How to Find Correlations Between Molecular Shape and Packing in a Molecular Crystal: Application of a Novel Strategy to Recognize *n*-Point Polyhedra in Three-Dimensional Space

Stefan Reichling[a] and Gottfried Huttner*[a]

Dedicated to Kurt Mislow, Princeton

Keywords: Crystal packing / Neural networks / Pattern recognition / Organometallic compounds / Close packing

The crystal structures of hundreds of thousands of molecular compounds have now been determined with the quantitative information being stored in large databases. Chemists are generally interested in that specific part of this information which refers to individual molecules. The packing of molecules and their spatial relation to each other is not the focus of interest, even though the material properties of a molecular crystal are determined by both the packing arrangement of the molecules and their specific properties. The lack of interest in packing has at least one obvious reason. It is difficult for the non-expert to find the appropriate packing categories when looking at a principally infinite lattice. A tool performing this search for individual categories of packing in an automatic manner would therefore be of general importance. The development and application of such a tool is described in this paper. Novel strategies to extract a set of n points forming a specific n-point polyhedron from a set of p points ($p > n$) is at its basis. Neural networks and second moment analysis are used as the methods of pattern recognition. In order to correlate packing and shape, second moment analysis of molecular shape is also used. This novel method is applied to three classes of organometallic compounds to determine whether the crystals formed by these compounds belong to any of the three classes *fcc*, *bcc*, or *hcp* in an idealized sense. A strong correlation is found between molecular shape and the membership or nonmembership of these classes. The corresponding program, including appropriate tools for visual representation, is available from the authors.

Introduction

The molecular concept is the basis of chemical reasoning. Even though it is quite complicated to define what a molecule really is, a well-trained chemist will immediately associate a structural representation with this idea and he or she may even take this representation as the idea itself. This kind of reduction has been, and still is, the basis of much of the development of the science of chemistry.

While in earlier days the structures of molecules had generally to be inferred from indirect reasoning, today there is an abundant wealth of structural information available, well-encoded in databases. As this information comes from three-dimensional diffraction analysis, the three-dimensional structures of all the molecules contained in such databases i.e. the coordinates of individual atoms are readily available, as is the mutual arrangement of these molecules within a crystal. While it is evident that such databases add much to the convenience with which a specific structure and even a specific substructure can be found, it is also clear that a collection of very many structures holds a lot of additional information. This additional information is encrypted in the database and one has to ask the proper questions and to apply the proper tools to extract it.

Perhaps the best known example for the validity of this statement is the mapping of a reaction pathway in the early and seminal work of Buergi and Dunitz.[1] The experimental basis for this elegant type of analysis is that the structures of many individual compounds are available within the class of compounds which are relevant to the problem. In another type of application, the conformations adopted by the many individual molecules within one class of compound may be used to find a mechanical model which correctly reproduces and also predicts the conformation of the members of this class.[2] In principle this type of database use builds on the existence or development of intelligent filters that are able to extract the desired information from the wealth of data available. The structure of these filters may be quite different, depending on whether one is looking for reaction pathways[1] or trying to develop a mechanical model.[2] Still, the basic logic is the same: look at the individual structures as snapshots of the geometry of a class of molecules and try to correlate the observed variations with some set of parameters.

It is clear from the above that extracting the hidden information calls for the appropriate tools. As far as molecular structures themselves are concerned, retrieval systems, such as the one built into the most abundant database, the Cambridge Structural Database[3] (hereafter named CSD), are well suited to select classes of compounds, specific kinds of interactions and also specific geometrical arrangements. The situation is much worse if the packing of individual molecules within a crystal is the focus of attention. Although plenty of graphical tools are available to visualize the arrangement of molecules within a three-dimensional

[a] Anorganisch Chemisches Institut der Universität Heidelberg, Im Neuenheimer Feld 270 D-69120 Heidelberg, Germany
Fax: (internat.) + 49-(0)6221/545-707
E-mail: g.huttner@indi.aci.uni-heidelberg.de

  1434−1948/00/0505−0857 $ 17.50+.50/0

lattice, it is very difficult for the non-expert to understand such packing diagrams. "Understanding" in this context means that one would be able to delineate the observed pattern into some archetypal arrangement and thus recognize the basic principles. Projecting this situation onto molecular chemistry: imagine someone who looks at the structure of hexacarbonylchromium and is not able to extract the archetypal octahedron, which is of course apparent to an educated chemist at first glance. The archetypes of arrangements within infinite lattices are more complicated to recognize, one reason being the translational symmetry, and the other being the lack of appropriate training. While solid-state chemists and crystallographers are well trained to reduce a given repetitive architecture to some standard pattern, molecular chemists are generally not so. Looking at any introduction to solid-state chemistry,[4] one will be aware of the power of categorizing compounds by idealizing their geometry such that they may be understood as derivatives of a few basic patterns. Closest packing, hexagonal close packing (*hcp*), cubic close packing (*fcc*) and body-centered packing (*bcc*), are at the basis of the most proliferate idealizations of this type. Referring to the packing of molecules within crystals it has been shown by Kitaigorodskii[5,6] that the idea of closest packing helps to rationalize the individual arrangements.

Definition of the Problem

With the development of what is called supramolecular chemistry,[7–9] i.e. with the increasing awareness of the importance of secondary interactions between molecules, any information about these secondary interactions is more important to the development of chemistry than ever. Structural databases[3] contain an as yet scarcely used and abundant source of information about secondary interactions. It is these secondary interactions which govern the arrangement of molecules within a crystal.

One way to extract part of this information relies upon reconstruction of the observed patterns of packing by making the appropriate choice for the individual terms of the intermolecular potential.[10,11] By doing this for a series of related structures the potential terms may be optimized.[12,13]

Choosing a good starting point for this kind of optimization is guaranteed by yet another method of systematic exploration of aspects of molecular packing: it is found that the centers of gravity of molecules tend to be close to specific compartments of a unit cell.[14] Helpful as this analysis is, it has no direct bearing on the packing of molecules since the position of points in space depends on their coordinates as well as on the metric system to which these coordinates refer. There are – not including translation – an infinite number of three-dimensional possibilities for a set of points if the fractional coordinates are all that is known, since the real position of a point in space is dependent on the lengths of the crystal axes and the angles between them.

Another way would entail the search for specific patterns of packing (c.f. *bcc*, *fcc*, ...) of the molecules as represented by their centers of gravity. Correlation of the appearance of these patterns with parameters describing the shape of the molecules may then be attempted. Basically, this type of approach has been taken by Kitaigorodskii,[6] albeit at a time where computer retrievable databases were not yet common. To make this approach feasible and efficient, tools have to be developed which recognize a specific three-dimensional repetitive pattern automatically. Such tools are still not available, and although it may appear a simple task to use today's pattern recognition techniques to develop such a tool, the problem is quite complicated.

While, given the metric system, the coordinates of n points completely define an n-point pattern, the values of these coordinates depend on translation as well as rotation. A direct search for a specific n-point pattern in three-dimensional space would hence be quite cumbersome. A solution to this problem is to transform the observed patterns into some representation that is independent of rotation and translation, i.e. a representation in internal coordinates. Since the choice of internal reference frame is intentional, this representation of a pattern formed by an ensemble of n points is again not unique and is thus not immediately accessible to analysis by pattern recognition techniques. To make the representations independent of the specific assignment of internal coordinates – i.e. independent of the $n!$ possible choices of labeling n individual points, each by a unique label – instead of analyzing the individual values, which the internal coordinates adopt for a specific arrangement of points, the distribution of these values may be analyzed. Such a distribution may immediately serve as an input to pattern recognition tools. While representing an ensemble of n points in three-dimensional space by a few histographic data does not in general allow reconstruction of the body represented by these points, it is a very efficient way to classify arrangements of n points with respect to symmetry. Different realizations of this type of analysis are described in this paper, i.e. different types of histographic mapping are used as an input to pattern recognition tools. Neural networks[15–18] are trained to differentiate between different types of bodies based on this histographic input.

The problem of analyzing for a specific type of packing within a repetitive pattern, i.e. crystal space, is dealt with in two steps: first the m next neighbors of a given point are determined; this subset of points is then analyzed for the presence of specific polyhedra which are a characteristic consequence of the kind of packing. Thus for instance octahedra are characteristic constituents in a close-packed situation.[19–21] In looking for octahedra, all six-point combinations out of the ensemble of $m + 1$ points (a good value of m being 16 in the search for octahedra in a close packed situation) are formed and automatically analyzed with respect to their potential octahedral arrangement. The few octahedra sorted out from the $\binom{k}{6}$ (read k over six; with $k = m + 1$) point combinations are then further analyzed with respect to their connectivity. From the type of connectivity found, the type of packing (*fcc*, *hcp*, *bcc*) which best approximates the given distribution of points in crystal space is derived.[19–21]

The procedure is applied to a subset of compounds crystallizing in space group *R*3 as well as to a family of tripod–metal complexes $RCH_2C(CH_2X)(CH_2Y)(CH_2Z)ML_n$ (R = H, alkyl, aryl; X, Y, *Z* = NR_2, OR, SR, PR_2, PRR′) and a selection of $Cr(CO)_5L$ compounds as registered in the CSD[3] (see Experimental Section). Packings which are close to *fcc*, *hcp*, *bcc* drop out automatically by this procedure. Attempts are made to correlate the shape of the molecules with the specific packing observed. It is found that describing the shape in terms of the eigenvalues of the tensor of inertia,[22] i.e. second moment analysis[23] of the molecular shape, produces some insight into the correlation between packing and shape. Second moment analysis is also used as a tool to differentiate different types of arrangements of *n* points in space, i.e. to recognize specific types of polyhedra. It is found to be efficient in this aspect as well and its use combined with neural network analysis produces an efficient filter for sorting out specific *n*-point polyhedra from a given ensemble of *p* points ($p > n$).

Results and Discussion

In this first attempt at automatic pattern recognition in crystal space, the search was limited to the three basic close packing motifs *fcc*, *hcp*, *bcc*. On the other hand, the pattern recognition machinery was designed such that it might also deal with any other type of packing motif.

From the two alternatives of searching in direct or reciprocal space, the first was preferred. If the only goal of analysis of packing motifs was the search for approximate close packed arrangements then search in reciprocal space might be quite effective. Replacing the molecules by atoms at their centers and calculating the diffraction pattern produced by these atoms would map the packing onto reciprocal space. A search for the approximate symmetry of the diffraction pattern and approximate rules of systematic extinction would then reveal the atomic packing pattern in direct space. This kind of analysis, while presumably efficient in the search for close packing motifs, would not easily adapt itself to the search of less specialized packing motifs. Searching for patterns in direct space was therefore preferred.

It might appear that with only three close-packed motifs in mind, a search for characteristic polyhedra such as the cube octahedron (*fcc*) or the anti-cube octahedron (*hcp*) might be the topic of focus. It is however found, as may be understood by imagining these polyhedra,[19–21] that differentiation between these two bodies as a whole is relatively poor. An alternative strategy has therefore been adopted, to search for octahedra and their connectivity from which the packings may be delineated.

Search for Polyhedra

Distance Histogram

An *n*-point polyhedron may be represented by *n*-coordinate triples in three-dimensional space. Since the values of these coordinates are dependent on translation and rotation, a representation in internal coordinates is preferred for pattern recognition purposes. An especially simple representation of this type consists of a vector for the [*n*(*n* – 1)/2] distances between all pairs of points of the *n*-point ensemble. This vector contains a full description of the polyhedron and may be back-transformed to a corresponding set of coordinates in three-dimensional space which represent the polyhedron. This fact is made use of in distance geometry calculations.[25,26] That the vector holds all the information necessary to reconstruct the body is immediately apparent from the observation that there are [*n*(*n* – 1)/2] distances and 3*n* – 6 degrees of freedom for a body with $n > 2$ points. The number of distances will be equal to the degrees of freedom for $n = 3$ and 4, and larger for $n > 4$, meaning that in these cases ($n > 4$) not all of the distances are independent. In each case, the number of independent distances is sufficient to reconstruct the polyhedron. Such a vector containing all [*n*(*n* – 1)/2] distances is, however, not suited for presentation to a simple pattern recognition algorithm since it may occur in *n*! different forms corresponding to the *n*! permutations of the *n* components in this vector. Only if the labeling scheme to which the vector refers is known, i.e. when the specific connectivity to which it refers to is known, will back-transformation to the coordinates of the *n* points be possible. To create a description of the polyhedron, which is independent of the specific numbering scheme, a histographic representation of

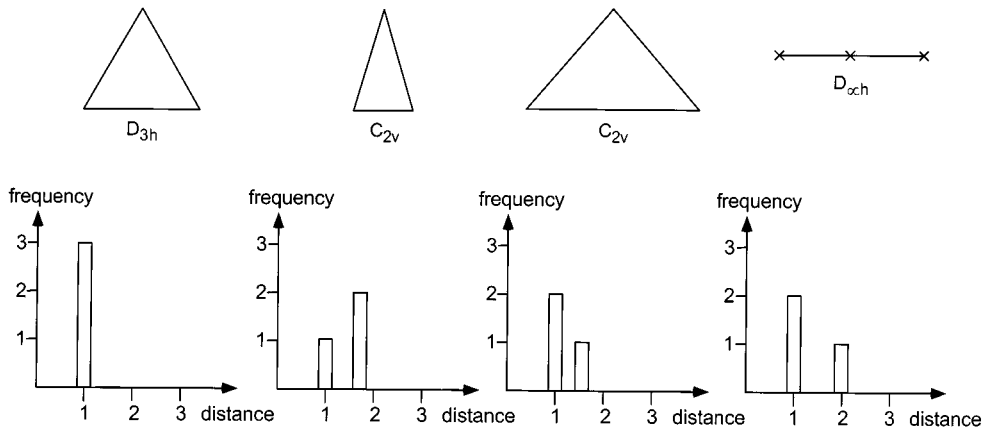


Figure 1. Histographic representation of the effects of symmetry on distance histograms for a set of three points

the vector may be chosen. That is, a histogram showing the frequency with which distances in a specific range occur may be used as input for pattern recognition machinery.

This type of histogram, even though it cannot in general be back-transformed to a three-dimensional representation of the body, is quite characteristic of the individual types of polyhedra. The reason for this is that the symmetry of the polyhedra is encoded in this type of histogram. As an example think of three points in space.

Any pattern formed by three such points is completely characterized by the three distances between these points. While in general these distances may map onto the histogram at any three positions, an arrangement of the points

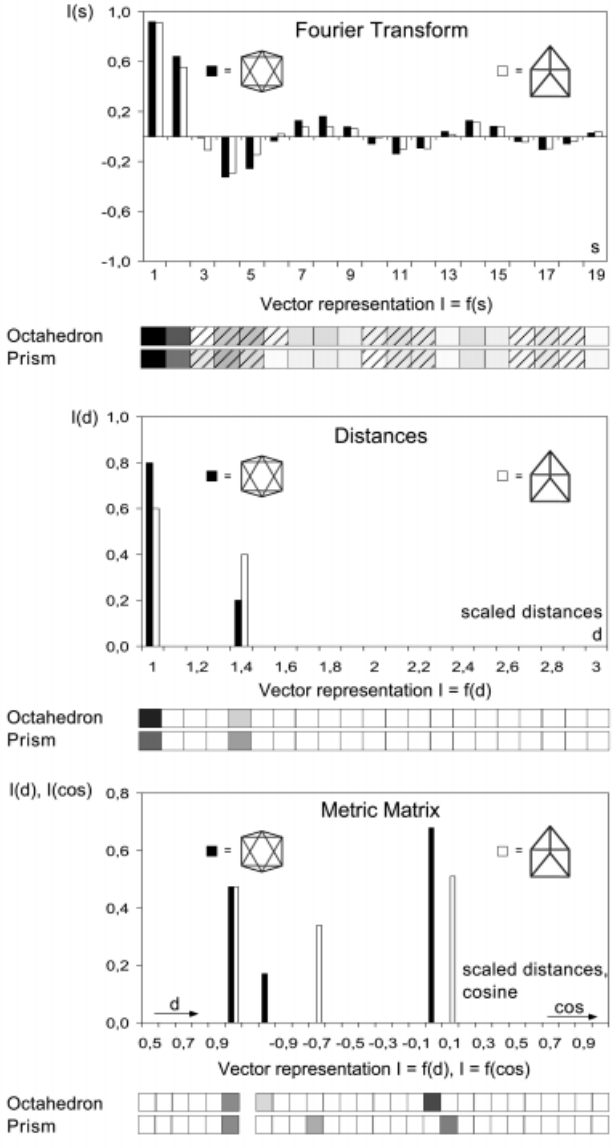


Figure 2. Three methods to represent three-dimensional bodies as a discrete function of just one independent variable. The procedures of obtaining the functions are described in the text. The functions are shown for the examples of a regular octahedron and a regular trigonal prism. The boxes below each function represent the individual components of the vector onto which the functions are mapped. The relative magnitude of the coefficients is indicated by gray shading with black being largest and white being smallest. Negative values of coefficients are indicated by dashed lines.

with any symmetry higher than C_1 will show a specific pattern. If the points conform to an equilateral triangle, the C_3 symmetry of this D_{3h} symmetric body will map itself into the histogram such that the shortest distance is observed three times and there is just one bar in the histogram (Figure 1 left). An isosceles triangle will be represented by two bars, the one corresponding to the two equal distances being twice as high as the one representing the third distance (Figure 1 middle). In the linear arrangement of the $D_{\infty h}$ borderline case of an isosceles triangle, the distance occurring only once will be twice as large as the shortest distance which occurs twice (Figure 1 right). This simple example shows how symmetry mirrors itself in the histographic representation. In this specific case of three points in space, the histographic representation is an unambiguous representation of the body itself. This is because the number of distances is equal to the number of degrees of freedom in this case. The same applies to four points in space, while with $n > 4$ the number of distances exceeds the number of degrees of freedom. In these cases ($n > 4$) the histogram is no longer unambiguous although it still responds in very specific ways to the symmetry of the arrangement of points in space. An example of this type of mapping is represented in Figure 2 (middle) for an octahedron and a trigonal prism.

Histogram for Encoding Distances and Angles: The Metric Matrix Approach

As has already been discussed, a complete set of [$n(n - 1)/2$] distances is a full representation of a body composed of n points in three-dimensional space if the connectivity is also known. It is a representation independent of translation and rotation. Another representation, which is also free of the ambiguities of translation and rotation, consists in the so-called metric matrix of the body (Figure 3 top). To construct this matrix the center of mass of the n points is taken as the origin. The scalar products of all the n vectors radiating from the origin to the n points may then be represented as a symmetric matrix that contains the squares of the distances on the diagonal, and the cosines of the angles sustained by each pair of vectors – each cosine multiplied by the product of the relevant distances – as the off-diagonal terms (Figure 3 top). The appearance of this matrix is also dependent on the scheme used to label the n points and the lines and columns of this matrix may be shuffled around by the $n!$ permutations of the labeling scheme. The matrix itself is thus not immediately suitable for input to a pattern recognition algorithm but the histogram derived from it is. The histogram now consists of two parts, one referring to the distances and the other one to the cosines of the angles (Figure 3 middle). The distances d are mapped onto the interval $0 < d \leq 1$ with the largest distance scaled to 1. Scaling of the cosines is unnecessary since their values are between –1 and +1. The frequencies are scaled such that the values fit in the range $0 \leq$ frequency ≤ 1 (see Experimental Section). The two parts of the histogram (distance, cosine) may be encoded as a vector once the resolution of its two parts (distance, cosine) is defined (Figure 3

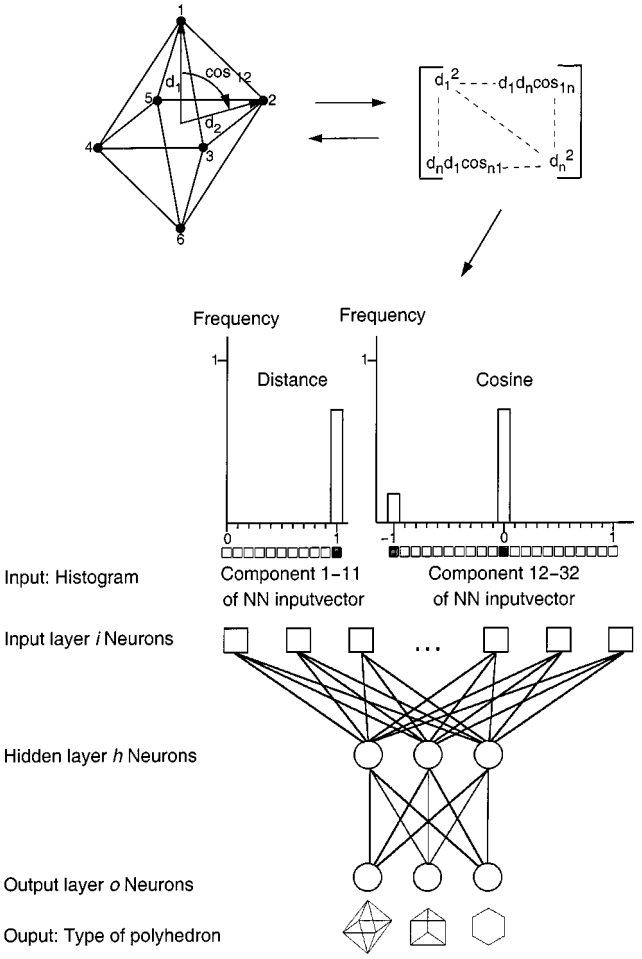


Figure 3. Sorting of *n*-point polyhedra into specific categories by neural networks. The top half of the diagram gives a shorthand illustration of how a vectorial representation of an ensemble of *n* points in space is obtained. An octahedron is deliberately selected as an example. The bottom half illustrates the type of network used. The input layer consists of $i = 32$ neurons to present the vector to the network. A hidden layer of $h = 3$ neurons connects this input layer to the output layer with $o = 3$ neurons. The network is trained to sort the bodies with respect to the class to which they have been assigned to (octahedron, trigonal prism, hexagon).

middle, Figure 2 bottom) and the vector may then be used as the input for a pattern recognition algorithm.

The histograms discussed are a projection of three-dimensional data onto a histographic function with just one independent variable. The advantage of this type of representation is that it is independent of translation and rotation of the body as well as of any permutation of the labeling scheme used to designate the points. Although much information is lost by projecting three-dimensional data this way, the patterns obtained do characterize the symmetry properties of the three-dimensional object (see above).

Fourier Transformation

A different type of projection from three-dimensional space onto a function with just one independent variable consists of a Fourier transformation of the distances spanned by the *n* points of the body. This type of transformation has a long-standing application in electron diffraction[27] and has recently been applied to the condensed description of molecular shape and constitution.[28,29] Application of this procedure is simple as it means to compute the sum

$$I(s) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} A_i A_j \frac{\sin sd_{ij}}{sd_{ij}}$$

where *s* refers to the "scattering angle". If the distances *d* are measured in Angstroms its dimension is $Å^{-1}$. Computing this sum for $1 \leq s \leq 50$ produces values of this function at each of these 50 points in *s* space. The function itself is a weighted sum of sine functions of different periodicity. It is found, that for the problems analyzed, as little as the first 19 values of this function are sufficient due to the almost exponential decay of $I(s)$ versus *s*. A vector containing these 19 values of $I(s)$ at $s = 1,2,..,19$ may serve as input to a pattern recognition algorithm (Figure 2 top, $A_i = A_j = 1$ for $i,j = 1...n$).

A comparison of the three methods described, which are able to create specific one-dimensional projections of a set of *n* points in three-dimensional space, is shown in Figure 2. The bodies compared in this example are a regular octahedron and a trigonal prism with all edges of equal length. It can be seen that all three methods are able to differentiate these two types of body. Using the distances only, the discriminative power of the histogram in direct space and its Fourier transform equivalent in reciprocal space do not appear to be very different (Figure 2 top and middle). In fact, it has been found that the simple histographic representation of the distances in direct space is already an efficient tool, so the Fourier transform version of the distance-based code has not been used to a large extent in this work. In addition to the fact that the discriminative power of the Fourier transform representation is not significantly superior to the representation by the distances themselves, it lacks the immediate interpretability of the histogram of distances. In a sense the histogram of distances corresponds to the radial distribution curve to which diffraction data are transformed in order to extract structural information, while the Fourier transform[27] representation corresponds to the diffraction pattern itself. As is seen in Figure 2 a histographic representation of the metric matrix encoding for distances and angles has the highest power of discrimination of the three methods analyzed and is by far preferred.

All the above one-dimensional projections may be seen as kind of spectral representation of the three-dimensional bodies they represent. Specific signatures of these spectra may be interpreted as a kind of fingerprint indicating that some specific body, say an octahedron, trigonal prism, hexagon, tetrahedron, etc. is the origin of this fingerprint. While it is easy to correctly interpret these spectra if the ideal symmetry-defined bodies are at their origin, it is not so simple to extract this type of information by mere inspection if a distorted polyhedron is the origin of the spectra. In fact, if we draw away from the scalar differences for the moment, there is only one octahedron, an ideal body defined by symmetry alone. In the real world, however, there are many arrangements of six points in space which are

close to an octahedral arrangement. Any set of rules defining the extent to which a body may be distorted and still be categorized as belonging to the octahedral class is artificial, and the rules may well change in response to the questions asked. Instead of defining a set of rules in mathematically expressed restrictions of angles and distances, we adopt a different philosophy. It is argued that no arrangement of six points not completely defined by octahedral symmetry is a true octahedron.[30] Nature confronts us with all kinds of six point arrangements but we are still intuitively able to classify these arrangements as say octahedral, trigonal prismatic, and hexagonal by mere inspection of the bodies. Instead of using artificial mathematical rules for the purpose of classification of these bodies we may well rely upon our intuitive interpretation: We look at a body and see it is octahedral. We look at another one and see it is close to a plane hexagon and so on. We then classify the bodies according to what we see. It may be argued that this way of classification lacks any objectivity. However, there is no objectivity in mathematically formulated and artificial classification rules. Adopting the totally subjective view mentioned above introduces the unconventional consideration of personal judgment in science. However, if "truth" is only in the symmetry-defined ideal objects, anything outside this is a matter of taste and purpose. One might like to introduce some kind of objectivity in the manner of social science or psychology: one might present the bodies to be categorized to as many people as time allows and then base classification on a common sense evaluation. Strange as this point of view may appear when uttered in a scientific paper, the thorough logical evaluation of the situation will show that it is completely appropriate to the purpose.[30] The purpose of classification described in this paper is to classify the packing of molecules with respect to categories, which again, not considering scale, are ideals, defined by symmetry alone. Any classification of an observed packing as a distorted version of the ideal is not objective in a scientific sense and is very much biased by our desire to find this relation to one or the other of the ideal patterns.[31]

Neural Network Analysis

Instead of defining a mathematical set of rules to explicitly sort the bodies into different categories, neural networks[15–18] have been trained to choose these categories on the basis of the input vectors of the spectra described above. To train the networks a training set has to be constructed which contains not only the ideal symmetry determined bodies, but also bodies derived from these by random perturbation. To this end, the points defining the ideal body were allowed to occupy positions at random within a perturbation sphere located at the corresponding point of the ideal body. The bodies thus generated where inspected by examining different orientations of their representations on a computer screen and then categorized by inspection. The perturbation spheres were enlarged in a stepwise manner until the chance of finding a perturbed structure showing an apparent relation to the ideal structure was impracticably low e.g. no chance in ten trials. In this way, a training set of 281 polyhedra altogether was produced. The training set comprised the following classes, $n = 4$ square planar, tetrahedral, $n = 5$ pentagon, square pyramidal, trigonal bipyramidal, $n = 6$ hexagon, octahedron, trigonal prism. To analyze the whole set of data with a single network, the vectorial representation of the histograms referring to distance (Figure 2 middle) was augmented by a single coefficient which presented the number of points n underlying the pattern to the network. $1/n$ was used to represent this number. These vectors were then fed into the neural network, which was trained to sort the input into the correct output categories. To validate the trained network function an independent set of 64 distorted polyhedra, comprising all categories to which the network had been trained, was generated as above. It was observed that the categories found by the network were equal to those defined by inspection in 78% of the cases. Some of the "faults" committed by the network are easily explained: A square pyramid may be interpreted as an octahedron due to the fact that neural networks are capable of reconstructing patterns when some of the input data is missing or at error.[17,18] The $1/n$ term (see above) in the input vector did not adequately counter this tendency to pattern reconstitution. Therefore it was decided to use specific pattern recognition machines for every number n of points analyzed and it appeared reasonable to also use the spectral input derived from the metric matrix (Figure 2 bottom). Due to its greater discriminative power, this might not only improve the quality of predictions of the trained network, but it should also abbreviate the time necessary for training the network. The procedure adopted is schematically shown in Figure 3.

For any given arrangement of six points in the data set the metric matrix is calculated. The distances obtained from the diagonal terms of the matrix and the cosines derived from the off-diagonal terms are projected onto histograms, one for the distances and a second for the cosines. A vector representation of these histograms is obtained by discretisation of the distance part (the distances are scaled between zero and one) into 11 intervals and the cosine part into 21 intervals (Figure 3 middle). The vector of these 32 components is used to feed the input layer of a neural network that contains 32 input units. The network consists of this input layer, one hidden layer and one output layer. Whenever an input vector representing a body in the training set is fed into the network, it is simultaneously informed which of the three categories the body belongs to, and how far it is from the ideal. The corresponding qualifier between zero and one (zero no fit, one ideal fit) was obtained by visual inspection (Figure 4) of the body.

Training the network requires feeding the network with the input data and the expected results over and over again until there is no more change between the expected output and the output generated by the network in successive training cycles. At this point of convergence the strength of the connections between the individual neurons (lines in the diagram Figure 3) has been optimized such that the network as a whole automatically sorts the bodies into the cat-

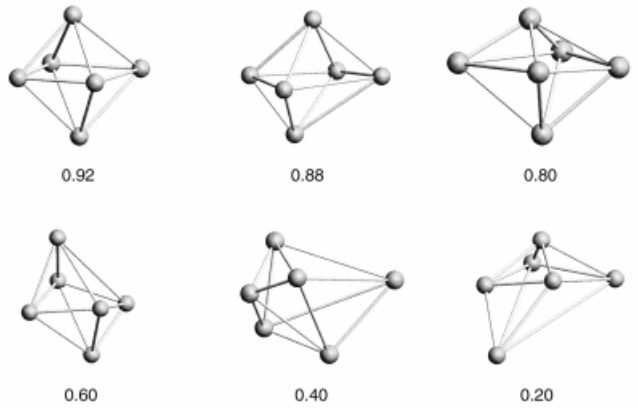


Figure 4. Distorted octahedra and their qualifiers. A few selected examples are shown. The qualifiers q were allotted by visual inspection. For the purpose of training the network only octahedra with qualifiers q greater than or equal to 0.8 were used.

egories in the best way it can. If the network is able to reproduce the artificial categorization of the bodies contained in the training set in an almost quantitative manner, it may be said that the network – with respect to the training bodies – "sees" what a human inspector would see when looking at the bodies themselves. The network is then able to reproduce the categories and qualifiers a human being would allocate by inspection, even for bodies it has never "seen" before. This means the machinery is capable of categorizing bodies in the same manner a human inspection of the bodies would. That is, the time-consuming task of inspecting a three-dimensional body by eye may be transferred to the neural network that produces the same result a human inspector would produce. This protocol was used for the subset of the polyhedra with $n = 6$. Not only was it found to lead to rapid convergence but it also led to a high quality of prediction. Based on a training set of 153 input vectors distributed over the three categories to be identified (octahedron, trigonal prism, hexagon) the square root of the mean standard error (see Experimental Section) after convergence of the training phase was only 0.017. To validate the trained network function, an independent set of 16 distorted polyhedra was generated as above, comprising all categories for which the network had been trained. It was observed that the categories found by the network were equal to those defined by inspection with a deviation between the predetermined and the calculated qualifiers of only 4.5%. Neural networks are thus found to be a sufficient filter to recognize the specific shape of an n-point body from a set of coordinates containing all kinds of n-point arrangements.

Second Moment Analysis

Since it had been found that the manner in which three-dimensional data is reduced to form a vector appropriate for input to the neural network had a large influence on the discriminative power of the trained network, another and a more traditional way of reducing the three-dimensional shape to just a few parameters has been analyzed. If a set of coordinate triples in three-dimensional space is considered to represent the coordinates of n points of equal mass, the tensor of inertia[32] associated with this set reflects the spatial distribution of these points. The lengths of the eigenvectors of this matrix of inertia characterize the shape of the body.[22,32–34] With respect to the bodies analyzed, both an ideal octahedron and an ideal tetrahedron will be characterized by three equal eigenvalues. The rotational symmetry of bodies such as the trigonal prism (C_3), trigonal bipyramid (C_3), and square pyramid (C_4) will produce two eigenvalues of the same size while the third may be different. A planar body such as a hexagon will give rise to one eigenvalue of zero. If the body has rotational symmetry, such as a hexagon with C_6 symmetry, the remaining two eigenvalues will be equal in size. The transformation of a set of points to the coordinate system defined by the eigenvectors of the matrix of inertia is a well established procedure in molecular crystallography, commonly used to define the "best plane" of a set of points.[22] In terms of pattern recognition this method is equivalent to describing the characteristics of a pattern defined by n points in three-dimensional space by its second moments.[23] Thus this method relates to a standard tool in the analysis of distributions which is also a type of pattern recognition technique.

If the masses of the n points of the ensemble analyzed are all taken to be one, the eigenvalues of the matrix of inertia are equal to the sum of the squares of the deviations of the points from the origin (center of mass) along the eigenvector axes. With respect to the bodies analyzed, this means that the eigenvectors will tend to increase with the number of points defining the bodies being analyzed. Thus the number of points is implicitly encoded in the size of the eigenvalues, given that the bodies are scaled such that they fit into a unit sphere and touch the sphere at at least one point. This condition is fulfilled for the training set as described above. In order to retain the information about the number of points n giving rise to a specific set of eigenvalues, and to make results comparable, the eigenvalues have to be scaled in an appropriate way. It has been found through a number of trials that with the conditions met in the training set, the discriminative power of this eigenvector analysis is optimized when each eigenvalue l is replaced by $l' = (1/3.5)\cdot l\cdot n$. To map the results onto a coordinate system defined by the eigenvectors in the interval zero to one, the coordinates are calculated as $l'^2/(l'^2 + 40)$. The results obtained for the training data set using this type of scaling are shown in Figure 5.

It can be seen that the different classes of bodies are, on average, neatly separated in this type of scaled eigenvector diagram. The eigenvalue filter has the advantage of projecting planar bodies onto the YZ plane (X representing the eigenvector axis associated with the smallest eigenvalue and Z corresponding to that with the largest eigenvalue). The only region where there is no clear separation is the one in which octahedra and trigonal prisms are projected (Figure 5). The degree of separation of trigonal bipyramids and square pyramids is also low. Since these bodies are well separated by the neural network technique (see above) it appears that a combination of the eigenvalue filter and the

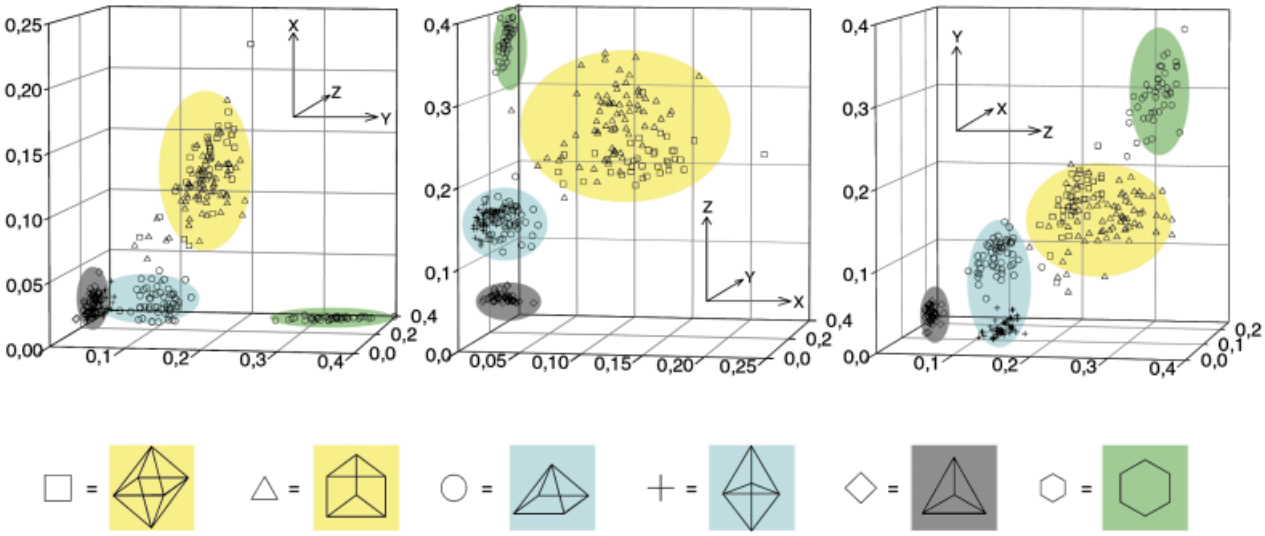


Figure 5. Sorting polyhedra in eigenvector space. The matrix of inertia of a body of n points is calculated with all points given equal weight. The eigenaxes of this matrix define the coordinate system X, Y, Z. The coordinate values are scaled in response to the eigenvalues and to the number of points comprising the body (see text). All three projections onto the eigenvector axes are shown.

neural network filter might be maximally discriminative. This method has therefore been applied as the basis of the work reported below.

Analysis of Packing

In this paper, analysis of the patterns formed by the geometric centers of the molecules within a crystal has been restricted to a search for the three basic types of closest packing. It is clear, however, that the methods developed are applicable to searching for any type of packing pattern.

The analysis has been conducted for the two patterns maximizing density (*fcc*, *hcp*) as well as for the *bcc* pattern, which maximizes the number of closest neighbors.[24] All three of these modes of packing may be described as networks of octahedra.[19–21] In the case of the two dense sphere packings, these octahedra are regular, while in the *bcc* packing mode they are tetragonally compressed. The connectivity of these octahedra, as well as the angles at which they are inclined with respect to each other, are different for the three types of packing (Figure 6; Table 1).

In the *fcc* mode the octahedra are connected via vertices and edges (Figure 6). The angles around the centers of the octahedra at the connecting points are decidedly different for these two types of connections (Table 1). In the *hcp* mode there are three types of connections, vertices, edges, faces (Figure 6) with the associated angles again being

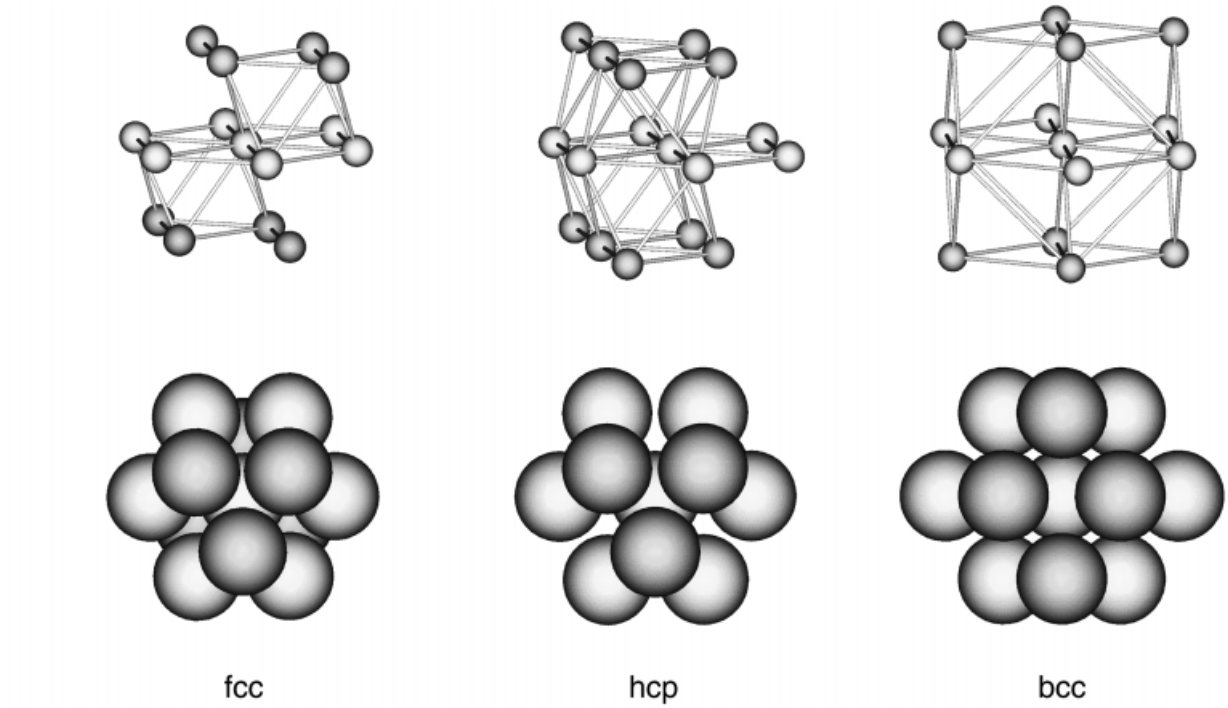


Figure 6. The three types of close packing shown as a network of octahedra (top) and as a stack of layers (bottom)

Table 1. Differentiation of octahedral networks by connection angles. The three basic packing types are described as networks made up of interconnecting octahedra. The octahedra are regular for the *fcc* and *hcp* cases but are tetragonally compressed for the *bcc* case. The values given refer to the angles subtended by the centers of the octahedra and their connecting points in the different modes of connection.

Packing type	Vertex	Edge	Face
fcc	180.0°	90.0°	–
hcp	131.8°	90.0°	70.5°
bcc	180.0°; 135.0°	90.0°	60.0°

within specific ranges (Table 1). The distorted octahedra making up a *bcc* packing also show vertex, edge and face bridging (Figure 6). The angles relevant to these bridging modes are again characteristic for the respective mode (Table 1). A similar type of description may be made with tetrahedra seen as the basic constituents.[19] It was felt that a description based on a network of octahedra[20] might be of higher discriminative power with respect to the problem at hand, since given a random repetitive pattern it is less probable to find six points making up an approximately octahedral shape than it is to find four points conforming to an approximately tetrahedral shape.

Examples

Compounds Crystallizing in Space Group R3

The procedure described above has been applied to a few classes of examples selected either by space group symmetry or by chemical composition. To start, an especially informative example was chosen. When analyzing the packing of a series of compounds which crystallize in the space group *R*3 and which have themselves crystallographic C_3 symmetry, some *bcc* and *fcc* packings should be found within the sample by applying the above methods. In this space group points lying at the C_3 axis will form ideal face-centered and body-centered cubic patterns at specific ratios of the lengths of the a and c axes of their hexagonal cells. A ratio of $a/c = 0.408$ results in an *fcc* packing and a ratio of $a/c = 1.633$ produces a *bcc* pattern.[35] Since the geometric centers of C_3 symmetric molecules necessarily lie on the crystallographic threefold axis, these ratios should also indicate ranges around which the corresponding type of packing will be observed. This means that an a priori assessment of the packing type is possible on the basis of this ratio.

On the other hand, the packing can be analyzed by the pattern recognition techniques described and the relevant packing modes can be delineated from this analysis in each case. If the pattern recognition techniques work well, it is expected that all structures for which the lattice constant ratio predicts a specific type of close packing are independently identified as the same kind of packing. It is in fact found that the predictions by the algorithm are in close agreement with predictions made on the basis of lattice constants (Table 2). This result is highly satisfying since it means that all distorted *fcc* and *bcc* patterns which are a consequence of only the actual axial ratio are found. As shown in Table 2, *bcc* and *fcc* packing modes are also found in ranges which are well outside the exact numbers of $a/c = 0.408$ and 1.633 respectively. The reason for choosing to analyze patterns of packing in space group *R*3 as the first example is also motivated by the idea that perhaps packing and molecular shape are correlated. In space group *R*3, with the centers of gravity of the C_3 symmetric molecules occupying positions at the crystallographic C_3 axis, layers are generated by the very symmetry of the space group. The stacking of these layers, which have the C_3 axis as the normal, is ABC, again solely by virtue of crystallographic symmetry.[35] Whether this kind of stacking of the densely packed layers will approximately correspond to a basic lattice type such as *fcc* depends on the dimensions of the lattice, which are again a consequence of the shape of the molecules.

Table 2. Data set analyzed for compounds crystallizing in space group *R*3. REFCODE refers to the name under which the compound is stored in the CSD. The numbers refer to the numbering scheme used in Figure 7. *a*/*c* gives the corresponding axial ratio with respect to hexagonal setting. The few cases for which a rhombohedral setting is stored in the database have been transformed to the corresponding hexagonal setting (see ref. [35]). Where no value for the *a*/*c* ratio is given the molecules do not occupy positions on a crystallographic C_3 axis. The designators in the columns "packing" refer to the idealized packing type as found by pattern analysis.

REFCODE	No.	*a*/*c*	Packing	REFCODE	No.	*a*/*c*	Packing
CIBKAZ	1	4.87	?	CETJUG	26	1.93	*bcc*
CUKGIY	2	3.00	?	CPPECO10	27	1.64	*bcc*
EFASCO01	3	0.70	?	DAMWET	28	1.92	*bcc*
FEBZIV	4	–	?	DIPYIK	29	1.08	*bcc*
PHPRHO10	5	3.59	?	DODZEB	30	1.80	*bcc*
RERLOP	6	4.60	?	DUSHAA	31	2.11	*bcc*
SUJPUI	7	–	?	FOTGUQ	32	1.19	*bcc*
SUJREU	8	3.58	?	LAMNES	33	1.61	*bcc*
VEKXEO	9	–	?	MSBFEC	34	1.09	*bcc*
VOCXAM	10	–	?	PAFVEX	35	1.54	*bcc*
VUSGIZ	11	2.77	?	PECCUV	36	1.18	*bcc*
VUSGOF	12	2.77	?	PICMET	37	1.74	*bcc*
VUSGUL	13	2.77	?	PPMEIR	38	1.61	*bcc*
YIHTIS	14	–	?	PPMERH	39	1.61	*bcc*
ZEMZIA	15	–	?	RARPEF	40	1.31	*bcc*
ZUDPIX	16	0.79	?	RARVAH	41	1.27	*bcc*
KIFSOH	17	0.87	?	TBZQCR	42	1.18	*bcc*
TEMTOU	18	0.92	?	TFORHI10	43	1.17	*bcc*
VENHIF	19	0.67	?	TICCAJ	44	2.40	*bcc*
TERQIQ	20	2.86	?	TIYLOC	45	2.67	*bcc*
DAJJIH	21	0.30	*fcc*	TONZOL	46	1.35	*bcc*
LIFMUI	22	0.33	*fcc*	VULVAZ	47	1.32	*bcc*
REDBAD	23	0.58	*fcc*	YENHOO	48	3.14	*bcc*
VUGHAG	24	0.67	*fcc*	ZULTOP	49	2.54	*bcc*
VIJGEA	25	1.09	*bcc*				

By using the eigenvalues of the tensor of inertia corresponding to the molecules, we can construct a graph which has the ratio of the length of the eigenaxis of intermediate size, *m*, to the largest one, *l*, (*m*/*l*) on one axis (horizontal axis in Figure 7) and the ratio of the smallest eigenvalue, *s*, to the intermediate one (*s*/*m*) on the other axis (vertical axis in Figure 7). This graph gives an idea of the overall shape of the molecules concerned.

Molecules projected close to the point (1,1) on this diagram are close to spherical in shape. Starting from this point (1,1) along the line $s/m = 1$, the molecular shape is

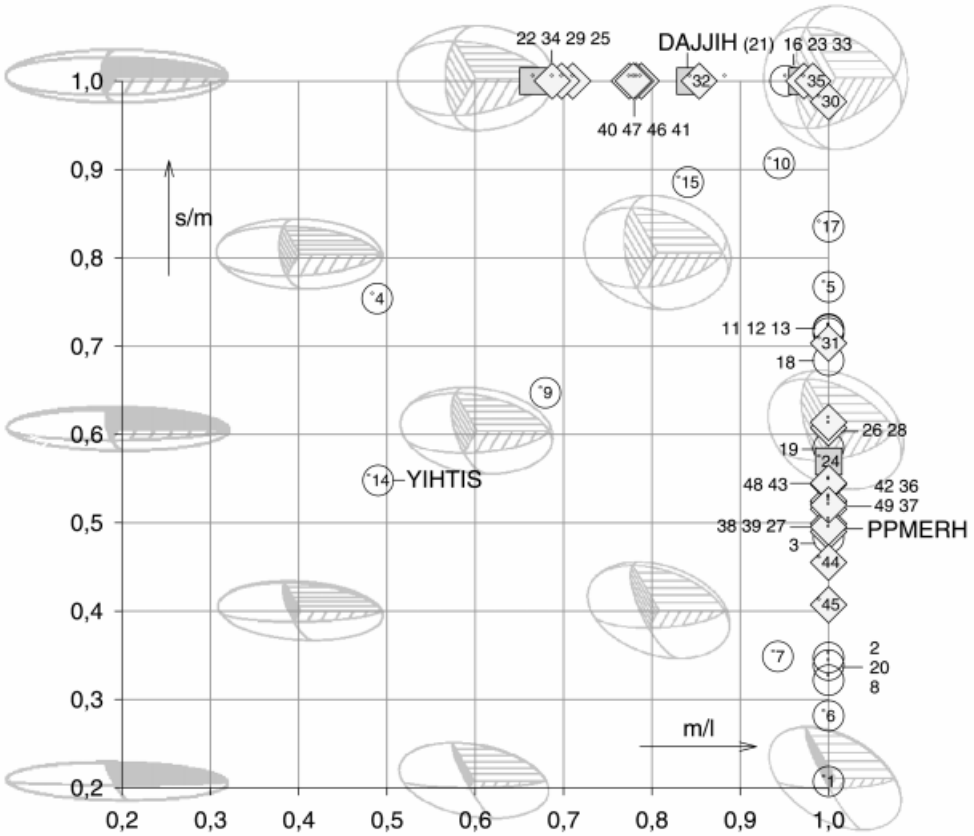


Figure 7. Molecular shape and packing of compounds crystallizing in space group *R*3. The diagram refers to the shape of the ellipsoid of inertia pertaining to each molecule analyzed. The vertical axis gives the ratio of the smallest eigenvalue of this ellipsoid to the intermediate one (*s*/*m*). The horizontal axis refers to the ratio of the intermediate eigenvalue to the largest one (*m*/*l*). The shape of the corresponding ellipsoids is shown in the background of the diagram. The numbers refer to the numbering scheme given in Table 2. In addition, three compounds are named by their REFCODE. The symbols have the following meanings: square *fcc*, diamond *bcc*, circle neither.

increasingly elongated along the C_3 axis. The smaller the value of *m*/*l* the more elongated the ellipsoid. Along the axis $m/l = 1$, again starting from the point (1,1), the molecules are increasingly compressed along the C_3 axes the smaller these *s*/*m* values are (Figure 7). Molecules which are not projected onto the axes ($s/m = 1$, $m/l = 1$) do not have rotational symmetry and all three eigenvalues of the ellipsoids are different. In space group *R*3 this means that these molecules (no 4, 7, 9, 10, 14, 15 see Figure 7, Table 2) are not located at special positions of the unit cell and are not necessarily C_3 symmetric.

The relation of the shape of the ellipsoid of inertia to the true molecular shape is a very reduced one. To present an idea of what this relation looks like Figure 8 shows three selected examples.

Compound no 21 (DAJJIH[36] Figure 7, Figure 8, left) is represented by an elongated ellipsoid. Compound no 39 (PPMERH[36] Figure 7, Figure 8, right) gives rise to a compressed ellipsoid. Both compounds are C_3 symmetric and the ellipsoids therefore show rotational symmetry. The third compound no 14 (YIHTIS[36] Figure 7, Figure 8, middle) gives rise to an ellipsoid with three different axial lengths. Even though the description of the true molecular shapes by their associated ellipsoids of inertia is a very simplified one this method has been used with success to classify the shape of compounds in a general way[33,34] and is also found to be useful in the present context.

When the algorithm described above is applied to the data set, a number of structures are sorted into either one of the two packings *bcc* (diamonds in Figure 7) or *fcc* (squares in Figure 7). Approximate *bcc* packing occurs more often than *fcc* arrangements in the data set analyzed (Figure 7, Table 2). It is to be noted that only structures containing C_3 symmetric molecules fall into these packing categories (*fcc*, *bcc*) and it may be argued that it is the rotational symmetry which allows for the formation of the closely packed layers.

Compounds $Cr(CO)_5L$

To further test the pattern recognition capabilities of the algorithms, a sample of $Cr(CO)_5L$ compounds, showing a wide structural variability and with no space group restrictions were selected for analysis. Among the 99 structures analyzed (Table 3) only 13 were found to fit into either one of the three search patterns *fcc*, *hcp*, *bcc*.

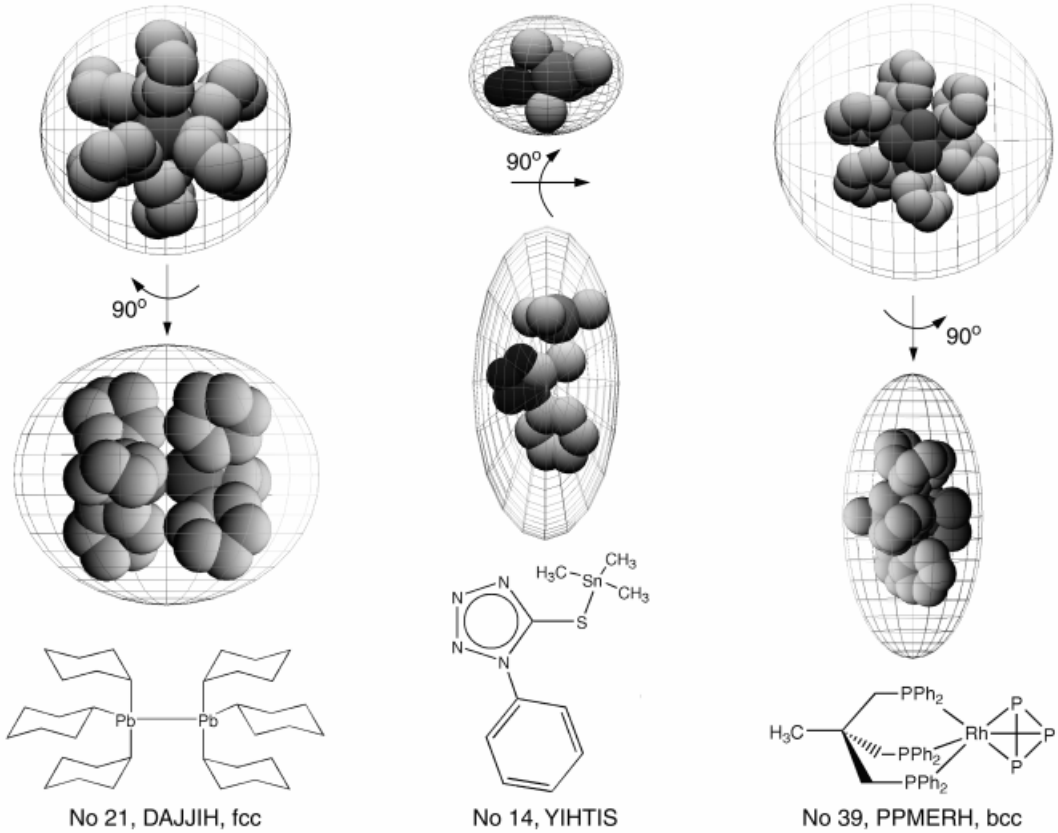


Figure 8. Description of molecular shape by ellipsoids of inertia. The numbers and designators refer to Figure 7 and Table 2. The ellipsoids are scaled such that the molecules fit into their envelope. The labels *fcc* and *bcc* refer to the idealized packing motif of the molecules.

The parent compound $Cr(CO)_6$, being octahedral in shape and thus being represented by a spherical envelope (no 99, coordinates 1,1 in diagram Figure 9) shows almost ideal hexagonal close packing. This has been previously noted by an appropriate visual inspection of the structure[37]. Close to the line $s/m = 1$, i.e. for ellipsoids which have two smaller eigenvalues of equal size, two additional *hcp* patterns are observed (no 90, 95 Figure 9). Close to this line *fcc* and *bcc* patterns are also found (no 83, 85, 89, 92 Figure 9). Out of the 99 structures only 13 are close to one of the idealized packing models, with *fcc* occurring more often than *hcp*, and *bcc* being observed only twice (no 83, 92 Figure 9). It is apparent from Figure 9 that the three basic packing motifs are only found in a range where either s/m or m/l, or both, are greater than 0.8, i.e. in a range where the ellipsoids are not distorted too far from rotational symmetry.

Figure 10 illustrates the relation between molecular shape and ellipsoidal envelope for three selected compounds (no 94, ZUKSED,[36] *fcc*; no 89, YIKSAM,[36] *fcc*; no 92, VOZMUS,[36] *bcc* in Figure 9) which pack in one of the modes analyzed.

Compound 94 gives an impression of how large the deviation from spherical symmetry may be to still allow for an approximate *fcc* type of packing. Compounds 89 and 92 are closer to spherical symmetry (see also diagram Figure 9) and their corresponding ellipsoids are closer to rotational symmetry than the one describing the shape of 94.

Compounds Containing a Tripod-Metal Entity

Due to the interest in the coordination chemistry of tripod metal templates, tripod transition metal compounds were selected as the next class for analysis. All compounds in the set are of the type $RCH_2C(CH_2X)(CH_2Y)(CH_2Z)ML_n$ (R = H, Alkyl, Aryl; X, Y, Z = NR_2, OR, SR, PR_2, PRR′). The constitutional C_3 symmetry of the tripod metal templates, which is exact when X = Y = Z, and may be still approximately so when X and Y and Z are similar, mirrors itself into the diagram Figure 11.

Obviously this constitutional C_3 symmetry is still dominant also in the structure of many of the tripod-ML_n compounds since they project close to the axes $s/m = 1$ and $m/l = 1$. As observed in the two previous examples, structures projecting close to these lines have a high probability of falling into one of the three idealized categories (*fcc*, *hcp*, *bcc*) which define the search pattern of this work. Out of the 44 examples (Table 4) 31 could be categorized this way. Only a few patterns of the *hcp* (no 38, 39, 40 in Figure 11) and *fcc* (no 36, 37 in Figure 11) types are observed. The

Table 3. Data set comprising compounds $Cr(CO)_5L$. REFCODE refers to the name under which the compound is stored in the CSD. The numbers refer to the numbering scheme used in Figure 9. The designators in the columns "packing" refer to the idealized packing type as found by pattern analysis. The designator *hl* assigned to structures 96, 97, and 98 means that the packing of the relevant compounds is characterized by layers of centered hexagons. This side result has been obtained by analyzing the data set on the basis of a filter searching for regular centered hexagons (see text).

REFCODE	No.	Packing	REFCODE	No.	Packing	REFCODE	No.	Packing
BALVIT	1	?	JONMEE	34	?	WAXKUB	67	?
BASTOE10	2	?	JUGWEN	35	?	WETGOR	68	?
BIVTOP	3	?	KAFMUZ	36	?	YESBAZ	69	?
BOVKEC	4	?	KARWEF	37	?	YIPMAL	70	?
CASCUU	5	?	KAXCUH	38	?	YUBTUK	71	?
CATYIF	6	?	KEBMOT	39	?	YUWHAZ	72	?
CHVICR	7	?	KIZWOF	40	?	ZATWAS	73	?
CMPHSC	8	?	KORGED	41	?	ZEPPEP	74	?
CTMTCR	9	?	KUJTOY	42	?	ZIJZOH	75	?
CUFPIC	10	?	LEWYAN	43	?	ZIKCUR	76	?
DAGPUW	11	?	LICBUU	44	?	ZIKDAY	77	?
DARFEH	12	?	LICCAB	45	?	ZITQEY	78	?
DAWXEE	13	?	MEAMCR10	46	?	ZOCCAV	79	?
DEKDUS	14	?	MESTCR	47	?	YETJUC	80	?
DEPTUN	15	?	MTCYCR	48	?	YIZBOY	81	?
DOPKUO	16	?	PCTHFC	49	?	GEWCAM	82	?
EAMNCR10	17	?	PCTXCR	50	?	SUBBIA	83	*bcc*
EMACCR	18	?	PEMSOP	51	?	CNETPA	84	*fcc*
FICZOG	19	?	PEVDAV	52	?	GABYUD	85	*fcc*
FIPCAI	20	?	PMPHCR	53	?	GABZAK	86	*fcc*
FOBPOB	21	?	RARFAR	54	?	LICCEF	87	*fcc*
FOPBIV	22	?	SAKSUS	55	?	YESBED	88	*fcc*
GABZEO	23	?	SASFAT	56	?	YIKSAM	89	*fcc*
GEHBAW	24	?	SEBJEO	57	?	PEZHOR	90	*hcp*
GEVFIW	25	?	TABNEP	58	?	PORGOS	91	*hcp*
GILCUZ	26	?	TACCRC	59	?	VOZMUS	92	*bcc*
HAYXAG	27	?	TAYMOV	60	?	ZIKCOL	93	?
HAYXEK	28	?	TEFXIL	61	?	ZUKSED	94	*fcc*
HEDVOB	29	?	TITHIN	62	?	FONHEV	95	*hcp*
HEPREZ	30	?	TMXTCR10	63	?	WECNIB	96	hl
HEPRID	31	?	VILFUR	64	?	WEDXIM	97	hl
HEZTUB	32	?	VOYJEY	65	?	YUWBIB	98	hl
HIGTUM	33	?	VOYJIC	66	?	FOHCOU	99	*hcp*

greater majority of the patterns correspond to the *bcc* type of packing. By comparing Figure 11 with Figure 7, it appears that *bcc* forms the predominant type of packing in compounds whose inertial tensors give rise to compressed ellipsoids of rotational symmetry with the ratio of eigenvalues *s*/*m* around 0.55.

An illustration of the relation between the molecular shape of tripod metal compounds and the ellipsoids representing them is given in Figure 12.

The two molecules (no 43, NAXHID[36] Figure 12 left; no 42, DUMSUZ[36] Figure 12 middle; see also Figure 11) give rise to ellipsoids with close to rotational symmetry and to a packing which can be classified as idealized *bcc*, while the one showing less apparent symmetry (no 4, LEMFUE[36] Figure 12 right; see also Figure 11) does not fall into one of the three categories with respect to packing.

Correlation Between Packing and Molecular Shape

By comparing the results obtained from trying to correlate the shape of molecules with their mode of packing within a molecular crystal it appears that there is a definitive correlation between the shape and the tendency to pack in one of the three basic packing types, *fcc*, *hcp*, *bcc*. An overlay of Figures 7, 9, 11 is shown in Figure 13.

The results described above for each set of compounds separately are collectively presented in Figure 13. It is apparent that the examples studied map into a range of around $0.3 \leq s/m \leq 1.0$, $0.3 \leq m/l \leq 1.0$. The data tend to accumulate along the lines $s/m = 1$ and especially $m/l = 1$ while the rest of the area $s/m \geq 0.4$, $m/l \geq 0.5$ is almost randomly populated. A total of 193 compounds is represented in Figure 13. The compounds whose packings do not conform to one of the three search patterns are marked by crosses, while the ones that fit are labeled by the appropriate symbols. The area to which the compounds are mapped, and for which assignment to one of the three packing modes is possible, is shaded in yellow in the diagram. Assignment is possible for 73 compounds out of 193. This area is characterized by the symmetry of the ellipsoids describing the shape of the molecules being not too far from rotational. The deviation from rotational symmetry which still allows for one of the three basic packing types may be up to 20% ($s/m = 0.8$, $m/l > 0.8$). When the deviation is larger than this, no *fcc*, *hcp*, or *bcc* packing is found in the sample of 193 compounds (Figure 13). This is a pleasing result as it correlates molecular shape and symmetry with the probability of forming one of the sphere packing modes within a crystal. A possible explanation may be that close packing

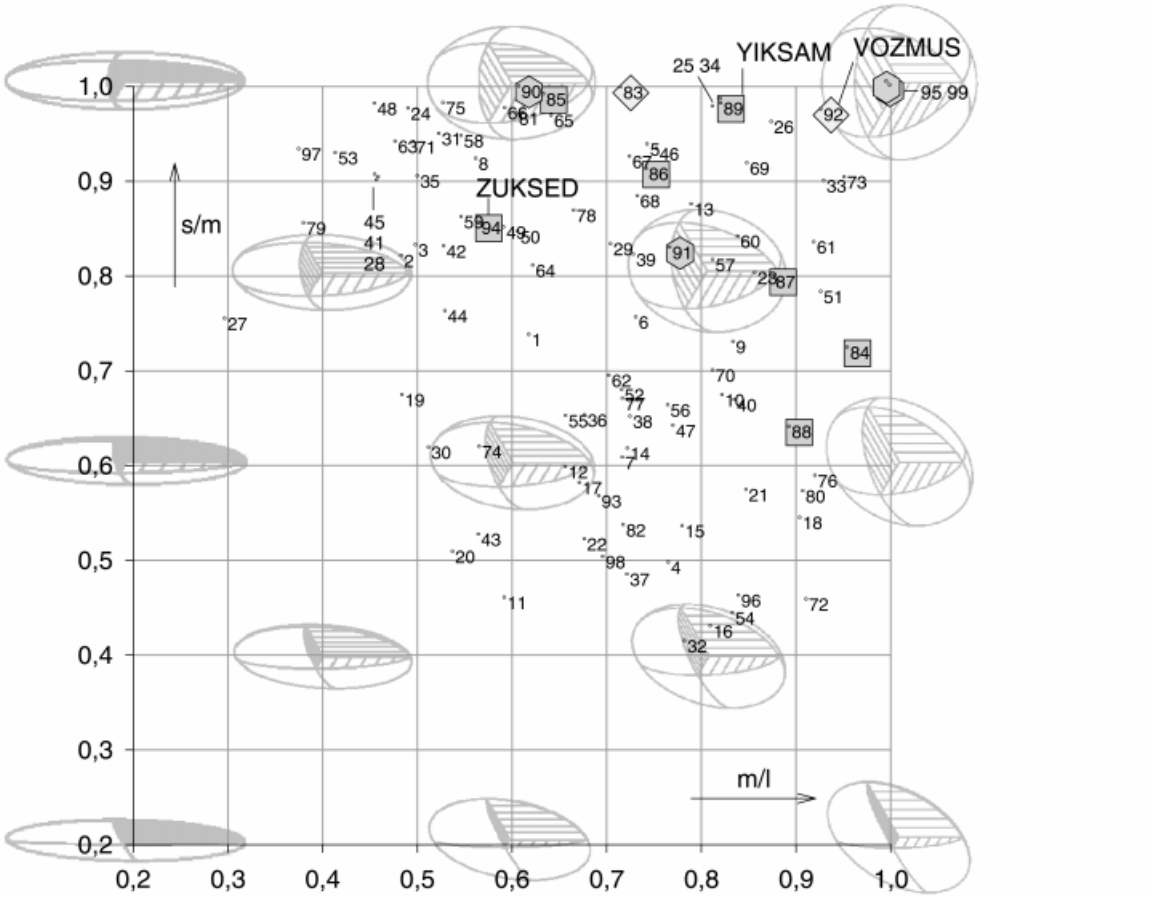


Figure 9. Molecular shape and packing of compounds $Cr(CO)_5L$. The diagram refers to the shape of the ellipsoid of inertia pertaining to each molecule analyzed. The vertical axis gives the ratio of the smallest eigenvalue of this ellipsoid to the intermediate one (*s*/*m*). The horizontal axis refers to the ratio of the intermediate eigenvalue to the largest one (*m*/*l*). The shape of the corresponding ellipsoids is shown in the background of the diagram. The numbers refer to the numbering scheme given in Table 3. In addition, three compounds are named by their REFCODE. The symbols have the following meanings: square *fcc*, diamond *bcc*, hexagon *hcp*.

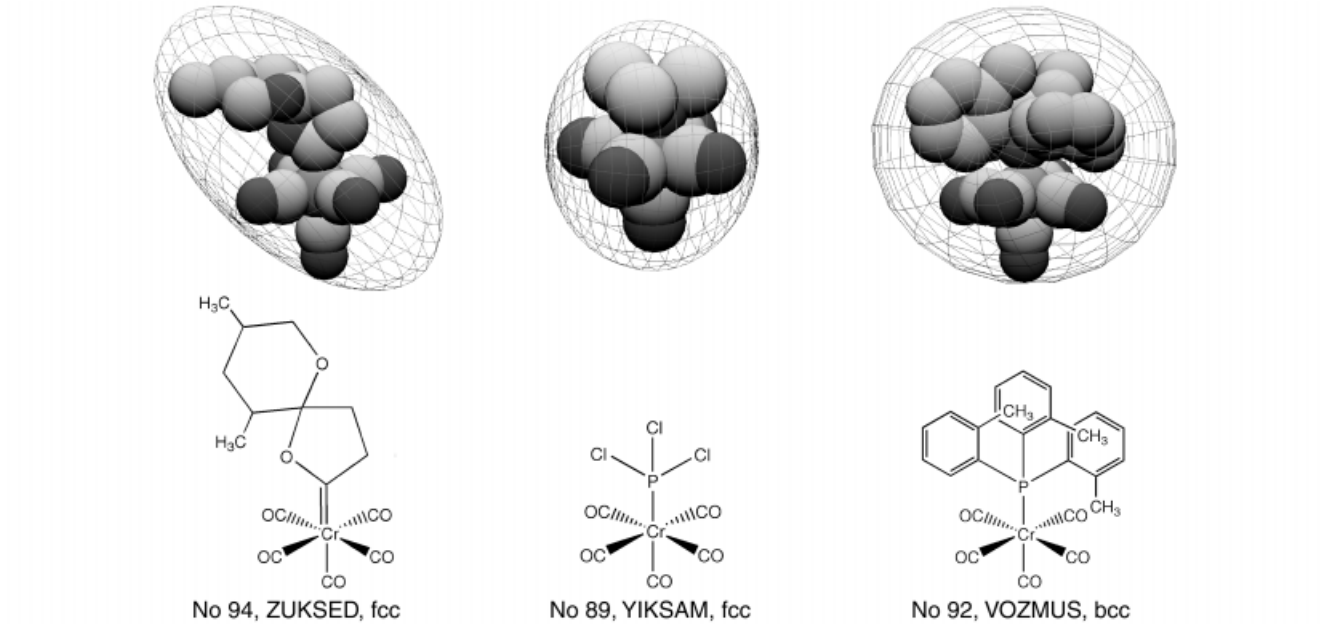


Figure 10. Molecular shape of three selected compounds $Cr(CO)_5L$. The numbers and designators refer to Figure 9 and Table 3. The ellipsoids are scaled such that the molecules fit into their envelope. The labels *fcc* and *bcc* refer to the idealized packing motif of the molecules.

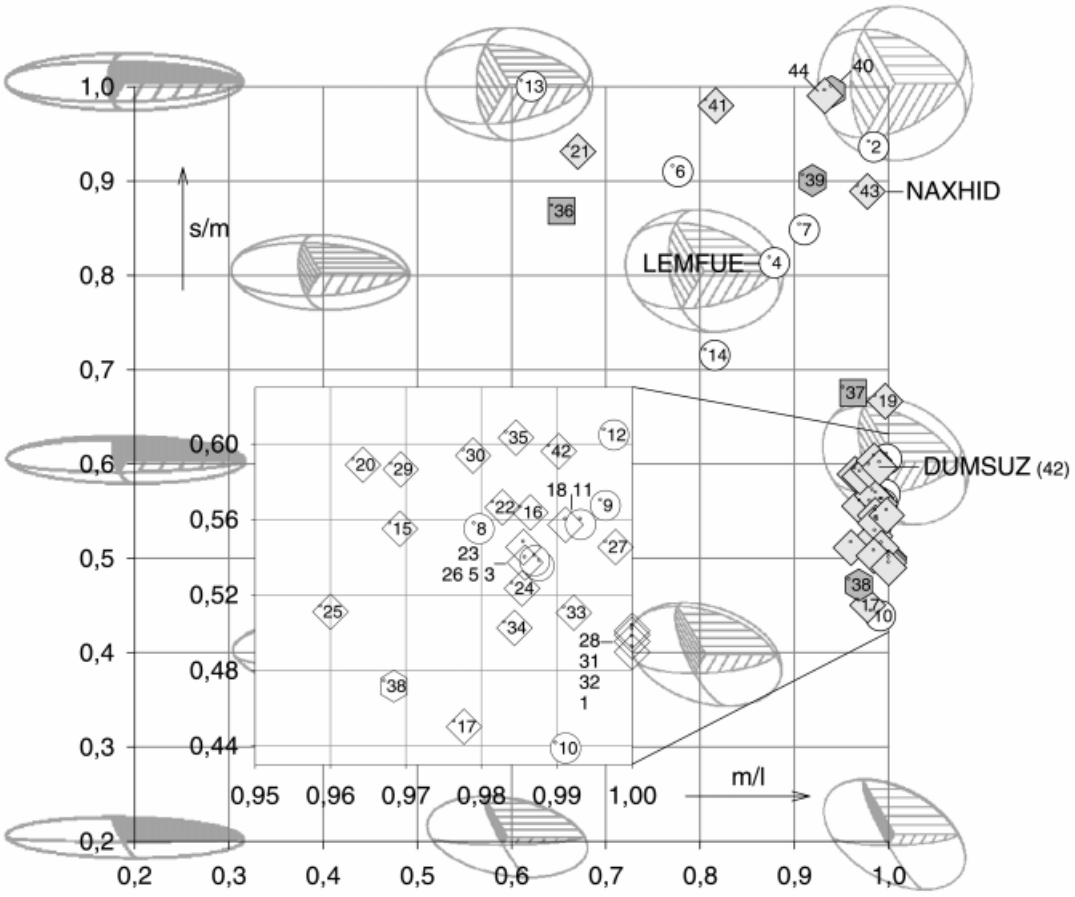


Figure 11. Molecular shape and packing of tripod-metal compounds. The diagram refers to the shape of the ellipsoid of inertia pertaining to each molecule analyzed. The vertical axis gives the ratio of the smallest eigenvalue of this ellipsoid to the intermediate one (s/m). The horizontal axis refers to the ratio of the intermediate eigenvalue to the largest one (m/l). The shape of the corresponding ellipsoids is shown in the background of the diagram. The numbers refer to the numbering scheme given in Table 4. In addition, three compounds are named by their REFCODE. The symbols have the following meanings: square *fcc*, diamond *bcc*, hexagon *hcp*, circle none. – The region around $m/l \approx 1$, $0.4 < s/m < 0.7$ is shown in magnified form in the inset.

Table 4. Data set comprising tripod-metal compounds. REFCODE refers to the name under which the compound is stored in the CSD. The numbers refer to the numbering scheme used in Figure 11. The designators in the columns "packing" refer to the idealized packing type as found by pattern analysis.

REFCODE	No.	Packing	REFCODE	No.	Packing	REFCODE	No.	Packing
CPPECO10	1	*bcc*	BOLJAN	16	*bcc*	PPMEIR	31	*bcc*
COTPMO	2	?	BOMPIC	17	*bcc*	PPMERH	32	*bcc*
LAKYUR	3	?	BURREL10	18	*bcc*	SXPMNI	33	*bcc*
LEMFUE	4	?	CODJIO	19	*bcc*	TPHCOA	34	*bcc*
PAKGOX	5	?	COGWOK	20	*bcc*	TUMYUV	35	*bcc*
RIPFAX	6	?	DOZWUK	21	*bcc*	HAJFED	36	*fcc*
TOTBAF	7	?	DPMENI10	22	*bcc*	HAJFIH	37	*fcc*
ZEPPUF	8	?	FIFBAX	23	*bcc*	DEPNER	38	*hcp*
SUHNAK	9	?	HBPMCO01	24	*bcc*	DUMVIQ	39	*hcp*
VEPSOY	10	?	KEDBIE	25	*bcc*	TOXBEN	40	*hcp*
BELTIV10	11	?	LATNUP	26	*bcc*	CUZTEW	41	*bcc*
ZIWRUS	12	?	PIBYUU	27	*bcc*	DUMSUZ	42	*bcc*
REZYOK	13	?	PICMET	28	*bcc*	NAHXID	43	*bcc*
ROQGUZ	14	?	PMECOS	29	*bcc*	ZIGCOH	44	*bcc*
BISWEF10	15	*bcc*	PMENSE	30	*bcc*			

within layers is optimal when the perimeter of the molecule within the layer is close to circular. The ratios in which the three types of packing *fcc*, *hcp*, *bcc* are observed in the sample shown in Figure 13 is 13:7:53. Even if the sample may be too small to derive any general rule from this finding, one of the reasons which might favor a *bcc* type of

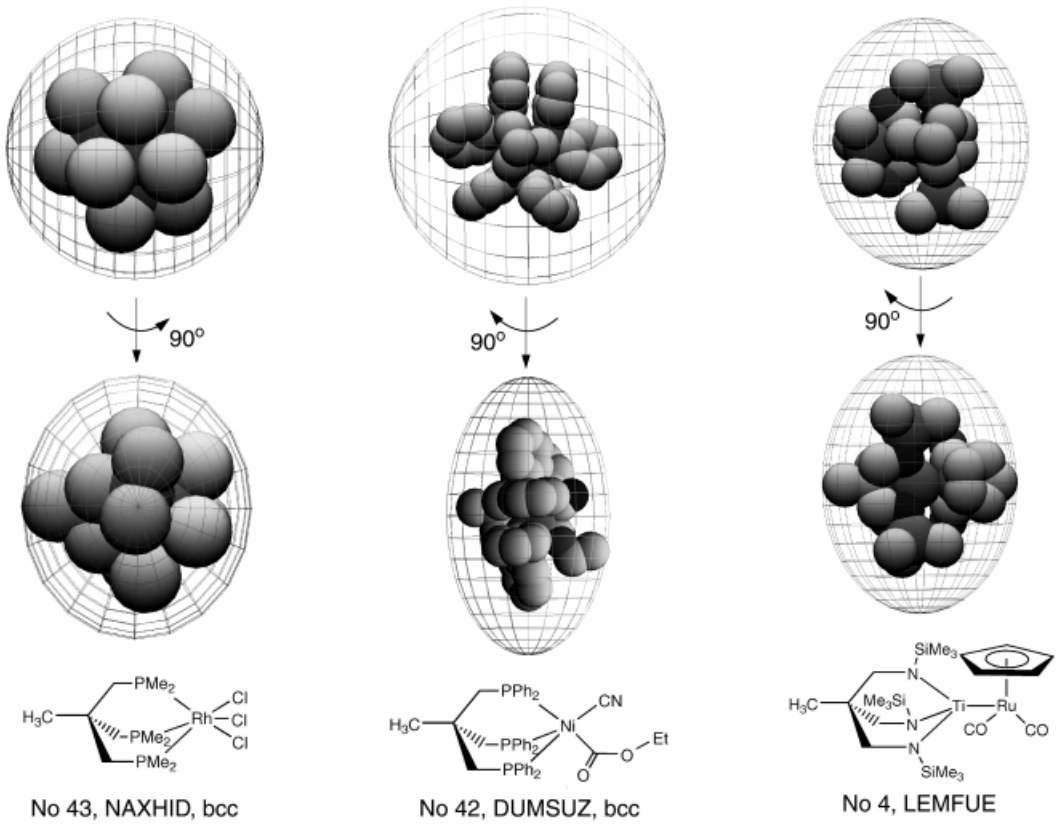


Figure 12. Molecular shape of selected tripod-metal compounds. The numbers and designators refer to Figure 11 and Table 4. The ellipsoids are scaled such that the molecules fit into their envelope. The label *fcc* refers to the idealized packing motif of the molecules.

packing over the other two packing modes is that, with respect to packing spheres, the *fcc* and *hcp* modes are the optimal solutions as far as efficient use of the available space is concerned. Body centered cubic packing, on the other hand, is the optimal solution with respect to maximizing the number of short direct contacts between neighboring spheres.[24] It may well be that maximizing the number of contacts does more to stabilize a molecular crystal than does the optimal use of space.

Conclusion

1. Pattern recognition techniques have been developed which allow for classification of a given set of *n* points in three-dimensional space as being similar to a predefined pattern of the same number of points. The predefined pattern may be any type of chemically interesting polyhedron or may be some other type of pattern that might be of interest. The techniques are such that the quality of fit between a given arrangement of points with one or the other of the predefined categories is quantified as a number between zero and one. Predefinition of patterns relies upon training a neural network to "see" what a human observer would see. Whenever a specific category has to be predefined, a set of training data has to be produced by randomly distorting the ideal pattern. The distorted patterns are inspected and ranked from zero to one according to similarity to and recognizability of the ideal pattern.

2. Several methods to encode three-dimensional structural data into a vector suitable as input to neural networks are reported and it is found that projecting the distance and angle data of the metric matrix of a structure in histographic form onto the vector is a highly efficient strategy.

In addition to this novel pattern recognition technique, the more traditional second moment analysis based upon the matrix of inertia and its eigenvalues is also used.

3. The packing of molecules within a crystal is analyzed with respect to whether the pattern observed might be described as a derivative of one of the three basic types of sphere packing *fcc*, *hcp*, *bcc*. To this end, the molecules are represented by the coordinates of their geometric centers. By symmetry expansion, a minimum number of *m* nearest neighbors within a sphere is generated and all six-point combinations for this set of points are analyzed for octahedral geometry. The octahedral arrangements found are evaluated with respect to their connectivity and to the angles that describe this connectivity. From the type of connectivity found, discrimination between the classes – *fcc*, *hcp*, *bcc*, none of these – is possible.

4. Three sets of compounds – containing a total of 193 organometallic structures – are analyzed with respect to whether the packing modes adopted within the crystals may be described as a variation of one of the three basic types

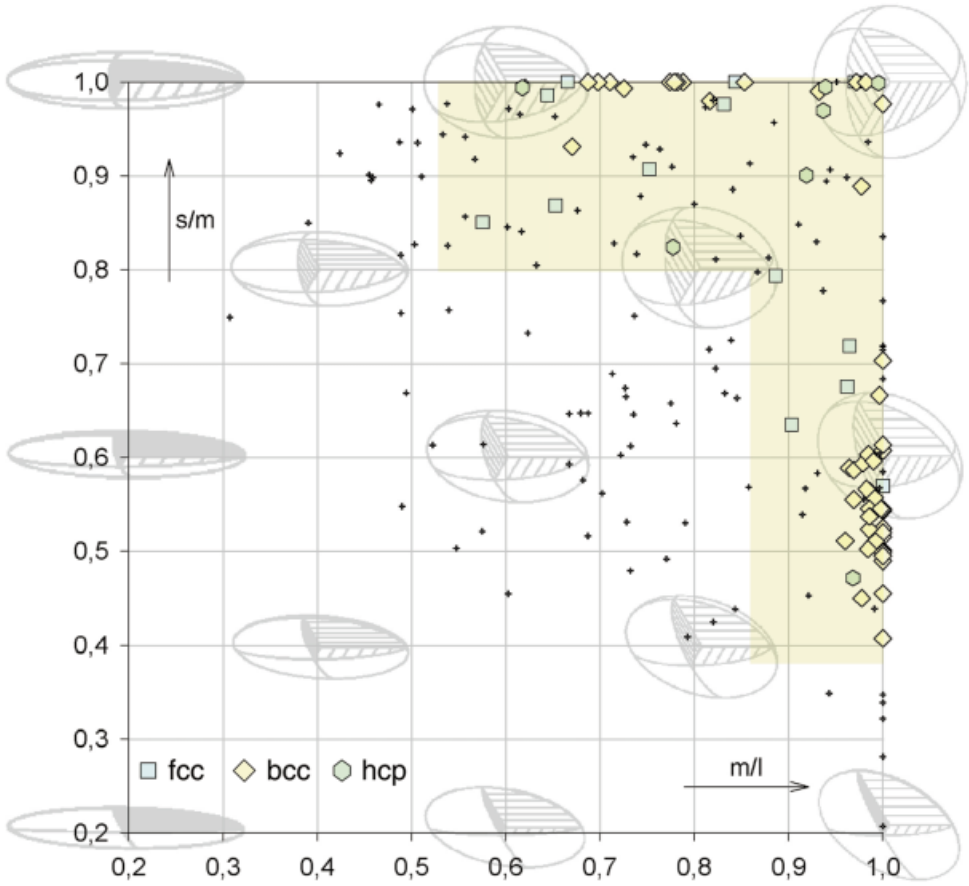


Figure 13. Relation between molecular shape and molecular packing over all data sets analyzed. The diagram refers to the shape of the ellipsoid of inertia pertaining to each molecule analyzed. The vertical axis gives the ratio of the smallest eigenvalue of this ellipsoid to the intermediate one (*s*/*m*). The horizontal axis refers to the ratio of the intermediate eigenvalue to the largest one (*m*/*l*). The shape of the corresponding ellipsoids is shown in the background of the diagram. Only molecules whose shape ellipsoids fall into the area shaded in yellow fall into one of the specified packing categories. Molecules whose shape indicators are shown as a cross do not fall into any of the packing categories analyzed.

of sphere packing. The relevant crystal data are taken from the CSD. One set comprises all organometallic compounds crystallizing in the space group *R*3. The second set contains compounds $Cr(CO)_5L$ and the third one all tripod metal derivatives $RCH_2C(CH_2X)(CH_2Y)(CH_2Z)ML_n$ (R = H, Alkyl, Aryl; X, Y, Z = NR_2, OR, SR, PR_2, PRR′). It is found that the probability of a structure falling into one of the three classes is connected with the shape of the molecules. If the shape is described by the proportions of the eigenvalues of the matrix of inertia corresponding to the molecule, the appropriate diagrams reveal that a packing mode close to one of the three types of sphere packing is most probable when the ellipsoid of inertia has rotational symmetry, i.e. if the compound has approximate C_3 or higher rotational symmetry. Compounds represented by ellipsoids devoid of rotational symmetry may still adopt one of the basic sphere packing types if the relevant eigenvalues do not differ by more than 20% from the mean. At greater ellipticities no match with the search pattern was found.

Experimental Section

Construction of a Data Set Comprising Specified Polyhedra: The data set for training the neural network and calibration of the second moment analysis was generated using a program that allows for the random distortion of ideal polyhedra. From the vertices of an ideal polyhedron, each new vertex of the distorted polyhedron is chosen by randomly selecting a point within a sphere around the vertex of the ideal polyhedron. The degree of distortion can be controlled by the radius of this sphere. As distortion of an ideal polyhedron performed in this way leads rapidly to geometries far from the ideal, a second parameter was introduced to control the distance between the vertices. For each distortion process of a given type of polyhedron, one selects the radius of the distortion sphere (see above), the percentage the distances are allowed to deviate from the ideal polyhedron and the number of distorted polyhedra to be generated. Using this procedure a data set of 381 polyhedra consisting of 37 squares, 43 tetrahedra, 36 pentagons, 55 square pyramids, 54 trigonal bipyramids, 50 hexagons, 73 trigonal prisms, and 38 octahedra including the ideal polyhedra was generated. The degree of distortion was evaluated by visual inspection of each polyhedron. The polyhedron was visualized on a computer screen in any desired orientation. The quality of fit of each polyhedron to the ideal was judged by inspection and a qualifier between $q = 0$ (no fit at all) and $q = 1$ (perfect fit) was allotted. Figure 4 gives an idea of how this subjective assignment of the quality of fit was done. With the specific pattern analysis problem in mind, only those polyhedra for which the qualifier was above 0.8 were incorporated into the training set, giving a total of 381 polyhedra together with the qualifiers (see above). Figure 14 shows six octahedra together with their qualifiers as an example.[38]

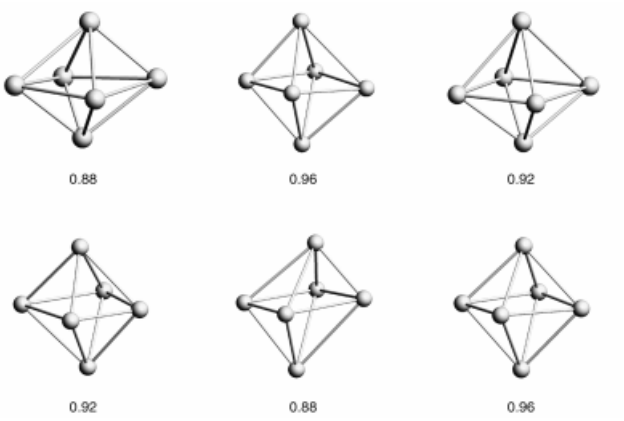


Figure 14. Example of six octahedra with qualifiers above 0.8. The octahedra were generated by randomly distorting an ideal octahedron. The qualifiers indicated by the numbers were allotted by inspection (see text).

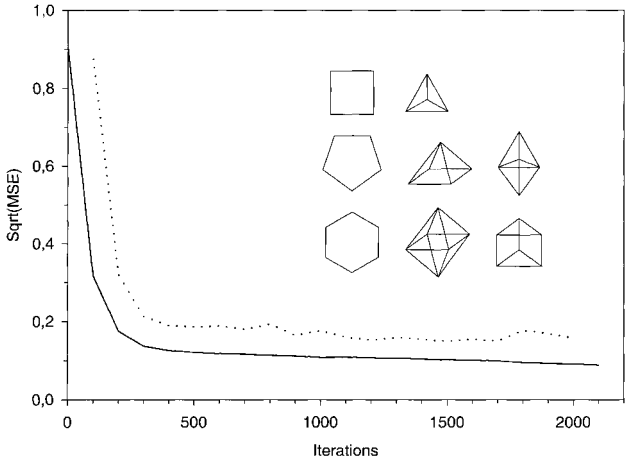


Figure 15. Training of a neural network. The network is designed (see text) to sort the types of polyhedra shown in the inset into their appropriate categories. It was trained on the basis of 281 reference polyhedra. The quality by which the learning pattern is reproduced is represented by the solid line. The quality of prediction for bodies not contained in the training data set is shown by the dashed line.

Sorting of Polyhedra With $4 \leq n \leq 6$ Vertices: The neural network simulator program SNNS,[39] version 4.1 was used. In this work only neural networks of the *feed forward* type were used.[40,41] Such a network consists of an input layer with the number of i input neurons depending on the dimension of the input vector; one or more hidden layers with a number of h hidden neurons and an output layer with the number of o output neurons depending on the dimension of the output vector (see Figure 3). The neurons between each layer are connected and each connection has a specific weight. During the learning process the individual patterns are presented to the network and the weights are adjusted such that the error between the given output vector and the calculated one is minimized.

Figure 15 (solid line) shows the progress of the learning procedure for a network with 27 input neurons, three hidden neurons and nine output neurons, which was trained with a data set containing a total of 281 polyhedra of the types shown in this figure. The shapes of these polyhedra were encoded into a vector based on the method outlined as described in Figure 3 (see above). In contrast to the example shown in Figure 3, the distance part of this vector was limited to only six components, since distances shorter than one half of the maximum distance did not occur in the data set as a consequence of the qualifier being set to q 0.8 (see above). The number of points comprising the individual types of polyhedra was implicitly encoded. Instead of normalizing the ordinate values (''frequency'' in Figure 3) in the histographic representations (for instance by setting the maximal frequency value to one in each case) the frequencies f were mapped onto an ordinate range $0 \leq$ ordinate < 1 by using $f^2/(f^2 + 40)$ as the ordinate value. This means that ordinate values will tend to be larger with larger numbers of points.

As far as convergence of the learning procedure is concerned, the dashed line in Figure 15 shows the root mean square error for a validation data set of 100 patterns containing the same kind of polyhedra as the training set. The root mean square error is defined as the square root of the mean standard error. The mean standard error (MSE) is the sum of the squared deviations between the calculated and the given components of the output vector for all training patterns divided by the number of training patterns. The root mean square error thus indicates the deviation of the output vector as a whole. The deviation of one component of the output vector is calculated as the sum of the squared distances between the calculated and the given output vectors for all training patterns, divided by the number of training patterns and by the number of vector components. The scale of the vertical axis in Figure 15 would then be reduced by a factor of nine (nine output neurons, eight for the bodies shown in Figure 15, a specific one for each body and one extra neuron introduced as a test. This ninth output neuron should only have large numbers $0 \leq q \leq 1$ allotted to it if the body analyzed has no relation to any of the types presented in the training set). The deviation of the individual components is on the order of 0.02 on a scale $0 \leq q \leq 1$. It can be seen that there is no large difference between the straight and the dashed lines. This means that the trained network cannot only recognize the patterns included in the training set but also recognizes patterns which have not been part of the training set.

Sorting of Polyhedra With $n = 6$ Vertices With Respect to Crystal Packing: For the particular problem of pattern recognition with respect to crystal packing, a neural network of the *feed forward* type was trained to filter hexagons, trigonal prisms, and octahedra from ensembles of six points in three-dimensional space. The data set comprised 153 patterns, 43 hexagons, 38 octahedra, and 73 trigonal prisms. This data set was split by random selection into a training set of 137 patterns (35 hexagons, 34 octahedra, 68 prisms) and a validation set of 16 patterns (7 hexagons, 4 octahedra, 5 prisms). The manner in which the data set is separated into different training and validation sets has no great influence on the result. The neural network contained 27 input neurons, six hidden neurons and three output neurons as is schematically shown in Figure 3 where only three hidden neurons are shown. Experiments with different modifications of this type of network suggested that six hidden neurons – as was finally used – is close to the optimum. Each output neuron represents one of the three trained classes (hexagon, octahedron, trigonal prism). The training parameters used with SNNS[39] were as follows:

Learning func: *Backprop Momentum*; Update func: *Topological Order*; Init func: *Randomize Weights*; Learning parameters: *0.6 0.1 0.05 0.05*; Iterations: *500*

Validation of the trained network with the validation data set resulted in a prediction capability of 95.5% with respect to the qualifiers q allocating the polyhedra to their corresponding categories.

The question of practical interest is whether any polyhedron was allocated to an incorrect class. It was found that the classification by the network was 100% correct in this respect, i.e. all polyhedra were correctly assigned to their corresponding class.

Second Moment Analysis: As described above, second moment analysis is another, more traditional method of pattern analysis. The method was used in this case to classify the polyhedra of the data set described at the beginning of the experimental section. For each polyhedron with p vertices one has to calculate the center of mass. The coordinates for each vertex are translated such that the center of mass is at the origin of an orthogonal coordinate system. The new orthogonal coordinates (x_i, y_i, z_i) of the p vertices build up a $3 \times p$ matrix **A**

$$\boldsymbol{A} = \begin{bmatrix} x_1 & x_2 & \cdots & \cdots & x_p \\ y_1 & y_2 & \cdots & \cdots & y_p \\ z_1 & z_2 & \cdots & \cdots & z_p \end{bmatrix}$$

which is multiplied with the transposed matrix $\mathbf{A}^T$ to give a new 3×3 matrix $\mathbf{B} = \mathbf{A}\mathbf{A}^T$. The matrix **B** can be expressed as $\mathbf{B} = \mathbf{X}\Lambda\mathbf{X}^T$, where Λ is the eigenvalue matrix holding the eigenvalues λ_1, λ_2, λ_3 as diagonal elements and **X** is the eigenvector matrix which holds the eigenvectors. The eigenvalues describe the shape of the ellipsoid of inertia which has some relation to the shape of the polyhedron. In order to retain information about the number of points giving rise to a specific set of eigenvalues and at the same time to make results comparable, the eigenvalues have to be scaled in an appropriate way. It has been found through a number of trials that, with the conditions met in the training set, the discriminative power of this eigenvector analysis with respect to separate bodies with n different vertices into individual categories is optimized when each eigenvalue λ_i is replaced by

$$r_i = \lambda_i \cdot \frac{n}{3.5} \quad i = 1,2,3$$

with the scaling function being

$$e_i = \frac{r_i^2}{r_i^2 + 40}$$

The results obtained for the training data set using this type of scaling are shown in Figure 5. The axes X, Y, Z represent the scaled eigenvalues e_i (i = 1, 2, 3). It can be seen that there is a fair separation for some kinds of polyhedra but also that octahedra and trigonal prisms as well as square pyramids and trigonal bipyramids are not well separated. The regions to which the different kinds of polyhedra are mapped to in this diagram are numerically shown in Table 5.

Table 5. Classification of polyhedra by their inertial tensors. The table refers to Figure 5. The coordinates X, Y, Z relate to the eigenvalues of their inertial tensor with a scaling function applied as described in the text. The following abbreviations are used: sq. py square pyramid; trigbpy trigonal bipyramid; oct octahedron; trig pr trigonal prism

Polyhedron	X	Y	Z
Square	0.00–0.0001	0.080–0.130	0.100–0.150
Tetrahedron	0.030–0.055	0.042–0.068	0.054–0.087
Pentagon	0.00–0.0001	0.200–0.242	0.240–0.274
Sq. py/trig bpy	0.031–0.110	0.082–0.183	0.156–0.242
Hexagon	0.00–0.0001	0.370–0.400	0.390–0.430
Oct/trig pr	0.005–0.228	0.156–0.244	0.277–0.510

Packing Analysis: Figure 16 illustrates the general strategy adopted. Given the coordinates of the atoms of the independent molecules within a unit cell and the dimensions and symmetry of the cell, the centers of gravity of all molecules contained in the volume of the unit cell are generated (Figure 16 I→II). The pattern thus obtained is expanded by translational symmetry (Figure 16 II→III). A point in the center of this expanded pattern is selected and its nearest neighbors (16 to start with) are found. Using the pattern recognition tools described above, all octahedra formed by any six of these 16+1 points which are above a certain quality threshold q are found. This qualifier is used to select those six point arrangements which most probably fit together to describe the packing by sharing vertices, edges and faces. In a first attempt, the value of the qualifier $q = 0.9$ is set as the lower limit and from the subgroup of octahedra transgressing this border, pairs of octahedra are selected according to the additional restriction that their qualifiers should be numerically as close together as possible. The connectivity between these pairs of octahedra is then analyzed. The process is repeated with all possible combinations. Based on the entire set of connectivities and their corresponding angles, the packing is classified as belonging to one of the three classes. If no octahedra that fulfill the above conditions are found the thresholds are reduced and the process is repeated. If no fit is found with the qualifiers still having values above $q = 0.8$, the number of points selected in the starting set is increased by one and the procedure is repeated. If no fit is found for up to 24 points in the starting set the packing is considered to not be one of the three types analyzed. In successful cases, this procedure is illustrated as step III→IV in Figure 16. While the thresholds and limits as defined do not have a theoretical foundation, they have been found to be practically sufficient to find any type of packing which has some apparent relation to the three basic packing types analyzed.

Since it is well-known that many patterns of packing involve densely packed layers of molecules, even though these layers are not stacked in such a way as to give rise to one of the three special packing modes discussed above (*fcc*, *hcp*, *bcc*), the search for densely packed layers appears especially worthwhile since the appropriate filters are available (see above). The search may be done by filtering for regular hexagons, independent whether they are centered or not. These hexagons will be filtered out as already described. As an alternative, centered hexagons may be filtered out solely by the application of second moment analysis. This search is done by momentum analysis (see second moment analysis) in order to qualify the seven point arrangements with respect to their planarity (length of the smallest eigenvector). This qualifier *eval* is later used as a parameter in the final filtering function. As a second parameter for this filter-function, the geometric center X of the seven points p_i (i = 1 ... 7) is calculated. The distances d_i from X to all seven points p_i are determined. The smallest of these, d_m, is discarded (the point p_m leading to the smallest distance d_m should be the idealized center of the hexagon). The mean, d_{mean}, of the six distances d_i ($i \neq m$) is calculated. The sum of the differences between the individual distances d_i ($i \neq m$) and their mean value, d_{mean}, is used as the parameter *ldiff*. For a close to regular centered hexagon *ldiff* should be as small as possible. In order to find the hexagons of minimum size the sum of the distances d_i ($i \neq m$) is chosen as the third parameter *lsum*. The filter function *over* is defined as

$$over_i = 1.5 \cdot \left(\frac{eval_i}{eval_{\max}}\right) + \left(\frac{ldiff_i}{ldiff_{\max}}\right) + \left(\frac{lsum_i}{lsum_{\max}}\right)$$

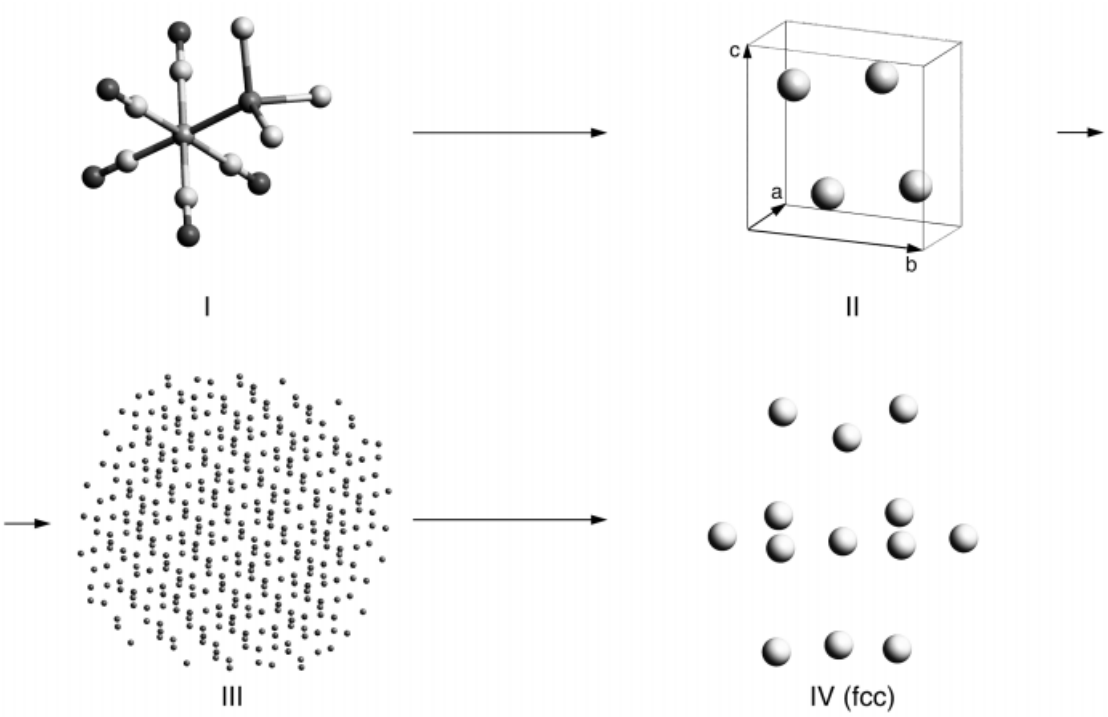


Figure 16. Schematic illustration of the strategy. Molecules within a crystal are replaced by their centers of geometry. The packing of these centers is analyzed by pattern search methods for its similarity to some predefined pattern.

$eval_{max}$ is the largest value of *eval* within the set of seven point arrangements analyzed, comprising a total of 16+1 over 7 individual arrangements. Correspondingly, $ldiff_{max}$ and $lsum_{max}$ are the maximum values found for any seven point combination within the set. Weighting the first term by 1.5 against a weight of 1 for the other terms means that planarity is taken as the most important criterion. The coefficients as well as the type of functions suitable for a filter function have been determined by trial an error. In this manner, any close packed layers are automatically found for any structure analyzed. The centered hexagon with the smallest value *over* is taken as the best plane in the search of an idealized unit cell.

Construction of a Unit Cell: The information about close packed layers obtained by the procedure described above is useful in several aspects. It may be used to project the structure – be it reduced to one central point per molecule or the complete structure – onto this densely packed plane. If the molecules are represented by points, the kind of pattern seen in this projection immediately illustrates the type of packing (see Figure 6). Layer structures with different types of stacking of the individual layers are also immediately identified (see Table 3).

In a second application, the information about layers containing centered hexagons may be used to construct an idealized unit cell. The hexagon is part of a (1,1,1) layer in the *fcc* case (layer stacking ABC) or a (1,1,0) layer in the *bcc* case (layer stacking ABA). In the latter case, even ideal hexagons do not have the full D_{6h} symmetry– which characterizes an *fcc* or *hcp* packing – but are elongated and have maximally D_{2h} symmetry. To find an idealized cube with the appropriate centering with the experimental points as its vertices and centers, the connectivity between the points of such a cube is automatically determined on the basis of the known connectivity of the ideal bodies and of their orientation with respect to the layers.

Figure 17 represents an example of such a body (data from YIKSAM,[36] no. 89 Table 3, Figure 10). In the center of this figure, the body is shown in a projection usually used for illustrating a face-centered cubic cell. At the sides of this figure, projections along the four space diagonals of this body are shown, illustrating that the packing and stacking of close packed planes is very similar in these four directions, as it should be for an *fcc* type packing, where four projections would be identical in the ideal case. The body shown in Figure 17, in addition to not being an exact cube, is not even a parallelepiped and will therefore not be suitable to

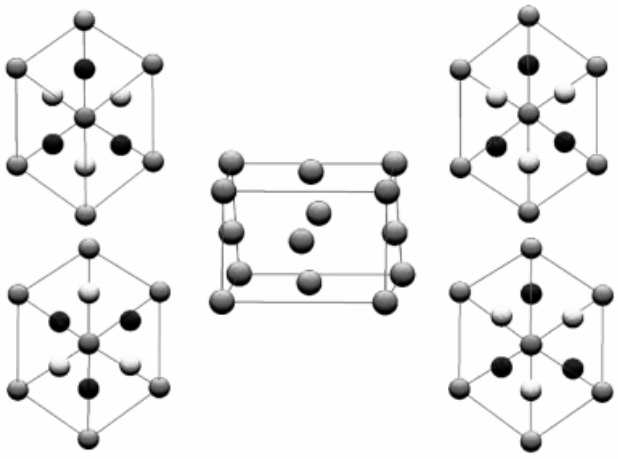


Figure 17. Idealized *fcc* arrangement of one of the specimens analyzed. The figure refers to YIKSAM (Table 3, Figure 10). A projection illustrating the approximate face-centering is shown in the middle. The four independent views of the packing along the four space diagonals of the body shown in the middle are shown at the sides.

describe a crystal, which is an object characterized by translational symmetry. The specific simple example (YIKSAM) is shown in Figure 18 in a two-dimensional projection.

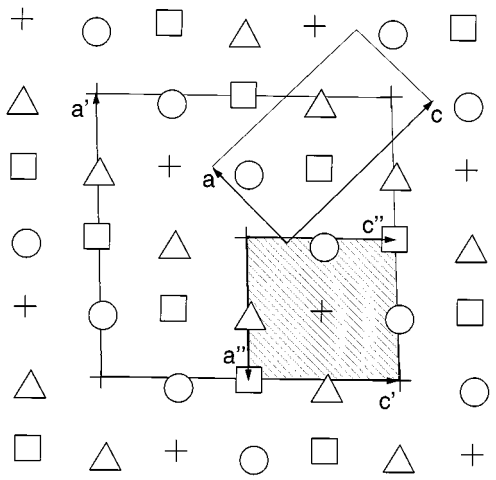


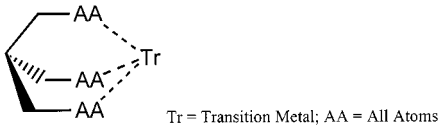
Figure 18. Relation between packing motif and unit cell. The data refer to YIKSAM (Table 3, Figure 10). The packing motif already shown in the middle of Figure 17 is projected onto one of its approximately square faces (shaded area). The axes defining this area are designated a′′, c′′. The axes of the original unit cell are indicated by a and c with the third axis b being perpendicular to a and c. While the shaded area is not yet a repeating unit, a unit cell defined by a′ and c′ is. The length of the axis b has to be doubled.

The crystallographic unit cell is indicated by the vectors a and c. The four points within this unit cell are designated by four different symbols relating to the different coordinates with respect to the third axis b, which is perpendicular to the projection. The idealized face-centered pattern is indicated by an approximately square shaded area. Since the points symbolized by squares and crosses are at approximately the same elevation, they may together form the face of a cube. The points symbolized by circles and triangles again have rather similar coordinates along the b axis but the elevation along this axis is different from that of the points forming the face of the cube. These points (circles, triangles) are the ones centering the faces parallel to the direction of projection. It can be seen that the shaded square is not yet a repetitive unit but that doubling the lengths of its axes a′′, c′′ will give a true repetitive unit, designated by a′ and c′ in the diagram. Once the packing is characterized as belonging to one of the cubic pattern modes, it is possible to construct a body with eight vertices and the respective centering with the experimental points. To find the corresponding repetitive unit the edges of this body, which are close to perpendicular with each other, are used to define a parallelepiped. The axes of this parallelepiped are then expressed in terms of the axes of the original unit cell and based on the restriction that only integer multiples of the original axes are capable of constructing a repetitive pattern, a new unit cell which is a true unit cell and contains the centered bodies as constituents is constructed. This cell may be several times larger than the basic crystallographic unit cell (Figure 18).

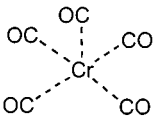
The algebraic details of this transformation are the following: Let a′′, b′′, c′′ be the three vectors which approximately correspond to a description of the body and let a, b, c be the base vectors of the crystallographic unit cell (Figure 18). We then search for a matrix **R**, which transforms the matrix **A** formed by the rows a, b, c into the matrix **B** composed of the rows a′′, b′′, c′′. Since both matrices are nonsingular by definition, the matrix **R** is equal to the product of **B** and the inverse of **A** ($\mathbf{R} = \mathbf{B}\cdot\mathbf{A}^{-1}$; $\mathbf{R}\cdot\mathbf{A} = \mathbf{B}$). The matrix **R** may have non-integer coefficients. Within one row of this matrix, the multipliers that make the coefficients integer, or at least close to integer, are determined and the row is multiplied by the product of these multipliers. Multiplying **A** by this modified matrix results in the best approximation to a cell that is at the same time a true repetitive unit and has the closest relation to the packing type.

Data Sets: All structures were retrieved from the August 1998 version of the CSD.[3] For all searches the parameters NRES = 1 and $Z' \leq 1$ were used. In the case of the structures in the space group *R*3, the parameters SPGN = 146 and atom type Tr were used.

For the molecules with tripodal ligands the search fragment was the following:



For the $Cr(CO)_5L$ structures the search fragment was:



Programs and Equipment: All calculations were carried out on a Silicon Graphics Indigo^2 R4400 workstation, 200 MHz, 128 MB RAM, operating under IRIX 5.3.

The described algorithm was implemented in the new program *crystpack* written in C (≈ 20,000 lines source code) using Motif[42] as graphical user interface and OpenGL[43] for rendering. The program is designed for use under Unix and Linux platforms and is available as source code via www.rzuser.uni-heidelberg.de/~il1/crystpack.html. The program makes use of the SGInfo[44] package. For the purpose of illustration the ZORTEP[45] version of ORTEP[46] and the program package InsightII[47] were used.

Acknowledgments

Fruitful discussions with the late Dr. Laszlo Zsolnai about crystallographic details are gratefully acknowledged, as are the helpful, and at times even enlightening discussions with colleagues Dr. J. Hunger, Dr. S. Beyreuther, and V. Schulz.

[1] H. B. Bürgi, J. D. Dunitz, *Structure Correlation Volume 1 Part II*, VCH, Weinheim, **1994** and references therein.
[2] See for instance: J. Hunger, S. Beyreuther, G. Huttner, K. Allinger, U. Radelof, L. Zsolnai, *Eur. J. Inorg. Chem.* **1998**, 693–702 and references therein.
[3] F. H. Allen, J. E. Davis, J. J. Galloy, O. Johnson, O. Kennard, C. F. Macrae, E. M. Mitchell, G. F. Mitchell, J. M. Smith, D. G. Watson, *J. Chem. Inf. Comput. Sci.* **1991**, 31, 187–204.
[4] See for instance: L. Smart, E. Moore, *Solid State Chemistry: An Introduction*, Chapman Hall, London, **1992**.
[5] A. I. Kitaigorodskii, *J. Phys.* **1945,** 9, 351–352.
[6] A. I. Kitaigorodskii, *Organic Chemical Crystallography*, Consultants Bureau, New York, **1961**.
[7] J. M. Lehn, *Proc. Robert A. Welch Found. Conf. Chem. Res.* **1995**, 11–19.
[8] G. R. Desiraju, C. V. K. Sharma, *Perspect. Supramol. Chem.* **1996**, 31–61.
[9] R. G. Desiraju in: *Compr. Supramol. Chem.* (Eds.: D. D. MacNicol, F. Toda, R. Bishop), Elsevier, Oxford UK, **1996**, Volume 6, p. 1–22.
[10] A. I. Kitaigorodskii, *Molekülkristalle*, Akademie, Berlin, **1979**.

[11] A. I. Kitaigorodskii, *Advan. Struct. Res. Diffr. Methods* **1970**, 3, 173–247.
[12] J. A. Pertsin, A. I. Kitaigorodskii, *The Atom-Atom Potential Method. Applications to Organic Molecular Solids*, Springer Series in Chemical Physics, Vol. 43, Springer, Berlin, **1987**.
[13] A. Gavezzotti, G. Filippini, *Mol. Cryst. Liq. Cryst. Sci. Technol., Sect. A* **1992**, 219, 37–41
[14] W. D. S Motherwell, *Acta Cryst.* **1997**, B53, 726–736.
[15] K. P. Kratzer, *Neuronale Netze – Grundlagen und Anwendungen*, Hanser, Hamburg, **1993**.
[16] H. Ritter, T. Martinez, K. Schulten, *Neuronale Netzwerke – Eine Einführung in die Neuroinformatik selbst-organisierender Netzwerke*, Addison-Wesley, Bonn, **1994**.
[17] A. Zell, *Simulation Neuronaler Netze*, Addison-Wesley, Bonn, **1994**.
[18] J. Zupan, J. Gasteiger, *Neural Networks in Chemistry*, VCH, Weinheim, **1993**.
[19] A. F. Wells, *Structural Inorganic Chemistry*, Fifth Edition, Clarendon Press, Oxford, **1984**.
[20] U. Müller, *Anorganische Strukturchemie*, Teubner, Stuttgart, **1992**.
[21] G. Harsch, R. Schmidt, *Kristallgeometrie – Packungen und Symmetrie in Stereodarstellungen,* Diesterweg, Frankfurt, **1981**.
[22] J. S. Rollett, *Computing Methods in Crystallography*, Pergamon Press, London, **1965**.
[23] G. W. Snedecor, W. G. Cochran, *Statistical Methods*, Iowa State University Press, Ames, **1968**.
[24] F. C. Frank, *Phil. Mag. Letters* **1992**, 66, 81–84.
[25] G. M. Crippen, T. F. Havel, *Distance Geometry and Molecular Conformation,* Research Studies Press Ltd., Sommerset, **1988**.
[26] J. M. Blaney, J. S. Dixon, in: *Reviews in Computational Chemistry Volume 5 Chapter 6* (Ed.: K. B. Lipkowitz, D. B. Boyd), VCH, Weinheim, **1994**.
[27] R. Wierl, *Ann. Phys.* **1931**, 8, 521–564.
[28] J. H. Schuur, P. Selzer, J. Gasteiger, *J. Chem. Inf. Comput. Sci.* **1996**, 36, 334–344.
[29] L. J. Soltzberg, C. L. Wilkins, *J. Am. Chem. Soc.* **1977**, 99, 439–443.
[30] A condensed and easy to read essay on the underlying philosophical problems is to be found in: D. H. Rouvray in: *Fuzzy Logic in Chemistry* (Ed.: D. F. Rouvray), Academic Press, **1997**, chapter 1, p. 1–29; the problem of measuring symmetry has been elegantly treated in: M. Pinsky, D Avnir, *Inorg. Chem.* **1998**, *37*, 5575–5582; D. Avnir, O. Katzenelson, S. Keinan, M. Pinsky, Y. Pinto, Y. Salomon, H. Zabrodsky Hel-Or, *Concepts in Chemistry* (Ed. D. H. Rouvray), Research Studies Press, Somerset, **1997**, chapter 9, p. 283–324.
[31] J. W. Goethe, *Goethes Werke, Vollständige Ausgabe letzter Hand*, Cottasche Buchhandlung, Stuttgart und Tübingen, **1829** Band 22 p. 238, Betrachtungen im Sinne der Wanderer. "Die Theorie an und für sich ist nichts nütze, als insofern sie uns an den Zusammenhang der Erscheinungen glauben macht"; translation: Theory in itself is of no use. It's only purpose is to make us believe in the coherence of phenomena.
[32] H. Goldstein, *Klassische Mechanik*, Akademische Verlagsgesellschaft, Frankfurt, **1963**, V. Kapitel.
[33] D. M. P. Mingos, A. L. Rohl, *J. Chem. Soc., Dalton Trans.* **1991**, 3419–3425.
[34] D. M. P. Mingos, A. L. Rohl, *J. Chem. Soc., Dalton Trans.* **1992**, 3541–3552.
[35] *International Tables for Crystallography*, Volume A 3rd edition, (Ed.: T. Hahn), Kluwer Academic Press, Boston, **1992**, p. 77.
[36] The name is the REFCODE under which the compound is stored in the CSD (see ref. [3]).
[37] D. Braga, F. Grepioni, P. Sabatino, *J. Chem. Soc., Dalton Trans.* **1990**, 3137–3142.
[38] Files containing the coordinates of the polyhedra and the corresponding qualifiers can be accessed over the internet at: www.rzuser.uni-heidelberg.de/il1/crystpack.html.
[39] A. Zell et al., *SNNS User Manual*, Version 4.1, **1995**, http://www.informatik.uni-stuttgart.de/ipvr/bv/projekte/snns/snns.html.
[40] S. Beyreuther, J. Hunger, G. Huttner, S. Mann, L. Zsolnai, *Chem. Ber.* **1996**, *129*, 745–757.
[41] J. Hunger, G. Huttner, *J. Comput. Chem.* **1999**, *20*, 455–471.
[42] M. Brain, *Motif Programming – The Essentials ... and More*, Digital, Newton, **1992**.
[43] M. J. Kilgard, *OpenGL – Programming for the X Window System*, Addison–Wesley, New York, **1996**.
[44] R. W. Grosse-Kunstleve, *Space Group Info's* © **1994–96**, http://www.csb.yale.edu/sginfo
[45] L. Zsolnai, G. Huttner, *University of Heidelberg*, **1994**, http://www.rzuser.uni-heidelberg.de/≈v54/xpm.html.
[46] C. K. Johnson, *ORTEP-II: A FORTRAN Thermal-Ellipsoid Plot Program for Crystal Structure Illustrations*, Oak Ridge National Laboratory Report ORNL-5138, **1976**, http://www.ornl.gov/ortep/ortep.html.
[47] Commercial program package from MSI see: http://www.msi.com.

Received September 9, 1999
[I99320]