Methods in Chemistry III – Part 1
Modul M.Che.1101 WS 2010/11 – 8
Modern Methods of Inorganic Chemistry

Mi 10:15-12:00, Hörsaal II

George Sheldrick

gsheldr@shelx.uni-ac.gwdg.de

The phase problem

•In order to calculate the electron density we need both the phases ϕ and the intensities I of the reflections h,k,ℓ.

•The information content of the phases is appreciably greater than that of the intensities.

•In real life, it is essentially impossible to measure the phases.

This is known as the *crystallographic phase problem* and appears to be impossible to solve!

In spite of this, nowadays small molecule crystal structures can be solved using the *Patterson function* and the *heavy atom method*, or (much more often) with *direct methods* (probability relations involving phases) without significant difficulties.

What are phases (1)?

The electron density is obtained by Fourier transformation of the structure factors F_{hkl}:

$$\rho_{xyz} = (1/V) \sum_{hk\ell} F_{hk\ell} \exp[-2\pi i(hx+ky+\ell z)]$$

F_{hkl} are complex numbers with amplitudes F and phases ϕ. F is proportional to the square root of the measured intensity I. The contribution of one atom has the phase $2\pi(hx+ky+\ell z)$. The inverse transformation, to calculate F_{hkl} from ρ_{xyz}:

$$F_{hk\ell} = {}_V\!\!\int \rho_{xyz} \exp[+2\pi i(hx+ky+\ell z)]\, dV$$

can also be represented as a sum over all atoms j:

$$F_{hk\ell} = \sum_{j=1}^{n} f_j \{ \cos[2\pi(hx_j+ky_j+\ell z_j)]+i\cdot\sin[2\pi(hx_j+ky_j+\ell z_j)] \}$$
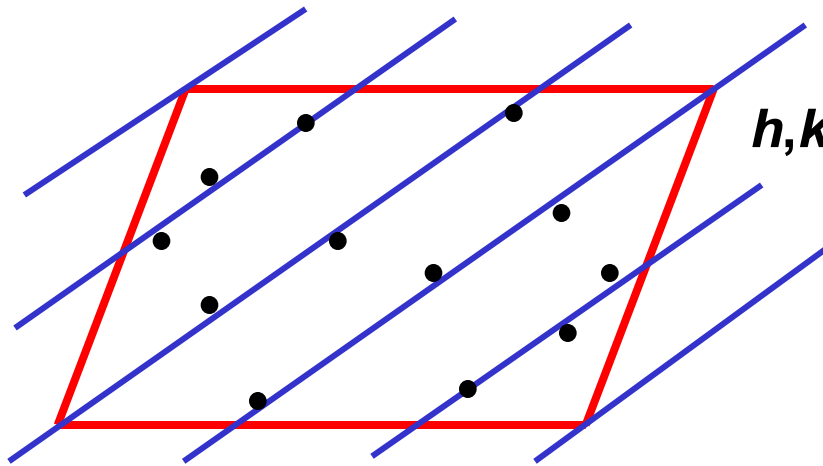
where f_j is the scattering factor of the atom j. For centrosymmetric structures the sine terms for x,y,z and $-x,-y,-z$ in the summation cancel each other out; ϕ is then always 0^0 or 180^0. In the special case of one atom on the origin, all phases are 0^0.
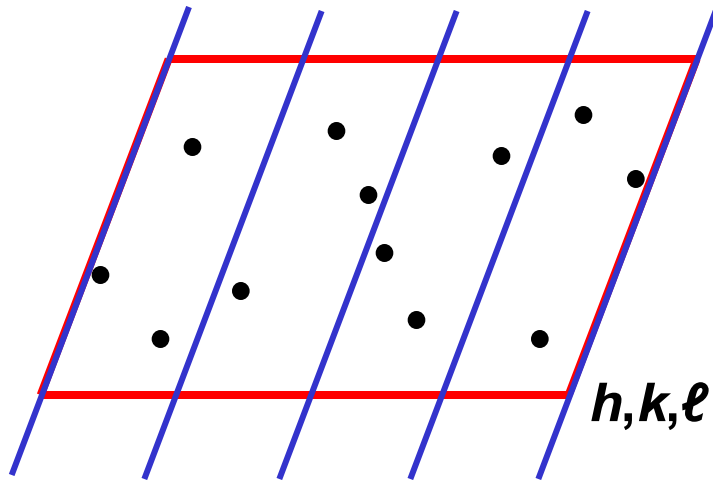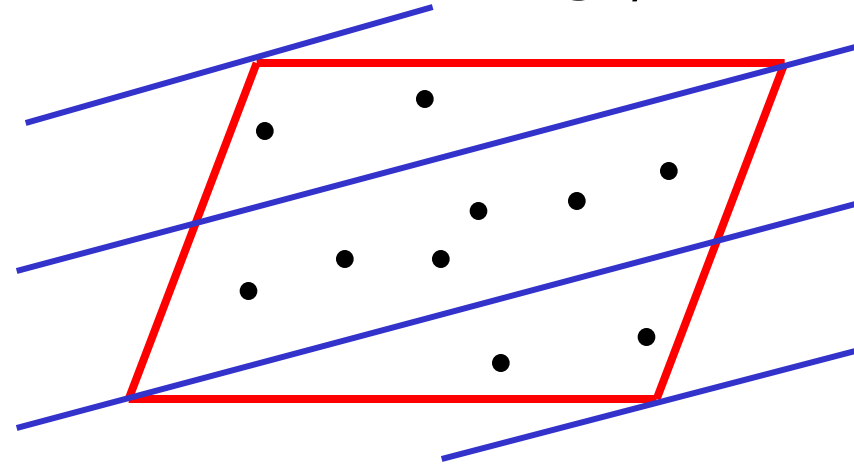
What are Phases (2)?

Each atom x_j, y_j, z_j in the unit cell makes a contribution with phase $2\pi(hx_j+ky_j+\ell z_j)$ and amplitude f_j to each reflection h,k,ℓ. The total structure factor $\mathbf{F_{hkl}}$ is the vector sum of these contributions. The atoms are not really point atoms; f_j also takes into account the electron distribution within an atom and the thermal motion.

$$f_j\{\cos[2\pi(hx_j+ky_j+\ell z_j)]+i\cdot\sin[2\pi(hx_j+ky_j+\ell z_j)]\}$$

imaginär

F

ϕ

reel
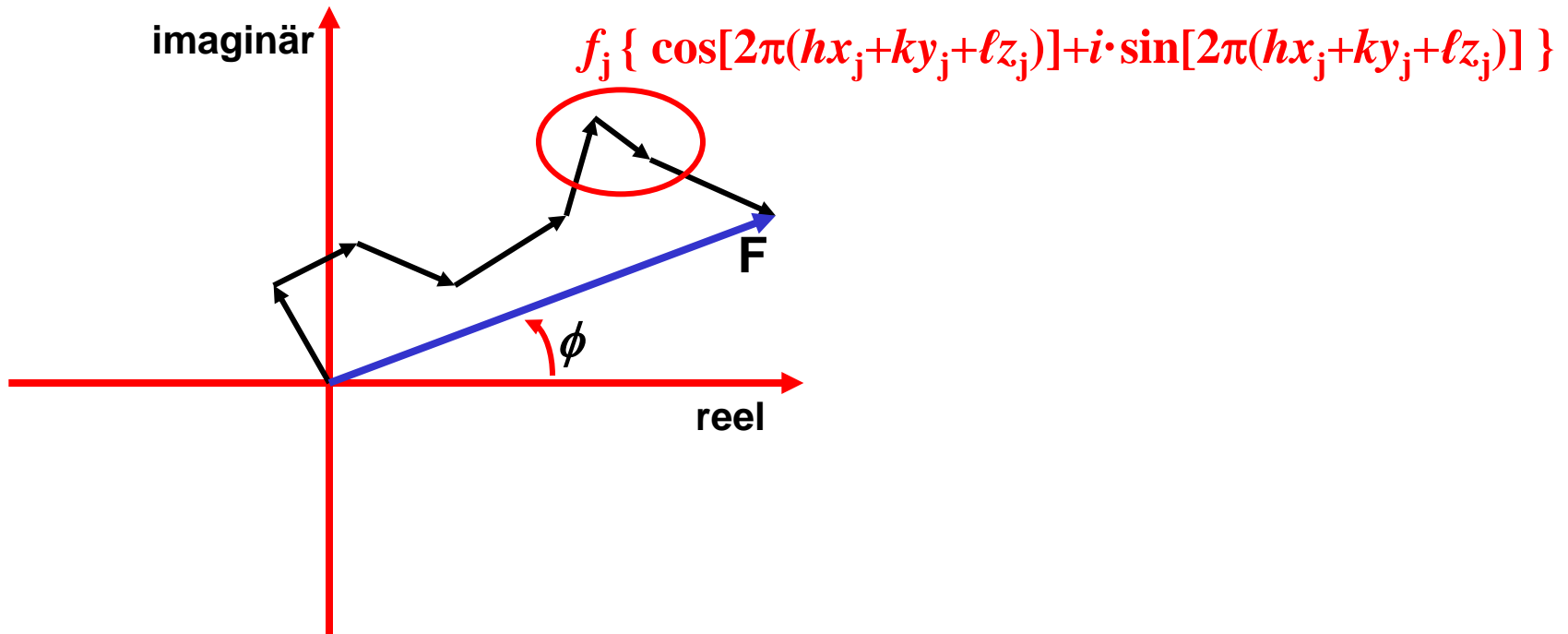
What are phases (3)?

$h,k,\ell = 2,3,0$; strong, $\phi = 0^0$

$h,k,\ell = 2,1,0$; strong, $\phi = 180^0$

$h,k,\ell = 0,4,0$; weak, $\phi = ?$

The phase ϕ of a reflection is determined by its relation to the origin of the cell. When the atoms are concentrated in planes that run through the origin, ϕ is about 0^0; when they predominantly lie half-way between such planes, ϕ is about 180^0.

The Sayre equation (1952)

The electron density (0 or Z) of a *point atom structure* with *equal atoms* is proportional to its square (0 or Z^2). The convolution theorem leads then directly to the *Sayre equation:*

$$F_h = q \sum_{h'} (F_{h'} F_{h-h'})$$

q is a constant that depends on $\sin(\theta)/\lambda$ of the reflection h ($hk\ell$). The summation is performed over all reflections h' ($h'k'\ell'$). Often F_h is *normalized* to E_h, in order to eliminate the effects of the electron distribution within an atom and the thermal motion. E_h corresponds to the structure factor of a *point atom structure.*

$$E_h^2 = (F_h^2/\varepsilon) / <F^2/\varepsilon>_{\text{resolution shell}}$$

The statistical factor ε is usually 1, but can be larger for special reflection classes (e.g. 00ℓ in tetragonal space groups).

The triplet equation

In the Sayre equation with *E* instead of *F*: $E_h = q \sum_{h'} (E_{h'} E_{h-h'})$

one can compare the phases on the left and right hand sides for one reflection h' only (instead of a summation over all reflections):

$$\phi_h = \phi_{h'} + \phi_{h-h'} \quad \text{(modulus 360°)}$$

If this were an exact equation, it would be possible to deduce all phases iteratively, but unfortunately it is subject to statistical fluctuations. With the help of suitable assumptions, for example that the structure consists of randomly distributed equal atoms (see the Sayre equation), one can deduce the probability distribution (Cochran, 1955):

$$P(\Phi) = g \exp (2|E_h E_{h'} E_{h-h'}| \cos(\Phi) / N^{1/2})$$

where $\Phi = \phi_h - \phi_{h'} - \phi_{h-h'}$, *g* is a normalizing factor and *N* is the number of atoms per lattice point. It follows that the phase relations are more reliable for large *E*-values and small structures!

The tangent formula (Karle & Hauptman, 1956)

The tangent formula, often in well disguised form, is still the key formula for structure solution with direct methods. This formula is nothing more than the standard formula to calculate the phase of a sum of complex numbers! It is used to calculate the phase of the summation on the right hand side of the Sayre equation:

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

In the calculation of the phase angle it is important to consider the signs of both the numerator and the denominator, so that angles can be calculated in the full range 0-360°. With computational brute force it is possible to determine small structures by applying this equation iteratively starting from random phases.

The Nobel prize for chemistry 1985 was awarded to J. Karle and H. A. Hauptman for their contributions to *direct methods*.

How does one find the minimum?

DIRECT METHODS IN PRACTICE ...
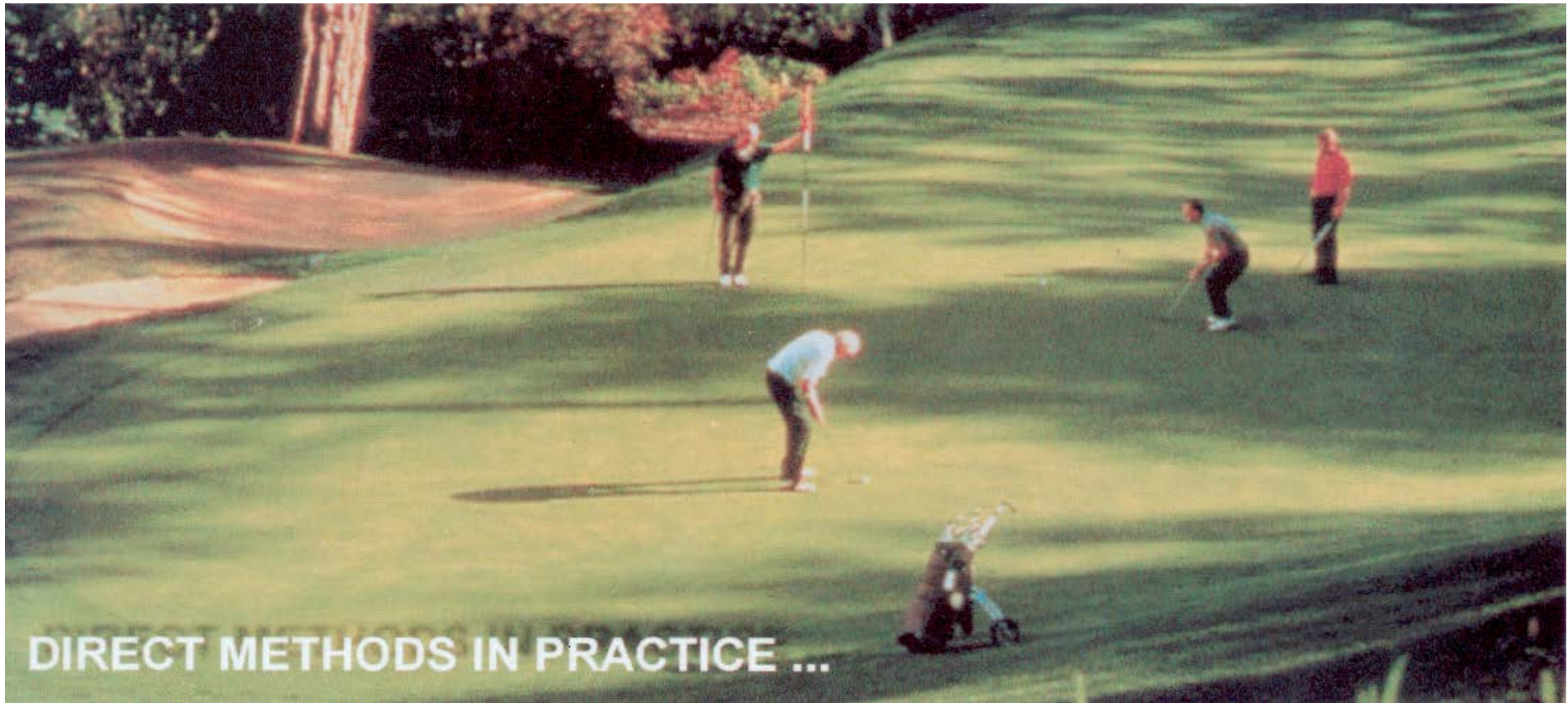
The phase problem is really a search problem; there are various structures (minima) that are reasonably consistent with the tangent formula or related equations, including both the correct solution and the statistically too perfect *uranium atom* pseudo-solution that has all phases zero!
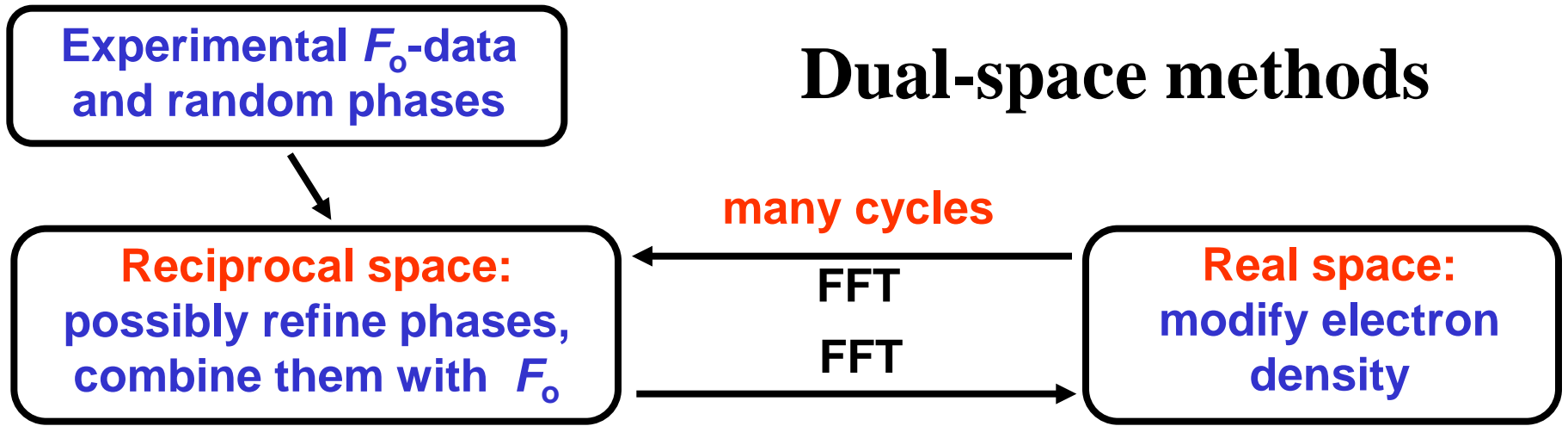
The limits of direct methods in reciprocal space

Conventional direct methods, using improved versions of the tangent formula, are computationally extremely efficient for the solution of structures of up to about 100 unique atoms. Small structures can even be solved in one or two seconds. In spite of this, these methods rapidly run out of steam for larger structures, and few structures larger than 200 atoms have been solved in this way.

The efficiency of the tangent formula as a search algorithm lies in its ability to relate phases of reflections that lie far apart in reciprocal space. The weaknesses of this formula – for example the tendency to produce a uranium atom solution – become more serious as the structures get larger.

It is necessary to restrict the phases so that they correspond to a chemically sensible structure. *Dual-space methods* are able to do this and can solve much larger structures (up to about 1000 atoms).

Dual-space methods

Experimental F_o-data and random phases	

↓

| **Reciprocal space:** possibly refine phases, combine them with F_o | ← **many cycles** FFT / FFT → | **Real space:** modify electron density |

Dual-space methods iterate between real and reciprocal space using *Fast Fourier Transforms*. In real space atoms may be found or (for example) the density ρ is replaced by $-\rho$ if it is lower than a preset threshold (*charge flipping*).

The dual-space iterations may be performed in the space group P1 and then, after the structure has been solved, the space group is determined (without the help of systematic absences).

All *direct methods* require in practice diffraction data to *atomic resolution* (1.2 Å or better). It is not yet clear whether the atoms must really be resolved from one another or whether the data-to-parameter ratio is the critical factor.

Exercises

1. For a structure in space group $P\bar{1}$, the reflections 1 1 0 und 2 2 0 were both strong. What can one deduce about their phases (a) graphically and (b) with the help of the triplet equation?

2. What are the phases of reflections (a) $h+k+\ell=2n$ and (b) $h+k+\ell\neq2n$ (where n is a whole number) when there is one heavy atom at x=½, y=½, z=½?

3. For a small structure in the space group P1 the reflections 1 3 2, 4 –1 1, 5 2 3 and 6 5 5 are all very strong. Estimate the phase of the reflection 6 5 5 when the phase of 1 3 2 is 40^0 and the phase of 4 –1 1 is 270^0. Why is this prediction less reliable when (a) the structure is larger or (b) the reflection 5 2 3 is weaker?