

# What are the possible axial point groups for a molecule of a certain size or formula?

Nike Dattani<sup>1,2,\*</sup> and Sichao Xiong<sup>1,†</sup>

<sup>1</sup>HPQC Labs, Waterloo, Canada

<sup>2</sup>HPQC College, Waterloo, Canada

(Dated: 22nd September 2023)

For each axial point group, we determine *all* possible structures in terms of the number of atoms and the chemical formula. We provide examples of molecules of each size and formula if they are known, and hence we determine the yet-to-be discovered molecules with under five atoms for each point group. We present a set of principles that can be used for proofs about the possibility or impossibility for a molecule of a certain size and/or formula to have a point group symmetry. We also resolve some inconsistencies in a table by P. Atkins, M. Child and C. Phillips, add missing entries to it, and improve it further in a number of ways.

## I. INTRODUCTION

One might wonder why, despite group theory being applied to molecules for about 100 years [19], the questions that are being addressed in this paper do not appear to have been answered yet. Indeed we have asked all living authors of [1–18, 36] if they have ever come across a list of all possible point groups for a 4-atom (tetraatomic) system, and in all cases we were explicitly told that they were unaware of such a list, or we received no response at all. We have determined all possible point groups for systems with *any* number of atoms, and we have gone a step even further by determining all possible point groups for structures with a given chemical formula.

After searching through all of [1–18, 36] thoroughly, the closest work that we found to the results in this paper, is the table of “shapes” found in the 2008 version of the famous “Tables for Group Theory” document by P. Atkins, M. Child and C. Phillips [1]. Nothing similar was found in any of [1–18, 36], and the table of “shapes” in [1] has several shortcomings which we will address and fix in Section III, before we begin presenting the main results of our work in the following sections.

For investigating the novelty of our work, we believe that our thorough examination through all of [1–18, 36] was a fair survey of the known literature on the topic. The 2007 paper [40] reported a thorough survey of known published works that provided character tables for the  $S_8$  and  $D_{8h}$  point groups, and our list [1–18, 36] includes all of those works, in addition to [40] itself, the papers that cited [40], and also thoroughly selected work on the topic that did not provide  $S_8$  and  $D_{8h}$  character tables.

Our paper focuses on *axial* point groups, which are those belonging to the series  $C_n$ ,  $C_{nv}$ ,  $C_{nh}$ ,  $S_{2n}D_n$ ,  $D_{nh}$ ,  $D_{nd}$  (including  $C_s$  which

is equivalent to  $C_{1h}$  and  $C_{1v}$ , and  $C_i$  which is equivalent to  $S_2$ ). For the *polyhedral* point groups  $T$ ,  $T_d$ ,  $T_h$ ,  $O$ ,  $O_h$ ,  $I$ ,  $I_h$ , we will present even more novel results in a forthcoming paper. For example, in all of the literature that we have seen (including [1–18, 36]), the shape with the smallest number of vertices (or the system containing the smallest number of atoms) for which the point group is  $O$ , is the snub cube which has 24 vertices (or atoms), but within the last month we have discovered that it is possible to construct an object with only 20 vertices (or for the atoms in a 20-atom system to be arranged) such that the point group is  $O$  [44]. Likewise, in all of the literature that we have seen, the shape with the smallest number of vertices for which the point group is  $I$ , is the snub dodecahedron which has 60 vertices, but within the last month we have discovered that it is possible to construct an object with only 50 vertices such that the point group is  $I$  [44].

The next section of this paper will explain the context which made *us* feel compelled to answer the questions that this paper addresses; for any readers that wish to skip the motivation section, the rest of the paper is self-contained and does not require knowledge of any of the material presented there. The second paragraph of this introduction already described the contents of Section III. The diagrams in Section III will make it easier to follow Section IV, which contains the foundation for obtaining our main results which are provided in Section V and summarized in Section VI. Overall our results are as follows (all results appear to be novel or unusually difficult to find in existing places):

- In Section III: A table that presents visual examples of structures for all aforementioned axial point groups for  $2 \geq n \geq 6$  and  $n = \infty$ ,
  - in a more clear way than in [1],
  - with inconsistencies in [1] resolved,
  - with more of the  $n = \infty$  shapes presented than in [1],
  - with structures requiring fewer atoms than in [1] for the  $C_n$  and  $D_n$  series,

\* [nike@hpqc.org](mailto:nike@hpqc.org)

† [sichao@hpqc.org](mailto:sichao@hpqc.org)

- with an extra column showing additional structures (in some cases these are more realistic than in [1]) for the  $n = 2$  cases, and
  - with a more logical order, especially for pedagogical reasons, than in [1].
- In Section IV: Properties of some point groups that are very helpful for obtaining or explaining new results, and for pedagogical purposes.
  - In Section V: For each axial point group, we determine
    - the minimum number of atoms (cardinality) required;
    - all possible cardinalities allowed (the minimum cardinality is a special case of this); and
    - all possible chemical formulas allowed (the list of all possible cardinalities is a special case of this).
  - In Section VI: As a consequence of the results in Section V,
    - we were able to list in table X all possible point groups allowed for systems with  $m$  atoms for  $m = 1 - 8$  (and a table with arbitrary values of  $m$  can be generated with the results from V but would take up too much space);
    - we also explicitly listed the specific point groups (and diagrams of example structures) that are possible for systems with the specific formulas  $A_3$ ,  $A_2B$ , and  $ABC$  for triatomic systems and  $A_4$ ,  $A_3B$ ,  $A_2B_2$ ,  $A_2BC$ , and  $ABCD$  for tetraatomic systems (analogous lists for arbitrary chemical formulas can also be generated with the results from V but would take up too much space);
    - we attempted to provide examples of known molecules with each possible point group for each formula  $A_3$ ,  $A_2B$ ,  $ABC$ ,  $A_4$ ,  $A_3B$ ,  $A_2B_2$ ,  $A_2BC$ , and  $ABCD$ , and this led us to the discovery of shapes for which no molecule/complex has yet been synthesized or found, despite the existence of such molecules/complexes being proven to be possible (this was not a surprise for larger structures, but we determined that the minimum number of atoms required for the point group  $D_2$  is 4 but we could not find any example of a 4-atom molecule with the  $D_2$  point group despite the apparent simplicity that would come with such a small molecular size).

Section VI A briefly mentions how the results of this work can be used for forthcoming projects.

## II. MOTIVATION FOR THIS WORK

Our need for answers to the question in the present paper's title, arose during other work. GW100 is a dataset of 100 systems (95 molecules and 5 atoms) which was introduced in a 2015 paper [38] and by the end of 2020, already more than 100 electronic structure methods/codes were used to calculate energies for all systems in the GW100 dataset. It has therefore become a monumental dataset for benchmarking, and efforts were made in 2021 [20] and 2023 [21] to improve the most accurate calculations that had been done on the system until then [22], and therefore to create one of the best known sets of benchmark data in existence for electronic structure calculations.

However, for many of the molecules in the GW100 dataset, calculations are done at geometries that were optimized using density functional theory, which means that the results are dependent on the basis set and density functional used; and it also had an effect on the comparisons that have been made between calculations that were done for the GW100 dataset and experimental data on the same dataset. For many of those molecules, very accurate bond lengths and angles that were determined from high-resolution spectroscopic experiments, are available in places such as the Landolt–Bornstein volumes and related books [23–35] and also in the original papers in which the geometric data was determined.

The books [23–35] provide only the *point group*, a 3D diagram, and usually the minimum number of bond lengths, bond angles and dihedrals necessary to generate all XYZ coordinates or a z-matrix. An example is the entry for  $H_2Si_2$  in [26] (this molecule is not in GW100 but it is useful for illustrating this point) which provides the point group ( $C_{2v}$ ), a 3D shape that is redrawn in Fig. of this paper, and *only* two bond lengths and a dihedral angle:  $r_{Si-Si} = (2.2154 \pm 0.0020)\text{\AA}$ ,  $r_{Si-H} = (1.6680 + 0.0030)\text{\AA}$ ,  $\varphi = (104.22 + 0.30)^\circ$  (the original experimental paper from which this data came [42] and the previous paper on the same molecule [41], also provided no additional bond lengths or angles). But to do electronic structure calculations we need either the XYZ coordinates of all atoms, or a z-matrix for the molecule, and there is no software for obtaining either of these from just the limited information given (point group, 3D diagram, two bond lengths and a dihedral angle). A software for generating a z-matrix from this data would be significantly simpler to write than a software for generating XYZ coordinates of all atoms, but the z-matrix for this system would need at least one bond angle, and the reader might find it surprising that the simplest formula for the H-Si-H angle in  $H_2Si_2$  is (in terms of the provided dihedral angle  $\varphi$ , the provided Si-Si bond length  $r_1$  and the provided Si-H bond length  $r_2$ ):

$$\theta = \arcsin \left( \frac{\sin \left( \frac{\varphi}{2} \right) \sqrt{4r_2^2 - r_1^2}}{r_2} \right). \quad (1)$$

Such a formula is not expected to be found in any compilation of formulas. Determining this formula can be accomplished with some high school level trigonometry (applications of the ordinary cosine law), and applications of the *dihedral cosine law* which is unlikely to be taught in high school or university-level curricula, but can be learned and applied to such small molecules within a day by most researchers; however with the pace of research today, the derivation of such a formula would ideally be done by software rather than by hand. Unfortunately, not even the most advanced symbolic computation or computer algebra software can provide a user with Eq. 1 based on the information given.

Furthermore, the interest today in machine learning and big data analysis has led to the introduction of much larger datasets than GW100, such as GW5000 and OE62 (both from Stuke *et al.* [39]) which involve more than 5000 and 61000 molecules respectively, but the geometries for the molecules in these datasets have issues similar to the ones for the GW100 dataset. This reinforces our motivation for constructing a software that could, for example, convert the thousands of geometries from the Landolt–Börnstein series and related compilations of spectroscopically determined geometry data, into z-matrices or XYZ coordinates.

We believe that as machine learning and big data analysis become more popular over the next century, there will be a rise in interest in being able to categorize molecules (e.g. for training neural networks) and the question of which point groups are possible for molecules of a certain size of formula, will be asked more and more. Anyone seeking the answers to these questions will be able to find them in this paper.

**A software to generate z-matrices (or even XYZ coordinates) for molecules with the data presented in experimental papers, such as for H<sub>2</sub>Si<sub>2</sub>, would require specific code blocks for each molecular shape.** It is possible for two molecules with the same point group and formula to have very different shapes (for example see Fig. X, in which two isomers of H<sub>2</sub>Si<sub>2</sub> are shown, both with the C<sub>2v</sub> point group, but one of them being pyramidal and the other one being planar), so a general-purpose “z-matrix constructor” or “XYZ-coordinates constructor” that uses only bond lengths, bond angles and dihedral angles, would need specific code blocks based on their point groups and formulas, in addition to other characteristics. However, a natural first step towards constructing such a software would be to categorize *n*-atom molecules based on their molecular formulas and point groups (the Landolt–Börnstein volumes and related books [23–35], the CCCBDB [43], and the CRC Handbook of Chemistry and Physics are some examples in which point groups are already used to help describe molecules), and then to write the relevant z-matrix-construction or

XYZ-construction code for each case, adding sub-cases when necessary (e.g. to distinguish between the two types of C<sub>2v</sub> shapes for A<sub>2</sub>B<sub>2</sub> molecules as depicted in Fig. X).

**For *n* atoms, how many cases do we need to consider in order to cover each point group at least once in the software?** That is the essence of the question in the present paper’s title, and as mentioned in the opening paragraph of this Introduction section, it appears that this question has not been answered before, or at least the answer is not easily available. We also go even one step further, by determining all possible *chemical formulas* for a particular point group of an *n*-atom system. For example, in order for a 7-atom molecule to have the point group C<sub>i</sub> its formula needs to be A<sub>2</sub>B<sub>2</sub>C<sub>2</sub>D (with any of the variables B, C, or D being allowed to be a duplicate of one of the previous letters) but a 7-atom molecule whose chemical formula does not have that structure cannot have the point group C<sub>i</sub> (e.g. for reasons that will be explained in this work, there is no arrangement of the atoms of C<sub>3</sub>H<sub>3</sub>N such that the point group of the molecule can be C<sub>i</sub>).

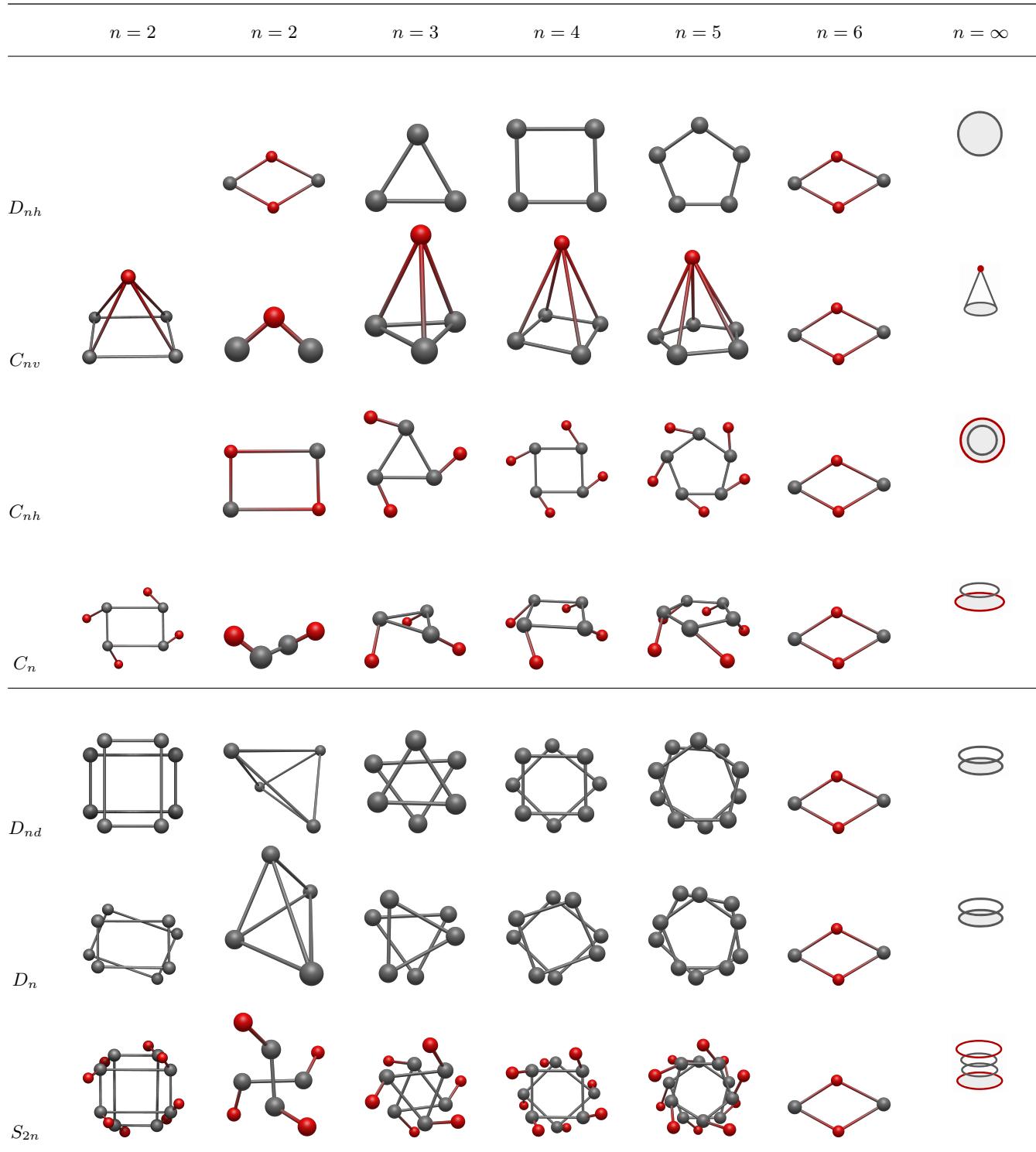
### III. TABLE OF EXAMPLE SHAPES FOR ALL AXIAL POINT GROUP SERIES

The 2008 version of [1] contains a table of “shapes” on Page 37, which appears to be the only table available of its kind (see Section I). For the point groups C<sub>n</sub>, D<sub>n</sub>, C<sub>nv</sub>, C<sub>nh</sub>, D<sub>nh</sub>, D<sub>nd</sub>, and S<sub>2n</sub>, with *n* from 2 to 6, schematic diagrams are presented in attempt to provide an impression of how a molecule with the point group would look. This table has some minor inconsistencies (e.g. one of the lines in the C<sub>4</sub> diagram is black instead of purple, the C<sub>6</sub> diagram has some extra unintended purple lines, and the C<sub>4</sub>, C<sub>5</sub> and C<sub>6</sub> diagrams have lines with inconsistent lengths compared to other lines in the same diagrams and compared to analogous lines in the C<sub>4h</sub>, C<sub>5h</sub> and C<sub>6h</sub> diagrams). Although any inconsistency in such a table can hamper the experience of its readers, and although in these cases some readers are likely to lose a bit of time wondering why some lines aren’t the same size and why there’s some extra lines, most readers will be comfortable to ignore these ones and to simply attribute them to human error.

However, the aforementioned table in [1] has more major consistencies which we would like to address here. For example, objects to which the C<sub>n</sub> and D<sub>nd</sub> series converge (respectively) n = ∞ case is shown

the C<sub>nv</sub> series is shown to converge to an n = ∞ case in which the shape of the object is a cone, but the n = ∞ case for D<sub>nh</sub> is not shown; instead the D<sub>nd</sub> series is shown to converge to what appears to be a single circle, which is the object to which the shown D<sub>nh</sub> series converges, but for the shown D<sub>nd</sub> series it would be more correct for the n = ∞ diagram to depict two equivalent circles on different parallel planes but with

Table I: An improved version of the table presented on Page 37 of [1], depicting example structures for each point group provided in the table. For  $n = 2$ , two families of structures are presented (one with a rectangular “base” and one with a diatomic “base”). For each point group series and for each value of  $n$ , more families of structures also exist, so the examples that are presented are not exhaustive. All objects that are depicted for the  $n = \infty$  point groups satisfy the definition of either  $D_{\infty h}$  or  $C_{\infty v}$ , so they are labeled as such. The colors of the atoms are not intended to have any meaning apart from showing that the grey/black atoms have to all be the same element, and the red atoms have to all be the same element (the red atoms in all of these diagrams can also be the same element as the black atoms without affecting the point group, but it is not necessary for them to be the same). No structure is intended to depict a real molecule.



their centers along the same axis. In our Tabel

#### IV. GENERAL PRINCIPLES FOR POSSIBLE MOLECULES

In this section we will state some general principles that we will use in the following section when formulating proofs for the required numbers of atoms and required formulas for each point group. Some of the statements made in this section may already be familiar or intuitive to the reader, but we have not seen them stated in textbooks about group theory for chemistry [1–18, 36], although the first statement is mentioned in other places, such as [37].

##### Ability to orient a molecule according to the valid point group operations on the molecule:

Any molecule with a rotation axis, called a  $C_n$ -axis, can be oriented in an xyz (Caterian) coordinate system such that for the largest possible  $n$  value in  $C_n$ , the  $z$ -axis is aligned with a  $C_n$  axis, and a  $\sigma_h$  plane is a mirror plane coinciding with the  $xy$ -plane in the coordinate system.

##### Existence of a $C_n$ axis:

The existence of a  $C_n$  axis is only possible if the molecule contains a subset of atoms that are vertices of an  $n$ -sided polygon with equal side-lengths (a regular  $n$ -gon), or if  $n < 3$  we simply need to have  $n$  identical atoms in the molecule. A  $C_n$  axis would go through the center of this polygon (or in the case of  $n = 2$ , it would go through the centre of the line joining the two identical atoms), and since we can call this axis the  $z$ -axis, all atoms that form vertices of the polygon (or the two identical atoms in the case of  $n = 2$ ), would have to have the same  $z$ -coordinates. Therefore, a minimum of  $n$  atoms is needed in order to have a  $C_n$  axis.

If a  $C_n$  axis is the  $z$ -axis, then after a  $C_n$  rotation, an atom that originally had coordinates  $(x, y, z)$ , will have coordinates  $(x', y', z)$  in which the  $z$ -coordinate does not change, because of the definition of a  $C_n$  rotation, and the new coordinates  $(x', y')$  can be different from the old ones  $(x, y)$  but will coincide with the original  $(x, y)$  coordinates of another identical atom in the molecule if the axis is indeed a  $C_n$  axis.

##### Necessary and sufficient conditions for the $D_{nh}$ point group when $n \geq 3$ :

A regular  $n$ -gon with uniform vertices (in a molecule, this would mean the presence of identical atoms at the vertices) has a point group of  $D_{nh}$  (this was already depicted visually in Table I), and any object with the  $D_{nh}$  point group must contain a subset of identical vertices that form a regular  $n$ -gon if  $n \geq 3$ .

##### Action of the $S_n$ operation:

An  $S_n$  operation is by definition a  $C_n$  operation followed by a  $\sigma_h$  reflection operation. If we define the associated  $C_n$ -axis to be the  $z$ -axis, and the  $\sigma_h$  plane to

be the  $xy$ -plane, then if an atom originally had coordinates  $(x, y, z)$ , it will have coordinates  $(x', y', z)$  after the  $C_n$  rotation, and coordinates  $(x', y', -z)$  after the  $\sigma_h$  reflection. So  $S_n$  operation moves atoms from  $(x, y, z)$  to  $(x', y', -z)$ . As described in the section about the existence of a  $C_n$  axis, the new coordinates  $(x', y')$  can be different from the old ones  $(x, y)$  but will coincide with the original  $xy$ -coordinates of another identical atom in the molecule if the molecule indeed has a  $C_n$  axis.

##### Action of the inversion operation, $i$ :

We can always choose center of inversion to be at the origin  $(0, 0, 0)$ , so if an atom is located at  $(x, y, z)$ , the inversion operator will move it to  $(-x, -y, -z)$ .

### V. POSSIBLE NUMBER OF ATOMS AND MOLECULAR FORMULAS FOR EACH AXIAL POINT GROUP

As usual, we only consider the maximum point group. All in order of number of atoms (except  $C_i$  is before others).

#### A. Point groups without axial rotational symmetry

##### $C_s$

Any two atoms form a line, and if all atoms of a molecule are on the same line, then the point group will be promoted to either  $C_{\infty v}$  or  $D_{\infty h}$ . Likewise, a single atom has the point group  $K_h$ , so it is not possible for a molecule with fewer than three atoms to have the point group  $C_s$ . Three atoms is enough for the point group of a molecule to be  $C_s$  though, as in the example ABC. **The minimum number of atoms required is three, and any formula is allowed.**

##### $C_1$

Any three atoms form a plane or a line, and if all atoms of a molecule are on the same plane, then that plane is a reflection/mirror plane. Therefore if a molecule has three or fewer atoms, it will at least have one reflection plane, and would be promoted from  $C_1$  to at least  $C_s$ . Four atoms is enough for the point group of a molecule to be  $C_1$  though, as in the example ABCD. **The minimum number of atoms required is four, and any formula is allowed.**

##### $C_i$

Any five atoms with an inversion center will lie on the same plane (or same line) and will therefore have a reflection/mirror plane. Since the  $C_i$  point group does not have any reflection elements, for any five atom molecule with an inversion center the maximal point group must be larger than  $C_i$ .

Table II: Possible point groups for systems with 1 to 8 atoms.

$N$	Possible groups (not including subgroups)	$K_h$
1		
2		
3	$C_s$	$C_{2v}$
4	$C_s - C_1 \ C_2$	$C_{2h} - C_{2v}C_{3v}$
5	$C_s - C_1 \ C_2$	$C_{2v}C_{3v} \ C_{4v}$
6	$C_s \ C_i \ C_1 \ C_2 \ C_3$	$C_{2h}C_{3h} \ C_{2v}C_{3v} \ C_{4v} \ C_{5v}$
7	$C_s \ C_i \ C_1 \ C_2 \ C_3$	$C_{2h}C_{3h} \ C_{2v}C_{3v} \ C_{4v} \ C_{5v} \ C_{6v}$
8	$C_s \ C_i \ C_1 \ C_2 \ C_3 \ C_4$	$C_{2h}C_{3h}C_{4h} \ C_{2v}C_{3v} \ C_{4v} \ C_{5v} \ C_{6v} \ C_{7v} \ S_4$
		$D_{3h}$
		$D_2 \ D_{2d} \ D_{2h}D_{3h}D_{4h}$
		$D_2 \ D_{2d} \ D_{2h}D_{3h}D_{4h}D_{5h}$
		$D_2 \ D_3 \ D_{2d}D_{3d}D_{2h}D_{3h}D_{4h}D_{5h}D_{6h}$
		$D_2 \ D_3 \ D_{2d}D_{3d}D_{2h}D_{3h}D_{4h}D_{5h}D_{6h}D_{7h}$
		$D_2 \ D_3 \ D_{2d}D_{3d}D_{2h}D_{3h}D_{4h}D_{5h}D_{6h}D_{7h}T_d$
		$O_h \ D_{\infty h}C_{\infty v}$
		$O_h \ D_{\infty h}C_{\infty v}$
		$O_h \ D_{\infty h}C_{\infty v}$
		$O_h \ D_{\infty h}C_{\infty v}$
		$O_h \ D_{\infty h}C_{\infty v}$

If we have three or fewer atoms, then there does not even need to be an inversion center for the atoms to be confined to a plane (or line).

If we have four atoms, and we make the inversion center the origin, the atoms must come in pairs according to the inversion operation  $i$  that was described earlier:  $A = (x_1, y_1, z_1)$  will be paired with  $-A = (-x_1, -y_1, -z_1)$  and  $B = (x_2, y_2, z_2)$  will be paired with  $-B = (-x_2, -y_2, -z_2)$ . Assuming that no two atoms will have the same  $xyz$ -coordinates, none of these four atoms can be on the origin.  $A$  and  $-A$  form a line, and  $B$  and  $-B$  form another line, and that these two lines both intersect at the origin, and two lines that cross are always co-planar.

If we have five atoms, the fifth one must be at the origin, because atoms need to either be paired or at the origin in order for the inversion operation to be valid on the system of atoms. However, the other four atoms form a plane that contains the origin (because the plane is formed by two lines that intersect at the origin). Therefore the fifth atom will also be on the same plane as the other four atoms, and the point group will be promoted beyond  $C_i$ .

Six atoms is enough for a molecule's maximal point group to be  $C_i$ , for example  $\text{C}_2\text{H}_2\text{Br}_2\text{Cl}_2$  with the two carbons removed. The minimum number of atoms required is six, and the formula must be of the form  $\text{A}_2\text{B}_2\text{C}_2\text{D}_2 \dots$  or  $\text{AB}_2\text{C}_2\text{D}_2 \dots$  depending on if the number of atoms is even or odd.

## B. Point groups with axial rotational symmetry

### $D_{nh}$

$$\underline{n=2}$$

If we have any fewer than three atoms, the point group of the system will be promoted to  $K_h$ ,  $D_{\infty h}$  or  $C_{\infty v}$ , however all  $D_{nh}$  point groups with even values of  $n$  (including  $D_{2h}$ ) have an inversion element  $i$ , and the only way for a three atom system to have an inversion element is for it to be linear, which would promote it either to  $D_{\infty h}$  or  $C_{\infty v}$ . If we choose the origin  $(0, 0, 0)$  to be the center of inversion, and one atom is at the position  $(x, y, z)$ , then an identical atom would need to be

at the positon  $(-x, -y, -z)$  and the third atom would need to be at the origin. We therefore would have two identical atoms on a line going through the origin, and a third atom in the center of that line. We therefore would need a minimum of four atoms for the  $D_{2h}$  point group, and indeed four atoms is enough as in the case of  $\text{Br}_2\text{Na}_2$ . The minimum number of atoms required is four, and the most general formula for the minimal structure is  $\text{A}2\text{B}2$ .

$$\underline{n \geq 3}$$

Since a regular  $n$ -gon with identical atoms at each vertex is both necessary and sufficient for the point group of a system of atoms to be  $D_{nh}$  for  $n \geq 3$ , we can conclude the the minimum number of atoms required is  $n$ , and the formula for this minimal structure would be  $\text{A}_n$ . An even number of additional atoms can be added along the  $C_n$  axis in such a way that no new valid symmetry operations are made possible, and such that none of the existing valid symmetry operations are invalidated, provided that these atoms are added in pairs on each side of the plane formed by the  $n$ -gon. One additional atom can also be added at the center of the  $n$ -gon. Therefore the most general formula for a structure with  $n$  atoms plus an even number of additional atoms would be  $\text{A}_n\text{B}_2\text{C}_2\text{D}_2 \dots$ , and for an odd number of additional atoms the most general formula would be  $\text{A}_n\text{BC}_2\text{D}_2 \dots$ .

### $C_{nv}$

$$\underline{n=2}$$

If we have any fewer than three atoms, the point group of the system will be promoted to  $K_h$ ,  $D_{\infty h}$  or  $C_{\infty v}$ , however three atoms is indeed enough for the point group of a molecule to be  $C_{2v}$ , as in the case of  $\text{H}_2\text{O}$ . The minimum number of atoms required is three, and the most general formula is  $\text{A}_2\text{BCD} \dots$  since the extra atoms  $\text{C}, \text{D}, \dots$  can be placed on the  $C_2$  axis without any loss of valid symmetry operations, and without allowing any new valid symmetry operations.

Table III: Possible number of atoms and possible chemical formulas for each point group or point group series.

Point groups	Number of atoms	Possible chemical formula										
		Minimum	1	2	3	4	5	6	7	8	Even # of atoms	Odd # of atoms
$K_h$	1											A
$D_{\infty h}$	2										$A_2B_2C_2D_2 \dots$	$AB_2C_2D_2 \dots$
$C_{\infty v}$	2										$ABCD \dots$	
$D_{nh}$	$n \geq 3$	$n$									$A_nB_2C_2D_2 \dots$	$AB_nC_2D_2 \dots$
$D_{nh}$	$n = 2$	4										
$D_{nd}$	$n \geq 2$	$2n$									$A_{2n}$	$AB_{2n}$
$D_n$	$n \geq 2$	$2n$									$A_{2n}$	$AB_{2n}$
$S_{2n}$	$n \geq 2$	$4n$									$A_{2n}B_{2n}C_2D_2 \dots$	$AB_{2n}C_{2n}D_2E_2 \dots$
$C_{nv}$	$n \geq 2$	$n + 1$									$AB_n$	
$C_{nh}$	$n \geq 2$	$2n$									$A_nB_n$	$AB_nC_n$
$C_n$	$n \geq 2$	$2n$									$A_nB_n$	$AB_nC_n$
$C_s$	3										$ABCD \dots$	
$C_i$	6										$(ABCD \dots)_2$	$A(BCD \dots)_2$
$C_1$	4										$ABCD \dots$	

$n \geq 3$

Since the point group of a regular  $n$ -gon with identical atoms at each vertex is required in order to have the  $C_n$  axis that forms part of the point group  $C_{nv}$  and we have already mentioned in the previous section that the point group of such a structure would be  $D_{nh}$ , we would need at least one additional atom to break any symmetries in  $D_{nh}$  that are not present in  $C_{nv}$ . It turns out that at least one additional atom is not only necessary, but only one atom is sufficient, since if this atom is placed on a  $C_n$  axis that goes through the center of the  $n$ -gon, but the atom is not placed on the plane of the  $n$ -gon, it will break the horizontal mirror plane and any  $S_n$  axes, while maintaining the  $n$  vertical mirror planes that are in  $C_{nv}$ . The only point group with a  $C_n$  axis and  $n$  vertical mirror planes, without having any horizontal mirror planes or  $S_n$  axes, is  $C_{nv}$ , so the pyramid that we have formed must have the point group  $C_{nv}$ . **The minimum number of atoms required is  $n+1$ , and the most general formula is  $A_nBCD \dots$ .**

$C_{nh}$

$n = 2$

We need a minimum of four atoms for the  $C_{2h}$  point group, for the same reason that we gave when we made this statement for the  $D_{2h}$  point group (because the  $C_{2h}$  point group has an inversion element  $i$ ). Indeed four atoms is enough for the point group of a molecule to be  $C_{2h}$ , as in the case of trans-H<sub>2</sub>N<sub>2</sub>. **The minimum number of atoms required is four, and the most general formula for the minimal structure is A2B2.**

$n \geq 3$

From Section IV we know that the system needs to contain a subset of atoms that form a regular  $n$ -gon, in order to have a  $C_n$  axis, but if we only have one regular  $n$ -gon, the point group would be  $D_{nh}$ , so we need to add at least one more atom in order to break the additional symmetries that  $D_{nh}$  compared to  $C_{nh}$ . Adding any number of atoms to the  $C_n$  axis (which we will make our  $z$ -axis) will not break the  $\sigma_v$  planes, so we are forced to add another regular  $n$ -gon. We can add this second regular  $n$ -gon on the same plane as the first one, as long as its vertices do not bisect the exterior angles at the first  $n$ -gon's vertices, and as long as the second  $n$ -gon's vertices are not on the edges of the first  $n$ -gon (both of these placements of the vertices would result in a failure to break the  $\sigma_v$  planes and would not successfully demote the point group from  $D_{nh}$  to  $C_{nh}$ ).

$C_n$   
 $n = 2$

If we have any fewer than three atoms, the point group of the system will be promoted to  $K_h$ ,  $D_{\infty h}$  or  $C_{\infty v}$ ; and if we have only three atoms, they form a plane which will be a mirror plane, meaning that the point group of the system which contains a  $C_2$  axis will be promoted to  $C_{2v}$  or  $D_{3h}$ . Four atoms would be enough for the point group of the system to be  $C_2$  though, as in the case of  $\text{H}_2\text{O}_2$ . **The minimum number of atoms required is four.** In terms of the possible formulas, if we have two atoms at positions  $A$  and  $B$ , then after rotation around the  $C_2$  axis (which we will make the  $z$ -axis, as we described in the previous section), these atoms will be at positions  $A' = (-x_1 - y_1, z_1)$  and  $B' = (-x_2 - y_2, z_2)$  respectively, and for the  $C_2$  operation to be valid, the atoms that were originally at positions  $A$  and  $A'$  need to be identical, and likewise for the atoms that were originally at  $B$  and  $B'$ . **The most general formula is therefore  $\text{A}_2\text{B}_2\text{CD}\dots$  since the extra atoms C,D,\dots can be placed on the  $C_2$  axis without any loss of valid symmetry operations, and if their positions are chosen carefully, no new valid symmetry operations would be introduced.**

$n \geq 3$

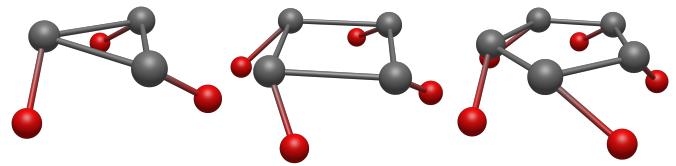
Since the point group of a regular  $n$ -gon with identical atoms at each vertex is required in order to have a  $C_n$  axis, and we have already mentioned in the previous section that the point group of such a structure would be  $D_{nh}$ , we would need at least one additional atom to break any symmetries in  $D_{nh}$  that are not present in  $C_n$ . In order for the  $C_n$  operation to remain valid, we can either add any number of atoms (including just one) on the  $C_n$  axis, or a full set of  $n$  more atoms that form a regular  $n$ -gon with the  $C_n$  axis going through its center and perpendicular to the plane formed by the  $n$ -gon. However, adding any number of atoms to the  $C_n$  axis will not break the  $\sigma_v$  planes, so by adding one or more atoms to the  $C_n$  axis that goes through the center of a regular  $n$ -gon, would form a structure with a  $C_{nv}$  point group, so instead of adding any number of atoms to the  $C_n$  axis, we must add at least one set of  $n$  more atoms forming a regular  $n$ -gon (and we know that this one set is enough for us to form a structure in which  $C_n$  is the point group, as in Fig. 1). The two parallel  $n$ -gons,  $\text{A}_n$  and  $\text{B}_n$ , must lie on different planes so that the molecular is not planar, and therefore would not get promoted to  $C_{nh}$ . **The minimum number of atoms required is  $2n$ , and the most general formula would be  $\text{A}_n\text{B}_n\text{CD}\dots$  since in addition to the  $\text{A}_n\text{B}_n$  structure, extra atoms C,D,\dots can be**

Figure 1: Structures with the point group  $C_n$  and with the minimum number of atoms,  $2n$ .

(a)  $n = 3$

(b)  $n = 4$

(c)  $n = 5$



placed on the  $C_n$  axis without any loss of valid symmetry operations, and if their positions are chosen carefully, no new valid symmetry operations would be introduced.

$D_{nd}$

$n = 2$

The  $D_{2d}$  group does not include any horizontal mirror planes, and since any three atoms can form a plane which can serve as a horizontal mirror plane, we need more than four atoms for the point group of the system to be  $D_{2d}$ . Four atoms is indeed enough for the point group of the system to be  $D_{2d}$  as depicted in the second  $n = 2$  column in Table III, in which **we have two diatomic molecules that are perpendicular to each other, but on different planes.**

$n \geq 3$

From Section IV we know that the system needs to contain a subset of atoms that form a regular  $n$ -gon, in order to have a  $C_n$  axis, but if we only have one regular  $n$ -gon, the point group would be  $D_{nh}$ , so we need to add at least one more atom in order to break the additional symmetries that  $D_{nh}$  compared to  $D_{nd}$ . Adding any number of atoms to the  $C_n$  axis (which we will make our  $z$ -axis) will not break the  $\sigma_v$  planes, and adding any number of atoms on the plane that contains our regular  $n$ -gon, will not break the  $\sigma_h$  plane, so we are forced to add another  $n$ -gon on a plane parallel to the first one, and with its center having the same  $(x, y)$  coordinates as the center of the first regular  $n$ -gon. We now have  $2n$  atoms, and this is indeed enough for the point group of the structure to be  $D_{nd}$ , since if the vertices of the two  $n$ -gons are separated by an angle of  $\pi/n$ , as in the examples provided in Table III, all elements of  $D_{nd}$  are present, without any additional elements.

Table IV: Possible point groups and example systems for all possible chemical formulas of triatomic systems.

Point group	Chemical formula type		
	A <sub>3</sub>	A <sub>2</sub> B	ABC
D <sub>∞h</sub>	C <sub>3</sub>	CO <sub>2</sub>	✗
C <sub>∞v</sub>		N <sub>2</sub> O	HOS
D <sub>3h</sub>	H <sub>3</sub> <sup>+</sup>	✗	✗
C <sub>2v</sub>	O <sub>3</sub>	H <sub>2</sub> O	✗
C <sub>s</sub>	H· · H <sub>2</sub>	HO <sub>2</sub>	HNO

D<sub>n</sub>

n = 2

The  $D_2$  group does not include any horizontal mirror planes, and since any three atoms can form a plane which can serve as a horizontal mirror plane, we need more than four atoms for the point group of the system to be  $D_{2d}$ . Four atoms is indeed enough for the point group of the system to be  $D_2$  as depicted in the second  $n = 2$  column in Table III, in which we have two diatomic molecules that are *not* perpendicular to each other, and are also on different planes.

n ≥ 3

We start with  $D_{nd}$  and remove the  $S_{2n}$  axis by rotating the n-gons by a bit.

S<sub>2n</sub>

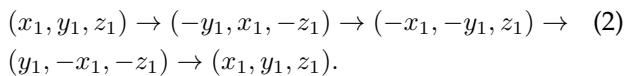
n = 2

From Section IV we know that a  $C_n$  axis can be chosen to be along the  $z$ -axis and a  $\sigma_h$  plane can be chosen to be the  $xy$ -plane in the same coordinate system. With that convention, a  $C_4$  operation will transform the point  $(x, y, z)$  to  $(-y, x, z)$  and the  $\sigma_h$  operation will

Table V: Possible point groups and example systems for all possible chemical formulas of tetratomic systems.

Group	Chemical formula type				
	A <sub>4</sub>	A <sub>3</sub> B	A <sub>2</sub> B <sub>2</sub>	A <sub>2</sub> BC	ABCD
T <sub>d</sub>		✗	✗	✗	✗
P <sub>4</sub>					
D <sub>∞h</sub>		✗		✗	✗
C <sub>4</sub>			C <sub>2</sub> H <sub>2</sub>		
C <sub>∞v</sub>		Unlikely	C <sub>3</sub> H	C <sub>2</sub> HF	HCNO
D <sub>4h</sub>		✗	✗	✗	✗
S <sub>4</sub> <sup>2+</sup>					
D <sub>3h</sub>			✗	✗	✗
S <sub>4</sub>		H <sub>3</sub> B			
D <sub>2h</sub>		✗		✗	✗
Br <sub>2</sub> Na <sub>2</sub>					
D <sub>2d</sub>		✗	✗	✗	✗
S <sub>4</sub>					
D <sub>2</sub>		✗	✗	✗	✗
C <sub>3v</sub>			✗	✗	✗
		H <sub>3</sub> N			
C <sub>2v</sub>			H <sub>2</sub> Si <sub>2</sub>	CH <sub>2</sub> O	✗
S <sub>4</sub>		IF <sub>3</sub>			
C <sub>2h</sub>		✗		✗	✗
			H <sub>2</sub> N <sub>2</sub>		
C <sub>2</sub>		✗		✗	✗
S <sub>4</sub> <sup>2-</sup>			H <sub>2</sub> O <sub>2</sub>		
C <sub>s</sub>				Cl <sub>2</sub> OS	CHFO
		HO <sub>3</sub> <sup>+</sup>	Cl <sub>2</sub> S <sub>2</sub>		
C <sub>1</sub>				H <sub>2</sub> OS	CHBrCl

transform the point  $(-y, x, z)$  to  $(-y, x, -z)$ ; so the  $S_4$  operation which is the  $C_4$  operation followed by the  $\sigma_h$  operation, will transform  $(x, y, z)$  to  $(-y, x, -z)$ . If our first atom is at the position  $(x_1, y_1, z_1)$  then the  $S_4$  transformation will move it to the position  $(-y_1, x_1, -z_1)$ , and an identical atom must be there for the  $S_4$  operation to be valid. Also after the same  $S_4$  transformation, the atom at the position  $(-y_1, x_1, -z_1)$  will move to the position  $(-x_1, -y_1, z_1)$  and an identical atom must be there too. Since no atom has replaced our first atom which was at the position  $(x_1, y_1, z_1)$ , we need to keep applying  $S_4$  transformations and keeping track of the atoms that are necessary to include in our system, until we do successfully replace the first atom. We show below that  $m = 4$  identical atoms are required in order for an  $m$ -atom system to undergo an  $S_4$  transformation and recover the same system:



Unfortunately the coordinates in ?? form the shape that we presented in the second  $n = 2$  column in our  $D_{nd}$  examples in Table III, because  $(x_1, y_1, z_1)$  and  $(-x_1, -y_1, z_1)$  form a line on the  $z = z_1$  plane, whereas  $(-y_1, x_1, -z_1)$  and  $(y_1, -x_1, -z_1)$  form a perpendicular line of the same size but on the  $z = -z_1$  plane. We can add at least one more atom (it does not necessarily need to be homonuclear in comparison to the first set of four atoms) in order to break the  $\sigma_d$  planes that are in  $D_{2d}$ , but adding any number of atoms on the  $C_4$  axis, or on the  $\sigma_h$  plane, or on the  $z = \pm z_1$  planes will not break the  $\sigma_d$  planes. Since we have chosen the convention in which the  $C_4$  axis is the  $z$ -axis, adding atoms on this axis would mean adding atoms with  $(x, y) = 0$ ; since we have chosen the convention in which the  $\sigma_h$  plane associated with the  $S_4$  operation is the  $xy$ -plane, adding atoms on this plane would mean adding atoms with  $z = 0$ ; and regardless of those conventions, any atom on one of the  $\pm z$  planes will have a  $z$ -coordinate of  $z$  or  $-z$ . To break the  $\sigma_d$  planes, we therefore need an atom at a different  $z$ -value such as  $z_2$  and with non-zero  $x$ - and  $y$ -values. Regardless of the  $x$ - and  $y$ -values of this new atom (the fifth atom), the pattern in 2 shows

that we will not recover the new atom's position by applying  $S_4$  operations unless we have another three atoms that are identical to the fifth atom.

The example that we have provided in the second  $n = 2$  column for  $S_{2n}$  in Table III is able to satisfy the above relationships among the coordinates of the atoms; since the  $S_4$  group is generated by the  $S_4$  operation only, and an  $S_4$  operation can successfully be applied to this example, and there are no other symmetry operations valid on this structure which are not in the  $S_4$  point group, so the point group of the 8-atom structure is  $S_4$ . **The minimum number of atoms required is 8 with the formula  $A_4B_4$ .**

$$\underline{n \geq 3}$$

For any positive integer  $n$ , the  $S_{2n}$  operation is a generator of the  $S_{2n}$  point group, so  $C_n$  is in the point group because  $S_{2n}^2 = C_n$ . This means we need a regular  $n$ -gon with the  $C_n$  axis perpendicular to it and passing through its center, but a regular  $n$ -gon is planar and therefore has a  $\sigma_h$  plane, which is an operation that does not exist in any of the  $S_{2n}$  point groups, so at least one atom will be needed to break the  $\sigma_h$  mirror plane.

will not satisfy the  $S_{2n}$  operation unless all atoms are at the origin. You need  $2n$  to have an  $S_{2n}$  axis, but then you'd get  $D_{nh}$ , so you need another  $2n$  to break the  $D_{nh}$ . **The minimum number of atoms required is  $4n$ .**

## VI. SUMMARY OF OUR RESULTS AND OUTLOOK

The results of the previous section are summarized in Tables II and III.

### A. Outlook

Apart from finding molecules to fill in the tables,  $z$ -matrix generator in OpenMolcas.

- 
- [1] P. W. Atkins, *Tables for Group Theory* (London, Oxford University Press, 1970).
  - [2] D. M. Bishop, *Group Theory and Chemistry* (Clarendon Press, Oxford, 1973).
  - [3] R. L. Carter, *Molecular Symmetry and Group Theory* (J. Wiley, New York, 1998).
  - [4] F. A. Cotton, *Chemical Applications of Group Theory* (Interscience, New York, 1963).
  - [5] G. Davidson, *Group Theory for Chemists*, Macmillan physical science (Macmillan, Basingstoke, 1991).
  - [6] M. Diem, *Introduction to Modern Vibrational Spectroscopy* (Wiley, New York, 1993).
  - [7] R. S. Drago, *Physical Methods in Chemistry*, Saunders golden sunburst series (Saunders, Philadelphia, 1977).
  - [8] J. Ferraro and J. Ziomek, *Introduction to Group Theory and Its Application to Molecular Structure* (Plenum Press, 1969).
  - [9] L. Hall, *Group Theory and Symmetry in Chemistry* (McGraw-Hill, 1969).
  - [10] D. Harris and M. Bertolucci, *Symmetry and Spectroscopy: An Introduction to Vibrational and Electronic Spectroscopy* (Oxford University Press, 1978).
  - [11] R. M. Hochstrasser, *Molecular Aspects of Symmetry* (W. A. Benjamin, New York, 1966).

- [12] J. M. Hollas, *Symmetry in Molecules*, Chapman and Hall chemistry textbook series (Chapman and Hall, London, 1972).
- [13] P. W. M. Jacobs, *Group Theory with Applications in Chemical Physics* (Cambridge University Press, Cambridge, 2005).
- [14] Y. Ohrn, *Elements of Molecular Symmetry* (Wiley Interscience, New York, 2000).
- [15] D. S. Urch, *Orbitals and Symmetry*, Penguin library of physical sciences. Chemistry. (Penguin Books, Harmondsworth, 1970).
- [16] J. P. Lowe, *Quantum Chemistry* (Academic Press, New York, 1978).
- [17] J. P. Lowe, *Quantum Chemistry*, 2nd ed. (Academic Press, Boston, 1993).
- [18] A. Vincent, *Molecular Symmetry and Group Theory : a Programmed Introduction to Chemical Applications* (Wiley, Chichester, 1977).
- [19] H. Bethe, *Annalen der Physik* **395**, 133 (1929).
- [20] F. Bruneval, N. Dattani, and M. Setten, *Frontiers in Chemistry* **9** (2021), 10.3389/fchem.2021.749779.
- [21] L. Monzel, C. Holzer, and W. Klopper, *The Journal of Chemical Physics* **158**, 144102 (2023).
- [22] K. Krause, M. E. Harding, and W. Klopper, *Molecular Physics* **113**, 1952 (2015).
- [23] L. E. Sutton and H. J. M. Bowen, *Tables of Interatomic Distances and Configuration in Molecules and Ions* (Chemical Society London, London, 1958).
- [24] L. E. Sutton, *Tables of Interatomic Distances and Configuration in Molecules and Ions* (Chemical Society London, London, 1965).
- [25] N. Vogt, *Structure Data of Free Polyatomic Molecules* (Springer, Cham, Switzerland, 2019).
- [26] K. Kuchitsu, *Structure Data of Free Polyatomic Molecules · Inorganic Molecules* (Springer, 1998).
- [27] K. Kuchitsu, *Structure Data of Free Polyatomic Molecules · Molecules containing One or Two Carbon Atoms* (Springer, 1999).
- [28] K. Kuchitsu, *Structure Data of Free Polyatomic Molecules · Molecules containing Three or Four Carbon Atoms* (Springer, 2001).
- [29] K. Kuchitsu, *Structure Data of Free Polyatomic Molecules · Molecules Containing Five or More Carbon Atoms* (Springer, 2003).
- [30] N. V. K. Kuchitsu, M. Tanimoto, *Structure Data of Free Polyatomic Molecules · Inorganic Molecules Supplement to Volume II/25* (Springer, 2006).
- [31] N. V. K. Kuchitsu, M. Tanimoto, *Structure Data of Free Polyatomic Molecules · Molecules containing One or Two Carbon Atoms Supplement to Volume II/25* (Springer, 2006).
- [32] N. V. K. Kuchitsu, M. Tanimoto, *Structure Data of Free Polyatomic Molecules · Molecules containing Three or Four Carbon Atoms Supplement to Volume II/25* (Springer, 2007).
- [33] N. V. K. Kuchitsu, M. Tanimoto, *Structure Data of Free Polyatomic Molecules · Molecules containing Five or More Carbon Atoms Supplement to Volume II/25* (Springer, 2007).
- [34] M. T. K. Kuchitsu, N. Vogt, *Structure Data of Free Polyatomic Molecules · Molecules Containing No Carbon Atoms and Molecules Containing One or Two Carbon Atoms* (Springer, 2014).
- [35] M. T. K. Kuchitsu, N. Vogt, *Structure Data of Free Polyatomic Molecules · Molecules Containing Three or Four Carbon Atoms and Molecules Containing Five or More Carbon Atoms* (Springer, 2014).
- [36] E. B. Wilson, *Molecular vibrations : the theory of infrared and Raman vibrational spectra* (McGraw-Hill, New York, 1955).
- [37] V. Bakken and T. Helgaker, *The Journal of Chemical Physics* **117**, 9160 (2002).
- [38] M. J. van Setten, F. Caruso, S. Sharifzadeh, X. Ren, M. Scheffler, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, C. Yang, F. Weigend, J. B. Neaton, F. Evers, and P. Rinke, *Journal of Chemical Theory and Computation* **11**, 5665 (2015), pMID: 26642984.
- [39] A. Stuke, C. Kunkel, D. Golze, M. Todorović, J. T. Margraf, K. Reuter, P. Rinke, and H. Oberhofer, *Scientific Data* **7**, 58 (2020).
- [40] R. B. Shirts, *Journal of Chemical Education* **84**, 1882 (2007).
- [41] M. Bogey, H. Bolvin, C. Demuynck, and J. L. Destombes, *Phys. Rev. Lett.* **66**, 413 (1991).
- [42] M. Bogey, H. Bolvin, M. Cordonnier, C. Demuynck, J. L. Destombes, and A. G. Császár, *The Journal of Chemical Physics* **100**, 8614 (1994).
- [43] R. D. Johnson, “*NIST Computational Chemistry Comparison and Benchmark Database*,” (2022), NIST Standard Reference Database Number 101.
- [44] Private conversation between Nike Dattani and Gabe Cunningham (2023).