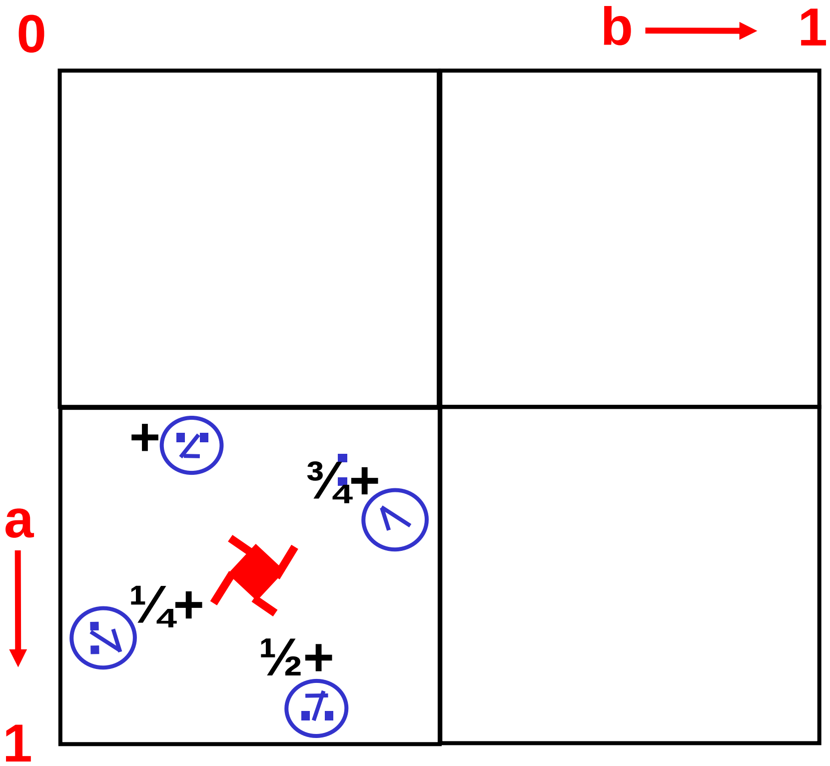
Lesson 20

- Solving the structure
  - Getting a good login
  - The Phase Problem
  - Charge Flipping

Exercises

- **Find the 11 enantiomorphic pairs of space groups (hint: all are of course chiral and possess 3_N, 4_N or 6_N axes).**
- **P31/P32 P41/P43 P61/P63**
- **P41212/P432121 P4122/P4322**
- **P3112/P3212 P3121/P3221 P62/P64**
- **P6122/P6322 P6222/P6422**
- **P4132/P4332**
- **The incomplete picture should show the space group $I4_1$. Why is there no space group $I4_3$? The I creates a 43 in the cell**



. Assign the space group of a tetragonal crystal with the systematic absences:

kℓ, $h+k+\ell = 2n$; $hk0$, $h = 2n$ or $k = 2n$; $hh\ell$, $2h+\ell = 4n$ I41/a

Electron Density in Crystals

- Since the crystal is made up of repeating unit cells, the electron density in the cell must be periodic!
- This is called the Charge Density Wave.
- In 1-dimension, the length of the charge density wave is the cell length.

The phase problem

- In order to calculate an electron density map, we need both the intensities (|*F* | is proportional to the square root of *I*) and the phases ϕ of the X-ray reflections *hkl*.

- The information content of the phases is significantly greater than that of the intensities.

- It is unfortunately not normally possible to determine the phases experimentally.

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

In spite of this, small molecule structure are solved routinely these days, usually without serious problems, with the help of the *Patterson function* and the *heavy atom method*, or (more commonly) by so-called *direct methods* based on probability relations involving phases.

The structure factor F and the electron density ρ

$$I_{hk\ell} \text{ is proportional to } |F_{hk\ell}|^2$$

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

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

$F_{hk\ell}$ and ρ_{xyz} are related to each other by means of these *Fourier transformations*. The electron density ρ is real and positive, but the structure factor F is a complex number: to calculate the electron density from the structure factors, we also need the *phase* (ϕ) of *F*.

Under normal conditions we can only measure the diffracted intensities *I* and not the phase ϕ of a reflection *h*,*k*,ℓ, so it appears that we are faced with an insoluble problem, the *crystallographic phase problem.*

What are phases (1)?

The electron density is obtained by Fourier transformation of the structure factors F_{hkl}. F_{hkl} are complex numbers with amplitudes and phases.

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

The reciprocal transformation to calculate F_{hkl} from ρ_{xyz}:

$$F_{hkl} = \int_V \rho_{xyz} \exp[+2\pi i(hx+ky+lz)]\, dV$$

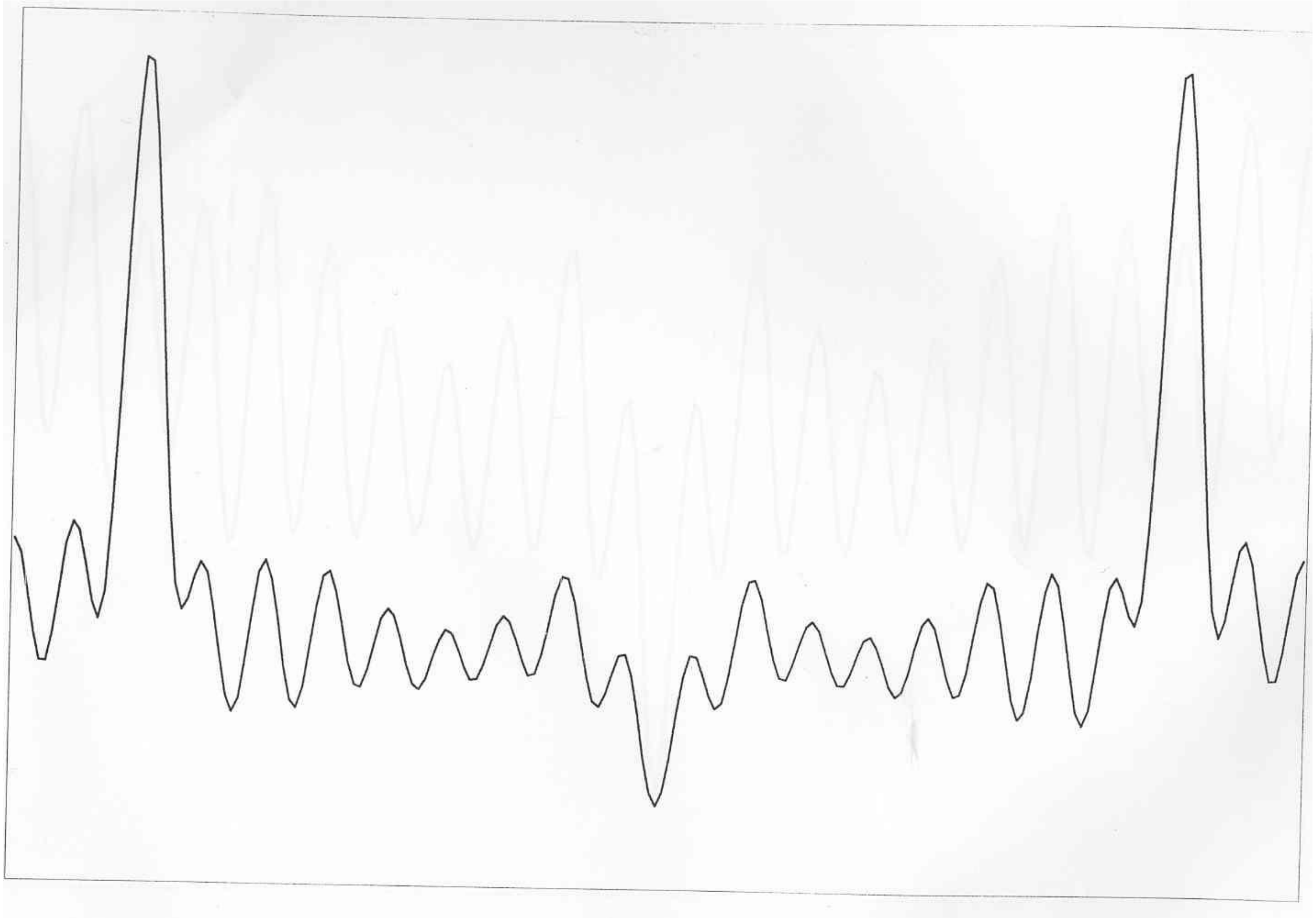
can also be represented as a sum over the *n* atoms in one unit-cell:

$$F_{hkl} = \sum_{j=1}^{n} f_j \exp[+2\pi i(hx_j+ky_j+lz_j)] \qquad \text{or}$$

$$F_{hkl} = \sum_{j=1}^{n} f_j \{ \cos[2\pi(hx_j+ky_j+lz_j)]+i\cdot\sin[2\pi(hx_j+ky_j+lz_j)] \}$$

where f_j is the scattering factor of atom j (including thermal motion). If the structure is centrosymmetric, the sine terms for x,y,z and $-x,-y,-z$ cancel, with the result that the phase of F_{hkl} must be 0º or 180º. In the special case of one atom on the origin, all phases must be 0º.

Correct Phasing



Incorrect Phasing

How to know correct phasing

- The electron density is always greater or equal to zero
- The electron density is highly localized at the atom positions.
- Since the distance between maxima represents bond distances, the peaks should be distinct.

Approaches to Structure Solution

- The process of overcoming the phase problem is referred to as structure solution.
- There are several approaches
  - Directly Determine a set of consistent phases by using probability relations—direct methods
  - Try to correctly place at least 15% of the density within the cell—Heavy atom methods
  - Some combination of the two
  - Charge flipping

Charge Flipping

This is the newest method (G. Oszlányi and A. Süto, *Acta Cryst.* (2004). A**60**, 134-141)

This is an *ab initio* method that does not require the formula or space group

The method is done as follows

1. Assign arbitrary phases to the observed data
2. Do a Fourier Transform to electron density space
3. Change the sign of any density which is less than some value delta (note: cannot use initially zero for delta as it will converge to a false solution)
4. Do a Fourier Transform to calculate data.
5. Apply the calculated phases to the observed data
6. Go to step 2.

Incorrect Phasing

Charge Flipping

- Advantages
  - Do not need either the space group or the formula
  - Very easy to program and understand
- Disadvantages
  - Very slow –two Fourier Transforms/cycle
  - Badly effected by noise (weak) data.
  - Provides the electron density in the entire cell—must then figure out the space group and element assignments to proceed.

PLATON

- Charge flipping is programmed in PLATON a crystallographic tool box program
- PLATON is written by Antony (Ton) Spek at the University of Utrecht in the Netherlands
- PLATON is free for academics and is available for WINDOWS and LINUX.
- PLATON is used for graphics, analysis, and other functions. It will be used for many operations in this course.

Superflip

The program superflip in its Shelx equivalent version has been incorporated into the solve gui.

This program is very fast as it uses extremely efficient Fourier transform methods.

The current software to mate it to the Purdue package is in the beta state and may have errors.

Direct Methods

- This is an attempt to create a consistent set of phases starting with a few arbitrary phases
- Makes use of several features of the electron density
  - Density is always positive
  - Density is localized
  - The average density is close to zero

Kasper-Harker Inequality

F_{000} is simply the sum of all the electrons in the unit cell. This is F for the beam at $\Theta=0°$

The unitary structure factor u_{hkl} is defined as $u_{hkl}=F_{hkl}/F_{000}$

u will be 1 if every atom in the unit cell is scattering in phase.

Obviously most u values are much less than 1

Kasper-Harker Inequality

In 1948 Kasper and Harker published the following inequality

$$u^2_{hkl} \leq 1/2 + 1/2u_{2h,2k,2l}$$

This means that if both u_{hkl} and $u_{2h,2k,2l}$ are both large that it is probable that the sign of $u_{2h,2k,2l}$ is +

$$u^2_{hkl} \leq 1/2 + 1/2u_{2h,2k,2l}$$

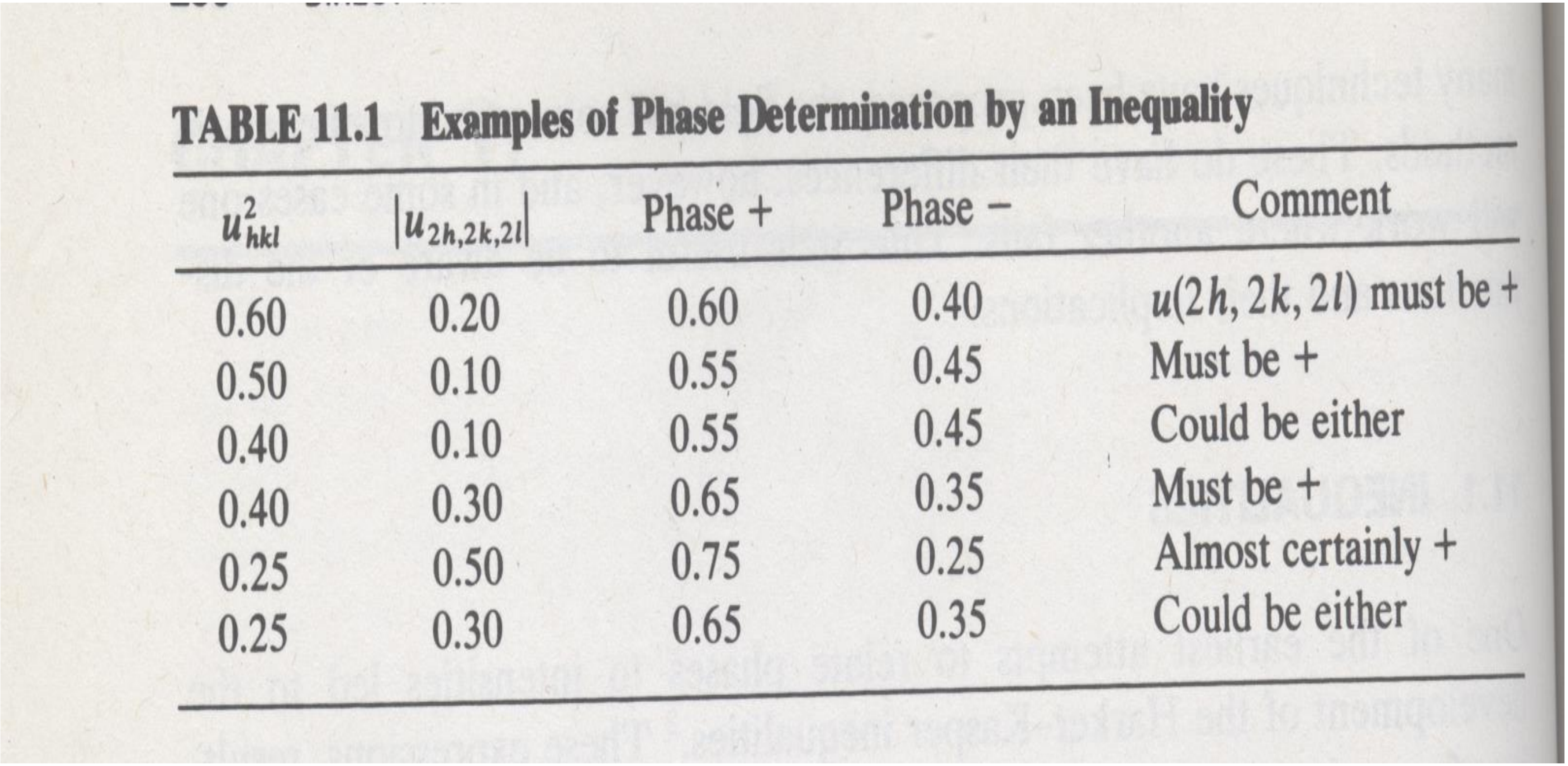


TABLE 11.1 Examples of Phase Determination by an Inequality

u^2_{hkl}	$\lvert u_{2h,2k,2l}\rvert$	Phase +	Phase −	Comment
0.60	0.20	0.60	0.40	$u(2h, 2k, 2l)$ must be +
0.50	0.10	0.55	0.45	Must be +
0.40	0.10	0.55	0.45	Could be either
0.40	0.30	0.65	0.35	Must be +
0.25	0.50	0.75	0.25	Almost certainly +
0.25	0.30	0.65	0.35	Could be either

What is the big deal

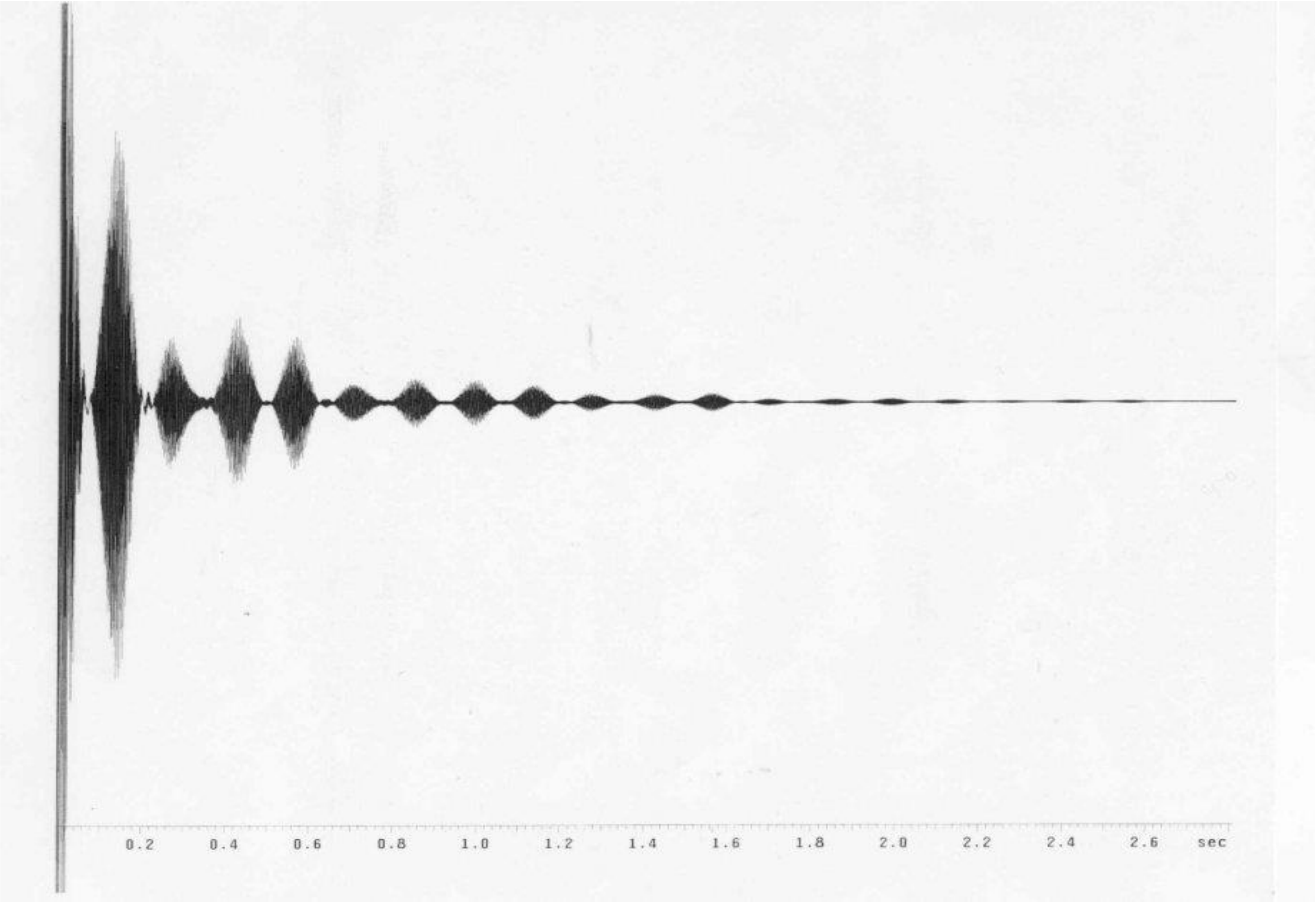
This suggests that it is possible to determine something about the phases from intensity information.

By itself the Kasper-Harker inequality is useless as there are very few cases where the two needed reflections are both strong.

FT-NMR

- The wavelength domain of an NMR has properties similar to that of real space electron density.
- Try to see what we can see in the fid that is like x-ray data and how we might learn something from it.
- Unlike an NMR fid the x-ray data is unsigned and not continuous

FID



FID Analysis

- It is clear that the extremes are the key.
- If the maximum is located then it can be assumed the data nearest to it share the same sign.
- However finding the extremes is more difficult than just finding the largest absolute values. Since the intensity falls off with time (or theta for x-rays) the largest values are not always a maximum

How to Find Intense Data

- If the expected value at a particular time (theta) can be found then data that are very much larger than the expected value suggest they are near a maximum.
- For the x-ray case the expectation value can be found for a particular theta from $<f>=k\epsilon\ \Sigma f_i^2\ \mathrm{exp}(-2B\sin\theta^2/\lambda^2)$
- This is equivalent to placing all the electrons in the cell in one average density

- $k\epsilon\ \Sigma f_i^2\ \exp(-2B\sin\theta^2/\lambda^2)$

- What are the symbols.
- The f's are the scattering factor for each atom which is summed over all the atoms in the cell. This requires at least an estimate of the formula and Z must be provided!! If the formula is way off then the expectation value will be incorrect.
- B is the isotropic adp

k is the scale factor

Є is a scaling factor for some classes of reflections

Scale Factor

- Obviously the calculated values for F (the corrected square root of the intensity of a given reflection) depends only on the structure.
- The observed data depends on the size of the crystal, the intensity of the x-ray beam, <u>etc.</u>
- To compare observed vs calculated data there needs to be a scale factor that places the observed data on the same scale as the calculated data

Initial Values of B and k

- To calculate the expectation value both B and k are needed.
- These can be calculated from a Wilson plot—see Massa for details.
- Note B is calculated from the linear slope of a function that is anything but linear.
- Bad values of B can badly effect direct methods results!

E's

- The real intense data can be located by dividing the observed value of F by the expectation value at that theta.
- The result are called normalized structure factors and are given the symbol E.
- After the E's are calculated they are renormalized so that the average value of E is 1.0
- Therefore E>1.0 are strong reflections.

The Sayre equation
(1952)

In the same issue of Acta Crystallographica in 1952, Sayre, Cochran and Zachariasen independently of each other proposed phase relations that could be summarized with the equation:

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

This is known today as the ***Sayre equation***. q is a constant that depends on sin(θ)/λ of the reflection h (hkl). The summation is performed over all reflections h' ($h'k'l'$). For a structure with equal point atoms, the electron density is proportional to its square. The convolution theorem then leads directly to the Sayre equation. This equation is under these conditions an exact relationship, provided that the summation is performed over *all* reflections, including F_{000} (which can not normally be measured).

Normalized structure factors

Direct methods turn out to be more effective when the observed structure factors F_h are modified to allow for the effects of the electron distribution within an atom and its thermal motion. The resulting normalized structure factors E_h correspond to the structure factors 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 greater for special reflection classes (e.g. 4 for 00*l* in tetragonal space groups). $<F^2/\varepsilon>$ can either be calculated directly or modeled with the help of an exponential function (Wilson plot).

The triplet relation

For the Sayre equation with *E* instead of *F*:

$$E_h = q \sum_{h'} (E_{h'} E_{h-h'})$$

one can compare the phases of the left and right sides for a single *h*' contribution instead of a summation:

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

If this equation were exact, it would be possible to use it to derive all the phases from a small number of starting phases. However it should be understood in a statistical sense. Starting from suitable assumptions, for example that the structure consists of equal, randomly distributed atoms, it is possible to derive a probability distribution for this equation (Cochran, 1955):

$$P(\Phi) = g \exp(2|E_h E_{h'} E_{h-h'}|\cos(\Phi)/N^{½})$$

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

The tangent formula (Karle & Hauptman, 1956)

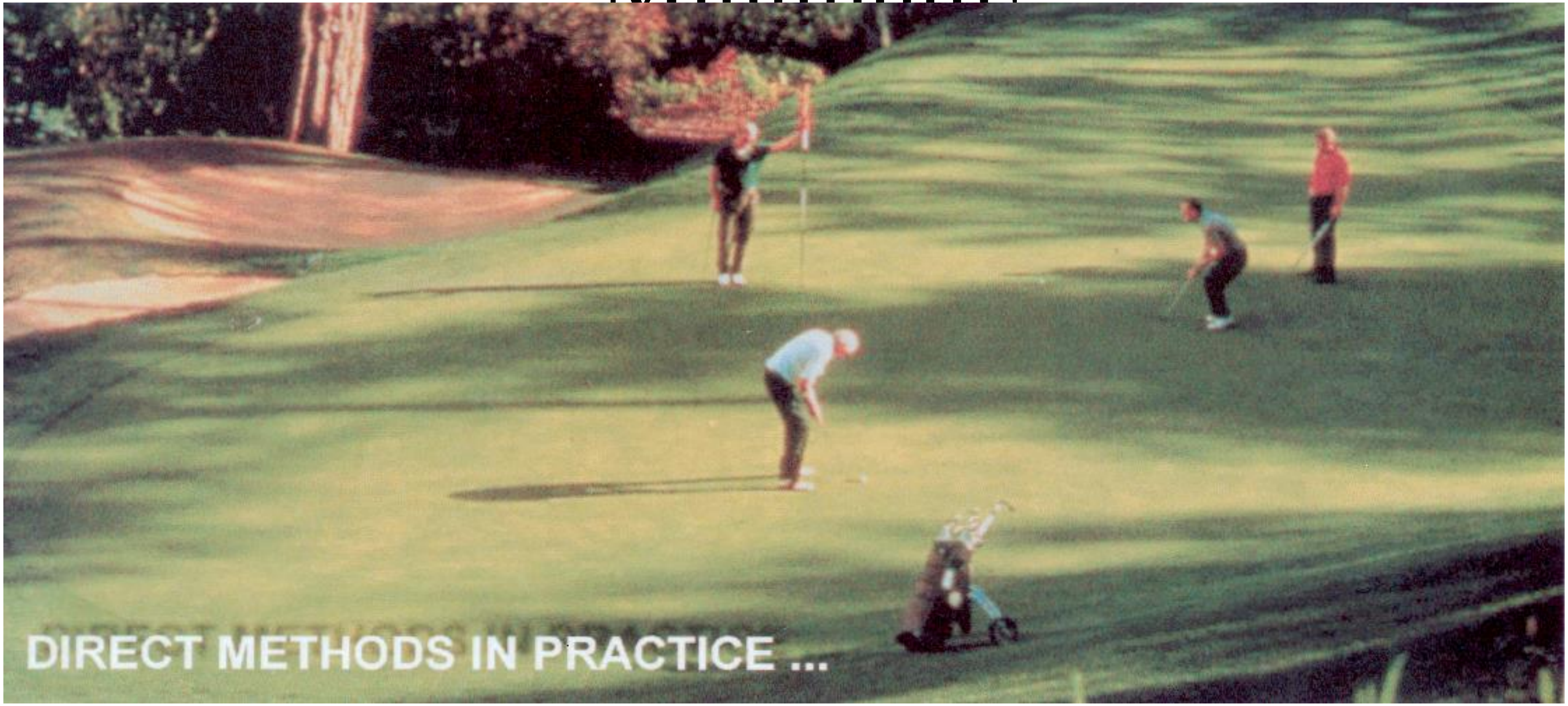
The tangent formula, often in a well disguised form, is still the key formula for the solution of structures by direct methods. The formula is nothing more than the standard method of calculating the phase of a sum of complex numbers; it is used to calculate the phase of the sum represented by the Sayre 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 phases, the signs of the numerator and denominator must both be taken into account, so that the resulting angle is in the full range 0–360°. With sufficient number crunching power it is possible to solve small structures starting from random phases by repeated application of this equation.

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

How can one find the Minimum?



The phase problem is really a search problem; there are various structures (minima) that are reasonably consistent with the tangent or related equations, including both the correct solution and the statistically too perfect *uranium atom* pseudo-solution which 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 structures larger than about 200 independent equal atoms, 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 tendancy to produce a uranium atom solution – become more serious as the structures get larger. It is necessary to constrain the phases in some way so that they correspond to a chemically sensible structure.

All such *direct methods* require data to a resolution of 1.2 Å or better (*atomic resolution*). Whether this is because the atoms have to be resolved from each other, or simply because the data to parameter ratio needs to be high enough, is not clear.

Direct Methods in Practice

- Calculate the E's
- Try to find a few reflections that can be used to produce the entire data set.
  - The parity groups of these reflections are important
  - eee, eeo, eoe, oee, ooe, oeo, eoo, ooo
  - Some programs normalize E's so that each parity group's average is 1.0
- Arbitrarily assign phases to these starting reflections and expand to derive as many phases as possible.

A Problem

- This approach produces many possible solutions. Obviously most are not correct.
- How can the quality be evaluated.
- Do a Fourier Map on the E's and analyze this for chemical correctness. This is way to slow.
- Check on consistency
  - do the expanding set of phases suggest common phases on other data
  - do the phases predict the weak relfections are weak

FOM

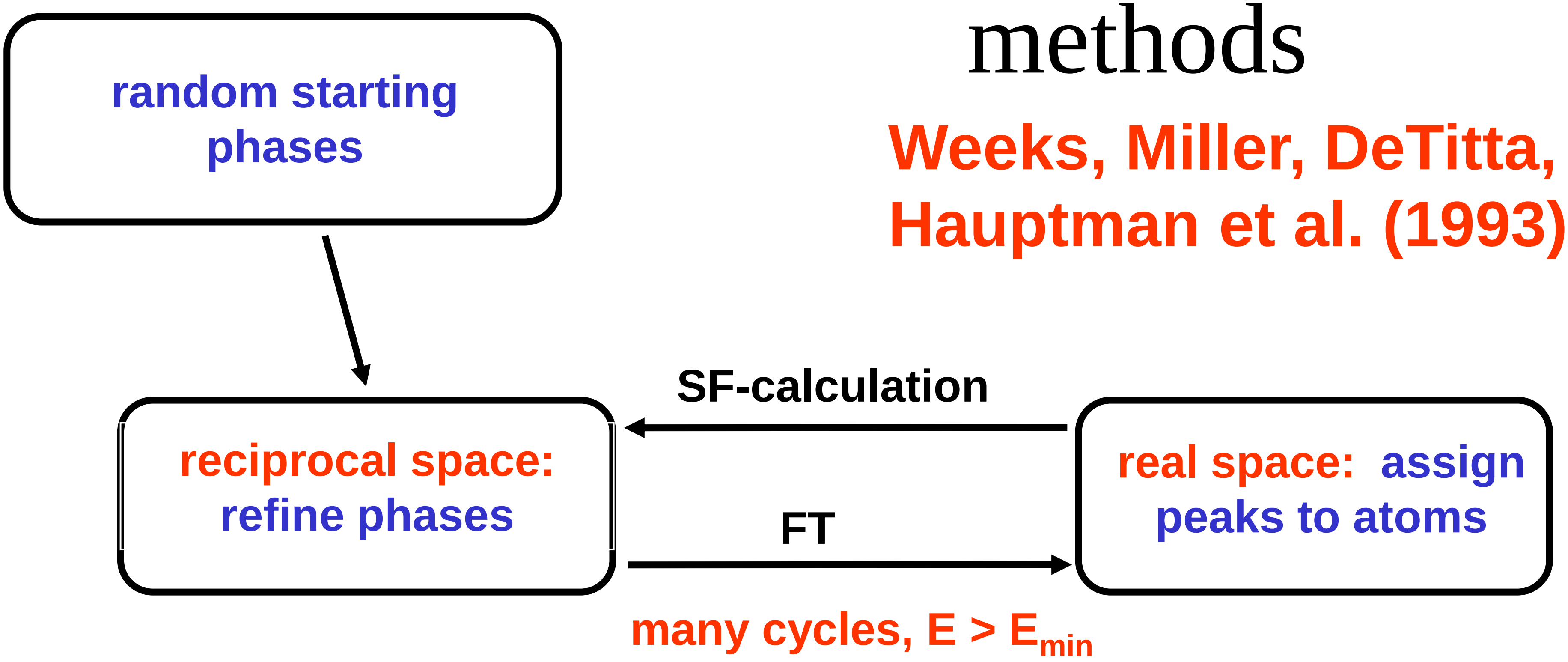
- The indicator used is the FOM –Figure of Merit.
- Its value depends on the program used.
- Need some familiarity with the individual program to have a feel for what good values are

Programs

- MULTAN –original program rarely used anymore
- SHELXS – part of SHELXTL –FOM should be as small as possible usually much less than one
- The Sir programs
  - SIR2002
  - SIR2004 Typically FOM is greater than 3 but since this package is so automatic user does not even notice.

Dual space methods

Weeks, Miller, DeTitta, Hauptman et al. (1993)



random starting phases

reciprocal space: refine phases

SF-calculation

FT

real space: assign peaks to atoms

many cycles, $E > E_{min}$

When the figures of merit indicate a good solution, it can be extended to the full structure by using all the data. Dual space methods were first implemented in the computer program *Shake and Bake*, and later in other programs. In this way it is possible to solve structures containing up to about 1000 independent atoms (not counting H). *Atomic resolution* is however still necessary.