Data processing with DIALS

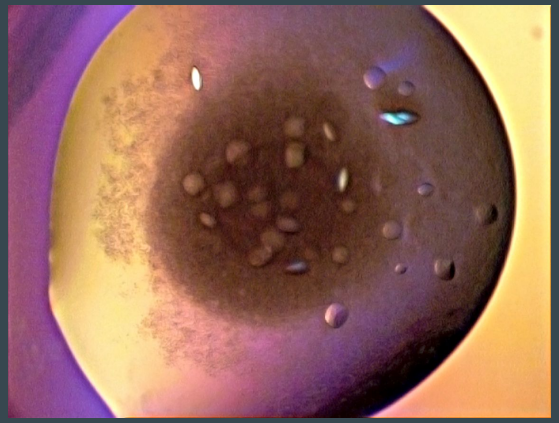
James Parkhurst

IFSC/CCP4 Molecular Crystallography School

November 2018

What are we doing and why are we doing it?

We have 1 or more crystals and want to determine the structure of our protein



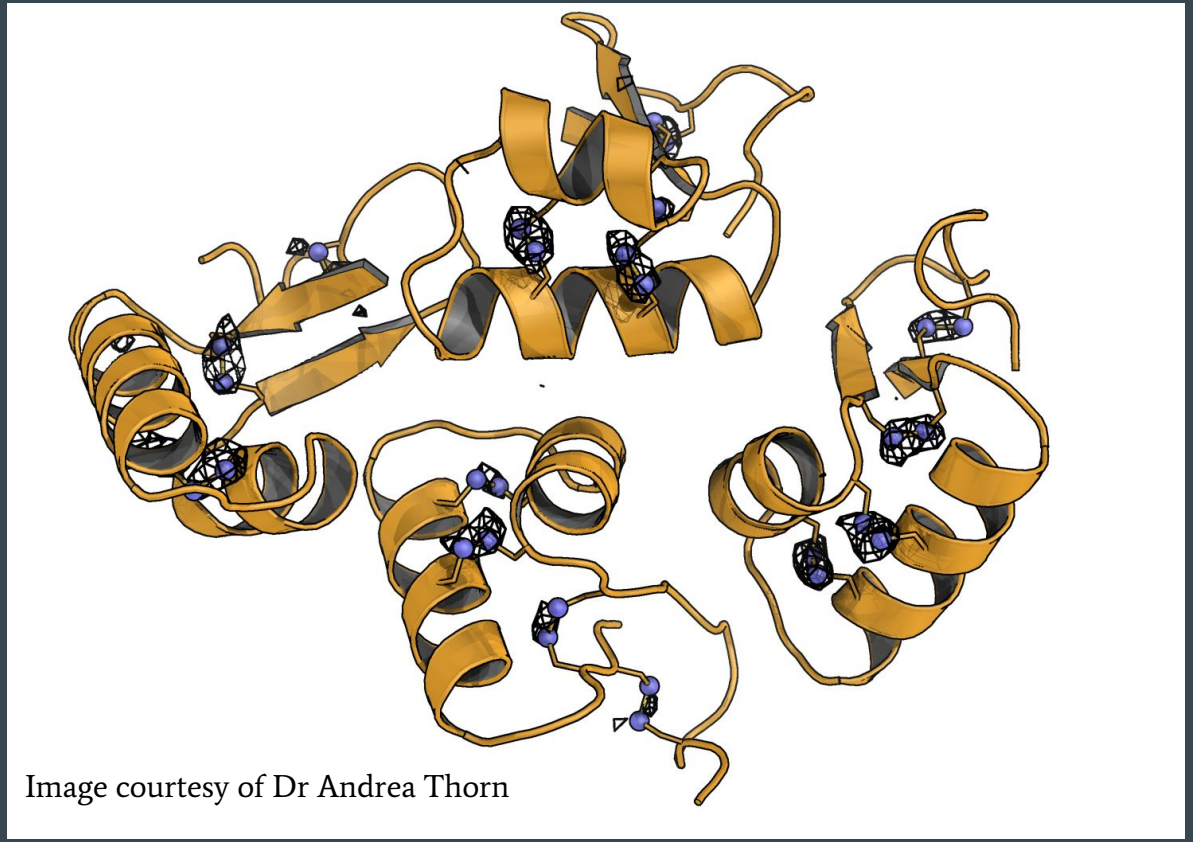
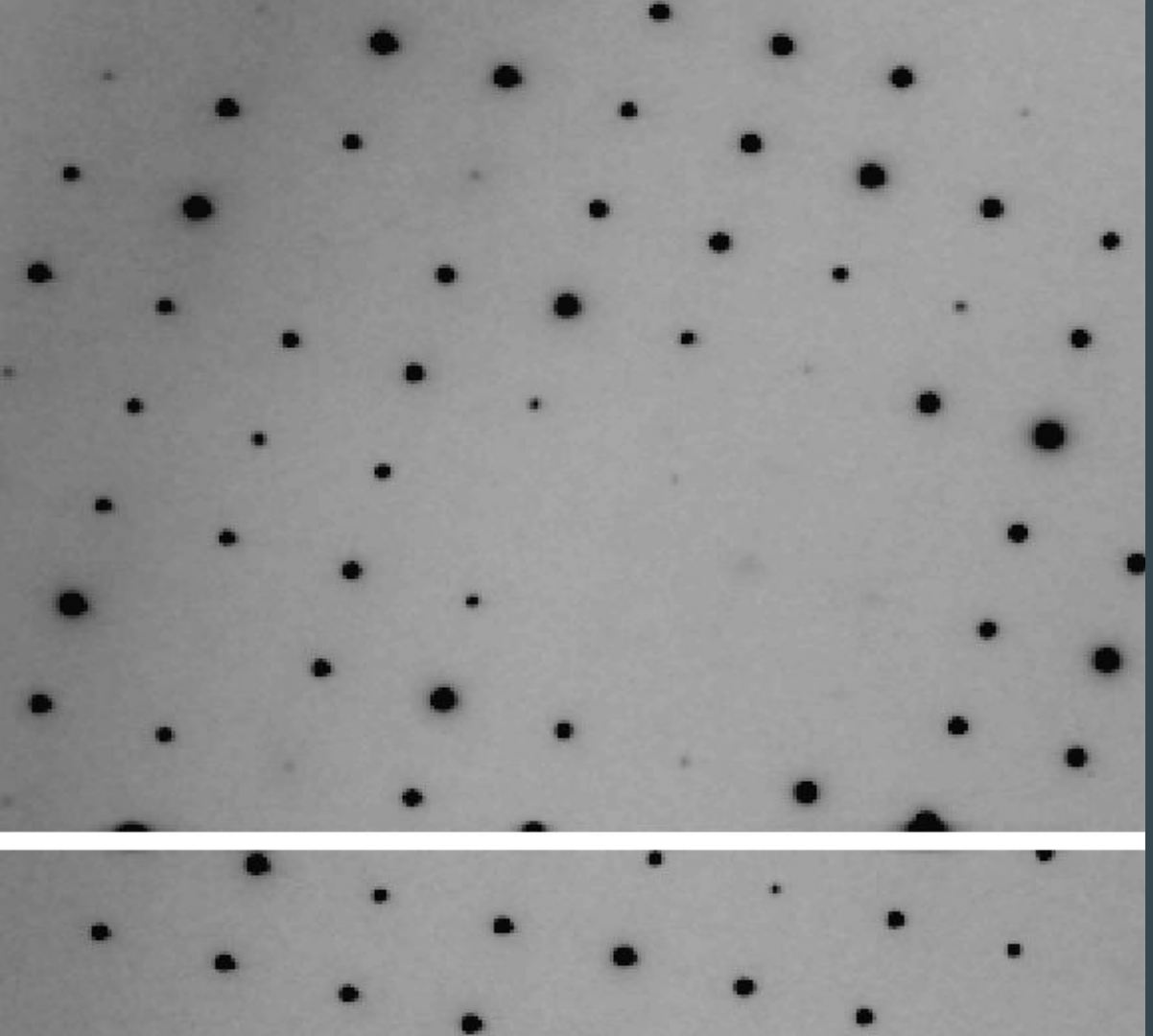
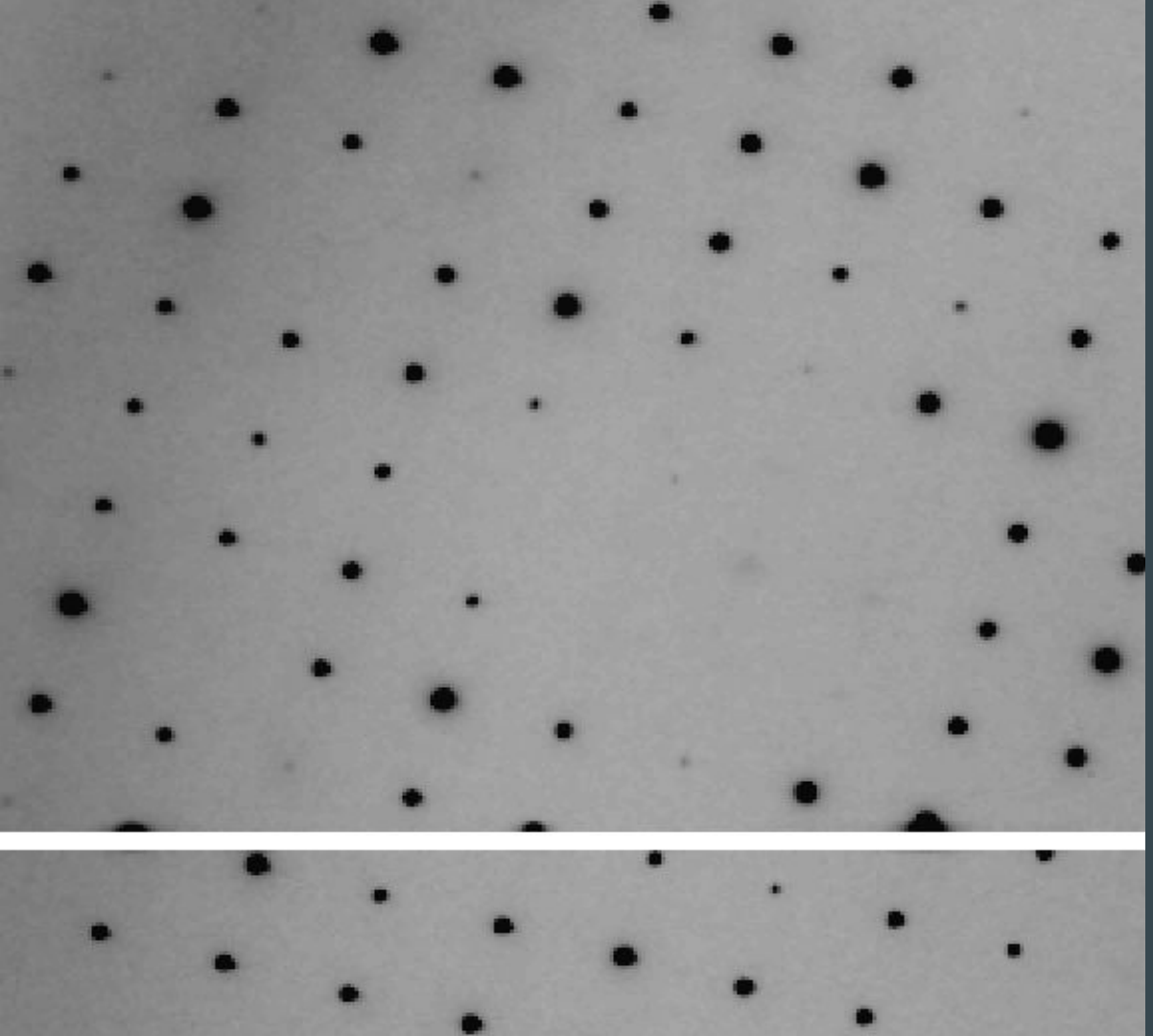


Image courtesy of Dr Andrea Thorn



However, the data we collect are diffraction images!

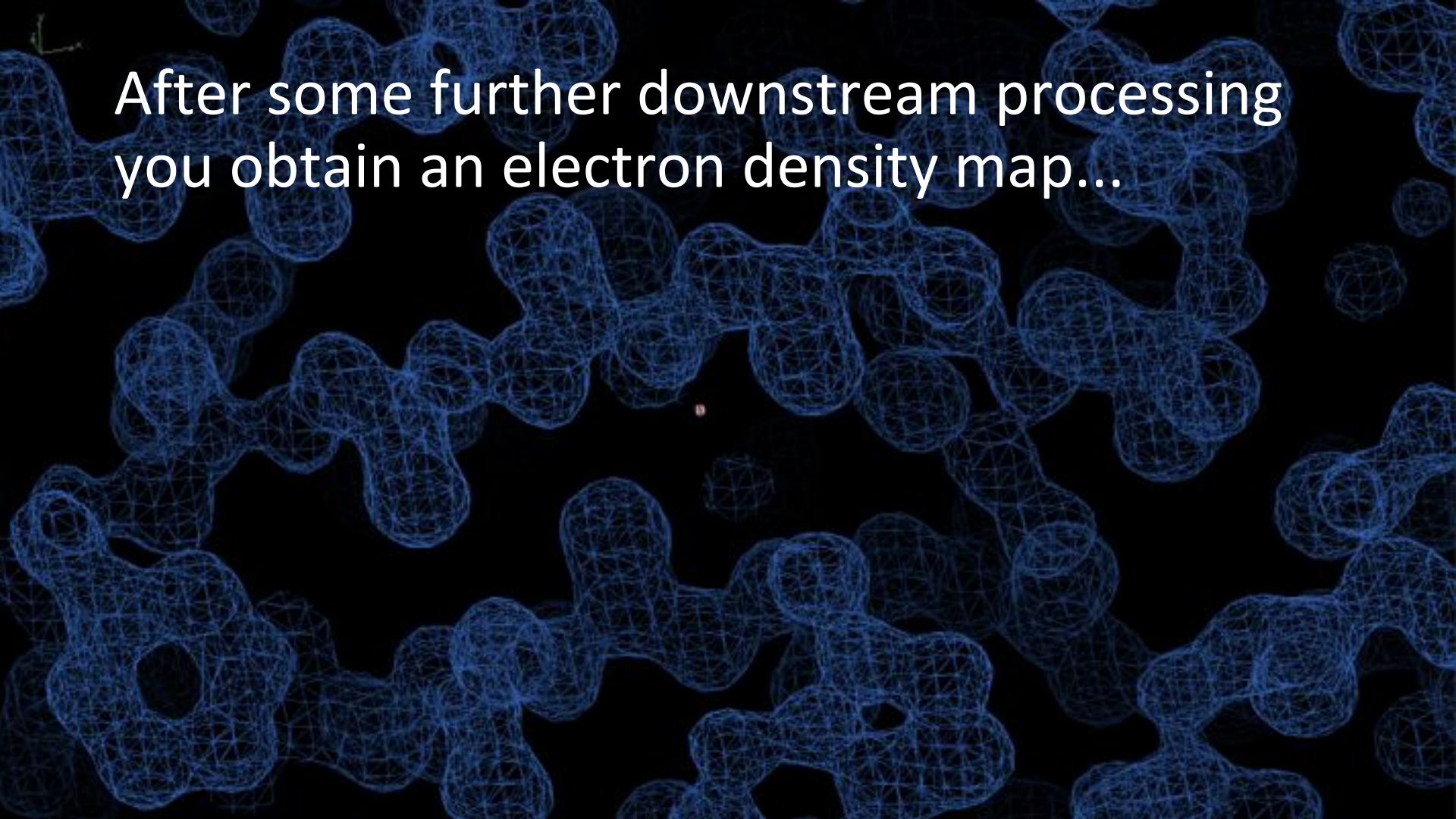


Compute the intensity of each Bragg spot in a set of diffraction images

$$|F_{hkl}| = \sqrt{\frac{K I_{hkl}}{Lp}}$$

K = constant for given crystal
L = Lorentz factor
p = polarization factor

After some further downstream processing you obtain an electron density map...



And after some model building you can determine your protein structure...

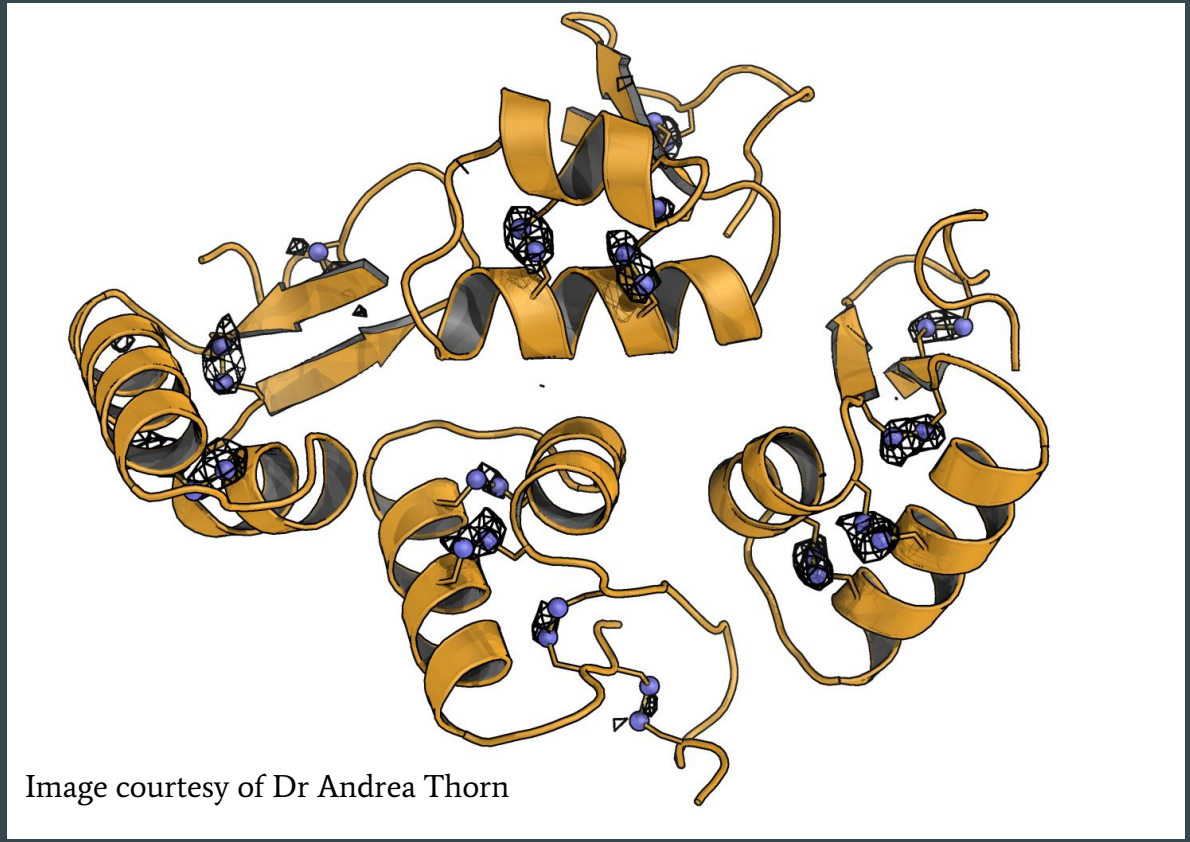
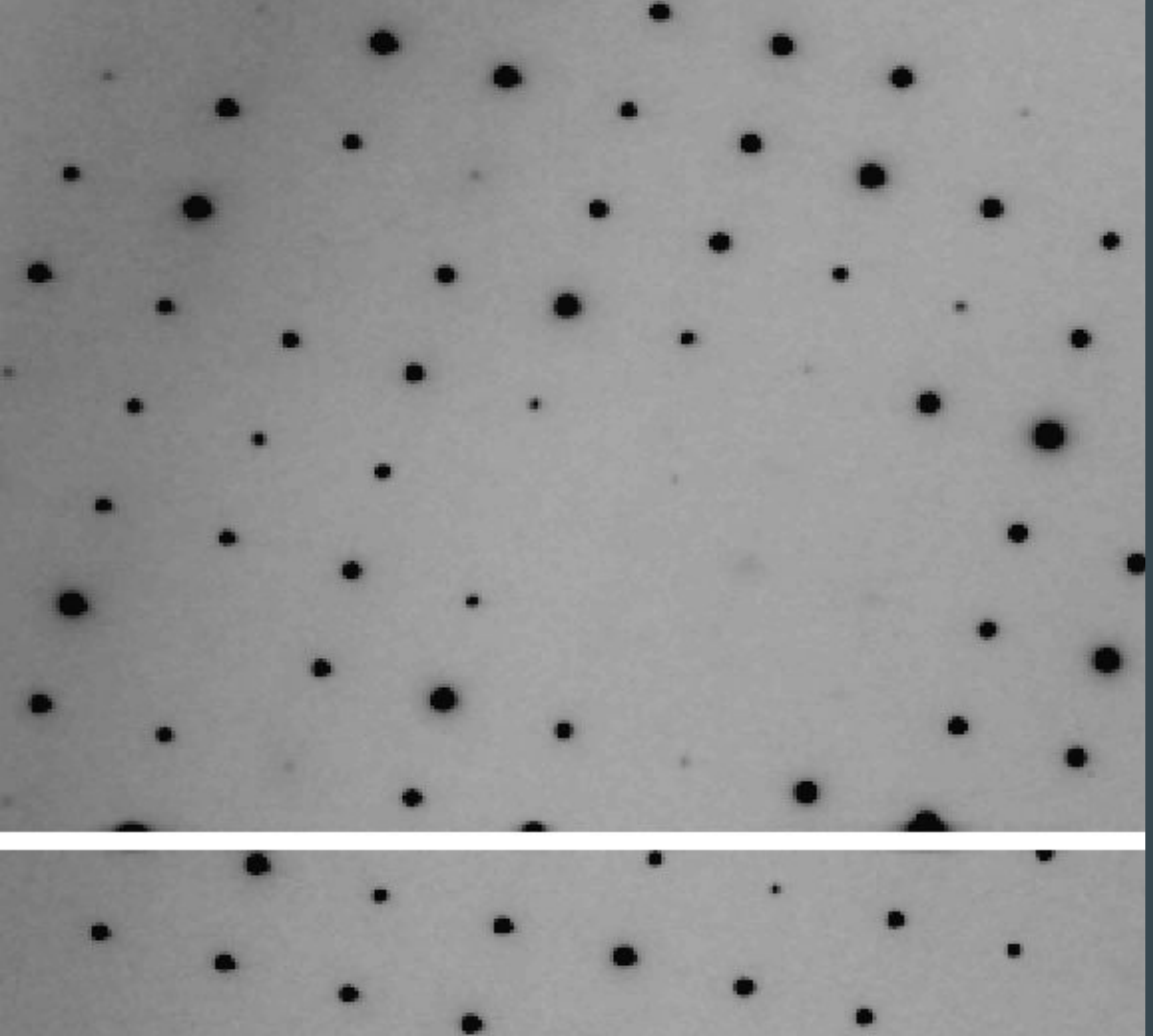


Image courtesy of Dr Andrea Thorn

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

Electron density at every point depends on the intensity of every reflection. We need to measure our intensities as well as possible!

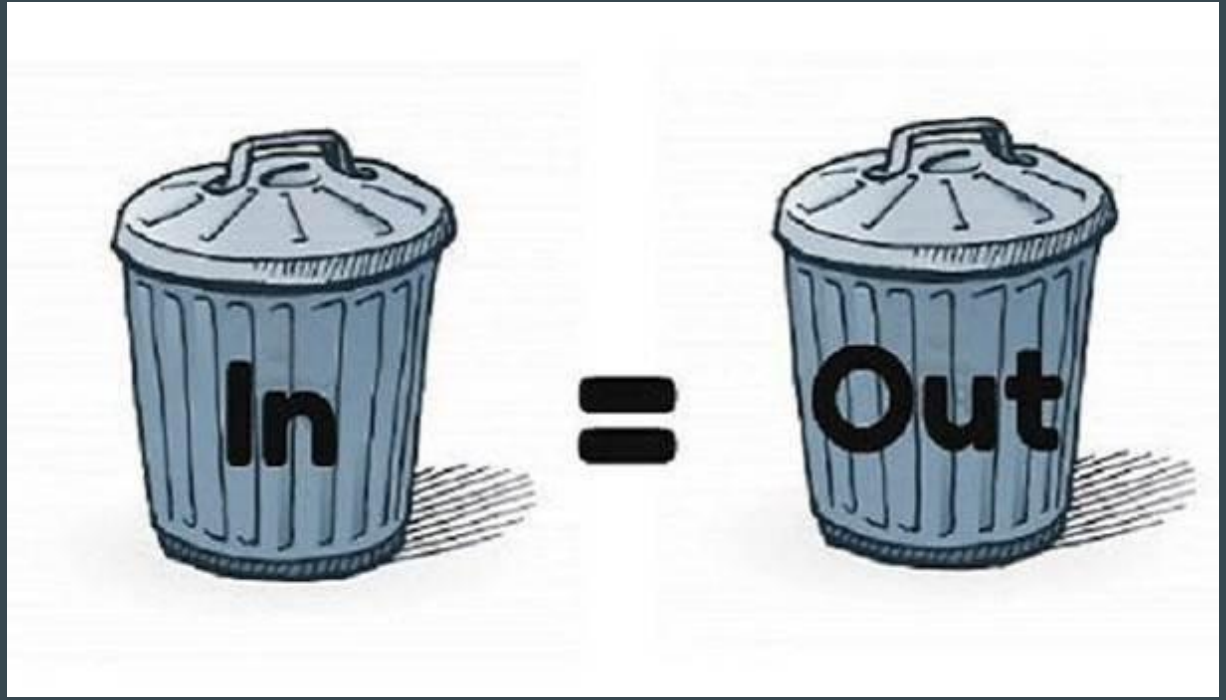


To compute spot intensities you need to predict their positions.

To predict their positions you need to know the crystal unit cell and space group.

To compute the unit cell and space group you need to find strong spots.

Warning: garbage in, garbage out



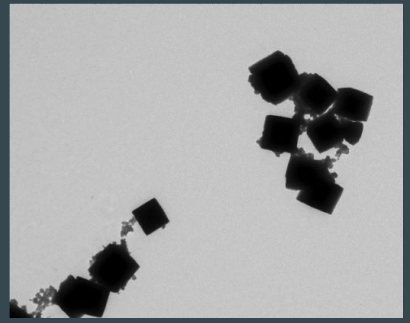
Data collection is the last experimental stage; if you collect bad data you are stuck with it! Data processing programs won't be able to rescue you!

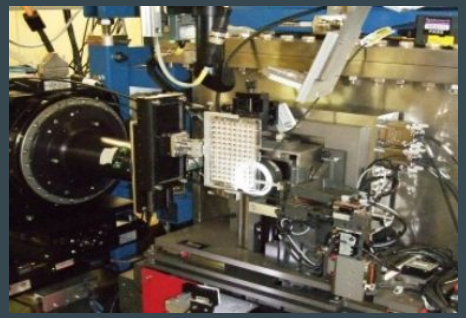
DIALS overview

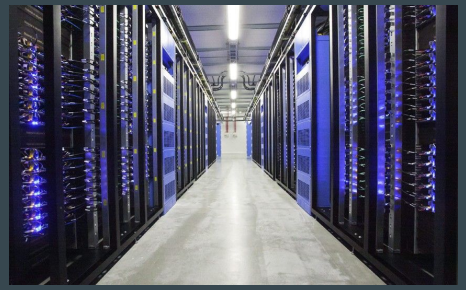
What is DIALS

Diffraction **I**ntegration for **A**dvanced **L**ight **S**ources

- Project began in late 2011 at Diamond Light Source and CCP4
- Additional funding by BioStruct-X and now Wellcome Trust
- Now an international collaborative development
- Aim to develop new data processing software to meet modern challenges







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 contr
package *XDS* for
described in the con
include automatic d
range and recognitio
Moreover, the limita
number of correction
pixel contents have b
been restructured fo
and completeness of
measurement.

1. Functional speci

The program packag
developed for the re
recorded on a plana
monochromatic X-ra
XDS accepts a
rotation images fron
and multiwire area
metrics and produce
of the reflections occ
way. The program a
positive amount of c
incident beam and cr
imposes no limitati
directions of the rot
oscillation range cov

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The finer things in

J. W. Pflugrath

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

Correspondence e-mail: jwp@msc.com

X-ray diffraction images from
sensitive detectors can be cha
depending on whether the rotatio
is greater than or less than the c
The expectations and consequen
and thin images in terms of spa
X-ray background and $I/\sigma(I)$
software suite for processing
introduced, and results from *d*
those from another popular pacl

1. Introduction

Two-dimensional position-sensiti
for many years in X-ray diffract
cular, data from crystals of macr
oligonucleotides and their co
acquired with an area detector
obsolete), a multi-wire system
recently commercialized char
coupled to a phosphor-coated fil
detectors, the crystal, centered in
oscillated around a single axis th
~2.0°, while counts from diffrac
for a specified time. At the en
detector is read out and the cou
two-dimensional array with each
to a distinct position on the d

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The integrati

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
produce from a set of
with their associated
uncertainties), togeth
crystal unit-cell param
reliable, but should i
intervention. The pro
three stages. The first
parameters and the c
parameters may indic
The second step is to r
parameters and also t
known as post-refiner
images, which consists
reflections on each in
intensity of each refle
out while simultaneou
parameters. Basic fea
each of these three s
with reference to the

1. Introduction

The collection of mac
gone dramatic advan
advent of two-dimensi
and CCDs, crystal cry
monochromatic and

Centre National de la Recherche Scientifique
Université Paris-Sud

Laboratoire pour l'Utilisation du Rayonnement Electromagnétique

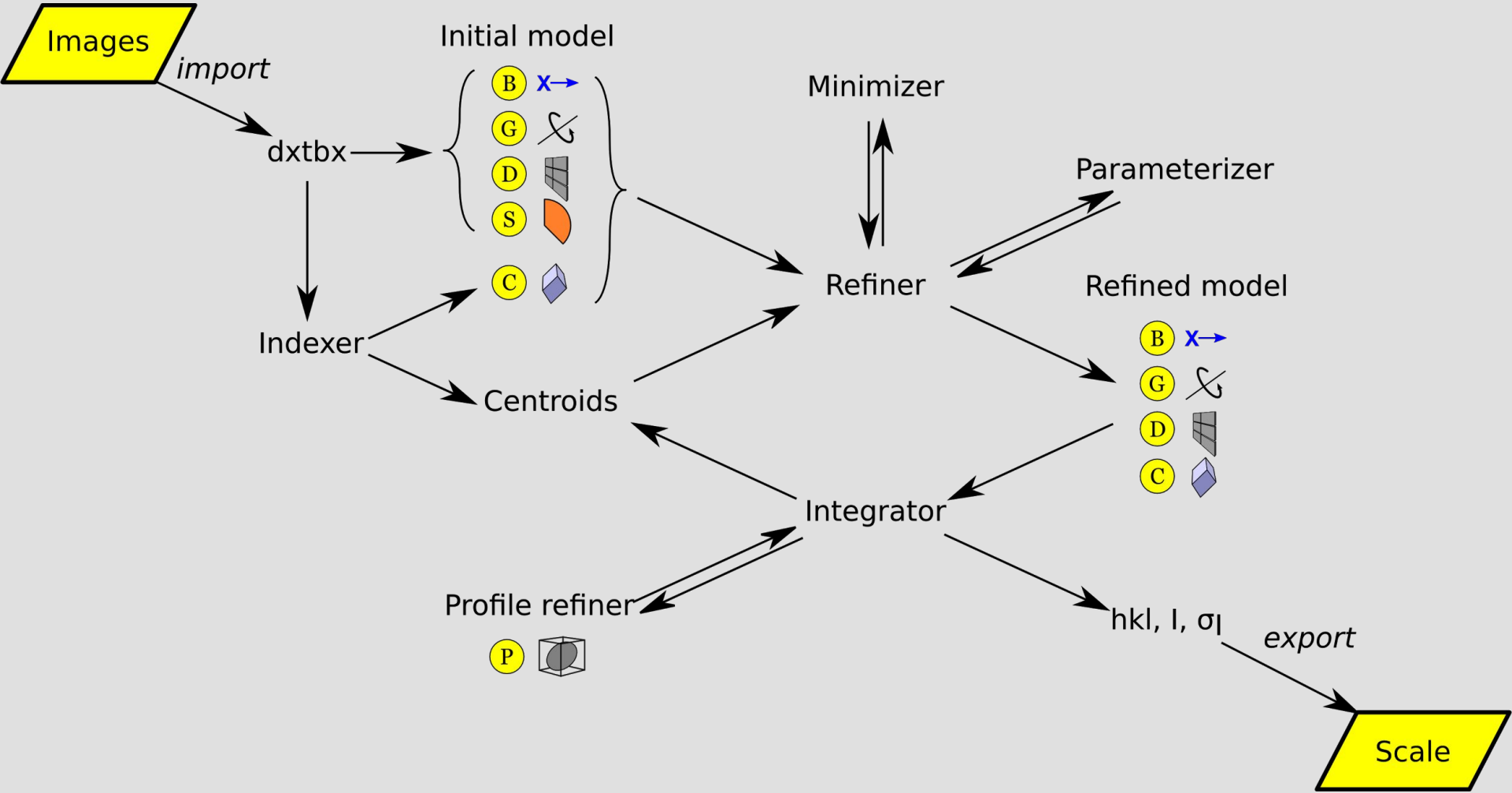
Proceedings

of the EEC Cooperative Workshop

on Position-Sensitive Detector Software

(Phases I & II)

held at L.U.R.E. from May 26 to June 7, 1986.


Images
import
dxtbx
Initial model
B
G
D
S
C
Indexer
Centroids
Minimizer
Refiner
Parameterizer
Refined model
B
G
D
C
Integrator
Profile refiner
P
hkl, I, σI
export
Scale

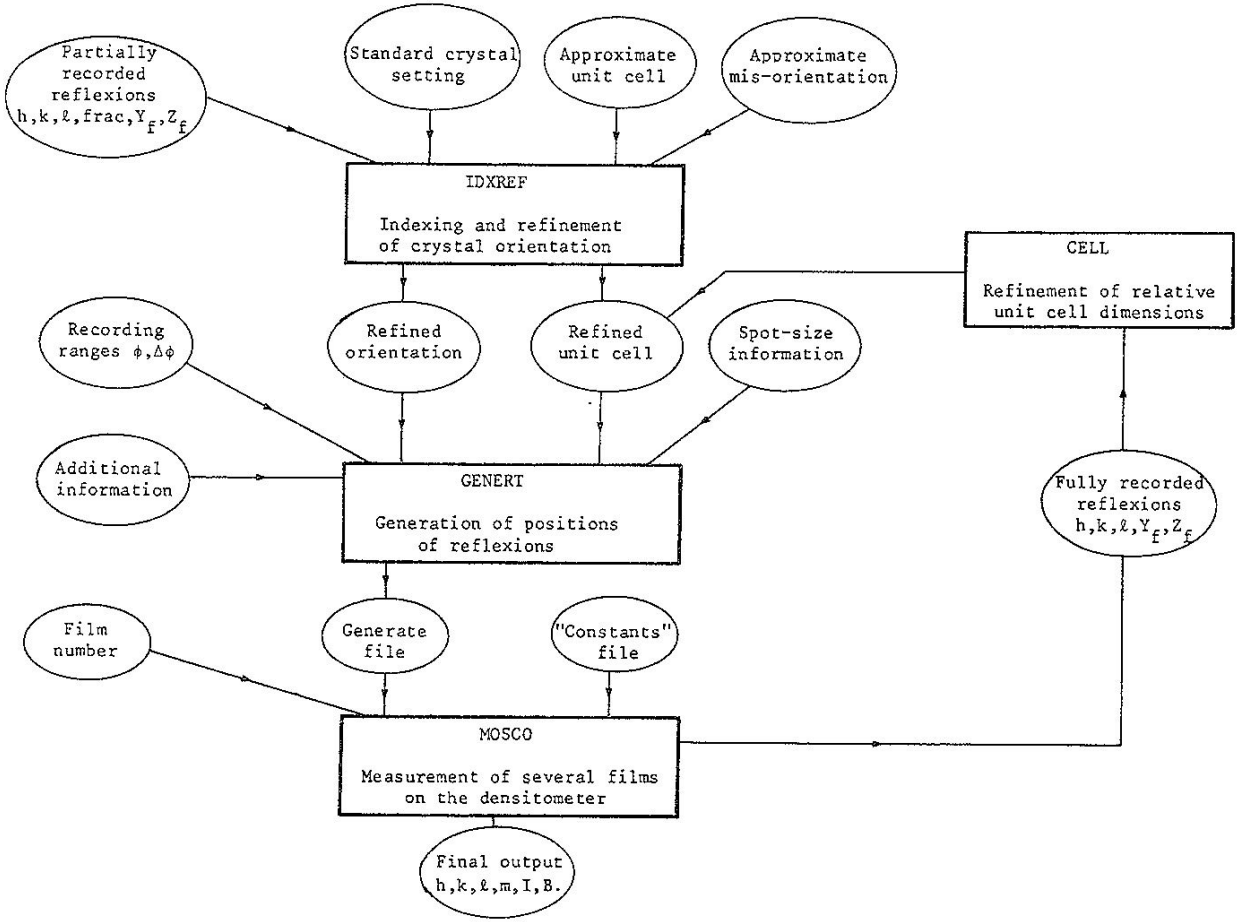
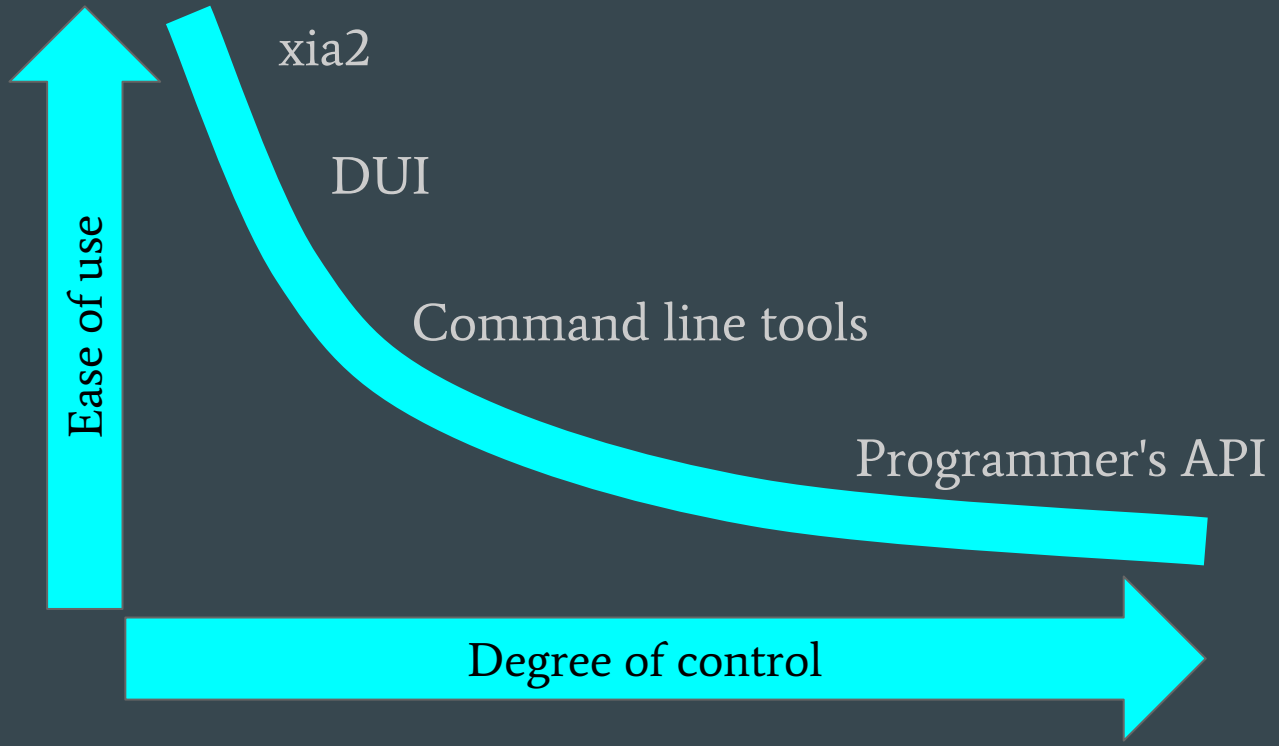


Fig. 10.1. Simplified flow-diagram of the Cambridge system, showing the inter-relation of the component programs, IDXREF, GENERT, MOSCO and CELL.

Philosophy: toolbox

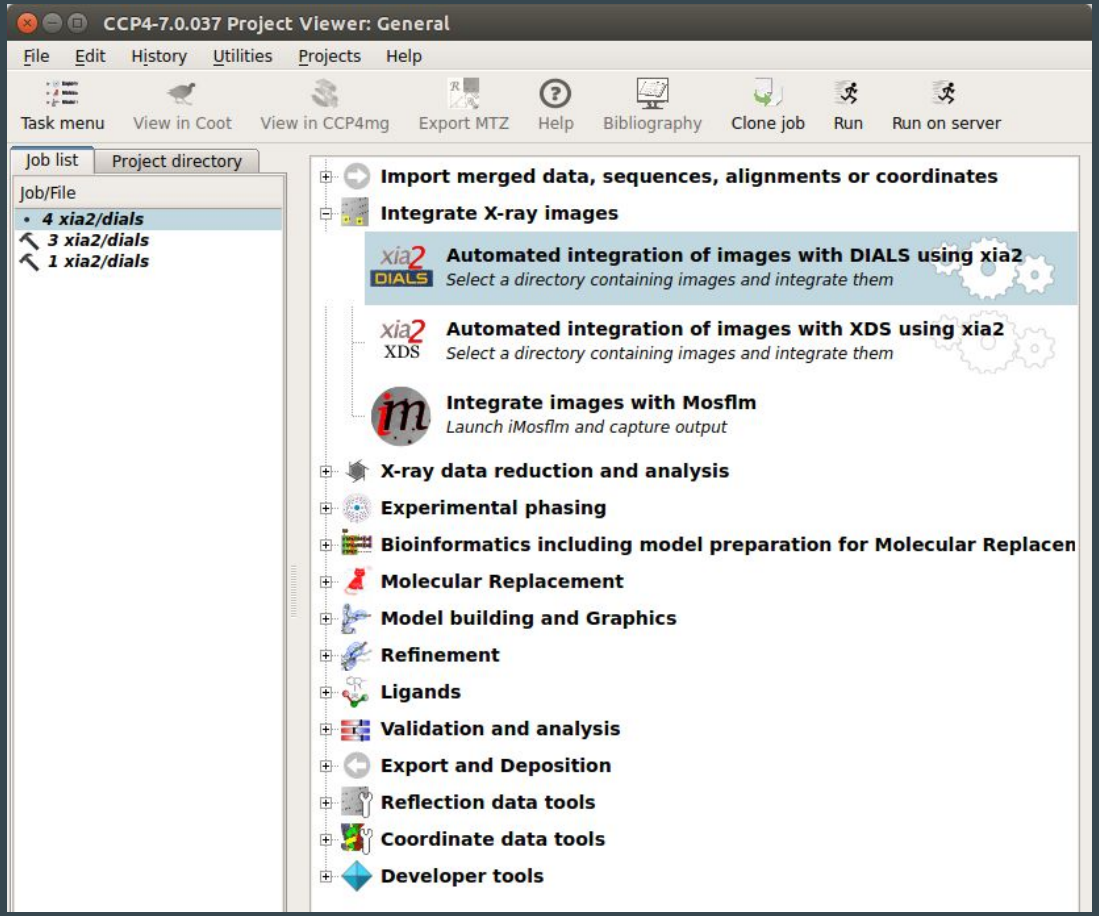
Philosophy: levels of interaction



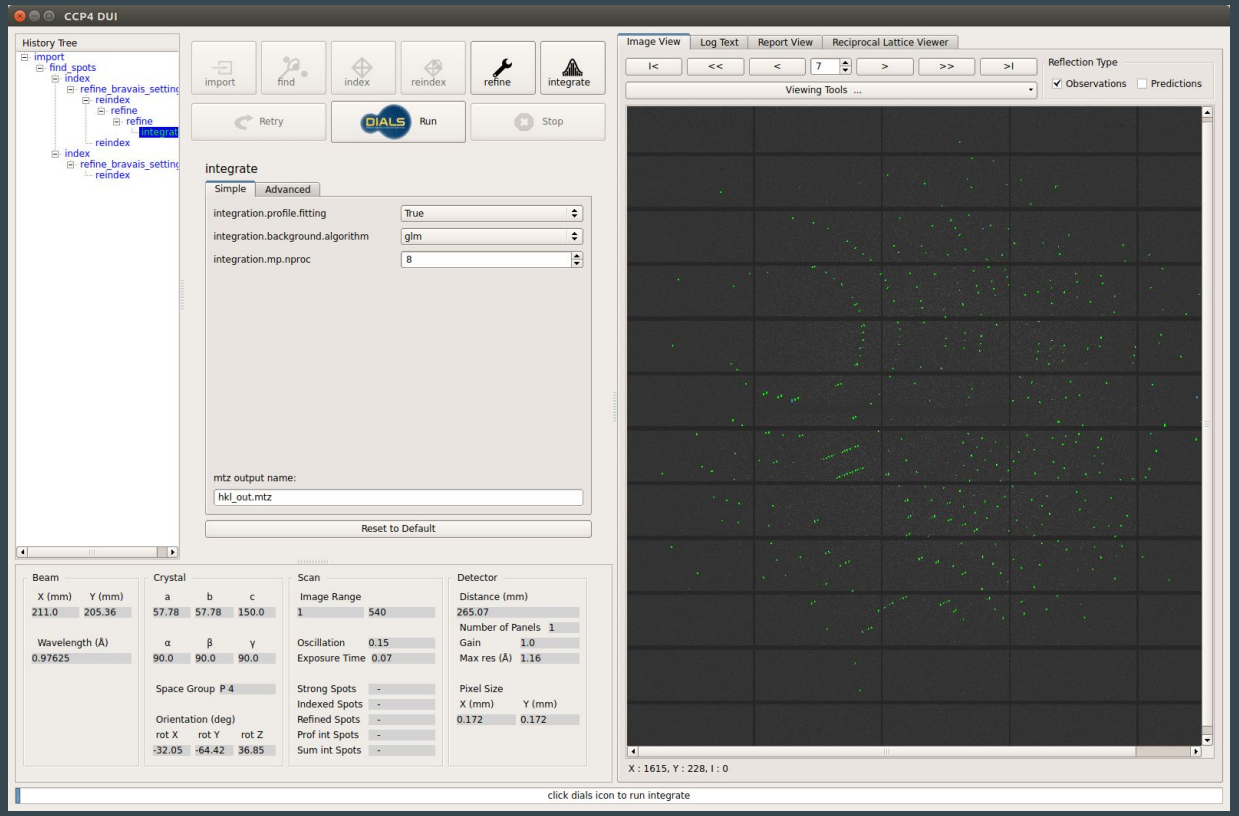
xia2 -dials ${data_directory}

```
For AUTOMATIC/DEFAULT/NATIVE
High resolution limit                  1.36   6.08   1.36
Low resolution limit                  53.92  53.92   1.40
Completeness                          99.8   99.9   97.8
Multiplicity                           5.3    4.9    3.1
I/sigma                               11.7   26.8    2.2
Rmerge                               0.061  0.026  0.370
Rmeas(I)                             0.075  0.032  0.521
Rmeas(I+/-)                          0.075  0.032  0.495
Rpim(I)                              0.032  0.014  0.279
Rpim(I+/-)                           0.042  0.018  0.326
CC half                              0.999  0.998  0.818
Wilson B factor                      8.913
Anomalous completeness                97.5  100.0   77.3
Anomalous multiplicity                 2.6    3.1    1.8
Anomalous correlation                 0.005  0.204 -0.019
Anomalous slope                      0.955  0.000  0.000
Total observations                   292123 3747   12262
Total unique                         55480  768    3919
Assuming spacegroup: P 41 21 2
Other likely alternatives are:
P 43 21 2
Unit cell:
57.781  57.781 149.995
90.000  90.000  90.000
```

DIALS/XIA2 in CCP4 I2



Future: DIALS GUI (see Luiso's talk tomorrow!)



Main DIALS programs

dials.import
dials.find_spots
dials.index
dials.refine_bravais_settings
dials.refine
dials.integrate
dials.export_mtz

dials.scale (new program)

More than 50 other dials.* commands

Then onwards into the CCP4 data processing pipeline:
POINTLESS → AIMLESS → CTRUNCATE...

DIALS on the command line

```
$ dials.import ${data_directory}/th_8_2_0*.cbf

$ dials.find_spots datablock.json nproc=8

$ dials.index datablock.json strong.pickle

$ dials.refine_bravais_settings experiments.json indexed.pickle

$ dials.reindex indexed.pickle change_of_basis_op=a,b,c

$ dials.refine bravais_setting_9.json reindexed_reflections.pickle outlier.algorithm=tukey
    use_all_reflections=true scan_varying=true output.experiments=refined_experiments.json

$ dials.integrate refined_experiments.json refined.pickle outlier.algorithm=null nproc=4

$ dials.export_mtz integrated.pickle refined_experiments.json hklout=integrated.mtz

$ pointless hklin integrated.mtz hklout sorted.mtz > pointless.log

$ aimless hklin sorted.mtz hklout scaled.mtz > aimless.log << eof
  resolution 1.3
  anomalous off
eof

$ ctruncate -hklin scaled.mtz -hklout truncated.mtz -colin '/*/*/[IMEAN,SIGIMEAN]' > ctruncate.log
```

Who needs a GUI?

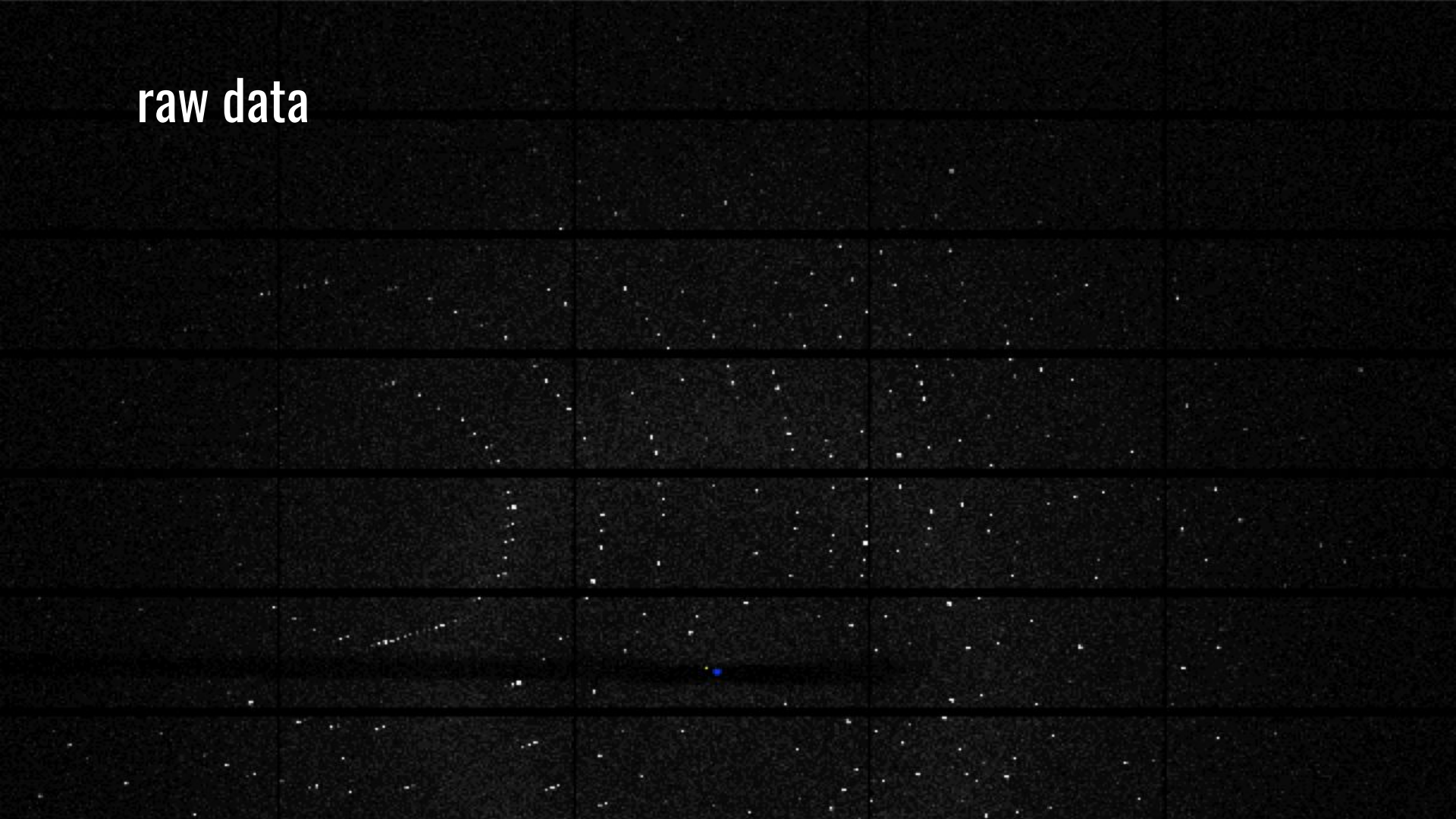
Spot finding

dials.find_spots

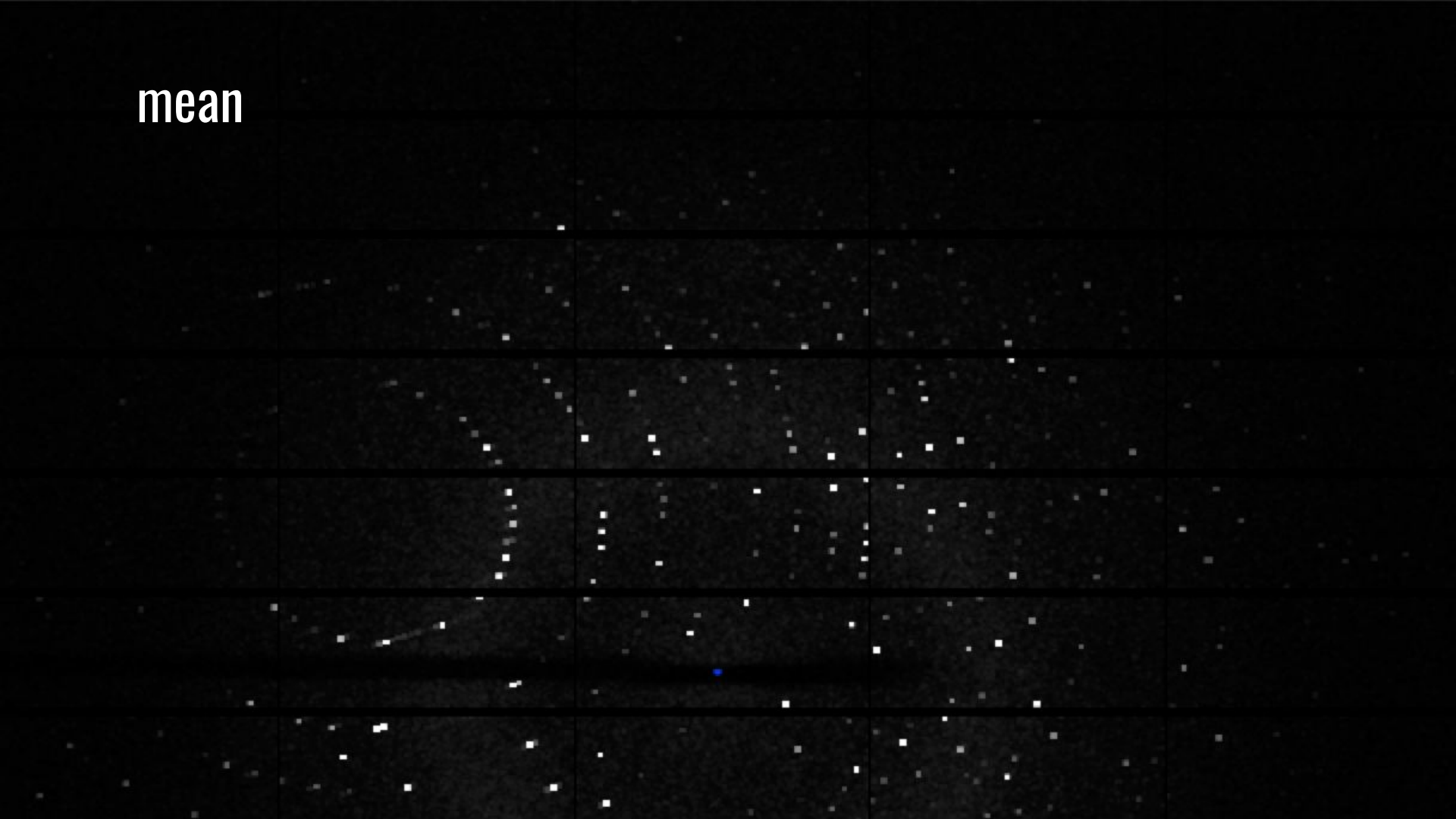
- Sequence of per-image filters to find strong pixels
- 3D analysis of strong pixels to identify strong spots
- Filter spots by
  - number of pixels
  - peak-centroid distance
  - resolution
  - ice rings
  - untrusted regions

```
$ dials.find_spots datablock.json nproc=8

…

Histogram of per-image spot count for imageset 0:
71268 spots found on 200 images (max 1897 / bin)
*
*
**** *                                             *       *
*********                            ** ******* ***** *  * *
************                 ** ****************************
***************** * ****************************************
************************************************************
************************************************************
************************************************************
************************************************************
1                         image                          200

---------------------------------------------------------------------
Saved 71268 reflections to strong.pickle
Time Taken: 88.113627
```

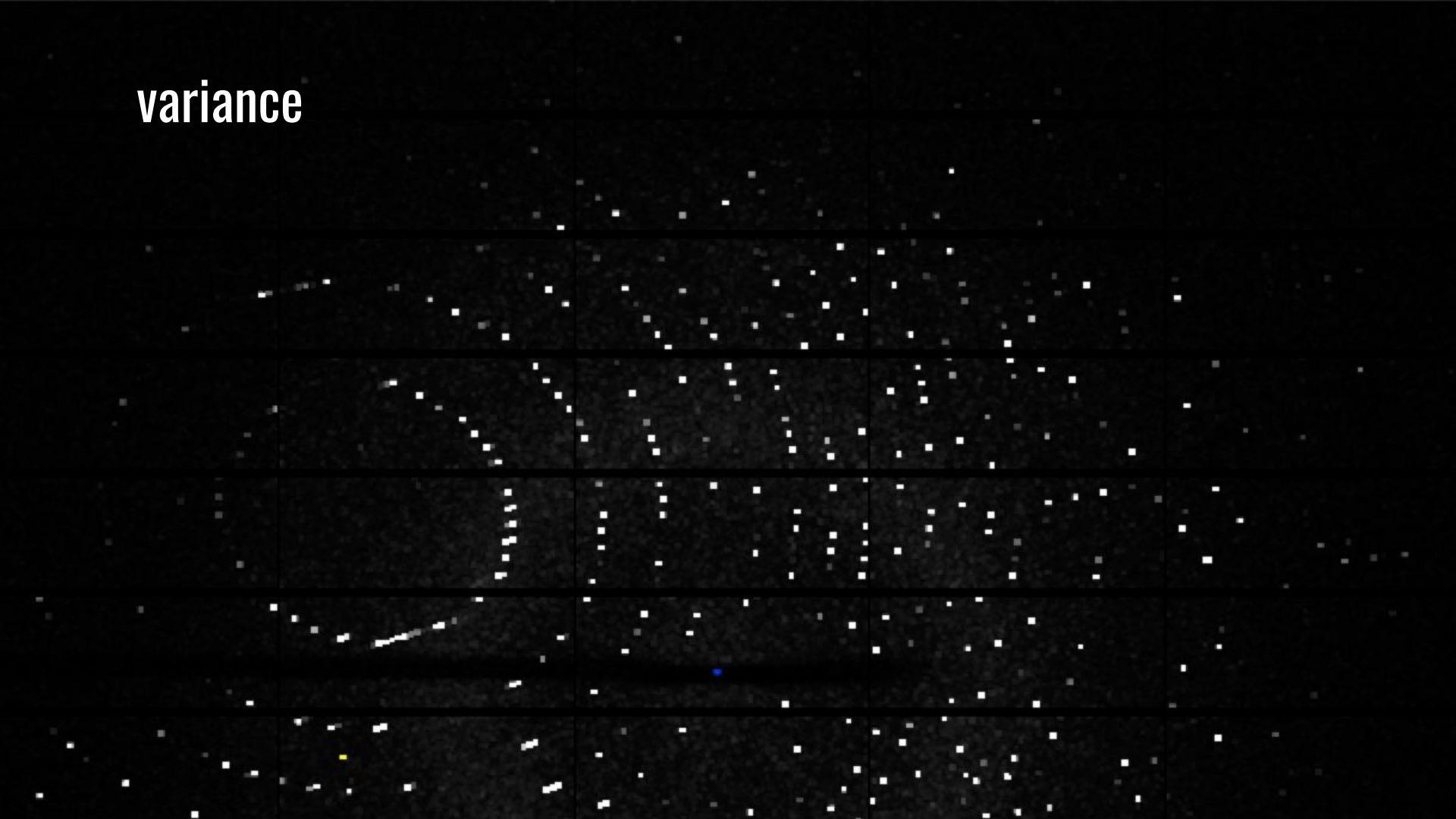
raw data



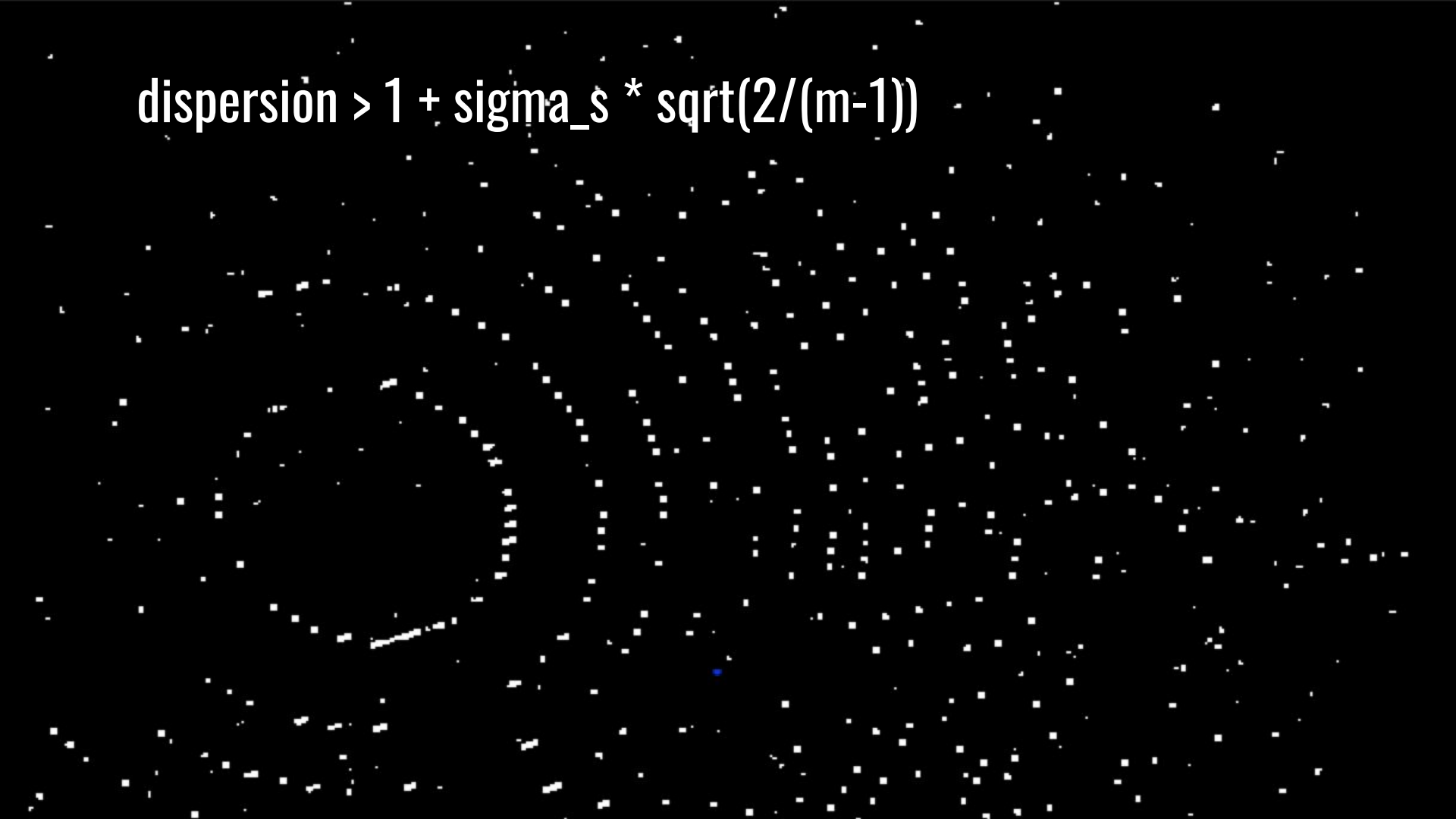
mean



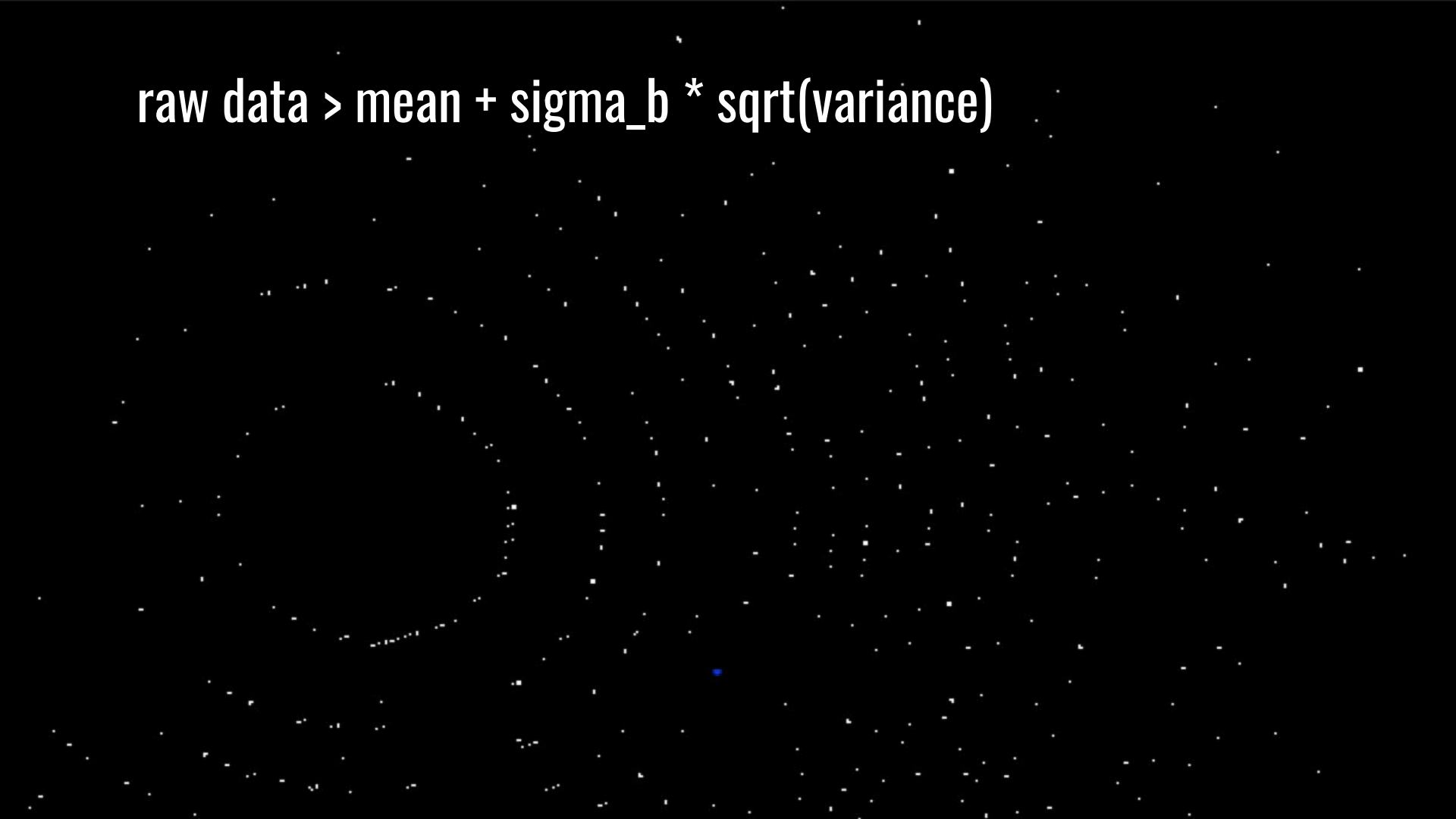
variance

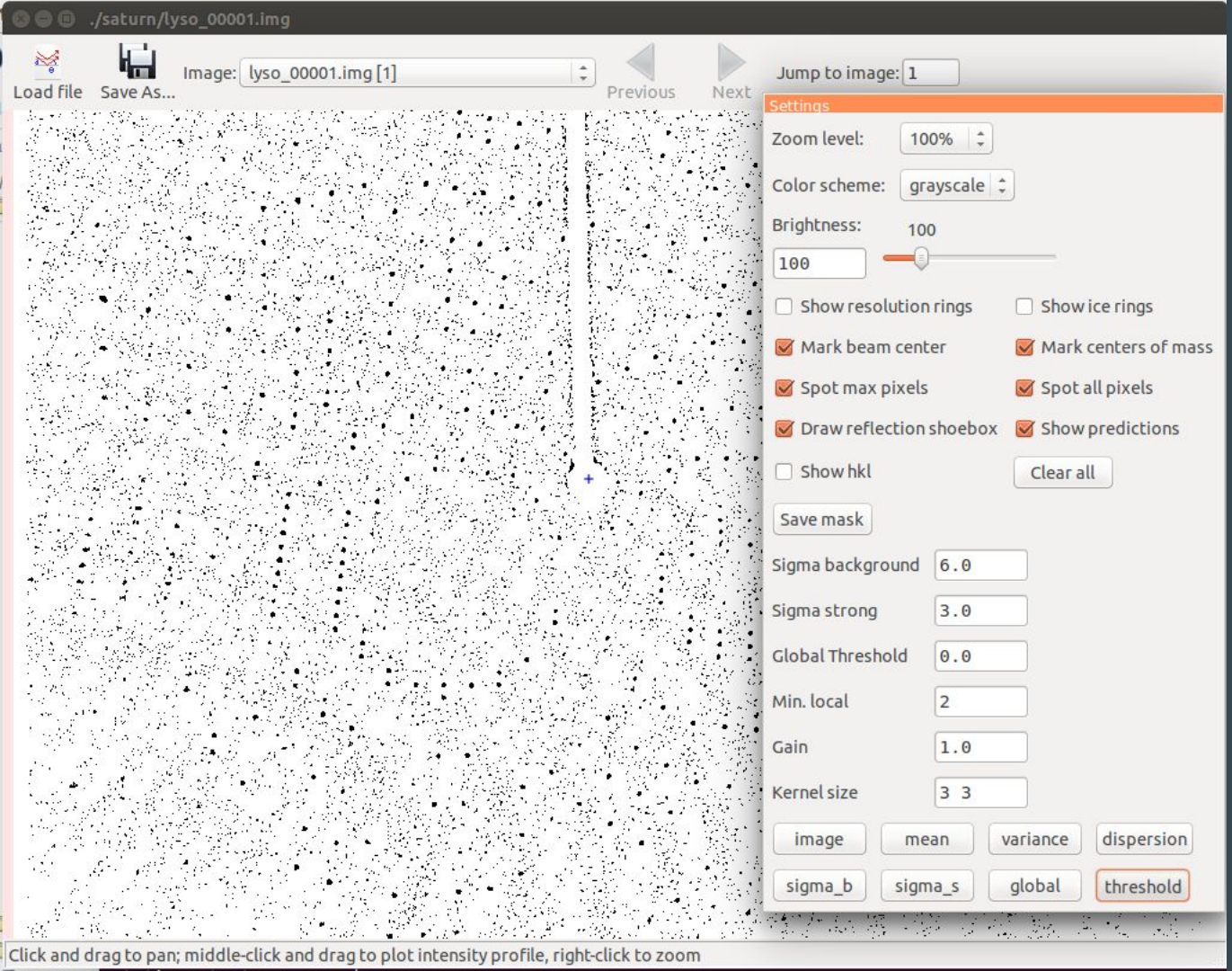


dispersion = variance / mean


dispersion > 1 + sigma_s * sqrt(2/(m-1))

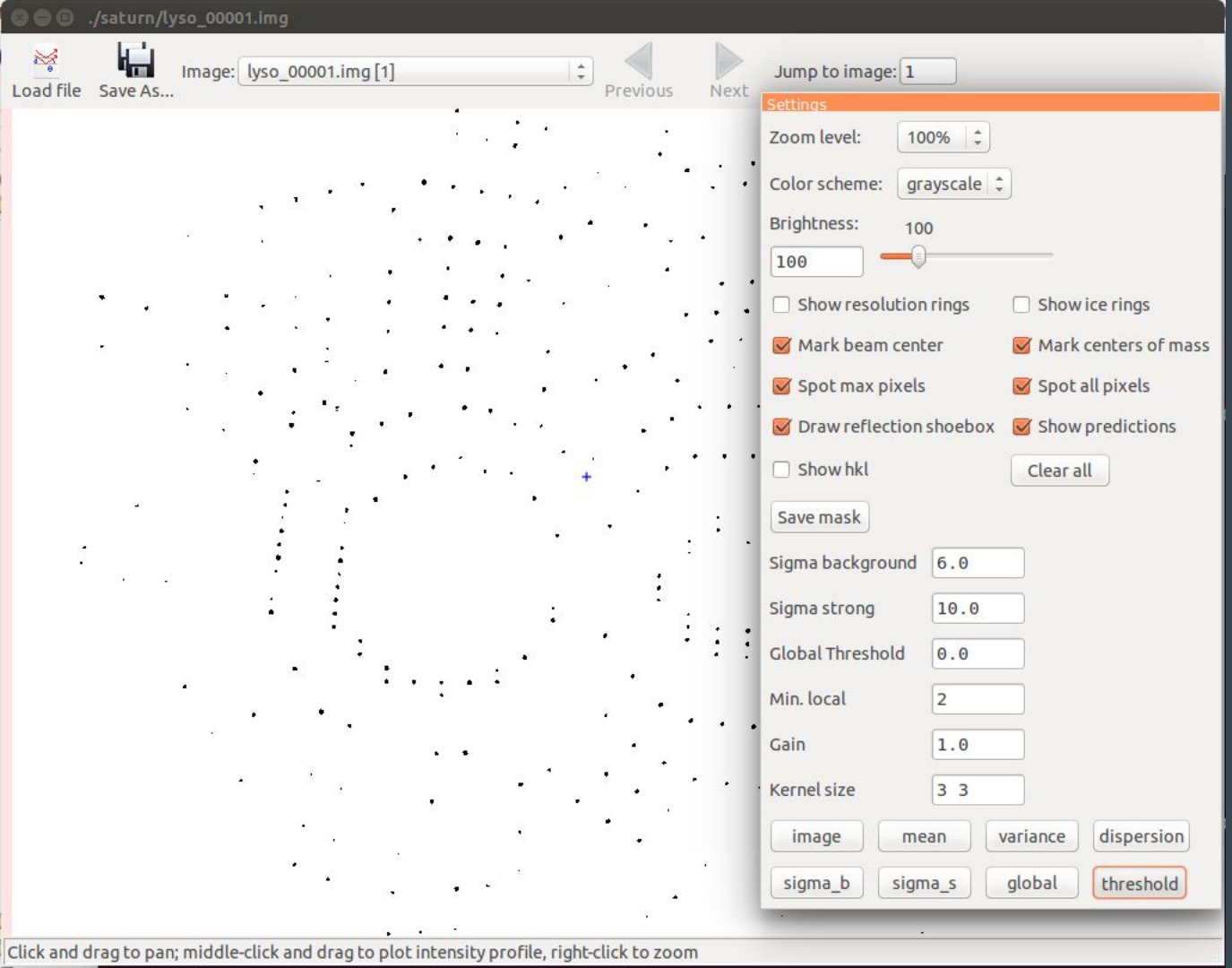
raw data > mean + sigma_b * sqrt(variance)



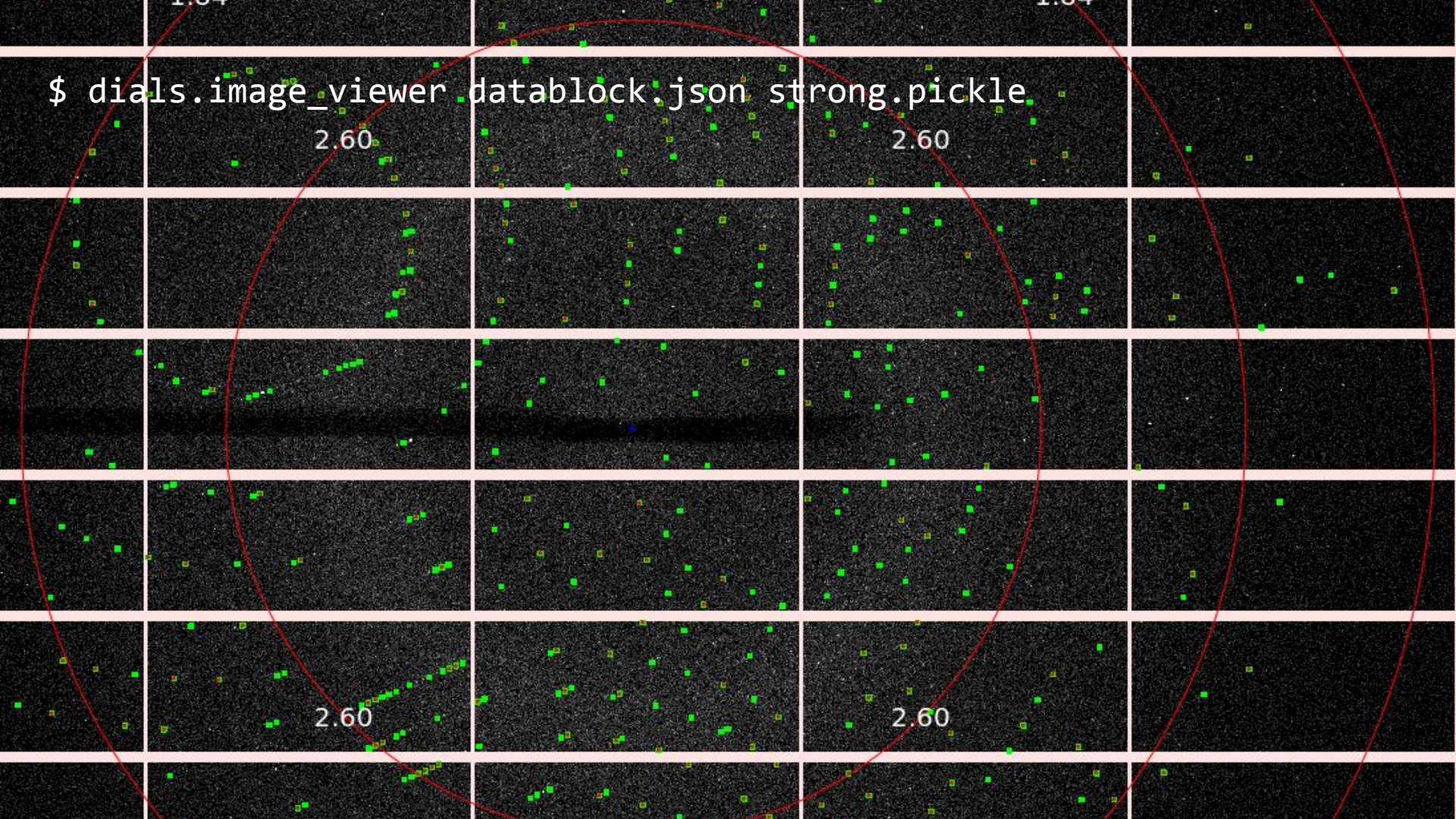


Default spot finding parameters are often not suitable for CCD images

Image is from Rigaku Saturn 92 detector



Default spot finding parameters are often not suitable for CCD images


$ dials.image_viewer datablock.json strong.pickle
2.60
2.60
2.60
2.60

Indexing

dials.index

Map centroids to reciprocal space, analyse for periodicity and determine basis vectors for the reduced cell

- Choice of method:
  - 1D FFT (DPS)
  - 3D FFT - default
  - new real space grid search algorithm
- Optionally provide known unit cell and space group

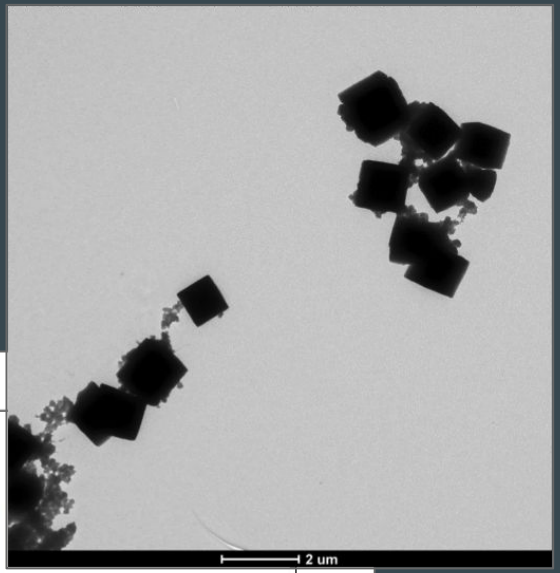
```
$ dials.index datablock.json strong.pickle
Found max_cell: 199.1 Angstrom
Setting d_min: 3.89
...
RMSDs by experiment:
-----------------------------------------
| Exp | Nref | RMSD_X | RMSD_Y  | RMSD_Z   |
|     |      | (px)   | (px)    | (images) |
-----------------------------------------
| 0   | 4049 | 0.2881 | 0.25838 | 0.17767  |
-----------------------------------------

Final refined crystal models:
model 1 (114690 reflections):
Crystal:
    Unit cell: (57.804, 57.782, 150.027, 90.009, 89.991,
   89.990)
    Space group: P 1
    U matrix:  {{ 0.3455, -0.2589, -0.9020},
                { 0.8914,  0.3909,  0.2292},
                { 0.2933, -0.8833,  0.3659}}
    B matrix:  {{ 0.0173,  0.0000,  0.0000},
                {-0.0000,  0.0173,  0.0000},
                {-0.0000,  0.0000,  0.0067}}
    A = UB:    {{ 0.0060, -0.0045, -0.0060},
                { 0.0154,  0.0068,  0.0015},
                { 0.0051, -0.0153,  0.0024}}

Saving refined experiments to experiments.json
Saving refined reflections to indexed.pickle
```

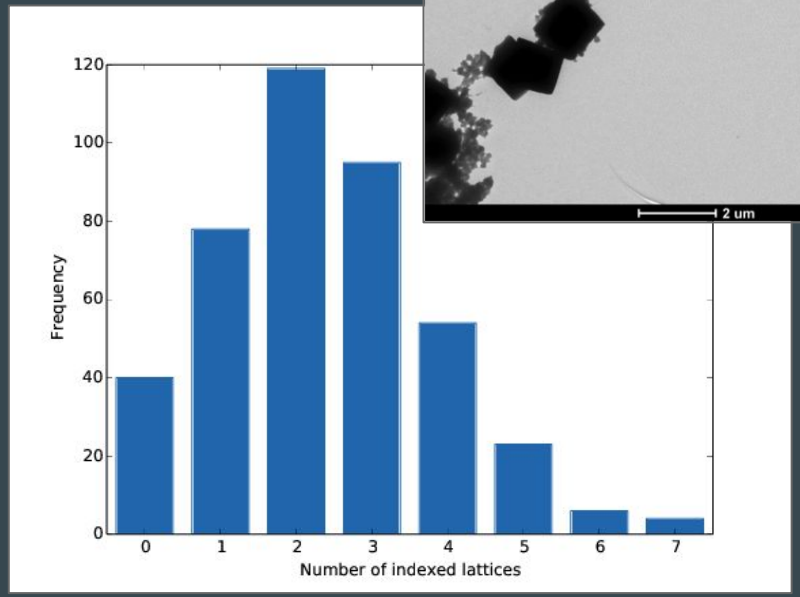
dials.index

DIALS contains an indexing algorithm that is very successful at identifying multiple lattices

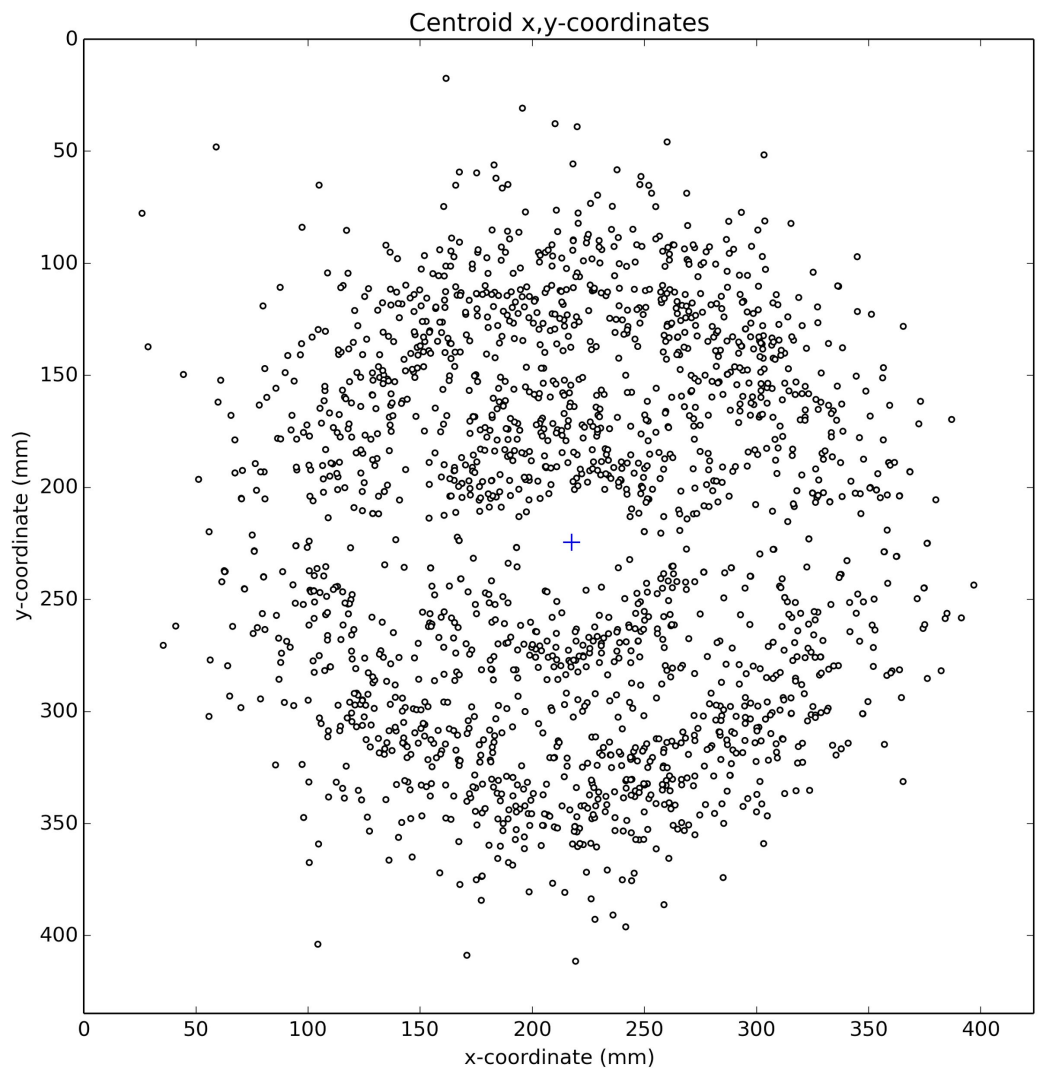
This even works when lattices diffract equally well, and only a narrow wedge of data is available

As additional lattices are found, joint refinement reduces correlations between crystal and detector parameters

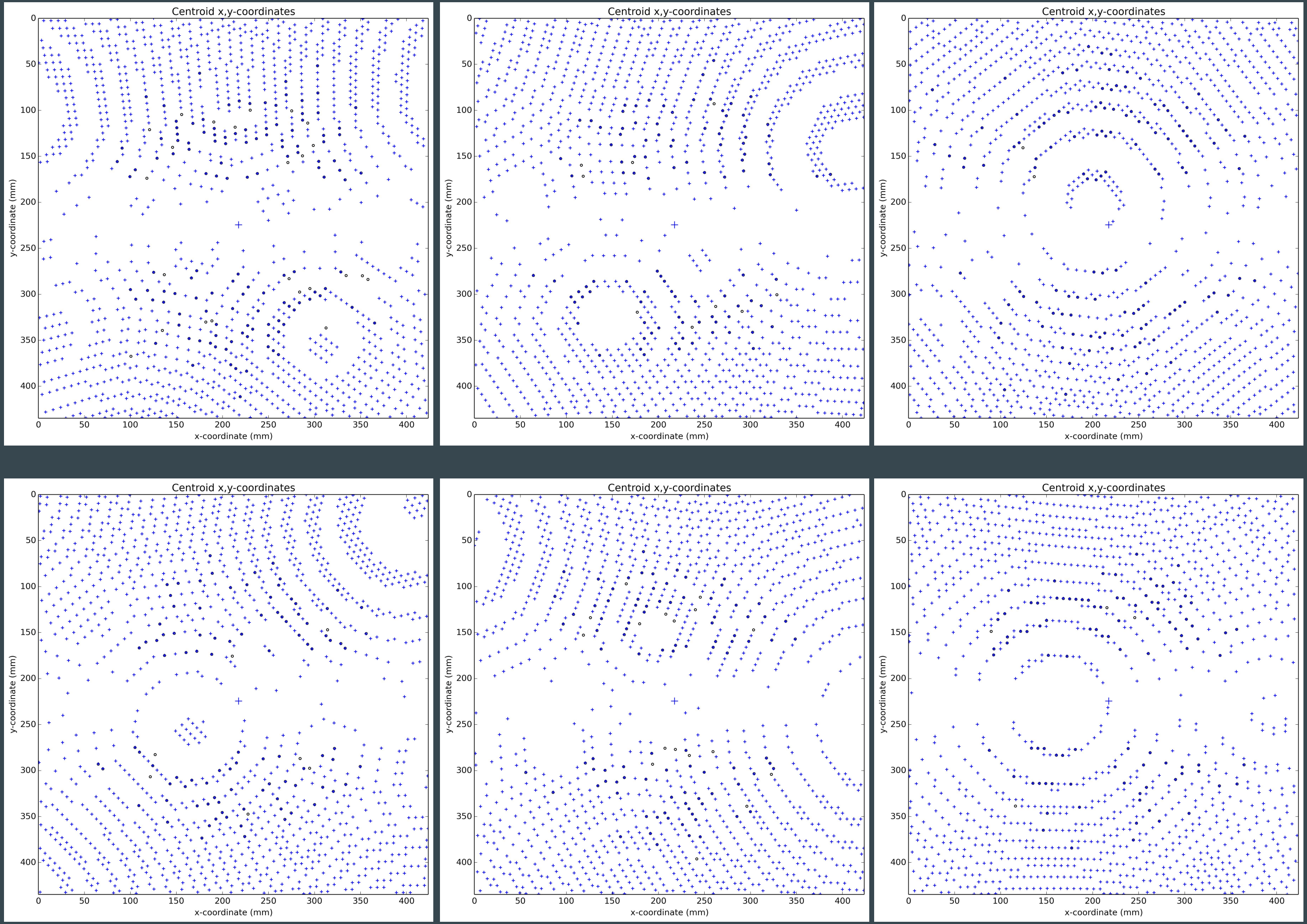




R. Gildea et al. (2014) Acta Cryst. D70, 2652-66



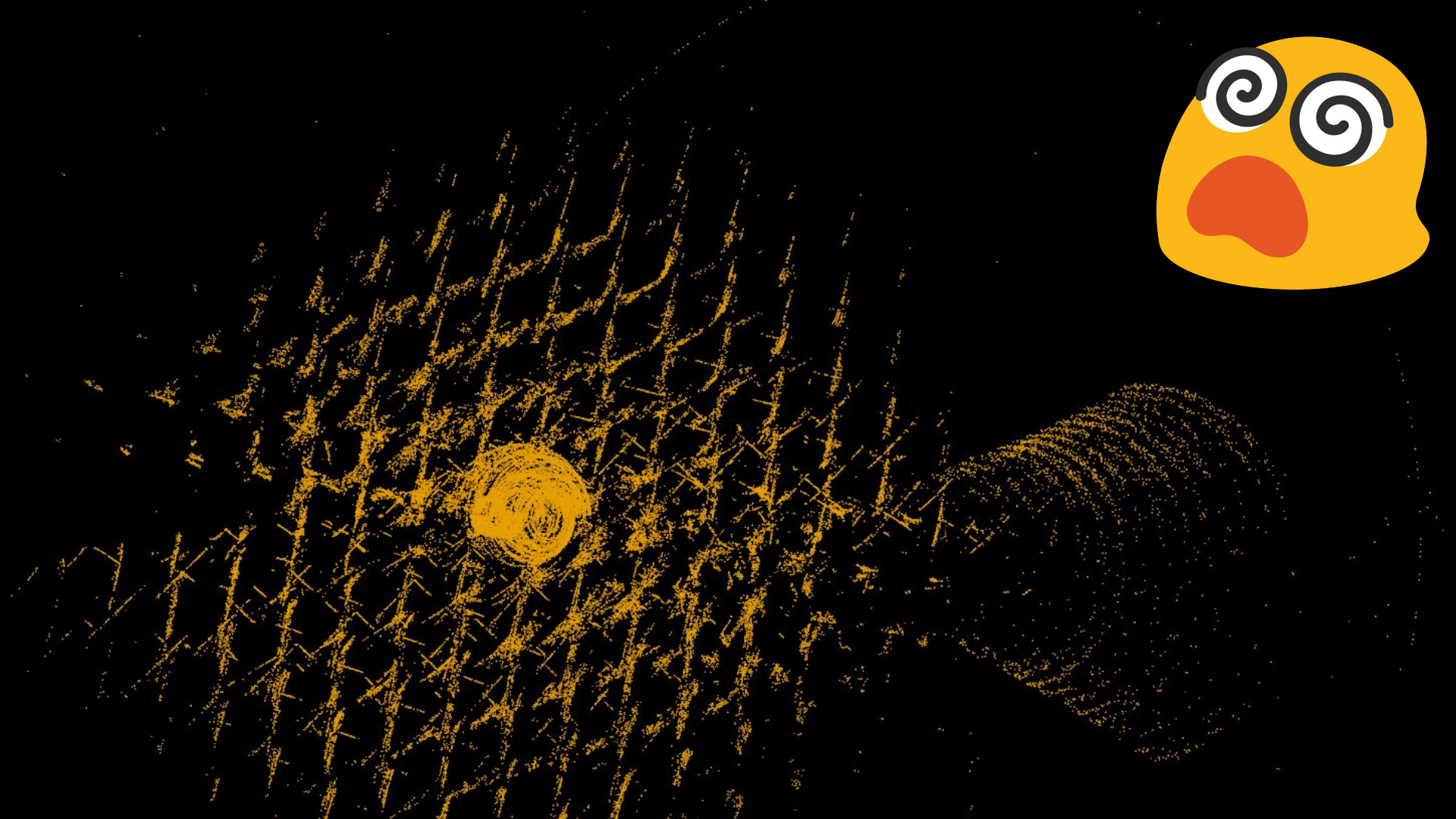
1° wedge of data
1858 spots

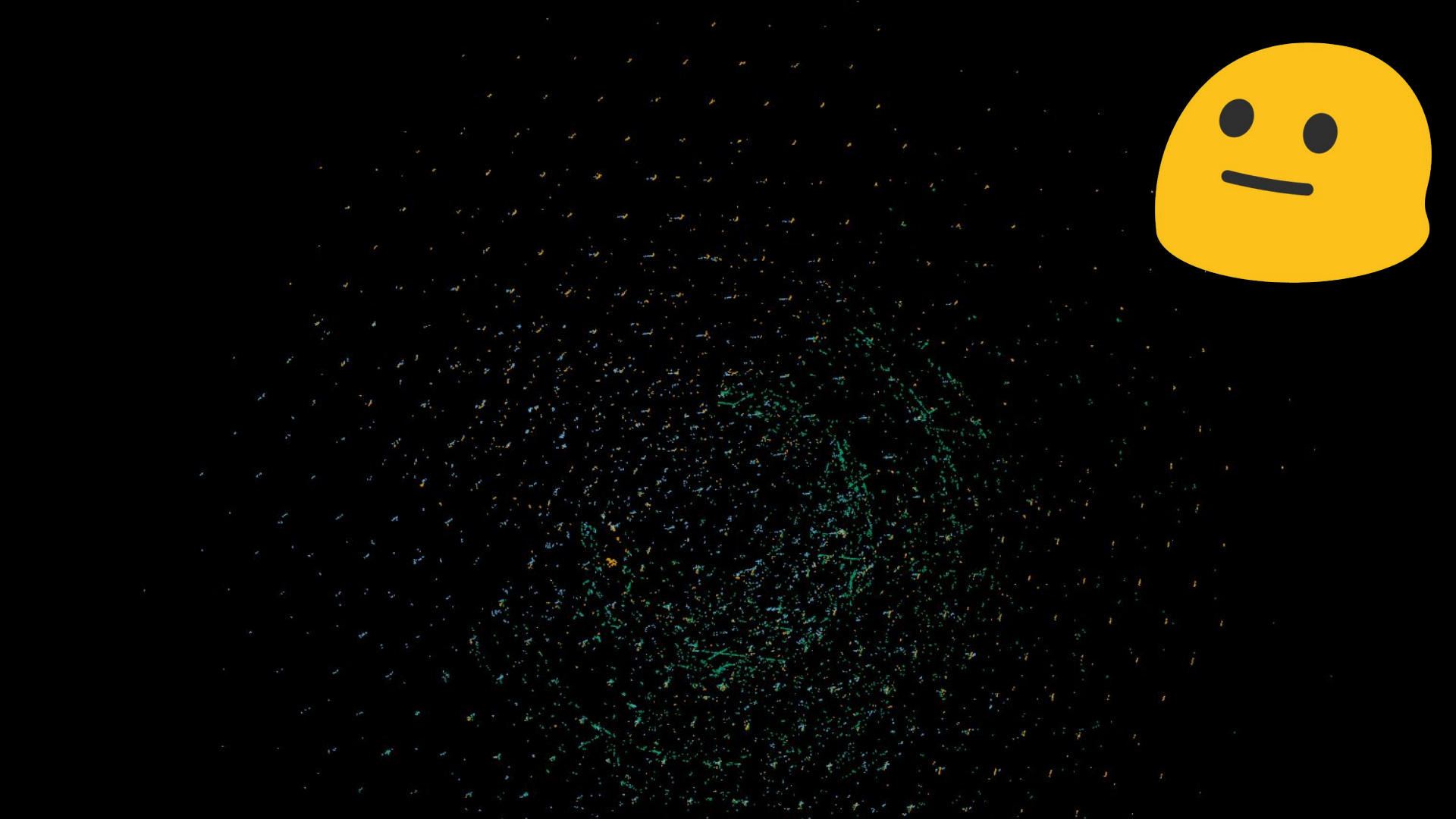


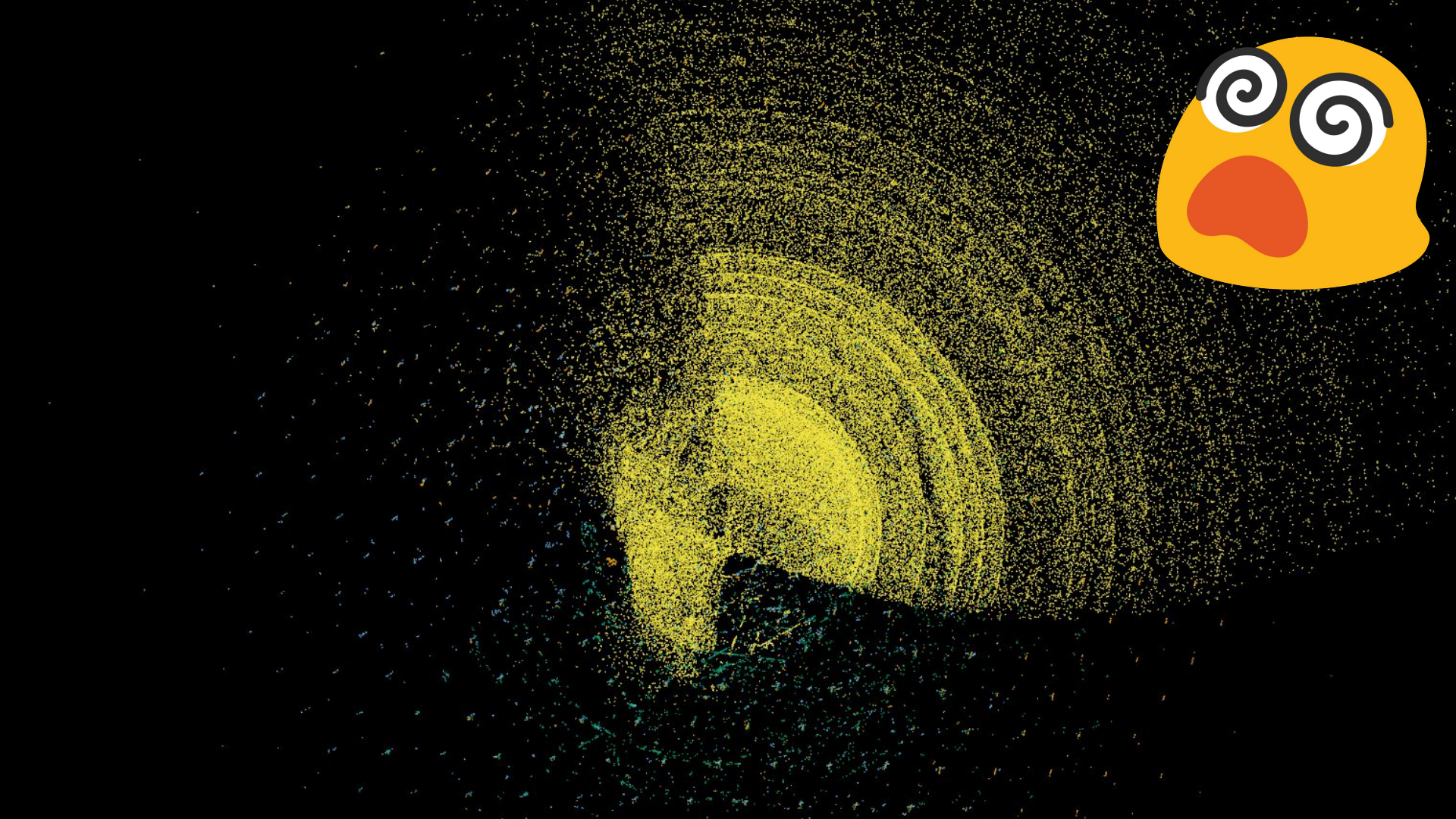
6 lattices identified

dials.reciprocal_lattice_viewer

```
$ dials.reciprocal_lattice_viewer
Usage: dials.reciprocal_lattice_viewer [options] datablock.json reflections.pickle

Options:
  -h, --help          show this help message and exit
  -c, --show-config   Show the configuration parameters.
  -a ATTRIBUTES_LEVEL, --attributes-level=ATTRIBUTES_LEVEL
                      Set the attributes level for showing configuration
                      parameters
  -e EXPERT_LEVEL, --expert-level=EXPERT_LEVEL
                      Set the expert level for showing configuration
                      parameters
  -v                  Increase verbosity


Visualise the strong spots from spotfinding in reciprocal space.

Examples:

  dials.reciprocal_lattice_viewer datablock.json strong.pickle

  dials.reciprocal_lattice_viewer experiments.json indexed.pickle
```







dials.refine_bravais_settings

- After indexing, look for lattice symmetry
- All compatible Bravais lattices are tested
- Metric fit score, refined RMSD and symmetry element CCs are reported
- The user chooses which solution to take further

```
$ dials.refine_bravais_settings experiments.json indexed.pickle
```

	Solution	Metric fit	rmsd	min/max cc	#spots	lattice	unit_cell						volume	cb_op
*	9	0.0311	0.063	0.800/0.857	8099	tP	57.78	57.78	150.00	90.00	90.00	90.00	500867	a,b,c
*	8	0.0311	0.063	0.800/0.969	8099	oC	81.72	81.73	150.01	90.00	90.00	90.00	1002008	a-b,a+b,c
*	7	0.0272	0.061	0.969/0.969	8099	mC	81.73	81.74	150.03	90.00	89.99	90.00	1002365	a-b,a+b,c
*	6	0.0311	0.062	0.805/0.805	8099	mC	81.73	81.72	150.02	90.00	89.99	90.00	1002012	a+b,-a+b,c
*	5	0.0154	0.061	0.800/0.906	8099	oP	57.79	57.76	149.99	90.00	90.00	90.00	500672	a,b,c
*	4	0.0147	0.060	0.821/0.821	8099	mP	57.77	57.80	150.01	90.00	90.02	90.00	500853	-b,-a,-c
*	3	0.0154	0.060	0.906/0.906	8099	mP	57.80	57.78	150.02	90.00	89.98	90.00	500945	a,b,c
*	2	0.0152	0.061	0.800/0.800	8099	mP	57.78	150.01	57.80	90.00	89.99	90.00	500925	b,c,a
*	1	0.0000	0.060	-/-	8099	aP	57.80	57.78	150.03	90.01	89.99	89.99	501086	a,b,c

* = recommended solution

dials.refine_bravais_settings

- After indexing, look for lattice symmetry
- All compatible Bravais lattices are tested
- Metric fit score, refined RMSD and symmetry element CCs are reported
- The user chooses which solution to take further

```
$ dials.refine_bravais_settings experiments.json indexed.pickle
```

	Solution	Metric fit	rmsd	min/max cc	#spots	lattice	unit_cell						volume	cb_op
	5	3.7053	1.048	0.199/0.438	1792	oP	11.52	13.51	29.38	90.00	90.00	90.00	4571	a,b,c
	4	3.7053	1.001	0.438/0.438	1755	mP	13.43	11.55	29.30	90.00	91.40	90.00	4543	-b,-a,-c
	3	3.7038	1.033	0.287/0.287	1811	mP	11.47	29.43	13.46	90.00	88.93	90.00	4542	-a,-c,-b
*	2	0.1091	0.065	0.199/0.199	1929	mP	11.63	13.55	30.14	90.00	93.69	90.00	4739	a,b,c
*	1	0.0000	0.060	-/-	1890	aP	11.63	13.55	30.15	89.94	93.70	90.06	4742	a,b,c

* = recommended solution

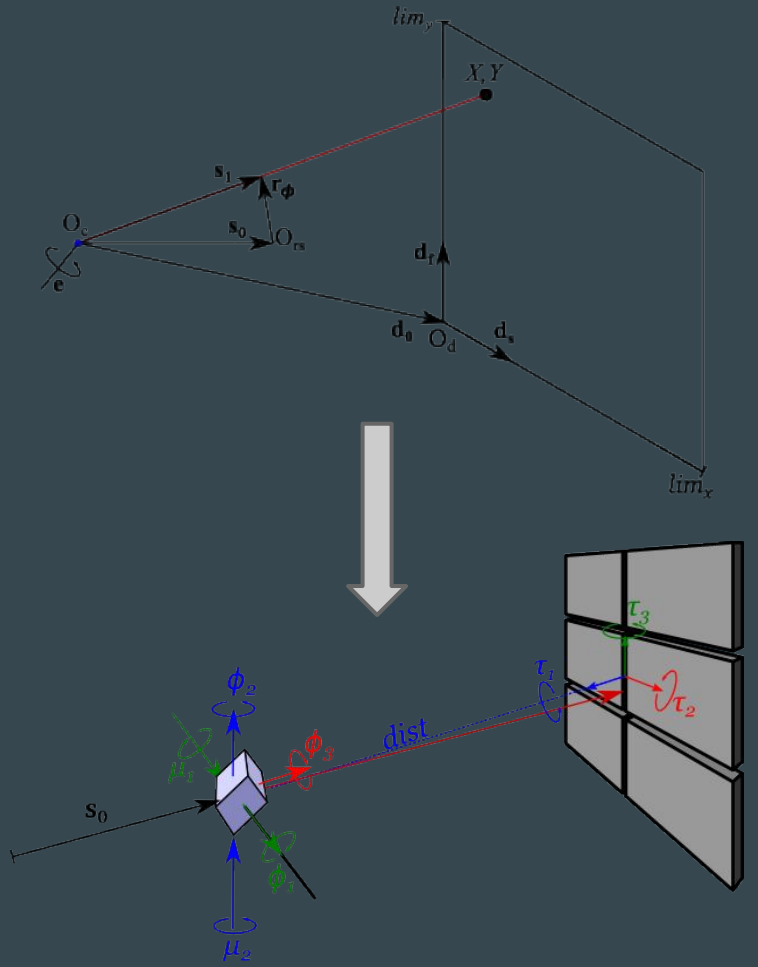
Refinement

dials.refine

Parameterise the reflection prediction equation, minimise the squared differences between predictions (X_c, Y_c, ϕ_c) and observations (X_o, Y_o, ϕ_o)

Model	State	Parameters
Beam	$\mathbf{s_0}$	2 orientation angles (μ_1, μ_2) 1 length of $\mathbf{s_0}$ (ν, wavenumber)
Crystal orientation	**U**	3 orientation angles (φ_1, φ_2, φ_3)
Crystal unit cell	**B**	6 elements of the metrical matrix $\mathbf{G}^*$
Detector	**d**	3 translations (Dist, t_1, t_2) 3 rotation angles (τ_1,τ_2,τ_3)

18 parameters in the $P1$ case. Usually ν and μ_1 are fixed



dials.refine

We approach this as a traditional non-linear least squares problem

Keep simple and robust by using the positional residual only (no post-refinement)

Do refinement globally, using strong spot data from the whole scan

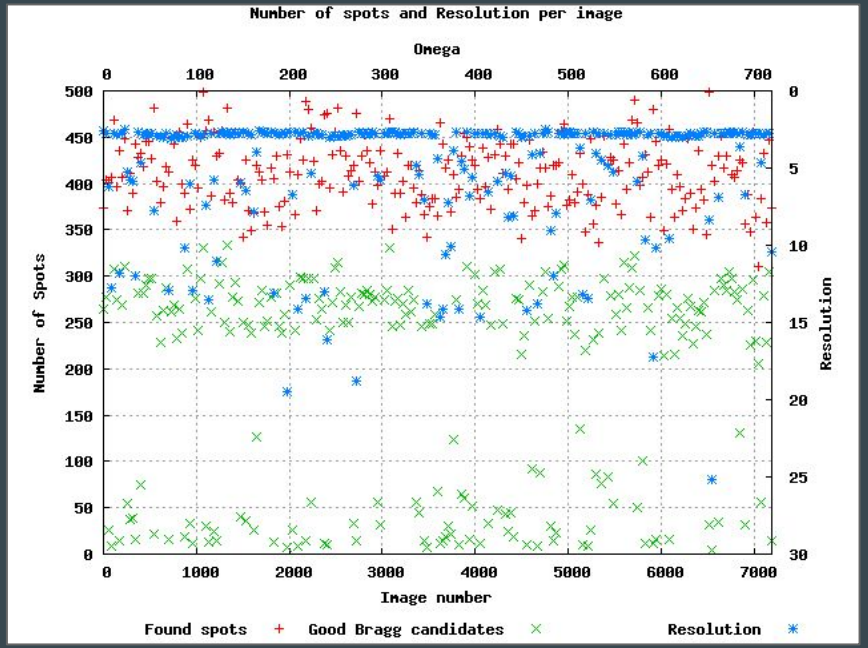
How do we model a crystal that changes during the data collection?

The crystal orientation and the unit cell may change due to

- sampling of different mosaic blocks during data collection
- crystal "slippage" (unlikely for cryo-cooled samples)
- radiation damage

dials.refine: example

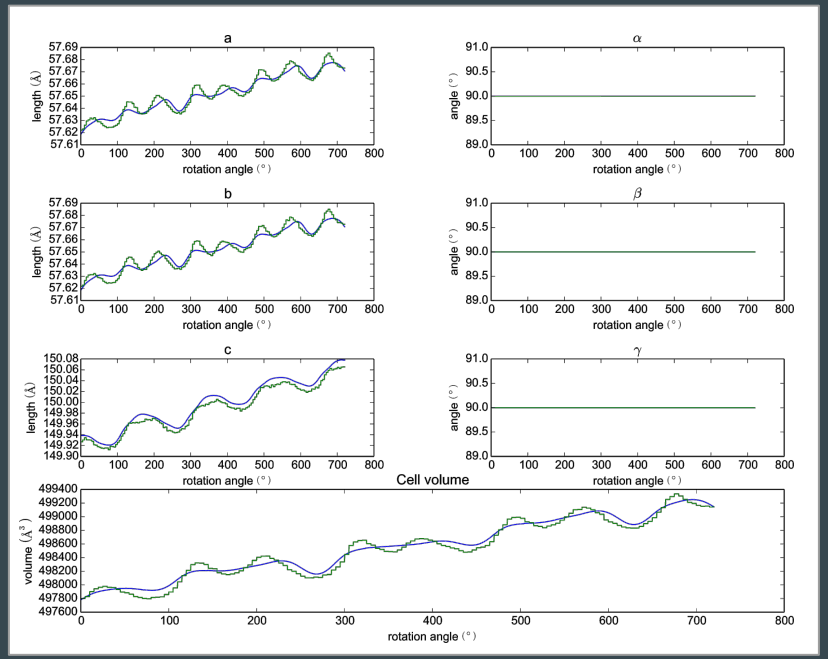
720° of tetragonal thaumatin data collected at 0.1°/image, 40Hz, 3% transmission at DLS I03

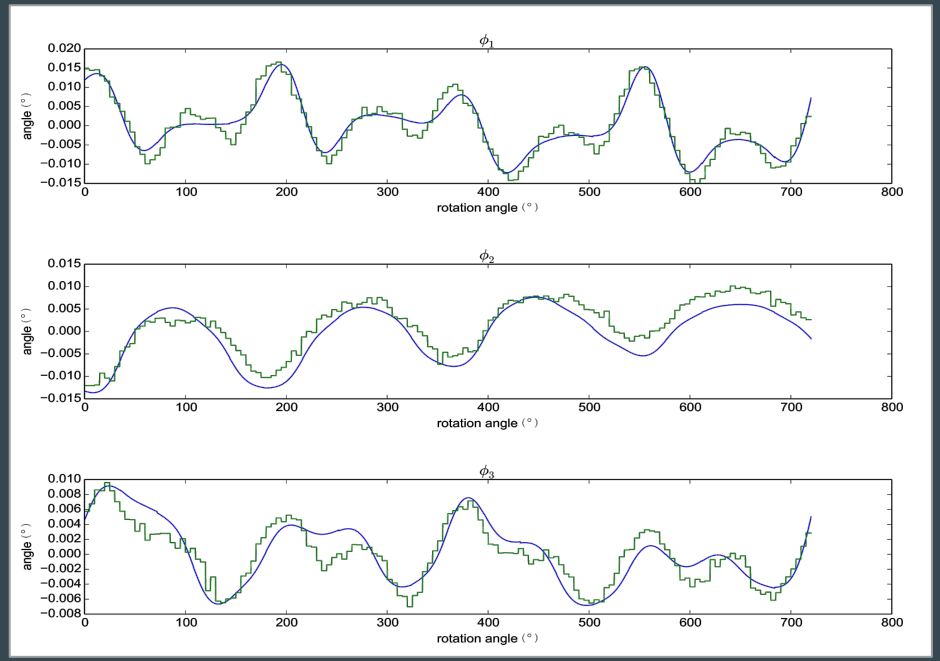


Scan-varying refinement

Global, scan-varying refinement with a Gaussian smoother inspired by AIMLESS

117 parameters in total: 6 detector, 1 beam, 22 "samples" of 3 crystal orientation and 2 unit cell parameters





dials.refine

Recently, the scan-varying refinement technique was extended to other parameters such as the beam orientation angles

This can be useful for certain difficult experiments, such as electron diffraction

More details in Clabbers *et al.* (2018) Acta Cryst. **D74**, 506-518.

research papers

STRUCTURAL BIOLOGY

ISSN 2059-7983

Electron diffraction data processing with *DIALS*

Max T. B. Clabbers,[a] Tim Gruene,[b] James M. Parkhurst,[c] Jan Pieter Abrahams[a,b] and David G. Waterman[d,e]*

[a]Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, University of Basel, Mattenstrasse 26, 4058 Basel, Switzerland, [b]Paul Scherrer Institute, 5232 Villigen PSI, Switzerland, [c]Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot OX11 0DE, England, [d]STFC, Rutherford Appleton Laboratory, Didcot OX11 0FA, England, and [e]CCP4, Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot OX11 0FA, England. *Correspondence e-mail: david.waterman@stfc.ac.uk

Received 26 March 2018
Accepted 23 May 2018

Edited by R. J. Read, University of Cambridge, England

Keywords: electron microscopy; electron crystallography; protein nanocrystals; diffraction geometry; *DIALS*.

Supporting information: this article has supporting information at journals.iucr.org/d

Electron diffraction is a relatively novel alternative to X-ray crystallography for the structure determination of macromolecules from three-dimensional nanometre-sized crystals. The continuous-rotation method of data collection has been adapted for the electron microscope. However, there are important differences in geometry that must be considered for successful data integration. The wavelength of electrons in a TEM is typically around 40 times shorter than that of X-rays, implying a nearly flat Ewald sphere, and consequently low diffraction angles and a high effective sample-to-detector distance. Nevertheless, the *DIALS* software package can, with specific adaptations, successfully process continuous-rotation electron diffraction data. Pathologies encountered specifically in electron diffraction make data integration more challenging. Errors can arise from instrumentation, such as beam drift or distorted diffraction patterns from lens imperfections. The diffraction geometry brings additional challenges such as strong correlation between lattice parameters and detector distance. These issues are compounded if calibration is incomplete, leading to uncertainty in experimental geometry, such as the effective detector distance and the rotation rate or direction. Dynamic scattering, absorption, radiation damage and incomplete wedges of data are additional factors that complicate data processing. Here, recent features of *DIALS* as adapted to electron diffraction processing are shown, including diagnostics for problematic diffraction geometry refinement, refinement of a smoothly varying beam model and corrections for distorted diffraction images. These novel features, combined with the existing tools in *DIALS*, make data integration and refinement feasible for electron crystallography, even in difficult cases.

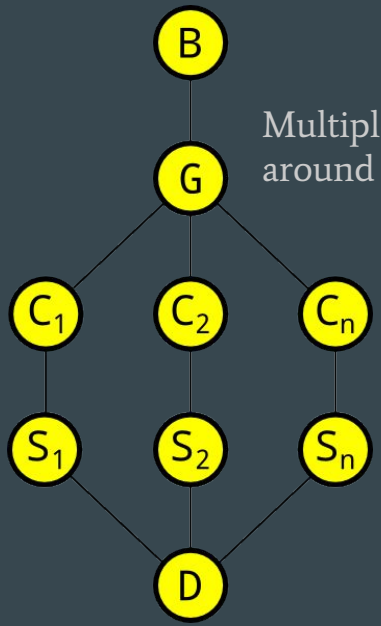
1. Introduction

Electron diffraction (ED) allows the structural analysis of nanometre-sized samples of crystalline material. Since the maximal radiation dose is proportional to the sample volume, electron diffraction of organic and macromolecular compounds was long limited to two-dimensional samples (Unwin & Henderson, 1975; Hovmöller, 2017). In contrast to X-ray crystallography, the three domains, inorganic, organic and macromolecular electron crystallography, developed rather independently of each other (Vainshtein, 1964; Dorset, 1995; Kolb *et al.*, 2007; Glaeser *et al.*, 2007; Zou *et al.*, 2011). Physical and instrumental limitations, such as miniature sample size or dynamic scattering effects and lens distortions, affect data precision. However, several studies show that the model accuracy compares with that of X-ray structures (Weirich *et al.*, 1996; Mugnaioli & Kolb, 2014; Dorset, 1995; Palatinus *et al.*, 2017). Only about one and a half decades ago, electron diffraction of three-dimensional crystals was pioneered with automated diffraction tomography (ADT) and was further refined with rotation electron diffraction (RED;

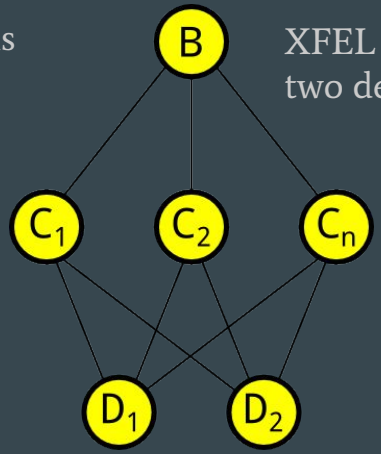
OPEN ACCESS

Multiple experiments

- Global refinement across datasets that share some models
- Typical use cases involve multiple crystals



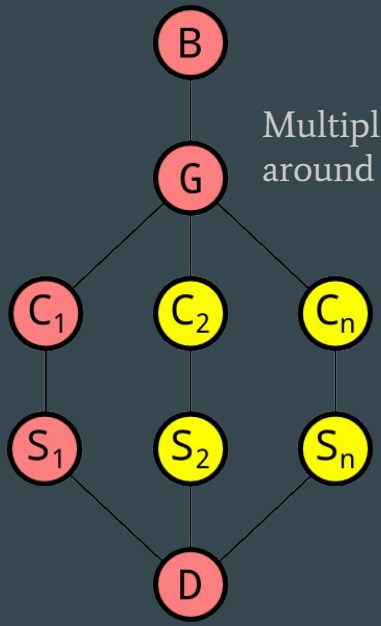
Multiple rotation scans around the same axis



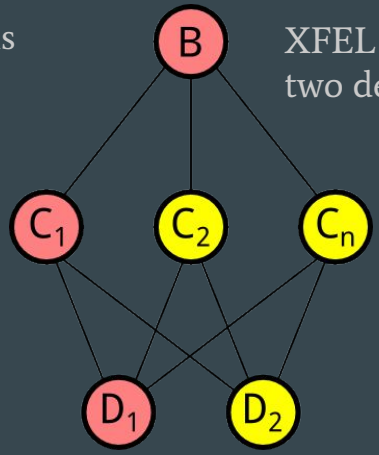
XFEL shots with two detectors

Multiple experiments

- Global refinement across datasets that share some models
- Typical use cases involve multiple crystals



Multiple rotation scans around the same axis

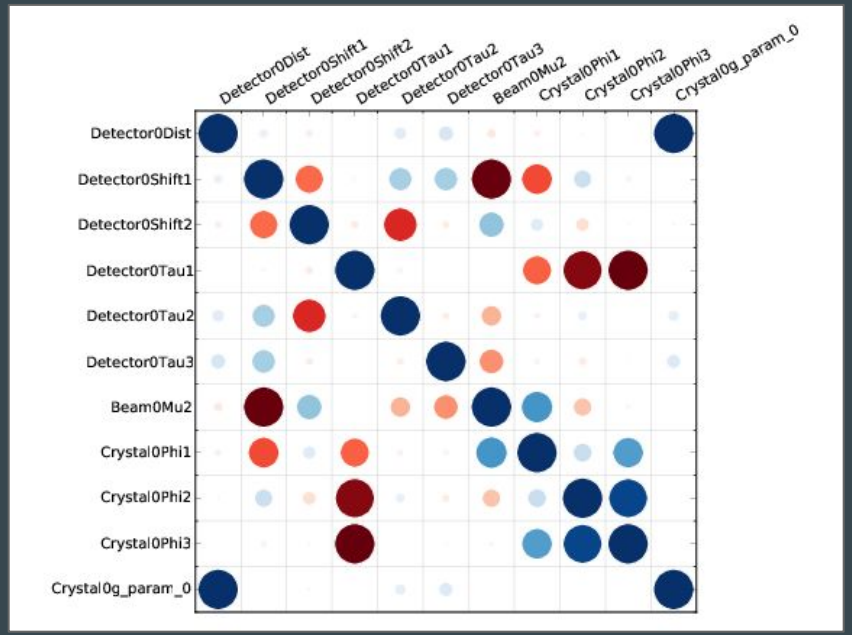


XFEL shots with two detectors

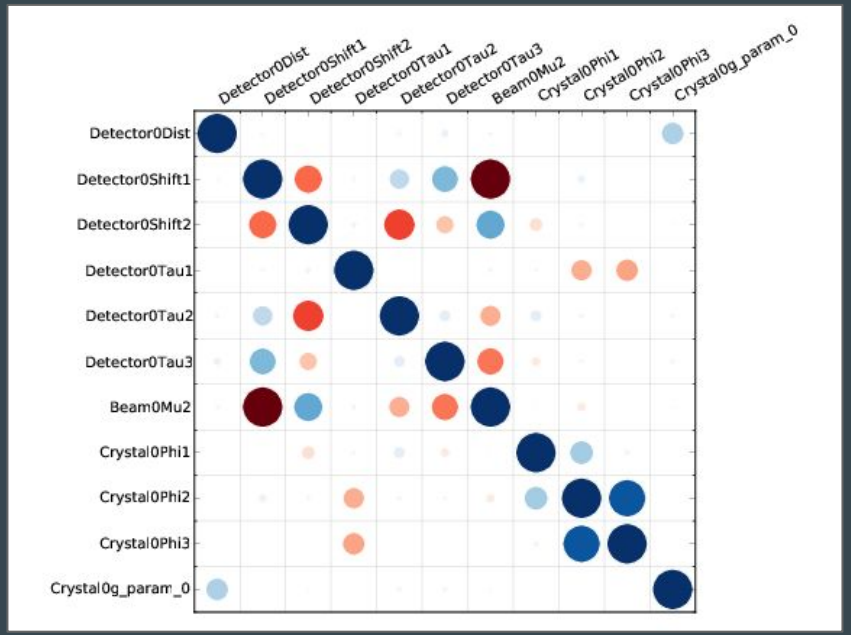
Multiple experiments

Cubic polyhedrin crystals, 1° scans

One lattice



5 sweeps (16 lattices)



Integration

Tasks in dials.integrate

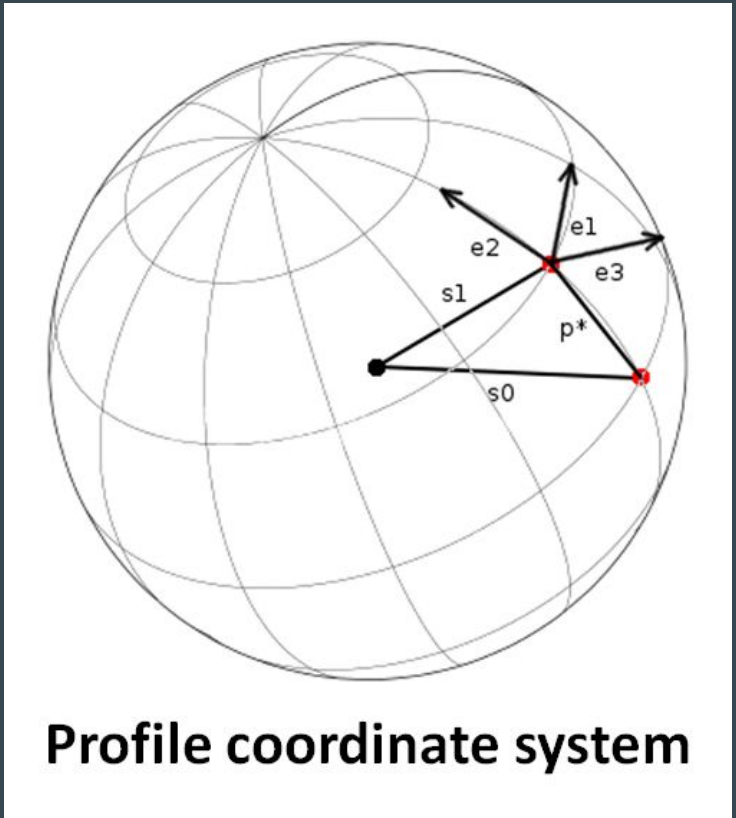
Calculate the bounding box parameters from strong reflections

Predict the positions of reflections on the images

Build reference profiles across all images

Integrate the reflections and save output

Computing reflection shoeboxes



Use the kabsch model of a normal distribution on the surface of the Ewald sphere

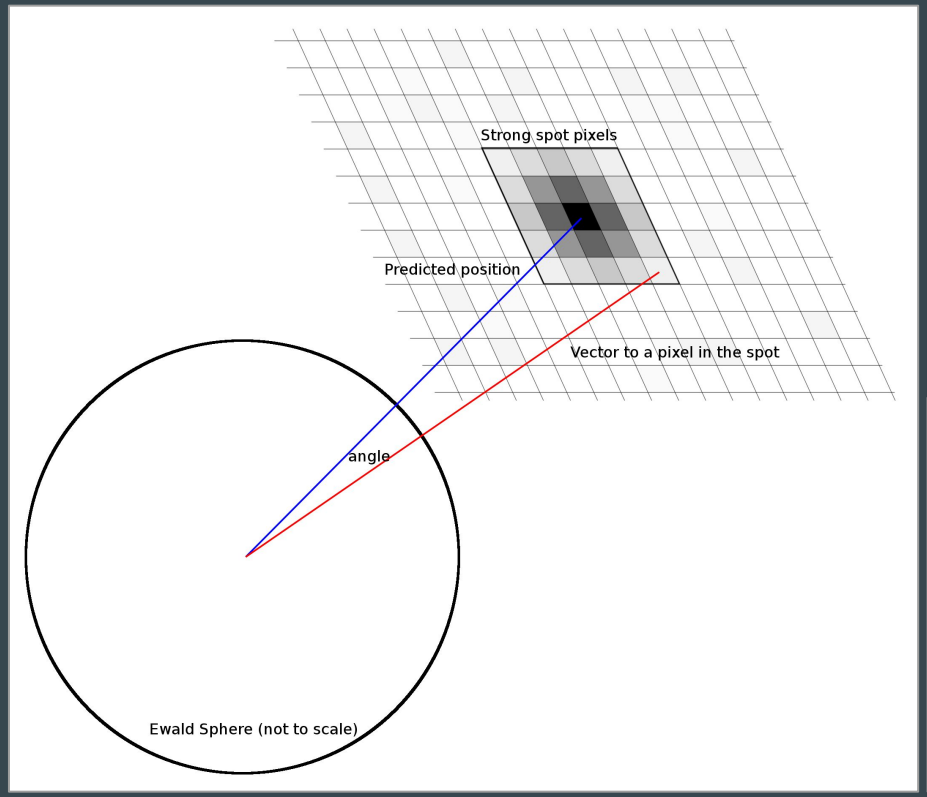
- Corrects for geometrical distortions
- Makes spots appear to have taken shortest path through Ewald sphere
- Model assumes a Gaussian profile in Kabsch coordinate system

$$\exp\left(\frac{-\epsilon_1^2}{2\sigma_D^2}\right)\exp\left(\frac{-\epsilon_2^2}{2\sigma_D^2}\right)\exp\left(\frac{-\epsilon_3^2}{2\sigma_M^2}\right)$$

2 parameters:

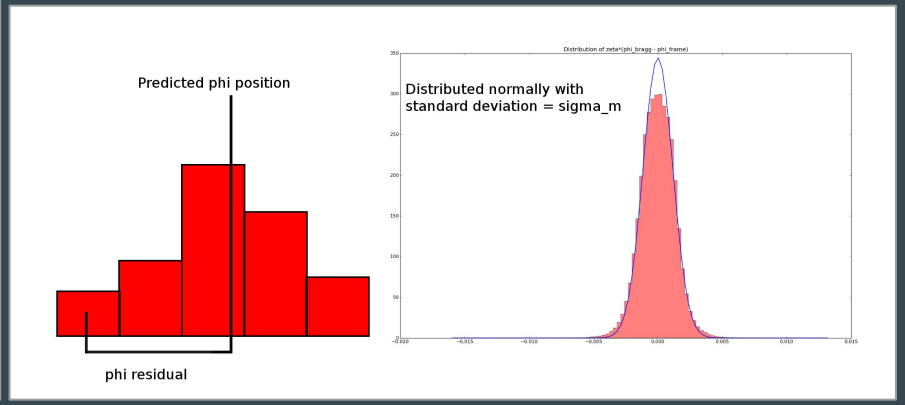
σ_D - roughly "beam divergence"
σ_M - roughly "mosaicity"

Computing reflection shoeboxes

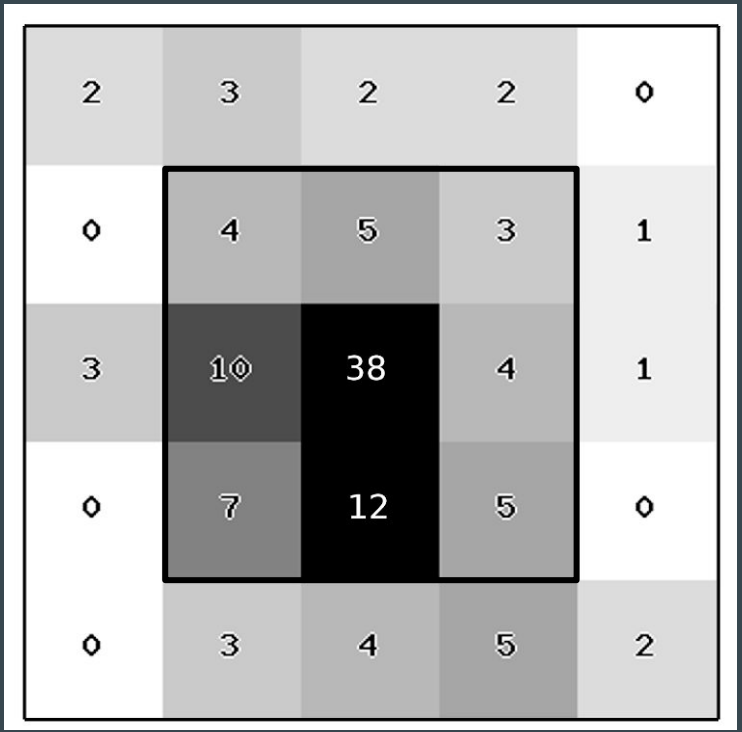


σ_D is calculated from the spread of angles between the predicted diffracted beam vector and the vector for each strong pixel in the spot

σ_M is calculated by maximum likelihood method assuming a normal distribution of phi residuals for each strong pixel in the spot



Background determination



Summation integration: estimate the reflection intensity by summing the counts contributing to the reflection and subtracting the background

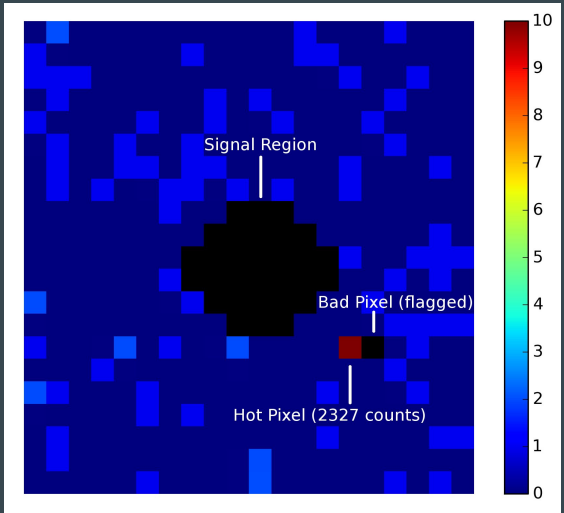
I = SUM(Counts - Background)

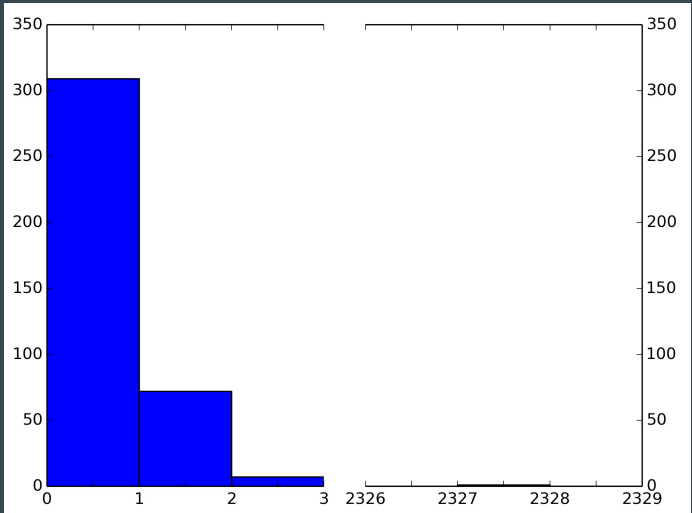
Profile fitting: fit a known profile shape to the reflection to estimate the intensity

Need to estimate background under reflection peak

Don’t know background in signal region so estimate from the surrounding pixels

Background outlier pixels





	With Hot Pixel	Without Hot Pixel
Mean	6.20	0.22
Variance/Mean	2237.90	0.926

~1 for Poisson distribution

Background modelling with outlier pixels

DIALS has multiple options for outlier pixel handling

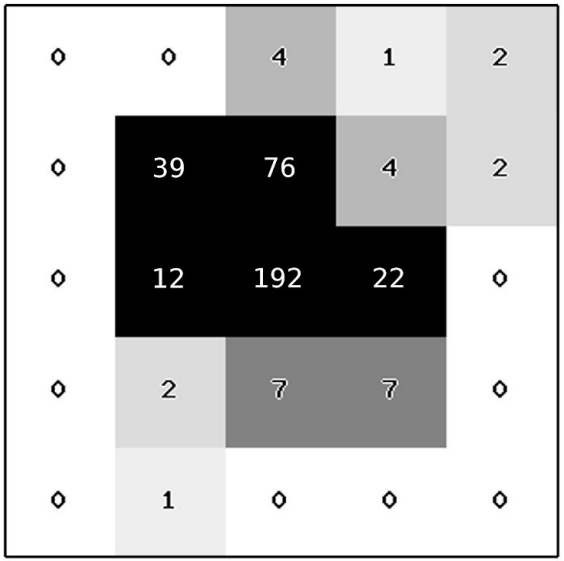
- Truncated - removed percentage of high and low valued pixels
- Normal - remove pixels above and below 3 STD around the mean
- Tukey - remove pixels based on interquartile range
- Plane - compute a plane and remove pixels based on deviation from plane (based on published mosflm algorithm)
- Normal - iteratively remove high valued pixels until they are approximately normally distributed (based on published XDS algorithm)

However, these methods assume a normal distribution and result in biases intensity estimates (particularly for low background)

Default algorithm in DIALS used a GLM algorithm

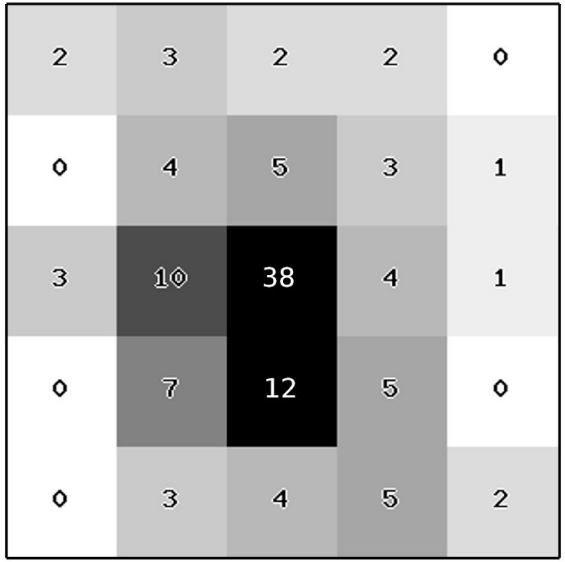
- assumes a Poisson distribution of pixel counts.
- Provides an unbiased estimate of the reflection background

Pixel array detectors: low background



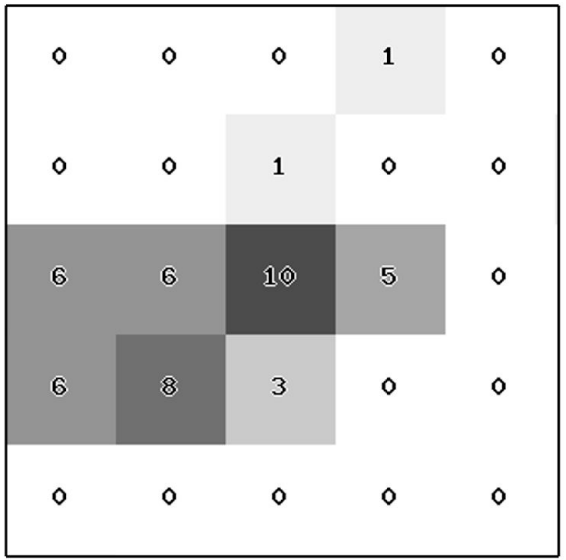
(a)

Thaumatin



(b)

DNA

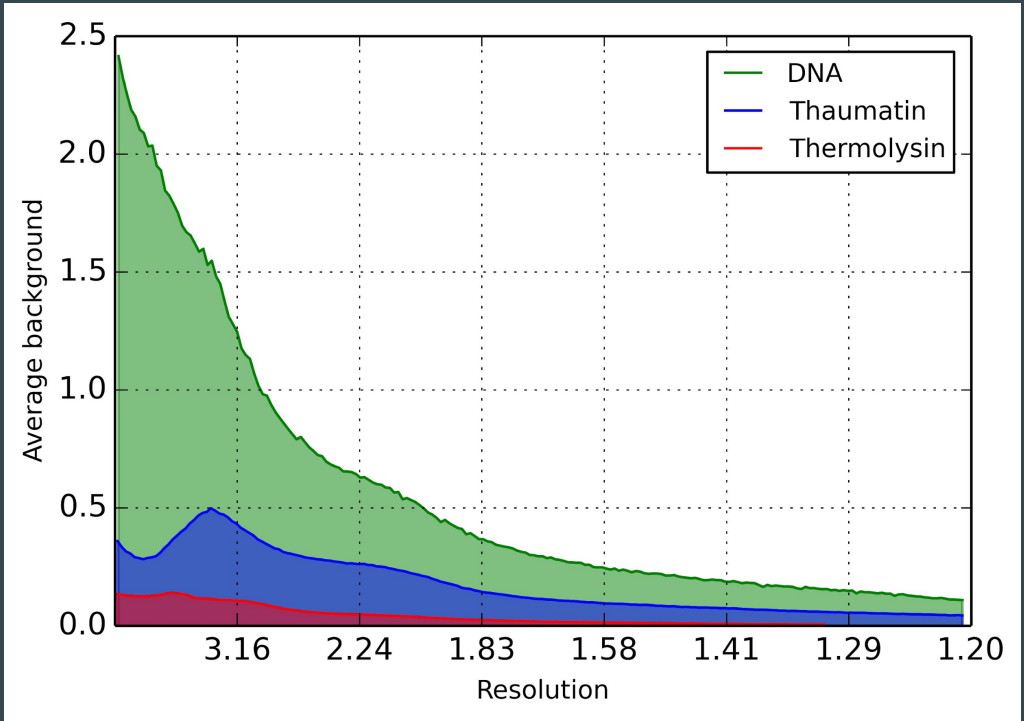


(c)

Thermolysin

Parkhurst et. al (2016) 'Robust background modelling in DIALS', *J. Appl. Cryst.* 49(6), 1912–1921.

Pixel array detectors: low background

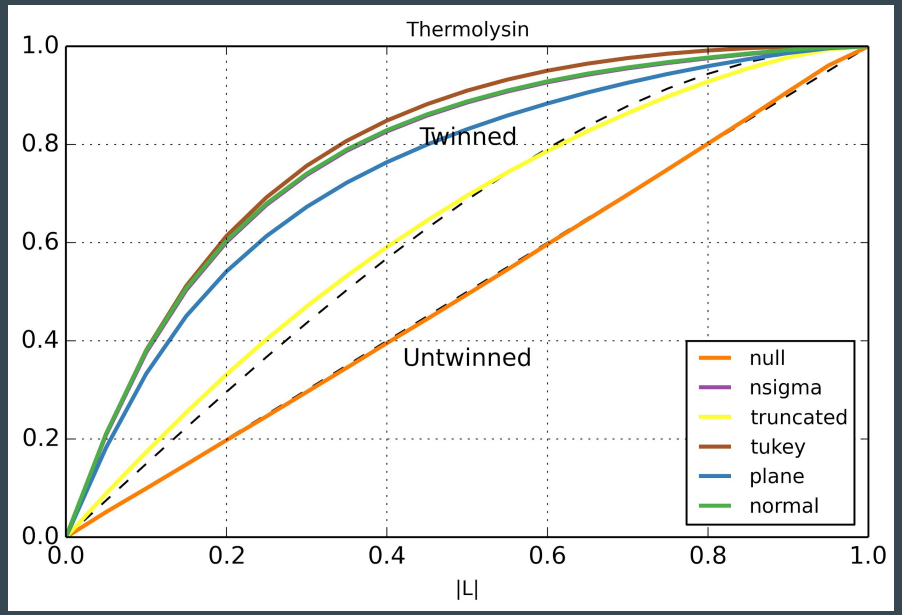


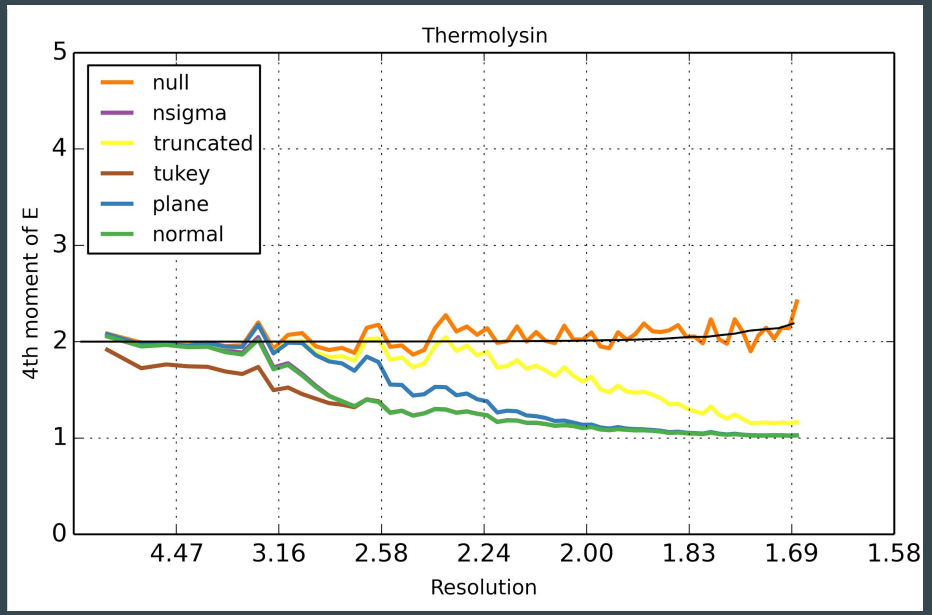
Each dataset has low background over entire resolution range.

Thaumatin and Thermolysin datasets have background less than 1 count per pixel over the whole resolution range

DNA dataset has background less than 1 count per pixel at high resolution

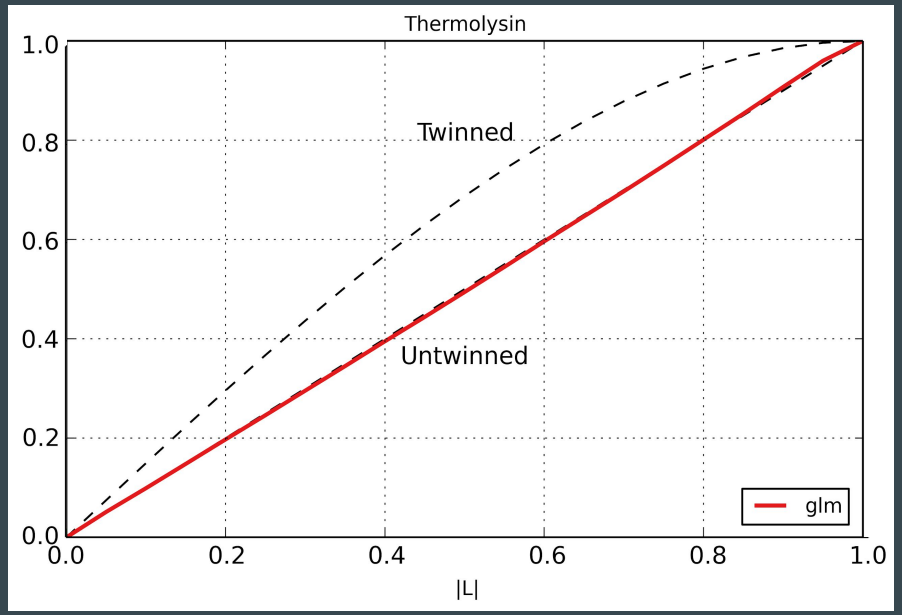
Bias in background determination

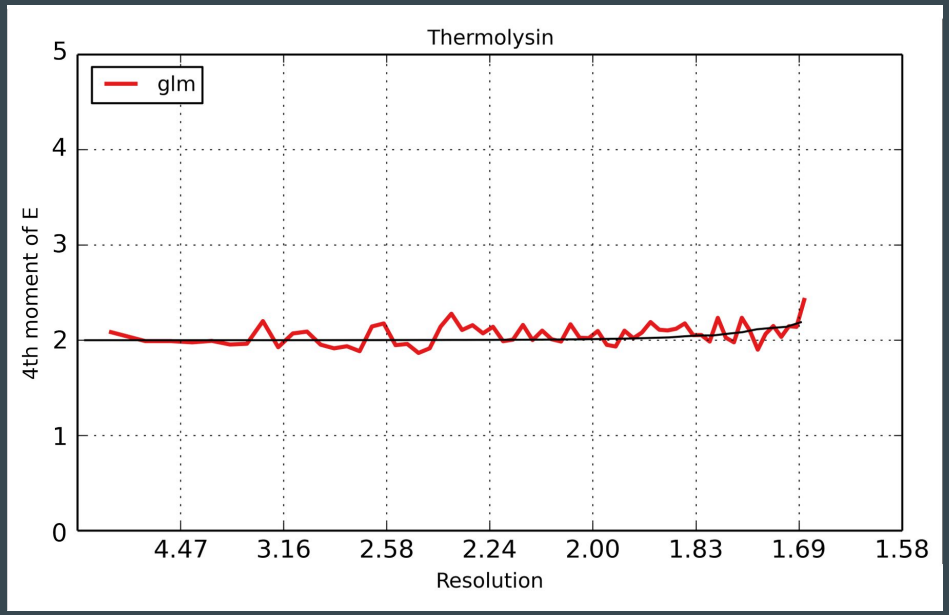




These methods all introduce bias. Background is *underestimated.* Integrated intensities are *overestimated.*

Bias in background determination



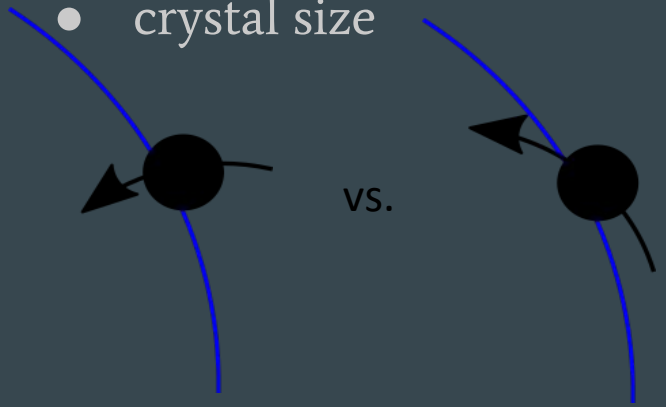


Using a Poisson GLM instead of a least-squares fit removes the bias and still allows outlier rejection

dials.integrate

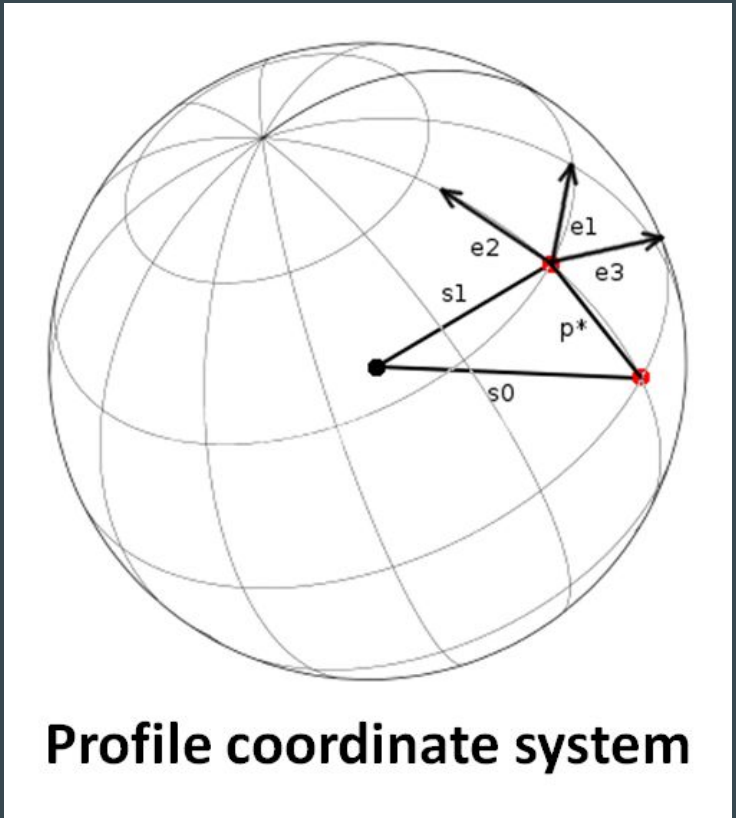
Most programs perform empirical profile modelling using local strong spots.
DIALS makes 3D profile models, like XDS.
The 3D method takes advantage of the Kabsch transform.

- **geometry**
- crystal mosaicity
- crystal shape
- crystal size



- beam size
- beam divergence
- beam spectral dispersion
- **obliquity**
- detector point spread

3D profile fitting coordinate system



Use the kabsch model of a normal distribution on the surface of the Ewald sphere

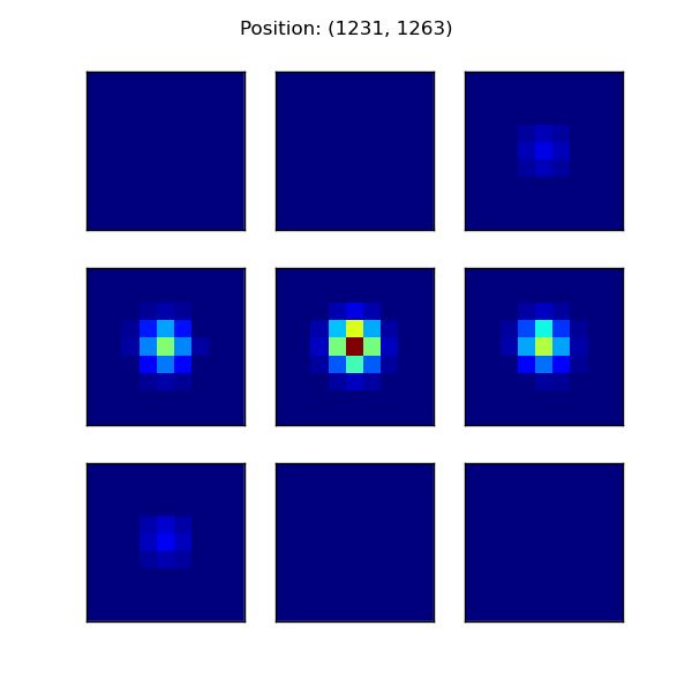
- Corrects for geometrical distortions
- Makes spots appear to have taken shortest path through Ewald sphere
- Model assumes a Gaussian profile in Kabsch coordinate system

$$\exp\left(\frac{-\epsilon_1^2}{2\sigma_D^2}\right)\exp\left(\frac{-\epsilon_2^2}{2\sigma_D^2}\right)\exp\left(\frac{-\epsilon_3^2}{2\sigma_M^2}\right)$$

2 parameters:

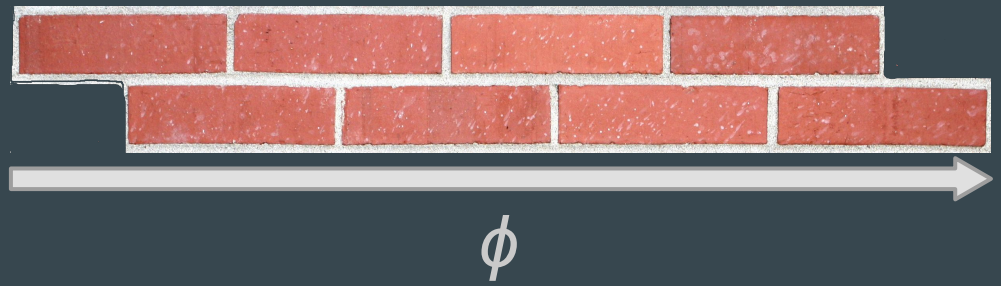
σ_D - roughly "beam divergence"
σ_M - roughly "mosaicity"

dials.integrate



3D profiles are formed for different regions of the detector and different blocks of the ϕ-scan.

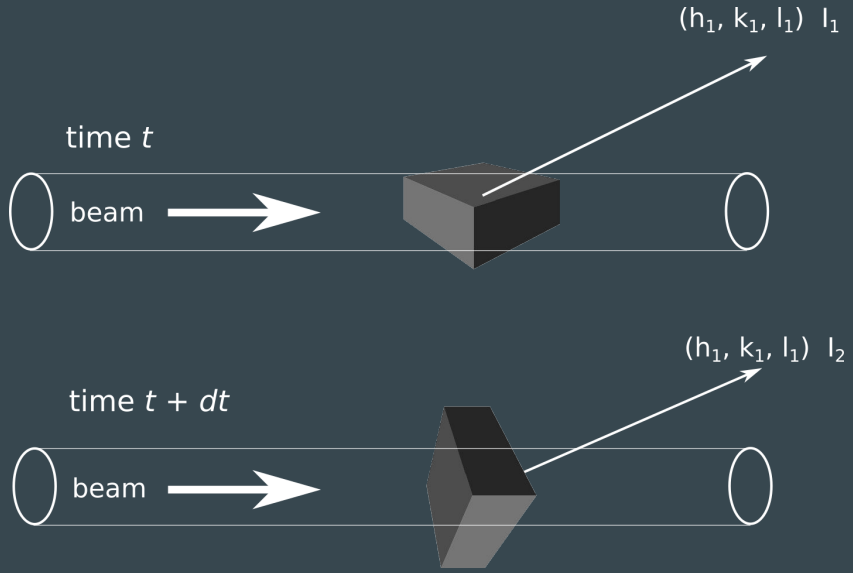
DIALS uses a rectangular grid on the detector, and overlapping blocks in ϕ, with size chosen so that each reflection is integrated fully in one block



dials.scale

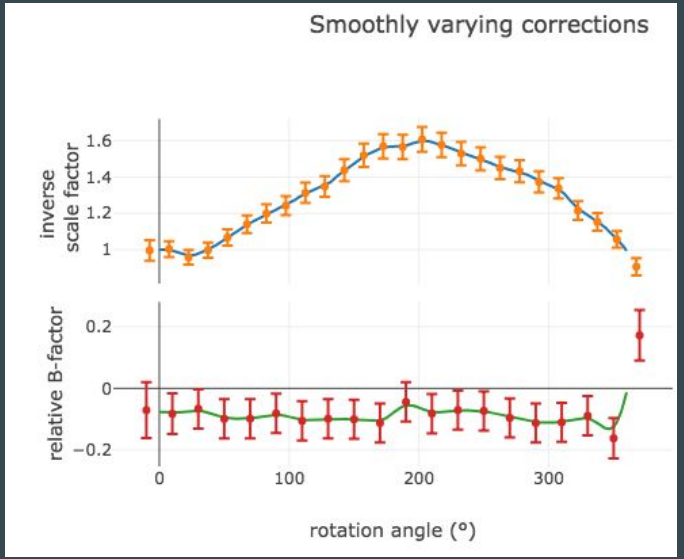
Scaling diffraction data

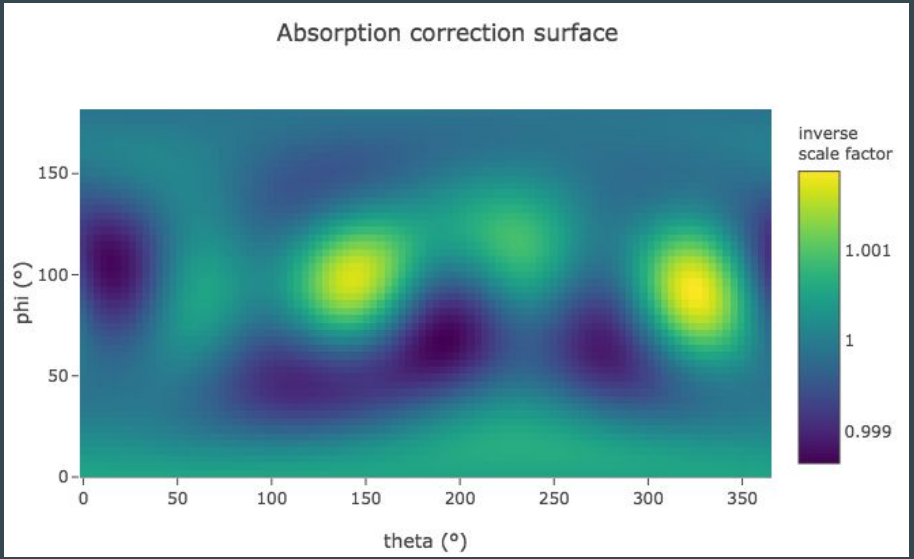
- Correct reflection intensities for experimental effects – beam illumination volume, absorption, radiation damage.
- Apply a correction factor to g_i to each reflection I_i to make symmetry-equivalent reflections most self-consistent across dataset.
- Refine a scaling model with ~50 free parameters. Important to have the right point group symmetry!




```
dials.symmetry integrated_experiments.json integrated.pickle
dials.scale reindexed_experiments.json reindexed_reflections.pickle
```

Example of refined scaling model





- Smoothly varying corrections of a scale factor and *B*-factor depending on rotation angle.
- Absorption correction based on direction of scattering for particular reflection.
- Correction factors g_i obtained by multiplying these three corrections.

dials.report

Experiments

Crystal: Space group: P 4 (No. 75)

Unit cell: $(57.785 \quad 57.785 \quad 150.006 \quad 90.000 \quad 90.000 \quad 90.000)$

U matrix:

$$\begin{pmatrix} 0.3455 & -0.2590 & -0.9020 \\ 0.8913 & 0.3911 & 0.2292 \\ 0.2935 & -0.8831 & 0.3660 \end{pmatrix}$$

B matrix:

$$\begin{pmatrix} 0.0173 & 0.0000 & 0.0000 \\ -0.0000 & 0.0173 & 0.0000 \\ -0.0000 & 0.0000 & 0.0067 \end{pmatrix}$$

A = UB:

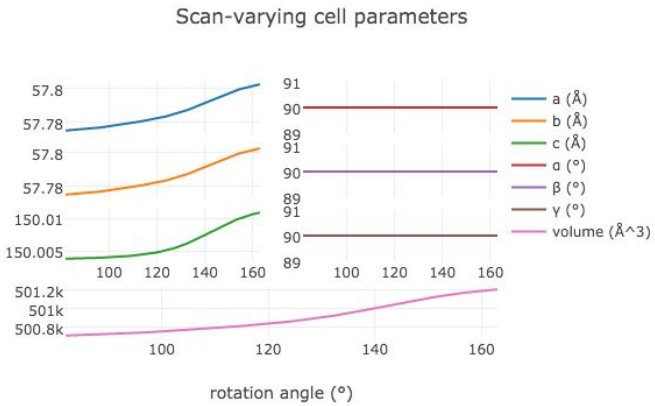
$$\begin{pmatrix} 0.0060 & -0.0045 & -0.0060 \\ 0.0154 & 0.0068 & 0.0015 \\ 0.0051 & -0.0153 & 0.0024 \end{pmatrix}$$

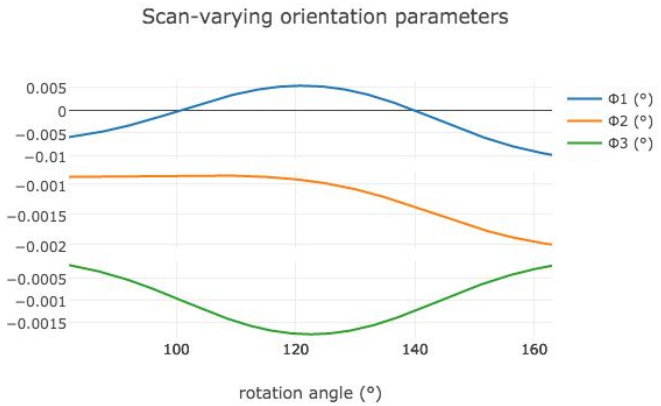
A sampled at 541 scan points

Average unit cell: $(57.786 \quad 57.786 \quad 150.006 \quad 90.000 \quad 90.000 \quad 90.000)$

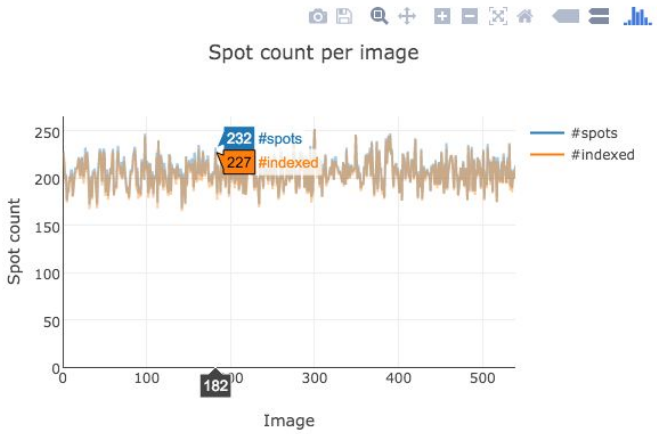
Experimental geometry

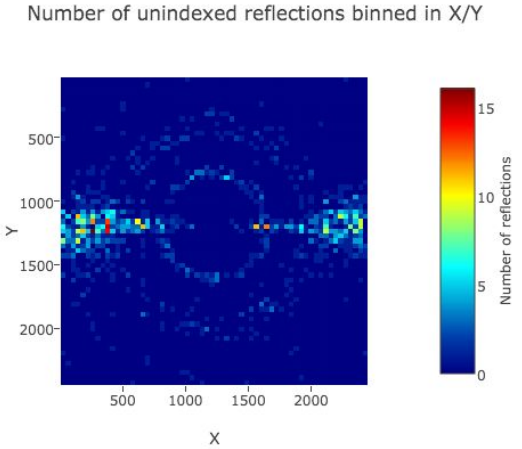
Analysis of scan-varying crystal model



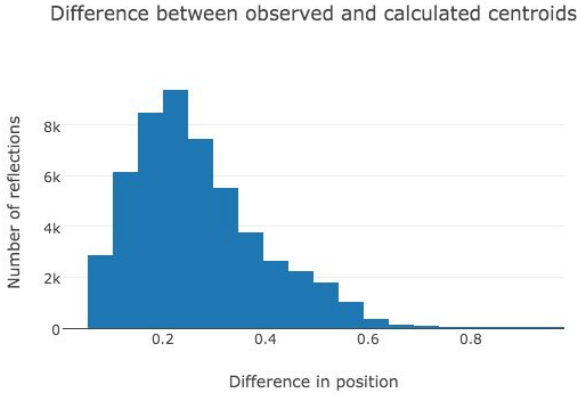


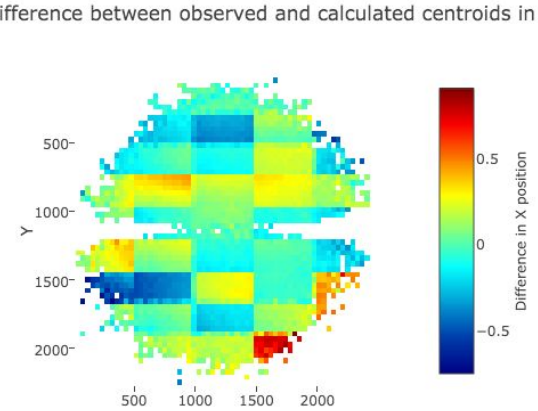
Analysis of strong reflections





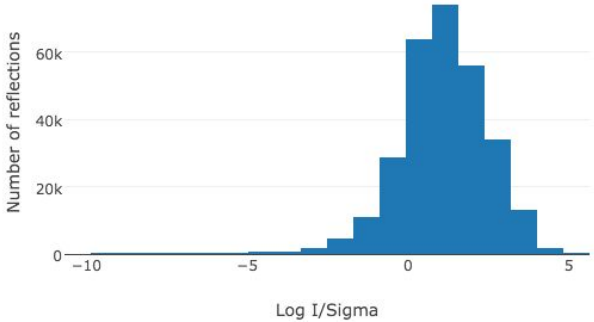
Analysis of reflection centroids



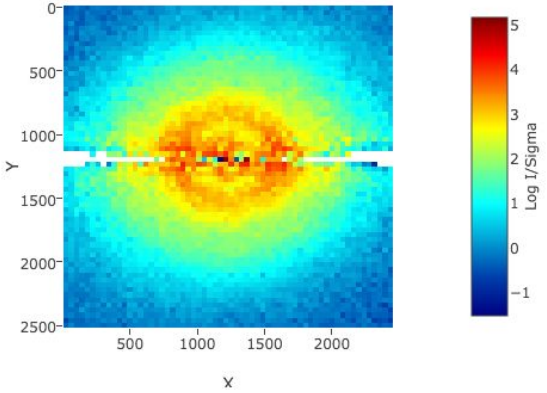


Analysis of reflection intensities

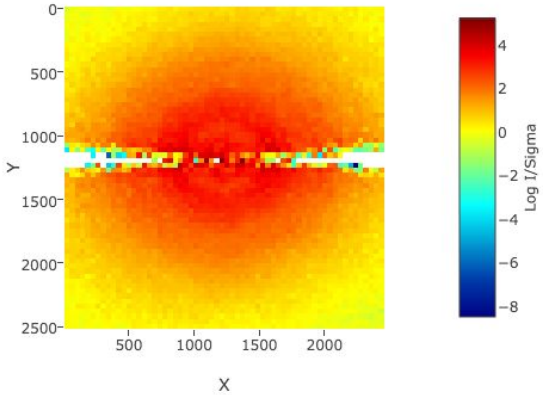
Log I/Sigma histogram



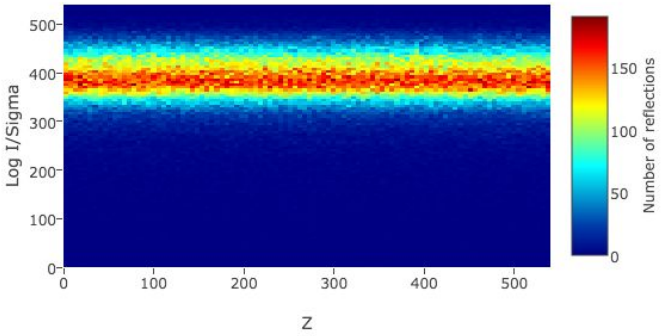
Distribution of I(sum)/Sigma vs X/Y



Distribution of I(prf)/Sigma vs X/Y



Distribution of I/Sigma vs Z



DIALS
Diffraction Integration for Advanced Light Sources

Navigation

About
Installation
Getting started
Documentation
Publications
Links
License

Processing in Detail

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 -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 Thaumatin dataset collected using beamline I04 at Diamond Light Source which is available for download from DOI 10.5281/zenodo.10271

Import

The first stage of step-by-step DIALS processing is to import the data - all that happens here is that the image headers are read, and a file describing their contents (datablock.json) is written. It's worth noting that if this file is changed subsequent processing can use this.

```
dials.import data/th_8_2_0*cbf
```

The output just describes what the software understands of the images it was passed, in this case one sweep of data containing 540 images.

```
The following parameters have been modified:

input {
  datablock = <image files>
}

--------------------------------------------------------------------------------
DataBlock 0
  format: <class 'dxtbx.format.FormatCBFMiniPilatusDLS6MSN100.FormatCBFMiniPilatusDLS6MSN100'>
  num images: 540
  num sweeps: 1
  num stills: 0
--------------------------------------------------------------------------------
Writing datablocks to datablock.json
```

Find Spots

Navigation

Quick start guide

If you don't like reading manuals and just want to get started, try:

```
xia2 -2d /here/are/my/images
```

or:

```
xia2 -3d /here/are/my/images
```

or:

```
xia2 -dials /here/are/my/images
```

(remembering of course -atom X if you want anomalous pairs separating in scaling.) If this appears to do something sensible then you may well be home and dry. Some critical options:

Option	Usage
-atom X	tell xia2 to separate anomalous pairs i.e. I(+) ≠ I(−) in scaling
-2d	tell xia2 to use MOSFLM and Aimless
-3d	tell xia2 to use XDS and XSCALE
-3dii	tell xia2 to use XDS and XSCALE, indexing with peaks found from all images
-dials	tell xia2 to use DIALS and Aimless

If this doesn't hit the spot, you'll need to read the rest of the documentation.

Summary

- DIALS is new software for diffraction image integration, combining the best parts of older software in a modular and extensible toolkit
- New algorithms are now being developed within the DIALS toolkit
- xia2 is the “friendly” DIALS user interface for synchrotron data, and is bundled with DIALS
- DUI is the new DIALS GUI (see Luiso’s talk tomorrow)
- Get DIALS 1.10.1 for Mac and Linux or the source code from http://dials.github.io/ under the BSD license
- DIALS is also available from CCP4 7.0 (includes Windows)

Reference

Please cite the main reference if you use DIALS in your work:

Winter *et al.* (2018) Acta Cryst. **D74**, 85-97.

STRUCTURAL BIOLOGY

ISSN 2059-7983

DIALS: implementation and evaluation of a new integration package

Graeme Winter,[a]* David G. Waterman,[c,d] James M. Parkhurst,[a,e] Aaron S. Brewster,[b] Richard J. Gildea,[a] Markus Gerstel,[a] Luis Fuentes-Montero,[a] Melanie Vollmar,[a] Tara Michels-Clark,[b] Iris D. Young,[b] Nicholas K. Sauter[b] and Gwyndaf Evans[a]*

[a]Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot OX11 0DE, England, [b]Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA, [c]STFC Rutherford Appleton Laboratory, Didcot OX11 0FA, England, [d]CCP4, Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot OX11 0FA, England, and [e]Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, England. *Correspondence e-mail: graeme.winter@diamond.ac.uk, gwyndaf.evans@diamond.ac.uk

Received 14 June 2017
Accepted 30 November 2017

Keywords: X-ray diffraction; data processing; methods development; *DIALS*.

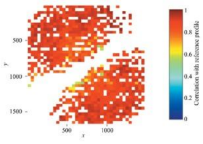
Supporting information: this article has supporting information at journals.iucr.org/d

The *DIALS* project is a collaboration between Diamond Light Source, Lawrence Berkeley National Laboratory and CCP4 to develop a new software suite for the analysis of crystallographic X-ray diffraction data, initially encompassing spot finding, indexing, refinement and integration. The design, core algorithms and structure of the software are introduced, alongside results from the analysis of data from biological and chemical crystallography experiments.

1. Introduction

X-ray crystallography is the dominant method for the determination of the atomic structure of biological macromolecules. Macromolecular crystallography (MX) has evolved over decades into an essentially routine method for the majority of structures being investigated. Incremental improvements in detector technology, X-ray sources, beamline instrumentation (both in optics and endstation) and automation of sample handling have contributed to the success of the method. The overwhelming majority of diffraction data resulting in PDB depositions over the last 2–3 decades have been analysed using just four programs: *XDS* (Kabsch, 2010*b*), *MOSFLM* (Leslie, 2006), *HKL*-2000/*DENZO* (Otwinowski & Minor, 1997) and *d*TREK* (Pflugrath, 1999). For chemical crystallography, *SAINT* (Bruker AXS Inc., Madison, Wisconsin, USA) and *EVAL* (Duisenberg *et al.*, 2003; Schreurs *et al.*, 2010) as well as *d*TREK* are in common use. Significant effort by a relatively small number of developers over this time has been critical to producing the diffraction-intensity data sets that are the raw material of structure determination.

In more recent years there has been a step change in MX throughput, driven principally by the availability of new X-ray sources and data-collection methodologies (Emma *et al.*, 2010; Ishikawa *et al.*, 2012; White *et al.*, 2012; Gati *et al.*, 2014; Stellato *et al.*, 2014; Sierra *et al.*, 2016; Fuller *et al.*, 2017), high-frame-rate pixel-array detectors (Henrich *et al.*, 2009), fast sample exchange (Russi *et al.*, 2016) and automated data analysis (Winter, 2010; Winter & McAuley, 2011; Vonrhein *et al.*, 2011). This allows larger numbers of smaller samples to be used, with correspondingly more challenging data. New algorithms and approaches to data analysis are therefore required to address the novel approaches to the measurement of diffraction data sets. The initial focus of the development of



OPEN ACCESS

Acknowledgements

DIALS East

Gwyndaf Evans, Graeme Winter, David Waterman, James Parkhurst, Richard Gildea, Luis Fuentes-Montero, Markus Gerstel, Nick Devenish, James Beilsten-Edmands, Ben Williams, Melanie Vollmar

DIALS West

Nick Sauter, Aaron Brewster, Tara Michels-Clark, Iris Young

Lots of other people

Garib Murshudov, Andrew Leslie, Phil Evans, Harry Powell, Takanori Nakane, Andrea Thorn


diamond


BioStructX


Research Complex
at Harwell


Science & Technology
Facilities Council

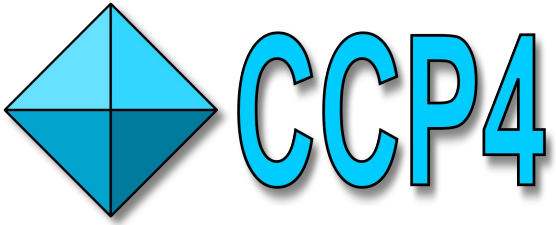

UNIVERSITY OF
CAMBRIDGE


DIALS
Diffraction Integration for Advanced Light Sources


wellcome


BERKELEY LAB
Lawrence Berkeley National Laboratory


MRC
Laboratory of
Molecular Biology


CCP4

Thanks for listening!

•••

https://dials.diamond.ac.uk