Chemical Crystallography and Structural Chemistry

VO 270287

Lecture № 6 — 19th May 2022

Dr. Tim Grüne
Centre for X-ray Structure Analysis
Faculty of Chemistry
University of Vienna

tim.gruene@univie.ac.at

Course Details

3rd	March	Lecture № 1	10th	March	Lecture № 2
17th	March	**no lecture**	24th	March	Exercise № 1
31st	March	Lecture № 3			
7th	April	Discussion Ex. 1	14th	April	**Easter break**
21st	April	**Easter break**	28th	April	Lecture № 4
5th	May	Power cut	12th	May	**no lecture**
19th	May	Lecture № 6	26th	May	Exercise № 3
2nd	June	Lecture № 7	8th	June	Lecture № 8
16th	June	Exercise № 4	23th	June	Lecture № 9
30th	June	**no lecture**			

Previous Lecture

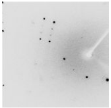
- Symmetry and Diffraction data
- 11 Laue groups
- Friedel's law
- Some aspects on data collection
- Indexing

Contents

1 Experimental procedure: from data collection to structure

From Data Collection to Structure

Data collection → **Data integration** → **Data Scaling**



several GB

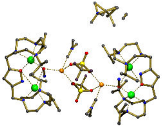
```
0 0 -1       2.7   0.9
0 0  1       4.0   1.0
0 0 -2 1'257.0  35.5
0 0 -2 1'600.0  42.7
```

several files, 100's MB

```
0 0 -1       2.8   0.55
0 0  1       3.8   0.63
0 0 -2 1'432.0  95.7
0 0 -2 1'282.0  85.9
```

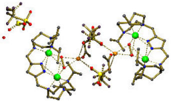
1 "hkl"-file, 50MB

→ **Phasing**



Starting model

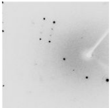
→ **Refinement**



Chemically sensible model

Data Collection

Data collection



several GB

Data integration

```
0 0 -1      2.7   0.9
0 0  1      4.0   1.0
0 0 -2 1'257.0  35.5
0 0 -2 1'600.0  42.7
```

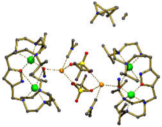
several files, 100’s MB

Data Scaling

```
0 0 -1      2.8  0.55
0 0  1      3.8  0.63
0 0 -2 1'432.0  95.7
0 0 -2 1'282.0  85.9
```

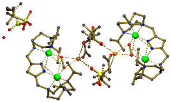
1 “hkl”-file, 50MB

Phasing



Starting model

Refinement



Chemically sensible model

How to collect good data

Data resolution

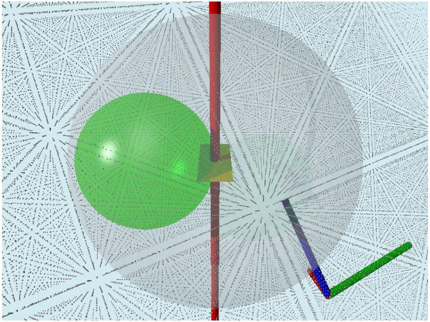
Data completeness

Data multiplicity

Data quality

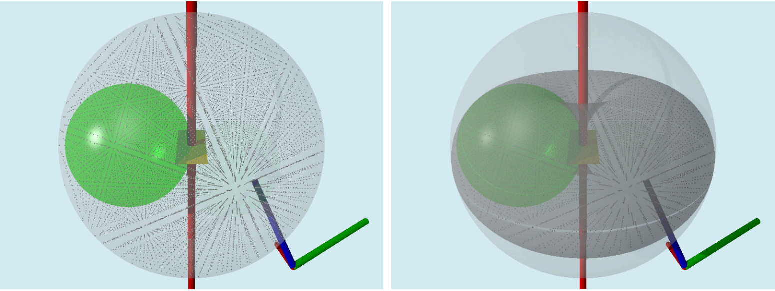
Data resolution

- highest possible resolution: $d_{\text{min}} > \lambda/2$
- good crystals diffract better than this limit (instrumentation limited resolution)
- bad crystals may not reach this resolution (sample limited resolution)
- NB: low resolution are also important



Data completeness

- Only reflections with resolution sphere can be collected (radius in reciprocal space: $r = 2./\lambda$)
- Laue equations permit to compute their location, and how to rotate the crystal into an appropriate position.
- Single rotation axis: torus shape of reflections.

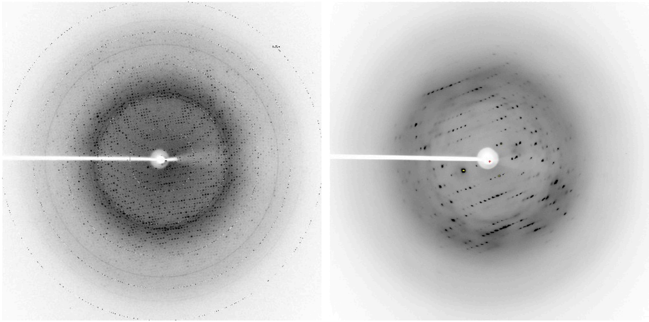


Data multiplicity

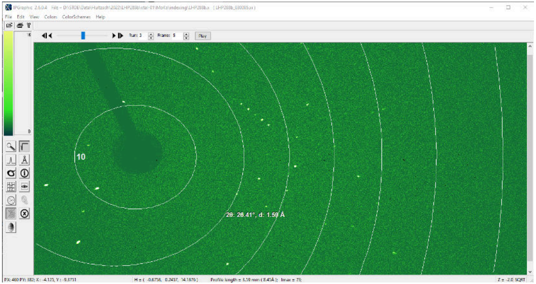
- Intensity $(I(hkl)$ of the identical reflection (hkl) can be measured multiple times
- Intensities $I(hkl)$ and $I'(R(hkl))$ symmetry equivalent reflections increase multiplicity
- Measuring on different positions on detector: reduction of systematic errors
- increasing multiplicity reduces uncertaintig $\sigma(I)$

Data quality

- "Diffraction data acquisition is the final experimental stage of the crystal structure analysis." (Z. Dauter, [1])
- Ensure proper data collection strategy (desired resolution, multiplicity; adjust detector distance to unit cell parameters, etc.)
- Crystal quality can be the limiting factor

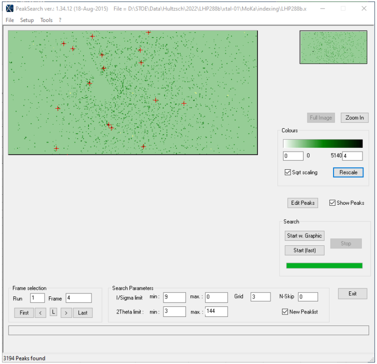


Steps of data integration: initial diffraction images



- typically four short runs:
- 2×10°–15°: 0°–10° and 90°–100°
- two detector settings each (low to mid resolution and mid to high resolution)

Steps of data integration: peak finding



PeakSearch ver.: 1.34.12 (18-Aug-2015) File = D:\STOE\Data\Hultzsch\2022\LHP288b\xtal-01\MoKa\indexing\LHP288b.x

Colours: 0 0 5140 4

☑ Sqrt scaling

☑ Show Peaks

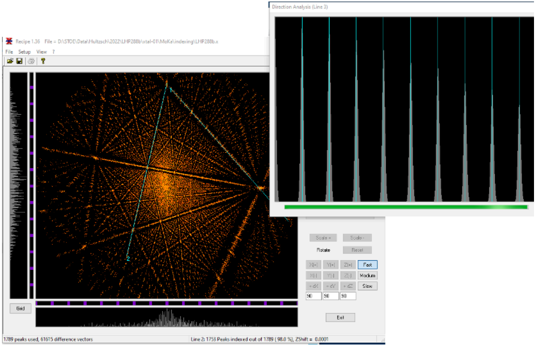
Frame selection: Run 1 Frame 4

Search Parameters:
I/Sigma limit min : 9 max. : 0
2Theta limit : min : 3 max. : 144
Grid 3
N-Skip 0
☑ New Peaklist

3194 Peaks found

- simple (non-crystallographic) pattern recognition algorithm: peak above background
- main parameter: desired peak strength (I/sigma)

Steps of data integration: finding unit cell



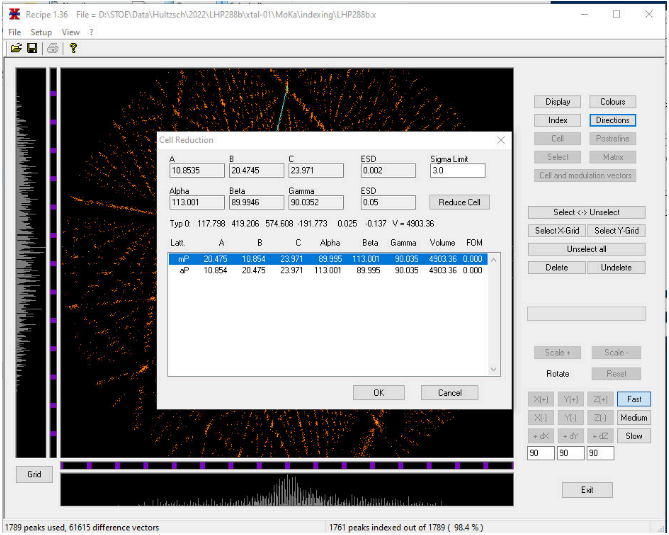
90 90 90

1789 peaks used, 61615 difference vectors

Line 2: 1753 Peaks indexed out of 1789 (98.0 %), ZShift = 0.0001

- known instrument geometry: project peaks into reciprocal space
- find a lattice basis

Steps of data integration: finding unit cell



Cell Reduction

A	B	C	ESD	Sigma Limit
10.8535	20.4745	23.971	0.002	3.0

Alpha	Beta	Gamma	ESD
113.001	89.9946	90.0352	0.05

Typ 0: 117.798 419.206 574.608 -191.773 0.025 -0.137 V = 4903.36

Latt.	A	B	C	Alpha	Beta	Gamma	Volume	FOM
mP	20.475	10.854	23.971	89.995	113.001	90.035	4903.36	0.000
aP	10.854	20.475	23.971	113.001	89.995	90.035	4903.36	0.000

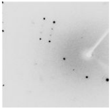
Rotation angles: 90 90 90

1789 peaks used, 61615 difference vectors

1761 peaks indexed out of 1789 (98.4 %)

2 Data Integration

Data collection → **Data integration** → **Data Scaling**



several GB

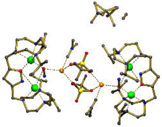
```
0 0 -1       2.7    0.9
0 0  1       4.0    1.0
0 0 -2 1'257.0   35.5
0 0 -2 1'600.0   42.7
```

several files, 100’s MB

```
0 0 -1       2.8   0.55
0 0  1       3.8   0.63
0 0 -2 1'432.0   95.7
0 0 -2 1'282.0   85.9
```

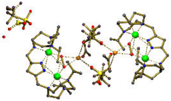
1 “hkl”-file, 50MB

→ **Phasing**



Starting model

→ **Refinement**



Chemically sensible model

Data Integration

Data integration comprises

1. Indexing: Determination of unit cell dimensions, orientation of the crystal, point group
2. Extraction of spot intensities from detector images.
3. Optimisation of experimental parameters

Programs for data integration (incomplete)

Saint Licensed by Bruker AXS. Specific to Bruker programs. Good for data from twinned crystals. Derived from XDS.

X-Area Software suite distributed with Stoe X-ray diffractometers. Free to use. Very good interactive indexing routine, good for twinned data.

XDS Free for non-commercial users (`http://xds.mpimf-heidelberg.mpg.de`). Supports nearly all detector formats, very well documented. Very fast.

DIALS Free for non-commercial users. Very active development. (`https://dials.diamond.ac.uk/`)

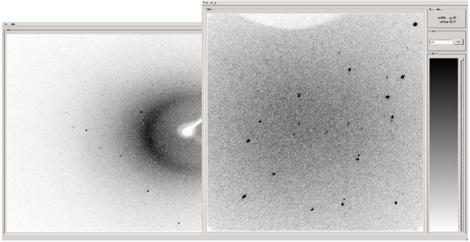
EVAL Suite Free for non-commercial users (`http://www.crystal.chem.uu.nl/distr/eval`). Can integrate *e.g.* incommensurate crystals

Crysalis Pro Licensed by Rigaku (`https://www.rigaku.com/products/smc/crysalis`)

iMosflm Free for non-commercial users, distributed with CCP4 (`http://www.ccp4.ac.uk`)

HKL3000 Very good visualisation GUI for fine-tuning of parameters. very popular in the US (`https://hkl-xray.com/`)

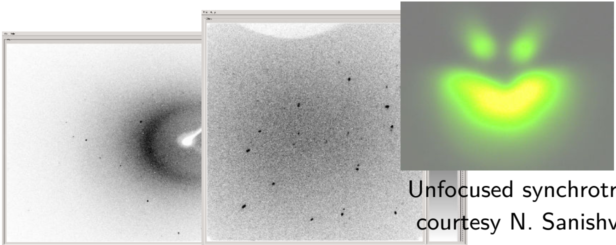
Indexing



“Ewald sphere backwards:”

- find 200–1000 **strong** spots
- backtransform into reciprocal space (Laue equations)
- find a lattice and a suitable basis
- basis for reciprocal lattice corresponds to reduced unit cell constants

Indexing



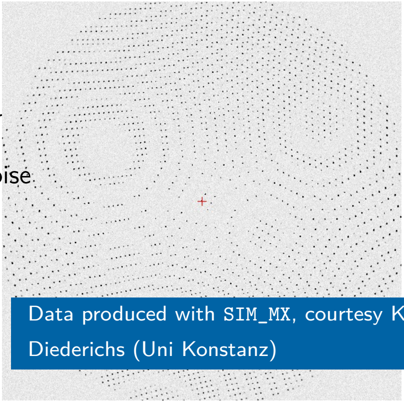
Unfocused synchrotron beam, courtesy N. Sanishvili, APS, Chicago, USA

Possible reasons for indexing problems:

- Incorrect parameters: detector distance, direction of rotation, wavelength, especially at synchrotrons
- Too few reflections
- Distorted spots (lattice defects, unfocused beam)
- Alien spots (ice, metal, contaminant)
- multiple lattices *twins*

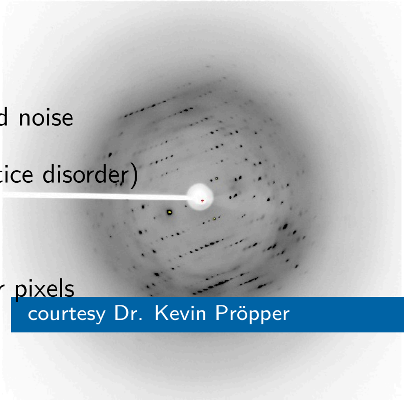
Ideal diffraction image

- no lattice disorder
- no background noise
- perfect beam



Data produced with `SIM_MX`, courtesy Kay Diederichs (Uni Konstanz)

Realistic diffraction image



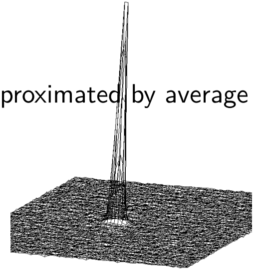
courtesy Dr. Kevin Pröpper

- strong background noise
- smeary spots (lattice disorder)
- spot overlap
- saturated detector pixels
- finite resolution

Signal extraction

1. Calculate reflex positions
2. Determine local background
3. Differentiate background from signal (spot volume, shape)
4. Different approach for strong spots and weak spots

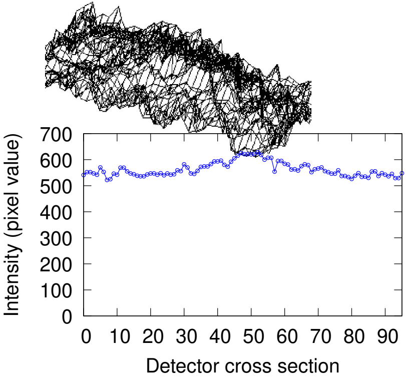
Signal extraction: strong spots



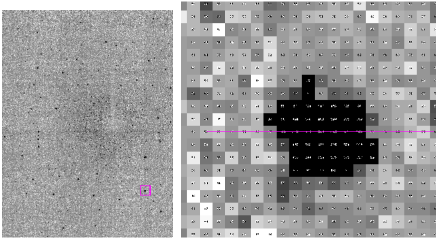
- High intensity ⇒ Small error from noise
- Spot covers large detector area: noise approximated by average
- Good spot separation
- Good spot profile (shape)

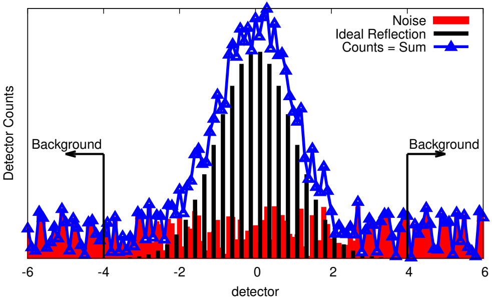
Signal extraction: weak spots

- Low intensity: ⇒ high error from noise
- Spot covers small detector area: large effect from noise
- Background difficult to determine



Cross section of a spot on the detector





Reflection profile

Most integration programs create a set of reflection profiles from strong and reliable reflections, *e.g.* a 3D Gauss function.
The profiles depend on the region on the detector and on the crystal orienation.
Advantages:

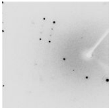
1. Measurement of weak reflections (fitting data for profile)
2. takes non-isotropic crystal shape into accoung
3. takes regions of varying detector sensitivity into account
4. produces a standard deviation of the reflection intensity: (h, k, l, I, σ_I)

Summary Data Integration

- Starts with indexing: crystal orientation, unit cell
- Look at all images per run
- Look only at calculated spot positions on detector
- Strong spots: sum pixel values, substract background
- Strong spots: determine average reflection profile
- Weak spots: extract data based on profiles

3 Scaling

Data collection → **Data integration** → **Data Scaling**



several GB

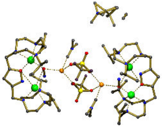
```
0 0 -1      2.7    0.9
0 0  1      4.0    1.0
0 0 -2 1'257.0   35.5
0 0 -2 1'600.0   42.7
```

several files, 100's MB

```
0 0 -1      2.8   0.55
0 0  1      3.8   0.63
0 0 -2 1'432.0   95.7
0 0 -2 1'282.0   85.9
```

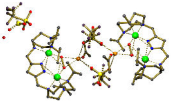
1 "hkl"-file, 50MB

→ **Phasing**



Starting model

→ **Refinement**



Chemically sensible model

Objective of the diffraction experiment

Structure elucidation of a chemical compounds

- chemical composition (*e.g.* purity after chromatography)
- Connectivity, distances between (non-) bonded atoms
- Configuration of stereochemical centres (R,S)

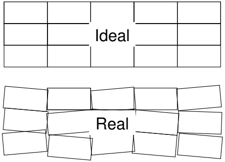
The values should be independent from the experimental setup.
Scaling makes the raw intensities (from data integration) independent from the experimental setup

Calculation of Intensities

Under consideration of the experimental, non-idealised setup, intensities are calculated as [2, Ch. Diffraction Intensities]

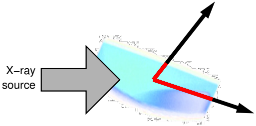
$$I_{exp}(hkl) = \frac{e^4}{{m_e}^2 c^4} \frac{\lambda^3 V_{crystal}}{V_{u.c.}^2} I_0 LPTE |F_{theor.}(hkl)|^2 \tag{1}$$

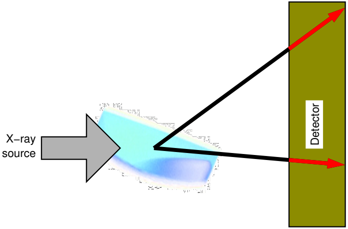
- I_0 incoming intensity (may vary with time)
- L Lorentz factor describes trajectory through the Eqwald sphere
- P Polarisation correction; $P = (1 + \cos^2 2\theta)/2$ for unpolarised source
- T Absorption correction (esp. heavy elements)
- E extinction correction (crystal defects, mosaicity)
- $F_{theor.}(hkl)$ structure factor (calculated from structure)



Scaling = Idealisation and correction

In order to make data as independent from the experiment as possible, data are "standardised". Some corrections are of numerical nature (polarisation), others are sample dependent (extinction: depends on elements in compounds). Two examples: absorption and angle of incidence





Absorption in the crystal depends on the path

Higher angle of incidence w.r.t. detector surface leads longer path through detector phosphor and thus to stronger signal

Basis for Scaling: Symmetry and multiple measurements

- Some corrections depend on the instrument and can be calibrated (polarisation, angle of incidence, Lorentz factor)
- Some corrections (*e.g.* absorption, extinction) are (also) sample dependent

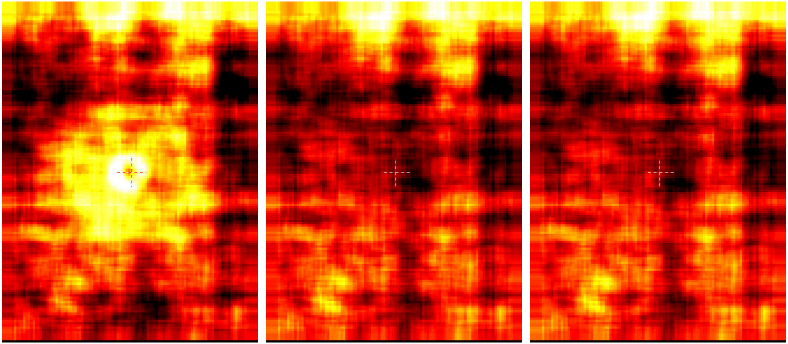
Idea: symmetry equivalent reflections, or multiply measured reflections, should have the same intensity

Scaling means

1. Determination of the measured intensity I_{obs} of a set of equivalent reflections
2. Determination of their standard uncertainties
3. Result: idealised data set

Examples for corrections

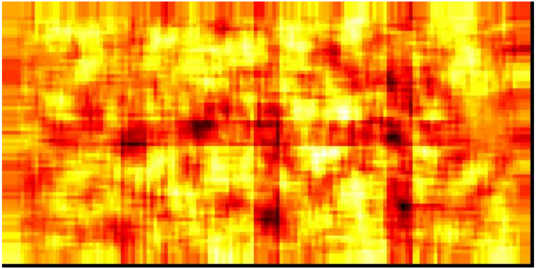
CMOS Detector (Bruker Photon II)



Detector “gain”: dark regions of the detector are more sensitive (factor between 0.97 and 1.3). Direct beam (left): indicates non-linear response at high intensity

Examples for corrections

Hybrid pixel detector (DECTRIS Eiger2 R500)

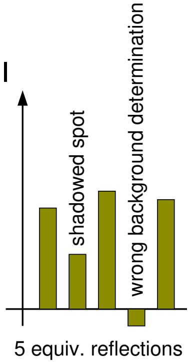


Hybrid pixel detectors are closer to ideal detectors for X-rays, with a smooth gain (factor between 0.96 and 1.02).

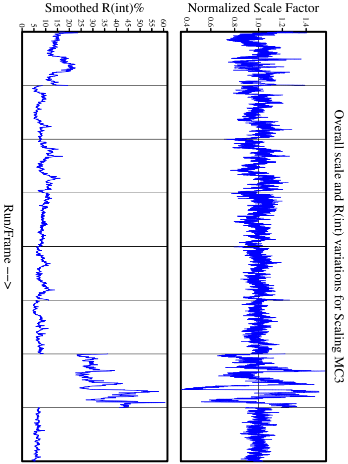
Scaling details

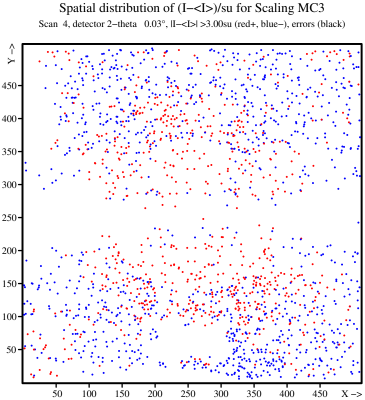
Options to find the mean intensity $< I >$ of a group of equivalent reflections

1. Statistical average: exaggerates outliers
2. Scaling of each group of equivalents independently from other groups: neglects systematic errors.
3. Instead: **One** scale factor for several groups of equivalent reflections (*e.g.* 100 groups) which are close together on the detector



Scaling plots (program SADABS)


Overall scale and R(int) variations for Scaling MC3


Spatial distribution of (I−<I>)/su for Scaling MC3
Scan 4, detector 2−theta 0.03°, |I−<I>| >3.00su (red+, blue−), errors (black)

Idealised intensities

Before Scaling

$$I(hkl) \quad = \frac{e^4}{{m_e}^2 c^4} \frac{\lambda^3 V_{\text{crystal}}}{V_{\text{u.c.}}^2} I_0 LPTE|F(hkl)|^2$$

```
0 -1  5  1.379E+03  2.516E+02
0 -1 -5  1.367E+03  2.726E+02
0  1  5  1.184E+03  2.610E+02
0  1 -5  1.347E+03  2.674E+02
0 -1  6  1.090E+04 -1.229E+03
0 -1 -6  4.677E+03  5.733E+02
0  1  6  4.286E+03  5.488E+02
0  1 -6  9.065E+03 -1.034E+03
0 -1  7  0.204E+02  0.571E+01
```

After Scaling

$$I(hkl) \quad = c|F(hkl)|^2 \qquad (2)$$

```
0 -1  5  7.014E+01  1.208E+01
0 -1 -5  6.812E+01  1.274E+01
0  1  5  5.987E+01  1.231E+01
0  1 -5  6.753E+01  1.258E+01
outlier removed
0 -1 -6  2.365E+02  2.856E+01
0  1  6  2.145E+02  2.689E+01
outlier removed
0 -1  7  1.404E+02  2.271E+01
```

Summary Scaling

- Scaling finds the mean intensity $< I >$ of symmetry equivalent reflections
- Scaling puts one or more data sets on a common scale
- Scaling removes outliers
- The result of scaling is a single (hkl)-file, which is used for all subsequent steps:
    1. Phasing, *i.e.* structure solution
    2. Model refinement

4 Intermezzo: The structure factor

The Structure Factor $F(hkl)$

- Context between atoms and diffraction intensities
- Describing the electron density with the independent atom model (“IAM”)
- Formfactor and the “fudge factor” ADP

The Structure Factor

The structure factor $F(hkl)$ is related to the **electron density** $\rho(x, y, z)$, i.e. the distribution of electrons inside the unit cell:

$$F(hkl) = \int_{\text{unit cell}} \rho(x, y, z) e^{2\pi i(hx+ky+lz)} d^3x \tag{3}$$

This equation is the Fourier transformation of the electron density.
Note: The term “Fourier transformation” is important mainly because computers are very fast in calculating Fourier transformations.

The Structure Factor

The *Independent Atom Model* (IAM, alias *isolated atom model*) is a powerful method to calculate the atomic structure factor $F(hkl)$

$$F(hkl) = \sum_{\text{in u.c.}}^{\text{atoms } j} f_j(\theta) e^{-8\pi^2 U_j(\theta,\lambda)} e^{2\pi i(hx_j+ky_j+lz_j)} \tag{4}$$

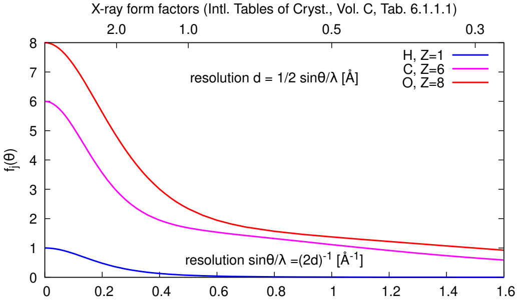
f_j atomic *form* factor. Dependent on atom element, decreases with decreasins scattering angle $\theta = \theta(hkl)$

$U_j(\theta, \lambda)$ atomic displacement parameter (ADP, alias Debye-Waller factor): models thermal vibration of atoms

$e^{2\pi i(hx_j+ky_j+lz_j)}$ phase shift of the atom relative to the origin of the unit cell

The form factor $f_j(\theta)$

Intensity of the scattered X-rays decreases with increasing scattering angle θ.



The wavelength λ is of the same order of magnitude as the size of the atoms: photons "see" the shape of atoms.

Note: hydrogen atoms do not contribute to data higher than 1 Å

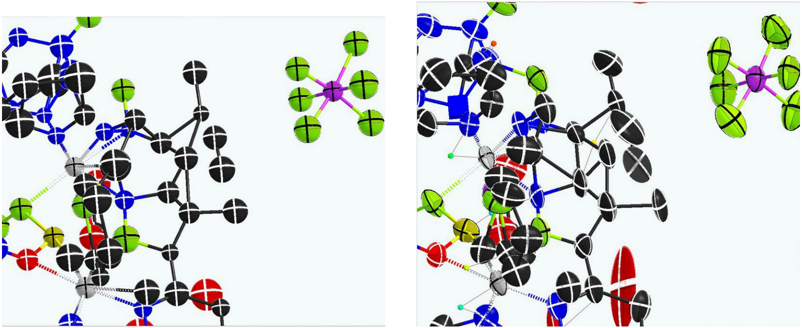
The atomic displacement parameter $U_j(\theta, \lambda)$

- Atoms vibrate at $T > 0K$
- Vibration leads to reduction of spot intensities, but not to change in spot shape
- At medium resolution: 1 parameter
- At high resolution: anisotropic description with 6 parameters as ellipsoids

Isotropic	**Anisotropic**
$U_j(\theta, \lambda) = 4U_{\text{iso}}\frac{\sin^2\theta}{\lambda^2}$	$U_j(\theta, \lambda) = (hkl)\begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{12} & U_{22} & U_{23} \\ U_{13} & U_{23} & U_{33} \end{pmatrix}\begin{pmatrix} a^* \\ b^* \\ c^* \end{pmatrix}$
1 parameter per atom	6 parameters per atom

- Name : ADP = isotropic or anisotropic *atomic displacement parameter*

Example images for ADP

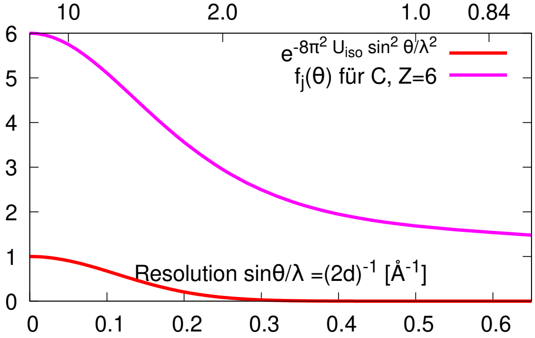


Refinement with isotropic ADPs

Refinement with anisotropic ADPs

Caveat: Anisotropic ADPs “wipe up” errors in the data.

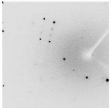
The ADP U: a fudge factor



- Sharp drop-off with resolution: can make wrongly placed atoms disappear.
- Similarity with form factor: confusion of atom types
- Risk of overfitting

5 Phasing

Data collection → **Data integration** → **Data Scaling**



several GB

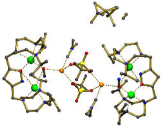
```
0 0 -1       2.7    0.9
0 0  1       4.0    1.0
0 0 -2 1'257.0   35.5
0 0 -2 1'600.0   42.7
```

several files, 100's MB

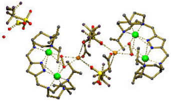
```
0 0 -1       2.8   0.55
0 0  1       3.8   0.63
0 0 -2 1'432.0   95.7
0 0 -2 1'282.0   85.9
```

1 “hkl”-file, 50MB

→ **Phasing** → **Refinement**



Starting model



Chemically sensible model

Phasing *alias* Solving the structure

- Ideally, we would measure data, run a single calculation, and get the molecular structure as result.
- our data: thousands of measured intensities $I(hkl)$.
- From Eq. 2 (p. 40) and Eq. 4 (page 45):

$$\begin{aligned} I(hkl) &= c|F(hkl)|^2 \\ &= c\left|\int_{\text{unit cell}} \rho(x,y,z)e^{2\pi i(hx+ky+lz)}d^3x\right|^2 \end{aligned}$$

- **If** we could invert this equation, we could calculate the coordinates x, y, z in one go. But we cannot ...

The phase problem

The inverse of the Fourier transformation

$$F(hkl) = \int_{\text{unit cell}} \rho(x,y,z) e^{2\pi i(hx+ky+lz)} d^3x$$

reads

$$\rho(x,y,z) = \sum_{(h,k,l)} F(hkl) e^{-2\pi i(hx+ky+lz)}$$

The phase problem

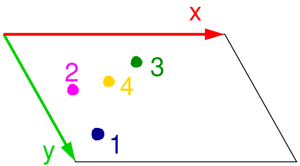
The structure factor $F(hkl)$ is a complex number. Therefore, it has

an amplitude $|F(hkl)| = \sqrt{I(hkl)/c}$

a phase $\phi(hkl) = ???$

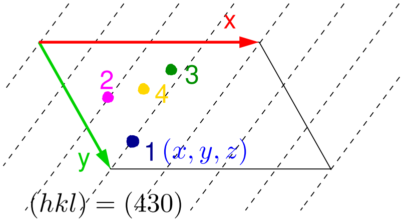
$F(hkl) = \sqrt{I(hkl)/c} \times e^{-i\phi(hkl)}$
We can measure the amplitude, but we cannot measure the phase. This is known as the **phase problem of crystallography**.

Relation between Phases and Molecule



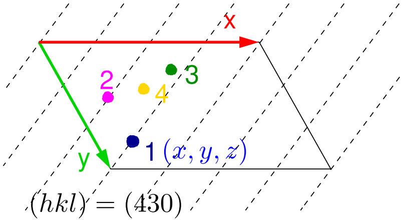
Unit cell with four atoms inside.

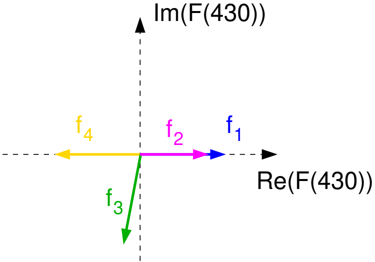
The phases are related to the chemical structure



Example reflection $F(430) = \sum_{\text{j=1}}^{4} f_j(\theta) e^{-8\pi^2 U_j(\theta,\lambda)} e^{2\pi i(4x_j + 3y_j + 0z_j)}$

The phases are related to the chemical structure





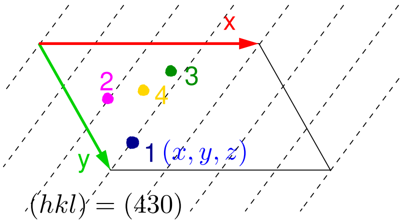
$$\phi_j = 2\pi(4x_i + 3y_i + 0z_i)$$

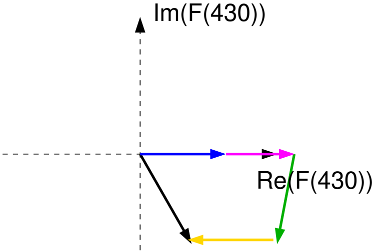
= rel. distance to origin

= rel. distance to lattice plane

$$F(430) = \sum_{\mathrm{j}=1}^{4} f_j(\theta) e^{-8\pi^2 U_j(\theta,\lambda)} e^{2\pi i(4x_j+3y_j+0z_j)}$$

The phases are related to the chemical structure

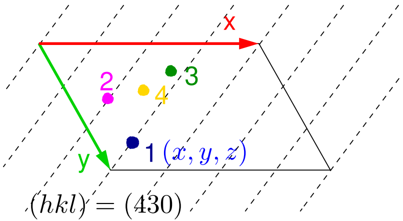


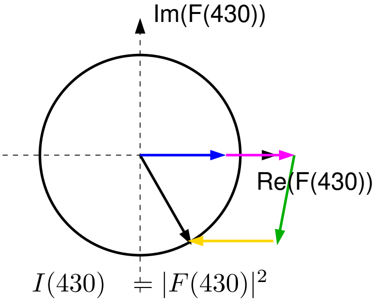


$$
\begin{aligned}
F(430) &= f_1(430)\\
&+ f_2(430)\\
&+ f_3(430)\\
&+ f_4(430)
\end{aligned}
$$

Total $F(430)$ = sum of individual vectors

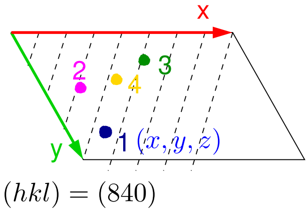
The phases are related to the chemical structure

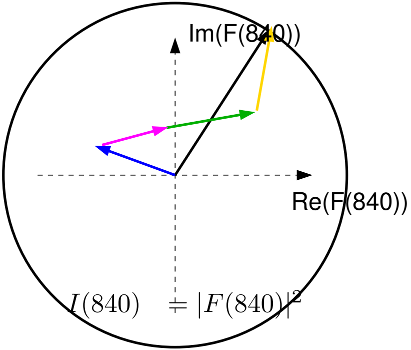




Measurement $I(430) = |F(430)|^2$: loss of phase information

The phases are related to the chemical structure





Same coordinates, *different* contributions per atom to $F(840)$

Phases $\phi(hkl)$

- The phase problem arises, because $F(hkl)$ is a complex number.
- $F(hkl) = \sum_{\text{all atoms}} f_j(\theta) e^{-8\pi^2 U_j(\theta,\lambda)} e^{2\pi i(hx_j+ky_j+lz_j)}$
- The length f_j for each atom is independent of (hkl).
- The phase contribution $e^{2\pi i(hx_j+ky_j+lz_j)}$ varies for each reflection.
- the phases $\phi(hkl)$ is an angle between 0° and 360°
- The total phase $\phi(hkl)$ contains convoluted information from each atom. Each contribution depends on the position of the atom **relative** to the lattice planes for the reflection (hkl), *i.e.*, there are as many phases as there are reflection spots.

References

[1] Zbigniew Dauter. ‘Collection of X-Ray Diffraction Data from Macromolecular Crystals’. In: *Protein Crystallography: Methods and Protocols*. Ed. by Alexander Wlodawer, Zbigniew Dauter and Mariusz Jaskolski. New York, NY: Springer New York, 2017, pp. 165–184. DOI: `10.1007/978-1-4939-7000-1_7`.

[2] C. Giacovazzo, ed. *Fundamentals of Crystallography*. Oxford University Press, 1985.