communications in mathematical
match
and in computer chemistry

no. 34, October 1996

ISSN 0340-6253

MATCDY (34) 123-155 (1996)

SPACE MODELS OF MOLECULES BASED ON INTERATOMIC DISTANCES AND POINT SYMMETRY GROUP *

Jens-Uwe Grabowski, Juliana Grell

Institut für Algebra, Technische Universität Dresden,
01062 Dresden, Germany
[grabo, grell]@math.tu-dresden.de

Lubov Zlatina†

Department of Mathematics and Computer Science,
Ben-Gurion University of the Negev,
84105 Beer Sheva, Israel

Abstract

A new approach to the elaboration of a three-dimensinal model of a chemical molecule is outlined. We propose an algorithm which uses as input interatomic geometric parameters, the molecular graph and a prescribed point symmetry group of the molecule. The output are three-dimensional coordinates of the atoms satisfying the prescribed combinatorial and metric requirements and admitting the prescribed symmetry group. The steps of the algorithm are demonstrated on a practical example. Relations between graph automorphism groups and point symmetry groups are explained.
Keywords: molecular model, molecular graph, graph automorphism group, point symmetry group, group embedding, Crippen algorithm (principal component analysis)

1 Introduction

1.1. This paper is concerned with a new mathematical method of building spatial models of molecules.
Since the structure of a molecule and the behavior of its spectrum are correlated, combinatorial and geometric molecular models are useful in order to identify

*This research was partially supported by the Grant No. I-0333-263.06/93 from the G.I.F., the German-Israeli Foundation for Scientific Research and Development

†Current affilation: BAIN, Haamazger 1, Raanana, Israel

structures from their spectra and, vice versa, to predict spectra of given structures, cf. [Gri83], [ESMZKKY91], [ZE92]. Spatial models allow the molecular design of new chemical compounds with given properties, they support the search for new ways of the synthesis of compounds. Especially, molecular spatial models play a crucial role in several applications of artificial intelligence systems in chemistry. We mention in this connection computer aided design, expert systems, QSAR (qualitative and quantitative structure - activity relations). Here, one of the most attractive goals is the prediction of biological activity on the base of some geometric and/or combinatorical descriptions of the molecules.
Experimental methods yielding the needed real data and forming the physico chemical background in this field are molecular spectroscopy, spectroanalysis, molecular mechanics and quantum chemistry. The obtainable data considered here are the number and kinds of atoms (brutto formula of the compound), presence of bonds, metric parameters defining near interatomic distances (bond lengths, valency angles, possibly dihedral angles) and supposed point symmetry group.
We call the model we want to get here the *symmetrized molecular model.* It has to unite the information of two classical models — the Crippen model and the molecular graph — with the symmetry group of the molecule.

1.2. Methods of construction of spatial models. There is a large number of papers concerned with the determination of three dimensional Cartesian coordinates from interatomic geometrical parameters. We cite as examples [Ess83], [GP72], [NBB85], [Cri77], [Cri78], [Cri81], [Dav86], [WH86], [BS83], [ZM87], [Hil69], [GSh73], [VN78], [Kle89].
All algorithms described in [Ess83], [GP72], [NBB85], [Cri77], [Cri78], [Dav86], [WH86], [BS83] do not take into account symmetry at all. Moreover, they do not allow a correction of calculated geometric parameters. Paradoxically, often, these algorithms lead to models where the geometric and the combinatorial part of information contradict each other, namely to models where cycles are not closed. So, these methods may only yield a first approximation and be used for non-cyclic or less complicated cyclic structures.
The Crippen algorithm yielding an essential contribution to our symmetrization algorithm has also the above mentioned disadvantages. The Crippen model is a molecular model given in Cartesian coordinates determined from internal geometric parameters by a principal component analysis. The real molecule's symmetry is not taken into account in the Crippen's algorithm. Sometimes this is not evident, but there are cases where the molecular model determined by Crippen's algorithm does not have the required symmetry. So for example, the spatial model of trans-decalin [EKDG69] obtained by this method has the symmetry $C_2 = 2$ instead of the real symmetry $C_{2h} = 2/m$ (for symbols of symmetry groups see 2.11) and the geometrical parameters become strongly deviated. The authors of [WS82] propose to use more advanced methods of quantum mechanics

and quantum chemistry to augment the exactness.
Consideration of symmetry makes easier to predict vibrational spectra. Methods described in [ZM87] and [Hil69] take into account the symmetry, but however, not in a systematic way. The determination of atoms' site symmetry and the orbits[1] considered in these papers is a procedure of trial and error to be fulfilled by the user and demanding some experience. So, for a practically working chemist this procedure might be hard to apply.
The GO-SHERAGA algorithm described in [GSh73] is able to determine only one symmetry element via the solution of about 100 transcendental equations. This algorithm is not universal and contains an extensive numerical procedure.

1.3. Our paper is based on the PhD thesis [Zla91] of one of the authors, L. Zlatina. This thesis was worked out within the scope of a rather general project: to elaborate a package of computer programs for the enumeration of possible stereoisomers to be used in the practical work of chemists. The systematical use of the information about the symmetry of the molecules was outlined from the early beginning as its essential feature (see [ZE92]).
The mathematical background of the initial approach was revised by the other two authors who also wrote new sections dealing with graph automorphism groups and symmetry groups. They feel that this impulse started from a physico-chemical problem generates more mathematically interesting questions being worth to have a solid mathematical foundation. We hope that our paper will attract the attention of the readers to deep algebraic problems which are on the edge between physico-chemical analysis and theoretical stereochemistry.

1.4. Our aim is to describe the main features of an algorithm — we call it the *symmetrization algorithm* — which elaborates a symmetrized molecular model. During a few years L. Zlatina was elaborating numerous subroutines which were supposed to become a part of a future software package. Unfortunately, the entire package is still far away from practical exploitation. However, numerous experimental use of some subroutines had helped the authors to outline the main steps of the algorithm more accurately.

1.5. Our paper can be divided into four parts:
After *part 1* — the introduction — *part 2* consisting of subsections A, B and C gives a comprehensive presentation of the theoretical background of the symmetrization algorithm.
In particular, subsections 2A and 2B are devoted to the group theoretical foundations: permutation groups as automorphism groups of graphs and general point symmetry groups and the relationships between them. Both kinds of groups are regarded in a large number of pure theoretical as well as applied works. We were mainly interested to show the links between these two kinds of groups describing

[1] all used notions are explained below

different aspects (combinatorial and geometric) of the symmetry of a molecule. In 2C we briefly present the Crippen algorithm. With this, we had less intention to demonstrate in detail this classical method here than to make plausible that it can yield only an approximative molecular model whenever the considered molecule is not very simple.
Part 3 consists of a detailed description of the symmetrization algorithm. Almost each step is illustrated with the respective calculations for Aluminium Borohydride $Al\,B_3H_{12}$ [AGH68] as an example.
Part 4 contains some final remarks and the bibliography.
This paper is written for both chemists and mathematicians. In order to read it no special mathematical background is necessary; the essential notions are explained in detail. However, some initial experience with mathematical notions, in particular, mathematical modelling in chemistry will be certainly helpful.

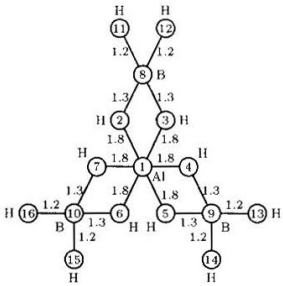
2 Theoretical background

A. Permutation groups and automorphism groups of graphs

2.1. Graphs and molecules. We understand molecules as consisting of atoms and bonds (balls and sticks are sometimes a helpful approximation of such idea). On a more rigorous level this leads to a combinatorial structure which is called graph. As one example to get a survey in the field of application of graphs to molecules we cite [KRRT95].
A (non-directed) *graph* $\Gamma = (V, E)$ consists of a set V of vertices and a set E of edges, where an edge $e \in E$ is a non-ordered pair $e = \{v, w\}$ of distinct vertices $v, w \in V$. Non-ordered means that we don't distinguish between $\{v, w\}$ and $\{w, v\}$. Commonly, vertices are labelled by natural numbers.
A *coloured graph* $\Gamma = (V, E, A, B)$ is a graph with additional label sets — the "colours" — A and B, where each vertex $v \in V$ is additionally labelled by an element $a \in A$ (symbolically $[v, a]$) and each edge $e = \{v, w\} \in E$ is labelled by its colour $b \in B$ (we write $e = \{v, b, w\}$ or $v \overset{b}{\text{---}} w$).

2.2. Example $Al\,B_3H_{12}$. Let $V = \{1, 2, \ldots, 16\}$ be the set of atoms of a molecule; E the set of bonds[2]. Let the colours for the atoms be the names of the atoms and the colours for the bonds be the bond lengths. Then we can represent the molecule of $Al\,B_3H_{12}$ with some additional information by the following coloured graph:

[2]Here by bonds we mean not necessarily chemical bonds but also some distinguished pairs of neighbouring atoms.



Generally, to each molecule there can be assigned a coloured graph in this or in a similar way (e.g. one can agree to label the multiplicity of bonds by colours). The graph above is called *coloured molecular graph* of the molecule.

2.3. Groups. Symmetry properties of graphs as well as of geometric figures can be described by a suitable set of mappings transforming the respective object into itself. These sets often have the algebraic structure of a group.
An abstract *group* $(G, \circ)$ consists of a set G and a binary operation $\circ$, such that the following conditions hold:

1. For all $g, h, k \in G$: $g \circ (h \circ k) = (g \circ h) \circ k$ (law of associativity)
2. There exists an element $id \in G$ such that $g \circ id = id \circ g = g$ holds for all $g \in G$ (existence of an identity element)
3. For any element $g \in G$ there exists an element $g^{-1} \in G$ such that $g \circ g^{-1} = g^{-1} \circ g = id$ holds (existence of inverse elements).

The element $g \circ h$, often denoted by gh, is called the *product* of g and h.
Examples for groups are the set of integers with the usual addition as operation, permutation groups (see below, 2.4) and symmetry groups (discussed in detail in section 2B). In the case of symmetry groups, the set G consists of motions in the 3- or n-dimensional Euclidean space (rotations, reflections etc.) and the binary operation $\circ$ means the consecutive application of two motions. Of course, the identity is the "motion" which leaves all points of the space unchanged and the inverse element of a motion is the reverse motion of it.
A subset $G' \subseteq G$ of a group G is called a *subgroup* of G, notation $G' \leq G$, if G' with the operation $\circ$ (as defined in G restricted to the elements in G') is also a group. This hols if and only if for any $g, h, k \in G'$ also g^{-1} and $h \circ k$ belong to G'.
If $G' \leq G$ and $h \in G$ then the subsets $G'h$, respectively, $hG' \subseteq G$ are called *right*,

respectively, *left cosets of G with respect to G'*. The number of right cosets and the number of left cosets of G w.r.t. G' are always equal and they are called (if finite) the *index of G' in G*.

Let $K \subseteq G$ be an arbitrary subset of a group G. The set G' of all elements obtained by repeated forming of products of elements of K and taking the inverses from K until no new elements occur is called the subgroup *generated by K*, denoted by $G' = \langle K \rangle$ and in such a case K is called a set of generators for G'. We stress that here $G' \leq G$.

Let $(G_1, \circ_1)$ and $(G_2, \circ_2)$ be two groups. A mapping φ from G_1 to G_2 is called a *homomorphism*, if for all $g, h \in G_1$

$$\varphi(g \circ_1 h) = \varphi(g) \circ_2 \varphi(h).$$

If, in addition, such a mapping is bijective, i.e. a one-to-one correspondence onto, then it is called an *isomorphism* and the groups $(G_1, \circ_1)$ and $(G_2, \circ_2)$ are *isomorphic*, notation $(G_1, \circ_1) \cong (G_2, \circ_2)$. Two isomorphic groups are "the same" up to the names of the elements.

If there is an isomorphism from G_1 into a subgroup of G_2 then this isomorphism is called an *embedding* of $(G_1, \circ_1)$ into $(G_2, \circ_2)$.

2.4. Permutations. A *permutation* g on a set V is a one-to-one mapping of V onto itself, i.e. g maps each element $v \in V$ onto an element v^g of V and for any element w there is exactly one element v which is mapped via g onto w. Note, that we use here the "exponential" notation v^g for the image of v instead of traditional $g(v)$. The *inverse permutation* g^{-1} of a permutation g maps v^g onto v. A *fixed point* of a permutation g is an element $v \in V$ which is mapped onto itself by g: $v^g = v$.

The *product* $g \circ h$, for short gh, of two permutations g and h is the consecutive application of the mappings g and h, i.e. $v^{gh} := (v^g)^h$.

A *cycle* $c = (v_1 v_2 v_3 \ldots v_r)$ (where the v_i are distinct elements of V) is the permutation which maps v_1 to v_2, v_2 to v_3, ... , v_r to v_1, and leaves all elements of V not occuring in c unchanged. A permutation g can be denoted by its *cycle representation*, i.e. a product $g = c_1 c_2 \ldots c_l$ of cycles $c_1, c_2, \ldots, c_l$, where different cycles involve disjoint sets of elements.

If G is a set of permutations on a set V and the product of two permutations from G as well as the inverse permutation of any permutation from G belong again to G, then G satisfies the conditions 1.-3. for groups and we call the pair (G, V) a *permutation group* on V. Let (G, V) be a permutation group and $v \in V$. The set $v^G := \{v^g \mid g \in G\}$ of all images of v under the permutations of G is called the *orbit* of v under G.

The *k-orbit* of a k-tuple of elements of V is defined analogously:

$$(v_1, v_2, \ldots, v_k)^G := \{(v_1^g, v_2^g, \ldots, v_k^g) \mid g \in G\}.$$

2.5. Automorphisms of graphs. Let $\Gamma = (V, E, A, B)$ be a coloured graph. An *automorphism* g of Γ is a permutation of the vertices V "respecting" all edges and all colours: More precisely, the image v^g of a vertex v under g must have the same colour as v and if there is an edge $e = \{v, w\}$ then $\{v^g, w^g\}$ must be an edge with the same colour as e, i.e.

$$[v, a] \quad \text{maps to} \quad [v^g, a] \quad \text{and} \quad v \overset{b}{\text{---}} w \quad \text{maps to} \quad v^g \overset{b}{\text{---}} w^g.$$

The set of all automorphisms of a coloured graph Γ forms a permutation group which is called the *automorphism group* of the graph, denoted by $\mathbf{Aut}\,(\Gamma)$.

2.6. Example $\mathrm{Al\,B_3H_{12}}$ **continued.** Let
$g_1 = (2\,4\,6)(3\,5\,7)(8\,9\,10)(11\,13\,15)(12\,14\,16)$, $g_2 = (2\,3)$ and $g_3 = (11\,12)$.
These are permutations of the vertices of Γ which are in fact automorphisms of Γ.
The set of fixed points of g_1 and g_2, respectively, are $\{1\}$ and $\{1, 4, 5, \ldots, 16\}$, respectively.
We have $g_1 g_2 = (2\,4\,6\,3\,5\,7)(8\,9\,10)(11\,13\,15)(12\,14\,16)$ and
$g_1^{-1} = (2\,6\,4)(3\,7\,5)(8\,10\,9)(11\,15\,13)(12\,16\,14)$.
The group generated by g_2 and g_3 is $G = \langle g_2, g_3 \rangle = \{id, g_2, g_3, g_2 g_3\} = \{id, (2\,3), (11\,12), (2\,3)(11\,12)\}$ and $G' = \{id, (2\,3)(11\,12)\}$ is a subgroup of G.
Let $\mathsf{R}_0, \mathsf{R}_1$, respectively, be the rotation about $0, \pi$, respectively, around an arbitrary axis in the space. Then $\mathsf{R}_0, \mathsf{R}_1$ form a group $\mathfrak{R}$ isomorphic to G'. In fact we have the one-to-one correspondence $h : id \mapsto \mathsf{R}_0$, $(2\,3)(11\,12) \mapsto \mathsf{R}_1$. By the same correspondence, $\mathfrak{R}$ can be embedded into G.
The orbits of G are $2^G = 3^G = \{2, 3\}$, $11^G = 12^G = \{11, 12\}$ and $v^G = \{v\}$ for all other $v \in V$. An example of a 2-orbit is given by $(6, 2)^G = \{(6, 2), (6, 3)\}$.
It turns out that altogether the molecular graph of $\mathrm{Al\,B_3H_{12}}$ has 384 automorphisms. A set generating this group is
$\{\,(8\,9)(2\,4)(3\,5)(11\,13)(12\,14), (8\,10)(2\,6)(3\,7)(11\,15)(12\,16), (2\,3), (11\,12)\,\}$.
We want to state here explicitely as a warning: most of such automorphisms do not need to correspond to a geometrical symmetry.

B. Symmetry groups

Our notations for groups, group elements, their multiplication, mappings etc. in this subsection follows as far as it seemed us possible the standard book about symmetry groups [Hah83] and that's why sometimes it deviates from the usual algebraic notation introduced in the preceeding subsection 2A. We think that we should not ignore the language of crystallographers who we expect among the audience of our paper and who are mostly occupied with this kind of groups and succeeded — often after long discussions — in creating a well recognized common system of notions and notations. So, we decided to give the general group theoretical facts (in 2A) in an algebraic, the special symmetry group facts

(in 2B) in a crystallographic language. In the last item 2.12 of 2B symmetry groups will be considered as permutation groups and notions as well as notations of both approaches are compared.

2.7. The Euclidean space. Crystals and molecules are objects in the three-dimensional (for short also 3D) space we live in. A general model for the mathematical treatment of this space is the Euclidean space $\mathbb{R}^n$ over the field $\mathbb{R}$ of real numbers represented here by

$$\mathbb{R}^n := \{(a_1, \ldots, a_n) | a_i \in \mathbb{R}\}.$$

The cases $n = 1, 2, 3$ lead to models seeming to be most closely connected with nature; but in some cases, e.g. with certain kinds of aperiodic crystal structures, also other natural numbers $n > 3$ make it possible to construct even more useful models.

In our paper we shall consider molecular models in $\mathbb{R}^3$, except in the Crippen algorithm where the molecule is originally assumed to live in $\mathbb{R}^n$.
We will interprete the elements $(a_1, \ldots, a_n) \in \mathbb{R}^n$ sometimes as points A, sometimes as vectors $\mathbf{a} := \vec{OA}$, where $O := (0, \ldots, 0)$ is the n-dimensional zero vector and write them

$$\text{as rows } (a_1, \ldots, a_n) \text{ or as columns } \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = (a_1, \ldots, a_n)^T.$$

In $\mathbb{R}^n$ (considered as a metric vector space over $\mathbb{R}$) we have the following operations:

(i) *addition* of any two vectors $\mathbf{a}$ and $\mathbf{b}$:

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) := (a_1 + b_1, \ldots, a_n + b_n),$$

(ii) *multiplication* of any vector by any scalar $r \in \mathbb{R}$:

$$r(a_1, \ldots, a_n) := (ra_1, \ldots, ra_n),$$

in particular we have for the *opposite vector*

$$-(a_1, \ldots, a_n) = (-a_1, \ldots, -a_n)$$

and

(iii) the *scalar product* of any two vectors:

$$(\mathbf{a}, \mathbf{b}) := a_1 b_1 + \cdots + a_n b_n.$$

In $\mathbb{R}^n$ *metric relations* are defined as follows:

- The *distance* between two points $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$ equals the length of the vector $\mathbf{b} - \mathbf{a}$, which also will be denoted by $\vec{AB} := \mathbf{b} - \mathbf{a}$,

$$|\vec{AB}| = |\mathbf{b} - \mathbf{a}| := \sqrt{(a_1 - b_1)^2 + \cdots + (a_n - b_n)^2}$$

and

- the *angle* $\angle(A, B, C)$ between $\vec{BA}$ and $\vec{BC}$ is obtained from

$$\cos\angle(A, B, C) = \cos\angle(\vec{BA}, \vec{BC}) = \frac{(\vec{BA}, \vec{BC})}{|\vec{BA}| \cdot |\vec{BC}|} = \frac{(\mathbf{a} - \mathbf{b}, \mathbf{c} - \mathbf{b})}{|\mathbf{a} - \mathbf{b}| \cdot |\mathbf{c} - \mathbf{b}|}.$$

2.8. Motions. A bijective mapping[3]

$$\mathsf{W}: \quad \mathbb{R}^n \to \mathbb{R}^n: \quad X \mapsto \tilde{X},$$

whereby all distances are left invariant, is called an *isometry*, *isometric mapping*, *symmetry operation* or *motion* of $\mathbb{R}^n$.
Any motion W may by represented as

$$\mathsf{W}(X) = \tilde{X} = \mathbf{W}X + W_t,$$

where $X = (x_1, \ldots, x_n)^T$, $W_t = (w_1, \ldots, w_n)^T \in \mathbb{R}^n$ and $\mathbf{W} = (w_{ij})$ is an $(n \times n)$-matrix consisting of elements $w_{ij} \in \mathbb{R}$. Then we write $\mathsf{W} = (\mathbf{W}, W_t)$.
For any isometry $\mathsf{W} = (\mathbf{W}, W_t)$ we have

$$det(\mathbf{W}) = 1 \text{ or } -1.$$

A point $X \in \mathbb{R}^n$ is called a *fixed point* of the mapping W if $\mathsf{W}(X) = X$. The origin O is a fixed point if and only if $W_t = O$.
The symmetry operations describing a molecule's symmetry under consideration in this paper are motions of $\mathbb{R}^3$ with at least one fixed point. Let be $\mathbf{I}$ the (3×3)-unit matrix and $-\mathbf{I}$ the (3×3)-matrix obtained from $\mathbf{I}$ by multiplication of all elements with -1. The following kinds of motions with fixed point occur in $\mathbb{R}^3$:
a) Proper motions, i.e. motions with $det(\mathbf{W}) = 1$. They are *rotations*.
b) Improper motions, i.e. motions with $det(\mathbf{W}) = -1$. They are called *inversions* if $\mathbf{W} = -\mathbf{I}$; *reflections* if $\mathbf{W}^2 = \mathbf{I}$ and $\mathbf{W} \neq -\mathbf{I}$; and *rotoinversions* in all other cases.
Let be k a natural number. A rotation, respectively, rotoinversion about rotation angle $2\pi/k$, is called a k-fold rotation (then $\mathbf{W}^k = \mathbf{I}$), respectively, a k-fold rotoinversion (then $\mathbf{W}^k = \mathbf{I}$ for even k and $\mathbf{W}^k = -\mathbf{I}$ for odd k).

2.9. Symmetry elements. For the visualization of (geometrical) symmetry within the Euclidean space the concept of *symmetry elements* is useful. In first approximation the symmetry element of a symmetry operation with fixed point is

[3]the arrow $\mapsto$ describes the mapping elementwise, i.e. $\tilde{X} = \mathsf{W}(X)$

the set of its fixed points, together with a geometric interpretation of the motion. This statement is an exact definition if the motion is a k-fold rotation, a reflection or an inversion, respectively. Then the corresponding symmetry element is equal to a *k-fold rotation axis*, a *mirror plane* or an *inversion center*, respectively.
A peculiar situation exists for k-fold rotoinversions (except when $k = 1$ where we have the inversion $\bar{1} = i$, and $k = 2$ which gives the reflection $\bar{2} \equiv m = \sigma$; for symbols cf. explanation in Table 2.9). A k-fold rotoinversion is the product of a k-fold rotation and an inversion. Its symmetry element consists of two components: the inversion point and the k-fold rotation axis passing through the inversion point. Both together define the *k-fold rotoinversion axis* as the corresponding symmetry element. Its single fixed point is the inversion center ($k \neq 2$). We remark that k-fold rotoinversions ($k = 1, 2, \ldots$) may also be considered as h-fold *rotoreflections*, i.e. as a product of a rotation about a h-fold rotation axis and a reflection on a plane perpendicular to it, where $h = k$ for $k = 4N$; $h = k/2$ for $k = 4N + 2$, and $h = 2k$ for $k = 2N + 1$. The corresponding symmetry element of the rotoreflection is a *rotoreflection axis* consisting of a h-fold rotation axis and a perpendicular to it mirror plane. The symmetry group concept of HERMANN and MAUGUIN bases on rotoinversions, that of SCHOENFLIES bases on rotoreflections[4], [Hah83].

For symmetry groups to be considered in the next items 2.10, 2.11 and symmetry elements various notations exist and are used for practical, historical or private reasons[5]. Most common are the *International symbols* — based on the full HERMANN-MAUGUIN symbols — and the SCHOENFLIES *symbols* for point groups. We introduce them both, as given in [Hah83]. Symmetry elements will also be denoted twice: by International symbols [Hah83] and symbols introduced in [Flu80] (close to the Schoenflies group symbols).
In TABLE 2.9 there are listed the symmetry elements in $\mathbb{R}^3$ corresponding to symmetry operations with fixed points. Each input in the table consists of a general description of the kind of the symmetry element (above) with all symbols and an oriented symmetry element, i.e. where the set of its points is indicated related to a Cartesian coordinate system $(O, \mathbf{x}, \mathbf{y}, \mathbf{z})$ (below).

[4] If the only symmetry operations under consideration are k-fold rotations and rotoreflections ($k = 1, 2, \ldots$) then the symmetry element of a motion can be defined as the set of points in special position (cf. 2.12) w.r.t. this motion together with an indication where these points are mapped to by this motion.

[5] "One would rather use a colleague's toothbrush then her / his notations." — Orally information from H. Wondratschek.

General International symbol / oriented	Symbol according to [Flu80]	Interpretation and points of the symmetry element
k $k = 1, 2, \cdots$ oriented: k_{xx_0, xy_0, xz_0}	C_k	k-fold rotation axis, rotation angle $\varphi = 2\pi/k$ $\{(xx_0, xy_0, xz_0) \mid x \in \mathbb{R}\}$
$\bar{k}$ $\bar{1}$ oriented: $\bar{k}_{xx_0, xy_0, xz_0}$, $(0,0,0)$	 S_k $(k = 4N)$ S_{2k} $(k = 2N+1)$ $S_{\frac{k}{2}}$ $(k = 4N+2)$ i	k-fold rotoinversion axis, product of rotation about angle $\varphi = 2\pi/k$ and inversion with common fixed point inversion center $\{(xx_0, xy_0, xz_0) \mid x \in \mathbb{R}\}$ inversion point: $(0,0,0)$
m $(= \bar{2})$ oriented: $m_{xx_1+yx_2, xy_1+yy_2, xz_1+yz_2}$ m_{xy0} m_{0yz}	σ σ_h σ_v	mirror plane $\{(xx_1 + yx_2, xy_1 + yy_2, xz_1 + yz_2) \mid x, y \in \mathbb{R}\}$ $\{(x, y, 0) \mid x, y \in \mathbb{R}\}$ $\{(0, y, z) \mid y, z \in \mathbb{R}\}$

TABLE 2.9. Symmetry elements with fixed points in $\mathbb{R}^3$.

We give some concrete symmetry elements, with mentioned orientation related to a Cartesian coordinate system, i.e. the set of points of the symmetry element and the corresponding transformation matrix $\mathbf{W}$ (the translation part W_t is O):

k_{00z}	$\{(0,0,z) \mid z \in \mathbb{R}\}$	$\begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$
2_{0y0}	$\{(0,y,0) \mid x \in \mathbb{R}\}$	$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$
$\bar{k}_{00z}$, $(0,0,0)$	$\{(0,0,z) \mid z \in \mathbb{R}\}$	$\begin{pmatrix} -\cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & -\cos\varphi & 0 \\ 0 & 0 & -1 \end{pmatrix}$
$\bar{1}_{000}$	$\{(0,0,0)\}$	$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$
$m_{x,y,0}$	$\{(x,y,0) \mid x,y \in \mathbb{R}\}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$
$m_{0,y,z}$	$\{(0,y,z) \mid y,z \in \mathbb{R}\}$	$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

2.10. Symmetry groups. Let $\mathsf{W} = (\mathbf{W}, W_t)$, $\mathsf{W}' = (\mathbf{W}', W_t')$. If the composition $\mathsf{W} \cdot \mathsf{W}'$ of two motions W and W'

$$(\mathsf{W} \cdot \mathsf{W}')(X) := \mathsf{W}(\mathsf{W}'(X)) = \mathbf{W}\mathbf{W}'(X) + \mathbf{W}(W_t') + W_t$$

is taken as multiplication, then all motions of $\mathbb{R}^n$ form a group $\mathfrak{B}$.
A *figure* F is a set of points of $\mathbb{R}^n$. A *symmetry operation of* F is a motion W which maps this figure onto itself, that is $\mathsf{W}(F) = F$. The set of all symmetry operations of F forms a subgroup of $\mathfrak{B}$ and is called the *symmetry group of* F.
The set of all motions of $\mathbb{R}^n$ with fixed point O also forms a group (the symmetry group of a single fixed point).
Various classifications of symmetry groups are useful. Two symmetry groups $\mathfrak{G}_1$ and $\mathfrak{G}_2$ belong to the same *type* if they are conjugated in $\mathfrak{B}$, i.e. if

$$\mathfrak{G}_1 = \mathbf{W}^{-1}\mathfrak{G}_2\mathbf{W}, \text{ for some } \mathbf{W} \in \mathfrak{B}.$$

2.11. General point groups. A *(general) point group* of $\mathbb{R}^n$ is a symmetry group, where all its elements (symmetry operations) have at least one fixed point $O \in \mathbb{R}^n$ in common.
Here we will consider point groups of finite order. This means in particular, that the rotations occuring in these point groups are only rotations about angles $2\pi/k$, where k is an integer.
We remark that the notion of point groups is often identified with *crystallographic point groups.* The latter are only those, which transform an n-dimensional (infinite) lattice into itself. This leads to certain restrictions to possible rotation angles $2\pi/k$, i.e. to k: in $\mathbb{R}^2$ and $\mathbb{R}^3$ only rotations about 1-, 2-, 3-, 4- and 6-fold axes are *crystallographically* possible.
In the 2D space $\mathbb{R}^2$ all finite point groups which can occur are the symmetry groups of regular k-gons and their subgroups.
In the 3D space $\mathbb{R}^3$ all finite point symmetry groups which can occur are the symmetry groups of k-gonal prisms, $k = 1, 2, \ldots$, of the five Platonian solids tetrahedron, cube, octahedron, dodecahedron, icosahedron and their subgroups (where cube and octahedron have the same symmetry group type and so do dodecahedron and icosahedron) [Wey52].
In TABLE 2.11, following on the next page, all types of finite general point groups in $\mathbb{R}^3$ are indicated each with a set of generators in symbols according to TABLE 2.9 and a concrete isomorphic permutation group.

General Int. symbol	Schoenflies symbol	Group order	Symmetry elements of generating symmetry operations	Isomorphic perm. group
$4N$-gonal system (single k-fold symmetry axis with $k = 4N$)				
k	C_k	k	k_{00z}	Z_k
$\bar{k}$	S_k	k	$\bar{k}_{00z}$	Z_k
k/m	C_{kh}	$2k$	$k_{00z},\ m_{xy0}$	$Z_k \times S_2$
$k22$	D_k	$2k$	$k_{00z},\ 2_{x00},\ 2_{x\cos(\frac{\varphi}{2}),x\sin(\frac{\varphi}{2}),z}$	D_k
kmm	C_{kv}	$2k$	$k_{00z},\ m_{x0z},\ m_{x\cos(\frac{\varphi}{2}),x\sin(\frac{\varphi}{2}),z}$	D_k
$\bar{k}2m$	$D_{\frac{1}{2}kd}$	$2k$	$\bar{k}_{00z},\ m_{x0z},\ 2_{x\cos(\frac{\varphi}{2}),x\sin(\frac{\varphi}{2}),z}$	D_k
k/mmm	D_{kh}	$4k$	$k_{00z},\ m_{xy0},\ m_{x0z},\ m_{x\cos(\frac{\varphi}{2}),x\sin(\frac{\varphi}{2}),z}$	$D_k \times S_2$
$(2N+1)$-gonal system (single k-fold symmetry axis with $k = 2N+1$)				
k	C_k	k	k_{00z}	Z_k
$\bar{k} = k \times \bar{1}$	C_{ki}	$2k$	$\bar{k}_{00z}$	$Z_k \times S_2$
$k2$	D_k	$2k$	$k_{00z},\ 2_{x00}$	D_k
km	C_{kv}	$2k$	$k_{00z},\ m_{x0z}$	D_k
$\bar{k}m$	D_{kh}	$4k$	$k_{00z},\ m_{xy0},\ m_{x0z}$	$D_k \times S_2$
$(4N+2)$-gonal system (single k-fold symmetry axis with $k = 4N+2$)				
k	C_k	k	k_{00z}	Z_k
$\bar{k}$	$C_{\frac{1}{2}kh}$	k	$\bar{k}_{00z}$	$Z_{\frac{k}{2}} \times S_2$
$\bar{2} \equiv m$	C_s	2	m_{xy0}	S_2
k/m	C_{kh}	$2k$	$k_{00z},\ m_{xy0}$	$Z_k \times S_2$
$k22$	D_k	$2k$	$k_{00z},\ 2_{x00},\ 2_{x\cos(\frac{\varphi}{2}),x\sin(\frac{\varphi}{2}),z}$	D_k
kmm	C_{kv}	$2k$	$k_{00z},\ m_{x0z},\ m_{x\cos(\frac{\varphi}{2}),x\sin(\frac{\varphi}{2}),z}$	D_k
$\bar{k}2m$	$D_{\frac{1}{2}kd}$	$2k$	$\bar{k}_{00z},\ m_{x0z},\ 2_{x\cos(\frac{\varphi}{2}),x\sin(\frac{\varphi}{2}),z}$	D_k
k/mmm	D_{kh}	$4k$	$k_{00z},\ m_{xy0},\ m_{x0z},\ m_{x\cos(\frac{\varphi}{2}),x\sin(\frac{\varphi}{2}),z}$	$D_k \times S_2$
Cubic system				
23	T	12	$2_{xxx},\ 3_{00z}$	A_4
$2/m\bar{3}$	T_h	24	$2_{00z},\ \bar{3}_{xxx}$	$A_4 \times S_2$
432	O	24	$4_{00z},\ 3_{xxx}$	S_4
$\bar{4}3m$	T_d	24	$3_{00z},\ \bar{4}_{x\sqrt{2},0,x}$	S_4
$4/m\,\bar{3}\,2/m$	O_h	48	$4_{00z},\ \bar{3}_{xxx}$	$S_4 \times S_2$
Icosahedral system				
235	I	60	$5_{00z},\ 3_{x\sin(37.38°),0,x\cos(37.38°)}$	A_5
$2/m\,\bar{3}\,\bar{5}$	I_h	120	$\bar{5}_{00z},\ \bar{3}_{x\sin(37.38°),0,x\cos(37.38°)}$	$A_5 \times S_2$

TABLE 2.11. The finite general point symmetry groups in $\mathbb{R}^3$.

The symbols[6] for concrete permutation groups in this table denote: Z_k — the cyclic group, D_k — the dihedral group (of order $2k$), S_k — the full symmetric group, A_k — the alternating group, each acting on k elements; $\times$ denotes the direct product. We only mention these permutation groups here. A more detailed understanding of the permutation group symbols is not obligatory for further considerations and their explanation would go beyond the scope of our paper (for details cf. e.g. [Wie64] or [KPR88]). We remark that point groups belonging to different types may be isomorphic (e.g. $422 = D_4$ (D_4 is here SCHOENFLIES symbol) and $4mm = C_{4v}$ are both isomorphic to the permutation group D_4). The classification of symmetry groups into types is a finer one than that by isomorphisms: two isomorphic groups may be of different types.

2.12. Symmetry groups and permutation groups. Any symmetry group $\mathfrak{G}$ may be considered as permutation group on the points of $\mathbb{R}^n$. This, however, is a set of cardinality of the continuum.
The orbit $A^{\mathfrak{G}}$, as introduced in 2.4, of a point $A \in \mathbb{R}^n$ under the action of $\mathfrak{G}$ is the set

$$A^{\mathfrak{G}} = \{B | B \in \mathbb{R}^n, \exists \mathsf{G} \in \mathfrak{G} : A^{\mathsf{G}} = B\}.$$

We wrote A^{G} instead of $\mathsf{G}(A)$ in order to emphasize that here the motion G is considered as a permutation.
Points belonging to the same orbit under the action of a symmetry group are said by crystallographers to have *equivalent positions*. One can restrict the action of a point symmetry group only to some (finitely many) of its orbits. Then we have a finite permutation group on a finite set of points. This will be done in further considerations. The greatest subgroup $\mathfrak{G}_A \leq \mathfrak{G}$ with the property

$$A^{\mathfrak{G}_A} = \{A\},$$

is what algebraists call the *stabilizer* and crystallographers the *site symmetry* of point A. If $\mathfrak{G}_A = 1 = C_1$ (C_1 is here SCHOENFLIES symbol), then the point A is said to be in *general position*, otherwise in a *special position*.
For point groups, the notion *multiplicity* of the corresponding site symmetry used in crystallographic symmetry group tables ([Hah83]) equals the *length* $|A^{\mathfrak{G}}|$ *of the orbit* $A^{\mathfrak{G}}$. It is equal to the index $|\mathfrak{G} : \mathfrak{G}_A|$ (LAGRANGE's theorem).The point A is in general position if and only if $|A^{\mathfrak{G}}| = |\mathfrak{G}|$.
We remark, that only certain subgroups of a symmetry group acting in $\mathbb{R}^n$ can figurate as site symmetry (stabilizer) of a point $A \in \mathbb{R}^n$. This becomes obvious by considering in some cyclic symmetry groups the corresponding symmetry elements and the possible positions of points of $\mathbb{R}^3$:

[6]**Warning !!!** As we know from our experience with foreign languages "false friends" can occur. This is what happens with Schoenflies symbol S_k, denoting a cyclic group of order $k = 4N$ and the permutation group symbol S_k for the full symmetric permutation group of order $k!$ on a set of k points — of course, they are non-isomorphic.

Let $\langle \mathsf{W}_k \rangle$ be the group generated by a k-fold rotation W_k, $k \geq 2$. Then

$|A^{\langle \mathsf{W}_k \rangle}| = 1$, if A lies on the axis, i.e. $\langle \mathsf{W}_k \rangle_A = \langle \mathsf{W}_k \rangle$ or

$|A^{\langle \mathsf{W}_k \rangle}| = k$, if A does not lie on the axis, i.e. A lies in general position.

With a k-fold rotoinversion W_k the situation is a little bit more delicate:

$|A^{\langle \mathsf{W}_k \rangle}| = 1$, if A is the fixed point of the rotoinversion or

$|A^{\langle \mathsf{W}_k \rangle}| = 2$, if A lies on the axis but is not the fixed point or

$|A^{\langle \mathsf{W}_k \rangle}| = |\langle \mathsf{W}_k^2 \rangle|$, if $k = 4N + 2$ and A lies not on the inversion point but on a mirror plane perpendicular to the axes through the inversion point or

$|A^{\langle \mathsf{W}_k \rangle}| = |\langle \mathsf{W}_k \rangle|$, if A lies in general position.

Finally, we summarize the most important pairs of occasions of terms.

Dictionary.

	Crystallographers:	Algebraists:
$A^{\mathfrak{G}}$	set of equivalent positions	orbit
$\lvert A^{\mathfrak{G}} \rvert$	number of equivalent positions, multiplicity[7]	orbit length
$\mathfrak{G}_A$	site symmetry of A	stabilizer of A
$\mathfrak{G}_A = 1$	A lies in general position	A has a trivial stabilizer $\neq \mathfrak{G}$
$\mathfrak{G}_A \neq 1$	A lies in a special position	A has a non-trivial stabilizer or is a fixed point

2.13. Automorphisms of the molecular graph and symmetry operations. The symmetrization algorithm presented in section 3 will yield an allocation of the atoms in the 3D space such that the given symmetry is satisfied, i.e. the group of symmetry operations of the desired figure belongs to a given type $\mathfrak{G}$.

We consider such allocation as a mapping φ which assigns to each atom (e.g. each vertex of the labelled molecular graph Γ) a triple of coordinates:

$$\varphi : V(\Gamma) \longrightarrow \mathbb{R}^3.$$

Of course, with this mapping we define an allocation not only for the atoms but also for the molecule, hence we assign to the labelled molecular graph Γ a figure F in the space (and we use for this assignment also the sign φ):

$$F = \varphi(\Gamma).$$

For computing such a φ we use the relations between the symmetry group $\mathfrak{G}_F$ of the molecule and the automorphism group $\mathbf{Aut}\,(\Gamma)$ of the molecular graph.

Let F be a figure corresponding to the molecule.

[7] This notion indicates the orbit length only for point groups. In general, the multiplicity indicates the number of equivalent positions within the unit cell.

Any motion W belonging to the symmetry group $\mathfrak{G}_F$ can be considered as a permutation on the set V of atoms of the molecule. We denote this permutation by w and the subset of $\mathbf{Aut}\,(\Gamma)$ of all permutations assigned in this way to the elements of $\mathfrak{G}_F$ by G_F.
The correspondence between the elements of $\mathfrak{G}_F$ and those of G_F should be one-to-one, but this is the case necessarily only for non-planar molecules: In the case of planar molecules it may happen $w = w'$ in spite of $\mathsf{W} \neq \mathsf{W}'$, and especially, it is possible that $w = id$ for some $\mathsf{W} \neq id$ (e.g. W is the reflection with respect to the plane the molecule lies in).
Moreover, any element of G_F permutes the atoms of the molecule preserving the names of the atoms and the bonds with their length and multiplicity etc. Hence, it is an automorphism of the labelled molecular graph: $w \in \mathbf{Aut}\,(\Gamma)$.
However, not every automorphism of the graph must correspond to a symmetry operation in $\mathbb{R}^3$, i.e. G_F is a subgroup of $\mathbf{Aut}\,(\Gamma)$ and
we can state:

The group G_F is a homomorphic image of $\mathfrak{G}_F$ and it is a subgroup of $\mathbf{Aut}\,(\Gamma)$*:*

$$\mathfrak{G}_F \longrightarrow G_F \leq \mathbf{Aut}\,(\Gamma).$$

In the case of non-planar molecules, G_F is isomorphic to $\mathfrak{G}_F$ and it is a subgroup of $\mathbf{Aut}\,(\Gamma)$*:*

$$\mathfrak{G}_F \cong G_F \leq \mathbf{Aut}\,(\Gamma).$$

Let now be given a type $\mathfrak{G}$ of the symmetry group $\mathfrak{G}_F$ and an automorphism group $\mathbf{Aut}\,(\Gamma)$, but let the coordinates of the figure F of the molecule in the $\mathbb{R}^3$ be unknown. We restrict our attention to the case of non-planar molecules.
We want to know, how the symmetry operations can act on the atoms of the molecule. From the type $\mathfrak{G}$ and the table in 2.11 we know a permutation group G isomorphic to $\mathfrak{G}_F$. Let $\Phi : G \longrightarrow \mathbf{Aut}\,(\Gamma)$ be a mapping assigning an automorphism of $\mathbf{Aut}\,(\Gamma)$ to each element of G. Generally, the mapping Φ is not uniquely determined. In order to define it, it is sufficient to assign an automorphism to each generator of G.

Some necessary conditions for such an assignment are easy to see, for instance, if an element of G corresponds to a symmetry operation W with $\mathsf{W}^k = id$, then one must assign to it a permutation $w \in \mathbf{Aut}\,(\Gamma)$ with $w^k = id$.
In principle, the problem of enumeration of all possible embeddings of the symmetry group $\mathfrak{G}_F$ into the automorphism group $\mathbf{Aut}\,(\Gamma)$ may be solved on the base of some standard group-theoretical techniques. Here the use of modern tools of computer algebra, for instance the computer program system GAP (see [Sch95]), is recommended. In general, the group embedding $\Phi : \mathfrak{G}_F \longrightarrow \mathbf{Aut}\,(\Gamma)$ does not guarantee that the image G_F of $\mathfrak{G}_F$ in $\mathbf{Aut}\,(\Gamma)$ is really realizable by a 3D model of the molecule. Additional conditions must be satified.

We mention one possibility: Given a faithful permutation representation, say (S, M),

of $\mathfrak{G}_F$ on a finite set $M \subseteq \mathbb{R}^3$ (which is invariant under the action of S), then there must exist a "faithful action homomorphism" from G_F into (S, M), that means there is a mapping $h : V \longrightarrow M$ of the vertex set V (not necessarily injective) such that

$$h(v^{\Phi(g)}) = h(v)^g, \quad \text{for all } g \in S \text{ and for all } v \in V$$

(of course, the action of g on $h(v)$ is the action in (S, M)).

2.14. Example $\mathrm{Al\,B_3H_{12}}$ **continued.** The symmetry group of $\mathrm{Al\,B_3H_{12}}$ is $\bar{6}2m = D_{3h}$ of order 12 generated by 3_{00z}, m_{xy0}, m_{0yz}. One embedding Φ : $\mathfrak{G}_F \longrightarrow \mathbf{Aut}\,(\Gamma)$ is given by the following assignment:

$$\begin{aligned} 3_{00z} &\mapsto (2\,4\,6)(3\,5\,7)(8\,9\,10)(11\,13\,15)(12\,14\,16) \\ m_{xy0} &\mapsto (2\,3)(4\,5)(6\,7) \\ m_{0yz} &\mapsto (2\,4)(3\,5)(6\,7)(8\,9)(11\,14)(12\,13)(15\,16) \end{aligned}$$

There exist a reasonably large number of possible assignments. Nevertheless only a small amount of them leads to essentially different embeddings of $\mathfrak{G}_F$ into $\mathbf{Aut}\,(\Gamma)$. Some of the possible embeddings may be omitted by the following elementary geometrical reasons : In some cases all four atoms of a component $\mathrm{Al}\begin{smallmatrix}\diagup \mathrm{H} \diagdown \\ \diagdown \mathrm{H} \diagup\end{smallmatrix}\mathrm{B}$ are stable under reflection on a horizontal ($m_{x,y,0} = \sigma_h$) and on a vertical mirror plane ($m_{x,0,z} = \sigma_v$); hence, they would lie in one line, which is impossible. In other cases there are pairs of atoms switched by a reflection on a horizontal mirror plane as well as by a reflection on a vertical mirror plane, what is again impossible. Thus finally only few possible embeddings are left to examine. We deliberately avoid here to discuss concrete details of such a group-theoretical search (if it is arranged via the use of a suitable computer package, a huge number of possibilities can be managed, intermediate details reducing the cases as early as possible are not significant). Here we restrict ourselves to the embedding given above which leads to the real model. The orbits of atoms w.r.t. this embedding are: $\{1\}$, $\{2,3,4,5,6,7\}$, $\{8,9,10\}$, $\{11,12,13,14,15,16\}$.

C. The Crippen algorithm.

2.15. The aim of Crippen's algorithm (cf. [Cri77], [Cri78], [Cri81]) is to construct a model CM of the molecule given by Cartesian coordinates in $\mathbb{R}^3$ using all interatomic distances as input data. Let n be the number of atoms in the molecule. The atoms are considerd as n points $A_1, \ldots, A_n$ in an n-dimensional space $\mathbb{R}^n$. The CM is an optimal projection of these n points in n dimensions into a three dimensional subspace, i.e. a projection where the distances between atoms are most similar to the distances in $\mathbb{R}^n$. The Crippen algorithm passes the following steps 2.16 - 2.19.

2.16. Collection of initial data. The input data are the $n \times n$ values d_{ij} of interatomic distances. Some of the d_{ij} are obtained from experiments. The remaining d_{ij} are searched in sufficient large intervals $[l_{ij}, u_{ij}]$, where lower and upper bounderies satisfy the triangle inequalities:

$$l_{ij} \geq l_{ik} - l_{kj} \quad \text{and} \quad u_{ij} \leq u_{ik} - u_{kj}.$$

Then the remaining $d_{ij} \in [l_{ij}, u_{ij}]$ are chosen with the help of a random number generator taking into account correlations between atom distances (again the triangle inequalities).

2.17. Calculation of n-dimensional atomic coordinates. Let $O \in \mathbb{R}^n$ be the geometrical center of the molecule. Its distance $d_{0i} := |\vec{OA_i}|$ from point (atom) A_i can be calculated as the difference between the geometrical means of the distances from point A_i and all distances:

$$d_{0i}^2 = \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 - \frac{1}{2n^2}\sum_{k=1}^{n}\sum_{j=1}^{n} d_{jk}^2.$$

Let us denote the searched coordinates of point A_i by

$$\mathbf{v}_i = (v_{i1}, \ldots, v_{in}) := \vec{OA_i} \quad \text{with} \quad |\mathbf{v}_i| = d_{0i}. \tag{1}$$

Then

$$(\mathbf{v}_i, \mathbf{v}_j) = d_{0i}d_{0j}\cos\angle(A_i, O, A_j) = (d_{0i}^2 + d_{0j}^2 - d_{ij}^2)/2$$

(via the cosine theorem).
Let

$$\mathbf{V} := \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{n1} & \cdots & v_{nn} \end{pmatrix} \quad \text{and} \quad \mathbf{G} = (g_{ij}) := \mathbf{V}\mathbf{V}^T, \tag{2}$$

where $\mathbf{V}^T$ is the transpose of the matrix $\mathbf{V}$ and thus $g_{ij} = (\mathbf{v}_i, \mathbf{v}_j)$. The eigenvalues λ_k and corresponding eigenvectors $\mathbf{w}_k$ of the matrix $\mathbf{G}$ can be determined e.g. by the partial searching method (cf. [FF63]). Moreover, the eigenvectors can be assumed to be found so that they are orthonormal, i.e.

$$(\mathbf{w}_k, \mathbf{w}_l) = \delta_{kl}, \tag{3}$$

where δ_{kl} is KRONECKER's symbol (i.e. $\delta_{kk} = 1$ and $\delta_{kl} = 0$ if $k \neq l$). Let $\mathbf{W}$ be the matrix which columns equal the eigenvectors of the matrix $\mathbf{G}$

$$\mathbf{W} := \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & & \vdots \\ w_{n1} & \cdots & w_{nn} \end{pmatrix}, \quad \mathbf{w}_k = \begin{pmatrix} w_{1k} \\ \vdots \\ w_{nk} \end{pmatrix},$$

and let

$$\mathbf{L} := \begin{pmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix} \quad \text{and} \quad \Lambda := \mathbf{LL}\ (= \mathbf{LL}^T).$$

Then

$$\mathbf{G} = \mathbf{W}\Lambda\mathbf{W}^T = \mathbf{WLL}^T\mathbf{W}^T = \mathbf{WL}(\mathbf{WL})^T$$

and according to (2) and (3) a solution for the desired n-dimensional coordinates satisfying (1) is determined by

$$\mathbf{V} = \mathbf{WL}, \quad \text{i.e.} \quad v_{ij} = w_{ij}\sqrt{\lambda_j}. \tag{4}$$

2.18. Determination of the 3D projection space. The eigenvalues λ_k, where $k = 1, \ldots, n$, are a measure of the correlation of deviations of atoms' k-th coordinates from 0. These eigenvalues are non-negative, because $\mathbf{G}$ is a symmetric matrix and positive semi definite (it can be shown, that its main minors are ≥ 0; cf. [Ant95], [Voe74]). The coordinate axes corresponding to the three greatest eigenvalues define the desired 3D subspace $\mathbb{R}^3$, allowing an optimal 3D projection of the molecule. Be $\mathbf{x}_i^{(0)} = (x_{i1}^{(0)}, x_{i2}^{(0)}, x_{i3}^{(0)})$, $i = 1, \ldots, n$, the coordinates of the n projected atoms considered as the first approximation of the 3D coordinates searced for the molecule.

2.19. Optimization of calculated atomic coordinates. Generally, the $\mathbf{x}_i^{(0)}$ would not yield sufficiently correct interatomic distances according to the given data. The larger the number n of participating atoms is the less reliable are calculated data. The deviations of atomic distances r_{ij} of the molecule's three dimensional projection from the input data $d_{ij} \in [l_{ij}, u_{ij}]$ (for the exactly given near interatomic distances we have $d_{ij} = l_{ij} = u_{ij}$) are decreased by minimization of the function F_{LU} of squares deviation:

$$F_{LU}(\mathbf{x}_1, \ldots, \mathbf{x}_n) := \sum_{\substack{1 \leq i < j \leq n \\ r_{ij} > u_{ij}}} (r_{ij}^2 - u_{ij}^2)^2 + \sum_{\substack{1 \leq i < j \leq n \\ r_{ij} < l_{ij}}} (r_{ij}^2 - l_{ij}^2)^2, \tag{5}$$

where $r_{ij} := \left[\sum_{k=1}^{3} (x_{ik} - x_{jk})^2\right]^{1/2}$.

The function F_{LU} — at least near its points of minima — is a smooth function and minimization methods described in [GMW81] can be applied.

The result are $n \times 3$ values x_{ik}; $i = 1, \ldots, n$; $k = 1, 2, 3$; representing the coordinates of the n atoms in the 3D space and describing a molecular model, which in the sequel will be called the Crippen model of the molecule and denoted by CM.

2.20. A program elaborated in the All Russian Research Institut of Organic Synthesis with an implementation of the Crippen algorithm was tested with representatives of different structure classes, mainly polycyclic compounds. Data were obtained from literature.

For some simple structures satisfactory results were found without optimization: it was just $F_{LU}(\mathbf{x}_1^{(0)}, \ldots, \mathbf{x}_n^{(0)}) \approx 10^{-5}$. This was the case with skeletons (models ignoring chemically bonded H-atoms) of molecules forming a tetrahedron, cube, 5-ring (cyclopentane), prism with triangle basis plane. If the hydrogene atoms are taken into account the number of atoms enlarges, of course, and the optimization step can't be neglected. Such molecules are represented in [ZE92].

2.21. The Crippen algorithm does not take into account the symmetry of the molecule at all. As usual, the resulting CM is not invariant with respect to the prescribed symmetry group type of the molecule, i.e. this model is not exact. The sources of unexactness of the Crippen algorithm are

- incomplete input data (most atom distances are chosen randomly respecting certain limitations),
- projecting the atomic coordinates from n dimensions into three, there is lost a reasonable part of information,
- occurence of rounded values via the implementation of linear algebra algorithms for the calculation of eigenvalues and eigenvectors.

Thus, it yields only an approximative model the less satisfying the larger the number of atoms.

3 The symmetrization algorithm

3.1. Input data. The following information about the molecule — available from physico-chemical experiments (spectroscopy) — serves as input data of the symmetrization algorithm:

$\boxed{1}$	Number n of atoms	$1 \quad \ldots \quad n$	
$\boxed{2}$	Kinds of atoms	$s_1 \quad \ldots \quad s_n$	
$\boxed{3}$	Bonds	$\begin{matrix} b_{11} & \ldots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \ldots & b_{nn} \end{matrix}$	$b_{ij} \in \{0,1\}$
$\boxed{4}$	Multiple bonds	$\begin{matrix} m_{11} & \ldots & m_{1n} \\ \vdots & & \vdots \\ m_{n1} & \ldots & m_{nn} \end{matrix}$	$m_{ij} = 0$, if $b_{ij} = 0$, $m_{ij} \in \{1,2,3\}$, if $b_{ij} = 1$
$\boxed{5}$	Bond lengths	$d_{ij} = \lvert \vec{A_i A_j} \rvert$, if $b_{ij} = 1$	

$\boxed{6}$	Valence angles	$w_{ijk} = \angle(A_i, A_j, A_k)$, if $b_{ij} = b_{jk} = 1$	
$\boxed{7}$	Type of symmetry group	$\mathfrak{G}$	(i.e. without orientation of symmetry elements in the space)

Wanted: A 3D model of the molecule satisfying all these data. The symmetrization algorithm brings together two models of the molecule and the symmetry group each reflecting only some of the points $\boxed{1}$ - $\boxed{7}$:

I. the molecular graph Γ — resulting from $\boxed{1}$, $\boxed{2}$, $\boxed{3}$ and $\boxed{4}$,

II. a model living in $\mathbb{R}^3$, obtained by Crippen's algorithm — using $\boxed{1}$, $\boxed{5}$, $\boxed{6}$,

III. the symmetry group type — point $\boxed{7}$.

3.2. Example $Al\,B_3H_{12}$ **continued.** The input data $\boxed{1}$, $\boxed{2}$, $\boxed{3}$ and $\boxed{4}$ can be obtained from the labelled molecular graph (2.2).
The bond lengths $\boxed{5}$ are:
$d_{1,2} = d_{1,3} = d_{1,4} = d_{1,5} = d_{1,6} = d_{1,7} = 1.801$Å,
$d_{2,8} = d_{3,8} = d_{4,9} = d_{5,9} = d_{6,10} = d_{7,10} = 1.283$Å,
$d_{8,11} = d_{8,12} = d_{9,13} = d_{9,14} = d_{10,15} = d_{10,16} = 1.196$Å.
The valence angles $\boxed{6}$ are:
$w_{2,1,3} = w_{4,1,5} = w_{6,1,7} = 73.4°$,
$w_{1,2,8} = w_{1,4,9} = w_{1,6,10} = 86.3°$,
$w_{2,8,11} = w_{2,8,12} = w_{3,8,11} = w_{3,8,12} = w_{4,9,13} = w_{4,9,14} =$
$w_{5,9,13} = w_{5,9,14} = w_{6,10,15} = w_{6,10,16} = w_{7,10,15} = w_{7,10,16} = 106.7°$
$w_{11,8,12} = w_{13,9,14} = w_{15,10,16} = 116.2°$.
The type of symmetry group $\boxed{7}$ is $\bar{6}2m = D_{3h}$ (group of order 12).

The symmetrization algorithm passes through the following steps:

3.3. Step 1: Determination of the graph automorphism group. The molecular graph Γ represents all atoms and all bonds of the investigated molecule. It is a graph, generally, with labelled vertices (vertex labels correspond to kinds of atoms) and labelled edges (edge labels represent multiplicities of bonds). Its automorphism group $\mathbf{Aut}(\Gamma)$ can be determined by hand or by computer using one of the known algorithms (e.g. [Hof82]).
In some cases it is easier to consider not the whole graph Γ but only its "skeleton" Γ' neglegting vertices with valency 1, i.e. the ordinarily chemically bonded hydrogene atoms (however such simplification will not allow to get a reasonable model for the whole molecule).

3.4. Step 2: Construction of the model CM using Crippen's algorithm. The input data are the available interatomic distances for the atoms of the molecule:

- $d_{ii} = 0$ for all $i = 1, \ldots, n$, where n is the number of atoms (3.1 $\boxed{1}$),
- d_{ij} for all bonded atoms A_i and A_j (3.1 $\boxed{5}$),
- d_{ik} for atoms A_i and A_k, i.e. belonging to a chain $A_i - A_j - A_k$; calculated from d_{ij}, d_{jk} and valence angle w_{ijk} (3.1 $\boxed{6}$).

Distances between all other, i.e. farther remote atoms are estimated (cf. 2.16). Then the Crippen algorithm works as described in 2C. The result are three dimensional coordinates $\mathbf{x}_i$, $i = 1, \ldots, n$, of the n atoms describing the Crippen model CM of the molecule. The molecule's symmetry is ignored and generally not satisfyed by the CM.

3.5. Step 3: Embedding of the symmetry group into the graph automorphism group. For the next steps we need to know how the generators of the symmetry group permute the atoms, i.e. we need the assignment Φ of automorphisms to the generators, cf. 2.13. This assignment yields a partition of the atom set into orbits w.r.t. the symmetry group and the determination of the site symmetries (stabilizers) of the atoms. In 2.13 we explained how to find the assignments satisfying conditions from the group theoretical point of view. But generally, this yields a huge number of possible assignments to be considered in the next steps. It can be decreased drastically taking into account some chemical and geometrical facts (cf. Example 2.14).
It is also possible to use the results from the Crippen algorithm for deciding which automorphism belongs to which symmetry group element, but this is dangerous because the Crippen algorithm can yield a spatial model deviating from the required symmetry.
For each of the remaining assignments — possibly more than one — we have to try to build up a space model satisfying the required symmetry and also the geometrical requirements passing steps 4.-7.
Once an automorphism is assigned to a generator of the symmetry group the orbits of this single automorphism are determined (to be more exact, the orbits of the subgroup generated by this automorphism): these are the cycles in the cycle representation of the automorphism. It is easy now to determine the decomposition of the set V of atoms into the orbits of the whole group G_F.

3.6. Step 4: Orient the Crippen model and the symmetry elements in the space. The atomic coordinates obtained by Crippen's method are used as initial approximation of the symmetrization algorithm. We want to place the symmetry elements in the space, so that the Crippen model is "as symmetric as

possible" w.r.t. the symmetry elements. A rotation axis of highest order will be called the main axis, all other rotation or rotoinversion axes are called side axes. Then we orient the model by an isometric transformation in such a way that the symmetry elements are in the following canonical positions:

a) the k-fold main axis coincides with the **z**-axis : k_{00z},

b) a mirror plane σ_h perpendicular to the main axis coincides with the plane $\{(x, y, 0)|x, y \in \mathbb{R}\}$: m_{xy0}
and one of mirror planes σ_v containing the main axis coincides with the plane $\{(0, y, z)|y, z \in \mathbb{R}\}$: m_{0yz},

c) a side axis C_2 coincides with the **y**-axis : 2_{0y0} (if this does not contradict b)),

d) one of the fixed points coincides with the origin O,

e) the side axes S_4 or S_6 of cubic groups are located at $\{(x\sqrt{2}, 0, x)|x \in \mathbb{R}\}$: $\bar{4}_{x\sqrt{2},0,x}$ or $3_{x\sqrt{2},0,x}$.

Knowing the orbits of the atoms w.r.t. G_F place the symmetry elements of the generators as follows:
If there is no main axis but an inversion center (i.e. in the case of point group $\bar{1} = C_i$) or a mirror plane ($m = C_s$), respectively, (then the orbits are actually single atoms or pairs of atoms) we approximate the center points of the orbits by a point or plane, respectively, using the least squares' method and transform it isometrically into canonical position.

If there is a main axis we proceed as follows:

- For the placement of the main axis, we calculate by the least squares' method the center points of the orbits assigned in step 3 to the main axis. Then we approximate again by the least squares' method these points by a line. In cases when at least two atoms are known to lie in the axis we can also take an approximation line for this points. An isometric transformation brings this axis in canonical position.

- If there is a unique fixed point of the symmetry group we approximate its coordinates by $(0, 0, z)$ where z is the arithmetic average of the third coordinates of all atoms of the molecule. Alternatively, instead of all atoms we can use those atoms which have the same third coordinate as the fixed point, i.e. the fixed points of $\bar{1} = i$, $m = \sigma_h$ etc. A shift parallel to the main axis brings this point into the origin.

- It remains to rotate the model around the main axis. To do this we choose a point not belonging to the main axis which is a point of

- the side axis $2 = C_2$ axis or
- the vertical mirror plane $m = \sigma_v$ plane or
- the side axis $\bar{3} = S_6$ or
- the side axis $3 = C_3$ or the side axis $\bar{4} = S_4$.

If there is more than one possibility we can use an arbitrary one. Only in the case of point group $\bar{4}3m = T_d$ we have to use a fixed point of a $3 = C_3$ axis, if there is one. Finally, we rotate the model around the main axis so that the choosen point lies in the plane $\{(0, y, z) | y, z \in \mathbb{R}\}$ (in the first case) or in the plane $\{(x, 0, z) | x, z \in \mathbb{R}\}$. (in the other three cases).

3.7. Example $Al\,B_3H_{12}$ **continued.** We orient the CM as follows.
Placing the main axis: The orbits corresponding to the main axis $3 = C_3$ are: $\{1\}$, $\{2, 4, 6\}$, $\{3, 5, 7\}$, $\{8, 9, 10\}$, $\{11, 13, 15\}$, $\{12, 14, 16\}$.
Placing the unique fixed point: The shortest orbit consists only of the Al-atom, hence we take its third coordinate from CM for z.
Rotate around the main axis: The atom 10 is a fixed point of $m = \sigma_v$, hence we finally rotate the model so that this atom lies in the in the **yz**-plane.

3.8. Step 5: Symmetrization of the model. The model obtained up to now (reoriented CM), in general does not satisfy the required symmetry. In this step we construct a new model such that its symmetry group coincides with the prescribed symmetry. For this, we select a suitable subset $A \subseteq V$ representing all orbits of atoms w.r.t. G_F, project each $a \in A$ into the symmetry element(s) (determined in step 3) corresponding to its site symmetry (stabilizer) $\mathfrak{G}_a \leq \mathfrak{G}_F$. Applying $\mathfrak{G}_F$ to the elements of A we obtain new atomic coordinates for all atoms taking into account $\mathfrak{G}_F$.
From each of the orbits we choose one representative as follows:
Choose an arbitrary atom as representative of a shortest orbit. For the remaining orbits we proceed by the following rule:
Let R be the set of all atoms of the orbits where no representative is chosen yet. Put R' the set of atoms from R which are bonded with the maximal number of already choosen representatives. In R' we choose an element with orbit of minimal length (if there is more than one atom with this property then we choose an arbitrary one among these). Let A be the set of the chosen representatives.

For $a \in A$ let Q_a be the set of common fixed points of all symmetry operations W for which the corresponding permutation $w = \Phi(\mathsf{W})$ of the atoms leaves a unchanged:

$$Q_a := \{\mathbf{x} \in \mathbb{R}^3 \mid \mathsf{W}(\mathbf{x}) = \mathbf{x} \text{ for all } \mathsf{W} \text{ with } a^w = a\}.$$

We obtain the coordinates of a in the symmetrized model by projecting the CM coordinates of a into the set Q_a. The coordinates of the non-representatives are

obtained simply by applying all symmetry operations to the new coordinates of the representatives. Note, that because of this projection step, we will obtain for each orbit as many points in $\mathbb{R}^3$ as the length of the orbit.

3.9. Example $Al\,B_3H_{12}$ **continued.**

orbit	representative	... is projected into the symmetry elements
$\{1\}$	1	all
$\{2,3,4,5,6,7\}$	2	$0, y, z$
$\{8,9,10\}$	8	$m_{x,y,0}$, $m_{0,y,z}$
$\{11,12,13,14,15,16\}$	11	$m_{x,y,0}$

All non representative atoms get new coordinates by applying the symmetry operations to the representatives. For instance, we get the coordinates of atom 7 by applying 3^{-1}_{00z} and then $m_{x,y,0}$ to the coordinates of atom 2.

3.10. Step 6: Optimization of the symmetrized model. The coordinates of the representatives determine the coordinates of all atoms of the molecule. Hence, it is required to optimize the coordinates of the representatives in such a way that the complete model obtained by applying the symmetry operations to the representatives satisfies the geometrical parameters. Thus, (if there is any proper symmetry) we get an optimization problem of lower dimension. We shall optimize not only the coordinates of the representatives but also of some other atoms and auxiliary points which are collected in a set E, see below.
For a point $p \in E$ the vector of its coordinates in the 3D space is denoted by $\mathbf{x}_p = (x_{p1}, x_{p2}, x_{p3}) \in \mathbb{R}^3$. Let the set E consist of the m points $p_1, p_2, \ldots, p_m$. In order to simplify the notation we construct a tuple $\mathbf{x}$ of $3m$ components from the coordinates of the points of E by

$$\mathbf{x} := (x_{p_1 1}, x_{p_1 2}, x_{p_1 3}, x_{p_2 1}, x_{p_2 2}, x_{p_2 3}, \ldots, x_{p_m 1}, x_{p_m 2}, x_{p_m 3}).$$

The optimization problem is formulated as follows:

$$F^*(\mathbf{x}) \longrightarrow \min \qquad (\mathbf{x} \in \mathbb{R}^{3m},\ bl_i \le x_i \le bu_i\ (i = 1, \ldots, 3m)),$$

where $F^*(\mathbf{x}) := F_{LU^*}(\mathbf{x}) + \sum \gamma_k(\mathbf{x})$,

$$F_{LU^*}(\mathbf{x}) := \sum_{\substack{1 \le i < j \le m \\ r_{ij} > u_{ij}}} (r_{ij}^2 - u_{ij}^2)^2 + \sum_{\substack{1 \le i < j \le m \\ r_{ij} < l_{ij}}} (r_{ij}^2 - l_{ij}^2)^2,$$

where $r_{ij} := |\mathbf{x}_{p_i} - \mathbf{x}_{p_j}|$ and the sum $\sum \gamma_k(\mathbf{x})$ runs over all terms $\gamma(\mathbf{x})$ defined troughout this subsection 3.10. The bounds bu_i and bl_i and the numbers u_{ij} and l_{ij} are described below. Let $\mathbf{bl}$ and $\mathbf{bu}$, respectively, be the tuples $(bl_1, \ldots, bl_{3m})$ and $(bu_1, \ldots, bu_{3m})$, respectively, and let LU^* be the matrix where the entry (ij) is u_{ij} and the entry (ji) is l_{ij} $(1 \le i < j \le m)$. The function F_{LU^*} is quite

similar to F_{LU} used in the Crippen algorithm (2.19, formula (5)), but the set of points, whose coordinates form $\mathbf{x}$ and hence the values l_{ij} and u_{ij} are different. Despite of the additional terms γ_k and the constraints **bl** and **bu** the minimization problem has similar properties and can be solved by analogous procedures as the optimization problem in the Crippen algorithm.
In the sequel we describe the construction of the set E, the matrix LU^*, the functions γ_k, the bounds **bl** and **bu** and the initial value for $\mathbf{x}$.

The set E:

We put $\tilde{A} := A \cup A'$ and $E := \tilde{A} \cup Q$ (A, A', Q are defined below).

- A consists of all representatives chosen in the previous step 5.
- The set Q consists of *auxiliary points* assigned to symmetry operations:

  $$Q := \{q_{a,w} \mid a \in A,\ w \in G_F,\ \text{the distance } d_{a\,w(a)} \text{ is exactly determined}^8\},$$

  where $q_{a,w}$ denotes the center point of the orbit $a^{\langle w \rangle}$ in the molecule. Two auxiliary points which a priori coincide can be identified.
- We want to find a set A' of atoms such that:
  if the space model has the required symmetry and all metrical requirements are satisfied on $\tilde{A} = A \cup Q \cup A'$ then all metrical requirements will be satisfied "automatically" for all atoms via the action of the symmetry group $\mathfrak{S}_F$. A first idea would be to choose A' so that in each 2-orbit of bonds there is a bond between two atoms from $\tilde{A}$ and in each 3-orbit of valence angles there is a valence angle formed by three atoms from $\tilde{A}$. But using auxiliary points we can form a smaller set A':

  The set of all pairs of bonded atoms is partitioned into 2-orbits under the action of the group G_F (i.e. the permutations of atoms assigned to the symmetry operations). Now we take each 2-orbit, one after the other, and do the following: Let R be the choosen 2-orbit. All pairs $(a, a') \in R$ which satisfy

  - there is an auxiliary point $q = q_{a,w}$ with $a' = w(a)$ and
  - the distance d_{qa} can be computed from the exactly determined d_{ij} (e.g. via cosine theorem from $d_{aa'}$).

  we put into a set T of "guaranteed distances". If there is no pair in R such that both its atoms already belong to $\tilde{A}$ and if there is no pair in $R \cap T$ then choose one pair of R such that one atom of it belongs to A and put the other atom into A'.

[8] i.e. can be calculated from 3.1 [5], [6]

Analogously, we partition the set of all triples of atoms forming a valence angle into 3-orbits under G_F. For each 3-orbit, taken one after the other, we proceed as follows: Let R be the choosen 3-orbit. For all triples $(b, a, a') \in R$ which satisfy

- atom $b \in \tilde{A}$
- there is an auxiliary point $q = q_{a,w}$ with $a' = w(a)$ or $a' = w(b)$ and
- the distances d_{qa} and d_{qb} can be computed from the exactly determined d_{ij}.

put the pair (b, a') into the set T. If we do not find a triple (a, b, c) in R such that the pair (a, c) is already completely contained in $\tilde{A}$ or belongs to T, then we will find one satisfying this property after adding one of its atoms to A'.

The functions F_{LU^*} and γ_k and the bounds bl and bu

Let $w \in G_F$ be the permutation of atoms assigned to a motion $\mathsf{W} \in \mathfrak{B}_F$ (cf. 2.13) and let $fix(\mathsf{W})$ be the set of fixed points of W. Note that if W is a k-fold rotoinversion, $k \neq 2$, then $fix(\mathsf{W})$ is a single point. For any other symmetry operation $fix(\mathsf{W})$ coincides with the symmetry element of W (cf. 2.12).
We set up the functions F_{LU^*} and γ_k and the bounds **bl** and **bu** in such a way that $F^*(\mathbf{x})$ is zero and $bl_i \leq x_i \leq bu_i$ $(i = 1, \ldots, 3m)$ if the allocation of the points of E given by the value of $\mathbf{x}$ satifies:

1. For all atoms $a \in \tilde{A}$ and all $w \in G_F$: if $w(a) = a$ then $\mathsf{W}(\mathbf{x}_a) = \mathbf{x}_a$ (i.e. $\mathbf{x}_a \in fix(\mathsf{W})$).

2. For all auxiliary points $q = q_{a,w}$: Let $\mathbf{c}$ be the center of the points $\mathbf{x}_a$, $\mathsf{W}(\mathbf{x}_a)$, $\mathsf{W}^2(\mathbf{x}_a), \ldots, \mathsf{W}^k(\mathbf{x}_a)$, then $\mathbf{x_q} = \mathbf{c}$. Note, that this implies $\mathsf{W}(\mathbf{x}_q) = \mathbf{x}_q$.

3. The atoms from $\tilde{A}$ placed by $\mathbf{x}$ satify the given geometrical parameters, i.e. the distance $|\mathbf{x}_a - \mathbf{x}_b|$ for bonded atoms $a, b \in \tilde{A}$ equals the given bond lenght d_{ab} and the angle $\angle(\mathbf{x}_a - \mathbf{x}_b, \mathbf{x}_c - \mathbf{x}_b)$ for atoms $a, b, c \in \tilde{A}$ forming a valence angle equals the given valence angle w_{abc}.

4. For all atoms $a' \in A'$ and $a \in A$: If there is a motion W with $a' = w(a)$ (i.e. $a' \in a^{G_F}$) then $\mathbf{x}_{a'} = \mathsf{W}(\mathbf{x}_a)$.

Requirement 1: Let $a \in \tilde{A}$ be a fixed point of $w \in G_F$.
If $fix(\mathsf{W})$ belongs to a coordinate plane (or to a coordinate axis or is the origin, respectively) from $\mathbf{x}_a \in fix(\mathsf{W})$ it follows that one (or two or three, respectively) coordinates of $\mathbf{x}_a$ are equal to zero. We put the respective components of **bl** and **bu** equal to zero.
If this is not sufficient for $\mathbf{x}_a \in fix(\mathsf{W})$, the following term γ is constructed:

In the case of an axis $fix(\mathsf{W})$: $\gamma = (\cos^2 \angle(\mathbf{b}, \mathbf{x}_a) - 1)^2$, where $\mathbf{b}$ is the direction vector of $fix(\mathsf{W})$ and in the case of a plane $fix(\mathsf{W})$: $\gamma = (\cos^2 \angle(\mathbf{b}, \mathbf{x}_a))^2$, where $\mathbf{b}$ is a vector perpendicular to $fix(\mathsf{W})$.

Requirement 2: Let $q = q_{a,w}$ and let $a' = w(a)$.
In the same way as for the first requirement we guarantee $\mathbf{x}_q \in fix(\mathsf{W})$.
In order to complete the second requirement we construct the following term γ:
In the case of an axis $fix(\mathsf{W})$: $\gamma = (\cos^2(b, \mathbf{x}_a - \mathbf{x}_q))^2$, where $\mathbf{b}$ is the direction vector of $fix(\mathsf{W})$ and in the case of a plane $fix(\mathsf{W})$: $\gamma = (\cos^2(b, \mathbf{x}_a - \mathbf{x}_q) - 1)^2$, where $\mathbf{b}$ is a vector perpendicular to $fix(\mathsf{W})$.
If $a' \in A'$ the analogous terms γ for a' instead of a are constructed.

Requirement 3: We put LU^* as follows:
The bounds $u_{aa'}$ and $l_{aa'}$ for distances of atoms $a, a' \in \tilde{A}$ in the matrix LU^* are taken from the matrix LU of the Crippen algorithm. Some distances d_{qa} ($a \in \tilde{A}$, $q \in Q$) involving auxiliary points are calculated during the construction of the set A', in these cases one sets $u_{qa} = l_{qa} = d_{qa}$. For the remaining distances the lower bounds are put to zero and the upper bounds sufficient large ($10\sqrt[3]{|E|}$).

Requirement 4: Let $a' = w(a)$. Put a term $\gamma = |\mathbf{x}_{a'} - \mathsf{W}(\mathbf{x}_a)|^2$.

The initial vector for x.

The initial coordinates for the atoms are taken from the symmetrized model obtained in step 5. For an auxiliary point q_{aw} the initial value of x_{qj} is the arithmetic average of x_{aj}, $x_{w(a)j}$, $x_{w^2(a)j}, \ldots$ ($j = 1, \ldots, 3$) and the coordinates x_{aj}, $x_{w(a)j}$, $x_{w^2(a)j}, \ldots$ are taken from the symmetrized model obtained in step 5.

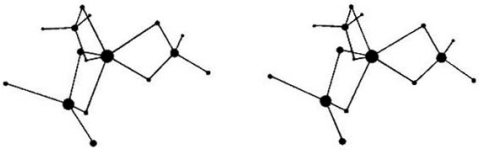
The coordinates resulting from the optimization step define the final 3D Cartesian coordinates of the symmetrized model.

3.11. Example AlB_3H_{12} continued.

The set Q:
The permutations corresponding to $m = \sigma_h$ and $2 = C_2$ both map $2 \mapsto 3$ and $11 \mapsto 12$ and the distances d_{23} and $d_{11\,12}$ are known. Hence, the points $q_1 = q_{2\,\sigma_h}$, $q_2 = q_{11\,\sigma_h}$, $q_3 = q_{2\,C_2}$, and $q_4 = q_{11\,C_2}$ are candidates for Q. From requirement 2 it follows $\mathbf{x}_{q_1} = \mathbf{x}_{q_3}$ and $\mathbf{x}_{q_2} = x_{q_4}$. Hence, $Q = \{q_1, q_2\}$.
The set A':
The set of bonds is decomposed into the 2-orbits:
$\{(1,2), (1,3), (1,4), (1,5), (1,6)(1,7)\}$,
$\{(2,8), (3,8), (4,9), (5,9), (6,10), (7,10)\}$ and
$\{(8,11), (8,12), (9,13), (9,14), (10,15), (10,16)\}$. In each of these orbits, there is a bond, where both atoms already are in A.
The set of valence angles is decomposed into four 3-orbits. In two of them we find a triple where all three atoms are in A ($\angle(1,2,8)$ and $\angle(2,8,11)$).
Knowing the bond length $d_{23} = 2d_{2q_1}$ and $d_{11\,12} = 2d_{11\,q_2}$ and the angles $\angle(2,1,3)$ and $\angle(11,8,12)$ we do not need to add atoms to A' during processing the two

remaining 3-orbits.
So, finally, the set A' is empty.

The symmetrized molecular model of $\mathrm{Al\,B_3H_{12}}$ is shown in the following two pictures. Squint and you obtain (with a little bit of training) a stereo picture:



3.12. Concluding remarks. The symmetrization algorithm allows the construction of spatial models of molecules with given near interatomic distances and assumed symmetry. The main advantages are:

- All information about the symmetry of the molecule is taken into account without expensive interactive participation of the user;
- In most cases, the result is more satisfactory than results obtained by methods which do not consider symmetry (e.g. the methods mentioned in the introduction);
- The optimization steps are the most extensive procedures in both, the Crippen algorithm and the symmetrization algoritm. In the symmetrization algorithm, the optimization is done with much less variables because of the choice of representatives w.r.t. symmetry. So, in the example of $\mathrm{Al\,B_3H_{12}}$, only 18 variables (6 points in R^3) instead of 48 coordinates of all atoms participate. One possible modification of the symmetrization algorithm is to omit the extensive optimization step in Crippen algorithm.

The central algebraic point of the theoretical background of the symmetrization algorithm is the relationship between the spatial symmetry of a molecule and the symmetry of its molecular graph. In this paper this topic is focused on the problem of embedding the group $\mathfrak{G}_F$ into the group $\mathbf{Aut}\,\Gamma$. Here, we gave some contributions to this problem. This problem seems to be especially worth for further mathematical investigation, probably involving also aspects of quantum mechanics and quantum chemistry.

3.13. Acknowledgements. L. Zlatina likes to express her deep gratitude to Prof. M. Elyashberg from the All-Russian Research Institute of Organic Synthesis (Moscow) where the PhD was performed for having given her the possibility of dealing with this very interesting problem and for a lot of useful advices during several years.

We are obligued to Prof. A. Kerber (Bayreuth) for his interest in this work and his personal support of it and to Prof. M. Klin (Beer Sheva) having initialized and supported this joint work and firstly called our attention to some theoretical problems in this field.
The two German authors want to dedicate their special thanks to Prof. R. Pöschel (Dresden) for many detailed, helpful and fruitful discussions.

References

[AGH68] L. ALMENNINGEN, G. GUNDERSET, A. HAALAND: On the Molecular Structure of Aluminium Borohydride $Al(BH_4)_3$. *Acta Chem. Scand.* 1968, **22**, No. 1, pp. 328-334.

[Ant95] H. ANTON: Lineare Algebra. Heidelberg 1985, pp. 510-513.

[BS83] J. BAWER, W. SCHUBERT: Computer Modelling of Molecular Structure. *Comp. and Chem.* 1983, **7**, No. 2, pp. 61-65.

[Cri77] G.M. CRIPPEN: A Novel Approach of Calculation of Conformation: Distance Geometry. *J. Comput. Phys.* 1977, **24**, No. 1, pp. 96-107.

[Cri78] G.M. CRIPPEN: Rapid Calculation of Coordinates from Distance Matrices. *J. Comput. Phys.* 1978, **26**, No. 3, pp. 449-452.

[Cri81] G.M. CRIPPEN: Distance Geometry and Conformational Calculations. New York 1981.

[Dav86] K. DAVIES: Computer Aided Molecular Modelling. *Manuf. Chem.* 1986, **57**, No. 3. pp. 30-31.

[Ess83] H. ESSEN: On the General Transformation from Molecular Geometric Parameters to Cartesian Coordinates. *J. Computational Chemistry* 1983, **4**, No. 1, pp. 136-140.

[EKDG69] M.R. ELYASHBERG, YU.Z. KARASEV, V.A. DEMENT'EV, L.A. GRIBOV: Interpreted Spectra of Hydrocarbons — Derivatives of Cyclohexane and Cyclopropane. Nauka 1969 (in Russian).

[ESMZKKY91] M.E. ELYASHBERG, V.V. SEROV, E.R. MATROSYAN, L.A. ZLATINA, YU.Z. KARASEV, V.N. KOLDASHOV, YU.YU. YAMPOLINSKIY: An Expert System for Molecular Structure Elucidation Based on Spectral Data. *J. Mol. Struct. (Theochem.)* 1991, **230**, pp. 191-203.

[FF63] D.K. FADDEEV, V.N. FADDEEVA: Computational Methods of Linear Algebra. San Francisco, Freeman & Co. 1963.

[Flu80] R.L. FLURRY: Symmetry Groups: Theory and Applications to Chemistry. Englewood Cliffs, NJ., Prentice Hall 1980.

[GP72] E. GAVUZZO, S. PAGLUICA: Generation and best Fitting of Molecular Models. *Acta Cryst. B* 1972, **28**, pp. 1968-1969.

[GMW81] PH. GILL, W. MURRAY, M. WRIGHT: Practical Optimization. London, Academic Press, 1981.

[GSh73] N. GO, H.A. SHERAGA: Ring Closure in Molecules with C, I, S Symmetry. *Macromolecules* 1973, **6**, No. 2, pp. 273-281.

[Gri83] L.A. GRIBOV, M.E. ELYASHBERG, V.N. KOLDASHOV, V.V. PLETNEV: A Dialogue Computer Program System for Structural Recognition of Complex Molecules by Spectroscopic Methods. *Anal. Chim.* Acta 1983, **148**, pp. 159-170.

[Hah83] TH. HAHN (ed.): International Tables for Crystallography, **A**. Dordrecht, Boston, D. Reidel Publ. Comp. 1983.

[Hil69] R.L. HILDEBRANDT: Cartesian Coordinates of Molecular Models. *J. Chem. Phys.* 1969, **51**, No. 5, pp. 1654-1659.

[Hof82] CHR.M. HOFFMANN: Group-Theoretic Algorithms and Graph Isomorphism. Springer, Berlin Heidelberg, 1982, p. 71.

[Kle89] G. KLETCH: Anwendung Kartesischer symmetrisierter Koordinaten zur Berechnung der Schwingungsspektren von Molekülen und von Festkörpern. *Z. Krist.* 1989, **186**, No. 1-4, pp. 163-164.

[KPR88] M. KLIN, R. PÖSCHEL, K. ROSENBAUM: Angewandte Algebra. VEB Deutscher Verlag d. Wiss.,Berlin, 1988.

[KRRT95] M. KLIN, CH. RÜCKER, G. RÜCKER, G. TINHOFER: Algebraic Combinatorics in Mathematical Chemistry I. Technische Universität München, 1995.

[NBB85] J.E. NORLANDER, A.F. BOND, M. BADER: *Computers and Chemistry* 1985, **9**, pp. 209-235.

[Sch95] M. SCHÖNERT et.al.: GAP — Groups, Algorithms and Programming. Lehrstuhl D für Mathematik, Rheinisch Westfälische Technische Hochschule, Aachen, Germany, 5-th ed., 1995.

[VN78] A.V. VANDURA, N.P. NOVOSELOV: Vychislenie dekartovykh koordinat atomov slozhnykh molekul i molekulyarnykh kompleksov. *Zh. S. Ch.* 1978, **19**, No. 2, pp. 357-358 (Russian).

[Voe74] V.V. VOEVODIN: Linejnaya algebra. izd. "Nauka", Moscow 1974 pp. 211-215 (Russian).

[WS82] J.C. WENGER, D.H. SMITH: Deriving Three Dimensional Representational of Molecular Structure from Connective Tables Augmented with Configuration Designations Using Distance Geometry. *J. Chem. Inf. Comput. Sci.* 1982, **22**, No. 1, pp. 29-34.

[Wie64] H. WIELANDT: Finite Permutation Groups. Academic Press, New York — London, 1964.

[Wey52] H. WEYL: Symmetry. Princeton, University Press, 1952.

[WH86] W.T. WIPKE, M.A. HAHN: Artificial Intelligence in Chemistry. *Symposium Series 306, American Chemical Society* 1986, pp. 136-146.

[ZM87] A.A. ZENKIN, V.I. MOTSHALOV: Universal'nyj algoritm rascheta dekartovykh koordinat atomov molekulyarnykh modelej proizvol'noj stepeni slozhnosti. *Vestnik Moskovskogo Universiteta* 1987, **28**, No. 1, pp. 88-92 (Russian).

[Zla91] L. ZLATINA: Matematicheskie metody generatsii stereoisomerov i postroeniya prostranstvennykh modelej molekul. *PhD thesis, Vsesoyuznyj nauchno-issledovatel'skij institut organicheskogo sinteza*, Moscow 1991 (Russian).

[ZE92] L. ZLATINA, M. ELYASHBERG: Generation of stereoisomers and their spatial models corresponding to the given molecular structure. *MATCH* 1992, **27**, pp. 191-207.