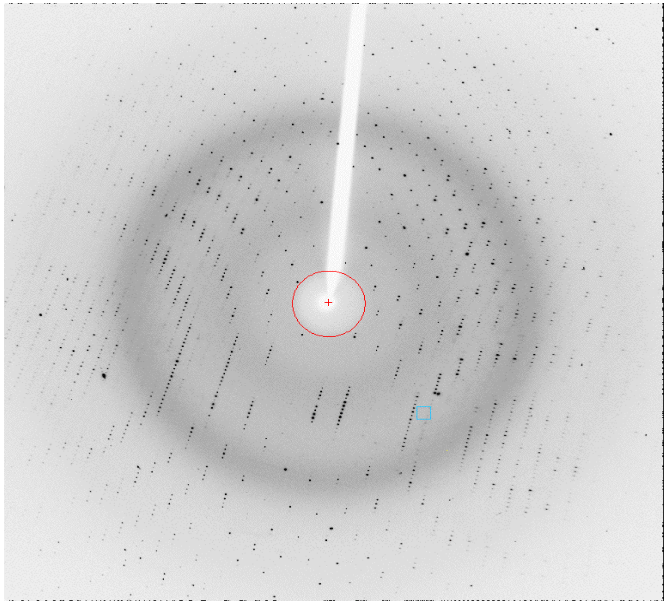
CCP4-BGU workshop 2020

The X-ray Diffraction Experiment

Diffraction Geometry and Data Collection Strategy



Ed Lowe, University of Oxford

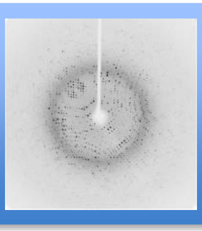
Department of Biochemistry

UNIVERSITY OF OXFORD

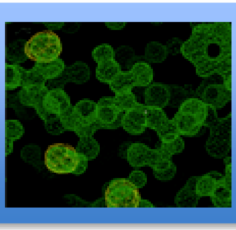
Processes in Crystallography

Crystallisation

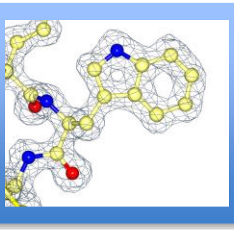
Data Collection



Map Calculation



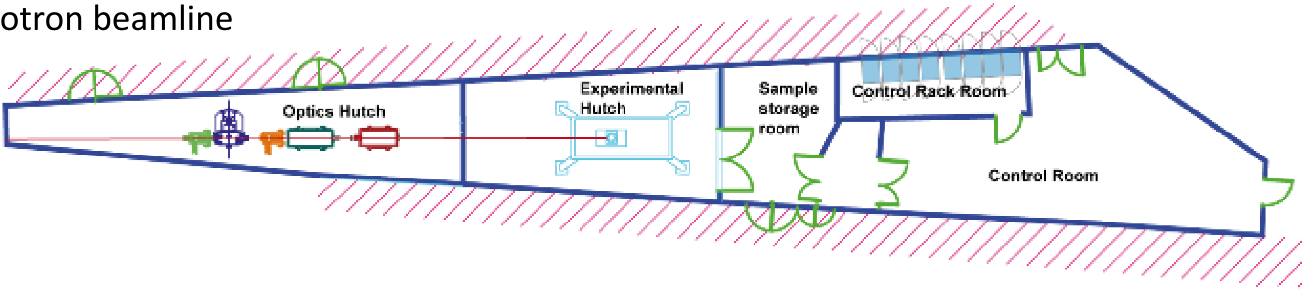
Map Interpretation

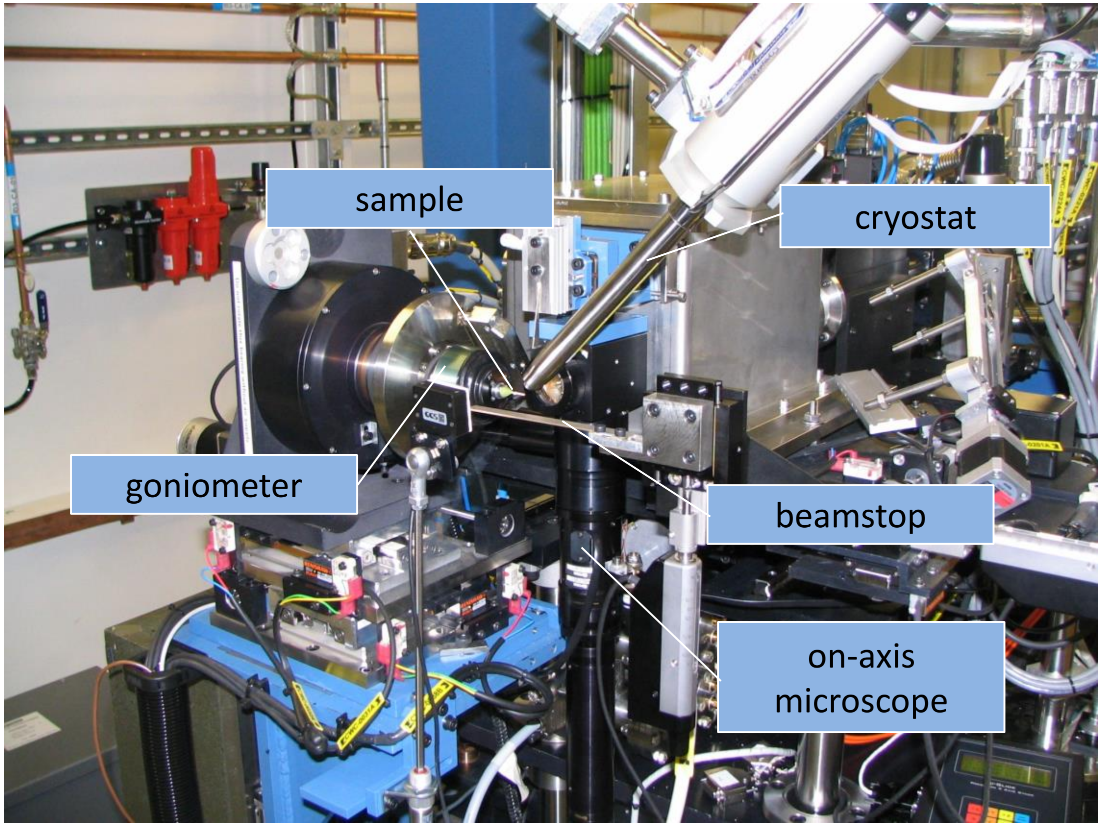


Validate and Publish Structure

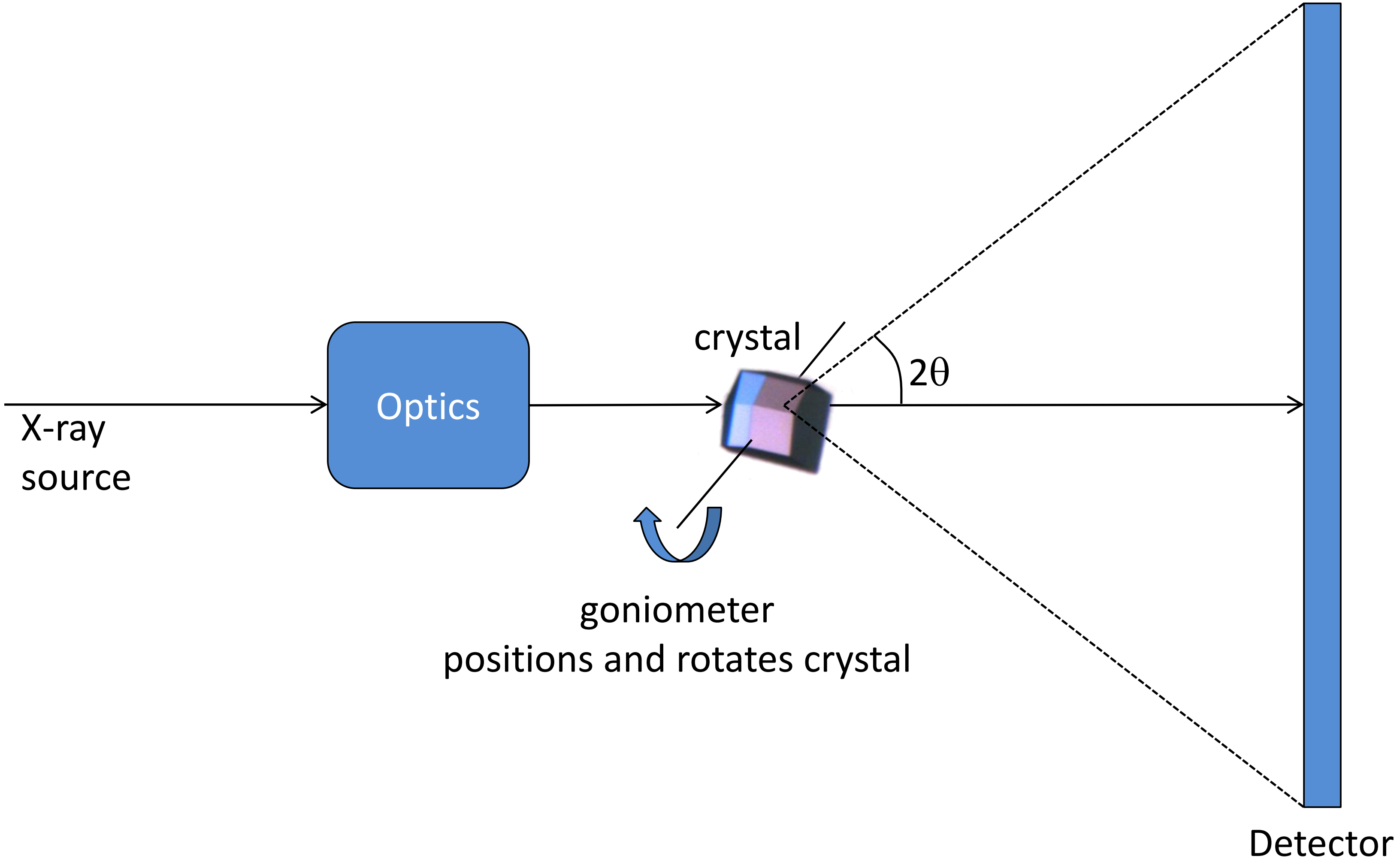
Collecting Diffraction Data

Layout of a synchrotron beamline
(Diamond I03)





Data Collection



Definition of a crystal and the unit cell

Crystal:

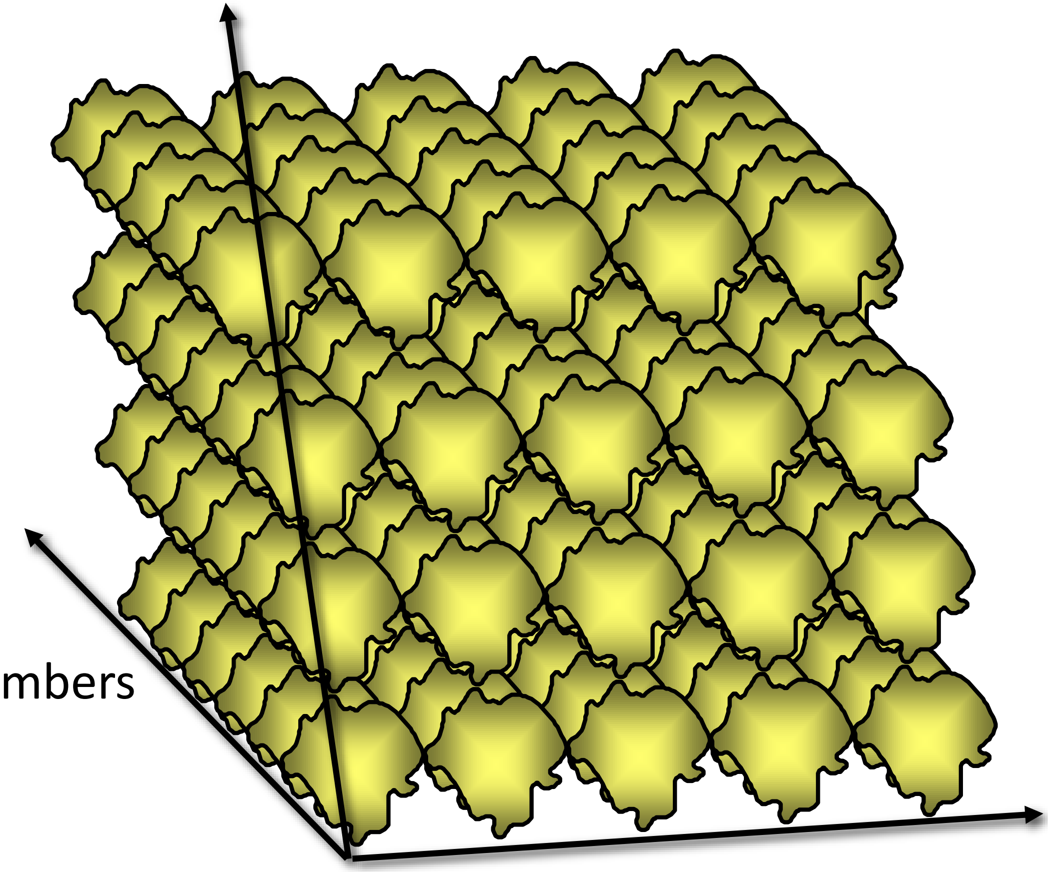
An ordered array of objects, related by a systematic combination of translations and rotations

Unit cell:

- the "bricks" of the crystal
- the unit cells are all oriented in the same way in the crystal
- the crystal is constructed by stacking unit cells together (only translational symmetry required)

The unit cell may be sub-divided into asymmetric units

Crystals are ordered arrays

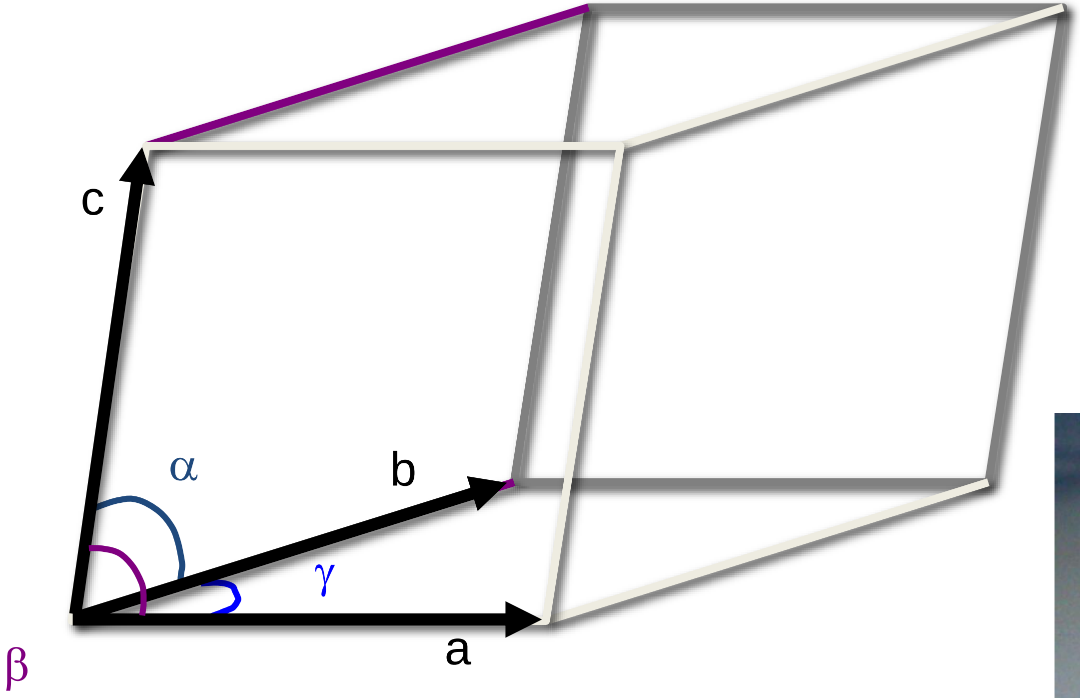


In a typical crystal there are huge numbers of molecules

Crystal = 100 µm Molecule = 100 Å

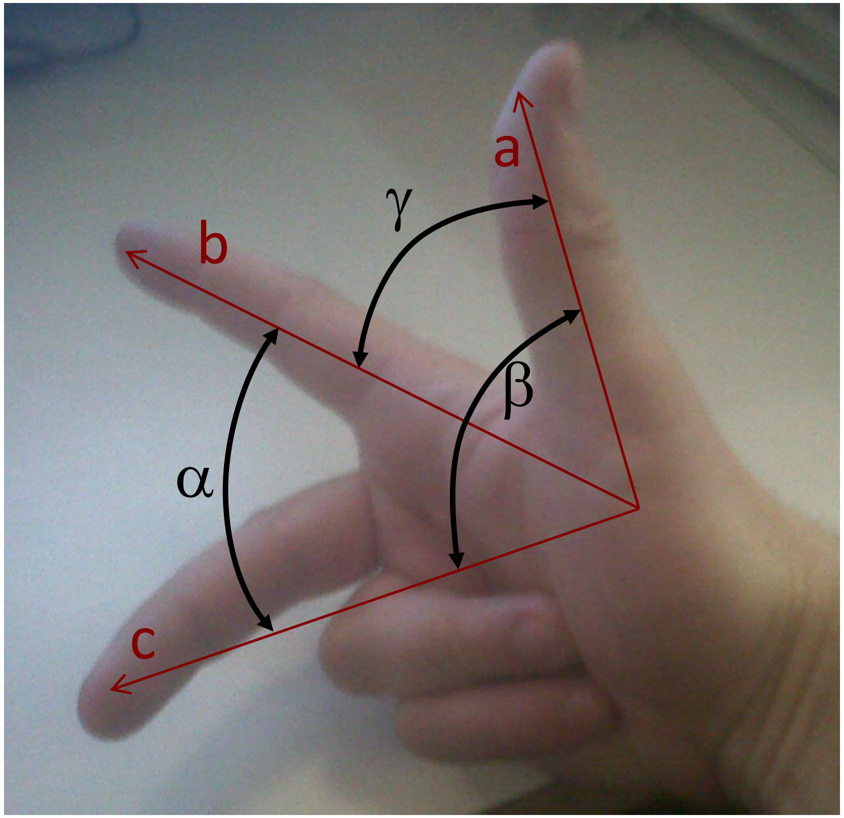
So we have 10,000 molecules along one side (and 1,000,000,000,000 in total)

Unit Cell Parameters

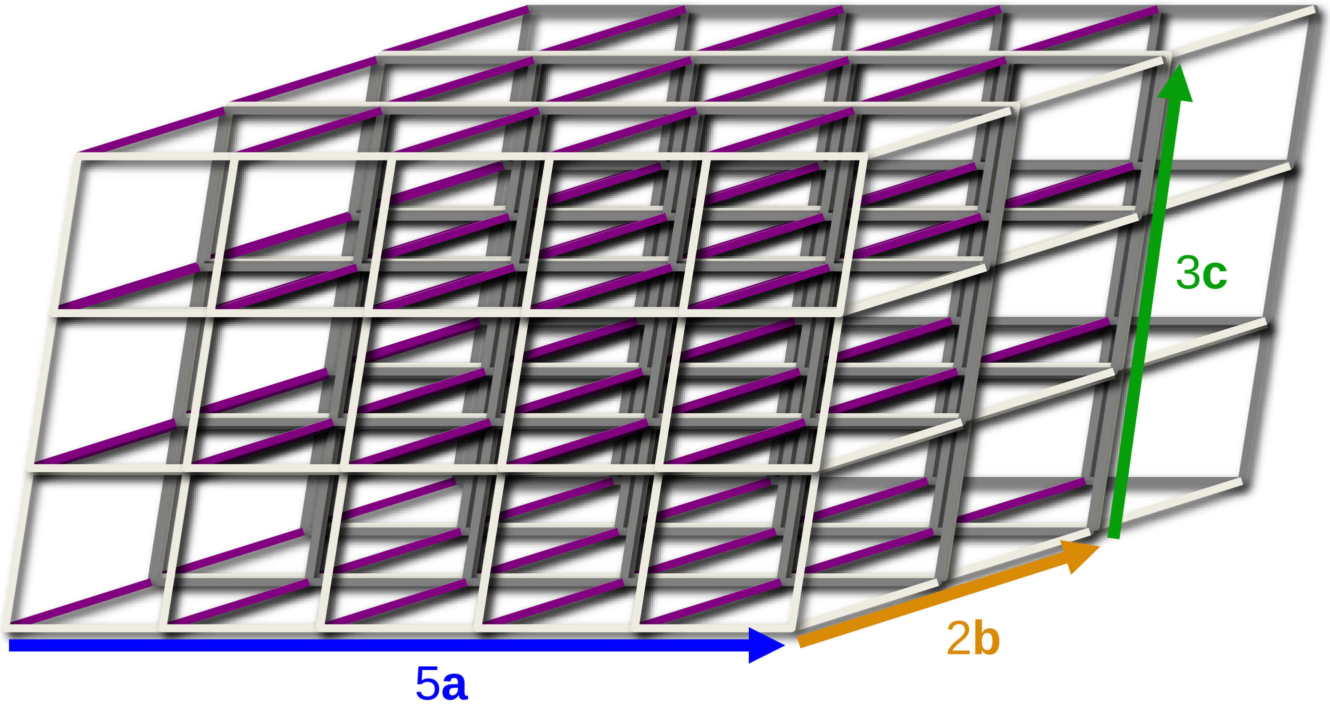


General shape: Parallelepiped

Lattice vectors: **a**, **b**, **c**
Lengths (cell edges): |**a**|, |**b**|, |**c**|
Angles between lattice vectors:
- α - angle between **b** and **c**
- β - angle between **a** and **c**
- γ - angle between **a** and **b**



A lattice of unit cells

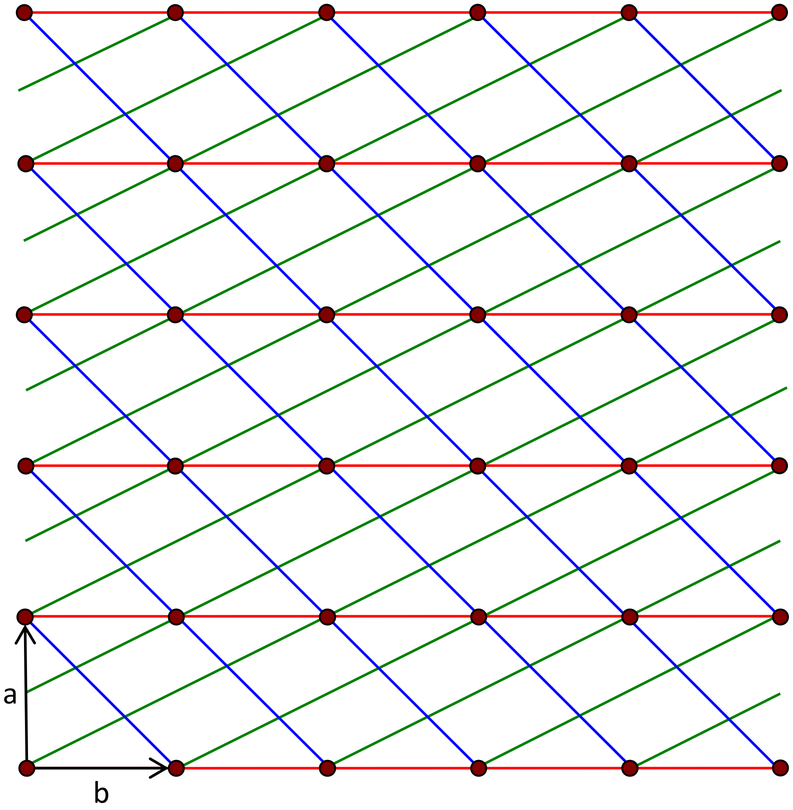


Lattice vectors can be used to define the position of any lattice point:

$r = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$ (u, v, w are integers)

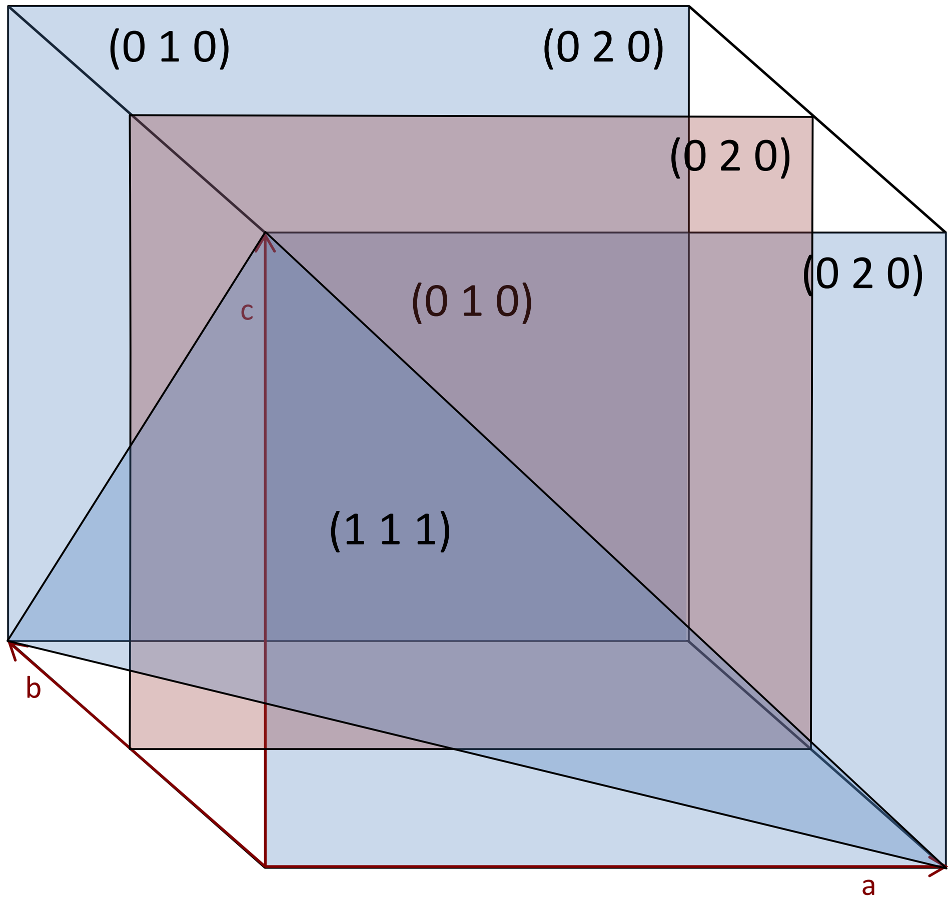
e.g. $r = 5\mathbf{a} + 2\mathbf{b} + 3\mathbf{c}$

Lattice Planes



- Start by considering a 2-dimensional lattice
- Planes cross a-axis 1 times per unit cell repeat and b-axis 0 times per repeat, thus the reflection from these planes is assigned the Miller indices h=1, k=0
- Planes cross a-axis 1 times per repeat (moving downwards) and b-axis 1 times per cell repeat, thus the reflection from these planes is therefore assigned the Miller indices h=-1, k=1
- Planes cross a-axis 2 times per repeat and b-axis 1 times per repeat, thus the reflection from these planes is therefore assigned the Miller indices h=2, k=1
- NB. Higher Miller indices indicate finer sampling (smaller d between planes) leading to a greater scattering angle and higher resolution

Lattice Planes in 3 Dimensions



The reciprocal lattice

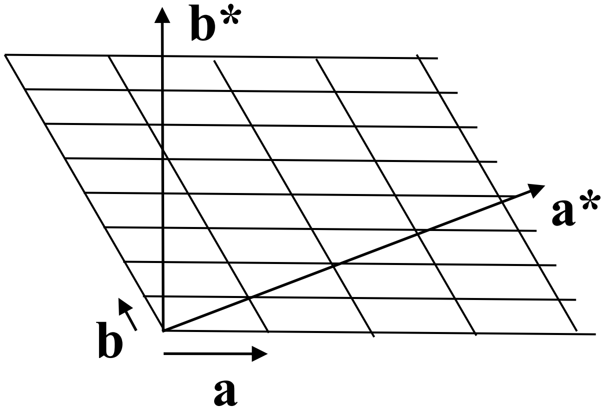
The reciprocal lattice is an abstract concept, unlike the crystal (real space) lattice, that is very useful in visualizing diffraction geometry when used together with the Ewald sphere construction.

The reciprocal lattice is defined by three vectors **a***, **b***, **c*** (in the same way that the crystal lattice is defined by the three vectors **a**, **b**, **c**) as follows:

$\mathbf{a^*} = (\mathbf{b} \times \mathbf{c})/V$ $\quad$ $\mathbf{b^*} = (\mathbf{c} \times \mathbf{a})/V$ $\quad$ $\mathbf{c^*} = (\mathbf{a} \times \mathbf{b})/V$ $\qquad$ V = unit cell volume

This definition implies:

$\mathbf{a.a^*} = 1$ $\qquad$ $\mathbf{b.b^*} = 1$ $\quad$ $\mathbf{c.c^*} = 1$



A two dimensional projection down the **c** axis (assumed to be orthogonal to **a** and **b**)

The reciprocal vectors are perpendicular to the *planes* of the real crystal lattice, so in 3D **a*** is perpendicular to (100), **b*** to (010) **c*** to (001)

|**a***| is the inverse of the spacing between the (100) planes,

|**b***| is the inverse of the spacing between the (010) planes,

|**c***| is the inverse of the spacing between the (001) planes

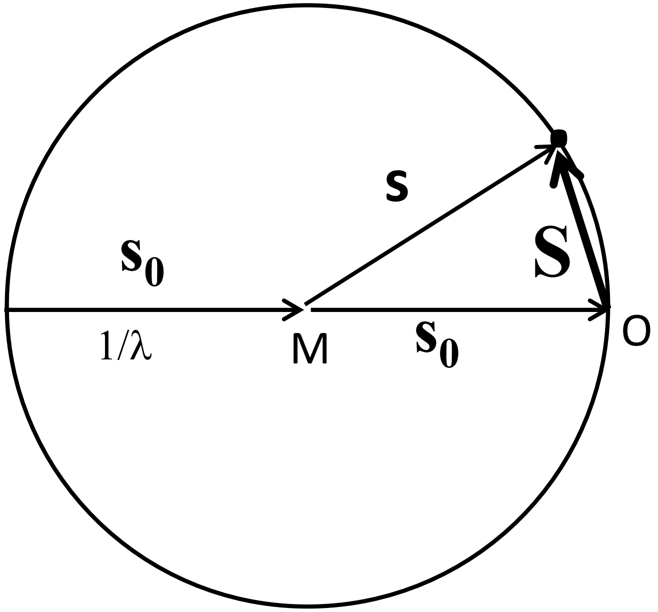
For the special case of an orthogonal lattice (only),

|**a***| = 1/|**a**|, |**b***| = 1/|**b**|, |**c***| = 1/|**c**| ; **a*** ||ˡ **a** **b*** ||ˡ **b** **c*** ||ˡ **c**

We define the scattering vector **S** (or **d***) for a particular reciprocal lattice point as: **S** = h **a*** + k **b*** + l **c***

Then |**S**| = |**d***| is the inverse of the inter-planar spacing of the Bragg (hkl) planes

The Ewald Sphere Construction



$\mathbf{s}_0$ is the incident beam vector

s is the diffracted beam vector

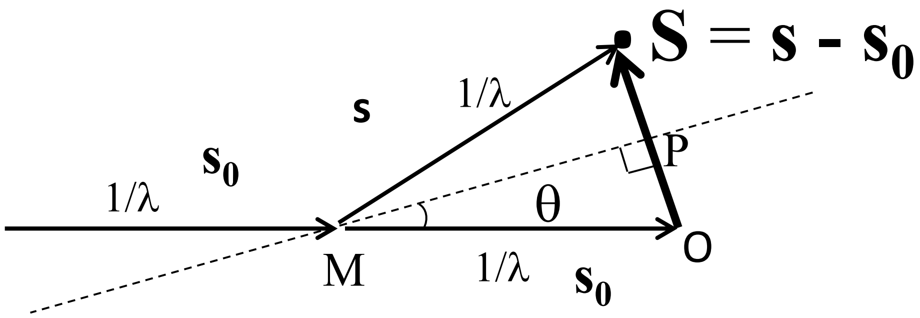
S is the scattering vector

Sphere radius $|\mathbf{s}_0| = |\mathbf{s}| = 1/\lambda$

$|\mathbf{S}| = d^* = 1/d$

The general condition for diffraction is illustrated by the vector equation:

$\mathbf{S} = \mathbf{s} - \mathbf{s}_0$



In the triangle MOP

$\mathrm{Sin}(\theta) = OP/OM = \lambda d^*/2 = \lambda/2d$

Re-arranging this we get: $\lambda = 2d\,\mathrm{Sin}(\theta)$

S is the scattering or diffraction vector in *reciprocal space*

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

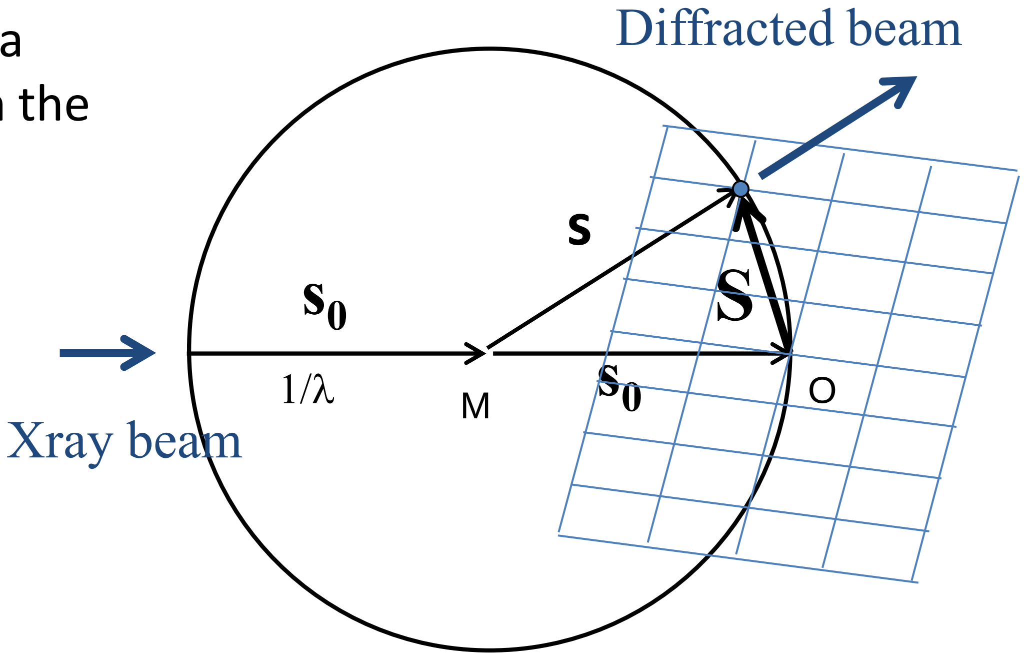
The Ewald sphere is only a construction but is very useful to understand the *geometry* of diffraction. Confusingly, it has two origins:-

M is the centre of the sphere, and may be considered as the position of the crystal, since this is the source of the diffracted beam **s**

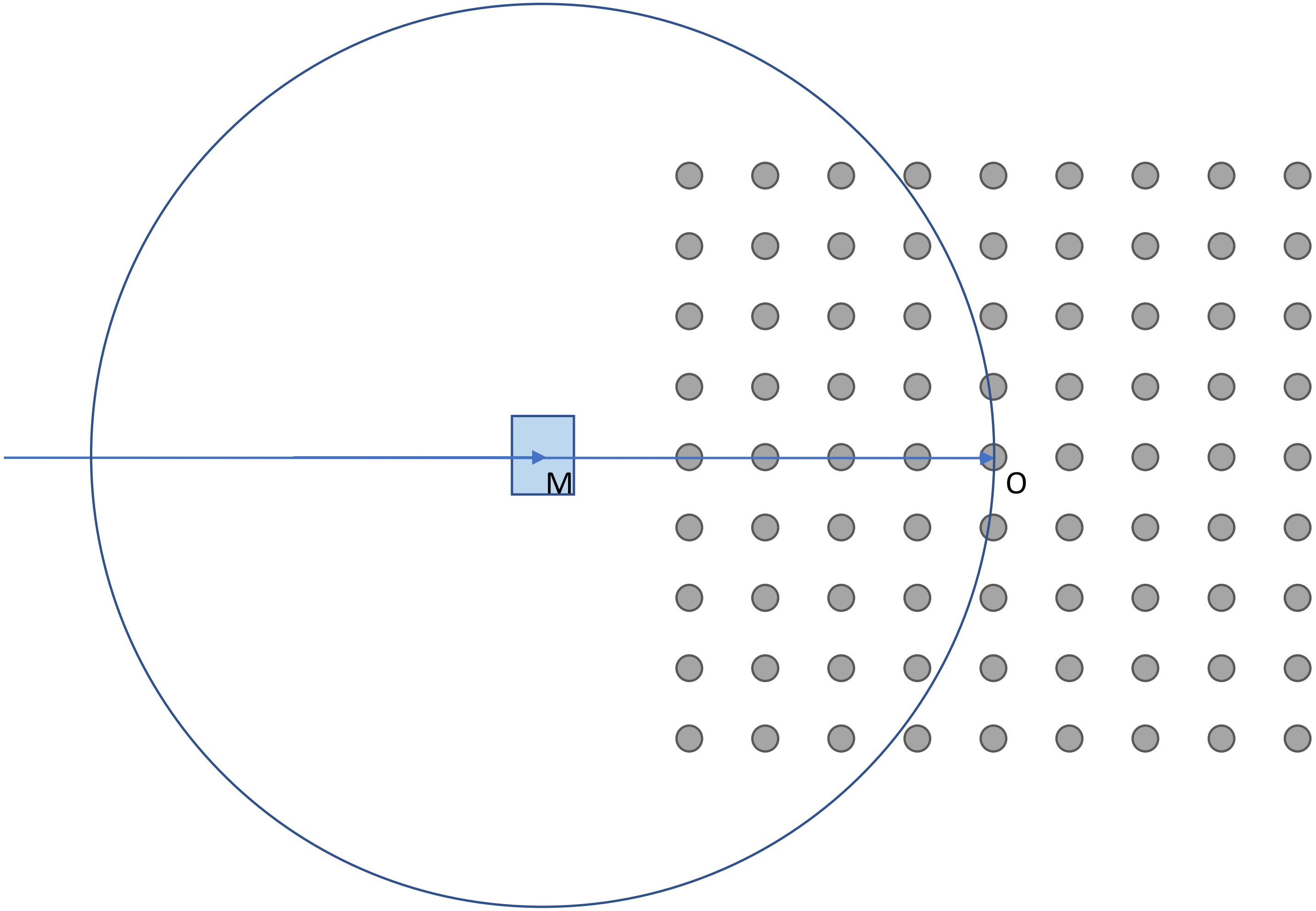
O is the origin of reciprocal space, the origin of the scattering or diffraction vector **S**, and the centre of the reciprocal lattice

As the crystal rotates, the reciprocal lattices rotates in exactly the same way

Diffraction only occurs when a reciprocal lattice point lies on the sphere

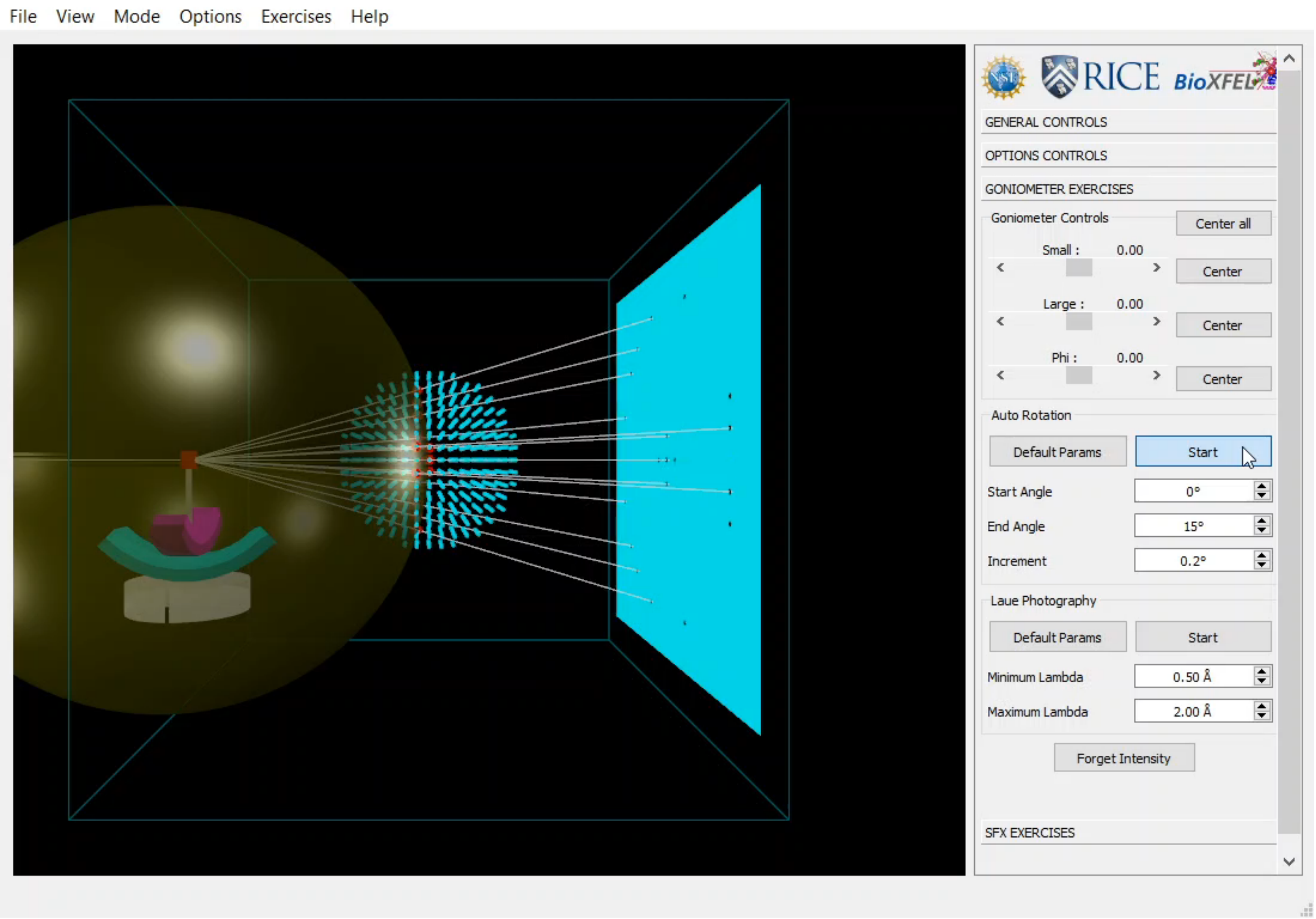


An example in 2 dimensions



In three dimensions

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



Small : 0.00

Large : 0.00

Phi : 0.00

Start Angle 0°

End Angle 15°

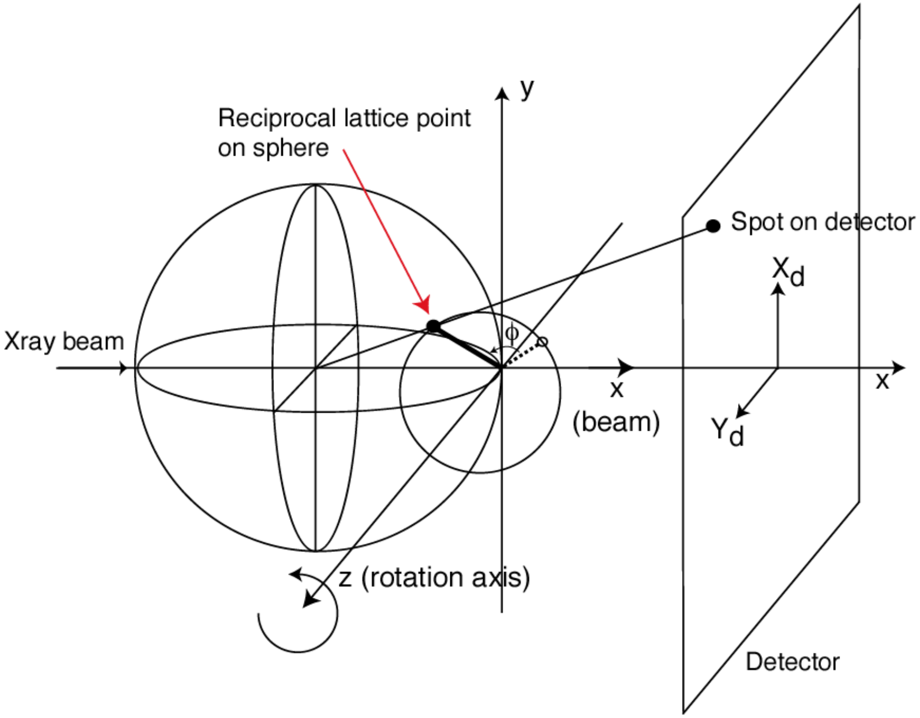
Increment 0.2°

Minimum Lambda 0.50 Å

Maximum Lambda 2.00 Å

http://www.phillipslab.org/downloads

What happens on the detector

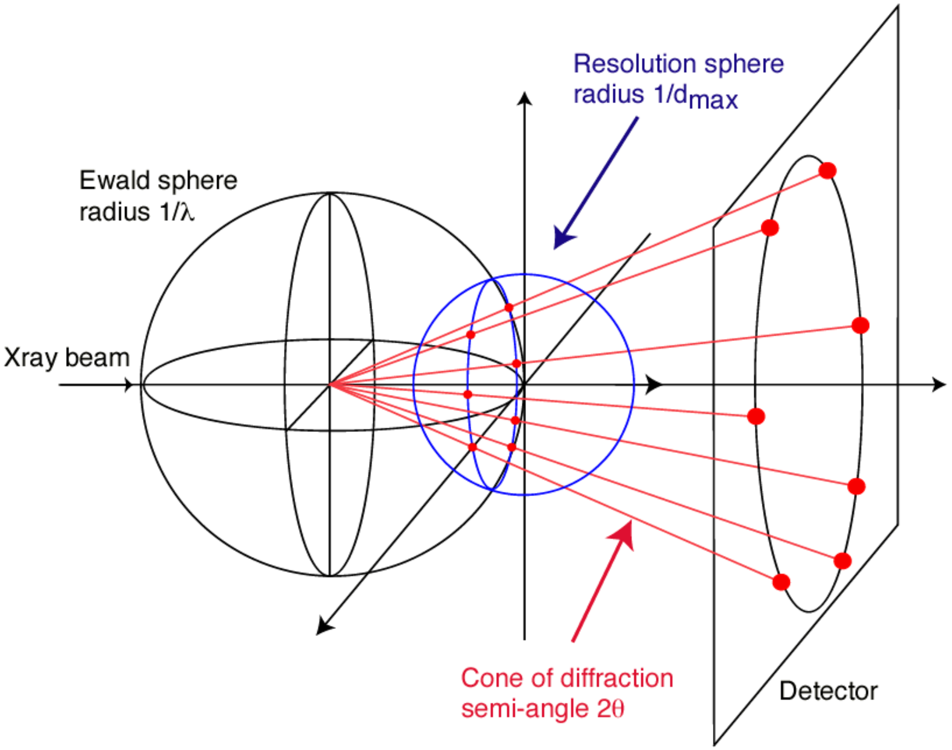


As the crystal rotates, each lattice point in turn passes through the sphere, and a spot is recorded on the detector.

We can use the Ewald construction to understand

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

Detector position



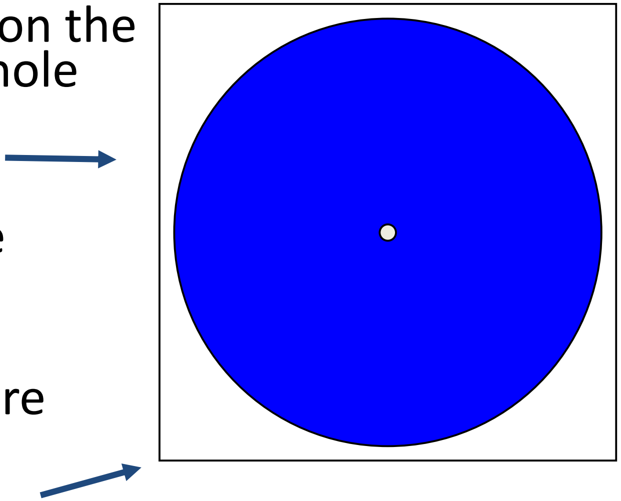
For a maximum resolution of d_{max}, all diffraction vectors **S** must lie within a resolution sphere of radius $1/d_{max}$

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

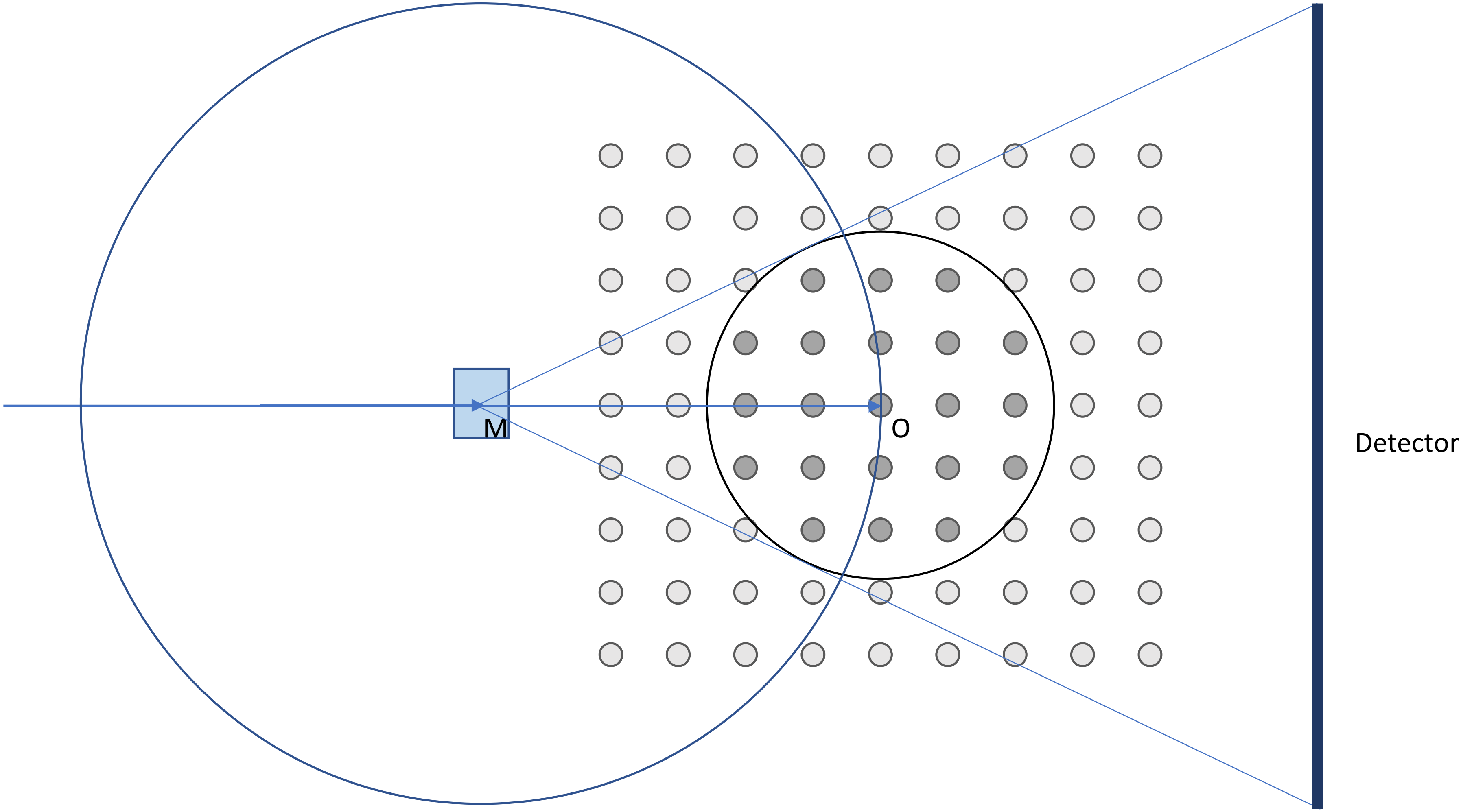
$$\lambda/d_{max} = 2 \sin \theta_{max}$$

A detector centered on the beam collects the whole cone

This gives optimum efficiency and simple strategy

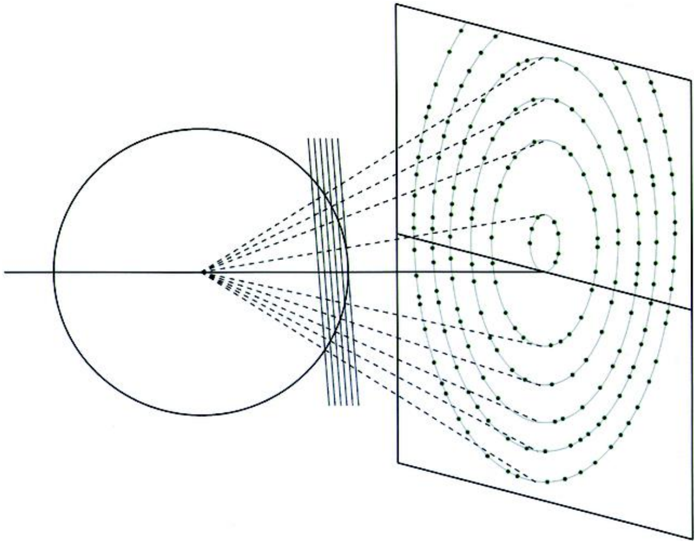


The corners of a square detector collect incomplete data

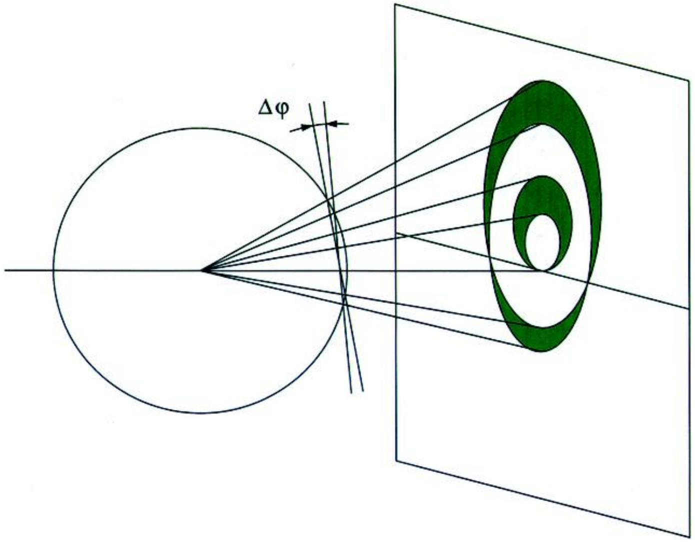

M
O
Detector

The appearance of diffraction images

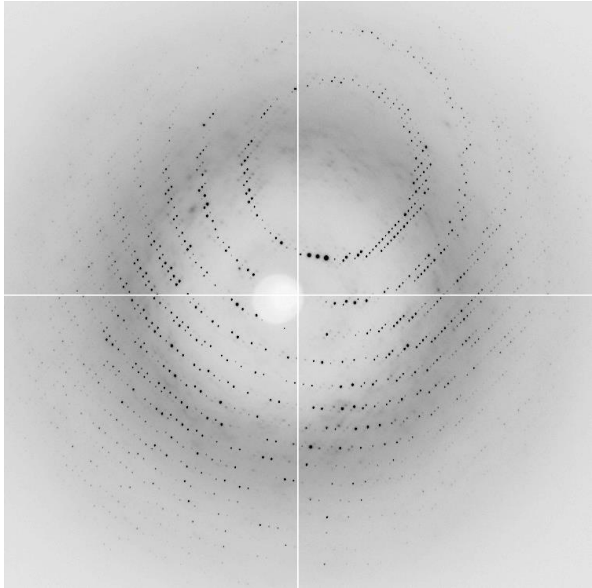
Reciprocal lattice points lie in layers (planes). Each plane intersects the sphere in a circle, and the spots projected on the detector lie in ellipses

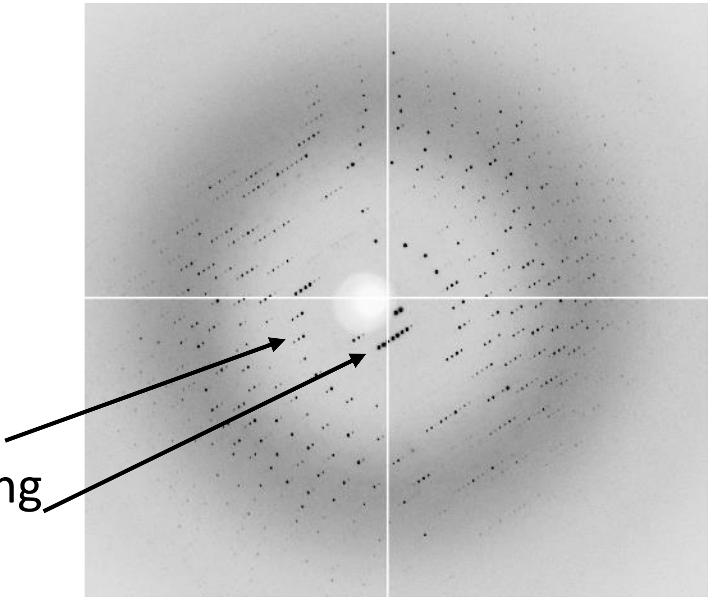


If the crystal is rotated through a small angle each circle is broadened into a ***lune***. All the spots in a lune belong to one plane of the reciprocal lattice (not necessarily a principal plane)



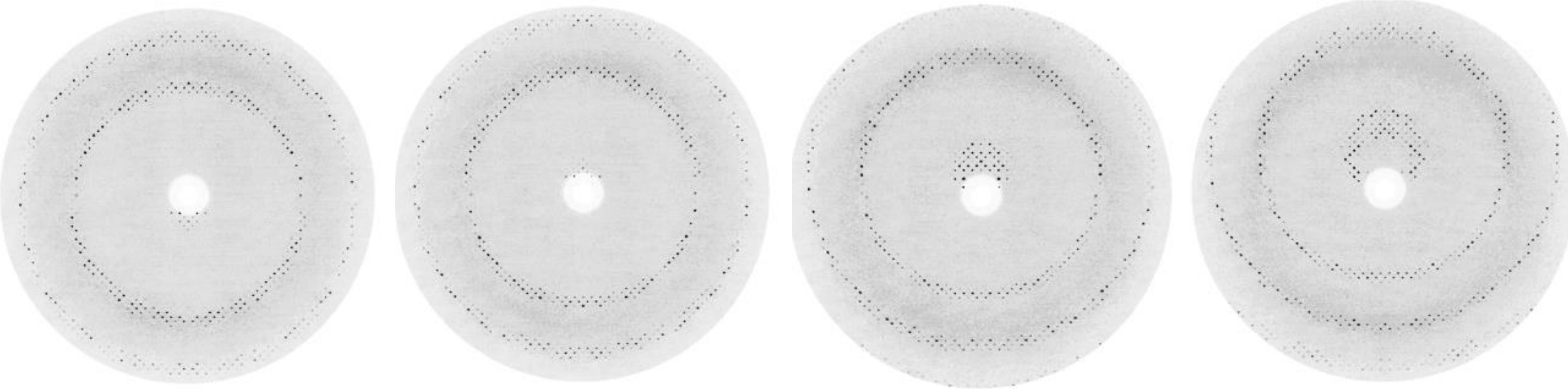
Examples of lunes in rotation images



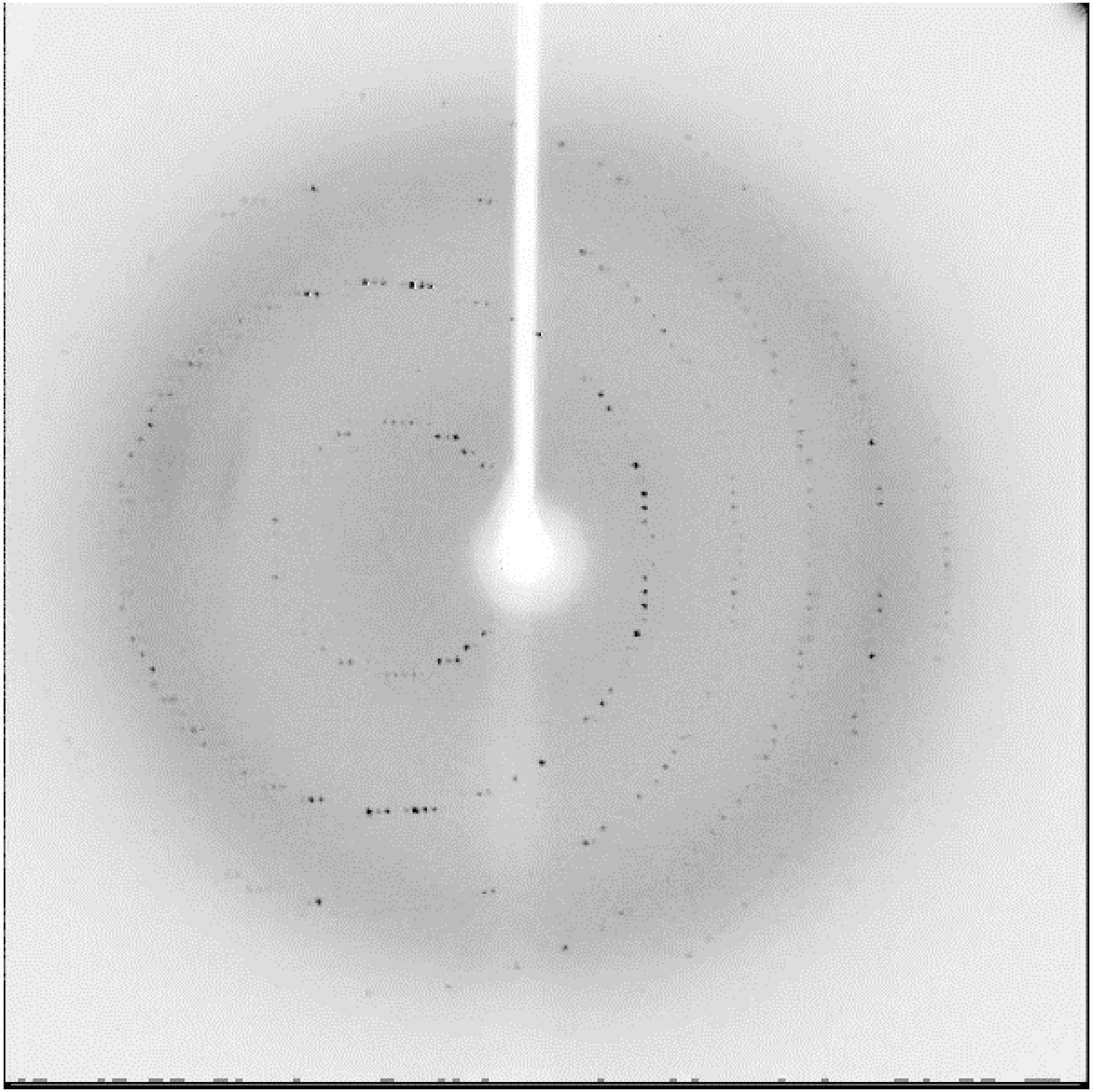


Lunes corresponding to different sets of planes

A series of successive rotation images showing the lunes moving up the image



A movie of diffraction from an aligned crystal



The size of spots in reciprocal space and on the detector

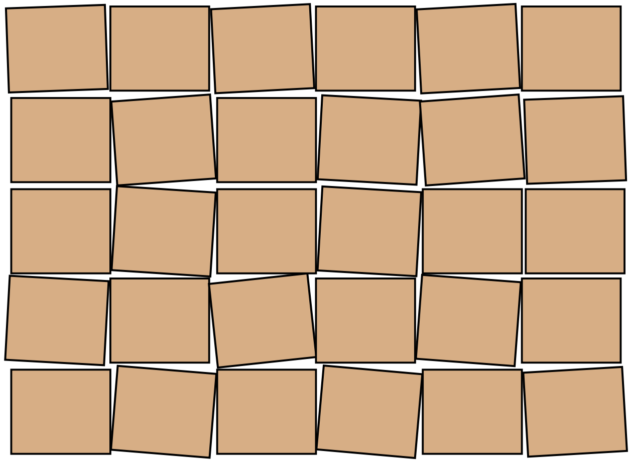
Real observed diffraction is complicated by the imperfections of real crystals and X-ray beams

The X-ray beam

- the incident beam has a finite width and is not exactly parallel ***(beam divergence)***
- the beam is not entirely monochromatic ***(dispersion).*** As a result the Ewald sphere (radius 1/λ) will have a finite thickness.

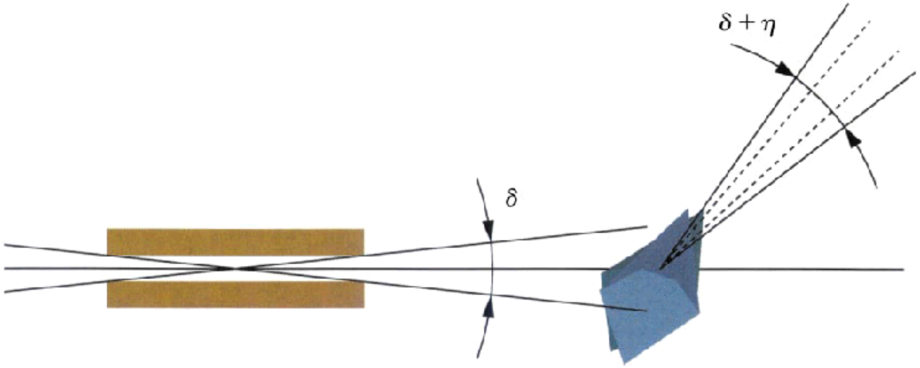
The crystal

- the crystal is not perfect, but may be considered a mosaic of blocks in slightly different orientations ***(mosaicity)***
- the crystal has a finite size (only has a significant effect for very small crystals (<5µm))

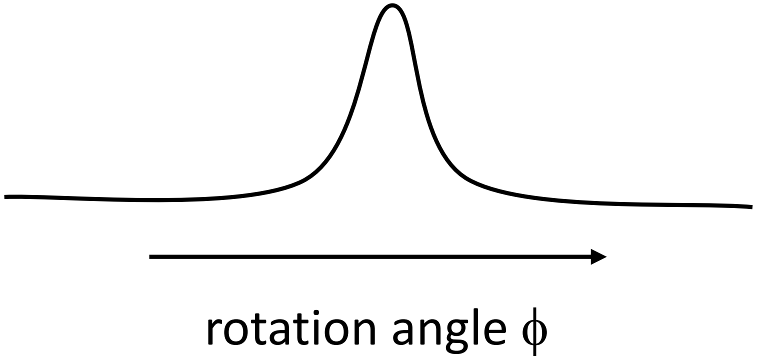


The effect of these factors can be considered as a broadening of the reciprocal lattice points, giving them a non-zero size and therefore a finite reflecting range

(Figure from James Holton)

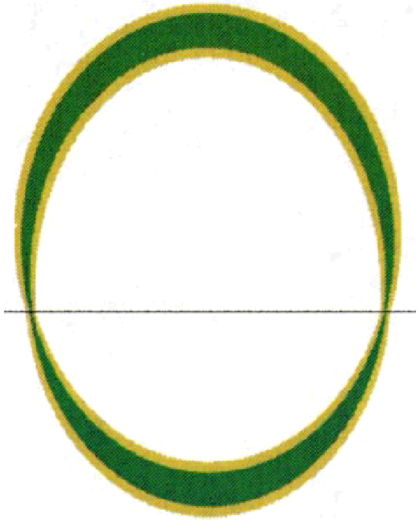


Beam divergence δ and mosaicity η add up to increase the angular width of the diffracted beam

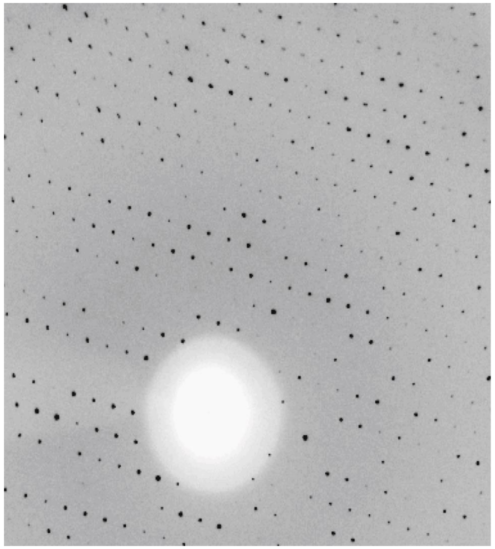


Reflection width $= \delta + \eta +$ geometric factor

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

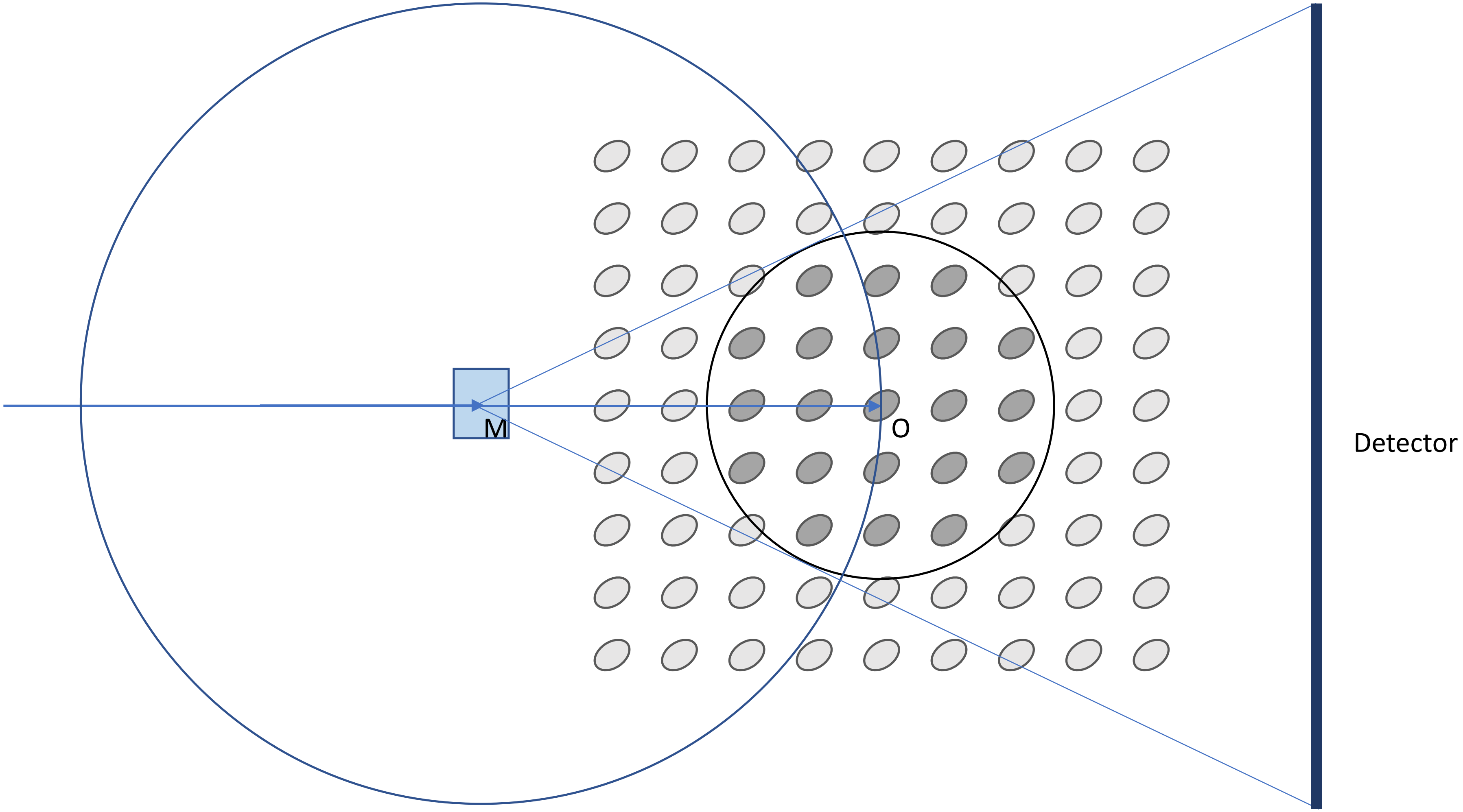


High mosaicity causes broadening of the lunes

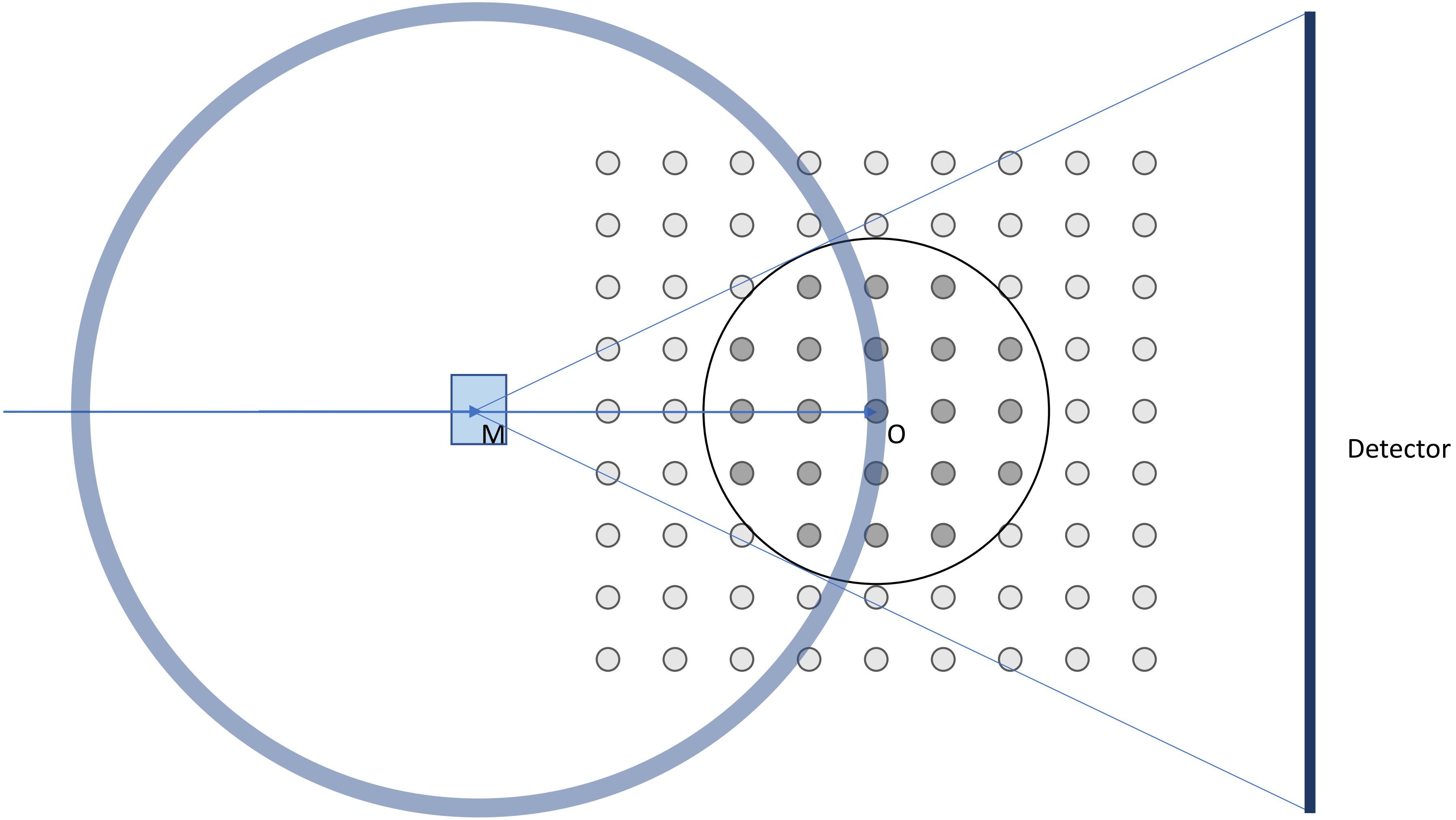


Sometimes no lunes can be distinguished.

High Mosaicity



Non-monochromatic (pink) beam

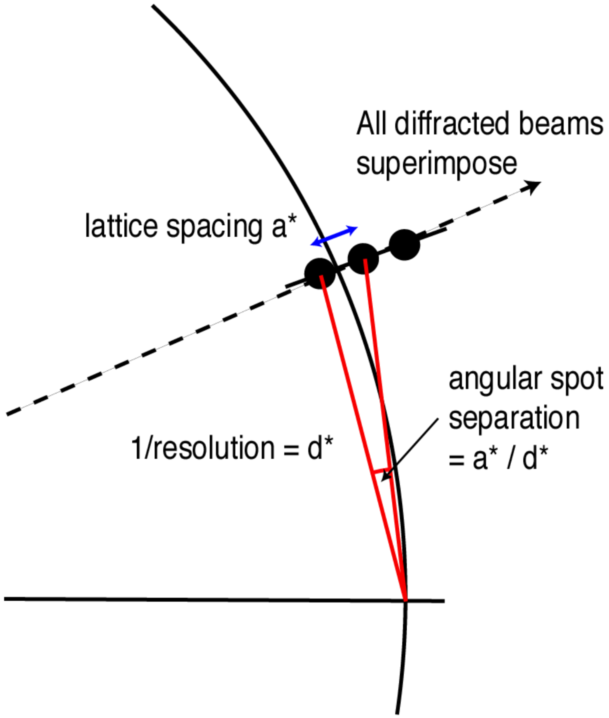


Overlaps and rotation range

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 of adjacent spots can be 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 = $(a^*/d^*) - w$

$= d/a - w$

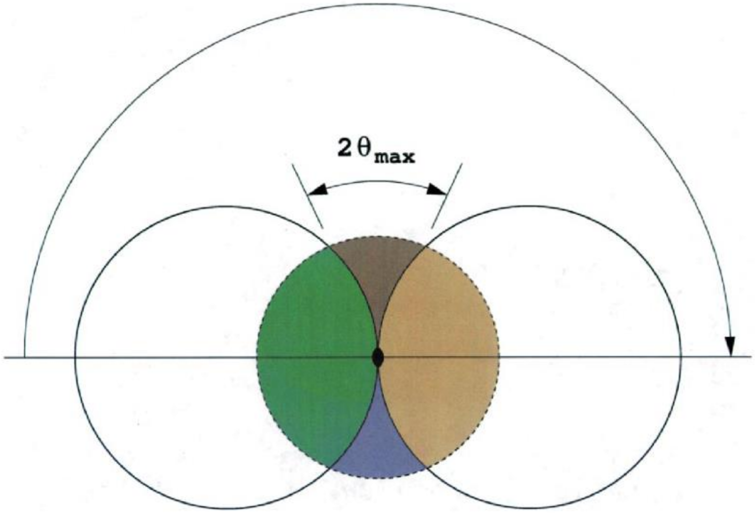
where w is the reflection width $= \delta + \eta$

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

Completeness: total rotation range and the blind region

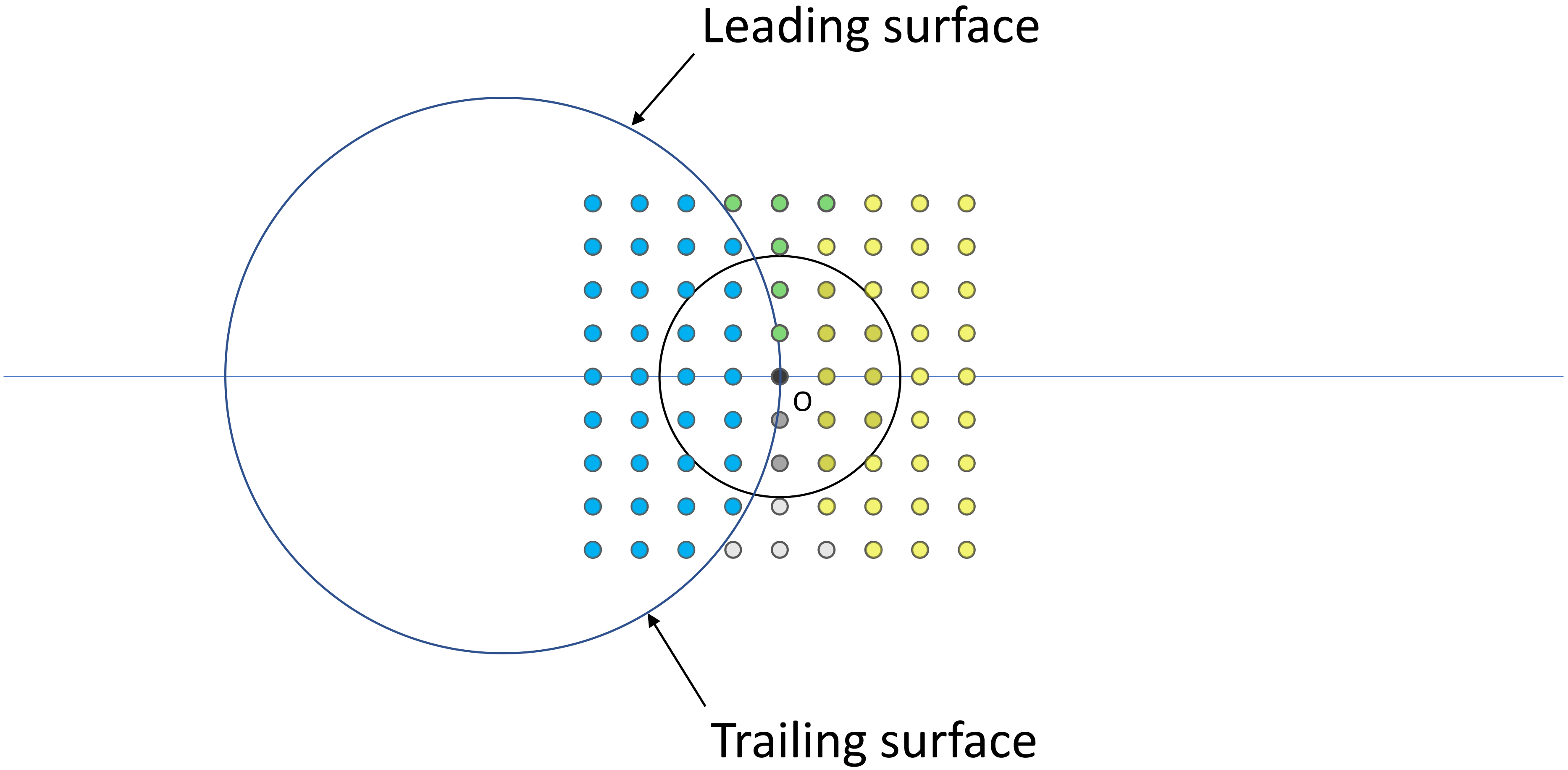


To use the Ewald sphere construction to understand which parts of reciprocal space are measured, it is easier to fix the “resolution sphere” of all reciprocal lattice points within a maximum resolution, and to rotate the Ewald sphere to correspond to the crystal rotation. The region collected is the 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 and dark brown, while the upper boundary sweeps out the dark brown and the light brown part. The blue part is not covered at all

Figure from Zbyszek Dauter

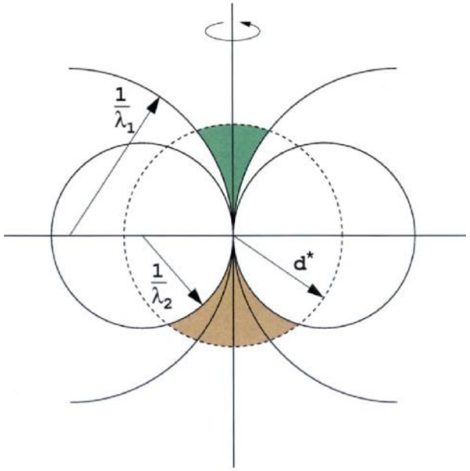
Because of Friedel’s law, this dataset is complete (apart from the *blind region*), but if complete anomalous differences are required for a triclinic cell (P1) then 180° + $2\theta_{max}$ is required.

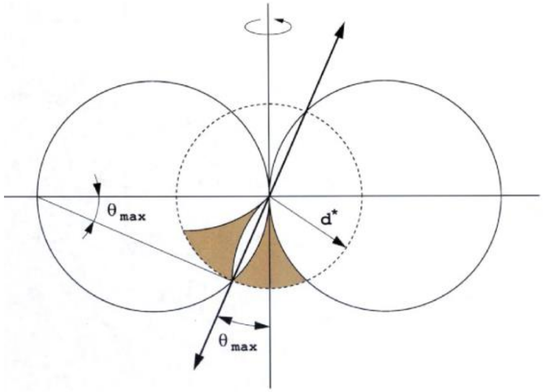

Leading surface
O
Trailing surface

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 θ_{max} or by symmetry (except in P1)

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

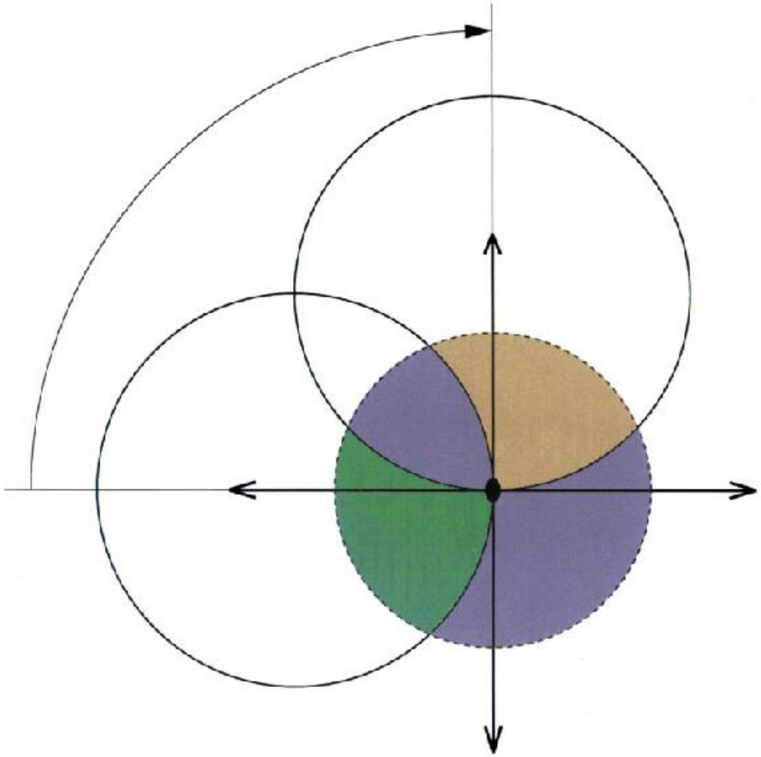
In practice, for commonly used wavelengths of ~1Å, and because the crystal is not usually aligned to rotate around a unit cell axis, the blind region is not normally an issue

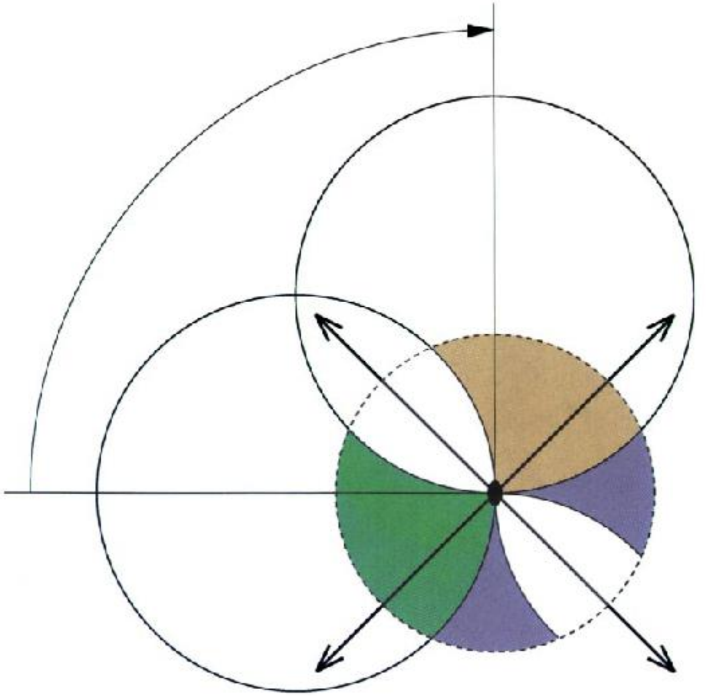
Figures from Zbyszek Dauter

Symmetry and total rotation range: an orthorhombic example

Rotation of an orthorhombic crystal by 90° starting from an axis aligned along the beam 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





Figures from Zbyszek Dauter

Data collection strategy

Compromise between getting good data 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

Issues to consider:

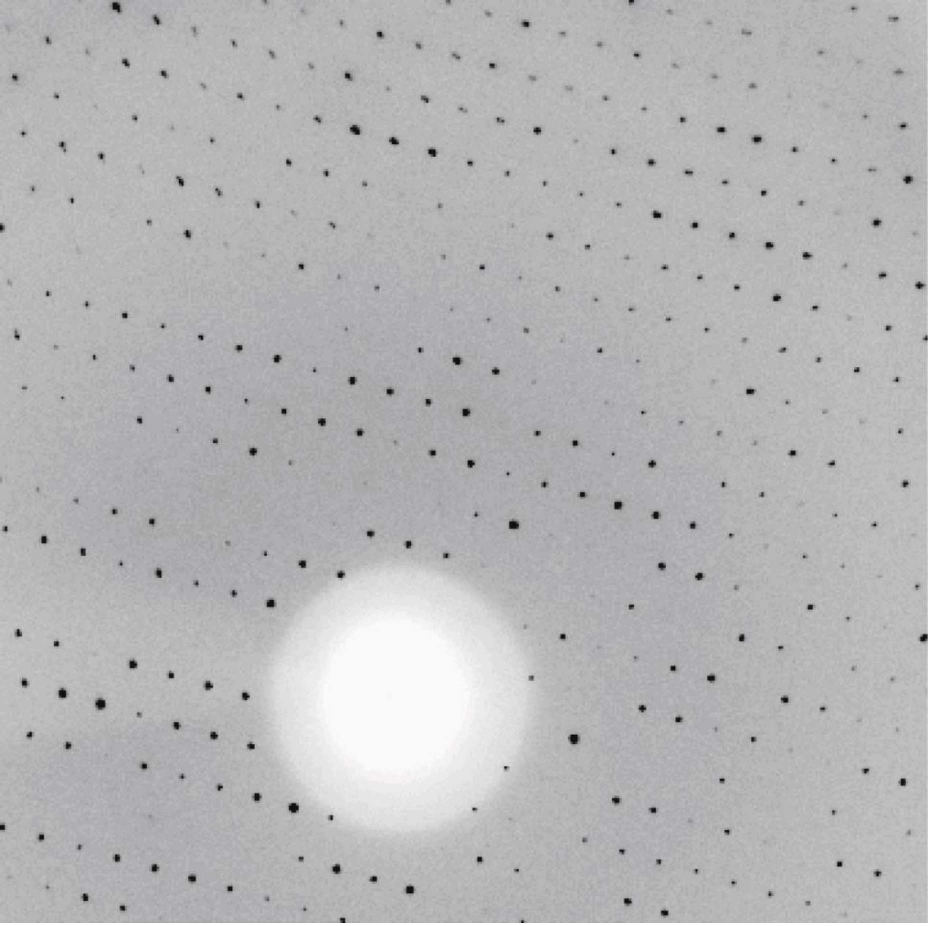
- selecting a good crystal
- image width
- exposure time
- start angle and total rotation range
- detector position (maximum resolution)
- wavelength (for anomalous phasing)
- beam size
- do you need multiple crystal positions or (isomorphous) crystals?

(several of these are linked)

What is a good crystal?

- Single: *only one lattice, check by eye or indexing pattern and looking for unpredicted spots*
- Diffracts to a useful resolution: *depends on the goal of the experiment*
- Low mosaicity: *improves signal to noise*
- 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 (100–200μm). Smaller crystals usually cryo-cool better (lower mosaicity)*
- Homogeneous: *Diffraction quality is uniform for the entire crystal volume. Can be checked by a raster scan with a small beam. If not homogeneous, use a small beam and scan crystal for best regions.*
- Good cryo-protection: *no ice, minimum amount of liquid (low background, hence improved signal to noise)*
- Well centred in the beam: *ensures crystal remains fully in the beam during rotation*
- The best that you have!

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



Phi = 0°

Phi = 90°

Always check diffraction in two orthogonal images !

Image width

Image plate detectors and older CCD models take a significant time to read-out, so for these detectors we have to close the shutter & stop the rotation (simultaneously!) during the read-out. To optimise use of beam-time, large image widths (0.3-1.0°) were commonly used.

Modern pixel detectors (eg Pilatus and Eiger) with very fast read-out (a few msec) can be used with continuous rotation with the shutter open (shutter-less data collection).

In both cases our sampling of the 3-dimensional reciprocal space is in consecutive slices, typically of between ~0.1° and 1°

For shutter-less data collection, the "optimum" image width is ~1/2 XDS width σ_{ϕ}, or ~1/5 Mosflm "mosaicity"

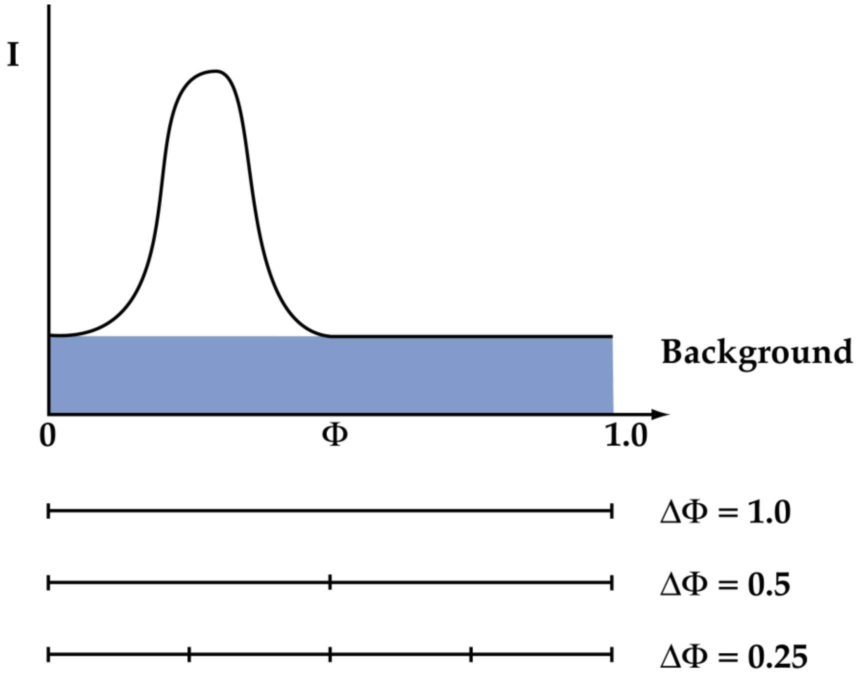
Mueller, Wang & Schulze-Briese, 2011. *Acta Cryst (2012). D68, 42-56*

1° is always wrong

For pixel detectors use 0.05-0.2°, depending on mosaicity

Advantages of fine phi slicing

Using an oscillation angle smaller than the reflection width (in phi) will improve signal to noise for weak reflections by minimising the background:



$$I_{spot} = I_{tot} - I_{bck}$$

$$\mathrm{var}(I_{spot}) = \mathrm{var}(I_{tot}) + \mathrm{var}(I_{bck})$$

$$= I_{tot} + I_{bck}$$

$$I/\sigma(I) = (I_{tot} - I_{bck})/(I_{tot} + I_{bck})^{1/2}$$

For weak reflections, the background will dominate I /σ(I).

Fine phi slicing *should* give better data but:

1) Assumes no errors in shutter synchronisation (demanding for very short (<0.5 sec) exposures).
2) There will be detector readout noise in each image (except Pilatus).

(see J. Pflugrath, Acta Cryst **D55**, 1718-1725, 1999).

Exposure time per image

Can be the most critical parameter and also (perhaps) the most difficult to choose

Radiation damage is the biggest issue!
Need to estimate the crystal lifetime.
Take advantage of available knowledge:

- Lifetime of a similar crystal if data have already been collected and processed (although different crystals can behave differently). Look at B factor variation.
- Software estimates (EDNA/BEST; RADDOSE).
- Advice from beamline scientist (especially if you have not used the beamline before).

Note: It is not always straightforward to estimate crystal lifetime from processing statistics.

Given an estimate of the **total** crystal lifetime t_{decay}, then need to decide how to assign the time for each image t_{exp}. This depends on:

- Total rotation range to be collected ϕ_{tot}
- Rotation angle per image $\Delta\phi$

$$t_{exp} = t_{decay} * \Delta\phi \,/\, \phi_{tot}$$

Start angle and total rotation range

Historically, a typical strategy was to collect the minimum rotation range required to collect a complete dataset
…. and then collect more if the crystal was still diffracting well.

The minimum rotation range depends on:

- The symmetry (Laue group) of the crystal
- The crystal orientation on the goniometer
- Whether anomalous completeness is required

These strategies generally also require a starting phi value (see previous slide for an orthorhombic example).

Strategy programs are available for this (providing the crystal has been indexed to provide the orientation).

Start angle and total rotation range (contd)

However, with fast read-out, noiseless detectors, shutter-less data collection and improved processing software (that can process very weak images), this is commonly being replaced with a strategy of collecting 180° or 360° (or even more).

This has the following advantages:

- Knowledge of the symmetry is not required
- No requirement for a specific starting angle
- If necessary, it should be possible to reject later images from the dataset if radiation damage is apparent **without** affecting the completeness of the dataset.

Comments:

- A total rotation range of 180° + $2\theta_{max}$ is required for complete anomalous data in spacegroup P1.
- If collecting more than 360°, it is preferable to change the crystal orientation or move the detector in order to minimise systematic errors.

Think hard about data collection efficiency.
Calculate a data collection strategy.

Then collect at least 180° of data.

Detector position

Increasing the crystal to detector distance will:

- Improve the signal to noise (the background falls of more rapidly than the diffracted intensity)
- Reduce the maximum resolution limit

Try to get a realistic estimate of the resolution limit of diffraction. This can be difficult if the exposure time per image is calculated for a total rotation range of 360°, so use a longer exposure time for the initial reference images.

Make sure to use the whole area of the detector to collect useful data, not just the central area!

Wavelength

If not collecting anomalous data, this is not critical, but ask the advice of the beamline scientist (the intensity can vary with wavelength).

Remember that the wavelength (with detector distance) determine the maximum resolution.

Most anomalous diffraction experiments are now done with a single wavelength (SAD). Inverse beam (collecting small wedges (eg 5°) at ϕ and 180+ϕ) will minimise the effect of radiation damage on the anomalous signal, but will take longer (ask the beamline scientist).

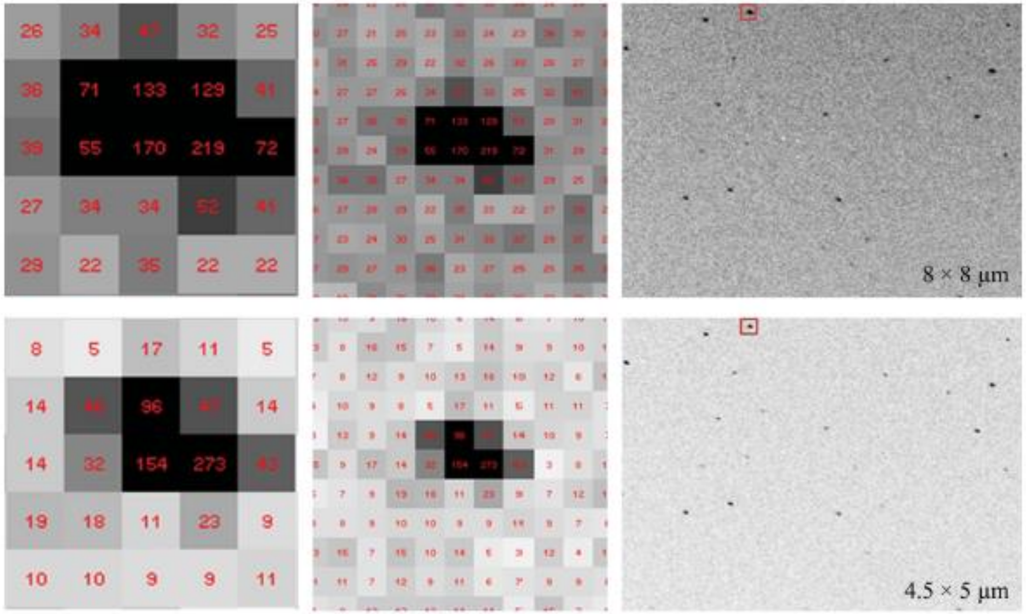
Note that radiation damage will be more serious if collecting data at a wavelength close to an absorption edge (eg for SeMet phasing).

If collecting sulphur SAD data, ask the beamline scientist for the optimum wavelength (usually around 2Å).

The importance of matching beam size to crystal size

Matching the beam size to the crystal size will optimise the signal to noise:

- A beam larger than the crystal will increase the background without increasing the strength of diffraction.
- A beam smaller than the crystal will reduce the strength of diffraction (which is proportional to the crystal volume in the beam) and if only using one position will reduce the effective crystal lifetime.



The upper figures represent the diffraction obtained using a 8x8 micron beam with a 5x5x5 micron crystal. In the lower figures, a 5x5 micron beam was used, greatly reducing the background. From Evans et al., Acta **D67**, 261-270, 2011.

Multiple positions on one crystal or multiple crystals

Helical Scans

In general, it is best to match beam size to the crystal size, but an exception is for needle shaped crystals, where it is preferable to set the beam size to the diameter of the needle and use a helical scan (where the crystal is translated along its length during data collection).

This assumes that the crystal is homogeneous in its diffraction properties.

Multiple crystals

It may not be possible to collect a complete dataset to a useful resolution from one crystal, either because the crystal is too small or because it is very weakly diffracting.

Strategy programs are available that will suggest how to collect data from the currently mounted (and indexed) crystal to maximise overall completeness taking into account data that has already been collected.
Non-isomorphism can be an issue in such cases.
This is a current research area.

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 (<1Å) 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.

The End

Thanks to Phil Evans, Zbigniew Dauter, James Holton and Andrew Leslie for figures and slides

An excellent overview is given in:
Z. Dauter, “Collection of X-ray diffraction data from macromolecular crystals”
Methods Mol. Biol. 2017, **1607**:165-184, doi:10.1007/978-1-4939-7000-1_7.
(Can be downloaded from HHS Public Access)