|epcc|

EPCC-SS98-02

Determination of Crystal Structures from Powder Diffraction Data

Sam T Carr

Abstract

In the past crystal structure determination from powder diffraction has mostly involved highly symmetric inorganic systems and has had limited application for organic molecules. This is primarily due to the fact that whereas single crystal reflection intensities are resolved in three dimensions in reciprocal space, powder diffraction patterns contain the same intensity information projected in two dimensions.

This project involves parallelising an existing code (SNIFFER) to reconstruct the structure of crystals from powder diffraction, and then extending the code to deal with a crystal structure containing more than one molecule.

Finally, the software, and computing power of the Cray are used to attempt to resolve the crystal structure of a urea/catechol complex.

1 Introduction

X-ray diffraction is a standard method to investigate the crystal structure of solids. In its simplest form, mono-chromatic X-rays are fired at a single crystal of the structure to be resolved, and the microscopic structure can be uniquely determined by taking the Fourier transform of the resulting 3D diffraction pattern.

Unfortunately, single crystals of many molecules can not be grown to carry out diffraction. In these cases powder diffraction may be used instead, where rather than one crystal, many are used in a powder which is modelled by theory as a continuous distribution of orientations of the crystals. X-ray diffraction then gives the same information as for the 3D case, but projected onto 1D as there is no orientation of the crystal and only the angle between the incoming and outgoing beam has any significance. In simple cases, the crystal structure can then be fairly easily built from this data, but all sorts of problems arise with more complicated structures. This is mainly due to peak overlap where information from two different reflections end up either in the same place, or too close to be resolved by the resolution of the measuring device.

Using computer modelling techniques the structure of these molecules can now be measured. The program models possible structures compatible with known information about the crystal structure, simulates the powder diffraction pattern from such a configuration and compares it to the experimental data. By making the search fine enough, the crystal structure can then be discovered, although this obviously requires high performance computers. Techniques such as Reitveld refinement, which takes a structure close to the real one and refines it, are used to finalise details of the structure but require a good initial approximation.

This technique has been used successfully for determining the structure of many asymmetric molecules. However, there are many more that have not been resolved due to lack of computing power. This project involves implementing this search on a parallel computer to try and resolve some more complicated structures.

For more details, see project portfolio entry in [5].

2 Background

The methodology which has been developed for solving crystal structures from powder diffraction can be summarised as follows:

1. Collection of high resolution powder diffraction data.
2. Determination of the unit cell parameters and space group of the crystal using a variety of programs.
3. Determination of the molecular configuration of the molecules present within the asymmetric unit of the crystal lattice.
4. Determination of trial crystal structures for Rietveld refinement (SNIFFER)
5. Rietveld refinement [7] of the trial crystal structure.

The code for SNIFFER was originally written by Dr Robert Hammond and Dr Robert Docherty in 1996. It has since been used in conjunction with Cerius2, a general purpose modelling suite

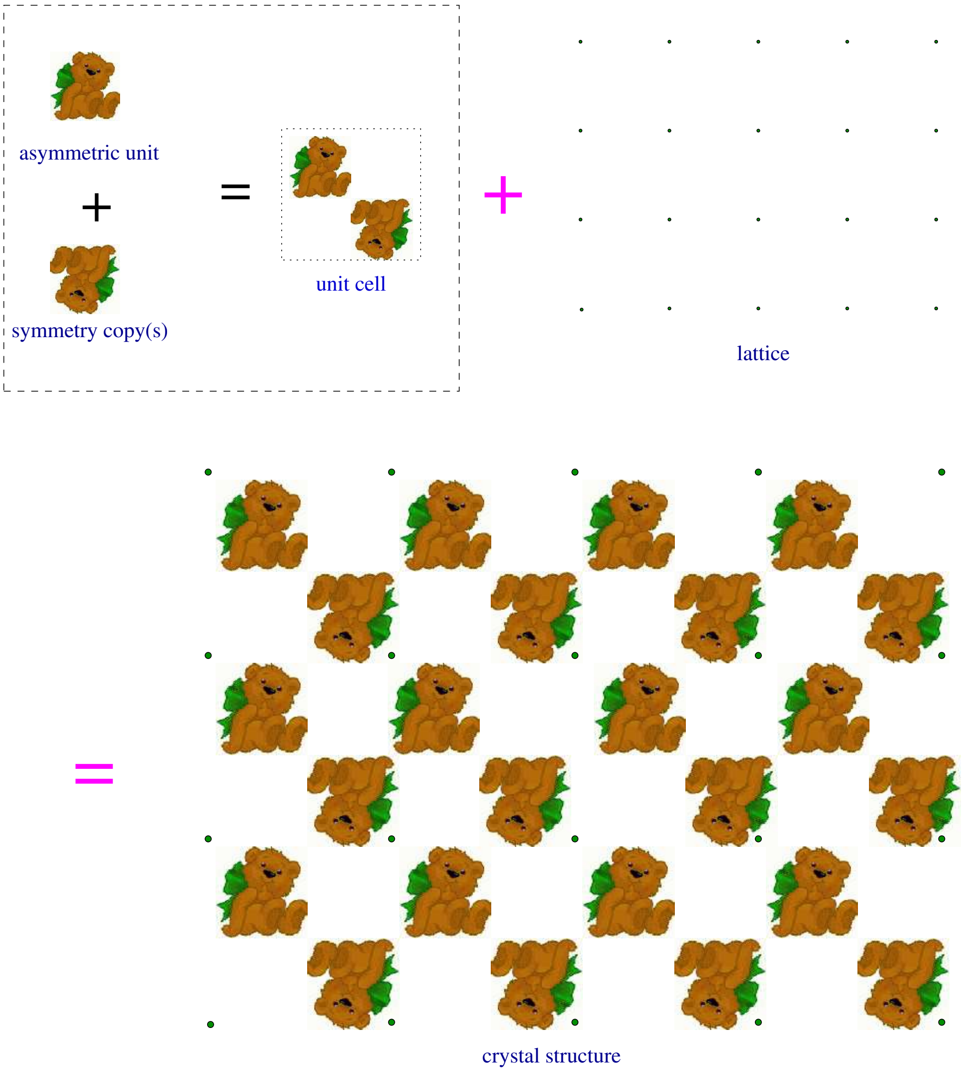

Figure 1: The structure of crystals

for investigating and visualisation of molecules, which includes a Reitveld refinement program. The crystal structures of molecules such as X-Form metal-free phthalocyanine [4] have already been successfully determined using SNIFFER.

A certain amount of work has also gone into resolving structures containing more than one molecule in the asymmetric cell - most notably, the approach taken in this project was originally developed by Lorna Smith (1997), with the program COMPACK1 which was successfully used to resolve the structure of rubidium dodecylsulphate [8].

For more details, see [3, 8].

3 Theory

3.1 Crystal Structure

The structure of a crystalline solid is built up from the original gaseous molecules in a certain way. Firstly, the structure is split into a lattice and a unit cell, the contents of the unit cell appearing on every lattice point.

There are seven different lattice point groups from which the structure can be built - cubic, tetragonal, orthorhombic, monoclinic, triclinic, trigonal and hexagonal. All crystal lattices can be reduced to one of these point groups on examination of the symmetries they contain.

The unit cell consists of a number of symmetry copies of the asymmetric unit, which are determined from the space group of the crystal, the space groups are different for each lattice type, and are obtained by systematically removing symmetry components from the lattice point symmetry group, depending on whether or not that symmetry is present in the unit cell. There are a total of 230 space groups.

The asymmetric unit then consists of the minimum repeatable unit. This is usually a single molecule of the gaseous state of the crystalline solid (or molecules if more than one substance is present) Figure 1 shows all of this diagrammatically. For more rigorous definitions, see [1, 2].

3.2 Diffraction Patterns

Electromagnetic radiation can be thought of as waves, and so a diffraction pattern is just the interference of the waves reflected off the crystal. The peaks in a diffraction pattern are where constructive interference occurs, where the waves combine in phase. X-rays are ideal for probing a crystal structure as their wavelength are of the same order of magnitude as the typical spacing between atoms in a crystal lattice which is about 1 angstrom.

It is the periodic nature of a crystal lattice that gives rise to non-uniform diffraction patterns. Taking just the crystal lattice and ignoring the unit cell at each lattice point, many lattice planes can be constructed, each of which reflects the incoming radiation. The peaks in the diffraction pattern can then be found by requiring a half-integral number of wavelengths between each plane, which is expressed in Bragg's law,

$$n\lambda = 2d \sin\theta \tag{1}$$

where the integer n is the order of the reflection, d is the spacing between the lattice planes, θ is the angle of the incoming radiation and λ is the wavelength of the X-ray. This theory is well described in [1].

It is the position of the peaks that determines the space group and cell parameters of a crystal. Although in powder diffraction, there is the complication of overlapping orders, these properties are usually uniquely determined from diffraction data.

The relative sizes of the peaks depends however not only on the space group but also on the orientation of the molecule within the unit cell. X-rays tend to reflect off electrons, so a large peak means high electron density on this plane. Hence taking the Fourier transform of single crystal diffraction data will directly give the density of electrons in a unit cell, from which it is easy to infer the molecular structure. However, no such analytic approach can be taken with powder diffraction data except for the simplest highly symmetric molecules, because overlapping peaks means there is not enough data. Hence a trial and error approach is adopted - implemented in SNIFFER.

3.3 Calculating Lattice Energy

The lattice energy of a crystal is defined as the binding energy of the crystal itself, ignoring individual molecular contributions. SNIFFER calculates the lattice energy per mole of asymmetric units, as opposed to Cerius2 which calculates energy per mole of unit cell. Transforming between the two is a simple matter of multiplication by the number of symmetry copies of the asymmetric unit in the unit cell.

The lattice energy is calculated by summing all of the atom-atom contributions to it. A crystal lattice has millions of atoms in it, but because of symmetry, only the interactions of one unit cell with the rest of the atoms needs to be considered. In calculating these interactions, the forces all fall off with distance, so only a certain number of neighbouring cells need be considered specified by a radial cut-off. The smaller the cut-off, the faster the calculation, but also the less accurate it is. This is shown diagrammatically in figure 2, where in this case the cut off would probably be set at b, to give a compromise between speed and accuracy.

Three forces are considered by SNIFFER. The first is the Van-der-Waals force law, which is a semi-empirical law to model the interactions between atoms. It is

$$U = \frac{A}{r^6} - \frac{B}{r^{12}} \tag{2}$$

where U is the energy contribution, A and B are constants depending on the elements in question, and r is the distance between the two atoms. In the substances examined in this project, this force gave the largest effect in calculating lattice energies.

Secondly, there is the Coulomb interaction. Although each individual atom on its own is neutral, the presence of other atoms changes the electron density distribution slightly so that, although the entire molecule is still neutral, each atom has a small charge. These are calculated using a program called MOPAC. SNIFFER uses these calculated charges to evaluate the electrostatic contribution to the lattice energy by using Coulombs law

$$U = \frac{q_1 q_2}{4\pi\epsilon_0 r} \tag{3}$$

Finally, hydrogen bonds are evaluated separately. The electron in a hydrogen atom can move about easily, so the presence of another molecule close by, even though not covalently bonded will change the state of the hydrogen slightly, giving a new effect, hydrogen bonding. This is modelled semi-empirically by

$$U = \left(\hat{\vec{a}} \cdot \hat{\vec{b}}\right)^4 \left(\frac{A}{|\vec{b}|^{12}} - \frac{B}{|\vec{b}|^{10}}\right) \tag{4}$$

where $\vec{a} \cdot \vec{b} < 0$ and $U = 0$ otherwise. Here, $\vec{a}$ is the vector to the H-bond doner which is the atom to which the hydrogen is bonded in the original molecule, and $\vec{b}$ is the vector to the H-bond acceptor, which is the vector between the two atoms whose energy is being evaluated. This is used instead of the typical Van-der-Waals formula.

The bonding of crystals is well explained in [6].

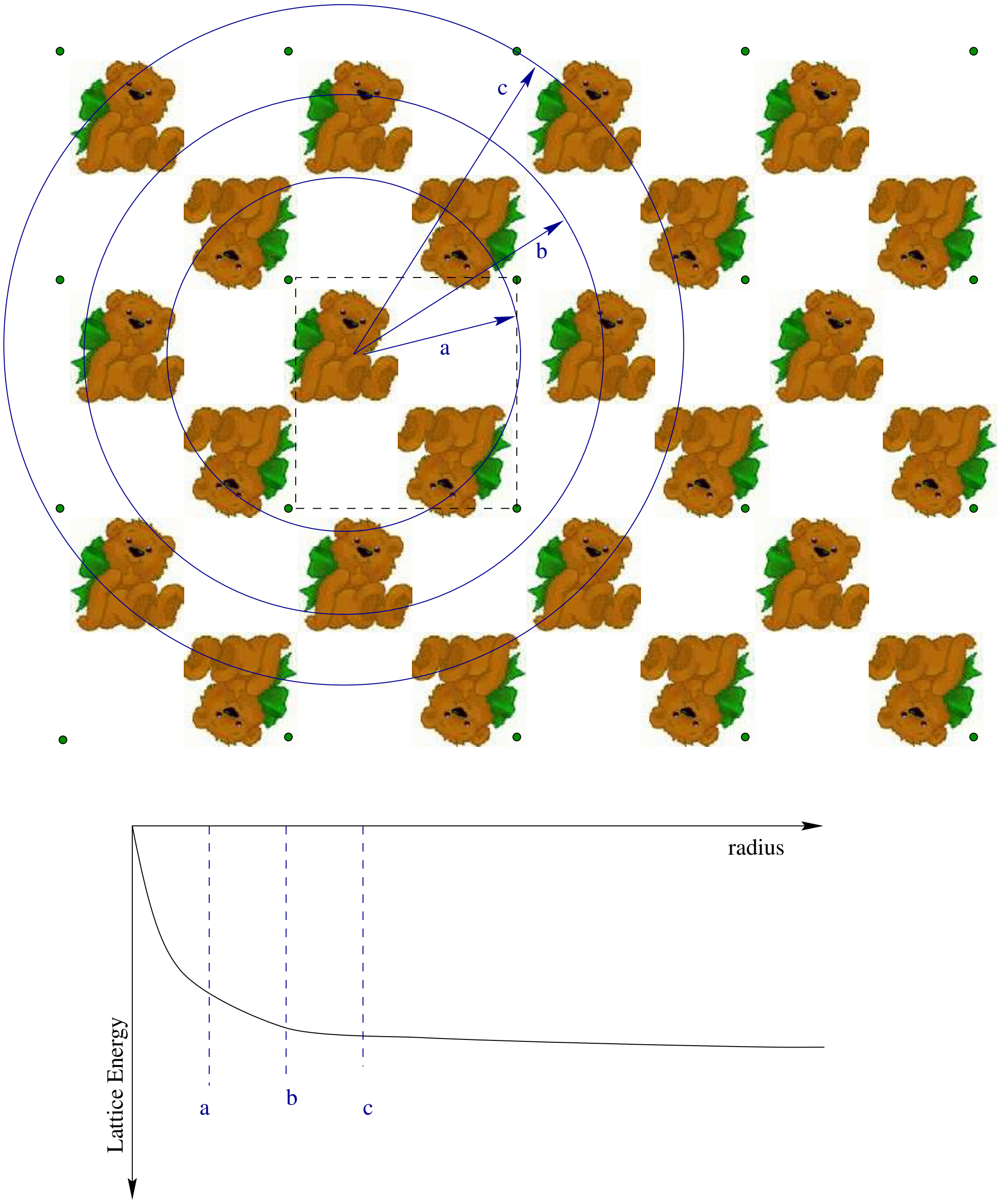

Figure 2: A plot of Lattice Energy against Cut-Off radius

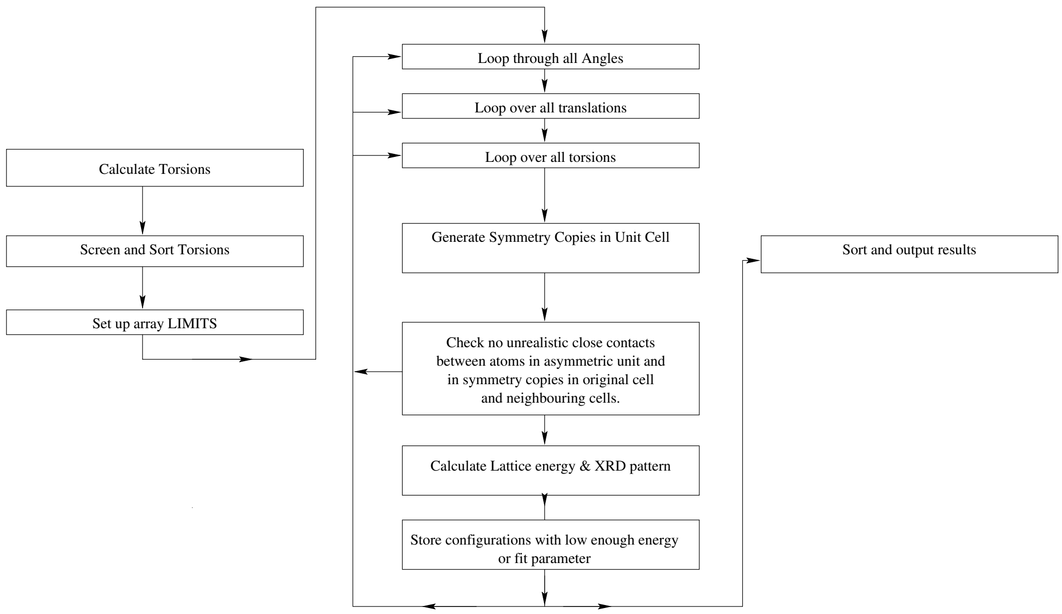

Figure 3: Schematic of the code SNIFFER

4 Parallelisation of Sniffer

Given the space-group, cell parameters, and known molecular structure SNIFFER searches each possible orientation and position of the molecule within the asymmetric unit of the symmetry group. It first discards unrealistic possibilities due to close contacts between atoms, and then calculates the lattice energy of each structure. For reasonable lattice energies, the X-ray diffraction pattern is calculated, and compared to the experimental data to give the closest fit. This is shown diagrammatically in figure 3. More details of the exact search mechanism are given in [3].

This searching mechanism is ideal for parallelisation. Each search is independent of each other one, giving no communications overhead, and hence there is the possibility of perfect speed-up. The method of parallelisation was however limited by the algorithm and memory size.

SNIFFER initially sets up a large array, LIMITS which is an array giving the atoms which could possibly collide with other ones. This is used to speed up the routine to check distances for each possible translation within the asymmetric unit. As this array is extremely large, and each translation requires only part of the array, this array should be distributed between processors - and parallelisation should therefore take place over all possible translations.

To avoid using dynamic load balancing, which would require a lot of costly communications to access correct elements of the array LIMITS, a cyclic decomposition of translations was used. This means that neighbouring elements were on different processors and hence no processor should have the bulk of the work, and the only unbalancing should be due to random fluctuations depending on the system.

As this is trivial parallelisation, the MPI calls do not greatly affect performance. Aside from three calls at the beginning to initialise MPI, no communications are required until the end when the results are collated and sorted. One processor only carries out this task as the number of

Processors	Time (mins)	Speed Up
4	40.224	4.00
8	26.591	6.05
16	10.056	16.0
32	5.300	30.4
64	3.536	45.5

Table 1: Timing Results from a test run with Benzephonone

possible results should be so small that communications costs in parallelisation would outweigh any parallelisation benefits.

4.1 Timing Results

The timing results for a test run (with benzephonone) are shown in the table 4.1.

The program with this degree of search would not run on less than four processors, so speed-up is calculated assuming perfect speed-up up to four processors. The parallel speed-up was very good on the whole, with any small discrepancies being due to imperfect load-balancing. For example there is no perfect way to distribute three processes amongst two processors, and there is no way of knowing in advance how long each piece of the calculation will take.

For running on more than 64 processors, it may be worth extending the program to distribute the unbalanced processes over the fastest processor. This would however give very little gain except in exceptional circumstances, and the communications costs may outweigh the gain.

5 Two Molecules in the Asymmetric Unit

When there are two molecules within the asymmetric unit, everything becomes more complicated. This can be simulated in the original version of SNIFFER by adding the second molecule onto the first by a pseudo-bond, and then using the torsional controls to move the second molecule around the first. This is not very satisfactory, as the second molecule is entirely independent of the first, and so many possible configurations are missed by using torsions.

On the other hand, in its full generality, the system has twelve degrees of freedom (three rotational and three translational for each molecule), so even searching just two units in each direction would lead to over four-thousand possibilities - a number which soars upwards when any realistic search is attempted. To solve these problems, I used a variation of the method used in COMPACK1, [8].

In COMPACK1 each molecule is treated individually and a distance check is carried out. Every configuration which passes the individual distance check is saved, and are combined with every configuration in the second which passed its own distance check. These are then fed back to a modified version of SNIFFER which then uses these as its search input data.

In this project, a parallel version of COMPACK1 within SNIFFER was implemented with the following addition. Every single molecule that passes on distance has its lattice energy evaluated as if it were the only molecule in the system. Each molecule then has a cut-off, and if the

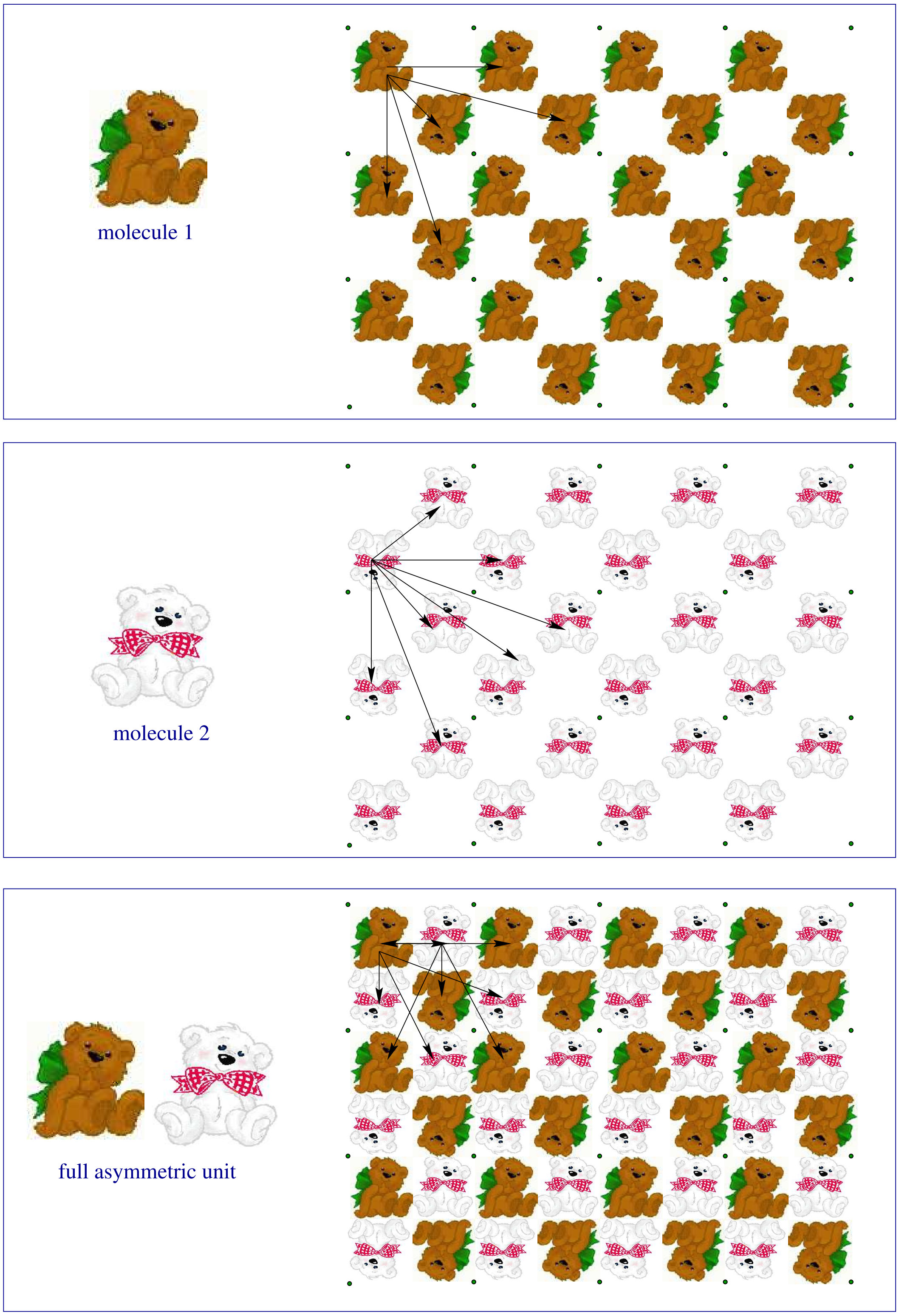

Figure 4: Building the lattice and calculating the energy of a crystal containing two molecules

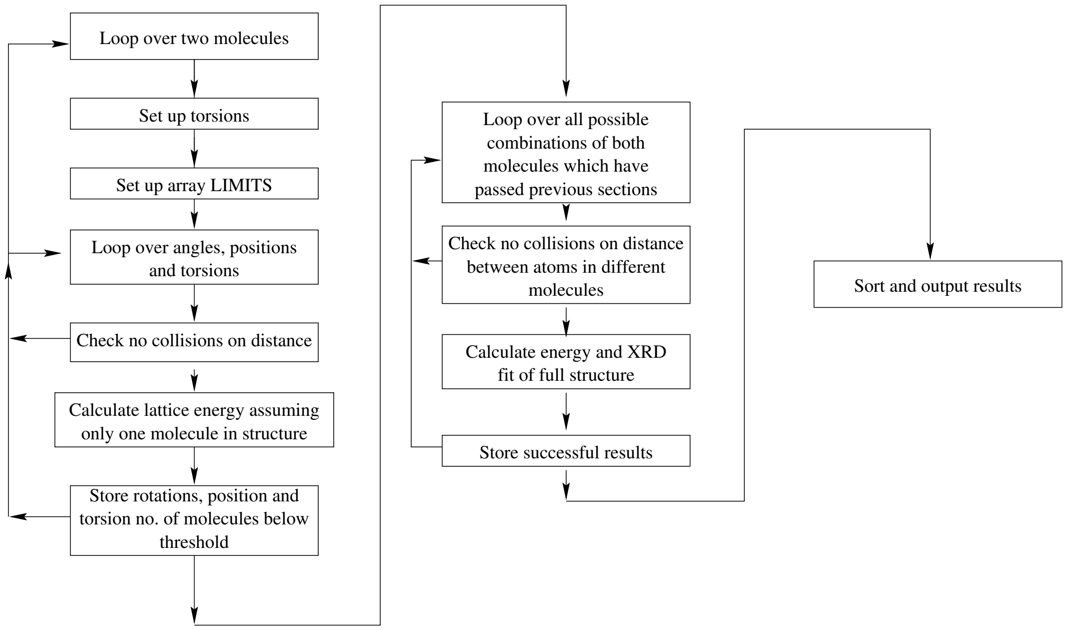

Figure 5: Structure of SNIFFER for dealing with two molecules

individual lattice energy is too high, the configuration is discarded. The flow chart for the program can be seen in figure 5

Care must be taken when setting these cut-offs because the tight binding of the two molecules within the asymmetric unit will often overcompensate for a higher lattice energy for each individual molecule. However, by investigating the breakdown of the lattice energies in a rough search, suitable values can be chosen for these cut-offs, giving a large time-saving, but with a very low risk of discarding a suitable configuration.

The other advantage of this method is that individual lattice energies are stored together with the configurations, so when the lattice energies for the full case are calculated, only the AB and BA cases need be evaluated, those of AA and BB having already been calculated. This halves the time required by the lattice energy routine. The steps in calculating the lattice energy are shown in figure 4.

5.1 Parallel techniques for Two Molecules

Parallelisation of the two-molecule version of the program is simpler than the one-molecule search. The program loops through all of the configurations cyclically, which aids load balancing and keeps communications to zero. The only communications costs are in gathering the results of each individual molecule search in the middle, and collating results at the end. This gives almost perfect speed-up, as expected from trivial parallelisation.

a	b	c	α	β	γ
19.218	6.249	7.179	90.000	97.883	90.000

Table 2: Lattice parameters for Urea/Catechol. Lengths are in angstroms, angles in degrees

Full Symbol	Unique Axis	Orientation Number	Lowest Fit Parameter
1 P21/c 1	b	1	0.85941
1 P21/n 1	b	2	0.91608
1 P21/a 1	b	3	0.92118
1 1 P21/b	c	1	0.92623
1 1 P21/n	c	2	0.91850
1 1 P21/a	c	3	0.88768

Table 3: Lowest fit parameters for a course run of SNIFFER with all sub-groups of P21/a

6 The search for the structure of Urea/Catechol complex

The crystal structure of a urea/catechol complex to date has not been determined by single crystal diffraction. Powder diffraction has been carried out, and the lattice parameters and space group have been determined. It is P21/a which is a monoclinic lattice. The lattice parameters are given in table 2.

A search with an angular step size of 15 degrees and a translation step size of 1 angstrom for each molecule were used.

6.1 Results

Unfortunately, no good results were found. Results were found with the expected lattice energy, but the lowest XRD fit parameter was 0.85, which is very high. This could be due to a number of reasons, the search could possibly not have been fine enough and the structure was missed. Alternatively, there may be a very large torsional effect caused by the tight interaction of the two molecules within the asymmetric unit. Other possibilities include the XRD data not being accurate enough, or a wrongly determined space group.

The best energy fit is shown in figure 6.

The space group was tested using SNIFFER. A short search for each of the monoclinic space-groups was carried out. Table 3 shows the lowest fit parameter for each of the sub-groups of P21/c. The search had an angular step of 30 degrees and a step size of 1.5 angstrom. The evidence is towards the group P21/a being correct, but this should be checked thoroughly.

7 Conclusion

The adaption of SNIFFER to run on parallel machines with almost perfect speed up was successful, managing to reproduce known results (such as the structure of Benzephonone), and is capable of determining many structures which were previously unknown due to lack of computing power.

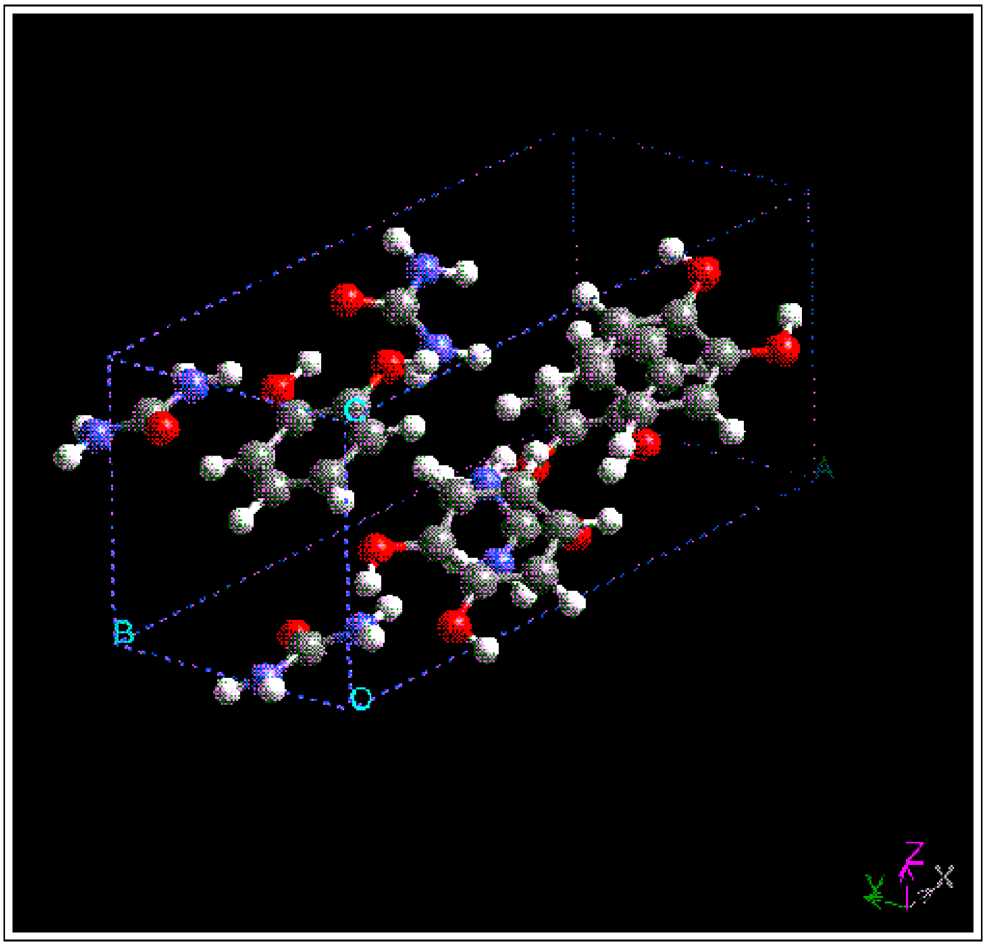

Figure 6: A possible structure for Urea/Catechol, found with SNIFFER and plotted with Cerius2

The extension of SNIFFER to deal with two independent molecules was also successful. It managed to reproduce the structures for urea/catechol already discovered using torsional methods. The structure of urea/catechol could however not be resolved, possibly due to the coarseness of the search.

References

[1] Ashcroft and Mermin. *Solid State Physics*. Saunders College Publishing, 1976.

[2] Walter Borchardt-Ott. *Crystallography*. Springer-Verlag, 1993.

[3] Hammond, Roberts, Docherty, and Edmondson. Computationally assisted structure determination for molecular materials from x-ray powder diffraction data. *Physical Chemistry B*, 1997.

[4] Hammond, Roberts, Docherty, Edmondson, and Gairns. X-form metal-free phthalocyanine: crystal structure determination using a combination of high resolution x-ray powder diffraction and molecular modelling techniques. *Chem. Soc. Perkin. Trans. 2*, 1996.

[5] Elspeth Minty. 1997 SSP Project Portfolio. EPCC Internal, June 1997. EPCC-SSP-PORTFOLIO-97.

[6] David Pettifor. *Bonding and Structure of Molecules and Solids*. Oxford University Press, 1995.

[7] H M Rietveld. Line profiles of neutron powder-diffraction peaks for structure refinement. *Acta Cryst*, 1967.

[8] Lorna Smith. *Crystallisation of alkyl metal sulfates: A molecular perspective*. PhD thesis, Strathclyde University, 1997.

"I am Sam
Sam I am
I do so like green eggs and ham."

Dr Suess

Well, that partly describes me. But as a slightly more physical description, I am a fourth year theoretical physics student at St Johns college, Oxford. Aside from physics, my major hobby is music - I play the piano, performing as a soloist and accompanist all over Oxford and am also co-founder of the St John's chamber players, having performed Schuberts 'Trout' Quintet and Brahms Piano Quintet with them since we founded two terms ago. I also founded and conduct the college orchestra, with an ambitious program planned for next term involving Rachmaninov's famous second piano concerto (used on the soundtrack of Brief Encounters) and Mozart's Requiem mass. My favourite drinks are good scotch malt whisky, or guiness.

My supervisors were

- Dr Lorna Smith, EPCC
- Dr Robert Hammond, Department of Chemical Engineering, Heriot-Watt University