PAPER • OPEN ACCESS

A Comparison of Medium-Sized Basis Sets for the Prediction of Geometries, Vibrational Frequencies, Infrared Intensities and Raman Activities for Water

View the article online for updates and enhancements.

A Comparison of Medium-Sized Basis Sets for the Prediction of Geometries, Vibrational Frequencies, Infrared Intensities and Raman Activities for Water

Brian C. Ferrari

Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando FL 32816

E-mail: `Brian.Ferrari@ucf.edu`

Chris J. Bennett

Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando FL 32816
Department of Chemistry, University of Central Florida, 4111 Libra Drive, Orlando FL 32816
Florida Space Institute, 12354 Research Parkway, Partnership 1 Building, Suite 214, Orlando FL 32826
Department of Chemistry & Biochemistry, Georgia Institute of Technology, 901 Atlantic Drive, Atlanta, GA 30332

E-mail: `Christopher.Bennett@ucf.edu`

Abstract.

Optimized geometries, vibrational frequencies, as well as infrared intensities and Raman activities were calculated for water (H_2O) utilizing popular quantum mechanical approaches. Here, density functional theory (DFT) calculations were performed using the B3LYP (Becke, three-parameter, Lee-Yang-Parr) functional, as well as *ab initio* calculations using second-order Møller-Plesset (MP2) perturbation theory and coupled-cluster with single, double and perturbative triple excitations [CCSD(T)] levels of theory were used. We assess and benchmark the performance of 69 different atomic orbital basis sets including various popular families of medium-sized basis sets typically of two to four zeta quality and differing levels of augmentation by polar and diffuse functions. The basis sets range from the commonly adopted Pople-style (6-31G & 6-311G), Dunning's correlation consistent (cc-pV(n+d)Z & aug-cc-pV(n+d)Z, as well as Truhlar's calendar variations, Jensen's polarization consistent (pc-n & aug-pc-n), Ahlrichs (def2-...), Sapporo's and Karlsruches as well as atomic natural orbitals (ANOs) such as NASA Ames (ANOn), Neese-style, and Roos-style. We also compare several basis sets specifically designed to calculate vibrational and electronic properties, including the Sadlej-pVTZ (and LPol-X families), as well as SNS families of Barone. The results are compared to experimental values where available, or calculations performed with 5 or 6 zeta-level (e.g., cc-pV6Z). The performance of each family of basis sets is discussed in terms of their accuracy (and pitfalls), as well as computational resource scaling and efficiency. The Def2 basis family performs very well overall, yielding more accurate results with lower runtimes than traditional basis sets. 'May' basis sets also provide accurate predictions of vibrational frequencies at significantly lower costs. Raman activities can be accurately calculated using MP2 under harmonic approximation with several 'spectroscopic' families performing well.

1. Introduction

Predictions of molecular properties, such as geometries, vibrational frequencies, infrared intensities and Raman activities have become powerful tools for chemists and physicists performing spectroscopic analysis of substances. Particularly involving the characterization

or quantification of unstable species, or radicals which are produced *in situ* during experiments, where experimental values are often not available [1, 2]. Commonly employed quantum mechanical approaches to calculation these properties include density functional theory (DFT) and *ab initio* methods, of which extensive studies have been done investigating accuracy of their predictions [3–11]. However, only a few studies exist systematically comparing the performance of a limited number of differing atomic orbital basis sets to accurately calculate these properties [12]. Moreover, careful selection from the plethora of basis set choices now easily accessible (for example through the EMSL database, https://bse.pnl.gov/bse/portal) can lead to optimization of the approach implemented based upon computational resources and required accuracy levels [12–14]. Since computational resources are often constrained, Pople-style basis sets are often employed since they are often not demanding, and widely available in commercial packages, however, they are often out-performed by other basis sets at similar computational cost. On the other hand, correlation consistent basis sets are widely available and - although computationally heavy - they do systematically converge on accurate results when either 5/6 zeta basis sets are used, or complete basis set extrapolation schemes are incorporated (where each zeta, ζ, typically approximately doubles the number of atomic orbitals). Since computational costs for these calculations scale with N^4 to N^7, the runtimes for even medium-sized molecules under 5/6 zeta become unfeasible. However, these large calculations don't necessarily produce the most accurate predictions. The results from several basis set families can be extrapolated to complete basis set (CBS) limit via a series of calculations performed at lower zeta level [15–17]. This study compares the results of several medium-sized basis sets across three levels of theory, in order to identify which basis sets can be reliably utilized to predict molecular properties (such as geometries, frequencies, IR intensities and Raman activities) maximizing the precision/accuracy for a given set of computational resources. Here, we only focus on preliminary results for water, but the findings are in line with those observed for other molecules we have investigated. Both DFT and *ab initio* calculations were perfomed, using Becke's three parameter exchange funtional [18] along with Lee-Yang-Parr's correlation functional [19] (B3LYP), second-order Møller-Plesset (MP2) perturbation theory [20] and coupled-cluster with single, double and perturbative triple excitations [CCSD(T)] [21, 22]. For each level of theory 69 different atomic orbital basis sets were used varying from Pople-style (6-31G & 6-311G) [23], Dunning's correlation consistent (cc-pV(n+d)Z & aug-cc-pV(n+d)Z) [24], as well as Truhlar's calendar variations [25–30], Jensen's polarization consistent (pc-n & aug-pc-n) [31–33], Alrich's (def2-...) [34], Sapporo's and Karlsruhe's, as well as atomic natural orbitals (ANOs) such as NASA Ames (ANOn), Neese-style [35], and Roos-style. We also compare several basis sets specifically designed for calculating vibrational and electronic properties, including the Sadlej-pVTZ (and LPol-X families) [36,37], as well as SNS & NO7 families of Barone [38,39] and a new generation of Thakkar DZ basis sets, NLO-1 [40]. Calculated values are compared to both experimental values [41–43] where available, as well as various convergences found either running large zeta calculations or through extrapolation. Though many medium-sized basis sets may converge on the solutions provided by extrapolation schemes and high-zeta results, this does not guarantee, and in fact rarely results in the prediction of molecular properties being accurately calculated due to inherent flaws in the level of theory taken. Therefore, minor improvements at the cost of significantly larger runtimes are rarely worthwhile. It is of note that this study does not account or attempt to correct for errors arising from inadequacies in the levels of theory described here, although it is possible to do so (e.g., by correcting for the truncation of the theoretical approaches, core-correlation effects, or relativistic effects which need to be taken into account to reproduce frequencies to within 1 cm^{-1}). However, due to fortuitous cancellation of errors, the CBS limits, or cc-pV6Z results are often within 5-10 cm^{-1} of the corrected values [44]. Here, we report on which basis sets can be reliably utilized to predict molecular properties (such as geometries, frequencies, IR intensities and Raman activities) maximizing the precision/accuracy

for a given set of computational resources. The need for high zeta basis sets, or extrapolated results can often be avoided, if the basis sets are more carefully considered for the property being evaluated, and more accurate results reliably obtained at a lower computational cost. We briefly address the benefits of using anharmonic corrections to determine vibrational frequencies versus the use of scaling factors for a few select cases.

2. Methods

Calculations were done using the General Atomic and Molecular Electronic Structure System, GAMESS(US) [45–47], using the 14 FEB 2018 (R1) version of Gamess(US). All calculations were run using spherical harmonics; Pople-style basis sets were typically ran using cartesian functionals unless they incorporated f-type polarization functions, where spherical harmonics were used. The segmented versions of polarization consistent basis sets were used, it should also be noted that within the Def2 family QZVPD is equivalent to QVZPPD and that QZVP is equivalent to QVZPP. Tight convergence criteria were used throughout (e.g., integral cutoff 10^{-12}, primitive cutoff 10^{-25}, 10^{-7} SCF convergence, gradient convergence 10^{-6} for geometries). Hessian and Raman calculations were run with semi-numerical (B3LYP and MP2) or fully-numerical gradients [CCSD(T)] under the double harmonic approximation with projected frequencies and two displacements in each cartesian direction during force field computations. The frozen-core approximation was used for both MP2 and CCSD(T) calculations. B3LYP calculations were run using an "army grade" pruned grid (JANS=2) with roughly 71,000 grid points and 155 radial points per atom. In general calculations were run on four processors, with only a few exceptions. Larger basis sets such cc-pV5Z, cc-pV6Z and Pcseg-4 were run with fully numerical gradients for all theory methods, and required calculations to be run on a single processor. For a few special cases, the anharmonic frequencies were calculated using Vibrational self-consistent field (VSCF) [48]. These calculations were run with third order mode couplings, ten grid points, with second order degenerate perturbation theory corrections. Extrapolated complete basis set (CBS) limits were obtained using the following two-point linear extrapolation formula generated by Halkier *et al* [49].

$$E_{X,Y} = \frac{E_X^{corr} X^3 - E_Y^{corr} Y^3}{X^3 - Y^3} \tag{1}$$

Where X and Y are the basis set's zeta value, with $Y = X - 1$, CBS limits were found using the highest zeta basis sets calculated per basis "family". Error analysis was done by taking the average error between the calculated values and either experimental data or CBS limits.

3. Results and Discussion

<div align="center">

Table 1: Molecular Geometry Predictions,
Bond Lengths are in Angstroms (Å), Bond Angles are in Degrees (°), {1} is fastest time rank

</div>

	Optimization	Bond Length{Time Rank}			Bond Angle		
	Experimental [42]	0.9578			104.4776		
#	Basis Set	B3LYP	MP2	CCSD(T)	B3LYP	MP2	CCSD(T)
1	cc-pV6Z	0.9606{67}	0.9582{68}	0.9582{69}	105.120	104.338	104.422
2	cc-pV5Z	0.9605{63}	0.9580{63}	0.9580{61}	105.091	104.286	104.373
3	Aug-cc-pVQZ	0.9608{55}	0.9589{55}	0.9589{55}	105.110	104.269	104.365
4	Jul-cc-pVQZ	0.9608{53}	0.9588{52}	0.9589{52}	105.109	104.263	104.360
5	Jun-cc-pVQZ	0.9608{50}	0.9586{48}	0.9587{49}	105.114	104.209	104.301
6	May-cc-pVQZ	0.9608{48}	0.9584{45}	0.9585{45}	105.122	104.219	104.306
7	cc-pVQZ	0.9605{45}	0.9577{43}	0.9579{43}	104.877	104.018	104.116
8	Aug-cc-pVTZ	0.9621{41}	0.9614{39}	0.9616{39}	105.082	104.109	104.180
9	Jul-cc-pVTZ	0.9620{39}	0.9614{37}	0.9616{37}	105.100	104.125	104.188
10	Jun-cc-pVTZ	0.9620{35}	0.9609{34}	0.9611{35}	105.110	104.063	104.118
11	May-cc-pVTZ	0.9620{32}	0.9603{31}	0.9604{31}	105.052	103.893	103.920
12	cc-pVTZ	0.9616{29}	0.9591{27}	0.9594{29}	104.517	103.518	103.582
13	Aug-cc-pVDZ	0.9651{21}	0.9659{20}	0.9665{21}	104.741	103.873	103.937
14	Jul-cc-pVDZ	0.9650{14}	0.9653{14}	0.9658{14}	105.035	104.232	104.256
15	Jun-cc-pVDZ	0.9686{8}	0.9665{8}	0.9672{8}	104.045	102.835	102.798
16	May-cc-pVDZ	0.9686{10}	0.9665{8}	0.9672{9}	104.045	102.835	102.798

17	cc-pVDZ	0.9689{4}	0.9649{5}	0.9663{6}	102.720	101.902	101.912
18	6-311++G(3df,3pd)	0.9614{34}	0.9589{33}	0.9592{34}	105.040	104.037	104.086
19	6-311G(2df,2pd)	0.9610{30}	0.9569{26}	0.9575{23}	103.936	103.109	103.211
20	6-311++G(2d,2p)	0.9612{14}	0.9583{17}	0.9584{20}	105.138	104.279	104.272
21	6-311G(2d,2p)	0.9617{13}	0.9573{13}	0.9578{12}	103.902	103.246	103.313
22	6-311++G(d,p)	0.9624{11}	0.9600{8}	0.9607{13}	105.043	103.358	103.299
23	6-311+G(d,p)	0.9624{6}	0.9600{5}	0.9607{10}	105.047	103.374	103.314
24	6-311G(d,p)	0.9624{6}	0.9586{5}	0.9596{4}	103.752	102.300	102.306
25	6-31+G(d,p)	0.9653{4}	0.9632{1}	0.9633{5}	105.744	105.421	105.338
26	pc4	0.9606{64}	0.9576{66}	0.9577{63}	105.128	104.347	104.430
27	aug-pc3	0.9606{57}	0.9581{57}	0.9583{59}	105.137	104.370	104.449
28	pc3	0.9606{49}	0.9578{49}	0.9580{48}	105.139	104.341	104.415
29	aug-pc2	0.9610{40}	0.9590{38}	0.9593{38}	105.145	104.306	104.340
30	pc2	0.9606{23}	0.9580{30}	0.9582{30}	105.149	104.177	104.165
31	aug-pc1	0.9661{19}	0.9662{17}	0.9662{18}	104.831	103.922	104.000
32	pc1	0.9686{3}	0.9650{1}	0.9654{3}	103.865	103.553	103.518
33	Def2-QZVPD	0.9607{53}	0.9585{51}	0.9586{47}	105.121	104.237	104.321
34	Def2-QZVP	0.9604{46}	0.9580{46}	0.9581{44}	105.099	104.223	104.301
35	Def2-TZVPPD	0.9613{37}	0.9598{36}	0.9600{34}	105.127	104.070	104.118
36	Def2-TZVPP	0.9610{26}	0.9586{27}	0.9589{26}	104.893	103.823	103.859
37	Def2-TZVPD	0.9636{31}	0.9644{31}	0.9650{32}	105.111	104.377	104.401
38	Def2-TZVP	0.9630{17}	0.9622{21}	0.9630{19}	105.236	104.597	104.548
39	Def2-SVPD	0.9644{18}	0.9643{14}	0.9649{15}	105.370	104.444	104.476
40	Def2-SVP	0.9670{2}	0.9625{1}	0.9638{2}	103.078	102.438	102.465
41	SPK-AQZP	0.9607{56}	0.9585{56}	0.9586{57}	105.157	104.351	104.440
42	SPK-QZP	0.9604{47}	0.9580{47}	0.9582{46}	105.082	104.234	104.321
43	SPK-ATZP	0.9612{42}	0.9598{40}	0.9602{40}	105.102	104.227	104.266
44	SPK-TZP	0.9612{33}	0.9588{35}	0.9593{33}	104.355	103.493	103.565
45	SPK-ADZP	0.9645{22}	0.9640{23}	0.9645{22}	104.790	103.996	104.047
46	SPK-DZP	0.9674{8}	0.9637{8}	0.9647{7}	102.772	102.267	102.287
47	KTZVPP	0.9610{24}	0.9586{27}	0.9589{26}	104.893	103.823	103.859
48	KTZV	0.9750{1}	0.9757{1}	0.9781{1}	109.965	110.567	110.162
49	Sadlej-LPolX-fl	0.9616{69}	—[a]	0.9621{69}	105.038	—[a]	104.364
50	Sadlej-LPolX-fs	0.9613{60}	0.9615{60}	0.9622{60}	105.048	104.431	104.489
51	Sadlej-LPolX-dl	0.9622{54}	0.9607{54}	0.9611{54}	105.055	104.420	104.454
52	Sadlej-LPolX-ds	0.9616{44}	0.9614{50}	0.9621{51}	105.038	104.265	104.296
53	Sadlej-pVTZ	0.9675{28}	0.9689{25}	0.9697{28}	104.592	103.541	103.610
54	N07T	0.9644{27}	0.9665{22}	0.9676{25}	104.754	103.609	103.632
55	N07D	0.9644{12}	0.9641{8}	0.9645{11}	104.643	103.679	103.733
56	SNST	0.9649{24}	0.9653{24}	0.9659{24}	104.702	103.853	103.896
57	SNSD	0.9644{20}	0.9643{17}	0.9648{17}	104.662	103.748	103.802
58	NLO-1	0.9751{16}	0.9840{14}	0.9847{16}	106.514	105.024	105.088
59	Roos-Aug-TZ-ANO	0.9607{61}	0.9581{61}	0.9583{62}	105.128	104.251	104.310
60	Roos-Aug-DZ-ANO	0.9616{43}	0.9609{44}	0.9613{50}	105.054	104.333	104.371
61	NASA-Ames-ANO2	0.9606{65}	0.9587{64}	0.9588{65}	105.042	104.249	104.340
62	NASA-Ames-ANO1	0.9629{49}	0.9614{58}	0.9617{56}	104.923	103.928	104.000
63	NASA-Ames-ANO0	0.9647{36}	0.9638{41}	0.9642{41}	104.393	103.662	103.602
64	Neese-ANO-AVQZ	0.9606{68}	0.9588{67}	0.9588{68}	105.144	104.330	104.414
65	Neese-ANO-VQZ	0.9608{66}	0.9589{65}	0.9589{66}	105.127	104.303	104.391
66	Neese-ANO-AVTZ	0.9609{62}	0.9581{62}	0.9583{64}	105.085	104.213	104.286
67	Neese-ANO-VTZ	0.9628{59}	0.9617{59}	0.9619{58}	105.087	104.069	104.130
68	Neese-ANO-AVDZ	0.9633{52}	0.9605{53}	0.9606{53}	105.053	104.275	104.335
69	Neese-ANO-VDZ	0.9648{38}	0.9645{42}	0.9646{42}	104.584	103.861	103.794

[a] Removed due to convergence issues

Table 2: Harmonic Vibrational Frequencies in Wavenumbers (cm^{-1})

Hes	ω_1{Time Rank}			ω_2			ω_3		
Exp [5]	3832.17			1648.47			3942.53		
The [44]	3835.89			1649.39			3946.65		
#	B3LYP	MP2	CCSD(T)	"	"	"	"	"	"
1	3806.0{67}	3845.0{68}	3837.3{ }[a]	1631.4	1633.0	1651.3[a]	3907.4	3971.2	3947.2[a]
2	3805.6{62}	3848.9{62}	3840.1{ }[a]	1630.0	1635.6	1653.4[a]	3906.9	3974.3	3949.3[a]
3	3801.6{55}	3839.8{55}	3831.2{54}	1628.4	1632.2	1649.6	3903.3	3965.6	3940.8
4	3802.2{53}	3840.5{52}	3831.8{51}	1628.8	1632.5	1649.9	3903.8	3966.4	3941.5
5	3801.7{51}	3843.4{50}	3834.4{49}	1628.4	1635.6	1652.5	3903.3	3968.8	3943.5
6	3801.9{47}	3845.0{47}	3836.1{45}	1628.3	1636.5	1653.4	3903.5	3970.3	3945.1
7	3803.1{45}	3855.2{43}	3844.5{42}	1633.9	1642.5	1658.8	3903.4	3977.8	3951.5
8	3793.8{40}	3822.0{39}	3810.9{38}	1626.4	1628.1	1645.4	3896.4	3947.8	3920.0
9	3798.2{38}	3824.6{37}	3813.7{36}	1627.3	1629.9	1647.4	3901.1	3952.0	3924.4
10	3799.6{34}	3832.6{33}	3821.9{32}	1626.1	1633.8	1650.8	3903.3	3960.0	3932.7
11	3799.9{33}	3843.4{32}	3833.6{31}	1625.9	1642.0	1661.0	3902.8	3967.0	3940.1
12	3797.8{27}	3855.5{29}	3841.0{29}	1638.8	1651.9	1668.4	3898.0	3975.8	3945.6
13	3791.7{19}	3803.4{19}	3787.0{20}	1617.8	1622.1	1637.7	3901.4	3937.6	3904.9
14	3794.9{14}	3818.9{13}	3804.2{13}	1624.2	1632.4	1648.8	3904.6	3952.7	3922.1
15	3772.7{8}	3842.6{8}	3824.1{8}	1617.5	1638.4	1655.8	3875.2	3964.2	3930.4
16	3772.7{8}	3842.6{8}	3824.1{9}	1617.5	1638.4	1655.8	3875.2	3964.2	3930.4
17	3748.4{5}	3851.8{5}	3821.7{5}	1658.1	1677.6	1689.8	3849.8	3971.1	3927.6
18	3810.8{39}	3869.8{34}	3855.9{34}	1625.7	1623.7	1640.5	3909.7	3989.1	3959.7
19	3805.3{27}	3890.6{26}	3871.5{25}	1655.0	1670.3	1685.6	3901.1	4005.0	3971.4
20	3816.4{21}	3861.2{19}	3848.9{21}	1637.8	1659.7	1678.6	3917.8	3982.7	3953.9
21	3803.7{16}	3872.5{16}	3852.7{15}	1666.0	1686.1	1701.1	3900.1	3988.6	3953.7
22	3813.3{13}	3880.1{11}	3860.8{12}	1601.6	1629.7	1648.0	3918.7	3999.6	3965.1
23	3813.2{11}	3879.6{7}	3860.2{11}	1601.8	1629.7	1647.9	3918.7	3999.4	3964.9
24	3804.3{7}	3896.6{5}	3868.9{6}	1636.8	1667.2	1681.7	3902.3	4007.4	3966.7
25	3806.8{4}	3867.1{3}	3855.8{4}	1602.5	1622.8	1639.4	3928.7	4013.9	3987.5
26	3805.5{64}	—[b]{64}	—[b]{61}	1630.8	1654.2	1677.8	3907.1	3968.6	3945.0
27	3802.6{68}	3822.4{58}	3811.5{58}	1628.0	1632.9	1649.5	3904.4	3948.2	3921.1

28	3802.9{49}	3822.7{49}	3811.5{46}	1627.6	1635.1	1651.7	3904.6	3948.7	3921.2
29	3805.3{41}	3846.2{38}	3832.8{37}	1626.3	1632.5	1649.8	3908.9	3972.2	3942.6
30	3806.6{27}	3859.3{28}	3846.8{30}	1623.5	1637.7	1656.3	3910.9	3982.8	3954.2
31	3780.2{18}	3805.5{18}	3797.2{18}	1615.1	1630.7	1648.5	3887.6	3940.3	3916.1
32	3762.0{7}	3854.7{2}	3840.2{3}	1627.5	1648.7	1664.7	3873.5	3991.1	3962.4
33	3805.0{50}	3849.3{51}	3840.0{48}	1628.4	1635.3	1652.2	3906.8	3975.6	3949.9
34	3807.2{46}	3855.9{45}	3846.5{43}	1629.6	1637.0	1654.1	3908.3	3980.7	3955.0
35	3808.2{37}	3838.1{35}	3826.0{35}	1627.2	1636.3	1653.3	3911.2	3963.5	3935.2
36	3807.3{31}	3853.7{29}	3841.1{28}	1631.9	1646.5	1664.1	3910.2	3976.3	3947.7
37	3783.4{26}	3798.0{27}	3783.1{26}	1622.9	1620.4	1636.9	3884.8	3925.2	3893.5
38	3782.9{19}	3819.0{21}	3801.3{19}	1616.2	1613.3	1629.4	3888.1	3946.1	3912.4
39	3816.8{15}	3843.4{14}	3828.0{14}	1584.2	1596.4	1612.3	3923.1	3974.9	3944.5
40	3788.9{3}	3895.3{3}	3867.3{2}	1638.2	1654.5	1666.9	3884.0	4011.3	3971.1
41	3805.6{57}	3845.2{56}	3836.5{56}	1627.5	1631.4	1648.7	3907.1	3971.4	3946.5
42	3807.1{47}	3852.9{46}	3843.2{44}	1629.8	1635.4	1652.5	3908.7	3978.2	3952.5
43	3799.9{42}	3835.5{42}	3821.3{41}	1625.4	1626.4	1643.6	3901.5	3961.3	3930.8
44	3799.1{36}	3852.4{36}	3835.5{33}	1648.0	1655.2	1670.8	3895.5	3971.9	3939.3
45	3778.4{22}	3797.3{22}	3780.4{22}	1604.9	1617.9	1634.3	3888.1	3931.8	3899.3
46	3759.5{6}	3844.6{12}	3817.3{7}	1653.6	1665.2	1679.3	3864.5	3972.5	3932.2
47	3807.3{30}	3853.7{29}	3841.1{27}	1631.9	1646.5	1664.1	3910.2	3976.3	3947.7
48	3610.8{1}	3598.2{1}	3551.4{1}	1567.0	1601.4	1610.4	3775.9	3787.2	3725.3
49	3798.6{69}	—c	4031.4{62}	1634.4	—c	2126.1	3898.0	—c	4126.8
50	3803.0{60}	3811.3{60}	3793.1{59}	1638.6	1642.1	1658.6	3904.4	3943.8	3910.5
51	3796.7{54}	3820.2{54}	3805.0{53}	1630.4	1629.3	1645.5	3899.0	3949.7	3918.3
52	3799.1{44}	3811.9{48}	3793.4{49}	1629.8	1640.1	1655.9	3900.3	3937.6	3902.9
53	3794.7{24}	3797.9{25}	3779.9{24}	1633.8	1642.9	1658.3	3908.4	3935.8	3900.8
54	3770.5{25}	3775.4{24}	—b	1631.4	1622.9	—b	3871.7	3901.6	—b
55	3812.6{8}	3841.5{8}	3829.4{10}	1640.9	1650.4	1665.3	3921.1	3975.9	3948.0
56	3785.8{23}	3810.4{23}	3794.3{23}	1629.8	1632.6	1647.6	3896.4	3946.3	3914.3
57	3808.6{17}	3834.6{17}	3820.6{17}	1633.6	1640.4	1655.4	3918.0	3968.3	3938.1
58	3716.6{12}	3620.4{15}	3597.6{16}	1600.3	1605.1	1623.5	3862.4	3790.4	3750.2
59	3807.0{61}	3850.1{61}	3838.1{60}	1627.6	1639.3	1656.2	3906.6	3974.4	3946.0
60	3807.9{43}	3828.9{44}	3812.6{47}	1635.6	1644.9	1660.5	3908.3	3954.9	3921.7
61	3803.7{65}	3843.7{65}	3835.1{65}	1631.6	1636.0	1652.7	3906.6	3972.5	3947.5
62	3798.1{56}	3838.6{57}	3827.2{55}	1629.5	1637.9	1653.3	3901.7	3969.1	3942.3
63	3810.2{32}	3854.4{40}	3835.0{39}	1638.1	1642.0	1659.1	3927.6	3991.6	3958.2
64	3804.4{68}	3840.9{67}	3832.9{64}	1628.6	1631.3	1649.0	3904.8	3967.5	3943.2
65	3802.1{66}	3842.2{66}	—b	1628.5	1634.2	—b	3907.0	3973.2	—b
66	3803.6{63}	3851.9{63}	3842.0{63}	1629.3	1640.0	1656.8	3904.8	3977.6	3951.6
67	3800.2{59}	3833.9{59}	3823.8{57}	1624.3	1633.2	1649.3	3904.2	3965.2	3939.0
68	3796.7{52}	3845.0{53}	3832.9{52}	1628.8	1637.9	1653.1	3900.3	3972.3	3944.1
69	3811.5{35}	3847.8{41}	3833.0{40}	1633.8	1635.3	1654.0	3931.3	3989.1	3959.2

[a] These results were taken from Tew *et al.* (2007)

[b] An error in the calculation of this vibrational frequency caused this particular result to be invalid

[c] Removed due to convergence issues

Table 3: Infrared Intensities in $Km \cdot mol^{-1}$ and Raman Activities in $\mathring{A}^4 amu^{-1}$

—	Infrared Intensities						Raman Intensities					
Exp [5]	2.93		62.5		41.7		111 ± 12		0.9 ± 0.2		19 ± 2	
—	IR_1		IR_2		IR_3		R_1		R_2		R_3	
#	B3LYP	MP2	"	"	"	"	"	"	"	"	"	"
3	4.63{57}	6.09{57}	76.10	73.17	63.27	78.36	98.26	109.50	0.81	0.83	25.4	22.1
4	5.20{52}	6.68{51}	76.19	73.04	63.65	77.85	87.63	97.56	0.71	0.75	25.0	21.9
5	5.21{55}	6.80{54}	76.03	73.50	63.67	78.52	87.75	96.98	0.72	0.75	25.0	21.8
6	5.16{49}	6.69{48}	76.56	72.64	63.16	77.23	88.55	97.72	0.71	0.79	25.4	22.2
7	3.89{45}	6.55{43}	72.79	69.11	51.46	68.30	78.51	78.28	2.75	2.78	26.0	23.5
8	4.59{41}	5.56{39}	75.71	71.68	62.83	75.46	96.82	107.62	1.02	1.09	25.6	22.8
9	5.86{36}	7.44{33}	77.62	72.50	62.87	74.21	80.43	87.64	0.86	0.97	26.1	23.6
10	5.91{39}	7.61{37}	77.20	73.81	63.08	75.77	79.33	87.53	0.86	0.91	25.8	23.1
11	6.53{34}	9.57{32}	69.58	63.43	60.65	73.11	78.84	81.66	3.79	4.00	28.5	26.2
12	3.16{29}	5.74{29}	69.46	64.52	40.54	55.20	71.40	69.55	4.34	4.25	26.1	24.8
13	4.00{19}	4.14{19}	71.28	67.46	60.41	67.02	96.28	104.28	1.86	1.96	25.5	24.2
14	11.75{8}	16.28{9}	65.74	60.84	59.66	66.22	80.40	81.55	4.52	4.82	37.5	36.3
15	6.33{14}	9.13{13}	81.06	79.23	55.36	65.70	72.34	74.73	1.35	1.39	29.9	27.6
16	11.75{9}	16.28{9}	65.74	60.84	59.66	66.22	80.40	81.55	4.52	4.82	37.5	36.3
17	2.84{4}	6.56{5}	55.64	56.98	19.70	32.64	73.03	68.32	6.24	5.83	33.3	33.8
18	4.45{37}	6.42{34}	72.10	67.09	59.30	72.70	94.11	96.06	2.20	2.28	26.5	23.5
19	2.72{26}	5.12{26}	68.08	64.04	33.76	49.29	69.05	65.45	4.11	4.13	23.8	22.8
20	7.23{22}	9.84{19}	70.95	65.33	61.76	73.54	82.31	82.95	4.54	4.74	28.3	26.1
21	2.06{17}	3.60{15}	65.23	61.36	29.89	43.09	73.50	69.74	4.71	4.75	24.5	23.8
22	9.01{13}	12.66{11}	66.60	56.65	56.58	61.67	83.54	81.66	7.45	7.94	34.7	34.1
23	9.48{11}	12.96{7}	67.05	57.13	57.07	61.97	80.38	78.95	7.77	8.09	34.2	33.8
24	3.41{7}	5.82{6}	57.32	50.17	24.14	32.17	71.51	66.12	7.90	7.92	30.6	31.1
25	6.34{5}	10.08{3}	91.20	92.87	57.07	66.68	80.34	78.73	3.36	3.48	35.1	34.0
27	4.64{60}	6.09{60}	76.26	72.91	63.30	78.90	97.55	111.47	0.67	0.75	25.5	21.8
28	5.07{50}	7.04{50}	74.88	71.04	62.08	78.03	86.39	95.65	1.63	1.76	26.4	23.0
29	4.66{40}	5.60{38}	75.92	71.88	63.41	75.58	97.96	109.88	0.74	0.81	25.7	22.4
30	5.76{28}	9.42{28}	71.38	66.09	54.89	68.70	76.92	76.53	4.31	4.38	29.9	27.7
31	4.57{16}	4.79{18}	75.72	72.47	65.99	73.73	96.78	104.35	0.69	0.65	25.3	22.3
32	7.35{2}	13.81{3}	74.19	78.31	42.98	63.56	76.17	70.14	5.08	4.57	37.1	35.7

33	4.61{54}	6.08{52}	76.07	71.96	63.09	77.54	97.95	111.63	0.77	0.84	25.4	22.1
34	5.47{46}	8.14{45}	74.45	70.25	58.55	74.74	77.32	82.65	2.49	2.57	26.6	23.7
35	4.66{38}	5.59{36}	75.81	70.19	62.89	74.20	96.54	107.74	0.78	0.89	25.5	22.6
36	4.46{30}	7.23{29}	69.75	64.39	48.72	62.46	76.52	75.31	4.51	4.44	27.8	26.0
37	4.08{32}	5.13{27}	77.76	76.03	61.03	74.42	98.89	108.40	0.66	0.73	26.1	23.3
38	4.38{21}	8.83{21}	80.64	80.97	43.76	60.96	78.80	74.43	5.09	4.98	32.9	31.5
39	7.05{15}	7.22{14}	77.96	73.99	75.28	82.02	93.91	103.84	0.92	0.90	24.5	22.5
40	4.79{2}	9.87{2}	55.38	58.08	26.65	42.65	74.23	69.19	7.01	6.44	32.7	33.1
41	4.74{59}	6.25{58}	76.38	73.41	63.66	79.15	94.50	109.18	0.81	0.83	25.3	22.0
42	4.97{48}	7.96{47}	74.87	71.14	56.06	73.35	78.04	80.52	2.58	2.61	26.8	23.9
43	4.73{42}	5.95{42}	75.66	71.85	63.25	76.44	90.30	110.88	0.87	0.89	25.6	22.4
44	3.08{35}	4.94{35}	68.75	64.64	44.39	59.51	75.62	78.28	3.78	3.88	25.1	23.2
45	4.03{20}	4.36{22}	71.41	68.02	60.28	68.26	92.34	103.12	1.28	1.27	25.8	23.6
46	4.40{6}	8.40{11}	60.07	62.63	28.94	44.97	72.22	69.20	5.38	5.10	31.1	31.4
47	4.46{31}	7.23{31}	69.75	64.39	48.72	62.46	76.52	75.31	4.51	4.44	27.8	26.0
48	6.91{1}	3.27{1}	94.34	103.74	19.53	29.17	98.06	94.75	9.97	9.17	37.2	39.0
50	4.55{62}	5.15{62}	76.26	72.75	62.69	75.00	106.73	121.62	0.64	0.83	25.7	22.6
51	4.57{56}	4.93{56}	76.56	72.06	63.43	73.14	103.55	106.04	0.85	0.95	26.0	23.2
52	4.79{44}	5.13{49}	76.05	71.78	63.30	72.97	95.17	118.33	0.76	0.91	25.9	23.1
53	4.66{23}	4.68{25}	73.71	69.16	61.79	68.04	102.57	110.91	1.08	1.00	24.8	22.8
54	4.63{24}	4.49{24}	75.96	69.42	60.95	67.08	98.22	109.62	0.86	0.95	25.7	23.8
55	5.65{10}	5.39{7}	78.46	74.25	62.45	69.30	72.52	76.56	0.87	0.81	21.1	20.3
56	5.09{25}	4.78{23}	76.43	71.70	62.76	69.52	95.51	108.23	0.88	0.92	25.5	23.4
57	4.56{17}	4.53{17}	74.88	69.79	61.51	67.60	97.11	105.17	0.81	0.83	25.7	23.8
58	3.95{11}	3.73{15}	81.39	73.79	69.13	70.28	92.47	99.53	2.86	3.26	24.2	24.6
59	4.65{63}	5.57{63}	76.27	72.09	63.68	76.90	90.80	111.12	0.96	0.96	25.3	22.1
60	3.99{43}	4.48{44}	73.68	69.35	59.44	69.17	95.38	113.06	1.06	1.30	24.7	22.5
61	4.36{65}	6.95{65}	73.82	71.56	56.02	73.41	72.29	86.85	2.44	2.24	24.4	22.8
62	4.51{58}	6.24{59}	73.82	69.50	47.62	61.96	66.56	72.05	3.74	3.51	21.8	21.2
63	7.78{27}	10.17{40}	77.07	76.05	47.81	61.06	68.46	65.61	5.11	4.96	25.9	27.3
64	6.37{67}	8.55{67}	75.87	73.08	63.32	78.99	73.72	87.14	2.25	2.17	25.1	22.6
65	5.00{66}	7.89{66}	74.08	71.86	58.17	76.13	74.70	85.16	2.57	2.38	25.0	23.2
66	5.12{64}	7.51{64}	73.25	70.42	59.13	74.38	78.62	83.35	2.71	2.71	25.6	23.4
67	5.10{61}	7.16{61}	74.30	69.80	51.45	65.63	65.97	73.44	3.69	3.49	22.0	21.5
68	6.32{51}	7.70{55}	71.70	67.09	53.11	64.38	68.66	73.80	4.15	3.93	23.5	23.0
69	9.09{33}	12.02{41}	78.73	77.57	51.62	65.64	65.21	67.87	5.25	5.15	26.7	27.9

Table 4: VSCF Calculations
for Anharmonic Frequencies in cm^{-1} and Infrared Intensities in $Km \cdot mol^{-1}$

VSCF	ω_1	ω_2	ω_3	IR_1	IR_2	IR_3
Exp [50]	3657	1595	3756	2.93	62.5	41.7
B3LYP/cc-pV6Z	3639.4	1546.2	3713.5	4.39	76.68	61.01
B3LYP/cc-pV5Z	3639.8	1546.9	3713.7	4.17	76.16	58.07
B3LYP/Aug-cc-pVQZ	3637.1	1545.8	3711.3	3.92	77.46	61.74
B3LYP/May-cc-pVTZ	3635.2	1542.4	3710.7	5.61	71.47	59.03
B3LYP/aug-pc3	3638.5	1545.5	3712.9	3.92	77.63	61.76
B3LYP/Def2-QZVPD	3639.5	1545.8	3713.8	3.90	77.46	61.57
MP2/cc-pV5Z	3680.3	1553.0	3779.6	6.53	72.86	74.81
MP2/Aug-cc-pVQZ	3671.5	1550.1	3771.2	5.32	74.24	76.47
MP2/May-cc-pVTZ	3677.3	1559.5	3775.3	8.49	65.02	71.19
MP2/aug-pc3	3669.3	1551.0	3769.4	5.33	74.05	76.98
MP2/Def2-QZVPD	3679.6	1552.5	3779.6	5.30	73.16	75.67
SCS-MP2/Def2-QZVPD	3673.7	1569.2	3768.0	5.29	73.31	76.07
CCSD(T)/Def2-QZVPD	3664.6	1569.4	3746.8	13.03	93.92	79.31
CCSD(T)/Def2-TZVPPD	3650.2	1569.5	3731.9	13.11	93.55	78.27
CCSD(2)$_T$/Def2-QZVPD	3670.9	1571.1	3752.2	3.64	73.08	58.12
CCSD(2)$_T$/Def2-TZVPPD	3656.4	1571.0	3737.2	3.04	70.65	54.36

3.1. Calculations Removed from Study

After careful consideration, certain combinations of basis set and theory methods were dropped from the study due to issues discussed in this section. The version of Gamess(US) used for this study, does not support basis sets with high angular momentum i-functions (5/6 zeta basis sets) to be utilizied for predicting IR intensities or Raman activities. Zeta 5/6 Hessian calculations for CCSD(T) were too computationally intensive and did not provide significant improvements on molecular property predictions to justify their extreme runtimes. Sadlej-LPolX-fl was also

dropped from most calculations (MP2 within optimization and hessian calculations, across all theories within Raman calculations) due to convergence issues.

3.2. Molecular Geometries

The floating integer values within Figures 1-4 throughout the following sections are the zeta values for the basis sets. As expected, CCSD(T) performs best for calculating molecular geometries yielding consistent results with few errors. B3LYP converges faster than MP2 and CCSD(T), with the latter providing only slightly more accurate bond lengths and angles, for water.
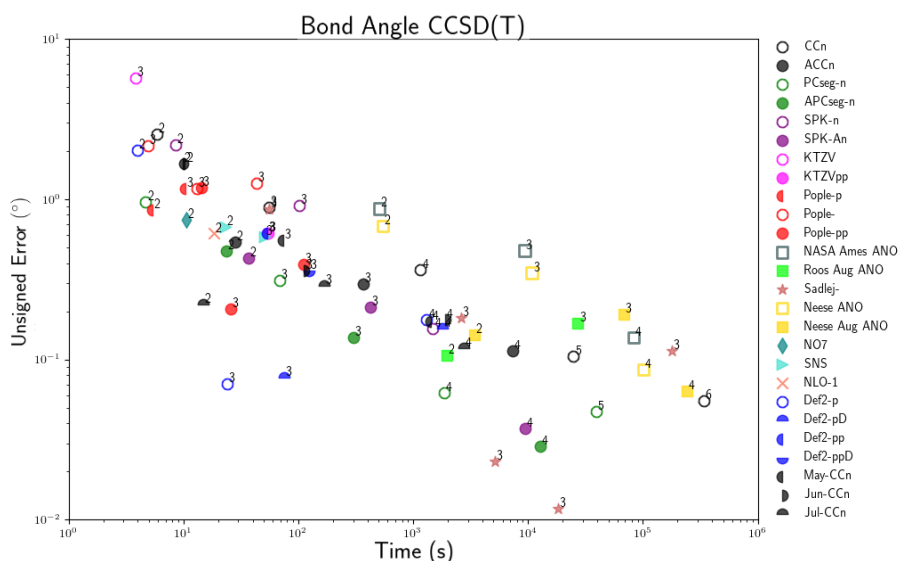
Figure 1: Unsigned Error of CCSD(T) Bond Angle Calculations Compared to Experimental Data

It is notable to mention that the Pople-style basis sets provide reasonably accurate geometries, and it requires considerably high zeta basis sets to converge on the accurate geometry; particularly for the bond angle of water. While bond length converges quickly, for all levels of theory studied here, the bond angle converges more slowly, as shown for CCSD(T) in Figure 1. In general, most basis families show linear trend lines of slower convergence as either diffuseness or zeta values increase; making DFT with pople basis sets particularly well-suited for optimizing the geometries of large molecules. The Def2 family performs well in both runtime and accuracy, similarly both the PCseg and APCseg yield high accuracy predictions with lower runtimes, particularly when compared to their CC counterparts.

3.3. Vibrational Frequencies

When comparing vibrational frequencies to experimental data CCSD(T) calculations yielded the most accurate results, as expected. Interestingly, the NO7 and SNS basis set families yield results within ten wavenumbers (here, we consider this a reasonable threshold for the characterization or discrimination of most molecular species) with minimal runtimes [11].
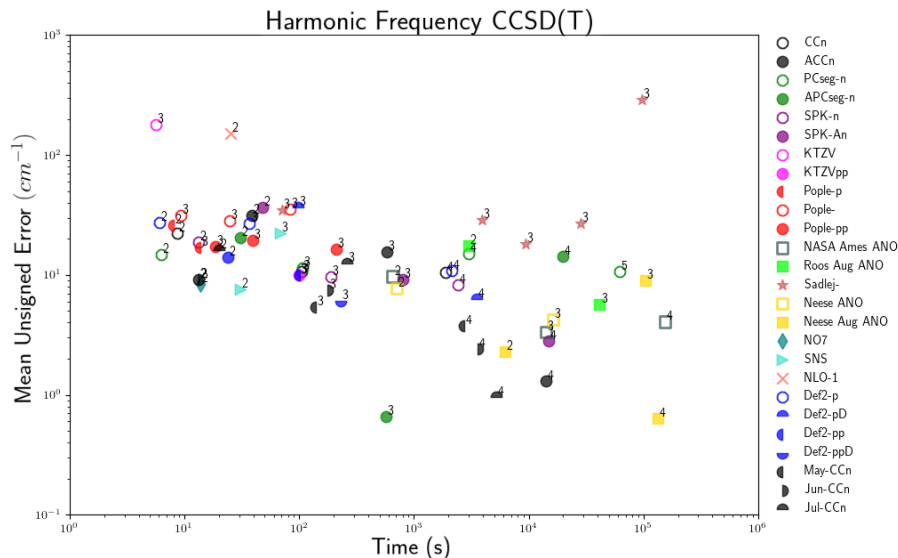
Figure 2: Mean Unsigned Error of CCSD(T) Frequency Calculations Compared to Experimental Data

The Def2 basis family consistently outperforms both correlation consistent and polarization consistent basis sets, at significantly reduced computational expense. Both the augmented and non-augmented SPK- basis sets also perform very well in calculating frequencies, producing similar accuracy and runtimes as Def2 sets. While, for B3LYP, the commonly used Pople family, converges above the 10 cm^{-1} threshold with similar runtimes. It is noteworthy to mention that fortuitous cancellation of errors often plays a large role here, particularly with smaller basis sets. However, if these errors consistently off-set a deficiency in the level of theory encountered at the basis set limit, or that is not overcome unless much larger basis sets are used, this may prove useful. As an example, for the prediction of vibrational frequencies, the 'may' calendar basis sets consistently out-perform even their fully augmented counterparts at only a minor additional computational expense compared to the CCn basis at all levels of theory studied. B3LYP is known to over-estimate bond-lengths, and under-estimate harmonic frequencies. Often, a scaling factor is applied to reduce the errors in predicting anhamonic values from calculations performed under the harmonic approximation [51–53]. In Table 4 we can see clearly that the anharmonic frequencies for B3YLP are all below the experimental values. Again in Table 4, the anharmonic frequencies are shown for a few select levels of theory and basis sets. Here, we can see that B3LYP consistently underestimates the experimental frequencies, and that even the may-cc-pVTZ basis set values lie within 5 cm^{-1} of the cc-pV6Z values. Similarly, we see for MP2 that the def2, may and pc families provide excellent agreement with the cc-pV5Z values. For each level of theory, the def2-QZVPD basis set approaches the accuracy of cc-pV5Z/cc-pV6Z much faster than most other basis sets of similar quality. Since anharmonic calculations are computationally expensive, CCSD(T) calculations were only performed using two of the def2 basis sets, where it could be seen that although an improvement over MP2 is observed, there is still not overall agreement with the experimental harmonic frequencies. This is likely because MP2 and CCSD(T) do not perform as well when used to calculate energetics at geometries far from the equilibrium position. To emphasize this point, calculations were additionally ran using the SCS-MP2 and CCSD(2)$_T$ levels of theory, which in each case show a marked improvement in the agreement with experimental anharmonic frequencies. All of the levels of theory considered here, struggle to reproduce the ν_2/ω_2 frequency. We note that the CCSD(2)$_T$/def2-QVZPD level of theory only has a MUE of $\approx$13 cm^{-1}.

8

3.4. Infrared Intensities

From Figure 3, we can see that though the Pople and other small basis sets can outperform larger basis sets at predicting some infrared intensities, however, caution is urged since this is not always the case. Both the B3LYP and MP2 levels of theory have significant errors when calculating infrared intensities under the harmonic approximation, even when extrapolated to their CBS limits. The B3LYP theory level convergence line appears to lie just above the ten percent error margin, some basis sets - particularly small basis sets, such as Pople - deviate from this producing more accurate results. In some cases where small basis sets are utilized, the errors arising actually compensates for the deficiencies in B3LYP, producing seemingly more accurate results. However, these results may occur sporadically, and caution is urged using these results, which could easily be taken out of context [54].
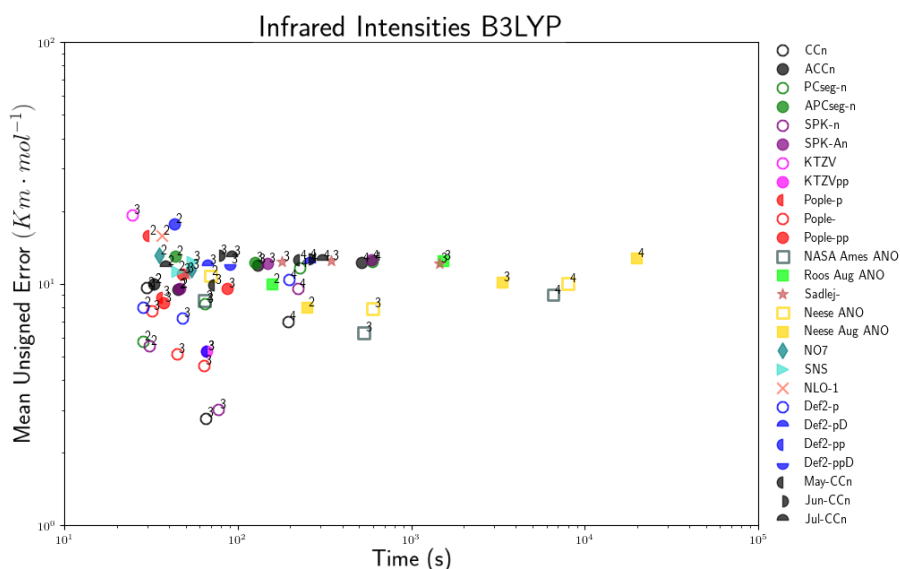
Figure 3: Mean Unsigned Error of B3LYP Infrared Intensity Calculations Compared to Experimental Data

Both the harmonic and anharmonic calculation of the infrared frequencies is subject to the accuracy of the initial Hessian used to project the geometries, as well as the accuracy of the level of theory and basis set used to calculate the energetics and dipole moment. As explained previously, the MP2 and CCSD(T) levels of theory are not as reliable at geometries required for anharmonic, or even harmonic calculations. As a result, infrared intensities are not as dependent upon diffuse and polar functions as Raman activities, but rather how accurately the relative energies are calculated at each geometry, which explains why some basis sets that do not incorporate large polarization and diffuse functions are only necessary to the extent that the energetics and dipole moment are calculated with sufficient accuracy. Table 4 shows that by utilizing SCS-MP2 [55] or $CCSD(2)_T$ [56] levels of theory (the former at no additional cost, the latter at twice the cost of regular CCSD(T) calculations), further improvements in the frequencies and IR intensities can be obtained.

3.5. Raman Activities

In contrast to the IR intensities, the Raman activities can be very accurately calculated under the harmonic approximation, in particular when the MP2 level of theory is used (in agreement with previous studies [3, 4]) and are extremely sensitive to the degree of polarization and diffusivity incorporated into the basis sets.
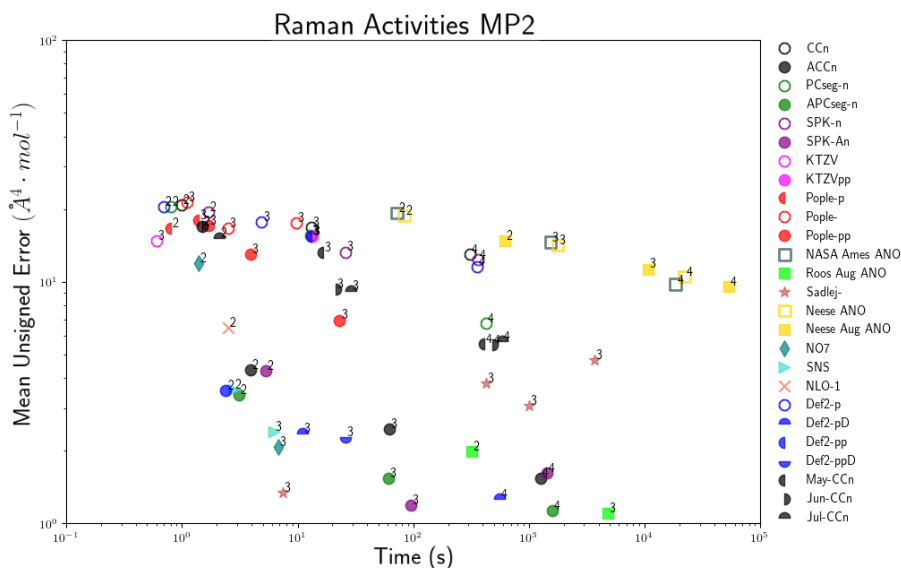
Figure 4: Mean Unsigned Error of MP2 Raman Activities Calculations Compared to Experimental Data

It is clear from Figure 4 that while increasing the zeta-level or overall size of basis set is helpful, those basis sets that incorporate higher polarization levels and more diffuse functions are generally more accurate at reproducing Raman activities. In particular, the spectroscopic basis sets all perform well here (e.g., Sadlej, N07, SNS). The (PPD versions of) Def2 and (aug) Roos basis sets perform very well here also. The NLO-1 basis set also performs fairly well in calculating Raman activities, comparable to many quadruple-zeta basis sets. The dependence on the level of diffusivity is well demonstrated from the correlation consistent and calendar basis sets (cc→ may→ jun→ jul→ aug).

4. Conclusion

From our investigation we have identified various trends for the predictions of water's molecular properties, of which the most notable is the Def2-n basis set family's overall performance. Pople-style basis sets tend to be very consistent with their predictions, typically producing minimal changes between each individual basis set's calculation. Pople basis sets are a reasonable choice for exploration, but there is little systematic improvement within their family, and are not recommended when results require high accuracy or quantitative values. Although they are occasionally somewhat accurate for infrared intensities when used with the B3LYP level of theory, they are not recommended to predict Raman activities. The Aug-CCn family is good for predicting Raman activities, particularly with the MP2 level of theory, however, they are computationally heavy, and outperformed by less computationally heavy options, such as aug-PCn, Def2-PD, and SPK augmented versions, as well as the majority of spectroscopic basis sets in this area. These basis sets have been shown to produce results comparable to six zeta basis sets, with minimal runtimes for the calculation of many optical properties. Another notable trend is the NO7 and SNS basis families success with calculating vibrational frequencies with harmonic approximation methods. Regarding the ANO series of basis sets, the NASA and Neese series offer high accuracy for geometries, as well as vibrational frequencies and infrared intensities, but lack sufficient diffuse functions to be able to accurately determine Raman activities (even the Aug-Neese sets). Only the Roos augmented ANO basis sets were able to accurately reproduce Raman activities, and represent an excellent all-round choice. We noticed that for the calculation of Raman activities basis sets which incorporated large numbers of diffuse functions always provided superior accuracy, however our scope was limited to the water molecule and further

investigations need to be conducted.We are currently increasing the scope of this study to include various other molecular species in the hopes of validating trends found in this study.

References

[1] Bennett C J, Ennis C P and Kaiser R I 2014 *The Astrophysical Journal* **782** 63
[2] Glaser R, Hodgen B, Farrelly D and McKee E 2007 *Astrobiology* **7** 455–470
[3] Zvereva E E, Shagidullin A R and Katsyuba S A 2010 *The Journal of Physical Chemistry A* **115** 63–69
[4] Jiménez-Hoyos C A, Janesko B G and Scuseria G E 2008 *Physical Chemistry Chemical Physics* **10** 6621–6629
[5] Huber K P 2013 *Molecular spectra and molecular structure: IV. Constants of diatomic molecules* (Springer Science & Business Media)
[6] Helgaker T, Jorgensen P and Olsen J 2014 *Molecular electronic-structure theory* (John Wiley & Sons)
[7] Halls M D and Schlegel H B 1998 *The Journal of chemical physics* **109** 10587–10593
[8] Halls M D and Schlegel H B 1999 *The Journal of chemical physics* **111** 8819–8824
[9] Scott A P and Radom L 1996 *The Journal of Physical Chemistry* **100** 16502–16513
[10] Buczek A, Kupka T and Broda M A 2011 *Journal of molecular modeling* **17** 2029–2040
[11] Galabov B, Yamaguchi Y, Remington R B and Schaefer H F 2002 *The Journal of Physical Chemistry A* **106** 819–832
[12] Hill J G 2013 *International Journal of Quantum Chemistry* **113** 21–34
[13] Jensen F 2013 *Wiley Interdisciplinary Reviews: Computational Molecular Science* **3** 273–295
[14] Kupka T and Lim C 2007 *The Journal of Physical Chemistry A* **111** 1927–1932
[15] Helgaker T, Klopper W, Koch H and Noga J 1997 *The Journal of chemical physics* **106** 9639–9646
[16] Truhlar D G 1998 *Chemical Physics Letters* **294** 45–48
[17] Petersson G and Al-Laham M A 1991 *The Journal of chemical physics* **94** 6081–6090
[18] Becke A D 1993 *The Journal of chemical physics* **98** 5648–5652
[19] Lee C, Yang W and Parr R G 1988 *Physical review B* **37** 785
[20] Møller C and Plesset M S 1934 *Physical Review* **46** 618
[21] Piecuch P, Kucharski S A, Kowalski K and Musiał M 2002 *Computer Physics Communications* **149** 71–96
[22] Olson R M, Bentz J L, Kendall R A, Schmidt M W and Gordon M S 2007 *Journal of Chemical Theory and Computation* **3** 1312–1328
[23] Hariharan P C and Pople J A 1973 *Theoretica chimica acta* **28** 213–222
[24] Woon D E and Dunning Jr T H 1993 *The Journal of chemical physics* **99** 3730–3737
[25] Papajak E and Truhlar D G 2010 *Journal of chemical theory and computation* **7** 10–18
[26] Dunning Jr T H 1989 *The Journal of chemical physics* **90** 1007–1023
[27] Woon D E and Dunning Jr T H 1994 *The Journal of chemical physics* **100** 2975–2988
[28] Kendall R A, Dunning Jr T H and Harrison R J 1992 *The Journal of chemical physics* **96** 6796–6806
[29] Papajak E and Truhlar D G 2010 *Journal of chemical theory and computation* **6** 597–601
[30] Papajak E, Zheng J, Xu X, Leverentz H R and Truhlar D G 2011 *Journal of chemical theory and computation* **7** 3027–3034
[31] Jensen F 2001 *The Journal of Chemical Physics* **115** 9113–9125
[32] Jensen F 2002 *The Journal of chemical physics* **116** 7372–7379
[33] Jensen F 2014 *Journal of chemical theory and computation* **10** 1074–1085
[34] Weigend F and Ahlrichs R 2005 *Physical Chemistry Chemical Physics* **7** 3297–3305
[35] Neese F and Valeev E F 2010 *Journal of chemical theory and computation* **7** 33–43
[36] Baranowska A and Sadlej A J 2010 *Journal of computational chemistry* **31** 552–560
[37] Benkova Z, Sadlej A J, Oakes R E and Bell S E 2005 *Journal of computational chemistry* **26** 145–153
[38] Barone V, Cimino P and Stendardo E 2008 *Journal of chemical theory and computation* **4** 751–764
[39] Barone V, Biczysko M and Bloino J 2014 *Physical Chemistry Chemical Physics* **16** 1759–1787
[40] Paschoal D, Costa M F and Dos Santos H F 2014 *International Journal of Quantum Chemistry* **114** 796–804
[41] Shimanouchi T 1977 *Journal of physical and chemical reference data* **6** 993–1102
[42] Hoy A and Bunker P R 1979 *Journal of Molecular Spectroscopy* **74** 1–8
[43] Culot F and Liévin J 1992 *Physica Scripta* **46** 502
[44] Tew D P, Klopper W, Heckert M and Gauss J 2007 *The Journal of Physical Chemistry A* **111** 11242–11248
[45] Gordon M S and Schmidt M W 2005 *Advances in electronic structure theory: GAMESS a decade later* (Amsterdam: Elsevier) pp 1167–1189
[46] Sayvetz A 1939 *The Journal of Chemical Physics* **7** 383–389
[47] Ishimura K, Acc S N T C, McMurchie L, Davidson E and Rys J 1993 *J. Comput. Chem* **14** 1347–1363
[48] Krasnoshchekov S V, Schutski R S, Craig N C, Sibaev M and Crittenden D L 2018 *The Journal of chemical physics* **148** 084102

[49] Halkier A, Helgaker T, Jørgensen P, Klopper W, Koch H, Olsen J and Wilson A K 1998 *Chemical Physics Letters* **286** 243–252
[50] Johnson B G, Gill P M and Pople J A 1993 *The Journal of chemical physics* **98** 5612–5626
[51] Merrick J P, Moran D and Radom L 2007 *The Journal of Physical Chemistry A* **111** 11683–11700
[52] Andersson M P and Uvdal P 2005 *The Journal of Physical Chemistry A* **109** 2937–2941
[53] Halls M D, Velkovski J and Schlegel H B 2001 *Theoretical Chemistry Accounts* **105** 413–421
[54] Hudson R L, Gerakines P A and Ferrante R F 2018 *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **193** 33–39
[55] Grimme S 2003 *The Journal of chemical physics* **118** 9095–9102
[56] Piecuch P, Włoch M, Gour J R and Kinal A 2006 *Chemical physics letters* **418** 467–474