

---

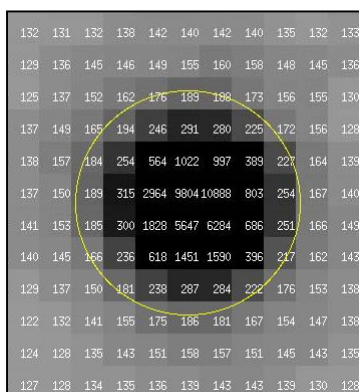
HKL2000/HKL3000

Dominika Borek  
UT Southwestern Medical Center at Dallas

---

## Data processing

- ▶ X-ray data processing = from the detector output to the estimate of structure factor amplitudes squared



$$\Rightarrow |\mathbf{F}|^2 \pm \sigma_{|\mathbf{F}|^2}$$

## Phase information

Isomorphous differences	$I_{hkl}^P - I_{hkl}^{PH}$
Anomalous differences	$I_{hkl} - I_{\bar{h}\bar{k}\bar{l}}$
Dispersive differences	$I_{hkl}^{\lambda_1} - I_{hkl}^{\lambda_2}$

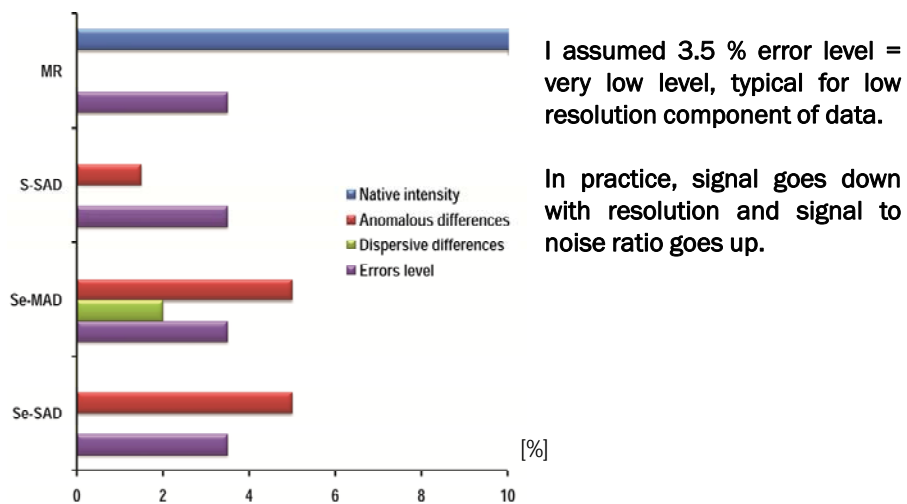
$$\text{information} = \frac{\text{signal}}{\text{noise}}$$

X-ray quantum nature	- random effects
Assumption and approximations	- systematic effects
	- outliers

▶ 3

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

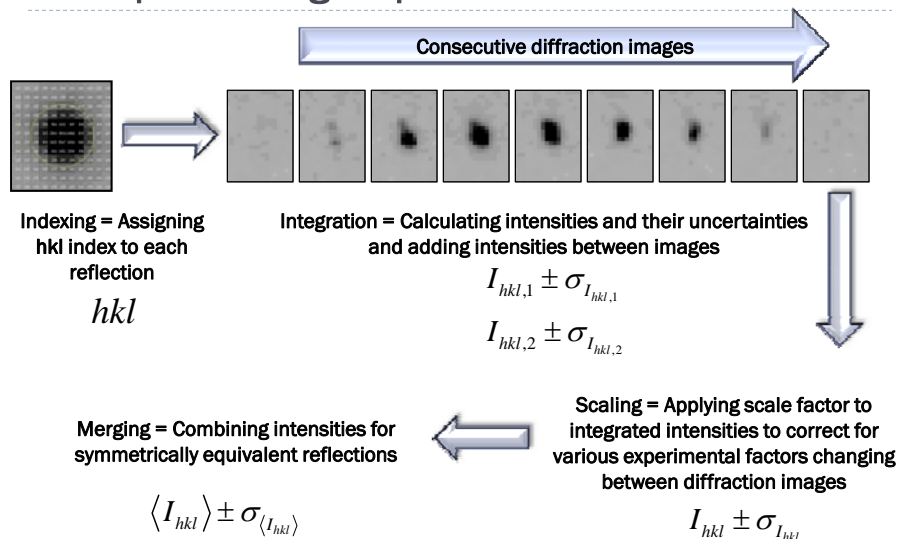
## Why do we care about signal to noise ratio?



▶ 4

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Data processing steps



▶ 5

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Data processing in HKL2000/HKL3000

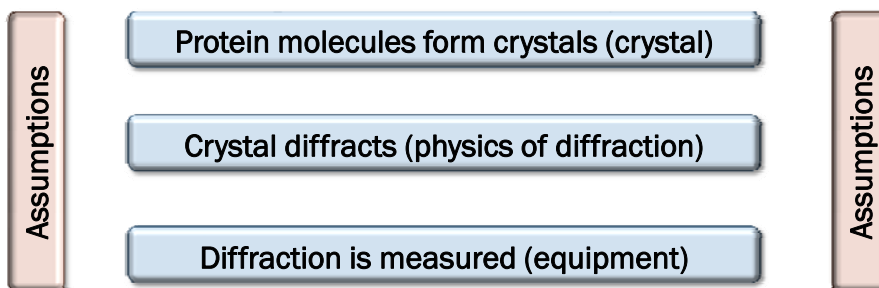
- I. Detector description (site file)
- II. Autoindexing (Denzo) and visual assessment (XDisplayF)
- III. Refinement of experimental parameters and optimization of integration parameters (Denzo)
- IV. Integration (Denzo)
- V. Scaling (Scalepack)
- VI. Merging and statistical assessment (Scalepack and HKL2000)

▶ 6

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Assumptions

---



---

▶ 7

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Data Model - Crystal

---

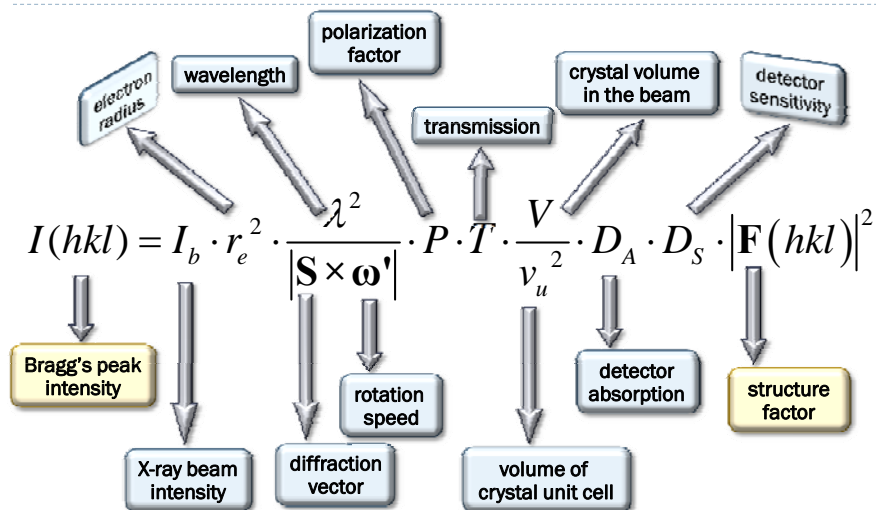
- ▶ Crystal  $\equiv$  ideal space group symmetry in a perfectly ordered infinite crystal lattice
- ▶ Deviations:
  - ▶ Finite crystal size
  - ▶ Ideally imperfect crystal (no double scattering and no extinction)
  - ▶ Observable mosaicity
  - ▶ Multiple lattices due to phase transition
  - ▶ Twinning
  - ▶ Pseudosymmetry

---

▶ 8

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Data model - diffraction

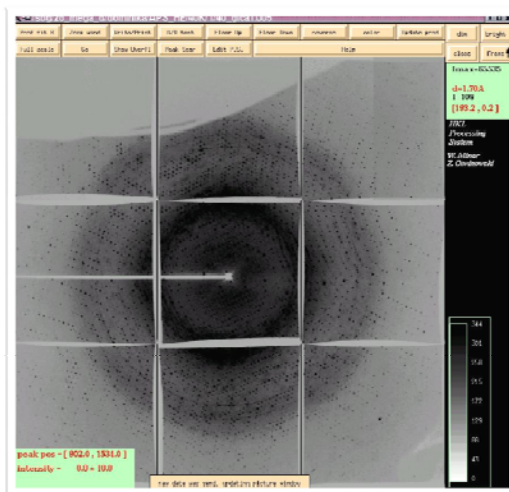


▶ 9

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Data Model – Experimental Setup

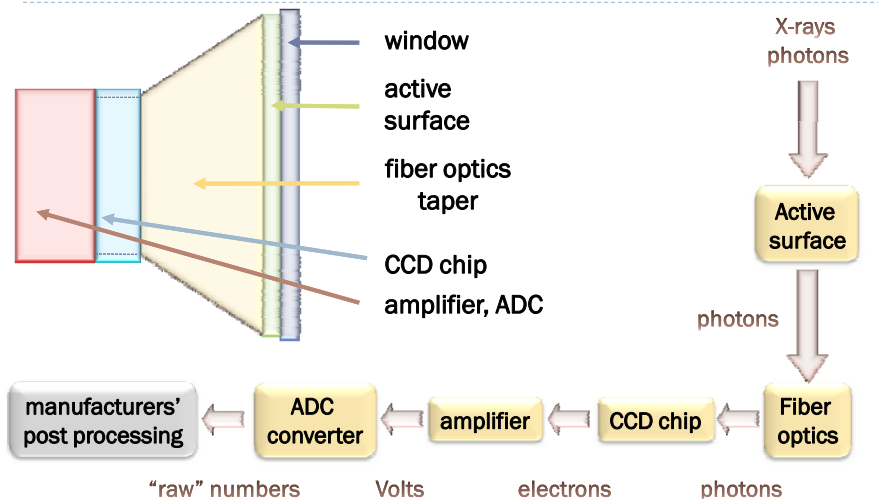
- ▶ Obscuration:
  - ▶ Beam stop
  - ▶ Cryo-cooling
  - ▶ Goniostat
- ▶ Always remove beam stop shadow!
- ▶ Goniostat shadow rotates with the crystal (use Reject Low Value)



▶ 10

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Data Model – Experimental Setup



▶ 11

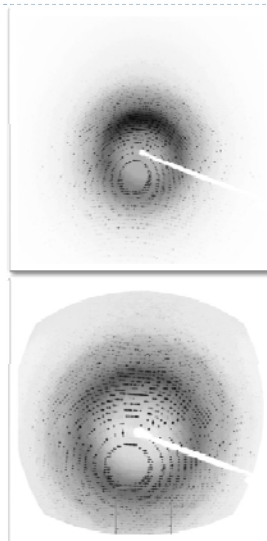
APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Detector description – Site file

The site file contains numerical parameters describing how reciprocal space is distorted in the diffraction image. These parameters belong to two groups: one describing the geometry of distortion and an optional second, describing sensitivity of each pixel in the detector.

### Wrong site file:

- ▶ misindexation, misprediction of spots' positions, wrong refinement of processing parameters
- ▶ wrong correction of intensities due to wrong values of pixels' sensitivity



▶ 12

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Indexing

---

### Assigning hkl index to diffraction maxima (spots)

#### REQUIREMENTS:

- ▶ approximate description of detector geometry
  - ▶ x beam, y beam !!!
  - ▶ distance
  - ▶ detector orientation
- ▶ list of peaks free of artifacts (peak search)
- ▶ proper procedure (spots separation, oscillation range)

---

▶ 13

*APS Data collection workshop and CCP4 school, June 7 - June 15, 2011*

## Autoindexing procedure

---

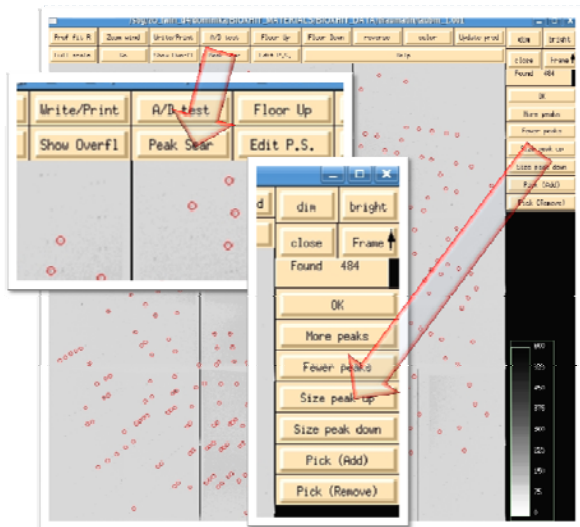
- ▶ peak search
- ▶ autoindexing in primitive lattice
- ▶ choice of Bravais lattice (lattice symmetry)
- ▶ reindexing to standard symmetry
- ▶ if more than one crystal involved – checking the consistency of indexing between crystals
  - ▶ needed only for some space groups
  - ▶ after separate scaling of data from crystals

---

▶ 14

*APS Data collection workshop and CCP4 school, June 7 - June 15, 2011*

## Peak Search Procedure



It finds the strongest intensity peaks

500 spots – very good

50 spots – so-so

10 spots – absolute minimum

Spot size in 'peak search' procedure is specific for the detector type

Sometimes it makes sense to increase the spot size, particularly for laboratory detectors when too small spot size causes rejection of peaks resulting in smaller number of peaks used in autoindexing procedure.

▶ 15

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Autoindexing in primitive lattice and choice of higher symmetry Bravais lattice (if possible)

Bravais Lattice	Parameters
primitive cubic	76.36%, 85.83, 86.25, 85.96, 105.22, 76.78, 86.69, 86.12, 86.13, 86.12, 86.00, 86.00, 86.00
I centred cubic	74.11%
F centred cubic	7.61%
primitive rhombohedral	2.38%
primitive hexagonal	15.07%, 85.80, 86.55, 85.98, 74.48, 76.70, 113.48, 86.24, 86.24, 85.98, 86.00, 86.00, 139.80
primitive tetragonal	70.27%, 80.89, 86.20, 89.83, 113.46, 109.22, 109.22, 86.20, 86.20, 85.82, 86.00, 86.00, 86.00
I centred tetragonal	1.16%, 89.51, 104.37, 94.70, 89.22, 89.84, 89.86, 101.84, 101.84, 94.70, 86.00, 86.00, 86.00, 86.00
primitive orthorhombic	20.28%, 85.80, 85.83, 86.55, 113.46, 74.48, 20.70, 85.80, 85.83, 86.55, 86.00, 86.00, 86.00, 86.00
C centred orthorhombic	13.18%, 83.81, 126.79, 85.98, 117.25, 81.76, 81.47, 85.83, 126.79, 85.98, 86.00, 86.00, 86.00, 86.00
I centred orthorhombic	0.28%, 94.70, 89.51, 104.37, 89.96, 89.32, 89.94, 94.70, 89.51, 104.37, 86.00, 86.00, 86.00
F centred orthorhombic	1.18%, 84.20, 144.16, 144.26, 87.27, 88.64, 88.45, 94.70, 144.16, 144.25, 86.00, 86.00, 86.00
primitive monoclinic	15.21%, 85.83, 85.80, 86.55, 105.52, 113.46, 109.22, 85.83, 85.89, 86.55, 86.00, 113.49, 86.00
C centred monoclinic	8.02%, 146.69, 89.51, 94.70, 86.06, 131.84, 86.00, 146.69, 89.51, 94.70, 86.00, 131.84, 86.00
primitive triclinic	0.00%, 85.50, 85.93, 86.55, 113.40, 105.52, 109.22

▶ Refine first parameters describing the experiment before moving to a higher symmetry Bravais lattice

▶ 16

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011



## Why autoindexing may fail?

### Peak search list problems:

- less than 10 diffraction peaks
- twin crystals, ice, satellite crystals



Resolution limits  
Manual editing  
 $I/\sigma(I)$  selection

### Detector description:

x beam, y beam



Use refined values from a previous experiment  
Direct beam exposure  
Center of powder diffraction (ice)

### Unknown system

- spindle direction
- correction files



Site documentation  
Four possibilities (only one works when indexing in 3D-mode)

▶ 17

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Why autoindexing may fail?

- ▶ Procedure problems:

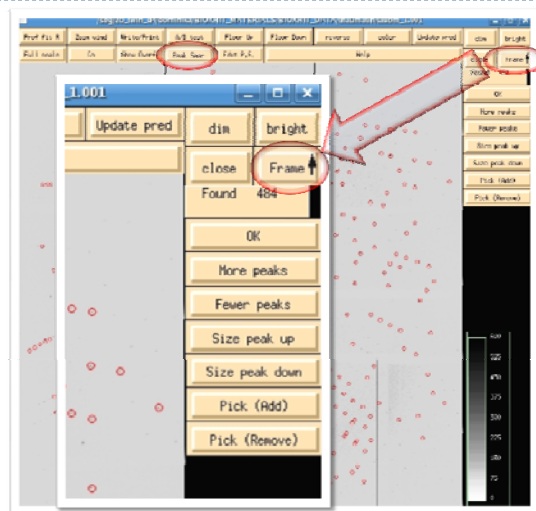
$$\textit{longest vector} = \frac{\textit{distance} \cdot \lambda}{\textit{spot size}}$$

- ▶ spot size - reduce spot radius
- ▶ distance - re-collect image at longer distance
- ▶ mosaicity too large - reorient the crystal if only one axis is affected
- ▶ rotation range too large - decrease for large unit cells, but even if indexing works there may be too many overlaps

▶ 18

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Multiframe indexing – Peak Search

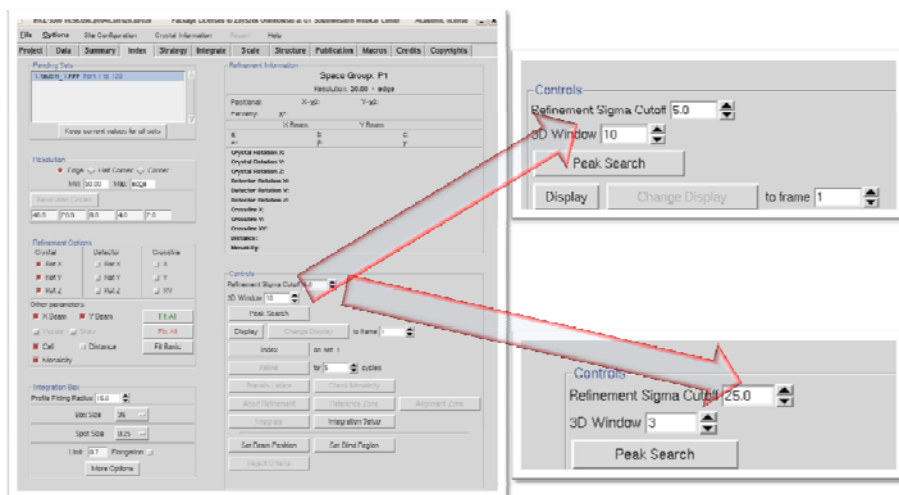


After pressing **Peak Search** in the first frame, move a cursor to the field **Frame** and press the middle mouse button. It will search peaks in the next frame and diffraction image will change to the next one. Repeat operation for the next frames if desired.

▶ 19

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Advanced option - Multiframe indexing



▶ 20

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Advance option - Multiframe indexing

---

- ▶ Peak search performed on multiple images
- ▶ 3D option in Index Tab (HKL2000) has to cover at least the same range of images as those used in peak search
- ▶ **Benefits:**
  - ▶ autoindexing possible if there are not enough spots in one frame
  - ▶ may resolve confusing diffraction patterns, e.g. multiple crystals, highly mosaic crystals etc.

---

▶ 21

*APS Data collection workshop and CCP4 school, June 7 - June 15, 2011*

## Refinement of parameters

---

- ▶ **Crystal:**
  - ▶ Orientation
  - ▶ Unit cell
  - ▶ Mosaicity
- ▶ **Beam:**
  - ▶ Focus parameters
- ▶ **Detector:**
  - ▶ Distance
  - ▶ Orientation
  - ▶ Position
  - ▶ Internal geometry

Parameters could be the same or different for consecutive images.

---

▶ 22

*APS Data collection workshop and CCP4 school, June 7 - June 15, 2011*

## Refinement - target

$$\chi_{total}^2 = \chi_x^2 + \chi_y^2 + \chi_p^2 \quad \leftarrow \text{Minimization of target function}$$

$$\chi_x^2 = \sum_{spots} \frac{(x_{predicted} - x_{observed})^2}{\sigma_x^2}; \quad \chi_y^2 = \sum_{spots} \frac{(y_{predicted} - y_{observed})^2}{\sigma_y^2}$$

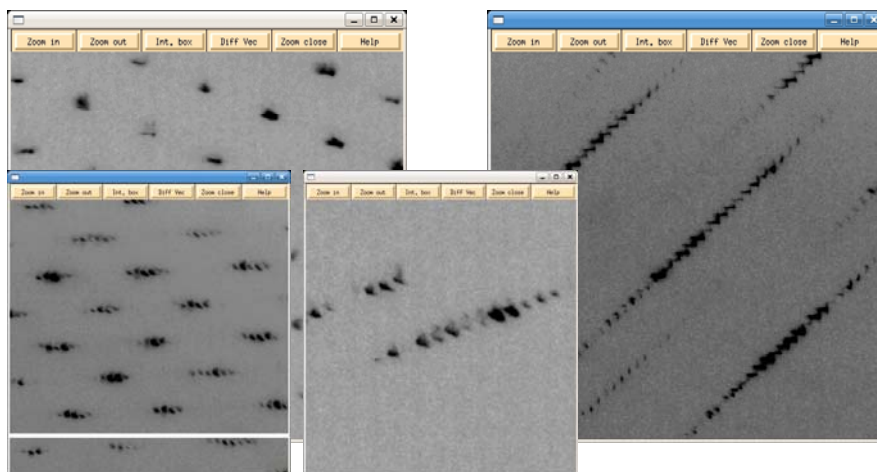
$$\chi_p^2 = \sum_{hkl} \frac{(P_{predicted} - P_{observed})^2}{\sigma_p^2}; \quad P = \frac{I_{hkl, frame}}{I_{hkl, total}}$$

The displayed values of  $\chi^2$  are divided by the number of observations

▶ 23

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Problems

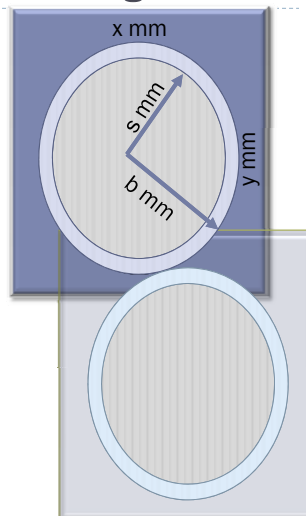


▶ 24

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Integration of diffraction peaks - Background

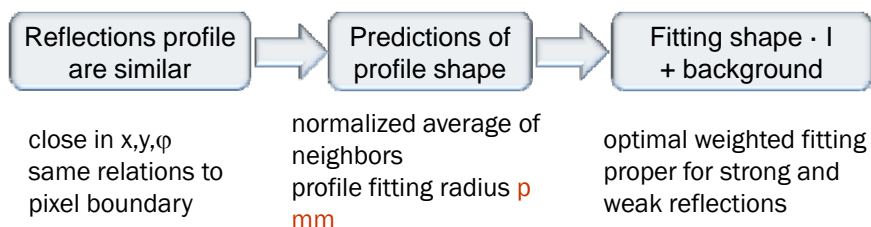
- ▶ - based on analysis of local environment of peaks – “box” (box  $x\_mm$   $y\_mm$  or  $i\_box$   $x\_pixels$   $y\_pixels$ )
  - ▶ Definition of spot area (spot radius  $s\_mm$ )
  - ▶ Background is outside of spot area (including other reflections) and outside of background radius  $b\_mm$
  - ▶ Background is analyzed for slope (linear variations with respect to position) and artifacts
- Spot and background are symmetric with respect to the center of the box.



▶ 25

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

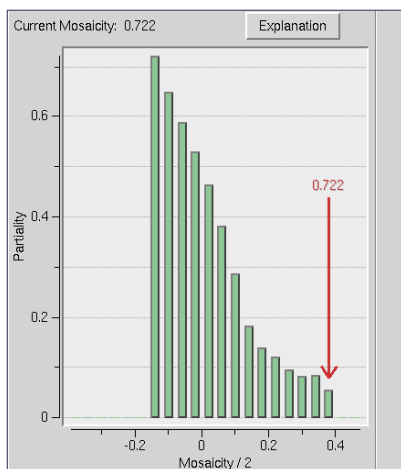
## Integration of diffraction peaks - Profiles



▶ 26

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Mosaicity



- ▶ reflections are sorted into 20 zones, which range from  $-(\text{mos}/2)$  to  $+(\text{mos}/2)$ .
- ▶ each zone represents the shortest angular distance of the center of the reflection from the surface of the Ewald sphere at the end of the oscillation range.
- ▶ only reflections that are single partials are included in the analysis.
- ▶ reflections in negative zones - the center of the reflection has already passed through the Bragg condition.
- ▶ the histogram should pass through 50% in the zero zone.
- ▶ if there are only a few reflections, the histogram may be quite choppy.
- ▶ weird shapes of histogram may indicate motor, spindle, or shutter problems.

▶ 27

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Scaling - definition

$$I(hkl) = I_b \cdot r_e^2 \cdot \frac{\lambda^2}{|\mathbf{S} \times \boldsymbol{\omega}'|} \cdot P \cdot T \cdot \frac{V}{v_u^2} \cdot D_A \cdot D_S \cdot |\mathbf{F}(hkl)|^2$$

Scale factor K

$$K = k_{\text{overall}} \cdot (k_{\text{Lorentz}} \cdot k_{\text{polarization}} \cdot k_{\text{detector}} \cdot k_{\text{absorption}} \cdot \dots)$$

From comparison of  
data to the atomic  
model

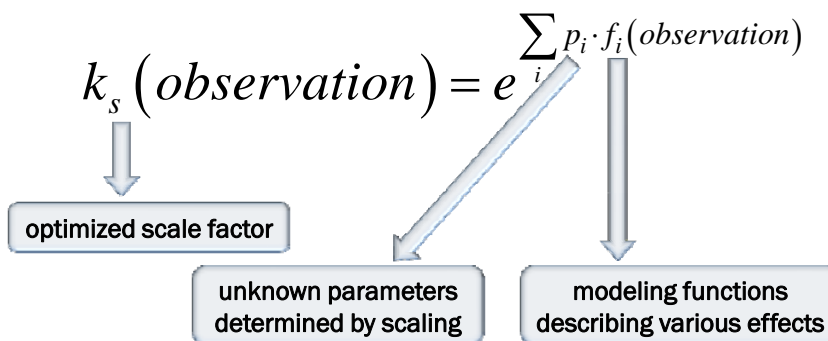
From calibration and  
diffraction geometry

From comparison of  
symmetry related  
reflections  
**SCALING**

▶ 28

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Scaling - exponential modeling



▶ 29

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Scaling - decay described by scaling B-factor

$$f_{pb,n} = \frac{|\mathbf{S} \cdot \mathbf{S}|}{2} \cdot \text{dose}^n$$

→ B-factor as a continuous function of accumulated dose

$$f_{b_j} = \frac{|\mathbf{S} \cdot \mathbf{S}|}{2} \text{ for data in batch } j$$

$$f_{b_j} = 0 \text{ for other data}$$

→ Separate B-factor for every batch

▶ 30

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Scaling - correction for absorption

Modeling functions (spherical harmonics)

$$f_{as,lm} = \frac{1}{2} \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} (P_{lm}(\cos\theta_i) \sin(2\pi m\Phi_i) + P_{lm}(\cos\theta_o) \sin(2\pi m\Phi_o))$$

$$f_{ac,lm} = \frac{1}{2} \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} (P_{lm}(\cos\theta_i) \cos(2\pi m\Phi_i) + P_{lm}(\cos\theta_o) \cos(2\pi m\Phi_o))$$

“Pure” absorption  $\Rightarrow$  odd coefficients zero

odd coefficients non-zero  $\Rightarrow$  ? - slowly changing function

▶ 31

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Scaling – Advanced options

Correction for uneven rotation or uneven/unstable beam or shutter error

- ▶ When to use it? When there is an indication of any of these problems.
- ▶  $\chi^2$  vs. intensity is going up towards large intensities

How to use it?

- ▶ In Macros under During Scaling, add macro:
  - absorption exposure [number] [separate] e.g **absorption exposure 1**
  - ▶ number = frame width/mosaicity (round it down to an integer number)
  - ▶ problem with the number of parameters [2 (sin, cos terms) \* [number] \* number of frames] if „separate” option is used

▶ 32

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011



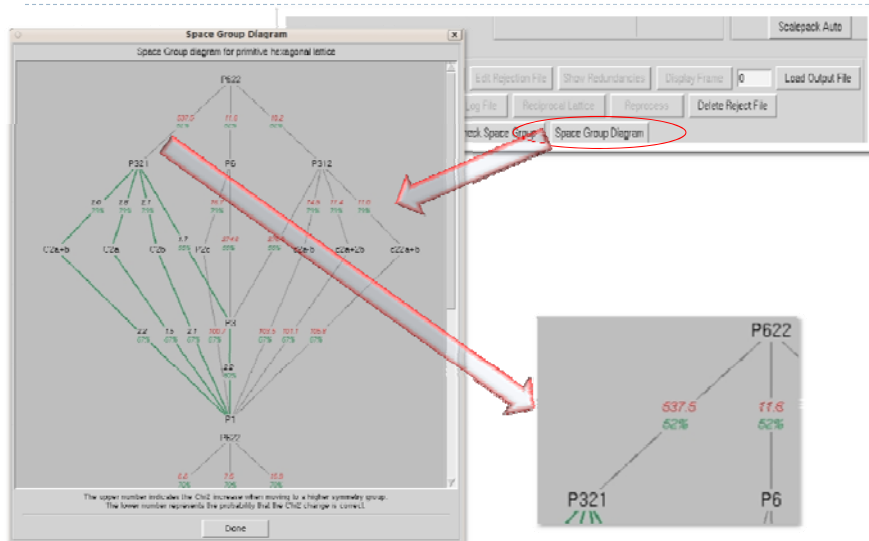
## Merging - analysis

- ▶ Determination of point group symmetry
  - ▶ metric pseudosymmetries and relative indexing of different crystals
- ▶ Parameters of error model (error scale factor, error systematic, rejection probability)
- ▶ Assessment of data quality
  - ▶ random events (signal-to-noise ratio)
  - ▶ non-random events (outliers, ice-rings, bad frames etc.)
  - ▶ non-isomorphism (radiation damage, pseudosymmetry)
- ▶ Assessment of data content (significance of anomalous signal, systematic absences, translational pseudosymmetry, pseudosystematic absences)

▶ 33

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Space Group Diagram



▶ 34

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Error model

Based on the  $\chi^2$  test we can adjust the error model:

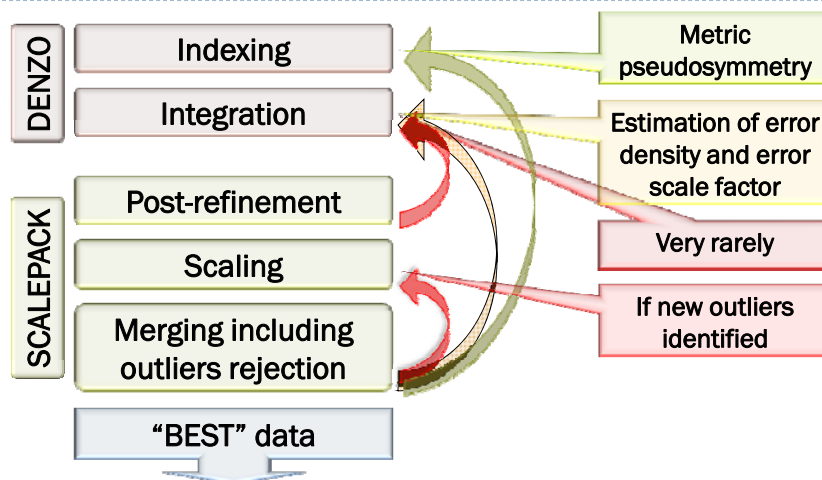
In HKL2000:

- ▶ error model (default value = 0.03)
  - ▶ change in resolution shells – be careful
  - ▶ if you have to go over 0.10 – something bad happened in the experiment
- ▶ scale factor (default value 1.3)
  - ▶ more impact at higher resolution
  - ▶ if you have to go over 2.0:
    - ▶ increase error density value in Denzo
    - ▶ non-isomorphism – accept  $\chi^2$

▶ 35

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Should we reprocess?



▶ 36

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

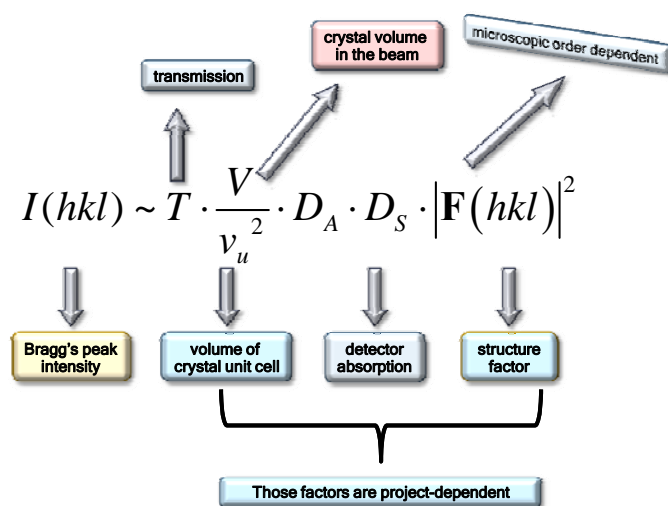
## Challenges in data processing

- ▶ Single crystal not sufficient to obtain complete data set
- ▶ Insufficient phasing signal
- ▶ Insufficient or anisotropic diffraction limit
- ▶ Non-isomorphisms
  - ▶ Induced by radiation damage
  - ▶ Induced by cryo-cooling within crystal
  - ▶ Between crystals
    - ▶ Mostly due to cryo-cooling variability, e.g. variable humidity and rate of cooling
- ▶ Problematic macroscopic order, e.g. twinning

▶ 37

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

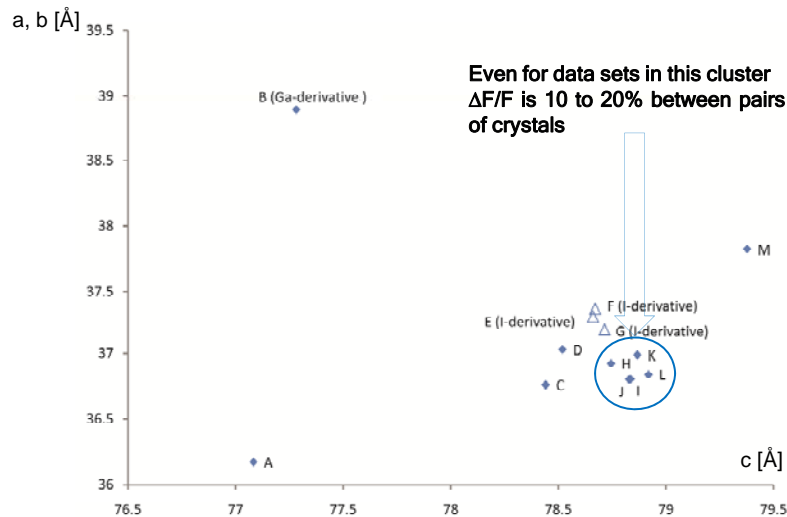
## Number of diffracted photons relative to the dose



▶ 38

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## “Easiest” – tetragonal lysozyme



▶ 39

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## How to identify non-isomorphism?

### Unit cell dimensions are not good indicator:

- ▶ In some cases the uncertainty in determination of unit cell parameters is large, so different unit cells may represent isomorphous crystals
- ▶ Exactly the same unit cell (within 0.01 Å for two axes in case of tetragonal lysozyme) may still have substantial non-isomorphism

### R-merge between data sets – better - but what kind of non-isomorphism???

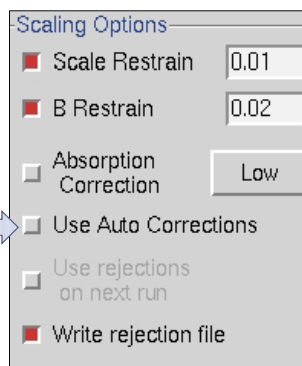
- ▶ It is a sum of contributions from experimental errors and non-isomorphisms. To estimate the non-isomorphism contribution, a very good model of experimental errors for data sets in question is needed.
- ▶ It does not define the source of non-isomorphism, e.g. does not differentiate between radiation-induced and crystal reproducibility non-isomorphisms.

▶ 40

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Automatic corrections option

- ▶ Works with complete data sets from macro-beam approach
- ▶ Data sets (if more than one) have to be processed in the same order as they were exposed, otherwise non-automatic.
- ▶ Still needs improvement for resolution worse than 3.0 Å (non-automatic)



▶ 41

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

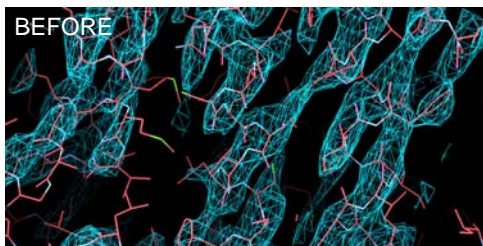
## Main features

- ▶ **Correction for anisotropic diffraction**
- ▶ **Informativity-based resolution limit (ellipsoidal-like resolution cuts)**
- ▶ **Radiation-damage correction**
  - ▶ **Extrapolation to zero-dose is not yet automatic**
- ▶ **Error-model adjusted automatically**
- ▶ **Estimates also internal non-isomorphism**
  - ▶ **If point group symmetry is too high it is equivalent to very high internal non-isomorphism, which at this point will be automatically estimated and included in error-model adjusting procedure.**
- ▶ **Much better outlier rejection (e.g. ice), particularly in the presence of radiation damage, anomalous signal *etc.***

▶ 42

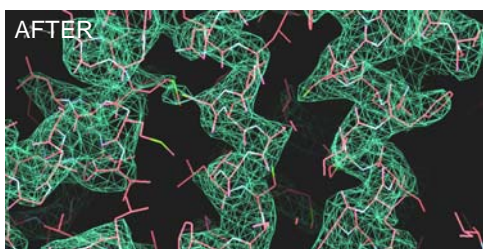
APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Anisotropy correction in action



Example of log file:

```
Anisotropic B factors
  0.000  -1.000  0.000
 -0.000  -0.000  -1.000
  1.000   0.000  0.000
 -57.888 -28.772 -19.262
```



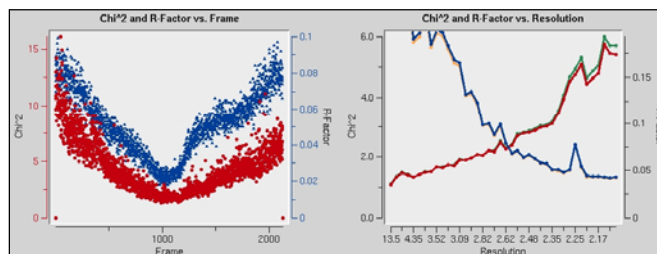
Works better than in other programs due to informativity-based cut, i.e. “bad” (outside elliptical resolution limit) reflections do not contribute.

▶ 43

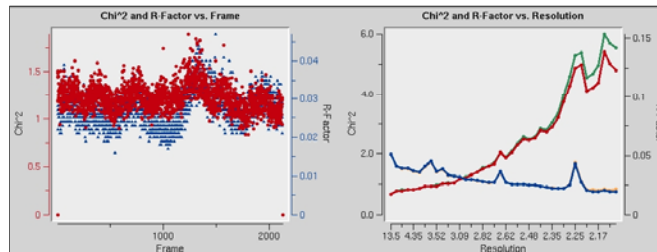
APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Error model adjusted automatically - I

BEFORE:



AFTER:

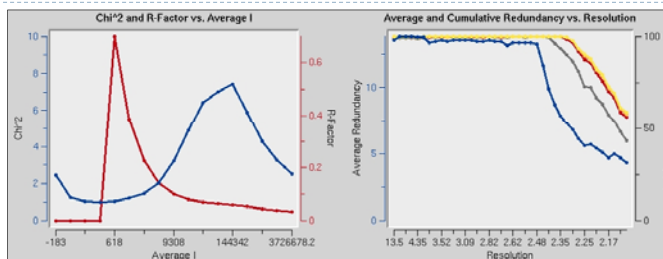


▶ 44

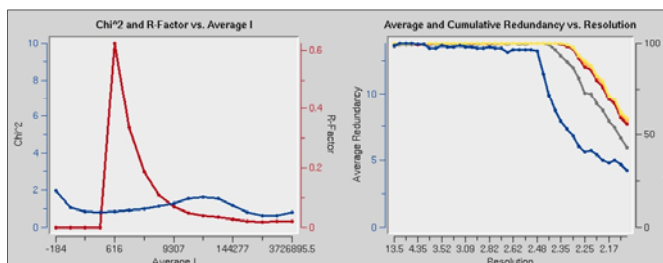
APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Error model adjusted automatically - II

BEFORE:



AFTER:



▶ 45

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

Resolution	Iso I	I z	I x	I y	Error	
50.00	6.15	4190.0	2055.7	2318.8	2629.2	240.4
6.15	4.88	2396.0	1497.7	1812.9	2212.8	134.3
4.88	4.27	3350.4	1172.9	1506.4	1955.9	187.7
4.27	3.88	2957.5	942.6	1276.4	1751.2	173.1
3.88	3.60	2353.6	769.6	1094.0	1579.1	144.1
3.60	3.39	1695.0	636.1	946.3	1432.3	114.2
3.39	3.22	1256.5	531.2	824.9	1305.7	97.7
3.22	3.08	882.1	447.3	723.8	1195.7	80.3
3.08	2.96	716.3	379.5	638.6	1099.1	73.6
2.96	2.86	541.3	323.9	566.1	1013.6	65.2
2.86	2.77	424.5	278.0	503.9	937.4	60.6
2.77	2.69	366.2	239.6	450.2	869.0	58.1
2.69	2.62	305.9	207.5	403.4	807.4	56.3
2.62	2.55	279.2	180.3	362.6	751.6	56.0
2.55	2.50	226.9	157.2	326.7	700.9	55.2
2.50	2.44	216.2	137.5	295.1	654.6	56.0
2.44	2.39	210.2	120.6	267.1	612.3	57.0
2.39	2.35	189.9	106.0	242.2	573.5	57.6
2.35	2.31	161.9	93.5	220.1	537.8	57.9
2.31	2.27	141.0	82.6	200.3	504.9	60.1
2.27	2.23	131.6	73.1	182.6	474.5	64.5
2.23	2.20	127.8	64.8	166.7	446.4	68.3
2.20	2.16	118.5	57.6	152.4	420.3	71.9
2.16	2.13	112.9	51.3	139.5	396.1	73.7
2.13	2.11	95.6	45.7	127.8	373.6	76.3
2.11	2.08	92.6	40.8	117.3	352.7	80.5
2.08	2.05	88.8	36.5	107.7	333.2	84.5
2.05	2.03	76.6	32.7	99.0	315.0	86.0
2.03	2.00	61.3	29.3	91.2	298.0	88.9
2.00	1.98	59.1	26.3	84.0	282.0	94.7
1.98	1.96	59.0	23.6	77.4	267.1	93.7
1.96	1.94	50.4	21.3	71.5	253.2	98.0
1.94	1.92	46.6	19.2	66.0	240.1	96.1
1.92	1.90	46.8	17.3	61.0	227.8	96.0
1.90	1.88	33.3	15.6	56.5	216.2	91.6
1.88	1.86	39.2	14.1	52.3	205.3	89.0
1.86	1.85	29.8	12.7	48.4	195.1	87.5
1.85	1.83	38.2	11.5	44.9	185.5	87.4
1.83	1.82	28.2	10.4	41.7	176.4	88.2
1.82	1.80	22.7	9.5	38.7	167.8	95.1

▶ 46

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Informativity-based resolution limit

Difference in each direction of overall B-factor is  $\sim 10 \text{ \AA}^2$

## The estimators of non-isomorphism level

		11 % - radiation-induced non-isomorphism	2.5 % - anomalous signal	0.1 % - internal non-isomorphism
		↓	↓	↓
1	1.0001	1.0368	0.1100	0.0010
2	1.4456	0.0311		
		↑		
		3.1 % - systematic errors		

▶ 47

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011

## Statistics - Work in progress

Two completeness statistics existing now:

- ▶ Scalepack statistics based on the Bragg law, i.e. every measured reflection is counted. It is useful to know to judge if some part of reciprocal space was not measured.
- ▶ Downstream applications, e.g. Refmac statistics, see only informative reflections – only reflections that make the cut are counted.
- ▶ Both statistics are needed, so they cannot be merged into one. However, historically only the first one was used and discussed, thus crystallographers still feel uncomfortable with reporting completeness below 100% (there is no way to achieve it in the case of anisotropic diffraction without sacrificing the map quality).

▶ 48

APS Data collection workshop and CCP4 school, June 7 - June 15, 2011



## Acknowledgments

---

- ▶ Marcin Cymborowski
- ▶ Zbyszek Otwinowski
- ▶ Wlodek Minor
- ▶ SBC beamline staff
- ▶ All the people who provided test data