

# What point groups can a molecule of a certain size or formula have?

Nike Dattani,<sup>1,2,\*</sup> Sichao Xiong,<sup>1,†</sup> and Hemanth H.<sup>1,‡</sup>

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

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

(Dated: 27 August 2023)

We list all possible point groups for 1, 2, 3, 4 and  $n$  atoms as well as for  $n$ -atom molecules with specified formulas. For each point group, we give examples of molecules of each size and formula if they are known, and hence we determine the yet-to-be discovered molecules with under 5 atoms for each point group. We establish a small 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.

## I. INTRODUCTION

One might wonder why, despite group theory being applied to molecules for about 100 years (Bethe 1929), 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 () 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.

Our need for answers to the question in the present paper’s title, arose during other work. GW100 (MJ van Setten 2015) is a dataset of 100 systems (95 molecules and 5 atoms) which was introduced in a 2015 paper and by the end of 2020, already more than 100 electronic structure methods/codes were used to calculate energies for all systems in the dataset. It has therefore become a monumental dataset for benchmarking, and efforts were made in (2021 Bruneval) and (2023 Monzel, Wim Klopper) to improve the most accurate calculations that had been done on the system until then (and used to benchmark the other 100+ calculations from various research groups) (Harding 2016). But the geometries were wrong and we need to be able to generate them from places like the Landolt-Bornstein volumes. Such volumes provide only the point group, a 3D diagram, and a minimal set of bond lengths, angles and dihedrals. For example H<sub>2</sub>Si<sub>2</sub> (not in GW100 but good for illustrating this point) gives only two bond lengths and a dihedral, 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. 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 you might find it surprising that the simplest formulas for the HHSi? angle in H<sub>2</sub>Si<sub>2</sub> is... Determining this bond angle can be accomplished with some high school level trigonometry (applications of the ordinary cosin law), and applica-

tions of the *dihedral cosine law* which is unlikely to be taught in high school or university-level curricula, and can be learned and applied to such small molecules within a day by most researchers, but with the pace of research today, the derivation of such a formulas as in Eqs. () would ideally be done by software rather than by hand.

Furthermore, the interest today in machine learning and big data analysis has lead to the introduction of much larger datasets such as GW5000, OE31? and OE62 (all from [Rinke]) which involve 5000, 31000 and 61000 molecules respectively, but they have the same geometry problem (crystal structures vs spectroscopic geometries). This reinforces our motivation for constructing a software that could, for example, convert the ~5000 geometries from Landolt-Bornstein 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 (and those seeking the answers 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 (r1=,r2=,phi=) 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 to have very different shapes (e.g. C<sub>2v</sub> for H<sub>2</sub>Si<sub>2</sub> and C<sub>2v</sub> planar and maybe another 4-atom C<sub>2v</sub>, all have 4 atoms but one is pyramidal, one is planar, etc.), so a “z-matrix generator” or “xyz coordinates generator” would need specific code blocks for n-atom molecules not just based on their point group, but also based on other characteristics (perhaps their graph, formula, etc.), but a natural first step towards constructing such a software would be to categorize n-atom molecules based on their point groups (as done in the Landolt Bornstein series and a lot of other places, check CRC), and to construct code for each case. **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 everyone tell us that it hasn’t been done yet. We take this even one step further, by looking at the possible formulas, for example for a 7-atom molecule to have the point group Ci you need its formula to be A2B2C2D (with any of B,C,D being allowed to be a duplicate of one of the previous letters) but a molecule with any other formula cannot have the point group Ci (for example A2B3C2 would no work unless B and C are A, which would bring us back to the previously mentioned formula in the case in which A=B=C=D).

In the initial stages of the development of our z-matrix constructor, we found that the Landolt-Bornstein series had all these molecules, but we realized that we missed the tadpole structure because we know that there’s 6 connected 4-vertex graphs. Likewise we tried to see if we had representation from all possible point groups, and this required us to first determine the possible point groups for 4-atom molecules (which surprisingly we couldn’t find, and we found that experts who wrote books on the topics were also unaware). In doing this, we discovered that D2 and D2d are possible but not yet found, so this may be an avenue for further experimental research.

Another motivation was geometry optimization. Roland Lindh and others have discovered that you can get fewer cycles, but if point group symmetry is included, then we can have a smaller Wilson-B matrix.

In this paper we provide general principles, 3-atom and 4-atom possible groups, formulas and examples, n-atom groups and formulas for minimal structures, all of which seems to be novel. We skip the polyhedral groups  $T, T_d, T_h, O, O_h, I, I_h$  since we discovered something with fewer vertices than the snub cube that has  $O$  symmetry.

## II. GENERAL PRINCIPLES FOR POSSIBLE MOLECULES

The following general principles will be used. They might seem obvious, but we don’t see them in textbooks (maybe the first one is in textbooks).

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

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

### Existence of a $C_n$ axis:

This 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). The  $C_n$  axis would be in the center of this polygon, and if it is the  $z$ -axis, then all atoms that form vertices of the polygon,

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 candidate  $C_n$  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  $xy$ -coordinates of another identical atom in the molecule if the molecule indeed has a  $C_n$  axis.

### Necessary and sufficient conditions for the $D_{nh}$ point group:

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  $D_{nh}$ , and any object with the  $D_{nh}$  point group must contain a subset of identical vertices that form a regular  $n$ -gon.

### 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  $(-x, -y, -z)$ .

## III. MINIMUM NUMBER OF ATOMS AND POSSIBLE MOLECULAR FORMULAS FOR EACH POINT GROUP

As usual, we only consider the maximum point group. All in order of number of atoms, except for D2h which is paired with Dnh. What about D2, C2, C2v, C2h, D2d?

$$C_s$$

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  $C_i$ .

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  $C_2H_2Br_2Cl_2$  with the two carbons removed. **The minimum number of atoms required is six, and the formula must be of the form  $A_2B_2C_2D_2 \cdots$  or  $AB_2C_2D_2 \cdots$  depending on if the number of atoms is even or odd.**

$$D_{nh}, n = 2$$

1 or two atoms would be linear, and a 3-atom molecule can't have an inversion center unless it's linear, so you need at least 4 atoms, as in the case of  $Br_2Na_2$ .

$$D_{nh}, n \geq 3$$

Regular n-gon

$$C_{nv}, n \geq 3$$

Pyramid formed from regular n-gon is  $D_{nh}$ , so add one more atom that's not on the same plane but still on the  $C_n$  axis that goes through the center of the n-gon to break the horizontal mirror plane and the  $S_n$  axis, while maintaining the  $n$  vertical mirror planes. 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  $C_{nv}$ . **The minimum number of atoms required is  $n+1$ .**

$$C_n, n \geq 3$$

Pyramid formed from regular n-gon is  $D_{nh}$ , we need to add at least one atom to break all mirror planes. To preserve the  $C_n$  axis we need to either add any number of atoms on the  $C_n$  axis, or at least  $n$  atoms that are not on the  $C_n$  axis (forming a regular n-gon with the  $C_n$  axis in the centre of it). However, adding any atoms on the axis will not break the vertical mirror planes, so we must add at least  $n$  more atoms and form a second concentric regular n-gon around the  $C_n$  axis, but it cannot be on the same plane as the original n-gon because then it would still have the horizontal mirror plane and the point group would be promoted at least to  $C_{nh}$ , and the new n-gon would have to be rotated from the original one in order to avoid  $D_{nh}$  in the case where the side-lengths are the same (horizontal mirror plane is just halfway between the planes of the two n-gons), and  $C_{nv}$  where the two regular n-gons have their own distinct side-lengths. **The minimum number of atoms required is  $2n$ .**

$$C_{nh}, n \geq 3$$

As in the case of  $C_n$ , any structure with fewer than  $2n$  atoms will not have enough atoms to break the vertical planes of a  $D_{nh}$  n-gon, and without maintaining the  $C_n$  axis. Having two concentric n-gons on the same plane, with one of them rotated from the other in order to avoid having  $D(2n)h$ , would be one way to construct

a  $C_{nh}$  molecule with the minimum number of atoms. Another way would be to add the extra  $n$  atoms to the plane containing the original regular  $n$ -gon, and no rotation would be necessary if we are careful about how we place the bonds. **The minimum number of atoms required is  $2n$ .**

$$D_n, n \geq 3$$

**The minimum number of atoms required is  $2n$ .**

$$D_{nd}, n \geq 3$$

**The minimum number of atoms required is  $2n$ .**

$$S_{2n}, n \geq 2$$

Since  $S_{2n}$  is a generator of the  $S_{2n}$  point group (or is it  $S_n$ ) then  $C_n$  is in the point group because  $C_n = S_n^2$ . This means we need a regular polygon, but that wouldn't satisfy the  $S_{2n}$  operation unless all atoms were to be 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  $4n$ .**

#### IV. SUMMARY OF OUR RESULTS

Point groups		Number of atoms							
		Minimum	1	2	3	4	5	6	7 8
$K_h$		1	✓	✓	✓	✓	✓	✓	✓
$D_{\infty h}$		2	✓	✓	✓	✓	✓	✓	✓
$C_{\infty v}$		2	✓	✓	✓	✓	✓	✓	✓
$D_{nh}$	$n \geq 3$	$n$	✓	✓	✓	✓	✓	✓	✓
	$n = 2$	4	✓	✓	✓	✓	✓	✓	✓
$D_{nd}$	$n \geq 2$	$2n$	✓	✓	✓	✓	✓	✓	✓
$D_n$	$n \geq 2$	$2n$	✓	✓	✓	✓	✓	✓	✓
$S_{2n}$	$n \geq 2$	$4n$	✓	✓	✓	✓	✓	✓	✓
$C_{nv}$	$n \geq 2$	$n + 1$	✓	✓	✓	✓	✓	✓	✓
$C_{nh}$	$n \geq 2$	$2n$	✓	✓	✓	✓	✓	✓	✓
$C_n$	$n \geq 2$	$2n$	✓	✓	✓	✓	✓	✓	✓
$C_s$		3	✓	✓	✓	✓	✓	✓	✓
$C_i$		6	✓	✓	✓	✓	✓	✓	✓
$C_1$		4	✓	✓	✓	✓	✓	✓	✓

#### A. Possible point groups for 3-atom (triatomic) molecules

Point group	Chemical formula type		
	$A_3$	$A_2B$	$ABC$
$D_{\infty h}$	$C_3$	$CO_2$	✓
$C_{\infty v}$	Unlikely	$N_2O$	HOS
$D_{3h}$	$H_3^+$	✓	✓
$C_{2v}$	$O_3$	$H_2O$	✓

#### B. Possible point groups for 4-atom (tetraatomic) molecules

Point group	Chemical formula type				
	$A_4$	$A_3B$	$A_2B_2$	$A_2BC$	$ABCD$
$T_d$	$P_4$	✓	✓	✓	✓
$D_{\infty h}$		✓	$C_2H_2$	✓	✓
$C_{\infty v}$	Unlikely			$H_2BN$	
$D_{4h}$	$S_4^{2+}$	✓	✓	✓	✓
$D_{3h}$		$H_3B$	✓	✓	✓
$D_{2h}$		Unlikely	$Br_2Na_2$	✓	✓
$D_{2d}$	Possible but unknown	✓	✓	✓	✓
$D_2$	Possible but unknown	✓	✓	✓	✓
$C_{3v}$		$H_3N$	✓	✓	✓
$C_{2v}$			$H_2Si_2$	✓	✓
$C_{2h}$		Unlikely	$H_2N_2$	✓	✓
$C_2$		Unlikely	$H_2O_2$	✓	✓
$C_s$				$Cl_2OS$	
$C_1$				$H_2OS$	

\* [nike@hpqc.org](mailto:nike@hpqc.org)  
† [sichao@hpqc.org](mailto:sichao@hpqc.org)  
‡ [hemanth@hpqc.org](mailto:hemanth@hpqc.org)

\* [nike@hpqc.org](mailto:nike@hpqc.org)  
† [sichao@hpqc.org](mailto:sichao@hpqc.org)  
‡ [hemanth@hpqc.org](mailto:hemanth@hpqc.org)

[illegible]