The Origin of Systematic Error in the Standard Enthalpies of Formation of Hydrocarbons Computed via Atomization Schemes

Gyula Tasi, a,b Róbert Izsák, a Gergely Matisz, a Attila G. Császár, a,b Mihály Kállay, a Branko Ruscic, d and John F. Stanton e

The combustion of hydrocarbon fuels involves many chemical species and reactions; for example, saturated hydrocarbons with up to about twelve and eighteen carbon atoms constitute the bulk of gasoline and diesel fuel, respectively. Knowledge of the (temperature-dependent) enthalpies of formation of alkanes is therefore of importance for developing reliable models of fuel combustion. Motivated partly by their industrial importance, several recent theoretical studies have dealt with the structural, electronic, and thermodynamic properties of small- to mid-sized alkanes. [3–8] Extremely accurate thermochemical calculations are now possible for small species with up to about five atoms (for recent examples, see refs. [9–16]). However, the severe associated computational requirements coupled with a steep scaling of cost with system size are such that compromises must be made for larger systems. It is necessary to resort to less demanding electronic structure techniques and to employ methods with relatively small basis sets and additivity approximations; often, somewhat ad hoc empirical corrections must also be used. One of the first successful families of approximate ab initio methods for first-principles thermochemistry were the BAC-MPX (X = 2 and 4) techniques. [17] Even more successful have been the Gaussian-n (Gn) approaches of Pople and co-workers, G1, [18] G2, [19] and G3, [20] and their modified versions, including G2(MP2,SVP), [21,22] the CBS methods, [23] as well as the Weizmann-n (Wn) variant W1. [24] Common to all of these model chemistries is a relatively wide range of applicability which derives from judicious approximations made in the electronic structure treatments. All of these methods should be distinguished from the more accurate model chemistries known as W2, [25,26] W3, [10] HEAT, [12] and the focal-point approach, [25,26] which can only be applied to small molecules, albeit with considerably higher accuracy.

The G2(MP2,SVP) model [21,22] is one of the simplest versions of the Gn series that can be applied to larger molecules. In this model enthalpies of formation of molecules at 0 K, \( \Delta H^0_{\text{f}} \), are calculated from theoretical atomization energies and accepted values of enthalpies of formation for gaseous atoms. To adjust these values to 298.15 K, that is, to obtain \( \Delta H^0_{\text{f,298}} \), one must carefully calculate thermal corrections to the enthalpies of molecules of interest are added via the standard formulas of statistical thermodynamics, using the rigid-rotor, harmonic oscillator (RRHO) approximation. [9,27] According to ref. [28], the G2(MP2,SVP) model supplies better standard enthalpies of formation for small noncyclic hydrocarbons than the seemingly more rigorous G2 model from which it derives, perhaps due to its excellent parameterization. Furthermore, literature studies show that accuracy of the G2(MP2,SVP) model chemistry improves when (a) one corrects the G2(MP2,SVP) energy of the carbon atom for the first-order spin–orbit (SPO) interaction (thereby lowering atomization energies, which has the effect of raising the molecular enthalpies of formation), [22b] and (b) the low-frequency internal rotations are treated as free rotors. [27b] Nevertheless, since the relevant higher-level corrections (HLC) of the G2(MP2) model chemistries have been fitted without consideration of SPO and using the CODATA value of \( \Delta H^0_{\text{f,C}} \), one must be careful when assessing such improvements.

As shown here, the remarkable accuracy of G2(MP2,SVP) for standard enthalpies of formation deteriorates for larger alkanes. Moreover, a decidedly systematic feature is noted: apart from methane, standard enthalpies of formation are underestimated for all species studied thus far.

This Communication reveals that the systematic degradation of G2(MP2,SVP) enthalpies of formation with molecular size derives to a large extent, from the value used for the enthalpy of formation of the carbon atom. This is not completely surprising, since the CODATA value of \( \Delta H^0_{\text{f,C}} \) has a relatively large uncertainty, ±0.45 kJ mol\(^{-1}\), and associated problems have been noted in previous publications. [12–14,30] Table 1 lists experimental as well as G2(MP2,SVP) standard enthalpies of formation for a number of C1–C6 alkanes. In the calculations, only the principal isotopic species was considered for each molecule, and the RRHO approximation was applied to obtain enthalpy corrections. For molecules with conformational flexibility, the number of unique conformers is listed in parentheses in Table 1; an ideal mixture of freely interconverting conformers was assumed and Boltzmann statistics were used to determine the population of each conformer. To the best of our knowledge, there are no other alkanes with a small number of conformers for which dependable experimental enthalpies of formation are available. Note here that the experimental enthalpies of formation of the alkanes C4 and C6 have the largest deviation from the appropriate values of the comm...
putational models investigated. Leaving out these data from Model A, that is, the original G2 calculations, excellent, independent of the choice of the model. However, and the rmsd.

cule contains, the larger the discrepancy generally becomes significantly to 0.987.

dicates that the linear correlation between the experimental and the computed enthalpies of formation is excellent, independent of the choice of the model. However, Model A, that is, the original G2(2P,2V) method, systematically underestimates the standard enthalpies of formation of alkanes, except for methane. The more carbon atoms the molecule contains, the larger the discrepancy generally becomes between the computed and the experimental values. The linear regression coefficient for Model A is $r = 0.971$.

Model B data reveal a significant improvement in the computed G2(2P,2V) enthalpies of formation. The 6.0 kJ mol$^{-1}$ rmsd of Model A decreases to 3.7 kJ mol$^{-1}$, and $r$ increases significantly to 0.987.

The inclusion of the SPO stabilization for the carbon atom (from the averaged state that is addressed in non-relativistic calculations) improves both the rmsd and $r$ values of Models A and B. The linear regression coefficient of Model D is an impressive 0.999. Furthermore, the rmsd between Models A and C, 1.7 kJ mol$^{-1}$, decreases substantially, to 0.7 kJ mol$^{-1}$, when Model B is improved to Model D. Thus, it is considerably more

Table 1. Experimental and G2(2P,2V) standard enthalpies of formation of alkanes [kJ mol$^{-1}$].

<table>
<thead>
<tr>
<th>C$_n$</th>
<th>Molecules$^a$</th>
<th>Expt.$^a$</th>
<th>Model A</th>
<th>Model B</th>
<th>Model C</th>
<th>Model D</th>
<th>Model E</th>
<th>Model F</th>
</tr>
</thead>
<tbody>
<tr>
<td>C$_1$</td>
<td>methane (1)</td>
<td>$-74.549 \pm 0.060$</td>
<td>$-74.50$</td>
<td>$-74.04$</td>
<td>$-74.16$</td>
<td>$-73.70$</td>
<td>$-73.68$</td>
<td>$-73.45$</td>
</tr>
<tr>
<td>C$_2$</td>
<td>ethane (1)</td>
<td>$-83.800 \pm 0.300$</td>
<td>$-83.89$</td>
<td>$-82.97$</td>
<td>$-83.21$</td>
<td>$-82.29$</td>
<td>$-82.25$</td>
<td>$-81.79$</td>
</tr>
<tr>
<td>C$_3$</td>
<td>propane (1)</td>
<td>$-104.700 \pm 0.500$</td>
<td>$-105.42$</td>
<td>$-104.04$</td>
<td>$-104.40$</td>
<td>$-103.02$</td>
<td>$-102.96$</td>
<td>$-102.46$</td>
</tr>
<tr>
<td>C$_4$</td>
<td>butane (2)</td>
<td>$-125.600 \pm 0.670$</td>
<td>$-126.53$</td>
<td>$-124.69$</td>
<td>$-125.17$</td>
<td>$-123.33$</td>
<td>$-123.24$</td>
<td>$-123.21$</td>
</tr>
<tr>
<td>C$_5$</td>
<td>isobutane (1)</td>
<td>$-135.007 \pm 0.410$</td>
<td>$-136.28$</td>
<td>$-134.44$</td>
<td>$-134.92$</td>
<td>$-133.08$</td>
<td>$-133.00$</td>
<td>$-132.37$</td>
</tr>
<tr>
<td>C$_6$</td>
<td>pentane (4)</td>
<td>$-146.400 \pm 0.670$</td>
<td>$-147.47$</td>
<td>$-145.17$</td>
<td>$-145.76$</td>
<td>$-143.46$</td>
<td>$-143.36$</td>
<td>$-143.97$</td>
</tr>
<tr>
<td>C$_7$</td>
<td>isopentane (2)</td>
<td>$-154.500 \pm 0.840$</td>
<td>$-156.05$</td>
<td>$-153.75$</td>
<td>$-154.34$</td>
<td>$-152.04$</td>
<td>$-151.94$</td>
<td>$-151.96$</td>
</tr>
<tr>
<td>C$_8$</td>
<td>neopentane (1)</td>
<td>$-168.500 \pm 1.000$</td>
<td>$-172.37$</td>
<td>$-170.07$</td>
<td>$-170.66$</td>
<td>$-168.36$</td>
<td>$-168.26$</td>
<td>$-167.25$</td>
</tr>
<tr>
<td>C$_9$</td>
<td>hexane (12)</td>
<td>$-167.200 \pm 0.790$</td>
<td>$-168.88$</td>
<td>$-166.12$</td>
<td>$-166.83$</td>
<td>$-164.07$</td>
<td>$-163.94$</td>
<td>$-165.23$</td>
</tr>
<tr>
<td>C$_{10}$</td>
<td>2,2-dimethylbutane (1)</td>
<td>$-186.100 \pm 1.000$</td>
<td>$-190.07$</td>
<td>$-187.31$</td>
<td>$-188.03$</td>
<td>$-185.26$</td>
<td>$-185.14$</td>
<td>$-184.84$</td>
</tr>
<tr>
<td>C$_{11}$</td>
<td>2,3-dimethylbutane (2)</td>
<td>$-178.300 \pm 1.000$</td>
<td>$-181.73$</td>
<td>$-178.97$</td>
<td>$-179.68$</td>
<td>$-176.92$</td>
<td>$-176.80$</td>
<td>$-176.97$</td>
</tr>
<tr>
<td>C$_{12}$</td>
<td>2,2,3-trimethylbutane (1)</td>
<td>$-204.800 \pm 1.100$</td>
<td>$-212.22$</td>
<td>$-209.00$</td>
<td>$-209.83$</td>
<td>$-206.61$</td>
<td>$-206.46$</td>
<td>$-206.54$</td>
</tr>
<tr>
<td>C$_{13}$</td>
<td>2,2,3,4-tetramethylbutane (1)</td>
<td>$-225.900 \pm 1.900$</td>
<td>$-238.35$</td>
<td>$-234.67$</td>
<td>$-235.62$</td>
<td>$-231.94$</td>
<td>$-231.77$</td>
<td>$-230.83$</td>
</tr>
<tr>
<td>C$_{14}$</td>
<td>di-tert-butylmethane (1)</td>
<td>$-241.569 \pm 0.832$</td>
<td>$-255.58$</td>
<td>$-251.44$</td>
<td>$-252.51$</td>
<td>$-248.37$</td>
<td>$-248.18$</td>
<td>$-248.20$</td>
</tr>
<tr>
<td>C$_{15}$</td>
<td>tri-tert-butylmethane (1)</td>
<td>$-235.738 \pm 2.460$</td>
<td>$-244.25$</td>
<td>$-238.27$</td>
<td>$-239.81$</td>
<td>$-233.83$</td>
<td>$-233.56$</td>
<td>$-234.56$</td>
</tr>
</tbody>
</table>

- **Statistical data included in the last four rows of the table:**
  - C: linear correlation coefficient;
  - r: linear regression coefficient referring to computed versus experimental enthalpies of formation (kJ mol$^{-1}$); and rmsd, root-mean-square deviation between experimental and computed enthalpies of formation (kJ mol$^{-1}$).
  - [a] Numbers in parentheses refer to the number of non-isomorphic conformers considered during the computation of the enthalpies of formation.
  - [b] The experimental value listed for methane comes from the ATC database of Ruscic et al.; the remaining experimental values are from the NIST WebBook.

![Figure 1](https://example.com/figure1.png)
important to employ a seemingly improved enthalpy of formation of C\textsubscript{gas} in the computation than the inclusion of the SPO correction. This is not unexpected due to the fact that the latter was effectively absorbed into the parameterization of the method.

Figure 1 reflects an important result of this Communication, as it shows the rmsd as a function of \(\Delta H^0(C_{\text{gas}})\) results in an estimate of 711.68 kJ mol\(^{-1}\) for \(\Delta H^0(C_{\text{gas}})\). The Model E data in Table 1 were obtained with this estimate.

An analysis similar to that of Model E was performed for Model F. This resulted in another, slightly higher estimate for \(\Delta H^0(C_{\text{gas}})\), 711.91 kJ mol\(^{-1}\). The Model F data in Table 1 were obtained with this estimate. Relative to Model E, the rmsd decreases only slightly and the change in the regression coefficient is marginal.

Significance tests for Models A through F suggest that Model A is significantly different from all the other models, Models B and C are somewhat similar, while differences in Models D, E, and F are statistically insignificant. Results from Models D, E, and F, based on the G2(MP2,SVP) technique without refitting the HLC, suggest a total correction of the reference dissociation asymptote of 0.80–1.05 kJ mol\(^{-1}\) per carbon atom. Out of this, about 0.35 kJ mol\(^{-1}\) can be ascribed to SPO, the rest of it is suggested to be due to the use of the (imprecise) CODATA value of \(\Delta H^0(C_{\text{gas}})\). Averaging Model-D, E, and F \(\Delta H^0(C_{\text{gas}})\) values gives the best empirical estimate of this study, \(\Delta H^0(C_{\text{gas}}) = (711.75 \pm 0.45)\) kJ mol\(^{-1}\).

In summary, the original G2(MP2,SVP) model can be improved slightly for alkanes by including the spin–orbit effect in carbon, as already noted in ref. [28]. It is also clear that treating the low-frequency internal rotations as free rotors—as suggested in ref. [27]—results in only a marginal (statistically insignificant) improvement in the predicted enthalpies of formation of larger alkanes. The most significant source of systematic error in enthalpies of formation for larger organic compounds based on atomization-energy schemes appears to be that associated with the enthalpy of formation of atomic carbon. The best empirical estimate of this study for the enthalpy of formation of C\textsubscript{gas} \((711.75 \pm 0.45)\) kJ mol\(^{-1}\) suggests that the CODATA value—\((711.194 \pm 0.45)\) kJ mol\(^{-1}\)—is too low by roughly 0.5 kJ mol\(^{-1}\). This estimate for \(\Delta H^0(C_{\text{gas}})\) is in good agreement with an ab initio value\([31]\) \((711.65 \pm 0.32)\) kJ mol\(^{-1}\), and with an interim Active Thermochemical Tables (ATcT)\([32]\) value of Ruscic\([33]\) quoted in ref. [12], \((711.79 \pm 0.21)\) kJ mol\(^{-1}\). The suggested increase in \(\Delta H^0(C_{\text{gas}})\) will increase the enthalpies of formation of the radicals CH\textsubscript{13} and CH\textsubscript{14} computed from accurate ab initio total atomization energies. These new values, in turn, considerably improve the agreement between enthalpies of formation of CH\textsubscript{2} computed via different isodesmic reactions within the HEAT protocol (see footnote 101 of ref.[12]). This Communication has focused on the G2(MP2,SVP) model chemistry because it is sufficiently simple to apply, so that calculations on relatively large molecules are possible, thereby allowing the problems pointed out here to be quantified numerically. The main conclusion of this paper supports a suggestion based on the ATcT approach\([30]\), namely, that establishing a new value for \(\Delta H^0(C_{\text{gas}})\) is needed. This will affect all theoretical-model chemistries based upon atomization energies and all modeling studies and chemical-engineering calculations employing, directly or indirectly, \(\Delta H^0(C_{\text{gas}})\).

**Acknowledgements**

G.T. performed his work as a János Bolyai research fellow. G.T., A.G.C., and M.K. received support from the Scientific Research Fund of Hungary (OTKA) through grants T042825, T047185, and D048583, respectively. This work was partly supported by the U.S. Department of Energy, Division of Chemical Sciences, Geosciences, and Biosciences of the Office of Basic Energy Sciences, under Contract No. W-31-109-ENG-38 (BR, Argonne). J.F.S. acknowledges support of the Robert A. Welch and United States National Science Foundations. This research also forms part of an effort by a Task Group of the International Union of Pure and Applied Chemistry (IUPAC, Project no. 2003-024-1-100) to determine structures, vibrational frequencies, and thermodynamic functions of free radicals of importance in atmospheric chemistry.

**Keywords:** carbon · enthalpy of formation · G2(MP2,SVP) · hydrocarbons · thermochemistry

---

[33] B. Ruscic, unpublished interim result from Active Thermochemical Tables version 1.25 using the Core (Argonne) Thermochemical Network version 1.048.

Received: February 4, 2006
Published online on July 14, 2006