Als Manuskript gedruckt

Technische Universität Dresden
Herausgeber: Der Rektor

# A Mathematical Formalization of Graph Set Analysis in Crystallography*

R. Pöschel (cf. page 1)

MATH–AL–10–2000 August 2000

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

The following manuscript has a long history. It is based on the joint work of J. Bernstein (Beer-Sheva), B. Ganter (Dresden), J. Grell (Berlin), D. Kuske (Dresden) and R. Pöschel (Dresden), who all made contributions and should be considered as coauthors.

It turned out that the present form is still not suitable for a final publication. Moreover, one of the collaborators mentioned above does not support anymore the methodology of the paper as it is developed here. This dissension led me to the conclusion to collect what was done up to July 1999 and to present this material as technical report. It may be used as background for further research.

The criticism which I got during the preparation of the paper mainly concerned the mathematical part. With the presentation as technical report I take the full responsibility for its content. Benefits which might arise from this report however should be assigned to every member of the above group.

Dresden, August 2000 Reinhard Pöschel

# A Mathematical Formalization of Graph Set Analysis in Crystallography

**Abstract**

Graph Set Analysis investigates "graph sets", i.e. special hydrogen bond patterns, in crystals and other arrays of molecules. In this paper, rigorous mathematical concepts and tools for Graph Set Analysis are presented. One crucial notion is that of a constructor graph, from which the graph sets and their descriptors can be determined.

## Introduction

The everyday language of chemistry contains many concepts and phrases taken from mathematics. The chemist's shorthand rendering of the structure of a molecule, usually referred to as the "structural formula", is often simply a graph in the mathematical sense, where atoms are represented by vertices and chemical bonds are represented by edges.

The utility of the structural formula lies in the fact that most organic and organometallic substances may be prepared and isolated as a collection of molecules all of which may be represented by the same structural formula. This language is useful because the bonding capacity of atoms is well characterized, and predictable. Such is not the case for arrays of molecules, which are best exemplified by crystals. In terms of energy, geometry and topology, interactions between the molecules that make up the crystal are much less well understood. Perhaps the dominant interaction between molecules in organic crystals is the *hydrogen bond*. The hydrogen bond is characterized by a fairly high energy (for interactions between molecules) and a fairly specific directionality. Hence the attempts to understand the patterns formed by hydrogen bonds. In pursuing that understanding, it is natural to employ the graphical language that is already common parlance among chemists. The hydrogen bond will constitute the edge in this language. Now, however, instead of considering atoms as vertices, the chemists will want to consider the entities that are connected by hydrogen bonds. In most cases these will be whole molecules or parts of single molecules.

Such notions were the basis for the development of the application of "Graph Set Analysis" to hydrogen-bonded arrays of molecules – in most cases crystals, but not necessarily restricted to crystals – by the late M.C. Etter and her colleagues ([10]) .

This approach turned out to be an extremely useful tool to describe, to compare and to classify structures containing hydrogen bonds. Among the many applications, its utility in comparing polymorphs – multiple crystal structures of a particular chemical entity – is particularly noteworthy ([3]). The approach was developed by chemists, mainly employing the language of chemistry but again adopting some of the lexicon of mathematics ([2]).

The initial empirical notions have been tested and have matured to the point where they now require firmer mathematical foundations to enable further development and implementation. Over the past few years we have been working together as mathematicians and chemists, to develop the common language necessary to provide that mathematical foundation.

This paper is a product of that collaboration; it outlines some mathematical concepts and tools that have been elaborated for the further development and application of graph theory to hydrogen bond patterns (first attempts can be found in the preliminary version [4]). What is presented here is a sound beginning, which is already serving as the basis for further developments (see [19] and [12, 13]).

# 1 Outline

In the following we give a short outline of the main ideas and developments in this paper. After this we give a short overview about the sections.

In Fig. 1 there are shown several levels of abstraction we have to traverse in developing and applying the graph set notions of hydrogen bonding patterns (the shaded parts belong to the chemical crystallographic background, which is not considered in this paper).

The starting point is a real chemical structure (crystal) determined by chemists (crystallographers). The result of a crystal structure analysis is a set of atomic coordinates and symmetry properties from which all the required information about the crystal can be derived (in principle). From a heuristic point of view chemists model such crystals as an array. This is a graph whose vertices (representing the atoms) are points (given by coordinates) in the 3D-space and whose edges represent bonds.

Once the symmetry group $G$ (group of isometries) is known, only a very small part (the crystallographic asymmetric unit) of the array need be presented. Every other vertex $v$ or edge $e$ is crystallographically equivalent to a given one, i.e., there is a symmetry operation $g \in G$ moving a vertex from the given small part to $v$ (analogously for edges).

Now we inquire as to what can and should be done from the mathematical point of view. First, mathematicians have to create a definition of an array which reflects all the necessary features of a crystallographer's array for the description of the hydrogen bonding and comparison with other structures, but suppresses all unnecessary information. From the point of view of Graph Set Analysis, bonds and their patterns are essential but the concrete position in 3D-space can be neglected. This gives the definition of an array as presented in 3.1.

However, crystallographic equivalence is always inherent in all chemical considerations (but mostly hidden in the background). Thus a rigorous mathematical treatment must make this explicit. Therefore we introduce $G$-arrays (arrays on which the symmetry

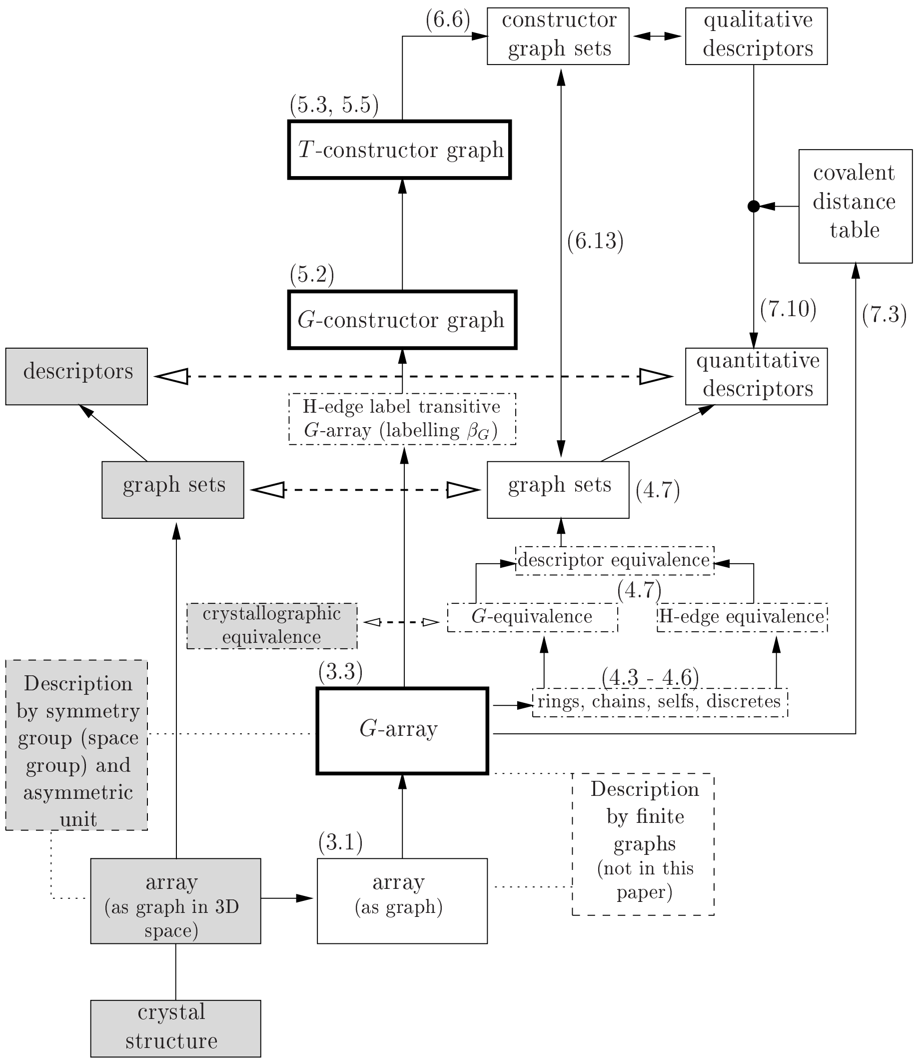



Figure 1: Notions in the development of graph sets to describe and compare hydrogen bonding patterns

group $G$ acts). Crystallographic equivalence is now a concept which can be expressed naturally in terms of group orbits: two objects, say $u$ and $v$, are ***crystallographically equivalent*** if and only if they belong to the same $G$-orbit, i.e., there exists a $g \in G$ which maps $u$ to $v$. At the same time crystallographic equivalence, i.e. group orbits, provides a labelling $\beta_G$ of the hydrogen edges: two hydrogen edges get the same label if and only if they are crystallographically equivalent. We shall therefore canonically label edges $e$ by their $G$-orbits (i.e., $\beta_G(e) := e^G$).

$G$-arrays are infinite graphs which can also be represented by a finite part taking into account the action of $G$. In this paper, however, we do not treat the subject of how to present infinite graphs by finite data (it will be shown in [5] how this can be done in an economical way independent of the group $G$).

$G$-arrays are an adequate mathematical description of what has been done and used by chemists so far. We now consider Graph Set Analysis, which deals with the detection and description of hydrogen bond patterns (rings, chains, selfs, discretes representing the so-called graph sets). That can be described by special paths in the $G$-array. For this analysis the internal structure of the molecules (e.g. atom connectivity) plays a minor role. From this point of view it is a disadvantage that the use of the $G$-array retains all the internal information about the molecules.

The next level of abstraction overcomes this by ignoring the internal molecular structure. Shrinking molecules to a single vertex but keeping the hydrogen edges as edges, a new graph – called the constructor graph – is obtained (to some extent this idea was already used for inorganic compounds in [20, pp. 294-315], and further developed in [15, pp. 19-21], [17] and [10]). Some information about the molecules is nevertheless needed in the analytical process and will therefore be stored in the so-called covalent distance table. Again, the action of the symmetry group $G$ is an essential feature, so more precisely we have a $G$-constructor graph.

However, in the $G$-constructor graph we are faced with an unpleasant phenomenon: the introduced labelling $\beta_G$ is no longer sufficient to locate (up to $G$-equivalence) paths of hydrogen edges by their labels only; mathematically we have a multigraph with possibly several equally labelled edges between two vertices. This property however will be needed for the description and computation of graph sets. Therefore we refine the labelling $\beta_G$ by considering a suitable subgroup $T$ of $G$ (e.g. the subgroup of translations). The constructor graph with the resulting refined labelling $\beta_T$ is called the $T$-constructor graph.

The labelling $\beta_T$ allows one to characterize graph sets by sequences of $\beta_T$-labels, which are the basis of what we shall call qualitative descriptors.

The situation now is as follows: with the $T$-constructor graph and the covalent distance table we have a simplified (and more abstract) model of the original $G$-array which nevertheless contains all the information needed for Graph Set Analysis. As in the $G$-array model we want to locate and describe graph sets and to compute their descriptors. In fact, this objective can be reached without relying on a lower level of abstraction (in particular without use of the original $G$-array).

This approach has conceptual advantages: we (or a computer algorithm) need to work only with those data which are inherent in the model under consideration, and the method will work for all graphs which fit into the definition of a $G$-array or $T$-constructor graph, respectively. In particular this opens the way for other applications not restricted to the analysis of simple hydrogen bond patterns (cf. 2.9).

However, there is a price to pay for these advantages. We must define all notions for $G$-arrays as well as for constructor graphs independently of each other. Further, because one can canonically assign to each $G$-array a $T$-constructor graph, we must ensure that in the latter case we do not get fewer or more graph sets than in the former (First Justification Statement 6.13), and that the computation of (quantitative) descriptors gives the same result in both models (Second Justification Statement 7.10). This is a rather formal procedure and nearly trivial from the mathematical point of view, but it should be done at least once for correctness. Therefore the reader who is not interested in mathematical details may simply skip those parts of this paper (some of the longer proofs have been relegated to the appendix).

The examples in Section 2 and in 7.11 may provide the reader some practical aquaintance with the overall approach. Moreover they may serve as a guide and point of reference throughout the paper which is organized as follows:

*Section 2* contains a statement of the basic principles of graph set analysis of hydrogen bonds and a worked example (anthranilic acid) that is meant to be comprehensible to every interested reader.

Subsequent sections are written in a more mathematical style. The aim of these sections is to formulate basic principles and results and to establish mathematical notions and notations for further research.

In *Section 3* we introduce *arrays* and *$G$-arrays.* A special labelling is introduced to distinguish between covalent bonds and hydrogen bonds (H-edges).

In *Section 4* we start to define mathematically those notions which are used by chemists in Graph Set Analysis: *rings, chains, selfs, discretes.* These are certain subgraphs with special H-edge patterns. They are representatives of so-called *graph sets* and can be classified by (quantitative) *descriptors.* In this section it is also shown how labellings of $G$-arrays can be refined by using a special subgroup $T$ of $G$.

In *Section 5* we introduce *constructor graphs*, which are crucial for the mathematical treatment. In $(T, G)$-constructor graphs one can describe paths by directed label sequences (up to $G$-equivalence).

In *Section 6*, rings, chains, selfs and discretes as well as graph sets in $T$-constructor graphs and their *qualitative descriptors* are introduced. Two descriptors may represent the same constructor graph set.It is clarified under which conditions this happens. The connection between the constructor graph sets and the graph sets of the underlying $G$-array is investigated and justified. Moreover we clarify which types of graphs sets can appear simultaneously in a $G$-array.

In *Section 7*, it is shown how quantitative descriptors can be computed from the qualitative ones (the latter contain much more information about the structure of graph sets than the former). Since some quantitative information is lost by switching from $G$-arrays to constructor graphs (namely the inner structure of the molecules), every constructor graph will be accompanied by a *covalent distance table*. A justification statement assures the correctness of the procedure. An example demonstrates the computational steps which lead from a given $G$-constructor graph via the $T$-constructor graph to the computation of quantitative descriptors of the underlying $G$-array.

Finally, in the Mathematical Appendix (*Section 8*) we fix the group- and graph-theoretical notions and notations used in the paper. Morover, some particular lemmata and lengthy proofs are given at the end of the appendix.

# 2 An Example

We start with an example which elucidates some of the most important features of the theory and the method which will be developed below in the language of mathematics. We hope that mathematicians will get an idea of the chemical and crystallographic origin of the mathematical notions, and that those chemists who find the mathematical language not particularly intuitive will be able to infer the main ideas from this example (for further examples we refer to [12, 13]). Those notions which appear later in a precise mathematical framework are marked by emphasis, e.g. *graph*.

The features of the method are readily demonstrated with the example of **anthranilic acid** (form II) which has been treated elsewhere in detail (e.g. [6], [2]).

Figure 2 is a typical representation of a crystal structure as presented by crystallographers (it was prepared with the drawing program PLUTO which is part of the software package of the Cambridge Structural Database, see e.g. [7]). Other representations of the same crystal structure can be found in [6] and [2].

A schematic picture, particularly suitable for subsequent discussion here, is given in Figure 3.

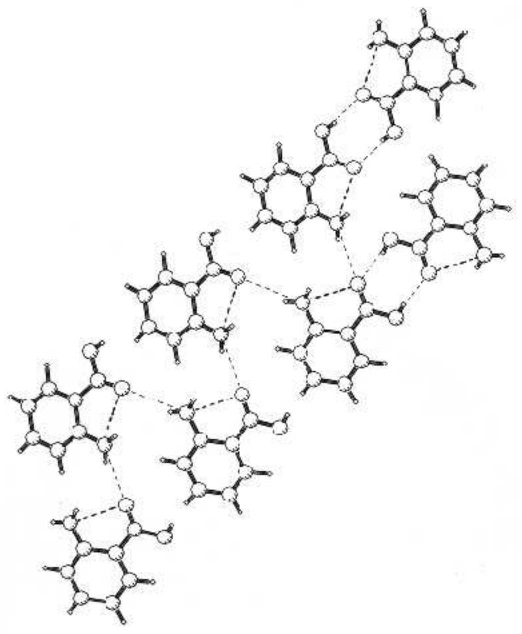


Figure 2: The structure of anthranilic acid (form II)

**2.1. Prerequisites (Input data).** In order to create the mathematical structure (graph), we require information which is available from the experimentally determined crystal structure, in particular from the following data:

(1) The molecules

   (a) the atoms in each molecule (number and kind, i.e. chemical element),

   (b) the covalent bonds between the atoms in a molecule.

(2) Parameters for the crystal structure

   (a) the atomic coordinates and the "lattice parameters" (i.e., parameters which define the dimension and the shape – directions, angles – of the unit cell),

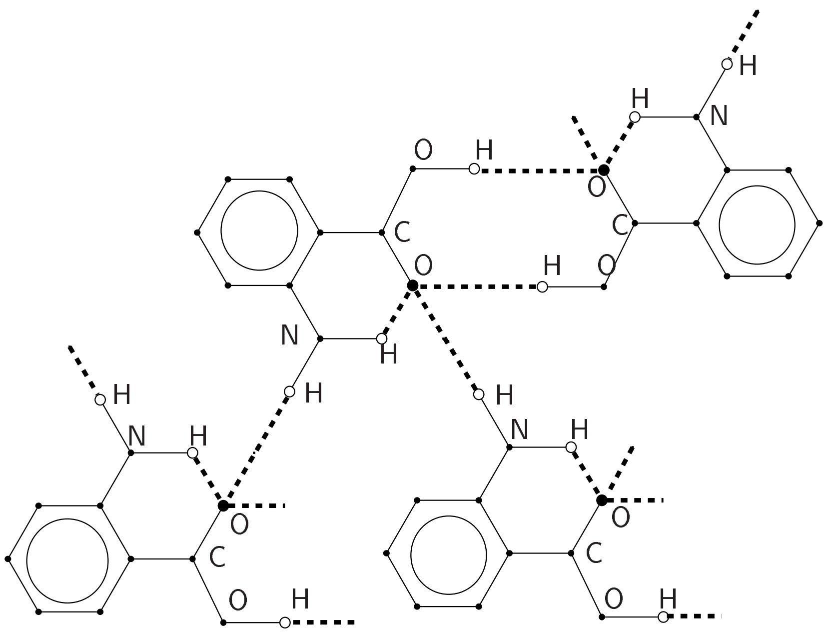



Figure 3: Schematic representation of the crystal structure of anthranilic acid (form II). Broken lines indicate the hydrogen bonds.

(b) the symmetry group $G$ (crystallographic space group) and how it transforms a selected set of atomic coordinates that represents the structure (e.g. the crystallographic asymmetric unit) and from which the entire crystal structure may be generated, using the symmetry operations of the space group including the translations.

(3) The hydrogen bonds.

Hydrogen bonds are generally recognized by the interatomic distance and range of angles, although a universally accepted working definition of a hydrogen bond is elusive ([16], [2]). All these components – atoms, covalent bonds, molecules, hydrogen bonds and symmetry group – are prerequisites for the mathematical approach developed in this paper. They can be classified by parameters, e.g.

by the kind of atoms which are involved or by their chemical or crystallographic equivalence. In this paper we mainly use crystallographic equivalence (cf. the outline in the introduction), i.e. the information from the symmetry group $G$.

**2.2. The array of anthranilic acid.** In our process of carrying out Graph Set Analysis there are several levels of abstraction and specification. To start with the most simple level, we encode the structure of anthranilic acid as a *labelled graph*. For this process the atoms are marked as points (they form the set $V$ of *vertices*), the covalent bonds between the atoms of any single molecule are marked as lines (*edges*), and each hydrogen bond is marked as a broken line (H-*edge*) between the hydrogen atom and the acceptor atom of the hydrogen bond. For each atom we note the respective element symbol at the corresponding vertex (*labelling* $\alpha$; in this specific case, C, H, O, or N). At this stage of abstraction, for some edge labelling $\beta$, the edges can be labelled just by specifying whether they are covalent or H-edges. We reserve a set $C$ for covalent edge labels (in all examples we use a line like N — O between atoms to indicate covalent edges). However, hydrogen bonds are crucial: they are not only marked by broken lines but will be designated by a label from a set $B$ indicating a feature characteristic of the bond, e.g. the atoms involved such as N–H···O and O–H···O. (Mathematically speaking, the labels of covalent edges and H-edges form two disjoint sets $C$ and $B$.) We stress that the hydrogen atoms play a special role and their labels (at this point simply H) are collected in a special label set $\mathrm{H}$.

All these ingredients (a set $V$ of vertices, a set $E$ of edges and the sets $A, B, C, \mathrm{H}$ of labels and the relations between them) together comprise what is called an *array*. In the mathematical definitions 3.1 one will find several restrictions which are obvious for a chemist, and which are due to the chemical origin of this notion (e.g. molecules are generally finite[1] and each H-edge connects a hydrogen atom ($\alpha$-label H) with a non-hydrogen atom). In addition we shall introduce *graph molecules* mathematically by defining them

[1]The treatment described here does not include polymers, which in the current chemical context may be considered to be essentially infinite.

as the covalent *connected components* of the (*covalent subgraph* of the) array (see Fig. 6).

In Figure 4 one finds a representative part of the array of the anthranilic acid under consideration. Some H-edges are fixed notationally $(e_1, \ldots, e_5, e'_1, \ldots, e''_3)$ in order to refer to them later. This notation indicates the edge itself and not its label (different edges may have the same label, e.g. $e_1$ and $e'_1$ are crystallographically equivalent and will therefore get the same label). Here, in addition, vertices which belong to an H-edge are marked by filled (acceptor atom) or unfilled (H-atom) circles; for simplicity the H-atoms in the (hexagonal) benzene rings are not drawn.

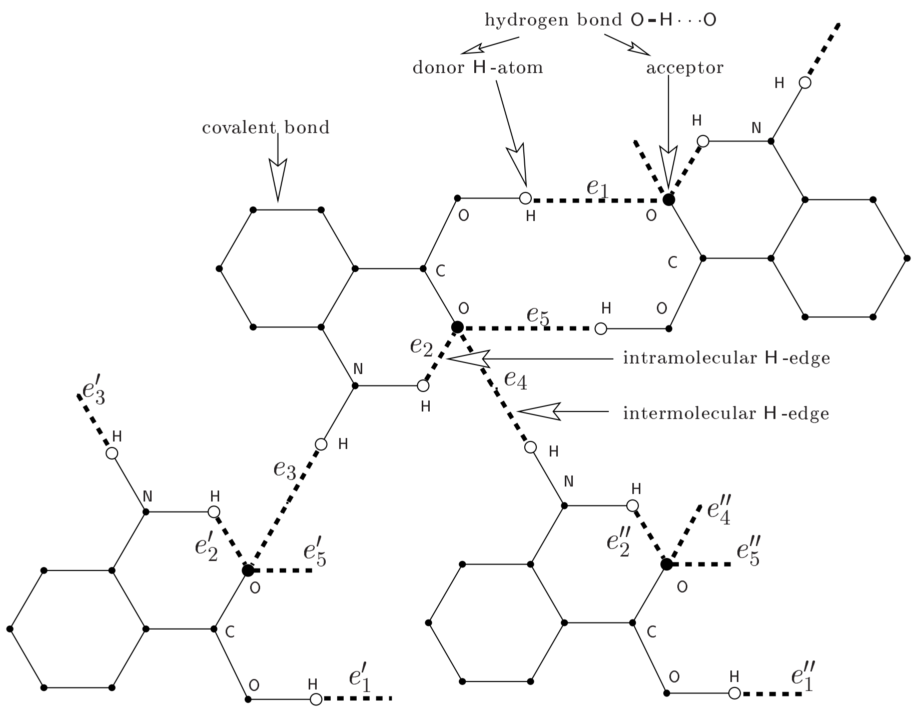



Figure 4: Part of the array of anthranilic acid (form II)

**2.3. The $G$-array of anthranilic acid.** On a microscopic scale we may consider the whole crystal structure of anthranilic acid to be infinite; it is composed of approximately $10^{17}$ molecules all bearing the same intermolecular relationship shown in Figure 4. Thus one can draw only a very restricted part which is representative of the structure.[2]

The symmetry information about the structure of anthranilic acid, which is required to generate the three-dimensional structure, is its symmetry group $G$, well-known to chemists and crystallographers as the crystallographic space group.

Taking into account the symmetry group will lead us to the notion of a $G$-array. Every element of the symmetry group describes a transformation of the three-dimensional space. We assume that we know how the atoms of anthranilic acid are moved by these symmetry operations. This gives us the *action* $\psi$ of the group $G$ on the set $V$ of vertices, e.g. if $v$ is an atom and $g$ a symmetry operation, then $v^g$ will denote the atom to which $v$ is moved under the action of $g$ (mathematically speaking, $g$ induces a permutation $\psi(g)$ on the set $V$). It is obvious that such an action is *compatible* with the array structure, i.e. covalent edges are moved to covalent edges, hydrogen bonds are moved to hydrogen bonds (3.3(3) means just this). An array with a fixed compatible action of $G$ on $V$ is called a *$G$-array*.

We add here some information from the International Tables of Crystallography, which is clear to chemists and crystallographers but may be less familiar to mathematicians. Form II of anthranilic acid crystallizes in the three-dimensional space group $Pbca$ ([14, p. 286]), which determines the symmetry group $G$. There are eight combined symmetry operations of this space group and they give rise to eight so-called equivalent positions, which when applied to the asymmetric unit can be used to generate the contents of the unit cell. The latter is the basic building block of the three-dimensional crystal structure, which in turn may be generated by translating this unit cell (and its contents, of course) in three directions along the unit cell axes.

[2] We have shown only a two-dimensional representation here. Clearly the structure is three-dimensional, but the principle of generating the structure is the same.

The symmetry group $G$ enables us to decide which parts of the $G$-array are crystallographically equivalent (i.e. can be moved to each other by a symmetry operation). This gives rise to a *refined labelling* of the array by so-called *$G$-orbits* (in the sense of group theory). All vertices, edges or molecules, respectively, which are crystallographically equivalent to (i.e., which are in the same $G$-orbit as) a given vertex $v$, edge $e$ or molecule $M$, are put in a set denoted by $v^G$, $e^G$ or $M^G$, which serves as a label for the vertices, edges ($\beta(e) := e^G$) or molecules ($\alpha_{\mathcal{M}}(M) := M^G$), respectively. This is of interest mainly for the H-edges of the array, for which we get the *labelling* $\beta_G$. The $G$-array with such defined labels has a useful property: two H-edges have the same label if and only if there is a symmetry operation moving one to the other (H-*edge label transitivity*, cf. 4.8).

Up to now there is no ambiguity in the notations, because they are independent of any individually chosen symbols or representatives. But of course in practice one cannot use a label which is an infinite set, although this is very convenient for the mathematical treatment . One therefore has to choose a representative or a special symbol for each of the above orbits. It is necessary to do this at least for H-edges and molecules.

In the anthranilic acid example there are three $G$-orbits for H-edges; representative edges are $e_1$, $e_2$ and $e_3$ in Figure 4. Of course, there are many other choices of representative edges, e.g., $e_1''$, $e_2'$ and $e_4$ would be a possible choice but not very convenient. Usually one prefers those which are connected with a small part of the array (e.g., a molecule).

Note that for the hydrogen bonds with $\beta$-label N–H···O we now need two $\beta_G$-labels since there are two classes (orbits) of such bonds: one with an intermolecular H-edge, the other with an intramolecular H-edge. As in [2] we choose the symbols $a, b, c$ to represent the three types of H-edges (cf. Fig. 5). Thus we obtain Table 1 for the H-edges *incident* with a molecule of anthranilic acid.

Remark: The choice of the symbols is rather arbitrary (cf. 5.6). One can use other symbols. The only requirement is that the labelling with these symbols must be H-edge label transitive as in

| representative | $\beta$-label | $\beta_G$-label | symbol for $\beta_G$-label | other symbol for $\beta_G$-label |
|---|---|---|---|---|
| $e_1$ | O–H···O | $e_1^G$ | $a$ | $O_1$–$H_1$···$O_2$ |
| $e_2$ | N–H···O | $e_2^G$ | $b$ | N–$H_6$···$O_2$ |
| $e_3$ | N–H···O | $e_3^G$ | $c$ | N–$H_7$···$O_2$ |
| $e_4$ | N–H···O | $e_4^G(= e_3^G)$ | $c$ | N–$H_7$···$O_2$ |
| $e_5$ | O–H···O | $e_5^G(= e_1^G)$ | $a$ | $O_1$–$H_1$···$O_2$ |

Table 1: H-edges and their labels

case of the labelling $\beta_G$. Table 1 may then serve as a "dictionary" to enable different people (or computers) to understand each other. The labelling $\beta_G$ is fixed whenever we know $G$, and one can then ignore the original labelling $\beta$. In practice, the old labelling $\beta$ also remains a useful parameter.

An analogous labelling as for H-edges can be done (via the group action of $G$) for vertices (*labelling* $\alpha_G$). We remark that for anthranilic acid the vertices also split (with repect to $G$) into several classes (i.e. group orbits of crystallographically equivalent atoms). In Figure 5 some H-atoms are marked; they fall into seven classes (but $H_2, \ldots, H_5$ belong to the benzene ring and are not drawn). In this structure only $H_1, H_6, H_7$ participate in hydrogen bonds; thus the special set of hydrogen symbols would now become $\mathrm{H} = \{H_1, H_6, H_7\}$. Again, two H-atoms with the same label ($\alpha_G$*-label*) are crystallographically equivalent. The last column in Table 1 specifies the $\beta$-label of the H-edges according to the vertex labelling and this might also be used instead of the labelling $\beta_G$.

Finally we consider the whole molecules: the molecules of anthranilic acid are all crystallographically equivalent (i.e. belong to the same orbit). Therefore we need only one symbol to denote the graph molecules, which all look as given in Fig. 6. The corresponding *labelling* $\alpha_{\mathcal{M}}$ (which assigns to each (graph) molecule $M$ its orbit $M^G$) is trivial (i.e. it is constant because there is only one group orbit). However this case (where we might speak of "the" graph molecule of the array) is not typical of more general struc-

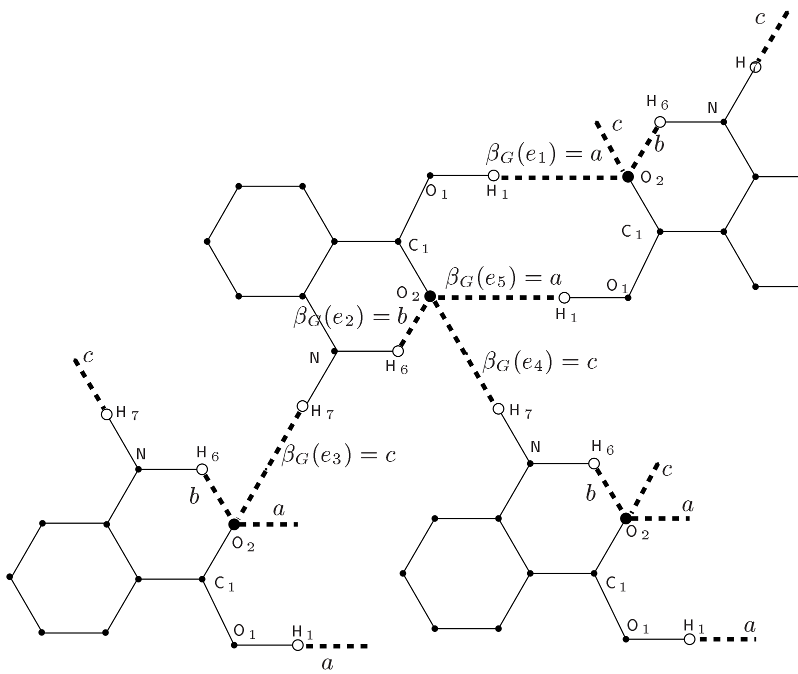



Figure 5: Part of the $G$-array of anthranilic acid (form II)

tures.

With the $G$-array and its labellings (in particular $\beta_G$), we have now reached the stage which is usually the starting point of Graph Set Analysis in crystallography. The mathematical notion of a $G$-array simply models some features of this starting point. Note that because of the graph-theoretical setting the position of vertices in the space or plane does not play any role: we kept only their connection via covalent or hydrogen bonds, e.g. Figure 6 presents a graph molecule corresponding to one molecule in Fig. 5.

**2.4. The constructor graph of anthranilic acid.** The crucial concept for our mathematical treatment of Graph Set Analysis is the *constructor graph*. This was developed as a mathematical tool to express and compute the hydrogen bond features usually considered in Graph Set Analysis.

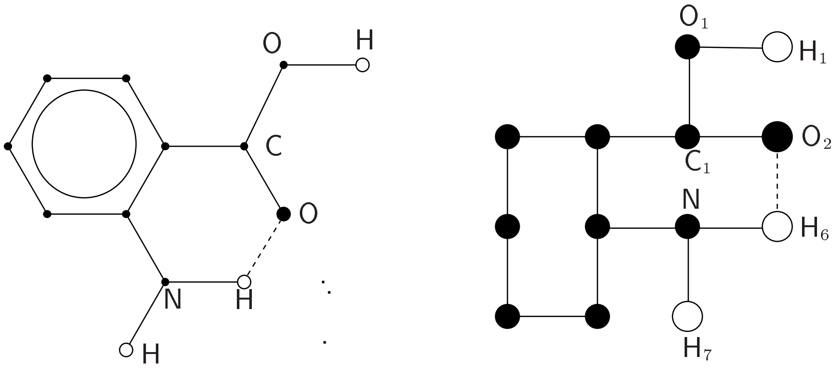



Figure 6: The molecule and the graph molecule of the $G$-array of anthranilic acid (form II)

The idea is very simple: we are no longer interested in the inner structure of a molecule and therefore just consider each molecule as a single "point", i.e. a molecule becomes a vertex of the constructor graph. The edges of the constructor graph are simply the old H-edges of the $G$-array together with their $\beta_G$-labels.

However, some difficulties occur: in the graph we cannot see now at which end of an edge is the H-vertex and at which end is the acceptor vertex; moreover, there might be different H-edges between the same molecules with the same label (mathematically, we have a *multigraph*). We overcome this obstacle in two steps.

First we indicate the molecules to which the acceptor atom and the donor H-atom of an H-edge belong. This will be done by using a directed edge (arrow) from the molecule with the donor H-atom to the molecule with the acceptor atom. Thus every H-edge has an initial vertex (*source)* and a terminal vertex (*sink*). Each intramolecular H-edge becomes a loop (we draw an arrow as a loop) at the vertex which represents the molecule to which the H-edge belongs. The part of the $G$-array of anthranilic acid shown in Fig. 5 thus shrinks to Figure 7 (more precisely, to the part of Fig. 7 which consists of the four lower left points and their edges; in Figure 7 we added a few more points). The result of this treatment is called a $G$-constructor graph (since the action of $G$ on the $G$-array also induces an action on the molecules and thus on the constructor graph, 3.2).

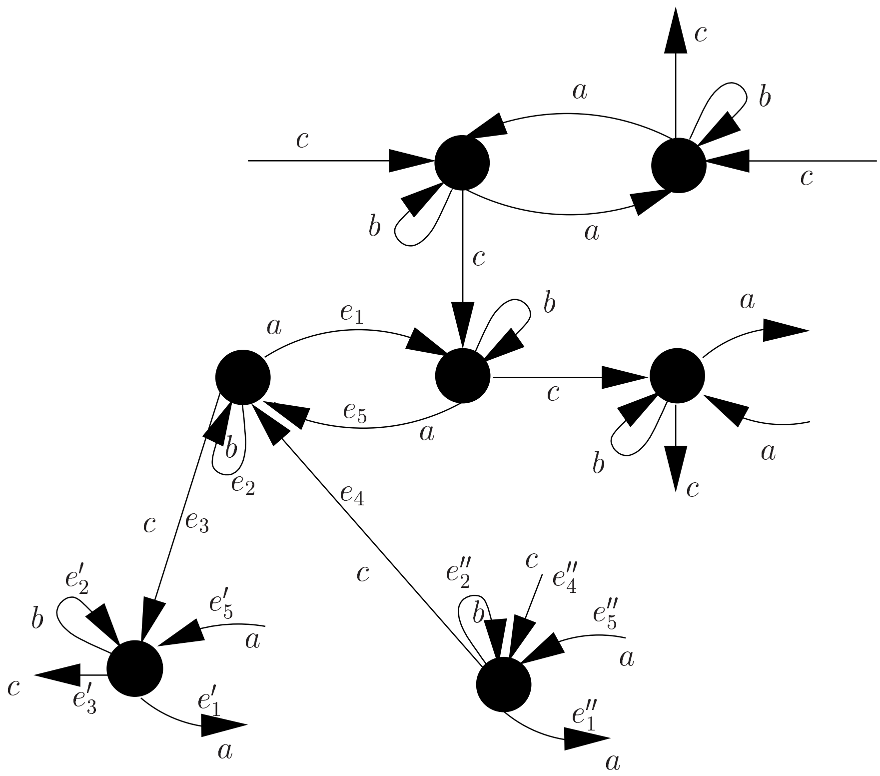



Figure 7: Part of the constructor graph of anthranilic acid (form II)

However, it can happen that after this procedure there exist several H-edges (now directed) which all start in the same molecule or end in the same molecule and still have the same label. This case does not appear in our example. Nevertheless we shall briefly discuss this case here, but in a more rough form, postponing rigorous definitions and explanations to the next sections (an example can be found in [12]). In that case, looking at the $G$-constructor graph, we are faced with subgraphs (e.g. cycles, chains) which include such "non-distinguishable" H-edges and which seem to be absolutely equivalent. Really however, such subgraphs are usually not equivalent if we take into account some additional information from the underlying $G$-array. Therefore we need to distinguish such H-edges which have the same label by some label refinement. It turns out that an appropriate refinement is provided by a suitable normal subgroup $T$ of $G$ (e.g. the subgroup of all translations will always be suitable) by adding the $T$-orbits as refined labels (*$\beta_T$-labels*) to the $\beta_G$-labels. Note that $G$ itself is also a normal subgroup of $G$; we may therefore take $T = G$ whenever a refinement is not necessary.

The resulting graph (more exactly, directed multigraph) with these new $\beta_T$-labels is called the $T$-constructor graph of a given $G$-array. This constructor graph might depend on the choice of $T$, therefore $T$ together with $G$ must be indicated. In our example we can take $T = G$ and we therefore have $\beta_T = \beta_G$.

**2.5. The covalent distance table of a constructor graph.** Some information is lost in the constructor graph, namely everything concerning the inner structure of the molecules. Since such information will sometimes be needed, the constructor graph will be associated with some data stored in the so-called *covalent distance table*.

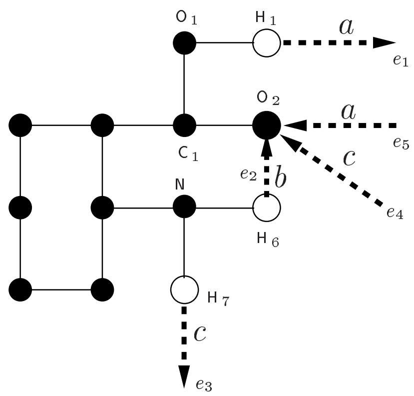
For every molecule $M$ (more precisely, for every $\alpha_{\mathcal{M}}$-label $M^G$) we list all H-edges $e$ (more precisely, their $\beta_T$-labels $b$) which start or end in the molecule $M$. Then we determine all distances (in the graph-theoretical sense, i.e. the length (= number of edges) of a shortest path) between all atoms of $M$ involved in H-edges. These distances will be called *covalent distances*[3].

We present these numbers in a slightly more sophisticated way (which will be useful later). The $\beta_T$-label $b$ of $e$ will get an additional mark (an arrow above the symbol): we can write $\overrightarrow{b}$ or $\overleftarrow{b}$. We imagine that we enter the molecule by an H-edge $e$ with $\beta_T$-label $b$. Since $e$ has a direction we can enter $M$ with the direction of $e$, in which case we use $\overrightarrow{b}$, or against the direction of $e$, then we use $\overleftarrow{b}$. We do the same thing if we imagine leaving the molecule via an H-edge $e$. The distance between the vertex where we enter the

[3]Note that these distances have nothing to do with distance in space.

molecule and the vertex where we leave the molecule becomes an entry in the table for the $\beta_T$-labels involved in $M$.

Figure 8 illustrates this for the molecule of anthranilic acid.




| $\mathsf{cd}(\sigma\sigma')$ | $\sigma' =$ $\overrightarrow{a}$ | $\overleftarrow{a}$ | $\overrightarrow{b}$ | $\overleftarrow{b}$ | $\overrightarrow{c}$ | $\overleftarrow{c}$ |
|---|---|---|---|---|---|---|
| $\sigma =$ $\overrightarrow{a}$ | 3 | 0 | 5 | 0 | 5 | 0 |
| $\overleftarrow{a}$ | 0 | 3 | 6 | 3 | 6 | 3 |
| $\overrightarrow{b}$ | 3 | 0 | 5 | 0 | 5 | 0 |
| $\overleftarrow{b}$ | 6 | 5 | 0 | 5 | 2 | 5 |
| $\overrightarrow{c}$ | 3 | 0 | 5 | 0 | 5 | 0 |
| $\overleftarrow{c}$ | 6 | 5 | 2 | 5 | 0 | 5 |

Figure 8: The molecule and its covalent distance table for anthranilic acid (form II)

E.g. entering the molecule against the H-edge $e_1$ (label $a$) and leaving along $e_3$ (label $c$) gives the entry $\mathsf{cd}(\overleftarrow{a}\overrightarrow{c}) = 6$ ($\mathsf{cd}$ stands for covalent distance), since there is a shortest path of length 6 connecting the entering atom $\mathsf{H}_1$ and the leaving atom $\mathsf{H}_7$.

In an analogous manner "entering"(i.e. starting) along the intramolecular H-edge $e_2$ (label $b$) and "leaving"(i.e. finishing) with the same $e_2$ gives $\mathsf{cd}(\overrightarrow{b}\overrightarrow{b}) = 5$, since the distance from the acceptor atom $\mathsf{O}_2$ of $e_2$ to the H-atom $\mathsf{H}_6$ of $e_2$ is 5.

Zero-valued entries indicate that we enter and leave the molecule at the same atom.

The entries in the table are highly redundant (half of the rows determines the other half, for more details see 7.3), e.g. $\mathsf{cd}(\overleftarrow{a}\overrightarrow{c}) = \mathsf{cd}(\overleftarrow{c}\overrightarrow{a})$, $\mathsf{cd}(\overrightarrow{a}\overrightarrow{b}) = \mathsf{cd}(\overleftarrow{b}\overleftarrow{a})$.

It can happen that entries have no meaning because, e.g., it might be impossible to enter the molecule against (or along) the direction of some H-edge with a label in the table. In such cases we use blank entries or the sign $\infty$ for the corresponding rows or columns (such rows or columns, however, make sense if we collect the covalent distance tables of several molecules together to form the covalent distance table of the $G$-array). In our example this situation does not arise; it would occur if there were, e.g., no H-edge starting from the atom $\mathsf{H}_7$. Then the row $\overleftarrow{c}$ and the column $\overrightarrow{c}$ would become meaningless.

**2.6. Graph sets and descriptors of anthranilic acid.** Graph Set Analysis deals with the determination of *descriptors* of graph sets of a given $G$-array. *Graph sets* can be represented by four hydrogen bond patterns: rings, chains, selfs and discretes. All such patterns which are equivalent in a particular sense (for details see 4.7) can be considered as a graph set. We very briefly mention here these patterns, which are intuitively more or less obvious in particular examples but (as we shall see in the next sections) require some effort for their mathematical formulation. The main advantage of the constructor graph is that all notions and computations are easier to handle in the constructor graph.

A ring in the constructor graph (*constructor ring*) is (the graph of) a simple cycle in the constructor graph; an example is shown in Figure 9.

If we recall the inner structure of the molecules (i.e. the vertices of the constructor graph), then a *ring* in the $G$-array can be defined similarly by choosing a shortest path going through the molecules corresponding to a ring in the constructor graph. See Figure 10, where there is only one shortest path inside each molecule. There might exist several such shortest paths, in which case any of them defines a ring, but all give the same descriptor and therefore only one of these H-*edge equivalent* rings is needed to determine the descriptors below.

The *(quantitative) descriptor* of a ring is a symbol $\mathbf{R}^{\mathrm{a}}_{\mathrm{d}}(n)$, where "a" and "d" are the numbers of acceptor atoms and donor H-atoms of the H-edges involved, respectively (they must be distinguished from the notations $a, \ldots$ for the $\beta_G$-labels), and $n$ is the length of the ring (= number of vertices = number of edges). As additional characteristics we add the set $L$ (called the *type*) of all H-edge labels involved and write $L : \mathbf{R}^{\mathrm{a}}_{\mathrm{d}}(n)$ in this situation. This, in some sense, extends and makes more precise the idea of "optional fields"

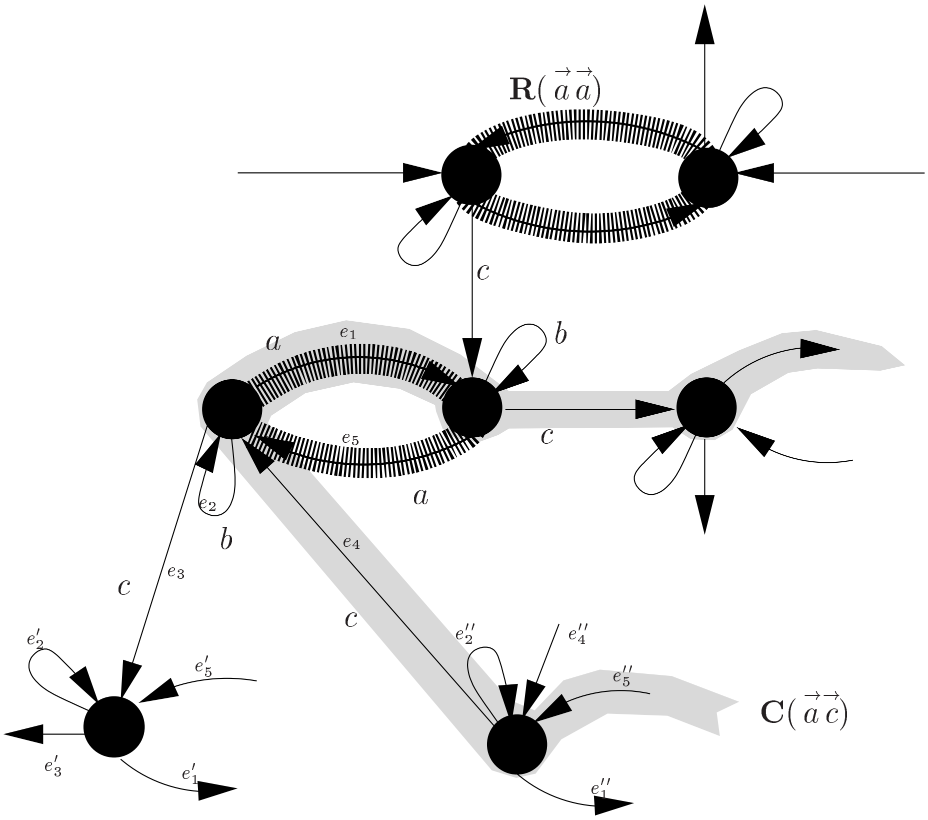



Figure 9: Two rings $\mathbf{R}(\vec{a}\,\vec{a})$ and a chain $\mathbf{C}(\vec{a}\,\vec{c})$ (notation see 2.7) in the constructor graph of anthranilic acid (form II)

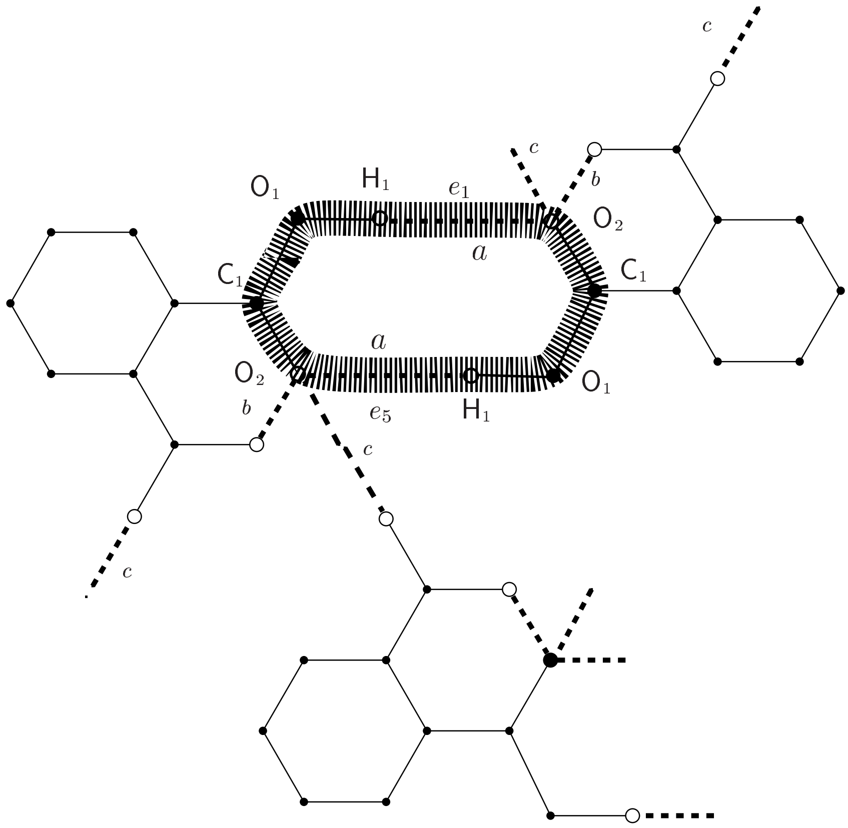



Figure 10: A ring in the $G$-array of anthranilic acid (form II)

considered in [18]. Thus the ring in Fig. 10 has the descriptor $\{a\} : \mathbf{R}_2^2(8)$.

*Chains* are (the graphs of) paths which are *periodic* with respect to some symmetry operation $g$ (of infinite order) of the symmetry group $G$.

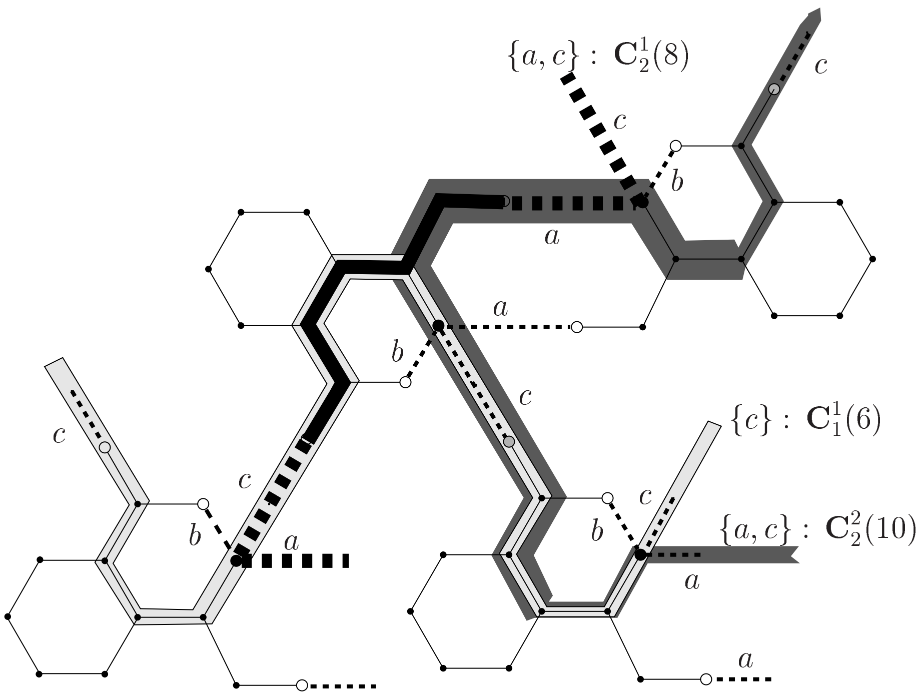



Figure 11: Chains in the $G$-array of anthranilic acid (form II)
$\mathbf{C}_2^1(8)$ (= $\mathbf{C}(\overrightarrow{a}\,\overleftarrow{c})$) in black, $\mathbf{C}_1^1(6)$ (= $\mathbf{C}(\overrightarrow{c})$) in light grey, $\mathbf{C}_2^2(10)$ (= $\mathbf{C}(\overrightarrow{a}\,\overrightarrow{c})$) in dark grey (thick and thin).

In Figure 11 one finds examples of chains, marked by differently shaded areas. In the chain $\{a, c\} : \mathbf{C}_2^2(10)$ the larger part indicates the period (i.e. a repeated segment) of length 10. This chain is also indicated in the constructor graph in Figure 9.

The descriptor $L : \mathbf{C}_{\mathrm{d}}^{\mathrm{a}}(n)$ of a chain is obtained from a (minimal) period by indicating the H-edge labels involved and counting the acceptor atoms, donor H-atoms and the length of a period (analogously as for rings). In our examples we get the descriptors $\{c\} : \mathbf{C}_1^1(6)$, $\{a, c\} : \mathbf{C}_2^2(10)$ and $\{a, c\} : \mathbf{C}_2^1(8)$ (cf. Fig. 11).[4]

An intramolecular H-edge together with a shortest path between the acceptor and donor H-atom is called an *elementary self* (symbol $L : \mathbf{S}_{\mathrm{d}}^{\mathrm{a}}(n)$). All elementary selfs of a molecule which have $\beta_G$-label in a set $L$ form a *self* (symbol $L : \mathbf{S}_{\mathrm{d}}^{\mathrm{a}}$).

For anthranilic acid there is only one intramolecular H-edge in a molecule (e.g. $e_2$), i.e. we have only elementary selfs. Figure 12 shows the self of anthranilic acid in the $G$-array as well as in the constructor graph (*constructor self*), where it becomes nothing but a vertex with a single loop.

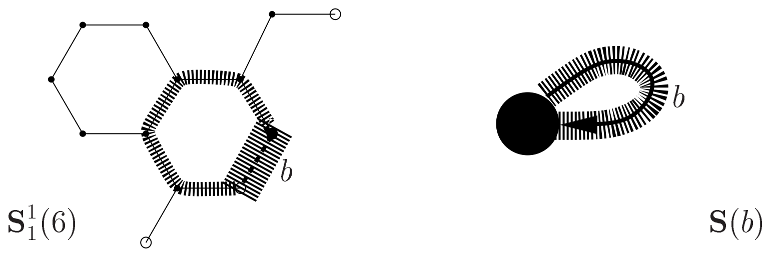



Figure 12: A self in the $G$-array and in the constructor graph of anthranilic acid (form II)

In the case of several loops the refined labelling $\beta_T$ (mentioned above) ensures that all loops at a single vertex in the constructor graph get different labels (and therefore can be listed as sets of labels).

Finally there is one more sort of graph set represented by so-called *discretes*. These are special H-edge patterns in a $G$-array which are neither rings, selfs nor chains. In the case of anthranilic acid they do not occur and we refer to the mathematical part for further discussion (Definitions 4.6 and 6.5).

**2.7. Qualitative descriptors in the constructor graph of anthranilic acid.** Hydrogen bond patterns and their descriptors were

[4]When both subscript and superscript have a value of one, then they are often omitted; here e.g. $\mathbf{C}(6)$.

described, for instance, in [11] and [2]. As seen above the constructor graph is – among other things – a suitable tool for finding such information (see also Figures 9 - 12).

We shall now characterize patterns such as constructor rings, chains, selfs and discretes by so-called *qualitative descriptors*. In contrast with the descriptors considered above, these qualitative descriptors retain much more information, e.g. which H-edges are involved in which order.

How shall we obtain a qualitative descriptor of a constructor ring, chain, self or discrete? Recall that such patterns (except non-elementary selfs) are paths in a graph. We walk along the path and record every H-edge visited by making a note of its $\beta_T$-label and the direction (with or against the arrow) in which we walk along this edge. The mathematical treatment then ensures that no confusion can appear, i.e. that crystallographically equivalent rings, chains etc. get the same qualitative descriptor while crystallographically different rings, chains etc. get different qualitative descriptors (and that the descriptors of these rings, chains etc. do not depend on the particular edges but only on the labels and directions of the edges).

Consider for instance the ring in Fig. 9 or 10. Walking along this ring we pass along two H-edges $e_1$ and $e_5$ with label $a$. If we walk counterclockwise (against the H-edge arrows) then we write $\overleftarrow{a}$ for every passage and the qualitative descriptor of this ring is

$$\mathbf{R}(\overleftarrow{a}\,\overleftarrow{a})\,.$$

If we walk clockwise we get $\mathbf{R}(\overrightarrow{a}\,\overrightarrow{a})$ (but only one of these representations will be used).

Looking at Fig. 9 the reader should also understand the nature of the walk (which is also a "ring" in a more general sense) described by the *directed label sequence* $(\overleftarrow{a}\,\overrightarrow{b}\,\overleftarrow{a})$: we walk through the H-edges $e_1$, $e_2$ and $e_5$ (always on a shortest path inside a molecule) with labels $a$, $b$ and $a$.

Likewise we can describe chains by the qualitative descriptor of a (minimal) period. The chains in Figure 11 thus have the descriptors (cf. also Fig. 9)

$$\mathbf{C}(\overrightarrow{c})\,,\quad \mathbf{C}(\overrightarrow{a}\,\overrightarrow{c})\quad\text{ and }\quad \mathbf{C}(\overrightarrow{a}\,\overleftarrow{c})\,.$$

One should note that from these qualitative descriptors it is easy to locate the corresponding rings or chains in the constructor graph (or in the $G$-array). The descriptor of a chain can be presented in several equivalent ways, e.g. $\mathbf{C}(\overrightarrow{c}) = \mathbf{C}(\overleftarrow{c})$, $\mathbf{C}(\overrightarrow{a}\,\overrightarrow{c}) = \mathbf{C}(\overrightarrow{c}\,\overrightarrow{a}) = \mathbf{C}(\overleftarrow{a}\,\overleftarrow{c}) = \mathbf{C}(\overleftarrow{c}\,\overleftarrow{a})$ and $\mathbf{C}(\overrightarrow{a}\,\overleftarrow{c}) = \mathbf{C}(\overleftarrow{c}\,\overrightarrow{a}) = \mathbf{C}(\overrightarrow{c}\,\overleftarrow{a}) = \mathbf{C}(\overleftarrow{a}\,\overrightarrow{c})$. This becomes clear if one imagines these chains as infinite strings

$$\ldots \overrightarrow{c}\,\overrightarrow{c}\,\overrightarrow{c}\,\overrightarrow{c} \ldots, \quad \ldots \overrightarrow{a}\,\overrightarrow{c}\,\overrightarrow{a}\,\overrightarrow{c}\,\overrightarrow{a}\,\overrightarrow{c}\,\overrightarrow{a}\,\overrightarrow{c} \ldots \text{ and } \ldots \overrightarrow{a}\,\overleftarrow{c}\,\overrightarrow{a}\,\overleftarrow{c}\,\overrightarrow{a}\,\overleftarrow{c}\,\overrightarrow{a}\,\overleftarrow{c} \ldots,$$

which may be followed from right to left or from left to right, while a period can start at any place.

Provided that we are interested only in H-edges (and ignore the covalent parts of the path, just choosing always shortest covalent paths) then the directed label sequences are very suitable and useful to describe paths in constructor graphs as well as in $G$-arrays (cf. 5.11).

**2.8. Descriptor computations.** There are simple formulas to enable one to compute the descriptors from the qualitative descriptors using the covalent distance table. Here we describe this procedure for the case where – as for anthranilic acid – the H-edge labelling $\beta_G$ need not be refined as in step 2 of 2.4, i.e. where $\beta_G = \beta_T$; otherwise we must divide some numbers by a special factor, see 7.9:

The type is just the set of $\beta_G$-labels occurring in the qualitative descriptor, the number $n$ is the sum of those entries in the covalent distance table which are the label covalent distances $\mathsf{cd}(\sigma_i\sigma_{i+1})$ of consecutive labels in the directed label sequence of the descriptor.

The number of acceptors is the number of labels (since each label represents an H-edge) but we have to subtract 1 every time two H-edges share the same acceptor (in these cases, mathematically, we introduce equivalence classes for *acceptor* H-*vertex equalities*, cf. 7.4). Such a case is found easily from the covalent distance table:

Specifically, if e.g. $\mathsf{cd}(\overrightarrow{a}\,\overleftarrow{c}) = 0$ then the H-edges (of a particular molecule) corresponding to these labels share the same acceptor (cf. Fig. 8).

Analogously, entries $\mathsf{cd}(\overleftarrow{x}\,\overrightarrow{y}) = 0$ indicate that H-edges of a molecule with label $x$ and $y$ share the same donor H-atom (this case does not appear for anthranilic acid as can be seen from the table in Fig. 8).

Table 2 shows how to do these computations for the qualitative descriptors of anthranilic acid considered above.

| Graph set | Ring | Chain | | | Self |
|---|---|---|---|---|---|
| label sequence $(\sigma_1\sigma_2\dots\sigma_m)$ | $(\vec{a}\,\vec{a})$ | $(\vec{c})$ | $(\vec{a}\,\vec{c})$ | $(\vec{a}\,\overleftarrow{c})$ | $\{b\}$ or $(\vec{b})$ |
| qualitative descriptor | $\mathbf{R}(\vec{a}\,\vec{a})$ | $\mathbf{C}(\vec{c})$ | $\mathbf{C}(\vec{a}\,\vec{c})$ | $\mathbf{C}(\vec{a}\,\overleftarrow{c})$ | $\mathbf{S}(\{b\})$ |
| $L =$ | $\{a\}$ | $\{c\}$ | $\{a, c\}$ | $\{a, c\}$ | $\{b\}$ |
| descriptor | $L : \mathbf{R}^{\mathrm{a}}_{\mathrm{d}}(n)$ | $L : \mathbf{C}^{\mathrm{a}}_{\mathrm{d}}(n)$ | | | $L : \mathbf{S}^{\mathrm{a}}_{\mathrm{d}}(n)$ |
| | $\mathbf{R}^2_2(8)$ | $\mathbf{C}^1_1(6)$ | $\mathbf{C}^2_2(10)$ | $\mathbf{C}^1_2(8)$ | $\mathbf{S}^1_1(6)$ |
| $n = m+$ $\mathsf{cd}(\sigma_1\sigma_2)+\dots$ $\dots+\mathsf{cd}(\sigma_m\sigma_1)$ | $m$+3+3 =8 | $m$+5 =6 | $m$+5+3 =10 | $m$+0+6 =8 | |
| $n = \mathsf{cd}(\vec{b}\,\vec{b}) + 1$ | | | | | 5+1=6 |
| a = (see 7.9) | $2-0=2$ | $1-0=1$ | $2-0=2$ | $2-1=1$ | $\mathrm{a}(\{b\}) = 1$ |
| d = (see 7.9) | $2-0=2$ | $1-0=1$ | $2-0=2$ | $2-0=2$ | $\mathrm{d}(\{b\}) = 1$ |

Table 2: Computing descriptors from qualitative descriptors

**2.9. Remarks.** We close this section with some remarks and problems. In principle, hydrogen bond patterns can be very complicated. One could mix chains, rings and selfs, etc. Even with our definitions there are infinitely many graph sets, for instance represented by chains: for anthranilic acid we can compose H-edges with label $a$ and $c$ to form chains of an arbitrarily large period, e.g. $\mathbf{C}(\vec{a}\,\overleftarrow{c}\,\overleftarrow{a}\,\overleftarrow{c})$ gives $\{a, c\} : \mathbf{C}^3_4(18)$ ($n = m + 0 + 5 + 3 + 6 = 18$, $\mathrm{a} = 4 - 1 = 3$, $\mathrm{d} = 4 - 0 = 4$, cf. the Table in Fig. 8).

In general we restrict ourselves to those graph sets which are of some chemical interest for Graph Set Analysis, e.g., for a given type, one can restrict to chains of shortest period (cf. [3]).

Moreover, one could think about the choice of a complete package of graph sets (i.e. a set of graph sets from which all other graph sets of the array can be deduced by well-defined rules). In general this is still an open problem which is currently under investigation.

Mathematically, a very challenging problem concerns the extent to which the symmetry group $G$ of a crystal structure restricts (or even determines) the graph sets and (qualitative) descriptors. This problem is currently under investigation.

Finally, the methods presented here are not restricted to the analysis of hydrogen bond patterns. The mathematical notions can be interpreted in very different situations. We merely need graphs (here arrays) with a group action (here symmetry operations), specified "interesting bonds", i.e. edges (here H-edges), and "interesting" patterns decribable by paths (here rings, chains, selfs and discretes). Moreover, for chemical applications, the vertices of the graph need not even be atoms: one could use other chemical structures forming new "vertices" (e.g. rings, ladders, chains in crystals).

# 3 Arrays

We now change the style of presentation and move into a more or less purely mathematical mode in order to elucidate the mathematical basis for the description in the previous section. All graph- and group-theoretical notions are defined in Section 8.

**3.1. Definitions.** Let H be some non-empty fixed set[5] of symbols. An ***array*** is an undirected labelled graph $\Gamma := (V, E, \alpha, \beta)$ such that

(1) $\alpha : V \to A$ is a mapping into a set $A$ with $\mathrm{H} \subset A$.

[5] This set usually consists of a single symbol for a hydrogen atom. However, we reserve a whole set of symbols in order to have more flexibility for further applications.

(2) $\beta : E \to B \cup C$ is a mapping into the union of disjoint sets $B$ and $C$, such that $\beta(\{u, v\}) \in B$ implies $\alpha(u) \in \mathrm{H}$ or $\alpha(v) \in \mathrm{H}$, but not both.

(3) Each connected component of the subgraph $(V, \{e \in E \mid \beta(e) \in C\})$ is finite.

$\beta(e)$ is called the ***type*** of an edge $e \in E$. For $D \subseteq B \cup C$, let

$$E_D := \{e \in E \mid \beta(e) \in D\}.$$

Note that $E = E_B \cup E_C$ is a disjoint union. The edges in the set $E_C$ are called ***covalent*** edges while the edges in $E_B$ are ***hydrogen edges*** or, for short, H-***edges***. A vertex $v$ with $\alpha(v) \in \mathrm{H}$ is called a ***hydrogen vertex***. Because of (2), each hydrogen edge $e$ has exactly one hydrogen vertex, called the ***donor*** H-***vertex*** of $e$. The other vertex is the ***acceptor vertex*** of $e$.

The connected components of the ***covalent subgraph*** $(V, E_C)$ are the ***graph molecules*** of $\Gamma$ (often just called molecules). (3) states that graph molecules must be finite. Edges between vertices of a graph molecule are ***intramolecular***, those between vertices of different graph molecules are ***intermolecular*** edges. Only hydrogen edges can be intermolecular.
The set of all graph molecules is denoted by $\mathcal{M}$. For graph molecules we usually use letters like $M$. If $u \in V$ belongs to the vertex set of a graph molecule $M \in \mathcal{M}$ then, for simplicity, we write $u \in M$ (instead of $u \in V(M)$). Moreover, we say that an H-edge $e = \{u, v\}$ is an intramolecular edge of a graph molecule $M$, whenever both vertices $u, v$ belong to $M$ (thus we can speak, e.g., of a cycle in $M$ containing $e$). An intermolecular H-edge $e$ is ***incident*** with a molecule $M$ if one of its vertices belongs to $M$.

Remark: Chemists also use two further notions ([2, 18]): any connected subgraph of an array is called a ***network***. If all H-edges in a network have the same label, say $\beta(e) = b \in E_B$, then the network is called a ***motif*** (i.e., it is a connected subgraph of $(V, E_{\{b\}})$).

**3.2. Compatible actions.** An ***automorphism*** of an array $\Gamma$ is, as usual, a permutation of the vertices that maps edges to edges, and preserves labels. The group of all automorphisms is denoted by $\mathrm{Aut}(\Gamma)$. Note that $\mathrm{Aut}(\Gamma)$ is a subgroup of $\mathsf{Sym}(V)$.

A ***compatible action*** of some group $G$ on an array $\Gamma$ is given by a group homomorphism $\psi : G \to \mathrm{Aut}(\Gamma)$. Therefore $G$ acts on $V$. For $g \in G$, $v \in V$, the ***action*** $\psi(g)(v)$ of $g$ on $v$ is denoted by $v^g$. In the following we shall always consider a fixed action $\psi$.

We write $(G, V)$ to emphasize that we consider an action of $G$ on $V$. This action induces a number of other actions of $G$ on sets of "parts" of $\Gamma$. First of all we get an action $(G, E)$ on the set of all edges, an action $(G, E_B)$ on the set of H-edges as well as an action $(G, \mathcal{M})$ on the set of graph molecules. Moreover, whenever we fix a "type" of parts of $\Gamma$ under consideration, for example paths or cycles, then we also get an action of $G$ on the set of all parts of that fixed "type" (for details see 8.2).

If $e$ is an edge and $M$ is a graph molecule, we write $e^g$ and $M^g$, resp., to denote their images under the corresponding action of $g \in G$.

$\mathrm{Orb}(G, V)$ denotes the set of ***orbits*** (see 8.3) of an action $(G, V)$. Two elements (e.g. vertices or edges or molecules or paths) of a set on which $G$ acts are called ***G-equivalent*** if they belong to the same orbit. (For chemists this just means that they are ***crystallographically equivalent***.)

**3.3.** $G$**-arrays.** Let $G$ be a group. A ***G-array*** consists of an array (cf. 3.1) $\Gamma = (V, E, \alpha, \beta)$ together with a compatible action of $G$ on $\Gamma$; more explicitly this means the following properties:

(1a) $v^{id} = v$ ($id$ denoting the neutral element of the group $G$),

(1b) $v^{g_1 g_2} = (v^{g_1})^{g_2}$ ($g_1 g_2$ denoting the product of $g_1$ and $g_2$ in $G$),

(2) $e^g \in E$ for all $e \in E$ and $g \in G$,

(3a) $\alpha(v^g) = \alpha(v)$ for all $v \in V$ and $g \in G$,

(3b) $\beta(e^g) = \beta(e)$ for all $e \in E$ and $g \in G$.

**3.4. Remarks.**

a) Conditions (1a,b) in 3.3 just express the fact that $\psi : G \to \mathsf{Sym}(V)$, defined by $\psi(g) : V \to V : v \mapsto v^g$, is an action (cf. 8.1, in particular $\psi(g)$ is a permutation), (2) says that $\psi(g)$ is an automorphism of the graph $(V, E)$, and (3a,b) means compatibility with the labellings of the array $\Gamma$.

b) It is a crucial point for the whole conception of our approach to understand the origin of the group $G$. Given a crystallographic array, to the set $V$ of atoms one can associate a set $\tilde{V}$ of points in 3D-space. It is convenient to think that there is a one-to-one correspondence between $V$ and $\tilde{V}$ (e.g., points in $\tilde{V}$ are regarded as "centers" of atoms in $V$). Let $Iso(\tilde{V})$ be the group of all symmetries (*isometries*) of the crystal given by $\tilde{V}$ in 3D-space. It is clear that $Iso(\tilde{V})$ preserves the set $\tilde{V}$ (i.e., is contained in the space symmetry group of the crystal defined by the set $\tilde{V}$). Usually the group $Iso(\tilde{V})$ is known to chemists as a result of the experimental determination of the crystal structure, moreover, this knowledge is exploited for the representation of an array in a form convenient for crystallographers (cf., e.g., [14]). In the following the reader may assume that $G = Iso(\tilde{V})$ in most cases, i.e., that our $G$-array is defined in terms of the symmetry group of a crystal.

   Evidently, the group $Iso(\tilde{V})$ acts on the set $\tilde{V}$, and this action, as a rule, may be identified with a certain permutation group acting on $V$. For practical purposes this permutation group $(G, V)$ is usually adequate for dealing with the symmetries of crystals. Nevertheless in our apprach we prefer to have more freedom for the group $G$, not necessarily identifying it with its permutational action. It is quite enough to treat the "concrete" group $G = Iso(\tilde{V})$ as an "abstract" group $G$. This approach has the advantange that we can consider several actions of the same group $G$ but on different objects connected with $\Gamma$ (edges, molecules, labels etc., cf., e.g., 3.2, 8.2, 5.4c); otherwise we would have to deal with several different permutation groups. These advantages will become even clearer in particular in the framework of a more formal definition of infinite periodic graphs (however, this question is out of the scope of the present paper, but we refer the reader to [5]).

c) The action under consideration, say $\psi : G \to \mathrm{Aut}(\Gamma)$, is fixed, and we will never consider two different actions of the same group on the same set at the same time. Therefore, in what follows, we simply speak of the $G$-array $\Gamma$, and $v^g$ denotes the action of $g$ on $v$.

   As mentioned above, for chemical applications the action $\psi$ is always given by the chemist as the action of the group $G$ of isometries on the crystal structure under consideration. Thus, e.g., if $v$ is some atom with coordinates $(x, y, z)$ in 3D-space then $v^g$ is the atom with coordinates obtained from $(x, y, z)$ by applying the isometry $g$ to this point.

d) Since in practice a group and a natural action of this group on $V$ always are given, in the following we shall consider $G$-arrays only. However, every definition also can be used, in principle, for arbitrary arrays: in fact, every array can be considered as a $G$-array for $G = \mathrm{Aut}(\Gamma)$ (with its natural action), if no other group is prescribed.

# 4 Graph sets and their quantitative descriptors

**4.1. Types and levels.** Let $\Gamma$ be a $G$-array. For a set $F \subseteq E$ of edges, let

$$\tau(F) := \{\beta(e) \mid e \in F, \beta(e) \in B\}$$

denote its ***hydrogen type set***, or, for short, its ***type***. Thus every type is a set of labels: the number $|\tau(F)|$ of elements is its ***level***. The type and level of subgraphs of $\Gamma$ as well as of paths are those of the corresponding edge sets.

**4.2. M-simple paths.** A path (cf. 8.7) in $\Gamma$ is ***molecule-simple*** or, for short, an ***M-simple path***, if all its H-edges are intermolecular and if it passes through each graph molecule at most once. The latter means, more precisely, that every molecule is incident (cf. 3.1) with at most two H-edges of the path and that for finite paths which are not cycles the first and the last vertex belong to different graph molecules.

An ***M-simple cycle*** is just a path which is both a cycle and M-simple. Cycles with exactly one H-edge will also be called M-simple.

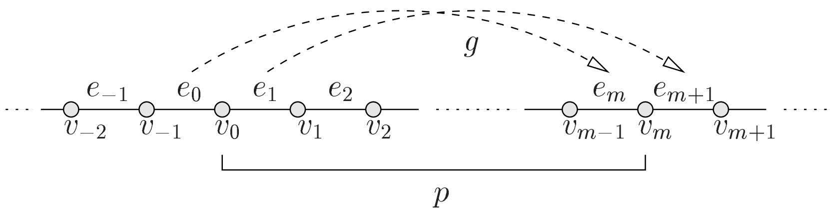
**4.3. Rings.** A ***ring*** $r$ is the graph of an M-simple cycle in $\Gamma$, such that there is no shorter M-simple cycle containing the same hydrogen edges as $r$. To each ring $r$, a symbol,

$$L : \mathbf{R}^{\mathrm{a}}_{\mathrm{d}}(n)$$

is associated. Here $n$ denotes the length of the ring (this coincides with the number of edges as well as with the number of vertices in the ring), $L$ its type, "d" the number of donor H-vertices, and "a" the number of acceptor vertices of hydrogen edges occurring in $r$.

**4.4. Chains.** For the next definition we need the notion of a $G$-periodic path. All mathematical details can be found in 8.8. Here we explain this notion more intuitively.

An infinite M-simple path $c = (e_i, v_i)_{i \in \mathbb{Z}}$ is ***G-periodic*** if it is mapped ("shifted") onto itself by some $g \in G$, i.e. $g : v_i \mapsto v_{i+m}$ for some $m \neq 0$, as shown in the following picture:




The $g$-period length $m$ is denoted by $\pi_g(c)$. The smallest possible $m > 0$ (with respect to all such $g \in G$) is the ***G-period***

***length*** $\pi_G(c)$ of $c$, and a connected piece $p$ of that length is called a ***G-period*** of $c$.

A ***chain*** is the graph of a $G$-periodic M-simple path $c$ for which there is no $G$-periodic M-simple path of shorter $G$-period length containing the same hydrogen edges.

To each chain we associate a symbol,

$$L:\ \mathbf{C}_{\mathrm{d}}^{\mathrm{a}}(n)\,,$$

where $L$ is the type of the chain, $n$ is the $G$-period length, "d" the number of donor H-vertices and "a" the number of acceptor vertices of hydrogen edges within one $G$-period of $c$.

**4.5. Selfs.** An ***elementary self*** is the graph of a cycle containing exactly one H-edge such that there is no shorter cycle with this unique H-edge. In particular this implies that the H-edge must be intramolecular i.e. the cycle belongs to a single molecule.

A ***self*** is the union $\bigcup S$ (cf. 8.4) of a non-empty set $S$ of elementary selfs within one graph molecule, such that with any intramolecular H-edge also all H-edges of the same type in the same molecule belong to $\bigcup S$. To each self $\bigcup S$, we associate a symbol,

$$L:\ \mathbf{S}_{\mathrm{d}}^{\mathrm{a}}$$

where $L$ is the type of $S$, "d" the number of donor H-vertices and "a" the number of acceptor vertices in hydrogen edges occurring in $\bigcup S$. For selfs $\bigcup S$ of level $|L| = 1$, the length $n$ of an elementary self in $S$ is given as an additional parameter, and the symbol

$$L:\ \mathbf{S}_{\mathrm{d}}^{\mathrm{a}}(n)$$

is associated with $\bigcup S$.

<u>Remarks:</u> In most crystals, as in the case of anthranilic acid (cf. Fig. 12), there appear only selfs of level 1, i.e. we can determine $n$. For higher level types (i.e., $|L| \geq 2$) it is still under discussion which number $n$ should be chosen to characterize the self. As soon as chemists provide a reasonable intramolecular structure of higher level selfs, the above definition can be changed adequately.

(This, however, cannot affect the treatment of selfs in the constructor graph, where they are simply looped graphs with one vertex, cf. 6.4.) A self $\bigcup S$ of type $L = \{a, b, \dots\}$ is a graph consisting of the union of the selfs of level 1 in $S$, say $\{a\} : \mathbf{S}^{\mathrm{a}_1}_{\mathrm{d}_1}(n_1)$, $\{b\} : \mathbf{S}^{\mathrm{a}_2}_{\mathrm{d}_2}(n_2), \dots$ . Thus it might be reasonable to use the symbol $L : \mathbf{S}^{\mathrm{a}}_{\mathrm{d}}(n_1, n_2, \dots)$ for the self $\bigcup S$ instead of $L : \mathbf{S}^{\mathrm{a}}_{\mathrm{d}}$ as above.

**4.6. Discretes.** For $L \subseteq B$ a molecule $M \in \mathcal{M}$ is called an ***L-end*** if it is incident with exactly one H-edge $e$ with $\beta(e) \in L$, and this $e$ is intermolecular. A ***discrete*** is the graph of an M-simple path $p$ between two vertices $u, v \in V$ such that the molecules to which $u$ and $v$ belong are $\tau(p)$-ends and there is no shorter path with exactly the same H-edges.

To each discrete $p$ we associate a symbol,

$$L : \mathbf{D}^{\mathrm{a}}_{\mathrm{d}}(n).$$

Here $L$ is the type of $p$, $n$ the length of $p$, "d" the number of donor H-vertices and "a" the number of acceptors in hydrogen edges occurring in $p$.

**4.7. Further Definitions and Remarks.**

(a) Rings, chains, selfs and discretes are special H-edge patterns of the array $\Gamma$, and the symbols $L : \mathbf{R}^{\mathrm{a}}_{\mathrm{d}}(n)$, $L : \mathbf{C}^{\mathrm{a}}_{\mathrm{d}}(n)$, $L : \mathbf{S}^{\mathrm{a}}_{\mathrm{d}}$, $L : \mathbf{D}^{\mathrm{a}}_{\mathrm{d}}(n)$ are called their ***descriptors*** (we refer to them also as the ***quantitative*** descriptors of the $G$-array because, in 6.6, we shall introduce *qualitative* descriptors ). If the type $L$ is known from the context, then the notations $\mathbf{R}^{\mathrm{a}}_{\mathrm{d}}(n)$, $\mathbf{C}^{\mathrm{a}}_{\mathrm{d}}(n)$, $\mathbf{S}^{\mathrm{a}}_{\mathrm{d}}$, $\mathbf{D}^{\mathrm{a}}_{\mathrm{d}}(n)$ are also used. Note that the definition of $n$ for a discrete differs (by one) from the definition in former graph set articles, e.g. [2], where $n$ denotes the number of vertices of the discrete. Mathematically it seems to be more convenient for $n$ always to be the length of the corresponding path (see 8.7). No difference occurs for cycles and infinite paths.

(b) Rings, chains, self and discretes, respectively, which have the same set of H-edges (and therefore the same descriptor) will be called H-***edge equivalent***. They differ only in their covalent parts. It is usual in Graph Set Analysis to consider rings etc. only up to H-edge equivalence. This is crucial for our consideration and will justify the use of the ($G$-)constructor graph (5.2), where all covalent parts will "disappear".

(c) Another equivalence relation which is used by us is $G$-equivalence ("crystallographic equivalence", see 3.2). Obviously, by compatibility of $G$ with the elements of a $G$-array, rings, chains, selfs or discretes, resp., which are $G$-equivalent also give the same descriptor. Again, in Graph Set Analysis one considers rings, chains etc. up to $G$-equivalence; this will justify our description of rings, chains etc. by qualitative descriptors (6.6).

(d) Now we combine both of the above equivalences and introduce the notion of "descriptor equivalence". Two rings, chains, selfs or discretes, resp., are called ***descriptor equivalent*** if their sets of H-edges are $G$-equivalent (thus, e.g., for two rings $\Gamma(p)$ and $\Gamma(p')$ we must have $(E(p) \cap E_B)^g = E(p') \cap E_B$ for some $g \in G$). In Graph Set Analysis, every ring, chain, self or discrete, resp., is considered as representative of its descriptor equivalence class. The corresponding descriptor characterizes the whole class since it is also the descriptor of every member of the class.

Because of (b) and (c), two descriptor equivalent rings have the same descriptor (in the sense of the definition in (a)), but the converse is not true. Therefore one might ask why we call such rings *descriptor* equivalent. However, later we shall introduce so-called qualitative descriptors with the property that one can associate the same qualitative descriptor to two rings if and only if they are descriptor equivalent. This analogously holds also for chains, selfs and discretes.

(e) Up to now we have not introduced the notion of a graph set. In chemical literature sometimes a ring, chain etc. is called a "graph set" (contrary to the mathematical notion of a set), sometimes the set of all these special hydrogen bond patterns forms the "graph set" of an array. Graph Set Analysis itself is understood as the determination and investigation of descriptors (also called graph set symbols). In [18, Ch. 4 (p. 145)] graph sets themselves are regarded as a kind of "descriptor" (like the organic chemist's empirical formula). Therefore, by (d), the following definition makes sense: a ***graph set*** of a $G$-array $\Gamma$ is any descriptor equivalence class of a ring, chain, self or discrete of $\Gamma$. The ***descriptor*** of a graph set is the descriptor of any of its representatives (in particular this also gives the type of a graph set).

After this precise definition one can return to the convention that rings, chains, selfs and discretes are also called graph sets (with the meaning that they are representatives of graph sets as just defined).

**4.8. Refined labellings.** As we shall see later the labelling of a $G$-array can be ambiguous for several reasons. For instance, while $G$-equivalent H-edges always have the same label (by definition of a $G$-array, cf. 3.3(3b)), the converse is not true in general, i.e., H-edges with the same label need not be $G$-equivalent. We therefore introduce the following notion.

A $G$-array $\Gamma = (V, E_B \cup E_C, \alpha, \beta)$ is called H-***edge label transitive*** (with respect to $G$) if

$$\beta(e_1) = \beta(e_2) \quad \text{implies} \quad e_1^g = e_2 \text{ for some } g \in G$$

for all hydrogen edges $e_1, e_2 \in E_B$. Notice that no conditions are given for the labels of covalent edges.

Obviously the H-edge labels in an H-edge label transitive $G$-array uniquely correspond to the $G$-orbits, which therefore provide a canonical labelling $\beta_G(e) := e^G$. We consider this in a slightly more general context.

Let $T$ be a subgroup of $G$. Then every action of $G$ can be restricted to an action of $T$, and in particular, every $G$-array is canonically also a $T$-array. The group $T$ allows us to refine the H-edge labelling of a given $G$-array as follows. Let $B_T := \mathrm{Orb}(T, E_B)$ and $C_T := C$. Then

$$\begin{aligned} \beta_T : \quad & E_B \to B_T : \quad e \mapsto e^T \\ & E_C \to C_T : \quad e \mapsto \beta(e) \end{aligned}$$

is called the ***T-orbit labelling of*** H-***edges***. We introduce the array

$$\Gamma_T := (V, E_B \cup E_C, \alpha, \beta_T),$$

which can easily be seen to be a $T$-array (with respect to the action of $T$ on $\Gamma$ given by $G$). Moreover, it is obvious that $\Gamma_T = (V, E_B \cup E_C, \alpha, \beta_T)$ is an H-edge label transitive $T$-array.

In particular we can consider the special case $T = G$. Thus $\Gamma_G = (V, E_B \cup E_C, \alpha, \beta_G)$ is a canonically defined H-edge label transitive $G$-array for any given $G$-array $\Gamma = (V, E_B \cup E_C, \alpha, \beta)$.

Therefore, without loss of generality, one always can assume — and we shall do so — that all $G$-arrays under consideration are H-edge label transitive with the $G$-orbit labelling $\beta_G$.

**4.9. Remark.** The mathematical essence of the above refinement of a labelling consists of the following. Let $(T, X)$ be an action of a group $T$ on a set $X$. Then the ***T-orbit labelling*** $\beta_T : X \to \mathrm{Orb}(T, X) : x \mapsto x^T$ is compatible and label transitive (with respect to $T$), i.e., we have

$$\beta_T(x) = \beta_T(x') \iff \exists\, g \in T : \; x' = x^g \; .$$

# 5 Constructor graphs

In this section we introduce a new level of abstraction. Since in Graph Set Analysis one is interested only in H-edge patterns, the structure of graph sets inside the molecules plays no role. We therefore introduce a graph whose vertices are the graph molecules (cf.

also 2.4). Since this loses the information about which end of an H-edge meets its H-vertex, a direction for H-edges is defined (via the M-incidence mapping in 5.1). This gives the notion of a $G$-constructor graph (5.2).

However, it can still happen that in a $G$-constructor graph there are several edges from a molecule $M$ (to other molecules) with the same label $\beta_G(e)$. This becomes an obstacle if we want to describe graph sets by label sequences. In order to exclude such ambiguous labellings we shall refine the H-edge labelling $\beta_T$ via a suitable subgroup $T$ of $G$. This leads to the notion of the $T$-constructor graph of a $G$-array $\Gamma$ (5.3) where this obstacle is finally overcome as shown in Proposition 5.7.

As a consequence it is now possible to describe paths in the $T$-constructor graph by directed label sequences (5.11). This description is unique up to $G$-equivalence, as will be shown in Proposition 5.13.

**5.1. M-incidence mapping.** Let $\Gamma = (V, E, \alpha, \beta)$ be an array. Every H-edge $e = \{u, v\} \in E_B$ can be naturally directed, namely from the donor H-vertex to the acceptor vertex; i.e. for $\alpha(u) \in \mathrm{H}$ we consider the ordered pair $\overset{\rightarrow}{e} = (u, v)$ (or $\overset{\leftarrow}{e} := (v, u)$).

Thus, to any H-edge $e = \{u, v\} \in E_B$ we can also assign the pair $(M, M')$ consisting of the two graph molecules which are incident with $e$ in such a way that the hydrogen vertex of $e$ belongs to $M$, i.e. for $\alpha(u) \in \mathrm{H}$, we choose $M, M' \in \mathcal{M}$ such that $u \in M$ and $v \in M'$.
This assignment defines a mapping,

$$J : E_B \to \mathcal{M} \times \mathcal{M} : \{u, v\} \mapsto (M, M'),$$

the ***molecule incidence***, for short ***M-incidence mapping*** of the array. Note that $J$ can be extended to the whole set $E$ by $J(\{u, v\}) := (M, M)$ for covalent (and thus intramolecular) edges with $u, v \in M$.

**5.2. The G-constructor graph.** Let $\Gamma = (V, E, \alpha, \beta)$ be a $G$-array. Consider the set $\mathrm{Orb}(G, \mathcal{M}) = \{M^G \mid M \in \mathcal{M}\}$ of all orbits of the action $(G,\mathcal{M})$ of $G$ on the set $\mathcal{M}$ of graph molecules (cf. 3.2) and let $\alpha_{\mathcal{M}} : \mathcal{M} \to \mathrm{Orb}(G,\mathcal{M}) : M \mapsto M^G$ be the $G$-orbit labelling (cf. 4.9), i.e., the mapping assigning to each graph molecule the $G$-orbit to which it belongs.

Given $\Gamma$, the labelled directed multigraph (cf. 8.6)

$$\Sigma(\Gamma_G) := (\mathcal{M}, E_B, J, \alpha_{\mathcal{M}}, \beta_G)$$

is called the ***(labelled) G-constructor graph*** of the $G$-array $\Gamma$.

Remarks:
(a) From the definitions it follows immediately that the induced action of $G$ on $\mathcal{M}$ is compatible with the $G$-constructor graph $\Sigma(\Gamma_G)$, i.e., for each $g \in G$ the mapping $\mathcal{M} \to \mathcal{M} : M \mapsto M^g$ is an automorphism of $\Sigma(\Gamma_G)$ (respecting the labels and $J$).
(b) Notice that $J$ provides a ***direction*** for H-edges: for $J(e) = (M, M')$ we say that $e$ is an ***edge from*** $M$ ***to*** $M'$ (***outgoing*** from $M$, ***incoming*** to $M'$). For intramolecular edges $e$ we have $J(e) = (M, M)$. (Then for technical reasons it is sometimes useful to use $\vec{e}$ as defined in 5.9 instead of $e$. The mapping $J$ is the same: $J(\vec{e}) := J(e)$ (analogously for $\beta_G$). We may identify $e$ with $\vec{e}$ in the constructor graph whenever necessary.)

**5.3. The T-constructor graph of a G-array.** Let $T$ be a subgroup of $G$ satisfying the following conditions:

(0) $T$ is a normal subgroup of $G$;

(1) no element $t \in T$, except those elements whose action on $E_B$ is the identity, has a fixed point in its induced action on $\mathcal{M}$ (for short, we call such actions H-***edge semiregular***[6]);

(2) the factor group $G/T$ is finite.

[6] An action $(T,\mathcal{M})$ is called *semiregular* if the existence of a fixed point for $t \in T$ implies that $t$ acts identically. Here we have a stronger condition: $t$ has to act identically on the set $E_B$ of H-edges. However, for those groups $T$ which usually will be considered for $G$-arrays of crystals, both concepts, semiregularity and H-edge semiregularity, coincide.

*$G$ and $T$ and the $G$-array $\Gamma$ will be fixed throughout the following.*

The (directed) multigraph

$$\Sigma(\Gamma_T) = (\mathcal{M}, E_B, J, \alpha_{\mathcal{M}}, \beta_T)$$

(together with the action of $G$ and therefore also of $T$) is called the ***T*-constructor graph** or simply the ***constructor graph of the G*-array $\Gamma$**.

**5.4. Motivation and Remarks.**

a) For chemical applications $G$ is just the symmetry group (space group) of the crystal (cf. 3.4(b,c)) and, for simplicity, only two choices for $T$ might be considered: either $T$ is chosen equal to $G$ (namely if the action $(G, \mathcal{M})$ is already H-edge semiregular as in the example of anthranilic acid, cf. 2.4) or $T$ is the normal subgroup of all translations in $G$ (which is known to satisfy 5.3(0)-(2)).

b) Since every $G$-orbit splits into $T$-orbits, the mapping $\beta_T$ refines the edge labelling $\beta_G$ (cf. 4.8, 5.2) of the $G$-constructor graph. In particular, for every $\beta_T$-label $\beta_T(e)$ we can consider its "$\beta_G$-label" defined by $\beta_G(\beta_T(e)) := \beta_G(e)$ (this is independent of the choice of $e$).

c) Condition 5.3(0) guarantees that there is a natural action $(G, B_T)$ on the set $B_T$ (cf. 4.8) of $T$-orbits, defined by

$$\beta_T(e)^g := \beta_T(e^g) \quad \text{for } g \in G\,.$$

Recall that we have $\beta_T(e) = e^T$ by definition, where $e^T$ is the $T$-orbit of $e$. Therefore, in order to prove that the above definition is correct, we have to show that $\beta_T(e^g)$ does not depend on the choice of $e$ in its $T$-orbit. Thus let $e_1, e_2$ be edges which belong to the same $T$-orbit, i.e., $e_1^T = e_2^T$. Then $\beta_T(e_1^g) = (e_1^g)^T = e_1^{(gT)} =_* e_1^{(Tg)} = (e_1^T)^g = (e_2^T)^g = e_2^{(Tg)} =_* e_2^{(gT)} = (e_2^g)^T = \beta_T(e_2^g)$, as required. Note that we had to use the normality property 5.3(0) twice (at places marked with $=_*$).

The kernel of this action of $G$ on $B_T$ contains $T$ (since every $t \in T$ acts identically on $B_T$). Therefore there is also an action $(G/T, B_T)$ which, for $[g] \in G/T, [g] = Tg = gT$, is defined by $\beta_T(e)^{[g]} = \beta_T(e^g)$. It is evident that this action of $[g]$ on $T$-orbits coincides with the action of $g$ on $B_T$ defined above.

d) Condition 5.3(1) is a sufficient and (almost) necessary condition that edges in the $T$-constructor graph are uniquely determined by an incident vertex, the direction and the $\beta_T$-label. Sufficiency will be shown in Proposition 5.7 below.

Necessity holds under the assumption that the $G$-array is connected (as is always implicitly assumed for real crystal structures) and this can be seen as follows: if the action of $T$ on $\mathcal{M}$ is not semiregular, then there exists a non-trivial $g \in T$ and a fixed point $M^g = M \in \mathcal{M}$. Because $\Gamma$ is connected and $g$ is non-trivial, there must be at least one H-edge $e \in E_B$, say $J(e) = (M', M'')$, with $e^g \neq e$ such that at least one of the incident molecules $M'$ or $M''$ is also a fixed point. Consequently, $e$ and $e^g$ are two different edges with the same $\beta_T$-label either both outgoing from $M'$ or both incoming to $M''$.

Finally we mention that later in 5.11 the reader will see that this condition can be used in practice for the description of paths in the constructor graph by directed label sequences.

e) Recall that a chain in a $G$-array is a periodic infinite subgraph, whose period is a finite subgraph determined by the orbits of a particular $g \in G$ (4.4, 8.3). Thus condition 5.3(2) will guarantee that if there exists a finite period in a $G$-array then there exists also a finite period (of perhaps larger size) in the corresponding $T$-array (i.e., chains in the $G$-array become chains in the sense of definition 6.2 in the $T$-constructor graph).

**5.5. T-constructor graphs.** In what follows we shall work mainly with the $T$-constructor graph of a $G$-array. However it turns out that for the purposes of Graph Set Analysis, $T$-constructor graphs can be used independently of the underlying $G$-array. More precisely we mean that the data listed below may be collected from the definition of a $T$-constructor graph so that we get the slightly more abstract notion of a $(T, G)$-constructor graph (the $G$-array "disappears" in the definition).

A $T$-constructor graph consists of three data: a *multigraph* $(V, E, J)$ (cf. 8.6), an *action* of a group $G$ on $V$ as well as on $E$ (compatible with $J$, i.e., $J(e) = (u, v) \implies J(e^g) = (u^g, v^g)$) and an H-edge semiregular normal *subgroup* $T$ of $G$ with finite factor group $G/T$ (this corresponds to conditions 5.3(0)-(2)). Here H-***edge semiregular*** means that whenever $g \in G$ has a fixed point in $V$ then $g$ acts identically on the set $E$ of edges (traditionally called H-edges).

Then the ***T-constructor graph*** (more precisely: ***(T, G)-constructor graph***) is the labelled multigraph $\Sigma = (V, E, J, \alpha_V, \beta_T)$, where $\alpha_V : v \mapsto v^G$ is the $G$-orbit labelling of vertices and $\beta_T : e \mapsto e^T$ is the $T$-orbit labelling of edges (cf. 4.9).

It is obvious that the $T$-constructor graph $\Sigma(\Gamma_T)$ of a $G$-array $\Gamma$ is a $(T, G)$-constructor graph in the sense of this definition.

**5.6. Remark.** One should note that up to now all labels used are just "named" by orbits. For practical applications the following question arises: what is the most convenient way to select names for such labels? Of course, we might identify the name with one of the representatives of the orbit; however with this we lose uniqueness, and other possible representatives will be "discriminated" against because they could be used as well. In this paper we just once (in Table 1 on page 17) used a "dictionary" saying which symbol is assigned to which orbit as its "name". As pointed out already in the remark in 2.3, it is evident that the following principle must be satisfied by such a dictionary. If $\beta^*(e)$ denotes the "name" of the orbit $\beta_T(e)$ then we have to ensure the property $\beta^*(e) = \beta^*(e') \iff \beta_T(e) = \beta_T(e')$. However, we shall use $e^T$ throughout the paper as "name" for the $T$-orbit.

A similiar practical question arises if we consider both $\beta_G$-labels and their refined $\beta_T$-labels. Then it is reasonable also to "refine" the names just by adding a number to each $T$-orbit of a given $G$-orbit, e.g., if b is the "name" of the orbit $e^G$, the different $T$-orbits in $e^G$ might get "names" b1, b2, b3, ... (as done e.g. in [12]).

The Property (°) in the next proposition was the goal for which the refined labelling $\beta_T$ was introduced. Corollary 5.8 is an immediate consequence of (°), and it shows that in the $T$-constructor graph the edges are uniquely determined by their vertices, their direction (i.e. by $J$) and by their $\beta_T$-labels.

**5.7. Proposition.** *Let $\Gamma$ be a $G$-array and $\Sigma(\Gamma_T)$ its constructor graph. Then:*

(°) *all edges in $\Sigma(\Gamma_T)$ outgoing from (incoming to, resp.) a vertex $M \in \mathcal{M}$ have pairwise different $\beta_T$-labels.*

PROOF. (a): Let $e_1, e_2 \in E_B$ with $J(e_1) = (M, M')$, $J(e_2) = (M, M'')$. Assume that $\beta_T(e_1) = \beta_T(e_2)$, i.e. $e_1^T = e_2^T$. Then there exists $t \in T$ with $e_1^t = e_2$, so (by compatibility with $J$, cf. 5.2(a)) we have $(M^t, {M'}^t) = (M, M'')$, and in particular $M^t = M$. Thus $t$ has a fixed point $M$ in its action on $\mathcal{M}$ and therefore $t$ acts as the identity (on edges) by 5.3(1), hence $e_1 = e_1^t = e_2$.
The proof is similar for $J(e_1) = (M', M)$, $J(e_2) = (M'', M)$. $\square$

**5.8. Corollary.** *With the notation of 5.7 we have for $e_1, e_2 \in E_B$:*

$$[\, J(e_1) = J(e_2) \;\; \textit{and} \;\; \beta_T(e_1) = \beta_T(e_2)\,] \iff e_1 = e_2\,.$$

**5.9. Notation.** Edges $M \stackrel{e}{—} M'$ of $\Sigma(\Gamma_T)$ with $J(e) = (M, M')$ and $\beta_T(e) = b$ in pictures from now on will also be denoted by

$$M \xrightarrow{b} M' \quad \text{or} \quad M' \xleftarrow{b} M\,.$$

Because of 5.8, this representation describes the edge $e \in E_B$ uniquely.

For intramolecular edges $e = \{u, v\}$ with H-vertex $u$ the directed edges $\vec{e} := (u, v)$ and $\overleftarrow{e} := (v, u)$ (formally introduced in 5.1) also allow us to prescribe a direction in $\Sigma(\Gamma_T)$: passing through the edge $M \xrightarrow{b} M'$ from left to right is understood to pass through the edge $e$ from the H-vertex to the acceptor vertex, while passing through $M' \xleftarrow{b} M$ from left to right would mean the opposite direction, i.e., we let $u \in M$, $v \in M'$ (but note that in the case of intramolecular edges we have $M = M'$).

In the remaining part of this section we extend this to the description of paths, with the aim of describing them uniquely by directed label sequences (this aim can be attained up to $G$-equivalence, see 5.13 below). We start with an example.

**5.10. Example.** Let $p = M_0 \overset{e_1}{\longrule} M_1 \overset{e_2}{\longrule} M_2 \overset{e_3}{\longrule} M_3$ be a path with edges directed according to $J(e_1) = (M_0, M_1)$, $J(e_2) = (M_2, M_1)$, $J(e_3) = (M_3, M_2)$ and let $\beta(e_i) = b_i$ for $i = 1, 2, 3$. Then we can also represent the path $p$ as (cf. 5.9)

$$p = (M_0 \xrightarrow{b_1} M_1 \xleftarrow{b_2} M_2 \xleftarrow{b_3} M_3)\,.$$

Note that $J$ automatically gives the direction for intermolecular H-edges. In the case of intramolecular H-edges $e_i$ one has to use $\overset{\rightarrow}{e}_1$, $\overset{\leftarrow}{e}_2$ or $\overset{\leftarrow}{e}_3$ (cf. 5.9 and Remark (b) in 5.2), however, our definition is of interest mainly for intermolecular H-edges, since in graph sets we shall not mix inter- and intramolecular edges).

To such a path we assign its *directed label sequence* defined by

$$\lambda(p) = (\overset{\rightarrow}{b}_1 \overset{\leftarrow}{b}_2 \overset{\leftarrow}{b}_3),$$

i.e., running through the path from $M_0$ to $M_3$, the arrows on the labels indicate in which direction (with or against the arrow) one travels along the edge with that label. This label sequence will be a very useful tool for describing and locating paths (or walks) in the constructor graph: the reader should have no difficulties in locating a path with directed label sequence e.g. $\overset{\rightarrow}{a}\overset{\rightarrow}{a}\overset{\leftarrow}{b}\overset{\rightarrow}{c}\overset{\rightarrow}{b}\overset{\leftarrow}{c}\overset{\leftarrow}{c}\overset{\leftarrow}{a}$ in the constructor graph in Figure 7 (e.g. using the edges $e_1, e_5, e_2, e_3, e_2', e_3, e_4, e_5''$ one after the other). Of course there may be several paths with the same directed label sequence; however, in 5.13 we shall see that these paths must then be $G$-equivalent.

We will now define such label sequences in full generality.

**5.11. Directed label sequence.** Consider a constructor graph $\Sigma = (\mathcal{M}, E, J, \alpha_{\mathcal{M}}, \beta_T)$ and a path $p$ with directed edges

$$p := (\, M_0 \overset{e_1}{\text{——}} M_1 \overset{e_2}{\text{——}} \dots \overset{e_n}{\text{——}} M_n \,)$$

(the direction is not indicated here notationally: it follows from $J$ for intermolecular edges, but must be given explicitly in the case of intramolecular H-edges, cf. 5.9, 5.10).

For $\beta_T(e_i) = b_i$, $i \in \{1, \dots, n\}$, let

$$\lambda_p(e_i) := \begin{cases} \overrightarrow{b}_i\,, & \text{if } \; e_i \text{ is an edge from } M_{i-1} \text{ to } M_i, \\ \overleftarrow{b}_i\,, & \text{if } \; e_i \text{ is an edge from } M_i \text{ to } M_{i-1}. \end{cases}$$

Then the ***directed $\boldsymbol{\beta_T}$-label sequence*** $\lambda_T(p)$ of $p$ is defined by

$$\lambda_T(p) := (\lambda_p(e_1)\lambda_p(e_2)\dots\lambda_p(e_n)).$$

For any set $S \subseteq E_B$ of loops in $\Sigma$ we define

$$\lambda_T(S) := \{\beta_T(e) \mid e \in S\}\,,$$

the set of all $\beta_T$-labels of the H-edges in $S$.

**5.12. Graph equivalence and $G$-equivalence.**

To every path a unique directed label sequence can be associated (according to 5.11). This is no longer true if we want to consider the graph $\Gamma(p)$ of a path $p$ (cf. 8.7). The graph of a path has several representations and therefore several directed label sequences can be associated to it. In the above example 5.10 the same graph $\Gamma(p)$ can be described by paths with the following directed label sequences: $(\overrightarrow{b}_1\overleftarrow{b}_2\overleftarrow{b}_3)$, $(\overrightarrow{b}_3\overrightarrow{b}_2\overleftarrow{b}_1)$. These label sequences will be called graph equivalent.

If e.g., in addition, $M_0 = M_3$ then the path is a cycle and the following label sequences are also equivalent to the two mentioned above: $(\overleftarrow{b}_2\overleftarrow{b}_3\overrightarrow{b}_1)$, $(\overleftarrow{b}_3\overrightarrow{b}_1\overleftarrow{b}_2)$, $(\overrightarrow{b}_2\overleftarrow{b}_1\overrightarrow{b}_3)$, $(\overleftarrow{b}_1\overrightarrow{b}_3\overrightarrow{b}_2)$.

Formally we define two directed label sequences, say $\sigma, \sigma'$, to be ***graph equivalent*** if there exist paths $p, p'$ such that $\sigma = \lambda_T(p)$, $\sigma' = \lambda_T(p')$ and $\Gamma(p) = \Gamma(p')$.

The action $(G, B_T)$ on $\beta_T$-labels described in 5.4(c) can obviously be extended to directed labels by $(\overrightarrow{b})^g := (\overrightarrow{b^g})$ , $(\overleftarrow{b})^g := (\overleftarrow{b^g})$ , and also to directed label sequences, e.g., $(\overrightarrow{b}_1 \overleftarrow{b}_2 \dots)^g := (\overrightarrow{b}_1^g \overleftarrow{b}_2^g \dots)$.

According to 3.2 two directed label sequences $\sigma, \sigma'$ are called ***G-equivalent*** if there exists $g \in G$ with $\sigma^g = \sigma'$.

Note that the subgroup $T$ acts trivially on $\beta_T$-labels (cf. 5.4c), i.e. $G$-equivalence is actually $G/T$-equivalence. Thus, if we are allowed to choose the case $G = T$ (cf. 5.4a) two directed label sequences are $G$-equivalent if and only if they are equal! Moreover, because $G/T$ is finite there are only finitely many (at most $|G/T|$) directed label sequences which are $G$-equivalent to each other.

Now we are ready to state that – up to $G$-equivalence – paths in the constructor graph are uniquely characterized by their directed label sequence.

**5.13. Proposition.** *Let $\Sigma(\Gamma_T) = (\mathcal{M}, E_B, J, \alpha_{\mathcal{M}}, \beta_T)$ be the constructor graph of a $G$-array $\Gamma$. For paths $p, p'$ in $\Sigma(\Gamma_T)$ we have:*

a) *$p$ and $p'$ are $G$-equivalent if and only if $\lambda_T(p)$ and $\lambda_T(p')$ are $G$-equivalent; moreover, in that case one can choose $g, g' \in G$ such that $p^g = p'$, $\lambda_T(p)^{g'} = \lambda_T(p')$ and $Tg = Tg'$ ;*

b) *there is $g \in G$ such that $\Gamma(p)^g = \Gamma(p')$ if and only if there is $g' \in G$ such that $\lambda_T(p)^{g'}$ and $\lambda_T(p')$ are graph equivalent* (cf. 5.12)*; moreover one can choose $g, g'$ satisfying $Tg = Tg'$.*
(For the proof see 8.9)

Remark: The definition of $\lambda_T(p)$ (5.11) and Proposition 5.13 are not restricted to finite paths and can easily be extended to infinite paths $p$ (cf. 8.7).

# 6 Qualitative descriptors

In this section we shall describe special H-edge patterns (rings, chains, selfs, discretes) in a $T$-constructor graph in full analogy to those in $G$-arrays. By the results in the previous section, with such constructor graph sets we can associate directed label sequences which give the so-called corresponding qualitative descriptor. Then, in 6.13, it is justified that there is a one-to-one correspondence between the graph sets of a $G$-array $\Gamma$ and the graph sets of the $T$-constructor graph $\Sigma(\Gamma_T)$.

Let $\Gamma = (V, E_B \cup E_C, \alpha, \beta)$ be an H-edge label transitive $G$-array (without loss of generality we can assume $\beta = \beta_G$, cf. 4.8), and let $\Sigma(\Gamma_T) = (\mathcal{M}, E, J, \alpha_{\mathcal{M}}, \beta_T)$ be its constructor graph. Instead of $\Sigma(\Gamma_T)$ an arbitrary $(T, G)$-constructor graph could also be considered.

**6.1. Constructor rings.** The graph of a simple cycle

$$r := (M_0 \overset{e_1}{\text{---}} M_1 \overset{e_2}{\text{---}} \dots \overset{e_{n-1}}{\text{---}} M_{n-1} \overset{e_n}{\text{---}} M_0)$$

in $\Sigma(\Gamma_T)$ of length $n \geq 2$ is called a ***constructor ring***. To each constructor ring $\Gamma(r)$ a symbol

$$\mathbf{R}(\lambda_T(r))$$

is associated.

**6.2. Constructor chains.** A ***constructor chain*** is the graph of an infinite simple $T$-periodic path $c$ in $\Sigma(\Gamma_T)$. To each constructor chain $\Gamma(c)$ we associate a symbol

$$\mathbf{C}(\lambda_T(p)),$$

where $p$ is a $T$-period (8.8) of $c$.

**6.3. Proposition.** *Let $\Sigma$ be a $(T, G)$-constructor graph. Then an infinite simple path $c$ in $\Sigma$ is $G$-periodic if and only if it is $T$-periodic* (cf. 8.8). *Moreover, the $T$-period length $\pi_T(c)$ is a multiple of the $G$-period length $\pi_G(c)$.* (Proof see 8.11)

**6.4. Constructor selfs.** A graph $\Gamma(S) := (\{M\}, S)$ consisting of a finite set $S = \{e_1, \ldots, e_n\}$ of loops (in $\Sigma(\Gamma_T)$) at a single vertex $M \in \mathcal{M}$ is called a ***constructor self*** if for each $e \in S$ every loop $e'$ with $\beta(e') = \beta(e)$ and $J(e') = (M, M)$ also belongs to $S$. To each constructor self $\Gamma(S)$ in $\Sigma(\Gamma_T)$ we associate the symbol

$$\mathbf{S}(\lambda_T(S)).$$

**6.5. Constructor discretes.** A ***constructor discrete*** is the graph of a simple path $p$ in $\Sigma(\Gamma_T)$ between two vertices $M, M' \in \mathcal{M}$ such that both $M$ and $M'$ are $\tau(p)$-ends (i.e. incident with exactly one edge whose type belongs to $\tau(p)$, cf. 4.6). To each constructor discrete $\Gamma(p)$, we associate the symbol

$$\mathbf{D}(\lambda_T(p)) .$$

**6.6. Constructor graph sets and qualitative descriptors.** Every $G$-equivalence class of constructor rings, chains, selfs or discretes, respectively, is called a ***constructor graph set*** of the constructor graph $\Sigma(\Gamma_T)$. The symbols $\mathbf{R}(\lambda_T(r))$, $\mathbf{C}(\lambda_T(p))$, $\mathbf{S}(\lambda_T(S))$, $\mathbf{D}(\lambda_T(p))$ are called ***qualitative descriptors*** of the underlying rings, chains, selfs and discretes, respectively, and also of the whole $G$-equivalence class.

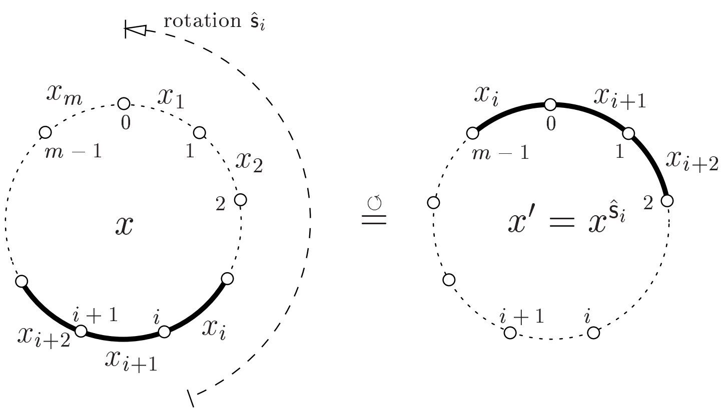
The symbols which can be associated with each ring, chain, etc. are by no means uniquely determined. As pointed out already in 5.12, a ring, chain, etc. may be described by different (graph equivalent) paths giving different directed label sequences. Moreover, if we consider different representatives of the same graph set (i.e. $G$-equivalence class), then we have still more possibilities to select a qualitative descriptor.

Nevertheless we want to describe and characterize graph sets in some sense *uniquely* by qualitative descriptors. We must therefore ask if a single qualititative descriptor can represent all elements of a constructor graph set, i.e., of a whole $G$-equivalence class. Moreover we have to ask how we can recognize whether two qualitative descriptors represent rings, chains, selfs or discretes which belong to the same graph set (this has practical applications if we want to give a repetition-free list of qualitative descriptors representing graph sets). The answers to these questions are given in the next Proposition 6.11. In preparation we still have to consider shifted and inverse sequences.

**6.7. Shifted sequences.** Let $x = (x_1 x_2 \dots x_m)$ be a word (sequence) of length $m$ over some alphabet $X$ (e.g., $x$ can be a path $p$ in a graph considered as sequence $p = (e_1 \dots e_m)$ of its edges, or $x$ can be a sequence $\sigma = (\sigma_1 \dots \sigma_m)$ of directed $\beta_T$-labels). Any word $x'$ of the form $x' = (x_{i+1} \dots x_m x_1 \dots x_i)$ is called a ***cyclic shift*** of $x$ ($i \in \{0, 1, \dots, m-1\}$). In this case we use the notation

$$x' \overset{\circlearrowleft}{=} x \quad \text{or (if } i \text{ is to be mentioned)} \quad x' = x^{\hat{\mathsf{s}}_i}\,.$$

If we represent $x$ as a clockwise labelled circle (with $x_1$ at the top) then $x'$ is just $x$ rotated anticlockwise by $i$ steps as shown in the following picture (and the above $\hat{\mathsf{s}}_i$ can be regarded as a permutation which performs this rotation):




**6.8. Inverse sequences.** Let $p$ be a finite path $(v_0, e_1, v_1, e_2, \dots, e_n, v_n)$ or an infinite path $(e_i, v_i)_{i \in \mathbb{Z}}$ in some (multi-)graph (cf. 8.7). Then the "reversed" path $(v_n, e_n, \dots, e_2, v_1, e_1, v_0)$ or $(e_{-i}, v_{-i-1})_{i \in \mathbb{Z}}$, respectively, will be denoted by $p^{-1}$ (i.e., we move through the path $p$ in the opposite direction).

Let $\sigma = (\sigma_1 \dots \sigma_m)$ be a sequence of directed $\beta_G$-labels (i.e., every $\sigma_i$ is $\vec{b}$ or $\overleftarrow{b}$ for some $b \in B_T$). Then $\sigma^{-1} := (\sigma_m^{-1} \dots \sigma_1^{-1})$ denotes the reverse sequence with reversed arrow directions, i.e., $\vec{b}^{-1} := \overleftarrow{b}$ and $\overleftarrow{b}^{-1} := \vec{b}$.

**6.9. Lemma.** *Let $\Delta$ be a qualitiative descriptor of $\gamma$ – a ring, chain, self or discrete, respectively. Let $\langle\Delta\rangle_{(T)}$ denote the set of all rings, chains, selfs or discretes, respectively, with which $\Delta$ can be associated as qualitative descriptor. Then $\langle\Delta\rangle_{(T)}$ is the $T$-equivalence class $\gamma^T = \{\gamma^t \mid t \in T\}$ to which $\gamma$ belongs.* (Proof see 8.13)

Thus two rings, chains, selfs or discretes, respectively, are $T$-equivalent if and only if there is a common qualitative descriptor which can be associated with both.

**6.10.** Since $\gamma^T \subseteq \gamma^G$, Lemma 6.9 ensures that any qualitative descriptor $\Delta$ represents a uniquely determined graph set, namely the graph set $\gamma^G = \{\gamma^g \mid g \in G\}$ to which $\gamma$ belongs. We shall denote this graph set also by $\langle\Delta\rangle^{\text{graph}}_{\text{set}}$ in the following proposition, where we shall clarify which descriptors represent the same graph set (here graph sets are always understood as constructor graph sets). If $\Delta$ is a qualitative descriptor $\mathbf{R}(\sigma)$, $\mathbf{C}(\sigma)$, $\mathbf{D}(\sigma)$, $\mathbf{S}(S)$, resp., then $\Delta^g$ will denote the qualitative descriptor $\mathbf{R}(\sigma^g)$, $\mathbf{C}(\sigma^g)$, $\mathbf{D}(\sigma^g)$, $\mathbf{S}(S^g)$, resp.

**6.11. Proposition.** (I) *Let $\Delta = \mathbf{R}(\sigma)$ and $\Delta' = \mathbf{R}(\sigma')$, or $\Delta = \mathbf{C}(\sigma)$ and $\Delta' = \mathbf{C}(\sigma')$, respectively, be two qualitative descriptors. Then they represent the same constructor graph set $\langle\Delta\rangle^{\text{graph}}_{\text{set}} = \langle\Delta'\rangle^{\text{graph}}_{\text{set}}$ if and only if $\sigma'$ is $G$-equivalent to a cyclic shift of $\sigma$ or $\sigma^{-1}$, i.e., there exists some $g \in G$ such that $\sigma' \overset{\circlearrowleft}{=} \sigma^g$ or $\sigma' \overset{\circlearrowleft}{=} (\sigma^{-1})^g$.*

(II) *Let $\Delta = \mathbf{D}(\sigma)$ and $\Delta' = \mathbf{D}(\sigma')$ be qualitative descriptors. Then they represent the same constructor graph set $\langle\Delta\rangle^{\text{graph}}_{\text{set}} = \langle\Delta'\rangle^{\text{graph}}_{\text{set}}$ if and only if $\sigma'$ is $G$-equivalent to $\sigma$ or $\sigma^{-1}$.*

(III) *Let $\Delta = \mathbf{S}(S)$ and $\Delta' = \mathbf{S}(S')$ be qualitative descriptors. Then they represent the same constructor graph set $\langle\Delta\rangle^{\text{graph}}_{\text{set}} = \langle\Delta'\rangle^{\text{graph}}_{\text{set}}$ if and only if $S$ and $S'$ are $G$-equivalent (i.e., there exists a $g \in G$ such that $S' = S^g$).*

(IV) *Let* $\Delta$ *be a qualititative descriptor and let* $G = g_1T \cup g_2T \cup \cdots \cup g_kT$ *be the partition of* $G$ *into* $T$*-cosets. Then we have:*

$$\langle \Delta \rangle^{\text{graph}}_{\text{set}} = \bigcup_{g \in G} \langle \Delta^g \rangle_{(T)} = \langle \Delta^{g_1} \rangle_{(T)} \cup \langle \Delta^{g_2} \rangle_{(T)} \cup \cdots \cup \langle \Delta^{g_k} \rangle_{(T)} .$$

(Proof see 8.16)

There is a correspondence between graph sets and constructor graph sets of a $G$-array $\Gamma$, induced by a mapping, here denoted by $\Xi$, which assigns to each path $p$ in a $G$-array $\Gamma$ a path $\Xi(p)$ in the multigraph $\Sigma(\Gamma_T)$, defined as follows:

**6.12. The mapping $\Xi$.** Let $\gamma = (V(\gamma), E(\gamma))$ be a subgraph (cf. 8.4) of the $G$-array $\Gamma$. Then we define

$$\Xi(\gamma) := (\mathcal{M}(\gamma), E_B(\gamma), J|_{E_B(\gamma)}) .$$

This is a sub(multi)graph of the constructor graph $\Sigma(\Gamma_T) = (\mathcal{M}, E_B, J, \alpha_{\mathcal{M}}, \beta_T)$; its vertex set $\mathcal{M}(\gamma)$ consists of all graph molecules that contain at least one vertex of $V(\gamma)$, $E_B(\gamma) := E(\gamma) \cap E_B$ is its edge set, and $J|_{E_B(\gamma)}$ is just the restriction of $J$ to $E_B(\gamma)$.

**6.13. First Justification Statement.** *Let* $\Gamma$ *be a (*H*-edge transitive)* $G$*-array.*

(a) *Let* $\gamma$ *be a ring, chain, self or discrete in* $\Gamma$. *Then* $\Xi(\gamma)$ *is respectively a constructor ring, chain, self or discrete in* $\Sigma(\Gamma_T)$.

(b) *For every constructor ring, chain, self or discrete in* $\Sigma(\Gamma_T)$, *say* $\delta$, *there exists a ring, chain, self or discrete* $\gamma$ *in* $\Gamma$, *respectively, such that* $\Xi(\gamma) = \delta$.

(c) *Let* $\gamma$ *and* $\gamma'$ *be rings, chains, selfs or discretes, respectively. Then*

(i) $\gamma$ *and* $\gamma'$ *are* H*-edge equivalent* (cf. 4.7b) *if and only if* $\Xi(\gamma) = \Xi(\gamma')$.

(ii) $\gamma$ *and* $\gamma'$ *are descriptor equivalent* (cf. 4.7d) *if and only if* $\Xi(\gamma)$ *and* $\Xi(\gamma')$ *are* $G$*-equivalent.*

(d) $\Xi$ *induces a one-to-one correspondence between the graph sets of* $\Gamma$ *and the constructor graph sets of* $\Sigma(\Gamma_T)$.

PROOF. In principle, the proof is straightforward and follows directly from the definitions. Nevetheless we give a few hints for checking each part of the statement.
For (a), first one can easily check that a path $p$ in the array $\Gamma$ is M-simple if and only if $\Xi(p)$ is simple. The remainder is routine comparison of the definitions.
For (b), note that $\Xi$ just shrinks every graph molecule $M$ of $\Gamma$ to the vertex $M$ of $\Sigma(\Gamma_T)$. The H-edges are not touched. Therefore, to see (b), for any vertex $M$ of $\delta$ (in $\Sigma(\Gamma_T)$) one has to choose a shortest covalent path in $\Gamma$ between the endpoints (again in $\Gamma$) of the incident H-edges.
(c) is just a reformulation of the definitions in 4.7(b,d), using $\Xi$.
(d) follows directly from c(ii); here the $\Xi$-image of a graph set is understood to be the constructor graph set (i.e. the $G$-equivalence class) of the $\Xi$-image of any representative. $\square$

At the end of this section we want to clarify which types of graph sets can appear simultaneously in a $G$-array. We formulate the answer for a $T$-constructor graph. But by 6.13, the following proposition also holds for an arbitrary (H-edge label transitive) $G$-array (since it holds for the corresponding constructor graph $\Sigma(\Gamma_T)$).

**6.14. Proposition.** *Let* $\Sigma = (\mathcal{M}, E_B, J, \alpha_{\mathcal{M}}, \beta_T)$ *be a* $(T, G)$*-constructor graph. Let* $L \subseteq \{\beta_G(e) \mid e \in E_B\}$ *be some non-empty type set (recall that type sets contain* $\beta_G$*-labels but not* $\beta_T$*-labels). Then exactly one of the following statements is true (and the others are false):*

I. *There exist rings or chains of type* $L$.

II. *There exist selfs of type* $L$.

III. *There exist discretes of type* $L$.

IV. *There are neither rings, chains, selfs nor discretes of type* $L$ *and* $|L| \geq 2$.

*In particular, every* H*-edge $e$ belongs to (a representative of) a graph set of the first level (i.e. of type $L = \{\beta_G(e)\}$).* (Proof see 8.10)

**6.15. Remark.** Proposition 6.14 says, in particular, that for $|L| = 1$ we always find a graph set of type $L$. The essence of (I) – (III) is that the existence of discretes, of selfs and of rings or chains of type $L$ excludes the other possibilities, which is not the case for the existence of rings and of chains: in fact, if we have a two-

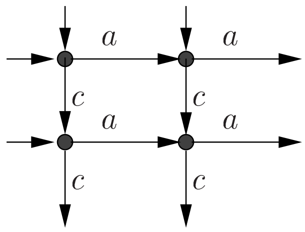



Figure 13: An array with rings and chains of the same type

dimensional periodic array with a constructor graph as in Fig. 13 then there exist both chains and rings of type $L = \{a, c\}$ (namely $\mathbf{C}(\vec{a}\,\vec{c})$ and $\mathbf{R}(\vec{a}\,\vec{c}\,\overleftarrow{a}\,\overleftarrow{c})$).

# 7 Computation of quantitative descriptors

Descriptors of a $G$-array $\Gamma$ can be computed from concretely given rings, chains, selfs or discretes. If we climb up one level of abstraction we get constructor rings, chains, selfs and discretes (the $\Xi$-images). Although different rings, chains, etc. may have the same $\Xi$-image, the descriptors must be uniquely determined by the $\Xi$-image (by 6.13) which in turn is characterizable by a qualitative descriptor.

There arises the question of whether the (quantitative) descriptors of a $G$-array can be computed directly from the qualitative descriptors of the $T$-constructor graph. The $T$-constructor graph was invented in order to get an affirmative answer to this question. In this section we shall present explicit formulas for these computations. Example 7.11 may help the reader to get a better understanding of the formal treatment. The section closes with some concluding remarks.

However, we have not told the whole story. In the $T$-constructor graph $\Sigma(\Gamma_T)$ all information about the inner structure of the molecules is lost. But some numerical information about the lengths of covalent paths between H-edges must be stored. This is the reason why every $T$-constructor graph must be accompanied by the so-called covalent distance table. Thus, in order to calculate quantitative descriptors, two data must be provided: a $T$-constructor graph and a covalent distance table.

In preparation we need some more notions. At first we introduce a "covalent distance" for two vertices. This notion can be extended to pairs of directed $\beta_T$-labels, which will constitute the covalent distance table.

**7.1. Covalent distances for vertices.** For a given array $\Gamma = (V, E, \alpha, \beta)$ and two vertices $u, v \in V$, let $d(u, v)$ denote the ***(covalent) distance*** from $u$ to $v$ (inside a graph molecule), i.e. the length of a shortest intramolecular path from $u$ to $v$. If there is no such intramolecular path from $u$ to $v$, i.e. if $u$ and $v$ belong to different graph molecules, we put $d(u, v) = \infty$. In a $G$-array $\Gamma$ we obviously have $d(u, v) = d(u^g, v^g)$ for every $g \in G$ (cf. 3.3), i.e., the covalent distance is invariant under the action of $G$.

**7.2. Covalent distances for labels.** Let $\Gamma$ be a $G$-array and let $\Sigma(\Gamma_T)$ be its constructor graph. Further, let $p = (M \overset{e}{\text{---}} M' \overset{e'}{\text{---}} M'')$ be a directed walk (of length 2) in $\Sigma(\Gamma_T)$ with $\beta_T(e) = b$, $\beta_T(e') = b'$, $e = \{u, v\}$, $e' = \{u', v'\}$, $u \in M$, $v, u' \in M'$ and $v' \in M''$. Then the covalent distance of the two vertices $v, u'$ which belong to the middle molecule $M'$ will give the ***covalent distance*** $\mathsf{cd}(\lambda_T(p))$ of the pair of directed $\beta_T$-labels which is given by $\lambda_T(p)$ (cf. 5.11) as follows (for notation see also 5.9, 7.1):

$$\mathsf{cd}(\lambda_T(p)) = \left\{ \begin{array}{c} \mathsf{cd}(\vec{b}\,\vec{b}') \\ \mathsf{cd}(\vec{b}\,\overleftarrow{b}') \\ \mathsf{cd}(\overleftarrow{b}\,\vec{b}') \\ \mathsf{cd}(\overleftarrow{b}\,\overleftarrow{b}') \end{array} \right\} := d(v,u') \left\{ \begin{array}{l} \text{if } p = (M \xrightarrow{b} M' \xrightarrow{b'} M''), \\ \text{if } p = (M \xrightarrow{b} M' \xleftarrow{b'} M''), \\ \text{if } p = (M \xleftarrow{b} M' \xrightarrow{b'} M''), \\ \text{if } p = (M \xleftarrow{b} M' \xleftarrow{b'} M''). \end{array} \right.$$

Note that the argument of $\mathsf{cd}$ in these defining formulas is always just $\lambda_T(p)$. To all pairs of labels (of intermolecular or intramolecular H-edges) which cannot be represented as above (namely as $\lambda_T(p)$ for some walk $p = (ee')$) the value $\infty$ is assigned as covalent distance.

With the last remark in 7.1 one can show that this definition does not depend on the choice of the path with H-edges $e, e'$; in fact, different paths with the same directed label sequence must be $G$-equivalent by 5.13.

**7.3. Covalent distance table.** Given $\Gamma$ and $\Sigma(\Gamma_T)$ we now consider the set $\overline{B}_T$ of all directed $\beta_T$-labels of H-edges (recall that $B_T = \mathrm{Orb}(T, E_B)$):

$$\vec{B}_T := \{\,\vec{b} \mid b \in B_T\}, \quad \overleftarrow{B}_T := \{\,\overleftarrow{b} \mid b \in B_T\}, \quad \overline{B}_T := \vec{B}_T \cup \overleftarrow{B}_T.$$

The matrix

$$\mathsf{Cd}(\Gamma) := \Big(\mathsf{cd}(\mu\nu)\Big)_{\mu,\nu \in \overline{B}_T}$$

is called the ***covalent distance table*** of the $G$-array $\Gamma$ with respect to the $\beta_T$-labelling.

<u>Remarks:</u> The values in $\mathsf{Cd}(\Gamma)$ are highly redundant. Note that different representatives of the same graph $\Gamma(p) = (M \overset{e}{\text{—}} M' \overset{e'}{\text{—}} M'')$ give graph equivalent (cf. 5.12) label sequences. Thus we have for arbitrary $b, b' \in B_T$:

$$\mathsf{cd}(\overleftarrow{b}\,\vec{b}') = \mathsf{cd}(\overleftarrow{b}'\,\vec{b}), \quad \mathsf{cd}(\overleftarrow{b}\,\overleftarrow{b}') = \mathsf{cd}(\vec{b}'\,\vec{b}), \quad \mathsf{cd}(\vec{b}\,\overleftarrow{b}') = \mathsf{cd}(\vec{b}'\,\overleftarrow{b}). \tag{a}$$

Moreover, if $b = b'$ and both edges $e, e'$ are either outgoing from $M'$ or incoming to $M'$ then we must have $e = e'$ by 5.7(°). Therefore passing through such a path we have to enter and to leave the molecule $M'$ on the same H-edge, i.e., we have trivially

$$\mathsf{cd}(\overrightarrow{b}\overleftarrow{b}) = \mathsf{cd}(\overleftarrow{b}\overrightarrow{b}) = 0\,. \tag{b}$$

Finally we get from 7.1

$$\mathsf{cd}(\sigma_1^g \sigma_2^g) = \mathsf{cd}(\sigma_1\sigma_2) \tag{c}$$

for arbitrary directed labels $\sigma_1, \sigma_2 \in \overline{B}_T$ and $g \in G$.

The covalent distance table of $\Gamma$ can be split into smaller tables, the covalent distance tables $\mathsf{Cd}(M)$ of the molecules, taking into account only H-edge labels of H-edges which are incident with $M$. Moreover, entries where $\mu$ and $\nu$ are labels of an intermolecular and an intramolecular H-edge could be omitted since we shall not use them further.

**7.4. Acceptor and donor** H-**vertex equalities.** While 7.3(b) trivially holds, the condition $\mathsf{cd}(\overrightarrow{b}\overleftarrow{b}') = 0$ or $\mathsf{cd}(\overleftarrow{b}\overrightarrow{b}') = 0$, resp., gives essential information for different $b, b'$: it means for a path $p = (M \xrightarrow{b} M' \xleftarrow{b'} M'')$ or $(M \xleftarrow{b} M' \xrightarrow{b'} M'')$, that the two edges in $p$ have the same acceptor vertex or donor H-vertex, resp. This motivates the following definition.

Let $Q$ be a set of $\beta_T$-labels which are the labels of loops at a single vertex $M \in \mathcal{M}$ in $\Sigma(\Gamma_T)$. We observe that $\mathsf{cd}(\overrightarrow{b}\overleftarrow{b}') = 0$ and $\mathsf{cd}(\overrightarrow{b}'\overleftarrow{b}'') = 0$ imply that $\mathsf{cd}(\overrightarrow{b}\overleftarrow{b}'') = 0$ (analogously for $\mathsf{cd}(\overleftarrow{b}\overrightarrow{b}') = 0$ etc.), where $b, b', b'' \in Q$. This allows us to define the following equivalence relations on $Q$:

$$\begin{aligned} b \overset{\mathrm{a}}{\sim} b' \quad &:\Longleftrightarrow \quad \mathsf{cd}(\overrightarrow{b}\overleftarrow{b}'') = 0 \text{ (“acceptor equality”)} \\ b \overset{\mathrm{d}}{\sim} b' \quad &:\Longleftrightarrow \quad \mathsf{cd}(\overleftarrow{b}\overrightarrow{b}'') = 0 \text{ (“donor equality”).} \end{aligned}$$

Let $\mathrm{a}(Q)$ and $\mathrm{d}(Q)$, denote the number of equivalence classes of $\overset{\mathrm{a}}{\sim}$ and $\overset{\mathrm{d}}{\sim}$, resp.

**7.5. The splitting factor for chains.** Let $\Sigma$ be a $(T, G)$-constructor graph and $c$ a $G$-periodic path. Then (for notation see 4.4 and 8.8)

$$f(c) := \frac{\pi_T(c)}{\pi_G(c)}$$

is called the ***splitting factor*** of $c$ (it is well defined because of 6.3).

Remark: If $\Gamma(c) = \Gamma(c')$ for different paths $c, c'$, then $c' = c^{\mathsf{s}_n}$ or $c' = (c^{\mathsf{s}_n})^{-1}$ for some $n$ ($\mathsf{s}_n$ shifts paths by $n$ steps, for $n \in \mathbb{Z}$; for notation see 8.7, 8.8, 6.8), and thus $f(c) = f(c')$. Therefore $f(c)$ is also called the splitting factor of the corresponding chain $\Gamma(c)$.

The splitting factor is just the number of $G$-periods into which every $T$-period of $c$ splits. This is the reason why the numbers in 7.9(ii) below must be divided by $f(c)$ (these numbers belong to the quantitative descriptor of a chain computed from a $T$-period). Because we want to deal with graph sets (i.e. $G$-equivalence classes) we have to ask whether $f(c)$ is independent of the choice of $c$ as representative of a graph set. Moreover, graph sets are characterized by qualitative descriptors and therefore there arises the question of whether and how it is possible to compute $f(c)$ from a qualitative descriptor of $c$. These questions are answered in the following. We present here an easy practicable way to compute the splitting factor. All mathematical details and proofs can be found in the appendix (Section 8, Part C).

**7.6. The splitting factor for label sequences.** Let $g \in G$ and let $\sigma = (\sigma_1 \sigma_2 \ldots \sigma_m)$ be a sequence of directed $\beta_T$-labels. If $\sigma^g \overset{\circ}{=} \sigma$ then define $o(\sigma, g)$ (the ***order*** of $g$ with respect to $\sigma$) to be the least positive natural number $n$ such that $\sigma^{g^n} = \sigma$. Otherwise we put $o(\sigma, g) := 0$. Let $g_1, \ldots, g_k$ be coset representatives of $G/T$: $G = Tg_1 \cup Tg_2 \cup \cdots \cup Tg_k$, $k := |G/T|$. Then the number

$$f(\sigma) := \max\{o(\sigma, g_1), \ldots, o(\sigma, g_k)\}$$

is called the ***splitting factor*** of $\sigma$.

Remark: This definition is efficient for computing $f(\sigma)$ but mathematically it looks "strange". Where it comes from? There is an action of $G$ on the $\beta_T$-labels (see 5.4c), and hence an induced action of $G/T$ on these labels. For any given $\sigma$, the elements $g \in G$ such that $\sigma^g \overset{\circlearrowleft}{=} \sigma$ form a subgroup $G(\sigma)$ of $G$ containing $T$. If $g \notin G(\sigma)$ then $\sigma^g \overset{\circlearrowleft}{\neq} \sigma$, and we have defined $o(\sigma, g) = 0$. If $g \in G(\sigma)$ then $g$ induces a rotation $\hat{\mathsf{s}}_n$ of $\sigma$, and $o(\sigma, g)$ is the order of this rotation (which can be computed also via $o(\sigma, g) = \frac{m}{\gcd(n,m)}$, cf. 8.19 and 8.20). Since $T$ acts trivially on $\beta_T$-labels, we can regard $o(\sigma, \cdot)$ as defined on the cosets of $T$. Now $G(\sigma)$ (and hence $G(\sigma)/T$) induces a group $C(\sigma)$ of rotations of $\sigma$; $C(\sigma)$ must be cyclic, so the order $f(\sigma)$ of this group is equal to the maximum order of its elements. Thus this order is the splitting factor $f(\sigma)$ as defined above in 7.6 (since $o(\sigma, g) = 0$ for all $g \notin G(\sigma)$). In particular, the splitting factor $f(\sigma)$ is independent of the choice of the coset representatives $g_1, \ldots, g_k$. Notice, moreover, that we have $1 \leq f(\sigma)$, and every non-zero $o(\sigma, g_i)$ divides $f(\sigma)$.

**7.7. Proposition.** *Let $c$ be a $G$-periodic path (in a $(T,G)$-constructor graph), $p$ a $T$-period of $c$ and $\sigma := \lambda_T(p)$. Then we have*

$$f(c) = f(\sigma)\,. \qquad \text{(Proof see 8.21)}$$

This proposition enables us to compute the splitting factor directly from a qualitative descriptor. Moreover, since $o(\sigma^{-1}, g) = o(\sigma, g)$ and $o(\sigma^{\hat{\mathsf{s}}_i}, g) = o(\sigma, g)$ (cf. 6.7, 6.8) we get from 6.11(I) the following:

**7.8. Corollary.** *Let $\Gamma(c)$ and $\Gamma(c')$ be $G$-equivalent chains in a $(T,G)$-constructor graph. Then $f(c) = f(c')$.* □

**7.9. Descriptor computation.** To every qualitative descriptor of a $T$-constructor graph $\Sigma(\Gamma_T)$ we associate in the following a (quantitative, 4.7) descriptor of the $G$-array $\Gamma = (V, E_B \cup E_C, \alpha, \beta_G)$ using only the covalent distance table $\mathsf{Cd}(\Gamma)$. Since $\beta_T$ is a refinement of $\beta_G$, for any $\beta_T$-label $b \in B_T$ (directed or not), we can define the values $\beta_G(b)$, $\beta_G(\vec{b})$ and $\beta_G(\overleftarrow{b})$ all to be equal to $\beta_G(e)$ for some edge $e$ with $b = \beta_T(e)$ (cf. 5.4b).

(i) **Rings:** Let $\mathbf{R}(\sigma)$ be a qualitative descriptor for a sequence $\sigma = (\sigma_1\sigma_2 \dots \sigma_m)$ of directed $\beta_T$-labels. We define the corresponding descriptor $\Psi(\mathbf{R}(\sigma))$ of $\Gamma$ as follows:

$$\Psi(\mathbf{R}(\sigma)) := L : \mathbf{R}_{\mathrm{d}}^{\mathrm{a}}(n)\,,$$

where

$$\begin{aligned}
L &= \{\beta_G(\sigma_i) \mid i = 1, \dots, m\}\,,\\
n &= \mathsf{cd}(\sigma_1\sigma_2) + \cdots + \mathsf{cd}(\sigma_{m-1}\sigma_m) + \mathsf{cd}(\sigma_m\sigma_1) + m\,,\\
\mathrm{a} &= m - |\{i \in \{1, \dots, m\} \mid (\sigma_i\sigma_{i+1}) = (\vec{b}\,\overleftarrow{b'}) \text{ and } \mathsf{cd}(\vec{b}\,\overleftarrow{b'}) = 0\}|\,,\\
\mathrm{d} &= m - |\{i \in \{1, \dots, m\} \mid (\sigma_i\sigma_{i+1}) = (\overleftarrow{b}\,\vec{b'}) \text{ and } \mathsf{cd}(\overleftarrow{b}\,\vec{b'}) = 0\}|
\end{aligned}$$

with the convention that $\sigma_{m+1} := \sigma_1$. Here, for short, e.g. $(\sigma_i\sigma_{i+1}) = (\vec{b}\,\overleftarrow{b'})$ stands for $\exists b, b' : (\sigma_i\sigma_{i+1}) = (\vec{b}\,\overleftarrow{b'})$.

(ii) **Chains:** Let $\mathbf{C}(\sigma)$ be a qualitative descriptor for a sequence $\sigma = (\sigma_1\sigma_2 \dots \sigma_m)$ of directed $\beta_T$-labels. We define the corresponding descriptor $\Psi(\mathbf{C}(\sigma))$ of $\Gamma$ as follows:

$$\Psi(\mathbf{C}(\sigma)) := L : \mathbf{C}_{\mathrm{d}}^{\mathrm{a}}(n)\,,$$

where

$$\begin{aligned}
L &= \{\beta_G(\sigma_i) \mid i = 1, \dots, m\}\,,\\
n &= \frac{1}{f(\sigma)} \cdot \Big(\mathsf{cd}(\sigma_1\sigma_2) + \cdots + \mathsf{cd}(\sigma_{m-1}\sigma_m) + \mathsf{cd}(\sigma_m\sigma_1) + m\Big)\,,\\
\mathrm{a} &= \frac{1}{f(\sigma)} \cdot \Big(m - \\
&\qquad |\{i \in \{1, \dots, m\} \mid (\sigma_i\sigma_{i+1}) = (\vec{b}\,\overleftarrow{b'}) \text{ and } \mathsf{cd}(\vec{b}\,\overleftarrow{b'}) = 0\}|\Big)\,,\\
\mathrm{d} &= \frac{1}{f(\sigma)} \cdot \Big(m - \\
&\qquad |\{i \in \{1, \dots, m\} \mid (\sigma_i\sigma_{i+1}) = (\overleftarrow{b}\,\vec{b'}) \text{ and } \mathsf{cd}(\overleftarrow{b}\,\vec{b'}) = 0\}|\Big)\,.
\end{aligned}$$

$f(\sigma)$ denotes the splitting factor (cf. 7.6).

(iii) **Discretes:** Let $\mathbf{D}(\sigma)$ be a qualitative descriptor for a sequence $\sigma = (\sigma_1\sigma_2 \ldots \sigma_m)$ of directed $\beta_T$-labels. We define the corresponding descriptor $\Psi(\mathbf{D}(\sigma))$ of $\Gamma$ as follows:

$$\Psi(\mathbf{D}(\sigma)) := L :\ \mathbf{D}^{\mathrm{a}}_{\mathrm{d}}(n)$$

where

$$\begin{aligned}
L &= \{\beta_G(\sigma_i) \mid i = 1, \ldots, m\}\,,\\
n &= \mathsf{cd}(\sigma_1\sigma_2) + \cdots + \mathsf{cd}(\sigma_{m-1}\sigma_m) + m\,,\\
\mathrm{a} &= m - \\
&\quad |\{i \in \{1, \ldots, m-1\} \mid (\sigma_i\sigma_{i+1}) = (\vec{b}\,\overleftarrow{b'}) \text{ and } \mathsf{cd}(\vec{b}\,\overleftarrow{b'}) = 0\}|\,,\\
\mathrm{d} &= m - \\
&\quad |\{i \in \{1, \ldots, m-1\} \mid (\sigma_i\sigma_{i+1}) = (\overleftarrow{b}\,\vec{b'}) \text{ and } \mathsf{cd}(\overleftarrow{b}\,\vec{b'}) = 0\}|\,.
\end{aligned}$$

(iv) **Selfs:** Let $\mathbf{S}(F)$ be a qualitative descriptor for a set $F$ of (undirected) $\beta_T$-labels. We define the corresponding descriptor $\Psi(\mathbf{S}(F))$ of $\Gamma$ as follows:

$$\Psi(\mathbf{S}(F)) := L :\ \mathbf{S}^{\mathrm{a}}_{\mathrm{d}} \quad \text{or} \quad L :\ \mathbf{S}^{\mathrm{a}}_{\mathrm{d}}(n)\,,$$

where (cf. 7.4)

$$\begin{aligned}
L &= \{\beta_G(b) \mid b \in F\}\,,\\
\mathrm{a} &= \mathrm{a}(F)\,,\\
\mathrm{d} &= \mathrm{d}(F)\,.
\end{aligned}$$

In the case of a one-element set $L = \{b\}$ we define

$$n = \mathsf{cd}(\vec{b}\,\vec{b}) + 1\,.$$

**7.10. Second Justification Statement.** *The set of the $\Psi$-images of all qualitative descriptors of a $G$-array $\Gamma$ equals the set of all quantitative descriptors of $\Gamma$.*
*More precisely: for every ring, chain, self or discrete $\gamma$ of a given $G$-array $\Gamma$ we have*

$$\Psi(\Delta_\Sigma(\Xi(\gamma))) = \Delta_\Gamma(\gamma)\,,$$

*where $\Delta_\Sigma$ denotes a qualitative descriptor and $\Delta_\Gamma$ denotes the quantitative descriptor of its argument.*

Proof. The proof is straightforward (checking the definitions). For chains we mention that every $T$-period of a chain $\mathbf{C}(\sigma)$ (in the constructor graph) splits into $f(\sigma)$ $G$-periods (see 6.3, 7.5, 7.7). In case of selfs one should notice that all loops at a vertex $M$ of the constructor graph have different $\beta_T$-labels by 5.7 (this guarantees that the numbers “a” and “d” cannot decrease under the mapping $\Xi$). $\square$

Remark: Contrary to the result in 7.10, the length and period length of paths and chains are not preserved under the mapping $\Xi$ (therefore we had to use the splitting factor).

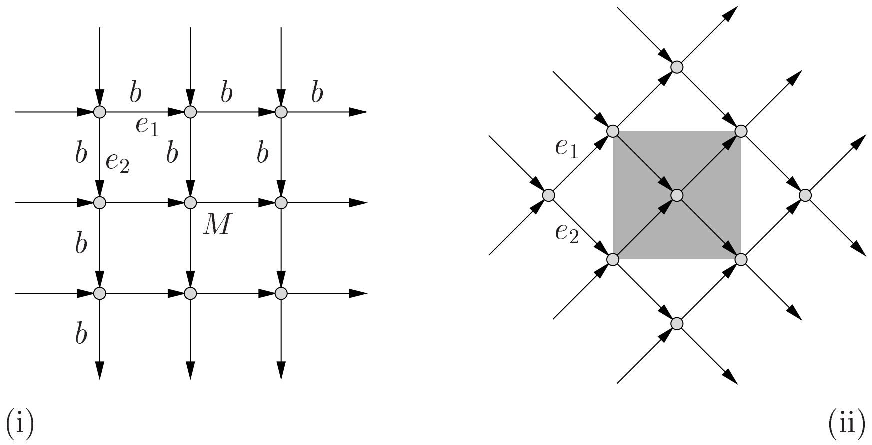
**7.11. Example.** For simplicity we treat here a hypothetical two-dimensional structure. Let us consider a $G$-array $\Gamma$ with $G$-constructor graph (cf. 5.2) as given in Figure 14(i).

In this example, the symmetry group $G$ can be generated by three operations: $G = \langle g_1, g_2, g_3 \rangle$ where $g_1$ and $g_2$ are horizontal and vertical translations, and $g_3$ is the reflection in a diagonal axis (see Fig. 14(iii)).

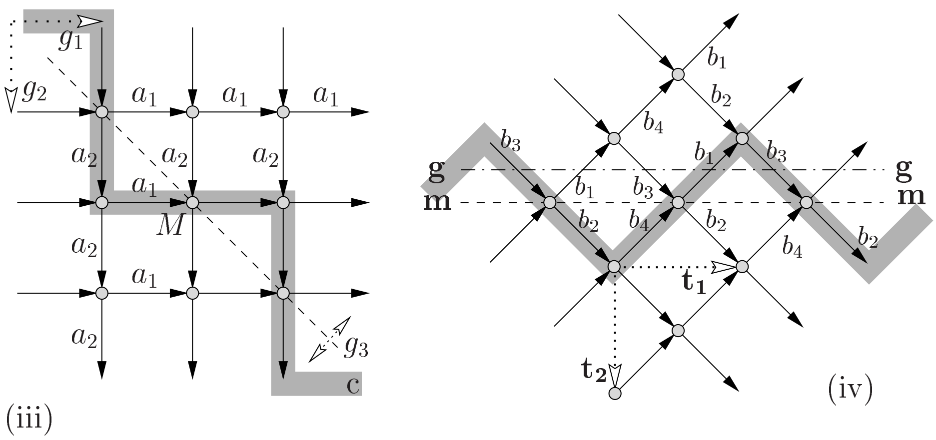
Although this generating system is obvious and natural from a mathematical point of view, crystallographers are more familiar with another generating system. Let us rotate the diagram through 45° for convenience. Then, from Fig. 14(ii), crystallographers will recognize what they call a centered lattice (for which one unit cell is shaded). As shown in Fig. 14(iv) there is a reflection $\mathbf{m}(= g_3)$ in a mirror axis (also indicated by $\mathbf{m}$) and a glide reflection $\mathbf{g}(= g_2 g_3)$ along a parallel glide axis (also indicated by $\mathbf{g}$), designated mirror and glide plane in crystallographer’s terminology.

Since $g_1 = \mathbf{mg}$, $g_2 = \mathbf{gm}$ and $g_3 = \mathbf{m}$, we see that $G$ is generated by $\mathbf{g}$ and $\mathbf{m}$, and that $G$ is the group denoted by $\mathsf{cm}$ in the list of 17 2-dimensional crystallographic plane groups (see [9, Tab. I, p. 494] for example).

Note that from the graph-theoretical point of view all graphs in Fig. 14 are the same (they could even be drawn with arrows of different lengths). The presentation here is given by such a symmetric pattern just to explain easily how the action of the symmetry operations works (in crystallographic terms). Otherwise it would have been difficult to explain explicitly the action of the group $G$.




$G$-constructor graph $\Sigma(\Gamma_G)$




$T'$-constructor graph $\Sigma(\Gamma_{T'})$

$$G = \langle\, g_1, g_2, g_3 \,\rangle$$
$$T' = \langle\, g_1, g_2 \,\rangle$$

$T''$-constructor graph $\Sigma(\Gamma_{T''})$

$$G = \langle\, \mathbf{m}, \mathbf{g} \,\rangle$$
$$T'' = \langle\, \mathbf{t_1}, \mathbf{t_2} \,\rangle$$

Figure 14: Constructor graphs of some $G$-array

The given $G$-constructor graph has many symmetries. All vertices are $G$-equivalent (each may be labelled by $M$). Note that this means that in the underlying $G$-array all graph molecules are $G$-equivalent. Further, all H-edges are also $G$-equivalent, i.e., there is only one $G$-orbit on the set $E_B$ of H-edges. Thus every H-edge has the same $\beta_G$-label, which we shall denote by $b := e_1^G = E_B$.

However, we see that there are two H-edges (e.g. $e_1$ and $e_2$) leaving each molecule, so when we introduce the $T$-constructor graph we cannot take $T = G$ because this would contradict 5.3(1) ($\mathbf{m}$ has fixed points). Instead we must choose $T$ to be some normal subgroup of finite index which acts H-edge semiregularly (see 5.3). For illustrative purposes, we will demonstrate the method with two different choices of $T$, namely the subgroup $T'$ generated by $g_1$ and $g_2$ (see Fig. 14(iii)), and the subgroup $T''$ generated by $\mathbf{t}_1 = g_1 g_2$ and $\mathbf{t}_2 = g_1^{-1} g_2$ (see Fig. 14(iv)); each small square in Fig. 14(i) is a unit cell for $T'$, while the shaded square in Fig. 14(ii) is a unit cell for $T''$. The use of $T = T'$ or $T''$ seems to be more natural to mathematicians or to crystallographers, respectively. $T''$ is the translational subgroup in the crystallographic setting of $\mathsf{cm}$ and so will be chosen by default according to Remark 5.4a.

The graph sets and quantitative descriptors for these two choices of $T$ remain the same, but, as we shall see, their descriptions by qualitative descriptors may differ. We shall discuss both cases in parallel in order to demonstrate this. It turns out that here the treatment via the $T''$-constructor graph is unnecessarily complicated in comparison with the treatment via the $T'$-constructor graph, the main reason being that $T''$ is a proper subgroup of $T'$.

We have $G = T'' \cup T''\mathbf{m} \cup T''\mathbf{g} \cup T''\mathbf{gm}$ and thus there are four $T''$-orbits of $\beta_G$-labels (cf. 5.4b), which we shall denote by $b_1 := e_1^{T''}$, $b_2 := e_2^{T''} = e_1^{T''\mathbf{m}}$, $b_3 := e_1^{T''\mathbf{g}}$, $b_4 := e_1^{T''\mathbf{gm}}$. The resulting $T''$-constructor graph is shown in Fig. 14(iv).

On the other hand we have $G = T' \cup T' g_3$. Thus there are two $T'$-orbits of $\beta_G$-labels, which we shall denote by $a_1 := e_1^{T'}$ and $a_2 := e_2^{T'} = e_1^{T' g_3}$ yielding the $T'$-constructor graph Fig. 14(iii).

The action of the coset representatives $g \in G$ on $\beta_{T'}$- and $\beta_{T''}$-labels $x \in B_{T'}$ and $y \in B_{T''}$ (as described in 5.4c) is given in

Table 3 ($\mathsf{id}$ denotes the identity mapping representing $T' = T'\mathsf{id}$ and $T'' = T''\mathsf{id}$).

| $x^g$ | $G/T'$ id | $g_3$ |
|---|---|---|
| $a_1$ | $a_1$ | $a_2$ |
| $a_2$ | $a_2$ | $a_1$ |

| $y^g$ | $G/T''$ id | $\mathbf{m}$ | $\mathbf{g}$ | $\mathbf{gm}$ |
|---|---|---|---|---|
| $b_1$ | $b_1$ | $b_2$ | $b_3$ | $b_4$ |
| $b_2$ | $b_2$ | $b_1$ | $b_4$ | $b_3$ |
| $b_3$ | $b_3$ | $b_4$ | $b_1$ | $b_2$ |
| $b_4$ | $b_4$ | $b_3$ | $b_2$ | $b_1$ |

Table 3: The action of $G$ on the $\beta_{T'}$- and $\beta_{T''}$-labels

There are still two data which we need in order to perform Graph Set Analysis (in particular to compute descriptors). We need the original $\beta_G$-labels of the $\beta_T$-labels introduced — this is trivial in our case because there is only one $\beta_G$-label $b$. Formally one has to go back to the definition of the $\beta_T$-labels, e.g. $b_1 = e_1^{T''}$, and then to define $\beta_G(b_1) := \beta_G(e_1)$ (see 5.4b). This gives Table 4.

| $x$ | $a_1$ | $a_2$ | $b_1$ | $b_2$ | $b_3$ | $b_4$ |
|---|---|---|---|---|---|---|
| $\beta_G(x)$ | $b$ | $b$ | $b$ | $b$ | $b$ | $b$ |

Table 4: The $\beta_G$-labels of $\beta_{T'}$- and $\beta_{T''}$-labels

Finally we need the covalent distance table. Let us assume hypothetically that the covalent distance table for the $T'$-constructor graph of a graph molecule $M$ (all graph molecules are $G$-equivalent to $M$) is as given in Table 5. We do not claim that there is any crystal structure with such molecules. We only want to demonstrate the formal treatment of our notions and computations.

According to 7.3(a,b,c) a few entries (shaded here) of the covalent distance table suffice to determine all the others.

The covalent distance table for the $T''$-constructor graph repeats the same information; however it looks more complicated since any $T'$-orbit splits into two $T''$-orbits (if we remember that the labels as orbits are sets then we can write: $a_1 = b_1 \cup b_4, a_2 = b_2 \cup b_3$).

Again, the shaded entries suffice to determine all the others (using 7.3(a,b,c)). Here for the first time we also have to use entries $\infty$ in order to show that there is no path with such a pattern in its qualitative descriptor. For instance, Fig. 14(iv) shows immediately that there is no path $p$ (and therefore also no graph set) in the $T''$-constructor graph containing $(\dots \vec{b}_1 \vec{b}_1 \dots)$ in its sequence $\lambda_{T''}(p)$ of $\beta_{T''}$-labels (cf. 5.11), i.e., $\mathsf{cd}(\vec{b}_1 \vec{b}_1) = \infty$.

for the $T'$-constructor graph:

| $\mathsf{cd}(\sigma\sigma')$ | $\sigma' =$ $\vec{a}_1$ | $\overleftarrow{a}_1$ | $\vec{a}_2$ | $\overleftarrow{a}_2$ |
|---|---|---|---|---|
| $\sigma = \vec{a}_1$ | 5 | 0 | 3 | 2 |
| $\overleftarrow{a}_1$ | 0 | 5 | 6 | 3 |
| $\vec{a}_2$ | 3 | 2 | 5 | 0 |
| $\overleftarrow{a}_2$ | 6 | 3 | 0 | 5 |

for the $T''$-constructor graph:

| $\mathsf{cd}(\sigma\sigma')$ | $\sigma' =$ $\vec{b}_1$ | $\overleftarrow{b}_1$ | $\vec{b}_2$ | $\overleftarrow{b}_2$ | $\vec{b}_3$ | $\overleftarrow{b}_3$ | $\vec{b}_4$ | $\overleftarrow{b}_4$ |
|---|---|---|---|---|---|---|---|---|
| $\sigma = \vec{b}_1$ | $\infty$ | 0 | $\infty$ | 2 | 3 | $\infty$ | 5 | $\infty$ |
| $\overleftarrow{b}_1$ | 0 | $\infty$ | 6 | $\infty$ | $\infty$ | 3 | $\infty$ | 5 |
| $\vec{b}_2$ | $\infty$ | 2 | $\infty$ | 0 | 5 | $\infty$ | 3 | $\infty$ |
| $\vec{b}_3$ | 3 | $\infty$ | 5 | $\infty$ | $\infty$ | 0 | $\infty$ | 2 |
| $\vec{b}_4$ | 5 | $\infty$ | 3 | $\infty$ | $\infty$ | 2 | $\infty$ | 0 |
| | | | | $\vdots$ | | | | |

Table 5: The (hypothetically given) covalent distance table for the graph molecule $M$ of the array in Fig. 14

Now we are ready for Graph Set Analysis. Let us examine some graph sets. In the constructor graph there are no selfs and no discretes. Obviously there is a ring in $\Sigma(\Gamma_{T'})$ with qualitative descriptor $\mathbf{R}(\vec{a}_1 \vec{a}_2 \overleftarrow{a}_1 \overleftarrow{a}_2)$. In $\Sigma(\Gamma_{T''})$ its descriptor is $\mathbf{R}(\vec{b}_1 \vec{b}_3 \overleftarrow{b}_4 \overleftarrow{b}_2)$. The formulas in 7.11(i) give $L = \{b\}$, $n = 18$, a = d = 4, i.e., the corresponding quantitative descriptor is $\{b\} : \mathbf{R}_4^4(18)$.

There are many chains, e.g. $\mathbf{C}(\vec{a}_1) = \mathbf{C}(\vec{b}_1\,\vec{b}_4)$, $\mathbf{C}(\vec{a}_2) = \mathbf{C}(\vec{b}_2\,\vec{b}_3)$, $\mathbf{C}(\vec{a}_1\,\vec{a}_2) = \mathbf{C}(\vec{b}_1\,\vec{b}_3)$. However we shall treat here only the chain $c$ shown shaded in Fig. 14.

In $\Sigma(\Gamma_{T'})$ it has the qualitative descriptor $\mathbf{C}(\vec{a}_1\,\vec{a}_1\,\vec{a}_2\,\vec{a}_2)$. In $\Sigma(\Gamma_{T''})$ we get $\mathbf{C}(\vec{b}_4\,\vec{b}_1\,\vec{b}_3\,\vec{b}_2)$.

In order to apply the formulas in 7.9(ii) we have to determine the splitting factor $f(\sigma)$. Recall that $G = T' \cup T'g_3 = T'\mathsf{id} \cup T'g_3$. We have $\sigma^{g_3} \overset{\circ}{=} \sigma$, in fact, $\sigma^{g_3} = (\vec{a}_1\,\vec{a}_1\,\vec{a}_2\,\vec{a}_2)^{g_3} = (\vec{a}_2\,\vec{a}_2\,\vec{a}_1\,\vec{a}_1)$ is a cyclic shift of $\sigma$ with $\sigma^{g_3^2} = \sigma$. Therefore $o(\sigma, g_3) = 2$. Trivially we have $o(\sigma, \mathsf{id}) = 1$. Thus $f(\sigma) = 2$ (cf. 7.6). By 7.9(ii) we have

$$
\begin{aligned}
L &= \{\beta_G(\vec{a}_1), \beta_G(\vec{a}_1), \beta_G(\vec{a}_2), \beta_G(\vec{a}_2)\} = \{b\}\,,\\
n &= \frac{1}{2}\cdot(\mathsf{cd}(\vec{a}_1\,\vec{a}_1) + \mathsf{cd}(\vec{a}_1\,\vec{a}_2) + \mathsf{cd}(\vec{a}_2\,\vec{a}_2) + \mathsf{cd}(\vec{a}_2\,\vec{a}_1) + 4)\\
&= \frac{1}{2}\cdot(5+3+5+3+4) = 10\,,\\
\mathrm{a} &= \frac{1}{2}\cdot 4 = 2, \quad \mathrm{d} = \frac{1}{2}\cdot 4 = 2\,.
\end{aligned}
$$

Consequently, $\{b\} : \mathbf{C}_2^2(10)$ is the corresponding quantitative descriptor.

One gets the same results with $\sigma = (\vec{b}_4\,\vec{b}_1\,\vec{b}_3\,\vec{b}_2)$ in the $T''$-constructor graph. Here $f(\sigma) = 2$ since $\sigma^{\mathbf{m}} = (\vec{b}_3\,\vec{b}_2\,\vec{b}_4\,\vec{b}_1)$ is a cyclic shift of $\sigma$ with $\sigma^{\mathbf{m}^2} = \sigma$, thus $o(\sigma, \mathbf{m}) = 2$ (note that $o(\sigma, \mathbf{g}) = 0$, $o(\sigma, \mathbf{gm}) = 0$ and trivially $o(\sigma, \mathsf{id}) = 1$). The descriptor is calculated analogously:

$$
\begin{aligned}
L &= \{\beta_G(\vec{b}_4), \beta_G(\vec{b}_1), \beta_G(\vec{b}_3), \beta_G(\vec{b}_2)\} = \{b\}\,,\\
n &= \frac{1}{2}\cdot(\mathsf{cd}(\vec{b}_4\,\vec{b}_1) + \mathsf{cd}(\vec{b}_1\,\vec{b}_3) + \mathsf{cd}(\vec{b}_3\,\vec{b}_2) + \mathsf{cd}(\vec{b}_2\,\vec{b}_4) + 4)\\
&= \frac{1}{2}\cdot(5+3+5+3+4) = 10\,,\\
\mathrm{a} &= \frac{1}{2}\cdot 4 = 2, \quad \mathrm{d} = \frac{1}{2}\cdot 4 = 2\,.
\end{aligned}
$$

**7.12. Concluding remarks.** The results presented so far are a mathematical background for Graph Set Analysis. In order to determine the graph sets of a given crystal structure (with given symmetry group), according to this approach one has to proceed as follows:

(1) Determine the $G$-constructor graph (with $\beta_G$-labels for H-edges) and determine the covalent distance table.

(2) Decide whether the action of $G$ is H-edge semiregular (cf. 5.3(1)) (equivalently, one can check the condition (°) in 5.7, cf. 5.4(d)). If so, let $T = G$, otherwise let $T$ be the subgroup of all translations in $G$.

(3) Determine the $\beta_T$-labels of the edges of the $G$-constructor graph (which has been computed in step (1)), yielding the constructor graph.

(4) Determine constructor graph sets and their qualitative descriptors up to $G$-equivalence (i.e. determine one representative constructor ring, chain, self or discrete for each $G$-equivalence class), cf. 6.1-6.6.

(5) Determine the corresponding quantitative descriptors (7.9).

It is clear that all these steps can be performed effectively. However, we have not dealt with the question of how to implement them, so it remains open to develop effective algorithms for any of the above steps. Some parts have already been implemented in the Cambridge Structural Database program PLUTO (see [19]) using the constructor graph methodology.

Moreover, the above steps include the question of a representation of $G$-arrays and constructor graphs that makes it possible to extract easily all the necessary information. Further research in this direction might be based on the methods developed in [8] where it is shown that infinite periodic nets can be represented by finite labelled graphs.

We place particular emphasis on the fact that once the constructor graph (with $\beta_G$- and $\beta_T$-labels) and the covalent distance table are known, no further knowledge about the original crystal structure or $G$-array is required.

All notions for the constructor graph are defined independently of the $G$-array (this will allow a modular approach to the algorithmic treatment). Therefore – in the meantime – we used for instance the notion of *constructor graph sets* for graph sets in the constructor graph. However we think that the constructor graph might become the basic tool for Graph Set Analysis, so that in future all notions should be described in terms of the constructor graph from the very beginning. We also refer to [12, 13] for a presentation which is much closer to applications from the crystallographic point of view.

# 8 Mathematical Appendix

## A. Groups and actions

**8.1. Group actions.** Given a group $G$, a ***(group) action*** of $G$ on a set $V$ is a group homomorphism

$$\psi : G \to \mathsf{Sym}(V)$$

of $G$ into the full symmetric group $\mathsf{Sym}(V)$ on $V$.

If the action $\psi$ is fixed, the image of an element $v \in V$ under the permutation $\psi(g)$ is denoted by $v^g$, and the action is denoted by $(G, V)$. The restriction $\psi|_T : T \to \mathsf{Sym}(V)$ of any action $\psi : G \to \mathsf{Sym}(V)$ to a subgroup $T$ of $G$ is an action of $T$ on $V$ denoted by $(T, V)$.

If $V$ is the base set of an algebraic or relational structure $\Gamma$, then $\psi$ is called ***compatible*** if $\psi(g)$ is an automorphism of $\Gamma$ for every $g \in G$.

**8.2. Induced actions.** For a given action $(G, V)$, there are canonically defined actions $(G, \mathfrak{P}(V))$ on the power set $\mathfrak{P}(V)$, and $(G, V^n)$ on the set of $n$-tuples, given by $M^g := \{v^g \mid v \in M\}$ for $M \subseteq V$ and $(v_1, \ldots, v_n)^g := (v_1^g, \ldots, v_n^g)$ for $(v_1, \ldots, v_n) \in V^n$. One often refers

to them as ***induced actions***. If $(G, V)$ is a compatible action on the vertex set $V$ of a graph $\Gamma = (V, E)$, then it also induces actions on paths, edge sets and subgraphs as follows:
$p^g := (v_0^g, e_1^g, \ldots, e_n^g, v_n^g)$ for a path $p = (v_0, e_1, \ldots, e_n, v_n)$, $F^g := \{e^g \mid e \in F\}$ for $F \subseteq E$ and $(U, F)^g := (U^g, F^g)$ for a subgraph $(U, F)$ of $(V, E)$.

**8.3. Orbits.** Let $(G, V)$ be a group action and $v \in V$. Then we define

$$v^G := \{v^g \mid g \in G\}.$$

These sets $v^G$ are called ***orbits*** of $G$ (or $G$-orbits). It is well known that

$$V = \bigcup_{v \in V} v^G$$

is a partition of $V$ into $G$-orbits. $\mathrm{Orb}(G, V)$ denotes the set of all orbits of an action $(G, V)$. For $g \in G$, the orbits of the cyclic group generated by $g$ (i.e. the sets $\{v^{g^n} \mid n \in \mathbb{Z}\}$) are simply called ***orbits of g***.

## B. Graph-theoretical notions

Since different textbooks on Graph Theory use slightly different definitions we felt some need to fix the graph-theoretical notions and notations in the form which is used in this paper.

**8.4. Graphs (undirected).** An ***undirected graph*** $\Gamma = (V, E)$ (without loops and multiple edges) consists of a set $V$ of ***vertices*** and a set $E$ of two-element subsets of $V$. A subset $\{u, v\} \in E$ is called an ***(undirected) edge*** between $u$ and $v$ (the ***end vertices*** of $e$), symbolized also by $u$ — $v$. For $e = \{u, v\}$ we say that $e$ is ***incident*** with $u$ and $v$ (and vice versa).

The number of edges between a vertex $u$ and the other vertices is called the ***valency*** or ***degree*** of $u$.

A graph $\Gamma' = (V', E')$ with $V' \subseteq V$, $E' \subseteq E$ is called a ***subgraph*** of $\Gamma = (V, E)$.

For a set $S = \{(V_i, E_i) \mid i \in I\}$ of graphs, the ***union*** is defined by $\bigcup S := (\bigcup_{i \in I} V_i, \bigcup_{i \in I} E_i)$.

**8.5. Directed graphs.** A ***directed graph*** $\Gamma = (V, E)$ consists of a set $V$ of vertices and a set $E \subseteq V \times V$ of pairs of vertices. A pair $\vec{e} = (u, v) \in E$ is called a ***(directed) edge*** from vertex $u$ to vertex $v$; we represent it by $u \longrightarrow v$ or $v \longleftarrow u$, and also by $u \overset{\vec{e}}{\text{——}} v$ or $v \overset{\overleftarrow{e}}{\text{——}} u$. Edges of the form $(u, u)$ are called ***loops***.

**8.6. Multigraphs.** A ***directed multigraph*** $\Sigma = (V, E, J)$ consists of a set $V$ of vertices, a set $E$ of edges and an ***incidence mapping***

$$J : E \to V \times V .$$

An edge $e \in E$ with $J(e) = (u, v)$ is called a (directed) edge from the vertex $u$ (the ***source*** of $e$) to the vertex $v$ (the ***sink*** of $e$), symbolized by $u \overset{e}{\longrightarrow} v$, or also by $u \overset{e}{\text{——}} v$ or $v \overset{e}{\text{——}} u$ (since the direction is already given via $J$ by the "label" $e$). If the direction is unimportant or clear from the context, we simply speak of an edge between $u$ and $v$.

Notice that there can be several edges between two vertices ("multiple edges").

A ***labelled*** multigraph $(V, E, J, \alpha, \beta)$ is a multigraph $(V, E, J)$ together with a vertex-labelling $\alpha$ and an edge-labelling $\beta$.

Every directed graph $\Gamma = (V, E)$, $E \subseteq V \times V$, can be considered as a directed multigraph by defining $J(e) = (u, v)$ for $e = (u, v)$.

**8.7. Paths and cycles.** A ***walk*** of ***length*** $n$ in a graph $\Gamma = (V, E)$ is an alternating sequence $p = (v_0, e_1, v_1, e_2, \ldots, e_n, v_n)$ of vertices and edges such that $e_i$ is an edge between $v_{i-1}$ and $v_i$ for $i \in \{1, \ldots, n\}$. Of course, $p$ is uniquely defined by the sequence of its edges only. Clearly, an ***infinite walk*** is a sequence

$(e_i, v_i)_{i \in \mathbb{Z}}$ such that $e_i$ is incident with $v_{i-1}$ and $v_i$ for all $i \in \mathbb{Z} = \{\dots, -1, 0, 1, 2, \dots\}$.

For a walk $p$ let $V(p)$ and $E(p)$ denote the set of vertices and edges occurring in $p$. To every walk $p$ we consider the graph $\Gamma(p) := (V(p), E(p))$ which we call the ***graph of p***. This is the subgraph of $\Gamma$ defined by the set of vertices and edges involved in $p$. Both the walk $p$ and its graph $\Gamma(p)$ will be symbolized by

$$(v_0 \overset{e_1}{\text{---}} v_1 \overset{e_2}{\text{---}} \dots \overset{e_n}{\text{---}} v_n).$$

A walk with pairwise different edges is called ***path***. A ***cycle*** of length $n$ is a path $p$ with $v_0 = v_n$. A path $p$ which is not a cycle is called ***simple*** if all vertices are different. A cycle $(v_0 \overset{e_1}{\text{---}} \dots \overset{e_{n-1}}{\text{---}} v_{n-1} \overset{e_n}{\text{---}} v_0)$ is called simple if the vertices $v_0, \dots, v_{n-1}$ are different.

Simple paths in (directed) graphs are already uniquely determined by the sequence of their edges. (This remains true for almost all paths except those which involve only two vertices, e.g., a cycle of length 2 where $p = (e_1 e_2)$ can be either $v_0 \overset{e_1}{\text{---}} v_1 \overset{e_2}{\text{---}} v_0$ or $v_1 \overset{e_1}{\text{---}} v_0 \overset{e_2}{\text{---}} v_1$.) This gives a shorter notation for paths, e.g., the above path $p$ will also be denoted by

$$(e_1 e_2 \dots e_n)\,.$$

Notice that, for example, the two paths

$$\begin{aligned} p &= (v_0 \overset{e_1}{\text{---}} v_1 \overset{e_2}{\text{---}} \dots \overset{e_n}{\text{---}} v_n)\,, \\ p' &= (v_n \overset{e_n}{\text{---}} v_{n-1} \overset{e_{n-1}}{\text{---}} \dots \overset{e_1}{\text{---}} v_0) \end{aligned}$$

are ***representatives*** of the same graph, i.e., $\Gamma(p) = \Gamma(p')$, just as the cycles

$$\begin{aligned} r_0 &= (v_0 \overset{e_1}{\text{---}} v_1 \overset{e_2}{\text{---}} \dots \overset{e_{n-1}}{\text{---}} v_{n-1} \overset{e_n}{\text{---}} v_0)\,, \\ r'_0 &= (v_0 \overset{e_n}{\text{---}} v_{n-1} \overset{e_{n-1}}{\text{---}} \dots \overset{e_2}{\text{---}} v_1 \overset{e_1}{\text{---}} v_0)\,, \\ r_1 &= (v_1 \overset{e_2}{\text{---}} v_2 \overset{e_3}{\text{---}} \dots \overset{e_{n-1}}{\text{---}} v_{n-1} \overset{e_n}{\text{---}} v_0 \overset{e_1}{\text{---}} v_1)\,, \\ r'_1 &= (v_1 \overset{e_1}{\text{---}} v_0 \overset{e_n}{\text{---}} v_{n-1} \overset{e_{n-1}}{\text{---}} \dots \overset{e_3}{\text{---}} v_2 \overset{e_2}{\text{---}} v_1) \end{aligned}$$

all represent the same graph $\Gamma(r_0) = \Gamma(r'_0) = \Gamma(r_1) = \Gamma(r'_1)$.

Often one does not want to distinguish between paths $p$ which have the same graph $\Gamma(p)$. Thus speaking of paths, cycles $p$, etc. one can have in mind the corresponding graph $\Gamma(p)$, taking the specific walk $p$ as a representative of $\Gamma(p)$.

**Remark:** All notions, which for simplicity are introduced above for graphs $\Gamma = (V, E)$, remain valid if we use directed graphs or multigraphs $\Gamma = (V, E, J)$ instead of graphs, just by interpreting $v \overset{e}{\text{---}} v'$ as a (directed) edge $e \in E$ from $v$ to $v'$ if $J(e) = (v, v')$, or from $v'$ to $v$ if $J(e) = (v', v)$.

**8.8. Periodic paths.** Let $\Sigma = (V, E, J)$ be a multigraph (graphs are included according to 8.6) with a compatible action of a group $G$ on $V$ as well as on $E$, i.e. satisfying the compatibility condition $\forall e \in E\, \forall g \in G : J(e^g) = J(e)^g$, i.e., $J(e) = (u, v)$ implies $J(e^g) = (u^g, v^g)$.

Let $g \in G$. An infinite path $c = (e_i, v_i)_{i \in \mathbb{Z}}$ is called ***g-periodic*** if there exists $m \in \mathbb{Z} \setminus \{0\}$ such that $e_i^g = e_{i+m}$ and consequently $v_i^g = v_{i+m}$ for all $i \in \mathbb{Z}$. The interpretation of this is that $g$ moves the path $c$ along itself $m$ steps as shown in the picture in 4.4.

If we want to distinguish between $m > 0$ and $m < 0$ then we call $c$ ***right*** $g$-periodic if $m > 0$, and ***left*** $g$-periodic if $m < 0$. A subpath $p$ consisting of $|m|$ consecutive edges is called a (right or left, resp.) ***g-period*** of $c$, e.g. $p = (e_1 e_2 \dots e_{|m|})$; $m$ is denoted by $\pi_g(c)$, and $|m|$ is the right or left, resp., ***g-period length*** of $c$. A path $c$ is called ***G-periodic*** if it is $g$-periodic for some $g \in G$. The ***G-period length*** $\pi_G(c)$ of a $G$-periodic path is the minimum of all right $g$-period lengths of $c$, and any subpath $p$ of this length is called a ***G-period*** of $c$. Since $G$ is a group one can restrict to such $g$ with $m > 0$ (note $\pi_{g^{-1}}(c) = -\pi_g(c)$).

Following a suggestion of M. Klin we add here some useful group-theoretical considerations concerning the period length.

(a) Let $G_c := \{g \in G \mid c \text{ is } g\text{-periodic or } g \text{ fixes } c\}$. Note that every $g \in G_c$ acts on the edge set $E(c) = \{\dots, e_{-1}, e_0, e_1, e_2, \dots\}$ of the path $c$ as a "shifting" permutation $\tilde{g}$. More precisely, $\tilde{g} : E(c) \to E(c) : e \mapsto e^g$ is a permutation of $E(c)$ and belongs to the cyclic permutation group $C_\infty := \langle \mathsf{s}_1 \rangle = \{\mathsf{s}_n \mid n \in \mathbb{Z}\}$ generated by the "1-shift" $\mathsf{s}_1 : e_i \mapsto e_{i+1}$ $(i \in \mathbb{Z})$, $\mathsf{s}_n := {\mathsf{s}_1}^n$ $(n \in \mathbb{Z})$. Consequently, the index $[C_\infty : \langle \tilde{g} \rangle]$ coincides with $|\pi_g(c)|$ and we have $\tilde{g} = \mathsf{s}_m$ for $m = \pi_g(c)$. In particular we have for a $g$-period $p$ of $c$:

$$p^g = p^{\tilde{g}} = p^{\mathsf{s}_m} .$$

(b) $G_c$ is closed under composition and we have $\widetilde{gg'} = \tilde{g}\tilde{g}'$ for $g, g' \in G_c$. Thus $G_c$ is a subgroup of $G$ and $\tilde{G}_c := \{\tilde{g} \mid g \in G_c\}$ is a homomorphic image. Since $\tilde{G}_c$ is a subgroup of a cyclic group (namely of $C_\infty$), it must be cyclic. Thus there exists a generating element $\mathsf{s}_{m_0} = \tilde{g}_0 \in \tilde{G}_c$ with $m_0 > 0$ and $\langle \tilde{g}_0 \rangle = \tilde{G}_c$. The index $[C_\infty : \tilde{G}_c]$ is just the $G$-period length of $c$ and coincides with the (right) $g_0$-period length $m_0$ of $c$.

(c) Since $\langle \tilde{g} \rangle$ is a subgroup of $\tilde{G}_c = \langle \tilde{g}_0 \rangle$ we have $[C_\infty : \langle \tilde{g}_0 \rangle] \cdot [\langle \tilde{g}_0 \rangle : \langle \tilde{g} \rangle] = [C_\infty : \langle \tilde{g} \rangle]$, in particular, the $g$-period length of $c$ is always a multiple of the $g_0$-period (=$G$-period) length.

## C. The remaining proofs and some further results

**8.9.** Proof **of 5.13.** The proof is formulated here for finite paths, but it works similarly for infinite paths.

(a): "$\Longleftarrow$": Assume that $\lambda_T(p)$ and $\lambda_T(p')$ are $G$-equivalent, i.e., we have $\lambda_T(p)^{g'} = \lambda_T(p')$ for a path $p = M_0 \xrightarrow{e_1} \ldots \xrightarrow{e_n} M_n$, a path $p' = M'_0 \xrightarrow{e'_1} \ldots \xrightarrow{e'_n} M'_n$ and some $g' \in G$. Thus, in particular, $\lambda_p(e_i)^{g'} = \lambda_{p'}(e'_i)$ (cf. 5.11), and the directions of the edges coincide in both paths. Moreover we have $\beta_T(e_i)^{g'} = \beta_T(e'_i)$. The latter means that $e_i^{Tg'} = {e'_i}^T$, consequently there exist $t_i \in T$ such that $e_i^{t_i g'} = e'_i$ (for $i \in \{1, \ldots, n\}$). Since incidence and the directions $J(e_i)$ are preserved by every compatible action, we have $M_i^{t_i g'} = M_i^{t_{i+1} g'}$, hence $M_i$ is a fixed point of $t_i t_{i+1}^{-1}$ for every $i \in \{1, \ldots, n-1\}$. By assumption 5.3(1), every element of $T$ with a fixed point must act trivially on the edges: thus $e^{t_i} = e^{t_{i+1}}$ for every $e \in E_B$. We can therefore take $t := t_1$ (acting on edges in the same way as $t_2, \ldots, t_n$) and get $p^{tg'} = p'$, i.e., we can take $g := tg'$ (obviously, $g \in Tg'$, i.e., $Tg = Tg'$).

"$\Longrightarrow$": Assume that $p$ and $p'$ are $G$-equivalent, i.e., $p^g = p'$ for some

$g \in G$. Recall, by 5.4(c), that we have $\beta_T(e^g) = \beta_T(e)^g$ for every edge $e \in E_B$. This immediately implies that $\lambda_T(p)^g = \lambda_T(p')$.

(b) is a direct consequence of (a) (cf. 5.12). $\square$

**8.10.** Proof **of 6.14.** Let $L \subseteq \{\beta_G(e) \mid e \in E_B\}$. We distinguish several cases.

Case 1: Suppose that there is a self $S$ of type $L$. Then (by definition 6.4) the edges $e$ in $S$ are all loops incident with a single vertex $M \in \mathcal{M}$. Consequently, $e^g$ is a loop at $M^g$ for every $g \in G$, in other words, every edge in the orbit $e^G$, i.e., every edge $e' \in E_B$ with the same $\beta_G$-label $\beta_G(e') = \beta_G(e)$, is a loop. Therefore only selfs can contain edges of this type; moreover $S$ is a self if and only if $S^g$ is a self.

Case 2: Suppose that there is a discrete $\Gamma(p) = (M \overset{e_1}{\text{---}} \dots \overset{e_n}{\text{---}} M')$ of type $L$. In particular we have $\beta_G(e_1) \in L$. By definition 6.5, $e_1$ is the only edge which is incident with $M$ and whose $\beta_G$-label belongs to $L$. By H-edge label transitivity this property holds for all $e$ with $\beta_G(e) = \beta_G(e_1)$. Thus no such edge $e$ can belong to a ring or chain of type $L$ (or self as already shown in case 1), i.e., no ring, no chain and no self is of type $L$.

Case 3: Suppose that $L = \{b\}$, and let $e = (M \overset{e}{\longrightarrow} M') \in E$ with $\beta_G(e) = b$. Assume further that

*there are no rings, chains, selfs or discretes of type $L$.* (*)

We consider the various subcases. For brevity an edge $e'$ with $\beta_G(e') = b$ will be called a $b$-edge.

(a) There is no further $b$-edge which is adjacent to $e$ (i.e., sharing a vertex with $e$). Then $e$ constitutes either a self or a discrete of type $L$, each contradicting (*).

(b) There is a further $b$-edge $M' \overset{e'}{\longrightarrow} M''$, i.e., we have a path $\overset{e}{\longrightarrow}\overset{e'}{\longrightarrow}$ of type $L$. By label transitivity there exists some $g \in G$ with $e^g = e'$. If $e^{g^n} = e$ for some $n \in \mathbb{N}$, then $(ee^g e^{g^2} \dots e^{g^{n-1}})$ is a simple cycle (cf. 8.7). Otherwise $(\dots e^{g^{-2}} e^{g^{-1}} e e^g e^{g^2} \dots)$ is an infinite periodic path with $g$-period length 1. Thus either a ring or a chain exists, contradicting (*).

The situation $\xrightarrow{e'}\xrightarrow{e}$ can be treated in the same manner (transposing the roles of $e$ and $e'$).

(c) There is a further $b$-edge $M' \xleftarrow{e_1} M''$, i.e., we have a path $\xrightarrow{e}\xleftarrow{e_1}$ of type $L$. If there is no other $b$-edge adjacent to $e_1$, then $(ee_1)$ is a discrete of type $L$. Otherwise we have subcase (b) for $e_1$ (instead of $e'$) or there exists a path of the form $\xrightarrow{e}\xleftarrow{e_1}\xrightarrow{e'}$. Now we can proceed as in case (b): there exists $g \in G$ with $e^g = e'$, which yields either a ring $(ee_1e^ge_1^ge^{g^2}\dots e^{g^{n-2}}e_1^{g^{n-2}}e^{g^{n-1}})$ or a chain $(\dots e^{g^{-2}}e_1^{g^{-2}}e^{g^{-1}}e_1^{g^{-1}}ee_1e^ge_1^ge^{g^2}e_1^{g^2}\dots)$ of $g$-period length 2.

The situation $\xleftarrow{e_1}\xrightarrow{e}$ can be treated similarly. Thus any subcase contradicts (*), i.e., (*) cannot hold.

Combining Cases 1–3, we obtain the required statement. □

**8.11.** Proof **of 6.3.** Since $T \subseteq G$, a $T$-periodic path $c$ is also $G$-periodic (by Definition 8.8). Conversely, let $c$ be $G$-periodic, i.e., $g$-periodic for some $g \in G$. Since $G/T$ is finite (cf. 5.3(2), 5.5) there exists $n \neq 0$ (e.g., $n = |G/T|$) with $g^n \in T$. Since $g$ acts on $c$ as a non-identity shift, so does $g^n$, i.e., $c$ is also $g^n$-periodic, so it is $T$-periodic.

Let $s := \pi_G(c)$ and $m := \pi_T(c)$. By applying 8.8 to $G$ and to $T$, we see that there exist $g_0 \in G$ and $t_0 \in T$ with $\pi_{g_0}(c) = s$, $\pi_{t_0}(c) = m$; moreover, $m$ is a multiple of $s$ (take $g = t_0$ in 8.8c). □

**8.12. Lemma.** *Let $t_0 \in T$ and $g \in G$.*

a) *Let $c = (e_i)_{i\in\mathbb{Z}}$ be a $t_0$-periodic path* (8.8). *Then $c^g = (e_i^g)_{i\in\mathbb{Z}}$ is $g^{-1}t_0g$-periodic and the period lengths of $c$ and $c^g$ coincide.*

b) *If $p$ is a $T$-period of $c$ then $p^g$ is a $T$-period of $c^g$.*

Proof. a) Let $\pi_{t_0}(c) = m$, i.e., $e_i^{t_0} = e_{i+m}$ for $i \in \mathbb{Z}$. Consequently $(e_i^g)^{(g^{-1}t_0g)} = (e_i^{t_0})^g = e_{i+m}^g$ for $i \in \mathbb{Z}$, and $\pi_{g^{-1}t_0g}(c^g) \leq \pi_{t_0}(c)$. The same argument with $g^{-1}$ instead of $g$ shows $\pi_{t_0}(c) \leq \pi_{g^{-1}t_0g}(c^g)$.

b) Since $T$ is a normal subgroup of $G$ we have $g^{-1}t_0g \in T$ for $t_0 \in T$. Thus (b) follows from (a). □

**8.13.** Proof **of 6.9.** Let $\gamma' = \gamma^t$ for some $t \in T$ and let the qualitative descriptor $\Delta = \mathbf{R}(\sigma)$, $\mathbf{C}(\sigma)$, $\mathbf{D}(\sigma)$, $\mathbf{S}(S)$, resp., be associated with $\gamma$. Then we can associate the qualitative descriptor $\Delta^t := \mathbf{R}(\sigma^t)$, $\mathbf{C}(\sigma^t)$, $\mathbf{D}(\sigma^t)$, $\mathbf{S}(S^t)$, resp. with $\gamma'$ (this is obvious for rings, discretes and selfs; for chains it follows immediately from Lemma 8.12b with $g = t$). However, $\Delta^t$ coincides with $\Delta$ since $T$ acts trivially on $\beta_T$-labels and thus also on the sequence $\sigma$ and on $S$. Consequently, $\gamma^T \subseteq \langle\Delta\rangle_{(T)}$.

Conversely, let $\Delta = \mathbf{R}(\lambda_T(r)) = \mathbf{R}(\lambda_T(r'))$ for simple cycles $r, r'$. Then $\lambda_T(r)^{\mathsf{id}_T} = \lambda_T(r) = \lambda_T(r')$, and by Prop. 5.13 there exists $g$ with $r' = r^g$ and $Tg = T\,\mathsf{id}_T$, i.e., $g \in T$. Thus $\Gamma(r') = \Gamma(r)^g$, i.e., $\Gamma(r)$ and $\Gamma(r')$ are $T$-equivalent.

The proof for selfs and discretes is similar.

In the case of chains, let $\Delta = \mathbf{C}(\lambda_T(p)) = \mathbf{C}(\lambda_T(p'))$ for $T$-periods $p = e_1 \ldots e_m$, $p' = e'_1 \ldots e'_m$ of infinite simple $T$-periodic paths $c = (e_i)_{i\in\mathbb{Z}}$, $c' = (e'_i)_{i\in\mathbb{Z}}$, resp. By definition of a $T$-period (8.8) there are $t, t' \in T$ such that $e_i^t = e_{i+m}$ and $(e'_i)^{t'} = e'_{i+m}$ for $i \in \mathbb{Z}$. Consequently $\lambda_T(e_i) = \lambda_T(e_{i+m})$, $\lambda_T(e'_i) = \lambda_Z(e'_{i+m})$. Thus, from $\lambda_T(e_i) = \lambda_T(e'_i)$ for $i \in \{1, \ldots, m\}$, we can conclude that $\lambda_T(c) = \lambda_T(c')$. As above (by Prop. 5.13) we get $c' = c^g$ for some $g \in T$, i.e. $\Gamma(c)$ and $\Gamma(c')$ are $T$-equivalent.

This shows that $\gamma^T \supseteq \langle\Delta\rangle_{(T)}$. □

**8.14. Lemma.** *Let $c$ be a $T$-periodic path with $T$-period $p$ of length $m$. Then we have*

$$\lambda_T(p^{\mathsf{s}_{n'}}) = \lambda_T(p)^{\hat{\mathsf{s}}_n}$$

*for arbitrary $n' \in \mathbb{Z}$ and $n := n' \bmod m$.*

Proof. Informally, this lemma follows immediately from the fact that $\lambda_T(c)$ is just the infinite repetition of $\lambda_T(p)$, so shifts $\mathsf{s}_{km}$ act trivially on $\lambda_T(c)$ ($k \in \mathbb{Z}$).

Formally, let $c = (e_i)_{i\in\mathbb{Z}}$, $p = (e_{i_0+1}e_{i_0+2} \ldots e_{i_0+m})$ and $\lambda_T(p) = (\sigma_1\sigma_2 \ldots \sigma_m)$. By 8.8 there is some $t_0 \in T$ such that $c$ is $t_0$-periodic with period length $m$; more precisely, we have $e_j^{t_0} = e_j^{\mathsf{s}_m} = e_{j+m}$

for $j \in \mathbb{Z}$. Consequently, $\lambda_T(e_j) = \lambda_T(e_j)^{t_0} = \lambda_T(e_j^{t_0}) = \lambda_T(e_{j+m})$. Thus $\lambda_T(p^{\mathsf{s}_{n'}}) = \lambda_T(e_{i_0+1+n'} \dots e_{i_0+m+n'}) = \lambda_T(e_{i_0+1+n} \dots e_{i_0+m+n}) = (\sigma_{n+1} \dots \sigma_m \sigma_1 \dots \sigma_n) = \lambda_T(p)^{\hat{\mathsf{s}}_n}$. $\square$

Different paths may have the same graph (see 8.7). The following lemma makes this more explicit for rings and chains.

**8.15. Lemma.** *Let $r, r'$ be simple cycles of length $m$, and let $c, c'$ be infinite simple $T$-periodic paths in a $(T, G)$-constructor graph. Then we have:*

(a) $\Gamma(r) = \Gamma(r') \iff \exists\, i \in \{0, 1, \dots, m-1\} :\ r' = r^{\hat{\mathsf{s}}_i}$ *or* $r' = (r^{-1})^{\hat{\mathsf{s}}_i}$.

(b) $\Gamma(c) = \Gamma(c') \iff \exists\, i \in \mathbb{Z} :\ c' = c^{\mathsf{s}_i}$ *or* $c' = (c^{-1})^{\mathsf{s}_i}$.

(c) *Let $\sigma := \lambda_T(r)$, $\sigma' := \lambda_T(r')$. Then $\sigma$ and $\sigma'$ are graph equivalent if and only if $\sigma' \overset{\circlearrowleft}{=} \sigma$ or $\sigma' \overset{\circlearrowleft}{=} \sigma^{-1}$.*

(d) *Let $\sigma := \lambda_T(p)$, $\sigma' := \lambda_T(p')$ for periods $p, p'$ of $c, c'$, respectively. Then $\lambda_T(c)$ and $\lambda_T(c')$ are graph equivalent if and only if $\sigma' \overset{\circlearrowleft}{=} \sigma$ or $\sigma' \overset{\circlearrowleft}{=} \sigma^{-1}$.*

Proof. The proof of (a) and (b) is obvious since a graph is uniquely determined by its vertices and edges. Since $r$ is simple, $r^{\hat{\mathsf{s}}_i}$ and $(r^{-1})^{\hat{\mathsf{s}}_i}$ are the only paths with the same vertices and edges as the path $r$ (and similarly for $c$).

(c): By 5.13b, $\sigma$ and $\sigma'$ are graph equivalent iff $\Gamma(r)^g = \Gamma(r')$ for some $g \in G$ with $Tg = T$, i.e., $g \in T$. Note that $\Gamma(r)^g = \Gamma(r^g)$ and $\lambda_T(r^g) = \lambda_T(r)^g = \lambda_T(r)$. Thus (c) follows from (a).

(d) We can proceed as in (c): from (b) (and 5.13b) we conclude that $\lambda_T(c)$ and $\lambda_T(c')$ are graph equivalent if and only if there exists $i \in \mathbb{Z}$ such that $\lambda_T(c') = \lambda_T(c)^{\mathsf{s}_i}$ or $\lambda_T(c') = (\lambda_T(c)^{-1})^{\mathsf{s}_i}$. By 8.14 we get the equivalent condition $\sigma' \overset{\circlearrowleft}{=} \sigma$ or $\sigma' \overset{\circlearrowleft}{=} \sigma^{-1}$. $\square$

**8.16.** Proof **of 6.11.** It follows from the definitions that the whole proposition is a combination of facts which we already know from 5.13, 8.12 and 8.15.

(I): Let $\sigma = \lambda_T(r)$ and $\sigma' = \lambda_T(r')$ for simple cycles $r, r'$. Then $\langle\Delta\rangle_{\text{set}}^{\text{graph}} = \langle\Delta'\rangle_{\text{set}}^{\text{graph}}$ if and only if $\Gamma(r)$ and $\Gamma(r')$ are $G$-equivalent, i.e., there is some $g \in G$ with $\Gamma(r') = \Gamma(r)^g$ (note that $\langle\Delta\rangle_{\text{set}}^{\text{graph}} = \Gamma(r)^G$ by Definition 6.10).

By 5.13b this is equivalent to the condition that $\sigma^{g'}$ and $\sigma'$ are graph equivalent for some $g' \in G$. By 8.15(c) we are done (note that $(\sigma^{-1})^{g'} = (\sigma^{g'})^{-1}$).

In the case of chains, let $\sigma := \lambda_T(p)$, $\sigma' := \lambda_T(p')$ for periods $p, p'$ of $c, c'$, respectively. Then $\langle\Delta\rangle_{\text{set}}^{\text{graph}} = \langle\Delta'\rangle_{\text{set}}^{\text{graph}}$ if and only if $\Gamma(c)$ and $\Gamma(c)$ are $G$-equivalent. Thus $\lambda_T(c)$ and $\lambda_T(c')$ are $G$-equivalent by 5.13b, and 8.15(d) finishes the proof.

(II) follows from 5.13b and the fact that $\Gamma(p) = \Gamma(p')$ for two simple paths (not cycles) if and only if $p' = p$ or $p' = p^{-1}$.

(III) follows immediately from the definitions. Note that $\langle\Delta\rangle_{\text{set}}^{\text{graph}} = (\bigcup S)^G$.

(IV): Let $\langle\Delta\rangle_{\text{set}}^{\text{graph}} = \gamma^G$ (for notation see 6.10), $\gamma$ being a ring, chain, self or discrete. Then we may consider the $T$-coset decomposition $\gamma^G = \gamma^{g_1T} \cup \cdots \cup \gamma^{g_kT}$. Since the descriptor $\Delta^{g_i}$ can be associated with $\gamma^{g_i}$ we have $\gamma^{g_iT} = (\gamma^{g_1})^T = \langle\Delta^{g_i}\rangle_{(T)}$ (cf. 6.9). $\square$

**8.17. Lemma.** *Let $c$ be a $T$-periodic path with $T$-period length $m$, let $p$ be a $T$-period of $c$ and $\sigma := \lambda_T(p)$. Then we have for arbitrary $g \in G$ and $n \in \{1, \ldots, m-1\}$:*

$$\sigma^g = \sigma^{\hat{\mathsf{s}}_n} \iff \exists\, h \in Tg : \textit{ c is h-periodic with } p^h = p^{\mathsf{s}_n}\,.$$

PROOF. "$\Leftarrow$": Let $c$ be $h$-periodic for $h = tg$ $(t \in T)$ with $p^h = p^{\mathsf{s}_n}$ (cf. 8.8). Then we have (note that $t$ acts trivially on $\sigma$): $\sigma^g = (\sigma^t)^g = \sigma^h = \lambda_T(p)^h = \lambda_T(p^h) = \lambda_T(p^{\mathsf{s}_n}) = \lambda_T(p)^{\hat{\mathsf{s}}_n} = \sigma^{\hat{\mathsf{s}}_n}$. The last but one equality follows from Lemma 8.14.

"$\Rightarrow$": Let $c = (e_i)_{i \in Z}$. Since $p$ is a $T$-period we have $\lambda_T(e_j) = \lambda_T(e_{j_0})$ for $j \in \mathbb{Z}$ and $j_0 := j \bmod m$. Because $\sigma^g = \sigma^{\hat{\mathsf{s}}_n}$ we also have: $\lambda_T(e_j^g) = \lambda_T(e_j)^g = \lambda_T(e_{j_0})^g = \sigma_{j_0}^g = \sigma_{j_0}^{\hat{\mathsf{s}}_n} = \sigma_{n+j_0} = \lambda_T(e_{n+j})$. This situation is illustrated in Figure 15.

In particular we have $\lambda_T(e_{n+1}) = \lambda_T(e_1^g)$. Therefore both $e_{n+1}$ and $e_1^g$ belong to the same $T$-orbit and there exists $t \in T$ with

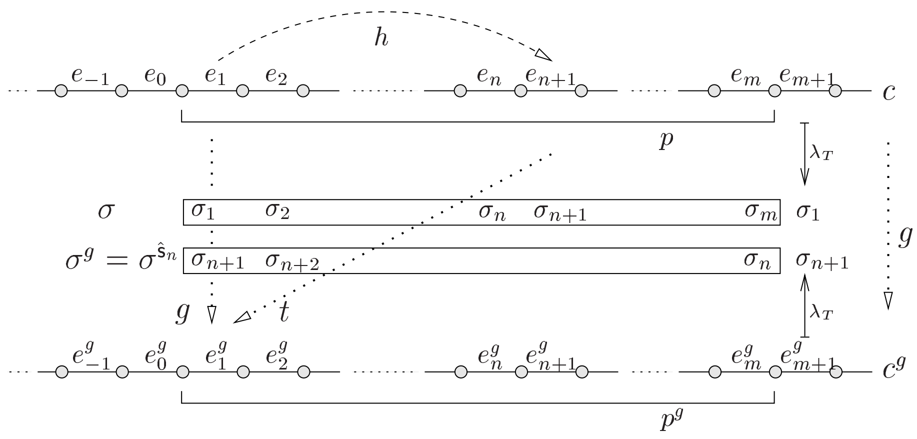



Figure 15: Action of $g$ on $c$, $p$ and $\sigma$

$e^t_{n+1} = e^g_1$ (cf. Fig. 15). Recall that $t \in T$ maps adjacent edges to adjacent edges and preserves $\beta_T$-labels and directions; moreover an edge is uniquely determined by an incident vertex and its $\beta_T$-label (cf. property 5.7($\circ$)). Thus also $e^t_{n+2} = e^g_2$ and so on, i.e., $t$ maps $c$ onto $c^g$, and $e^t_{n+j} = e^g_j$ for $j \in \mathbb{Z}$. Define $h := gt^{-1} \in gT = Tg$. Then we have $e^h_j = (e^g_j)^{t^{-1}} = (e^t_{n+j})^{t^{-1}} = e_{n+j}$ for $j \in \mathbb{Z}$, i.e., $c$ is $h$-periodic with period length $n$: $p^h = p^{\mathsf{s}_n}$. $\square$

Remark: The right-hand side in 8.17 is also equivalent to

$$\exists\, h \in Tg : \ c \text{ is } h\text{-periodic with } p^h = p^{\mathsf{s}_{n'}} \text{ and } n = n' \bmod m$$

(since for such an $h$ there always exists a suitable power $t := t_0^i$ such that $c$ has $th$-period length $n$; here $t_0$ is some element with $\pi_{t_0}(c) = \pi_T(c) = m$, cf. 8.8). Thus Lemma 8.17 also shows that any $h \in G$ (for which $c$ is $h$-periodic) acts on $c$ as the shift $\mathsf{s}_{n'}$ if and only if the coset $[h] \in G/T$ acts on $\sigma$ as the rotation $\hat{\mathsf{s}}_n$ (recall that the action of $G$ on $\beta_T$-labels is in fact the action of $G/T$, cf. 5.4c).

**8.18. Corollary.** *Let $c$, $p$, $m$ and $\sigma$ be as in* 8.17. *If $\sigma' \overset{\circlearrowleft}{=} \sigma$ then there exists a unique $i \in \{0, 1, \ldots, m-1\}$ such that $\sigma' = \sigma^{\hat{\mathsf{s}}_i}$.* (In particular we have $\sigma^{\hat{\mathsf{s}}_i} \neq \sigma$ for $i \in \{1, \ldots, m-1\}$.)

PROOF. If $\sigma^{\hat{\mathsf{s}}_i} = \sigma^{\hat{\mathsf{s}}_j}$ for $1 \le i < j \le m-1$ then $\sigma^{\mathsf{id}} = \sigma = \sigma^{\hat{\mathsf{s}}_{j-i}}$. By 8.17 there exists $h \in T\mathsf{id} = T$ such that $c$ is right $h$-periodic with $p^h = p^{\mathsf{s}_{j-i}}$, so $m = \pi_T(c) \le j-i$ in contradiction to $j-i < m$. Consequently $i = j$. $\square$

**8.19. The period length of $\boldsymbol{\sigma}$.** Because of 8.18 the following definition makes sense (for notation see 8.17). If $\sigma^g = \sigma^{\hat{\mathsf{s}}_n}$ ($n \in \{1, \ldots, m-1\}$) then $\sigma$ is called ***g*-periodic** and $\pi_g(\sigma) := n$ is the (right) ***g*-period length of $\boldsymbol{\sigma}$**. If $\sigma^g = \sigma$ (e.g. for $g \in T$) we formally define $\pi_g(\sigma) := m$.

The minimum of all $g$-period lengths is called the $G$-period length of $\sigma$, and this will be denoted by $\pi_G(\sigma)$. Analogously one defines $\pi_{G'}(\sigma)$ for a subgroup $G'$ of $G$ to be the minimum of all $\pi_g(\sigma)$ with $g \in G'$ for which $c$ is $g$-periodic. Note that $g$ acts on $\sigma = (\sigma_1 \ldots \sigma_m)$ (represented as a circle, see 6.7) as a rotation $\hat{g} := \hat{\mathsf{s}}_n$. These rotations form a cyclic group $C_m := \{\hat{\mathsf{s}}_0, \hat{\mathsf{s}}_1, \ldots, \hat{\mathsf{s}}_{m-1}\}$ generated by $\hat{\mathsf{s}}_1$ (with multiplication $\hat{\mathsf{s}}_i\hat{\mathsf{s}}_j := \hat{\mathsf{s}}_{(i+j \bmod m)}$). Since $\langle \hat{\mathsf{s}}_n \rangle = \langle \hat{\mathsf{s}}_{\gcd(m,n)} \rangle$, the order of $\hat{\mathsf{s}}_n$ coincides with the order of $\hat{\mathsf{s}}_{\gcd(m,n)}$, and we have for the index $[C_m : \langle \hat{\mathsf{s}}_n \rangle] = \gcd(m,n)$. Thus, if $\sigma^g = \sigma^{\hat{\mathsf{s}}_n}$ then $\pi_{\langle g \rangle}(\sigma) = [C_m : \langle \hat{\mathsf{s}}_n \rangle]$. Moreover, $\hat{G}_\sigma := \{\hat{g} \mid g \in G$ and $\sigma^g = \sigma$ or $\sigma$ is $g$-periodic$\}$ is a subgroup of $C_m$ and is therefore generated by some $\hat{g}_0 = \hat{\mathsf{s}}_{n_0}$ with $\pi_{g_0}(\sigma) = \pi_G(\sigma) = [C_m : \hat{G}_\sigma]$.

**8.20. Lemma.** *Let $c$, $p$, $m$ and $\sigma$ be as in* 8.17*, and suppose that $c$ is a right $g$-periodic path for some $g \in G$ with $\pi_g(c) \le m$. Then $\pi_g(c) = \pi_g(\sigma)$ and $o(\sigma, g) = \frac{m}{\pi_{\langle g \rangle}(\sigma)}$ (for notation see* 7.6*). Moreover, $\pi_G(c) = \pi_G(\sigma)$.*

PROOF. Let $n := \pi_g(c)$. If $n < m$ then $\pi_g(c) = \pi_g(\sigma)$ follows directly from 8.17. If $n = m$ then $g$ acts on $\sigma$ as the identity $\hat{\mathsf{s}}_0$ (see 8.14), and we have $\pi_g(\sigma) = m$ by definition. Thus $\pi_g(c) = \pi_g(\sigma)$. Because $\pi_G(c)$ is defined as the minimum of some $\pi_g(c)$ we conclude that $\pi_G(c) = \pi_G(\sigma)$.

Since $\sigma^g = \sigma^{\hat{\mathsf{s}}_n}$ we have $\sigma^{g^i} = \sigma^{\hat{\mathsf{s}}_n^i}$. Thus $o(\sigma, g)$, the least $i > 0$ for which $g^i$ acts on $\sigma$ as the identity, equals the least $i$ with $\hat{\mathsf{s}}_n^i = \hat{\mathsf{s}}_0$, i.e., $o(\sigma, g) = |\langle \hat{\mathsf{s}}_n \rangle| = \frac{m}{\pi_{\langle g \rangle}(\sigma)}$ (cf. 8.18, 8.19). $\square$

**8.21.** PROOF **of 7.7.** Let $c$, $p$, $m$ and $\sigma$ be as in 8.17 and let $g_0 \in G$ such that $\pi_{g_0}(c) = \pi_G(c)$ (the existence of $g_0$ follows from 8.8). Since $f(\sigma)$ is independent of the choice of coset representatives (cf. 7.6) we can choose $g_0$ from among the coset representatives $g_1, \ldots, g_k$. Since $\frac{m}{o(\sigma,g_0)} = \pi_{\langle g_0\rangle}(\sigma) = \pi_{\langle g_0\rangle}(c) \leq \pi_{\langle g\rangle}(c) = \pi_{\langle g\rangle}(\sigma) = \frac{m}{o(\sigma,g)}$ for all other $g \in G$ (with right $g$-periodic path $c$), and since $\pi_{g_0}(\sigma) = \pi_{g_0}(c)$ by 8.20, we finally get $f(\sigma) = \max\{o(\sigma, g_1), \ldots, o(\sigma, g_k)\} = \frac{m}{\pi_{g_0}(\sigma)} = \frac{m}{\pi_{g_0}(c)} = \frac{\pi_T(c)}{\pi_G(c)} = f(c)$. $\square$ .

Remark. The different groups used in the preceding proofs and their connections may be visualized with the aid of the following diagram (for notation $\tilde{G}$ and $\hat{G}_\sigma$ see 8.8(b) and 8.19, resp.). The horizontal arrows indicate epimorphisms, the vertical lines show inclusion and their labels indicate the corresponding indices.

$$\begin{tikzcd}
{} \arrow[dd, dash, "\pi_T(c)"'] & C_\infty \arrow[r] \arrow[d, dash, "\pi_G(c)"] & C_m \arrow[d, dash, "\pi_G(\sigma)"] & {} \arrow[dd, dash, "m"] \\
& \tilde{G}_c = \langle \tilde{g}_0 \rangle \arrow[r] \arrow[d, dash, "f(c)"] & \langle \hat{g}_0 \rangle = \hat{G}_\sigma \arrow[d, dash, "o(\sigma, g_0)"] & \\
{} & \tilde{T}_c = \langle \tilde{t}_0 \rangle \arrow[r] & \{\mathsf{id}\} & {}
\end{tikzcd}$$

**Acknowledgements.** The research reported in this paper was carried out under Grant No. I-0333-263.06/93 from the G.I.F., the German-Israeli Foundation for Scientific Research and Development. The Foundation's support served as the catalyst for the success of this interdisciplinary program. Our main thanks are due to the initiator of this project, M. Klin (Beer-Sheva), for bringing mathematicians and chemists together to develop a common understanding and language, and for his profound input through all stages of the work. Special thanks are due also to G. Tinhofer (München) and G.A. Jones (Southampton) for their incisive and helpful criticism, many useful hints and for careful reading of several versions of the manuscript.

J.-O. Henck (Innsbruck) took part in many discussions, providing helpful insights and criticisms.
U. Hengst (Dresden), who took part in the preliminary version [4], contributed many parts of the anthranilic acid example in Section 2.
A. Ellern and O. Navon (Beer-Sheva) contributed valuable technical assistance.

# References

[1] J. Bernstein, *Polymorphism of L-Glutamic Acid: Decoding the $\alpha$-$\beta$ Phase Relationship* via *Graph-Set Analysis.* Acta Cryst. B**47**(1991), 1004-1010.

[2] J. Bernstein, R.E. Davis, L. Shimoni, N. Chang, *Patterns in Hydrogen Bonding: Functionality and Graph Set Analysis in Crystals.* Angew. Chem. Int. Ed. Engl. 34(1995), 1555-1573.

[3] J. Bernstein, M.C. Etter, J.C. MacDonald, *Decoding Hydrogen-bond Patterns. The case of Iminodiacetic Acid.* J. Chem. Soc. Perkin Trans. **2**(1990), 695-698.

[4] J. Bernstein, B. Ganter, J. Grell, U. Hengst, D. Kuske, R. Pöschel, *Mathematical basics of graph set analysis.* Technical Report MATH-AL-17-1997, TU Dresden, Dec. 1997 (48 pages).

[5] U. Brehm, M. Klin, R. Pöschel, *Finite representation of infinite periodic graphs.* Manuscript in preparation.

[6] C.D.G. Boone, J.L. Derissen, J.C. Schoone, *Anthranilic Acid* II *(o-Aminobenzoic Acid).* Acta Cryst. B**33**(1977), 3205-3206.

[7] Cambridge Structural Database (CSD), see CCDC Home page: `http://www.ccdc.cam.ac.uk`

[8] S.J. Chung, Th. Hahn, W.E. Klee, *Nomenclature and generation of three-periodic nets: the vector method.* Acta Cryst. **A40**(1984), 42-50.

[9] H.S.M. Coxeter, *Unvergängliche Geometrie.* Birkhäuser Verlag, Basel u. Stuttgart, 1963.

[10] M.C. Etter, *Encoding and Decoding Hydrogen-Bond Patterns of Organic Compounds.* Acc. Chem. Res. **23**(1990), 120-126.

[11] M.C. Etter, *Hydrogen Bonds as Design Elements in Organic Chemistry.* J. Phys. Chem. **95**(1991), 4601-4610.

[12] J. Grell, J. Bernstein, G. Tinhofer, *Graph Set Analysis of Hydrogen Bond Patterns. Some Mathematical Concepts.* Acta Cryst. **B55**(1999), 1030-1043.

[13] J. Grell, J. Bernstein, G. Tinhofer, *Investigation of Hydrogen Bond Patterns: The Graph Set Approach. Part I: General Ideas and Their Application to Crystal Structures with exactly one Molecule in the Asymmetric Unit.* Report TUM-M9910, 1999.

[14] Th. Hahn (ed.), International Tables of Crystallography, Vol. A. 3rd edition, published for the International Union of Crystallography by Kluwer Academic Publishers, Dordrecht, 1992.

[15] W.C. Hamilton, J.A. Ibers, *Hydrogen Bonding in Solids.* W.A. Benjamin: New York, 1968.

[16] G.A. Jeffrey, W. Saenger, *Hydrogen Bonding in Biological Structures.* Springer, New York, 1991.

[17] L.N. Kuleshova, P.M. Zorky, *Graphical Enumeration of Hydrogen-Bonded Structures.* Acta Cryst. **B36**(1980), 2113-2115.

[18] J.C. MacDonald, *Hydrogen-bonded aggregates: Imidazole as a hydrogen-bond director with applications toward the design of solid-state materials.* PhD Thesis, University of Minnesota, 1993.

[19] W.D.S. Motherwell, G.P. Shields, F.H. Allen, *Visualisation and Characterization of Non-Covalent Networks in Molecular Crystals: Automated Assignment of Graph Set Descriptors for Asymmetric Molecules.* (Manuscript 1999, submitted to Acta Cryst.)

[20] A.F. Wells, *Structural Inorganic Chemistry.* Clarendon Press, Oxford, 1962.

# Index

Any correspondence concerning this paper should be addressed to

Reinhard Pöschel
Institut für Algebra
Technische Universität Dresden
D - 01062 Dresden
Germany
e-mail: poeschel@math.tu-dresden.de

Addresses of members of the group mentioned at page 1:

J. Bernstein
Dept. of Chemistry
Ben-Gurion University of the Negev
Beer Sheva 84120
Israel
e-mail: yoel@bgumail.bgu.ac.il

J. Grell
Moosdorfstr. 13
D - 12435 Berlin
Germany
e-mail: juliana.grell@okay.net

B. Ganter, D. Kuske
Institut für Algebra
Technische Universität Dresden
D - 01062 Dresden
Germany
e-mail: [ganter, kuske]@math.tu-dresden.de