RECIPROCAL SPACE METHODS SOLUTION

Rosanna Rizzi

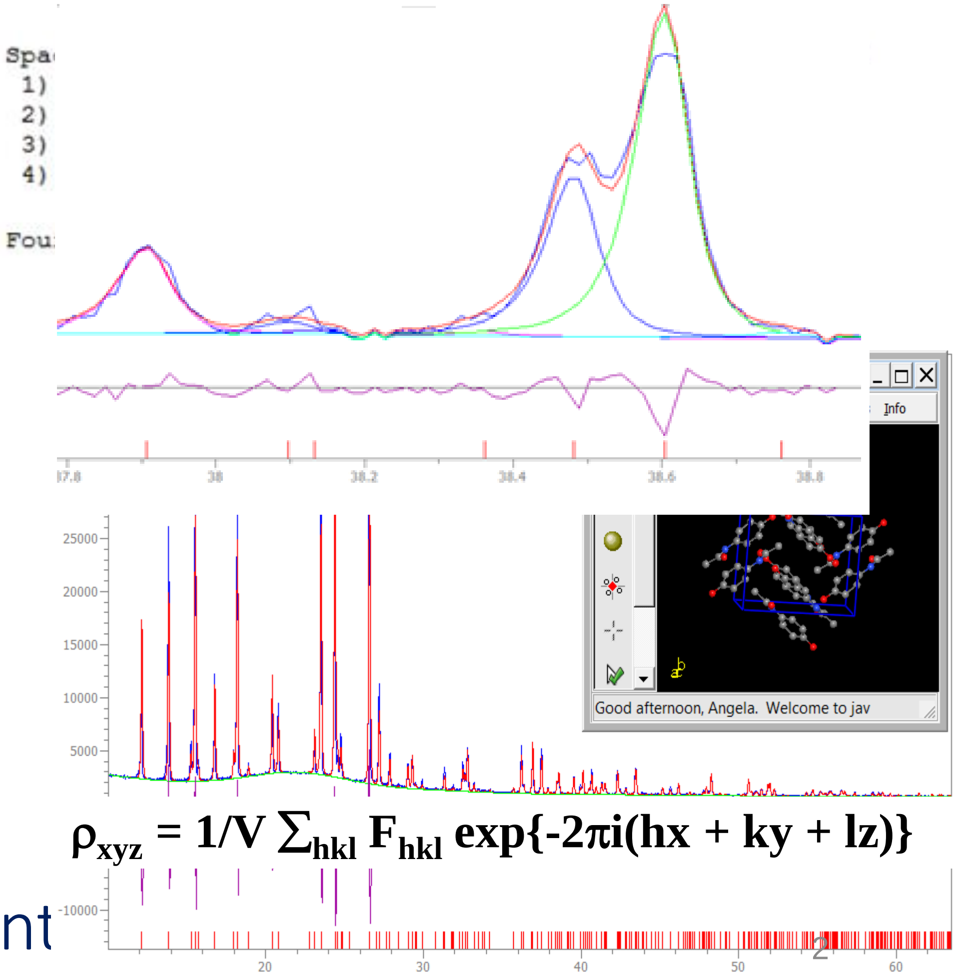
Institute of Crystallography (CNR-Bari)

rosanna.rizzi@ic.cnr.it

Structure determination in reciprocal space through powder diffraction is a sequential process with clearly defined stages:

Starting information: **chemical formula and experimental profile**

- Indexation
- Space group determination
- Profile decomposition and intensity extraction
- Structure solution
- Structure Rietveld refinement



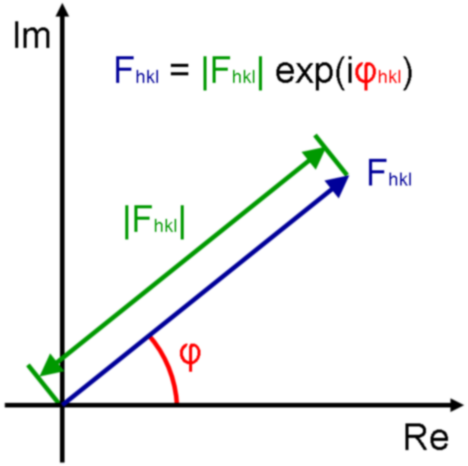
Good afternoon, Angela. Welcome to jav

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

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

Need to solve the equation above for all x, y, z in the unit cell

$$F_{hkl} = |F_{hkl}| \exp(i\varphi_{hkl})$$



What do we know?

$|F_{hkl}|$ ⟶ collected in the X-ray experiment. The measured intensities are proportional to the amplitudes of the coefficients used in the electron density equation ($I_{hkl} \propto |F_{hkl}|^2$)

What we still need?

φ_{hkl} ⟶ Not collected in the X-ray experiment. Hidden among the measured $|F_{hkl}|$ ⟶ **The phase problem**

Unfortunately the phases contain the bulk of the structural information

Karle amplitudes with Karle phases

Hauptman amplitudes with Hauptman phases

Karle amplitudes with Hauptman phases

Hauptman amplitudes with Karle phases

Phases dominate the image!

This is also why incorrect phases can cause big problems

Main steps of the Reciprocal Space approach for solving powder crystal structures

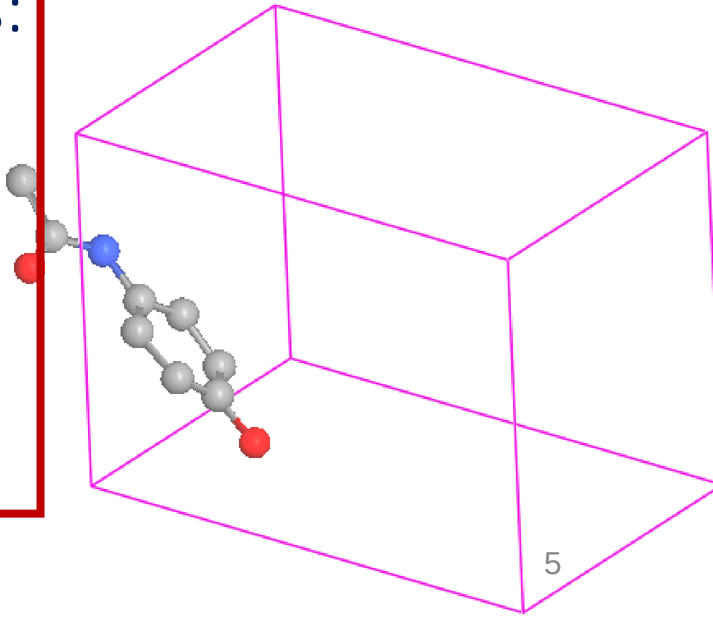
Starting information: **chemical formula**
experimental profile

- Indexation
- Space group determination
- Profile decomposition and intensity extraction

Structure solution by *ab initio* methods:

- Direct Methods
- Patterson method
- Maximum Entropy
- Charge Flipping

$\left.\right\}$ **Solving the phase problem**

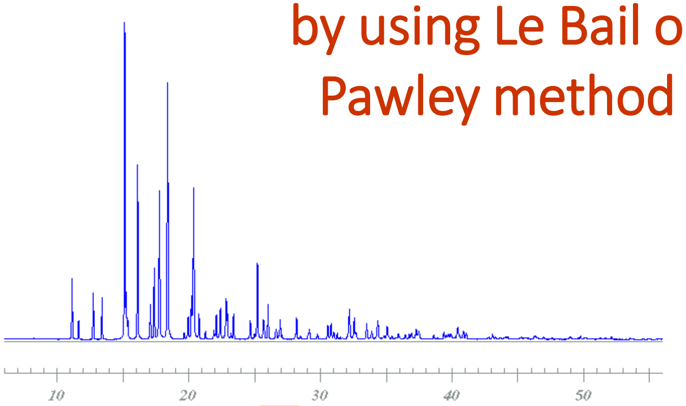


- Structure Rietveld refinement

Direct Methods

Statistical approach to the phase problem, work in the reciprocal space with a two step method

First step



by using Le Bail or Pawley method

the experimental pattern is decomposed into single integrated intensities $I_{hkl} \propto |F_{hkl}|^2$ and a list of structure factor moduli $|F_{hkl}|$ associated to each hkl reflection, is obtained

Second step

By using the experimental moduli

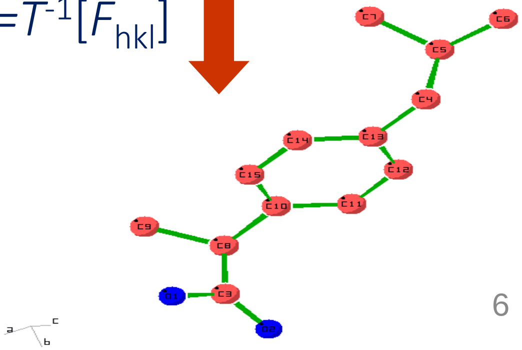
$$I_{hkl} \propto |F_{hkl}|^2$$

Direct Methods

are applied to calculate the phases φ_h of structure factors

$$F_{hkl}=|F_{hkl}|\exp(i\varphi_{hkl})$$

$$\rho=T^{-1}[F_{hkl}]$$



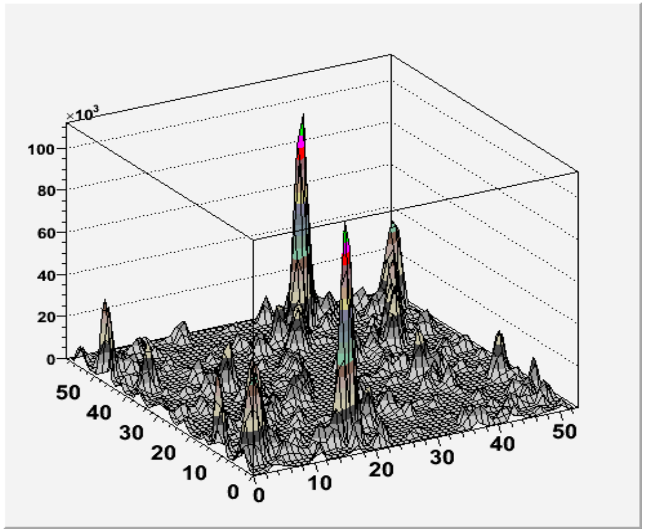
Second Step:

solution of the phase problem

determination of the reflection phases directly from structure factor magnitudes

$$I_h \longrightarrow |F_h| \Longrightarrow \varphi_h$$

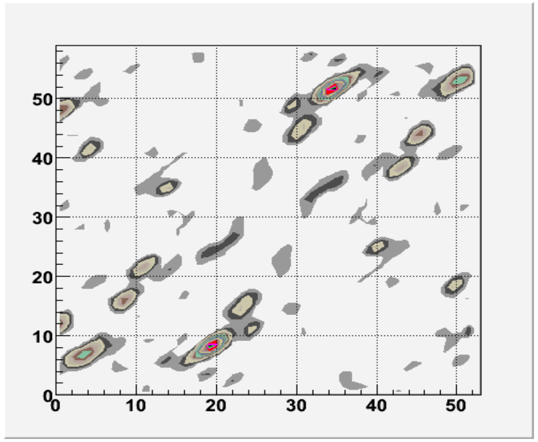
by using the following assumption on ρ(r):



1) **Positivity:** in the unit cell the electron density can never be less than 0

$$\rho(\mathbf{r}) \geq 0 \;\Rightarrow\; f > 0$$

2) **Atomicity:**
the electrons are concentrated around the nuclei and not dispersed in the unit cell



3) **Uniform distribution of the atoms in the unit cell**

Some considerations

This information, apparently trivial, is very useful to succeed in all the steps of a modern Direct Methods procedure:

1) Scaling of the observed intensities and normalization of the structure factors

2) Estimate of the structure invariants

3) Application of the tangent formula

4) Crystal structure complation and refinement

Some considerations

experimental pattern $\longrightarrow$ $I_h = \boxed{k} |F_h|^2$ the experimental structure factors are measured on a relative scale

positivity and atomicity of the electron density map $\longrightarrow$

$$<|F_{\mathbf{h}}|^2>_\theta = \sum_{j=1}^{N} f_j^2(\theta)$$

$\downarrow$

resolution dependent (f_j varies with ϑ) $\longrightarrow$ reflections measured at different θ values can not be compared directly

temperature dependent:

$$f_j = f_j^0 \exp(-\boxed{B_j} \sin^2\theta / \lambda^2)$$

The k = scale factor and B = average isotropic thermal parameter is determined by using the Wilson method

EXPO:
DIRECT METHODS IN ACTION

%……

%normalization (scaling of the observed intensities and normalization of the experimental structure factor moduli)

%invariants

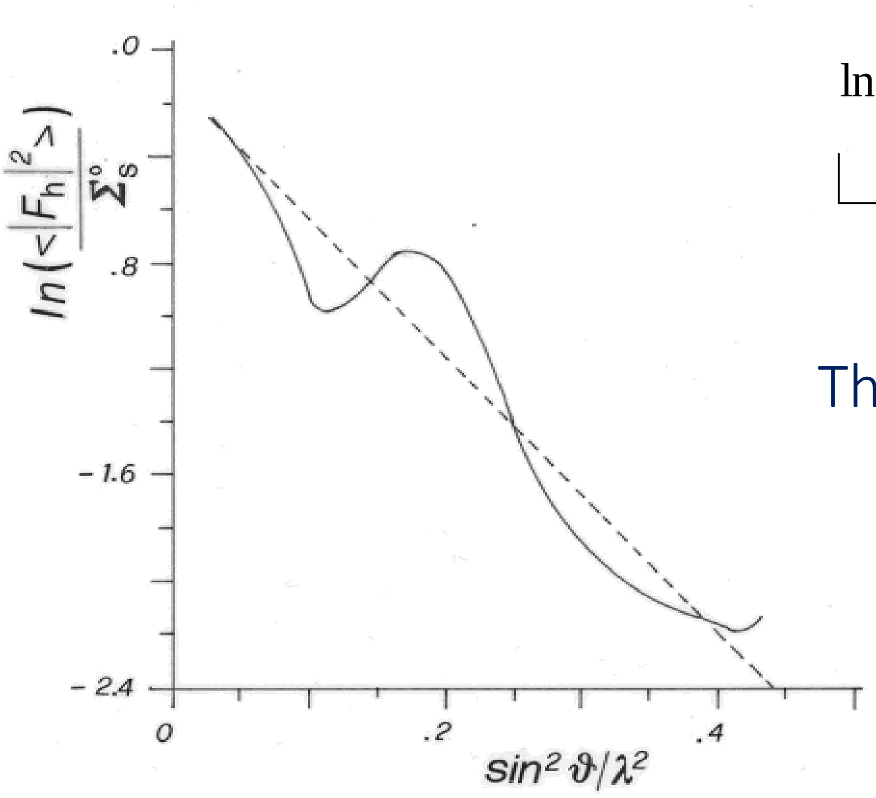
%phase

%fourier

%continue

THE WILSON PLOT

Determination of the unknown constants K and B using the extracted intensities



$$\underbrace{\ln\left(\frac{<|F_h|^2_{obs}>}{\Sigma_s^0}\right)}_{y} = \ln K - 2B\underbrace{\sin^2\theta/\lambda^2}_{x}$$

The plot of $\ln\left(\frac{<|F_h|^2_{obs}>}{\Sigma_s^0}\right)$ *vs.* $\sin^2\theta/\lambda^2$

represents a straight line:

slope: 2B

intercept: ln[k]

Direct Methods turn out to be more effective if the observed structure factors are modified to take out the effects of the electron density distribution in an atom and its thermal motion

Normalized structure factors

$$E_{\mathbf{h}} = F_{\mathbf{h}} / (\sum_{j=1}^{N} f_j^2)^{1/2}$$

$$|E_h| = \left\{ \frac{|F_h|^2_{\mathrm{obs}}}{\mathrm{K} \exp\{-2\mathrm{Bs}^2\} \varepsilon \sum_{\mathrm{j=1}}^{\mathrm{N}} {}^{o} f_j^2} \right\}^{1/2}$$

They represent an ideal point atom structure because no θ dependent factors are present

Some considerations

This information, apparently trivial, is very useful to succeed in all the steps of a modern Direct Methods procedure:

1) Scaling of the observed intensities and normalization of the structure factors

2) Estimate of the structure invariants

3) Application of the tangent formula

4) Crystal structure complation and refinement

EXPO:
DIRECT METHODS IN ACTION

%……

%normalization

%invariants (determination of the structure invariants)

%phase

%four

%continue

STRUCTURAL INFORMATION

$$|F_h|^2 = \Sigma f_j^2 + \Sigma f_i f_j \,\mathrm{exp}[2\pi \mathrm{ih}(\mathrm{r_i}-\mathrm{r_j})]$$

Observed amplitudes → Interatomic vectors

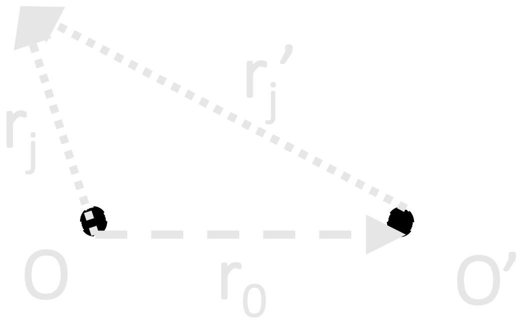
Independent on the origin chosen by the user

Structure factor phases → Image of the structure

Dependent on the choice of origin

STRUCTURAL INFORMATION

$$r_j = r_0 + r_j'$$



$$F_{\boldsymbol{h}} = \sum_{j=1}^{N} f_j \exp\left(2\pi i \boldsymbol{h}\boldsymbol{r}_j\right) = \sum_{j=1}^{N} f_j \exp\left(2\pi i \boldsymbol{h}(\boldsymbol{r}_0 + \boldsymbol{r}_j')\right)$$

$$= \exp\left(2\pi i \boldsymbol{h}\boldsymbol{r}_0\right) \sum_{j=1}^{N} f_j \exp\left(2\pi i \boldsymbol{h}\boldsymbol{r}_j'\right) = \exp\left(2\pi i \boldsymbol{h}\boldsymbol{r}_0\right) F_{\boldsymbol{h}}'$$

$$F_h \exp\left(-2\pi i \boldsymbol{h}\boldsymbol{r}_0\right) = F_{\boldsymbol{h}}'$$

$$\left|F_{\boldsymbol{h}}\right| \exp i\left(\phi_h - 2\pi \boldsymbol{h}\boldsymbol{r}_0\right) = \left| F_{\boldsymbol{h}}' \right| \exp i(\phi_h')$$

$$\left\{\begin{array}{l} |F_h'| = |F_h| \longrightarrow \text{structure invariants} \\ \phi_h' = \phi_h - 2\pi h r_0 \longrightarrow \text{not structure invariants} \end{array}\right.$$

The phase values depend on the origin chosen by the user: not structure invariants

The moduli are independent on the origin chosen by the user: structure invariants

Can we derive the phases from the diffraction moduli ?

Evidently, from the observed moduli is possible to derive information **only** on **single phases or linear combinations of phases which are independent on the choice of origin,** which are **structure invariants**

STRUCTURE INVARIANTS ARE

Products of structure factors satisfying the condition that the sum of the indices is zero

↓

The simplest invariant: triplet invariant $F_h F_k F_{-h-k}$

Owing to the shift of origin r_0, the single phase ϕ_h changes according to:

$$\phi_h{}' = \phi_h - 2\pi \mathbf{h} \mathbf{r}_0$$

But the product calculated respect to the new origin, does not change

$$F'_{\boldsymbol{h}} F'_{\boldsymbol{k}} F'_{-\boldsymbol{h}-\boldsymbol{k}} = |F_{\boldsymbol{h}}| \exp i\,(\phi_{\boldsymbol{h}} - 2\pi \boldsymbol{h}\boldsymbol{r}_0)\,|F_{\boldsymbol{k}}| \exp i\,(\phi_{\boldsymbol{k}} - 2\pi \boldsymbol{k}\boldsymbol{r}_0)$$
$$|F_{-\boldsymbol{h}-\boldsymbol{k}}| \exp i[\phi_{-\boldsymbol{h}-\boldsymbol{k}} + 2\pi(\boldsymbol{h}+\boldsymbol{k})\boldsymbol{r}_0]$$

The sum ($\varphi_h + \varphi_k + \varphi_{-h-k}$) is called **triplet phase invariant**

STRUCTURE INVARIANTS: EXAMPLES

doublet invariant : $F_h\, F_{-h} = |F_h|^2$

triplet invariant : $F_h\, F_k\, F_{-h-k}$

quartet invariant : $F_h\, F_k\, F_l\, F_{-h-k-l}$

quintet invariant : $F_h\, F_k\, F_l\, F_m\, F_{-h-k-l-m}$

…………………

Some considerations

This information, apparently trivial, is very useful to succeed in all the steps of a modern Direct Methods procedure:

1) Scaling of the observed intensities and normalization of the structure factors

2) Estimate of the structure invariants

3) Application of the tangent formula

4) Crystal structure complation and refinement

EXPO:
DIRECT METHODS IN ACTION

%……

%normalization

%invariants

%phase (phases estimation)

%fourier

%continue

DIRECT METHODS PROCEDURE

$I_h \rightarrow |F_h| \rightarrow |E_h|$

Selection of the reflections with $|E_h| > 1.0$ (N_{large})

Setting up triplet relationships among N_{large} reflections

To apply the phasing procedure we need to know at least one pair ($\varphi_k + \varphi_{h-k}$); much better if we know more pairs ($\varphi_k + \varphi_{h-k}$)

Application of the Tangent formula to calculate and refine the phases

DIRECT METHODS IN PRACTICE

The phase determination process usually leads to more than one trial solution. Obviously most are not correct

FOM calculation → Allows an a priori estimate of the goodness of each phase set

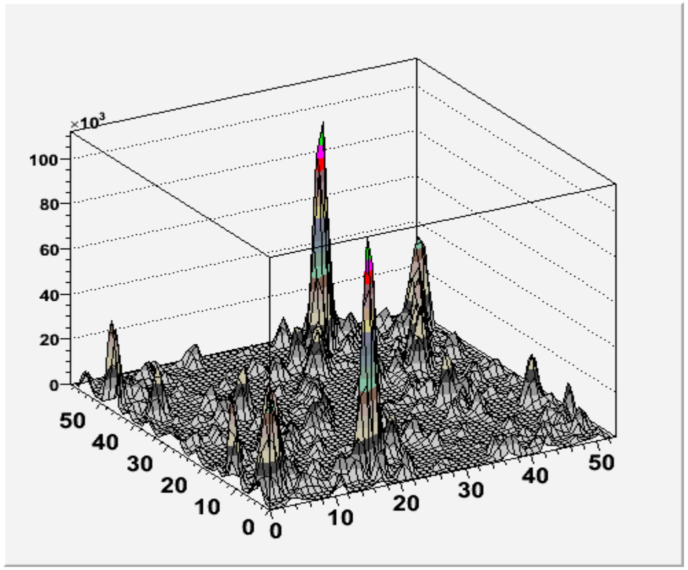
List of sorted solutions and selection of the most reliable one

$\{|E|, \varphi\}$

Electron density map calculation

Rietveld refinement process

$$\rho(r)=T^{-1}[E_h]$$



Automatic interpretation

EXPO:
DIRECT METHODS IN ACTION

%……

%normalization

%invariants

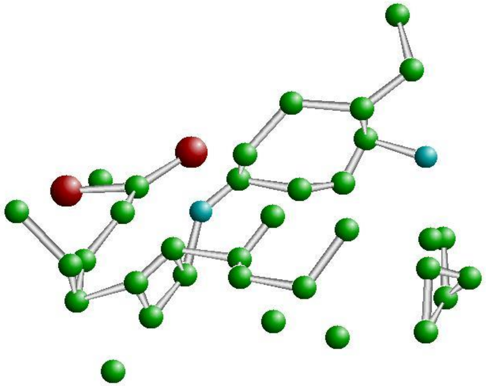
%phase

%fourier (E-map calculation)

%continue

AT THE END OF DIRECT METHODS……

E-map: generally the obtained structure model is incomplete and distorted, and represents only a rough approximation of the real structure



The final model can be uninterpretable

↓

before the final Rietveld refinement

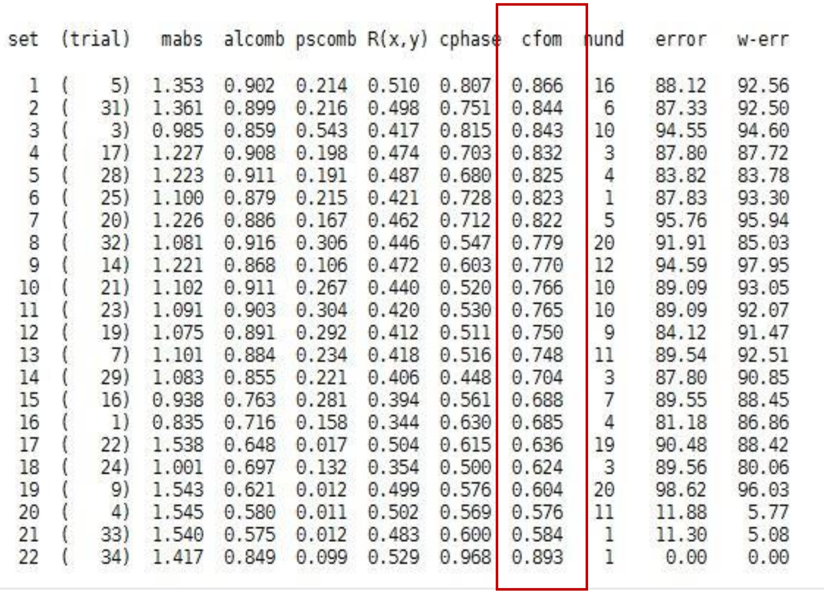
necessary to complete and refine the structure

→ MODEL OPTIMIZATION

At the end of Direct Methods procedure, the most plausible sets of phases are ranked according to *CFOM*

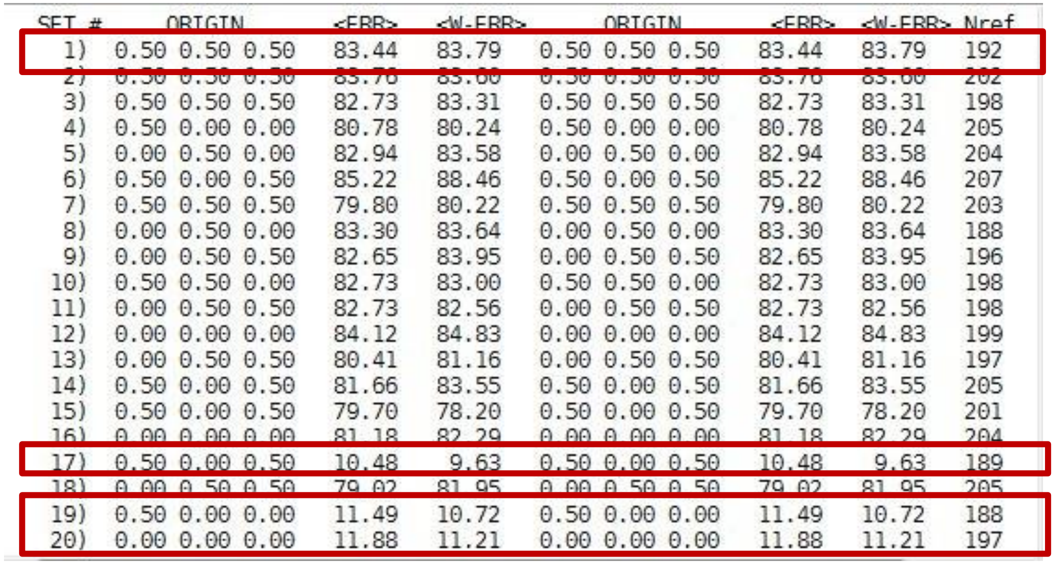
CFOM exploits the integrated intensities values ➡

it can be meaningless and the correct set of Direct Methods phases could not be the first one in the list



```
set  (trial)   mabs  alcomb pscomb R(x,y) cphase  cfom  nund    error    w-err

  1  (     5)  1.353  0.902  0.214  0.510  0.807  0.866  16     88.12    92.56
  2  (    31)  1.361  0.899  0.216  0.498  0.751  0.844   6     87.33    92.50
  3  (     3)  0.985  0.859  0.543  0.417  0.815  0.843  10     94.55    94.60
  4  (    17)  1.227  0.908  0.198  0.474  0.703  0.832   3     87.80    87.72
  5  (    28)  1.223  0.911  0.191  0.487  0.680  0.825   4     83.82    83.78
  6  (    25)  1.100  0.879  0.215  0.421  0.728  0.823   1     87.83    93.30
  7  (    20)  1.226  0.886  0.167  0.462  0.712  0.822   5     95.76    95.94
  8  (    32)  1.081  0.916  0.306  0.446  0.547  0.779  20     91.91    85.03
  9  (    14)  1.221  0.868  0.106  0.472  0.603  0.770  12     94.59    97.95
 10  (    21)  1.102  0.911  0.267  0.440  0.520  0.766  10     89.09    93.05
 11  (    23)  1.091  0.903  0.304  0.420  0.530  0.765  10     89.09    92.07
 12  (    19)  1.075  0.891  0.292  0.412  0.511  0.750   9     84.12    91.47
 13  (     7)  1.101  0.884  0.234  0.418  0.516  0.748  11     89.54    92.51
 14  (    29)  1.083  0.855  0.221  0.406  0.448  0.704   3     87.80    90.85
 15  (    16)  0.938  0.763  0.281  0.394  0.561  0.688   7     89.55    88.45
 16  (     1)  0.835  0.716  0.158  0.344  0.630  0.685   4     81.18    86.86
 17  (    22)  1.538  0.648  0.017  0.504  0.615  0.636  19     90.48    88.42
 18  (    24)  1.001  0.697  0.132  0.354  0.500  0.624   3     89.56    80.06
 19  (     9)  1.543  0.621  0.012  0.499  0.576  0.604  20     98.62    96.03
 20  (     4)  1.545  0.580  0.011  0.502  0.569  0.576  11     11.88     5.77
 21  (    33)  1.540  0.575  0.012  0.483  0.600  0.584   1     11.30     5.08
 22  (    34)  1.417  0.849  0.099  0.529  0.968  0.893   1      0.00     0.00
```



```
  SET #        ORIGIN          <ERR>    <W-ERR>        ORIGIN          <ERR>    <W-ERR>  Nref
    1)   0.50 0.50 0.50    83.44    83.79    0.50 0.50 0.50    83.44    83.79    192
    2)   0.50 0.50 0.50    83.76    83.60    0.50 0.50 0.50    83.76    83.60    202
    3)   0.50 0.50 0.50    82.73    83.31    0.50 0.50 0.50    82.73    83.31    198
    4)   0.50 0.00 0.00    80.78    80.24    0.50 0.00 0.00    80.78    80.24    205
    5)   0.00 0.50 0.00    82.94    83.58    0.00 0.50 0.00    82.94    83.58    204
    6)   0.50 0.00 0.50    85.22    88.46    0.50 0.00 0.50    85.22    88.46    207
    7)   0.50 0.50 0.50    79.80    80.22    0.50 0.50 0.50    79.80    80.22    203
    8)   0.00 0.50 0.00    83.30    83.64    0.00 0.50 0.00    83.30    83.64    188
    9)   0.00 0.50 0.50    82.65    83.95    0.00 0.50 0.50    82.65    83.95    196
   10)   0.50 0.50 0.00    82.73    83.00    0.50 0.50 0.00    82.73    83.00    198
   11)   0.00 0.50 0.50    82.73    82.56    0.00 0.50 0.50    82.73    82.56    198
   12)   0.00 0.00 0.00    84.12    84.83    0.00 0.00 0.00    84.12    84.83    199
   13)   0.00 0.50 0.50    80.41    81.16    0.00 0.50 0.50    80.41    81.16    197
   14)   0.50 0.00 0.50    81.66    83.55    0.50 0.00 0.50    81.66    83.55    205
   15)   0.50 0.00 0.50    79.70    78.20    0.50 0.00 0.50    79.70    78.20    201
   16)   0.00 0.00 0.00    81.18    82.29    0.00 0.00 0.00    81.18    82.29    204
   17)   0.50 0.00 0.50    10.48     9.63    0.50 0.00 0.50    10.48     9.63    189
   18)   0.00 0.50 0.50    79.02    81.95    0.00 0.50 0.50    79.02    81.95    205
   19)   0.50 0.00 0.00    11.49    10.72    0.50 0.00 0.00    11.49    10.72    188
   20)   0.00 0.00 0.00    11.88    11.21    0.00 0.00 0.00    11.88    11.21    197
```

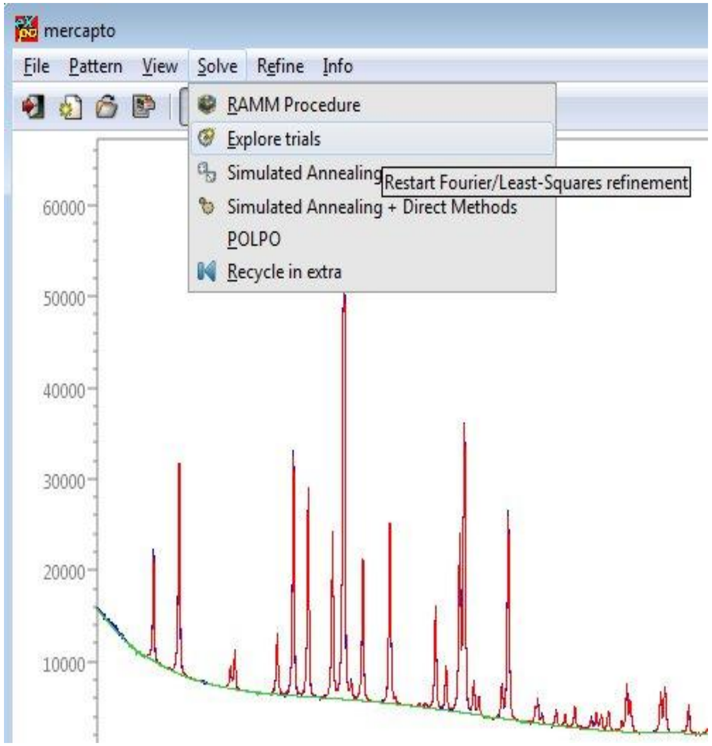
No solution on set n. 1

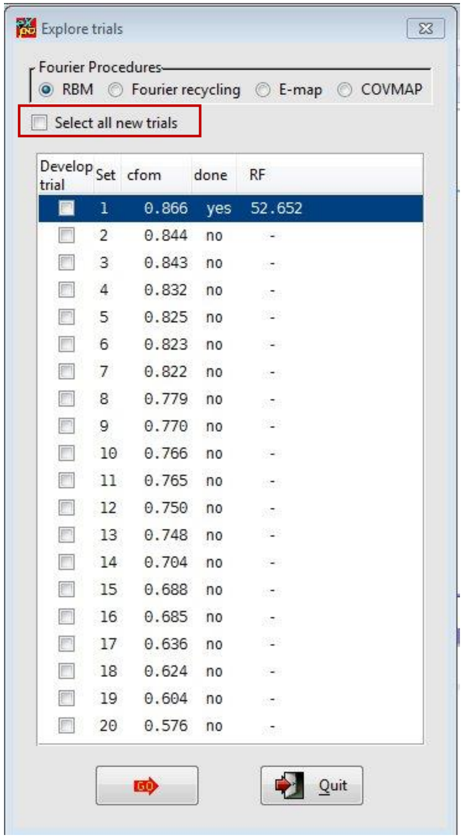
3 Solutions on sets n. 17, 19 and 20

ALLTRIALS PROCEDURE

When no solution is found by using the default choice of *EXPO*
(set of phases n. 1)

The exploration of all the 20 stored trials can automatically executed by *EXPO* through the non-default *ALLTRIALS* procedure. It automatically carries out, for each of the stored trials, the structure solution process

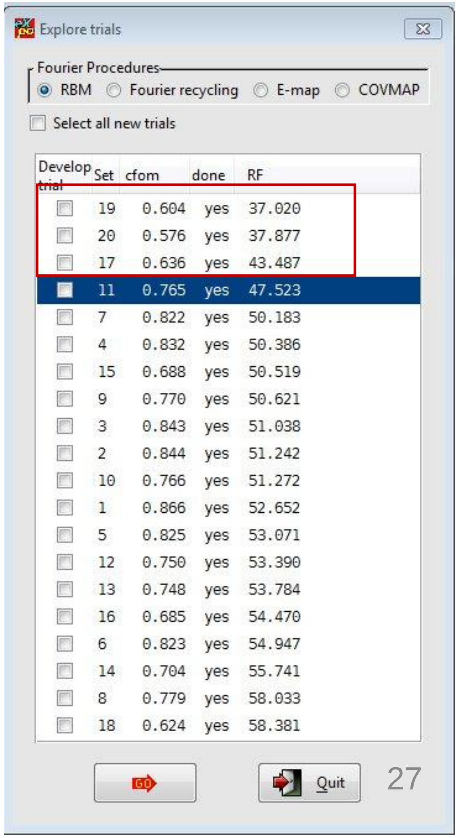




Fourier Procedures
● RBM ○ Fourier recycling ○ E-map ○ COVMAP

☐ Select all new trials

Develop trial	Set	cfom	done	RF
☐	1	0.866	yes	52.652
☐	2	0.844	no	-
☐	3	0.843	no	-
☐	4	0.832	no	-
☐	5	0.825	no	-
☐	6	0.823	no	-
☐	7	0.822	no	-
☐	8	0.779	no	-
☐	9	0.770	no	-
☐	10	0.766	no	-
☐	11	0.765	no	-
☐	12	0.750	no	-
☐	13	0.748	no	-
☐	14	0.704	no	-
☐	15	0.688	no	-
☐	16	0.685	no	-
☐	17	0.636	no	-
☐	18	0.624	no	-
☐	19	0.604	no	-
☐	20	0.576	no	-



Fourier Procedures
● RBM ○ Fourier recycling ○ E-map ○ COVMAP

☐ Select all new trials

Develop trial	Set	cfom	done	RF
☐	19	0.604	yes	37.020
☐	20	0.576	yes	37.877
☐	17	0.636	yes	43.487
☐	11	0.765	yes	47.523
☐	7	0.822	yes	50.183
☐	4	0.832	yes	50.386
☐	15	0.688	yes	50.519
☐	9	0.770	yes	50.621
☐	3	0.843	yes	51.038
☐	2	0.844	yes	51.242
☐	10	0.766	yes	51.272
☐	1	0.866	yes	52.652
☐	5	0.825	yes	53.071
☐	12	0.750	yes	53.390
☐	13	0.748	yes	53.784
☐	16	0.685	yes	54.470
☐	6	0.823	yes	54.947
☐	14	0.704	yes	55.741
☐	8	0.779	yes	58.033
☐	18	0.624	yes	58.381

Main steps of the Reciprocal Space approach for solving powder crystal structures

Starting information: **chemical formula**

experimental profile

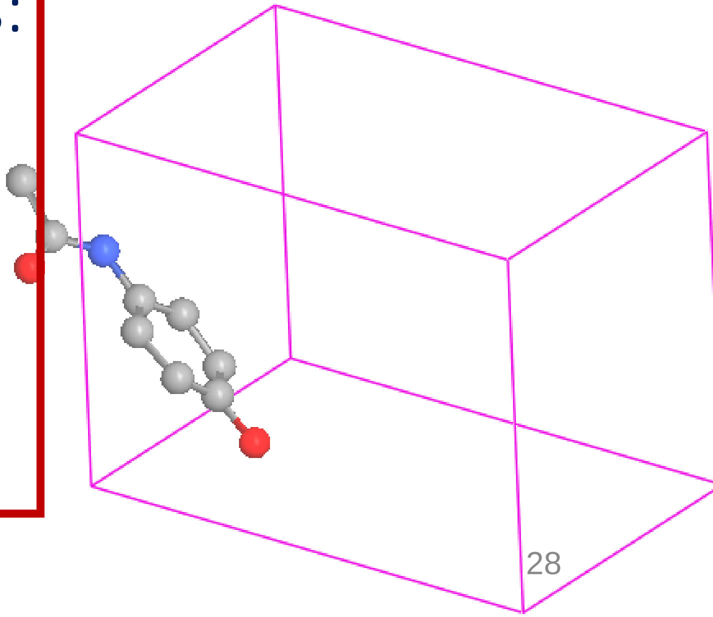
- Indexation
- Space group determination
- Profile decomposition and intensity extraction

Structure solution by *ab initio* methods:

- Direct Methods
- Patterson method
- Maximum Entropy
- Charge Flipping

Solving the phase problem

- Structure Rietveld refinement



PATTERSON METHOD

There are three very important points about the Patterson method:

- The Patterson map can be calculated from the diffraction data *without* knowing the phases

$$P(\mathrm{u})=T^{-1}[\,|\mathrm{F}(\mathrm{r})|^2]$$

- Requires the presence of a heavy atom in the structure, e.g. Fe, Cl, S, etc

- The Patterson is a **vector map** of the structure

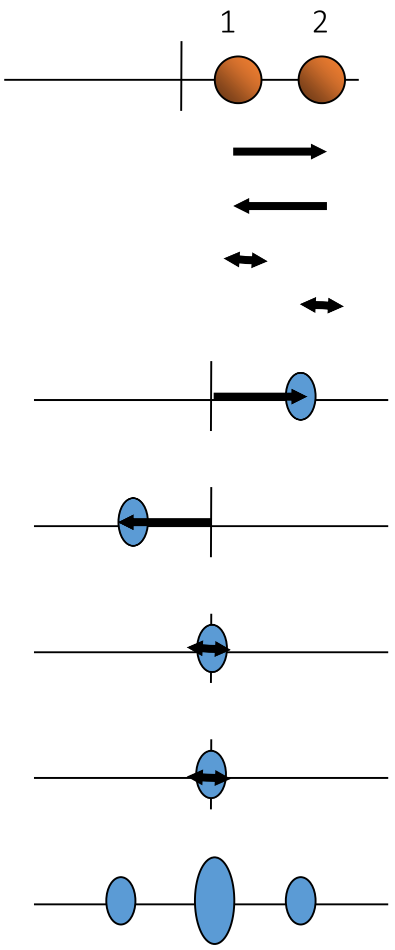
↓

We can get the relative (not absolute) positions of atoms with respect to each other

THE VECTOR MAP OF TWO ATOMS

What is the complete set of vectors between two atoms?

There are four vectors, two equal and opposite interatomic vectors and two self vectors



Vector atom 1 to atom 2

Vector atom 2 to atom 1

Vector atom 1 to atom 1

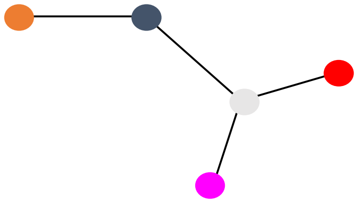
Vector atom 2 to atom 2

Vector map

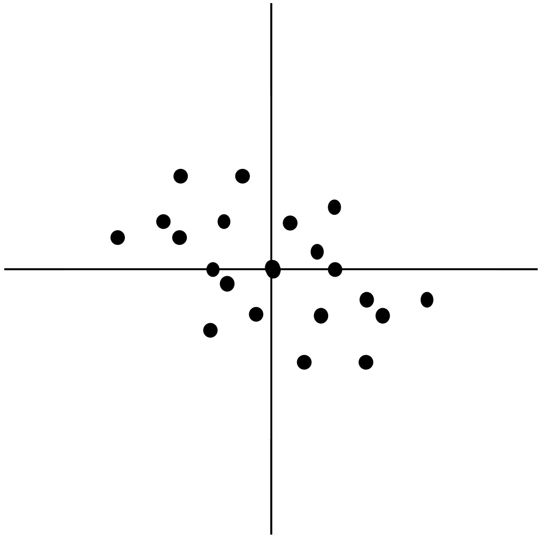
The vector map has a large peak at the origin and two lower peaks on either side of it, separated from the origin by the distance between the two atoms

THE VECTOR MAP OF SOME ATOMS

We can generate a vector map of a molecule by putting each atom in succession at the origin

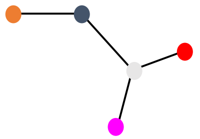


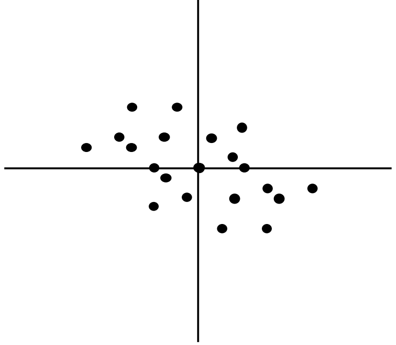
molecule



Patterson map

PATTERSON MAP





The Patterson map of N atoms contains N^2 peaks:

- N peaks at the origin, representing the vector between each atom and itself
- *N*(*N*-1) peaks within the unit cell

The map can be very congested and difficult to interpret. In addition, the Patterson map is centrosymmetric, which further complicates cells for non-centrosymmetric lattices.

The intensity of each peak is proportional to the product of the atomic numbers of the two atoms to which the vector refers

Patterson function particularly effective for the identification of the position of heavy atoms when most of the other atoms are light

PATTERSON METHOD

Patterson methods are often the first choice for the solution of structures containing a few heavy atoms

If they have a sufficiently high atomic number, they can be used as a good initial model to which one can apply the so-called *Method of Fourier Recycling*, to obtain the light atom positions and then to recover the complete structure

EXPO:
PATTERSON METHOD IN ACTION

%……

%normalization

%patterson

inverse →

%continue

To activate the procedure of calculating the |F| values from an inverted suitably modified Patterson map

The values are then used as starting point in the Le Bail algorithm for extracting new structure factor moduli from the experimental pattern

Main steps of the Reciprocal Space approach for solving powder crystal structures

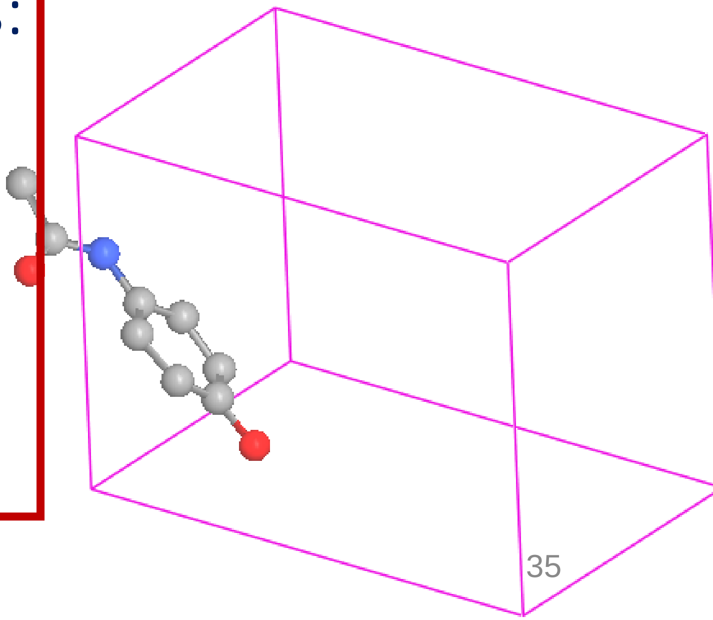
Starting information: **chemical formula**
experimental profile

- Indexation
- Space group determination
- Profile decomposition and intensity extraction

Structure solution by *ab initio* methods:

- Direct Methods
- Patterson method
- Maximum Entropy
- Charge Flipping

Solving the phase problem



- Structure Rietveld refinement

CHARGE FLIPPING: THE PRINCIPLE

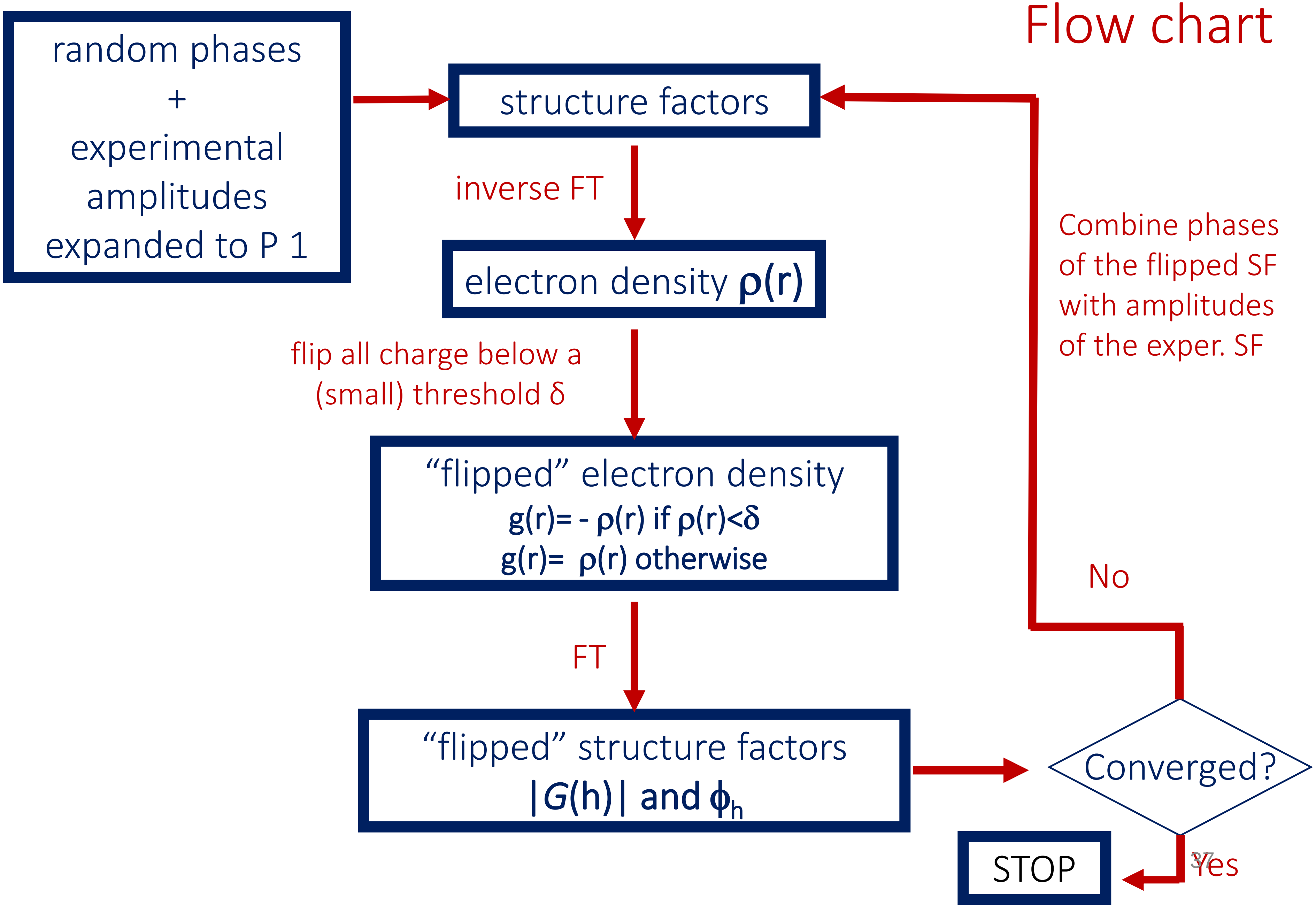
Proposed for single crystal data, has been also adapted to powder diffraction data

It is an *ab initio* dual-space iterative method switching back and forth between real and reciprocal spaces for the determination of an approximate electron density

Needs only lattice parameters and the set of the experimental structure factors moduli; neither chemical composition nor symmetry information is required. The structure is solved in *P*1

CHARGE FLIPPING: THE PRINCIPLE

Flow chart



random phases
+
experimental amplitudes
expanded to P 1

structure factors

inverse FT

electron density **$\rho(r)$**

flip all charge below a (small) threshold δ

"flipped" electron density

$g(r) = -\rho(r)$ if $\rho(r) < \delta$

$g(r) = \rho(r)$ otherwise

FT

"flipped" structure factors

$|G(h)|$ and ϕ_h

Converged?

No

Combine phases of the flipped SF with amplitudes of the exper. SF

Yes

STOP

CHARGE FLIPPING

Charge flipping reconstructs the density always in P1 where the maxima can appear anywhere in the cell

In higher symmetry the choice is limited with a lower effectivity

Advantages:

- No need to know the symmetry, which can be read out from the result
- It is extremely simple and easy to implement

Disadvantages:

- The structure is randomly shifted in the cell: it is necessary to locate the origin of the space group
- suffers from the inaccuracy of the extracted reflection intensities

CHARGE FLIPPING METHOD

It requires atomic resolution to work: better than 1.2 Å for organic compounds, 1.6 Å for heavy atom structure

The problem has been faced by introducing in the algorithm some modifications like the histogram matching loop

CONCLUSION

Crystal structure solution by powder diffraction data is not easy.

Great experimental, methodological and computing progress has been reached.

In spite of that, powder solution is still a challenge in many cases.

The effort is to make solution by powder comparable with the single crystal case: *sophisticated software in which, constantly evolving theories are implemented, are on the market*

crystals

IMPACT FACTOR 2.061

Special Issue

Crystal Structure Characterization by Powder Diffraction

Guest Editor:

Dr. Angela Altomare

Dr. Rosanna Rizzi

Institute of Crystallography, National
Research Council-CNR, Bari, Italy

E-mail: powder@ic.cnr.it

Submission Deadline: 31 December 2019

We invite contribute of papers that, while discussing the followed computational, methodological, and/or experimental strategies, point out the essential and advanced contribution of powder diffraction in identifying the unknown crystal structure of a compound.

MDPI
Academic Open Access Publishing
since 1996

Crystals Editorial Office
MDPI AG
St. Alban-Anlage 66
4052 Basel, Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18
www.mdpi.com
crystals@mdpi.com

THANK YOU FOR YOUR ATTENTION