

# What point groups can a molecule of a certain size 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: 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.

The

the questions that are addressed in this paper were not answered in earlier work, despite

cess, and was first released in September 2014) processors (28 cores per node). [Check if I used 768G at any point.. I did for xvcc part of 9z CCSD(T) and same for ncc 9z for CCSDT, the 9z MRCC CCSD(T) was done with 512G on wob149, 8Z cc files need to be found, everything else either requested for only 511G, or only required at most 511G, but best way to double-check would be to look at the jobID's (though not necessary if it's too much trouble, since I just checked everything (except 8z still to do) and am certain enough that 9z was the only place where > 512G was needed ... but wait .. some calcs did use 32G, so specific nodes may need to be checked].

## B. Software

For the small basis sets (cc-pVXZ for  $X = 2$  to 6), the following programs from the CF0UR v2.00beta package [? ], which was released in March 2014, were used: vmo1 for the 1- and 2-electron integrals, vscf for the Hartree-Fock calculations, vtran for the AO to MO transformation, and intprc and int for processing and printing the integrals into ASCII format for the software package MRCC [? ] to read and do the coupled cluster calculations. The July 2016 release of MRCC was used.

For the cc-pV7Z, cc-pV8Z, and cc-pV9Z basis sets, the aforementioned programs in CF0UR v2.00beta cannot treat the necessary  $k$ -type,  $l$ -type and  $m$ -type Gaussian-type orbitals involved, so development versions of vmo1, vscf, and vtran written mainly by Devin Matthews, were de-bugged and tested as part of this work. MRCC was still used to do CCSD and CCSD(T) calculations for all three of these basis sets, but due to RAM limitations (? or size of fort.55?), CCSDT (which was only done for cc-pV7Z), was done using the program ncc from the CF0UR v2.1 package [? ], which was released in July 2019. CCSD and CCSD(T) calculations for all three of these basis sets were also done with ncc, ecc, and vcc of CF0UR v2.1 in addition to the aforementioned calculations done in MRCC, and the difference among the different implementations of CCSD and CCSD(T) was never more than  $2 nE_h$  (nano-

## II. IMPLEMENTATION

### A. Hardware

Despite the largest calculations in this work being state-of-the-art by a large margin, they were actually done on relatively old hardware (all calculations were done in 2019 on hardware from 2014). While this would have made the calculations significantly slower than they could have been (in terms of wall clock time), it meant that access to the hardware was more readily available than if using any of the newer (and more in-demand by other scientists) resources available to the author. This was especially important, since there was a lot of de-bugging involved, as the 1- and 2-electron integrals, SCF (self-consistent field) calculations, and AO to MO transformation were done with a new code, which had not been tested on such basis sets before. Furthermore, the particular nodes on which the calculations were done, happened to have 55T of node local storage, which was in fact necessary for the calculations using the biggest basis sets, and is not available on any of the other resources available to the author.

All calculations were done on Ontario's SHARC-NET cluster called wobbie, on compute nodes containing two sockets, each with a 14-core Xeon E5-2697 v3 (from the Haswell-EP model, which used a 22nm pro-

Hartrees), which makes us certain that our use of `ncc` for CCSDT with cc-pV7Z and MRCC for all smaller basis sets, was not problematic.

### C. Data

The cc-pVXZ basis sets for  $X = 2$  to 6, were obtained from the EMSL Basis Set Exchange before January 2019 (the input file containing the exact exponents and contraction coefficients used, and the file’s version history, can be found in the file `GENBAS` on my [GitHub page](#)). For  $X = 2$  to 5 the basis sets were originally published in [? ], and for  $X = 6$  the basis sets were originally published in [? ].

For  $X = 7$  to 9, the basis sets were obtained privately from David Feller, who built them for [1]. They were then converted to CFOUR’s input format and inserted into the same `GENBAS` file located in [1].

## III. METHODOLOGY AND RESULTS

### A. Hamiltonian

The standard non-relativistic molecular electronic Hamiltonian with the clamped nucleus (Born-Oppenheimer) approximation and the point-sized nucleus approximation, was used (this Hamiltonian has been called NR-CPN (non-relativistic, with clamped and point-sized nucleus) in the past [1]) but with the two deepest core electrons frozen and only the remaining 8 electrons correlated. This constitutes a very standard “frozen core approximation” which can be corrected *a posteriori* since the difference between the all-electron and the frozen-core energies tends to converge more rapidly with respect to basis set size and electron correlation treatment than the total frozen-core energy does.

### B. Basis sets

The electronic wavefunction was modeled using a linear combination of atomic orbitals (LCAO) in which Gaussian-type basis sets were used for the atomic orbitals. A series of these basis sets was used, from the cc-pVXZ family, which was first proposed by Dunning [? ]. The references for the specific cc-pVXZ basis sets used, were given in Section II C.

The basis sets were contracted using a general contraction scheme (`CONTRACTION = GENERAL` in CFOUR), and the total number of spatial orbitals for each basis set (after contraction) are given in Table II.

### C. 1- and 2-electron integral calculations

The transformed integrals were calculated with the CFOUR keyword `XFORM_TOL` set to 18 as opposed to the default value of 11, meaning that only integrals with values smaller than  $10^{-18}$  would be neglected as opposed to the default tolerance threshold of  $10^{-11}$  (effectively, no integrals are ignored).

### D. Hartree-Fock calculations

RHF (restricted Hartree-Fock) calculations were done with the spin multiplicity set to 1 (to obtain a singlet state). A core Hamiltonian initial guess was used, with no restriction on the occupation numbers in terms of spatial symmetry within the  $C_{2v}$  point group which was used, but at convergence the final occupation numbers were always 6 for the  $A_1$  irrep (3 electrons with spin  $+\frac{1}{2}$  and 3 with  $-\frac{1}{2}$ ), 2 for the  $A_2$  and 2 for the  $B_1$  irreps (in each case, 1 electron with each spin direction), and 0 for  $B_2$ .

The CFOUR keyword `SCF_CONV` was set to 7 for the cc-pV9Z basis set (and maybe 8z), meaning that SCF iterations were made until the maximum change in a density matrix element between iterations was smaller than  $10^{-7}$ , and `SCF_CONV` was set even higher for the smaller basis sets (meaning that the convergence requirement was even more strict for the smaller basis sets). The energies were converged even better... =9 for 7Z, but don’t know 8z so can’t say it was >9 for the rest. probably it was 8 for 8Z, it was 10 for 6Z, 11 for 5Z, 12 for 4Z, 3Z and 2Z. I wonder why it changed by literally 1 value for each Z until 4Z, was it because it wasn’t converging otherwise?).

All other settings, and details of how well the energies were converged, can be found in the output files (which include the corresponding input files at the beginning) which are included at `AI_Energies` (symbolic link from `Paper_H2O` folder to `AI_Energies` folder so that if files at `AI_Energies` change paths it’s okay?). The final SCF energies for each basis set are given in Table I.

### E. Electron correlation calculations

Unrestricted coupled cluster calculations (and perturbative corrections as defined in [? ]) were done with as many excitations included in the cluster operator as possible with a given basis set and within 768GiB of RAM. All iterative coupled cluster calculations calculated in MRCC were converged such that the norm of the residual vector was below  $7 \times 10^{-6}$  in the last iteration, and in all cases the final iteration resulted in the energy changing by no more than  $0.13 \mu E_h$ .

Table I.

	cc-pV2Z	cc-pV3Z	cc-pV4Z	cc-pV5Z	cc-pV6Z	cc-pV7Z	cc-pV8Z	cc-pV9Z
RHF	-76.026 778 80	-76.057 139 16	-76.064 804	-76.067 060	-76.067 376	-76.067 426	-76.067 449	-76.067 455
CCSD	-76.237 998 10	-76.324 545 75	-76.350 801	-76.359 519	-76.362 323	-76.363 594	-76.364 296	-76.364 659
CCSD(T)	-76.241 033 59	-76.332 192 23	-76.359 794	-76.369 041	-76.372 019	-76.373 362	-76.374 100	-76.374 481
CCSDT	-76.241 195 72	-76.332 268 49	-76.359 807	-76.368 997	-76.371 941	-76.373 264	-76.373 990	-76.374 364
CCSDT(Q)	-76.241 679 91	-76.332 623 49	-76.360 217	-76.369 431	-76.372 384	-76.373 706	-76.374 432	-76.374 806
CCSDTQ	-76.241 650 58	-76.332 587 70	-76.360 183	-76.369 394	-76.372 345	-76.373 667	-76.374 393	-76.374 768
CCSDTQ(P)	-76.241 665 00	-76.332 602 77	-76.360 189	-76.369 401	-76.372 352	-76.373 674	-76.374 400	-76.374 774
CCSDTQP	-76.241 665 92	-76.332 606 22	-76.360 198	-76.369 410	-76.372 361	-76.373 683	-76.374 409	-76.374 783
CCSDTQP(H)	-76.241 667 98	-76.332 606 77	-76.360 199	-76.369 410	-76.372 361	-76.373 683	-76.374 410	-76.374 784
CCSDTQPH	-76.241 668 32	-76.332 606 94	-76.360 199	-76.369 410	-76.372 361	-76.373 683	-76.374 410	-76.374 784
CCSDTQPH(S)	-76.241 668 38	-76.332 606 99	-76.360 199	-76.369 410	-76.372 361	-76.373 683	-76.374 410	-76.374 784
CCSDTQPHS	-76.241 668 43	-76.332 607 05	-76.360 199	-76.369 410	-76.372 362	-76.373 684	-76.374 410	-76.374 784
CCSDTQPHS(O)	-76.241 668 43	-76.332 607 05	-76.360 199	-76.369 410	-76.372 362	-76.373 684	-76.374 410	-76.374 784
FCI = CCSDTQPHSO	-76.241 668 45	-76.332 607 07	-76.360 199	-76.369 410	-76.372 362	-76.373 684	-76.374 410	-76.374 784

Table II. Basis sets used and their sizes after contraction

Basis set	Number of spatial orbitals		
	O	H	H <sub>2</sub> O
cc-pV2Z	14	5	24
cc-pV3Z	30	14	58
cc-pV4Z	55	30	115
cc-pV5Z	91	55	201
cc-pV6Z	140	91	322
cc-pV7Z	204	140	484
cc-pV8Z	285	204	693
cc-pV9Z	385	286	957

The converged coupled-cluster energies are given in bold text in Table I, while energies given in plain text indicate the result of adding the energy difference between the given row and the preceding row, in the preceding column, to the preceding energy of a given column (the energies added are either 0 or shown as a non-zero value in Table III). The energies in bold text are exact whereas the energies in plain text are approximate.

In Table IV, final energy estimates for various levels of coupled-cluster calculations are shown for the largest basis set used (cc-pV9Z). **They should be the same as in Table 1 but they're not!.** The RHF and CCSDT energies come from direct calculations using only the cc-pV9Z basis set, whereas the rest of the rows contain energies that were obtained by adding the numbers in Table III that were in boldface text, and any digits which are uncertain due to the approximation described in the previous paragraph, are underlined.

#### F. Estimation of $\pm 2\mu E_h$ uncertainty for FCI energy with cc-pV9Z

Now we describe the patterns.

## IV. DISCUSSION

### A. Bottlenecks:

Bottlenecks:

- speed of CCSDTQ for 7Z (cfour will take 150 days? with 768G or would require 1.5T which doesn't allow more than 30 days, and NCC doesn't allow restarts ... not sure about MRCC... Hanrath program unavailable.)

- speed of CCSDTQP for 5Z (lack of hand-coded CCSDTQP, lack of parallel I/O in MRCC, lack of Hanrath code, speed of CC-QMC to get error down).

- ability to do CCSDTQ(P) for 5Z/6Z, or CCSDT(Q) for 7Z, 8Z, 9Z (MRCC can't restart perturbative calculations).

- getting higher than CCSDTQP is not necessary for any basis sets that it's not already possible for, and getting higher than CCSDTQ does not seem necessary for anything beyond 7Z?

---

\* [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)

Table III. Explain what's in the brackets

	cc-pV2Z	cc-pV3Z	cc-pV4Z	cc-pV5Z	cc-pV6Z	cc-pV7Z	cc-pV8Z	cc-pV9Z
CCSD(T)-CCSD	-3035.	-7646.	-8993.	-9522.	-9696.	-9768.	-9805	-9822
CCSDT-CCSD(T)	-162.	-76.	-13.	44.	77.	98.	110	117
CCSDT(Q)-CCSDT	-484.	-355.	-410.	-434.	-442.			
CCSDTQ-CCSDT(Q)	29.3	35.8	34.4	36.7	38.4	<b>38.9(0.5)</b>		
CCSDTQ(P)-CCSDTQ	-14.4	-15.1	-6.5					
CCSDTQP-CCSDTQ(P)	-0.9	-3.5	-9.0					
CCSDTQP-CCSDTQ	-15.3	-18.5	<b>-15.5(3.0)</b>					
CCSDTQP(H)-CCSDTQP	-2.1	-0.5						
CCSDTQPH-CCSDTQP(H)	-0.3	-0.2						
CCSDTQPH-CCSDTQP	-2.4	<b>-0.7(1.7)</b>						
CCSDTQPH(S)-CCSDTQPH	-0.053							
CCSDTQPHS-CCSDTQPH(S)	-0.053							
CCSDTQPHS(O)-CCSDTQPHS	-0.003							
CCSDTQPHSO-CCSDTQPHS(O)	-0.019							
CCSDTQPHSO-CCSDTQP	<b>-0.128</b>							

Table IV.

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

	cc-pV9Z
RHF	-76.067 455 2
CCSDT	-76.374 364 2
CCSDTQ	-76.373 961
CCSDTQP	-76.373 976
CCSDTQPH	-76.373 977
FCI = CCSDTQPHO	-76.373 977