Low cost estimation of the contribution of post-CCSD excitations to the total atomization energy using density functional theory calculations

H.R. Sánchez, R. Pis Diez *

Departamento de Química, CEQUINOR, Centro de Química Inorgánica (CONICET, UNLP), Facultad de Ciencias Exactas, UNLP, CC 962, 1900 La Plata, Argentina

A B S T R A C T

Based on the A_λ diagnostic for multireference effects recently proposed [U.R. Fogueri, S. Kozuch, A. Karton, J.M. Martin, Theor. Chem. Acc. 132 (2013) 1], a simple method for improving total atomization energies and reaction energies calculated at the CCSD level of theory is proposed. The method requires a CCSD calculation and two additional density functional theory calculations for the molecule. Two sets containing 139 and 51 molecules are used as training and validation sets, respectively, for total atomization energies. An appreciable decrease in the mean absolute error from 7–10 kcal mol^{-1} for CCSD to about 2 kcal mol^{-1} for the present method is observed. The present method provides atomization energies and reaction energies that compare favorably with relatively recent scaled CCSD methods.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

In the theoretical calculation of molecular thermodynamic properties the total atomization energy (TAE) becomes usually the main contribution and the harder to calculate.

In the context of ab-initio methods, the calculation of TAE is usually carried out within the coupled-cluster (CC) framework [1]. However, accurate results require the calculation of higher excitations and large enough basis sets, which are very demanding from a computational point of view when they are compared with other alternatives.

The first available choice is the CC method including single and double excitations and perturbative corrections to account for triple excitations, CCSD(T) [2]. The CCSD(T) method is robust and accurate enough when the system under study is dominated by a single reference and, moreover, it is available in almost any software package devoted to quantum chemical calculations. For most applications, the main drawback of this method lies in its computational cost, which scales as $\mathcal{N}^6$ and as $\mathcal{N}^7$ for CCSD [3,4] and CCSD(T), respectively, $\mathcal{N}$ being the system size. Due to such scalings with the system size, many approximations to CCSD(T) have been developed in recent years [5–10].

When post Hartree–Fock methods based on a single Slater determinant, as those derived from the single-reference CC framework, provide results with relatively large errors when they are compared with methods that include higher excitations it is said that multireference effects are present. This is due to the fact that the Slater determinant obtained with the Hartree–Fock method presents not enough flexibility to account for the energy lowering known as static correlation energy. When this is the case, multiconfiguration methods could be used to construct a wavefunction that could recover the static component of the correlation energy [11].

It is then important to quantify those multireference effects to judge the reliability of the post Hartree–Fock calculation. Several tests have been proposed to diagnostic the multireference character in a CCSD calculation. The so-called $\mathcal{T}_1$ diagnostic is based on the norm of the t_1 vector, which is scaled by the number of correlated electrons in the closed-shell system [12]. When $\mathcal{T}_1$ is greater than 0.02 the CCSD method is probably unreliable to obtain accurate results. Janssen and Nielsen proposed another diagnostic for closed-shell systems, D_1, which is based on the matrix 2-norm of the t_1 vector [13]. Extensions of the D_1 diagnostic to open-shell configurations were also proposed [14,15].

More recently, two new diagnostics have been proposed, which are based on the fact that the static component of the correlation energy is related to triple and higher excitations within the CC framework. Karton and coworkers proposed that the percentage contribution of quadruple and quintuple excitations to the TAE, %TAEe[$T_4 + T_5$], is a good diagnostic of multireference effects

* Corresponding author.
E-mail addresses: hernan.sanchez@quimica.unlp.edu.ar (H.R. Sánchez), pis_diez@quimica.unlp.edu.ar (R. Pis Diez).

[16]. Moreover, the authors also found that the corresponding percentage contribution of perturbative triple excitations to the TAE, %TAEe[(T)], exhibits a good linear correlation with %TAEe[$T_4 + T_5$]. This fact suggests that %TAEe[(T)] could also be used as a multireference diagnostic with the additional advantage of being more cheaper than the diagnostic derived from higher excitations.

Following similar ideas, Fogueri and coworkers proposed a new diagnostic for multireference effects based on the density functional theory (DFT) [17]. The diagnostic is defined as $A_\lambda = \lambda^{-1}(1 - \text{TAEe}[\lambda]/\text{TAEe}[0])$, where TAEe[$\lambda$] is the TAE calculated with a given hybrid exchange-correlation functional, in which λ is the fraction of exact exchange, and TAEe[0] is the corresponding TAE obtained with the pure version of that exchange-correlation functional, that is, without exact exchange. The authors found that A_λ shows a good linear correlation with %TAEe[$T_4 + T_5$]. Furthermore, A_λ seems to be very insensitive to the choice of the exchange-correlation functional according to results reported in Table 2 of Ref. [17].

Interestingly, it can be argued a priori that as both A_λ and %TAEe[(T)] correlate well with %TAEe[$T_4 + T_5$], they must also correlate well to each other. This correlation was in fact characterized with a squared correlation coefficient, R^2, greater than 0.9 [17].

Then, it is proposed in the present Letter that A_λ can be used to estimate the contribution of post-CCSD excitations to the TAE. That estimation requires only two additional DFT calculations, which implies a negligible increase in the overall computational cost because DFT calculations scale as $\mathcal{N}^3 - \mathcal{N}^4$. Moreover, it is pertinent to remark that accurate TAE values allow the reliable calculation of some reaction energies. Thus, economical and accurate TAE values would expand the calculation of reaction energies to more complex reactions.

2. Method

The W4-11 set for computational thermochemistry [16,17] is used to test our proposal. The dataset has a total of 140 very diverse molecules. As it has been done elsewhere, [16,17] the Be_2 molecule is excluded from the test set due to its extremely anomalous behavior. The post-CCSD contributions to the electronic part of the TAE values from different levels of theory are obtained from the Supporting Information found in Refs. [16,17]. In particular, those contributions are calculated as %TAEe[post-CCSD] = %TAEe[(T)] + %TAEe[$T_3 - (T)$] + %TAEe[$T_4 + T_5$], where (T) is the contribution of the triple excitations calculated perturbatively after the CCSD calculation, whereas $T_3 - (T)$ and $T_4 + T_5$ are the post-CCSD(T) contribution. It is worth mentioning that TAEe values were obtained using the frozen-core and non-relativistic approximations and, moreover, they include neither zero point energy corrections nor thermal corrections.

Ordinary least squares regression is employed to correlate %TAEe[post-CCSD] with the $A_{0.25}$ values obtained with the hybrid GGA PBE0 functional [18] and its pure counterpart, the GGA PBE functional [19]. The regression parameters, together with the corresponding $A_{0.25}$ value, are then used to estimate the post-CCSD percentage contribution to the TAEe with an equation of the form %TAEe[post-CCSD] = $\alpha + \beta A_{0.25}$. Finally, that value is used in conjunction with the TAEe calculated at CCSD level of theory, TAEe[CCSD], to obtain a final estimation for the TAEe according to TAEe = 100 TAEe[CCSD]/(100−%TAEe[post-CCSD]).

To get a quantitative estimate of the predictive power of the present model an internal validation is carried out. The internal validation is used to avoid the utilization of any other test set that requires a very high level of theory to achieve accurate enough reference values. On the other hand, although the W4-11 dataset could be split into a validation set and a test set, the diversity

of the molecules in the original dataset could easily lead to a wrong choosing of those sets. For the internal validation process the Leave-One-Out method [20] is employed. Within this method, 139 different linear models are fitted. Each model is formed using 138 of the original 139 molecules, whereas the remaining molecule is used to validate the model.

A subset of 51 molecules from the overall set of 212 molecules used by Feller to calibrate the CCSD(T*) method for the calculation of TAE was used as a further validation set [21]. Despite the fact that the reference values reported by that author neglect excitations higher than triple excitations, they are accurate enough for the purpose of validation. For the present calculations, geometry optimizations were performed at the PBE/def2-QZVP and PBE0/def2-QZVP levels of theory, and the total electronic energies were those corresponding to the optimized geometries. Both TAEe[λ] and TAEe[0] are then calculated by subtracting the total electronic energies of the corresponding constituent atoms in their ground states. Finally, TAEe[CCSD] values were calculated at the CCSD/aug-cc-pV(5+d)Z//PBE0/def2-QZVP level of theory.

To explore further the performance of the present proposal, a set of 14 TAE values and a set of 19 reaction energies obtained with the spin-component scaled CCSD method (SCS-CCSD) [22] are used for comparison. Present reaction energies were calculated as the difference between the TAE of the reactants and the TAE of the products.

All the statistical calculations were carried out with the SciPy [23] and the scikit-learn software packages [24]. Geometry optimizations and electronic structure calculations were carried out with the G09 program [25].

3. Results and discussion

From the linear regression analysis we obtained a value of 16.901 for α and a value of 0.090053 for β, with a squared correlation coefficient (R^2) of 0.91. The entire W4-11 dataset is used for fitting the model.

The comparison of the model built using the original 139 molecules in the fitting set with the 139 different models obtained through the Leave-One-Out method shows a mean absolute difference of only 0.030 kcal mol^{-1}, being the largest absolute difference of 0.28 kcal mol^{-1}. Although the differences are small enough for practical purposes, the individual errors of the Leave-One-Out procedure are considered for making assertions and comparisons concerning the present method.

Table 1 shows the reference values of the TAEe of every molecule in the W4-11 dataset, obtained as the contribution of TAEe[SCF], TAEe[CCSD], TAEe[(T)], TAEe[$T_3 - (T)$], and TAEe[$T_4 + T_5$], as reported in Ref. [17]. The errors of the CCSD and CCSD(T) methods, as well as the errors of the method proposed in the present work, are also shown in the table. The relationship between the reference values of the TAEe and the TAEe values obtained with the present method is shown in Figure 1. An impressive decrease of the mean absolute error (MAE) from 9.82 kcal mol^{-1} for the CCSD level of theory to 1.99 kcal mol^{-1} for the present method is found. The MAE for the CCSD(T) method, on the other hand, is only 0.38 kcal mol^{-1}. The maximum absolute errors (MAX) are 30.8, 14.0 and 3.5 kcal mol^{-1} for the CCSD, the present proposal and the CCSD(T) methods, respectively. The two-sided p-value that compares the CCSD method and the present method with the Wilcoxon test is about 10^{-24}, indicating that our method is a clear improvement over the CCSD level of theory for the calculation of total atomization energies.

It is also of interest to compare the performance of the present method with the results obtained with other methods for the whole W4-11 dataset. As can be seen in Table 6 of Ref. [17], only

Table 1

Reference values (RV) for the total atomization electronic energies of the 139 molecules in the test set considered in this work. Individual errors for the CCSD method (ϵ_{CCSD}), for our proposed estimation (ϵ^m_{CCSD}) and for the CCSD(T) method ($\epsilon_{CCSD(T)}$) are given. Mean absolute error (MAE) and maximum absolute error (MAX) are also provided. Both reference values and errors are given in kcal mol^{-1}.

Molecule	RV	ϵ_{CCSD}	ϵ^m_{CCSD}	$\epsilon_{CCSD(T)}$	Molecule	RV	ϵ_{CCSD}	ϵ^m_{CCSD}	$\epsilon_{CCSD(T)}$	Molecule	RV	ϵ_{CCSD}	ϵ^m_{CCSD}	$\epsilon_{CCSD(T)}$
Oxirene	453.6	14.5	0.0	−0.1	CH$_2$NH$_2$	480.4	6.9	0.4	0.0	BH$_3$	280.1	0.8	−1.0	0.0
Oxirane	649.0	11.6	−0.4	0.2	CH$_3$NH	473.1	6.3	0.3	−0.1	BN	104.0	19.2	−5.1	0.3
Dioxirane	408.9	15.0	2.0	−0.3	CH$_3$NH$_2$	580.5	7.2	−0.3	0.1	BF	181.8	4.0	1.4	0.1
Ketene	530.7	13.2	0.6	−0.1	CF$_2$	258.3	9.3	3.0	−0.1	NH	83.0	1.3	0.8	−0.1
Acetaldehyde	675.3	11.6	0.1	0.0	N$_2$H	224.0	9.9	2.2	−0.4	NH$_2$	182.2	2.6	1.1	−0.1
Formic	500.2	12.5	2.5	0.1	t-N$_2$H$_2$	295.7	10.2	2.2	−0.3	HCN	311.7	9.5	0.9	−0.3
Acetic	801.1	16.1	0.7	0.1	N$_2$H$_4$	437.1	8.4	1.7	0.0	HOF	158.5	8.6	3.4	−0.4
Methanol	512.0	6.8	0.5	0.1	FO$_2$	134.7	23.7	−2.4	−3.0	AlH	73.7	0.1	0.2	0.0
Ethanol	808.6	10.3	−1.0	0.2	FOOF	152.6	27.6	1.1	−1.8	AlH$_3$	213.8	0.2	−1.1	0.0
Glyoxal	632.4	17.3	1.9	−0.2	AlF$_3$	431.0	7.5	3.7	0.4	AlF	163.2	3.1	1.9	0.2
t-HCOH	321.6	7.4	1.4	−0.2	Si$_2$H$_6$	535.9	2.9	−4.1	0.0	AlCl	122.3	2.9	−1.5	0.1
c-HCOH	316.8	7.6	1.5	−0.2	P$_4$	288.8	21.2	−11.4	0.6	SiH	73.9	0.5	−0.1	−0.1
HCNH	334.7	9.0	0.7	−0.2	SO$_2$	259.6	16.2	1.9	−0.4	SiH$_4$	325.0	1.0	−1.8	−0.1
HOCN	408.0	13.7	2.2	−0.1	SO$_3$	345.7	20.6	2.7	−0.3	SiO	192.0	8.5	1.0	−0.2
HONC	348.5	14.3	2.2	−0.2	OCS	334.3	14.8	−0.9	−0.3	SiF	142.3	3.5	1.7	0.1
HNCO	432.5	14.7	1.9	−0.1	CS$_2$	279.1	16.2	−5.0	−0.5	CS	171.4	10.2	−2.5	−0.5
HCNO	362.6	17.7	1.8	−0.4	S$_2$O	208.0	17.7	−2.3	−0.8	H$_2$	109.5	0.0	0.3	0.0
c-N$_2$H$_2$	290.4	10.4	2.3	−0.3	S$_3$	167.7	18.3	−6.7	−1.2	OH	107.0	1.7	1.0	−0.1
HNNN	330.2	19.4	1.9	−0.8	S$_4$-C$_{2v}$	233.7	30.8	−14.0	−2.4	HF	141.4	2.1	1.3	0.1
HNC	296.7	8.9	0.5	−0.1	BeCl$_2$	223.6	4.9	−3.1	0.0	H$_2$O	232.6	3.5	1.3	0.0
t-HONO	312.1	17.5	4.0	−0.8	CCl$_2$	176.8	11.1	−2.4	−0.5	CH	84.1	1.0	0.2	−0.1
c-HONO	311.7	17.3	3.9	−0.7	AlCl$_3$	312.9	6.8	−5.8	0.3	CH$_2$-trip	189.9	1.0	−0.3	0.0
NH$_2$Cl	247.5	6.9	0.4	−0.1	ClCN	283.6	12.6	−0.1	−0.2	CH$_3$	306.8	1.9	−0.5	0.0
t-HOOO	233.1	20.8	3.5	−2.3	OClO	127.7	17.9	−0.1	−0.6	CH$_4$	419.2	2.9	−0.8	0.0
c-HOOO	232.9	19.1	4.3	−1.9	ClOO	126.4	22.9	−5.5	−3.5	CCH	264.0	8.2	−1.1	−0.4
CH$_3$F	421.8	5.2	0.9	0.1	Cl$_2$O	101.3	11.4	0.3	−0.3	C$_2$H$_2$	403.0	8.4	−1.0	−0.1
CH$_2$F$_2$	436.6	7.8	2.4	0.2	BN$_3$PI	104.7	7.2	−0.4	−0.8	NH$_3$	297.4	3.9	0.9	0.0
CF$_4$	477.7	12.9	4.3	0.4	CF	132.4	5.1	1.8	−0.1	C$_2$	145.7	19.7	−3.9	−0.4
SiH$_3$F	382.7	2.8	−0.1	0.3	BeF$_2$	306.9	5.7	2.7	0.1	N$_2$	227.6	9.8	2.2	−0.4
SiF$_4$	576.9	9.4	3.3	0.7	CH$_2$C	358.1	7.2	−0.5	−0.3	CO	258.7	8.1	1.2	−0.1
C$_2$H$_3$F	571.4	10.2	0.3	0.0	CH$_2$CH	443.9	6.9	−1.1	−0.2	CN	180.2	11.3	0.7	−0.9
C$_2$H$_5$F	719.1	8.8	−0.6	0.2	C$_2$H$_4$	561.7	7.4	−1.6	0.0	NO	152.3	9.9	1.9	−0.4
FCCF	383.1	13.7	2.7	0.0	CH$_2$NH	437.8	8.6	0.5	−0.1	O$_2$	120.5	9.8	1.7	−0.6
HCCF	395.8	11.1	1.0	0.0	HCO	278.2	8.6	1.3	−0.2	OF	53.1	7.7	1.8	−0.9
HCOF	402.4	11.1	2.9	0.0	H$_2$CO	373.3	8.0	1.5	0.0	F$_2$	39.1	8.3	1.5	−0.7
F$_2$CO	419.1	13.4	3.9	0.2	CO$_2$	388.3	13.9	2.6	−0.1	PH$_3$	241.9	2.2	−0.6	−0.2
CH$_2$-sing	181.1	2.2	−0.3	−0.3	HNO	205.4	10.5	2.8	−0.5	HS	87.6	1.1	−0.2	−0.1
Propane	1004.3	9.9	−3.9	0.2	NO$_2$	227.1	20.4	0.7	−1.1	H$_2$S	183.6	2.3	−0.6	−0.1
Propene	857.9	11.1	−3.1	0.1	N$_2$O	269.6	19.5	2.0	−0.7	HCl	107.3	1.5	−0.4	0.0
Propyne	701.8	11.9	−2.3	0.1	O$_3$	147.3	28.5	−4.2	−2.9	SO	125.9	8.6	0.3	−0.2
Allene	700.4	12.2	−1.8	0.1	HOO	175.3	8.6	3.6	−0.5	ClO	65.2	7.0	0.9	−0.7
B$_2$H$_6$	604.1	4.7	−3.5	−0.2	HOOH	268.7	9.2	3.5	−0.2	ClF	62.7	5.5	1.6	−0.2
BHF$_2$	409.3	5.9	2.2	0.2	F$_2$O	93.9	14.6	2.9	−0.9	P$_2$	116.7	10.3	−3.6	−0.7
BF$_3$	469.0	8.0	3.3	0.3	HOCl	165.9	6.9	1.3	−0.1	S$_2$	103.8	8.0	−2.9	−0.3
C$_2$H$_6$	710.6	6.3	−2.3	0.1	SSH	164.7	6.5	−1.7	−0.2	Cl$_2$	59.6	4.7	−1.0	−0.1
H$_2$CN	342.2	7.8	0.8	−0.3	B$_2$	66.7	11.2	−5.2	−1.4	**MAE**		9.82	1.99	0.38
NCCN	498.6	21.1	0.9	−0.6	BH	84.8	0.5	0.1	−0.1	**MAX**		30.8	14.0	3.5

double hybrid density functionals provide MAE values around or below 2 kcal mol^{-1}. On the other hand, composite thermochemical procedures, which include up to triple excitations within the CC methodology, consistently show MAE values below 2 kcal mol^{-1}. The G3 [26], G4 [27] and CBS-QB3 [28,29] methods provide MAE values in the range of 1.5 and 1.7 kcal mol^{-1}. The G4(MP2)-6X [30] procedure presents a mean absolute error of only 0.89 kcal mol^{-1}. The correlation consistent composite approach (ccCA) methods [31,32], exhibit MAE values between 0.77 and 0.84 kcal mol^{-1}, whereas the W1 procedures [33] become the most accurate methods as they predict total atomization energies with MAE values between 0.56 and 0.66 kcal mol^{-1}. The present proposal compares well with the most accurate double hybrid density functionals and with the G3, G4 and CBS-QB3 methods.

The calculated TAE for the subset of 51 molecules formed from the set of 212 molecules reported in Ref. [21] are shown in Table 2. Figure 2 represents the relationship between the reference values of the total atomization energies and the values obtained with the present method. It can be seen from the table that the present method exhibits a MAE of 2.00 kcal mol^{-1}, whereas the

MAE for the CCSD method is 7.83 kcal mol^{-1}. Interestingly, the MAE of 2.00 kcal mol^{-1} is almost identical to the MAE of 1.99 kcal mol^{-1} obtained previously for the W4-11 set. MAX errors are 25.7 and 6.7 kcal mol^{-1} for the CCSD and the present methods, respectively.

The present method is also compared with the relatively recent spin-component scaled CCSD method of Sherrill et al. [22]. A subset of 14 TAE values and a subset of 19 reaction energies were used in the comparison. It is important to remark that experimental values and CCSD/cc-pVQZ results were taken as reference values for TAE and reaction energies, respectively, in Ref. [22]. In the former case, spin-orbit corrections were also taken into account. Thus, due to the different reference values the comparison should only serve to gain an insight into the relative performance of the different methods. Table 3 lists the errors of the SCS-CCSD method and the present method for TAE and reaction energies. An important decrease in the MAE from 10.6 kcal mol^{-1} for the SCS-CCSD method to 1.6 kcal mol^{-1} for the present method is found for the total atomization energies. The corresponding MAX errors are 16.2 and 3.5 kcal mol^{-1} for the SCS-CCSD method and the present proposal, respectively. It is argued that the large difference in the
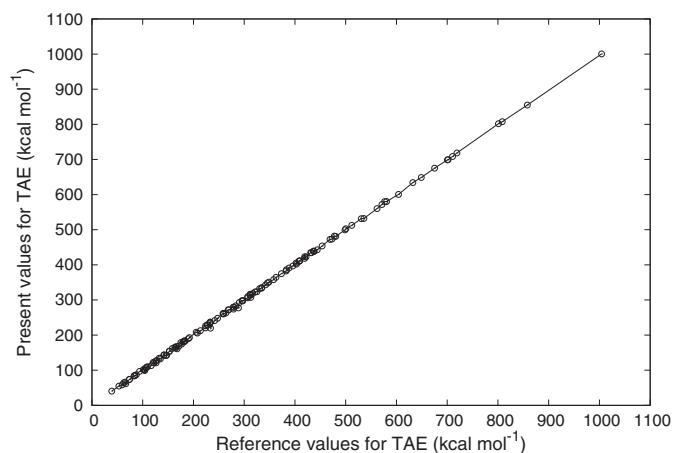
Figure 1. Relationship between the reference values of the electronic total atomization energies of the W4-11 dataset and the values calculated in the present work.
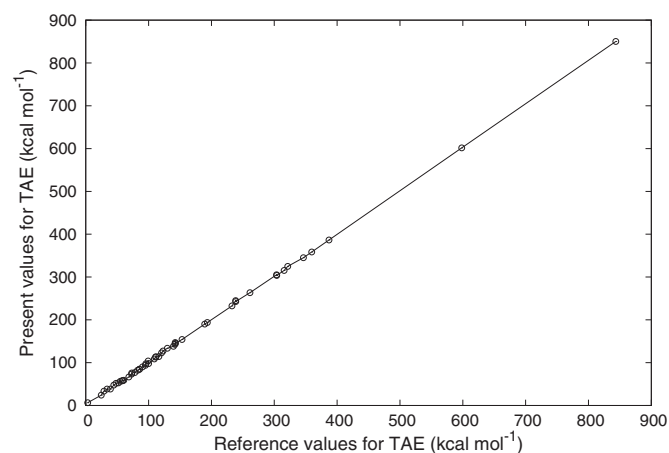
Figure 2. Relationship between the reference values of the electronic total atomization energies of the validation set formed by 51 molecules and the values calculated in the present work.

Table 2
Reference values (RV) for the total atomization electronic energies of the 51 molecules in the validation set. Individual errors for the CCSD (ϵ_{CCSD}) and the present (ϵ_{CCSD}^{m}) methods are given. Mean absolute error (MAE) and maximum absolute error (MAX) are also provided. Both reference values and errors are given in kcal mol^{-1}.

Molecule	RV	ϵ_{CCSD}	ϵ_{CCSD}^{m}	Molecule	RV	ϵ_{CCSD}	ϵ_{CCSD}^{m}
CCl	97.3	−6.7	−1.7	LiN	38.4	−2.7	0.8
PH	73.3	−0.9	−0.3	NF	77.0	−6.1	1.2
SiS	146.9	−8.7	−4.6	ZnO	38.4	−12.9	−4.5
Si$_2$	75.9	−5.3	−2.8	HBr	93.7	0.2	1.2
Al$_2$	33.3	−6.4	−4.3	BO	193.9	−8.0	−0.4
CBr	82.2	−5.0	0.2	NS	113.2	−11.7	−2.2
LiF	138.2	−3.6	1.3	PN	145.0	−13.1	−2.3
AlN	58.3	−5.7	0.4	LiCl	113.6	−2.8	−1.9
PF	108.6	−4.5	0.5	LiO	84.4	−3.2	1.1
BCl	122.8	−5.3	−2.2	H$_2$S$_2$	242.8	−8.1	−4.1
AsF	97.9	−3.7	2.1	PH$_2$	154.1	−2.0	−0.9
PO	142.6	−9.8	−0.7	ClO$_2$	127.1	−20.1	−4.1
Li$_2$	24.1	−0.1	0.6	ClNO	190.2	−18.6	−0.9
AlP	51.8	−6.3	−3.6	CCl$_3$	244.9	−13.8	−6.3
BrCl	56.5	−4.7	−1.4	C$_3$H$_6$	850.2	−13.1	−6.7
FS	84.8	−5.6	0.5	C$_2$H$_5$	601.6	−6.9	−3.4
LiH	57.7	−0.1	0.4	C$_2$H	263.4	−8.8	−2.1
AsN	114.2	−10.6	2.1	NCO	315.3	−15.2	0.3
BrO	58.8	−6.2	1.7	FCN	303.8	−14.0	0.1
Br$_2$	52.8	−3.6	−0.6	HNO$_3$	386.6	−25.7	0.5
I$_2$	47.6	−5.3	−2.9	H$_2$CS	324.8	−9.3	−3.4
AsH	66.5	1.2	2.1	CF$_3$	345.1	−11.9	1.3
As$_2$	90.1	−6.8	0.2	NH$_2$OH	358.6	−9.8	1.0
ZnP	6.5	−3.5	−3.4	FNC	232.6	−14.7	0.1
PS	103.8	−9.8	−4.6	H$_2$SiO	305.2	−9.1	−1.6
BS	133.8	−7.8	−4.0	**MAE**		7.83	2.00
				MAX		25.7	6.7

Table 3
Errors for the SCS-CCSD method ($\epsilon_{SCS-CCSD}$) and for our proposed estimation (ϵ_{CCSD}^{m}) for the total atomization energies of a set of 14 molecules and for the reaction energies of a set of 19 reactions. Mean absolute error (MAE) and maximum absolute error (MAX) are also provided for each set. Errors are given in kcal mol^{-1}. See text for a discussion on the reference values.

Molecule	$\epsilon_{SCS-CCSD}$	ϵ_{CCSD}^{m}	Reaction	$\epsilon_{SCS-CCSD}$	ϵ_{CCSD}^{m}
HF	5.9	1.3	F$_2$ + H$_2$ → 2HF	0.0	0.8
CO	5.8	1.2	F$_2$O + H$_2$ → F$_2$ + H$_2$O	1.2	0.4
N$_2$	10.3	2.2	O$_3$ + 3H$_2$ → 3H$_2$O	5.4	7.2
F$_2$	6.4	1.5	H$_2$O$_2$ + H$_2$ → 2H$_2$O	0.3	1.2
CH$_2$	9.5	0.1	CO + H$_2$ → CH$_2$O	1.1	0.0
H$_2$O	10.8	1.3	CO$_2$ + 3H$_2$ → CH$_4$ + H$_2$O	2.4	3.0
NH$_3$	14.4	0.9	N$_2$ + 3H$_2$ → 2NH$_3$	1.7	1.3
CH$_4$	15.0	0.8	CH$_2$ + H$_2$ → CH$_4$	0.4	0.8
HNO	13.2	2.8	N$_2$O + H$_2$ → N$_2$ + H$_2$O	2.8	1.2
CO$_2$	8.1	2.6	C$_2$H$_2$ + H$_2$ → C$_2$H$_4$	0.4	0.7
CH$_2$O	12.8	1.5	BH$_3$ + 3HF → BF$_3$ + 3H$_2$	1.4	1.3
H$_2$O$_2$	16.0	3.5	CO + H$_2$O → CO$_2$ + H$_2$	1.9	0.4
C$_2$H$_2$	8.8	1.4	C$_2$H$_2$ + HF → C$_2$H$_3$F	0.3	0.4
C$_2$H$_4$	16.2	1.6	HCN + H$_2$O → CO + NH$_3$	1.0	0.1
			HCN + NH$_3$ → N$_2$ + CH$_4$	0.4	0.4
MAE	10.6	1.6	O$_3$ + CH$_4$ → 2H$_2$O + CO	3.0	8.8
MAX	16.2	3.5	2BH$_3$ → B$_2$H$_6$	0.7	1.5
			2CH$_2$ → C$_2$H$_4$	0.0	1.0
			^{3}CH$_2$ → ^{1}CH$_2$	4.6	0.0
			MAE	1.53	1.61
			MAX	5.4	8.8

4. Conclusions

An economic method based on a CCSD calculation and two DFT calculations is proposed to obtain reliable total atomization energies for molecules at their equilibrium geometries. The two DFT calculations are combined to estimate post-CCSD contributions in a very simple manner.

The method was trained with a dataset containing 139 molecules and then it was validated using a dataset of 51 molecules. The mean absolute errors for total atomization energies decrease from 8–10 kcal mol^{-1} for the CCSD method to about 2 kcal mol^{-1} for the present method.

The method was also compared with atomization energies and reaction energies obtained with the SCS-CCSD method. The mean absolute errors of both methods are very similar for reaction energies. In the case of atomization energies, the present method clearly outperforms the SCS-CCSD method for the test set considered. Those results, however, could be due to the fact that SCS-CCSD results were obtained with relatively small basis sets.

absolute errors is due to the fact that SCS-CCSD results are obtained with considerably smaller basis sets. On the other hand, the performance of the SCS-CCSD method and the present proposal are very similar for reaction energies, for which reference values were obtained with the CCSD(T) method and the CC method including up to quintuple excitations, respectively. Differences in the reference values are due to exact triple, quadruple and quintuple excitations on the one hand, and to basis set extensions, on the other hand. It is also interesting to note that the effect of basis set variations is more notorious in atomization energies than in reaction energies due to the differences in speed of convergence in atoms and in molecules. Thus, MAE errors 1.53 and 1.61 kcal mol^{-1} for the SCS-CCSD method and the present proposal, respectively, whereas the corresponding MAX errors are 5.4 and 8.8 kcal mol^{-1}.

Thus, to the best of our knowledge this is the first proposal in which DFT calculations are used to predict the effects of higher excitations in post-Hartree Fock calculations.

Acknowledgments

References

[1] R.J. Bartlett, M. Musiał, Rev. Mod. Phys. 79 (2007) 291.
[2] K. Raghavachari, G.W. Trucks, J.A. Pople, M. Head-Gordon, Chem. Phys. Lett. 157 (1989) 479.
[3] G.D. Purvis, R.J. Bartlett, J. Chem. Phys. 76 (1982) 1910.
[4] P.J. Knowles, C. Hampel, H. Werner, J. Chem. Phys. 99 (1993) 5219.
[5] C. Sosa, J. Geertsen, G.W. Trucks, R.J. Bartlett, Chem. Phys. Lett. 159 (1989) 148.
[6] A.G. Taube, R.J. Bartlett, Czech. Chem. Commun. 70 (2005) 837.
[7] W. Li, P. Piecuch, J. Gour, S. Li, J. Chem. Phys. 131 (2009) 114109.
[8] C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese, J. Chem. Phys. 139 (2013) 134101.
[9] A.E. DePrince, C.D. Sherrill, J. Chem. Theory Comput. 9 (2013) 2687.
[10] A.E. DePrince, C.D. Sherrill, J. Chem. Theory Comput. 9 (2013) 293.
[11] P.G. Szalay, T. Müller, G. Gidofalvi, H. Lischka, R. Shepard, Chem. Rev. 112 (2012) 108.
[12] T.J. Lee, P.R. Taylor, Int. J. Quantum Chem. 36 (1989) 199.
[13] C.L. Janssen, I.M.B. Nielsen, Chem. Phys. Lett. 290 (1998) 423.
[14] M.L. Leininger, I.M.B. Nielsen, T.D. Crawford, C.L. Janssen, Chem. Phys. Lett. 328 (2000) 431.
[15] T.J. Lee, Chem. Phys. Lett. 372 (2003) 362.
[16] A. Karton, S. Daon, J.M. Martin, Chem. Phys. Lett. 510 (2011) 165.
[17] U.R. Fogueri, S. Kozuch, A. Karton, J.M. Martin, Theor. Chem. Acc. 132 (2013) 1.
[18] C. Adamo, V. Barone, J. Chem. Phys. 110 (1999) 6158.
[19] J.P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. 77 (1996) 3865.
[20] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, 2nd edn., Springer, 2009.
[21] D. Feller, J. Phys. Chem. A 119 (2015) 7375.
[22] T. Takatani, E.G. Hohenstein, C.D. Sherrill, J. Chem. Phys. 128 (2008) 124111.
[23] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open Source Scientific Tools for Python, 2001.
[24] F. Pedregosa, et al., J. Mach. Learn. Res. 12 (2011) 2825.
[25] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford, CT, 2010.
[26] A.G. Baboul, L.A. Curtiss, P.C. Redfern, K. Raghavachari, J. Chem. Phys. 110 (1999) 7650.
[27] L.A. Curtiss, P.C. Redfern, K. Raghavachari, J. Chem. Phys. 126 (2007) 84108.
[28] J.A. Montgomery Jr., M.J. Frisch, J.W. Ochterski, G.A. Petersson, J. Chem. Phys. 110 (1999) 2822.
[29] G.P.F. Wood, L. Radom, G.A. Petersson, E.C. Barnes, M.J. Frisch, J.A. Montgomery Jr., J. Chem. Phys. 125 (2006) 094106.
[30] L.A. Curtiss, P.C. Redfern, K. Raghavachari, J. Chem. Phys. 127 (2007) 124105.
[31] N.J. DeYonker, T.R. Cundari, A.K. Wilson, J. Chem. Phys. 124 (2006) 114104.
[32] N.J. DeYonker, T. Grimes, S. Yockel, A. Dinescu, B. Mintz, T.R. Cundari, A.K. Wilson, J. Chem. Phys. 125 (2006) 104111.
[33] J.M.L. Martin, G. de Oliveira, J. Chem. Phys. 111 (1999) 1843.