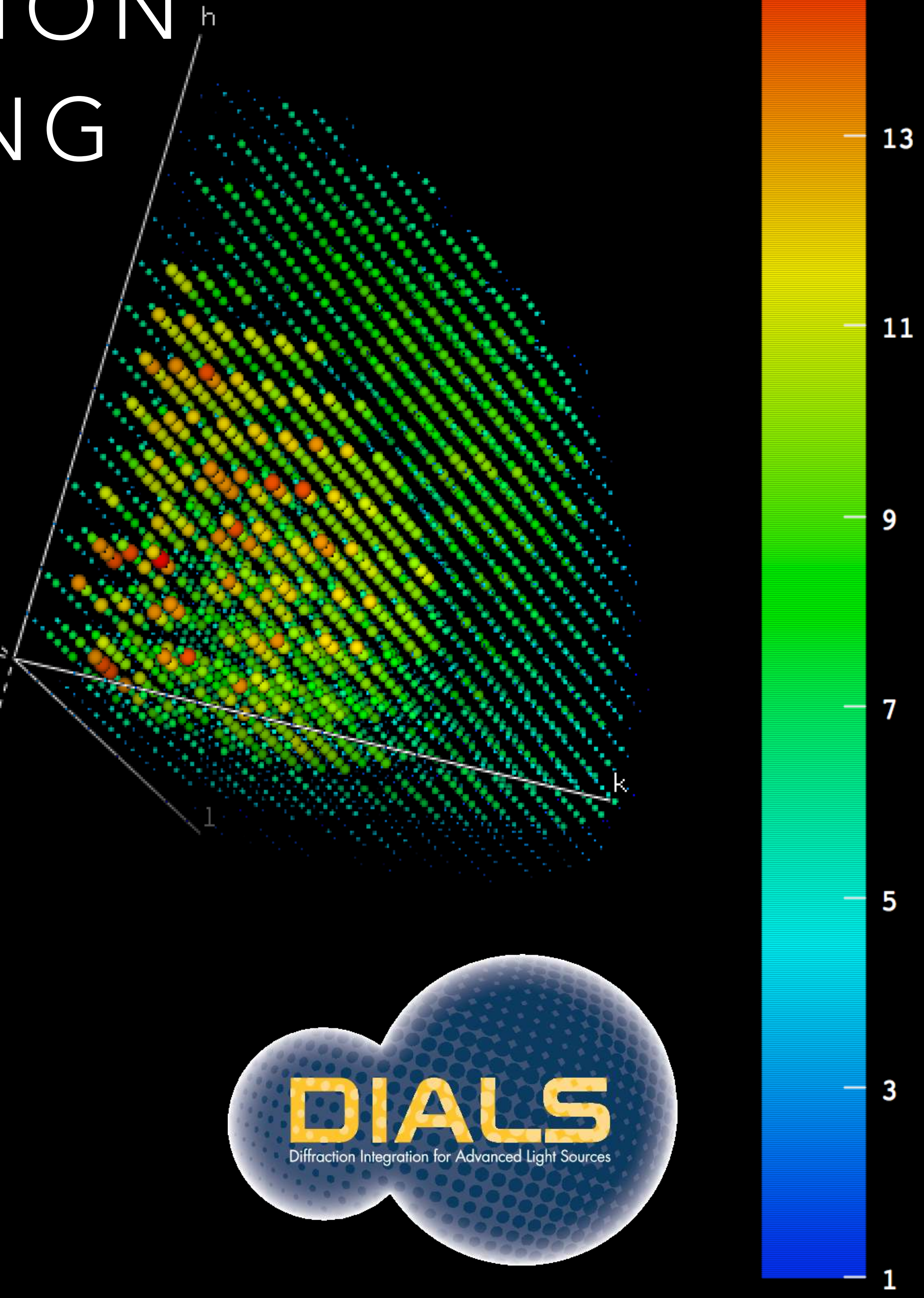
X-RAY DIFFRACTION DATA PROCESSING WITH DIALS & XIA2



DIAMOND / CCP4 WORKSHOP 2021
GRAEME WINTER
DIAMOND LIGHT SOURCE

diamond

DIALS
Diffraction Integration for Advanced Light Sources

PROJECT SUPPORT

DIALS
Diffraction Integration for Advanced Light Sources

diamond

wellcome

BERKELEY LAB
Lawrence Berkeley National Laboratory

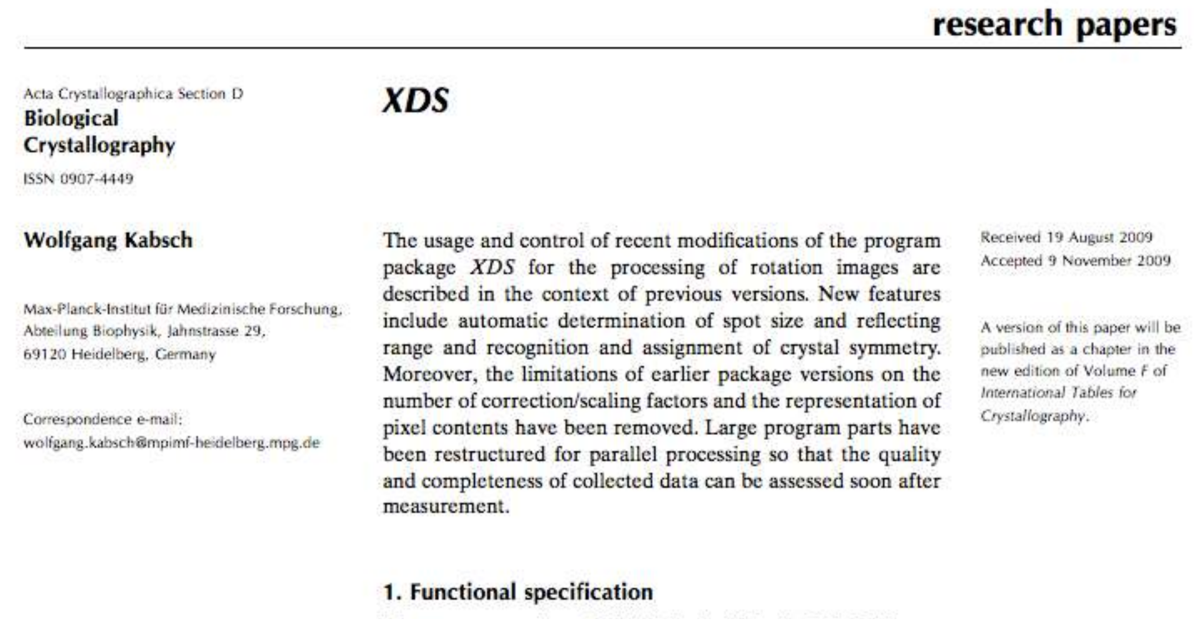
BioStructx

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

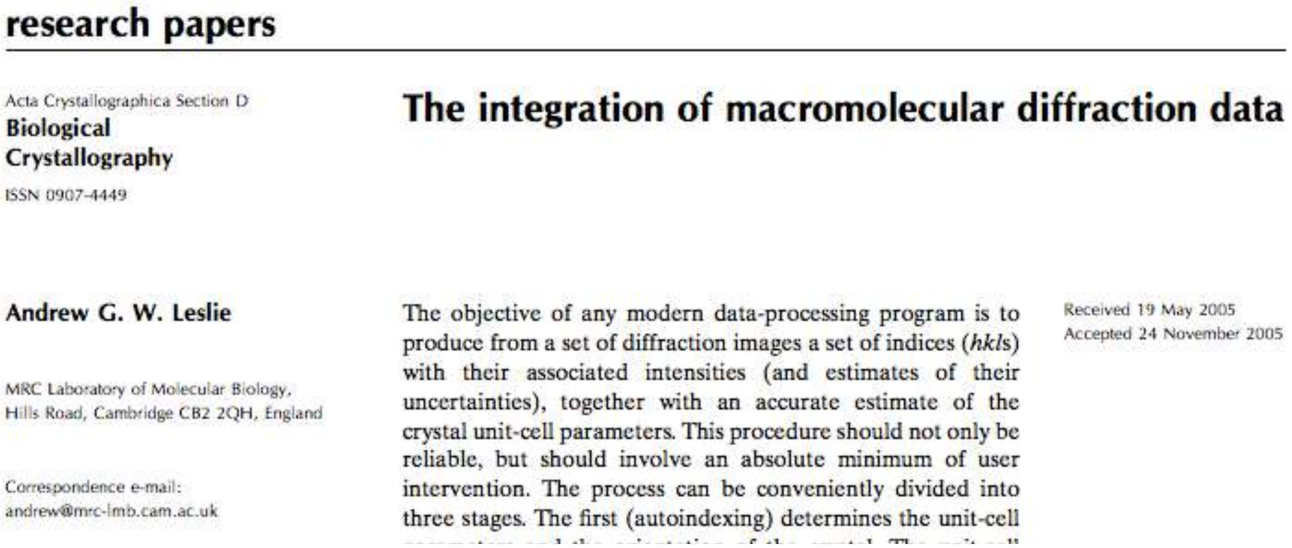
Correspondence e-mail: wolfgang.kabsch@mpimf-heidelberg.mpg.de

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 been restructured for parallel processing so that the quality and completeness of collected data can be assessed soon after measurement.

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*.

1. Functional specification



research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The integration of macromolecular diffraction data

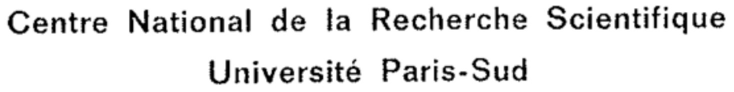
Andrew G. W. Leslie

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

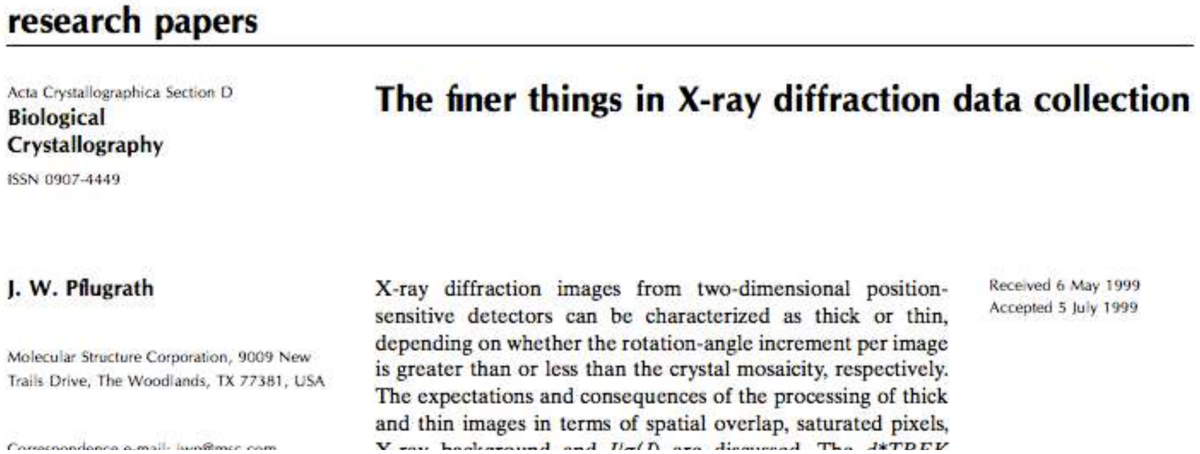
Correspondence e-mail: andrew@mrc-lmb.cam.ac.uk

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

Received 19 May 2005
Accepted 24 November 2005



Centre National de la Recherche Scientifique
Université Paris-Sud



research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

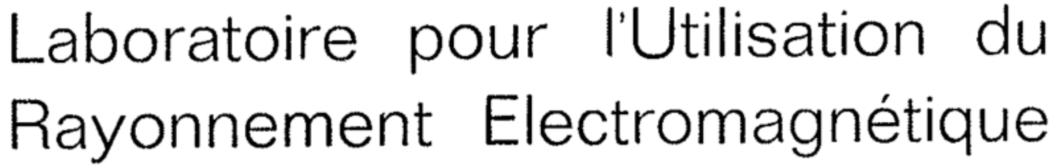
The finer things in X-ray diffraction data collection

J. W. Pflugrath

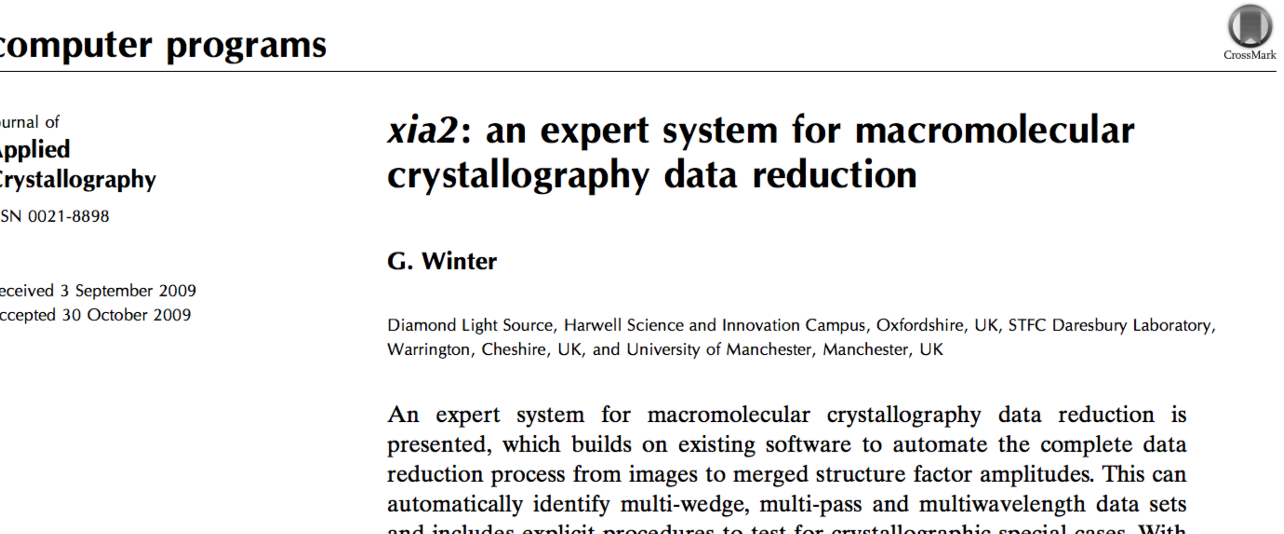
Molecular Structure Corporation, 9009 New Trails Drive, The Woodlands, TX 77381, USA

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,

Received 6 May 1999
Accepted 5 July 1999



Laboratoire pour l'Utilisation du
Rayonnement Electromagnétique



computer programs

urnal of
pplied
rystallography
SN 0021-8898

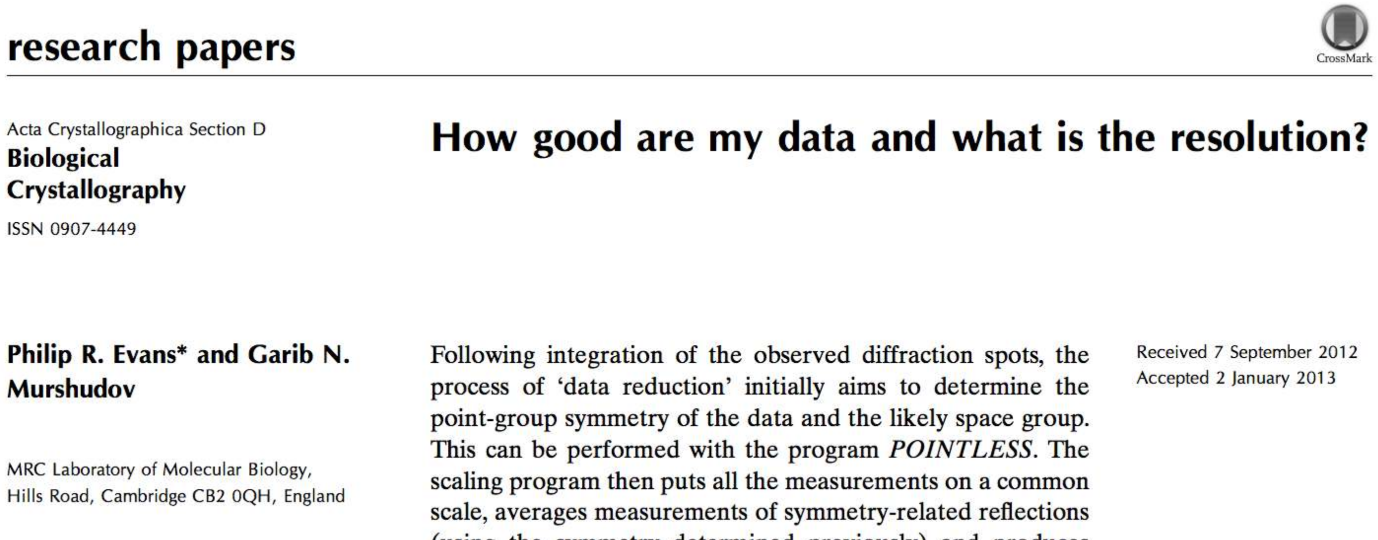
eceived 3 September 2009
ccepted 30 October 2009

xia2: an expert system for macromolecular crystallography data reduction

G. Winter

Diamond Light Source, Harwell Science and Innovation Campus, Oxfordshire, 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



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

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

Received 7 September 2012
Accepted 2 January 2013

SOLUTION TO MOST PROCESSING PROBLEMS - COLLECT BETTER DATA!

- Poorly performed experiments are the biggest problem with data processing - radiation damage etc.
- With a pixel array detector (pilatus, eiger) no readout noise so be sparing with photons (Elspeth yesterday)
- Collect more, weaker data in preference - lousy R_{merge} but great R_{pim} (Kay's presentation shortly)
- Mount your crystals carefully - make sure most of the atoms in the beam are your crystal not loop, drop, … and cryo-cool your samples carefully

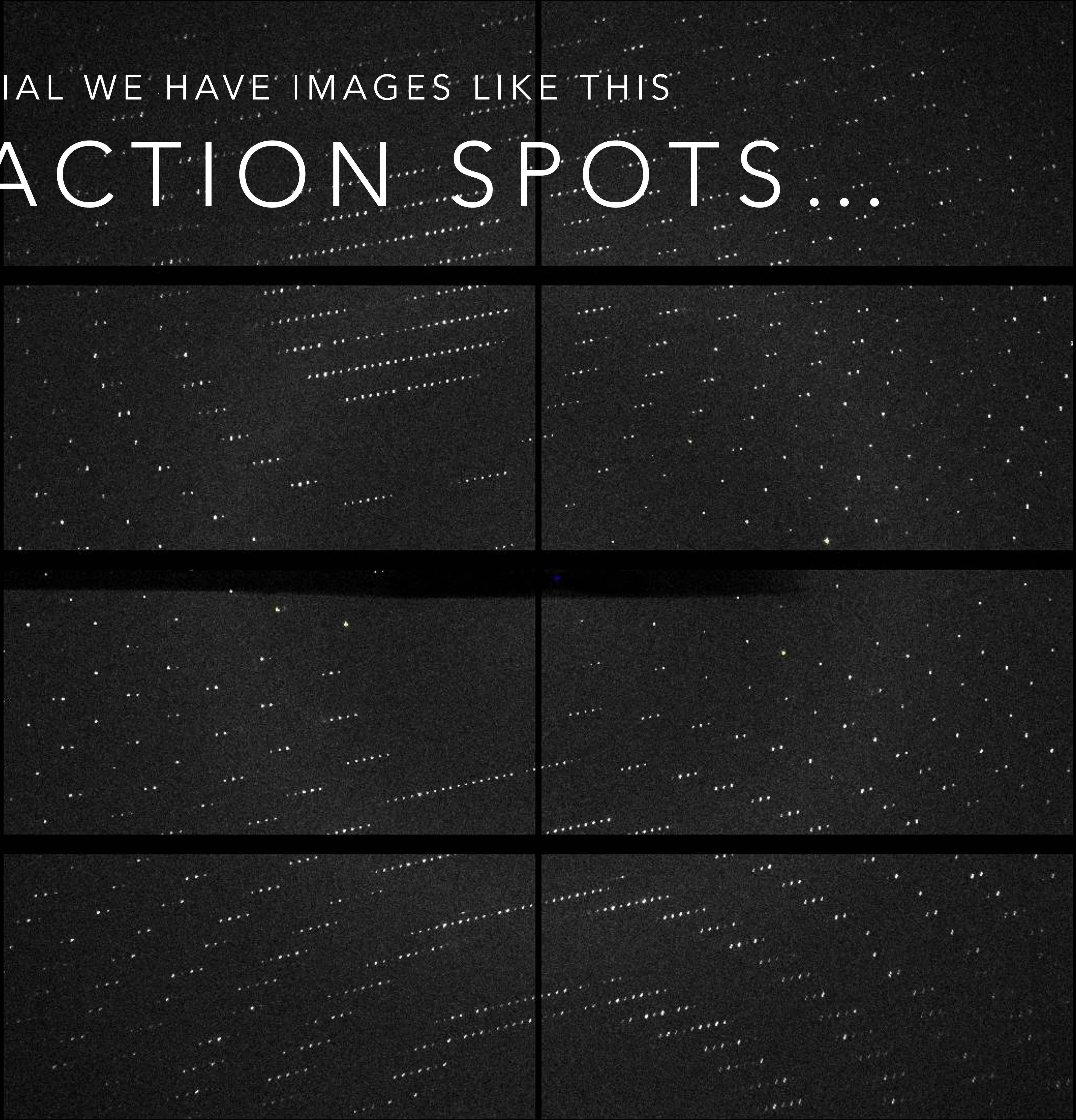
RUNNING DIALS

- Tutorials will have running individual DIALS steps because this is good to know
- In reality I run DIALS with xia2
- You can also run XDS through xia2
- You can also use DUI for interactive processing

DATA INTEGRATION AND SCALING (IN SIMPLE TERMS)

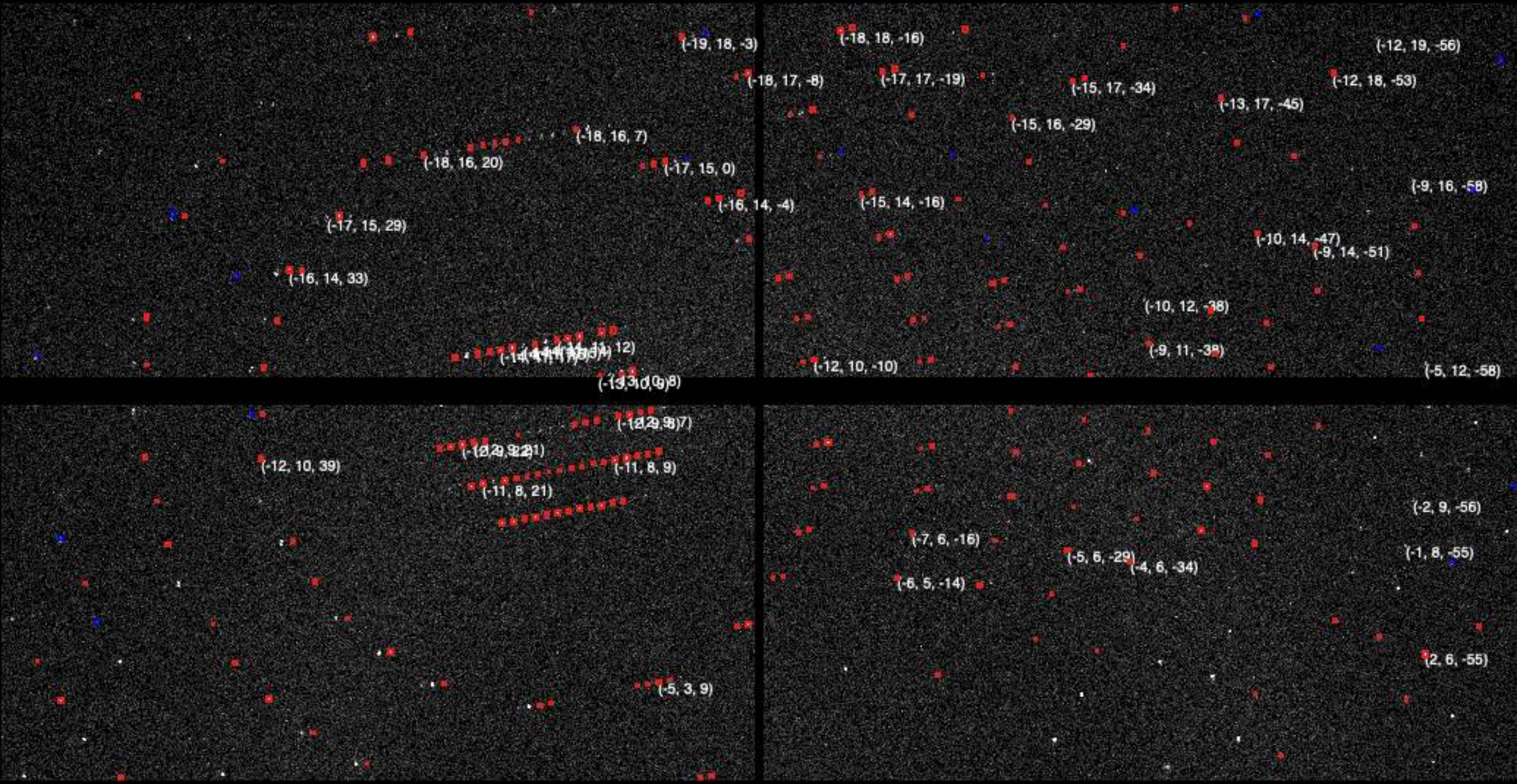
FOR THE TUTORIAL WE HAVE IMAGES LIKE THIS

DIFFRACTION SPOTS...

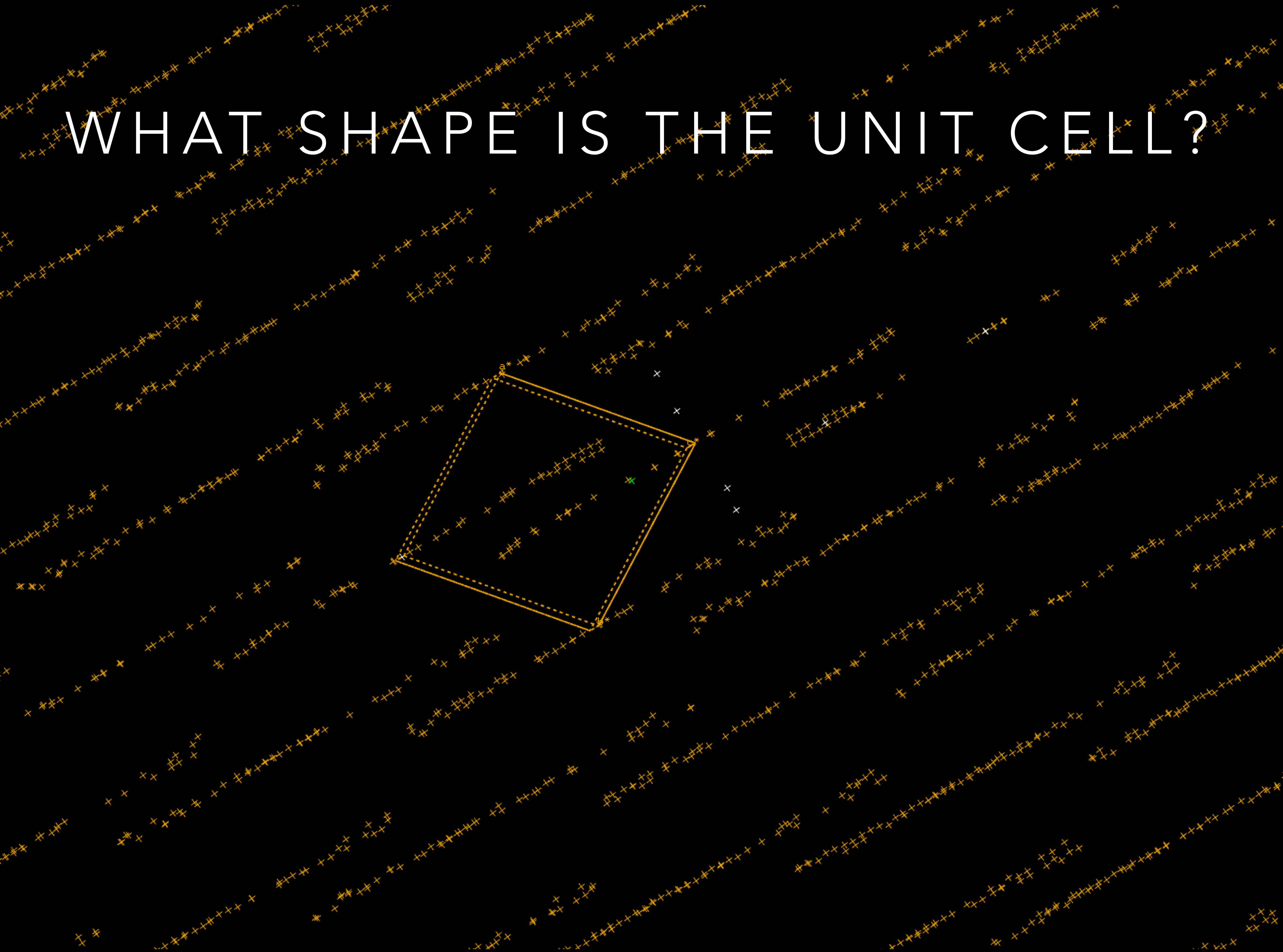


NAME THAT SPOT (MILLER INDICES)

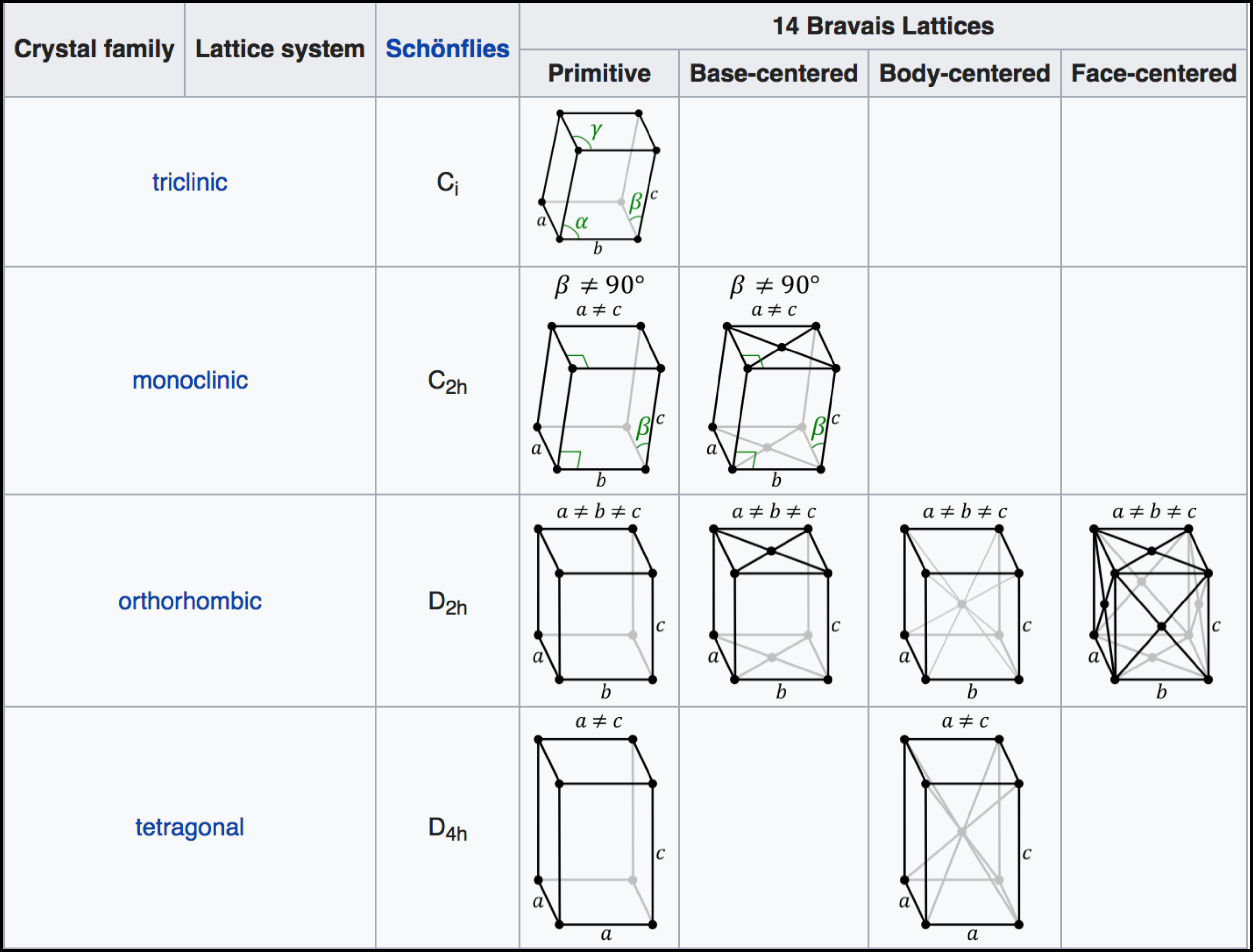
WHICH SPOT IS WHICH?



WHAT SHAPE IS THE UNIT CELL?



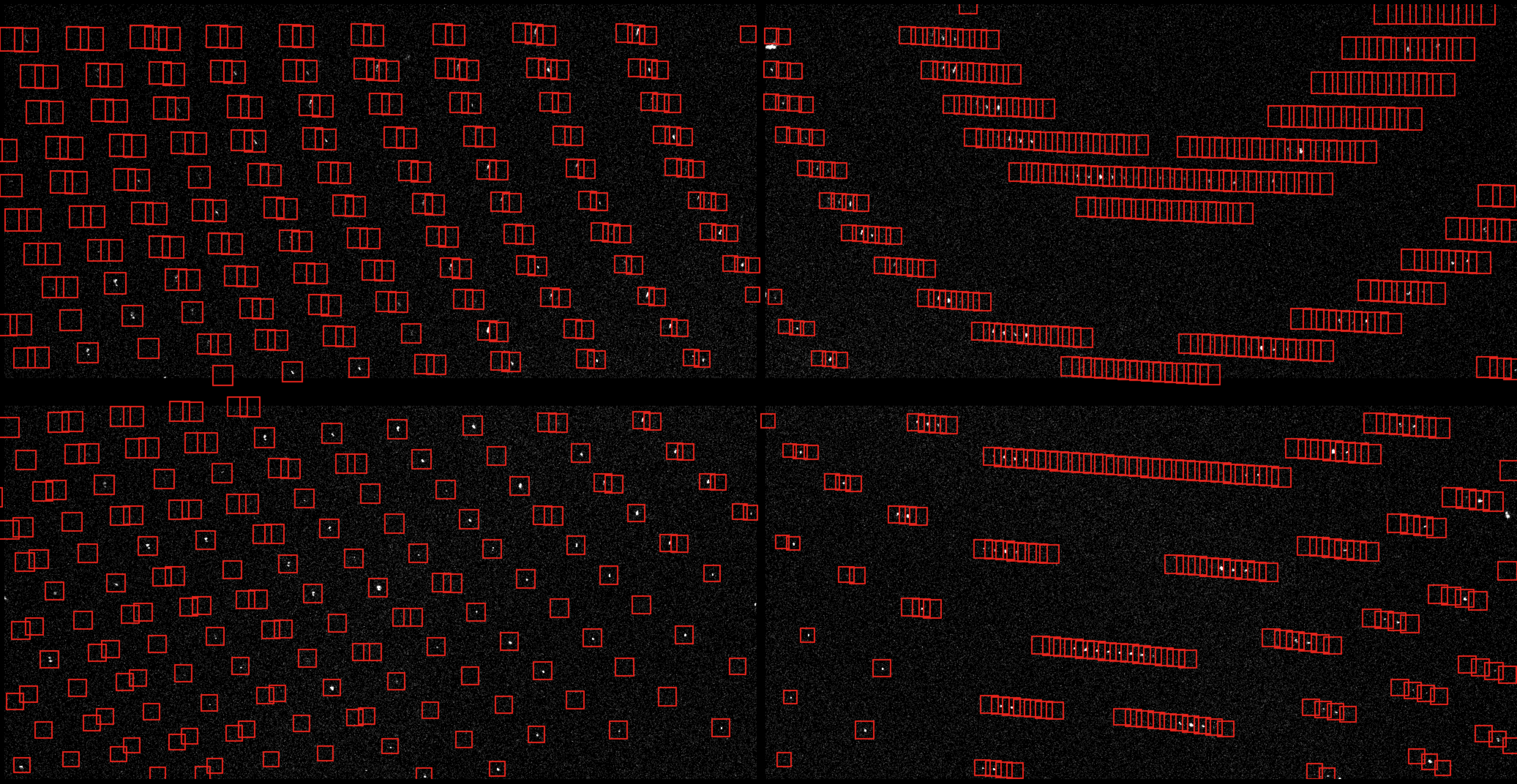
WHAT IS THE CORRECT BRAVAIS LATTICE?



Crystal family	Lattice system	Schönflies	14 Bravais Lattices			
			Primitive	Base-centered	Body-centered	Face-centered
triclinic		C_i				
monoclinic		C_{2h}	$\beta \neq 90°$ $a \neq c$	$\beta \neq 90°$ $a \neq c$		
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}	$a \neq c$		$a \neq c$	

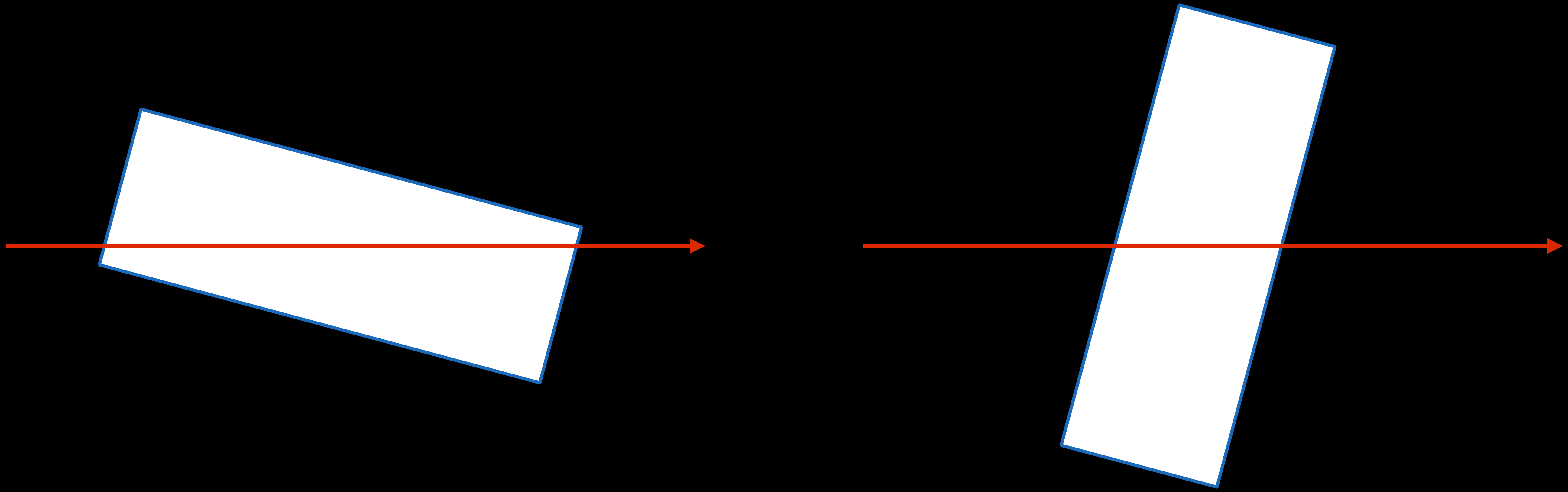
DATA INTEGRATION

HOW MANY COUNTS?



CORRECTIONS FOR EXPERIMENTAL EFFECTS

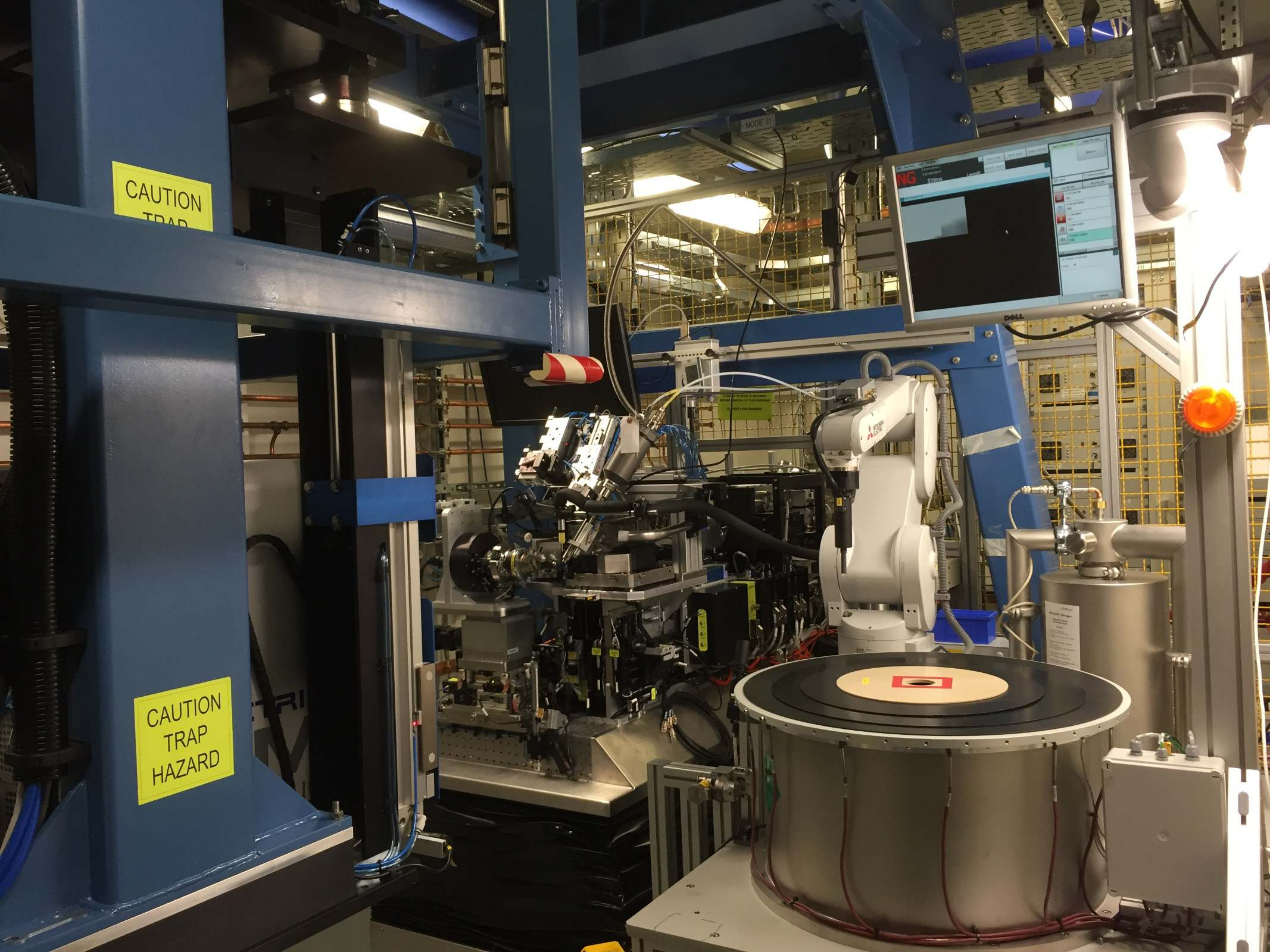
PLACE ON COMMON SCALE

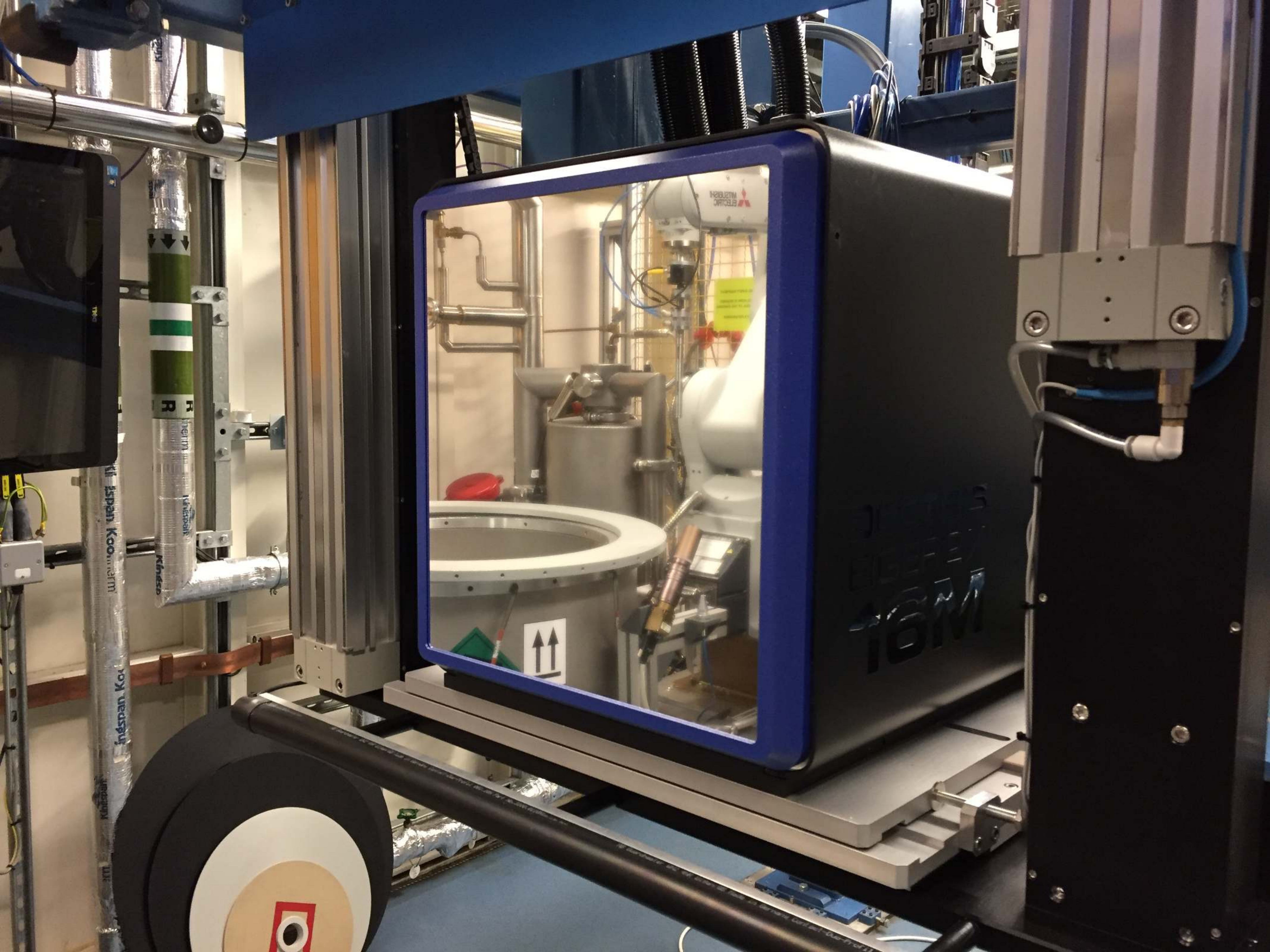


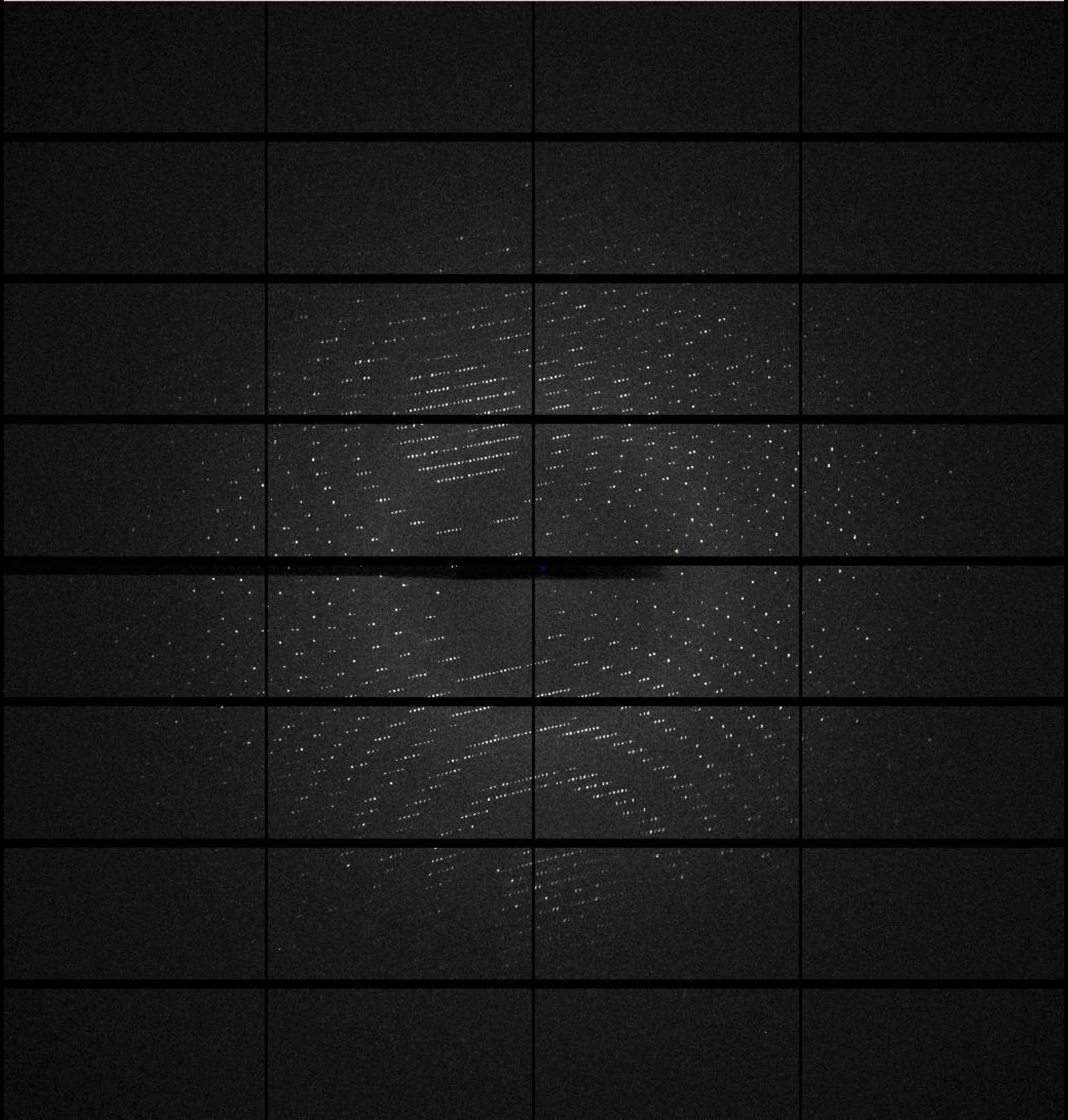
TERMINOLOGY (UNIVERSAL)

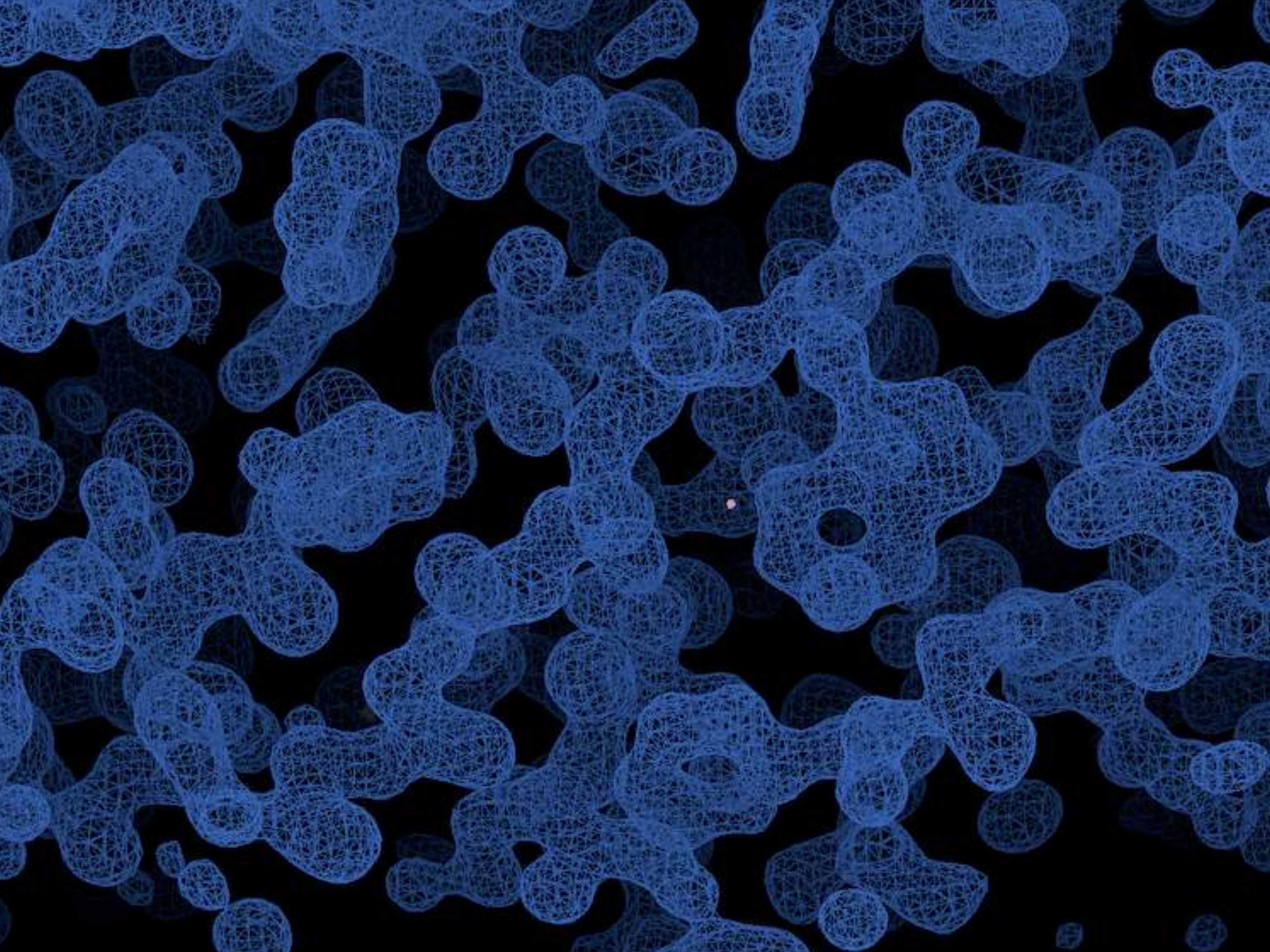
- Spot finding: locate spots on some or all of the data
- Indexing: looking at the positions of the spots and calculating a model which explains those positions (may involve choices about lattice or space group, *optional*)
- Refinement: improving the model
- Integration: record how many counts in each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

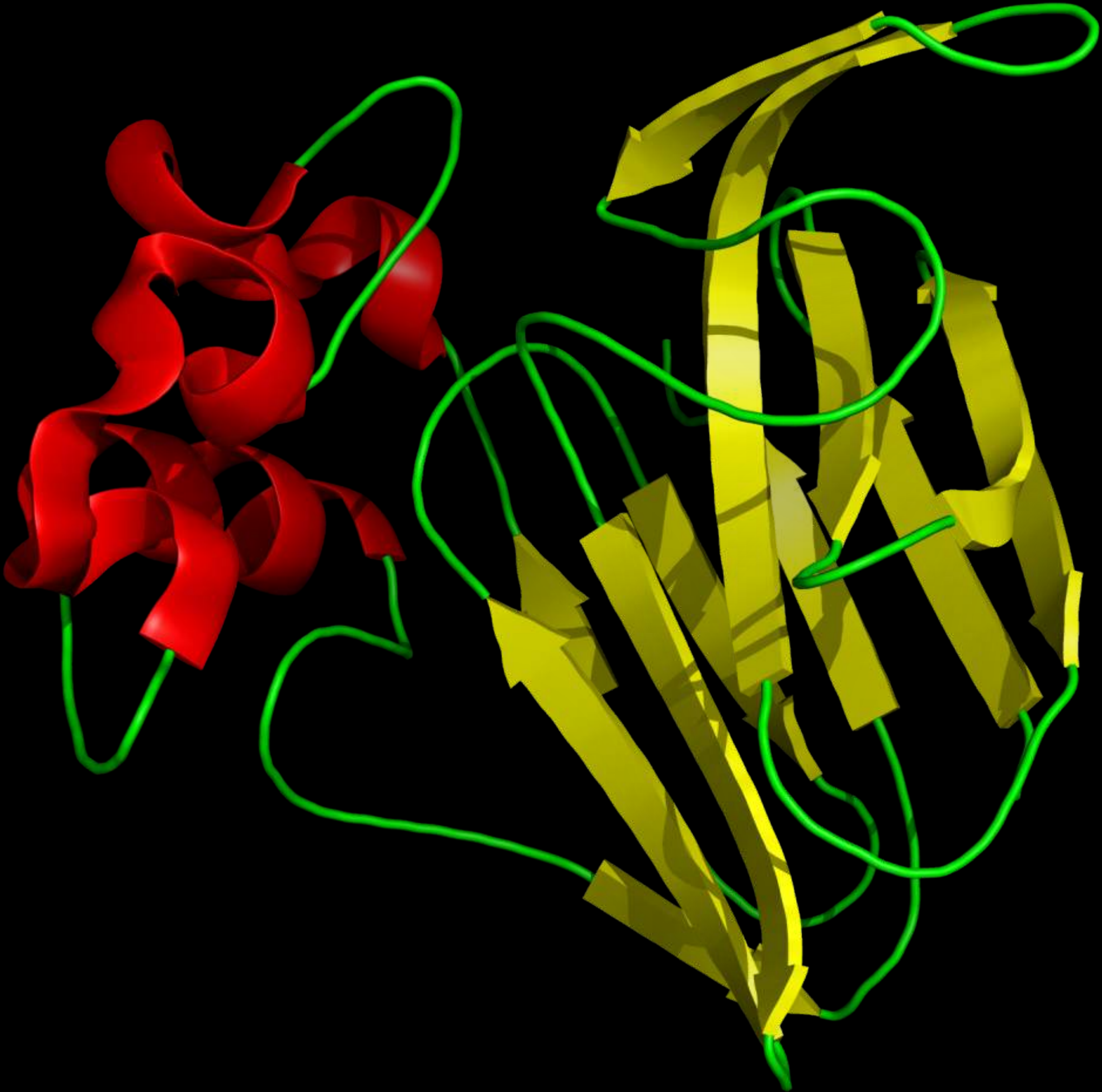
EXPERIMENT


CAUTION TRAP
NODE 31
DELL
MITSUBISHI ELECTRIC
CAUTION TRAP HAZARD


16M







EASIEST ROUTE FROM IMAGES TO INTENSITIES

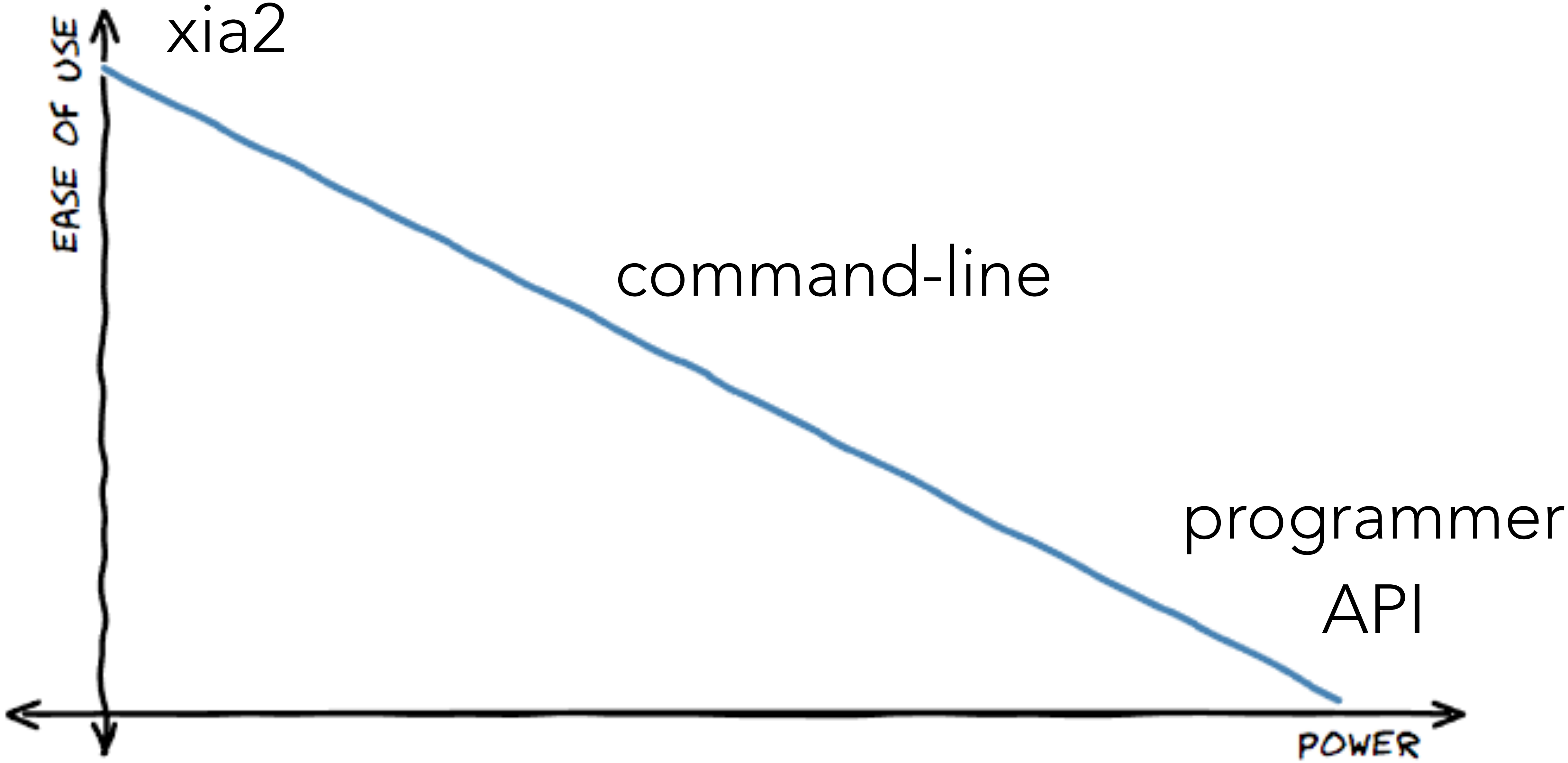
- Automation!

- generate_XDS.INP / autoPROC / xia2 / fast_dp / …. most facilities have *something*

- For dials: xia2 =>

```
xia2 atom=Se image=../SeThau_1_1_master.h5
```

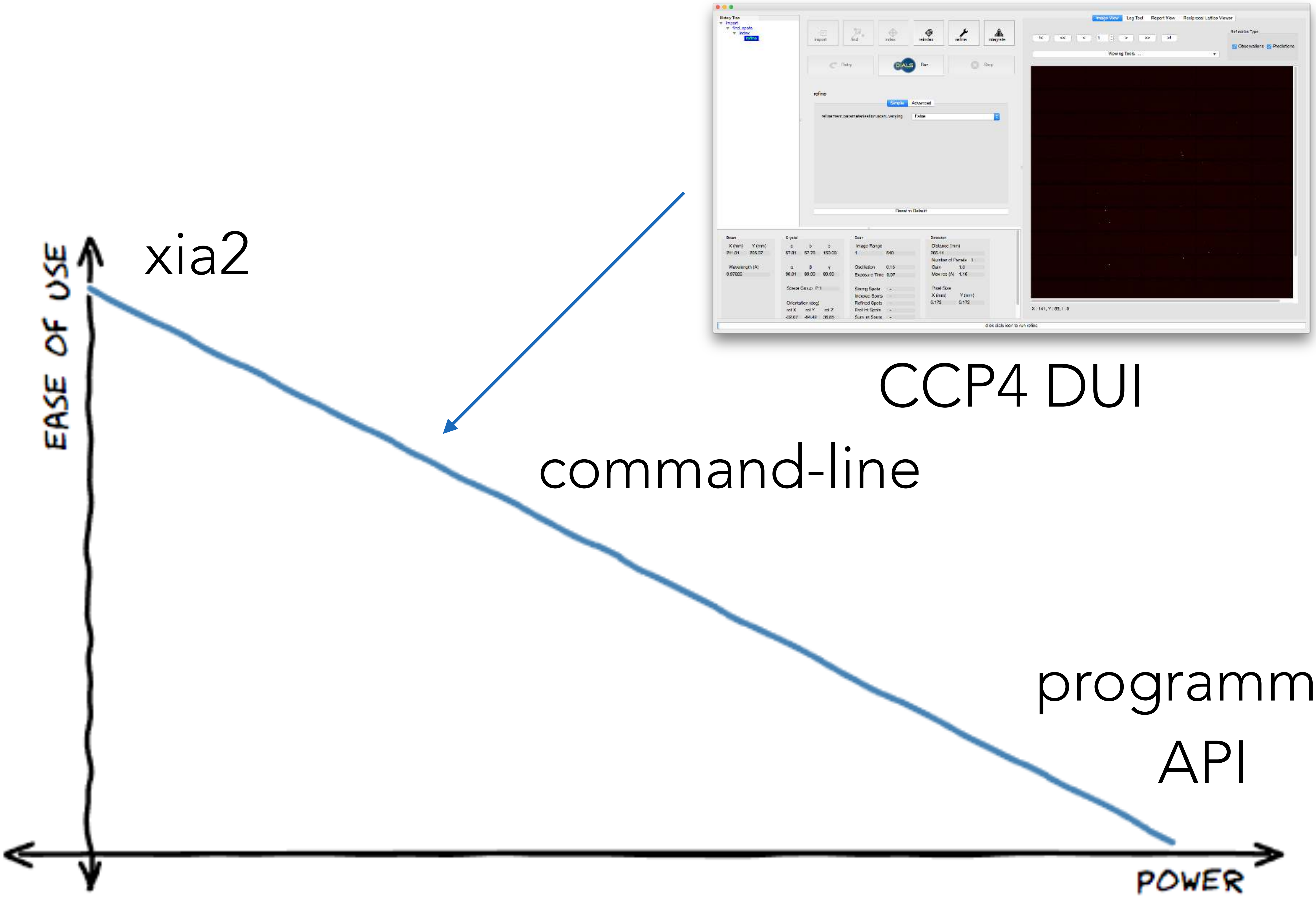
- You would learn *nothing*



xia2

command-line

programmer
API



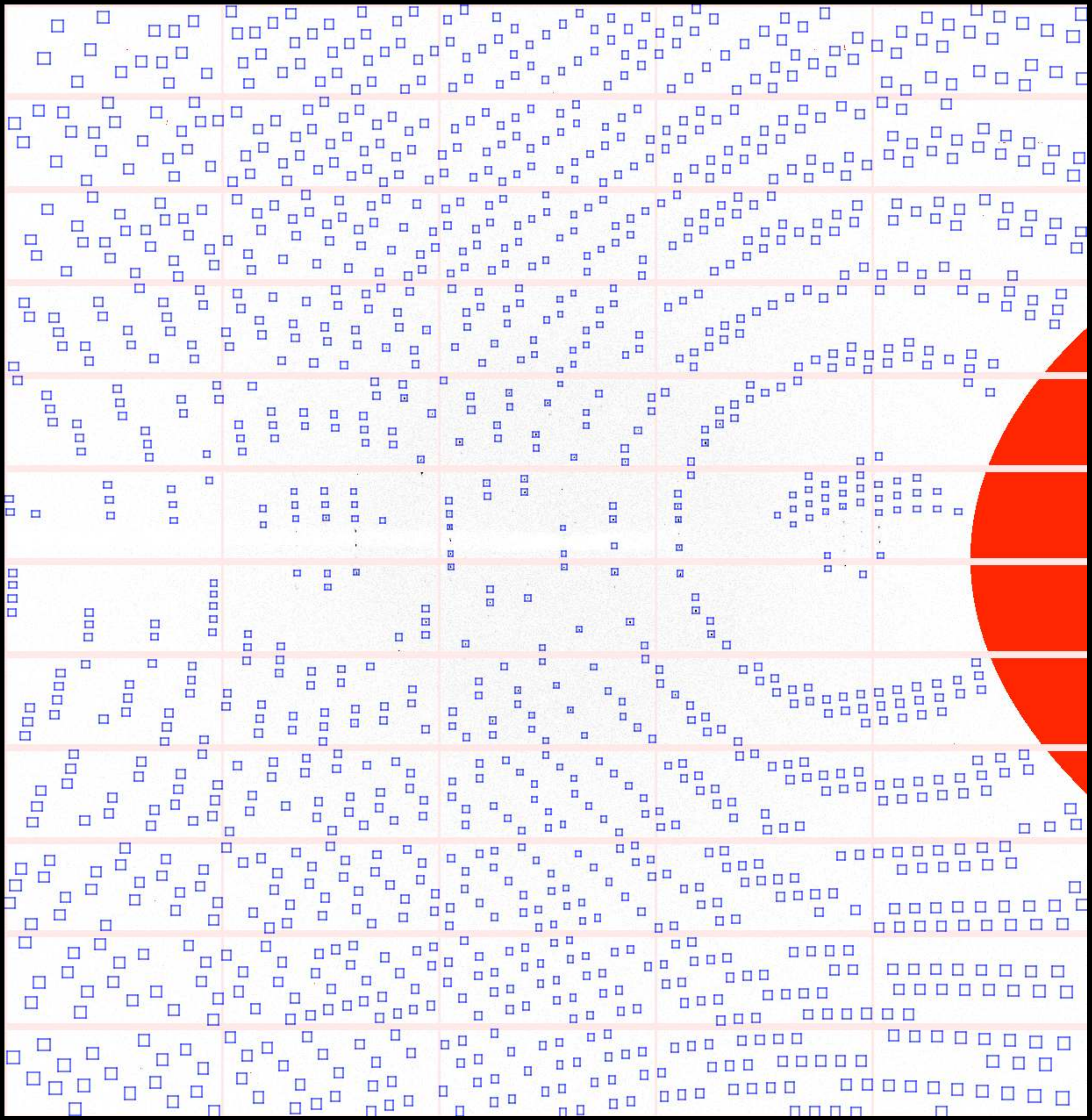
xia2

CCP4 DUI

command-line

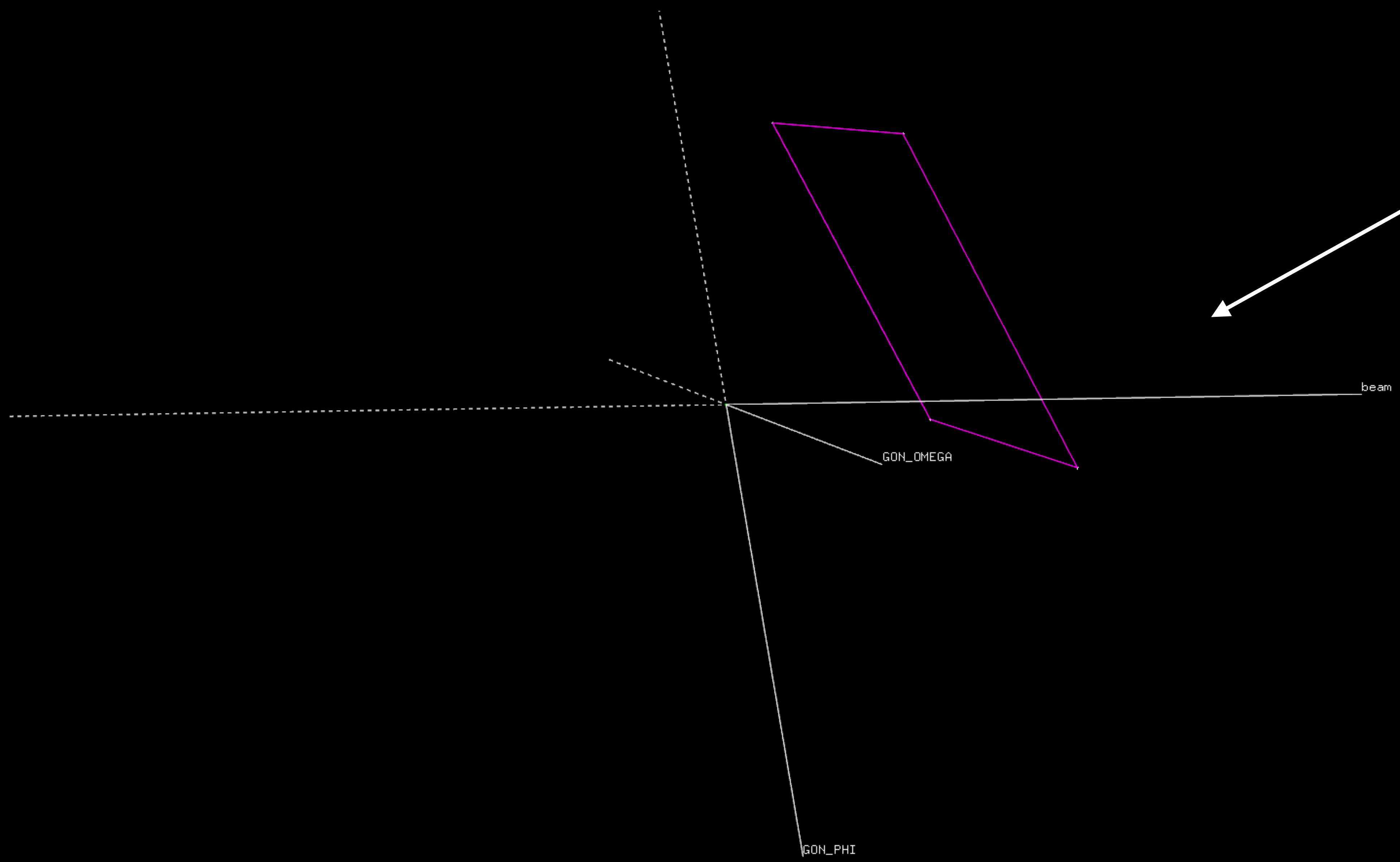
programmer API

WHAT CAN DIALS COPE WITH?



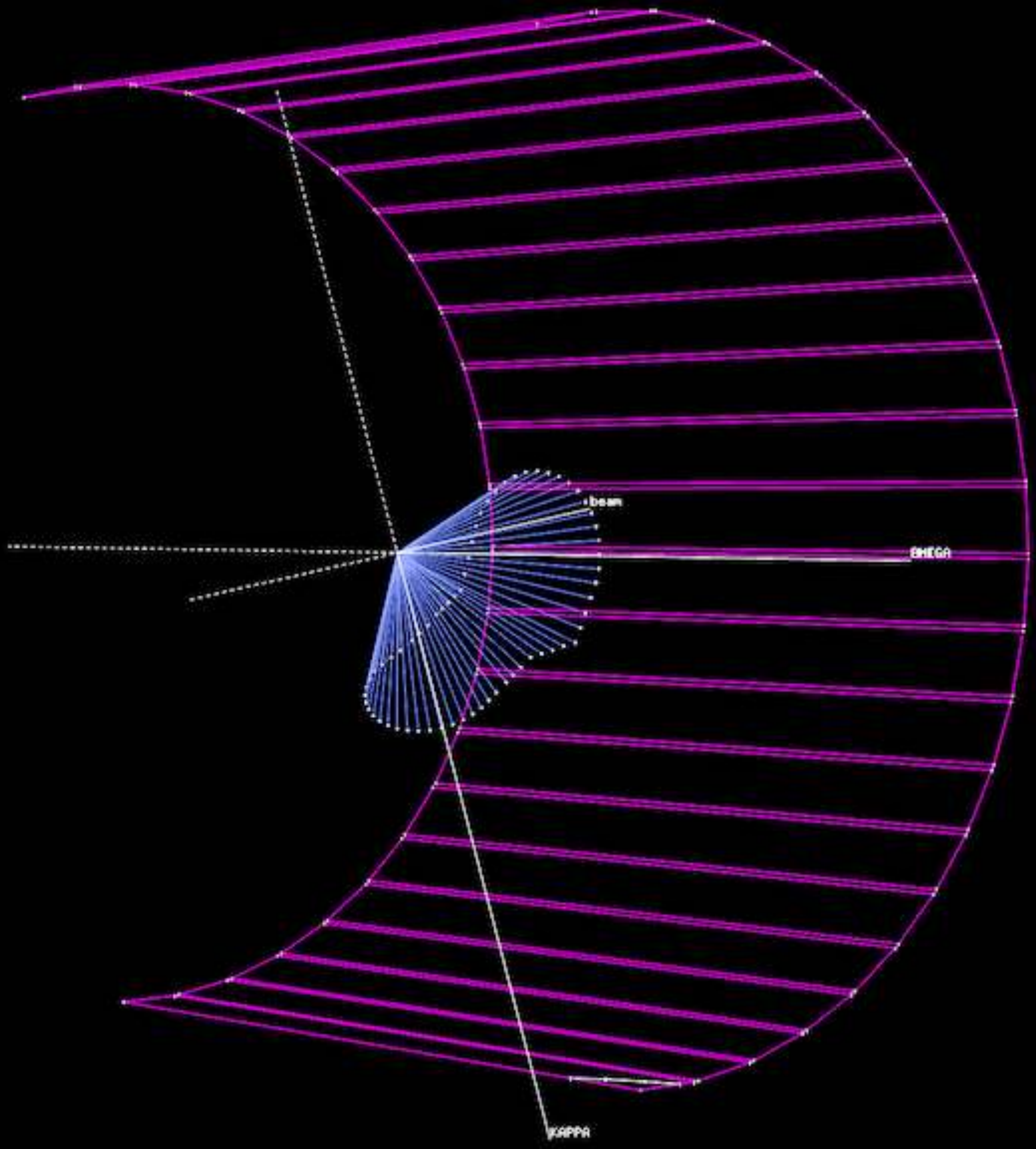
Moving shadows!

WHAT CAN DIALS COPE WITH?



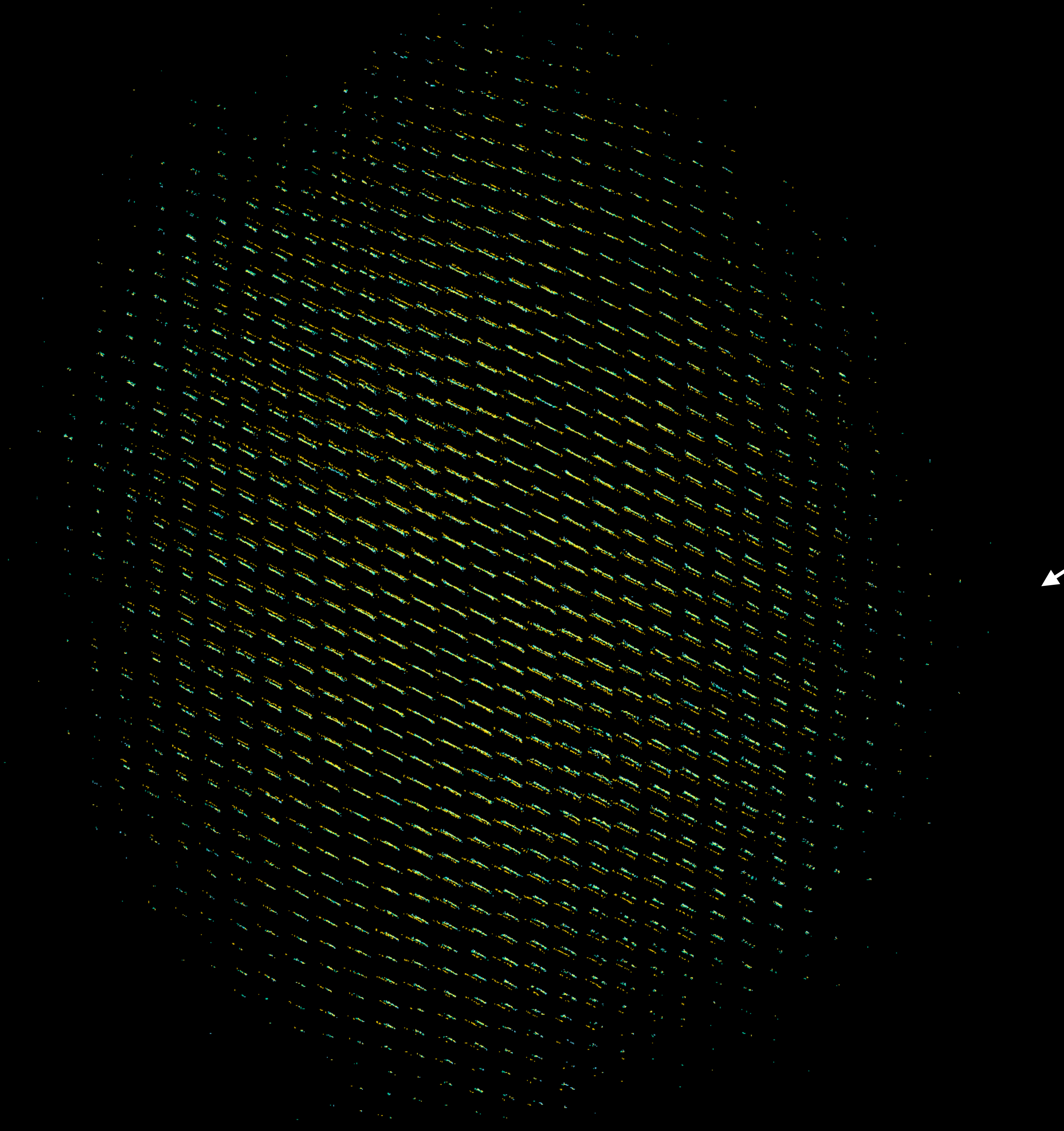
Complex geometry

WHAT CAN DIALS COPE WITH?



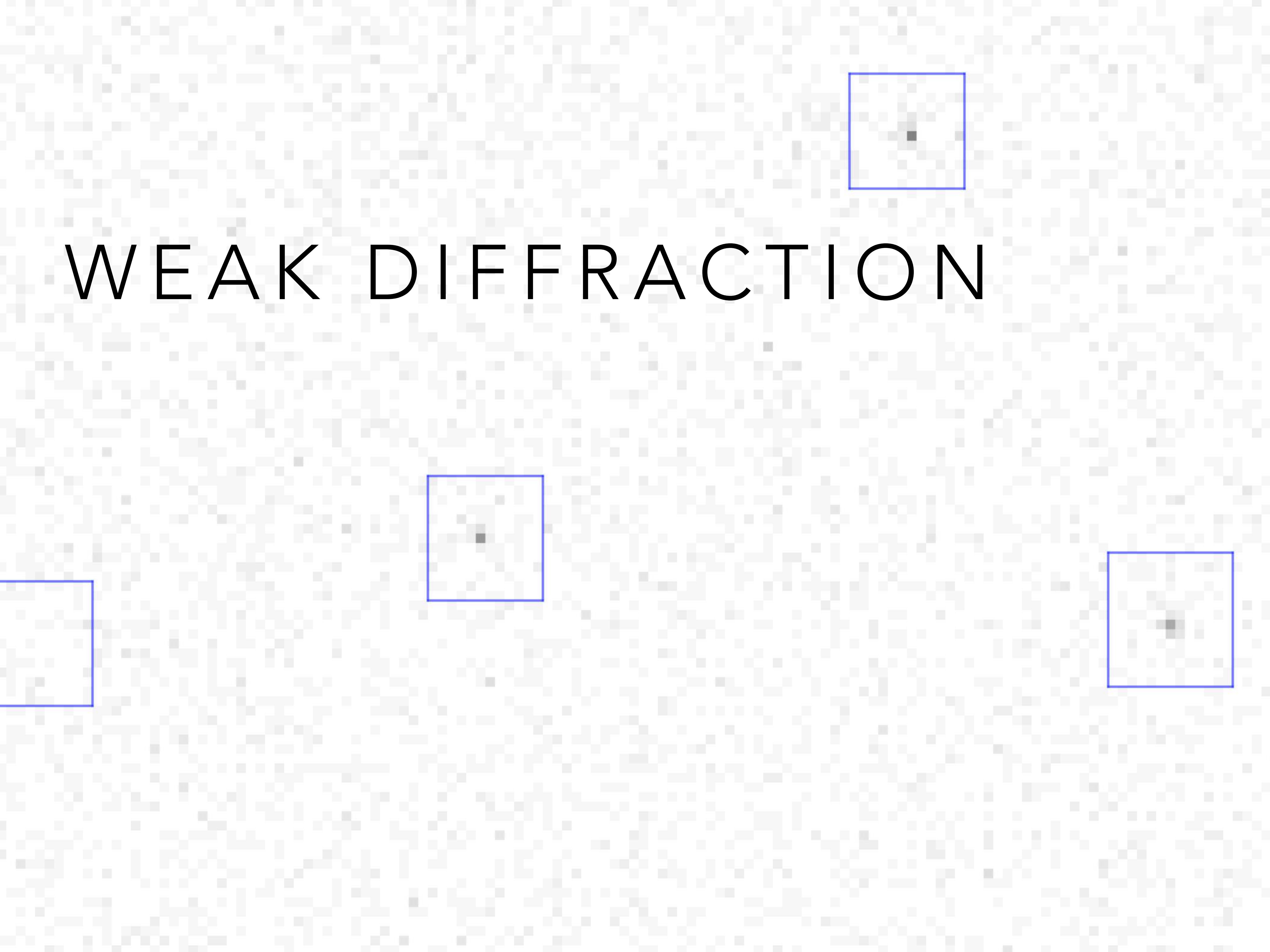
Complex detectors

WHAT CAN DIALS COPE WITH?



Multiple orientations

WEAK DIFFRACTION



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

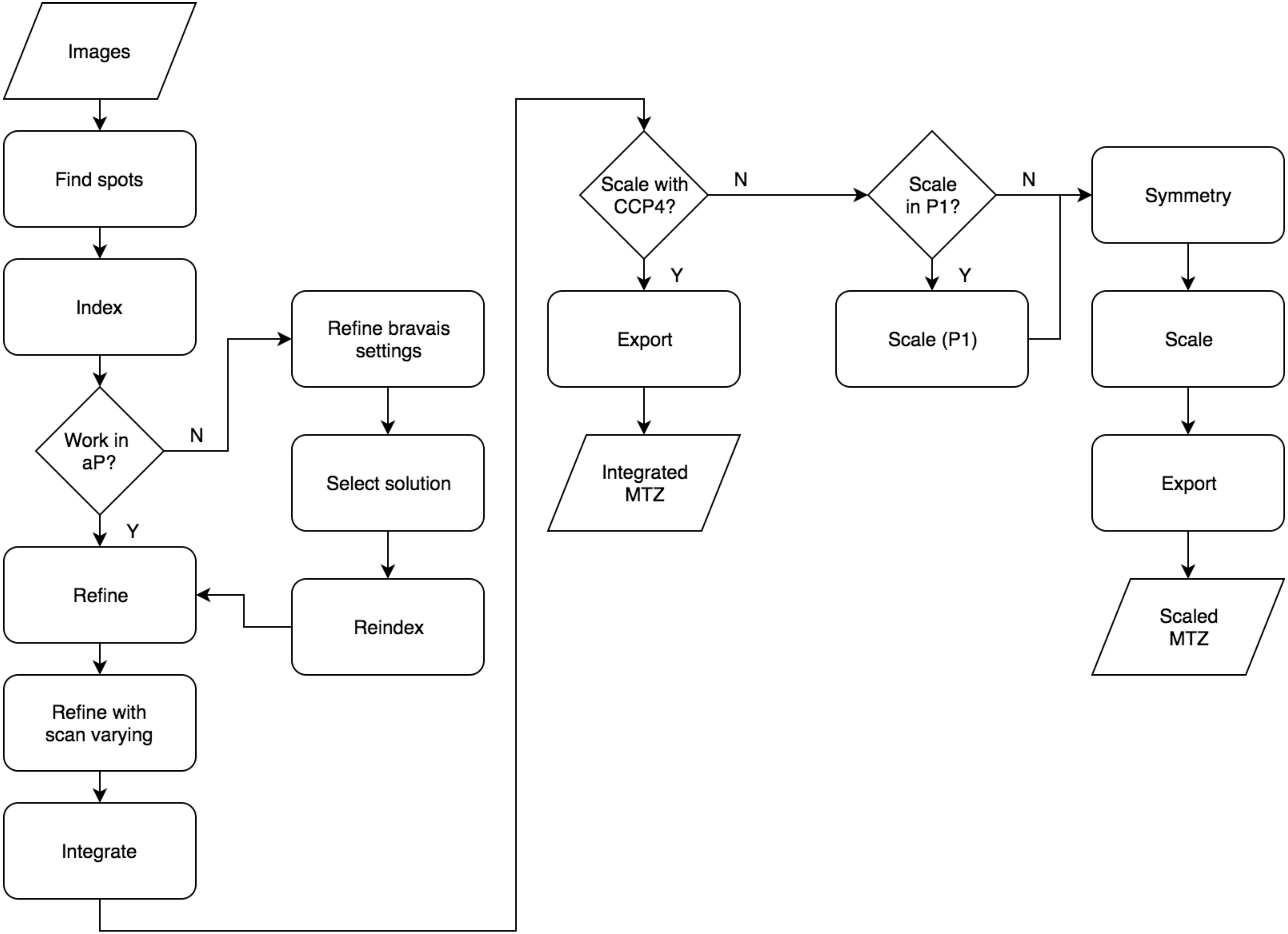
KEY STEPS WITH DIALS

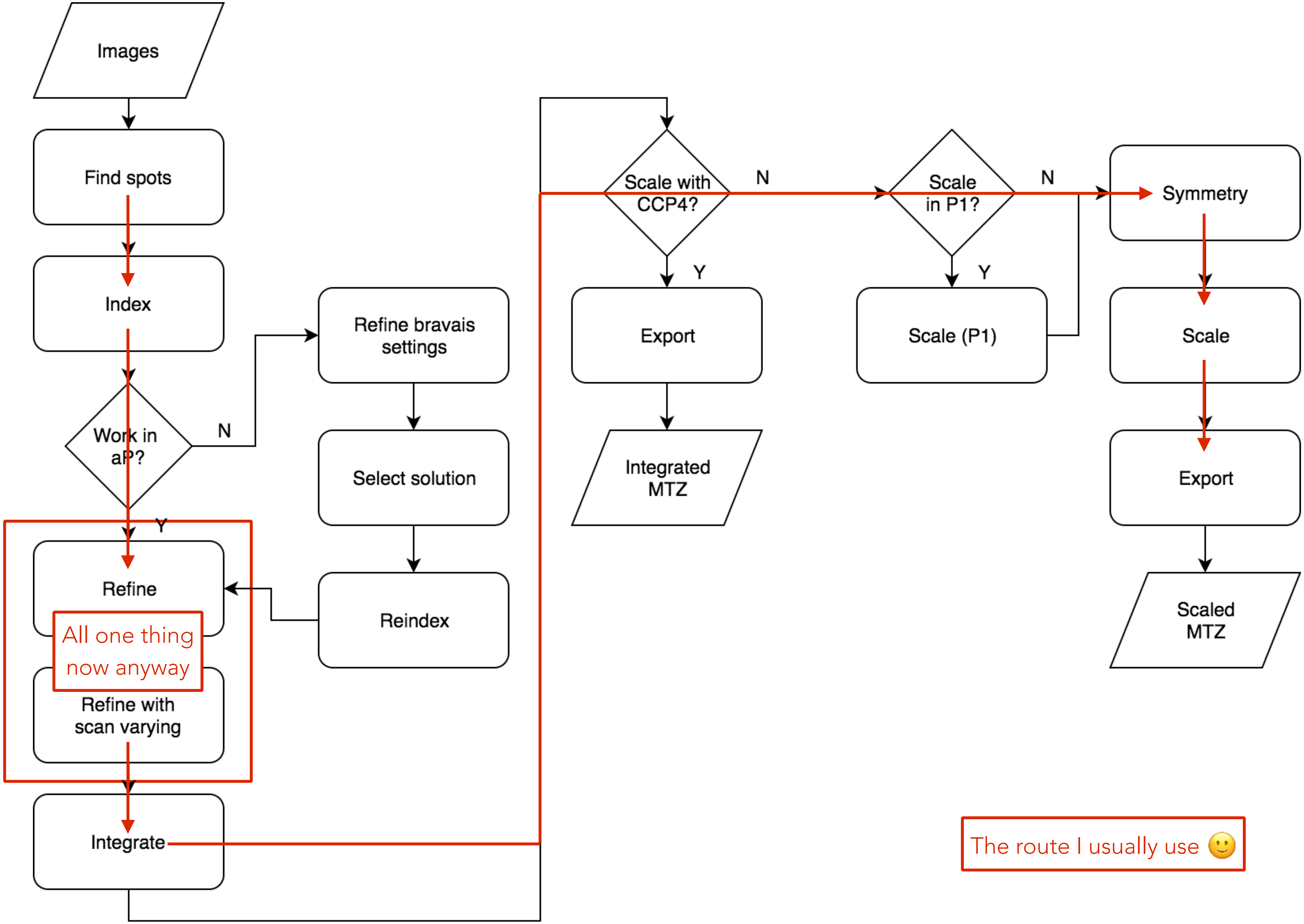
```
dials.import

dials.find_spots imported.expt [nproc=8]

dials.index imported.expt strong.refl [space_group=..] [unit_cell=..]

[dials.refine_bravais_settings indexed.refl indexed.expt]

dials.refine

dials.integrate

dials.symmetry

dials.scale

dials.export
```

KEY STEPS WITH DIALS

```
dials.import

dials.find_spots imported.expt

dials.index imported.expt strong.refl [space_group=..] [unit_cell=..]

[dials.refine_bravais_settings indexed.refl indexed.expt]

dials.refine

dials.integrate

dials.symmetry

dials.scale

dials.export
```

```
dials.image_viewer imported.expt

dials.reciprocal_lattice_viewer imported.expt strong.refl
```


Images
Find spots
Index
Work in aP?
N
Y
Refine bravais settings
Select solution
Reindex
Refine
Refine with scan varying
Integrate
Scale with CCP4?
N
Y
Export
Integrated MTZ
Scale in P1?
N
Y
Scale (P1)
Symmetry
Scale
Export
Scaled MTZ



The route I usually use 🙂

RUNNING DIALS

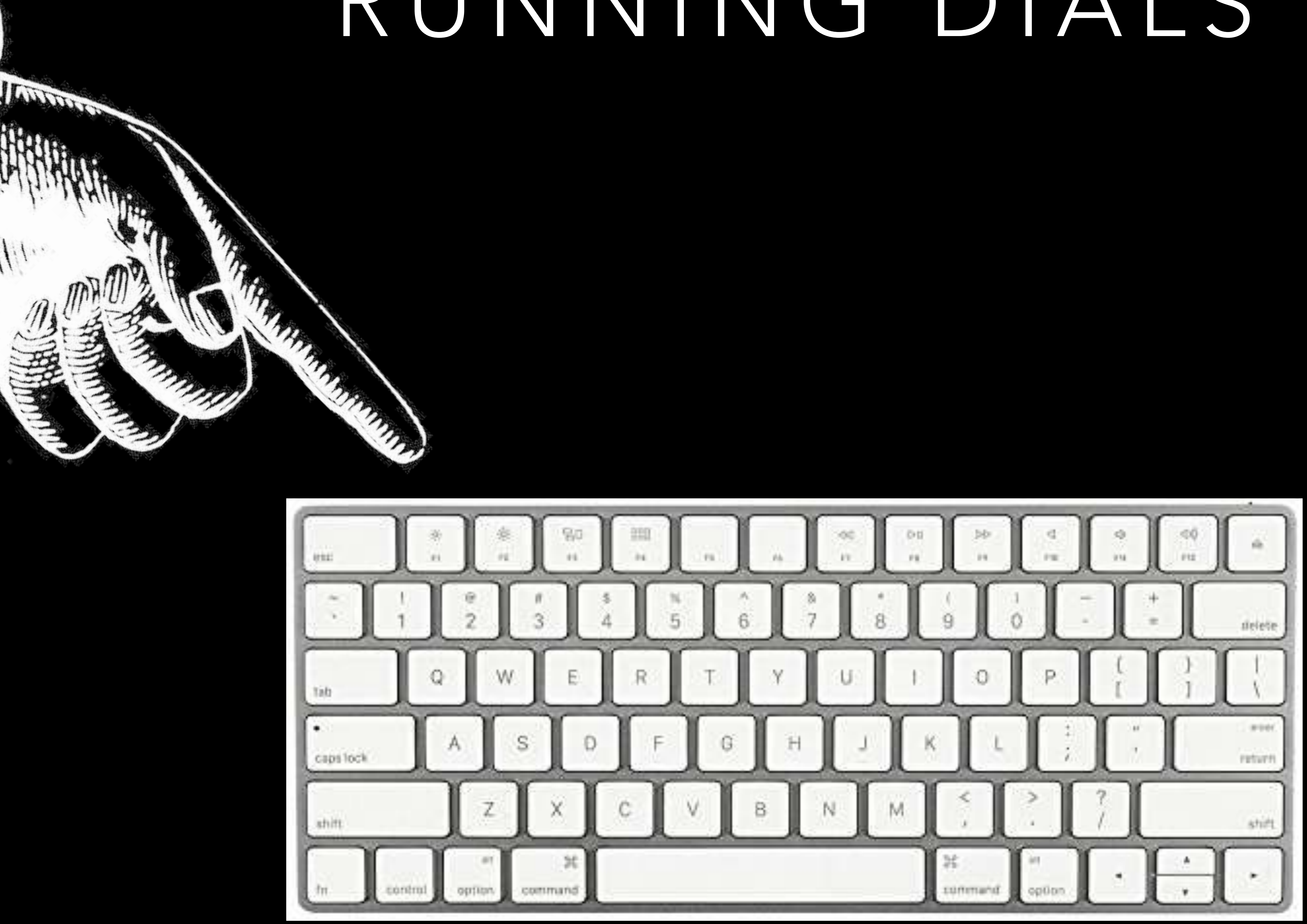
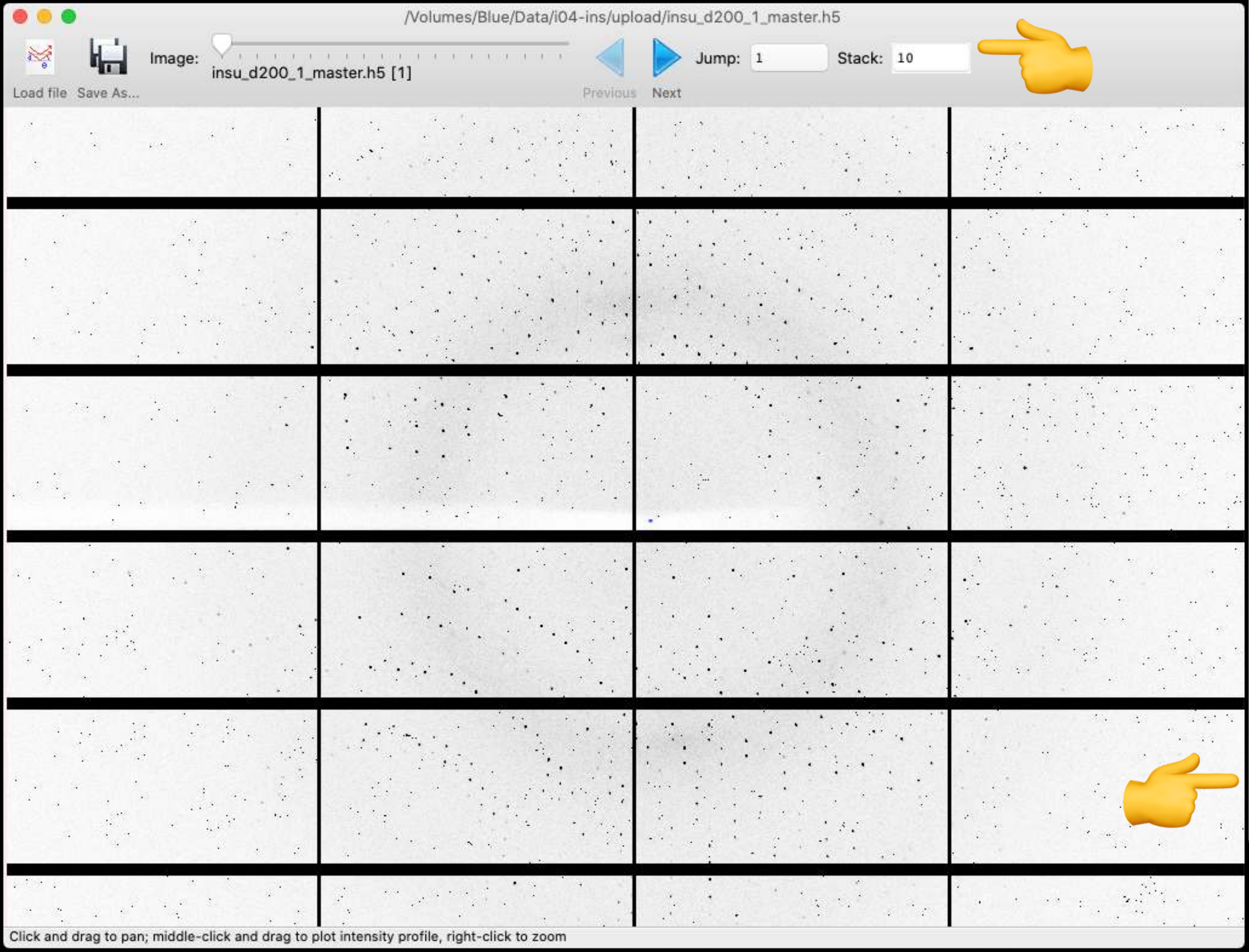


IMAGE VIEWER



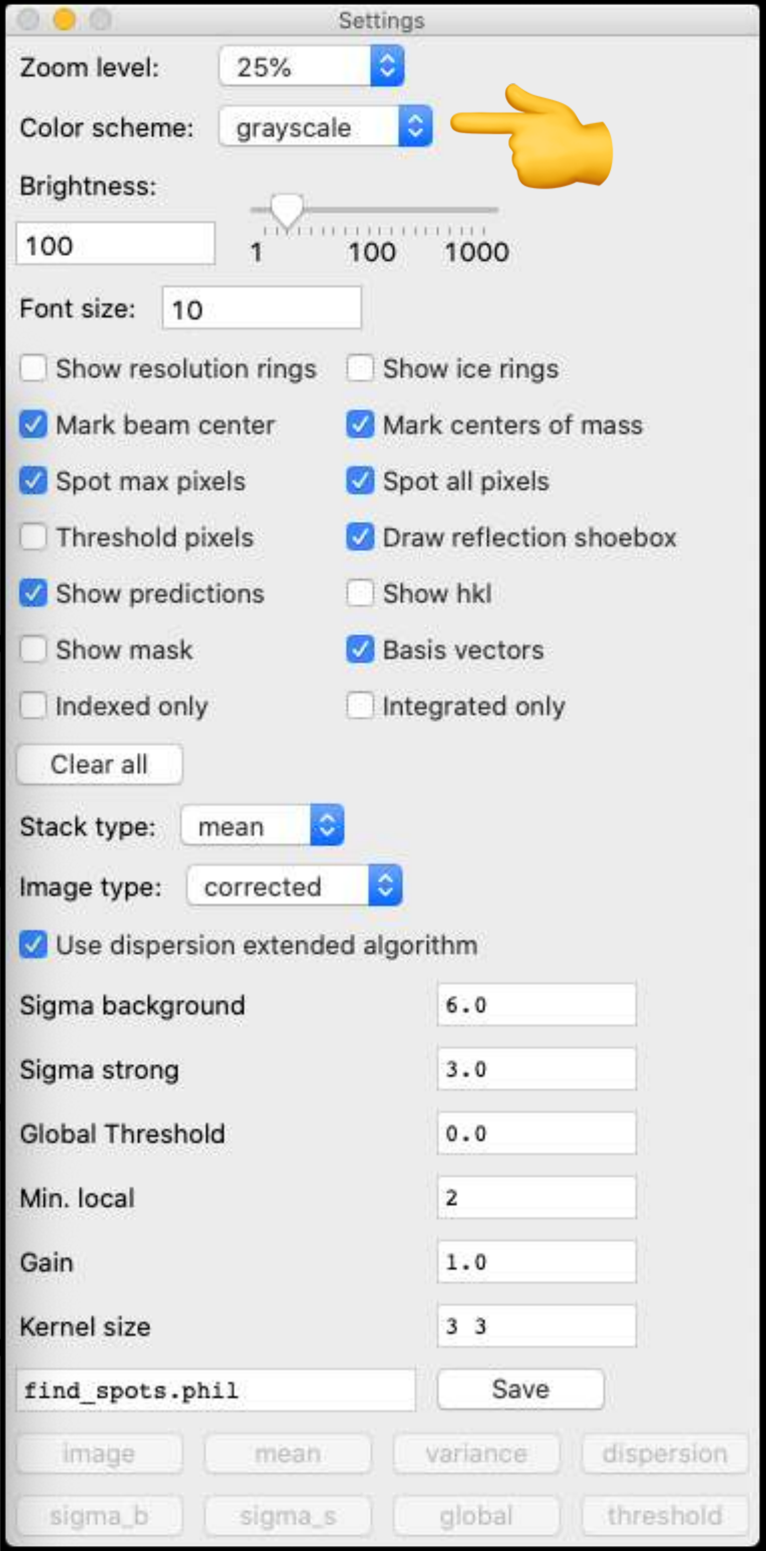
/Volumes/Blue/Data/i04-ins/upload/insu_d200_1_master.h5

Image: insu_d200_1_master.h5 [1]

Jump: 1

Stack: 10

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



Settings

Zoom level: 25%

Color scheme: grayscale

Brightness: 100

Font size: 10

☐ Show resolution rings
☐ Show ice rings
☑ Mark beam center
☑ Mark centers of mass
☑ Spot max pixels
☑ Spot all pixels
☐ Threshold pixels
☑ Draw reflection shoebox
☑ Show predictions
☐ Show hkl
☐ Show mask
☑ Basis vectors
☐ Indexed only
☐ Integrated only

Stack type: mean

Image type: corrected

☑ Use dispersion extended algorithm

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

DIALS FILES

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

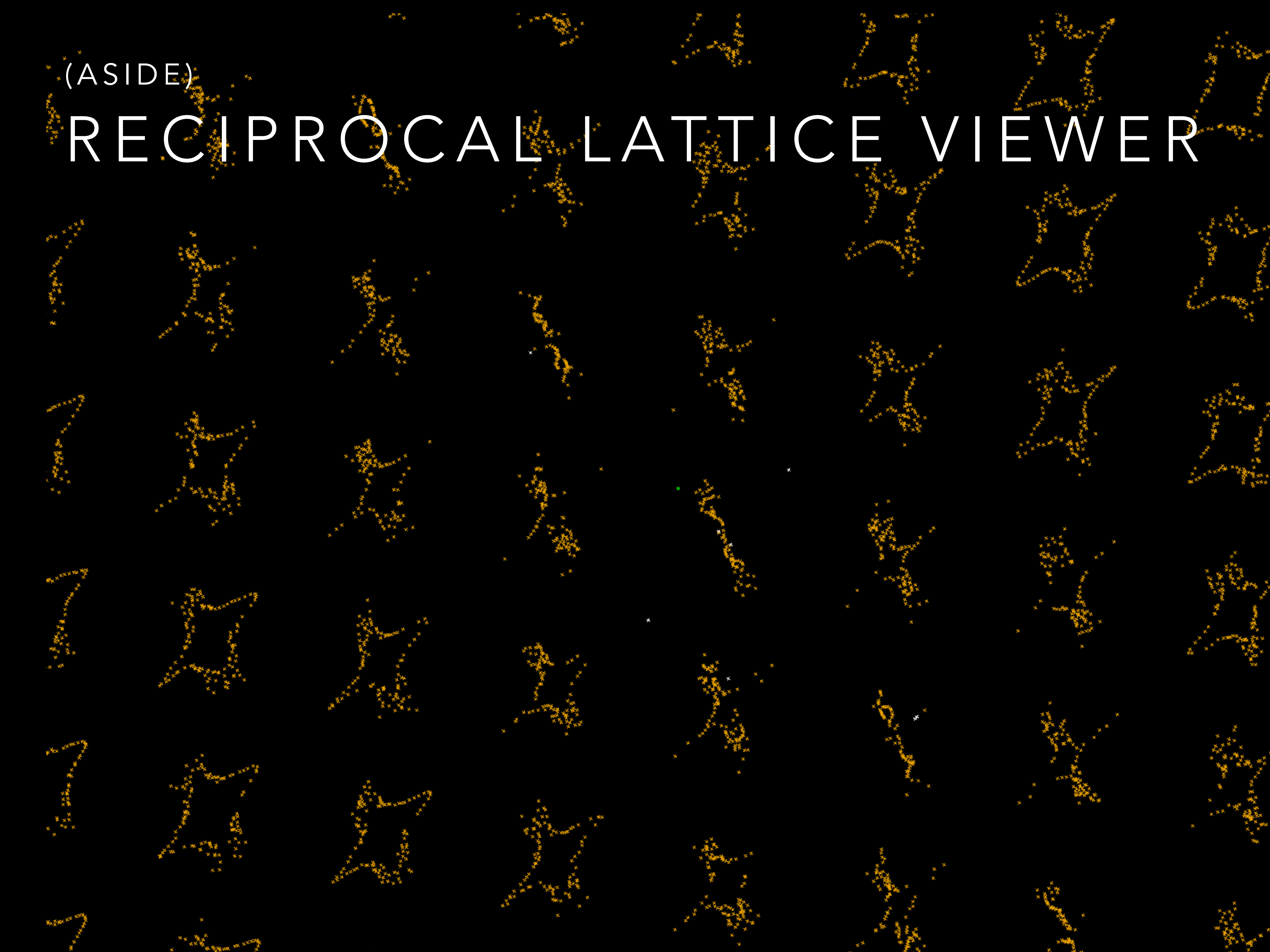
ALGORITHMS

SPOT FINDING

MAP TO RECIPROCAL SPACE

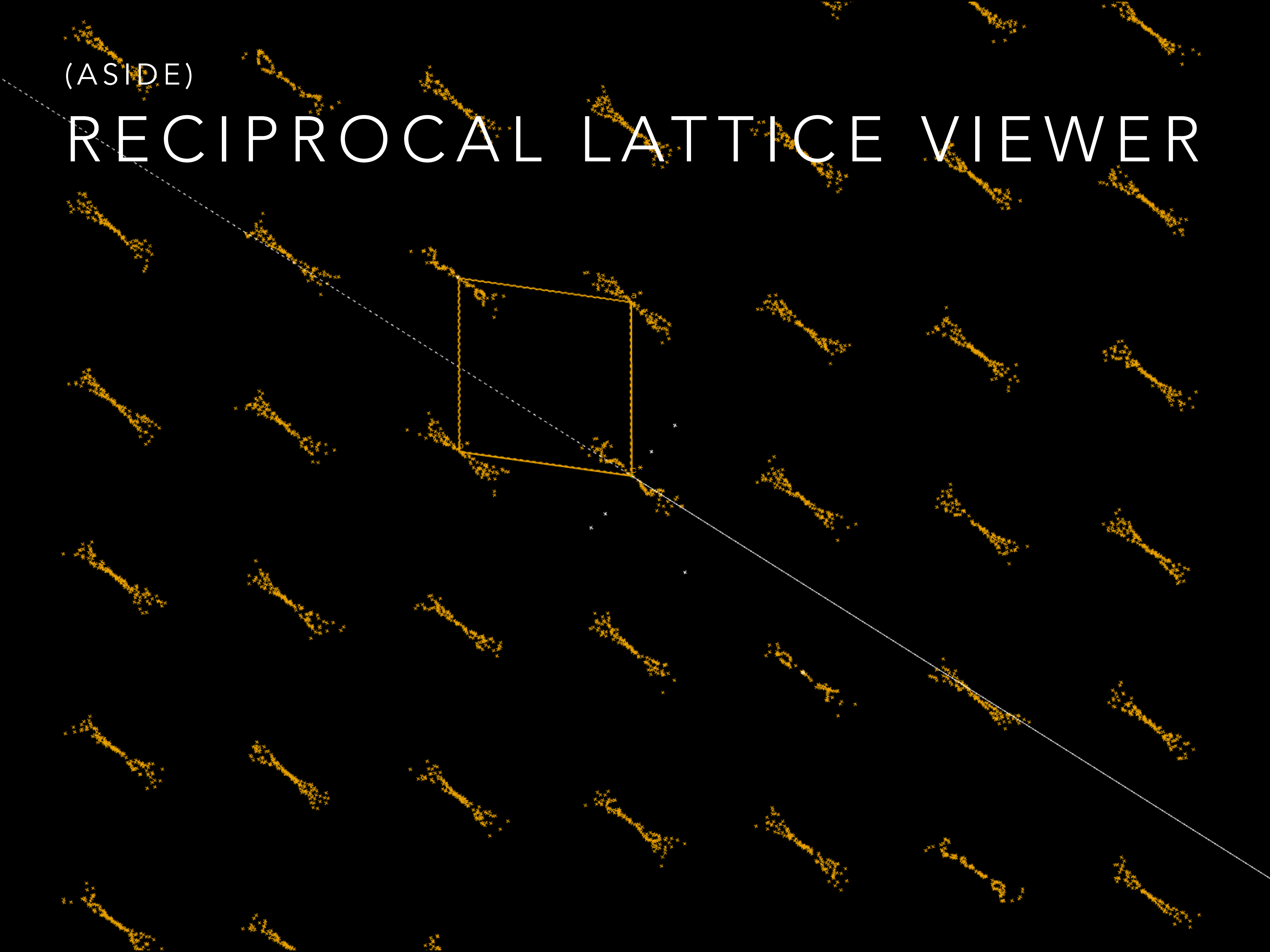
(ASIDE)

RECIPROCAL LATTICE VIEWER

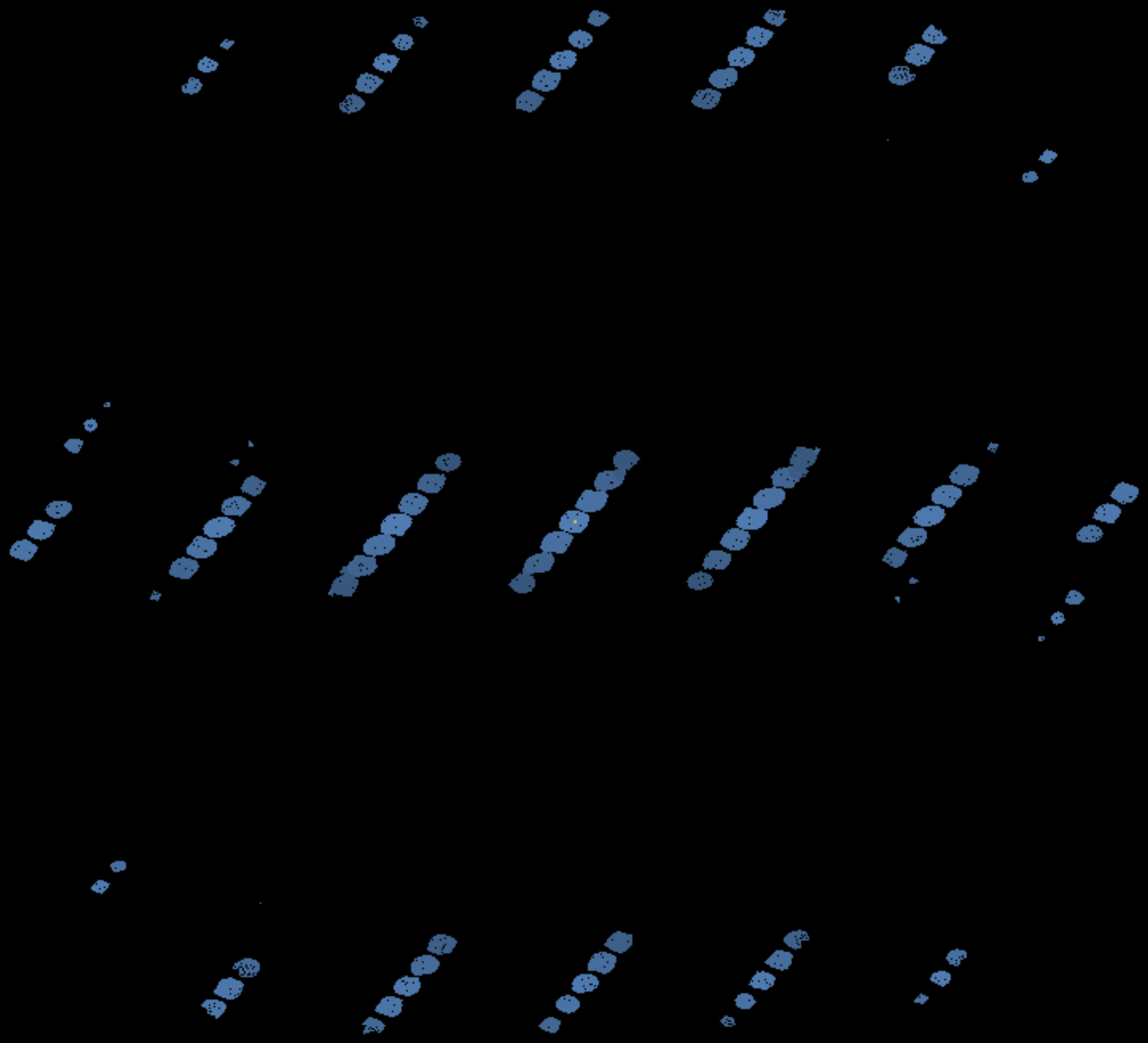


(ASIDE)

RECIPROCAL LATTICE VIEWER



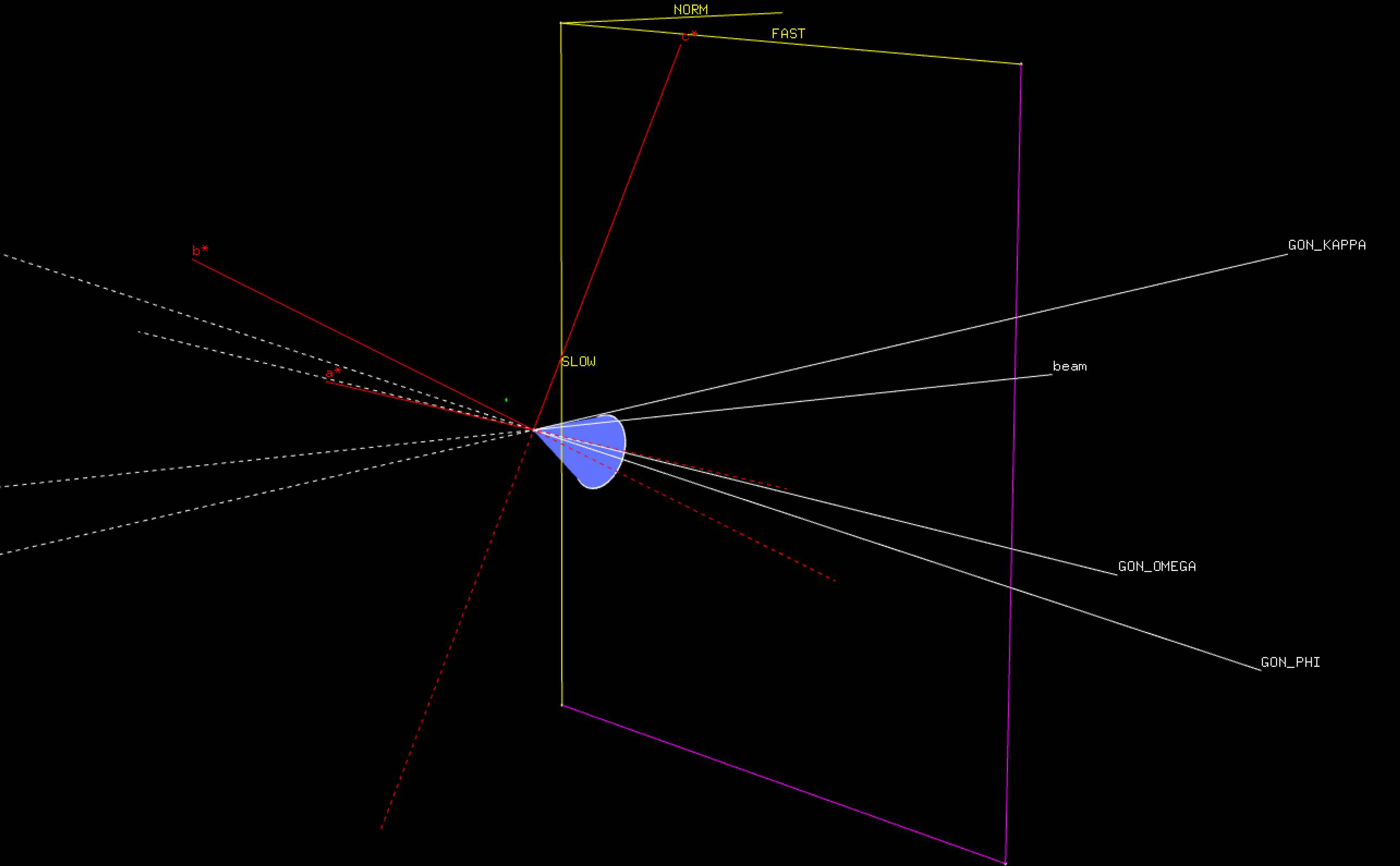
INDEXING (3D FFT)



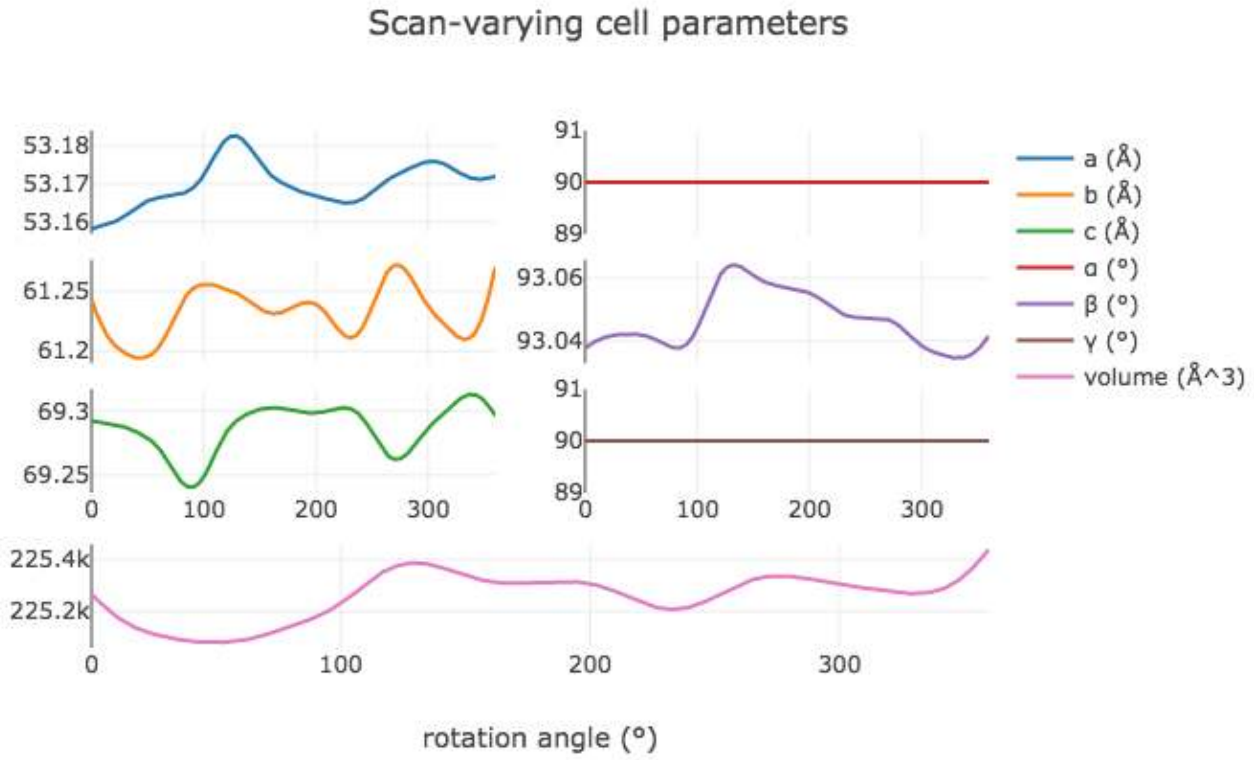
(OPTIONAL!) LATTICE SELECTION

```
---------------------------------------------------------------------------------------------------------
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
---------------------------------------------------------------------------------------------------------
```

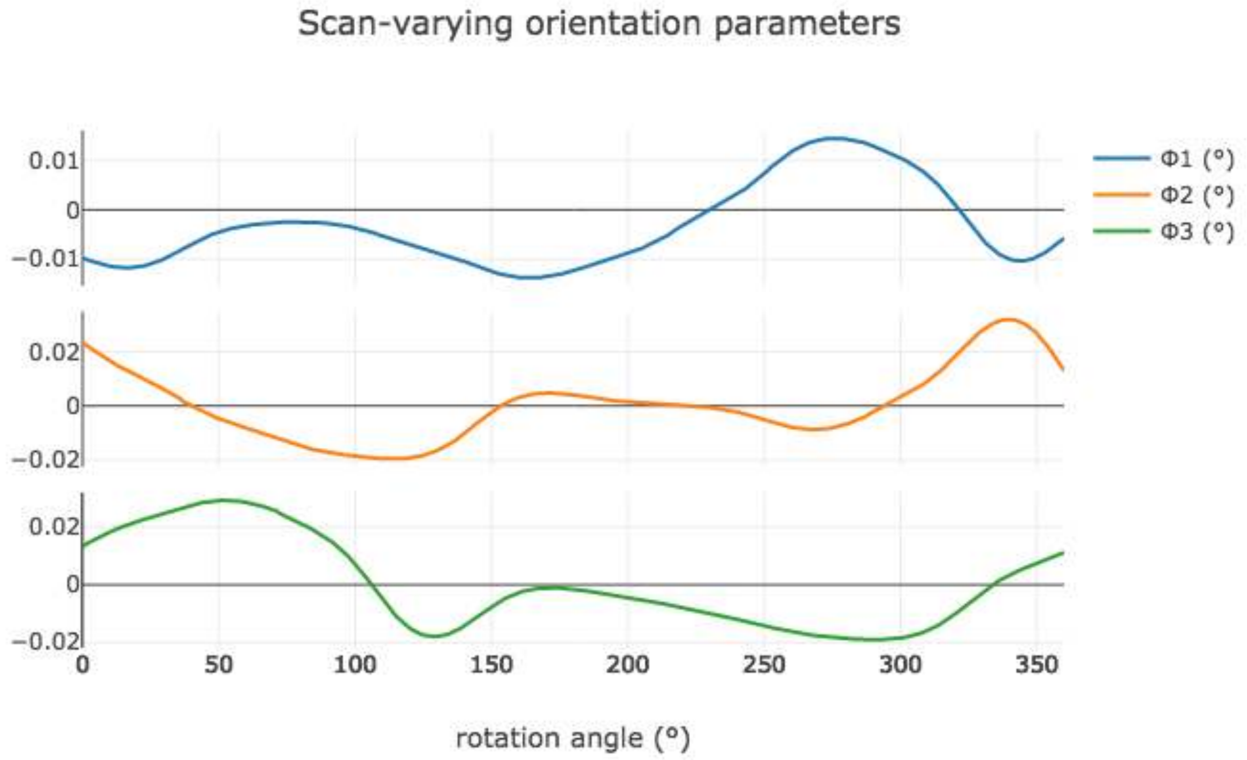
REFINEMENT



REFINEMENT



Scan-varying cell parameters

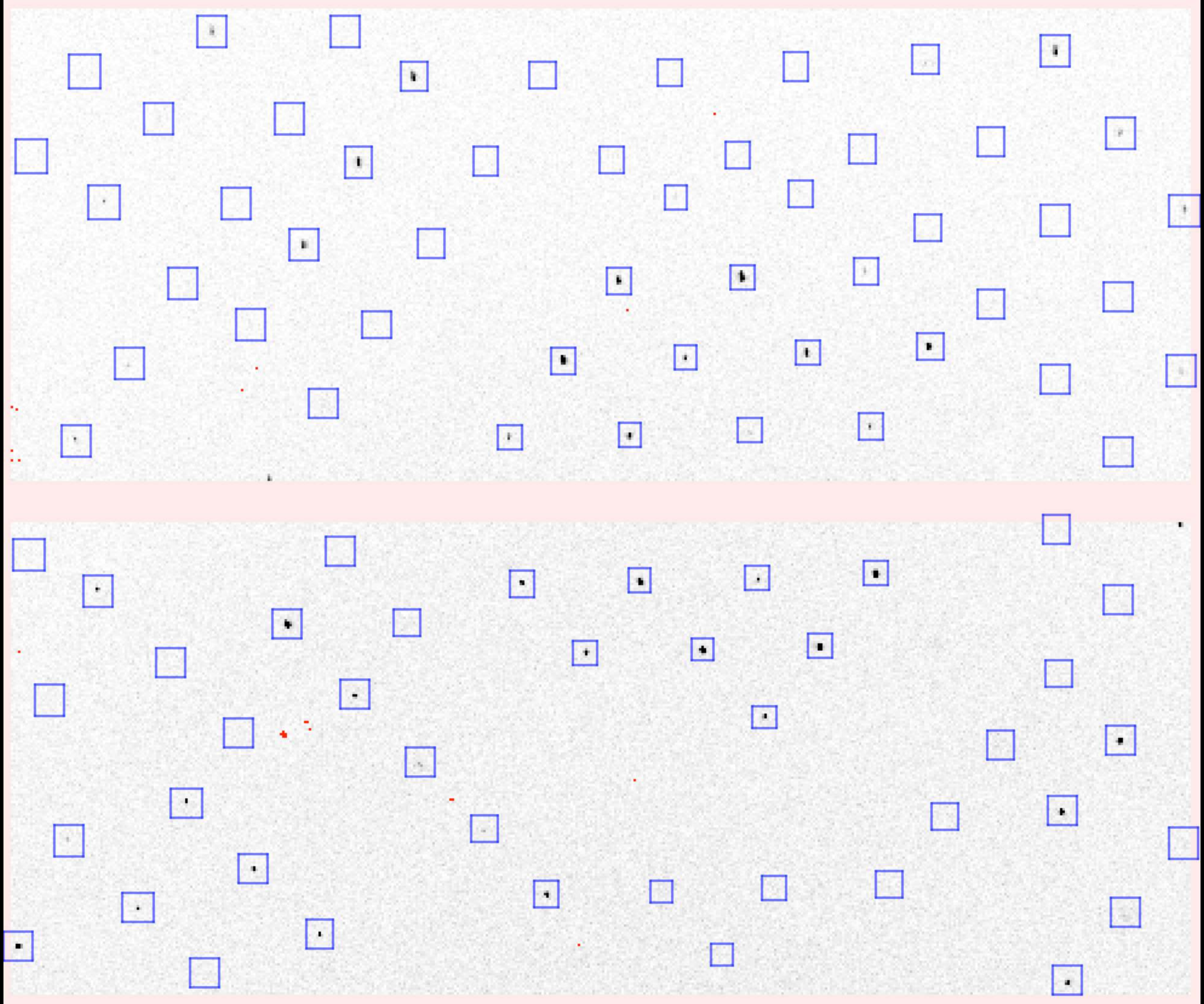


Scan-varying orientation parameters

REFINEMENT FEATURES

- Possible to do *joint refinement* - one model for geometry for e.g. 50 data sets
- True errors on unit cell parameters - critical for assessing isomorphism

INTEGRATION



SCALING

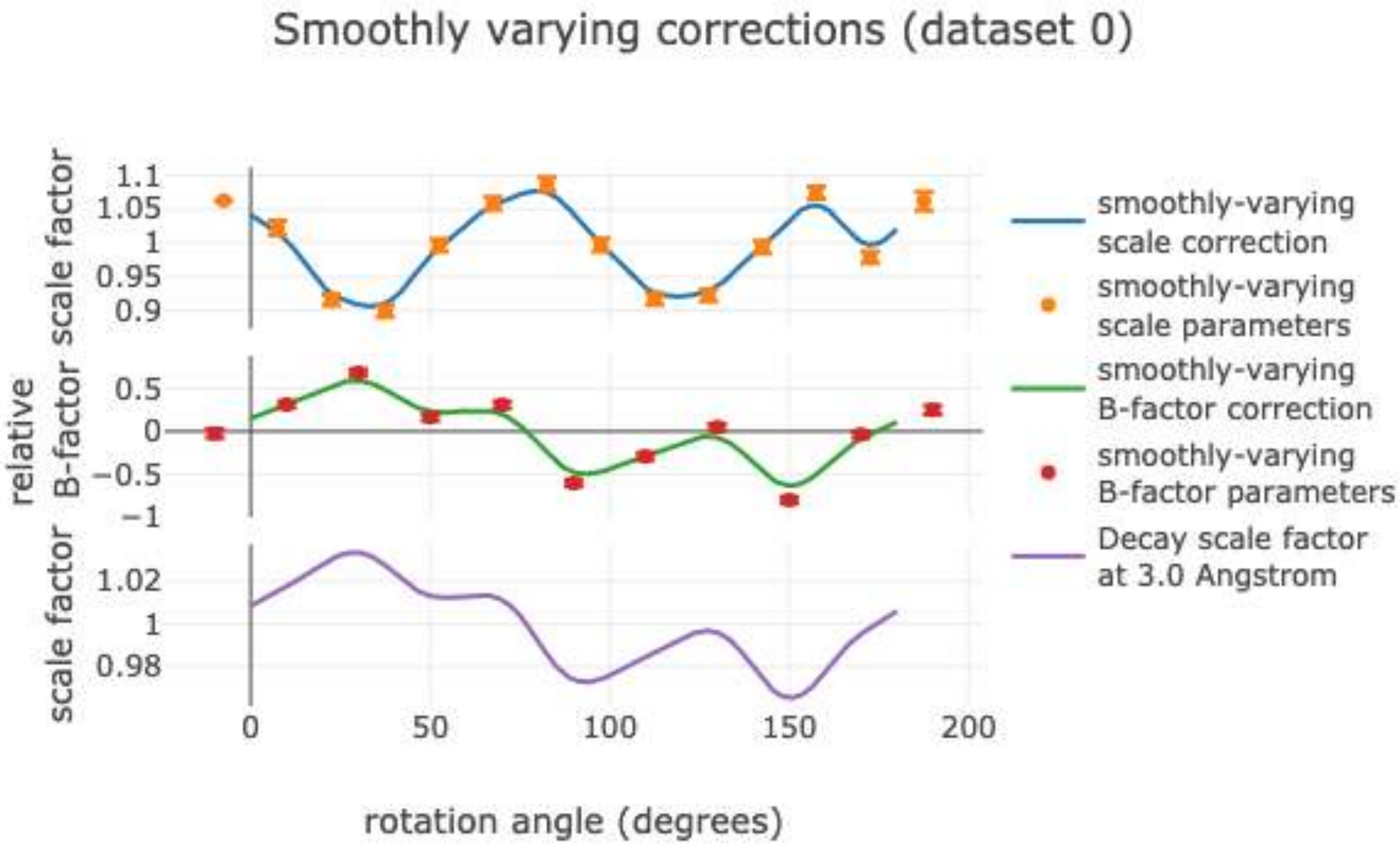
- dials.symmetry - determine the Patterson symmetry of the data
- dials.scale - correct for experimental effects using symmetry-related experiments

SYMMETRY

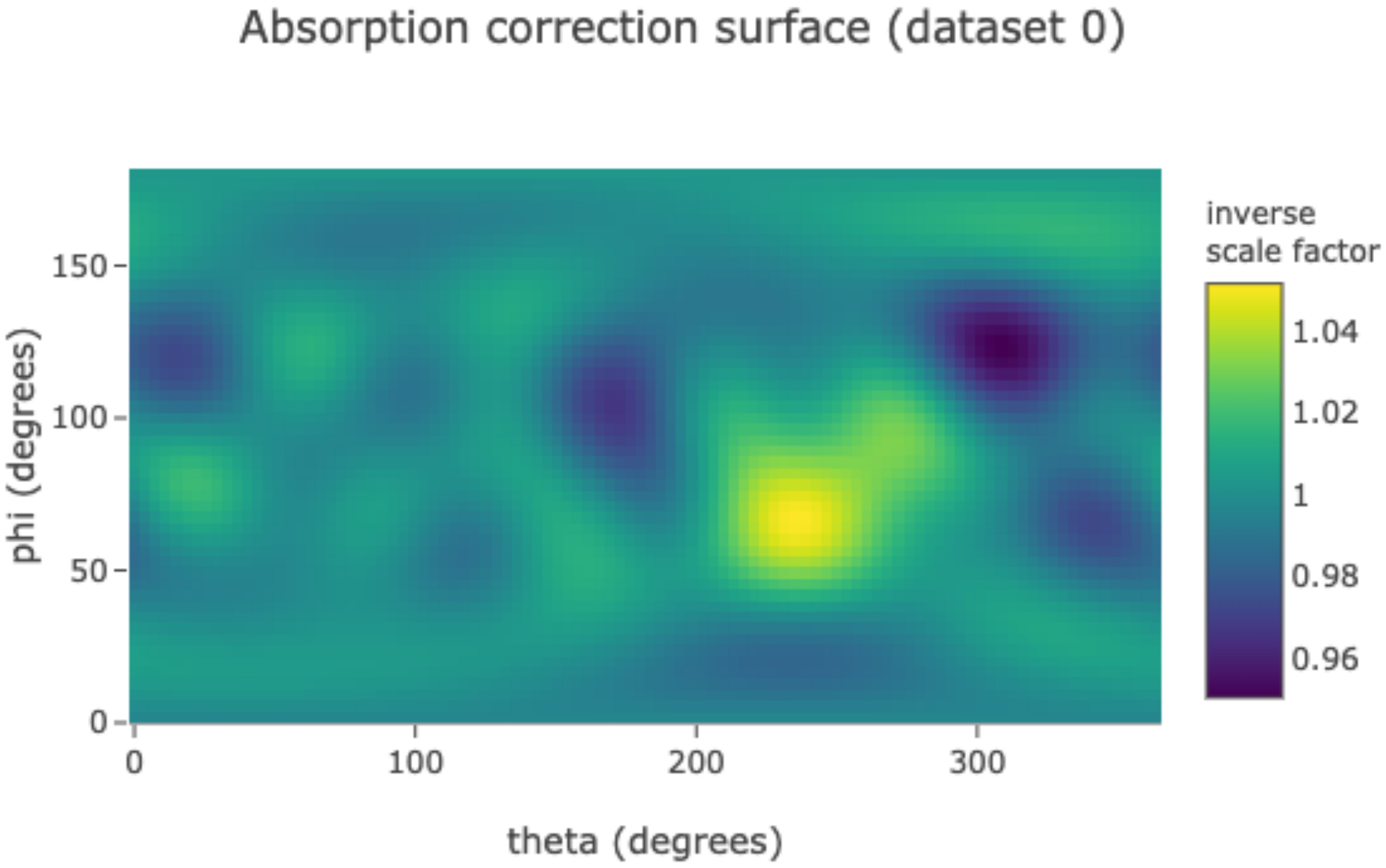
```
Scoring individual symmetry elements

+--------------+--------+------+--------+-----+---------------+
|   likelihood |   Z-CC |   CC |      N |     | Operator      |
|--------------+--------+------+--------+-----+---------------|
|        0.94  |   9.9  | 0.99 | 362986 | *** | 1 |(0, 0, 0)  |
|        0.94  |   9.92 | 0.99 | 751640 | *** | 4 |(0, 0, 1)  |
|        0.941 |   9.92 | 0.99 | 373768 | *** | 2 |(1, 0, 0)  |
|        0.941 |   9.93 | 0.99 | 393952 | *** | 2 |(0, 1, 0)  |
|        0.94  |   9.91 | 0.99 | 364832 | *** | 2 |(0, 0, 1)  |
|        0.94  |   9.9  | 0.99 | 389620 | *** | 2 |(1, 1, 0)  |
|        0.94  |   9.91 | 0.99 | 367896 | *** | 2 |(-1, 1, 0) |
+--------------+--------+------+--------+-----+---------------+

--------------------------------------------------------------------------------

Scoring all possible sub-groups

+-------------------+-----+--------------+----------+--------+--------+------+-------+---------+--------------------+
| Patterson group   |     |   Likelihood |   NetZcc |   Zcc+ |   Zcc- |   CC |   CC- |   delta | Reindex operator   |
|-------------------+-----+--------------+----------+--------+--------+------+-------+---------+--------------------|
| P 4/m m m         | *** |        0.999 |     9.91 |   9.91 |   0    | 0.99 |  0    |     0.1 | a,b,c              |
| P m m m           |     |        0     |     0.01 |   9.92 |   9.91 | 0.99 |  0.99 |     0.1 | a,b,c              |
| C m m m           |     |        0     |    -0.02 |   9.91 |   9.92 | 0.99 |  0.99 |     0.1 | a+b,a-b,-c         |
| P 4/m             |     |        0     |    -0.01 |   9.91 |   9.92 | 0.99 |  0.99 |     0.1 | a,b,c              |
| P 1 2/m 1         |     |        0     |     0    |   9.91 |   9.91 | 0.99 |  0.99 |     0   | a,b,c              |
| P 1 2/m 1         |     |        0     |     0    |   9.91 |   9.91 | 0.99 |  0.99 |     0.1 | b,-a,c             |
| C 1 2/m 1         |     |        0     |    -0.01 |   9.91 |   9.92 | 0.99 |  0.99 |     0.1 | a-b,-a-b,-c        |
| P 1 2/m 1         |     |        0     |    -0.01 |   9.91 |   9.92 | 0.99 |  0.99 |     0   | -a,c,b             |
| C 1 2/m 1         |     |        0     |    -0.01 |   9.9  |   9.92 | 0.99 |  0.99 |     0.1 | a+b,a-b,-c         |
| P -1              |     |        0     |    -0.01 |   9.9  |   9.91 | 0.99 |  0.99 |     0   | a,b,c              |
+-------------------+-----+--------------+----------+--------+--------+------+-------+---------+--------------------+
```

SCALING

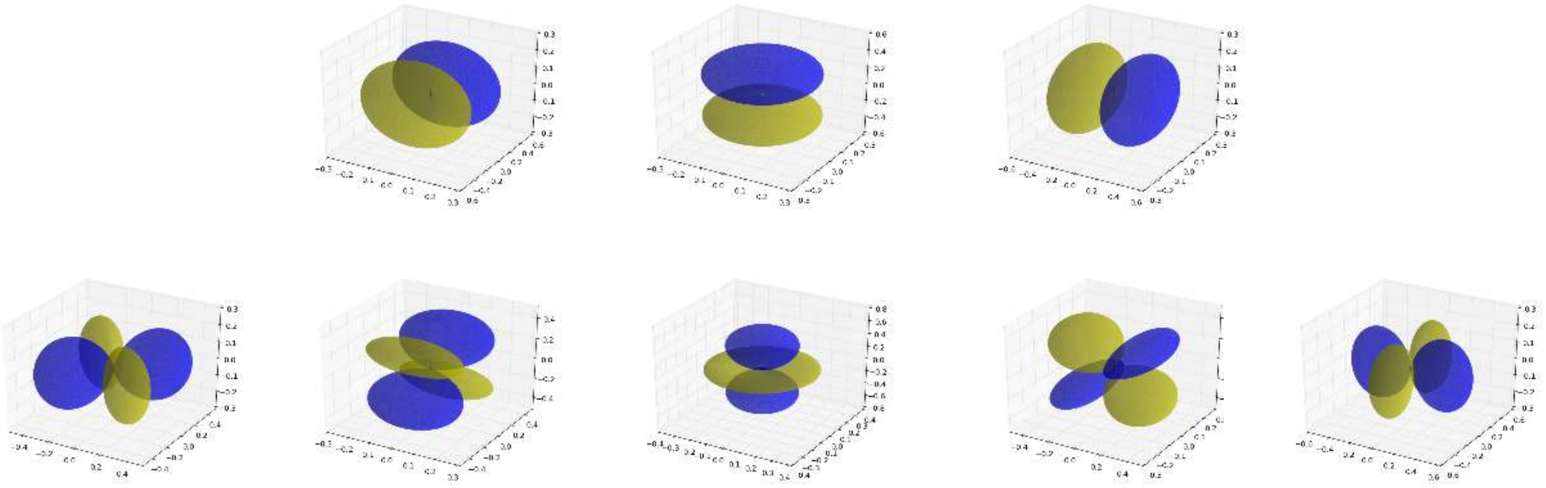
kB (smoothed)



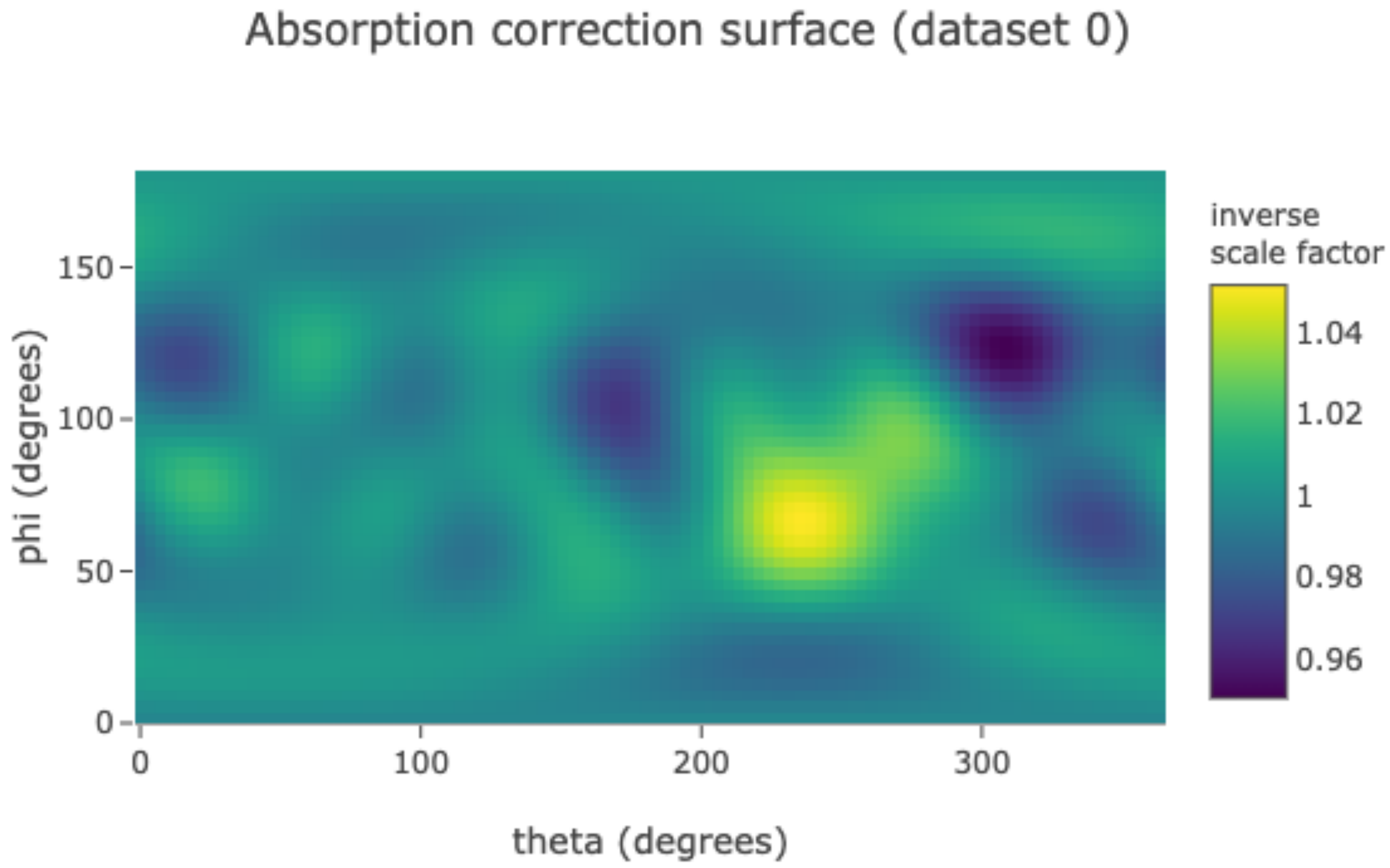
Absorption surface (as spherical harmonics)



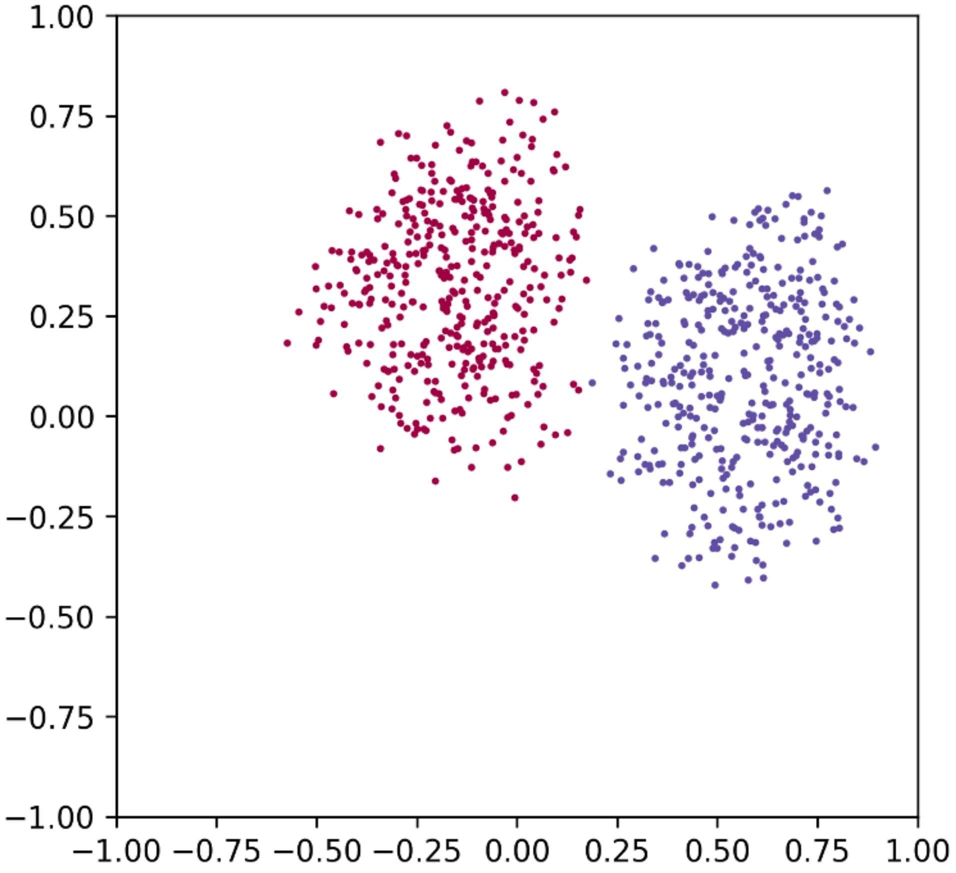
SCALING

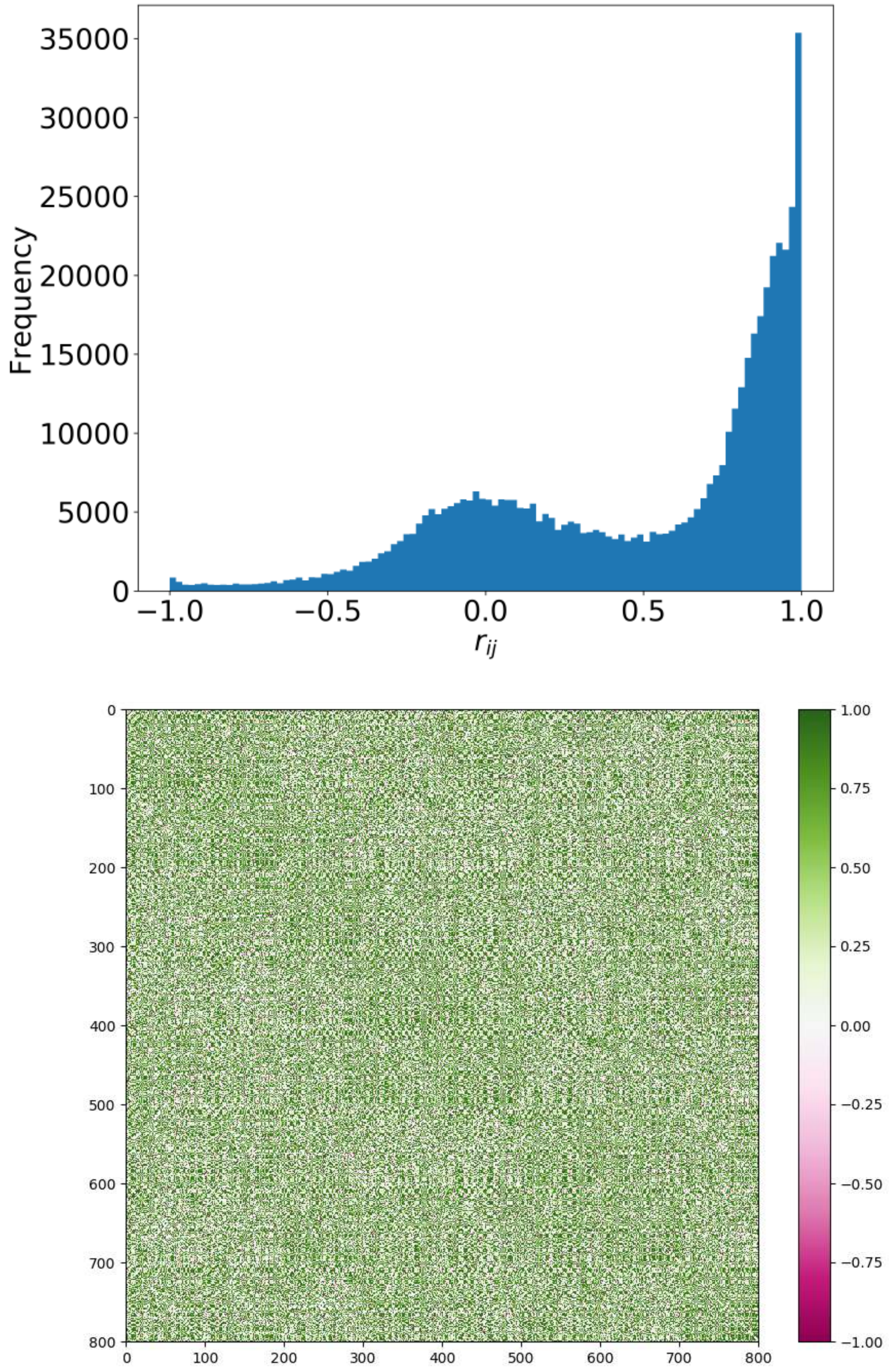


Absorption surface (as spherical harmonics)



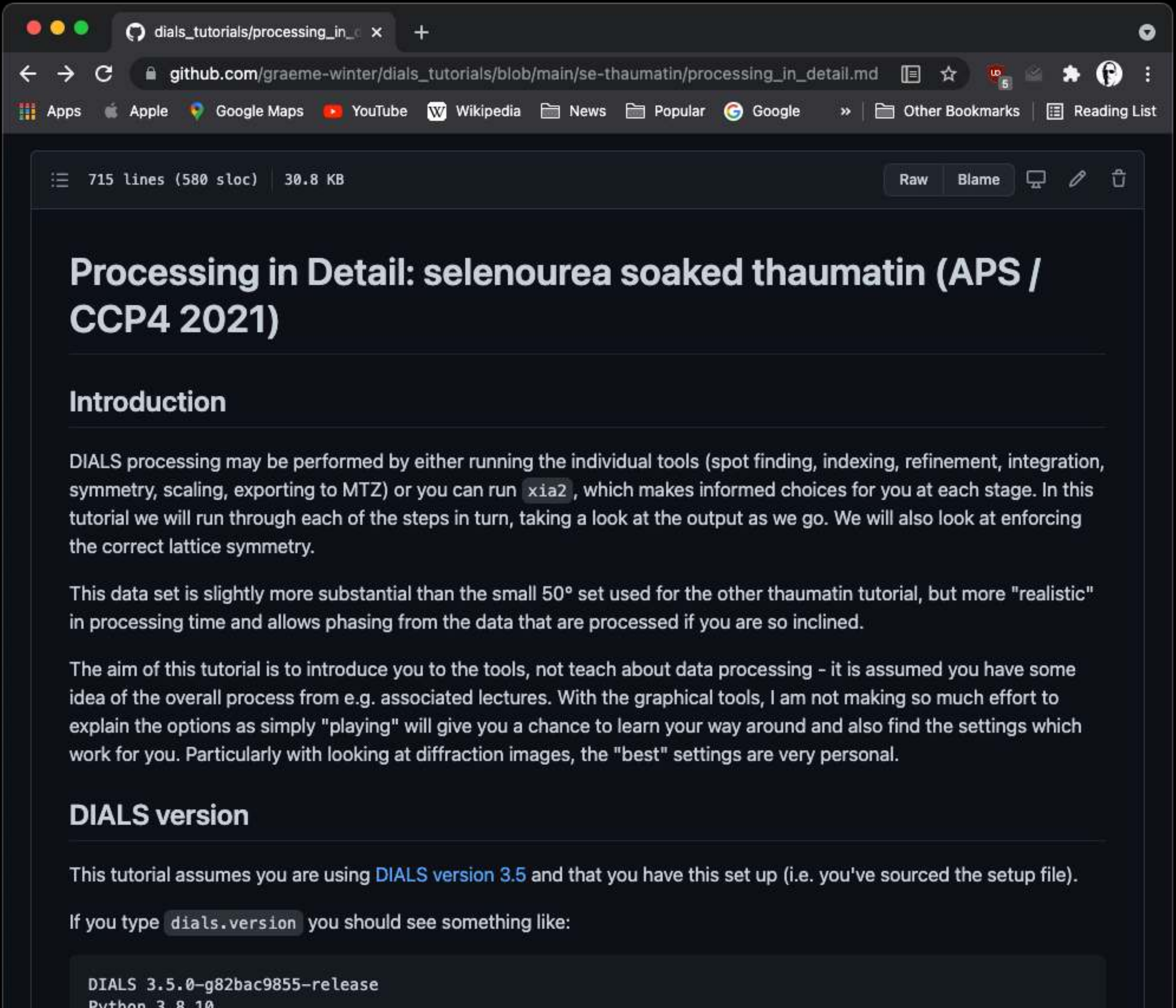
SYMMETRY & ISOMORPHISM - COSYM





COMMENTARY

- 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 - there are no black boxes
- If ever you really want to understand X-ray diffraction, write an integration program 🙂



Processing in Detail: selenourea soaked thaumatin (APS / CCP4 2021)

Introduction

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

This data set is slightly more substantial than the small 50° set used for the other thaumatin tutorial, but more "realistic" in processing time and allows phasing from the data that are processed if you are so inclined.

The aim of this tutorial is to introduce you to the tools, not teach about data processing - it is assumed you have some idea of the overall process from e.g. associated lectures. With the graphical tools, I am not making so much effort to explain the options as simply "playing" will give you a chance to learn your way around and also find the settings which work for you. Particularly with looking at diffraction images, the "best" settings are very personal.

DIALS version

This tutorial assumes you are using DIALS version 3.5 and that you have this set up (i.e. you've sourced the setup file).

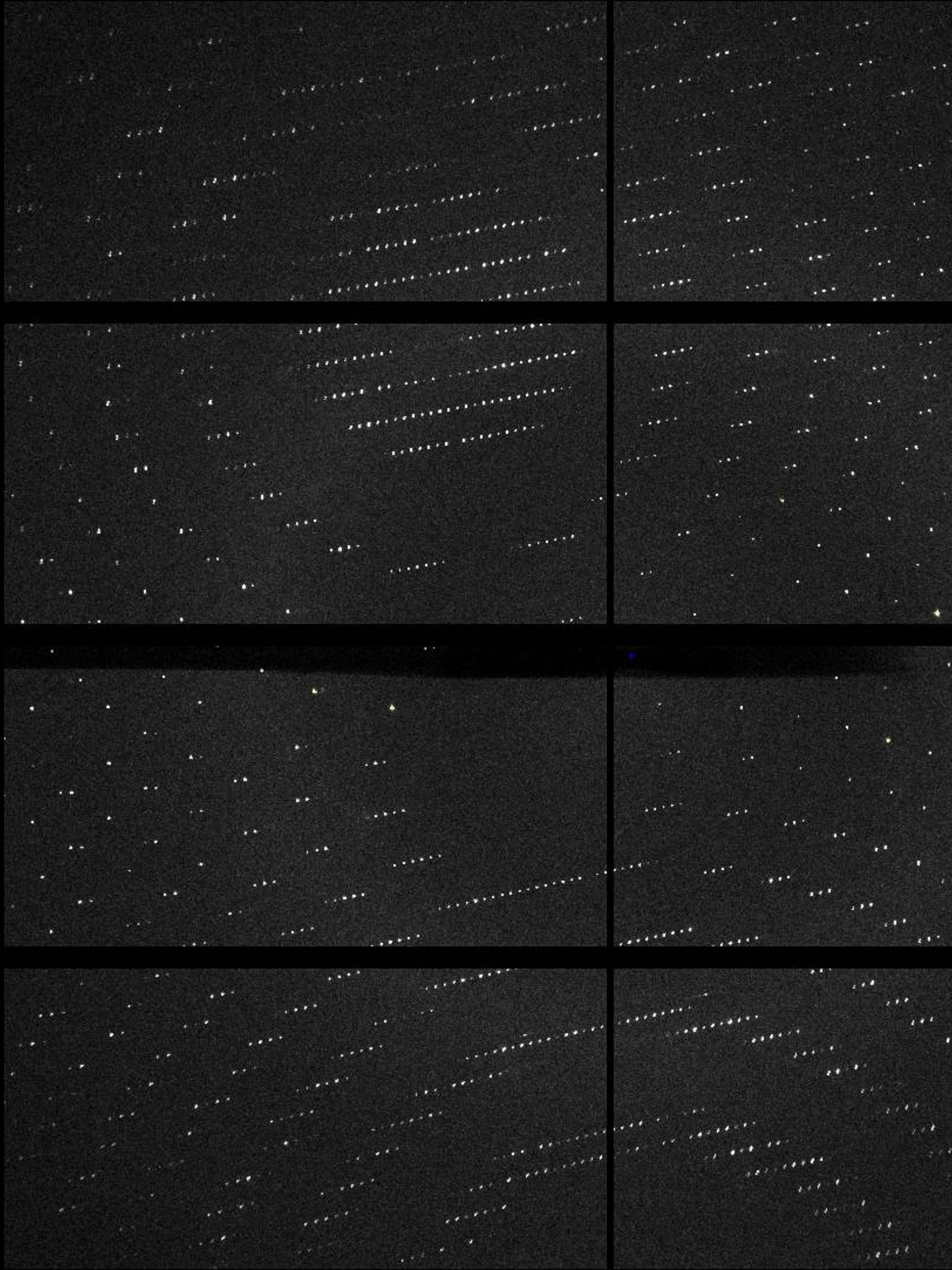
If you type `dials.version` you should see something like:

```
DIALS 3.5.0-g82bac9855-release
Python 3.8.10
```

TUTORIAL

SE-UREA THAUMATIN

- Single sweep of data measured for routine testing
- 1800 x 0.1° images
- Simple to process through to refined structure
- Good for learning how to drive DIALS



SPOT FINDING

```
Histogram of per-image spot count for imageset 0:
234130 spots found on 1800 images (max 4530 / bin)
*     ******
********************              **********             ***
*************************** ************************ *******
************************************************************
************************************************************
************************************************************
************************************************************
************************************************************
************************************************************
************************************************************
1                         image                         1800
```

RECIPROCAL LATTICE VIEW

INDEXING

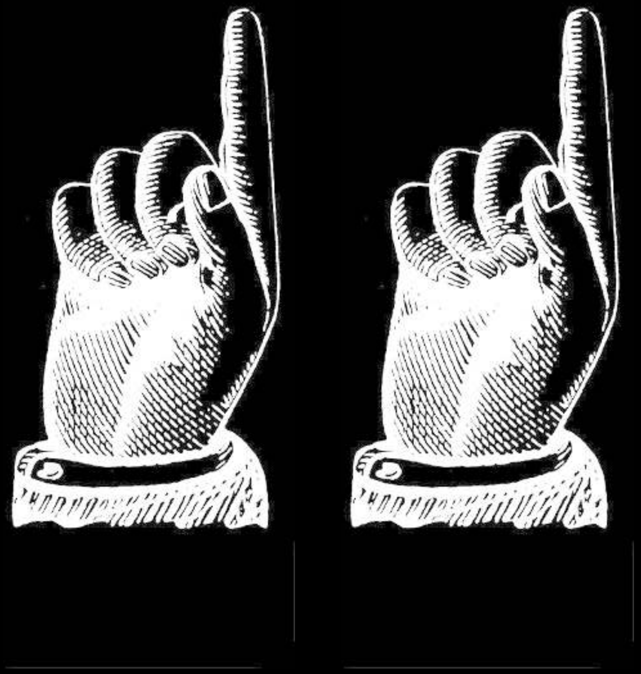
```
+-------+--------+----------+----------+------------+
|   Exp |   Nref |   RMSD_X |   RMSD_Y |     RMSD_Z |
|    id |        |     (px) |     (px) |   (images) |
|-------+--------+----------+----------+------------|
|     0 |  18000 |  0.50174 |  0.69553 |    0.76912 |
+-------+--------+----------+----------+------------+
```

```
+------------+-------------+---------------+-------------+
|   Imageset |   # indexed |   # unindexed | % indexed   |
|------------+-------------+---------------+-------------|
|          0 |      178887 |         55243 | 76.4%       |
+------------+-------------+---------------+-------------+
```

LOOK AT THE IMAGES

REFINE BRAVAIS SETTINGS

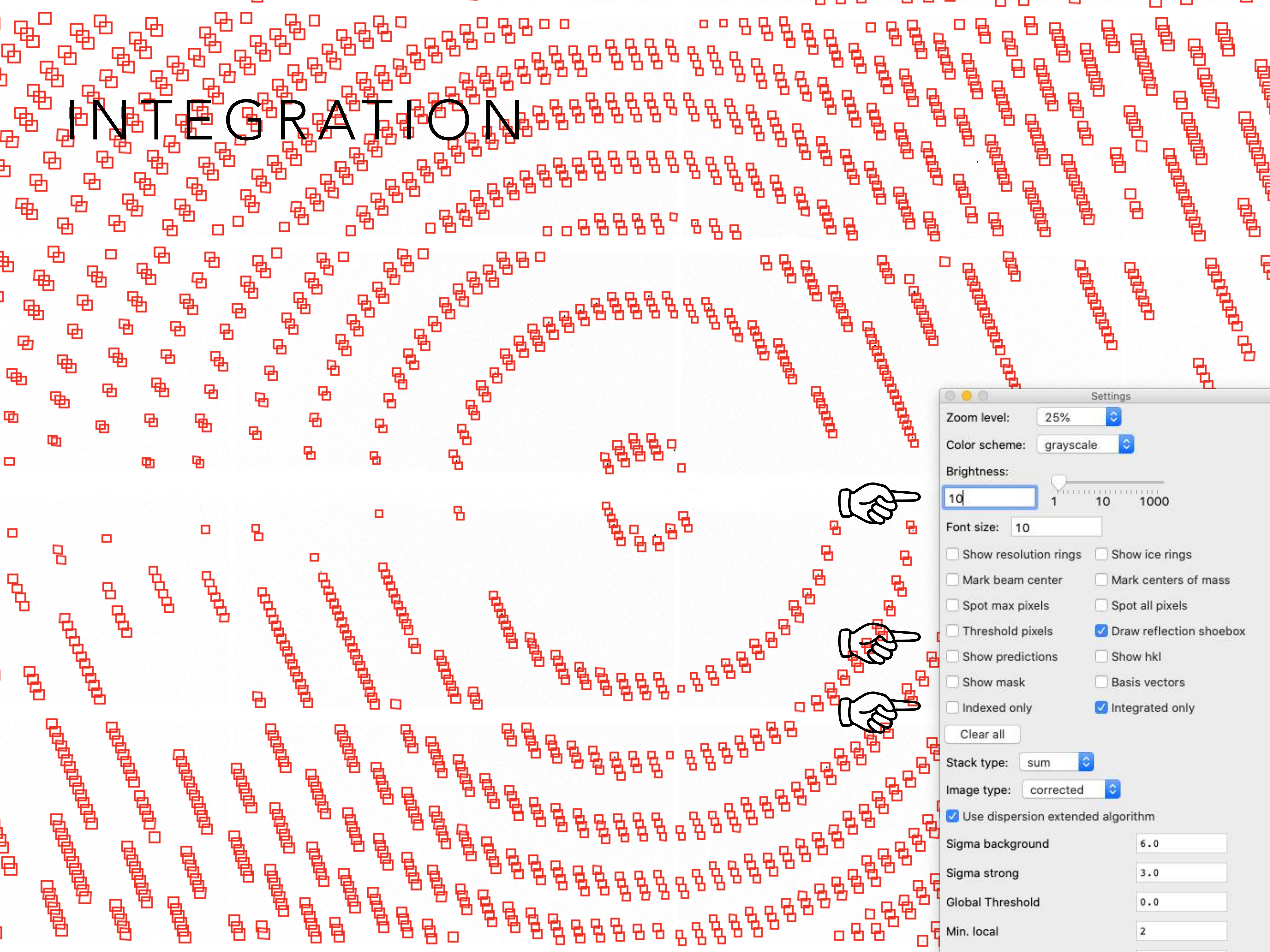
```
+------------+-------------+--------+--------------+----------+-----------+------------------------------
|   Solution |   Metric fit |   rmsd | min/max cc   |   #spots | lattice   | unit_cell
|------------+-------------+--------+--------------+----------+-----------+------------------------------
|   *      9 |      0.1075 |  0.083 | 0.941/0.944  |    18000 | tP        | 57.86  57.86 149.97  90.00  9
|   *      8 |      0.1075 |  0.085 | 0.935/0.941  |    18000 | oC        | 81.81  81.87 149.99  90.00  9
|   *      7 |      0.1032 |  0.084 | 0.941/0.941  |    18000 | mC        | 81.87  81.82 149.99  90.00  9
|   *      6 |      0.1075 |  0.084 | 0.935/0.935  |    18000 | mC        | 81.80  81.85 149.97  90.00  9
|   *      5 |      0.0531 |  0.064 | 0.940/0.957  |    18000 | oP        | 57.85  57.95 150.03  90.00  9
|   *      4 |      0.0531 |  0.064 | 0.940/0.940  |    18000 | mP        | 57.95  57.84 150.02  90.00  8
|   *      3 |      0.0399 |  0.062 | 0.957/0.957  |    18000 | mP        | 57.84  57.95 150.01  90.00  9
|   *      2 |      0.0391 |  0.067 | 0.941/0.941  |    18000 | mP        | 57.85 150.03  57.95  90.00  9
|   *      1 |      0      |  0.065 | -/-          |    18000 | aP        | 57.84  57.95 150.02  89.99  9
+------------+-------------+--------+--------------+----------+-----------+------------------------------
```



REFINEMENT (WITH SCAN VARYING)

```
From indexing:
+-------+--------+----------+----------+------------+
|   Exp |   Nref |   RMSD_X |   RMSD_Y |     RMSD_Z |
|    id |        |     (px) |     (px) |   (images) |
|-------+--------+----------+----------+------------|
|     0 |  18000 |  0.50174 |  0.69553 |    0.76912 |
+-------+--------+----------+----------+------------+
From refinement:
+-------+--------+----------+----------+------------+
|   Exp |   Nref |   RMSD_X |   RMSD_Y |     RMSD_Z |
|    id |        |     (px) |     (px) |   (images) |
|-------+--------+----------+----------+------------|
|     0 | 114818 |  0.31991 |  0.50414 |    0.45294 |
+-------+--------+----------+----------+------------+
```

INTEGRATION



Zoom level: 25%

Color scheme: grayscale

Brightness: 10

Font size: 10

☐ Show resolution rings ☐ Show ice rings
☐ Mark beam center ☐ Mark centers of mass
☐ Spot max pixels ☐ Spot all pixels
☐ Threshold pixels ☑ Draw reflection shoebox
☐ Show predictions ☐ Show hkl
☐ Show mask ☐ Basis vectors
☐ Indexed only ☑ Integrated only

Stack type: sum

Image type: corrected

☑ Use dispersion extended algorithm

Sigma background 6.0

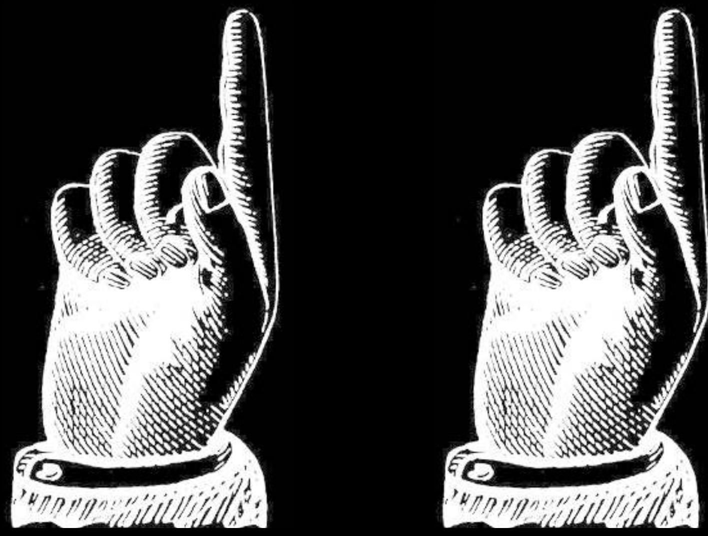
Sigma strong 3.0

Global Threshold 0.0

Min. local 2

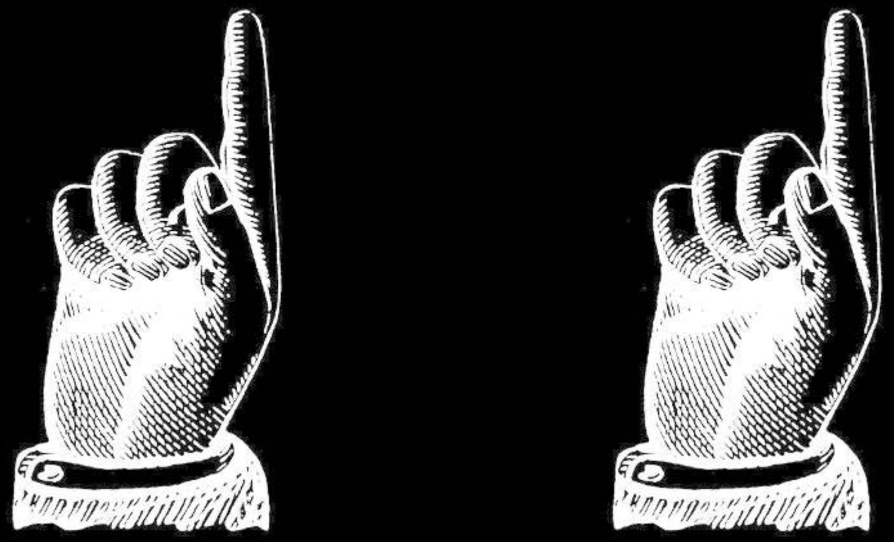
SYMMETRY

```
+--------------+--------+------+--------+-----+---------------+
|   likelihood |   Z-CC |   CC |      N |     | Operator      |
|--------------+--------+------+--------+-----+---------------|
|        0.94  |   9.9  | 0.99 | 362986 | *** | 1 |(0, 0, 0)  |
|        0.94  |   9.92 | 0.99 | 751640 | *** | 4 |(0, 0, 1)  |
|        0.941 |   9.92 | 0.99 | 373768 | *** | 2 |(1, 0, 0)  |
|        0.941 |   9.93 | 0.99 | 393952 | *** | 2 |(0, 1, 0)  |
|        0.94  |   9.91 | 0.99 | 364832 | *** | 2 |(0, 0, 1)  |
|        0.94  |   9.9  | 0.99 | 389620 | *** | 2 |(1, 1, 0)  |
|        0.94  |   9.91 | 0.99 | 367896 | *** | 2 |(-1, 1, 0) |
+--------------+--------+------+--------+-----+---------------+
```

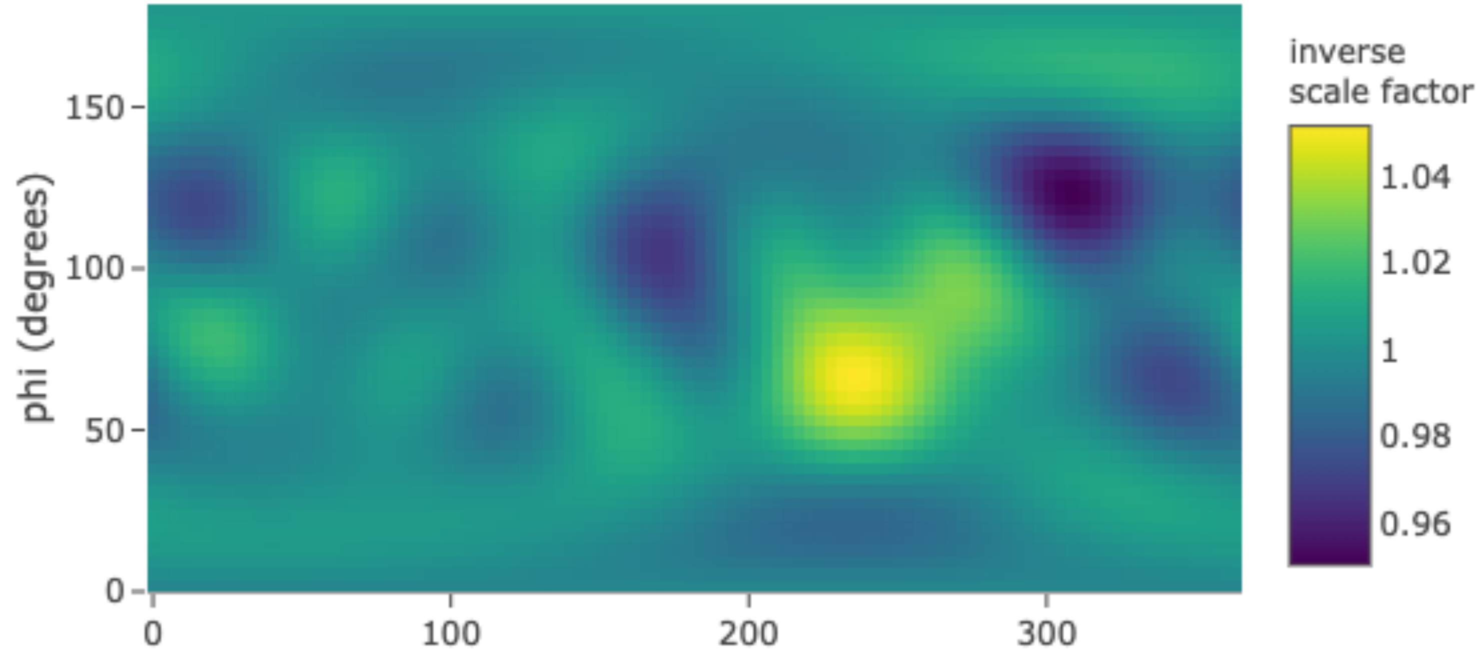


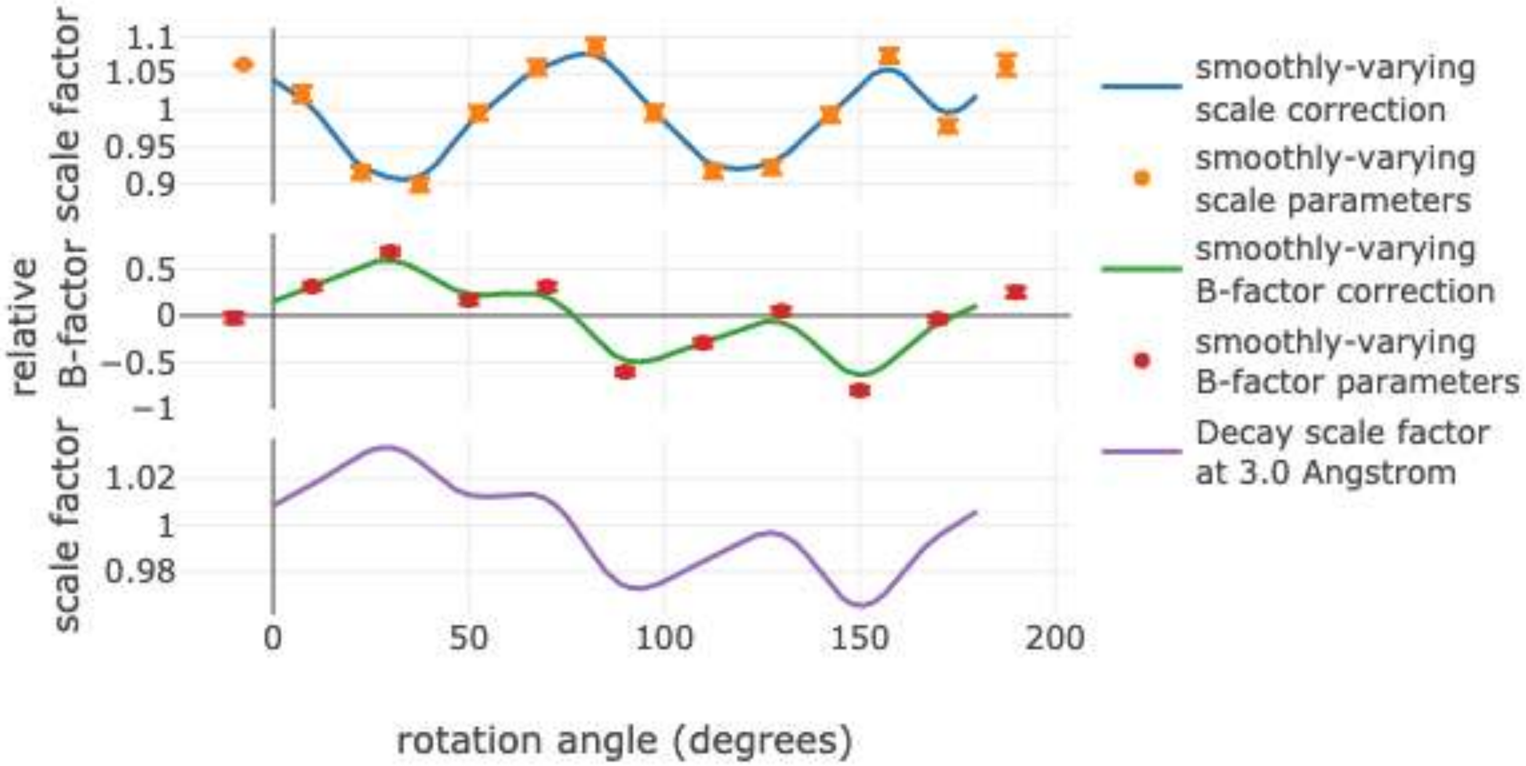
SYMMETRY

```
+-------------------+-----+--------------+----------+--------+--------+------+-------+---------+----------
| Patterson group   |     |   Likelihood |   NetZcc |   Zcc+ |   Zcc- |   CC |   CC- |   delta | Reindex
|-------------------+-----+--------------+----------+--------+--------+------+-------+---------+----------
| P 4/m m m         | *** |        0.999 |     9.91 |   9.91 |   0    | 0.99 |  0    |     0.1 | a,b,c
| P m m m           |     |        0     |     0.01 |   9.92 |   9.91 | 0.99 |  0.99 |     0.1 | a,b,c
| C m m m           |     |        0     |    -0.02 |   9.91 |   9.92 | 0.99 |  0.99 |     0.1 | a+b,a-b,-
| P 4/m             |     |        0     |    -0.01 |   9.91 |   9.92 | 0.99 |  0.99 |     0.1 | a,b,c
| P 1 2/m 1         |     |        0     |     0    |   9.91 |   9.91 | 0.99 |  0.99 |     0   | a,b,c
| P 1 2/m 1         |     |        0     |     0    |   9.91 |   9.91 | 0.99 |  0.99 |     0.1 | b,-a,c
| C 1 2/m 1         |     |        0     |    -0.01 |   9.91 |   9.92 | 0.99 |  0.99 |     0.1 | a-b,-a-b,
| P 1 2/m 1         |     |        0     |    -0.01 |   9.91 |   9.92 | 0.99 |  0.99 |     0   | -a,c,b
| C 1 2/m 1         |     |        0     |    -0.01 |   9.9  |   9.92 | 0.99 |  0.99 |     0.1 | a+b,a-b,-
| P -1              |     |        0     |    -0.01 |   9.9  |   9.91 | 0.99 |  0.99 |     0   | a,b,c
+-------------------+-----+--------------+----------+--------+--------+------+-------+---------+----------
```

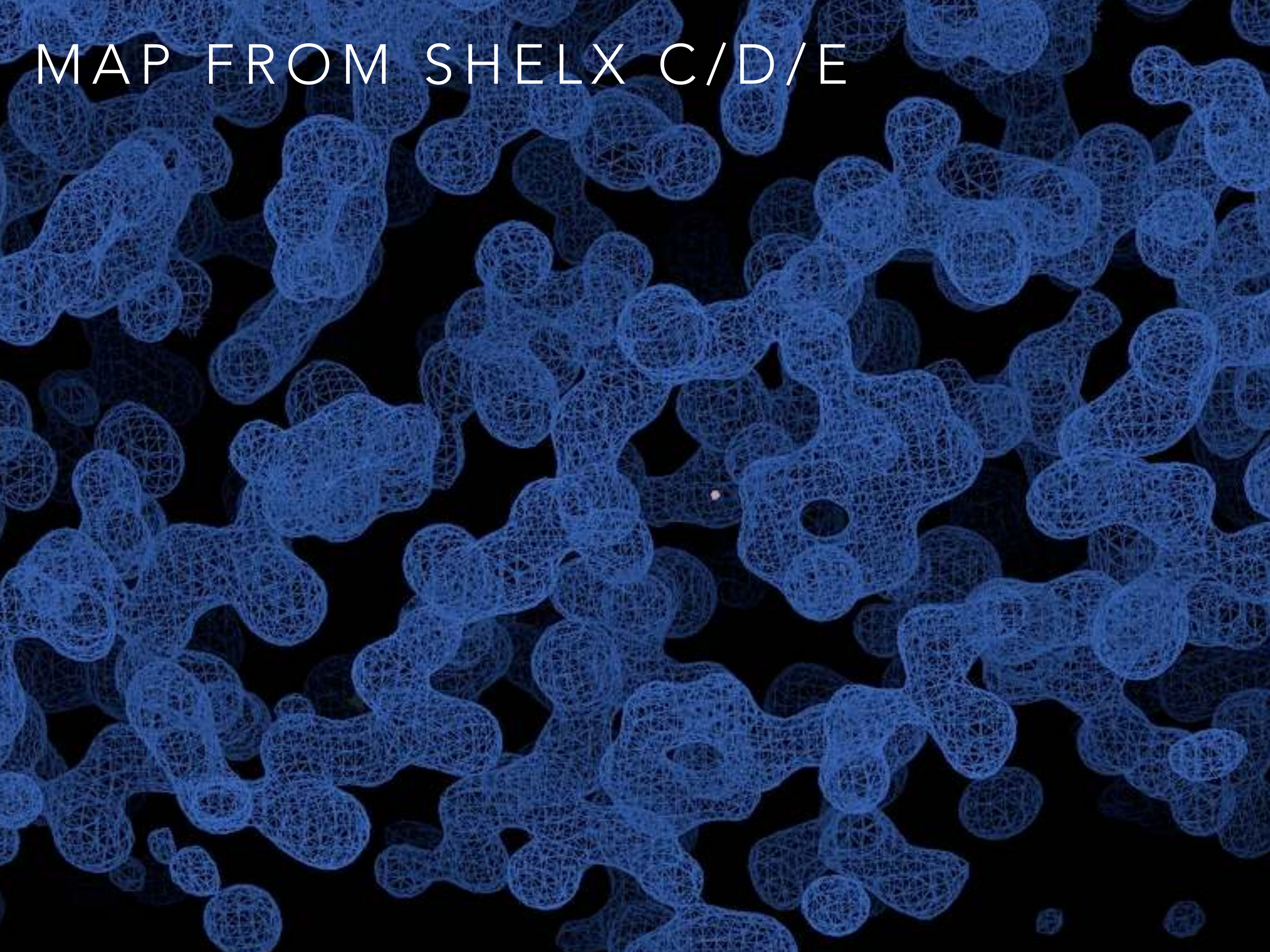


SCALING


Absorption correction surface (dataset 0)


Smoothly varying corrections (dataset 0)

MAP FROM SHELX C/D/E



DIALS.GITHUB.IO
XIA2.GITHUB.IO