Accurate Complete Basis Set Extrapolation of Direct Random Phase Correlation Energies

Pál D. Mezei,[*,†] Gábor I. Csonka,[*,†] and Adrienn Ruzsinszky[‡]

[†]Department of Inorganic and Analytical Chemistry, Budapest University of Technology and Economics, H-1521 Budapest, Hungary
[‡]Department of Physics, Temple University, Philadelphia, Pennsylvania 19122, United States

Ⓢ Supporting Information

ABSTRACT: The direct random phase approximation (dRPA) is a promising way to obtain improvements upon the standard semilocal density functional results in many aspects of computational chemistry. In this paper, we address the slow convergence of the calculated dRPA correlation energy with the increase of the quality and size of the popular Gaussian-type Dunning's correlation consistent aug-cc-pVXZ split valence atomic basis set family. The cardinal number X controls the size of the basis set, and we use $X = 3-6$ in this study. It is known that even the very expensive $X = 6$ basis sets lead to large errors for the dRPA correlation energy, and thus complete basis set extrapolation is necessary. We study the basis set convergence of the dRPA correlation energies on a set

dRPA correlation energy vs. X; 65 hydrocarbons

of 65 hydrocarbon isomers from CH_4 to C_6H_6. We calculate the iterative density fitted dRPA correlation energies using an efficient algorithm based on the CC-like form of the equations using the self-consistent HF orbitals. We test the popular inverse cubic, the optimized exponential, and inverse power formulas for complete basis set extrapolation. We have found that the optimized inverse power based extrapolation delivers the best energies. Further analysis showed that the optimal exponent depends on the molecular structure, and the most efficient two-point energy extrapolations that use $X = 3$ and 4 can be improved considerably by considering the atomic composition and hybridization states of the atoms in the molecules. Our results also show that the optimized exponents that yield accurate $X = 3$ and 4 extrapolated dRPA energies for atoms or small molecules might be inaccurate for larger molecules.
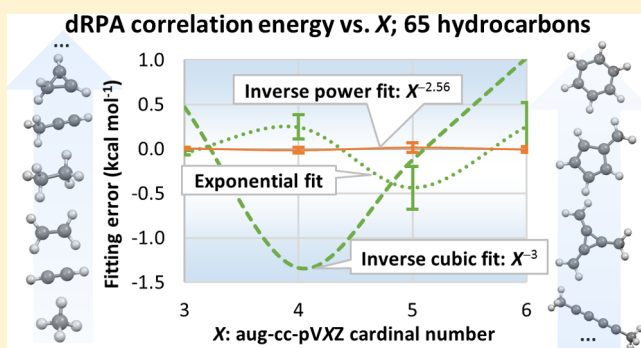
INTRODUCTION

The direct random phase approximation (noted as RPA or dRPA)[1−3] is a promising way to obtain improvements upon the standard semilocal density functional results. It can be constructed either in the framework of density functional theory or in the coupled cluster formalism.[4] RPA is accurate for intra- and intermolecular noncovalent interactions,[5] for adsorption,[6,7] for interlayer interactions,[8] and for van der Waals bonded solids.[9] RPA captures the nonpairwise-additive feature of dispersion interactions missed by many pairwise *a posteriori* molecular mechanics-like dispersion corrections. For covalent bond breaking reaction energies, RPA performs moderately.[4] However, RPA describes the short-range correlation poorly. It fails seriously in situations that lead to short-range rearrangement of the electronic structure. It also fails for ionization and atomization processes, where the number of electron pairs are changing. RPA overestimates the ionization potentials[10] and underestimates the atomization energies of molecules[11] and solids.[12] Furthermore, RPA suffers from self-interaction and static correlation error, which lead to the underestimation of the barrier heights[13] and to the overestimation of the energy of breaking bonds.[14] These errors can be understood from the perspective of fractional charges

and fractional spins.[15] Beyond RPA methods, like RPA+,[16] RPA ++,[17] RPA plus second-order screened exchange (SOSEX),[18] RPA plus renormalized single excitations (rSE),[13] and several other approaches, are applied to correct these errors.[19−21]

Another well-documented problem with the RPA calculations is the slow convergence of the RPA correlation energy with the increase of the basis set size. The slow convergence can be traced back to the poor description of the electron–electron Coulomb cusps by smooth orbital products.[22,23] The complete basis set (CBS) extrapolation method provides a simple way to deal with this problem. There exist more sophisticated, but very expensive methods, like the explicitly correlated wave function,[22,24−27] transcorrelated,[28,29] diffusion quantum Monte Carlo,[30−32] and other methods that treat this error.[33−35] The CBS extrapolations are based on the convergence of the energy terms with respect to the increasing completeness of a series of basis sets. The CBS energy limit of the infinite basis set can be extrapolated from convergent energy series calculated with systematically increasing quality Gaussian-type[36] or plane-wave[37] basis sets. One such widely
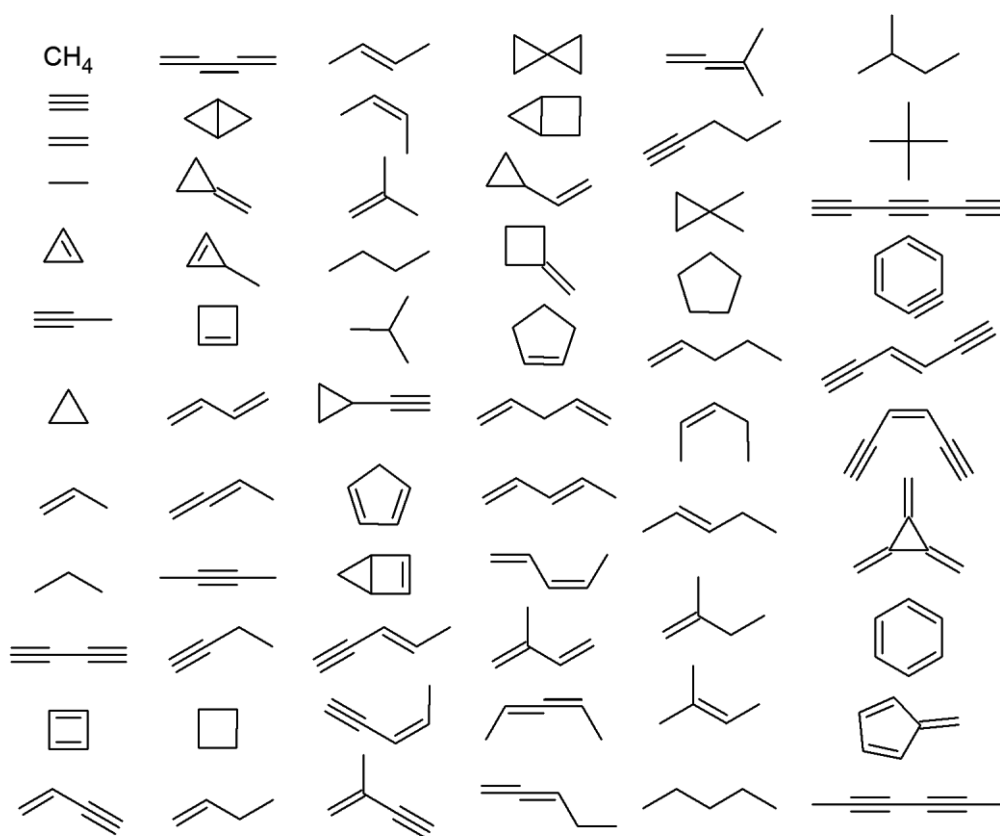
Figure 1. Test set of 65 hydrocarbon isomers from CH_4 to C_6H_6. The B3LYP/6-31G(2df,p) geometries were taken from the NIST Computational Chemistry Comparison and Benchmark Database.[51]

used Gaussian-type basis set is Dunning's correlation consistent aug-cc-pVXZ split valence atomic basis set family (noted here as AXZ) characterized by the X cardinal number. More complete basis sets are characterized with larger cardinal numbers, increasing computational times, and convergent series of energies or other properties.[38,39]

The Hartree–Fock (HF) energy converges exponentially with respect to the increase of the cardinal number X.[40,41] A two point extrapolation form of eq 5 in ref 41 yields HF energy limits with an RMS error of 0.01 millihartree using the A5Z and A6Z basis sets (noted as CBS(6/5)). It was also found that the inverse fifth power form is close to optimal for the so-called W1 calculations.[41]

The Møller–Plesset second order perturbation (MP2) correlation energies converge much slower with the cardinal number.[42] Feller, Peterson, and Hill compared the exponential, mixed exponential/Gaussian, inverse cubic[42] and shifted quartic[43] CBS schemes for the coupled cluster correlation energies.[44] For the random-phase approximation (RPA) correlation energy, Eshuis and Furche used the inverse cubic formula.[45] Fabiano and Della Sala examined shifted inverse cubic or quartic formulas and inverse power functions for the basis set extrapolation of RPA correlation. At the CBS(7/6) and CBS(6/5) levels, all these approaches yield similar results.[46] For intermolecular correlation energy in dimers, the counterpoise (CP) correction might improve the basis set convergence (but its application for systems of more than two entities is problematic).[47] For a detailed explanation about the errors and for further information about the development of CP corrections, see refs 48 and 49. Recently, it was shown that

CP corrections yield poor results for anion-π interaction energies calculated with MP2.[50]

COMPUTATIONAL DETAILS

We examine the basis set convergence of the dRPA correlation energies on a set of 65 hydrocarbon isomers (Figure 1) from CH_4 to C_6H_6. The molecular geometries are optimized with B3LYP/6-31G(2df,p) and taken from the NIST Computational Chemistry Comparison and Benchmark Database.[51]

The dRPA correlation energy (dRPAc) is given by eq 1

$$E_c^{dRPA} = \frac{1}{2}\text{tr}[\boldsymbol{BT}] \tag{1}$$

where $\boldsymbol{B}$ is the nonantisymmetrized two electron repulsion matrix, and $\boldsymbol{T}$ is the double excitation amplitude matrix. The dRPAc[52] is calculated using the efficient algorithm of Heßelmann[53] which is based on the CC-like form (by using antisymmetrized two-electron integrals in the Riccati equation) of the dRPA equations. The equations are solved in an $O(n^4)$-scaling iterative procedure by the density-fitted (DF) form of electron repulsion integrals and Cholesky decomposition of the orbital energy denominators. Recently an efficient linear-scaling implementation of the dRPA is applied for molecules with more than 1000 atoms and 10 000 basis functions.[54]

We calculate the iterative DF-dRPAc energies on self-consistent HF orbitals, called as dRPAc:HF, using the MRCC[55] (30/12/2014 release) program. We apply Dunning's correlation consistent aug-cc-pVXZ basis sets (noted here as AXZ, where X = 3, 4, 5, 6) with aug-cc-pVXZ-RI-JK and aug-cc-pVXZ-RI auxiliary basis sets for the self-consistent field (SCF) and correlation calculations, respectively. We call these

calculations dRPAc:HF/AXZ or for short dRPAc/AXZ. (We used RI approximation instead of RI-JK for A6Z and consider these results as $X = 5.99$ based on the difference between the RI and RI-JK approximations for small molecules such as methane, ethyne, ethene, and ethane.) We note that we excluded double-ζ basis sets from this study because the random errors of such popular basis sets are disturbingly large, and extrapolations that include such basis sets are considerably less reliable.[43]

We use the inverse cubic function (eq 2)[43] suggested in the literature for dRPAc energy extrapolation.[45] In addition, we test the exponential decay (eq 3) function and the generalized inverse power function (eq 4)[43]

$$E_{CBS}^{dRPAc} = E_{AXZ}^{dRPAc} - AX^{-3} \tag{2}$$

$$E_{CBS}^{dRPAc} = E_{AXZ}^{dRPAc} - Ae^{-nX} \tag{3}$$

$$E_{CBS}^{dRPAc} = E_{AXZ}^{dRPAc} - A(X + d)^{-\alpha} \tag{4}$$

where A, n, and α are optimized parameters, and d is a shift parameter that modifies the cardinal numbers. For the simple inverse power fit, $d = 0$. We define the average fitting error as the average of the errors yielded by eqs 2–4 using optimized parameters A, n, d, α, and E_{CBS}^{dRPAc} (e.g., for eq 2 the error = $E_{CBS}^{dRPAc} + AX^{-3} - E_{AXZ}^{dRPAc}$), compared to the calculated E_{AXZ}^{dRPAc} energy values for every molecule in the test set, where $X = 3-6$. Notice that optimal α in eq 4 linearizes the E_{AXZ}^{dRPAc} vs $X^{-\alpha}$ function and yields the E_{CBS}^{dRPAc} as the intercept and A as the slope (*vide infra*).

Eqs 2–4 are one-point extrapolation formulas. Here, we apply a two-point extrapolation formula for the dRPAc energy extrapolation similar to the two-point extrapolations applied in refs 56 and 57:

$$E_{CBS(X,X-1)}^{dRPAc} = E_{AXZ}^{dRPAc} + C_{X,X-1}^{dRPAc}(E_{AXZ}^{dRPAc} - E_{A(X-1)Z}^{dRPAc}) \tag{5}$$

The $C_{X,X-1}^{dRPAc}$ extrapolation coefficients can be calculated from eq 4 ($d = 0$) as follows:[56,57]

$$C_{X,X-1}^{dRPAc} = \frac{1}{\left(\frac{X}{X-1}\right)^{\alpha} - 1} \tag{6}$$

■ RESULTS AND DISCUSSION

The average fitting errors for dRPAc energies for 65 hydrocarbons using cardinal numbers $X = 3$, 4, 5, 6 (see Table S1 for the energies) and eqs 4–6 are shown in Figure 2. As the E_{AXZ}^{dRPAc} energy values predicted by the inverse cubic formula do not fit well to the calculated E_{AXZ}^{dRPAc} energy values, for $X = 3-6$, systematic errors occur. The predicted and calculated energies agree relatively well for $X = 3$ and 5 but not for $X = 4$ (large, systematic negative deviations) and $X = 6$ (large, systematic positive deviations). This behavior clearly shows that the inverse cubic form is not suitable for accurate extrapolation of dRPAc energies. Thus, the inverse cubic function form yields very large and oscillating average fitting errors (over 1 kcal mol^{-1}) and large individual fitting errors (up to almost 2 kcal mol^{-1}) for cardinal numbers 4 and 6, respectively (cf. Figure 2). Figure 2 shows that the fitted exponential function form, eq 3, is considerably better. Clearly the best performance is shown by the optimized inverse power functions, eq 4 with average $\alpha = 2.5608$, as the predicted energies are 100-times more accurate than those of calculated from the inverse cubic formulas, eq 2 as shown in Figure 2. Notice that the small average fitting errors for eq 6 are around
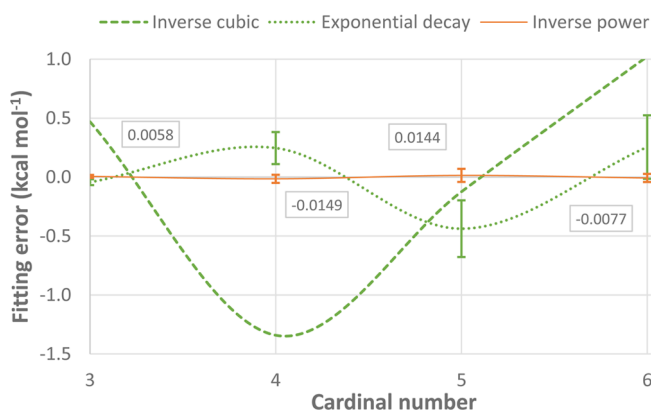

Figure 2. The average fitting errors of the inverse cubic function (eq 2), the exponential decay (eq 3), and the inverse power function (eq 4, $d = 0$) for the dRPAc energies calculated with aug-cc-pVXZ basis sets ($X = 3-6$) for 65 hydrocarbon isomers in Figure 1. The 95% confidence intervals are also shown for the exponential decay eq 3, and the inverse power function (eq 4, $d = 0$). The inverse cubic function is so imprecise that the 95% confidence interval is around ±0.4 kcal mol^{-1}. We used the RI approximation instead of the RI-JK for A6Z and consider these results as $X = 5.99$. This is based on the energy difference between the RI and RI-JK approximations for small molecules (methane, ethyne, ethene, and ethane). (The very small average fitting errors of the inverse power function are shown numerically.)

0.01 kcal mol^{-1}, thus they are practically invisible in Figure 2; this is why we show the numerical values.

The inverse power function ($d = 0$ in eq 4) performs similarly to its shifted versions ($d = \pm 0.5$ in eq 4), in agreement with the results in Figure 4 of ref 43. The average fitting errors for shifted versions compared to the inverse power fitting errors are shown in Figure S1 of the Supporting Information. The best fitting can be obtained by $d = 0.165$ in eq 4. We show these results in Table S2 and Figures S3–S5. The inverse power function form fits particularly well for the ATZ and AQZ energies (cf. Figure 2, $X = 3$ and 4), so the CBS energy can be estimated from these two energies with $C_{4,3}^{dRPAc:HF} = 0.9183$ from eq 6.

Detailed analysis shows that the extrapolation coefficient in eq 5 and the α exponent in eq 6 are slightly different for every molecule in the test set (cf. Supporting Information Figures S2 and S3). Figure S4 shows the calculated dRPA:HF correlation energies for ethyne as a function of X^{-3} and as a function $X^{-2.529}$ for $X = 3$, 4, 5, and 6. The nonlinearity shown for the inverse cubic formula (eq 2) leads to poorly extrapolated CBS energy. Approximating the parabolic function by a linear fit for the inverse cubic formula causes systematic negative deviations for $X = 4$ and positive deviations for $X = 6$ as shown in Figure 2. The optimal $\alpha = -2.529$ for ethyne molecule linearizes the correlation energy function vs $X^{-\alpha}$ and yields more consistently and accurately extrapolated CBS energy (cf. Figures 2 and S4). Using the inverse cubic formula will yield different CBS(4/3), CBS(5/4), and CBS(6/5) extrapolated dRPA correlation energies, while the optimal α yields practically equal extrapolated correlation energies independent of the cardinal numbers used for the extrapolation.

Notice that the dRPAc:HF energy of free C atoms converges with the X as $\alpha = 2.303$ in eq 6. We observed that the optimal exponent for methane and H_2, $\alpha = 2.66$ and 2.70, respectively, are considerably larger than the optimal exponent for 1,3,5-hexatriyne ($\alpha = 2.495$) or C atom. Notice that the H:C ratio for

methane is 4:1 and for 1,3,5-hexatriyne is 1:3. While the dRPAc energy for 1,3,5-hexatriyne converges significantly slower than that of the methane, it converges significantly faster than that of the free C atom. The origin of this behavior might be the considerable sensitivity of the E_{AXZ}^{dRPAc} on the cusp error. For less compact electron densities in the hydrogen rich molecules, the basis set convergence might be faster as the cusp error might be smaller for these compounds compared to the heavy atom rich compounds. This observation also shows that the basis set convergence for atoms is not transferable to molecules without significant loss of accuracy. Analysis of Figures S2 and S3 yields that increasing the number of C atoms in the molecule slows down the basis set convergence by decreasing the value of α. We have found that the exponents for eq 4 ($d = 0$) can be estimated from the number of carbon and hydrogen atoms in the molecule (eq 7 with the parameters of Table 1) or from the

Table 1. Average and Atomic Parameters Values for Estimated Inverse Power Exponents of the Basis Set Convergence, Energy in Eqs 7 and 8

	average		atomic contributions (P_{atom}^A)		hybridization state contributions ($P_{hybr}^{A(state)}$)
m	2.5608	m_{atom}	2.6184	m_{hybr}	2.5508
		P_{atom}^H	0.0268	P_{hybr}^H	0.1663
		P_{atom}^C	−0.1829	$P_{hybr}^{C(sp)}$	−0.1393
				$P_{hybr}^{C(sp^2)}$	−0.1852
				$P_{hybr}^{C(sp^3)}$	−0.2972

hybridization states of the carbon atoms (eq 8 with the parameters of Table 1)

$$\alpha \approx m_{atom} + \frac{N^H}{N^H + N^C}P_{atom}^H + \frac{N^C}{N^H + N^C}P_{atom}^C \quad (7)$$

$$\alpha \approx m_{hybr} + \frac{N^H}{N^H + N^C}P_{hybr}^H + \frac{N^{C(sp)}}{N^H + N^C}P_{hybr}^{C(sp)}$$
$$+ \frac{N^{C(sp^2)}}{N^H + N^C}P_{hybr}^{C(sp^2)} + \frac{N^{C(sp^3)}}{N^H + N^C}P_{hybr}^{C(sp^3)} \quad (8)$$

where α is the exponent in eq 6, m and P are fitting parameters, and N is the number of the specific atoms (e.g., C atoms in a specific hybrid state). In the test set of 65 hydrocarbon molecules we have 450 H atoms, 48 C(sp), 110 C(sp²), and 133 C(sp³) atoms, altogether 291 C atoms. Notice that the classification of the C atoms according to their hybrid states is not always straightforward. Inspection of the parameter values in Table 1 suggests that it is enough to distinguish the C(sp³) atoms from all the rest of the C atoms (eq 9), as the parameters for C(sp) and C(sp²) agree quite well. This leads to the following simpler equation

$$\alpha \approx m_{hybr} + \frac{N^H}{N^H + N^C}P_{hybr}^H + \frac{N^C}{N^H + N^C}P_{hybr}^C$$
$$+ \frac{N^{C(sp^3)}}{N^H + N^C}\Delta P_{hybr}^{C(sp^3)} \quad (9)$$

where m_{hybr} and P_{hybr}^H can be found in Table 1; the weighted average of the C(sp) and C(sp²) parameters can be used for the carbon atoms, $P_{hybr}^C = (48P_{hybr}^{C(sp)} + 110P_{hybr}^{C(sp^2)})/158 = −0.1712$, with the C(sp³) correction of $\Delta P_{hybr}^{C(sp^3)} = P_{hybr}^{C(sp^3)} - P_{hybr}^C = −0.1260$.

Inspection of the results shows that increasing the ratio of H atoms increases the basis set convergence, increasing the ratio of C atoms slows that down, and this effect is particularly large for C(sp³) atoms in the molecules. The exponent calculated for 1,3,5-hexatriyne that contains 6 C(sp) and only two H atoms is the smallest, eqs 7 and 8 give 2.488 in good agreement with the optimized $\alpha = 2.495$ given above. For methane eqs 7−9 do not work that well as the optimal exponent is $\alpha = 2.653$, but the predicted exponents are 2.603 and 2.624 for eqs 7 and 8, respectively (see also Figure 3). From the parameter values in
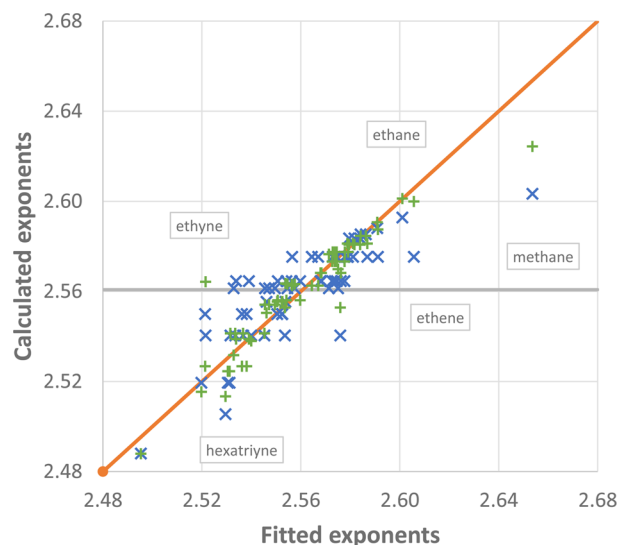

Figure 3. The correlation between the calculated and the fitted exponents of the power functions describing the basis set convergence of dRPAc:HF energy for the 65 hydrocarbon molecules shown in Figure 1. Legend: Average means that we use the average $\alpha = 2.5608$ exponent in eq 6. Atomic contribution (x) means that we calculate the value of α for every molecule according to eq 7 using the parameters from Table 1. Hybridization states (+) means that we calculate the value of α for every molecule according to eq 8 using the parameters from Table 1.

Table 1, we can predict the α values for larger saturated hydrocarbons. For C_8H_{18}, $C_{10}H_{22}$, and $C_{12}H_{26}$ we obtain $\alpha = 2.574$, 2.572, and 2.571, respectively, from hybrid state contributions, and we obtain very similar values from atomic contributions. For large saturated linear hydrocarbons, as the C:H ratio converges to 1:2, the α converges to 2.563.

The correlation between the calculated and the fitted exponents is shown in Figure 3. Inspection of the figure shows that consideration of the hybridization states (cf. eq 8 and Table 1) leads to better agreement (see also Figures S2 and S3) between the calculated and fitted exponents. The agreement is poorest for methane as discussed before, and this error is well visible in Figure 3. The two other problematic predictions are for ethyne and ethene.

Our present results show that the exponents working well for atoms or small molecules are not well applicable for dRPAc of larger molecules. In our examples the basis set convergence of large hydrocarbons is slower than that for methane, thus the dRPAc/CBS(4/3) energies that use $\alpha = 2.654$ for such large molecules might be in serious endothermic error. For CO and C_2 molecules we obtained $\alpha = 2.188$ and 2.348, respectively. Notice that this latter value is not far from the value of $\alpha = 2.303$ for the C atom shown above. For NH_3, H_2O, and HF

molecules we obtained similar α = 2.235, 2.203, and 2.189, respectively. These exponents are much smaller than the exponents that control the basis set convergence for larger hydrocarbons (see above α around 2.563). Application of such small inverse power exponents for large hydrocarbons would lead to very large exothermic CBS extrapolation error. These examples clearly show that the basis set convergence is sensitive to the molecular structure, but this sensitivity decreases with the increase of the size of the molecules. While for small molecules, like methane, water, ammonia, HF, N_2, CO, and C_2, the optimal α values are very different (vary from 2.18 to 2.65), for the larger molecules the optimal α values are similar and predictable (cf. eqs 7–9 and see also Figures S2a, S2b, and S3). Notice also that all the dRPAc energies in this study converge considerably slower with the cardinal number than the inverse cubic formula suggested in the literature.[6,11]

Next we analyze how the basis set convergence is influenced by the reference orbitals for the dRPA correlation energy calculation. Replacing the HF orbitals by PBE orbitals might improve the basis set convergence of the E_{AXZ}^{dRPAc}, as the potentially less compact PBE molecular electron density might lead to smaller cusp error. For methane we obtained α = 2.80 vs 2.65 for PBE vs HF reference orbitals, respectively ($C_{CBS(4/3)}^{dRPAc:PBE}$ = 0.807 and $C_{CBS(4/3)}^{dRPAc:HF}$ = 0.876). In our previous paper for PBE, PBE0.25, and HF reference orbitals applied for ethyne and ethene we obtained $C_{CBS(4/3)}^{dRPAc:PBE}$ = 0.856, $C_{CBS(4/3)}^{dRPAc:PBE0.25}$ = 0.867, and $C_{CBS(4/3)}^{dRPAc:HF}$ = 0.917.[57] For the more compact electron density of N_2 the basis set convergence is slower but less dependent on the reference orbitals, leading to α = 2.20 vs 2.18 for PBE vs HF reference orbitals, respectively. For this latter molecule the difference between the basis set convergences is practically negligible (cf. $C_{CBS(4/3)}^{dRPAc:HF}$ = 1.144 and $C_{CBS(4/3)}^{dRPAc:PBE}$ = 1.132). Using the published RPAc energies calculated with SCF PBE orbitals for the N_2 molecule (cf. Table 1 in ref 11), we obtained α = 2.08, leading to a slightly slower basis set convergence.

We have also analyzed the basis set convergence for the MP2 correlation energy. We have selected methane and 1,3,5-hexatriyne molecules for the analysis and obtained α = 2.88 and 2.83, respectively. Our results show that the basis set convergence of the correlation energy is considerably faster for the MP2 correlation energy than for the dRPAc energy, and these power values are much closer to the α = 3 used extensively in the literature.[1,5,6,11] It can be noticed that the difference between the negative power values for different molecules is also much smaller for MP2 than for dRPA.

Next we analyze the improvements yielded by applying eqs 7 and 8 over the simple average α for CBS extrapolation of the dRPAc energies. The individual deviations for the simple average α = 2.5608 are calculated as a difference: $E_{CBS(X/(X-1))}^{dRPAc}$ (α = 2.5608) − E_{CBS}^{dRPAc} (α = optimal). The individual deviations for the predicted α-s given by eqs 7 or 8 are calculated as a difference: $E_{CBS(X/(X-1))}^{dRPAc}$ (α = predicted) − E_{CBS}^{dRPAc} (α = optimal). Figure 4 shows that the mean absolute deviation (MAD) of the dRPAc/CBS(4/3) calculated with the average α = 2.5608 is equal to 0.40 kcal mol^{-1}. The application of atomic corrections in eq 7 almost halves this deviation to MAD = 0.21 kcal mol^{-1} as shown in Figure 4. The MAD is further reduced by considering the hybrid states in eqs 8 (MAD = 0.11 kcal mol^{-1}) and 9 (MAD = 0.15 kcal mol^{-1}). The improvements for dRPAc:HF/CBS(5/4) extrapolation caused by the use of eqs 7 and 8 are well visible in Figure 4. The MAD = 0.22 kcal mol^{-1} for dRPAc:HF/CBS(5/4) is reduced to 0.10 kcal mol^{-1} by applying eq 7, and it is further reduced to 0.08 kcal mol^{-1} by
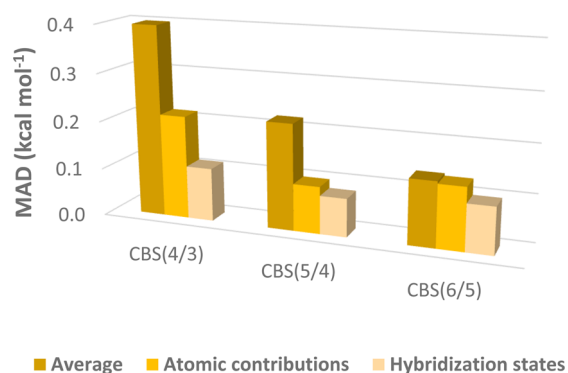

Figure 4. The mean absolute deviations of the predicted from the fitted complete basis set extrapolated dRPAc:HF/CBS(X/(X-1)) energies, where X = 4, 5, 6 for the 65 hydrocarbon molecules shown in Figure 1. Legend: Average means that we use the average α = 2.5608 exponent in eq 6. Atomic contribution means that we calculate the value of α for every molecule according to eq 7 using the parameters from Table 1 and use this exponent in eq 6. Hybridization states means that we calculate the value of α for every molecule according to eq 8 using the parameters from Table 1 and use this exponent in eq 6. The $C_{X,X-1}^{dRPAc}$ extrapolation coefficients obtained this way are used in eq 5 to obtain dRPAc:HF/CBS(X/(X-1)) energies.

applying eq 8. For CBS(5/4) extrapolation the application of the hybrid corrections (eq 8 or 9) does not lead to significant further improvements over the atomic corrections (eq 7). For dRPAc:HF/CBS(6/5) extrapolation the application of the average α yields MAD = 0.13 kcal mol^{-1}. Application of the eqs 7 and 8 does not improve significantly the results, yielding MAD = 0.13 kcal mol^{-1} for eq 7 with MAD = 0.10 kcal mol^{-1} for eqs 8 and 9. The main conclusion is that the accuracy and precision of the dRPA/CBS(4/3) extrapolation can be considerably improved by the application of eq 8, and CBS(6/5) quality results can be reached from relatively quick A3Z and A4Z calculations. This is a huge speed up and increase of the efficiency of the method. The results in Figure 4 show that the CBS(4/3) extrapolation can be very efficiently corrected, while the CBS(6/5) extrapolation does not. This means that the energies calculated with smaller, less saturated A3Z or A4Z basis sets depend considerably on the molecular structure, while the energies calculated with larger, much more saturated A5Z and A6Z basis sets are quite independent of the molecular structure, thus the empirical correction leads to limited improvements, and the average exponent works reasonably well. Notice also that the error of the density fitting is also included in these calculations which also limits the accuracy.

Figure 5 presents the detailed statistical analysis of the improvements yielded by eqs 7–9: the minimal and maximal deviations from the mean deviation (MD), the corrected sample standard deviations (CSSD) multiplied by 3, and the MD values of the two-point CBS(4/3), CBS(5/4), and CBS(6/5) extrapolated dRPAc:HF energies for the test set of 65 hydrocarbons. Figure 5 shows that the application of the average α exponent gives an asymmetric error distribution. The largest positive deviation is observed for 1,3,5-hexatriyne (C_6H_2). The values of those deviations are 1.62 and 1.28 kcal mol^{-1} for CBS(4/3) and CBS(5/4), respectively (cf. Figure 5). These errors are very efficiently decreased to almost zero by atomic (eq 7) or hybridization (eq 8 correction of the exponent. Figure 5 shows that application of eqs 7 and 8 leads to a symmetric error distribution with greatly reduced 3CSSD
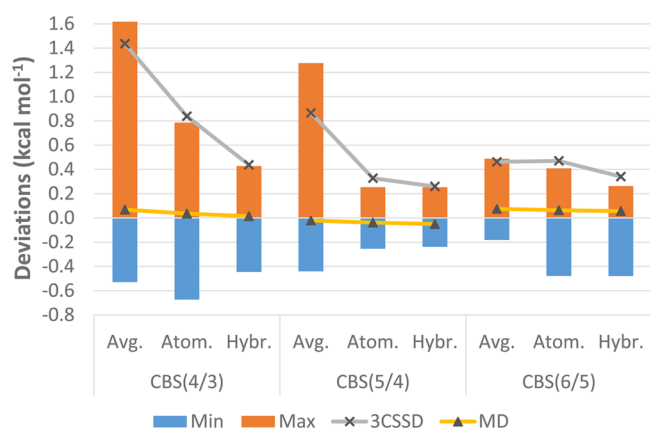
Figure 5. The minimum, maximum, mean (MD), and corrected sample standard deviations multiplied by 3 (3CSSD) of the two-point CBS(4/3), CBS(5/4), and CBS(6/5) extrapolated dRPAc:HF energies for the test set of 65 hydrocarbons. Legend: Avg. means that we use the average $\alpha = 2.5608$ exponent in eq 6. Atom means that we calculate the value of α for every molecule according to eq 7 using the parameters from Table 1 and use this exponent in eq 6. Hybr. means that we calculate the value of α for every molecule according to eq 8 using the parameters from Table 1 and use this exponent in eq 6. The $C_{X,X-1}^{dRPAc}$ extrapolation coefficients obtained this way are used in eq 5 to obtain dRPAc:HF/CBS($X/(X-1)$) energies.

(down from 1.44 to 0.44 kcal mol^{-1}). Figure 5 shows that application of the hybridization corrected exponents via eq 8 improves the dRPAc/CBS(4/3) energy extrapolation in every statistical aspect and makes it practically equivalent to the very expensive CBS(6/5) energy extrapolation for the present test set (see also Figures S5 and S6). Figure 5 also shows that for dRPAc/CBS(5/4) extrapolation the atomic correction of the exponents is sufficient (eq 7), and no further improvements can be obtained from hybridization corrections (eqs 8 or 9). For dRPAc/CBS(6/5) extrapolation, the atomic and hybridization corrections of the exponents yield only small but well appreciable improvements as shown in Figure 5.

For practical CBS(6/5) extrapolation of the dRPAc energy we suggest using the average $\alpha = 2.5608$ exponent in eq 6 for larger hydrocarbons. For CBS(5/4) extrapolations we suggest using the α values calculated from eq 7 in eq 6. For CBS(4/3) extrapolations we suggest using the α values calculated from the eqs 8 or 9 in eq 6.

CONCLUSION

Accurate and precise calculation of the dRPA energy requires very large basis sets, as observed earlier in the literature. In this paper, we have selected a test set of 65 hydrocarbon molecules in order to provide insight into the basis set convergence problems. The DF-dRPA correlation (dRPAc) energy is calculated using an efficient algorithm which is based on the CC-like form by using antisymmetrized two-electron integrals in the Riccati equations. These equations are solved in an iterative procedure by the density-fitted form of electron repulsion integrals and Cholesky decomposition of the orbital energy denominators ($O(n^4)$-scaling).

The accurate CBS extrapolation of the HF energy is given in the literature. In this paper, we focus on the CBS extrapolation of dRPAc energies using Dunning's correlation consistent aug-cc-pVXZ basis sets (noted here as AXZ) with aug-cc-pVXZ-RI auxiliary basis sets for the correlation calculations ($X = \{3, 4, 5, 6\}$), respectively, where X is the cardinal number of the basis

set. We use the self-consistent HF wave function for dRPAc energy calculation noted as dRPA:HF.

The dRPAc energy converges slowly, and even the dRPAc/A6Z energies are far from being converged. We have tested 3 functions for the complete basis set (CBS) energy extrapolation of dRPAc energy with respect to the X cardinal number of the AXZ basis sets. We have found that the popular inverse cubic extrapolation formula fits poorly to dRPAc/AXZ energies, giving well over 1 kcal mol^{-1} error for the CBS extrapolated energies for the tested molecules. The optimized exponential extrapolation gives a considerably better fit. However, the best fit was obtained for an inverse power formula with optimized exponent. We tested the accuracy and the precision of the dRPAc/CBS energies for an averaged inverse power exponent $\alpha = 2.5608$.

We observed that the C atom and the hydrocarbon molecules show different convergence for dRPAc, and a simple atomic correction of the inverse power exponent leads to considerable improvement for the dRPAc/CBS(4/3) or CBS(5/4) energy extrapolations. The dRPAc/CBS(4/3) energies improved further by considering the hybrid states of the C atoms in the molecules: the mean absolute deviation (MAD) obtained by the averaged inverse power exponent is reduced to one-third (MAD = 0.13 kcal mol^{-1}). The other statistical parameters, like minimum and maximum deviations from the mean deviation and the corrected sample standard deviations, improve greatly by considering atomic hybridization states in the molecules. dRPAc/CBS(4/3) energies that account for the hybridization information in the molecule via eq 8 reach the accuracy and precision of the very expensive dRPAc/CBS(6/5) energies that use the averaged inverse power exponent $\alpha = 2.5608$.

Our results show that basis set convergence of the dRPAc energy also depends on the reference orbitals. In general, using PBE determinant instead of the HF determinant speeds up the basis set convergence for molecules containing a large fraction of hydrogen atoms. A large fraction of carbon or nitrogen atoms slows down the basis set convergence and also diminishes the difference between the application of the HF or PBE determinants. The MP2 correlation energy converges considerably faster with the basis set quality increase, and the difference between the molecules is considerably smaller in this respect. For methane and 1,3,5-hexatriyne molecules, we obtained exponents that are close to the widely used inverse cubic.

Our results suggest that A3Z and A4Z basis sets have some atom specific convergence properties in the dRPAc/CBS energy extrapolations and that can be corrected by the method proposed here. Possibly redesigned AXZ basis sets also could solve this problem. We noticed that atomic or hybrid corrections of the inverse power exponent do not improve considerably the dRPAc/CBS(6/5) energies, as these A5Z and A6Z basis sets are considerably more complete than A3Z. Our results also show that the optimized exponents that yield accurate results for atoms or small molecules might be inaccurate for larger molecules, as the basis set convergence depends on the size and structure of the molecules for dRPAc/CBS(4/3) extrapolations. For practical CBS(6/5) extrapolation of the dRPAc energy of hydrocarbons we suggest using the average $\alpha = 2.5608$ exponent in eq 6. For CBS(5/4) extrapolations we suggest using the α values calculated from eq 7 in eq 6. For CBS(4/3) extrapolations we suggest using the α values calculated from eqs 8 or 9 in eq 6.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

Figures are provided for the average fitting errors calculated with aug-cc-pVXZ basis sets ($X = 3-6$) for 65 hydrocarbon isomers, the actual and the predicted exponents of the inverse power function, for the correlation between the predicted and the fitted exponents of the slightly shifted ($d = 0.165$) power function, and for the mean absolute deviations of the predicted from the fitted complete basis set extrapolated dRPA correlation energies using the slightly shifted ($d = 0.165$) inverse power function formula. Tables are provided for the atomic compositions, the hybrid states, and for the dRPAc energies calculated with HF SCF wave function using AXZ basis set, where $X = 3-6$, and the extrapolated CBS energies, and for the average and atomic parameters values for estimated slightly shifted ($d = 0.165$) inverse power function exponents. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00269.

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: csonkag@mail.bme.hu.
*E-mail: mezei.pal@mail.bme.hu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Furche, F. *J. Chem. Phys.* **2008**, *129*, 114105.
(2) Kubo, R. *Rep. Prog. Phys.* **2002**, *29*, 255−284.
(3) Scuseria, G. E.; Henderson, T. M.; Sorensen, D. C. *J. Chem. Phys.* **2008**, *129*, 231101.
(4) Eshuis, H.; Bates, J. E.; Furche, F. *Theor. Chem. Acc.* **2012**, *131*, 1084.
(5) Eshuis, H.; Furche, F. *J. Phys. Chem. Lett.* **2011**, *2*, 983−989.
(6) Ren, X.; Rinke, P.; Scheffler, M. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2009**, *80*, 1−8.
(7) Schimka, L.; Harl, J.; Stroppa, A.; Grüneis, A.; Marsman, M.; Mittendorfer, F.; Kresse, G. *Nat. Mater.* **2010**, *9*, 741−744.
(8) Lebègue, S.; Harl, J.; Gould, T.; Ángyán, J. G.; Kresse, G.; Dobson, J. F. *Phys. Rev. Lett.* **2010**, *105*, 1−4.
(9) Harl, J.; Kresse, G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 1−8.
(10) Jiang, H.; Engel, E. *J. Chem. Phys.* **2007**, *127*, 1−11.
(11) Furche, F. *Phys. Rev. B* **2001**, *64*, 195120.
(12) Harl, J.; Schimka, L.; Kresse, G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *81*, 115126.
(13) Paier, J.; Ren, X.; Rinke, P.; Scuseria, G. E.; Grüneis, A.; Kresse, G.; Scheffler, M. *New J. Phys.* **2012**, *14*, 043002.
(14) Henderson, T. M.; Scuseria, G. E. *Mol. Phys.* **2010**, *108*, 2511−2517.
(15) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Science* **2008**, *321*, 792−794.
(16) Yan, Z.; Perdew, J. P.; Kurth, S. *Phys. Rev. B* **2000**, *61*, 16430−16439.
(17) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I. *J. Chem. Theory Comput.* **2010**, *6*, 127−134.
(18) Grüneis, A.; Marsman, M.; Harl, J.; Schimka, L.; Kresse, G. *J. Chem. Phys.* **2009**, *131*, 154115.
(19) Ren, X.; Tkatchenko, A.; Rinke, P.; Scheffler, M. *Phys. Rev. Lett.* **2011**, *106*, 153003.
(20) Heßelmann, A. *J. Chem. Phys.* **2011**, *134*, 204107.
(21) Paier, J.; Janesko, B. G.; Henderson, T. M.; Scuseria, G. E.; Grüneis, A.; Kresse, G. *J. Chem. Phys.* **2010**, *132*, 094103.
(22) Noga, J.; Kutzelnigg, W.; Klopper, W. *Chem. Phys. Lett.* **1992**, *199*, 497−504.
(23) Tew, D. P.; Klopper, W.; Helgaker, T. *J. Comput. Chem.* **2007**, *28*, 1307−1320.
(24) Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. *Chem. Rev.* **2012**, *112*, 4−74.
(25) Kong, L.; Bischoff, F. A.; Valeev, E. F. *Chem. Rev.* **2012**, *112*, 75−107.
(26) Kutzelnigg, W.; Klopper, W. *J. Chem. Phys.* **1991**, *94*, 1985.
(27) Klopper, W. *Mol. Phys.* **2001**, *99*, 481−507.
(28) Handy, N. C. *Mol. Phys.* **1971**, *21*, 817.
(29) Umezawa, N.; Tsuneyuki, S. *Phys. Rev. B* **2004**, *69*, 31−35.
(30) Ceperley, D. M.; Alder, B. J. *Phys. Rev. Lett.* **1980**, *45*, 566−569.
(31) Lester, W. A.; Mitas, L.; Hammond, B. *Chem. Phys. Lett.* **2009**, *478*, 1−10.
(32) Lüchow, A.; Anderson, J. B. *Annu. Rev. Phys. Chem.* **2000**, *51*, 501−526.
(33) Nguyen, H. V.; De Gironcoli, S. *Phys. Rev. B* **2009**, *79*, 1−12.
(34) Nguyen, H.-V.; Galli, G. *J. Chem. Phys.* **2010**, *132*, 044109.
(35) Nakatsuji, H.; Nakashima, H.; Kurokawa, Y.; Ishikawa, A. *Phys. Rev. Lett.* **2007**, *99*, 240402.
(36) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.
(37) Shepherd, J. J.; Grüneis, A.; Booth, G. H.; Kresse, G.; Alavi, A. *Phys. Rev. B* **2012**, *86*, 1−14.
(38) Kendall, R., Jr.; Dunnin, T.; Harrison, R. *J. Chem. Phys.* **1992**, *96*, 6796.
(39) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *99*, 1914−1929.
(40) Jensen, F. *Theor. Chem. Acc.* **2005**, *113*, 267−273.
(41) Karton, A.; Martin, J. M. L. *Theor. Chem. Acc.* **2006**, *115*, 330−333.
(42) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639.
(43) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243−252.
(44) Feller, D.; Peterson, K. A.; Hill, J. G. *J. Chem. Phys.* **2011**, *135*, 044102.
(45) Eshuis, H.; Furche, F. *J. Chem. Phys.* **2012**, *136*, 084105.
(46) Fabiano, E.; Della Sala, F. *Theor. Chem. Acc.* **2012**, *131*, 1278.
(47) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.
(48) Mentel, M.; Baerends, E. J. *J. Chem. Theory Comput.* **2014**, *10*, 252−267.
(49) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873−1885.
(50) Mezei, P. D.; Csonka, G. I.; Ruzsinszky, A.; Sun, J. *J. Chem. Theory Comput.* **2015**, *11*, 360−371.
(51) NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, Release 16a, August 2013, Johnson, R. D., III, Ed. http://cccbdb.nist.gov/ (accessed January 8, 2015).
(52) Kállay, M. *J. Chem. Phys.* **2014**, *141*, 244113.
(53) Heßelmann, A. *Phys. Rev. A* **2012**, *85*, 012517.
(54) Kállay, M. *J. Chem. Phys.* **2015**, *142*, 204105.
(55) Mrcc, a quantum chemical program suite written by Kállay, M.; Rolik, Z.; Csontos, J.; Ladjánszki, I.; Szegedy, L.; Ladóczki, B.; Samu, G.. See also Rolik, Z.; Szegedy, L.; Ladjánszki, I.; Ladóczki, B.; Kállay, M. *J. Chem. Phys.* **2013**, *139*, 094105, as well as www.mrcc.hu (accessed December 30, 2014).
(56) Csonka, G. I.; Kaminsky, J. *J. Chem. Theory Comput.* **2011**, *7*, 988−997.
(57) Mezei, P. D.; Csonka, G. I.; Kállay, M. *J. Chem. Theory Comput.* **2015**, *11*, 2879−2888.