Data reduction and analysis

Karen Friese & Andrzej Grzechnik

Departamento Física Materia Condensada, Universidad del País Vasco,
Bilbao, Spain

karen.friese@ehu.es

Specialities of high pressure data

- the resolution in $\sin\theta/\lambda$ is limited

- Access to reciprocal space is limited in certain directions
  (limits in h,k,l resolution)

- intensities are affected by errors
  - diamond dips
  - reflection intensities falsified due to powder rings
    (gasket, backing plates)

- Outliers may be present

  - overlap with diamond reflections
  - shadowed reflections

→ High pressure data are of poor quality

The internal R-value

$$R_{int} = \sum_i \sum_j \frac{(I_j - \bar{I}_i)}{\bar{I}_i}$$

where i runs over all independent reflections and j over all symmetry equivalent reflections corresponding to the i-th independent reflection.

Rule of thumb: the final agreement factor for the refinement should be below the internal R-value

Redundancy

Ratio of the number of measured reflections to the number of crystallographically independent reflections

the more reflections are merged, the smaller the importance of outliers

→ measure as many reflections as possible!

Parameter to data ratio

Rule of thumb: 10 data points for 1 parameter

in high pressure experiments it is very often not possible
to reach this relation

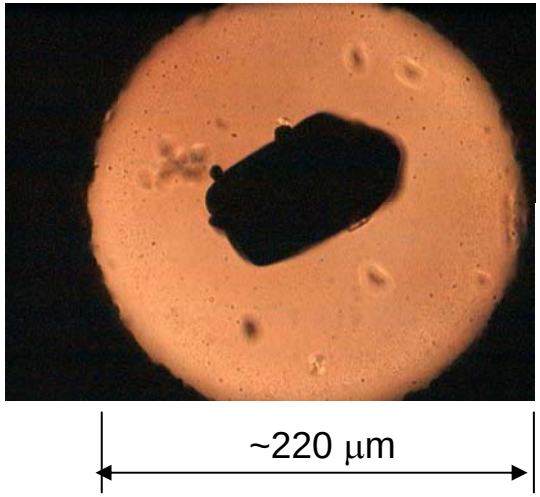
Two solutions:

- increase the number of data points

  Shift to shorter wavelengths: synchrotron
  Choose a cell with a maximum opening angle
  Reduce the number of bad reflections

- limit the number of parameters

Example dataset:
(Laboratory source; Mo Kα, 5.82 GPa, 2 runs, Almax-type DAC, 1:1 pentane to isopentane)



BaV_6O_{11} (Z=2)

Lattice parameter: a ≈ 5.8 Å, c ≈ 13.2 Å

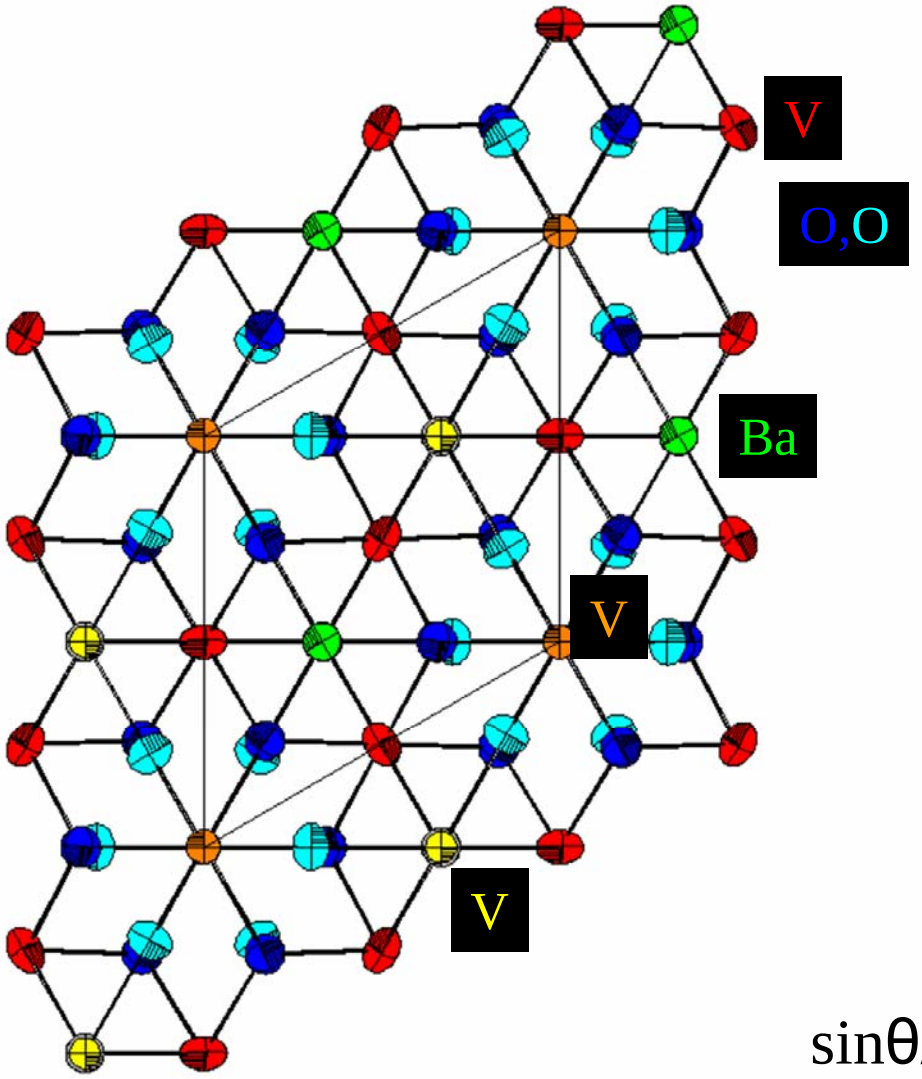
Space group $P6_3/mmc$ at ambient conditions:
1 Barium (56), 3Vanadium (23), 3 Oxygen (8)

Space group $P6_3mc$ at high pressures (> 3 GPa)

1 Barium, 4 Vanadium, 5 Oxygen

Influence of data resolution on atomic displacement parameters

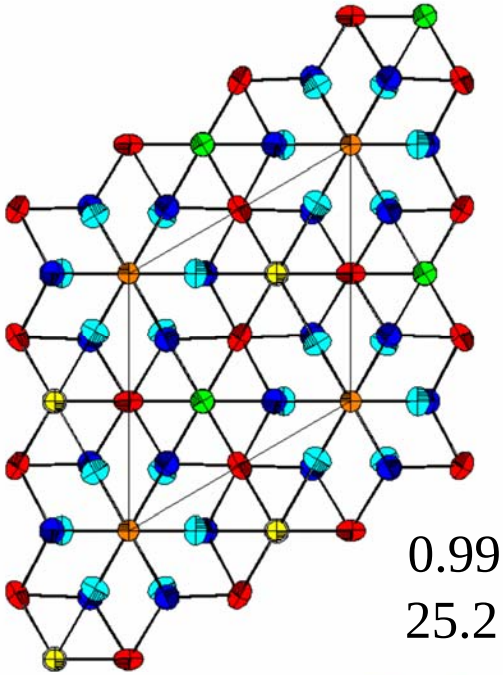
Ambient pressure



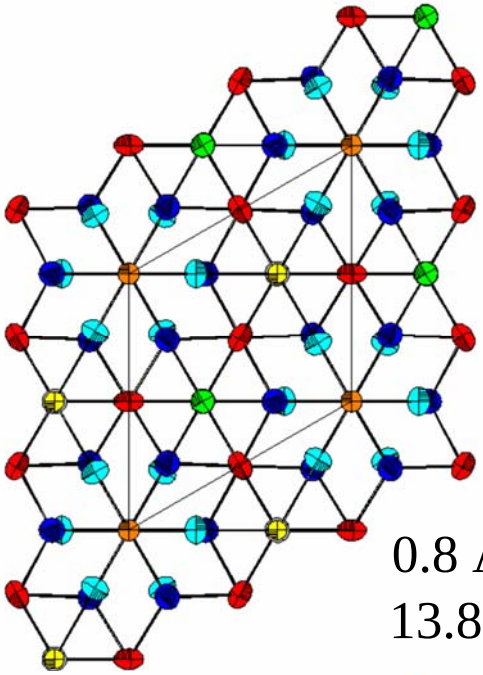
$\sin\theta/\lambda = 0.99$ Å^{-1}

data/parameter = 25.2

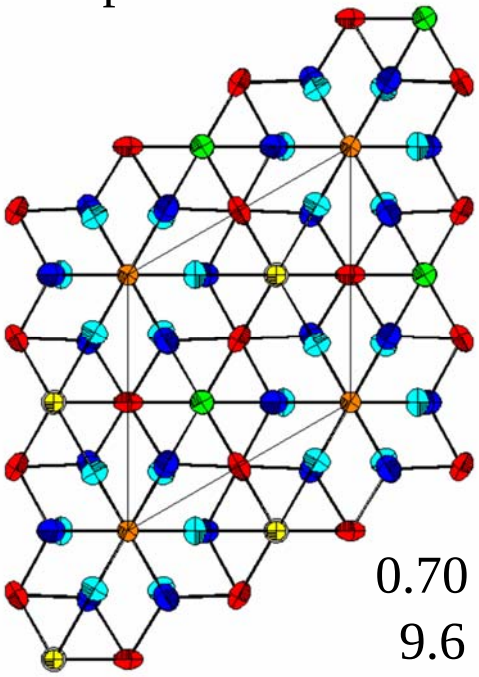
Influence of data resolution on atomic displacement parameters



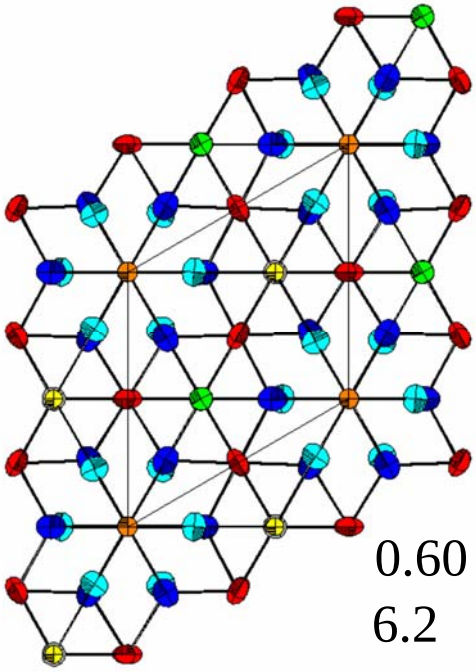
0.99 Å^{-1}
25.2



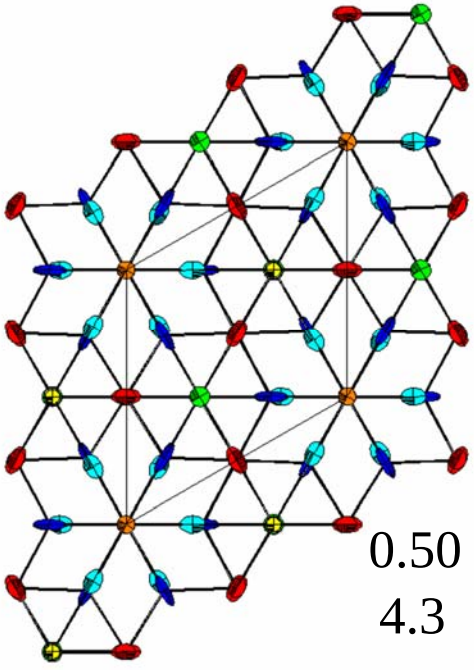
0.8 Å^{-1}
13.8



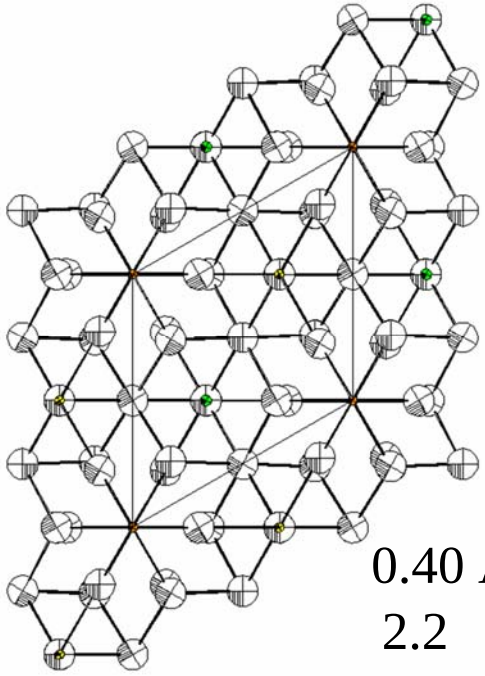
0.70 Å^{-1}
9.6



0.60 Å^{-1}
6.2

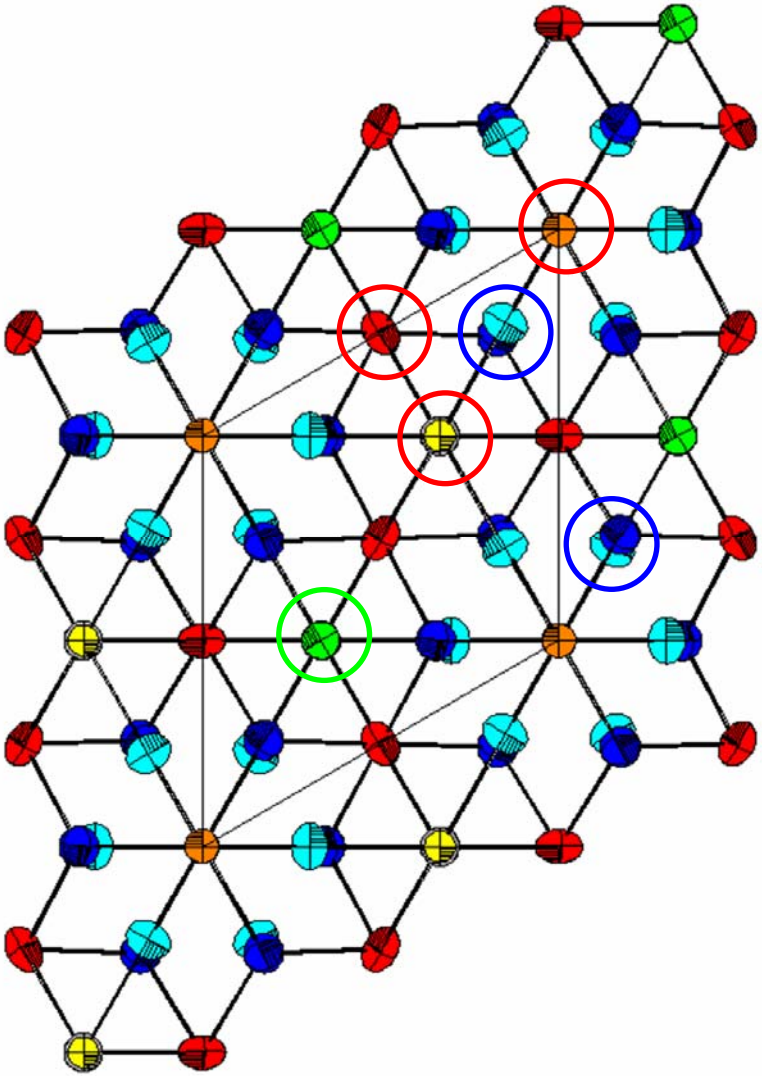


0.50 Å^{-1}
4.3

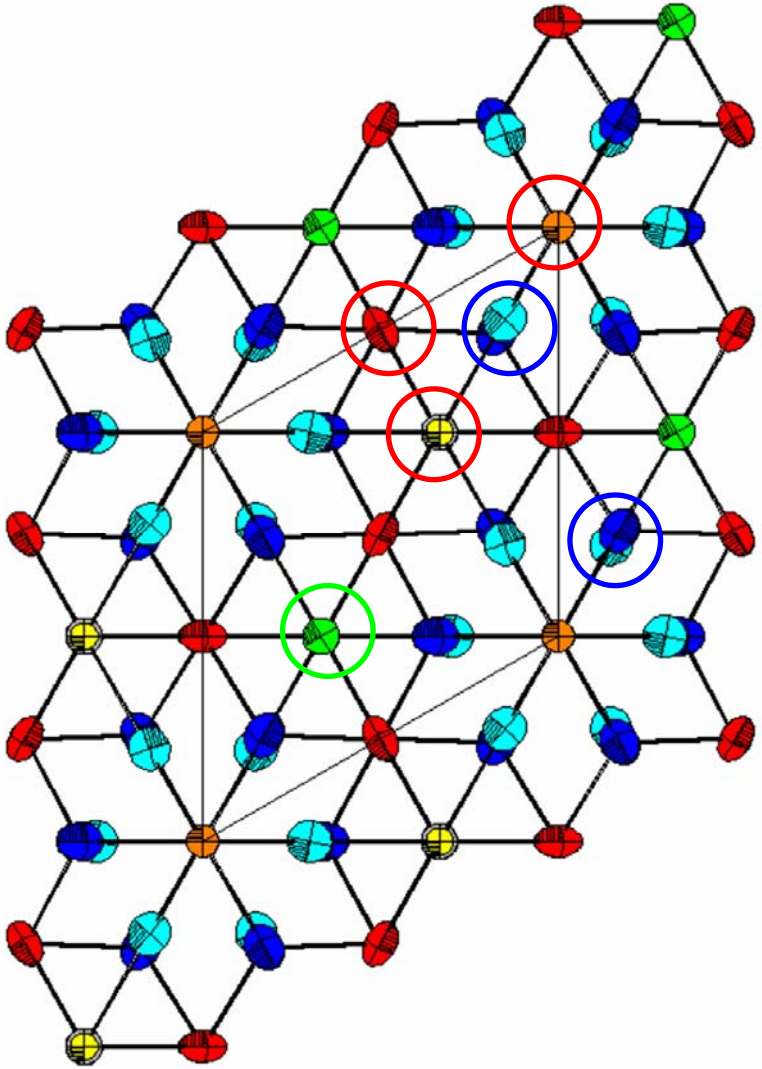


0.40 Å^{-1}
2.2

Influence of data resolution on atomic displacement parameters



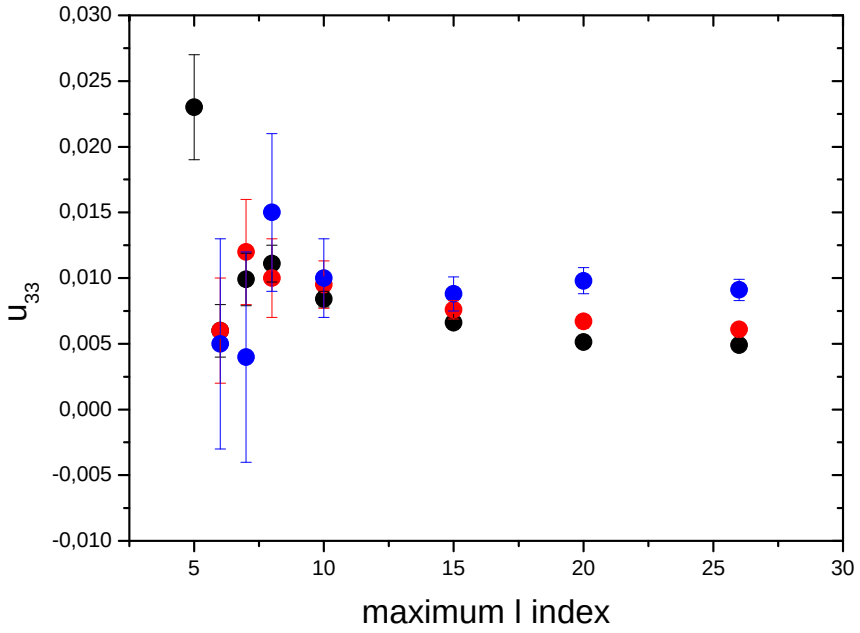
Ambient pressure: 0.99/25.2

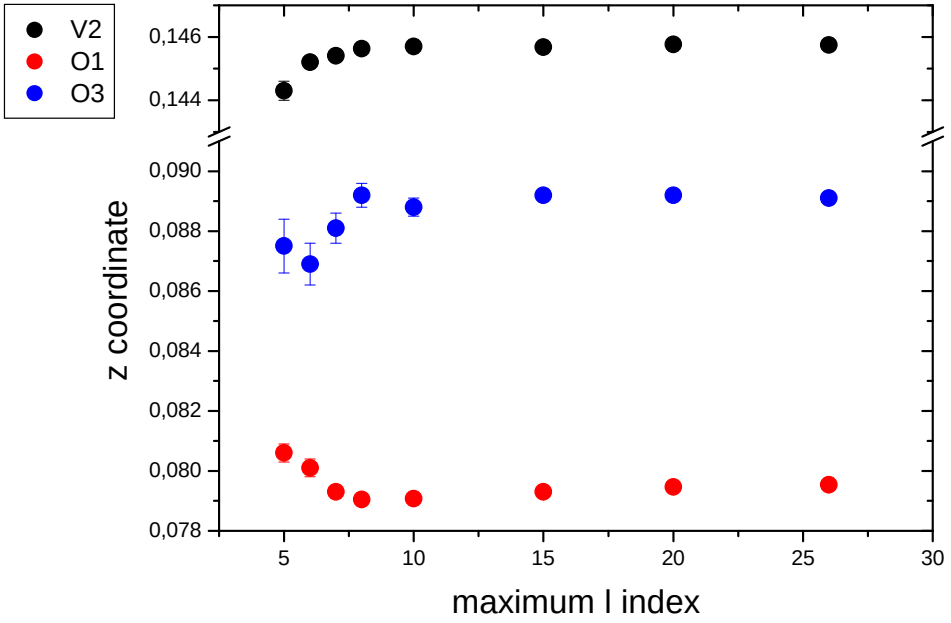


1.18 GPa: 0.67/12.6

Influence of limited resolution in certain directions

Example: limited resolution in c*

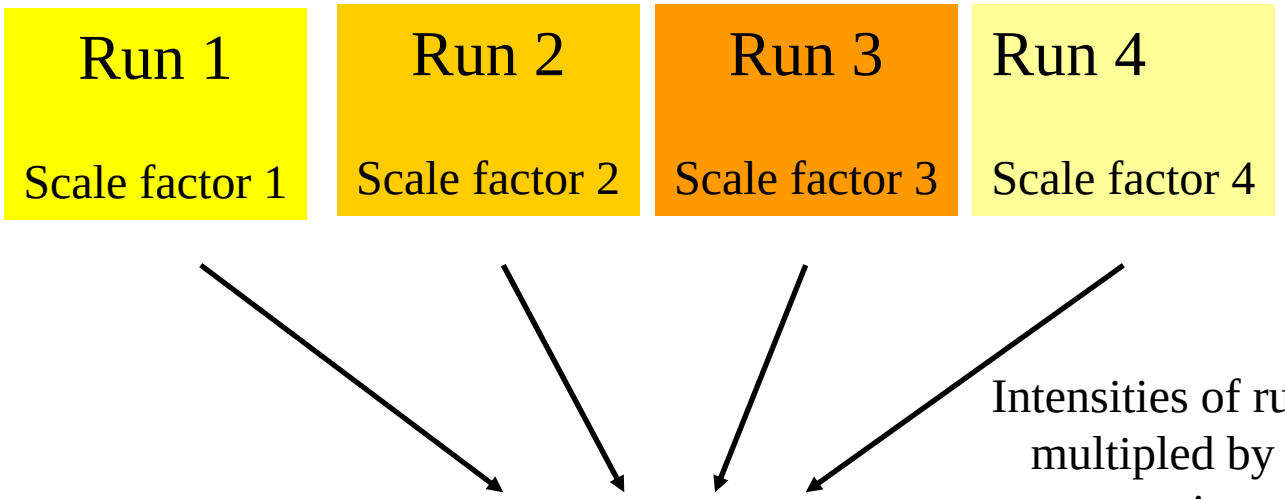




anisotropic displacement parameters

atomic coordinates

Scaling of datasets



Intensities of run 2, 3, and 4 have to be multipled by scale factors to fit the intensities of run 1

Advantage: increase of redundancy

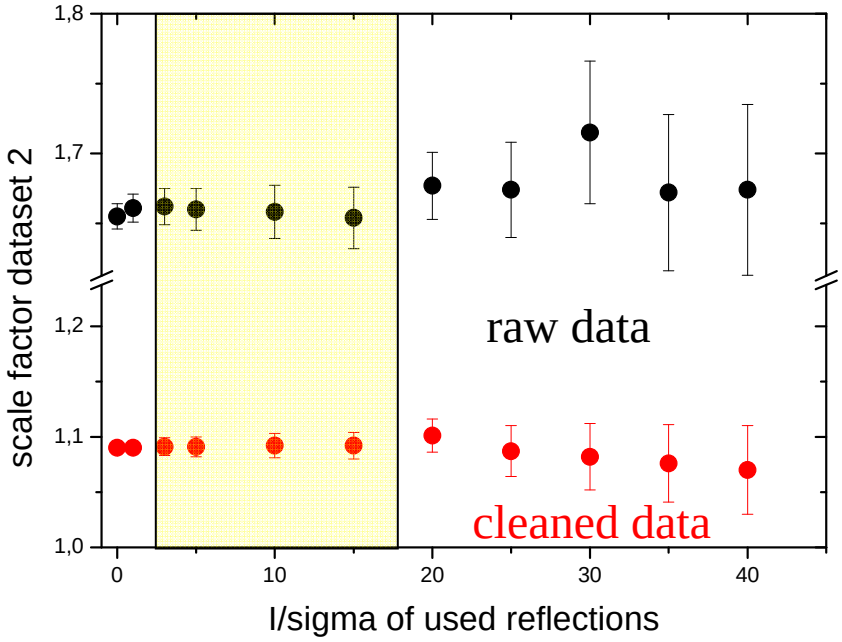
scaling is usually done via common reflections in the partial datasets

only reflections over a certain I/σ ratio limit are chosen

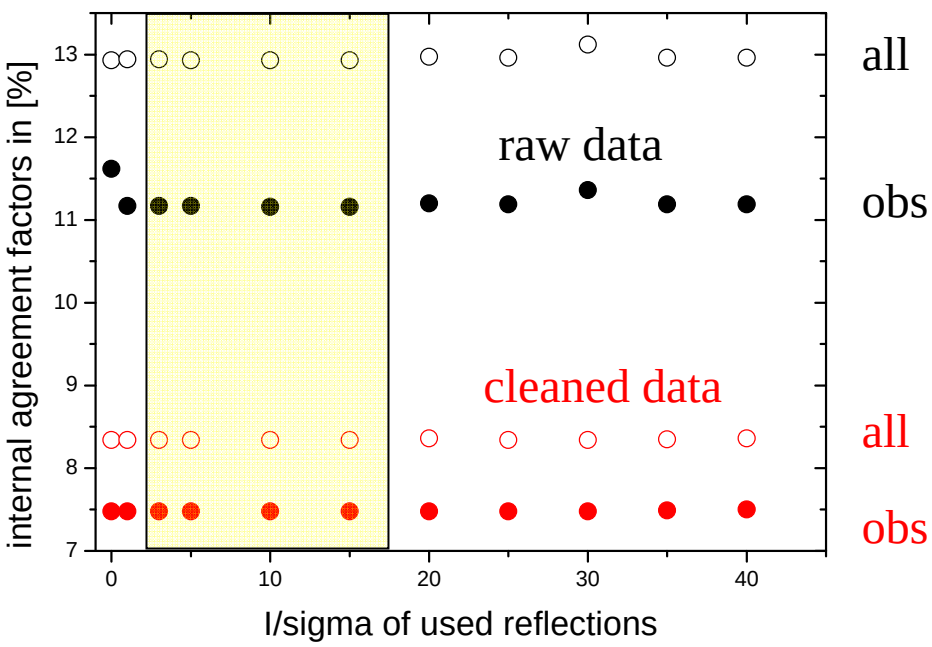
falsified reflection intensities introduce errors in the scale factors
any error in the scale factor influences all reflections in the run !!!

Which I/σ limit should be used for choosing the reflections for scaling?

Scale factor in dependence of I/σ of chosen reflections



Internal agreement factors in dependence of I/σ of chosen reflections



According to our experience I/σ should be between 3 to 15

Improving the dataset

obs/all

11.16/12.93

10.12/11.34

10.04/11.25

- Internal R-value after integration
- after correction for diamond anvils (no shadowing by gasket)
- after correction for absorption of crystal

Identification of outliers

- On the basis of symmetry equivalent reflections

  the more reflections are averaged, the easier to find the outliers

  → the higher the symmetry and redundancy, the better

  in the initial stages one can use “approximate” symmetries to make identification of outliers easier (e.g. Laue symmetry)

- On the basis of the refinement

  F(obs)/F(calc) plots

First stage: reflections with I-I(average)>25σ(I(ave))

			I	σ(I)
0	5	-8	473549.1	9500.2
-5	5	-8	926865.1	16891.2
5	-5	-8	20233.0	8700.6

?

- Check on the original frames
- Increase redundancy
- Check sinθ/λ (diamond or gasket?)

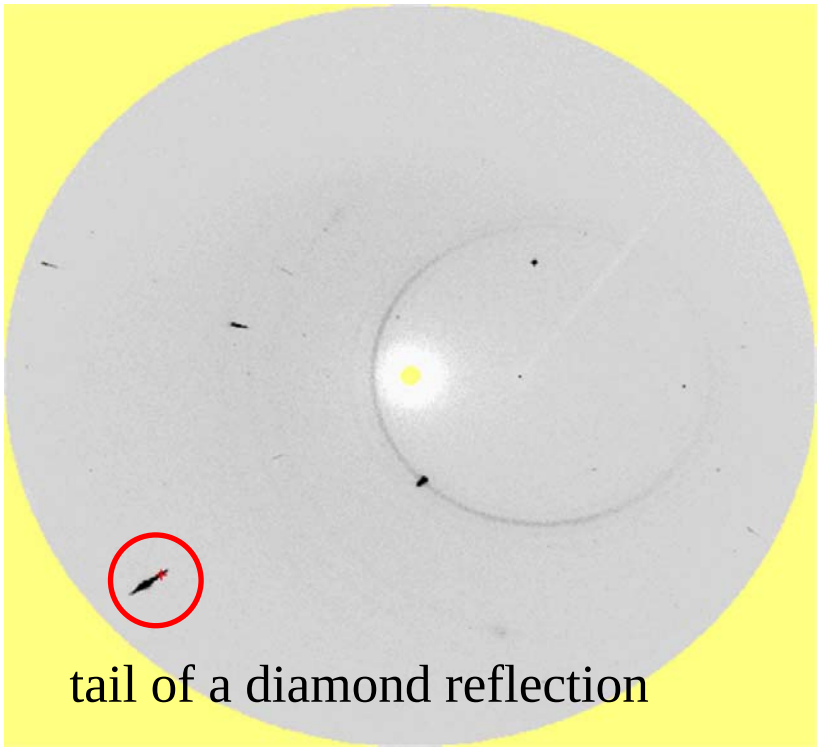
-1	2	-2	235417.7	4534.3
-2	1	-2	307915.0	9683.5
-2	1	-2	290463.4	8889.1
-1	2	-2	1271.9	7205.8
-2	1	-2	342020.5	10210.5

strongest reflection in the dataset

0	0	8	395706.3	3593.2
0	0	8	393350.8	14390.3
0	0	8	415274.4	9004.3
0	0	8	369744.4	10256.1
0	0	8	385701.7	12387.9
0	0	8	438368.4	14201.6
0	0	8	391667.8	12969.5
0	0	8	450817.9	10454.7
0	0	8	411253.3	8985.0
0	0	8	301623.6	8961.2
0	0	8	399260.8	10202.7

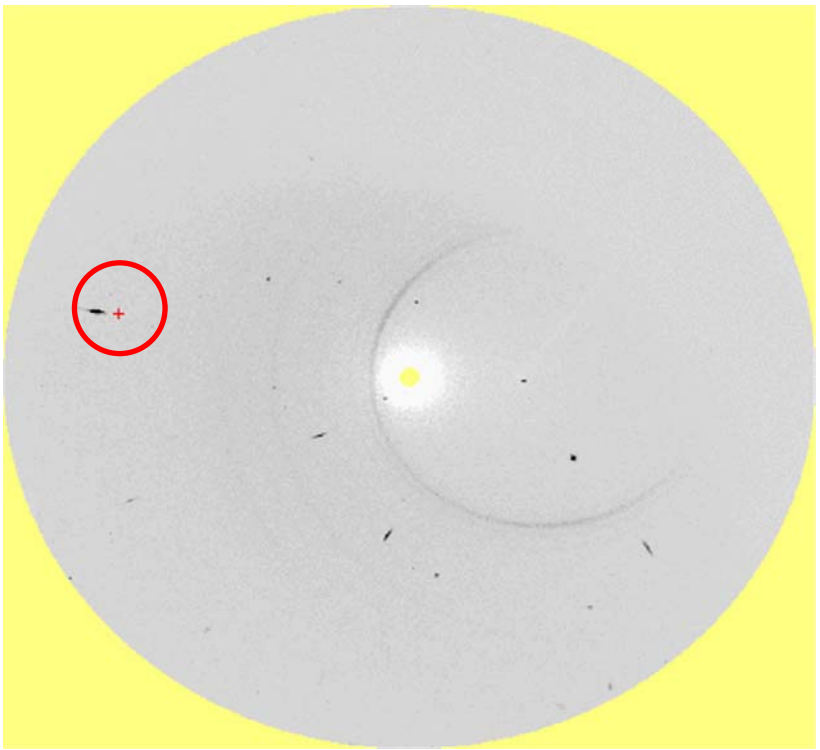
shadowed reflections

Checking reflections on the original frames



$\bar{5}5\bar{8}$

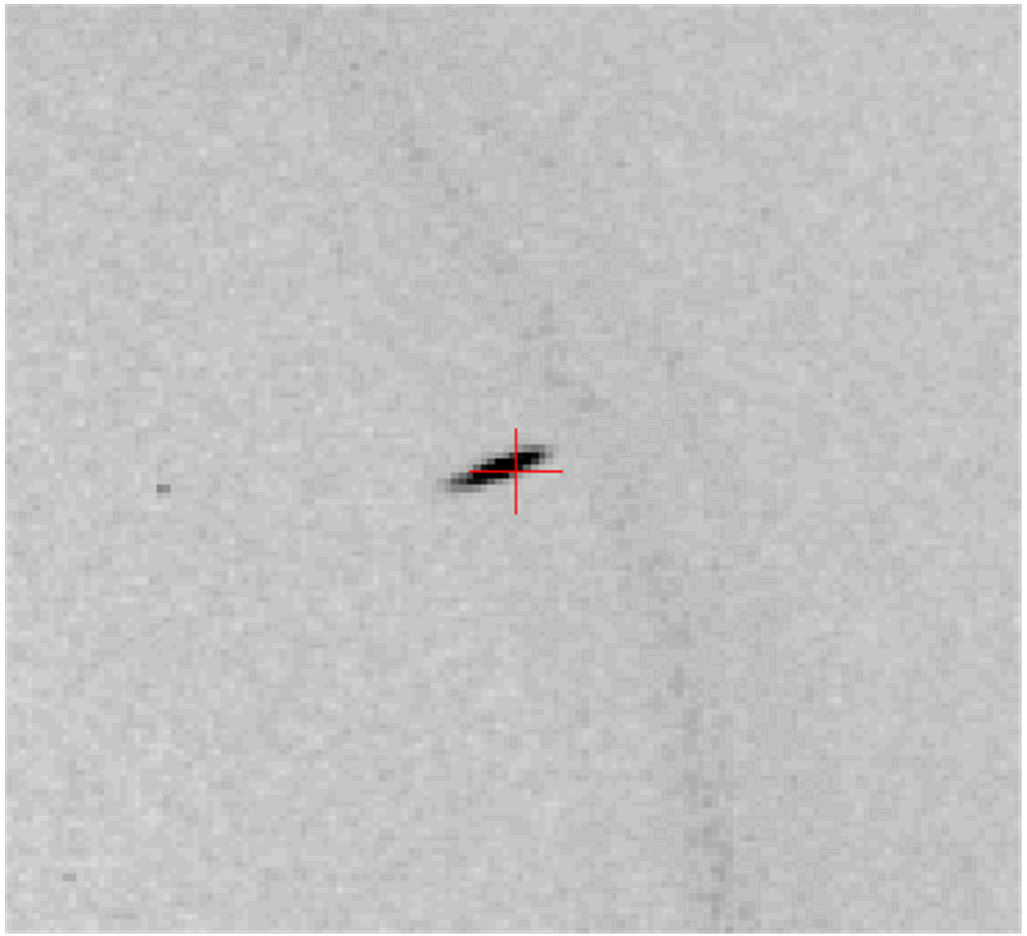
I=926865.1 σ(I)=16891.2



$5\bar{5}\bar{8}$

I=20233.0 σ(I)= 8700.6

Tail of a diamond reflection



Increasing redundancy: adding a center of symmetry

$P6_3mc$				
0	5	-8	473549.1	9500.2
-5	5	-8	926865.1	16891.2
5	-5	-8	20233.0	8700.6

$P6_3/mmc$				
0	5	8	237917.1	5993.2
5	-5	8	1046.5	11860.2
5	-5	-8	20290.2	8725.2
-5	5	-8	929489.1	16939.0
-5	5	8	842.5	8424.3

Improving the dataset

obs/all

11.16/12.93

- Internal R-value after integration

10.12/11.34

- after correction for diamond anvils (no shadowing by gasket)

10.04/11.25

- after correction for absorption of crystal

8.28/9.57

- initial exclusion of falsified reflections (3 shadowed + 1 diamond)

Next stages: reflections with I-I(average)>xxσ(I(ave))

- **Change the criteria and repeat**

h	k	l	I	σ(I)
h	k	l	…….	…….
…….			…….	…….
……..			…….	…….

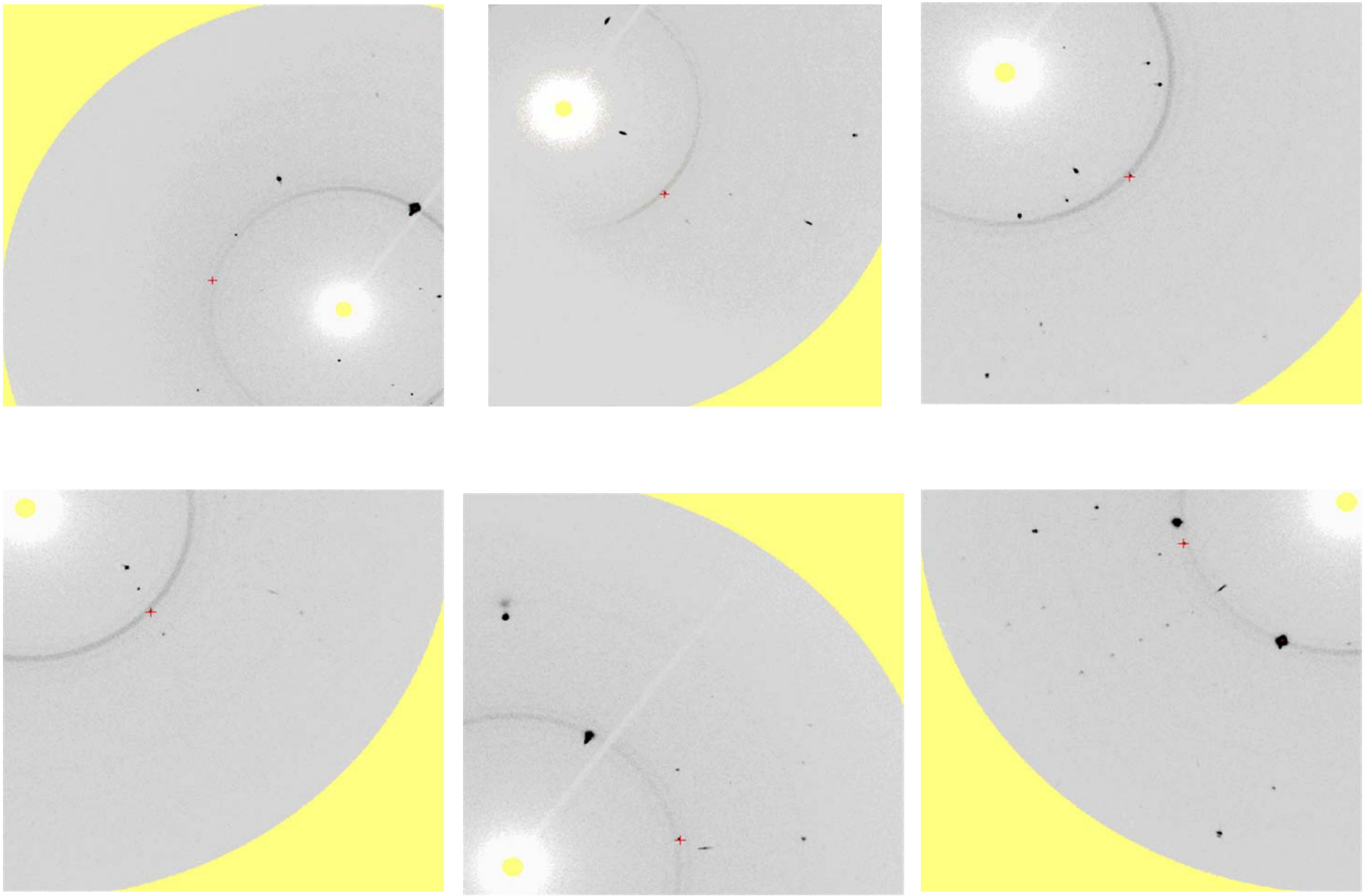
h	k	l	I	σ(I)
h	k	l	…….	…….
…….			…….	…….
…….			…….	…….
…….			…….	…….

Reflections falling on a gasket ring

First gasket ring $\sin\theta/\lambda = 0.245$ Å^{-1}

$\sin\theta/\lambda$ 0.249Å^{-1}									
	0	1	-6	37999.4	2508.3				
	1	0	-6	43012.5	10080.6	0	-1	-6	33644.9	8301.3
	1	0	-6	42179.6	8953.9	1	0	-6	36127.9	9828.2
	1	0	-6	736.9	7368.7	0	-1	-6	39565.2	9850.8
	-1	1	-6	60834.1	10070.0	0	-1	-6	40667.8	9114.2
	0	-1	-6	41443.9	9130.0	0	1	-6	40718.1	9734.4
	-1	0	-6	48138.3	8794.1	0	-1	-6	41116.5	11286.6
	1	0	-6	35725.2	8622.5	0	1	-6	28081.3	9637.7

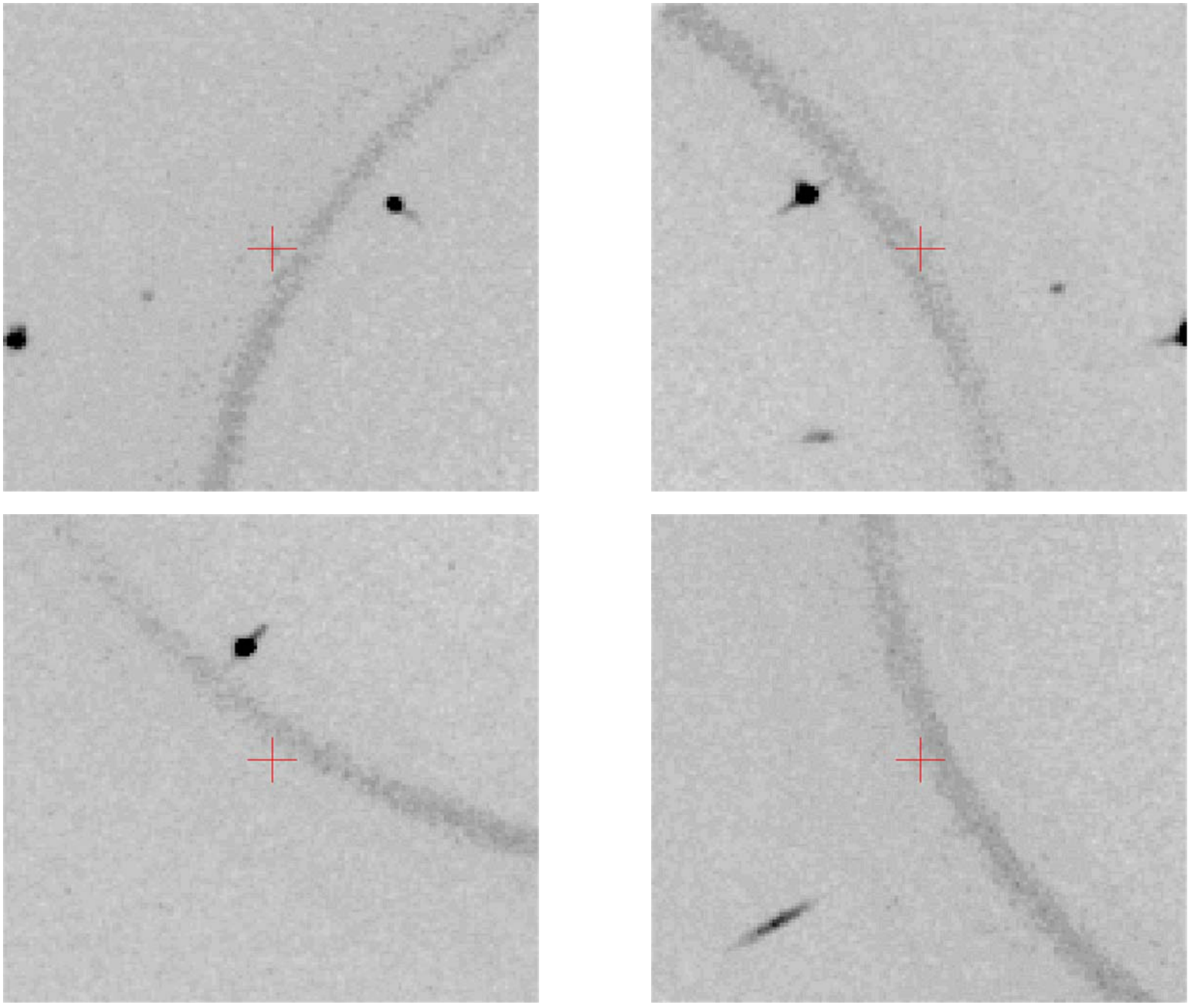
Reflections on the measured frames



Weak or unobserved reflections on gasket

0	2	4	6221.3	2515.8					
2	-2	4	742.4	7424.6	-2	2	4	23129.5	8921.0
-2	2	4	745.6	7456.2	2	-2	4	1066.8	10667.7
-2	2	4	10668.6	15560.3	2	-2	4	3319.6	7047.4
2	-2	4	703.0	7029.9	2	-2	4	11725.8	9544.0
2	-2	4	932.5	9325.4	-2	2	4	16548.5	7099.7
2	-2	4	944.1	9440.0	-2	2	4	6599.6	6936.5
-2	2	4	3751.0	7659.5					

Reflections on the measured frames



Identification of outliers

- On the basis of symmetry equivalent reflections

  the more reflections are averaged, the easier to find the outliers

  → the higher the symmetry and redundancy, the better

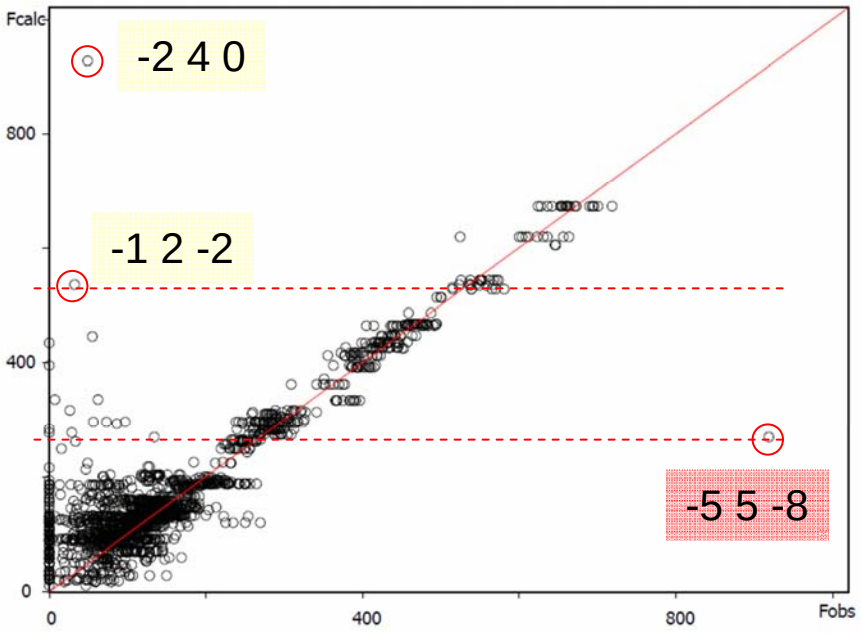
  in the initial stages one can use “approximate” symmetries to make identification of outliers easier (e.g. Laue symmetry)

- On the basis of the refinement

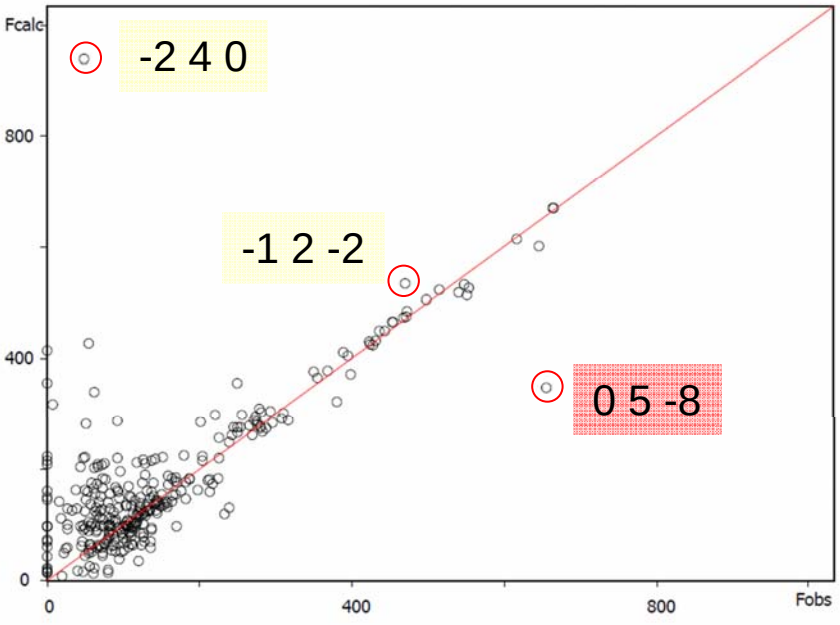
  F(obs)/F(calc) plots

Outliers from refinement: F(obs)/F(calc) plots

Shadowed reflections:
F(obs)<<F(calc)



Diamond reflections:
F(obs)>>F(calc)



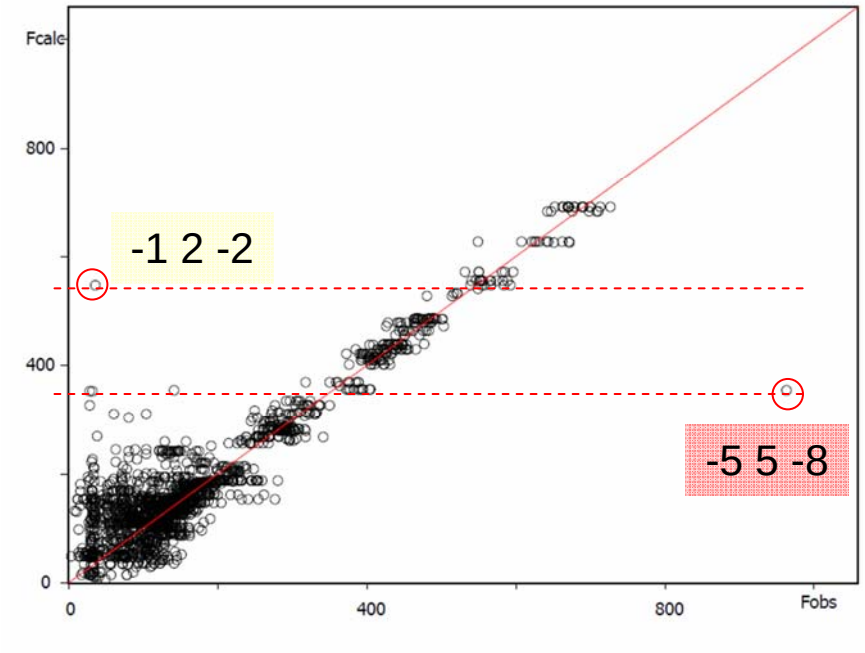
raw data

non-averaged
6.61/21.62

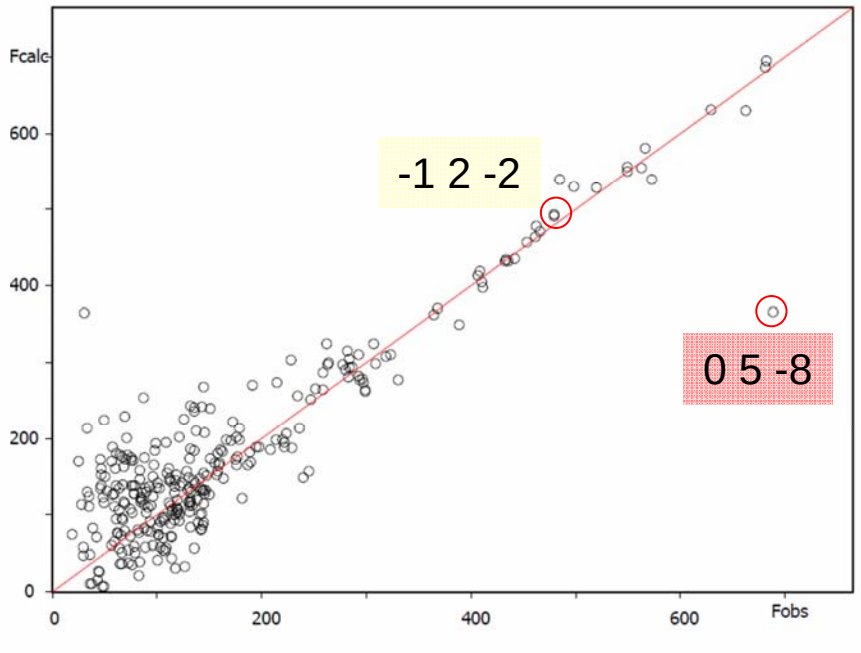
averaged
10.56/33.26

Outliers from refinement: F(obs)/F(calc) plots

Shadowed reflections:
F(obs)<<F(calc)



Diamond reflections:
F(obs)>>F(calc)

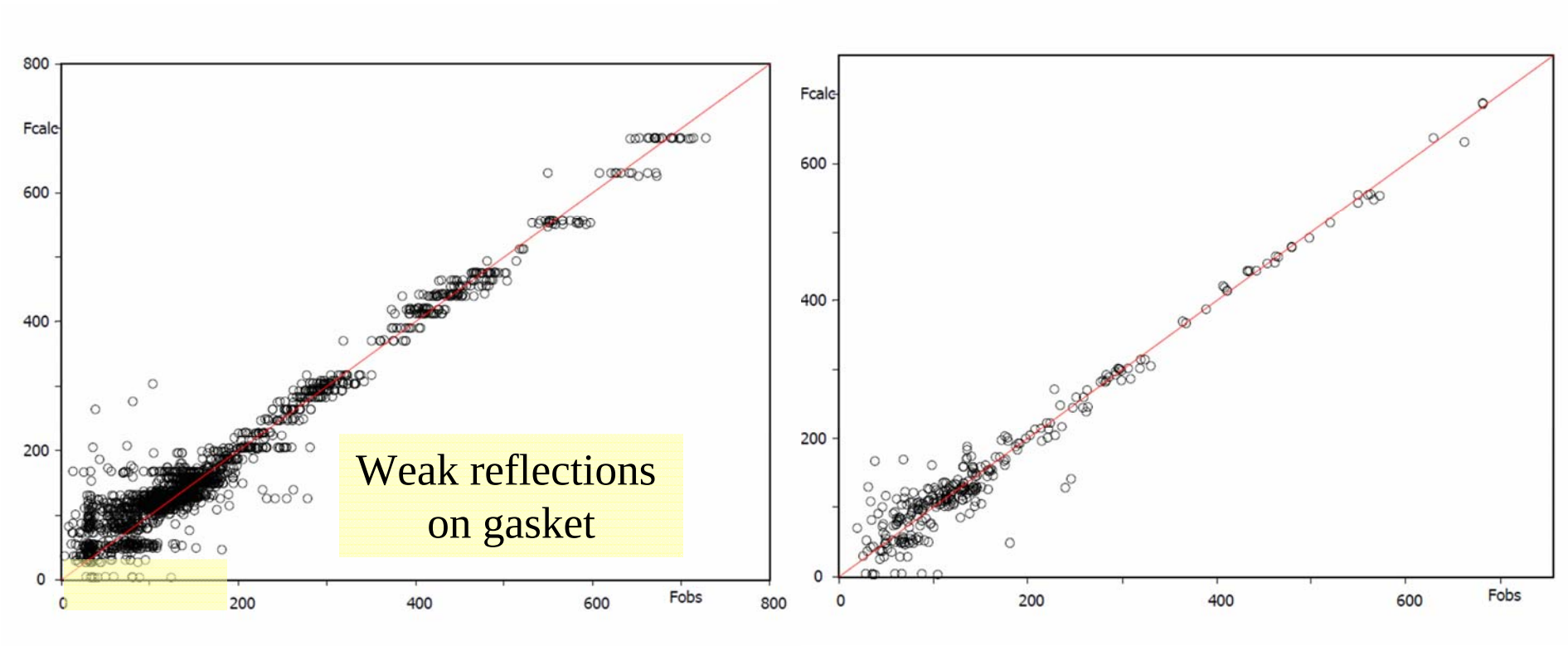


Absorption corrected data

non-averaged
6.25/19.23

averaged
10.73/25.10

Outliers from refinement: F(obs)/F(calc) plots



Absorption corrected data without biggest outliers

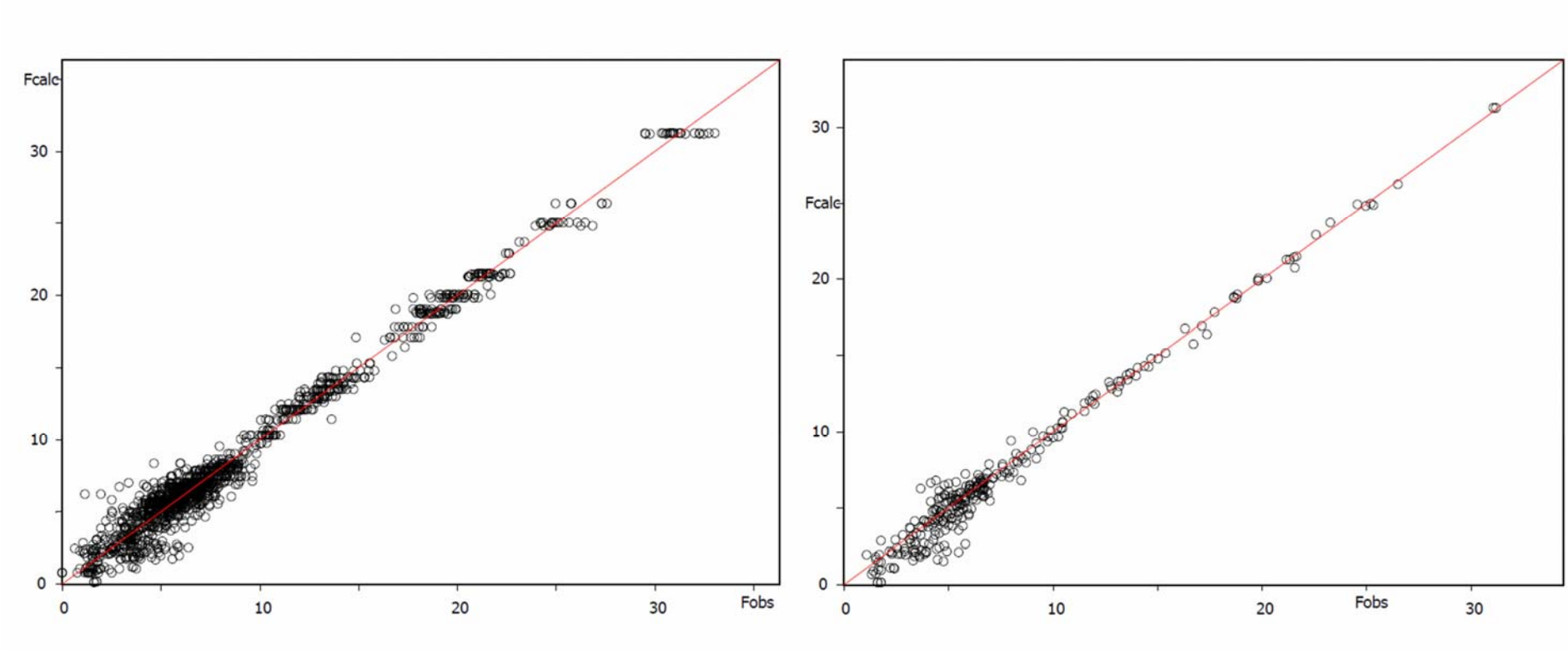
non-averaged

4.48/11.65

averaged

4.66/11.22

Outliers from refinement: F(obs)/F(calc) plots



Final dataset without outliers

non-averaged

3.63/7.93

averaged

2.51/8.23

Improving the dataset

obs/all

11.16/12.93

- Internal R-value after integration

↓

10.12/11.34

- after correction for diamond anvils (no shadowing by gasket)

↓

10.04/11.25

- after correction for absorption of crystal

↓

8.28/9.57

- initial exclusion of falsified reflections (3 shadowed + 1 diamond)

↓

7.48/8.34

- Further rejection of outliers

Limiting the number of refinable parameters

Displacement Parameters:

- use isotropic displacement parameters instead of anisotropic ones
- use higher pseudosymmetry (if present) to restrict the number of parameters
- TLS refinement
- fix the displacement parameters to reasonable values

Geometrical constraints:

- restrict bond lengths
- restrict molecular/polyhedral geometry

Approximate the structure (serious cases)

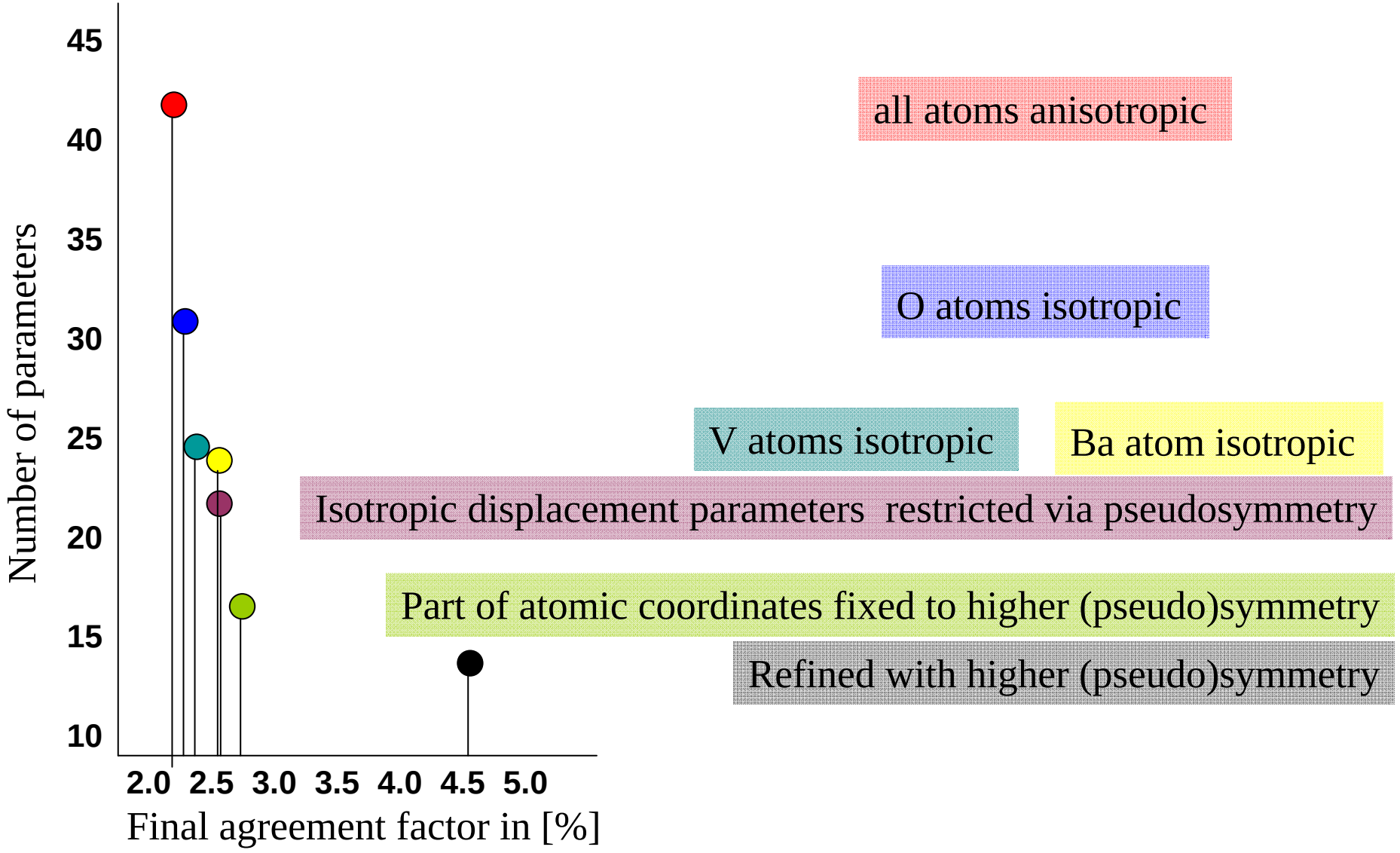
- refine an average structure with higher symmetry (if present)
- fix part of the atomic positions

In the case of a phase transition:

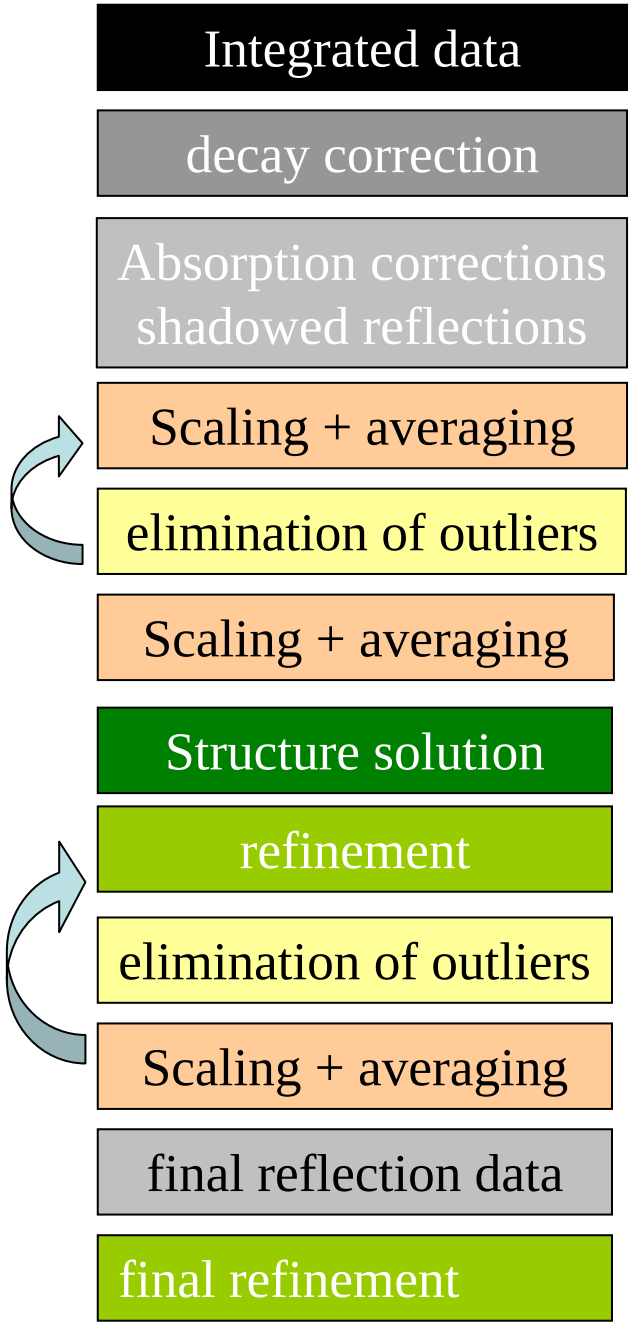
- symmetry mode analysis

(Perez-Mato et. al., Acta Cryst. A66, 2010, 558-590; Grzechnik et.al., J.Phys. Cond. Matter 20,(2008), 285208.

Limiting the number of parameters



The whole procedure



Absorption correction (crystal),
Shadowed reflections

Only if synchrotron

Absorb-program

use higher (pseudo)symmetry
if necessary and possible

use real symmetry

if there is no structural model

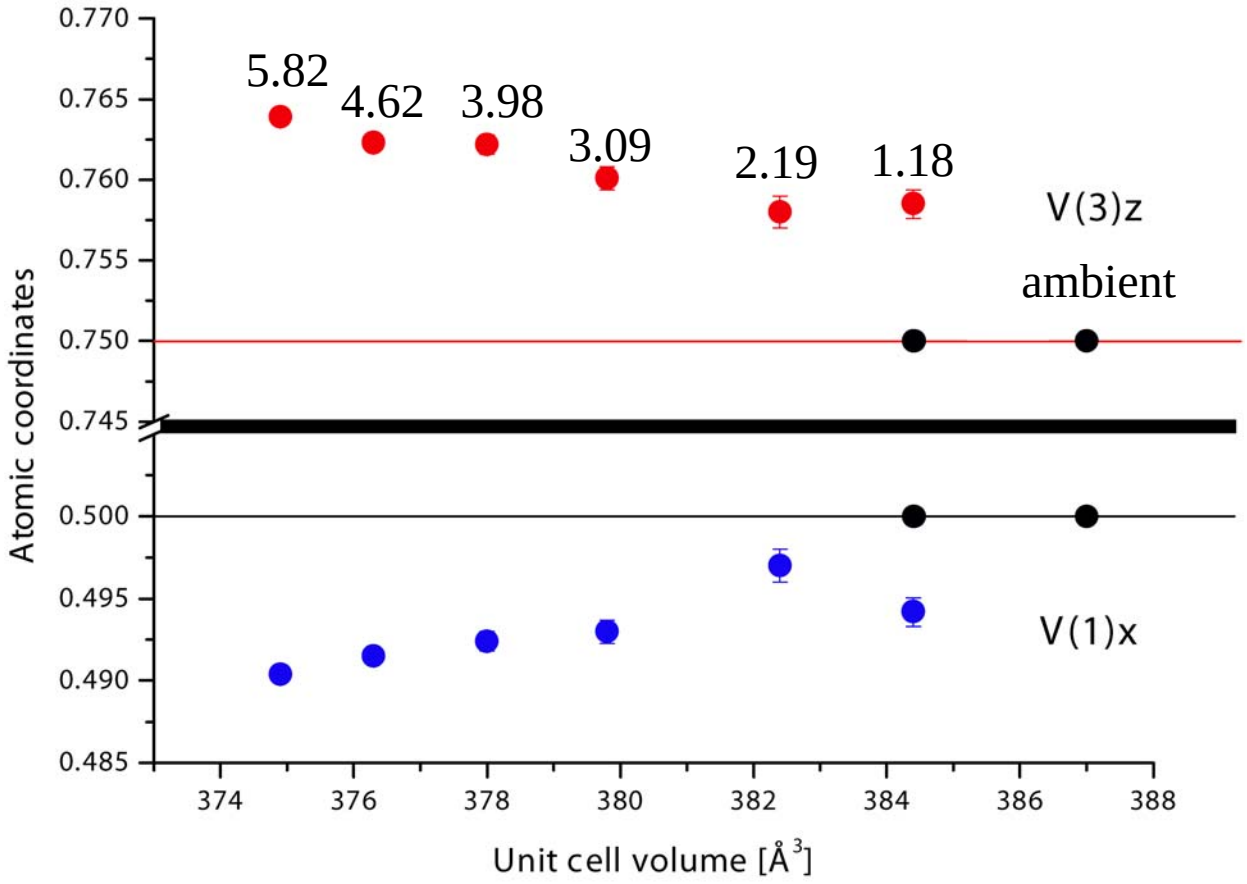
limit the number of parameters

F(obs)/F(calc) plots

Example: BaV_6O_{11}
comparison of the results with good and bad datasets

	wR(all) [%] uncleaned	wR(all) [%] cleaned
• Solution (Sir97)	17.16	8.34
	2 oxygen missing	1 oxygen missing
• Scale factor	18.27	7.05
• Coordinates	18.18	4.05
• Inclusion of missing atom(s)	13.74	2.84
• Coordinates of missing atom	14.45	2.64
	refinement unstable	
• U_{iso} Ba, V (1V negative)	13.26	2.52
	refinement unstable	
• U_{iso} O (1V and 1 O negative)	10.57	2.45
	1V and 3 O negative	1V and 1 O negative
• Ba aniso (1V negative)	10.57	2.38
• Trial: V aniso (2 V negative)	10.23	2.29
• U_{iso} of part of V/O set equal	11.17	2.50
	4 O negative	

Looking at data trends



$P6_3mc$

$P6_3/mmc$

$P6_3/mmc$

$P6_3mc$

Comparing different structural models : Hamilton's test

W.C. Hamilton, Acta Crystallogr. 18, 502-510 (1965)
Significance tests on the Crystallographic R-Factor

Does the increase of parameters to a model lead to a significant improvement of the model?

Comparison of an R-factor ratio to tabulated values

R-factor ratio: wR(model B)/wR(model A)
Model B: model with restriction
Model A: model without restrictions

If the R-factor ratio is larger than the tabulated value
→ the hypothesis can be rejected

Examples for the use of the Hamilton test

- independent structure refinements
- different structural models
  e.g. anisotropic/isotropic/partially anisotropic
- structural models with refined and fixed (=estimated) coordinates
- comparison of two absolute configurations
- two refinements: one with fixed molecular geometry, the other with free geometry
- refinements with different space group symmetries

Some points which have to be observed

- Test is based on wR(F)
- the number of reflections in the two model has to be equal
- If you use geometrical constraints, think carefully about the number of parameters
- the tabulated values correspond to a certain probability level

e.g. $R_{b,n-m,0.50}$ indicates that the hypothesis cannot be rejected (can be rejected) at the 50% level i.e. we are wrong half the time if we reject (or accept) a hypothesis at this level.

Example: What is the correct space group at a pressure of 5.82 GPa?

From the refinment:

Number of reflections n=292

Model A ($P6_3/mmc$): 14 Parameters = m_A R_A = wR(all) = 0,0311

Model B ($P6_3mc$): 22 Parameters = m_B R_B = wR(all) = 0,0255

- Hypothesis: Model A is better than model B

  Dimension of the hypothesis $m_B - m_A = 8$

  Number of degrees of freedom $n - m_B = 292 - 22 = 270$

Interpolated value at a 0.005 significance level:

$$R_{8,270,0.005} \cong 1 + 120/270(R_{8,120,0.005} - 1) = 1 + 120/270(1.093 - 1) = 1.0413$$

$$R = R_A/R_B = 1.219 > 1.0413$$

- Hypothesis can be rejected at a 0.005 probability level
  → model B is better → the structure is acentric

Conclusions

- Invest time and effort in the experiment
- Collect data at different pressure points
- Make reconstructions of reciprocal space
- Check carefully for outliers
- Refine carefully and stepwise: make sure adding parameters improves the model
- Limit the number of parameters
- Be critical about the data