

# Accuracy of basis-set extrapolation schemes for DFT-RPA correlation energies in molecular calculations

E. Fabiano · F. Della Sala

Received: 2 July 2012 / Accepted: 7 September 2012 / Published online: 11 October 2012  
© Springer-Verlag 2012

**Abstract** We construct a reference benchmark set for atomic and molecular random phase approximation (RPA) correlation energies in a density functional theory framework at the complete basis-set limit. This set is used to evaluate the accuracy of some popular extrapolation schemes for RPA all-electron molecular calculations. The results indicate that for absolute energies, accurate results, clearly outperforming raw data, are achievable with two-point extrapolation schemes based on quintuple- and sextuple-zeta basis sets. Moreover, we show that results in good agreement with the benchmark can also be obtained by using a semiempirical extrapolation procedure based on quadruple- and quintuple-zeta basis sets. Finally, we analyze the performance of different extrapolation schemes for atomization energies.

**Keywords** RPA correlation · Basis-set extrapolation · Complete basis-set limit

## 1 Introduction

The random phase approximation (RPA) for electron correlation was originally developed in the context of the many-body perturbation treatment of the uniform electron gas [1, 2] and later reformulated in the framework of

density functional theory (DFT) via the adiabatic-connection fluctuation-dissipation theorem [3]. In the last years, RPA correlation witnessed an increasing interest in quantum chemistry [4–27], due to its ability to describe with good accuracy a rather large variety of reaction and interaction energies. In this context, RPA is generally employed as a post-Kohn-Sham (KS) approach where the exact exchange (EXX) and the correlation energies are evaluated using the eigenvalues and eigenfunctions from a KS calculation with conventional exchange-correlation (XC) functionals. EXX KS orbitals [28–30] also have been employed via the optimized effective potential (OEP) method, while for a fully variational approach self-consistent EXX-RPA orbitals are required [31–38]. In all cases, practical applications of the method are often hindered by the very slow convergence of the RPA correlation energy with the basis-set dimension [4, 39, 40], which is related to the electron cusp problem [4], in analogy with wave function methods. The cusp problem is strongly reduced for range-separated RPA [11–13, 41, 42], whereas for full-range RPA approaches extrapolation to the complete basis set (CBS) must be employed.

Basis-set extrapolation to the CBS limit has been studied for long in post-Hartree–Fock correlated methods, with the aim of reducing as much as possible the basis-set truncation error without resorting to the brute force employment of extremely large basis sets. Numerical work [43–49] on various correlation treatments of the helium atom has shown that the basis-set incompleteness error is approximately proportional to the inverse third power of the maximum angular momentum, provided that all angular momenta are radially saturated. However, the ideal conditions of fully saturated angular momenta in an atomic system are rather distant from those encountered in real applications on molecules with finite basis sets.

E. Fabiano (✉) · F. Della Sala  
National Nanotechnology Laboratory (NNL), Istituto di  
Nanoscienze-CNR, Via per Arnesano 16, 73100 Lecce, Italy  
e-mail: eduardo.fabiano@nano.cnr.it

F. Della Sala  
Center for Biomolecular Nanotechnologies (CBN) of the Italian  
Institute of Technology (IIT), Via Barsanti,  
73010 Arnesano, Italy

Nevertheless, the correlation energy has been empirically observed to converge regularly enough to allow the development of various accurate extrapolation schemes which have been assessed in several studies in literature [50–54]. On the other hand, no such study has been performed for RPA correlation energies.

In this paper, we aim at filling this gap and shed light on the usefulness and limits of different extrapolation schemes for RPA correlation energies. In this respect, the goal of the present work is to: (1) create a set of benchmark CBS RPA absolute correlation energies to be used as reference for future assessments; (2) validate some extrapolation schemes, usually applied in the context of coupled cluster theory, for the calculation of CBS RPA absolute energies, including semiempirical techniques to approach the extrapolation from relatively small basis sets; (3) test the convergence behavior of different basis sets and the reliability of the extrapolations schemes for atomization energies, which exploit a significant error cancellation effect. Absolute correlation energies will be considered, as in many corresponding studies of correlated methods, because these converge regularly with the basis set and provide an ideal quantity for the study of extrapolation schemes. In this paper, we focus on light elements because they provide a simpler convergence of the correlation energy and can be more easily investigated with large basis sets.

## 2 Computational details

RPA calculations were performed using the eigenvalues and the orbitals from DFT calculations based on the Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional [55]. In all calculations, Dunning’s correlation-consistent basis sets [56–59] (cc-pVnZ,  $n = D, 5, 6, 7$ ) augmented with core and core-valence basis functions [60–62] were employed. These are hereafter denoted as VnZcv with  $n = 4, 5, 6, \text{ and } 7$ . Augmented diffuse basis functions were not considered in the present work since they were shown to slow down RPA basis-set convergence without bringing substantial benefits in most cases [40]. Also, no basis-set superposition error correction was considered for atomization energies. For molecules, experimental geometries were considered [63–76]. All calculations were performed with the TURBOMOLE program package [77] using the implementation described in [4].

## 3 Reference data

The construction of a set of benchmark CBS-limit RPA energies is a fundamental step before any possible assessment work on extrapolation schemes. Unfortunately, in the case of RPA correlation energies, this is a difficult task.

In fact, unlike for MP2 or coupled cluster methods, no explicitly correlated [78] RPA data exist that can serve as an accurate reference. At the same time, a brute force strategy, based on the use of very large basis sets, is hampered by the extremely slow convergence rate of the RPA calculations. Thus, in order to approach the true CBS limit as close as possible and provide accurate reference CBS energies, some form of extrapolation is necessary. However, this introduces inevitably an undesirable degree of uncertainty in the data that can be anyway reasonably reduced by an appropriate control of the possible sources of errors.

In this work, we construct our set of reference RPA correlation energies based on two leading criteria: (1) we employ as a basis for the extrapolation the best results at our disposal (i.e., those from V5Zcv to V7Zcv calculations), in order to recover as much as possible the “theoretical” asymptotic converge of the correlation energy that underlies all extrapolation formulas and thus reduce the extrapolation error. For this reason, we prefer to exclude data from V4Zcv calculations (or lower) that may increase the computational noise due to the incompleteness of the basis set. (2) We fix our best estimate of the RPA correlation energy by averaging over the data obtained from several flexible three-parameter extrapolation formulas, in order to avoid the possible bias characteristic of a specific extrapolation scheme. A similar procedure was recently applied for the construction of a benchmark set of CCSD(T) CBS energies [51]. In addition, we consider in detail the evolution of different extrapolated energies with dimension of the basis sets to fix reasonable bounds to our uncertainty.

In more detail, we consider the popular extrapolation formula [51, 79, 80]

$$E_n = E_\infty + \frac{A}{(n+d)^3} \quad (1)$$

and its variant proposed by Martin [81]

$$E_n = E_\infty + \frac{A}{(n+d)^4}, \quad (2)$$

where  $A$  and  $d$  are the parameters and  $E_\infty$  is the CBS limit of the correlation energy. The free parameter  $d$  accounts for the incompleteness of the basis-set angular momentum saturation [52] and also partially introduces effectively higher-order contributions to the asymptotic correlation expansion, as is readily recognized by considering the alternative form for Eq. (1)

$$E_n = E_\infty + \frac{A}{(n)^3} + \sum_{i=1}^{\infty} (-1)^i \binom{n}{i} \frac{Ad^i}{n^{3+i}}. \quad (3)$$

In addition, we use the extrapolation formula proposed by Bakowies [52, 53]

$$E_n = E_\infty + \frac{A}{n^\alpha}, \quad (4)$$

with  $A$  and  $\alpha$ , the adjustable parameters. Again,  $\alpha$  provides an effective correction for deviations from the leading asymptotic behavior  $\propto n^{-3}$  [52], by partially adding high-order terms to the expansion.

The RPA correlation energies of several atoms and molecules as resulting from V5Zcv, V6Zcv, V7Zcv calculations and the extrapolations with formulas (1–4) are reported in Table 1, together with our best estimate for the CBS-limit RPA correlation energy. All the extrapolation methods yield very close results, with differences  $<1$  mHa, justifying our assumption of taking as best estimate their average value.

We define in addition, in the last column of Table 1, the error  $\delta$  associated with of our benchmark energies in the following way. First, we note that numerical evidence shows that for the two limiting cases of an inverse cubic extrapolation (i.e., using Eq. (1) with  $d = 0$ ) and of an exponential extrapolation, the exact CBS RPA correlation energy is approached from below and above, respectively. This is shown for some representative cases in Fig. 1, where the results of the two extrapolations with increasingly large basis sets are reported, and can be rationalized as follows: on the one hand, the exponential extrapolation is well known to underestimate the CBS limit [50] because of the too fast rate of variation with the basis set (exponential rather than with a power law) and on the other hand, the simple inverse cubic extrapolation is likely to overestimate the CBS limit because the  $1/n^3$  behavior is strictly valid only asymptotically, while for real operative conditions a faster rate of variation with the basis set shall be expected [hence the  $d$  parameter in Eq. (1)]. The best extrapolated values obtained with these two procedures can thus be considered as upper and lower bounds for the true CBS RPA correlation energy. Then, after an analysis of the trends of Fig. 1, we assume as an estimation of the error half of the maximum difference between our best estimate and any of the two bounds for the energy.

The reference energies of Table 1 have thus an estimated accuracy of about 1 mHa or less, except for the last four entries of the table, which display slightly larger errors, with a maximum of 2.4 for  $F_2$ . We note, however that these systems also have the largest absolute correlation energies. In fact, the relative errors for all systems are very low, ranging between 0.11 and 0.25 %, with a mean absolute relative error of 0.16 %.

To further validate our results, we considered for the atoms also RPA calculations with an extremely large basis set (V7Zcv+), constructed by adding to the V7Zcv basis set centered on the atom additional uncontracted basis functions (up to angular momentum  $g$ ) centered at six

points located on the Cartesian axes at  $\pm 1$  Bohr of distance from the atom, for a total of 1,113 basis functions per atom (723 for the H atom). The resulting basis set allows to describe effectively very high angular momentum contributions, thanks to the composition of momenta from functions at different centers, without the need to effectively include basis functions with such high values of angular momentum into our basis set and partially avoiding problems related to the linear dependence of basis functions. The results of the V7Zcv+ calculations, reported in Table 1, show that indeed by using a very large basis set the estimated RPA CBS energy are approached well, with deviations that are always close or lower than the expected uncertainty of the reference values. This result provides a good support for the accuracy of our estimation of the CBS limit of the RPA correlation as well as for the the corresponding errors.

The procedure outlined in this section, and whose results are reported in Table 1, allows therefore to produce an accurate set of CBS RPA correlation energies, improving of several mHa (about 5–10) with respect to V7Zcv results (see bottom part of Table 1). Even more important improvements, up to 20 mHa, are found with respect to V6Zcv calculations. The proposed set, despite its limited dimension, provides in addition a representative selection of molecules constituted from first-row atoms, including doublet and triplet ground states. Thus, it offers a valuable tool for the benchmarking of RPA correlation energies.

#### 4 Two-point extrapolations

In this section, we seek for globally optimized parameters  $d$  and  $\alpha$  to be used in Eqs. (1), (2), and (4), capable of yielding accurate RPA energies with respect to reference values. The use of global parameters in the extrapolation formulas has in fact several advantages in practical applications: (1) it allows to compute the CBS energy through two-point extrapolation formulas instead of using a fitting procedure; (2) with a careful optimization of the global parameters, it allows to effectively include (on an average) high-order effects into the extrapolation, so that smaller basis sets can be used; (3) it produces more systematic errors, favoring error cancelation in many applications. The search for global parameters is also motivated in the present study by the observation that in the individual fittings of RPA correlation energies performed in the previous section, we observed that for each formula all the optimal values of  $d$  or  $\alpha$  were contained in a rather narrow range. Thus, it is conceivable that “average” values of the parameters may exist which produce accurate estimates of the CBS limit.

**Table 1** RPA correlation energies of several atoms and molecules for different basis sets and extrapolation schemes

System	V5Zcv	V6Zcv	V7Zcv	V7Zcv+	$1/(n+d)^3$	$1/(n+d)^4$	$1/n^\alpha$	Best est.	$\delta$
H	-20.3	-20.6	-20.7	-20.9	-21.0	-21.0	-21.0	-21.0	0.1
C	-284.9	-288.5	-290.1	-292.0	-292.6	-293.0	-292.9	-292.8	0.7
N	-327.8	-332.8	-334.9	-336.8	-337.2	-337.5	-337.1	-337.3	0.6
O	-416.5	-423.4	-426.7	-430.4	-431.5	-432.0	-431.9	-431.8	0.9
F	-504.1	-513.1	-517.5	-522.0	-523.9	-524.6	-524.4	-524.3	1.3
Ne	-579.2	-590.5	-595.6	-600.5	-601.9	-602.6	-602.0	-602.2	1.3
H <sub>2</sub>	-80.0	-80.5	-80.7		-81.2	-81.1	-81.2	-81.2	0.2
NH	-399.3	-404.8	-407.2		-410.4	-410.7	-410.4	-410.5	0.6
NH <sub>2</sub>	-468.1	-474.1	-476.8		-480.4	-480.9	-480.6	-480.6	0.7
CH <sub>4</sub>	-491.4	-496.3	-498.6		-501.4	-501.7	-501.4	-501.5	0.6
NH <sub>3</sub>	-533.4	-539.3	-541.8		-544.5	-544.9	-544.5	-544.6	0.7
H <sub>2</sub> O	-565.7	-573.8	-577.4		-581.7	-582.3	-581.8	-581.9	0.9
FH	-582.1	-591.9	-596.3		-602.0	-602.7	-602.2	-602.3	1.1
C <sub>2</sub> H <sub>2</sub>	-756.5	-765.3	-769.4		-774.7	-775.3	-774.9	-775.0	1.1
CN	-774.5	-783.9	-787.8		-792.3	-792.9	-792.3	-792.5	1.0
HCN	-799.8	-809.6	-813.8		-818.7	-819.3	-818.7	-818.9	1.0
CO	-821.6	-832.8	-837.5		-842.8	-843.5	-842.7	-843.0	1.2
N <sub>2</sub>	-833.6	-844.6	-849.3		-855.1	-855.8	-855.2	-855.4	1.2
C <sub>2</sub> H <sub>4</sub>	-833.4	-842.5	-846.7		-852.3	-852.9	-852.5	-852.6	1.1
HCO	-865.9	-877.5	-882.5		-888.4	-889.1	-888.4	-888.6	1.2
H <sub>2</sub> CO	-909.8	-921.6	-926.7		-932.7	-933.5	-932.7	-933.0	1.2
O <sub>2</sub>	-973.1	-987.6	-993.9		-1,001.5	-1,002.4	-1,001.6	-1,001.8	1.5
H <sub>3</sub> COH	-984.0	-996.6	-1,002.1		-1,008.7	-1,009.5	-1,008.8	-1,009.0	1.3
HOOH	-1,077.2	-1,092.6	-1,099.4		-1,107.6	-1,108.6	-1,107.7	-1,108.0	1.6
F <sub>2</sub>	-1,125.2	-1,143.9	-1,152.3		-1,163.0	-1,164.3	-1,163.4	-1,163.6	2.4
ME	17.9	9.0	5.1		0.2	-0.4	0.1		
MAE	17.9	9.0	5.1		0.2	0.4	0.1		
MARE	2.72 %	1.40 %	0.80 %		0.04 %	0.05 %	0.02 %		
MAD	38.4	19.7	11.3		0.5	0.7	0.3		
Std. Dev.	8.7	4.4	2.5		0.1	0.2	0.1		

The last two columns report the best estimate (average of the extrapolated values) and the estimated uncertainty  $\delta$ . At the bottom of the table, we report for the raw data from different basis sets the mean error (ME), mean absolute error (MAE), mean absolute relative error (MARE), maximum absolute deviation (MAD), and standard deviation with respect to the best estimate. All results are in mHa

For our purpose, we consider the two-point extrapolation formulas

$$E_\infty = \frac{E_n(n+d_{opt})^3 - E_m(m+d_{opt})^3}{(n+d_{opt})^3 - (m+d_{opt})^3} \quad (5)$$

$$E_\infty = \frac{E_n(n+d_{opt})^4 - E_m(m+d_{opt})^4}{(n+d_{opt})^4 - (m+d_{opt})^4} \quad (6)$$

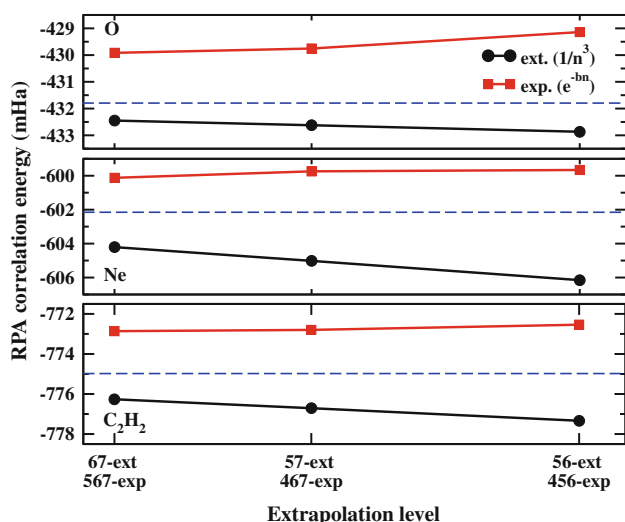
$$E_\infty = \frac{E_n n^{\alpha_{opt}} - E_m m^{\alpha_{opt}}}{n^{\alpha_{opt}} - m^{\alpha_{opt}}}, \quad (7)$$

which are derived directly from Eqs. (1), (2), and (4), respectively, considering two basis sets of cardinal number  $n$  and  $m$ .

The optimization of the global parameters  $d_{opt}$  and  $\alpha_{opt}$  was performed using either  $n = 7, m = 6$  or  $n = 6,$

$m = 5$ , and minimizing the mean absolute error with respect to the reference data of Table 1. The use of a smaller basis set ( $n = 5, m = 4$ ) resulted instead into a failure of the minimization procedure, because no reasonable compromise value was found for  $d_{opt}$  and  $\alpha_{opt}$ , reflecting the difficulty of the V4Zcv basis set to correctly describe the RPA correlation.

The values of the optimized parameters as well the corresponding results for the test systems and global statistics are reported in Table 2. We see that all methods yield similar results in good agreement with the reference values. The differences between the various extrapolation formulas as well as between data obtained using a 67-extrapolation and a 56-extrapolation are negligible ( $<0.4$  mHa). This indicates that the optimized global parameters



**Fig. 1** Extrapolated RPA correlation energies (mHa) of the O and Ne atoms and the C<sub>2</sub>H<sub>2</sub> molecule as a function of different combinations of basis sets. The label *nm-ext* denotes an extrapolation based on Eq. (1) with  $d = 0$  and using VnZcv and VmZcv basis sets; the label *nmk-exp* denotes an exponential extrapolation using VnZcv, VmZcv, and VkZcv basis sets. The *dashed lines* indicate our best estimate for the CBS RPA energy. The *x-axis* uses an inverse cubic scale

can in fact describe very well the correlation convergence behavior, effectively taking into account higher-order contributions. Concerning general trends, we observe that 56-extrapolations appear to slightly underestimate in general the RPA correlation energy, while this effect is reduced for 67-extrapolations. We stress however that because the differences between different methods and with reference data are well below our estimated accuracy for the benchmark set, no quantitative conclusions can be drawn from the results of Table 2, and to practical purposes, all the methods must be considered completely equivalent.

As a final note, it is instructive to highlight that the global optimization of the parameters provides also valuable indications on the convergence behavior of different formulas. Thus, the values obtained from the optimization procedure can be used to shed some light on the quality of different extrapolation schemes. For example, in [52, 53], it is indicated how  $\alpha_{opt}$  provides a measure for the rate of convergence of the extrapolation. In our case, the values found for  $\alpha_{opt}$  are greater than the ideal asymptotic value of 3, indicating the need for an overweight of the energy obtained with the largest basis set with respect to the other one in the extrapolation formula (7), similarly to that found for CCSD calculations [52]. This means that for the considered basis sets, the rate of change of the correlation energy with increasing basis set is larger than it should be asymptotically, that is, the basis sets are rather inadequate to describe the CBS limit. This is not surprising for RPA

calculations that are well known to converge very difficultly to the CBS limit and nicely explains the fact that for 56-extrapolation a slightly larger value of  $\alpha_{opt}$  was found than for 67-extrapolation.

For the other two formulas (Eqs. 5, 6), similar considerations apply. In fact, it is possible to see that all the optimized parameters provide, in the extrapolation procedure, the same relative (over)weight of the energy obtained with the largest basis set with respect to the other one in all formulas. This is made evident in Fig. 2, where we plot the ratios

$$\frac{(n+d)^3}{(m+d)^3}; \quad \frac{(n+d)^4}{(m+d)^4}; \quad \frac{n^\alpha}{m^\alpha} \quad (8)$$

for  $n = 7, m = 6$  (top panel),  $n = 6, m = 5$  (lower panel), at several values of  $d$  or  $\alpha$ . The coincidence of the relative weights for different formulas provides thus a rationale for the fact that all formulas yield practically the same results (note that input energies are the same in all formulas at the same level of extrapolation). Inspection of the figure also indicates that the relative weights obtained from optimized expressions (1.78 and 1.98 for 67-extrapolation and 56-extrapolation, respectively) are not far from the “ideal” ones, obtained considering  $\alpha = 3$  (1.60 and 1.74, respectively).

## 5 Semiempirical extrapolation from small basis sets

As discussed in the previous section, for smaller basis sets, it is not possible to find accurate global parameters to extrapolate the RPA correlation energy to the CBS limit. This problem is clearly related to the inadequacy of the V4Zcv basis set which makes the progression to V5Zcv results not regular enough to be extrapolated with a global formula. The problem can be however partially circumvented by employing a suitable technique specifically developed for extrapolation from small basis sets.

In wave function theory, for example, for CCSD(T) calculations, two main strategies are used to this end: either considering effective empirical parameters for the extrapolation [52, 53, 82, 83] or considering the relatively inexpensive MP2/CBS limit plus a  $\Delta$ CCSD(T) correction evaluated with a smaller basis set [84]. This latter approach is however not straightforward to extend to RPA calculations, because it is based on the fact that CCSD(T) and MP2 energies display a very similar basis-set dependence. This similarity is not verified by RPA and MP2 energies, which instead have a markedly different basis-set behavior (e.g., for the Ne atom passing from the V3Zcv to the V4Zcv basis set, the energy changes by 60, 32, and 31 mHa for RPA, CCSD(T), and MP2, respectively;

**Table 2** RPA correlation energies of test atoms and molecules for different basis sets and global extrapolation schemes

Optimized parameters	67-extrapolation			56-extrapolation			Ref.
	$1/(n+d)^3$ $d_{opt} = -1.33$	$1/(n+d)^4$ $d_{opt} = 0.37$	$1/n^\alpha$ $\alpha_{opt} = 3.78$	$1/(n+d)^3$ $d_{opt} = -1.17$	$1/(n+d)^4$ $d_{opt} = 0.25$	$1/n^\alpha$ $\alpha_{opt} = 3.82$	
H	-20.9	-20.9	-20.9	-20.8	-20.8	-20.8	-21.0 ± 0.1
C	-292.1	-292.1	-292.1	-292.0	-292.0	-292.0	-292.8 ± 0.7
N	-337.5	-337.5	-337.5	-337.8	-337.8	-337.8	-337.3 ± 0.6
O	-431.0	-431.0	-431.0	-430.2	-430.2	-430.2	-431.8 ± 0.9
F	-523.1	-523.1	-523.1	-522.1	-522.1	-522.1	-524.3 ± 1.3
Ne	-602.0	-602.0	-602.0	-601.8	-601.8	-601.8	-602.2 ± 1.3
H <sub>2</sub>	-81.1	-81.1	-81.1	-81.0	-81.0	-81.0	-81.2 ± 0.2
NH	-410.4	-410.4	-410.4	-410.2	-410.2	-410.2	-410.5 ± 0.6
NH <sub>2</sub>	-480.3	-480.3	-480.3	-480.0	-480.0	-480.0	-480.6 ± 0.7
CH <sub>4</sub>	-501.4	-501.4	-501.4	-501.3	-501.3	-501.3	-501.5 ± 0.6
NH <sub>3</sub>	-544.9	-544.9	-544.9	-545.3	-545.3	-545.3	-544.6 ± 0.7
H <sub>2</sub> O	-581.9	-581.9	-581.9	-581.8	-581.8	-581.8	-581.9 ± 0.9
FH	-602.0	-602.0	-602.0	-601.7	-601.7	-601.7	-602.3 ± 1.1
C <sub>2</sub> H <sub>2</sub>	-774.5	-774.5	-774.5	-774.0	-774.0	-774.0	-775.0 ± 1.1
CN	-792.8	-792.8	-792.8	-793.2	-793.2	-793.2	-792.5 ± 1.0
HCN	-819.2	-819.1	-819.2	-819.4	-819.4	-819.4	-818.9 ± 1.0
CO	-843.4	-843.4	-843.4	-843.9	-843.8	-843.9	-843.0 ± 1.2
N <sub>2</sub>	-855.4	-855.4	-855.4	-855.5	-855.4	-855.5	-855.4 ± 1.2
C <sub>2</sub> H <sub>4</sub>	-852.1	-852.0	-852.0	-851.5	-851.5	-851.5	-852.6 ± 1.1
HCO	-888.8	-888.8	-888.8	-889.0	-889.0	-889.0	-888.6 ± 1.2
H <sub>2</sub> CO	-933.1	-933.1	-933.1	-933.3	-933.3	-933.3	-933.0 ± 1.2
O <sub>2</sub>	-1,001.9	-1,001.9	-1,001.9	-1,002.0	-1,001.9	-1,001.9	-1,001.8 ± 1.5
H <sub>3</sub> COH	-1,009.0	-1,009.0	-1,009.0	-1,009.1	-1,009.0	-1,009.0	-1,009.0 ± 1.3
HOOH	-1,108.0	-1,107.9	-1,108.0	-1,108.0	-1,107.9	-1,108.0	-1,108.0 ± 1.6
F <sub>2</sub>	-1,163.0	-1,163.0	-1,163.0	-1,162.4	-1,162.3	-1,162.4	-1,163.6 ± 2.4
ME	0.1	0.2	0.2	0.2	0.3	0.2	
MAE	0.3	0.3	0.3	0.6	0.6	0.6	
MARE	0.07 %	0.07 %	0.07 %	0.13 %	0.13 %	0.13 %	
MAD	1.1	1.2	1.1	2.2	2.2	2.2	
SD	0.4	0.4	0.4	0.8	0.8	0.8	

The last lines report the mean error (ME), the mean absolute error (MAE), the mean absolute relative error (MARE), the maximum absolute deviation (MAD) from reference, and the standard deviation of each data set. All results are in mHa

changing from the V4Zcv to the V5Zcv level, the change is 25, 11, and 12 mHa for RPA, CCSD(T), and MP2, respectively).

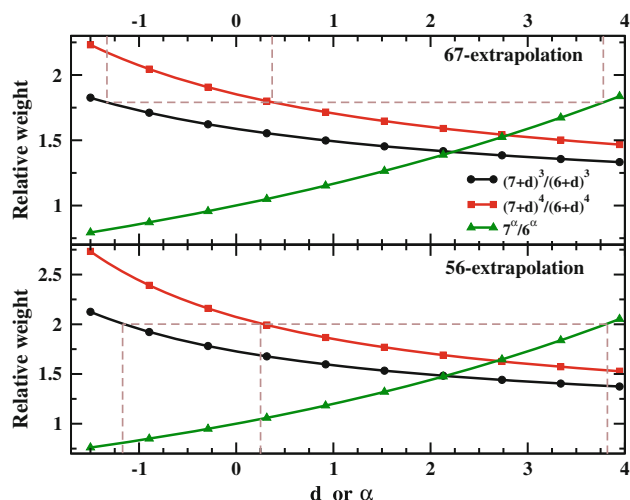
Therefore, in this paper, we follow the first strategy and, following a similar approach as that proposed by Bakowies [52, 53] for MP2 and CCSD calculations, we introduce atom-dependent parameters which account for the irregular behavior in the progression of V4Zcv–V5Zcv energies. The CBS energy is defined as

$$E_\infty \simeq \frac{E_n n^\gamma - E_m m^\gamma}{n^\gamma - m^\gamma} \quad (9)$$

where  $n = 5$ ,  $m = 4$ , and

$$\gamma = \frac{\sum_{i=1}^N n_i \gamma_i}{\sum_{i=1}^N n_i} \quad (10)$$

with the index  $i$  running over all atoms in the system,  $N$  the total number of atoms, and  $n_i$  the number of electrons of the  $i$ -th atom. The atomic parameters  $\gamma_i$  are fixed by fitting Eq. (9) for single atoms and a minimal set of homodimers (H<sub>2</sub>, N<sub>2</sub>, O<sub>2</sub>, F<sub>2</sub>) except for carbon where C<sub>2</sub>H<sub>2</sub> was considered. Considering only atomic data is in fact not sufficient to yield accurate and balanced results for all the systems (the mean absolute error and the mean relative errors are twice as big as in the case of the minimal training



**Fig. 2** Relative weights (Eq. 8) of the energies used in Eqs. (5), (6), and (7), for different values of the  $d$  and  $\alpha$  parameters. The dashed lines indicate the location of optimized parameters

set and the maximum deviation is 4.7 mHa for HOOH), while a large training set would be likely to lead to artifacts in the fitting. The minimal set considered here provides instead a good accuracy limiting as much as possible an arbitrary selection of reference systems.

The optimized parameters and the resulting RPA correlation energies are reported in Table 3.

The method performs indeed rather well, yielding a mean absolute error of only 1 mHa, in line with the estimated accuracy of the reference set, a mean absolute relative error of 0.16 %. Only in 9 cases over 25, the extrapolation errors exceed the intrinsic accuracy of the reference energies (i.e.,  $\Delta E^{ref}/\delta > 1$ ), while in all cases the extrapolated energies improve over the raw data obtained at the V5Zcv level and even with respect to the raw V7Zcv data.

The present semiempirical extrapolation scheme can thus be a valuable tool for RPA calculations on large systems, as it allows to achieve a good accuracy at a relatively small computational cost. We recall in fact that, since RPA correlation calculations scale as  $O(N^6)$  (which can be reduced to  $O(N^4 \log N)$  using the resolution of the identity (RI) technique [17]), where  $N$  is the number of basis functions, and a cc-pVnZ basis set contains approximately

$$N = \frac{(n+1)(n+\frac{3}{2})(n+2)}{3} \quad (11)$$

basis functions [80], the computational time required by the semiempirical approach is about  $204^6/91^6 \sim 127$  (30 when RI is used) times faster, that is, two orders of magnitude less, than a V7Zcv calculation and still  $140^6/91^6 \sim 13$  (6 when RI is used) times faster, that is, one order of magnitude less, than a V6Zcv calculation.

## 6 Atomization correlation energies

To conclude our work, we provide in this section a short discussion on RPA correlation atomization energies that are defined as energy differences between a molecule and its constituent atoms. These can be easily constructed from the data of Tables 1, 2, and 3 or alternatively, for methods using a two-point extrapolation formula with global (i.e., system independent) parameters, they can be obtained by applying the extrapolation formula directly to the raw atomization energies (i.e., computed with two given basis sets). The statistics of atomization energies, using as reference data the atomization energies obtained from the best estimated RPA energies of Table 1 are reported in Table 4.

The inspection of Table 4 reveals that several approaches are capable to yield very good results, with mean absolute errors close to 1 mHa and mean relative errors smaller than 1 %. In particular, all the calculations using raw basis-set results (except V4Zcv) perform remarkably well, showing a significantly better performance with respect to the case of absolute energies, thanks to the cancelation of systematic errors in the present case. Interestingly, little or no benefit comes instead from the use of extrapolation techniques. In fact, 67-extrapolation results are in line with V7Zcv results, while 56-extrapolation atomization energies are even worse than V5Zcv ones and of similar quality as the V4Zcv results. This result shall be rationalized in terms of an inevitable small but random computational noise that affects the extrapolation data (especially if small basis sets are used as a base for the extrapolation). The small inaccuracies of extrapolated data can be in fact almost irrelevant for absolute energies (below 1 mHa), but can easily sum to several mHa in the case of atomization energies (especially for many atoms molecules). Finally, we remark the good performance of the semiempirical extrapolation method, which yields atomization energies in good agreement with the reference ones, with a mean absolute relative error of only 0.75 % and very small maximum absolute deviation and standard deviation.

The results of Table 4 show that the quality of each extrapolation procedure to yield accurate atomization energies cannot be directly inferred from the results obtained for absolute energies, because atomization energies are energy differences and thus a complex error propagation can occur. As we showed this effect is more important for low-level approaches and is somehow amplified for extrapolated results, because of the presence of unsystematic errors due to the numerical procedure. It is thus finally important to note that this issue applies also for our reference atomization energies, that were obtained from the best estimated RPA absolute energies of Table 1, although in this case, we may expect the effect to be rather limited due to the high quality of the extrapolation.

**Table 3** Atomic parameters ( $\gamma_i$ ), extrapolated RPA correlation energies ( $E_c^{RPA}$ ), differences with respect to reference energies ( $\Delta E^{ref}$ ), and ratio between  $\Delta E^{ref}$  and the expected accuracy of each reference energy ( $\delta$ ), for the systems of the test set of Table 1

System	$\gamma_i$	$E_c^{RPA}$	$\Delta E^{ref}$	$\Delta E^{ref} / \delta$
H	3.10	-20.9	0.1	1.3
C	3.25	-292.2	0.6	0.9
N	3.35	-337.7	-0.5	-0.8
O	3.23	-430.9	1.0	1.1
F	3.15	-523.3	1.0	0.7
Ne	3.28	-602.2	0.0	0.0
H <sub>2</sub>		-81.2	0.0	-0.1
NH		-410.4	0.1	0.1
NH <sub>2</sub>		-480.5	0.2	0.2
CH <sub>4</sub>		-502	-0.5	-0.8
NH <sub>3</sub>		-547.2	-2.5	-3.6
H <sub>2</sub> O		-583.3	-1.3	-1.5
FH		-603.4	-1.1	-1.0
C <sub>2</sub> H <sub>2</sub>		-774.9	0.1	0.1
CN		-793.4	-1.0	-1.0
HCN		-819.8	-0.9	-0.9
CO		-844.8	-1.8	-1.5
N <sub>2</sub>		-855.2	0.2	0.1
C <sub>2</sub> H <sub>4</sub>		-852.5	0.1	0.1
HCO		-890	-1.4	-1.2
H <sub>2</sub> CO		-934.7	-1.7	-1.4
O <sub>2</sub>		-1,003.1	-1.3	-0.9
H <sub>3</sub> COH		-1,010.7	-1.7	-1.3
HOOH		-1,110.1	-2.1	-1.3
F <sub>2</sub>		-1,165.1	-1.5	-0.6
ME			-0.7	
MAE			0.9	
MARE			0.16 %	
MAD			2.6	
SD			1.0	

The last lines report the mean error (ME), the mean absolute error (MAE), the mean absolute relative error (MARE), the maximum absolute deviation (MAD) from reference, and the standard deviation of the extrapolated energies. All results are in mHa

Nevertheless, we have to remark that possibly errors of few mHa are plausible for some systems, so that in general, the set cannot be fully considered as an accurate benchmark set of atomization energies, although it could be safely considered as reference for a qualitative discussion of the trends obtained from different extrapolation methods.

## 7 Conclusions

The slow basis-set convergence of the RPA correlation energy is a major problem for the development and the

application of the method in quantum chemistry and solid-state physics. In fact, a careful assessment of the RPA methodology and related methods (e.g., RPA+ [14], RPA + SE [19]) requires the existence of a set of benchmark absolute and relative RPA correlation energies as well as the availability of accurate extrapolation techniques. Practical applications, on the other hand, would greatly benefit from efficient extrapolation schemes with well-calibrated levels of confidence.

Surprisingly however only little effort has been dedicated in literature to study the CBS limit of RPA correlation energies and assess the effectiveness of different extrapolation procedures in this context. In this paper, we aimed at pushing this work one step further and considered the problem in a more systematic way, constructing a benchmark set of reference RPA correlation energies and studying the behavior of different extrapolation schemes in the RPA framework. Our work can be summarized in the following conclusions:

- The extremely slow convergence of the RPA correlation energy with the basis-set dimension prevents the computation of highly accurate reference energies. Even at the V7Zcv level in fact the energies can be incorrect up to 10 mHa. Nevertheless, we showed that a careful extrapolation procedure allows to construct an accurate benchmark set with a well-defined estimation of the uncertainty. In the present work, we estimate our reference data to have an accuracy close to 1 mHa for most of the systems.
- We analyzed several two-point extrapolation formulas against our benchmark set of absolute RPA correlation energies. It turns out that good results can only be achieved if global parameters are optimized in Eqs. (5), (6), and (7), as to effectively reproduce high-order terms in the theoretical asymptotic energy expansion ( $\propto 1/n^3$ ). When this procedure is followed, all the considered formulas provide equivalent results. Moreover, with a careful optimization of the global parameters, very good results can also be achieved from 56-extrapolation procedures.
- RPA correlation energies evaluated at the V4Zcv level appear to be of too low quality to serve as a basis in an extrapolation procedure with globally optimized parameters. This fact indicates that in the case of 45-extrapolations, high-order terms beyond the  $1/n^3$  one play a crucial role and are thus dominant in the asymptotic expansion. Therefore, RPA energies obtained at the quadruple-zeta level of theory cannot be considered well converged with respect to the one-particle expansion in basis set and shall not be used in applications. Nevertheless, the introduction of atomic-based parameters within a semiempirical extrapolation



**Table 4** Statistics for the atomization energies computed with different approaches with respect to the best estimated RPA results (see Table 1)

Method	ME	MAE	MARE (%)	MAD	SD
V4Zcv	-1.5	1.7	1.27	5.0	1.6
V5Zcv	0.2	1.2	0.93	3.0	1.4
V6Zcv	0.9	1.3	1.05	2.6	1.4
V7Zcv	0.9	1.2	0.95	2.3	1.1
56-extr.	1.5	1.8	1.38	3.4	1.5
67-extr.	1.0	1.1	0.84	2.0	0.8
Semiempirical	1.0	1.0	0.75	2.2	0.6

All results are in mHa. For two-point extrapolation schemes, the results are reported only once, since all the procedures are essentially superimposable for atomization energies

*ME* mean error, *MAE* mean absolute error, *MARE* mean absolute relative error, *MAD* maximum absolute deviation, *SD* standard deviation

scheme can strongly reduce the problem and provide extrapolated RPA absolute energies of good quality (within 2 mHa from the reference). The semiempirical extrapolation method can thus be considered as a good tool for the CBS extrapolation in RPA applications on large systems.

- When energy differences are considered, an unpredictable error propagation occurs and accurate CBS results are even more hard to achieve than the absolute correlation energies. However, because an error cancellation occurs for the systematic errors (e.g., systematic underestimation of the energy by incomplete basis sets), we finally found that most approaches yield results that agree within 2 mHa. Because no highly accurate reference data exist to assess such small differences, no clear preference can be expressed for any of the considered approaches and 2 mHa ( $\sim 1.2$  kcal/mol,  $\sim 0.05$  meV) must be considered the level of confidence of state-of-the-art RPA calculations on atomization energies. Therefore, caution must be always employed when small energy differences are considered.

We mention finally that, following the encouraging results of the present study, further investigations will be needed to assess in detail the basis-set dependence of RPA energies and the accuracy of extrapolation schemes for energy differences (e.g., atomization energies), which are a main quantity in many practical computational studies. In fact, when energy differences are considered, an unpredictable error propagation may occur and accurate CBS results are harder to achieve than the absolute correlation energies, despite the former can be directly derived as an algebraic sum of the latter. From our calculations, it can be easily seen indeed that all the approaches considered in this work yield very similar atomization energies for the molecules of the test set, with differences mostly below 2–3 mHa. However, no clear trend can be identified between the different approaches, so that no highly

accurate reference data can be established to assess such small differences. This traces back to the fact that systematic errors (e.g., systematic underestimation of the energy by incomplete basis sets) may be expected to cancel out in this case, but extrapolation procedures may introduce in general unsystematic errors which worsen the results with respect to the raw data obtained from the corresponding basis sets.

**Acknowledgments** We thank TURBOMOLE GmbH for providing us with the TURBOMOLE program package, and M. Margarito for technical support. This work was funded by the ERC Starting Grant FP7 Project DEDOM, Grant Agreement No. 207441.

## References

1. Bohm D, Pines D (1953) Phys Rev 92:609
2. Gell-Mann M, Brueckner KA (1957) Phys Rev 106:364
3. Langreth DC, Perdew JP (1977) Phys Rev B 15:2884
4. Furche F (2001) Phys Rev B 64:195120
5. Fuchs M, Gonze X (2002) Phys Rev B 65:235109
6. Furche F, Van Voorhis T (2005) J Chem Phys 122:164106
7. Dobson JF, Wang J, Dinte BP, McLennan K, Le HM (2005) Int J Quantum Chem 101:579
8. Scuseria GE, Henderson TM, Sorensen DC (2008) J Chem Phys 129:231101
9. Grüneis A, Marsman M, Harl J, Schimka L, Kresse G (2009) J Chem Phys 131:154115
10. Janesko BG, Scuseria GE (2009) J Chem Phys 131:154106
11. Janesko BG, Henderson TM, Scuseria GE (2009) J Chem Phys 130:081105
12. Toulouse J, Gerber IC, Jansen G, Savin A, Ángyán JG (2009) Phys Rev Lett 102:096404
13. Toulouse J, Zhu W, Ángyán JG, Savin A (2010) Phys Rev A 82:032502
14. Ruzsinszky A, Perdew JP, Csonka GI (2010) J Chem Theor Comput 6:127
15. Nguyen HV, Galli G (2010) J Chem Phys 132:044109
16. Paier J, Janesko BG, Henderson TM, Scuseria GE, Grüneis A, Kresse G (2010) J Chem Phys 132:094103
17. Eshuis H, Yarkony J, Furche F (2010) J Chem Phys 132:234114
18. Jansen G, Liu RF, Ángyán JG (2010) J Chem Phys 133:154106
19. Ren X, Tkatchenko A, Rinke P, Scheffler M (2011) Phys Rev Lett 106:153003

20. Eshuis H, Furche F (2011) *J Phys Chem Lett* 2:983
21. Ruzsinszky A, Perdew JP, Csonka GI (2011) *J Chem Phys* 134:114110
22. Ángyán JG, Liu RF, Toulouse J, Jansen G (2011) *J Chem Theory Comput* 7:3116
23. Heßelmann A (2011) *J Chem Phys* 134:204107
24. Eshuis H, Bates JE, Furche F (2012) *Theor Chem Acc* 131:1084
25. Paier J, Ren X, Rinke P, Scuseria GE, Grüneis A, Kresse G, Scheffler M (2012) *New J Phys* 14:043002
26. Ren X, Rinke P, Blum V, Wiefierink J, Tkatchenko A, Sanfilippo A, Reuter K, Scheffler M (2012) *New J Phys* 14:053020
27. Ren X, Rinke P, Joas C, Scheffler M (2012) *J Mater Sci*. doi: [10.1007/s10853-012-6570-4](https://doi.org/10.1007/s10853-012-6570-4)
28. Heßelmann A, Görling A (2010) *Mol Phys* 108:359
29. Heßelmann A, Görling A (2011) *Phys Rev Lett* 106:093001
30. Heßelmann A, Görling A (2011) *Mol Phys* 109:2473
31. Godby RW, Schlüter M, Sham LJ (1986) *Phys Rev Lett* 56:2415
32. Kotani T (1998) *J Phys Cond Matt* 10:9241
33. Niquet YN, Fuchs M, Gonze X (2003) *Phys Rev A* 68:032507
34. Grüning M, Marini A, Rubio A (2006) *J Chem Phys* 124:154108
35. Hellgren M, von Barth U (2007) *Phys Rev B* 76:075107
36. Hellgren M, von Barth U (2010) *J Chem Phys* 132:044101
37. Hellgren M, Rohr DR, Gross EKV (2012) *J Chem Phys* 136:034106
38. Verma P, Bartlett RJ (2012) *J Chem Phys* 136:044105
39. Harl J, Kresse G (2008) *Phys Rev B* 77:045136
40. Eshuis H, Furche F (2012) *J Chem Phys* 136:084105
41. Janesko BG, Henderson TM, Scuseria GE (2009) *J Chem Phys* 131:034110
42. Zhu W, Toulouse J, Savin A, Ángyán JG (2010) *J Chem Phys* 132:244108
43. Schwartz C (1962) *Phys Rev* 126:1015
44. Schwartz C (1963) *Methods Comput Phys* 2:241
45. Carroll DP, Silverstone HJ, Metzger RM (1979) *J Chem Phys* 71:4142
46. Schmidt HM, von Hirschhausen H (1983) *Phys Rev A* 28:3179
47. Hill RN (1985) *J Chem Phys* 83:1173
48. Kutzelnigg W, Morgan JD III (1992) *J Chem Phys* 96:4484
49. Schmidt HM, Linderberg J (1994) *Phys Rev A* 49:4404
50. Müller T (2006) Basis set accuracy and calibration in quantum chemistry. In: Grotendorst J, Blugel S, Marx D (eds) *Computational nanoscience: do it yourself!* John von Neumann Institute for Computing, Julich, NIC Series, vol 31, ISBN 3-00-017350-1, pp 19-43
51. Feller D, Peterson KA, Grant Hill J (2011) *J Chem Phys* 135:044102
52. Barkowies D (2007) *J Chem Phys* 127:084105
53. Barkowies D (2007) *J Chem Phys* 127:164109
54. Varandas AJC (2007) *J Chem Phys* 126:244105
55. Perdew JP, Burke K, Ernzerhof M (1996) *Phys Rev Lett* 77:3865
56. Dunning TH Jr (1989) *J Chem Phys* 90:1007
57. Woon DE, Dunning TH Jr (1994) *J Chem Phys* 100:2975
58. Feller D, Peterson KA (1999) *J Chem Phys* 110:8384
59. Feller D, Sordo JA (2000) *J Chem Phys* 113:485
60. Woon DE, Dunning TH Jr. (1995) *J Chem Phys* 103:4572
61. Peterson KA, Dunning TH Jr (2002) *J Chem Phys* 117:10548
62. Grant Hill J, Mazumder S, Peterson KA (2010) *J Chem Phys* 132:054108
63. Boyd DRJ (1955) *J Chem Phys* 23:922
64. Venkateswarlu P, Gordy W (1955) *J Chem Phys* 23:1200
65. Redington RL, Olson WB, Cross PC (1962) *J Chem Phys* 36:1311
66. Herzberg G (1966) *Electronic spectra and electronic structure of polyatomic molecules*. Van Nostrand, New York
67. Sverdlov LM, Kovner MA, Krainov EP (1974) *Vibrational spectra of polyatomic molecules*. Wiley, New York
68. Tsuboi M, Overend J (1975) *J Mol Spect* 52:256
69. Hirota E (1979) *J Mol Spect* 77:213
70. Hoy AR, Bunker PR (1979) *J Mol Spect* 74:1
71. Huber KP, Herzberg G (1979) *Molecular spectra and molecular structure. IV. Constants of diatomic molecules*. Van Nostrand Reinhold Co, New York
72. Herbst E, Messer JK, DeLucia FC (1984) *J Mol Spect* 108:42
73. Junttila ML, Lafferty WJ, Burkholder JB (1994) *J Mol Spect* 164:583
74. Gurvich LV, Veyts IV, Alcock CB (1989) *Thermodynamic properties of individual substances*, 4th edn. Hemisphere Pub Co, New York
75. Kuchitsu, K (eds) (1998) *Structure of free polyatomic molecules—basic data*. Springer, Berlin
76. NIST Diatomic Spectral Database. <http://www.physics.nist.gov/PhysRefData/MolSpec/Diatomic/index.html>
77. TURBOMOLE V6.2 2010, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>
78. Kutzelnigg W (1985) *Theor Chem Acta* 68:445
79. Helgaker T, Klopper W, Koch H, Noga J (1997) *J Chem Phys* 106:9639
80. Halkier A, Helgaker T, Jørgensen P, Klopper W, Koch H, Olsen J, Wilson AK (1998) *Chem Phys Lett* 286:243
81. Martin JML (1996) *Chem Phys Lett* 259:669
82. Truhlar DG (1998) *Chem Phys Lett* 294:45
83. Fast PL, Sanchez ML, Truhlar DG (1999) *J Chem Phys* 111:2921
84. Pitoňák M, Riley KE, Neogrády P, Hobza P (2008) *Chem Phys Chem* 9:1636