)
- Refinement: improving the model
- Integration: record how many counts in each spot, subtract background, usually fit reflection profiles
- Scaling: correction for experimental effects i.e. beam intensity, illuminated volume, sample absorption...

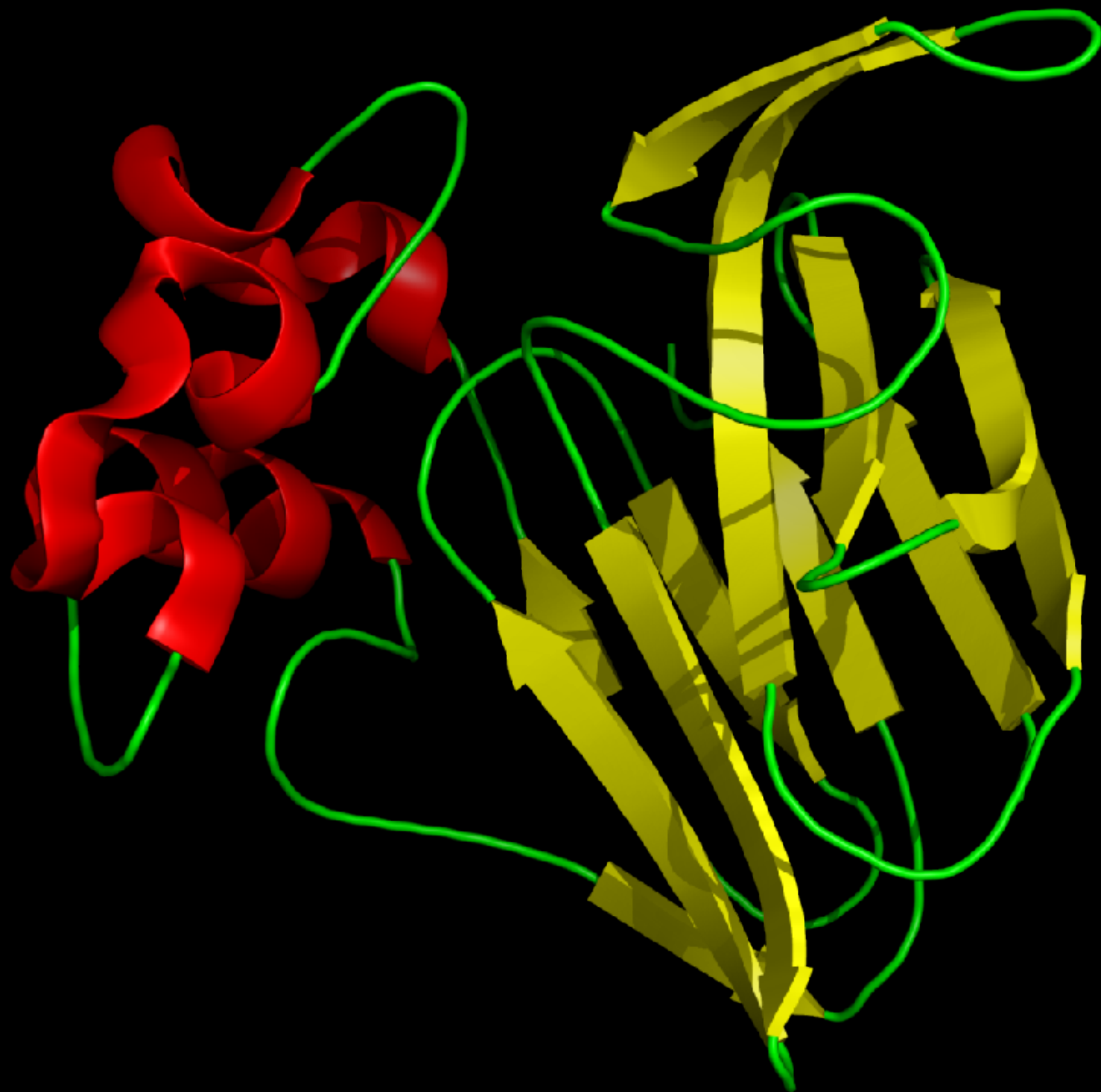
EXPERIMENT

CAUTION
TRAP

CAUTION
TRAP
HAZARD







DATA PROCESSING PROBLEMS

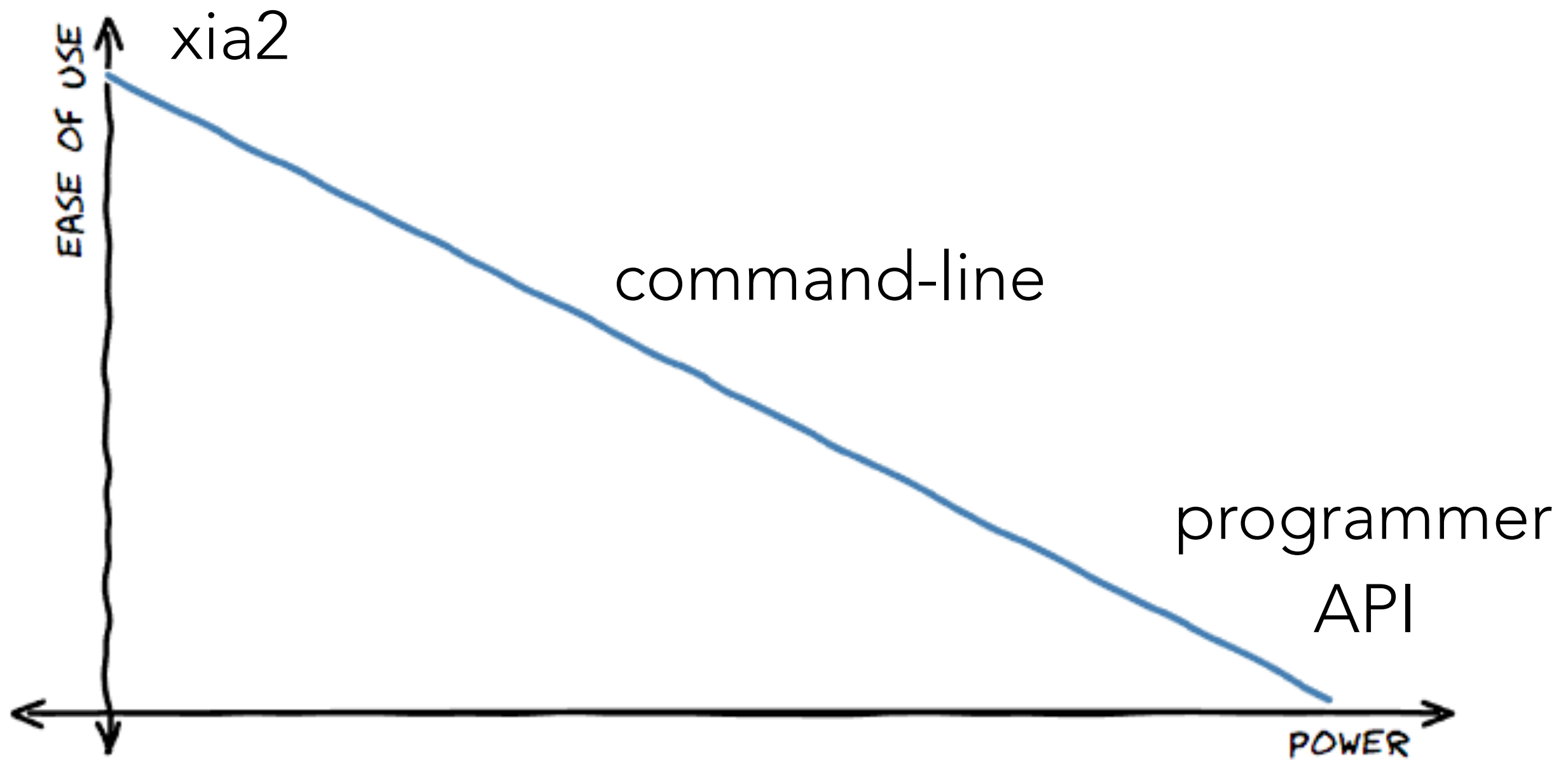
Many can be solved before you run a single program -

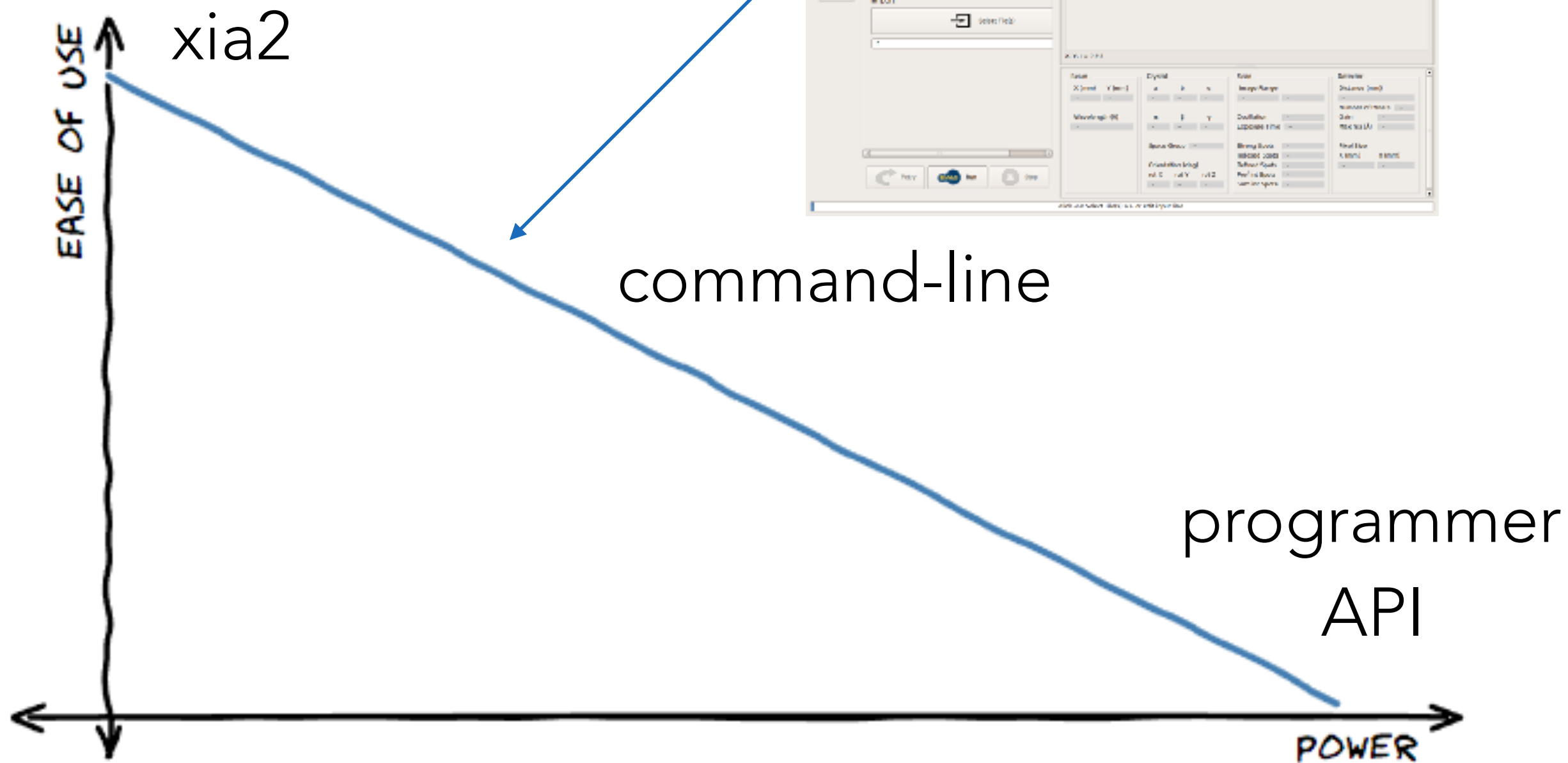
- mount your sample carefully
- collect your data carefully
- avoid / minimise radiation damage

FASTEST ROUTE FROM IMAGES TO INTENSITIES

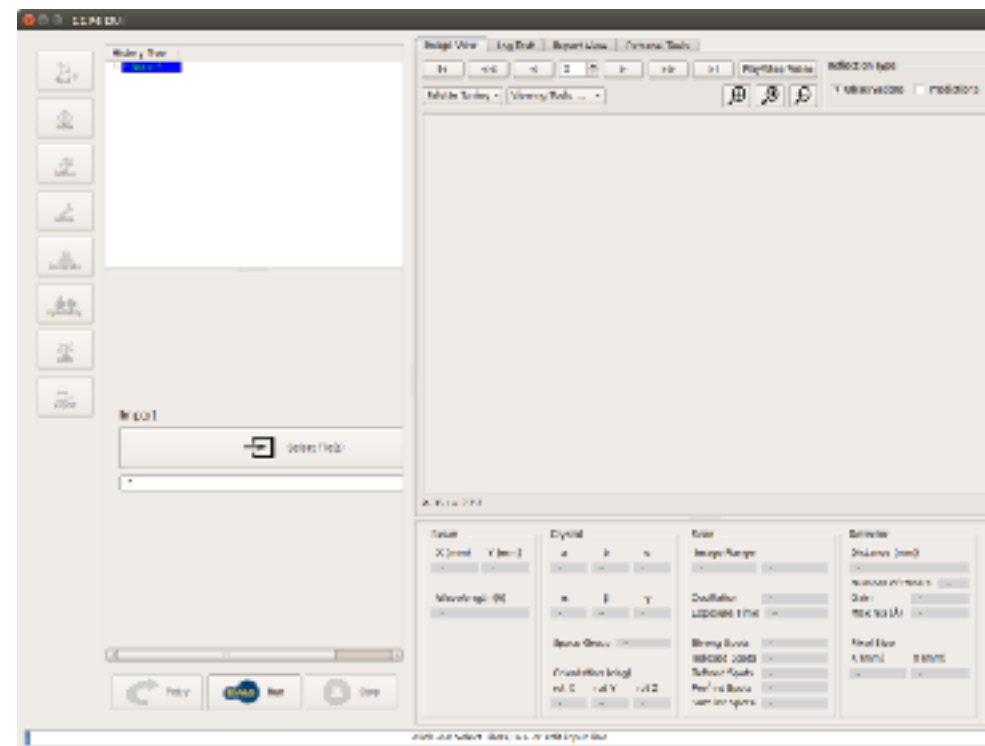
- Automation!
- generate_XDS.INP / autoPROC / xia2 / fast_dp /
most facilities have *something*
- For dials: xia2 =>

xia2 atom=Zn image=../data/C2_sum_0001.cbf.gz
- You would learn *nothing*

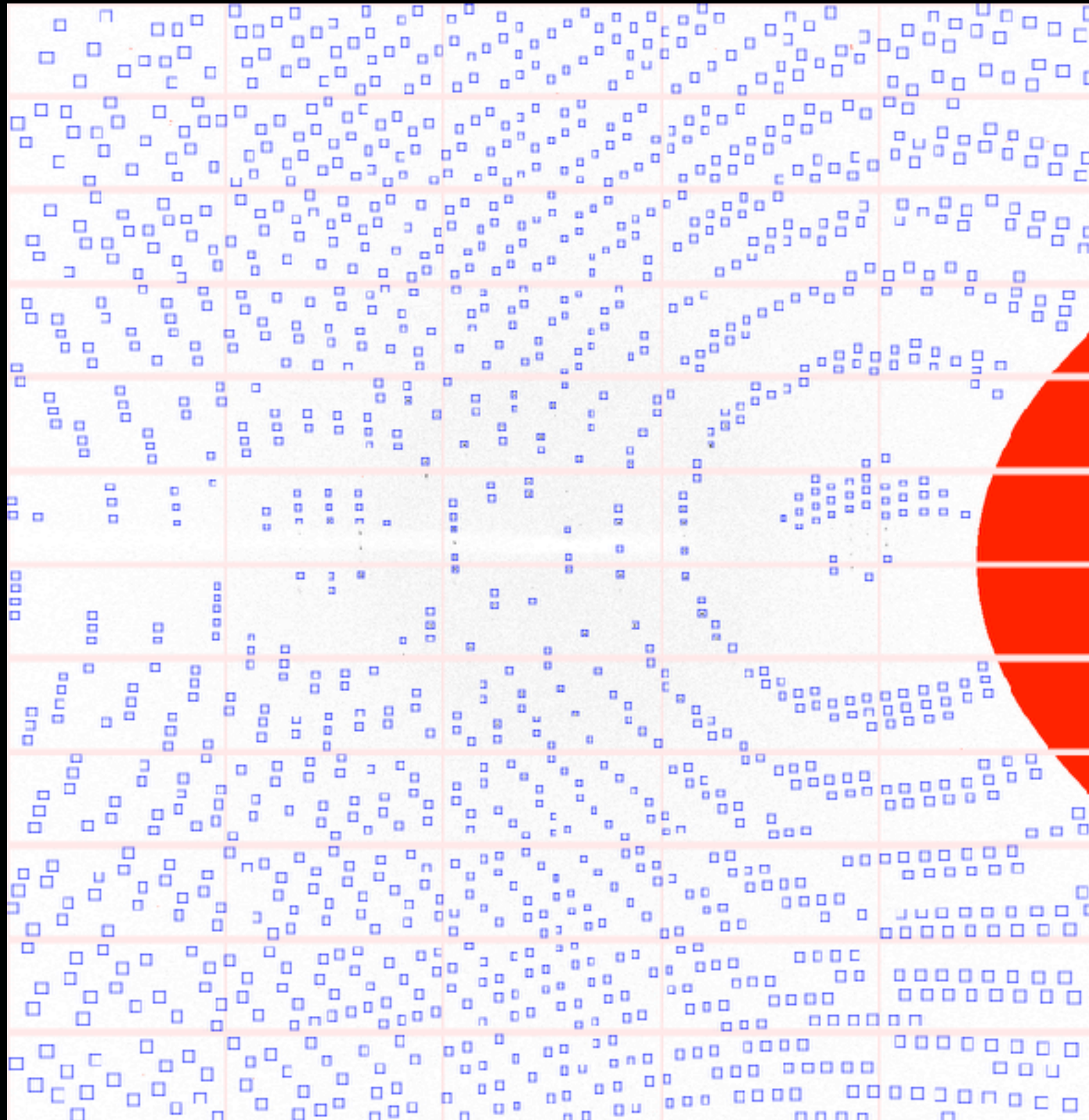




CCP4 DUI

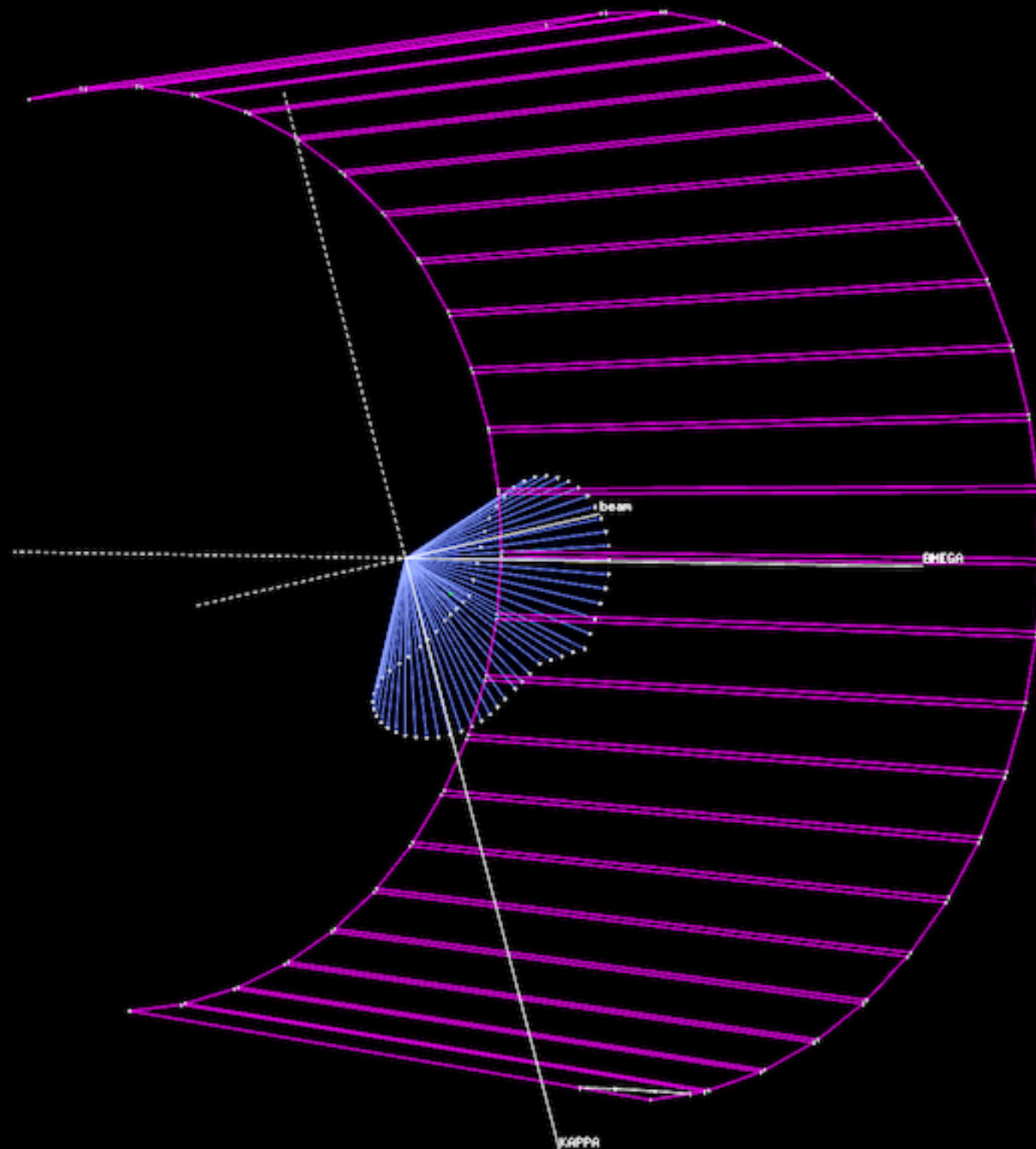


WHAT CAN DIALS COPE WITH?

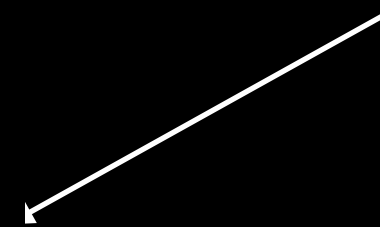


Shadows!

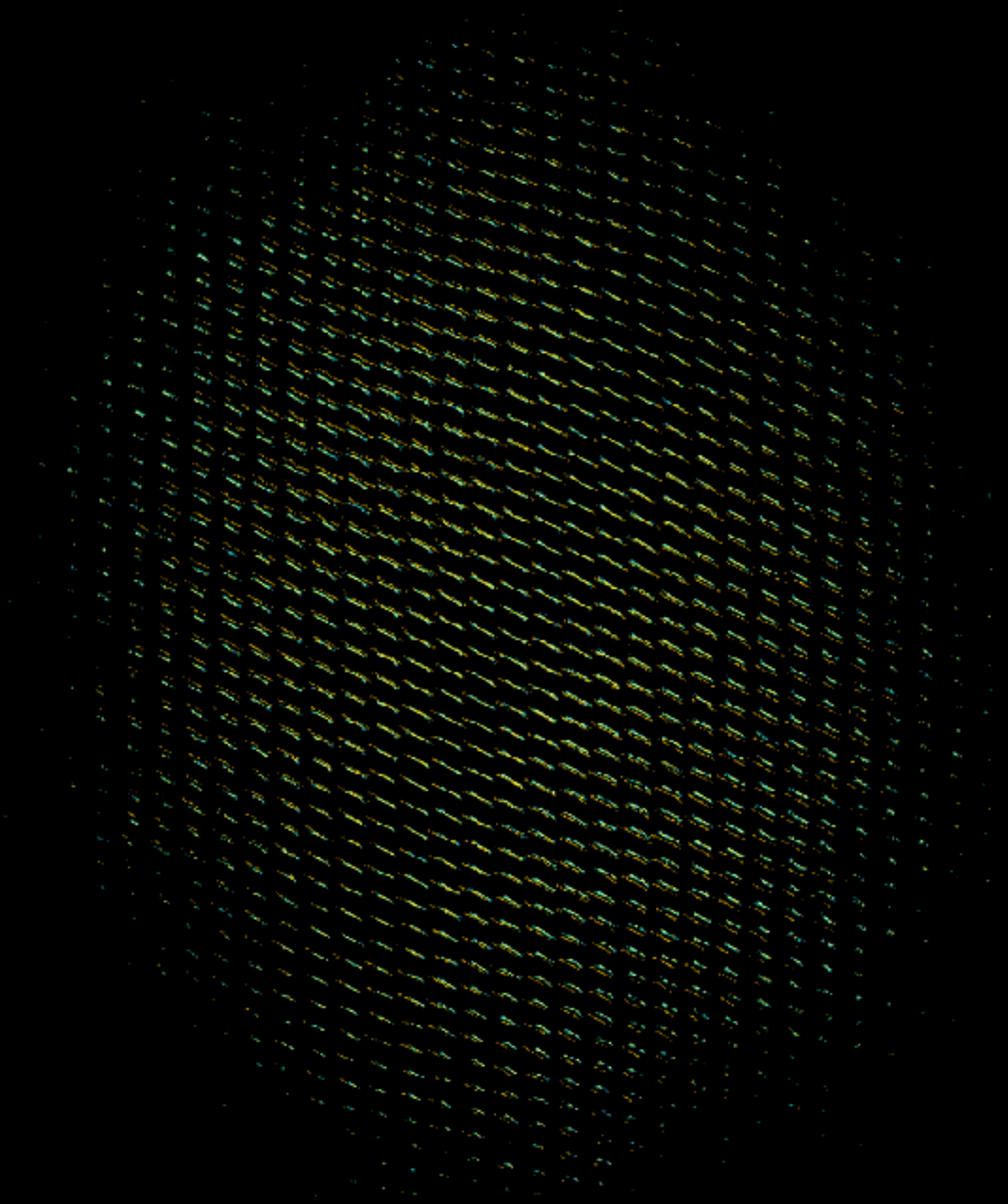
WHAT CAN DIALS COPE WITH?



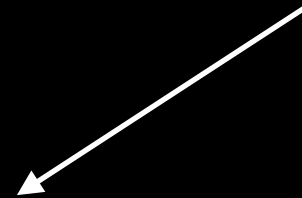
Complex
detectors



WHAT CAN DIALS COPE WITH?



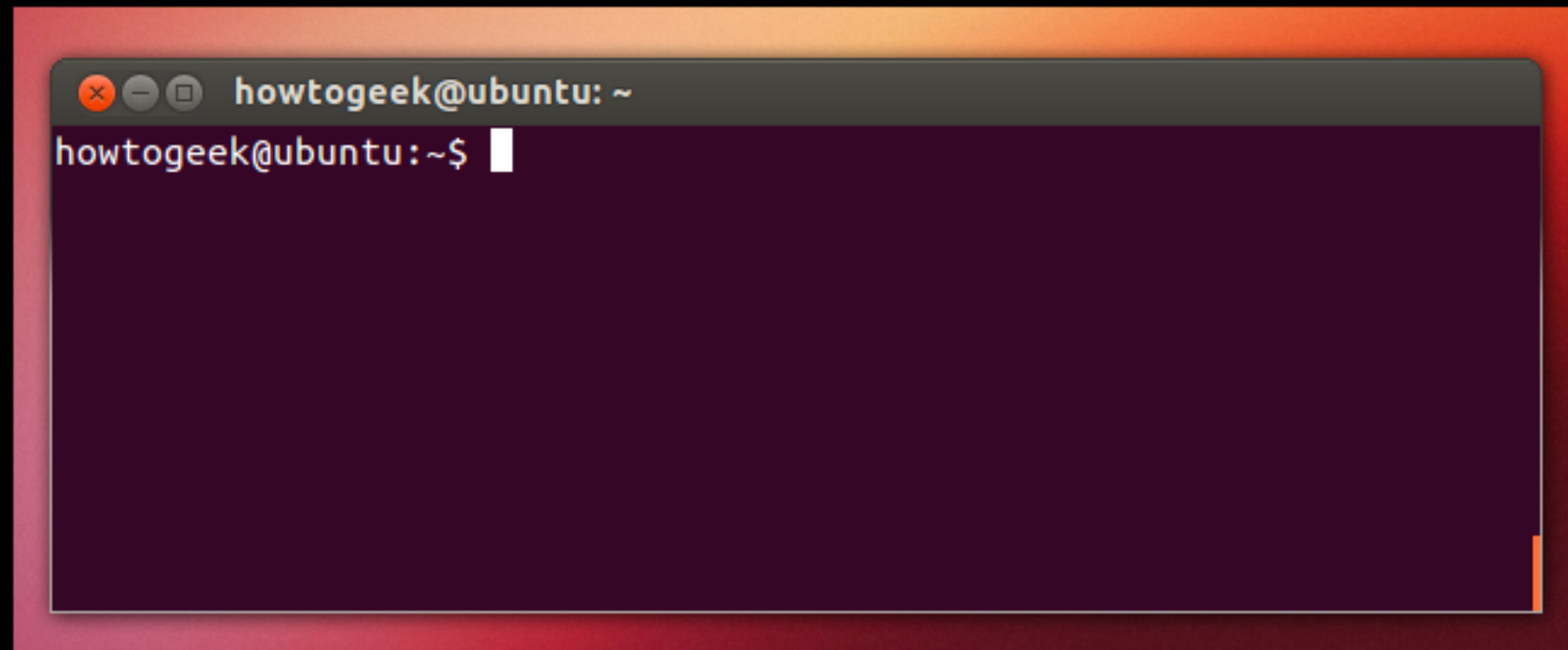
Multiple
orientations



WEAK DIFFRACTION



RUNNING DIALS



DIALS FILES

- CBF / HDF5 files contain X-ray diffraction images
- .pickle files contain reflection data
- .json files contain model information
- MTZ files are probably what you want at the end for structure determination and refinement

KEY STEPS IN DATA PROCESSING

- Look at your images <—— people often miss this one
- Spot finding
- Indexing (optionally selecting lattice)
- Refinement (may be part of indexing)
- Integration
- Post integration refinement
- Point group identification
- Scaling

KEY STEPS WITH DIALS

dials.import

dials.find_spots datablock.json [nproc=8]

dials.index datablock.json strong.pickle [space_group=..] [unit_cell=..]

[dials.refine_bravais_settings indexed.pickle experiments.json]

dials.refine

dials.integrate

dials.export ==> CCP4i / i2 for pointless / aimless - or -

dials.symmetry / dials.scale / dials.export

KEY STEPS WITH DIALS

dials.image_viewer datablock.json [sum_images=5]

dials.reciprocal_lattice_viewer datablock.json strong.pickle

IMAGE VIEWER

File Actions

Load file Save As...

Image: C2sum_1_0001.cbf.gz [1] Previous Next Jump to image: 1

Settings

Zoom level: 25% Color scheme: grayscale

Brightness: 100

☐ Show resolution rings ☐ Show ice rings

☒ Mark beam center ☒ Mark centers of mass

☒ Spot max pixels ☒ Spot all pixels

☒ Draw reflection shoebox ☒ Show predictions

☐ Show hkl ☐ Show mask

☒ Basis vectors ☐ Indexed only

☐ Integrated only

Clear all

Sigma background 6.0

Sigma strong 3.0

Global Threshold 0.0

Min. local 2

Gain 1.0

Kernel size 3 3

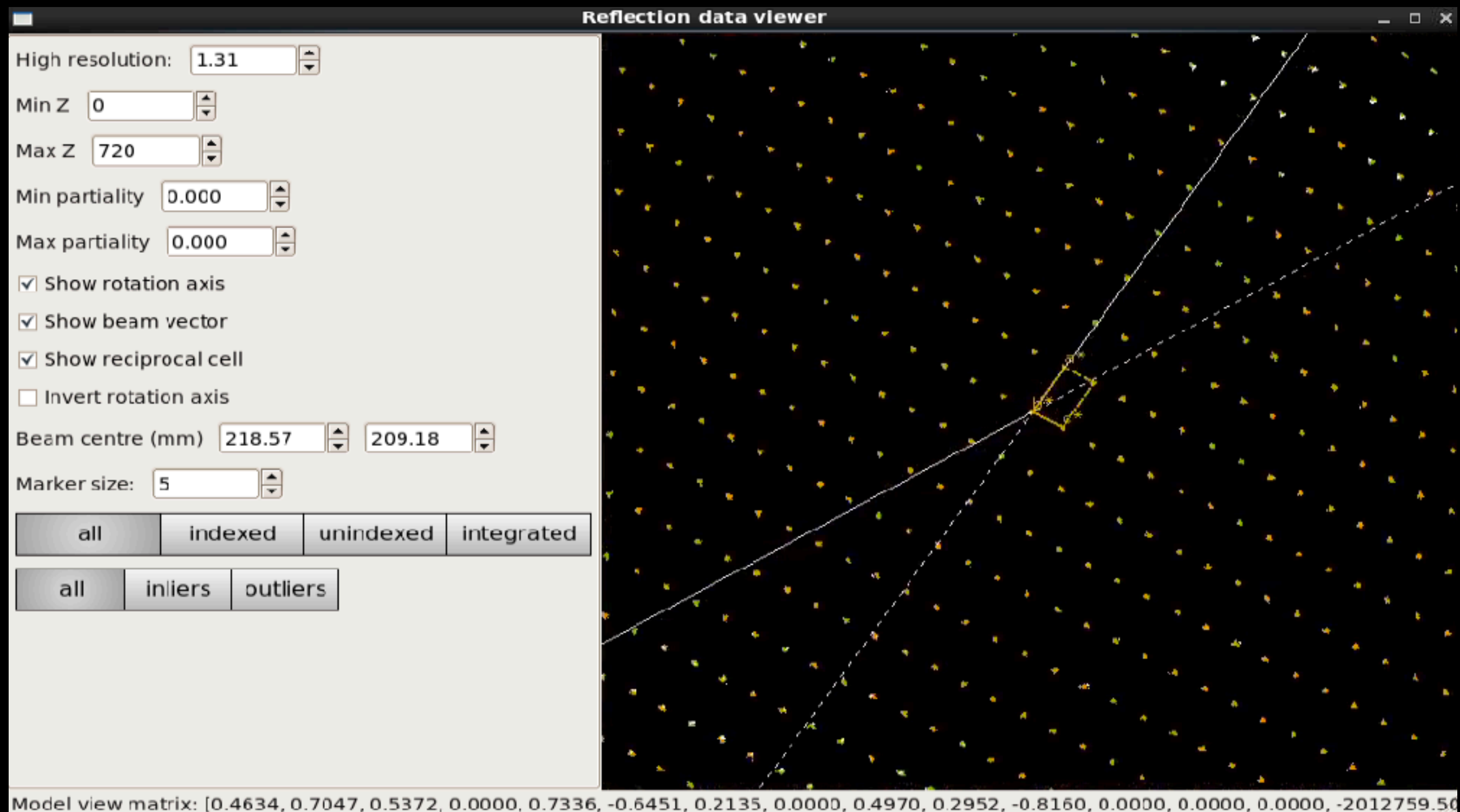
find spots.phil Save

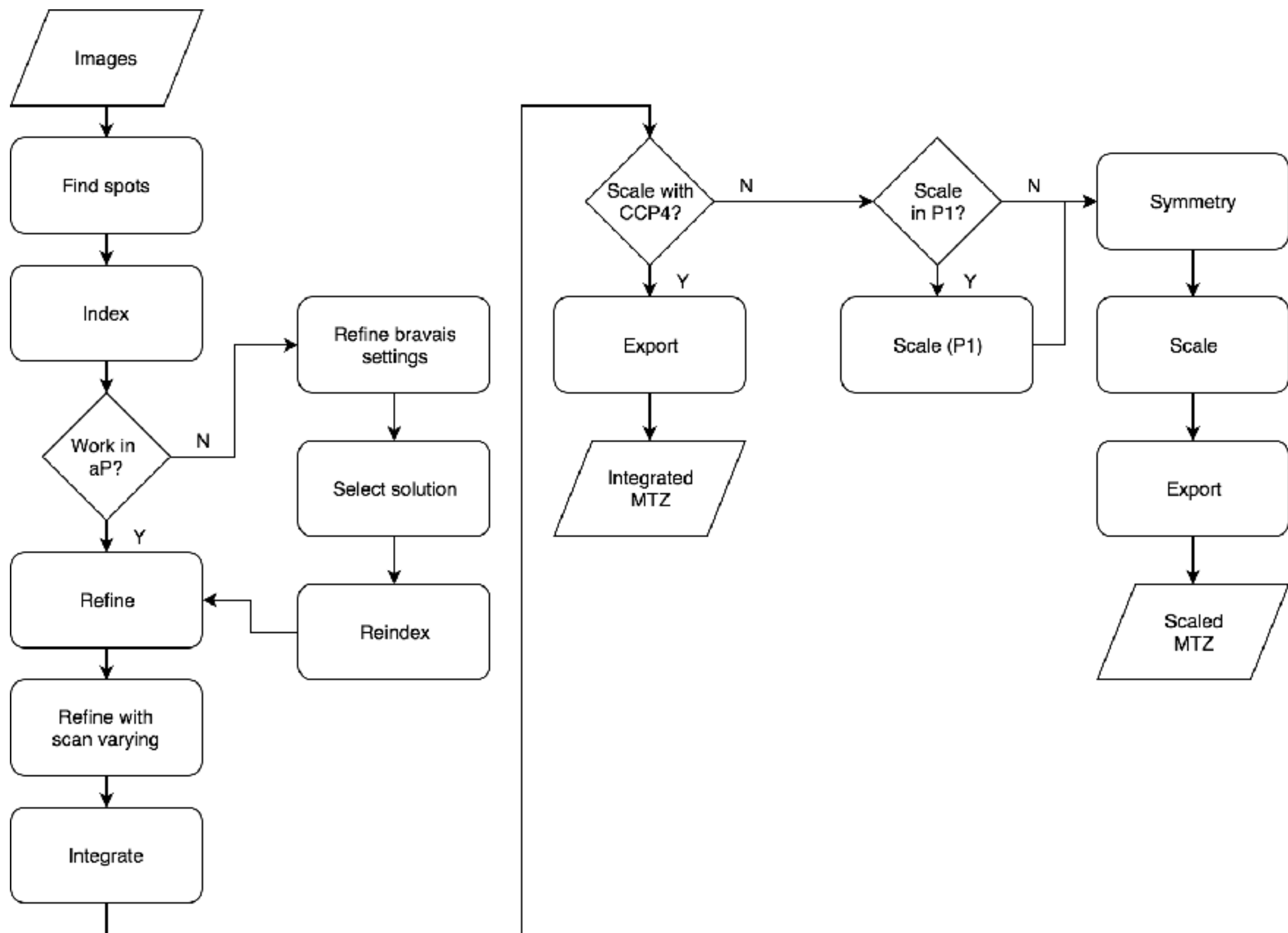
image mean variance dispersion

sigma b sigma s global threshold

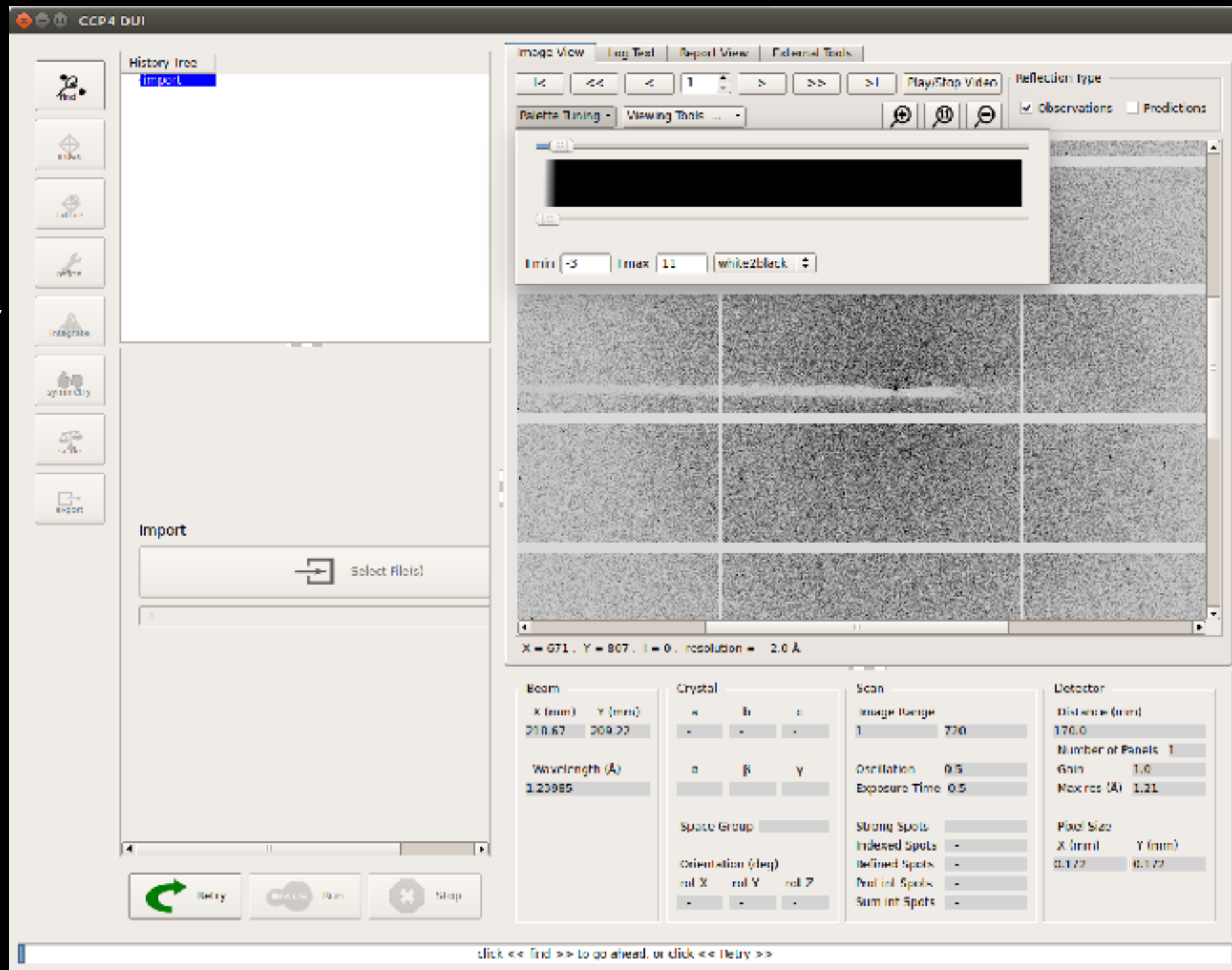
Click and drag to pan; middle-click and drag to plot intensity profile, right-click to zoom

RECIPROCAL LATTICE VIEWER





RUNNING THROUGH CCP4 DIALS USER INTERFACE (DUI)



ALGORITHMS

SPOT FINDING

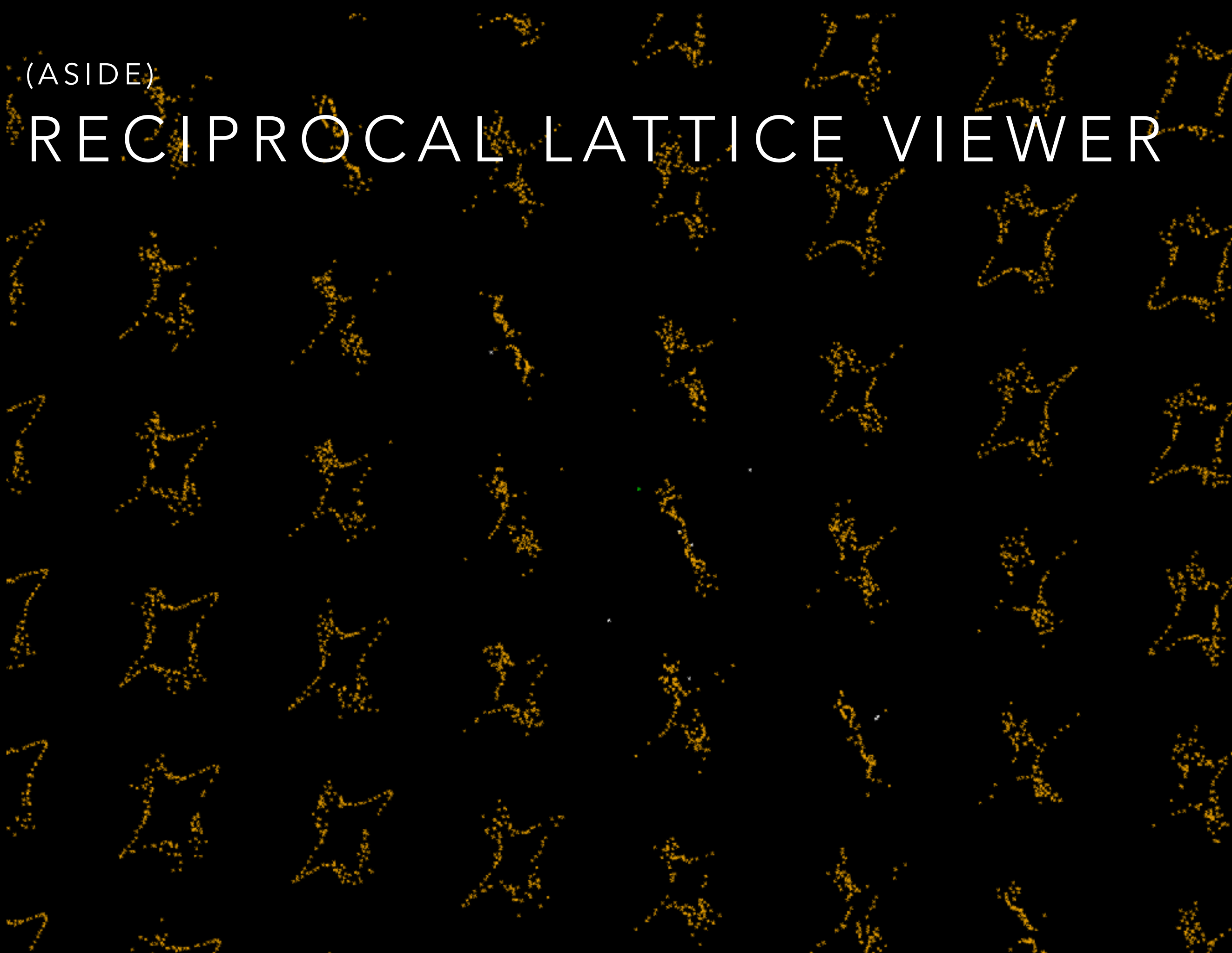


MAP TO RECIPROCAL SPACE



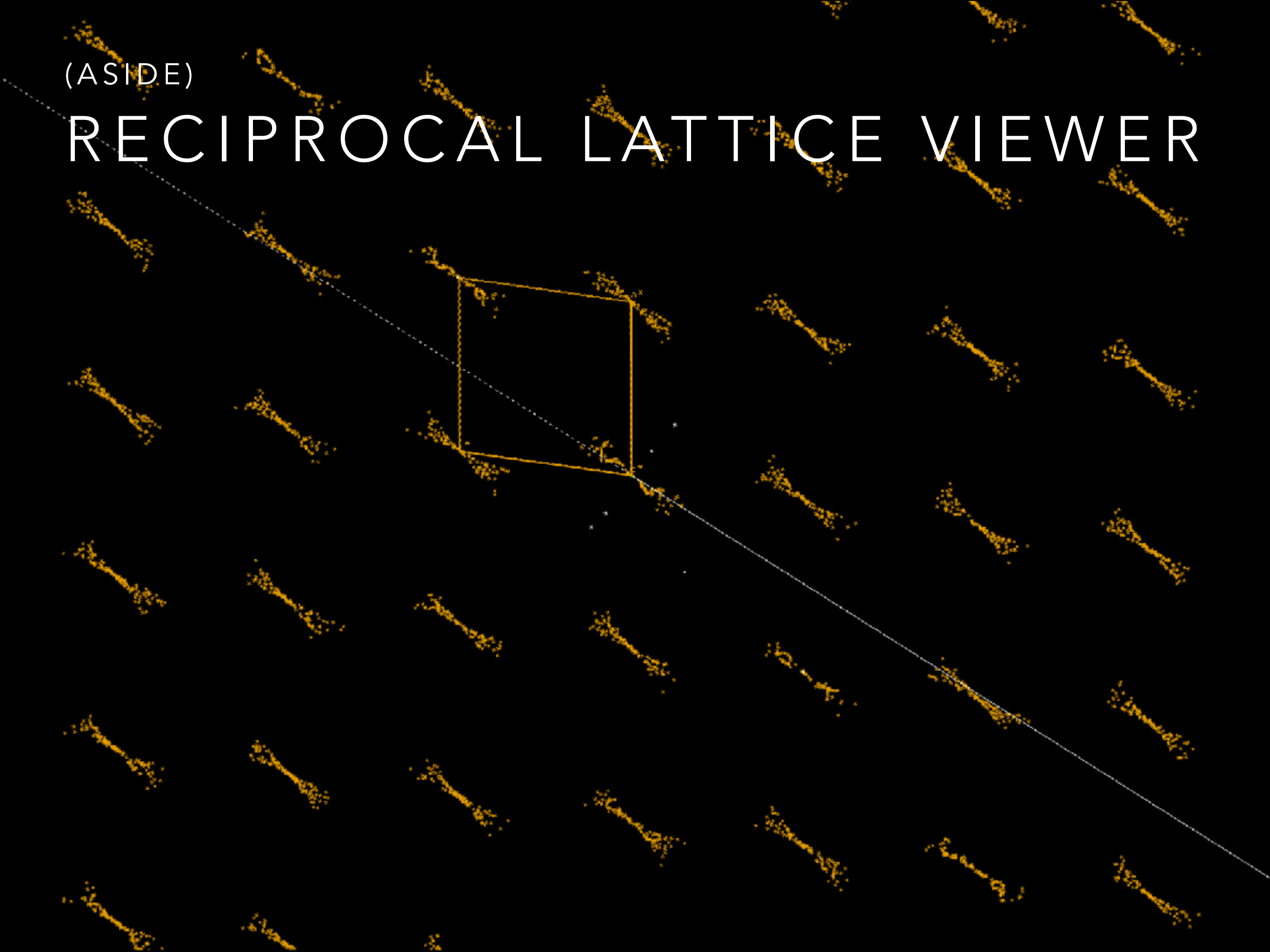
(ASIDE)

RECIPROCAL LATTICE VIEWER

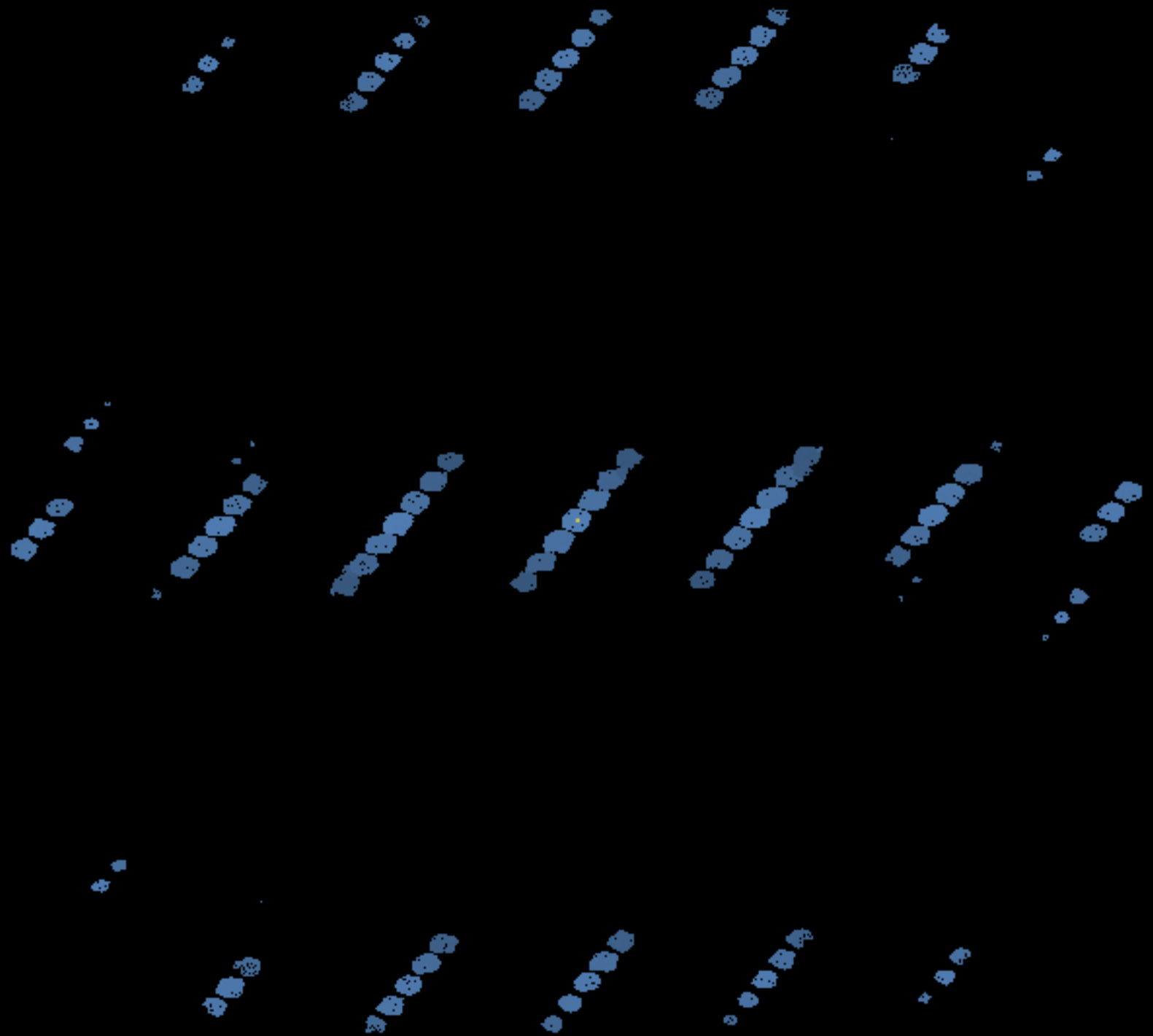


(ASIDE)

RECIPROCAL LATTICE VIEWER



INDEXING (3D FFT)



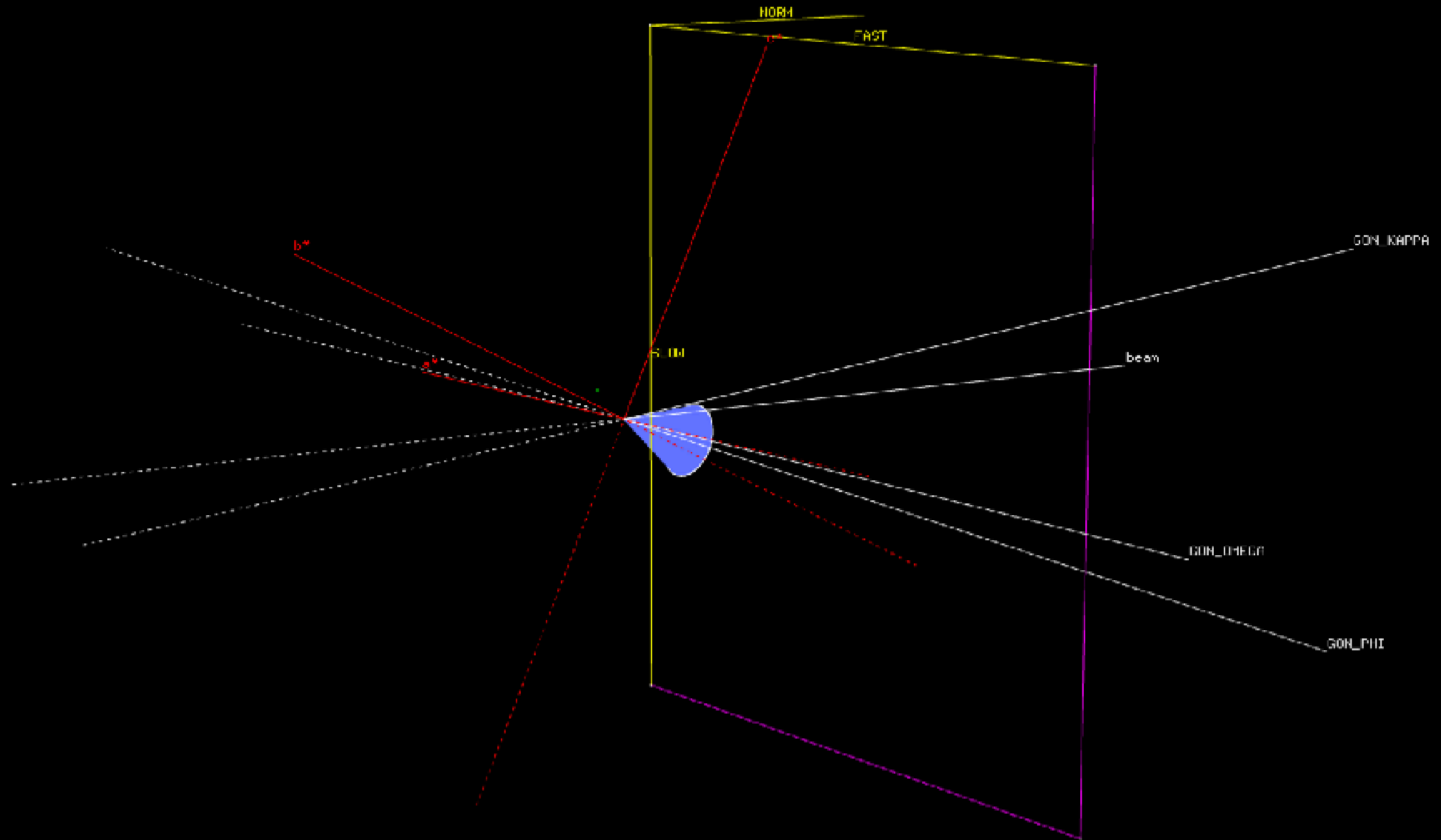
SOLUTION PICKING WITH REFINEMENT BRAVAIS SETTINGS

Solution	Metric	fit	rmsd	min/max	cc	#spots	lattice			unit_cell			volume	cb_op

5	3.0470	2.331	0.619/0.965	36000	oC	52.67	60.74	68.55	90.00	90.00	90.00	219311	a+b, -a+b, c	
4	3.0469	2.325	0.621/0.621	36000	mC	60.77	52.69	68.57	90.00	90.04	90.00	219554	a-b, a+b, c	
3	3.0470	2.346	0.619/0.619	35999	mP	40.27	68.59	40.25	90.00	98.17	90.00	110071	b, c, a	
* 2	0.0313	0.072	0.965/0.965	36000	mC	53.18	61.25	69.29	90.00	93.05	90.00	225383	a+b, -a+b, c	
* 1	0.0000	0.070		- / -	36000	aP	40.56	40.55	69.30	91.97	92.02	98.07	112701	a, b, c



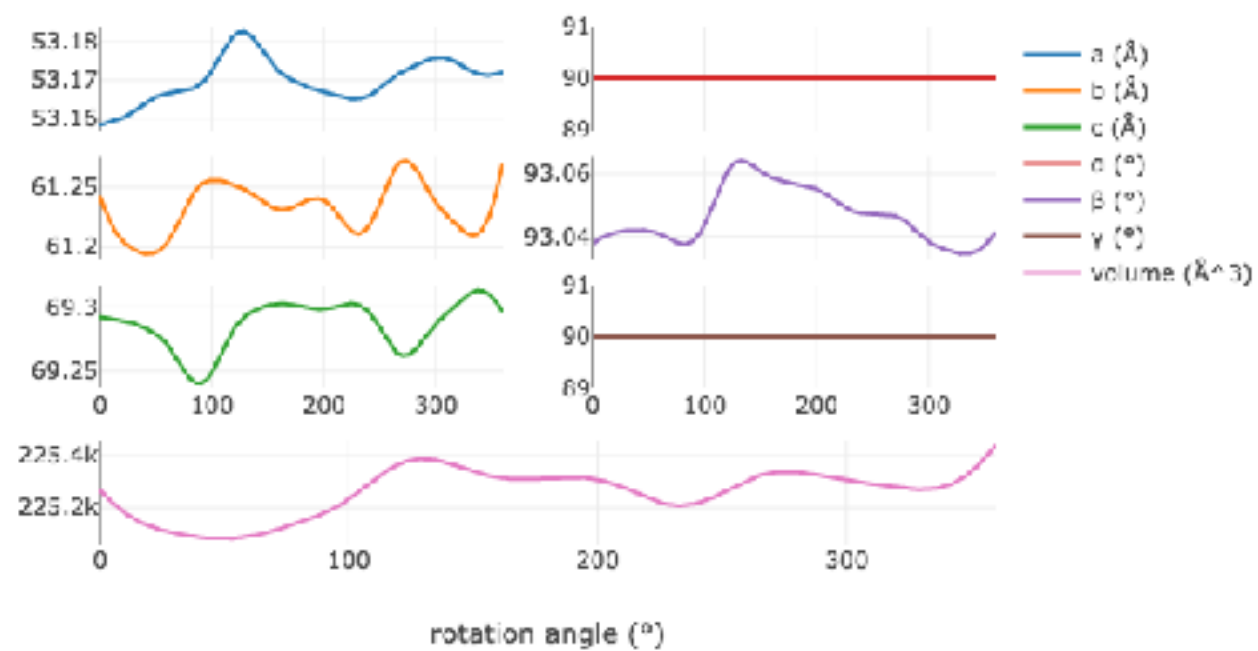
REFINEMENT



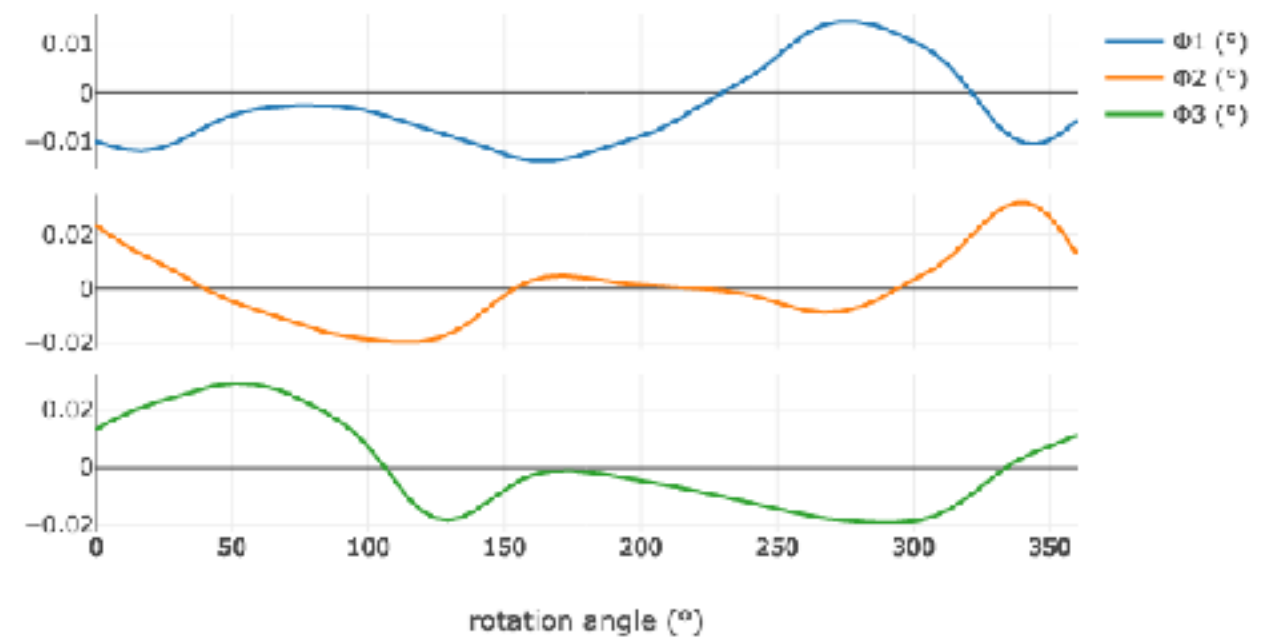
- Globally, using all reflections
- Possibility of multi-dataset joint refinement

REFINEMENT

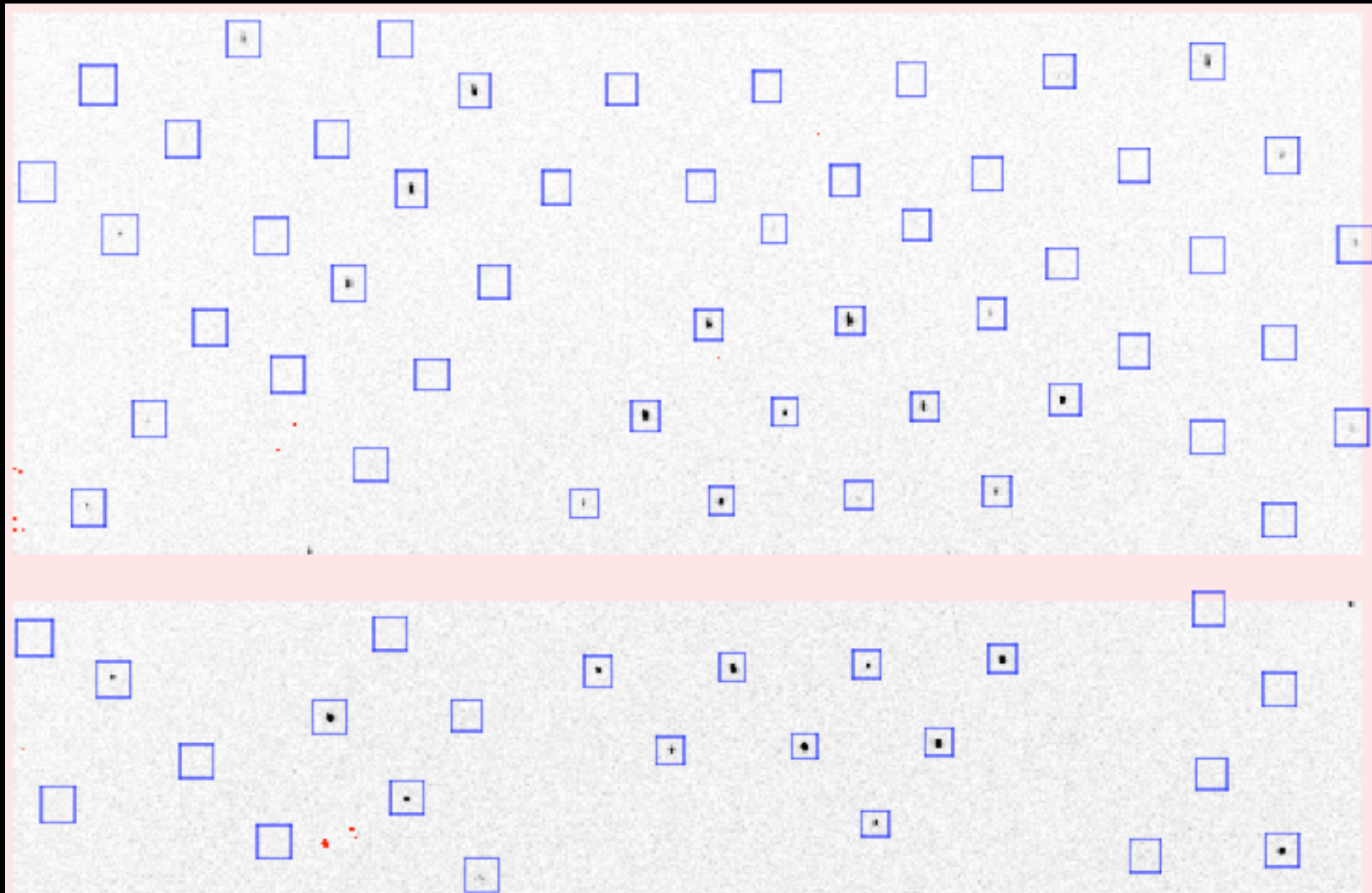
Scan-varying cell parameters



Scan-varying orientation parameters



INTEGRATION

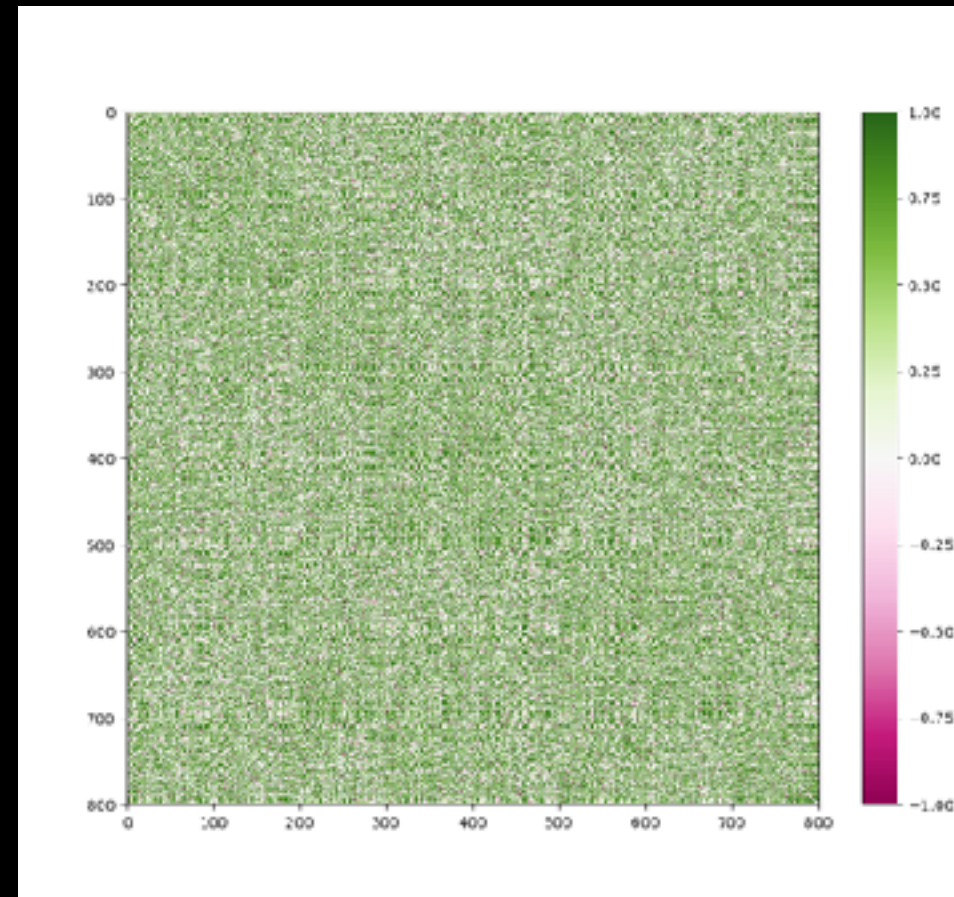
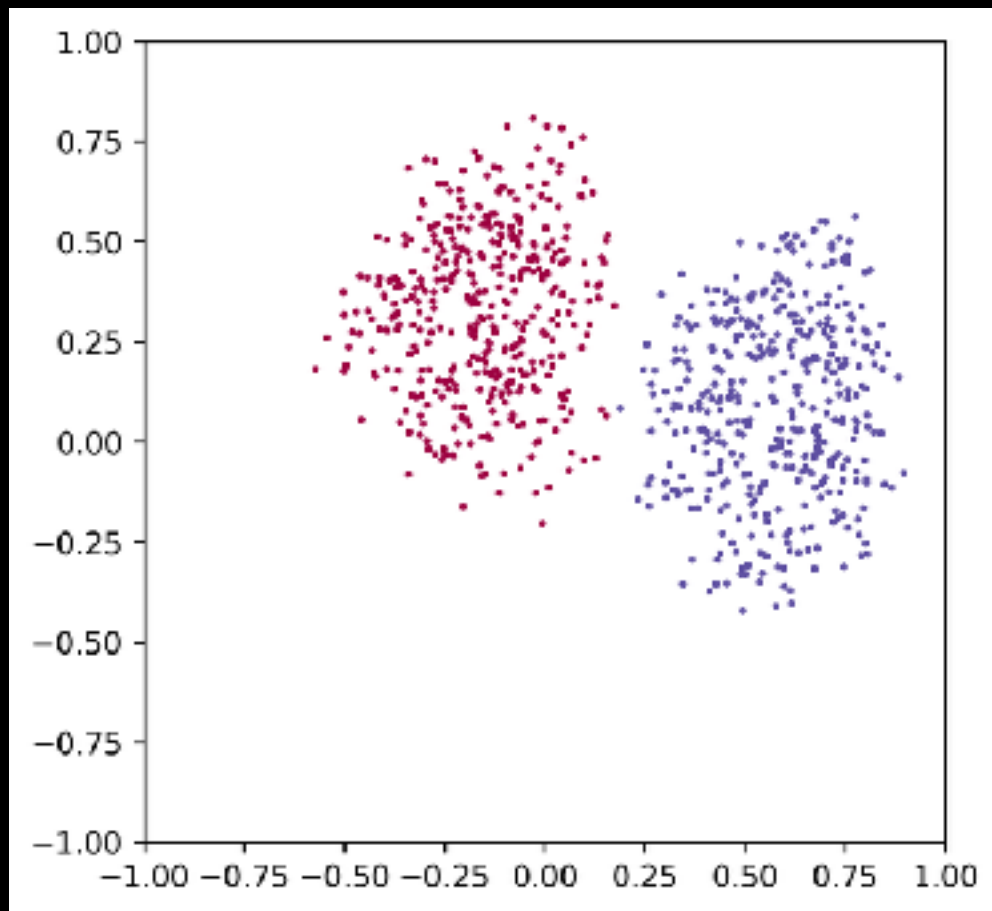


- 1) Determine profile parameters
- 2) Model background
- 3) Perform summation and profile integration

SYMMETRY & SCALING

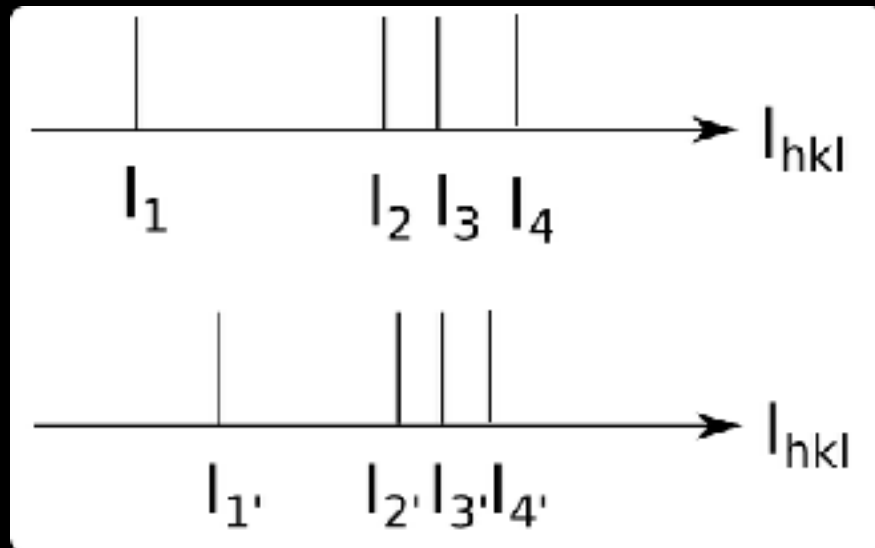
- With dials: `dials.symmetry` and `dials.scale`: new!
- With CCP4: `pointless hklin dials_integrated.mtz hklout sorted.mtz`; `aimless hklin sorted.mtz hklout scaled.mtz`

SYMMETRY & ISOMORPHISM; DIALS.COSYM DIALS.SYMMETRY

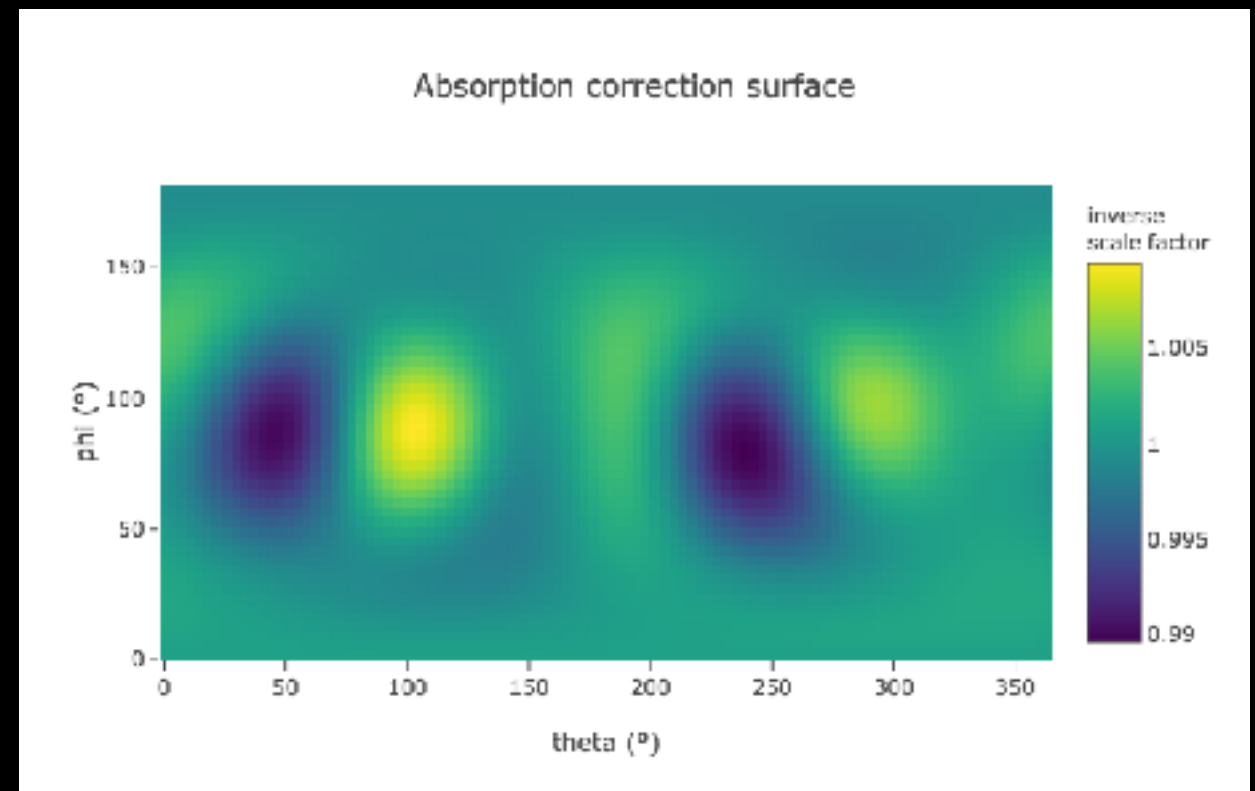
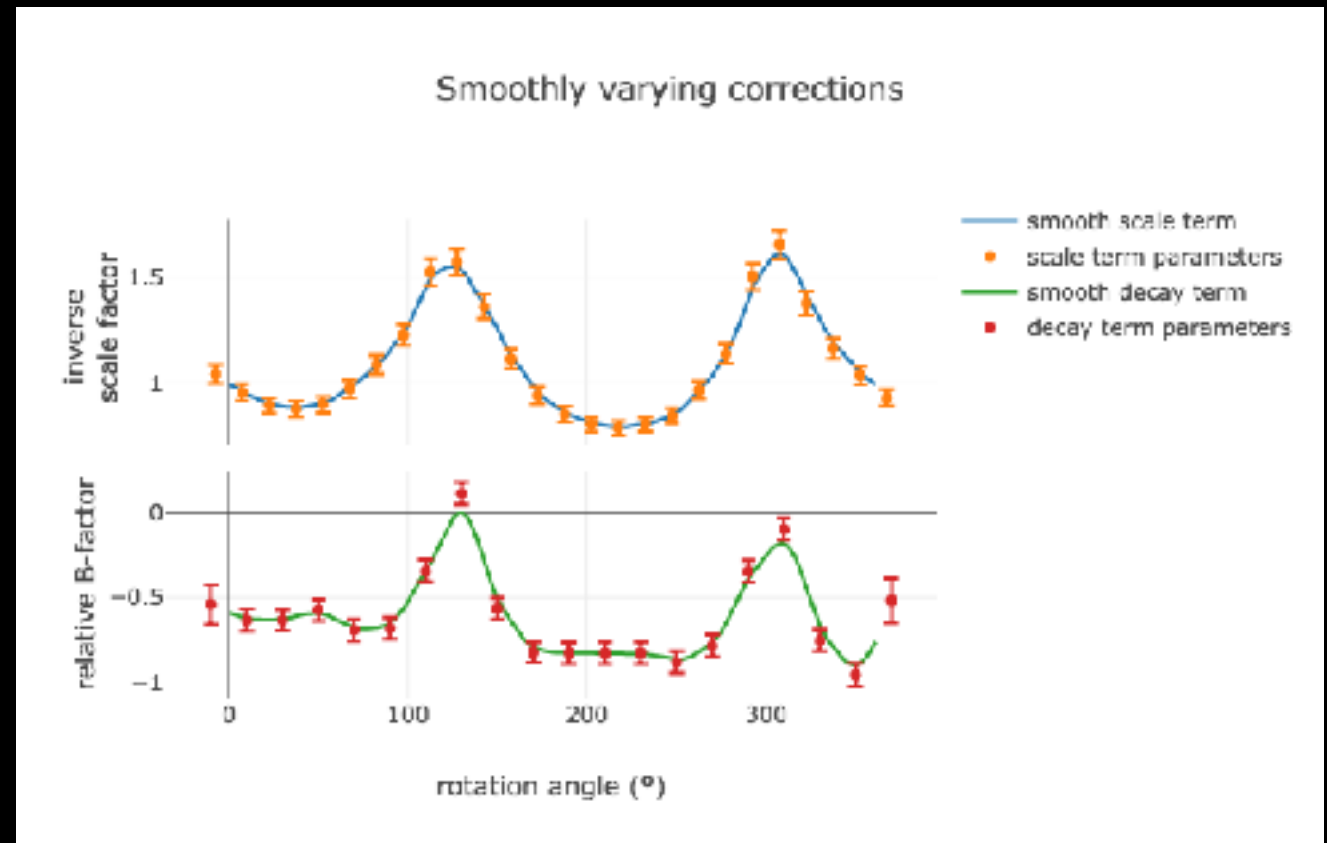


- 1) Score likelihood of symmetry operations
- 2) Combine symmetry operations into space groups

SCALING



- Do outlier rejection
- Analyse uncertainties (errors) on intensities



TUTORIAL



Diffraction Integration for Advanced Light Sources

[Home](#)

[About](#)

[Installation](#)

[Documentation](#)

[Tutorials](#)

[How-to](#)

[Workshops](#)

[Publications](#)

[Links](#)

[License](#)

Processing in Detail - CCP4 Workshop Edition

Introduction

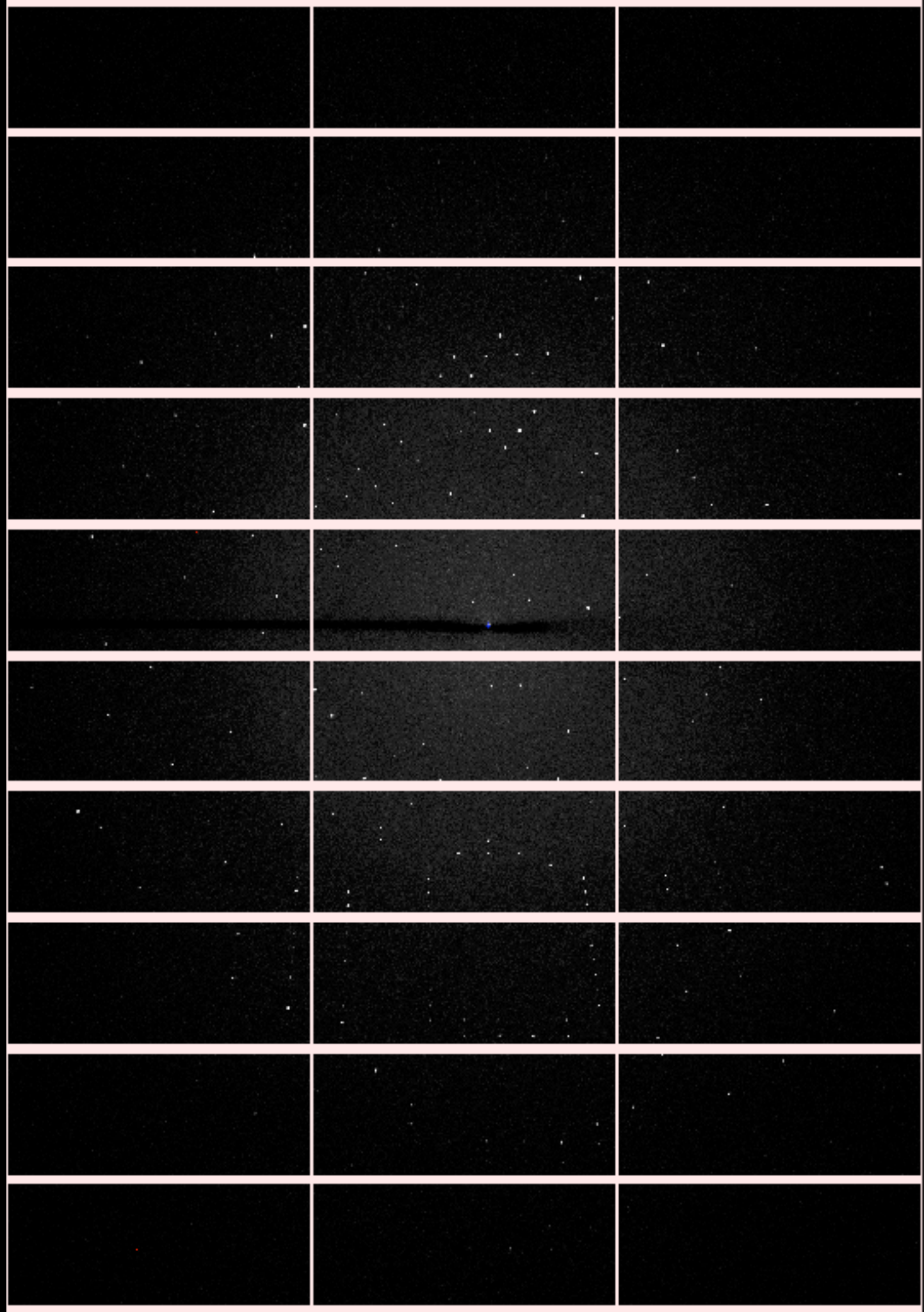
DIALS processing may be performed by either running the individual tools (spot finding, indexing, refinement, integration, exporting to MTZ) or you can run `xia2 pipeline=dials`, which makes informed choices for you at each stage. In this tutorial we will run through each of the steps in turn, checking the output as we go. We will also enforce the correct lattice symmetry.

Tutorial data

The following example uses a Beta-Lactamase dataset collected using beamline I04 at Diamond Light Source, and reprocessed especially for these tutorials. The data is available for download from [DOI 10.5281/zenodo.1014387](https://doi.org/10.5281/zenodo.1014387), but is also available for the workshop in the beamline data directory of `/dls/i03/data/2017/mx19576-1/tutorial_data/summed/`. We'll only be using the first run of data in this tutorial.

BETA LACTAMASE

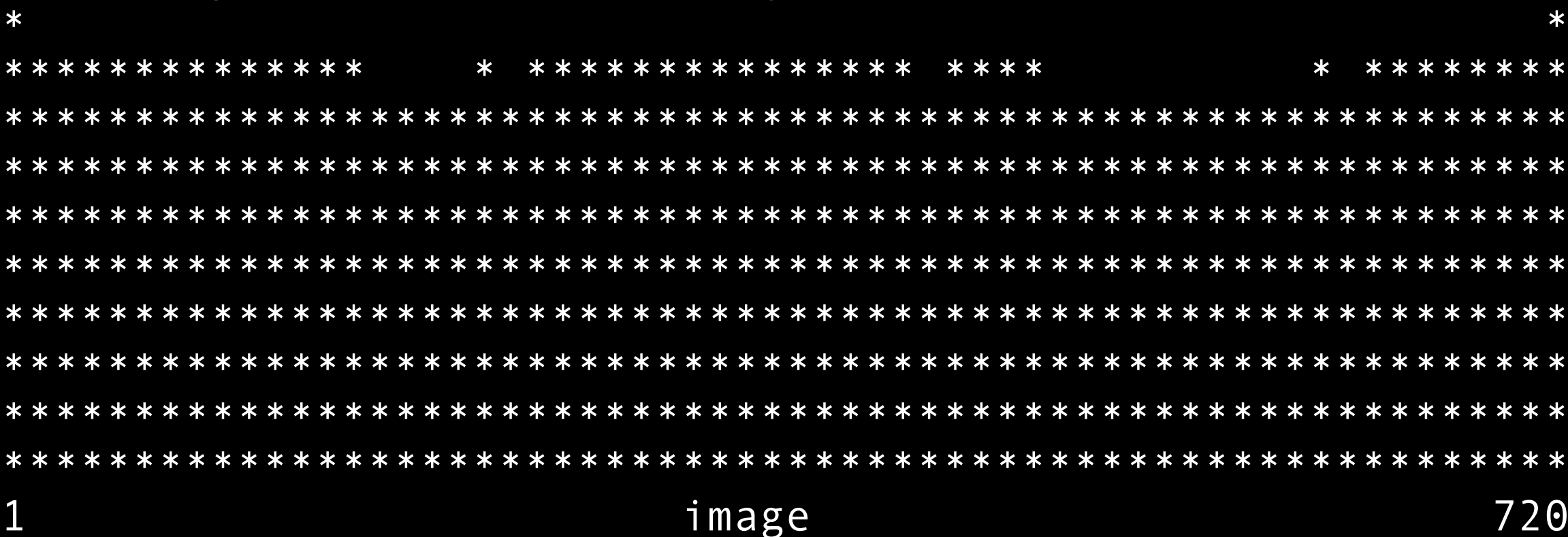
- Single sweep of data - was 3,600 images - summed to 720 x 0.5 degrees for speed
- Simple to process through to refined structure
- Good for learning how to drive DIALS

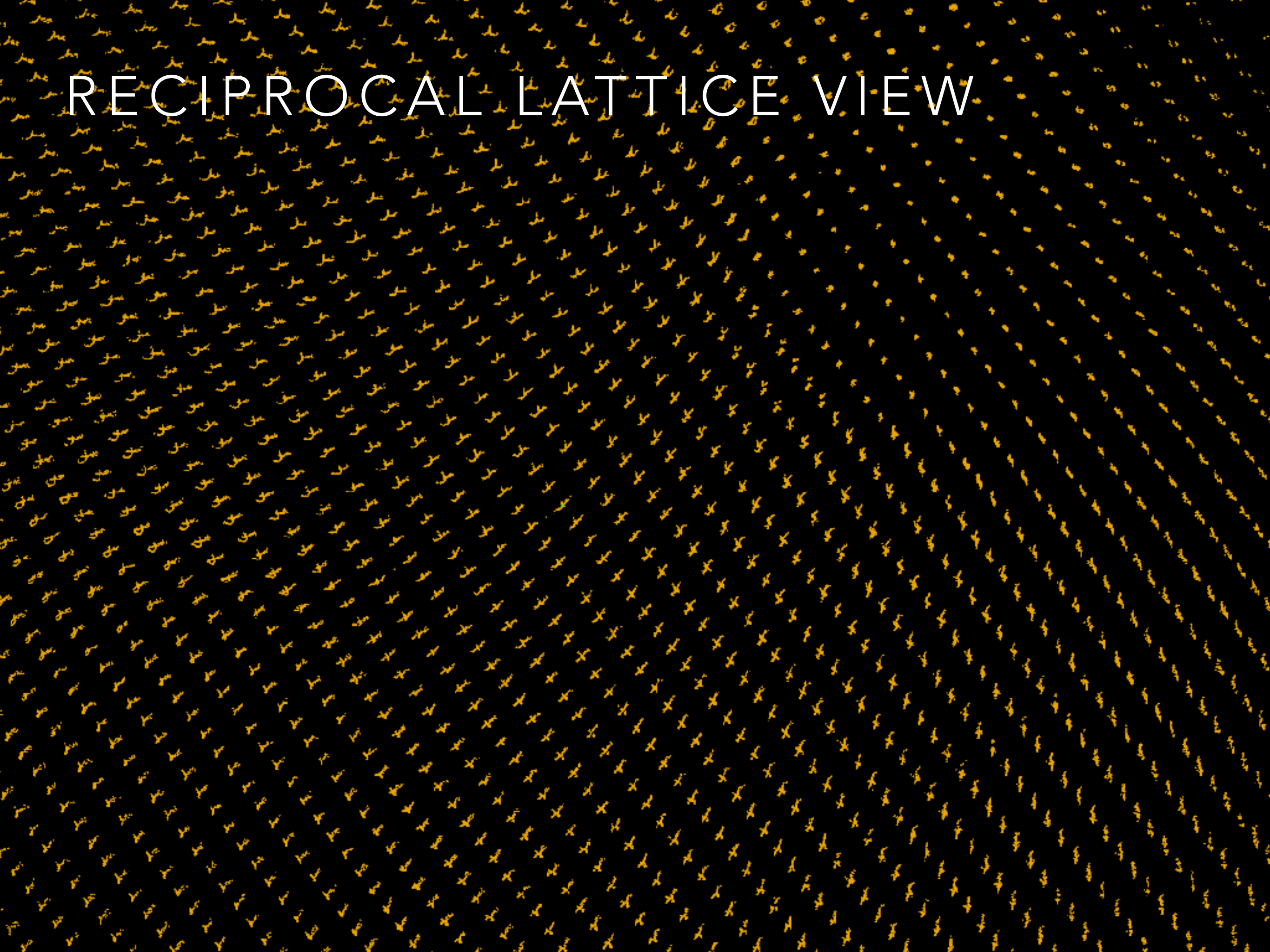


SPOT FINDING

Extracted 178377 spots
Removed 74450 spots with size < 3 pixels
Removed 13 spots with size > 100 pixels
Calculated 103914 spot centroids
Calculated 103914 spot intensities
Filtered 103780 of 103914 spots by peak-centroid distance

Histogram of per-image spot count for imageset 0:
103780 spots found on 720 images (max 1982 / bin)





RECIPROCAL LATTICE VIEW

INDEXING

Exp	Nref	RMSD_X	RMSD_Y	RMSD_Z
id		(px)	(px)	(images)
0	36000	0.28972	0.28327	0.20855

Imageset	#indexed	#unindexed
0	102151	1629



LOOK AT THE IMAGES



REFINE BRAVAIS SETTINGS

Solution	Metric	fit	rmsd	min/max	cc	#spots	lattice			

	5	3.0470	2.331	0.619/0.965	36000		oC	52.67	60.74	68
	4	3.0469	2.325	0.621/0.621	36000		mC	60.77	52.69	68
	3	3.0470	2.346	0.619/0.619	35999		mP	40.27	68.59	40
*	2	0.0313	0.072	0.965/0.965	36000		mC	53.18	61.25	69
*	1	0.0000	0.070	-/-	36000		aP	40.56	40.55	69

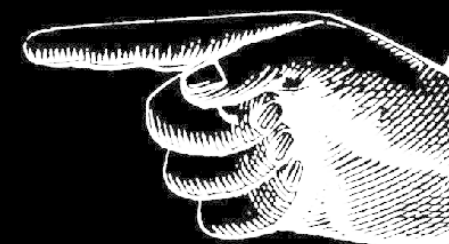
REFINEMENT (WITH SCAN VARYING)

From indexing:

Exp	Nref	RMSD_X	RMSD_Y	RMSD_Z
id		(px)	(px)	(images)
0	36000	0.28972	0.28327	0.20855

From refinement:

Exp	Nref	RMSD_X	RMSD_Y	RMSD_Z
id		(px)	(px)	(images)
0	90667	0.23769	0.23129	0.19998



INTEGRATION

Settings

Zoom level: 100%

Color scheme: grayscale

Brightness: 10

☐ Show resolution rings ☐ Show ice rings

☐ Mark beam center ☐ Mark centers of mass

☐ Spot max pixels ☐ Spot all pixels

☒ Draw reflection shoebox ☐ Show predictions

☐ Show hkl ☐ Show mask

☐ Basis vectors ☐ Indexed only

☒ Integrated only

Clear all

Sigma background 6.0

Sigma strong 3.0

Global Threshold 0.0

Min. local 2

Gain 1.0

Kernel size 3 3

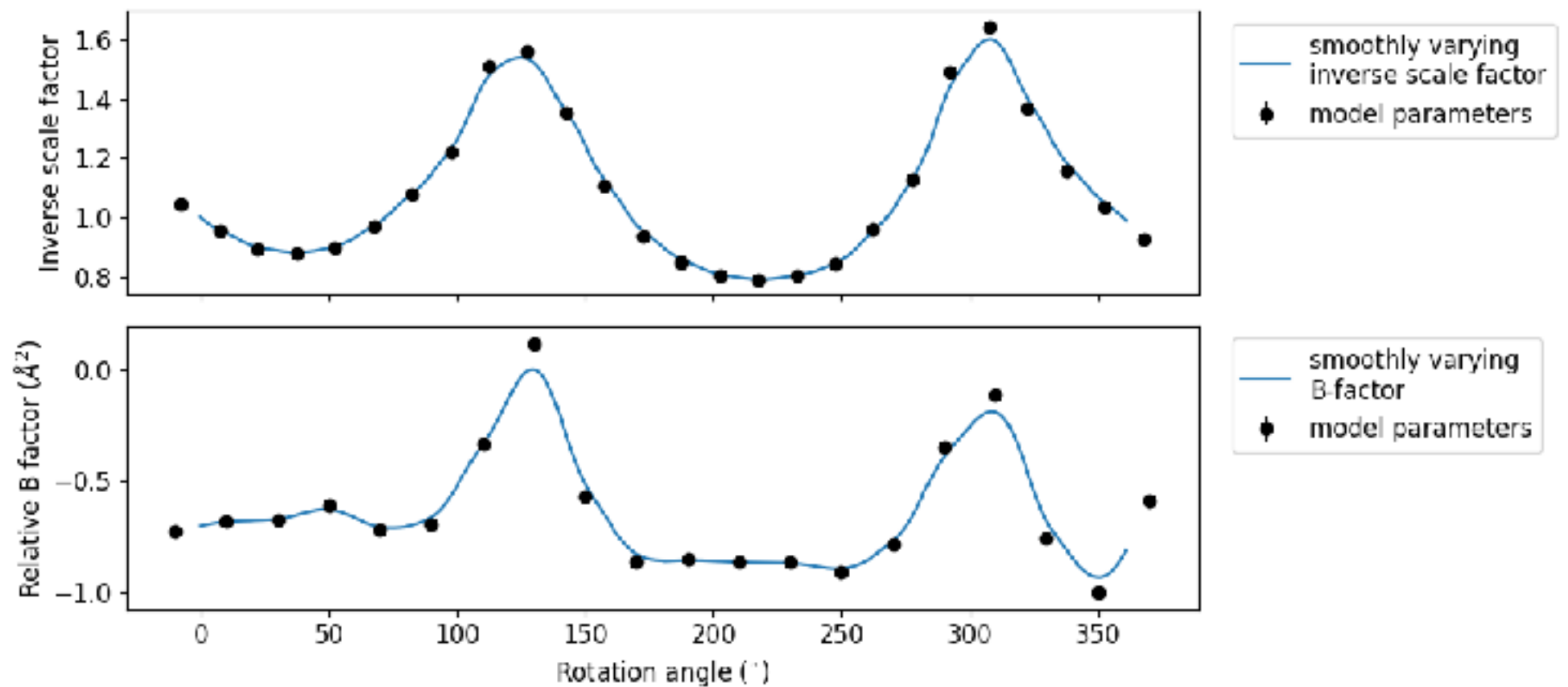
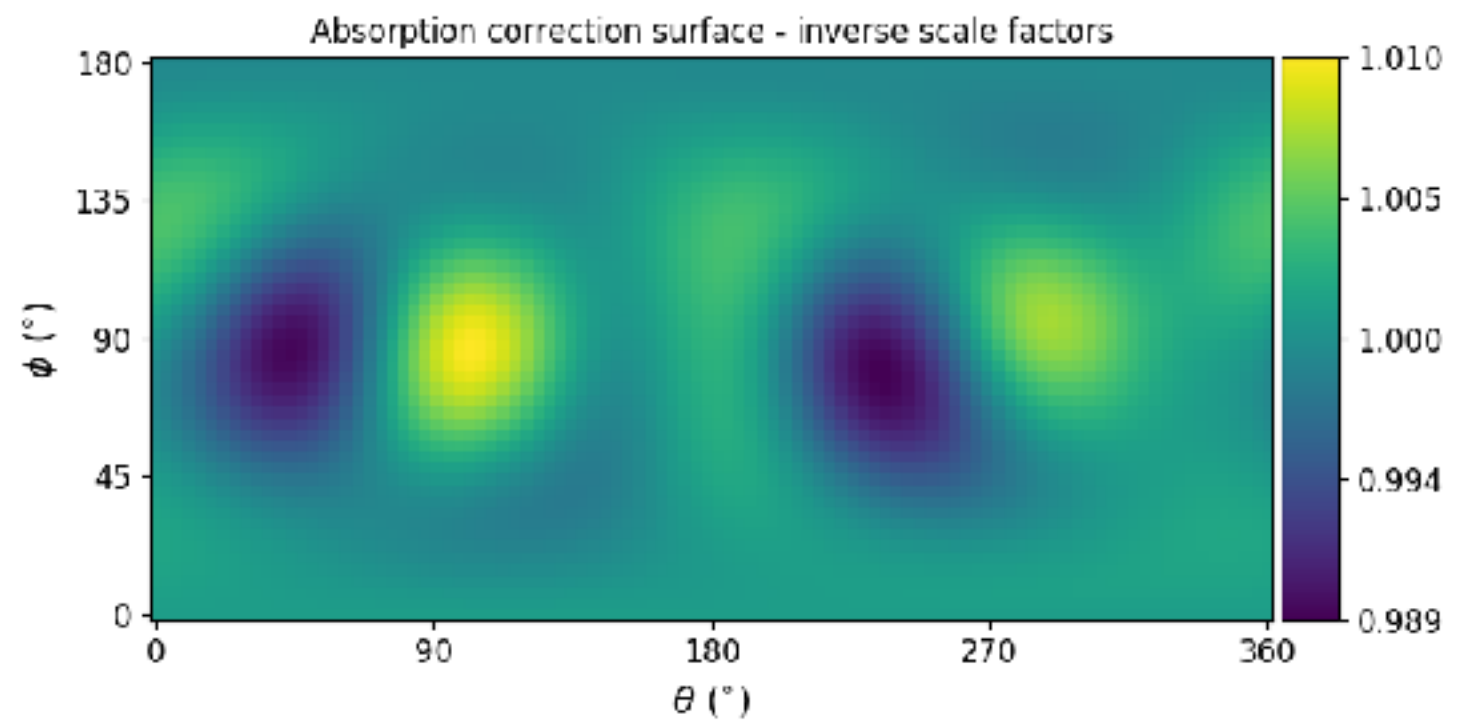
find_spots.phil

Save

image mean variance dispersion

sigma_b sigma_s global threshold

SCALING



DIALS.GITHUB.IO
XIA2.GITHUB.IO