



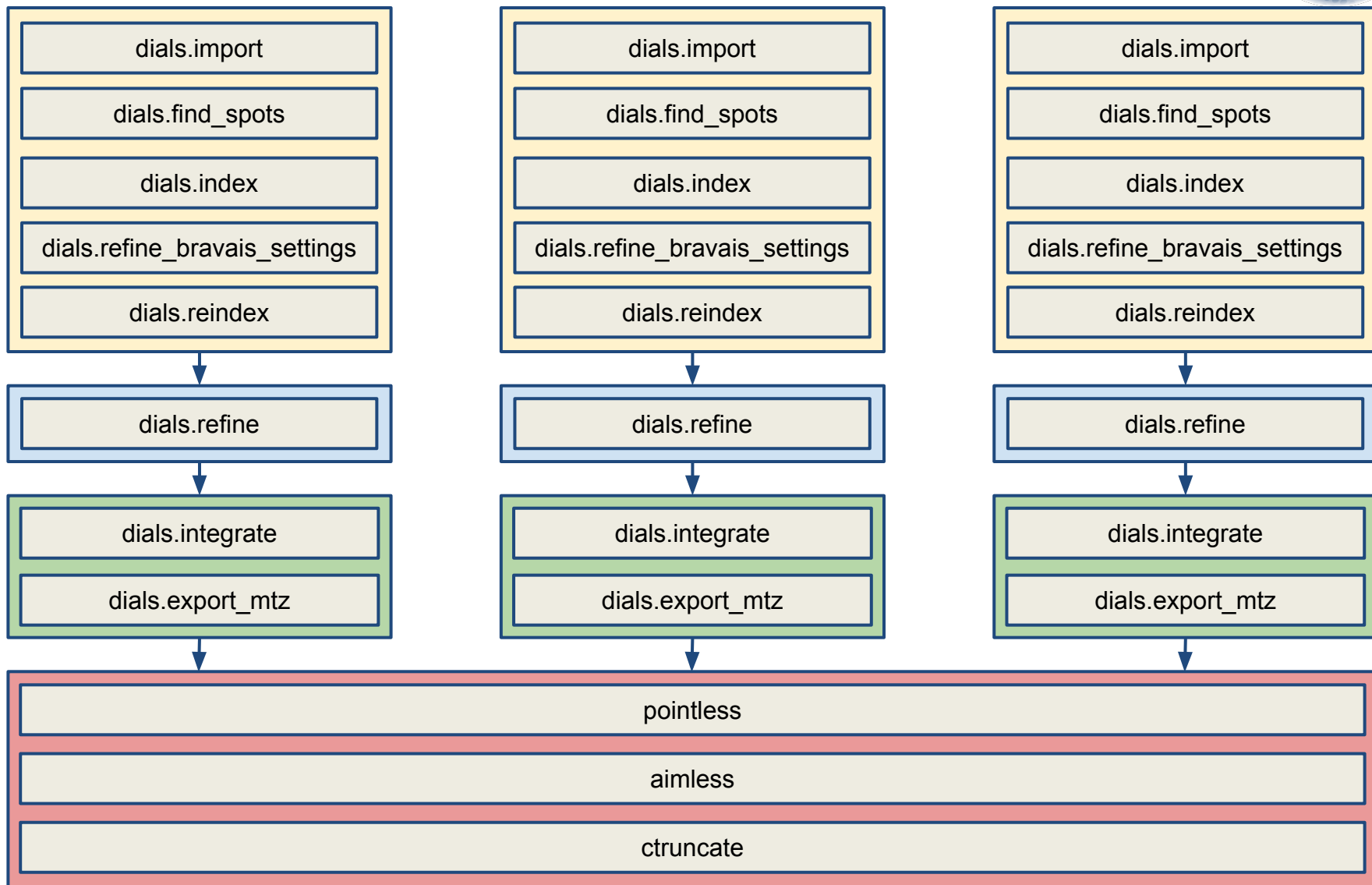
Indexing diffraction patterns: dials.index (and friends)

Richard Gildea

6th DIALS Workshop:
Deployment at Light Source Facilities
May 27, 2015



xia2 -dials



dials.import

Input: images

Output: datablock.json

```
$ dials.import /path/to/i04_bag_training/th_8_2_*.cbf
```

```
DataBlock 0
```

```
  format: <class 'dxtbx.format.FormatCBFMiniPilatusDLS6MSN100.FormatCBFMiniPilatusDLS6MSN100'>
  num images: 540
  num sweeps: 1
  num stills: 0
```

```
$ dials.import template=/path/to/i04_bag_training/th_8_2_####.cbf
```

dials.show_models

```
$ dials.show_models datablock.json
/path/to/i04_bag_training/th_8_2_####.cbf
Detector:
Panel:
  pixel_size:{0.172,0.172}
  image_size: {2463,2527}
  trusted_range: {-1,161977}
  thickness: 0.32
  material: Si
  mu: 39.2203
  fast_axis: {1,0,0}
  slow_axis: {0,-1,0}
  origin: {-210.76,205.277,-265.27}

Beam:
  wavelength: 0.97625
  sample to source direction : {0,0,1}
  divergence: 0
  sigma divergence: 0
  polarization normal: {0,1,0}
  polarization fraction: 0.999

Scan:
  image range:      {1,540}
  oscillation:      {82,0.15}

Goniometer:
  Rotation axis:      {1,0,0}
  Fixed rotation:      {1,0,0,0,1,0,0,0,1}
  Setting rotation:{1,0,0,0,1,0,0,0,1}
```

dials.image_viewer



```
$ dials.image_viewer datablock.json
```

The screenshot shows the dials.image_viewer application window. The main area displays a diffraction image with a grid of spots. The top toolbar includes buttons for 'Load file', 'Settings', 'Zoom', 'Previous', 'Next', and 'Save As...'. The 'Settings' panel on the right contains various controls: 'Zoom level' (Auto), 'Color scheme' (grayscale), 'Brightness' (50), and checkboxes for 'Show resolution rings', 'Show ice rings', 'Mark beam center', 'Mark centers of mass', 'Spot max pixels', 'Spot all pixels', 'Draw reflection shoebox', 'Show predictions', and 'Show hkl'. Below these are input fields for 'Sigma background' (6.0), 'Sigma strong' (3.0), 'Global Threshold' (0.0), 'Min. local' (2), 'Gain' (1.0), and 'Kernel size' (3 3). A vertical stack of buttons on the right allows switching between different views: 'image', 'mean', 'variance', 'dispersion', 'global', 'sigma_b', 'sigma_s', and 'threshold'. An 'Update display' button is at the bottom right. The status bar at the bottom left reads: 'Click and drag to pan; middle-click and drag to plot intensity profile, right-click to zoom'.

dials.find_spots



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

```
Setting spotfinder.filter.min_spot_size=3  
Configuring spot finder from input parameters
```

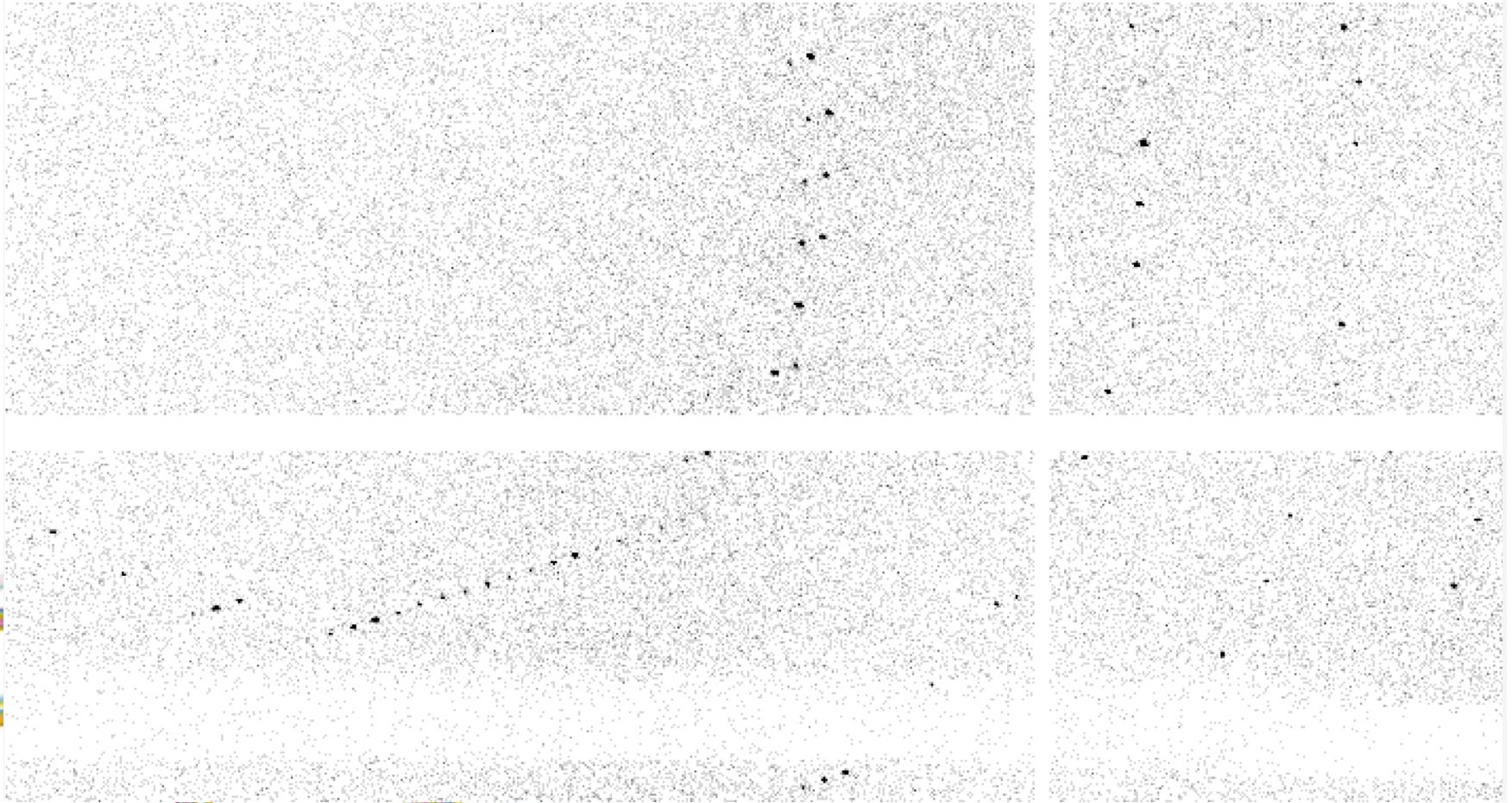
```
-----  
Finding strong spots in imageset 0  
-----
```

```
Finding spots in image 1 to 540...  
Extracting strong pixels from images (may take a while)  
Extracted strong pixels from images  
Merging 8 pixel lists  
Merged 8 pixel lists with 922120 pixels  
Extracting spots  
Extracted 219125 spots  
Calculating 219125 spot centroids  
Calculated 219125 spot centroids  
Calculating 219125 spot intensities  
Calculated 219125 spot intensities  
Found 1 possible hot spots  
Found 1 possible hot pixel(s)  
Filtering 219125 spots by number of pixels  
Filtered 116321 spots by number of pixels  
Filtering 116321 spots by peak-centroid distance  
Filtered 116082 spots by peak-centroid distance
```

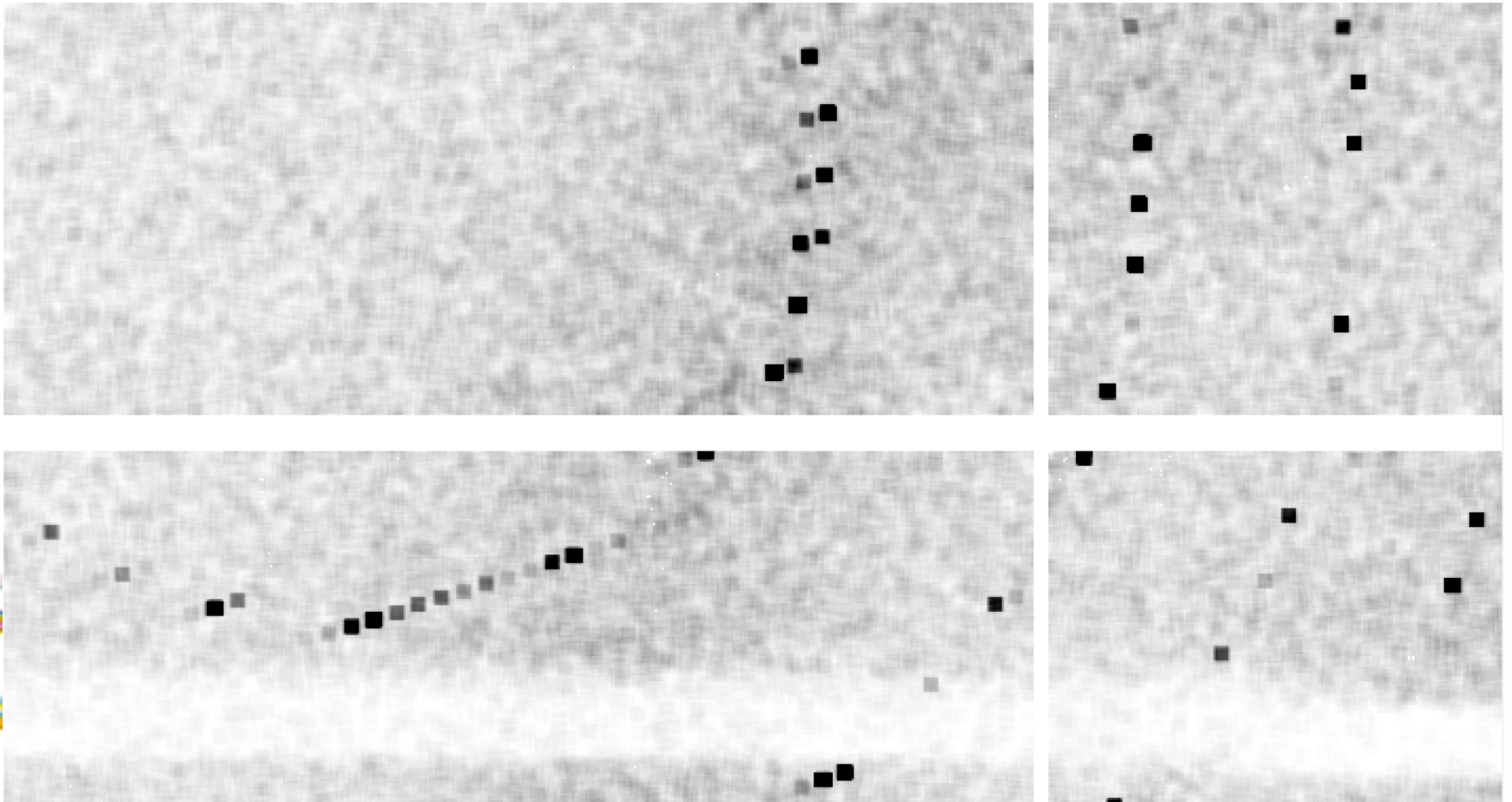
```
-----  
Saving 116082 reflections to strong.pickle  
Saved 116082 reflections to strong.pickle  
Time Taken: 31.768495
```



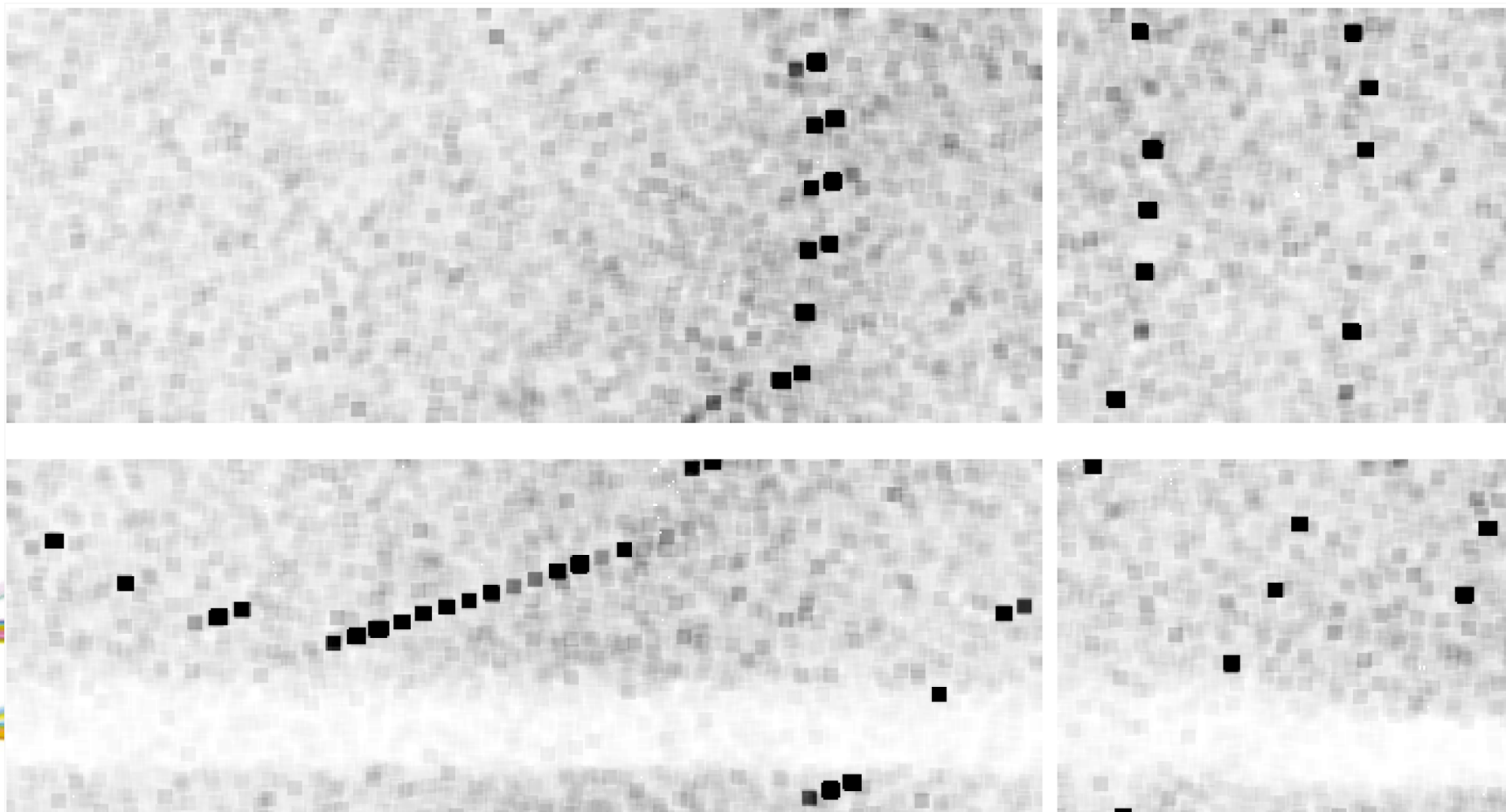
Raw image



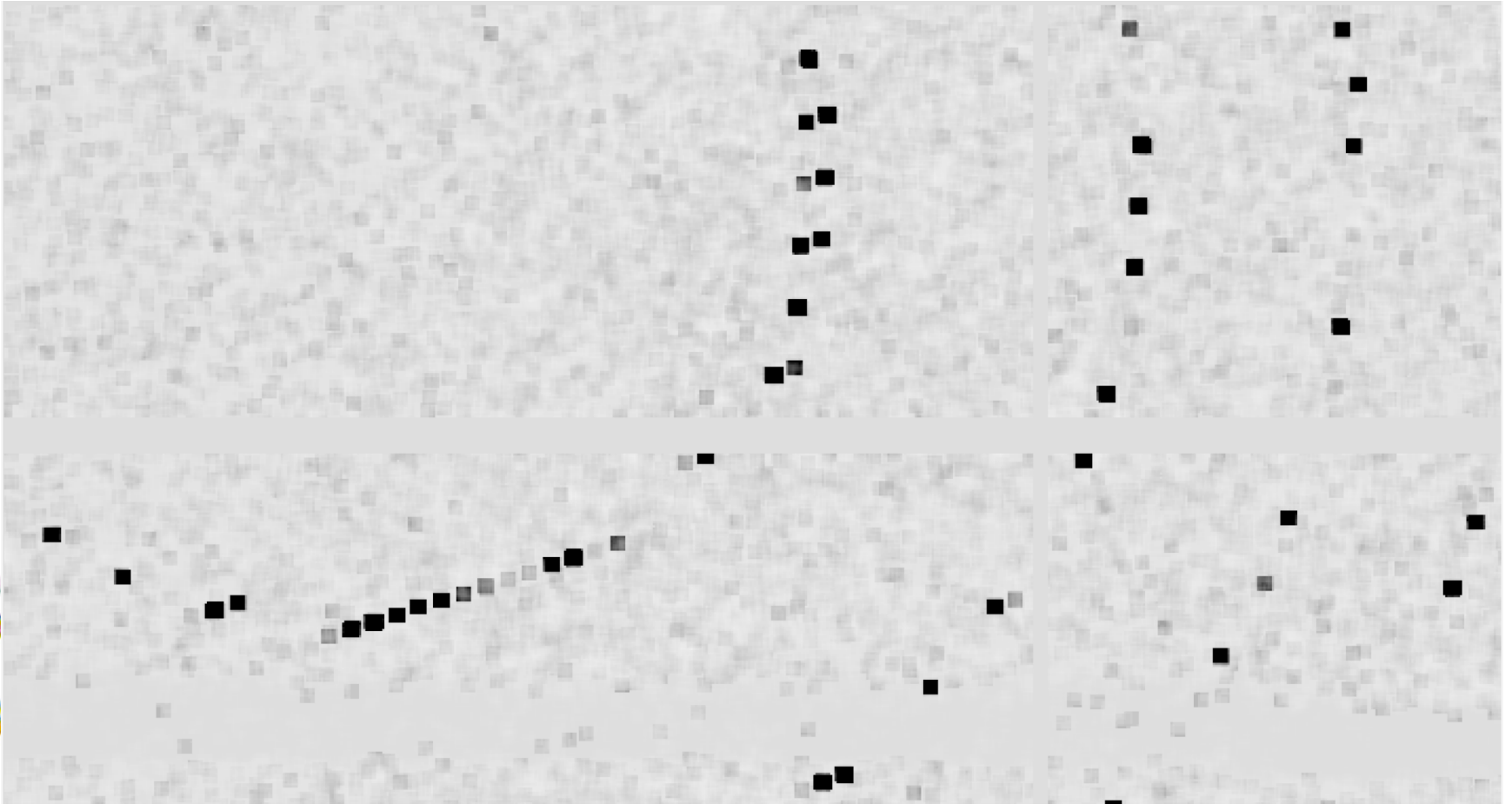
Local mean map



Local variance map



Dispersion index map



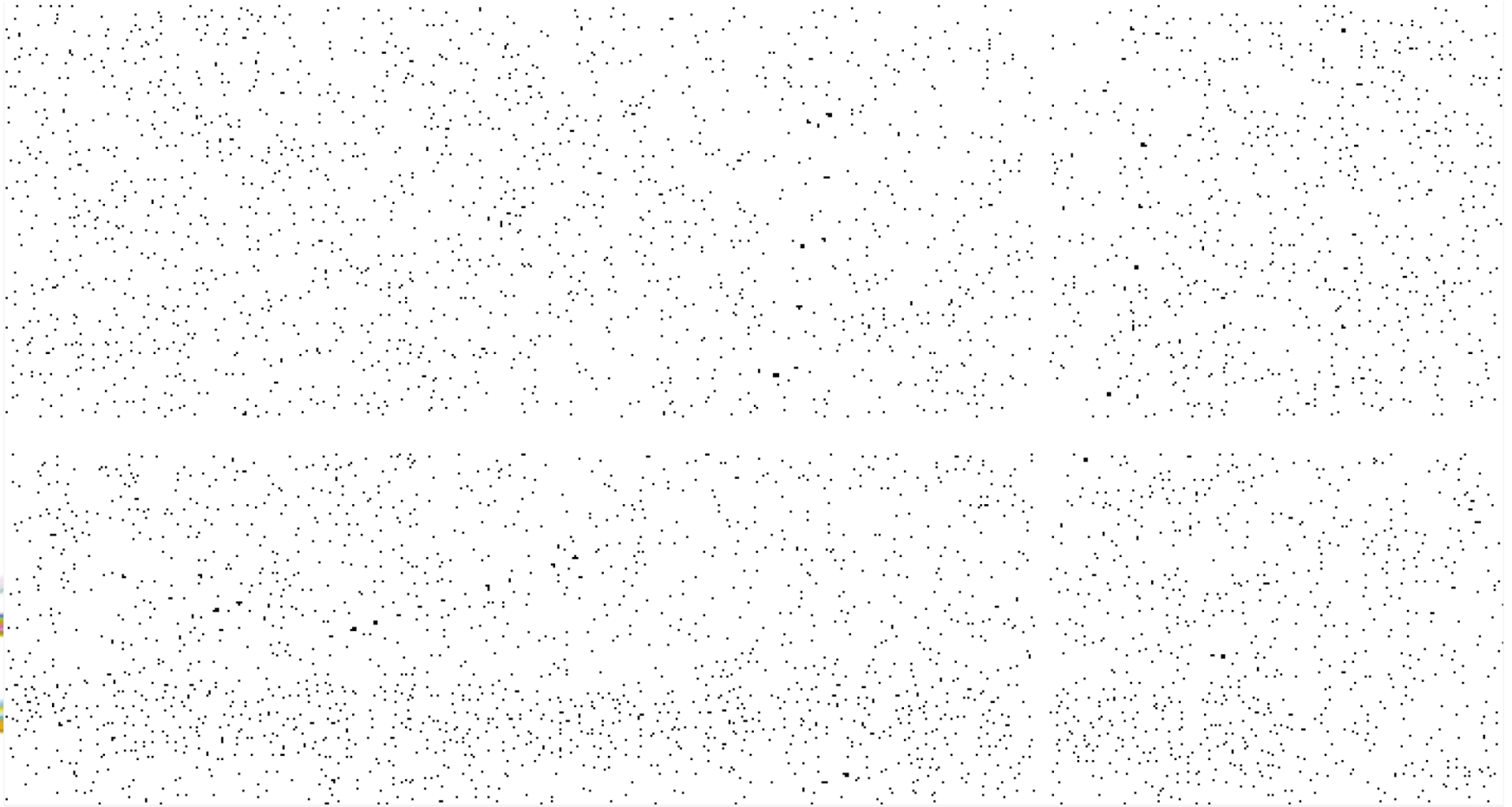
Global threshold filter



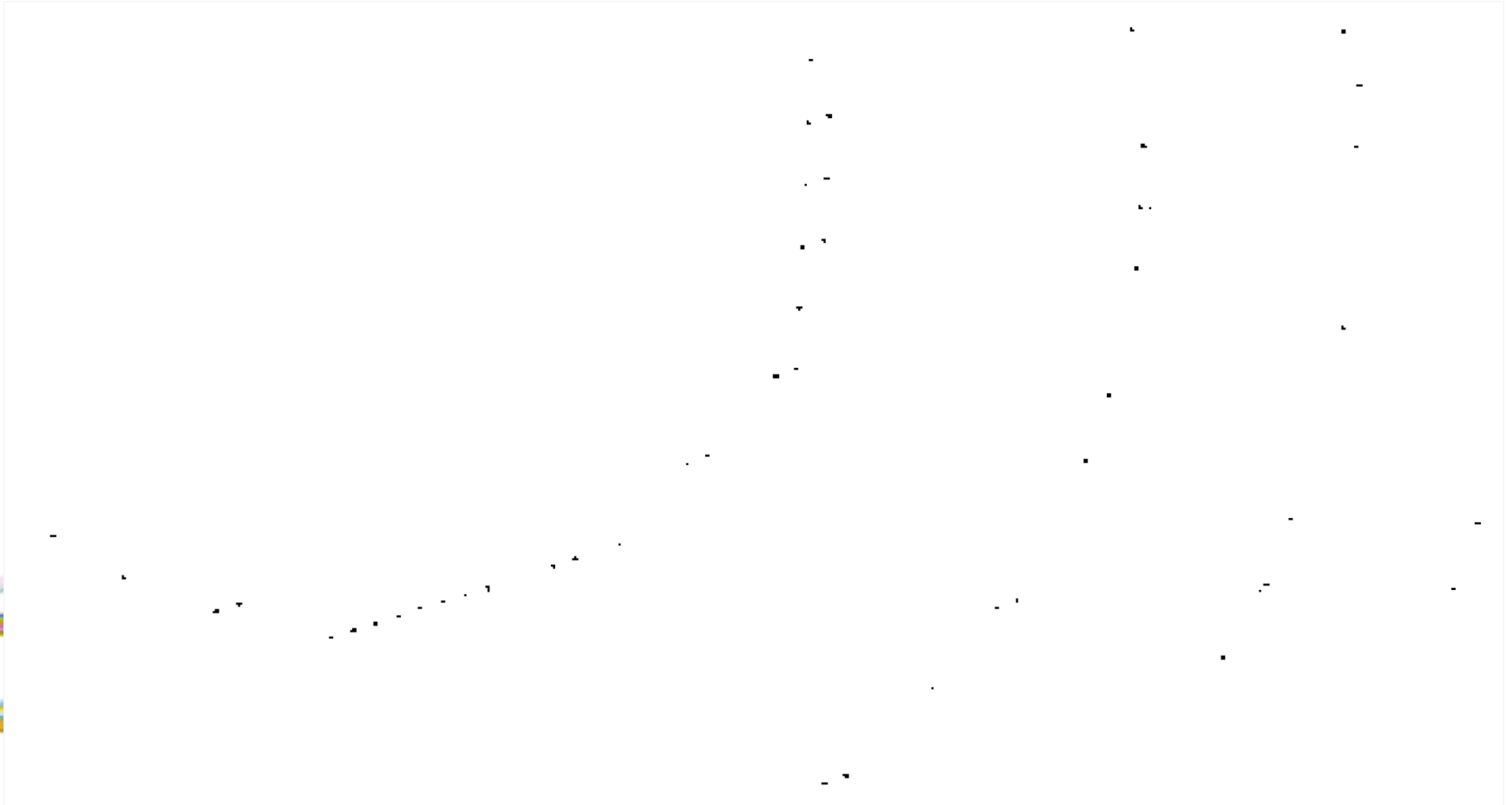
sigma_b filter



sigma_s filter



Threshold filter



dials.find_spots

Further filters:

- Number of pixels
 - filter.min_spot_size=3
- Peak-centroid distance
 - filter.max_separation=2
- Resolution
 - filter.d_min, filter.d_max
- Ice rings
 - filter.ice_rings=true
- Untrusted regions
 - filter.untrusted_polygon

Also available as server/client:

- Now running at Diamond as alternative to DISTL
- Improved spot-finding for small spots on weak backgrounds compared to DISTL

dials.image_viewer



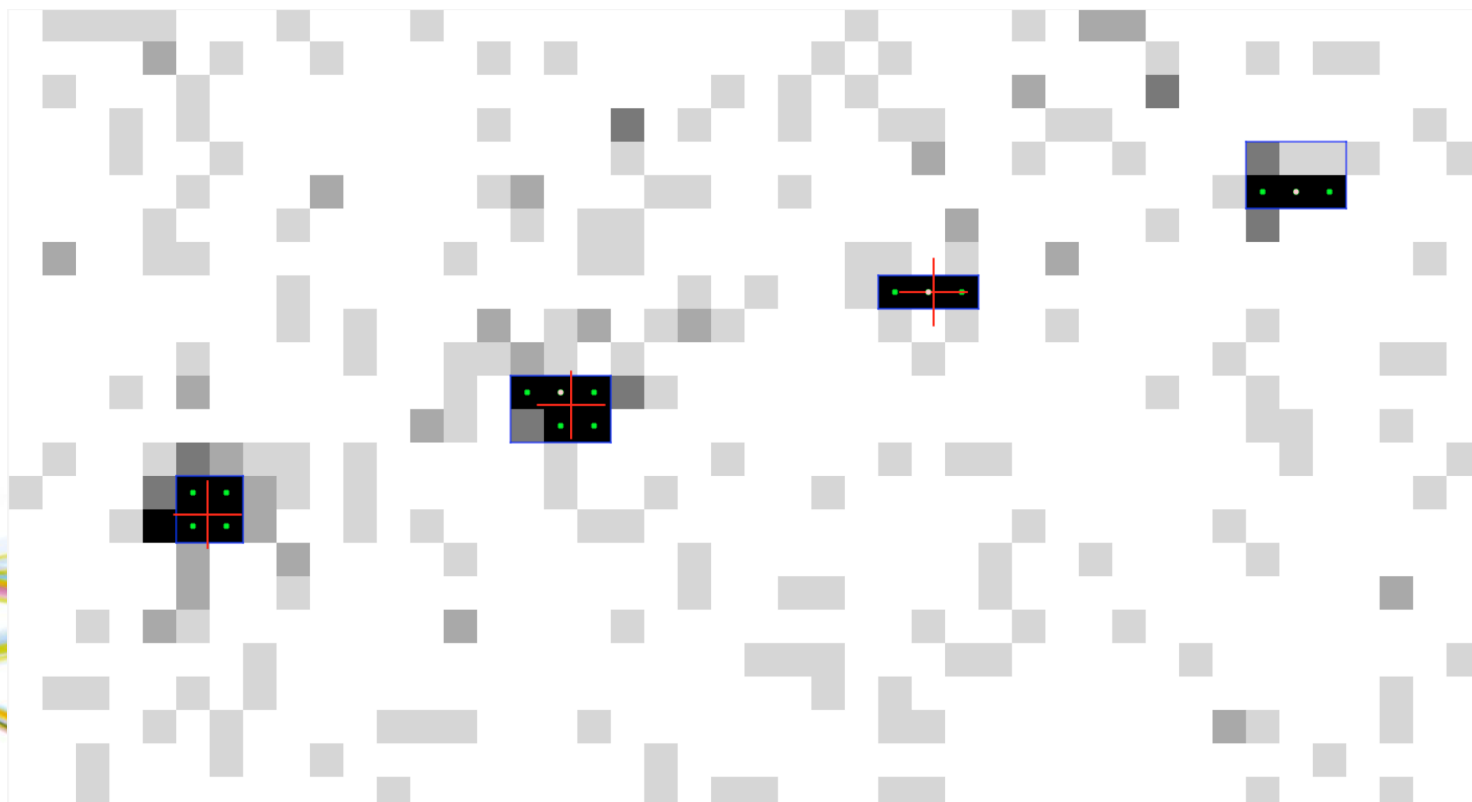
```
$ dials.image_viewer datablock.json strong.pickle
```



dials.image_viewer



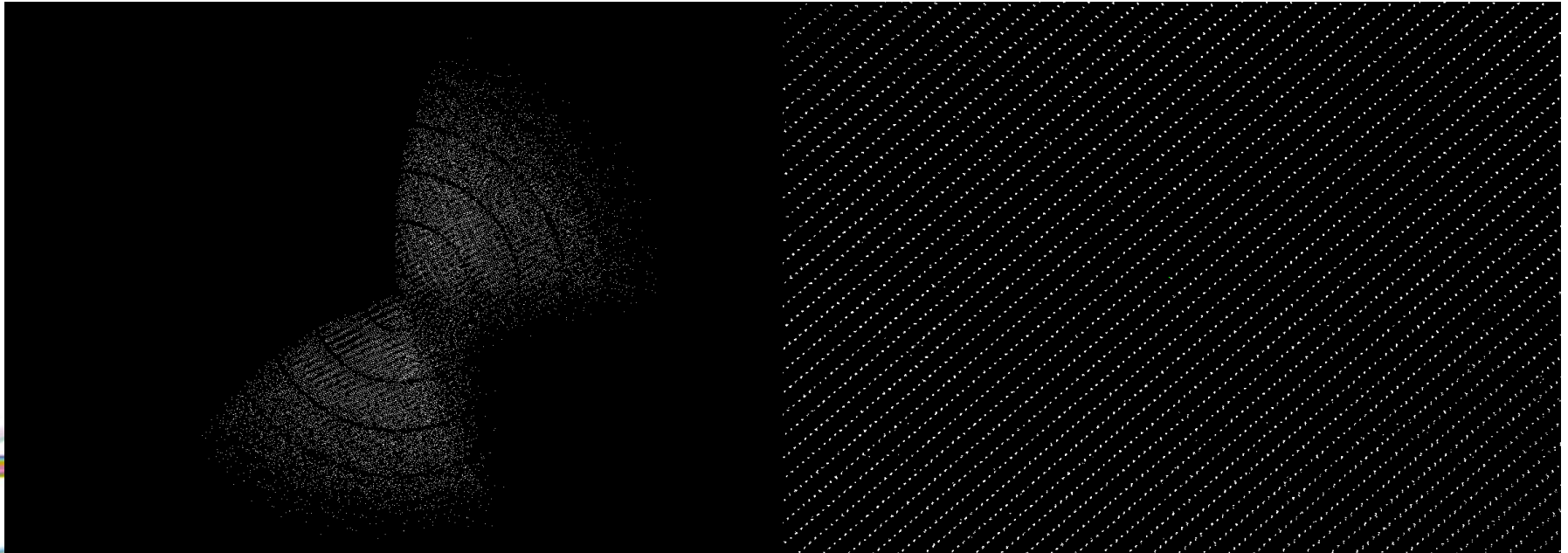
```
$ dials.image_viewer datablock.json strong.pickle
```



dials.reciprocal_lattice_viewer



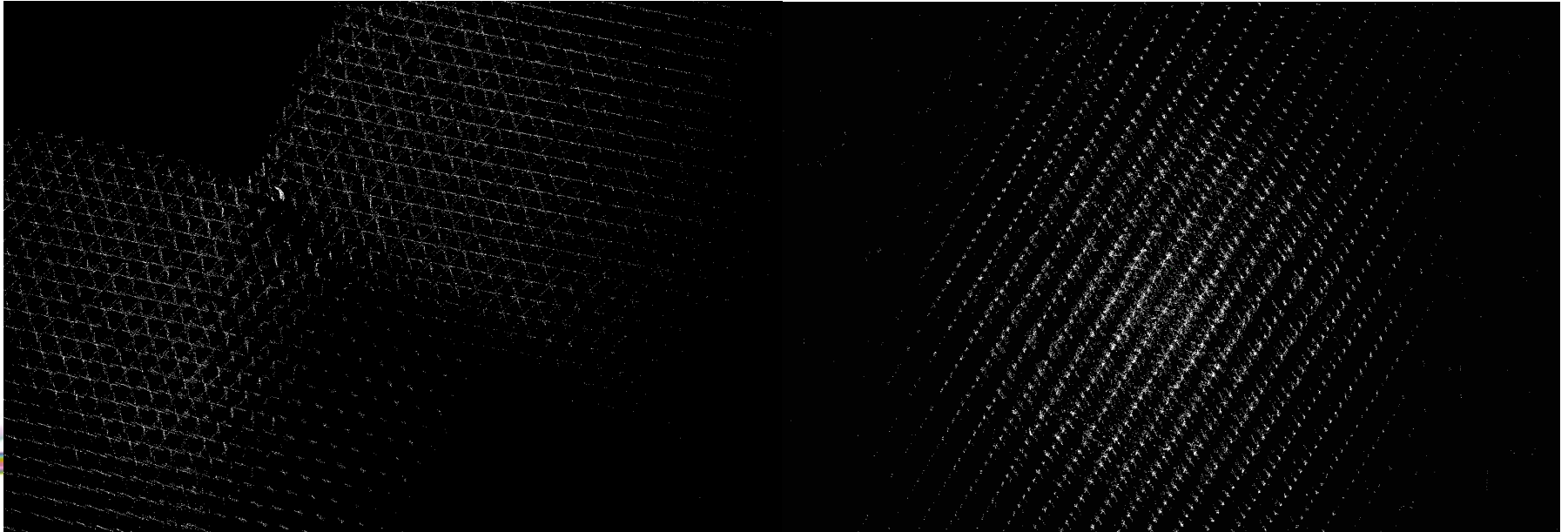
```
$ dials.reciprocal_lattice_viewer datablock.json strong.pickle
```



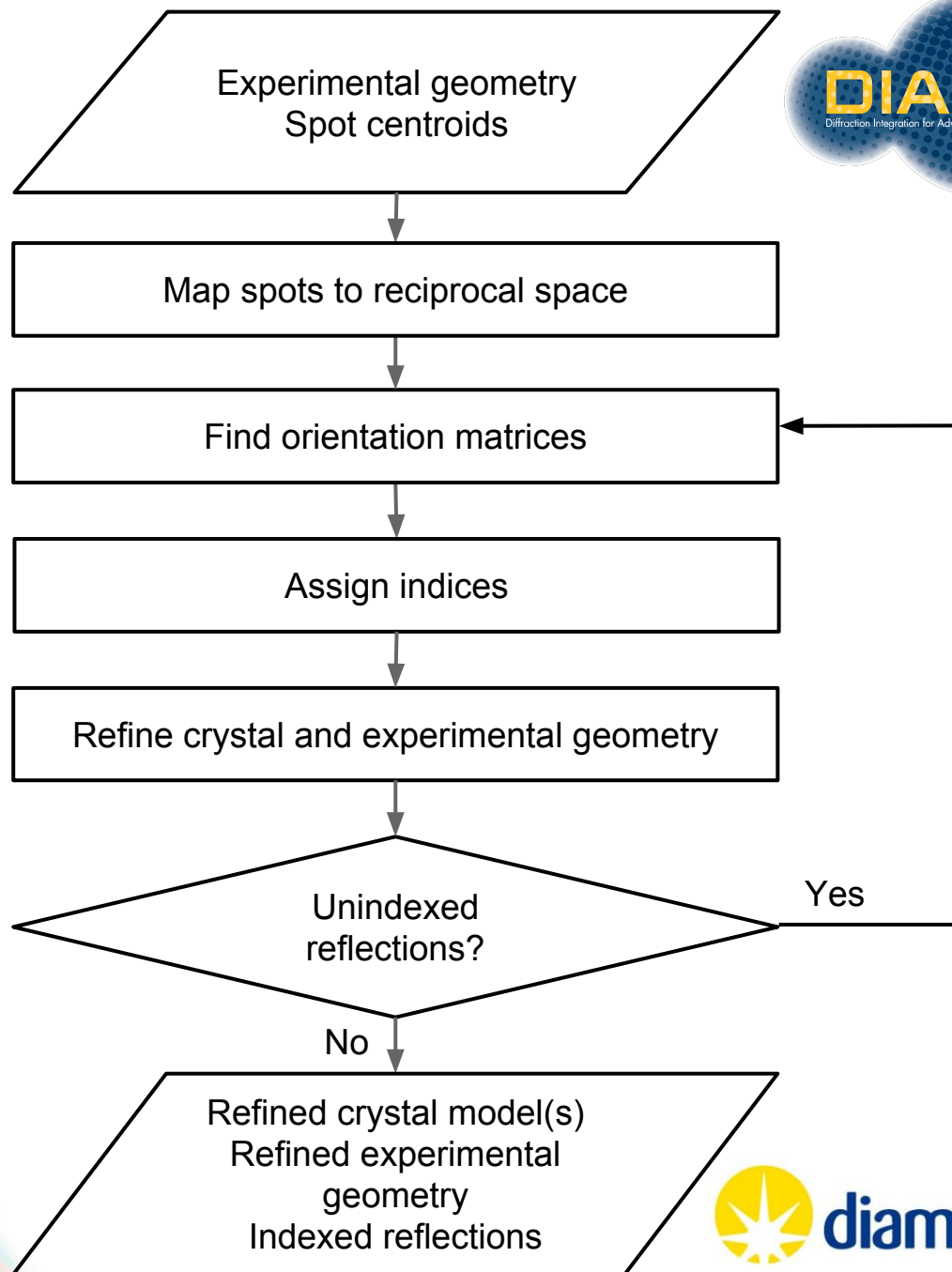
dials.reciprocal_lattice_viewer



```
$ dials.reciprocal_lattice_viewer datablock.json strong.pickle
```



dials.index



dials.index

```
$ dials.index datablock.json strong.pickle
```

```
Found max_cell: 199.1 Angstrom
```

```
Setting d_min: 3.89
```

```
...
```

```
RMSDs by experiment:
```

Exp	Nref	RMSD_X (px)	RMSD_Y (px)	RMSD_Z (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}}
```

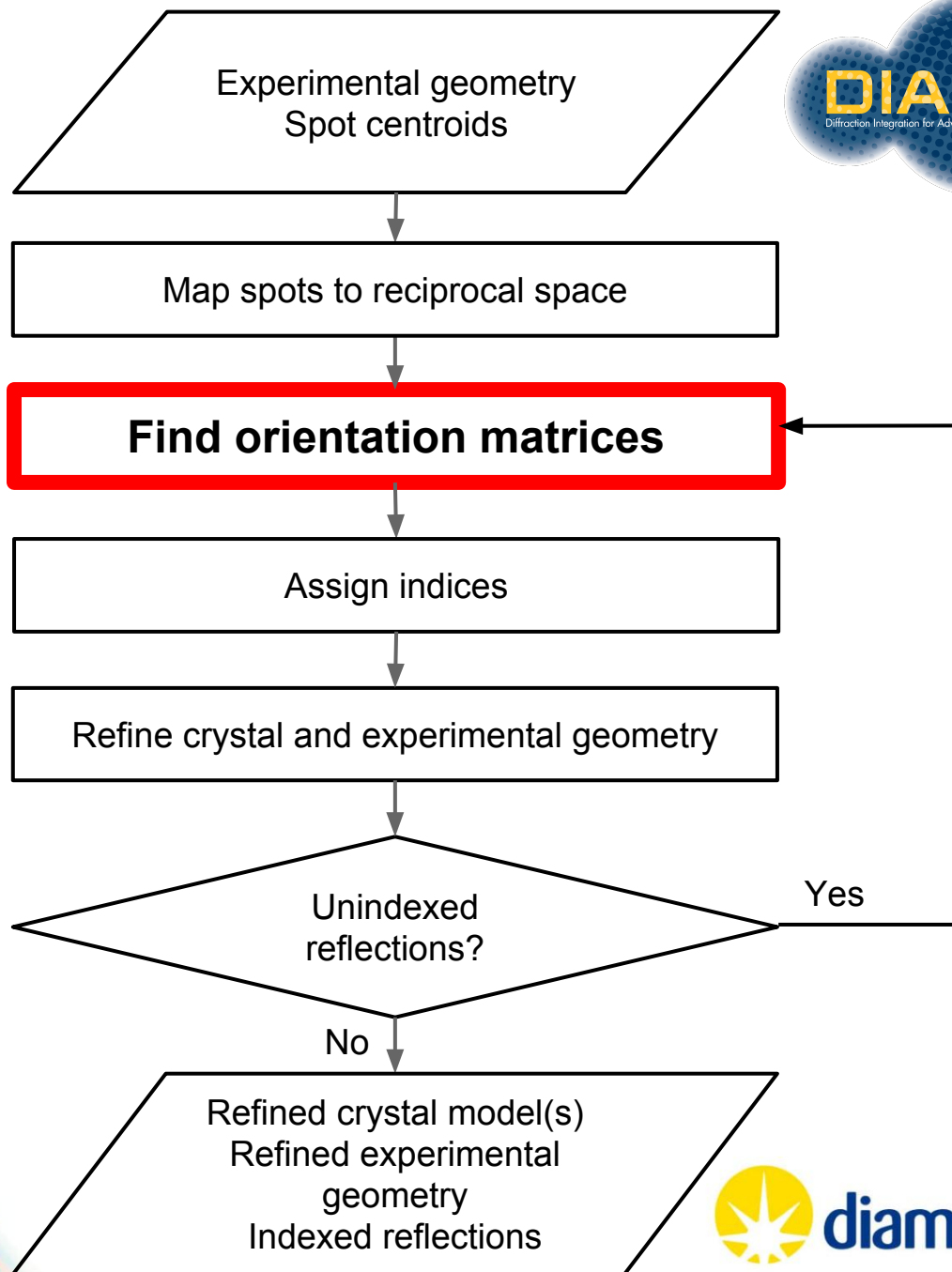
```
A = UB:   {{ 0.0060, -0.0045, -0.0060},  
           { 0.0154,  0.0068,  0.0015},  
           { 0.0051, -0.0153,  0.0024}}
```

```
Saving refined experiments to experiments.json
```

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

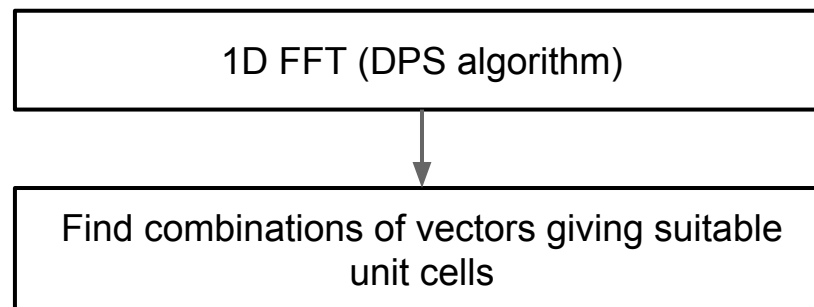
- Choice of 1D & 3D FFT methods or new real space grid search algorithm
- Optionally provide known unit cell and space group

dials.index



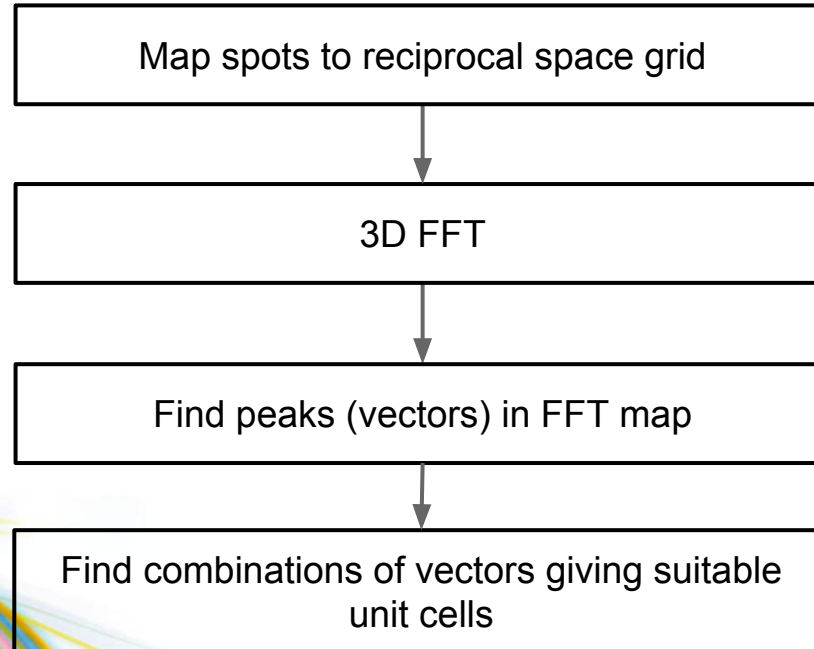
1D FFT (DPS) algorithm

- Use cctbx implementation of DPS algorithm (as used by LABELIT) to populate list of candidate basis vectors (credit: NKS)



3D FFT algorithm

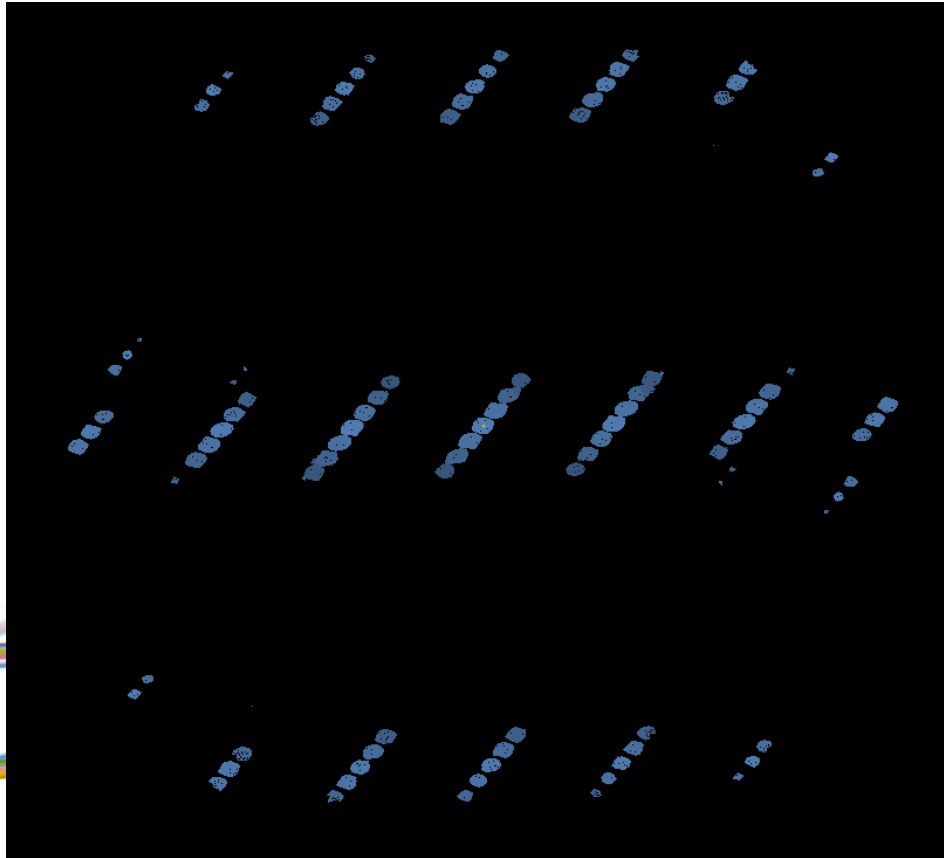
- New implementation of 3D FFT algorithm



Bricogne, G. (1986). *Proceedings of the EEC Cooperative Workshop on Position-Sensitive Detector Software (Phase III)*, p. 28. Paris: LURE.

Campbell, J. W. (1998). *J. Appl. Cryst.* 31, 407-413.

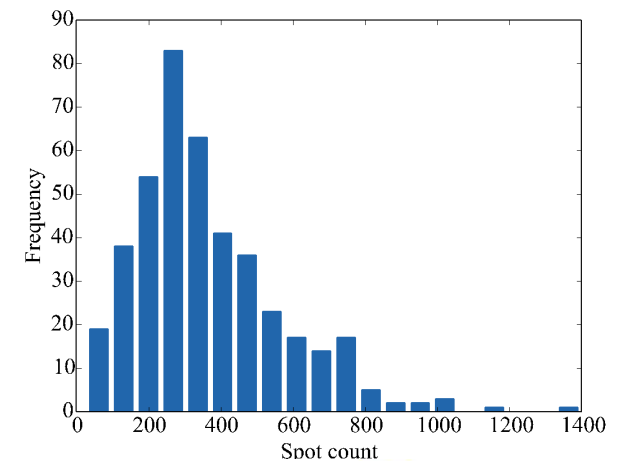
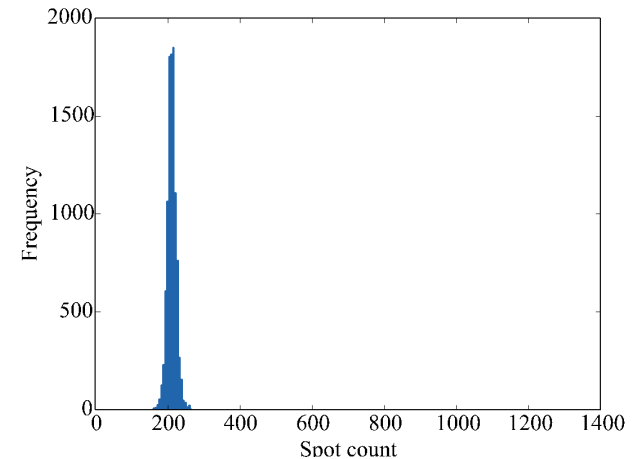
Peak-finding in the 3D FFT map



- Set all grid points below certain rmsd level (e.g. 15 sigma) to zero
- Identify separate 'blobs' in the map using flood-fill algorithm
- Peak positions are the centres of mass of each 'blob'

Aside: multiple lattices

- Polyhedra crystals
 - $\sim 1\ \mu\text{m}$, unit cell $\approx 100\ \text{\AA}$
- Data collection at DLS I24
 - Grid-mounted samples
 - Grid scans to determine best place to collect data
 - $4 \times 4\ \mu\text{m}$ beam (DLS I24)
 - Narrow wedges (1°)
- Data processing troublesome
 - Weak data
 - Multiple lattices
 - Hundreds of datasets



Real space grid search



- In many cases the unit cell is known - why not make use of this information?
- 1D FFT (DPS) and 3D FFT algorithms try to determine the magnitude and direction of the basis vectors simultaneously
- If the unit cell is already known, then we only need to determine the direction of the basis vectors



Real space grid search

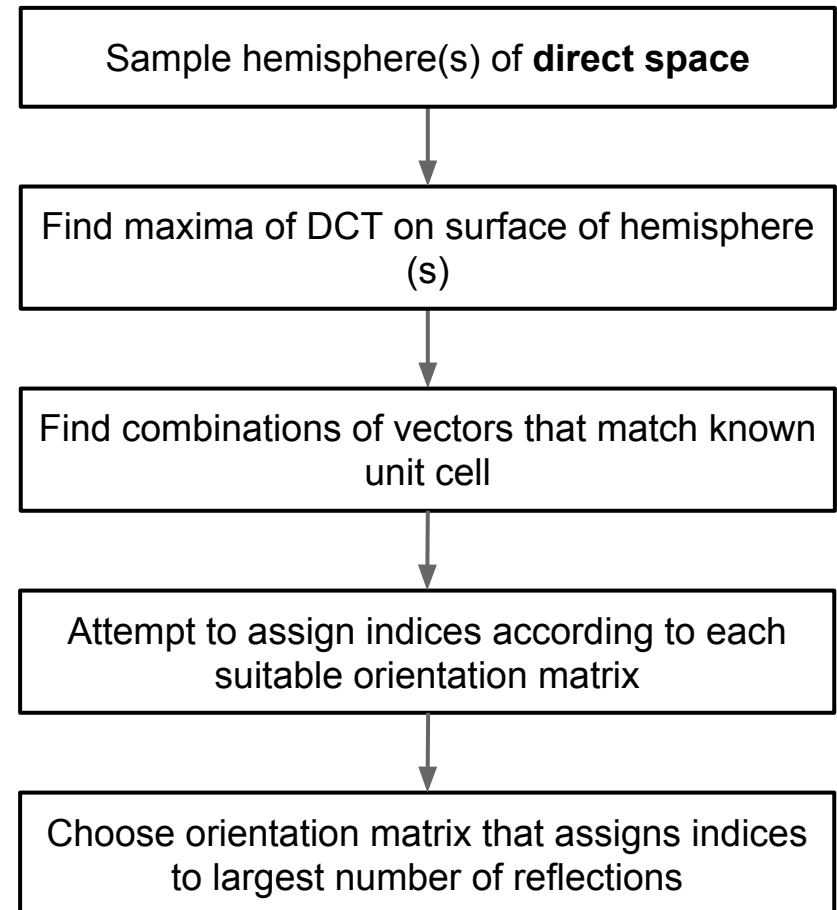
The maxima of the discrete cosine transform (DCT)

$$\sum_i \cos(2\pi \mathbf{S}_i \cdot \mathbf{a}_0)$$

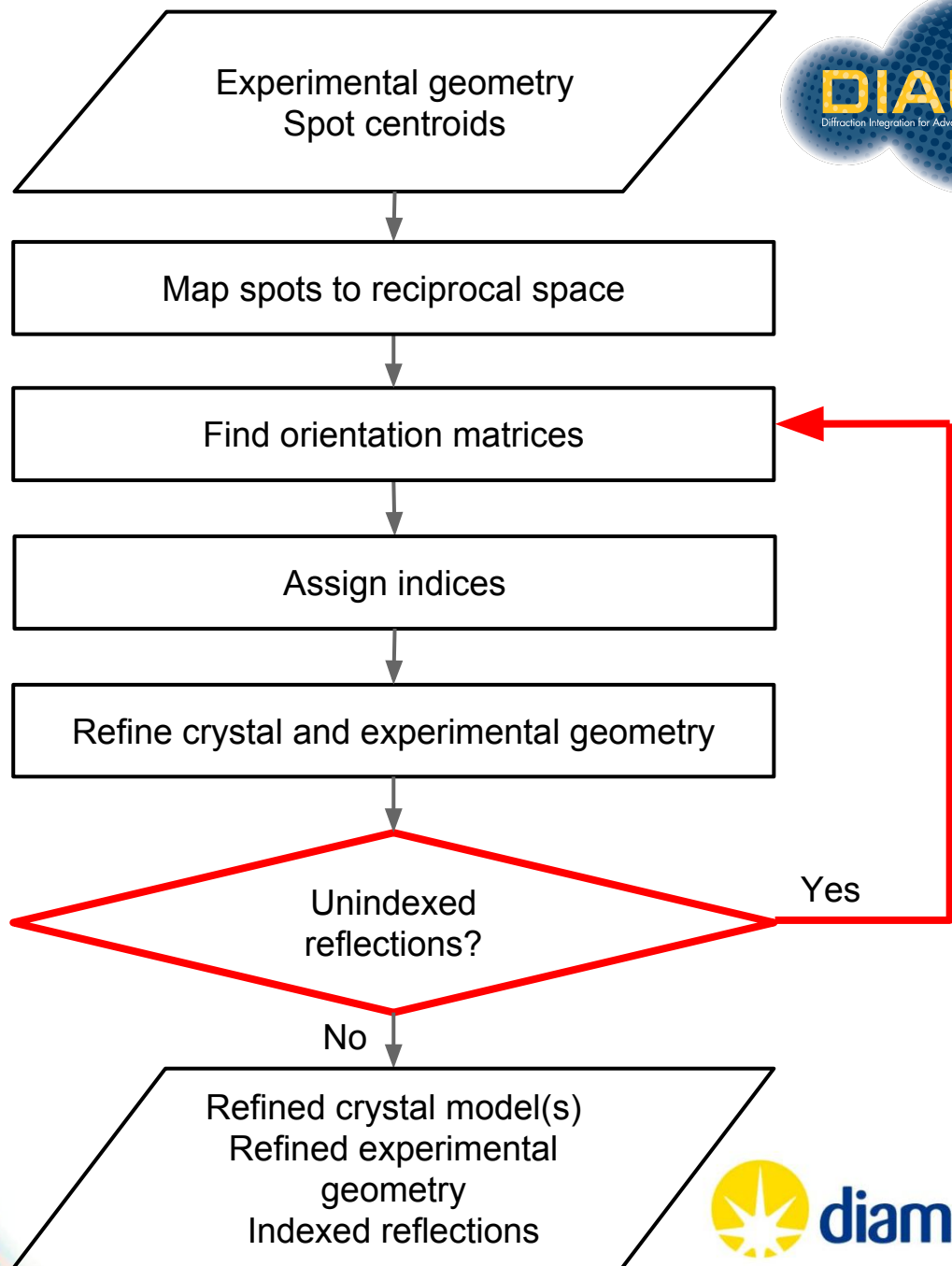
are the integer solutions to the equation

$$\mathbf{S} \cdot \begin{pmatrix} \mathbf{a}_0 \\ \mathbf{b}_0 \\ \mathbf{c}_0 \end{pmatrix} = \mathbf{h}$$

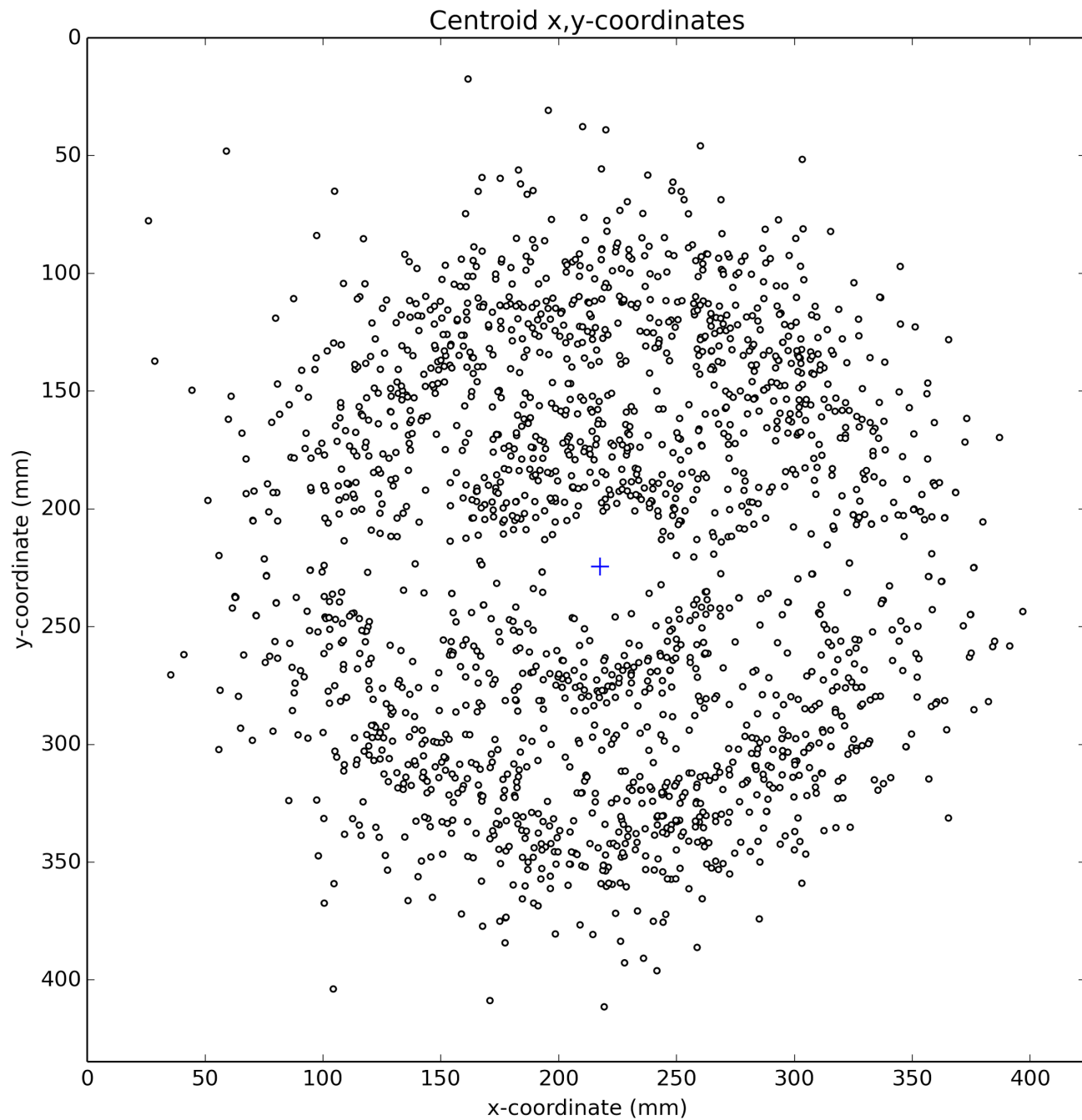
Perform real space grid search of function by varying direction of $\mathbf{a}_0$ only - equivalent to searching surface of a sphere

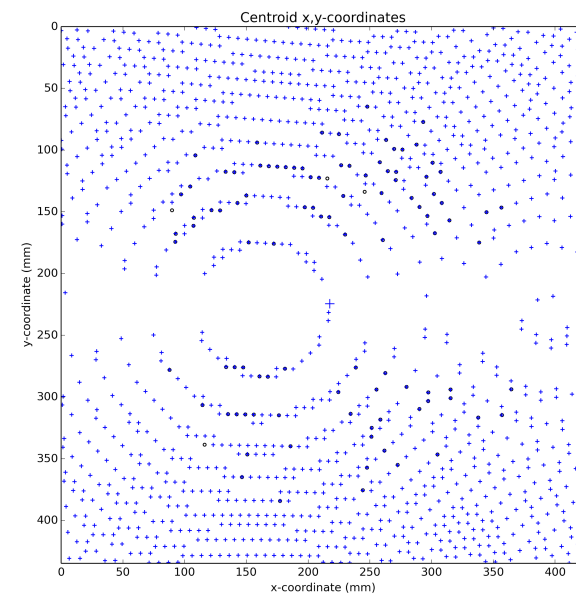
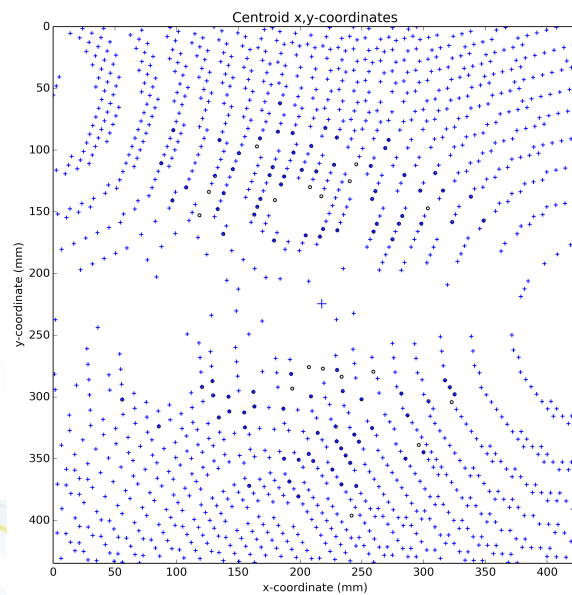
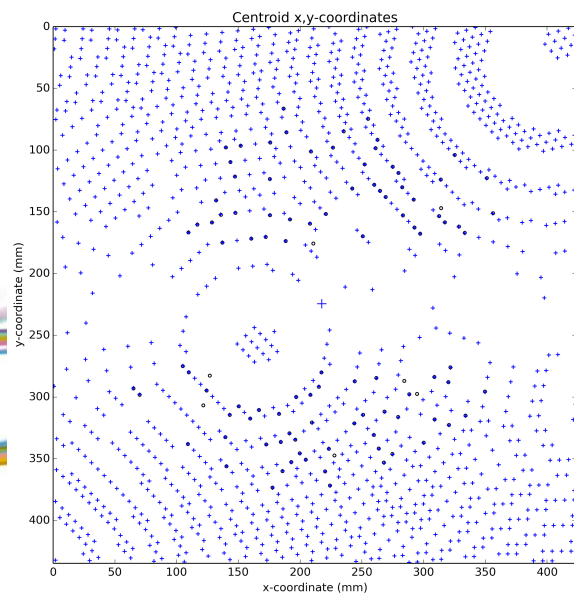
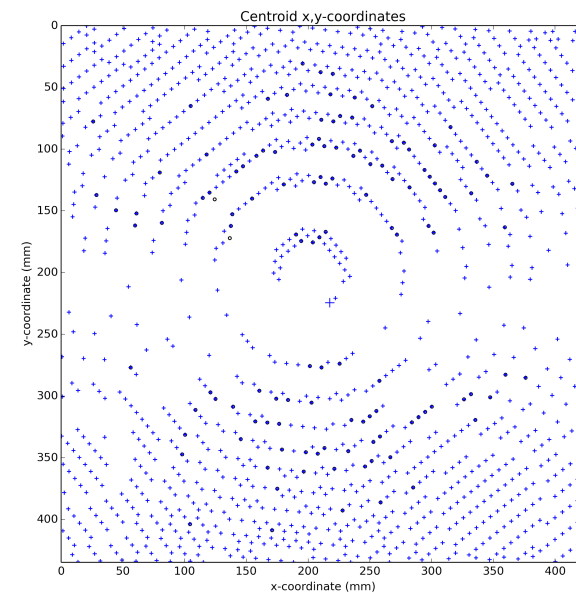
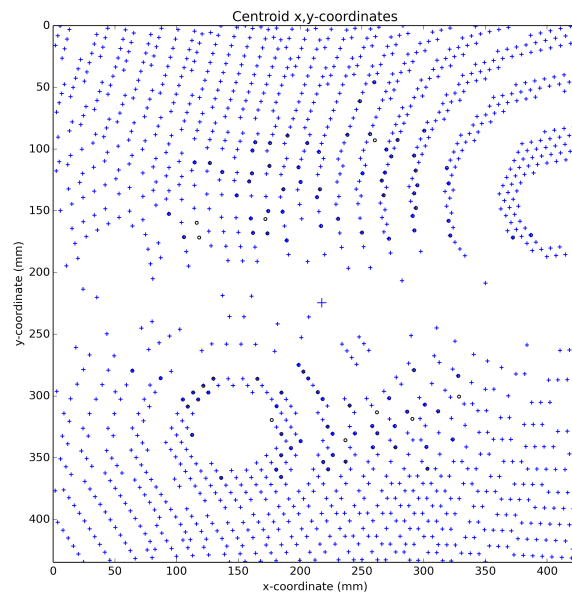
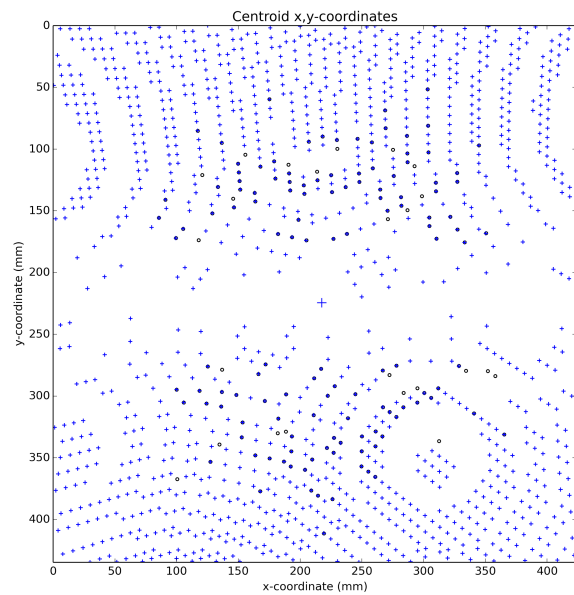


dials.index



- 1° wedge of data
- 1858 spots

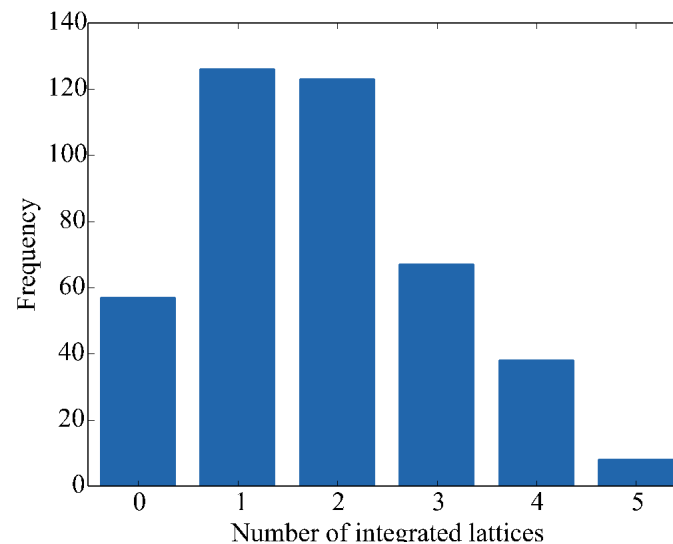
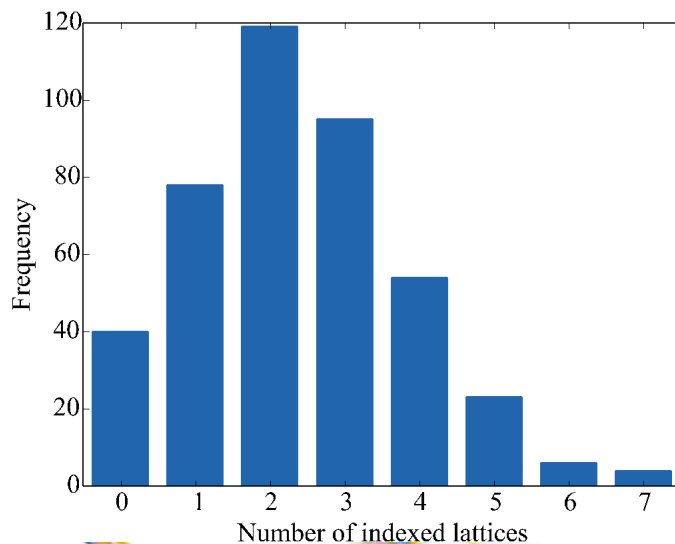




6 lattices identified

Multiple lattices: results

- From 420 sweeps
 - nearly 1000 lattices identified by dials.index
 - 768 successfully integrated with XDS



Choosing the “best” solution



- Given an initial list of candidate basis vectors wish to find the combination of 3 vectors that best accounts for the observed spot centroids
 - Robust against noise (e.g. split crystals, ice rings, multiple lattices)
- Simple version: choose the solution that indexes the most reflections:
 - Works well most of the time, but not always, and can favour larger supercells as they are guaranteed to ‘index’ more reflections.



Choosing the “best” solution



- For each potential solution perform quick refinement cycles
 - More reliable estimates of rmsds
- A sequence of filtering steps based on various heuristics
 - Prefer solutions that give a better fit to the indexed reflections
 - Definition of “model likelihood”: $1 - \text{rmsd}(xy)$, units: mm
 - Filter out solutions whose likelihood is $< 80\%$ of the best likelihood
 - Prefer solutions with a smaller unit cell
 - Filter out unit cells whose volume is $> 125\%$ of the minimum cell volume
 - Prefer solutions that account for more of the diffracted spots
 - Filter out unit cells that account for $< 90\%$ of the number of spots as the solution that indexes the most spots
 - Final choice is the solution with the highest model likelihood from those remaining.

dials.refine_bravais_settings



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

Solution	Metric	fit	rmsd	min/max	cc	#spots	lattice	unit_cell			volume	cb_op
9	0.0250	0.073	0.787/0.848	4049	tP	57.78	57.78 149.99	90.00	90.00	90.00	500681	a,b,c
8	0.0250	0.072	0.787/0.970	4049	oC	81.71	81.73 150.00	90.00	90.00	90.00	1001813	a-b,a+b,c
7	0.0133	0.071	0.787/0.899	4049	oP	57.78	57.76 149.98	90.00	90.00	90.00	500537	a,b,c
6	0.0217	0.071	0.970/0.970	4049	mC	81.72	81.74 150.02	90.00	89.99	90.00	1002178	a-b,a+b,c
5	0.0250	0.072	0.795/0.795	4049	mC	81.73	81.71 150.01	90.00	89.99	90.00	1001809	a+b,-a+b,c
4	0.0131	0.070	0.807/0.807	4049	mP	57.76	57.79 149.99	90.00	90.01	90.00	500676	-b,-a,-c
3	0.0133	0.070	0.899/0.899	4049	mP	57.79	57.77 150.00	90.00	89.99	90.00	500732	a,b,c
2	0.0125	0.071	0.787/0.787	4049	mP	57.77	149.99 57.79	90.00	89.99	90.00	500744	b,c,a
1	0.0000	0.070	-/-	4049	aP	57.80	57.77 150.01	90.01	89.99	89.99	500927	a,b,c

```
$ ls bravais_setting_*.json
```

```
bravais_setting_1.json bravais_setting_2.json bravais_setting_3.json bravais_setting_4.json bravais_setting_5.json  
bravais_setting_6.json bravais_setting_7.json bravais_setting_8.json bravais_setting_9.json
```



dials.refine_bravais_settings



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

Solution	Metric	fit	rmsd	min/max	cc	#spots	lattice	unit_cell					
9	0.9632	0.653	0.502/0.639			9000	tI	78.95	78.95	67.95	90.00	90.00	90.00
8	0.9632	0.663	0.502/0.639			9000	oF	67.98	111.81	111.60	90.00	90.00	90.00
7	0.9632	0.593	0.502/0.502			9000	mI	65.42	112.23	66.09	90.00	117.38	90.00
6	0.9632	0.523	0.508/0.508			9000	mI	64.71	111.12	65.28	90.00	117.34	90.00
5	0.9471	0.554	0.639/0.856			9000	oI	67.66	78.09	79.05	90.00	90.00	90.00
4	0.9471	0.566	0.639/0.639			9000	mI	78.13	67.68	79.09	90.00	89.90	90.00
3	0.9471	0.551	0.712/0.712			9000	mI	67.60	78.02	78.99	90.00	89.94	90.00
2	0.0000	0.101	0.856/0.856			9000	mI	67.61	78.98	78.03	90.00	90.96	90.00
1	0.0000	0.100		-/-		9000	aP	64.63	64.64	67.59	59.14	59.17	75.29



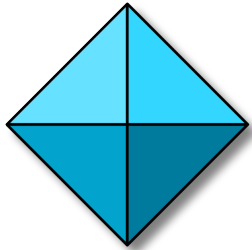
Acknowledgements

- DIALS East
 - Gwyndaf Evans, Graeme Winter, David Waterman, James Parkhurst, Luis Fuentes-Montero, Markus Gerstel
- DIALS West
 - Nick Sauter, Aaron Brewster, Tara Michels-Clark, Iris Young
- MRC-LMB
 - Phil Evans, Andrew Leslie, Garib Murshudov, Harry Powell
- Users and Diamond staff
- Program authors for publishing algorithms



diamond

Research Complex
at Harwell



CCP4



Science & Technology
Facilities Council



BioStruct_x

The End

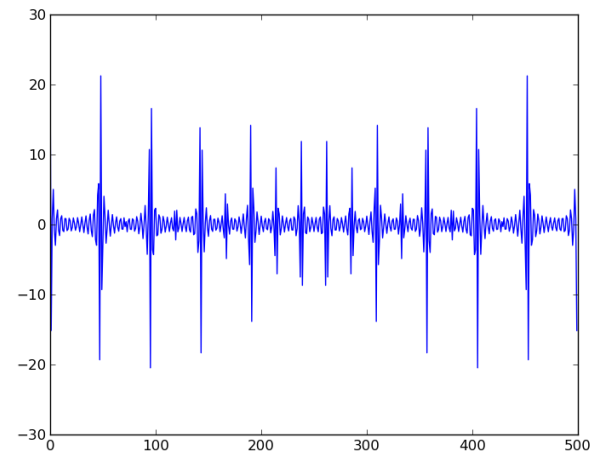
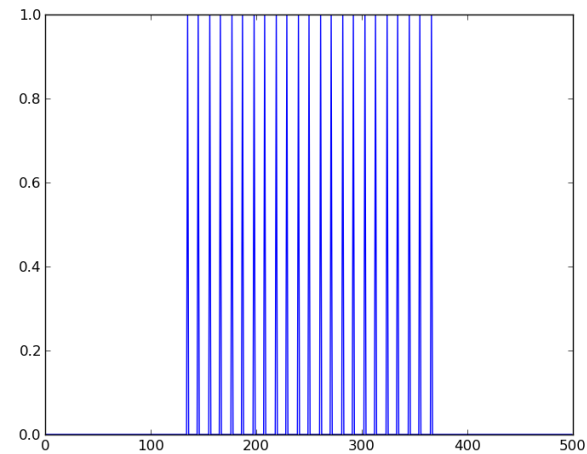
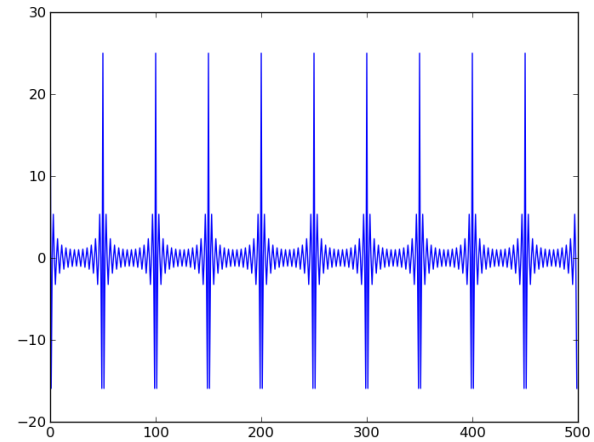
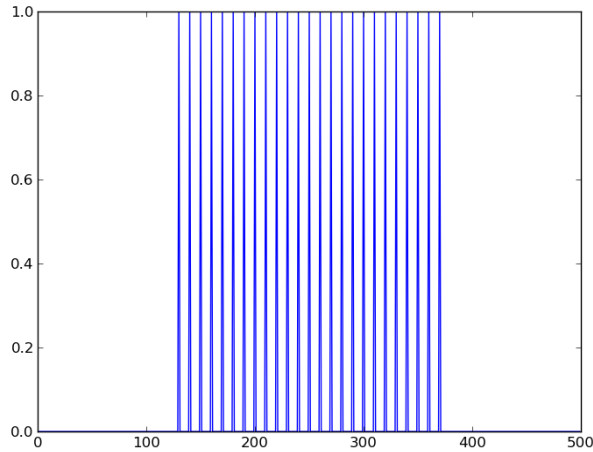


Maximum cell determination

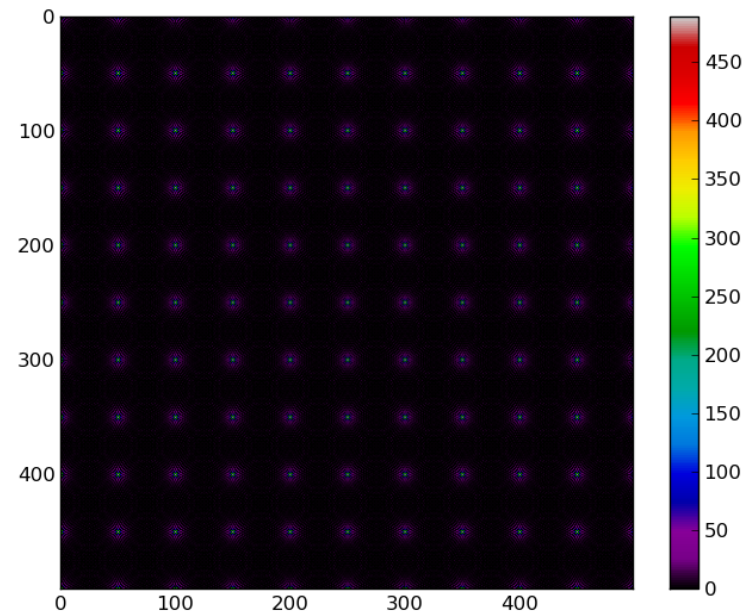
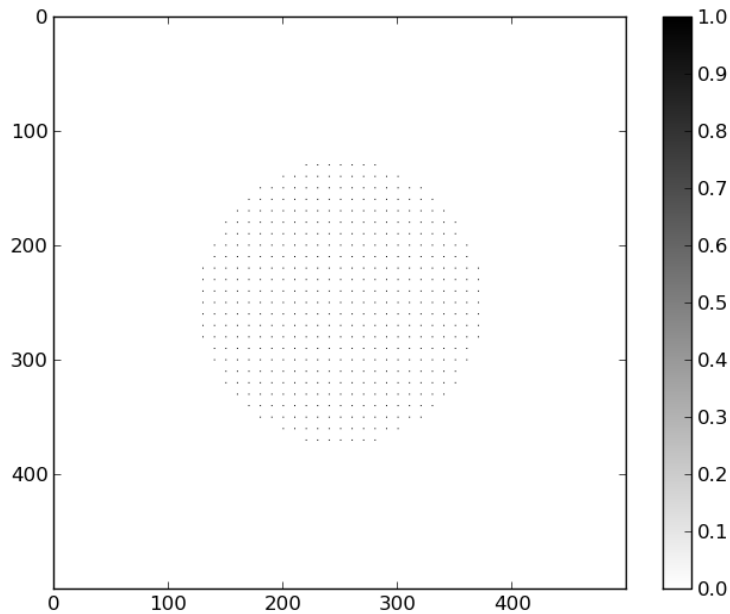


- Important for fft1d and fft3d
- Determines d_min for inclusion of reflections in initial indexing
 - Particularly important for fft3d:
 - default grid size (n_points) = 256 x 256 x 256
 - fft_cell should be greater than twice max_cell, so:
 - $\text{fft_cell} = \text{n_points} * \text{d_min}/2$
 - $\text{fft_cell} = 2.5 * \text{max_cell}$
 - $\text{d_min} = 5 * \text{max_cell}/\text{n_points}$
 - n_points = 256
 - max_cell = 50, d_min = 0.98
 - max_cell = 100, d_min = 1.95
 - max_cell = 400, d_min=7.81
 - n_points = 512
 - max_cell = 400, d_min=3.91

Wrap-around effect: 1D FFT



Wrap-around effect: 2D FFT



Wrap-around effect: 2D FFT

