Pepinsky's Machine

Version α, Release 0.5

Nicholas M Glykos, 2001

If you ever need to reference this program in your publications *(but note that you are not required to do so)*, please use the following citation : Glykos, N.M. (1999),'Pepinsky's Machine : an interactive, graphic-based Fourier synthesis program withapplications in teaching and research.", *J. Appl. Cryst.*,Vol. **32**, 821–823.

Please send comments, suggestions and bug reports to `glykos@crystal2.imbb.forth.gr`
This document was prepared with LaTeX2e. The graphs were prepared either with locally written programs (*ps* or *Conv2D*), or with *pluto* and *pltdev* from the CCP4 suite of programs.

The latest version of the program is available from the following www address :
`http://origin.imbb.forth.gr/~glykos/`

Contents

1 Recent additions

1.1 Version α, Release 0.5

- Fixed a memory-leak problem.
- Support for solaris added.
- Limited support for (macromolecular) CCP4 `.mtz` files added. The major limitation is that only the following space groups are recognised by the current version of the program : $P2, P2_1, C2, P222, C222_1, P2_12_12_1, P2_12_12, P3_221$.
- Tidy-up and update documentation in the hope that this will be a long-lived release.

1.2 Version α, Release 0.4

The program now uses the FFTW library to perform the FFTs. This has four significant consequences :

- I can now distribute the source code, and so, the program is indeed free software.
- Each time the program encounters a new grid size for the FFT, there will be a short pause while FFTW is learning how to best perform the given FFT. This is only done once and the results are saved permanently in a hidden ('.FFTW_wisdom') file in the user's home directory.
- FFTW is not limited to grid sizes that are a power of 2, and is thus more efficient for large transforms.
- FFTW is approximately 10 times faster than the previously adopted (radix-2) algorithm (but it does not matter because the limiting step is contouring the map and not doing the FFT).

1.3 Version α, Release 0.3

Thanks to Ygl, the program now runs on Linux and OSF platforms (in addition to the Irix platforms). Pre-compiled binaries are now included for all three architectures.

1.4 Version α, Release 0.2

- Add support for two extra command-line flags to allow for different contouring of Patterson and superposition functions ([Patterson] — [2]).
- Fix a number of p2-specific bugs : the Fourier synthesis plot had wrong proportions for gamma angles significantly different from 90 degrees, the fractional coordinates reported were wrong, etc.
- Pepinsky will now write out files that can be fed to GraphEnt for calculation of a MaxEnt map.
- Phase combinations can now be stored in (and restored from) memory by assigning them to the function keys (F1-F12). This should become the standard way to select (and work with) promissing phase combinations from the permutation mode.

Pepinsky's Machine.

User's guide and a discussion of the example files.

This document contains a quick guide to the "Pepinsky's Machine" program and a discussion of the example data files distributed with it. For a general description of the program, you can print the postscript file `Paper.ps` in the `/Doc` directory of the distribution (this paper has been published in *J. Appl. Cryst.*). Pressing the H key from within the program will give you the on-line help page.

2 User's guide

2.1 The input file

All you need is an ASCII file containing cell parameters, plane group number and the list of reflections ($h, k, F, \sigma(F)$), as shown below :

```
CELL     81.40  30.40    90.00
SYMMETRY 7
   2    0     237.10      3.30
   4    0     264.90      3.40
   6    0     346.80      4.90
                .....
```

Pepinsky's machine will only cooperate if the plane group is centrosymmetric : *p2, pmm, pmg, pgg, cmm, p4, p4m, p4g, p6, p6m*.

2.2 Possible space-group depended complications

The program uses the conventional plane group settings as given in the International Tables for X-ray Crystallography, 1952. This means that if you are using real data (indexed as required for the space group of your crystals), it may be neccessary to exchange and/or divide Miller indeces and the corresponding unit cell translations. To give an example : the plane group of the (010) projection of $C222_1$ is *pmg* but with the glide line along y (instead of x) and the a translation doubled (due to C-centering). You would have to exchange the h and k indeces (and the corresponding translations), and to divide h (and a) by 2. Actually, you could have leaved h (and a) doubled, but then the fractional coordinates along (the new) y given by the program should also be doubled. To refer the coordinates back to the conventional space group setting you will have to invert the procedure, *but taking also into account a possible origin shift* (the origin for all supported plane groups is on a centre of symmetry).

2.3 Starting the program

Give `pepinsky my_file.in` to run the program with the data set contained in the file `my_file.in`. You can also use just `pepinsky`, but in this case your data must be in a file named `pepinsky.in`.

2.4 The default mode

The program will start by plotting the amplitude-weighted reciprocal lattice, and, since no phase information is available when a new session starts, it will also plot the corresponding Patterson function (with contours every 4% of the origin peak). It then informs the user of the average and maximum observed amplitude, the size of the grid for the FFT and the number and parities of the origin-fixing reflections.

You can then place the graphics cursor on any of the displayed reflections and set its phase to be positive, negative or unknown by pressing the corresponding mouse buttons (see table below). The program will then update the phases of all symmetry-related reflections (depending on the plane group symmetry), re-draw the weighted reciprocal lattice representation (the reflections are colour-coded accordingly to their phase), and then calculate and plot the Fourier transform using only those reflections which have been assigned a trial phase. A series of graphical buttons (located below the weighted reciprocal lattice area, see table below) will allow you to increase or decrease the size (radius) of the reflections, the size of the density map, the number of contour levels, the size of the grid used for FFT, to clear all phase assignments and go back to the Patterson function, and, finally, to write out a postscript file containing the current map, weighted reciprocal lattice and phase assignments. While zooming is active, the electron density map can be moved with the arrow keys, thus allowing you (and once the origin is fixed), to concentrate on an area of interest. A summary of the commands available in the default mode is given below (this list is also available from within the program by pressing the H key) :

```
Press  <ESC>  to exit the program
Press <SPACE> to enter/exit permutation mode
Left   mouse button to make a reflection (+)
Right  mouse button to make a reflection (-)
Middle mouse button to un-phase a reflection
Use ++REFL-- to adjust the size of reflections
Use ++MAP--  to zoom in/out the map
Use ++CONT-- to increase/decrease contours
Use ++GRID-- to change grid size for FFT
Use PS       to write out postscript & the
             current phase set.
Use CLEAR    to discard all phase assignments

To change the first contour level, change the
value of F000 by clicking on it.

While zooming is active, move the map using
the arrow keys.

Click on the electron density map to get the
fractional coordinates corresponding to the
current position of the cursor.
```

The PS button produces not only a postscript file, but also an ASCII file containing $h, k, F, \sigma(F), \phi$, with $\sigma(F) = 0.0$ for reflections outside the (reciprocal) asymmetric unit. Also note that the postscript files can be quite large (depending on the grid size) and that the program will *not* overwrite previous versions of the output files.

2.5 The permutation syntheses mode

The program supports a second mode of operation, termed the "permutation syntheses mode". In this mode, the user selects (by clicking on them) a number of reflections that their phases will systematically be permuted (see the paper preprint for details), the program calculates all their unique sign combinations, combine these with the fixed phases (if any) and uses the resulting phase sets to calculate and plot the corresponding electron density maps. The permutation maps can either be shown one-by-one (with an option for the user to characterise and save the corresponding phase combination as "good", "bad" or "ignore"), or in a non-stop fashion for quickly estimating the clarity of the resulting maps.

To enter the permutation mode, press the space bar. The program will automatically give you the following help screen (which will also appear when pressing the H key and whilst you are in the permutation mode) :

```
You have entered the permutation mode :

Select the reflections to be systematically
permuted by clicking the middle mouse button
on them. Finish by pressing the space bar
again. You will then be asked to chose
between a non-stop display of all maps, or a
one-by-one examination of the permutations.
In the one-by-one mode you can mark the maps
you see as 'good' or 'bad'. When all maps have
been examined, you will be shown a table of
phases for all 'good' and 'bad' maps to help
you decide on putatively correct phase
combinations.
```

Once inside this mode, the you can't use any of the commands available in the default mode, so you better adjust the map magnification, position, grid, etc from beforehand.

3 The teaching exercises

3.1 The /Plane_groups directory

This directory contains the files that were used for confirming that the plane-group specific information in the program was correct. I only include them for the shake of completeness.

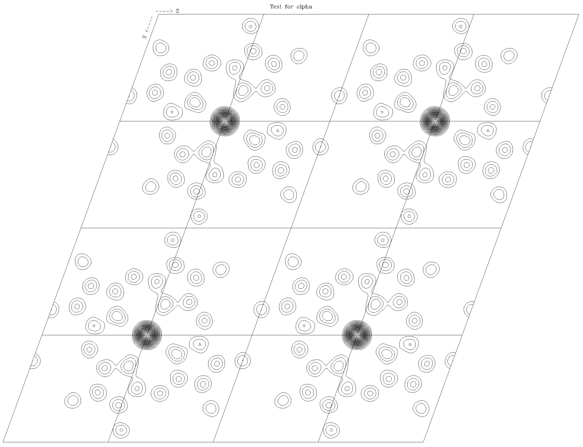
3.2 The /Examples directory

This contains three data sets :

```
heme_on_origin.in
heme_on_origin_4xF.in
simple_p4.in & simple_p4_Es.in
```

All three data sets have been calculated from hypothetical (non-existent) model structures based on the structure of the heme group but with a fluoride instead of an iron atom. Determining these structure from first principles ranges from extremely easy (heme_on_origin_4xF.in with four (???) fluoride atoms overlapping at the origin, almost all strong reflections +), to relatively easy (simple_p4.in with one fluoride and four carbon atoms (which can also be solved directly with trial-and-error (in the phase space)), to quite difficult (heme_on_origin.in). What follows is a short discussion of these examples. All maps shown extend from -0.5 to 1.5 on both axes and the origin is on the top-left-hand corner.

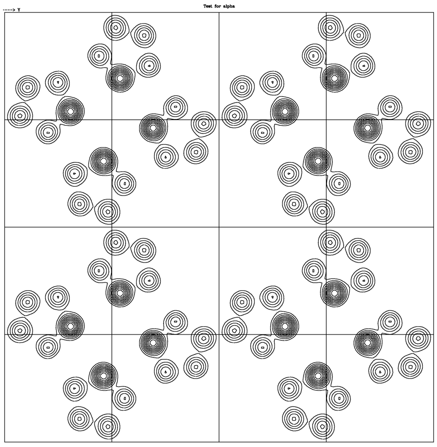
3.2.1 The heme_on_origin_4xF.in example



Plane group p2, with a=13.0Å, b=12.0Å, γ=110.0°. All strong reflections positive.

3.2.2 The simple_p4.in example

Plane group p4, with a=10.0Å, with four carbon and one fluoride atoms per asymmetric unit lying on the corners of a pentagon (???). The chemical impossibility aside, this is



probably the most useful example showing the complications arising with even such a simple crystal structure. For this reason, I will discuss it in some detail. To start the program (from the `/pepinsky` directory), do `pepinsky Examples/simple_p4.in`. The program will plot the weighted reciprocal lattice and the Patterson function. The equivalent positions are (x, y), $(\bar{x}, \bar{y})$, $(y, \bar{x})$, $(\bar{y}, x)$ and the unique self Patterson vectors (ignoring vectors related by Patterson symmetry) are $(2x, 2y)$ (single weight) and $(x - y, x + y)$ (double weight). The Patterson function contains two large peaks : one at u=0.376, v=0.219 and the other at u=0.204, v=0.140. The first peak corresponds to the vector connecting the centres of the "molecules" and arises due to their symmetry and special orientation. The second peak is indeed the double weight F-F vector, giving $x_{\mathrm{F}} = 0.172$ and $y_{\mathrm{F}} = -0.032$. Note that the single weight F-F vector (at $(2x, 2y)$) is not visible in the F^2 synthesis, but can be seen in the E^2 sharpened Patterson function (do `pepinsky Examples/simple_p4_Es.in` to see it). With the coordinates of the F atom in hand ($x = 0.184$, $y = -0.035$ from the sharpened Patterson function), making progress is not too difficult :

$$A = 4 \cos \pi[(h-k)x + (h+k)y] \cos \pi[(h+k)x - (h-k)y]$$

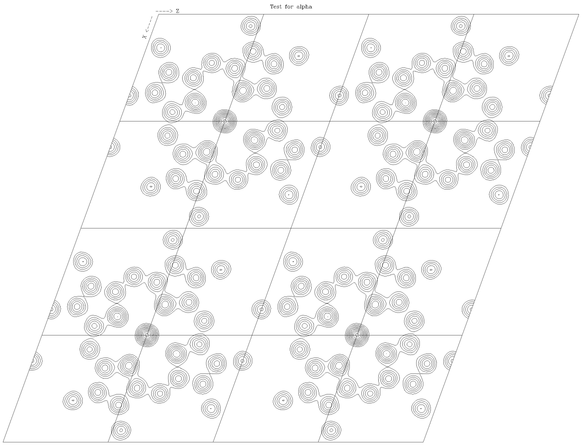
for *p*4, and ignoring the temperature factor (B=5Å^2 for all atoms), we have

hk	$\mathbf{F}_{\mathrm{F}}$	F_o	Correct phase	$\sin\theta/\lambda$	$4f_{\mathrm{F}}$	$f_{\mathrm{F}}/4f_{\mathrm{C}}$
01	+24.0	34	+	0.05	35.2	0.382
11	+13.6	23	−	0.07	34.4	0.389
21	−8.8	36	−	0.11	32.8	0.414
33	−20.0	16	−	0.20	26.8	0.472
24	+6.4	19	+	0.22	25.6	0.485
42	−19.2	21	−	0.22	25.6	0.485
43	−12.0	20	−	0.25	23.2	0.492
50	+15.2	19	+	0.25	23.2	0.492
64	+16.8	11	+	0.35	17.6	0.501
81	−13.6	10	−	0.42	14.4	0.472
47	−13.5	11	−	0.42	14.4	0.472

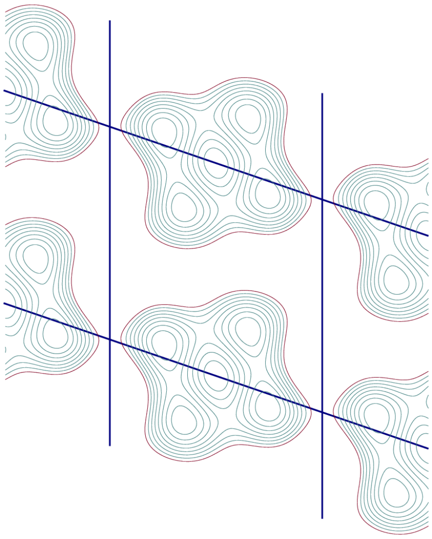
With the exception of the 11 reflection, all other phases agree with the indication from the fluoride atom (although only few of those reflections can have their signs fixed with certainty).

Alternative determination : suppose for a minute that we knew from beforehand that (i) the five atoms in the asymmetric unit do form a covalently linked molecule, (ii) that Z=4 (no atoms on special positions), (iii) that the atoms are resolved (no overlap), and, (iv) that this is a real-like crystal with crystal packing contacts between the molecules. Then the structure can be determined as follows : make reflection 21$(-)$ to fix the origin. Try both signs for 01 : it has to be $(+)$, otherwise we have density of the 4-fold axis. Try both signs for 11 : has to be $(-)$, otherwise we get unconnected density. Try 42 : $(-)$ or we get density on the 2-folds. Note that with only 3+1 reflections correctly phased, the "molecule" starts appearing. Try 43 : $(-)$, otherwise we get unconnected density. Try 33 : $(-)$, because we expect to see one resolved fluoride atom. Try all four combinations of 50 & 24 : for the $(+)(+)$ combination we can see all five atoms.

3.2.3 The `heme_on_origin.in` example



Plane group p2, with a=13.0Å, b=12.0Å, γ=110.0°. The fluoride atom on the origin can't really help. The determined student can easily reach to a low resolution image by fixing the origin (with, say, the 01 & 11 reflections), and then systematically permuting the other strong low resolution reflections ($1\bar{2}$, $1\bar{1}$, 10, 20). The correct map is difficult to miss :



Examination of permuation maps with uniformly distributed higher resolution reflections is a pleasant exercise, but phase errors will almost certainly build up. This is more an exercise for postdocs I should say.

4 LATEST ADDITIONS

Reading fractional coordinates off the map : Clicking with the left mouse button on the map section of the window gives you the fractional coordinates of the corresponding point along u,v or x,y.

Patterson superposition functions : Clicking the middle or right mouse button on the Patterson function, will form the sum or minimum Patterson superposition function corresponding to that point.

Storing and retrieving phase combinations : You can know store any phase combination that you like to any of the function keys (F1-F12). This you can do both in the default mode, and in permutation mode. You restore any stored combination by pressing the corresponding function key (while in the default mode). Take care with the following : the program will not make any checks whether you have already assigned a combination to any given function key. This means that *Pepinsky* will happily overwrite any of your previously stored combinations. Worse : if you accidentally press any of the un-assigned function keys, you will be taken back to the Patterson function.

Ready-to-go file for "GraphEnt" : When postscript output is requested, *Pepinsky* will also produce a file of the form `MAXENT_###.in` which can be fed directly to yet another of our localy available programs (*GraphEnt*, for the calculation of maximum entropy maps). You can safely delete it.

Flags for Patterson function contouring : If you intend to spend time looking at the Patterson function, there are two new command line flags to change the way *Pepinsky* contours Patterson functions. The first is 'P[atterson]' (and you call the program with something like `pepinsky Patt my_file.in`), and the second is '2' (and you call the program with something like `pepinsky 2 my_file.in`). The first mode will give you a logarithmic interval for contour plotting (with emphasis on low density features). The second mode is linear but will plot contours every 2% of the origin peak (instead of the default of 4%).

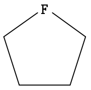
Limited support for CCP4 .mtz files : It is now feasible to run *Pepinsky* directly with a CCP4 .mtz file as input : `pepinsky [low-res] [high-res] my.mtz`, where `low-res` and `high-res` are resolution limits (Bragg spacings in Å). If you do not give resolution limits, all reflections will be used. The currect version of the program only supports some of the most common (macromolecular) space groups. If your space group is not supported, you will be told so. If your space group is supported and only has one centrosymmetric projection, the program will go ahead and use it. If there are more than one centrosymmetric projections, you will be asked to select a projection (before going into graphics mode). *Please note that there is no such thing as column assignments, etc. Pepinsky always assumes that the first column (after the indeces) is F, and the second is $\sigma(F)$.*

Pepinsky's Machine.

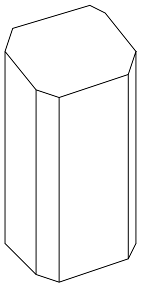
5 First (and most probably the last) exercise.

If what you need is a consistent, meaningful crystallographic exercise that will aim to teach you one thing only, this is not the right place to look for it. There is a whole lot of things below that are just mentioned on passing, are mixed-up with other seemingly unrelated bits of information, some of the basic principles are either assumed to be familiar to the user or they are not clearly spelled-out, etc. Worse : some of the ideas that are mentioned below are only correct under certain circumstances, but this is not stated clearly in the text. Instead, finding these “correct under certain circumstances” statements, is part of the exercise.

In a near-by galaxy of a universe with a different chemistry, a somewhat peculiar organic chemist purified this compound :

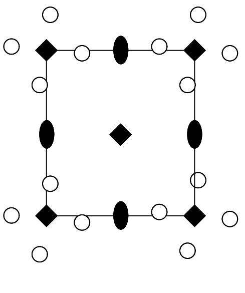


which, as you might have expected, had some very weird properties. In-between other things, this compound liked to form nice, big tetragonal crystals with a marvelous greenish colour, which diffracted X-rays to the respectably high (even for our galaxy) resolution of 0.8Å. Our poor crystallographer-turned-chemist determined the space group of his crystals, and found it to be : $P4$, with $a = b = 10.0$Å, $c = 3.5$Å, and Z=4 (ie, with only one molecule in the asymmetric unit). After some thought, he decided that the c axis is so short that not only there can't be any atoms overlapping along z, but also, that his molecule was completely flat with all its atoms lying on a plane perpendicular to c.



With these thoughts in mind, he went on to a two-century vacation to relax and think about his next move, when it occurred to him, that a completely flat molecule can be described by only two parameters per atom (x and y coordinates, the z coordinate can be set to any constant value). Another two centuries passed-by, before a crystallographer-turned-physicist discovered what is known to us as the projection theorem : the Fourier transform of a projection of an object is identical to a plane from the Fourier transform of the object that passes through the origin of the transform and is perpendicular to the projection axis. In less than two microseconds, our chemist realised all implications of the theorem : the object can be any object can be a crystal. The projection can be any projection can be a projection along a crystallographic axis. So, the Fourier transform of the projection of the electron density of a crystal along the crystallographic z axis, is identical with a reciprocal lattice plane which is passing thought the origin of the reciprocal lattice and is perpendicular to the z axis. It took him another two microseconds to realise that this was not totally correct (but could be said to be totally incorrect) : what he needed was not the reciprocal lattice plane (this he could calculate from the unit cell dimensions alone), but all the structure factors (both amplitudes and phases) that belonged to this plane.

After yet another century-long think-tank, he could finally formulate his problem : because his molecule was flat and its plane was perpendicular to the z axis, all he needed was to determine the projection of his crystals along z. To do this, he needed amplitudes and phases for all structure factors that belong to a reciprocal lattice place which is perpendicular to z and passes from the origin of the reciprocal lattice, ie all structure factors with indices of the type $hk0$. Because he have heard of the difficulty in measuring the phase of a wave that oscillates about $2\dot{1}0^{18}$ times in a second (the so called phase problem), he decided to make a good start by measuring only the amplitudes of these structure factors to a resolution of 0.8Å. While he was collecting this data set, one of their spies got back from our planet with a copy of the International Tables for X-ray Crystallography in hand. Over-excited, the chemist realised that our universe had been considering two-dimensional crystals (or, equivalently, projections of three-dimensional crystals) for quite some time now. He even found the following description of the symmetry elements for a projection of a tetragonal crystal along the 4-fold axis :



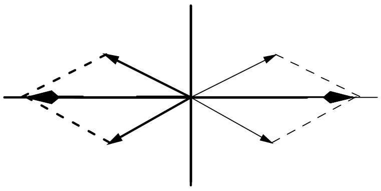
with black squares denoting 4-fold axes, ellipses 2-fold axes, and the open circles positions in the projection that are equivalent by symmetry (with coordinates (x, y), $(\bar{x}, \bar{y})$, $(y, \bar{x})$, $(\bar{y}, x)$). What was even more exciting was that these guys (us), had somehow calculated that the structure factors for a structure that belongs to this plane group are real numbers (instead of complex numbers), meaning that they can only be positive or negative, that is, they can only have two possible phases : 0° or 180°. After a few hundred mail messages between him and a crystallographer-turned-biologist, the answer to why this was so, became obvious : For a unit cell containing N atoms, each atom with an atomic scattering factor $f_j, j \in [1, N]$, the structure factor in the hk direction is given by

$$\mathbf{F}(hk) = \sum_{j=1}^{N} f_j \exp 2\pi i(hx_j + ky_j)$$

Because for every atom at (x, y), there exists one atom at $(\bar{x}, \bar{y})$, we have that

$$\mathbf{F}(hk) = \sum_{j=1}^{N/2} f_j [\exp 2\pi i(hx_j + ky_j) + \exp 2\pi i(-hx_j - ky_j)]$$

but, $\exp 2\pi i(-hx_j - ky_j) = \exp -2\pi i(hx_j + ky_j)$, which means that all terms of the summation have the form $\exp i\phi + \exp -i\phi$, ie they are all sums of vectors of the type shown below :



It was clear now, to both the crystallographer-turned-chemist and the crystallographer-turned-biologist that all they had to do was to finish off the data collection and then find which reflections were positive, and which were negative. After two years (and four broken X-ray generators), they actually collected a complete and quite respectful data set containing 122 reflections. It actually took another 231 years to calculate exactly how many are the 2^{122} combinations of signs for these structure factors. By the end of this time, one of our spies managed to get his hands on their computers and faxed us the complete list of amplitudes and directions for these reflections. You can see these reflections (and you can change their signs) by typing `pepinsky_first_exercise` on your terminal (ignore the nonsense about something called "Patterson function". Just start clicking the mouse buttons on the reflections to get a feeling for how things work). Press the `H` key to get a list of the commands available.

QUESTIONS :

1. How many earth-years old is the crystallographer-turned-chemist, given that he grew his first crystal at the age of 47 ?

2. The graph showing the symmetry elements in the projection along the 4-fold axis does not have the x and y axes marked on it. Why ?

3. The structure factor equation in three-dimensions is

   $$\mathbf{F}(hkl) = \sum_{j=1}^{N} f_j \exp 2\pi i(hx_j + ky_j + lz_j)$$

   Can you use this equation to show that the information about the projection of the structure along z is contained in the $hk0$ reflections only ?

4. Why they said "...the structure factor in the hk direction is given ..." when we all know that h and k are simply two integers ?

5. The first step for showing that the phases in a projection along a 4-fold axis can either be 0 or π was to note that for each and every atom at (x, y) there exists one atom at $(\bar{x}, \bar{y})$. What is the silent assumption here ? The same question in different words : under what circumstances the phases for this projection will be 0 or π degrees ?

6. Take my word for it : if you determine the signs of the 21 (ie $h = 2, k = 1$), 01, 11 and 42 reflections correctly, you will see the molecule blurred, but it will be there. If you also manage to find the correct signs for the 43, 33, 50 and 24 reflections you will be able to see all five atoms in the electron density map. The question is : why with only 8 out of 122 reflections it is still possible to solve the structure ?

7. Which is the correct phase combination for the 21, 01, 11, 42, 43, 33, 50 and 24 reflections ? Is there only one correct phase combination ?

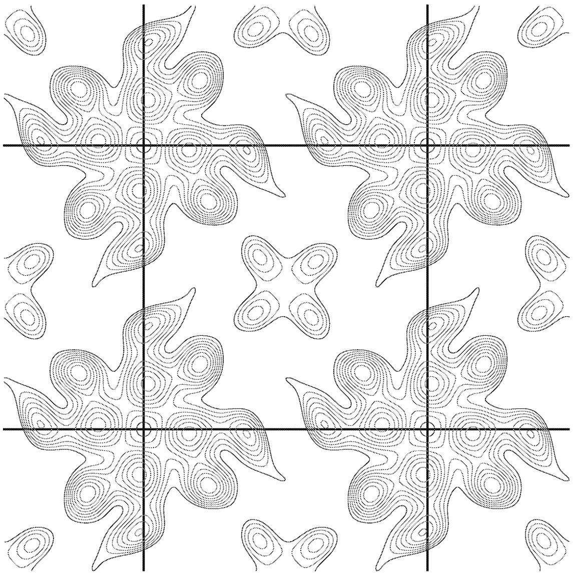
8. The total number of unique phase combinations for this plane group is only 2^{121} instead of 2^{122} (ie half). Why ?

9. Can you determine phases for any of the higher resolution reflections ?

10. Find the x and y coordinates of the five atoms in the asymmetric unit. Mail your answer to `glykos@crystal2.imbb.forth.gr`, and I will forward it to the good old `chemist@another.universe`.

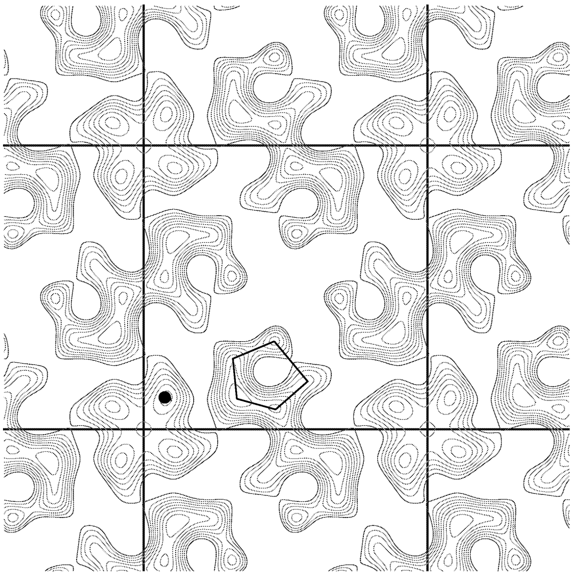
Philosophical aspects

Knowing just the amplitudes for 122 centrosymmetric reflections defines a set of 2^{121} electron density maps that are in agreement with the observed data. You could only determine what you think is the correct structure because of your expectations of how a non-overlapping projection of a molecule should look like and because you somehow had an idea of how your molecule should look like. To take this argument to its extreme, the amplitudes of, say, the 21, 01, 11, 42, 43, 33, 50 and 24 reflections are also consistent with the following projection of a big flower and a smaller flower :

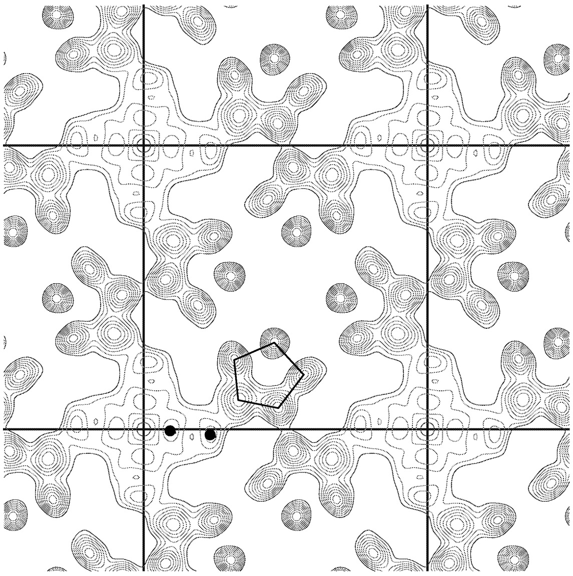


The only difference between the "chemically correct" map and the "flowering garden" map is, of course, that the phases used for the calculation were different (**QUESTION :** can you determine the phases corresponding to the projection of garden ?).

Under normal circumstances, you can't be fooled to think that a garden-looking map is your molecule. But phases are subtle beasts, and things can get messy. Look at this map :



If you were not certain whether a water molecule, or a Mg^{++} ion, or a SO_4^{--} could also be present in your crystals, you could have interpreted the above map as a projection of your molecule, together with an ordered water molecule or ion or something near the 4-fold. And this could take you all the way to a wrong structure determination. If you do not take my word for it, see how this wrong map can be beautified by including more reflections with suitably wrong signs :



The problem, of course, is not that you included a couple of non-existent water molecules, but that the molecule's bond lengths and angles and density are wrong. With the above (wrong) map in hand, and if you are desperate enough, you could even start making theories about the lack of bonding electrons for one atom and the possibility of a non-covalent interaction and "the discovery of a new interatomic force" and other nonsense.

The take-home message of preceding discussion in the form of a crystallographic Murphy's law is : if you think that something can be wrong, it is.