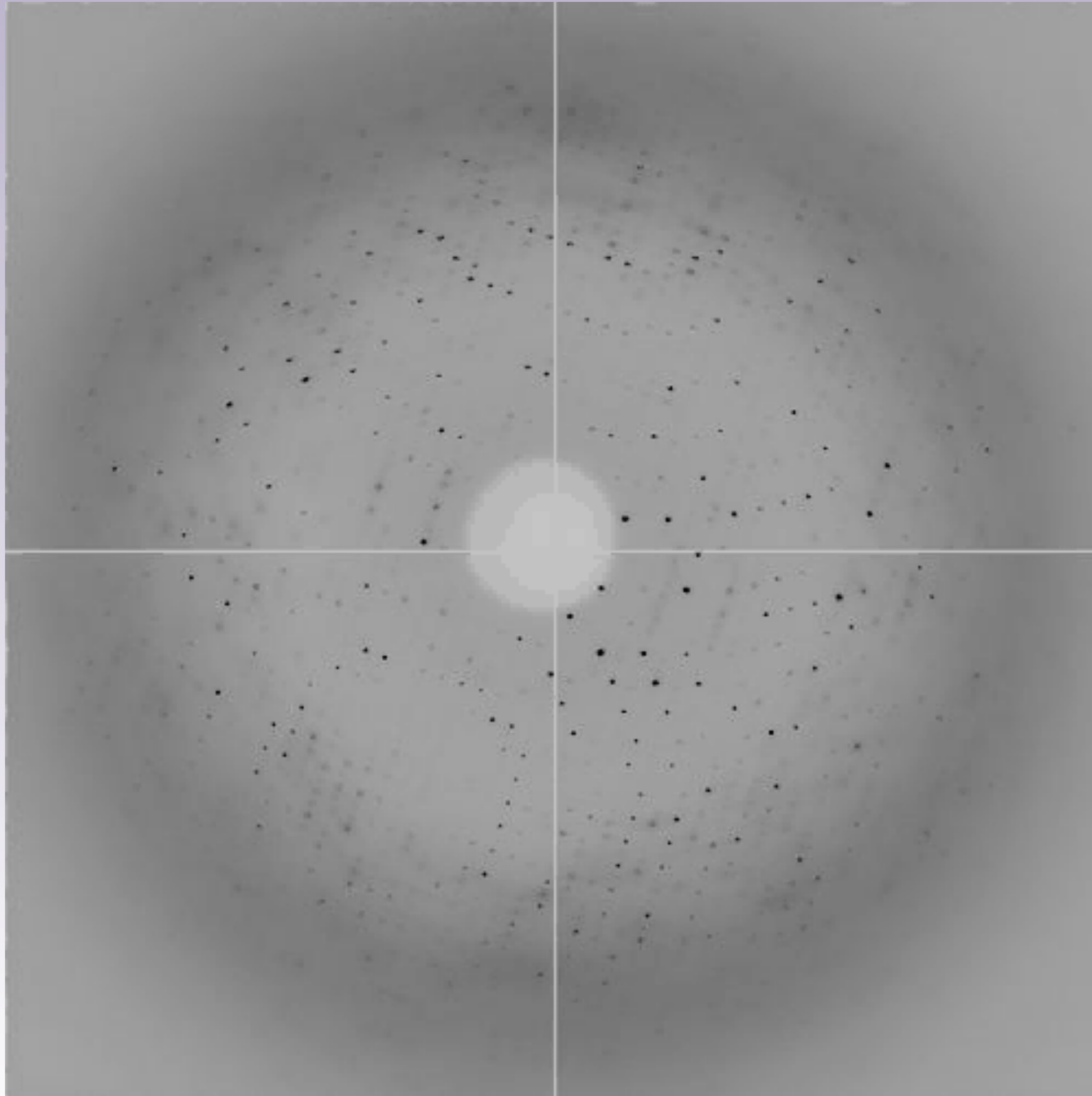


Data Collection Strategy and Processing (continued)

Phil Evans Tsukuba November 2014

*MRC Laboratory of Molecular Biology
Cambridge UK*



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.

Data collection strategy

Compromise between statistics (enough photons/reflection, and multiplicity) and radiation damage. **Radiation damage is the big problem.**

Radiation damage controls the total time available for crystal exposure.

Strategy: how to collect good data

Choices:

- a good crystal and index the pattern
- total rotation range
- start point
- image width
- exposure time
- detector position (maximum resolution)
- wavelength (for anomalous phasing)
- do you need multiple crystal positions or (isomorphous) crystals?

(some of these are linked)

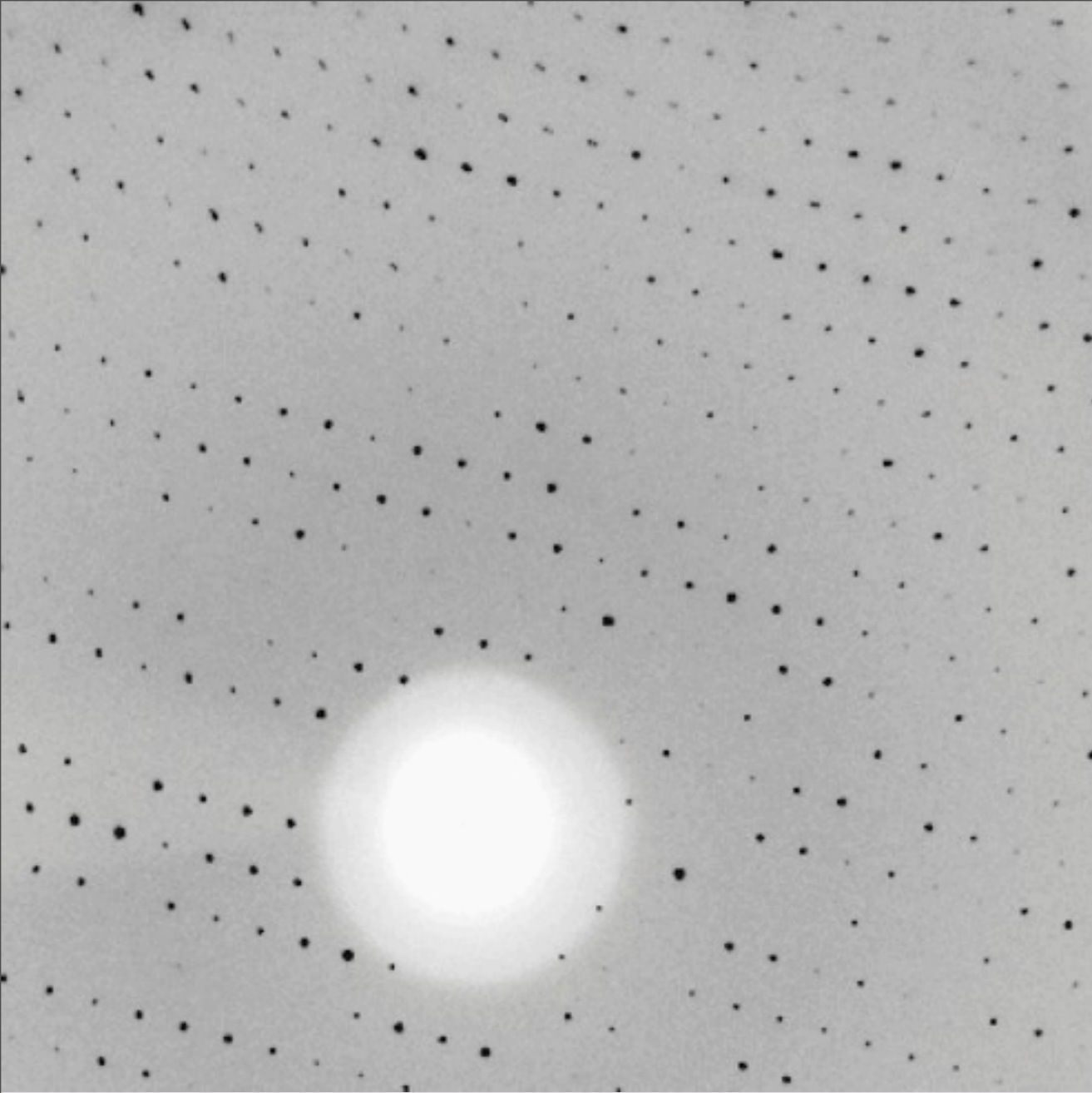
What is a good crystal?

- Single only one lattice
check by indexing pattern and looking for unpredicted spots
- Diffracts to high angle
- Low mosaicity
- Large *the diffracted intensity is proportional to the number of unit cells in the beam: not much gain for a crystal much larger than beam (100–200 μ m). Smaller crystals may freeze better (lower mosaicity)*
- Good freeze *no ice, minimum amount of liquid (low background)*
Optimise cryo procedure, and worry about crystal transfer procedures
- The best that you have! (the least worst)

Beware of pathological cases (twinning etc)

The quality of the crystal determines the quality of the dataset

You can get bad data from a good crystal,
but you can't get good data from a bad crystal

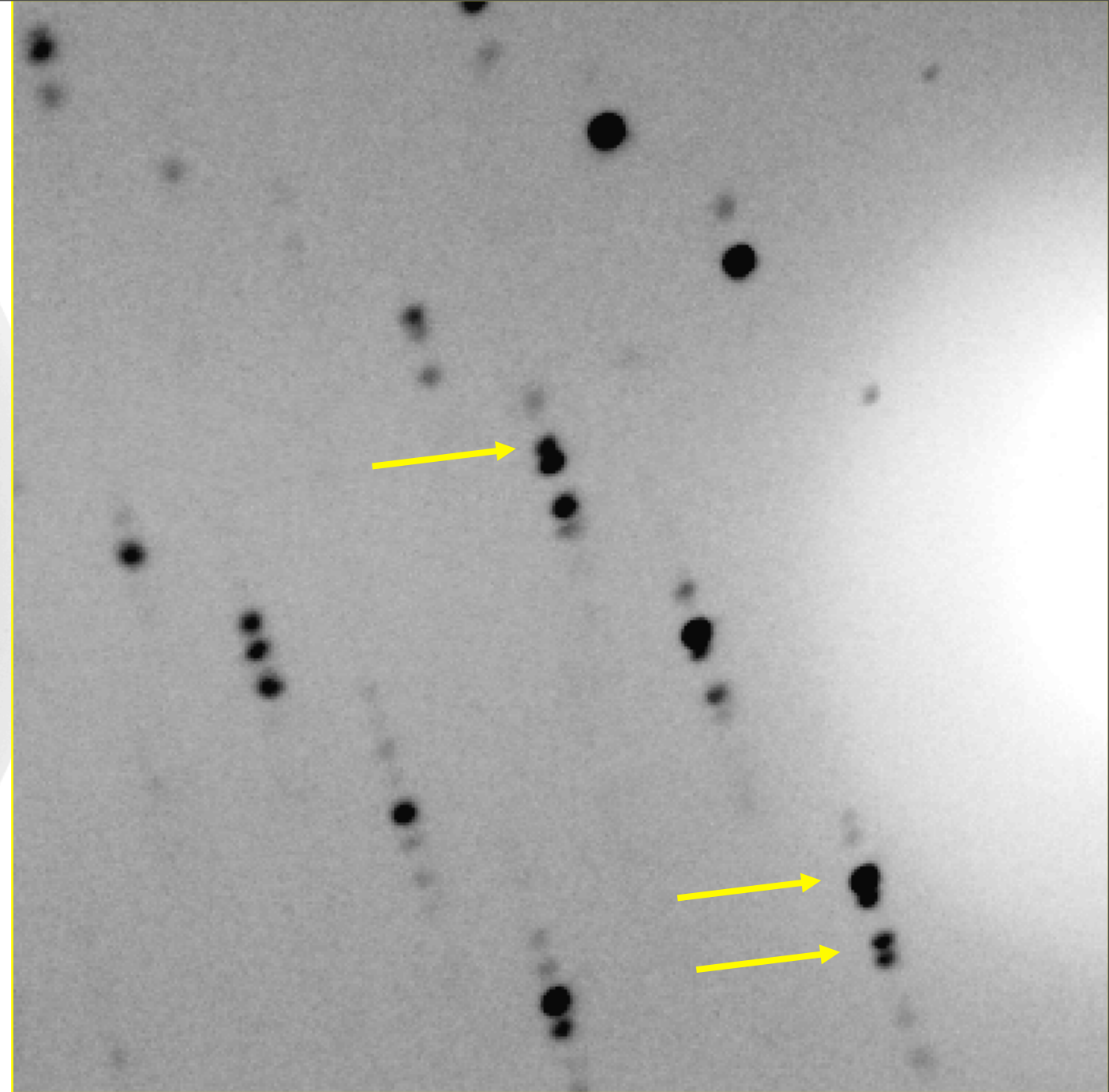
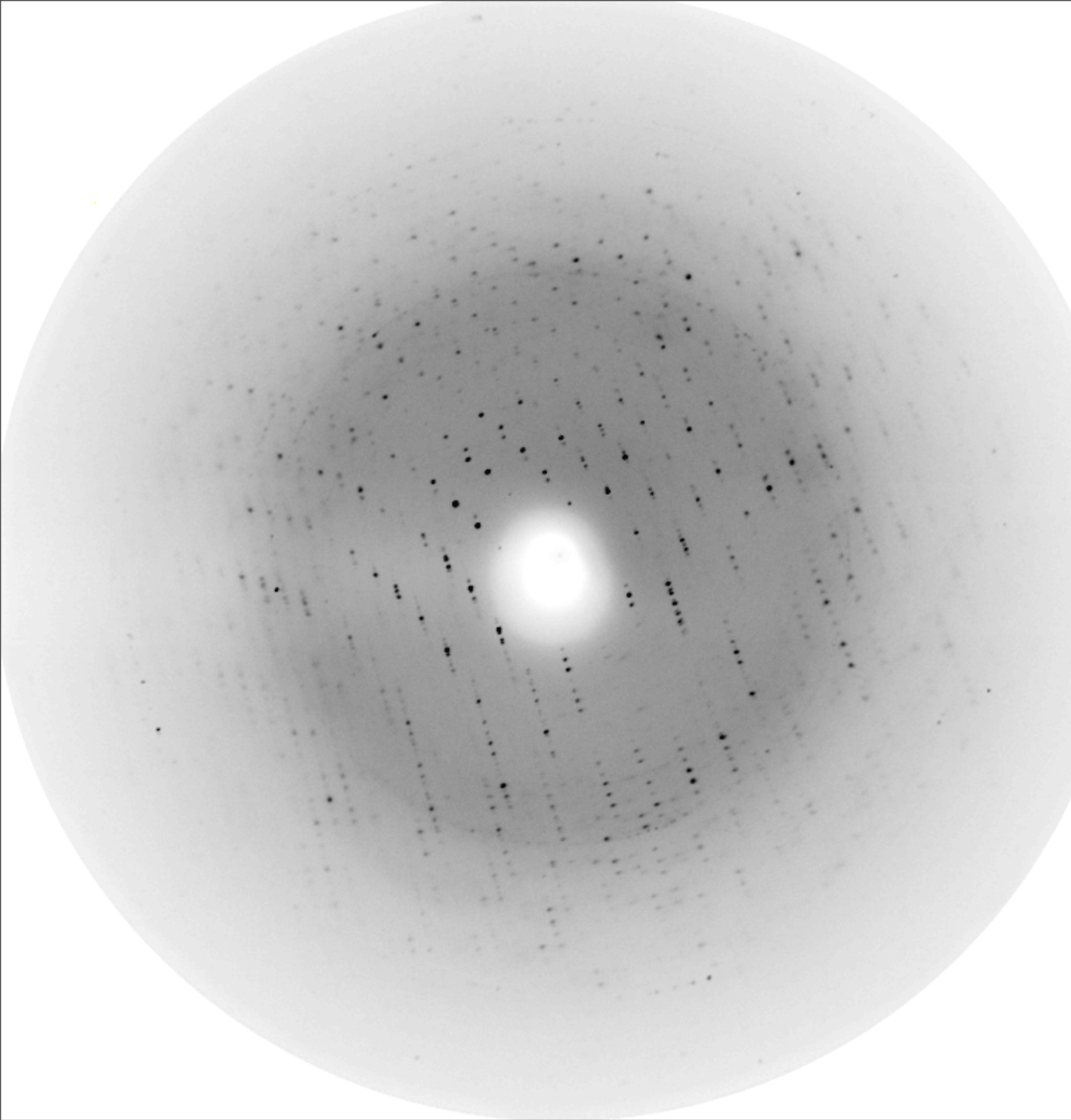


$\Phi = 0^\circ$

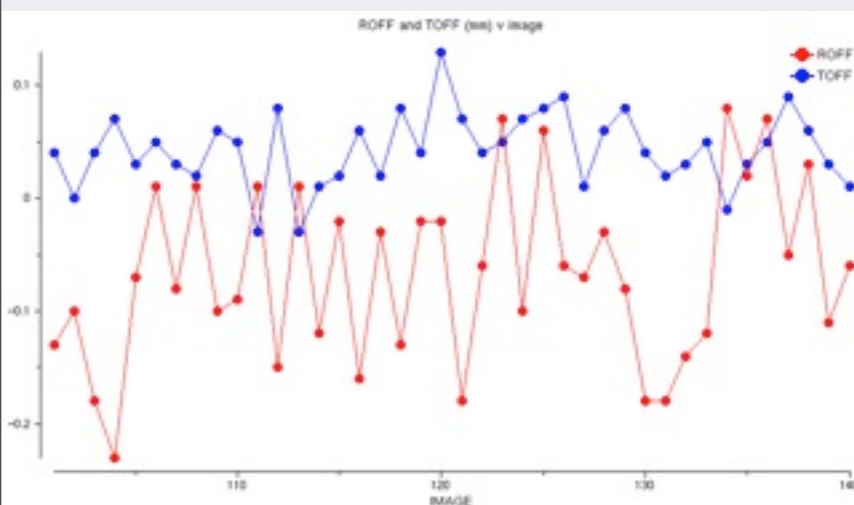
$\Phi = 90^\circ$

Always check diffraction in two
orthogonal images !

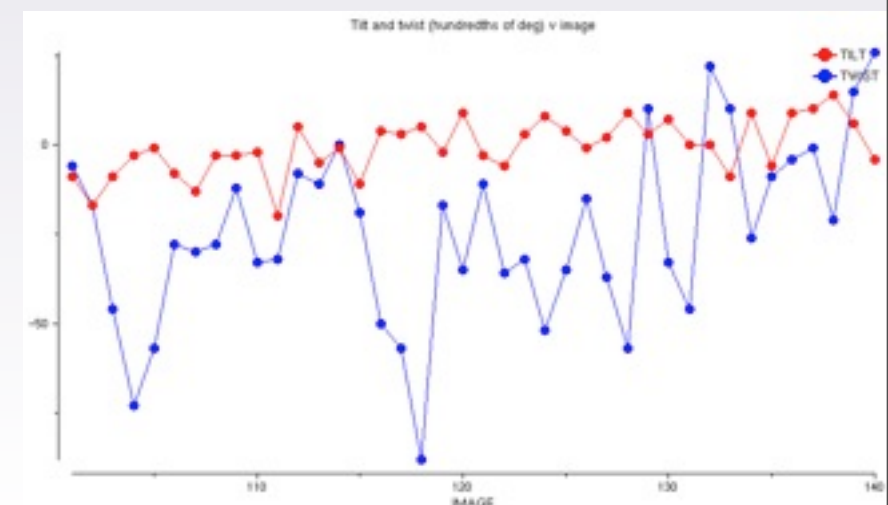




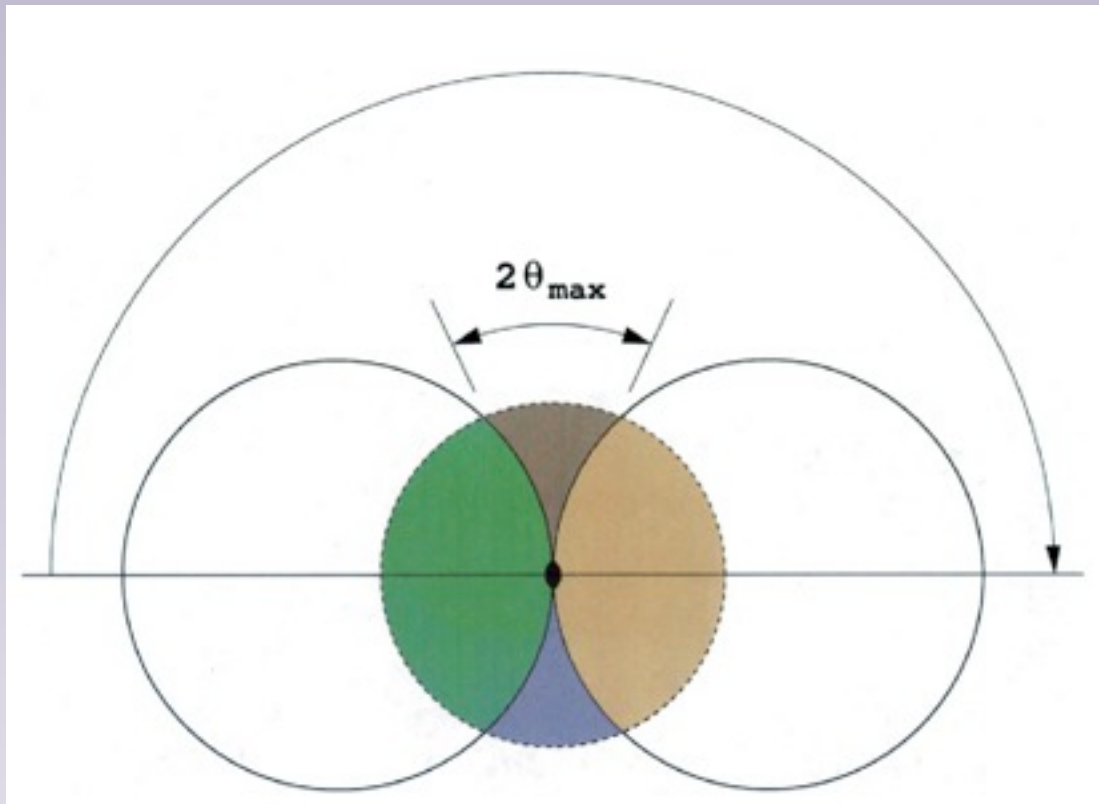
Additional spots present, not resolved.



Results in instability in refinement of detector parameters.



Completeness: total rotation range and the blind region

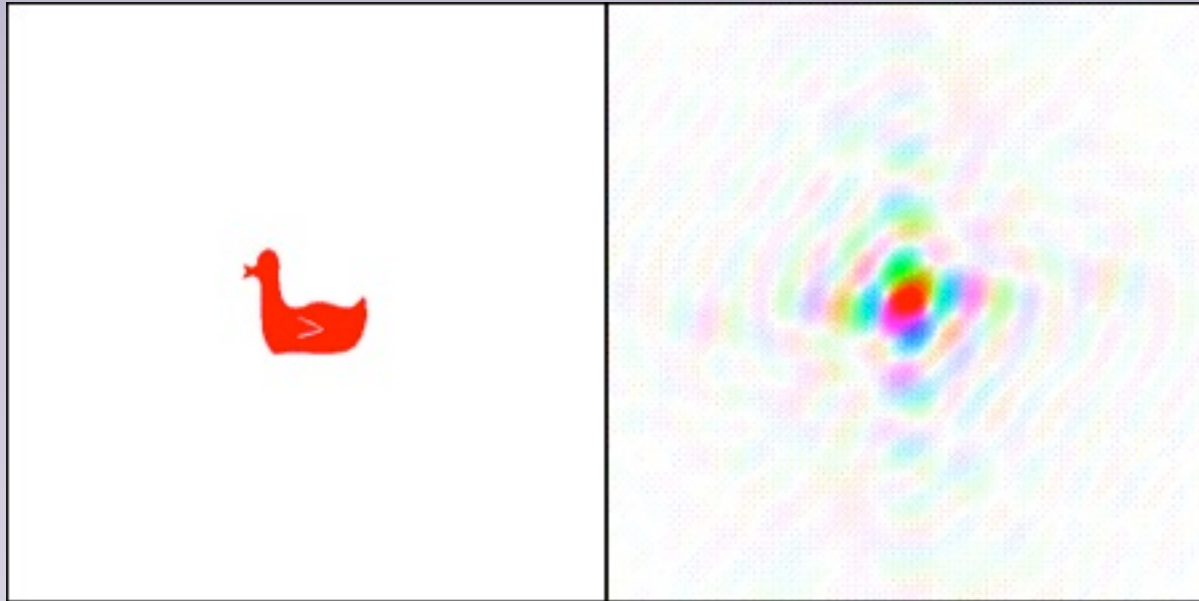


To use the Ewald sphere construction to understand which parts of reciprocal space are measured, it is easier to fix the “resolution sphere” of all reciprocal lattice points within a maximum resolution, and to rotate the Ewald sphere. The region collected is the volume swept out by the leading and trailing surfaces of the sphere

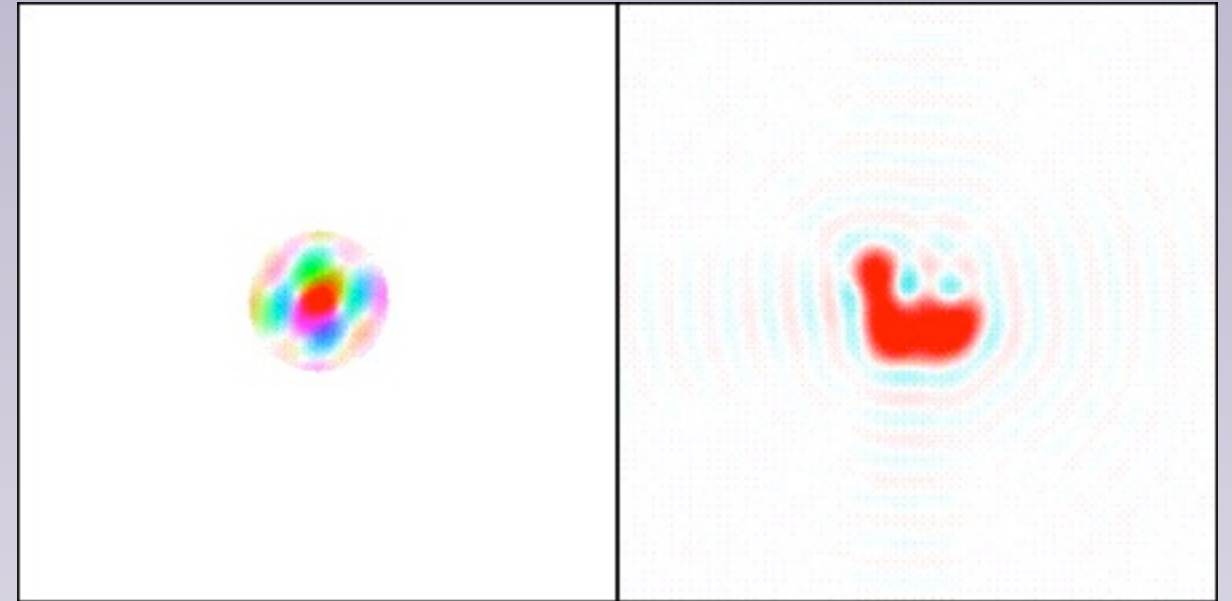
In a rotation of 180° above, the lower boundary of the initial sphere sweeps out the volume coloured green & the upper boundary the light brown part. The dark brown part is measured twice, and the blue part not at all

Because of Friedel’s law, this dataset is complete (apart from the *blind region*), but if complete anomalous differences are required, then $180^\circ + 2\theta_{\max}$ is required (unless there is symmetry)

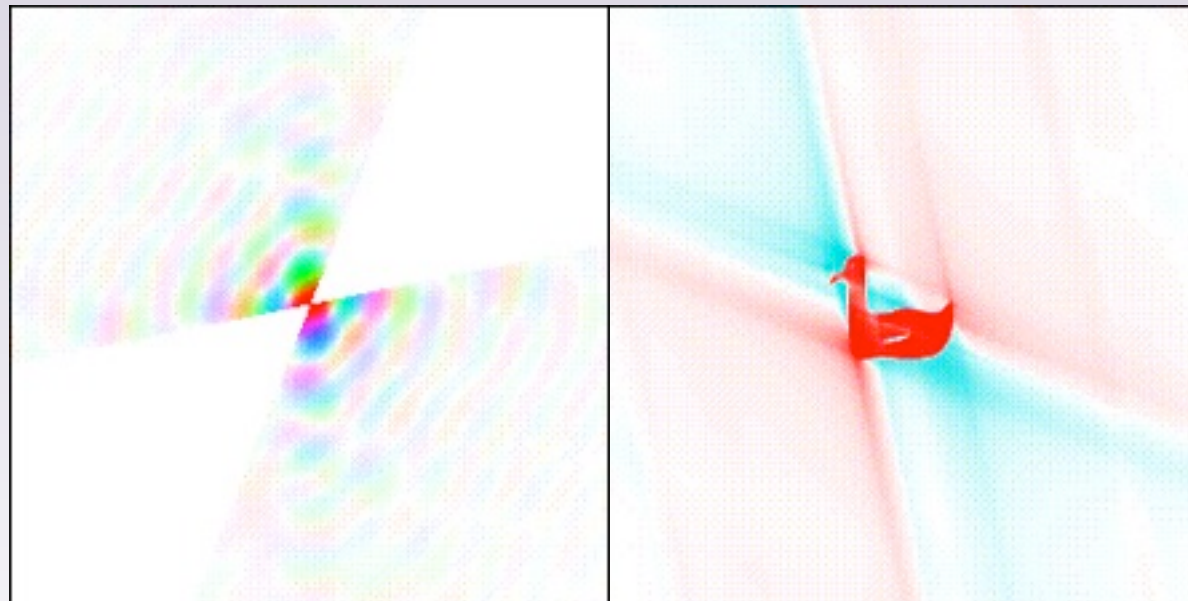
The importance of data completeness



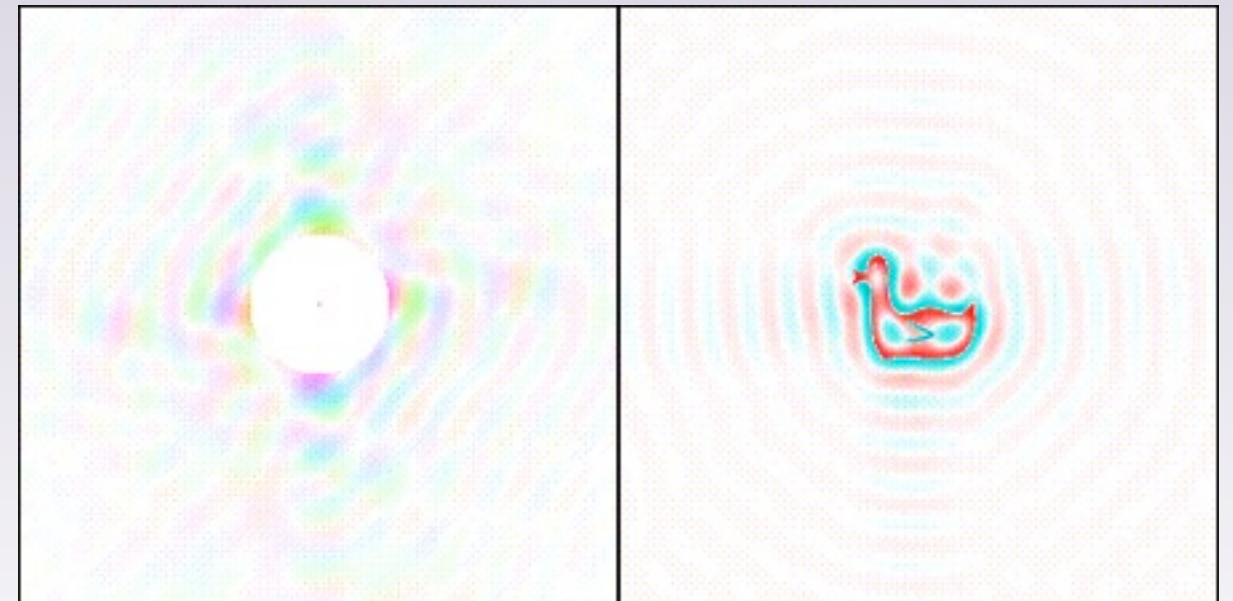
A duckand its Fourier transform



A low-resolution duck



Incomplete data: missing wedge



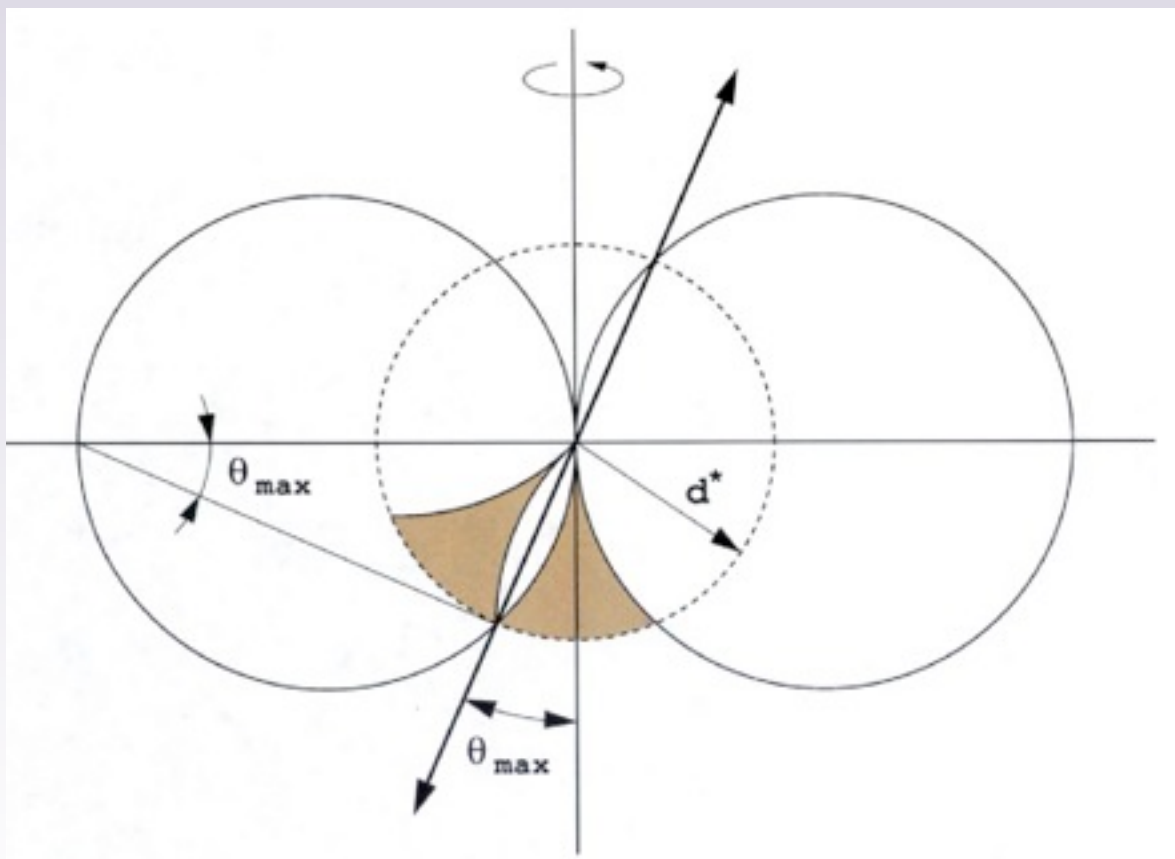
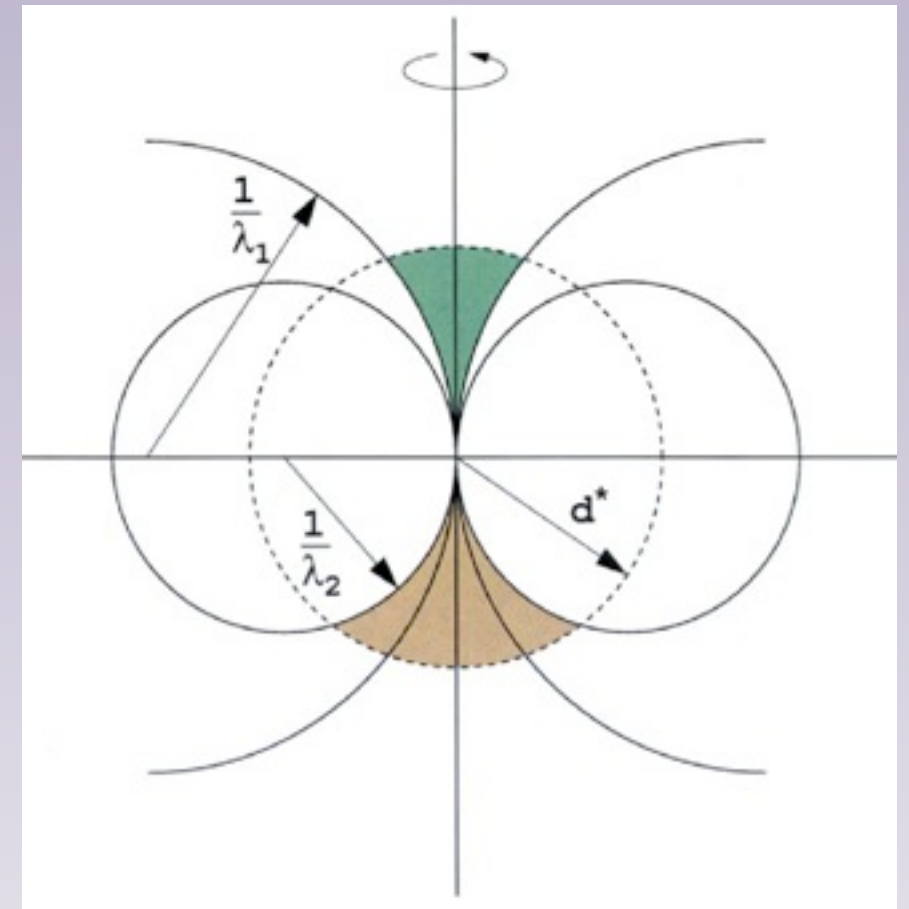
A duck without low-resolution reflections

from Kevin Cowtan's "Book of Fourier"

The blind region

Diffraction vectors close to the rotation axis will never pass through the sphere, even in a 360° rotation

The blind region is smaller for short wavelengths, as the Ewald sphere is flatter

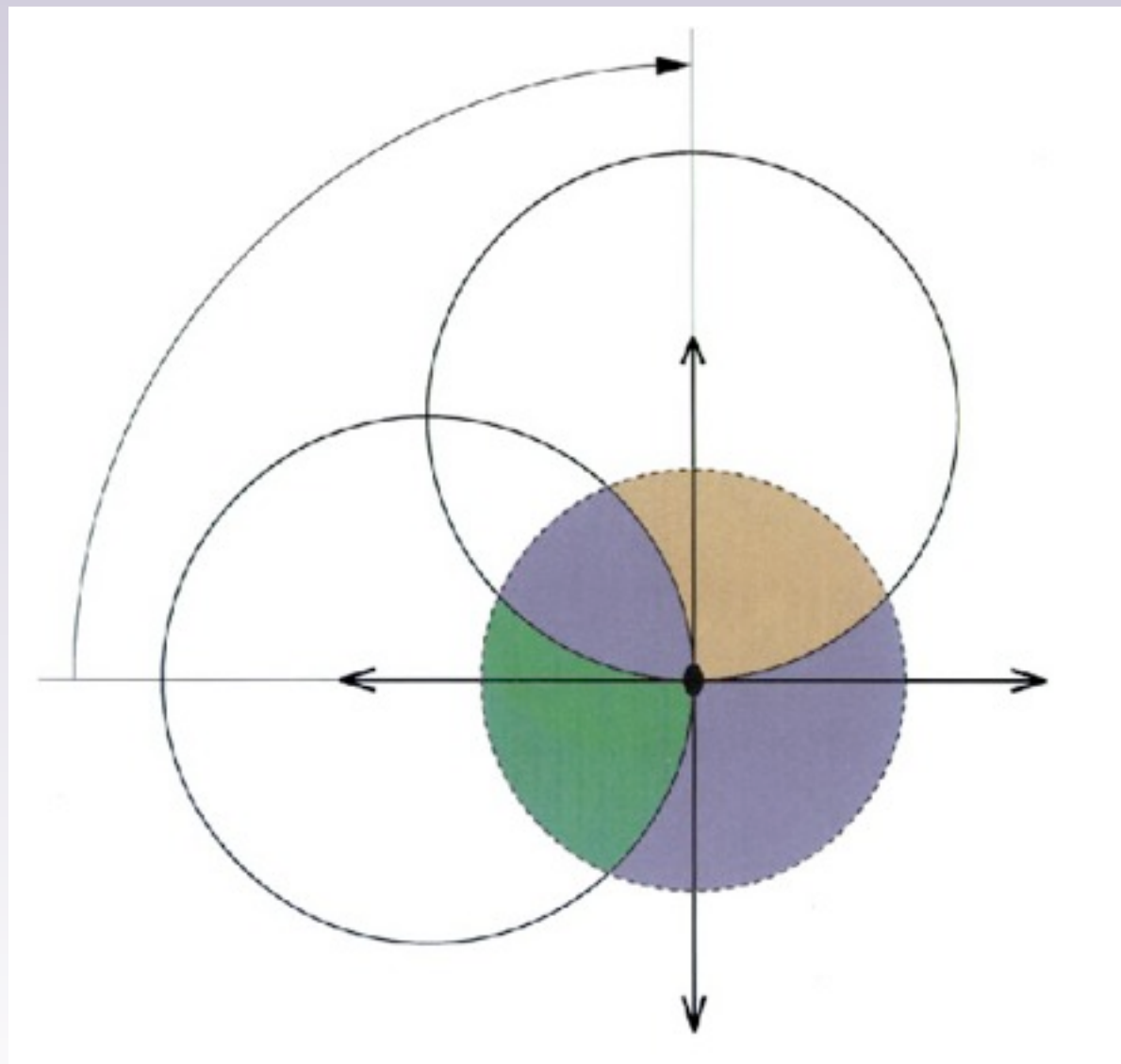


The blind region may be filled in by collecting a second set of data, offsetting the crystal by at least $\theta_{\max}$ or by symmetry (except in PI)

If there is symmetry, offsetting from an axis can remove or reduce the blind region for a single setting

Symmetry and total rotation range: an orthorhombic example

Rotation of an orthorhombic crystal by 90° starting from an axis gives a complete dataset (except for the blind region)



A 90° rotation starting at a diagonal collects the same 45° twice, and gives incomplete data

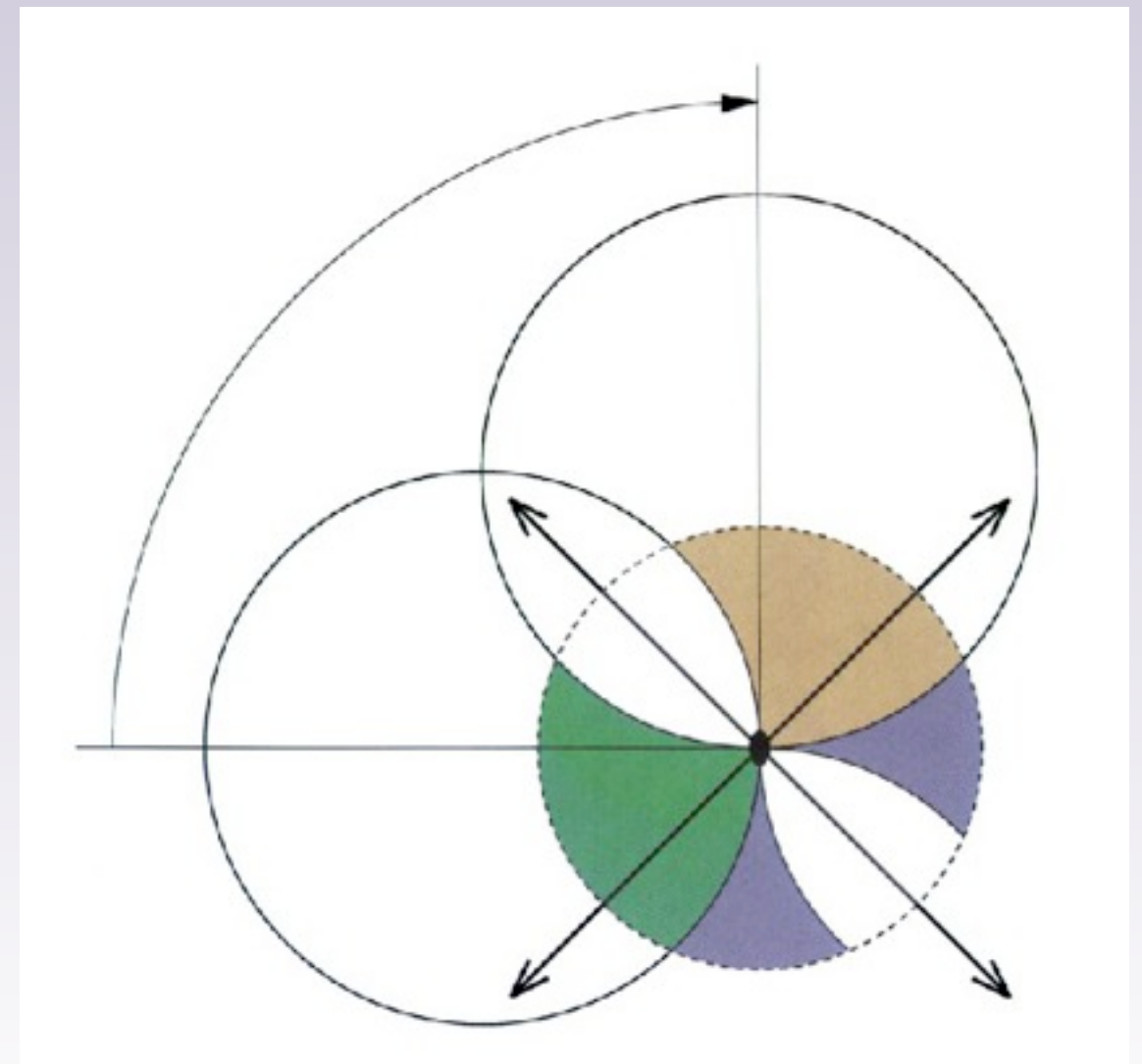


Image width

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°

For shutterless data collection, the “optimum” image width is $\sim 1/2$ XDS width σ_ϕ , or $\sim 1/5$ Mosflm “mosaicity”

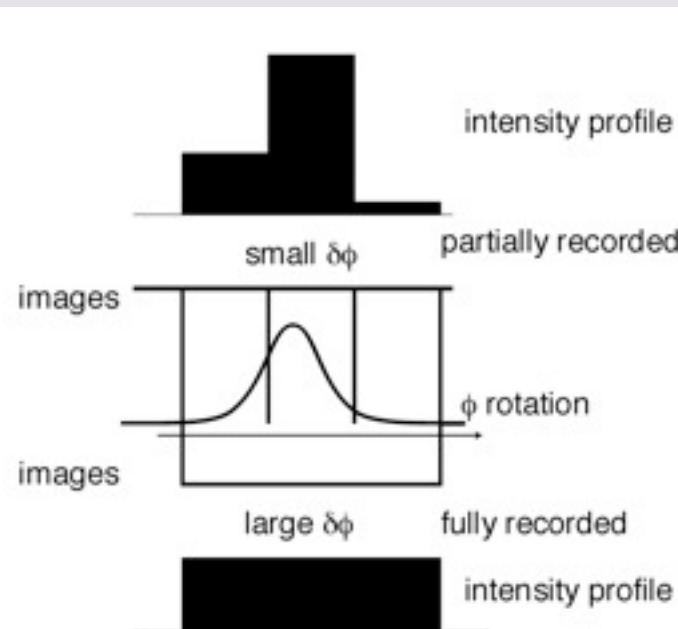
Mueller, Wang & Schulze-Bries, 2011.
Acta Cryst (2012). D68, 42-56

1° is always wrong

(almost always)

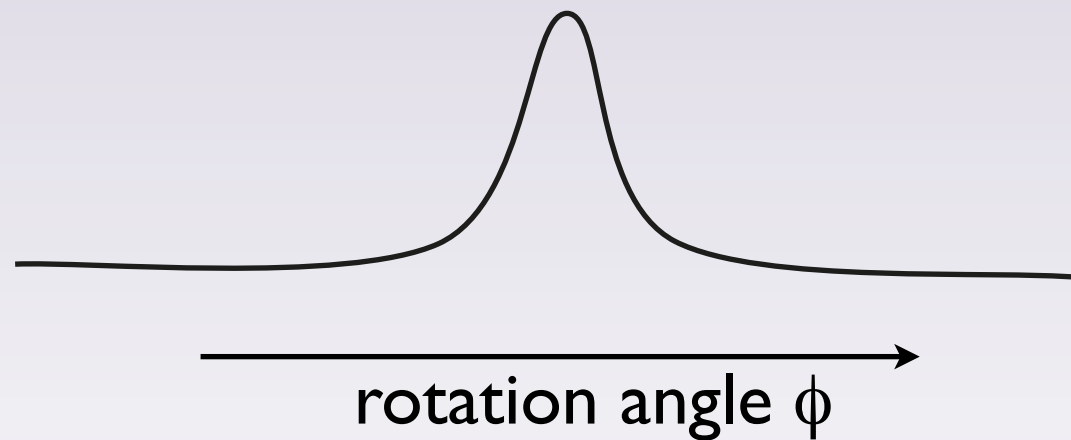
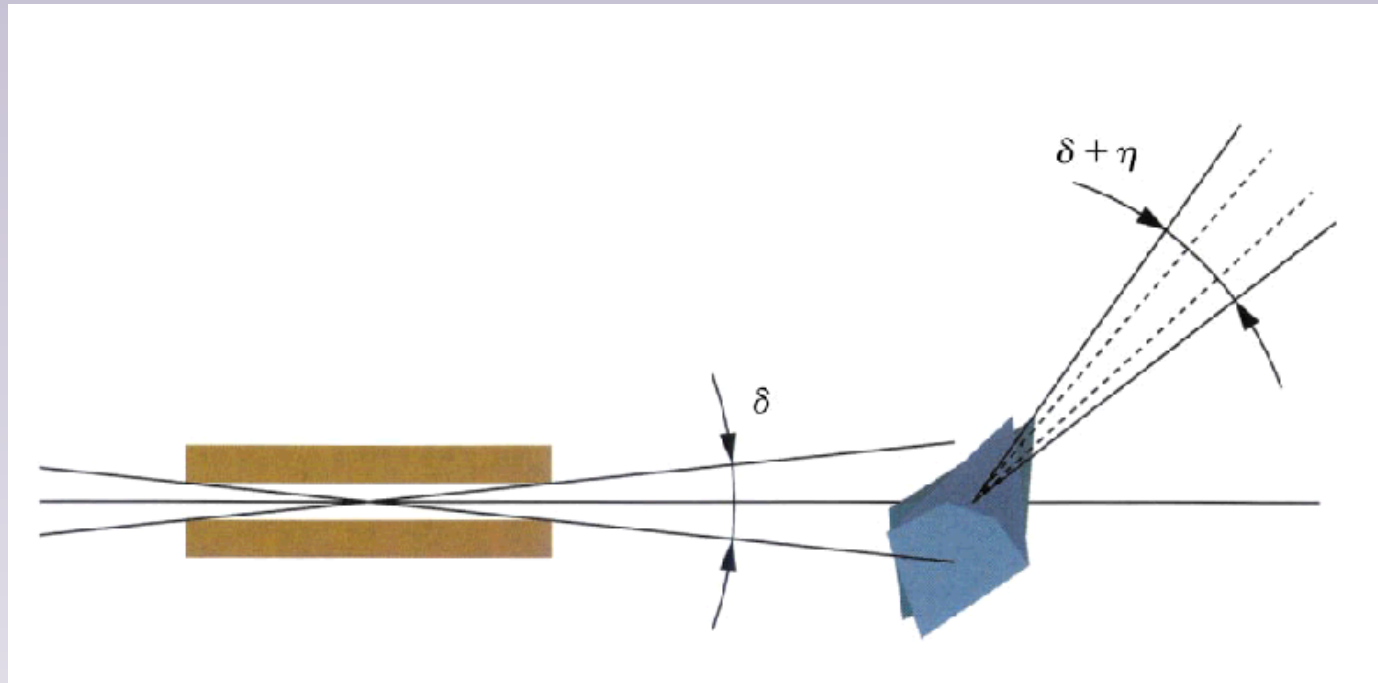
use $0.1^\circ - 0.5^\circ$

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*)



Reflection width

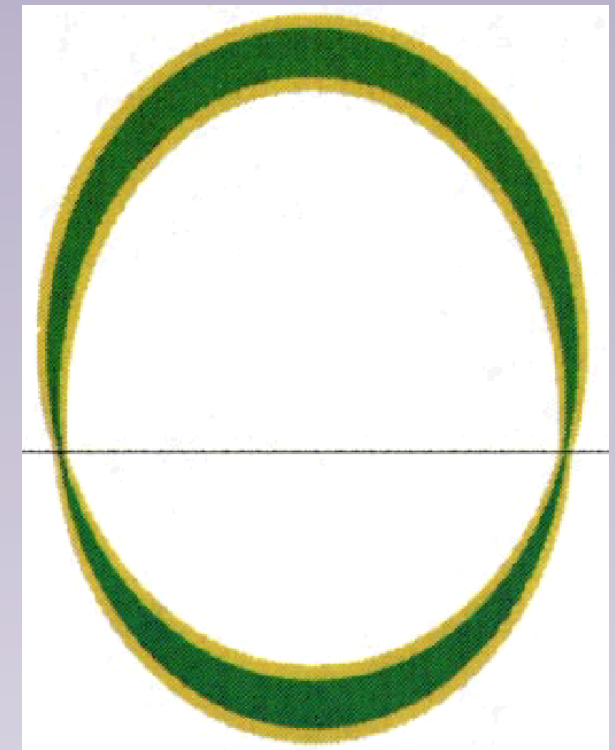
Beam divergence δ and mosaicity η add up to increase the angular width of the diffracted beam



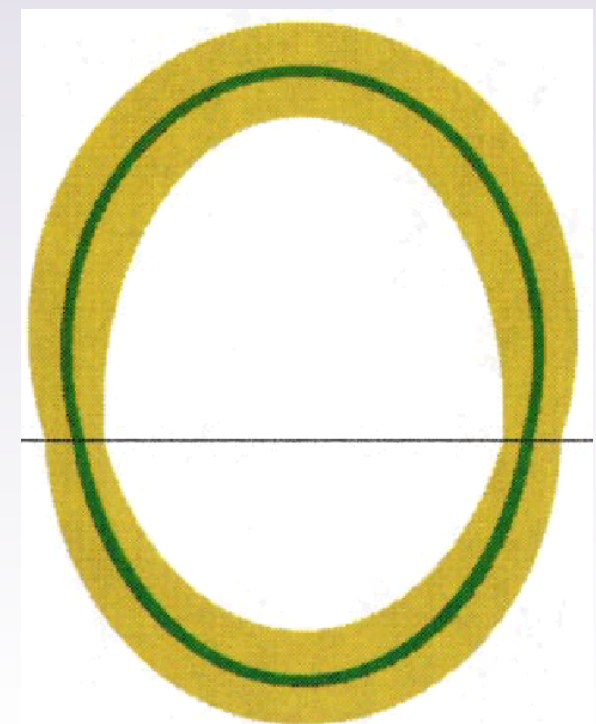
Reflection width in rotation

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

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



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

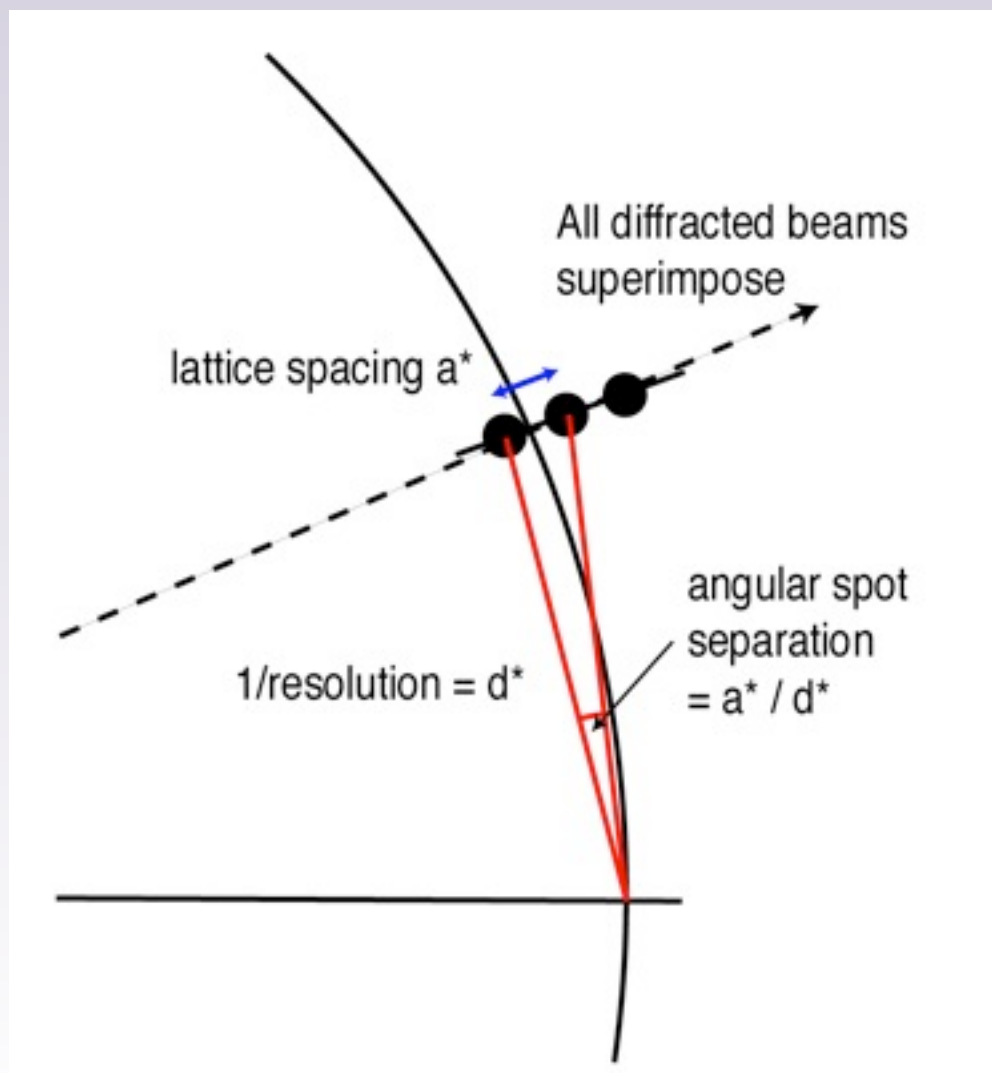


Overlaps and rotation range

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

Maximum slice width = $(a^*/d^*) - w$

$$= d/a - w$$

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

eg cell = 200Å, resolution = 2Å, width = 0.3°

Maximum Slice = 0.27°

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

Two Cases:

Anomalous scattering, MAD

High multiplicity is better than long exposures (eliminates outliers)

Split time between all wavelengths, be cautious about radiation damage, reduce time & thus resolution if necessary

Collect Bijvoet pairs close(ish) together in time: align along dyad or collect inverse-beam images

Maybe recollect first part of data at end to assess radiation damage

Data for refinement

Maximise resolution: longer exposure time (but still beware of radiation damage)

High multiplicity less important, but still useful

Use two (or more) passes with different exposure times (ratio ~ 10) if necessary to extend range of intensities (high & low resolution)

Short wavelength ($< 1 \text{ \AA}$) to minimise absorption

Collect symmetry mates at different times and in different geometries, to get best average (even with higher R_{merge} !). Rotate about different axes.

Summary of strategy choices

Detector position

- Place detector far enough away to resolve spots (or reduce beam size)
- Use the whole detector area (don't have blank region around edge)
- Don't use an offset detector unless desperate for spot resolution. If you have to offset the detector, be **very careful** in strategy planning

Total rotation range

- If possible, collect 180° (360° in P1 with anomalous). High multiplicity is *Good*, provided that radiation damage is not serious
- When rotating around (or close to) a symmetry axis of order n , the minimum rotation is about $180^\circ/n$ for 2- & 4-folds, $360^\circ/n$ for 3- & 6-folds (more complicated in dihedral or cubic symmetry)
- With an offset detector, a larger rotation range is needed, as only one surface of the Ewald sphere is active rather than two

Crystal orientation and rotation start point

- To remove the blind region, avoid rotating exactly around a symmetry axis
- To optimise anomalous differences with respect to absorption, rotate exactly around a symmetry axis (even-fold only)
- Use a strategy program (eg MOSFLM) to determine range and start point
- Collect 180° or 360° and start anywhere

Exposure time/image

- Compromise between high resolution (good) and radiation damage (bad)
- Process & assess first data collection, then revise strategy
- Use estimates from eg EDNA/BEST
- NB the effective exposure is set by the rotation **rate**
ie $0.2 \text{ sec}/0.1^\circ$ is equivalent to $0.5 \text{ sec}/0.5^\circ$

Image rotation range (slicing)

- Use a strategy program to determine optimum width
- **1° is (almost) always wrong! Typical values are 0.1° – 0.5°**
- Set width $< (\text{maximum resolution})/(\text{longest axis not along spindle}) - \text{spotwidth}$
- Process data & check for overlaps
- Fine-slicing is more sensitive to errors in synchronisation of shutter opening and rotation (this potentially adds an error for each image)
Not a problem for shutterless continuous collection with fast read-out detector

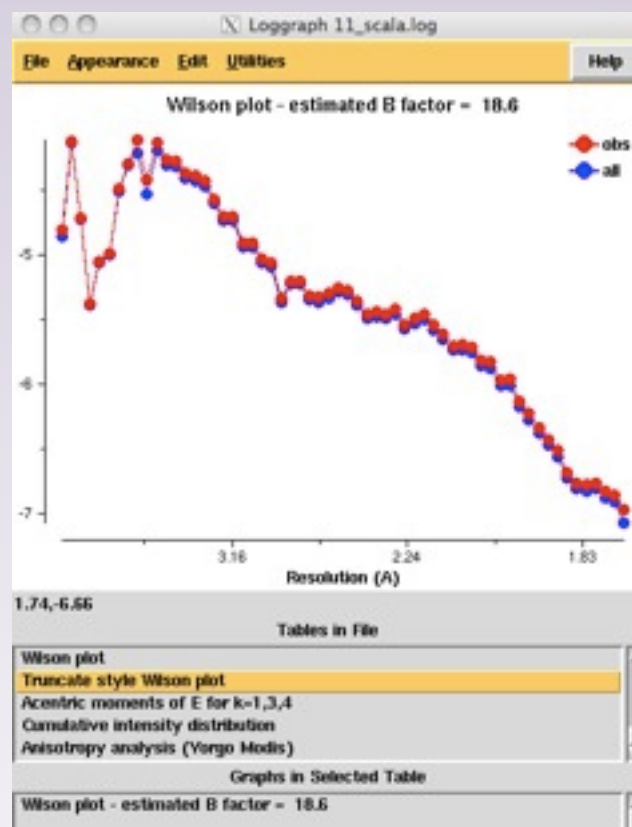
Intensity statistics

We need to look at the distribution of intensities to detect twinning

Assuming atoms are randomly placed in the unit cell, then

$$\langle I \rangle(s) = \langle F F^* \rangle(s) = \sum_j g(j, s)^2$$

where $g(j, s)$ is the scattering from atom j at $s = \sin\theta/\lambda$



Average intensity falls off with resolution, mainly because of atomic motions (B-factors)

For the purposes of looking for crystal pathologies, we are not interested in the variation with resolution, so we can use “normalised” intensities which are independent of resolution

$$\langle I \rangle(s) = C \exp(-2 B s^2)$$

Wilson plot: $\log(\langle I \rangle(s))$ vs s^2

This would be a straight line if all the atoms had the same B-factor

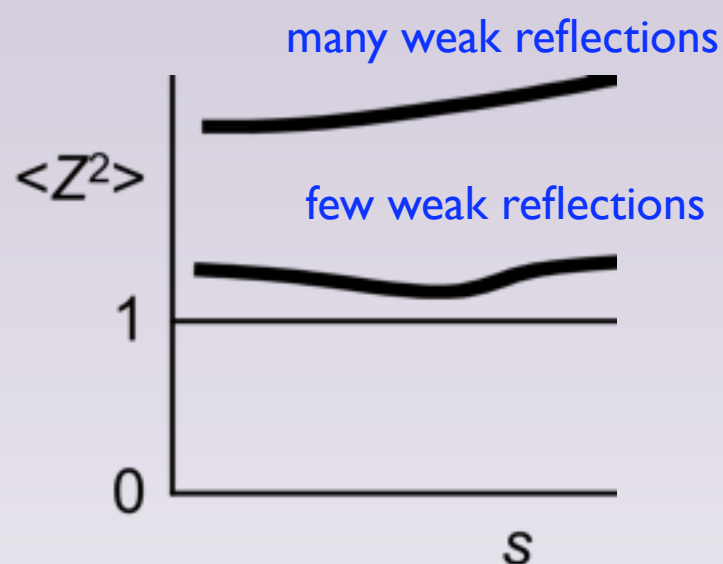
Normalised intensities: relative to average intensity at that resolution

$$Z(h) = I(h)/\langle I(s) \rangle \approx |E|^2$$

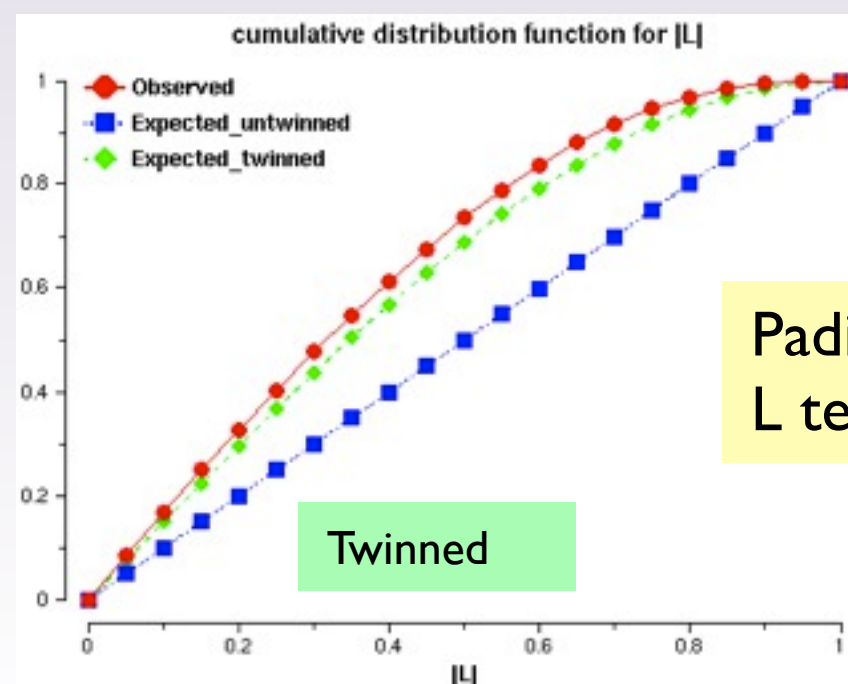
$$\langle Z(s) \rangle = 1.0 \text{ by definition}$$

$$\langle Z^2(s) \rangle > 1.0 \text{ depending on the distribution}$$

$\langle Z^2(s) \rangle$ is larger if the distribution of intensities is wider: it is the 2nd moment
ie the *variance* (this is the 4th moment of E)

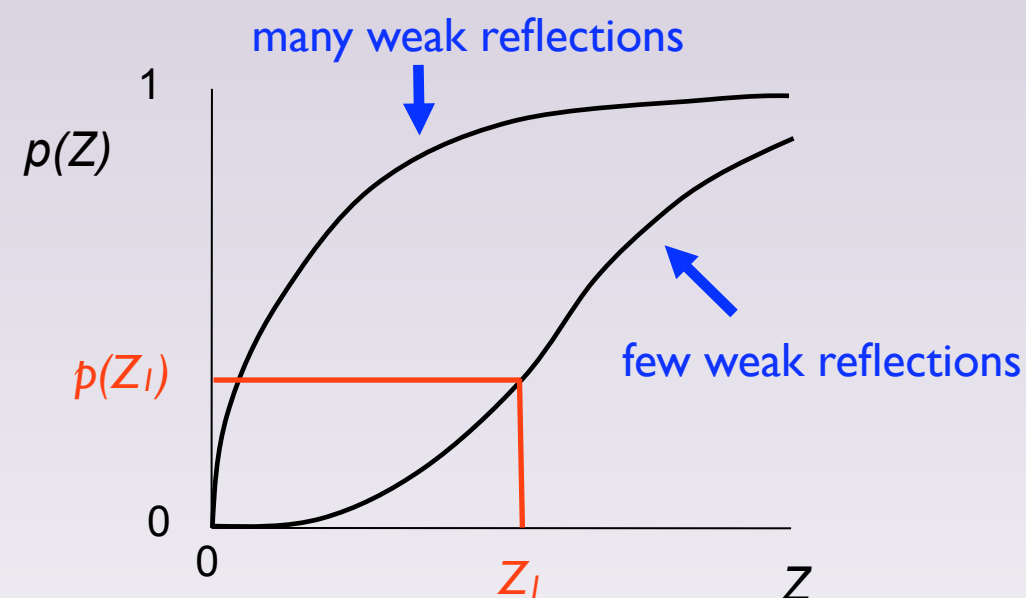


Moments



Padilla – Yeates
L test

Cumulative distribution of Z : $p(Z)$ vs. Z



$p(Z_1)$ is the proportion of
reflections with $Z < Z_1$

Other features of the intensity distribution which may obscure or mimic twinning

Translational non-crystallographic symmetry:

whole classes of reflections may be weak

eg h odd with a NCS translation of $\sim 1/2, 0, 0$

$\langle I \rangle$ over all reflections is misleading, so Z values are inappropriate

The reflection classes should be separated (not yet done)

Anisotropy: $\langle I \rangle$ is misleading so Z values are wrong

ctruncate applies an anisotropic scaling before analysis

Overlapping spots: a strong reflection can inflate the value of a weak neighbour, leading to too few weak reflections

this mimics the effect of twinning