Validation of the refinement

The CIF (file) and refinement details

The refinement - step by step

1. Reading of the *.hkl file

```
example.lst:

  h   k   l        Fo^2        Sig

  0   0   1       104.55        5.
  2   0   1        74.15        6.
etc.
  20975  Reflections read, of wh

-10 =< h =< 10,    -13 =< k =< 
    23  Systematic absence vio

Inconsistent equivalents etc.
  h   k   l       Fo^2    Sigma(Fo^2)   N   Esd of mean(Fo^2)
  2   0   0    127128.20     518.46     4     3194.33
  0   2   0     65677.31     243.31     6     1223.07
etc.
    14  Inconsistent equivalents
  2847  Unique reflections, of which        0  suppressed

R(int) = 0.0275     R(sigma) = 0.0111      Friedel opposites merged
```
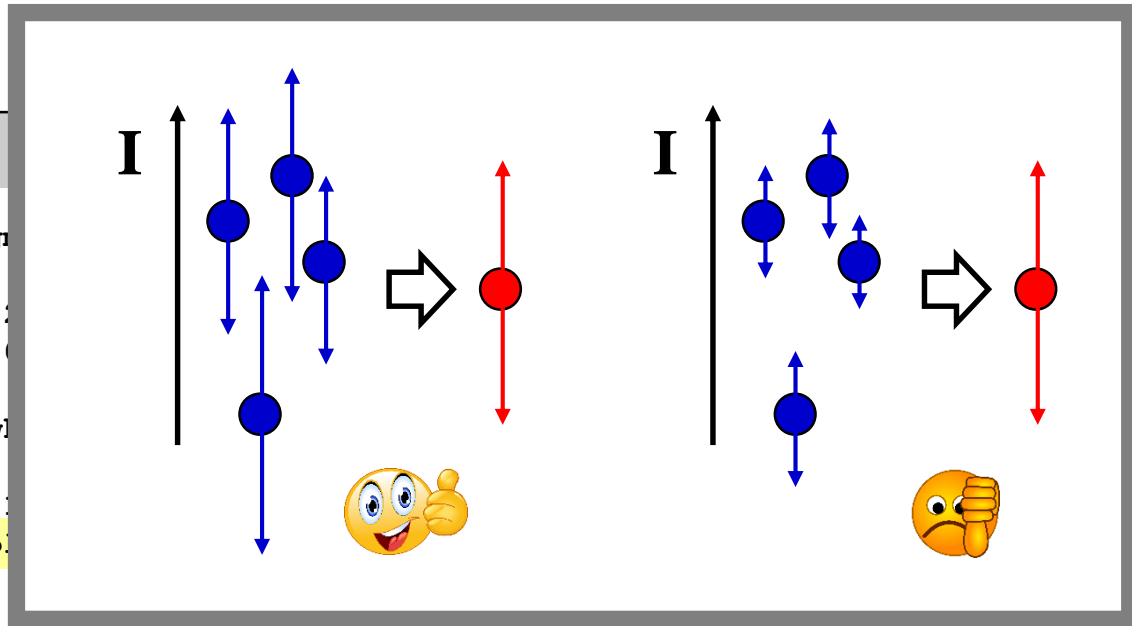

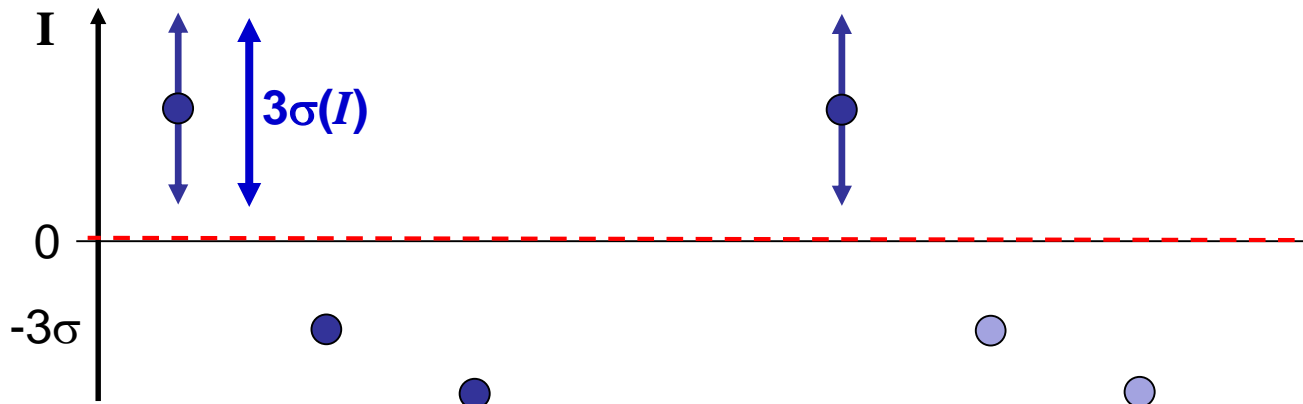Inconsistent equivalents:
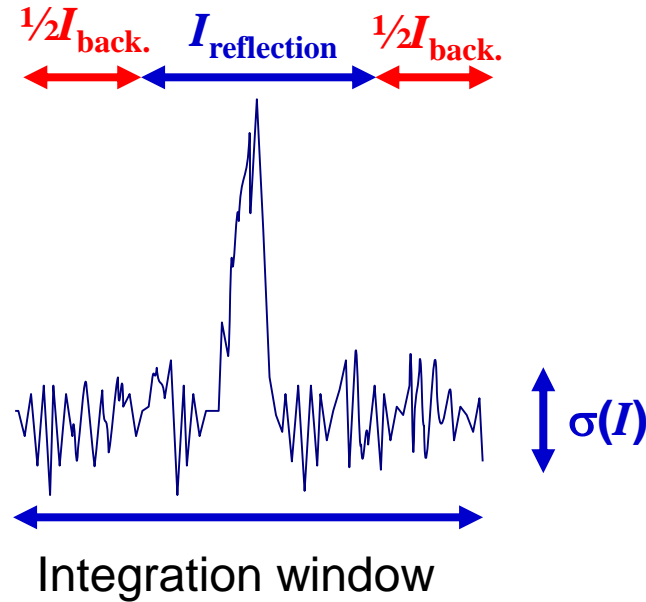σ of averaging >> σ of measure

Rejected reflections: • **Systematic absences violations**
• **Inconsistent equivalents**
• **Reflections with I < −3·σ(I)**

- During the integration of the reflections, their intensity is calculated by

$$I = I_{\text{reflection}} - I_{\text{background}}$$

½$I_{\text{back.}}$ $I_{\text{reflection}}$ ½$I_{\text{back.}}$

$\sigma(I)$

Integration window

I

$3\sigma(I)$

0

-3σ

Why use reflections with negative intensity?

- During the integration of the reflections, their intensity is calculated by $I = I_{reflection} - I_{background}$

- For weak reflections, it can happen that the result is negative, due to a large peak in the noise of the background.
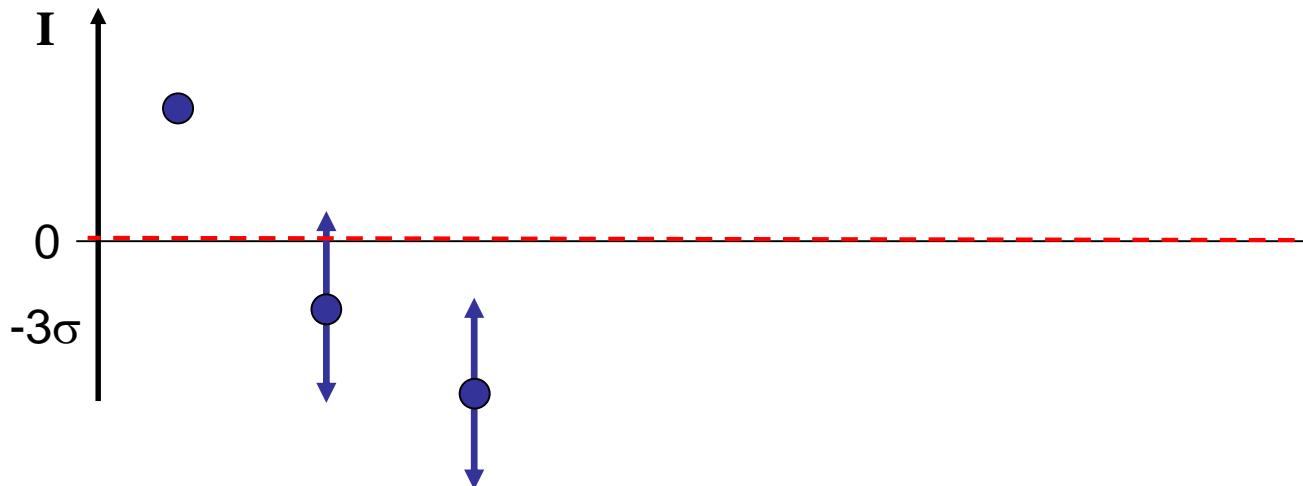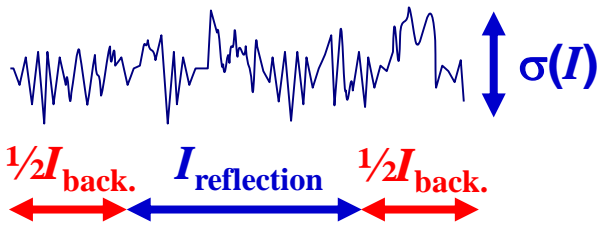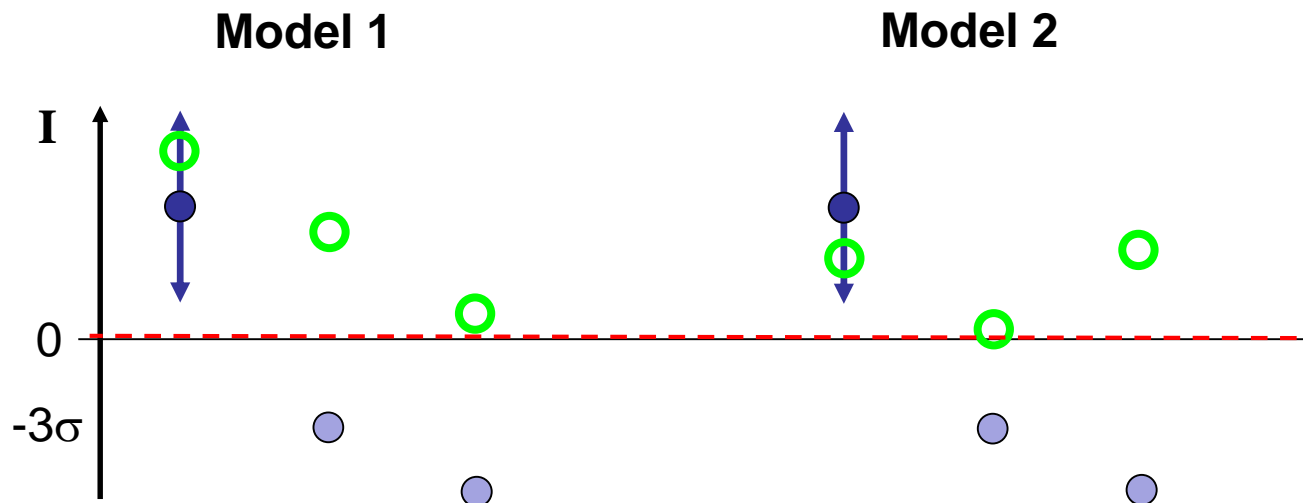
Why use reflections with negative intensity?

- During the integration of the reflections, their intensity is calculated by $I = I_{\text{reflection}} - I_{\text{background}}$

- For weak reflections, it can happen that the result is negative, due to a large peak in the noise of the background.

If we consider only reflections with $I > 0$, the two models 1 and 2 have the same probability / the same agreement with the experimental data.

Model 1   **Model 2**
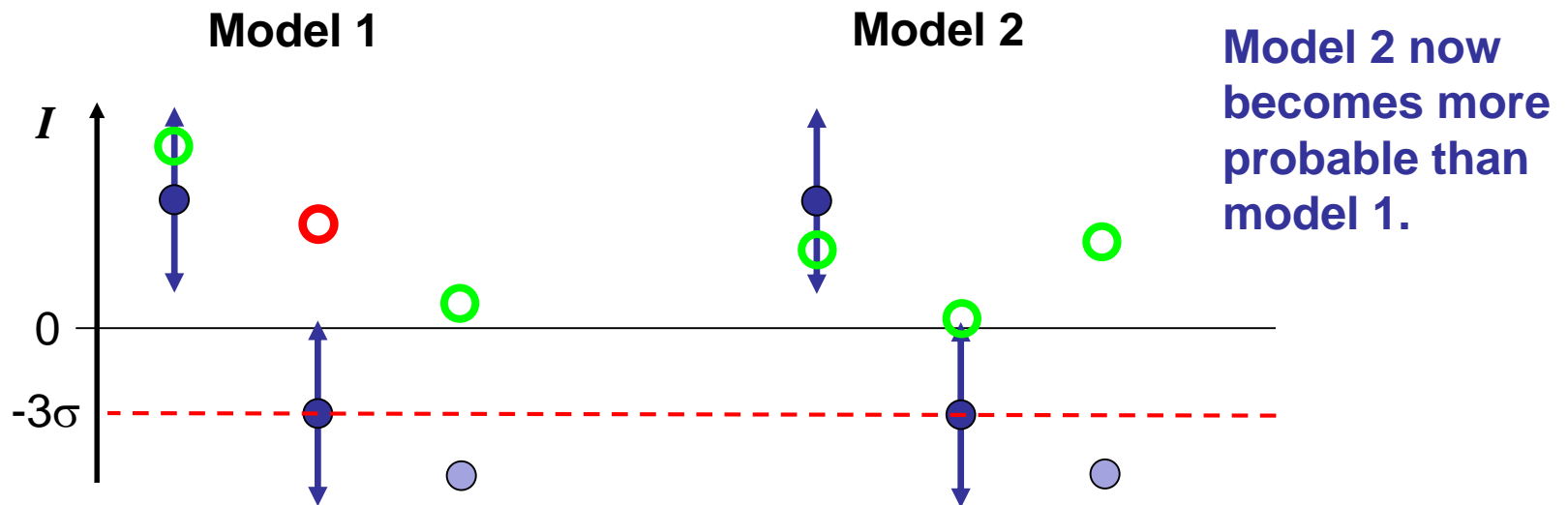
Why use reflections with negative intensity?

- During the integration of the reflections, their intensity is calculated by $I = I_{\text{reflection}} - I_{\text{background}}$

- For weak reflections, it can happen that the result is negative, due to a large peak in the noise of the background.

- Reflections with $0 < I < -3\sigma(I)$ contain important information, since $I = 0$ is in the range of the experimental error.

Model 1

Model 2

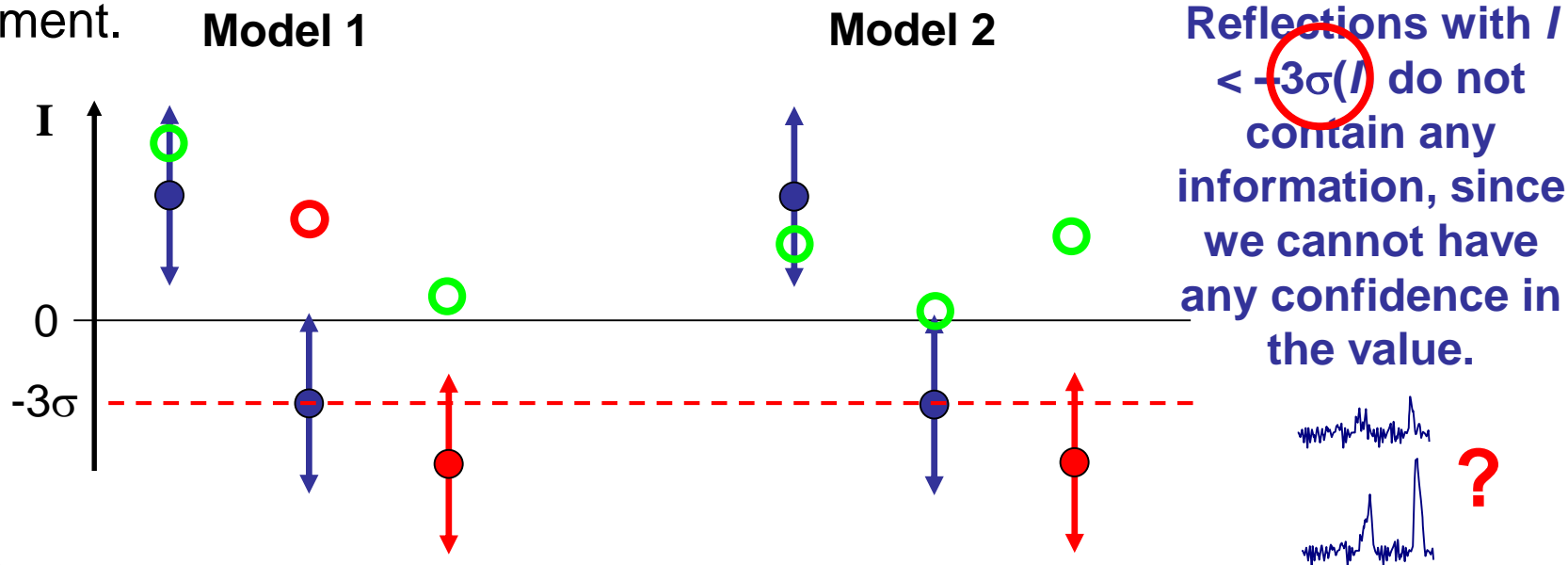**Model 2 now becomes more probable than model 1.**

I

0

-3σ

Why use reflections with negative intensity?

- During the integration of the reflections, their intensity is calculated by $I = I_{reflection} - I_{background}$

- For weak reflections, it can happen that the result is negative, due to a large peak in the noise of the background.

- Reflections with $0 < I < -3\sigma(I)$ contain important information, since $I = 0$ is in the range of the experimental error.

- Reflections with $I < -3\sigma(I)$ are impossible, since all possible values in the error range are negative. Thus, there was a problem with the determination of this reflection and **it needs** to be excluded from the refinement.

Model 1

Model 2

Reflections with $I < -3\sigma(I)$ do not contain any information, since we cannot have any confidence in the value.
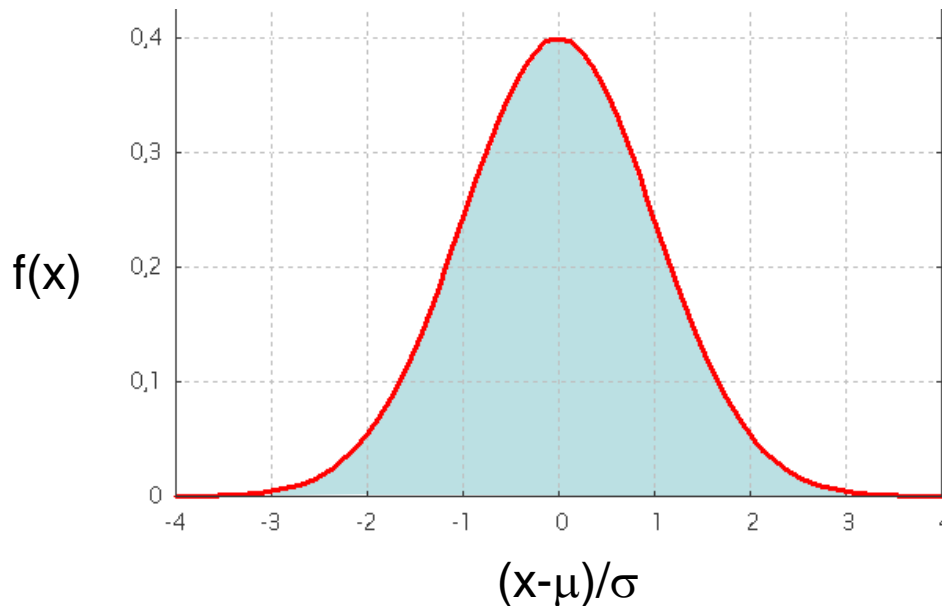
3σ: Why always ±3σ ?

An arbitrary distribution of errors has a Gaussian distribution.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2}$$

μ: correct value
σ: standard deviation

f(x)

(x-μ)/σ

$$\int_{\mu-\sigma}^{\mu+\sigma} f(x) = 0.68 = 68\%$$

$$\int_{\mu-2\sigma}^{\mu+2\sigma} f(x) = 0.96 = 96\%$$

$$\int_{\mu-3\sigma}^{\mu+3\sigma} f(x) = 0.99 = 99\%$$

99.7% probability that x is found in the interval μ ± 3σ

a = 1.234(5): *a*=1.234; σ=**0.005** → *a* = 1.234 **± 0.015**

The refinement – Merging of reflections

1. Reading of the *.hkl file with the reflection intensities

Number of measured reflections

```
example.lst:

20975   Reflections read, of which   685   rejected

 -10 =< h =< 10,     -13 =< k =< 13,     -18 =< l =< 18,    Max. 2-theta =  137.96
       23   Systematic absence violations


 Inconsistent equivalents etc.
    h    k    l        Fo^2    Sigma(Fo^2)  N  Esd of mean(Fo^2)
    2    0    0    127128.20     518.46     4    3194.33
    0    2    0     65677.31     243.31     6    1223.07
 etc.
      14   Inconsistent equivalents
    2847   Unique reflections, of which       0   suppressed

 R(int) = 0.0275      R(sigma) = 0.0111      Friedel opposites merged
```

- Number of **independent** measured reflections (neither repeated measurements, nor related by symmetry)

- Merging is ruled by the MERG command (see SHELXL manual). We need to change this only in special cases (twinning or merging of (meaningless) Friedel pairs)

The refinement – R$_{sigma}$

1. Reading of the *.hkl file with the reflection intensities

---

example.lst:

20975 Reflections read, of which   685 rejected

…

2847 Unique reflections, of which   0 suppressed

R(int) = 0.0275   R(sigma) = 0.0111   Friedel opposites merged

---

½$I_{back.}$ $I_{reflection}$ ½$I_{back.}$

σ(I)

Integration window

$$R_{sigma} = \frac{\sum \sigma(F_o^2)}{\sum F_o^2}$$

- **Measure of the signal/noise ratio**

- **In first approximation, R1** (error of our model, see later) **cannot be much better than R$_{sigma}$**

The refinement – «Merging» – R_{int}

1. Reading of the *.hkl file with the reflection intensities

```
example.lst:

20975  Reflections read, of which    685  rejected
...
 Inconsistent equivalents etc.
   h   k   l      Fo^2   Sigma(Fo^2)  N   esd of mean(Fo^2)
   2   0   0   127128.20    518.46    4    3194.33
   0   2   0    65677.31    243.31    6    1223.07
 etc.
     14   Inconsistent equivalents
   2847   Unique reflections, of which      0   suppressed

 R(int) = 0.0275      R(sigma) = 0.0111      Friedel opposites merged
```
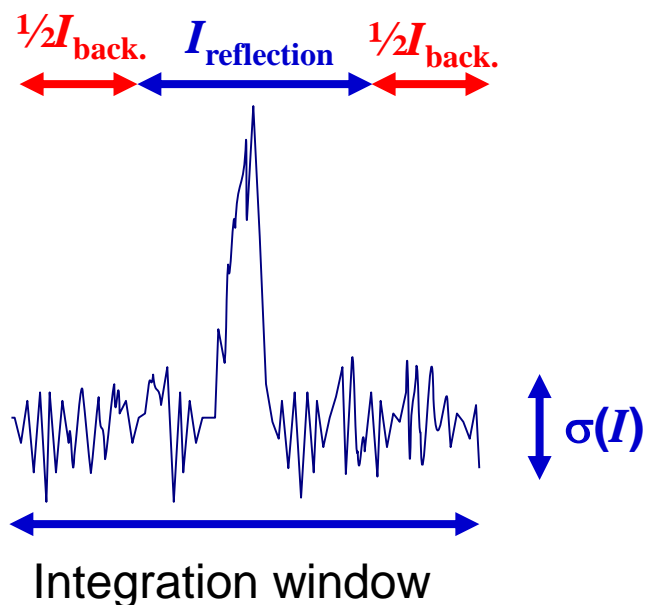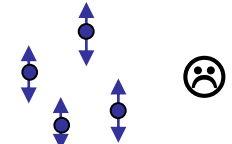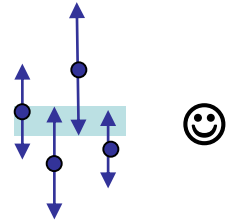
R_{int}

☺

☹

Inconsistent equivalents

$$R_{int} = \frac{\sum \left| F_o^2 - F_o^2(mean) \right|}{\sum F_o^2}$$

R_{int}: Error of merging (measure of precision / reproducibility of the measurement)

Possible reasons for high R_{int} (R_{int} is high if $R_{int} > 2\text{-}3 \cdot R_{sigma}$):

- Incorrect Laue group (crystal system)
- Missing or wrong absorption correction
- Crystal decomposition
- Twinning
- Goniometer problems (hidden reflections, bad alignment, etc.)

Confidence factors

Final Structure Factor Calculation for vivan6 in P2(1)/c

 Total number of l.s. parameters =   227 Maximum vector length = 623   Memory required =   49 /   2946 /   29281

 wR2 = 0.0800 before cycle   6 for   2847 data and   0 /   227 parameters

 GooF = S =   1.041;   Restrained GooF =   1.041 for   0 restraints

 Weight = 1 / [sigma^2(Fo^2) + (0.0457 * P)^2 +   0.57 * P]   where P = (Max (Fo^2, 0) + 2 * Fc^2) / 3

 wR2 = 0.0800, GooF = S = 1.041, Restrained GooF = 1.041 **for all data**
 R1 = 0.0307 for 2752 Fo > 4sig(Fo) and 0.0312 for all 2847 data

SHELXL minimises M, the difference of observed and calculated intensity:

$$M = \sum w \left(F_o{}^2 - F_c{}^2 \right)^2$$

(Thus the smaller is M, the better is the agreement between our model and the experimental data.)

Confidence factors

SHELXL minimises:
$$M = \sum w \left(F_o^{\,2} - F_c^{\,2} \right)^2$$

M increases with the number of reflections and their intensity. M is thus useful to optimise the structure, but not to compare the quality of the structure for two different experiment or to judge the quality of the model. We require thus a value independent from the number and intensity of the reflections.

Confidence factor, residual, R factor:

R2 →
$$wR_2 = R_w(F^2) = \sqrt{\frac{\sum w \left(F_o^{\,2} - F_c^{\,2} \right)^2}{\sum w (F_o^{\,2})^2}}$$

For purely statistical reasons, refinement against F^2 results in R factors twice as high than refinement against F. To facilitate comparisons (and to increase acceptance of the new refinement against F^2), SHELXL also calculates a factor R (or R1) based on F.

$$R_1 = \frac{\sum \left\| F_o \right| - \left| F_c \right\|}{\sum \left| F_o \right|} \quad \longleftarrow \textbf{R1}$$

R1 cannot be calculated with the weighing scheme optimized for F².
*Traditionally, it should also be calculated **only for reflections with I > 2s(I)**.*

A small detour – which variable to optimise?

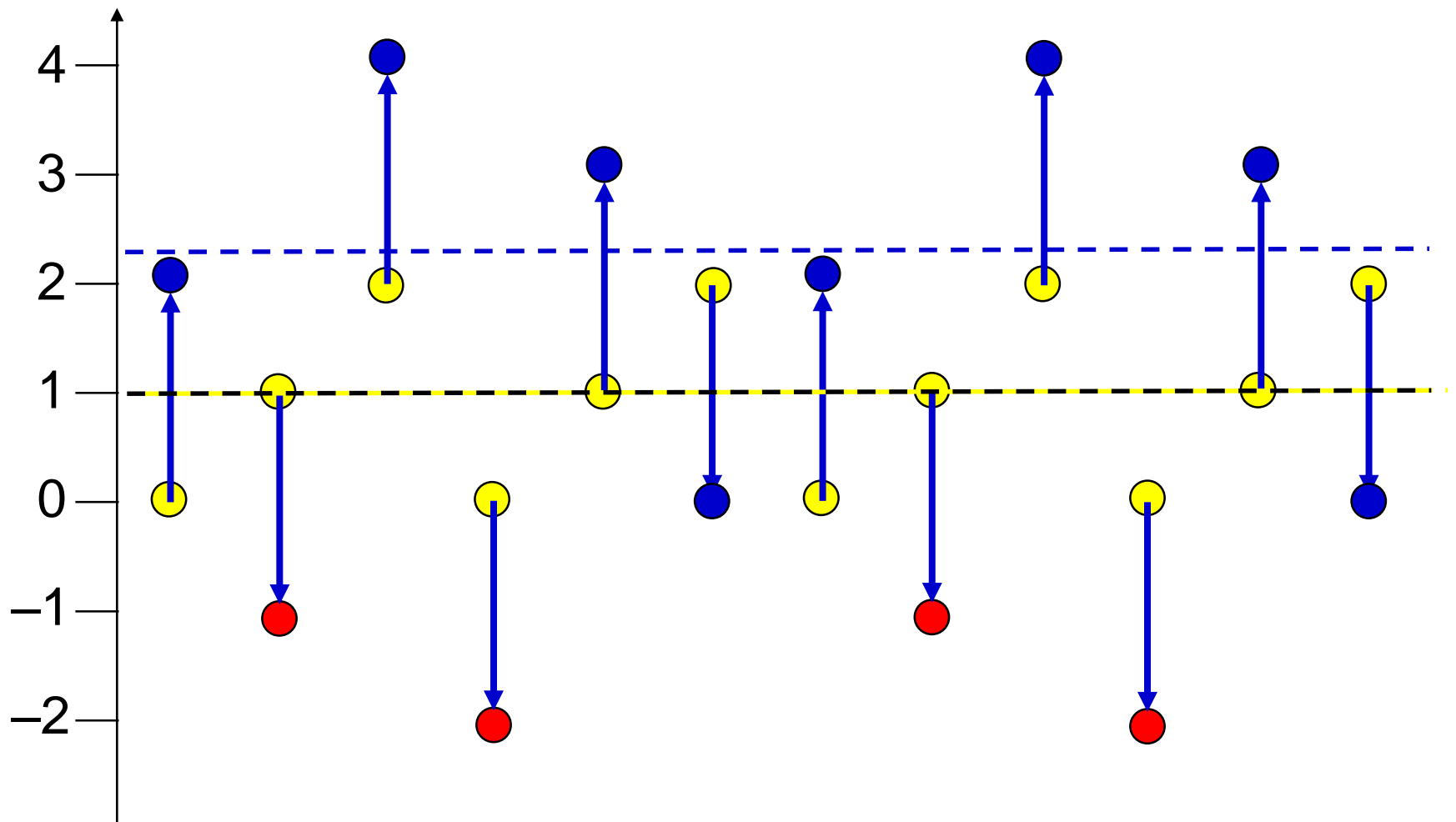$$M = \sum w\left(Y_o - Y_c\right)^2 \quad : \quad Y = F, F^2 \text{ or } I$$

All refinement programs minimise the difference between Y_c, calculated from our model, and the experimental value Y_o. But which value of Y to use?

Before 1993, the minimised value was typically the amplitude $|F|$ of the structure factor. This was logical, since this was the value obtained by the inverse Fourier transformation of our model $\rho_c \rightarrow F_{hkl,c}$. But refinement against F had certain problems:

- F_o is obtained from the square-root of the intensity: $F_o = \sqrt{I_o}$
- Since a square-root is not defined for negative values, all reflections with negative intensities had to be rejected.
- However, rejecting all negative intensities $I < 0$, introduced a bias towards higher intensities.

Literature:
Structure refinement: some background theory and practical strategies David Watkin *J. Appl. Cryst.* **2008**, *41*, 491–522.

- Assume a dataset with values between 0 and 2 and an average of 1 (yellow).
- The experimental values have an error of ±2 (blue and red).
- The average of the experimental values is of course also 1.
- However, if we eliminate all negative values (red), the average becomes now 2.25 (blue line).

A small detour – which variable to optimise?

$$M = \sum w(Y_o - Y_c)^2 \quad : \quad Y = F, F^2 \text{ or } I$$

All refinement programs minimise the difference between Y_c, calculated from our model, and the experimental value Y_o. But which value of Y to use?

Before 1993, the minimised value was typically the amplitude $|F|$ of the structure factor. This was logical, since this was the value obtained by the inverse Fourier transformation of our model $\rho_c \rightarrow F_{hkl,c}$. But refinement against F had certain problems:

- F_o is obtained from the square-root of the intensity: $F_o = \sqrt{I_o}$
- Since a square-root is not defined for negative values, all reflections with negative intensities had to be rejected.
- However, rejecting all negative intensities $I < 0$, introduced a bias towards higher intensities.
- To avoid this bias, we had to reject also all reflections with $F_o < 3\sigma(F)$.
- Given that the intensity diminishes with θ and that the reflections at high θ provide the best resolution, we lost a lot of important information in the high resolution reflections.

Some modern programs define $F_o = \text{sig}(I_o) \cdot \sqrt{\text{ABS}(I_o)}$, thus F_o retains the sign of I_o, which allows the use of negative intensities. However, the problem remains that the square-root is not a linear transformation and that it is difficult to correctly calculate $\sigma(F)$ from $\sigma(I)$.

A small detour – which variable to optimise?

$$M = \sum w\left(Y_o - Y_c\right)^2 \quad : \quad Y = F, F^2 \text{ or } I$$

With SHELXL93, Sheldrick introduced refinement against F^2 (initially proposed by Rollet) which is obtained by corrections and scaling from the intensity I: Lorentz-polarisation correction, absorption correction, etc.

- Refinement against F^2 allows to include negative intensities and $\sigma(F^2)$ is directly obtained from $\sigma(I)$.
- Refinement against F^2 is reported to be advantageous for twinned samples and for determining chirality.
- Refinement against F^2 is thus currently the most used method.
- Due to the square, reflections with weak intensities have a lower leverage in refinements against F^2 than in those against F. A good weighing scheme is thus essential for refinement against F^2.

Rollett, J. S. in *Crystallographic Computing 4*, edited by N. W. Isaacs & M. R. Taylor, **1988**, pp. 149–166. Oxford University Press.

A small detour – which variable to optimise?

$$M = \sum w(Y_o - Y_c)^2 \quad : \quad Y = F, F^2 \ or \ I$$

Comparable to refinements against F^2, we can refine directly against the measured intensity I. The advantage is that one actually refines against experimental data. However, the reverse of all corrections applied in data reduction (LP correction, absorption correction, etc.) needs to be applied to the amplitudes F_c^2 to calculate I_c.

- Refinement against I allows to include the parameters of the corrections into the refinement process. This poses, however, statistical problems due to the strong correlation between the absorption correction and the thermal displacement factors. Also, some crystallographers are uncomfortable to refine an external parameter (absorption = crystal form) together with the model (= structure).
- If correction parameters are not refined, refinement against I is identical to refinement against F^2, but requires additional calculations.

Although some advantages are associated with refinement against F^2, a committee of the IUCr did not arrive at a final conclusion which type of refinement is preferable. (Schwarzenbach et al. *Acta Cryst. A* **1989**, *45*, 63. In the end, if carefully done, all refinements should yield the same result.

Goodness of Fit

The GoF or GooF is another value describing the quality of our model:

$$GooF = S = \sqrt{\frac{\sum w(F_o^2 - F_c^2)^2}{N_{Ref.} - N_{Par.}}}$$

$N_{Ref.}$: number of independent reflections, $N_{Par.}$: number of parameters

Contrary to the R factor, which also depends on the noise, S is relatively independent of the noise. S, despite its way of calculation above, is the ratio of expected vs. observed deviations (errors).

S should have a value of appr. 1.

S > 1: errors larger than expected. Bad model or bad data/parameter ratio
S < 1: errors are smaller than expected. Typically indicates incorrect $\sigma(I)$) :
    problems with the absorption correction or the space group.

Criteria for a good structure

SHELXL calculates 4 confidence factors :

- **wR2 (all data)**
- **wR2 (observed data, I>2σ(I))**
- **R1 (all data)**
- **R1 (observed data, I>2σ(I))**

Refinement against F^2 requires an optimized weighing scheme, but this scheme cannot be used for calculation of R1.

The important factors are **wR2 (all data)** (since all reflections are used in refinement) and **R1 (observed data)**, for comparison purposes, since typically only reflections with $F < 3\sigma(F)$, above the observation threshold, were used in refinements against F. ($F < 3\sigma(F) \approx I < 2\sigma(I)$).

	Good	Acceptable	Problematic	Really problematic
R1	< 5%	< 7%	>10%	>15%
wR2	< 12%	< 20%	>25% (or » 2*R1)	>35%
S	0.9-1.1	0.8-1.5	<0.8 or >2	<0.6 or >4

These ranges are just providing a measuring stick. There are completely fine, publishable structures with R1=17% and unpublishable … stuff with R=4%. It depends on the *reason* for the bad values.

Correlation matrix elements

test.lst:
Largest correlation matrix elements

0.853 U11 Fe1 / OSF   **0.728 U11 S2 / U11 Fe1**   **0.524 U11 S1 / U11 Fe1**
0.771 U11 S2 / OSF   **0.588 U11 S1 / OSF**   **0.543 U12 S1 / U23 S1**

The correlation matrix, X, describes the correlation (interdependence) between a series of parameters in a refinement. The correlation of a variable with itself, X_{ii}, is always 1. The correlation between two variables, X_{ij}, is 1 (or –1) if they are linearly dependant and zero if they are independent.

The covariance matrix is essentially the same, but the correlation matrix is unit-free and scaled so that $X_{ij} = -1\ldots1$.

There is always some interdependency between variables. **Values in the correlation matrix > 0.5 , however, indicate significant correlation between variables and need to be inspected carefully.** They might indicate missed symmetry (or other problems due to overparametrization).

Correlation matrix elements

test.lst:
Largest correlation matrix elements

0.853 U11 Fe1 / OSF	**0.728 U11 S2 / U11 Fe1**	**0.524 U11 S1 / U11 Fe1**
0.771 U11 S2 / OSF	**0.588 U11 S1 / OSF**	**0.543 U12 S1 / U23 S1**

Values in the correlation matrix > 0.5 , however, indicate significant correlation between variables and need to be inspected carefully.

Some correlations are, however, anticipated and are no reason to worry:

- **Correlations between the ADPs of heavy atoms and the OSF or between the ADPs of heavy atoms** : A large quantity (if not the majority) of the electrons is concentrated in the heavy atoms. A diminution of the ADPs (=smaller integration area) can thus be compensated by an increase of the overall scale factor (OSF).

- **Correlation between ADPs of the same atom** : A rotation of 90°of the ADP ellipsoid around axis 1, transforms U_{22} into U_{33} and vice versa. A certain correlation between ADPs is thus normal, in particular if the ADP are relatively isotropic.

Correlation matrix elements

test.lst:

Largest correlation matrix elements

0.661 z C11 / z C61	0.650 x C7 / x C57	0.638 x C20 / x C70
0.660 y C12 / y C62	0.648 x C2 / x C52	0.636 z C3 / z C53
0.659 x C14 / x C64	0.647 x C6 / x C56	0.633 x C8 / x C58
0.657 x C11 / x C61	0.647 z C6 / z C56	0.633 y C11 / y C61
0.655 x C15 / x C65	0.647 x C16 / x C66	0.632 x C12 / x C62
0.654 x C17 / x C67	0.647 z C10 / z C60	0.631 x F11 / x F61
0.653 z C15 / z C65	-0.640 U11 C16 / U11 C66	0.631 z C9 / z C59
0.652 x C10 / x C60	0.639 z C12 / z C62	0.630 x C18 / x C68

Correlations between the atomic position of independent atoms or between ADPs of light atoms typically indicate a problem. In most cases, a missed symmetry (= wrong space group).

- Verify manually which atoms are correlated. It there a symmetry operation which could possibly correlate all these atoms?
- Check for other possible space groups?
- Use PLATON/ADDSYM
- Was another, more symmetric unit cell, offered during integration?

Thermal parameters

With the exception of wrong space group assignment, nearly all other problems of a structure refinement affect the thermal ellipsoids stronger than atom positions. For weak/bad structures, the thermal parameters thus become the garbage bin of the refinement. (Which is why crystallographers want to show ADPs with 50% probability while chemists prefer ball-and-stick pictures or smaller ellipsoids.)

In general :

- ADPs (= thermal movement) should be similar for similar atoms.

- The direction and size of the thermal ellipsoid should be in agreement with the lowest-energy vibration expected.
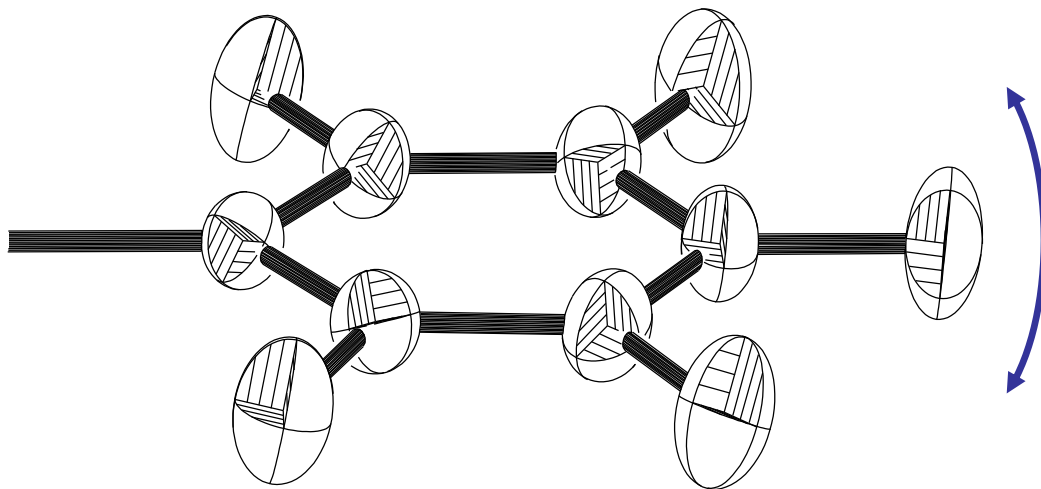
Manual inspection of thermal parameters

*.lst:
 Principal mean square atomic displacements U
[…]
  0.3098 0.0893 0.0464 C4   may be split into 0.6218 0.2673 0.2408 and 0.6118 0.2471 0.2666
  0.3100 0.0924 0.0392 C5   may be split into 0.5976 0.3191 0.3424 and 0.5834 0.3017 0.3597

Elongated ADPs often suggest unrefined disorder
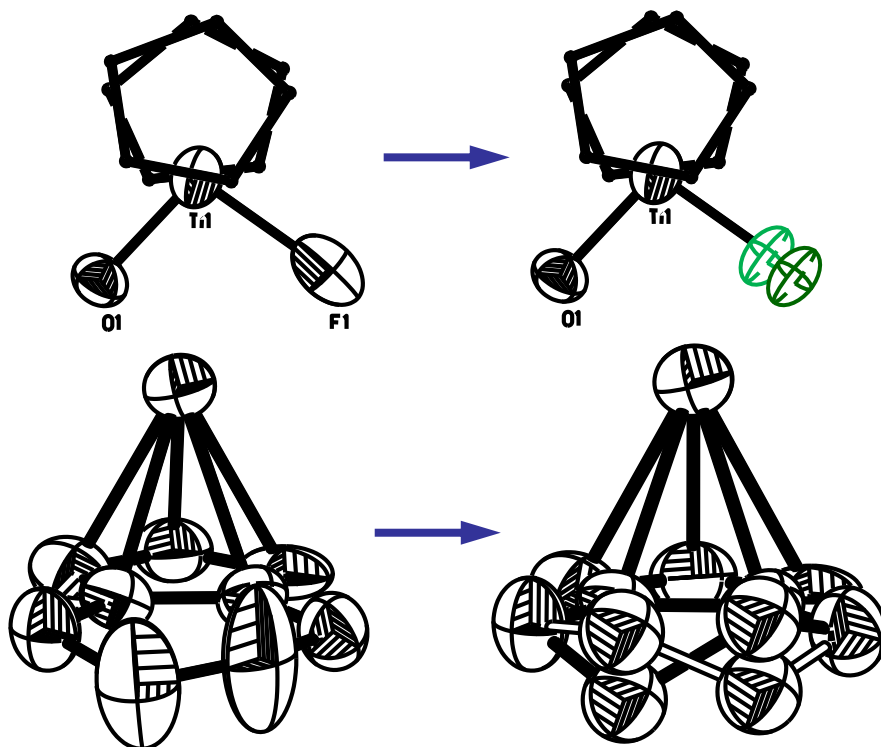
Of course this has to be confirmed by manual inspection

Most Disagreeable Reflections

Most Disagreeable Reflections (* if suppressed or used for Rfree).
Error/esd is calculated as sqrt(wD^2/<wD^2>) where w is given by the weight
formula, D = Fo^2-Fc^2 and <> refers to the average over all reflections.

h	k	l	Fo^2	Fc^2	Error/esd	Fc/Fc(max)	Resolution(A)
-1	0	4	34915.33	49953.90	6.93	1.000	3.80
0	5	4	44.72	18.59	6.14	0.019	1.96
-1	5	5	12.66	36.45	5.55	0.027	1.84
3	5	1	89.93	52.46	5.42	0.032	1.75
0	6	1	104.57	68.97	4.51	0.037	1.90
2	3	3	7.90	0.95	4.49	0.004	2.34
[…]							

Regardless how good your dataset, one reflection has to show the biggest deviation between calculated and observed intensity. **Most** disagreeable does not have to mean **very** disagreeable.

Look out for :

- A small number of reflections with much higher (at least 2x) the error of the others.
- Systematic tendencies (h = 3n, all high resolution, all low resolution…)
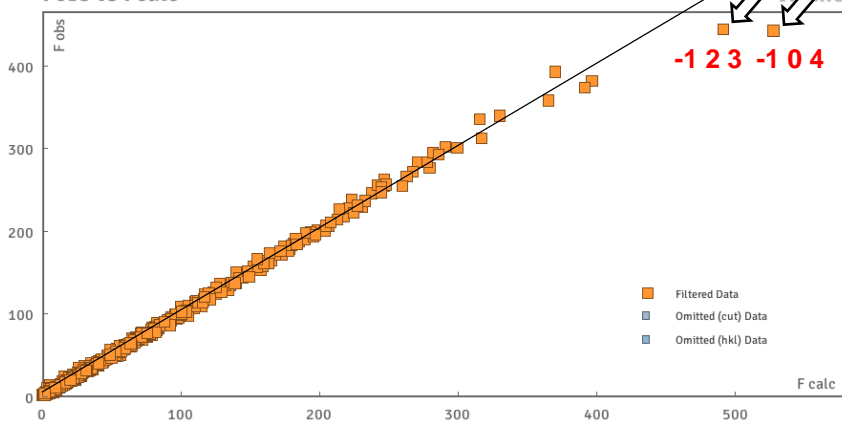
Most Disagreeable Reflections – OLEX2

The same list is available in OLEX2 under INFO-Bad reflections. This list is good to spot systematic tendencies (h=3n). Other systematic errors are better spotted with the statistics plots in OLEX2.

Normal Probability Plot   vivan6

$y = 1.666x + 0.151$
$R = 0.990$

0 5 4
3 5 1
-2 1 4
-1 5 5

Fobs vs Fcalc   vivan6

-1 2 3 -1 0 4

- Filtered Data
- Omitted (cut) Data
- Omitted (hkl) Data
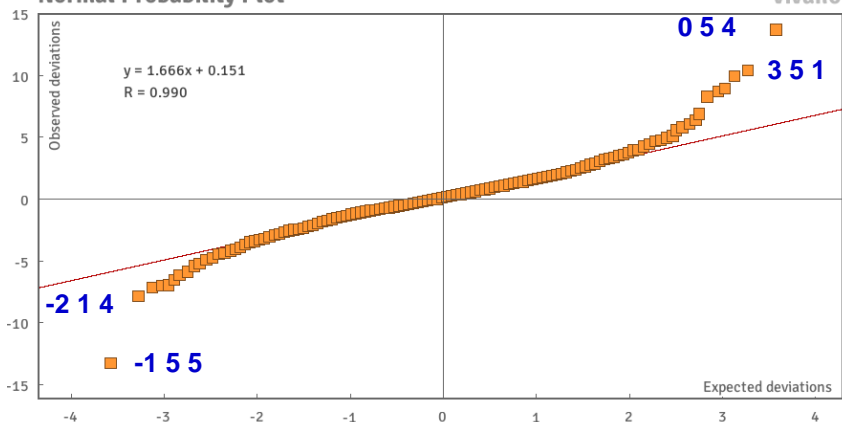
F calc

| Home | Work | View | Tools | **Info** |

Recent Files

Electron Density Peaks

Refinement Indicators

Bad Reflections

| OMIT | all reflections where |Error/esd| > than | 10 | **ClearOMITs** |

Unsigned sorting by Error/esd (refine to update)

H	K	L	Error/esd	d/A	
-1	0	4	6.93	3.80	omit Edit...
0	5	4	-6.14	1.96	omit Edit...
-1	5	5	5.55	1.84	omit Edit...
3	5	1	-5.42	1.75	omit Edit...
0	6	1	-4.51	1.90	omit Edit...
2	3	3	-4.49	2.34	omit Edit...
1	6	1	-4.35	1.84	omit Edit...
-3	2	3	4.30	2.52	omit Edit...
-1	2	3	4.20	3.72	omit Edit...

Click to omit reflections from refinement

*.ins
OMIT -1 2 3
OMIT -1 0 4

R1 :
3.07% → 2.95%

Residual electron density

```
*.lst
Electron density synthesis with coefficients Fo-Fc

Highest peak    0.20  at  0.4190  0.6862  0.1604  [  0.69 A from N3 ]
Deepest hole   -0.33  at  0.7000  0.3938  0.1491  [  0.38 A from P1 ]
```

If our model is correct, we should have correctly described all electron density. Thus the remaining electron density ($\rho - \rho_{modèle}$) should be zero. However, we still get notable maxima due to remaining noise, electron density in bonds, Fourier truncation errors, insufficient description of disorder, etc.

Acceptable values for remaining electron density :
- For light-atom structures (H – F) : **< 0.5 e$^-$/Å^3**
- For structures with heavy atoms : up to 10% of the electron density of the heavy atom (although rarely that high) can be found in a distance of around 1 Å of the heavy atom (Fourier truncation errors).
- Remaining electron density > 0.5 e$^-$/Å^3 can be accepted if due to a "good reason", i. e. nothing to be done about it. But it needs to be inspected and explained (comment in CIF, see later).
- Do look also at the biggest hole! If electron density is simply due to remaining noise, the whole should be as big as the maximum.

Fourier truncation errors

Fourier truncation errors appear if the number of terms of the infinite sum of the Fourier transformation is limited. They appear as "wiggles" in places where electron density changes drastically.
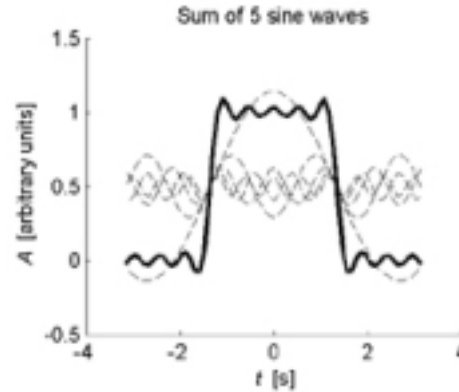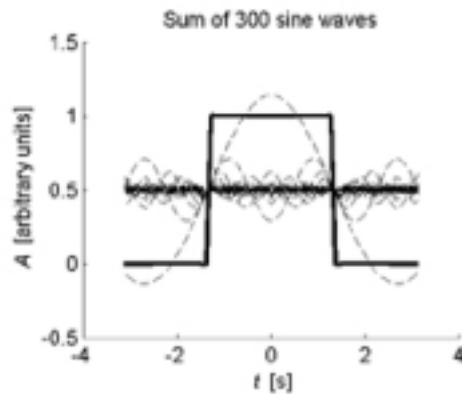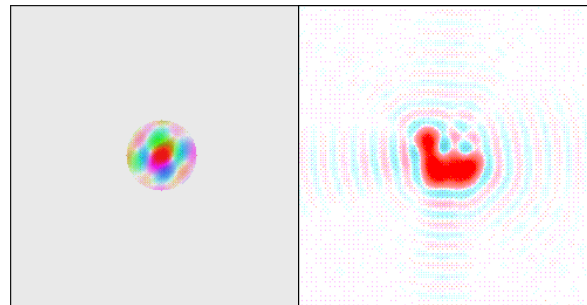

Image: http://www.theses.ulaval.ca/2005/23016/apd.html

High θ suppressed

Low θ suppressed

© Kevin Cowtan, York : http://www.ysbl.york.ac.uk/~cowtan

Fourier truncation errors
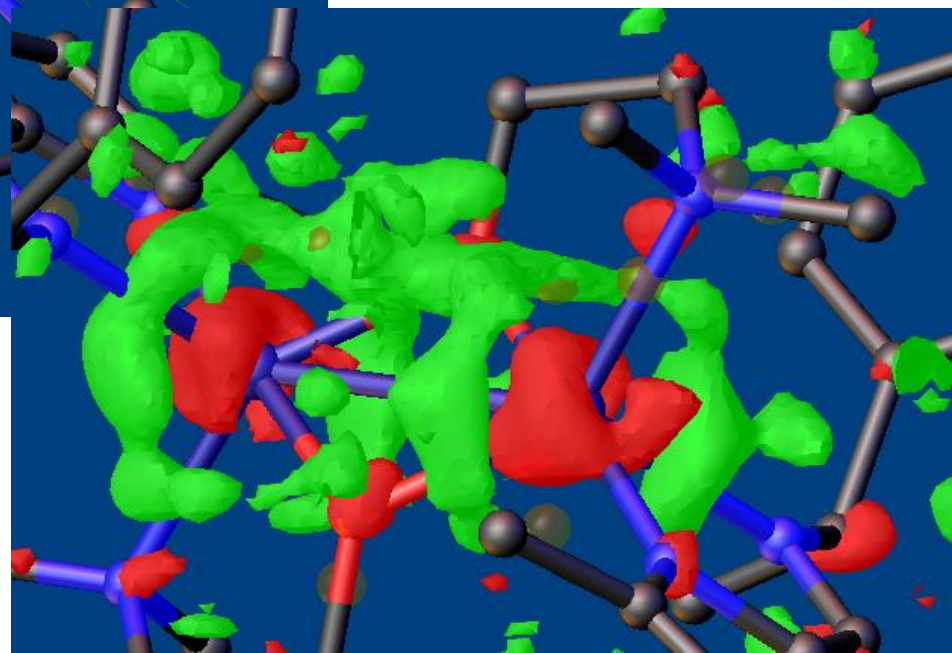
How to find and what to do with residual electron density

OLEX2 : info Q1

test.cif :

_refine_diff_density_max 0.425
_refine_diff_density_min -0.560

test.lst :

Electron density synthesis with coefficients Fo-Fc
 Highest peak   0.42 at 0.2140 0.0398 0.5135 [0.40 A from C2]
 Deepest hole -0.56 at 0.1424 0.4014 0.6865 [0.53 A from F3]

Possible reasons and consequences for $\rho > 0.5$ e$^-$/Å^3

- Missing/misassigned hydrogen atoms → Correct the problem.

- Fourier truncation errors → If close (appr. 1 Å) to a heavy atom or on a special position: explain in CIF. If not, it's not a Fourier truncation error.

- Bad absorption correction → Correct the problem.

- Disorder → Correct the problem. If not possible, explain in CIF.

- Twinning → Correct the problem.

- Electron density in bonds → Should not be > 0.5 e$^-$/Å^3

- Unknown → Explain in CIF why you can exclude all of the above and admit not knowing the reason. This, of course, reduces the quality of the structure.