

Chemical Crystallography and Structural Chemistry

(VO 270287)

Lecture 6

7th May 2020

Dr. Tim Grüne

Centre for X-ray Structure Analysis

Faculty of Chemistry

University of Vienna

tim.gruene@univie.ac.at

Previous Lecture

1. Overview: from data to structure
2. Integration: strong and weak reflections
3. Scaling: independence of the data from the experiment

Today's Lecture

1. Phasing: Solving a crystal structure
2. direct methods
3. Patterson map

Solving the Structure: Phasing

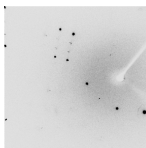
Data collection



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
100's MB

files,

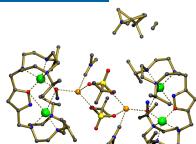
1 "hkl"-file, 50MB



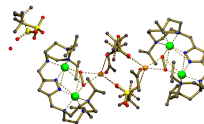
Phasing



Refinement



Starting model



Chemically sensible model

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”)
- Form factor 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:

$$F(hkl) = \int_{\text{unit cell}} \rho(x, y, z) e^{2\pi i(hx+ky+lz)} d^3x \quad (1)$$

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 $F(hkl)$

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

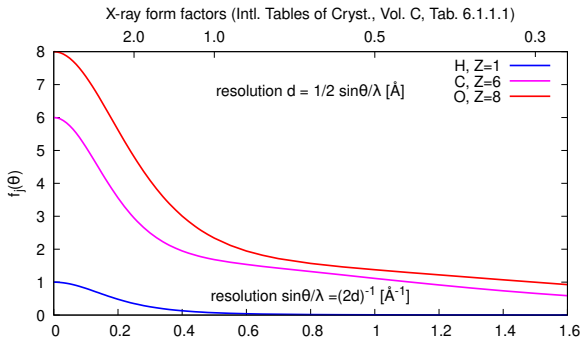
f_j atomic *form factor*. Dependent on atom element, decreases with decreasing scattering angle $\theta = \theta(hkl)$

$U_j(\theta, \lambda)$ 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 spots decreases at 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

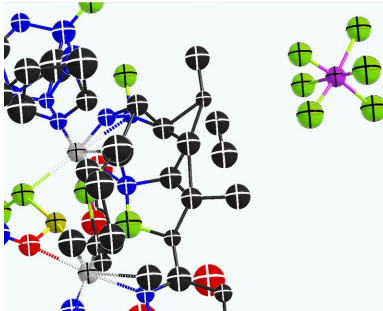
The atomic displacement parameter $U_j(\theta, \lambda)$

- 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

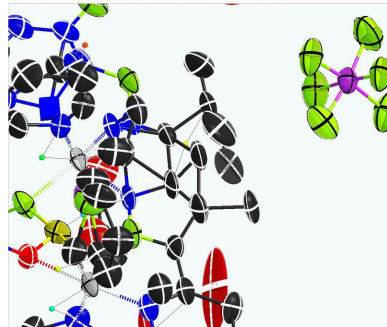
Isotropic		Anisotropic
$U_j(\theta, \lambda) = 4U_{\text{iso}} \frac{\sin^2 \theta}{\lambda^2}$	$U_j(\theta, \lambda) = (hkl) \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{12} & U_{22} & U_{23} \\ U_{13} & U_{23} & U_{33} \end{pmatrix} \begin{pmatrix} a^* \\ b^* \\ c^* \end{pmatrix}$	$\begin{pmatrix}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

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

Example images for ADP

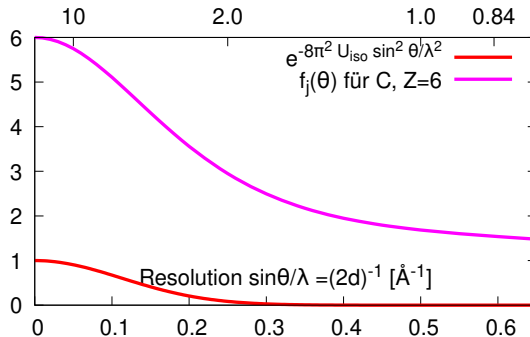


Refinement with isotropic ADPs



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

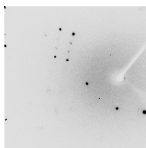
Data collection



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
100's MB

files,

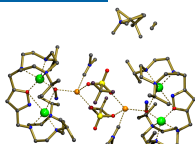
1 "hkl"-file, 50MB



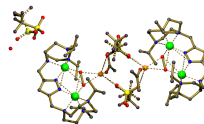
Phasing



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 $I(hkl)$.
- Previous lecture:

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

- **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

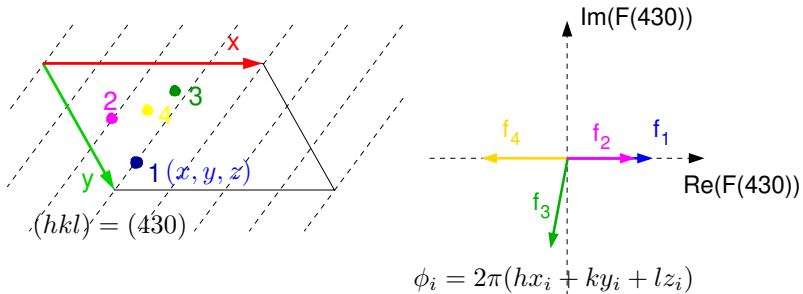
an amplitude $|F(hkl)| = \sqrt{I(hkl)}/c$

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

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

We can measure the amplitudes $|F(hkl)| = \sqrt{I(hkl)}/c$, but we cannot measure the phase angles $\phi(hkl)$. This is known as the **phase problem of crystallography**.

The phases are related to the chemical structure



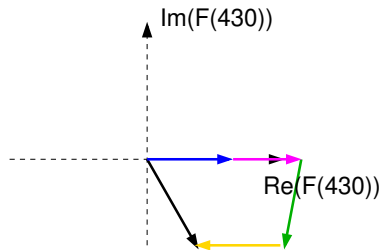
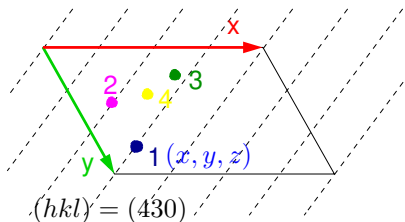
$$\phi_i = 2\pi(hx_i + ky_i + lz_i)$$

= rel. distance to origin

= rel. distance to lattice plane

Each atoms contributes to $F(430)$

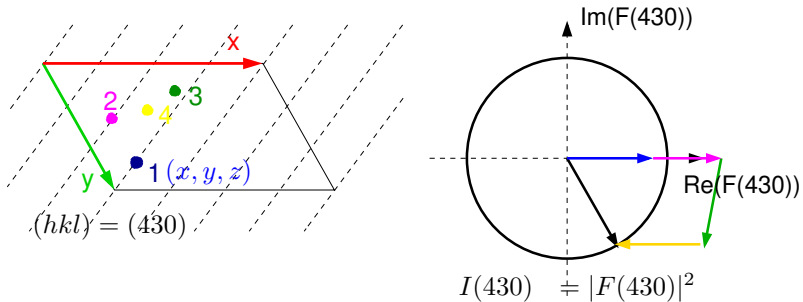
The phases are related to the chemical structure



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

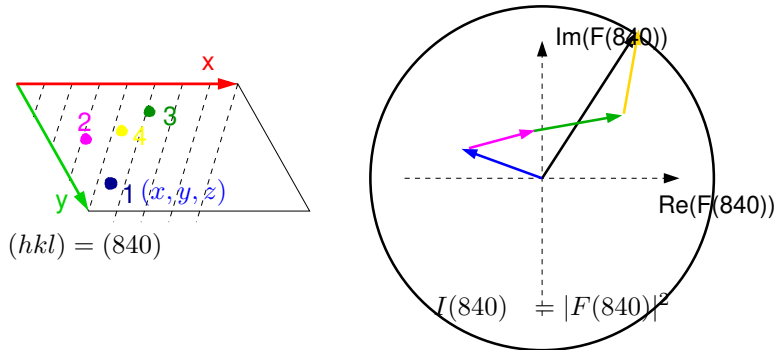
Contributions per atom to $F(430)$

The phases are related to the chemical structure



Contributions per atom to $F(430)$

The phases are related to the chemical structure



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

Phases $\phi(hkl)$ - 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

Solving the structure = Solving the phase problem

- The phase problem prevents us from calculating the electron density map directly from our data
- Phases can be calculated from a chemical model (its coordinates)
- “Phasing” means to find a model close enough to the proper model
- Once a good enough molecule has been found, it needs to be improved: “model building” and “refinement”

Phasing methods

Phasing in small molecule crystallography uses primarily one of two methods:

1. Patterson map
2. Direct methods

The Patterson map

- since 1934, Arthur Lindo Patterson (1902–1966)
- good for very small structures with some heavy elements
- direct determination of atom positions

Calculation of the Patterson map

The Patterson map ignores phases and calculates the Fourier transformation from the intensities:

$$P(uvw) = \sum_{(hkl)} I(hkl) e^{-2\pi i(hu+kv+lw)}$$

This can be calculated without knowing the phases $\phi(hkl)$, only from the measured intensities.

It turns out this map is the “auto-convolution” of the electron density with itself ($(uvw) = \vec{u}$):

$$\begin{aligned} P(uvw) &= \rho(\vec{x}) \star \rho(\vec{x} - \vec{u}) \\ &= \int_{\text{unit cell}} \rho(\vec{x}) \rho(\vec{x} - \vec{u}) d^3x \end{aligned}$$

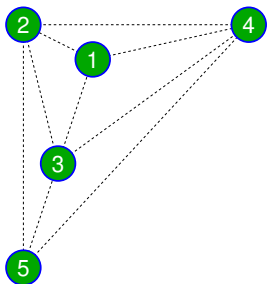
Meaning of the Patterson map

It can be shown that the Patterson map

$$\begin{aligned} P(uvw) &= \rho(\vec{x}) \star \rho(\vec{x} - \vec{u}) \\ &= \int_{\text{unit cell}} \rho(\vec{x}) \rho(\vec{x} - \vec{u}) d^3x \end{aligned}$$

has its peaks at vector (positions) \vec{u} that corresponds to the connecting vector between two atoms in the molecule in the unit cell.

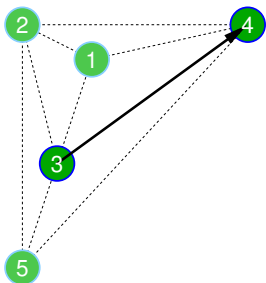
Illustration of the Patterson map in 2D



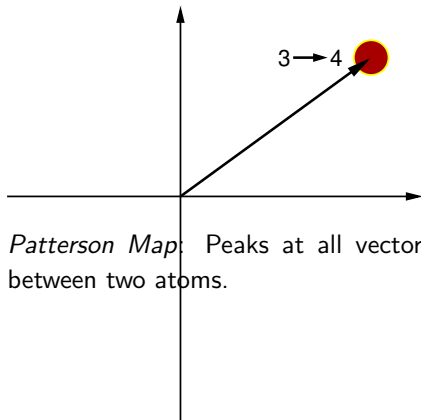
“2D molecule, 5 atoms”

peaks of Patterson map

Illustration of the Patterson map in 2D



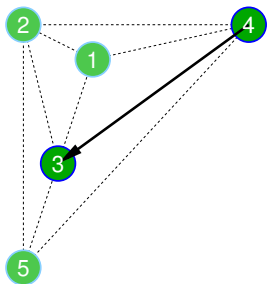
“2D molecule, 5 atoms”



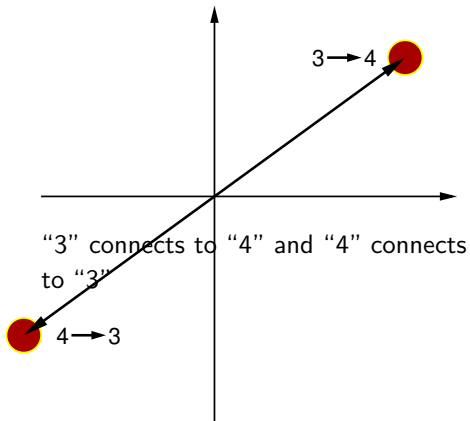
Patterson Map: Peaks at all vectors
between two atoms.

peaks of Patterson map

Illustration of the Patterson map in 2D

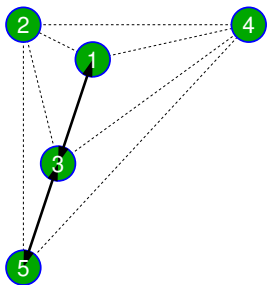


"2D molecule, 5 atoms"

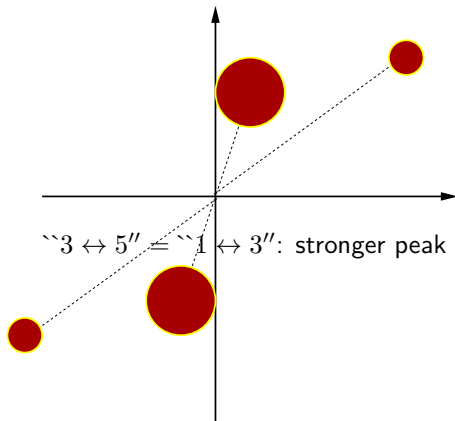


peaks of Patterson map

Illustration of the Patterson map in 2D

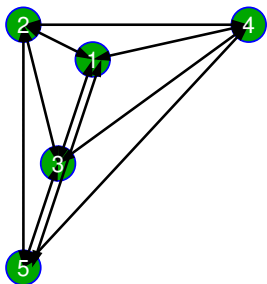


"2D molecule, 5 atoms"



peaks of Patterson map

Illustration of the Patterson map in 2D



“2D molecule, 5 atoms”

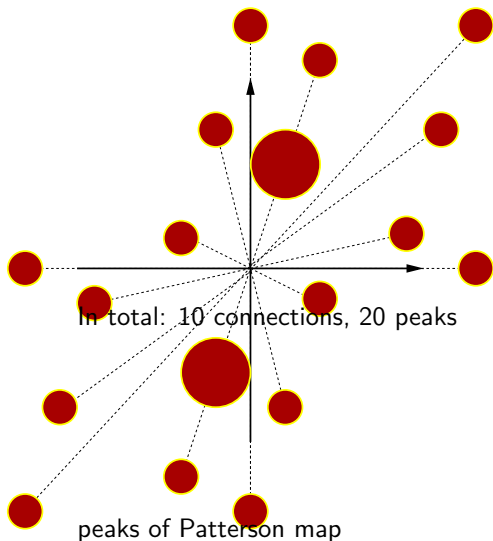
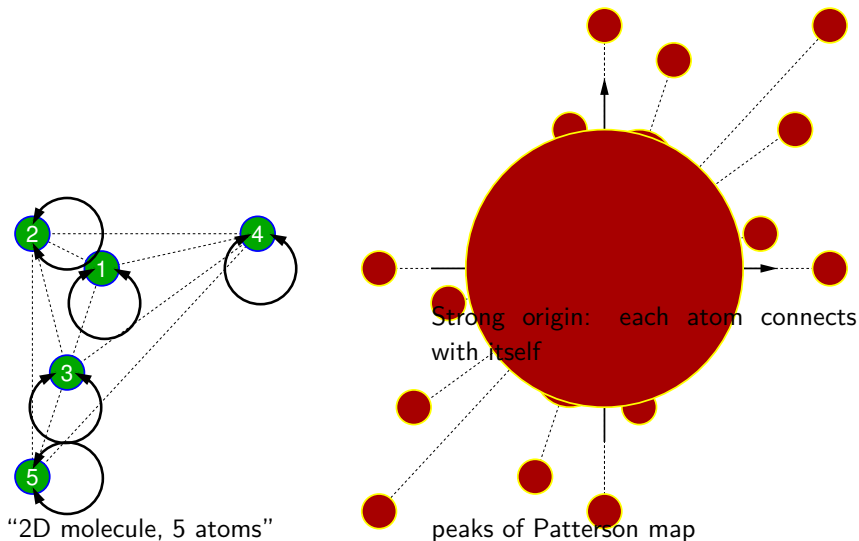


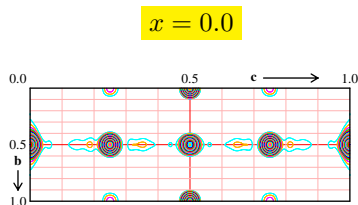
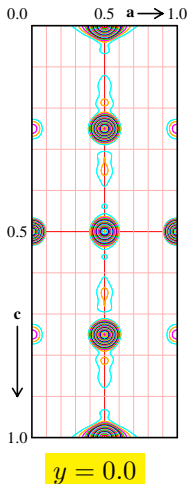
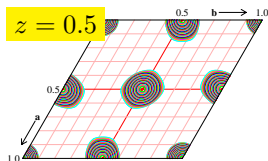
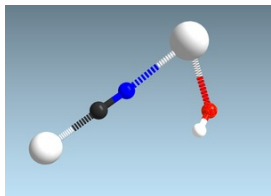
Illustration of the Patterson map in 2D



Patterson map observations

- heavy elements have stronger peaks (high density $\rho(x, y, z)$)
- with too many atoms: origin peak overwhelms: non-interpretable
- Patterson map always centro-symmetric (peak at $(x, y, z) \Leftrightarrow$ peak at $(-x, -y, -z)$)

Patterson map for $La[Au(CN)_2]_3 \cdot 3H_2O$



- La(III): $54e^-$, Au(I): $78e^-$, O^{2-} : $10e^-$
- $54 \cdot 78 = 4'212 \gg 78 \cdot 10 = 780$: 1 dominating peak
- Patterson maps: typically "origin peak removed"

Summary Patterson map

- Patterson map calculated from intensities, without phases
- Patterson map corresponds to convolution of density $\rho(x, y, z)$
- Peaks correspond to connecting vectors between atoms
- Peak height corresponds to product of number of electrons
- Atom coordinates can be deduced from map in case of few atoms, or few heavy atoms
- The more atoms (of similar weight), the harder to interpret

Direct methods

Direct methods

- Well suited with molecules of similar atom types (organic compounds with *C, N, O, ...*)
- Can work with thousands of atoms
- Requires atomic resolution, better than 1.2 Å (Sheldrick's rule) (Morris & Bricogne, "Sheldrick's 1.2 Å rule and beyond" *Acta Cryst.* (2003), D59, 615–617)

Concept of direct methods

1. Generate roughly the number of expected atoms at arbitrary positions
2. Calculate phases of this pseudo-molecule
3. Improve phases based on **tangent formula**
4. Improved phases produce an improved electron density map
5. Peak picking from improved map
6. Repeat
7. Best solution: assign atom types

Direct methods: the tangent formula

Tangent formula¹ was derived by H. A. Hauptman and J. Karle — chemistry Nobel prize 1985

$$\tan(\phi_{\mathbf{h}}) \approx \frac{\sum_{\mathbf{h}'} |E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'}| \sin(\phi_{\mathbf{h}'} + \phi_{\mathbf{h}-\mathbf{h}'})}{\sum_{\mathbf{h}'} |E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'}| \cos(\phi_{\mathbf{h}'} + \phi_{\mathbf{h}-\mathbf{h}'})}$$

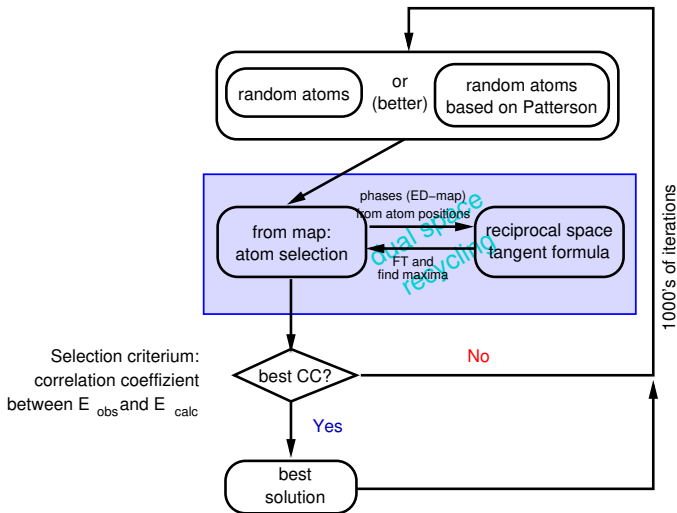
“entangles” phases with one another. Historically based on Sayre-Equation (1952)

$$F(hkl) = q(\sin \theta / \lambda) \sum_{(h'k'l')} F(h'k'l') * F(h - h', k - k', l - l')$$

Sayre equation is exact for cases of only one atom type in crystal (diamond, silicon, etc). That is why the tangent formula works best for similar-atoms-compounds.

¹ $E(hkl)$: normalised structure factors, derived from measured $F(hkl)$

Direct methods: dual space recycling



Amazingly, it works

Structure Refinement

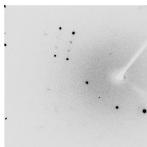
Data collection



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

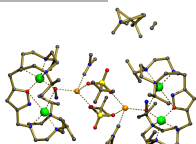
several
100's MB

files,

1 "hkl"-file, 50MB



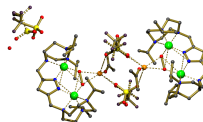
Phasing



Starting model



Refinement



Chemically sensible model

End of lecture