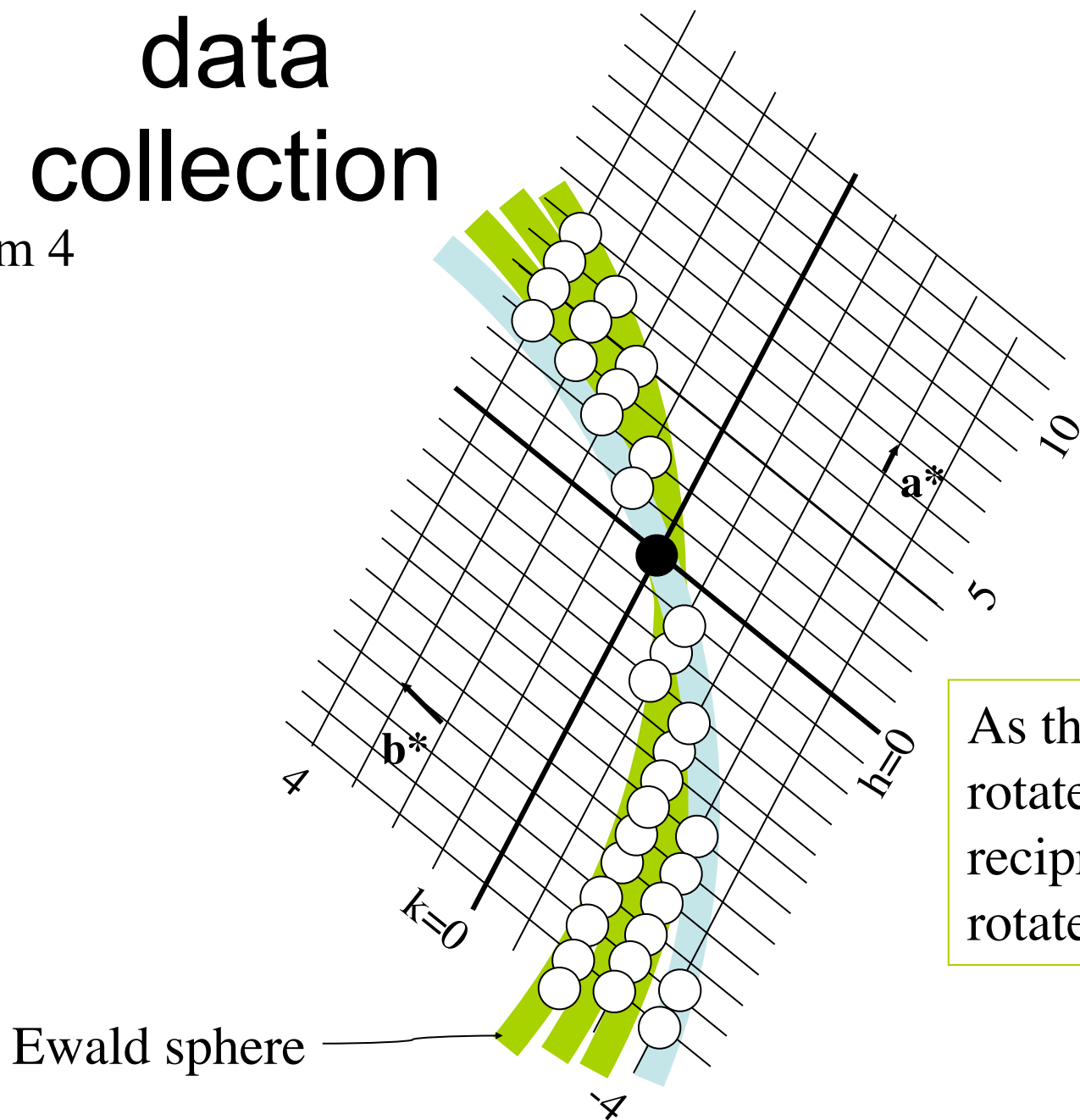


Protein Structure Determination

lecture 6 -- Ewald sphere, data
collection, scaling, B-factors

data collection

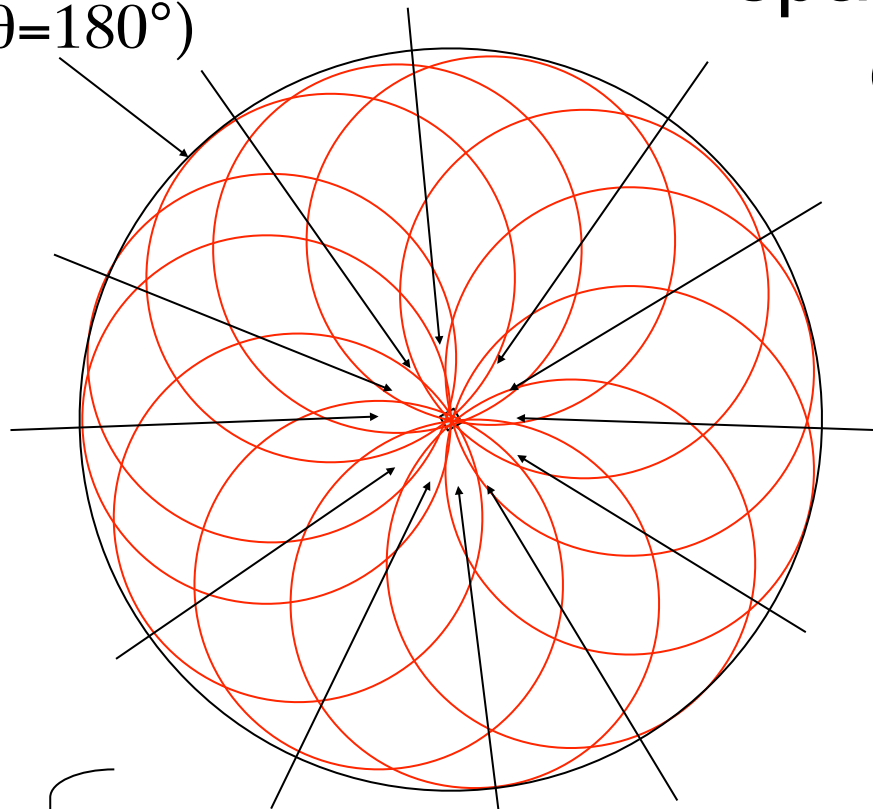
film 4



As the crystal is
rotates, the
reciprocal lattice
rotates.

limit of visible
reciprocal space = $2/\lambda$
($2\theta=180^\circ$)

The visible part of reciprocal space depends on the diffraction limit.



The set of all vectors S (red), given all possible directions of the beam (black arrows), is called reciprocal space.

Note: In real life, we find it easier to move the crystal, not the beam.

The reciprocal lattice moves with the crystal.

Watch the data collection movie. See the intersection of a sphere and a lattice?

<http://ucxray.berkeley.edu/~jamesh/movies/diffraction.mpeg>

Data collection

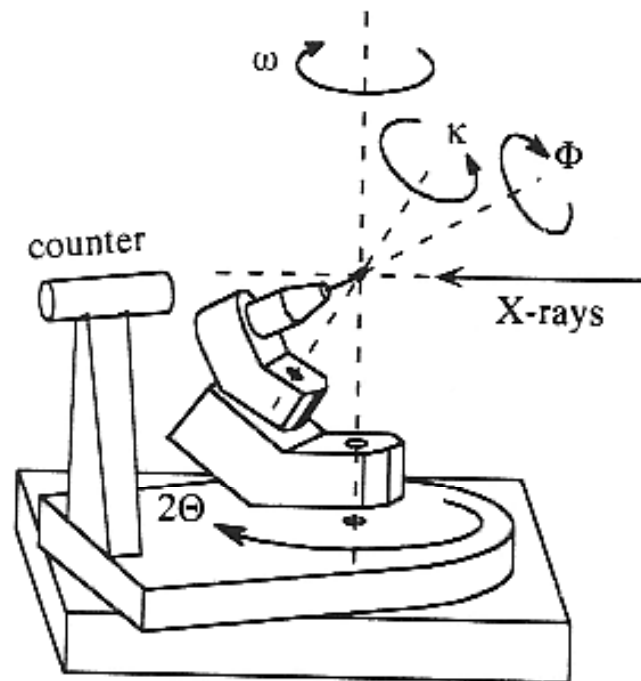
F = Square-root of the measured intensity of each reflection.

h, k, l = Scattering vector (Miller indices)

σ = standard deviation over syms

h k l F sigma

Diffractometers yesterday and today



Counter moves in 2θ .

Crystal moves in 3 angles ω , κ , and ψ

Single photon counter (photo multiplier tube)

Collecting data on photographic film

still widely used!

oscillation image

Raw images are scanned into digital images.
Each image has three angle associated with it
(κ, ψ and ω). A series of films, each with a
different ω angle, are collected and digitized.

Image plates

image plate  crystalline phosphorescent material with color centers

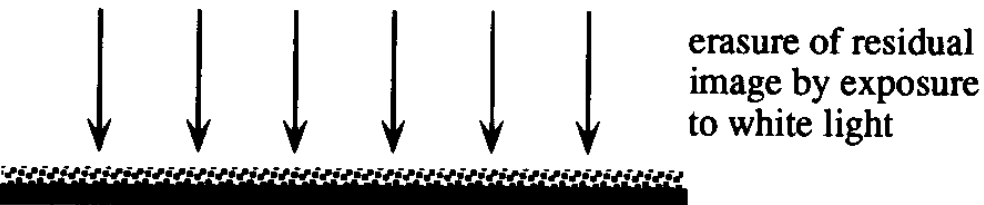
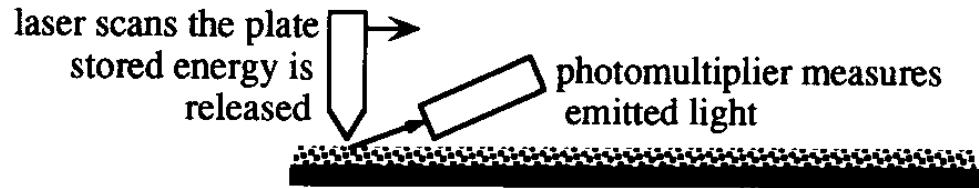
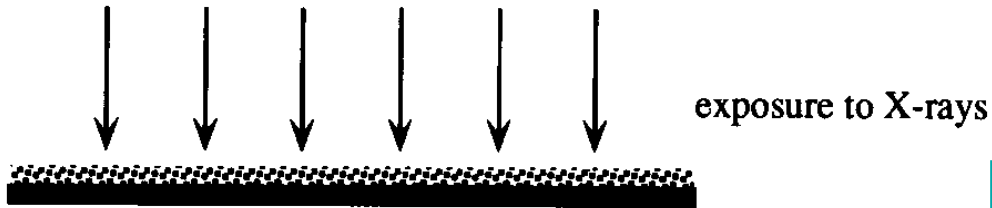
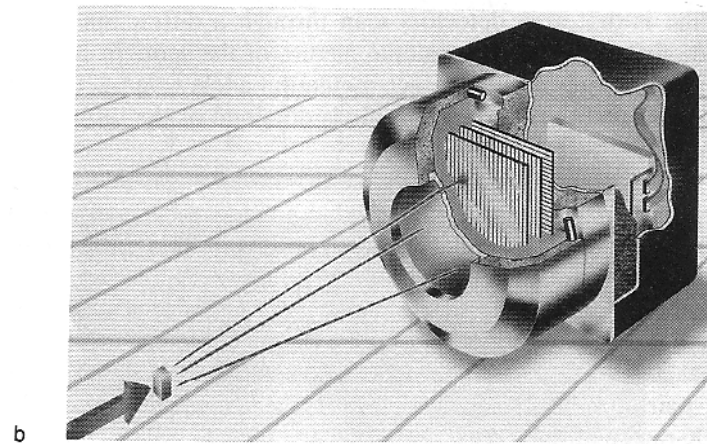
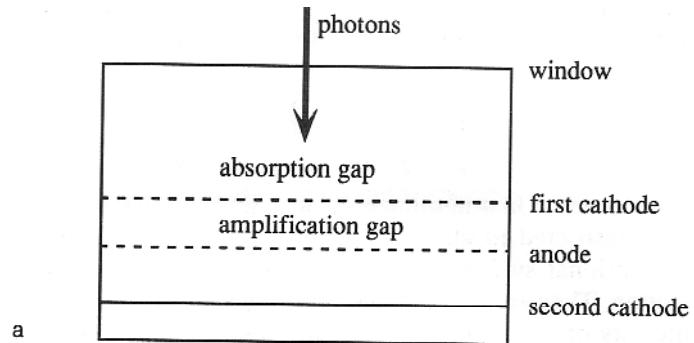


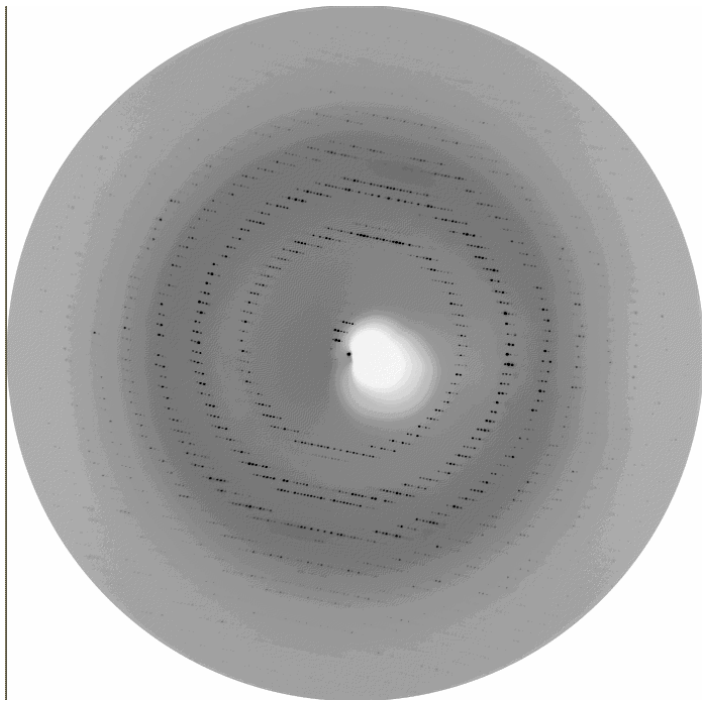
Image plates are ultra-sensitive, reusable films. Data collection is done the same way as for photographic film.

CCD area detectors



Position sensitive X-ray detectors give a 3D image of each spot, where film or image plates give 2D images.

Data reduction



raw images

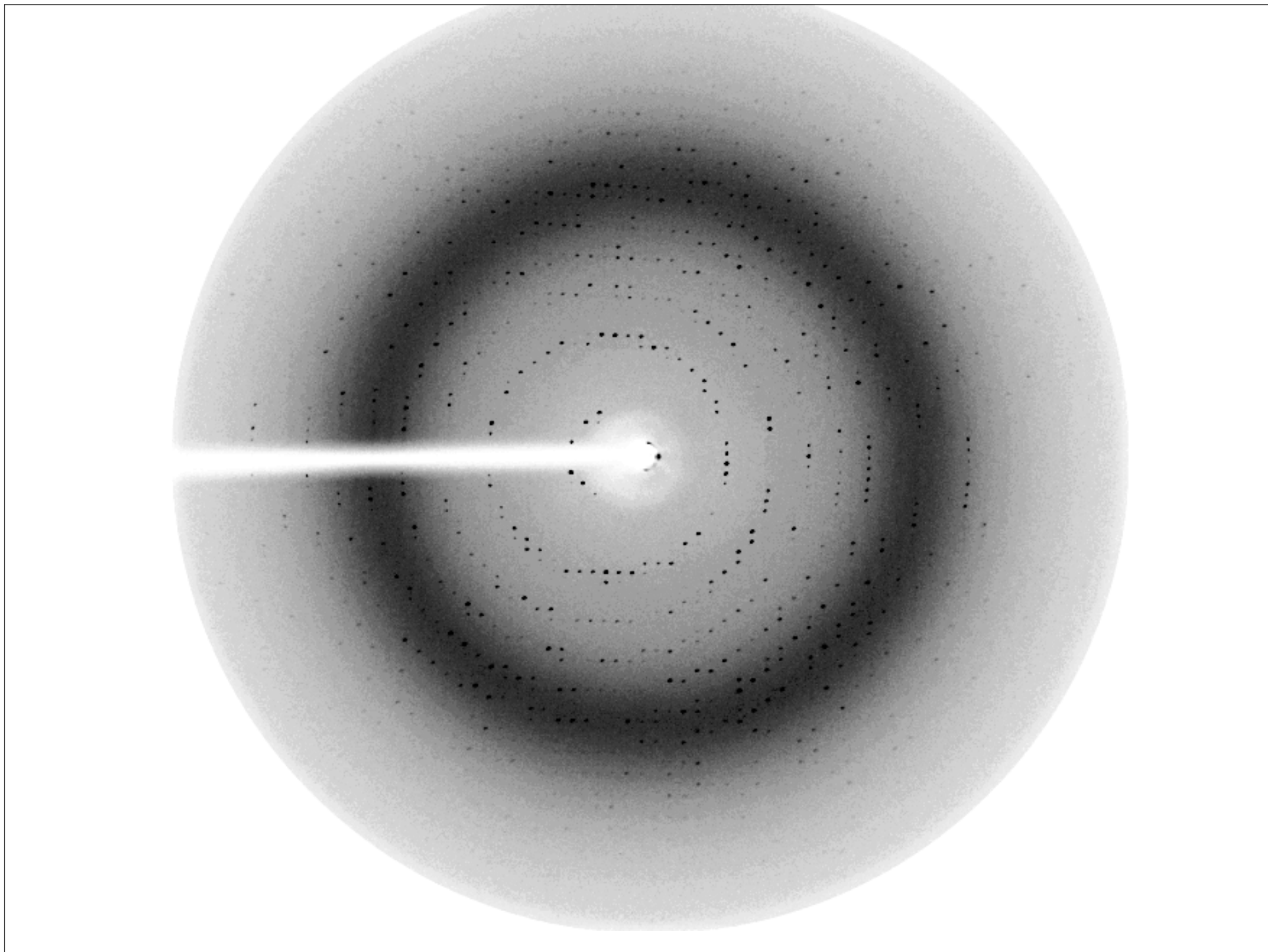
-
- indexing
 - background estimation
 - integration of spots
 - merging of partials
 - scaling
 - merging of syms

hkl	F	σ
200	99.0	0.2
210	65.1	0.3
201	78.5	0.2
220	6.3	0.1
221	19.9	0.2
222	88.1	0.2
...		

reflection data
"Structure factors"

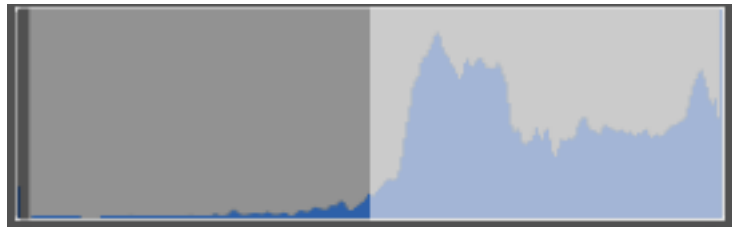
The Data reduction process: images to $hklF\sigma$

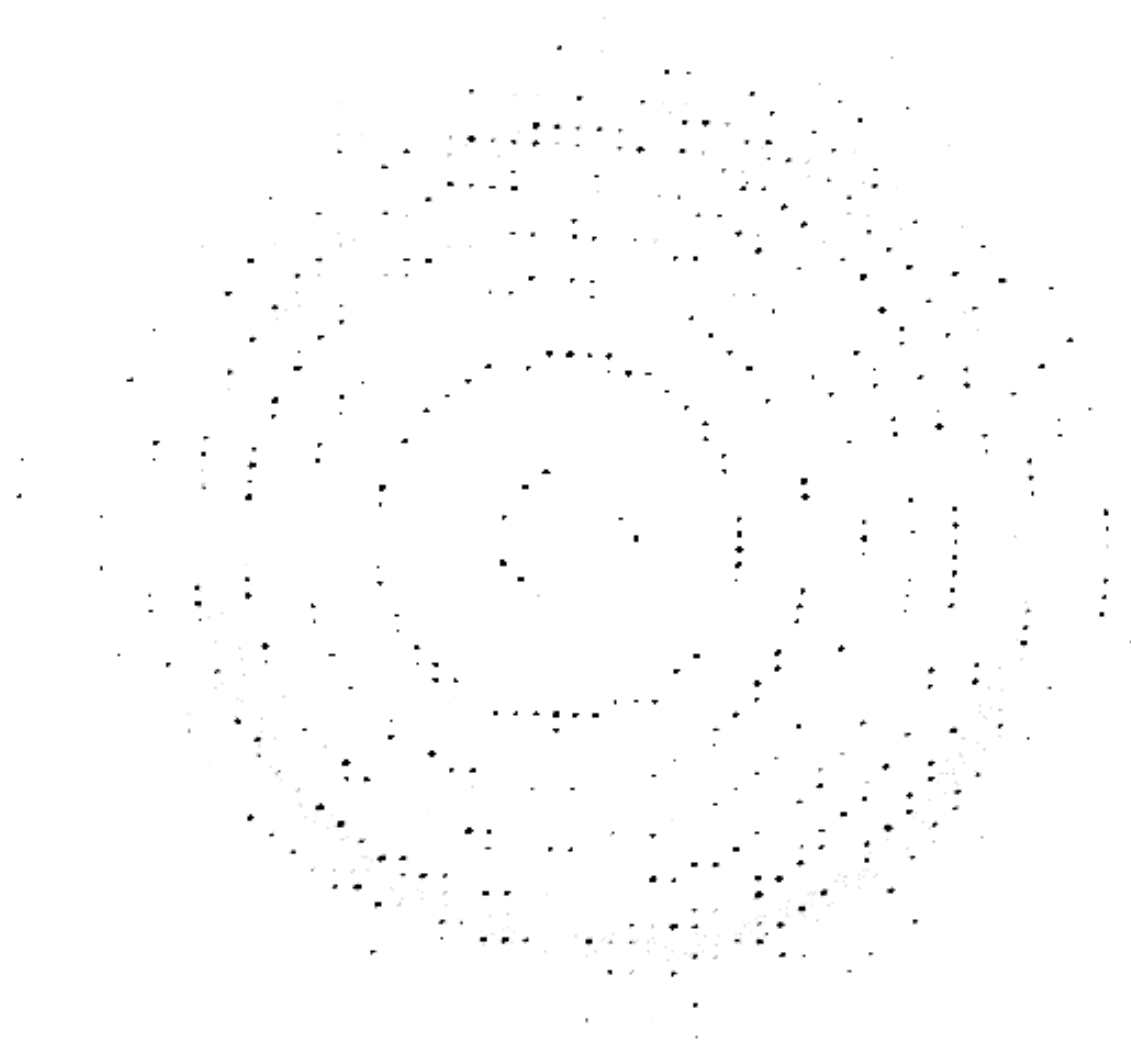
- **indexing** = finding the location of each reciprocal lattice point hkl
- **background estimation** = like subtracting the baseline, in 2D
- **integration of spots** = intensity is proportional to F^2
- **merging of partials** = One reflection may be split between two films.
- **Scaling** = If there is significant decay, then data is scaled in blocks of time.
- **Averaging of syms** = Symmetry-related reflections are averaged



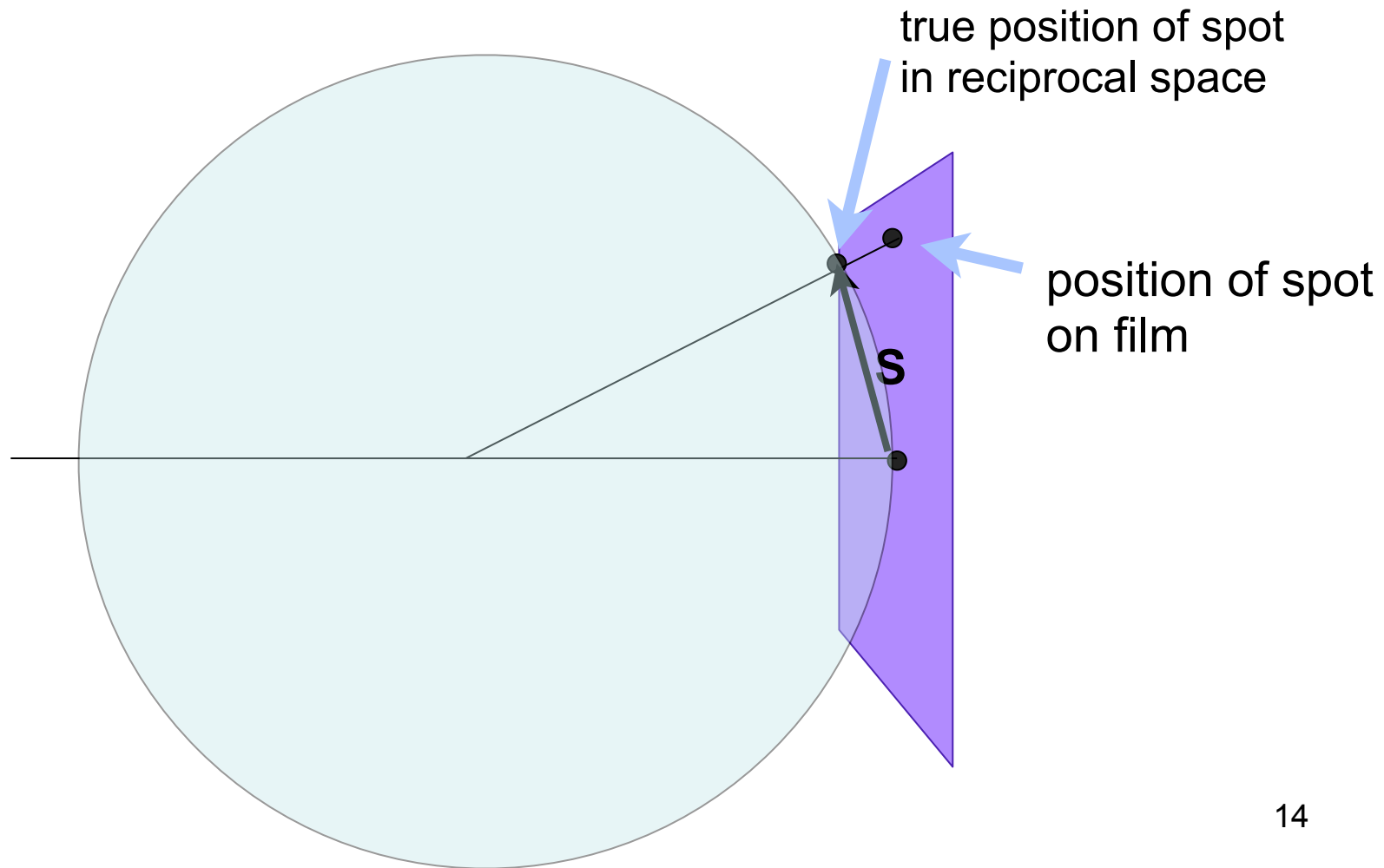
Background removal

- Identify the location of each spot
- Sum the average intensity around the spot.
- Subtract the average intensity.
- Integrate over the spot to get F_{hkl}





Calculating reciprocal lattice position from film and beam position



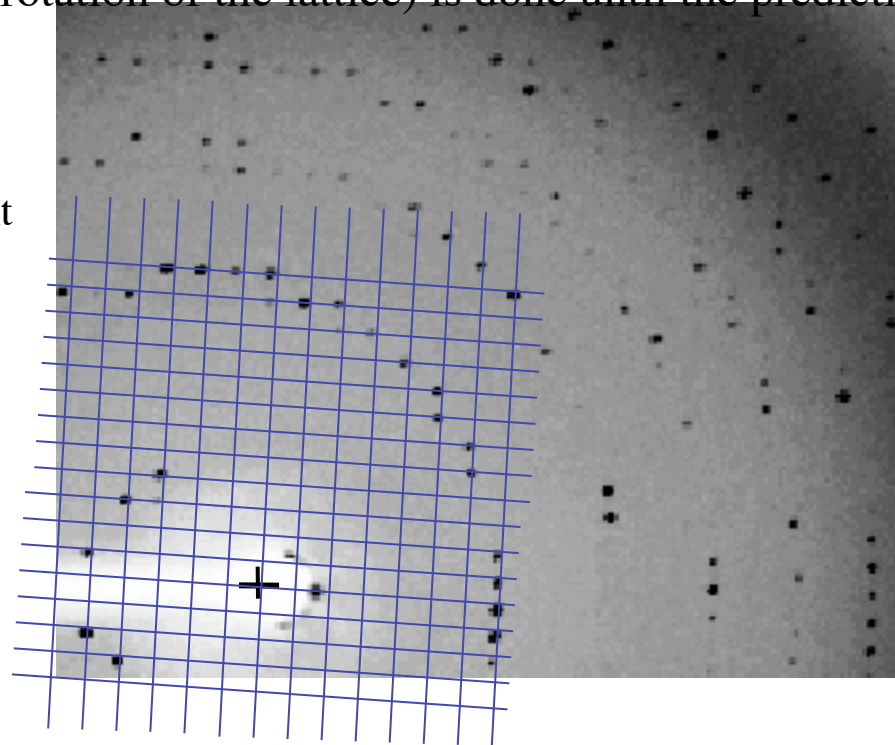
Indexing the data

A reciprocal lattice is initialized using the known cell dimensions.

Spots are predicted to be at the places where the lattice intersects the Ewald sphere.

A systematic search (rotation of the lattice) is done until the predictions match the observations.

Small refinements in the beam position might be required.

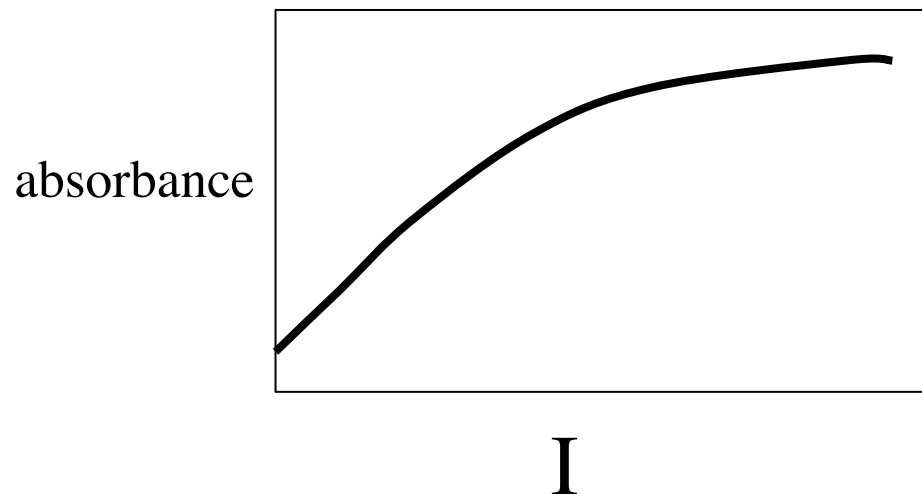


When the solution is found, every spot has an index (hkl).

Calibrating the film or detector.

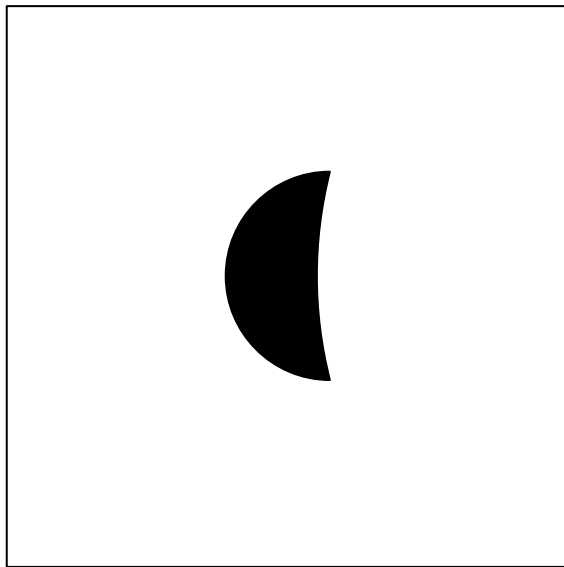
For photographic film, or any type of X-ray counter, a calibration curve has been pre-calculated.

The pixels are counted, multiplied by “I” from the calibration curve, to get $I(hkl)$ for each spot.



Merging partials

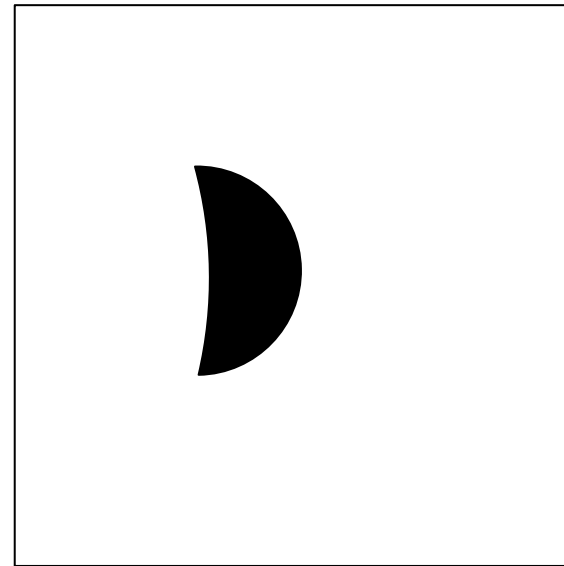
If films were switched while a spot was on the Ewald sphere, both copies (“partials”) are summed together to get $I(hkl)$.



Film 1

$\omega=153.0^\circ$

First half of spot hits the Ewald sphere.



Film 2

$\omega=153.5^\circ$

Other half of spot passes through.

Spot shape and mosaicity

Reflections are not *points in reciprocal space*, but *volumes*.

Reflections *spread* in all three **S** directions (a^* , b^* , c^*), because the crystal lattice is imperfect. This spread is called *mosaicity*.

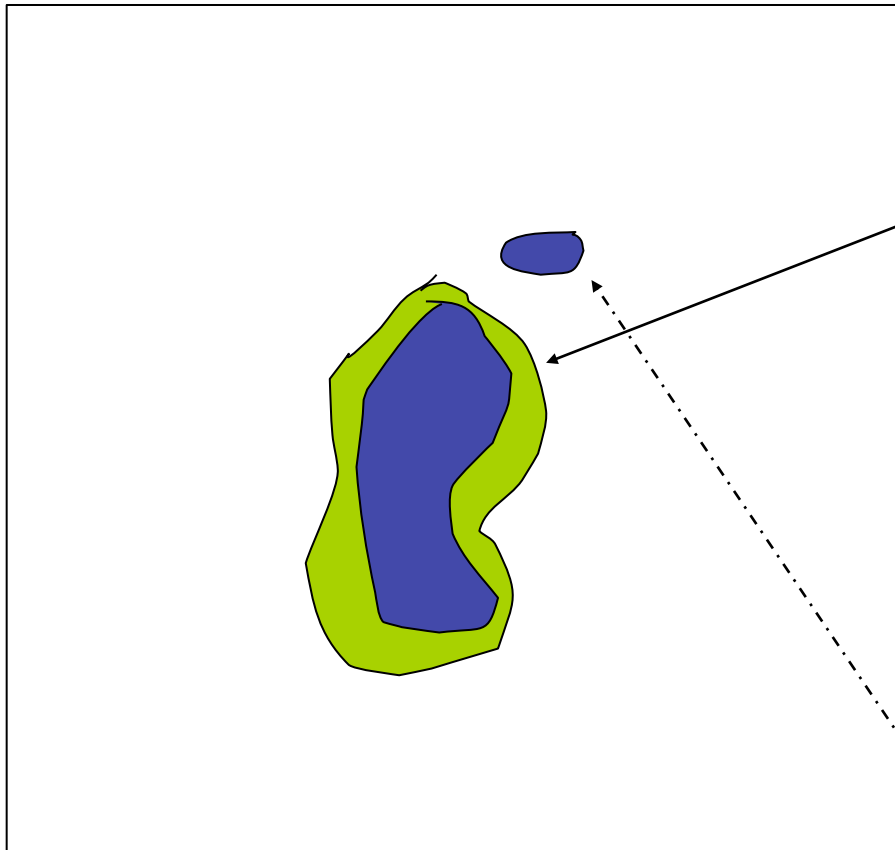
Reflections have *size* and *shape* on the film (or detector), because the beam and the crystal have size and shape.



beam

The spot shape is an image of the intersection of the beam and crystal.

Spot profiles in 3D⁻¹



Green is the “profile”
of the average spot,
summed over all spots
within a range in
diffractometer coords
 $\kappa, 2\theta, \omega$

Intensity of each spot
(blue) is the integrated only
within the spot profile.

This prevents
counting spurious
data like this.

Scaling syms within a dataset

Reflections may have errors in amplitude within a dataset because:

- Xray intensity varied.
- Film/detector sensitivity varied.
- Crystal orientation/ cross section varied with w .
- Crystal decayed over time.
- Exposure time varied.
- Background radiation varied.

Scaling assumes:

- (1) Symmetry-related reflections have the same amplitude
- (2) Reflections that were collected together are scaled together (i.e. applied the same scale factor).

$$R_{sym} = \frac{\sum_{hkl} |w_{hkl} F(hkl) - w_{R \cdot (hkl)} F(R \cdot (hkl))|}{\sum_{hkl} |w_{hkl} F(hkl)|}$$

Quality of the data set =

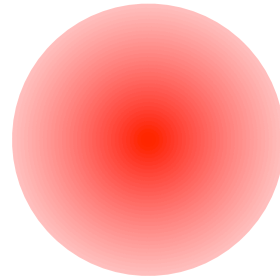
a sym op

Example Structure Factor file

```
data_rlpkqsf
#-----
_audit.revision_id      1_0
_audit.creation_date    2003-07-15
_audit.update_record    'Initial release'
#
loop_
_refln.wavelength_id
_refln.crystal_id
_refln.index_h
_refln.index_k
_refln.index_l
_refln.F_meas_au
_refln.F_meas_sigma_au
_refln.status
1 1  -39   0  26      70.300      34.700      0
1 1  -39   0  27     158.300      25.740      0
1 1  -39   1   1     156.000      15.800      0
1 1  -39   1  25      54.100      23.690      0
1 1  -39   1  26     201.400      11.450      0
1 1  -39   2  25     151.900      11.970      0
1 1  -39   3  22     202.800      22.730      0
1 1  -38   0  26      75.900      37.400      0
```

Structure Factors are deposited in the PDB (www.rcsb.org) along with the atomic coordinates.

Scattering factor of an atom



An atom is a spherically symmetrical cloud of electron density which is densest in the center.

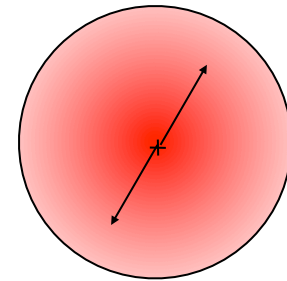
By integrating over the electron cloud, we get the Fourier transform of the atom.

$$f(S) = \int_{atom} \rho(r) e^{i2\pi S \cdot r} dr$$

If we define r to be a vector relative to the center of the atom, then $f(S)$ can be thought of as a single wave coming from the center of the atom.

Scattering from a centro-symmetric object has phase of the center position

The imaginary part cancels out because
 $\sin(-2\pi\mathbf{S}\cdot\mathbf{r}) = -\sin(2\pi\mathbf{S}\cdot(-\mathbf{r}))$



Proof

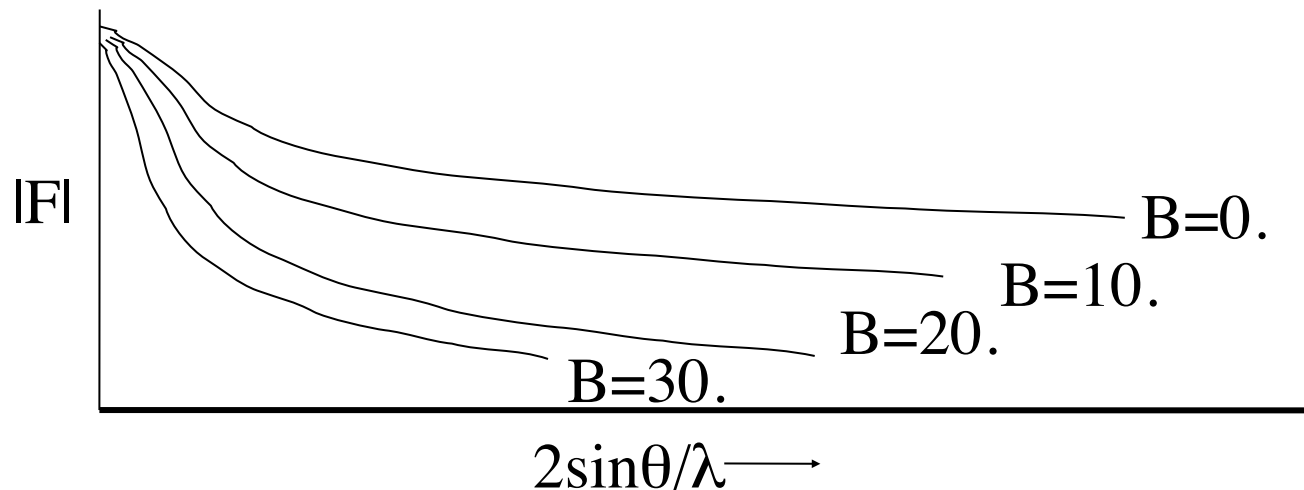
$$\begin{aligned} e^{i2\pi\mathbf{S}\cdot\mathbf{r}} &= \cos 2\pi\mathbf{S}\cdot\mathbf{r} + i \sin 2\pi\mathbf{S}\cdot\mathbf{r} \\ + e^{i2\pi\mathbf{S}\cdot(-\mathbf{r})} &= \cos 2\pi\mathbf{S}\cdot\mathbf{r} - i \sin 2\pi\mathbf{S}\cdot\mathbf{r} \\ \hline e^{i2\pi\mathbf{S}\cdot\mathbf{r}} + e^{i2\pi\mathbf{S}\cdot(-\mathbf{r})} &= 2\cos 2\pi\mathbf{S}\cdot\mathbf{r} \end{aligned}$$

.....
net sine part is zero!

The amplitude of an atom depends on how much it moves.

The sharper the electron density distribution, the broader the scattering factor. The temperature factor, B , modifies the scattering factor by spreading out the electron density.

Correction factor for atomic scattering factor: $e^{-B \frac{\sin^2 \theta}{\lambda^2}}$



B is called the “Temperature factor”, even though temperature has little to do with it!

Fourier transform with B-factors

$$F(S) = \sum_{atoms\ g} f(g) e^{-B_g \frac{\sin^2 \theta}{\lambda^2}} e^{i2\pi S \cdot r_g}$$

We sum one wave for each atom. The amplitude of the scattering factor $f(g)$ depends on how many electrons that atom has. Each scattering factor, $f(g)$, is a positive and *real* (i.e. not complex) function of $\sin\theta/\lambda$.

Same equation, using Miller indices **h**:

$$F(h) = \sum_{atoms\ g} f(g) e^{-B_g \frac{\sin^2 \theta}{\lambda^2}} e^{i2\pi h \cdot x_g}$$