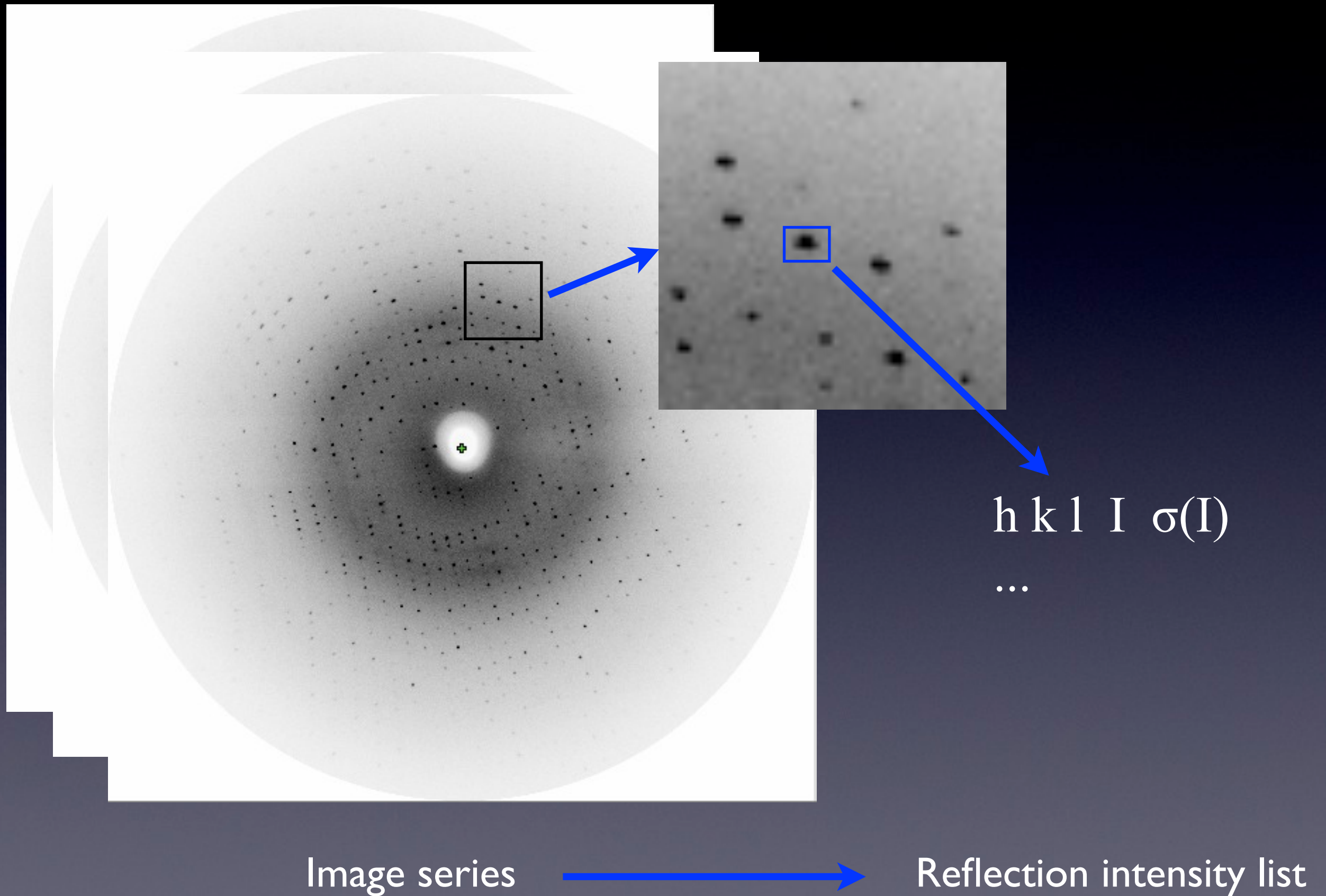


Diffraction geometry and integration of diffraction images

Phil Evans Tsukuba November 2014

*MRC Laboratory of Molecular Biology
Cambridge UK*

Integration



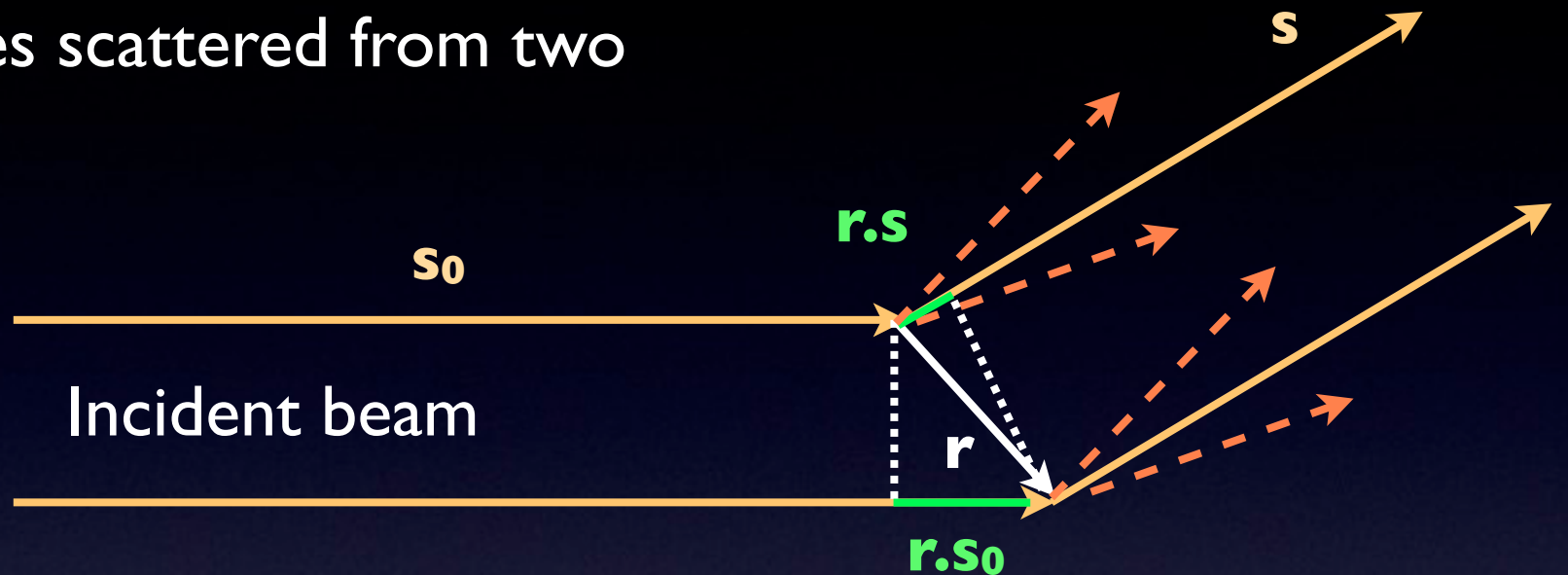
Diffraction Geometry

- Diffraction from a crystal - Laue equations
- Reciprocal lattice
- Ewald construction
- Data collection strategy

Scattering from two electrons

Path length difference for waves scattered from two points $\mathbf{r}$ apart in direction $\mathbf{s}$

$$\delta L = \mathbf{r} \cdot \mathbf{s} - \mathbf{r} \cdot \mathbf{s}_0 = \mathbf{r} \cdot (\mathbf{s} - \mathbf{s}_0)$$



Phase shift corresponding to path length difference δL for a wave with wavelength λ

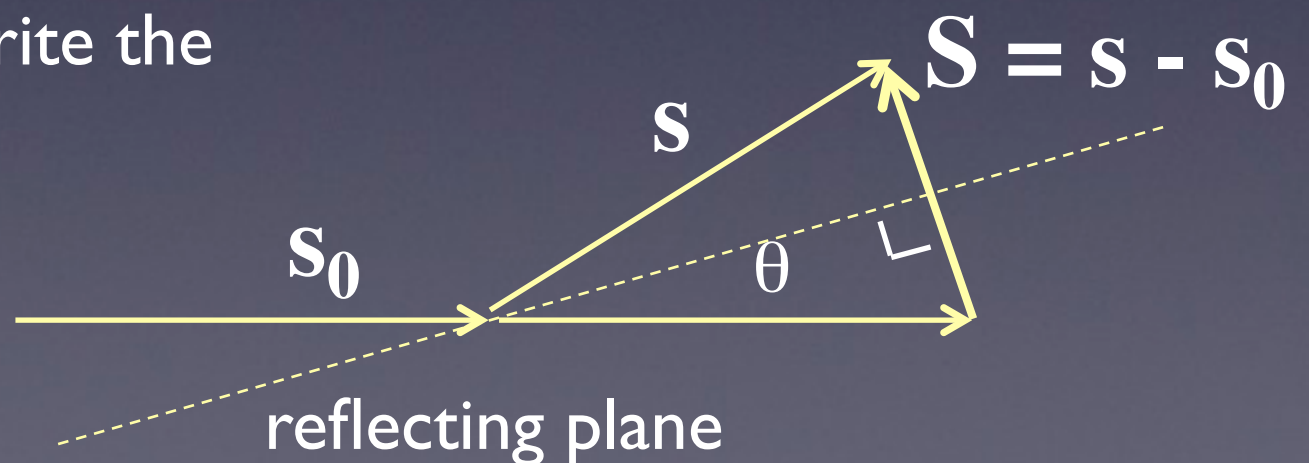
$$= 2\pi (\text{path difference}) / \lambda$$

$$= 2\pi \delta L / \lambda = 2\pi \mathbf{r} \cdot (\mathbf{s} - \mathbf{s}_0) / \lambda$$

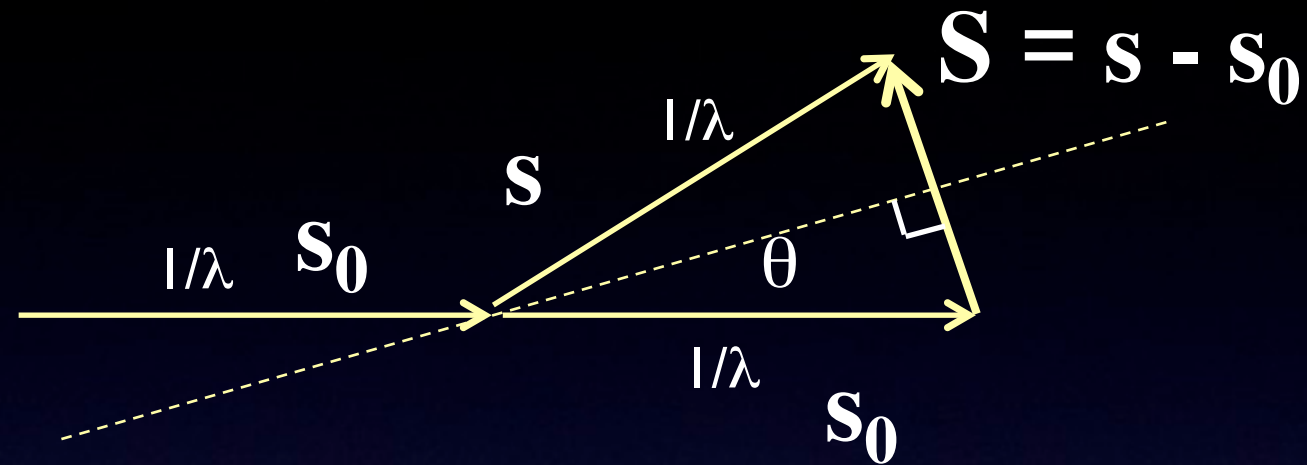
If we make the length of the wave vectors $\mathbf{s}_0$ and $\mathbf{s} = 1/\lambda$, ie $|\mathbf{s}_0| = |\mathbf{s}| = 1/\lambda$, then we can write the phase shift $= 2\pi \mathbf{r} \cdot \mathbf{S}$ where $\mathbf{S} = \mathbf{s} - \mathbf{s}_0$

$\mathbf{S}$ is the perpendicular to an imaginary “reflecting plane” with $|\mathbf{S}| = 2 \sin \theta / \lambda$

X-rays scattered in all directions



The Ewald Sphere Construction



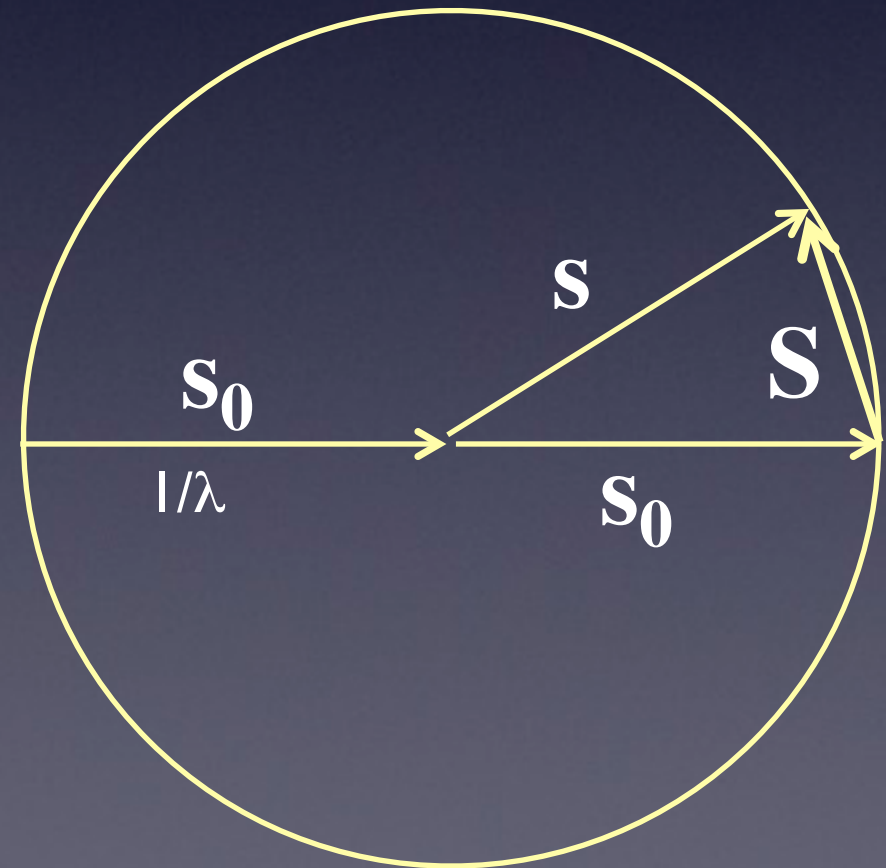
The general condition for diffraction is illustrated by the vector equation

$$S = s - s_0$$

Because s_0 and s have the same length ($1/\lambda$), we can generalise this diagram by drawing a sphere of radius $|s_0| = |s| = 1/\lambda$

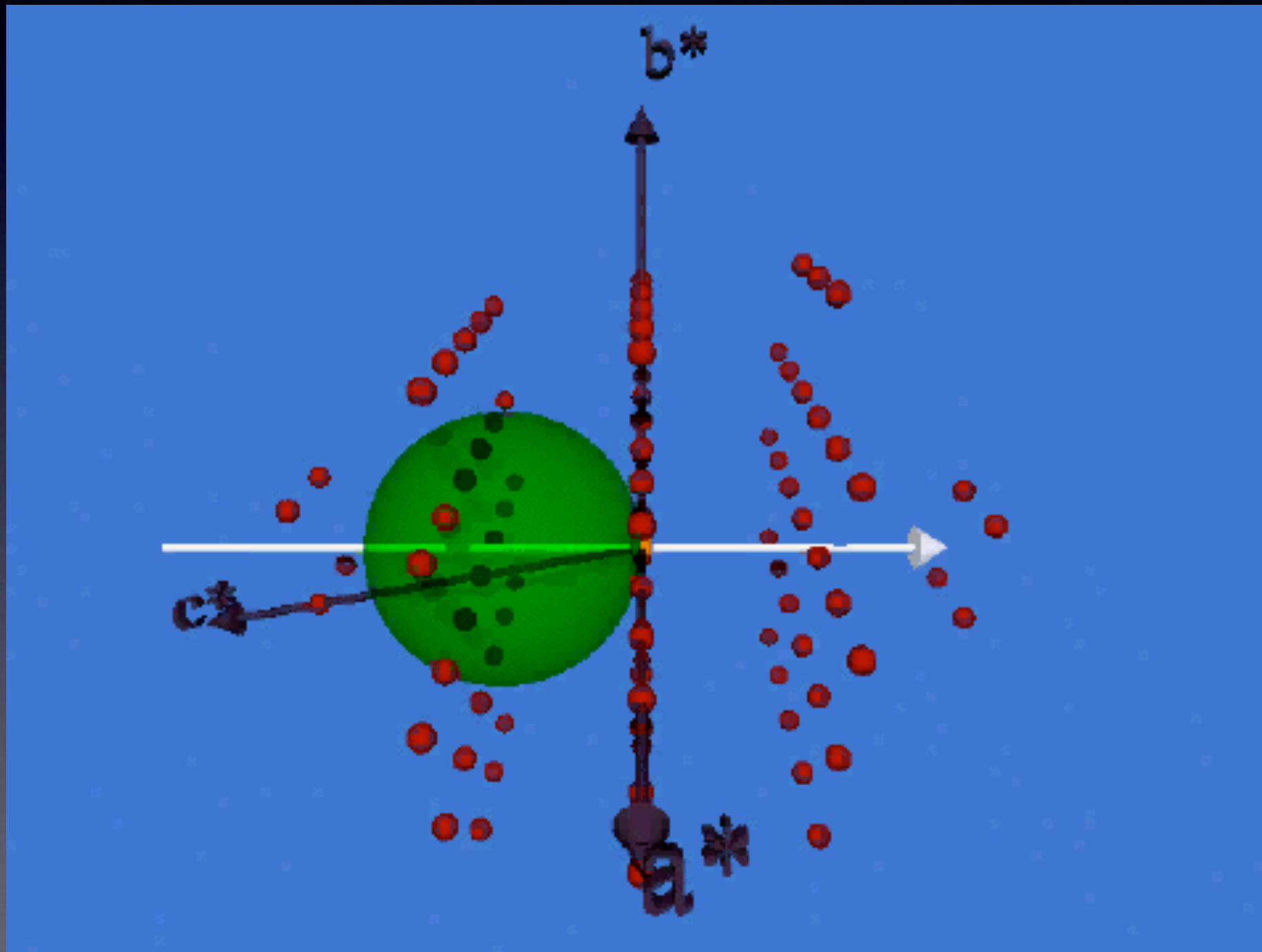
S is the diffraction vector in *reciprocal space*

For a crystal, S may only take certain values,
 $S = h a^* + k b^* + l c^*$

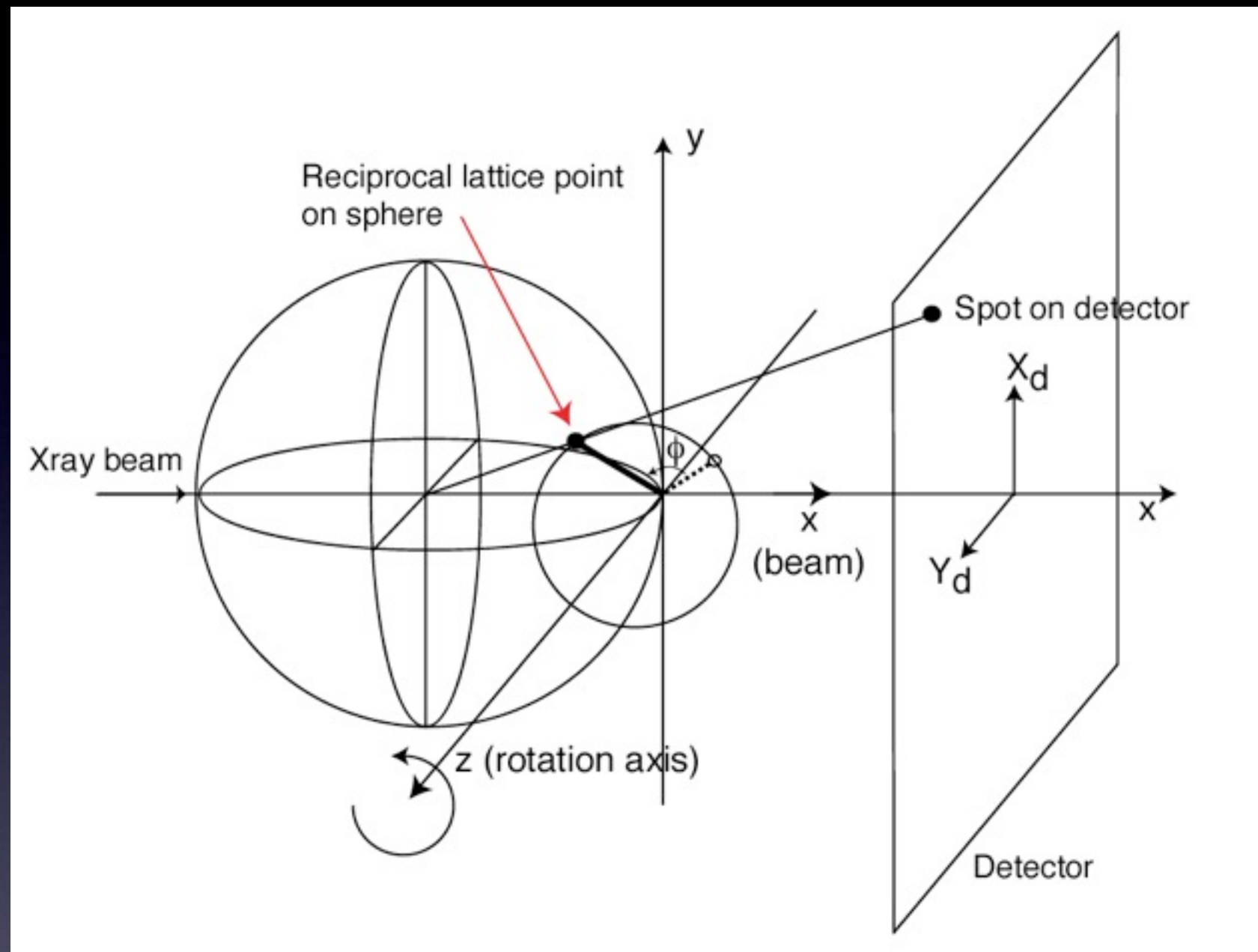


As the crystal rotates, so does the the reciprocal lattice

As a reciprocal lattice point passes through the Ewald sphere, a diffracted beam is observed along the line from the sphere centre to the reciprocal lattice point



The part of the reciprocal lattice which intersects the sphere is projected on to the detector



As the crystal rotates, each lattice point in turn passes through the sphere, and a spot is recorded on the detector

We can use the Ewald construction to understand

- what diffraction images look like
- how to collect a complete dataset without missing bits

We can imagine the reciprocal lattice sitting on the crystal on the camera, and rotating as the crystal rotates

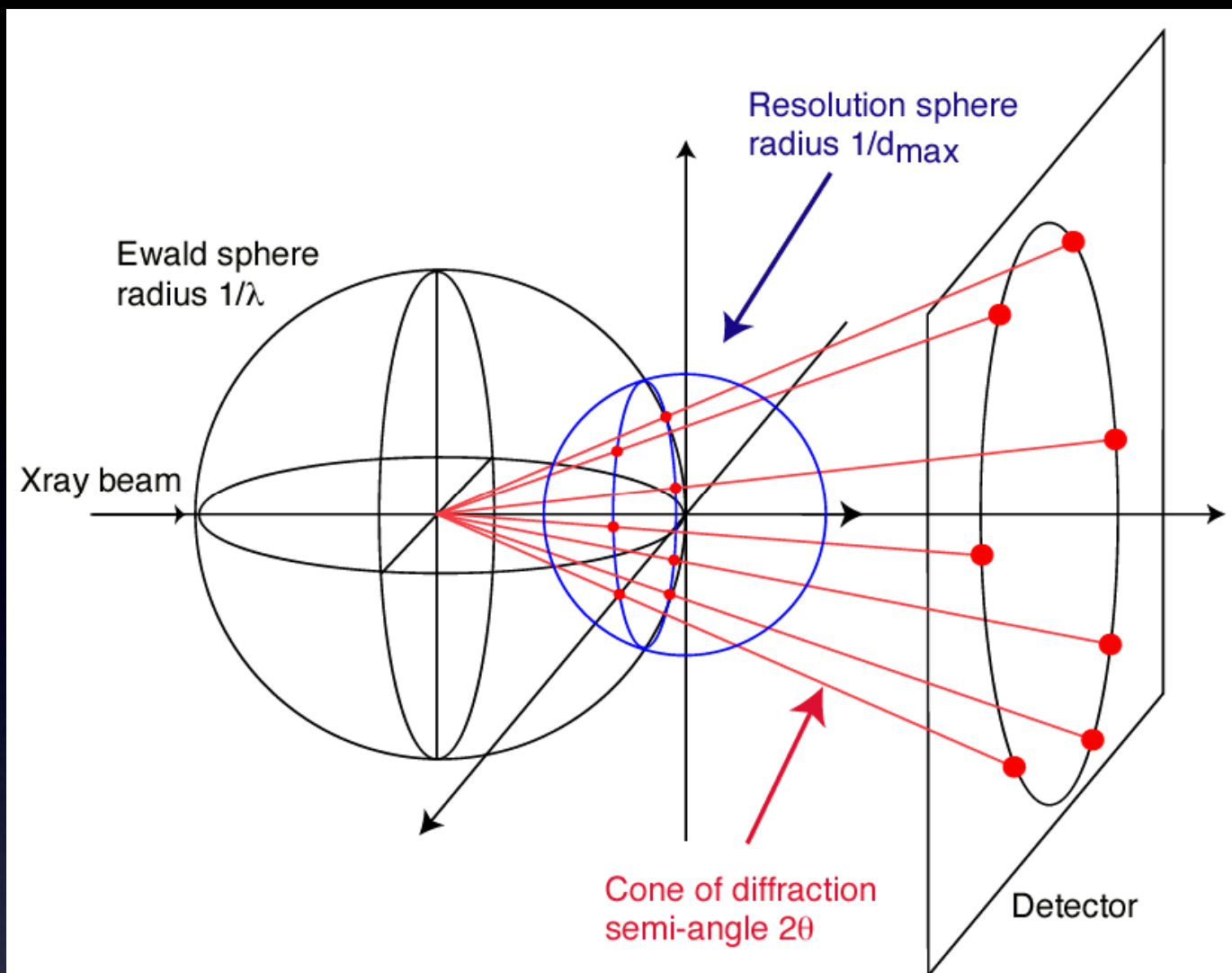
More about data collection strategy tomorrow

Detector position

For a maximum resolution of $d_{\max}$, all diffraction vectors S must lie within a resolution sphere of radius $1/d_{\max}$

As the crystal rotates, the diffracted beams all lie within a cone of semi-angle $2\theta_{\max}$

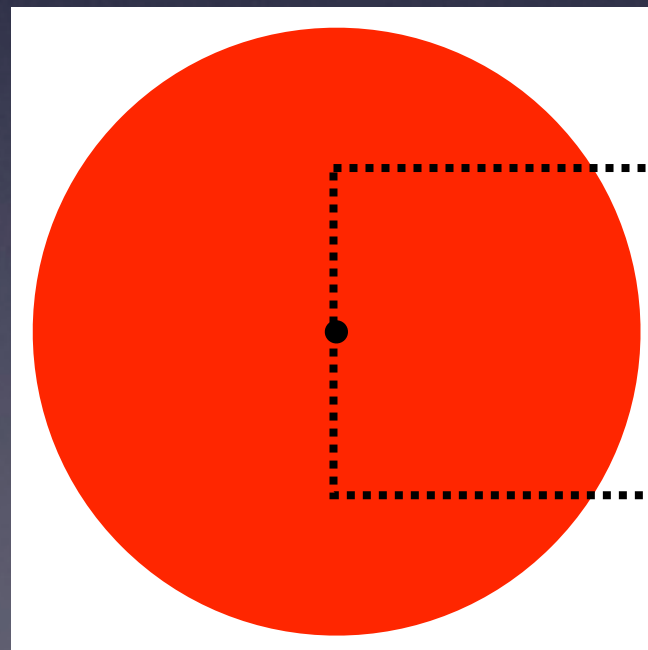
$$\lambda/d_{\max} = 2 \sin \theta_{\max}$$



A detector centered on the beam collects the whole cone

This gives optimum efficiency and simple strategy

The corners of a square detector collect incomplete data

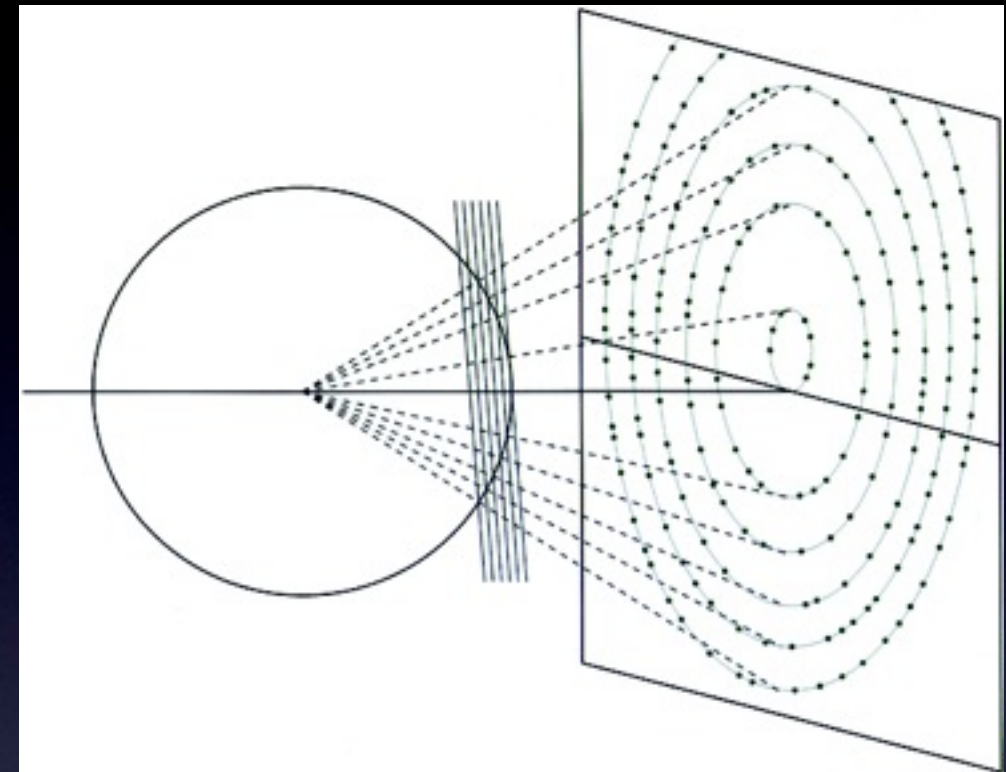


For long axes (close spot separation) it may be necessary to use a long detector distance and an offset detector

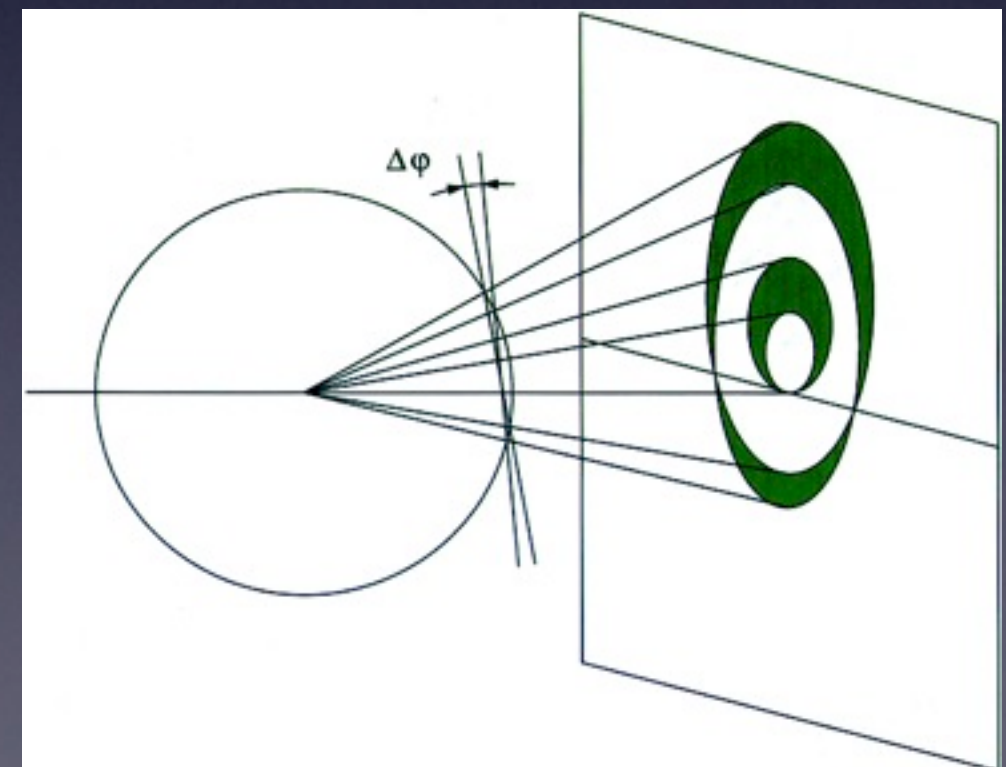
This gives a lower efficiency, and to get complete data requires a complicated strategy

The appearance of diffraction images

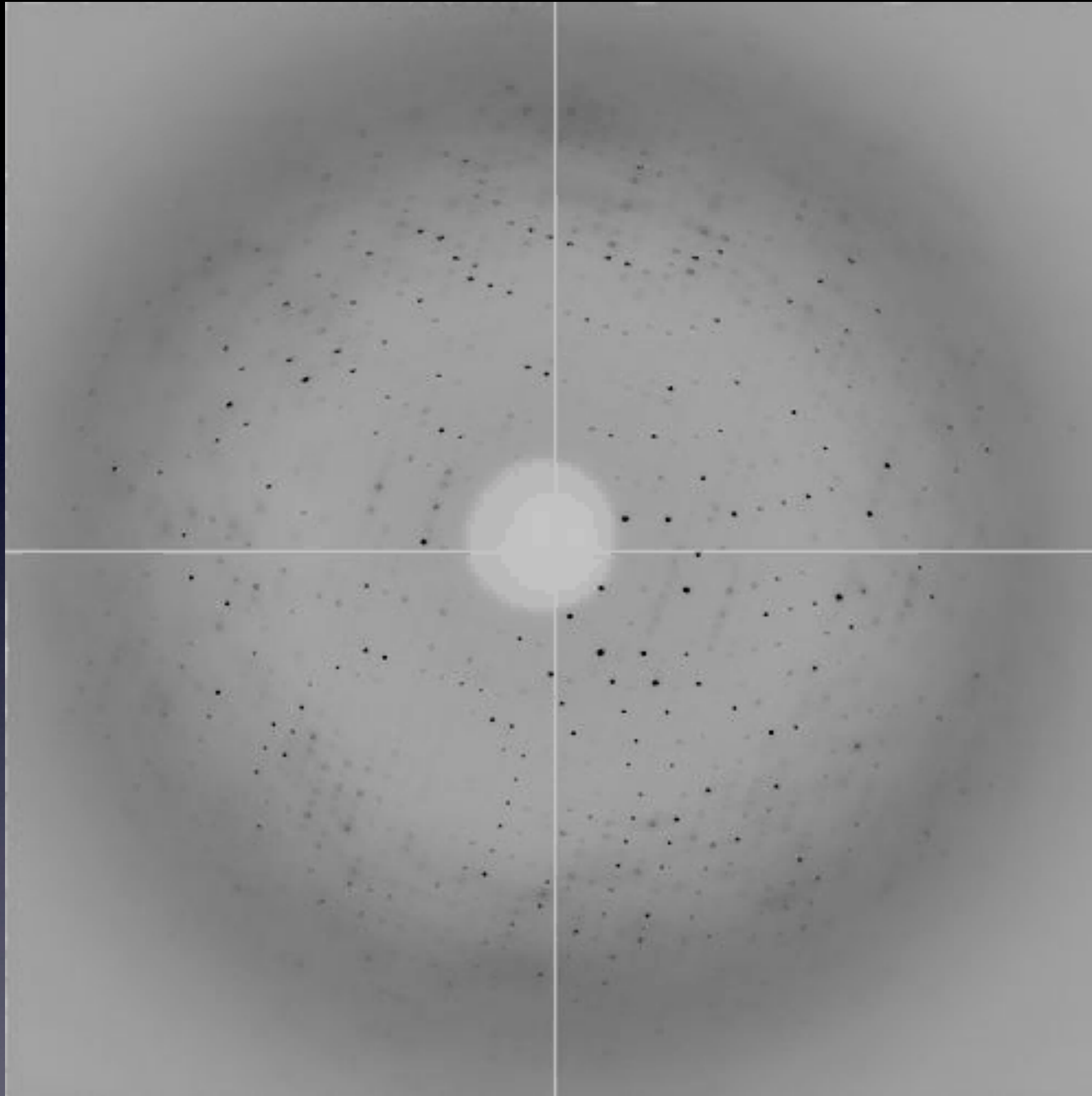
Reciprocal lattice points lie in layers (planes). Each plane intersects the sphere in a circle, and the spots projected on the detector lie in ellipses



If the crystal is rotated through a small angle, each circle is broadened into a *lune*. All the spots in a lune belong to one plane of the reciprocal lattice (not necessarily a principal plane)



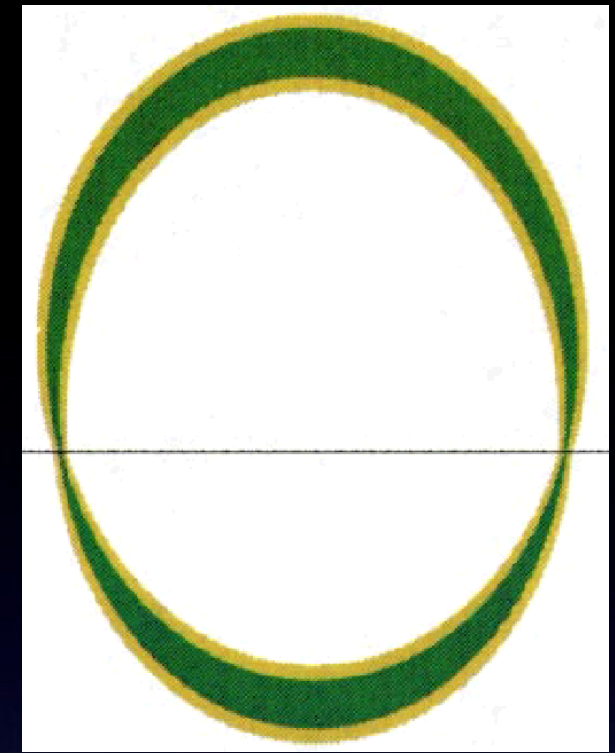
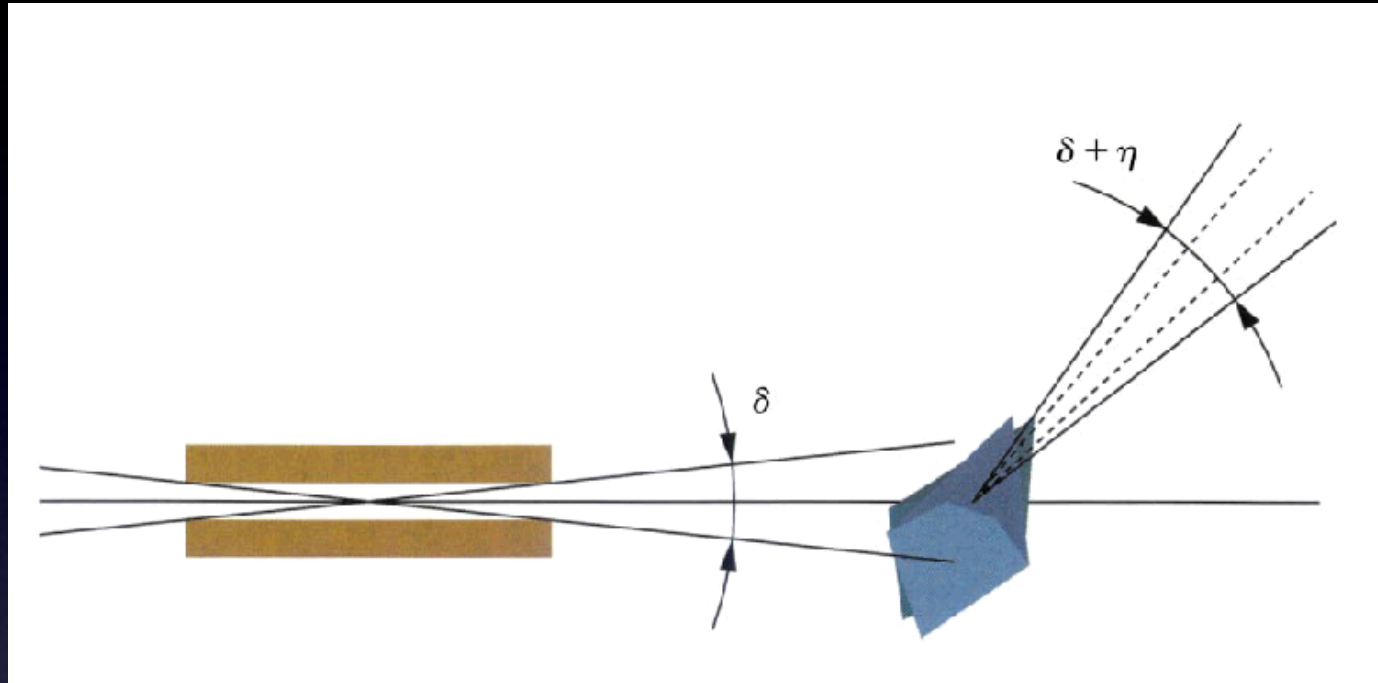
illustrations from Zbyszek Dauter



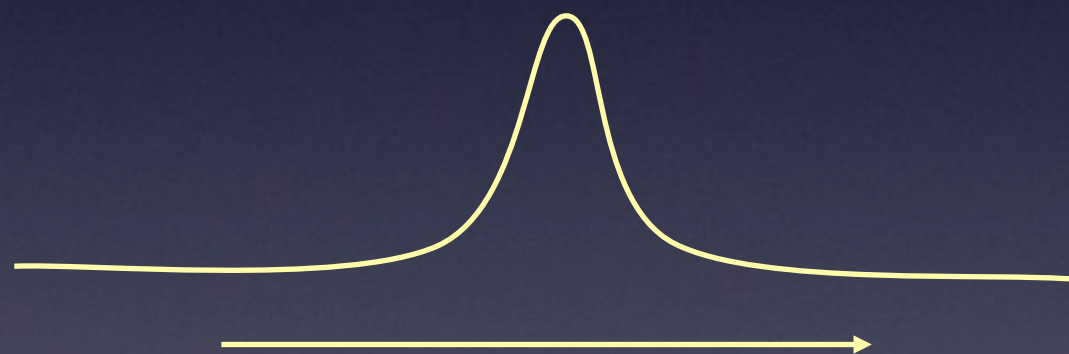
The full diffraction pattern (ie the reciprocal lattice) is 3-dimensional, and we want to measure the whole sphere to the maximum resolution (radius) available.

The dataset should also be complete in dynamic range, including weak & strong spots, ie avoiding too many *overloads*, since the structure of the asymmetric unit is inferred from the measured intensities.

Beam divergence δ and mosaicity η add up to increase the angular width of the diffracted beam



High mosaicity causes broadening of the lunes
Most obvious along the rotation axis

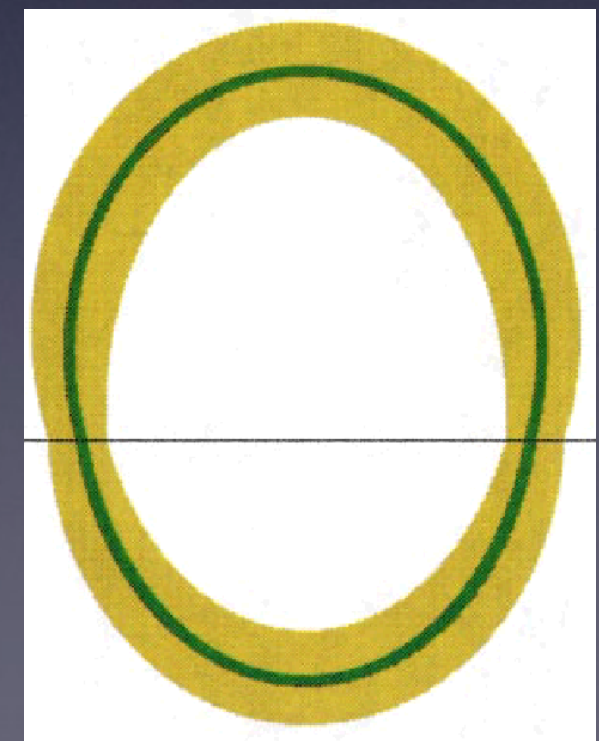


rotation angle ϕ

Reflection width in rotation

$$= \delta + \eta + \text{geometric factor}$$

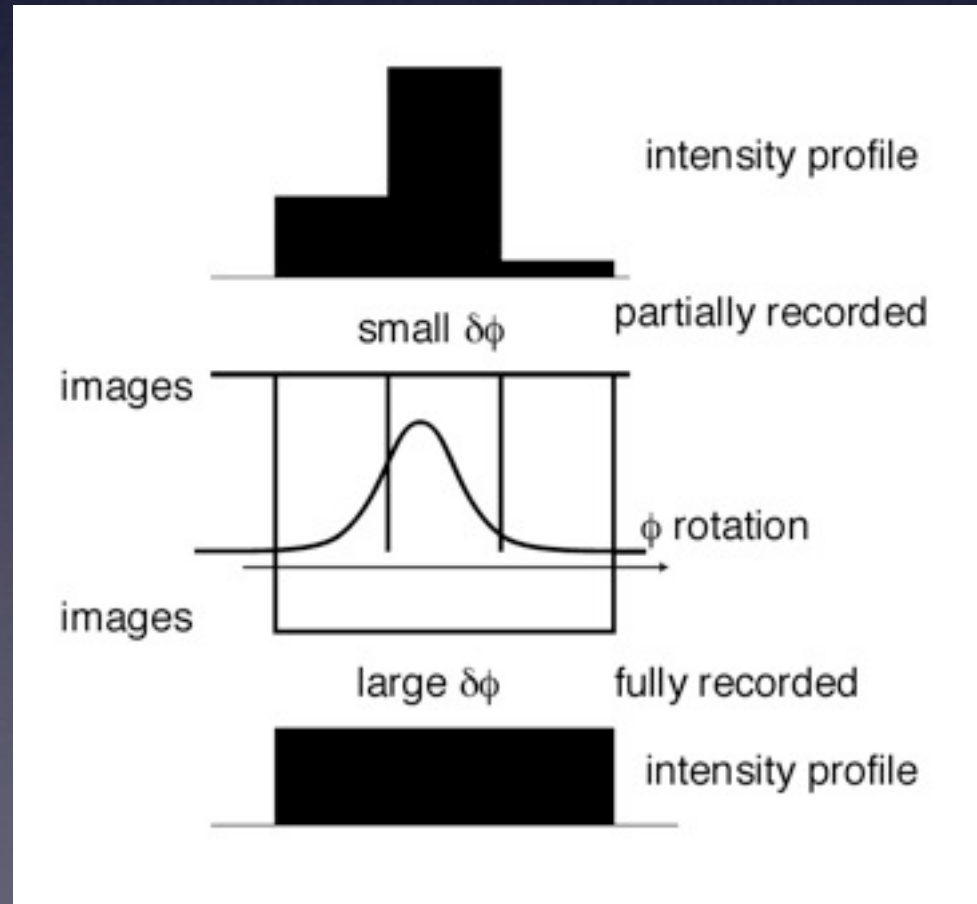
(geometric factor depends on angle between the rotation axis & S)



Images: fully recorded and partially recorded reflections

We want to determine the intensity of a reflection, integrated over its extent in reciprocal space by rotating the crystal so that the extended reciprocal lattice point passes through the sphere.

CCD and image plate detectors take a significant time to read out, so for these have to close the shutter & stop the rotation (simultaneously!). Pixel detectors (eg Pilatus) with very fast read-out can be used with continuous rotation with the shutter open. In both cases our sampling of the 3-dimensional reciprocal space is in consecutive slices, typically of between about 0.1° and 1°



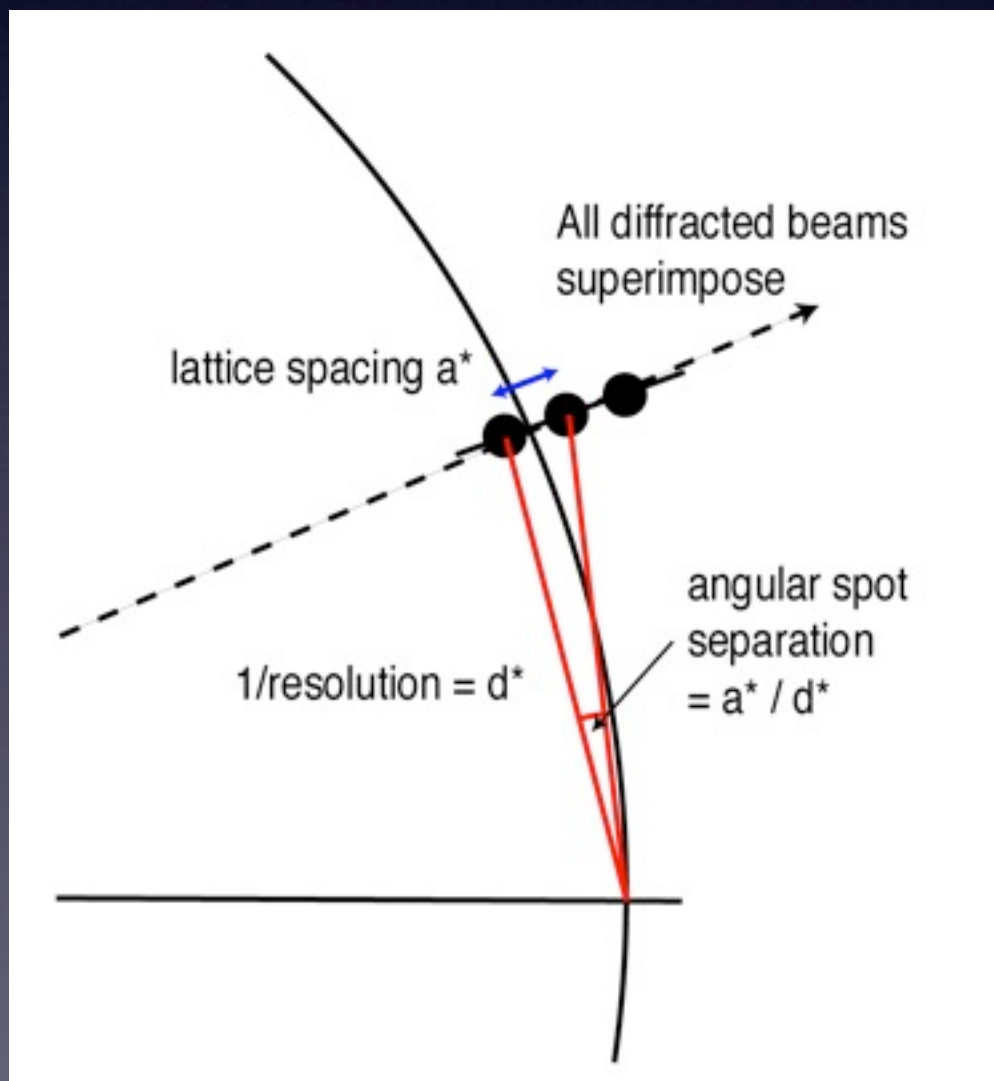
Depending on the slice width and the reflection width a reflection may occur on one image (*full or fully recorded*) or on several (*partial or partially recorded*)

Overlaps and rotation range

Current integration programs assume that spots are resolved, both on the detector and on rotation ϕ . This means that the intensity goes down to background all round the spot

Resolution is a problem for large unit cells, high mosaicity and high resolution

Overlap between spots on the detector is easy to see, but to understand overlap on ϕ we need to look in reciprocal space



When a closely-spaced row of spots (eg along a^*) is moving perpendicularly into the sphere, their diffracted beams almost coincide. The spots are on top of each other on the detector, and are only separated on ϕ

$$\text{Maximum slice width} = (a^*/d^*) - w$$

$$= d/a - w$$

$$w = \text{reflection width} = \delta + \eta$$

$$\text{eg cell} = 200\text{\AA}, \text{ resolution} = 2\text{\AA}, \text{ width} = 0.3^\circ$$

$$\text{Maximum Slice} = 0.27^\circ$$

If possible, orient a long axis along the rotation axis to minimise overlap problems

Integration of diffraction images

Integration

Images $\rightarrow$ hkl I $\sigma(I)$

Two distinct methods:

- 2-D: integrate spots on each image, add together partially recorded observations in the scaling program. MOSFLM, DIALS, DENZO, HKL2000, etc
- 3-D: integrate 3-dimensional box around each spot, from a series of images. XDS, D*TREK, SAINT, DIALS etc

For today: MOSFLM

Starting Point: A series of diffraction images, each recorded on a 2D area detector while rotating the crystal through a small angle (typically 0.2 - 1.0° per image) about a fixed axis (the Rotation/Oscillation Method).

Outcome: A dataset consisting of the indices (h,k,l) of all reflections recorded on the images with an estimate of their intensities and the standard uncertainties of the intensities: $h, k, l, I(hkl), \sigma(I)$

some integration slides from Andrew Leslie

New things in and around CCP4

DIALS - <http://dials.sf.net>

What is it? New data integration software suite including indexing, refinement and integration, for rotation experiments and XFEL snapshots.

Why are we doing this? Modern detectors & experiments pushing software: also want to develop new algorithms. It is time for 21st century software.

Using DIALS: processing steps explicit, find spots, index, refine, integrate, export to MTZ, scale with AIMLESS

Gwyndaf Evans, Graeme Winter, David Waterman, James Parkhurst, Luis Fuentes-Montero – DIAMOND

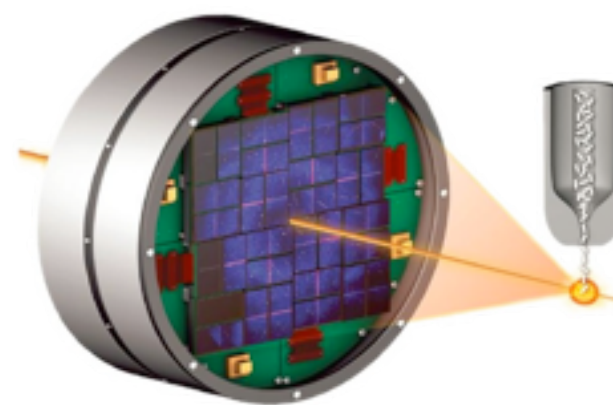
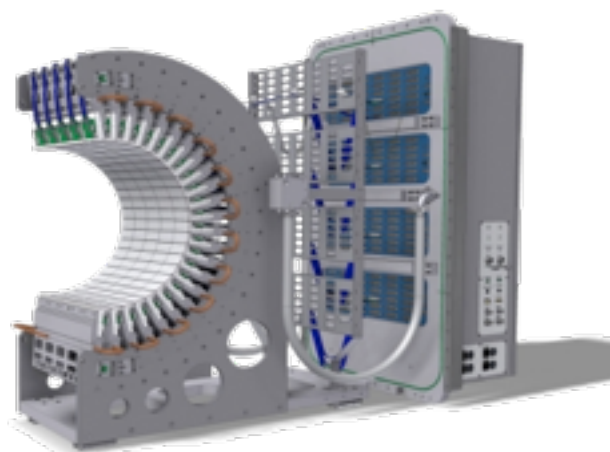
Nicholas Sauter, Aaron Brewster, Johan Hattne – Lawrence Berkeley National Laboratory



DIALS

What, no GUI? We are focussing on algorithms, GUI about same amount of effort

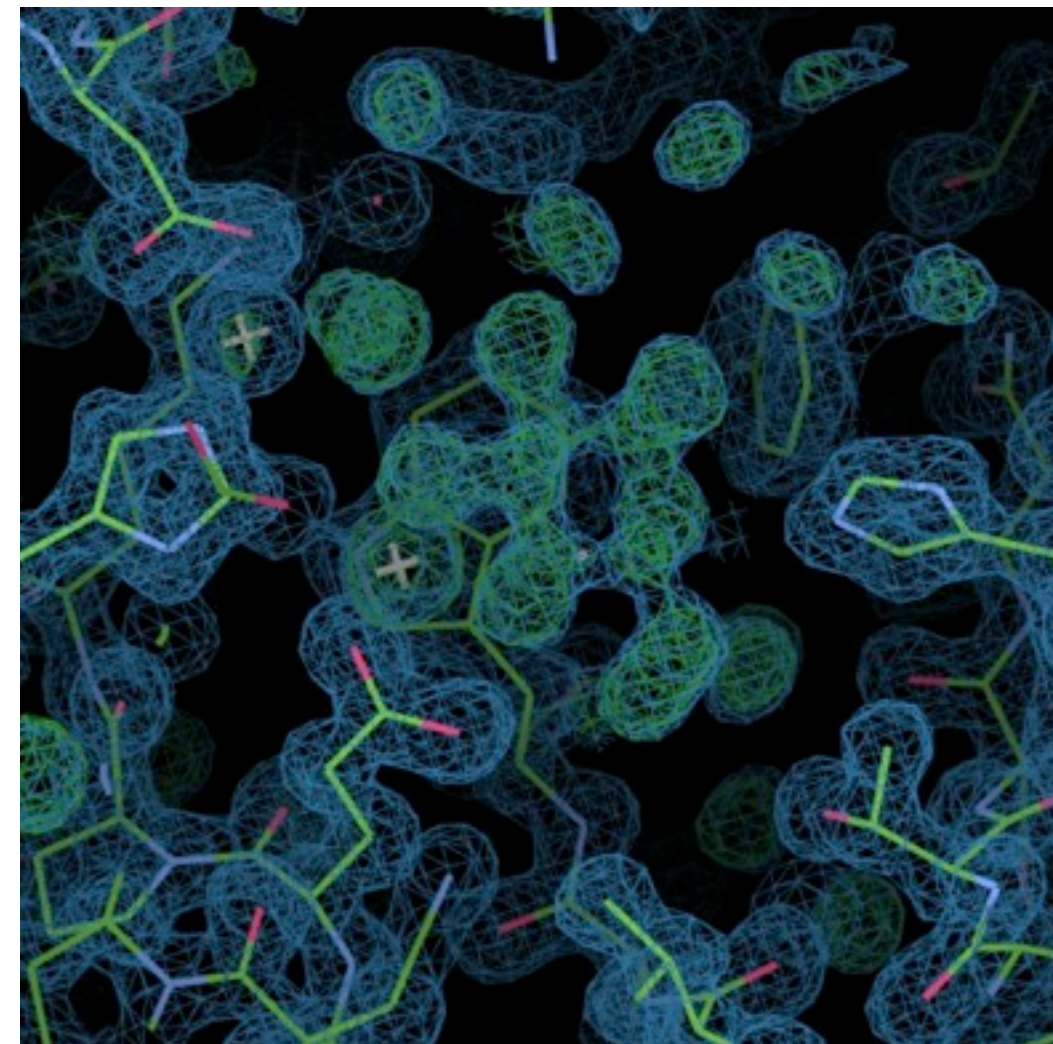
Results: Data quality today consistent with leading packages.
Currently working on chasing XDS in terms of data quality. Also we have support for modern detectors e.g. Pilatus 12M on Diamond's I23 and CS-PAD at the LCLS



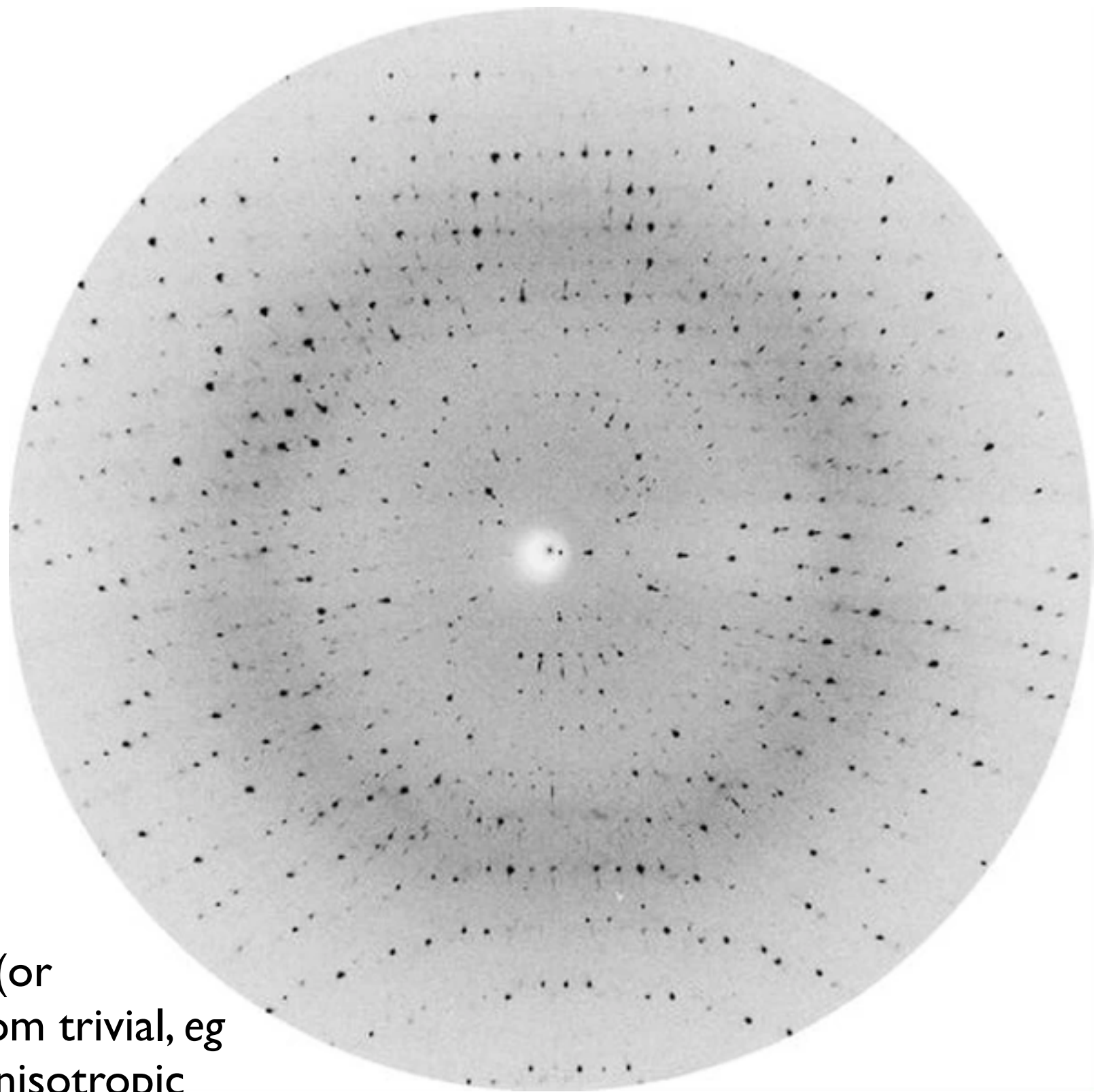
DIALS

It does work! (but still very much under development)

Low resolution limit	71.03	71.03	1.32
High resolution limit	1.30	7.12	1.30
Rmerge (all I+ and I-)	0.087	0.037	0.887
Rmeas (all I+ & I-)	0.096	0.040	0.982
Rpim (all I+ & I-)	0.040	0.017	0.415
Total number of observations	649690	4514	30623
Total number unique	115431	793	5655
Mean((I)/sd(I))	10.6	34.4	1.7
Mn(I) half-set CC(1/2)	0.998	0.999	0.688
Completeness	100.0	100.0	99.9
Multiplicity	5.6	5.7	5.4
R*	23.50		
Rfree*	24.31		



Note that a series of images samples the full 3-dimensional reciprocal space, Bragg diffraction and any other phenomena, all scattering from crystal and its environment.



In practice, defects in the crystals (or detectors) make integration far from trivial, eg weak diffraction, crystal splitting, anisotropic diffraction, diffuse scattering, ice rings/spots, high mosaicity, unresolved spots, overloaded spots, zingers/cosmic rays, etc, etc.

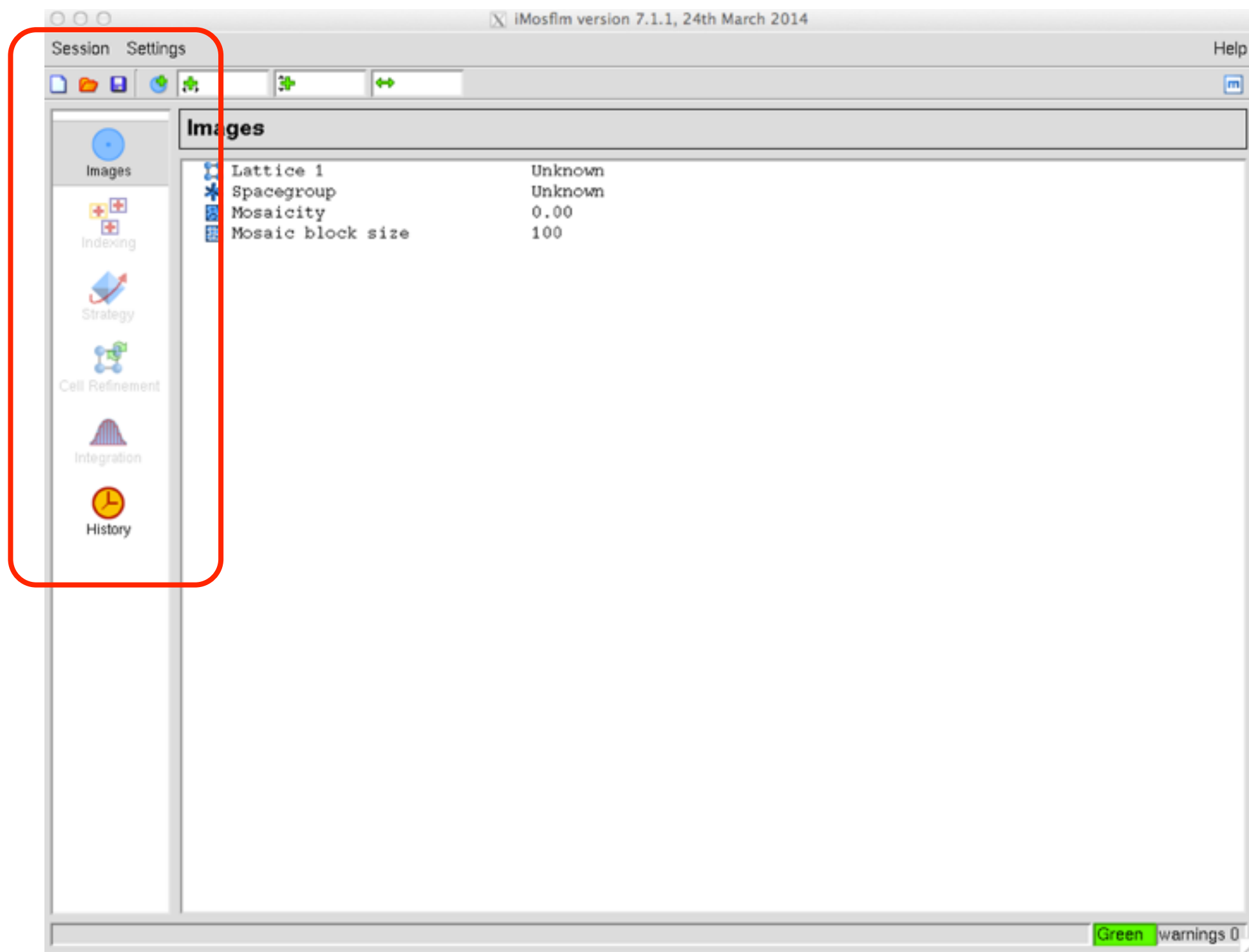
Integration

We want to calculate the intensity of each spot: then working backwards –

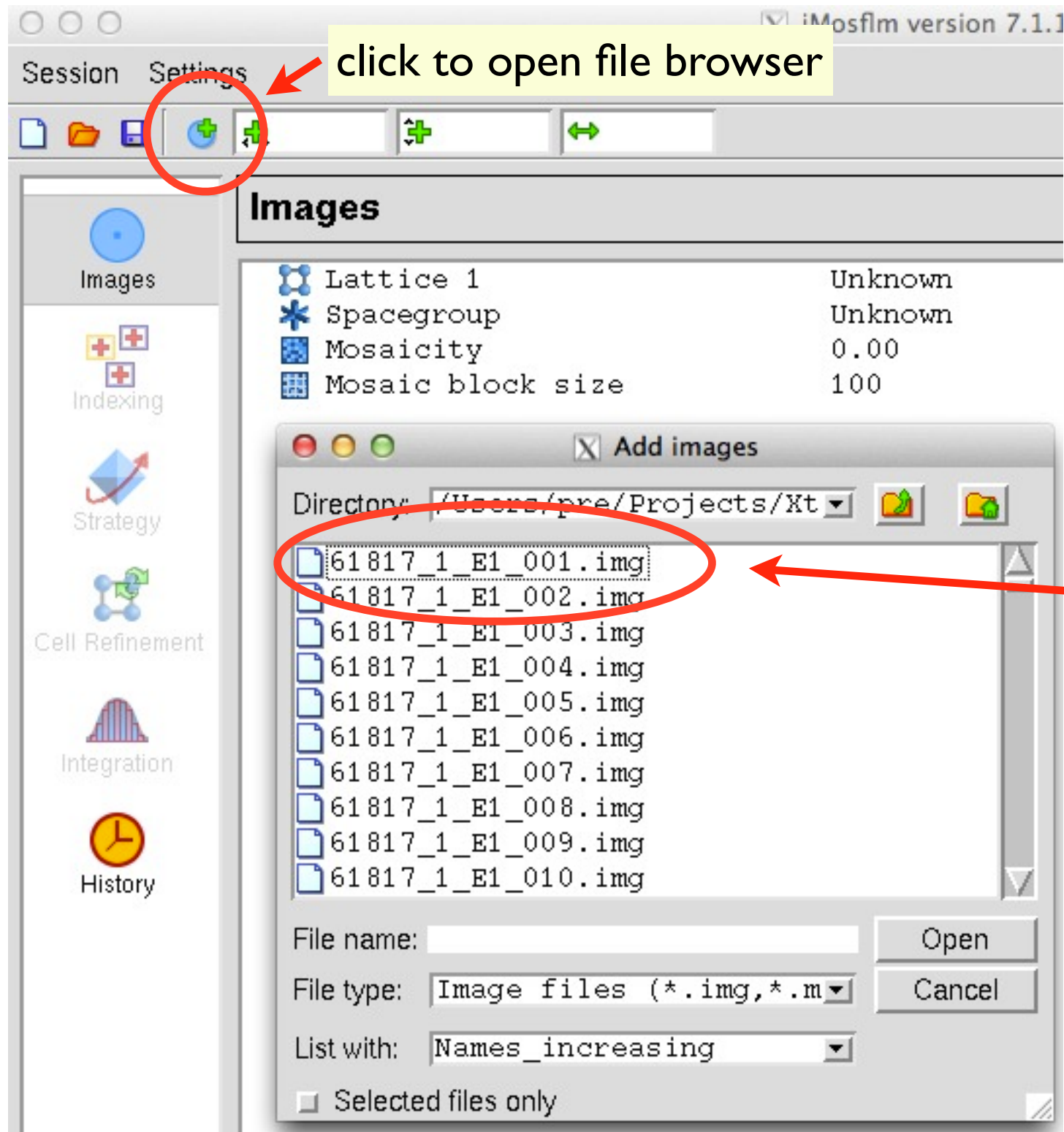
- The simplest method is draw a box around each spot, add up all the numbers inside, & subtract the background (or better, fit profile)
- To do this, we must know **where** the spot is: this needs
 - the unit cell of the crystal
 - the orientation of the crystal relative to the camera
 - the exact position of the detector
- To find the unit cell and crystal orientation, we must index the diffraction pattern
 - this can be done by finding spots on one or more images

iMosflm window

successive stages
of analysis



Select and load images



double-click on any image to import the whole series

This also opens the image display window for the first image

MOSFLM recognises *most* image formats, and tries to extract useful information from the header

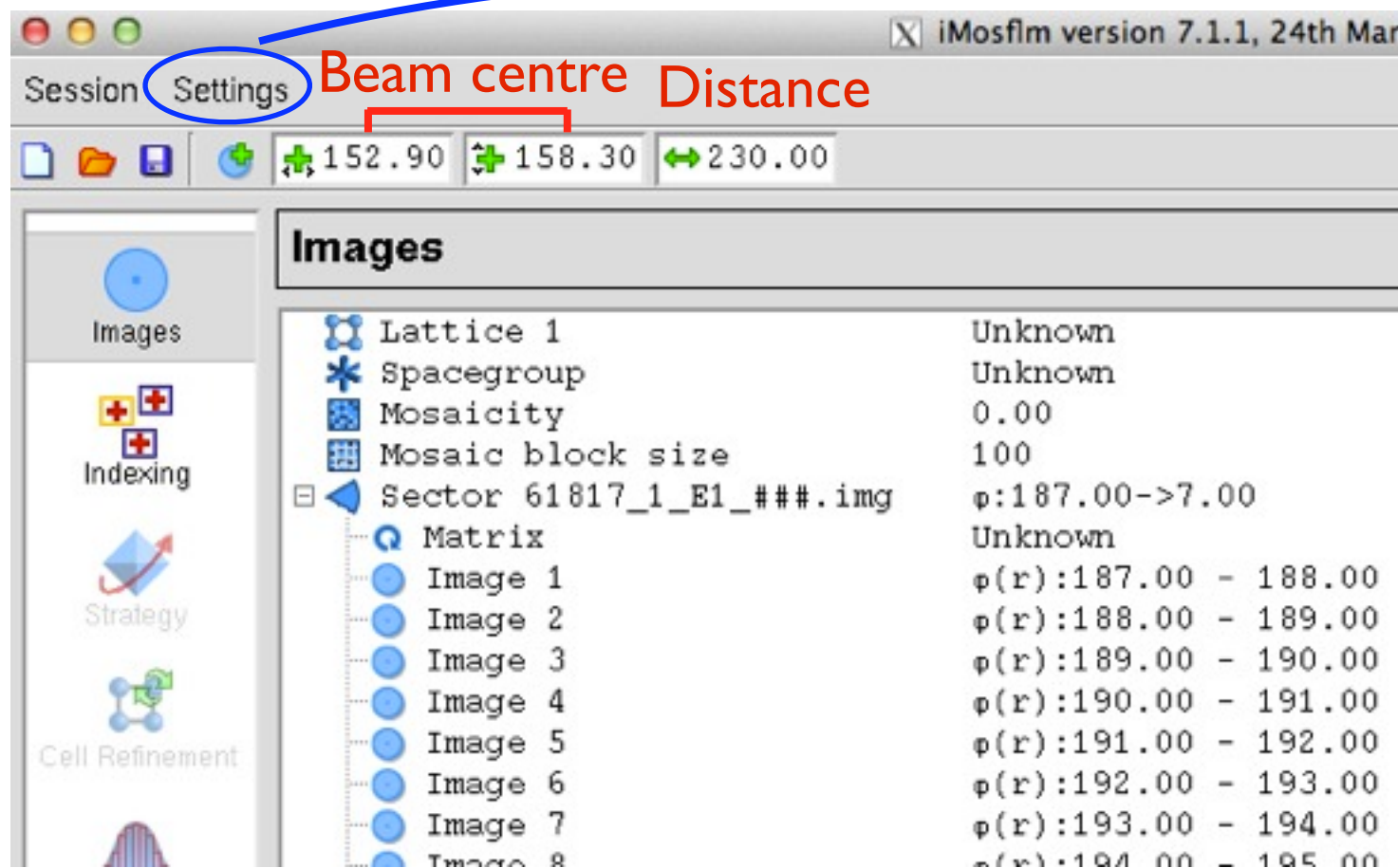
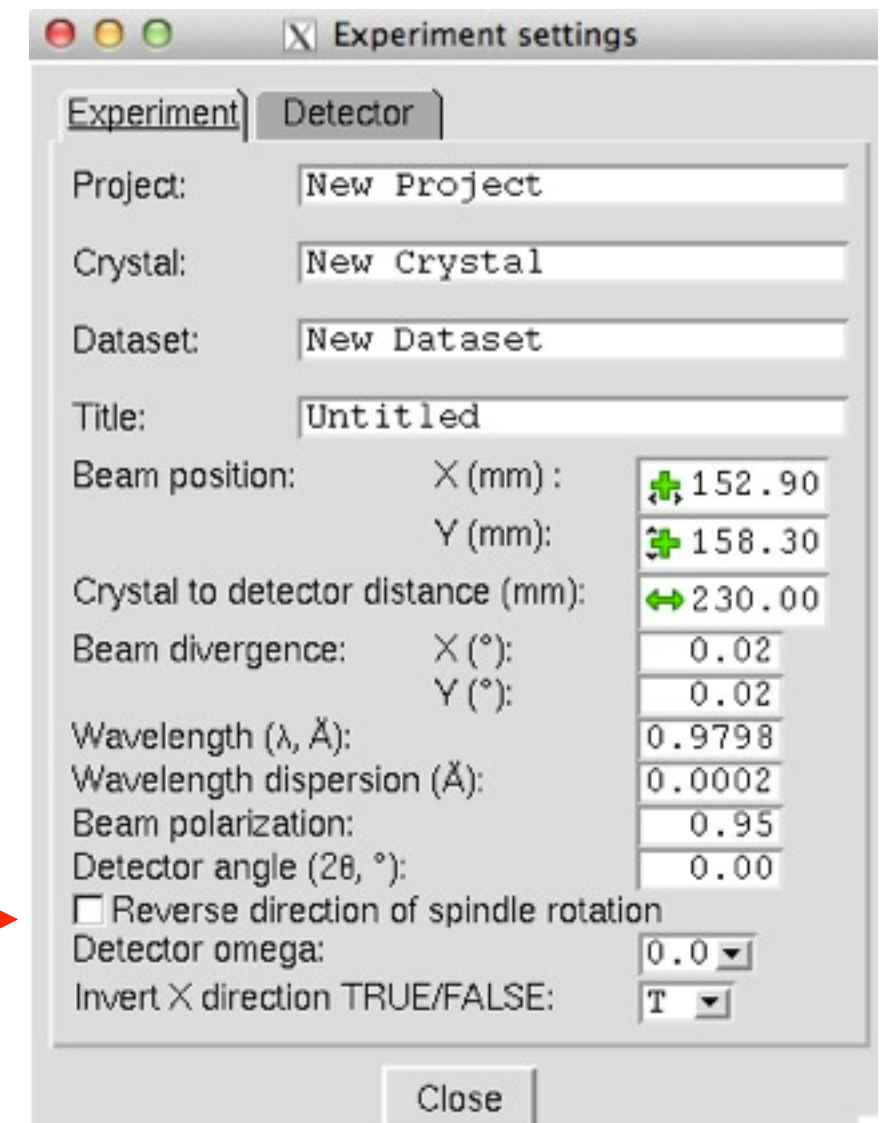


Image phi ranges

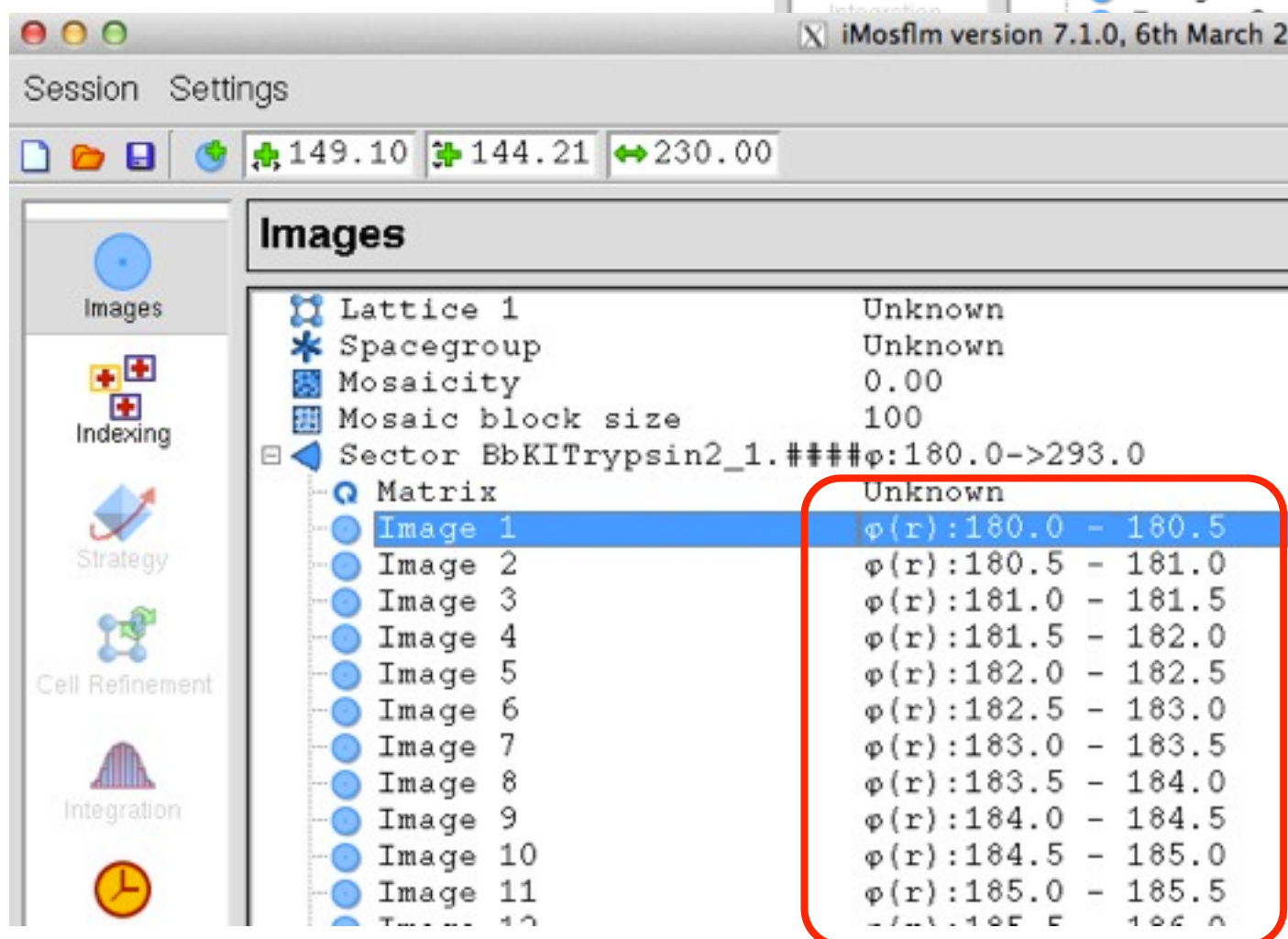
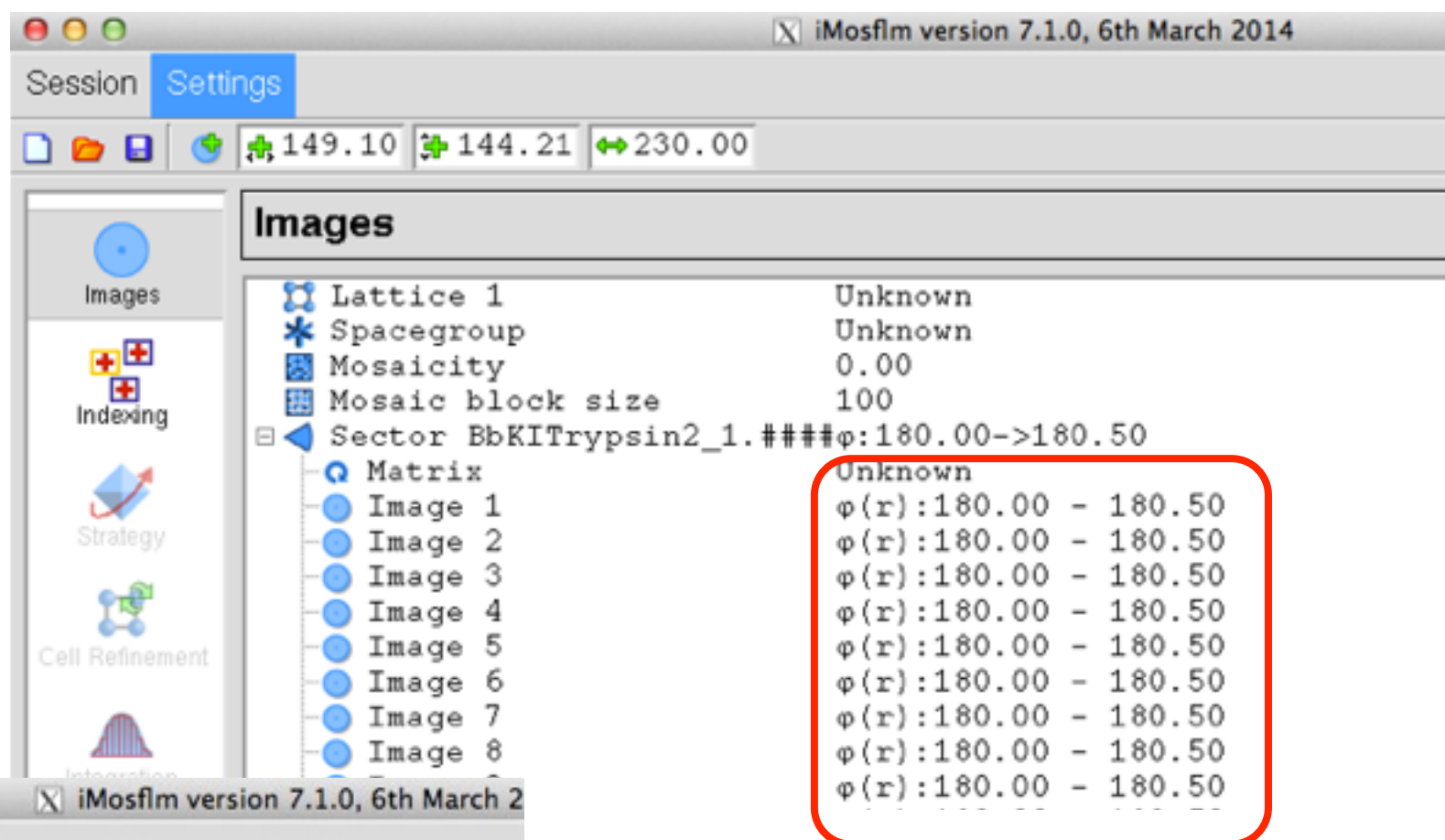
Parameters may be reset in the main window, or in the Settings → Experiment settings window

A few beamlines have a “reversed” spindle: this can be indicated here



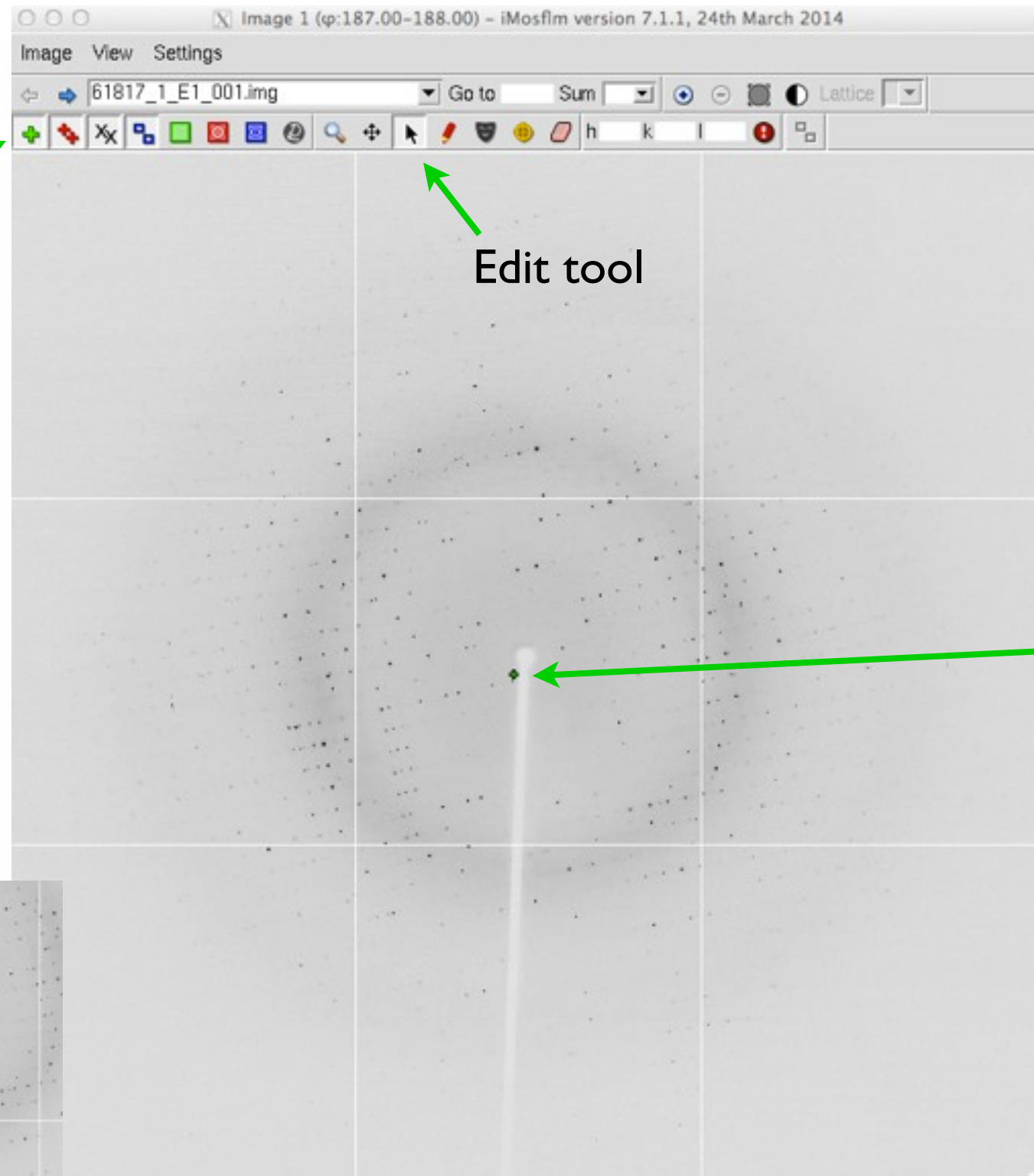
What can go wrong?

Some image sets have wrong Phi ranges, eg all the same



click first range, edit, and all will change to consecutive ranges

Image display window: see Tutorial for details



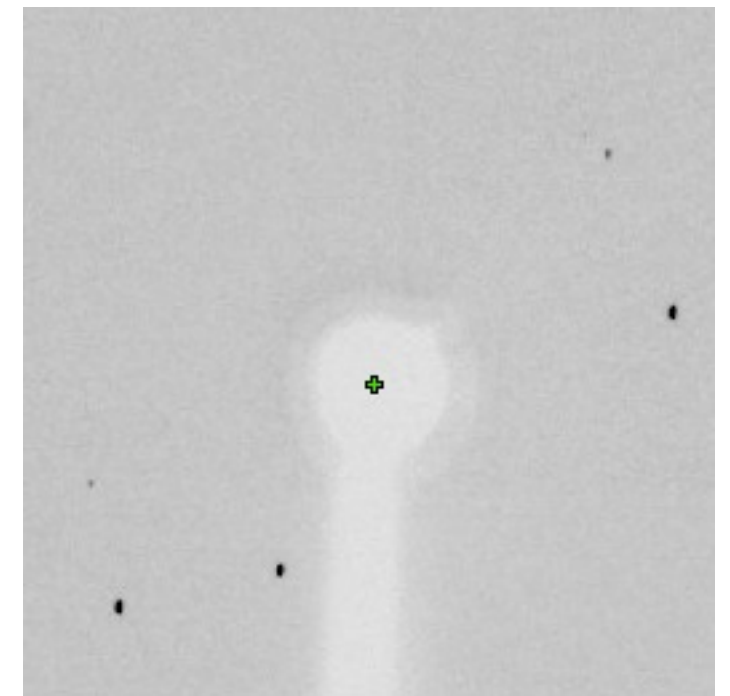
Display main beam position

Edit tool

Controls:

- zoom & pan
- choose overlays eg spots
- move main beam
- edit backstop mask
- etc

Main beam in wrong position

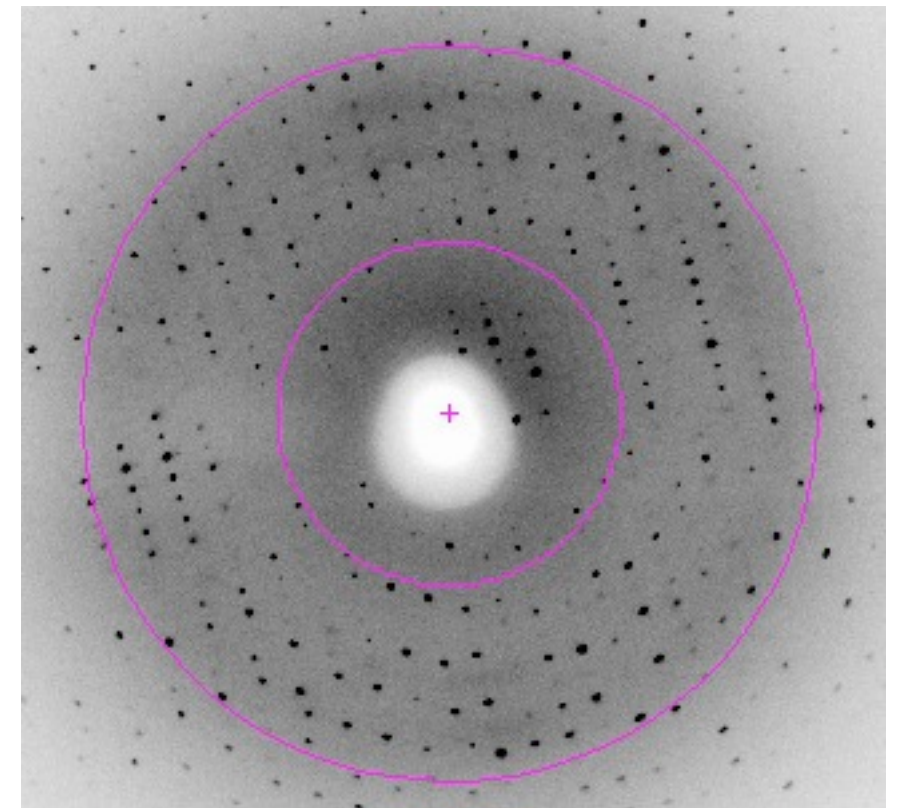
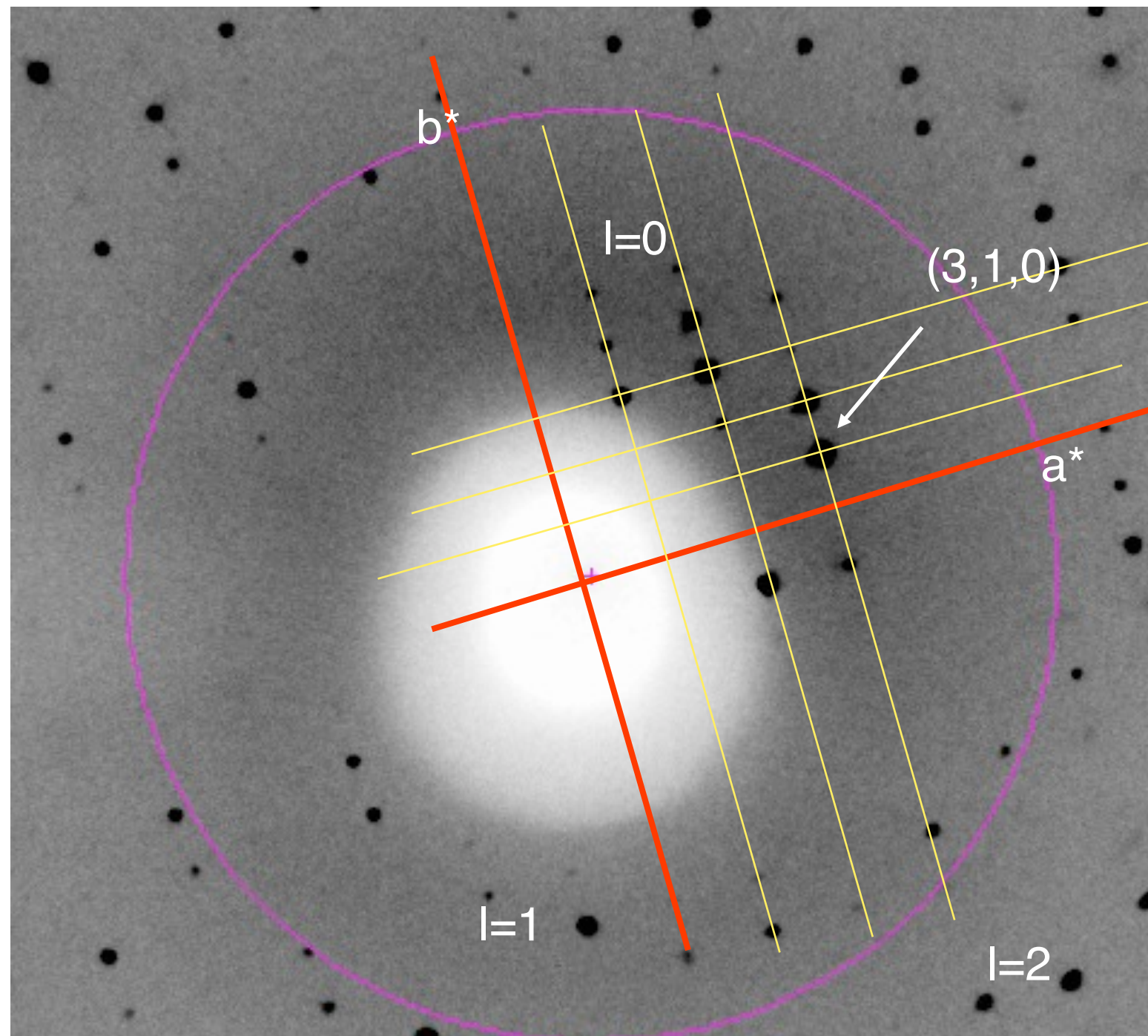


Moved to roughly right position

Note manually drawn mask for beam-stop shadow

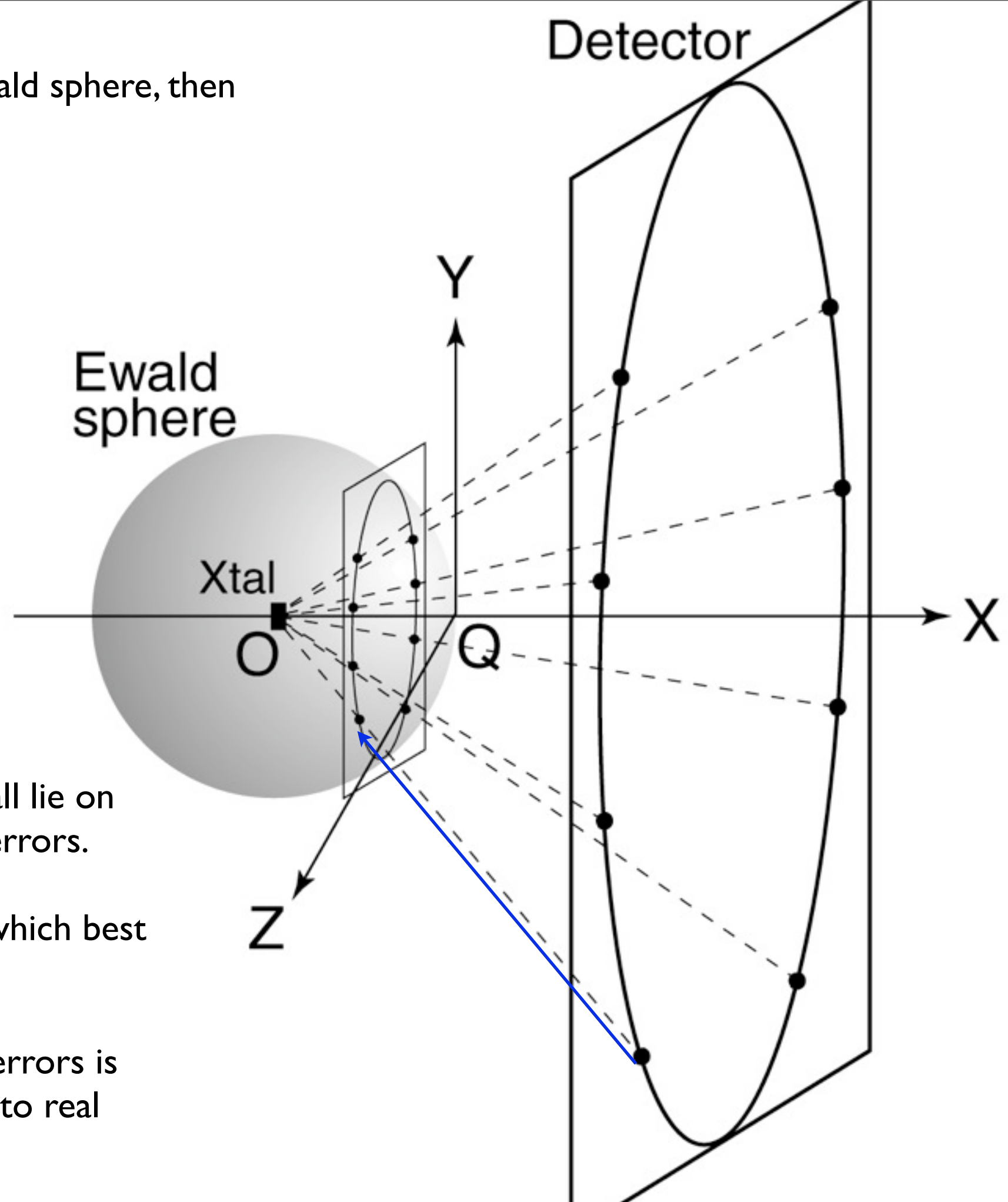
Indexing

If we know the main beam position on the image, we can count spots from the centre



To do it properly, we need to put the spots into 3 dimensions, knowing the rotation of the crystal for this image

Back-project each spot on to Ewald sphere, then rotate back into zero- φ frame

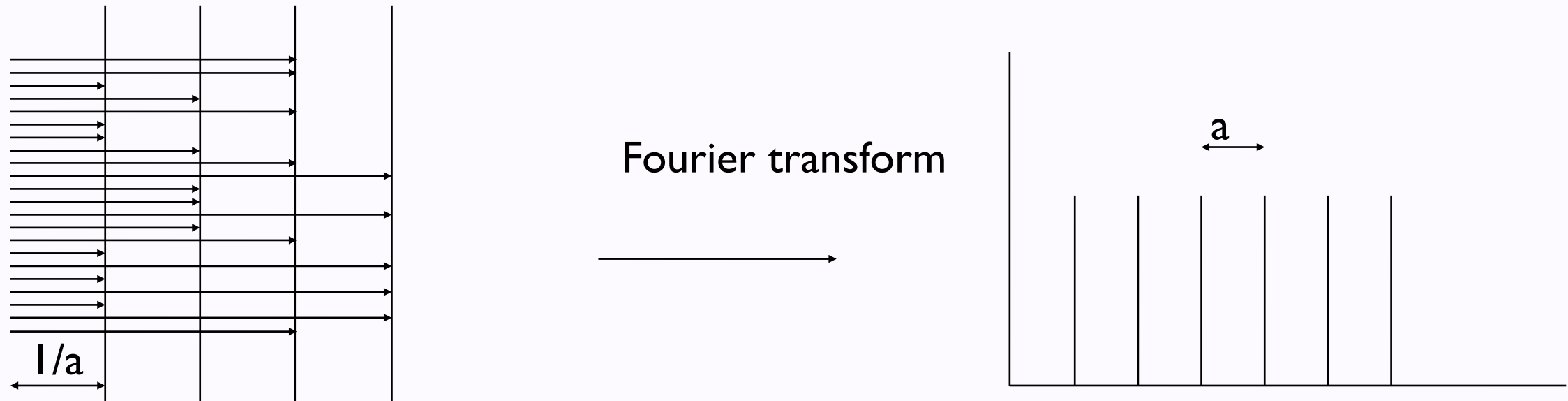


This gives a list of vectors which all lie on the reciprocal lattice, with some errors.

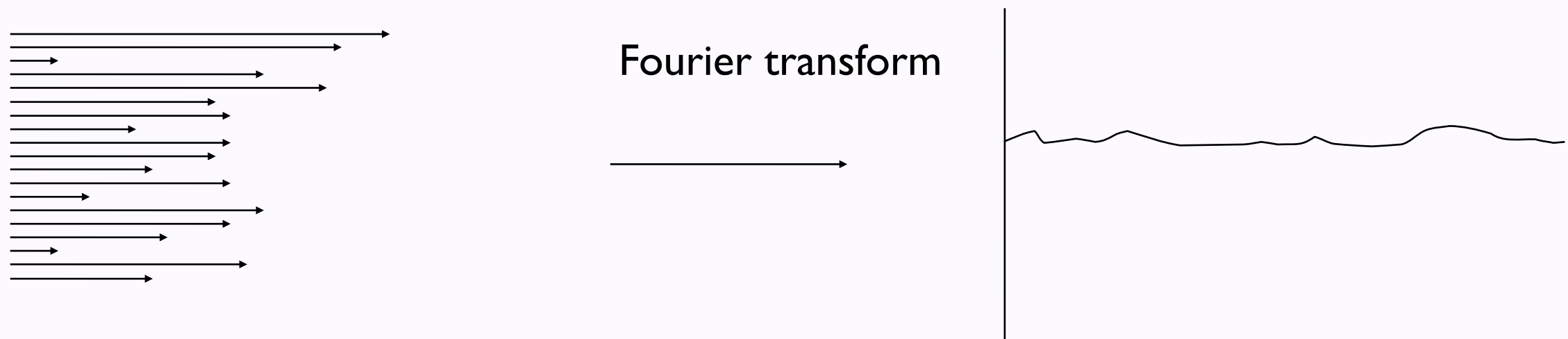
We then want to find a **lattice** which best fits these vectors.

The best way to average out the errors is to use a **Fourier transform** into real space

Consider every possible direction in turn as a possible **real-space** axis, ie perpendicular to a reciprocal lattice plane. Project all observed vectors on to this axis

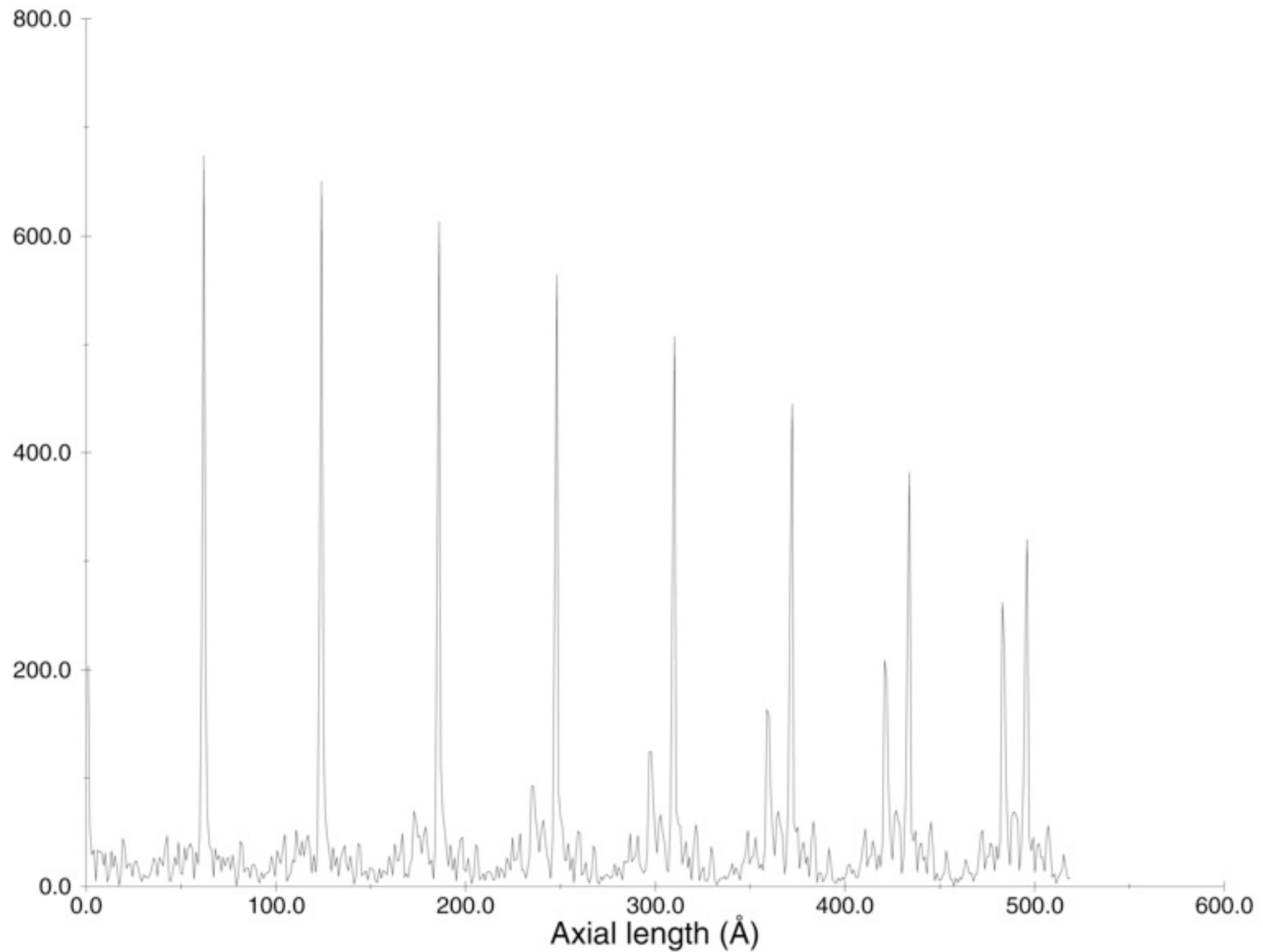


Lattice plane normal to lattice plane: vectors cluster at lengths which are multiples of the lattice spacing. Fourier transform shows sharp peaks



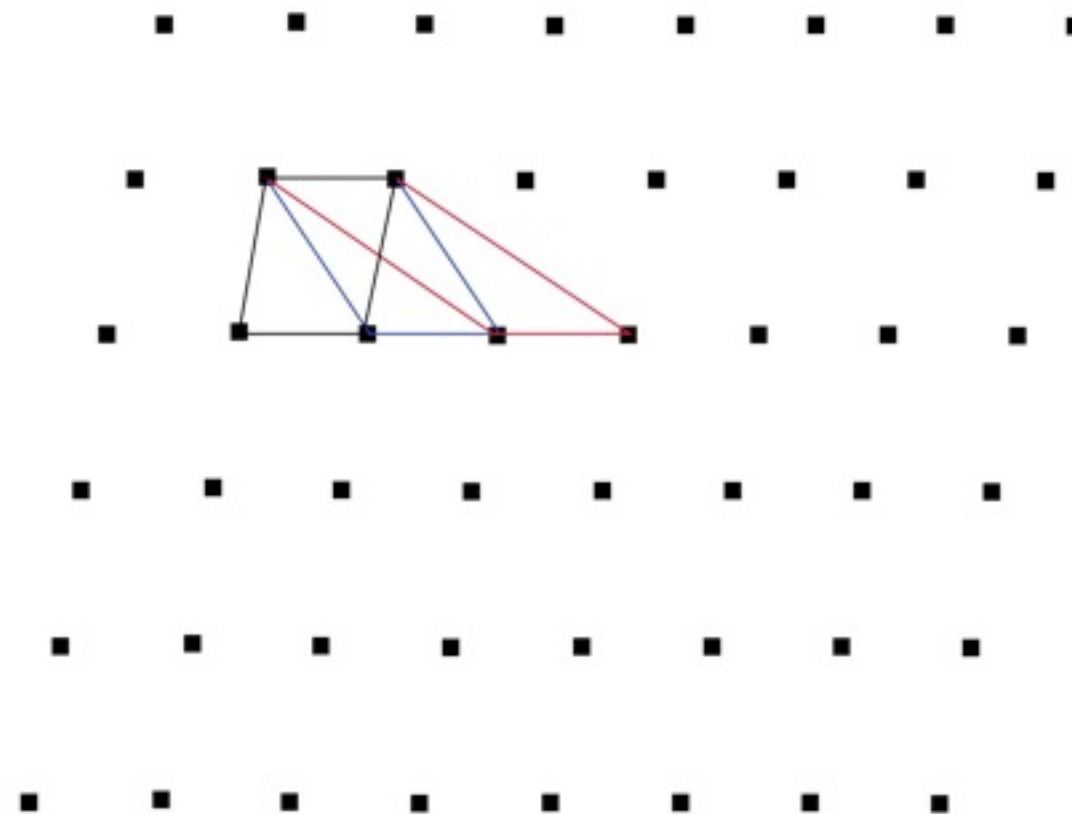
Non-lattice direction, random length. No peaks in Fourier transform

1D Fourier transform of projected scattering vectors



Pick three non-coplanar directions which have the largest peaks in the Fourier transforms to define a lattice.

This is not necessarily the simplest lattice (the “reduced cell”)



In the 2D example shown, the black cell corresponds to the reduced cell, while the red or blue cells may have been found in the autoindexing.

Indexing

iMosflm version 7.1.1, 24th March 2014

Session Settings

156.17 152.84 230.00 5.00 10.0 0.51 0.61 0.00 20 366 2.5

Autoindexing

61817_1_E1_###.img: 1, 90

change images to use

Image	φ range	Auto	Man	Del	> I/σ(I)	Find	Use
1	187.00 - 188.00	611	0	0	320		☒
90	276.00 - 277.00	640	0	0	293		☒

Finds spots on 2 images ~90° apart, according to set criteria

Total 1251 0 0 613

Lattice 1

Solution	Lat.	Pen.	a	b	c	α	β	γ	σ(x,y)	σ(φ)	δ beam
1 (ref)	aP	0	35.3	44.2	76.8	104.0	90.2	89.9	0.81	1.17	1.20 (0.4)
2 (ref)	aP	0	35.3	44.2	76.8	76.0	89.8	89.9	0.81	1.17	1.20 (0.4)
3 (ref)	mP	1	44.2	35.3	76.7	90.0	104.0	90.0	0.82	1.16	1.20 (0.2)
4 (ref)	mC	35	149.9	44.1	35.4	90.0	89.4	90.0	0.81	1.04	1.20 (0.8)
5			148.5	35.3		90.0	90.1	90.0	0.99	1.20	1.10 (0.7)
5			148.6	35.3		90.0	90.0	90.0	1.00	1.20	1.10 (0.7)
3			44.2	76.9		90.0	90.2	90.0	-	-	-
3			76.9	44.2		90.0			-	-	-
3			44.2	76.9		90.0			-	-	-
7			35.3	44.2		90.0			-	-	-
11 (reg)	mC	148	95.2	35.3	76.9	90.0	104.0	90.0	-	-	-
12 (reg)	mI	179	56.6	153.3	35.3	90.0	104.0	90.0	-	-	-
13 (reg)	oI	179	35.3	44.2	153.3	90.0	90.0	90.0	-	-	-
14 (reg)	mC	180	56.6	56.6	76.9	90.0	101.1	90.0	-	-	-
15 (reg)	mC	180	56.6	56.6	76.9	90.0	101.1	90.0	-	-	-
16 (reg)	tP	182	39.8	39.8	76.9	90.0	90.0	90.0	-	-	-
17 (reg)	oC	182	56.6	56.6	76.9	90.0	90.0	90.0	-	-	-

In this case there is a high positional residual

Shows lattices summary

Spacegroup: P2

Mosaicity: 0.55

Estimate

change chosen spacegroup

Search beam-centre [+]

Green warnings 0

Mosaicity estimation

Image: 61817_1_E1_001.img

Intensity

Estimates reflection width (mosaicity)

beam centre search

What can go wrong?

If indexing fails:

- check main beam position

- try main beam search (button)

- try different images – try single images or more than two

- change maximum cell edge

- change spot selection criterion or resolution range

- if indexing works on one image, but not on two, consider reversed phi

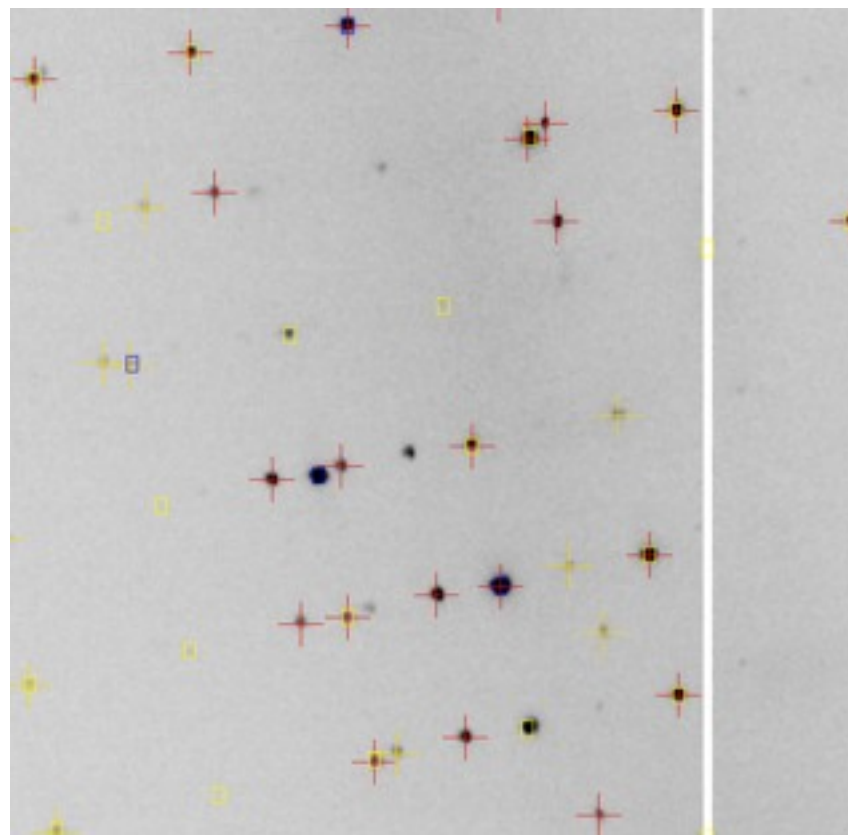
How do you know if it has worked?

Examine critically the predicted pattern superimposed on the image

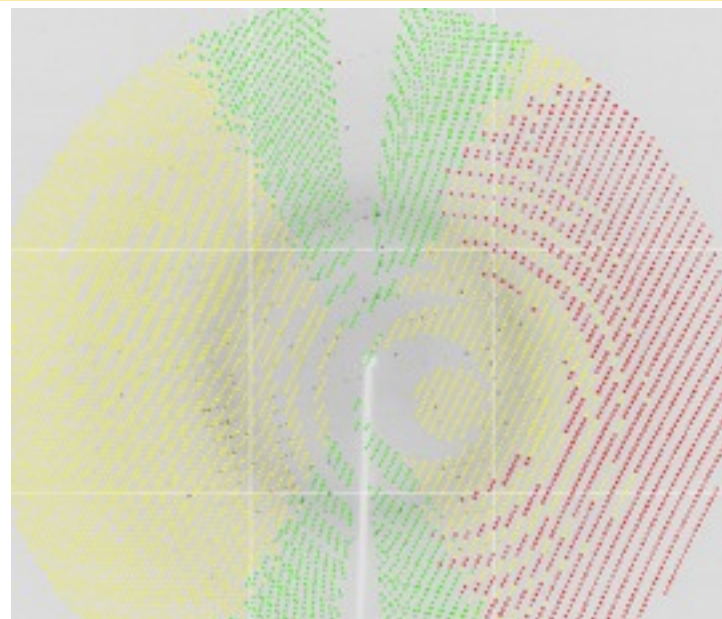
- the prediction should match the observed pattern without extra unpredicted spots

Low value of penalty and $\sigma(x,y)$

NB the true symmetry cannot be determined until after integration



In this case the predicted pattern (blue and yellow boxes) does not explain all the found (observed) spots (red crosses)
ie there are two lattices



Note the overall pattern of prediction: red reflections are overlapped, green too wide to measure. Overlaps come from too wide image width. Both are due to high mosaicity, and are unmeasurable (maybe fake mosaicity and spot size to improve completeness)

Indexing – choosing the “best” solution

Solution	Lat.	Pen.	a	b	c	α	β	γ	$\sigma(x,y)$	$\sigma(\phi)$	δ beam
1 (ref)	aP	0	81.3	81.9	82.3	90.0	90.1	90.0	0.17	0.43	0.01 (0.0)
2 (ref)	nP	0	81.3	81.9	82.3	90.0	90.1	90.0	0.16	0.43	0.00 (0.0)
3 (ref)	aP	0	81.3	81.9	82.3	90.0	89.9	90.0	0.17	0.43	0.01 (0.0)
4 (ref)	nP	1	81.3	82.3	81.8	90.0	90.0	90.0	0.18	0.44	0.00 (0.0)
5 (ref)	nP	1	81.8	81.3	82.3	90.0	90.0	90.0	0.18	0.44	0.00 (0.0)
6 (ref)	oP	1	81.3	81.8	82.3	90.0	90.0	90.0	0.18	0.44	0.00 (0.0)
7 (ref)	nC	3	116.2	115.8	81.4	90.0	90.2	90.0	0.18	0.48	0.00 (0.0)
8 (ref)	tP	3	82.1	82.1	81.3	90.0	90.0	90.0	0.22	0.45	0.00 (0.0)
9 (ref)	oC	3	116.0	116.1	81.3	90.0	90.0	90.0	0.21	0.46	0.00 (0.0)
10 (ref)	nC	3	116.2	115.8	81.4	90.0	90.2	90.0	0.18	0.48	0.00 (0.0)
11 (ref)	nC	4	115.5	115.3	82.2	90.0	90.1	90.0	0.24	0.43	0.01 (0.0)
12 (ref)	oC	4	115.5	115.3	82.3	90.0	90.0	90.0	0.24	0.42	0.01 (0.0)
13 (ref)	tP	4	81.6	81.6	82.3	90.0	90.0	90.0	0.24	0.42	0.01 (0.0)
14 (ref)	nC	4	115.5	115.3	82.2	90.0	90.1	90.0	0.24	0.43	0.01 (0.0)
15 (ref)	hR	6	115.5	115.5	142.4	90.0	90.0	120.0	0.28	0.49	0.01 (0.0)
16 (ref)	hR	6	115.7	115.7	141.7	90.0	90.0	120.0	0.35	0.44	0.01 (0.0)
17 (ref)	cP	7	81.8	81.8	81.8	90.0	90.0	90.0	0.34	0.44	0.01 (0.0)
18 (reg)	nC	248	81.3	183.3	81.9	90.0	90.0	90.0	-	-	-
19 (reg)	oC	248	81.3	183.3	81.9	90.0	90.0	90.0	-	-	-

Spacegroup: P23
Mosaicity: 0.92 Estimate

Note that the list of solutions given by Mosflm are in fact all the **same** solution, with different lattice symmetries imposed, so that if the triclinic solution (number 1) is wrong, then all the others are too

Normally the solution with the highest symmetry from the group of solutions with low penalties is correct. However, beware of pseudosymmetry !

Note here that the chosen solution (cubic) has a $\sigma(xy)$ value of 0.34mm while for the (correct) orthorhombic solution this value is 0.18mm

The true symmetry can only be determined by integrating some of the images

If known, the true space group can be selected from the drop down menu

Indexing multiple lattices

iMosflm version 7.1.1, 24th March 2014

Session Settings Help

156.27 152.86 230.00 5.00 10.0 0.51 0.61 0.00 20 363 2.5

Autoindexing

61817_1_E1_###.img : 1, 90

Image	p range	Auto	Man	Del	> I/ σ (I)	Find	Use
1	187.00 - 188.00	572	0	0	296		☒
90	276.00 - 277.00	573	0	0	261		☒

Index
Index multiple lattices

choose multilattice option from pull-down

lattice tabs

Total 1145 0 0 557

Lattice 1 Lattice 2

Solution	Lat.	Pen.	a	b	c	α	β	γ	$\sigma(x,y)$	$\sigma(q)$	δ beam
1 (ref)	aP	0	35.3	44.1	77.0	104.4	90.0	90.0	0.07	0.31	0.04 (0.0)
2 (ref)	mP	0	44.1	35.3	77.0	90.0	104.4	90.0	0.07	0.31	0.04 (0.0)
3 (ref)	aP	0	35.3	44.1	77.0	75.6	90.0	90.0	0.07	0.31	0.04 (0.0)
4 (ref)	mC	31	44.2	151.3	35.3	90.0	90.0	90.0	0.31	0.45	0.15 (0.1)
5 (ref)	oC	31	44.2	151.3	35.3	90.0	90.0	90.0	0.31	0.45	0.15 (0.1)
6 (ref)	mC	31	151.4	44.2	35.3	90.0	90.1	90.0	0.29	0.44	0.17 (0.1)
7 (reg)	mP	98	35.3	44.1	77.0	90.0	90.0	90.0	-	-	-
8 (reg)	mP	98	35.3	77.0	44.1	90.0	90.0	90.0	-	-	-
9 (reg)	oP	98	35.3	44.1	77.0	90.0	90.0	90.0	-	-	-
10 (reg)	mC	145	95.1	35.3	77.0	90.0	103.3	90.0	-	-	-
11 (reg)	mC	145	157.9	35.3	44.1	90.0	104.0	90.0	-	-	-

Lattice	Symbol	Penalty	a	b	c	α	β	γ	$\sigma(x,y)$	$\sigma(q)$	Split
1	mP	1	44.1	35.3	76.8	90.0	104.1	90.0	0.10	0.43	43.8
2	mP	0	44.1	35.3	77.0	90.0	104.4	90.0	0.07	0.31	-

summary of lattice solutions

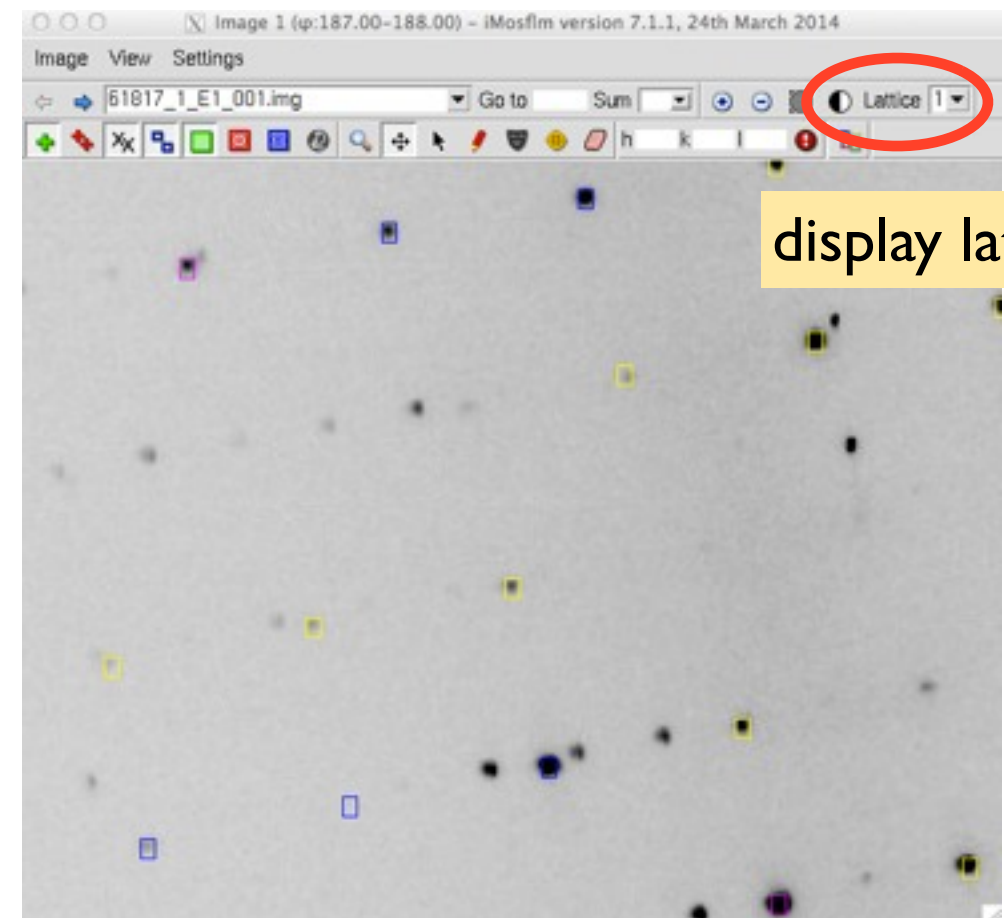
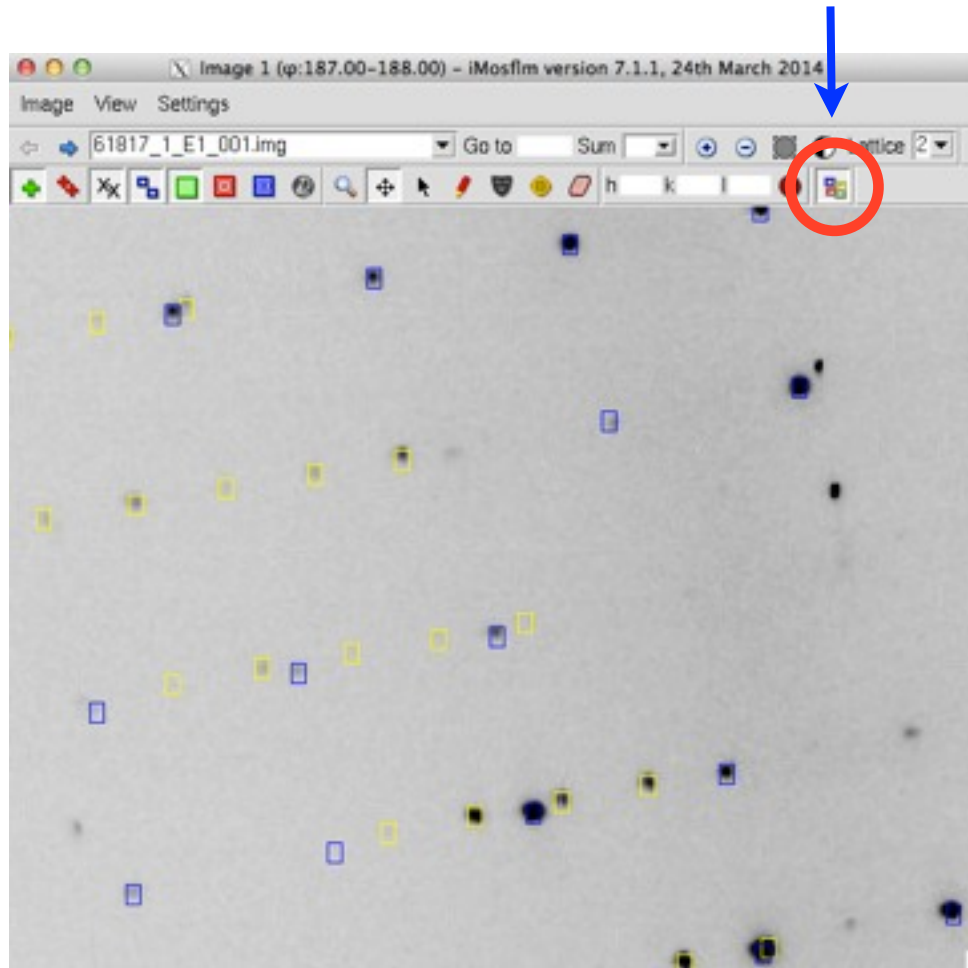
Show lattices summary[-]
Spacegroup: P2
Mosaicity: 0.50 Estimate

Search beam-centre[+]

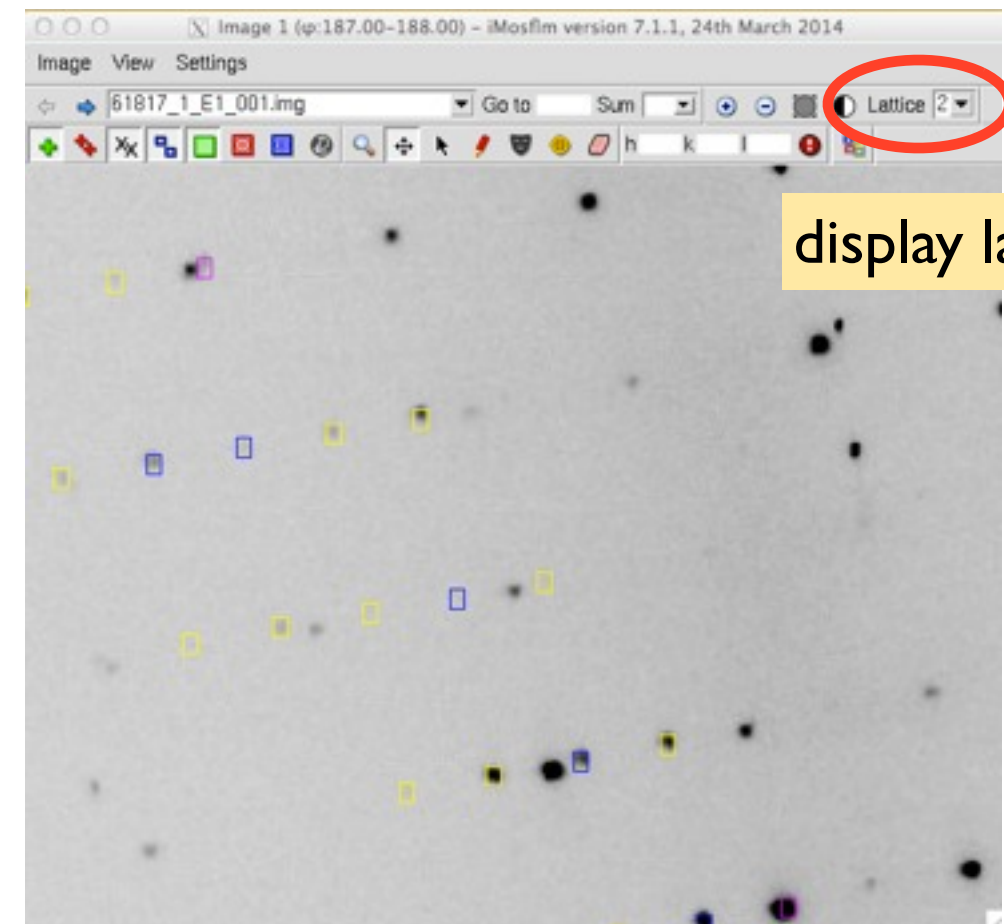
Green warnings 0

Display of multiple lattice solutions

display all lattices, in different colours



display lattice 1



display lattice 2

Processing multiple lattices: warning

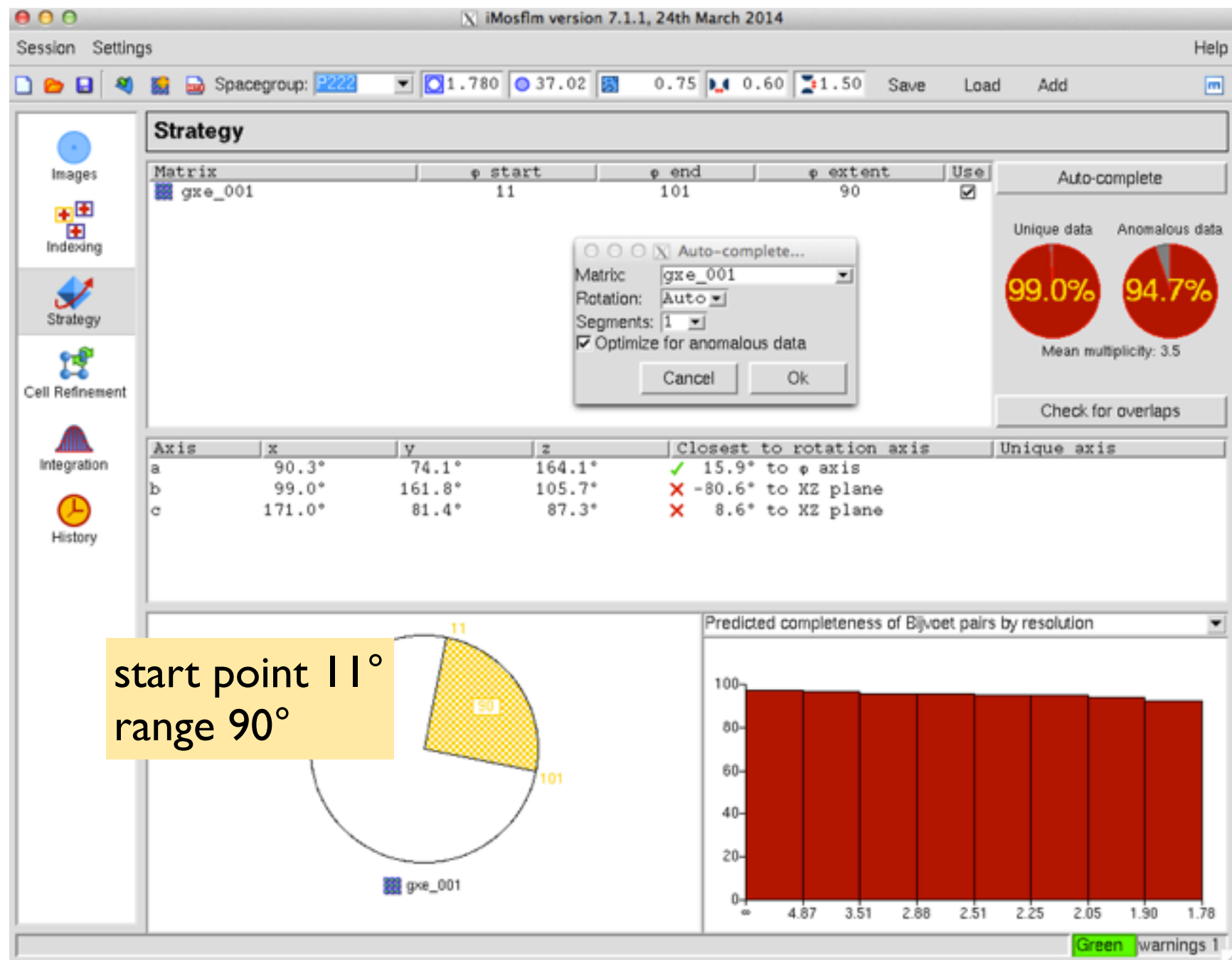
- Very much under development and test
- It may be better to ignore a weak 2nd lattice
- Caveat emptor!

Strategy

Useful before data collection,
after collecting (2 or 3)
orientation images and
indexing

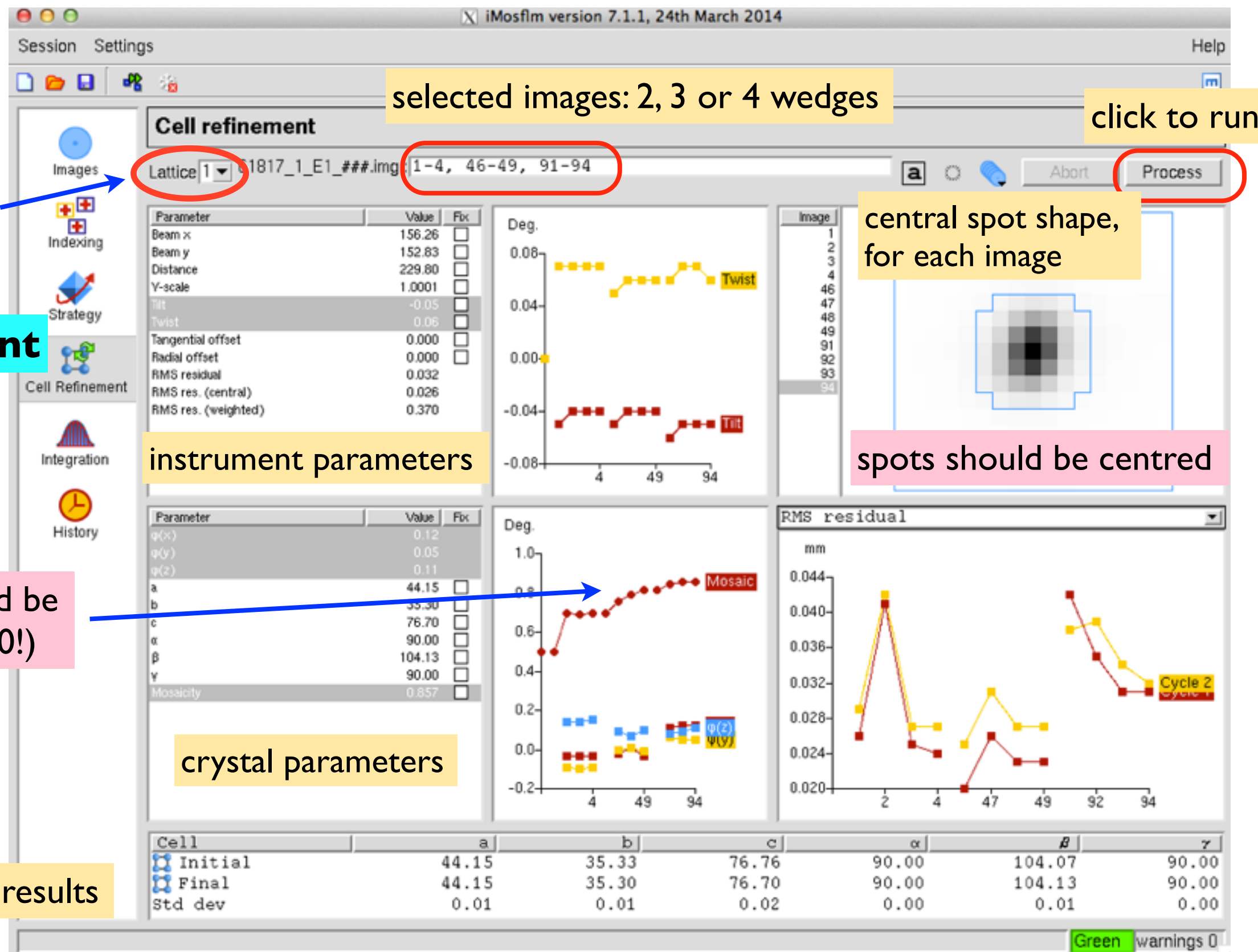
- choose start point and total rotation range
- explore maximum rotation range/image (NB never 1°!) to avoid or minimise overlaps
- find range to collect on eg 2nd crystal to complete partial data from first
- multi-segment strategies

see tutorial for more details, also tomorrow's talk

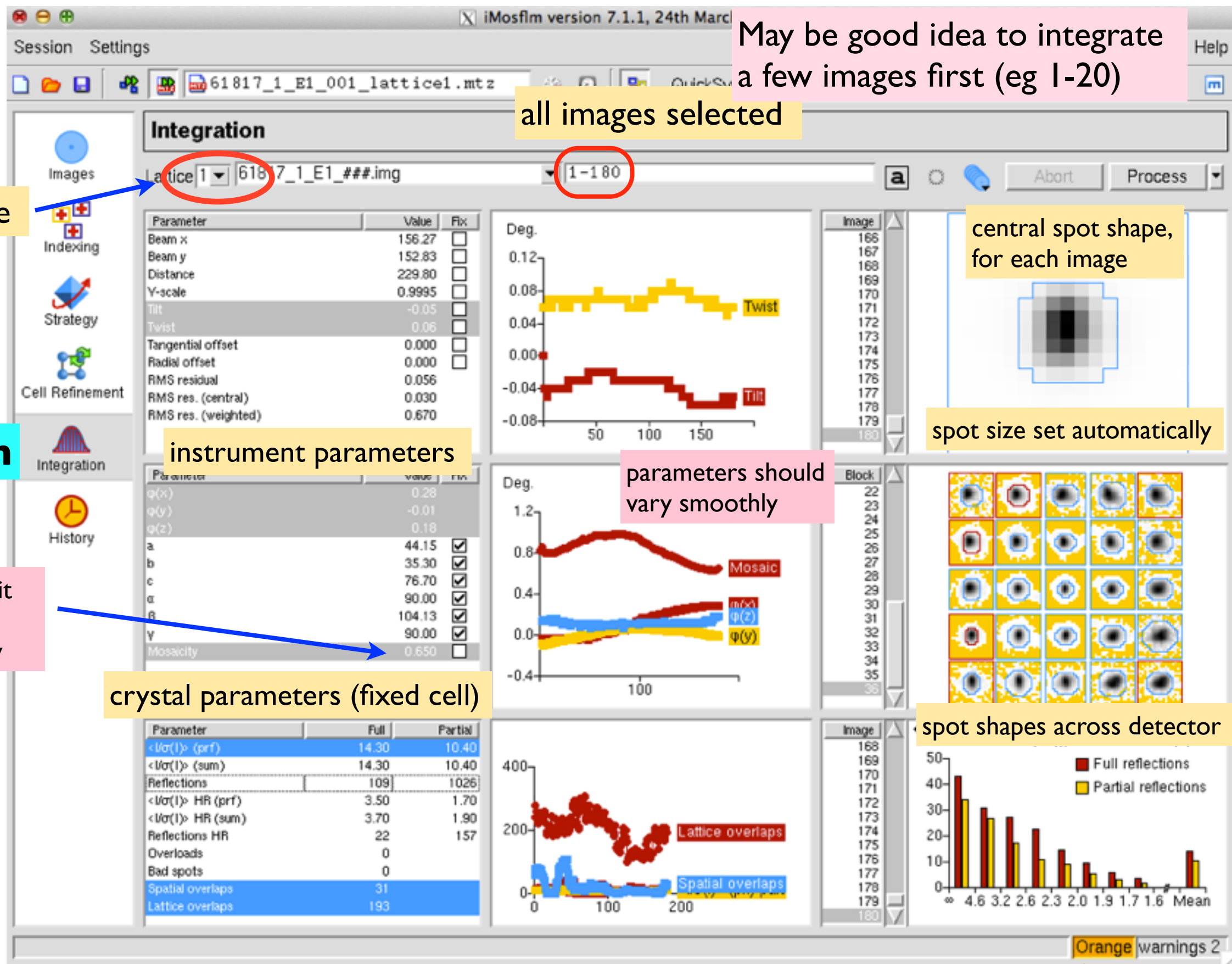


Sensitive to parameters from indexing:

- point group
- mosaicity
- maximum resolution

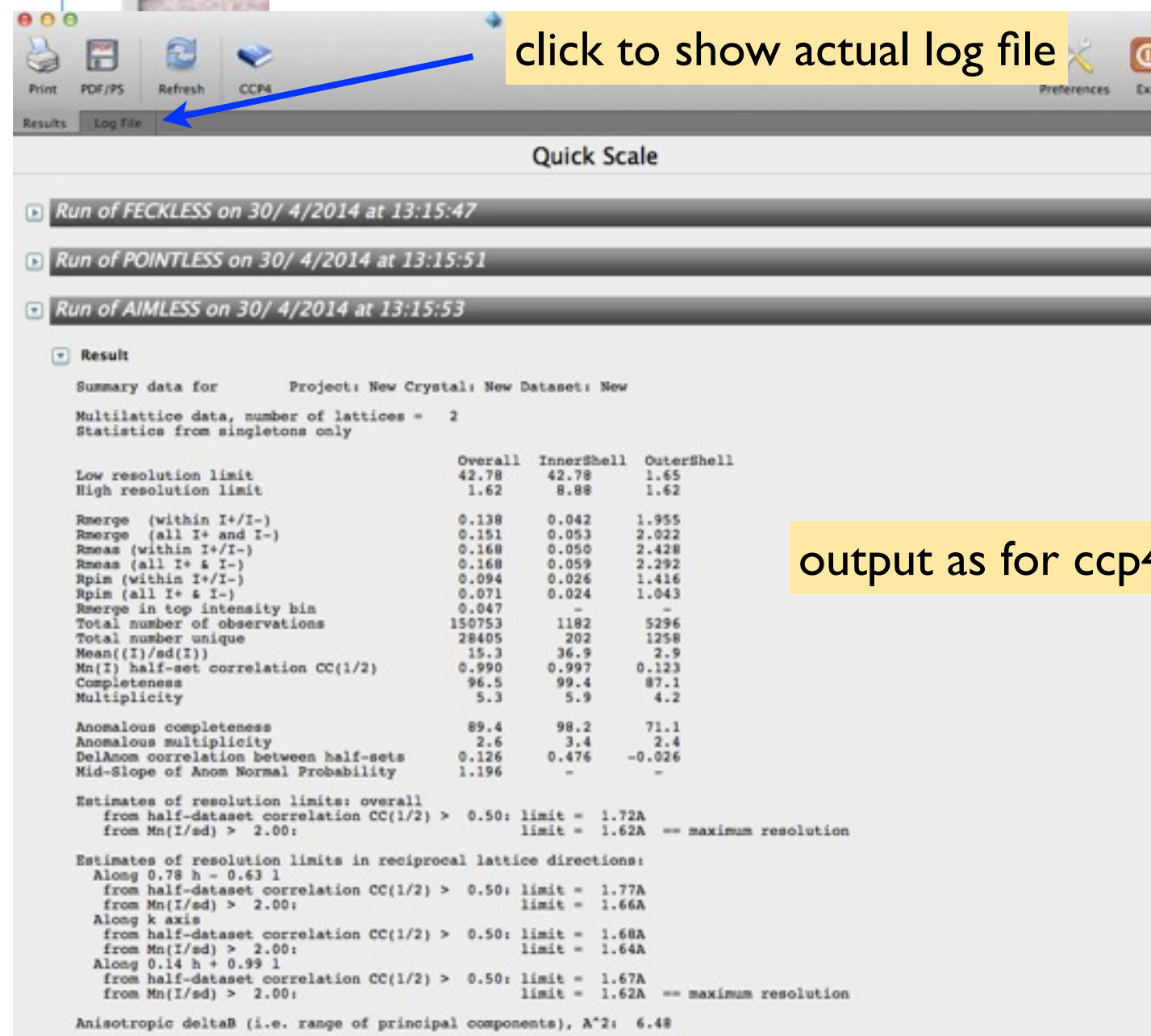
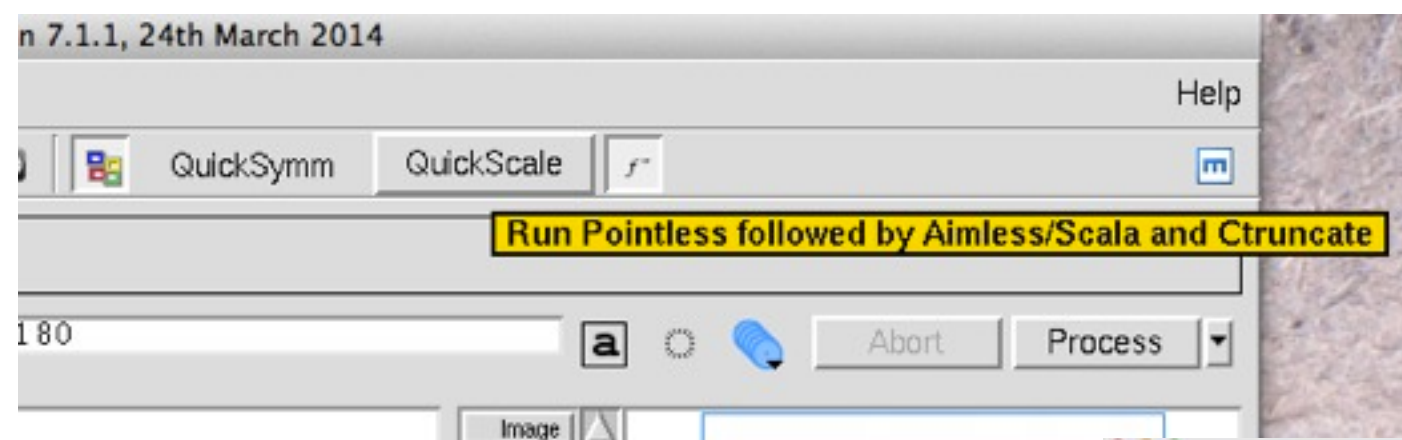


It is best to refine the cell from 2 or more widely-spaced wedges, then fix it during integration. Refinement works badly at low resolution ($> \sim 3.5\text{\AA}$), particularly in low symmetry: just index from several images (eg 5 or 6)



For multiple lattices, run for each lattice, writing a different MTZ file

QuickSymm and QuickScale



QuickScale runs:

[• If there are multiple lattices, FECKLESS to combine lattice files]

- POINTLESS to determine the point group and maybe the space group
- AIMLESS to scale and merge the data
- CTRUNCATE to convert I to F and detect twinning
- FREERFLAG to generate a Free-R set

Options may be set from Settings
(more options are available from ccp4i)

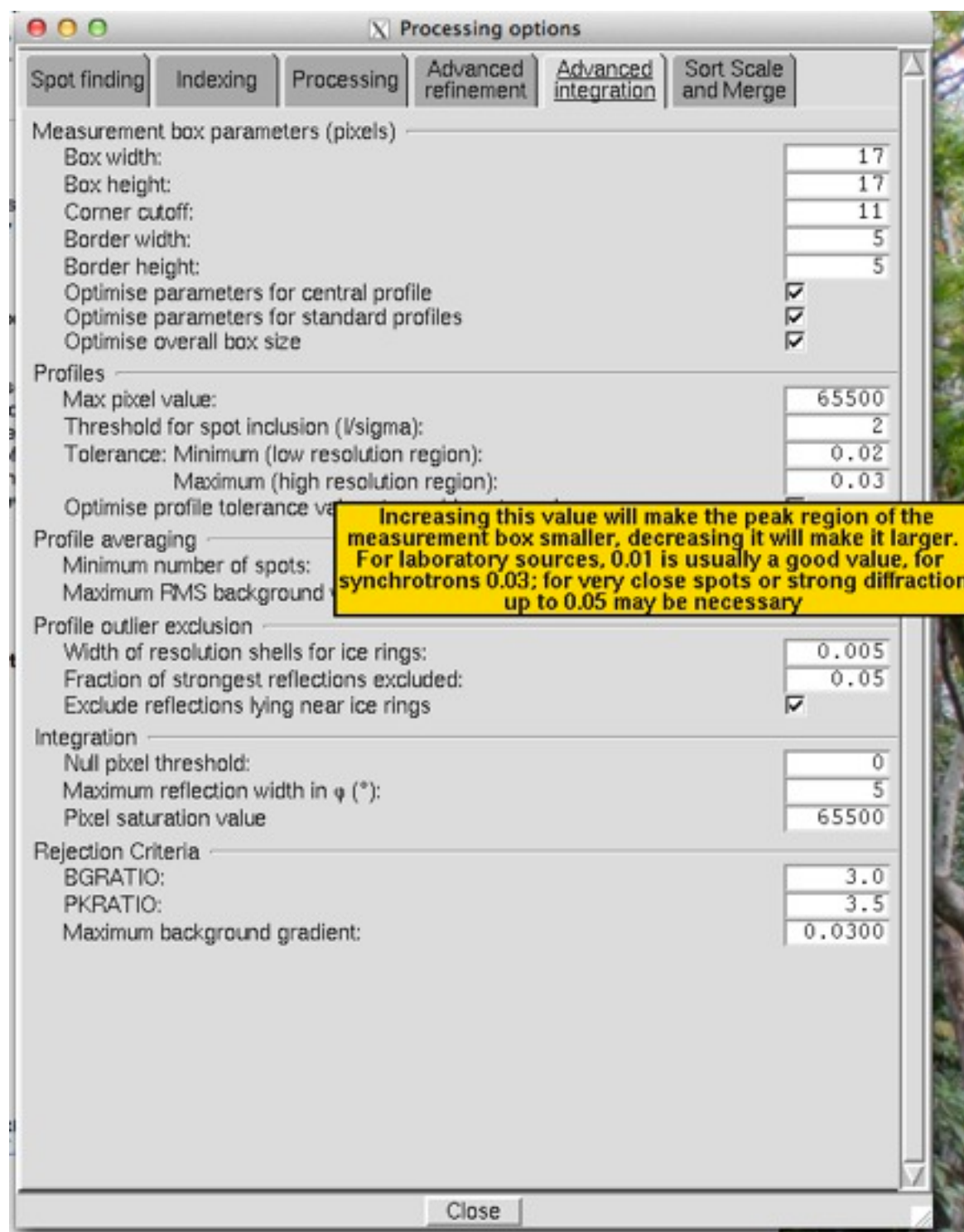
output as for ccp4i

Parameter options

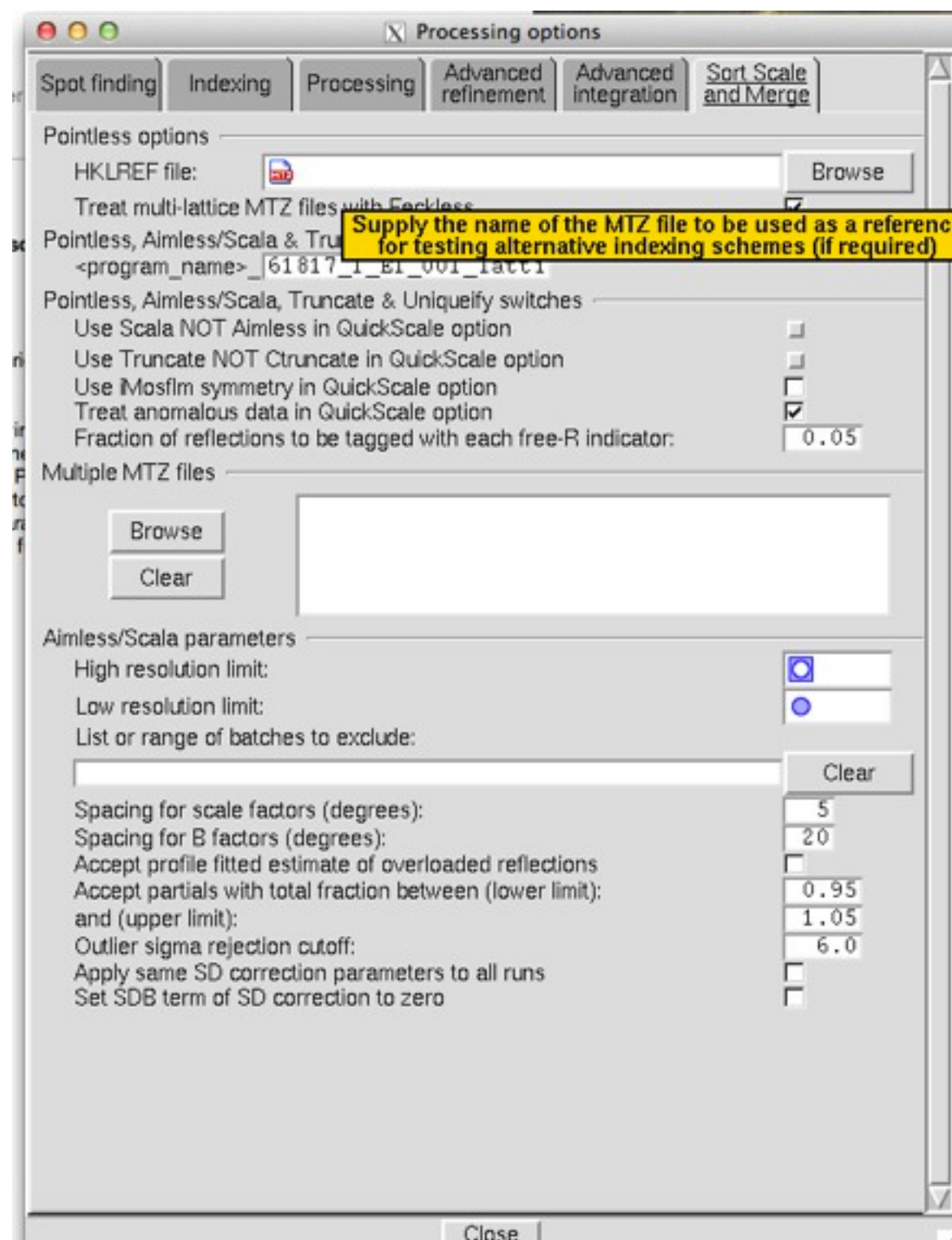
The Settings menu allows many parameters to be changed

Integration parameters

eg to control the box size to reduce overlaps



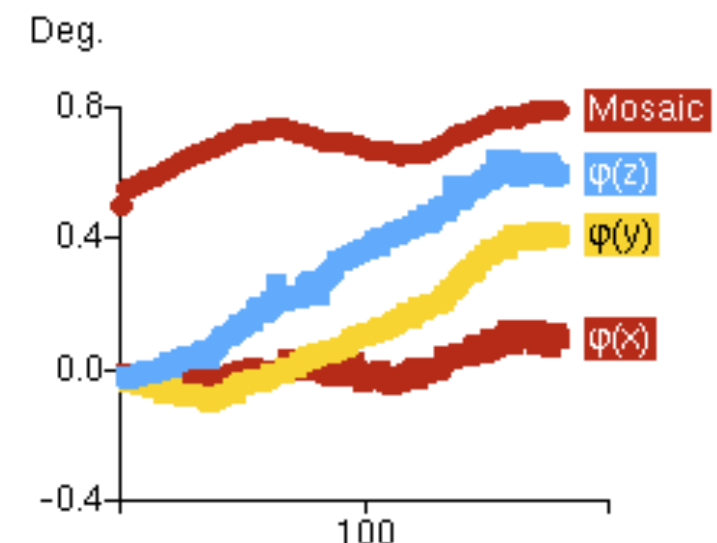
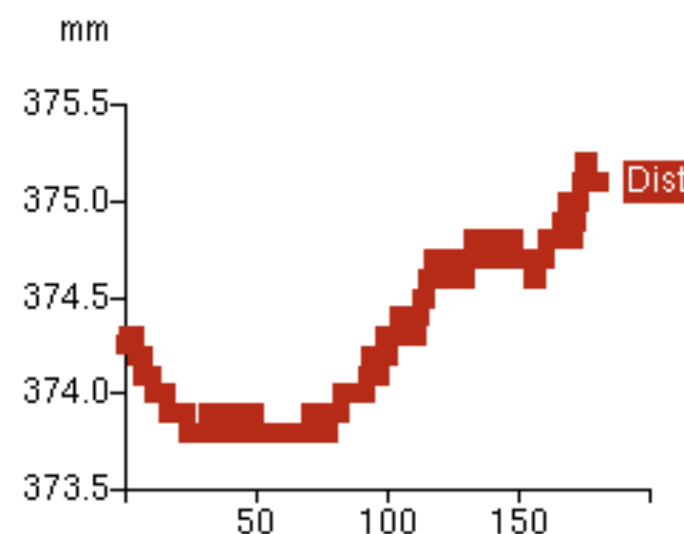
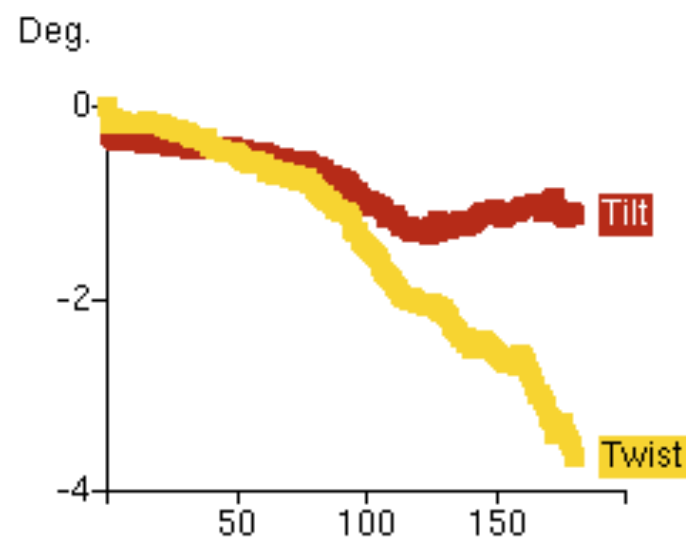
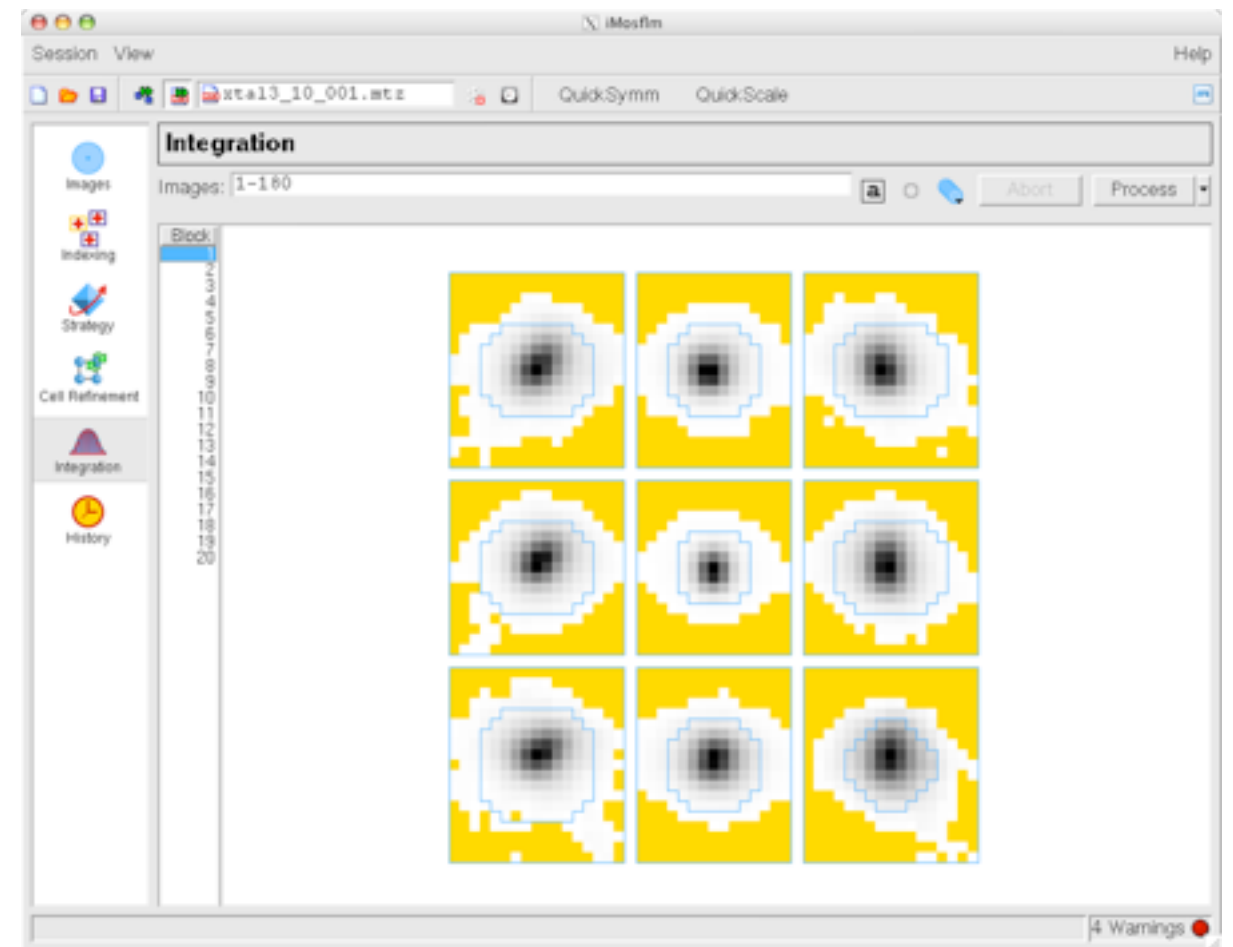
Options for QuickSymm & QuickScale



Integration – what should NOT happen

3Å weakly diffracting data,
monoclinic cell

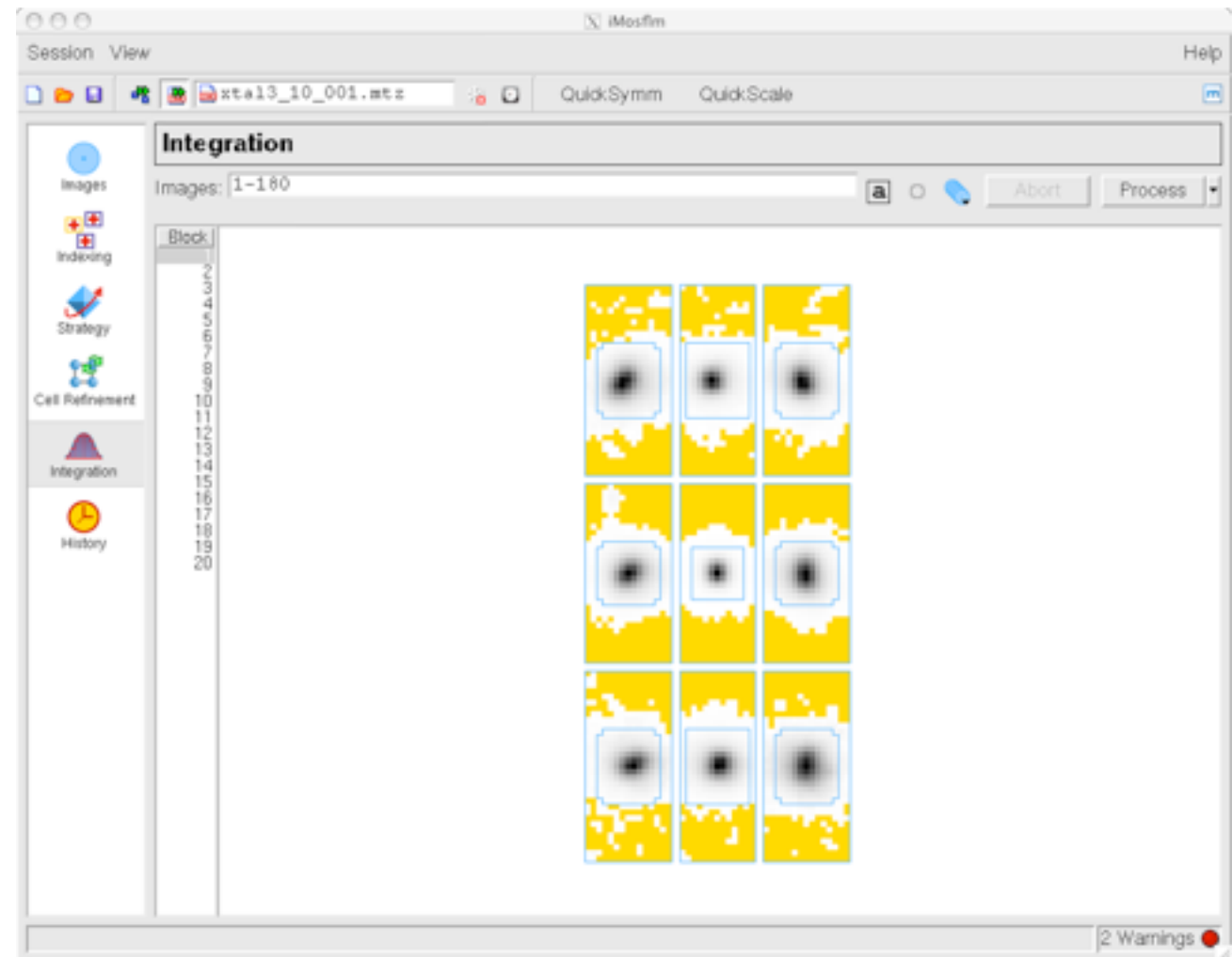
indexed from the first image only



Excessive variation in detector Tilt/Twist and distance also indicate a problem

Integration – what should NOT happen

The situation improves significantly if the cell is refined (using 2 segments of data)

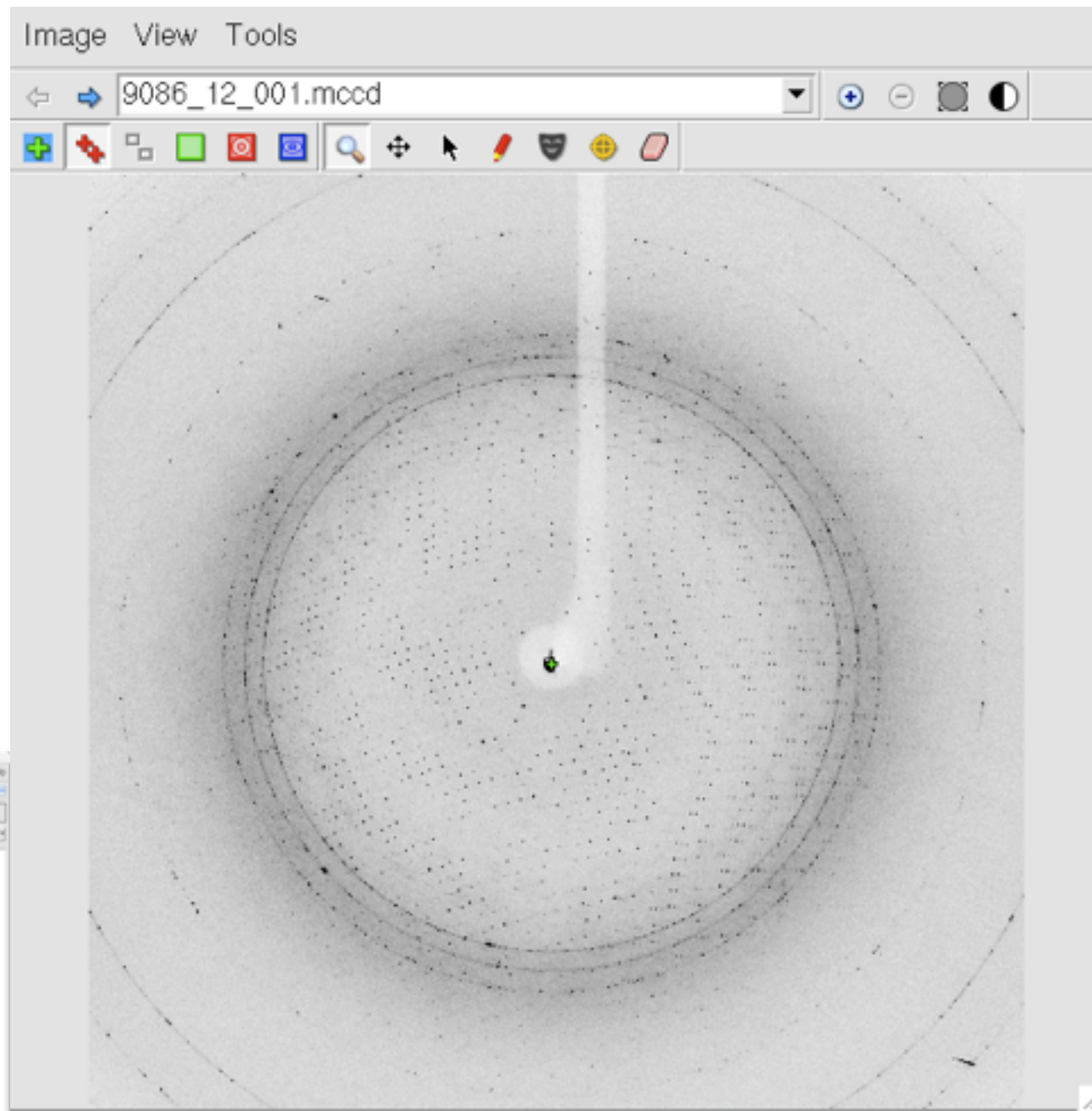
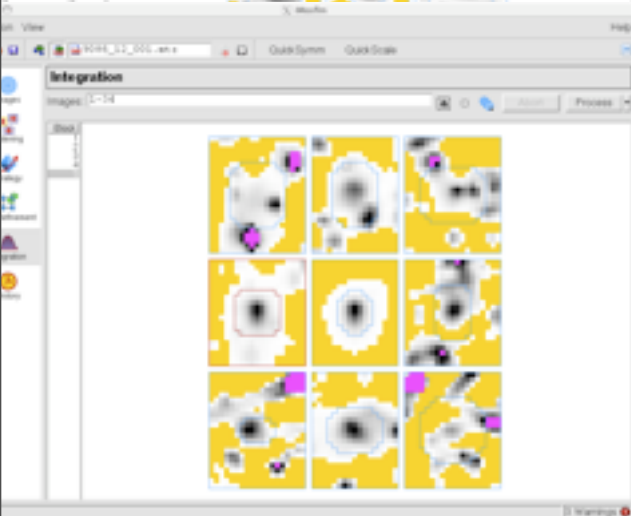
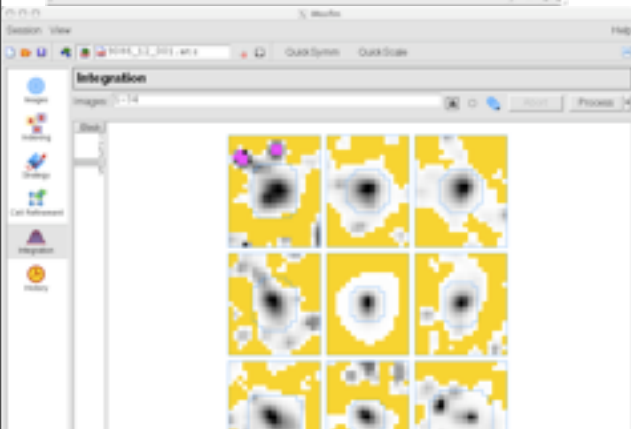
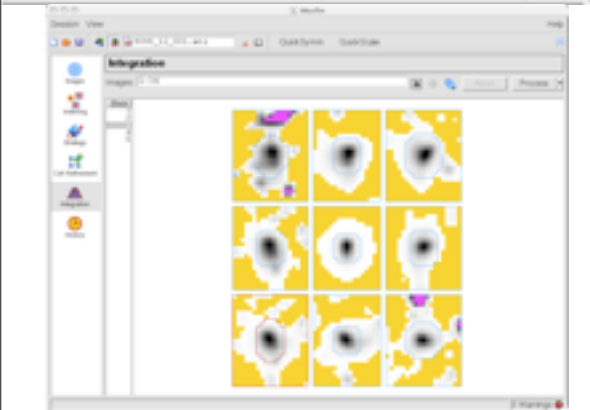
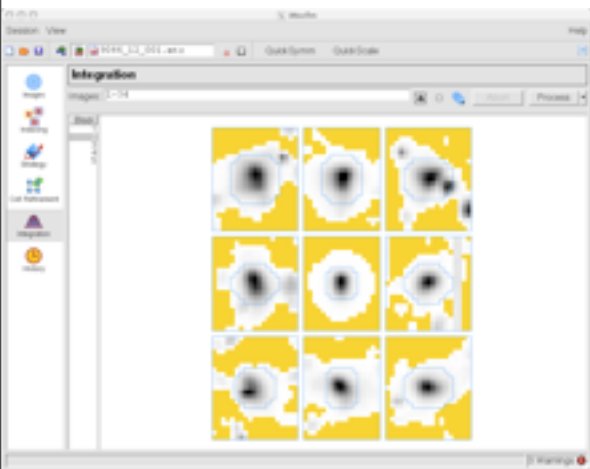


Cell parameters:

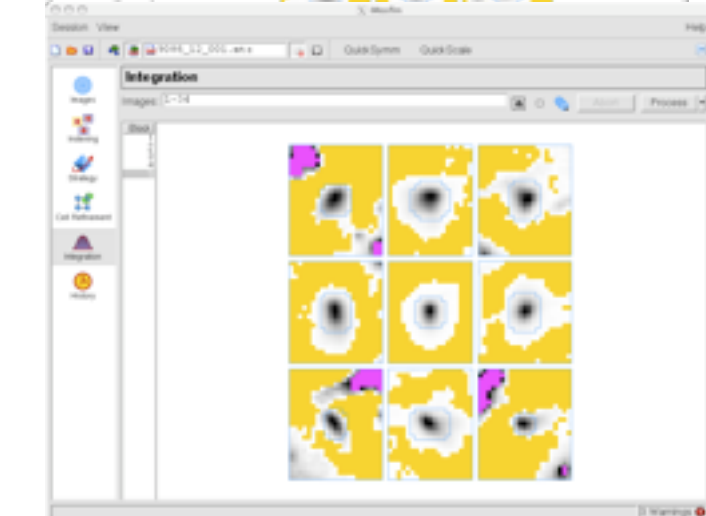
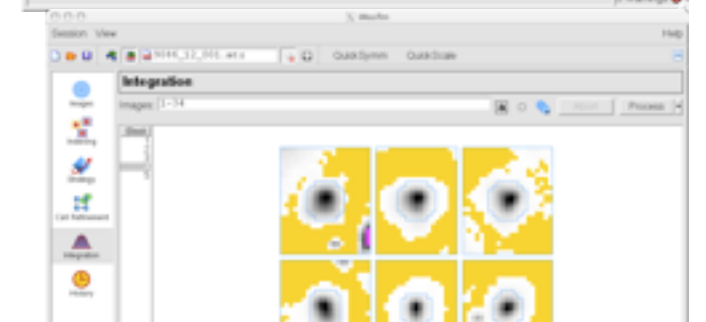
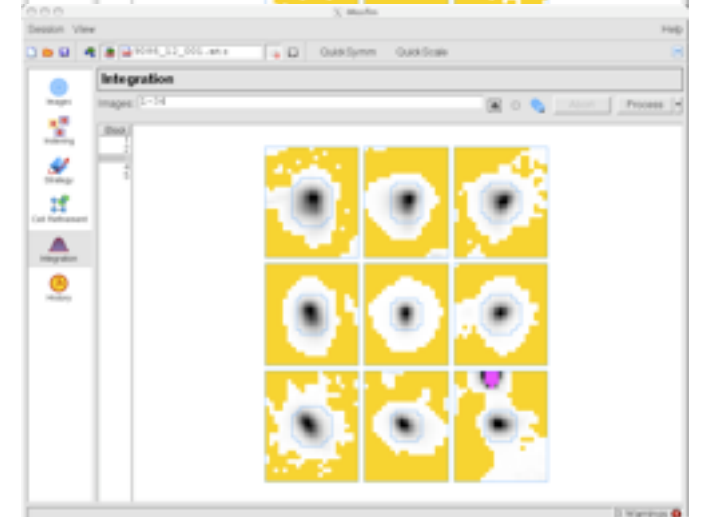
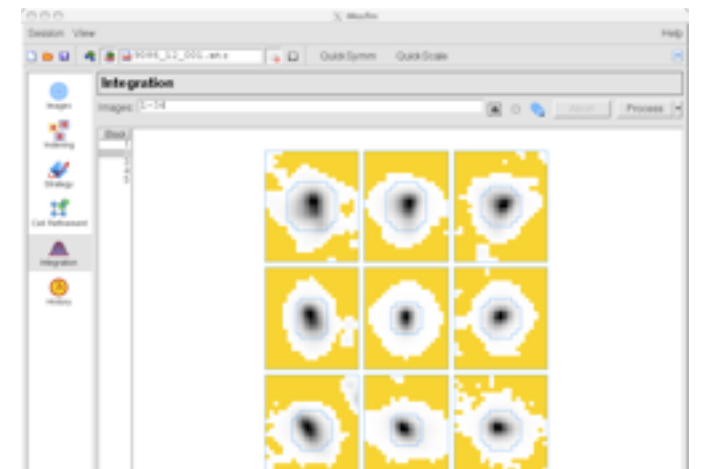
Initial $a=148.4$ $b=130.8$ $c=209.6$ $\beta=107.3$

Refined $a=147.8$ $b=130.8$ $c=210.4$ $\beta=108.0$

Icy Issues



Excluding spots in the resolution shells corresponding to the rings helps improve the profiles, but leads to lower completeness (eg 98->89% at 2.6 Å)



Acknowledgements



The Mosflm team

Andrew Leslie



Owen Johnson



Harry Powell



Geoff Battye*



Luke Kontogiannis*

** past team members*

Thanks to Andrew for some slides