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

Nike Dattani,[1,2,*] Sichao Xiong,[1,†] and Hemanth H.[1,‡]

[1]*HPQC Labs, Waterloo, Canada*
[2]*HPQC College, Waterloo, Canada*

(Dated: 26 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 H2Si2 (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 softawre 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 H2Si2 is... Determining this bond angle can be accomplished with some high school level trigonometry (applications of the ordinary cosin law), and applioca-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 softawre 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.

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 H2Si2 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. C2v for H2Si2 and C2v planar and maybe another 4-atom C2v, 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 symemtry 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 (Caterian) coordinate system such that for the largest possible n value in C_n, the z-axis is aligned wtih a C_n axis, and the σ_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 identaical 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 σ_h reflection operation. If we define the associated C_n-axis to be the z-axis, and the σ_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 coordiantes $(x', y', -z)$ after the σ_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 the new coordinates (x', y') can be different from the old ones (x, y) but will coincide with the original xy-coordinates of another identaical atom in the molecule if the molecule indeed has a C_n axis.

Action of the inversion operation, i:

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

* nike@hpqc.org
† sichao@hpqc.org
‡ hemanth@hpqc.org

* nike@hpqc.org
† sichao@hpqc.org
‡ hemanth@hpqc.org