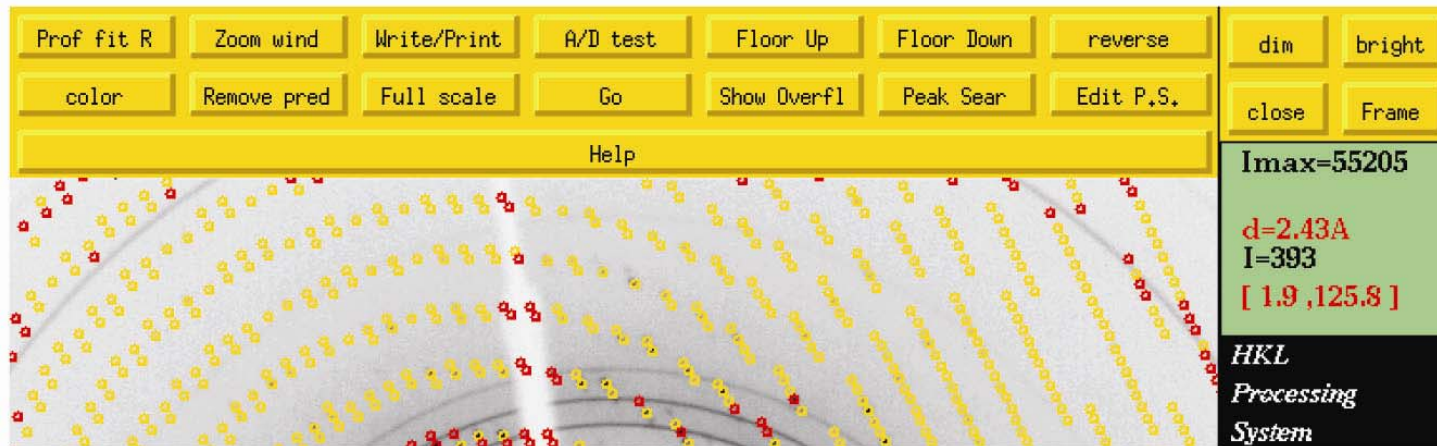


# X-ray Data Collection

Bio5325  
Spring 2006



# Obtaining $I_{hkl}$ and $\alpha_{(Ihkl)}$ from Frame Images

**Braggs Law** -predicts conditions for in-phase scattering by equivalent atoms lying in planes that transect a crystal.

**Model** –predict the occurrence of particular reflections (hkl's) on specific data frames and their locations.

**Estimate** –the intensity ( $I$ ) and associated error ( $\sigma$ ) for each predicted spot. Diffraction data are reduced to a list of integrated intensities and sigma values identified by the indices  $I_{(hkl)}$ .

**Merge and Scale** –symmetry-equivalent intensities are typically merged, yielding one estimate of  $I_{(hkl)}$  per unique reflection.

# The Oscillation Method

**For Each Frame-** predict reflections that will be present. Only those currently satisfying Braggs law—lying on the 2D surface of the Ewald sphere.

**Predict Where Reflections Will Occur on the Screen:**

- X center, Y center
- Detector distance from crystal
- Detector orientation w.r.t. incident beam
- Convention for mapping detector pixels onto lab frame of reference
- Extrapolate along vector from crystal (center of Ewald sphere) to reflection on sphere surface, continuing to the detector surface

# The Oscillation Method

## **Predict Which Reflections Will Be Present:**

- $1/\lambda$  gives curvature of the Ewald sphere
- $a^*$ ,  $b^*$ ,  $c^*$  and  $\alpha$ ,  $\beta$ ,  $\gamma$  give spacings of reciprocal lattice
- 3 angular parameters describe crystal orientation in real space
- Mosaicity describes “volume” of each lattice point

## **Refine Parameters to Fit Predictions With Observed Reflections—observables used to guess parameters:**

- Layer lines—measure spacings, adjust mosaicity
- Lunes
- Zones

# The Oscillation Method

## Auto-indexing of Frames:

- Start with strongest reflections, check predictions for these and other reflections during parameter refinement
- Project ( $X_{\text{frame}}$ ,  $Y_{\text{frame}}$ , distance) back onto Ewald's sphere in order to guess  $x, y, z$  (real space) by trial and error
- Pick 3 spots in different quadrants of the detector. Assign all possible (hkl) values consistent with guessed cell parameters
- Calculate 3x3 matrix that maps 9 variables ( $x, y, z$ ;  $x_2, y_2, z_2$ ;  $x_3, y_3, z_3$ ) onto 9 Miller indices ( $h, k, l$ ;  $h_2, k_2, l_2$ ;  $h_3, k_3, l_3$ )
- Apply this trial matrix to all observed reflections --> hkl values
- Judge success by how close to INTEGER VALUES the resulting hkl's are for each guess of unit cell parameters and orientation.

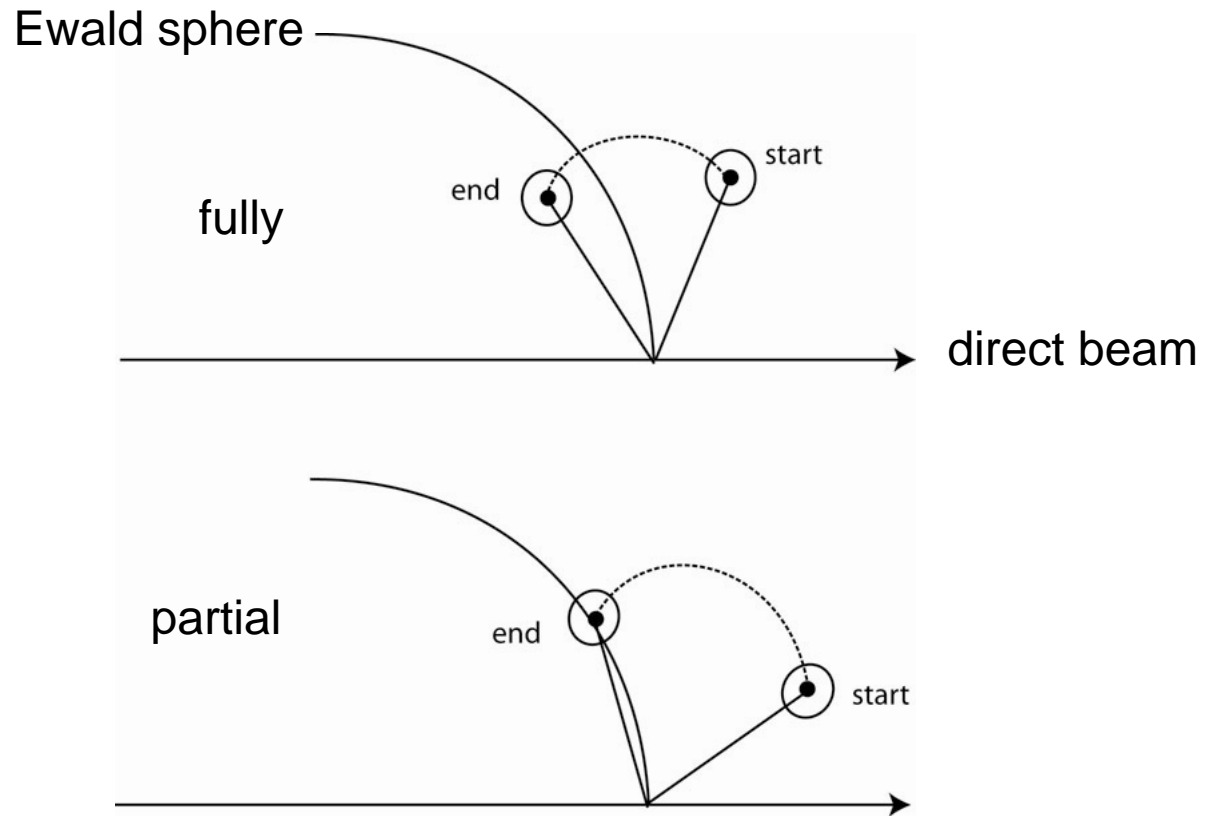
# The Oscillation Method

## Auto-indexing of Frames (cont.):

- Indexing matrix can be decomposed into a rotation matrix (crystal orientation) and orthogonalization matrix (unit cell parameters)
- If – cell parameters and indexing scheme (9 hkl numbers) are close to  $\alpha=\gamma=90$  and  $\beta=90$  or  $\beta=120$ 
  - and perhaps  $a=b$
  - and perhaps  $a=b=c$then can consider a higher symmetry Laue group
- Higher symmetry means fewer independent parameters to refine.
- Can now calculate  $x,y,(frame)$  for all possible hkl's

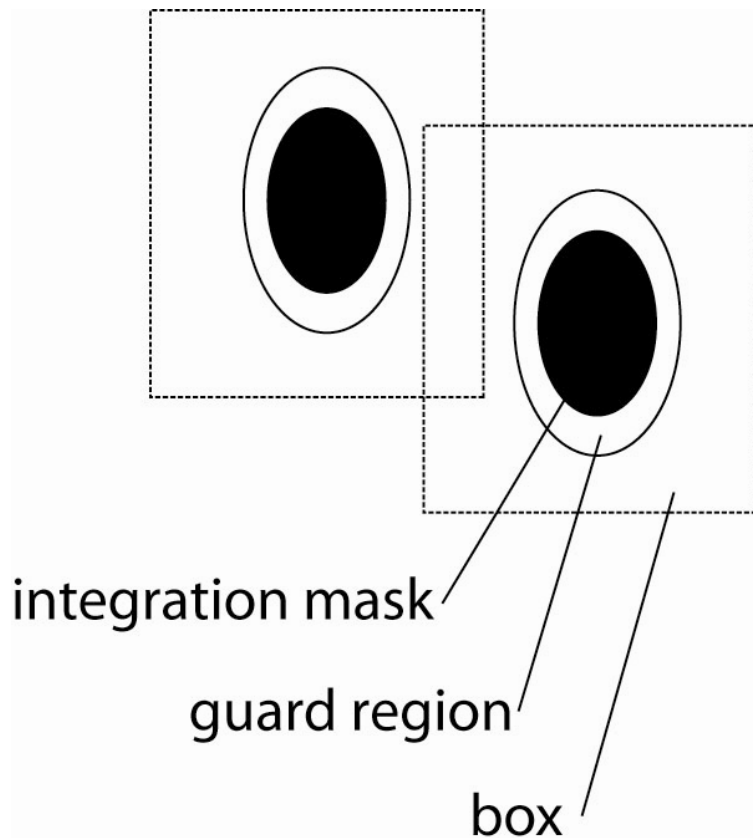
# The Oscillation Method

**Not all reflections are fully recorded on a single frame:**



# The Oscillation Method

## Integration of Reflections:



**Background-** the average pixel intensity in the box, excluding the guard region and reflection (assuming no surrounding reflections in the box).

**Total Counts-** sum of pixel values within the integration mask.



# The Oscillation Method

## Reflection Integration:

Simple:  $I = T - B$ , preferred for strong reflections

Profile fitting: shapes of neighboring reflections similar and dependent on geometry, crystal disorder, detector distortion, etc.

DENZO and other modern integration programs learn reflection profiles of strong reflections in various regions of the image. Applies local profile shape (integration mask) during integration of reflections in each detector region.

# The Oscillation Method

**DENZO: a widely used x-ray data integration program.**

- Obtain estimates of  $I_{hkl}$  and  $\sigma_{(hkl)}$ .
- Refines cell, orientation, and laboratory parameters (exp. geometry) based on fit of reflection centroids to predicted positions.

**SCALEPACK: scales and merges equivalent reflections to produce a “reduced” data set.**

- Refines scale factors  $ke^{-BS^2}$  per frame, and (optionally) the unit cell parameters and mosaicity on a per frame or “per pack” basis.
- Refinement target: does each scaled  $I_{hkl}$  multiplied by its calculated partiality agree with the average  $I_{hkl}$  of equivalent reflections?

# The Oscillation Method

**Error (sigma) Weighted Mean Intensity of Equivalent Reflections:**

$$\bar{I}_{hkl} = \frac{\sum \frac{1}{\sigma_{meas}^2 \cdot k^2 \cdot e^{-S^2 B_{frame}}} \cdot I_{meas}^{hkl} \cdot k_{frame} \cdot e^{-S^2 B_{frame}}}{\sum \frac{1}{\sigma_{meas}^2 \cdot k^2 \cdot e^{-S^2 B_{frame}}} \cdot I_{meas}^{hkl} \cdot k_{frame} \cdot e^{-S^2 B_{frame}}}$$

**Weighted error (sigma):**

$$\sigma^2 = \frac{\sum \frac{1}{\sigma_{meas}^2} \cdot \sigma_{meas}^2}{\sum \frac{1}{\sigma_{meas}^2}} = \frac{N}{\sum \frac{1}{\sigma_{meas}^2}}$$

# The Oscillation Method

## Lorentz and Polarization Corrections:

- Lorentz Correction—reciprocal lattice points further from the rotation axis travel through Ewald sphere faster than low angle reflections. Consequently, high angle reflections are in the reflecting condition for a shorter % of exposure time.
- Polarization—the resolution dependence of scattering efficiency is affected by polarization of the x-ray source, for example, by a crystal monochromator. Beam-specific parameters are known for individual beamlines at synchrotron light sources and specified on input to DENZO and other data reduction programs.

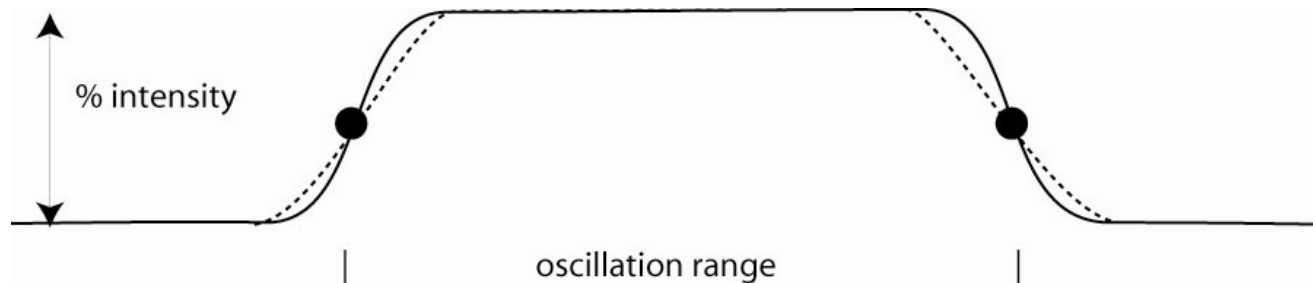
# The Oscillation Method

## Frame-specific scale factors $k$ and $B$ :

- $k$  is frame-specific global scale factor for intensities.  $B$  is the resolution dependent fall off in average reflection intensity ( $e^{S^2 \cdot B}$ ).
- Minimize unweighted agreement between average and measured  $I_{hkl}$   
e.g.,

$$\min Q = \sum_{hkl} \sum_{meas} \left( k_{frame} \cdot I_{meas}^{hkl} \cdot e^{-S^2 B_{frame}} - \overline{I_{hkl}} \right)^2$$

## Post-refine Cell Parameters and Mosaicity:



- dotted line is higher mosaicity.

# The Oscillation Method

## How good is our statistical model of the error?

- DENZO returns empirical/Poisson estimates of sigma ( $\sigma$ ).
- What is the expected difference between two measurements of equivalent reflection intensities?

$$\sigma^2(I_1^{meas} - I_2^{meas}) = \sigma^2(I_1) + \sigma^2(I_2)$$

- Chi square statistic:

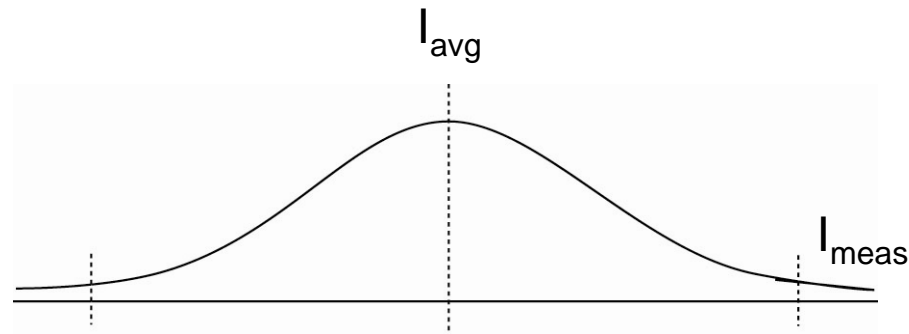
$$\chi^2 = \frac{1}{N} \sum \frac{(\bar{I}_{hkl} - I_{hkl}^{meas})^2}{\sigma^2 - \sigma_{meas}^2}$$

- Adjust scale factor of sigmas so that Chi square  $\approx 1.0$  for all (hkl)

# The Oscillation Method

## Rejecting Outliers:

- If statistical model predicts errors correctly then we are entitled to reject individual measurements that are unlikely to be true samples from the same population (i.e., outliers).



- Gaussian distribution with  $\sigma = \sqrt{\sigma^2 + \sigma_{meas}^2}$
- Reject Reflections if:  $\frac{|I_{hkl} - I_{meas}|}{\sqrt{\sigma^2 + \sigma_{meas}^2}} > 3.5$  ,for example

# The Oscillation Method

## Reported Statistics:

- Multiplicity- total number of observations divided by number of unique (hkl)'s. Reported as function of resolution.

- % completeness

- $R_{\text{sym}}$  (linear): 
$$\frac{\sum_{hkl} \sum_{meas} | \overline{I_{hkl}} - I_{meas}^{hkl} |}{\sum_{hkl} \sum_{meas} \overline{I_{hkl}}}$$

- I/sigma vs. resolution—a signal-to-noise ratio
- $R_{\text{sym}}$  (quadratic) is the R-factor per resolution bin.
- If  $R_{\text{sym}}$  is poor and intensity images/shapes look good, question the assigned symmetry and/or the indexing solution for the (hkl)'s.
- Data quality is related to  $R_{\text{sym}}/\text{SQRT}(N)$ .
- Rule of thumb: Error on F  $\approx$  1/2 error on I, Error on  $\Delta F \approx$  error on I.



# The Oscillation Method

Output of X-ray Data Integration and Scaling:

- Merged data with one  $I_{(hkl)}$  and  $\sigma_{(hkl)}$  per unique reflection.
- Unmerged data preserves original signed (hkl) values and scaled measurements for anomalous scattering calculations.

Convert Intensities to Structure Factor Amplitudes (F):

- TRUNCATE (CCP4):

$$\sqrt{I_{hkl}} \rightarrow F_{hkl} \qquad \frac{\sigma(I)}{2F} \rightarrow \sigma(F)$$