Powder diffraction data and crystal chemical information combined in an automated structure determination procedure for zeolites

R.W. Grosse-Kunstleve, L.B. McCusker & Ch. Baerlocher*
Laboratory of Crystallography, ETH Zentrum, CH-8092 Zurich, Switzerland

Abstract

The FOCUS method, in which both crystal chemical information and powder diffraction data are included in the structure determination process, is presented. FOCUS combines automatic Fourier recycling with a specialized topology search specific to zeolites, which can be described as having 3-dimensional, 4-connected framework structures. The capabilities of FOCUS have been tested with seven examples of medium to high complexity. The method was then applied to three novel zeolite structures and a promising model could be obtained in each case. Experience shows that the approach of using chemical and geometrical knowledge can compensate for some of the information that is lost as a result of the overlap problem. At the same time, there is an intrinsic disadvantage: any method based on assumptions of certain structural properties is also limited to materials which conform to these assumptions. Examples which show the consequences of relaxing the structural assumptions are also discussed.

Introduction

Over the last five years, structure solution from powder diffraction data has developed rapidly and in many directions. In particular, some ingenious approaches to the unraveling of the relative intensities of overlapping reflections have been devised. These include both computational (e.g. David (1987), David (1990), the programs SIRPOW (Cascarano, Favia & Giacovazzo, 1992), DOREES (Jansen, Peschar, & Schenk, 1992) and FIPS (Estermann & Gramlich, 1993) and experimental (e.g. the exploitation of differential thermal expansion (David,

---

* Present address: Yale University, New Haven, Connecticut 06520-8114, U.S.A.

W.I.F., Ibberson, R.M., Shankland, K. & Sivia, D.S., in preparation) and texture effects (Hedel, Bunge & Reck, 1994)) methods. Alternatives to traditional direct methods of structure solution, tailored to the problems inherent to powder diffraction data, have also been developed. For example, Bricogne (1991) and Gilmore, Henderson & Bricogne (1991) have applied the principles of maximum entropy, and Rius, Sane, Miravitlles, Gies, Marler & Oberhagemann (1995) have solved a relatively complex framework structure using Patterson search techniques with purposely low-resolution (2.3 Å) data. Morris, Harrison, Nicol, Wilkinson & Cheetham (1992) have also demonstrated the advantages of combining X-ray with neutron data.

Although these advances are largely responsible for the exponential growth in the number of structures determined from powder data and for the increasing structural complexity that can be handled, there remains a large gap between the size of structure that can be refined using Rietveld techniques (ca. 60 atoms) and that that can be solved *ab initio* from powder data (ca. 20 atoms). Often the only alternative for the determination of a complex structure is model building. Over the years, this has proved to be a powerful, albeit time-consuming and uncertain, approach. Unfortunately, the multifaceted and intuitive thought processes involved are difficult to translate into the strict logic of a computer program. Nonetheless, the incorporation of at least some of the information used in model-building into an automated structure determination process should allow more complex structures to be solved. In particular, the types and number of atoms in the unit cell, their expected coordination numbers, typical bond distances and angles, and minimum distances between non-bonded atoms could lend themselves to such an approach.

Of course, the use of such crystal chemical information to supplement the powder diffraction data requires that certain assumptions be made, and that an algorithm be specific to a class of materials. Since our particular interest is in the structure analysis of novel zeolite and zeolite-like molecular sieves, whose structures lie near the current limits of powder diffraction capabilities, our efforts focussed on these materials. A general feature of zeolites and their analogs is that all have open 3-dimensional, 4-connected framework structures in which tetrahedrally coordinated atoms (T-atoms) are bridged by oxygens. The connectivity of these T-atoms is referred to as the framework topology.

Initial attempts to apply a genetic algorithm (Goldberg, 1989) combined with Fourier recycling to the problem looked quite promising in one dimension, so the procedures were expanded to three dimensions. However, estimations of the number of evaluations needed to obtain convergence of the "gene pool" revealed that the attempted procedure would require several orders of magnitude more computing time than is practically available (Goldberg & Segrest, 1987). Fortunately, quite a lot of the experience gained during this phase could be adapted to an alternative approach that eventually developed into the program system FOCUS. In the following sections, the FOCUS algorithms are presented, the results of several test cases summarized, and the application of this approach to solve a few previously unknown structures that had resisted other methods described.

The FOCUS method

The FOCUS method can be viewed as a tool that can be added to the set of conventional structure determination techniques. It is itself a combination and adaptation of classical methods. The core of the program system consists of automatic Fourier recycling, topology search, and topology classification and sorting algorithms.

The FOCUS environment

Figure 1 shows a flowchart of the complete structure determination procedure, and indicates where FOCUS is applied. Standard procedures are used to collect the data, search for peak

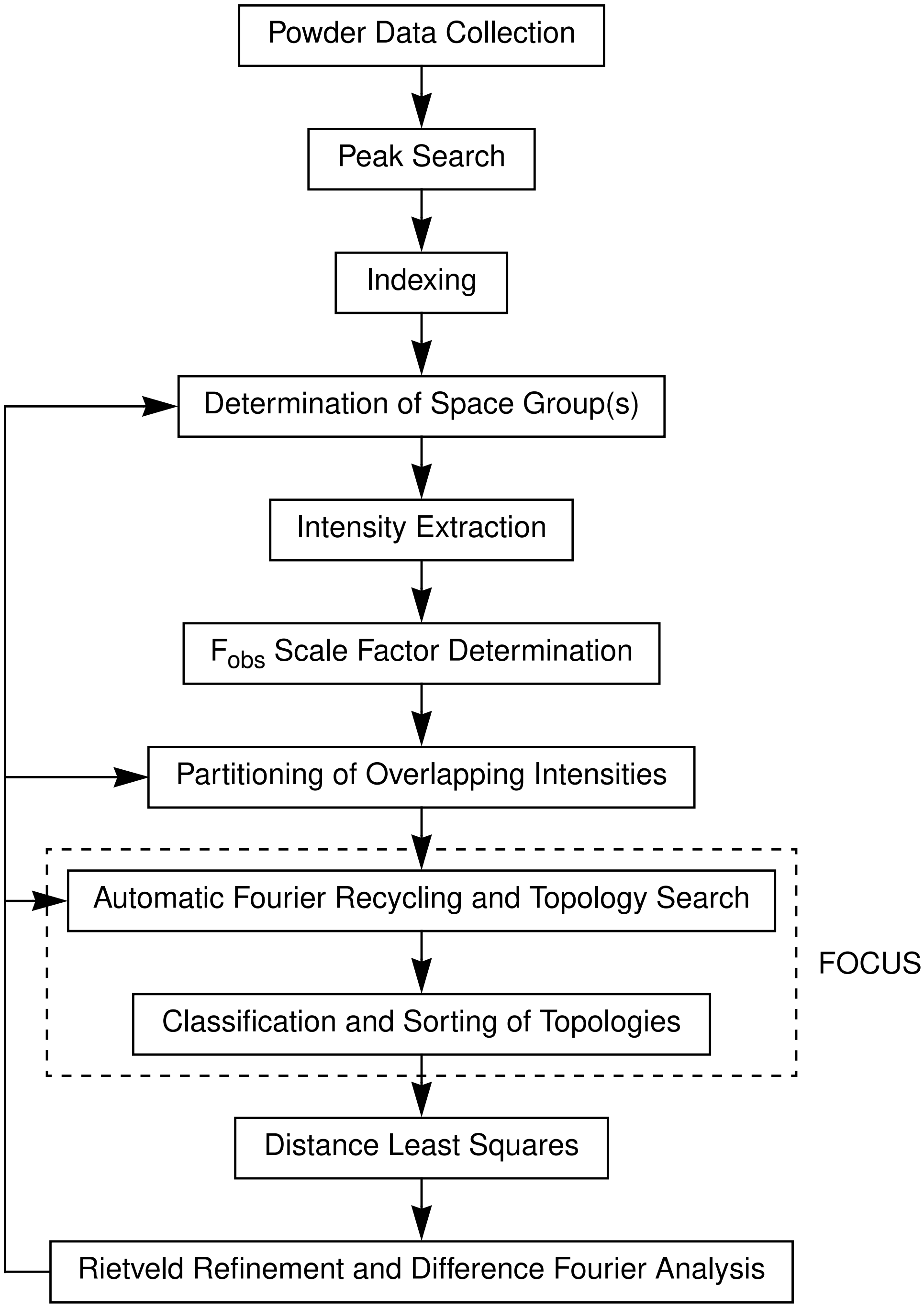

Figure 1: The Focus Environment

positions and index the pattern. The next step, the determination of possible space groups, deserves a word of caution. While space group determination with single crystal data is generally straightforward, this is not the case with powder data. The degree of reflection overlap in a powder pattern, whether due to sample quality or to structure complexity, has a significant potential to obscure the symmetry, and this can severely hamper the solution process. As will be shown elsewhere (applications paper in preparation), FOCUS can sometimes help to resolve space group ambiguities.

The implementation of Le Bail's iterative, model-free Rietveld refinement technique for the extraction of individual reflection intensities from powder patterns (Le Bail, Duroy & Fourquet, 1988) in most Rietveld programs, has made the extraction process almost routine. The extracted intensities are normalized by means of a Wilson plot, but, for zeolites and zeolite-like materials, it is very common for the Wilson plot to deviate significantly from the ideal straight line. Based on experience gathered while working on test cases, a pragmatic approach was developed to cope with this problem: the overall temperature factor was held fixed at $U_{overall} = 0.025$ Å^2, and the straight line shifted parallel to the y-axis until it intersected with the observed data at about $\sin\theta/\lambda = 0.15$. While this simple procedure gives satisfactory results in most cases, it should be mentioned that Estermann (1995) has recently presented a more elaborate and promising approach for the normalization of diffraction data from structures, which significantly violate the random atom expectation, on which the Wilson plot is based. However, this new approach was not applied here.

After scaling, the extracted intensities need further processing. The minimum treatment is the equipartitioning of overlapping intensities. That means that a sensible "overlap factor" (*of*) is chosen, typically 0.3, and the intensities of all groups of reflections with

$$2\theta_2 - 2\theta_1 < \frac{FWHM_1 + FWHM_2}{2} \, of \qquad (1)$$

are averaged (2θ = reflection position in the powder pattern, *FWHM* = full width at half maximum). Averaging of the N_g Fourier magnitudes $|F|^2_{\vec{H}}$ in a particular overlap group is then performed using the equation suggested by Estermann (1991)

$$|F|^2_{\vec{H},i} = \frac{\sum_{j=1}^{N_g} m_{\vec{H},j} |F|^2_{\vec{H},j}}{N_g \cdot m_{\vec{H},i}} \qquad i = 1, 2, \ldots, N_g \qquad (2)$$

(which results in equal $m \cdot |F|^2$, m = reflection multiplicity).

In recent years, methods for more sophisticated partitioning of overlapping intensities have been devised (David, 1987, David 1990, Jansen, Peschar & Schenk, 1992, Estermann & Gramlich 1993). In cases where a solution attempt with equipartitioned data is unsuccessful, the application of these methods can be helpful.

At this stage, the *pseudo* single crystal data are input to the FOCUS procedure. To complete the overall picture before going into details, it is sufficient to know that FOCUS produces a list of "solutions" in the form of framework topologies (connectivities of T-atoms) ranked by their frequency of occurrence, which can be interpreted as a measure of their likelihood of correctness.

For the most likely topologies, bridging oxygen atoms are inserted at the center of all node-node connections, and the resulting completed framework is subjected to a distance least-squares refinement with the DLS-76 program (Baerlocher, Hepp & Meier, 1977). After careful inspection of the DLS-76 residuals and the refined bond lengths and angles, the most promising structure can be selected as a starting model for a conventional Rietveld refinement with difference Fourier analysis to find missing atoms (i.e. non-framework atoms). In cases where the refinement does not converge, any other reasonable structures from the list can be tried, or parts of the whole procedure can be repeated. For example, a different space group can be selected, the partitioning of overlapping reflections can be varied, or the parameters for the FOCUS procedure can be changed.

The FOCUS Algorithms

The first step in the FOCUS procedure involves the generation and interpretation of Fourier maps using the *pseudo* single crystal data set produced by the normalization and partitioning steps. The Fourier recycling can be initiated either with random starting phases or with phases from some other source. The latter, for example, might be taken from a promising direct methods solution or calculated from a partial structural model. In this study, mainly the first approach, i.e. the use of random starting phases, has been employed. The phase set is then used together with experimentally determined Fourier magnitudes ($|F|$'s) to calculate an electron density map.

This electron density map is then subjected to a peak search algorithm. If random starting phases are used, the resulting peaklist can also be viewed as the corresponding "random starting model". In other words, starting with random phases or with a random model is essentially equivalent. Since it is technically easier to set up a random phase set than a random model, only the former was used.

The automatic Fourier recycling loop

Prerequisites

The automatic Fourier recycling is initialized by:

- selecting a subset of reflections for active use.
- defining structural properties, namely approximate unit cell contents and a minimum distance for each pair of atom types.
- defining technical parameters like grid spacings for the electron density map or maximum number of peaks in the electron density peaklist.

For the selection of the subset of reflections to be used in the recycling, the reflections (hkl, normalized and partitioned Fourier magnitudes) are sorted in descending order with respect to magnitude times multiplicity. Two selection procedures are possible: (a) a prescribed number of the strongest reflections are selected, or (b) the sum of all magnitudes, weighted by the multiplicities, is taken to be 100%, and the strongest reflections are selected from the sorted list until a prescribed percentage of the total sum is accumulated.

The description of the approximate unit cell contents is simply a list of expected atom types, the number of atoms per unit cell for each type, an isotropic displacement factor and an occupancy factor. In addition, structural information can be supplied by defining whether a certain type is expected to be a framework node, an atom bridging two framework nodes, or a general type. Further structural information is given in the form of minimum distances for pairs of atom types.

Initialization of a new trial and Fourier transform

The automatic Fourier recycling loop is illustrated in Fig. 2. A single trial is initialized by assigning starting phases to the selected reflections. The next step is a Fourier transform of magnitudes and phases to produce an electron density map.

Peak search

The first processing step of the electron density map is a peak search in the asymmetric unit of the unit cell. A histogram of the peak heights found is maintained throughout the search. After all grid points in the asymmetric unit have been scanned, the histogram is used to determine the height cut off, such that a preset maximum number of peaks is not exceeded.

Peak interpolation

Since the peak maxima do not generally coincide exactly with a grid point, the positions of

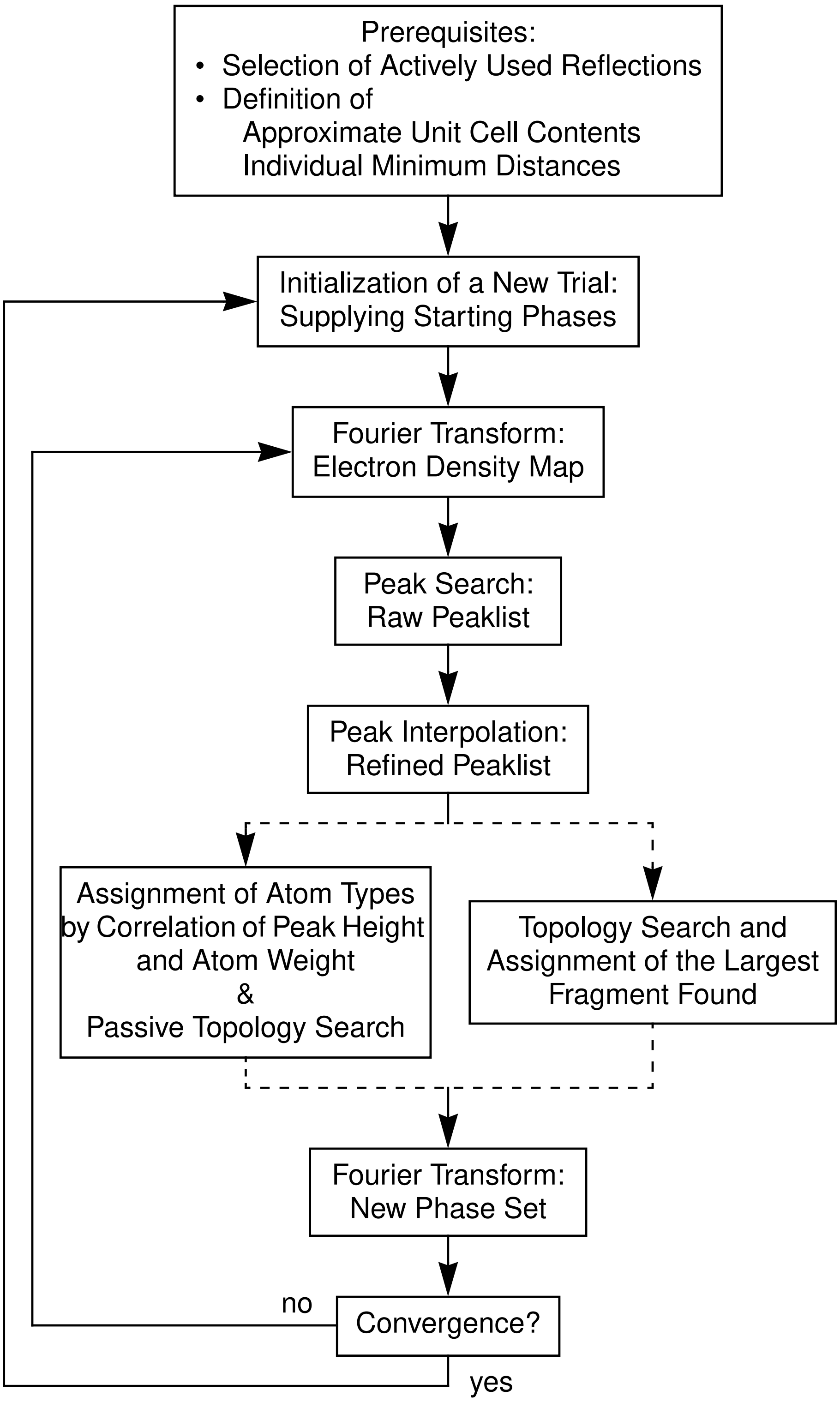

Figure 2: Automatic Fourier Recycling and Topology Search

the peaks are determined, or refined, with a formalism found in Rollet (1965), pp. 35-37. After a peak position has been refined, the shortest distance to all symmetrically equivalent positions (self-distance) is computed. If the self-distance is smaller than a prescribed minimum distance (e.g. for a position too close to a mirror plane), the peak is moved onto the symmetry element which is responsible for the close contact. After the shift, the self-distance calculation is repeated. Under certain conditions, the peak position will be corrected more than once.

In the next processing step, the list of interpolated peak positions is sorted in descending order by one of these criteria:

(a) the peak height found at the central grid point.
(b) the peak height $\rho_{calc}(xyz_{max})$ calculated for the interpolated position of the maximum.
(c) the analytical integral $\int_V \rho_{calc}(xyz)dV$ (Grosse-Kunstleve, 1996).

Experience has shown that the peak shapes in the electron density maps produced by the automatic recycling procedure are often very distorted and poorly approximated by the chosen mathematical models, and frequently introduce numerical instabilities. Therefore the simplest approach – use of the peak height found at the central grid point – turned out to give the best results. The last treatment of the refined peaklist is to set an "N-marker" (node marker) for each entry which can satisfy the node atom requirements (i.e. point symmetry is compatible with a tetrahedral coordination geometry).

Construction of a structural model

At this point there are two alternatives:

(a) <u>Assignment of atom types by correlation of peak height and atomic number</u>

The outer assignment loop steps over the defined atom types, which are sorted in order of descending atomic number. The inner loop steps over the unassigned entries of the refined peaklist, trying to find a position that fulfills the criteria for a pivot atom. The pivot atom type is assigned to a previously unassigned entry if (i) the N-marker is set (for atom types of class "node"), (ii) the multiplicity of the entry is not greater than the number of pivot atoms remaining to be assigned, and (iii) the prescribed minimum distances to all assigned atoms are not violated. The inner loop is terminated when the prescribed number of atoms per unit cell of the pivot atom type have been assigned, or the end of the refined peaklist is reached.

Independently, an exhaustive topology search among the 50…60 highest peaks in the asymmetric unit is performed and any topologies found are written to a file. This search procedure is described in more detail in the "Topology search" section below.

(b) Topology search and assignment of the largest fragment found

An exhaustive topology search similar to the one of alternative (a) above is used to find the largest framework fragment that can be built from a subset of the peaks in the refined peaklist with the N-marker set. The selection criterion is the total number of node-node bonds in the fragment divided by the number of active node positions. Of fragments with equal number of bonds and node positions, the one with the greatest sum of peak-heights is selected.

At the end of the topology search, atom types of class "node" are assigned to the fragment positions with an algorithm similar to that of alternative (a): the outer loop steps over the atom types of class "node" – again sorted in descending order of atomic number – and the inner loop searches for an unassigned fragment position. However, distances do not need to be checked, because the topology search has already taken care of these.

Fourier Transform and convergence test

The recycling loop is closed by a straightforward Fourier transform (see for example Giacovazzo (1992), "Calculation of the structure factor") of the structural model constructed through one of these processes, and a new phase set is generated. By means of a convergence test, which is based on the F-weighted ratio of phase changes, the decision is made as to whether the new phase set is used to calculate a new electron density map, or, in the case of convergence, a new trial is initialized by supplying new starting phases.

Topology search

The topology search is an essential part of the FOCUS method, and therefore it is dealt with in more detail here. With it, the additional structural information is supplied to the structure determination process.

Each time an electron density map is produced in the Fourier recycling step, the peaklist is examined to see if a 3-dimensional 4-connected net of T-atoms with appropriate distances and angles can be constructed. The topology search is an application of the well known backtracking algorithm (see for example Wirth, 1986) and operates on the refined peaklist. To make the topology search efficient, it was divided into two stages: the preparation of a list of potential node-node bonds ("bondlist") for each entry of the refined peaklist, and the actual backtracking which then operates on these bondlists.

Creation of the bondlists

For the creation of the bondlists, a *minimum node distance* (ND_{min}) and a *maximum node distance* (ND_{max}) is prescribed. Values typically used for SiO_2 frameworks were ND_{min} = 2.6 Å and ND_{max} = 3.6 Å, which allows for a tolerance of 0.5Å around the "ideal" node distance ND_{ideal} = 3.1 Å.

In a first scan through the refined peaklist, entries are marked as "Inactive" if the N-marker is not set, or the self-distance is less than ND_{min}. In the second scan, potential node-node bonds with distances in the range ND_{min} through ND_{max} are tabulated for each peak. If the distance between two nodes is less than ND_{min}, or if two peaks form more than the *maximum number of node-node bonds* (NN_{max}), they cannot be present together in the type of framework sought, and an “Exclusive” marker is set. In the next scan, all entries with less than the *minimum number of node-node bonds* (NN_{min}) are eliminated by setting the “Inactive” marker. Of course, the number of bondlist entries of peaks which had potential node-node bonds to those just eliminated is thereby reduced. Therefore, the last scan has to be repeated until no further changes are necessary. Finally, the refined peaklist is resorted by means of the number of active bondlists per entry, and the bondlists themselves are also sorted such that the order is optimized for the backtracking. Fig. 3 gives an example of the final bondlists of a refined peaklist. Position number one in the refined peaklist has four active bondlists, and one “Exclusive” marker, which indicates that positions one and four in the peaklist cannot occur together in a framework. The asterisk after a distance signals that this bond is symmetrically equivalent to the previous bond. In addition to the distances, the bond vectors (in Cartesian coordinates), pointing from the pivot peak to the corresponding bonded peaks, are also stored for use in the actual backtracking procedure.

The backtracking procedure

The first level of the backtracking procedure consists of an outer loop which steps over the active peaklist entries. Each pivot entry is used as a “seed node” to initialize a set of “present” framework positions (“F-set”). On the next level, a *connectivity completion procedure* (CCP), which loops all possibilities for the construction of NN_{min} through NN_{max} bonds for the pivot entry, is called. In these constructions, refined peaklist entries with indices less than the index of the pivot entry have to be omitted in order to avoid redundancy. For each possible bond configuration, a test, which checks its geometrical validity, is carried out. If the geometry proves to be acceptable, the positions which are newly bonded to the pivot position are added to the F-

No. in refined peaklist	No. of active bondlists or marker	Bondlists	
		Bond to No. in refined peaklist	Distance(s) [Å] or marker
0	5	0	3.3114
		1	3.3015
		2	3.4472 3.0204
		3	3.4349
		4	3.2393
1	4	1	3.1951 2.8726
		0	3.3015
		3	3.4216
		5	3.5401 3.2220
		4	Exclusive
2	3	0	3.4472 3.4472* 3.0204 3.0204*
		4	2.9356
		5	3.3422
3	3	0	3.4349 3.4349*
		1	3.4216 3.4216*
		4	3.4111
4	3	0	3.2393 3.2393*
		2	2.9356
		3	3.4111
		1	Exclusive
5	2	1	3.5401 3.5401* 3.2220 3.2220*
		2	3.3422
6	Inactive		
7	Inactive		

Figure 3: Final bondlists of a refined peaklist

set. Then the enlarged F-set is searched for the first entry which is not already a pivot-element (in a previous level) and the CCP is *recursively* called with this entry as the new pivot element. If all elements of the F-set have NN_{min} through NN_{max} bonds, a framework topology which meets the prescribed criteria has been found and it is written to a file.

Two basic types of backtracking algorithms are known: the algorithm which terminates as soon as a solution has been found, and the alternative algorithm which searches for all possible solutions and writes a protocol. The implementation discussed here is of the second kind. This means that the only condition on which the recursive CCP returns to the previous level is, that the possibilities for the construction of the desired connectivities for a given pivot position are depleted.

Selecting truly 3-dimensional frameworks

Experience revealed that another geometry filter is necessary to reduce the number of obviously useless frameworks produced by the search procedure. Very frequently, heavily distorted “layer structures” appeared. In an attempt to suppress all but truly 3-dimensional frameworks, a simple algorithm to test whether or not a path from an arbitrary starting node in the unit cell to all other nodes in the same cell exists was introduced (Grosse-Kunstleve, 1996).

Modified topology search: “two color” frameworks

There are a large number of zeolite frameworks with two types of strictly alternating node atoms, for example Si-Al, Al-P or Ga-P. While the node-node distances of pure silicon frameworks are always such that the (four) nodes bonded through bridging oxygen are also the next (four) neighboring nodes, this is not always true for other types of node atom pairs. For example, the gallophosphate ULM-5 (Loiseau & Férey, 1994) has one gallium in the asymmetric unit which is bonded to four phosphorous atoms through oxygen, and to another gallium, again through oxygen, at a distance smaller than the largest Ga-P distance. By ignoring this Ga-Ga bonded oxygen and also the four fluorine atoms per asymmetric unit, ULM-5 can still be viewed as tetrahedral framework with strict alternation of Ga and P. However, since the smallest Ga-Ga distance of the special gallium is smaller than the largest Ga-P distance, the topology search will not recover this framework.

To overcome this problem, the search algorithm was modified for frameworks with strictly alternating occupation of the nodes: a “color”, say white, is assigned to the seed node, which is set in the outer loop. In the CCP, all positions which are connected to the pivot position are assigned the “opposite color”, say black. Node distances smaller than ND_{min} are still not allowed, but bonds are created only between positions of different color.

This simple modification is sufficient to recover the tetrahedral topology of ULM-5 (given the correct peak positions). Furthermore, this modification also acts as a filter which allows only strictly alternating topologies to be accepted, and thereby reduces the number of unfeasible topologies that have to be investigated in the subsequent steps.

Identifying and sorting the topologies

A fast and efficient way of classifying and sorting the frameworks produced by the backtracking procedure was developed for the next stage. It is based on the evaluation of the site multiplicities, loop configurations (LC) and coordination sequences (CS). While the multiplicities were available immediately, because they were needed in several of the preceding steps, the determination of LC’s and CS’s is more involved.

Determination of a CS: a node counting algorithm

The notion of CS was formally introduced by Brunner & Laves (1971) in order to investigate the topological identity of frameworks and of atomic positions within a framework. The CS is a number sequence in which the k-th term is the number of atoms in “shell” k that are bonded to atoms in “shell” k-1. Shell 0 consists of a single atom, and the number of atoms in the first shell is the conventional coordination number.

The CS determination algorithm used here can be described as a *node counting algorithm* or a *coordination shell algorithm*. The algorithm is started by selecting an *initial node* (k = 0). In the next step, all nodes bonded to the initial node are determined (k = 1). For $k \geq 2$, all characteristics of the algorithm become evident: those nodes, which are bonded to the "new nodes of the previous step (k-1)", but have not been counted before, are counted.

Determination of a LC: modification of the node counting algorithm

The term LC as used here follows a definition of Fisher (1973) (where the term "Maschensymbol" is used) and is a generalization of the LC as defined in Meier, Olson & Baerlocher (1996). The LC of a framework node N_i with NN_i *node-node bonds* is understood as a set of $\binom{NN_i}{2}$ (binomial coefficient) pairs of integer numbers. Each pair characterizes the angle described by node N_i in the center and two bonded nodes. Fig. 4 gives an illustration of the six angles found for a node which is coordinated by four neighboring nodes.

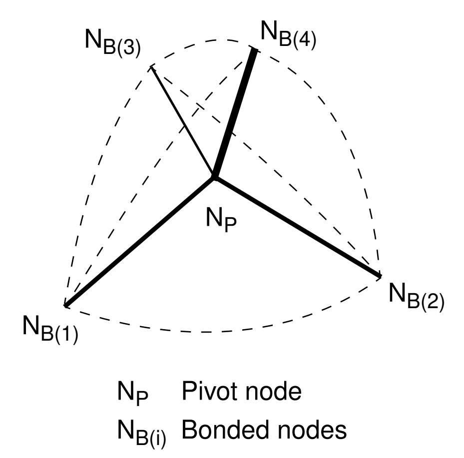

Figure 4: Six tetrahedral angles

The first integer of a pair is the number of nodes in the shortest loop which contains the corresponding angle. The second integer gives the number of loops with that number of nodes. For example, the loop configuration "4 1 4 1 5 1 5 1 6 1 7 2" says that two (of the six) angles are each part of single loops with four nodes, two angles are each part of single loops with five nodes, one angle is part of a loop with six nodes, and one angle is part of two distinct loops each with seven nodes.

The LC determination algorithm is very similar to the CS algorithm. The modified node counting algorithm is surrounded by an outer loop which steps over NN_i-1 bonded nodes. Let j be the index in the list of bonded nodes (the first entry has index 0), such that $N_{b(j)}$ is the pivot node in this loop. Each pivot node is taken as the *initial node* ($k = 0$), and the algorithm works its way through the coordination shells until all *target nodes* $N_{b(j+1)} \ldots N_{b(NN_i-1)}$ are visited. The crucial modification of the CS algorithm is that bonds to the center node N_i are never followed.

Each time a target node is hit, k+2 gives the number of nodes in the corresponding loop. If the target was not hit before, this number is recorded and the counter for the number of loops is set to one. If the target was hit before in the same shell (that means with the same loop size) the counter is advanced by one.

After all integer pairs are obtained, they are sorted in ascending order to give the final LC for the node N_i.

Combined evaluation of multiplicities, LC's and CS's

A characteristic "fingerprint" of a framework topology is obtained by constructing a sequence of integers for each node in the asymmetric unit, by merging site multiplicity, LC and CS. The LC consists of $\binom{4}{2} = 6$ pairs of integers, and the CS is computed up to the 10^{th} member. Altogether one 4-connected node position is described by 23 integer numbers.

Two frameworks – as produced by the search algorithm – are considered to be equivalent if the sets of lexically sorted integer sequences are equal. It should be mentioned that Fischer (1974) has derived four pairs of distinct sphere packings which cannot be distinguished by comparing the integer sequences, but these examples look unrealistic for crystal structures, and no example is known where two crystal structures cannot be distinguished by this “fingerprint”.

Applications

The FOCUS procedure has been applied to seven test cases of different complexity (including ZSM-5 with the most complex zeolite topology known) and to three previously unknown structures. Characteristic data for the ten structures is summarized in Tab. 1. In all cases, the full procedure outlined in Fig. 1 was followed. The whole-profile intensity extraction was carried out on measured data using the GSAS program suite in “Le Bail extraction mode” with “CW Peak profile type no. 2” (Larson & von Dreele, 1995). The refined profile parameters were used to prepare the overview of the overlap situation shown in Fig. 5. The overlap factor (eq. 1) used was 0.3, which means that reflections which are less than about 30% of their FWHM apart are put into the same overlap group. The plot shows how the ratio of overlapping and non-overlapping reflections develops with increasing resolution. For example, down to a d-spacing of 5.0 Å all reflections for EMC-2 are single, at a resolution of 3.0 Å, about 24% of all reflections overlap, and finally at 1.3 Å, the degree of overlap has reached 83%.

In all cases, data up to a resolution of 1.3 Å were used (indicated by the dashed line in Fig. 5). Variation of the FOCUS input parameters shows that, in general, the best recycling technique is a strict alternation of framework-fragment recycling and atom recycling, and to omit non-framework atoms in the model building procedure, but to include framework oxygen along with the node atoms in atom recycling mode (at the moment, oxygen cannot be included in framework fragment recycling mode). The test examples were all solved successfully, and in each case, the topology most frequently produced by FOCUS proved to be the correct solution. For ZSM-5, with 12 T-atoms in the asymmetric unit, feasible topologies were found rather slowly, but the result was clear nonetheless: only the correct topology was produced.

The method was then applied to three novel zeolite structures – the two zincosilicates VPI-9 and VPI-10, and the beryllosilicate B2 – and a promising model was obtained in all cases. Preliminary Rietveld refinements of the VPI-9 and VPI-10 structures indicate that the proposed models are correct (Grosse-Kunstleve, 1996). The structure of VPI-9 has since been confirmed with a full Rietveld refinement (McCusker, Grosse-Kunstleve, Baerlocher, Yoshikawa & Davis, 1996), and the Structure Commission of the International Zeolite Association has assigned the code **VNI** to that topology. Refinements of VPI-10 and B2 are still in progress.

Name & Ref.[a]	**#T**[b]	**Chemical Formula**	**Space group**	**Unit cell**	**Volume**	**Overlap**[c]
Dodecasil-1H [1]	4	$Si_{34}\ O_{68}$	P 6/m m m (No. 191)	a = 13.798 Å c = 11.211 Å	1848 $Å^3$	15 %
NU-3 [2]	2	$[Si_{54}\ O_{108}]\ \cdot(C_{10}H_{15}NH_2)_6$	$R\,\bar{3}\,m$ (No. 166)	a = 13.184 Å c = 22.221 Å	3345 $Å^3$	42 %
RUB-17 [3]	9 (5)	$K_4\ Na_{12}\ [Si_{28}\ Zn_8\ O_{72}]\ \cdot 18\ H_2O$	C m (No. 8)	a = 7.239 Å b = 40.562 Å c = 7.309 Å β = 91.84°	2145 $Å^3$	52 %
SAPO-40 [4]	4	$[(Si,Al,P)_{32}\ O_{64}]\ \cdot 2((CH_3CH_2CH_2)_4NOH)$	P m m n (No. 59)	a = 22.041 Å b = 13.698 Å c = 7.122 Å	2150 $Å^3$	64 %
Zeolite-A [5]	2 (1)	$Na_{96}\ [Al_{96}\ Si_{96}\ O_{384}]\ \cdot 150\ H_2O$	$F\,m\,\bar{3}\,c$ (No. 226)	a = 24.558 Å	14811 $Å^3$	67 %
ZSM-5 [6]	12	$[\ Si_{96}O_{192}]\ \cdot 4(CH_3CH_2CH_2)_4N$	P n m a (No. 62)	a = 20.063 Å b = 19.938 Å c = 13.409 Å	5364 $Å^3$	74 %
EMC-2 [7]	4	$Na_{11}\ [(Si,Al)_{96}\ O_{192}]\ \cdot 6\ H_2O$	$P\ 6_3/m\ m\ c$ (No. 59)	a = 17.378 Å c = 28.344 Å	7413 $Å^3$	83 %
VPI-9 [8, 9]	7	$(NH_4^+)_{24}\ [Si_{48}\ Zn_{12}\ O_{120}]\ \cdot 24H_2O$ [d]	$P\ 4_2/n\ c\ m$ (No. 138)	a = 9.895 Å c = 36.872 Å	3610 $Å^3$	47 %
VPI-10 [9, 10]	7	$(NH_4^+)_{16}\ [Si_{28}\ Zn_8\ O_{72}]\ \cdot 28\ H_2O$ [d]	I 2 m m (No. 44)	a = 12.599 Å b = 21.810 Å c = 7.022 Å	1930 $Å^3$	80 %
B2 [10, 11]	8	$K_4\ Na_4\ [Si_{16}\ Be_4\ O_{40}]\cdot 16\ H_2O$ [d]	$P\ 2_1\ m\ a$ (No. 26)	a = 13.173 Å b = 7.126 Å c = 12.678 Å	1190 $Å^3$	34 %

Table 1: Summary of characteristic data of the structures presented.
The first six are test cases and the last three novel structures.

[a] References: [1] Gerke & Gies (1984); [2] McCusker (1993); [3] Rohrig & Gies (1995); [4] Estermann, McCusker & Baerlocher, (1992); [5] Deroche, Marler, Gies, Kokotailo & Pennartz (1992); [6] Meier, Olson & Baerlocher (1996); [7] Baerlocher, McCusker & Chiappetta (1994); [8] McCusker, Grosse-Kunstleve, Baerlocher, Yoshikawa & Davis (1996); [9] Annen & Davis (1993); [10] Grosse-Kunstleve (1996); [11] Ueda, Koizumi, Baerlocher, McCusker & Meier (1986)

[b] #T = Number of T-atoms per asymmetric unit in the space group used. If different, the number in the topological symmetry is also given in parentheses.

[c] 100*N(overlap)/N(total) at a resolution of 1.3 Å

[d] estimated formula

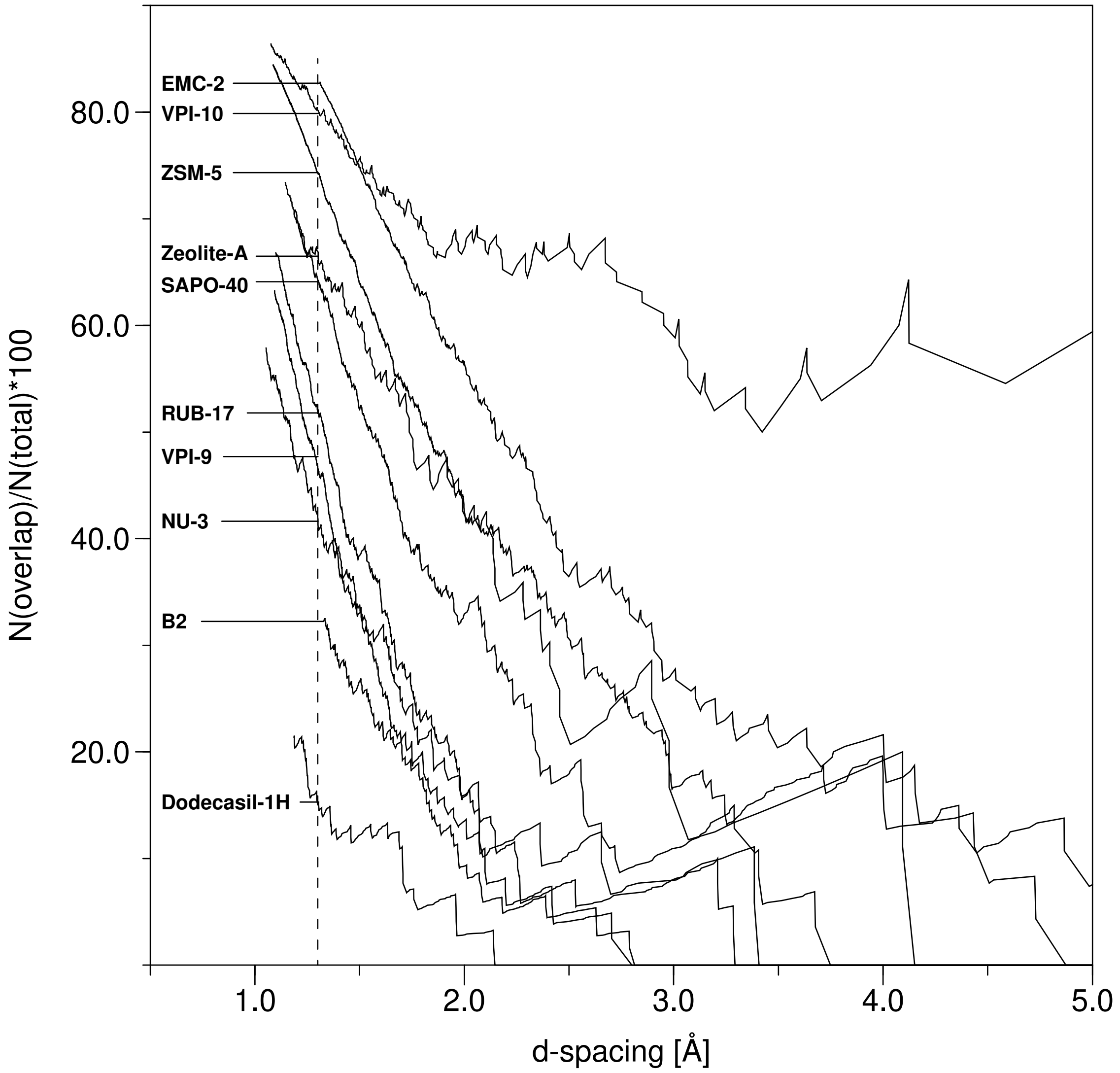

Figure 5: Overview of the overlap situation

A test example: SAPO-40

To show a structure determination in more detail, SAPO-40 can serve as an example. The synthesis of the SAPO-40 sample, the collection of data on a STOE laboratory diffractometer, and the structure solution from powder data with a combination of “Fast Iterative Patterson Squaring” and direct methods is described by Estermann, McCusker & Baerlocher (1992).

Integrated intensities were extracted up to a resolution of 1.19 Å using GSAS. Once the intensity scaling factor had been determined, a FOCUS input file was prepared (for a detailed description see Grosse-Kunstleve, 1996). Since the scattering powers of Si, Al, and P are only slightly different, only Si was used in the recycling. This is, in general, a proper approach for aluminophosphates. Only after the structure is known, can one introduce the strict Al-P alternation, which in many cases reduces the symmetry (see e.g. McCusker & Baerlocher, 1996, Simmen, McCusker, Baerlocher & Meier, 1991). Therefore the use of the “two color” framework search method is not recommended for aluminophosphates. Instead, FOCUS offers the `EvenLoopSizesOnly` option, which takes care of the fact that only even loop sizes are possible for structures with a strictly alternating occupation of the node positions.

The results of the runs are summarized in Tab. 2 and in the histogram in Fig. 6. It can be seen that the highest ranked topology was found more than 300 times, whereas none of the others occurred more than 5 times. Topology number 1 is, of course, the correct one.

Number of actively used reflections	314
Number of trials	1000
Number of Fourier transforms computed during the run	10975
Total number of acceptable frameworks found	338
Number of unique frameworks found	15
Total number of rejected frameworks	37756
Total computing time in minutes (MIPS R4400 CPU, 150 MHz clock rate)	1582
% of computing time spent for framework search	96

Table 2: FOCUS Results for SAPO-40 (AFR)

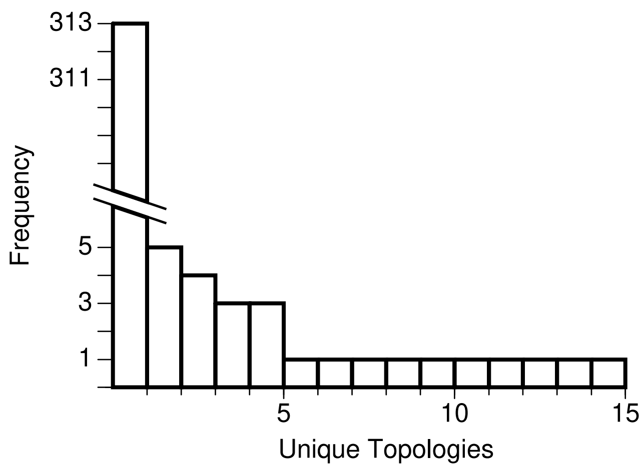

Figure 6: FOCUS Results for SAPO-40 (AFR)

Further development

Searching for non-tetrahedral node connectivities

All ten examples have one feature in common: a 3-dimensional 4-connected network of nodes was sought. To a certain degree, this is a consequence of the main idea which inspired the design of FOCUS: the integration of structural knowledge into the solution process. However, as FOCUS has been described, the specialization is extreme. To investigate the consequences of relaxing the structural assumptions, two further tests were conducted.

Searching for interruped frameworks

A FOCUS input file with simulated (equipartitioned) intensities was composed to determine the structure of Roggianite (**-RON**) (Giuseppetti, Mazzi, Tadini & Galli, 1991, IZA Structure Commission Report, 1994). The dash preceding the structure type code is used for “interrupted frameworks”, i.e. frameworks that are not fully 4-connected but have one or more nodes in the asymmetric unit which are connected to only three neighboring nodes.

100 trials were calculated in about 20 minutes (MIPS R4400 CPU, 150 MHz clock rate). The histogram shows a relatively clear discrimination between the most frequently occurring topologies, and the first histogram bar represents the **-RON** topology. No more tests have been carried out, but, based on this example, it can be assumed that a search for an interrupted framework is just as likely to give a solution as is a search for a fully 4-connected framework of the same complexity. However, in this simple example, the time spent for the framework search increased by a factor of 2.4, compared with that required for the search for a fully 4-connected net under similar conditions.

Searching for 3, 4, and 6-fold connectivities

A simple test of the feasibility of using the FOCUS algorithms for frameworks with node connectivities other than 4 was made with a gallophosphate structure known to have 3-, 4- and 6-connected atoms (Ga, P). The structure of this material had been solved from single crystal data by Chippindale, Walton & Turner (1995), and was picked at random during a search of the literature for open framework structures with *n*-connected nodes, where *n* is not only 4.

For testing purposes, FOCUS offers the possibility of bypassing the Fourier transform and peak search procedures, and starting with arbitrary peak positions in the peaklist. The backtracking procedure then works with the externally supplied positions. This feature was used for these tests. A series of peaklists were generated using the atomic coordinates from the single crystal refinement. In the first list only the nine node atoms were included, and then the number of positions were gradually increased until the last peaklist contained the whole structure (56 atoms).

To obtain a reference point for the search times, the search was first restricted to fully 4-connected frameworks. Then all three connectivity types were permitted, and the test series repeated.

It turned out that in both test series, the time required for the topology search increased approximately exponentially with the length of the peaklist. However, as a consequence of allowing connectivities other than four, the time to search a peaklist of the same size increased by a factor of about 30. Since the run times for complex structures are currently better measured in days rather than in hours, this factor increases the computing time from one day to one month. One way of overcoming such overwhelming time requirements is to work with a smaller peaklist, but then, of course, the success rate of the topology search drops accordingly.

Another weak point that should not be forgotten is that the possible node connectivities have to be prescribed *before* the structure is solved. Often one cannot be certain what connectivities to expect, and consequently has to allow for a wider range of possibilities. This results in an even further increase in computing time. To solve a structure like the gallophosphate from powder data using FOCUS would certainly require tremendous computing capabilities and effort. However, the “massive parallel” computers, equipped with several thousand processors, that are currently emerging would be very well suited for the algorithmic approach adopted by FOCUS, and might render attempts to determine structures of this complexity level successful.

Possible developments of FOCUS algorithm

FOCUS represents a purely algorithmic, heavily computer-based method. A source of information which was left untapped (except for a few tentative tests not reported here), is the exploitation of the statistics of the Fourier magnitudes, which is the foundation of direct methods. Rius & Miravitlles (1989) derived a new tangent formula, and, in contrast to that used in conventional direct methods, this has recently been shown to be applicable to low resolution (with respect to d-spacings) data (Rius, Vortmann & Gies, 1995). A combination of the “Fourier refinement” (recycling) of FOCUS and phase refinement with the new tangent formula offers tantalizing possibilities. An interesting aspect related to this is the fact that the proposed combination resembles recent developments in direct methods aiming at the determination of larger structures (e.g. “small proteins”) from single crystal data. In the “Shake-and-Bake” procedure presented by DeTitta, Weeks, Thuman, Miller & Hauptman (1994) and Weeks, DeTitta, Hauptman, Thuman & Miller (1994), phase refinement (“shake”) alternates with Fourier refinement “bake”. Similarly, Sheldrick & Gould (1995) have presented a procedure with alternating phase refinement and “peaklist optimization” (which they classify as “half baked” with reference to the Shake-and-Bake procedure). However, the powder specific difference between these procedures and the proposed combination of FOCUS and the Rius’ tangent formula, is a stronger enforcement of a prescribed class of structures at the Fourier refinement stage, and a significantly weaker demand for high resolution at the phase refinement stage.

At present, FOCUS only recycles phases derived from the automatically constructed models. However, it would also be possible to derive a new partitioning of overlapping intensities from the models. Experience shows that intensities play a vital role in the success rate. Obviously, those models that are in best agreement with the intensities have the highest chance of reproducing themselves. Of course, the correct model has no more chance of being randomly created than any other model, but once parts of it are present in the electron density map, the automatic Fourier recycling is likely to enforce it, while incorrect models are more likely to disintegrate. It is an open and highly interesting question, whether repartitioning of overlapping intensities during the recycling process would help to enforce the correct model, or whether it is more likely to “dilute” the already fragile intensity information extractable from a powder pattern.

Conclusions

The aim of this project was to incorporate some of the crystal chemical information used intuitively in model building into an automated structure determination procedure. It was hoped that this would allow more complex structures to be solved from powder diffraction data. Structural information, such as the types and numbers of atoms present, the expected connectivities, coordination numbers, and interatomic distances and bond-angles has been exploited to this end.

The FOCUS method was developed for the integration of zeolite-specific information. It makes extensive use of modern computer technology, and many substeps involve well established techniques, such as the conversion of powder data to a *pseudo* single crystal data set. The conventional treatment of the *pseudo* single crystal data is replaced, or enhanced, by a combination of automatic Fourier recycling and a topology search. Finally, the usefulness of the

FOCUS procedure has been demonstrated by its successful application in the structure determination of three complex novel zeolite structures, where only powder data were available.

Experience gathered during the course of this project shows that the methodologically attractive approach of using chemical and geometrical knowledge can compensate for some of the information lost as a result of the overlap problem. At the same time, there is an intrinsic disadvantage: any method based on assumptions of certain structural properties is also limited to materials which conform to these assumptions. Unlike direct methods, which only make assumptions valid for all X-ray diffraction experiments, the consideration of more specific structural information also introduces a certain specialization. However, from the outset it has been foreseen that the basic idea – the integration of structural assumptions into the solution process – should also be applicable to other classes of materials. Two short examples have been presented, which show the consequences of relaxing the structural assumptions to allow solution attempts for non-4-connected frameworks. It was found that the computing time requirements of FOCUS grow very rapidly with the number of different possible connectivity types. Suggestions for further developments to overcome this problem are outlined, and it is hoped that some of the experience gathered in the development of FOCUS contributes to the evolution of a more generalized mechanism.

Acknowledgments

The powder diffraction data used for VPI-9, VPI-10 and ZSM-5 were collected on the Swiss-Norwegian Beamline at the European Synchrotron Radiation Facility in Grenoble and those for B2 at the HASYLAB facility in Hamburg. We thank Phil Pattison, Andy Fitch, Kenneth Knudsen and Thomas Wroblewski for their assistance with these measurements. This work was supported in part by the Swiss National Science Foundation.

References

Annen, M.J. & Davis, M.E. (1993). Microporous Materials 1, 57-65.

Baerlocher, Ch., Hepp, A., Meier, W.M. (1977). DLS-76 – A Program for the Simulation of Crystal Structures by Geometric Refinement, Institute of Crystallography and Petrography, ETH Zurich, Switzerland.

Baerlocher, Ch., McCusker, L.B. & Chiappetta, R. (1994). Microporous Materials 2, 269-280.

Bricogne, G. (1991). Acta Cryst. A47, 803-829.

Brunner, G.O. & Laves, F. (1971). Wiss. Z. Techn. Univers. Dresden 20, 387-390.

Cascarano, G., Favia, L. & Giacovazzo, C. (1992). J. Appl. Cryst. 25, 310-317.

Chippindale, A.M., Walton, R.I. & Turner, C. (1995). J. Chem. Soc., Chem. Commun., 1261-1262.

David, W.I.F. (1987). J. Appl. Cryst. 20, 316-319.

David, W.I.F. (1990). Nature 346, 731-734.

Deroche, C., Marler, B., Gies, H., Kokotailo, G.T. & Pennartz, P.U. (1992). Z. Krist. Suppl. 5, 46.

DeTitta, G.T., Weeks, C.M., Thuman, P., Miller, R. & Hauptman, H.A. (1994). Acta Cryst. A50, 203-210.

Estermann, M.A. (1991). Ph.D. thesis, ETH Zurich, Switzerland.

Estermann, M.A., McCusker, L.B., & Baerlocher, Ch. (1992). J. Appl. Cryst. 25, 539-543.

Estermann, M.A. & Gramlich, V. (1993). J. Appl. Cryst. 26, 396-404.

Estermann, M.A. (1995). Optimised Wilson normalisation of integrated Bragg intensities from powder diffraction data lacking atomic resolution. Presented at: IUCr Workshop "Structure determination from Powder Data", Wadham College, Oxford, U.K.

Fischer, W. (1973). Z. Krist. 138, 129-146.

Fischer, W. (1974). Z. Krist. 140, 50-74.

Gerke, H. & Gies, H. (1984). Z. Krist. 166, 11-22.

Giacovazzo, C. (Ed.) (1992). Fundamentals of Crystallography. IUCr/Oxford University Press.

Gilmore, C.J., Henderson, K. & Bricogne, G. (1991). Acta Cryst. A47, 830-841.

Giuseppetti, G., Mazzi, F., Tadini, C. & Galli, E. (1991). Neues Jahrbuch für Mineralogie 7, 307-314.

Goldberg, D.E. & Segrest, P. (1987). Finite Markov chain analysis of genetic algorithms; In: John J. Grefenstette, Ed.; Genetic algorithms and their applications: proceedings of the Second International Conference on Genetic Algorithms, July 28-31, 1987, at the Massachusetts Institute of Technology, Cambridge, MA, Hillsdale, N.J.: Lawrence Erlbaum.

Goldberg, D.E. (1989). Genetic algorithms in search, optimization, and machine learning, Reading, MA: Addison-Wesley.

Grosse-Kunstleve, R.W. (1996). Ph.D. thesis, ETH, Zurich, Switzerland.

Hedel, R., Bunge, H.J. & Reck, G. (1994). Mat. Sci. Forum 157-162, 2067-2074.

Larson, A.C. & von Dreele, R.B. (1995). GSAS - General Structure Analysis System, Los Alamos National Laboratory, NM 87545, USA.

IZA Structure Commission Report (1994); Zeolites 14, 389-392.

Jansen, J., Peschar, R. & Schenk, H. (1992). J. Appl. Cryst. 25, 237-243.

Le Bail, A., Duroy, H. & Fourquet, J.L. (1988). Mat. Res. Bull. 23, 447-452.

Loiseau, T. & Férey, G. (1994). J. Solid State Chem. 111, 403-415.

McCusker, L.B. (1993). Mat. Sci. Forum 133-136, 423-434.

McCusker, L.B. & Baerlocher, Ch. (1996). Microporous Materials 6, 51-54.

McCusker, L.B., Grosse-Kunstleve, R.W., Baerlocher, Ch., Yoshikawa, M. & Davis, M.E. (1996). Microporous Materials 6, 295-309.

Meier, W.M., Olson, D.H. & Baerlocher, Ch. (1996). Atlas of Zeolite Structure Types, 4th ed., London: Elsevier (Zeolites 17, 1-230).

Morris, R.E., Harrison, W.T.A., Nicol, J.M., Wilkinson, A.P & Cheetham, A.K. (1992). Nature 359, 519-522.

Rius, J. & Miravitlles, C. (1989). Acta Cryst. A45, 490-494.

Rius, J., Sane, J., Miravitlles, C., Gies, H., Marler, B. & Oberhagemann, U. (1995). Acta Cryst. A51, 840-845.

Rius, J., Vortmann, S. & Gies, H. (1995). Solution of zeolite precursors from low-resolution X-ray powder diffraction data. Poster abstract presented at: IUCr Workshop "Structure determination from Powder Data", Wadham College, Oxford, U.K.

Rohrig, C. & Gies, H. (1995). Angewandte Chemie – Int. Ed. English 34, 63-65.

Rollet, J.S. (Ed.) (1965). Computing Methods in Crystallography, Oxford: Pergamon Press.

Sheldrick, G.M. & Gould, R.O. (1995). Acta Cryst. B51, 423-431

Simmen, A., McCusker, L.B., Baerlocher, Ch. & W.M. Meier (1991). Zeolites 11, 654-661.

Ueda, S., Koizumi, M., Baerlocher, Ch., McCusker, L.B. & Meier, W.M. (1986). Preprints of Poster Papers, The 7th International Zeolite Conference, p. 23.

Weeks, C.M., DeTitta, G.T., Hauptman, H.A., Thuman, P. & Miller, R. (1994). Acta Cryst. A50, 210-220.

Wirth, N. (1986). Algorithms and data structures, Englewood Cliffs, N.J.: Prentice-Hall