The Failure of Correlation to Describe Out-of-Plane Carbon=Carbon Bending

Ryan C. Fortenberry,[*] Timothy J. Lee,[†] Joshua P. Layfield.[‡]

November 2, 2017

Abstract

Carbon-carbon multiply bonded systems are improperly described with standard correlation methods and basis sets. For computations of vibrational modes, the out-of-plane bends can be reported as imaginary at worst or simply too low at best. Utilizing the simplest of aromatic structures (cyclopropenylidene) and various levels of theory, this work diagnoses this known behavior for the first time. A combined 1-particle and n-particle basis set effect conspire to produce these non-physical results. When moving from sp^2 to sp^3 hybridization in the carbon atoms, the larger number of basis functions overcorrects the energy. This is exacerbated by correlation methods. These allow for occupation of the π and π^* orbitals in the expanded wave function that combine with the hydrogen s orbitals. As a result, the improperly described space can be further and non-physically stabilized by post-Hartree-Fock correlation. This represents a fundamental problem with at least Hartree-Fock based methods of all flavors in describing carbon. Beyond being a flaw in quantum chemical theory, other repercussions will be present in computations regarding spectroscopy as well as energy and environmental studies where highly-accurate hydrocabon vibrational transitions or thermochemical data are needed.

Keywords: Quantum Chemistry; Correlation Methods; Basis Sets; Carbon

Quantum chemical correlation methods cannot properly describe what happens to carbon atoms as they bend out-of-plane from one another. The "best" electronic structure methods have what appears to be a serious flaw for carbon, the building block of life and most higher chemistry. For at least the past two decades, non-physical irregularities have arisen in the computation of vibrational frequencies for hydrocarbons, specifically π-bonded and aromatic species. The out-of-plane modes exhibit this problem. None of the in-plane modes seem affected[1]. This problem was first identified in correlated computations of harmonic frequencies for $C-C$ multiply-bonded systems that contain inversion symmetry, notably ethylene, acetylene, and benzene[2–4]. Recently, this problem has been identified in $C-C$

[*]Department of Chemistry & Biochemistry, Georgia Southern University, Statesboro, Georgia 30460-8064, U.S.A, rfortenberry@georgiasouthern.edu

[†]Division of Space Science & Astrobiology NASA Ames Research Center, MS 245-3 Moffett Field, California, 94035, U.S.A

[‡]Department of Chemistry, University of St. Thomas, St. Paul, Minnesota, 55105, U.S.A.

multiply-bonded systems that do not specifically contain an inversion symmetry operator in its point group albeit not as extreme as those that do. Nucleobases are a prime example of the former case[5]. The extreme example of this erroneous behavior is that often Møller-Plesset perturbation theory at second-order (MP2)[6] will produce imaginary harmonic frequencies for out-of-plane bending (OPB) modes[5,7,8]. However, the problem is much more widespread and less detectable than this would indicate.

Coupled cluster theory[9,10] has been labeled as the most accurate and even "gold standard" of quantum chemistry, especially in its singles, doubles, and perturbative triples [CCSD(T)] implementation[11–13]. The initial guess to the CCSD equations, necessary to compute CCSD(T) or any response properties, is MP2. While the iterative procedure and expanded wave function of CCSD improve the description of energetically-derived properties, the failure to describe the OPBs of hydrocarbons is still present and potentially more worrisome since the errors are hiding in plain sight. Real-valued harmonic frequencies are produced, though they may be too low in energy without any clear indication of being so upon initial analysis.

Previous work has suggested the problem is related to a one-particle basis set inadequacy. The suggestions have ranged from an imbalance between saturation in the *spd* space versus including higher-angular momentum functions, to an intramolecular basis set superposition error (BSSE), to a basis set linear dependency.[3–5,7,8,14]. To combat the problem, suggestions of using "balanced" basis sets have been proposed such as versions of the correletion-consistent basis sets or, especially, atomic natural orbitals (ANOs) instead of Pople basis sets[4]. While the correlation-consistent basis sets for post-MP2 methods minimize the problem for harmonic frequencies, extrapolation to the one-particle basis set limit can still be problematic[14], especially for anharmonic frequencies[1,15]. ANO basis sets seem to improve upon the correlation consistent basis sets for this problem, but non-physical behavior is not completely absent[4]. Alternatively, explicitly including BSSE counterpoise corrections could be included, but only if the proper fragments of the molecule are chosen, which can be *ad hoc* and filled with guesswork[14,16]. Most recently, Samala and Jordan have suggested the error in the OPBs is due to basis set linear dependencies[8]. They minimized the problem by eliminating eigenfunctions of the overlap matrix with small eigenvalues.

However, such actions as any of these still appear to be simply treatment without diagnosis. The purpose of the present work is to show that the OPB error is still present and may not be found until anharmonic frequencies are examined in C−C multiply-bonded systems without inversion symmetry. Additionally, a hypothesis as to the origins of this phenomenon is also offered.

Recent quantum chemical computations have pushed the accuracy of anharmonic fundamental vibrational frequency prediction to the 1.0 cm^{-1} level[17–22]. Such approaches for computing the vibrational potential energy surface involve composite energy schema for quartic force fields (QFFs). These accurate fourth-order Taylor series expansions of the internuclear potential contain energies for complete basis set (CBS) extrapolations, core electron correlation, and scalar relativity to give the so-called CcCR QFF[23]. This QFF is then coupled to vibrational perturbation theory at second order (VPT2)[24,25]. Consequently, errors in the OPB frequencies render these expensive computations ineffective for describing such modes.

The computation of accurate vibrational frequencies extends beyond spectroscopy where highly-accurate determinations of thermochemical enthalpic data depend upon good anhar-

Table 1: The c-C_3H_2 Anharmonic Vibrational Frequencies from Different Levels of Theory and from Experiment.

		Ar Matrix Exp.[26]	LHD[27]	CcCR	MP2-F12/apVDZ	MP2-F12/apVTZ	HF/apVDZ	HF/apV5Z(spd)
$\omega_1(a_1)$	C-H Symm. Stretch		3290.7	3290.9	3314.9	3314.2	3432.1	3437.8
$\omega_2(b_2)$	C-H Antisymm. Stretch		3256.5	3258.4	3280.3	3280.4	3392.1	3397.8
$\omega_3(a_1)$	C-C-C Bend/C=C Stretch		1629.9	1631.6	1622.0	1624.3	1745.6	1746.0
$\omega_4(a_1)$	C-C Symm. Stretch		1312.4	1314.2	1323.8	1327.5	1407.5	1413.8
$\omega_5(b_2)$	C-C Antisymm. Stretch		1093.1	1097.3	1095.5	1100.6	1180.4	1192.8
$\omega_6(a_2)$	Antisymm. OPB/Twist		995.7	997.5	996.0	1003.7	1083.8	1070.8
$\omega_7(a_1)$	H-C-C Symm. Bend		911.4	914.8	911.5	916.1	1001.4	1013.6
$\omega_8(b_2)$	H-C-C Antisymm. Bend		904.4	904.3	903.5	906.0	982.3	991.8
$\omega_9(b_1)$	Symmetric OPB		789.4	788.8	797.1	803.3	873.8	876.4
$\nu_1(a_1)$	C-H Symm. Stretch		3145.0	3152.6	3198.6	3194.4	3306.5	3310.1
$\nu_2(b_2)$	C-H Antisymm. Stretch		3123.5	3119.8	3151.0	3148.2	3271.3	3275.9
$\nu_3(a_1)$	C-C-C Bend/C=C Stretch		1600.2	1599.0	1594.3	1590.5	1717.4	1717.7
$\nu_4(a_1)$	C-C Symm. Stretch	1278.8	1279.4	1277.5	1288.9	1287.3	1377.2	1378.8
$\nu_5(b_2)$	C-C Antisymm. Stretch	1061.5	1063.0	1064.6	1066.8	1069.0	1150.9	1161.3
$\nu_6(a_2)$	Antisymm. OPB/Twist		976.3	952.1	974.8	922.6	1048.0	1058.9
$\nu_7(a_1)$	H-C-C Symm. Bend	886.4	889.2	887.2	889.1	878.3	972.7	982.3
$\nu_8(b_2)$	H-C-C Antisymm. Bend		882.2	873.9	878.9	872.3	958.5	961.2
$\nu_9(b_1)$	Symmetric OPB	787.4	776.0	753.2	782.4	731.2	838.2	835.9

monic vibrational and rovibratioanl computations. Furthermore, improper descriptions of the π cloud for the single element on which a majority of chemistry is based, carbon, are troublesome at a fundamental level for practitioners of the theoretical chemistry arts since robust, *ab initio* predictions are the ultimate goal of wave function theory development.

Cycloprenylidene (c-C_3H_2) is often called the simplest aromatic hydrocarbon since it follows Hückel's $4n + 2$ rule with $n = 0$. However, CcCR QFF VPT2 computations on the vibrational frequencies of c-C_3H_2 appear to possess the same behavior where an explicit inversion center is not present. Table 1 gives the vibrational frequencies for c-C_3H_2 with several different QFFs as well as argon-matrix experimental results[26]. Previous, high-level computations have been examined at the harmonic frequency level to determine whether the OPB issue is present for cyclopropenylidene[27] and the fundamental frequencies have been found to compare well with the available experiment. Hence, the Lee, Huang, and Dateo (LHD) results from Ref. 27 are considered the standard for this study.

Clearly the symmetric OPB, ν_9 of b_1 symmetry in Table 1, is well described by the LHD QFF since it produces a 776.0 cm^{-1} fundamental frequency with the experimental results at 787.4 cm^{-1} where a small shift compared to gas phase results is likely. The antisymmetric OPB, ν_6 of a_2 symmetry, is likely similarly well defined by LHD at 968.6 cm^{-1} but will have no intensity for experiment to be observed due to symmetry. However, the CcCR results fall short of these marks at 753.2 cm^{-1} and 952.1 cm^{-1} even though the other modes, especially those involving the heavy atoms, coincide nicely. While the a_2 mode is not as bad as the b_1 and will be spectroscopically dark, highly-accurate thermochemical computations require proper descriptions. The main difference between the two approaches is that the earlier LHD work is based upon CCSD(T)/aug-cc-pVTZ (apVTZ) energies further corrected for core correlation and the present CcCR approach has the one-particle CBS extrapolation giving it more basis functions. As a result, the CcCR a_2 and b_1 fundamental frequencies are noticeably lower than the LHD values and are probably too low. They are not imaginary but are artificially too low. The interesting and novel point here, however, is that the CcCR harmonic frequencies do not suggest a problem – in this case it is strictly due to the anharmonic correction. The LHD and CcCR harmonic frequencies differ by no more than 4.0 cm^{-1}, but the anharmonic fundamentals for this simple, C$-$C multiply-bonded system are clearly not performing as well.

Figure 1: The symmetric (blue) and antisymmetric (red) OPB scans of c-C_3H_2 (with H_2) versus c-C_3H_4.
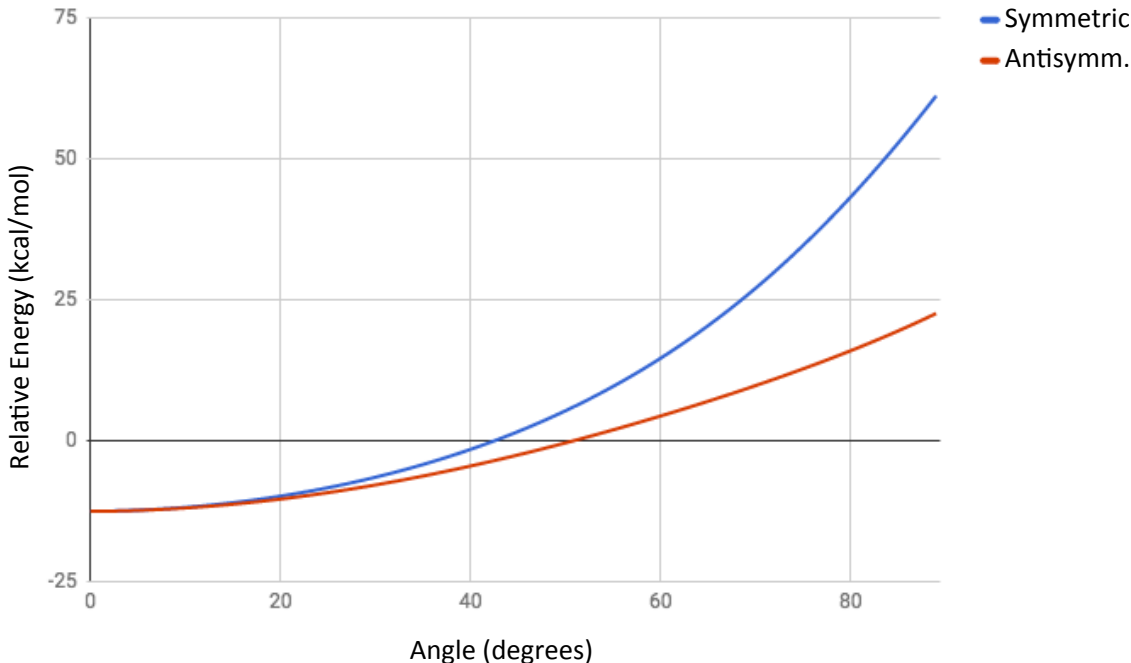
Another solution for this problem besides seemingly arbitrarily clipping the basis set is to use explicitly correlated methods, notably CCSD(T)-F12 or MP2-F12[28]. The explicit treatment of electron-electron repulsion has a better chance of describing the proper electronic behavior than standard methods especially for vibrational frequencies[15]. While the MP2-F12/apVDZ heavy atom modes are very well described with this seemingly simple yet fortuitous Pauling point[29], the harmonic frequencies for MP2-F12/apVTZ are nearly identical. This continues for the anharmonic frequencies, save for the ν_6 and ν_9 OPB fundamentals where the larger basis, again, leads to improper stabilization of the OPB motion. Hence, proper selection of basis set that a method can handle can reduce this concern, but such is still merely treatment and not diagnosis.

The hypothesis of this work is that the very act of adding correlation is the culprit, or in other words, this effect is a combined 1- and n-particle basis set effect. Hence, it is a method AND basis set effect. In examining the CCSD wave function for c-C_3H_2, the largest contributor to the singles is the delocalized occupied π orbital being substituted by the virtual π^* orbital containing a single vertical node. In other words, this contributor is the particle-on-a-ring $n = 2 \leftarrow n = 1$ transition. Consequently, the OPB modes bring the hydrogen s orbitals into constructive interference with both the $n = 1$ and $n = 2$ π clouds of the occupied and virtual orbitals.

In order to test how this affects the wave function, the potential of cycloprenylidene is scanned in both OPB coordinates with CCSD(T)/apVTZ. Adding the constant of the H_2 energy to this scan energy allows for comparison to the same scan in the saturated, nonaromatic c-C_3H_4 molecule which has no π electrons (Figure 1). Initially, c-C_3H_4 is lower in energy than c-$C_3H_2 + H_2$ which is indicated by the negative relative energy. However,

4

at 43°, cyclopropenylidene becomes more favorable in the symmetric OPB (b_1) and 51° for the antisymmetric OPB (a_2). As a result, the π cloud interactions with the hydrogen s orbitals stabilizes the OPB. As the displacement grows for higher derivative levels beyond the harmonic level, the artificial overlap is enhanced. This behavior is further exaggerated by the presence of the equally-favorable π^* orbital in the correlated wave function. Hence, the actual act of including electron correlation as an expansion of the Hartree-Fock (HF) wave function leaves the computation open to such overcorrection.

As a final test, non-correlated QFFs are constructed for c-C_3H_2 at the HF level with the apVDZ and the apV5Z basis sets but with the latter only containing s, p, and d functions in order to test the behavior of such orbital overlap for the valence orbitals. These values are also contained in Table 1. The vibrational frequencies are nearly the same for each mode between these two QFFs, and the b_1 and a_2 frequencies actually increase somewhat with the larger basis set. As a result of the lack of correlation, the increase in basis does not produce improperly treated OPB ν_6 and ν_9 fundamental frequencies as has been produced with coupled cluster theory and MP2.

As a final test, computations with MP2 using the apVDZ and apV5Z (only s, p, and d functions included) were performed for the harmonic frequencies since it has previously been shown that saturating the spd space without including higher angular momentum functions exacerbates the OPB issue[3]. In these calculations heinous errors are found, most notably for the a_2 mode where the MP2/apVDZ harmonic frequency is 945.6 cm^{-1} and the MP2/apV5Z(spd) harmonic frequency is 878.7 cm^{-1}, further confirming the role of electron correlation in the OPB error.

In summary, while pruning the basis set, utilizing ANO bases, or employing properly weighted explicitly correlated methods can treat the problem of artificially lowered OPBs, the diagnosis for the error appears to be fundamental to expanding post-HF wave functions. The artificial lowering of the OPB energy for carbon-carbon π bonded systems is the result of both basis set and correlation effects. The model molecule c-C_3H_2 has no explicit inversion center, and the CCSD(T)-based correlation-consistent basis harmonic frequencies are well-behaved, *but the inclusion of anharmonicity causes underestimates of the two OPBs.* Reduction of the basis set still exhibits this problem, but removal of correlation does not. Hence, coupling between the 1- and n-particle expansions is creating the OPB problem.

By way of outlook, the root for this problem is that basis sets for carbon are built around the ground electronic state of the atom which must be in atomic orbitals. However, C$-$C multiply-bonded systems are sp^2 hybridized. The simplest closed-shell, sp^2 hybridized hydrocarbon, methylidene (CH$_2$), has an empty p orbital, but this is actually a low-lying excited state roughly 0.5 eV above the triplet ground state which reverts to sp^3 hybridization with all orbitals containing at least one electron. Therefore, spin recoupling when going from sp^2 to sp^3 hybridization, as is strongly present in OPB motions, is not properly treated. This recoupling is a product of the electron density in the initial $2p$ carbon orbitals in the sp^2 hybridized atoms in planar or linear structures. As a result, more work within the quantum chemistry community is required to develop basis sets that can properly represent the orbital space contained once carbon atoms engage in multiple bonds with themselves. Such will be required before pure *ab initio* data can be fully trusted for OPB mode frequencies for pure hydrocarbons with applications stretching across spectroscopy, combustion, and environmental science in addition to fundamental quantum chemical theory.

The explicitly correlated methods described in this work utilize the MOLPRO 2015.1 program[30]; all else are from PSI4[31].

Acknowledgements

References

[1] J. M. L. Martin, T. J. Lee, and P. R. Taylor, J. Chem. Phys. **108**, 676 (1998).

[2] T. J. Lee, W. D. Allen, and H. F. Schaefer, III, J. Chem. Phys. **87**, 7062 (1987).

[3] E. D. Simandiras, J. E. Rice, T. J. Lee, R. D. Amos, and N. C. Handy, J. Chem. Phys. **88**, 3187 (1988).

[4] J. M. Martin, P. R. Taylor, and T. J. Lee, Chem. Phys. Lett. **275**, 414 (1997).

[5] D. Asturiol, M. Duran, and P. Salvador, J. Chem. Theory Comput. **9**, 2574 (2009).

[6] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[7] D. Moran, A. C. Simmonett, F. E. Leach III, W. D. Allen, P. v. R. Schleyer, and H. F. Schaefer III, J. Am. Chem. Soc. **128**, 9342 (2006).

[8] N. R. Samala and K. D. Jordan, Chem. Phys. Lett. **669**, 230 (2017).

[9] T. D. Crawford and H. F. Schaefer III, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (Wiley, New York, 2000), vol. 14, pp. 33–136.

[10] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory* (Cambridge University Press, Cambridge, 2009).

[11] T. J. Lee and G. E. Scuseria, in *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*, edited by S. R. Langhoff (Kluwer Academic Publishers, Dordrecht, 1995), pp. 47–108.

[12] T. Helgaker, T. A. Ruden, P. Jørgensen, J. Olsen, and W. Klopper, J. Phys. Org. Chem. **17**, 913 (2004).

[13] K. Raghavachari, G. W. Trucks., J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[14] R. M. Balabin, J. Chem. Phys. **129**, 164101 (2008).

[15] X. Huang, E. F. Valeev, and T. J. Lee, J. Chem. Phys. **133**, 244108 (2010).

[16] N. Kobko and J. J. Dannenberg, J. Phys. Chem. A **105**, 1944 (2001).

[17] X. Huang, P. R. Taylor, and T. J. Lee, J. Phys. Chem. A **115**, 5005 (2011).

[18] D. Zhao, K. D. Doney, and H. Linnartz, Astrophys. J. Lett. **791**, L28 (2014).

[19] R. C. Fortenberry, X. Huang, J. S. Francisco, T. D. Crawford, and T. J. Lee, J. Chem. Phys. **136**, 234309 (2012).

[20] X. Huang, R. C. Fortenberry, and T. J. Lee, J. Chem. Phys. **139**, 084313 (2013).

[21] R. C. Fortenberry, X. Huang, T. D. Crawford, and T. J. Lee, J. Phys. Chem. A **118**, 7034 (2014).

[22] R. C. Fortenberry, T. J. Lee, and H. S. P. Müller, Molec. Astrophys. **1**, 13 (2015).

[23] R. C. Fortenberry, X. Huang, J. S. Francisco, T. D. Crawford, and T. J. Lee, J. Chem. Phys. **135**, 134301 (2011).

[24] I. M. Mills, in *Molecular Spectroscopy - Modern Research*, edited by K. N. Rao and C. W. Mathews (Academic Press, New York, 1972), pp. 115–140.

[25] J. F. Gaw, A. Willets, W. H. Green, and N. C. Handy, in *Advances in Molecular Vibrations and Collision Dynamics*, edited by J. M. Bowman and M. A. Ratner (JAI Press, Inc., Greenwich, Connecticut, 1991), pp. 170–185.

[26] J. W. Huang and W. R. M. Graham, J. Chem. Phys. **93**, 1583 (1990).

[27] T. J. Lee, X. Huang, and C. E. Dateo, Mol. Phys. **107**, 1139 (2009).

[28] H.-J. Werner, T. B. Adler, and F. R. Manby, J. Chem. Phys. **126**, 164102 (2007).

[29] R. F. Fink, J. Chem. Phys. **145**, 184101 (2016).

[30] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, et al., *Molpro, version 2015.1, a package of ab initio programs* (2015), see http://www.molpro.net.

[31] J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, et al., Wiley Interdisciplinary Reviews: Computational Molecular Science **2**, 556 (2012).