Chemical Crystallography and Structural Chemistry

(VO 270287)

Lecture 5

23rd April 2020

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

tim.gruene@univie.ac.at

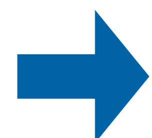
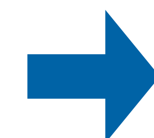
Previous Lecture

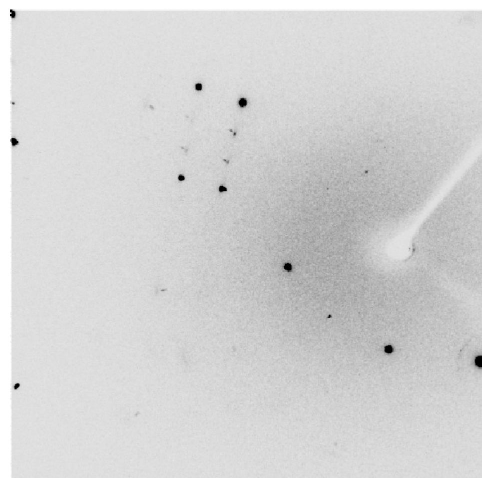
1. Space groups and crystallographic point groups
2. Choice of unit cell
3. 7 Crystal systems and 14 Bravais lattices
4. Symmetry in reciprocal space
5. Space group determination

Today's Lecture

1. Overview: from data to structure
2. Programs for data reduction
3. Scaling
4. Phasing = solving the structure

From Data Collection to Structure

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



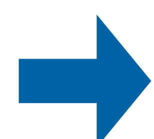
```
0 0 -1      2.7    0.9
0 0  1      4.0    1.0
0 0 -2 1'257.0   35.5
0 0 -2 1'600.0   42.7
```

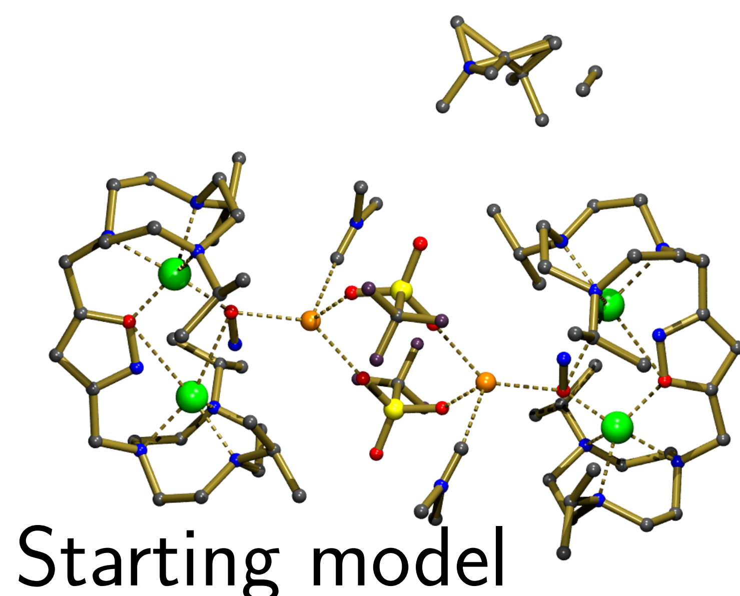
```
0 0 -1      2.8   0.55
0 0  1      3.8   0.63
0 0 -2 1'432.0   95.7
0 0 -2 1'282.0   85.9
```

several GB

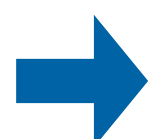
several files, 100's MB

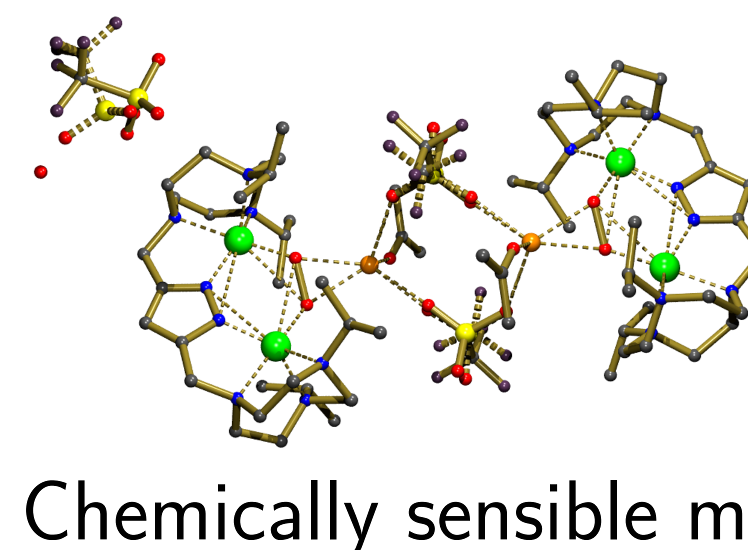
1 "hkl"-file, 50MB

 **Phasing**



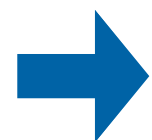
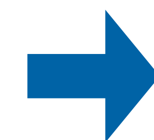
Starting model

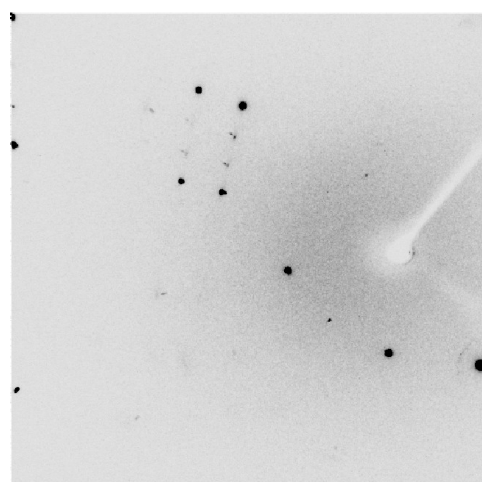
 **Refinement**



Chemically sensible model

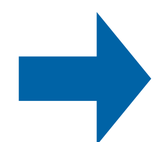
Data Collection

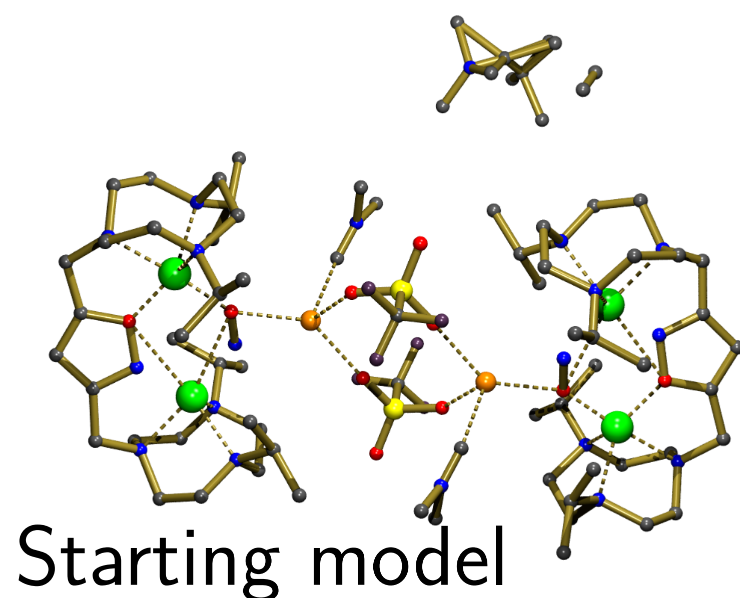
Data collection  **Data integration**  **Data Scaling**



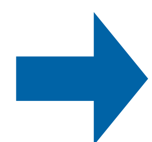
0 0 -1	2.7	0.9		0 0 -1	2.8	0.55
0 0 1	4.0	1.0		0 0 1	3.8	0.63
0 0 -2	1'257.0	35.5		0 0 -2	1'432.0	95.7
0 0 -2	1'600.0	42.7		0 0 -2	1'282.0	85.9

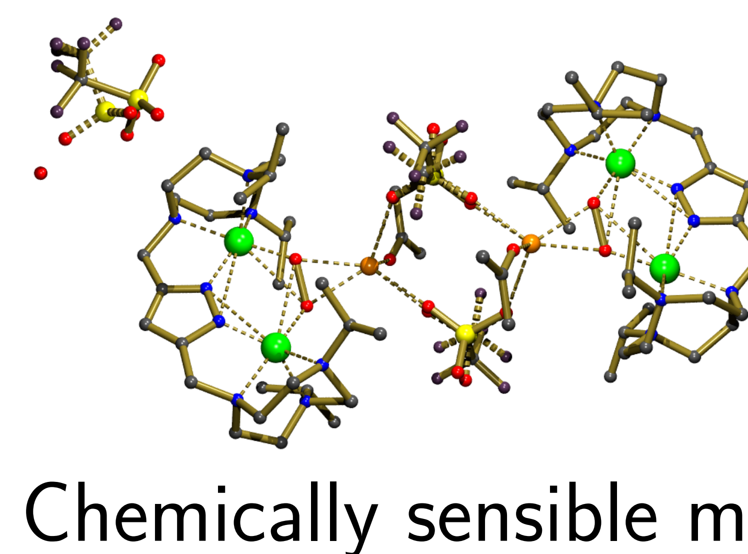
several GB — several files, 100's MB — 1 "hkl"-file, 50MB

 **Phasing**



Starting model

 **Refinement**

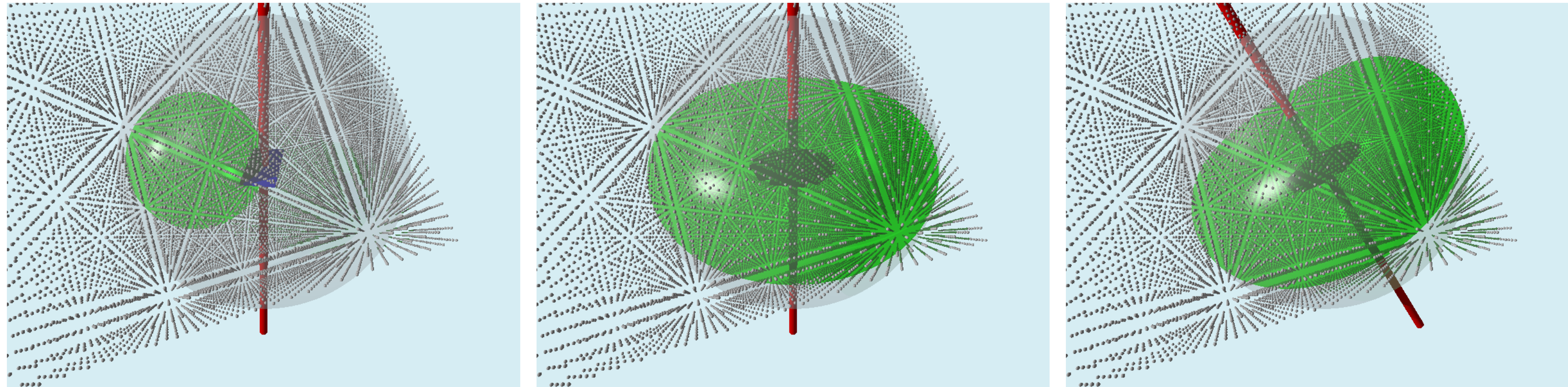


Chemically sensible model

How to collect good data

- Resolution limit $d_{\text{min}} = \lambda/2$ (or worse for poor quality crystals, *e.g.* protein crystals)
- Typical resolution limit $d_{\text{min}} = 0.84$ Å (resolution limit for publishing in Acta Crystallographica C)
- Reflections that can be measured theoretically: all Miller indices (hkl) with $\|h\vec{a^*} + k\vec{b^*} + l\vec{c^*}\| \leq 1/d_{\text{min}}$ ($\vec{a}^*, \vec{b}^*, \vec{c}^*$: reciprocal unit cell vectors)
- Multiple measurements per reflections improve data quality

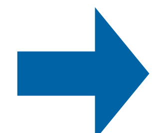
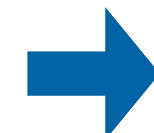
Data completeness and multiplicity

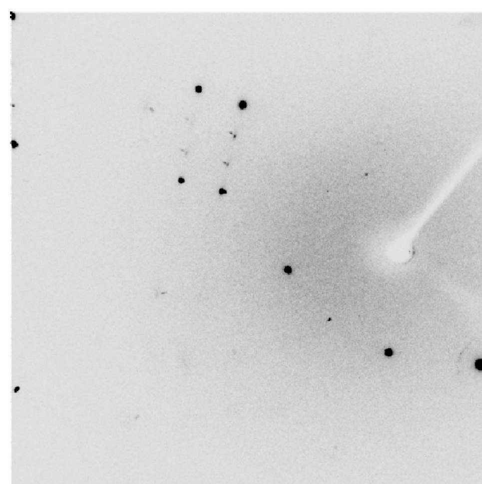


- grey dots: reciprocal lattice
- grey sphere: resolution limit, radius $2/\lambda$
- green sphere: Ewald sphere, radius $1/\lambda$
- red bar: rotation axis of crystal

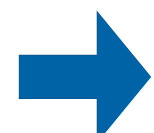
Rotation about a single axis by 360° captures all reflections inside green torus. Capturing all reflections inside grey "resolution shell" requires several orientations.

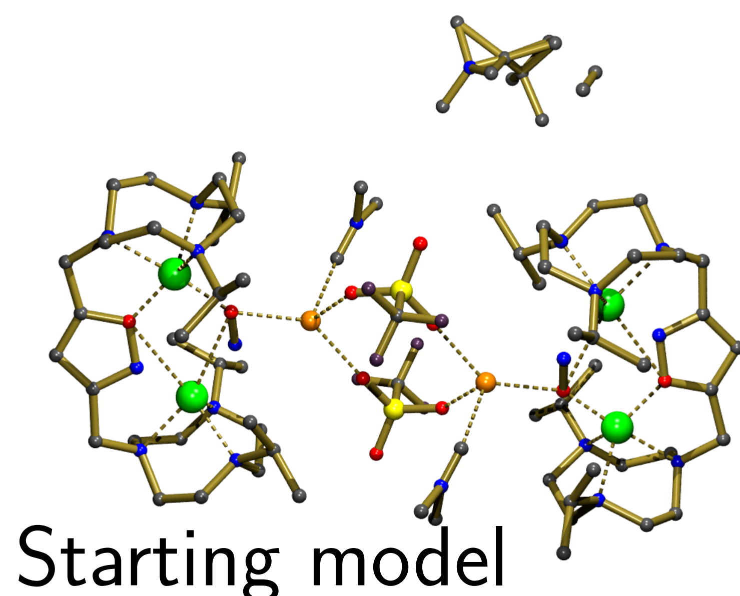
Data Integration

Data collection  Data integration  Data Scaling

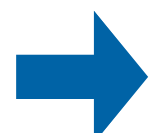


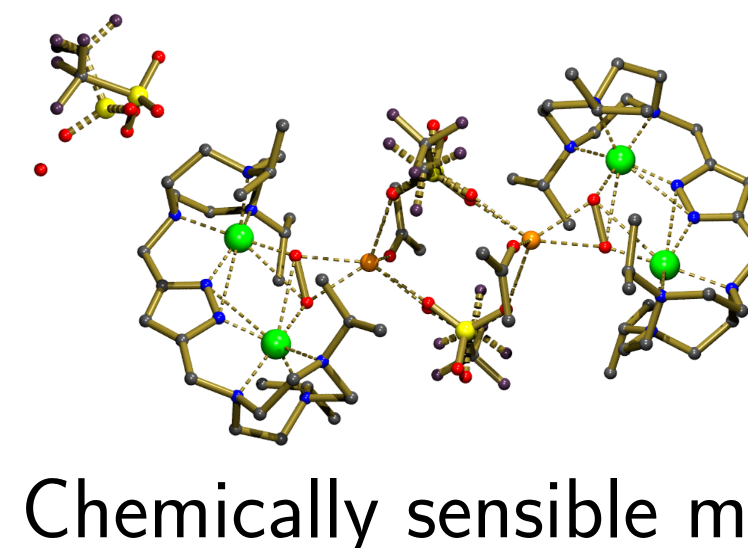
Data collection	Data integration	Data Scaling
	0 0 -1 2.7 0.9	0 0 -1 2.8 0.55
	0 0 1 4.0 1.0	0 0 1 3.8 0.63
	0 0 -2 1'257.0 35.5	0 0 -2 1'432.0 95.7
	0 0 -2 1'600.0 42.7	0 0 -2 1'282.0 85.9
several GB	several files, 100's MB	1 "hkl"-file, 50MB

 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. Very good for data from twinned crystals. Derived from XDS.

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/`)

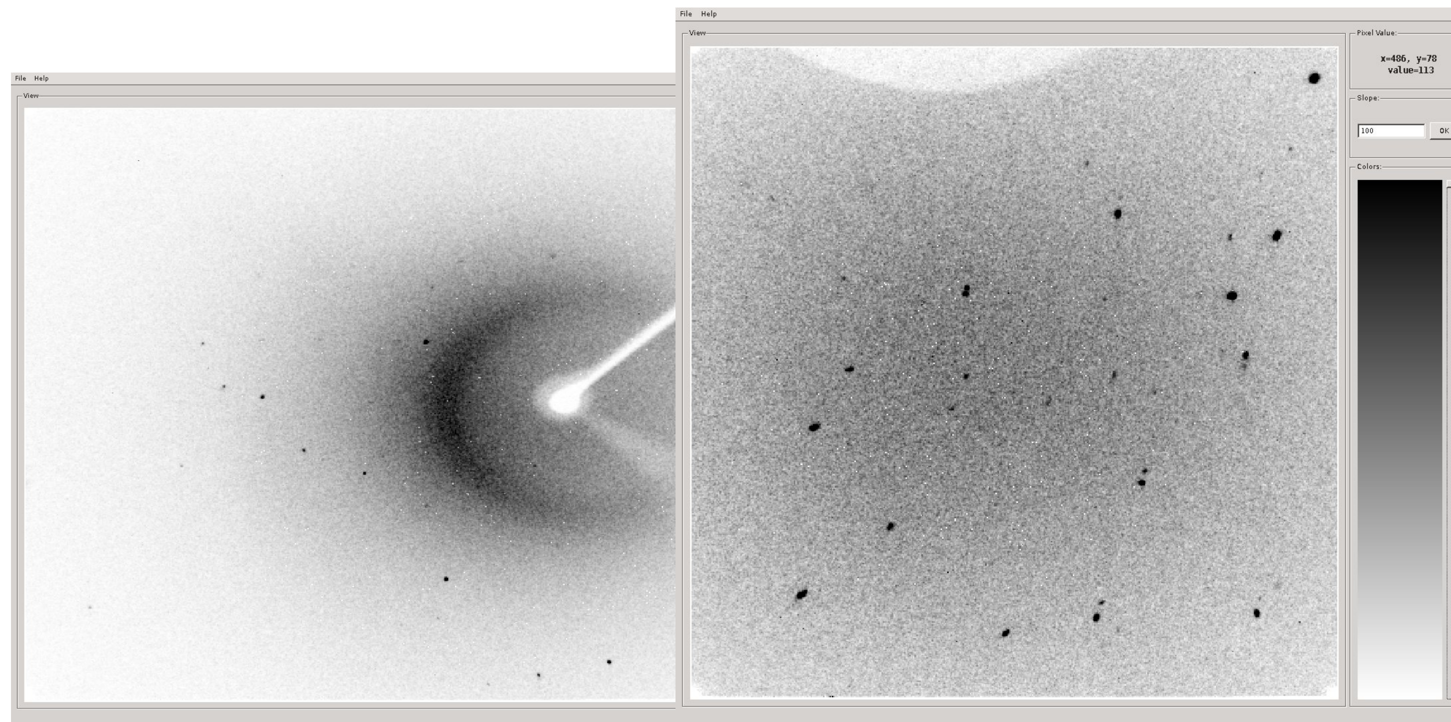
EVAL Suite Free for non-commercial users (`http://www.crystal.chem.uu.nl/distr/eval`). Can integrate *e.g.* incommensurate crystals

Crysalis Pro Licensed by Rigaku (`https://dials.diamond.ac.uk/`)

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/`)

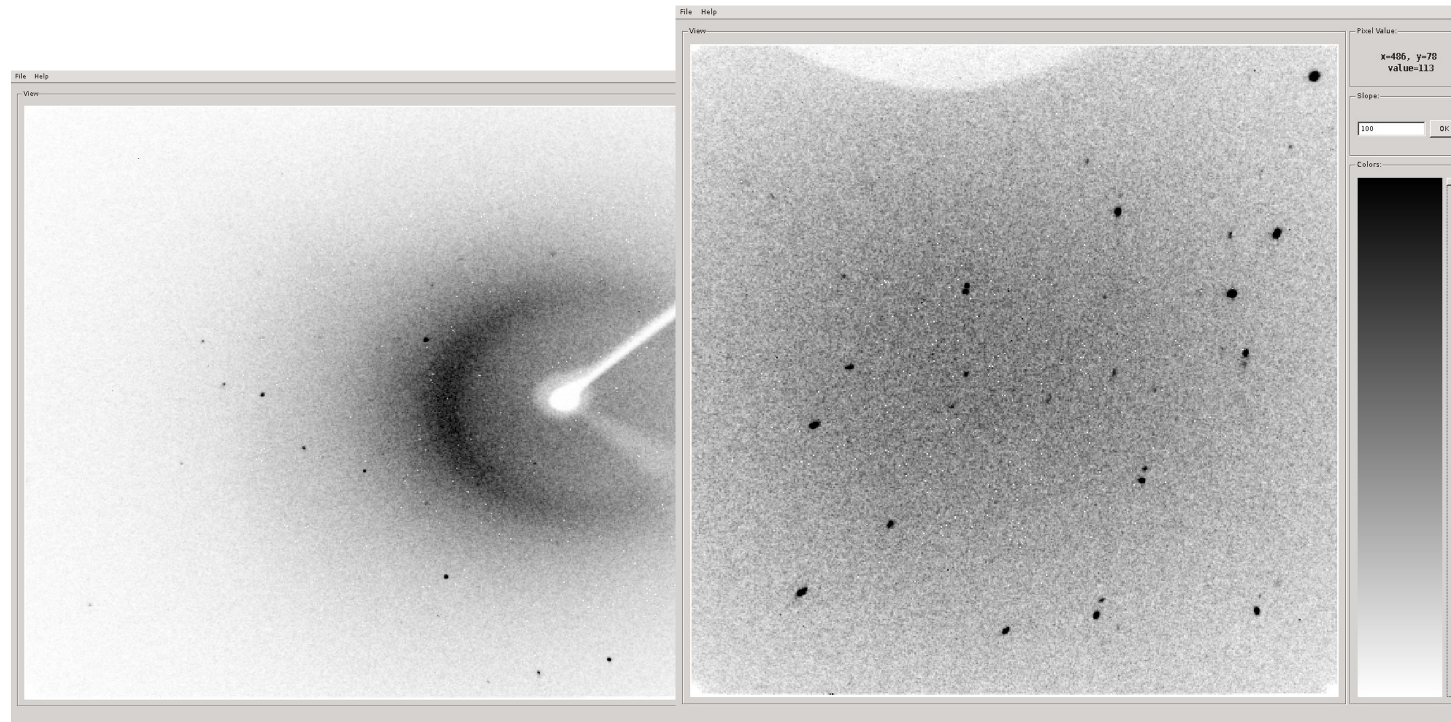
Indexing

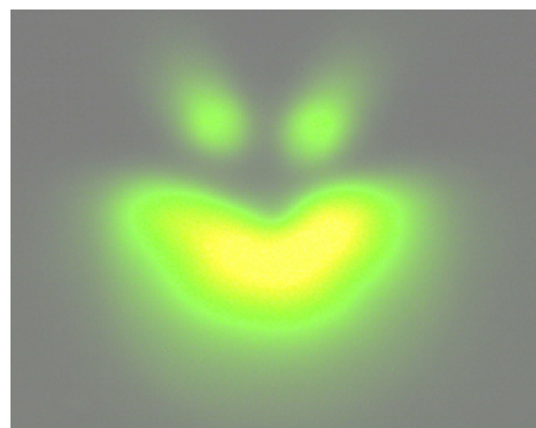


"Ewald sphere backwards:"

- 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

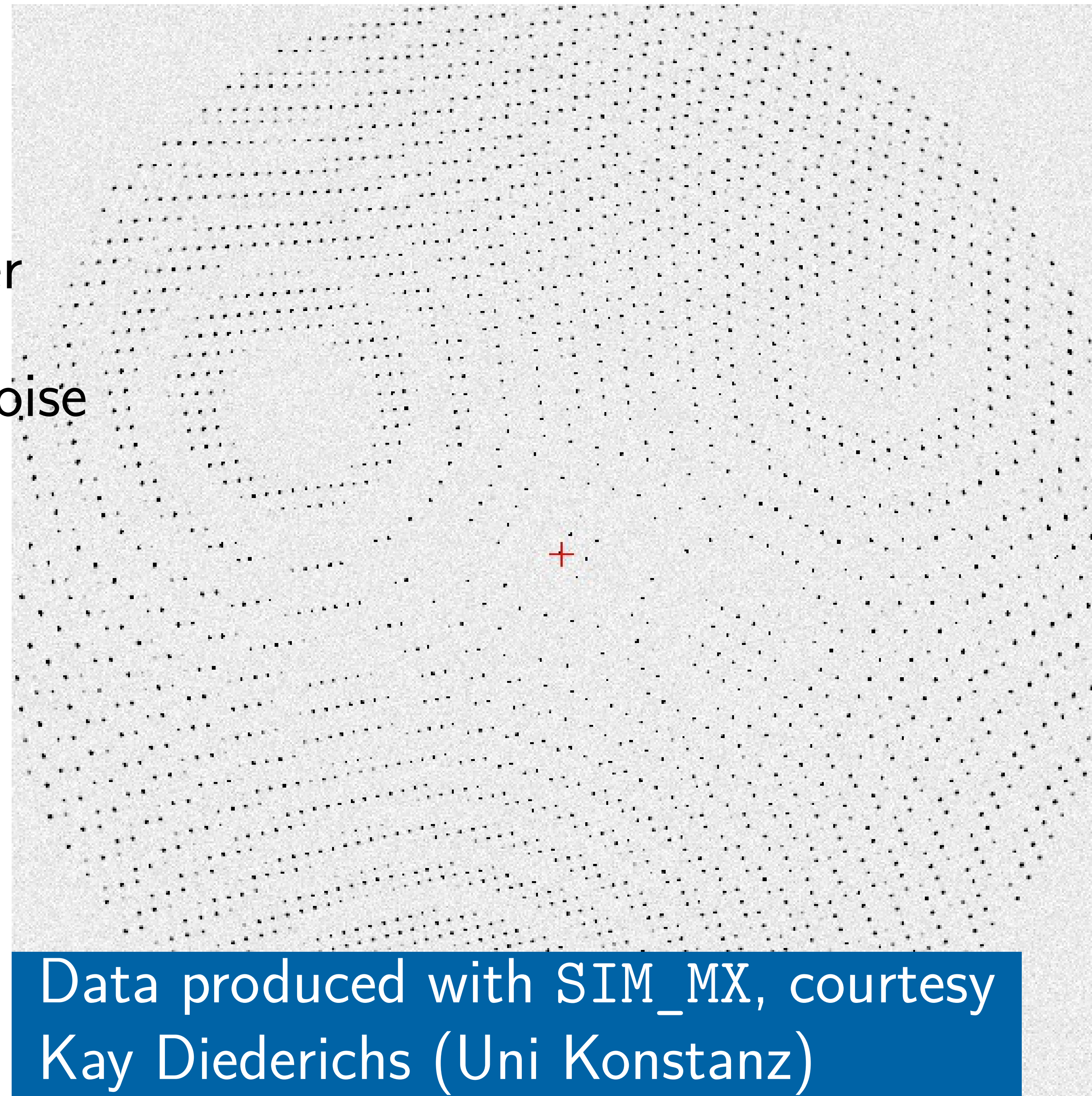
Possible reasons for indexing problems:

- 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*

Spot Intensity

Ideal diffraction image

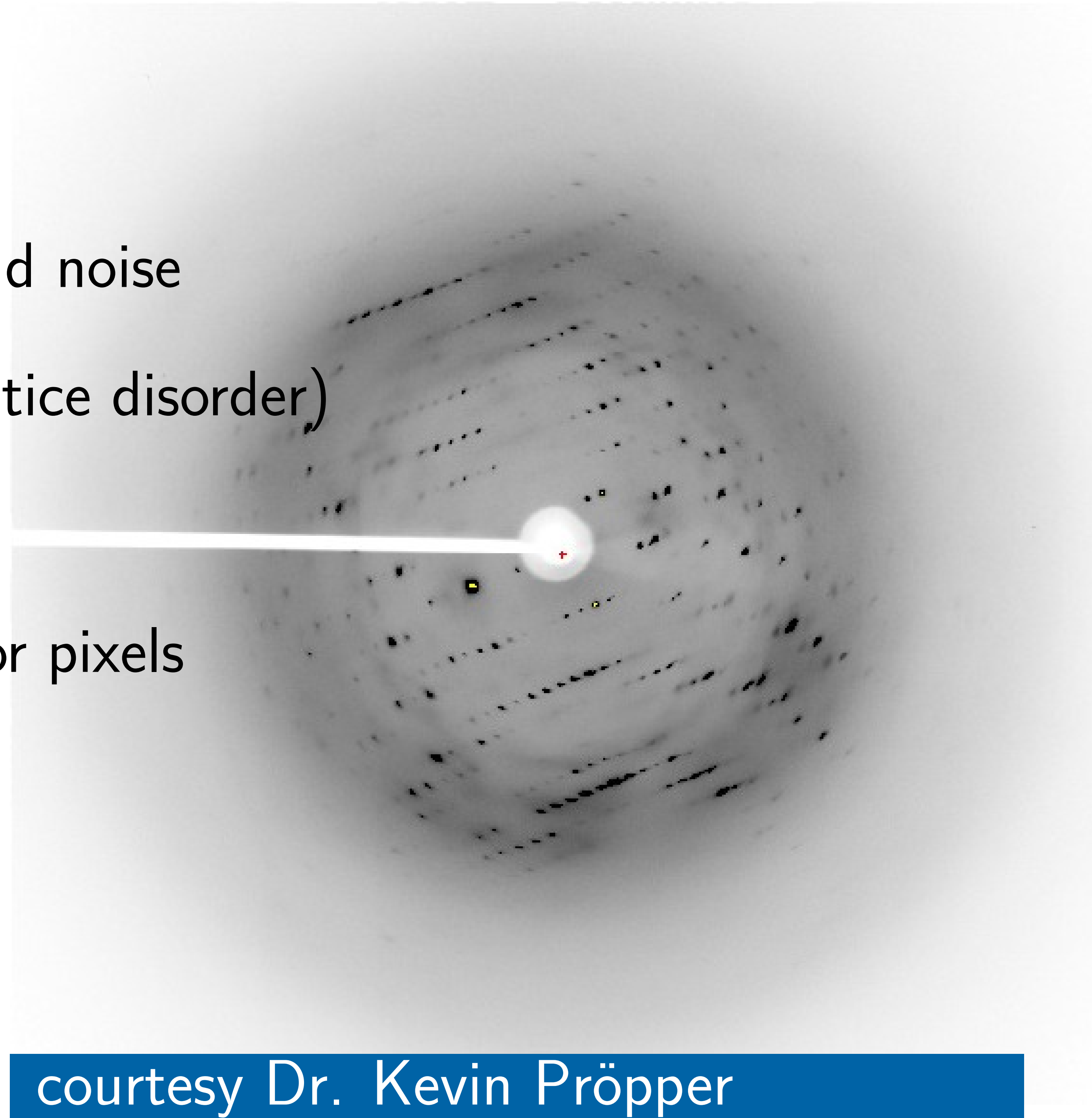
- no lattice disorder
- no background noise
- perfect beam



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

Realistic diffraction image

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



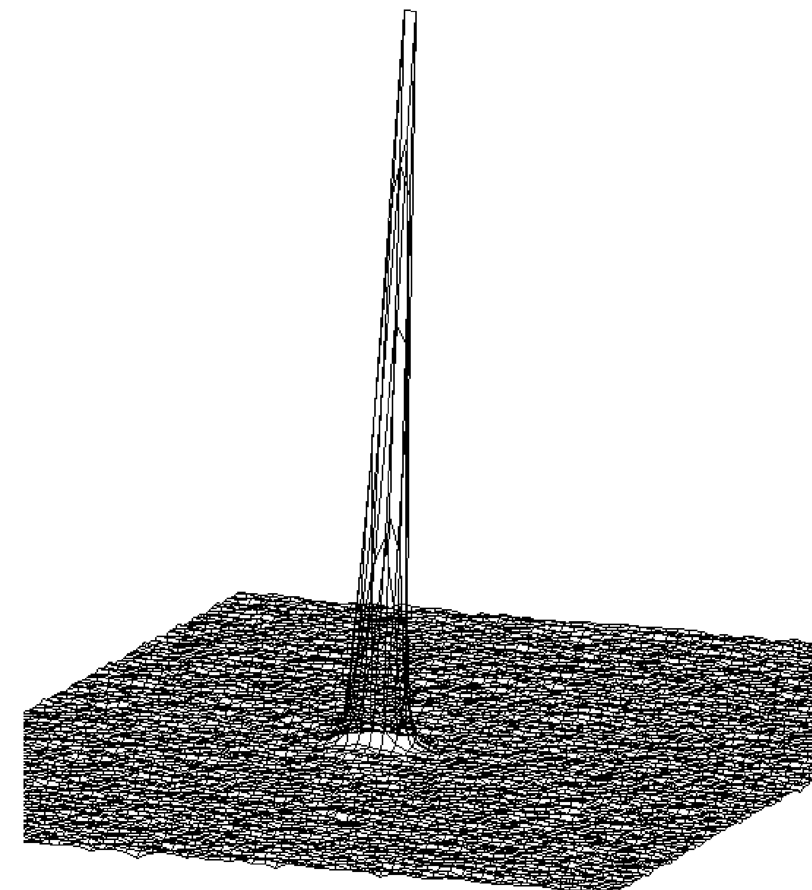
courtesy Dr. Kevin Pröpper

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

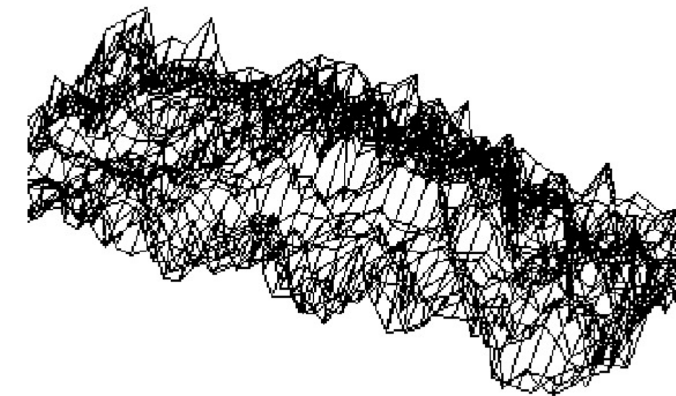
Signal extraction: strong spots

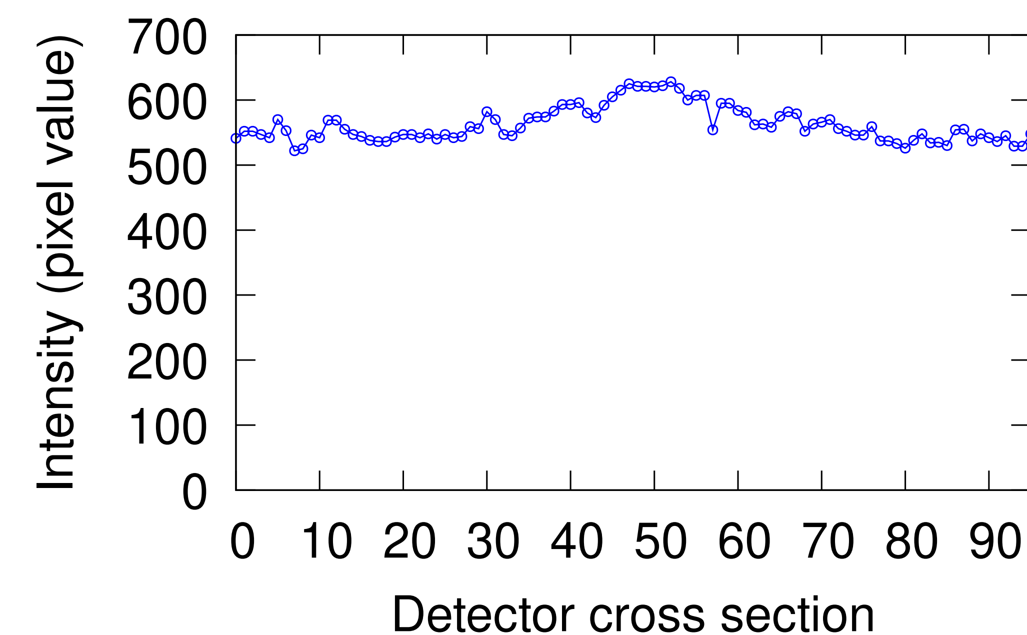
- High intensity ⇒ Small error from noise
- Spot covers large detector area: noise approximated by average
- Good spot separation
- Good spot profile (shape)



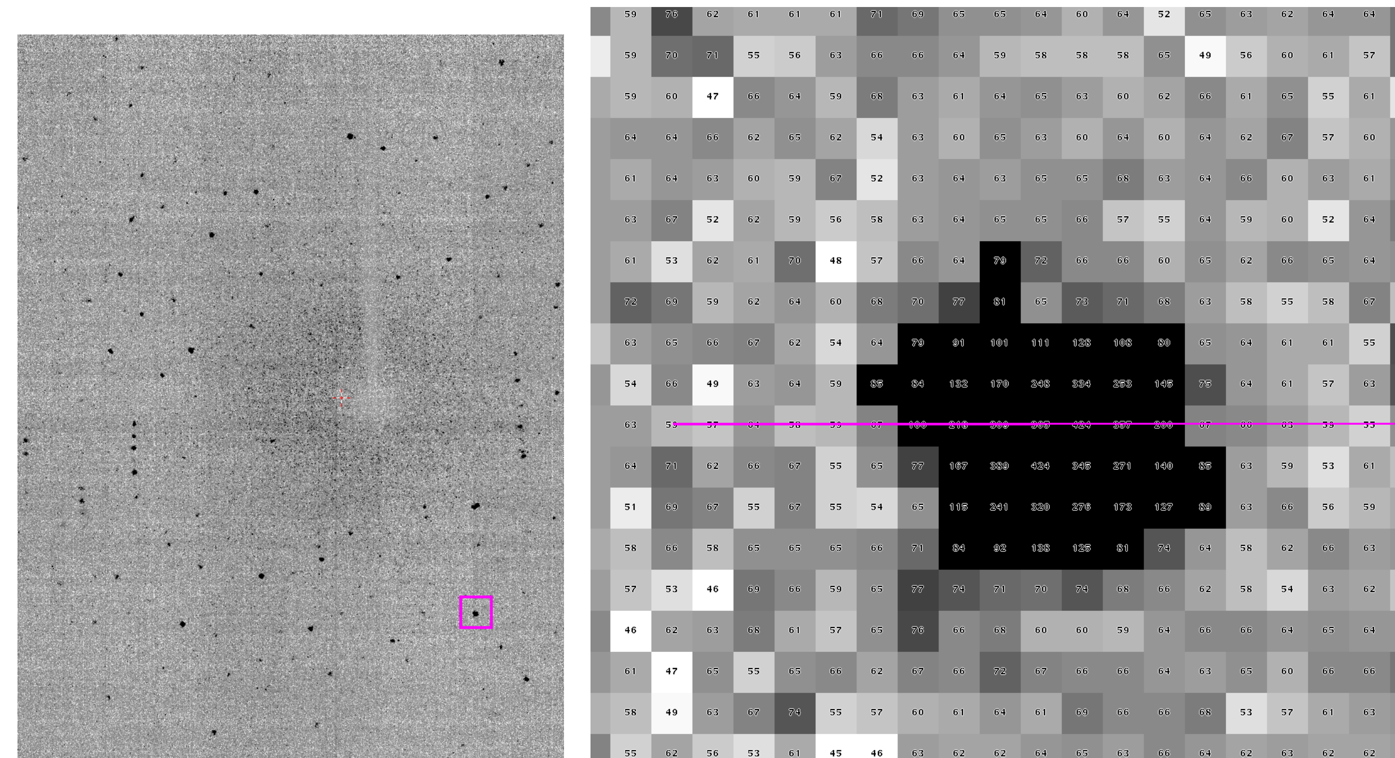
Signal extraction: weak spots

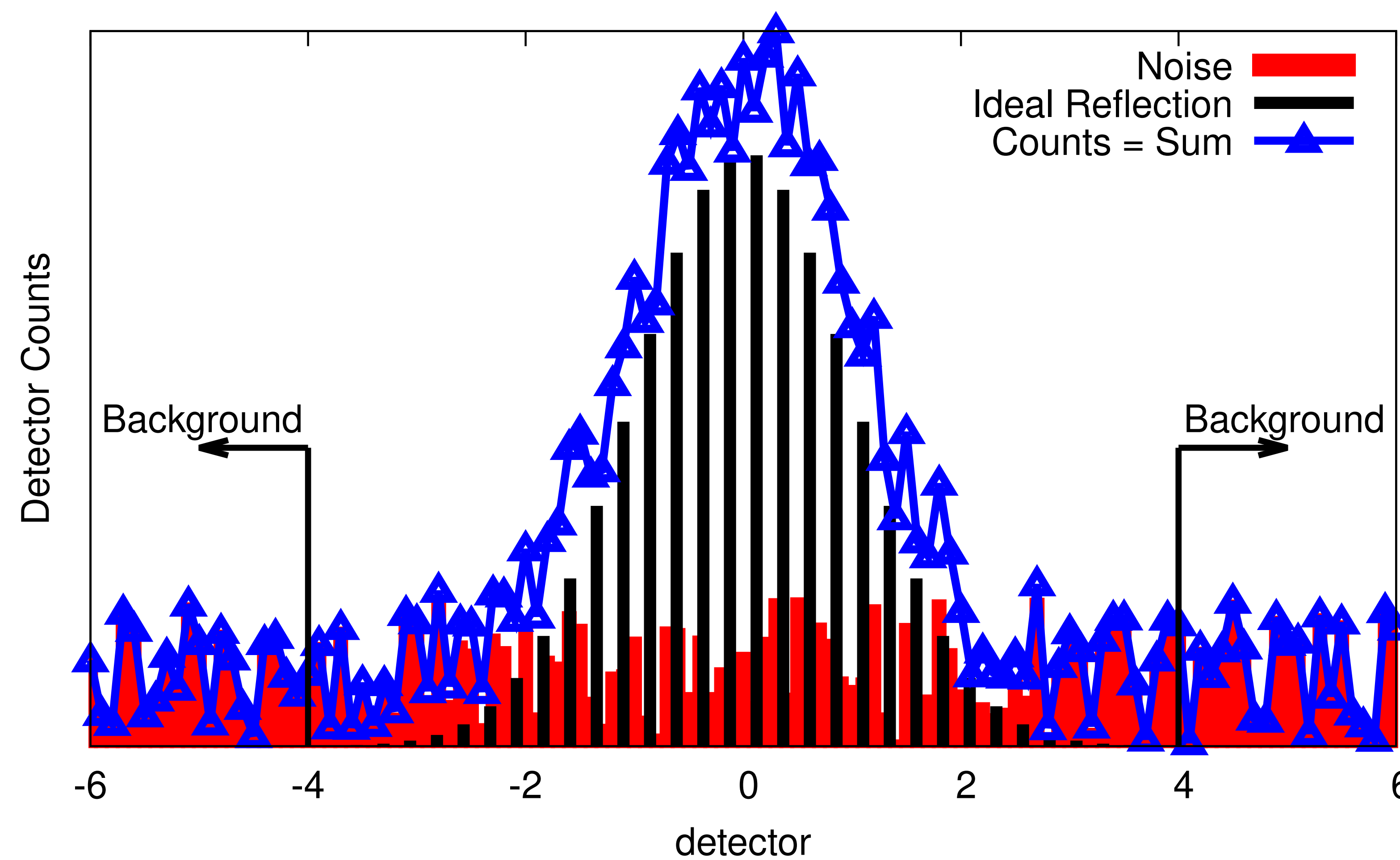
- Low intensity: $\Rightarrow$ high error from noise
- Spot covers small detector area: large effect from noise
- Background difficult to determine





Cross section of a spot on the detector





Reflection profile

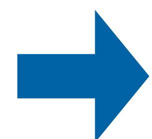
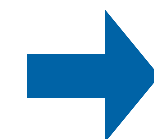
Most integration programs create a set of reflection profiles from strong and reliable reflections, *e.g.* a 3D Gauss function.
The profiles depend on the region on the detector and on the crystal orienation.
Advantages:

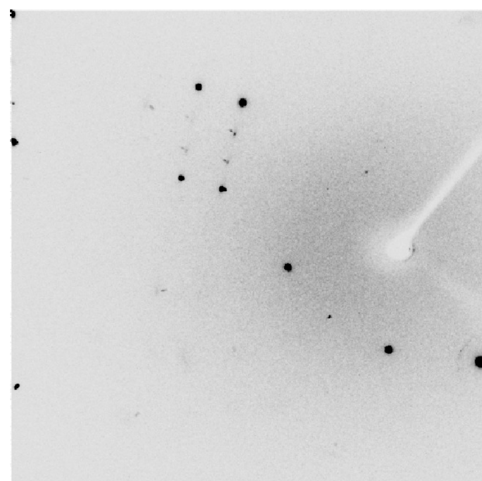
1. Measurement of weak reflections (fitting data for profile)
2. takes non-isotropic crystal shape into accoung
3. takes regions of varying detector sensitivity into account
4. produces a standard deviation of the reflection intensity: (h, k, l, I, σ_I)

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

Scaling

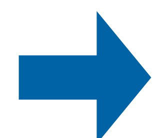
Data collection  Data integration  **Data Scaling**

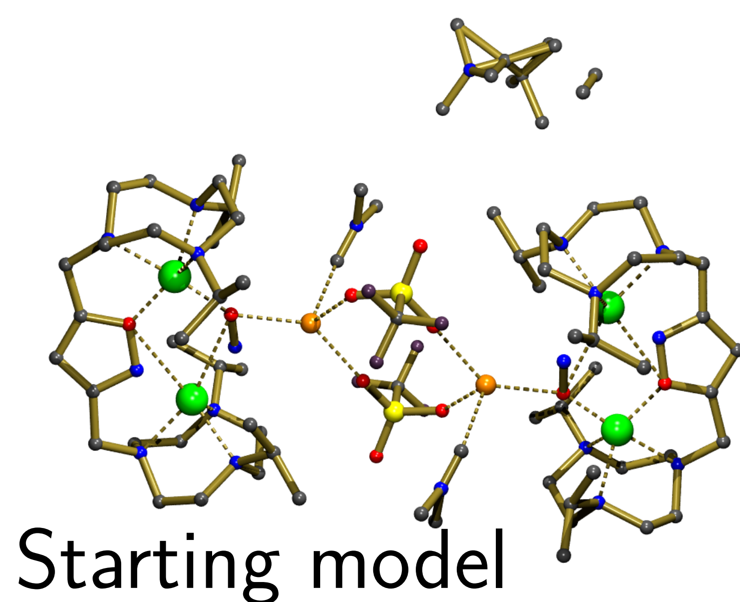


0	0	-1	2.7	0.9
0	0	1	4.0	1.0
0	0	-2	1'257.0	35.5
0	0	-2	1'600.0	42.7

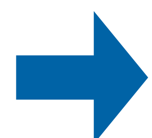
0	0	-1	2.8	0.55
0	0	1	3.8	0.63
0	0	-2	1'432.0	95.7
0	0	-2	1'282.0	85.9

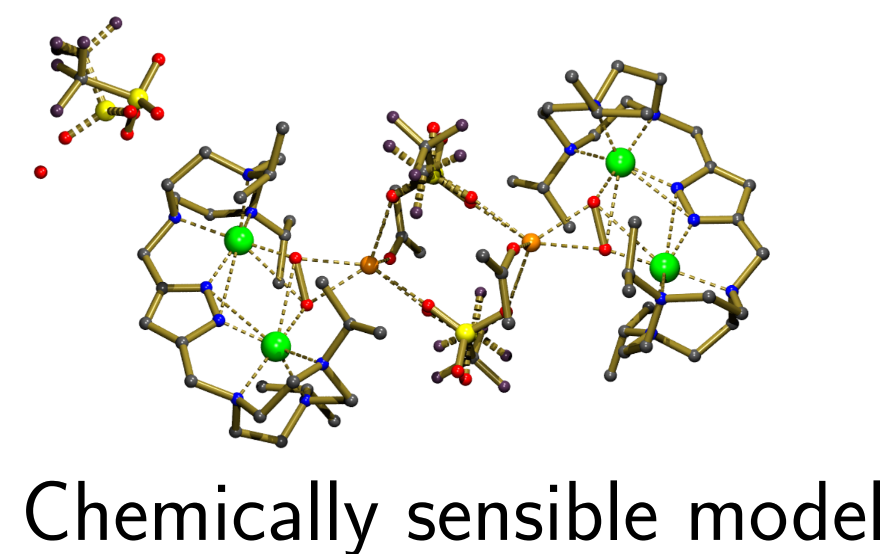
several GB — several files, 100's MB — 1 "hkl"-file, 50MB

 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)

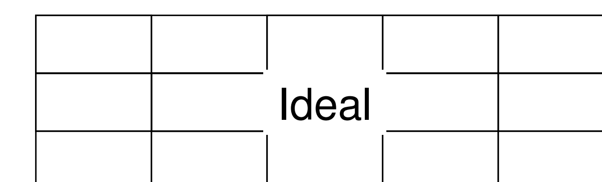
The values should be independent from the experimental setup.
Scaling makes the raw intensities (from data integration) independent from the experimental setup

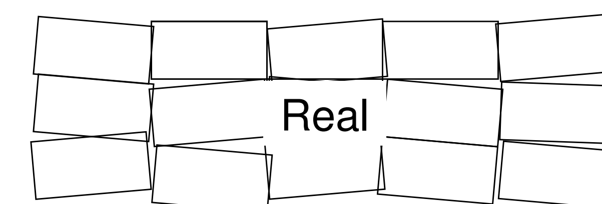
Calculation of Intensities

Under consideration of the experimental, non-idealised setup, intensities are calculated as [1]

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

- 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)

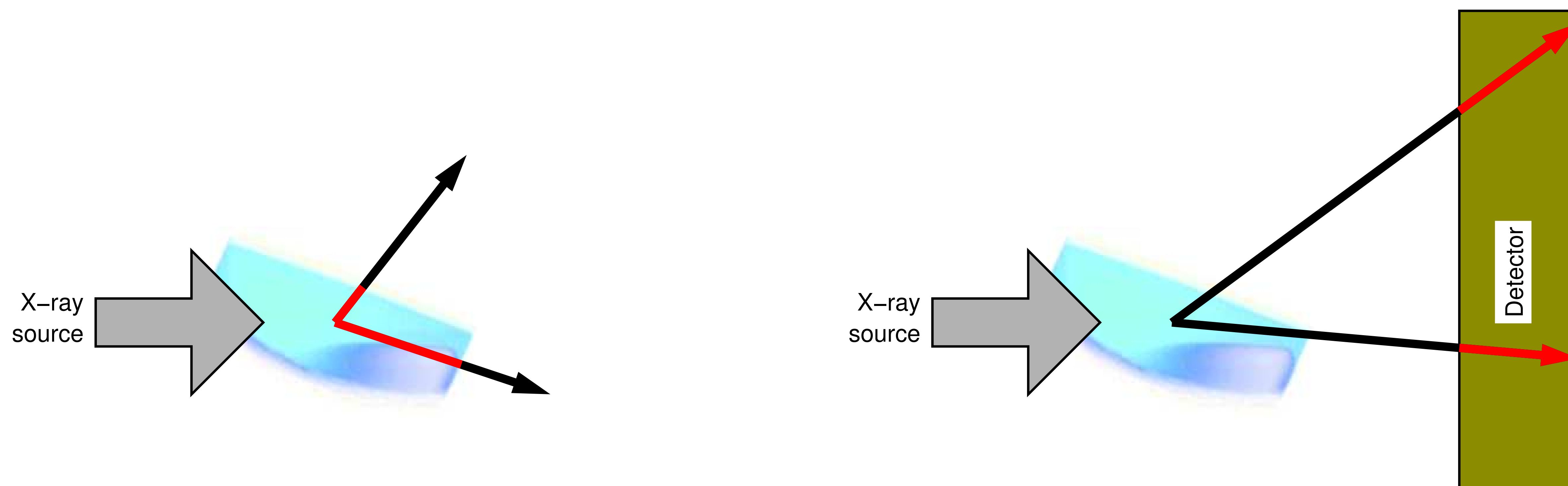




[1] Giacovazzo et el., "*Fundamentals of Crystallography*" (IUCr Texts on Crystallography), 1985, Chapter "Diffraction Intensities"

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

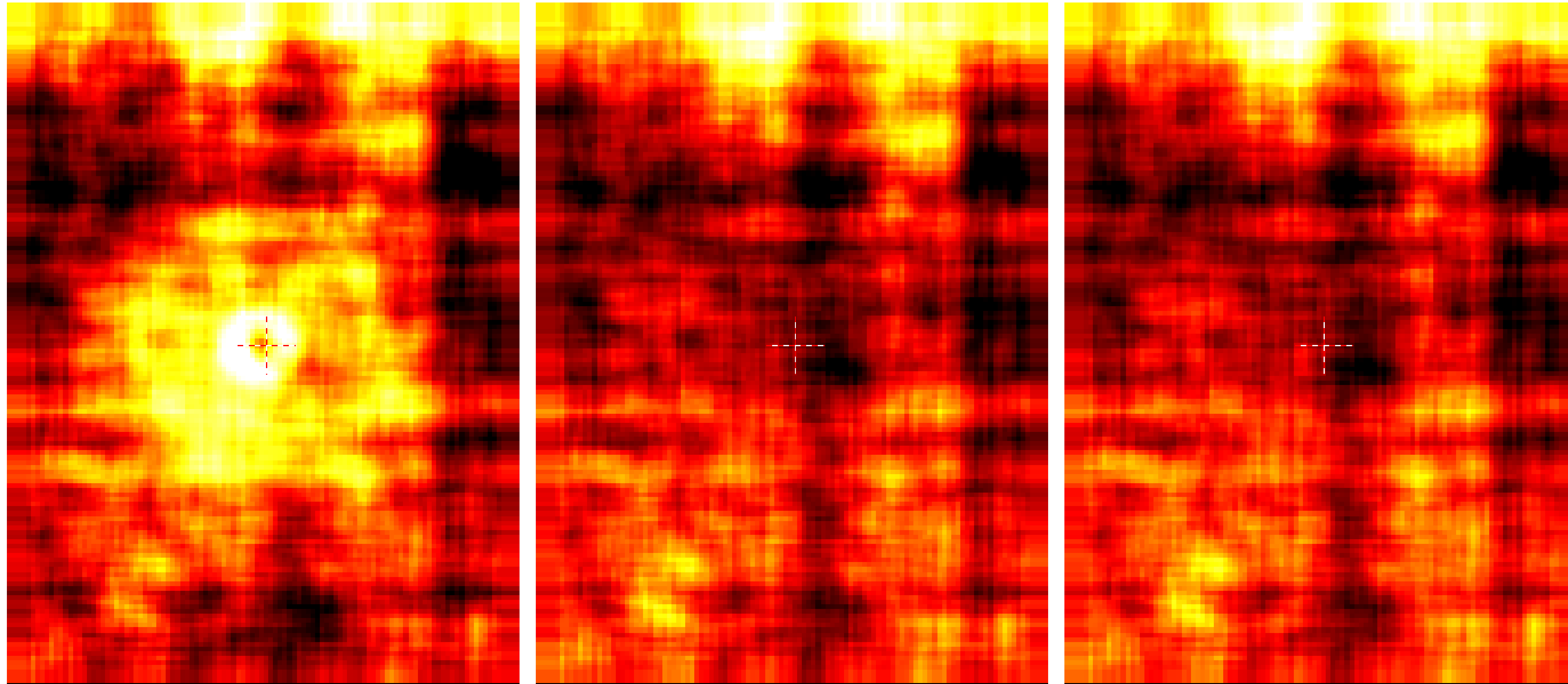
- Some corrections depend on the instrument and can be calibrated (polarisation, angle of incidence, Lorentz factor)
- Some corrections (*e.g.* absorption, extinction) are (also) sample dependent

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

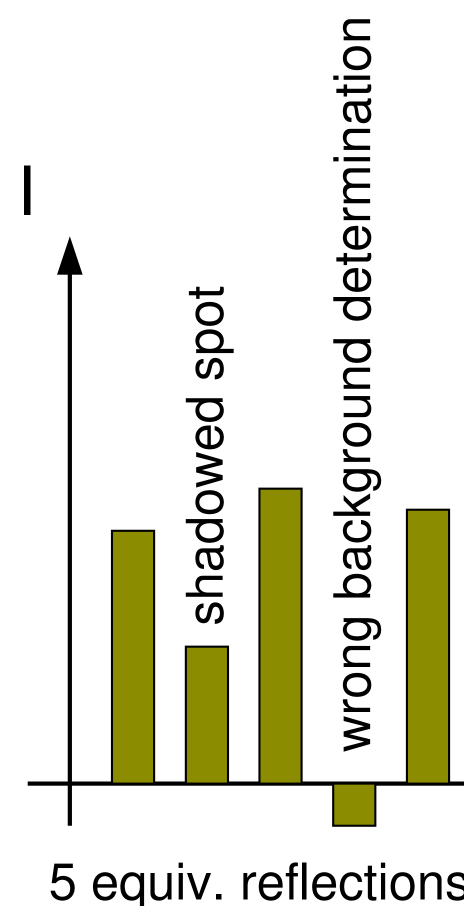
Examples for corrections



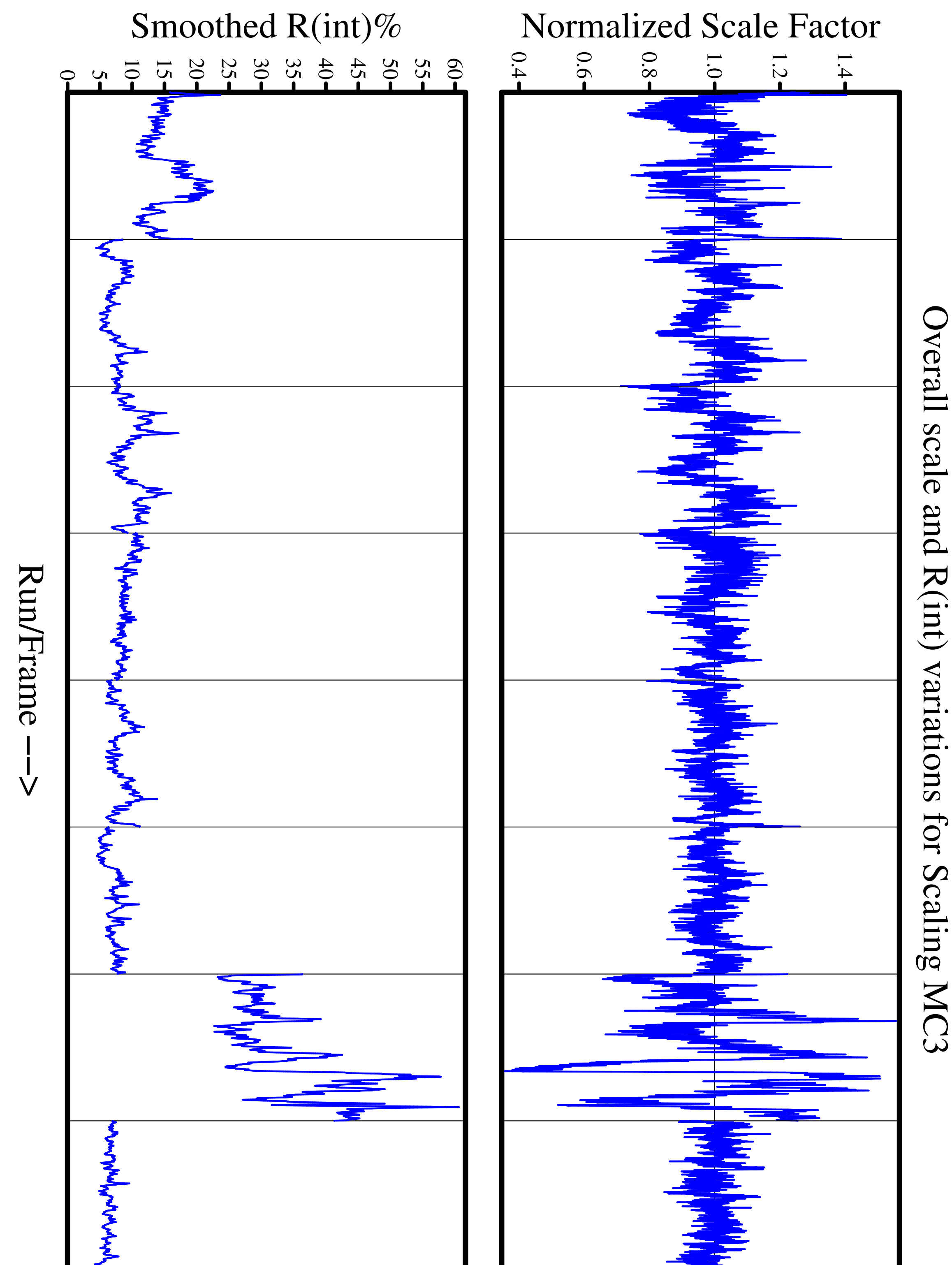
Detector “gain”: dark regions of the detector are more sensitive (factor between 0.97 and 1.3). Direct beam (left): indicates non-linear response at high intensity

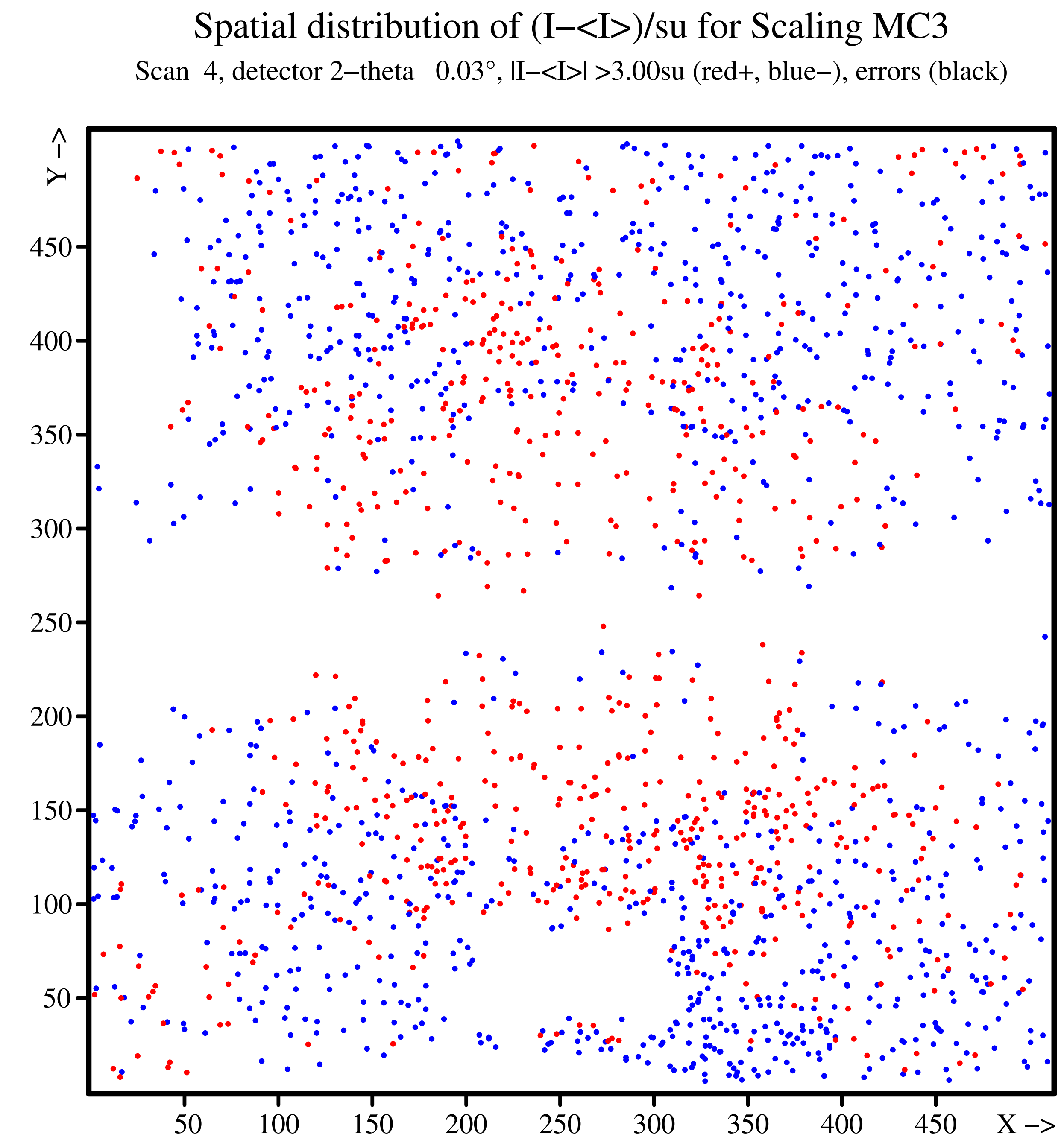
Scaling details

- Statistical average: exaggerates outliers
- Scaling of each group of equivalents independently from other groups: neglects systematic errors.
- 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)





Idealised intensities

Before Scaling

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

After Scaling

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

Before Scaling	After Scaling
0 -1 5 1.379E+03 2.516E+02	0 -1 5 7.014E+01 1.208E+01
0 -1 -5 1.367E+03 2.726E+02	0 -1 -5 6.812E+01 1.274E+01
0 1 5 1.184E+03 2.610E+02	0 1 5 5.987E+01 1.231E+01
0 1 -5 1.347E+03 2.674E+02	0 1 -5 6.753E+01 1.258E+01
0 -1 6 1.090E+04 -1.229E+03	outlier removed
0 -1 -6 4.677E+03 5.733E+02	0 -1 -6 2.365E+02 2.856E+01
0 1 6 4.286E+03 5.488E+02	0 1 6 2.145E+02 2.689E+01
0 1 -6 9.065E+03 -1.034E+03	outlier removed
0 -1 7 0.204E+02 0.571E+01	0 -1 7 1.404E+02 2.271E+01

Detour before Phasing: The Structure factor $F(hkl)$

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 $F(hkl)$

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

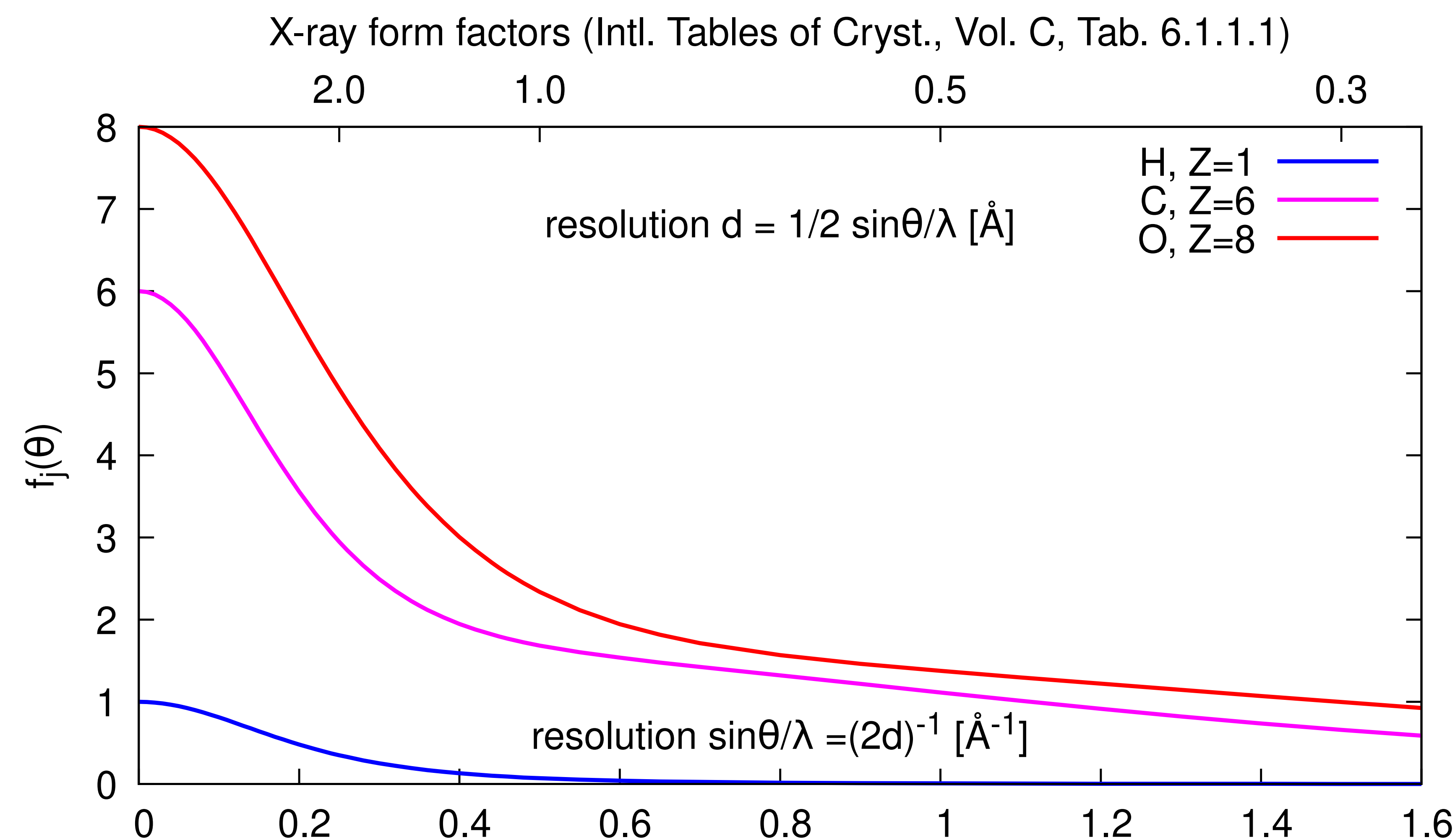
f_j atomic *form* factor. Dependent on atom element, decreases with decreasins scattering angle $\theta = \theta(hkl)$

$U_j(\theta, \lambda)$ 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 $f_j(\theta)$

The 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

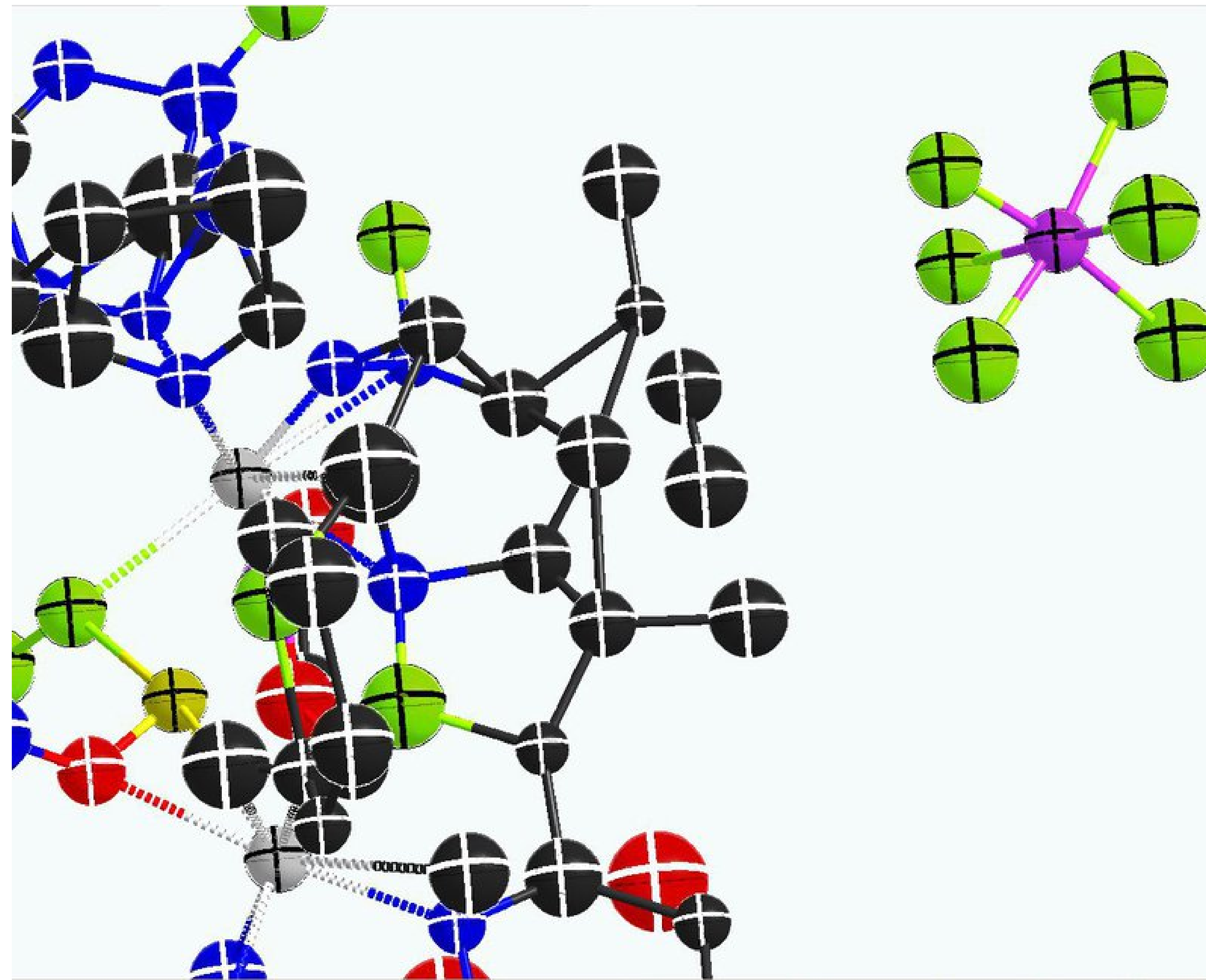
The atomic displacement parameter $U_j(\theta, \lambda)$

- 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

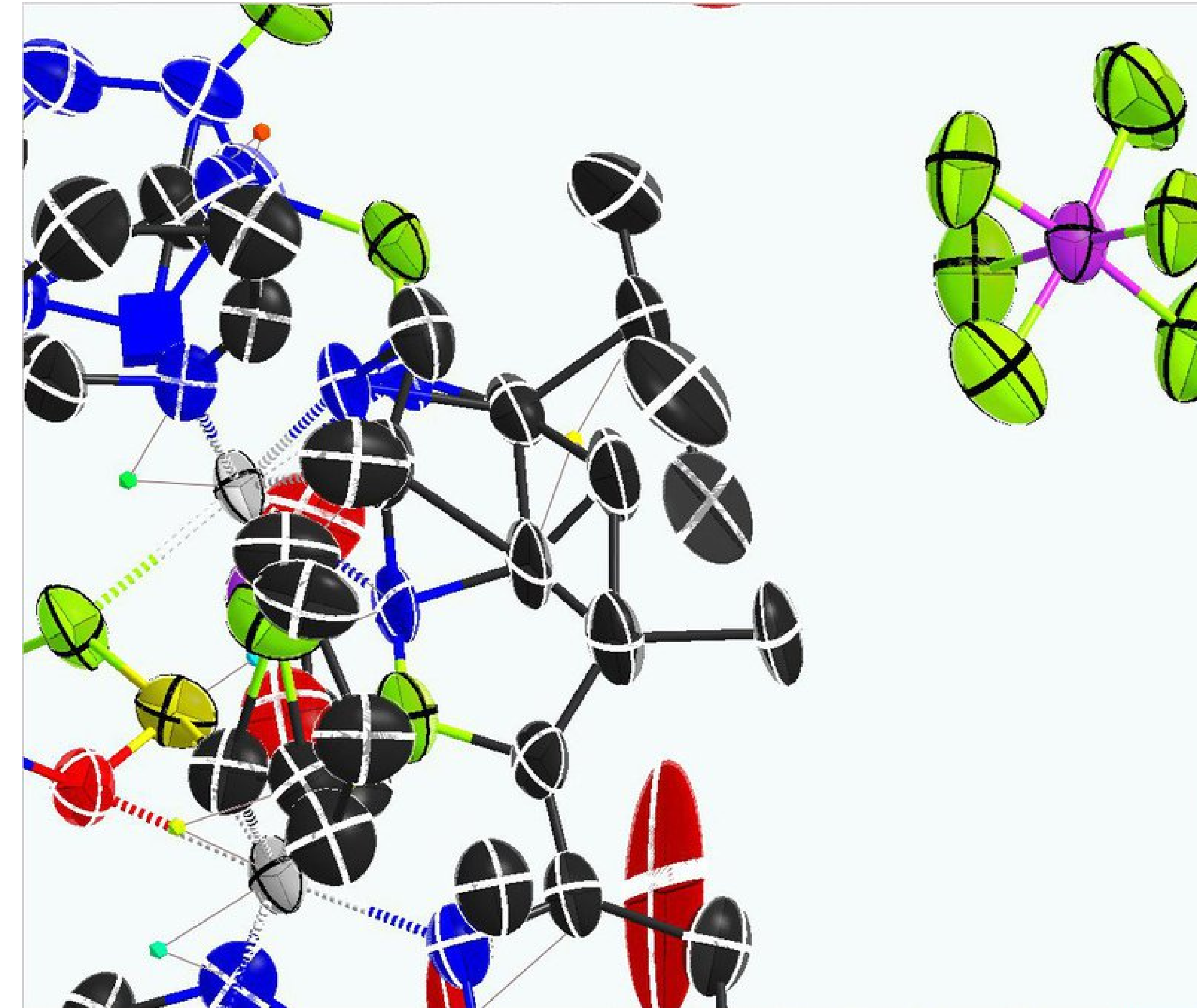
Isotropic	**Anisotropic**
$U_j(\theta, \lambda) = 4U_{\text{iso}}\frac{\sin^2\theta}{\lambda^2}$	$U_j(\theta, \lambda) = (hkl)\begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{12} & U_{22} & U_{23} \\ U_{13} & U_{23} & U_{33} \end{pmatrix}\begin{pmatrix} a^* \\ b^* \\ c^* \end{pmatrix}$
1 parameter per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

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

Example images for ADP

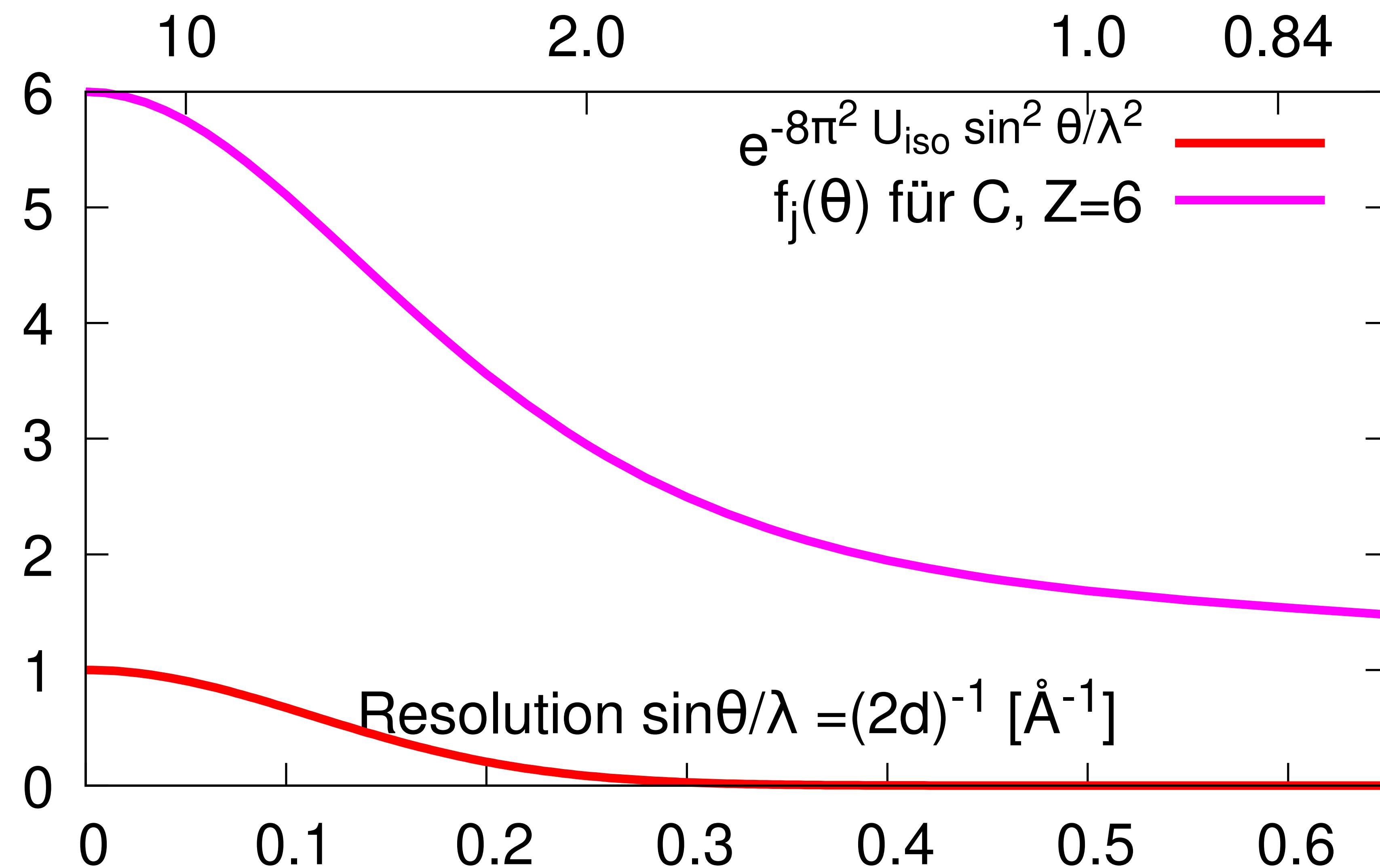


Refinement with isotropic ADPs



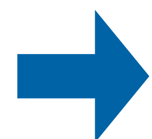
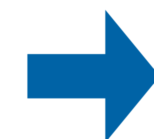
Refinement with anisotropic ADPs

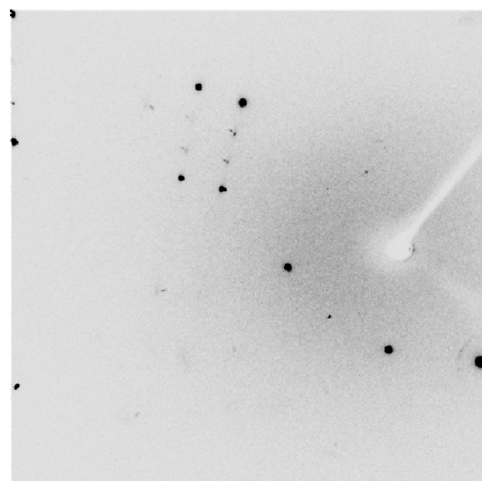
The ADP U: a fudge factor



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

Solving the Structure: Phasing

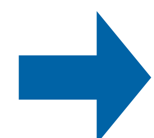
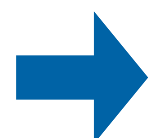
Data collection  Data integration  Data Scaling

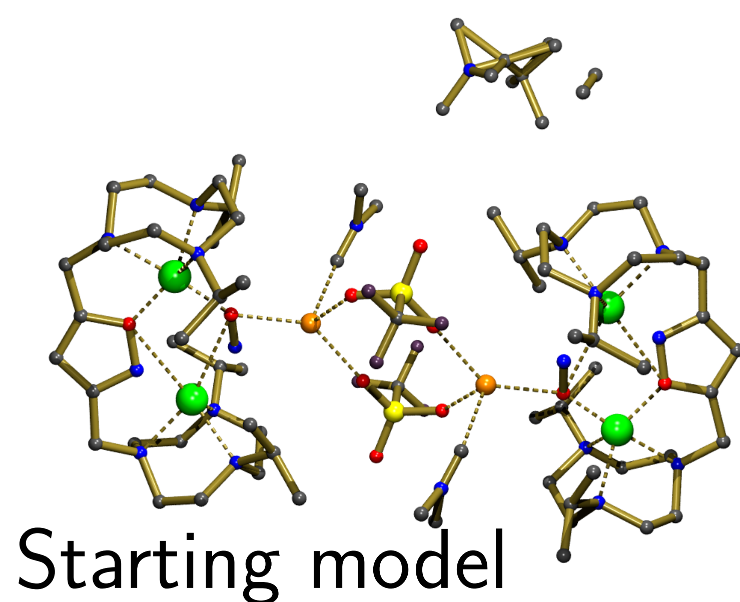


```
0 0 -1       2.7    0.9
0 0  1       4.0    1.0
0 0 -2 1'257.0   35.5
0 0 -2 1'600.0   42.7
```

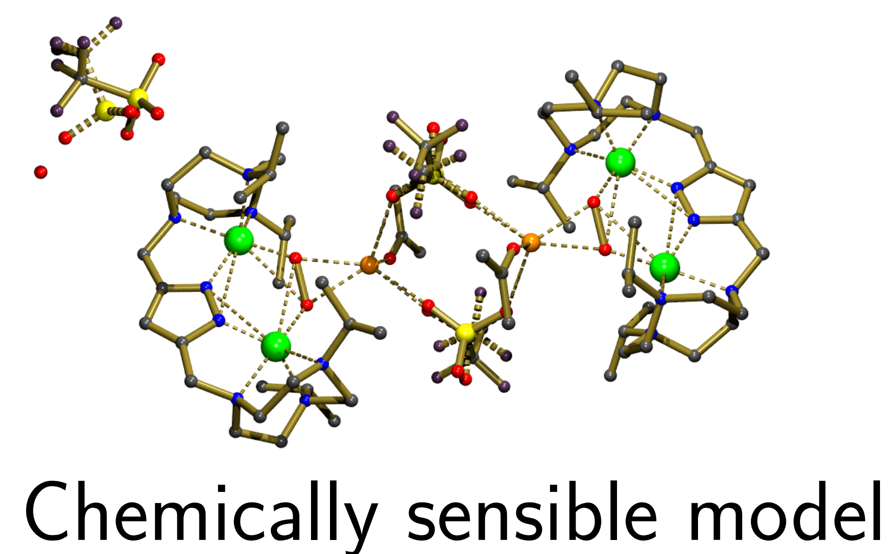
```
0 0 -1       2.8   0.55
0 0  1       3.8   0.63
0 0 -2 1'432.0   95.7
0 0 -2 1'282.0   85.9
```

several GB — several files, 100's MB — 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. 30) and Eq. 4 (page 34):

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

- **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

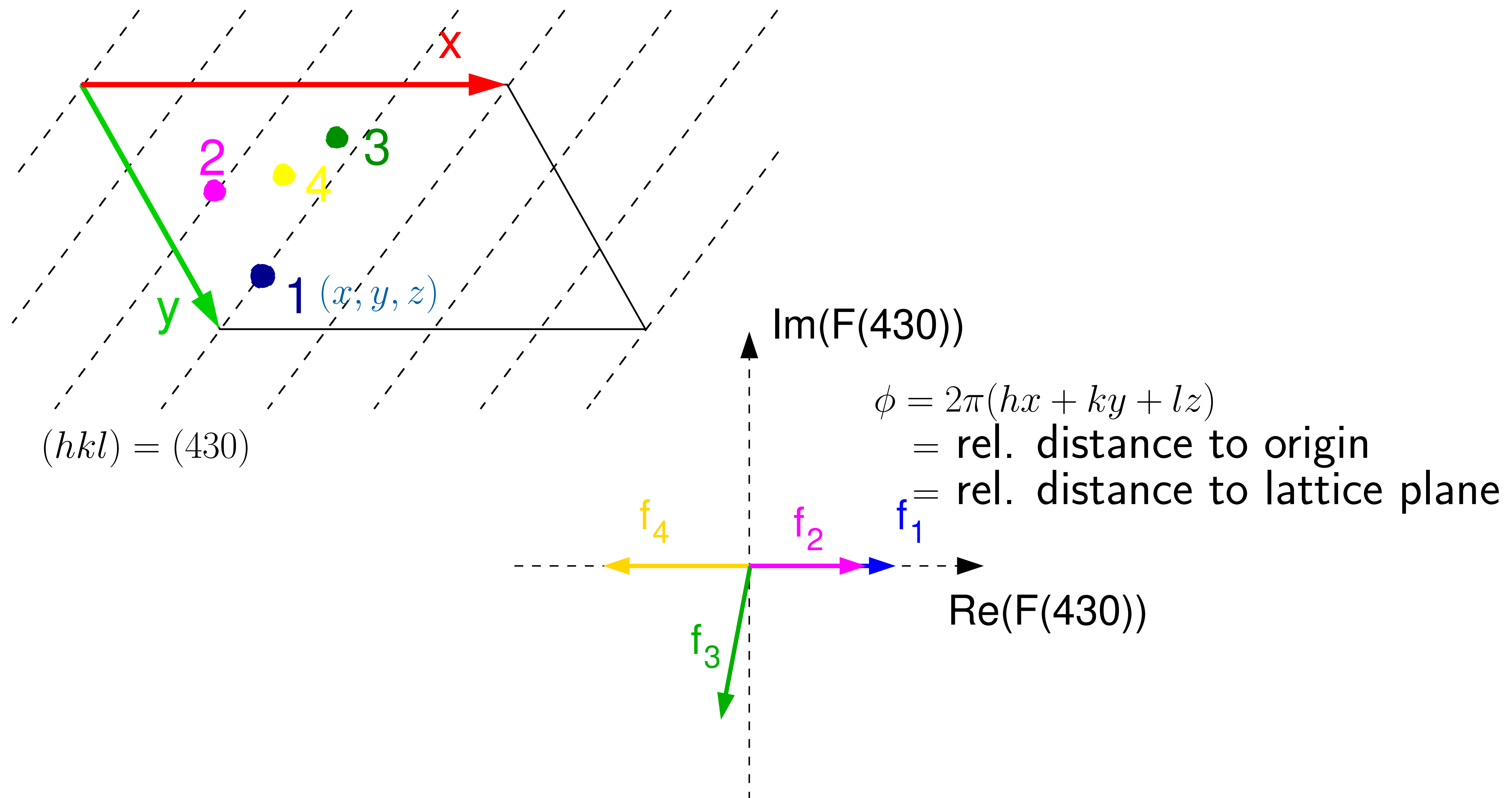
an amplitude $|F(hkl)| = \sqrt{I(hkl)/c}$

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

$F(hkl) = \sqrt{I(hkl)/c} \times e^{-i\phi(hkl)}$

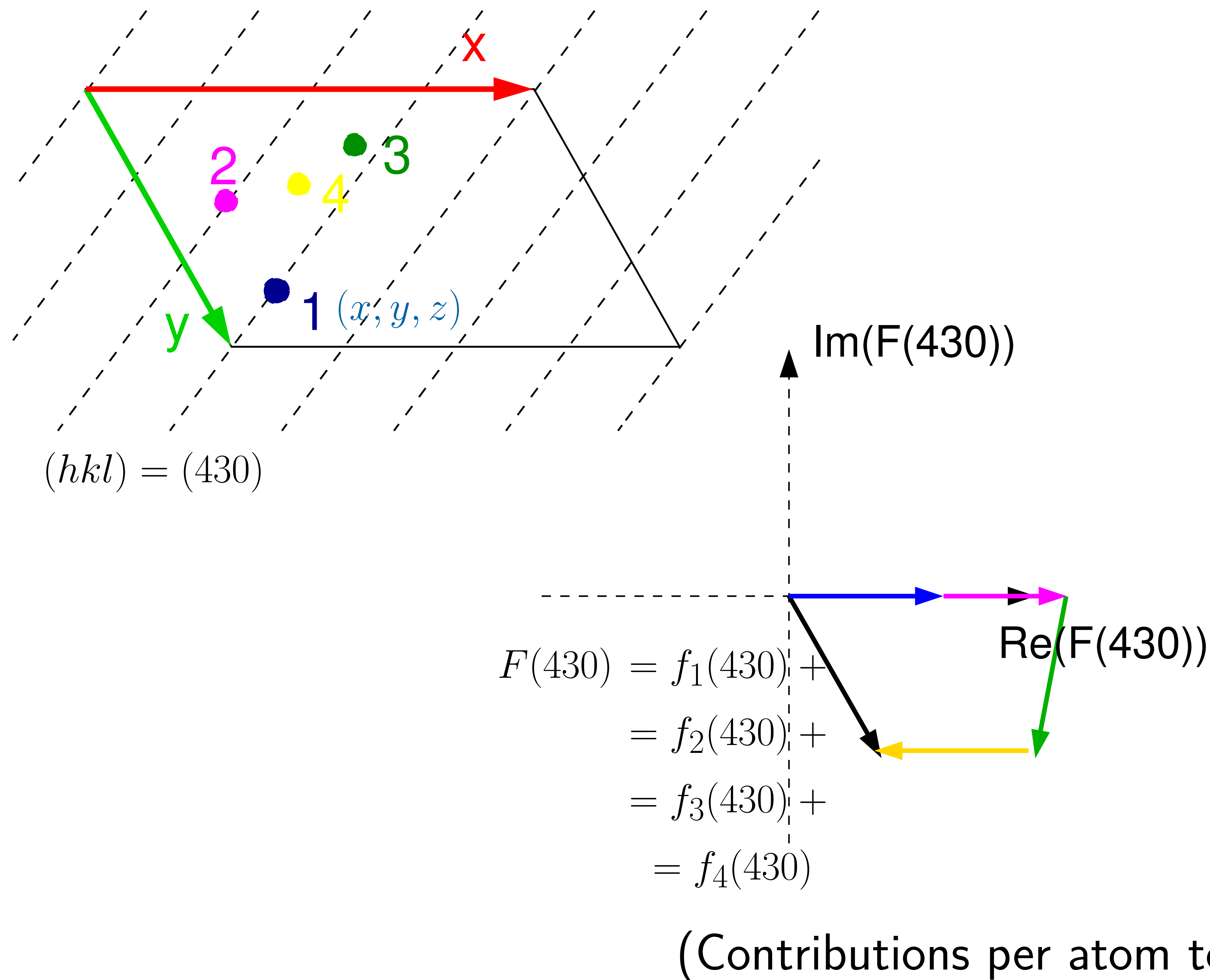
We can measure the amplitude, but we cannot measure the phase. This is known as the **phase problem of crystallography**.

A different view of the phase



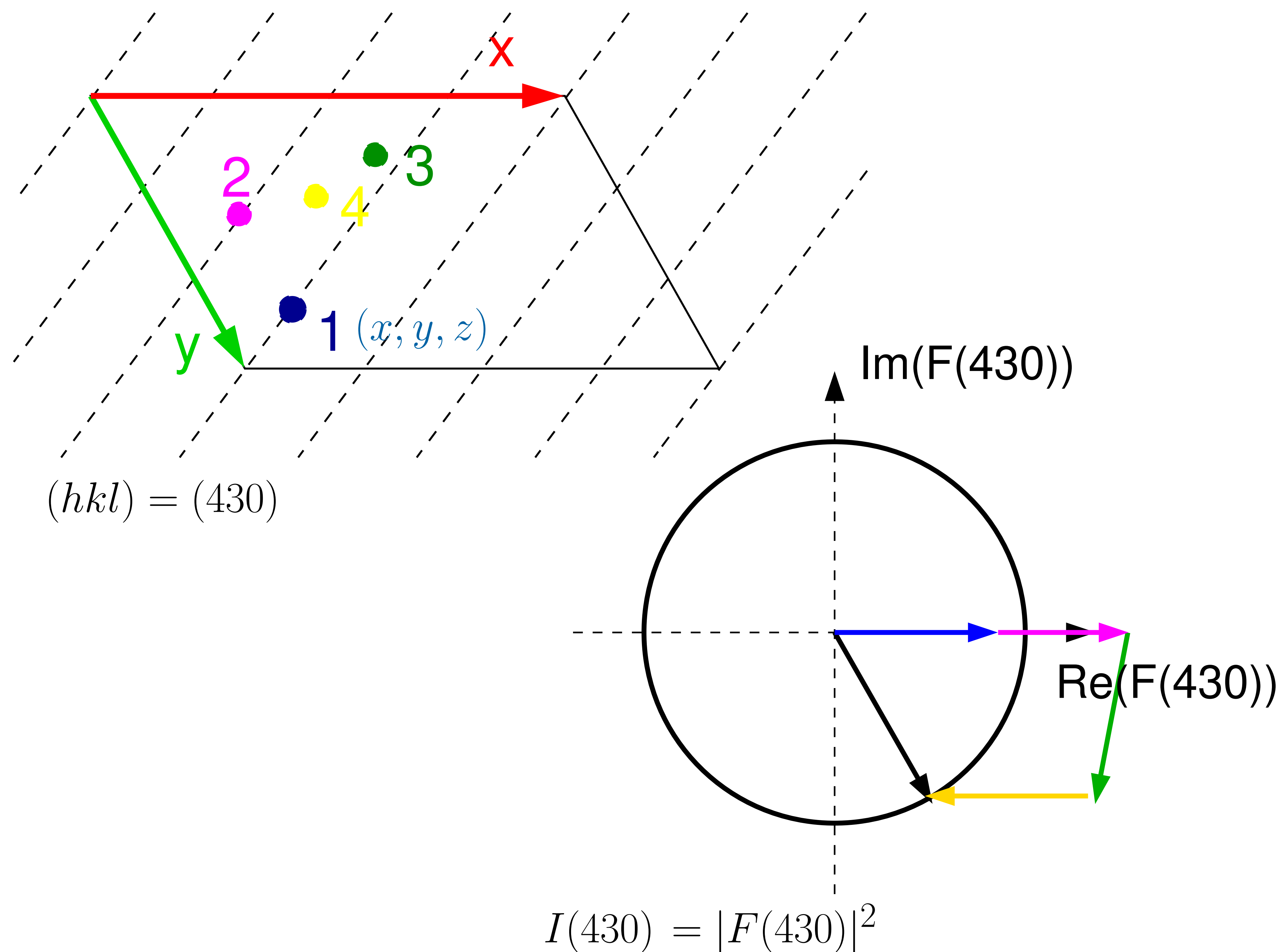
(Contributions per atom to $F(430)$)

A different view of the phase



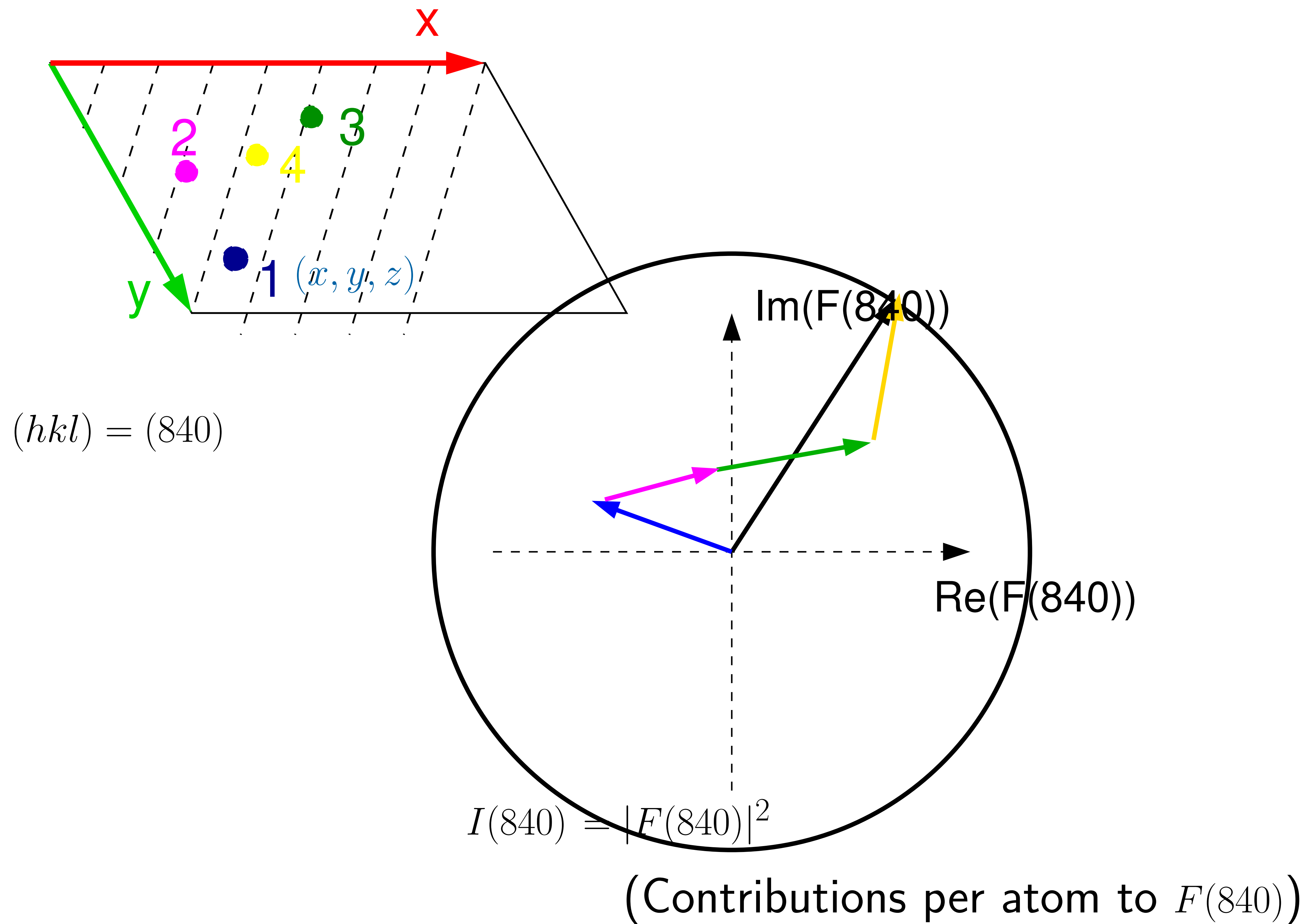
(Contributions per atom to $F(430)$)

A different view of the phase



(Contributions per atom to $F(430)$)

A different view of the phase



(Contributions per atom to $F(840)$)

A different view of the phase - Summary

- 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 total phase $\phi(hkl)$ contains convoluted information from each atom

End of lecture